Chapter 2

Probability

What is probability? Probability measures the *chance* of a specific event happening in the future. Probability is fundamental in developing an understanding of inferential statistics. This chapter will focus on the mechanics of probability and key definitions while future chapters will explore how it assists in effective statistical inference.

2.1 Just a Slurp of Set Thoery

Before diving into probability, we need to quickly recall some basic concepts in set theory. Sets are simply groups of objects. The objects can be numerical or categorical in nature. Sets are often notated using capitol letters such as F,Y, or I.

The objects contained within sets are referred to as *elements*. The number of elements in a set is defined as the sets *cardinality*. For a set A the cardinality is notated $|A|$.

Example 1: How many elements are in M?

$$M = \{m_1, m_2, ..., m_n\}$$

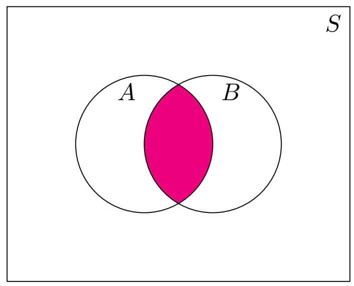
There are clearly n objects in M so we say the cardinality is n.

To visualize sets in an effective manner, Euler diagrams may be used. A box represents the universal set, while circles represent respective subsets. Intersections are shown through overlap.

Universal Set: The set containing all events of consideration. Although the letter U is often reserved for the universal set we will use S which in the context of probability will stand for the *sample space*.

The Empty Set: The empty set is denoted $\emptyset$ and contains no elements. $\emptyset = \{\}$. How many empty sets are in the universal set?

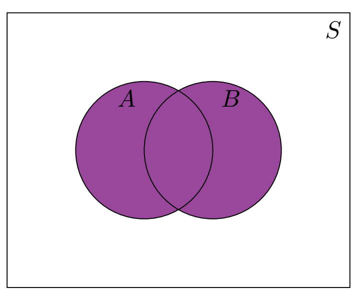
Intersection: For two sets A and B the intersection is denoted $A \cap B$ and includes elements within A **and** B



Example 2: $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{1, 2, 14\}$, what is $A \cap B$?

$$A \cap B = \{1, 2\}$$

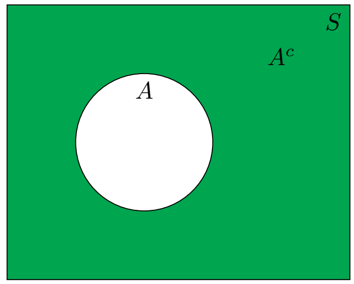
Union: For two sets A and B the union is denoted $A \cup B$ and includes elements within A **or** B



Example 3: $A = \{\text{Apple}, \text{Banana}, \text{Orange}\}$, $B = \{\text{Pineapple}, \text{Squash}, \text{Potato}\}$. What is $A \cup B$

$$A \cup B = \{\text{Apple}, \text{Banana}, \text{Orange}, \text{Pineapple}, \text{Squash}\}$$

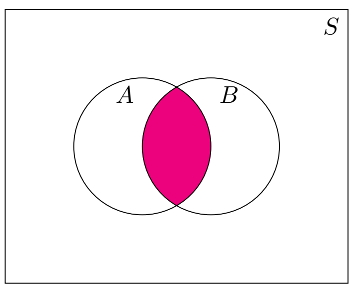
Compliment: For a set A the compliment is denoted A^c and includes elements from the universal set that are **not** in A.

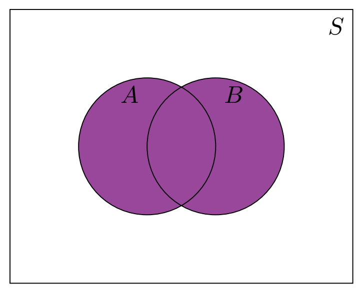


There are several laws of sets we will apply to probability.

***Commutative Laws* for sets.**

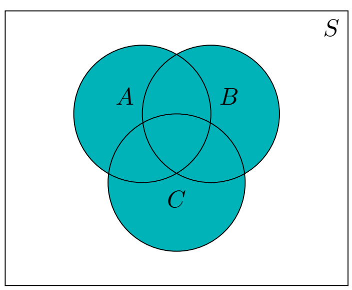
$$
\begin{aligned}
A \cup B &= B \cup A \\
A \cap B &= B \cap A
\end{aligned}
$$

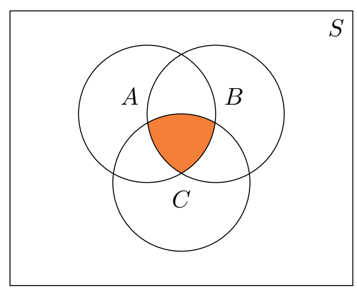




***Associative Laws* for sets**

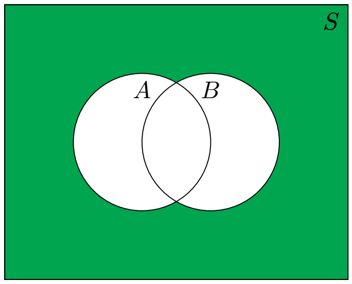
$$
\begin{aligned}
(A \cup B) \cup C &= A \cup (B \cup C) \\
(A \cap B) \cap C &= A \cap (B \cap C)
\end{aligned}
$$



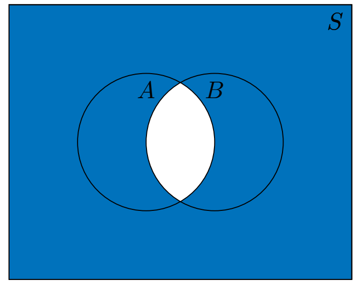


Example 4: For sets A, and B fill in an Euler diagram to represent the following

$(A \cup B)^c$



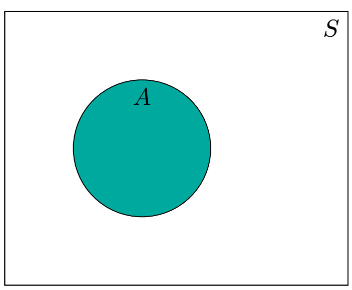
$(A \cap B)^c$



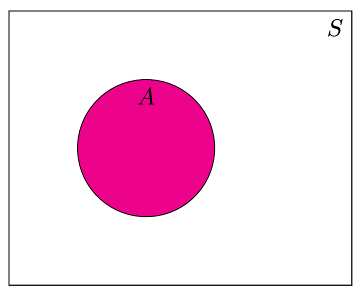
***DeMorgan's Laws* for sets.**

$$\begin{aligned}(A \cup B)^c &= (A^c \cap B^c)\\ (A \cap B)^c &= (A^c \cup B^c)\end{aligned}$$

$A \cup \emptyset$



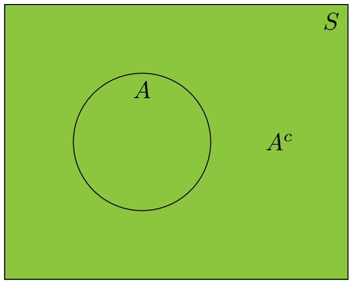
$A \cap S$



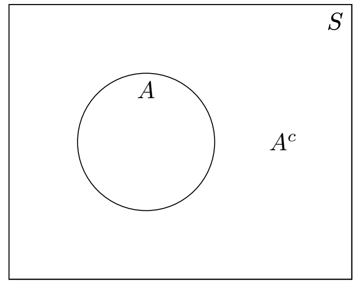
***Identity Laws* for sets**

$$\begin{aligned}A \quad \cup \quad \emptyset &= A\\ A \quad \cap \quad S &= A\end{aligned}$$

$A \cup A^c$



$A \cap A^c$

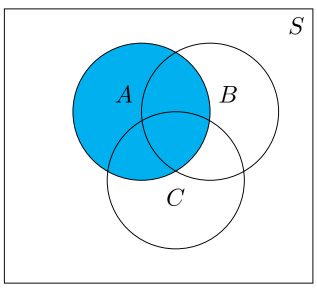


***Compliment Laws* for sets**

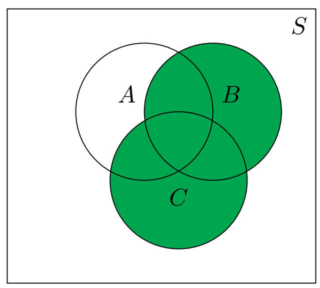
$$
\begin{aligned}
A \cup A^c &= S \\
A \cap A^c &= \emptyset \\
(A^c)^c &= A
\end{aligned}
$$

Example 5: For sets A, and B fill in an Euler diagram to represent the following

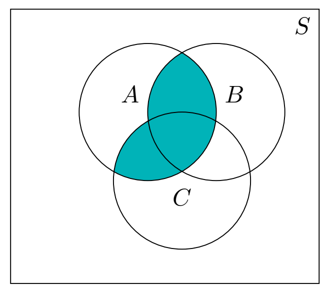
A



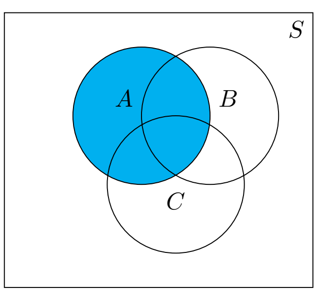
$B \cup C$



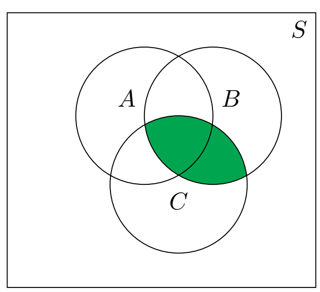
$A \cap (B \cup C)$



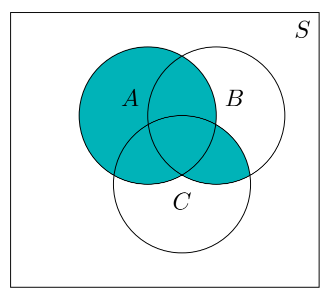
A



$B \cap C$



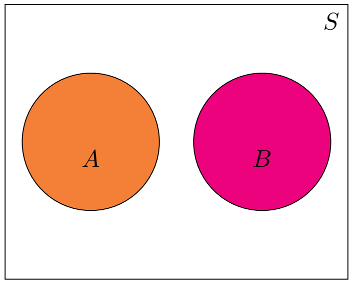
$A \cup (B \cap C)$



***Distributive laws* for sets.**

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

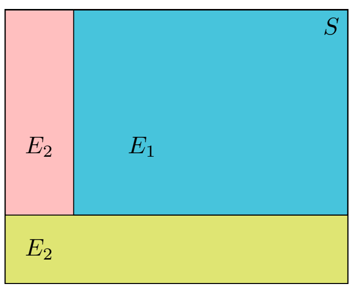
Example 6: Draw two sets such that $A \cap B = \emptyset$



Disjoint Sets: Two sets A and B are disjoint if $A \cap B = \emptyset$

Example 7: Split all the elements in the set S up into three disjoint sets.

Partition of S



A Partition of a set
If we partition a set into subsets E_1,E_2,...,E_n we are simply splitting **all the elements** of the set into n subsets such that that all the pieces are disjoint (no overlap).

Example 8: Consider the set $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$. Partition the set into $n = 3$ different subsets.

Lets choose $E_1 = \{2\}$, $E_2 = \{3\}$ and $E_3 = \{4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$

2.2 Probability and Sets

Random Experiment: A random experiment is a process that leads to a specific outcome or observation. Since the experiment is random, we do not know what the outcome of the experiment will be but we do know one unique outcome will occur.

Sample Space: For a random experiment, the sample space is the set of all possible outcomes or *universal set*. The sample space is often denoted S. A sample space may be *discrete* (countable number of elements) or *continuous* (non-countable number of elements)

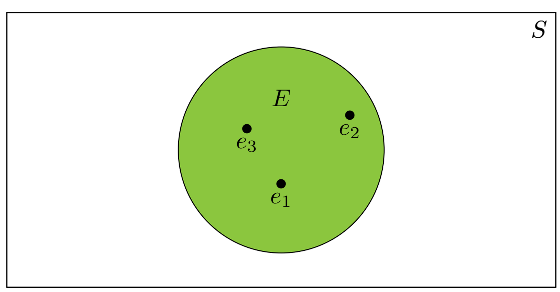
Sample Points: If the sample space is a set of outcomes, a sample point is a particular element within that set.

Simple Events: Simple events correspond to a singular sample point within the sample space.

Compound Events: Compound events correspond to a collection of sample points within the sample space.

Probability: For an event X, the probability of X happening is denoted $P(X)$ and measures the likelihood that the event will occur when an experiment happens.

Set notation gives us a convenient way to represent probability for a random experiment. The set of all possible outcomes (universal set) is the sample space. Events are simply subsets of the sample space. We will notate simple events with e and can think of them singular points or elements in S. We notate compound events with E and think of them as groups of simple events in S. Our laws for sets also correspond nicely with probability laws.



The Axioms Of Probability
For an event A, and Sample space S.

$$0 \leq P(A) \leq 1$$

$$P(S) = 1$$

For *pairwise* mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(A_1 \cup A_2 \cup + \cdots + A_n) = P(A_1) + P(A_2) + \cdots P(A_n)$$

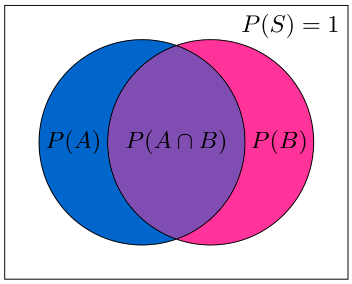
Probability for the *Union* of two events

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

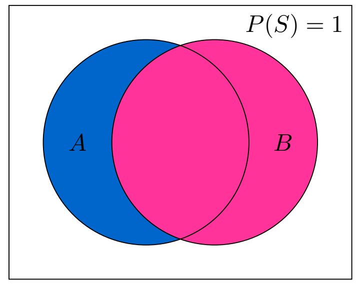
This represents the Probability of event A **or** event B occurring.
Also notice if A and B are mutually exclusive $P(A \cup B) = P(A) + P(B)$

We can prove this heuristically by examining an Euler Diagram.

$P(A) + P(B)$



$P(A \cup B)$



Notice that if we take $P(A) + P(B)$, we count $P(A \cap B)$ twice. We must subtract it off once, so that we only have $P(A \cup B)$.

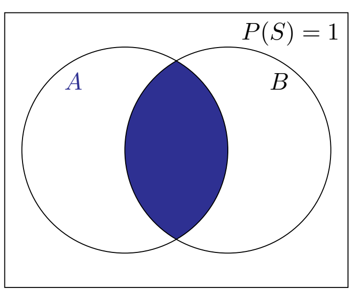
$$\begin{aligned} P(A) + P(B) &= P(A \cup B) + P(A \cap B) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Law of Total Probability

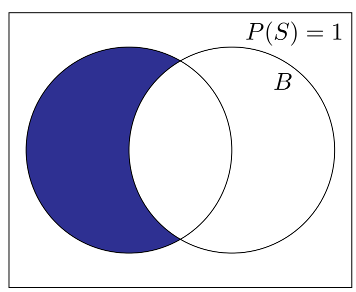
$$P(A) = P(A \cap B) + P(A \cap B^c)$$

If we would the the 'total' probability for event A We want the probability of $A \cap B$ **or** $A \cap B^c$. In probability **OR** often corresponds to **ADDING** probabilities.

$P(A \cap B)$



$P(A \cap B^c)$



Probability for the *Intersection* of two events

$$P(A \cap B)$$

This represents the probability of **both** events A **and** B occurring.

Other laws of sets may also be applied to probability.

DeMorgan's Laws **for probability**

$$P(A \cup B)^c = P(A^c \cap B^c)$$

$$P(A \cap B)^c = P(A^c \cup B^c)$$

Distribution Laws **for probability**

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

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

Two events A and B are **Mutually Exclusive** if $P(A \cap B) = 0$. This means that it is **impossible** to observe **both** A and B during the experiment. In Euler diagrams mutually exclusive events are represented as *disjoint* sets.

Example: If a six sided dice is rolled the possible events are **1, 2, 3, 4, 5, and 6**. Observing a one and two cannot both happen at once. $P(\text{Roll one} \cap \text{Roll two}) = 0$, so these are mutually exclusive events.

Example 1: When rolling a six sided die what is $P(\text{Roll a 2})$?

$$S = \{1, 2, 3, 4, 5, 6\}$$

Our event of interest is 2. Notice this is a simple event. Our probability of the event occurring is $P(\text{Roll 2}) = \frac{1}{6}$.

Example 2: When rolling a six sided die what is $P(\text{Roll an even number})$?

$$S = \{1, 2, 3, 4, 5, 6\}$$

Our events of interest are 2, 4, and 6. Notice this is a compound event. Our probability of the event occurring is $P(\text{Roll Even}) = \frac{3}{6} = \frac{1}{2}$.

Example 3: When rolling two three sided die what is $P(\text{Rolling a two})$?

$$S = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$$

Our events of interest are 12, 21, 22, 23, 32. Notice this is a compound event. Our probability of the event occurring is $P(\text{Rolling a two}) = \frac{5}{9}$.

Example 4: When flipping three coins what is $P(\text{Observe a head})$?

$$S = \{\text{HHH,HTH,HTT,HHT,THH,TTH,THT,TTT}\}$$

Our events of interest are every event with the exception of TTT. Notice this is a compound event. Our probability of the event occurring is $P(\text{Observing a head}) = \frac{7}{8}$.

But How Many Elements Are In The Sample Space?

In order to determine probability, we must be able to find both the number of elements in the sample space as well as the number of elements in the event of interest. In mathematics this process is referred to as *counting*. You might find it more challenging that the counting you did in Kindergarten.

> ***Multiplicative rule* for counting**
> The multiplicative rule is used when you are repeating an experiment with m outcomes n times. Recall that the cardinality of S is denoted $|S|$ and refers to the number of elements in the sample space.
> $$|S| = m^n$$

Example 1: Suppose you roll 11, six-sided die. How many outcomes are in the sample space?

Here there are $6^{11} = 362797056$ different outcomes.

Example 2: Six digits are selected at random in order to guess an IPhone password. What is the cardinality of the sample space?

Here there are $10^6 = 531441$ different outcomes in the sample space, so the cardinality is 531441.

Example 3: Suppose you attempt to shoot a basketball into the basket 18 times. How many outcomes are possible?

Here you may score or not on each shot, so there are $2^{18} = 262144$ different outcomes.

> ***Factorials***
>
> Factorials are used to represent the number of ways n objects can be arranged.
> $$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$

Example 4: You want to mount four different paintings in a row on the wall. How many different ways could you mount the paintings?

There will be $4! = 24$ different ways to mount the paintings.

Example 5: There is a red, green, and blue marble in a bag. If you pull the marbles out of the bag at random, how many different ways could you pull out the three marbles. Write out each way the marbles may be ordered.

You can pull out the marble $3! = 6$ different ways. RGB, RBG, BRG, BGR, GBR, GRB.

Permutations
Permutations describe the number of possible arrangements when n objects are selected from a total of m objects. It is important to note that the **order matters** in rearrangements.

$$_nP_m = \frac{n!}{(n-m)!}$$

Example 6: There are 10 books you are interested in reading, but you only have time to read 2 over the next few weeks. How many ways can you select a first, second, and third book to read?

Here there are $_9P_2 = 9 \cdot 8 = 72$ different rearrangements for the books

Example 7: There are 12 people competing in race. How many possible ways can the first, second, and third medals be awarded?

Here you are selecting 3 people from a total of 12 so there are $_{12}P_3 = 1320$ different arrangments.

Example 8: How many 4 letter words, with or without meaning, that may be created with 8 unique letters.

Here we have $_8P_4 = 1680$ different four letters words that may created with eight unique letters.

Combinations
Combinations describe the number of different groups of size m that can be selected from a set of n objects. Combinations are very similar to permutations, but here the order **does not** matter.

$$_nC_m = \binom{n}{m} = \frac{n!}{m! \cdot (n-m)!}$$

Example 9: There are 10 books you are interested in reading, but you only have time to read 2 over the next few weeks. How many ways can you select 2 different books?

Here you may notice the order does not matter. You may select $\binom{10}{2} = 40$ different groups of two books you can make.

Example 10: From a group of 12 people you must select a committee of 4. How many different committee's could you select?

You can select $\binom{12}{4} = 495$ different committees.

Example 11: A random sample of n people are selected from a population of N people. How many different samples are possible?

Here you want to know how many groups of size n can be selected from a population of size N. There are $\binom{N}{n}$ different combinations.

Arrangments with repeats
How many ways could I arrange the numbers $\{1, 1, 2, 3, 4\}$? Notice that the one repeats twice. To find the total number of arrangements we need to 'divide out' the number of ways the two items can be rearranged.

$$\#\text{ of arrangments} = \frac{5!}{2!}$$

In general if you have m objects and one of them repeats k times, the number of arrangements is

$$\#\text{ of arrangments} = \frac{m!}{k!}$$

Example 12: Suppose you have the letters A,A,A,D,E,F. How many different ways can you arrange the letters?

You can arrange the letters $\frac{6!}{3!} = 120$ different ways.

Example 13: You attempt to shoot 12 baskets. How many ways can you shoot 3 baskets in your twelve shots?

Here just think of a swish as an S, and a miss as an M. Now, using the letters you have S,S,S,M,M,M,M,M,M,M,M,M. The number of ways you can make 3 baskets is $\frac{12!}{3!9!}$. Notice this is $\binom{12}{3} = 220$ different ways.

Example 14: There are 8 different locations where drills will test for oil. How many different ways can the drills strike oil in 5 different locations.

Using the same method as the last problem there are $\binom{8}{5} = 56$ different ways to strike oil.

Example 15: A library has 4 copies of *Harry Potter*, 3 copies of *The Hobbit*, and 6 copies of *Eragon*. How many ways can the librarian order the books on a self?

There are $\frac{13!}{4!3!6!} = 60060$ different ways.

Probability for an event E
Recall that for a set A, $|A|$ is the cardinality (number of elements in the set).

In order to determine probability of an event E occurring we take the number of elements in the event, $|E|$, and divide by the number of elements in the sample space $|A|$.

$$P(E) = \frac{\#\text{ of elements in } E}{\#\text{ of elements in } S} = \frac{|E|}{|S|}$$

Probability for E using the compliment
Sometimes it is difficult to determine the number of elements in E, but easy to find the number of elements in E^c.

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

Example 16: An experiment consists of tossing a pair of dice. What is the probability that the **sum** of the dice is equal to seven?

Here we may summarize our results nicely in a two by two table.

+	1	2	4	3	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Here the sample space consists of 36 events. Our event of interest consists of seven elements. So it is clear that the probability is $(6/36) = (1/6)$

Example 17: An experiment consists of rolling three 3-sided die. What is the probability that the **product** is greater than two?

Unfortunately with three die a table doesn't work as nicely for summarizing the sample space. However, we know there are $3^3 = 27$ different options. Let's find the number of outcomes with products less than two.

There is only one way to roll three ones ($\{1, 1, 1\}$). Next with one two and two ones there are $\frac{3!}{2!} = 3$ options ($\{2, 1, 1\}$,$\{1, 2, 1\}$,$\{1, 1, 2\}$). In total there are 4 options such that the product≤ 2 so we know the remaining options are greater than two.

$$P(\text{Product} > 2) = \frac{27 - 4}{27} = \frac{23}{27}$$

Simulating Probabilities In An Experiment

Up to this point we have focused on *theoretical probabilities* that are derived using mathematics. *Theoretical probability* does not describe exactly what will happen in an experiment. We can also *estimate* probabilities for an experiment using simulation. This is done by conducting an experiment, and estimating the the probability for a certain outcome using the results.

Example 1: In a game of *Unstable Unicorns* each player is dealt 7 random cards, from a deck that contains 20 instant cards, 25 upgrade cards, 25 downgrade cards, 5 magic cards, 10 magical unicorn cards, and 15 basic unicorn cards. Mr. Merrick is surprised when he is dealt 7 unicorn cards. (Assume he is dealt the first 7 cards from the top of the deck). Estimate the probability of this happening *using simulation*

Here we are in interested in how likely being dealt a hand of 7 unicorns would be. There are 100 cards in total in the deck, we may assign digits to represent outcomes as follows:

- 0-19 represent instant cards
- 20-44 represent upgrade cards
- 45-69 represent downgrade cards
- 70-74 represent magic cards
- **75-84 represent magical unicorn cards**
- **85-99 represent basic unicorn cards**

Notice that we have assigned a unique number to each card in the deck. Also notice that we cannot pull out the same card twice so we must disregard any repeats in our simulation for each trial. Now we may simulate the experiment using a random number table.

Simulation 1: 38 16 42 13 98 48 14
Simulation 2: 62 86 38 38 44 19 98
Simulation 3: 75 77 48 94 64 67 71
Simulation 4: 94 74 25 46 74 56 71
Simulation 5: 63 83 78 37 86 14 46
Simulation 6: 41 29 44 21 74 69 40
Simulation 7: 15 43 35 85 21 48 15

I have ordered the random numbers to show 7 iterations of the experiment. In none of these cases do all observations fall in the range 75-99 so we would estimate the probability of of having a hand of all magical unicorns as 0/5=5%. Only five iterations will not create an extremely likely estimate. Using a computer we may run millions of iterations to find a better estimate.

The more iterations we simulate, the closer our simulated estimate will become to the true theoretical probability. Next we are interested in how many element compose our event of interest pulling 7 magical unicorn cards.

The **law of large numbers** suggests that as we repeat an experiment a large number of times the experimental probability will approach the theoretical probability. If flip a coin 1000000 times or 10 times which will provide a better estimate for the probability of getting a head?

Example 2: What is the theoretical probability of being dealt 7 unicorns?

Here we must fist understand what the sample space is. In this example the set of all possible outcomes is the total number of ways that 7 cards may be dealt without replacement.

First we need the number of elements in the sample space for the experiment. When drawing 7 cards at random from a population of 100 there are $\binom{100}{7} = 16007560800$ different possible hands.

Now we must determine the number of elements in our event of interest. We want to know the number of ways we can pull 7 unicorn cards. We will divide our population into unicorn cards (25), and non-unicorn cards (75). We want 7 unicorn cards and 0 non-unicorn cards. We can draw 7 unicorn cards $\binom{25}{7} = 480700$ ways from the set of unicorn cards. We can draw 0 non-unicorn $\binom{75}{0} = 1$ way. The draws are independent and we are drawing 7 unicorn **and** 0 non-unicorn cards so number of events of interest is

$$\binom{25}{7}\binom{75}{0} = 480700$$

Now putting it all together

$$P(7 \text{ Unicorns}) = \frac{\binom{25}{7}\binom{75}{0}}{\binom{100}{7}} = \frac{480700}{16007560800} = 0.00003003$$

We can summarize the steps for simulating an experiment as follows:

1. State the problem. Clearly define the random phenomenon you are trying to simulate.
2. State any assumptions you are making for the experiment
3. Assign digits to represent outcomes
4. Simulate a large number of iterations of the experiment and estimate the probability for the event
5. Use simulations to draw conclusions about the experiment

Example 3: In *Harry Potter and the Philosopher's Stone* Ronald Weasley introduces Harry Potter to wizard cards. Every chocolate frog box contains one wizard card. When Harry opens his frog to reveal an Albus Dumbledore card Ron says "I got about 6 of him". Suppose that the company that makes the frogs suggests that about 12% of boxes contain Albus Dumbledore cards. Run a simulation to figure out how many boxes it takes to get 2 Albus Dumbledore cards.

1. Here we would like the simulate opening a box of chocolate frogs until 7 Dumbledore cards are revealed.
2. We are assuming that 70% of boxes contain Dumbledore cards. We are also assuming that each trial is independent.
3. We will assign numbers 0-12 as Dumbledore cards and 13-99 as other cards.
4. Next we will select random digits to represent the number of boxes it takes to get 2 Dumbledore cards. (Generate random numbers on a TI-84 by using Math→Prob→randint)

   Simulation One: 30 35 36 85 59 76 80 82 25 99 26 73 72 71 61 9 96 3 63 11

   Simulation Two: 53 43 5 34 33 63 7

   Simulation Three: 89 31 61 9 81 83 23 41 81 97 14 11

5. It took 18 boxes in the first simulation, 7 boxes in the second simulation, and 12 boxes in the third simulation. Using 3 simulations we would estimate the average number of boxes you need to open to get a Dumbledore card to be 12.33333 boxes.

   Running this same simulation 10000 times using R, and got an average of 16.6325 boxes.

Example 4: Human resource data suggests that Jim pranks Dwight on 83% of work-breaks. Suppose the staff at Dunder Mifflin Paper Co. takes two breaks a day. Using Simulation, estimate the chance that during breaks Jim will prank Dwight 5 or more times in a five day week.

1. Here we are interested in estimating the probability of Dwight getting pranked 5 times or more on a 5 day week. This includes 10 breaks.
2. We are assuming that the probability of Dwight getting pranked is 83% for each break and that each break is independent of each other
3. As there is an 83% chance of Dwight getting pranked we will assign 0-82 to breaks where he is pranked and 83-99 to breaks where he is not pranked
4. Now we simulate random numbers

   Simulation One: 99 **16 42 12 73 29 44 9 36** 87

   Simulation Two: 95 **33 43** 84 **48 26** 89 **40** 98 **2**

   Simulation Three: 40 18 36 99 **44 3 70 57 16 21**

5. In simulation one, two, and three Dwight is pranked 8, 6, and 9 days respectively. We estimate the probability that Dwight is pranked five or more days to be 1.0 as $\frac{3}{3}$ times he was pranked more than five days. We could also estimate the average number of days he is pranked in a week by taking the average of all simulations. Running the simulation 10000 times in R gives an estimated probability of .998.

Note that R script is included for all simulation examples. We will use simulation in upcoming chapters to estimate averages and probabilities for random variables.

2.5 Conditional Probability

Conditional probability considers the probability of a specific event occurring given another event has already occurred. Here are several examples:

- What's the probability someone is taller than $6'4''$ given they play in the National Basketball League?
- Given someone is a heavy smoker, what is the probability they have lung cancer?
- What is the probability of passing a test, given you studied more than 5 hours?

Notice how in all of these examples there is a condition that restricts the sample space of interest. In the first example you are only considering NBA players, rather than all people. In the second example you are only considering heavy smokers, not all people.

Imagine that there is a town inhabited by 100 people. Forty of those people are male, and sixty of those people are female. Out of those 100 people, 16 have competed in an international math competition.

Suppose you select someone at random from the town to help you with your statistics homework.

Example 1: What is the probability someone is a female given they have competed in a competition?

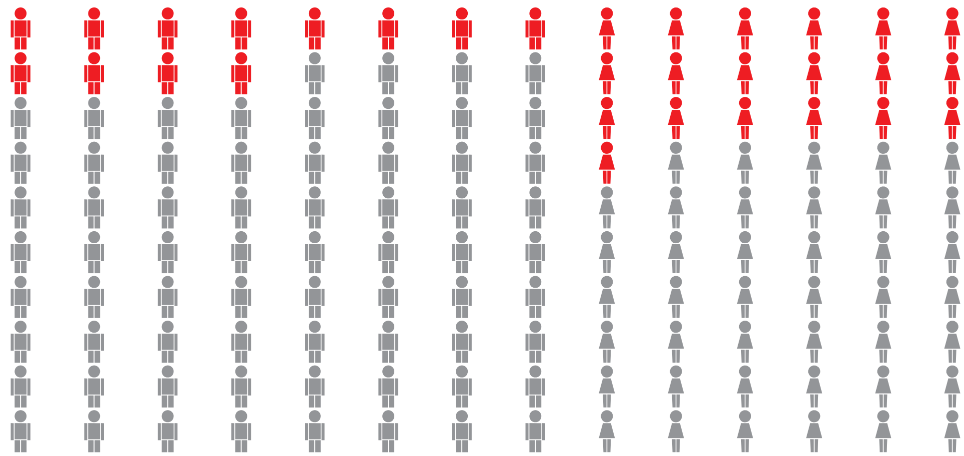
Here we are only interested in the 16 people who have competed in a math competition. Out of these 16 people, 12 people are female so $P(\text{female} \mid \text{competition}) = \frac{12}{16} = \frac{3}{4}$

Example 2: What is the probability of haven competed in a math competition, given the person selected is a male.

Now out of the 40 males we are interested in the 4 who have competed in an international math competition. $P(\text{competition} \mid \text{male}) = \frac{4}{40} = \frac{1}{10}$

You can see in both cases when we condition, we are restricting the total number of outcomes to a smaller subset of outcomes.

Suppose you are in a shopping mall in the late evening when the power goes out. You can see NOTHING. In the mall there are 80 males, and 60 females. What you don't know is 31 of the people are zombies!



Suppose you are feeling around in the dark and the bump into someone **at random**.

Example 3: What is the probability that the person is a zombie?

Out of the total 140 people, there are 31 zombies, so $P(\text{Zombie}) = \frac{31}{140}$

Example 4: What's the probability of the person you bump into being a zombie and a female?

Out of the 140 total possible people to bump into there are 19 who are both zombie **and** female so $P(\text{Female} \cap \text{Zombie}) = \frac{19}{140}$

Example 5: Given you bump into a zombie, what is the probability of them being a female?

Out of the 31 possible zombies, 19 are female so $P(\text{Female} \mid \text{Zombie}) = \frac{19}{31}$

Example 6: What's the probability of you bumping into a man that is a zombie?

Out of the 140 possible people 12 are men **and** zombies, so $P(\text{Male} \cap \text{Zombie}) = \frac{12}{140} = \frac{3}{35}$

Example 7: Given you bump into a man, what is the probability they are a zombie?

Now out of the 80 total possible males, 12 are zombies so $P(\text{Zombie} \mid \text{Male}) = \frac{12}{80} = \frac{3}{20}$

Example 8: A Statistician is interested in what sports people are interested in. In sample of 324 people 37 said they prefer hockey and the remainder prefer soccer. Out of the people who prefer hockey, 32 are from Canada. There were a total of 48 people sampled from Canada, and the remainder were from Mexico.

(a) What is the probability someone preferred Hockey given they were from Mexico?

We can summarize the question using a *contingency table.*

	Canada	Mexico	
Hockey	32	5	37
Soccer	16	271	287
	48	276	324

If we are looking at the probability someone prefers hockey given they are from Mexico, we are only interested in the subset of 324-48=276 people from Mexico. Out of this group 5 preferred hockey so

$$P(\text{H}|\text{M}) = \frac{5}{276}$$

(b) What is the probability someone is both from Canada **and** prefers Hockey?
For this questions let's take the previous contingency table and divide every cell by the sample size., 324. This converts each cell from a count to a probability.

	Canada	Mexico	
Hockey	(32/324)	(5/324)	(37/324)
Soccer	(16/324)	(271/324)	(287/324)
	(48/324)	(276/324)	1

It is clear that out of the 324 people 32 prefer hockey **and** are from Canada.

$$P(C \cap H) = (32/324)$$

(c) Given a person prefers soccer, what is the probability they are from Canada?

We can solve this method using the same method as part (a), however, let's instead use the probability contingency table from part (b).

	Canada	Mexico	
Hockey	(32/324)	(5/324)	(37/324)
Soccer	(16/324)	(271/324)	(287/324)
	(48/324)	(276/324)	1

It is clear that out of the 287 people that like soccer, 17 are from Canada. So

$$P(C|S) = \frac{17}{287}$$

Notice we arrive at the same answer using the same ratio of cells in the probability table.

$$P(C|S) = \frac{P(H \cap S)}{P(S)} = \frac{(17/324)}{(287/324)} = \frac{17}{287}$$

This leads to the formula for conditional probability.

Conditional Probability
Conditional probability is noted $P(A|B)$ represents the probability of event A happening, given event B has already happened.

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

Note: the difference between **"Given"** vs. **"And"** (Condition vs. Intersection) is very important. They **do not** mean the same thing.

- $P(A \cap B)$ is the probability of an element belonging to **both** A and B.
- $P(A|B)$ is the probability of an element belonging to A after you restrict your sample space to elements that belong to B
- $P(B|A)$ is the probability of an element belonging to B after you restrict your sample space to elements that belong to A

Independent Events
Two events A and B are said to be **Independent** if the occurrence of one does not effect the other. We have several formulas to test whether or not two events are independent.

First off it is intuitive that events A and B are **independent** if $\mathbf{P(A|B) = P(A)}$ or $\mathbf{P(B|A) = P(B)}$.

Next as $P(A \cap B) = P(A|B) \cdot P(B)$,

$$\begin{aligned} P(A \cap B) &= P(A|B) \cdot P(B) \\ &= P(A) \cdot P(B) \end{aligned}$$

So two events A and B are independent if and only if $\mathbf{P(A \cap B) = P(A) \cdot P(B)}$. If two events are not independent they are said to be **dependant**.

Example 9: If I flip two coins, what is the probability the second coin lands heads if the first one lands tails?

$$P(\text{Second Heads} \mid \text{First Tails}) = \frac{P(\text{Second Heads} \cap \text{First Tails})}{P(\text{First Tails})} = \frac{(1/2)^2}{(1/2)} = (1/2)$$

It is clear that the occurrence of landing tails on the first flip does not effect the probability of landing heads on the second flip so these events are **independent**

As $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$ we can derive **Baye's Theorem** for conditional probability.

Baye's Theorem

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

2.5.1 Contingency Tables and Trees

We have two great strategies for questions involving conditional probability; **tree diagrams** and **contingency tables**. We will start with contingency tables

Contingency Tables

In contingency tables the inner cells represent *intersection sets*. Consider two events A and B

	A	A^c	
B	$\lvert A \cap B \rvert$	$\lvert A^c \cap B \rvert$	$\lvert B \rvert$
B^c	$\lvert A \cap B^c \rvert$	$\lvert A^c \cap B \rvert$	$\lvert B^c \rvert$
	$\lvert A \rvert$	$\lvert A^c \rvert$	$\lvert S \rvert$

Probability is a simple extension.

	A	A^c	
B	$P(A \cap B)$	$P(A^c \cap B)$	$P(B)$
B^c	$P(A \cap B^c)$	$P(A^c \cap B)$	$P(B^c)$
	$P(A)$	$P(A^c)$	$P(S) = 1$

If two events are **independent**, we know $P(A \cap B) = P(A) \cdot P(B)$. So for independent events we have an effective method for finding the intersection between events.

Contingency tables also can be generalized beyond two events. *Partition* the sample space into events $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ and create a contingency table.

	X_1	X_2	$\ldots$	X_n	
Y_1	$P(X_1 \cap Y_1)$	$P(X_2 \cap Y_1)$	$\ldots$	$P(X_n \cap Y_1)$	$P(Y_1)$
Y_2	$P(X_1 \cap Y_2)$	$P(X_2 \cap Y_2)$	$\ldots$	$P(X_n \cap Y_2)$	$P(Y_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Y_m	$P(X_1 \cap Y_m)$	$P(X_2 \cap Y_m)$	$\ldots$	$P(X_n \cap Y_m)$	$P(Y_m)$
	$P(X_1)$	$P(X_2)$	$\ldots$	$P(X_n)$	1

Example 10: A random sample of 124 grade 10, 11, and 12 students from three difference schools are asked what the most important component of school is. The results are summarized in the following table

	School A	School B	School C	
Academics	81	15	32	108
Sports	33	4	51	99
Socialization	9	18	13	49
	123	37	96	256

(a) What is the probability a student thinks academics is the most important and is from School A?

$$P(\text{School A } \cap \text{ Academics}) = \frac{81}{256} \approx 32\%$$

(b) What is the probability a student thinks academics is the most important given the they are from School A?

$$P(\text{Academics} \mid \text{School A}) = \frac{81}{123} \approx 66\%$$

(c) What is the probability a student is from School A given they think academics is the most important?

$$P(\text{Academics} \mid \text{School A}) = \frac{81}{108} = 75\%$$

(d) What is the probability that a student is in School B given they prefer Socialization?

$$P(\text{Academics} \mid \text{School A}) = \frac{18}{49} \approx 37\%$$

(e) What is the probability a Student prefers sports given they are from School C?

$$P(\text{Academics} \mid \text{School A}) = \frac{51}{96} \approx 53\%$$

(f) Are School and importance preference independent? Explain.
Here $P(\text{School A}) = \frac{123}{256}$, $P(\text{Academics}) = \frac{108}{206}$, and $P(\text{School} \cap \text{Academics}) = \frac{81}{256}$

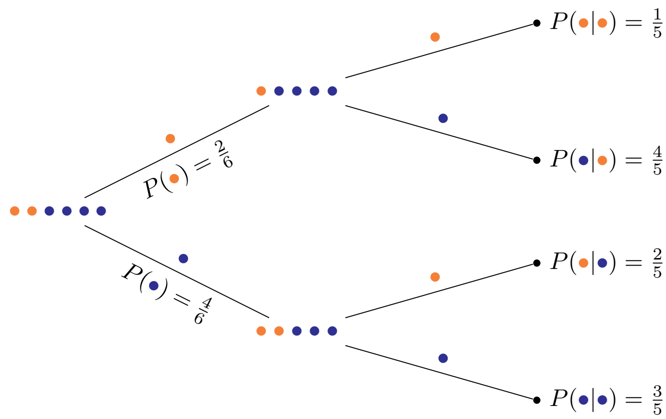
$$P(\text{School} \cap \text{Academics}) \neq P(\text{School A}) \cdot P(\text{Academics})$$

So the events are not independent.

Trees

Let's explore how tree diagrams can be used to visualize conditional probability though a few examples

Example 11: There are 6 marbles in a bag, 2 are orange and 4 are blue. You draw two marbles at random without replacement. We can visualize the experiment using a tree diagram.



(a) Determine $P(\text{Pull Orange Second} \mid \text{Pull Blue First})$.

From our diagram we can see clearly that $P(\text{Orange Second} \mid \text{Blue First}) = \frac{4}{5}$

(b) $P(\text{Blue Second} \mid \text{Blue First})$

From our diagram we can see clearly that $P(\text{Orange Second} \mid \text{Blue First}) = \frac{3}{5}$

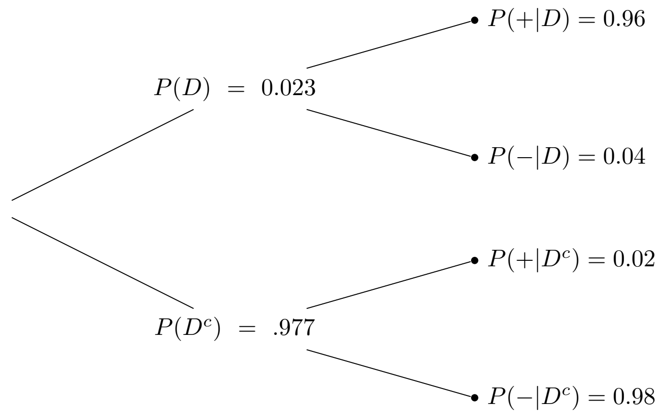
(c) $P(\text{Pull 2 Blue Marbles})$

Since $P(A \cap B) = P(A) \cdot P(B|A)$ we may obtain $P(\text{Blue First } \cap \text{ Blue Second})$ by 'multiplying down the branch'.

$$\begin{aligned} P(\text{Pull Two Blue Marbles}) &= P(\text{Blue Marble First}) \cdot P(\text{Blue Second} \mid \text{Blue First}) \\ &= \frac{4}{6} \cdot \frac{3}{5} = 12/30 \end{aligned}$$

Example 12: Suppose the probability of carrying a certain disease is 0.023. A diagnostic test has been developed to test for the presence of the disease. If an individual has the disease, the test shows "positive" for the disease with a probability of 0.96. However, if the individual does not have the disease, the test shows "positive" for the disease with a probability of 0.02.

(a) Construct a tree diagram for the problem



(b) What is the probability a person has the disease and tests negative?

Using the same method as the marbles problem $P(D \cap -) = P(D) \cdot P(-|D) = (0.023)(0.04)$

(c) What is the probability a person will test negative for the disease?

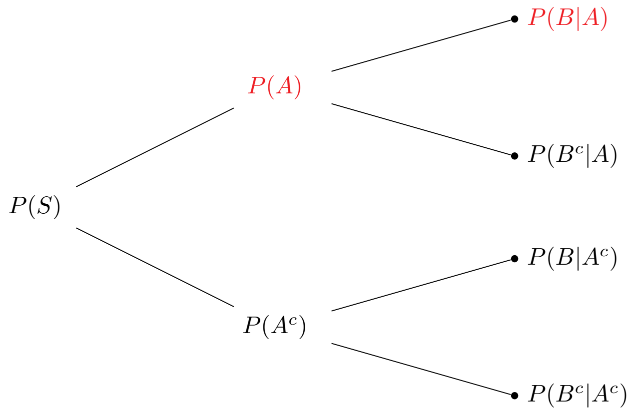
Here we must use the law of total Probability, $P(A) = P(A \cap B) + P(A \cap B^c)$

$$\begin{aligned} P(-) &= P(- \cap D) + P(- \cap D^c) \\ &= (.023)(.04) + (.977)(.98) \\ &= 0.95838 \end{aligned}$$

(d) What is the probability a person will test positive for the disease?

$$\begin{aligned} P(+) &= P(+ \cap D) + P(+ \cap D^c) \\ &= (.023)(.96) + (.977)(.02) \\ &= 0.04162 \end{aligned}$$

We can construct a tree diagram for two events A and B as follows.



Notice that as $P(A \cap B) = P(A|B) \cdot P(B)$. We may get $P(A \cap B)$ by multiplying down the the first branch.

Using the tree we have an easy way of finding $\mathbf{P(A)}$, and $\mathbf{P(B)}$ by 'multiplying down branches'

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

So now when calculating conditional probability.

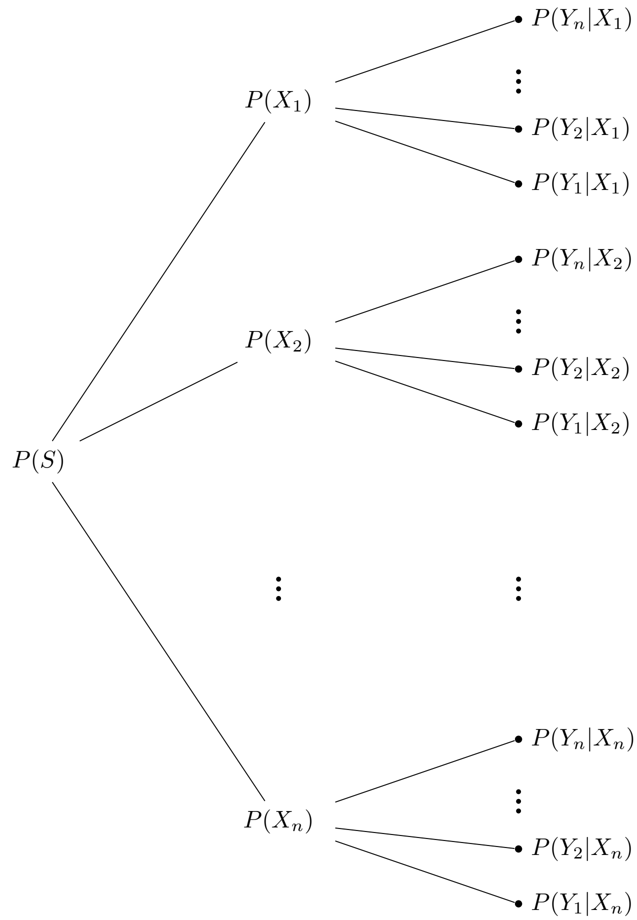
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(B \cap A) + P(B \cap A^c)}$$

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

Also notice that each 'layer' of branches is a partition of S and thus sums to one:

- $P(A^c) + P(A) = 1$
- $P(B|A) + P(B|A^c) = 1$
- $P(B|A^c) + P(B^c|A^c) = 1$

We can also extend this idea to multiple events.



Using the tree we have an easy way of finding $\mathbf{P(X_i)}$, and $\mathbf{P(Y_i)}$

$$P(X_i) = P(X_i \cap Y_1) + P(X_i \cap Y_2) + \cdots + P(X_i \cap Y_m)$$

$$P(Y_j) = P(Y_j \cap Y_1) + P(Y_j \cap Y_2) + \cdots + P(Y_j \cap Y_n)$$

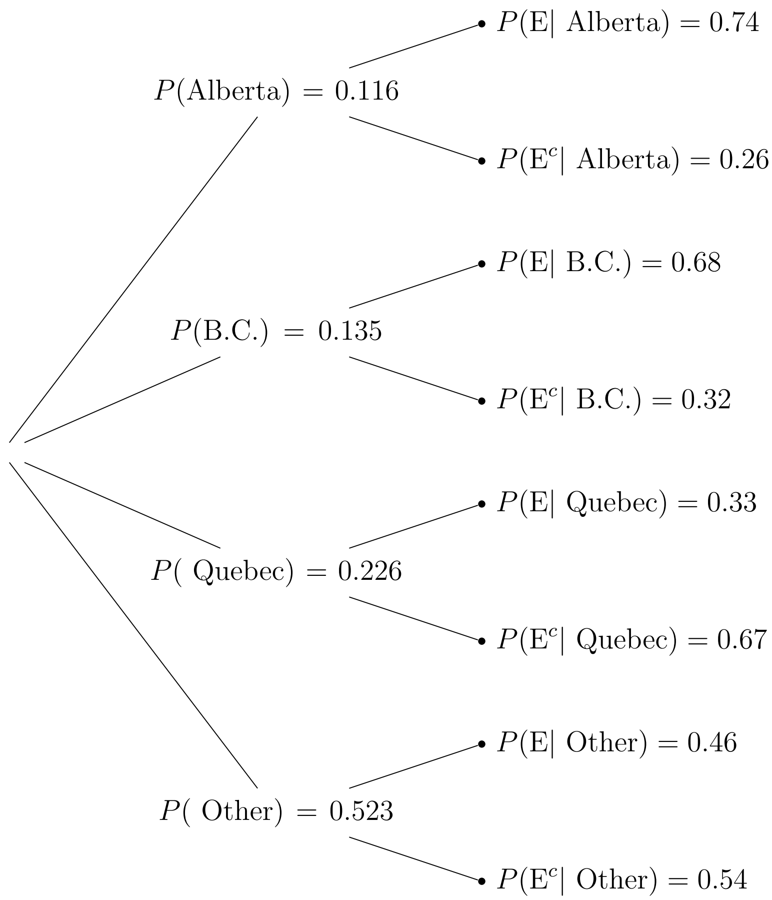
And similar to the simple case

$$P(X_i|Y_j) = \frac{P(X_i \cap Y_i)}{P(Y_i \cap Y_1) + P(Y_i \cap Y_2) + \cdots + P(Y_i \cap Y_n)}$$

$$P(Y_j|X_i) = \frac{P(X_i \cap Y_j)}{P(X_i \cap Y_1) + P(X_i \cap Y_2) + \cdots + P(X_i \cap Y_m)}$$

Example 13: 11.6% of Canadians live in Alberta, 13.5% live in B.C. and 22.6% live in Quebec. 74% of Albertans enjoy alpine skiing, 68% of B.C. residents enjoy alpine skiing, and 33% of Quebec residents enjoy alpine skiing. If a person is not from Alberta, B.C, or Quebec, there is a 54% chance they do not enjoy alpine skiing.

(a) Draw a tree diagram to visualize the probabilities in the problem.



(b) What is the probability a Canadian does not enjoy alpine skiing?

To calculate, we can multiply down each branch of interest to find $P(E^c)$

$$\begin{aligned} P(E^c) &= P(\text{Alberta } \cap \text{E}^c) + P(\text{B.C. } \cap \text{ E}^c) + P(\text{Quebec } \cap \text{ E}^c) + P(\text{Other } \cap \text{ E}^c) \\ &= (0.116)(0.26) + (0.135)(0.32) + (0.226)(0.67) + (0.523)(0.54) \\ &= 0.5072 \approx 51\% \end{aligned}$$

(c) What is the probability someone is from Alberta given the do not enjoy Alpine skiing?

$$\begin{aligned} P(\text{Alberta} \mid \text{E}^c) &= \frac{P(\text{Alberta } \cap \text{ E}^c)}{P(\text{E}^c)} \\ &= \frac{(0.116)(0.26)}{0.5072} \\ &= 0.05946372 \approx 6\% \end{aligned}$$