ITCS 121 Statistics

Lecture 4

Probability

Why care about probability?

- Suppose your friend wants to play a game.
- He wants to roll a 6-sided dice. If the result is 6 you win, and he says he will pay you $10.
- If the result is odd he wins, and he wants you to pay him $10.
- Is it a good idea to play the game?
- Most people would say no, because it’s more likely that the result is odd than 6.
- So your friend will probably get your money.

Another game

- Suppose your other friend wants to play a different game.
- She also wants to roll a 6-sided dice. If the result is 6 you win, and she says she will pay you $40.
- If the result is odd she wins, and she wants you to pay her $10.
- Is it a good idea to play the game?
- This is harder.
- It’s more likely the result will be odd than 6.
- But, your friend is offering to pay you more money if you win.
- It turns out this is a good game for you to play, as we will see later, so long as you can afford to pay in case you lose.

What is probability theory for?

- It was to answer questions about gambling games that the first results in what we now call probability theory were proved.

- This was done by Gerolamo Cardano in the 16th century, and later by Pierre de Fermat and Blaise Pascal in the 17th century.

- Some study of probability had also been done by Arab mathematicians much earlier in the medieval period.

What is probability anyway?

- While most people would agree you have more chance of rolling an odd number than rolling a 6, what does that really mean?
- If I say, "If you work hard you will probably pass the 125 statistics course", what does that mean?
- These are not easy questions.

> *Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means.*
>
> Bertrand Russell, 1929

The classical approach - experiments

- The **classical** approach was taken by some of the early pioneers of modern probability theory.
- We define an **experiment** to be some action that can have one or more possible basic results.
- E.g. rolling a 6-sided dice has 6 basic results.

The classical approach – assigning probabilities

- The classical theory says that without any extra information, each of these possible basic results should be equally probable.
- So we can calculate the probability of an event by counting up the number of ways it can happen, and then dividing by the total number of basic results.
- E.g. There are 3 ways we can roll an odd number with a 6-sided dice, so the probability of rolling an odd number is $\frac{3}{6} = 0.5$.

Problems with the classical view

- The classical approach works very well for finite games of chance, which are essentially designed so that the classical assumptions are correct.
- This is maybe one reason why the mathematicians interested in gambling liked it!
- But, not every situation is suitable for the classical assumptions.

Problems with the classical view – basic results have equal probabilities?

- What if my dice is not fair? In other words, what if I have designed it so it is more likely to roll 6 than anything else (e.g. so I can cheat in games)?
- The classical assumptions do not hold, but we would still like to talk about probabilities in dice games with this dice.
- Another problem is that it is not always clear what the ‘basic outcomes’ are in a given situation, and different choices can produce different probabilities for the same event.
  - Google ‘Bertrand’s paradox’ if you are interested in this.

The frequentist approach

- The **frequentist** approach was taken by many mathematicians in the 19th and 20th centuries, including several of the big names in the development of mathematical statistics.
  - E.g. Fisher, Neyman, Pearson.
- In the frequentist approach, if we want to define the probability of an outcome of an experiment, we imagine doing the experiment over and over again.
- We keep recording the proportion of times out of the total so far we have seen the outcome we're interested in.

The frequentist approach - limits

- This proportion will change a lot, but as we repeat the experiment more and more times, the proportion should converge to some limit.
- This limit is the probability of that outcome.
- E.g. I keep rolling my unfair dice and record the proportion of 6s so far after each roll. The limit as the number of rolls tends to infinity is the probability of rolling a 6 on this dice.

Problems with the frequentist view – do limits exist?

- One problem with the frequentist view is that we cannot be sure that this ‘limit as the number of experiments tends to infinity’ exists.
- Even if it does exist, no finite amount of information tells us what it is.
- However, we might still be able to use frequentist ideas to try to guess a probability (like with the unfair dice on the previous slide).

Problems with the frequentist view - limits, really?

- But, to *define* probability surely we shouldn't need to hypothesize the existence of limits to infinite sequences of experiments.
- Another problem comes if we want to ask, e.g. *what is the probability that Julius Caesar enjoyed walks on the beach?*
- Or, *what is the probability that there is life on another planet in our galaxy?*
- How does this even make sense in the frequentist framework?

The Bayesian approach

- The **Bayesian** approach is popular with many modern statisticians and scientists.
- It is named after Bayes' Theorem, which we will see next class.
- The idea is that probabilities reflect degrees of rational belief.
- These are inherently subjective, but to be rational a person's beliefs must fit together according to logical rules.

The Bayesian approach - updating

- In particular, Bayesians say beliefs should be updated in a mathematically precise way when new evidence is uncovered.
- E.g. Maybe I think most people enjoy walking on beaches, so I estimate the probability that Julius Caesar enjoyed walking on beaches at 80%.
- If I later discover that Caesar hated getting sand in his sandals, I would revise that estimate down.
- If I discover he loved the sea I would revise it upwards.

Problems with the Bayesian view - subjectivity

- The main problem with the Bayesian view comes from its subjectivity.
- E.g. I may believe that the probability of rolling a 6 on a fair dice is $\frac{1}{2}$, but *if* I believe that, I'm going to lose a lot of dice games.
- Many people think there must be some kind of objective truth to probabilities.

Problems with the Bayesian view – seeing what you want to see

- If someone's initial beliefs are extreme enough, the mathematical rule for updating beliefs with new evidence makes very little difference.
- So a true Bayesian can, in theory, believe almost anything no matter what the evidence!
- This seems like it could be a problem in science.
- To try and avoid these problems, most Bayesians add extra rationality conditions to initial beliefs.
  - E.g. beliefs about games of chance must agree with classical probabilities.

Probabilities in games of chance – sample spaces

- In games of chance the classical approach works best.
- We imagine an experiment with a finite number of possible outcomes (call these **simple events**).
- We call this set of simple events the **sample space**.
- E.g. rolling a dice gives the sample space $\{1,2,3,4,5,6\}$.
- E.g. Tossing a coin twice gives the sample space $\{HH, HT, TH, TT\}$.

Probabilities in games of chance – events

- We define an **event** to be a subset of the sample space.
- E.g. “rolling an odd number” corresponds to the event $\{1,3,5\}$.
- E.g. “getting at least one tails” corresponds to the event $\{HT, TH, TT\}$.

Probabilities for events in games of chance

- If Ω is a sample space, and if E is an event, we define the probability of E by

$$P(E) = \frac{|E|}{|\Omega|}$$

Size of E

Size of Ω

- Since simple events are essentially events with size 1, the probability of every simple event is $\frac{1}{|\Omega|}$.
- **Probabilities will always be between 0 and 1!**

Probabilities for events in games of chance - counting

- So, in gambling games, calculating probabilities is all about counting.
- E.g. the probability of getting at least one tail when tossing two coins is $\frac{3}{4}$.
- Because, there are 4 possible simple events ($\{HH, HT, TH, TT\}$).
- And 3 of these simple events have at least one tail ($\{HT, TH, TT\}$).

Class activity 1

A jar contains three coins: 1-baht, 5-baht, and 10-baht. Two coins are randomly selected from the jar.

a) List the simple events in S.
b) What is the probability that the selection will contain the 1-baht coin?
c) What is the probability that the total amount drawn will equal 12 baht or more?

Counting rules

- To calculate probabilities in games of chance we need to calculate $|\Omega|$ and $|E|$.

- This can be very hard.

- We will need some techniques for counting.

The *mn*-rule

- Imagine an experiment done in two stages.
- If there are *m* possible outcomes to the first part of the experiment, and for every outcome from the first stage there are *n* possible outcomes in the 2^{nd} stage, then there are *mn* possible outcomes to the experiment.
- E.g. tossing two coins is like a 2-stage experiment. At each stage there are 2 possible outcomes.
- So there are a total of 4 possible outcomes.
- In other words, in this case $|\Omega| = 4$.

The *mn*-rule - generalizing

- This generalizes to experiments with *k* stages.
- I.e. If the number of outcomes at the stages is given by $n_1, n_2, \dots, n_k$, then the total number of outcomes is $n_1 \times n_2 \times \cdots \times n_k$.
- E.g. there are $2 \times 2 \times 2 = 8$ possible results when we toss three coins.

Examples

- If you roll three 6-sided dice, the total number of simple events is $6 \times 6 \times 6 = 216$.
- If you take two balls at random from a bag containing ten balls, the total number of simple events is $10 \times 9 = 90$.
  - Why? Because there are ten possible outcomes for the 1^{st} selection, then 9 for the 2^{nd}.

Counting permutations

- Suppose I have a bag containing *n* objects.
- Suppose I take *k* objects from the bag.
- How many ways can I do this, if the order of selection is important?
- We have a formula for this

$$P_k^n = P(n, k) = \frac{n!}{(n-k)!}$$

Counting permutations – deriving the formula

- There are *n* possibilities for the 1st selection.
- There are then *n-1* objects left in the bag, so *n-1* possibilities for the 2nd selection.
- Then there are *n-2* possibilities for the 3rd selection.
- Etc.
- This continues till we have made all *k* choices.
- So we have $n \times (n-1) \times (n-2) \ldots \times (n-(k-1))$ total possibilities.
- This gives us the formula $P_k^n = P(n,k) = \dfrac{n!}{(n-k)!}$

Example: permutations

- A lock consists of five parts and can be assembled in any order. A quality control engineer wants to test each order for efficiency of assembly. How many orders are there?
  - The formula gives $\frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5!}{1} = 120$.

- Six people are waiting to see the doctor. It is late so the doctor can only see two of them. Since two people don't see the doctor at the same time, how many possible ways can this happen if the order matters?
  - The formula gives $\frac{6!}{(6-2)!} = \frac{6!}{4!} = 6 \times 5 = 30$.

Counting combinations

- Suppose again I have a bag containing *n* objects.
- Suppose I take *k* objects from the bag.
- How many ways can I do this, if the order of selection is *not* important?
- We have a formula for this too

$$C_k^n = C(n,k) = \frac{n!}{(n-k)!\,k!}$$

Counting combinations – deriving the formula

- We can work this out by starting with the formula for counting permutations.
- There are $\frac{n!}{(n-k)!}$ possibilities if order is important, but if order is not important many of these will be equivalent.
- Given a selection of *k* objects, any reordering of these *k* objects is equivalent if order isn't important.
- So to get the combination formula we have to divide by *k!*
- This gives the formula $C_k^n = C(n,k) = \dfrac{n!}{(n-k)!\,k!}$

Example: Combinations

- A box contains six M&Ms, four red and two green. A child selects two M&Ms at random. What is the probability that exactly one is red?
- How many simple events are there?
  - This is the number of selections where order is not important.
  - I.e. $C(6,2) = \frac{6!}{(6-2)!2!} = \frac{6!}{4!2!} = 15.$
- There are $C(4,1) = 4$ ways to select the one red M&M.
- There are $C(2,1) = 2$ ways to select the one green M&M.
- So, using the *mn* rule there are $4 \times 2 = 8$ ways we can select exactly one red M&M.
- So the probability of getting exactly one M&M is $\frac{8}{15}$.

Class activity 2

1. In how many ways can you select five people from a group of eight if the order of selection is important?

2. In how many ways can you select two people from a group of twenty if the order of selection is NOT important?

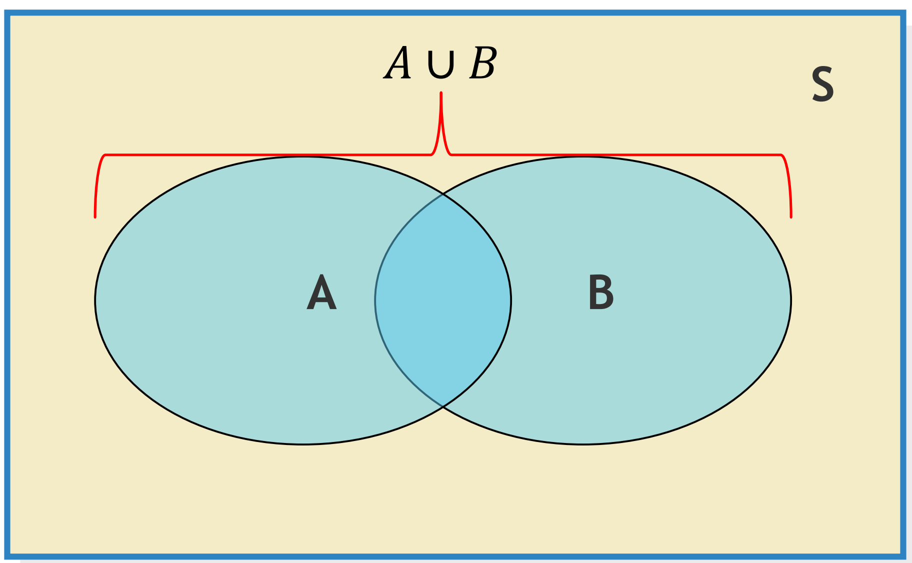
3. Three dice are rolled. How many simple events are in the sample space?

Class activity 3

- For a lottery you choose 6 numbers between 1 and 48 (with no repeats). In the lottery 6 numbers between 1 and 48 are generated at random (no repeats). If exactly 5 of your chosen numbers match any 5 of the numbers generated by the lottery then you win 2nd prize. What is the probability that you win 2nd prize?

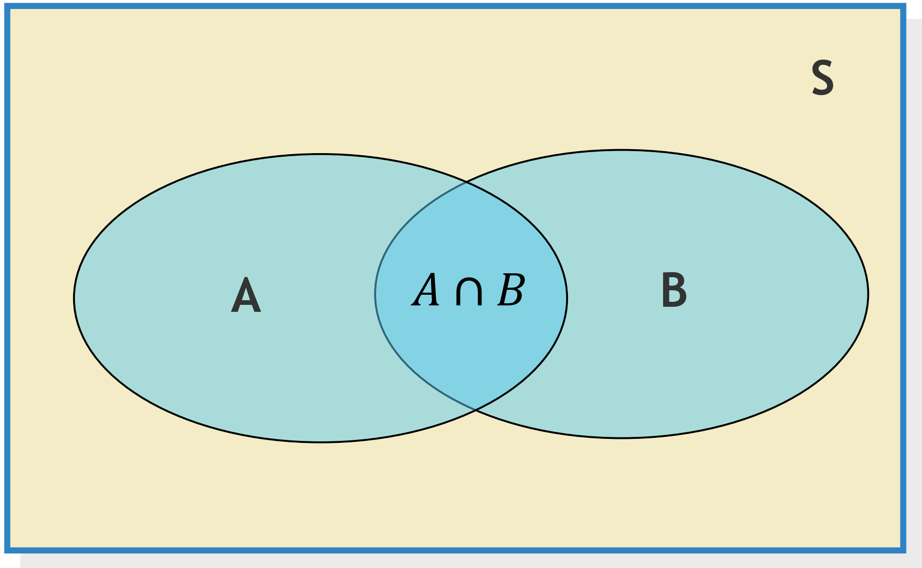
Event relations - unions

- We can combine events to create new events.
- If A and B are events, we can form the **union** of A and B, which we denote $A \cup B$.
- $A \cup B$ is the event that either A or B or both occurs.
- If we think of A and B as sets of simple events this is just the set theoretic union.



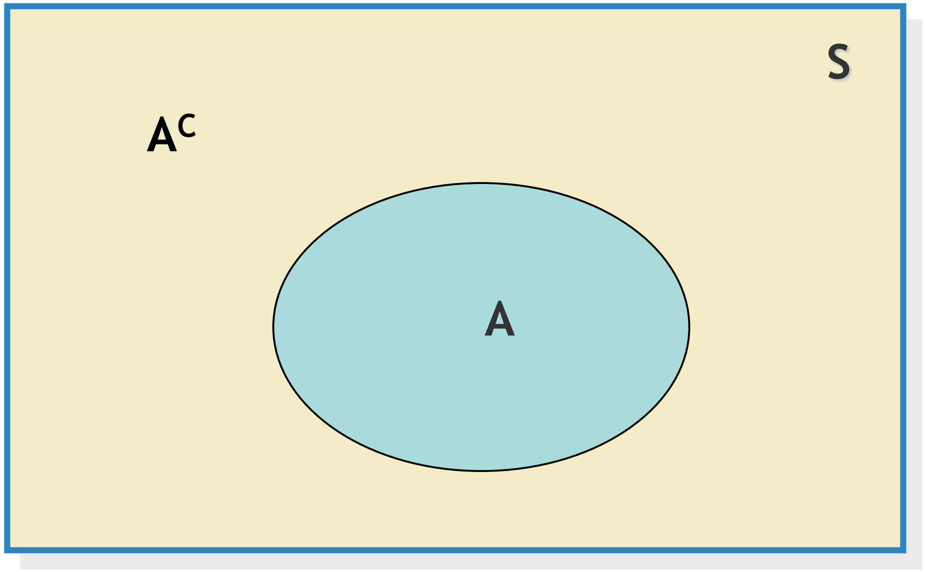
Event relations - intersections

- We can also form the **intersection** of A and B, which we denote $A \cap B$.
- $A \cap B$ is the event that both A and B occur.
- In terms of simple events this is the set theoretic intersection of A and B.
- A and B are **mutually exclusive** if they cannot both be true.
  - I.e. $A \cap B = \emptyset$.



Event relations - complements

- We can form the **complement** of A, which we denote A^c.
- This is the event that A does not occur.
- It's the set theoretic complement of A.



Example – event combinations

- Select a person at random.
  - A is the event this person has brown hair.
  - B is the event this person was born after 1989.
  - C is the event this person was born before 1974.

- $A \cup B$ is the event that the person either has brown hair, or was born after 1989, or both.
- B^c is the event that the person was not born after 1989.
- $B \cap C$ is the event that the person was born after 1989 and was born before 1974. This is not possible, so B and C are mutually exclusive.

Abstract probability

- To apply probability theory beyond simple games of chance where we can do everything with counting arguments, we need to take an abstract view.
- In the abstract, the probability of an event is a number between 0 and 1.
- Events with probability 0 are sometimes considered to be impossible.
- Events with probability 1 are sometimes considered to be certain.

Abstract probability – event relationships

- A probability assignment gives all events we are interested in a number in the range [0,1].

- But, events can be logically related to each other, so we also need these probability values to connect to each other in a sensible way.

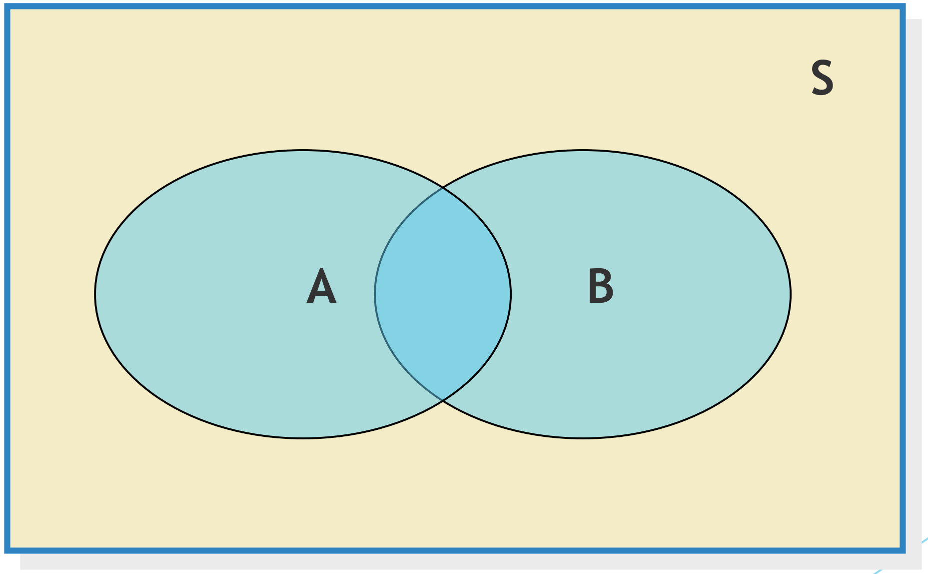
- Also, the sample space itself is an event, and we must have $P(\Omega) = 1$ (something must happen).

- Similarly, we must have $P(\emptyset) = 0$ (it's not possible that nothing happens).

Probability and unions

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

- We can think of probabilities like areas.
- To get the probability of the union of A and B we add the probability of A to the probability of B, but this double counts the probability of $A \cap B$, so we have to make a correction.



Probability and intersections

- If we know $P(A), P(B), P(A \cup B)$, we can rearrange the unions formula to find $P(A \cap B)$.

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

- In special circumstances we can calculate $P(A \cap B)$ using

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

- This only works when A and B are **independent**, in other words, when A being true does not affect the probability that B is true, and vice versa.

Probability and complements

- We require that

$$P(A^c) = 1 - P(A)$$

- This reflects the idea that either A happens or it does not.

Example - Students

- There are 120 students on a course. They can be classified as follows:

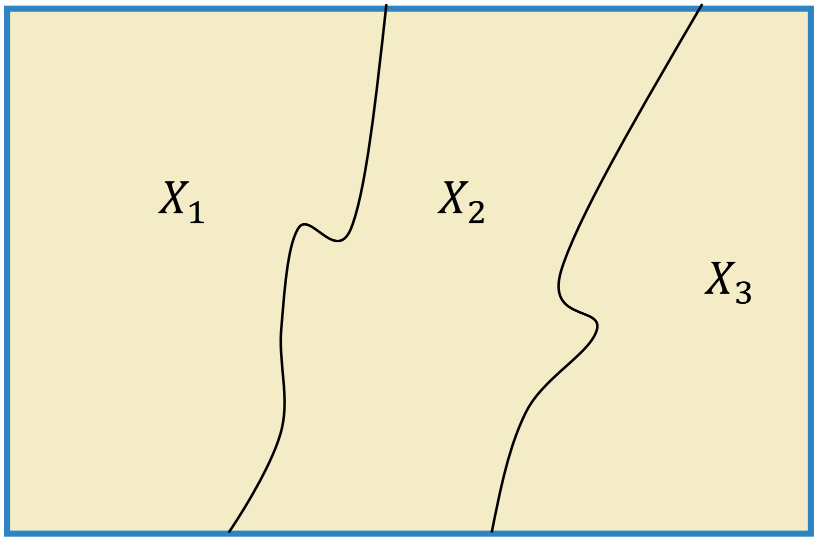
	Brown hair	Not brown hair
Born in 2000 or later	20	40
Born before 2000	30	30

- A is the event 'student has brown hair'.
- B is the event 'student born before 2000'.
- $P(A) = \frac{50}{120} = \frac{5}{12}$, $P(B) = \frac{60}{120} = \frac{1}{2}$.
- $P(A \cap B) = \frac{30}{120} = \frac{1}{4}$.
- $P(A \cup B) = \frac{50+30}{120} = \frac{80}{120} = \frac{3}{4}$.
- Also, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{50+60-30}{120} = \frac{3}{4}$.

equal

Partitions

- A **partition** of a set X is a division of X into two or more mutually exclusive parts $X = X_1 \cup X_2 \ldots \cup X_n$.
- E.g.



Partitioning the probability space

- We say a set of mutually exclusive events $E_1, \dots, E_n$ **partitions the sample space** if their union is equal to the whole sample space.
- In other words, if they cover all possible outcomes.
- If $E_1, \dots, E_n$ partition the sample space then we must have

$$P(E_1 \cup \cdots \cup E_n) = P(E_1) + \cdots + P(E_n) = 1$$

Class activity 4

We partition the sample space into 5 mutually exclusive events E_1, E_2, E_3, E_4, E_5 with these probabilities:

$$P(E_1) = P(E_2) = 0.15,\ P(E_3) = 0.4,\ P(E_4) = 2\ P(E_5)$$

a) Find the probabilities for events E_4 and E_5.

b) Find the probabilities for these two events:

$$A = E_1 \cup E_3 \cup E_4,\ B = E_2 \cup E_3$$

Conditional probability

- Let A and B be events.
- Often, the probabilities of A and B are related.
- If A is true, then this often makes it more or less likely that B is true.
- We use $P(A|B)$ to denote the probability of A given B. In other words, the probability that A is true assuming that B is true.
- For example, the probability that I will sleep well is affected by whether my neighbor is having a party.

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

Independence again

- Remember that if A and B are independent then $P(A \cap B) = P(A)P(B)$.
- This is because, to say that A and B are independent is to say that $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
- So, when A and B are independent, from the formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$ we get

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

- It can be fairly easily proved that if $P(A|B) = P(A)$ then we must also have $P(B|A) = P(B)$, and vice versa.
- In other words, A is independent from B if and only if B is independent from A.

Example – heart attacks

In a certain population, 10% of the people can be classified as being high risk for a heart attack. Two people are randomly selected from this population. What is the probability that exactly one of the two are high risk?

Define H = high risk, define N = not high risk

$$P(exactly\ one\ high\ risk) = P((H_1 \cap N_2) \cup (H_2 \cap N_1))$$

Using rule for mutually exclusive events

$$= P(H_1 \cap N_2) + P(H_2 \cap N_1)$$

$$= P(H_1)P(N_2) + P(H_2)P(N_1)$$

Using rule for independent events

$$= (0.1)(0.9) + (0.1)(0.9)$$

$$= 0.18$$

Example – more heart attacks

Suppose we have additional information in the previous example. We know that only 49% of the population are female. Also, 8% of women are high risk. A single person is selected at random. What is the probability that it is a high risk female?

Define H: high risk F: female

$$P(F) = 0.49 \qquad P(H|F) = 0.08$$

$$P(H \cap F) = P(H|F)P(F) = (0.08)(0.49) = 0.0392$$

Class activity 5

Suppose that $P(A) = 0.4$ and $P(A \cap B) = 0.12$

a) Find $P(B|A)$
b) Are events A and B mutually exclusive?
c) If $P(B) = 0.3$, are events A and B independent?

Class activity 6

Suppose the probability that a drive lasts at least k years before it fails is given by

$$P(L \geq k) = e^{-k}$$

So e.g. the probability that a drive fails in the third year is given by e^{-2}.

What is the probability that a drive fails in its 3rd year, assuming that it has survived two years already?