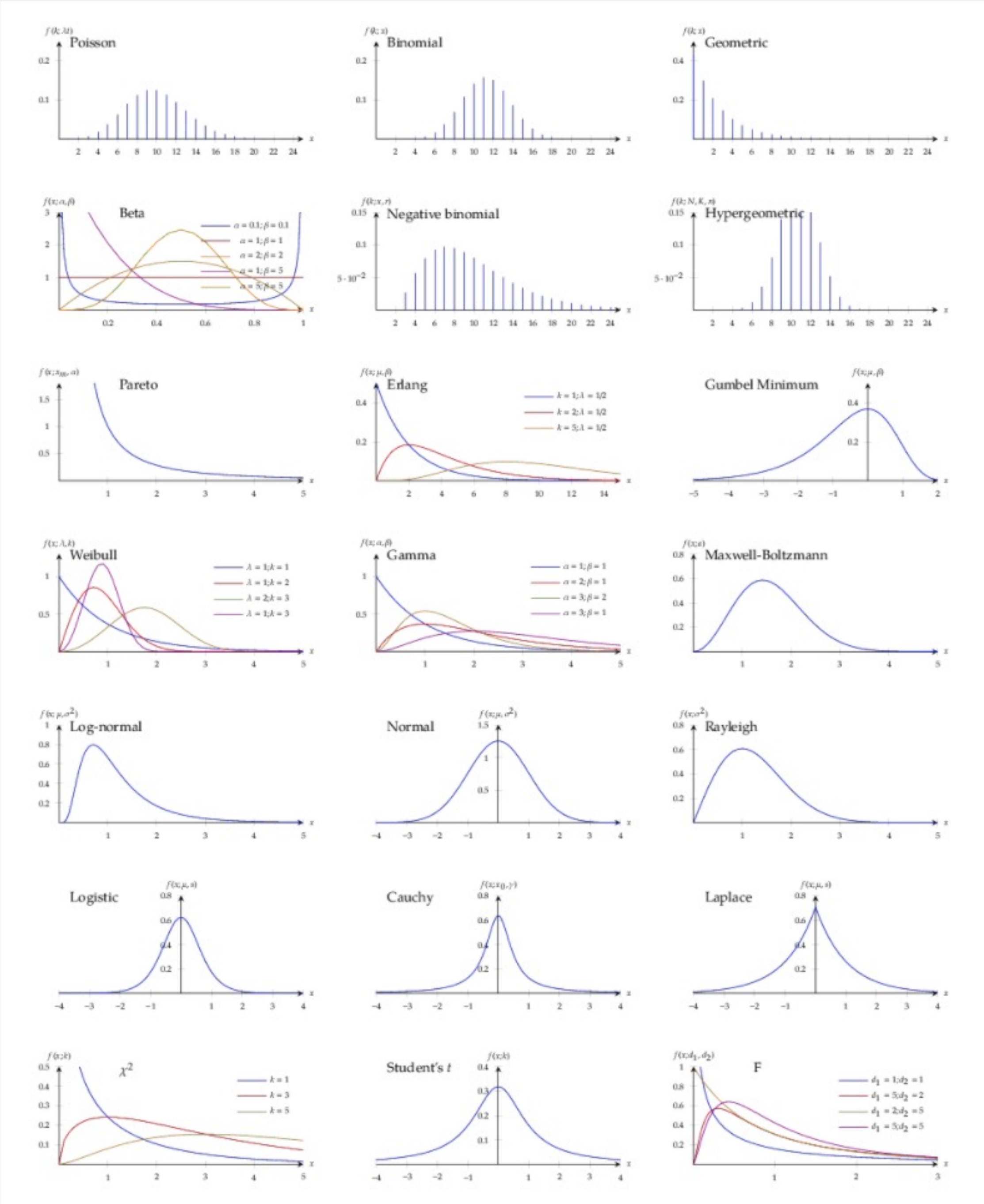
PROBABILITY

Chapter 2



- Probability and more specifically probability distributions are used to model data sets.

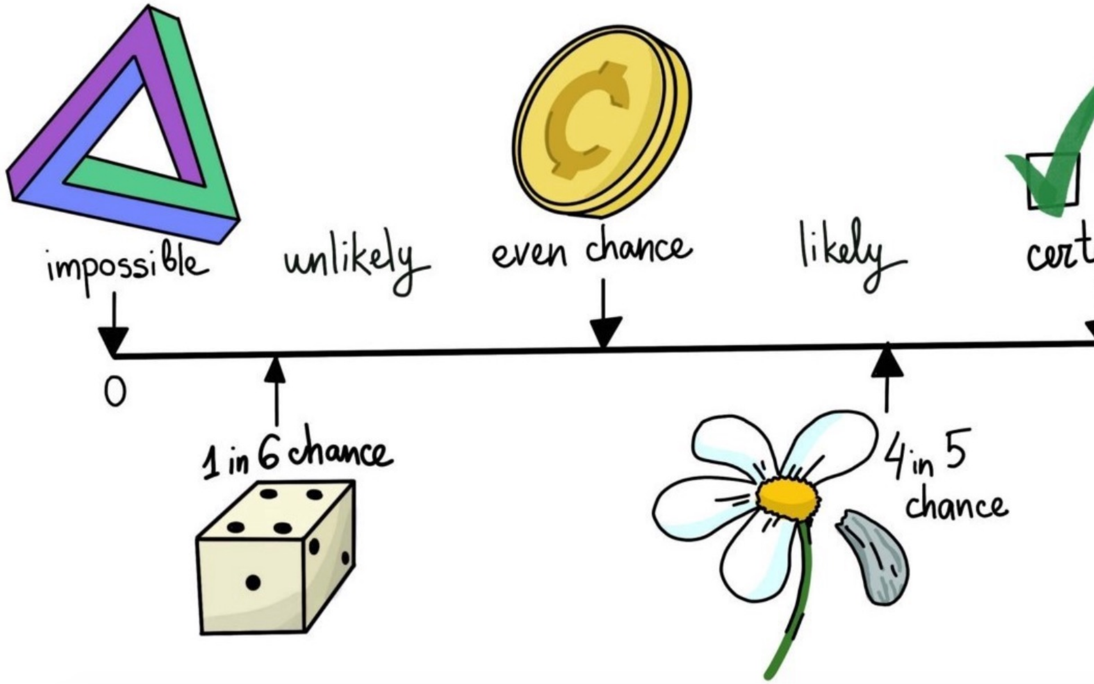
- We can study and/or ‘assign’ a probability distribution to a set of data in an attempt to discover patterns and relationships.

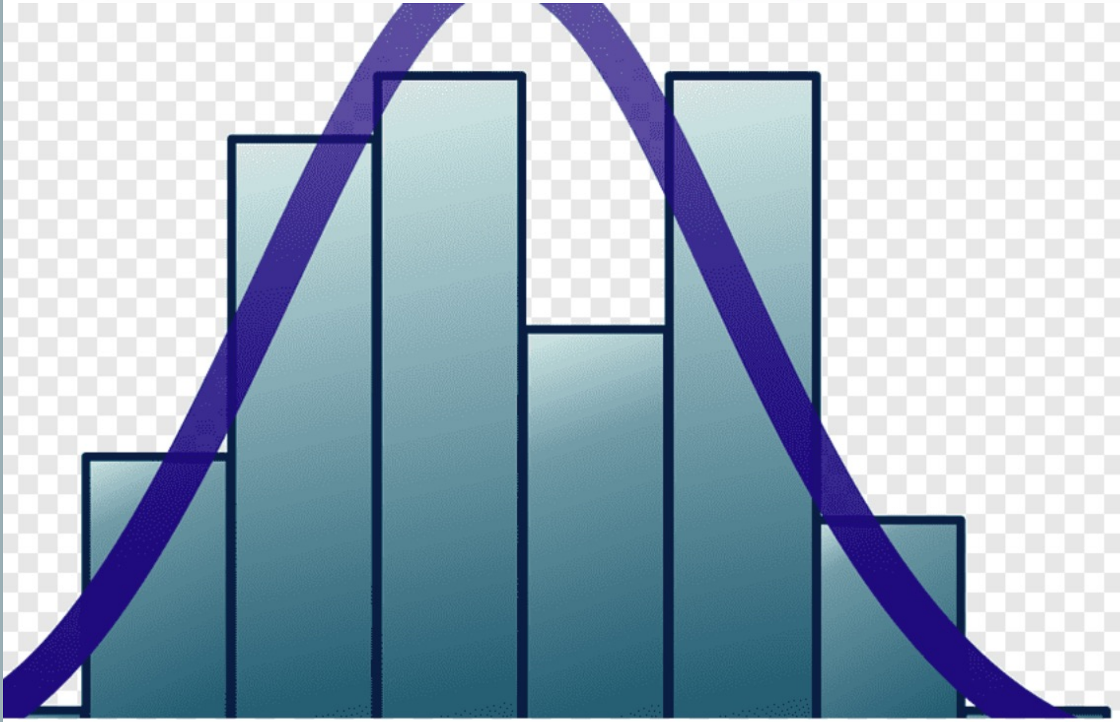
Essentially, we are studying phenomena via datasets.

* We perceive the phenomena as random,
  or unpredictable.
* We will model these phenomena as outcomes
  of an experiment.
* The outcomes are elements of a sample space Ω
* Subsets of Ω are called events.
* Events will be assigned a probability
  (a number between 0 and 1)

Hence our study of probability.

This slide show is a refresher of probability and the terminology and ideas used in this area of study.





Vocabulary

Experiment: A procedure that can be repeated and has well defined outcomes.
Outcomes: results of an experiment
Sample Space Ω: The name of the set of outcomes.
Events: Subsets of Ω that are assigned likelihood, $P(event)$, of occurrence. This likelihood or probability p is always a number $0 \leq p \leq 1$.

Example

Experiment:	Flipping a coin
Outcomes:	could land on 'heads' or 'tails'
Sample Space Ω:	$\Omega = \{H, T\}$
Event:	Coin lands on 'Tails' has probability 0.5 for a fair coin.

Example 2

Your birthday is in February.

You want to know who has a birthday within 1 month of yours on either side.

You collect data by asking people on Pearl Street their Birthday month.

Experiment:

Outcomes:

Sample Space Ω:

Event:

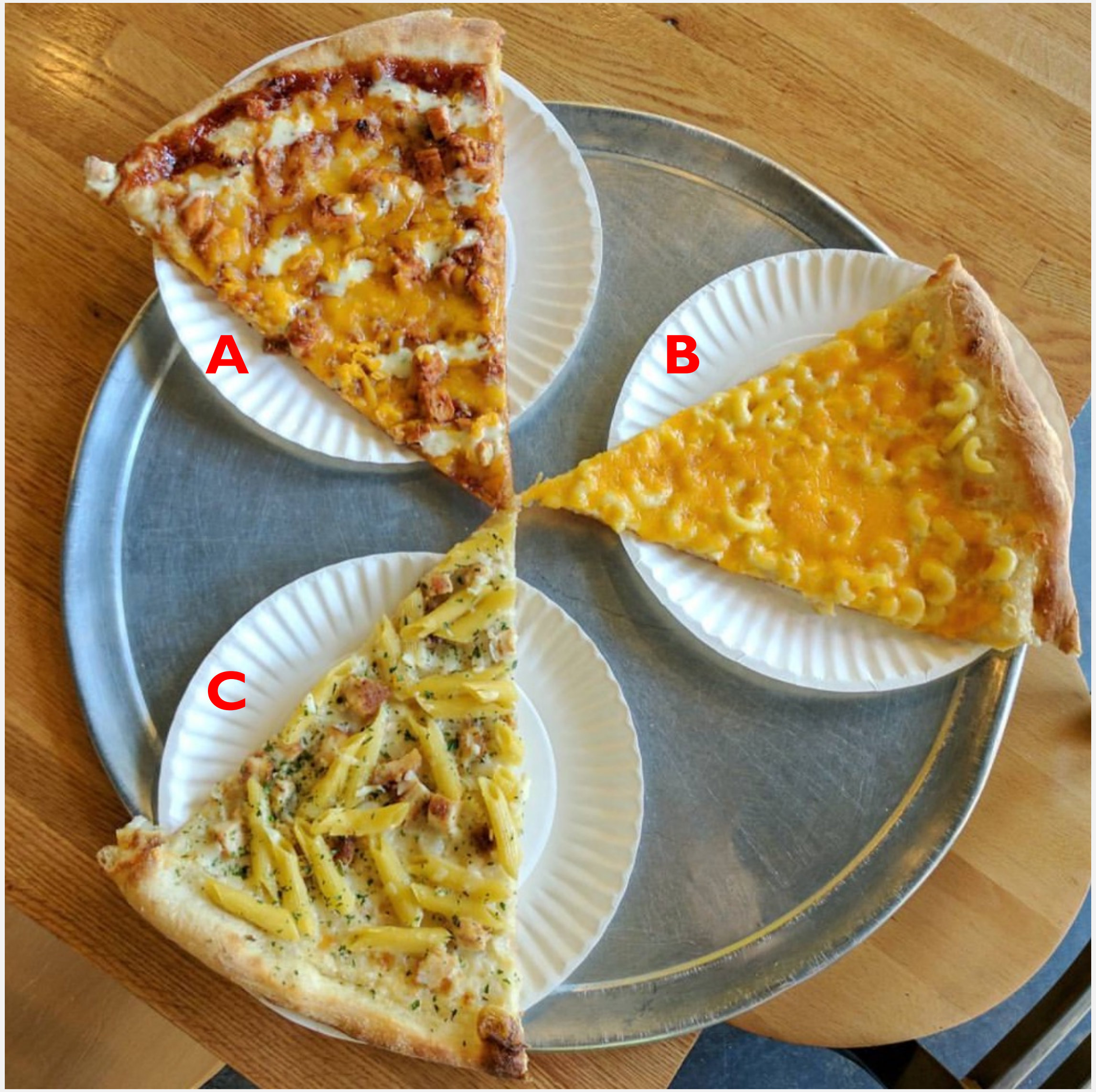
Example 3

- Experiment: You call for pizza deliveries from 3 different locations at the same time in order to see the order in which they arrive.

- $\Omega = \{ABC, ACB, BAC, BCA, CAB, CBA\}$

- In general, the order in which n different objects can be placed is $n!$

- Q: What is $|\Omega|$ if you called 4 pizza deliveries?
- A:



Events can be combined

Suppose our sample space contains the names of the months:

$$\Omega = \{\text{Jan}, \text{Feb}, \text{March}, \text{April}, \text{May}, \text{June}, \text{July}, \text{Aug}, \text{Sept}, \text{Oct}, \text{Nov}, \text{Dec}\}$$

We can create events, or subsets of Ω

$S = \{Feb, April, June, Sept, Nov\}$ The subset of short months

$L = \{Jan, March, May, July, Aug, Oct, Dec\}$ The subset of long months

$T = \{Jan, Feb, March\}$ The subset of months close to February

$R = \{Jan, Feb, March, April, Sept, Oct, Nov, Dec\}$ The subset of months with 'R'

$T \cup S$

$T \subset R$

$S \cup L$

Ω^c

$R \cap S$

$S \cap L$

Events can be combined

Suppose our sample space contains the names of the months:

$$\Omega = \{\text{Jan, Feb, March, April, May, June, July, Aug, Sept, Oct, Nov, Dec}\}$$

We can create events, or subsets of Ω

$S = \{Feb, April, June, Sept, Nov\}$ The subset of short months

$L = \{Jan, March, May, July, Aug, Oct, Dec\}$ The subset of long months

$T = \{Jan, Feb, March\}$ The subset of months close to February

$R = \{Jan, Feb, March, April, Sept, Oct, Nov, Dec\}$ The subset of months with 'R'

$T \cup S = \{Jan, Feb, March, April, June, Sept, Nov\}$

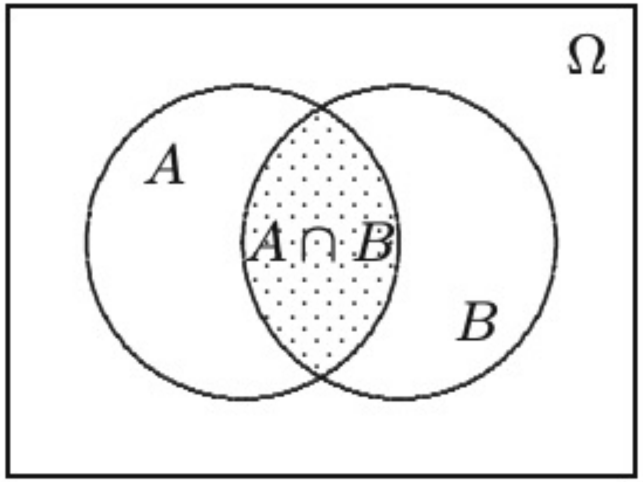
$T \subset R$ is read 'T implies R'

$S \cup L = \Omega$

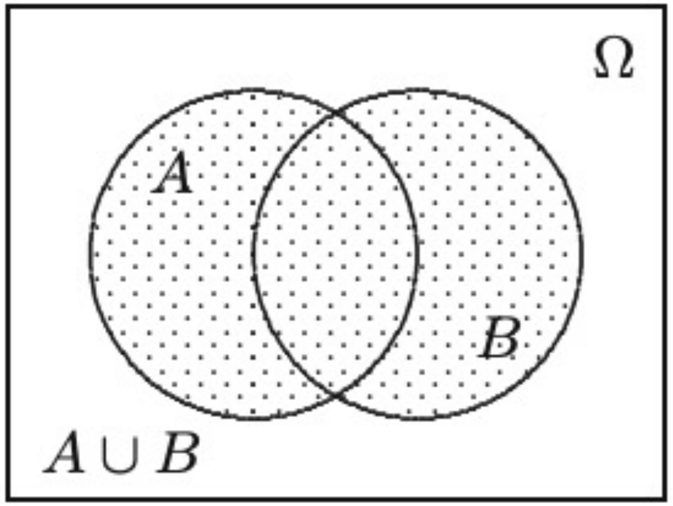
$\Omega^c = \emptyset$

$R \cap S = \{Feb, April, Sept, Nov\}$

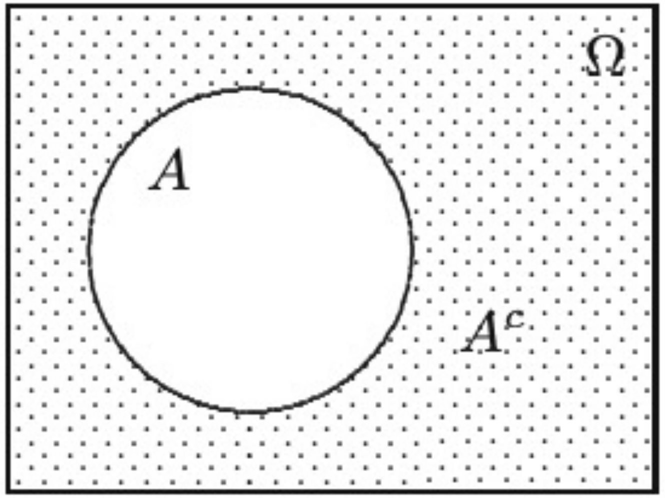
$S \cap L = \emptyset$ Disjoint or mutually exclusive sets.



Intersection $A \cap B$



Union $A \cup B$



Complement A^c

GRAPHIC REPRESENTATIONS OF SET NOTATION

Do these sentences mean the same thing?

- John or Mary is to blame, or both.

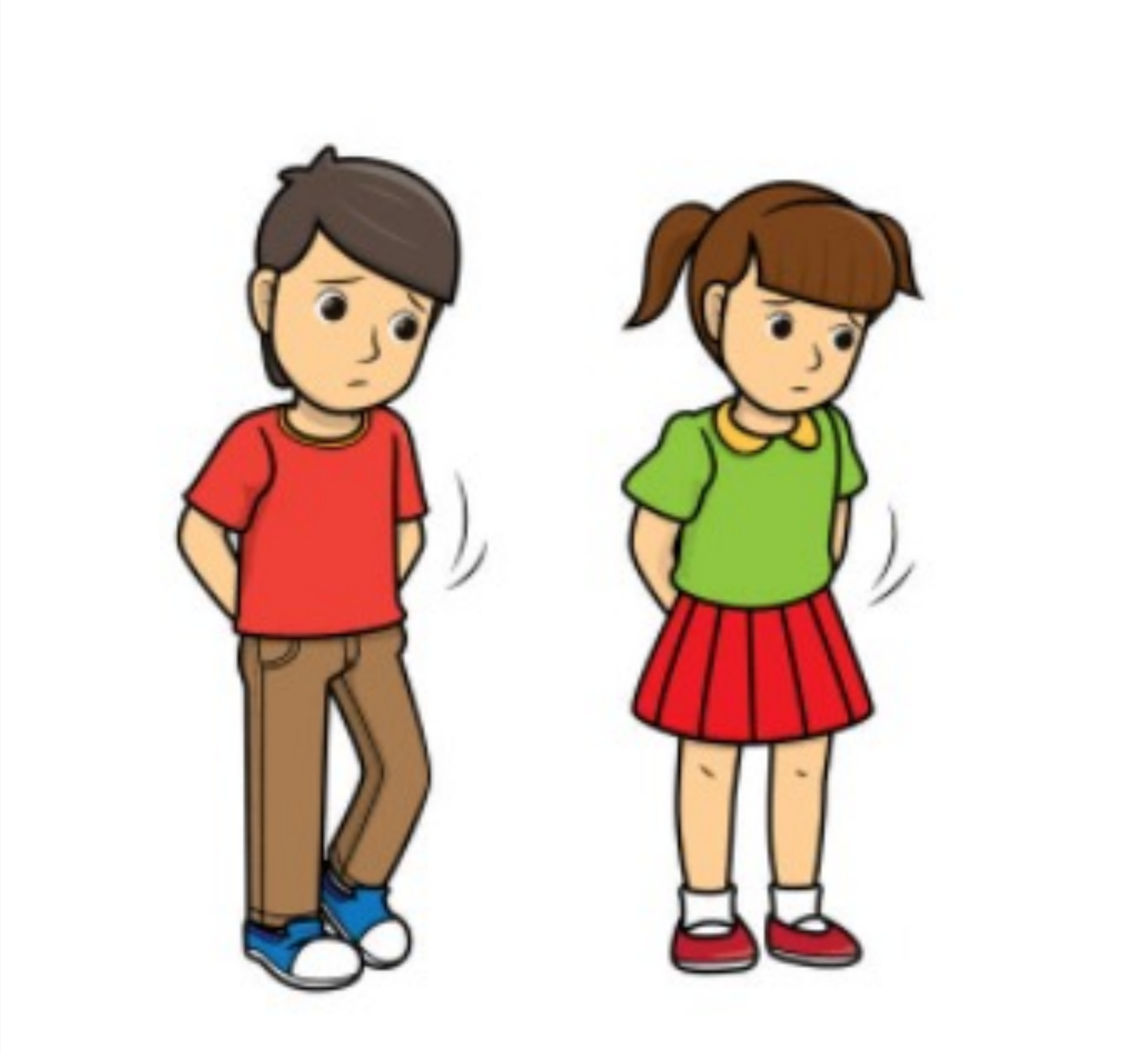
- It is certainly not true that neither John nor Mary is to blame.

This mental operation can be done with manipulation of sets (events)

DeMorgan's Laws

$$(A \cup B)^C = A^C \cap B^C$$

$$(A \cap B)^C = A^C \cup B^C$$



J:"John is to blame" M:"Mary is to blame"

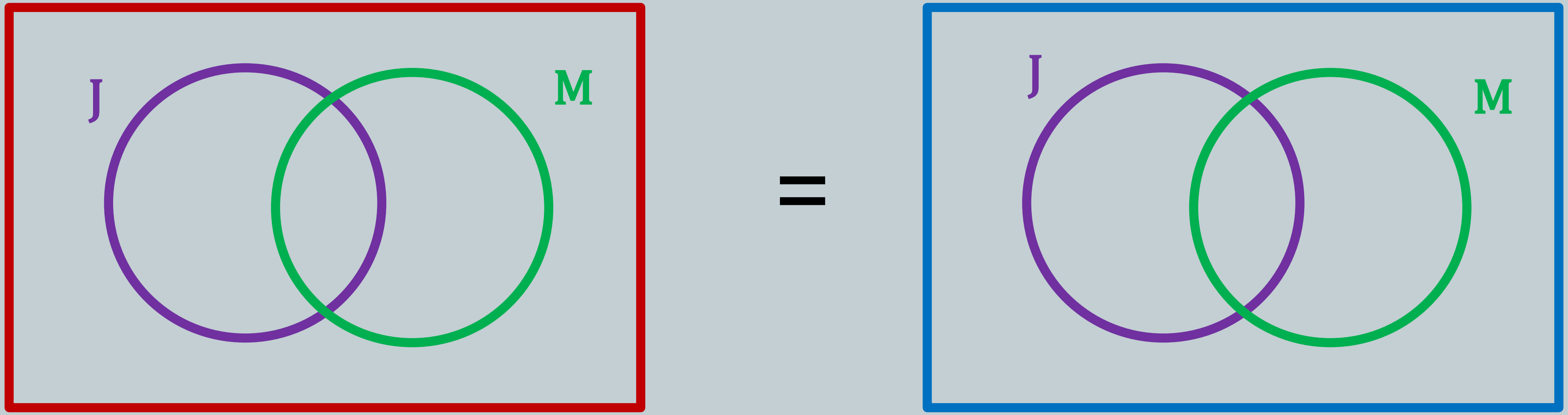
J^c:"John is not to blame" M^c: "Mary is not to blame"

It is certainly not true that neither John nor Mary is to blame.

is the same as

John or Mary is to blame, or both.

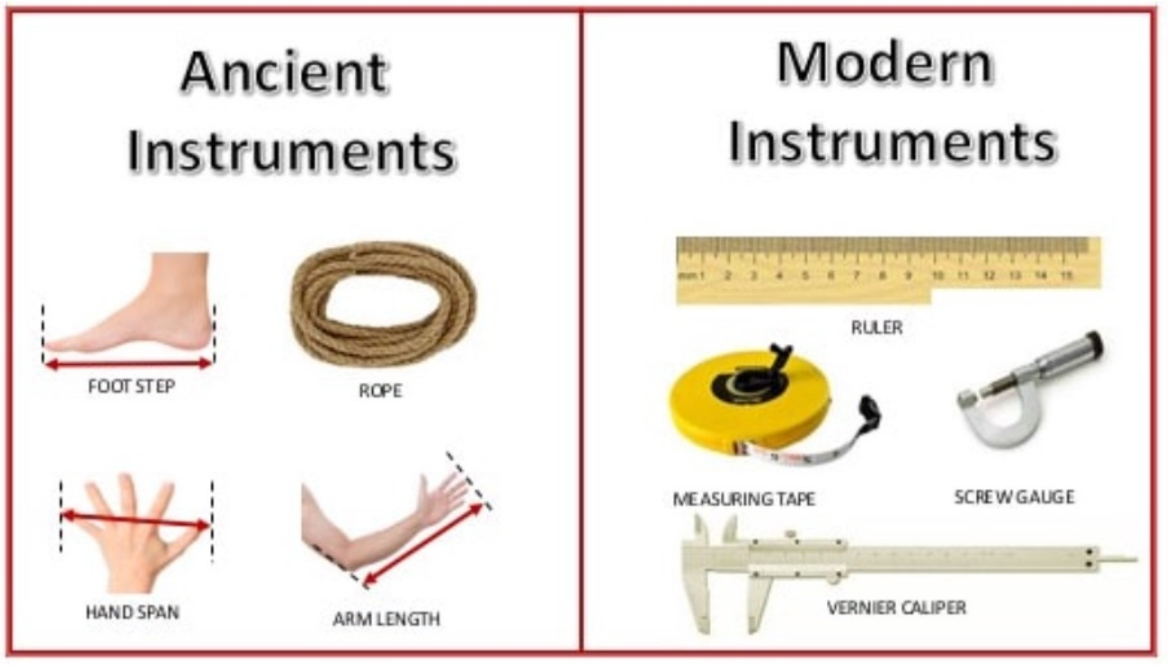
$$(J^c \cap M^c)^c = J \cup M$$



We like to measure things!

And everything we measure has its own scale or metric.

- ‘12’ is a shoe size
- ‘XL’ is a shirt size
- ‘AAA’ is a sports rating
- ‘B+’ is a grade
- ◆◆ is a ski slope rating
- ‘5.10’ is a climb rating
- ‘Lieutenant’ is a ranking



Now we want to express how likely it is that an event will occur.
To do so, we will assign a probability, ‘ x ’, where $0 \leq x \leq 1$.

The assignment of a probability to an event, E, will be done with a function.
For instance, we might say $P(E) = .66$

When assigning probabilities, we need to satisfy two basic properties:

A probability function, P, on a finite sample space, Ω,
assigns to each event A in Ω a number $P(A)$ in $[0, 1]$ such that

(i) $P(\Omega) = 1$, and

(ii) $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint.

The number $P(A)$ is called the probability that A occurs.

The domain of probability functions are events.
For shorthand purposes though we will write $P(H)$ instead of $P(\{H\})$.

So, in the experiment of flipping a coin we could write $P(H) = 0.5$,
unless the coin was slightly bias, and then we might write $P(H) = 0.4999$

Assigning probabilities can range from the simple to the complex.

- Experiment: Roll a fair 6-sided die.
- Event: Roll a multiple of 3
- $P(E) = ?$

- Experiment: Ask someone their birth month
- Event: Being born in February
- $P(E) = ?$

Assigning probabilities can range from the simple to the complex.

Experiment: Call three pizza deliveries and see who delivers first.

Event 1: The delivery is in the order CBA

Event 2: Company A delivers first.

$P(E1) = ?$

$P(E2) = ?$

Experiment: Ask the birth month of an individual.

Event: The birth month was a short month or a month with 'R' in it.

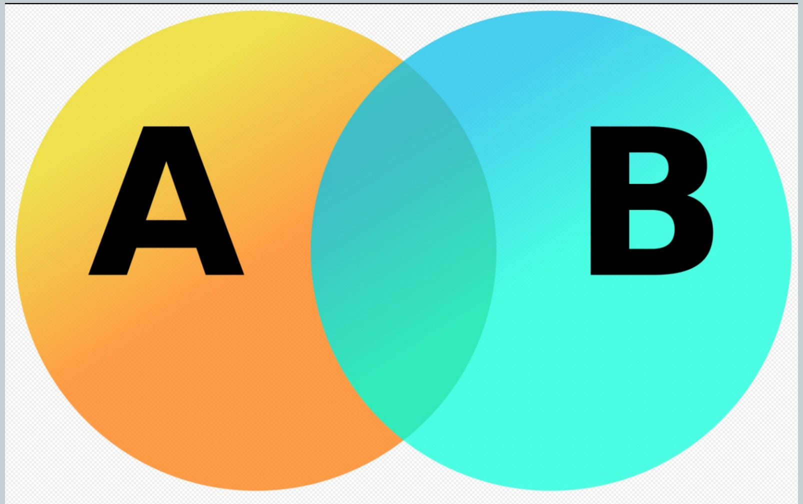
$P(E) = ?$

$S = \{Feb, April, June, Sept, Nov.\}$

$R = \{Jan, Feb, March, April, Sept, Oct, Nov, Dec\}$

$P(S) + P(R) = \frac{13}{12}$??? $13 - 4 = 9$

The previous example points out the inclusion/exclusion principle



The probability of a union

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

This leads to another useful idea

If $A \cup A^c = \Omega$, then $P(A^c) = 1 - P(A)$

Experiment: Flip a coin 5 times.

Event: Getting <u>at least one</u> 'tails'.

$P(\boldsymbol{E}) = ?$

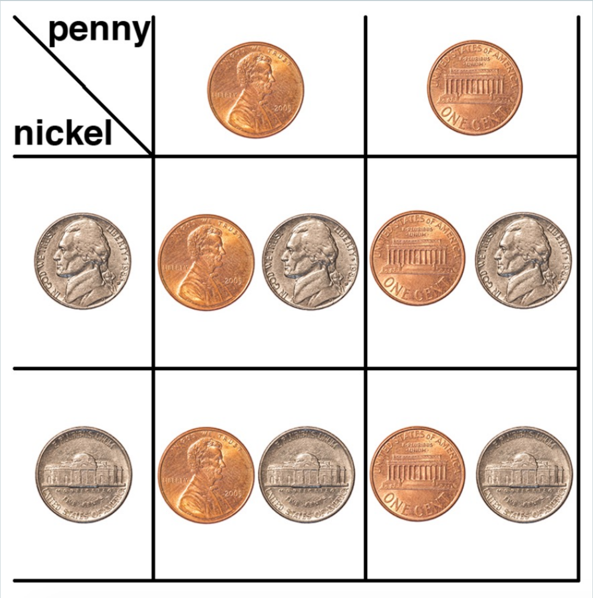
A basic reason for the study of probability in data analysis is that we usually do not consider ONE experiment, but rather we perform the same experiment several times. Then we have a dataset to study randomness.

What is the sample space associated with tossing a coin twice?

$$\Omega = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$$

And if the coin is fair,

then all four outcomes are equally likely; $P(x, y) = \frac{1}{4}$.



In general,

If we consider two experiments with sample spaces Ω_1 and Ω_2 ,

then the combined experiment has as its sample space the set:

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) | \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

And if Ω_1 has r elements and Ω_2 has s elements, then $\Omega_1 \times \Omega_2$ has rs elements.

If all outcomes are equally likely to occur, then $P(\omega_1, \omega_2) = \frac{1}{rs}$ *

* As long as the two experiments do not influence each other in any way

---

Example: A four-sided die has its sides marked 'A', 'B', 'C', and 'D'.

A six-sided die has its sides marked 1 thru 6.

What is the probability of rolling the dice and getting $(B, 4)$?

Recall we spoke of running an experiment n times.

Experiment: flipping a coin twice

Event: Getting two 'Tails'.

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

Run the experiment 5 times: $\{1,0\}\times\{1,0\}\times\{1,0\}\times\{1,0\}\times\{1,0\}$.

A is the event "exactly one experiment was a success"

$$p(A) = ?$$

A: Exactly one experiment was a success

$$A = \{(0,0,0,0,1), (0,0,0,1,0), (0,0,1,0,0), (0,1,0,0,0), (1,0,0,0,0)\}$$

$$P(A) = 5 \cdot (1-p)^4 \cdot p$$

with success having probability p, and failure having probability $1-p$.

Recall the probability of success is the probability of $\{T, T\}$ or $p = \frac{1}{4}$, therefore

$$P(A) = 5 \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right) = 0.3955$$

Experiment: flipping a coin twice

Event B: Getting two 'Tails'.

What is the probability of the event "exactly <u>two experiments</u> were successful"?

Experiment: flipping a coin twice

Event B: Getting two 'Tails'.

What is the probability of the event "exactly two experiments were successful"?

For example $(0, 1, 0, 1, 0)$ or $(1, 1, 0, 0, 0)$ or …

$$P(B) = 10(1 - p)^3 p^2$$

$$p(B) = 10 \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^2 = 0.26367$$

What happens when we come across an experiment with (potentially) infinitely many outcomes?

Experiment: Toss a coin repeatedly until the first head turns up.

The outcome of the experiment is the number of tosses it takes to have this first occurrence of a head.

$$\Omega = \{1, 2, 3, 4, \ldots\}$$

What is the probability function for this experiment?

What is the probability function for this experiment?

Suppose the probability of landing 'heads' is p, then $p(1) = p$

$$
\begin{aligned}
p(1) &= (1-p)^0 p && (H) \\
p(2) &= (1-p)^1 p && (T, H) \\
p(3) &= (1-p)^2 p && (T, T, H) \\
p(4) &= (1-p)^3 p && (T, T, T, H) \\
&\vdots \quad\quad \vdots \\
p(n) &= (1-p)^{n-1} p
\end{aligned}
$$

Is $p(n)$ a probability function?

We need $p(\Omega) = 1$ and additivity.

$$\sum_{n=1}^{\infty} (1-p)^{n-1} \cdot p = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$$

You are an avid bird watcher. Every day you spend 30 minutes searching the nearby woods for the elusive red-belly woodpecker.

You search every day of the week until you spot the bird.
The probability of seeing this bird is p.
You start looking on Monday.

What is the probability that you finally see the bird on the next Sunday?

What is the probability that you finally see the bird on the next Sunday?

This happens if an only if the searching fails on Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and is spotted on Sunday.

$$\underbrace{(1-p)}_{\text{Monday}} \cdot \underbrace{(1-p)}_{\text{Tuesday}} \cdot \underbrace{(1-p)}_{\text{Wednesday}} \cdot \underbrace{(1-p)}_{\text{Thursday}} \cdot \underbrace{(1-p)}_{\text{Friday}} \cdot \underbrace{(1-p)}_{\text{Saturday}} \cdot \underbrace{p}_{\text{Sunday}}$$

$$(1-p)^6 p$$

Next time: Chapter 3 – Conditional Probability