# FA-E1 R 7.1

Espiritu, Joseph Raphael M.

2025-02-17

---

## Problem Statement 1:

A **malicious spyware** can infect a computer system through either the **Internet** or through **email**. The probabilities for the **entry points** and the **detection rates** are as follows:

- **70%** of spyware infections occur through the **Internet**.
- **30%** of spyware infections occur through **Email**.

If the spyware enters via the **Internet**, the anti-virus detector will detect it with a probability of **60%**. If it enters via **Email**, the anti-virus will detect it with a probability of **80%**.

These probabilities can be expressed as:

- **70%** of all spyware comes through the **Internet**: $P(\text{Internet}) = 0.70$
- **30%** of all spyware comes through **Email**: $P(\text{Email}) = 0.30$
- The probability of detection if the spyware is from the **Internet**: $P(\text{Detected}|\text{Internet}) = 0.60$
- The probability of detection if the spyware is from **Email**: $P(\text{Detected}|\text{Email}) = 0.80$
- The probability of undetected spyware from the **Internet**: $P(\text{Undetected}|\text{Internet}) = 0.40$
- The probability of undetected spyware from **Email**: $P(\text{Undetected}|\text{Email}) = 0.20$

### Questions:

(a) What is the probability that this **spyware infects the system**?

(b) If the spyware is **detected**, what is the **probability that it came through the Internet**?

### Solution

- The probability of detected spyware is:

$$P(\text{Detected}) = P(\text{Detected}|\text{Internet})P(\text{Internet}) + P(\text{Detected}|\text{Email})P(\text{Email})$$

```
# Given Probabilities
P_Internet <- 0.70
P_Email <- 0.30
P_Detected_Internet <- 0.60
P_Detected_Email <- 0.80
P_Undetected_Internet <- 0.40
P_Undetected_Email <- 0.20

# Solving for Detected Spyware
P_Detected <- P_Detected_Internet * P_Internet + P_Detected_Email * P_Email
```

```r
# Show Value
P_Detected
```

```
## [1] 0.66
```

Using **law of total probability**, the probability of receiving **Undetected spyware** is:

$$P(\text{Undetected}) = P(\text{Undetected}|\text{Internet})P(\text{Internet}) + P(\text{Undetected}|\text{Email})P(\text{Email})$$

Using **Bayes' theorem**, the probability that a **spyware detected from the internet** is:

$$P(\text{Internet}|\text{Detected}) = \frac{P(\text{Detected}|\text{Internet})P(\text{Internet})}{P(\text{Detected})}$$

---

```r
# Solving for Spyware Undetected
P_Undetected <- P_Undetected_Internet * P_Internet + P_Undetected_Email * P_Email

# Solving for Spyware detected from Internet
P_Internet_Detected <- P_Detected_Internet * P_Internet / P_Detected

# Show Values
P_Undetected
```

```
## [1] 0.34
```

```r
P_Internet_Detected
```

```
## [1] 0.6363636
```

---

**Thus, Undetected Spyware appears 34% and Detected spyware from the Internet appears around 63.64% in any computer system.**

## Problem Statement 2:

Of the **emails you receive, 20% are spam** on average. Your **spam filter** is able to detect **90%** of them but also **misclassifies as spam 15% of the genuine emails**.

These probabilities can be expressed as:

- Probability that an email is spam: $P(\text{Spam}) = 0.20$
- Probability that an email is legitimate: $P(\text{Legitimate}) = 0.80$
- Probability that the spam filter detects spam: $P(\text{Marked}|\text{Spam}) = 0.90$
- Probability that the spam filter misclassifies a legitimate email as spam: $P(\text{Marked}|\text{Legitimate}) = 0.15$
- Probability that the spam filter does not mark a spam email as spam: $P(\text{Not Marked}|\text{Spam}) = 0.10$
- Probability that the spam filter does not mark a legitimate email as spam: $P(\text{Not Marked}|\text{Legitimate}) = 0.85$

**Questions:**

(a) If an **email arrives** and is **marked spam**, what is the **probability that it really is spam**?
(b) If an **email arrives** and is **not marked spam**, what is the **probability that it is legitimate**?

**Solution**

- The probability of Marked Spam is:

$$P(\text{Marked}) = P(\text{Marked}|\text{Spam})P(\text{Spam}) + P(\text{Marked}|\text{Legitimate})P(\text{Legitimate})$$

- The probability of Not Marked Spam is:

$$P(\text{Not Marked}) = P(\text{Not Marked}|\text{Spam})P(\text{Spam}) + P(\text{Not Marked}|\text{Legitimate})P(\text{Legitimate})$$

---

```
# Given Probabilities
P_Spam <- 0.20
P_Legitimate <- 0.80
P_Marked_Spam <- 0.90
P_Marked_Legitimate <- 0.15
P_NotMarked_Spam <- 0.10
P_NotMarked_Legitimate <- 0.85

# Solving for Marked
P_Marked <- P_Marked_Spam * P_Spam + P_Marked_Legitimate * P_Legitimate

# Solving for Not Marked
P_NotMarked <- P_NotMarked_Spam * P_Spam + P_NotMarked_Legitimate * P_Legitimate

# Show Value
P_Marked
```

```
## [1] 0.3
```

```
P_NotMarked
```

```
## [1] 0.7
```

---

Therefore,

- Probability for anything to be Marked is: $P(\text{Marked}) = 0.3$
- Probability for anything to be Not Marked: $P(\text{NotMarked}) = 0.70$

Using **Bayes' theorem** for both questions we have the following equations,

$$P(\text{Spam}|\text{Marked}) = \frac{P(\text{Marked}|\text{Spam}) \cdot P(\text{Spam})}{P(\text{Marked})}$$

$$P(\text{Legitimate}|\text{Not Marked}) = \frac{P(\text{Not Marked}|\text{Legitimate}) \cdot P(\text{Legitimate})}{P(\text{Not Marked})}$$

---

```
# Solving for Spam when its Marked
P_Spam_Marked <- P_Marked_Spam * P_Spam / P_Marked

# Solving for Legitimate when its Not Marked
P_Legitimate_NotMarked <- P_NotMarked_Legitimate * P_Legitimate / P_NotMarked

# Show Values
P_Spam_Marked
```

```
## [1] 0.6
```

```
P_Legitimate_NotMarked
```

```
## [1] 0.9714286
```

---

**Thus, if an email arrives, the probability it is spam is about 60%, and the probability that a legitimate email is not marked as spam is around 97.14%.**

---