37161 Probability and Random Variables

Lecture 1

Random Experiments

- A random experiment is one whose outcome is not determined in advance.
- Any action which may have more than one possible outcome can be considered to be a random experiment.
- The set of all possible outcomes of a random experiment is called its **sample space**, usually denoted by S or Ω .

Random Experiments

- For example, if the random experiment consists of rolling one die and recording the number shown, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

- If the random experiment consists of rolling three dice and recording if the numbers shown are odd or even, the sample space is odd or even, the sample space is

$$\Omega = \{OOO, OOE, OEO, OEE, EOO, EOE, EEO, EEE\}$$

where, for example, OOE represents the first two dice showing an odd number and the third an even.

Discrete and Continuous Sample Spaces

- The examples on the previous slide (coin flipping and die rolling) have **discrete** sample spaces.

- That is, we can write a list of possible outcomes as separate points.

- For other experiments, this is not possible as we have a **continuous** sample space.

- For example, if the experiment consists of measuring *t*, the time in seconds until the next train arrives is simply all non-negative values of t , $\Omega = \{t : t \geq 0\}$.

Discrete and Continuous Sample Spaces

- Both continuous and discrete sample spaces can be either finite or infinite.

- A discrete sample space is finite if it contains a finite number of points.

- An example of an infinite sample space would result from an experiment recording the mass of an object to the nearest gram $\Omega = \{0,1,2,3,4,...\} = \mathbb{R}^{\geq 0}$.

- A continuous finite sample space needs to have both an upper and lower bound. For example, if we wanted to measure the lifespan, L, to date or until failure (in years) of a lightbulb manufactured 10 years ago, then $\Omega = \{L : 0 \leq L \leq 10\}$.

Events

- A subset of a sample space is called an **event**.

- We say that event A occurs if and only if the outcome of a random experiment is one of the points in A.

- For example if an experiment consists of rolling two dice, the sample space contains 36 elements:

$$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, \ldots, 64, 65, 66\}$$

11	21	31	41	51	61
12	**22**	32	42	52	62
13	23	**33**	43	53	63
14	24	34	**44**	54	64
15	25	35	45	**55**	65
16	26	36	46	56	**66**

- If we define the event A to correspond to "the two dice show the same number" then this is denoted by the subset of Ω given by $A = \{11, 22, 33, 44, 55, 66\}$.

Events

- If an experiment consists of measuring how long, t (in minutes), until the next train arrives, the sample space is an infinite continuous interval $\Omega = \{t : t \geq 0\}$.

- If we define the event B to correspond to “a train arrives within the next ten minutes”, this is denoted by the subset of Ω given by $B = \{t : 0 \leq t < 10\}$.

- Note that since sample spaces can be discrete or continuous and finite or infinite, subsets of these (i.e. events) can also be classified in the same way.
- There can sometimes be possible outcomes of a random experiment which fall into more than one event

Randomness and Uncertainty

- In a random experiment, there is always some degree of uncertainty or unknowability about whether or not an event will occur.

- Being very precise, very few systems we encounter are inherently random. Even the outcome of the flip of a coin is decided by the exact angle it leaves the hand, the height it is allowed to fall, whether or not it will bounce on the landing surface etc.
- Common sources of “randomness” include:

- Incomplete information: not knowing the exact angle/height etc for a coin flip.
- Belief: “I’m 90% sure that it happened.”
- Measurement error: In whole years, he/she gave his/her age as 20, meaning his/her exact age in years is an uncertain value belonging to the interval $[20, 21)$.

Which is the Better Option?

- Assume you are ill and are offered two different new options - Treatment A or Treatment B, both of which are improvements on the old treatment.

- For every 100 patients who use Treatment A instead of the old treatment, on average, 20 more make a full recovery.
- The full recovery rate for Treatment B is 5 times as high as that for the old treatment.

- Which would you choose? Which is the better treatment?

- Answer… I have no idea!

Relative Risk vs Absolute Risk

- It all depends on what the absolute values are.

- If treatment A cures 22 patients per hundred (as opposed to 2 with the old treatment), then it is better than treatment B (which would cure 5 x 2 = 10 patients).
- If treatment A cures 40 patients per hundred (as opposed to 20 with the old treatment), then it is worse than treatment B (which would cure 5 x 20 = 100 patients).
- If treatment A cures 25 patients per hundred (as opposed to 5 with the old treatment), then it is exactly the same treatment B (which would cure 5 x 5 = 25 patients).

- Relative risks alone tell us very little.

Relative Risks Alone Tell Us Very Little

Cholesterol drugs raise diabetes risk by 9 percent

Friday, April 23, 2010 by: Ethan A. Huff, staff writer
Tags: cholesterol drugs, diabetes, health news

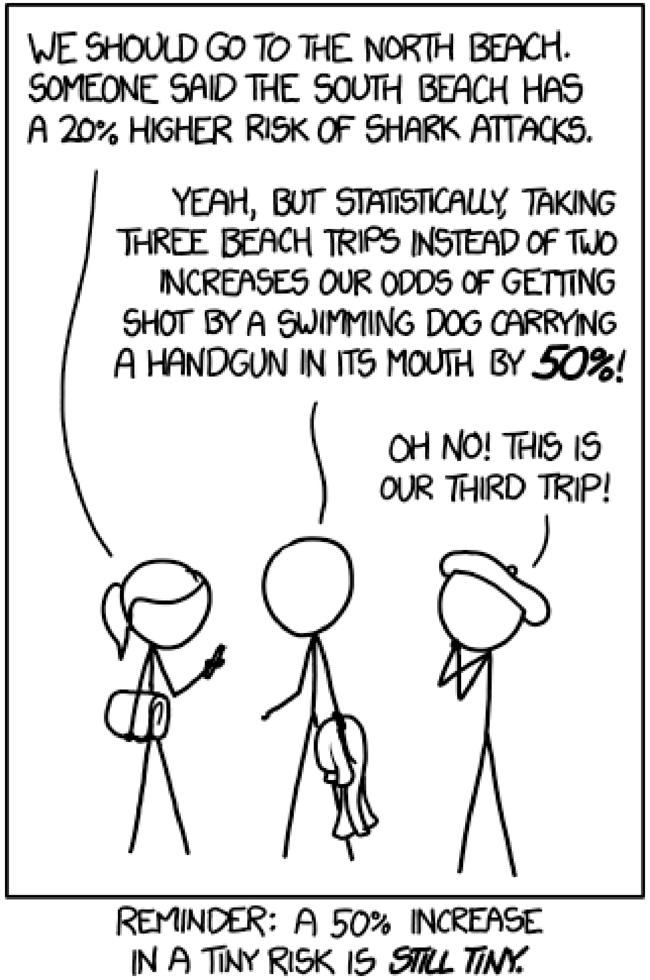
Why taking an afternoon nap 'raises risk of an EARLY DEATH by a third'

- Adults who nap are more likely develop deadly respiratory issues
- Scientists say napping can trigger inflammation in the body
- The findings from say dozing could be a symptom of lung disease

By PAT HAGAN

PUBLISHED: 22:50 GMT, 11 April 2014 | **UPDATED:** 07:03 GMT, 12 April 2014

Share Tweet Pin it +1 Share **9,149** shares 679 View comments



- Annoyingly) common to see articles quoting only relative risk ("*x*% increase") or sometimes no risk at all ("increase").

Probability

- In order to fairly assess the chance of random events occurring, we need some absolute measure of chance.

- We can define a probability through relative frequencies. That is, if we can look at a large number of "identical situations", $\text{Probability of } A = P(A) \approx \dfrac{\text{Number of times event } A \text{ occurs}}{\text{Number of trials}}$.

- If we flipped a coin a million times and observed Tails on half a million flips, we might estimate $P(Tails) \approx \dfrac{1}{2}$.

- This is known as the **frequentist** interpretation of probability.

Probability

- The classical definition of probability is simply that if all individual outcomes are equally likely, then the probability of a given event occurring is simply the proportion of the total number of outcomes that correspond to that event.
- That is, $P(A) = \frac{|A|}{|\Omega|}$.

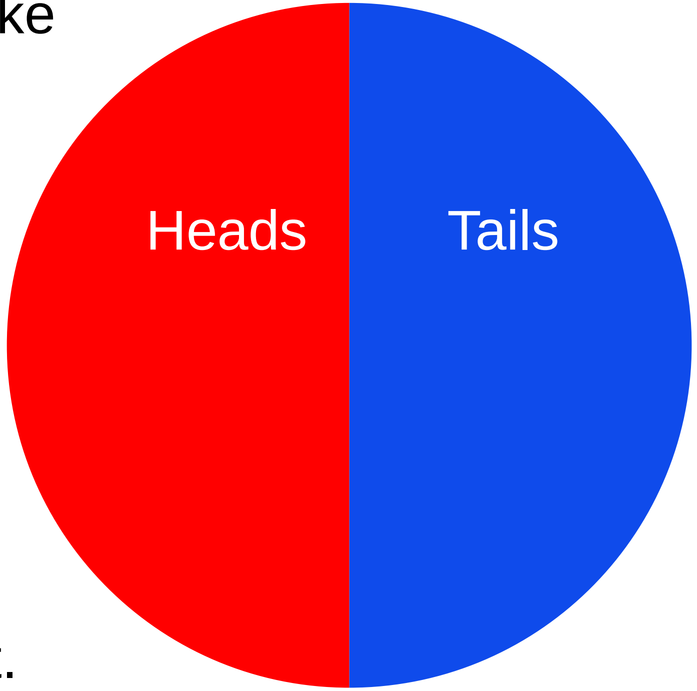
- For example, a (European style) roulette wheel contains the numbers 0-36.
- The probability a single spin of the wheel results in an odd number is therefore $P(Odd) = \frac{|\{1,3,5,7,...,33,35\}|}{|\{0,1,2,3,...,35,36\}|} = \frac{18}{37}$

Probability

- Thinking of probability as a proportion, we can see that it must take a value between 0 (the event never happens) and 1 (it always happens).

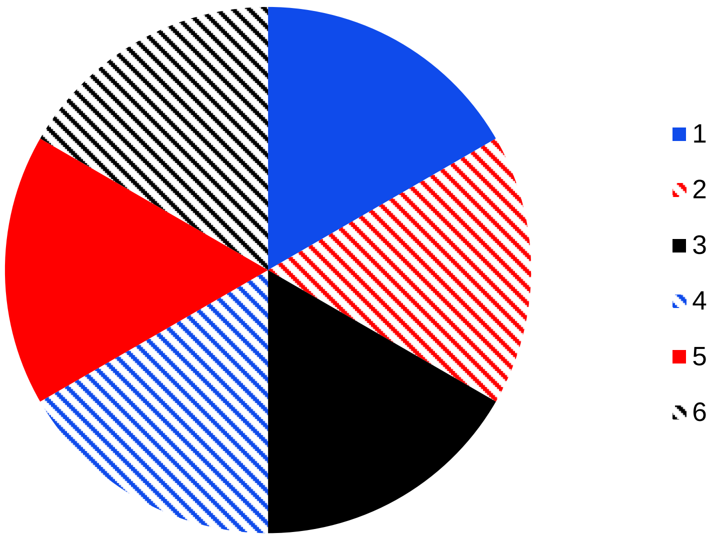
- Probabilities can be expressed as a decimal (e.g. 0.25), a fraction (e.g. 1/4), a percentage (25%) or, less commonly, odds (3 to 1 against).

- One way of picturing probabilities is as proportions on a pie chart. For example, the probabilities of getting Heads and Tails on a single flip of a fair coin are represented by:

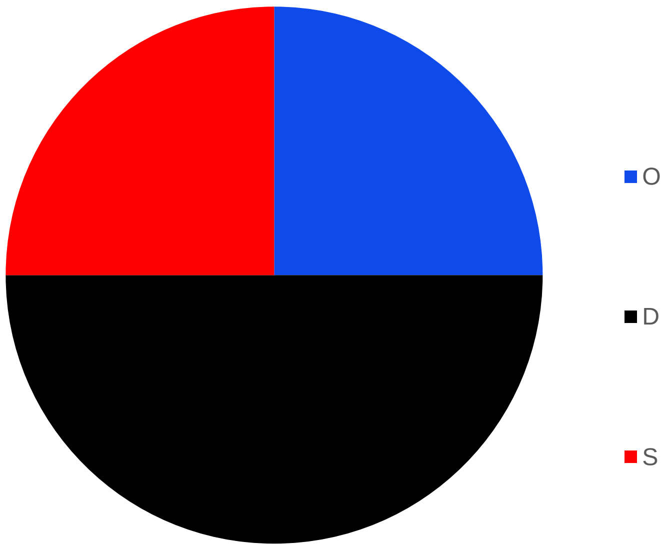


Probability

- What would the following probabilities look like as pie charts?



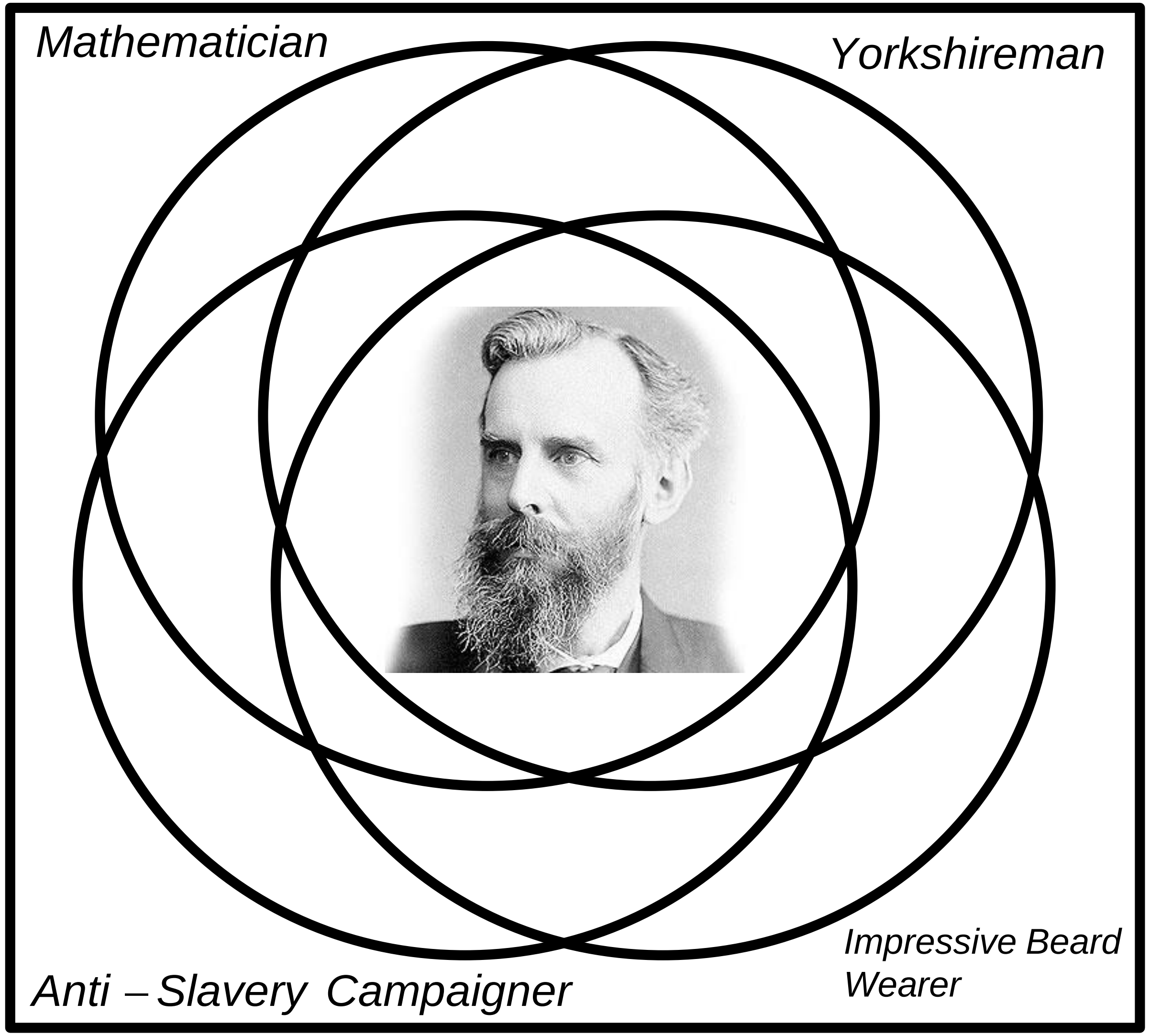
- Each possible outcome when rolling a fair regular die?



- Each possible outcome when selecting one letter from the word ODDS with four letters equally likely to be selected?

Venn Diagrams

- One of the most common ways of representing points which may or may not fall into multiple possible sets is a Venn diagram.
- Sets are represented by (generally overlapping) circles and points placed to lie in all circles corresponding to the sets to which they belong.
- Conceived by John Venn (1834-1923).



Venn Diagrams: Example

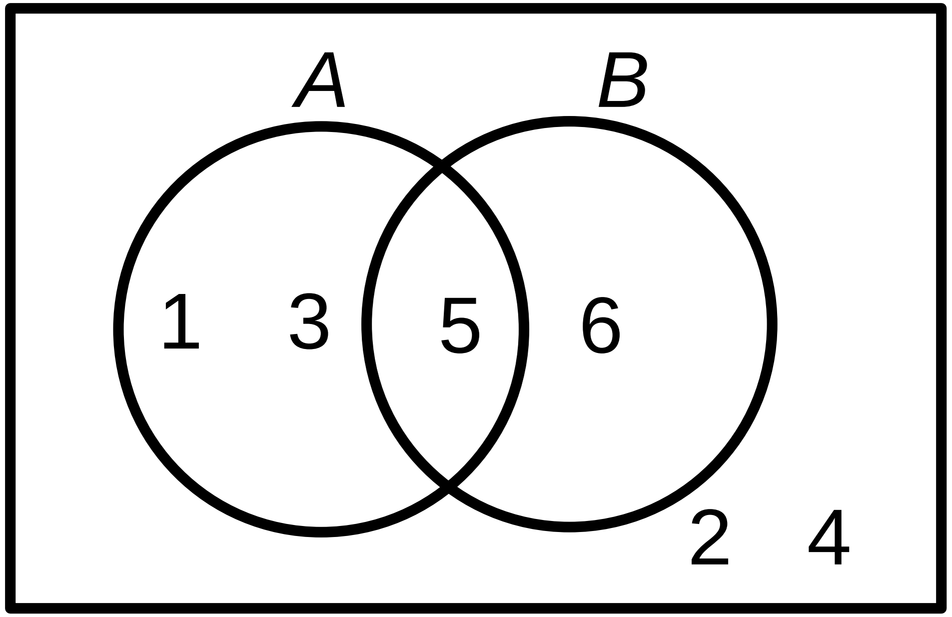
- Consider rolling one die. Let *A* denote the event that the result is an odd number and *B* be the event that the result is greater than four.

- This gives a sample space of $\Omega = \{1,2,3,4,5,6\}$.

- The subsets (events) *A* and *B* are therefore given by $A = \{1,3,5\}$ and $B = \{5,6\}$.

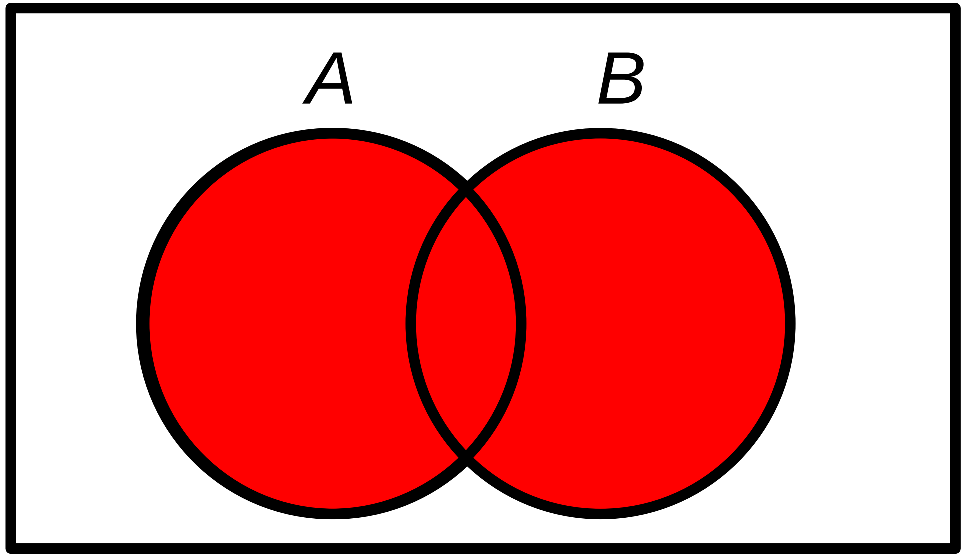
- The outcome of rolling a 5 is therefore in both *A* and *B*.

- The outcomes of rolling a 2 or a 4 are in neither event.

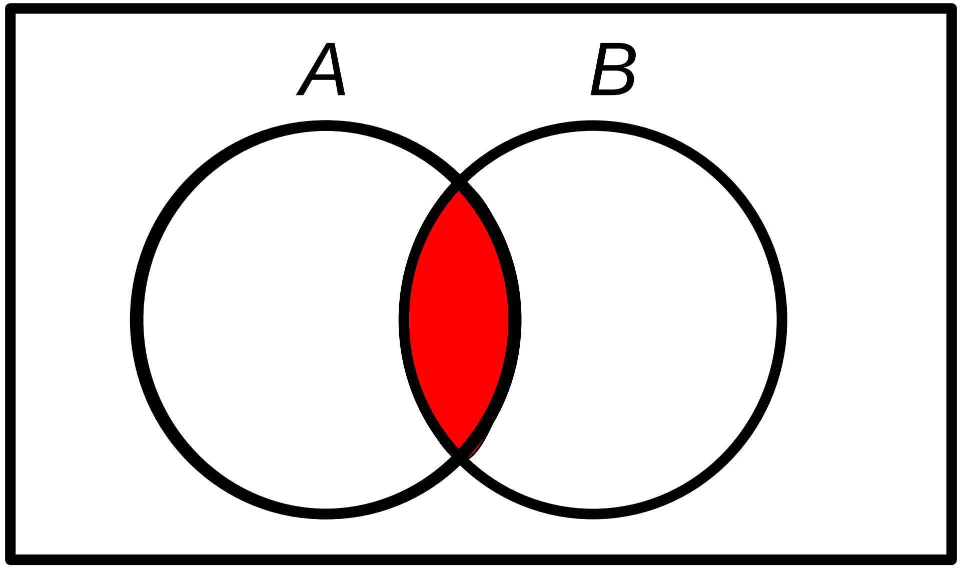


Set Operations

- Since events are subsets of the sample space, we can define the following set operations. Let *A* and *B* be two events in the sample space.



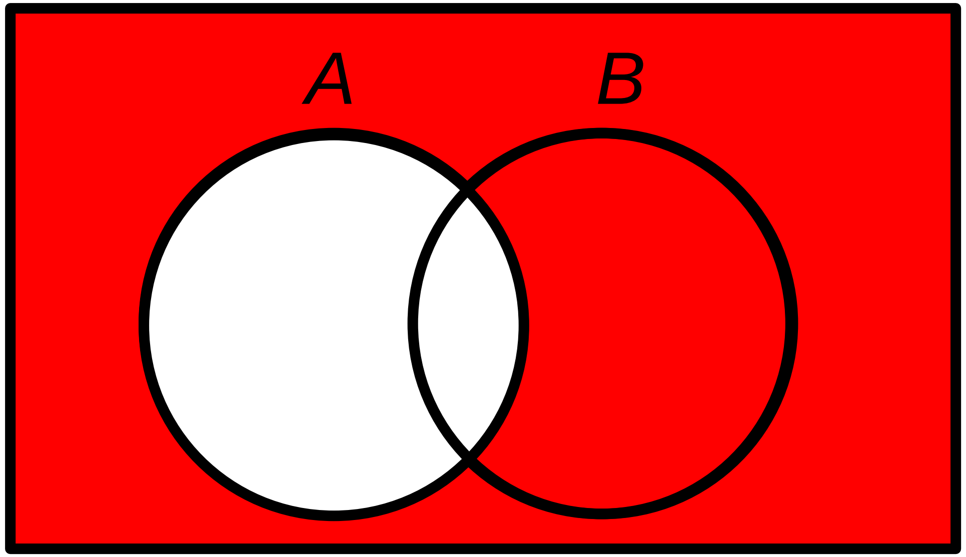
- The **union** of *A* and *B*: $x \in A \cup B$ if and only if $x \in A$ or $x \in B$ (or both).



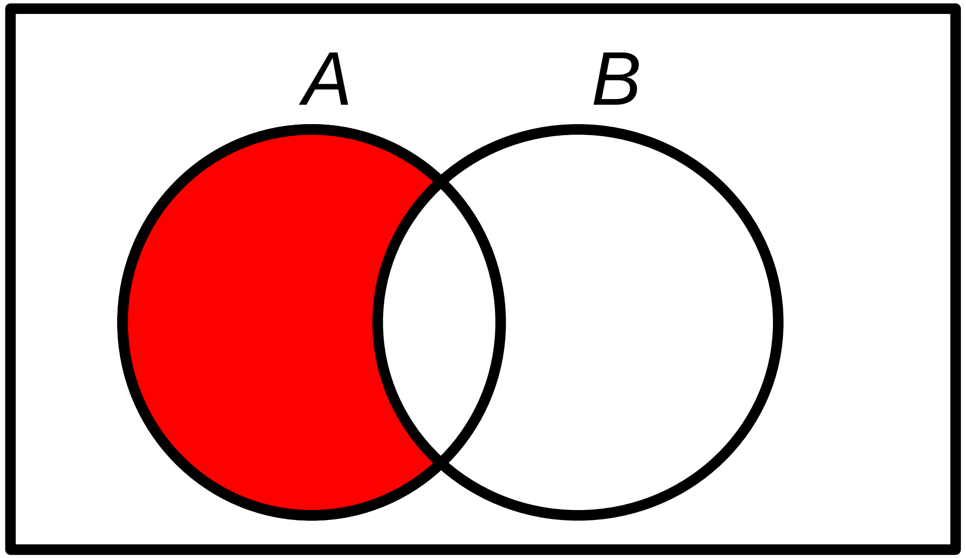
- The **intersection** of *A* and *B*: $x \in A \cap B$ if and only if $x \in A$ and $x \in B$.

Set Operations

- Since events are subsets of the sample space, we can define the following set operations. Let A and B be two events in the sample space.



- The complement of A : $x \in A^c$ if and only if $x \notin A$.



- The difference of A and B: $x \in A - B$ if and only if $x \in A$ and $x \notin B$.

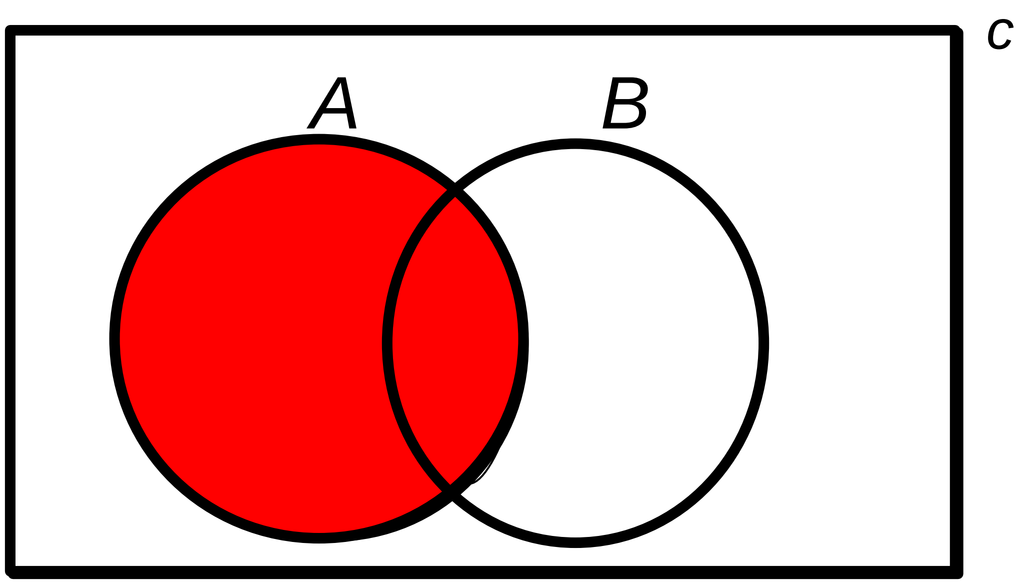
Set Operations

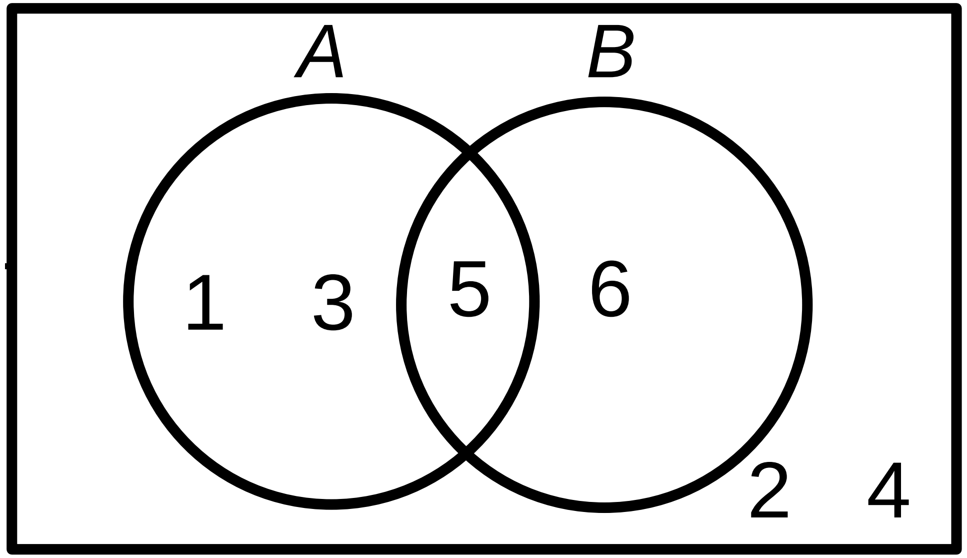
- It is easy to see that $\left(A^c\right)^c = A$

- Consider again the random experiment of rolling one die. Let *A* denote the event that the result is an odd number and *B* be the event that the result is greater than four.

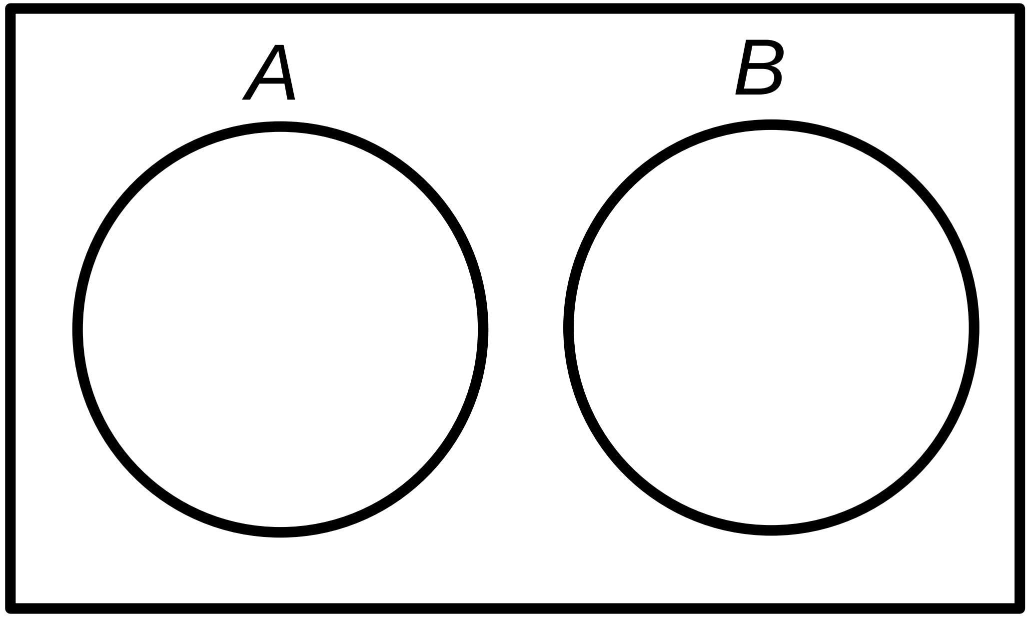
- Using set notation, we have $A = \{1,3,5\}$ and $B = \{5,6\}$.
- We also get $A^c = \{2,4,6\}$ $B^c = \{1,2,3,4\}$

$$A \cup B = \{1,3,5,6\}$$
$$A \cap B = \{5\}$$





Mutual Exclusivity and Partitions



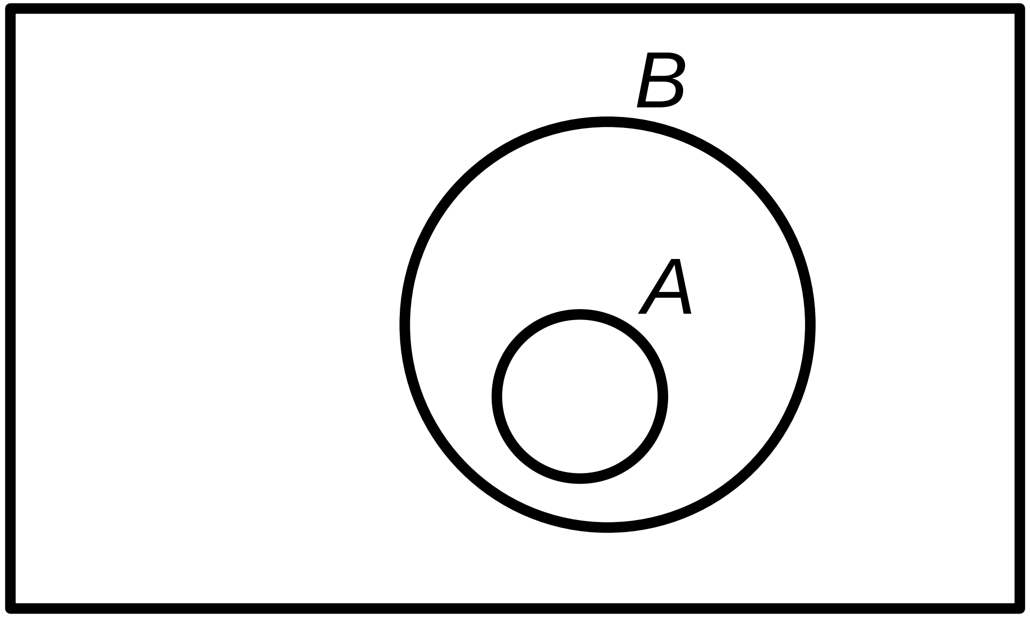
- Two events A and B are **mutually exclusive** if they cannot simultaneously occur i.e. do not overlap, that is if $A \cap B = \varnothing$.

- Two events A and B **partition** a sample space Ω if every possible outcome falls into <u>exactly</u> one of A or B.

- A partition means that $A \cup B = \Omega$ and A and B are mutually exclusive.

- These definitions also extend to more than two non-overlapping sets and to more than two sets forming a partition of the entire sample space.

Subsets

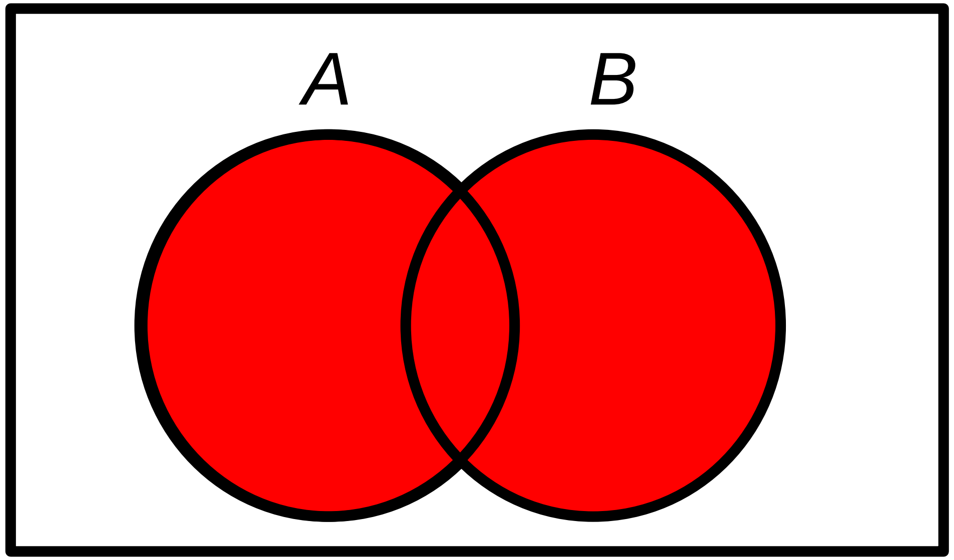


- A set A is a **subset** of set B if and only if every element in A is also in B. This is written $A \subseteq B$.
- This is equivalent to $A \cap B^c = \varnothing$.
- If A is a subset of B, there can still be elements of B which are not in A.
- If, however, all elements in A are in B and all in B are in A i.e. then A and B contain exactly the same elements, so $A = B$.
- For example, the set of humans is a subset of the set of mammals, but the two sets are not equal, since all humans are mammals, but not all mammals are humans.

Basic Probabilistic Calculations

- For any experiment, we are certain to see an outcome from the sample space. That is, $P(\Omega)=1$
- An event cannot happen if it contains no possible outcomes. That is, $P(\varnothing)=0$.
- If two events *A* and *B* are mutually exclusive, they cannot both happen, so $P(A \cap B)=P(\varnothing)=0$
- For any event, *A* and its complement A^c , we have a partition, so $P(A)+P(A^c)=P(\Omega)=1$. This therefore gives $P(A^c)=1-P(A)$. For example, if the probability that *A* occurs is 0.2, the probability that it does not occur is 0.8.

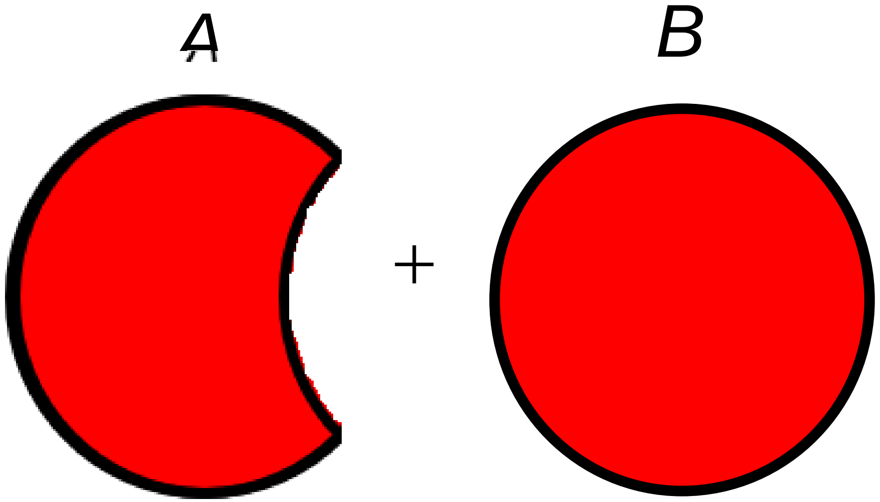
Inclusion-Exclusion Principle



- It is simple to see that

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

- For example, the number of people who own a car or a bike is equal to the number of car owners plus the number of bike owners. People who own both, however, have been counted twice and need to be subtracted once to avoid over-counting.



- This can easily be seen visually through Venn diagrams.

Inclusion-Exclusion Principle

- The inclusion-exclusion principle also extends to 3 (or more) events.
- For example, $P(A \cup B \cup C)$

$$
\begin{aligned}
&= P((A \cup B) \cup C) \\
&= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\
&= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cup B) \cap C)
\end{aligned}
$$

- Now, since a point is in (*A* or *B*) and *C* if and only if it is in (*A* and *C*) or (*B* and *C*).

$$
\begin{aligned}
P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) &= P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C)) \\
&= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)
\end{aligned}
$$

- Combining these gives

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

Inclusion-Exclusion Principle

- Consider the problem of selecting one card at random from a standard deck of 52, assuming all cards equally likely to be chosen.

- Let *HC* be the event that the card is either a Heart or a Club
- Let *R* be the event that the card colour is red
- Let *V* be the event that the card is even numbered

(Take the picture cards – Jack, Queen, King to be neither odd nor even).

- What is the probability that at least one of the three possible events, *HC*, *R* and *V* occurs?

Inclusion-Exclusion Principle

- We have that

$$P(HC) = \frac{26}{52}$$

$$P(R) = \frac{26}{52}$$

$$P(V) = \frac{20}{52}$$

- Working out the pairs, we have that

$$P(HC \cap R) = \frac{13}{52}$$

$$P(HC \cap V) = \frac{10}{52}$$

$$P(R \cap V) = \frac{10}{52}$$

- Similarly, $P(HC \cap R \cap V) = \frac{5}{52}$.

- The inclusion-exclusion principle therefore tells us that

$$P(HC \cup R \cup V) = \frac{26}{52} + \frac{26}{52} + \frac{20}{52} - \frac{13}{52} - \frac{10}{52} - \frac{10}{52} + \frac{5}{52} = \frac{44}{52} = \frac{11}{13}$$