MAT220 Lecture 6: More on Probability Theory

Contents

In Lecture 5, we introduced the basic vocabulary of probability: simple events, sample spaces, events, complements, intersections, unions, conditional probability and independence. We also saw that classical probability is especially nice because all outcomes are equally likely, so probabilities can be calculated by counting via our Golden Formula. Today we will put this theory into practice by studying some problems related to classical probabilities and the counting techniques of Lecture 4. After that, we will begin moving beyond classical probability, to a general setting in which outcomes are not necessarily equally likely. This leads naturally into probability distributions, and eventually, back around into descriptive statistics.

1 Problems Involving Counting

Today we will spend some time reviewing some exercises in classical probability, and then we will move away from this setting by studying situations in which outcomes are not equally likely. Before doing this, we will first briefly review the theory from last lecture.

1.1 A brief review of the formalism

Throughout Lecture 5, we introduced the basic formalism of probability theory. There are several key notions from our discussion that are worth recalling:

1. An *outcome* is something that can happen as a result of an experiment or operation.
2. The collection of all outcomes is known as the *sample space*. Formally, this is a set S containing all the outcomes ω.
3. An event is loosely defined to be a statement that we can make about our system. Formally, events are defined to be subsets A of the sample space S.
4. Simple events are special events which only contain one outcome, i.e. $A = \{\omega\}$ for some outcome ω in S.
5. The probability of an event, written $P(A)$, is a number between 0 and 1, where 0 represents impossibility, and 1 represents certainty.

Throughout Lecture 5, we worked with *classical probability*, which makes the extra assumption that all outcomes are equally likely, and the sample space S consists of finitely many outcomes in total. If the sample space has size $|S| = n$, then the probability of each outcome is $\frac{1}{n}$. As a matter of fact, we can write this in several equivalent ways. Technically, P is a function that needs events as inputs, *not* outcomes. So, the probability of an outcome happening can be written in terms of simple events as:

$$P(\{\omega\}) = \frac{1}{n}.$$

The central formula from the previous lecture was the Golden Formula for classical probability. We first introduced this naively, and then carefully studied it based on the properties of sets. The Golden Formula is stated below.

The Golden Formula

If the sample space is finite and all outcomes are equally likely, then

$$P(A) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}} = \frac{|A|}{|S|}.$$

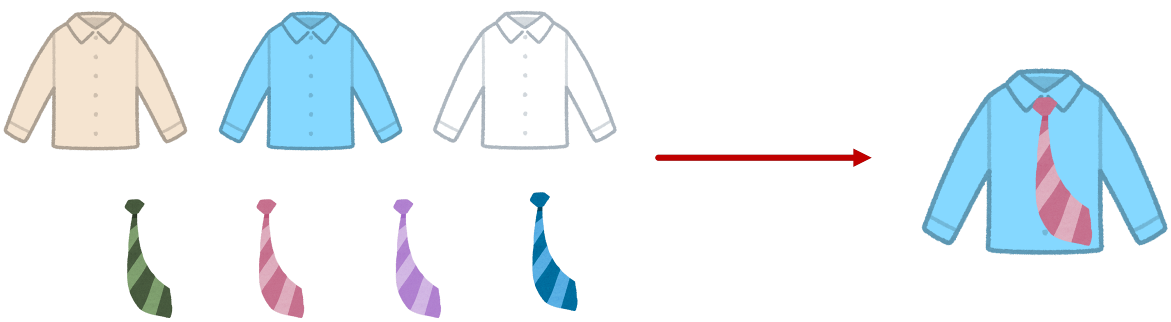
In particular, notice that if we take a simple event $A = \{\omega\}$, then we can apply the Golden Formula to recover the probability $\frac{1}{n}$ for each outcome.

The Golden Formula only works for classical probability, that is, it works when the sample space S is finite and all outcomes in S are equally likely. Often, the main difficulty is not the computation itself, but the act of *counting everything up*. We will now carefully walk through three examples of ways in which we may apply the Golden Formula in different contexts. These are based on the discussion of Lecture 4:

- choices made from several categories,
- arrangements of objects into different orders,
- random selections from a large collection of objects.

1.2 Example 1: picking shirts and ties

Recall our example from Lecture 4: we have a wardrobe with 3 shirts and 4 ties, and we want to pick one of each to make an outfit. For example, one such configuration is pictured below.



Using this set up, we can start picking objects at random, and therefore make our outfit a question of probability. Consider the following three questions.

Exercise 1

Suppose that we pick a shirt and then a tie at random, so that each shirt and each tie has an equal chance of being selected.

(a) What is the probability of choosing the brown shirt?

(b) What is the probability of choosing a white shirt with a red tie?

(c) What is the probability of choosing an outfit that has blue in it?

In order to answer this question, we will first create a sample space. According to the set up, the outcomes are the full choices of outfits. As we saw in Lecture 4, these can be arranged into a grid:

As you can see, there are $12 = 3 \times 4$ possible outfits. This means that the sample space S of this system has size 12, i.e. $|S| = 12$.

(a) We can answer the first question in a few separate ways. Firstly, we may apply the Golden Formula on an informal level to perform the count. Let A be the event "picking a brown shirt and any tie". Then

$$P(A) = \frac{\text{number of outfits with a brown shirt}}{\text{total number of outfits}} = \frac{4}{12} = \frac{1}{3}.$$

Alternatively, on a practical level we can simply observe that the choice of shirts is entirely separate from the choice of ties, and therefore we can just view the choice of shirts as a probabilistic system in its own right. We can write a new sample space

$$S' = \{\text{white shirt}, \text{blue shirt}, \text{brown shirt}\}.$$

There are three shirts and each is chosen with the same probability, meaning that the chance of picking the brown shirt is simply $\frac{1}{3}$. Selecting a tie afterwards does nothing to the chance of selecting a brown shirt.

(b) For the next question, we are being asked to find the probability for an individual outcome: the white shirt and red tie. We may use the Golden Formula directly to see that

$$P(\text{white shirt with red tie}) = \frac{\text{number of white shirt with red tie outfits}}{\text{total number of outfits}} = \frac{1}{12}.$$

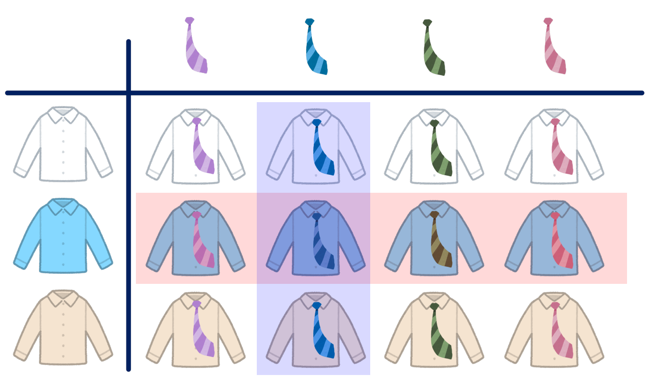
Alternatively, we can define the event A to mean "selecting a white shirt" and event B to mean "selecting a red tie". Then the statement "selecting a white shirt with a red tie" can be reinterpreted as

$$A \text{ and } B: \text{ selecting a white shirt and a red tie.}$$

Since these two events are independent, we can use the multiplication formula from before, to conclude that

$$P(A \text{ and } B) = P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

(c) Finally, we calculate the probability of selecting something blue. Again, there are several ways to approach the solution. Perhaps the simplest method is to examine the sample space and simply highlight all those outfits that happen to have blue in them:



As you can see, there are 6 outfits. Thus, we can naively apply the Golden Formula to conclude that

$$P(\text{something blue}) = \frac{\text{number of outfits with blue in it}}{\text{total number of outfits}} = \frac{6}{12} = \frac{1}{2}.$$

Alternatively, we may define two events

$$B = \{\text{outfits with a blue shirt}\} \quad \text{and} \quad C = \{\text{outfits with a blue tie}\}.$$

The subset B is pictured above in red, and the subset C is pictured above in blue. We note that these two events are *not* mutually exclusive, since there is an outfit consisting of a blue shirt with blue tie. This means that we must subtract this when computing the or statement:

$$P(B \text{ or } C) = P(B) + P(C) - P(B \text{ and } C) = \frac{4}{12} + \frac{3}{12} - \frac{1}{12} = \frac{6}{12} = \frac{1}{2}.$$

Of course, all of these questions can be answered more formally by actually writing out the event in symbols. We can write our system as a collection of ordered pairs corresponding to the choices of shirt/tie and the resulting outfits:

	t_1	t_2	t_3	t_4
s_1	(s_1, t_1)	(s_1, t_2)	(s_1, t_3)	(s_1, t_4)
s_2	(s_2, t_1)	(s_2, t_2)	(s_2, t_3)	(s_2, t_4)
s_3	(s_3, t_1)	(s_3, t_2)	(s_3, t_3)	(s_3, t_4)

According to the diagram above, the brown shirt is third and corresponds to s_3. We can write the sample space as the collection of all of the ordered pairs in the grid above:

$$S = \{(s_1, t_1), (s_1, t_2), \ldots, (s_3, t_4)\},$$

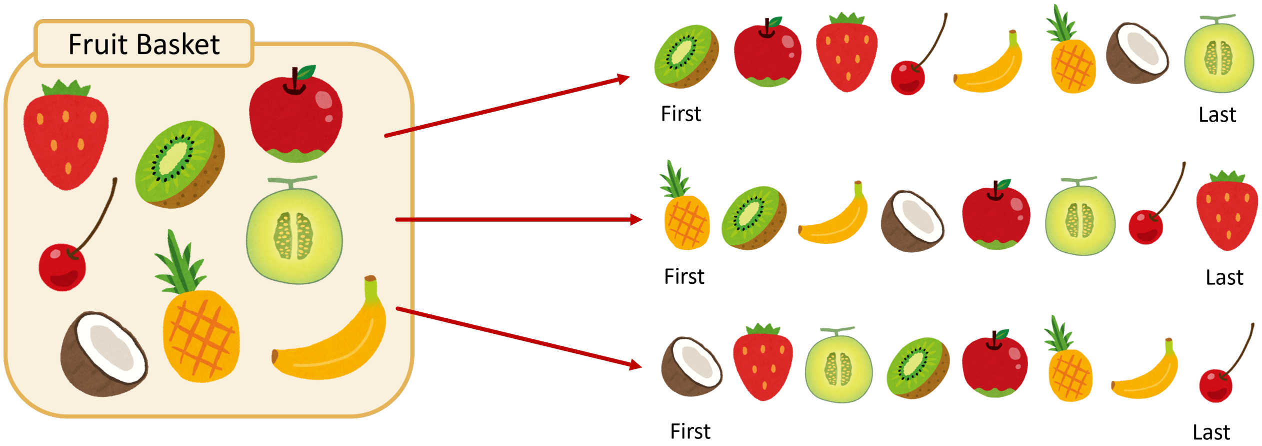
and the event containing all of the outfits with a brown shirt is

$$A = \{(s_3, t_1), (s_3, t_2), (s_3, t_3), (s_3, t_4)\}.$$

Using the formula $P(A) = \frac{|A|}{|S|}$, we recover $P(A) = \frac{1}{3}$ from before.

1.3 Example 2: ordering fruit

Recall from Lecture 4 the example of the fruit basket. There, we had 8 distinct fruit arranged in a basket and we selected them one-by-one. As we saw, there were $8! = 40320$ different ways to do this, three of which are pictured below:



Exercise 2

Suppose that we randomly pick out the fruit in order. In other words, suppose that we randomly select a permutation of the 8 fruit in the basket.

(a) What is the probability of choosing a kiwi first?

(b) What is the probability of selecting an ordering which has the strawberry third?

(c) What is the probability of choosing an ordering which has a melon in the second slot and a cherry in the third slot?

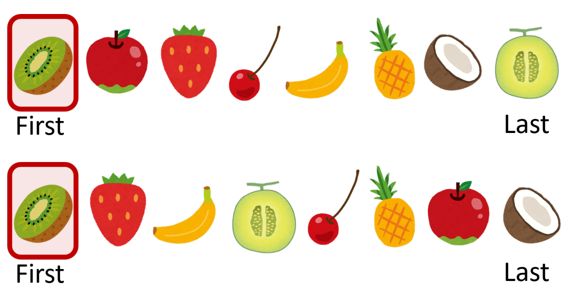
(a) We will argue this in two separate ways. Firstly, perhaps the easiest method is to simply observe: for the first position in the ordering, each of the 8 fruit in the basket has an equal probability of being selected. There are 8 fruit, and only one is a kiwi, therefore the probability of selecting the kiwi first ought to be

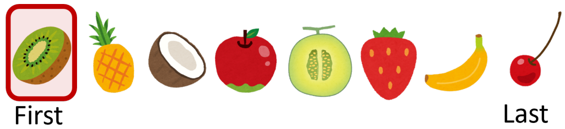
$$P(\text{kiwi first}) = \frac{\text{number of kiwis}}{\text{total number of fruit}} = \frac{1}{8}.$$

This can be confirmed if we instead answer the question on the level of full permutations. As the question states, each ordering has an equally likely chance, so we may equally calculate the probability of selecting a kiwi first by taking the ratio

$$P(\text{kiwi first}) = \frac{\text{number of permutations that have a kiwi first}}{\text{total number of permutations}}.$$

For the denominator, we saw in Lecture 4 that there are $8! = 40320$ unique ways to order the 8 pieces of fruit. Regarding the numerator above, we may calculate this by simply fixing the kiwi in place. For example, below are three of the possible orderings with the kiwi fixed:

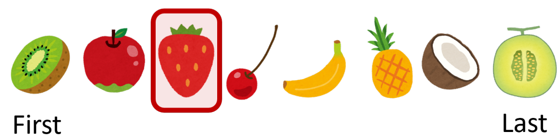


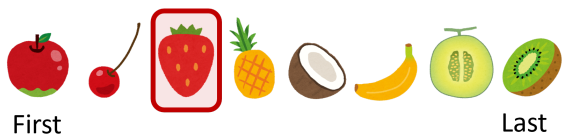


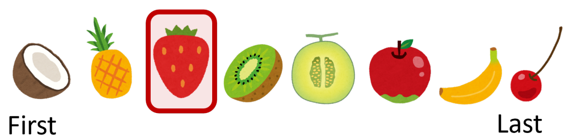
To calculate the total number of ways to make these orderings, we notice that there are 7 pieces of fruit that need to be ordered and there are 7 slots in which to order them. So, the number of total possible orderings is 7!, which cancels down to return the same probability we calculated earlier:

$$\begin{aligned} P(\text{kiwi first}) &= \frac{\text{number of permutations that have a kiwi first}}{\text{total number of permutations}} \\ &= \frac{7!}{8!} = \frac{\cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{8 \times \cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}} = \frac{1}{8}. \end{aligned}$$

(b) By the same logic, we can fix the strawberry in place and permute everything around it. For example, three such orderings could be:



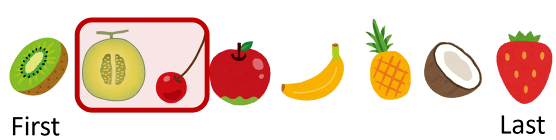


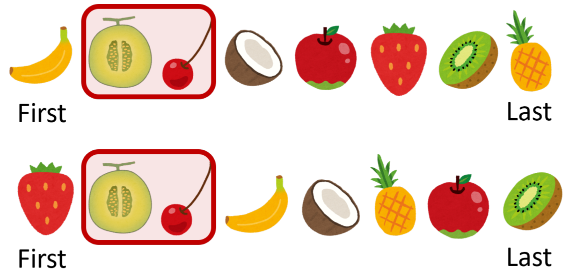


As with the previous question, there are 7 fruits that need to be ordered around the strawberry, and there are 7! ways to do this. The probability again reduces to

$$\begin{aligned} P(\text{strawberry third}) &= \frac{\text{number of permutations that have a strawberry third}}{\text{total number of permutations}} \\ &= \frac{7!}{8!} = \frac{\cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{8 \times \cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}} = \frac{1}{8}. \end{aligned}$$

(c) Finally, we are being asked to find the probability of picking a permutation with the melon in the second position and the cherry in the third. We will make two different arguments. Firstly, we can operate on a pure count as before. However, in contrast to our first two questions, here we have *two* slots already occupied. For example, some permutations may look like:





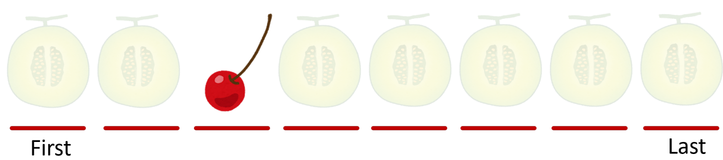
In this case, we are effectively being asked to count the number of ways to reorder the 6 fruits around the melon and cherry. Since there are 6 fruits, the total number of orderings will be 6!, and therefore our probability is

$$\begin{aligned} P(\text{melon second, cherry third}) &= \frac{\text{number of permutations with melon second, cherry third}}{\text{total number of permutations}} \\ &= \frac{6!}{8!} = \frac{\cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{8 \times 7 \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}} = \frac{1}{8 \times 7} = \frac{1}{56}. \end{aligned}$$

Alternatively, we can interpret our question as a compound event. If we let

$$A = \{\text{melon second}\}, \qquad B = \{\text{cherry third}\},$$

then our desired probability is $P(A \text{ and } B)$. According to our formula for "and" statements, we need to calculate the conditional probability $P(A|B)$. Our discussion in parts (a) and (b) tells us that the probability of finding a cherry in the third slot is $\frac{1}{8}$, since there are 8 potential slots and each is equally likely to contain the cherry. In contrast, to calculate the conditional probability $P(A|B)$, we need to calculate the chance of the melon being picked second, given that we *know* that the cherry will land in the third slot. Since we know the cherry will be third, this reduces the number of available slots from 8 to 7:



There is an equal probability for the melon to land in each of these positions. In particular, we conclude that the chance of finding the melon in the second position is $\frac{1}{7}$. Putting this all together, we can calculate the compound probability by multiplying this conditional probability with $P(B)$:

$$P(A \text{ and } B) = P(A|B) \cdot P(B) = \frac{1}{7} \cdot \frac{1}{8} = \frac{1}{56}.$$

1.4 Example 3: a royal flush

There are several different ways to play poker. One simple version is called *Five-Card Draw*, where, as the name may suggest, each player gets given 5 cards in their hand. The players are free to rearrange these cards in their hand as they like, so the ordering does not matter at all. According to the rules, the best possible hand is the "Royal Flush", which consists of the top 5 cards: 10, Jack, Queen, King and Ace, all in the same suit.

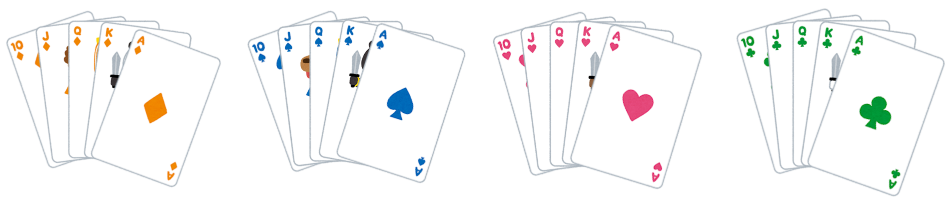
Exercise 3

What is the probability that a player receives a royal flush?

To answer this question, we will use the Golden Formula together with some combinatorics. The probability ought to be

$$P(\text{royal flush}) = \frac{\text{number of possible royal flushes}}{\text{number of possible hands}}.$$

Fortunately, the numerator is very easy to calculate. There are 4 suits, therefore there are 4 possible royal flushes:



To calculate the denominator, we recall that a standard deck of cards has 52 cards. According to the rules of Five-Card Draw, an individual poker hand consists of 5 cards and the order of the cards in the player's hand does not matter. This means that we are looking to count the number of ways that we can select 5 cards from 52, without caring about their order. To do so, we use the $C_{n,r}$ formula from Lecture 4, where here $n = 52$ and $r = 5$:

$$C_{52,5} = \frac{52!}{5!47!} = 2,598,960.$$

With this in hand (pun intentional), we can calculate the probability to be

$$P(\text{royal flush}) = \frac{\text{number of possible royal flushes}}{\text{number of possible hands}} = \frac{4}{2,598,960} = \frac{1}{649,740}.$$

So, the chance of getting a royal flush is 1 in $649,740$. This example is a good reminder that the Golden Formula itself is very simple. The difficult part is often the denominator: before we can calculate the probability, we need to understand the structure of the sample space well enough to count it.

2 Beyond Classical Probability

2.1 The hidden assumption in classical probability

So far, almost every probability calculation we have done has used the Golden Formula:

$$P(A) = \frac{|A|}{|S|}.$$

When applicable, this formula is extremely useful. But, its use is limited, because it assumes that the n outcomes all carry the same probability of occurring (namely $\frac{1}{n}$). This assumption is often justified because of some symmetry in the system we are working with: a fair coin has two opposite sides, a 6-sided dice is a cube, and so on. However, in reality many systems aren't perfectly symmetric like this. In these cases, we can still discuss questions of probability, but we have to relax our assumption that every outcome has the same chance of happening.

For example, suppose that a coin has an uneven probability of landing on either heads or tails, say:

$$P(H) = 0.4, \qquad P(T) = 0.6.$$

Here $P(H)$ means $P(\{H\})$, and similarly for T. Here the sample space is still $S = \{H, T\}$, however the probabilities are no longer equal at 50%. Put differently, the two outcomes are no longer equally likely. Notice here that we cannot use our simple Golden Formula to determine the probability of events:

$$P(\text{landing on heads}) \neq \frac{\text{number of heads in } S}{\text{size of } S} = \frac{1}{2}.$$

This is just a simple example, but it motivates the need for a more general treatment of probabilities. We will now discuss this in detail. Throughout this section we will assume that all of the rules of probability theory still apply, however, we will *not* suppose that every outcome is equally likely.

2.2 Assigning probabilities to simple events

Let S be a finite sample space, say:

$$S = \{\omega_1, \omega_2, \ldots, \omega_n\}.$$

Here the ω_i are the n unique outcomes of the system. Instead of assuming that each ω_i has a fixed probability of $\frac{1}{n}$, we will relax this assumption and instead simply write the probability of each individual outcome as:

$$p(\omega_i),$$

which is just a shorthand notation for $P(\{\omega_i\})$.

In order to comply with the other rules of probability theory, we still need to require that:

1. each probability $p(\omega_i)$ is a number between 0 and 1, and
2. the sum of all their probabilities equals 1:

$$\sum_{i=1}^{n} p(\omega_i) = p(\omega_1) + p(\omega_2) + \cdots + p(\omega_n) = 1.$$

As you can verify, our example of the unfair coin satisfies these properties. Using this setup, we can create a new "Golden Formula" for this more general setting.

Finite probability model

Let $S = \{\omega_1, \omega_2, \dots, \omega_n\}$ be a finite sample space of size n. A probability model assigns a number $p(\omega_i)$ to each outcome ω_i, such that

$$0 \leq p(\omega_i) \leq 1$$

and

$$\sum_{\omega \in S} p(\omega) = 1.$$

Then, for any event $A \subseteq S$, we define

$$P(A) = \sum_{\omega \in A} p(\omega).$$

This last formula is the general replacement for the Golden Formula. It says that in order to calculate the probability of an event A, we have to look at all of the outcomes ω contained in A and simply add up their probabilities.

As a matter of fact, our Golden Formula for classical probability is simply an *instance* of this more general idea. To see this, suppose that we have an event A containing say m outcomes, i.e.

$$A = \{\omega_1, ..., \omega_m\}.$$

If we assume for a moment that each ω_i carries the same probability of $\frac{1}{n}$ where n is the size of the sample space S, then some simple algebra reveals:

$$P(A) = \sum_{\omega \in A} p(\omega) = p(\omega_1) + ... + p(\omega_m) = \underbrace{\frac{1}{n} + \cdots + \frac{1}{n}}_{m \text{ many}} = \frac{1 + ... + 1}{n} = \frac{m}{n} = \frac{|A|}{|S|}.$$

So, we see that our Golden Formula for classical probability has been swallowed into a larger scheme.

2.3 The familiar formulas still work

Even though we have moved beyond classical probability, the main probability formulas from Lecture 5 still work. For completeness, they are stated again below.

Probability formulas

For events A and B:

$$P(\text{not } A) = 1 - P(A).$$

For and-statements:

$$P(A \text{ and } B) = \begin{cases} P(A) \cdot P(B), & \text{if } A \text{ and } B \text{ are independent,} \\ P(A|B) \cdot P(B), & \text{if } A \text{ and } B \text{ are dependent (and } P(B) \neq 0). \end{cases}$$

For or-statements:

$$P(A \text{ or } B) = \begin{cases} P(A) + P(B), & \text{if } A \text{ and } B \text{ are mutually exclusive,} \\ P(A) + P(B) - P(A \text{ and } B), & \text{in general.} \end{cases}$$

Let us briefly see why this is reasonable.

For complements, the sample space splits into two non-overlapping pieces:

$$S = A \cup A^c.$$

Everything either belongs to A, or does not belong to A. Therefore

$$P(A) + P(A^c) = 1,$$

so

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

For or-statements, the problem is overlap. If we add $P(A)$ and $P(B)$, then everything in $A \cap B$ has been counted twice. Therefore we subtract one copy:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

For and-statements, the general rule comes from conditional probability. If $P(A) \neq 0$, then

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}.$$

Rearranging gives

$$P(A \text{ and } B) = P(A) \cdot P(B|A).$$

If A and B are independent, then knowing that A happened does not change the probability of B. In that case,

$$P(B|A) = P(B),$$

so

$$P(A \text{ and } B) = P(A) \cdot P(B).$$

2.4 Example: an unfair dice

Suppose that we have an unfair dice with the following probabilities:

$$p(1) = 0.1, \quad p(2) = 0.1, \quad p(3) = 0.1, \quad p(4) = 0.1, \quad p(5) = 0.5, \quad p(6) = 0.1.$$

This is not a fair dice, since the probability of landing on a 5 is much higher than each of the other outcomes. However, it is still a perfectly valid probability model, because each individual probability is between 0 and 1, and their overall sum equals 1 exactly:

$$0.1 + 0.1 + 0.1 + 0.1 + 0.5 + 0.1 = 1.$$

We will now use our new formula to calculate some probabilities associated to this system.

Exercise 4

Using the unfair dice above, calculate the following:

(a) What is the probability of landing on a number different from 5?

(b) What is the probability of landing on an even number?

(c) What is the probability of landing on a number less than 3?

Solution

(a) Let A be the event "land on 5". Then the event "land on a number different from 5" is the complement of A. Therefore

$$P(\text{not } 5) = 1 - p(5) = 1 - 0.5 = 0.5.$$

(b) The even outcomes are 2, 4 and 6. Therefore

$$P(\text{even}) = p(2) + p(4) + p(6) = 0.1 + 0.1 + 0.1 = 0.3.$$

(c) The outcomes less than 3 are 1 and 2. Therefore

$$P(\text{less than } 3) = p(1) + p(2) = 0.1 + 0.1 = 0.2.$$

2.5 Example: an unfair coin

Suppose that we return to our unfair coin, where:

$$P(H) = 0.4, \qquad \text{and} \qquad P(T) = 0.6.$$

Exercise 5

Suppose that we flip this unfair coin independently several times in a row.

(a) What is the probability of getting two heads in a row?

(b) What is the probability of getting one head and one tail, in any order?

(c) What is the probability of getting three tails in a row?

Solution

(a) Let A be the event "heads on the first flip" and B be the event "heads on the second flip". The flips are independent, so we may simply multiply to yield:

$$P(HH) = P(H) \cdot P(H) = 0.4 \times 0.4 = 0.16.$$

(b) To get one head and one tail in any order, there are two possible sequences:

$$HT \qquad \text{or} \qquad TH.$$

These are mutually exclusive, so we add their probabilities:

$$P(HT \text{ or } TH) = P(HT) + P(TH).$$

Since the coin flips are independent,

$$P(HT) = 0.4 \times 0.6 = 0.24,$$

and

$$P(TH) = 0.6 \times 0.4 = 0.24.$$

Therefore

$$P(HT \text{ or } TH) = 0.24 + 0.24 = 0.48.$$

(c) The probability of getting three tails in a row is

$$P(TTT) = P(T) \cdot P(T) \cdot P(T) = 0.6 \times 0.6 \times 0.6 = 0.216.$$

As an extension of part (b), we can write down the full sample space for two flips:

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

The four probabilities are obtained by multiplication:

Outcome	HH	HT	TH	TT
Probability	0.16	0.24	0.24	0.36

These are unequal, and generally the probability has shifted towards TT and away from HH. To confirm, these probabilities still add up to 1 as they should:

$$0.16 + 0.24 + 0.24 + 0.36 = 1.$$

3 Discrete Probability Distributions

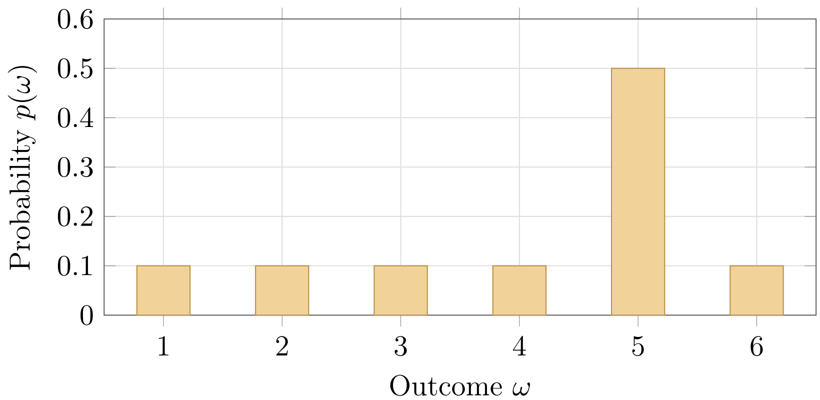
3.1 From probability weights to distributions

In the unfair dice example of Section 2.4, we assigned probabilities to each outcome:

$$p(1) = 0.1, \quad p(2) = 0.1, \quad p(3) = 0.1, \quad p(4) = 0.1, \quad p(5) = 0.5, \quad p(6) = 0.1.$$

We are now going to make a conceptual leap: this collection of information can be organised like data, meaning that we can write it out in a table and plot it as a histogram:

ω	$p(\omega)$
1	0.1
2	0.1
3	0.1
4	0.1
5	0.5
6	0.1



This is very reminiscent of our discussion in Lecture 2. However, now we are carrying some probabilistic context: the heights of the bars are not pure counts, but *probabilities*. The histogram, therefore, tells us how the probability of our entire system is distributed across the individual outcomes. Thus we have discovered *probability distributions*.

Probability distribution

A **probability distribution** tells us how probability is spread across the possible outcomes or values of a random experiment.

We will now briefly touch on the nature of the horizontal axis of a probability distribution.

3.2 Random variables

In the above example, the x axis of the histogram is labelled by the sample space – the numbers from 1 to 6. In usual mathematics, we might label this axis with a *variable*, which is a mystery letter x that is allowed to take on values on the real number line. Ordinary graphs are then described by sliding the x variable along the number line and computing its various outputs. In the case of our probability distributions, we are working with a situation in which we do not simply "slide" our variable around, but instead, we allow our variable to take some value in the sample space *randomly*. For that reason, we interpret the x-axis as showing the possible values of a random variable. We often write these random variables with capital letters, like X.

We will discuss random variables at length in later lectures – especially in our treatment of the Binomial Distribution in Lecture 8. For now, it is enough to understand the following definition.

Definition: random variable

A **random variable** is a rule that assigns a numerical value to each outcome of a random experiment.

The idea here is that a given sample space might not consist of quantitative data that we can plot on a histogram. So, random variables give us a way to turn the outcomes into numbers, so that we

can use descriptive statistics on the resulting probability distributions.
As an example, suppose that we have a fair coin, and we flip it twice. The sample space is qualitative:

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

and the resulting probabilities are

$$P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}.$$

From this information we can define a random variable X by taking a count, for example:

$$X = \text{the number of heads.}$$

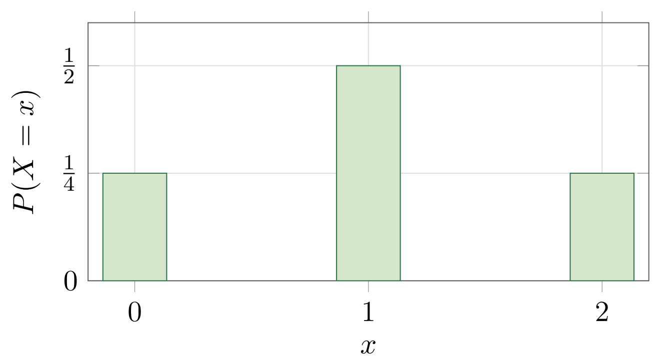
This allows us to describe the same system using numbers:

$$X(HH) = 2, \qquad X(HT) = 1, \qquad X(TH) = 1, \qquad X(TT) = 0.$$

and from here we can re-evaluate the probabilities to:

$$P(X = 0) = P(TT) = \frac{1}{4}, \quad P(X = 1) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \quad P(X = 2) = P(HH) = \frac{1}{4}.$$

From this information, we obtain a genuine probability distribution:



3.3 Discrete probability distributions

Recall that ordinary numbers, like 3, may have either a discrete or a continuous interpretation: we may think of 3 as counting up some quantity of things, or we may think of 3 as describing the length of a continuous interval. Throughout this lecture we will only deal with *discrete* random variables, since our sample spaces are finite sized.

Definition: discrete probability distribution

If X is a finite discrete random variable with possible values

$$x_1, x_2, \ldots, x_n,$$

then a **discrete probability distribution** assigns a probability to each value:

$$P(X = x_1), \quad P(X = x_2), \quad \ldots, \quad P(X = x_n).$$

These probabilities must satisfy

$$0 \leq P(X = x_i) \leq 1$$

and

$$P(X = x_1) + P(X = x_2) + \cdots + P(X = x_n) = 1.$$

Once we have the distribution, probability questions become addition questions. For example,

$$P(X \geq 1) = P(X = 1) + P(X = 2) = \frac{2}{4} + \frac{1}{4} = \frac{3}{4}.$$

In words, this is the probability of getting at least one head.

Similarly,

$$P(X < 2) = P(X = 0) + P(X = 1) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}.$$

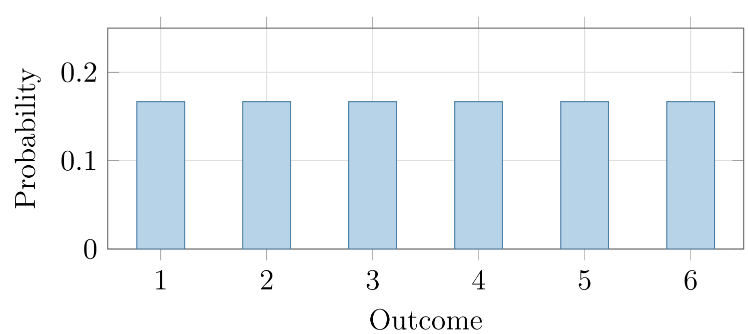
How to use a discrete distribution

To calculate the probability of a condition involving X, add up the probabilities of all values of x that satisfy the condition.
For example,

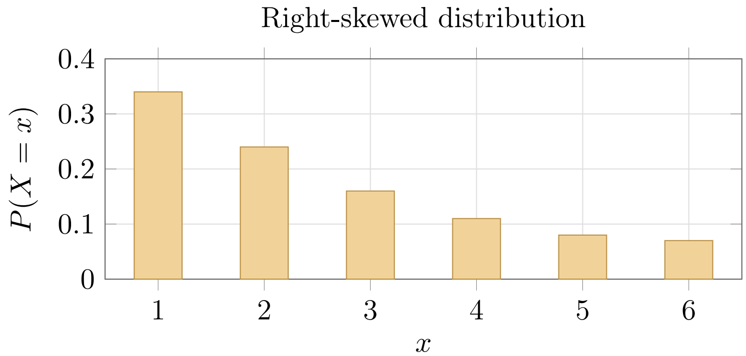
$$P(X \geq 1) = \sum_{x \geq 1} P(X = x).$$

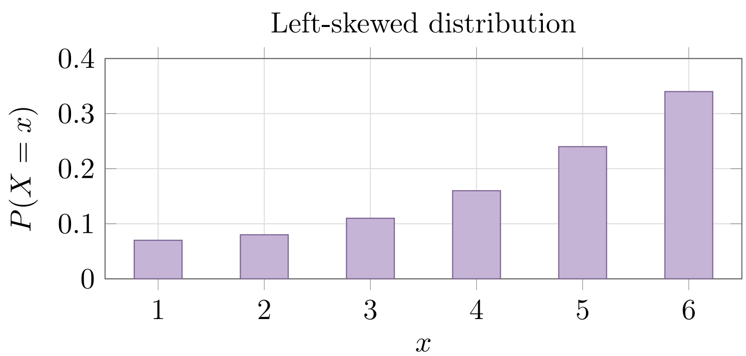
3.4 The shape of a distribution

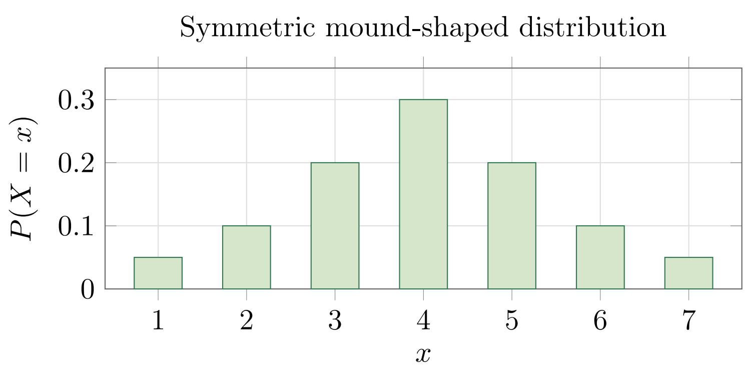
As we saw in Lecture 3, the shape of a histogram can efficiently inform us of the overall structure of a distribution. The same idea applies here with probability distributions. Perhaps the simplest case is that of classical probability. Here each outcome has the same exact probability, and therefore the resulting distribution will be uniform:



However, other systems may admit different shapes. For example, the probability may be skewed to the left or to the right, or even symmetric mound-shaped:







Throughout Lectures 8 and 9 we will talk extensively about discrete probability distributions. As we will see, binomial distributions are a particularly interesting setting for probability.

Waaa! I ran out of time to type up the bit on expected values! I will add this at the weekend.