Unit #2: Probability

Sections 3.1 - 3.3

Learning Objectives

At the end of this unit, students should be able to:

1. Define the terms *sample space*, *outcome*, and *event* (simple and compound).
2. Write out the axioms of probability theory.
3. Prove basic theorems of probability theory.
4. Calculate the probability of events arising from simple experiments (e.g., coin flipping, dice rolling, simple real-world examples). This will require the ability to perform basic counting tasks (e.g., permutations and combinations).
5. Define a *random variable* and calculate the probability that a random variable takes on a particular value.
6. Articulate the importance of an *interpretation* of probability and identify the two main interpretations.
7. Define *conditional probability* and calculate conditional probabilities.
8. Apply the multiplication rule.
9. Define *independence,* and calculate the probability of independent events.
10. Prove Bayes' theorem and the Law of Total Probability; use them in applications.

What is Probability?

What is Probability?

One main objective of statistics/data science is to help make good decisions under conditions of uncertainty.

Examples?

In studying a disease, we might want to know its prevalence in a population. We learn this by studying a sample. The inference from sample to pop. is uncertain.

Probability is one way to quantify outcomes that cannot be predicted with certainty.

Sample Space

Definition: A *probabilistic process* is system/ experiment whose outcomes are uncertain.

Definition: An *outcome* is a possible result of a probabilistic process .

Definition: A *sample space* of a probabilistic process is the set of all possible outcomes of that process.

Set theory

- A set is a collection of objects. outcomes

Notation: $A = \{a_1, a_2, \dots, a_n\}$.

$|A|$ = # of elements / objects in A

Ex. $\mathbb{N} = \{0, 1, 2, 3, \dots\}$

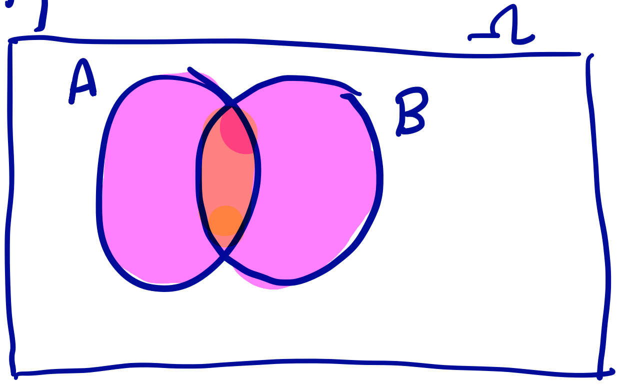
$\mathbb{R}, \mathbb{Z}, \mathbb{Q}, \mathbb{C}$.

- $A = \{1, 2, 3, 5, 7\}$, $B = \{1, 3, 5, 7, 9\}$.

$A \cup B = \{1, 2, 3, 5, 7, 9\}$. $A \cap B = \{1, 3, 5, 7\}$

Set operations

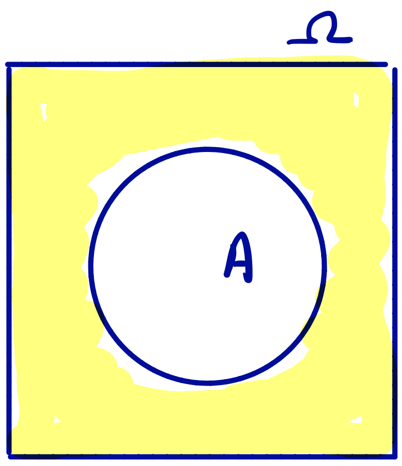
- $A \cup B$ = All elements that are in A <u>or</u> B



- $A \cap B$ = All elements that are in both A <u>and</u> B.

- A^c = All elements <u>not</u> in A (but in the sample space)

Ω



Sample Space

Ω

Sample spaces for:

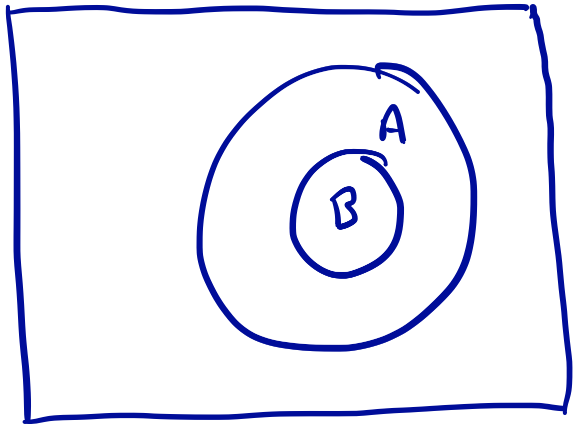
Tossing a coin

$\Omega = \{H, T\}$, $|\Omega| = 2$

Selecting a card from a deck

$\Omega = \{2\heartsuit, 2\diamondsuit, 2\spadesuit, 2\clubsuit, \ldots A\heartsuit, A\diamondsuit, AS, AC\}$, $|\Omega| = 52$

Measuring your commute time on a particular morning



Events

Definition: An *event* is any collection (subset) of outcomes from the sample space.

An event is *simple* if it consists of exactly one outcome and *compound* if it consists of more than one outcome.

When an experiment is performed, an event A is said to occur if the resulting experimental outcome is contained in A.

Events

Example: Suppose that we flip a coin 3 times. [A kid at Starbucks just dropped a bunch of coins while I was typing that sentence.]

Sample space: $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

$|\Omega| = 8$

- $A \cup C = \Omega$
- $A \cap C = \{HTT, THT, TTH\}$

$(A \cup C)^c = \emptyset$

Possible Event(s):

A = At least Two tosses are tails $= \{HTT, THT, TTH, TTT\}$

B = At most one head $= A$

C = At least one head $= \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$

Events

Example: Suppose we'd like to study how accurately baristas prepare drinks. We randomly sample 20 small drinks and weigh them. We want to see how heavy they are on average, and how much variation there is in the weight.

Sample space: $\Omega = (0, 12oz)$

Possible Event(s):

$\Omega = \{H, T\}$

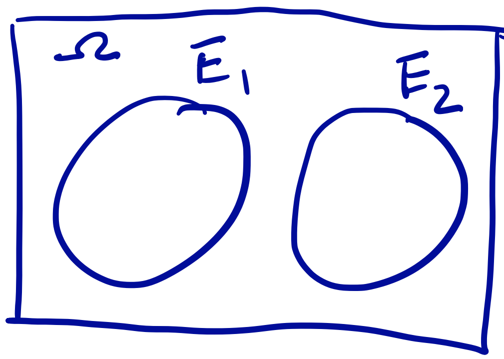
Note: $\mathcal{P}(\Omega)$ = the power set = set of all possible subsets

Rules (Axioms) of Probability

Probability is a function that takes in sets (and later, we'll see, *random variables*) and outputs numbers according to the following rules:

Let E and F be sets in $\underline{\underline{\mathcal{E} \subset \mathcal{P}(\Omega)}}$. Let $P : \mathcal{E} \to \mathbb{R}$, where

P satisfies:



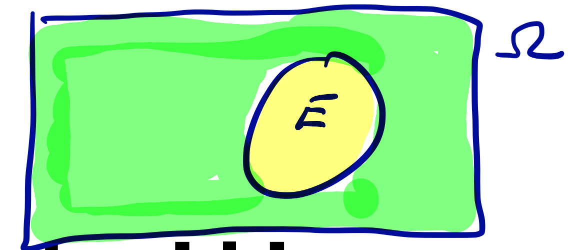
1.) $P(\Omega) = 1$.

2.) For any event $E \in \mathcal{E}$, $P(E) \geq 0$

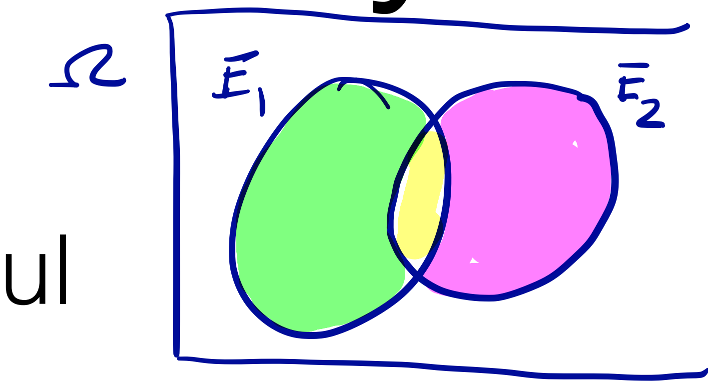
3.) Let E_1 and E_2 be events such that $E_1 \cap E_2 = \emptyset$.
Then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

$$E^c = \Omega \setminus E$$

$\cdot\ E \cup E^c = \Omega, \quad E \cap E^c = \emptyset$



Theorems of Probability



From the axioms, we can prove many useful theorems of probability.

1.) For any event E, $P(E^c) = 1 - P(E)$.

$$\underbrace{1 = P(\Omega)}_{\text{axiom \#1}} = \underbrace{P(E \cup E^c) = P(E) + P(E^c)}_{\text{axiom \#3}} \Rightarrow 1 = P(E) + P(E^c)$$

$$\Rightarrow P(E^c) = 1 - P(E).$$

2.) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$. Consider $E_1 \setminus [E_1 \cap E_2]$, $E_1 \cap E_2$ and $E_2 \setminus [E_1 \cap E_2]$. These sets are disjoint.

$$P(E_1 \cup E_2) = P\Big[E_1 \setminus [E_1 \cap E_2] \cup [E_1 \cap E_2] \cup E_2 \setminus [E_1 \cap E_2]\Big] = P(E_1 \setminus [E_1 \cap E_2]) + P(E_1 \cap E_2) + P(E_2 \setminus [E_1 \cap E_2])$$

$$= P(E_1) - P(E_1 \cap E_2) + P(E_1 \cap E_2) + P(E_2) - P(E_1 \cap E_2) = \text{done!}$$

Random Variables

Definition: a *random variable* is a (measurable) function that maps events to the real numbers (or, more generally, to a measurable space…whatever that is!)

Examples:

X = the sum of the faces on 3 dice. $(X \in \{3, 4, \ldots, 18\})$

Y = # of T in 10 flips of a coin $(Y \in \{0, 1, \ldots, 10\})$

Z = height of a randomly selected CU Student $(Z \in [a, b]$ for $a, b \in \mathbb{R})$

Probabilities of Random Variables

Let X = # of heads in three tosses of a fair coin.

What is the underlying probabilistic process?

Coin Toss x 3

What is the sample space?

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

What are the possible values for X?

$X \in \{0, 1, 2, 3\}$

$\frac{3}{8} = \frac{|A|}{|\Omega|}$

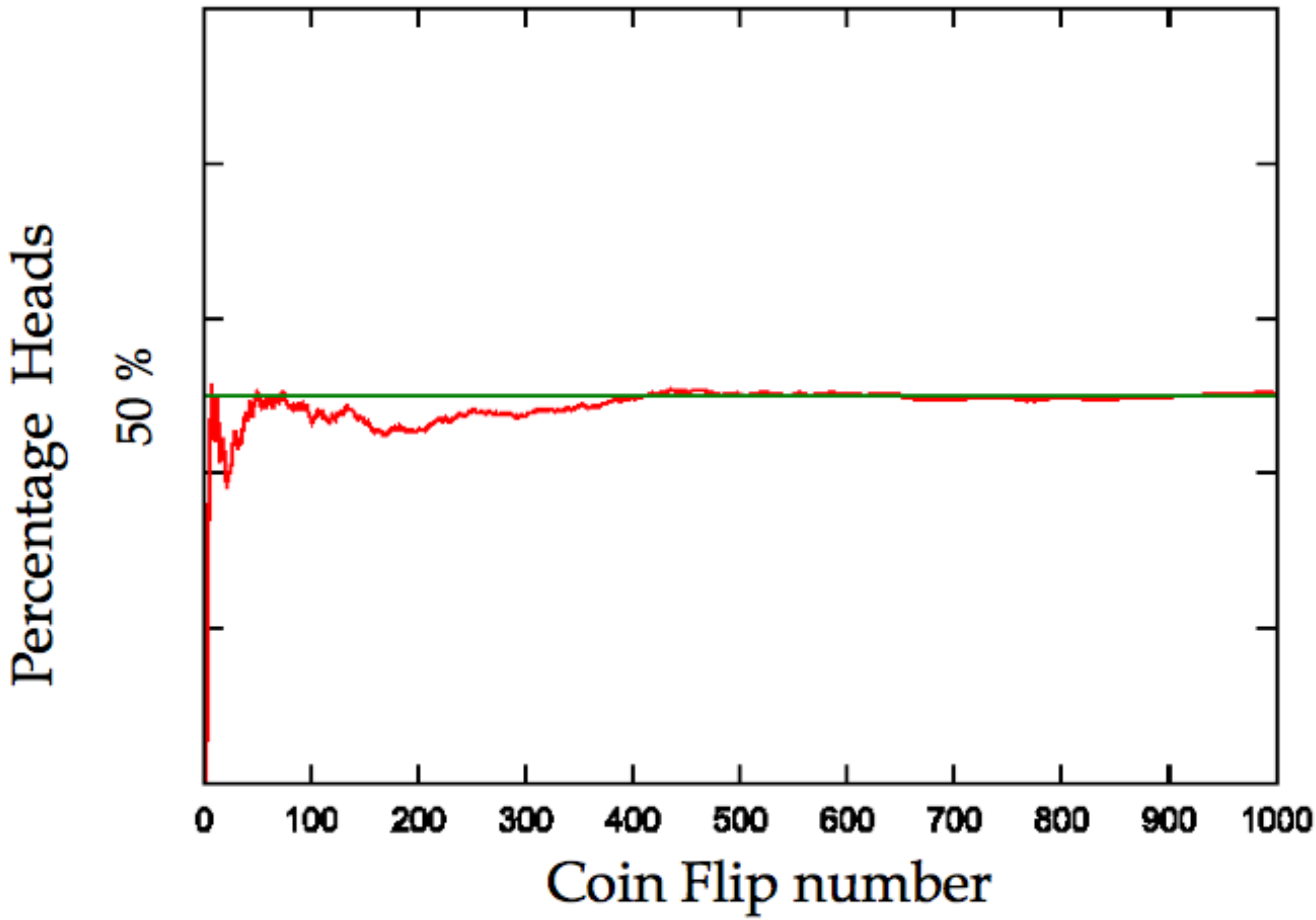
What is the probability that X is equal to 1: $\boxed{P(X = 1)}$?

Interpretations of Probability

Although P is well-defined mathematically, how we ought to *interpret* P in real world situations is not always clear. E.g., coin flips vs. P(rain).

In the last example, we relied on the **relative frequency interpretation** of probability to calculate $P(X = 1)$.

This interpretation says that P is just long run relative frequency.


Percentage Heads
50 %
0
100
200
300
400
500
600
700
800
900
1000
Coin Flip number

Interpretations of Probability

Another interpretation:

The **subjective interpretation** of probability is also popular.

This interpretation says that P represents one's "subjective degree of belief" in a claim about a random process.

There are other interpretations, and reasonable people disagree about the best interpretation. This has real consequences for statistical practice!

(For more, take my philosophy of statistics course!)

Examples

1.) Suppose that your Music app contains 100 songs, 10 of which are by the Beatles. Using the shuffle feature, what is the probability that the first Beatles song heard is the fifth song played?

w/ out replacement:

$$P(\text{'1st B song is 5th song'}) = \frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 10}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = 0.0679$$

2.) Probability of a flush (5 card hand)

② K K * K *

③ * * K K K

3.) Probability of 3 kings (5 card hand)

① K K K * *

$$P(\text{'three kings'}) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} \cdot {}_5C_3 = 0.0017\ldots$$

$${}_nC_k = \frac{n!}{(n-k)!\,k!}$$

P('roll 6 given even') = $\frac{1}{3}$

Conditional Probability

The purpose of conditional probability is to examine how the information "an event *B* has occurred" affects the probability assigned to *A.*

Example: We might refer to an individual having a particular disease in the presence of certain symptoms. If a blood test (an imperfect one…) is performed on the individual and the result is negative, then the updated probability of disease will be different than if the test result was positive.

Conditional Probability

We will use the notation $P(A|B)$ to represent the conditional probability of event A given that the event B has occurred. B is the "conditioning event."

Conditional probability is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

Example

Specific computer parts are assembled in a plant that uses two different assembly lines, *A* and *A'*. Line *A* uses older equipment than *A'*, so it is somewhat slower and less reliable.

Suppose that, on a given day, line *A* has assembled 8 parts, whereas *A'* has produced 10.

From the 8 parts from *A*, 2 were defective and 6 as not defective. From the 10 parts from *A'*, 1 was defective and 9 not defective.

$P(D) = \frac{3}{18} = \frac{1}{6}$

Example

The manager chose a part that turned out to be defective – ie, the event *D={defective part}* has occurred.

What is the chance that it was made by the line *A*?

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{1/9}{1/6} = \frac{2}{3}$$

	D	$\sim D$
A	2	6
A'	1	9

How is this different than the probability that line A makes a defective part?

The Multiplication Rule

The definition of conditional probability yields the following result:

$$P(A \cap B) = P(A|B)P(B)$$

(comes from the def. of conditional prob)

Independent Events

Definition: Two events A and B are *independent* if

$$P(A|B) = P(A)$$

and are dependent otherwise.

Multiplication Rule for Independent events:

$$P(A \cap B) = P(A)P(B)$$

($\underline{K}$ $\underline{K}$ $\underline{K}$ $\underline{*}$ $\underline{*}$ w/ replacement uses mult rule for dep. events)

Independence for More than Two Events

Definition: Events $A_1, \ldots, A_n$ are *mutually independent* if for every k ($k = 2, 3, \ldots, n$) and every subset of indices $i_1, i_2, \ldots, i_k$:

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$$

Example

You a roll a 6-sided die twice and hope for two fours. What is the probability you roll two fours?

$$P(\langle 4,4\rangle) = P(\text{'4'})\underbrace{P(\text{'4'}\mid\text{'4'})}_{=P(\text{'4'})} = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$$

If your first roll is a four, what is the probability that your second roll is a four?

$1/6$ (independent events)

M = Students taking math

S = seniors

$M \cap S$

$P(S) = \frac{25}{120}$;

Example

$P(M \cap S) = \frac{40}{1200}$

In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. One student is selected at random. Let *S* represent the a senior is chosen, and *M* represent that a student taking math is chosen.

If the randomly chosen student is a senior, then what is the probability that they are taking math?

$$P(M|S) = \frac{P(M \cap S)}{P(S)} = \frac{4}{120} \cdot \frac{120}{25} = \frac{4}{25}$$

Are these events independent?

$$P(M) = \frac{150}{1200} \neq \frac{4}{25}$$

Law of Total Probability

$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$

So,

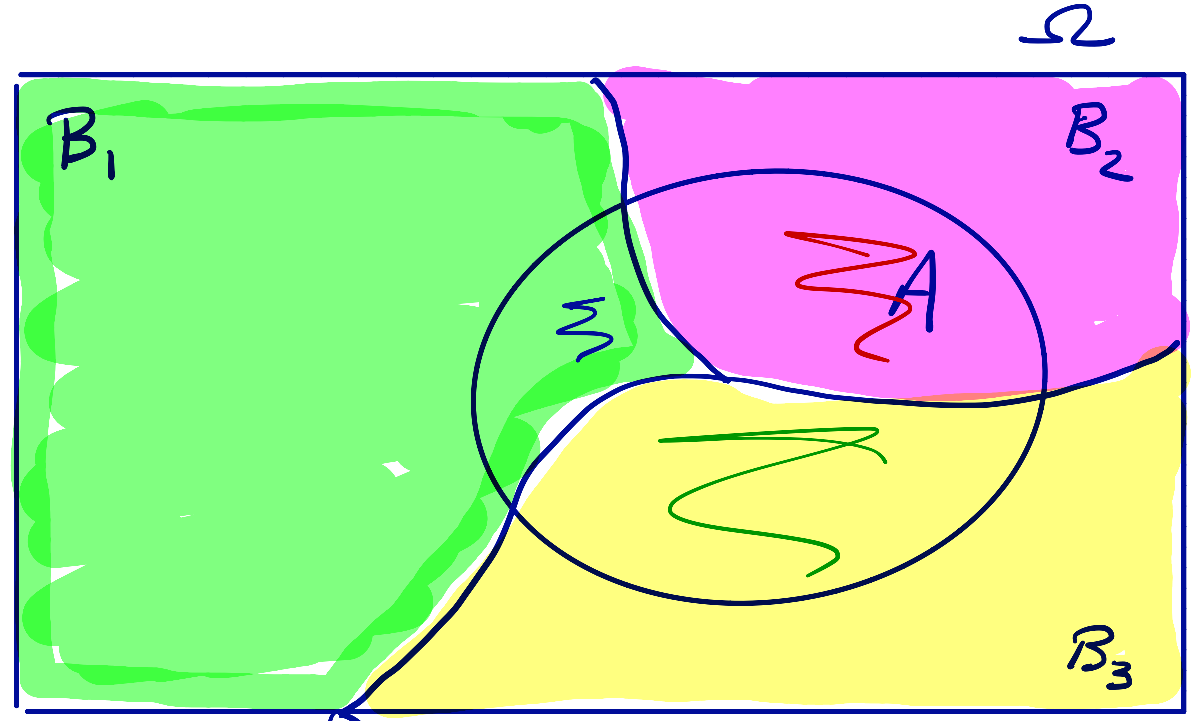
$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$

holds by the 3rd axiom.

Note that $P(A|B_i) = \dfrac{P(A \cap B_i)}{P(B_i)}$

$\Rightarrow P(A \cap B_i) = P(A|B_i)P(B_i)$.

Thus,



Partition: ① Note: $B_1 \cup B_2 \cup B_3 = \Omega$ $\left(\bigcup_{i=1}^{3} B_i = \Omega\right)$

② $B_i \cap B_j = \emptyset$, $i \neq j$ $(i, j \in \{1, 2, 3\})$

$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$

In general:

$$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \quad \text{where } \bigcup_{i=1}^{n} B_i = \Omega, \text{ and } B_i \cap B_j = \emptyset \ \ i \neq j.}$$

Note: $P(A|B) \neq P(B|A)$

Bayes' Theorem

Let A and B be events such that $P(B) > 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Note: ① $P(A|B)$ is called the posterior probability

② $P(B|A)$ is called the likelihood.

③ $P(A)$ is called the prior prob.

Now, let $A_1, \ldots, A_k$ be a partition.

Then,

$$\boxed{P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}}$$

$P(A_1) = 0.7$

$P(A_2) = 0.2$

$P(A_3) = 0.1$

Example

An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

$P(S|A_1) = 0.01$, $P(S|A_2) = 0.02$, $P(S|A_3) = 0.05$

$$P(S) = P(S|A_1)P(A_1) + P(S|A_2)P(A_2) + P(S|A_3)P(A_3) = 0.016.$$

LoTP

Example

An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively.

Say she randomly selected a message and it was indeed spam. What is the probability that it came from account #1?

$$P(A_1|S) = \frac{P(S|A_1)P(A_1)}{P(S)} = \frac{(0.01)(0.7)}{0.016} = 0.4375\ldots$$

$P(A) = 0.8$

$P(B) = 0.15$

$P(C) = 0.05$

Example

$P(D|A) = 0.04$

$P(D|B) = 0.06$

$P(D|C) = 0.09$

An aircraft emergency locator transmitter (ELT) is a device designed to transmit a signal in the case of a crash. The Altigauge $\leftarrow A$ Manufacturing Company makes 80% of the ELTs, the Bryant Company makes 15% of them, and the Chartair Company makes the other 5%.

The ELTs made by Altigauge have a 4% rate of defects, the Bryant ELTs have a 6% rate of defects, and the Chartair ELTs have a 9% rate of defects.

A randomly selected ELT is tested and is found to be defective. What is the probability that it was made by the Altigauge Manufacturing Company? $P(A|D) = \dfrac{P(D|A)P(A)}{P(D)}$ where $P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)$

$0.7032\ldots$

Which company do you suspect has the lowest market share? Why?

Let’s do some work in R!