Binary Hypothesis Testing

- Probability is a foundation for principled decision making from partial, noisy observations. This is known as detection theory or hypothesis testing.
- Key Idea: Decide which of two (mutually exclusive) events occured using an observation.
  - → Ex: Digital Communication: Based on the received voltage, was the transmitted bit 0 or 1?
  - → Ex: Cancer Detection: Based on a CT scan, is a tumor present or absent?
  - → Ex: Quality Control: Based on a measurement, is a manufactured part defective or not?
- Here, we focus on two events and scalar observations. Extending this framework is simple, as we will see later.

• Binary Hypothesis Testing Framework:

→ There are two hypotheses H_0 and H_1, which are events that partition the sample space Ω. "State of Nature"

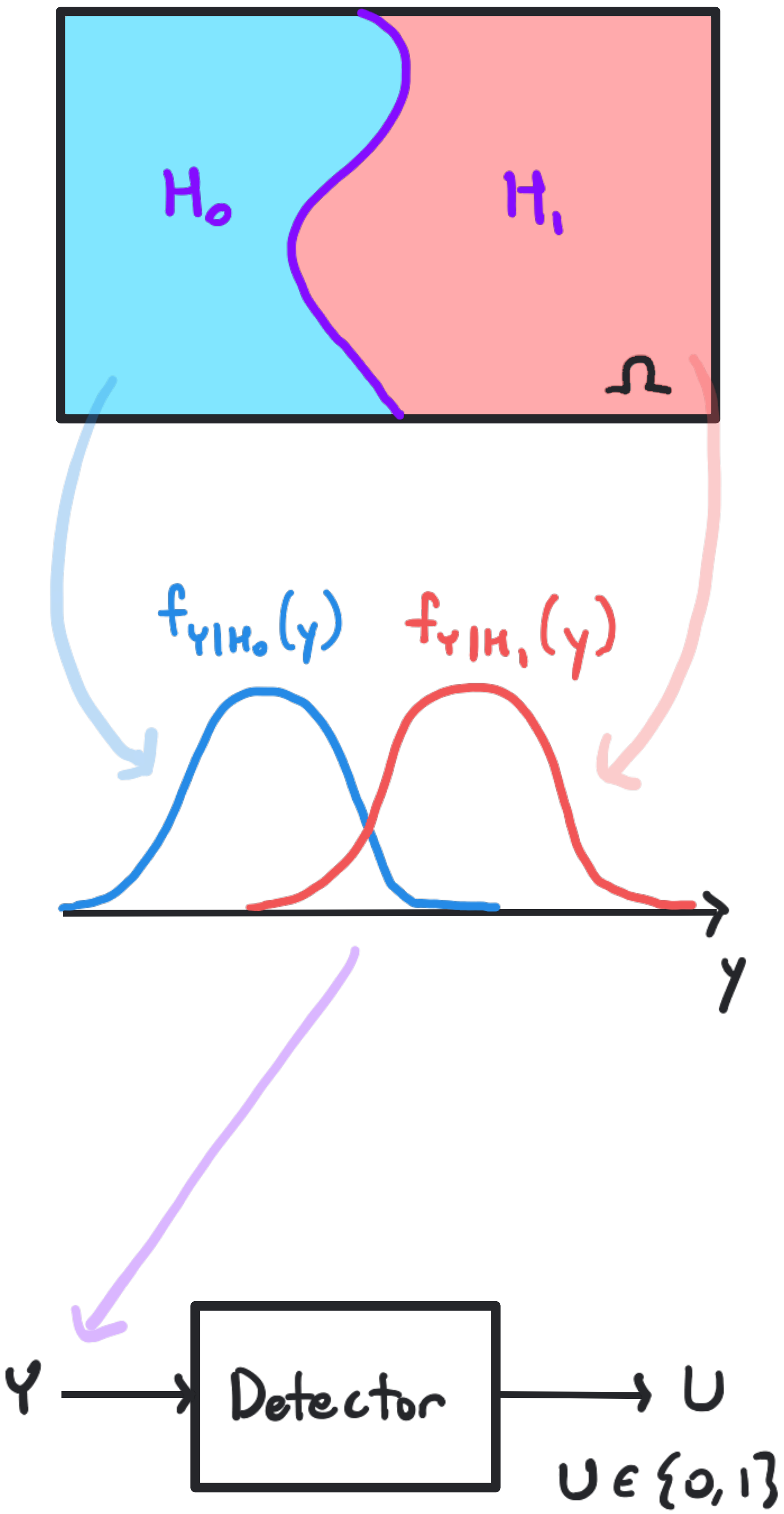
→ We obtain a measurement (or observation), which is a random variable Y whose values are distributed according to

(We model this.)

Discrete Case	Continuous Case
$\boxed{P_{Y\mid H_0}(y)}$ if H_0 occurs	$\boxed{f_{Y\mid H_0}(y)}$ if H_0 occurs
$\boxed{P_{Y\mid H_1}(y)}$ if H_1 occurs	$\boxed{f_{Y\mid H_1}(y)}$ if H_1 occurs

likelihoods

→ There is a detector (or decision rule), which is a function $U = D(Y)$ that outputs 0 if it decides H_0 occurred and 1 if it decides H_1 occurred.
(We design this.)



- The decision rule creates a partition of R_Y, the range of the measurement Y:

$$A_0 = \{y \in R_Y : D(y) = 0\}$$
$$A_1 = \{y \in R_Y : D(y) = 1\}$$

- An error occurs if we decide H_1 but H_0 actually occurred or if we decide H_0 but H_1 actually occurred.

→ as an event: $\{\text{error}\} = (\{Y \in A_1\} \cap H_0) \cup (\{Y \in A_0\} \cap H_1)$

- One measure of performance is the probability of error P_e

$$\boxed{\begin{aligned} P_e = \mathbb{P}[\{\text{error}\}] &= \mathbb{P}[\{\text{error}\} \mid H_0]\,\mathbb{P}[H_0] + \mathbb{P}[\{\text{error}\} \mid H_1]\,\mathbb{P}[H_1] \\ &= \mathbb{P}[\{Y \in A_1\} \mid H_0]\,\mathbb{P}[H_0] + \mathbb{P}[\{Y \in A_0\} \mid H_1]\,\mathbb{P}[H_1] \end{aligned}}$$

Discrete Case

$$\mathbb{P}[\{Y \in A_1\} \mid H_0] = \sum_{y \in A_1} P_{Y|H_0}(y)$$

$$\mathbb{P}[\{Y \in A_0\} \mid H_1] = \sum_{y \in A_0} P_{Y|H_1}(y)$$

Continuous Case

$$\mathbb{P}[\{Y \in A_1\} \mid H_0] = \int_{A_1} f_{Y|H_0}(y)\,dy$$

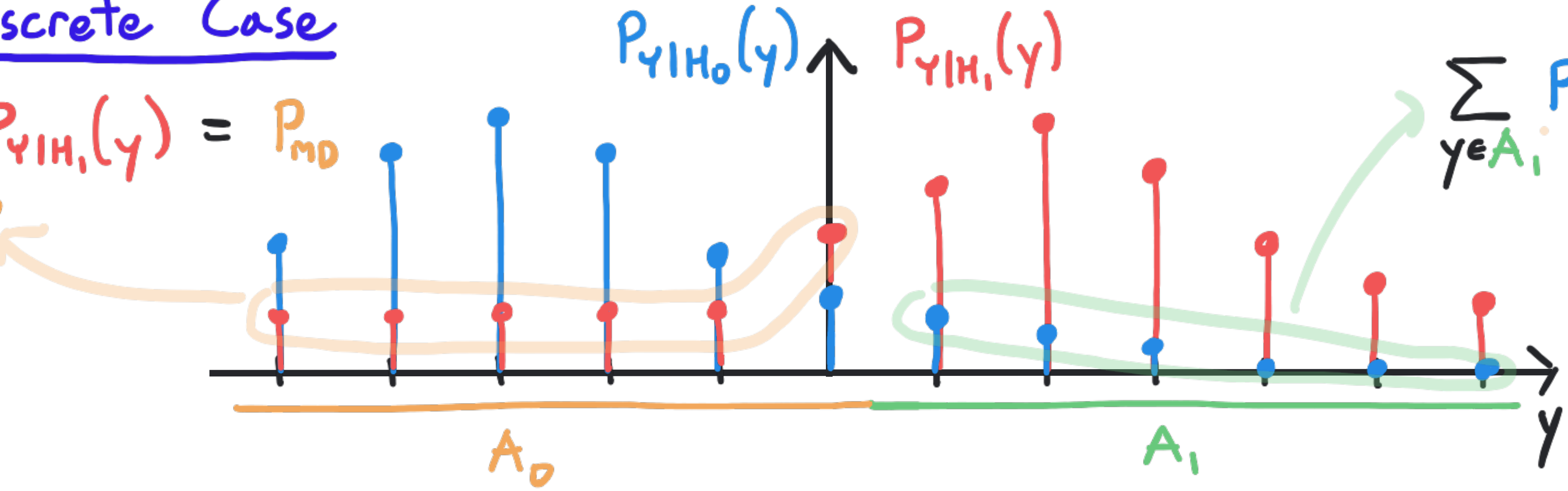
$$\mathbb{P}[\{Y \in A_0\} \mid H_1] = \int_{A_0} f_{Y|H_1}(y)\,dy$$

- One of the motivations for detection theory was detecting an aircraft using radar, which lead to the terminology:
  - → Probability of False Alarm $P_{FA} = \mathbb{P}[\{Y \in A_1\} \mid H_0]$
  - → Probability of Missed Detection $P_{MD} = \mathbb{P}[\{Y \in A_0\} \mid H_1]$

Discrete Case

$$\sum_{y \in A_0} P_{Y|H_1}(y) = P_{MD}$$

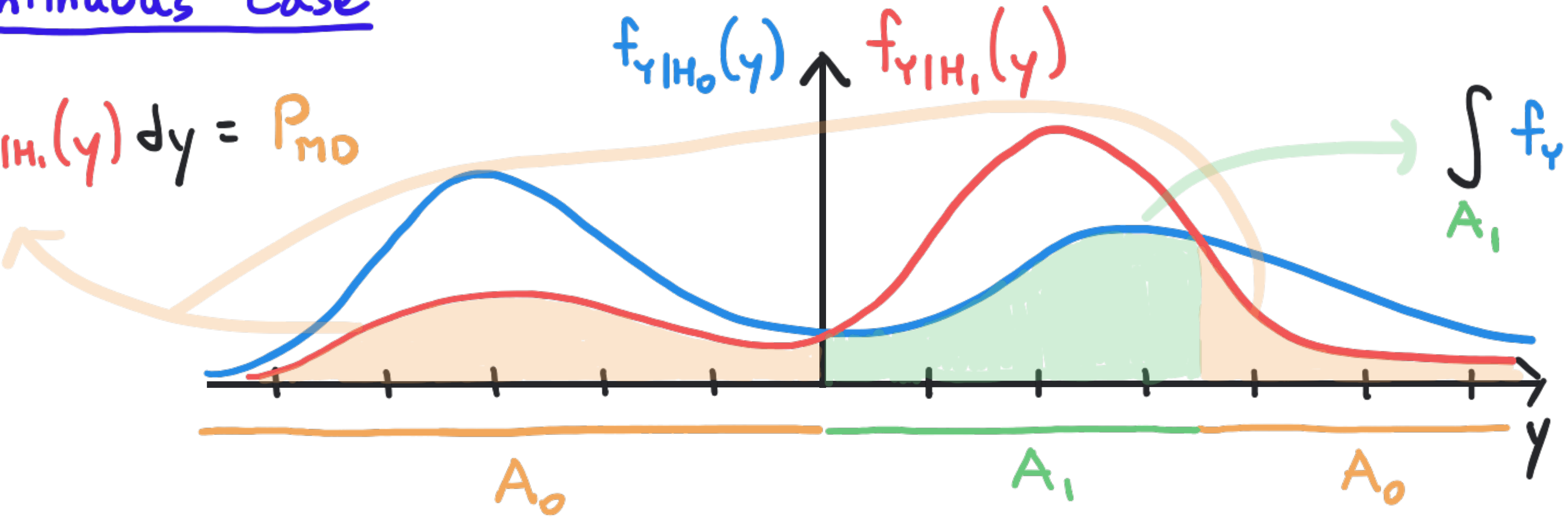
$$\sum_{y \in A_1} P_{Y|H_0}(y) = P_{FA}$$



Continuous Case

$$\int_{A_0} f_{Y|H_1}(y)\,dy = P_{MD}$$

$$\int_{A_1} f_{Y|H_0}(y)\,dy = P_{FA}$$



$$\boxed{P_e = P_{FA}\,\mathbb{P}[H_0] + P_{MD}\,\mathbb{P}[H_1]}$$

- Overall, our job is to design a decision rule $D(y)$ that attains a low probability of error P_e (or, better yet, the smallest possible P_e).

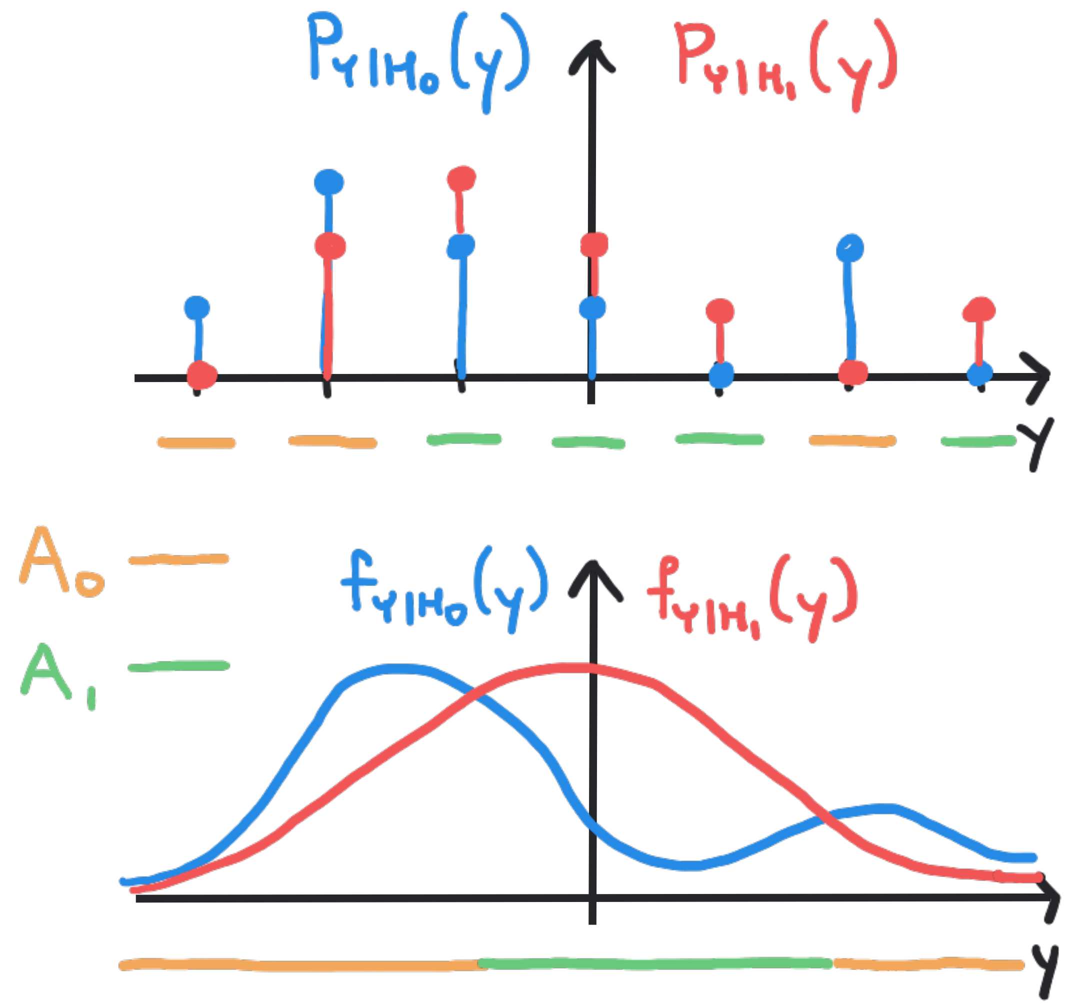
- The maximum likelihood (ML) rule $D^{ML}(y)$ selects the hypothesis with the highest likelihood value

Discrete Case

$$D^{ML}(y) = \begin{cases} 1, & P_{Y|H_1}(y) \geq P_{Y|H_0}(y) \\ 0, & P_{Y|H_1}(y) < P_{Y|H_0}(y) \end{cases}$$

Continuous Case

$$D^{ML}(y) = \begin{cases} 1, & f_{Y|H_1}(y) \geq f_{Y|H_0}(y) \\ 0, & f_{Y|H_1}(y) < f_{Y|H_0}(y) \end{cases}$$



→ Ties $P_{Y|H_1}(y) = P_{Y|H_0}(y)$ can be broken arbitrarily, we map them to 1.

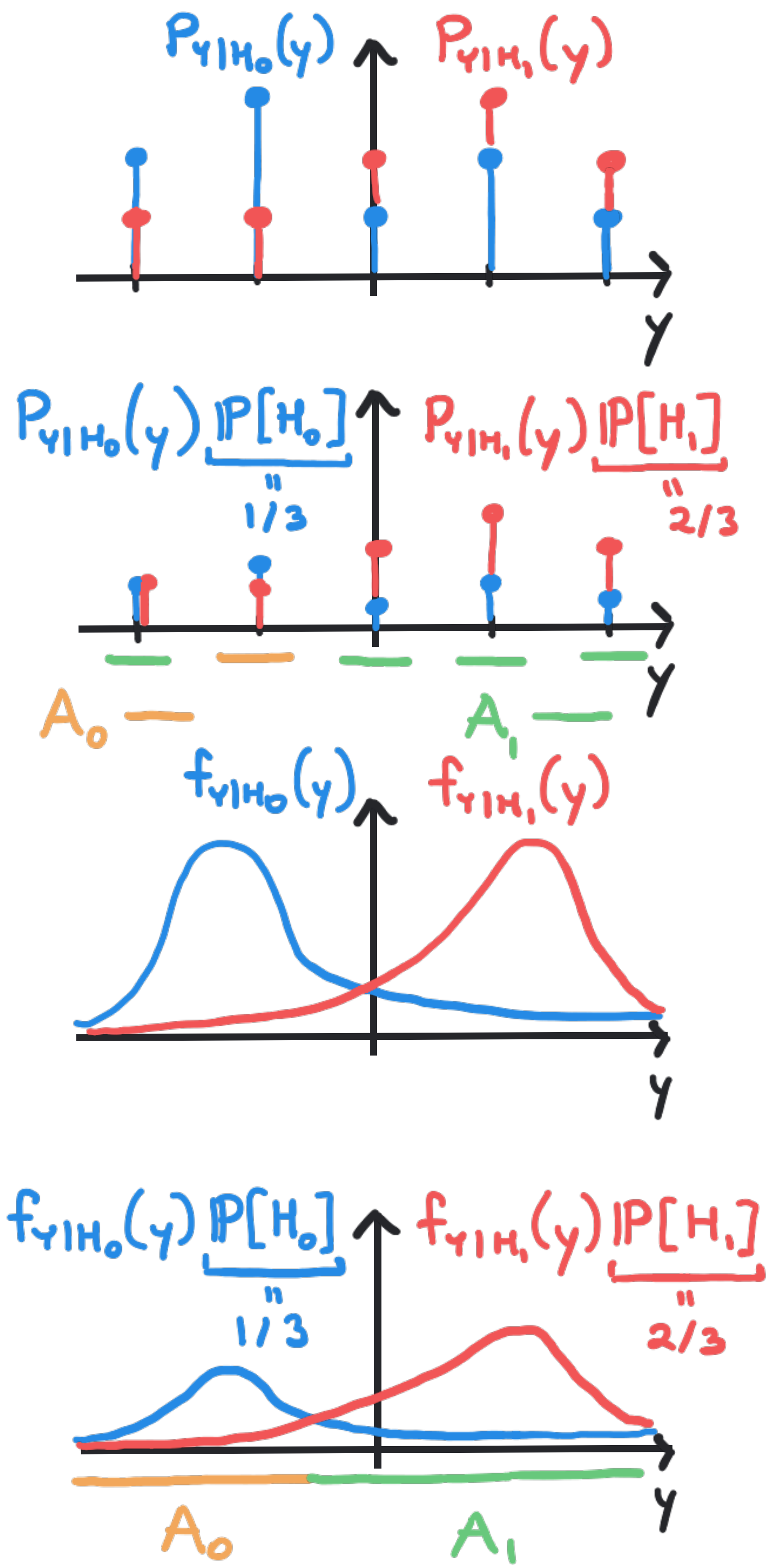
• The maximum a posteriori (MAP) rule $D^{MAP}(y)$ selects the hypothesis that is most likely given the observation.

Discrete Case

$$D^{MAP}(y) = \begin{cases} 1, & P_{Y|H_1}(y)\,\mathbb{P}[H_1] \geq P_{Y|H_0}(y)\,\mathbb{P}[H_0] \\ 0, & P_{Y|H_1}(y)\,\mathbb{P}[H_1] < P_{Y|H_0}(y)\,\mathbb{P}[H_0] \end{cases}$$

Continuous Case

$$D^{MAP}(y) = \begin{cases} 1, & f_{Y|H_1}(y)\,\mathbb{P}[H_1] \geq f_{Y|H_0}(y)\,\mathbb{P}[H_0] \\ 0, & f_{Y|H_1}(y)\,\mathbb{P}[H_1] < f_{Y|H_0}(y)\,\mathbb{P}[H_0] \end{cases}$$



→ How did we arrive at this rule? (Discrete)

$$\mathbb{P}[H_1 \mid \{Y=y\}] \geq \mathbb{P}[H_0 \mid \{Y=y\}]$$

Choose H_1 if it is more likely given Y.

$$\frac{\mathbb{P}[\{Y=y\} \mid H_1]\,\mathbb{P}[H_1]}{\cancel{\mathbb{P}[\{Y=y\}]}} \geq \frac{\mathbb{P}[\{Y=y\} \mid H_0]\,\mathbb{P}[H_0]}{\cancel{\mathbb{P}[\{Y=y\}]}}$$

Bayes' Rule

$$P_{Y|H_1}(y)\,\mathbb{P}[H_1] \geq P_{Y|H_0}(y)\,\mathbb{P}[H_0]$$

- The MAP rule is optimal: it attains the smallest possible probability of error.

→ Why?

$$P_e = \sum_{y \in R_Y} P_Y(y)\, \mathbb{P}[\{\text{error}\} \mid \{Y=y\}] \qquad \mathbb{P}[\{\text{error}\} \mid \{Y=y\}] = \begin{cases} \mathbb{P}[H_0 \mid \{Y=y\}] & D(y)=1 \\ \mathbb{P}[H_1 \mid \{Y=y\}] & D(y)=0 \end{cases}$$

$$= \sum_{y \in R_Y} P_Y(y) \left(D(y)\, \mathbb{P}[H_0 \mid \{Y=y\}] + (1-D(y))\, \mathbb{P}[H_1 \mid \{Y=y\}] \right)$$

$$= \sum_{y \in R_Y} P_Y(y) \left(\mathbb{P}[H_1 \mid \{Y=y\}] + D(y) \underbrace{\left(\mathbb{P}[H_0 \mid \{Y=y\}] - \mathbb{P}[H_1 \mid \{Y=y\}] \right)} \right)$$

If $\mathbb{P}[H_0 \mid \{Y=y\}] < \mathbb{P}[H_1 \mid \{Y=y\}]$, this term contributes negatively to the sum. Keep it by setting $D(y)=1$.
If $\mathbb{P}[H_0 \mid \{Y=y\}] > \mathbb{P}[H_1 \mid \{Y=y\}]$, this term contributes positively to the sum. Drop it by setting $D(y)=0$.

This is the MAP rule! (Ties can go either way.)

- Why not use the MAP rule exclusively?
  - → We may not know $\mathbb{P}[H_0]$ and $\mathbb{P}[H_1]$: ML rule still works.
  - → The costs of false alarm and missed detection may be unequal.

  <u>Ex</u>: Deciding a malignant tumor is benign is clearly worse.