ELE530: Minimax Risk Hypothesis Testing

Fernando Pérez-Cruz

Princeton University and University Carlos III in Madrid

February 9$^{\text{th}}$, 2009

Minimax Risk

- We need to minimize our maximum risk

$$\delta^{\text{minmax}} = \arg\min_{\delta} \max_{p(y)} \sum_{y=1}^{M} \int L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) p(y) d\mathbf{x}$$

- How do we do it?

Minimax Risk

- We need to minimize our maximum risk

$$\delta^{\text{minmax}} = \arg\min_{\delta}\max_{p(y)}\sum_{y=1}^{M}\int L(\delta(\mathbf{x}),y)p(\mathbf{x}|y)p(y)d\mathbf{x}$$

- For a binary problem $\mathcal{Y}=\{0,1\}$, we are given a decision function $\delta(\mathbf{x})=\delta'(\mathbf{x})$ that it is optimal for $p(y=0)=\pi'$.

- The risk as a function of π is given by:

$$R(\delta') = L_{00}\pi\int_{\mathcal{X}'_0} p(\mathbf{x}|y=0)d\mathbf{x} + L_{10}\pi\int_{\mathcal{X}'_1} p(\mathbf{x}|y=0)d\mathbf{x}$$
$$+ L_{01}(1-\pi)\int_{\mathcal{X}'_0} p(\mathbf{x}|y=1)d\mathbf{x} + L_{11}(1-\pi)\int_{\mathcal{X}'_1} p(\mathbf{x}|y=1)d\mathbf{x}$$

Minimax Risk

- We need to minimize our maximum risk

$$\delta^{\text{minmax}} = \arg\min_{\delta} \max_{p(y)} \sum_{y=1}^{M} \int L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) p(y) d\mathbf{x}$$

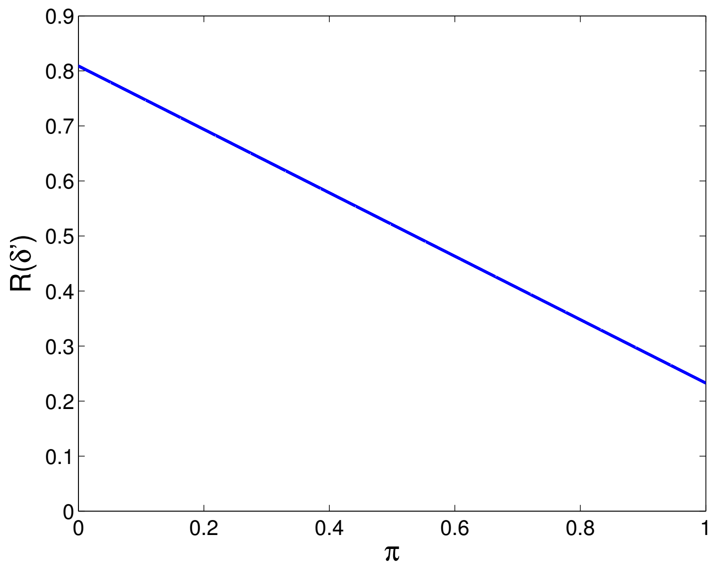
- For a binary problem $\mathcal{Y} = \{0, 1\}$, we are given a decision function $\delta(\mathbf{x}) = \delta'(\mathbf{x})$ that it is optimal for $p(y = 0) = \pi'$.
- The risk as a function of π is given by:

$$R(\delta') = L_{00} P_{00} \pi + L_{10} P_{10} \pi + L_{01} P_{01} (1 - \pi) + L_{11} P_{11} (1 - \pi),$$

which is a linear function of π.

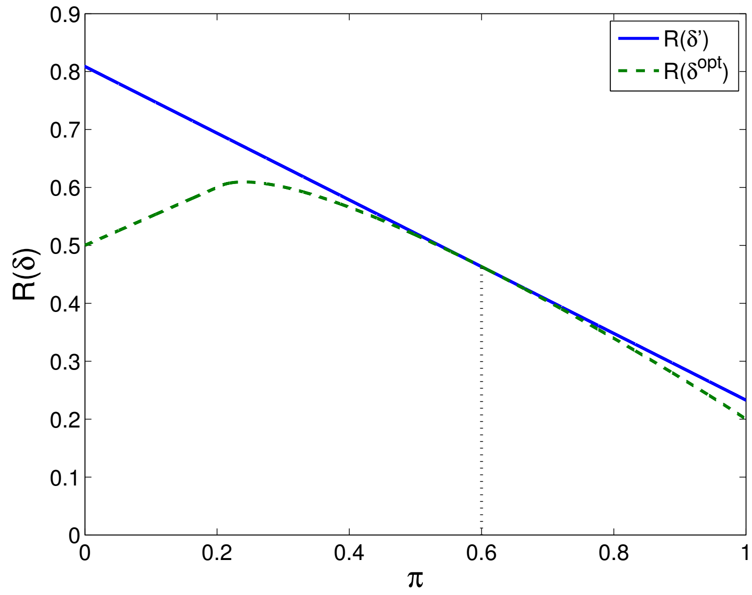
Example

- We observe a real-valued quantity $p(x|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x-\mu_y).^2}{2\sigma_y^2}\right)$.
- $\mu_1 = 0$, $\mu_0 = -1$ and $\sigma_1 = 2\sigma_0 = 2$.
- $L_{00} = 0.2$, $L_{11} = 0.5$ and $L_{10} = L_{01} = 1$.



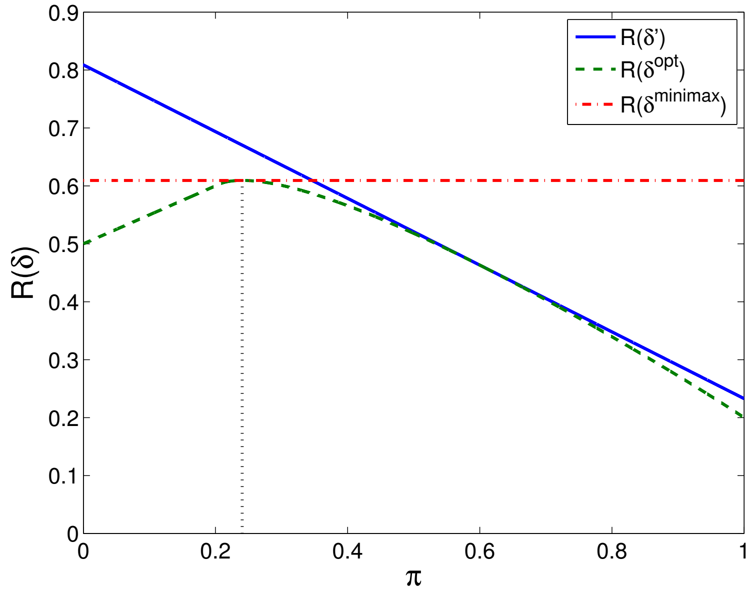
Example

- Now we compute the optimal $R(\delta^{\text{opt}})$ for all values of π,
- which is a concave function of π and touches $R(\delta')$ at $\pi = 0.6$.



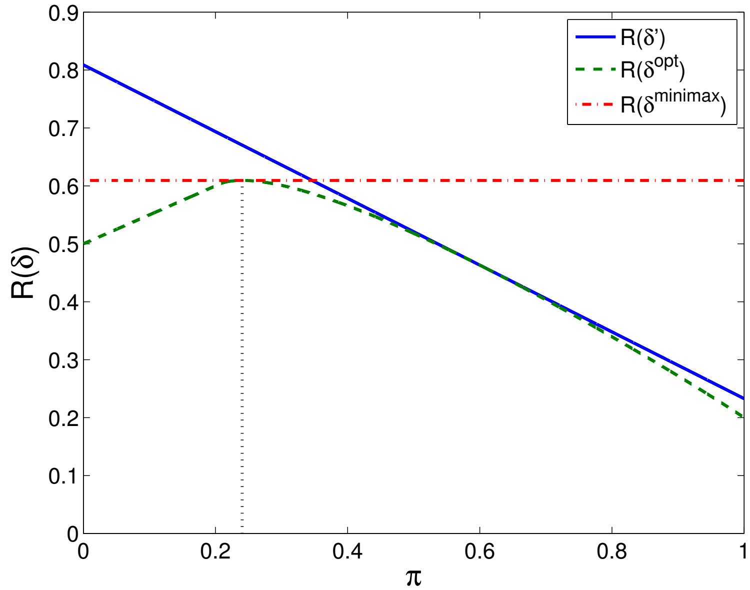
- How would you choose the minimax prediction: δ^{minmax}?

Example: Minimax Risk



- What property does $R(\delta^{\text{minmax}})$ has?

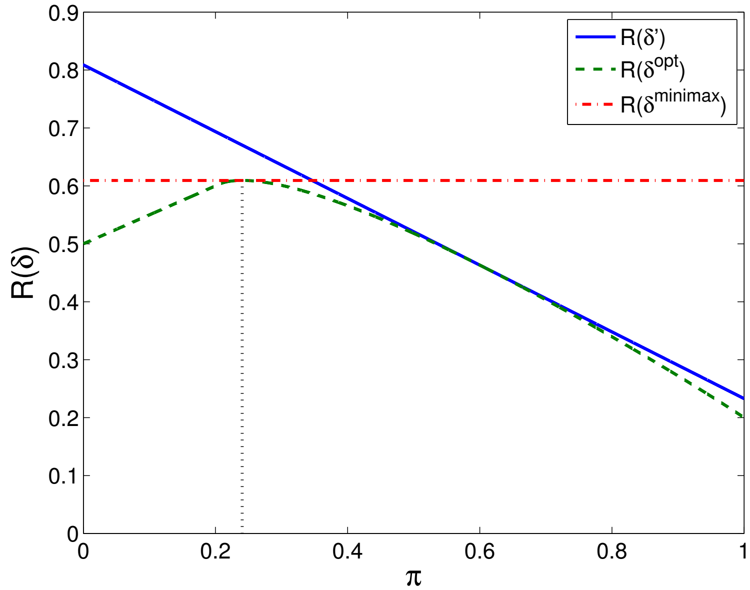
Example: Minimax Risk



- What property does $R(\delta^{\text{minmax}})$ has?

$$L_{00}P_{00} + L_{10}P_{10} = L_{01}P_{01} + L_{11}P_{11}$$

Example: Minimax Risk



- What property does $R(\delta^{\text{minmax}})$ has?

$$L_{00}P_{00} + L_{10}P_{10} = L_{01}P_{01} + L_{11}P_{11}$$

$$R(\delta^{\text{minmax}}|y = 0) = R(\delta^{\text{minmax}}|y = 1)$$

Minimax Rule

- We need to minimize our maximum risk

$$\delta^{\mathsf{minmax}} = \arg\min_{\delta} \max_{p(y)} \sum_{y=1}^{M} \int L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) p(y) d\mathbf{x}$$

- We have found that we need:

$$R(\delta^{\mathsf{minmax}}|y=1) = R(\delta^{\mathsf{minmax}}|y=2) = \cdots = R(\delta^{\mathsf{minmax}}|y=M)$$

- Is this always possible?

Minimax Rule

- We need to minimize our maximum risk

$$\delta^{\text{minmax}} = \arg\min_{\delta} \max_{p(y)} \sum_{y=1}^{M} \int L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) p(y) d\mathbf{x}$$

- We have found that we need:

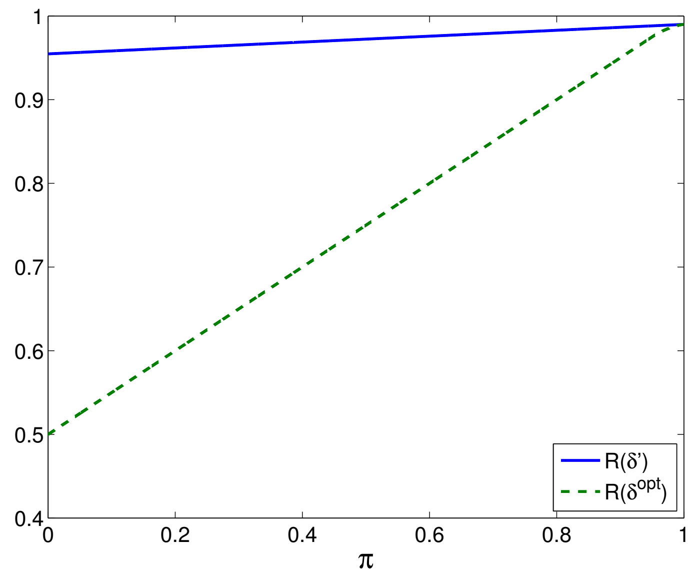
$$R(\delta^{\text{minmax}}|y=1) = R(\delta^{\text{minmax}}|y=2) = \cdots = R(\delta^{\text{minmax}}|y=M)$$

- Is this always possible? No
  - Special case I: The maximum is at an extreme point.
  - Special case II: The observation space is finite.
- Still the solution is the least favorable prior.

Special Case I

- In this case there is no δ^{minmax} that satisfies:

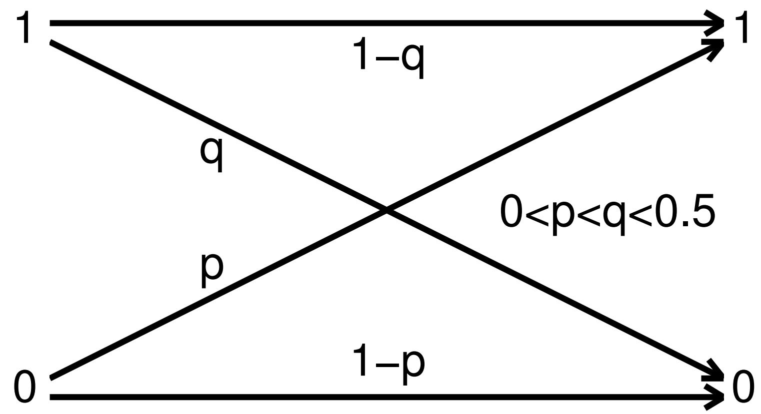
$$R(\delta^{\text{minmax}}|y=0) = R(\delta^{\text{minmax}}|y=1)$$



- Previous example with $L_{00} = 0.99$.

Special Case II: Example

- We have a binary channel:



with prior probabilities $p(y = 0) = \pi$ and $p(y = 1) = 1 - \pi$

- 4 possible detection rules:
- Which are they?

Special Case II: Example

- Decision Rules

$$
\begin{array}{lll}
\delta_0 : & \mathcal{X}_0 = \{0,1\} & \mathcal{X}_1 = \emptyset \\
\delta_1 : & \mathcal{X}_0 = \emptyset & \mathcal{X}_1 = \{0,1\} \\
\delta_2 : & \mathcal{X}_0 = \{0\} & \mathcal{X}_1 = \{1\} \\
\delta_3 : & \mathcal{X}_0 = \{1\} & \mathcal{X}_1 = \{0\}
\end{array}
$$

Special Case II: Example

- Decision Rules and their associated Risks

$$\delta_0: \quad \mathcal{X}_0 = \{0,1\} \quad \mathcal{X}_1 = \emptyset \quad R_0(\pi) = 1-\pi$$

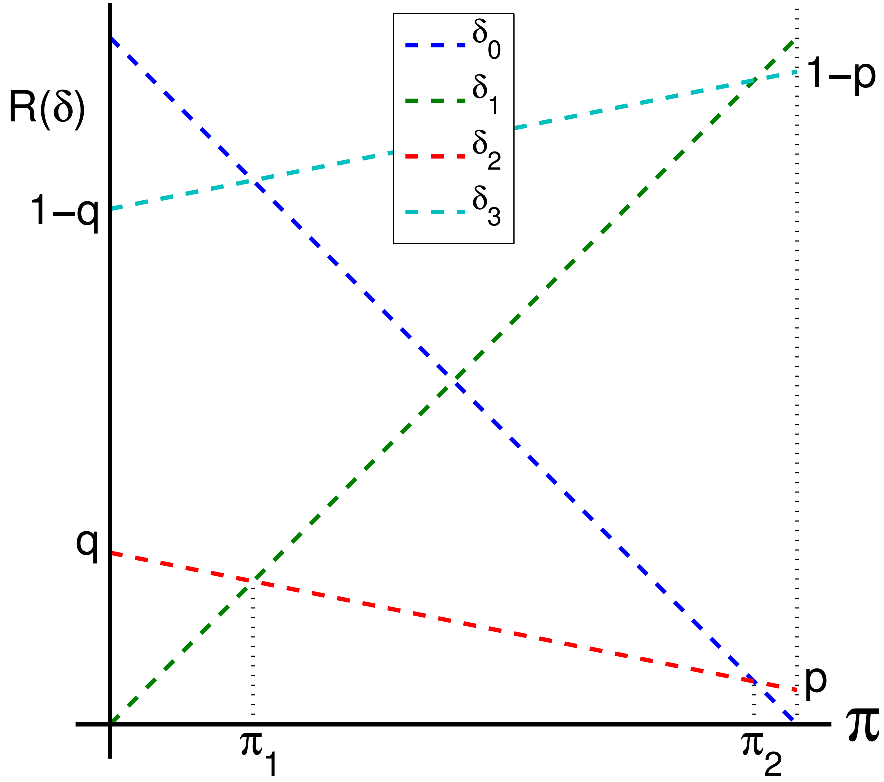
$$\delta_1: \quad \mathcal{X}_0 = \emptyset \quad \mathcal{X}_1 = \{0,1\} \quad R_0(\pi) = \pi$$

$$\delta_2: \quad \mathcal{X}_0 = \{0\} \quad \mathcal{X}_1 = \{1\} \quad R_0(\pi) = p\pi - q(1-\pi)$$

$$\delta_3: \quad \mathcal{X}_0 = \{1\} \quad \mathcal{X}_1 = \{0\} \quad R_0(\pi) = (1-p)\pi - (1-q)(1-\pi)$$

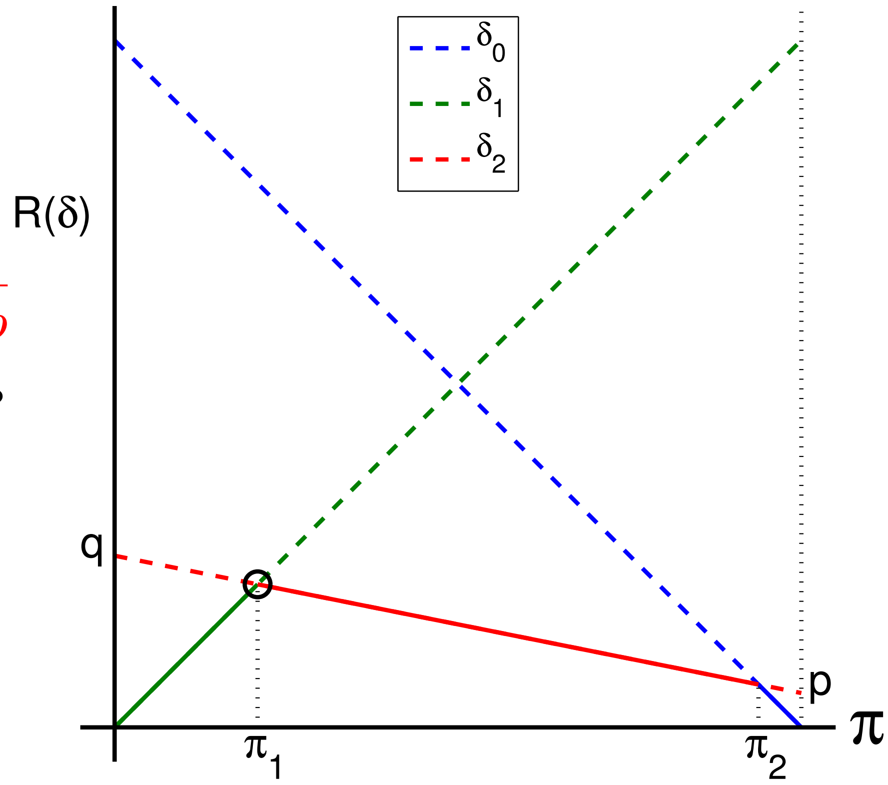
Risks for the four rules

- $\pi_1 = \dfrac{q}{1+q-p}$
- $\pi_2 = \dfrac{1-q}{1-q+p}$



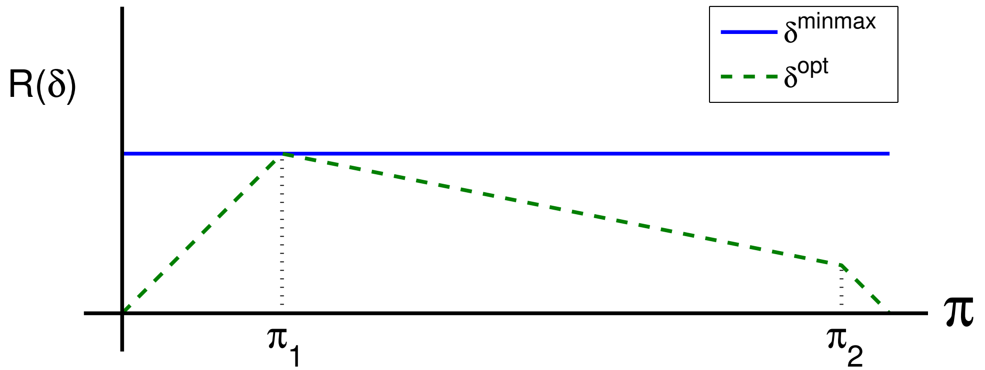
Risks for the four rules

- $\pi_1 = \dfrac{q}{1+q-p}$
- $\pi_2 = \dfrac{1-q}{1-q+p}$
- $R(\delta^{\text{minmax}}) = \pi_1 = \dfrac{1-q}{1-q+p}$
- How do we get $R(\delta^{\text{minmax}})$?



How can get the Minimax Rule?

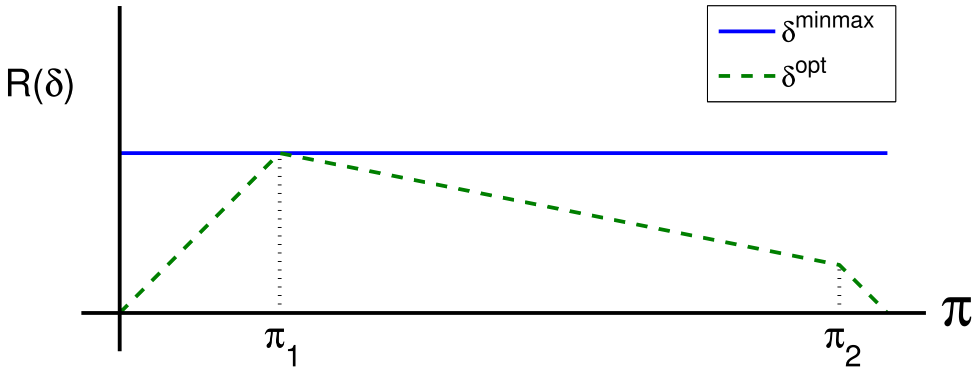
- Neither rule provides the minimax criterium:
  - $R(\delta_2) > R(\delta^{\text{minmax}})$ for $\pi < \pi_1$.
  - $R(\delta_1) > R(\delta^{\text{minmax}})$ for $\pi > \pi_1$.
  - $R(\delta_0) > R(\delta^{\text{minmax}})$ for $\pi < (1-p)/(1+q-p)$.



Randomize Rules

- Neither rule provides the minimax criterium:
  - $R(\delta_2) > R(\delta^{\text{minmax}})$ for $\pi < \pi_1$.
  - $R(\delta_1) > R(\delta^{\text{minmax}})$ for $\pi > \pi_1$.
- We use a randomized rule:

$$\delta^{\text{minmax}} = \begin{cases} \delta_1, & a \\ \delta_2, & 1-a \end{cases}$$



- How do we compute a?

$$aR_1(\pi) + (1-a)R_2(\pi) = \frac{q}{1+q-p} \qquad \forall \pi$$

$$a = \frac{q-p}{1+q-p}$$

Randomize Rules

- Another expression for randomize minimax rules:

$$\delta^{\text{minmax}}(\mathbf{x}) = \begin{cases} 1, & \mathcal{L}(\mathbf{x}) > \tau \\ a, & \mathcal{L}(\mathbf{x}) = \tau \\ 0, & \mathcal{L}(\mathbf{x}) < \tau \end{cases}$$

  where $\delta^{\text{minmax}}(\mathbf{x}) = a$ means that we choose $y = 1$ with probability a.

- We compute τ as the minimum Risk rule with the worst prior.

- For our example:

$$\tau = \frac{\pi}{1-\pi}\frac{L_{10}-L_{00}}{L_{01}-L_{11}} = \frac{\pi_1}{1-\pi_1}\frac{1-0}{1-0} = \frac{\frac{q}{1+q-p}}{1-\frac{q}{1+q-p}} = \frac{q}{1-p}$$

Bayesian Risk Minimization

- Minimax Risk minimization worse case analysis:
  - Maximum risk for all $p(y)$.
- Is there something else?
  - Assume that we are given a probability distribution over $p(y)$.

$$\pi = p(\pi)$$

  - Can we get a rule that minimizes the mean error?
- Can we view π as a random variable?

Bayesian Risk Minimization

- Minimum Risk:

$$\delta^{\text{opt}} = \arg\min_{\delta} \sum_{y} \int_{\mathcal{X}} L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) \pi d\mathbf{x}$$

- π is unknown and we are given a probability distribution over it.
- We can minimize the mean Risk:

$$\delta^{\text{Baymin}} = \arg\min_{\delta} \sum_{y} \int_{\mathcal{X}} \int_{\pi} L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) \pi p(\pi) d\mathbf{x}$$

- We can minimize with respect to each $\mathbf{x}$ individually:

$$\delta^{\text{Baymin}}(\mathbf{x}) = \arg\min_{\delta(\mathbf{x})} \sum_{y} \int L(\delta(\mathbf{x}), y) p(\mathbf{x}|y) \pi p(\pi)$$

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$R(\delta(\mathbf{x}) = 0) = \int [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)]\pi p(\pi)d\pi$$

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$R(\delta(\mathbf{x}) = 0) = \int [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)]\pi p(\pi)d\pi$$
$$= [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)] \int \pi p(\pi)d\pi$$

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$
\begin{aligned}
R(\delta(\mathbf{x}) = 0) = \int & [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\pi p(\pi) d\pi \\
=& [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)] \int \pi p(\pi) d\pi \\
=& [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)] \mu_\pi
\end{aligned}
$$

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$
\begin{aligned}
R(\delta(\mathbf{x}) = 0) = \int [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\pi p(\pi)d\pi \\
=[L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)] \int \pi p(\pi)d\pi \\
=[L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\mu_\pi
\end{aligned}
$$

- For $\delta(\mathbf{x}) = 1$, we get a similar result:

$$
R(\delta(\mathbf{x}) = 0) = [L_{10}p(\mathbf{x}|y=0) + L_{11}p(\mathbf{x}|y=1)]\mu_\pi
$$

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$R(\delta(\mathbf{x}) = 0) = \int [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)]\pi p(\pi)d\pi$$
$$= [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)] \int \pi p(\pi)d\pi$$
$$= [L_{00}p(\mathbf{x}|y = 0) + L_{01}p(\mathbf{x}|y = 1)]\mu_\pi$$

- For $\delta(\mathbf{x}) = 1$, we get a similar result:

$$R(\delta(\mathbf{x}) = 0) = [L_{10}p(\mathbf{x}|y = 0) + L_{11}p(\mathbf{x}|y = 1)]\mu_\pi$$

- These equations are identical to the Minimum Risk for $\pi = \mu_\pi$.

Bayesian Risk Minimization

- For a given $\mathbf{x}$, if we set $\delta(\mathbf{x}) = 0$:

$$\begin{aligned} R(\delta(\mathbf{x}) = 0) = & \int [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\pi p(\pi)d\pi \\ = & [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\int \pi p(\pi)d\pi \\ = & [L_{00}p(\mathbf{x}|y=0) + L_{01}p(\mathbf{x}|y=1)]\mu_\pi \end{aligned}$$

- For $\delta(\mathbf{x}) = 1$, we get a similar result:

$$R(\delta(\mathbf{x}) = 0) = [L_{10}p(\mathbf{x}|y=0) + L_{11}p(\mathbf{x}|y=1)]\mu_\pi$$

- These equations are identical to the Minimum Risk for $\pi = \mu_\pi$.
- In this case, the Bayes optimal rule is the optimal rule for the mean $p(y)$.