Lecture 4

Law of Total Probability & Bayes' Rule

STAT 330 - Iowa State University

Tree Diagram

Tree Diagram

Example 1: Suppose you randomly select one of 3 boxes, and then randomly select a coin from inside the box. The contents of the boxes are ...

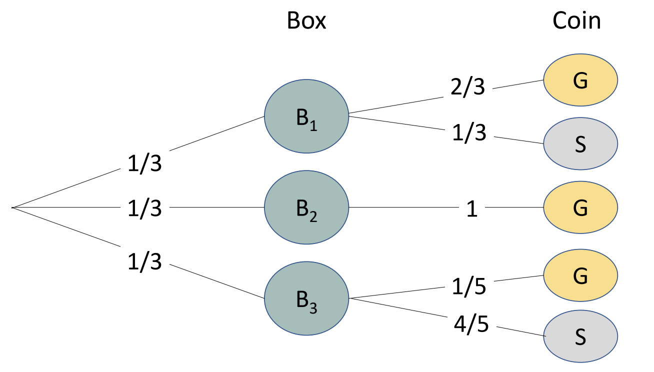
- Box 1: 2 gold coins, 1 silver coin
- Box 2: 3 gold coins
- Box 3: 1 gold coin, 4 silver coins

Let events $B_i = i^{th}$ box is selected for $i = 1, 2, 3$,
$G =$ gold coin selected, and $S =$ silver coin selected.

We can visualize this *step-wise procedure* with a *tree diagram*.

Using a Tree Diagram

A tree diagram shows all possible outcomes of step-wise procedures



$P(B_i) = \frac{1}{3}$ for $i = 1, 2, 3$

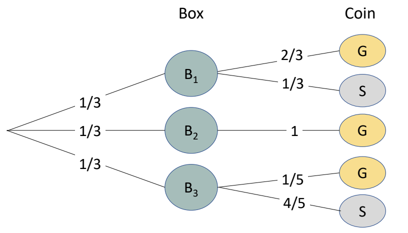
$P(G|B_1) = \frac{2}{3}$, $P(S|B_1) = \frac{1}{3}$

$P(G|B_2) = 1$

$P(G|B_3) = \frac{1}{5}, P(S|B_3) = \frac{4}{5}$

Using a Tree Diagram Cont.

What is the probability of choosing a gold coin $P(G)$?



- What are the *"total"* different paths to get to gold coin?

  $(B_1 \cap G)$ or $(B_2 \cap G)$ or $(B_3 \cap G)$

- These are disjoint events

$$
\begin{aligned}
P(G) &= P(B_1 \cap G) + P(B_2 \cap G) + P(B_3 \cap G) \\
&= P(B_1)P(G|B_1) + P(B_2)P(G|B_2) + P(B_2)P(G|B_2) \\
&=
\end{aligned}
$$

This calculation is done using *Law of Total Probability.*

Law of Total Probability

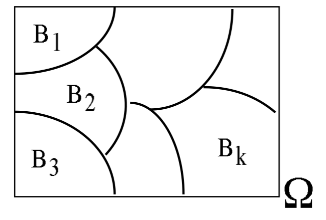
Cover/Partition

Definition:

A collection of events $B_1, \ldots B_k$ is a *cover* or *partition* of Ω if

1. the events are pairwise disjoint ($B_i \cap B_j = \emptyset$ for $i \neq j$), and
2. the union of the events is Ω ($\bigcup_{i=1}^{k} B_i = \Omega$).

We can represent a cover using a Venn diagram:



Note: In a tree diagram, the branches of the tree form a cover.

Law of Total Probability

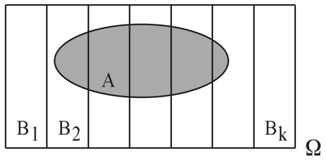
Theorem (Law of Total Probability)

If the collection of events $B_1, \ldots, B_k$ is a cover of Ω, and A is an event, then

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

Proof

- $A = (B_1 \cap A) \cup \ldots \cup (B_k \cap A)$
- $P(A) = P(B_1 \cap A) + \ldots + P(B_k \cap A)$
  $= P(A|B_1)P(B_1) + \ldots + P(A|B_k)P(B_k)$



Bayes' Rule

Bayes' Rule

Theorem (Bayes' Rule)

If $B_1, \ldots, B_k$ is a cover or partition of Ω, and A is an event, then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

Why?

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

- Bayes rule $\rightarrow$ way to "flip" conditional probabilities.
- If we know $P(A|B_j)$ and $P(B_j)$, then we can obtain $P(B_j|A)$
- Extremely useful for real world applications!

Applying Bayes Rule

Example 2:

My email is divided into 3 folders: Normal, Important, Spam. From past experience, the probability of emails belonging to these folders is 0.2, 0.1, and 0.7 respectively.

- Out of normal emails, "free" occurs with probability 0.01.
- Out of important emails, "free" occurs with probability 0.01.
- Out of spam emails, "free" occurs with probability 0.9.

My spam filter reads an email that contains the word "free". What is the probability that this email is spam?

Applying Bayes Rule Cont.

Define events:

N = email is normal, I = email is important, S = email is spam
F = email contains "free", $\overline{F}$ = email doesn't contain "free"

Given:

$P(N) = 0.2$, $P(I) = 0.1$, $P(S) = 0.7$
$P(F|N) = 0.01$
$P(F|I) = 0.01$
$P(F|S) = 0.9$

$P(S|F) =$? (This is what we want to know)

Applying Bayes Rule Cont.

What is the probability that my email is spam given that it contains the word "free"?

$$
\begin{aligned}
P(S|F) &= \frac{P(S \cap F)}{P(F)} \\
&= \frac{P(S)P(F|S)}{P(S)P(F|S) + P(I)P(F|I) + P(N)P(F|N)} \\
&=
\end{aligned}
$$

Applying Bayes Rule Cont.

Conceptual understanding:

- Before knowing anything
  $\rightarrow$ probability that email is spam was $P(S) = 0.7$.
- After knowing that the email contains the word "free"
  $\rightarrow$ update probability based on this knowledge.
- After knowing the email contains "free"
  $\rightarrow$ probability of the email being spam is $P(S|F) = 0.995$.
- Since this probability is more than 50%, we can *classify* this email as spam.
- In machine learning/statistics, this procedure is called a *naive Bayes classifier*.

Example

Bayes' and LOTP Example

Example 3: Approximately 1% of women aged 40-50 have breast cancer. A woman with breast cancer has 90% chance of testing positive for cancer from a mammogram. A woman without breast cancer has a 5% chance of testing positive for cancer (called a "false positive"). What is the probability that a woman has breast cancer given that she tested positive?

Bayes' and LOTP Example Cont.

Bayes' and LOTP Example Cont.