BAYES RISK DETECTION

Consider a <u>binary</u> hypothesis testing problem involving <u>simple</u> hypotheses:

$$H_0 : \underline{X} \sim f_0(\underline{x})$$

$$H_1 : \underline{X} \sim f_1(\underline{x})$$

We assume for every observation $\underline{x}$, exactly one of the two models is true.

Let's view the "active hypothesis" as a random event:

$$\pi_0 := \text{probability that } H_0 \text{ is in effect}$$

$$\pi_1 := \quad " \quad " \quad H_1 \quad " \quad " \quad "$$

Obviously we have

$$\pi_0 + \pi_1 = 1.$$

How should we measure the performance of a decision rule?

Suppose we have a decision rule defined by the *decision regions* R_0 and R_1.

$$\underline{x} \in R_0 \iff \text{declare } H_0 \text{ is in effect}$$

$$\underline{x} \in R_1 \iff \text{declare } H_1 \text{ is in effect}$$

There are four possible outcomes:

decision		H_0 (truth)	H_1 (truth)
	$\underline{x} \in R_0$	$(0,0)$	$(0,1)$
	$\underline{x} \in R_1$	$(1,0)$	$(1,1)$

Suppose we are able to specify

$$c_{ij} := \underline{\text{cost}} \text{ of declaring } H_i \text{ when } H_j \text{ true}$$

To be sensible, we should have

$$c_{ii} < c_{ij} \quad , \quad i \neq j$$

Define

$$\bar{C} = \text{expected cost of a decision}$$

$$= \sum_{i,j=0}^{1} C_{ij} \cdot P(\text{declare } H_i,\ H_j \text{ true})$$

$$= \sum_{i,j=0}^{1} C_{ij} \cdot P(H_j \text{ true}) \cdot P(\text{declare } H_i \mid H_j \text{ true})$$

$$= \sum_{i,j=0}^{1} C_{ij}\, \pi_j\, P(H_i \mid H_j)$$

where

$$P(H_i \mid H_j) := \text{probability that } \underline{X} \in R_i$$
$$\text{when } \underline{X} \sim f_j(\underline{x})$$

We will design a detector that minimizes the expected cost of a decision.

First, let's think about the probabilities $P(H_i / H_j)$.

Example | Consider a scalar observation

$$H_0 : X \sim \mathcal{N}(-1, 1)$$

$$H_1 : X \sim \mathcal{N}(1, 1)$$

If our decision regions are

$$R_0 = (-\infty, 0]$$

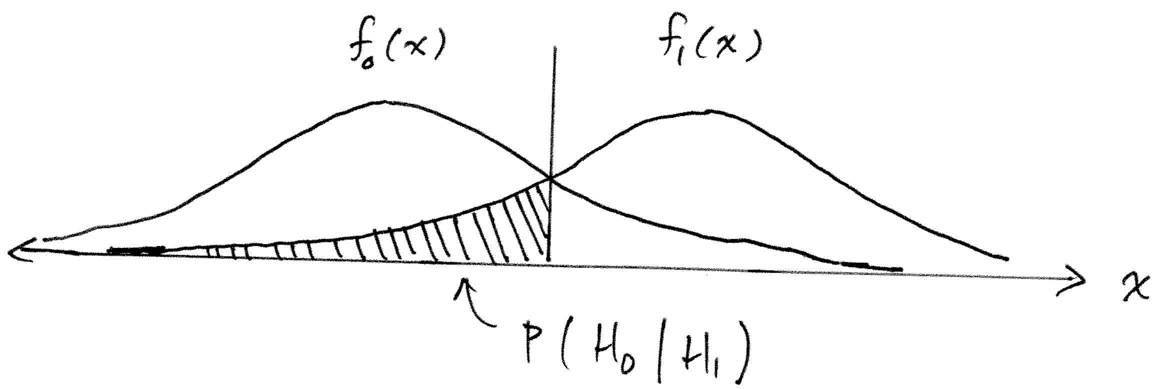
$$R_1 = (0, \infty)$$

then

$$P(H_0 \mid H_1) = P(X \in R_0 \mid H_1)$$

(a) $=$

Picture:



Detection as Bayesian Estimation

The detection problem described above can easily be viewed as a Bayesian estimation problem.

The parameter of interest is the true hypothesis, $\phi \in \{0, 1\}$.

The estimator $\hat{\phi}$ is the detector.

The cost is the loss

$$c_{ij} = L(i, j)$$

The expected cost of a decision is the Bayes risk

$$\bar{C} = E_{(\phi, \underline{X})}\left[L(\phi, \hat{\phi}(\underline{X})) \right]$$

The prior on ϕ is given by π_0, π_1.

The decision rule minimizing the Bayes risk is called the <u>Bayes risk detector</u>

Bayes Risk Detector: Continuous Case

Assume $\underline{X}$ is continuous under both H_0 and H_1. That is, assume $f_0(\underline{x})$ and $f_1(\underline{x})$ are densities.

We may write

$$\bar{C} = \sum_{i,j=0}^{1} c_{ij}\, \pi_j\, P(H_i \mid H_j)$$

$$= \sum_{i,j} c_{ij}\, \pi_j \int_{R_i} f_j(\underline{x})\, d\underline{x}$$

$$= \int_{R_0} \left(c_{00}\, \pi_0\, f_0(\underline{x}) + c_{01}\, \pi_1\, f_1(\underline{x}) \right) d\underline{x}$$

$$+ \int_{R_1} \left(c_{10}\, \pi_0\, f_0(\underline{x}) + c_{11}\, \pi_1\, f_1(\underline{x}) \right) d\underline{x}$$

How should we choose R_0 and R_1 to minimize this expression?

Recall:

$$\left.\begin{aligned} R_0 \cap R_1 &= \emptyset \\ R_0 \cup R_1 &= \mathbb{R}^N \end{aligned}\right\} \text{Partition of } \mathbb{R}^N$$

So every $\underline{x} \in \mathbb{R}^N$ is in one and only one R_i.

To minimize the Bayes risk, therefore, choose $\underline{x} \in R_i$ when the corresponding <u>integrand</u> is smaller.

That is, choose

$$\underline{x} \in R_0 \iff c_{00}\,\pi_0\, f_0(\underline{x}) + c_{01}\,\pi_1\, f_1(\underline{x}) < c_{10}\,\pi_0\, f_0(\underline{x}) + c_{11}\,\pi_1\, f_1(\underline{x})$$

$$\iff \frac{f_1(\underline{x})}{f_0(\underline{x})} < \frac{\pi_0}{\pi_1} \cdot \frac{(c_{10} - c_{00})}{(c_{01} - c_{11})}$$

More concisely, we may express the Bayes risk detector as:

$$\boxed{\frac{f_1(\underline{x})}{f_0(\underline{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1} \cdot \frac{(c_{10} - c_{00})}{(c_{01} - c_{11})}}$$

Likelihood Ratio Tests

The Bayes risk detector is an example of a likelihood ratio test (LRT)

A LRT has the form

$$\Lambda(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where

$$\Lambda(\underline{x}) := \frac{f_1(\underline{x})}{f_0(\underline{x})}$$

is the likelihood ratio and $\eta > 0$ is a threshold.

Bayes Risk Detector: Discrete Case

Now suppose f_0 and f_1 are mass functions, and let $\mathcal{X}$ denote the domain of $\underline{X}$.

Then

$$\bar{c} = \sum_{i,j} c_{ij}\, \pi_j\, P(H_i \mid H_j)$$

(b)
$$= \sum_{i,j} c_{ij}\, \pi_j$$

$$= \sum_{\underline{x} \in \mathcal{X} \cap R_0} \left(c_{00}\, \pi_0\, f_0(\underline{x}) + c_{01}\, \pi_1\, f_1(\underline{x}) \right)$$
$$+ \sum_{\underline{x} \in \mathcal{X} \cap R_1} \left(c_{10}\, \pi_0\, f_0(\underline{x}) + c_{11}\, \pi_1\, f_1(\underline{x}) \right)$$

Choosing R_0, R_1 to minimize this expression we once again obtain

$$\boxed{\frac{f_1(\underline{x})}{f_0(\underline{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1} \cdot \frac{(c_{10} - c_{00})}{(c_{01} - c_{11})}}$$

Minimum Probability of Error Detector

An important special case of the Bayes detector occurs when

$$c_{ij} = 1 - \delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

Then the Bayes risk is

$$\begin{aligned} \bar{C} &= \sum_{i,j} c_{ij} \, P(\text{declare } H_i, H_j \text{ true}) \\ &= P(\text{declare } H_0, H_1 \text{ true}) \\ &\quad + P(\text{declare } H_1, H_0 \text{ true}) \\ &= P(\text{decision} \neq \text{truth}) \\ &= \underline{\text{probability of error}} =: P_E \end{aligned}$$

The "min P_E" detector is therefore

Example 1 Consider the problem of detecting a DC signal with amplitude $A > 0$ in additive white Gaussian noise.

$$H_0: \ X_i = W_i, \qquad i = 1, \dots, N$$

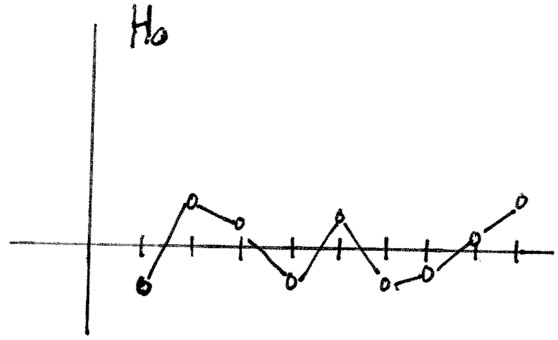
$$H_1: \ X_i = A + W_i, \qquad i = 1, \dots, N$$

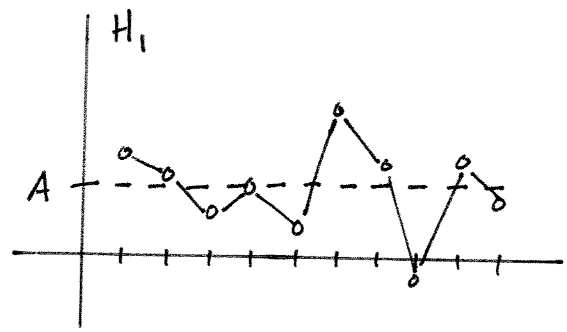
where $W_i \overset{iid}{\sim} N(0, \sigma^2)$ and A, σ^2 are known.

Equivalently we could write

$$H_0: \ \underline{X} \sim N(\underline{0}, \sigma^2 I)$$

$$H_1: \ \underline{X} \sim N(A \cdot \underline{1}, \sigma^2 I)$$





What is the Bayes risk detector? Any guesses?

We have

$$\Lambda(\underline{x}) = \frac{f_1(\underline{x})}{f_0(\underline{x})} = \frac{\prod_{n=1}^{N} f_1(x_n)}{\prod_{n=1}^{N} f_0(x_n)}$$

$$= \frac{\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(x_n-A)^2}{2\sigma^2}\right\}}{\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x_n^2}{2\sigma^2}\right\}}$$

$$= \frac{\exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(x_n-A)^2\right\}}{\exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}x_n^2\right\}}$$

$$= \exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(-2x_nA + A^2)\right\}$$

$$= \exp\left\{\frac{A}{\sigma^2}\sum_{n=1}^{N}x_n - \frac{NA^2}{2\sigma^2}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Can we simplify the detector further?

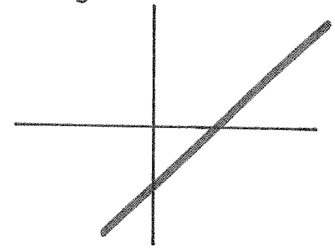
Monotonic Transformations

If we apply a monotonically increasing function to both sides of the LRT, the decision regions remain the same.

Examples

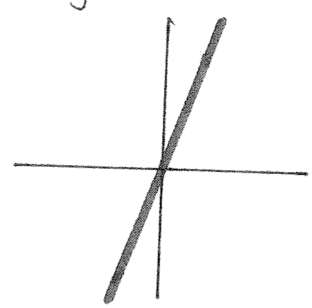
1. adding a number

$y = x + c, \; c \in \mathbb{R}$



2. multiplying by a positive number

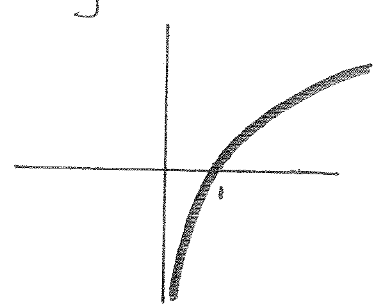
$y = cx, \; c > 0$



3. natural logarithm

Commonly used, since many densities and mass functions have an exponential form

$y = \ln(x)$



(DC signal detection, continued)

$$\exp\left\{ \frac{A}{\sigma^2} \sum_{n=1}^{N} x_n - \frac{NA^2}{2\sigma^2} \right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

$$\Updownarrow$$

$$\frac{A}{\sigma^2} \sum_{n=1}^{N} x_n - \frac{NA^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\eta)$$

$$\Updownarrow$$

$$\frac{1}{N} \sum_{n=1}^{N} x_n \underset{H_0}{\overset{H}{\gtrless}} \frac{\sigma^2}{NA} \ln(\eta) + \frac{A}{2} \equiv \gamma$$

Thus, the detector reduces to a simple <u>thresholding test</u> involving the sample mean.

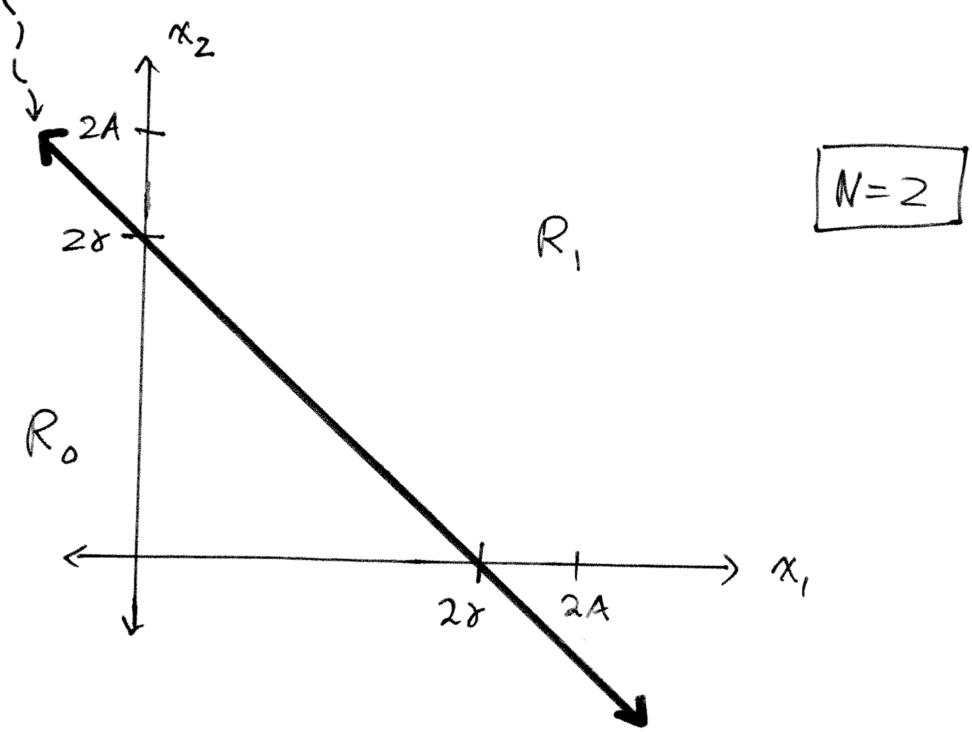
<u>Note</u> If $\eta = 1$ $(\pi_0 = \pi_1 = \frac{1}{2})$, then $\gamma = \frac{A}{2}$, and σ^2 need not be known.

Where did we use the assumption $A > 0$?

Note that the Bayes risk detector is a <u>linear</u> detector, meaning it is obtained by thresholding a linear function of the data.

Equivalently, the <u>decision boundary</u> is a hyperplane

$$\text{decision boundary} = \left\{ \underline{x} \in \mathbb{R}^N : \langle \underline{x}, \tfrac{1}{N}\underline{1} \rangle = \gamma \right\}$$



Calculating Error Probabilities

How can we calculate $P(H_i \mid H_j)$ to assess the performance of a detector?

For example, the probability of error is

$$P_E = P(H_0, H_1) + P(H_1, H_0)$$

$$= \pi_1 P(H_0 \mid H_1) + \pi_0 P(H_1 \mid H_0)$$

$$= \pi_1 \int_{R_0} f_1(\underline{x})\, d\underline{x} + \pi_0 \int_{R_1} f_0(\underline{x})\, d\underline{x}$$

We must compute <u>N-dimensional integrals</u>.

This is a daunting task, even for the relatively simple case of Gaussian noise and linear decision boundaries.

Fortunately, we can use

- simplified test statistics
- monotone transformations

to make our lives easier.

In the previous example

$$t = \frac{1}{N} \sum_{i=1}^{N} x_i$$

is an example of a <u>test statistic</u>.

More generally, a test statistic is simply a statistic (i.e., a function of the data) that is used in a test/detector.

The importance of test statistics for error calculation is that often they

- are 1-dimensional
- have known distributions

and can therefore be used in place of the N-dimensional data.

Example (continued)

We had

$$\frac{1}{N}\sum_{n=1}^{N} x_n = t \underset{H_0}{\overset{H_1}{\gtrless}} \gamma = \frac{\sigma^2}{NA}\ln(\eta) + \frac{A}{2}$$

Recall

$$\underline{x} \sim N(\underline{0}, \sigma^2 I) \quad \text{under } H_0$$

$$\underline{x} \sim N(A\underline{1}, \sigma^2 I) \quad \text{under } H_1$$

Now

$$t = B\underline{x}, \quad B = \left[\tfrac{1}{N} \cdots \tfrac{1}{N}\right]$$

(d) So

$T \sim$

$\sim$ under H_0

and

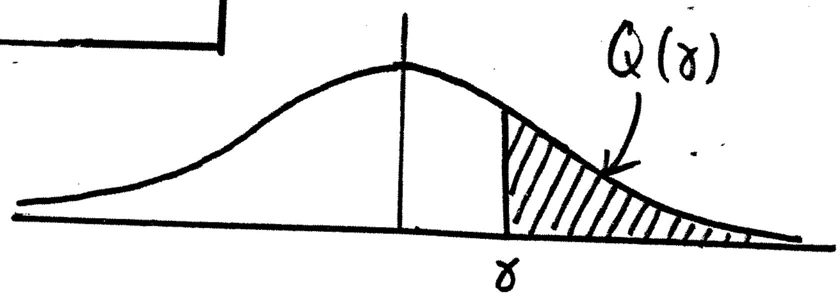
$T \sim$

$\sim$ under H_1

The Q-function

Let $X \sim N(0,1)$. Define

$$\boxed{\begin{aligned} Q(\gamma) &\equiv P(X \geq \gamma) \\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \end{aligned}}$$



If $X \sim N(\mu, \sigma^2)$, then

$$P(X \geq \gamma) = Q\left(\frac{\gamma - \mu}{\sigma}\right)$$

show this by change of variables argument

Note: $Q : \mathbb{R} \to (0,1)$ is monotonically decreasing, so it has an inverse.

In Matlab

$$Q(\gamma) = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{\gamma}{\sqrt{2}}\right)\right)$$

$$Q^{-1}(\alpha) = \sqrt{2}\ \mathrm{erfinv}(1 - 2\alpha)$$

Under H_0, $T \sim N(0, \frac{\sigma^2}{N})$, so

so

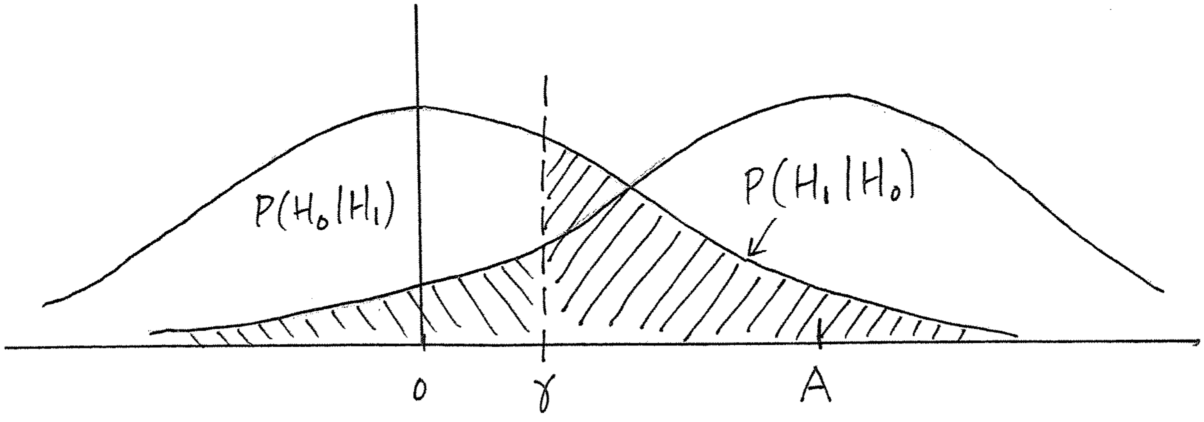
$$P(H_1 | H_0) = P(T > \gamma \mid H_0)$$
$$= Q\left(\frac{\gamma}{\sigma/\sqrt{N}}\right)$$

Under H_1, $T \sim N(A, \frac{\sigma^2}{N})$, so

$$P(H_0 | H_1) = P(T < \gamma \mid H_1)$$
$$= 1 - Q\left(\frac{\gamma - A}{\sigma/\sqrt{N}}\right).$$

Therefore,

$$P_E = \pi_0 P(H_1 | H_0) + \pi_1 P(H_0 | H_1)$$
$$= \pi_0 Q\left(\frac{\gamma}{\sigma/\sqrt{N}}\right) + \pi_1 \left(1 - Q\left(\frac{\gamma - A}{\sigma/\sqrt{N}}\right)\right)$$



Recall $\gamma = \frac{\sigma^2}{NA} \ln(\eta) + \frac{A}{2}$.

If $\pi_0 = \pi_1 = 1/2$ $(\eta = 1)$, then

$$\boxed{P_E = Q\left(\frac{A\sqrt{N}}{2\sigma}\right)}$$

For this problem we may define the <u>signal to noise ratio</u>

$$SNR = \frac{A^2 N}{\sigma^2}$$

Then

$$\text{smaller } P_E \Longleftrightarrow \text{larger } SNR$$

The MAP Detector

Instead of minimizing P_E, we could maximize P_c:

P_c = probability of a correct decision

$$= P(H_0, H_0) + P(H_1, H_1)$$

$$= \pi_0 \int_{R_0} f_0(\underline{x})\, d\underline{x} + \pi_1 \int_{R_1} f_1(\underline{x})\, d\underline{x}$$

So we would choose

$$\underline{x} \in R_i \iff \pi_i f_i(\underline{x}) \text{ is maximal}$$

<u>Note</u>: this just another way of writing the LRT

By Bayes rule,

$$P(\mathcal{H}_i | \underline{x}) = \frac{P(\mathcal{H}_i) \cdot f(\underline{x} | \mathcal{H}_i)}{f(\underline{x})}$$

$$= \frac{\pi_i f_i(\underline{x})}{f(\underline{x})}$$

same thing, different notation

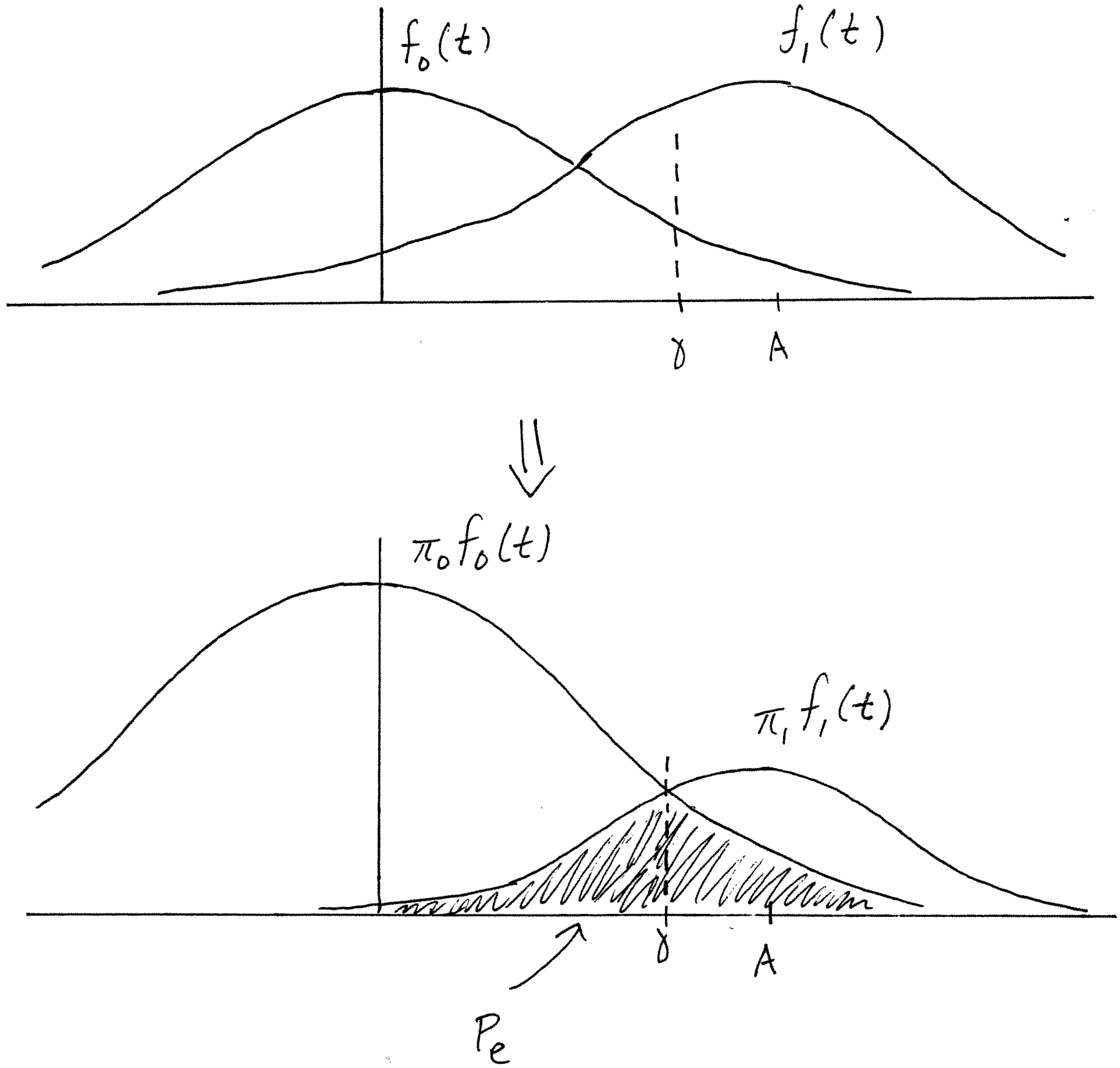
We call $P(\mathcal{H}_i | \underline{x})$ the <u>a posteriori</u> (or <u>posterior</u>) <u>probability</u> of hypothesis $\mathcal{H}_i$.

Since $f(\underline{x})$ is independent of i, maximizing $\pi_i f_i(\underline{x})$ is equivalent to maximizing $P(\mathcal{H}_i | \underline{x})$. This gives rise to the <u>maximum a posteriori probability</u> (MAP) detector:

$$\underline{x} \in R_i \iff P(\mathcal{H}_i | \underline{x}) \text{ is maximal}$$

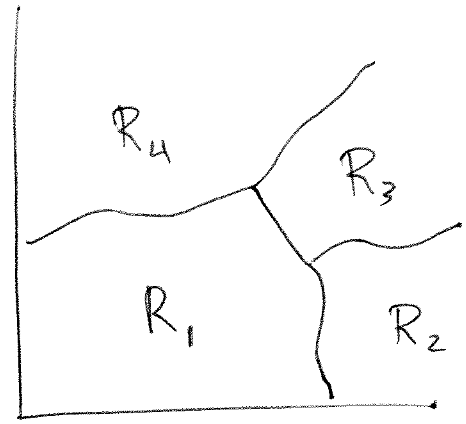
Example 1

Assume $\pi_0 > \pi_1$



Multiple Hypotheses

Consider



$$\mathcal{H}_1 : \underline{X} \sim f_1(\underline{x})$$

$$\vdots$$

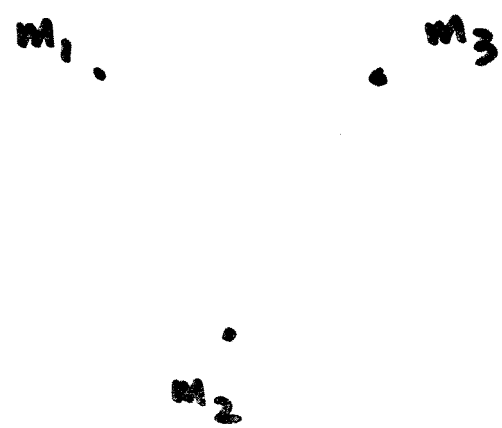
$$\mathcal{H}_M : \underline{X} \sim f_M(\underline{x})$$

Then $\quad P_e = 1 - P_c = 1 - \left(\sum_i \int_{R_i} \pi_i f_i(\underline{x}) \, d\underline{x} \right)$

$\Rightarrow$ MAP detector is optimal:

$$\boxed{x \in R_i \iff \pi_i f_i(\underline{x}) \text{ is maximal}}$$

Example $\quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N(\underline{m}_i, I), \quad \pi_i = \frac{1}{3}$



Summary

- Bayes detector: minimizes Bayes risk
  = expected cost of a decision

- Min P_e detector = special case of
  Bayes detector

- LRT = form of Bayes detector
  for <u>binary</u> tests

- MAP rule: form of Min P_e detector
  for $M \geq 2$ hypotheses

- All the above rules assume
  $\pi_i = P(H_i)$ is known.

- Next lecture: what if π_i is not known?

Key

a. $\int_{-\infty}^{0} f_1(x)\,dx$

b. $\sum_{\underline{x} \in \mathcal{X} \cap R_i} f_j(\underline{x})$

c. $\dfrac{f_1(\underline{x})}{f_0(\underline{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \dfrac{\pi_0}{\pi_1}$

d. $T \sim N(B \cdot \underline{0}, \; B \cdot \sigma^2 I \cdot B^T)$

$\sim N\left(0, \frac{\sigma^2}{N}\right)$ under H_0

$T \sim N(B \cdot A\underline{1}, \; B \cdot \sigma^2 I \cdot B^T)$

$\sim N\left(A, \frac{\sigma^2}{N}\right)$ under H_1