Lecture 10
Probability

COS10003 Computer Logic and Essentials (Hawthorn)

Semester 1 2021

Today

1. Basics
2. Events
3. Conditional
4. Expected value
5. Distributions

The basics of probability

Working with conditional events

Solving problems using distributions

Why statistics and probability?

- Probability is the study of random non-deterministic experiments.
- Statistics is the collection, organisation, analysis, interpretation and presentation of data.
- These are important for computer scientists:
  - to understand randomised algorithms and to analyse algorithms
  - to work with large amounts of data
  - to calculate chances of failure or reliability measures
  - etc. etc.

Where has the stats gone?

We did previously do some stats in this unit, but it was very basic (means and medians etc.)
It is assumed that you know how to calculate mean, median and standard deviation.
We will focus on probability.

What is probability?

Probability theory is a mathematical model of random events that assigns the limiting values of the relative frequencies, or probabilities to all possible outcomes.
Probability is relevant to logical reasoning because it provides a framework for making decisions and choosing between alternatives in situations of uncertainty.

Sample spaces

- A set S or Ω, the sample space, defines all the possible outcomes of an experiment.
- A particular outcome is an element of S.
- The sum of the probabilities of each outcome is equal to 1. Probabilities cannot be negative.

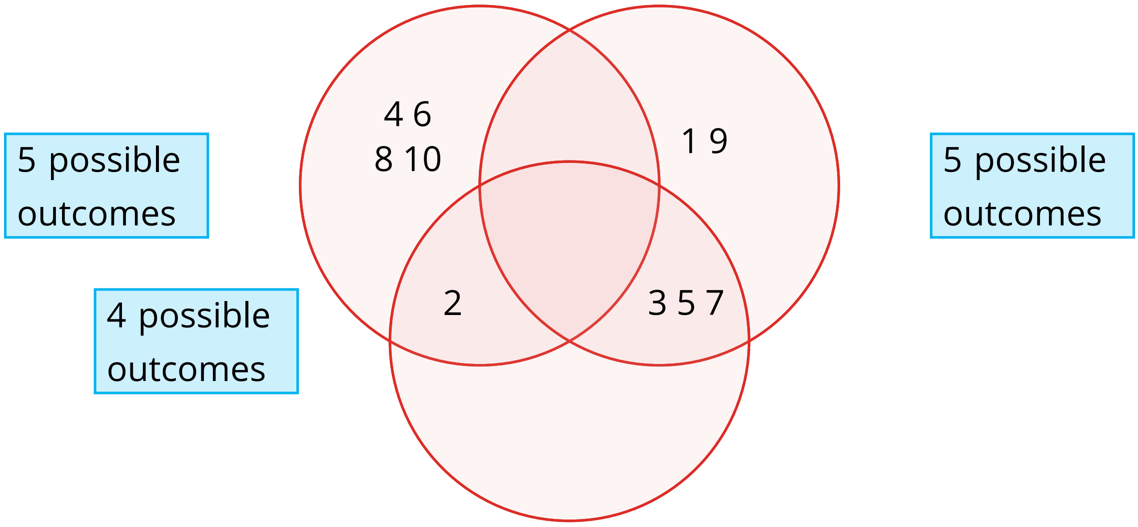
Events

- An event A is a set of outcomes where $A \subseteq S$. This could be a single item, or many items.
- The empty set $\varnothing$ is sometimes referred to as the impossible event.
- Events can be combined to form new events using the operations of sets because events are sets.
  - $A \cup B$ is the event that occurs whenever A, B or both A and B occur.
  - $A \cap B$ is the event that occurs whenever both A and B occur.
  - A^C is the event that occurs whenever A does not occur.
  - If $A \cap B = \varnothing$, events A and B are mutually exclusive.

Example

The numbers 1 to 10 can be drawn in a lottery. The sample space consists of the 10 possible numbers: $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, a total of 10 possible events.

- A is the event than an even number occurs
- B is the event that an odd number occurs
- C is the event that a prime number occurs



P(A) = 5/10 = 50% , P(B) = 5/10 = 50% , P(C) = 4/10 = 40%

For you to do

What is the probability of the following if one number is drawn?

- $P(A \cup C)$ is the event that an even or prime number occurs?
- $P(B \cap C)$ is the event that an odd prime occurs?
- $P(C^C)$ is the event that a prime number does not occur?
- Describe an event that is impossible.

Finite Probability Spaces

- S is some finite sample space S = $\{a_1, a_2, ..., a_n\}$.
- A finite probability space is obtained by assigning each sample a_i (which is an element of S) a real number p_i called the probability of a_i such that $1 \geq p_i \geq 0$ and the sum of all the p_i values ($p_1, p2, ..., p_n$) is equal to 1.
- The probability of an event A, denoted P(A) or p(A), is the sum of the probabilities of all the samples in A.

Equi-probable Spaces

Commonly, the situation arises that each sample in a finite sample spaces has an equal probability of occurring. Such a probability space is called an equi-probable space. In such cases, we can define:

$$P(A) = \frac{\text{number of ways that event A can occur}}{\text{number of ways that the sample space S can occur}}$$

Equi-probable spaces are typically associated with the words “at random”.

Equi-probable Spaces

Consider a card selected at random from a deck of 52 cards. A denotes the card is a club, B denotes the card is a jack, queen or king.
The fact that we are dealing with a equi-probable space means that:

- $P(A)$ = number of clubs / number of cards = 13/52 = 1/4
- $P(B)$ = number of jack, queen and king (face) cards / number of cards = 12/52 = 3/13
- $P(A \cap B)$ = number of club face cards / number of cards = 3/52

Axioms of Finite Probability Spaces

The axioms of finite probability spaces are:

- **Non-negativity:** for every event A, $P(A) \geq 0$
- **Normalization:** P(S) = 1
- **Finite additivity:** If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$ and $P(A \cap B) = 0$.

These axioms ensure that the probability of an event is the sum of the elementary events composing them.

Properties of Finite Probability Spaces

From these axioms it is possible to prove :

- $P(\varnothing) = 0$
- $P(A^C) = 1 – P(A)$
- $A \subseteq B$ implies $P(A) \leq P(B)$
- $P(A) < 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

The last result says that if A and B are not independent events, we must subtract the probability arising from the dependency. It can be generalized for any number of events.

Independence

Two events A_1 and A_2 are independent of each other if the following condition is satisfied:
$P(A_1 \cap A_2) = P(A_1)P(A_2)$
If this condition is not satisfied, the events are said to be dependent. This is useful in determining the dependency of events.

Example

The outcome of tossing a coin two times is:

$S = \{HH, HT, TH, TT\}$

Consider the following events:

A = first toss is heads, B = second toss is heads, C = both tosses are heads

A and B are independent of each other: $P(A \cap B)$ = 1/4 = $P(A)P(B)$ = 1/2 $\times$ 1/2

C is dependent on A and B both occurring; checking dependence of A and C:

$P(A \cap C) = 1/4 \neq P(A)P(C) = 1/2 \times 1/4$

Conditional probability

We are often interested in the probability of an event occurring given a set of circumstances.

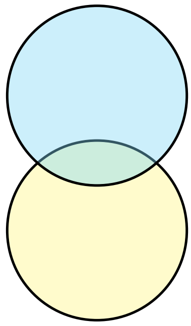
Let E be an event in our sample space that has a non-zero probability of occurring.

The probability that another event A occurs once E has occurred $P(A \mid E)$, i.e., the conditional probability of A given E is obtained from:

$$P(A \mid E) = \frac{P(A \cap E)}{P(E)}$$

Conditional Probability

This formula indicates that $P(A \mid E)$ depends on the amount of overlap between A and E. If we are dealing with an equi-probable space, then we can interpret this as
the number of ways that A and E can occur
divided by
the number of ways that E can occur.



$$P(A \mid E) = \frac{P(A \cap E)}{P(E)}$$

Conditional Probability

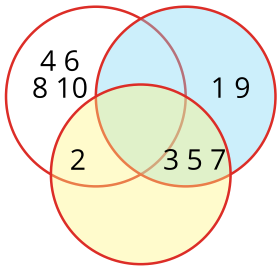
Recall our lottery from earlier. The probability of an even number was P(A) = 5/10 = 1/2.
What is the probability of a prime number given an even number?

$$P(C \mid A) = \frac{P(C \cap A)}{P(A)} = \frac{\frac{1}{10}}{\frac{1}{2}} = \frac{1}{10}\frac{2}{1} = \frac{2}{10} = \frac{1}{5}$$

Compare this with the probability of drawing 2 (= 1/10).

For you to do

How about the probability of a prime number given an odd number?



Basics Events Conditional Expected value Distributions Next

Conditional probability again

An alternative way of writing the conditional probability is

$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1)$

Or more generally:

$P(A_1 \cap A2... \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)...P(A_n \mid A_1 \cap A_2... \cap A_{n-1})$

This equation can be simply interpreted as the product of successive probabilities, each of which accounts for previous outcomes.

It is in fact an application of the multiplication principle.

Example

There are 12 diamonds, 4 of which are flawed. If three diamonds are picked randomly, what is the probability that all three are flawless?

$$
\begin{aligned}
P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \\
&= \frac{8}{12}\frac{7}{11}\frac{6}{10} \\
&= \frac{14}{55}
\end{aligned}
$$

Lecture 10
Probability

COS10003 Computer Logic and Essentials (Hawthorn)

SWINBURNE UNIVERSITY OF TECHNOLOGY

Semester 1 2021

Expectations: a game

Choose one game and one game only to play

Game	Reward	Odds of Winning	Expectation
Game #1	\$100	50%	
Game #2	\$1000	10%	
Game #3	\$10000	2%	

Chance to win $\neq$ Expectation

Great Expectations!

Often there are several possible net outcomes associated with a probability. In view of this what, on average, is the expected value E?

$E = m_1p_1 + m_2p_2 + m_3p_3 + ... + m_np_n$

That is, E is the probability weighted value of the possible outcomes, where m_n is a value and p_n is the probability of that value occurring.

Example

There are 1500 tickets in a lottery sold at \$1 each, with a first prize of \$100 and a second prize of \$50. If you buy a ticket, then what is your expectation (that is, how much money on average are you likely to get)?

Example

$$\begin{aligned} E &= \overset{\text{value of winning}}{(\$100-\$1)} \times 1/1500 + \overset{\text{value of second prize}}{(\$50-\$1)} \times 1/1500 + \overset{\text{value of not being 1st or 2nd}}{(\$0-\$1)} \times 1498/1500 \\ &= \$99/1500 + \$49/1500 + -\$1 \times 1498/1500 \\ &= -\$0.90 \end{aligned}$$

For you to do

How about if there were 100 tickets in the lottery sold at $2 each, with a first prize of $100 and a second prize of $50.

If you buy a ticket, then what is your expectation (that is, how much money on average are you likely to get)?

Fair Expectations?

A fair expectation would be if $E = 0$, that is in the long run you neither win or lose.

What would be a fair price (x) for the original lottery tickets?

$$
\begin{aligned}
E = 0 &= \frac{\$100 - x}{1500} + \frac{\$50 - x}{1500} + (-x)\frac{1498}{1500} \\
&= \$100 - x + \$50 - x - 1498x \\
&= \$150 - 1500x \\
x &= \$0.10
\end{aligned}
$$

Two outcomes

We have already encountered repeated trials in the previous examples of the lecture.
However, there is a special type of repeated trial that is characterised by only two outcomes, typically described as "failure" and "success."
Such situations give rise to various probability distributions.

Binomial or Bernoulli Distribution

If p is the probability of success, $q = 1-p$ is the probability of failure. The probability of k successes in n repeated trials is given by

$$P(X = k; n, p) = \binom{n}{k} p^k q^{n-k}$$

where X is the random variable corresponding to X = "number of times a successful outcome occurs".
The first term in the right hand side of this equation should be familiar to you.
It is the binomial coefficient described in the lecture on combinatorial analysis.

Example

A coin is tossed 6 times and heads is designated as a success.
We have $n = 6$, $q = p = \frac{1}{2}$.
Using the above formula we can determine various possibilities:

- The probability of exactly two heads ($k = 2$) is $P(2; 6, 1/2) = C(6, 2) \times 0.5^2 \times 0.5^4 = 15/64 = 0.234$.
- The probability of at least four heads (k = 4, 5 or 6) is $P(4; 6, 1/2) + P(5; 6, 1/2) + P(6; 6, 1/2) = 15/64 + 6/64 + 1/64 = 22/64 = 11/32 = 0.344$.
- The probability of no heads (k = 0) is $P(0; 6, 1/2) = (1/2)^6 = 1/64 = 0.016$.

Binomial or Bernoulli Distribution

The Bernoulli distribution is often used in real world decision-making strategies.
Airlines deliberately overbook seats, safe in the knowledge that past experience has shown that there will always be “no show” passengers. Let us say that an aircraft has a capacity of 150. The airline has a policy of selling 160 tickets to protect against no-show passengers based on past experience that the probability of a passenger being a no show is 0.1. What is the probability that at least one passenger will have to be ‘bumped’ from the flight?

Binomial or Bernoulli Distribution

This problem can be treated as 160 Bernoulli trials with a success rate of 0.9. Using the above formula we calculate

$$\begin{aligned} P(X > 150) =& P(X = 151) + P(X = 152) + ... + P(X = 160) \\ =& P(151; 160, 0.9) + P(152; 160, 0.9) + ... + P(160; 160, 0.9) \\ =& \binom{160}{151} 0.9^{151} 0.1^{9} + \binom{160}{152} 0.9^{152} 0.1^{8} + ... + \binom{160}{160} 0.9^{160} 0.1^{0} \\ =& 0.0186 + 0.0099 + ... + 0.0000 \\ =& 0.0359 \end{aligned}$$

This means than the risk of bumping a passenger is only 3.59%.

Poisson Distribution

Processes that have a small probability of success are described by the Poisson distribution, which in the limiting value of the Binomial distribution.

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

where $\lambda = np$ is the expected value and e = 2.71828... (Euler's number)
The usefulness of the Poisson distribution is that it is not necessary to know the number of trials (n) or the probability of success (p). We only need to know the product of these values ($\lambda = np$).
The Poisson distribution is particularly helpful for rare events (n is relatively large and p is relatively small).

Poisson Distribution: example

In a given year, Melbourne expects to experience 8 severe thunderstorms.
Last year, it experienced 12.
Is this evidence of climate change?
As severe thunderstorms or other natural disasters are rare events, it is legitimate to use the Poisson distribution. Using $\lambda = 8$ and k = 12, we find

$$P(X = 12) = e^{-8}\frac{8^{12}}{12!} = 0.0481$$

Another example: Poisson distribution

Consider the example of number of customers browsing in a shop. The owner observes that 20 customers per hour enter the shop. What is the probability that there will be some customers in the shop during a particular quarter hour?
In this case λ = 20 per hour or 5 per quarter hour, such that $P(X = k) = e^{-5}\frac{5^k}{k!}$
The statement “there will be some customers” implies that there will be one or more customers. Therefore
$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-5}\frac{5^0}{0!} = 1 - e^{-5} = 0.9933$
so there is a 99.33% chance of a customer in any 15-minute period.

Another example: Poisson distribution

How about if the owner observes that 3 customers per hour enter the shop. What is the probability that there will be no customers in the shop during a particular 20 minute period?

Hypergeometric Distribution

Another common process involving two possible outcomes is the so-called urn problem which is an example of sampling without replacement involving only two possibilities.
Consider an urn containing either red (R) or white (W) balls. The task is to randomly select n balls out of the urn without replacing any. We wish to determine the probability that r balls of the selected n balls will be red?

Hypergeometric Distribution

The probability is given by the hypergeometric distribution:

$$P(X = r) = \frac{C(R,r)C(W,n-r)}{C(R+W,n)}$$

where $C(R,r)C(W,n-r)$: the number of ways to pick r red balls from R+W balls
$C(R+W,n)$: the number of ways to pick n balls from R+W balls
Notice that all of the contributing terms to the hypergeometric distribution are binomial coefficients.

Hypergeometric Distribution: example

A so-called “baker’s dozen” consists of 13 (not 12 items) to account for a faulty item. What is the probability of picking the bad egg from a baker’s dozen after 6 attempts?
This is clearly a hypergeometric distribution because the pool (the 13 eggs) is not replaced and we can identify two types of eggs, namely good eggs (W = 12) and bad eggs (R = 1).
Applying the above formula with r = 1 and n = 6, we obtain
$P(X = 1) = \frac{C(1,1)C(12,5)}{C(13,6)} = \frac{792}{1716} = 0.4615$

For you to do

From a pack of cards, we consider clubs only. We draw 3 cards without replacement. What is the probability that 2 are face cards?

Basics Events Conditional Expected value Distributions Next

Geometric Distribution

What about counting the number of trials needed to achieve success? For example, the number of coin flips needed to see a head.
The geometric distribution is used for this: given the probability of success p, and X is the number of trials needed:

$$P(X = v) = (1 - p)^{v-1}p \text{ (get v-1 failures before one success)}$$

where v = $\{1, 2, 3, 4...\}$
The expected value is $E(X) = 1/p$.

Geometric distribution: example

There is a 20% chance of a COS10003 student attending the previous lecture. If I email students one at a time to ask them a question about the class, what is the chance of finding someone with the 5th email?

Seeking $P(X = 5)$, $p = 0.2$, $1 - p = 0.8$

$P(X = 5) = (0.8^4)(0.2) = 0.082$

So there is just over 8% chance that the 5th person emailed attended the previous lecture.

Geometric distribution: example

There is a 20% chance of a COS10003 student attending the previous lecture. If I email students one at a time to ask them a question about the class, what is the chance of finding someone within three emails?

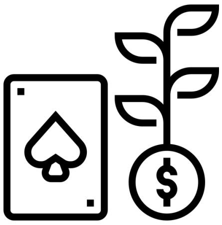
Seeking $P(X \leq 3), p = 0.2, 1 - p = 0.8$

$P(X = 1) + P(X = 2) + P(X = 3) = 0.2 + 0.16 + 0.128 = 0.488$

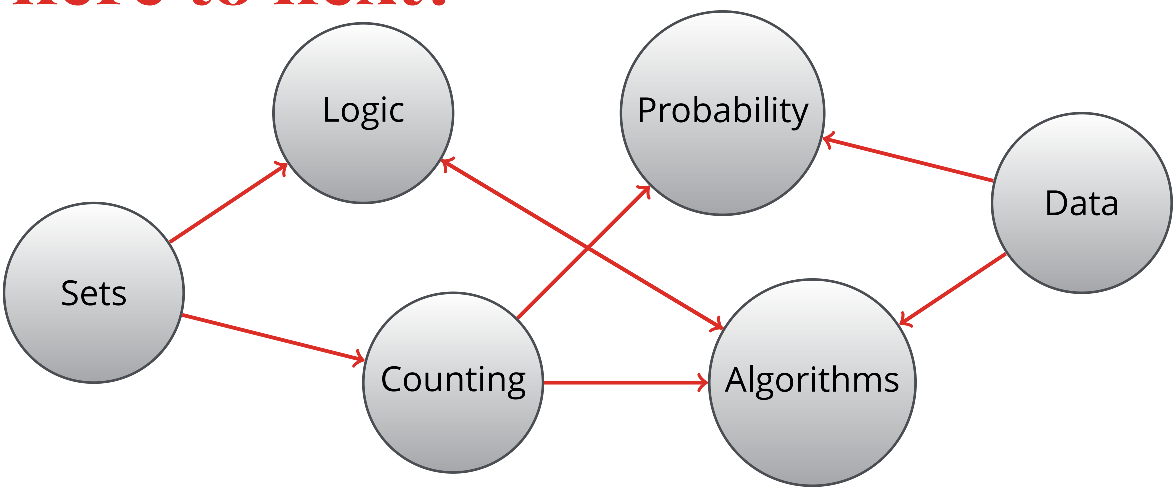
So there is just under 49% chance that it will take less than three emails to find someone who attended the previous lecture.

Reflecting

- What are the basic principles of probability?
- How do we work with conditional probability?
- How is expected value calculated?
- What are the characteristics of each probability distribution?



Where to next?



In which we look at applications of what we've learned.

Lecture 10
Probability

COS10003 Computer Logic and Essentials (Hawthorn)

Semester 1 2021

Basics Events Conditional Expected value Distributions Next

Questions I still have

Topics I need to review