Part 3 Section 1: Probability Trees

Aaron Robotham

Libraries needed for this chapter: magicaxis, DiagrammeR.

```r
library(magicaxis, quietly=TRUE)
library(DiagrammeR, quietly=TRUE)
```

So everybody running this chapter gets the same outputs, we should first set the seed.

```r
set.seed(1)
```

A Roll of the Dice

Everybody is familiar with dice, so that is always an intuitive way to start a statistics course. Let's throw an **R** die once.

```r
sample(x=1:6, size=1)
```

```
## [1] 1
```

The input x represents the number of sides on our die *and* the labels. We could also roll a six-sided alphabetical die.

```r
sample(x=c('a','b','c','d','e','f'), size=1)
```

```
## [1] "d"
```

Or given **R** already knows about letters we can just select from the vector provided.

```r
sample(x=letters[1:6], size=1)
```

```
## [1] "a"
```

It should be intuitive to anybody playing a board game that, unless something is wrong with the manufacture, the chances of rolling and particular number 1–6 is equal. We can check this by running the **sample** function with replacement.

```r
sample(x=1:6, size=100, replace=TRUE)
```

```
##   [1] 2 5 3 6 2 3 3 1 5 5 2 6 6 2 1 5 5 1 1 6 5 5 2 2 6 1 4 1 4 3 6 2 2 6 4 4 4
##  [38] 2 4 1 6 1 4 1 6 2 3 2 6 6 2 5 2 6 6 6 1 3 3 6 4 6 3 1 4 5 1 1 6 4 5 5 4 6
##  [75] 5 4 4 1 5 5 6 1 1 3 6 2 2 3 6 2 4 3 5 2 2 1 3 3 2 2
```

To make the output easier to process (and to have more samples) we can assign it to a new object.

```r
roll1000 = sample(x=1:6, size=1000, replace=TRUE)
```
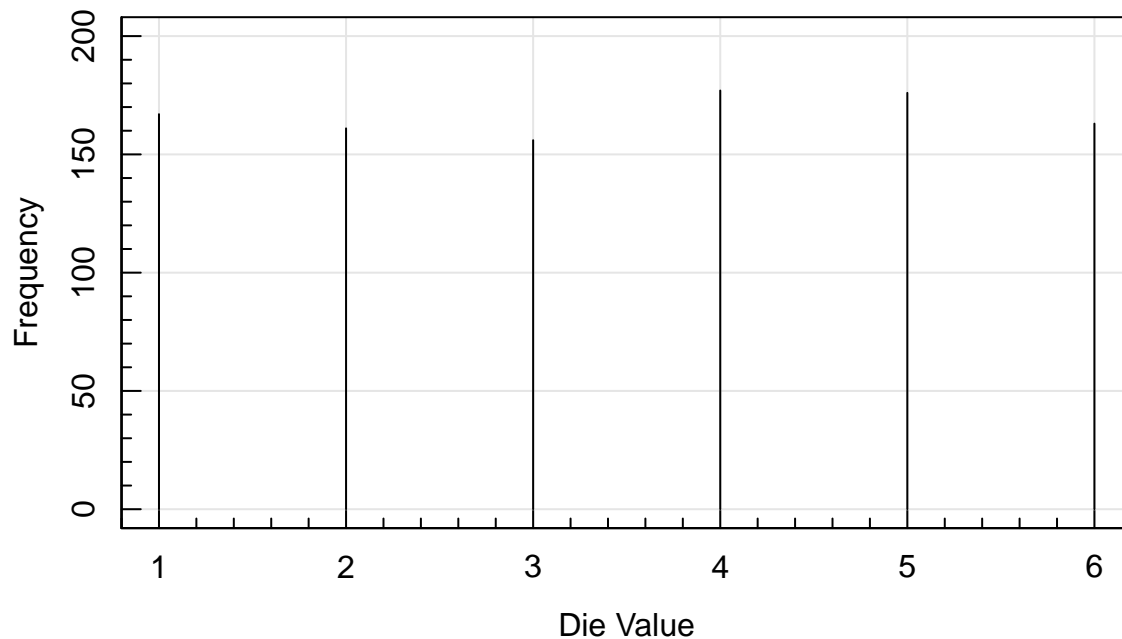
It is then possible to table the output:

```r
table(roll1000)
```

```
## roll1000
##   1   2   3   4   5   6
## 167 161 156 177 176 163
```

Or plot a histogram:

```r
magplot(table(roll1000), ylim=c(0,200), xlab='Die Value', ylab='Frequency')
```

In this case there is a visibly equal chance of rolling any one of six virtual die numbers, i.e. the chance of rolling any one number is 1/6, or more formally $P(x = 1 : 6) = 1/6$.

So what about rolling two dice? Clearly these should be seen as independent events, so the chance of rolling 2 (1 and 1) is simply $1/6 \times 1/6 = 1/36$. But what about 3 or 4? There are 2 ways to make 3 (1/2 and 2/1) and 3 ways to make 4 (1/3, 3/1 and 2/2). **R** has a useful utility function to expand out possible combinations of such events called **expand.grid**. You will find yourself using this for all sorts of reasons, but this is a good example.

```
die2 = expand.grid(1:6,1:6)
die2
```

```
##    Var1 Var2
## 1     1    1
## 2     2    1
## 3     3    1
## 4     4    1
## 5     5    1
## 6     6    1
## 7     1    2
## 8     2    2
## 9     3    2
## 10    4    2
## 11    5    2
## 12    6    2
## 13    1    3
## 14    2    3
## 15    3    3
## 16    4    3
## 17    5    3
## 18    6    3
## 19    1    4
## 20    2    4
## 21    3    4
## 22    4    4
## 23    5    4
## 24    6    4
## 25    1    5
## 26    2    5
```

```
## 27     3     5
## 28     4     5
## 29     5     5
## 30     6     5
## 31     1     6
## 32     2     6
## 33     3     6
## 34     4     6
## 35     5     6
## 36     6     6
```

This shows all the combinations we can expect when throwing 2 die (36 in all). We can find the sums easily.

```
die2sum = rowSums(die2)
die2sum
```
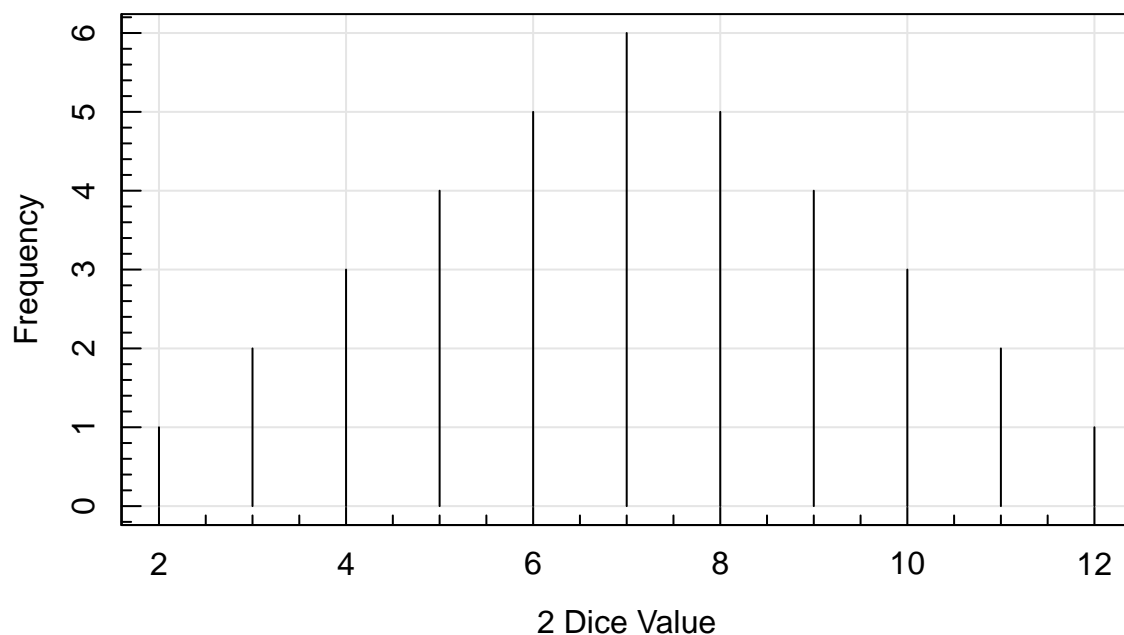
```
## [1]   2   3   4   5   6   7   3   4   5   6   7   8   4   5   6   7   8   9   5   6   7   8   9 10   6
## [26]   7   8   9 10 11   7   8   9 10 11 12
```

```
table(die2sum)
```

```
## die2sum
##  2  3  4  5  6  7  8  9 10 11 12
##  1  2  3  4  5  6  5  4  3  2  1
```
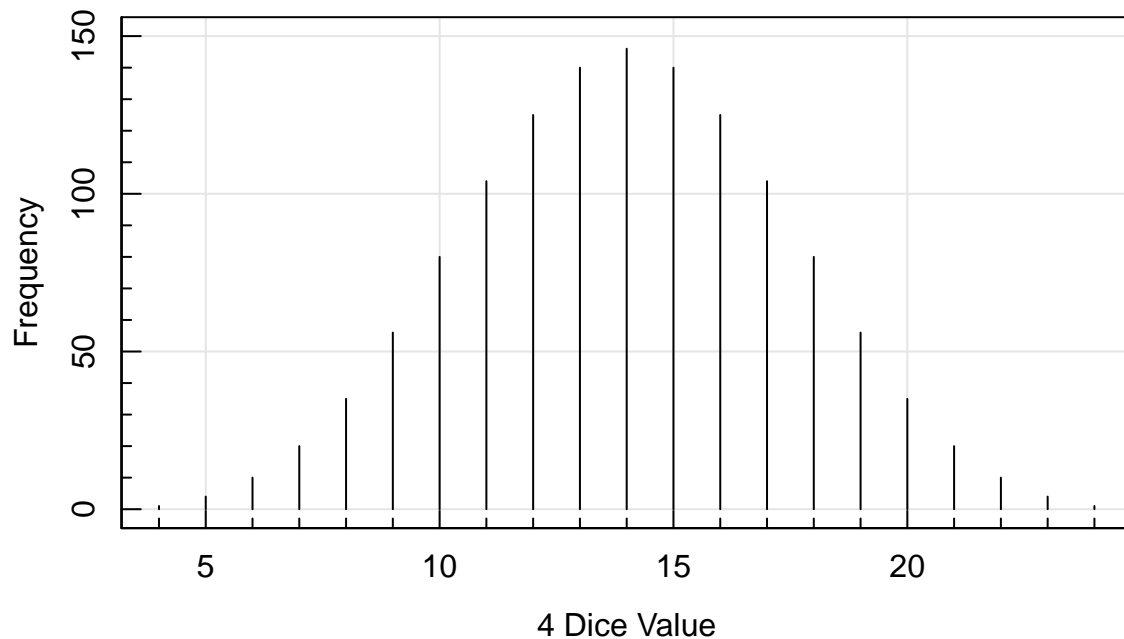
It is useful to plot this to see how it compares to single die event.

```
magplot(table(die2sum), ylim=c(0,6), xlab='2 Dice Value', ylab='Frequency')
```


This distribution is clearly not Uniform anymore. We can emphasise this by rolling 4 dice instead.

```
magplot(table(rowSums(expand.grid(1:6,1:6,1:6,1:6))), ylim=c(0,150), xlab='4 Dice Value',
        ylab='Frequency')
```
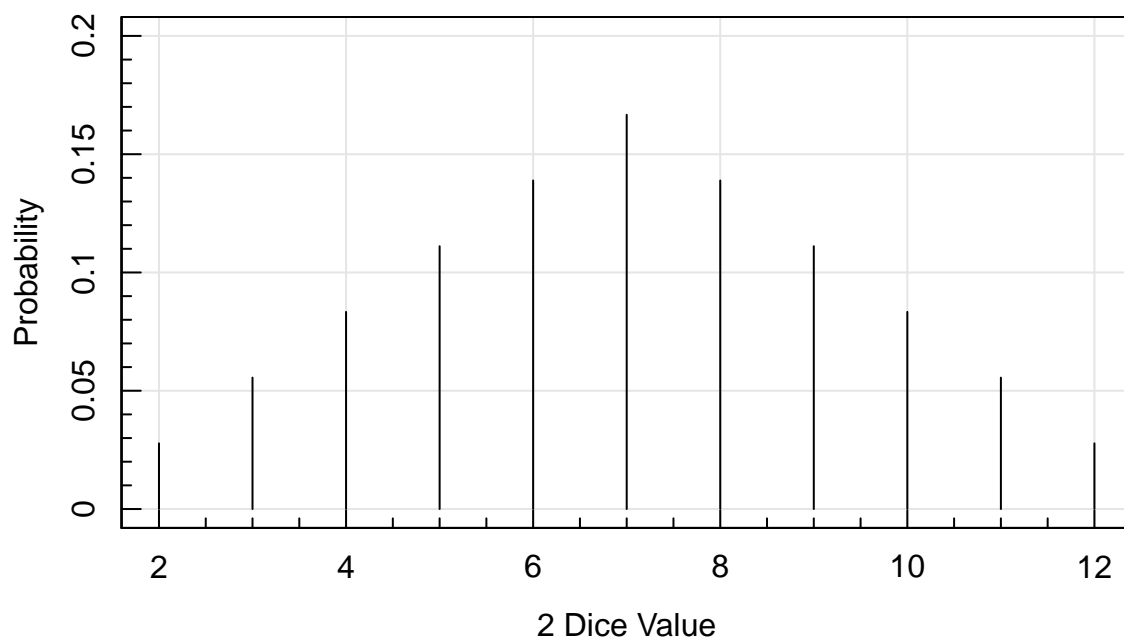
Going back to our 2 dice data, we can convert this to a true probability density function, where the y-axis values should represent probability of the event and the total sum should be 1, by re-normalising.

```
die2PDF = table(die2sum)/length(die2sum)
```

We can plot this like above.

```
magplot(die2PDF, ylim=c(0,0.2), xlab='2 Dice Value', ylab='Probability', type='h')
```


This plot allows us to quickly read off certain outcomes, e.g. the chance of 2 dice summing to exactly 5 is a little over 0.1.


Giving a Toss

Whilst coin tossing might seem a more simple problem to solve than dice rolling, it opens up an avenue to explore one of the most important distributions in statistics: the binomial distribution (more on that later).

Similar to the examples above, we can ask a number of questions regarding tossing a fair coin. A very simple example is what is the chance of a tossed coin coming up heads (H) or tails (T). In both cases it is simply 1/2, or $P(H) = P(T) = 1/2$. A more complicated question, which you should have a sense for from our dice examples above, is how likely is it for a coin tossed twice to come up two heads, or one head and one tail in any order?

Much like the dice example, where there are more ways to roll 7 (6 combinations) with 2 dice than 12 (1 combination), there are more ways to generate 1 head and 1 tail (2: HT or TH) than 2 heads (1: HH) or 2 tails (1: TT). Since the coin tosses are independent events the chance of any specific two coin combination is $1/2 \times 1/2 = 1/4$, but since there are two ways to get a head and tail outcome the probability is then $1/4 + 1/4 = 1/2$. This structure of sequential events is commonly called a probability tree. It is helpful to visualise this probability tree, which can be done with the **DiagrammeR** package and **mermaid** diagram function[1]. Here we will show the both the tree and the tree generating code, but later we will only show the tree (keen students should try to recreate the tree themselves).

```
mermaid("
  graph LR
    Start --> H[H, P=0.5]
    Start --> T[T, P=0.5]
    H --> HH[H, P=0.5]
    H --> HT[T, P=0.5]
    T --> TH[H, P=0.5]
    T --> TT[T, P=0.5]
    HH --> HHout[HH, P=0.25]
    HT --> HTout[HT, P=0.25]
    TH --> THout[TH, P=0.25]
    TT --> TTout[TT, P=0.25]
")
```
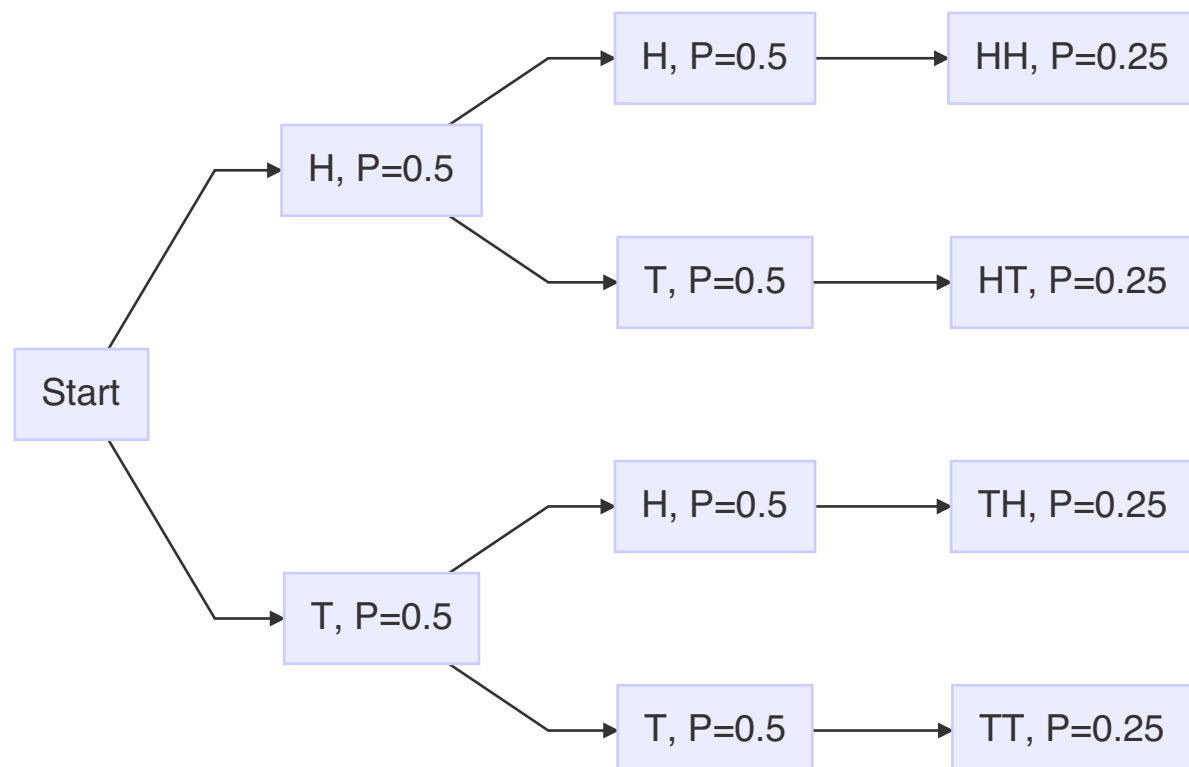

These graphs rapidly get quite complex given the large number of potential outcomes, e.g. for 3 tosses.

---

[1]To display these trees in PDF format you will need to install the **webshot** package and run `webshot::install_phantomjs()`.
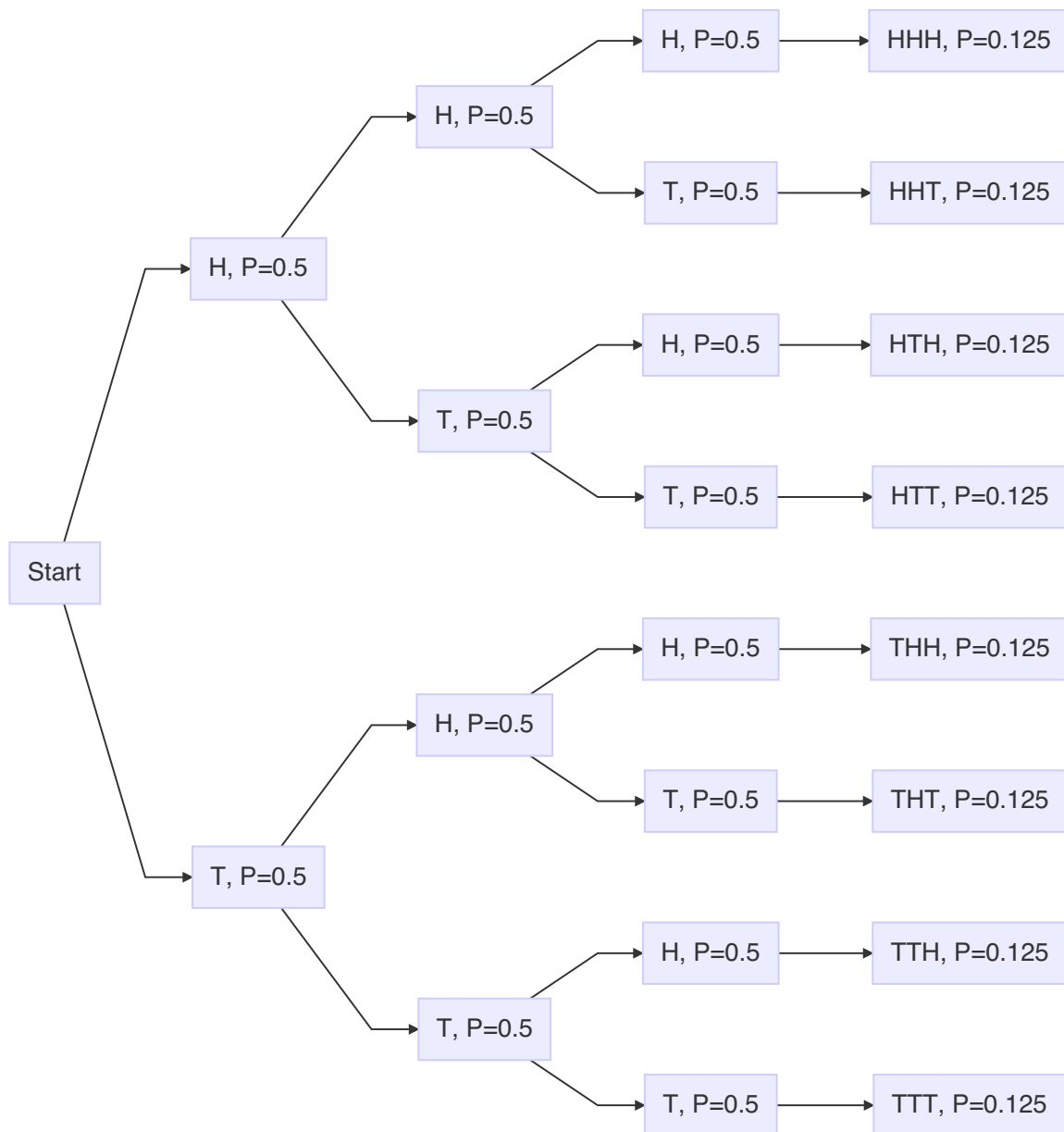
Using the above we can find out the chance of getting two heads and one tail (in any order) in 3 coin tosses: $1/8 \times 3 = 3/8$. Doing this took a bit of manually checking on our part though.

To take things a step further, consider what happens if the coin we are tossing is biased: heads come up with probability P=0.6 and tails with probability P=0.4. This is more complicated since the 8 possible outcomes are not all trivially the same (1/8).

```
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  H, P=0.6    │───────▶│  HHH, P=0.216    │
                                          └──────────────┘        └──────────────────┘
                          ┌──────────────┐
                          │  H, P=0.6    │
                          └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  T, P=0.4    │───────▶│  HHT, P=0.144    │
                                          └──────────────┘        └──────────────────┘
        ┌──────────────┐
        │  H, P=0.6    │
        └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  H, P=0.6    │───────▶│  HTH, P=0.144    │
                                          └──────────────┘        └──────────────────┘
                          ┌──────────────┐
                          │  T, P=0.4    │
                          └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  T, P=0.4    │───────▶│  HTT, P=0.096    │
                                          └──────────────┘        └──────────────────┘
┌──────────┐
│  Start   │
└──────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  H, P=0.6    │───────▶│  THH, P=0.144    │
                                          └──────────────┘        └──────────────────┘
                          ┌──────────────┐
                          │  H, P=0.6    │
                          └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  T, P=0.4    │───────▶│  THT, P=0.096    │
                                          └──────────────┘        └──────────────────┘
        ┌──────────────┐
        │  T, P=0.4    │
        └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  H, P=0.6    │───────▶│  TTH, P=0.096    │
                                          └──────────────┘        └──────────────────┘
                          ┌──────────────┐
                          │  T, P=0.4    │
                          └──────────────┘
                                          ┌──────────────┐        ┌──────────────────┐
                                          │  T, P=0.4    │───────▶│  TTT, P=0.064    │
                                          └──────────────┘        └──────────────────┘
```
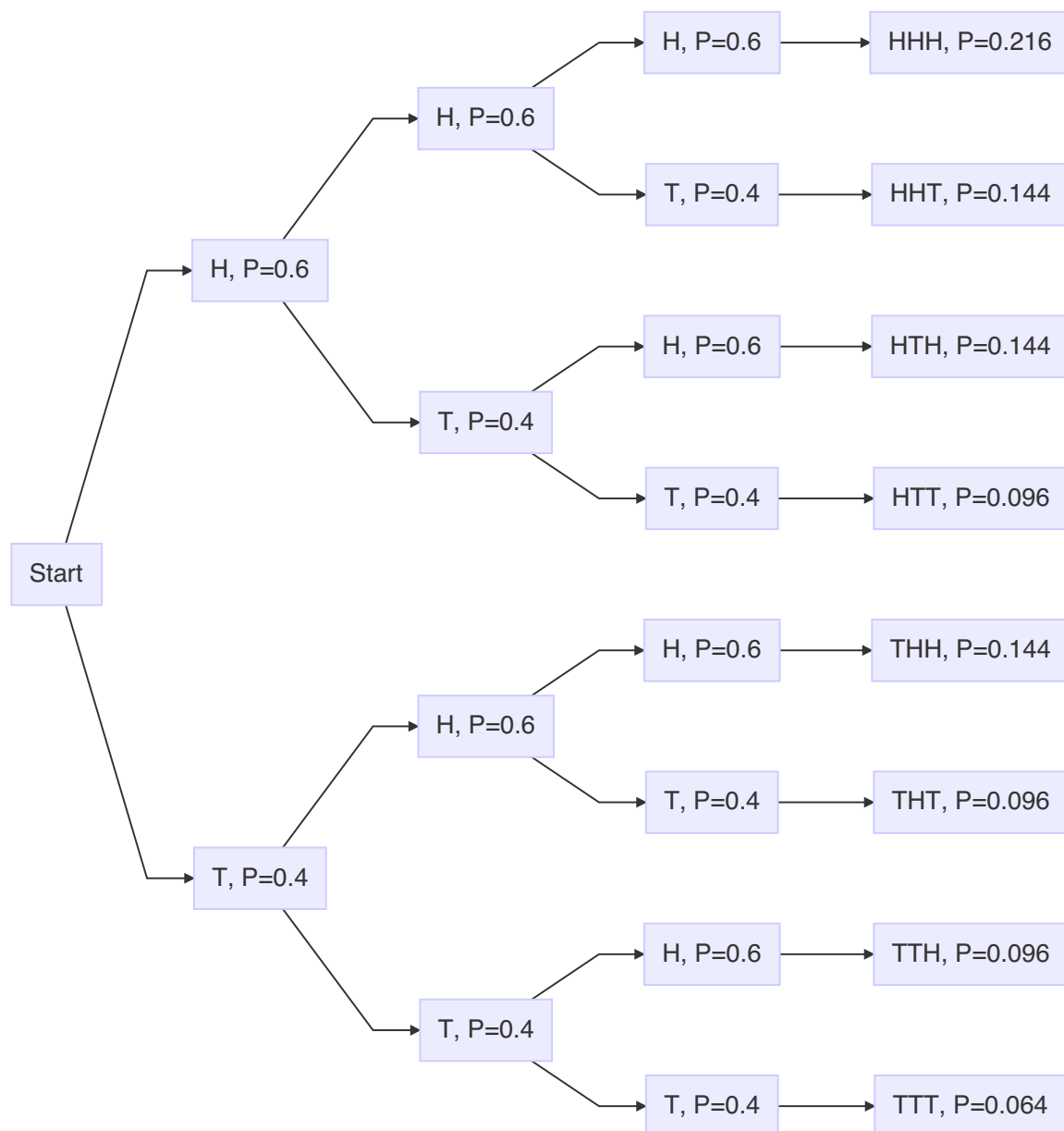
From summing the outcomes, we can see that the chance of getting two heads and one tail (in any order) is $0.144 \times 3 = 0.432$. But this approach is clearly unwieldy for, say, 8 tosses of a coin with biased outcomes. This is where the binomial distribution comes in: it explicitly computes the likelihood of different combinations of outcomes where the chance of success is any probability between 0 and 1. E.g. to recompute the chance of 2 heads out of 3 tosses with the biased coin we can compute the following using the built in **dbinom** function (where x is the number of trial successes, $size$ is the number of trials and $prob$ is the probability of a trial success).

```
dbinom(x=2, size=3, prob=0.6)
```

```
## [1] 0.432
```

We can compute the chance another way (where success is now defined as getting a tail on a toss).

```
dbinom(x=1, size=3, prob=0.4)
```

```
## [1] 0.432
```

We can now trivially answer an 'unwieldy' example of the chance of 3 heads from 8 tosses.

```
dbinom(x=3, size=8, prob=0.6)
```

```
## [1] 0.123863
```

Or 10 heads from 40 tosses where the chance of getting a head is 0.3.

```
dbinom(x=10, size=40, prob=0.3)
```

```
## [1] 0.1128173
```

The 'binary' nature of the binomial distribution is encoded in the name: it only allows for two outcomes from a given event, effectively success of failure. This means whilst it can be used for examples like the above, it is not flexible to different probabilities after each toss (e.g. the coin gets fairer with time), or a different number of outcomes for each event (6, in the case of rolling a standard die).

Conditional Probability

Sequences of independent events are very important in statistics (like coin tossing or dice rolling), but so are sequences of conditional events. Conditional events may have extreme asymmetries. E.g., I will give you $10 if your favourite team wins and you toss a head with a fair coin, or $20 if your favourite team loses and you roll 1 or 2 with a fair die. These are conditional events, and their conditional nature clearly affects the expected financial outcome we can expect. If the chance of our the team winning is 1/5 (they suck) then we can compute our expected financial outcome. A reason to do this is if I say that you will have to pay $5 to take my bet and play the game- should you pay this to play?
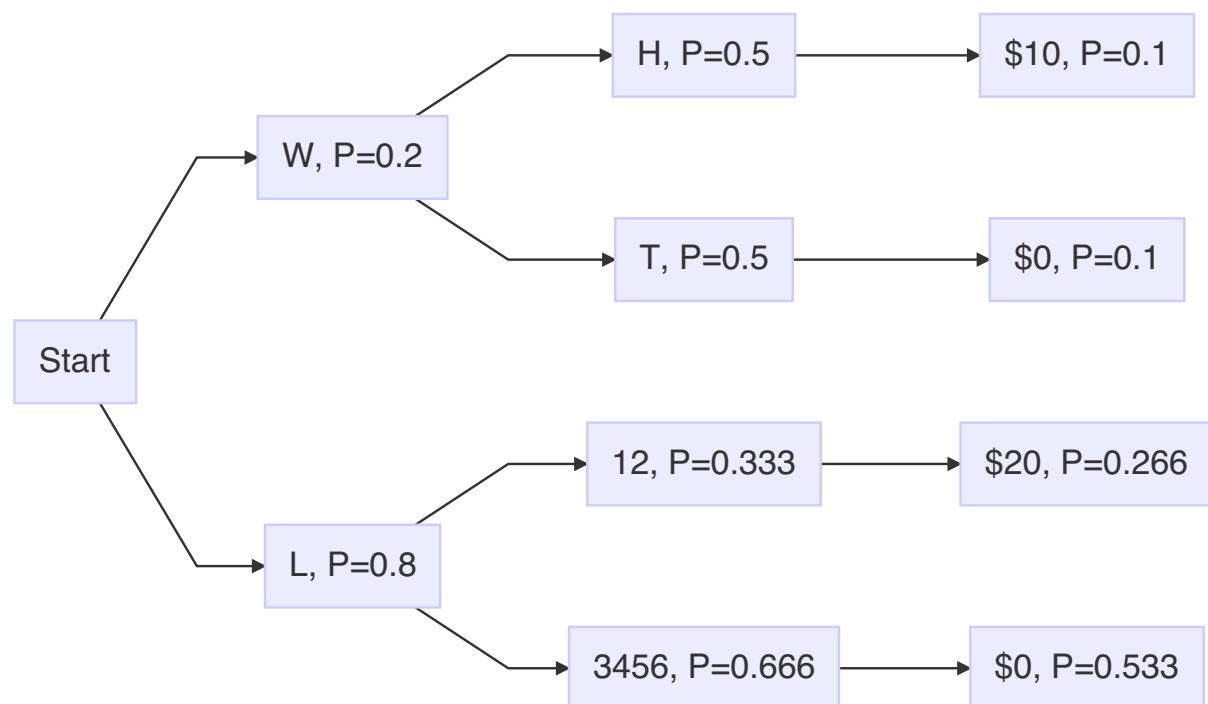
From the above we can then compute the expected return via $10 \times 0.1 + 20 \times 0.266 = 6.32$. So the expected return from taking the bet is $6.32, which is more than the $5 it will cost to play, i.e. we should play the game (if gambling is your thing, of course). This is the case even though we will only win something $0.1 + 0.266 = 0.366$ of the time.

Conditional Sampling

Imagine a bag with 5 coins, 4 are fair and 1 has two heads. If we take a coin out and toss it three times and get three heads, what is the probability the coin is fair? Perhaps you can intuit the answer, but we can draw out the useful part of conditional probability tree to do this quantitatively. This means we ignore the branches that do not produce the outcomes we have observed.
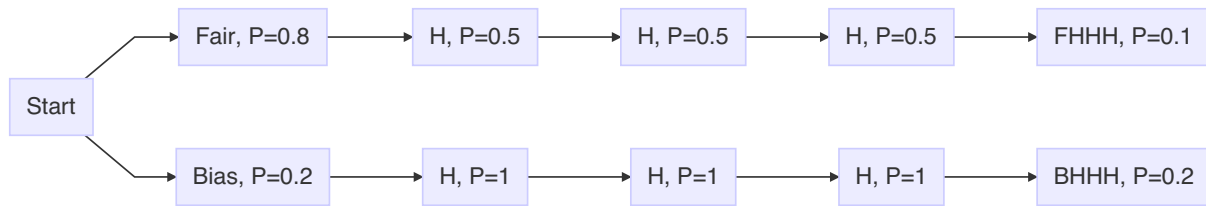
On each branch we observe three heads in a row (as stated in the initial problem). In the top branch the probability of selecting a fair coin is $P = 0.8$ and each subsequent toss has a $P = 0.5$ chance of coming up heads, giving a total probability for this branch of $P = 0.8 \times 0.5^3 = 0.1$. The bottom branch starts of with a less probable event, a $P = 0.2$ chance of picking the biased double headed coin, but each subsequent toss is certain to be a head with $P = 1$, giving a total probability for this branch of $P = 0.2 \times 1^3 = 0.2$.

To answer the question what we need to know is how much more likely the observed coin is to be fair or biased. This is now just simply the ratio between FHHH and BHHH, so $0.1/0.2 = 0.5$. I.e. it is twice as likely to be biased as fair, or stated differently 2 out of 3 times we do this experiment we would expect to find that the coin was biased.

The Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say door No. 1, and the host, who knows what's behind the doors, opens another door, say door No. 3, which has a goat. He then says to you, "Do you want to switch your pick to door No. 2?" Is it to your advantage to switch your choice? In this version of the problem the host *always* opens one losing (goat) door at random, if there is only one option he chooses that. This is a rephrased version of a question sent to Marilyn vos Savant (self styled "smartest person in the world" in Parade Magazine (1990).

First of all, let us look at the full tree of outcomes where the car is first placed randomly behind one of the three doors, we then randomly pick one of the three doors as our chosen door, and then the host randomly shows us what is behind one of the three doors (the constraint here being he will never show us what is behind the door we have picked, and he will never open the door revealing the car).
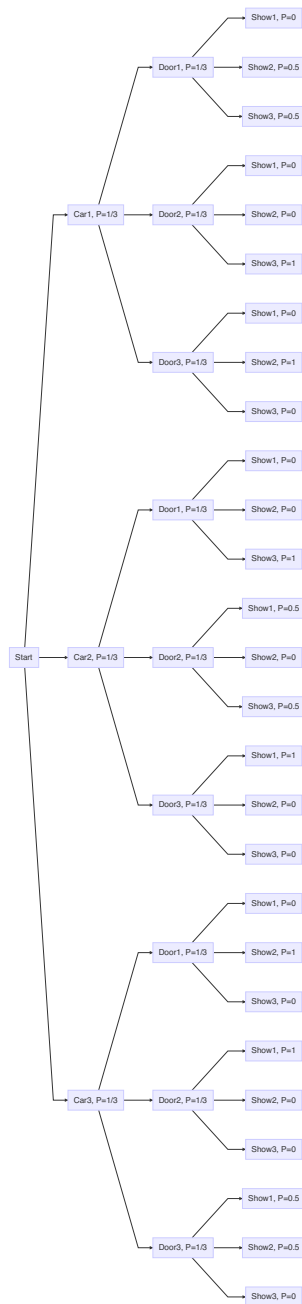
This question is famous for causing a great deal of confusion. But it is perfectly possible to create a conditional tree to guide us to the correct answer. As stated, we will start from the point where we have already picked door 1 (so pruning out other scenarios). There are two scenarios we need to consider then, one where the second goat is behind door 1, and the other where it is behind door 2. In both of these scenarios the host ends up picking door 3.
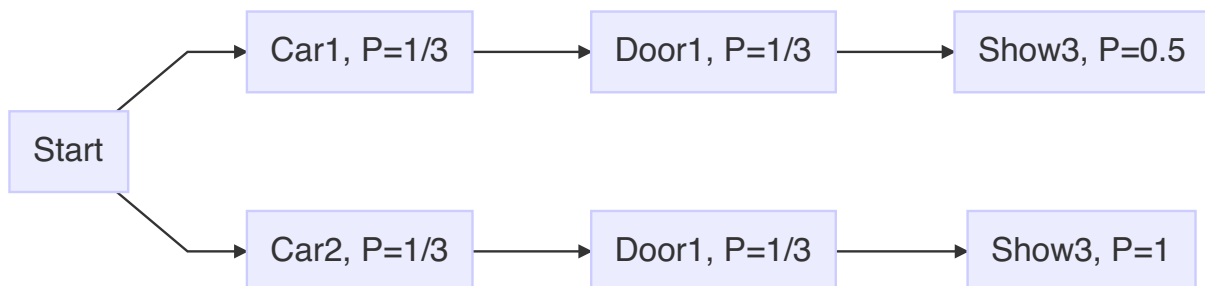
The above is a highly pruned version of the problem. Like the 3-heads problem above, we only consider

the route of conditional outcomes that could produce the observed state. Given the question states we pick door 1 and are shown a goat behind door 3, the only question left is the chance of the two routes that could produce this outcome. In one the other goat is behind door 1 (your door), and in the other it is behind door 2. If the goat is behind your door then it is certain ($P = 1$) that the host will have to show you the goat behind door 3 (since he cannot show you the car). If the goat is instead behind door 2 then there is a $P = 0.5$ chance, i.e. he could equally have shown you the goat behind door 2, it was just chance that meant he showed you door 3.

Given our aim is to maximise the chance of choosing the door with the car, we can use our tree to determine whether or not we switch doors. Since the first two parts of the tree always have the same probabilities (the chance of the car being behind a specific door and the chance of us picking a specific door are always 1/3), the conditional chance of the car being behind our current door (door 1) is determined by the probability fractions of the last branch alone: $P = 1/3$ (0.5/1.5), and the chance of the car being behind door 2 is $P = 2/3$ (1/1.5). Since the latter is more likely, we should swap to door 2. On 2/3 occasions swapping will be the correct strategy.

A more extreme way to view the problem is to increase the number of door, say to 1,000. Then say instead of the host randomly opening 1 door with a goat behind it, he instead randomly opens 998 doors with goats behind leaving 1 closed and the original door that you choose. It should be quite intuitive that the only reason not to swap is if you had originally guessed the correct door (with 1/1000 chance), i.e. in 999/1000 occasions swapping would be the correct strategy. This is also true with our original problem. The only reason not to swap would be if we had randomly guessed the correct door with 1/3 chance, so in 2/3 occasions the correct solution is to swap (as the pruned conditional probability tree above already revealed.

There are a lot of unhelpful discussions online about the "right' way to think about this problem, but the above is the simplest correct form of logic to apply. If this is making sense to you then good news: you are thinking in a Bayesian manner (even though you might not know what that means yet, it is a good thing).

A Rare Disease

There is a new disease called Frequentitus that is entirely fatal to anyone who contracts it. 0.1% of people are affected. Luckily there is a test administered by Perth Bayes Health. This test is 99.9% accurate at correctly determining you are positive, and 99% accurate at correctly determining you are negative. Unfortunately, you test positive. How worried should you be?

One way of representing this is a truth table:

Compare	Reality.Positive	Reality.Negative
Measured-True	True Positive	False Positive (Type I)
Measured-False	False Negative (Type II)	True Negative

Notice the off diagonal parts are also known as Type-I error (false-positive) and Type-II error (false-negative). You need to be careful in the naming here, since you could interpret "false-negative" " to mean you have measured false, and the reality is negative (i.e. a correct outcome), but instead the first part refers to whether the measurement was a success. This means true-positive and true-negative are both good, and false-positive and false-negative are both bad. You see reference to the error "Type" " a lot in reference material, but here we will avoid it from now because people find it hard to remember which is which! The other key part is that by construction the columns of the truth table must sum to 1 (the reality probabilities)

We can now add some numbers to this table from the original question.

Compare	Reality.Positive	Reality.Negative
Measured-True	0.999	0.01
Measured-False	0.001	0.99

The diagonals look very healthy (both close to 1), so this sounds ominous for our prognosis, but we need to weight the outcomes by how likely they are given that 0.1% of the population are actually affected. Again we can make a pruned probability tree.
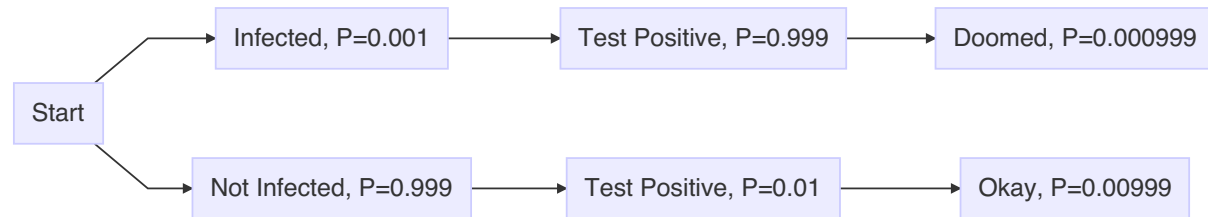
So the good news is that given how rare the disease actually is in the population, if we test positive then 10/11 times we will be okay (so we will have suffered a stressful false-positive). This is a regular headache for criminal trials, where the jury can struggle to understand that even if a test sounds like it is extremely accurate (99.99% etc), it is often much more likely a given person is a false-positive if the population is rare.

COVID-19

Status at August 2020 (i.e. below phrasing was correct at that time) Since this course was first presented we have had an outbreak of a serious disease: COVID-19 caused by the SARS-CoV-2. A hot topic is the veracity and utility of antibody testing, where the best current (as of August 2020) antibody test has a true-negative rate (chance of being negative for C19 antibodies when testing negative) of 99%, and a true-positive rate (chance of being positive for C19 antibodies when testing positive) of 90%.

We can recreate our truth table:

Compare	Reality.Positive	Reality.Negative
Measured-True	0.9	0.01
Measured-False	0.1	0.99

Imagine you are living mainland Europe which is reporting baseline infection rates of around 5% currently (again, as of August 2020) and you take an antibody test and test positive, what is the chance you truly have the antibodies?
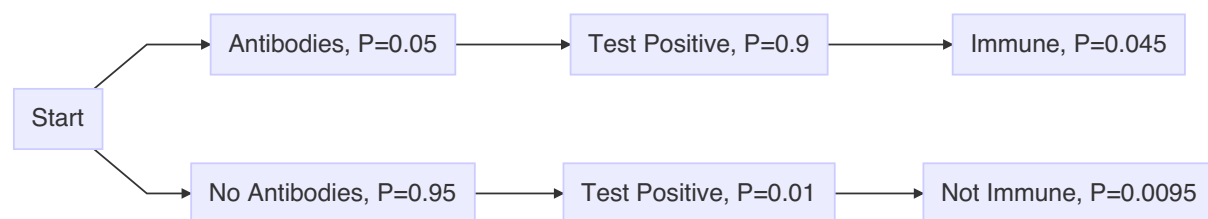

So you are about ~83% likely to have C19 antibodies when testing positive. The caveat is that more generally until the false-positive rate of the test is lower than the proportion of people who have antibodies you will be more likely to not have antibodies when testing positive for them. This makes it difficult in countries like Australia to use antibody test as a justification to loosen restrictions (since the prevalence rates are very low, much lower than 1%). Even in Europe, with 5% infection rates, allowing the 17% of people who do not have antibodies to behave as if they do (no social distancing etc) might be considered unethical given the mortality rate of the disease (1-3%).

The positive side is that even ~80% immunity rates might be high enough to give the community at large herd immunity. The key numbers are R_0 (the reproduction rate in a society with no immunity and no restrictions) and R_0 (the reproduction rate in a society in practice, i.e. with vaccines, immunity and restrictions etc). The virus would not spread efficiently and $R_f < 1$ (the average number of people infected by each positive person. R_f is easy to calculated approximately- if in a completely non-immune population $R_0 = 3$ then all things being equal $R_f = 3/5 = 0.6$ if 80% are immune, and this is why for the

original variant the target herd immunity fraction for C19 is estimated to be $1 - 1/3 \sim 67\%$.

How about instead of testing positive you test negative?
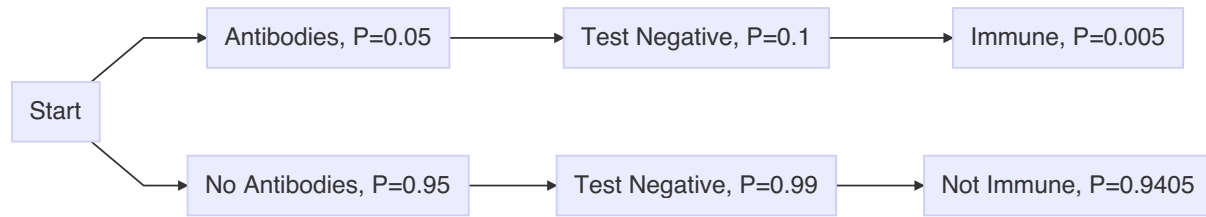
This means if you test negative you almost certainly (99.5% likely) do not have any antibodies. This really highlights that in the case of *rare* diseases it is the false-positive and true-negative rates (where one is 1 minus the other) that critically matter.

Status at August 2021 (i.e. below phrasing was correct at that time) The above I have left untouched from August 2020 as a sort of notes time capsule. The comments regarding efficacy against infectiousness of vaccines and the target for "herd immunity" has proved to be, sadly, extremely optimistic based on the much less infectious original strain of SARS-CoV-2. At that time there had not been any serious mutations that dramatically changed the baseline infectiousness of the virus, but that all changed towards with end of 2020 with the emergence first of all of the "Greek Letter" variants.

As of August 2021 the R_0 of the Delta variant is somewhere in the range 4–8, and it appears vaccines have only about 50% efficacy at stopping you being infected and infectious (this is different to be being sick with C19, where the efficacy is much higher). This means with Delta even if 100% of the population is vaccinated your R_f (without other restrictions) will be around 2. Even under stay-at-home orders in NSW with R_f is hovering around 1.3.

Currently there is no chance we will achieve herd immunity for Delta, and this is now very clear from looking at counties with very high vaccination rates (e.g. Israel and the UK), the very serious positive is many fewer people are getting seriously sick, hospitalised or dying thanks to most people being vaccinated.

Given this will be a multi-year drama, I will make some statistically motivated prediction for next year's version of the course (and we shall see how close these predictions are, I promise not to cheat!)

1) No country will achieve $R_f < 1$ for Delta with vaccines alone, i.e. without additional restrictions.

2) A new variant will emerge which will have $R_f > 1$ even with a highly vaccinated population and additional restrictions.

The above should be considered a very strong and positive case for getting vaccinated personally- there is almost no chance anybody will be able to rely on "herd immunity" to protect them statistically (like is possible with flu each year due to its very low $R_0 \sim 1.5$), so everybody will need the personal protection of a vaccine. The game changer *might* be that vaccines are developed that offer full immunity even to infection- this is the case with the measles vaccine, which is critical since it has an unbelievably high $R_0 \sim 20$, i.e. if vaccines did not generate true immunity then it would be constantly endemic everywhere (as it used to be before vaccines of course). The counter to this is that there are three very different types of vaccines currently (viral vector, mRNA and protein sub-unit) and a large number of each made by competing companies and countries, and yet none seem to offer more than 50% efficacy against infection.

In summary, the statistical case is very clear: get vaccinated as soon as you can with any vaccine that has been approved. The "as soon as you can" is key since the optimal second jab time is between 8-12 weeks after the first for all currently approved vaccines in Australia with a further week to take full effect. A single jab of any approved vaccine currently gives very low levels of protection. You cannot safely wait until you are in a serious outbreak to protect against it since you will be at serious risk for 9-13 weeks after your first jab given $R_f > 1$ for Delta even under reasonably strict restrictions.

Status at August 2022 (i.e. below phrasing was correct at that time) Well here we are in 2022, and let us have a look back on those predictions:

1) TRUE. Only China and New-Zealand fully squashed significant (truly exponential) Delta outbreaks with pretty extreme interventions (and in New-Zealand's case even the better vaccines could not

avoid this).

2) TRUE. Omicron (and children) has arrived and brought us a variant with $R_0 \sim 12$ and $R_f > 1$ even with high vaccination and previous infection rates.

For the future the question is really not will Covid disappear (it will not), but will it become endemic. In this usage we mean endemic in the sense that it is always present in the population, but at low levels and without majorly disruptive waves. There are only two countries left that are taking highly interventionist paths (China and North Korea), so my prediction for PHYS5513 2023 are:

1) Mandates and restrictions will only decrease around the world, not ramp up (this is just sociology speaking, not science).

2) Rather than a new Greek letter variant, we will be dealing with direct children of Omicron that are increasingly more transmissible and milder. This is informed by most previous major pandemics getting milder over time (since viruses optimise for total spread rather than illness and death).

Fingers crossed all our classes in 2023 are in person, and neither of the lecturers have to isolate for Covid!

Status at August 2023 (i.e. below phrasing was correct at that time) Looks like I nailed it this year:

1) TRUE

2) TRUE

I don't have any strong predictions for next year. It's almost certain everybody will have caught Covid at least once by then (are there even any hold outs now?)

Status at August 2024 Seems like Covid is yesterday's news now. An interesting aside is that we are still dominated by children of Omicron (e.g. KP.3.1.1 variant). That is probably a further sign that we are in the endemic stages of the disease, and hopefully that also means no massive future disruption (just pulsing waves).

Facial Detection Case Study

A quote from The Guardian:

"South Wales police began trialling the technology in June last year [2018] in an attempt to catch more criminals. The cameras scan faces in a crowd and compare them against a database of custody images. As 170,000 people arrived in the Welsh capital for the football match between Real Madrid and Juventus, 2,470 potential matches were identified. However, according to data on the force's website, 92% (2,297) of those were found to be false-positives." A scary sounding statistic, but is this statement fair?

The correct way to interpret the above is to say that out of 170,000 people 2,297 were false-positives ($2297/170000 = 0.0135$), and 173 were true-positives. To make a truth table we have to guess at the true-positive and the true/false-negatives, but we can be optimistic and state that the true-negative rate was 100% and the false-negative rate was 0% (so everybody who should have been identified was).

Compare	Reality.Positive	Reality.Negative
Measured-True	1	0.0135
Measured-False	0	0.9865

So the facial recognition test is probably actually very accurate, the issue is that a vast minority of people are criminals (which is probably a positive takeaway message). For the absolute rate of true-positives to dominate over the false-positives, the scanning software would need to be 99.9% accurate at avoiding false-positives (compared to the 98.65% it currently is). The other route is to use it in areas where the latent rate of criminality is a factor ~ 10 higher! The misleading part of the story was to compare the false-positive results to the true-positive results (which is not a common way to state the false-positive rate). It is still true to say that only 8% of people positively identified had anything to feel guilty about.

Bayes' Theorm

Many independent minds have formulated a Bayes-like theory:

- Thomas Bayes (provider of the name)
- James Bernoulli
- Pierre Simon Laplace (father of much of modern statistics)

It formulates ideas that are elementary in origin that we have basically introduced in the examples above:

- The probability that A occurs after B having occurred is P(A|B). i.e. the likelihood we are in AnB (the intersection of event A and event B) given we are in B (i.e. event B has definitely occurred already).
- The outright probability of AnB is therefore the probability that B occurs times P(A|B):
- P(AnB) = P(A|B) P(B), and similarly:
- P(AnB) = P(B|A) P(A)

We can visualise this easily as a Venn diagram:
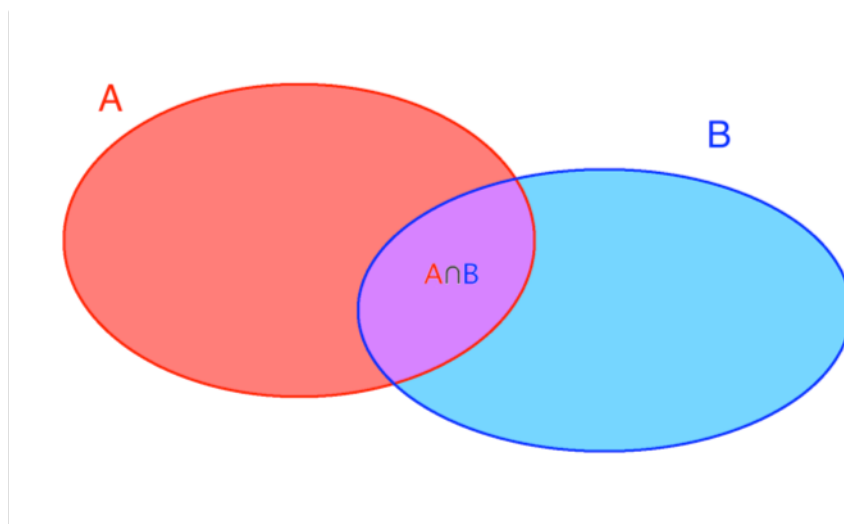
Figure 1: Discrete Bayes' theorem as a Venn diagram

We can generalise this a bit to ask what is the chance we observe the event A given different events B. This might look like a more complex Venn diagram.

The case above is the simplest, where we know what the possible B probabilities and intersections with A are.

So we can see that P(A) = P(AnB$_1$) + P(AnB$_2$) ... + P(AnB$_k$)

A real example is a bag with 5 coins, 4 are fair and 1 has two heads (recognise this setup?). If we take a coin out and toss it 3 times and get 3 heads, what is the probability the coin is fair? You might be able to answer this "intuitively", in which case you're a natural Bayesian, for the rest we can work through it.

Above is a Venn schematic of the problem. This is much simpler than the complicated example on the previous slide. It should also look familiar from the pruned probability version we drew earlier.

We are not interested in the chance of picking a biased coin, or observing 3 heads, rather we want to know the likelihood the coin is biased *given that we've already observed* 3 heads. This method of thinking is key to Bayes. We do not care how probable an observation is, just what the *most probable cause of the observation* was. As an aside, this is an idea that juries find famously hard to grasp: given 5 premature deaths on a ward the most likely explanation is usually a random sampling of a parent (nationwide) statistical distribution, rather than murder. 5-simga ('one in a million', although more like 0.57 in a million) events have to occur somewhere in populations much larger than a million, after all.
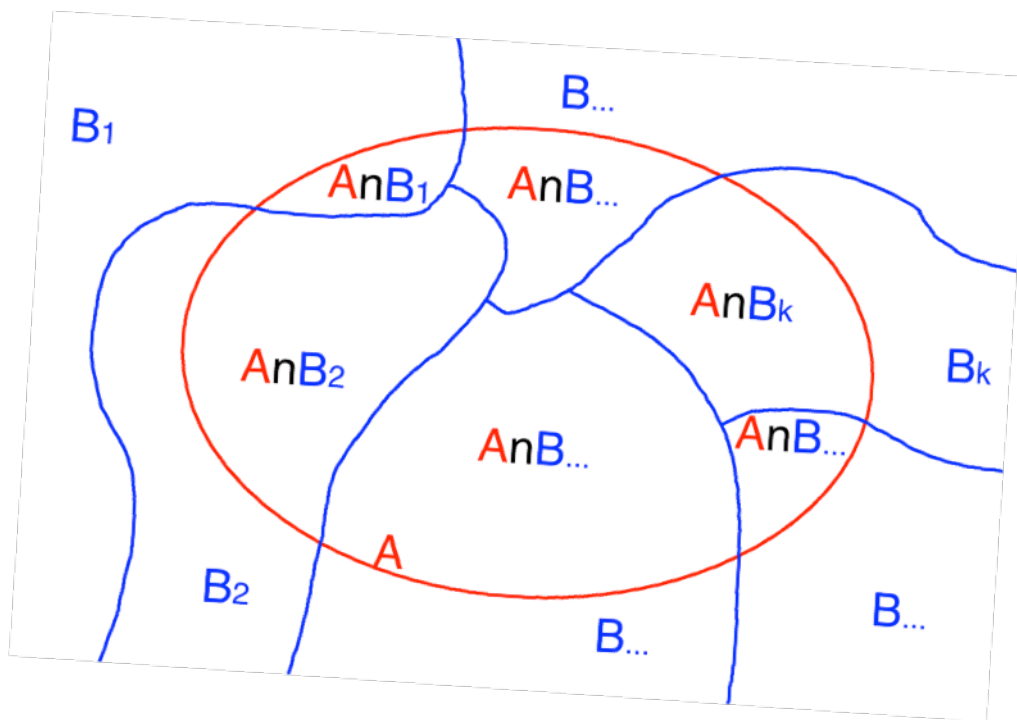
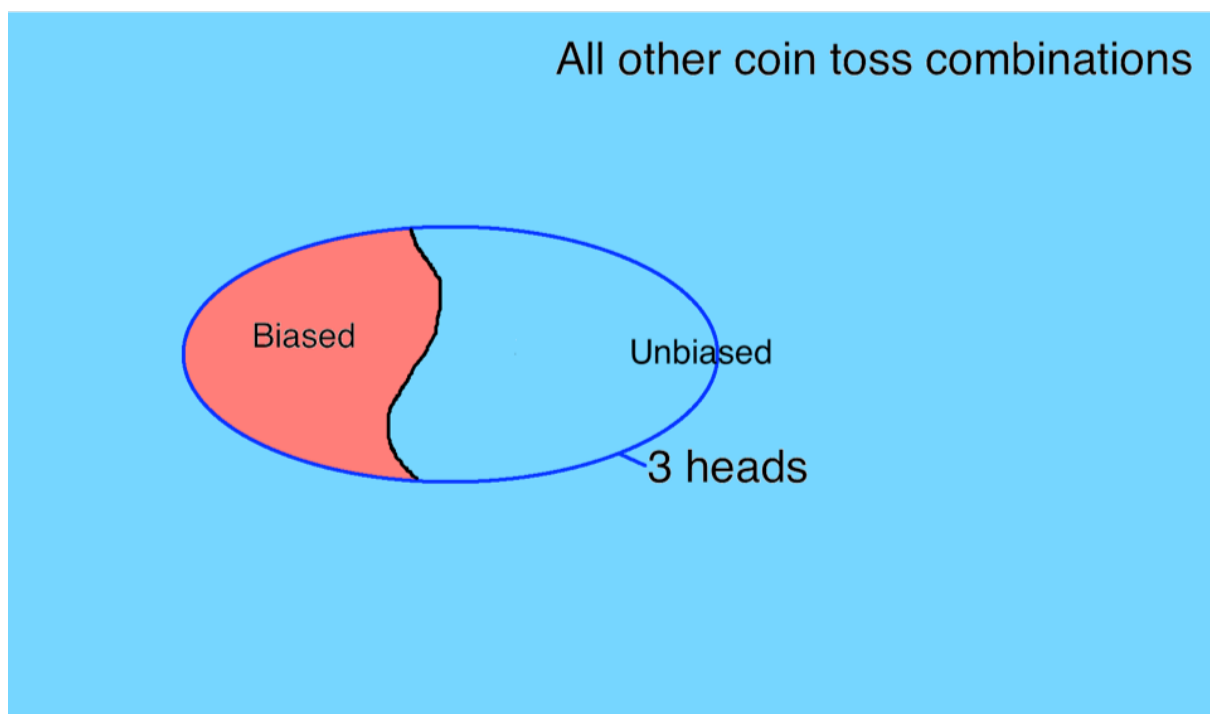Figure 2: Multiple condition Bayes' theorem as a Venn diagram

Figure 3: Simplified version of coin problem

The answer we want is what is the probability of our coin being fair {P(fair)} given that we already know it has been tossed 3 times and come up heads each. i.e.:

- P(fair|3head), which we know equals P(fair n 3head)/P(3head)
- We now use P(3head) = P(3head n fair) + P(3head n biased)
- So, P(3head) = P(3head|fair)P(fair) + P(3head|biased)P(biased)
- P(3head) = $(1/2)^3(4/5) + (1)^3(1/5) = 1/10 + 2/10 = 3/10$, so 7/10 times we would expect to pick a coin, toss it 3 times, and *not* get 3 heads in a row.
- P(3head n fair) = P(3head|fair)P(fair) = 1/10 (as above)
- P(fair|3head) = (1/10) / (3/10) = 1/3, so a one third chance, i.e. it is more likely the coin is biased (twice as likely, in fact).

For the keen student, how many heads in a row do we need to toss until we are confident the coin will be biased (two headed) 99% of the time?

Bayesian Monte Hall

This is a famous puzzle for being non-intuitive (for most people at least), but this is the Bayesian approach. We saw the probability tree approach before, which is strongly related to the Bayesian thinking behind it. These are the more Bayesian steps we can now write down. To simplify the approach here we state that we certainly decided to pick door 1 coming into the game ($P = 1$), other outcomes of playing (picking door 2 or door 3) can then just be argued by symmetry.

We want to know the probability that door 2 contains the car given we know door 3 is a goat, i.e. P(D2car|D3goat), to decide whether to swap:

- P(D2car|D3goat) = P(D2car n D3goat) / P(D3goat)

We can calculate the intersection easily:

- P(D2car n D3goat) = P(D3goat|D2car)P(D2car) = (1, since that is the only possibility)(1/3, initial random chance of the car being behind any door) = 1/3

The probability that Monty shows us a goat behind door 3 is P(D3goat), and requires us to consider 3 possibilities of where the car might, one of which is clearly impossible:

- P(D3goat) = P(D3goat|D1car)P(D1car) + P(D3goat|D2car)P(D2car) + P(D3goat|D3car)P(D3car)
- P(D3goat) = (1/2)(1/3) {Randomly chooses D3} + (1)(1/3) {No choice with door} + (0)(1/3) {N/A} = 1/2

The first 1/2 fraction takes thought- we need it because Monty could have picked D2 or D3 if the car truly was behind door 1. So finally

- P(D2car|D3goat) = (1/3) / (1/2) = 2/3
- Currently have P(D1car) = 1/3, so yes, we should definitely swap, twice as likely to win!

The simplest way to arrive at an 'intuitive' answer is to realise that you only lose when the car is behind your original door, and this only happens 1/3 of the time, so 2/3 of the time you win by swapping. This is as we suggested earlier when looking at the explicit tree based approach.

And Now for the Magic

You might have noticed that when assessing the better of two options we end up doing a calculation that looks like: * $(X_1/Y)/(X_2/Y)$, where $Y = X_1 + X_2 + X_3 + \ldots X_k$

In cases where we only care about calculating the relative likelihood of two models, calculating Y is clearly unnecessary (it divides out), we can just skip to calculating X_1/X_2. This equivalent to saying:

- P(A|B) = P(B|A) P(A) / P(B), so
- P(A|B) $\propto$ P(B|A) P(A)

This means we can avoid the expensive, difficult, and in practice often impossible calculation of P(B). The ratio of different P(A|B) is known as Bayes factor (BF from here). We actually made a simpler shortcut (with less justification) when using the tree based approach, where we commented that all branches except the last one have the same probability ($P = 1/3$) so we can ignore their contribution to the conditional probability ratios since it will cancel out to give 1 when dividing the two branches we were considering.

Monte Hall again

Using this insight we can simplify the problem to:

- BF = P(D2car|D3goat) / P(D1car|D3goat)
- BF = P(D3goat|D2car)P(D2car) / P(D3goat|D1car)P(D1car)

Being smart we know that P(D1car) = P(D2car), so. . .

- BF = P(D3goat|D2car) / P(D3goat|D1car) = (1) / (1/2) = 2

This means door 2 having a car is twice as *likely* as door 1 (our current door) to have the car behind it. We do not know what the exact probability is because we have made some short cuts, but we have arrived at the correct insight that we should swap, and we know the correct degree of increased likelihood of winning (a factor of 2). If we are comparing models, relative likelihood is all we really care about. This is ultimately even simpler than our probability tree approach from earlier.

We have now looked at the Monty Hall problem using a number of Bayesian perspectives and different degrees of thoroughness, so this should be plenty to get you thinking along the right track with similar types of problems.

Bayes Equation

With all this ground work we can now write down Bayes' equation for discrete probability events:

$$P(\theta, y) \propto P(y, \theta) \times P(\theta)$$

This equation can be read in a natural (and hopefully now intuitive) manner:

The probability of a particular model given by the data ($P(\theta, y)$) is proportional to ($\propto$) the probability of the data given the model ($P(y, \theta)$) times ($\times$) the probability of the model ($P(\theta)$).