# ECE531 Lecture 3: Minimax Hypothesis Testing

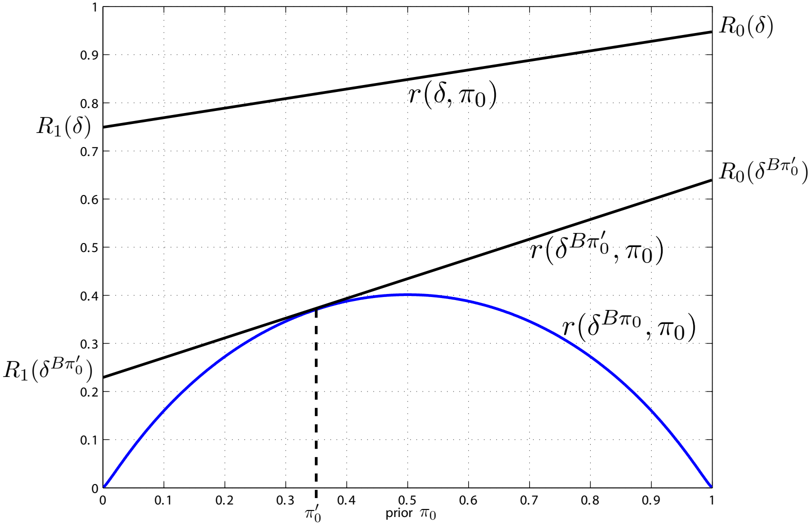
D. Richard Brown III

Worcester Polytechnic Institute

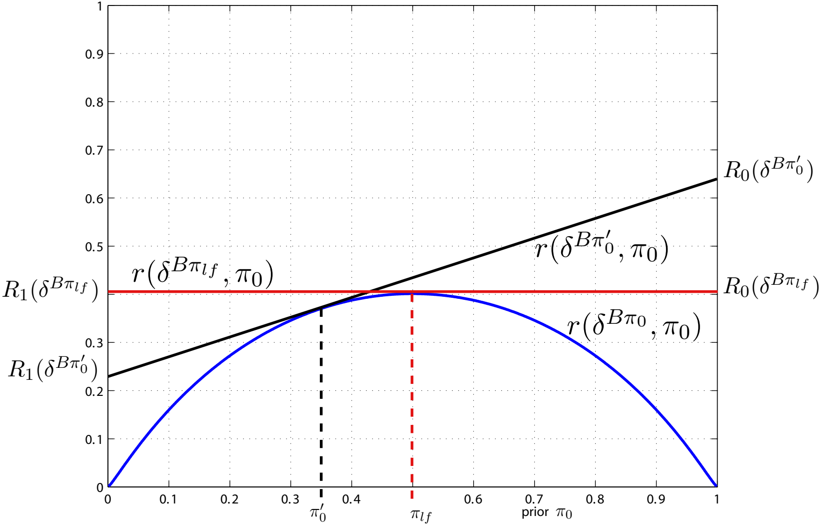
05-February-2009

# Simple Binary Bayesian Risks Under Different Priors

$$r(\delta, \pi_0) = \pi_0 R_0(\delta) + (1 - \pi_0) R_1(\delta)$$



# Least Favorable Prior State Distribution



# Minimax Hypothesis Testing

Definition:

$$\rho^{mm} \quad := \quad \arg\min_{\rho}\max_{j} R_j(\rho)$$

Remarks:

- No single decision rule minimizes the weighted average, e.g. Bayes, risk for every possible prior state distribution.
- A conservative approach is to minimize the worst case risk over all possible prior state distributions.
- Intuitively, there should be a least favorable prior. Does it always exist? Is it unique?
- Intuitively, the minimax decision rule should be the Bayesian decision rule with constant Bayesian risk over the priors. Is this always true?

# Minimum Bayesian Risk as a Function of the Prior

Let $V(\pi) := r(\delta^{B\pi}, \pi)$ be the minimum Bayesian risk for the prior $\pi$.

**Theorem**

*The minimum Bayesian risk $V(\pi)$ is concave and continuous over the space of priors satisfying $\pi_j \geq 0$, $j = 0, 1, \ldots, N-1$, and $\sum_j \pi_j = 1$. Hence, there exists a unique least favorable prior*

$$\pi_{lf} = \arg\max_{\pi} V(\pi).$$

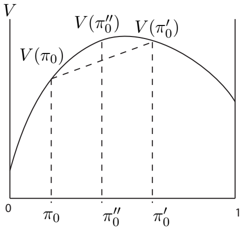
# Concavity of the Minimum Bayesian Risk

A function is **concave** if, for any $\{x, y\}$ in the domain of $f$ and any $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y)$.

Denote a pair of priors as $\pi$ and $\pi'$ and a third prior $\pi'' = \alpha\pi + (1-\alpha)\pi'$. We can write

$$\begin{aligned} V(\pi'') &= \pi''^\top R(\delta^{B\pi''}) \\ &= \alpha\pi^\top R(\delta^{B\pi''}) + (1-\alpha)\pi'^\top R(\delta^{B\pi''}) \\ &\geq \alpha V(\pi) + (1-\alpha)V(\pi') \end{aligned}$$

hence $V(\pi)$ is concave.



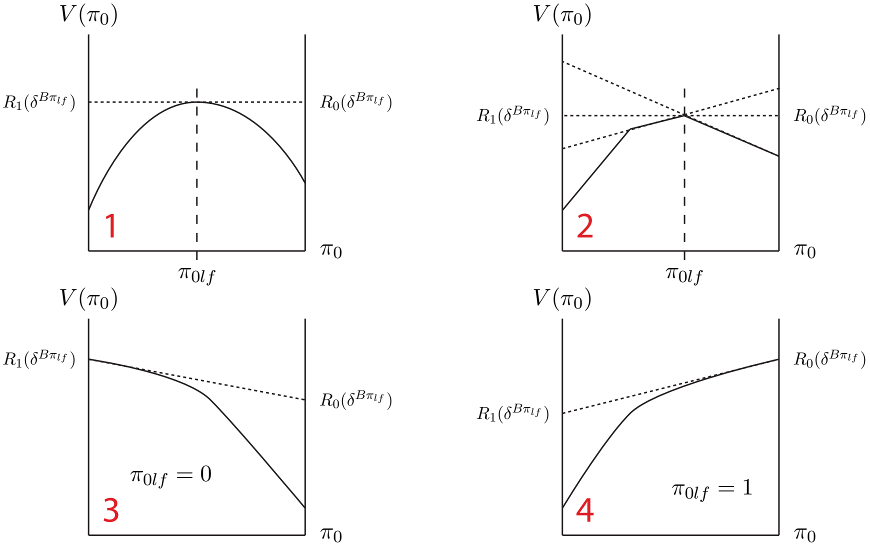
# Continuity of the Minimum Bayesian Risk

**Theorem ("A First Course in Optimization Theory" by R.K. Sundaram)**

*Let $f : \mathcal{D} \to \mathbb{R}$ be a concave function. Then, if $\mathcal{D}$ is open, $f$ is continuous on $\mathcal{D}$. If $\mathcal{D}$ is not open, $f$ is continuous on the interior of $\mathcal{D}$.*

Note that continuity does not imply differentiability.

# The Four Possibilities



# Case 1: Differentiable Interior Maximum Risk

## Theorem

*If there exists a prior* $\pi'$ *such that the conditional risks satisfy* $R_0(\delta^{B\pi'}) = R_1(\delta^{B\pi'})$ *then* $\pi'$ *is a least favorable prior and the minimax decision rule is* $\rho^{mm} = \delta^{B\pi'}$.

## Proof.

Given a $\pi'$ satisfying $R_0(\delta^{B\pi'}) = R_1(\delta^{B\pi'})$. For any $\delta$,

$$\begin{aligned}
\max\{R_0(\delta), R_1(\delta)\} &\geq \max_{\pi_0 \in [0,1]} \pi_0 R_0(\delta) + (1-\pi_0)R_1(\delta) \\
&\geq \pi' R_0(\delta) + (1-\pi')R_1(\delta) \\
&\geq \pi' R_0(\delta^{B\pi'}) + (1-\pi')R_1(\delta^{B\pi'}) \\
&= R_0(\delta^{B\pi'}) = R_1(\delta^{B\pi'})
\end{aligned}$$

Moreover, for any $\pi_0 \in [0,1]$,

$$V(\pi') = \pi' R_0(\delta^{B\pi'}) + (1-\pi')R_1(\delta^{B\pi'}) = \pi_0 R_0(\delta^{B\pi'}) + (1-\pi_0)R_1(\delta^{B\pi'}) \geq V(\pi_0)$$

□

# A Procedure for Finding the Minimax Decision Rule

1. Find a Bayesian decision rule $\delta^{B\pi}$ as a function of the prior $\pi$.
2. See if Case 1 holds by solving for the unique least favorable prior $\pi_{lf}$ using the equalizer rule:

$$R_0(\delta^{B\pi_{lf}}) = R_1(\delta^{B\pi_{lf}})$$

3. If the solution exists, then set

$$\rho^{mm} = \delta^{B\pi_{lf}}$$

4. If there is no solution to the equalizer rule, then see if Case 3 or 4 holds by computing the risk at the endpoints $\pi_{0lf} = 0$ and $\pi_{0lf} = 1$.
5. If neither endpoint is least favorable, then we must be in Case 2. In this case we must create a randomized minimax decision rule as a convex function of two deterministic Bayes decision rules.

# Example: Coherent Detection of BPSK

Our Bayes decision rule for coherent BPSK with prior $\pi_0$, $\pi_1 = 1 - \pi_0$ is

$$\delta^{B\pi}(y) = \begin{cases} 1 & \text{if } y > \gamma \\ 0/1 & \text{if } y = \gamma \\ 0 & \text{if } y < \gamma. \end{cases}$$

where $\gamma := \frac{a_0+a_1}{2} + \frac{\sigma^2}{a_1-a_0} \ln \frac{\pi_0}{\pi_1}$.
The conditional risks are

$$\begin{aligned} R_0(\delta^{B\pi}) &= Q\left(\frac{\gamma - a_0}{\sigma}\right) \\ R_1(\delta^{B\pi}) &= Q\left(\frac{a_1 - \gamma}{\sigma}\right) \end{aligned}$$

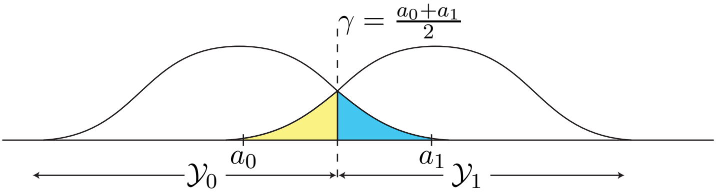
where $Q(x) := \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$.

Let's try the equalizer rule. What value of $\gamma$ gives us $R_0(\delta^{B\pi}) = R_1(\delta^{B\pi})$?

# Example: Coherent Detection of BPSK

Answer: $R_0 = R_1$ when $\gamma = \frac{a_0+a_1}{2}$. Hence

$$\rho^{mm}(y) = \begin{cases} 1 & \text{if } y > \frac{a_0+a_1}{2} \\ 0/1 & \text{if } y = \frac{a_0+a_1}{2} \\ 0 & \text{if } y < \frac{a_0+a_1}{2}. \end{cases}$$

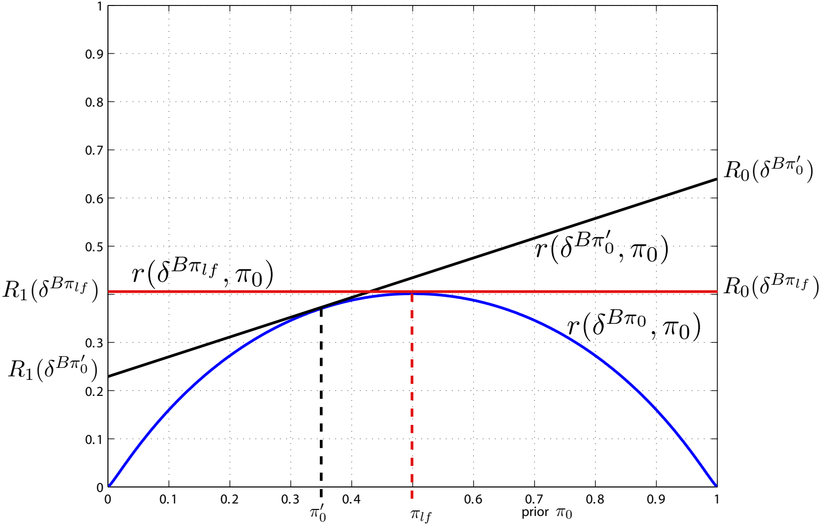



What does this imply about the least favorable prior?

Answer: $\pi_0 = \pi_1 = \frac{1}{2}$.

Given $a_0$, $a_1$, and $\sigma$, the minimax rule allows you to guarantee a worst-case risk over all priors.

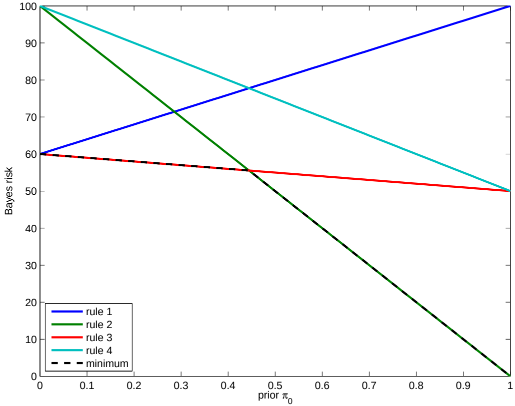
# Example: Coherent Detection of BPSK



# Cases 3-4: Maximum Risk Occurs at Boundary

Let's return to our coin flipping problem from Lecture 1 ($\mathcal{H}_0 \leftrightarrow x_0 = \text{HT}$ and $\mathcal{H}_1 \leftrightarrow x_1 = \text{HH}$) with a modified cost matrix

$$C = \begin{bmatrix} 0 & 100 \\ 100 & 60 \end{bmatrix}$$



# Cases 3-4: Maximum Risk Occurs at Boundary

Remarks:

- Rules 2 and 3, depending on the prior, minimize the Bayes risk.
- Note that the equalizer rule would give no solution to this problem:

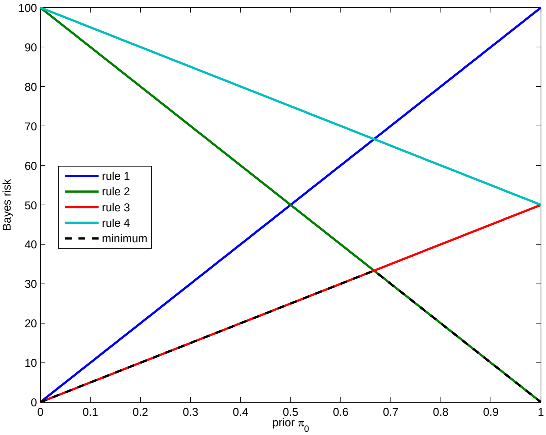
$$\begin{aligned} R_0(D_1) &= 100 \quad \text{and} \quad R_1(D_1) = 60 \\ R_0(D_2) &= 0 \quad \text{and} \quad R_1(D_2) = 100 \\ R_0(D_3) &= 50 \quad \text{and} \quad R_1(D_3) = 60 \\ R_0(D_4) &= 50 \quad \text{and} \quad R_1(D_4) = 100 \end{aligned}$$

  No decision rule gives $R_0 = R_1$.
- In this example, the least favorable prior (maximizing the minimum risk) is $\pi_0 = 0$, or that the coin is always HH. This should make sense.
- The minimax decision rule is Rule 3: *observe T, decide the coin is fair; observe H, decide the coin is unfair.*
- You can guarantee a worst-case risk of $60 by using Rule 3.

# Case 2: Non-Differentiable Interior Maximum

Back to our original coin flipping problem with cost matrix

$$C = \begin{bmatrix} 0 & 100 \\ 100 & 0 \end{bmatrix}$$



# Case 2: Non-Differentiable Interior Maximum

Remarks:

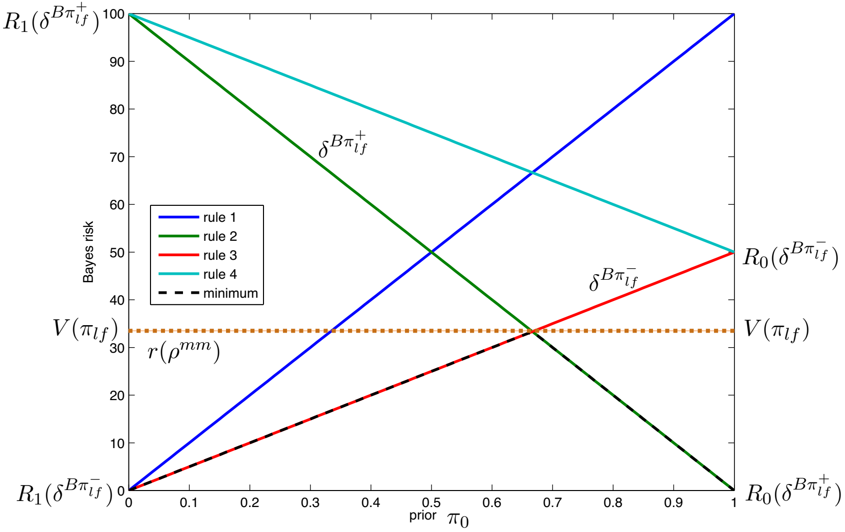
- Rules 2 and 3, depending on the prior, minimize the Bayes risk.
- Again, the equalizer rule gives no deterministic solution since

$$\begin{aligned} R_0(D_1) &= 100 \quad \text{and} \quad R_1(D_1) = 0 \\ R_0(D_2) &= 0 \quad \text{and} \quad R_1(D_2) = 100 \\ R_0(D_3) &= 50 \quad \text{and} \quad R_1(D_3) = 0 \\ R_0(D_4) &= 50 \quad \text{and} \quad R_1(D_4) = 100 \end{aligned}$$

- In this example, the least favorable prior (maximizing the minimum risk) is $\pi_0 = \frac{2}{3}$, or that the coin is HT with probability $\frac{2}{3}$.
- The minimax decision rule is **neither Rule 2 or Rule 3**.
- You can guarantee a worst-case risk of $\frac{\$100}{3}$ by using a randomized decision rule that is a combination of Rules 2 and 3.

# Case 2: Non-Differentiable Interior Maximum

Problem: Find a randomized decision rule that satisfies the equalizer rule



# Case 2: Non-Differentiable Interior Maximum

Our randomized minimax decision rule is the

$$\rho^{mm} = \alpha\delta^{B\pi_{lf}^{-}} + (1-\alpha)\delta^{B\pi_{lf}^{+}}$$

We can calculate the randomization $\alpha \in [0,1]$ by applying the equalizer rule:

$$\alpha R_0(\delta^{B\pi_{lf}^{-}}) + (1-\alpha)R_0(\delta^{B\pi_{lf}^{+}}) = \alpha R_1(\delta^{B\pi_{lf}^{-}}) + (1-\alpha)R_1(\delta^{B\pi_{lf}^{+}})$$

which gives the solution

$$\begin{aligned}\alpha &= \frac{R_1(\delta^{B\pi_{lf}^{+}}) - R_0(\delta^{B\pi_{lf}^{+}})}{(R_0(\delta^{B\pi_{lf}^{-}}) - R_1(\delta^{B\pi_{lf}^{-}})) - (R_0(\delta^{B\pi_{lf}^{+}}) - R_1(\delta^{B\pi_{lf}^{+}}))} \\ &= \frac{V'(\pi_{lf}^{+})}{V'(\pi_{lf}^{+}) - V'(\pi_{lf}^{-})}\end{aligned}$$

# Case 2: Non-Differentiable Interior Maximum

What is the minimax decision rule in our example?

$$\begin{aligned} V'(\pi_{lf}^{+}) &= -100 \\ V'(\pi_{lf}^{-}) &= 50 \end{aligned}$$

hence $\alpha = \frac{2}{3}$. If $\mathcal{H}_0$ is the hypothesis that the coin is HT, $\mathcal{H}_1$ is the hypothesis that the coin is HH, the observations $y_0 = \text{T}$, $y_1 = \text{H}$, then our deterministic decision rules 2 and 3 can be written as

$$D_3 = \delta^{B\pi_{lf}^{-}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } D_2 = \delta^{B\pi_{lf}^{+}} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

The minimax decision rule is then given by

$$\begin{aligned} \rho^{mm}(y = \text{T}) &= \frac{2}{3}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (\text{T} \rightarrow \text{always decide HT}) \\ \rho^{mm}(y = \text{H}) &= \frac{2}{3}\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} \quad \left(\text{H} \rightarrow \text{randomize} \rightarrow \begin{cases} \text{Pr(decide HT)=1/3} \\ \text{Pr(decide HH)=2/3} \end{cases}\right) \end{aligned}$$

# Final Remarks on Minimax Hypothesis Testing

1. The objective of minimax hypothesis testing is to minimize your worst-case (maximum) risk over all possible prior state probabilities.
2. Conservative approach but useful in scenarios when:
   - the prior is unknown and/or
   - you need to provide a maximum risk guarantee.
3. Try the equalizer rule first!
4. Minimax risk at the endpoints only occurs in weird cases.
5. Finite observation space $\mathcal{Y}$ implies that the minimum Bayes risk curve $V$ is not going to be differentiable everywhere. Randomization is often necessary to obtain the minimax decision rule in these cases.
6. Composite hypotheses:
   - The equalizer rule is still valid.
   - Checking "endpoints" is still valid (but more points to check).
   - In the case of a non-differentiable interior maximum, finding the randomization can be difficult.