

ECE531 Lecture 2b: Bayesian Hypothesis Testing

D. Richard Brown III

Worcester Polytechnic Institute

29-January-2009

Minimizing a Weighted Sum of Conditional Risks

We would like to minimize all of the conditional risks $R_0(D), \dots, R_{N-1}(D)$, but usually we have to make some sort of tradeoff. A standard approach to solving a **multi-objective optimization problem** is to form a **linear combination** of the functions to be minimized:

$$r(D, \lambda) := \sum_{j=0}^{N-1} \lambda_j R_j(D) = \lambda^\top R(D)$$

where $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$.

The problem is then to solve the **single-objective optimization problem**:

$$\begin{aligned} D^* &= \arg \min_{D \in \mathcal{D}} r(D, \lambda) \\ &= \arg \min_{D \in \mathcal{D}} \sum_{j=0}^{N-1} \lambda_j c_j^\top D p_j \end{aligned}$$

where we are focusing on the case of finite \mathcal{Y} for now.

Minimizing a Weighted Sum of Conditional Risks

The problem

$$D^* = \arg \min_{D \in \mathcal{D}} r(D, \lambda) = \arg \min_{D \in \mathcal{D}} \sum_{j=0}^{N-1} \lambda_j c_j^\top D p_j$$

is a **finite-dimensional linear programming** problem since the variables that we control (D_{ij}) are constrained by linear inequalities

$$D_{ij} \geq 0 \quad \forall i \in \{0, \dots, M-1\} \text{ and } j \in \{0, \dots, L-1\}$$

$$\sum_{i=0}^{M-1} D_{ij} = 1$$

and the objective function $r(D, \lambda)$ is linear in the variables D_{ij} . For notational convenience, we call this minimization problem $\text{LPF}(\lambda)$.

Decomposition: Part 1

Since

$$R_j(D) = c_j^\top D p_j = \sum_{i=0}^{M-1} C_{ij} \sum_{\ell=0}^{L-1} D_{i\ell} P_{\ell j}$$

the problem $\text{LPF}(\lambda)$ can be written as

$$\begin{aligned} D^* &= \arg \min_{D \in \mathcal{D}} \sum_{j=0}^{N-1} \lambda_j \sum_{i=0}^{M-1} C_{ij} \sum_{\ell=0}^{L-1} D_{i\ell} P_{\ell j} \\ &= \arg \min_{D \in \mathcal{D}} \underbrace{\sum_{\ell=0}^{L-1} \left(\sum_{i=0}^{M-1} D_{i\ell} \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} P_{\ell j} \right] \right)}_{r_\ell(d, \lambda)} \end{aligned}$$

Note that we can minimize each term $r_\ell(d, \lambda)$ separately.

Decomposition: Part 2

The ℓ th subproblem is then

$$\begin{aligned} d^* &= \arg \min_{d \in \mathbb{R}^M} r_\ell(d, \lambda) \\ &= \arg \min_d \sum_{i=0}^{M-1} d_i \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} P_{\ell j} \right] \end{aligned}$$

subject to the constraints that $d_i \geq 0$ and $\sum_i d_i = 1$.

- ▶ Note that $d_i \leftrightarrow D_{i\ell}$ from the original $\text{LPF}(\lambda)$ problem.
- ▶ For notational convenience, we call this subproblem $\text{LPF}_\ell(\lambda)$.

Some Intuition: The Hiker

Suppose you are going on a hike with a one liter drinking bottle and you must completely fill the bottle before your hike. At the general store, you can fill your bottle with any combination of:

- ▶ Tap water: \$0.25 per liter
- ▶ Ice tea: \$0.50 per liter
- ▶ Soda: \$1.00 per liter
- ▶ Red Bull: \$5.00 per liter

You want to minimize your cost for the hike. What should you do?

We can define a per-liter cost vector

$$h = [0.25, 0.50, 1.00, 5.00]^T$$

and a proportional purchase vector $\nu \in \mathbb{R}^4$ with elements satisfying $\nu_i \geq 0$ and $\sum_i \nu_i = 1$. Clearly, your cost $h^T \nu$ is minimized when $\nu = [1, 0, 0, 0]^T$.

Solution to the Subproblem $\text{LPF}_\ell(\lambda)$

Theorem

For fixed $h \in \mathbb{R}^n$, the function $h^\top \nu$ is minimized over the set $\nu \in \{\mathbb{R}^n : \nu_i \geq 0 \text{ and } \sum_i \nu_i = 1\}$ by the vector ν^* with $\nu_m^* = 1$ and $\nu_k^* = 0$, $k \neq m$, where m is any index with $h_m = \min_{i \in \{0, \dots, n-1\}} h_i$.

Note that, if two or more commodities have the same minimum price, the solution is not unique. Back to our subproblem $\text{LPF}_\ell(\lambda)$...

$$d^* = \arg \min_{d \in \mathbb{R}^M} \sum_{i=0}^{M-1} d_i \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} P_{\ell j} \right]$$

subject to the constraints that $d_i \geq 0$ and $\sum_i d_i = 1$. The Theorem tells us how to easily solve $\text{LPF}_\ell(\lambda)$. We just have to find the index

$$m_\ell = \arg \min_{i \in \{0, \dots, M-1\}} \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} P_{\ell j} \right]$$

and then set $d_{m_\ell} = 1$ and $d_i = 0$ for all $i \neq m_\ell$.

Solution to the Problem $\text{LPF}(\lambda)$

Solving the subproblem $\text{LPF}_\ell(\lambda)$ for $\ell = 0, \dots, L - 1$ and assembling the results of the subproblems into $M \times L$ decision matrices yields a non-empty, finite set of **deterministic** decision matrices that solve $\text{LPF}(\lambda)$.

All deterministic decision matrices in the set have the same minimal cost:

$$r^*(\lambda) = \sum_{\ell=0}^{L-1} \min_i \left\{ \sum_{j=0}^{N-1} \lambda_j C_{ij} P_{\ell j} \right\}$$

It is easy to show that any randomized decision rule in the convex hull of this set of deterministic decision matrices also achieves $r^*(\lambda)$, and hence is also optimal.

Example Part 1

Let $G_{il} := \sum_{j=0}^{N-1} \lambda_j C_{ij} P_{lj}$ and $G \in \mathbb{R}^{M \times L}$ be the matrix composed of elements G_{il} . Suppose

$$G = \begin{bmatrix} 0.3 & 0.5 & 0.2 & 0.8 \\ 0.4 & 0.2 & 0.1 & 0.5 \\ 0.5 & 0.1 & 0.7 & 0.6 \end{bmatrix}$$

- ▶ What is the solution to the subproblem $\text{LPF}_0(\lambda)$? $d = [1, 0, 0]^\top$.
- ▶ What is the solution to the problem $\text{LPF}(\lambda)$?

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- ▶ Is this solution unique? Yes.
- ▶ What is $r^*(\lambda)$? $r^*(\lambda) = 1.0$.

Example Part 2

- ▶ What happens if

$$G = \begin{bmatrix} 0.3 & 0.5 & 0.2 & 0.8 \\ 0.4 & 0.2 & 0.2 & 0.5 \\ 0.5 & 0.1 & 0.7 & 0.6 \end{bmatrix} ?$$

The solution to $\text{LPF}_2(\lambda)$ is no longer unique. Both

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ or } D = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

solve $\text{LPF}(\lambda)$ and achieve $r^*(\lambda) = 1.1$.

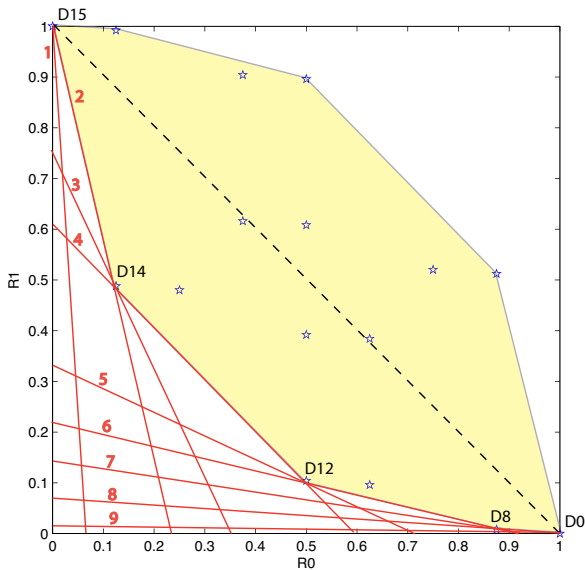
- ▶ In fact, any decision matrix of the form

$$D = \begin{bmatrix} 1 & 0 & \alpha & 0 \\ 0 & 0 & 1 - \alpha & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

for $\alpha \in [0, 1]$ will also achieve $r^*(\lambda) = 1.1$.

Remarks

- ▶ Note that the solution to the subproblem $\text{LPF}_\ell(\lambda)$ always yields at least one deterministic decision rule for the case $y = y_\ell$.
- ▶ If n_ℓ elements of the column G_ℓ are equal to the minimum of the column, then there are n_ℓ deterministic decision rules solving the subproblem $\text{LPF}_\ell(\lambda)$.
- ▶ There are $\prod_\ell n_\ell \geq 1$ deterministic decision rules that solve $\text{LPF}(\lambda)$.
- ▶ If there is more than one deterministic solution to $\text{LPF}(\lambda)$ then there will also be a corresponding convex hull of randomized decision rules that also achieve $r^*(\lambda)$.

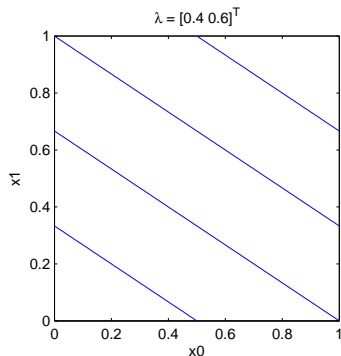
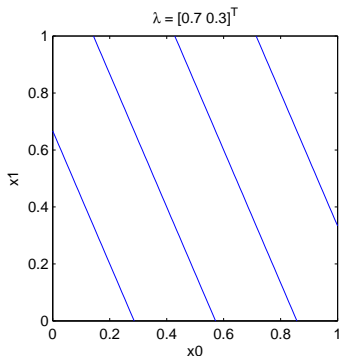
Working Example: Effect of Changing the Weighting λ 

Weighted Risk Level Sets

Notice in the last example that, by varying our risk weighting λ , we can create a family of “level sets” that cover every point of the optimal tradeoff surface.

Given a constant $c \in \mathbb{R}$, the **level set** of value c is defined as

$$L_c^\lambda := \{x \in R^N : \lambda^\top x = c\}$$



Infinite Observation Spaces: Part 1

When the observations are specified by conditional pdfs, our conditional risk function becomes

$$R_j(\rho) = \int_{y \in \mathcal{Y}} \left[\sum_{i=0}^{M-1} \rho_i(y) C_{ij} \right] p_j(y) dy$$

Recall that the randomized decision rule $\rho_i(y)$ specifies the probability of deciding \mathcal{H}_i when the observation is y .

We still have $N < \infty$ states, so, as before, we can write a linear combination of the conditional risk functions as

$$r(\rho, \lambda) := \sum_{j=0}^{N-1} \lambda_j R_j(\rho)$$

where $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$.

Infinite Observation Spaces: Part 2

The problem is then to solve

$$\begin{aligned} \rho^* &= \arg \min_{\rho \in \mathcal{D}} r(\rho, \lambda) \\ &= \arg \min_{\rho \in \mathcal{D}} \sum_{j=0}^{N-1} \lambda_j \int_{y \in \mathcal{Y}} \left[\sum_{i=0}^{M-1} \rho_i(y) C_{ij} \right] p_j(y) dy \end{aligned}$$

where \mathcal{D} is the set of all valid decision rules satisfying $\rho_i(y) \geq 0$ for all $i = 0, \dots, M-1$ and $y \in \mathcal{Y}$, as well as $\sum_{i=0}^{M-1} \rho_i(y) = 1$ for all $y \in \mathcal{Y}$.

- ▶ This is an infinite dimensional linear programming problem.
- ▶ We refer to this problem as $\text{LP}(\lambda)$.

Infinite Observation Spaces: Decomposition Part 1

As before, we can decompose the big problem $LP(\lambda)$ into lots of little subproblems, one for each observation $y \in \mathcal{Y}$. We rewrite $LP(\lambda)$ as

$$\begin{aligned} \rho^* &= \arg \min_{\rho \in \mathcal{D}} \sum_{j=0}^{N-1} \lambda_j \int_{y \in \mathcal{Y}} \left[\sum_{i=0}^{M-1} \rho_i(y) C_{ij} \right] p_j(y) dy \\ &= \arg \min_{\rho \in \mathcal{D}} \int_{y \in \mathcal{Y}} \left(\sum_{i=0}^{M-1} \rho_i(y) \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} p_j(y) \right] \right) dy \end{aligned}$$

To minimize the integral, we can minimize the term

$$(\dots) = \left(\sum_{i=0}^{M-1} \rho_i(y) \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} p_j(y) \right] \right)$$

at each fixed value of y .

Infinite Observation Spaces: Decomposition Part 2

The subproblem of minimizing the term (...) inside the integral when $y = y_0$ is a finite-dimensional linear programming problem (since \mathcal{X} is still finite):

$$\nu^* = \arg \min_{\nu} \sum_{i=0}^{M-1} \nu_i \left[\sum_{j=0}^{N-1} \lambda_j C_{ij} p_j(y_0) \right]$$

subject to $\nu_i \geq 0$ for all $i = 0, \dots, M-1$ and $\sum_{i=0}^{M-1} \nu_i = 1$.

We already know how to do this. We just find the index that has the lowest commodity cost

$$m_{y_0} = \arg \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} \lambda_j C_{ij} p_j(y_0)$$

and set $\nu_{m_{y_0}} = 1$ and all other $\nu_i = 0$. Same technique as solving $\text{LPF}_{\ell}(\lambda)$.

Infinite Observation Spaces: Minimum Weighted Risk

As we saw with finite \mathcal{Y} , solving $LP(\lambda)$ will yield at least one deterministic decision rule that achieves the minimum weighted risk

$$\begin{aligned}
 r^*(\lambda) &= \int_{y \in \mathcal{Y}} \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} \lambda_j C_{ij} p_j(y) dy \\
 &= \int_{y \in \mathcal{Y}} \min_{i \in \{0, \dots, M-1\}} g_i(y, \lambda) dy \\
 &= \sum_{i=0}^{M-1} \int_{y \in \mathcal{Y}_i^*} g_i(y, \lambda) dy
 \end{aligned}$$

1. The existence of at least one deterministic decision rule solving $LPF(\lambda)$ or $LP(\lambda)$ implies the existence of an **optimal partition** of \mathcal{Y} .
2. Since the problems $LPF(\lambda)$ and $LP(\lambda)$ always yield at least one deterministic decision rule, another way to think of these problems is: "Find a partition of the observation space that minimizes the weighted risk".

Infinite Observation Spaces: Putting it All Together

A decision rule that solves $LP(\lambda)$ is then

$$\rho_i^*(y) = \begin{cases} 1 & \text{if } i = \arg \min_{m \in \{0, \dots, M-1\}} g_m(y, \lambda) \\ 0 & \text{otherwise} \end{cases}$$

where $\rho^*(y) = [\rho_0^*(y), \dots, \rho_{M-1}^*(y)]^\top$.

Remarks:

1. The solution to $LP(\lambda)$ may not be unique.
2. If there is more than one deterministic solution to $LP(\lambda)$ then there will also be a corresponding convex hull of randomized decision rules that also achieve $r^*(\lambda)$.
3. Note that the standard convention for deterministic decision rules is $\delta : \mathcal{Y} \mapsto \mathcal{Z}$. An equivalent way of specifying a deterministic decision rule that solves $LP(\lambda)$ is

$$\delta^*(y) = \arg \min_{i \in \{0, \dots, M-1\}} g_i(y, \lambda).$$

The Bayesian Approach

We assume a prior state distribution $\pi \in \mathcal{P}_N$ such that

$$\text{Prob}(\text{state is } x_j) = \pi_j$$

Note that this is our model of the state probabilities **prior to the observation**.

We denote the Bayes Risk of the decision rule ρ as

$$r(\rho, \pi) = \sum_{j=0}^{N-1} \pi_j R_j(\rho) = \sum_{j=0}^{N-1} \pi_j \int_{y \in \mathcal{Y}} p_j(y) \sum_{i=0}^{M-1} C_{ij} \rho_i(y) dy.$$

This is simply the weighted overall risk, or average risk, given our prior belief of the state probabilities. A decision rule that minimizes this risk is called a Bayes decision rule for the prior π .

Solving Bayesian Hypothesis Testing Problems

We know how to solve this problem. For finite \mathcal{Y} , we just find the index

$$m_\ell = \arg \min_{i \in \{0, \dots, M-1\}} \underbrace{\sum_{j=0}^{N-1} \pi_j C_{ij} P_{\ell j}}_{G_{i\ell}}$$

and set $D_{m_\ell, \ell}^{B\pi} = 1$ and $D_{i, \ell}^{B\pi} = 0$ for all $i \neq m_\ell$.

For infinite \mathcal{Y} , we just find the index

$$m_{y_0} = \arg \min_{i \in \{0, \dots, M-1\}} \underbrace{\sum_{j=0}^{N-1} \pi_j C_{ij} p_j(y_0)}_{g_i(y_0, \pi)}$$

and set $\delta^{B\pi}(y_0) = m_{y_0}$ (or, equivalently, set $\rho_{m_{y_0}}^{B\pi}(y_0) = 1$ and $\rho_i^{B\pi}(y_0) = 0$ for all $i \neq m_{y_0}$).

Prior and Posterior Probabilities

The conditional probability that \mathcal{H}_j is true given the observation y :

$$\pi_j(y) := \text{Prob}(\mathcal{H}_j \text{ is true} \mid Y = y) = \frac{p_j(y)\pi_j}{p(y)}$$

where

$$p(y) = \sum_{j=0}^{N-1} \pi_j p_j(y).$$

- ▶ Recall that π_j is the **prior** probability of choosing \mathcal{H}_j , before we have any observations.
- ▶ The quantity $\pi_j(y)$ is the **posterior** probability of choosing \mathcal{H}_j , conditioned on the observation y .

A Bayes Decision Rule Minimizes The Posterior Cost

General expression for a deterministic Bayes decision rule:

$$\delta^{B\pi}(y) = \arg \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} \pi_j C_{ij} p_j(y)$$

But $\pi_j p_j(y) = \pi_j(y) p(y)$, hence

$$\begin{aligned} \delta^{B\pi}(y) &= \arg \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} C_{ij} \pi_j(y) p(y) \\ &= \arg \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} C_{ij} \pi_j(y) \end{aligned}$$

since $p(y)$ does not affect the minimizer. What does this mean?

$\sum_{j=0}^{N-1} C_{ij} \pi_j(y)$ the average cost of choosing hypothesis \mathcal{H}_i given $Y = y$, i.e. the **posterior** cost of choosing \mathcal{H}_i . **The Bayes decision rule chooses the hypothesis that yields the minimum expected posterior cost.**

Bayesian Hypothesis Testing with UCA: Part 1

The uniform cost assignment (UCA):

$$C_{ij} = \begin{cases} 0 & \text{if } x_j \in \mathcal{H}_i \\ 1 & \text{otherwise} \end{cases}$$

The conditional risk $R_j(\rho)$ under the UCA is simply the probability of not choosing the hypothesis that contains x_j , i.e. the probability of error when the state is x_j . The Bayes risk in this case is

$$r(\rho, \pi) = \sum_{j=0}^{N-1} \pi_j R_j(\rho) = \text{Prob}(\text{error}).$$

Bayesian Hypothesis Testing with UCA: Part 2

Under the UCA, a Bayes decision rule can be written in terms of the posterior probabilities as

$$\begin{aligned}
 \delta^{B\pi}(y) &= \arg \min_{i \in \{0, \dots, M-1\}} \sum_{j=0}^{N-1} C_{ij} \pi_j(y) \\
 &= \arg \min_{i \in \{0, \dots, M-1\}} \sum_{x_j \notin \mathcal{H}_i} \pi_j(y) \\
 &= \arg \min_{i \in \{0, \dots, M-1\}} \left[1 - \sum_{x_j \in \mathcal{H}_i} \pi_j(y) \right] \\
 &= \arg \max_{i \in \{0, \dots, M-1\}} \sum_{x_j \in \mathcal{H}_i} \pi_j(y)
 \end{aligned}$$

Hence, for hypothesis tests under the UCA, **the Bayes decision rule is the MAP (maximum a posteriori) decision rule**. When the hypothesis test is simple, $\delta^{B\pi}(y) = \arg \max_i \pi_i(y)$.

Bayesian Hypothesis Testing with UCA and Uniform Prior

Under the UCA and a uniform prior, i.e. $\pi_j = 1/N$ for all $j = 0, \dots, N - 1$

$$\begin{aligned}
 \delta^{B\pi}(y) &= \arg \max_{i \in \{0, \dots, M-1\}} \sum_{x_j \in \mathcal{H}_i} \pi_j(y) \\
 &= \arg \max_{i \in \{0, \dots, M-1\}} \sum_{x_j \in \mathcal{H}_i} \pi_j p_j(y) \\
 &= \arg \max_{i \in \{0, \dots, M-1\}} \sum_{x_j \in \mathcal{H}_i} p_j(y)
 \end{aligned}$$

since $\pi_j = 1/N$ does not affect the minimizer.

- ▶ In this case, $\delta^{B\pi}(y)$ selects the **most likely** hypothesis, i.e. the hypothesis which best explains the observation y .
- ▶ This is called the **maximum likelihood** (ML) decision rule.
- ▶ Which is better, MAP or ML?

Simple Binary Bayesian Hypothesis Testing: Part 1

We have two states x_0 and x_1 and two hypotheses \mathcal{H}_0 and \mathcal{H}_1 . For each $y \in \mathcal{Y}$, our problem is to compute

$$m_y = \arg \min_{i \in \{0,1\}} g_i(y, \pi)$$

where

$$g_0(y, \pi) = \pi_0 C_{00} p_0(y) + \pi_1 C_{01} p_1(y)$$

$$g_1(y, \pi) = \pi_0 C_{10} p_0(y) + \pi_1 C_{11} p_1(y)$$

We only have two things to compare. We can simplify this comparison:

$$\begin{aligned} g_0(y, \pi) \geq g_1(y, \pi) &\Leftrightarrow \pi_0 C_{00} p_0(y) + \pi_1 C_{01} p_1(y) \geq \pi_0 C_{10} p_0(y) + \pi_1 C_{11} p_1(y) \\ &\Leftrightarrow p_1(y) \pi_1 (C_{01} - C_{11}) \geq p_0(y) \pi_0 (C_{10} - C_{00}) \\ &\Leftrightarrow \frac{p_1(y)}{p_0(y)} \geq \frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})} \end{aligned}$$

where we have assumed that $C_{01} > C_{11}$ to get the final result. The expression $L(y) := \frac{p_1(y)}{p_0(y)}$ is known as the **likelihood ratio**.

Simple Binary Bayesian Hypothesis Testing: Part 2

Given $L(y) := \frac{p_1(y)}{p_0(y)}$ and

$$\tau := \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})},$$

our Bayes decision rule is then simply

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

Remark:

- ▶ For any $y \in \mathcal{Y}$ that result in $L(y) = \tau$, the Bayes risk is the same whether we decide \mathcal{H}_0 or \mathcal{H}_1 . You can deal with this by always deciding \mathcal{H}_0 in this case, or always deciding \mathcal{H}_1 , or flipping a coin, etc.

Performance of Simple Binary Bayesian Hypothesis Testing

- ▶ Since the observation Y is a random variable, so is the likelihood ratio $L(Y)$. Assume that $L(Y)$ has a conditional pdf $\Lambda_j(t)$ when the state is x_j .
- ▶ When the state is x_0 , the probability of making an error (deciding \mathcal{H}_1) is

$$P_{\text{false positive}} = \text{Prob}(L(y) > \tau \mid \text{state is } x_0) = \int_{\tau}^{\infty} \Lambda_0(t) dt$$

- ▶ Similarly, when the state is x_1 , the probability of making an error (deciding \mathcal{H}_0) is

$$P_{\text{false negative}} = \text{Prob}(L(y) \leq \tau \mid \text{state is } x_1) = \int_{-\infty}^{\tau} \Lambda_1(t) dt$$

What is the overall (unconditional) probability of error?

Simple Binary Bayesian Hypothesis Testing with UCA

Uniform cost assignment:

$$C_{00} = C_{11} = 0$$

$$C_{01} = C_{10} = 1$$

In this case, the discriminant functions are simply

$$g_0(y, \pi) = \pi_1 p_1(y) = \pi_1(y) p(y)$$

$$g_1(y, \pi) = \pi_0 p_0(y) = \pi_0(y) p(y)$$

and a Bayes decision rule can be written in terms of the posterior probabilities as

$$\delta^{B\pi}(y) = \begin{cases} 1 & \text{if } \pi_1(y) > \pi_0(y) \\ 0/1 & \text{if } \pi_1(y) = \pi_0(y) \\ 0 & \text{if } \pi_1(y) < \pi_0(y). \end{cases}$$

In this case, it should be clear that the Bayes decision rule is the MAP (maximum a posteriori) decision rule.

Example: Coherent Detection of BPSK

Suppose a transmitter sends one of two scalar signals and the signals arrive at a receiver corrupted by zero-mean additive white Gaussian noise (AWGN) with variance σ^2 . We want to use Bayesian hypothesis testing to determine which signal was sent.

Signal model conditioned on state x_j :

$$Y = a_j + \eta$$

where a_j is the scalar signal and $\eta \sim \mathcal{N}(0, \sigma^2)$. Hence

$$p_j(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - a_j)^2}{2\sigma^2}\right)$$

Hypotheses:

$$\mathcal{H}_0 : a_0 \text{ was sent, or } Y \sim \mathcal{N}(a_0, \sigma^2)$$

$$\mathcal{H}_1 : a_1 \text{ was sent, or } Y \sim \mathcal{N}(a_1, \sigma^2)$$

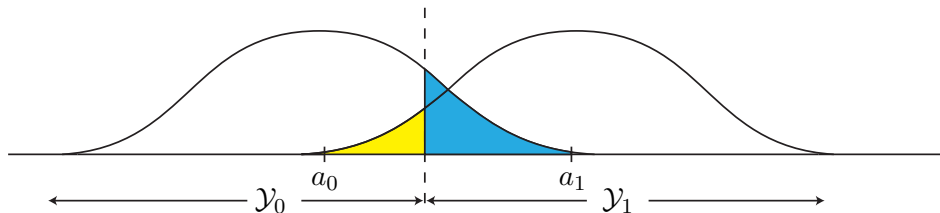
Example: Coherent Detection of BPSK

This is just a simple binary hypothesis testing problem. Under the UCA, we can write

$$R_0(\rho) = \int_{y \in \mathcal{Y}_1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - a_0)^2}{2\sigma^2}\right) dy$$

$$R_1(\rho) = \int_{y \in \mathcal{Y}_0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - a_1)^2}{2\sigma^2}\right) dy$$

where $\mathcal{Y}_j = \{y \in \mathcal{Y} : \rho_j(y) = 1\}$. Intuitively,



Example: Coherent Detection of BPSK

Given a prior π_0 , $\pi_1 = 1 - \pi_0$, we can write the Bayes risk as

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

The Bayes decision rule can be found by computing the likelihood ratio and comparing it to the threshold τ :

$$\begin{aligned} \frac{p_1(y)}{p_0(y)} > \tau &\Leftrightarrow \frac{\exp\left(\frac{-(y-a_1)^2}{2\sigma^2}\right)}{\exp\left(\frac{-(y-a_0)^2}{2\sigma^2}\right)} > \frac{\pi_0}{\pi_1} \\ &\Leftrightarrow \frac{(y-a_0)^2 - (y-a_1)^2}{2\sigma^2} > \ln \frac{\pi_0}{\pi_1} \\ &\Leftrightarrow y > \frac{a_0 + a_1}{2} + \frac{\sigma^2}{a_1 - a_0} \ln \frac{\pi_0}{\pi_1} \end{aligned}$$

Example: Coherent Detection of BPSK

Let $\gamma := \frac{a_0 + a_1}{2} + \frac{\sigma^2}{a_1 - a_0} \ln \frac{\pi_0}{\pi_1}$. Our Bayes decision rule is then :

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

Using this decision rule, the Bayes risk is then

$$r(\delta^{B\pi}, \pi) = \pi_0 Q\left(\frac{\gamma - a_0}{\sigma}\right) + (1 - \pi_0) Q\left(\frac{a_1 - \gamma}{\sigma}\right)$$

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

- ▶ When $\pi_0 = \pi_1 = \frac{1}{2}$, the decision boundary is simply $\frac{a_0 + a_1}{2}$, i.e. the midpoint between a_0 and a_1 and $r(\delta^{B\pi}, \pi) = Q\left(\frac{a_1 - a_0}{2\sigma}\right)$.
- ▶ When σ is very small, the prior has little effect on the decision boundary. When σ is very large, the prior becomes more important.

Composite Binary Bayesian Hypothesis Testing

We have $N > 2$ states x_0, \dots, x_{N-1} and two hypotheses \mathcal{H}_0 and \mathcal{H}_1 that form a valid partition on these states. For each $y \in \mathcal{Y}$, we compute

$$m_y = \arg \min_{i \in \{0,1\}} g_i(y, \pi), \text{ where } g_i(y, \pi) = \sum_{j=0}^{N-1} \pi_j C_{ij} p_j(y)$$

As was the case of simple binary Bayesian hypothesis testing, we only have two things to compare. The comparison boils down to

$$g_0(y, \pi) \geq g_1(y, \pi) \Leftrightarrow J(y) := \frac{\sum_j \pi_j C_{0j} p_j(y)}{\sum_j \pi_j C_{1j} p_j(y)} \geq 1.$$

Hence, a Bayes decision rule for the composite binary hypothesis testing problem is then simply

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

Composite Binary Bayesian Hypothesis Testing with UCA

For $i = 0, 1$ and $j = 0, \dots, N - 1$, the UCA is:

$$C_{ij} = \begin{cases} 0 & \text{if } x_j \in \mathcal{H}_i \\ 1 & \text{otherwise} \end{cases}$$

Hence, the discriminant functions are

$$g_i(y, \pi) = \sum_{x_j \in \mathcal{H}_i^c} \pi_j p_j(y) = 1 - \sum_{x_j \in \mathcal{H}_i} \pi_j p_j(y) = 1 - \sum_{x_j \in \mathcal{H}_i} \pi_j(y) p(y)$$

Under the UCA, the hypothesis test comparison boils down to

$$g_0(y, \pi) \geq g_1(y, \pi) \Leftrightarrow J(y) := \frac{\sum_{x_j \in \mathcal{H}_1} \pi_j(y)}{\sum_{x_j \in \mathcal{H}_0} \pi_j(y)} \geq 1.$$

and the Bayes decision rule $\delta^{B\pi}(y)$ is the same as before. Note that

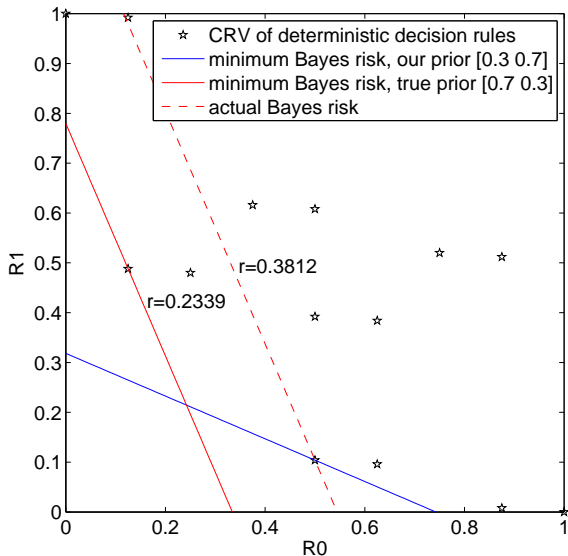
$$J(y) := \frac{\sum_{x_j \in \mathcal{H}_1} \pi_j(y)}{\sum_{x_j \in \mathcal{H}_0} \pi_j(y)} = \frac{\text{Prob}(\mathcal{H}_1|y)}{\text{Prob}(\mathcal{H}_0|y)}$$

so, as before, the Bayes decision rule is the MAP decision rule.

Final Remarks on Bayesian Hypothesis Testing

1. A Bayes decision rule minimizes the Bayes risk $r(\rho, \pi) = \sum_j \pi_j R_j(\rho)$ under the prior state distribution π .
2. At least one deterministic decision rule $\delta^{B\pi}(y)$ achieves the minimum Bayes risk for any prior state distribution π .
3. The prior state distribution π can also be thought of as a parameter.
 - ▶ By varying π , we can generate a family of Bayes decision rules $\delta^{B\pi}$ parameterized by π .
 - ▶ This family of Bayes decision rules can achieve any operating point on the optimal tradeoff surface (or ROC curve).
4. In some applications, a prior state distribution is difficult to determine.

Bayesian Hypothesis Testing with an Incorrect Prior



Conclusions and Motivation for the Minimax Criterion

- ▶ If the true prior state distribution π_{true} is far from our belief π , a Bayes detector $\delta^{B\pi}$ may result in a significant loss in performance.
- ▶ One way to select a decision rule that is robust with respect to uncertainty in the prior is to minimize $\max_{\pi} r(\rho, \pi) = \max_j R_j(\rho)$.
- ▶ The “minimax” criterion leads to a decision rule that minimizes the Bayes risk for the **least favorable prior** distribution.
- ▶ This “minimax” criterion is conservative, but it allows one to guarantee performance over all possible priors.