Detection Theory

A Modern Theory of Detection and Estimation.
Block-2: Analytical Detection

Emilio Parrado-Hernández, emilio.parrado@uc3m.es

October 24, 2022

Index

Example: Tipping riders

You usually order dinner to your favourite restaurant. The ordering app lets you fix a time slot to get your food delivered. They always serve within this range because if they serve off-range you get your order free. Since the pandemia the service is contactless: the riders leave the packet at your door, ring and rush to serve the next order in time. So you don't know who brought the food.
Since you have had a **long term relationship with them** you know your area is served by 3 riders with different quality of service:

- Early: delivers most frequently at the beginning of the time slot.
- Flatty: delivers indistinctly at any point in the time slot.
- L'80: delivers most frequently by the end of the slot.

The app lets you tip the riders according to their service, and you have made up a tipping policy oriented to getting the best possible service

Tipping riders: Fixed Tipping policy

Your tipping policy reflects you value of your long lasting relationship with the riders, that is you don't tip based on each independent order but on a long term good quality. Therefore you tip (through the app, as the service is contactless)

- Anytime you think that the rider was Rider Early you'd like to tip 20%
- Anytime you think that the rider was Rider Flatty you'd like to tip 15%
- Anytime you think that the rider was Rider L'80 you'd like to tip 10%

You want to design a **tipping strategy** that helps you maximize the goal of your tipping policy: rewarding each rider with what you consider they deserve. The tipping strategy depends on the **observation of the deliver instant**. Since the time slot for deliver is variable, you consider a normalized slot so $x \in [0, 1]$.

At the core of your tipping strategy lies a **decider** $d = \phi(x)$ that takes x as input and outputs the exact percentage of your tip depending on its guess about the correct rider.

So you define a **cost policy** to evaluate the quality of your decisions.

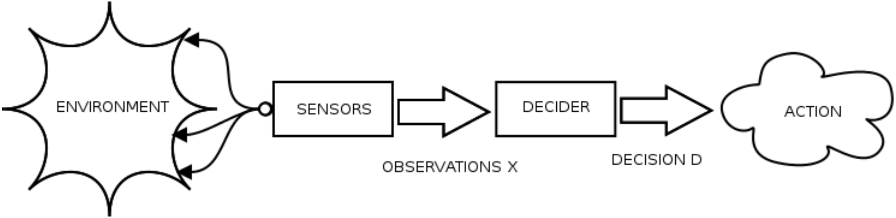
Examples of decision problems

Decision problems

Choose one of out several **hypothesis** or possible ways of explaining the observations

- Digital Communications: H_0 a zero was sent. H_1 a one was sent. $\mathbf{X}$ is the received signal
- Radar. H_0 no target. H_1 target. $\mathbf{X}$ is the received pulse
- Medical diagnosis. H_0 healthy. H_1 patient. $\mathbf{X}$ is the outcome of the test.
- Spam filtering. H_0 regular mail. H_1 spam. $\mathbf{X}$ is the email text.
- OCR. A hypothesis for each character. $\mathbf{X}$ is the written character
- Speech recognition. A hypothesis for each phoneme. $\mathbf{X}$ is the audio clip.
- News classification. A hypothesis for each newspaper section. $\mathbf{X}$ is a text.
- Image classification. A hypothesis for each class. $\mathbf{X}$ are the pixels in the image.

Elements of a detection problem



- **Hypothesis:** Discrete random variables that represent the several options that explain observations. **Disjoint, exhaustive and finite**.
- **Observations:** Random vector that contains the information recorded by sensors. Statistically related to the hypothesis.
- **Detector:** Implements the **discriminant function** $D = \phi(\mathbf{X})$. Mathematical function of the observation that assigns each observation to a decision.
- **Decision:** Discrete random variable D. Deterministic given the observation, i.e., each observation always leads to the same decision
- **Decision region:** part of the input space formed by observations that lead to the same decision

Decision regions

- **Each observation always leads to the same decision**.
- Input space partitioned into **categories:** regions formed by observations that lead to the same decision
- Decision region for category d:

$$\mathcal{X}_d = \{\mathbf{x} \in \mathcal{X} | \phi(\mathbf{x}) = d\}$$

- **Decision boundaries:** separation between regions.
- Every decider **induces a partition** of the input space into decision regions $\mathcal{X} = \bigcup_{d=0}^{M-1} \mathcal{X}_d$
  - This partition completely characterizes the decision function
  - It's equivalent to design a decision function or to design the partition of the input space

Design of decision functions

- **Analytical methods:** Problem is defined in terms of a complete statistical characterization of the involved random variables. This lecture and the next one
- **Machine learning:** Problem defined in terms of a set of labelled examples: observation and right decision. In two lectures time.

Index

Statistical modeling of a decision problem

- **Likelihood** of each hypothesis $p_{\mathbf{X}|H}(X = x|H = h)$. Generation of observations under each hypothesis. In a case with L hypothesis: $p_{\mathbf{X}|H}(x|H = 0)$, $p_{\mathbf{X}|H}(x|H = 1)$, $\ldots$, $p_{\mathbf{X}|H}(x|H = L - 1)$.
- **Prior probability of each hypothesis** $P_H(H = h)$. Note H is a discrete random variable. $\sum_{h=0}^{L-1} P_H(h) = 1$.
- **Prior distribution of the observations** $p_{\mathbf{X}}(\mathbf{x})$.
- **Joint distribution of observations and hypothesis** $p_{\mathbf{X},H}(x, h) = p_{\mathbf{X}|H}(x|h)P_H(h)$
- **Posterior of each hypothesis** $P_{H|\mathbf{X}}(h|X = x)$.

Tipping riders: math modelling of the problem (I)

The **observation**, x, is the precise moment of the time slot you got your dine delivered.

You choose to model this as a multiclass decision problem with **3 hypotheses** with the following **likelihoods**:

$$\begin{array}{llll} H=0: & \text{Flatty delivered} & p_{X|H}(x|0) = & 1 & 0<x<1 \\ H=1: & \text{Early delivered} & p_{X|H}(x|1) = & 2(1-x) & 0<x<1 \\ H=2: & \text{L'80 delivered} & p_{X|H}(x|2) = & 2x & 0<x<1 \end{array}$$

The prior probabilities (according to your past experiences/beliefs) are $P_H(0) = 0.4$ y $P_H(1) = P_H(2) = 0.3$

Risk

Evaluation of the performance of a decider.
Decisions involve **costs**

Costs

- Quantification of the consequences of each decision.
- $c(D, H) \in \mathbb{R}$ assigns a penalty c_{dh}, with $c_{dh} > c_{hh} \geq 0$, $\forall d \neq h$, to the fact of deciding $D = d$ when the right hypothesis was $H = h$
- We usually call **cost policy** to the set of the c_{dh}

Risk of a decider $\phi(\mathbf{x})$

Risk: Expected cost

$$
\begin{aligned}
r_\phi =& \mathbb{E}\{c(D,H)\} = \sum_{d=0}^{M-1} \sum_{h=0}^{L-1} c_{dh} P_{D,H}\{D = d, H = h\} \\
=& \sum_{d=0}^{M-1} \sum_{h=0}^{L-1} c_{dh} P_H(h) P_{D|H}(d|h) = \sum_{d=0}^{M-1} \sum_{h=0}^{L-1} c_{dh} P_H(h) \int_{\mathcal{X}_d} p_{\mathbf{x}|H}(\mathbf{x}|h) d\mathbf{x}
\end{aligned}
$$

Tipping riders: math modelling of the problem (II)

The possible **decisions** output by the decider are

$$\phi(x) = \begin{cases} d = 0 & \text{tip } 15\% \\ d = 1 & \text{tip } 20\% \\ d = 2 & \text{tip } 10\% \end{cases}$$

The last ingredient you need to design your decider is to stablish a cost policy that captures your perception of the impact of the consequences of the decisions. As a first step, you keep it simple $c_{hh} = 0, \ h = 0, 1, 2$ y $c_{dh} = 1, \ d \neq h$

Tipping riders: evaluate the performance of a generic decider

Compute the risk of the decider $\phi(x)$:

$$\phi(x) = \begin{cases} 1, & x < 0.5 \\ 2, & x > 0.5 \end{cases}$$

Solution

$$r_\phi = \quad 0.4 \cdot 0.5 + 0.4 \cdot 0.5 + 0.3 \cdot 0.25 + 0.3 \cdot 0.25 = 0.55$$

Conditional Risk

Evaluate the quality of a decision given the observation

$$\mathbb{E}\{c(d,H)|\mathbf{x}\} = \sum_{h=0}^{L-1} c_{dh} P_{H|\mathbf{X}}(h|\mathbf{x})$$

The conditional risk relates to the risk or overall risk

$$r_\phi = \mathbb{E}\{c(D,H)\} = \int \mathbb{E}\{c(d,H)|\mathbf{x}\} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

Example continued

Conditional risk for each decision:

$$\mathbb{E}\{c(d,H)|x\} = c_{d0}P_{H|X}(0|x) + c_{d1}P_{H|X}(1|x) + c_{d2}P_{H|X}(2|x)$$

After the application of the Bayes rule and some math we arrive at

- If $d = 0$:

$$\begin{aligned}\mathbb{E}\{c(0,H)|x\} &= c_{00}P_{H|X}(0|x) + c_{01}P_{H|X}(1|x) + c_{02}P_{H|X}(2|x)\\ &= 0\cdot 0.4 + 1\cdot 0.6(1-x) + 1\cdot 0.6x = 0.6\end{aligned}$$

- If $d = 1$:

$$\begin{aligned}\mathbb{E}\{c(1,H)|x\} &= c_{10}P_{H|X}(0|x) + c_{11}P_{H|X}(1|x) + c_{12}P_{H|X}(2|x)\\ &= 1\cdot 0.4 + 0\cdot 0.6(1-x) + 1\cdot 0.6x = 0.4 + 0.6x\end{aligned}$$

- If $d = 2$:

$$\begin{aligned}\mathbb{E}\{c(2,H)|x\} &= c_{20}P_{H|X}(0|x) + c_{21}P_{H|X}(1|x) + c_{22}P_{H|X}(2|x)\\ &= 1\cdot 0.4 + 1\cdot 0.6(1-x) + 0\cdot 0.6x = 1 - 0.6x\end{aligned}$$

Index

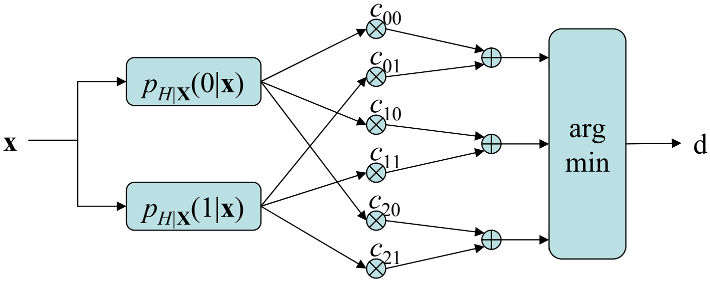
Bayesian Decision Theory

Minimum Expected Risk

Bayesian Decider: choose the decision of minimum risk

$$\phi^*(\mathbf{x}) = \arg\min_d \sum_{h=0}^{L-1} c_{dh} P_{H|\mathbf{X}}(h|\mathbf{x})$$

For instance, in a case with two hypotheses and three possible decisions

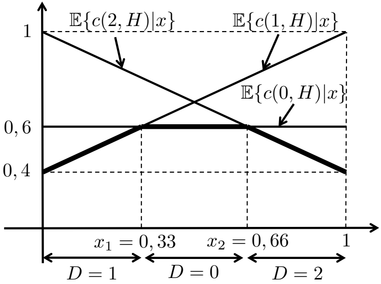


Example: Bayesian decider of the previous case

We had the expected cost per decision

$$\begin{aligned}\mathbb{E}\{c(0,H)|x\} &= \quad 0.6\\ \mathbb{E}\{c(1,H)|x\} &= \quad 0.4+0.6x\\ \mathbb{E}\{c(2,H)|x\} &= \quad 1-0.6x\end{aligned}$$

analyze, for each observation x, which term is the minimum:



Computing x_1:
$0.4 + 0.6x_1 = 0.6$
$x_1 = 0.33$

Computing x_2:
$1 - 0.6x_2 = 0.6$
$x_2 = 0.66$

Solution:

$$\phi(x) = \begin{cases} 1, & 0 < x < 0.33 \\ 0, & 0.33 < x < 0.66 \\ 2, & 0.66 < x < 1 \end{cases}$$

Bayesian Decider with likelihoods and priors

Applying Bayes' Rule

$$P_{H|\mathbf{X}}(h|\mathbf{x}) = \frac{p_{\mathbf{X}|H}(\mathbf{x}|h)P_H(h)}{p_{\mathbf{X}}(\mathbf{x})}$$

$$\phi^*(\mathbf{x}) = \arg\min_d \sum_{h=0}^{L-1} c_{dh} \frac{p_{\mathbf{X}|H}(\mathbf{x}|h)P_H(h)}{p_{\mathbf{X}}(\mathbf{x})}$$

since the denominator does not depend on the decision

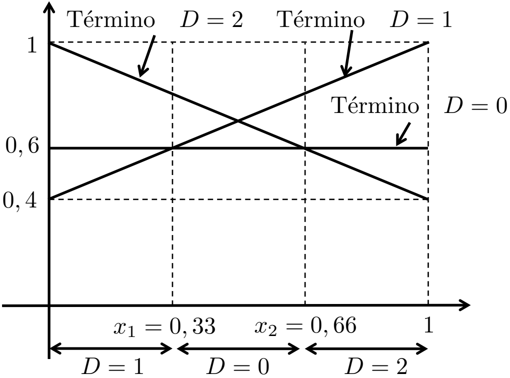
$$\phi^*(\mathbf{x}) = \arg\min_d \sum_{h=0}^{L-1} c_{dh} p_{\mathbf{X}|H}(\mathbf{x}|h)P_H(h)$$

Example continued

- If $D = 0$:
  $\sum_{h=0}^{2} c_{0h} p_{X|H}(x|h) P_H(h) = 0 \cdot 1 \cdot 0.4 + 1 \cdot 2(1-x) \cdot 0.3 + 1 \cdot 2x \cdot 0.3 = 0.6.$
- If $D = 1$:
  $\sum_{h=0}^{2} c_{1h} p_{X|H}(x|h) P_H(h) = 1 \cdot 1 \cdot 0.4 + 0 \cdot 2(1-x) \cdot 0.3 + 1 \cdot 2x \cdot 0.3 = 0.4 + 0.6x.$
- If $D = 2$:
  $\sum_{h=0}^{2} c_{2h} p_{X|H}(x|h) P_H(h) = 1 \cdot 1 \cdot 0.4 + 1 \cdot 2(1-x) \cdot 0.3 + 0 \cdot 2x \cdot 0.3 = 1 - 0.6x.$

Minimizing the cost for every x (after plotting each term as a function of x)

Example continued



Computing x_1:
$0,4 + 0,6x_1 = 0,6$
$x_1 = 0,33$

Computing x_2:
$1 - 0,6x_2 = 0,6$
$x_2 = 0,66$

Solution:

$$\phi(x) = \begin{cases} 1, & 0 < x < 0.33 \\ 0, & 0.33 < x < 0.66 \\ 2, & 0.66 < x < 1 \end{cases}$$

Index

MAP Decision

What if we have no access to the cost policy?

- Assume a decision corresponds to each hypothesis (number of decisions equal to number of hypotheses).
- Assume cost policy given by

$$c_{dh} = \left[\begin{array}{ll} 1, & \text{si } d \neq h \\ 0, & \text{si } d = h \end{array} \right] = 1 - \delta_{d-h}$$

MAP decider is Bayesian only if costs are $c_{dh} = 0$ if $d = h$ and $c_{dh} = 1$ if $d \neq h$

MAP Decider

Risk equal to probability of error

$$r_\phi = \sum_{d \neq h} P\{D = d, H = h\} = P\{D \neq H\}$$

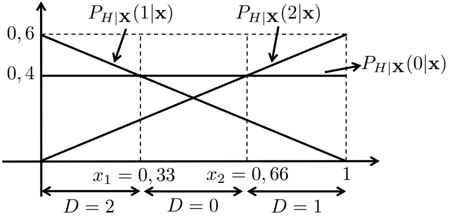
Minimum risk means maximum posterior probability

Risk of deciding $D = d$: $\mathbb{E}\{C(d, H)|\mathbf{x}\} = \sum_{h \neq d} P_{H|\mathbf{X}}(h|\mathbf{x}) = 1 - P_{H|\mathbf{X}}(d|\mathbf{x})$

$$\phi_{\mathrm{MAP}}(\mathbf{x}) = \arg\max_h P_{H|\mathbf{X}}(h|\mathbf{x})$$

Example continued

$$P_{H|X}(0|x) = 0.4$$
$$P_{H|X}(1|x) = 0.6(1-x)$$
$$P_{H|X}(2|x) = 0.6x$$



Solution:

$$\phi_{\text{MAP}}(x) = \begin{cases} 1, & 0 < x < 0.33 \\ 0, & 0.33 < x < 0.66 \\ 2, & 0.66 < x < 1 \end{cases}$$

Index

ML Decider

What if we have no access to the cost policy nor the prior probabilities of the hypotheses?

ML rule

$$\phi^*_{\mathrm{ML}}(\mathbf{x}) = \arg\max_h p_{\mathbf{X}|H}(\mathbf{x}|h)$$

ML and MAP are sometimes equivalent

MAP and ML deciders are equivalent when the hypotheses are **equiprobable**: $P_H(h) = 1/L \ \forall h$

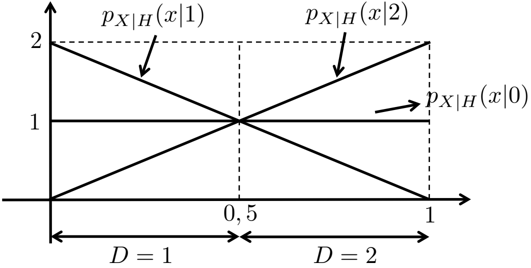
ML is Bayesian?

ML is Bayesian when

- Hypotheses are equiprobable
- Cost policy is a minimum error probability case

Example continued

$$
\begin{array}{lll}
p_{X|H}(x|0) = & 1 & 0 < x < 1 \\
p_{X|H}(x|1) = & 2(1-x) & 0 < x < 1 \\
p_{X|H}(x|2) = & 2x & 0 < x < 1
\end{array}
$$



Decision regions:

$$
\phi(x) = \begin{cases} 1, & 0 < x < 0.5 \\ 2, & 0.5 < x < 1 \end{cases}
$$

Index

Index

Binary decision setup

- Two Possible Decisions $D = \{0, 1\}$
- Two Possible Hypotheses $H = \{0, 1\}$
- Correct decision (hit): $\{D = 0, H = 0\}$ or $\{D = 1, H = 1\}$
- Wrong decision (error): $\{D = 1, H = 0\}$ or $\{D = 0, H = 1\}$
- Special names for some joint events:
  - Detection: $\{D = 1, H = 1\}$
  - Missing target: $\{D = 0, H = 1\}$
  - False Alarm: $\{D = 1, H = 0\}$
- Design Bayesian binary deciders using a minimum risk criterion.
- Plus some alternatives based in other criteria different from risk minimization.

Cost policy and risk

In binary detection the cost policy is defined by 4 constants: c_{00}, c_{11}, c_{01}, c_{10}. Costs are positive. The cost of an error must be larger than the cost of the corresponding correct decision: $c_{10} > c_{00}$ and $c_{01} > c_{00}$.
The risk of a binary decider is given by:

$$\begin{aligned} r_\phi =& c_{00} P\{D=0, H=0\} + c_{01} P\{D=0, H=1\} \\ & + c_{10} P\{D=1, H=0\} + c_{11} P\{D=1, H=1\} \\ =& c_{00} P_H(0) P_{D|H}(0|0) + c_{01} P_H(1) P_{D|H}(0|1) \\ & + c_{10} P_H(0) P_{D|H}(1|0) + c_{11} P_H(1) P_{D|H}(1|1) \end{aligned}$$

Probability of False Alarm and probability of Missing Target

Probability of False Alarm

$$P_{FA} = P_{D|H}(1|0) = \int_{\mathcal{X}_1} p_{\mathbf{x}|H}(\mathbf{x}|0)d\mathbf{x}$$

Probability of Missing Target

$$P_M = P_{D|H}(0|1) = \int_{\mathcal{X}_0} p_{\mathbf{x}|H}(\mathbf{x}|1)d\mathbf{x}$$

Risk in terms of P_{FA} and P_M

$$
\begin{aligned}
r_\phi =& c_{00} P_H(0) P_{D|H}(0|0) + c_{01} P_H(1) P_{D|H}(0|1) \\
& + c_{10} P_H(0) P_{D|H}(1|0) + c_{11} P_H(1) P_{D|H}(1|1) \\
=& c_{00} P_H(0)(1 - P_{D|H}(1|0)) + c_{01} P_H(1) P_{D|H}(0|1) \\
& + c_{10} P_H(0) P_{D|H}(1|0) + c_{11} P_H(1)(1 - P_{D|H}(0|1)) \\
=& c_{00} P_H(0)(1 - P_{\mathrm{FA}}) + c_{01} P_H(1) P_{\mathrm{M}} + c_{10} P_H(0) P_{\mathrm{FA}} + c_{11} P_H(1)(1 - P_{\mathrm{M}}) \\
=& (c_{01} - c_{11}) P_H(1) P_{\mathrm{M}} + (c_{10} - c_{00}) P_H(0) P_{\mathrm{FA}} + (c_{00} P_H(0) + c_{11} P_H(1))
\end{aligned}
$$

Risk is sum of three components:

- $(c_{00} P_H(0) + c_{11} P_H(1))$ minimum risk corresponding to an ideal detector $P_M = 0$ and $P_{FA} = 0$ and hits with probability 1.
- $(c_{01} - c_{11}) P_H(1) P_M$ contribution of missing targets to the risk
- $(c_{10} - c_{00}) P_H(0) P_{FA}$ contribution of false alarms to the risk.

Notice that as long as $p(\mathbf{x}|H = 0)$ and $p(\mathbf{x}|H = 1)$ present some overlapping the decider will incur in errors.

Discriminant function

Every binary decision can be expressed as comparing a function of the observations with a threshold η

Discriminant function

$$g(\mathbf{x}) \underset{D=0}{\overset{D=1}{\gtrless}} \eta$$

Since $\mathbf{X}$ is a random variable, $g(\mathbf{X})$ is a random variable itself. We call $\Lambda = g(\mathbf{X})$

P_{FA} and P_M using Λ

$$P_{\text{FA}} = P_{D|H}(1|0) = P\{\Lambda > \eta | H = 0\} = \int_{\eta}^{\infty} p_{\Lambda|H}(\lambda|0) d\lambda$$

$$P_M = P_{D|H}(0|1) = P\{\Lambda > \eta | H = 1\} = \int_{-\infty}^{\eta} p_{\Lambda|H}(\lambda|1) d\lambda$$

Notice that even when $\mathbf{X}$ could be a random vector, Λ is always a **scalar**.

Index

Bayesian binary deciders

Design a binary decider means **find a discriminant function** $g(\mathbf{X})$ and a **threshold** η that define a decision rule of minimum risk

$$c_{10} P_{H|\mathbf{X}}(0|\mathbf{x}) + c_{11} P_{H|\mathbf{X}}(1|\mathbf{x}) \underset{D=1}{\overset{D=0}{\gtrless}} c_{00} P_{H|\mathbf{X}}(0|\mathbf{x}) + c_{01} P_{H|\mathbf{X}}(1|\mathbf{x})$$

Grouping terms

$$(c_{10} - c_{00}) P_{H|\mathbf{X}}(0|\mathbf{x}) \underset{D=1}{\overset{D=0}{\gtrless}} (c_{01} - c_{11}) P_{H|\mathbf{X}}(1|\mathbf{x})$$

since $c_{10} > c_{00}$ y $c_{01} > c_{11}$,

$$\frac{P_{H|\mathbf{X}}(1|\mathbf{x})}{P_{H|\mathbf{X}}(0|\mathbf{x})} \underset{D=0}{\overset{D=1}{\gtrless}} \frac{c_{10} - c_{00}}{c_{01} - c_{11}}$$

$g(\mathbf{X})$ is the ratio of posterior probabilities, η is the ratio of the cost increment.

Likelihood Ratio Tests

Application of the Bayes' rule to the previous result:

$$\frac{p_{\mathbf{x}|H}(\mathbf{x}|1)}{p_{\mathbf{x}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\gtrless}} \frac{(c_{10}-c_{00})P_H(0)}{(c_{01}-c_{11})P_H(1)}$$

Index

MAP decider

- Risk equal to probability of error
- Cost policy given by

$$c_{dh} = \left[\begin{array}{ll} 1, & \text{if } d \neq h \\ 0, & \text{if } d = h \end{array} \right] = 1 - \delta_{d-h}$$

$$P_{H|\mathbf{x}}(1|\mathbf{x}) \underset{D=0}{\overset{D=1}{\gtrless}} P_{H|\mathbf{x}}(0|\mathbf{x})$$

Using Bayes' rule:

$$\frac{p_{\mathbf{x}|H}(\mathbf{x}|1)}{p_{\mathbf{x}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\gtrless}} \frac{P_H(0)}{P_H(1)}$$

The probability of error is

$$P_e = r_{\phi_{\mathrm{MAP}}} = P\{D \neq H\} = P_H(1)P_{\mathrm{M}} + P_H(0)P_{\mathrm{FA}}$$

Binary ML Decider

ML rule

$$\phi^*_{\text{ML}}(\mathbf{x}) = \arg\max_h p_{\mathbf{X}|H}(\mathbf{x}|h)$$

ML binary case

$$p_{\mathbf{X}|H}(\mathbf{x}|1) \underset{D=0}{\overset{D=1}{\gtrless}} p_{\mathbf{X}|H}(\mathbf{x}|0)$$

Remember

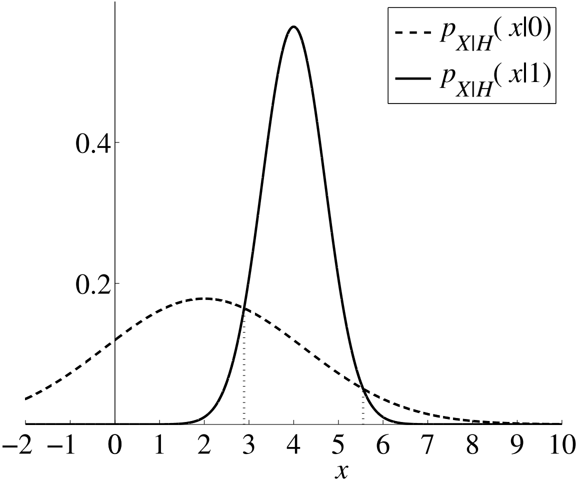
Binary MAP and ML deciders are equivalent when the hypotheses are **equiprobable**: $P(H=0) = P(H=1) = 1/2$

Binary ML is Bayesian if

- $P(H=0) = P(H=1)$
- Cost policy is a minimum error probability case

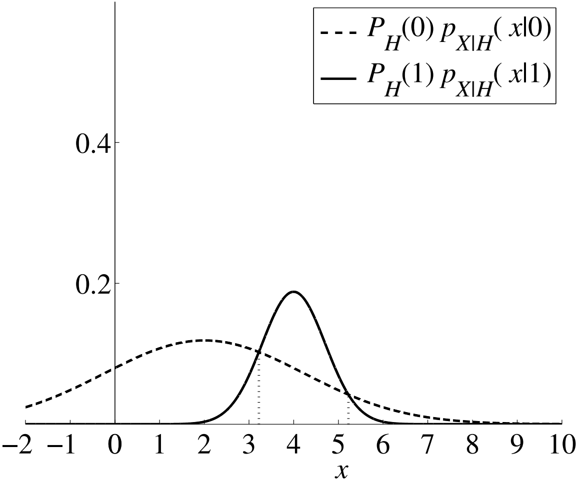
Example Binary decider

Consider a binary decision problem with Gaussian likelihoods with means $m_0 = 2$ and $m_1 = 4$ and variances $v_0 = 5$ and $v_1 = 0.5$



Example Binary decider

Consider a binary decision problem with Gaussian likelihoods with means $m_0 = 2$ and $m_1 = 4$ and variances $v_0 = 5$ and $v_1 = 0.5$
And Priors $P_H(0) = 2/3$ and $P_H(1) = 1/3$



Example Binary decider

Consider a binary decision problem with Gaussian likelihoods with means
$m_0 = 2$ and $m_1 = 4$ and variances $v_0 = 5$ and $v_1 = 0.5$
And Priors $P_H(0) = 2/3$ and $P_H(1) = 1/3$
And costs $c_{10} = 2$, $c_{01} = 1$ and $c_{00} = c_{11} = 0$