Decision Theoretic Framework

MIT 18.655

Dr. Kempthorne

Spring 2016
1 Decision Theoretic Framework
   • I. Basic Elements of a Decision Problem
Decision Problems of Statistical Inference

- **Estimation**: estimating a real parameter \( \theta \in \Theta \) using data \( X \) with conditional distribution \( P_\theta \).
- **Testing**: Given data \( X \sim P_\theta \), choosing between two hypotheses (deciding whether to accept or reject \( H_0 \))
  \[
  H_0 : P_\theta \in \mathcal{P}_0 \text{ (a set of special } \mathcal{Ps})
  \]
  \[
  H_1 : P_\theta \not\in \mathcal{P}_0
  \]
- **Ranking**: rank a collection of items from best to worst
  - Products evaluated by consumer interest group
  - Sports betting (horse race, team tournament, division championship, etc.)
- **Prediction**: predict response variable \( Y \) given explanatory variables \( Z = (Z_1, Z_2, \ldots, Z_d) \).
  - If know joint distribution of \((Z, Y)\), use \( \mu(Z) = E[Y \mid Z] \)
  - With data \( \{(z_i, y_i), i = 1, 2, \ldots, n\} \), estimate \( \mu(Z) \).
    - If \( \mu(Z) = g(\beta, Z) \), then use \( \hat{\mu}(Z) = g(\hat{\beta}, Z) \)
Basic Elements of a Decision Problem

\( \Theta = \{ \theta \} : \) The “State Space”
- \( \theta = \) state of nature (unknown uncertainty element in the problem)

\( A = \{ a \} : \) The ”Action Space”
- \( a = \) action taken by statistician

\( L(\theta, a) : \) The “Loss Function”
- \( L(\theta, a) = \) loss incurred when state is \( \theta \) and action \( a \) taken
- \( L : \Theta \times A \rightarrow R \)

Example: Investing money in an uncertain world
- \( \Theta = \{ \theta_1, \theta_2 \} \) where \( \theta_1 = \) good economy/market
  \( \theta_2 = \) bad economy/market
- \( A = \{ a_1, a_2, \ldots, a_5 \} \) (different investment programs)
I. Basic Elements of a Decision Problem

**Loss function:**

\[
L(\theta, a) : \begin{array}{c|c|c|c|c|c}
 & a_1 & a_2 & a_3 & a_4 & a_5 \\
\hline
\theta_1 \text{ (good economy)} & -4 & -4 & -1 & 2 & 4 \\
\theta_2 \text{ (bad economy)} & 4 & 0 & -1 & -6 & -4 \\
\end{array}
\]

Note:
- \(a_1\) does well in good market (negative loss)
- \(a_5\) does well in bad market (negative loss)
- \(a_3\) gains in either market (e.g., risk-free bond)

**Problem:** How to choose among investments?
Additional Elements of a Statistical Decision Problem

\( X \sim P_\theta \): Random Variable (Statistical Observation)
- Conditional distribution of \( X \) given \( \theta \)
- Sample space \( \mathcal{X} = \{x\} \)
- Density/pmf function of conditional distribution:
  \( f(x \mid \theta) \) or \( f_X(x \mid \theta) \)

\( \delta(X) \): A “Decision Procedure”
- Observe data \( X = x \) and take action \( a \in A \)
- \( \delta(\cdot): \mathcal{X} \to A \).

\( \mathcal{D} \): Decision Space (class of decision procedures)
- \( \mathcal{D} = \{ \text{decision procedures } \delta: \mathcal{X} \to A \} \)

\( R(\theta, \delta) \): Risk Function (performance measure of \( \delta(\cdot) \mid \theta \))
- \( R(\theta, \delta) = E_X[L(\theta, \delta(X)) \mid \theta] \)
- Expectation of loss incurred by decision procedure \( \delta(X) \) when \( \theta \) is true.
- For no-data problem (no \( X \)), \( R(\theta, a) = L(\theta, a) \)
Statistical Estimation Problem

- $X \sim P_{\theta} = N(\theta, 1), \ -\infty < \theta < \infty$.
- $A = \Theta = R$.
- Squared-error loss:
  $$L(\theta, a) = (a - \theta)^2$$
- Decision procedure: for finite constant $c : 0 < c \leq 1$
  $$\delta_c(X) = cX.$$  
- Risk function:
  $$R(\theta, \delta_c) = E_X[(\delta(X) - \theta)^2 | \theta]$$
  $$= \text{Var}(\delta(x)) + [E_X[\delta(x) | \theta] - \theta]^2$$
  $$= c^2 + (c - 1)^2 \theta^2$$

Special cases: consider $c = 1, 0, \frac{1}{2}$

- $\delta_1(X) = X : R(\theta, \delta_1) = 1$ (independent of $\theta$)
- $\delta_0(X) \equiv 0 : R(\theta, \delta_0) = \theta^2$ (zero at $\theta = 0$, unbounded)
- $\delta_{0.5}(X) = X/2 : R(\theta, \delta_{0.5}) = \frac{1}{4} \times (1 + \theta^2)$.

What about $\delta_c$ for $c > 1$? (or for $c < 0$)?
Mean-Squared Error: Estimation Risk (Squared-Error Loss)

- \( X \sim P_\theta, \theta \in \Theta \).
- Parameter of interest: \( \nu(\theta) \) (some function of \( \theta \))
- Action Space: \( A = \{ \nu = \nu(\theta), \theta \in \Theta \} \)
- Decision procedure/estimator: \( \hat{\nu}(X) : \mathcal{X} \to A \)
- Squared Error Loss: \( L(\theta, a) = [a - \nu(\theta)]^2 \)
- Risk equal to Mean-Squared Error:
  \[
  R(\theta, \hat{\nu}(X)) = E[L(\theta, \hat{\nu}(X)) \mid \theta]
  = E[(\hat{\nu}(X) - \nu(\theta))^2 \mid \theta] = MSE(\hat{\nu})
  \]

**Proposition 1.3.1** For an estimator \( \hat{\nu}(X) \) of \( \nu(\theta) \), the mean-squared error is

\[
MSE(\hat{\nu}) = Var[\hat{\nu}(X) \mid \theta] + [Bias(\hat{\nu} \mid \theta)]^2
\]

where \( Bias(\hat{\nu} \mid \theta) = E[\hat{\nu}(X) \mid \theta] - \nu(\theta) \)

**Definition:** \( \hat{\nu} \) is **Unbiased** if \( Bias(\hat{\nu} \mid \theta) = 0 \) for all \( \theta \in \Theta \).
Examples of Statistical Decision Problems

**Statistical Testing Problem** (Two-Sample Problem)

- $X_1, \ldots, X_m$ iid $N(\mu, \sigma^2)$, (response under control treatment)
- $Y_1, \ldots, Y_n$ iid $N(\mu + \Delta, \sigma^2)$ (response under test treatment)
- where $\mu \in R$, $\sigma^2 \in R_+$ unknown
- and $\Delta \in R$, is unknown treatment effect.

- Let $P(X, Y | \mu, \Delta, \sigma^2)$ denote the joint distribution of $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_n)$

- Define two hypotheses:
  \[ H_0 : P \in \{ P : \Delta = 0 \} = \{ P_{\theta}, \theta \in \Theta_0 \} \]
  \[ H_1 : P \in \{ P : \Delta \neq 0 \} = \{ P_{\theta}, \theta \notin \Theta_0 \} \]

- $A = \{0, 1\}$ with 0 corresponding to accepting $H_0$ and 1 to rejecting $H_0$. 
Statistical Testing Problem

- Construct decision rule accepting $H_0$ if estimate of $\Delta$ is significantly different from zero, e.g.,
  \[ \hat{\Delta} = \bar{Y} - \bar{X} \] (difference in sample means)
  \[ \hat{\sigma}: \text{an estimate of } \sigma \]

\[
\delta(X, Y) = \begin{cases} 
0 & \text{if } |\hat{\Delta}| < c \quad \text{(critical value)} \\
1 & \text{if } |\hat{\Delta}| \geq c
\end{cases}
\]

Apply decision theory to specify $c$ (and $\hat{\sigma}$)

- Zero-One Loss function

\[
L(\theta, a) = \begin{cases} 
0 & \text{if } \theta \in \Theta_a \quad \text{(correct action)} \\
1 & \text{if } \theta \notin \Theta_a \quad \text{(wrong action)}
\end{cases}
\]

- Risk function

\[
R(\theta, \delta) = L(\theta, 0)P_{\theta}(\delta(X, Y) = 0) + L(\theta, 1)P_{\theta}(\delta(X, Y) = 1)
\]
\[
= P_{\theta}(\delta(X, Y) = 1), \quad \text{if } \theta \in \Theta_0
\]
\[
= P_{\theta}(\delta(X, Y) = 0), \quad \text{if } \theta \notin \Theta_0
\]
Terminology of Statistical Testing

- Using r.v. $X \sim P_\theta$ with sample space $\mathcal{X}$ and parameter space $\Theta$, to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$

- **Critical Region** of a test $\delta(\cdot)$
  $$ C = \{ x : \delta(x) = 1 \} $$

- **Type I Error**: $\delta(X)$ rejects $H_0$ when $H_0$ is true
- **Type II Error**: $\delta(X)$ accepts $H_0$ when $H_0$ is false

- Risk under zero-one loss:
  $$ R(\theta, \delta) = P_\theta(\delta(X) = 1 \mid \theta), \text{ if } \theta \in \Theta_0 $$

  $$ = \text{Probability of Type I Error} $$

  and

  $$ R(\theta, \delta) = P_\theta(\delta(X) = 0 \mid \theta), \text{ if } \theta \notin \Theta_0 $$

  $$ = \text{Probability of Type II Error (function of } \theta \text{)} $$

- **Neyman-Pearson** framework:
  Constrained optimization of risks:
  Minimize: $P(\text{Type II Error})$
  subject to: $P(\text{Type I Error}) \leq \alpha$ ("significance level")
Interval Estimation and Confidence Bounds

**VAR: Value-at-Risk**

- Let $X_1, X_2, \ldots$ be the change in value of an asset over independent fixed holding periods and suppose they are i.i.d. $X \sim P_\theta$ for some fixed $\theta \in \Theta$.
- For $\alpha = 0.05$, say, define $\text{VAR}_\alpha$ (the level $-\alpha$ Value-at-Risk) by
  $$P(X \leq -\text{VAR}_\alpha | \theta) = \alpha$$
- Consider estimating the $\text{VAR}$ of $X_{n+1}$ given $X = (X_1, \ldots, X_n)$
  Determine an estimator $\widehat{\text{VAR}}(X)$:
  $$P_\theta(X \leq -\widehat{\text{VAR}}(X)) \leq \alpha, \text{ for all } \theta \in \Theta.$$ 
- The outcome $X_{n+1}$ exceeds $\text{VAR}_\alpha$ to the downside with probability no greater than $\alpha (= 0.05)$. 
Lower-Bound Estimation

- $X \sim P_{\theta}, \theta \in \Theta$.
- Parameter of interest: $\nu(\theta)$ (some function of $\theta$)
- Action Space: $A = \{\nu = \nu(\theta), \theta \in \Theta\}$
- Estimator: $\hat{\nu}(X) : \mathcal{X} \rightarrow A$
- Objective: bounding $\nu(\theta)$ from below
- Lower-Bound Estimator: $\hat{\nu}(X)$ is good if
  
  $P_{\theta}(\hat{\nu}(X) \leq \nu(\theta))$ has high probability
  $P_{\theta}(\hat{\nu}(X) > \nu(\theta))$ has low probability

  $\implies$ Define the loss function
  
  $L(\theta, a) = 1$, if $a > \nu(\theta)$; zero otherwise

- Risk function under zero-one loss $L(\theta, a)$:
  
  $R(\theta, \hat{\nu}(X)) = E[L(\theta, \hat{\nu}(X)) | \theta] = P_{\theta}(\hat{\nu}(X) > \nu(\theta))$.

- The Lower-Bound Estimator $\hat{\nu}(X)$ has Confidence Level $(1 - \alpha)$ if
  
  $P_{\theta}(\hat{\nu}(X) \leq \nu(\theta)) \geq 1 - \alpha$, for all $\theta \in \Theta$. 
Interval (Lower and Upper Bound) Estimation

- \( X \sim P_\theta, \theta \in \Theta. \)
- Parameter of interest: \( \nu(\theta) \) (some function of \( \theta \))
- Define \( \mathcal{V} = \{ \nu = \nu(\theta), \theta \in \Theta \} \)
- Objective: Interval estimation of \( \nu(\theta) \)
- Action Space: \( \mathcal{A} = \{ a = [a, \bar{a}] : a < \bar{a} \in \mathcal{V} \} \)
- Estimator: \( \hat{\nu}(X) : \mathcal{X} \rightarrow \mathcal{A} \)
  \[ \hat{\nu}(X) = [\hat{\nu}_{\text{LOWER}}(X), \hat{\nu}_{\text{UPPER}}(X)] \]
- Interval Estimator: \( \hat{\nu}(X) \) is good if
  \[ P_\theta(\hat{\nu}_{\text{LOWER}}(X) \leq \nu(\theta) \leq \hat{\nu}_{\text{UPPER}}(X)) \text{ is high} \]
  \[ P_\theta(\hat{\nu}_{\text{LOWER}}(X) > \nu(\theta) \text{ or } \hat{\nu}_{\text{UPPER}}(X) < \nu(\theta)) \text{ is low} \]

**NOTE:** \( \theta \) is non-random; the interval is random given \( \theta \).
We need Bayesian models to compute:
\[ P(\nu(\theta) \in [\hat{\nu}_{\text{LOWER}}(X), \hat{\nu}_{\text{UPPER}}(X)] | X = x) \]
Define the loss function
\[ L(\theta, (a, \bar{a})) = \begin{cases} 
1, & \text{if } a > \nu(\theta) \text{ or } \bar{a} < \nu(\theta) \\
0, & \text{otherwise.} 
\end{cases} \]

Risk function under zero-one loss \( L(\theta, a) \):
\[ R(\theta, \hat{\nu}(X)) = E[L(\theta, \hat{\nu}(X)) \mid \theta] \\
= P_\theta(\hat{\nu}_{\text{LOWER}}(X) > \nu(\theta) \text{ or } \hat{\nu}_{\text{UPPER}}(X) < \nu(\theta)) \\
= 1 - P_\theta(\hat{\nu}_{\text{LOWER}}(X) \leq \nu(\theta) \leq \hat{\nu}_{\text{UPPER}}(X) \mid \theta) \]

The Interval Estimator \( \hat{\nu}(X) \) has **Confidence Level** \((1 - \alpha)\) if
\[ P_\theta(\hat{\nu}_{\text{LOWER}}(X) \leq \nu(\theta) \leq \hat{\nu}_{\text{UPPER}}(X) \mid \theta) \geq (1 - \alpha) \text{ for all } \theta \in \Theta \]

Equivalently:
\[ R(\theta, \hat{\nu}(X)) \leq \alpha, \text{ for all } \theta \in \Theta. \]
Choosing Among Decision Procedures

Admissible/Inadmissible Decision Procedures

- On basis of performance measured by the Risk function $R(\theta, \delta)$, some rules obviously bad
- A decision procedure $\delta(\cdot)$ is inadmissible if $\exists \delta'$ such that $R(\theta, \delta') \leq R(\theta, \delta)$ for all $\theta \in \Theta$
  with strict inequality for some $\theta$.

Examples:
- In no-data investment problem: actions $a_1$ and $a_5$ are inadmissible
- In $N(\theta, 1)$ estimation problem: decisions $\delta_c(\cdot)$ with $c \notin [0, 1]$ are inadmissible

Objectives:
- Restrict $D$ to exclude inadmissible decision procedures
- Characterize “Complete Class” (all admissible procedures)
- Formalize ‘best’ choice amongst all admissible procedures
**Selection Criteria for Decision Procedures**

**Approaches to Decision Selection**

- Compare risk functions by global criteria
  - Bayes risk
  - Maximum risk (Minimax approach)

- Apply sensible constraint on the class of procedures:
  - Unbiasedness (estimators and tests)
  - Upper limit for level of significance (tests)
  - Invariance under scale transformations

E.g., Given $X \sim P_\theta$ where $\theta = E[X \mid \theta]$, if $\delta(X)$ is used to estimate $\theta$ then $\delta(\cdot)$ should satisfy

$$\delta(cX) = c\delta(X).$$

(same estimator applied if transform $X$ to $Y = cX$.)

See e.g., Ferguson (1967), Lehmann (1997)
Bayes Criterion for Selecting a Decision Procedure

Basic Elements of Decision Problem (as before)

\( X \sim P_\theta \): Random Variable (Statistical Observation)

- Distribution of \( X \) given \( \theta \) with sample space \( \mathcal{X} = \{x\} \)

\( \delta(X) \): A "Decision Procedure" \( \delta(\cdot) : \mathcal{X} \to \mathcal{A} \).

\( D \): Decision Space (class of decision procedures)

- \( D = \{\text{decision procedures } \delta : \mathcal{X} \to \mathcal{A}\} \)

\( R(\theta, \delta) \): Risk Function (performance measure of \( \delta(\cdot) | \theta \) )

- \( R(\theta, \delta) = E_X[L(\theta, \delta(X)) | \theta] \)

Additional Elements of Bayesian Decision Problem

\( \theta \sim \pi \): Prior Distribution for parameter \( \theta \in \Theta \).

\( r(\pi, \delta) \): Bayes Risk of \( \delta \) given prior distribution \( \pi \)

- \( r(\pi, \delta) = E_{\theta^*} R(\theta^*, \delta(X)) \),
  taking expectation with respect to \( \theta^* \sim \pi \).

Bayes rule \( \delta^* \): Decision procedure that minimizes the Bayes risk

\[ r(\pi, \delta^*) = \min_{\delta \in D} r(\pi, \delta) \]
Bayesian Decision Problem: Oil Wildcatter

Problem: An oil wildcatter owns rights to drill for oil at a location. He/she must decide whether to Drill, Sell the rights, or Sell partial rights.

State Space: $\Theta = \{\theta_1, \theta_2\}$
A location either contains oil ($\theta_1$) or not ($\theta_2$).

Action Space: $A = \{a_1(Drill), a_2(Sell), a_3(PartialRights)\}$

Loss Function: $L(\theta, a) : \Theta \times A \rightarrow R$ given by the following table:

<table>
<thead>
<tr>
<th>$\theta$ \ $a$</th>
<th>(Drill)</th>
<th>(Sell)</th>
<th>(Partial Rights)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>12</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>
Oil Wildcatter Problem

Random Variable: Rock formation $X \sim P_\theta$

- Sample Space: $\mathcal{X} = \{0, 1\}$
- Conditional pmf function:

| $\theta$ \( | \) $x$ | 0 | 1 |
|---|---|---|
| (Oil) $\theta_1$ | 0.3 | 0.7 |
| (No Oil) $\theta_2$ | 0.6 | 0.4 |

Note:
- rows sum to 1 (conditional distributions!)
- $X = 1$ supports $\theta_1$ (Oil)
- $X = 0$ supports $\theta_0$ (No Oil)
Oil Wildcatter Problem

\[ \mathcal{D} : \text{Class of all possible Decision Rules} \]

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( \delta(X = 0) )</th>
<th>( \delta(X = 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_1 )</td>
<td>( a_1 )</td>
<td>( a_1 )</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>( a_1 )</td>
<td>( a_2 )</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>( a_1 )</td>
<td>( a_3 )</td>
</tr>
<tr>
<td>( \delta_4 )</td>
<td>( a_2 )</td>
<td>( a_1 )</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>( a_2 )</td>
<td>( a_2 )</td>
</tr>
<tr>
<td>( \delta_6 )</td>
<td>( a_2 )</td>
<td>( a_3 )</td>
</tr>
<tr>
<td>( \delta_7 )</td>
<td>( a_3 )</td>
<td>( a_1 )</td>
</tr>
<tr>
<td>( \delta_8 )</td>
<td>( a_3 )</td>
<td>( a_2 )</td>
</tr>
<tr>
<td>( \delta_9 )</td>
<td>( a_3 )</td>
<td>( a_3 )</td>
</tr>
</tbody>
</table>

Note:
- \( \delta_4 \) Drills or Sells consistent with \( X \)
- \( \delta_2 \) Drills or Sells discordant with \( X \)
- \( \delta_1, \delta_5 \) and \( \delta_9 \) ignore \( X \).
Oil Wildcatter Problem

Risk Function: $R(\theta, \delta) = E[L(\theta, \delta(X) \mid \theta]$

$= \sum_{i=1}^{3} L(\theta, a_i) P(\delta(X) = a_i \mid \theta)$

Risk Set: $S = \{ \text{risk points } (R(\theta_1, \delta), R(\theta_2, \delta)), \text{ for } \delta \in D \}$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\delta(X = 0)$</th>
<th>$\delta(X = 1)$</th>
<th>$R(\theta_1, \delta)$</th>
<th>$R(\theta_2, \delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_1$</td>
<td>$a_1$</td>
<td>$a_1$</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>7</td>
<td>7.6</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>$a_1$</td>
<td>$a_3$</td>
<td>3.5</td>
<td>9.6</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>$a_2$</td>
<td>$a_1$</td>
<td>3</td>
<td>5.4</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>$a_2$</td>
<td>$a_2$</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>$a_2$</td>
<td>$a_3$</td>
<td>6.5</td>
<td>3</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>$a_3$</td>
<td>$a_1$</td>
<td>1.5</td>
<td>8.4</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>$a_3$</td>
<td>$a_2$</td>
<td>8.5</td>
<td>4.0</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>$a_3$</td>
<td>$a_3$</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: When $\Theta$ is finite with $k$ elements, the whole risk function of a procedure $\delta$ is represented by a point in $k$-dimensional space.
Bayes Risk: For prior distribution $\pi : r(\pi, \delta) = \sum_\theta \pi(\theta)R(\theta, \delta)$

Consider e.g., $\pi(\theta_1) = 0.2$ and $\pi(\theta_2) = 0.8$

$$r(\pi, \delta) = \pi(\theta_1) \times R(\theta_1, \delta) + \pi(\theta_2) \times R(\theta_2, \delta)$$

$$= 0.2 \times R(\theta_1, \delta) + 0.8 \times R(\theta_2, \delta)$$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\delta(X = 0)$</th>
<th>$\delta(X = 1)$</th>
<th>$R(\theta_1, \delta)$</th>
<th>$R(\theta_2, \delta)$</th>
<th>$r(\pi, \delta)$</th>
<th>$\max_\theta R(\theta, \delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_1$</td>
<td>$a_1$</td>
<td>$a_1$</td>
<td>0</td>
<td>12</td>
<td>9.6</td>
<td>12</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>7</td>
<td>7.6</td>
<td>7.48</td>
<td>7.6</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>$a_1$</td>
<td>$a_3$</td>
<td>3.5</td>
<td>9.6</td>
<td>8.38</td>
<td>9.6</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>$a_2$</td>
<td>$a_1$</td>
<td>3</td>
<td>5.4</td>
<td>4.92</td>
<td>5.4</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>$a_2$</td>
<td>$a_2$</td>
<td>10</td>
<td>1</td>
<td>2.8</td>
<td>10</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>$a_2$</td>
<td>$a_3$</td>
<td>6.5</td>
<td>3</td>
<td>3.7</td>
<td>6.5</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>$a_3$</td>
<td>$a_1$</td>
<td>1.5</td>
<td>8.4</td>
<td>7.02</td>
<td>8.4</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>$a_3$</td>
<td>$a_2$</td>
<td>8.5</td>
<td>4.0</td>
<td>4.9</td>
<td>8.4</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>$a_3$</td>
<td>$a_3$</td>
<td>5</td>
<td>6</td>
<td>5.8</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: $\delta_5$ is Bayes rule for prior $\pi$ – it achieves the minimum Bayes risk.
Computing Bayes Risks

- Bayes risk for discrete priors:
  \[ r(\pi, \delta) = \sum_{\theta} \pi(\theta)R(\theta, \delta) \]

- Bayes risk for continuous priors:
  \[ r(\pi, \delta) = \int_{\Theta} \pi(\theta)R(\theta, \delta) d\theta \]

Identifying Bayes Procedures

- Identification of Bayes rule does not require exhaustive search
- **Posterior analysis** specifies Bayes rule(s) directly
- Apply **Posterior Distribution** of \( \theta \) given \( X \) to minimize risk a posteriori.

Limits of Bayes Procedures

- Bayes-risk comparisons can be useful when \( \pi(\theta) \) improper
  i.e., \( \int_{\Theta} \pi(\theta) d\theta = \infty \) (e.g., uniform prior on \( \mathcal{R} \))
- Such comparisons relate to the consideration of limits of Bayes procedures.
Minimax Criterion: 
- Prefer $\delta$ to $\delta'$ if 
  \[
  \sup_{\theta \in \Theta} R(\theta, \delta) < \sup_{\theta \in \Theta} R(\theta, \delta')
  \]
- A procedure $\delta^*$ is called minimax if 
  \[
  \sup_{\theta \in \Theta} R(\theta, \delta^*) = \inf_{\delta \in D} \sup_{\theta \in \Theta} R(\theta, \delta)
  \]

Game-Theoretic Framework: Two-Person Games
- Player I (Nature chooses $\theta$)
- Player II (Statistician chooses $\delta$)
- Player II pays Player I $R(\theta, \delta)$.
- Minimax Theorem: von Neumann (1928)
  Subject to regularity conditions (e.g., “perfect information” and “zero-sum” payoffs), there exists a pair of strategies: 
  - $\pi^*$ for nature and 
  - $\delta^*$ for the Statistician 
  which allows each to minimize his/her maximum losses.
Elements of Decision Problems: Randomization

Randomized States of Nature

- State of Nature: \( \theta \sim \pi(\cdot) \)
- Prior Distribution for \( \theta \in \Theta \).

Randomized Decision Rules

- \( \mathcal{D} = \) Class of all (non-randomized) decision procedures.
- \( \mathcal{D}^* = \) Class of randomized decision procedures.
- Consider \( \delta^* \in \mathcal{D}^* : \)
  - Set of non-randomized procedures: \( \{\delta_1, \delta_2, \ldots, \delta_q\} \)
  - \( \delta^* : P(\delta^* = \delta_i) = \lambda_i, \ i = 1, \ldots, q \) (with \( \sum_{i=1}^q \lambda_i = 1 \))
  - Extend definitions of Risk and Bayes risk:
    \[
    R(\theta, \delta^*) = \sum_{i=1}^q R(\theta, \delta_i)
    \]
    \[
    r(\pi, \delta^*) = \sum_{i=1}^q r(\pi, \delta_i)
    \]
Elements of Decision Problems: Randomization

Risk Set $S^*$

- $k$–dimensional parameter space $\Theta = \{(\theta_1, \ldots, \theta_k) \in R^k\}$
- The risk set of non-randomized procedures $\mathcal{D} = \{\delta\}$ is
  $S = \{(R(\theta_1, \delta), R(\theta_2, \delta), \ldots, R(\theta_k, \delta)), \delta \in \mathcal{D}\}$
- The risk set of randomized procedures $\mathcal{D}^* = \{\delta^*\}$ is
  $S^* = \{(R(\theta_1, \delta^*), R(\theta_2, \delta^*), \ldots, R(\theta_k, \delta^*)), \delta^* \in \mathcal{D}^*\}$
- $S^*$ is the convex hull of $S$

Example: Oil Wildcatter Problem

- $\Theta = \{\theta_1(\text{Oil}), \theta_2(\text{No Oil})\}$
- Prior distribution $\pi : \pi(\theta_1) = \gamma$ and $\pi(\theta_2) = 1 - \gamma$
- Contour of constant Bayes risk ($= r_0$)

\[
S_{r_0}^{**} = \{(R(\theta_1, \delta), R(\theta_2, \delta)) : \gamma R(\theta_1, \delta) + (1 - \gamma) R(\theta_2, \delta) = r_0\}
\]

\[
= \{(x, y) : \gamma x + (1 - \gamma) y = r_0\}
\]

\[
= \{(x, y) : y = \frac{r_0}{1-\gamma} - \frac{\gamma}{1-\gamma} x\}
\]

(Line with slope $-\gamma/(1 - \gamma)$)
Bayes and Minimax Procedures in Risk Sets

Bayes Procedures

- Bayes rule(s): find risk point $s \in S^*$ that intersects $S_{r_0}^{**}$ with the smallest value of Bayes risk $r_0$.
- Lower-left convex hull of $S$ identifies all Bayes procedures. (Points with tangents having negative slope, including $-\infty$)
- If the tangent/intersection is a single point, the Bayes rule is unique and non-randomized.
- If the tangent/intersection is a line, then the Bayes rules are any whose risk point lies on the line.
  Such points correspond to randomized procedures between two non-randomized procedures
- For any prior, there is a non-randomized Bayes rule.

Minimax Procedures

- Minimax rule(s): find risk point $s \in S^*$ that intersects $Q(c^*) = \{(x, y) : x \leq c^* \text{ and } y \leq c^*\}$ lower-left quadrant with smallest value $c^*$.
Theoretical Results of Decision Theory

Results for Finite $\Theta$
- If minimax procedures exist, then they are Bayes procedures.
- All admissible procedure are Bayes procedures for some prior.
- If a Bayes prior has $\pi(\theta_i) > 0$ for all $i$ then any Bayes procedure corresponding to $\pi$ is admissible.

Results for Non-Finite $\Theta$
- If a Bayes prior $\pi$ has density $\pi(\theta) > 0$ for all $\theta \in \Theta$, then any Bayes procedure corresponding to $\pi$ is admissible.
- Under additional conditions, all admissible procedures are either Bayes procedures, or limits of Bayes procedures.

Key References:
- Wald, A. (1950). *Statistical Decision Functions*
Problems

Problem 1.3.3 Testing problem with three hypotheses.

Problem 1.3.4 Stratified sampling – evaluating MSEs of different estimators.

Problem 1.3.8 Variance estimation: deriving unbiased estimator; lowering MSE with biased estimator.

Problem 1.3.14 Convexity of the risk set.

Problem 1.3.18 Sampling inspection example 1.1.1 with asymmetric loss function.