Math 1228A/B Online

Lecture 1:

Using Sets

(text reference: Section 1.1, pages 1 - 2)

1 Techniques of Counting

Counting seems pretty basic. What we're going to be learning about is how to count the number of ways in which something can be done (or can happen), without having to list and physically count all the different ways. To see that counting can be more complicated than just, for instance, counting heads in the classroom, let's look at a few of the examples we'll be doing in this chapter:

- A class contains 200 students. 130 of the students are from Social Science and the rest are from other faculties. Furthermore, 120 of the 200 students are women, and 50 of the Social Science Students are men. How many of the women are not in Social Science?
- You have decided to have pie and ice cream for dessert. There are 3 kinds of pie available: apple, blueberry and cherry. As well, there are 2 flavours of ice cream available: vanilla and heavenly hash. In how many ways can you choose 1 kind of pie and 1 flavour of ice cream?
- How many subsets of the set $\{0, 1, 2, 3, 4, 5, 6\}$ are there?
- In how many ways can the positions of President, Vice President, Treasurer and Secretary be filled by election of members from a 15-person Board of Directors?

For some of these problems, maybe even most or all of them, you may feel that given the opportunity to look at the problem and think about it, you could probably come up with the answer, even without learning anything in this course. And you're probably right. What we'll be learning is how to approach these problems in an organized way, and how to arrive at the right answer *without* having to make a list of the possibilities in order to count them up.

1.1 Basic Set Theory

This should mostly be review for you, I think. Set notation can be very helpful in organizing the counting we want to do. We start with the following definitions:

A **set** is an unordered collection of objects. We use capital letters to denote sets, for instance the set S or the set A.

We use $x \in S$ to say that object x is in the set S, i.e. that x is an **element** of the set S. Similarly, $y \notin S$ says that y is *not* in set S.

If we have 2 sets, A and B, then $A \subseteq B$ says that set A is a **subset** of set B, i.e. that everything which is in A is also in B. But B might have more things in it, that aren't in A.

If we have $A \subseteq B$ and also $B \subseteq A$, then everything in A is also in B, and everything in B is also in A, so there's nothing in either set which isn't also in the other set. That is, the 2 sets contain exactly the same objects. In that case, we say that $A = B$.

A special set which is often useful is the **empty set** or **null set**. This is the set which contains *no objects*. We denote this set by $\emptyset$.

Usually, we have some set which contains all of the objects we're concerned with in a particular problem, and we want to count the number of objects in some subset of this set. We call this master set the **Universal Set**. That is, the Universal Set U is the set containing *all* objects which could potentially be in any subset that we might be interested in. For instance, if we want to talk about characteristics of students who are taking this course, e.g. how many are men, or how many are in Social Science, or how many are registered to write exams in London, then we would use as the universal set the set containing all of the students registered in the course.

Set Operations

$A \cup B$ (pronounced A **union** B) is the set containing all objects which are in either set A or set B. i.e. $x \in A \cup B$ if $x \in A$, or $x \in B$, or both.

$A \cap B$ (pronounced A **intersect** B) is the set containing those objects which are in *both* set A and set B. i.e. $x \in A \cap B$ if $x \in A$ *and also* $x \in B$.

A^c (pronounced A **complement**, or the complement of A) is the set of all objects (in the universal set, of course) which are *not* in set A. i.e. $x \in A^c$ if $x \in U$ but $x \notin A$.

$n(A)$ (pronounced the **number in** A) counts the number of elements in A. i.e. $n(A) = k$ if A is a set containing k distinct objects.

Example 1.1. *Consider the set of all students in a certain probability class. Let S be the subset of the students who are in Social Science, W be the set containing all women in the class, and A, B, and C be the sets of first year, second year and upper years students, respectively.*

(a) Express each of the following subsets in set notation:

- *(i) the set of women Social Science students in the class;*
- *(ii) the set of all students (in the class) who either are in first year or are in Social Science (or both);*
- *(iii) the set of all students who are neither women nor in second year;*
- *(iv) the set of all students who are not first year women students.*

(b) Give a verbal description of each of the following sets:

- *(i)* $W \cup S$;
- *(ii)* $(A \cap W)^c \cup S$.

(c) Are female first year Social Science students in the set $(A \cap W)^c \cup S$ (from (b)(ii))?

Note: It's often useful to name sets using a letter which helps you remember what objects the set contains, such as is done here in using W to name the set of women and S to name the set of Social Science students.

(a) We translate each verbal description into set notation.

(i) A woman social science student is a student who is a woman and is also a Social Science student, so if x is a woman social science student, then $x \in W$ and also $x \in S$, i.e. we have $x \in W \cap S$. That is, the women Social Science students in the class are the students in the *intersection* of set W and set S, so the subset we are looking for here is $W \cap S$.

(ii) This subset contains all the first year students, (i.e. all of set A), and also contains all the Social Science students (i.e. set S). The subset which contains everything that is in set A as well as everything that is in set S, i.e. all the students who are either in A or in S (some of whom may be in both), is the *union* of the two sets, $A \cup S$.

(iii) The set of all women students is W, so the set of all students who are not women is W^c. Likewise, B is the set of all second year students, so B^c is the set of all students who are *not* in second year.

We want the set of all students who are *neither* in W *nor* in B, i.e. the students who are not in W *and also* are not in B, which is the *intersection* of these 2 complements. That is, the set of students who are neither women nor in second year is $W^c \cap B^c$.

(iv) The set of all first year women students is $W \cap A$. (Just as the set of all women Social Science students is $W \cap S$, in part (a)(i).) Therefore the students who are *not* first year women students are (all of) the students who are *not* in this set, i.e. are in the complement of this set. So the set of all students who are not first year women students is $(W \cap A)^c$.

(b) We must translate each expression in set notation into words.

(i) $W \cup S$ is the set containing all students who are women or who are in Social Science (or both). That is, this set contains all of the women in the class as well as all of the Social Science students in the class.

(ii) The set $A \cap W$ contains all of the first year women in the class, so the set $(A \cap W)^c$ contains all students in the class who are *not* women in first year (as in (a)(iv)). Thus the set $(A \cap W)^c \cup S$ contains all students who are not first year women or who are in Social Science.

(c) The question is "*Are female first year Social Science students in the set* $(A \cap W)^c \cup S$ *(from (b)(ii))?*" Yes. *All* of the Social Science students in the class are in this set. Notice that $(A \cap W)^c$ contains all students in the class except the first year women. When we add (all) the Social Science students into the set (i.e. take the union of $(A \cap W)^c$ with S), we end up with a set containing all of the students in the class *except* the ones who are first year women who are in some faculty other than Social Science.

Venn Diagrams

Often, it is useful to draw pictures of sets. We do this using something called a **Venn diagram**. A Venn diagram consists of a *box*, which corresponds to the universal set, and circles, which correspond to subsets of U. The circles overlap, to show that the subsets (may) have elements in common.

For instance, Figure 1 shows a Venn diagram depicting 2 subsets, A and B. The box contains all of the elements of U. Each circle contains only those elements which are in the corresponding set.

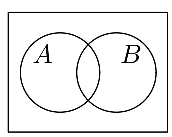


Figure 1: Sets A and B.

We shade in parts of a Venn diagram to show a particular subset. For instance, to show the subset $A \cup B$, we shade all of A and all of B, whereas for $A \cap B$ we shade only the part where the 2 circles overlap. (See Figures 2 and 3.) Likewise, to show the set B^c, we shade in everything that's *not* part of the circle representing set B. (See Figure 4.)

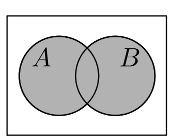


Figure 2: $A \cup B$

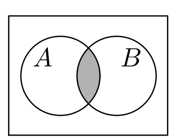


Figure 3: $A \cap B$

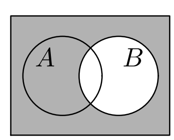


Figure 4: B^c

When there are 3 sets of interest, we use Venn diagrams which have 3 overlapping circles, as shown in Figure 5.

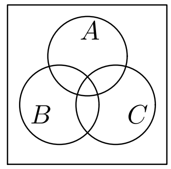


Figure 5: A Venn diagram showing 3 sets, A, B and C.

When you need to figure out what a complex statement involving set unions, intersections and/or complements is saying, or you need to determine how to express a complicated set description in set notation, it is often useful to use a Venn diagram or a series of Venn diagrams. Start with the basics. Remember, whenever you draw a Venn diagram, you *always* need to put the box around it, representing the Universal set, so start with a box. Then draw a circle for each different set name. Now build things up gradually, using as many diagrams as you need. Draw each subset you need to work with (e.g. $A \cap B$, or A^c) separately and then find their union/intersection/complement/whatever and make a clean diagram of that.

To find the union of the sets represented in 2 Venn diagrams involving the same (named) sets, you shade everything that's shaded in either of the diagrams, whereas to find the intersection of the sets represented in 2 Venn diagrams, you shade only the parts that are shaded in both diagrams.

Tip: If you use different colours for the shading, or different shading patterns, then after you draw 2 separate diagrams (or instead of doing so), you can combine them into 1 which shows both shadings. After that, for clarity, you might want to draw another Venn diagram which more clearly shows the set you were trying to find. If it's the union of the 2 diagrams, then it includes everything that's shaded with either kind of shading, whereas if it's the intersection of the 2 diagrams, it will include only those areas which are shaded with *both* kinds of shading.

We will see more examples of how this works in the next lecture, to demonstrate certain properties of the set operations we have been talking about.

Math 1228A/B Online

Lecture 2:
Properties of Set Operations

(text reference: Section 1.1, pages 2 - 5)

Properties of Union and Intersection

1. Both $\cup$ and $\cap$ are commutative
   i.e. for any sets A and B, $A \cup B = B \cup A$ and $A \cap B = B \cap A$

2. Both $\cup$ and $\cap$ are associative
   i.e. for any sets A, B and C,
   $A \cup (B \cup C) = (A \cup B) \cup C$ (so no brackets are needed)
   and $A \cap (B \cap C) = (A \cap B) \cap C$ (again, brackets aren't necessary)

3. Each of $\cup$ and $\cap$ is distributive over the other
   i.e. for any sets A, B and C
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ *Note:* Brackets are
   and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *absolutely necessary* here

You can see easily that the first two of these are true, either simply by thinking about it a little bit or by using Venn diagrams. For instance, the commutative properties just say it doesn't matter whether we start with set A or with set B – once we put all elements of both sets into one set (or select only those elements which are in both sets) we end up with the same thing. Likewise, the associative properties are also saying that the order in which we do things doesn't matter, as long as we're only dealing with *one* kind of operator, i.e. $\cup$ or $\cap$, but not both in the same statement.

The results in the third property are less obvious, and are worth looking at in more detail. (Also, this will demonstrate how we sort things out using Venn diagrams.) This property has to do with using both $\cup$ and $\cap$ in the same statement.

To see that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, we draw the 2 subsets separately, to see that they are identical. In order to draw $A \cup (B \cap C)$, we draw Venn diagrams of A and $B \cap C$ separately, and then bring the 2 diagrams together to find the union of these sets. Remember, when we take the union of 2 sets using Venn diagrams, we shade *everything* that's shaded in *either* diagram.

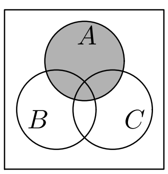


Figure 6(a):
Set A.

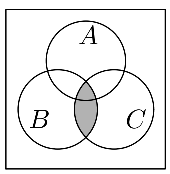


Figure 6(b):
The subset $B \cap C$.

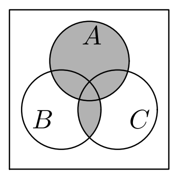


Figure 6(c)
$A \cup (B \cap C)$ is everything which is shaded in 6(a) or 6(b).

Next, we draw $(A \cup B) \cap (A \cup C)$. Again, we draw separate diagrams of $A \cup B$ and $A \cup C$, and then combine them into a single Venn diagram showing their intersection. Remember, when we take an intersection of 2 sets using Venn diagrams, we shade *only* the parts which are shaded in *both* diagrams.

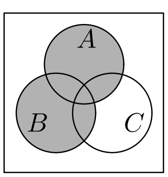


Figure 7(a):
$A \cup B$

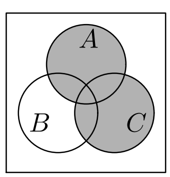


Figure 7(b):
$A \cup C$.

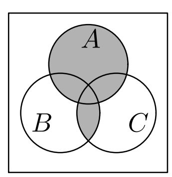


Figure 7(c)
$(A \cup B) \cap (A \cup C)$ is only the parts which are shaded in *both* 7(a) and 7(b).

Comparing Figures 6(c) and 7(c), we see that they are identical, so it is true that the sets they depict are equal.

Showing that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ can be done in a very similar fashion and you should do that ***as an exercise***. You will need to follow the same basic steps that we've followed here.

Properties of Set Complement

1. For any sets A and B, $(A \cap B)^c = A^c \cup B^c$
   That is, the complement of the intersection of 2 sets is the same as the union of their complements. For instance, from Example 1.1 (a)(iv) in Lecture 1 (see pg. 3), the set of all students who are not first year women students contains all the men students and also all the students who aren't in first year. That is, anyone who is either not a woman or not a first year student is not a first year woman student, i.e. anyone either in W^c or in A^c is in $(W \cap A)^c$, so $(W \cap A)^c = W^c \cup A^c$.

2. $(A \cup B)^c = A^c \cap B^c$
   That is, the complement of the union of 2 sets is the intersection of their complements. For instance, from Example 1.1 (a)(iii) in Lecture 1 (see pg. 2), the students who are neither women nor in second year are not women, so they're not in W, and also are not in second year, so they're not in B. But if they're not in W and they're not in B, then they're not in $W \cup B$, so they must be in $(W \cup B)^c$ and we see that $W^c \cap B^c = (W \cup B)^c$.

Note: We can think of these 2 properties as follows: when we distribute the ^{c} through the bracket, the $\cup$ or $\cap$ inside the bracket flips over (to become the other one).

Again, we can show the first of these properties using Venn diagrams. We draw the set $A \cap B$ first, in order to see what its complement is. Then, we draw A^c and B^c in order to see what the their union is.

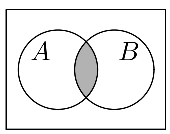


Figure 8(a): $A \cap B$

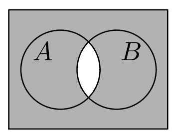


Figure 8(b): $(A \cap B)^c$

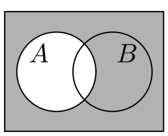


Figure 9(a): A^c

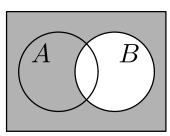


Figure 9(b): B^c

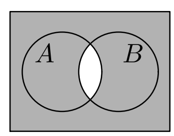


Figure 9(c): $A^c \cup B^c$
(all shaded parts from (a) and (b))

We see that Figures 8(b) and 9(c) are identical.

Exercise: Show that $(A \cup B)^c = A^c \cap B^c$ ***using Venn diagrams.***

Properties of *number in* a set

1. $n(\emptyset) = 0$
   explanation: the number of elements in the empty set (the set with no elements) is 0

2. For any sets A and B,
   $$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$
   explanation: to find the number of elements in the union of sets A and B, count the elements in A and the elements in B and add these numbers together, but then subtract off the number of objects which you counted twice (because they were in both sets), so that they're counted only once.

3. For any subset A of a universal set U,
   $n(A^c) = n(U) - n(A)$
   explanation: we can find the number of objects which are *not* in set A by subtracting the number of objects which *are* in A from the total number of objects in the universal set.

4. *set partitioning:* For any subsets A and B of the same universal set,
   $n(A) = n(A \cap B) + n(A \cap B^c)$
   explanation: the elements of a set A can be partitioned into 2 sets – the ones that are also in some other set B, and the ones that aren't.

5. *partitioning into more parts:* If sets B_1, B_2, ..., B_k together comprise all of the universal set U without overlap, so that $B_1 \cup B_2 \cup ... \cup B_k = U$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$, then

$$n(A) = n(A \cap B_1) + n(A \cap B_2) + ... + n(A \cap B_k)$$

   explanation: if we have a number of sets, B_1, ... B_k, such that each element of the universe is in one and only one of these sets, then we can partition the elements of some other set A according to which of the sets B_1, B_2, ..., B_k they are in.

To understand what the last property is saying, let U be the set of cards in a deck, and let B_1 be the set of all Hearts, B_2 be the set of all Spades, B_3 be the set of all Clubs and B_4 be the set of all Diamonds. Since every card in the deck is in exactly one of these suits, then any subset of the cards, i.e. any $A \subseteq U$, can be thought of as the Hearts that are in A, the Spades that are in A, the Clubs that are in A and the Diamonds that are in A. We can count all of A by counting how many are in each of these disjoint subsets of A. For instance, if A is a subset of the cards containing 3 Hearts, 4 Spades, 2 Clubs and no Diamonds, then A must contain $3+4+2+0 = 9$ cards in total.

(See the document entitled ***Coins, Dice and Cards*** on the *More Useful Stuff* page of the course web site if you're not familiar with a standard North American deck of playing cards. It will be assumed throughout the course that you know about the cards in a deck.)

Let's look at an example of using some of these properties.

Example 1.2. *Suppose that the class in Example 1.1 contains 200 students. 130 of the students are from Social Science and the rest are from other faculties. Furthermore, 120 of the 200 students are women, and 50 of the Social Science students are men. How many of the women are not in Social Science?*

We have the sets U, W and S defined in Example 1.1 (see Lecture 1, pg. 2), and now we are told that $n(U) = 200$, $n(S) = 130$, $n(W) = 120$ and $n(S \cap W^c) = 50$, since the set of all men in the class is the complement of the set of women in the class (i.e. W^c is the set of all men in the class). We are asked to find $n(W \cap S^c)$. Partitioning the Social Science students into women and men, and then rearranging, we have

$$\begin{aligned} n(S) &= n(S \cap W) + n(S \cap W^c) \\ \Rightarrow n(S \cap W) &= n(S) - n(S \cap W^c) = 130 - 50 = 80 \end{aligned}$$

That is, since we know that there are 130 Social Science students, and 50 of them are men, then the other 80 of them must be women. Similarly, we have

$$\begin{aligned} n(W) &= n(W \cap S) + n(W \cap S^c) \\ \Rightarrow n(W \cap S^c) &= n(W) - n(W \cap S) = 120 - 80 = 40 \end{aligned}$$

That is, since 80 of the Social Science students are women, and there are 120 women in total, there must be 40 women who are not in Social Science.

Example 1.3. *Consider again the class in the previous example. If the class contains 48 women who are in first year Social Science and 17 women who are in second year Social Science, how many of the women students in the class are in upper years of a Social Science degree?*

Here, we are dealing only with women Social Science students, and we have already seen that there are 80 of them. That is, we found that $n(W \cap S) = 80$.

Just to make the notation less complicated, let's define a new set, T, the set of women Social Science students in the class. That is, we have $T = W \cap S$ so $n(T) = 80$. The sets A, B and C form a partition of the class, because by definition each of the students must be in exactly 1 of these 3 sets (first year, second year or upper years). Thus we can partition the women Social Science students according to what year they are in. That is, we have

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

with no overlap between these sets, so that

$$n(T) = n(T \cap A) + n(T \cap B) + n(T \cap C)$$

The new information we are given here is that there are 48 of the women Social Science students who are in first year, and 17 who are in second year, so $n(T \cap A) = 48$ and $n(T \cap B) = 17$. We are asked to find the number of women in upper years of a Social Science degree, which is $n(T \cap C)$. We have:

$$\begin{aligned} n(T) &= n(T \cap A) + n(T \cap B) + n(T \cap C) \\ \Rightarrow n(T \cap C) &= n(T) - [n(T \cap A) + n(T \cap B)] \\ &= 80 - (48 + 17) = 80 - 65 = 15 \end{aligned}$$

Therefore there are 15 women in the class who are in upper years of a Social Science degree.

Direct Product

The set operations union and intersection allow us to combine sets which contain similar objects – uniting them in one set or picking out only those objects which are common to both sets. That is, these operations act on sets which are subsets of the same universal set. Often, though, we are interested in ways of combining elements of 2 sets of *different* kinds of objects – not from the same universal set. We combine them into a new kind of set which contains *both* kinds of objects, but still keep them separate.

Definition:

The **Direct Product** of set A and set B, written $A \times B$, is the set of all *pairs* of objects, with the first object in the pair being from set A and the second object in the pair being in set B. That is, every element of the set $A \times B$ is an object of the form (a, b), where $a \in A$ and $b \in B$. The set $A \times B$ contains *all* such pairs.

Theorem: $n(A \times B) = n(A) \cdot n(B)$

Note: The $\times$ in $n(A \times B)$ is the direct product operator, operating on 2 sets. The $\cdot$ in the expression $n(A) \cdot n(B)$ denotes multiplication of 2 numbers. (We could have used $\times$ again, instead of $\cdot$, but it would still be indicating the multiplication of numbers, not the direct product of 2 sets.)

Example 1.4. *You have decided to have pie and ice cream for dessert. There are 3 kinds of pie available: apple, blueberry and cherry. As well, there are 2 flavours of ice cream available: vanilla and heavenly hash. In how many ways can you choose 1 kind of pie and 1 flavour of ice cream?*

Let A, B and C represent Apple, Blueberry and Cherry pie, respectively. Also, let V and H represent the ice cream flavours Vanilla and Heavenly Hash. We can define P to be the set of pies, so that $P = \{A, B, C\}$ and define I to be the set of ice cream flavours, so that $I = \{V, H\}$.

Each choice of a kind of pie and a flavour of ice cream is a pair from the set $P \times I$. For instance, (A, V) is the pair corresponding to having Apple pie with Vanilla ice cream. We are asked to determine how many such pairs there are in total, i.e. in how many ways we can pair a kind of pie with a flavour of ice cream. Since $n(P) = 3$ and $n(I) = 2$, we see that

$$n(P \times I) = n(P) \cdot n(I) = 3 \times 2 = 6$$

That is, there are 6 different combinations of 1 kind of pie and 1 flavour of ice cream available.

Note: We could list the set $A \times B$ as:

$$A \times B = \{(A, V), (A, H), (B, V), (B, H), (C, V), (C, H)\}$$

and count the elements. However, the theorem above allows us to determine the number of elements in the set *without* having to list and then count the elements.

Math 1228A/B Online

Lecture 3:
Counting with Trees

(text reference: Section 1.2)

1.2 Tree Diagrams

A tree diagram is another way of organizing set information. These can be very useful in counting problems. A tree diagram consists of points and edges (or branches). There is usually a sense of direction in a tree diagram, in that there is a starting point, called the **root** of the tree, and there are various **paths** that can be taken along branches of the tree. Each path leads from the root to a different **terminal point**, from which there are no more branches to follow.

One of the ways in which tree diagrams are useful is in representing a sequence of **decisions** which must be made. Each time there is a decision to be made, we put a branch for each of the **choices** available for that decision. Each different decision corresponds to a different level of branching in the tree.

For instance, suppose we have 2 decisions to be made, where the second decision must be made no matter which choice was chosen in the first decision, and the choices available for the second decision do not depend on which choice was made in the first decision.

The root of the tree corresponds to the decision point for the first decision, and has a branch growing from it for each of the choices available in this decision. Then, *each* of these branches on the first level leads to more branches on the second level, with branches corresponding to each of the choices avaiable for the second decision.

When we are finished the branchings for each decision, each terminal point in the tree represents a different combination of choices that could be made. If we want to know *how many* combinations of choices are possible, we just count up the number of terminal points in the tree.

To see how this works, we can use a tree to model the situation we had in Example 1.4 (See Lecture 2, pg. 8.)

Example 1.4 Revisited
There are 2 decisions to be made: choose a kind of pie, and choose a flavour of ice cream. If we consider the pie decision first, there are 3 choices available: apple, blueberry or cherry. So growing from the root, we have 3 branches, corresponding to these choices (labelled A, B and C). This gives us the first level of the tree.

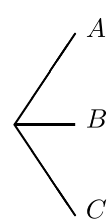


Figure 1: Partial tree (first level)
for Example 1.4

The second decision, flavour of ice cream, gives us a second level in the tree. No matter which kind of pie is chosen, there are 2 flavours of ice cream available: vanilla or heavenly hash. So *each* of the 3 branches on the first level leads to 2 more branches, one for each ice cream flavour.

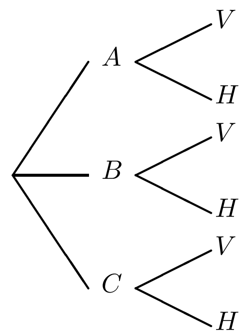


Figure 2: Completed tree for Example 1.4

We see that the completed tree has 6 terminal points (i.e. path ends from which we can go no further), corresponding to the 6 ways of choosing 1 kind of pie and 1 flavour of ice cream. The terminal points in this tree correspond to the elements of the set $P \times I$ from when we previously looked at this problem.

Drawing a tree is not necessarily helpful for *this* kind of problem - we already knew how to find the answer more easily. However, trees can have different kinds of structures. There may be some aspects of a later decision that depend on how an earlier decision was made. For instance, there may be different choices available at level 2, depending on which choice was made at level 1, so that the level 2 branchings are not always the same. Or there might be no need to make a later decision if an earlier decision was made a different way (i.e. no level 2 branching at all for some level 1 branches).

Consider the next example.

Example 1.5. *Fred and Barney plan to play 3 games of darts. However, Fred is a poor sport and always quits in a huff as soon as he loses a game. Draw a tree to determine the number of ways their dart-playing could turn out.*

Notice: When we talk about a *decision*, we don't necessarily mean that someone will actually make a choice. Sometimes, the "decision" will be made by chance. For instance, in this problem, the "decisions" which will be made are of the form "who will win the current game of darts?". Most likely, neither Fred nor Barney will consciously choose who wins. However, we still consider this to be a decision, with 2 choices available - either Fred wins or Barney wins. (We'll assume that ties are not possible.)

Since they plan to play 3 games, there are 3 decisions to be made, and hence 3 levels to the tree. For the first decision, i.e. the first game, there are 2 choices of who wins: Fred (F) or Barney (B).

We are told that, in spite of the plan to play 3 games, if Fred loses he'll quit playing. So if Barney wins the first game, then no more games will be played. That is, in this case there will be no more decisions made, so the branch corresponding to Barney winning leads to a terminal point. We can put an $\times$ there, to remind us that the path stops there.

However, if Fred wins the first game, then they continue with the plan and play a second game. This results in a second decision (*who wins the second game?*) and again there are 2 choices available – Fred or Barney. So the F branch on the first level leads to another branching, again F or B.

Of course, if Fred loses the second game, he will quit in a huff and no third game will be played. So the B branch on the second level also leads to a terminal point. But if Fred wins, they play a third game, corresponding to a third decision, and thus a third branching (F or B again). Both of these branches lead to terminal points, because now they have finished playing 3 games.

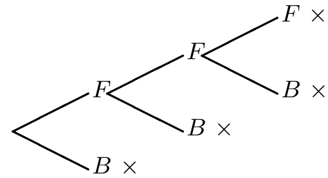


Figure 3: Example 1.5 tree

We see that there are 4 terminal points in the tree (1 on the first level, one on the second and 2 on the third), so there are 4 different ways that their dart playing could turn out.

Counting Trees

Sometimes, it is useful to put numbers on the branches of a tree. If the branches of a tree correspond to subsets of some universal set, then we make a **counting tree** by putting the number of elements in the subset on the corresponding branch.

Note: In this kind of tree, branches growing from some earlier branch correspond to **intersections** of subsets.

To see how this works, consider Example 1.2 again. (See Lecture 2, pg. 7)

Example 1.2 Revisited
We had the universal set corresponding to the students in some class, with $n(U) = 200$. We also had S, the set of Social Science students, with $n(S) = 130$, and W, the set of women students, with $n(W) = 120$. As well, we were told that $n(S \cap W^c) = 50$, and we need to find $n(W \cap S^c)$.

In this kind of tree, we have a level for each characteristic we're interested in. Here, the students in the class either do or don't have the characteristic of being in Social Science, i.e. are either in S or in S^c. We can show this characteristic on the first level of the tree, which gives 2 choices (branches) at the first level: S or S^c. Likewise, each student in the class either is or isn't a woman - so we have a second level for this characteristic, on which there are again 2 choices: W or W^c. Since *all* students in the class may be in either W or W^c, then *both* of the first level branches branch again at the second level.

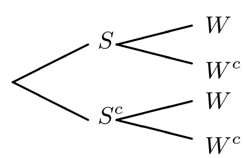


Figure 4: Tree structure for Example 1.2 tree

Notice: In other situations, we could have more than 2 choices at a particular level. For instance, if we were also interested in whether students are in first, second or upper years, we could add another level, with branches for A, B and C growing from each of the second level branches we already have.

The topmost path through the tree leads to a terminal point which corresponds to the set $S \cap W$. That is, this path corresponds to the students who are in S and also are in W. We have four terminal points in this tree, corresponding to (reading down) $S \cap W$, $S \cap W^c$, $S^c \cap W$ and $S^c \cap W^c$.

Think of the students (i.e. the elements of the set U) starting out at the root of the tree and each travelling along the path that describes them. That is, each student takes the branches that correspond to characteristics which they have. So, for instance, a male social science student starts out at the root, chooses the S branch on the first level, and then takes the W^c branch on the second level.

The numbers we put on the branches of the tree correspond to the number of elements (students) who travel along that branch. We have $n(U) = 200$, so we put the 200 students in the class at the root of the tree. At the first level, we are dividing these 200 students into S and S^c, so we put $n(S)$ on the S branch, and $n(S^c)$ on the S^c branch. We know that $n(S) = 130$, so we have $n(S^c) = 200 - 130 = 70$. That is, of the 200 students in the class, 130 are in Social Science, and travel along the S branch, which means that the other $200 - 130 = 70$ students travel along the S^c branch.

Notice: Since, at the second level, we are dividing the Social Science students into women and men, and dividing the non-Social Science students into women and men, there is no single branch in the tree which corresponds to the set W. That is, the second level W branches correspond to $S \cap W$ and $S^c \cap W$, not just W.

We know that $n(S \cap W^c) = 50$. We have a branch for $S \cap W^c$ on the second level, so we put 50 on that branch. That is, there are 50 students in Social Science who are not women, and all of these students (and only these students) will travel along the W^c branch that follows the S branch.

Look at what we have. 130 students went along the S branch, and of these, 50 of them then go along the W^c branch, so the other $130 - 50 = 80$ must go along the W branch. In this way, our counting tree tells us that $n(S \cap W) = 80$, as we found before.

To find the number to put on the $S^c \cap W$ branch, we use the fact that $n(W) = 120$. That is, we know that there are 120 women in the class, so in total there are 120 students who travel along W branches in the tree. Therefore, the numbers on all of the W branches must add up to 120. What we are doing here is recognizing that $n(W) = n(S \cap W) + n(S^c \cap W)$. So the number on the W branch growing from S^c is $120 - 80 = 40$.

Now we can fill in the last number on the tree, because we see that of the 70 students who go along the S^c branch, 40 of them then take the W branch, so the other $70 - 40 = 30$ of them must take the W^c branch. That is, 30 of the non-Social Science students are men.

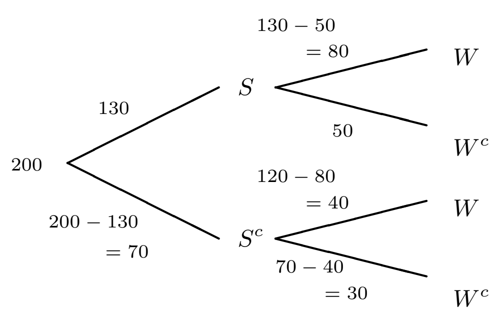


Figure 4: Completed Counting Tree for Example 1.2 tree

Notice: In this example, we were asked to find the number of women who are not in Social Science. This is the number on the $S^c \cap W$ branch, i.e. 40. We had that before filling in the number on the $S^c \cap W^c$ branch, so in this case we could have stopped before the counting tree was complete.

Characteristics of Counting Trees

1. Every element in the universe starts out at the root and ends up at some terminal point (i.e. follows some path through the tree)

2. The number arriving at some intermediate point must equal the number leaving that point. That is, unless it's a terminal point, everyone proceeds to the next level. Thus, the sum of the numbers on the branches leaving any (non-terminal) point equals the number on the branch leading to that point.

3. On each level of the tree, the sum of the numbers on all branches is $n(U)$ minus the number who stopped at terminal points on earlier levels.

4. For each characteristic (i.e. set), the sum of the numbers on all branches labelled with that set must be the number in that set.

   For instance, in the example, we have $n(W \cap S) + n(W \cap S^c) = n(W)$, i.e. the sum of the numbers on all W branches is $n(W)$. *On the tree, we have our idea of* ***set partitioning*** *in action.*

5. For *any* set which can be expressed in terms of the sets represented on the tree, the number in the set can be found by adding up the numbers at all terminal points which correspond to that set. This means that for any combination of characteristics, the number who have that combination of characteristics can be found by summing the numbers on all terminal branches in paths containing branches corresponding to these characteristics.

For example, suppose we have a larger tree, with 3 sets, A, B and C, represented on the first, second and third levels, respectively. Then there is no single branch in the tree representing, for instance, the set $(A \cup B^c) \cap C$. However, we can find the number in this set by adding up the numbers on all of the terminal branches in paths which contain either an A or a B^c branch (or both) and also a C branch. Figure 5 shows a tree like this. The paths which contain an A or a B^c branch as well as a C branch are indicated with arrows. Adding up the numbers on the terminal branches of all these paths, we see that $n\left((A \cup B^c) \cap C\right) = 3 + 2 + 1 = 6$.

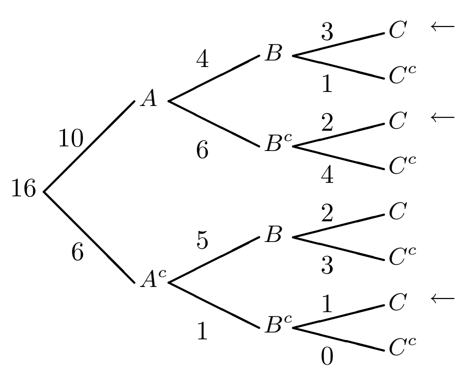


Figure 5: A counting tree with 3 sets, for counting $n\left((A \cup B^c) \cap C\right)$.

Note: If you run into problems with the homework in this section (or even the previous one) and it's because you're having difficulty defining the subsets appropriately, peek ahead to the ***Tip*** at the end of Lecture 12 (pg. 64). Replace the words "sample space" with "universal set" and the word "event" by "subset" and the same principle applies.

Math 1228A/B Online

Lecture 4:
Some Other Counting Tools

(text reference: Section 1.3)

1.3 Fundamental Counting Principle

Theorem *(The FCP for 2 decisions)*

> If 2 decisions are to be made, where there are m choices available for the first decision and n choices available for the second decision, no matter what happens on the first decision, then the number of different ways of making both decisions is $m \cdot n$.

Notice: This isn't actually new to us. If we let A be the set of choices available in the first decision, and B be the set of choices available in the second decision, then this theorem says the same thing as our previous result that

$$n(A \times B) = n(A) \cdot n(B)$$

Example 1.4 Again! (See Lecture 2, pg. 8)
In this problem we had 2 decisions – kind of pie, with 3 choices available (i.e. $m = 3$) and flavour of ice cream, with 2 choices available (i.e. $n = 2$), no matter which pie was chosen. So the Fundamental Counting Principal tells us that the number of ways of choosing pie and ice cream is $m \cdot n = 3 \times 2 = 6$, just as we found before.

We can extend this same result to more than just 2 decisions:

Theorem *(The General FCP)*

> If there are k decisions to be made, where there are m_i choices available for the i^{th} decision, for $i = 1, ..., k$, (no matter what has happened on earlier decisions) then the number of different ways of making all k decisions is $m_1 \cdot m_2 \cdot ... \cdot m_k$ (i.e., multiply all the m_i's together).

Example 1.6. *A certain children's toy consists of a number of pieces. There are red, blue, green and yellow pieces. There are round, square and triangular pieces. Each piece has either a cat or a dog on it. A full set contains 1 piece of each combination of colour, shape and animal. How many pieces are in a full set?*

Here, we have 3 characteristics, which we can think of as decisions. (For instance, if you were going to choose 1 piece, you would need to *decide*: what colour? what shape? which animal?.) There are 4 choices available for the colour decision, 3 choices available for the shape decision, and 2 choices available for the animal decision. That is, we have $k = 3$ decisions, with $m_1 = 4$, $m_2 = 3$ and $m_3 = 2$. So the total number of ways of making all 3 decisions, i.e. the total number of ways of combining these 3 characteristics, which is the total number of pieces in the set, is given by

$$m_1 \cdot m_2 \cdot m_3 = 4 \cdot 3 \cdot 2 = 24$$

Notice: We can extend the idea of direct product in a similar way. For instance, here we had 3 sets, one each for colour, shape and animal: $C = \{\text{red}, \text{blue}, \text{green}, \text{yellow}\}$, $S = \{\text{round}, \text{square}, \text{triangular}\}$ and $A = \{\text{dog}, \text{cat}\}$. We would say that the set $C \times S \times A$ contains ordered triples of the form (colour, shape, animal), and we have

$$n(C \times S \times A) = n(C) \cdot n(S) \cdot n(A) = 4 \cdot 3 \cdot 2 = 24$$

Example 1.7. *How many 4-digit numbers less than 6000 and divisible by 5 can be formed using only odd digits:*

(a) *if repetition is allowed?*

(b) *if repetition is not allowed?*

(a) We are to form 4-digit numbers, using only the digits 1, 3, 5, 7 or 9. This corresponds to making 4 decisions, one for each digit of the number (i.e. 'what will the first digit be?', 'what will the second digit be?', etc.). That is, there are 4 positions in the number we are forming, and we must make a decision for each position, of the form 'what number will go in this position?'.

The number we form must be less than 6000, so it cannot start with a 7 or a 9. Therefore there are 3 choices of a number to put in the first position: 1 or 3 or 5. For both the second and third position in the 4-digit number, we may use any of the odd digits. Therefore there are 5 ways to make the second decision, and also 5 ways to make the third decision. We are told that the 4-digit number we form must be divisible by 5. In general, numbers which are divisible by 5 end in either 5 or 0. But since we can only use odd digits, we can't form any numbers ending in 0. Therefore there is only 1 choice available for the fourth position in the number: 5.

By the FCP, we see that the total number of 4-digit numbers less than 6000 and divisible by 5 which can be formed using only odd digits, with repetition allowed, is

$$3 \times 5 \times 5 \times 1 = 75$$

(b) This time, we are not allowed to use the same digit again in the number. For instance, in part (a) we counted numbers like 5555 and 1715, but this time these numbers must be excluded from our count.

In this situation, it is important that we deal with the most restrictive condition first. That is, in part (a) we could have considered the 4 decisions in any order. But here, we need to know how many of the digits in certain subsets have not yet been used, which complicates things. The least complicated way to deal with this is to consider the various decisions in an order which eliminates, or at least minimizes, the need to consider "but how many (if any) of the elements of this subset are really still available?". The most restrictive condition we have in this particular problem is that the number must be divisible by 5, which means (as we saw before) that the number must end in a 5. That means that 5 *must be* available to go in the fourth position of the number, so it is not available for any of the other positions. That is, we start by considering the fourth position, and observe that there is only 1 choice for which number goes in this position. Once we have filled the fourth position, the 5 is not available for any of the other positions.

The next most restrictive condition is that the number must be less than 6000. This means that, since the 5 is no longer available, the first digit of the number must be either 1 or 3. That is, we next consider the first position decision, and we see that there are only 2 choices available.

When we choose a number to put in the second position, the only restriction is that we cannot repeat a digit we've already used. Of the 5 odd digits, we've already used 2, the 5 and either the 1 or the 3, leaving 3 to choose from. Thus there are 3 ways of choosing a number for the second position.

Finally, the third position may contain any of the odd numbers we have not yet used. We've already used 3 of the 5 odd numbers (for the fourth, first and second positions), so there are 2 choices available for the third position.

In total, the number of 4-digit numbers less than 6000 and divisible by 5 which can be formed using only odd digits, without repetition, is

$$1 \times 2 \times 3 \times 2 = 12$$

Considering different cases

Sometimes, we need to use more than the FCP alone to answer a question. Often, we need to consider *different cases*. When we have two or more different cases which could occur, so that, for instance, we can do something either *this* way or *that* way, the total number of ways in which the thing can be done is found by **adding together** the numbers of ways in which each case can occur. For instance, consider the next example.

Example 1.8. *Recall Example 1.4 (see Lecture 2, pg. 8). Instead of pie and ice cream, you might have a bowl of fruit salad. There are 2 different kinds of fruit salad available: "Melon Madness" or "Tropical Delight". How many different ways can you have dessert now?*

Here, we consider 2 different cases: that you have pie and ice cream, or that you have fruit salad. As we have already seen, there are 6 different pie-and-ice-cream choices available. And there are 2 different choices available if you have fruit salad. Therefore in total there are $6 + 2 = 8$ different dessert choices available to you.

Notice: We could analyze this new situation using a tree. On the first level, the decision is: pie and ice cream ($P\&I$) or fruit salad (F)? Following the $P\&I$ branch, the entire tree from before is replicated. Following the F branch, there is a branching for Melon Madness (M) or Tropical Delight (T).

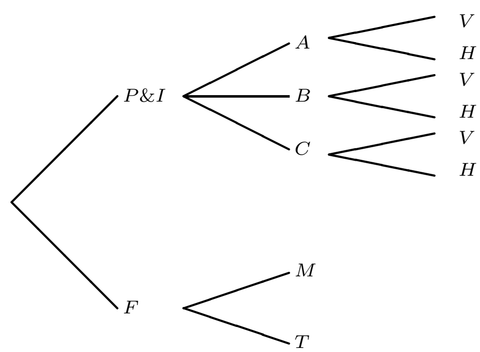


Figure 1: Tree modelling Example 1.9

As before, we can find the total number of ways of making the dessert decision by counting the number of terminal points. Again, we see that there are 8 ways the dessert decision can be made. The branches on the first level of this tree correspond to the 2 distinct cases discussed previously.

Tip: A good way to determine whether you need to be multiplying or adding is to remember that for *and* you multiply, but for *or* you add. That is, if we need to "do this" *and* "do that", the FCP tells us that the total number of ways to "do this and do that" is the number of ways we can "do this" *times* the number of ways we can "do that". However, if we are going to "do this" *or* "do that", the total number of ways of doing "this *or* that" is the number of different ways we could "do this" *plus* the number of different ways we could "do that".

Indirect Counting

Sometimes, it is easier to include in your counting some things which you didn't want to count, and then determine how many of the things you counted were ones you shouldn't have counted. You can then determine how many you *did* want to count by simply subtracting the second number from the first number. This is often true if the FCP can be directly applied to a problem, but includes some things we were supposed to omit, provided that it is easy to count the number of things we didn't want to include.

Consider, for instance, the following problem.

Example 1.9. *Recall Example 1.4. How many choices of dessert are available if you don't have to have both pie and ice cream, but you do have to have dessert?*

Note: This refers to the original problem, i.e. no fruit salads.

We could solve this problem using cases. There are 3 different things that could happen. You might have only pie, or you might have only ice cream, or you might have both. This gives 3 different cases to consider.

<u>Case 1</u>: Pie only
There are 3 different kinds of pie, so there are 3 ways to have pie only for dessert.

<u>Case 2</u>: Ice cream only
There are 2 flavours of ice cream, so there are 2 ways to have only ice cream for dessert.

<u>Case 3</u>: Both pie and ice cream
We have already seen that there are $3 \times 2 = 6$ different ways to have both pie and ice cream for dessert.

Total: There are $3 + 2 + 6 = 11$ different ways to have only pie *or* only ice cream *or* both pie and ice cream for dessert, i.e. 11 ways to have pie and/or ice cream for dessert.

<u>Another Approach:</u> An easier way to do it
It's easy here to find the total number of dessert choices available using the FCP, if we include the possibilities of not having pie and not having ice cream. We simply expand our set of pie choices to include the choice *"no pie"* and the set of ice cream choices to include the choice *"no ice cream"*. As in the original problem (Example 1.4) we're making 2 decisions: a pie decision and an ice cream decision. We now have 4 choices available for the pie decision (apple, blueberry, cherry or none) and 3 choices available for the ice cream decision (vanilla, heavenly hash or none). So there are $4 \times 3 = 12$ ways to make these 2 decisions. However, these 12 combinations include the combination "no pie and no ice cream", i.e. no dessert. Since we were told that this is not an option ('you do have to have dessert'), we didn't want to count this one combination. Therefore we see that there are $12 - 1 = 11$ ways to choose pie and/or ice cream for dessert.

<u>Number of subsets of a set</u>
One application of the FCP is to count how many subsets of a set there are. Consider the following questions.

Example 1.10.

(a) *How many subsets of* $\{0, 1\}$ *are there?*

(b) *How many subsets of* $\{0, 1, 2\}$ *are there?*

(c) *How many subsets of* $\{0, 1, 2, 3, 4, 5, 6\}$ *are there?*

(d) *How many subsets of* $\{0, 1, 2, 3, 4, 5, 6\}$ *are there which include both 1 and 2 but not 5?*

(e) *How many subsets of* $\{0, 1, 2, 3, 4, 5, 6\}$ *are there which include either 0 or 6 but not both?*

(a) This seems easy enough. We can just list them: $\emptyset$, $\{0\}$, $\{1\}$ and $\{0, 1\}$.
As long as we remember that the empty set is a subset of *every* set, and every subset is a subset of itself, i.e. for any set A, $\emptyset \subseteq A$ and also $A \subseteq A$, we can easily see that there are 4 subsets of $\{0, 1\}$.

(b) Now, the set is $\{0, 1, 2\}$.
First approach: same as before - list them and count how many there are.

We get $\emptyset$, $\{0\}$, $\{1\}$, $\{2\}$, $\{0,1\}$, $\{0,2\}$, $\{1,2\}$ and $\{0,1,2\}$, so we see that this set has 8 subsets.

A different approach:
When we form a subset of a set, we can think of making a decision for each element of the set: is this element in the subset? In this case, we have a set containing 3 elements. For each of these, we make a decision in which there are 2 choices available: *yes*, include this element in the subset, or *no*, don't include this element. That is, we have 3 decisions, and each has 2 choices available. Therefore, by the fundamental counting principle, the number of ways to make this sequence of decisions is:

$$2 \times 2 \times 2 = 2^3 = 8$$

This last approach shows us how to figure out how many subsets *any* set has. We simply think of it as making a decision for each of the n elements of the set, where there are 2 choices (*in* or *not in*) available for each of these decisions.

Theorem:

A set containing n elements has 2^n subsets.

(c) Now, consider the set $\{0,1,2,3,4,5,6\}$.
Listing all of the subsets of this set would be onerous. But we can find out how many there are easily, using the theorem. We have a set with $n = 7$ elements, so it has $2^n = 2^7 = 128$ subsets.

As well, the FCP allows us to answer more complicated questions about subsets of a set.

(d) How many subsets of $\{0,1,2,3,4,5,6\}$ are there which include both 1 and 2 but not 5?
We can use an approach similar to the one that gave us the theorem. We have 7 elements in the set, so there are 7 decisions to be made. But some of these have already been made. For 3 of these elements, there is only one choice available, either it must be in (for 1 and 2), or it must be not in (for 5). For each of the other $7 - 3 = 4$ elements, there are still 2 choices available (in or not in). Therefore, by the FCP, there are

$$1 \times 1 \times 1 \times 2 \times 2 \times 2 \times 2 = 1^3 \times 2^4 = 16$$

ways to make the 7 decisions. That is, there are 16 different subsets of the desired form.

(e) How many subsets of $\{0,1,2,3,4,5,6\}$ are there which include either 0 or 6 but not both?

Approach 1: Use cases
Case 1: 0 is in, 6 is not in
We simply need to decide which of the other 5 elements are in the set. That is, the subsets we can construct are all of the form $\{0\} \cup A$, where A is some subset of $\{1,2,3,4,5\}$. Since this set has 5 elements and thus 2^5 subsets, there are 2^5 different subsets of $\{0,1,2,3,4,5,6\}$ which contain 0 but not 6. (Alternatively, there is just 1 choice – *in* – for 0 and just 1 choice – *not in* – for 6, and there are 2 choices available for each of the other 5 elements, so there are $1^2 \times 2^5 = 2^5$ different subsets.)
Case 2: 0 is not in, 6 is in
This is the same. Now we want subsets of the form $\{6\} \cup A$, where A is once again any subset of $\{1,2,3,4,5\}$, so in this case, once again, there are 2^5 subsets.
Total: There are $2^5 + 2^5 = 2(2^5) = 2^6 = 64$ subsets of $\{0,1,2,3,4,5,6\}$ which contain either 0 or 6 but not both.

Approach 2: Indirect Counting
We know that in total the set $\{0,1,2,3,4,5,6\}$ has 2^7 subsets. How many of these don't we want

here? We only wanted the ones which *do* have 0 or 6 but *don't* have both, so we don't want the ones that contain neither 0 nor 6 and we also don't want the ones that have both. There are 2^5 subsets which contain neither 0 nor 6 (these are all the subsets of $\{1, 2, 3, 4, 5\}$) and there are 2^5 more that contain *both* 0 and 6 (i.e. $\{0, 6\} \cup A$, for $A \subseteq \{1, 2, 3, 4, 5\}$). Therefore in total there are $2 \times 2^5 = 2^6$ subsets of $\{0, 1, 2, 3, 4, 5, 6\}$ which we didn't want to count, so the number of subsets we *did* want to count is

$$2^7 - 2^6 = 2 \times 2^6 - 2^6 = 2 \times 2^6 - 1 \times 2^6 = (2 - 1) \times 2^6 = 2^6$$

Approach 3: Direct Counting – use the FCP
Instead, we can think of 2 decisions: *Which one of 0 or 6 is in the subset?* and *Which other elements are in the subset?*. There are 2 choices for the first decision (0, or 6) and 2^5 choices for the second decision (the subsets of $\{1, 2, 3, 4, 5\}$), so by the FCP there are $2 \times 2^5 = 2^6$ subsets which contain either 0 or 6 but not both.

Let's look at one more example.

Example 1.11.

(a) In how many distinct ways can 5 identical candies be distributed between 2 children?

(b) In how many ways can 5 distinct candies be distributed between 2 children if there are no restrictions on how many either child may receive?

(c) In how many ways can the candies be distributed if each child must receive at least 1?

(a) Here, the candies are identical, so it doesn't matter *which* candies a child receives, only *how many*. We just need to decide how many candies each child will get. We can think of 2 decisions, one per child, of the form 'how many candies will this child receive?'. The first decision can be made in any of 6 different ways: the first child could be given 0 or 1 or 2 or 3 or 4 or all 5 of the 5 identical candies. Because all of the candies are to be distributed, once this first decision has been made, there is only 1 way to make the second decision: give all remaining candies to the second child. Therefore we see that there are $6 \times 1 = 6$ ways to distribute the candy between the 2 children.

(b) This time, we have a completely different situation. We are distributing 5 **distinct** (i.e. *different*) candies among the 2 children. This time, we need to think about not only *how many* candies each will get, but also *which* candies each child receives.

If we try to count the number of ways of distributing the candies by thinking of the 2 children as decisions (as we did in part (a)), things get quite complicated. We need to consider many different cases – in fact, there is 1 case for each of the 'ways' we counted in part (a). For instance, we would need to consider a case in which child 1 receives 2 candies (so that child 2 receives 3 candies) and then determine the number of choices available for the subsequent decision: *which 2 candies does child 1 receive?*

There is a *much* easier way to approach this problem. Instead of making a decision for each child, we make a decision for each of the candies. That is, since each of the 5 different candies will be given to one of the two children, we can consider 5 decisions of the form 'which child will *this* candy be given to?'. *Notice:* Once these 5 decisions have been made, we know both *how many* candies each child will receive and *which* candies each child will receive.

For each of the 5 decisions, we have 2 choices available: give the candy to child 1 or child 2. Therefore, by the FCP, the number of different ways to make these 5 decisions is:

$$2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$$

(c) What's different now? We're not allowed to give all of the candies to one child.

If we have 5 identical candies (as in (a)), this means that child 1 cannot receive 0 or 5 candies, so the available choices for the first decision are 1, 2, 3 or 4. That is, there are 4 different ways to make the decision 'how many candies does child 1 get?', and then (as in (a)) just 1 way to decide 'how many does child 2 get?' (i.e. all remaining candies), for a total of $4 \times 1 = 4$ ways to distribute 5 identical candies between 2 children so that each child receives at least 1 candy.

What if we have 5 distinct candies? The easiest way to approach this counting problem is by indirect counting, i.e., count too many things and then subtract off the ones we didn't want to count.

From (b), we know that there are 32 different ways to distribute 5 distinct candies beween 2 children when there are no restrictions on how the candies may be distributed.

How many of these ways are no longer allowed? Just 2: the one in which all of the candies are given to child 1, and the one in which all of the candies are given to child 2. All of the other ways of distributing the candies satisfy our requirement that each child should receive at least one candy.

Another way to think of this is as a new counting problem. We want to count the number of ways of distributing the candies which violate the condition 'each child must receive at least 1', i.e. the number of ways in which the candy can be distributed so that there is a child who receives no candy. This corresponds to a single decision: 'which child will get no candy?' for which there are 2 choices. Therefore once again we see that 2 of the ways we previously counted are no longer allowed.

Either way, we conclude that there are $32 - 2 = 30$ ways to distribute 5 distinct candies between 2 children so that each child receives at least one candy.

Notice: This indirect counting approach is often useful when we have some restriction on the ways in which something can be done. We count the number of ways the thing could be done if there was no restriction, and then subtract off the number of ways which violate the restriction. (That's what we did in Examples 1.9, too.) As long as it's easier to count the number of ways of violating the restriction than the number of ways of abiding by the restriction, this will be the easiest approach. (That wasn't actually true in Example 1.10 (e). We found a direct counting approach that was easier.)

Math 1228A/B Online

Lecture 5:
Arranging Objects in a Line

(text reference: Section 1.4, pages 25 - 28)

1.4 Permutations

A fundamental question in counting problems is "Does order matter?". If we're talking about a *set* of objects, then order doesn't matter. The set $\{0, 1\}$ can also be written as $\{1, 0\}$, but it's still the same set. (Remember, a set is an *unordered* collection of objects.)

In many counting problems, though, we're interested in *arrangements* of objects, so order matters. For instance, consider the following 3 different questions:

1. How many sets of 3 letters are there?
2. How many 3-letter 'words' are there?
3. How many 3-letter 'words' are there in which no letter is repeated?

The first question is concerned with the number of subsets of 3 objects that can be made from the 26 letters in the alphabet. This is the subject of section 1.5, which we'll get to soon.

The second question is an application of the FCP – we've already done problems like this. *Note:* By 'word' we mean a string of 3 letters, not necessarily distinct. It does not have to be an actual *word* in English or any other language.

The third question requires that each letter be distinct, so in this case, each of the 'words' uses 3 *different* letters. The question is asking how many *arrangements* of 3 out of 26 objects there are. This is also an application of the FCP and we've done similar problems before, but this is a special kind of application that arises a lot. We have special terminology and notation we use for these arrangements of distinct objects. This is the kind of problem we look at in the current section.

Define:

A **permutation** of a set of objects is an arrangement, or ordering, of the objects.

Let's start by looking at an example of a problem involving counting the number of permutations of *all* of the elements of a set.

Example 1.12. *In how many ways can 6 people be arranged in a line to have their photograph taken?*

We want to arrange 6 people in a line. We can think of the line as having 6 positions. We can approach this using the FCP, by considering each position in the line as a decision to be made, of the form 'who will stand in this position in the line?'. (Obviously, each person can stand in only one position in the line.)

To begin with, we have 6 choices of who to put in the first position in the line. Once we have someone in position 1, there are only 5 people left to choose among for position 2, so there are now 5 choices for the second decision. Next, we have 4 people left to choose from for position 3, so we have 4 choices for the third decision. Similarly, we have 3 choices for the fourth decision and 2 choices for the fifth position. Finally, when the first 5 positions in the line have been filled, we have just 1 person left, who must take the sixth position, so there's only 1 choice for the last decision. We see that, by the FCP, there are

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

ways to make the 6 decisions. That is, there are 720 different ways to line up 6 people for the photo.

We can easily see how this would extend to the more general question: *In how many ways can n distinct objects be arranged?*. Using the same reasoning, the answer is:

$$n \times (n-1) \times (n-2) \times ... \times 2 \times 1$$

Because permutations occur frequently in mathematics, especially in counting problems, we have special notation for this. We define:

> **n**! is pronounced **n factorial** and is defined to be the product of the integers from n down to 1. That is,
> $$n! = n \times (n-1) \times (n-2) \times ... \times 2 \times 1$$

For instance, we saw that $6! = 720$. Similarly, $3! = 3 \times 2 \times 1 = 6$ and $1! = 1$.

There's one more factorial we need which isn't covered by our previous definition. For convenience, we define:

$$0! = 1$$

Notice: We can think of this in the same way. Consider: In how many ways can 0 objects be arranged? Answer: there's only 1 way, which is not to arrange anything, i.e., do nothing.

Using this factorial notation, along with our reasoning from before, we see that:

> **Theorem:** The number of permutations of n objects is $n!$.

So we could have approached Example 1.12 as follows:
Each different way of arranging the 6 people in a line is a different permutation of the 6 objects (i.e. people). There are $6! = 720$ such permutations.

We can look at more complicated problems involving permutations.

Example 1.13. *In how many ways can 6 people be arranged in a line to have their photograph taken if there are 3 men and 3 women, and men and women must alternate in the line?*

This problem is different. If the men and women alternate in the line, then the line must look either like $MWMWMW$ or like $WMWMWM$ (where, of course, M representa a man and W represents a woman).

<u>One Approach:</u> (similar to previous)
We have 6 choices for who to put in the first position in the line. Once that person is chosen, it must be a person of the opposite sex who goes in the second position, and there are 3 of those, so we have 3 choices of who to put in the second position. Next, the person in the third position must be the same sex as the person in the first position - and there are 2 people of that sex not yet positioned, so there are 2 choices for the third position. For position 4, we go back to the opposite sex - and we again have 2 people of that sex left, so there are 2 choices for position 4. For position 5, the one remaining person of the same sex as the person first in line is the only choice. And then the last person, who is of the opposite sex, must take the sixth position.

We see that there are
$$6 \times 3 \times 2 \times 2 \times 1 \times 1 = 72$$
ways to make the 6 decisions, i.e. 72 ways to arrange the 3 men and 3 women so that men and women alternate in line.

<u>Another Approach:</u> (Make only 3 decisions)
Since we must have either $MWMWMW$ or $WMWMWM$, we can first decide: *Will it be a man or a woman in the first position?*. There are 2 ways to make this decision.

Now, we know the sex to be put in each position in the line. All that remains is: *What order will the men stand in?* and *What order will the women stand in?* Since there are 3 men, there are 3! ways to arrange the men. Similarly, there are 3! ways to arrange the 3 women. That is, there are

3! ways to make the second decision (order of men) and 3! ways to make the third decision (order of women). So by the FCP, there are

$$2 \times 3! \times 3! = 2 \times 6 \times 6 = 72$$

ways to make the 3 decisions. That is, again we see that there are 72 ways the 3 men and 3 women can be arranged so that men and women alternate in line.

Notice: The second approach is easier and also is more easily extended to bigger, i.e. more complicated, problems.

Let's look at another problem in which we take a similar approach:

Example 1.14. *There are 5 different math books, 4 different chemistry books and 6 different biology books to be put on a bookshelf. If the books are to be grouped by subject, in how many ways can this be done?*

What decisions do we need to make here? There are 4 of them:

1. order of math books — 5 books $\Rightarrow$ 5! permutations
2. order of chemistry books — 4 books $\Rightarrow$ 4! permutations
3. order of biology books — 6 books $\Rightarrow$ 6! permutations
4. order of *subjects* — 3 subjects $\Rightarrow$ 3! permutations

We see that there are 5!4!6!3! ways in total to put the books on the shelf so that they are grouped together by subject.

When we talk about arrangements, we don't have to be talking about physical arrangements. For instance, consider the next problem:

Example 1.15. *Jimmy, Johnny, Nicky and Sandy, all aged 8, have formed a club. They know that a club should have officers, so they have decided to elect themselves to the positions of President, Vice President, Vice Vice President and Chairman of the Board. In how many ways can these positions be filled by the four children, if each child will hold exactly one position?*

Here, we want to count the number of ways in which the 4 children can be assigned to the 4 offices, with each child assigned to a different position. Although we are not physically arranging the children, we are still looking for the number of arrangements of the 4 children. That is, the number of ways the 4 positions can be filled is simply the number of permutations of the 4 children, which is $4! = 4 \times 3 \times 2 \times 1 = 24$.

The next example is a bit different.

Example 1.16. *In how many ways can the positions of President, Vice President, Treasurer and Secretary be filled by election of members from a 15-person Board of Directors? (Of course, any director may hold at most one position.)*

What's different here? In Example 1.15, we had 4 children and wanted to arrange them all. This time, we have 15 people, but we only want to arrange 4 of them. We can use the FCP.

We have 4 decisions to make: one for each position to be filled. There are 15 choices for who should fill the first position (say, President). Once that choice has been made, that person cannot hold another position, so there are only 14 choices left for who should hold the second position. Similarly, we have 13 choices left when we get to the third position, and 12 choices left for the fourth position. Thus, we see that there are

$$15 \times 14 \times 13 \times 12$$

ways in which to make these four decisions. That is, there are $15 \times 14 \times 13 \times 12$ ways in which the 4 officers can be elected from the 15-person board.

Let's look at this answer some more. We were arranging 4 people from a set of 15, so we used the first 4 terms of 15!. The product $15 \times 14 \times 13 \times 12$ is kind of messy. We can easily see how we would extend this to a larger problem, like arranging 20 out of 100 objects, but it would be even messier.

Maybe there's a less messy way that we can *express* this. We have the first four terms of 15!. What if we both multiply and divide by the missing terms? (If we multiply and divide by the same thing, then we are actually multiplying by 1, so of course this won't change the value of the answer, just the *appearance*.) We can take the product we have so far and continue multiplying by successively smaller integers, but also dividing by them, so that really we're not changing anything. We found the answer to be the product of the integers from 15 down to 12, so if we continue with 11 and the rest, we will multiply and also divide by:

$$11 \times 10 \times 9 \times ... \times 2 \times 1 = 11!$$

That is, if we take our $15 \times 14 \times 13 \times 12$ and multiply it by 1, expressed as $\frac{11!}{11!}$, then we get:

$$\begin{aligned} 15 \times 14 \times 13 \times 12 &= 15 \times 14 \times 13 \times 12 \times 1 \\ &= 15 \times 14 \times 13 \times 12 \times \frac{11!}{11!} \\ &= \frac{15 \times 14 \times 13 \times 12 \times 11!}{11!} \\ &= \frac{15 \times 14 \times 13 \times 12 \times 11 \times ... \times 2 \times 1}{11!} \\ &= \frac{15!}{11!} \end{aligned}$$

Hmm. 15! over 11!. Clearly, the 15 in 15! comes from the fact that there are 15 people on the board. Where did the 11 in 11! come from? That was the number of people we *weren't* arranging. We wanted to arrange only 4 out of 15 people, leaving $15 - 4 = 11$ people not arranged. That is, of the 15 decisions we would make if we were arranging all 15 people, we didn't make $15 - 4$ of them. So we see that

$$15 \times 14 \times 13 \times 12 = \frac{15!}{(15-4)!}$$

This allows us to see how to extend the result we found in Example 1.16 to the general case, in which we have some number n distinct objects, and we need to arrange only some number $k < n$ of them.

That is, we see how to answer the question
In how many ways can k out of n objects be arranged?
In our example we were arranging $k = 4$ out of $n = 15$ objects, and we saw that this can be done in $\frac{15!}{(15-4)!} = \frac{n!}{(n-k)!}$ ways. This is true in general:

Theorem: The number of permutations of k out of n distinct objects is $\frac{n!}{(n-k)!}$.
(*Note:* We must, of course, have $k \leq n$.)

Some observations:

1. We write $\frac{15!}{11!}$ as shorthand for $15 \times 14 \times 13 \times 12$, but to actually calculate this, we would not usually calculate 15! and 11!. We would just use $15 \times 14 \times 13 \times 12$. (However, if using a calculator which has a ! button, it's probably easier to actually do $15! \div 11!$.)

2. Sometimes we evaluate an expression like this, and sometimes we just leave it in this format. The general practice used in this course is to calculate the actual number if it is reasonably easy to do so. For instance, if the question was "In how many ways can 3 out of 6 objects be arranged?", the answer

$$\frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \times 5 \times 4 \times \cancel{3} \times \cancel{2} \times \cancel{1}}{\cancel{3} \times \cancel{2} \times \cancel{1}} = 6 \times 5 \times 4$$

   can easily be calculated as 120, so we would probably say that the answer was 120. However, something like $15 \times 14 \times 13 \times 12$ is not so easily calculated, so we would just leave it like that, or as $\frac{15!}{11!}$, rather than give the answer as 32,760.

3. In the example, we used the fact that $15! = 15 \times 14 \times 13 \times 12 \times 11!$. Each factorial has embedded in it *every* smaller factorial, so we can always factor a smaller factorial out of a larger factorial (although we may not want to, to make it easier to express). For instance, suppose we have $\frac{n!}{(n-1)!}$. We can express this more simply by factoring the $(n-1)!$ out of the $n!$. That is, we have

$$\frac{n!}{(n-1)!} = \frac{n \times \cancel{(n-1)!}}{\cancel{(n-1)!}} = n$$

   Similarly, above we looked at $\frac{6!}{(6-3)!}$. We could have simply said $\frac{6!}{(6-3)!} = \frac{6 \times 5 \times 4 \times 3!}{3!}$, without expanding the 3!'s before cancelling them out.

Let's look at one further complication.

Example 1.17. *In how many ways can the positions of President, Vice President, Treasurer and Secretary by filled by election of members from a 15-person Board of Directors, if only 3 of the directors have the necessary qualifications to hold the position of Treasurer?*

This is a situation in which we can make the problem very complicated if we don't approach it carefully. If some positions have already been filled before the Treasurer position, then we will need to consider different cases, based on how many of the 3 qualified directors are still available, i.e., have not yet been appointed to positions. In fact, if we consider the Treasurer position last, there is the possibility that there will not be *any* qualified person available. However, we can find the answer in a very straightforward manner if we simply make sure we consider the Treasurer position first. (Remember, we should always tackle the most restrictive condition first.)

We can approach this by considering 2 decisions: *who will be Treasurer?* and then *how will the other 3 positions be filled?*. There are 3 directors qualified to be Treasurer, so there are 3 ways in which the Treasurer decision can be made. Once the Treasurer has been chosen, there are 14 directors remaining, and we need to arrange 3 of them in the positions of President, Vice President and Secretary, which can be done in $\frac{14!}{(14-3)!} = \frac{14!}{11!}$ ways. Therefore the total number of ways of making the 2 decisions, i.e., of filling the 4 positions, is $3 \times \frac{14!}{11!} = 3 \times 14 \times 13 \times 12$.

The permutations we've been talking about so far are called **linear** permutations, because we can consider these as arranging things in a line (even when we're not physically arranging them). Sometimes, arranging objects does not quite correspond to arrangements in a line, because all we're really concerned with is which object is next to which, with no first and no last – like in a circle. For this, there's another kind of permutation that we sometimes need. We'll learn about that in the next lesson.

Math 1228A/B Online

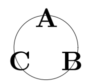
Lecture 6:
Arranging Objects in a Circle

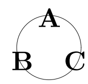
(text reference: Section 1.4, pages 28 - 30)

Circular Permutations

Question: In how many ways can 3 people be arranged in a circle?

Suppose the people are A, B and C. They could be arranged in the order ABC as we look *clockwise* around the circle, or in the order ABC as we look *counterclockwise* around the circle. That is, we could have either of the following:





Are there any other circular arrangements which are distinct from these 2? No. A circle has no beginning and no end, i.e. has no first position and no last position. The only thing that distinguishes one circular arrangement from another is the *relative* position of the objects, i.e. which is on the right or the left of which. For our problem of arranging 3 people in a circle, every arrangement we can make either has:

B to the left of A and C to the right of A

or else has:

C to the left of A and B to the right of A

so there are only 2 ways to arrange 3 people in a circle.

How can we relate this to permutations? That is, can we find a way to relate this to the 3! linear permutations of 3 people, in order to get some insight as to how to generalize this result?

Suppose that the 3 people are arranged in a circle, and we want to arrange them in a line, so that they keep the same relative order that they had in the circle. The circle A-B-C *clockwise* could form a line as ABC, as BCA or as CAB. That is, there are 3 ways we could choose where to break the circle in order to straighten it into a line (i.e., 3 choices of who will be first in the line), but after that, the other positions must be filled in the order in which the people were arranged in the circle. Similarly, the circle A-B-C *counterclockwise*, or A-C-B *clockwise*, could form a line as ACB, CBA or BAC.

We see that each circular permutation of 3 objects gives rise to 3 different linear permutations of the objects, which arise from the 3 different ways to choose which of the objects is first in the linear permutation. That is, we have:

$$\#\text{ of linear permutations} = 3 \times \#\text{ of circular permutations}$$

which we can rewrite as:

$$\#\text{ of circular permutations} = \frac{1}{3} \times \#\text{ of linear permutations}$$

Notice: There are $3! = 6$ ways to arrange 3 people in a line, and we saw that there are $2 = \frac{1}{3} \times 6$ ways to arrange 3 people in a circle.

In the same way, if we have n people arranged in a circle, then there are n different places we could break the circle to straighten it out into a line, i.e. n different people who might be first in the line, but then the rest of the line is fixed in the same order as the circle, so each circular permutation gives n distinct linear permutations. Therefore, when we are arranging n objects,

$$\#\text{ of linear permutations} = n \times \#\text{ of circular permutations}$$

so we get

$$\#\text{ of circular permutations} = \frac{1}{n} \times \#\text{ of linear permutations}$$

But of course, we know that the number of linear permutations of n objects is $n!$. So we have

$$\begin{aligned}\text{\# of circular permutations} &= \frac{1}{n} \times \text{\# of linear permutations} \\ &= \frac{1}{n} \times n! = \frac{n!}{n} = \frac{n \times (n-1)!}{n} = (n-1)!\end{aligned}$$

We see that the following holds in general:

Theorem: The number of circular permutations of n distinct objects is $(n-1)!$.

Notice:

1. For $n = 3$, we get $(n-1)! = (3-1)! = 2! = 2 \times 1 = 2$, as we saw.
2. The circle doesn't have to physically be a circle. For instance, if we arrange 3 people in a circle, they're actually forming a triangle, not a circle. Similarly, if we arrange 4 people around a small 4-sided table, with one on each side, again what we have is a circular permutation, because all that matters is the relative positions – who is to the right or left of whom. Circular permutations apply any time we are concerned only with the positioning of the objects *relative* to one another, without any sense of first or last.

Let's look at some examples involving circular permutations.

Example 1.18. *(a) King Arthur and 6 of his knights will be sitting around The Round Table while discussing plans for their next battle. In how many different ways could the seven men be arranged at the table?*

Here, we just want to know the number of circular permutations of 7 objects, which we know is $(7-1)! = 6! = 720$.

Of course, we can have more complicated problems. For instance, we may want some of the objects to be in a specified position relative to one another.

Example 1.18. *(b) Queen Guinevere will be joining them for their next discussion. If Guinevere must be seated across from Arthur, how many different seating arrangements are possible?*

We can start by arranging Arthur and the 6 knights. Then we just need to have Guinevere come and sit across from Arthur. There are $(7-1) = 6!$ ways to arrange Arthur and the knights, and then there's only 1 way to decide where Guinevere will sit – she must sit across the table from Arthur. So there are $6! \times 1 = 720$ arrangements of Arthur, Guinevere and 6 knights, with Guinevere sitting across from Arthur.

Example 1.18. *(c) If Queen Guinevere must sit beside King Arthur, instead of opposite him, how many seating arrangements are there now?*

One Approach:
Again, we can arrange the 7 men, and then have Guinevere come in and take her place at the table. As before, there are 6! circular permutations of the 7 men. But now, there are 2 choices of where Guinevere can sit – either on Arthur's right or on Arthur's left. So we see that this time, the number of possible seating arrangements is:

$$6! \times 2 = 720 \times 2 = 1440$$

Another Approach: (i.e. look at it in a slightly different way)
Consider Arthur and Guinevere as a single object. We think about arranging 7 objects in a circle, where the 7 objects consist of 6 knights, and 1 object which is "Arthur and Guinevere". Then arrange that special object, i.e. arrange Arthur and Guinevere.

There are $(7-1)! = 6!$ ways to arrange the 7 objects, one of which is "Arthur and Guinevere", and then there are 2! ways to arrange the 2 objects, Arthur and Guinevere, within this bigger arrangement. Again we see that there are $6! \times 2! = 720 \times 2 = 1440$ different seating arrangements.

Notes:

1. The second approach generalizes more easily. For instance, we could easily use this same approach if we had 20 people going to sit in a circle, where some 5 of them all had to sit together. We would consider arranging $(20-5)+1 = 16$ objects in a circle, and then arrange the 5 who must all be together, so there are $15! \times 5!$ possible arrangements.

2. When we arrange the Arthur-and-Guinevere object *within* the cricle, or we arrange the 5 objects which must be together in the circle of 20 objects, we're *not* arranging them "in a circle". That is, we don't have a circle as part of a larger circle, so we use linear permutations. (There *is* a first and last within this group of objects.) So there are 2! ways, not $(2-1)! = 1$ way, to arrange Arthur and Guinevere, and likewise there are 5!, not $(5-1)!$, ways to arrange the 5 objects which are together in the circle of 20 objects.

3. We can have this same kind of complication with linear permutations, too. For instance, if we want to arrange Arthur, Guinevere and 6 knights in a line, with Arthur and Guinevere required to be side-by-side, we can consider 2 decisions: first, arrange 7 objects (6 knights and one Arthur-and-Guinevere object) in a line, which can be done in 7! different ways, and then arrange the Arthur-and-Guivevere object, which can be done in 2! ways. This gives a total of $7!2! = 5040 \times 2 = 10080$ different possible arrangements.

Consider the next problem:

Example 1.19. *King Arthur always supplies wine at these discussions with his knights. The servant who is setting up the room for a discussion involving Arthur and 6 of his knights must place a flagon at each place. There is a set of 12 flagons, each with a different design engraved on it. In how many ways can the servant set 7 of these flagons around The Round Table?*

What we are asked to find here is the number of circular permutations of 7 out of 12 objects. We can use the same reasoning we used before, considering the relationship between the number of circular permutations and the number of linear permutations, by thinking about straightening out the circle into a line, starting the line with any of the 7 flagons. Or, expressed another way, we find the number of linear permutations by considering 2 decisions: 'which circular permutation will be used?' and 'which flagon is first in the line?'. We see that

$$\#\text{ of linear permutations} = \#\text{ of circular permutations} \times 7$$

which we can rearrange to

$$\#\text{ of circular permutations} = \frac{1}{7} \times \#\text{ of linear permutations}$$

This time, of course, we are not arranging all of the objects in some set. Instead, we are arranging 7 out of 12 objects from a set. But we know that the number of permutations, i.e. linear permutations, of k out of n objects is $\frac{n!}{(n-k)!}$, so we have

$$\#\text{ of linear permutations} = \frac{12!}{(12-7)!} = \frac{12!}{5!}$$

Thus we get

$$\begin{aligned} \text{\# of circular permutations} &= \frac{1}{7} \times \text{\# of linear permutations} \\ &= \frac{1}{7} \times \frac{12!}{5!} \end{aligned}$$

Of course, we can generalize this to a formula that we can use anytime we need to arrange some k out of n objects in a circle. We get:

Theorem:

The number of circular permutations of k out of n distinct objects is $\frac{1}{k} \times \frac{n!}{(n-k)!}$.

(However, after the next lecture there will be an easier way to find this than by memorizing a separate formula.)

Math 1228A/B Online

Lecture 7:
Choosing Objects

(text reference: Section 1.5)

1.5 Combinations: Subsets of a Given Size

In counting problems, it is always necessary to ask yourself *'does order matter?'*. Permutations are all about order, i.e. ordering or arranging items. We know that with sets, order is not important. So the question *'How many subsets of k objects can be formed from a set of n objects?'* is very different from the question '*How many arrangements of k out of n objects are there?*'. However, these 2 questions are related to one another.

Definition: A subset of a specified size is called a **combination**.

One way of finding all of the permutations of k out of n objects would be to start by finding all of the different subsets of size k, and then for each of these, list or count all of the $k!$ permutations of the objects in that subset. That is, we can consider the problem of permuting k out of n objects as 2 decisions which have to be made: *choose a subset of size k* and then *arrange the k chosen objects.* Thus we see that:

\# of permutations of k out of n objects = # of combinations of k out of n objects $\times k!$

Of course, we can rearrange this to

\# of combinations of k out of n objects = $\frac{1}{k!}\times$ # of permutations of k out of n objects

And since we know that the number of permutations of k out of n objects is $\frac{n!}{(n-k)!}$, then we see that the number of *combinations* of k out of n objects is

$$\frac{1}{k!} \times \frac{n!}{(n-k)!} = \frac{n!}{k!(n-k)!}$$

These combinations come up a lot, so we have special notation we use.

Definition:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

We pronounce this as "n choose k". It is sometimes referred to as a *binomial coefficient.* (*Notice:* we must have $k \leq n$.)

Theorem: The number of different ways of choosing a subset of size k from a set of n distinct objects is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Example 1.20. *A certain committee has 15 members. A 4-member subcommittee is to be set up. In how many ways can the members of the subcommittee be selected?*

Here, we have a set of 15 people, from whom we want to choose 4 to be on the subcommittee. Since these 4 people are not being given specific roles, but will just be on the subcommittee together, the order of the 4 people is not important. That is, what we want is to choose a *subset* of 4 of the 15 people, not an *arrangement* of 4 of the 15 people. Therefore we see that the number of ways in which the subcommittee can be chosen is

$$\binom{15}{4} = \frac{15!}{4!(15-4)!} = \frac{15!}{4!11!} = \frac{15\times \not{14}^7 \times 13\times \not{12}\times \not{11}!}{\not{4}\times \not{3}\times \not{2}\times \not{11}!} = 15\times 7\times 13 = 1365$$

Notice: This is very different from the answer to Example 1.16, in which we were *arranging* 4 out of 15 people. The answer there was $15 \times 14 \times 13 \times 12$. Make sure you understand the difference between these 2 problems.

Properties of $\binom{n}{k}$

1. For any $n \geq 0$, $\binom{n}{0} = 1$

   Proof: Recall that we defined $0! = 1$. This gives

   $$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{\not{n}!}{1\times \not{n}!} = \frac{1}{1} = 1$$

   This result says that there's only 1 way to choose no objects from a set, which is to not choose anything.

2. For any $n \geq 1$, $\binom{n}{1} = n$

   Proof: (This just says that there are n different ways to choose 1 out of n objects.)

   $$\binom{n}{1} = \frac{n!}{1!(n-1)!} = \frac{n\times \cancel{(n-1)!}}{1\times \cancel{(n-1)!}} = n$$

3. For any $n \geq 1$, $\binom{n}{n-1} = n$

   Proof: (There are also n ways to choose $n-1$ out of n objects, because there are n ways to choose which 1 object to leave out.)

   $$\binom{n}{n-1} = \frac{n!}{(n-1)!(n-(n-1))!} = \frac{n\times \cancel{(n-1)!}}{\cancel{(n-1)!}(n-n+1)!} = \frac{n}{1!} = n$$

4. For any $n \geq 0$, $\binom{n}{n} = 1$

   Proof: (There's only 1 way to choose all n out of n objects, which is to choose them all.)

   $$\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{\not{n}!}{\not{n}!0!} = 1$$

5. For any k with $0 \leq k \leq n$, $\binom{n}{k} = \binom{n}{n-k}$

   Proof: (Again, choosing a subset of k out of n objects is the same as choosing $n-k$ of the objects not to put in the subset.)

   $$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!(n-n+k)!} = \frac{n!}{(n-k)!k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Whenever we choose *some* of a group of objects we can always look at it the other way around and focus on the ones we're not choosing, without really changing anything. What we're really doing here is separating the objects into 2 groups: the 'chosen' group and the 'not chosen' group. For instance, instead of choosing 4 out of 15 people to be on a subcommittee, we could choose 11 out of 15 people *not* to be on the subcommittee, and we'd really be doing the same thing. When we select 4 people for the subcommittee, we've separated the 15 people into 4 who will be on the subcommittee and 11 who won't.

We can have complications in combinations problems, just as we can in permutations problems.

Example 1.21. *Suppose the 15-member committee from Example 1.20 is made up of 6 men and 9 women.*
(a) In how many ways can the subcommittee be formed, if it must consist of 2 men and 2 women?
(b) In how many ways can the subcommittee be formed if at least 2 men must be selected?
(c) In how many ways can the subcommittee be formed if at least one woman must be selected?

(a) Here, we have 2 decisions to make: 'which 2 of the 6 men will be on the committee?' and 'which 2 of the 9 women will be on the subcommittee?'. There are

$$\binom{6}{2} = \frac{6!}{2!4!} = \frac{6\times 5\times \cancel{4!}}{2!\,\cancel{4!}} = \frac{\cancel{6}^3\times 5}{\cancel{2}} = 15$$

ways to choose 2 of the 6 men, and

$$\binom{9}{2} = \frac{9!}{2!7!} = \frac{9\times 8\times \cancel{7!}}{2!\,\cancel{7!}} = \frac{9\times \cancel{8}^4}{\cancel{2}} = 36$$

ways to choose 2 of the 9 women, so there are $15 \times 36 = 540$ ways to choose 2 men and 2 women to form the subcommittee.

(b) We need to find the number of possible subcommittees which have *at least* 2 men. A subcommittee of 4 people of which at least 2 are men must have 2 or 3 or 4 men. This gives us 3 cases:

<u>Case 1:</u> 2 men $\Rightarrow$ same as (a) because the other 2 must be women. So there are 540 ways to form a subcommittee with exactly 2 men.

<u>Case 2:</u> 3 men $\Rightarrow$ 1 woman

$$\binom{6}{3}\times\binom{9}{1} = \frac{6!}{3!3!}\times 9 = \frac{\cancel{6}\times 5\times 4}{\cancel{3}\times\cancel{2}\times 1}\times 9 = 20\times 9 = 180$$

We see that there are 180 possible subcommittees with exactly 3 men.

<u>Case 3:</u> 4 men $\Rightarrow$ 0 women

$$\binom{6}{4} = \frac{6!}{4!2!} = \frac{\cancel{6}^3\times 5}{\cancel{2}} = 15 \text{ or } \binom{6}{4}\times\binom{9}{0} = 15\times 1 = 15$$

so there are 15 possible subcommittees with exactly 4 men (and therefore 0 women).

In total, there are $540 + 180 + 15 = 735$ ways to form a subcommittee with at least 2 men. (Remember, if we do this *and* do that, then we multiply, e.g. m ways to do this and n ways to do that so we can do this *and* that in $m \cdot n$ ways. But if we do this *or* that, then we add, so there are $m + n$ ways we can do this *or* do that. Here, we choose a subcommittee with 2 *or* 3 *or* 4 men.)

(c) Now, we want to count the subcommittees which have at least one woman. We could do this using exactly the same approach as we used in (b), except now we would need to consider 4 cases:

1 or 2 or 3 or 4 women on the subcommittee. However, this question is much easier if we consider the complementary situation. That is, determine how many subcommittees *don't* have at least one woman, and then subtract that from the total we found in example 1.20.

We know (from example 1.20) that there are, in total, 1365 possible subcommittees that could be formed, if we don't care how many are men or women. We saw in part (b) of the current example (Case 3) that there are 15 ways to form a subcommittee which has 4 men and therefore has no women. Since 15 of the 1365 possible subcommittees don't include any women, then there must be $1365 - 15 = 1350$ subcommittees which do include at least one woman.

Notice: Whenever it's necessary to consider several cases, think about whether the complementary situation involves fewer cases, i.e., is less work.

Example 1.22. *Suzie will be turning 10 next week, and her mother has told her that she may invite 10 - 12 of her friends to her birthday party. Suzie has made a list of all the people she would like to invite and there are 15 names on this master list. She now wants to make a guest list of those (selected from the master list) who will be invited to the party.*
(a) How many different guest lists for the party could Suzie make?
(b) Three of the names on Suzie's master list are those of the Rodriguez triplets. Suzie knows that it wouldn't be very nice to invite only 1 or 2 of the triplets to the party, so she has decided that she should invite either all or none of them. How many different guest lists are possible now?

(a) There are 15 friends that Suzie is considering inviting. However we don't know how many she will actually decide to invite. She might put 10 or 11 or 12 of her 15 friends on the guest list. Therefore the number of different possible guest lists is given by

$$\binom{15}{10} + \binom{15}{11} + \binom{15}{12} = \frac{15!}{10!5!} + \frac{15!}{11!4!} + \frac{15!}{12!3!} = 3003 + 1365 + 455 = 4823$$

(b) Now, Suzie is considering 2 separate cases: she might invite the triplets, or she might not. If Suzie does invite the triplets, then that's 3 of the 10, 11 or 12 people to go on the guest list, so she just needs 7, 8 or 9 more, which will be chosen from her other $15 - 3 = 12$ friends. We see that the number of possible guest lists which *do* include the Rodriguez triplets is

$$\binom{12}{7} + \binom{12}{8} + \binom{12}{9} = \frac{12!}{7!5!} + \frac{12!}{8!4!} + \frac{12!}{9!3!} = 792 + 495 + 220 = 1507$$

On the other hand, if Suzie decides not to invite the triplets, then she will put 10 or 11 or 12 of her other 12 friends on the guest list, so the number of possible guest lists which *do not* include the Rodriguez triplets is

$$\binom{12}{10} + \binom{12}{11} + \binom{12}{12} = \frac{12!}{10!2!} + \frac{12!}{11!1!} + \frac{12!}{12!0!} = 66 + 12 + 1 = 79$$

In total, there are $1507 + 79 = 1586$ possible guest lists which include either all or none of the Rodriguez triplets.

<u>A final thought on permutations of k out of n objects:</u>
Now that we have a formula for the number of ways of choosing k out of n distinct objects, we don't need to specifically remember formulas for the number of permutations of k out of n objects, for either the linear or the circular case. We can always find the answer by thinking of the problem as:

first, choose k objects, then arrange the chosen objects. For instance, suppose we need to count the number of linear permutations of 4 out of 6 objects. Rather than remembering that the formula is $\frac{6!}{(6-4)!}$, we could count the number of ways of choosing 4 out of 6 objects, which is $\binom{6}{4}$, and the number of arrangements of the 4 chosen objects, which is 4!. This way, we find the number of possible arrangements to be

$$\binom{6}{4} \times 4! = \frac{6!}{\not{4}!2!} \times \not{4}! = \frac{6!}{2!}$$

Likewise, if we needed to count the number of circular permutations of 4 out of 6 objects, we could take the same approach, but count the number of circular permutations of the 4 chosen objects, instead of the number of linear permutations. That is, instead of using the formula $\frac{1}{k} \times \frac{n!}{(n-k)!} = \frac{1}{4} \times \frac{6!}{(6-4)!}$, we could instead use

$$\binom{6}{4} \times (4-1)! = \frac{6!}{4!2!} \times 3! = \frac{6!}{(4\times \not{3}!)2!} \times \not{3}! = \frac{6!}{4 \times 2} = \frac{1}{4} \times \frac{6!}{2!}$$

(*Note:* the rearrangements shown here serve to show that we get the same answer as when we use the previously learnt formlas. We wouldn't necessarily *do* those rearrangements to actually calculate the answers.)

The point here is that we need to know formulas for the number of combinations of k out of n objects, and the number of linear and circular permutations of n objects. But there's no need to clutter up our heads with the formulas for linear or circular permutations of k out of n objects, because we can easily derive those when we need them (i.e., get the answer another way), using the other formulas that we have to know anyway. We're going to have lots of formulas that we *do* need to remember. There's no need to make life more difficult by trying to remember extra formulas that we don't really need.

Note: Now that we know about combinations, we can answer the 3 questions posed in Lecture 5 (intro to section 1.4), on page 22. The answers are as follows:

1. How many sets of 3 letters are there?

$$\binom{26}{3} = \frac{26!}{3!23!} = \frac{26 \times 25 \times \not{24}^4 \times \not{23}!}{\not{3} \times \not{2} \times 1 \times \not{23}!} = 26 \times 25 \times 4 = 2600$$

2. How many 3-letter 'words' are there?

$$26 \times 26 \times 26 = (26)^3 = 17,576$$

3. How many 3-letter 'words' are there in which no letter is repeated?

$$\frac{26!}{(26-3)!} = \frac{26 \times 25 \times 24 \times 23!}{23!} = 26 \times 25 \times 24 = 15,600$$

OR

$$\binom{26}{3} \times 3! = \frac{26!}{3!23!} \times 3! = \frac{26!}{23!} = 26 \times 25 \times 24 = 15,600$$

Math 1228A/B Online

Lecture 8:

Two Problems

That are Really the Same

(text reference: Section 1.6, pages 41 - 42)

1.6 Labelling Problems

Consider the following problem:

Example 1.23. *Recall Example 1.22. (See Lecture 7, pg. 34.) Suzie ended up inviting 11 friends to her party (including the Rodriguez triplets) and they all came. The 12 children (Suzie and her 11 guests) are going to go out to play in Suzie's large back yard, which has been set up for croquet, badminton and basketball. There are only 5 croquet mallets and 4 badminton racquets. In how many ways can the children divide themselves up into groups to play these games, assuming that 5 will play croquet, 4 will play badminton and 3 will play basketball?*

We can find the answer using what we've already learned. We need to decide

1. which children will play croquet,
2. which children will play badminton, and
3. which children will play basketball.

That is, there are 3 decisions to be made. We simply need to determine in how many ways each decision can be made and then apply the FCP.

There are 12 children, and some 5 of them will play croquet. The order of the children is not important. That is, we don't need to select a first croquet player, a second croquet player, etc., we simply need to *choose* which 5 will play croquet, i.e. we want to choose a subset of 5 of the 12 children. There are $\binom{12}{5}$ combinations of 5 out of 12 objects, so there are $\binom{12}{5}$ ways to choose 5 of the children to play croquet.

Once these 5 children have been selected, there are $12 - 5 = 7$ children left, of whom 4 will be chosen to play badminton. Of course, this can be done in $\binom{7}{4}$ ways.

Now, there are just $7 - 4 = 3$ children remaining, who will play basketball. There are $\binom{3}{3}$ ways to choose 3 of the 3 children to play basketball. (Of course, we know that $\binom{3}{3} = 1$, but let's think of it as $\binom{3}{3}$ for now.)

All together, we see that there are $\binom{12}{5} \times \binom{7}{4} \times \binom{3}{3}$ ways in which the children can be divided up to play the 3 games. Expanding and simplifying this, we see that the number of ways is given by:

$$\binom{12}{5} \times \binom{7}{4} \times \binom{3}{3} = \frac{12!}{5!\,\cancel{7!}} \times \frac{\cancel{7!}}{4!\,\cancel{3!}} \times \frac{\cancel{3!}}{3!0!} = \frac{12!}{5!4!3!}$$

Let's think about what we ended up with here. The 12 in the numerator is the total number of children. The 5 and 4 and 3 in the denominator are the sizes of the groups they're being divided up into. We had some other terms, but they all cancelled out. Notice what happened. Every time we chose some k objects from the n objects we had available, we had an $(n-k)!$ in the denominator of $\binom{n}{k}$, and then there were only $n-k$ objects left to choose the next group from, so we had $(n-k)!$ in the next numerator. So each time, the $(n-k)!$ in the denominator of one fraction is cancelled out by the $(n-k)!$ in the numerator of the next fraction. Except, of course, the last fraction, where the $(n-k)!$ part of the denominator is 0!, because we've run out of objects. (And of course, $0! = 1$.)

This same kind of cancellation is going to happen *every time* we have this kind of problem, in which we are dividing a set of objects into a bunch of groups in such a way that each object ends up in a group. That is, as long as the total of the groups is the same as the number of objects we started with, we're going to have this kind of cancellation taking place, and we will end up with

$$\frac{\text{(total number of objects) factorial}}{\text{product of factorials of each of the group sizes}}$$

Again, this is a situation which comes up a lot in counting problems, so we have special notation.

Definition: $\binom{n}{n_1\ n_2\ ...n_k} = \dfrac{n!}{n_1!n_2!...n_k!}$ where $n_1 + n_2 + ... + n_k = n$

This notation is called a *multinomial coefficient*.

Note: We **must** have the sum of the numbers on the bottom equal to the number on the top. That is, this multinomial coefficient is *only* defined when the numbers across the bottom sum to the number on the top. It must be true that *each* object ends up in some group.

We see that what we found in the last example was actually the multinomial coefficient $\binom{12}{5\ 4\ 3}$. That is, in Example 1.23, the children can divide themselves up in $\binom{12}{5\ 4\ 3}$ ways.

Stating what we found above in general terms, we get:

> **Theorem:** The number of ways in which a set of n distinct objects can be divided up into k distinct groups of sizes n_1, n_2, ..., n_k, where $n_1 + n_2 + ... + n_k = n$, is given by
>
> $$\binom{n}{n_1\ n_2\ ...\ n_k} = \frac{n!}{n_1!n_2!...n_k!}$$

Notice: Think for a moment about a situation in which we are dividing a set of objects up into 2 groups. Suppose there are n objects in total, to be divided up into a group of k and a group of $n-k$. From the above theorem, we see that the number of ways of doing this is $\binom{n}{k\ (n-k)}$. How is this different from simply choosing k of the objects? It's not. When we choose k out of n objects, we're really (as observed in Lecture 7, see pg. 33) dividing the set up into 2 groups: the chosen objects and the objects which weren't chosen. So choosing k out of n objects is the same as dividing a set of n objects up into a group of size k and another group of size $n-k$. And of course we have

$$\binom{n}{k\ (n-k)} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

For instance, in Example 1.20 (see Lecture 7, pg. 31), the number of ways of choosing 4 out of 15 people to be on the subcommittee could be expressed as $\binom{15}{4\ 11}$, the number of ways of dividing the 15 people into a group of 4 who will be on the subcommittee and a group of 11 who will not.

Now, let's look at Example 1.23 (above) again, but think of it in a slightly different way. We can think of the problem as follows:

> We have 12 children, and we want to give 5 of them *labels* that say croquet, give 4 of them *labels* that say badminton and give 3 of them *labels* that say basketball. Then all of the children who have the same kind of label will go off and play that game.

Looked at this way, we have 12 labels, of 3 different types:

5 identical labels of the first type: *croquet*
4 identical labels of the second type: *badminton*
3 identical labels of the third type: *basketball*

Every different way in which the children can be divided up into groups corresponds to a different way of *arranging* these labels among the 12 children. (For instance, imagine that the children are standing in a line. When we give out the labels, we're arranging the labels in the line.) Therefore it must be true that the number of distinct *permutations* of these 12 labels is the same as the number of ways of dividing up the 12 children to play the games. So we see that the number of distinct permutations of 12 labels, where we have 5 identical labels of one type, 4 identical labels of a second type and 3 identical labels of a third type must be given by $\binom{12}{5\ 4\ 3}$. This gives us another way of looking at, or another use for, the multinomial coefficients we defined earlier.

Theorem: The number of distinct permutations of n objects, where there are k distinguishable types of objects, with n_1 identical objects of type 1, n_2 identical objects of type 2, ..., and n_k identical objects of type k, with $n_1 + n_2 + ... + n_k = n$, is given by

$$\binom{n}{n_1\ n_2\ ...n_k} = \frac{n!}{n_1!n_2!...n_k!}$$

By thinking of the problem of dividing the set of children into groups as a *labelling problem*, we see that it is inherently the same as other problems which, at first glance, can appear to be quite different. For instance, consider the next example:

Example 1.24. *In how many distinct ways can the letters of the word MISSISSAUGA be arranged?*

The problem of arranging a given set of letters doesn't sound like it's the same sort of thing as the problem of dividing a set of children into different groups. However, it is.

First of all, notice that we have a set of 11 letters, including some repetitions. That is, the letters are not all distinct. If we had 11 distinct letters, we could arrange them in 11! ways. However, having repeated letters complicates things. We can't tell the difference between one S and another, etc., so if we listed the 11! permutations of the letters as if they were all distinct, there would be a lot of repetition on the list.

To see why this problem is inherently the same as the previous problem, think of it this way: Every arrangement of the letters is an 11-letter "word". We have 11 positions in the word, which we must fill. The 11 letters we're going to use consist of 4 $S's$, 2 $I's$, 2 $A's$, 1 M, 1 U and 1 G. So we want to **label** the 11 positions of the word, using 4 S labels, 2 I labels, 2 A labels, 1 M label, 1 U label and 1 G label.

Thought of this way, we see that the problem of *arranging* objects, when we have groups of identical objects, is fundamentally the same problem as *dividing* a set of distinct objects into a number of groups. Although these 2 kinds of problems *appear* to be different, they are, in fact, exactly the same. (In our example, we want to divide the positions of the 11-letter word into an S group, an I group, etc.)

Using our second multinomial result, which doesn't actually refer to labelling, we can easily see how it applies to the arrangement of the letters in the word. We have 11 objects (letters) of 6 different types, where there are 4 of type S, 2 of type I, 2 of type A, and 1 each of types M, U and G. Therefore the number of distinct arrangements of these 11 letters is

$$\binom{11}{4\ 2\ 2\ 1\ 1\ 1} = \frac{11!}{4!2!2!1!1!1!} = \frac{11!}{4!2!2!}$$

Notice: We can omit the 1!'s from the final fraction, since multiplying by 1 doesn't matter. However we **can't** omit the three 1's from the multinomial coefficient, because the numbers on the bottom *must* add up to the number on the top.

Any time we have a problem about dividing objects into groups, *or* about arranging objects when some of the objects are identical, we use the multinomial coefficient. But for the first of these types of problems, an additional complication can arise. Here, we have assumed that the groups are *distinguishable*, i.e. that there's something different about each group. What if that's not the case? That will be the focus of the next lecture.

Math 1228A/B Online

Lecture 9:
Are the Groups *Distinguishable?*

(text reference: Section 1.6, pages 42 - 44)

It's very important to understand what's the same (i.e. identical, or *indistinguishable*) and what's different (i.e. distinct or *distinguishable*) in a labelling problem. Objects of the same type are indistinguishable. However the different *types* of objects must be distinguishable. That is, in the "dividing objects into groups" type of problem, it must be true that the different *groups* are distinguishable, in order for the multinomial coefficient to tell us in how many ways the division can be done. This is really just the question of "does order matter?" again. This time, the ordering applies to the *groups*, not to the individual objects.

Think about the following 3 problems:

1. In how many ways can 4 objects be divided into a group of 3 and a group of 1?
2. In how many ways can 4 objects be divided up, with 2 given to John and 2 given to Mary?
3. In how many ways can 4 objects be divided into 2 groups of 2?

In the first case, we are dividing 4 objects into 2 distinct groups – a group of 3 and a group of 1. These groups are *distinguishable* because they are different sizes. No matter what is going to be done with the groups, we can always tell the difference between a group of 3 and a group of 1. There are $\binom{4}{3\ 1} = \frac{4!}{3!1!} = 4$ distinct ways to ways to divide the objects up.

In the second case, we are again dividing the 4 objects up into 2 distinct groups – even though the 2 groups are the same size, one group is John's and the other group is Mary's. We can *distinguish* between these groups because they are going to be doing different things – being given to John, or being given to Mary. That is, giving objects A and B to John and giving C and D to Mary is a *different* division of the objects than giving C and D to John and giving A and B to Mary. The number of distinct ways of dividing the 4 objects up is again given by the multinomial coefficient. This time we have $\binom{4}{2\ 2} = \frac{4!}{2!2!} = \frac{4\times3\times2!}{2!\times2!} = \frac{4\times3}{2} = 6$ different ways in which the 4 objects can be divided into the 2 groups.

The third case is different. We are simply dividing the 4 objects into 2 groups of 2. We don't have a *first* group and a *second* group which are somehow distinct, we just have 2 groups the same size. There is no way to distinguish between the 2 groups. Each group is going to be doing the same thing – just being in a group together. So making a group containing objects A and B and another group containing objects C and D is the same as making a group with C and D and another group with A and B. Changing the order in which we mention the groups doesn't change the way the division has been done.

When we use the multinomial coefficient to divide 4 objects into 2 groups of 2, it counts the number of ways of dividing them into 2 *distinguishable* groups. That is, we're counting the number of ways to make a *first* group and a *second* group. So in using the multinomial coefficient, we *impose order* on the groups, by assuming that they are *distinguishable*. When what we have are actually *indistinguishable* groups, the multinomial coefficient *overcounts* the ways of performing the division into groups, by counting once for each possible ordering of the groups. Therefore, in order to find the number of ways to divide the set of objects up into *indistinguishable* groups, we must correct for this overcounting, by dividing by the number of possible orderings of the indistinguishable groups.

In the case of dividing 4 objects up into 2 groups of 2, we have 2 indistinguishable groups, so there are 2! possible arrangements of these groups. Therefore the number of ways of dividing the objects up is given by

$$\binom{4}{2\ 2} \div 2! = \frac{4!}{2!2!} \times \frac{1}{2!} = \frac{4!}{2!2!2!} = \frac{4\times3\times2}{2\times2\times2} = 3$$

(*Notice:* If the objects are A, B, C and D, these 3 distinct ways of dividing up the objects are: A with B and C with D; A with C and B with D; or A with D and B with C.)

In general, we need to recognize that it might not be *all* groups that are indistinguishable, but perhaps only *some* of the groups.

> **Theorem:** If a set of n distinct objects is to be divided up into k groups of sizes n_1, n_2, ..., n_k, with $n_1 + n_2 + ... + n_k = n$, where t of the groups are the same size and are indistinguishable, then the number of ways in which the objects can be divided into the groups is
>
> $$\frac{n!}{n_1!n_2!...n_k!} \div t!$$

Remember: Different sized groups are *always* distinguishable. Groups of the same size may be distinguishable or indistinguishable, depending on whether they are going to be doing the same or different things.

(*And Note:* This issue doesn't come up in the "permutations where some objects are identical" type of problem, because it doesn't make sense to say, for instance, that there are n_1 objects of type 1 and n_2 objects of type 2, but type 1 and type 2 are indistinguishable. If the types are indistinguishable then what we really have is $n_1 + n_2$ objects which are *all the same type.*)

Example 1.25. *(a) Recall Example 1.23. One of the croquet mallets is broken, so instead the children are going to divide into 3 groups of 4 to play the 3 games. In how many ways can this be done?*

Remember, in this example (see Lecture 8, pg. 36) we had the 12 children at Suzie's birthday party being divided into groups to play croquet, badminton and basketball. This time, we're dividing the 12 children into 3 groups of 4. Although the groups are the same size, they are going to be doing different things – one group will play croquet, one group will play badminton and one group will play basketball. Therefore we have *distinguishable* groups, so the number of ways in which the children can be divided up is given directly by the multinomial coefficient. We see that the number of distinct ways in which this division can be done is

$$\binom{12}{4\ 4\ 4} = \frac{12!}{4!4!4!} = \frac{12!}{(4!)^3}$$

Example 1.25. *(b) After playing outside for a while, the children go back inside to play Crazy Eights. Since there are 3 decks of cards, it has been decided to have 3 games of Crazy Eights, with 4 players in each game. In how many ways can the children be divided up to play cards?*

Again we are dividing the 12 children up into 3 groups of 4. However this time the 3 groups are all going to be doing the same thing – playing Crazy Eights. We have 3 groups which are all the same size and which will all be doing the same thing, so these groups are indistinguishable. Therefore we find the number of ways of dividing the children up by dividing the multinomial coefficient by the number of permutations of the 3 indistinguishable groups, i.e. 3!. This time, the number of distinct ways of dividing up the children is

$$\binom{12}{4\ 4\ 4} \div 3! = \frac{12!}{4!4!4!} \times \frac{1}{3!} = \frac{12!}{(4!)^3 3!}$$

Notice: If the groups were different sizes, then even though they are going to do the same thing, they would still be distinguishable. It is only because these groups are the same size *and* doing the same thing that they are indistinguishable.

To differentiate between the 2 situations, think about what we see once the groups have been formed and the children are playing. In the earlier problem, if Suzie's Mom goes outside and looks around, she see: a group of 4 children playing croquet, a group of 4 children playing badminton and a group of 4 children playing basketball. She sees a difference between the groups, even though they're all the same size. (Perhaps she says to herself "I see Suzie is in the group that's playing badminton, not in the croquet group or the basketball group.) But later, if she goes into the Rec Room where the kids are playing Crazy Eights, she sees: here's a group of 4 kids playing Crazy Eights ... and here's another ... and here's yet another. There's no real difference between the groups. (That is, "I see Suzie is in the group playing Crazy Eights, not the group playing Crazy Eights or the group playing Crazy Eights" is just silly.)

Example 1.25. *(c) Later, the children will choose between 2 activities. There are 3 identical sets of face paints, so 6 of the children will pair up to paint each other's faces. The remaining children will go back outside and play a game of '3 on 3' basketball. In how many ways can the children divide themselves up for these activities?*

The 12 children are going to divide up into 3 groups of 2 (i.e., 3 pairs, where each pair will take a set of face paints and paint one another's faces) and 2 groups of 3 (i.e., 2 teams for the game of 3 on 3 basketball).

Are these groups distinguishable or indistinguishable? Certainly the face painting groups are distinguishable from the basketball playing groups, both because they will engage in different activities and because the groups are of different sizes (groups of 2 versus groups of 3). However, the 3 groups of 2 are indistinguishable from one another, since these are groups of the same size which will all be doing the same thing (painting each other's faces). Likewise, the 2 groups of 3 are indistinguishable from one another, again being groups of the same size which will be doing the same thing (being on a basketball team together). So this time, we have *2 sets* of indistinguishable groups, although the sets are themselves distinguishable. When counting the number of distinct ways in which the 12 children can divide themselves into 3 groups of 2 and 2 groups of 3, we need to compensate for the fact that the 3 groups of 2 are indistinguishable from one another, and also for the fact that the 2 groups of 3 are indistinguishable.

There are $\binom{12}{2\,2\,2\,3\,3}$ ways in which the 12 children can divide themselves up into 3 distinguishable groups of size 2 and 2 distinguishable groups of size 3. But then we need to divide by the 3! possible arrangements of the groups of size 2 and *also* by the 2! possible arrangements of the groups of size 3, in recognition of the fact that the groups of the same size are actually, in this case, indistinguishable from one another. Therefore, the number of ways in which the children can divide themselves up for these activities is

$$\binom{12}{2\,2\,2\,3\,3} \div (3!2!) = \frac{12!}{2!2!2!3!3!} \times \frac{1}{3!2!} = \frac{12!}{(2!)^4(3!)^3} = \frac{12!}{2^4 6^3}$$

Another Approach: We could have taken a slightly different approach (with more steps) here, thinking about dividing the 12 children into 2 distinguishable groups of 6, a face painting group and a basketball playing group. Once the children have chosen their activity (i.e. have divided themselves into the 2 groups of 6), we next need to divide up each group. The face painting group must divide itself up into 3 indistinguishable groups of 2, and the basketball playing group must divide into 2 indistinguishable teams of 3. There are $\binom{12}{6\,6}$ ways in which the children can choose their activities. Then there are $\binom{6}{2\,2\,2} \div 3!$ ways in which the face painting group can pair off and $\binom{6}{3\,3} \div 2!$ ways in which the teams can be chosen for basketball (i.e., in which the basketball players can divide into 2 indistinguishable groups of 3). By the FCP, we see that the total number of ways in which the children can divide into the groups is

$$\binom{12}{6\,6} \times \left[\binom{6}{2\,2\,2} \times \frac{1}{3!}\right] \times \left[\binom{6}{3\,3} \times \frac{1}{2!}\right] = \frac{12!}{6!6!} \times \frac{6!}{2!2!2!} \times \frac{1}{3!} \times \frac{6!}{3!3!} \times \frac{1}{2!} = \frac{12!}{(2!)^4(3!)^3}$$

Once we've done all the cancelling out, we get the same answer as before.

Notice: It will not always be true that every group is in some set of indistinguishable groups. It may be only some. For instance, we might have had 3 indistinguishable pairs of face painters, but with the other 6 either all playing in a single group, or playing in 2 *distinguishable* groups.

Let's look at a couple of other examples, just to make sure these concepts are clear.

Example 1.26. *(a) The 4 members of a debating team are going to hone their skills by debating whether university tuition should be free for everyone. Two people will argue in favour of free tuition while the other two will present the arguments against. In how many different ways can the debating team be divided into the 2 sides for this debate?*

Here, we're dividing 4 people into 2 distinguishable groups: a 'for' group and an 'against' group. That is, the groups are doing different things. Therefore there are $\binom{4}{2\ 2} = \frac{4!}{2!2!} = \frac{4\times 3}{2} = 6$ ways to divide up the debating team.

Example 1.26. *(b) Four people are going to play a game of* euchre*, a card game in which two teams of 2 players each play against each other. In how many distinct ways can the teams be formed?*

Here, we're again dividing 4 people into 2 groups of 2, but this time the groups will do the same thing: be together on a team, playing euchre against the 'other' team. Therefore the groups are indistinguishable, so there are $\binom{4}{2\ 2} \div 2! = \frac{4!}{2!2!} \times \frac{1}{2!} = 3$ ways to form the 2 teams.

Example 1.27. *There are 9 students in a small class. The students are required to do a group project. The professor has decreed that there shall be 3 groups, with 3 students in each group.*

(a) In how many different ways can the students divide themselves into groups for the project?

Here, each group will "work together on a project". The groups are all the same size and are doing the same thing, so they are indistinguishable. (And there are 3 such groups.) Therefore the number of ways the students can divide themselves up is:

$$\binom{9}{3\ 3\ 3} \div 3! = \frac{9!}{3!3!3!} \times \frac{1}{3!} = \frac{\not{9} \times 8 \times 7 \times \not{6} \times 5 \times \not{4} \times \not{3}!}{\not{3} \times \not{2} \times 1 \times \not{3} \times \not{2} \times 1 \times \not{3} \times \not{2} \times 1 \times \not{3}!} = 8 \times 7 \times 5 = 4 \times 7 \times 10 = 280$$

Example 1.27. *(b) Suppose the professor has announced one topic, and all groups are to do a project on this same topic. In how many distinct ways can the project groups be formed?*

Now, each groups will "work together on a project about the assigned topic". The groups are still indistinguishable (i.e. still all doing the same thing), so nothing has changed from part (a). The groups can be formed in 280 ways.

Example 1.27. *(c) Suppose the professor has announced three topics. In how many ways can the groups be formed if*

(i) each groups will do a project on a different topic?

(ii) the class will vote to select one of the topics and then all groups will do projects on that topic?

(iii) each group may select any one of the topics (so more than 1 group may select the same topic)?

(i) *Approach 1:*
Now, each group works on a different topic, so we're forming distinguishable groups. That is, the students will divide themselves up into "a group for topic A", "a group for topic B" and "a group for topic C". The number of ways in which these groups can be formed is:

$$\binom{9}{3\ 3\ 3} = \frac{9!}{3!3!3!} = \frac{\not 9 \times 8 \times 7 \times 6 \times 5 \times \not 4 \times \not 3!}{\not 3 \times \not 2 \times 1\ \not 3 \times \not 2 \times 1 \times \not 3!} = 8 \times 7 \times 6 \times 5 = 1680$$

Approach 2:
Form indistinguishable groups as in part (a) (or part (b)) and then assign topics, i.e. arrange the 3 topics among the 3 groups. This can be done in $280 \times 3! = 280 \times 6 = 1680$ different ways.

(ii) There are 3 different ways the class might select the one topic that all groups will do a project on. That is, the groups are indistinguishable, as in (a), but we also have to account for the selection of the topic. By the FCP, there are $3 \times 280 = 840$ different ways in which the class can select 1 of the 3 topics and divide into indistinguishable groups.

(iii) This is actually quite complicated. We have to consider different cases, based on how many of the topics are chosen by groups.

Case 1: All groups choose the same topic
This is the same as part (c)(ii). There are 3 different topics that all 3 groups might choose and then the groups, all working on that same topic, are indistinguishable. So there are 840 different ways in which the groups can all work on the same topic.

Case 2: All groups choose different topics
This is the same as part (c)(i), in which each group works on a different topic. So there are 1680 ways to do this.

Case 3: 2 groups choose 1 topic and 1 group chooses another
First, there are 3 ways to choose which topic 2 of the groups will do, and then 2 ways to choose which (other) topic only one group does. (Or, $\binom{3}{2} \times \binom{2}{1}$ ways to choose which 2 topics will be done and then choose which one is done by 2 groups, which is again 3×2; or $\binom{3}{1\ 1\ 1} = \frac{3!}{1!1!1!} = 3!$ ways to divide the 3 topics into 3 distinguishable groups – 1 which 2 groups will do, 1 which 1 group will do and 1 which no groups will do.) So there are 6 ways to decide how many groups do each topic.

Then, the students will divide up into 2 indistinguishable groups of 3 (the 2 groups that do the same topic) and one different group (that does a different topic), so the number of ways of forming the groups is $\binom{9}{3\ 3\ 3} \div 2! = \frac{1680}{2!} = 840$ (using the fact that in (c)(i) we found that $\binom{9}{3\ 3\ 3} = 1680$). (*Notice:* We could also have found this by realizing that $\binom{9}{3\ 3\ 3} \div 2! = \left[\binom{9}{3\ 3\ 3} \div 3!\right] \times 3$ which is just the same calculation we did in (c)(ii).) So in total for this case there are $6 \times 840 = 5040$ ways to decide how many groups do each topic and then form the groups.

Altogether, we see that there are $840 + 1680 + 5040 = 7560$ distinct ways in which groups can be formed and topics chosen when each group is free to choose any of the 3 suggested topics.

Example 1.27. *(d) Repeat part (c), all parts, for the situation in which the prof announces 4 topics, instead of just 3.*

So we still have the 9 students forming 3 groups of size 3 for the project, and we know that they can form indistinguishable groups is 280 distinct ways. Now, the groups need topics, and there are 4 topics to choose among.

(i) If each group will do a project on a different topic, then we can think of forming indistinguishable groups and then assigning (i.e. arranging) 3 of the 4 topics among them. Therefore groups can be formed and topics chosen in $280 \times \frac{4!}{(4-3)!} = 280 \times 4! = 280 \times 24 = 6720$ different ways.

(ii) After forming indistinguishable groups, there are 4 different ways in which the class could select the one topic which all groups will do a project on, so there are $280 \times 4 = 1120$ different ways the groups can be formed and the single topic be chosen.

(iii) If repetition is allowed, but not required, among project topics, then we have the same 3 cases as before. If all 3 groups do different topics, there are 6720 distinct ways to form indistinguishable groups and then each group choose a different topic (from (d)(i)). If all 3 groups do the same topic, there are 1120 ways in which indistinguishable groups can be formed and the same topic chosen by all groups. If each groups is free to select any topic, with repetition allowed but not required, then as in (c) the only other possibility is that 2 groups do the same topic and 1 does a different topic. After forming indistinguishable groups, there are $4 \times 3 = 12$ ways to pick one of the 4 topics to be done by 2 groups and another to be done by the other group, and then $\binom{3}{2} = 3$ ways to decide which 2 groups will both do the topic which will be done by 2 groups. So we see that there are $280 \times 12 \times 3 = 10080$ distinct ways in which indistinguishable groups can be formed and then the topics chosen and assigned (with exactly 2 groups doing the same topic).

Therefore in total the number of ways in which groups can be formed and project topics chosen, when each group is free to choose any of the 4 suggested topics, is given by

$$6720 + 1120 + 10080 = 17920$$

Math 1228A/B Online

Lecture 10:
Free Distributions, and Some Review

(text reference: Section 1.7)

1.7 Miscellaneous Counting Problems

There is not very much that's new here. For the most part, in this section the text presents problems which are a bit more complicated, and which may require that we use several of our counting techniques in the same problem.

There is, though, one new idea in this section. We learn how to approach a slightly different kind of counting problem.

Free Distributions

Recall Example 1.11, which we looked at in section 1.3 (Lecture 4, pg. 20): *In how many distinct ways can 5 identical candies be distributed between 2 children?*

We approached this previously by recognizing that, since all 5 candies are to be distributed and there are only 2 children, child 2 will receive all of the candies that aren't given to child 1. Therefore the number of ways of distributing the candies is given by the number of different choices for the decision 'how many candies will be given to child 1?'. Since there are 5 candies, child 1 may be given 0, 1, 2, 3, 4 or 5 candies and hence there are 6 distinct ways to distribute the candies. (That is, after the 'child 1' decision is made, there is only 1 way to decide how many candies child 2 gets: give child 2 all the remaining candies.)

There are 2 points of interest here. First, the distinguishing characteristic of this problem is that the 5 candies are all identical. If we had 5 *distinct* candies to distribute, it would be a quite different problem, because we would also have to decide *which* candies are given to each child, as well as *how many.* Since the candies are all the same here, we are *only* concerned with *how many* candies each receives.

The other important point is that we were able to find an easy way to count the number of possible distributions because there were only 2 children. It is easy to see how to generalize this approach to problems involving a different number of identical objects to be distributed between 2 children/people/locations or whatever. (Let's say, between 2 *labels.*) If there are k identical objects, then there are $k+1$ choices of how many to allocate to one label (i.e. 0 or 1 or 2 or ... or k), and then the rest will be allocated to the other. But if there are more than 2 labels (i.e. potential recipients of the objects), the problem becomes much more complex.

Consider, for instance, the problem of counting the number of ways in which 5 identical candies can be distributed among 3 children. There are still 6 choices for the decision 'how many will child 1 receive?'. However, after that decision has been made there is another (non-trivial) decision still to be made: 'how will the remaining candies be distributed between child 2 and child 3?'. Unless all of the candies were given to child 1, this subsequent counting problem is just as complicated as the original (2 child) problem. The number of ways of distributing the 'remaining' candies depends on how many there are, so we would need to consider 6 different cases, one for each of the choices of 'how many will child 1 receive?'. If child 1 receives i candies (which can happen in only 1 way – give any i of the identical candies to child 1), then there are $k-i$ candies remaining, so there are $k-i+1$ ways to distribute these candies (i.e., child 2 may receive 0 or 1 or ... or $k-i$ of them, and then the rest will be given to child 3). By considering the 6 different cases (child 1 receives 0 or 1 or ... or 5 candies, leaving 5 or 4 or ... or 1 or 0 to distribute between child 2 and child 3), we see that there are

$$(5+1)+(4+1)+(3+1)+(2+1)+(1+1)+(0+1)=6+5+4+3+2+1=21$$

different ways in which the candy might be distributed among the 3 children.

But now what if there were more children? Using this approach, each additional child gives rise to an extra level of cases to consider. For instance, with 4 children, each of the 6 cases based on the number of candies which might be given to the first child leaves a 3 child problem, giving rise to a set of subcases ($k - i + 1$ of them, if child 1 receives i of the k candies), based on the number of candies given to child 2. Each of these subcases is now an easy 2 child problem (how should the remaining candies be distributed between child 3 and child 4?), but the need to enumerate all of the various cases and sub-cases makes the problem very complex. And that's with only 4 children. Imagine how difficult it would be to apply this approach to a problem involving 10 children!

We see that the approach used in the simple problem cannot easily be extended to a workable approach for the more general problem *'In how many ways can k identical objects be distribued among r different labels?'*. That is, we would like to find a way of counting the number of different ways in which k identical objects may be assigned labels of r distinct types, where each label type may be used as many times as desired. Clearly the approach we used in the original problem is not what we want to use for this more general problem.

Note: Although we express this problem as one of 'labelling' the objects, this is *not* a labelling problem of the sort we studied in the previous section. There, we had distinct objects which were to be labelled and we knew how many of each label were to be used. Here, we are labelling *identical* objects, and may assign any number of the objects to the same label.

So let's go back to our original problem and look for a different approach which will generalize more easily. We have 5 identical objects (candies) and 2 labels (child 1 and child 2). The problem is to count the number of different ways of determining how many objects are assigned to each label. (In the current problem, we are counting the number of different ways of determining how many candies are given to child 1, i.e. are assigned to the 'child 1' label, and how many candies are given to child 2, i.e. are assigned to the 'child 2' label.)

Suppose we make a list. The list starts with a label, 'child 1'. After this label, we will put an $\times$ for each candy which child 1 will receive. Then we put a 'child 2' label, followed by an $\times$ for each candy which child 2 receives. For instance, one possible list is

$$\text{child 1}, \times, \times, \times, \text{child 2}, \times, \times$$

which corresponds to the distribution 'child 1 gets 3 candies and child 2 gets 2'. Each different distribution of the candies gives rise to a different list. If a child receives no candies, the corresponding list has no $\times$'s after that child's label. For instance, the list

$$\text{child 1}, \text{child 2}, \times, \times, \times, \times, \times$$

represents the allocation 'child 2 gets all 5 candies'.

The problem now is to count how many such lists can be formed, so we need to think about the characteristics of these lists. Each lists consists of 2 labels (child 1 and child 2) and 5 $\times$'s (for the 5 identical candies). Thus there are 7 positions on each list, and we are concerned with the different possible ways of filling these positions.

One position, though, is the same in all of the lists. The list *always* starts with the child 1 label. It is only the other 6 positions which change from one list to another. So what we really need to count is the number of ways in which the 5 $\times$'s and 1 other label can be put into the 6 other positions on the list. (*Notice:* If we also consider lists which start with the child 2 label, we will be double counting. For instance, the lists 'child 1, $\times, \times, \times$, child 2, $\times, \times$' and 'child 2, $\times, \times$, child 1, $\times, \times, \times$' both correspond to the *same* allocation, 'child 1 gets 3 candies and child 2 gets 2', so we don't want to count both of these lists, only 1 of them. This means that we need to consider only one *ordering* of the labels – always the 'child 1' label at the beginning and the 'child 2' label somewhere later in the list.)

Since the $\times$'s are all the same (i.e. the candies are all identical), all we really need to do to make a list is to decide 'which 5 positions in the list will contain $\times$'s?'. Once we have chosen 5 of the 6 available places to put the 5 $\times$'s, we have fully determined what the list looks like, because the child 1 label is (always) in the first position and the child 2 label goes in the only position left in the list. In this way, we see that the number of possible lists is given by $\binom{6}{5} = 6$. That is, we see that there are 6 ways to distribute the candy. (Although we used a very different approach, we of course found the same answer as before.)

Okay, now what about the 3 child problem? Can we use this approach for that? Yes, and it's just as easy as the 2 child problem. We still have 5 $\times$'s (i.e. 5 identical candies), but now we have 3 distinct labels, one for each of the 3 children. Of course, the list must still start with the child 1 label, so although we have $5+3=8$ positions in the list, the first is already filled, so we only need to place 5 identical $\times$'s and $3-1=2$ labels in the remaining 7 positions on the list. Also, we only want to consider lists in which the *order* of the labels is fixed, to avoid overcounting. That is, if we choose 5 (out of 7) places in which to put $\times$'s, then we will always put the child 2 label in the first remaining space and the child 3 label in the last remaining space, for the same reason that it's always the child 1 label that goes at the start of the list. Permuting the children (i.e. the labels) does *not* give rise to different ways of allocating the candies, it simply counts each possible distribution multiple times, which we don't want to do. So as soon as we have decided which positions in the list will contain $\times$'s, we have fully determined the distribution of the candies. Since there are 7 positions available to put the 5 $\times$'s in, and the $\times$'s are all the same, the number of possible lists is $\binom{7}{5}$. That is, the number of different ways to distribute 5 identical candies among 3 children is given by

$$\binom{7}{5} = \frac{7!}{5!2!} = \frac{7 \times 6 \times 5!}{5! \times 2 \times 1} = 21$$

Now, let's consider the most general problem. Suppose we have k identical candies to distribute among r children. In how many ways can this be done? Again, we think about the number of different possible lists we can make. Each list will contain k $\times$'s and r labels. However the list always starts with the child 1 label, and the order of the other $r-1$ labels is fixed, so the number of different $k+r$-position lists is simply the number of ways to choose k of the $k+r-1$ positions after the child 1 label in which to put the $\times$'s. Once we have determined the placement of the $\times$'s, we just put the other labels, in the pre-determined order, into the empty spaces and we know how many candies each child will receive. The number of different ways of distributing the candies is simply the number of different lists we can form in this way, i.e. the number of different ways of choosing where to put the $\times$'s. Since there are k $\times$'s, and $k+r-1$ available positions (because the first position in the list cannot contain an $\times$), we see that there are $\binom{k+r-1}{k}$ different possible lists, i.e. different ways to distribute the candy.

Using this approach, the 10 child problem is just as easy as the 2 child problem. To count the number of ways in which 5 identical candies can be distributed among 10 children, we simply think about forming lists involving 5 $\times$'s and 10 labels, where each list must start with the child 1 label and the other labels always appear in the same order on all lists. There are 15 positions on the list, but only 14 of them are available to put the 5 $\times$'s in. Therefore the number of possible lists, i.e. the number of ways of distributing the candy, is $\binom{5+10-1}{5} = \binom{14}{5}$.

Of course, we don't need to be thinking about distributing identical candies among a group of children. What we have here is a general approach which can be used (easily) for any counting problem involving k identical objects to be distributed among r different labels, where each label may be used any number of times (i.e., where any number of the objects can be assigned to a label). We call these problems *Free Distribution* problems.

> **Theorem: *Free Distributions*** The number of distinct ways of distributing (all of) k identical objects among r different labels is given by
>
> $$\binom{k+r-1}{k}$$

Let's look at another example involving free distributions.

Example 1.28. *A family of 4 sits down to a breakfast of pancakes, sausages and fruit. There are 10 (identical) pancakes on the platter before they start, and by the time they finish, all of the pancakes have been eaten.*
(a) In how many different ways might the pancakes have been distributed among the 4 people?
(b) In how many ways might the pancakes have been distributed if it is known that Mary didn't have any pancakes and the others had at least 2 each?

(a) We have to determine the number of ways to distribute 10 identical pancakes among 4 different people. We don't know anything about how many pancakes each person ate. Maybe little Johnny ate all of them and the rest just ate sausages and fruit. Or maybe Dad and Mary had 3 each, Johnny ate 4 and Mom didn't have any. There are many ways the 10 pancakes might have been eaten by the 4 people. We need to figure out how many.

Since the pancakes are all identical, this is a free distribution problem. The number of copies of the identical objects, in this case the pancakes, is what we call k, so we have $k = 10$. And the number of people we're going to distribute them to is r, so here $r = 4$. Now, we just use the formula. The number of ways the pancakes might have been distributed is

$$\binom{10+4-1}{10} = \binom{13}{10} = \frac{13!}{10!3!} = \frac{13 \times 12 \times 11}{3 \times 2 \times 1} = 13 \times 2 \times 11 = 286$$

(b) Now, we're told that Mary didn't have any pancakes, while the others had at least 2 each, and once again we need to determine the number of different ways the pancakes might have been eaten.

If Mary didn't have any pancakes, then we just won't distribute any to her. That is, the pancakes were actually only distributed among 3 people. Also, we know that each of these 3 people ate at least 2 pancakes. That accounts for 2×3 of the pancakes, so there are only $10 - 6 = 4$ pancakes unaccounted for. We just need to count the number of ways of distributing these $k = 4$ pancakes among the $r = 3$ people, which is

$$\binom{4+3-1}{4} = \binom{6}{4} = \frac{6!}{4!2!} = \frac{6 \times 5}{2} = 15$$

<u>... And Some Review</u>

The rest of this section can be thought of as 'Chapter 1 Review'. For each problem, we need to determine what kind of problem it is, i.e. which of our counting approaches we need to use. The problems may require more than one approach, or may involve complications that require us to use an approach in a slightly different way.

Example 1.29. *On a certain television game show, some audience members are chosen to participate in particular individual challenges. At the taping of today's show, there were 100 people in the audience, and 8 challenges were available. On this game show, each chosen person faces a different challenge. Each challenger either succeeds or fails at his or her challenge. There is only time for 5 challenges during a show. In terms of who won or lost what challenge, how many different possibilities were there for today's show?*

We are interested in who won or lost what challenge when today's show was taped. For instance, a description of one of the possibilities might be something like

> 'John succeeded at the spelling challenge, Mary beat the clock on the water challenge and Susan found the hidden word, but Clayton was unable to solve the logic problem and Sandy failed the strength challenge.'

Thinking about a description like this, though, may make the counting problem seem unnecessarily complicated. The best way to approach a complicated problem is often to think about 'what decisions have to be made?' and determine the number of ways in which each can be made, i.e., the number of choices available for each decision. As previously observed, we are interested in who won or lost what challenge when today's show was taped. We can view this as 3 decisions that were made: *who participated?*, *what challenge did each face?*, and *did each challenger win or lose?*.

Decision 1: Who were the challengers?
We are told that there is time for 5 challenges during a show, so there must have been 5 challengers who participated. These challengers were selected from the 100 audience members, so there were $\binom{100}{5}$ different possible combinations of challengers that could have been selected. That is, when the decision 'who will participate in the challenges?' was made, there were $\binom{100}{5}$ different choices available.

Decision 2: What challenge did each face?
One way to approach this is to break it down further into 2 sub-decisions: 'which 5 challenges were used?' and 'which person faced each of these challenges?'. There were 8 challenges available, of which 5 different challenges were used. Therefore there are $\binom{8}{5}$ ways to select the 5 challenges used. Once the 5 challenges and the 5 challengers had been chosen, we simply need to think about how the challenges were assigned to the challengers. The ways this could have been done are the various possible permutations of the challenges (or the challengers). Any time we need to assign n different objects/tasks/people, etc., among n different people/positions/activities, etc., there are $n!$ ways in which it can be done. Think of the order of 1 group as being fixed and assign the members of the other group to them. For instance, if we think of the 5 challengers in some specific order, and of assigning the 5 different challenges to them, there are 5! ways to pair up the 5 challenges with the 5 challengers.

We see that there were $\binom{8}{5} \times 5! = \frac{8!}{5!3!} \times 5! = \frac{8!}{3!}$ different ways in which to make the 2 decisions 'which 5 of the 8 challenges?' and 'who faces which challenge?', i.e. to decide what challenge each of the 5 chosen people would face.

Of course, this is just the number of different permutations of 5 of the 8 challenges. That is, we could also have counted the number of ways of determining what challenge each of the chosen people faced more directly, by thinking of it as assigning 5 of the 8 challenges to the 5 people, i.e., by recognizing that it is an arrangement of 5 of the 8 challenges. We know that arranging k out of n objects can be done in $\frac{n!}{(n-k)!}$ ways, so there are $\frac{8!}{(8-5)!} = \frac{8!}{3!}$ ways to assign 5 of the 8 available challenges to the 5 chosen challengers. Therefore at the time of determining 'what challenge will each participant face?', there were $\frac{8!}{3!}$ choices available.

Notice: We have *not* thought about the order in which the challengers appeared on the show. That is, we have 'arranged' the challenges among the challengers, but we have not arranged the challengers. This is because the question we were asked specified that we were concerned with 'who won or lost what challenge', and did not ask us to consider in what order the challengers appeared on the show. If we wanted to also take that into consideration, for instance if we were asked for the 'number of different shows possible', then we would need another 5! multiplier, because we would have another decision, 'in what order did the contestants appear?', which could be made in any of 5! ways.

Decision 3: Who won and who lost?
So far, we have determined who the challengers were and what challenge each faced. We need to consider whether each of these challengers won or lost his or her challenge. For each challenger there are 2 possible outcomes: the challenger succeeded or the challenger failed. That is, this decision was actually 5 sub-decisions, one for each challenger, of the form 'does this challenger succeed or fail?'. With 2 choices for each of these 5 sub-decisions, we see that there are $2 \times 2 \times 2 \times 2 \times 2 = 2^5$ different possible results from the 5 challengers facing their assigned challenges. That is, for the 'decision' of 'which challengers won and which lost?', there were 2^5 choices available.

Conclusion:
Since all 3 'decisions' needed to be made, we use the FCP to find the number of possibilities that there were for today's show. In total, the number of ways of choosing challengers, assigning challenges and observing whether each succeeds or fails, i.e. the number of possibilities for today's show, was

$$\binom{100}{5} \times \frac{8!}{3!} \times 2^5$$

Example 1.30. *Recall Example 1.19, in which a servant must arrange some 7 of a set of 12 different flagons around the Round Table, for use by Arthur and 6 of his knights while they discuss strategy for an upcoming battle. The servant has selected and placed the flagons, and the men are now coming into the room. In how many distinct ways can the men position themselves around the table so that each is sitting in front of one of the flagons?*

Ignore the flagons for a moment. Suppose that all the servant had done to set up the room was place 7 identical chairs around the Round Table. Then the number of possible arrangements of the 7 men around the table would simply be the number of circular permutations of the 7 men, which is 6!.

How does the problem change when the 7 distinct flagons are on the table? For each of the 6! circular arrangements of the 7 men, there are 7 different assignments of flagons possible. That is, since the arrangement of the flagons is fixed, then once we have 'chosen' a circular arrangement of the men, we need to determine 'who will sit in front of this one particular flagon?', but only for one of the flagons. For instance, suppose one of the flagons bears a picture of a dragon. There are 7 choices for which man will have the flagon with the dragon. After that, there are no more decisions to be made, because we have already determined both the order of the flagons and the order of the men. If Arthur sits in front of the flagon with the dragon, then whoever is to be on his left will have the flagon which has been placed to the left of the dragon flagon, and so on. Therefore there are $6! \times 7 = 7!$ ways in which the 7 men can position themselves in front of the 7 flagons around the table.

Notice what this means: the number of ways of arranging the 7 men is 7!, i.e. we are counting *linear* permutations, not *circular* permutations. By distinguishing among the 7 places around the table (by putting a different flagon at each place), the problem loses its circular aspect. What matters here is *not* just 'who is next to whom' at the table, but rather which *specific* place at the table each occupies. If we need to arrange n distinct objects (e.g. 7 men) in n distinct positions (e.g. assigned to 7 different flagons), then there are $n!$ ways in which it can be done, even if the n distinct positions happen to be in a circle rather than in a line.

To finish up the chapter, let's go back to Suzie's birthday party and consider some more complications.

Example 1.31. *Recall Example 1.23 in which the children at Suzie's party were going to divide themselves up into groups to play croquet (5 children), badminton (4 children) and basketball (3 children).*

(a) If the Rodriguez triplets insist on all playing the same game, in how many ways can the children be divided up to play the games?

Here, we need to consider 3 different cases. That is, we could approach this by saying that first there is a decision about which game the triplets will play, and then there is a decision about how the other 9 children are divided up. However, this is not a situation in which we can use the FCP, because the number of ways in which the second decision can be made depends on how the first decision is made. That is, the sizes of the groups the remaining 9 children will be divided up into changes, depending on which game the triplets are playing. Therefore we need to consider several cases, one for each way of making this first decision.

If the triplets play croquet, then we need 2 more croquet players, 4 badminton players and 3 basketball players. In this case, the number of ways of dividing up the rest of the children is $\binom{9}{2\ 4\ 3}$.

If the triplets play badminton, then we still need 5 croquet players, 1 badminton player and 3 basketball players, so the number of ways of dividing up the other children is $\binom{9}{5\ 1\ 3}$.

If the triplets play basketball, then we simply need to divide the remaining children up into 5 to play croquet and 4 to play badminton, which can be done in $\binom{9}{5\ 4}$ ways.

In total, the number of ways to divide up the children to play the games is

$$\binom{9}{2\ 4\ 3} + \binom{9}{5\ 1\ 3} + \binom{9}{5\ 4}$$

Notice: We could also write the last term here as $\binom{9}{5\ 4\ 0}$, or even just as $\binom{9}{5}$ or $\binom{9}{4}$. All of these variations have the same value, $\frac{9!}{5!4!}$.

Example 1.31. *(b)In how many ways can the children divide themselves up if the triplets insist on all playing different games?*

At first glance, you might think this is a more complicated problem, because now we need to think about 'which triplet plays which game?'. However, this problem is actually easier than the last one. This time, no matter how the triplets are assigned, as long as they are assigned one to each game, then we always need to divide the other 9 children up into 4 to play croquet, 3 to play badminton and 2 to play basketball. That is, this time the decision 'how will the other 9 children be divided up?' is *not* dependent on how the decision 'how are the triplets assigned to the games?' is made, so the FCP applies.

There are 3! ways to assign the 3 triplets to the 3 games (one per game) and $\binom{9}{4\ 3\ 2}$ ways to divide up the rest of the children, so the total number of ways the children can divide themselves up is

$$3! \times \binom{9}{4\ 3\ 2} = \frac{3!9!}{4!3!2!} = \frac{9!}{4!2!}$$

Example 1.32. *While the children at Suzie's party were playing outside, Mary and Jenny had an argument and are mad at one another. It is now time for the birthday cake. The 12 children are going to sit around a single table. In order to prevent further trouble, Suzie's mother doesn't want*

Mary and Jenny to be sitting either beside one another or directly across from one another.

(a) How many acceptable seating arrangements are there?

We need to determine the number of ways in which the 12 children at the party can be arranged at the table so that Mary and Jenny are not sitting either side-by-side or directly across from one another. Actually counting the number of seating arrangements which satisfy these criteria might be difficult, so we won't take that approach. Instead, we'll count the total number of possible seating arrangements, and also count the number which are unacceptable.

There are $(12 - 1)! = 11!$ different circular permutations of the 12 children, i.e., 11! possible seating arrangements at the table (if we ignore the restrictions about where Mary and Jenny may sit relative to one another).

How many of these arrangements have Mary and Jenny sitting side-by-side? We consider Mary and Jenny as a single object, so there are 11 objects to arrange around the table, and then we arrange the one Mary-and-Jenny object. There are $(11 - 1)! = 10!$ circular permutations of 11 objects (10 other children plus the one Mary-and-Jenny object), and $2! = 2$ ways to arrange Mary and Jenny within their object (i.e., Mary to the left of Jenny or Jenny to the left of Mary). Therefore there are $10! \times 2$ seating arrangements in which Mary and Jenny are sitting beside one another.

Now, how many of the seating arrangements have Mary and Jenny sitting directly across from one another? We did something like this before, with Arthur and Guinevere. Using the same approach we used there, there are $(11 - 1)! = 10!$ ways to arrange Mary and the other 10 children at the table, and then only 1 way to position Jenny – directly across from Mary.

Alternatively, we could think of this in the same way as we approached the side-by-side problem. If Mary and Jenny are considered as a single object, then there are $(11 - 1)! = 10!$ ways to arrange the 10 other children and the Mary-and-Jenny object at the table. Now, however, we must be very careful. There's only 1 way to arrange the Mary-and-Jenny object – across from one another. (If you think that there are 2 ways to do this then you are double counting, because we've already taken that into account. That is, switching the places of Mary and Jenny is just the same as moving the Mary-and-Jenny object directly across the circle. And that's another one of the circular permutations that we already counted.)

By either of these approaches, we see that there are 10! seating arrangements in which Mary and Jenny are sitting across the table from one another. Therefore the total number of *unacceptable* seating arrangements is

$$10! \times 2 + 10! = 10! \times 2 + 10! \times 1 = 10! \times (2 + 1) = 10! \times 3$$

Therefore, subtracting these from the total number of possible seating arrangements, we see that the number of *acceptable* seating arrangements is

$$11! - 10! \times 3 = 10! \times 11 - 10! \times 3 = 10! \times (11 - 3) = 10! \times 8$$

Notice: It is important to realize that we are assuming here that when the 12 children sit around the table, it is not possible for 2 of them to be sitting side-by-side and at the same time be sitting across from one another. At a round table, or a rectangular table at which people are sitting on all 4 sides, this is true. If this assumption is not true, for instance if it's a long table with 6 children on each side and nobody on the ends, then we've overcounted the number of unacceptable arrangements, because if Mary and Jenny sit at the same end of the 2 sides (i.e. one on each side, at the very end) then they are sitting across from one another but are also, in circular terms, sitting beside one another. In this situation, we would need to consider things more carefully, considering 2 separate cases - Mary sitting at an end of a side, or Mary sitting not on an end. But we don't need to go there right now.

Also Notice: Having seen that the total number of unacceptable seating arrangements is $10! \times 3$, we can (perhaps more easily now) see how we could have directly counted the number of acceptable seating arrangements. Let Mary sit anywhere. This leaves 11 positions at the table. Of these, 3 are positions which Jenny can't take - to Mary's left, to Mary's right, and directly across from Mary. The other 8 positions are all places where Jenny could sit, so once Mary is seated, there are 8 choices for 'where can Jenny sit?'. Now, there are 10 places remaining, and 10 children to arrange in them. Since Mary and Jenny have already been placed, the remaining 10 places do not (themselves) form a circle, so the different arrangements of the other children are linear permutations, not circular permutations. Therefore there are 10! ways to arrange the other children, for a total of $10! \times 8$ acceptable seating arrangements.

Example 1.32. *(b) Is the answer to part (a) any different if Suzie is going to sit in a special 'Birthday Throne'?*

The real question here is whether it makes a difference, when counting circular permutations, if one (only) position in the circle is distinguished in some way. The answer is no.

To see this, let's think about a smaller problem. When we started looking at (i.e. counting) circular permutations, we first looked at the question 'in how many ways can 3 people be arranged in a circle?'. We drew 2 arrangements, ABC and ACB, reading clockwise around the circle, and observed that those were the only possibilities. When we drew these circles, we showed A at the top of each, but that wasn't important because we were only concerned with the circular ordering of the 3 people. The question about Suzie and the Birthday Throne is analogous to the question "in how many ways can A, B and C be arranged in a circle so that A is at the 'top' of the circle?". This is not a different question than the simpler one that didn't refer to the top of the circle. We can see that there are (still) 2 arrangements. Likewise, if there's a special chair for Suzie, that doesn't make any difference to the number of ways the 12 children can be arranged at the table. (We could imagine that we first have the children arrange themselves at the table, and then we simply move the Birthday Throne to wherever Suzie is. Or if the throne is already in place, imagine that the children form a circle, and then shuffle around the table/circle, maintaining their relative positions, until Suzie is at the place with the Birthday Throne.)

Distinguishing 1 position in a circle, or anchoring 1 object in a circular arrangement, does not change the problem in any significant way. In fact, for our general problem of counting circular permutations, we could think of the problem in the following way: Have any one of the n objects placed in any one position in the circle (e.g. have Suzie sit in the Birthday Throne). Now, arrange the other $n-1$ objects in the remaining $n-1$ positions. Because 1 object is fixed, the arrangement of the other $n-1$ objects is now a linear permutation, not a circular permutation. There are $(n-1)!$ ways to arrange the remaining objects.

Notice: We could have approached the 'Mary and Jenny across from one another' problem in this way. That is, we can think of having Mary and Jenny sit across from one another anywhere at the table, and then having the other 10 children arrange themselves in the remaining 10 (now distinct) places, which can be done in 10! ways.

Math 1228A/B Online

Lecture 11:
Basic Concepts of Probability

(text reference: Section 2.1)

2 Probability

For the next while, and actually for the entire rest of the course, we're going to be talking about probability. We need to start off by defining some basic concepts.

2.1 Sample Spaces and Events

In studying probability, we look at situations in which we don't *know* what's going to happen – but we *do* know all of the things that *could* happen.

In a **probabilistic experiment**, some procedure is performed in which some characteristic is observed or measured. *Chance* determines which one of a number of possibilities occurs. Some examples of probabilistic experiments are:

- toss 2 dice and observe what the sum of the numbers is
- draw a card from a deck and observe what card it is
- choose a student at random from the class and measure the student's height
- measure how long it takes to drive from home to school on a particular day

In each of these experiments, there is an element of *chance* which makes the result *uncertain*.

An **outcome** of an experiment is a particular result, i.e. observation or measurement, which *could* occur when the experiment is performed. For instance, if we draw a card from a standard deck, one possible outcome is "the King of diamonds was drawn". Another possible outcome is "the 3 of spades was drawn".

A **sample space** for an experiment is a set containing (descriptions of) all possible outcomes for the experiment. We always use S to denote the sample space.

In defining a sample space, there are 2 characteristics that the set S must have:

1. The set must be **complete**. That is, it must contain *every* possible outcome of the experiment. There must be no possibility of observing some outcome which does not appear anywhere in the sample space (i.e. is not described by any element of the set).
2. The elements of the set S must be **mutually exclusive**. That is, there must be *no overlap* between the elements of the set. It must not be possible to observe some outcome which is described by *more than one* element of the set.

These 2 conditions together ensure that each possible outcome of the experiment is described by one and only one (i.e. by exactly one) of the elements of S.

The elements of a sample space are called **sample points**.

To try to get a better understanding of what sets are and are not potential sample spaces for an experiment, let's look at a specific probabilistic experiment and consider a number of different sets we might define. For each, we will look at whether or not the set has the characteristics required in order to be a possible sample space for the experiment.

Example 2.1. *Consider the experiment "toss 2 dice and observe what the sum of the numbers is". Which of the following sets could be used as a Sample Space for this experiment?*
(a) $S_0 = A \times A$, *where* $A = \{1, 2, 3, 4, 5, 6\}$
(b) $S_1 = \{2, 3, ..., 12\}$
(c) $S_2 = \{$ *"sum is even", "sum is odd"*$\}$
(d) $S_3 = \{sum < 7, sum = 7, sum > 7\}$
(e) $S_4 = \{sum \leq 5, sum \geq 8\}$
(f) $S_5 = \{sum < 8, sum > 5\}$
(g) $S_6 = \{$ *"sum is even"*, $sum > 5\}$

We must consider each of the given sets separately, and determine whether each is a complete set of mutually exclusive descriptions of the possible outcomes when 2 dice are tossed.

(a) $S_0 = A \times A$, where $A = \{1, 2, 3, 4, 5, 6\}$.

Recall that the direct product of 2 sets is a set containing ordered pairs. We have the set

$$S_0 = \{(1,1), (1,2), ..., (1,6), (2,1), (2,2), ..., (2,6), ..., (6,1), ..., (6,6)\}$$

When we toss 2 dice, each die comes up as a number between 1 and 6. If we let (i, j) represent that the number on the first die is i and the number on the second die is j, then each of these pairs is a possible outcome of the experiment "toss 2 dice", and given the information which pair was observed, we can determine what the sum of the 2 dice is.

Question: When we toss 2 dice (together), is $(1, 2)$ a different outcome from $(2, 1)$?

The answer to this question is Yes. To see that this is true, consider tossing 2 dice, where there is a red die and a white die. If we define that (i, j) denotes the outcome 'i is the number showing on the red die and j is the number showing on the white die', then the outcome $(1, 2)$ occurs when the red die shows 1 and the green die shows 2. This is clearly a different outcome than $(2, 1)$, which occurs when the red die shows 2 and the white die shows 1.

What if the 2 dice are identical in appearance? Does that change anything? No. For instance, we could distinguish between the 2 dice by considering which lands further to the left when the dice are tossed. If we define (i, j) so that i is the number showing on the left die and j is the number showing on the right die, then again we see that $(1, 2)$ and $(2, 1)$ are different outcomes.

It doesn't matter how, or whether, we distinguish between the 2 dice. When we define the outcomes of a toss of 2 dice as (i, j), $(1, 2)$ and $(2, 1)$ are always considered to be *different* outcomes, and likewise for any $i \neq j$.

Now, let's get back to the question of whether the set of all outcomes defined in this way could be a sample space for the experiment. Clearly the set S_0 contains *all* of the possibilities when we toss 2 dice, and (as discussed above) each possible outcome is described by only one of these pairs. So S_0 is a complete set of mutually exclusive descriptions of the possible outcomes of the experiment, and therefore set S_0 could be used as a Sample Space for the experiment.

(b) $S_1 = \{2, 3, ..., 12\}$

When we toss 2 dice and observe the sum, that sum is always an integer. The lowest sum we can possibly get is 2 and the highest sum we can get is 12. That is, we're adding up 2 numbers, each of which is an integer between 1 and 6, so $1 + 1 = 2$ is the smallest possible value and $6 + 6 = 12$ is the highest possible value. (All integer values between these are possible, too.)

The set S_1 is *complete* because any time we toss 2 dice, the numbers on the dice always produce a sum which is a number contained in this set. Also, the elements of this set are *mutually exclusive* because the sum of the numbers on the 2 dice on any particular toss cannot be more than one of these values at the same time. (We can have more than one toss giving the same sum, but we cannot have a single toss giving more than one sum. For instance, tossing a 1 and a 3, tossing a 3 and a 1 or tossing two 2's all give a sum of 4. Whenever any one of these is tossed, **only** '4' describes what the sum is.) We see that S_1 is a complete set of mutually exclusive possibilities, so S_1 is a possible Sample Space for the experiment.

(c) $S_2 = \{$"sum is even", "sum is odd"$\}$

Every possible outcome of tossing two dice produces a sum which is either even or odd, and clearly cannot be both. That is, each possible outcome of the experiment fits one, and only one, of the 2 descriptions 'sum is even' and 'sum is odd'. Therefore S_2 is another *complete* set of *mutually exclusive* descriptions of the possible outcomes and can be used as a Sample Space for the experiment.

(d) $S_3 = \{sum < 7, sum = 7, sum > 7\}$

When we toss 2 dice, the sum of the numbers will always fit exactly one of these descriptions. It is not possible to observe a sum for which none of these descriptions applies, and we cannot have a sum for which more than one of these descriptions applies. So this is yet another *complete* set of *mutually exclusive* descriptions of the possible outcomes and therefore is another possible Sample Space for this experiment.

(e) $S_4 = \{sum \leq 5, sum \geq 8\}$

This set contains *mutually exclusive* descriptions, because clearly when we toss 2 dice, the result cannot be a sum which is both less than or equal to 5 and also greater than or equal to 8. However, this set is *not complete*, because when we toss 2 dice, we could get a sum which doesn't fit either of these descriptions. For instance, if we roll a 1 and a 6 then the sum is 7, which is neither less than or equal to 5 nor greater than or equal to 8. Therefore this set *cannot* be used as a Sample Space for this experiment.

(f) $S_5 = \{sum < 8, sum > 5\}$

This set *is complete*, because every possible sum is included. That is, there is no sum we could get which is *not* less than 8 and *also is not* greater than 5. (Every sum which is not less than 8 must be at least as big as 8 and hence bigger than 5. Similarly, every sum that is not greater than 5 must be at most 5 and therefore less than 8.) However, the elements of S_5 are *not mutually exclusive* because some possible outcomes fit both descriptions. For instance, a sum of 7 is both less than 8 and greater than 5. Therefore this set is *not* a possible Sample Space for the experiment.

(g) $S_6 = \{$"sum is even"$, sum > 5\}$

This set is *not complete*. For instance, we could roll a 1 and a 2 for a sum of 3, which is neither even nor greater than 5. Also, this set is *not mutually exclusive*. For instance, if we roll two 3's, we get a sum of 6, which is both even and greater than 5. So this set has neither of the properties it would need to possess in order to be eligible to be used as a Sample Space for the experiment.

We have seen here that there can be more than one way to define a suitable sample space for a particular experiment. Which one we choose depends on what we're trying to do. If *all* we care about is whether the sum is even or odd, then S_2 could be used. But if we may want to know about other characteristics of the sum, then S_1 would be better. And if we might want to know about the particular numbers rolled, not just the sum, then we would want to use S_0. Also, in section 2.3 (Lecture 13), we will see that in some situations we need to describe the outcomes in the most basic possible terms, and use a sample space in which there is a different sample point for each possible outcome, in order to be able to calculate probabilities.

An **event** in an experiment is a subset of the sample space S for the experiment. Since these are sets, we use capital letters to denote events. Often we use E for one event, or E and F if we are interested in 2 events. For several events, we could use E_1, E_2, E_3, ... However, as usual, it's often useful to use a letter which will help you remember what the event is. For instance, in the experiment "Choose a student from the class and observe which student it is", we could use the set of all students as the sample space. One event we might be interested in is the set of all male students in the class. If we use M to denote this set, it helps us remember that this is the event "the chosen student is male".

Since events are, by definition, subsets of a universal set which is the sample space, then we can talk about complementary events. That is, for any event E, E^c is the event that contains all of the sample points which are *not* in E, i.e. the event that E does *not* occur. For instance, in the experiment mentioned above, M^c would be the event that the chosen student is not male, i.e. is female.

Example 2.2. *Suppose we perform the experiment "toss 2 dice and observe what the sum of the numbers is". Using the sample space $S = \{2, 3, ..., 12\}$, find the subset of S corresponding to each of the following events:*
(a) the sum is even;
(b) the sum is at least 9;
(c) the sum is not an odd number less than 9;
(d) the sum is an even number and is at least 9.

Recall that with this sample space, the sample points represent the possible sums which might be observed.

(a) the sum is even;
Let E be the event that the sum is even. Then E is the subset of S containing all even numbers that appear in S, so $E = \{2, 4, 6, 8, 10, 12\}$.

(b) the sum is at least 9;
Let N be the event that the sum is at least 9, i.e. $sum \geq 9$. Then N is the subset of S containing all of the numbers in S which are 9 or bigger, so $N = \{9, 10, 11, 12\}$.

(c) the sum is not an odd number less than 9;
We could define a new event, F, the event that the sum is not an odd number less than 9. Then F contains all of the numbers in S except the ones which are odd numbers less than 9, i.e. all except 3, 5 and 7. So $F = \{2, 4, 6, 8, 9, 10, 11, 12\}$. That is, we can easily determine that F^c, the set containing all sample points which *are* odd numbers less than 9, is $F^c = \{3, 5, 7\}$ and therefore $F = \{2, 4, 6, 8, 9, 10, 11, 12\}$.

On the other hand, we could define this event in terms of events we have previously defined. If the outcome is not an odd number less than 9, then it must be an even number, or be at least 9 (or both). But then we could also describe this as the union of those other 2 events. So we see that $F = E \cup N$. That is, we don't need to give this event a new name; we can simply refer to it as the event $E \cup N$, and determine that

$$E \cup N = \{2, 4, 6, 8, 10, 12\} \cup \{9, 10, 11, 12\} = \{2, 4, 6, 8, 9, 10, 11, 12\}$$

(d) the sum is even and at least 9;
This is the event containing only those sample points which are in event E and are also in event N, so this is the event

$$E \cap N = \{2, 4, 6, 8, 10, 12\} \cap \{9, 10, 11, 12\} = \{10, 12\}$$

Let's look at one more example.

Example 2.3. *Joan and Fred each toss a coin. Define a Sample Space for this experiment and find the subsets corresponding to the following events:*
(a) both tosses had the same result;
(b) Joan tosses heads.

When we toss a coin, there are 2 possible outcomes, *heads* or *tails*, which we can denote by H and T, respectively. When 2 coins are tossed, we can represent the possible outcomes as pairs of letters, where each letter is an H or a T. For instance, here we can define our sample points to be pairs of letters, where the first letter denotes the result of Joan's toss and the second letter denotes the result of Fred's toss. In this way, we get $S = \{HH, HT, TH, TT\}$.

(a) both tosses had the same result
Let E be the event that both tosses had the same result. Then $E = \{HH, TT\}$.

(b) Joan tosses heads
Let F be the event that Joan tosses heads. Then F contains all the sample points in which the *first* letter is H, since that's how we defined the sample points. So $F = \{HH, HT\}$.

Notice: When the sample points are complex like this, we must be sure to remember how we defined them when we write the subsets corresponding to certain events. For instance, here we wanted the set $\{HH, HT\}$, **not** the set $\{HH, TH\}$, which is the event that *Fred* tosses heads.

Math 1228A/B Online

Lecture 12:
Properties of Probabilities

(text reference: Section 2.2)

2.2 Basic Properties of Probability

Our purpose in defining sample spaces, outcomes and events is to allow us to determine the **probability** that on a particular performance of an experiment, the **outcome** of the experiment will belong to some particular **event** in which we are interested.

We use **Pr**[**E**] to denote the probability that event E occurs.

There are a number of properties of probabilities that must *always* be true.

Property 1:
A probability is *always* a number between 0 and 1 (inclusive).

So we could have $Pr[E] = .5$ or $Pr[E] = \frac{2}{3}$ or $Pr[E] = 0$ or $Pr[E] = 1$, but we can **never** have $Pr[E] = 1.3$ or $Pr[E] = 5002$ or $Pr[E] = -\frac{1}{2}$. Those numbers **cannot** be probabilities.

We can also express probabilities as percentages (between 0% and 100% inclusive). For instance, if $Pr[E] = .5$, then we could also say that event E has a 50% chance of occurring, or that E occurs 50% of the time (on average).

Property 2:
An event that is *certain* to occur has probability 1.

That is, if we know that event E *must* happen, *every time* the experiment is performed, then $Pr[E] = 1$.

For instance, when we perform any experiment, we must define a sample space which is complete. So whenever we perform an experiment, the outcome we observe will always be in the set S. That is, we are certain that the event "the outcome is in S" will occur, so the probability of this event is 1. That is, one important consequence of Property 2 is:

Property 2(a): $Pr[S] = 1$ for any sample space S.

But Property 2 applies in other kinds of situations, as well. For instance, if we toss a single die, the result will be a number between 1 and 6, so if we define E to be the event that the number tossed is less than 10, then E is an event which is certain to occur, so $Pr[E] = 1$.

Property 3:
The probability of any *impossible* event is 0.

For instance, if E is the event that the outcome of the experiment is not in the sample space, then $Pr[E] = 0$ because this is impossible. (If it wasn't impossible, S would not be a sample space.) Also, if we have two events E and F which have no sample points in common, then they cannot both occur at the same time, so the probability of observing an outcome that is in both event E and event F is 0, because there are no such outcomes. Recalling that we use $\emptyset$ to denote the *empty set*, we have:

Property 3(a): For any events E and F defined on the same sample space,

$$\text{if } E \cap F = \emptyset, \text{ then } Pr[E \cap F] = 0$$

Define: If $E \cap F = \emptyset$, we say that E and F are **mutually exclusive** events.

Of course, an event and its complement have no sample points in common, so $E \cap E^c = \emptyset$. Therefore events E and E^c must always be mutually exclusive events (i.e. cannot both happen at the same time), so we have:

Property 3(b): For any event E, $Pr[E \cap E^c] = 0$.

Property 4:
For any events E and F defined on the same sample space,

$$Pr[E \cup F] = Pr[E] + Pr[F] - Pr[E \cap F]$$

Notice the similarity to $n(A \cup B) = n(A) + n(B) - n(A \cap B)$ (in Chapter 1). The reasoning is similar. If we add the probability of event E and the probability of event F, then if there are some outcomes in which both events occur, we have 'overcounted' the probability that E or F occurs, and must subtract off the probability that both occur together.

Now, consider any event defined on some sample space S. The event is a subset of S, and contains some sample points, i.e. some of the elements of S. We can think of this event as being the union of several smaller subsets, where each of these smaller subsets contains only one sample point.

For instance, suppose that $S = \{0, 1, 2, 3\}$ is a sample space for some experiment. Then we can define events which each contain only a single (different) sample point, such as $E_1 = \{0\}$, $E_2 = \{1\}$, $E_3 = \{2\}$ and $E_4 = \{3\}$. Now, any other event defined on S can be expressed as the union of some of these events, E_1, E_2, E_3 and E_4. For example, the event $F = \{1, 2\}$ can be expressed as $F = E_2 \cup E_3$.

Property 4 tells us that since $F = E_2 \cup E_3$, then we can calculate $Pr[F]$ as $Pr[F] = Pr[E_2 \cup E_3] = Pr[E_2] + Pr[E_3] - Pr[E_2 \cap E_3]$. But of course events E_2 and E_3 do not overlap, since each contains a different (single) sample point and sample points must, by definition, be mutually exclusive. Therefore $Pr[E_2 \cap E_3] = Pr[\emptyset] = 0$ (by property 3(a)), so we see that $Pr[F] = Pr[E_2] + Pr[E_3]$. And of course, $Pr[E_2]$ and $Pr[E_3]$ are just the probabilities of the sample points 1 and 2. Therefore we can find the probability of this event F by simply adding up the probabilities associated with the sample points which are contained in this event.

The same is true for any event defined on any sample space. For instance, if some event E is defined on some sample space S, and E contains, say, 4 different sample points, then because the sample points are themselves mutually exclusive events, we can find $Pr[E]$ by summing the probabilities associated with these 4 sample points.

In general, we have:

Property 4(a): To find the probability of any event E, add up the probabilities of all sample points which are contained in E.

Note: We can think of the probabilities of the particular sample points as weights. These weights must all be non-negative and together must sum to 1. So then the probability of event E is the total weight in that event, which is found by summing the weights of the sample points that are in E.

Of course, we can also apply Properties 3(a) and 4 together to *any* pair of mutually exclusive events, which gives:

Property 4(b): If E and F are mutually exclusive events, then

$$Pr[E \cup F] = Pr[E] + Pr[F]$$

Finally, since E and E^c are always mutually exclusive events, then by property 4(b) we have $Pr[E \cup E^c] = Pr[E] + Pr[E^c]$. But E^c contains all of the sample points in S which are not in event E, so the union of these 2 events contains *all* of the sample points, i.e. $E \cup E^c = S$. And as we have already seen (property 2(a)), $Pr[S] = 1$. Thus we see that for any event E, $Pr[E] + Pr[E^c] = 1$, or expressed another way, we have:

Property 5 For any event E, $Pr[E^c] = 1 - Pr[E]$.

Again, notice the similarity to $n(E^c) = n(U) - n(E)$ (since the sample space for a probabilistic experiment acts as the universal set).

Let's look at some examples of using these properties.

Example 2.4. *We have the Sample Space $S = \{a, b, c, d\}$ for some experiment. If it is known that $Pr[a] = .2$, $Pr[b] = .3$ and $Pr[c] = .15$, what is $Pr[d]$?*

Approach 1:
We have $S = \{a\} \cup \{b\} \cup \{c\} \cup \{d\}$. And since a, b, c and d are sample points, then they are mutually exclusive, so (as observed in Property 4(a)) we can find $Pr[S]$ by adding up the probabilities of these sample points. That is, we have

$$Pr[S] = Pr[a] + Pr[b] + Pr[c] + Pr[d]$$

But we know that $Pr[S] = 1$, so we have

$$\begin{aligned} 1 &= .2 + .3 + .15 + Pr[d] \\ \Rightarrow Pr[d] &= 1 - (.2 + .3 + .15) = 1 - .65 = .35 \end{aligned}$$

Approach 2:
Let $E = \{a, b, c\}$. Then (by property 4(a)) $Pr[E] = Pr[a] + Pr[b] + Pr[c] = .3 + .2 + .15 = .65$ and so (by property 5) $Pr[E^c] = 1 - Pr[E] = 1 - .65 = .35$. However, the set E^c is the set of all of the sample points in S which are not in E, and E contains all sample points except d, so $E^c = \{d\}$. Therefore $Pr[d] = Pr[E^c] = .35$.

Example 2.5. *A certain donut shop sells only coffee and donuts. It has been observed that when a customer enters the donut shop, the probability that the customer orders only coffee is .3, while the probability that the customer buys only donuts is .25 . If every customer makes a purchase, what is the probability that a customer orders both coffee and donuts?*

We are interested here in the purchasing habits of the customers of this donut shop. We can think of the probabilistic experiment "observe a donut shop customer and notice what he or she buys". The outcome depends on *which* customer we observe, so we can think of these customers as the sample points. That is, we let S be the set of all customers.

Now, what events are we interested in, or do we know something about? We're told the probability that a customer only buys coffee, so we'd better have an event for that. That is, we are interested in the subset of customers who only buy coffee. Likewise, another subset that we're interested in (and whose probability we know) is the subset of customers who only buy donuts. Therefore we let C be the event that a customer buys only coffee and D be the event that a customer buys only donuts. Then we have $Pr[C] = .3$ and $Pr[D] = .25$.

Notice: A customer cannot "buy only coffee" and also, on the same visit, "buy only donuts". Because of the *meanings* of these events, we can see that C and D are mutually exclusive events.

What other possibilities are there? We are told that all customers buy something, and that the shop sells only coffee and donuts. So any customer who doesn't buy either only coffee or only donuts must buy *both* coffee and donuts. That is, letting B be the event that a customer buys both coffee and donuts, we see that since every customer who is not in either of C or D must be in B, then $B = (C \cup D)^c$. And since C and D are mutually exclusive events, so that $C \cap D = \emptyset$ and $Pr[C \cap D] = 0$, then we have

$$Pr[C \cup D] = Pr[C] + Pr[D] - Pr[C \cap D] = Pr[C] + Pr[D] = .3 + .25 = .55$$

Thus, we see that

$$Pr[B] = Pr[(C \cup D)^c] = 1 - Pr[C \cup D] = 1 - .55 = .45$$

Therefore, the probability that a customer buys both coffee and donuts is .45

Example 2.6. *At this donut shop, 40% of all customers stay more than 5 minutes, but 90% of all customers stay less than half an hour. What is the probability that a customer is in the donut shop for more than 5 minutes but less than 30 minutes?*

We are still looking at the donut shop, but this time we're performing a different experiment, observing a different characteristic. In the previous example, our experiment was "observe a customer coming into the donut shop and notice what he or she buys". This time, our experiment still involves observing a customer who comes into the donut shop, but now we "notice how long he or she stays".

The sample space can still be considered to be the set of all customers. Let E be the event that a customer stays more than 5 minutes (i.e. the subset of S containing those customers who stay more than 5 minutes) and F be the event that a customer stays less than half an hour (i.e. less than 30 minutes). We are asked to find the probability that a customer stays more than 5 minutes and less than 30 minutes, i.e. is in both E and F, so we are looking for $Pr[E \cap F]$.

We are told that $Pr[E] = .4$ and $Pr[F] = .9$, so we see that $Pr[E^c] = 1 - .4 = .6$ and also that $Pr[F^c] = 1 - .9 = .1$.

Think about what these events E^c and F^c are. E^c is the event that a customer does not stay more than 5 minutes. So E^c is the event that a customer stays at most 5 minutes. Similarly, F^c is the event that a customer does not stay less than 30 minutes, i.e. the event that a customer stays at least 30 minutes. By thinking about the meanings of these events, we can easily see that they are mutually exclusive, because a customer clearly cannot "stay at most 5 minutes" and at the same time "stay at least 30 minutes". So we see that $E^c \cap F^c = \emptyset$ and so $Pr[E^c \cap F^c] = 0$.

But we know (see Properties of Set Complement, Lecture 2 pg. 6) that for any sets E and F, $E^c \cap F^c = (E \cup F)^c$, so we see that $Pr[(E \cup F)^c] = Pr[E^c \cap F^c] = 0$, and this tells us that $Pr[E \cup F] = 1 - Pr[(E \cup F)^c] = 1$. (That is, since $(E \cup F)^c$ is an impossible event, then its complement, the set $E \cup F$, is a certainty.)

Now we can see (by rearranging property 4) that

$$Pr[E \cap F] = Pr[E] + Pr[F] - Pr[E \cup F] = .4 + .9 - 1 = 1.3 - 1 = .3$$

That is, we see that 30% of all customers stay more than 5 minutes but less than half an hour.

Notice: We could also have found that $Pr[E \cup F] = 1$ by recognizing that this event is a certainty, i.e. by thinking about the meaning of E and of F to see that between them these events contain all customers, so that $E \cup F = S$. That is, we have $E^c \subseteq F$ because any customer who does not stay more than 5 minutes obviously stays less than half an hour. So then $E \cup F$ contains everything in E and also everything not in E, i.e. contains the whole sample space.

Similarly, we could get the same result by observing that $F^c \subseteq E$, because everyone who does not stay less than half an hour stays at least half an hour, and therefore clearly stays at least 5 minutes. So $E \cup F$ contains all of F and also all of F^c and hence comprises the whole sample space.

A Different Approach:
Define events E and F as above so that once again $Pr[E^c] = .6$ and $Pr[F^c] = .1$, and think again about what these mean. E^c is the event that a customer stays at most 5 minutes and F^c is the event that a customer stays at least 30 minutes. What customers are not in either of these 2 events? The ones who stay more than 5 minutes but less than 30 minutes, which are the ones we're interested in here. That is, we can express the probability we've been asked to find as $Pr[(E^c \cup F^c)^c]$. (Again, by the properties of set complement and using the fact that the complement of the complement of a set is just the set itself, we have $(E^c \cup F^c) = (E^c)^c \cap (F^c)^c = E \cap F$.)

As we saw before, because of their meanings, the events E^c and F^c are mutually exclusive events. Therefore we can calculate

$$Pr[E^c \cup F^c] = Pr[E^c] + Pr[F^c] = .6 + .1 = .7$$

But then the probability that a customer stays more than 5 minutes but less than 30 minutes, which we've seen is given by $Pr[(E^c \cup F^c)^c]$, can be found as

$$Pr[(E^c \cup F^c)^c] = 1 - Pr[E^c \cup F^c] = 1 - .7 = .3$$

Partitioning

Remember: Events are sets, i.e. subsets of the sample space S. When we say "let E be the event that such and such happens" we're really saying that E is the subset of S containing those sample points which have characteristic "such and such".

Also recall: We can partition any set A with respect to membership or non-membership in some other set B, using the fact that $A = (A \cap B) \cup (A \cap B^c)$ where $A \cap B$ and $A \cap B^c$ do not overlap, to see that $n(A) = n(A \cap B) + n(A \cap B^c)$. Similarly, we can partition the sample points in any *event* E according to membership or non-membership in some other event F. And since $E \cap F$ and $E \cap F^c$ are clearly mutually exclusive events, because no sample point can be both in F and in F^c, then we see (from Property 4(b)) that

Theorem: For any events E and F defined on the same sample space,

$$Pr[E] = Pr[E \cap F] + Pr[E \cap F^c]$$

More generally, if events F_1, F_2, ..., F_k are mutually exclusive events which together comprise the whole sample space, i.e. if $F_1 \cup F_2 \cup ... \cup F_k = S$ with $F_i \cap F_j = \emptyset$ for all pairs i, j with $i \neq j$, then for any other event E we have

$$Pr[E] = Pr[E \cap F_1] + Pr[E \cap F_2] + ... + Pr[E \cap F_k]$$

Notice: These are precisely analagous to the corresponding properties of "number in a set", in which we count the elements of a set by partitioning a set into 2 or more pieces (see Lecture 2, pg. 7).

We apply these partitioning ideas in the next example.

Example 2.7. *In the donut shop from Example 2.5, the probability that a customer buys at least 6 donuts and doesn't buy coffee is .15, while the probability that a customer buys 6 or more donuts and also buys coffee is .05 . What is the probability that a customer buys at least 6 donuts?*

We previously defined D to be the event that a customer only buys donuts, i.e. doesn't buy coffee. Now, we also define A to be the event that a customer buys at least 6 donuts. Recall that this particular donut shop only sells coffee and donuts.

If a customer "buys at least 6 donuts and doesn't buy coffee", then that customer buys at least 6 donuts and buys only donuts, so a customer who does this is in event $A \cap D$, and therefore we know that $Pr[A \cap D] = .15$. Similarly, the event that a customer buys 6 or more donuts and also buys coffee, i.e. buys at least 6 donuts and doesn't 'only buy donuts' is $A \cap D^c$. So we have $Pr[A \cap D^c] = .05$.

By partitioning the customers who "buy at least 6 donuts" into those who only buy donuts and those who don't, we see that

$$Pr[A] = Pr[A \cap D] + Pr[A \cap D^c] = .15 + .05 = .2$$

so the probability that a customer buys at least 6 donuts is .2

Alternatively, we could partition A into more pieces, using D, B and C. That is, we know that every customer is in exactly one of D (only donuts), B (both coffee and donuts) or C (only coffee), so these 3 sets form a partition of the sample space. Therefore, we have

$$Pr[A] = Pr[A \cap D] + Pr[A \cap B] + Pr[A \cap C]$$

Of course, it is impossible for a customer to buy at least 6 donuts and at the same time buy only coffee, so $Pr[A \cap C] = 0$. Clearly, every customer who buys at least 6 donuts and also buys coffee buys both coffee and donuts, so these customers are all in B, i.e. are the customers in the set $A \cap B$. Thus we have $Pr[A \cap B] = .05$ and we get

$$Pr[A] = Pr[A \cap D] + Pr[A \cap B] + Pr[A \cap C] = .15 + .05 + 0 = .2$$

A tip about defining events:
When approaching a problem, define the sample space very broadly (as in "the set of all customers") and then define subsets (events) broadly as well, defining any particular event in terms of only *one* characteristic. For instance, in Example 2.7 we were told the probability that a customer "buys at least 6 donuts and doesn't buy coffee". There are 2 things going on there – 2 characteristics – one concerning *how many* donuts are bought and the other about *whether or not* coffee is bought. If we simply say "let E be the event that a customer buys at least 6 donuts and doesn't buy coffee", that's not very helpful. Instead, we use a different event for each of these characteristics and express the event "buys 6 donuts and doesn't buy coffee" as the intersection of these events. That is, as shown above, we define a new event for the new characteristic "buys at least 6 donuts" and use the previously defined event for the characteristic "doesn't buy coffee" (i.e. only buys donuts). In this way, we can sort out the interactions of these characteristics, which is much more difficult if we have confounded the 2 characteristics into one event.

Notice: This same tip applies to defining subsets in a counting problem.

Math 1228A/B Online

Lecture 13:

Using Sample Spaces in which

All Sample Points are Equally Likely

(text reference: Section 2.3)

2.3 Equiprobable Sample Spaces

Example 2.8. *A fair coin is tossed. What is the probability that it comes up Heads?*

I'm sure that even if you've never studied any probability before, you probably know the answer to this, without too much thought. It's $\frac{1}{2}$. But *why* is the probability of tossing Heads $\frac{1}{2}$?

When we toss a coin and observe what side is facing up, there are 2 possibilities, Heads or Tails, so the sample space for this experiment is $S = \{H, T\}$. When we say "a fair coin", we mean a coin which is *equally likely* to come up Heads as Tails. That is, we mean that

$$Pr[H] = Pr[T]$$

But also, since H and T are the only sample points, we know that

$$Pr[S] = Pr[H] + Pr[T]$$

So if we let $x = Pr[H]$, then we have $Pr[T] = x$ as well, so that

$$Pr[S] = Pr[H] + Pr[T] = x + x = 2x$$

And of course, we also know that $Pr[S] = 1$, so we get

$$2x = 1 \Rightarrow x = \frac{1}{2}$$

That is, we have $Pr[H] = \frac{1}{2}$, so that $Pr[T] = \frac{1}{2}$ too.

This is an example of a special kind of sample space. We define:

A sample space is called **equiprobable** if every sample point is equally likely to occur.

For instance, since Heads and Tails are equally likely with a fair coin, then $S = \{H, T\}$ is an equiprobable sample space for the experiment "toss a fair coin".

Suppose S is any equiprobable sample space. Then we have $n(S)$ sample points, each equally likely to occur. But since it is always true that $Pr[S]$ is given by the sum of the probabilities of the various sample points in S, and also that $Pr[S] = 1$, then we see that each of the $n(S)$ different sample points must occur with probability $\frac{1}{n(S)}$. That is, if each of the $n(S)$ equally likely sample points occurs with probability p, then

$$\underbrace{p + p + \cdots + p}_{n(S) \text{ of these}} = n(S) \times p = 1 \qquad \text{so} \qquad p = \frac{1}{n(S)}$$

Therefore we have:

Theorem: If S is an equiprobable sample space, then for any sample point a,

$$Pr[a] = \frac{1}{n(S)}$$

Notice: For $S = \{H, T\}$, we have $n(S) = 2$ and we found that

$$Pr[H] = Pr[T] = \frac{1}{2} = \frac{1}{n(S)}$$

We can use this knowledge to easily find $Pr[E]$ for any event E which is defined on an equiprobable sample space. We know that we can find $Pr[E]$ by adding up the probabilities of all the sample points which are contained in the event E. When S is an equiprobable sample space, these sample points all have the same probability, $\frac{1}{n(S)}$. Of course E contains $n(E)$ sample points, and since each has probability $\frac{1}{n(S)}$, then when we add up $n(E)$ of these, we get

$$Pr[E] = \underbrace{\frac{1}{n(S)} + \cdots + \frac{1}{n(S)}}_{n(E) \text{ of these}} = n(E) \times \frac{1}{n(S)} = \frac{n(E)}{n(S)}$$

Therefore we see that

Theorem: For any event E defined on an equiprobable sample space S, $Pr[E] = \frac{n(E)}{n(S)}$.

Example 2.9. *A single die is tossed. What is the probability that an even number comes up?*

When we toss a single die, the outcome, i.e. the number that comes up, will be 1, 2, 3, 4, 5 or 6. So we can use $S = \{1, 2, 3, 4, 5, 6\}$ as a sample space for this experiment. Of course, assuming that what we are tossing is a fair die, then each of these outcomes (i.e. numbers) is equally likely to occur, so this is an *equiprobable sample space*. It contains $n(S) = 6$ sample points.

Let E be the event that an even number comes up. Then we have $E = \{2, 4, 6\}$, with $n(E) = 3$. Thus we have

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

Notice: It might be tempting, when looking at Example 2.9, to say "well, every toss either comes up even or comes up odd, so $S = \{\text{even}, \text{odd}\}$ is a sample space for this experiment, with $n(S) = 2$, so $Pr[\text{even}] = \frac{1}{n(S)} = \frac{1}{2}$.". This does work in *this* case, but *only because* 3 of the 6 possible outcomes are even, and 3 are odd. We must be careful about doing something like this. Suppose we tossed a **seven sided** die, with faces numbered 1 through 7. Then it would still be true that each toss will be either even or odd, so the set $\{\text{even}, \text{odd}\}$ is a sample space for the experiment. However it is **not** an equiprobable sample space. $S = \{1, 2, 3, 4, 5, 6, 7\}$ is, assuming we have a fair die, an equiprobable sample space for the experiment, with $n(S) = 7$. Letting E be the event that an even number comes up, we have $E = \{2, 4, 6\}$, so $n(E) = 3$ and $Pr[E] = \frac{3}{7} \neq \frac{1}{2}$.

When defining a sample space for a problem in which we need to calculate probabilities by counting sample points, we must always be *very careful* to define a sample space in which each sample point *is* equally likely to occur. We usually accomplish this by expressing the possible outcomes of the experiment in the most basic possible way (such as "what number comes up on the die?" or "which card is drawn?", etc.) and considering each of these possible outcomes to be a different sample point, i.e. defining the sample space as the set containing each of these individual possible outcomes.

Let's look at some more examples.

Example 2.10. *A single letter is chosen at random from the word '*bassoonist*'. What is the probability that the chosen letter is an 'o'?*

This 10-letter word only contains the letters b, a, s, o, n, i and t, so the outcome of this experiment will be one of these letters. Therefore the set {b, a, s, o, n, i, t} would be a possible sample space for this experiment. This sample space contains 7 sample points. Does that mean that the probability of getting an 'o' is 1/7? No. This is *not* an equiprobable sample space. Since we have more s's than o's and more o's than any of the other letters, then when we pick a letter at random, we are more likely to get an s than an o, and more likely to get an o than a b, a, n, i or t.

Let's think about how we could define a different sample space for this experiment. We could restate the experiment as

> "Randomly choose 1 of the positions in the 10-letter word 'bassoonist' and observe what letter is in that position."

Looked at this way, it is easy to see how to find an equiprobable sample space for this experiment. Consider the sample points to be the *positions* of letters in the 10-letter word. That is, define the sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. When we choose one of the 10 positions at random, each position is equally likely to be chosen – that's what 'choose one at random' means. So *this* set is an *equiprobable* sample space for the experiment.

We have o's in 2 of the 10 positions (i.e. positions 5 and 6), so if we define E to be the event that the chosen position contains an 'o', we have $E = \{5, 6\}$. We see that

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{2}{10} = \frac{1}{5}$$

so the probability that the chosen letter is an 'o' is $\frac{1}{5}$.

Example 2.11. *Two dice are tossed. What is the probability that the sum is 7?*

We have seen that the set $\{2, 3, 4, ..., 12\}$ is a possible sample space for the experiment "toss 2 dice and observe the sum". However this is not an equiprobable sample space. How do we know that it isn't? Well, that doesn't really matter. What matters is that we don't know that it *is*. We can never assume that a sample space is equiprobable. We have to *know* that it is.

In fact, if we think about this a little bit, we can see that this set wouldn't be an equiprobable sample space because the possible sums are not equally likely to occur. For instance, there is only one way to get a sum of 2 – roll two ones. But there are lots of ways to get a sum of 6 – a 1 and a 5, a 2 and a 4, two 3's, etc. So some of the possible sums are more likely to be observed than others.

To find the probability of getting a sum of 7, we need to define an equiprobable sample space. Remember that for an equiprobable sample space, we need the sample points to be individual outcomes of the experiment, in the most basic possible form. Then performing the experiment, which corresponds to choosing one of the outcomes at random, translates to simply choosing one of the sample points at random, ensuring that the sample points each occur with the same probability.

In section 2.1, we saw that the set $S = A \times A$, where $A = \{1, 2, ..., 6\}$, is another possible sample space for this experiment, because the ordered pairs we get represent the various possible outcomes of tossing two dice, in the form (a, b) where a is the number on one die and b is the number on the other die (where the 2 dice are considered to be distinct).

Here, S is the set explicitly containing each of the possible outcomes of tossing 2 dice, with each as a separate sample point. When we toss 2 (fair) dice, each die is equally likely to come up any of 1 through 6, so each of these outcomes is equally likely to occur. Therefore S is an equiprobable sample space for any experiment involving tossing 2 dice. Of course, $n(S) = n(A \times A) = n(A) \times n(A) = 6 \times 6 = 36$.

Using this set S as the sample space, let E be the event that the sum of the 2 dice is 7. Then E contains all of the outcomes, i.e. all of the ordered pairs, in which the 2 numbers add up to 7, so we have

$$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

and we see that $n(E) = 6$. Therefore we get

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

The counting methods from Chapter 1 come up a lot in equiprobable sample space problems. For any question of the form "what is the probability that a randomly chosen one of the ways of doing *something-or-other* has *such-and-such* a characteristic?", we define an equiprobable sample space containing each of the distinct ways of "doing *something-or-other*" as a separate sample point, and consider the event that when we do the something-or-other we observe "characteristic *such-and-such*". That is, we look at the number of ways we can do whatever it is, and how many ways we can do it so that it has the desired characteristic. Therefore for any equiprobable sample space problem of this type we have 2 counting problems – counting the number of ways of doing whatever is being done, and counting the number of ways it can be done so that it has the characteristic whose probability we need to find.

Let's look at some examples.

Example 2.12. *If 3 men and 3 women line up at random to have their picture taken, what is the probability that men and women alternate in the line?*

Let S be the set of all ways that the 3 men and 3 women can stand in a line. Since the people line up at random, each of these ways of lining up is equally likely, so S is an equiprobable sample space. There are 6 people lining up, so there are 6! different permutations of the people, i.e. ways they can line up, so we have $n(S) = 6! = 720$.

Let E be the event that men and women alternate in the line. Then the number of distinct ways in which event E can occur is the number of distinct ways in which 3 men and 3 women can alternate in line. As we saw in example 1.13 (see Lecture 5, pg. 23), there are $2 \times 3! \times 3! = 72$ distinct ways that this can be done. (Remember: there are 2 choices for which sex is first, and then 3! ways to arrange the men and 3! ways to arrange the women.) We have $n(E) = 72$ and we get

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{72}{720} = \frac{1}{10}$$

Example 2.13. *Arthur, Guinevere and 6 knights sit down at The Round Table. What is the probability that Arthur and Guinevere are sitting together?*

Notice: We must assume that the 8 people sit *randomly* around the table. If Arthur and Guinevere make a point of sitting together, then there's nothing to calculate, because them sitting together is a certainty. But we're not told that they are going to do that, so we assume that the seats are chosen at random.

Let S be the set of all ways that the 8 people could sit around The Round Table. Since the 8 people sit randomly at the table, each seating arrangement is equally likely and S is an equiprobable sample space. Then $n(S)$ is just the number of distinct ways in which 8 objects can be arranged in a circle, which is $n(S) = (8-1)! = 7!$.

Let E be the event that Arthur and Guinevere are sitting together. As we saw in Example 1.18 (see Lecture 6, pg. 28), there are $n(E) = 6! \times 2$ different ways that Arthur, Guinevere and 6 knights could be arranged at the table, with Arthur and Guinevere sitting together. (i.e. there are 6! ways to arrange Arthur and the 6 knights in a circle, and then 2 choices for where Guinevere sits.) Therefore we have

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{6! \times 2}{7!} = \frac{\cancel{6!} \times 2}{7 \times \cancel{6!}} = \frac{2}{7}$$

So provided that they all sit down at the table at random, the probability that Arthur and Guinevere are sitting together is $\frac{2}{7}$.

Notice: Another way to get this is to think of Arthur sitting anywhere at the table, which has 8 places at it. Now, Guinevere comes in and picks a seat at random. Since she can't pick the seat that Arthur's already in, there are 7 choices of seat available to her, of which 2 are beside Arthur. That is, we let S be the set of all ways Guinevere could choose a seat. Since Guinevere chooses a seat at the table at random, she is equally likely to be in any position relative to Arthur, so thinking of Arthur's place as already fixed, Guinevere is equally likely to choose any of the other 7 seats. Now let E be the event that she chooses a seat next to Arthur. Then we have $n(S) = 7$ and $n(E) = 2$, so that $Pr[E] = \frac{2}{7}$ (assuming she chooses a seat at random). How the 6 knights arrange themselves in the remaining seats is not important.

Example 2.14. *A 4-person subcommittee is chosen randomly from a 15-member committee. What is the probability that Jane Smith, one of the committee members, is on the subcommittee?*

Let S be the set of all ways to choose a 4-member subcommittee from a 15-person committee. Then we have $n(S) = \binom{15}{4}$. Since the subcommittee members are chosen randomly, each of the different possible subcommittees is equally likely to occur, so this is an equiprobable sample space.

Let E be the event that Jane Smith is on the sub-committee. Then E contains all of the possible sub-committees which have Jane Smith as one of the members. Each of these has Jane as one member, and some subset of 3 other people as the other members. There are $\binom{14}{3}$ ways to choose 3 of the other 14 people to be on the sub-committee with Jane, so we have $n(E) = \binom{14}{3}$. (We could also consider this as $\binom{1}{1}$ ways to choose Jane and $\binom{14}{3}$ ways to choose 3 others so that $n(E) = \binom{1}{1}\binom{14}{3} = \binom{14}{3}$.) Thus we get

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{\binom{14}{3}}{\binom{15}{4}} = \frac{\frac{14!}{3!11!}}{\frac{15!}{4!11!}} = \frac{14!}{3!\cancel{11!}} \times \frac{4!\cancel{11!}}{15!} = \frac{\cancel{14!} \times 4 \times \cancel{3!}}{\cancel{3!} \times 15 \times \cancel{14!}} = \frac{4}{15}$$

Notice: Again, we can do this another way. One method for selecting a subcommittee would be to have the 15 committee members arrange themselves randomly in a line and then specify that the people in 4 particular positions in the line, e.g. the first 4 in line, will be on the subcommittee. When the 15 people line up at random, there are 15 positions that Jane could be in. Let S be the set containing these 15 positions (each of which is equally likely to be the position Joan is in) and let E be the event that Jane is in one of the 4 positions designated as subcommittee members. Then $n(S) = 15$ and $n(E) = 4$, so the probability that Jane is on the subcommittee is $Pr[E] = \frac{4}{15}$.

Example 2.15. *Seven paintings are hung at random in a row along a wall in an art gallery. There are 3 paintings by Picasso and 4 by Monet. What is the probability that all of the Picassos are hanging side by side and all of the Monets are also hanging side by side?*

Let S be the set containing all the different ways the paintings could be hung on the wall. Since the paintings are hung in a row, each possible way of hanging the paintings is simply a permutation of the seven paintings, so $n(S) = 7!$. And since the paintings are hung at random, each permutation is equally likely, so S is an equiprobable sample space.

Let E be the event that the 3 Picasso paintings are hung side by side on the wall and also the 4 Monet paintings are hung side by side. We need to count the number of ways of hanging the paintings like this. The Picassos must be grouped together, but may be in any of the 3! permutations of the 3 paintings. Likewise, the 4 Monets, hanging all together, may be in any of 4! different arrangements. And of course, we don't know the order of the painters. That is, the Monets could be either to the left or to the right of the Picassos, so we must also take into account the $2! = 2$ possible orderings of the painters.

Taking all 3 decisions into account, i.e. order of Picassos, order of Monets and order of painters, we see that $n(E) = 3!4!2!$ and therefore

$$Pr[E] = \frac{3!4!2!}{7!} = \frac{\cancel{3} \times \cancel{2} \times \cancel{4!} \times 2}{7 \times \cancel{6} \times 5 \times \cancel{4!}} = \frac{2}{35}$$

Example 2.16. *Two couples (Pat and Lee; Kim and Sandy) choose seats at random at a card table, to play bridge. What is the probability that each player is partnered with his or her significant other? (Note: square table; one player sits on each side; players sitting opposite one another are partnered for the game.)*

Approach 1: consider the arrangement of the players around the table
Let S be the set containing all the different ways that the 2 couples, i.e. 4 people, could sit around the table. Having 4 people sit around a square table (one per side) is effectively the same as having 4 people sit around a round table. That is, the 4 positions are not distinguished in any way, there is no first or last, so each seating arrangement is a circular permutation of the 4 people. Therefore $n(S) = (4-1)! = 3! = 6$. Since the 4 sit randomly around the table, S is an equiprobable sample space.

Let E be the set of all ways that the 4 could sit at the table with each partnered with, i.e. sitting across from, his or her 'significant other'. Then Pat must be sitting across from Lee, and Kim must be sitting across from Sandy. It must be true that Kim is sitting beside Lee, but Kim could be sitting either to Lee's right or to Lee's left. We see that there are 2 arrangements of the people contained in event E, i.e. that $n(E) = 2$. (That is, if Pat and Lee take places at the table facing one another, there are $2! = 2$ possible arrangements of the other 2 people, Kim and Sandy, in the other 2 places.)

$$\text{Therefore } Pr[E] = \frac{n(E)}{n(S)} = \frac{2}{6} = \frac{1}{3}$$

Approach 2: consider only who is partnered with whom
Let S be the set of all partners Pat could have. Since there are 3 other people, any of whom could be partnered with (i.e. sit across from) Pat, then $n(S) = 3$. And since the 4 people choose seats at random, then each of the other 3 is equally likely to be partnered with Pat, so S is an equiprobable sample space.

Let E be the event that Pat is partnered with Lee. Whenever this occurs, it must also be true that Kim is partnered with Sandy and thus the event that each is partnered with his or her significant other is precisely this event E. Of course, defined as a subset of S as defined here, E is the set containing only Lee and hence $n(E) = 1$. Thus we see that $Pr[E] = \frac{1}{3}$.

Example 2.17. *An urn contains 5 red balls and 8 black balls. Two of the balls are selected at random.*

(a) *What is the probability that one ball of each colour is selected?*

(b) *What is the probability that both of the selected balls are red?*

(c) *What is the probability that at least one of the selected balls is red?*

(d) *What is the probability that at least one of the selected balls is black?*

(a) There are some red balls and some black balls in the urn, and we need to find the probability that one ball of each colour is selected when 2 balls are drawn at random.

Let S be the set of all ways of choosing 2 balls from the urn. Since the urn contains 5 red balls and 8 black balls, there are 13 balls in the urn. S contains all of the different ways of choosing 2 of these 13 balls, so $n(S) = \binom{13}{2}$. Since the 2 balls are selected at random, this is an equiprobable sample space.

Notice: Although it would seem that we do not necessarily need to distinguish between balls of the same colour, we do need to take into account how many there are. All we're interested in here is the colours of the 2 drawn balls, so the set {both red, both black, one of each} is a possible sample space for the experiment. However, it is *not* an *equiprobable* sample space because the sample points are not equally likely to occur. Because there are more black balls than red balls, it is more likely that 2 black balls will be drawn than that 2 red balls will be drawn, i.e. we have

$$Pr[\text{both black}] > Pr[\text{both red}]$$

In order to take this into account, we do need to *treat* the red balls as being distinct from one another, and the same for the black balls. Thus we consider that there are 13 *distinct* balls in the urn, giving $\binom{13}{2}$ distinct ways to select 2 of them. (This is similar to the situation where we must treat the 2 dice as distinct even when they may be identical, in order to have an equiprobable sample space for an experiment in which 2 dice are tossed.)

Now, let E be the event that one ball of each colour is selected. There are $\binom{5}{1}$ ways to choose one of the 5 red balls and $\binom{8}{1}$ ways to choose one of the 8 black balls, so there are $\binom{5}{1} \times \binom{8}{1}$ ways to choose one red ball and choose one black ball, i.e. to choose one ball of each colour. Therefore we see that $n(E) = \binom{5}{1}\binom{8}{1}$ and we get

$$Pr[E] = \frac{n(E)}{n(S)} = \frac{\binom{5}{1}\binom{8}{1}}{\binom{13}{2}} = \frac{5 \times 8}{\frac{13!}{2!11!}} = 40 \times \frac{2!11!}{13!} = \frac{40 \times \cancel{2} \times \cancel{11!}}{13 \times \cancel{12}_6 \times \cancel{11!}} = \frac{\cancel{40}^{20}}{13 \times \cancel{6}_3} = \frac{20}{39}$$

(b) We need to find the probability that both of the chosen balls are red. We can, of course, use the same sample space as in (a), with $n(S) = \binom{13}{2}$.

Let R be the event that both of the chosen balls are red. There are $\binom{5}{2}$ ways to choose 2 of the 5 red balls (and none of the black balls), so we have $n(R) = \binom{5}{2}$. We see that

$$Pr[R] = \frac{n(R)}{n(S)} = \frac{\binom{5}{2}}{\binom{13}{2}} = \frac{\frac{5!}{2!3!}}{\frac{13!}{2!11!}} = \frac{5!}{\cancel{2}!3!} \times \frac{\cancel{2}!11!}{13!} = \frac{5 \times \cancel{4} \times \cancel{3!} \times \cancel{11!}}{\cancel{3!} \times 13 \times \cancel{12}_3 \times \cancel{11!}} = \frac{5}{39}$$

(c) We are asked to find the probability that at least one of the 2 selected balls is red. In order to have at least one red ball, we may have either 1 or 2 red balls chosen. We know (from (a) and (b)) that the probability of choosing exactly one red ball is $\frac{20}{39}$ and that the probability of choosing 2 red balls is $\frac{5}{39}$. Therefore the probability of choosing at least one red ball, when 2 balls are drawn at random from this urn, is $\frac{20}{39} + \frac{5}{39} = \frac{25}{39}$.

(That is, if F is the event that at least one red ball is drawn, then F contains all the outcomes in which exactly one red ball is drawn, as well as all the outcomes in which both of the drawn balls are red. Thus $F = E \cup R$, and since E and R are mutually exclusive (cannot happen together), then $n(E \cap R) = 0$, so $n(F) = n(E) + n(R) + n(E \cap R) = n(E) + n(R)$. This approach gives $n(F) = \binom{5}{1}\binom{8}{1} + \binom{5}{2} = 40 + 10 = 50$ and since $n(S) = \binom{13}{2} = 78$, we get $Pr[F] = \frac{50}{78} = \frac{25}{39}$.)

(d) This time, we wish to find the probability that at least one of the selected balls is black. There are various approaches we could take to this. For instance, we could calculate the probability that at least one black is drawn as the sum of the probability that exactly one black is drawn plus the probability that exactly 2 blacks are drawn. Or we could count the number of outcomes in the event that at least one black ball is drawn, and use that to directly calculate the probability that this event occurs. (That is, we could do this similarly to the way (c) was done.). Or we could determine the number of outcomes in this event by counting the number of outcomes in the complement of this event. However, the single quickest and easiest way to find this probability is to use the probability of, rather than the number of elements in, the complementary event, which is a probability we have already calculated.

If there is *not* at least one black selected then there must be no black ball drawn. And if neither of the 2 balls is black, then they must both be red. Therefore the event that at least one ball is black is the complement of the event that both balls are red. So we are looking for $Pr[R^c]$, where R is as defined in (b), and we use the value for $Pr[R]$ calculated there to see that

$$Pr[R^c] = 1 - Pr[R] = 1 - \frac{5}{39} = \frac{39}{39} - \frac{5}{39} = \frac{34}{39}$$

Math 1228A/B Online

Lecture 14:

Calculating Probabilities when we Already Know Something About the Outcome

(text reference: Section 2.4, pages 78 - 81)

2.4 Conditional Probability

When an experiment is performed, knowing whether one event occurs can change our assessment of the probability that another event occurs. That is, knowing *something* about what outcome has been observed can change the probability of an event. To see this, consider the following pair of problems:

Example 2.18. *(a) A single card is drawn at random from a standard deck. What is the probability that the card is a heart?*

There are 52 cards in the deck, of which 13 are Hearts. So if we let S be the set of all ways in which one card can be drawn from the deck, and H be the set of all ways of drawing a heart, then we have $n(S) = \binom{52}{1} = 52$ and $n(H) = \binom{13}{1} = 13$, i.e. 13 of the 52 cards are Hearts. Since one card is drawn at random, each card is equally likely to be drawn, so S is an equiprobable sample space and thus $Pr[H] = \frac{13}{52} = \frac{1}{4}$.

Example 2.18. *(b) A single card has been drawn from a standard deck and it is known that the card is red. What is the probability that the card is a heart?*

This time, we know that the card is one of the 26 red cards. Knowing that a red card has been drawn, we can think of the experiment as 'choose one of the red cards at random', and we can use the set of red cards as the equiprobable sample space. Of course, since all of the Hearts are red, then 13 of the 26 red cards are Hearts, so now the probability that the red card which was drawn is a Heart is $\frac{13}{26} = \frac{1}{2}$.

We see that knowing that the card is red *changes* the probability that the card is a Heart.

We define:

> The **conditional probability of event** E**, given event** F, written $\mathbf{Pr[E|F]}$, is the probability that an outcome is in event E, when it is known that the outcome is in event F.

In the previous example, if we let R be the event that a red card is drawn when we draw a single card from a standard deck, then the probability we found in part (b) was $Pr[H|R]$, whereas in (a) what we found was $Pr[H]$.

There's a simple formula which we can use to find the conditional probability of E given F:

> **Theorem:** For any events E and F defined on the same sample space, with $Pr[F] \neq 0$, the conditional probability of E given F can be found as

$$Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]}$$

That is, the probability that E occurs, given that F has occured, is the probability that the events both occur, divided by the probability that event F occurs.

Notice: For the situation in example 2.18 (b), we have $Pr[H \cap R] = Pr[H] = \frac{1}{4}$, because all Hearts are red, so if a Heart is drawn, it must also be true that event R has occurred. In this case, the formula gives $Pr[H|R] = \frac{1/4}{1/2} = \frac{1}{4} \times \frac{2}{1} = \frac{1}{2}$, which is what we found before.

Let's look at some more examples.

Example 2.19. *Recall Examples 2.5 and 2.7, about the purchasing habits of the customers at the donut shop.*
(a) Find the probability that a customer who comes into the donut shop to buy only donuts buys at least 6 donuts.
(b) Find the probability that a customer who comes in to buy at least 6 donuts does not buy any coffee.

In Example 2.5 (see Lecture 12, pg. 61), we defined D to be the event that a customer buys only donuts, with $Pr[D] = .25$. In Example 2.7 (see Lecture 12, pg. 64), we defined A to be the event that a customer buys at least 6 donuts, and we had $Pr[A \cap D] = .15$ and $Pr[A \cap D^c] = .05$, so that $Pr[A] = .2$.

(a) We want to find the probability that a customer who only buys donuts buys at least 6 of them, i.e. the probability that a customer buys at least 6 donuts (i.e. is in event A), given the information that this is a customer who only buys donuts (i.e. is in event D). Therefore we are looking for $Pr[A|D]$. Using the formula, we get:

$$Pr[A|D] = \frac{Pr[A \cap D]}{Pr[D]} = \frac{.15}{.25} = \frac{15}{25} = \frac{3}{5} = .6$$

That is, 60% of the people who only buy donuts buy at least 6 of them.

(b) Now, we want to find the probability that a customer who buys at least 6 donuts doesn't buy coffee. That is, we need to find the probability that a customer buys only donuts (event D), when we know that the customer buys at least 6 donuts (event A). So this time, we are looking for $Pr[D|A]$. We get:

$$Pr[D|A] = \frac{Pr[D \cap A]}{Pr[A]} = \frac{.15}{.2} = \frac{15}{20} = \frac{3}{4} = .75$$

That is, 75% of the customers who buy at least 6 donuts do not buy any coffee.

Note: Although the numerators are the same in these 2 calculations, the denominators, and thus also the final answers, are different. It's very important to accurately identify which event is 'given', i.e. which is known to happen. **That** is the event whose probability is in the denominator.

We can rearrange the formula for conditional probability to allow us to find the probability that E and F both occur, when we know $Pr[E|F]$. We have:

$$Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]} \qquad \Rightarrow \qquad Pr[E \cap F] = Pr[E|F] \times Pr[F]$$

Example 2.20. *A certain polling company reported that in a poll of 200 people, 75% said that they intend to vote in an upcoming plebiscite. Among those who intend to vote, 60% said that they are in favour of the proposal. How many people said both that they intend to vote and that they are in favour of the proposal?*

Consider the experiment "Choose a person at random from those polled". Let V be the event that the person said that they intend to vote, and let F be the event that the person said that they are in favour of the proposal. We know that 75% of those polled said that they intend to vote, so $Pr[V] = .75$. Also, we are told that "among those who intend to vote, 60% said they are in favour". This means that if we restrict our attention to those who intend to vote, the probability that a person is in favour of the proposal is .6 . That is, .6 is the probability that someone is in favour of the proposal, *given that* they are among those who said that they intend to vote, so we have $Pr[F|V] = .6$

We need to find the probability that someone said both that they intend to vote and that they are in favour of the proposal, i.e. $Pr[F \cap V]$. We get:

$$Pr[F \cap V] = Pr[F|V] \times Pr[V] = (.6)(.75) = .45$$

Therefore 45% of the 200 people, i.e. $(.45)(200) = 90$ people, said that they intend to vote and also said that they are in favour of the proposal.

Example 2.21. *A hand of three cards has been dealt from a standard deck. It is known that exactly 1 of the cards is black. What is the probability that the hand contains exactly 1 heart?*

We're going to need to calculate probabilities here, so the first thing we need to do is define an equiprobable sample space. Let S be the set of all 3-card hands which could be dealt from a standard deck. Then $n(S) = \binom{52}{3}$. Assuming the deal is random, this set S is an equiprobable sample space for any experiment in which a 3-card hand is dealt.

We know that exactly one card in the hand is black and want to know about the probability that the hand contains exactly one Heart. We need to define some events. Let B be the event that exactly one of the 3 cards in the hand is black, and let H be the event that the hand contains exactly one Heart.

Notice: We must be careful in defining events here. First of all, we need to be precise. We don't just say 'Let B be the event that a card is black' because that's not a 3-card hand – it's not a subset of S. Any event we define must describe a 3-card hand. An event *must* be a subset of the sample space. Also, we're not interested in the event that the hand 'contains some black cards', but specifically that it contains exactly 1 black card. Likewise, we are interested specifically in the event that a 3-card hand contains exactly 1 red card. Furthermore, althought what we're really interested in here is the probability that 'a hand which is known to contain exactly 1 black card also contains exactly 1 heart', we **don't** use this as the definition of an event. Instead, we must recognize that this is a conditional probability, and define 2 events which allow us to express this conditional probability.

Since we know that B has occurred, and are asked to find the probability that H has occurred, we are looking for $Pr[H|B]$. Of course, we know that

$$Pr[H|B] = \frac{Pr[H \cap B]}{Pr[B]}$$

so we need to find $Pr[B]$ and $Pr[H \cap B]$.

If the 3-card hand contains *exactly* one black card, then it must also contain 2 red cards. That is, B is the set of all possible 3-card hands which contain 1 of the 26 black cards and 2 of the 26 red cards from the deck. Thus we have

$$n(B) = \binom{26}{1} \times \binom{26}{2} \Rightarrow Pr[B] = \frac{\binom{26}{1} \times \binom{26}{2}}{\binom{52}{3}}$$

In order for a hand to be in event B and also in event H, the hand must contain exactly 1 black card and exactly 1 Heart. Since there are 3 cards in the hand, then the third card in the hand must be a card which is neither black nor Hearts, which means it must be a Diamond (since these are the only non-black, non-Heart cards in the deck). That is, a hand in the event $B \cap H$ must contain 1 of the 26 black cards, 1 of the 13 Hearts and 1 of the 13 Diamonds, so we have

$$n(B \cap H) = \binom{26}{1} \times \binom{13}{1} \times \binom{13}{1} \Rightarrow Pr[B \cap H] = \frac{\binom{26}{1} \times \binom{13}{1} \times \binom{13}{1}}{\binom{52}{3}}$$

Notice: When we count $n(B)$ and $n(B \cap H)$, we **must** remember that we are counting 3-card hands and account for all 3 cards. That is, not only do we need to be careful to define events as subsets of the sample space, perhaps even more importantly we need to be sure that we are *counting* elements of a subset of the sample space, i.e. that the objects we count are the same kind of objects that we counted when we counted the objects in the sample space – in this case, hands of 3 cards.

Now we're ready to find Pr$[H|B]$. We get:

$$Pr[H|B] = \frac{Pr[H \cap B]}{Pr[B]} = \frac{\frac{\binom{26}{1}\times\binom{13}{1}\times\binom{13}{1}}{\binom{52}{3}}}{\frac{\binom{26}{1}\times\binom{26}{2}}{\binom{52}{3}}} = \frac{\cancel{\binom{26}{1}} \times \binom{13}{1} \times \binom{13}{1}}{\cancel{\binom{52}{3}}} \times \frac{\cancel{\binom{52}{3}}}{\cancel{\binom{26}{1}} \times \binom{26}{2}} = \frac{13 \times 13}{\frac{\cancel{26}^{13} \times 25}{\cancel{2}}} = \frac{\cancel{13} \times 13}{\cancel{13} \times 25} = \frac{13}{25}$$

Notice: What did we do with $n(S)$ here? We never actually used it. All we did was cancel it out. This same thing happens *whenever* we are trying to find the conditional probability of one event, given that another event occurred, *when the events are defined on an equiprobable sample space.* That is, if E and F are defined on an equiprobable sample space S, we get:

$$Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]} = \frac{\frac{n(E \cap F)}{n(S)}}{\frac{n(F)}{n(S)}} = \frac{n(E \cap F)}{\cancel{n(S)}} \times \frac{\cancel{n(S)}}{n(F)}$$

and so the two $n(S)$'s cancel one another out. Thus we get:

Theorem: If E and F are events defined on the same equiprobable sample space, with $Pr[F] \neq 0$, then

$$Pr[E|F] = \frac{n(E \cap F)}{n(F)}$$

Notice: This is actually what we did in Example 2.18(b), where we were looking for the probability that a randomly selected card is a heart when it is known that the card is red. We restricted our attention to only the red cards and determined what proportion of those were (also) hearts.

Example 2.22. *An urn contains 10 red balls, 15 white balls and 25 black balls. Three balls are drawn (together) from the urn. What is the probability that exactly one of the balls drawn was red, given that:*
(a) exactly 2 black balls were drawn?
(b) at most 1 black ball was drawn?

The first thing we need to do is define an equiprobable sample space. Or do we? According to the previous result, we won't actually use $n(S)$ in our calculations. Does this mean that we don't need to define S? **NO**. We still need to be sure that we define the events of interest in terms of (i.e. as subsets of) an equiprobable sample space. Otherwise, we can't use the formula. To clarify our thoughts about what the possible outcomes of the experiment are, to ensure that we count properly the number of them in each event of interest, we should always start by stating what the sample space is and ensuring that it is an equiprobable sample space.

In the experiment, the possible outcomes are sets of 3 balls which might be drawn. Therefore we let S be the set of all ways of drawing 3 balls from the urn. As long as the balls are drawn randomly, S is an equiprobable sample space for the experiment. Because the urn contains $10 + 15 + 25 = 50$ balls, then $n(S) = \binom{50}{3}$. We won't actually need that number, but noting what it is may help us to define the events properly. (For instance, it may help us to remember that every event must always have 3 balls chosen, i.e. we must always be counting ways of drawing 3 balls.)

Next, we need to define the events of interest and express what we are trying to find. In both parts of the problem, we are interested in the event that exactly one of the balls drawn was red. Call this event R.

(a) In the first part of the problem, we are looking for the probability that exactly one red ball was drawn, given that exactly 2 black balls were drawn. Let E be the event that exactly 2 black balls were drawn. Then what we are asked to find is $Pr[R|E]$. Since we are working with an equiprobable sample space, we know that this can be calculated as $Pr[R|E] = \frac{n(R\cap E)}{n(E)}$, so we need to find both $n(R \cap E)$ and $n(E)$.

If exactly 2 of the 3 balls drawn were black, then there must have been one ball drawn which was not black, i.e. was either red or white. The urn contains 10 red balls and 15 white balls, for a total of 25 non-black balls. Therefore the number of different ways of drawing 2 of the 25 black balls and one of the 25 non-black balls is

$$n(E) = \binom{25}{2}\binom{25}{1}$$

Of course, $R \cap E$ is the set of all ways that exactly one red ball and also exactly 2 black balls can be drawn. That is, when the 3 balls were drawn, the draw must have consisted of 1 of the 10 red balls and 2 of the 25 black balls. The number of way in which this could happen is

$$n(R \cap E) = \binom{10}{1}\binom{25}{2}$$

Now, we can find the conditional probability we are asked for as

$$Pr[R|E] = \frac{n(R\cap E)}{n(E)} = \frac{\binom{10}{1}\cancel{\binom{25}{2}}}{\cancel{\binom{25}{2}}\binom{25}{1}} = \frac{10}{25} = \frac{2}{5}$$

(b) This time, we need to find the probability that exactly one red ball was drawn, given that *at most* 1 black ball was drawn. In order to have at most one black ball drawn, we could either have no black balls drawn or else have 1 black ball drawn. Let F be the event that at most one black ball is drawn. Then F contains all the ways of drawing no black balls as well as all the ways of drawing exactly 1 black ball. The conditional probability we've been asked to find is $Pr[R|F]$.

If no black ball is drawn, then all 3 of the balls drawn must have come from among the 25 non-black balls in the urn. There are $\binom{25}{3}$ different ways to draw 3 of these balls. Similarly, if exactly 1 of the 25 black balls was drawn, then there must also have been 2 of the 25 non-black balls drawn. This can happen in $\binom{25}{1}\binom{25}{2}$ ways. Therefore we have

$$\begin{aligned} n(F) &= \binom{25}{3} + \binom{25}{1}\binom{25}{2} = \frac{25!}{3!22!} + \left(25 \times \frac{25!}{2!23!}\right) = \frac{25 \times \cancel{24}^{4} \times 23}{\cancel{3} \times \cancel{2}} + \frac{25 \times 25 \times \cancel{24}^{12}}{\cancel{2}} \\ &= 25 \times (4 \times 23 + 25 \times 12) = 25 \times (92 + 300) = 25 \times 392 \end{aligned}$$

Since $Pr[R|F] = \frac{n(R\cap F)}{n(F)}$, we must also find $n(R \cap F)$. $R \cap F$ is the event that exactly one red ball at at most one black ball were drawn. Again, we must consider 2 possibilities: there could be 1 red ball and no black balls, or there could be 1 red ball and 1 black ball. Of course, if there is 1 red ball and no black ball, then 2 of the white balls must have been drawn. There are $\binom{10}{1}\binom{15}{2}$ different ways to draw 1 of the 10 red balls and 2 of the 15 white balls (and therefore 0 of the 25 black balls). Similarly, if one red ball and one black ball are drawn, then there must also have been one white ball drawn. There are $\binom{10}{1}\binom{25}{1}\binom{15}{1}$ ways to draw 1 of the 10 red balls, 1 of the 25 black balls and 1 of the 15 white balls. We see that

$$\begin{aligned} n(R \cap F) &= \binom{10}{1}\binom{15}{2} + \binom{10}{1}\binom{25}{1}\binom{15}{1} = \left(10 \times \frac{15!}{2!13!}\right) + (10 \times 25 \times 15) \\ &= \left(10 \times \frac{15 \times \cancel{14}^{7}}{\cancel{2}}\right) + (10 \times 15 \times 25) = 10 \times 15 \times (7 + 25) = 10 \times 15 \times 32 \end{aligned}$$

Now, we can find $Pr[R|F]$.

$$Pr[R|F] = \frac{n(R \cap F)}{n(F)} = \frac{\cancel{10}^{2} \times \cancel{15}^{3} \times 32}{\cancel{25} \times 392} = \frac{\cancel{2} \times 3 \times 32}{\cancel{392}^{196}} = \frac{96}{196} == \frac{\cancel{4} \times 24}{\cancel{4} \times 49} = \frac{24}{49}$$

Math 1228A/B Online

Lecture 15:

Independent Events

(text reference: Section 2.4, pages 81 - 82)

Independence of Events

Consider the following problem:

Example 2.23. *A card has been drawn from a standard deck. If it is known that the card is an Ace, what is the probability that the card is a club?*

Our experiment is to draw one card from a deck. The possible outcomes are the 52 different cards in the deck, each of which is equally likely to be chosen. That is, if we let S be the set of cards in the deck, then S is an equiprobable sample space, with $n(S) = 52$.

We know that the drawn card is an Ace, and need to find the probability that the card is a club. Defining A to be the event that the chosen card is an Ace and C to be the event that the chosen card is a Club, we are looking for $Pr[C|A]$.

Since S is an equiprobable sample space, we know that $Pr[C|A] = \frac{Pr[C\cap A]}{Pr[A]} = \frac{n(C\cap A)}{n(A)}$. What is the event $C \cap A$? This is the event that the chosen card is a Club and is also an Ace. Of course, there's only one card in the deck that is both a Club and an Ace, and that's the Ace of Clubs, so we have $n(C \cap A) = 1$. Also, A is the set containing the 4 Aces in the deck, so $n(A) = 4$. We get:

$$Pr[C|A] = \frac{Pr[C \cap A]}{Pr[A]} = \frac{n(C \cap A)}{n(A)} = \frac{1}{4}$$

We see that if we know that an Ace has been drawn, the probability of getting a Club is $\frac{1}{4}$. Suppose that we hadn't known that an Ace had been drawn. What would the probability of getting a Club be then? i.e. What is $Pr[C]$? Since 13 of the 52 cards in the deck are Clubs, then $n(C) = 13$, so we have:

$$Pr[C] = \frac{n(C)}{n(S)} = \frac{13}{52} = \frac{1}{4}$$

But this is just the same as $Pr[C|A]$! That is, we have $Pr[C|A] = Pr[C]$.

In this case, we see that the additional information that the card which had been drawn was an Ace *did not change* the probability that a club was drawn.

When knowledge of whether or not one event occurs has no effect on the probability that another event also occurs, we say that the events are **independent** events. The way we actually define independent events is a bit different from this, though.

Suppose we have two events, E and F, for which knowing whether or not F occurs doesn't change the probability that E occurs, so that we have $Pr[E|F] = Pr[E]$. Then we have

$$\frac{Pr[E \cap F]}{Pr[F]} = Pr[E] \qquad \Rightarrow \qquad Pr[E \cap F] = Pr[E] \times Pr[F]$$

This is the relationship we use to define the concept of *independent events.*

Definition: Two events, E and F, are called **independent events** if

$$Pr[E \cap F] = Pr[E] \times Pr[F]$$

Note: If 2 events are *not* independent events, we say that they are **dependent events**.

Also Note: Although this relationship, $Pr[E \cap F] = Pr[E] \times Pr[F]$ seems, for some reason , to be assumed by many students to be a general rule, this is **not the case**. It is **not** generally true. It is *only* true if E and F are independent events and can *only* be assumed to be true if it is **known** that the events are independent. That is, although we use the rule "*and* means multiply" in **counting**, because of the FCP, that rule **does not apply, in general, to probabilities**. (Actually, in counting we also need some sort of independence to be in effect in order for the FCP to apply, because it only applies when the number of choices for the second decision is the same no matter how the first decision is made.)

We can use the relationship $Pr[E \cap F] = Pr[E] \times Pr[F]$ to *check*, i.e, *determine* whether 2 events, E and F are independent. (We must, though, know or calculate $Pr[E \cap F]$ by some means that does not *assume* independence, in order to do this.) For instance, in our example, we determined the meaning of the event $A \cap C$ and used this, together with the fact that we were using an equiprobable sample space, to (effectively) calculate $Pr[A \cap C] = \frac{n(A \cap C)}{n(S)} = \frac{1}{52}$. We calculated $Pr[A] = \frac{4}{52}$ and $Pr[C] = \frac{13}{52}$ in a similar way. In this case, we see that

$$Pr[A] \times Pr[C] = \frac{\not{4}}{52} \times \frac{\not{1}3}{\not{5}2} = \frac{1}{52} = Pr[A \cap C]$$

so (as expected) A and C in this case *do* satisfy the definition of independent events.

Most events, though, are dependent events. That is, 2 events being independent is a special relationship that is relatively rare. Consider, for instance, the next couple of examples.

Example 2.24. *A single card is drawn at random from a standard deck. Let A be the event that the card is an Ace and F be the event that the card is a face card (i.e. is a Jack, Queen or King).*
(a) Find $Pr[A \cap F]$. *(b) Find $Pr[A \cap F^c]$.*
(c) Are A and F independent events? *(d) Are A and F^c independent events?*

Here S is the set of all cards which might be drawn, i.e. is the set of all cards in the deck, so $n(S) = 52$.

(a) Since an Ace is not a face card, then it is not possible to draw an Ace and also draw a face card when a single card is drawn. That is, there is no card which is both an Ace and a face card. Therefore A and F are mutually exculsive events, and $Pr[A \cap F] = 0$. (That is, the set F containing all the face cards in the deck does not contain any Aces, so $A \cap F = \emptyset$ and $Pr[A \cap F] = 0$.)

(b) $A \cap F^c$ is the set containing all the cards which are Aces and are not face cards. Since all Aces are not face cards, this set contains all the Aces (and no other cards). That is, since the set of all Aces is contained within the set of all cards which are not face cards, i.e. $A \subseteq F^c$, then taking the intersection of these sets simply picks all of set A from the set F^c, so $A \cap F^c = A$. Therefore $Pr[A \cap F^c] = Pr[A] = \frac{4}{52} = \frac{1}{13}$ (since there are 4 Aces in the 52 card deck).

(c) Are A and F independent events? No. *Mutually exclusive events can never be independent events* (unless one of the events is actually impossible). That is, if 2 events, E_1 and E_2 are mutually exclusive events, then $Pr[E_1 \cap E_2] = 0$. And it is not possible to have $Pr[E_1] \times Pr[E_2] = 0$ unless either $Pr[E_1] = 0$ or $Pr[E_2] = 0$ (or both). (That is, the only way that the product of 2 numbers can be 0 is if one or both of the numbers is 0.)

In this case, we know that A and F are mutually exclusive events, so that $Pr[A \cap F] = 0$. Also, we have already calculated $Pr[A] = \frac{1}{13}$. Of course, the event F contains 4 Jacks, 4 Queens and 4

Kings, so $n(F) = 12$ and $Pr[F] = \frac{12}{52} = \frac{3}{13}$. We see that

$$Pr[A] \times Pr[F] = \frac{1}{13} \times \frac{3}{13} = \frac{3}{169} \neq 0$$

so $Pr[A \cap F] \neq Pr[A] \times Pr[F]$. A and F are not independent events, they are dependent events.

(d) Now, we need to determine whether A and F^c are independent events. We have seen that $Pr[A \cap F^c] = Pr[A] = \frac{1}{13}$. And since $Pr[F] = \frac{3}{13}$, then $Pr[F^c] = 1 - \frac{3}{13} = \frac{10}{13}$. (That is, there are 10 non-face cards, Ace and 2 through 10, in each of the 4 suits, so 40 of the 52 cards are not face cards.) We see that

$$Pr[A] \times Pr[F^c] = \frac{1}{13} \times \frac{10}{13} \neq \frac{1}{13}$$

so $Pr[A \cap F^c] \neq Pr[A] \times Pr[F^c]$ and we see that A and F^c are not independent events. They are dependent events.

Notice: Be sure you understand the very different meanings of *mutually exclusive* events and *independent* events. If E and F are any 2 (non-empty) events defined on the same sample space, then they might be mutually exclusive events, or they might be independent events, or they might be neither of these. (But as observed in part (c), they can never be both.)

Of course, sometimes we know probabilities through some means other than calculating them using an equiprobable sample space. In those cases, we simply use the known probabilities to determine whether 2 events are independent of one another.

Example 2.25. *In the donut shop from before, according to the information in Examples 2.5 and 2.7 , are the events "a customer buys at least half a dozen donuts" and "a customer buys only donuts" independent events or dependent events?*

In Example 2.5 (see Lecture 12 page 61), we defined D to be the event that a customer buys only donuts, and we had $Pr[D] = .25$. In Example 2.7 (see Lecture 12 page 64), we defined A to be the event that a customer buys at least half a dozen donuts, and we found that $Pr[A] = .2$. To determine whether or not these events are independent, we need to check whether $Pr[A \cap D] = Pr[A] \times Pr[D]$.

We have $Pr[A] \times Pr[D] = (.2)(.25) = .05$, but in Example 2.7 we were told that $Pr[A \cap D] = .15$, so we see that $Pr[A] \times Pr[D] \neq Pr[A \cap D]$. Therefore A and D are *not* independent events, they are dependent events.

In some situations, we *know* that 2 events are independent, either by being told, or by our own understanding of what the events are. In those situations, we can *use* that knowledge to find the probability that they both happen.

Example 2.26. *Consider again the donut shop from the previous example. If it has been determined that the events "a customer buys only coffee" and "a customer stays more than 5 minutes" (ref. examples 2.5 and 2.6) are independent events, then what proportion of customers stay longer than 5 minutes and only buy coffee?*

In Example 2.5 (see Lecture 12 pg. 61), we defined C to be the event that a customer buys only coffee, with $Pr[C] = .3$, i.e. 30% of the customers buy only coffee. Also, in Example 2.6 (see Lecture 12 pg. 62), we defined E to be the event that a customer stays more than 5 minutes, with $Pr[E] = .4$, i.e. 40% of customers stay more than 5 minutes.

We are told that 'it has been determined that' these are independent events and are asked to find the proportion of customers who stay longer than 5 minutes and buy only coffee, i.e. to find $Pr[E \cap C]$. Since E and C are known to be independent, we see that

$$Pr[E \cap C] = Pr[E] \times Pr[C] = .4 \times .3 = .12$$

Therefore 12% of all customers stay more than 5 minutes and buy only coffee.

Example 2.27. *It is known that A and B are independent events defined on a sample space S. If $Pr[A] = .4$ and $Pr[B] = .3$, find*

(a) $Pr[A \cup B]$ *(b) $Pr[A|B]$*

(a) We know that $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$. We are told that $Pr[A] = .4$ and $Pr[B] = .3$, and also that A and B are independent events. This tells us that

$$Pr[A \cap B] = Pr[A] \times Pr[B] = (.4)(.3) = .12$$

and so we get

$$Pr[A \cup B] = .4 + .3 - .12 = .70 - .12 = .58$$

(b) We are asked to find $Pr[A|B]$. We have:

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{.12}{.3} = \frac{12}{30} = \frac{4}{10} = .4$$

Of course, this is just $Pr[A]$. Remember, A and B being independent events means that knowing whether or not B occurs does not change the probability that A occurs, so whenever A and B are independent events, it is always true that $Pr[A|B] = Pr[A]$. That is, when A and B are independent events, $Pr[A \cap B] = Pr[A] \times Pr[B]$, so we have

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{Pr[A] \times \cancel{Pr[B]}}{\cancel{Pr[B]}} = Pr[A]$$

All we did in the calculations in part (b) here was to undo the calculation $.4 \times .3 = .12$ that we did in part (a), in order to cancel out the .3's.

Example 2.28. *A single die is tossed twice in succession. What is the probability that a 2 comes up both times?*

Approach 1:
We have seen that the set S containing all ordered pairs (i, j) where each of i and j is a number between 1 and 6 is an equiprobable sample space for any experiment involving tossing two dice, or tossing a single die twice. We have $n(S) = 36$, and $(2, 2)$ is the only outcome which has a 2 coming up on both tosses, so the probability that this occurs is $\frac{1}{36}$.

Approach 2: (a bit easier)
We *know* that the number which comes up the first time a die is tossed has no effect on what number comes up when we toss the die a second time. That is, we know that the 2 tosses are independent of one another, so the events E: a 2 comes up on the first toss, and F: a 2 comes up on the second toss, are independent events. Of course, we have $Pr[E] = Pr[F] = \frac{1}{6}$, so knowing that E and F are independent tells us that the probability that a 2 comes up both times is

$$Pr[E \cap F] = Pr[E] \times Pr[F] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Remember, if we know from our understanding of the situation that 2 events are independent, then we can use that information. However, we must be careful when dealing with more complicated events. Not all events involving the appearance of 1 or more 2's when a die is tossed twice are independent events. For instance, here we have E and F defined as the events that a 2 comes up on the first or second toss, respectively. As we have seen, these are independent events. However, if we have a more complicated event, such as A: the event that *at least one* 2 comes up, then A and E are not independent events (nor are A and F). The set A contains all the tosses in which a 2 comes up on the first die as well as all the tosses in which a 2 comes up on the second die, so $n(A) = 11$ and $Pr[A] = \frac{11}{36}$ (i.e. 6 of each type of outcome, but we don't count (2, 2) twice). Of course $A \cap E = E$ (since whenever the first die comes up 2, at least one 2 comes up) and likewise $A \cap F = F$, so we have $Pr[A \cap E] = Pr[A \cap F] = \frac{1}{6}$. Therefore $Pr[A] \times Pr[E] = \frac{11}{36} \times \frac{1}{6} \neq Pr[A \cap E]$ and likewise $Pr[A] \times Pr[F] \neq Pr[A \cap F]$.

Math 1228A/B Online

Lecture 16:
Probability Trees and Their Properties

(text reference: Section 2.5, pages 86 - 88)

2.5 Stochastic Processes

A **stochastic process** is a sequence of probabilistic experiments. This may involve several repetitions of the *same* experiment, or different experiments performed one after another.

Examples of stochastic processes:

1. Flip a coin several times.
2. Randomly choose one of several containers which contain some balls, then draw one or more balls from the chosen container.
3. Draw a card from a standard deck. Then draw another card, without putting the first one back (i.e. draw 2 cards from a standard deck, *without replacement*).
4. Draw a card from a standard deck. If the card is a heart, put it back and draw another card.

When we need to analyze a stochastic process, it is often useful to draw a **probability tree**. A probability tree is a *model* of a stochastic process. These trees look very much like the counting trees we have already used. However, there are some important differences.

A probability tree contains *a level for each experiment* in the stochastic process. The *branches* on a particular level correspond to the various *possible outcomes or events* that could be observed when the experiment corresponding to that level is performed.

Tip: The branchings on a particular level should represent events of the corresponding experiment, defined as broadly as possible while still giving sufficient detail for the required analysis. For instance, if the expermiment being modelled on a particular level of the tree is “toss a die”, and we are *only* interested in whether the number that comes up is odd or even, use branchings for the events “odd” and “even”, rather than for the individual sample points / possible outcomes “toss a 1”, “toss a 2”, ..., “toss a 6”. Similarly, if the experiment is “draw a card” and we only care whether or not a Heart is drawn, use branchings for the 2 events “draw a Heart” and “draw a card which is not a Heart”, rather than branchings for all 4 suits, or for all of the 52 cards which might be drawn. This approach produces a smaller tree which is much easier to draw and to work with.

The set of outcomes which might be observed when an experiment is performed *may* depend on the outcomes of previous experiments which occur earlier in the tree. For instance, suppose we have 2 urns, one containing white balls and black balls and the other containing white balls and red balls. If we pick an urn at random and then draw a ball from that urn, the the event “draw a black ball” can only occur if we picked the urn that has some black balls in it, i.e. if a particular event occurred on the earlier experiment.

Another possibility is that the same outcomes may be possible, but the *probabilities* of observing these outcomes may depend on the outcome of an earlier experiment. For instance, if we draw 2 cards from a deck without replacement, then the probability that the second card is red depends on whether the first card was red or black. When we draw the second card, there are 51 cards in the deck. If the first card was black, then 26 of the 51 remainiing cards are red. But if the first card was red, then only 25 of the remaining 51 cards are red.

Sometimes, whether or not a later experiment is performed at all depends on the outcome of some earlier experiment. For instance, for the stochastic process “Draw a card from a standard deck. If the card is a Heart, put it back and draw another card.”, the second experiment is *only* performed if the event “a Heart was drawn” occurred on the first experiment.

Consider a simple case in which we have a stochastic process involving 2 probabilistic experiments to be performed one after the other. On the first experiment, there are 3 possible outcomes:

A, B and C. On the second experiment, we observe whether or not some event D occurs.

The probability tree modelling this stochastic process has 2 levels: one for the first experiment and one for the second. On the first level, there are branches for the 3 possible outcomes, A, B and C. On the second level, for *each* of the possible outcomes of the first experiment, we have 2 second level branches, for whether D does or doesn't occur.

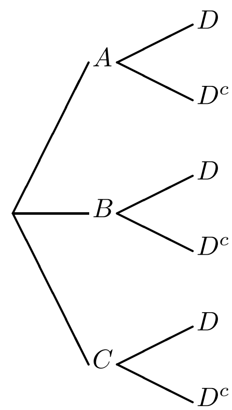


Figure 1: Structure of the probability tree

So far, this could be a counting tree. And just as with counting trees, there are numbers associated with each of the branches in a probability tree. However the numbers on a probability tree are different from the numbers on a counting tree, and they have *different properties*.

On a probability tree, the number associated with a particular branch is the *probability* that the outcome or event corresponding to that branch will be observed, *if*, in the course of the stochastic process, it transpires that we are in a position in which that branch might be taken. That is, we think about travelling through the tree as we perfrom the experiments in the stochastic process. When we perform the first experiment, one of the possible outcomes will be observed, so we will travel along a particular first level branch. At the second level, the only branches which we could possibly continue along are the ones growing from the end of the branch we took on the first level. So the probability that we will take any particular one of these second-level branches is the probability that the corresponding outcome or event will occur, *given that* the first-level outcome or event leading to that branch has occurred. Therefore after the first level of the tree, the probabilities associated with other levels are always *conditional* probabilities.

In our example, the probabilities associated with the first level branches are simply the probabilities that the corresponding outcomes occur when we perform the first experiment, i.e. $Pr[A]$, $Pr[B]$ and $Pr[C]$. When we get to the second level, which branches we might travel along depends on which outcome was observed in the first experiment, i.e. which branch we travelled along at the first level.

For instance, the top pair of D and D^c branches can *only* be taken if we observed outcome A on the first experiment. That is, we will not be in a position to take one of these branches unless we go along the A branch at level 1. So the probability that we take this D branch is the conditional probability that event D occurs in the second experiment, given that outcome A occurred in the first experiment, i.e. $Pr[D|A]$. By the same reasoning, the probability associated with the D^c branch of this pair is $Pr[D^c|A]$. Similarly, the probabilities associated with the other second level branches are (reading down the tree) $Pr[D|B]$, $Pr[D^c|B]$, $Pr[D|C]$ and $Pr[D^c|C]$. The completed tree is shown on the next page.

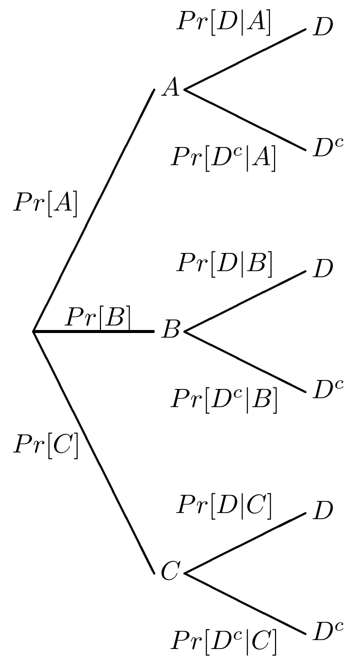


Figure 2: Completed probability tree

Properties of Probability Trees

Let's think about the probabilities on this tree. When the first experiment is performed, it is a certainty that one of the 3 outcomes on the first level of the tree will be observed, and only one of them can occur at a time, so the sum of the probabilities on the first level branches must be 1.

Now, suppose that we have performed the first experiment and observed outcome A. Then we are, in effect, standing at the end of the A branch, waiting to see which branch to travel along next. Again, it is a certainty that one (and only one) of D and D^c will occur, so the sum of the probabilities on the two branches ahead of us must again be 1. Likewise, each of the other second level *pairs* of branches must have probabilities which sum to 1. That is, we see that:

> **Property 1:**
> In any probability tree, the probabilities on all branches growing out of a single point in the tree *must always* sum to 1.

The various *paths* through the tree correspond to the various compound outcomes which could be observed when the experiments are performed. The terminal point in a path corresponds to the events on all of the branches leading to that terminal point having occurred. For instance, one of the paths through our sample tree involves taking the A branch on the first level and then taking the D branch (leading from the A branch) on the second level. So the only way that we could end up at the terminal point at the end of this branch is if outcome A is observed on the first experiment, and then event D occurs on the second experiment. That is, events A and D must both occur, so the event corresponding to this terminal point is the event $A \cap D$, and the probability that we take this particular path when we travel through the tree is $Pr[A \cap D]$.

Of course, we know that $Pr[A \cap D]$ can be calculated as $Pr[D|A] \times Pr[A]$. And these are exactly the probabilities on the branches in the path leading to this terminal point. So the probability that we will take this particular path when we travel through the tree can be found by multiplying the probabilities on the branches in the path.

In general, we define:

> *Definition:* Each terminal point in a probability tree corresponds to the intersection of the events along the path leading to that terminal point. The probability of this event is the probability that this path will be taken through the tree, which is called the **path probability**.

and we have:

> **Property 2:**
> The path probability of any path through the tree is the product of the probabilities on the branches along that path.

For convenience, we sometimes write the path probabilities at the ends of the paths.

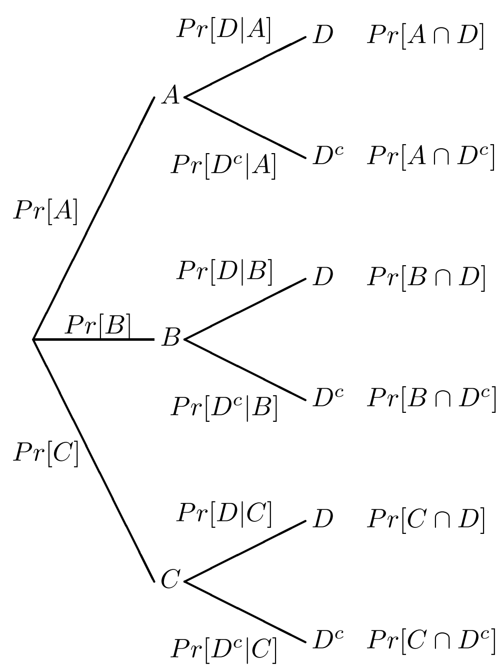


Figure 3: Probability tree with Path Probabilities shown

Of course, these terminal points correspond to *all* of the possible outcomes which could be observed when the sequence of experiments, i.e. the stochastic process, is performed. That is, when we perform the experiments, *exactly* one of these paths will be taken, so exactly one of these terminal point events will occur. Therefore, we have the following

> **Property 3:**
> In any probability tree, the sum of all path probabilities *must always* be 1.

Now, suppose we need to find the probability of some particular event which could occur during the stochastic process. For instance, suppose we need to find the probability that event D occurs in the stochastic process modelled by the tree we have drawn above. Notice that there is no single path or single branch corresponding to the event D. In the tree, this second experiment event is partitioned according to which outcome occurred on the first experiment. This partitioning tells us how to find $Pr[D]$. That is, since *any* outcome of the stochastic process will fit into one and only one (i.e. exactly one) of the descriptions "A occurred on experiment 1", "B occurred on experiment 1" or "C occurred on experiment 1", the we can partition the event "D occurred on experiment 2" into disjoint pieces corresponding to these various first experiment outcomes, so that

$$Pr[D] = Pr[D \cap A] + Pr[D \cap B] + Pr[D \cap C]$$

And these probabilities that we need to sum are simply the path probabilities for all of the paths on which there is a D branch.

Similarly, if we needed to find the probability of the event E: "neither A nor D occurred", we can partition "A did not occur" into "B occurred" and "C occurred", and recognize that event E occurs whenever D does not occur (i.e. D^c occurs) on experiment 2 and either B or C occurs on experiment 1, so that

$$Pr[E] = Pr[B \cap D^c] + Pr[C \cap D^c]$$

Once again, the probabilities we need to sum are the *path probabilities* for *all paths on which event E occurs* (because D occurred and either B or C occurred).

We can *always* do this sort of thing, to find the probability of *any* event which can be defined in terms of possible outcomes of the stochastic process, as represented by branches on the tree. So we have:

Property 4:
For any event E which may occur in a stochastic process, $Pr[E]$ is given by the sum of the path probabilities for all paths in which the event E occurs.

Let's put some actual numbers on the tree we've been looking at. Suppose that when experiment 1 is performed, outcome A occurs half the time and B occurs one-fifth of the time (with C of course occurring the rest of the time. That is, suppose that $Pr[A] = \frac{1}{2}$ and $Pr[B] = \frac{1}{5}$, so that $Pr[C] = 1 - \left(\frac{1}{2} + \frac{1}{5}\right) = 1 - \left(\frac{5}{10} + \frac{2}{10}\right) = \frac{10}{10} - \frac{7}{10} = \frac{3}{10}$. Also, suppose that event D occurs three-fifths of the time when A has occurred, half of the time when B has occurred and two-thirds of the time when C has occurred. Then we have $Pr[D|A] = \frac{3}{5}$, $Pr[D|B] = \frac{1}{2}$ and $Pr[D|C] = \frac{2}{3}$. We put these probabilities on the tree and then use **Property 1** (that the probabilities at any branching must sum to 1) to fill in the missing probabilities and complete the tree (see *Figure 3a*, next page. (*Note:* Property 1 is what we used to find $Pr[C] = \frac{3}{10}$.)

Now, we can use **Property 2** to find the path probabilities, by multiplying the probabilities along the path. As before, for convenience we write the path probabilities at the ends of the corresponding paths. (See *Figure 3b:*, next page.)

Notice that **Property 3** holds, i.e. the sum of all of the path probabilities is 1:

$$\frac{3}{10} + \frac{2}{10} + \frac{1}{10} + \frac{1}{10} + \frac{2}{10} + \frac{1}{10} = \frac{10}{10} = 1$$

We can use **Property 4** and simply sum certain path probabilities to find $Pr[D]$ and $Pr[E]$. (*Recall:* E is the event that neither A nor D occurs.) To find $Pr[D]$, we simply add up all path probabilities for paths in which there is a D branch:

$$Pr[D] = \frac{3}{10} + \frac{1}{10} + \frac{2}{10} = \frac{6}{10} = \frac{3}{5}$$

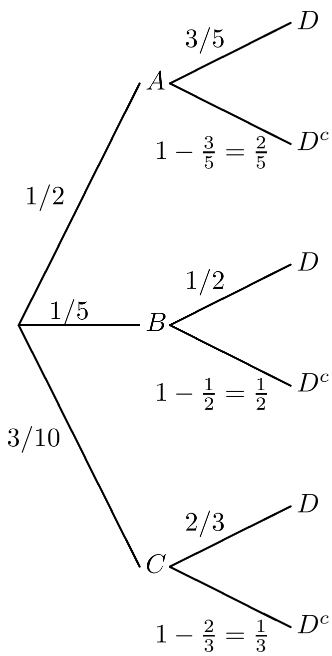


Figure 3a:
Probability tree with actual probabilities

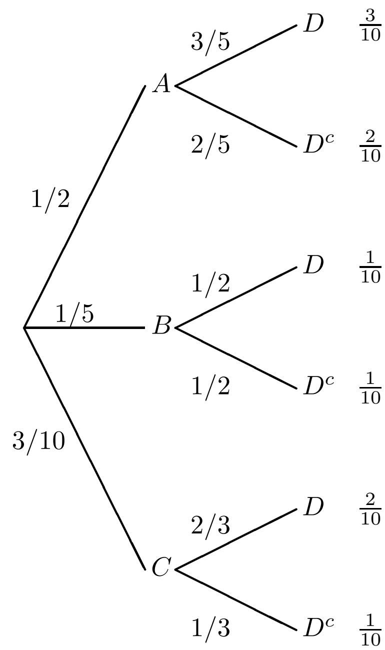


Figure 3b:
... and with path probabilities shown

To find $Pr[E]$, we identify all paths on which E occurs, i.e. on which neither A nor E occurs, so we need all paths in which there is no A branch and also no D branch. Looking at the tree we can identify these paths:

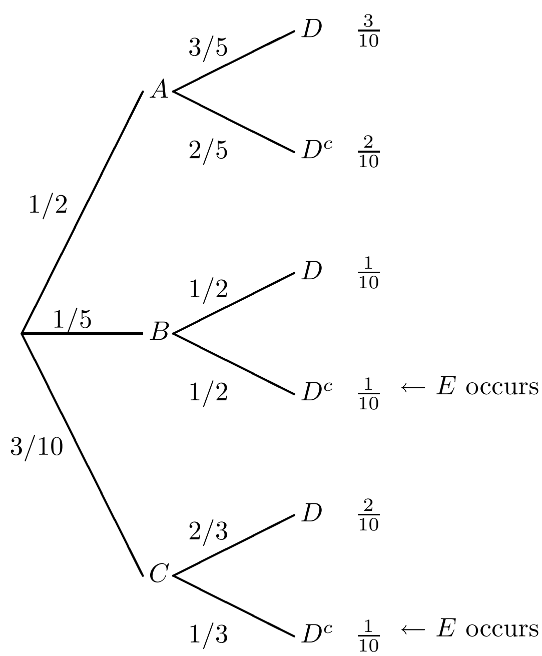


Figure 4: Probability tree with paths for event E shown

We see that $Pr[E] = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \frac{1}{5}$.

Let's look at how we use a probability tree in an actual example.

Example 2.29. *There are 2 identical-looking urns containing red and blue balls. Urn A contains 2 blue balls and 3 red balls. Urn B also contains 2 blue balls, but contains only 1 red ball. One of the urns is chosen at random and a single ball is drawn at random from that urn.*
(a) Draw a probability tree to model this stochastic process and find all path probabilities.
(b) What is the probability that the ball which is drawn is blue?
(c) When the experiment was performed, the ball which was drawn was blue. What is the probability that it was urn A from which the ball was drawn?

(a) The stochastic process described here consists of 2 experiments:
Experiment 1: Choose an urn.
Experiment 2: Pick a ball from that urn
Therefore the probability tree modelling this stochastic process has 2 levels, one for each of these experiments. In the first experiment, either Urn A or Urn B will be chosen, so there are 2 possible outcomes, A and B. (That is, let A be the event that urn A is chosen and B be the event that urn B is chosen.) This means that we need 2 branches on the first level of the tree.

In the second experiment, no matter which urn was chosen, there are 2 possible outcomes:
R: a red ball is drawn
R^c: a blue ball is drawn
This gives 2 second-level branches growing from the end of each of the first-level branches.

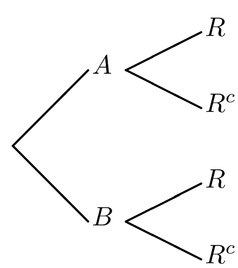


Figure 5: Structure of the tree

Now, we need to put probabilities on the branches of the tree. Consider the first experiment. Since one of the 2 urns is chosen at random, these 2 outcomes are equally likely to occur, so we have $Pr[A] = Pr[B] = \frac{1}{2}$. These are the probabilities associated with the A and B branches on the first level of the tree. (Notice that the sum of the probabilities on all level 1 branches is 1.)

The probabilities of the two second experiment outcomes depend on which urn we're drawing from.

<u>Urn A:</u>
There are 2 blue balls and 3 red balls (a total of 5 balls) in Urn A. Since 3 of the 5 balls are red, the probability of drawing a red ball is $\frac{3}{5}$. Similarly, 2 of the 5 balls are blue, so the probability of drawing a blue ball is $\frac{2}{5}$. That is, the probabilities on the R and R^c branches growing out of the A branch are $Pr[R|A] = \frac{3}{5}$ and $Pr[R^c|A] = \frac{2}{5}$.
<u>Urn B:</u>
On the other hand, in Urn B, there is only 1 red ball in with the 2 blue balls (3 balls in total), so the probability that we get a red ball is $\frac{1}{3}$, while the probability that we get a blue ball is $\frac{2}{3}$. That is, we have $Pr[R|B] = \frac{1}{3}$ and $Pr[R^c|B] = \frac{2}{3}$.

Notice that in both cases, the probabilities sum to 1.

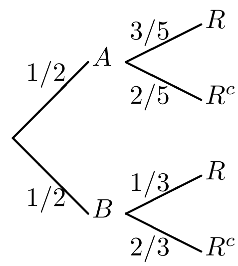


Figure 6: Completed probability tree

We find the path probabilities by multiplying the probabilities on the branches in the path. For instance, the topmost path, with an A branch followed by an R branch, has probabilities $\frac{1}{2}$ and $\frac{3}{5}$ showing on these branches, so the path probability is $\frac{1}{2} \times \frac{3}{5} = \frac{3}{10}$. Of course, this path probability is the probability that when then experiment is performed, this path will be taken, which is the probability that urn A is chosen and a red ball is chosen. So what we have calculated is $Pr[A \cap R] = \frac{3}{10}$. We calculate the other path probabilities in a similar way.

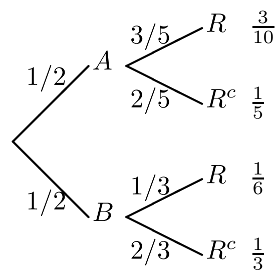


Figure 7: Probability tree with path probabilities shown

Notice that the sum of all of the path probabilities is

$$\frac{3}{10} + \frac{1}{5} + \frac{1}{6} + \frac{1}{3} = \frac{3}{10} + \frac{2}{10} + \frac{1}{6} + \frac{2}{6} = \frac{5}{10} + \frac{3}{6} = \frac{1}{2} + \frac{1}{2} = 1$$

(b) We are asked to find the probability that a blue ball is drawn. We partition the event "the ball is blue" (i.e. R^c) with respect to which urn is chosen. That is, we need to sum the path probabilites for all paths in which R^c occurs. We get:

$$Pr[R^c] = Pr[A \cap R^c] + Pr[B \cap R^c] = \frac{1}{5} + \frac{1}{3} = \frac{3}{15} + \frac{5}{15} = \frac{8}{15}$$

Notice: In this situation, we have a total of 8 balls, of which 4 (i.e. half) are blue. However the probability of drawing a blue ball is *not* $\frac{1}{2}$. This is because the balls are not equally likely to be drawn, since the urns contain different numbers of balls. That is, each ball in urn B, which has fewer balls, is more likely to be drawn than those in urn A.

(c) We are told that a blue ball was drawn and are asked to find the probaility that urn A was chosen. So what probability are we looking for here? We want the probability that Urn A was chosen, *given that* we know that a blue ball was drawn, i.e. $Pr[A|R^c]$. This probability does not appear anywhere on the tree. We find this conditional probability in the usual way, using the *definition* of conditional probability, and the probabilities we've already found. We have

$$Pr[A|R^c] = \frac{Pr[A \cap R^c]}{Pr[R^c]} = \frac{\frac{1}{5}}{\frac{8}{15}} = \frac{1}{\cancel{5}} \times \frac{\cancel{15}^{3}}{8} = \frac{3}{8}$$

Math 1228A/B Online

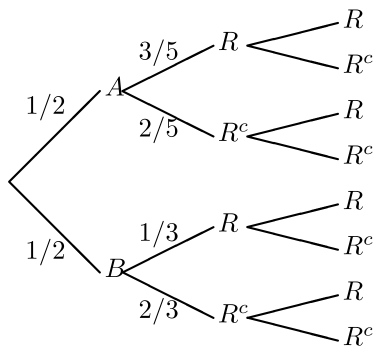
Lecture 17:

More Examples of Using Probability Trees

(text reference: Section 2.5, pages 88 - 91)

Example 2.30. *Recall the urns from Example 2.29 . This time, after the urn is selected at random, 2 balls are drawn from the chosen urn, without replacement.*
(a) What is the probability that both balls are the same colour?
(b) What is the probability that both balls are blue, given that they are the same colour?

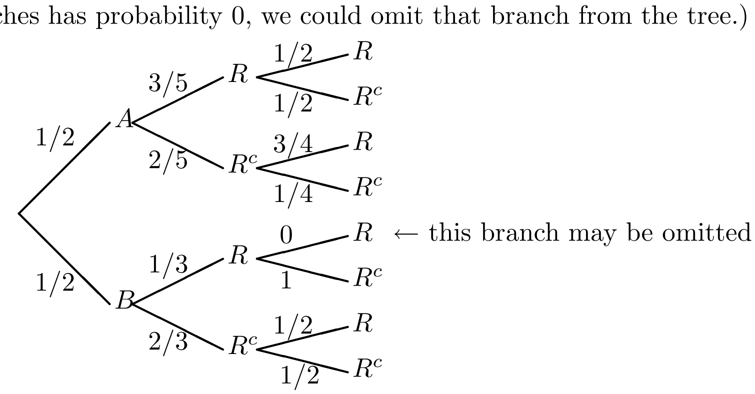
This time, we have a different stochastic process, involving 3 experiments:
Experiment 1: Choose an urn.
Experiment 2: Draw a ball from that urn
Experiment 3: Draw another ball from the same urn.
We can construct the probability tree which models this stochastic process. It has, of course, 3 levels, one for each experiment. The first 2 levels of the tree look just like our previous tree, even having the same probabilities on the branches.



Incomplete probability tree

To find the probabilities to put on the third-level branches, we need to think about what balls are in the urn when the second ball is drawn. We are drawing without replacement, so 1 ball has been removed from the urn before the second draw is performed. For instance, if we are drawing from Urn A, then the urn contains only 4 balls at the time of the second draw. Suppose we draw a red ball on the first draw. Then at the time of the second draw, the urn contains 2 blue balls and 2 red balls, so the probability of getting a red ball and the probability of getting a blue ball are both $\frac{2}{4}$. On the other hand, if we draw a blue ball on our first draw from urn A, then when we draw the second ball we are drawing from 3 red balls and 1 blue ball, so the probability of getting a red ball is $\frac{3}{4}$, while the probability of getting a blue ball is only $\frac{1}{4}$. Similarly, if we are drawing from Urn B and draw a red ball on the first draw, then at the time of the second draw, the urn contains *only* blue balls, so drawing a blue ball is a certainty, occurring with probability 1 (and drawing a red ball is impossible, occurring with probability 0). On the other hand, if the first draw produced a blue ball, then at the time of the second draw the urn contains one ball of each colour. Thus the probability of drawing a red and the probability of drawing a blue are both $\frac{1}{2}$.

We complete the tree by putting these probabilities on the level 3 branches. (*Note:* Now that we know that one of the branches has probability 0, we could omit that branch from the tree.)

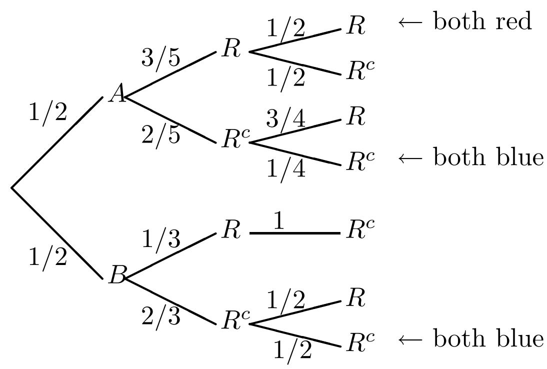


Completed probability tree

Now we're ready to look at the specific questions we were asked.

(a) What is the probability that both balls are the same colour?

Let E be the event that both balls are the same colour. We find $Pr[E]$ by adding up the path probabilities for all paths in which 2 balls the same colour are drawn.



Identifying the paths which make up event E

We see that there are 3 such paths (and if we included the branch which is taken with probability 0 there would be a fourth such path, but its path probability would be 0). We find the path probabilities, of course, by calculating the products of the probabilities on the branches in the paths. We get:

$$\begin{aligned} Pr[E] &= \left(\frac{1}{\not{2}} \times \frac{3}{5} \times \frac{\not{2}}{4}\right) + \left(\frac{1}{\not{2}} \times \frac{\not{2}}{5} \times \frac{1}{4}\right) + \left(\frac{1}{\not{2}} \times \frac{\not{2}}{3} \times \frac{1}{2}\right) \\ &= \frac{3}{20} + \frac{1}{20} + \frac{1}{6} = \frac{\not{4}}{\not{20}^{5}} + \frac{1}{6} = \frac{1}{5} + \frac{1}{6} = \frac{6}{6 \times 5} + \frac{5}{6 \times 5} = \frac{11}{30} \end{aligned}$$

(b) What is the probability that both balls are blue, given that they are the same colour?

Let F be the event that both balls are blue. We can find $Pr[F]$ by adding up the path probabilities for all paths in which two R^c branches occur. We get

$$Pr[F] = \left(\frac{1}{2} \times \frac{2}{5} \times \frac{1}{4}\right) + \left(\frac{1}{2} \times \frac{2}{3} \times \frac{1}{2}\right) = \frac{1}{20} + \frac{1}{6} = \frac{3}{20 \times 3} + \frac{10}{6 \times 10} = \frac{13}{60}$$

However, this is not the probability we were asked to find. We need to find the conditional probability $Pr[F|E]$, which by definition is $Pr[F|E] = \frac{Pr[F \cap E]}{Pr[E]}$. What is $Pr[F \cap E]$? This is the probability that both balls are blue and are also the same colour. But in order for both balls to be blue, they must be the same colour. That is, in this case we have $F \subseteq E$, i.e. F can *only* occur when E occurs, so that $Pr[F \cap E^c] = 0$ and we have $Pr[F \cap E] = Pr[F]$. (That is, $Pr[F] = Pr[F \cap E] + Pr[F \cap E^c]$, so since $Pr[F \cap E^c] = 0$, then $Pr[F] = Pr[F \cap E]$.) Thus we see that

$$Pr[F|E] = \frac{Pr[F \cap E]}{Pr[E]} = \frac{Pr[F]}{Pr[E]} = \frac{\frac{13}{60}}{\frac{11}{30}} = \frac{13}{\not{60}_2} \times \frac{\not{30}}{11} = \frac{13}{22}$$

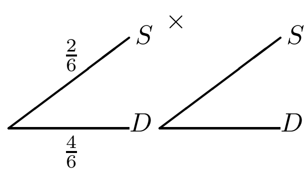
Let's look at a problem involving a different kind of stochastic process, with a different kind of tree structure.

Example 2.31. *A certain professor brings a box of pencils to a multiple choice exam which is going to be marked by computer. However, the professor doesn't realize that the box contains only 6 pencils, 4 of which have not been sharpened and are too dull to use. A student asks for a pencil. The professor draws pencils out of the box at random, without replacement, until a sharp pencil is found. What is the probability that the professor does not find a sharp pencil on the first 3 tries?*

This stochastic process is different from the ones we've looked at so far, because we don't know how many experiments will be performed, i.e. how many pencils will be drawn from the box before a sharp pencil is found. The probability tree modelling this stochastic process has a level for each experiment which *might* be performed, i.e. each time the professor might draw a pencil.

On any particular performance of the experiment, let S be the event that a sharp pencil is drawn, and D be the event that a dull pencil is drawn. Then each level of the tree contains an S branch and a D branch. The professor only draws another pencil from the box if no sharp pencil has been drawn yet. Therefore an S branch always leads to a terminal point, whereas a D branch always leads to another branching. If another pencil is being drawn from the box, all of the sharp pencils must still be in the box, so the number of sharp pencils in the box is the same every time the experiment is performed. However the number of dull pencils in the box changes on each repetition of the experiment.

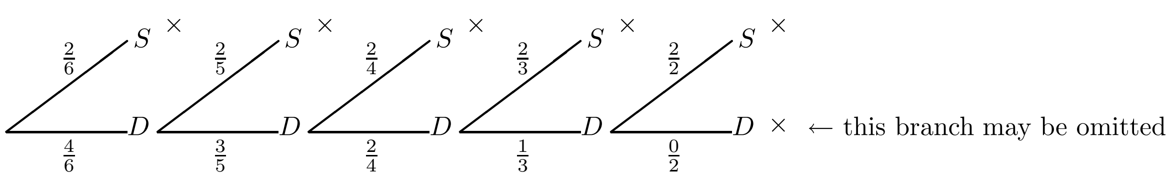
Initially, the box contains 6 pencils: 2 sharp and 4 dull. So the first time the professor draws a pencil from the box, the probability that a sharp pencil is drawn is $\frac{2}{6}$ and the probability that a dull pencil is drawn is $\frac{4}{6}$. If the professor draws a sharp pencil, no further experiments are performed (i.e. this is a terminal point of the tree). On the other hand, if the first pencil is dull, then the professor draws another pencil, i.e. the experiment is repeated.



Partial Tree (× denotes a terminal point)

On the second draw, the box contains only 5 pencils, 2 of which are sharp. So the probability of drawing a sharp pencil this time is $\frac{2}{5}$, and the probability of drawing a dull pencil is $\frac{3}{5}$. Of course, drawing a sharp pencil ends the process, but if a dull pencil is drawn, the experiment continues.

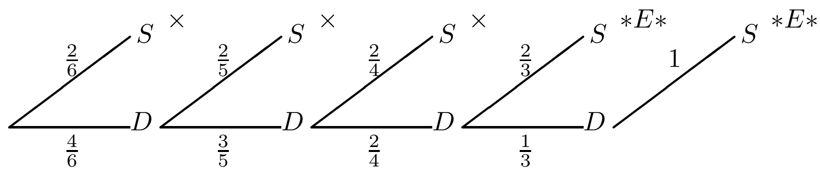
The tree continues in this way until we get to the point at which all of the dull pencils have already been removed from the box, so that a sharp pencil *must* be drawn.



Completed probability tree

Notice that the tree contains 5 levels. In the worst case scenario, the professor draws all 4 dull pencils before finding a sharp pencil, so the experiment could be performed as many as 5 times.

Now that we have the completed tree, we are ready to determine the probability that the professor does not find a sharp pencil in the first 3 tries. Let E be the event that no sharp pencil has been found after 3 draws have been made. Then this event E occurs when it takes either 4 draws or 5 draws to get a sharp pencil.

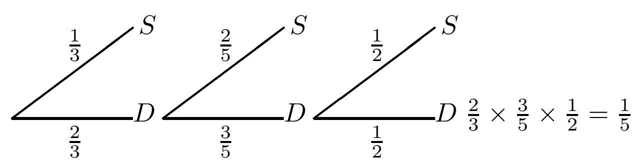


Paths corresponding to event E are shown with $*E*$ at the end

This gives 2 paths through the tree in which event E occurs, so we add up the path probabilities for these 2 paths. We get:

$$Pr[E] = \left(\frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3}\right) + \left(\frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \times \frac{2}{2}\right) = \frac{2}{15} + \frac{1}{15} = \frac{3}{15} = \frac{1}{5}$$

Notice: In this case, rather than modelling the whole stochastic process with the tree, we could have drawn a tree which simply showed the first 3 draws. That is, since we are only interested in whether or not a sharp pencil is found in the first 3 draws, we can consider the stochastic process to involve up to only 3 draws, instead of continuing until a sharp pencil is found. In this case, the tree would end after the third level, and $Pr[E]$ would be found as simply the path probability for the path with 3 D's in it.



Probability tree modelling a process involving at most 3 draws

Of course, this gives the same result as before.

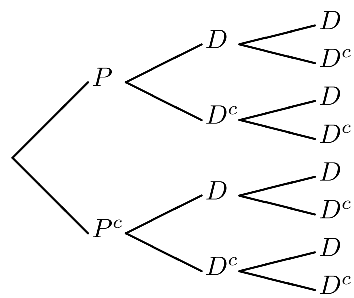
Another common type of stochastic process arises in the next example.

Example 2.32. *A certain machine must be set up every day prior to beginning the production run. The set up is very complex and 10% of the time is done improperly. When the machine is improperly set up, 20% of the parts produced are defective. When the machine is properly set up, only 2% of the parts produced are defective. Two parts have been selected randomly from today's production.*
(a) What is the probability that both parts are satisfactory (i.e. not defective)?
(b) If both parts are found to be satisfactory, what is the probability that the machine was set up improperly today?

Before we try to answer any questions, we should model what's going on here. Every day, the machine is set up. Some days it's done properly, but other days the set up is improper. So on any particular day (like today), there are 2 possible outcomes of the probabilistic experiment 'set up the machine' – proper set up and improper set up. (And presumably we cannot easily tell which has occurred until the machine is used and we see how many defective parts are produced.) We know that an improper set up occurs with probability .1, so a proper set up occurs with probability .9 (i.e. the other 90% of the time).

Once the machine has been set up for the day, it is used to produce 'parts'. The production run presumably involves producing many of these parts. Each part produced on the machine might be defective or (preferably) non-defective, i.e. a satisfactory part. So producing a part, or randomly selecting a part which has been produced by the machine, and determining whether or not it is defective, is another probabilistic experiment.

The stochastic process we need to model involves setting up the machine and then testing 2 randomly chosen parts produced by the machine. Thus we have 3 experiments performed (where the last 2 are actually the same experiment, performed twice). Let P denote a proper set up and D denote a defective part. Then on the first experiment, either P or P^c occurs and on each of the second and third experiments, either D or D^c occurs. Therefore the structure of the probability tree which models this stochastic process is:

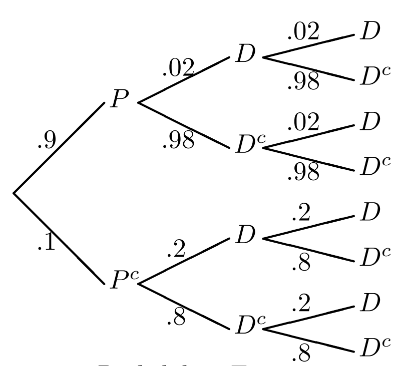


Tree Structure

Now, we need to determine the probabilities which go on the branches of this tree. We have already seen that $Pr[P] = .9$ and $Pr[P^c] = .1$.

We know that when the machine is improperly set up, 20% of the parts produced are defective. That is, given that the set up was improper, each part produced has probability .2 of being defective. Therefore $Pr[D|P^c] = .2$ and this is the probability associated with every D branch which follows the P^c branch. Of course, this also means that each part has probability .8 of being non-defective (when the set up is improper), so $Pr[D^c|P^c] = .8$ is the probability on each D^c branch following the P^c branch.

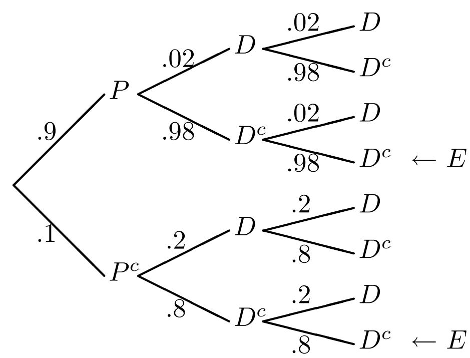
Similarly, the fact that when the set up is proper, 2% of the parts are defective (so that 98% are satisfactory) tells us that $Pr[D|P] = .02$ and $Pr[D^c|P] = .98$, and these probabilities go on all of the D branches, and D^c branches, respectively, which follow the P branch. This gives us the completed probability tree.



Probability Tree

Now, we are ready to think about the questions we were asked.

(a) We need to find the probability that both of the tested parts are found to be non-defective. Let E be the event that both parts are satisfactory. Then event E occurs whenever 2 D^c branches are taken. That is, $Pr[E]$ can be found by summing the path probabilities for all paths which include two D^c branches. There are 2 such paths.



Paths marked E are in event E – both parts are satisfactory

Therefore the probability that both parts are satisfactory is

$$Pr[E] = [(.9)(.98)(.98)] + [(.1)(.8)(.8)] = .86436 + .064 = .92836$$

(*Note:* In a *no calculators* environment, easier numbers would be used.)

(b) In this kind of manufacturing problem, it is often the case that we cannot tell whether the set up was done properly or improperly until we see how many defective parts the machine is producing. However, even proper set ups produce some defective parts, and even when the machine is improperly set up, it is still true that most of the parts produced are satisfactory. So testing a couple of parts does not tell us definitively whether or not the machine was improperly set up. However, by testing 2 parts, we can obtain a better idea of how likely it is that the machine was improperly set up (i.e. better than the initial assessment that there's a 10% chance).

We are asked to determine the probability that the set up was improper if 2 parts have been tested and found to be satisfactory. That is, we need to find $Pr[P^c|E]$. We use the definition:

$$Pr[P^c|E] = \frac{Pr[P^c \cap E]}{Pr[E]}$$

In part (a), we found that $Pr[E] = .92836$, and of course $Pr[P^c \cap E]$ is the path probability for the one path through the tree containing both the P^c branch and 2 D^c branches:

$$Pr[P^c \cap E] = (.1)(.8)(.8) = .064$$

Thus we see that

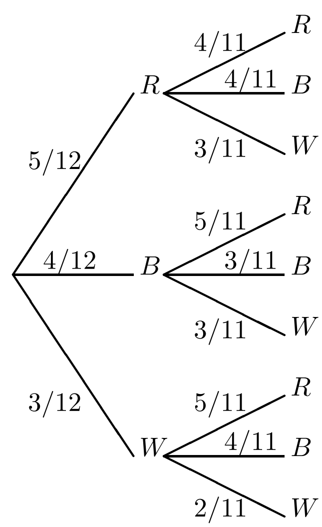
$$Pr[P^c|E] = \frac{Pr[P^c \cap E]}{Pr[E]} = \frac{.064}{.92836} \approx .06894$$

We see that even if 2 randomly chosen parts are found to be satisfactory, there is still about a 6.9% chance that the machine was improperly set up. (But the fact that no defective parts were found has decreased this probability from the original 10% chance.)

Let's look at one last example.

Example 2.33. *An urn contains 5 red balls, 4 black balls and 3 white balls. 2 balls are drawn from the urn without replacement.*
(a) Find the probability that both balls are red, given that the first ball is red.
(b) Find the probability that the first ball is red, if it is known that at least one red ball was drawn.

We can start by modelling the stochastic process with a probability tree. We've drawn similar trees previously. We are drawing 2 balls, so we are performing 2 experiments and the tree has 2 levels. Let R, B and W, respectively, represent drawing a red, black or white ball, when one ball is drawn from the urn. These are the 3 possible outcomes each time the experiment is performed, so each level has these 3 kinds of branches. The probabilities on the branches are found by considering what combination of balls are in the urn at the time of the draw. We get:



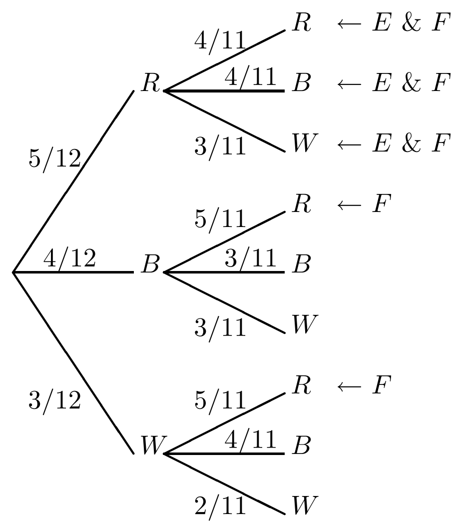
Probability Tree

(a) We are asked to find the probability that both balls are red, given that the first draw produces a red ball. ***Note: we do not need the probability tree to answer this.*** We have already found this probability, and put it on the tree. This is simply the probability that the second ball drawn is red, given that the first ball drawn was red. If the first ball is red, then at the time of the second draw, the urn contains 4 red, 4 black and 3 white balls, so the probability of drawing a red ball is $\frac{4}{11}$. This is the probability that we calculated to put on the second level R branch that follows the first-level R branch.

(b) Now we need to find the probability that the first ball is red, given that at least one red ball is drawn. For this, the probability tree *is* helpful.

Let E be the event that the first ball drawn is red and F be the event that at least 1 of the 2 balls drawn is red. Then we need to find $Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]}$. Event E, of course, simply corresponds to drawing a red ball on the first draw, which happens with probability $\frac{5}{12}$, so $Pr[E] = \frac{5}{12}$. What is the event $E \cap F$? This is the event that the first draw produces a red ball, and at least one of the 2 balls drawn is red. But if the first ball drawn is red, then it is *always* true that at least one red ball is drawn. That is, event F must happen whenever event E happens, so E and F both occur whenever E occurs. Thus $E \cap F = E$, so $Pr[E \cap F] = Pr[E] = \frac{5}{12}$.

Note: We could have added up path probabilities to get this, but we didn't need to. We can mark the paths corresponding to events E and F on the tree. Event E corresponds to the paths in which the first branch is an R branch. Event F corresponds to all paths which contain at least one R branch. We get:



Probability Tree - Events E and F marked

Notice that all the E paths are marked with both E and F (because F must occur whenever E occurs). Also notice that the only paths marked E are (of course) those which start with the first level R branch. We get

$$Pr[E \cap F] = \left(\frac{5}{12} \times \frac{4}{11}\right) + \left(\frac{5}{12} \times \frac{4}{11}\right) + \left(\frac{5}{12} \times \frac{3}{11}\right) = \frac{5}{12}\left(\frac{4}{11} + \frac{4}{11} + \frac{3}{11}\right) = \frac{5}{12} \times 1 = \frac{5}{12}$$

Now, we still need to find $Pr[F]$. We add up path probabilities for all paths marked with F in the tree above. Of course, we've already seen that the first 3 add up to $\frac{5}{12}$, so we just need to add 2 more path probabilities to this. We get:

$$\begin{aligned} Pr[F] &= \frac{5}{12} + \left(\frac{4}{12} \times \frac{5}{11}\right) + \left(\frac{3}{12} \times \frac{5}{11}\right) = \frac{5}{12} + \left(\frac{5}{12} \times \frac{4}{11}\right) + \left(\frac{5}{12} \times \frac{3}{11}\right) \\ &= \frac{5}{12} \times \left(1 + \frac{4}{11} + \frac{3}{11}\right) = \frac{5}{2 \times \cancel{6}} \times \frac{\cancel{18}^{3}}{11} = \frac{15}{22} \end{aligned}$$

Therefore we get

$$Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]} = \frac{\frac{5}{12}}{\frac{15}{22}} = \frac{\cancel{5}}{\cancel{12}_6} \times \frac{\cancel{22}^{11}}{\cancel{15}_3} = \frac{11}{18}$$

Math 1228A/B Online

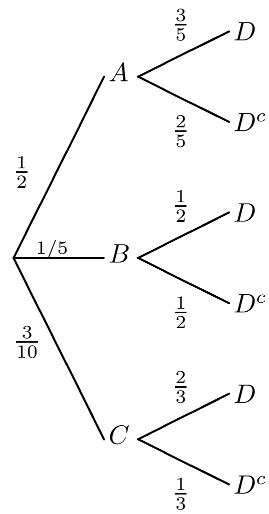
Lecture 18:

There's Nothing Really New Here!

(text reference: Section 2.6)

2.6 Bayes' Theorem

Consider a stochastic process in which there are 2 experiments. In the first experiment, outcomes A, B and C occur with probabilities 1/2, 1/5 and 3/10 respectively. In the second experiment, event D either does or does not occur, with $Pr[D|A] = 3/5$, $Pr[D|B] = 1/2$ and $Pr[D|C] = 2/3$. We can construct a probability tree to model this stochastic process:



Question: If it is known that event D occurred, what is the probability that the outcome of the first experiment was A?

We want to find $Pr[A|D]$. Of course, we know that $Pr[A|D] = \frac{Pr[A \cap D]}{Pr[D]}$. Also, we know that we can find $Pr[A \cap D]$ using

$$Pr[A \cap D] = Pr[D|A] \times Pr[A]$$

(i.e. the path probability for the 'A and then D' path). As well, we can find $Pr[D]$ by considering the partition of D into $D \cap A$, $D \cap B$ and $D \cap C$. And these, of course, are the events at the terminal points of the various paths in which D occurs. That is, we know that we can find $Pr[D]$ by adding up the path probabilities for all paths in which there is a D branch. And those path probabilities are found by multiplying the probabilities along the path.

Looking at what calculations we would need to do here, we have:

$$\begin{aligned} Pr[D \cap A] &= Pr[D|A] \times Pr[A] \\ Pr[D \cap B] &= Pr[D|B] \times Pr[B] \\ Pr[D \cap C] &= Pr[D|C] \times Pr[C] \end{aligned}$$

which gives:

$$Pr[D] = Pr[D|A] \times Pr[A] + Pr[D|B] \times Pr[B] + Pr[D|C] \times Pr[C]$$

Using these formulas, we get:

$$\begin{aligned} Pr[A|D] &= \frac{Pr[A \cap D]}{Pr[D]} \\ &= \frac{Pr[D|A] \times Pr[A]}{Pr[D|A] \times Pr[A] + Pr[D|B] \times Pr[B] + Pr[D|C] \times Pr[C]} \qquad (***) \\ &= \frac{\frac{1}{2} \times \frac{3}{5}}{\left(\frac{1}{2} \times \frac{3}{5}\right) + \left(\frac{1}{5} \times \frac{1}{2}\right) + \left(\frac{3}{10} \times \frac{2}{3}\right)} = \frac{\frac{3}{10}}{\frac{3}{10} + \frac{1}{10} + \frac{2}{10}} = \frac{3/10}{6/10} = \frac{3}{6} = \frac{1}{2} \end{aligned}$$

Notice that:

1. There was absolutely nothing new here. In fact, we did a calculation just like this in part (c) of Example 2.29 (see Lecture 16, pg. 90), and a quite similar calculation in part (b) of Example 2.32 (see Lecture 17, pg. 95).

2. We got an ugly looking formula (***), which really just expressed how to find $Pr[A|D]$ using the probabilities on the branches of the tree.

The ugly formula is called *Bayes' Formula*. It is given in the following theorem, which tells how to find the probability that a particular outcome was observed on an *earlier* experiment, when it is know what outcome occurred on a *later* experiment.

Bayes' Theorem: Let E_1, E_2, ..., E_n be n mutually exclusive events which together comprise the whole sample space for a stochastic process. Let A be any other event. Then for any value i with $1 \leq i \leq n$,

$$Pr[E_i|A] = \frac{Pr[A|E_i] \times Pr[E_i]}{Pr[A|E_1] \times Pr[E_1] + Pr[A|E_2] \times Pr[E_2] + ... + Pr[A|E_n] \times Pr[E_n]}$$

Translation: If a stochastic process involves 2 experiments, where the possible outcomes on the first experiment are E_1, E_2, ..., E_n, and A is some event which may occur on the second experiment, then for any outcome E_i on the first experiment, the probability that this outcome occurred, given the information that event A occurred on the second experiment, is given by

$$Pr[E_i|A] = \frac{\text{path prob. of the path containing both } E_i \text{ and } A}{\text{sum of path prob.'s of all paths in which } A \text{ occurs}}$$

Remember: The only thing new here is the name of the formula. All we have done is to pull together some ideas which we already knew, and which, in fact, we had previously put together, when we did Example 2.29(c). There, we didn't write it out as a single formula. Instead, we had previously calculated the numerator probability ($Pr[A \cap R^c]$), and the denominator probability ($Pr[R^c]$), so we put them directly into the formula for $Pr[A|R^c]$. However, the calculations we did there were exactly those we would have done if we had used Bayes' Theorem.

Let's look at a couple of examples in which we use Bayes' Theorem directly.

Example 2.34. *In a certain town, 1% of all children have a particular disease. There is a test which can be used to detect this disease. When a child who has the disease is given the test, a positive test result occurs 95% of the time. When a child who does not have the disease is given the test, a positive test result occurs 2% of the time. One of the children from the town is tested and a positive test result occurs. What is the probability that the child has the disease?*

We have 2 different things happening here. There is a disease which some children have, and there is a test that a child can be given, which is supposed to determine whether or not the child has the disease. However the test is not 100% reliable. Some children who have the disease do not get a positive test result (only 95% show positive). Also, some children who do not have the disease *do* get a positive test result (2% of them). (This is not unusual. Many medical and other types of tests are not quite 100% reliable.)

Let D be the event that a child has the disease and let P be the event that a child has a positive result when tested.

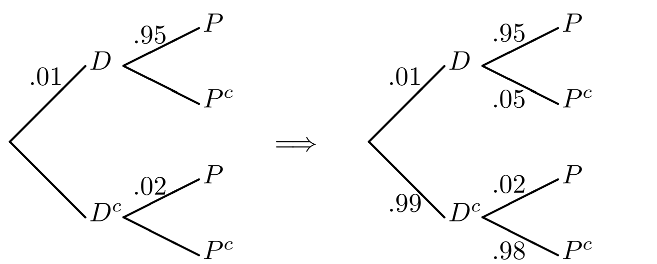
We are told that:

1. 1% of all children have the disease
   i.e. if a child is chosen at random, there is a 1% chance that the chosen child has the disease, so $Pr[D] = .01$
2. When a child with the disease is tested, there is a 95% chance of a positive test result
   i.e. $Pr[P|D] = .95$
3. When a child without the disease is tested, there is a 2% chance of a positive test result (called a *false positive*)
   i.e. $Pr[P|D^c] = .02$

Consider the stochastic process:

1. Pick a child at random and determine whether the child has the disease. (i.e. D or D^c)
2. Give that child the test and see whether a positive test result occurs. (i.e. P or P^c)

We can draw the probability tree for this stochastic process. We fill in the other probabilities using the fact that the probabilities on branches growing from the same point must sum to 1. We get:



Tree structure *Completed Probability Tree*

Notice: In reality, when we perform the experiments, we have some child and we don't know whether or not the child has the disease. That's why we give him or her the test. That's also why we need to know the probability that a child with a positive test result has the disease, i.e. $Pr[D|P]$, which is what we've been asked to find. The tree which we've drawn, and which corresponds to the information available, is actually backwards to the tree which models 'test the child' and later 'find out whether the child actually had the disease'. This is a fairly typical example of a situation requiring Bayes' Theorem.

Using Bayes' formula, we get

$$Pr[D|P] = \frac{\text{path prob. for path with } D \text{ and } P}{\text{sum of path prob.'s for all paths with } P \text{ in them}}$$

$$= \frac{(.01)(.95)}{(.01)(.95) + (.99)(.02)} = \frac{.0095}{.0095 + .0198} = \frac{.0095}{.0293} = \frac{95}{293} \approx .3242$$

We see that even though the test is quite accurate, there's less than a 1/3 chance that a child with a positive test result actually has the disease.

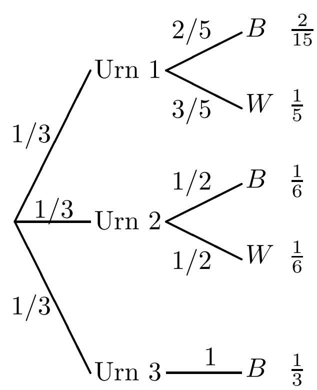
Note: This happens because although the incidence of "false positive" is low, it is also true that the incidence of the disease is low, so the population of disease-free children is large. For instance, suppose that there are 10,000 children in the town. Then about 100 of them have the disease, of whom about 95 will test positive. But of the other roughly 9,900 disease-free children, about 2%, or 198 of them, will falsely test positive. Therefore we get more than twice as many false positives as true positives.

Example 2.35. *There are 3 identical-looking urns. One contains 20 black balls and 30 white balls. The second contains 20 black balls and 20 white balls. The third urn contains 20 black balls. An urn is chosen at random, and a ball is drawn from that urn.*
(a) What is the probability that the third urn was chosen, if the ball drawn is black?
(b) What is the probability that the second urn was chosen, if the ball drawn is black?
(c) What is the probability that the third urn was chosen, if the ball drawn is white?

We start by drawing the probability tree for this stochastic process. The stochastic process involves 2 experiments: choose an urn, and draw a ball from the chosen urn. Since the urn is chosen at random, each of the 3 is equally likely to be chosen, so the probability of choosing any particular urn is $\frac{1}{3}$. When a ball is drawn, the outcome is either a black ball (B) or a white ball (W). The probability of each outcome depends, of course, on which urn the ball is drawn from.

Notice that since the third urn contains only black balls, there is only one possible outcome when that urn is chosen. (That is, drawing a white ball is impossible and will occur with probability 0, so we don't need to show a branch for this outcome.)

Since we're asked to calculate several things, we might as well calculate all of the path probabilities while preparing the tree. As usual, we can show them at the ends of the paths.



Probability Tree for Example 2.35

We're ready now to tackle the various questions we were asked. Notice that in each case we're asked to find the probability that a particular urn was chosen, given that a particular colour of ball was drawn. That is, we're being asked the probability of a particular outcome on the first experiment, given a particular observation on the second experiment. These are the kinds of probabilities that Bayes' formula calculates.

(a) What is the probability that the third urn was chosen, if the ball drawn is black?

We need to find $Pr[\text{Urn } 3|B]$. We have:

$$\begin{aligned} Pr[\text{Urn } 3|B] &= \frac{\text{path prob. for (Urn 3} \cap B\text{) path}}{\text{sum of path prob.'s for all paths with } B} \\ &= \frac{\frac{1}{3}}{\frac{2}{15}+\frac{1}{6}+\frac{1}{3}} = \frac{\frac{1}{3}}{\frac{2}{15}+\frac{1}{2}} = \frac{\frac{1}{3}}{\frac{4}{30}+\frac{15}{30}} = \frac{1/3}{19/30} = \frac{1}{\cancel{3}} \times \frac{\cancel{30}^{10}}{19} = \frac{10}{19} \end{aligned}$$

(b) What is the probability that the second urn was chosen, if the ball drawn is black?

This time, we need to find $Pr[\text{Urn } 2|B]$. Again, we use Bayes' Theorem:

$$Pr[\text{Urn } 2|B] = \frac{\text{path prob. for (Urn } 2 \cap B\text{) path}}{\text{sum of path prob.'s for all paths with } B}$$

$$= \frac{\frac{1}{6}}{\frac{2}{15}+\frac{1}{6}+\frac{1}{3}} = \frac{1/\cancel{6}}{19/\cancel{30}_5} = \frac{5}{19}$$

(c) What is the probability that the third urn was chosen, if the ball drawn is white?

This looks like the same sort of thing, so we probably need to use Bayes' formula again. Attempting to use Bayes' Theorem, we get:

$$Pr[\text{Urn } 3|W] = \frac{\text{path prob. for (Urn } 3 \cap W\text{) path}}{\text{sum of path prob.'s for all paths with } W}$$

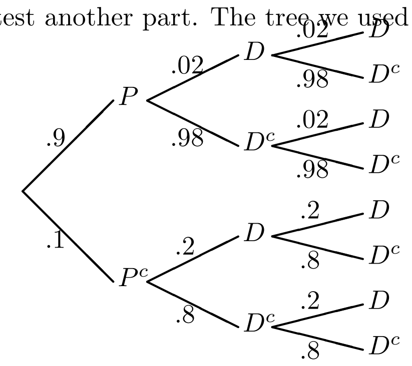
$$= \;???$$

Oops. There's no Urn 3 $\cap W$ path. That's because this is an impossible event. Urn 3 contains no white balls, so we *can't* draw a white ball from Urn 3. That is, given that we have drawn a white ball, we know that we cannot possibly be drawing from Urn 3, so $\Pr[\text{Urn } 3|W] = 0$. (*Recall:* the reason we don't have an Urn 3 $\cap W$ path is because $Pr[W|\text{Urn } 3] = 0$, so the path probability for this path is 0, which is why we omitted it.)

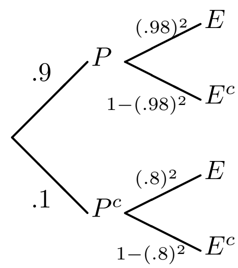
Example 2.32 Revisited

Earlier in this lecture, it was asserted that what we did in part (b) of Example 2.32 (see Lecture 17, pg. 95) was "quite similar" to using Bayes' formula. Let's have another look at that. The problem said *"A certain machine must be set up every day prior to beginning the production run. The set up is very complex and 10% of the time is done improperly. When the machine is improperly set up, 20% of the parts produced are defective. When the machine is properly set up, only 2% of the parts produced are defective. Two parts have been selected randomly from today's production. If both parts are found to be satisfactory (i.e. not defective), what is the probability that the machine was set up improperly today?"*

When we looked at this originally, we modelled the stochastic process involving 3 experiments: set up the machine; test a part; test another part. The tree we used looked like this:



However, we *could* consider the stochastic process to involve only 2 experiments: set up the machine; test 2 parts. Also, since the only event we're interested in on the second experiment is 'both parts are satisfactory', we could define this to be event E and consider only whether or not E occurs. From the tree above, we can easily see that $Pr[E|P] = (.98)^2$ and $Pr[E|P^c] = (.8)^2$. That is, if we collapse the second and third level branchings into a single level of branching, and collapse all but the D^c then D^c branch into a single branch, we have a branch for E, with probability equal to the product of the probabilities on the two D^c branches, and a branch for E^c. Modelling the problem in this way, the tree is:



Now, using Bayes' formula, we see that

$$Pr[P^c|E] = \frac{\text{path prob. for path with both } P^c \text{ and } E}{\text{sum of path prob.'s for all paths with } E} = \frac{(.1)(.8)^2}{(.9)(.98)^2 + (.1)(.8)^2}$$

which is the same calculation we ended up doing before.

Math 1228A/B Online

Lecture 19:
Bernoulli Trials

(text reference: Section 2.7, pages 102 - 106)

2.7 Independent Trials

Consider the following problem:

Example 2.36. *Three standard decks of cards are shuffled (separately). A single card is drawn from each deck. What is the probability that exactly 2 of the drawn cards are Hearts?*

We can solve this problem by analyzing the stochastic process:

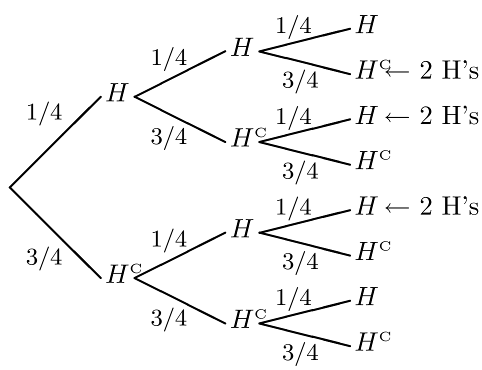
Experiment 1: Draw a card at random from deck 1.

Experiment 2: Draw a card at random from deck 2.

Experiment 3: Draw a card at random from deck 3.

All we are interested in here is whether or not each of the drawn cards is a Heart. For the experiment 'draw a single card from a deck', let H be the event that a Heart is drawn. Then for each deck, we have $Pr[H] = \frac{1}{4}$. That is, no matter which deck it is, and no matter what we may have drawn from other decks, each time we draw we are drawing from a full deck, so the probability of drawing a Heart is always $\frac{1}{4}$. Of course, this means that (for each deck) the probability of not drawing a Heart is $Pr[H^c] = \frac{3}{4}$.

We can draw the probability tree which models this stochastic process.



Notice that each pair of branches looks exactly the same: an H branch with $Pr[H] = \frac{1}{4}$ and an H^c branch with $Pr[H^c] = \frac{3}{4}$. Thus the path probabilities are all products of 3 numbers, each of which is either $\frac{1}{4}$ or $\frac{3}{4}$.

To find the probability that exactly 2 of the drawn cards are hearts, we add up the path probabilities for all paths in which there are exactly 2 H branches. Letting E be the event that exactly 2 Hearts are drawn, we get:

$$\begin{aligned} Pr[E] &= \left(\frac{1}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{1}{4}\right) \\ &= 3 \times \left[\left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)\right] = 3 \times \left(\frac{3}{64}\right) = \frac{9}{64} \end{aligned}$$

Of course, since we are adding up path probabilities for paths which contain exactly 2 H branches, and therefore also 1 H^c branch, and since $Pr[H]$ is the same for each experiment, then all of these paths have the same path probability: the product of two $\frac{1}{4}$'s and one $\frac{3}{4}$, i.e. $\left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^1$.

How many of these paths were there? There was the HHH^c path, the HH^cH path and the H^cHH path. But this is just all of the permutations of 2 H's and 1 H^c. There are $\binom{3}{2}$ ways to decide which 2 of the 3 positions contain H's, and then the one remaining position must contain the H^c. That is, there are $\binom{3}{2}$ ways to "choose" 2 of the 3 decks to draw a Heart from, and then a non-Heart would have to be drawn from the other deck.

Looked at this way, we see that what we found was:

$$Pr[E] = \binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1$$

What were the characteristics of this problem which made the answer have this form? We are performing the *same* experiment several times. And every time we perform the experiment, we are interested in the same event. Also, the result on one performance of the experiment was *independent* of the results observed on previous experiments, so that the probability of the event we are interested in is the *same* on each repetition of the experiment.

We can see that any time we look at a problem with these characteristics, we'll get the same kind of thing happening, i.e. the answer will have the same basic form. We have a special name for this kind of stochastic process. We say that we are performing a number of **Bernoulli Trials**.

Definition: A stochastic process consists of performing n **Bernoulli trials** whenever:

1. the same experiment is performed n times (i.e. we perform n *trials* of the experiment),
2. each time, we observe whether or not some specific event E occurs (i.e. whether E or E^c occurs), and
3. the trials are *independent*, so that $p = Pr[E]$ is the same on each trial (i.e. the outcome of a trial is not affected by the outcomes of the previous trials).

The event of interest, E, is referred to as a **success**. The probability $p = Pr[E]$ is referred to as the **probability of success**. Similarly, the **probability of failure** is $q = 1 - p$.

Suppose that, as in our example, we want to know the probability of observing exactly k successes during a set of n Bernoulli trials. Then if we were to construct a probability tree to model the n trials, we would find the answer by adding up the path probabilities for all paths in which there are exactly k E branches, and therefore also $(n-k)$ E^c branches. Since each E branch has probability $Pr[E] = p$ and each E^c branch has probability $Pr[E^c] = 1 - p = q$, then for each of these paths, the path probability is going to be the product of k copies of p and $(n-k)$ copies of q, i.e. each of the paths we're interested in has path probability $p^k q^{n-k}$. How many of these paths are there? One for each different permutation of k E's and $(n-k)$ E^c's. That is, there is one such path for each way in which we can choose k of the trials to be successes so that the other $n-k$ trials are failures. So there are $\binom{n}{k}$ of these paths.

But if we know all that without drawing the tree, then there's no need to actually draw it. We have a simple formula which gives the probability of observing exactly k successes on n trials. We have:

Theorem: The probability of observing exactly k successes in n Bernoulli trials in which the probability of success is p is given by:

$$\binom{n}{k} p^k q^{n-k}$$

Let's look at some examples of using this formula.

Example 2.37. *A single die is tossed 4 times.*
(a) What is the probability that no 6 is tossed?
(b) What is the probability that exactly one 6 is tossed?
(c) What is the probability that at most one 6 is tossed?
(d) What is the probability that more than one 6 is tossed?
(e) What is the probability that at least one 6 is tossed?

Here, we are performing the experiment "toss a die" four times. Each time we toss the die, we are interested in whether or not the event "a six is tossed" occurs. Of course, each time a die is tossed, the probability of getting a 6 is the same, $\frac{1}{6}$. So we have $n = 4$ Bernoulli trials. Defining success to be that a 6 is tossed, we have the probability of success being $p = \frac{1}{6}$ on each trial, which means that the probability of failure is $q = \frac{5}{6}$.

(a) We are asked to find the probability that no 6 is tossed, i.e. the probability of observing $k = 0$ successes in the $n = 4$ trials. Using the formula, we see that the probability of tossing no 6's is

$$Pr[\text{0 successes}] = \binom{n}{k} p^k q^{n-k} = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{4-0} = 1 \times 1 \times \left(\frac{5}{6}\right)^4 = \frac{5^4}{6^4} = \frac{625}{1296}$$

(b) This time, we want the probability of tossing exactly one 6. This corresponds to having exactly $k = 1$ success. Using the formula again, the probability that this occurs is

$$Pr[\text{1 success}] = \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{4-1} = 4 \times \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)^3 = 4 \times \frac{1}{6} \times \frac{5^3}{6^3} = 4 \times \frac{5^3}{6^4} = \frac{4 \times 125}{1296} = \frac{500}{1296}$$

(c) Now, we need to find the probability of having *at most* one success during the four trials. Of course, having at most one success means either having no successes or having exactly one success. We have already found the probabilities of both of these events. We simply need to add them together. So the probability of getting at most one 6 is

$$Pr[\text{at most 1 success}] = Pr[\text{0 successes}] + Pr[\text{1 success}] = \frac{625}{1296} + \frac{500}{1296} = \frac{1125}{1296}$$

(d) We need to calculate the probability that more than one 6 is tossed, i.e. the probability of observing more than one success in the n trials. Since we are performing 4 trials, more than one success means 2 or 3 or 4 successes. We could calculate the probabilities of each of these and add them up, but there's an easier way to do this. Observing more than one success is the *complement* of observing at most one success. That is, we observe more than one successs whenever we *don't* observe 0 or 1 successes. Therefore we can find this using our answer from (c).

$$\begin{aligned} Pr[\text{more than one success}] &= Pr[(\text{at most 1 success})^c] \\ &= 1 - Pr[\text{at most 1 success}] \\ &= 1 - \frac{1125}{1296} = \frac{1296 - 1125}{1296} = \frac{171}{1296} \end{aligned}$$

(e) Now, we need the probability of tossing at least one 6. Getting at least one 6 in four tosses means getting exactly 1 or exactly 2 or exactly 3 or exactly 4 sixes. But it's easier to realize that getting at least one 6 is the complement of getting no 6's. And we have seen that the probability of getting no 6's is $\frac{625}{1296}$. Therefore, the probability that at least one 6 is tossed is

$$\begin{aligned} Pr[\text{at least one success}] &= 1 - Pr[\text{no successes}] \\ &= 1 - \frac{625}{1296} = \frac{671}{1296} \end{aligned}$$

Another Approach: Use the answers to (b) and (d)
We can also find the probability of tossing at least one 6 as follows:

$$\begin{aligned} Pr[\text{at least one success}] &= Pr[\text{exactly one success}] + Pr[\text{more than one success}] \\ &= \frac{500}{1296} + \frac{171}{1296} = \frac{671}{1296} \end{aligned}$$

Notice: It's important that you realize the meanings of and the differences between the expressions *more than* and *at least*, and also between *fewer than* and *at most*. As well, it's often useful to think about whether it's easier to calculate the required probability directly or to find it by calculating the probability of the complementary event. (Especially, but not only, when some calculations have already been done.)

Often, we need to *decide* how many trials should be performed in order to have the probability of observing some outcome at least once have at least some given value. That is, we often want to answer a question of the form 'how many trials should be performed in order to have at least an x% chance of observing at least one success?'. Consider, for instance, the next example.

Example 2.38. *As an advertising promotion, a cereal company has put a \$2 coin into "1 out of every 10" boxes of cereal. How many boxes of cereal would you need to buy in order to be sure that you have more than a 1/3 chance of getting at least 1 coin?*

What do we need to do here? Well, before we think about that, let's just think about what's being described to us. You are going to buy some boxes of cereal, and then see whether or not you got any \$2 coins. That is, for every box of cereal you buy, you want to see whether or not the event C: *got a \$2 coin* occurs. As long as the coins were distributed among the boxes at random at the factory, and/or the boxes you buy are chosen at random, then the boxes are independent of one another and each has a 1 in 10 chance of containing a coin. So buying n boxes of cereal corresponds to performing n Bernoulli trials in which success is 'got a coin', with $p = \frac{1}{10} = .1$.

Now, how many successes do we want to observe? At least one. We want to consider 'the chance of getting at least one coin', which means the probability of getting at least one success. Of course, this is most easily expressed as the complement of having no successes. So if we perform n trials, the probability of getting at least one coin will be:

$$Pr[\text{at least 1 success}] = 1 - Pr[0 \text{ successes}] = 1 - \left[\binom{n}{0}(.1)^0(.9)^n\right] = 1 - [1 \times 1 \times (.9)^n] = 1 - (.9)^n$$

Now, what is it that we want to be true here? We want your chance of getting at least one coin to be more than 1/3. That is, we need the value of this probability to be bigger than 1/3:

$$Pr[\text{at least 1 success}] > \frac{1}{3} \quad \Rightarrow \quad 1 - (.9)^n > \frac{1}{3}$$

We can rearrange this to get

$$1 - \frac{1}{3} > (.9)^n \quad \Rightarrow \quad \frac{2}{3} > (.9)^n \quad \Rightarrow \quad (.9)^n < \frac{2}{3}$$

We were asked to determine how many boxes of cereal you should buy, in order for this to be true. That is, we want to find how many trials should be performed, i.e. how big does n need to be, in order for the relationship above to be true. There are 2 ways we can figure this out:
(1) If you're using a calculator and you know how to use logarithms, you get

$$n \log(.9) < \log(2/3) \qquad \Rightarrow \qquad n > \frac{\log(2/3)}{\log(.9)} \approx 3.8484$$

Since n must be bigger than 3.8484, and of course n must be an integer, we see that n must be at least 4. (Don't forget that the logarithm of a number smaller than 1 is negative, so when we divide through by log(.9), the inequality changes direction.)

But if you don't know how logarithms work, or aren't using a calculator, that's okay because there's another approach:
(2) Trial and Error. We need to find n such that $(.9)^n < \frac{2}{3}$, i.e. we need to find the smallest value of n for which this is true. Let's see, obviously $(.9)^1 > \frac{2}{3}$, and for $n = 2$ we get $(.9)^2 = (.9)(.9) = .81$, so that's still bigger than $\frac{2}{3}$. For $n = 3$ we have $(.9)^3 = (.81)(.9) = .729$, so we're not down to $\frac{2}{3}$ (which is about .6667) yet. For $n = 4$ we get $(.9)^4 = (.729)(.9) = .6561$ – **aha!** We see that $(.9)^4 < \frac{2}{3}$, so the the smallest value of n for which $(.9)^n < \frac{2}{3}$ is $n = 4$. You would need to buy at least 4 boxes of cereal in order to be sure that you have more than a 1/3 chance of getting at least one $2 coin.

Notice: In some of the homework, you may need to use a calculator. If you don't know how to use logarithms, use trial and error like this – but you don't need to try **every** value along the way. If q^2 isn't very close, jump up to, say, q^5, and then maybe jump to q^{10}. It's only once you get close that you need to calculate each successive value of n – except, of course, when you're doing the calculations by hand (and not always then).

Math 1228A/B Online

Lecture 20:

Independent Trials with

More than Two Possible Outcomes

(text reference: Section 2.7, pages 107 - 108)

Sometimes, there's more than one event that we're interested in counting the occurrences of. For instance, we might want to know the probability that some event E occurred exactly once and some other event F occurred exactly twice in 5 independent trials, so that there must have been 2 trials on which neither E nor F occurred. To see how we can find a probability like this, consider the next example.

Example 2.39. *Recall Example 2.36. What is the probability that exactly 1 Heart and exactly 1 black card are drawn?*

Example 2.36 (see lecture 19, pg. 106) was the one in which we are drawing one card from each of 3 different decks. If we draw the 3 cards and get exactly one Heart and exactly 1 black card, then we must also draw exactly one card which is neither a Heart nor a black card, i.e. is a Diamond. Let H be the event that a Heart is drawn, B be the event that a black card is drawn, and D be the event that a Diamond is drawn, when 1 card is drawn from a deck. Each time we draw a card from a full deck, we have $Pr[H] = \frac{1}{4}$, $Pr[B] = \frac{1}{2}$ and $Pr[D] = \frac{1}{4}$. These probabilities are the same on each repetition of the experiment, since we are drawing from a full deck each time.

We could draw a probability tree to help us solve this problem. (Feel free to do so.) It would start with a level for the first deck, with branches for events H, B and D. Then, there would be a second level, corresponding to the second deck, on which each of the first level branchs sprouts 3 new branches, again for events H, B and D. Finally, on the third level of the tree, which corresponds to the third deck of cards, each of the 9 second level branches has 3 more branches growing from the end of it, once again H, B and D branches. (There are 27 terminal points / paths in this tree. It would be obnoxious to draw.) Throughout the tree, every H branch has probability $\frac{1}{4}$, every B branch has probability $\frac{1}{2}$, and every D branch has probability $\frac{1}{4}$.

Every path through this tree consists of 3 branches. We are interested in the ones which correspond to 1 Heart and 1 black card, and therefore also 1 Diamond, being drawn. That is, all of the paths corresponding to the event of interest have one H branch, one B branch and one D branch in them. Since these events occur with the same probability each time, the path probability of each path like this is $\frac{1}{4} \times \frac{1}{2} \times \frac{1}{4} = \frac{1}{32}$.

How many of these paths are there in the tree? One for each possible arrangement of 1 H, 1 B and 1 D. Of course, there are $\binom{3}{1\ 1\ 1} = \frac{3!}{1!\,1!\,1!} = 6$ different ways of labelling the 3 positions in the path with one H, one B and one D. So there are 6 paths through the tree with exactly one H branch and one B branch, and each of these paths has path probability $\frac{1}{32}$. Therefore, the probability that we draw exactly one Heart and exactly one black card is $\not{6}^3 \times \frac{1}{\not{32}_{16}} = \frac{3}{16}$. (*Notice:* we didn't need to actually look at the tree to get this.)

We can generalize what we just did, to get a formula which can be applied whenever we have this kind of situation, i.e. a series of independent trials with more than one event of interest.

Theorem:
Suppose we perform n independent trials in which, on each trial, exactly 1 of events E_1, E_2, ... E_t will occur, where the probability that event E_i occurs, p_i, is the same on each trial, for $i = 1, ..., t$ with

$$p_i + p_2 + ... + p_t = 1$$

Then the probability that event E_1 occurs exactly n_1 times, event E_2 occurs exactly n_2 times, etc., and E_t occurs exactly n_t times, with $n_1 + n_2 + ... + n_t = n$, is given by

$$\binom{n}{n_1\ n_2\ ...\ n_t}(p_1)^{n_1}(p_2)^{n_2}...(p_t)^{n_t}$$

In the example we did, we were performing $n = 3$ trials, on each of which there were $t = 3$ different possible outcomes or events of interest. We had $p_1 = \frac{1}{4}$, $p_2 = \frac{1}{2}$ and $p_3 = \frac{1}{4}$. Also, we had $n_1 = n_2 = n_3 = 1$, and we saw that the probability was $\binom{3}{1\ 1\ 1}\left(\frac{1}{4}\right)^1\left(\frac{1}{2}\right)^1\left(\frac{1}{4}\right)^1$, just as the formula tells us. Let's look at another example and work directly with the formula.

Notice: The Bernoulli formula is just a special case of this formula, where $t = 2$ because we just have the 2 events E and E^c (which could have been called E_1 and E_2), occurring with probabilities p and q (which could have been called q_1 and q_2). If we call let $n_1 = k$ and $n_2 = n - k$, and express $\binom{n}{k} = \binom{n}{k\ (n-k)}$ as $\binom{n}{n_1\ n_2}$, then the 2 formulas are identical.

Example 2.40. *An urn contains 50 black balls, 20 red balls and 30 white balls. Five balls are drawn from the urn* **with replacement**. *What is the probability that exactly 2 black balls and 1 red ball are drawn?*

Notice: If we draw repeatedly, *without* replacement, then the probabilities change on each draw, because some objects have been removed. In that situation, we are **not** performing independent trials. However, if we draw *with* replacement, then we put the object we have drawn back, before making the next draw, and so the probabilities are the same on each draw. For instance, in this case, because we replace the drawn ball before drawing again, the urn contains 50 black balls, 20 red balls and 30 white balls *every time* we draw a ball. Whenever we perform several draws *with replacement*, the probabilities are the same on each draw, so we are performing independent trials.

In this case, we're drawing $n = 5$ balls. On each draw, we have $t = 3$ possible outcomes: B: a black ball is drawn, R: a red ball is drawn, and W: a white ball is drawn, which occur with probabilities $Pr[B] = \frac{50}{100} = \frac{1}{2}$, $Pr[R] = \frac{20}{100} = \frac{1}{5}$ and $Pr[W] = \frac{30}{100} = \frac{3}{10}$.

We want to know the probability that exactly 2 black balls and 1 red ball are drawn when these 5 independent trials are performed. Of course since we drew 5 balls, if *exactly* 2 black balls and 1 red ball were drawn, there must also have been 2 white balls drawn. We see that we're looking for the probability of getting exactly $n_B = 2$ black, $n_R = 1$ red and $n_W = 2$ white balls, where $p_B = \frac{1}{2}$, $p_R = \frac{1}{5}$ and $p_W = .3$. Using the formula, this probability is given by:

$$\begin{aligned}\binom{5}{2\ 1\ 2}\left(\frac{1}{2}\right)^2\left(\frac{1}{5}\right)^1\left(\frac{3}{10}\right)^2 &= \frac{5!}{2!\ 1!\ 2!} \times \frac{1}{2^2} \times \frac{1}{5} \times \frac{3^2}{10^2} \\ &= \frac{5 \times 4 \times 3}{2} \times \frac{1}{4} \times \frac{1}{5} \times \frac{9}{100} = \frac{27}{200} = .135\end{aligned}$$

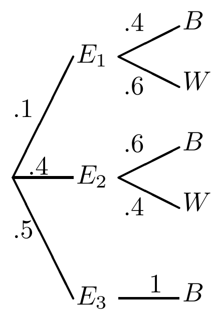
Let's finish off with one last example.

Example 2.41. *There are many identical-looking urns in a room. Ten percent of the urns are "type 1" urns, which contain 20 black balls and 30 white balls. Forty percent of the urns are "type 2" urns, containing 30 black balls and 20 white balls. The remaining 50% of the urns are "type 3" urns containing 50 black balls.*
(a) An urn is chosen at random and one ball is drawn from that urn. What is the probability that the chosen urn is a "type 3" urn, if the ball drawn is black?
(b) An urn is chosen at random, and 3 balls are drawn from that urn, with replacement. What is the probability that the chosen urn is a "type 3" urn if all 3 draws produced black balls?

(a) In this first part, we are performing a fairly simple stochastic process consisting of 2 experiments:
1: Choose one of the urns at random.
2: Draw 1 ball from the chosen urn.

Let E_1, E_2 and E_3 be the events that a type 1 urn, a type 2 urn or a type 3 urn is chosen, respectively. We don't know how many urns there are, but we do know the percentage of the total which are of each of the various types. Of course, if $k\%$ of the urns are of type i, then when we choose one urn at random, there is a $k\%$ chance of choosing a type i urn, so $Pr[E_i] = \frac{k}{100}$. That is, we see that $Pr[E_1] = .1$, $Pr[E_2] = .4$ and $Pr[E_3] = .5$.

Also, let B be the event that a black ball is drawn and W be the event that a white ball is drawn. We know how many balls of each colour are in each type of urn, so we can calculate $Pr[B|E_i]$ and $Pr[W|E_i]$ for i =1, 2 and 3. These probabilities are shown on the probability tree modelling the stochastic process. Notice that since type 3 urns contain only black balls, then B is the only possible outcome if a type 3 urn is chosen.



We need to find $Pr[E_3|B]$. This is a "probability of earlier outcome given later outcome" situation, so we use Bayes' Theorem.

$$\begin{aligned} Pr[E_3|B] &= \frac{\text{path prob. for } (E_3 \cap B) \text{ path}}{\text{sum of path prob.'s for all paths with } B} \\ &= \frac{.5 \times 1}{(.1 \times .4) + (.4 \times .6) + (.5 \times 1)} = \frac{.5}{.04 + .24 + .5} = \frac{.5}{.78} \approx 0.6410 \end{aligned}$$

We see that there is about a .64 probability that we are drawing from a type 3 urn if we draw a black ball.

(b) Now, the stochastic process involves drawing 3 balls, with replacement. We could consider this as a 4-experiment stochastic process, where the first experiment is "choose an urn" and the last 3 experiments are all "draw a ball from the urn" (where of course we must put the ball back into the urn after each experiment). The tree modelling this stochastic process would be quite large, though. And because we are drawing *with replacement*, the last 3 experiments are all identical, so the tree will look quite repetitive.

Instead, we can consider this as a 2-experiment stochastic process, where the first experiment is still "choose an urn at random", but now the second experiment is:

> Perform 3 Bernoulli trials of 'draw 1 ball from the chosen urn', where success is defined to be that a black ball is drawn.

In this second experiment, the only event we are interested in is the event that 3 black balls are drawn, i.e. that 3 successes are observed on the 3 Bernoulli trials. Let's call this event F. We can model this stochastic process with a probability tree which is no more complicated than the one we used in part (a). Of course, if event E_3 occurs, i.e. if an urn of type 3 is chosen, then (since the urn contains only black balls) event F will occur with certainty, so we can omit the F^c branch that would normally follow the E_3 branch.

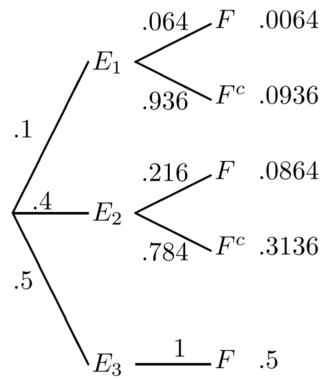
The tree structure is actually just the same as before, but on the second level the branchings are for whether or not event F occurs.

We have events E_1, E_2 and E_3 each occurring with the same probability as before (since we have the identical first experiment). That is, we still have $Pr[E_1] = .1$, $Pr[E_2] = .4$ and $Pr[E_3] = .5$. To find the other probabilities, we simply need to find, in each case, the probability of observing $k = 3$ successes in $n = 3$ Bernoulli trials, where the probability of success, p, depends on which type of urn we're drawing from. If p_i is the probability of drawing a black ball from a type i urn, we have

$$Pr[F|E_i] = \binom{3}{3}(p_i)^3(1-p_i)^0 = 1 \times (p_i)^3 \times 1 = (p_i)^3$$

(Remember, we are drawing *with* replacement, so after a ball is drawn, it is put back in the urn before another ball is drawn, and the probability of drawing a black ball is the same on each draw.)

When a type 1 urn has been chosen, the probability of drawing a black ball is $p_1 = .4$ so we have $Pr[F|E_1] = (.4)^3 = .064$. And this means that $Pr[F^c|E_1] = 1 - .064 = .936$. Similarly, for a type 2 urn, we have $p_2 = .6$ so $Pr[F|E_2] = (.6)^3 = .216$ and $Pr[F^c|E_2] = 1 - .216 = .784$. And of course since a type 3 urn only has black balls in it, $p_3 = 1$ and so $Pr[F|E_3] = 1$ as noted above. These probabilities allow us to complete the tree. Also, we can calculate the path probabilities and show them on the ends of the paths.



Now, what is it we were supposed to be finding? Oh yes, the probability that we are drawing from a type 3 urn, given that 3 black balls have been drawn, i.e. $Pr[E_3|F]$. Again, we use Bayes' Theorem.

$$\begin{aligned} Pr[E_3|F] &= \frac{\text{path prob. for } (E_3 \cap F) \text{ path}}{\text{sum of path prob.'s for all paths with } F} \\ &= \frac{.5}{.0064 + .0864 + .5} = \frac{.5}{.5928} \approx 0.8435 \end{aligned}$$

Notice: Before we drew any balls, there was a 50% chance that we were drawing from a type 3 urn. We saw in (a) that after drawing 1 black ball, that probability had gone up to about 64%. Unless we draw a white ball, which would tell us that we are definitely *not* drawing from a type 3 urn, we can never be absolutely certain whether we're drawing from a type 3 urn. However, after drawing 3 black balls and no white balls, there's a better than 84% chance that we're drawing from a type 3 urn. If we were to continue drawing, and continue getting only black balls, the probability that we are, in fact, drawing from a type 3 urn would continue to increase.

To finish off this topic, go back and look again at Example 2.32 Revisited, at the end of Lecture 18. What we did there was just like what we did here. In that case, what we called event E was really the event that $k = 2$ successes (non-defective parts) were observed on $n = 2$ Bernoulli trials, so the probabilities on the E branches had the form $\binom{2}{2}p^2q^0 = 1 \times p^2 \times 1 = p^2$, where p is the probability of producing a non-defective part, which depends on whether the machine was set up properly ($p = .98$) or improperly ($p = .8$).

Math 1228A/B Online

Lecture 21:

Discrete rv's, pdf's, cdf's, etc.

(text reference: Section 3.1, pages 112 - 117)

3 Discrete Random Variables

3.1 Probability Distributions and Random Variables

Often, when we perform a probabilistic experiment, or a series of probabilistic experiments, what we're interested in is not so much the actual outcome observed, but rather some number, some numerical value, associated with the outcome.

We define:

> A **random variable** (or **r.v.**) is a function which associates some real number with each element of a sample space.

And also:

> A **discrete random variable** is a random variable for which the set of possible values is finite, i.e. only a finite number of different numerical values are used.

Note: If a random variable has infinitely many possible values, we call it **continuous**. We will study continuous random variables in Chapter 4.

We use capital letters, usually X or Y or Z, to denote r.v.'s. (The same letter, in lower case, is used to denote an unspecified possible value of the r.v. – for instance we would use x to represent a possible value of the random variable X.) We denote the event that some r.v. X has some particular possible value x by $(X = x)$, and the probability of this event by $Pr[X = x]$. Let's look at an example of a discrete random variable, it's possible values, events and probabilities.

Example 3.1. *A coin is tossed 3 times. Define X to be a random variable corresponding to the number of times Heads is observed. For each possible event $(X = x)$, list the outcomes associated with the event and find $Pr[X = x]$.*

First, we need to determine a sample space for this experiment. We toss a coin 3 times, so we can use

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

The random variable X is defined to be the number of times heads comes up. When we toss a coin 3 times, we might get 0, 1, 2 or 3 heads, so these are the possible values of X. Since we can count the possible values, i.e. there are finitely many of them, X is a discrete random variable. We can identify the X-value associated with each sample point by simply counting the H's it contains.

There is one event for each possible value. That is, we have events $(X = 0)$, $(X = 1)$, $(X = 2)$ and $(X = 3)$. *Note:* We have defined an equiprobable sample space, with $n(S) = 8$, so we have

$$Pr[X = x] = \frac{n(X = x)}{n(S)} = \frac{n(X = x)}{8}$$

We can make a table to show the outcomes corresponding to each event, and the probability of each event.

Event	Outcomes	Probability
$(X = 0)$	TTT	$\frac{1}{8}$
$(X = 1)$	HTT, THT, TTH	$\frac{3}{8}$
$(X = 2)$	HHT, HTH, THH	$\frac{3}{8}$
$(X = 3)$	HHH	$\frac{1}{8}$

We said that a random variable is a function associating a numeric value with each outcome of the experiment. That's what we have in this example. The function that is the random variable X associates the number 0 with the outcome TTT, the number 1 with each of the outcomes HTT, THT and TTH, etc.

Note: There are other ways we could have approached this problem. For instance, we could have drawn a probability tree to model the 3 coin tosses. We would find all the path probabilities, and also note, at the end of each path, the value of X for that path – by counting the H branches along the path. Alternatively, we could realize that we are performing $n = 3$ Bernoulli trials in which the probability of success (getting Heads) is $\frac{1}{2}$. Then X is simply counting the number of successes, so for each of $x = 0, 1, 2, 3$ we get

$$Pr[X = x] = \binom{3}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{3-x} = \binom{3}{x}\left(\frac{1}{2}\right)^3$$

There's another function which is associated with a discrete random variable. This is the function which associates a *probability* with each possible value of the random variable. We call this the **probability distribution function** of the r.v., often abbreviated **pdf**. That is, we define:

> The **probability distribution function**, or **pdf**, of a discrete random variable X gives $Pr[X = x]$ for each possible event $(X = x)$, i.e. for each possible value x of X.

We found the pdf of the random variable in the previous example, by finding the probabilities of the various possible values, i.e. for the events $(X = x)$ for $x = 0, 1, 2, 3$.

There are 2 ways in which pdf's are often presented. One is to use a table. Usually the table would have column headings x and $Pr[X = x]$, i.e. a column listing the possible values and a column showing the probability of that value being observed. For instance, for the r.v. in the previous example, we would usually present the pdf as:

x	$Pr[X = x]$
0	1/8
1	3/8
2	3/8
3	1/8

The other way we can present a pdf is by using a bar graph, called a **histogram**. This bar graph has a bar for each possible value of X. Each bar has width 1, with the bar for event $(X = x)$ being centred at x, which is shown on the horizontal axis. The vertical axis measures probability. The height of the bar centred at x is $Pr[X = x]$. *Notice:* This means that the area of the bar associated with the event $(X = x)$ is width $\times$ height $= 1 \times Pr[X = x]$, so the area is $Pr[X = x]$.

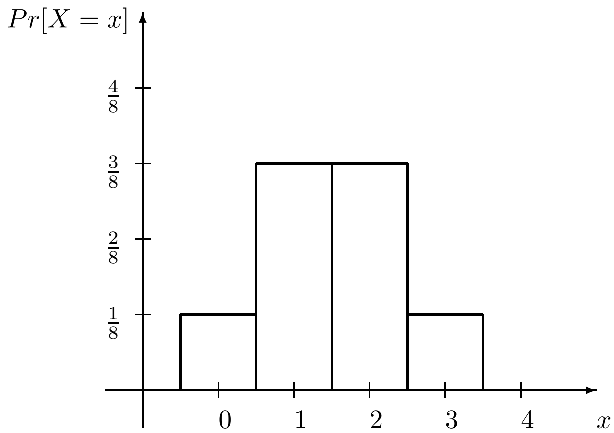
Example 3.2. *Draw the histogram of the pdf of the random variable X from Example 3.1.*

The histogram is a graph in which the horizontal axis is labelled x and the vertical axis is labelled $Pr[X = x]$. Since X has possible values 0, 1, 2 and 3, these are the values shown on the x axis. The histogram has a bar for each of these possible values of x.

Each bar has width 1 and is centred at the possible value. For instance, the bar for the event $(X = 0)$ is centred at $x = 0$, with width 1, so it runs (horizontally) from $x = -\frac{1}{2}$ to $x = \frac{1}{2}$. Similarly, there is a bar for $(X = 1)$ which goes from $x = \frac{1}{2}$ to $x = \frac{3}{2}$, and the bars for $(X = 2)$ and $(X = 3)$ span $x = \frac{3}{2}$ to $x = \frac{5}{2}$ and $x = \frac{5}{2}$ to $x = \frac{7}{2}$, respectively.

The height of each bar is the probability of the corresponding event. For instance, since $Pr[X = 0] = \frac{1}{8}$, the bar for event $(X = 0)$, centred at $x = 0$, has height $\frac{1}{8}$. Similarly, the bars for $(X = 1)$ and $(X = 2)$ each have height $\frac{3}{8}$ and the bar for $(X = 3)$ has height $\frac{1}{8}$.

(*Note:* We need the vertical axis to show on the histogram, but it does not have to be shown at $x = 0$, as it would on most graphs with x on the horizontal axis. In a histogram, we usually show the vertical axis over to the left of the first bar, so that it is not getting in the way of any of the bars.)



Histogram for the r.v. X in Example 3.1

There is also one more function associated with a random variable, called the **cumulative distribution function**, or **cdf**. We define:

> For any random variable X, the **cumulative distribution function**, or **cdf**, of X, denoted $F(x)$, is defined to be
>
> $$F(x) = Pr[X \leq x]$$

For any discrete random variable, we can find the values of $F(x)$ by adding up the values of the pdf, for all possible values up to and including x. The cumulative distribution function for a discrete random variable can be displayed in a table similar to that used for the pdf.

Example 3.3. *Make a table showing the cumulative distribution function of the random variable X from Example 3.1.*

First, of course, we need to find the values of the cdf of X. That is, we want to find and tabulate the values of $Pr[X \leq x]$ for each possible value. So we need to find $Pr[X \leq 0]$, $Pr[X \leq 1]$, $Pr[X \leq 2]$ and $Pr[X \leq 3]$.

For $Pr[X \leq 0]$, we see that $(X \leq 0) = (X = 0)$; that is, since 0 is the lowest possible value of X, the event $(X < 0)$ is the empty set so the event $(X \leq 0)$ is identical to the event $(X = 0)$. This gives us

$$F(0) = Pr[X \leq 0] = Pr[X = 0] = \frac{1}{8}$$

Because the only values X can have which satisfy $X \leq 1$ are 0 and 1, then we see that $(X \leq 1) = (X = 0) \cup (X = 1)$. And of course those are mutually exclusive events, so for $Pr[X \leq 1]$ we have

$$F(1) = Pr[X \leq 1] = Pr[X = 0] + Pr[X = 1] = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}$$

Likewise, for $Pr[X \leq 2]$, we have $(X \leq 2) = (X = 0) \cup (X = 1) \cup (X = 2)$, and so we get

$$\begin{aligned} F(2) = Pr[X \leq 2] &= Pr[X = 0] + Pr[X = 1] + Pr[X = 2] \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8} \end{aligned}$$

Alternatively, we could view this as $(X \leq 2) = (X \leq 1) \cup (X = 2)$, and use $F(1) = Pr[X \leq 1]$, which we alredy found, to get

$$Pr[X \leq 2] = Pr[X \leq 1] + Pr[X = 2] = \frac{4}{8} + \frac{3}{8} = \frac{7}{8}$$

Finally, for $Pr[X \leq 3]$, we have $(X \leq 3) = (X = 0) \cup (X = 1) \cup (X = 2) \cup (X = 3)$, and so

$$\begin{aligned} F(3) = Pr[X \leq 3] &= Pr[X = 0] + Pr[X = 1] + Pr[X = 2] + Pr[X = 3] \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1 \end{aligned}$$

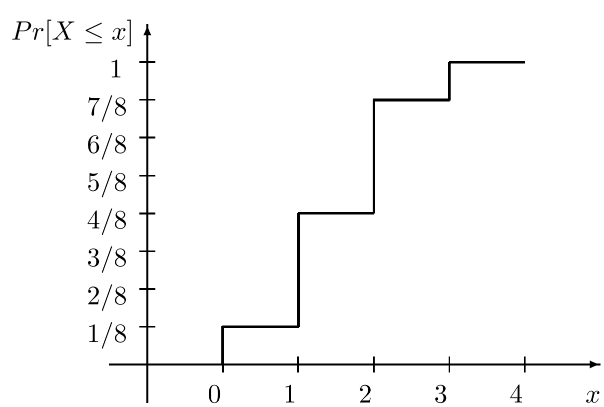
Or alternatively, $(X \leq 3) = (X \leq 2) \cup (X = 3)$ which gives

$$Pr[X \leq 3] = Pr[X \leq 2] + Pr[X = 3] = \frac{7}{8} + \frac{1}{8} = 1$$

Thus we have:

x	$F(x) = Pr[X \leq x]$
0	1/8
1	1/2
2	7/8
3	1

Notice: The cdf for a random variable can also be displayed in a graph. It is not a bar graph (like the histogram of the pdf) but instead is a step graph – a line graph in which the value jumps up each time we get to a different possible value of the random variable. For instance, the graph of the cdf of X in our example shows a line at height $\frac{1}{8}$ from 0 to 1 (on the horizontal or x axis), then jumps up by $\frac{3}{8}$ to height $\frac{1}{2}$ from x-values 1 to 2, jumps by $\frac{3}{8}$ again, to $\frac{7}{8}$, at 2, and finally jumps up $\frac{1}{8}$, to height 1, at $x = 3$. That is, at each possible value, x, the line jumps by $Pr[X = x]$ to a new height which is the (new) value of $F(x)$. For this example, we have:



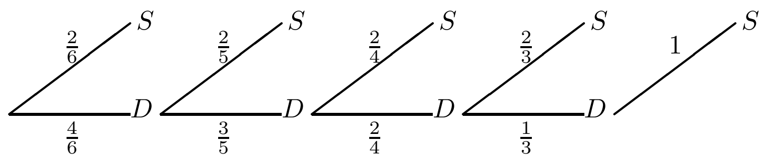
The cdf of the r.v. X in Example 3.1

We will not be much concerned with the graphs of cdf's in this course. We will normally present the cdf of a discrete r.v. in tabular form, or simply state values of $F(x)$.

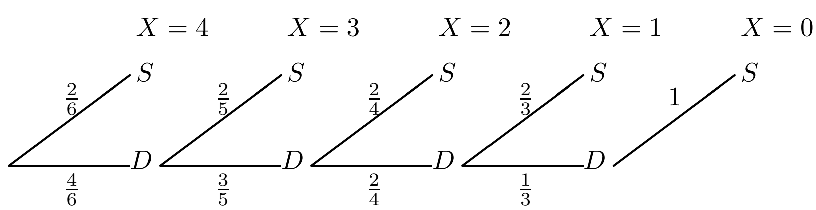
Next lecture, we'll discuss various properties which the function $F(x)$ has. For now, let's just look at another example.

Example 3.4. *Recall Example 2.31 in which a professor has a box containing 6 pencils, only 2 of which are sharp. The professor draws pencils from the box one by one until a sharp pencil is found. Let X be the number of dull pencils still in the box after the professor has found a sharp pencil. Find the cdf of X.*

In Example 2.31 (see Lecture 17, p. 94), we drew a probability tree to model the stochastic process described here. We had:



We can put the value of X at the end of each path through this tree. For any path, each D branch in the path represents a dull pencil which has been removed from the box. The box started out with 4 dull pencils, so we can find the number of dull pencils left in the box at the end by subtracting the number of D branches in the path from 4. For instance, along the D-then-S path, 1 dull pencil has been drawn, so at the end of this path, the experiment has ended with $4 - 1 = 3$ dull pencils left in the box.



In this situation, each path through the tree has a different X-value, so each value of the pdf is a different path probability. We get:

x	$Pr[X = x]$
0	$\frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} = \frac{1}{15}$
1	$\frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} = \frac{2}{15}$
2	$\frac{4}{6} \times \frac{3}{5} \times \frac{2}{4} = \frac{1}{5} = \frac{3}{15}$
3	$\frac{4}{6} \times \frac{2}{5} = \frac{4}{15}$
4	$\frac{2}{6} = \frac{1}{3} = \frac{5}{15}$

Now we can find the cdf by summing the pdf:

x	$F(x) = Pr[X \leq x]$
0	$\frac{1}{15}$
1	$\frac{1}{15} + \frac{2}{15} = \frac{3}{15} = \frac{1}{5}$
2	$\frac{3}{15} + \frac{3}{15} = \frac{6}{15} = \frac{2}{5}$
3	$\frac{6}{15} + \frac{4}{15} = \frac{10}{15} = \frac{2}{3}$
4	$\frac{10}{15} + \frac{5}{15} = \frac{15}{15} = 1$

Notice: In this situation, each path corresponded to a different value of X. This is not normally the case (although it always is in the kind of situation we had here). For instance, if we drew the tree for the discrete r.v. in Example 3.1, there would be 8 different paths through the tree, but there are only 4 different possible X-values. Some X-values appear on more than one path. Of course, when the event $(X = x)$ occurs on more than one path, we find $Pr[X = x]$ by summing the path probabilities for all such paths.

Also Notice: There are various discrete r.v.'s which could be defined in this kind of situation. Instead of counting the number of dull pencils left in the box, we could count the number of dull pencils drawn, or the total number of pencils drawn (or left in the box), i.e. including the sharp pencil that is the last pencil drawn in each case. The tree is the same, and so the path probabilities don't change, but in each of these situations X would be counting something different, so the possible values **do** change, i.e. the same path would correspond to a different value of X. Always make sure you are using the correct random variable in these kinds of problems.

Math 1228A/B Online

Lecture 22:

Properties of the cdf,

... And Binomial random variables

(text reference: Section 3.1, pages 117 - 118 and also 119 - 120)

There are certain properties which a cumulative distribution function, $F(x)$, must always have. You will have noticed some of these already, while we were finding $F(x)$ in the examples. In order to express these properties in general terms, we need to order the (unspecified) possible values of some generic discrete r.v. X. This is done *only* for the purpose of expressing the properties. We will not normally use this kind of notation. (When we are dealing with a specific r.v., the possible values are specific numbers, so we don't need special notation in order to know which is bigger or smaller than which.) In the following, Example 3.3 is referred to, so you may want to review that example before, or have it handy while, reading the following.

Properties of F(x)

Let X be a discrete random variable whose possible values are, in ascending order, $x_1, x_2, ..., x_n$. That is, we have $x_1 < x_2 < ... < x_n$.

1. $F(x_1) = Pr[X = x_1]$
   That is, for the smallest possible value of X, the pdf and cdf have the same value, because $Pr[X < x_1] = 0$. In Example 3.3, we saw that $F(0) = Pr[X \leq 0] = Pr[X = 0]$, since X cannot have any value less than 0.

2. $F(x_j) > F(x_i)$ for any possible values x_i and x_j with $j > i$.
   That is, the cdf is an increasing function, i.e. gets bigger with each bigger possible value. In the example, we found that $F(3) > F(2) > F(1) > F(0)$.

3. $F(x_n) = 1$
   That is, for the largest possible value of X, the value of the cdf must be 1 (because $Pr[X > x_n] = 0$ and X must have *some* value, so the total probability must be 1). For instance, in the example we had $F(3) = 1$.

4. $Pr[X < x_k] = Pr[X \leq x_{k-1}] = F(x_{k-1})$
   That is, the probability that X is strictly less than some particular possible value is the same as the probability that X is no bigger than the *next smaller* possible value. For instance, in our example, since the only way that X can be strictly less than 3 is to be no bigger than 2, we get $Pr[X < 3] = Pr[X \leq 2] = F(2)$.

5. $Pr[X > x_k] = 1 - F(x_k)$ and $Pr[X > x_k] = Pr[X \geq x_{k+1}]$
   That is, the complement of the event $(X > x_k)$ is the event $(X \leq x_k)$. Also, the probability that X is strictly bigger than some particular possible value is the probability that X is at least as big as the next larger possible value. For instance, in our example, if X is bigger than 1, it must be at least 2, so that $Pr[X > 1] = Pr[X \geq 2] = Pr[(X = 2) \cup (X = 3)] = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}$. Likewise, if X is *not* bigger than 1, it must be at most 1, so $(X > 1)^c = (X \leq 1)$ and we have $Pr[X > 1] = 1 - Pr[(X > 1)^c] = 1 - Pr[X \leq 1] = 1 - F(1) = 1 - \frac{1}{2} = \frac{1}{2}$.

6. $Pr[X = x_k] = F(x_k) - F(x_{k-1})$
   That is, we can find $Pr[X = x_k]$ from the cdf, by subtracting the cdf value for the next smaller value. This is because we have

   $$Pr[X \leq x_k] = Pr[X < x_k] + Pr[X = x_k] = Pr[X \leq x_{k-1}] + Pr[X = x_k]$$

   We see that $F(x_k) = F(x_{k-1}) + Pr[X = x_k]$. We used this relationship in calculating some of the cdf values in the example. That is, we took the previous cdf value and added the probability for the next possible value. We simply rearrange this relationship to get $Pr[X = x_k] = F(x_k) - F(x_{k-1})$, which is what the property states.

What this property means is that if we "lose" the pdf and have only the cdf, we can "recover" the pdf values from the cdf values. That is, given a pdf, we can find the cdf, but also, since no information is lost in doing so, then given a cdf, we are able to find the pdf.

7. $Pr[x_i \leq X \leq x_j] = F(x_j) - F(x_{i-1})$ for any $x_i < x_j$.
That is, if we want to find the probability that X has a value between 2 particular possible values, inclusive, we take the probability that X *is not more than the larger value*, and then subtract off the part we didn't want to include, which is the probability that X *is strictly smaller than the smaller value*. And of course, (as noted in Property 4) the probability that X is strictly smaller than a particular value is the probability that it's not more than the next smaller possible value.

   For instance, suppose that a and b are possible values of X, with $b > a$. Then $(X \leq b) = (X < a) \cup (a \leq X \leq b)$, i.e. we partition the event $(X \leq b)$ into the part that's also less than a, and the part that's at least as big as a. So we get

   $Pr[X \leq b] = Pr[X < a] + Pr[a \leq X \leq b] \Rightarrow Pr[a \leq X \leq b] = Pr[X \leq b] - Pr[X < a]$

   From here, the above property (as stated) is just re-expressing this by realizing that $Pr[X \leq b]$ is just the cdf value of b, and $Pr[X < a]$ is (by Property 4) the same as the cdf value of the next smaller possible value from a.

The next example uses some of these properties.

Example 3.5. *$F(x)$ is the cumulative distribution function for some random variable X which assumes only integer values. Given that $F(6) = \frac{5}{12}$, $F(7) = \frac{7}{12}$ and $F(9) = \frac{5}{6}$, find the following probabilities:*
(a) the probability that the random variable X has the value 7;
(b) the probability that X has a value less than 10;
(c) the probability that X is at least 10;
(d) the probability that X has a value not less than 7 and not more than 9;
(e) the probability that X has a value less than 7 or greater than 9.

(a) The probability that X has the value 7 is $Pr[X = 7]$. We know that $F(7) = \frac{7}{12}$ and that $F(6) = \frac{5}{12}$. Also, we are told that all of the possible values of X are integers, so there can't be any other possible value of X between 6 and 7, and we see that 6 must be the "next smaller" possible value from 7. So by Property 6, we see that

$$Pr[X = 7] = F(7) - F(6) = \frac{7}{12} - \frac{5}{12} = \frac{2}{12} = \frac{1}{6}$$

(b) We are asked to find the probability that X has a value less than 10, i.e. $Pr[X < 10]$. We don't know whether or not 10 is a possible value of X, but we do know that 9 is, and that there are no other possible values of X that are bigger than 9 but less than 10, since X has only integer values. That is, we know that 9 is the next smaller possible value from 10. Therefore, by Property 4, we have

$$Pr[X < 10] = Pr[X \leq 9] = F(9) = \frac{5}{6}$$

(c) We want to find the probability that X is at least 10, i.e. $Pr[X \geq 10]$. We know that the next smaller possible value of X (from 10) is 9. Using both parts of Property 5, we get

$$Pr[X \geq 10] = Pr[X > 9] = 1 - F(9) = 1 - \frac{5}{6} = \frac{1}{6}$$

That is, the event $(X \geq 10)$ is the complement of the event $(X < 10)$ whose probability we found in part (b).

(d) We need to find the probability that X has a value not less than (i.e. at least) 7 and not more than (i.e. at most) 9. Of course, X being at least 7 is the event $(X \geq 7)$ and X being at most 9 is the event $(X \leq 9)$, so we are looking for the probability of $(X \geq 7) \cap (X \leq 9) = (7 \leq X \leq 9)$. (That is, for X to be at least 7 and at most 9, we must have X between 7 and 9 inclusive.) As in Property 7, we use the fact that

$$Pr[X \leq 9] = Pr[X < 7] + Pr[7 \leq X \leq 9] \text{ where we have } Pr[X < 7] = Pr[X \leq 6]$$

to get

$$Pr[7 \leq X \leq 9] = F(9) - F(6) = \frac{5}{6} - \frac{5}{12} = \frac{10-5}{12} = \frac{5}{12}$$

That is, the probability that X is between 7 and 9 inclusive is that total probability that X is no bigger than 9, less the probability that X is smaller than 7 and is therefore no bigger than 6, the next smaller possible value. (*Note:* Instead of thinking it through as we have done here, we could have simply applied Property 7 directly to get $Pr[7 \leq X \leq 9] = F(9) - F(6) = \frac{5}{12}$.)

(e) We want to find the probability that X has a value less than 7 or greater than 9, i.e. the probability of the event $(X < 7) \cup (X > 9)$. The easiest way to find this is to realize that this is the complement of the event from part (d). That is, if X is less than 7 or greater than 9, then X is *not* between 7 and 9 inclusive, so we have

$$\begin{aligned} Pr[(X < 7) \cup (X > 9)] &= Pr[(7 \leq X \leq 9)^c] \\ &= 1 - Pr[7 \leq X \leq 9] \\ &= 1 - \frac{5}{12} = \frac{7}{12} \end{aligned}$$

Note: Rather than trying to memorize the formulas stated in the list of properties of the cdf, you should try to understand what the properties are saying. They're really all just fairly obvious consequences of simple relationships.

Final Comment on cdf's: With discrete random variables, we must *always* be very careful in expressing the complement of an event involving an inequality. Complementation changes both the *direction* and the *inclusiveness* of an inequality. That is, $(X \geq k)^c = (X < k)$ and $(X \leq k)^c = (X > k)$. *At least* k is the complement of *strictly less than* k, and *at most* k is the complement of *strictly more than* k. Think carefully about it logically, and express it in English to help you see how to express it in Math. And think about which event the value $(X = k)$ is in, because it can *never* be in both, but must *always* be in one of them, i.e. either in the event of interest or in its complement.

For the rest of this lecture, we talk about a particular kind of discrete random variable ...

Binomial Random Variables

There is a special kind of discrete random variable called a Binomial random variable. These are the most frequently used discrete r.v.'s, because they are found in any real-world application which involves repeated independent trials of the same experiment.

Definition:
If the random variable X counts the number of successes in a series of Bernoulli trials, then X is called a **Binomial random variable** and we say that X has a **Binomial distribution**.

We have special notation we use for Binomial random variables. If X is a Binomial r.v., counting the number of success on n Bernoulli trials in which the probability of success is p, then rather than expressing this in words, we use the expression $B(n, p)$. That is, we have

Definition:
$\mathbf{B(n, p)}$ means the Binomial random variable which counts the number of successes in n Bernoulli trials with probability of success p.

For instance, $X = B(100, .3)$ says that X is a discrete random variable whose value is given by observing the number of successes which occur when 100 independent trials of the same experiment are performed, where there is, on each trial, a .3 probability of observing the event which is considered to be 'a success'.

Of course, we know that the probability of observing k successes in n Bernoulli trials with probability of success p is given by $\binom{n}{k}p^k(1-p)^{n-k}$. That is, we have

$$Pr[B(n,p) = k] = \binom{n}{k}p^k(1-p)^{n-k}$$

Example 3.6. *Let X be the number of times 2 comes up when a fair die is tossed 4 times.*
(a) Make tables of the pdf and the cdf of X.
(b) What is the probability that at most two 2's are observed?

Repeatedly tossing a die corresponds to performing a series of independent trials. Counting the number of 2's which come up in 4 tosses of a die corresponds to performing $n = 4$ Bernoulli trials in which the probability of success is $p = \frac{1}{6}$ (because when a fair die is tossed, the probability that a 2 comes up is 1/6). Since the r.v. X is counting the number of successes in a series of Bernoulli trials, then X is a Binomial random variable. That is, we see that $X = B\left(4, \frac{1}{6}\right)$.

(a) To find the pdf of X, we repeatedly use the Bernoulli formula, i.e. we find $Pr[B\left(4, \frac{1}{6}\right) = x]$ for each possible value of x. Of course, when a die is tossed 4 times, we could observe 0, 1, 2, 3 or 4 two's, so these are the possible values of X. Therefore we need to find $Pr[B\left(4, \frac{1}{6}\right) = x]$ for $x = 0$, 1, 2, 3 and 4. Using

$$Pr[B(n,p) = x] = \binom{n}{x}p^x(1-p)^{n-x} \text{ so that for } X = B\left(4, \frac{1}{6}\right), \quad Pr[X = x] = \binom{4}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{4-x},$$

then for $x = 0$ we have

$$Pr[X = 0] = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{4-0} = 1 \times 1 \times \left(\frac{5}{6}\right)^4 = \frac{625}{1296}$$

Similarly, we get

$$\begin{aligned}
Pr[X=1] &= \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{4-1} = 4\times\left(\frac{1}{6}\right)^1\times\left(\frac{5}{6}\right)^3 = \frac{500}{1296} \\
Pr[X=2] &= \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{4-2} = 6\times\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^2 = \frac{150}{1296} \\
Pr[X=3] &= \binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{4-3} = 4\times\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^1 = \frac{20}{1296} \\
Pr[X=4] &= \binom{4}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{4-4} = 1\times\left(\frac{1}{6}\right)^4\times 1 = \frac{1}{1296}
\end{aligned}$$

We tabulate these values:

pdf for $X = B\left(4, \frac{1}{6}\right)$

x	$Pr[X = x]$
0	$\frac{625}{1296}$
1	$\frac{500}{1296}$
2	$\frac{150}{1296}$
3	$\frac{20}{1296}$
4	$\frac{1}{1296}$

To find the cdf of X, we simply find successive sums of the pdf, as always. That is, we need to find, for each possible value x,

$$F(x) = Pr[X \leq x] = Pr[X = 0] + Pr[X = 1] + ... + Pr[X = x]$$

For instance, we have

$$F(1) = Pr[X \leq 1] = B(1; 4, 1/6) = Pr[B(4, 1/6) = 0] + Pr[B(4, 1/6) = 1] = \frac{625}{1296} + \frac{500}{1296} = \frac{1125}{1296}$$

We get:

cdf for $X = B\left(4, \frac{1}{6}\right)$

x	$F(x)$
0	$\frac{625}{1296}$
1	$\frac{1125}{1296}$
2	$\frac{1275}{1296}$
3	$\frac{1295}{1296}$
4	$\frac{1296}{1296} = 1$

(b) To find the probability of tossing at most two 2's, we simply need to choose the right number from the right table from part (a). Having at most two 2's come up, i.e. at most 2 successes, is the event $(X \leq 2)$. For $F(2) = Pr[X \leq 2]$, we look in the table showing the cdf. We see that $F(2) = \frac{1275}{1296}$.

Note: If we hadn't already found the full pdf oand cdf of X, we would either use

$$F(2) = Pr[X \leq 2] = Pr[X = 0 \text{ or } 1 \text{ or } 2] = Pr[X = 0] + Pr[X = 1] + Pr[X = 2]$$

and use the Bernoulli formula 3 times, or else use

$$F(2) = Pr[X \leq 2] = 1 - Pr[X > 2] = 1 - Pr[X = 3 \text{ or } 4] = 1 - (Pr[X = 3] + Pr[X = 4])$$

and use the Bernoulli formula only twice.

Notice: We have already found the pdf and cdf of another Binomial rv, without realizing that was what it was. In Example 3.1, the value of X was the number of times Heads was observed on 3 tosses of a coin. Of course, tossing a coin 3 times corresponds to performing $n = 3$ independent trials, and considering success to be that Heads comes up, we see that $X = B(3, .5)$. That is, in Example 3.1 and Example 3.3 what we found was the pdf and cdf of $B(3, .5)$.

Properties of $\mathbf{B(n, p)}$

1. If $X = B(n, p)$, the possible values of X are 0, 1, ..., n.
2. If $X = B(n, p)$, then for each possible value x,
$$Pr[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$$
3. If $X = B(n, p)$, then for each possible value x,
$$F(x) = Pr[X \leq x] = Pr[X = 0] + Pr[X = 1] + ... + Pr[X = x]$$
(where each $Pr[X = k]$ term in the sum is found using $Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$).
4. If $X = B(n, p)$, then for each possible value x,
$$Pr[X > x] = Pr[X = x + 1] + Pr[X = x + 2] + ... + Pr[X = n]$$
Note: If we're not finding the whole pdf of $B(n, p)$, then for any possible value x which is closer to n than to 0 it is easiest (i.e. few Bernoulli calculations involved) to find $Pr[X \leq x]$ using the complement. For instance, for $B(10, p)$, the easiest way to find $F(8)$ is by
$$F(8) = Pr[X \leq 8] = 1 - Pr[X > 8] = 1 - (Pr[X = 9] + Pr[X = 10])$$
which requires using the Bernoulli formula only twice, instead of 9 times.
5. Of course, in any series of Bernoulli trials, the number of successes and the number of failures must sum to the total number of trials. That is, having k successes on n trials also means having $n - k$ failures. The random variable that counts the number of *failures* on a series of Bernoulli trials is *also* a Binomial rv, based on the same number of trials, with probability $1 - p$ instead of p. So
$$B(n, 1-p) = n - B(n, p) \text{ and } Pr[B(n, p) = k] = Pr[B(n, 1-p) = n - k]$$
For instance, for $X = B(10, .2)$ we can also consider $Y = B(10, .8)$ and it must be true that $X + Y = 10$, i.e. we have $Y = n - X$. And having at most x successes can also be considered as having at least $n - x$ failures, so
$$Pr[X \leq x] = Pr[Y \geq n - x]$$

Next, consider the following:

> **Fact:** Selecting a small random sample from a **very large** population can be considered as performing independent trials.

Notice: If the population is not "very large" then this isn't true. Once we have selected one or more members of the population to be in the sample, the probability that the next randomly selected member (who must be someone not already selected) has the characteristic we're interested in changes, depending on how many of the members already selected (effectively, removed from the population) have the characteristic. That is, really what we're doing here is sample **without replacement**, which we know does **not** correspond to performing independent trials. (For instance, when we draw several balls from an urn or select several cards from a deck.) However, when the population is **very large**, sampling without replacement is *effectively* the same as sampling **with** replacement, because the probabilities do not change *significantly*.

For instance, suppose we have a (large) urn containing 200 red balls and 800 black balls, and we draw 2 balls without replacement. On the first draw, the probability of getting a red ball is $\frac{200}{1000} = .2$ and the probability of getting a black ball is $\frac{800}{1000} = .8$. On the second draw, if the first ball was black, then the probaiblity of drawing a red ball is $\frac{200}{999} \approx .2002002002...$ (i.e. the 200 keeps repeating), and the probability of drawing a black ball is $\frac{799}{999} \approx .79979979...$ (i.e. the 799 keeps repeating). And if the first ball was red, then on the second draw the probability of getting another red ball is $\frac{199}{999} \approx .199199199...$ and the probability of getting a black ball is $\frac{800}{999} \approx .800800800....$ We see that no matter what colour ball was drawn first, on the second draw the probability of getting a red ball is very close to .2, and the probability of getting a black ball is very close to .8.

In terms of the whole sample of 2 balls, the actual probabilities of getting 0, 1 or 2 red balls (i.e. getting 2 black, 1 red and 1 black, or 2 red) are given by

$$\begin{aligned} Pr[\text{no red}] &= \frac{\binom{200}{0}\binom{800}{2}}{\binom{1000}{2}} \approx .6398 \\ Pr[\text{1 red}] &= \frac{\binom{200}{1}\binom{800}{1}}{\binom{1000}{2}} \approx .3203 \\ Pr[\text{2 red}] &= \frac{\binom{200}{2}\binom{800}{0}}{\binom{1000}{2}} \approx .0398 \end{aligned}$$

If we were sampling *with* replacement, these probabilities would be

$$\begin{aligned} Pr[\text{no red}] &= \binom{2}{0}(.2)^0(.8)^2 = (.8)^2 = .64 \\ Pr[\text{1 red}] &= \binom{2}{1}(.2)^1(.8)^1 = 2(.2)(.8) = .32 \\ Pr[\text{2 red}] &= \binom{2}{2}(.2)^2(.8)^0 = (.2)^2 = .04 \end{aligned}$$

The difference between the 2 sets of probability values is very small. (Compare this with the situation in which there are 2 red and 8 black balls in the urn. In that case, the actual probabilities are $Pr[\text{no red}] \approx .622222222...$, $Pr[\text{1 red}] = .3555555555...$ and $Pr[\text{2 red}] = .022222222222...$, so the difference between the probabilities when sampling with or without replacement is much more pronounced. Even with 100 balls (20 red, 80 black), there is a noticeable difference. But with 1000 balls one could almost say the difference is negligible.)

Example 3.7. *In a very large first year Calculus course, 30% of the students failed the final exam. A random sample of 10 students is selected from the students in the course.*

(a) What is the probability that exactly 3 of the students in the sample failed the exam?
(b) What is the probability that at most 2 of the students failed?
(c) What is the probability that at least 9 of the students passed?

Since we are told that the class is very large, and the sample is quite small, we can consider the sample of 10 students as performing 10 independent trials of the experiment "randomly choose a student from the course and see whether the chosen student passed or failed the final exam". In fact, since we're not told exactly how many students are in the course, we don't have enough information to approach this problem any other way.

(a) We need to find the probability that exactly 3 of the 10 students in the sample failed the exam. Let X be the number of students in the sample who failed. Then, considering "failed the exam" as success (an odd definition of success, but remember with Bernoulli trials 'success' only means 'the outcome whose occurrences we want to count'), we have $X = B(10, .3)$ so the probability that exactly 3 of the students failed is

$$Pr[X = 3] = \binom{10}{3}(.3)^3(.7)^7 \approx .2668$$

(Alternatively, we could define Y to be the number who passed, so that $Y = B(10, .7)$ and recognize that having exactly 3 fail corresponds to having exactly 7 of the 10 pass, so the probability that exactly 3 of the students failed is $Pr[Y = 7] = \binom{10}{7}(.7)^7(.3)^3$. This is, of course, exactly the same calculation.)

(b) Using $X = B(10, .3)$ as in (a), the probability that at most 2 of the students in the sample fail is

$$\begin{aligned} Pr[X \leq 2] &= Pr[X = 0] + Pr[X = 1] + Pr[X = 2] \\ &= \binom{10}{0}(.3)^0(.7)^{10} + \binom{10}{1}(.3)^1(.7)^9 + \binom{10}{2}(.3)^2(.7)^8 \\ &\approx .02825 + .12106 + .23347 = .38278 \end{aligned}$$

(Alternatively, using $Y = B(10, .7)$ and recognizing that having at most 2 fail corresponds to having at least 8 of the 10 pass, we find $Pr[Y \geq 8]$, which once again gives the same calculations, but in reverse order if we do it as $Pr[Y = 8] + Pr[Y = 9] + Pr[Y = 10]$.)

(c) Having at least 9 of the students pass corresponds to having at most 1 of them fail, so using $X = B(10, .3)$ we see that the probability this happens is

$$Pr[X \leq 1] = Pr[X = 0] + Pr[X = 1] \approx .02825 + .12106 = .14931$$

Or, since we are asked about the number who pass, it is more obvious here that the r.v. counting those who pass is relevant. As mentioned above, this is $Y = B(10, .7)$, so using this to count the students who pass, we directly get

$$Pr[Y \geq 9] = Pr[Y = 9] + Pr[Y = 10] \approx .12106 + .02825 = .14931$$

Math 1228A/B Online

Lecture 23:
Mean, Variance and Standard Deviation

(text reference: Section 3.3, pages 126 - 129)

3.3 The Mean and Standard Deviation

There are various statistical measures which *summarize* and *give partial information about* a probability distribution. Although they do not give complete information about the distribution, the information they do provide is sufficient for some purposes. The most commonly used statistical measures are the mean, variance and standard deviation.

> *Definition:* The **mean**, also called the **expected value**, of a discrete random variable, X, is the weighted average of the possible values of X, where each possible value is weighted by its probability. We denote the mean by μ or $\mathbf{E(X)}$ (the E stands for expected value).

This means that:

> **Formula:** If X has possible values x_1, x_2, ..., x_n, then taking the weighted average, where each value is weighted by it's probability, gives:
>
> $$\mu = E(X) = (x_1 \times Pr[X = x_1]) + (x_2 \times Pr[X = x_2]) + ... + (x_n \times Pr[X = x_n])$$
>
> which we can also write as
>
> $$\mu = \sum_{i=1}^{n} (x_i Pr[X = x_i])$$

Note: The terms *mean* and *expected value* are used interchangeably, as are μ and $E(X)$.

Also Note: The notation $\sum_{i=1}^{n} f(x_i)$ means take the sum, for integer values of i going from $i = 1$ to $i = n$, of the expression $f(x_i)$, i.e. some function involving x_i.

Example 3.8. *Find the expected value of the random variable X whose value is the number observed when a single die is tossed.*

In this case, X is a discrete random variable with possible values 1, 2, ..., 6, where each value is equally likely to occur so that $Pr[X = x] = \frac{1}{6}$ for each possible value. Thus we get

$$\begin{aligned} E(X) &= (1 \times Pr[X = 1]) + (2 \times Pr[X = 2]) + (3 \times Pr[X = 3]) \\ &\quad +(4 \times Pr[X = 4]) + (5 \times Pr[X = 5]) + (6 \times Pr[X = 6]) \\ &= 1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + 3 \times \left(\frac{1}{6}\right) + 4 \times \left(\frac{1}{6}\right) + 5 \times \left(\frac{1}{6}\right) + 6 \times \left(\frac{1}{6}\right) \\ &= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2} \end{aligned}$$

We see that the expected value when a die is tossed is 3.5. Since rolling a 3.5 is impossible, clearly this doesn't mean the we "expect" to roll a 3.5 when we toss a die. So what *does* the expected value, or mean, of a discrete random variable mean? Basically, it means that if we were to repeat the experiment *very many* times, and then take the *average* of the observed outcomes, we expect *this average* to be about 3.5.

Example 3.9. *Find μ for the discrete random variable X defined in Example 3.1 .*

In Example 3.1, we had the experiment *toss a coin 3 times*, and we defined X to be the number of times that Heads comes up. We found the probability distribution of X to be:

x	$Pr[X = x]$
0	1/8
1	3/8
2	3/8
3	1/8

Thus we get:

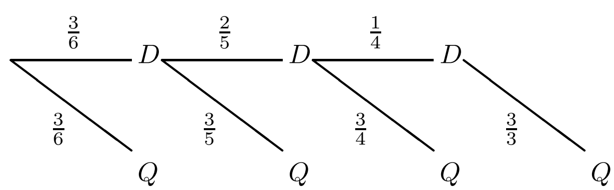
$$\begin{aligned} \mu &= \left(0 \times \frac{1}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(2 \times \frac{3}{8}\right) + \left(3 \times \frac{1}{8}\right) \\ &= \frac{0}{8} + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} \\ &= \frac{12}{8} = \frac{3}{2} \end{aligned}$$

We see that if we tossed 3 coins, over and over again, we would expect the average number of times Heads came up in the 3 tosses to be around 1.5.

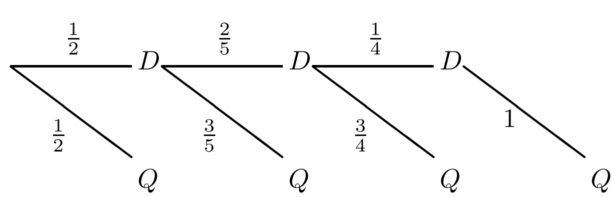
Example 3.10. *I have 3 dimes and 3 quarters in my pocket. I draw coins from my pocket one at a time, at random, without replacement, until I find a quarter. What is the expected number of coins I will draw out of my pocket?*

Let X be the number of coins I draw when I perform this experiment. To see what the possible values of X are, and what their corresponding probabilities are, we can construct a probability tree to model what's going on here. The tree has a level for each time I might reach into my pocket. Whenever I reach into my pocket, the possible outcomes are D: I pull out a dime, and Q: I pull out a quarter.

The probabilities of these outcomes depend on how many coins are still in my pocket, and how many of them are dimes. Notice that the process stops as soon as I find a quarter, so there are always 3 quarters in my pocket when I reach in. It is only the number of dimes remaining which changes with each draw. We get:



We can write the path probabilities at the ends of the paths. Also, at each terminal point we can show the value which X has when the experiment turns out in the way corresponding to that path. (Remember, we have defined X to be the total number of coins drawn.)



path prob.:	1/2	3/10	3/20	1/20
value of X:	$X = 1$	$X = 2$	$X = 3$	$X = 4$

We see that the probability distribution of X is:

x	$Pr[X = x]$
1	1/2
2	3/10
3	3/20
4	1/20

We get:

$$\begin{aligned} E(X) &= \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{3}{10}\right) + \left(3 \times \frac{3}{20}\right) + \left(4 \times \frac{1}{20}\right) \\ &= \frac{10}{20} + \frac{12}{20} + \frac{9}{20} + \frac{4}{20} \\ &= \frac{35}{20} = \frac{7}{4} \end{aligned}$$

Therefore, the expected number of coins I will draw from my pocket is 1.75.

Notice: Recall (see Lecture 21, pg. 120) that if we instead want to count the number of dimes drawn, or the number of dimes/coins **not** drawn (i.e. left in my pocket at the end of the experiment), the probability values in the pdf are the same, but the values of the random variable change, which means the expected value also changes. As before, it is important to be sure you're using the **right** random variable.

Example 3.11. *Consider the following 3 games:*
Game 1:
Toss a coin. If Heads comes up, win \$1; otherwise, lose \$1.
Game 2:
Draw 1 card from a standard deck. If the card is the Ace of Spades, win \$510; otherwise, lose \$10.
Game 3:
Draw 1 card from a standard deck. If the card is black, win \$1,000,000; otherwise, lose \$1,000,000.

Let X, Y and Z be random variables representing a player's winnings (in dollars) on games 1, 2 and 3, respectively. Find $E(X)$, $E(Y)$ and $E(Z)$.

Game 1
In this game, you either win \$1 or lose \$1. The value of X is the number of dollars won, so the possible values of X are 1 and -1. (That is, a loss is a negative win.) The value is determined by the toss of a coin, so both values are equally likely. That is, we have $Pr[X = 1] = Pr[X = -1] = \frac{1}{2}$. The expected value of X is

$$E(X) = (1) \times \left(\frac{1}{2}\right) + (-1) \times \left(\frac{1}{2}\right) = \frac{1}{2} - \frac{1}{2} = 0$$

That is, on average, if you played the game many times, you should break even playing this game.

Game 2
In this game, you either win \$510 or lose \$10, so the possible values of Y, the amount won, are 510 and -10. The value is determined by the draw of one card from a standard deck, so we can define an equiprobable sample space S which is the set of all cards in the deck, with $n(S) = 52$. The event $(Y = 510)$ contains only one of the sample points, the Ace of Spades, so $Pr[Y = 510] = \frac{1}{52}$. Also, the event $(Y = -10)$ contains all of the other 51 cards, so we have $Pr[Y = -10] = \frac{51}{52}$. We see that the expected value of the winnings in this case is

$$E(Y) = (510) \times \left(\frac{1}{52}\right) + (-10) \times \left(\frac{51}{52}\right) = \frac{510}{52} - \frac{510}{52} = 0$$

Once again, you would expect to break even in the long run.

Game 3
Here, you either win \$1,000,000 or lose \$1,000,000, so the possible values of Z are $1,000,000$ and $-1,000,000$. Since red and black cards are equally likely (26 of each) when a single card is drawn from a deck, we have $Pr[Z = 1,000,000] = Pr[Z = -1,000,000] = \frac{26}{52} = \frac{1}{2}$. Therefore we have

$$E(Z) = (1,000,000) \times \left(\frac{1}{2}\right) + (-1,000,000) \times \left(\frac{1}{2}\right) = 500,000 - 500,000 = 0$$

On this game, too, you would expect, on average, to break even.

All of the random variables in Example 3.11 have the same mean. However, these games are very different and the random variables have very different distributions. Thus we see that the mean alone doesn't actually tell us very much about a r.v.'s pdf. It's important to know where the distribution's "balance point" is, which is what the mean tells us. But it's also important to have some idea of how spread out the possible values are. For this purpose, we have a couple of other, related, statistical measures.

Definition: The **variance** of a discrete r.v. X, denoted $V(X)$, or sometimes $\sigma^2(X)$, is given by:

$$V(X) = \sigma^2(X) = E[(X - \mu)^2]$$

That is, the variance is the expected value of the squared *deviation from the mean.* To find the variance of X using this formula, we need to

1. Find the mean, μ.
2. Find the possible values of $(X - \mu)$ and then $(X - \mu)^2$.
3. Use this to find the expected value of $(X - \mu)^2$.

We also define:

Definition: The **standard deviation** of X, denoted $\sigma(X)$, is the positive square root of the variance of X. i.e. $\sigma(X) = \sqrt{V(X)}$

Notice: To find the standard deviation, we **must** first find the variance.

We can look at an example of finding the variance and standard deviation using these formulas.

Example 3.12. *Find $V(X)$ and $\sigma(X)$ for the random variable X defined in Example 3.1.*

Recall that in Example 3.1, we defined X to be the number of times Heads comes up when 3 coins are tossed. The pdf of X is

x	$Pr[X = x]$
0	1/8
1	3/8
2	3/8
3	1/8

From Example 3.9, the mean of X is $\mu = \frac{3}{2}$. We can add a couple of extra columns to the table showing the pdf, in which we put the values of $(X - \mu)$ and $(X - \mu)^2$ corresponding to each of the possible values of X. For instance, when $X = 0$, the value of $X - \mu$ is $0 - \frac{3}{2} = -\frac{3}{2}$ and so the value of $(X - \mu)^2$ is $\left(-\frac{3}{2}\right)^2 = \frac{9}{4}$.

x	$Pr[X = x]$	$x - \mu$	$(x - \mu)^2$
0	$\frac{1}{8}$	$0 - \frac{3}{2} = -\frac{3}{2}$	$\left(-\frac{3}{2}\right)^2 = \frac{9}{4}$
1	$\frac{3}{8}$	$1 - \frac{3}{2} = -\frac{1}{2}$	$\left(-\frac{1}{2}\right)^2 = \frac{1}{4}$
2	$\frac{3}{8}$	$2 - \frac{3}{2} = \frac{1}{2}$	$\left(\frac{1}{2}\right)^2 = \frac{1}{4}$
3	$\frac{1}{8}$	$3 - \frac{3}{2} = \frac{3}{2}$	$\left(\frac{3}{2}\right)^2 = \frac{9}{4}$

To get the expected value of the last column, we sum, over all possible values of X, the product of $(x - \mu)^2$ and $Pr[X = x]$. That is, we multiply the values in the second and fourth columns, and add these products. We get:

$$\begin{aligned} E[(X-\mu)^2] &= \sum\{(x-\mu)^2 \times Pr[X=x]\} \\ &= \left(\frac{9}{4}\right)\left(\frac{1}{8}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{8}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{8}\right) + \left(\frac{9}{4}\right)\left(\frac{1}{8}\right) \\ &= \frac{9}{32} + \frac{3}{32} + \frac{3}{32} + \frac{9}{32} = \frac{24}{32} = \frac{3 \times 8}{4 \times 8} = \frac{3}{4} \end{aligned}$$

That is, the variance of X is $V(X) = \frac{3}{4}$, and so the standard deviation of X is given by

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{\sqrt{4}} = \frac{\sqrt{3}}{2}$$

The formula $V(X) = E[(X - \mu)^2]$ that we used here is the actual *definition* of variance. However there is another formula which can easily be shown to be equivalent to this, but which is much easier to use in practice.

Theorem: The variance of a discrete r.v., X, can be found using the formula

$$V(X) = E(X^2) - \mu^2$$

To see how this formula is used, and that it does in fact give the same answer, we can use it to find $V(X)$ in the example we just did. This time, we just need one extra column, for the square of the possible value:

x	$Pr[X = x]$	x^2
0	1/8	$(0)^2 = 0$
1	3/8	$(1)^2 = 1$
2	3/8	$(2)^2 = 4$
3	1/8	$(3)^2 = 9$

We see that

$$\begin{aligned} E(X^2) &= \sum(x^2 \times Pr[X=x]) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} \\ &= \frac{0}{8} + \frac{3}{8} + \frac{12}{8} + \frac{9}{8} = \frac{24}{8} = 3 \end{aligned}$$

Now, we find $V(X)$ by subtracting the square of the mean. We have $\mu = \frac{3}{2}$, so $\mu^2 = \left(\frac{3}{2}\right)^2 = \frac{9}{4}$ and we get

$$V(X) = E(X^2) - \mu^2 = 3 - \frac{9}{4} = \frac{12}{4} - \frac{9}{4} = \frac{3}{4}$$

Of course, we get the same value as before, but there was less work required to get there.

Note: If we use the notation $E(X)$ instead of μ in this alternative formula for $V(X)$, we have $V(X) = E(X^2) - [E(X)]^2$. It is very important that you understand the difference between $E(X^2)$, which is the expected value of the square of X, and $[E(X)]^2$, which is the square of the expected value of X. That is, for $E(X^2)$, the squaring is done *before* taking the expected value, whereas for $[E(X)]^2$, the squaring is done *after* the expected value is calculated.

The variance and standard deviation are giving us information about how spread out the values of the random variable are. If the possible values are all fairly close together, then $V(X)$ and $\sigma(X)$ are relatively small. When the possible values of are spread over a wide range, the variance, and thus also the standard deviation, are much larger. To see this, consider again the 3 games we looked at earlier.

Example 3.13. *Find the variance and standard deviation for each of the random variables in Example 3.11 .*

This was the example in which we had X, Y and Z defined to be the winnings on 3 different games. We found the pdf for each variable. All 3 have expected value 0.

For X, we have:

x	$Pr[X = x]$	x^2
1	1/2	1
-1	1/2	1

So we get

$$E(X^2) = \left(1 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{2}\right) = 1$$

and subtracting off $[E(X)]^2 = 0^2 = 0$ we get

$$V(X) = E(X^2) - [E(X)]^2 = 1 - 0 = 1$$

so that $\sigma(X) = \sqrt{V(X)} = \sqrt{1} = 1$.

For Y, we get:

y	$Pr[Y = y]$	y^2
510	$\frac{1}{52}$	$(510)^2 = 260,100$
-10	$\frac{51}{52}$	$(10)^2 = 100$

So we have

$$E(Y^2) = \left(260,100 \times \frac{1}{52}\right) + \left(100 \times \frac{51}{52}\right) = \frac{260,100 + 5,100}{52} = \frac{265,200}{52} = 5100$$

and subtracting off $[E(Y)]^2 = 0^2 = 0$ we get $V(Y) = 5100$ and so $\sigma(Y) = \sqrt{5100} = 10\sqrt{51} \approx 71.4143$
.

Finally, for Z we have:

z	$Pr[Z=z]$	z^2
1,000,000	$\frac{1}{2}$	$(1,000,000)^2 = (10^6)^2 = 10^{12}$
$-1,000,000$	$\frac{1}{2}$	$(-1,000,000)^2 = 10^{12}$

So we get

$$E(Z^2) = \left(10^{12} \times \frac{1}{2}\right) + \left(10^{12} \times \frac{1}{2}\right) = 10^{12}$$

and subtracting off $[E(Z)]^2 = 0^2 = 0$ we get $V(Z) = 10^{12}$ and so $\sigma(Z) = \sqrt{10^{12}} = 10^6 = 1,000,000$.

Although these 3 random variables all have the same mean, the values of the variance, and therefore also the standard deviation, are very different. As expected, we see that for a r.v. like X, with possible values 1 and -1, which are very close together, the variance is small. But as the possible values get to be more spread out, the variance gets bigger.

Let's look at one more example of calculating variance.

Example 3.14. *What is the variance of the number obtained by tossing a single die?*

Let X be the number obtained when a single die is tossed. We can make a table showing the pdf's of X and X^2.

x	$Pr[X=x]$	x^2
1	1/6	1
2	1/6	4
3	1/6	9
4	1/6	16
5	1/6	25
6	1/6	36

We get

$$\begin{aligned} E(X^2) &= 1 \times \left(\frac{1}{6}\right) + 4 \times \left(\frac{1}{6}\right) + 9 \times \left(\frac{1}{6}\right) + 16 \times \left(\frac{1}{6}\right) + 25 \times \left(\frac{1}{6}\right) + 36 \times \left(\frac{1}{6}\right) \\ &= \frac{1}{6} \times (1+4+9+16+25+36) = \frac{91}{6} \end{aligned}$$

Also, in Example 3.8 we found that for this r.v., the mean is $\mu = \frac{7}{2}$, so we have

$$\begin{aligned} V(X) = E(X^2) - \mu^2 &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \\ &= \frac{91 \times 2}{6 \times 2} - \frac{49 \times 3}{4 \times 3} = \frac{182 - 147}{12} = \frac{35}{12} \end{aligned}$$

That is, the variance of the number obtained by tossing a single die is $\frac{35}{12}$.

Mean and Variance of a Binomial random variable

In Examples 3.9 and 3.12, we found the mean, variance and standard deviation of the random variable X which counts the number of times Heads comes up when a coin is tossed 3 times. Of course, as we have observed before, tossing a coin repeatedly and counting the number of times Heads comes up corresponds to performing Bernoulli trials with probability of success $p = 1/2$. Therefore, this r.v. X was counting the number of successes in $n = 3$ Bernoulli trials with $p = 1/2$, so we have $X = B(3, 1/2)$.

We used the usual formulas for calculating the mean and variance of X. Notice, though, what we found.

$$\text{For } X = B(3, 1/2), \quad E(X) = \frac{3}{2} = 3\left(\frac{1}{2}\right) \quad \text{and } V(X) = \frac{3}{4} = 3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

For $B(n, p)$ in general, especially for larger values of n, calculating the mean and variance by the usual formulas would be a lot of work. Fortunately, there is a very easy way to calculate $E(X)$ and $V(X)$ for any Binomial r.v., $X = B(n, p)$.

Let's think about what we found for $B(3, 1/2)$. When we perform 3 trials, with probability of success 1/2 on each trial, the expected number of successes is 3/2. That is, we expect that (on average) half of the 3 trials will be successes. This same reasoning works for any Binomial r.v. For instance, for $B(10, 1/2)$, we would expect that on average, 1/2 of the 10 trials will be successes, so we would "expect" 5 successes. Similarly, for $B(100, 1/4)$, we would expect that on average, 1/4 of the 100 trials will be successes, so we would "expect" to observe 25 successes. That is, it makes sense that the expected number of successes in n trials with probability of success p on each trial would be $n \times p$. This is exactly right. (Remember, though, that "expected value" doesn't actually mean we "expect" that will be the value. When I say we "would expect" to observe 25 successes, for instance, I simply mean that we expect that if we did many sets of 100 trials, the average number of successes observed would probably be close to 25.)

Theorem:
For any Binomial random variable, $X = B(n, p)$, the mean of X is $\mu = E(X) = np$.

Notice: Of course, this also means that the "expected" number of failures is $nq = n(1 - p)$.

There isn't as easy an explanation for the value of the variance of a Binomial random variable, but the formula is just as easy.

Theorem
For any Binomial random variable, $X = B(n, p)$:
the variance of X is given by $V(X) = \sigma^2(x) = npq = np(1 - p)$,
and so the standard deviation of X is $\sigma(X) = \sqrt{npq} = \sqrt{np(1 - p)}$.

Example 3.15. *A fair die is tossed 300 times. Find the mean and the standard deviation of the number of ones tossed.*

Let X be the number of times that 1 comes up when a fair die is tossed 300 times. Then, defining success to be that a 1 is tossed, X is counting the number of successes on $n = 300$ independent trials in which the probability of success is $p = 1/6$ each time. Therefore we have $X = B(300, 1/6)$. Of course, the mean of $X = B(300, 1/6)$ is given by

$$\mu = np = 300\left(\frac{1}{6}\right) = 50$$

Also, the variance of $X = B(300, 1/6)$ is given by

$$V(X) = np(1-p) = 300 \times \frac{1}{6} \times \frac{5}{6} = \frac{250}{6}$$

and so the standard deviation is

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{250}{6}} = \sqrt{25 \times \frac{10}{6}} = \sqrt{25} \times \sqrt{\frac{10}{6}} = 5 \times \sqrt{\frac{5}{3}} = \frac{5\sqrt{5}}{\sqrt{3}}$$

(or about 6.455).

Example 3.16. *In a certain manufacturing process, 1% of all items produced are defective. Find the expected number of defective items in a batch of 5000 items.*

Manufacturing 5000 items corresponds to performing 5000 trials of the same experiment. Assuming that whether or not one item is produced with a defect does not affect the probability that any other item is produced with a defect, these are independent trials. *Note:* We take the phrase "1% of all items produced are defective" to mean that *each* item, when produced, has probability .01 of being defective, *independently* of the other items produced.

Defining X to be the number of defective items in a batch of 5000 items, we see that X is counting the number of 'successes' in $n = 5000$ independent trials in which, with 'success' defined to be that an item is defective, the probability of success is $p = .01$, so we have $X = B(5000, .01)$. Of course, the expected number of defective items in the batch is the expected value of X, given by

$$E(X) = np = 5000(.01) = 50$$

Math 1228A/B Online

Lecture 24:
Functions of Random Variables

(text reference: Section 3.3, pages 129 - 133)

Functions of Random Variables

Sometimes, we are interested in a r.v. which can be expressed as a function of some other r.v., or perhaps a function of more than one r.v. For *some* particular relationships, i.e. some kinds of functions, we can find the mean of such a r.v. very easily. Next, we learn about which functions of r.v.s allow us to use easy formulas to find the mean and variance from the mean and variance of the other random variable(s).

Theorem: Consider any random variables X and Y and let a and b be any real-valued constants.

(a) If $W = aX + b$, then $E(W) = a[E(X)] + b$.

(b) If $Z = aX + bY$, then $E(Z) = a[E(X)] + b[E(Y)]$.

Also, in one of these cases, we can easily find the variance and standard deviation as well.

Theorem: Consider any random variable X and let a and b be any real-valued constants.

If $W = aX + b$ then $V(W) = a^2V(X)$ and so $\sigma(W) = |a|\sqrt{V(X)}$.

Notice:

1. The $+b$ term goes away in this case. Remember, variance is measuring how spread out the possible values are. If we add the same value, b, to each possible value, this has no effect on how spread out the values are – the distances between them remain the same, so the variance and standard deviation are not affected.

2. Recall that in the variance calculation, the values of the r.v. are squared. So if the r.v. is multiplied by a constant, during the variance calculation this constant also gets squared. Therefore when X is multiplied by a, the effect on the variance is to multiply it by a^2.

3. To get the variance, we square the a coefficient, and then to get the standard deviation, we take the square root. When we square a number and then take it's square root, the negative, if there was one, goes away, so it's $|a|$, not a, that we use to find the standard deviation.

Let's look at some examples.

Example 3.17. *I have 1 dime and 2 quarters. I toss each coin once. Let X be the number of dimes which come up Heads and Y be the number of quarters which come up Heads.*
(a) Find $E(X)$ and $E(Y)$.
(b) Find the expected value of the total number of coins which come up Heads.
(c) Find the expected value of the value of the coins which come up Heads.

(a) We have X being the number of dimes that come up Heads. There's only 1 dime being tossed, so the possible values of X are 0 and 1. Of course, these are equally likely, so we have $Pr[X = 0] = Pr[X = 1] = \frac{1}{2}$ and we get

$$E(X) = \left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{2}\right) = \frac{1}{2}$$

(That is, we have $X = B(1, 1/2)$, so $E(X) = 1 \times 1/2 = 1/2$. Thinking about performing just 1 "independent trial", i.e. $B(1, p)$, is a little silly, but the formula still works.)

Also, Y is the number of quarters which come up Heads. Since there are 2 quarters being tossed, Y has possible values 0, 1 and 2. Using the equiprobable sample space $S = \{HH, HT, TH, TT\}$, we see that $Pr[Y = 0] = \frac{1}{4}$, $Pr[Y = 1] = \frac{2}{4} = \frac{1}{2}$ and $Pr[Y = 2] = \frac{1}{4}$. So we have

$$E(Y) = \left(0 \times \frac{1}{4}\right) + \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{4}\right) = 0 + \frac{1}{2} + \frac{1}{2} = 1$$

(That is, we have $Y = B(2, 1/2)$, so $E(Y) = 2 \times 1/2 = 1$.)

(b) We need to find the expected value of the total number of coins which come up Heads. Let W be the r.v. whose value is the number of coins which come up Heads. Then W is given by adding the number of dimes which came up Heads and the number of quarters that come up Heads, so we have $W = X + Y$. This has the form $W = aX + bY$, with $a = 1$ and $b = 1$, so we can use the formula $E(W) = a[E(X)] + b[E(Y)]$ to get

$$E(W) = (1 \times E(X)) + (1 \times E(Y)) = E(X) + E(Y) = \frac{1}{2} + 1 = \frac{3}{2}$$

(Of course, W is just counting the number of Heads which come up when 1 dime and 2 quarters, i.e. 3 coins, are tossed, so we have $W = B(3, 1/2)$ and $E(W) = 3 \times 1/2 = \frac{3}{2}$.)

(c) We need to find the expected value of the *value* of the coins which come up Heads. We know that X is the number of dimes which come up Heads (0 or 1) and Y is the number of quarters that come up Heads (0, 1 or 2). Of course, the value of each dime is 10 cents and the value of each quarter is 25 cents. Therefore if we let Z be the value (in cents) of the coins which come up Heads, we have $Z = 10X + 25Y$, and we see that the expected value of Z is

$$E(Z) = E(10X + 25Y) = 10[E(X)] + 25[E(Y)] = 10(1/2) + 25(1) = 30$$

i.e. the expected value of the value of the coins which come up Heads is 30 cents.

(*Notice:* Although X and Y are Binomial random variables, Z is *not* a Binomial random variable. That is, Z is *not* counting successes in a series of Bernoulli trials.)

Example 3.18. *George is a widget salesman. George earns a salary of \$500 per week, plus a commission of \$300 for each widget he sells during the week. George never sells more than 4 widgets in a week. The probability that George sells n widgets in any week is given by the formula* $\frac{5-n}{15}$*, for any integer value of n from 0 to 4. Find the mean and standard deviation of George's weekly earnings.*

Because George earns commission, in addition to a base salary, his weekly earnings change from week to week, depending on how many widgets he sells that week. If we let X be the number of widgets George sells in a particular week, then we can define Y to be George's weekly earnings (in dollars) and we have

$$Y = 300X + 500$$

We are asked to find $E(Y)$ and $\sigma(Y)$. Of course, since we see that Y has the form $Y = aX + b$, with $a = 300$ and $b = 500$, we know that

$$E(Y) = E(aX + b) = a \times E(X) + b = 300E(X) + 500$$

and also that

$$\sigma(Y) = \sigma(aX + b) = |a|\sigma(X) = 300\sigma(X)$$

Therefore we need to find $E(X)$ and $\sigma(X)$. (We could find the pdf of Y and find $E(Y)$ and $\sigma(Y)$ directly, but the calculations for $E(X)$ and $\sigma(X)$ are easier, i.e. involve easier numbers.)

We are told that the probability that George sells n widgets in any week is given by $\frac{5-n}{15}$. Also, we know that George never sells more than 4 widgets in one week, and of course the number of widgets sold has to be a non-negative integer, so (as stated in the question) the possible values of X are 0, 1, 2, 3 and 4. For each of these possible values, we have $Pr[X = x] = \frac{5-x}{15}$. We use this to find the pdf of X.

x	$Pr[X = x]$
0	$\frac{5-0}{15} = \frac{5}{15}$
1	$\frac{5-1}{15} = \frac{4}{15}$
2	$\frac{5-2}{15} = \frac{3}{15}$
3	$\frac{5-3}{15} = \frac{2}{15}$
4	$\frac{5-4}{15} = \frac{1}{15}$

From this, we see that the expected number of widgets George sells in a week is

$$\begin{aligned} E(X) &= \sum_{x=0}^{4} (x \times Pr[X = x]) \\ &= \left(0 \times \frac{5}{15}\right) + \left(1 \times \frac{4}{15}\right) + \left(2 \times \frac{3}{15}\right) + \left(3 \times \frac{2}{15}\right) + \left(4 \times \frac{1}{15}\right) \\ &= \frac{0+4+6+6+4}{15} = \frac{20}{15} = \frac{4}{3} \end{aligned}$$

Therefore we see that the amount of George's expected weekly earnings is given by

$$E(Y) = 300E(X) + 500 = 300 \times \frac{4}{3} + 500 = 400 + 500 = 900$$

That is, the mean of George's weekly earnings is about \$900.

Of course, to find the standard deviation of X we must first find the variance, $V(X)$. We know that $V(X) = E(X^2) - [E(X)]^2$, where

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{4} \left(x^2 \times Pr[X = x]\right) \\ &= \left(0^2 \times \frac{5}{15}\right) + \left(1^2 \times \frac{4}{15}\right) + \left(2^2 \times \frac{3}{15}\right) + \left(3^2 \times \frac{2}{15}\right) + \left(4^2 \times \frac{1}{15}\right) \\ &= \frac{0+4+12+18+16}{15} = \frac{50}{15} = \frac{10}{3} \end{aligned}$$

Thus we get $V(X) = \frac{10}{3} - \left(\frac{4}{3}\right)^2 = \frac{10}{3} - \frac{16}{9} = \frac{30}{9} - \frac{16}{9} = \frac{14}{9}$ and so $\sigma(X) = \sqrt{\frac{14}{9}} = \frac{\sqrt{14}}{3}$. This gives

$$\sigma(Y) = 300\sigma(X) = 300\left(\frac{\sqrt{14}}{3}\right) = 100\sqrt{14} \approx 374.17$$

We see that the standard deviation of George's weekly earnings is about \$374.17.

Example 3.19. *X is a discrete random variable with $E(X) = 5$ and $V(X) = 16$. Random variable Y is related to X according to the relationship $Y = 1 - \frac{3}{2}X$.*
(a) Find $E(Y)$ and $\sigma(Y)$.
(b) Find $V(X+Y)$.
(c) Find $E(XY)$.

(a) We have $Y = 1 - \frac{3}{2}X = \left(-\frac{3}{2}\right)X + 1$ which has the form $Y = aX + b$, with $a = -\frac{3}{2}$ and $b = 1$. Using $E(aX+b) = aE(X) + b$, we see that

$$E(Y) = E\left(-\frac{3}{2}X + 1\right) = -\frac{3}{2}[E(X)] + 1 = \left(-\frac{3}{2}\right)(5) + 1 = -\frac{15}{2} + 1 = -\frac{13}{2} = -6.5$$

Also, using $\sigma(aX+b) = |a|\sigma(X)$ we get

$$\sigma(Y) = \sigma\left(-\frac{3}{2}X + 1\right) = \left|-\frac{3}{2}\right| \times \sigma(X) = \frac{3}{2} \times \sqrt{16} = \frac{3}{2} \times 4 = 6$$

(b) We want to find $V(X+Y)$. *Notice:* We don't have a formula telling us about the value of $V(X+Y)$. That is, we have a formula for $V(aX+b)$, but we *don't* have one for $V(aX+bY)$. However, since we have a formula relating Y to X, in this case we have

$$X + Y = X + \left(1 - \frac{3}{2}X\right) = X - \frac{3}{2}X + 1 = -\frac{1}{2}X + 1$$

which has the form $aX+b$, with $a = -\frac{1}{2}$ and $b = 1$, so using $V(aX+b) = a^2V(X)$ we get

$$V(X+Y) = V\left(-\frac{1}{2}X + 1\right) = \left(-\frac{1}{2}\right)^2 V(X) = \frac{1}{4} \times 16 = 4$$

Notice: If we take $V(X) + V(Y)$ we get $16 + (6)^2 = 16 + 36 = 52$ which is *not* the value of $V(X+Y)$. Although the *expected value* of the sum of 2 r.v.'s is given by the sum of the expected values, the same idea does *not* usually work for *variance*. Here, because of the interaction beween X and Y, the variance of $X+Y$ is much less than the sum of the variances of these random variables.

(c) Now, we want to find $E(XY)$. We don't have a formula for that, but once again, in this case we can re-express the product XY using the relationship between X and Y. We have

$$XY = X \times \left(1 - \frac{3}{2}X\right) = X - \frac{3}{2}(X^2)$$

If we let W be a random variable such that $W = X^2$, then we have $XY = X - \frac{3}{2}W$, which has the form $aX + bW$, so we see that

$$E(XY) = E\left(X - \frac{3}{2}W\right) = E(X) + \left(-\frac{3}{2}\right)E(W)$$

How can we find the value of $E(W) = E(X^2)$? Remember, we know that

$$V(X) = E(X^2) - [E(X)]^2$$

But we know that $V(X) = 16$ and $E(X) = 5$, so we get

$$16 = E(X^2) - (5)^2 \Rightarrow E(X^2) = 16 + 5^2 = 16 + 25 = 41$$

so we have $E(W) = E(X^2) = 41$. Plugging this into the formula we found for $E(XY)$, we get

$$E(XY) = E(X) + \left(-\frac{3}{2}\right)E(W) = 5 - \frac{3}{2} \times 41 = 5 - \frac{123}{2} = -\frac{113}{2}$$

Notice: $E(X) \times E(Y) = 5 \times -\frac{13}{2} = -\frac{65}{2} \neq E(XY)$. Although we can find the expected value of the *sum* of 2 r.v.'s by summing their individual expected values, this same idea does *not* generally work for the *product* of the r.v.'s.

Example 3.20. *Whenever a certain experiment is performed, exactly 1 of 3 possible outcomes, t_1, t_2, or t_3, is observed. X and Y are 2 discrete random variables defined for this experiment. When outcome t_1 occurs, X has the value 1 and Y has the value -1. When outcome t_2 occurs, X has the value 2 and Y has the value 0. When outcome t_3 occurs, X has the value 5 and Y has the value 1. Outcome t_1 occurs $\frac{1}{3}$ of the time and outcome t_2 occurs $\frac{1}{2}$ the time. Find $\sigma(X+Y)$.*

We are told that $S = \{t_1, t_2, t_3\}$ is a sample space for the experiment, so it must be true that $Pr[t_1] + Pr[t_2] + Pr[t_3] = 1$. And we know that $Pr[t_1] = \frac{1}{3}$ and $Pr[t_2] = \frac{1}{2}$, so we also have

$$Pr[t_3] = 1 - \left(\frac{1}{3} + \frac{1}{2}\right) = 1 - \left(\frac{2}{6} + \frac{3}{6}\right) = 1 - \frac{5}{6} = \frac{1}{6}$$

We can tabulate the pdf's of X and Y as shown here:

outcome	probability	x	y
t_1	$\frac{1}{3}$	1	-1
t_2	$\frac{1}{2}$	2	0
t_3	$\frac{1}{6}$	5	1

We're asked to find $\sigma(X+Y)$, that is, the standard deviation of the random variable whose value is given by $X+Y$. Let $W = X+Y$. For any particular outcome of the experiment, we can find the corresponding value of W by summing the values of X and Y. In this way we can find the pdf of the r.v. W – we can add an extra column to the table to show the value of W:

outcome	probability	x	y	$w = x + y$
t_1	$\frac{1}{3}$	1	-1	0
t_2	$\frac{1}{2}$	2	0	2
t_3	$\frac{1}{6}$	5	1	6

We see that $E(W) = \left(0 \times \frac{1}{3}\right) + \left(2 \times \frac{1}{2}\right) + \left(6 \times \frac{1}{6}\right) = 0 + 1 + 1 = 2$.

(*Notice:* $E(X) = \left(1 \times \frac{1}{3}\right) + \left(2 \times \frac{1}{2}\right) + \left(5 \times \frac{1}{6}\right) = \frac{1}{3} + 1 + \frac{5}{6} = \frac{2}{6} + \frac{6}{6} + \frac{5}{6} = \frac{13}{6}$ and $E(Y) = \left((-1) \times \frac{1}{3}\right) + \left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{6}\right) = -\frac{1}{3} + 0 + \frac{1}{6} = -\frac{2}{6} + \frac{1}{6} = -\frac{1}{6}$, so $E(X) + E(Y) = \frac{13}{6} + \left(-\frac{1}{6}\right) = \frac{13}{6} - \frac{1}{6} = \frac{12}{6} = 2$. As always, $E(X+Y) = E(X) + E(Y)$.)

Also, $E(W^2) = \left(0^2 \times \frac{1}{3}\right) + \left(2^2 \times \frac{1}{2}\right) + \left(6^2 \times \frac{1}{6}\right) = 0 + 2 + 6 = 8$, so we get

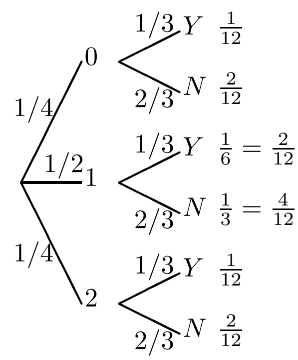
$$V(W) = E(W^2) - [E(W)]^2 = 8 - 2^2 = 8 - 4 = 4$$

Therefore $\sigma(X+Y) = \sigma(W) = \sqrt{V(W)} = \sqrt{4} = 2$.

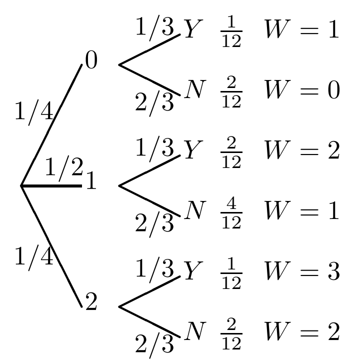
(*Notice:* We can calculate $V(X)$ and $V(Y)$. We get $E(X^2) = \frac{13}{2}$, so that $V(X) = \frac{65}{36}$ and $E(Y^2) = \frac{1}{2}$, so that $V(Y) = \frac{17}{36}$. Therefore $V(X) + V(Y) = \frac{82}{36} = \frac{41}{18}$. As before, we see that $V(X+Y) \neq V(X) + V(Y)$.)

Example 3.21. *A certain game involves tossing 2 coins and then tossing 1 die. You receive \$1 for each coin that comes up Heads, and you receive (another) \$1 if a 5 or a 6 comes up on the die. Find the expected value and the variance of the amount you win on this game.*

Let W be the amount you win playing the game. A tree will be helpful in finding the distribution of W. To make the tree as small as possible, we'll use a tree with only 2 levels. The first level represents whether 0, 1 or 2 of the coins came up Heads, and the second level represents whether a 5 or a 6 was tossed (Y, for yes) or not (N, for no). When 2 coins are tossed, the probabilities of getting 0 or 2 Heads are both $\frac{1}{4}$ and the probability of getting 1 Heads is $\frac{1}{2}$. (That is, the number of coins which come up Heads is $X = B(2, 1/2)$, and the Bernoulli formula gives the probabilities mentioned.) And when a single die is tossed, the probability of getting a 5 or a 6 is $Pr[Y] = \frac{2}{6} = \frac{1}{3}$ and the probability of getting some other number is $Pr[N] = \frac{4}{6} = \frac{2}{3}$. (These probabilities apply no matter how many of the coins came up Heads.) The tree is:



The path probabilities are shown at the ends of the paths. We can also put the value of W corresponding to each path on the tree. For instance, if 0 coins come up Heads and Y occurse on the die toss, you receive \$0 for the coins, plus \$1 for the die toss, for a total of \$1, so $W = 1$. Likewise, if 1 coin comes up Heads and Y occurs on the die toss, you receive \$1 for the coins and \$1 for the die toss, so $W = 2$. We have:



We see that the pdf of W is:

w	$Pr[W = w]$
0	$\frac{2}{12}$
1	$\frac{1}{12} + \frac{4}{12} = \frac{5}{12}$
2	$\frac{2}{12} + \frac{2}{12} = \frac{4}{12}$
3	$\frac{1}{12}$

Now, we can find $E(W)$ as:

$$E(W) = \left(0 \times \frac{2}{12}\right) + \left(1 \times \frac{5}{12}\right) + \left(2 \times \frac{4}{12}\right) + \left(3 \times \frac{1}{12}\right) = 0 + \frac{5}{12} + \frac{8}{12} + \frac{3}{12} = \frac{5+8+3}{12} = \frac{16}{12} = \frac{4}{3}$$

and $E(W^2)$ as:

$$\begin{aligned} E(W^2) &= \left(0^2 \times \frac{2}{12}\right) + \left(1^2 \times \frac{5}{12}\right) + \left(2^2 \times \frac{4}{12}\right) + \left(3^2 \times \frac{1}{12}\right) = 0 + \frac{5}{12} + \left(4 \times \frac{4}{12}\right) + \left(9 \times \frac{1}{12}\right) \\ &= \frac{5+16+9}{12} = \frac{30}{12} = \frac{5}{2} \end{aligned}$$

so that:

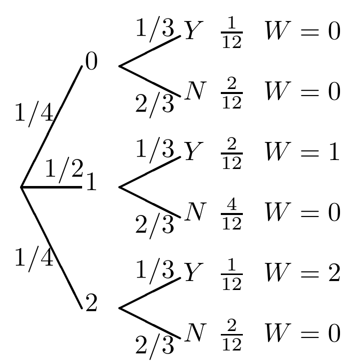
$$V(W) = E(W^2) - [E(W)]^2 = \frac{5}{2} - \left(\frac{4}{3}\right)^2 = \frac{5}{2} - \frac{16}{9} = \frac{5 \times 9}{2 \times 9} - \frac{16 \times 2}{9 \times 2} = \frac{45}{18} - \frac{32}{18} = \frac{13}{18}$$

That is, the expected value of your winnings is $\$\frac{4}{3}$ and the variance is $\frac{13}{18}$ (in $\$^2$).

(*Notice:* We have $X = B(2, 1/2)$, so $V(X) = npq = 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$. Also, we can think of the die toss as $Z = B(1, 1/3)$, with $V(Z) = 1 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$. But then $V(X) + V(Z) = \frac{1}{2} + \frac{2}{9} = \frac{9+4}{18} = \frac{13}{18}$. Oh! But $W = X + Z$ (i.e. the amount you win on the coins plus the amount you win from the die). So here we have $V(W) = V(X + Z) = V(X) + V(Z)$. Hmm. That's not usually true! But it is in this case.)

Example 3.22. *Suppose you play the game in the previous example again, but this time you only win \$1 for each coin which comes up Heads, and* **only** *if a 5 or a 6 comes up on the die toss. Find the expected value and variance of the amount you win this time.*

We have the same tree as before, but the values of W change. For instance, if 0 coins compu up Heads, you win nothing, no matter what comes up on the die. (That is, if a number other than 5 or 6 comes up, you don't win, and if a 5 or a 6 comes up, you *do* win, but the amount you win is \$0.) If 1 coin comes up Heads, you win \$1 if a 5 or a 6 comes up, but you don't win, i.e. win nothing, if any other number comes up. And similarly for 2 Heads: you win \$2 if a 5 or a 6 comes up, \$0 otherwise. We have:



This time, the pdf of W is:

w	$Pr[W = w]$
0	$\frac{1}{12} + \frac{2}{12} + \frac{4}{12} + \frac{2}{12} = \frac{9}{12}$
1	$\frac{2}{12}$
2	$\frac{1}{12}$

so $E(W) = \left(0 \times \frac{9}{12}\right) + \left(1 \times \frac{2}{12}\right) + \left(2 \times \frac{1}{12}\right) = 0 + \frac{2}{12} + \frac{2}{12} = \frac{4}{12} = \frac{1}{3}$
and $E(W^2) = \left(0^2 \times \frac{9}{12}\right) + \left(1^2 \times \frac{2}{12}\right) + \left(2^2 \times \frac{1}{12}\right) = 0 + \frac{2}{12} + \frac{4}{12} = \frac{6}{12} = \frac{1}{2}$
which gives $V(W) = \frac{1}{2} - \left(\frac{1}{3}\right)^2 = \frac{1}{2} - \frac{1}{9} = \frac{9}{18} - \frac{2}{18} = \frac{7}{18}$.

That is, this time your expected winnings is \$$\frac{1}{3}$, with variance $\frac{7}{18}$.

(*Notice:* In terms of $X = B(2, 1/2)$ which counts the number of coins which come up Heads and $Z = B(1, 1/3)$ which counts the number of dice which come up 5 or 6, this time your winnings can be expressed as $W = XZ$. And $E(X) = 2 \times \frac{1}{2} = 1$, while $E(Z) = 1 \times \frac{1}{3} = \frac{1}{3}$, so that $E(X) \times E(Z) = 1 \times \frac{1}{3} = \frac{1}{3}$. But ... that's the same as $E(W)$! That is, this time we have $E(XZ) = E(X) \times E(Z)$, whereas we know that's not usually true. Again, hmm.)

Math 1228A/B Online

Lecture 25:

Independence of rv's and
The Joint Distribution of X and Y

(text reference: Section 3.4, pages 136 - 139)

3.4 Independent Random Variables

Example 3.23. *Recall Example 3.17. Find $V(X)$, $V(Y)$ and $V(X+Y)$.*

In Example 3.17 (see Lecture 24, pg. 138), we have 1 dime and 2 quarters and are tossing each coin once. X and Y are defined to be the number of dimes and quarters, respectively, which come up Heads. As we observed before, both X and Y are binomial random variables, and we know that the variance of $B(n,p)$ is given by $np(1-p)$. Since $X = B\left(1, \frac{1}{2}\right)$ and $Y = B\left(2, \frac{1}{2}\right)$, we have $V(X) = 1 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = \frac{1}{4}$ and also $V(Y) = 2 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = \frac{1}{2}$. Of course, as we saw before, $X+Y$ is simply counting the total number of Heads which come up on the 3 coins, so $X + Y = B\left(3, \frac{1}{2}\right)$ and we see that $V(X+Y) = 3 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = \frac{3}{4}$.

Notice: In this case, we have $V(X+Y) = V(X) + V(Y)$. As we saw in the last section, this relationship is not generally true. And the same thing was true in Example 3.21 (see Lecture 24, pg. 143), but there the r.v.'s were called X and Z. Is there something about the r.v.'s in these 2 examples that makes this relationship hold? Or is it just a fluke?

In fact, it's not a fluke. The random variables X and Y in Example 3.23 (above) have a very special relationshiip – that they are totally unrelated. That is, what happens when we toss a dime has no effect on the outcome of tossing 2 quarters, so the value of X observed and the value of Y observed are independent of one another. (Likewise, in Example 3.21, the die toss is not affected by the coin tosses.) We have a name for this kind of relationship between 2 r.v.'s.

> *Definition:* Two random variables, X and Y, are called **independent r.v.'s** if
>
> $$Pr[(X = x) \cap (Y = y)] = Pr[X = x] \times Pr[Y = y]$$
>
> for every combination of a possible value x and a possible value y. That is, X and Y are independent r.v.'s if the events $(X = x)$ and $(Y = y)$ are independent events, for each possible combination of possible values x and y.

To determine whether 2 variables X and Y are independent, we need to determine whether or not this relationship holds. We do this by looking at the **joint distribution** of X and Y.

> *Definition:*
> For any discrete random variables X and Y, the **joint distribution of X and Y** is the function of $Pr[(X = x) \cap (Y = y)]$ values for all combinations of a possible value x and a possible value y.

This means that if we want to know whether or not 2 random variables, X and Y, are independent r.v.'s, we need to find the joint distribution of X and Y, i.e. the values of $Pr[(X = x) \cap (Y = y)]$ and determine whether or not, in each case, this is equal to $Pr[X = x] \times Pr[Y = y]$.

We can make a table showing the joint distribution of X and Y. Each row corresponds to a possible value of X, and each column corresponds to a possible value of Y. The table entry in the row for $(X = x)$ and the column for $(Y = y)$ is the value of $Pr[(X = x) \cap (Y = y)]$.

If we add up the entries in the $(X = x)$ row, we get $Pr[X = x]$, because all we're really doing is partitioning the event $(X = x)$ into the various possible values for Y. Likewise, we've partitioned $(Y = y)$ according to the value of X, so if we add the entries in the $(Y = y)$ column, we get $Pr[Y = y]$.

We can add an extra column for the row sums, and an extra row for the column sums, to also show the values of $Pr[X = x]$ and $Pr[Y = y]$ in the table. These are referred to as **marginal probabilities**, because we write them in the margins of the joint distribution table. Putting the marginal probabilities in the table makes it easy to check whether or not it is true that $Pr[(X = x) \cap (Y = y)]$ is equal to $Pr[X = x] \times Pr[Y = y]$ for each combination.

This will all be easier to understand looking at an example.

Example 3.24. *Recall Example 3.17. Show that X and Y are independent random variables.*

We need to find the joint distribution of X, the number of times that Heads is observed when one dime is tossed, and Y, the number of times Heads is observed when 2 quarters are tossed. We can describe the possible outcomes of the experiment 'toss 1 dime and 2 quarters' as sequences of H's and T's, where each outcome is a sequence of 3 letters, with the first denoting the dime and the next 2 denoting the first and second quarters, respectively. For instance, the outcome HTT represents that the dime came up Heads while both quarters came up Tails. Similarly, THT denotes that the dime was Tails, the first quarter was Heads and the second quarter was Tails. An equiprobable sample space for the experiment is

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

We use this sample space to find the joint distribution of X and Y, i.e. to find the values of $Pr[(X = x) \cap (Y = y)]$. Since X has 2 possible values, 0 and 1, and Y has 3 possible values, 0, 1 and 2, there are 6 such values we need to find. For each, we must think about what the event $(X = x) \cap (Y = y)$ is, and how many of the sample points from the equiprobable sample space each contains. (Note that we have $n(S) = 8$.)

The event $(X = 0)$ occurs when none of the dimes come up Heads, i.e. when the dime comes up Tails. Similarly, the event $(Y = 0)$ occurs when 0 of the quarters come up Heads, i.e. when both quarters come up Tails. So the event $(X = 0) \cap (Y = 0)$ corresponds to the dime and also both quarters coming up Tails. This happens only when outcome TTT is observed, so we see that

$$Pr[(X = 0) \cap (Y = 0)] = \frac{1}{8}$$

The event $(X = 0) \cap (Y = 1)$, of course, occurs when 0 of the dimes and 1 of the quarters comes up Heads; that is, the dime comes up Tails, while one quarter comes up Heads and the other comes up Tails. This corresponds to 2 sample points, THT and TTH, so we see that

$$Pr[(X = 0) \cap (Y = 1)] = \frac{2}{8} = \frac{1}{4}$$

We use similar reasoning to find all of the joint distribution values, i.e. to find the probabilities of all combinations of events: $(X = 0) \cap (Y = 0)$, $(X = 0) \cap (Y = 1)$, $(X = 0) \cap (Y = 2)$, $(X = 1) \cap (Y = 0)$, $(X = 1) \cap (Y = 1)$ and $(X = 1) \cap (Y = 2)$. We make a table showing the joint distribution, as described earlier. We have a row for each possible value of X, i.e. rows for events $(X = 0)$ and $(X = 1)$. Likewise, we have a column for each possible value of Y, i.e. columns for $(Y = 0)$, $(Y = 1)$ and $(Y = 2)$.

	$(Y = 0)$	$(Y = 1)$	$(Y = 2)$
$(X = 0)$	1/8	1/4	1/8
$(X = 1)$	1/8	1/4	1/8

Notice: If we add up all of the table entries, they of course sum to 1.

Adding up each row and also each column gives the marginal probabilities, which we put into the table (in the margins).

	$(Y=0)$	$(Y=1)$	$(Y=2)$	$Pr[X=x]$
$(X=0)$	1/8	1/4	1/8	1/2
$(X=1)$	1/8	1/4	1/8	1/2
$Pr[Y=y]$	1/4	1/2	1/4	

Now, we check whether or not it is true in *every* case that the entry in the main body of the table is equal to the product of the marginal entries in the corresponding row and column. In this case, it *is* true. For instance, we see that we have $Pr[(X=0)\cap(Y=0)]=\frac{1}{8}$, while (from the right margin) $Pr[X=0]=\frac{1}{2}$ and (from the lower margin) $Pr[Y=0]=\frac{1}{4}$. This gives

$$Pr[X=0]\times Pr[Y=0]=\frac{1}{2}\times\frac{1}{4}=\frac{1}{8}=Pr[(X=0)\cap(Y=0)]$$

Likewise, $Pr[(X=1)\cap(Y=1)]=\frac{1}{4}$, and we see that $Pr[X=1]=\frac{1}{2}$ and $Pr[Y=1]=\frac{1}{2}$, so that

$$Pr[X=1]\times Pr[Y=1]=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}=Pr[(X=1)\cap(Y=1)]$$

Note: We must check *every* row and column in this way.

Here, we see that $Pr[(X=x)\cap(Y=y)]$ *does* equal $Pr[X=x]\times Pr[Y=y]$ for every single combination of a value of X and a value of Y. Therefore, according to our definition, X and Y are independent random variables.

Let's look at another example in which we need to determine whether or not 2 random variables are independent.

Example 3.25. *A single die is tossed. X has the value 1 if the number showing on the die is odd, and 2 if the number showing is even. Y is defined to be 0 if the die shows 1 or 2 or 3, and has the value 1 if 4 or 5 or 6 comes up. Are X and Y independent random variables?*

Once again, we need to find the joint distribution of X and Y, and investigate whether or not it is always true that $Pr[(X=x)\cap(Y=y)]$ is equal to $Pr[X=x]\times Pr[Y=y]$. To find the joint distribution of X and Y in this case, we consider each possible outcome of tossing a single die and determine what the values of X and Y are in each case:

outcome	1	2	3	4	5	6
probability	1/6	1/6	1/6	1/6	1/6	1/6
Value of X	1	2	1	2	1	2
Value of Y	0	0	0	1	1	1

Now, we find $Pr[(X=x)\cap(Y=y)]$ for each combination by looking at which (how many) outcomes have that combination. We see, for instance, that both the outcome 1 and the outcome 3 have $X=1$ and $Y=0$, i.e. these 2 outcomes (and only these) are in the event $(X=1)\cap(Y=0)$, so $Pr[(X=1)\cap(Y=0)]=\frac{2}{6}=\frac{1}{3}$. Likewise, 2 is the only outcome with $X=2$ and $Y=0$, so $Pr[(X=2)\cap(Y=0)]=\frac{1}{6}$. In a similar manner, we find that $Pr[(X=1)\cap(Y=1)]=\frac{1}{6}$ and also that $Pr[(X=2)\cap(Y=1)]=\frac{1}{3}$. We tabulate these values and find the marginal probabilities:

	$(Y=0)$	$(Y=1)$	$Pr[X=x]$
$(X=1)$	1/3	1/6	1/2
$(X=2)$	1/6	1/3	1/2
$Pr[Y=y]$	1/2	1/2	

We can easily see from this table that

$$Pr[(X = x) \cap (Y = y)] \neq Pr[X = x] \times Pr[Y = y]$$

For instance, $Pr[(X = 1) \cap (Y = 0)] = 1/3$, whereas we have $Pr[X = 1] \times Pr[Y = 0] = (1/2) \times (1/2) = 1/4$. Therefore X and Y are *not* independent random variables.

Notice: As soon as we find *one* combination of a possible value x and a possible value y for which it is not true that $Pr[(X = x) \cap (Y = y)] = Pr[X = x] \times Pr[Y = y]$, we can immediately conclude that X and Y are *not* independent random variables.

Example 3.26. *Recall Example 3.20. Find the joint distribution table for X and Y. Are these independent random variables?*

This was the example in which there were 3 possible outcomes t_1, t_2 and t_3 for some experiment, with specific values of X and Y defined for each of these outcomes (see Lecture 24, pg. 142). We had:

outcome	probability	x	y
t_1	$\frac{1}{3}$	1	-1
t_2	$\frac{1}{2}$	2	0
t_3	$\frac{1}{6}$	5	1

The possible values of X are 1, 2 and 5, so our joint distribution table will have rows for $(X = 1)$, $(X = 2)$ and $(X = 5)$. Similarly, the possible values of Y are -1, 0 and 1, so the joint distribution table has columns for $(Y = -1)$, $(Y = 0)$ and $(Y = 1)$. This gives a table with 9 cells:

	$(Y = -1)$	$(Y = 0)$	$(Y = 1)$	$Pr[X = x]$
$(X = 1)$				
$(X = 2)$				
$(X = 5)$				
$Pr[Y = y]$				

How do we find the values to put into the table? Easy. When outcome t_1 occurs, $X = 1$ and $Y = -1$, and this is the only time this combination of values occurs, so $Pr[(X = 1) \cap (Y = -1)] = Pr[t_1] = \frac{1}{3}$. Similarly, when t_2 occurs we have $X = 2$ and $Y = 0$, a combination which doesn't happen any other time, so $Pr[(X = 2) \cap (Y = 0)] = Pr[t_2] = \frac{1}{2}$. And when t_3 occurs, $X = 5$ and $Y = 1$, which doesn't happen under any other circumstance, so $Pr[(X = 5) \cap (Y = 1)] = Pr[t_3] = \frac{1}{6}$. So far we have:

	$(Y = -1)$	$(Y = 0)$	$(Y = 1)$	$Pr[X = x]$
$(X = 1)$	1/3			
$(X = 2)$		1/2		
$(X = 5)$			1/6	
$Pr[Y = y]$				

What about the other cells? What other values do we put in the table? For instance, what is $Pr[(X = 1) \cap (Y = 0)]$? Well, when does this happen? It never happens. We only have $X = 1$ when t_1 occurs, and when that happens, $Y \neq 0$. So $(X = 1)$ and $(Y = 0)$ never occur together. Having them occur together is an *impossible* event, so $Pr[(X = 1) \cap (Y = 0)] = 0$. And the same is true for all the other combinations we haven't yet filled in values for.

Notice: In the joint distribution table, the sum of all the entries in (the main body of) the table is 1. And the probabilities we already filled in sum to 1, so all the other table entries must be 0.

Also Notice: We already know the pdf's of X and Y, from before. And the probability values in the pdf are the marginal probabilities, i.e. the row sums and the column sums. For instance, we previously use the fact that $Pr[X = 1] = \frac{1}{3}$ (when we calculated $E(X)$, before). But we already have a $\frac{1}{3}$ in the $(X = 1)$ row, so for the row sum to be only $\frac{1}{3}$, it must be the case that all other entries in this row are 0. A similar observation holds for each of the other rows, and for each of the columns.

The completed joint distribution table, with marginal probabilities, is:

	$(Y = -1)$	$(Y = 0)$	$(Y = 1)$	$Pr[X = x]$
$(X = 1)$	1/3	0	0	1/3
$(X = 2)$	0	1/2	0	1/2
$(X = 5)$	0	0	1/6	1/6
$Pr[Y = y]$	1/3	1/2	1/6	

We can easily see that $Pr[(X = x) \cap (Y = y)] \neq Pr[X = x] \times Pr[Y = y]$, for *any* of the possible values x and y. For instance, we have $Pr[(X = 1) \cap (Y = 0)] = 0 \neq \frac{1}{3} \times \frac{1}{2}$. Likewise, we see that $Pr[(X = 1) \cap (Y = -1)] = \frac{1}{3} \neq \frac{1}{3} \times \frac{1}{3}$. In this table, the product of the row sum and the column sum *never* gives the cell entry, so X and Y are not independent rv's, they are dependent. (But remember, we don't need this to never be true, we only need to find *any* instance in which it is not true in order to conclude that the random variables are not independent.)

Tip: Whenever X and Y are both defined by the same outcome, in such a way that only certain combinations of an X-value and a Y-value occur together, X and Y are dependent r.v.'s. Likewise, in *any* situation in which some particular possible value of X and some particular possible value of Y can *never* occur together, X and Y are dependent r.v.'s.

Math 1228A/B Online

Lecture 26:

Formulas Which Hold **ONLY** When X and Y are Independent rv's

(text reference: Section 3.4, pages 139 - 141)

If we know that 2 random variables are independent r.v.'s, then there are certain functions of the r.v.'s for which nice properties hold:

> **Theorem 1:** If X and Y are independent random variables, then
> $$V(X+Y) = V(X) + V(Y)$$
> and $E(XY) = E(X) \times E(Y)$.

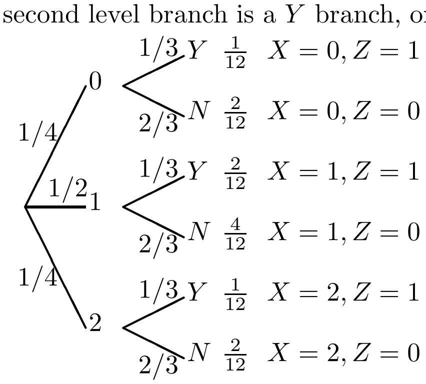
Recall: These relationships are *not* true in general. We can **ONLY** use these formulas when we **KNOW** that X and Y are independent r.v.'s.

Notice: In Example 3.23 (see Lecture 25, pg. 146) we saw that for the random variables X and Y in Example 3.17 (see Lecture 24, pg. 138) we get $V(X+Y) = V(X) + V(Y)$. And in Example 3.24 (see Lecture 25, pg. 147), we found that in this case X and Y are independent random variables. That's *why* it turns out that $V(X+Y) = V(X) + V(Y)$.

Example 3.27. *Recall Example 3.21 in which you play a game where 2 coins and 1 die are tossed. You receive \$1 for each coin that comes up Heads and \$1 if the die shows a 5 or a 6. If X is the number of coins which come up Heads and Z is the number of dice which show 5 or 6 (i.e. is 1 if a 5 or a 6 comes up and is 0 if any other number comes up), are X and Z independent random variables?*

When we did Example 3.21 (see Lecture 24, pg. 143), we observed that for X and Z as defined here, we get $V(X+Z) = V(X)+V(Z)$. However, knowing that this relationship (which is known to hold for independent r.v.'s) holds in this case does **not** necessarily mean that these r.v.'s are independent. We don't know that the relationship **only** holds for independnet r.v.'s. Maybe there could be 2 r.v.'s X and Z which are *not* independent, but for which (because of a fluke) $V(X+Y)$ happens to be equal to $V(X)+V(Y)$. So we need another way to determine whether X and Z are independent.

Approach 1: Use the tree from Example 3.21
We can put X and Z values at the ends of the paths through the probability tree we used when we did Example 3.21. The value of X is just the outcome (0, 1, or 2) on the first level branch in the path. The value of Z is 1 if the second level branch is a Y branch, or 0 if it is a N branch. We get:



In this case, each different path through the tree gives a different combination of X and Z values. Also, each of the possible values of X (0, 1 or 2) occurs with every possbile value of Y (0 or 1). Therefore the path probabilities give us all the values of the joint distribution. (If some combinations appeared on more than one path, we would need to sum path probabilities to get the corresponding values of the joint distribution. And if some combinations of a possible value of X and a possible value of Z didn't appear on the tree, we would need to put some 0's in the joint distribution table.) We get:

	$(Z=0)$	$(Z=1)$	$Pr[X=x]$
$(X=0)$	2/12	1/12	1/4
$(X=1)$	4/12	2/12	1/2
$(X=2)$	2/12	1/12	1/4
$Pr[Z=z]$	2/3	1/3	

Now, we can easily confirm that *in every case* we have

$$Pr[(X=x)\cap(Z=z)] = Pr[X=x]\times Pr[Z=z]$$

so X and Z *are* independent random variables.

Approach 2: simply recognize what we already know
Notice that when we check whether $Pr[X=x]\times Pr[Z=z]$ is equal to $Pr[(X=x)\cap(Z=z)]$ in the table above, the calculations we do are the exact same calculations we did in finding the path probabilities. Hmm. Why is that?

It's because when we put the probabilities on the tree in the first place, using the same value on all Y branches, and likewise on all N branches, regardless of how many coins came up Heads, it was because we **know** that the outcome of the die toss is independent of the outcome of the coin tosses. And since the value of X is determined (only) by the coin tosses and the value of Z is determined (only) by the toss of the die, then for any possible value x and for any possible value z, the events $(X=x)$ and $(Z=z)$ **must** be independent. Therefore X and Z are indpendent r.v.'s.

Notice: In Example 3.22 (see Lecture 24, pg. 144), we had the same random variables X and Z, so it's not surprising that we also found that $E(XZ)=E(X)\times E(Z)$. Notice, though, that even in this case $V(XZ)\neq V(X)\times V(Z)$.

Also Notice: In Example 3.24 (see Lecture 25, pg. 147), we could have used this kind of reasoning to observe that the number of dimes which come up Heads and the number of quarters which come up Heads *must* be independent.

Example 3.28. *For the random variables defined in Example 3.17, find $E(XY)$.*

Recall that this was the example in which 1 dime and 2 quarters are tossed. X is the number of dimes which come up Heads and Y is the number of quarters which come up Heads. (See Lecture 24, pg. 138.)

Approach 1: Find the distribution of XY
Let $W=XY$. Since X has possible values 0 and 1, and Y has possible values 0, 1 and 2, then $W=XY$ also has possible value 0, 1 and 2. To find the distribution of W, we can again consider the equiprobable sample space we used in Example 3.24:

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

where the first letter denotes the outcome of the dime toss and the next 2 are the first and second dime, respectively. For each sample point, we determine the value of X, the value of Y and the corresponding value of W.

sample point	x	y	$w = xy$
HHH	1	2	2
HHT	1	1	1
HTH	1	1	1
HTT	1	0	0
THH	0	2	0
THT	0	1	0
TTH	0	1	0
TTT	0	0	0

Of course, since each sample point occurs with probability $\frac{1}{8}$, we get $Pr[W = 0] = \frac{5}{8}$, $Pr[W = 1] = \frac{2}{8}$ and $Pr[W = 2] = \frac{1}{8}$. Therefore we have

$$E(XY) = E(W) = \left(0 \times \frac{5}{8}\right) + \left(1 \times \frac{2}{8}\right) + \left(2 \times \frac{1}{8}\right) = 0 + \frac{2}{8} + \frac{2}{8} = \frac{1}{2}$$

Approach 2: The easy way (since we have shown they're independent)
In Example 3.24, we determined that X and Y are independent random variables. Therefore, according to Theorem 1, $E(XY) = E(X) \times E(Y)$. We found (in Example 3.17 – see Lecture 24, pg. 138) that $E(X) = \frac{1}{2}$ and $E(Y) = 1$, so

$$E(XY) = E(X) \times E(Y) = \frac{1}{2} \times 1 = \frac{1}{2}$$

Note: Approach 2 is, of course, much easier. However, it is only because we **know** that X and Y are independent random variables that we are able to use it.

Example 3.29. *For the random variables X and Y defined in Example 3.25, find $E(XY)$.*

In this example (see Lecture 25, pg. 148), we had a single die being tossed. X is 1 when an odd number comes up, and 2 when an even number is tossed. Y has the value 0 when a 1, 2 or 3 is tossed, and the value 1 when a 4, 5 or 6 comes up. We can find the distribution of XY easily by extending the table we used to find the distributions of X and Y. For each outcome, the value of XY is founds (of course) as the product of the value of X and the value of Y.

outcome	1	2	3	4	5	6
probability	1/6	1/6	1/6	1/6	1/6	1/6
Value of X	1	2	1	2	1	2
Value of Y	0	0	0	1	1	1
Value of XY	0	0	0	2	1	2

This gives:

xy	Pr[XY = xy]
0	1/2
1	1/6
2	1/3

so we see that $E(XY) = \left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{6}\right) + \left(2 \times \frac{1}{3}\right) = 0 + \frac{1}{6} + \frac{2}{3} = \frac{5}{6}$.

Notice: In this case, as we have shown in the original example, X and Y are *not* independent random variables, and in this case we see that $E(XY) \neq E(X) \times E(Y)$. We have $E(X) = \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{2}\right) = \frac{3}{2}$ and $E(Y) = \left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{2}\right) = \frac{1}{2}$, so $E(X) \times E(Y) = \frac{3}{2} \times \frac{1}{2} = \frac{3}{4}$, whereas we found that $E(XY) = \frac{5}{6}$.

There is one more useful fact that we should realize.

Theorem 2: If X and Y are independent random variables, then aX and bY are also independent random variables, for any real values of a and b.

Example 3.30. *Find the variance of the value of the coins which come up Heads when 1 dime and 2 quarters are tossed.*

As we have used before (e.g. see Example 3.17, Lecture 24, pg. 138), let X be the number of dimes, and Y be the number of quarters, respectively, which come up Heads. Then as we saw in Example 3.17, the value of the coins which come up Heads is given by $Z = 10X + 25Y$. We need to find $V(Z)$.

Since, as we saw in Example 3.24 (see Lecture 25, pg. 147), X and Y are independent random variables, then, according to Theorem 2, $10X$ and $25Y$ are also independent r.v.'s, so (by Theorem 1) we know that

$$V(Z) = V(10X + 25Y) = V(10X) + V(25Y)$$

Of course, we also know that $V(10X) = (10)^2V(X)$ and $V(25Y) = (25)^2V(Y)$. Furthermore, in Example 3.23 (see Lecture 25, pg. 146), we found that $V(X) = \frac{1}{4}$ and $V(Y) = \frac{1}{2}$. Thus we see that

$$\begin{aligned} V(Z) &= V(10X + 25Y) = V(10X) + V(25Y) \\ &= (10)^2V(X) + (25)^2V(Y) = 100\left(\frac{1}{4}\right) + 625\left(\frac{1}{2}\right) \\ &= 25 + 312.5 = 337.5 \end{aligned}$$

Example 3.31. *Recall Example 3.19. Are X and Y independent random variables?*

<u>Approach 1:</u>
This was the example (see Lecture 24, pg. 141) in which we have a formula expressing a relationship between X and Y: $Y = 1 - \frac{3}{2}X$. Because Y is determined by X according to this formula, then each possible value of Y is associated with a specific value of X and vice versa. But then it cannot possibly be true that X and Y are independent random variables, because we cannot have $Pr[(X = x) \cap (Y = y)]$ being equal to $Pr[X = x] \times Pr[Y = y]$ for all combinations of X and Y. For instance, suppose that x_1 is one of the possible values of X, so that $Pr[X = x_1] \neq 0$. Then when $X = x_1$, the value of Y is $y_1 = 1 - \frac{3x_1}{2}$. Let y_2 be any other possible value of Y, so that $Pr[Y = y_2] \neq 0$. Then since Y *must be* y_1 whenever X has the value x_1, the event $(X = x_1) \cap (Y = y_2)$ can never occur, so $Pr[(X = x_1) \cap (Y = y_2)] = 0$. However, since $Pr[X = x_1] \neq 0$ and $Pr[Y = y_2] \neq 0$, then it must be true that $Pr[X = x_1] \times Pr[Y = y_2] \neq 0$. Thus, we see that X and Y are *not* independent random variables, they are dependent random variables. (*Notice:* Whenever each value of X occurs with only 1 particular value of Y, or vice versa, X and Y are dependent r.v.'s. That is, this kind of situation is an instance where the **Tip** from the end of Lecture 25 (see pg. 150) tells us that X and Y are not independent r.v.'s.)

<u>Approach 2:</u>
This example was also the one we used to observe that in general $V(X + Y) \neq V(X) + V(Y)$, by noticing that it wasn't true in this case. Since we know that, if X and Y are independent random variables then we *must* have $V(X + Y) = V(X) + V(Y)$, then knowing that this relationship does not hold in this case shows that X and Y cannot be independent random variables. They must be dependent random variables.

Example 3.32. *X and Y are random variables defined on the same sample space. X has possible values 0 and 1, while Y has possible values 1 and 2. It is known that $Pr[(X = x) \cap (Y = y)] = \frac{x+y}{8}$.*
(a) Are X and Y independent random variables?
(b) If $Z = X + Y$, find $V(Z)$.

(a) In this case, we are told the possible values of X and Y, and are given a formula for calculating the joint distribution of X and Y. We use the formula to construct the table of the joint distribution and then calculate the marginal probabilities. These give us the probabilities for the various possible values of X, and for the various possible values of Y. We then use the table to determine whether X and Y are independent or dependent random variables.

We consider each possible value of X paired with each possible value of Y to find the joint distribution. We have

$$\begin{aligned} Pr[(X = 0) \cap (Y = 1)] &= \frac{0+1}{8} = \frac{1}{8} \\ Pr[(X = 0) \cap (Y = 2)] &= \frac{0+2}{8} = \frac{1}{4} \\ Pr[(X = 1) \cap (Y = 1)] &= \frac{1+1}{8} = \frac{1}{4} \\ Pr[(X = 1) \cap (Y = 2)] &= \frac{1+2}{8} = \frac{3}{8} \end{aligned}$$

The table showing the joint distribution of X and Y is:

	$(Y = 1)$	$(Y = 2)$	$Pr[X = x]$
$(X = 0)$	1/8	1/4	3/8
$(X = 1)$	1/4	3/8	5/8
$Pr[Y = y]$	3/8	5/8	

From this table we can easily see that X and Y are *not* independent random variables, because, for instance, we have

$$Pr[X = 0] \times Pr[Y = 1] = \frac{3}{8} \times \frac{3}{8} = \frac{9}{64} \neq Pr[(X = 0) \cap (Y = 1)]$$

(b) Since X and Y are dependent random variables, we don't have a formula relating $V(X + Y)$ to $V(X)$ and $V(Y)$. In order to find $V(Z)$ where $Z = X + Y$, we must find the distribution of Z. For each combination of a possible value of X and a possible value of Y, we calculate the value of Z. The probability that Z has this value is given by the probability that this combination of X and Y values is observed (from the joint distribution of X and Y). We get:

x	y	$z = x + y$	$Pr[(X = x) \cap (Y = y)]$
0	1	1	1/8
0	2	2	1/4
1	1	2	1/4
1	2	3	3/8

Since there are 2 events that give $Z = 2$ (i.e. $(X = 0) \cap (Y = 2)$ and $(X = 1) \cap (Y = 1)$), the total probability that event $(Z = 2)$ occurs is the sum of the probabilities of these 2 events. We can make a simplified table showing the pdf of Z and also Z^2, so that we can find $E(Z)$ and $E(Z^2)$, to find $V(Z)$. (*Notice:* Instead, we could use the 4 events shown above to find $E(Z)$ and $E(Z^2)$ directly.)

z	$Pr[Z = z]$	z^2
1	1/8	1
2	1/2	4
3	3/8	9

We get:

$$\begin{aligned} E(Z) &= \left(1\times\frac{1}{8}\right)+\left(2\times\frac{1}{2}\right)+\left(3\times\frac{3}{8}\right)=\frac{9}{4} \\ E(Z^2) &= \left(1\times\frac{1}{8}\right)+\left(4\times\frac{1}{2}\right)+\left(9\times\frac{3}{8}\right)=\frac{11}{2} \\ V(Z) &= E(Z^2)-[E(Z)]^2=\frac{11}{2}-\left(\frac{9}{4}\right)^2=\frac{88}{16}-\frac{81}{16}=\frac{7}{16} \end{aligned}$$

Let's look at one last example.

Example 3.33. *Three \$1 coins and two \$2 coins are tossed. You will win \$1 for every \$1 coin which comes up Heads, but you will* lose *\$2 for ever \$2 coin which comes up Heads. Find the expected value and variance of the amount you will win.*

Let X be the number of \$1 coins which come up Heads. Then $X = B(3, 1/2)$, so $E(X) = 3/2$ and $V(X) = 3/4$. Let Y be the number of \$2 coins which come up Heads. Then $Y = B(2, 1/2)$, so $E(Y) = 1$ and $V(Y) = 1/2$. Let W be the amount you win when you play this game. Then since you win \$1 $\times X$ for the \$1 coins and lose \$2 $\times Y$ for the \$2 coins, we have $W = 1 \times X - 2 \times Y = X - 2Y = X + (-2)Y = aX + bY$, with $a = 1$ and $b = -2$.

Of course, it is always true that $E(aX + bY) = aE(X) + bE(Y)$, so we see that

$$E(W)=E(X-2Y)=E(X)+(-2)E(Y)=E(X)-2E(Y)=\frac{3}{2}-2(1)=-\frac{1}{2}$$

Clearly, the number of \$1 coins which come up Heads has no effect on the number of \$2 coins which come up Heads, since individual coin tosses are always independent of one another, so X and Y are independent r.v.'s. Therefore we know that $V(aX + bY) = a^2V(X) + b^2V(Y)$. So we get:

$$V(W)=V(X-2Y)=V(X+(-2)Y)=V(X)+(-2)^2V(Y)=V(X)+4V(Y)=\frac{3}{4}+4\left(\frac{1}{2}\right)=\frac{3}{4}+2=\frac{11}{4}$$

Math 1228A/B Online

Lecture 27:

Continuous Random Variables
– Using Areas to Find Probabilities

(text reference: Section 4.1, pages 146 - 149)

4 Continuous Random Variables

4.1 Probability Density Functions

We defined a *discrete* random variable to be one which has a finite number of possible values. Some r.v.'s can take on *any real value*, perhaps within a limited range. This is often true of a random variable involving measurement.

Consider the following measurements:

- a person's exact height
- the precise amount of draft beer poured by bars in London in one day
- the actual weight of a steak
- the temperature outside right now

Measurements like these are usually rounded – that is, we use *discrete measurements*. We might state a person's height to the nearest cm, measure draft beer to the nearest ml, refer to the weight of a steak to the nearest gram, and state the temperature to the nearest degree, or tenth of a degree. However all of these are only *approximations* of the actual measurement. The precise measurement could be *any real number* within some reasonable interval. For instance, when the thermometer says the temperature outside is 21 degrees Celsius, it might actually be 21.36947.... or maybe 20.8974593...

> *Definition:* A **continuous random variable** is a random variable whose possible values are some interval of the real number line.

For any continuous random variable X, the probability that X has a value of precisely some specific number x is so small as to be effectively 0. For instance, the probability that the temperature outside is *exactly* $21°C$, and not 21.001 or 20.98634, etc., is infinitesimally small. Because there are an infinite number of possible values, the probability is spread so thinly that the probability of any particular value is 0.

> *Baffling Fact:* If X is a continuous random variable,
> then $Pr[X = x] = 0$ for all possible values x.

However, it is still true that the probability that a continuous r.v. X has *some* value is 1. That is, for any continuous r.v. X, $Pr[-\infty < X < \infty] = 1$. And if all of the possible values of X are between some values a and b, then $Pr[a < X < b] = 1$.

When we work with continuous r.v.'s, we are only interested in events which involve the value of X being in some specified range of values. For instance, we have seen that if X is a random variable measuring the temperature outside at this moment, then $Pr[X = 21] = 0$. However, the probability that X is *near* 21 is not necessarily 0; for instance, on a spring day in London, $Pr[20.995 < X < 21.005]$ would have some non-zero value. Similarly, we might want to evaluate something like $Pr[19.5 < X < 20.5]$ or $Pr[15 < X < 25]$.

The fact that for a continuous random variable X, the event that X has precisely some specific value is effectively impossible means that some things are less complicated with a continuous random variable than they would be with a discrete random variable. Recall from Chapter 3 that if X is a discrete r.v. and x is any possible value of X, then the events $(X \leq x)$ and $(X < x)$ are different events and much care must be taken in defining events, complements of events, etc., to ensure precision about whether or not the bounding value is included in the event. This problem does not arise with continuous r.v.'s.

Fact: Since $Pr[X = x] = 0$ for all x when X is a continuous r.v.,
then $Pr[X \leq x] = Pr[X < x]$ and $Pr[X \geq x] = Pr[X > x]$.

That is, with a continuous r.v., we do not need to distinguish between $<$ and $\leq$. This is part of what makes continuous r.v.'s much easier to work with than discrete r.v.'s. And since the probability of any exact value being observed is 0 for a continuous r.v., then we *only* need to concern ourselves with the cumulative distribution function, $F(x) = Pr[X < x]$. (*Recall:* The definition of a probability distribution function specifies that it applies to a *discrete* random variable, whereas the definition of the cumulative distribution applies to *any* random variable.) If we know the cdf for a continuous r.v. X, we can find any probability we may be interested in, using only

$Pr[X < x] = F(x)$
$Pr[X > x] = 1 - F(x)$
$Pr[a < X < b] = F(b) - F(a)$

There is, though, another kind of function associated with a continuous random variable, which is in some ways analogous to the pdf of a discrete r.v., in that it can be used to calculate values of the cdf. We express the cdf for a continuous r.v. using something called a **density function**.

Definition: The **probability density function** of a continuous random variable X is a function $f(x)$ such that, for all real numbers a and b, $Pr[a < X < b]$ is equal to the area under the graph of $y = f(x)$ between $x = a$ and $x = b$.

It is important to clarify exactly what is meant by the "area under the graph of $y = f(x)$". Students who have taken calculus will be familiar with this concept, but others may not be.

Definition: For any function $f(x)$, the **area under** the graph of $y = f(x)$ means the area of the region which lies below the curve $y = f(x)$ and above the x-axis.

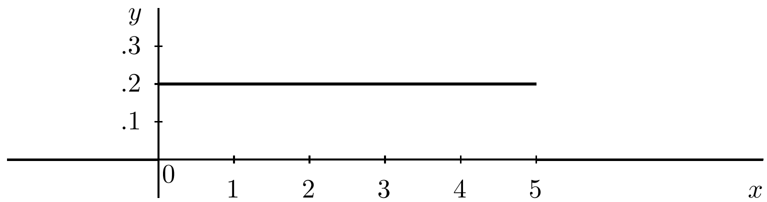
We can look at an example of how a probability density function allows us to find probabilities for a continuous random variable.

Example 4.1. *A continuous random variable X has the probability density function $f(x) = .2$ for $0 < x < 5$ and $f(x) = 0$ otherwise.*
(a) Find $F(3)$.
(b) Find $Pr[1 < X < 2.5]$.

To begin with, we need to draw the graph of this function $f(x)$. *Notice:* We always consider the probability density function of a continuous r.v. X to be defined everywhere, i.e. for all real values of X. However, it is often, as in this case, defined to be equal to 0 outside some interval of x-values.

In this case, we have
$f(x) = 0$ everywhere to the left of $x = 0$,
$f(x) = .2$ on the interval $0 < x < 5$,
$f(x) = 0$ everwhere to the right of $x = 5$.

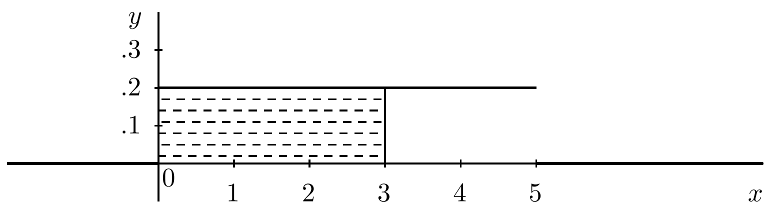
The graph of $y = f(x)$ looks like:



(a) We need to find $F(3) = Pr[X < 3]$, i.e. the probability that X has a value to the left of 3 on the horizontal axis. We can partition this event into the part that's also to the left of 0, and the part that isn't, i.e. is to the right of 0 but to the left of 3. That is, since $(X < 3) = (X < 0) \cup (0 < X < 3)$, then we have $Pr[X < 3] = Pr[X < 0] + Pr[0 < X < 3]$. We use the fact that probabilities involving X can be found by finding areas under the density function.

Since $f(x) = 0$ everywhere to the left of $x = 0$, the density function runs along the x-axis from $-\infty$ to 0, so there's no area under the curve on this interval. That is, within this interval, there's no region which lies below $y = f(x)$ but above the x-axis, so we have $Pr[X < 0] = 0$.

To find $Pr[0 < X < 3]$, we must find the area under $y = f(x)$ from $x = 0$ to $x = 3$.



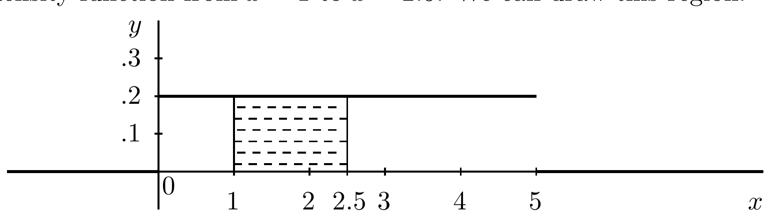
In this case, the region which lies below the density function and above the x-axis between $x = 0$ and $x = 3$ is a rectangle. This rectangle has width 3 units and height .2 units, so its area is

$$\text{width} \times \text{height} = 3 \times .2 = .6$$

Therefore $Pr[0 < X < 3] = .6$, so we have

$$F(3) = Pr[X < 3] = Pr[X < 0] + Pr[0 < X < 3] = 0 + .6 = .6$$

(b) Now, we need to find $Pr[1 < X < 2.5]$. We use the fact that this probability is given by the area under the density function from $x = 1$ to $x = 2.5$. We can draw this region.



Again, the region is a rectangle. The width of the rectangle is the distance from 1 to 2.5, which is $2.5 - 1 = 1.5$, and the height of the rectangle is .2, so we see that

$$Pr[1 < X < 2.5] = (1.5)(.2) = .3$$

Not just any function can be a density function. There are certain properties which a function must have in order to be a density function for some continuous r.v.

Properties of density functions:
For any continuous r.v. X with density function $f(x)$,

1. $f(x) \geq 0$ for all x
   i.e. the density function can never lie below the x-axis anywhere
2. The total area under $y = f(x)$ is 1
   i.e. the total probability, as always, must be 1
3. As x gets very large or very small, $f(x)$ gets very close to (and may coincide with) the x-axis.

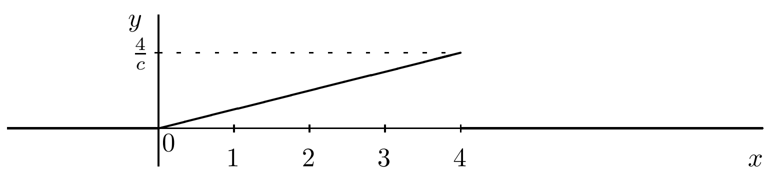
These properties of the density function have implications for the cdf.

Properties of $F(x)$ for a continuous r.v.
For any continuous r.v. X,

1. $F(x)$ is a non-decreasing function
   i.e. as x gets bigger, $F(x)$ may stay the same or get bigger, but can never get smaller
2. as x gets closer to $-\infty$, $F(x)$ gets to, or gets close to, 0
3. as x gets closer to ∞, $F(x)$ gets to, or gets close to, 1

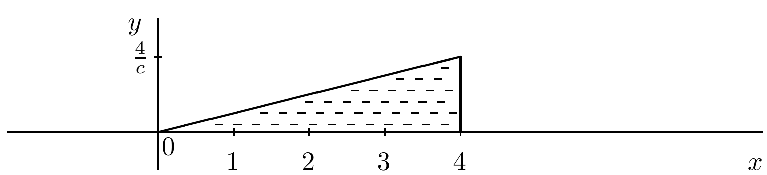
Example 4.2. *The probability density function of a certain continuous random variable, X, is given by $f(x) = \frac{x}{c}$ if $0 < x < 4$, and $f(x) = 0$ otherwise, for some constant c.*

Before we look at the things we're asked to find, let's draw this density function. What do we know here? The function $f(x) = \frac{x}{c}$ is a straight line with slope $\frac{1}{c}$. When $x = 0$, we have $\frac{x}{c} = \frac{0}{c} = 0$, and when $x = 4$, we get $\frac{x}{c} = \frac{4}{c}$. That is, we have $f(0) = 0$ and $f(4) = \frac{4}{c}$, so on the interval from $x = 0$ to $x = 4$, the function $y = f(x)$ is the line segment joining the points $(x, y) = (0, 0)$ and $(x, y) = (4, 4/c)$. Also, we know that $f(x) = 0$ everywhere outside this interval. So we get:



Example 4.2. *(a) What is the value of c?*

What do we know that will allow us to find the value of c? We know that the total area under $y = f(x)$ is 1. Of course, since $f(x)$ coincides with the x-axis outside of $0 < x < 4$, then the area under the curve outside of this interval is 0. That is, to the left of $x = 0$ and to the right of $x = 4$, there is no area under $y = f(x)$. Between these values, we see that the region under $y = f(x)$ is a triangle.



Recall: The area of a triangle is given by $\frac{1}{2}\times$ base $\times$ height

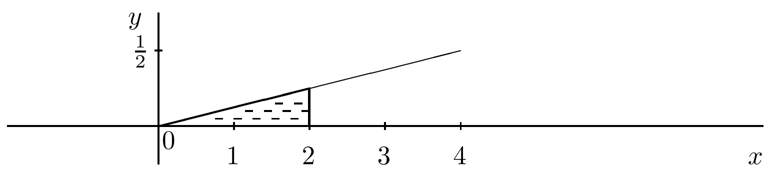
The triangle has base (i.e. width) 4 units and height $\frac{4}{c}$ units. The area of the triangle must be 1, so we need

$$\frac{1}{2} \times 4 \times \frac{4}{c} = 1$$
$$\Rightarrow \quad \frac{8}{c} = 1$$
$$\textit{(multiply through by c)} \qquad \Rightarrow \quad c = 8$$

Notice: This means that the height of the triangle above is $\frac{4}{c} = \frac{4}{8} = \frac{1}{2}$. Also, the density funtion is $f(x) = \frac{x}{8}$ if $0 < x < 4$, and $f(x) = 0$ otherwise.

Example 4.2. *(b) Find $F(2)$.*

Now that we know the value of c, we can easily find probabilities by finding areas of regions under the curve $y = f(x)$. We know that $F(2) = Pr[X < 2]$ is given by the area under the curve to the left of $x = 2$. Since $f(x) = 0$ everywhere to the left of $x = 0$, then $Pr[X < 0] = 0$, so we have $F(2) = Pr[0 < X < 2]$, which is again given by the area of a triangle.

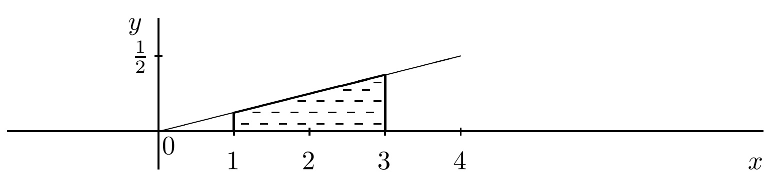


When $x = 2$, we have $f(2) = \frac{2}{c} = \frac{2}{8} = \frac{1}{4}$. Thus the height of the function $y = f(x)$ at $x = 2$, and therefore the height of the triangle, is 1/4. The base of the triangle is 2. Again using the fact that area $= \frac{1}{2}\times$ base $\times$ height, we get

$$F(2) = Pr[0 < X < 2] = \frac{1}{2} \times 2 \times \frac{1}{4} = \frac{1}{4}$$

Example 4.2. *(c) Find $Pr[1 < X < 3]$.*

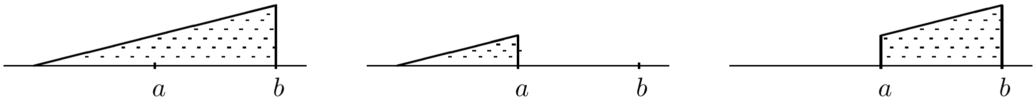
This probability is given by the area under the density curve from $x = 1$ to $x = 3$.



This shape is called a trapezoid. There are various methods that can be used to find the area of a region like this.

To find $Pr[a < X < b]$ when the region under $f(x)$ from $x = a$ to $x = b$ is a trapezoid:

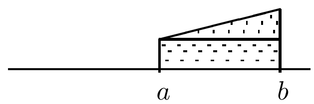
Approach 1:
Use the fact that $Pr[a < X < b] = F(b) - F(a)$. For this kind of situation, for both $F(a)$ and $F(b)$, the region whose area gives the probability is a triangle, where the triangle giving $F(b)$ is bigger than the one giving $F(a)$. (Notice that the trapezoid is a triangle with one end cut off, i.e. with a smaller triangle taken away.) We find the area of the trapezoid as the difference between the area of the larger triangle and the area of the smaller triangle. What we have is:



Triangle B   Triangle A   Trapezoid has area given by area of B − area of A

Notice: Triangle B has height $f(b)$, while triangle A has height $f(a)$.

Approach 2:
A trapezoid can be viewed as a triangle sitting on top of a rectangle



The rectangle has width $b-a$ and height $f(a)$. The small triangle sitting on top of it has base $b-a$ and height $f(b)-f(a)$. The area of the trapezoid is the sum of the areas of these two regions.

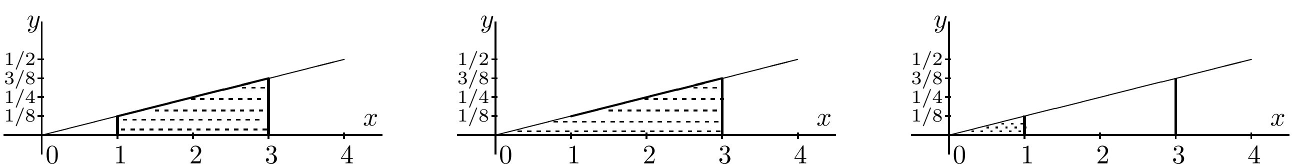
Approach 3:
There is a fairly easy formula for (directly) finding the area of a trapezoid.

> **Formula:** The area of a trapezoid with width w, left height h_L and right height h_R is given by width times average height, i.e.
>
> $$w \times \frac{h_L + h_R}{2}$$

Of course, the trapezoid defined by the area under the density function on the interval $a < x < b$ has $h_L = f(a)$ and $h_R = f(b)$, and has width $b-a$.

We can find the area of the trapezoid in Example 4.2 (c) using each of these different approaches. (Of course, normally we would use only one of the 3 approaches.)

Approach 1: We have:



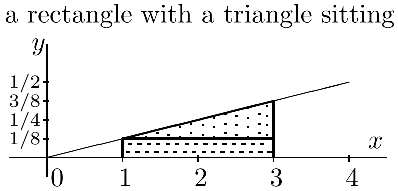
Area of this trapezoid   equals   area of this triangle   minus   area of this triangle

For the larger triangle, on the interval $0 < x < 3$, the base is 3 and since the density function is $f(x) = \frac{x}{8}$, we see that the height is $f(3) = \frac{3}{8}$. Therefore the area is $\frac{1}{2} \times 3 \times \frac{3}{8} = \frac{9}{16}$.

For the smaller triangle, on the interval $0 < x < 1$, the base is 1 and the height is $f(1) = \frac{1}{8}$ so the area is $\frac{1}{2} \times 1 \times \frac{1}{8} = \frac{1}{16}$. Therefore the area of the trapezoid, given by area of larger triangle minus area of smaller triangle, is

$$\frac{9}{16} - \frac{1}{16} = \frac{8}{16} = \frac{1}{2}$$

Approach 2:
We consider the trapezoid as a rectangle with a triangle sitting on top of it. We have:



The width of the rectangle and the base of the triangle are both given by $3 - 1 = 2$. The height of the rectangle is $f(1) = \frac{1}{8}$. The height of the triangle is the vertical distance from $\frac{1}{8}$ to $f(3)$, so the height is $f(3) - \frac{1}{8} = \frac{3}{8} - \frac{1}{8} = \frac{2}{8} = \frac{1}{4}$. Therefore the area of the trapezoid, given by the area of the rectangle plus the area of the triangle, is

$$\left(2 \times \frac{1}{8}\right) + \left(\frac{1}{2} \times 2 \times \frac{1}{4}\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Approach 3:
We find the area of the trapezoid using the formula. In this case, we have width $w = 3 - 1 = 2$, left height $h_L = f(1) = \frac{1}{8}$ and right height $h_R = f(3) = \frac{3}{8}$, so we get

$$w \times \frac{h_L + h_R}{2} = 2 \times \frac{\frac{1}{8} + \frac{3}{8}}{2} = \frac{4}{8} = \frac{1}{2}$$

Therefore (no matter which of these approaches we use), we see that since $Pr[1 < X < 3]$ is given by the area of the trapezoid, we have

$$Pr[1 < X < 3] = \frac{1}{2}$$

Notice: The quickest approach is, of course, to use the formula which gives us the area of the trapezoid directly. However this requires remembering yet another formula, and also (for the trapezoids we will encounter) evaluating a fraction whose numerator is a sum of fractions (or decimal values between 0 and 1). Some students may wish to avoid these "complications" by using one of the other 2 approaches. Each student should choose whichever one of the 3 approaches seems easiest for them, and use that approach whenever it is necessary to find the area of a region whose shape is a trapezoid.

Math 1228A/B Online

Lecture 28:

Using Continuous Random Variables To Approximate Discrete Random Variables

(text reference: Section 4.1, pages 149 - 151)

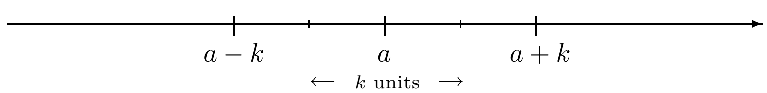
Approximating Discrete Random Variables

We talked earlier about the fact that for a continuous random variable such as "X: the height of a randomly chosen student in the class", in practice, we use discrete measurements rather than exact measurements, i.e. we usually round this kind of thing. So why don't we just use a discrete r.v. instead for something like that? That is, instead of using a continuous r.v., why don't we use a discrete r.v. that is *approximately* the same as the continuous r.v.?

Answer: Because continuous r.v.'s are *much easier* to work with.

In fact, rather than approximating a continuous r.v. by a discrete r.v., we do exactly the opposite. We often approximate a discrete r.v. by a continuous r.v., because values of the cdf are easier to find that way.

Suppose that we have some discrete random variable X whose possible values are evenly spaced, k units apart. Let a be a possible value of X. How could we have $Pr[X = a]$ approximated by some event in terms of some continuous random variable, Y? Clearly we can't approximate $Pr[X = a]$ by $Pr[Y = a]$, since $Pr[Y = a] = 0$ when Y is continuous. We need to be dealing with an interval of values when we work with a continuous random variable. What interval of Y-values would we use to approximate the event $(X = a)$? We have possible values of X occuring every k units along the number line, so we use an interval of width k. And we want the interval to be centred at a, so we go from $\frac{k}{2}$ units to the left of a to $\frac{k}{2}$ units to the right of a.



Possible values $a-k$, a and $a+k$ are evenly spaced k units apart.
An interval of width k centred at a runs from $a - \frac{k}{2}$ to $a + \frac{k}{2}$

Therefore, we approximate the event $(X = a)$ by the event $\left(a - \frac{k}{2} < Y < a + \frac{k}{2}\right)$.

So when can we do this? Well, whenever it works. Whenever it gives us the same probabilities.

Definition: Let X be a discrete r.v. whose possible values are evenly spaced, k units apart. The continuous r.v. Y is called a **good approximation for** X if

$$Pr\left[a - \frac{k}{2} < Y < a + \frac{k}{2}\right] = Pr[X = a]$$

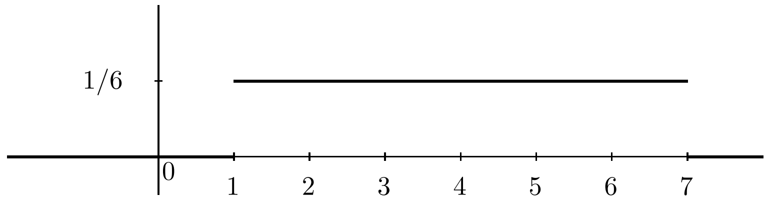
for each value a such that a is a possible value of X.

Let's look at an example of how this works.

Example 4.3. *Discrete random variable X has possible values 2, 4 and 6, each occurring with equal probability. Continuous random variable Y has probability density function $f(y) = \frac{1}{6}$ if $1 < y < 7$ and $f(y) = 0$ otherwise. Show that Y is a good approximation for X.*

Here, X is a discrete random variable. And since there are only 3 equally likely possible values, then we have $Pr[X = 2] = Pr[X = 4] = Pr[X = 6] = \frac{1}{3}$. Notice that the possible values of X are evenly spaced $k = 2$ units apart.

We also have continuous random variable Y. We can graph the density function of Y in the usual way.

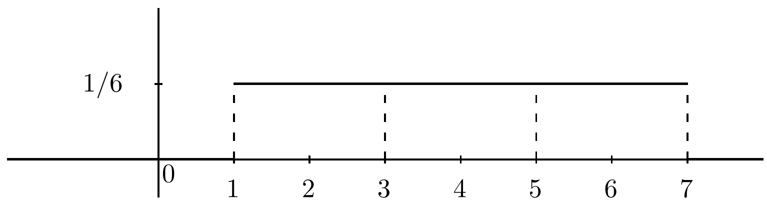


We need to show that Y is a good approximation for X. To do this, we first need to identify the intervals in terms of Y which are going to be used to approximate each of the possible values of X. We have $k = 2$, so $\frac{k}{2} = \frac{2}{2} = 1$ and so we approximate $(X = 2)$ by the event

$$\left(2 - \frac{k}{2} < Y < 2 + \frac{k}{2}\right) = (2 - 1 < Y < 2 + 1) = (1 < Y < 3)$$

Likewise, we approximate $(X = 4)$ by $(4 - 1 < Y < 4 + 1) = (3 < Y < 5)$ and $(X = 6)$ by $(6 - 1 < Y < 6 + 1) = (5 < Y < 7)$.

Notice: The events approximating the possible values of X span the whole range of values for which the density function of Y is non-zero.



Looking at these events on the graph of the pdf of Y, we see that

$$\begin{aligned} Pr[1 < Y < 3] &= \text{area under } f(y) \text{ from } y = 1 \text{ to } y = 3 \\ &= \text{area of rectangle with width } 3 - 1 = 2 \text{ and height } 1/6 \\ &= (2) \times (1/6) = 1/3 \end{aligned}$$

But then since we know that $Pr[X = 2] = \frac{1}{3}$, we see that $Pr[1 < Y < 3] = Pr[X = 2]$. Similarly, $Pr[3 < Y < 5]$ and $Pr[5 < Y < 7]$ are each given by the area of a region which is rectangular in shape, with width 2 and height $\frac{1}{6}$. So we get $Pr[3 < Y < 5] = Pr[5 < Y < 7] = 1/3$, and we see it is also true that $Pr[3 < Y < 5] = Pr[X = 4]$ and that $Pr[5 < Y < 7] = Pr[X = 6]$. Since, for each possible value a of the discrete random variable X, we have $Pr\left[a - \frac{k}{2} < Y < a + \frac{k}{2}\right] = Pr[X = a]$, we see that Y *is* a good approximation for X.

Notice that what we have done, in finding an interval of values of the continuous r.v. Y with which to approximate the event $(X = a)$, is to *extend the event by* $\frac{k}{2}$ *units in each direction.* We can refer to this as a **continuity correction**.

When we want to find an approximation for the probability that the discrete r.v. X has a value within some interval of values, i.e. the probability of an event involving 2 or more consecutive possible values of X, we find the approximating interval by again applying the same kind of continuity correction. That is, once again, we extend the event by $\frac{k}{2}$ units in each direction. (Of course, for a one-sided interval, we only need to extend it in one direction.) This approach gives the following general rules:

Theorem: Let X be a discrete r.v. whose possible values are evenly spaced k units apart, and let continuous r.v. Y be a good approximation for X. Let $F_X(x)$ and $F_Y(y)$ denote the cumulative distribution functions of X and Y, respectively. Then

$$(1)\ F_X(a) = Pr\left[Y < a + \tfrac{k}{2}\right] \qquad \text{i.e. } Pr[X \leq a] = Pr\left[Y < a + \tfrac{k}{2}\right]$$

$$(2)\ Pr[a \leq X \leq b] = Pr\left[a - \tfrac{k}{2} < Y < b + \tfrac{k}{2}\right]$$

$$\Rightarrow (3)\ Pr[a \leq X \leq b] = F_Y(b + k/2) - F_Y(a - k/2)$$

Before we look at an example in which we apply these ideas to actually *find* probabilities of a discrete random variable using a continuous random variable which is a good approximation for that discrete random variable, let's focus on simply finding the approximating intervals.

Example 4.4. *X is a discrete random variable which has possible values 1, 2, 3, 4 and 5. Continuous random variable Y is known to be a good approximation for X.*
(a) For each of the following, find the event in terms of Y which approximates the given event in terms of X.
(i) $(X = 1)$ *(ii)* $(X \leq 2)$ *(iii)* $(2 \leq X \leq 3)$ *(iv)* $(1 < X < 5)$
(b) What probability, in terms of X, is approximated by $Pr[Y > 2.5]$?

We have a discrete r.v. whose possible values are consecutive integers, i.e. are evenly spaced $k = 1$ unit apart. We also have a continuous r.v. Y which is a good approximation for X. In order to approximate probabilities involving X using events expressed in terms of Y, we will need to use a continuity correction of $\frac{k}{2} = \frac{1}{2}$.

(a) (i) We want to approximate the event $(X = 1)$. To find the event in terms of Y which is used to approximate this event, we extend the "interval" (which is currently only the value 1) by $\frac{k}{2}$ units in each direction. That is, we find the lower bound on the interval of y-values by subtracting $\frac{k}{2} = \frac{1}{2}$, and the upper bound by adding $\frac{1}{2}$. We get:

$$(X = 1) \approx (1 - 1/2 < Y < 1 + 1/2) = (1/2 < Y < 3/2)$$

That is, we use the event $\left(\frac{1}{2} < Y < \frac{3}{2}\right)$ to approximate $(X = 1)$.

(ii) Now, we want to find the approximation for the event $(X \leq 2)$. This is a one-sided interval involving X, so we extend the interval by $k/2$ in one direction:

$$(X \leq 2) \approx \left(Y < 2 + \frac{1}{2}\right) = \left(Y < \frac{5}{2}\right)$$

(iii) This time, we have a 2-sided interval of X-values, which we extend in both directions:

$$(2 \leq X \leq 3) \approx \left(2 - \frac{1}{2} < Y < 3 + \frac{1}{2}\right) = \left(\frac{3}{2} < Y < \frac{7}{2}\right)$$

(iv) For the event $(1 < X < 5)$, we need to be careful because it isn't expressed in the '$\leq$' form which we know how to deal with. The first thing we need to do is re-express the event in terms of X in a form which uses $\leq$. Since the next larger value of X, from $x = 1$, is 2, the event $(1 < X)$, i.e. $(X > 1)$, can be expresses as the event $(X \geq 2)$, i.e. $(2 \leq X)$. Likewise, for $x = 5$, the next smaller value of X is 4, so $(X < 5) = (X \leq 4)$. Therefore we have

$$(1 < X < 5) = (2 \leq X \leq 4)$$

Now that we have the right form, we can apply the continuity correction to find the approximating event.

$$(1 < X < 5) = (2 \le X \le 4) \approx \left(2 - \frac{1}{2} < Y < 4 + \frac{1}{2}\right) = \left(\frac{3}{2} < Y < \frac{9}{2}\right)$$

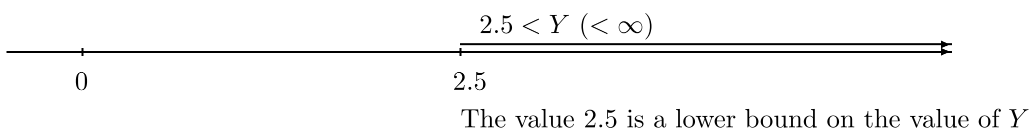
Remember: Although $\le$ and $<$ are the same when working with a continuous random variable, they are very different for a discrete random variable. We must always make sure we have the discrete r.v. expressed in $(a \le X \le b)$ form before applying the continuity correction 'extend the interval $\frac{k}{2}$ units in each direction'.

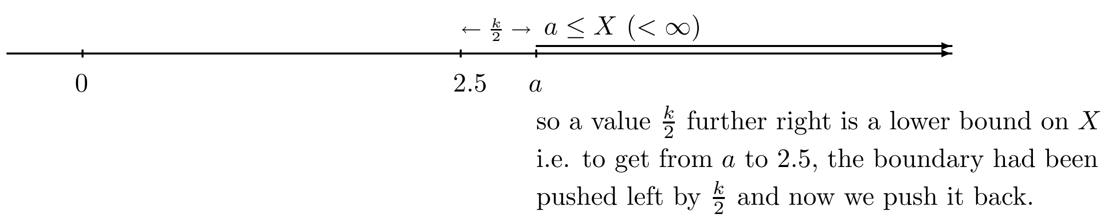
(b) We need to determine what probability, in terms of X, is approximated by $Pr[Y > 2.5]$. To do this, we need to apply the continuity correction in reverse. That is, a continuity correction of $\frac{k}{2} = \frac{1}{2}$ has been applied to get this event in terms of Y, so we must 'unapply' it to find the event in terms of X which is being approximated.

Has the continuity correction been added or subtracted? What we have here is a lower bound on Y, i.e. the left end of the interval. Perhaps it will be helpful to write this event with a $<$ instead of a $>$ in order to see this:

$$Pr[Y > 2.5] = Pr[2.5 < Y] \qquad\qquad (\text{or } Pr[2.5 < Y < \infty])$$

Since this is a left endpoint, it has been extended further left by subtracting $k/2$. That is, to undo the continuity correction, we need to *contract* the interval by $k/2$, and we make this interval smaller by moving the endpoint further right, i.e. by adding $k/2$. We see that the value 2.5 has been obtained by *subtracting* the continuity correction, so we counteract this by adding.





Since we have $\frac{k}{2} = \frac{1}{2} = .5$, and $a - \frac{k}{2} = 2.5$, then we get $a - .5 = 2.5 \Rightarrow a = 2.5 + .5 = 3$. That is, we see that since the 2.5 would have been obtained using the approximation

$$(a \le X) \approx \left(a - \frac{k}{2} < Y\right)$$

then we have

$$\begin{aligned} (Y > 2.5) &= (2.5 < Y) = (3 - .5 < Y) \\ &\approx (3 \le X) = (X \ge 3) \end{aligned}$$

That is, $Pr[Y > 2.5] = Pr[2.5 < Y] = Pr[a - .5 < Y] = Pr[a \le X] = Pr[X \ge a]$ and since $a - .5 = 2.5$ then $a = 3$, so $Pr[Y > 2.5] = Pr[X \ge 3]$.

Let's look at one more example:

Example 4.5. *Random variable X has possible values 10, 13, 16 and 19. Continuous random variable Y is a good approximation for X.*
(a) In what range of y-values does the density function lie above the horizontal axis?
(b) What probability, in terms of X, is approximated by $Pr[Y < 11.5]$?

Notice: This time, the possible values of X are evenly spaced $k = 3$ units apart, so the continuity correction is $\frac{k}{2} = \frac{3}{2} = 1.5$.

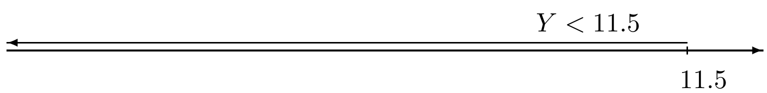
(a) What is this question asking us? We want to know the range of values of Y in which all of the probability lies. That is, we need to find the event in terms of Y which corresponds to the event 'X has one of its possible values', which is the event $(10 \leq X \leq 19)$. As always, we need to extend this interval by $k/2$ in each direction. We get

$$(10 \leq X \leq 19) \approx (10 - 1.5 < Y < 19 + 1.5) = (8.5 < Y < 20.5)$$

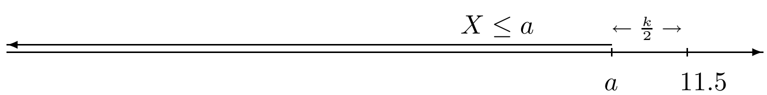
We see that the region which lies below the density curve and above the horizontal axis extends from $y = 8.5$ to $y = 20.5$. The density curve will coincide with the horizontal axis outside of this interval, i.e. $f(y)$ has non-zero values (and hence lies above the horizontal axis) only on $8.5 < y < 20.5$.

Notice: Since $Pr[10 \leq X \leq 19] = 1$, and we know that Y is a good approximation for X, then it must be true that $Pr[8.5 < Y < 20.5] = 1$ too, so outside of this interval the density function of Y must coincide with the horizontal axis.

(b) We are asked to find the probability, in terms of X, which is approximated by $Pr[Y < 11.5]$. To do this, we again need to *undo* the continuity correction. As before, we have a one-sided interval, but this time it is the right side. That is, 11.5 is the right endpoint of the interval, so in order to contract the interval by $k/2$, we need a value less far right, i.e. further left, so we subtract 1.5.



The value 11.5 is an upper bound on the value of Y.



We must contract the interval by moving this upper bound $\frac{k}{2}$ units to the left.

That is, the upper bound on Y must have been obtained using the approximation

$$(X \leq a) \approx \left(Y < a + \frac{k}{2}\right) \qquad \text{i.e. } Pr[X \leq a] = Pr\left[Y < a + \frac{k}{2}\right]$$

so, with $\frac{k}{2} = 1.5$ and $a + \frac{k}{2} = 11.5$, we see that this time we must have $a = 10$, i.e. that

$$(Y < 11.5) = (Y < 10 + 1.5) \approx (X \leq 10)$$

That is,

$$Pr[Y < 11.5] = Pr\left[Y < a + \frac{k}{2}\right] = Pr[X \leq a], \text{ so } Pr[Y < 11.5] = Pr[X \leq 11.5 - 1.5] = Pr[X \leq 10]$$

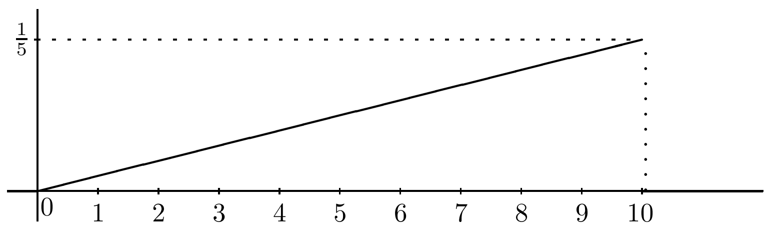
Let's look at an example of actually *using* a continuous r.v. which is a good approximation of a discrete r.v. to find probabilities involving the discrete random variable.

Example 4.6. *Discrete random variable X has possible values 1, 3, 5, 7 and 9. Continuous random variable Y has density function $f(y) = \frac{y}{50}$ if $0 < y < 10$ and $f(y) = 0$ otherwise. It is known that Y*

is a good approximation for X*.*
(a) Find $Pr[X = 1]$*.*
(b) Find $Pr[3 \le X \le 7]$*.*
(c) Find $Pr[X < 7]$*.*

X is a discrete random variable whose possible values are evenly spaced $k = 2$ units apart. Since Y is a good approximation for X, we can use areas under the density function of Y to find probabilities involving X. When we do this, we need to apply a continuity correction of $\frac{k}{2} = 1$.

We start by graphing the density function of Y. Of course, $f(y)$ coincides with the horizontal axis when $y < 0$ and when $y > 10$. Within the interval $0 < y < 10$, the density function has values given by $f(y) = \frac{y}{50}$. At the left end of this interval, when $y = 0$, the height of the density function is $f(0) = \frac{0}{50} = 0$. At the right end of the interval, the height of the density function is $f(10) = \frac{10}{50} = \frac{1}{5}$.



For each probability we are asked to find, we need to

1. find the event in terms of Y which approximates the given event in terms of X, then
2. find the probability of this approximating event by finding the area of the corresponding region under the density function of Y.

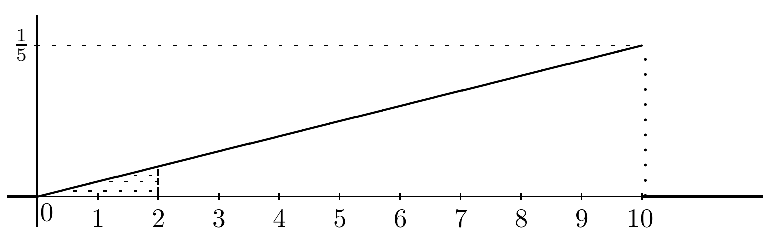
(a) To find $Pr[X = 1]$, we must extend the event $(X = 1)$ by $\frac{k}{2} = 1$ unit in both directions. That is, we know that

$$(X = a) \approx \left(a - \frac{k}{2} < Y < a + \frac{k}{2}\right)$$

so we have

$$Pr[X = 1] = Pr[1 - 1 < Y < 1 + 1] = Pr[0 < Y < 2]$$

We find this probability as the area under $f(y)$ from $y = 0$ to $y = 2$.



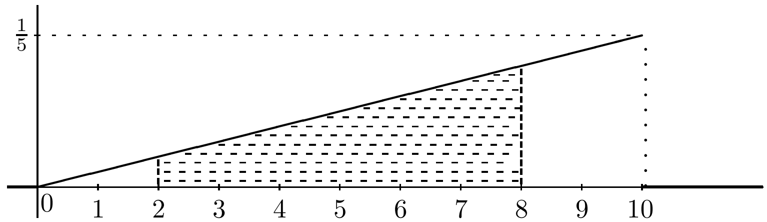
This region is a triangle with base 2 and height $f(2) = \frac{2}{50} = \frac{1}{25}$. Thus we see that

$$Pr[X = 1] = Pr[0 < Y < 2] = \frac{1}{2} \times 2 \times \frac{1}{25} = \frac{1}{25} = .04$$

(b)Again, we find $Pr[3 \le X \le 7]$ by extending the interval by 1 unit in both direction. We see that

$$(3 \le X \le 7) \approx (3 - 1 < Y < 7 + 1) = (2 < Y < 8)$$

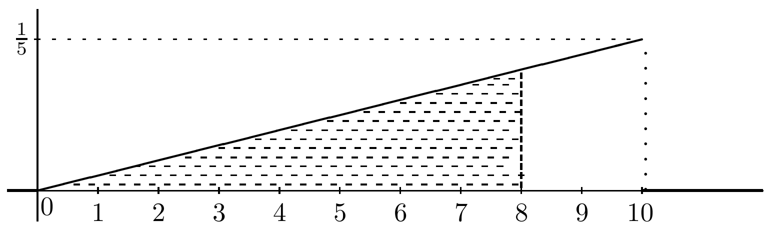
and so $Pr[3 \le X \le 7] = Pr[2 < Y < 8]$. We have



Since we have already found the area of the region to the left of $y = 2$, the easiest way to find the area of this trapezoid is to use the 'difference of triangles' approach. That is, we use

$$Pr[2 < Y < 8] = Pr[Y < 8] - Pr[Y < 2] = Pr[0 < Y < 8] - Pr[0 < Y < 2]$$

(since $Pr[Y < 0] = 0$). From part (a), we know that $Pr[0 < Y < 2] = \frac{1}{25}$. For $Pr[0 < Y < 8]$, we need to find the area of another triangle:



This larger triangle has base 8 and height $f(8) = \frac{8}{50} = \frac{4}{25}$, so we have

$$Pr[0 < Y < 8] = \frac{1}{2} \times 8 \times \frac{4}{25} = \frac{16}{25}$$

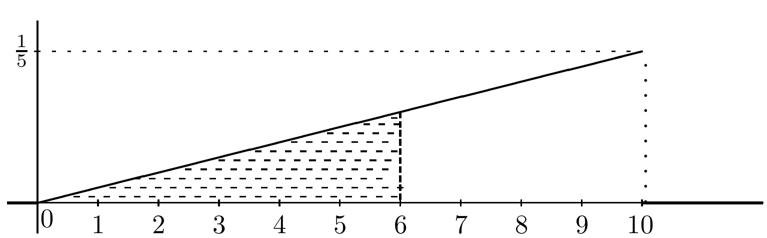
Therefore we see that:

$$\begin{aligned} Pr[3 \leq X \leq 7] &= Pr[2 < Y < 8] \\ &= Pr[0 < Y < 8] - Pr[0 < Y < 2] \\ &= \frac{16}{25} - \frac{1}{25} = \frac{15}{25} = \frac{3}{5} \end{aligned}$$

(c) We need to find $Pr[X < 7]$. We must be careful here. Since the possible values of X are 1, 3, 5, 7 and 9, if X is strictly less than 7, it must be no bigger than 5. That is, in this case, we have $(X < 7) = (X \leq 5)$, so

$$Pr[X < 7] = Pr[X \leq 5]$$

Now, to find $Pr[X \leq 5]$, we extend this one-sided interval, for which what we have is a right-side boundary, 1 unit further right, and find that the approximation for the event $(X \leq 5)$ is the event $(Y < 5 + 1) = (Y < 6)$. We find $Pr[Y < 6]$ as the area of another triangle.



That is, we have $Pr[Y < 6] = Pr[0 < Y < 6]$, since $Pr[Y < 0] = 0$. We see that this probability is given by the area of a triangle with base 6 and height $f(6) = \frac{6}{50} = \frac{3}{25}$, so we have

$$Pr[X < 7] = Pr[X \leq 5] = Pr[Y < 6] = \frac{1}{2} \times 6 \times \frac{3}{25} = \frac{9}{25} = .36$$

Another Approach: (a slightly different view of what we did)
We need to find $Pr[X < 7] = Pr[X \leq 5]$. Since the smallest possible value of X is 1, then we can express $(X \leq 5)$ as $(1 \leq X \leq 5)$. We approximate this 2-sided event by extending the interval 1 unit in each direction, so we have

$$(X \leq 5) = (1 \leq X \leq 5) \approx (1 - 1 < Y < 5 + 1) = (0 < Y < 6)$$

This identifies the same triangle as we used in the previous approach. That is, we have

$$\begin{aligned} Pr[X < 7] &= Pr[X \leq 5] \\ &= Pr[1 \leq X \leq 5] \\ &= Pr[0 < Y < 6] \\ &= \text{area of triangle with base 6 and height } 6/50 \\ &= .36 \end{aligned}$$

In this approach, we do exactly the same calculations. The only difference is that we recognized the lower bound on X, which translates in to the lower bound on Y, instead of waiting until we have expressed the event in terms of Y before recognizing the existence of the lower bound.

Note: Whenever discrete r.v. X has a smallest possible value and a largest possible value, we can always express one sided intervals as 2-sided intervals, which may make it easier to see how the continuity correction is applied. However, when we come back to these methods at the end of the course, we'll be approximating discrete r.v.'s by continuous r.v.'s which have no upper or lower bound. In those cases, trying to use this second approach may actually be a hindrance, rather than a help.

Math 1228A/B Online

Lecture 29:
The Standard Normal Random Variable Z

(text reference: Section 4.2)

4.2 The Standard Normal Random Variable

There is a very important *family* of continuous r.v.'s which are called **Normal random variables**. These r.v.'s all have the same basic shape to their density functions. These density functions are called **normal distributions**. You're probably already somewhat familiar with these density functions, but you call them *bell curves*.

Some Facts About the Normal Distributions

The probability density function for a normal r.v. is a bell-shaped curve which

1. is symmetric about its mean
2. extends infinitely in both directions
   i.e. never actually touches the horizontal axis.

Because of (2), a normal r.v. can, theoretically at least, have *any* real value. That is, there's no limit on how big or how small the value can be. (See, for instance, text p. 154, figure 4.4. Note that the curve does not actually touch the x-axis, but extends infinitely in both directions, getting closer and closer to, but never actually reaching, the axis.)

In the next section, we'll talk about the general family of normal r.v.'s. In this section, we first will be learning about *one particular* normal random variable.

> *Definition:* The **Standard Normal Random Variable** is the normal r.v. which has mean $\mu = 0$ and standard deviation $\sigma = 1$. We always use Z to denote the standard normal random variable.

Since *any* normal distribution is symmetric about its mean, the probability density function of Z is symmetric about the line $z = 0$. (We'll use this symmetry *a lot.*)

The formula for the density function of the standard normal distribution is ugly. (If you're curious, it's on p. 153 in the text. But we really have no need to look at the mathematical statement of this function.) Regions under this curve are not triangles or rectangles, or even trapezoids, so we can't find probabilities involving Z using the methods we used in section 4.1. (In fact, not even the usual methods of Calculus can be used to find areas under this density function.)

Instead, we use a table, which gives values of the cdf for Z. This table is on p. 180 in the text. It gives values of $F(z) = Pr[Z < z]$ (accurate to 4 decimal places) for values of z from 0 to 3.50 (accurate to 2 decimal places).

We know that Z can also have negative values – why aren't those tabulated? Because we can easily find any probabilities we want using only positive z-values, using the symmetry of the distribution.

Using the Symmetry of the Z-Distribution

Remember, the density function is symmetric about the line $z = 0$. This means that for any region under the curve which is bounded on only one side, there is a corresponding region (i.e. one whose area is identical) which is bounded by a value equidistant from 0, but on the other side of 0.

First, consider the 2 equal regions to either side of this line of symmetry, $z = 0$. Since $f(z)$ is a probability density function, the total area under this curve is, of course, 1. And since it is symmetric about $z = 0$, then half the area lies to the left of 0 and half the area lies to the right of 0, so we see that

$$Pr[Z < 0] = Pr[Z > 0] = 0.5$$

Now, let's think about other regions under this curve, bounded only on one side. For instance, consider the region under the standard normal density function which lies to the left of some negative value, $-k$. This region is identical in size to the region which lies to the right of the value $|-k| = k$. (See text p. 154, figure 4.5. The area of the 'left tail', to the left of $z = -k$, is identical to the area of the 'right tail', to the right of $z = k$.) Therefore $Pr[Z < -k] = Pr[Z > k]$, so we can find $Pr[Z < -k]$, where $-k$ is negative, by finding $Pr[Z > k]$, where k is positive.

Similarly, the region under the standard normal density function which lies to the right of some negative value, $-k$, is identical in size to the region which lies to the left of the positive value $|-k| = k$. (Look at text p. 154, figure 4.4. Find $z = -k$, the same distance from 0 as $z = k$, but to the left of 0 instead of to the right. The region to the right of this $z = -k$ value is the same size as the shaded region, i.e. the region to the left of $z = k$.) So we see that $Pr[Z > -k] = Pr[Z < k]$, and therefore we can find $Pr[Z > -k]$, where $-k$ is negative, by finding $Pr[Z < k]$, where k is positive.

Thus we see that any time we need to find a probability involving a negative value of the standard normal random variable Z, we can find this probability by finding the probability of an event expressed in terms of a positive value of Z, so there is no need to (i.e. it would be needlessly repetitive to) tabulate values of the cdf of Z, $F(z) = Pr[Z < z]$, for negative z-values.

Note: The table shows values of $Pr[Z < z]$ for z-values from 0 to 3.5. Of course, one might also wonder 'why only to 3.5?'. The answer is because $Pr[Z < 3.5]$ is very close to 1. That is, by $z = 3.5$ we have captured almost all of the probability. In fact, at the bottom of the table, values of $Pr[Z < z]$ are also given for a few larger values of z (4, 4.5, 5). In this region, the values of $Pr[Z < z]$ are changing very slowly and are all *approximately* 1. In practice we would rarely need Z-values bigger than these, and if we did we would say that $Pr[Z < z] \approx 1$ for any $z > 5$.

Finding values in the Z-table

The setup of the Z-table is a little unusual. The table has a *row* for each z-value, correct to the *first* decimal place, and then a *column* for each possible second decimal place. So if we want to look up $Pr[Z < a.bc]$, we find the row whose label is $a.b$, and then look in the column whose label is $.0c$. The table entry in that row and column is the value of $Pr[Z < a.bc]$.

For instance, to find $Pr[Z < 2.34]$, we go to the 2.3 row and the .04 column (see table, text p. 180). The number we find there is .9904, so we have $Pr[Z < 2.34] = .9904$. Similarly, we find the number .5199 in the 0.0 row and the .05 column, so $.5199 = Pr[Z < 0.05]$.

Example 4.7. *If Z is the standard normal random variable, find the following values:*
(a) $Pr[Z < 1.23]$.

This one's easy. We just look in the table, in the 1.2 row and the .03 column. We find .8907, so $Pr[Z < 1.23] = .8907$.

Example 4.7. *(b) $Pr[Z > 0.12]$.*

We know that (since Z is a continuous r.v.) the event $(Z > 0.12)$ is the complement of the event $(Z < 0.12)$. That is, since $Pr[Z = 0.12] = 0$, then if Z is not smaller than 0.12, it must be larger than 0.12. So we have $Pr[Z > 0.12] = 1 - Pr[Z < 0.12]$. From the table, we find that $Pr[Z < 0.12] = .5478$, and we get $Pr[Z > 0.12] = 1 - .5478 = .4522$.

Example 4.7. *(c)* $Pr[Z < -0.12]$.

Because the Z-distribution is symmetric about $z = 0$, then we know that $Pr[Z < -k] = Pr[Z > k]$. So here, we get

$$Pr[Z < -0.12] = Pr[Z > 0.12] = .4522 \quad \text{from (b)}$$

Example 4.7. *(d)* $Pr[Z > -1.23]$.

Again, we use the symmetry of $f(z)$.

$$Pr[Z > -1.23] = Pr[Z < 1.23] = .8907 \quad \text{from (a)}$$

Example 4.7. *(e) Find k, if it is known that* $Pr[Z < k] = .8944$ *.*

This time, we need to use the table *backwards*. We look in the table until we find the value .8944, and then see what row and column it's in. We find .8944 in the 1.2 row and the .05 column, so we see that $.8944 = Pr[Z < 1.25]$, i.e. we have $k = 1.25$.

Note: $F(z)$ is, of course, a (strictly) increasing function, since $f(z)$ never reaches the horizontal axis. This means that if $b > a$, then $F(b) > F(a)$. As we read across each row of the table in turn, moving down the table, we see that each tabulated value is larger than the one before. The only exceptions are near the bottom of the table, where the increases in $F(z)$ are too small to be noticed when we round the value to 4 decimal places. Therefore, table values do not repeat, except when the same value appears as several consecutive entries (only near the bottom of the table).

This means that when you find a value in the table, as we did in part (e) here, this is the *only* place it appears in the table (although it may appear for a consecutive group of table entries). Remembering that the table values are increasing also, of course, facilitates finding a specific value among the table entries.

Example 4.7. *(f) Find k, where* $Pr[Z > k] = .3228$.

Remember, the table gives values of $Pr[Z < k]$, not $Pr[Z > k]$. We need to have something of the form $Pr[Z < k] = a$ before we can use the table (either to find a, when we know k, or to find k when we know a). Of course, we know that $Pr[Z < k] = 1 - Pr[Z > k]$, so if $Pr[Z > k] = .3228$, then we must have

$$Pr[Z < k] = 1 - .3228 = .6772$$

Now, we look for .6772 in the table. We see that $.6772 = Pr[Z < 0.46]$, so we have $k = .46$.

Example 4.7. *(g) Find k, where* $Pr[Z > k] = .9222$ *.*

If we try to do what we did in (f), we get

$$Pr[Z < k] = 1 - .9222 = .0778$$

However, .0778 doesn't appear in the table anywhere. Remember: since half the area lies to the left of 0, and the table only shows $Pr[Z < k]$ for $k \geq 0$, then all of the table values are 0.5 or bigger, that is there are no table entries which are less than .5.

The fact that $Pr[Z < k] < 0.5$ is telling us that k is negative. That is, less than half the area lies to the left of k, and we know that (a full) half of the area lies to the left of 0, so it must be true that k lies further left than 0. In fact, we could have realized that *before* we calculated $1 - .9222 = .0778$ – since $Pr[Z > k] > .5$, then *more than* half of the area under $f(z)$ lies to the right of k, so k must be to the left of 0, i.e. k must be negative.

Now that we realize that $k < 0$, we can find k using symmetry. We know that

$$Pr[Z > k] = Pr[Z < -k] \quad \text{so we have} \quad Pr[Z < -k] = .9222$$

(Of course, since we know that k is negative in this case, then $-k$ is positive.) From the table we find that $.9222 = Pr[Z < 1.42]$, so we get $-k = 1.42$ and therefore $k = -1.42$.

When we need to find a probability of the form $Pr[a < Z < b]$, we use the fact,that

$$Pr[a < Z < b] = Pr[Z < b] - Pr[Z < a]$$

Notice: What we're doing here has nothing to do with symmetry. This relationship is true for *any* continuous r.v., and in fact for *any r.v. X* if we write it as

$$Pr[a \leq X \leq b] = Pr[X \leq b] - Pr[X < a]$$

We have used this relationship before, for both discrete and continuous random variables. (Recall, we said that for any continuous r.v. X, $Pr[a < X < b] = F(b) - F(a)$. That's what we're using here. And we had a similar result for discrete r.v.'s, but it's more complicated because it requires "the next possible value down from a".)

Example 4.8. *Find each of the following values, where Z is the standard normal random variable: (a)* $Pr[0.50 < Z < 1.02]$

For this one, everything is straightforward.

$$\begin{aligned} Pr[0.50 < Z < 1.02] &= Pr[Z < 1.02] - Pr[Z < 0.50] \\ &= .8461 - .6915 \qquad \text{(from } Z\text{-table)} \\ &= .1546 \end{aligned}$$

Example 4.8. *(b)* $Pr[-0.50 < Z < 1.02]$

This will be slightly more complicated, but we just use methods we've already seen in Example 4.7.

$$Pr[-0.50 < Z < 1.02] = Pr[Z < 1.02] - Pr[Z < -0.50]$$

Since the table doesn't give $Pr[Z < -0.50]$, we use symmetry to express an event involving the corresponding positive Z-value. That is, we use the fact that $Pr[Z < -0.50] = Pr[Z > 0.50]$

$$Pr[-0.50 < Z < 1.02] = Pr[Z < 1.02] - Pr[Z > 0.50]$$

Now, however, we have 'greater than' form, instead of 'less than' form. We use the fact that $(Z > 0.50)$ is the complement of $(Z < 0.50)$. After that, everything is straightforward.

$$\begin{aligned} Pr[-0.50 < Z < 1.02] &= Pr[Z < 1.02] - Pr[Z < -0.50] \\ &= Pr[Z < 1.02] - (1 - Pr[Z < 0.50]) \\ &= Pr[Z < 1.02] + Pr[Z < 0.50] - 1 \\ &= .8461 + .6915 - 1 \\ &= 1.5376 - 1 \\ &= .5376 \end{aligned}$$

Note: For a calculation like this, i.e. any time we have $num_1 - (1 - num_2)$, the easiest way to do the arithmetic is always to add num_1 and num_2 and then just not 'carry the 1'. (Of course, that's only true when $num_1 + num_2$ is bigger than 1, but if this calculation is giving a probability, that has to be true.)

Example 4.8. *(c)* $Pr[-0.50 < Z < 0.50]$

Once again, we will need to use symmetry and also complementation.

$$\begin{aligned} Pr[-0.50 < Z < 0.50] &= Pr[Z < 0.50] - Pr[Z < -0.50] \\ &= Pr[Z < 0.50] - Pr[Z > 0.50] \qquad \text{(by symmetry)} \\ &= Pr[Z < 0.50] - (1 - Pr[Z < 0.50]) \qquad \text{(use complement)} \\ &= Pr[Z < 0.50] + Pr[Z < 0.50] - 1 \\ &= 2 \times Pr[Z < 0.50] - 1 \\ &= 2(.6915) - 1 \\ &= 1.3830 - 1 \\ &= .3830 \end{aligned}$$

Notice: It may be useful to remember that because of the symmetry of the normal distribution, for any number k, (following the same steps as above) we have:

$$Pr[-k < Z < k] = 2 \times Pr[Z < k] - 1$$

Example 4.8. *(d) Find* k*, if* $Pr[0.12 < Z < k] = .3992$

We have
$$Pr[0.12 < Z < k] = Pr[Z < k] - Pr[Z < 0.12] = Pr[Z < k] - .5478$$

So we know that

$$.3992 = Pr[Z < k] - .5478 \qquad \Rightarrow \qquad Pr[Z < k] = .3992 + .5478 = .9470$$

Now we look in the table for .9470 ... but it's not there. What's wrong? Nothing. We're in between the table values. We only have table values for k precise to 2 decimal places, not for all possible values (since there are infinitely many of them).

In the text book, they talk about how to estimate probabilities between the table values and also how to find the approximate value of k, correct to more than 2 decimal places, when we're looking for the value of k where $Pr[Z < k]$ is between 2 values in the Z-table. (The method used is called *interpolation*.) We don't bother with that in this course. That is, **we don't use interpolation in this course**. We just find the *nearest* tabulated value and use that to estimate k. Similarly, if we need to find $Pr[Z < k]$ for a k-value involving more than 2 decimal places, we simply round k to 2 decimals (using normal rounding – i.e. 5 or higher is rounded up, 4 or lower is rounded down) and use the corresponding probability as an estimate of the probability we're looking for.

Here, when we look in the table, we can't find .9470, but we can find 2 table entries on either side of it. The table tells us that

$$Pr[Z < 1.61] = .9463 \qquad \text{and} \qquad Pr[Z < 1.62] = .9474$$

Since .9474 is closer to .9470 than .9463 is (i.e. .9463 is .0007 away, whereas .9474 is only .0004 units away) then we say that $k \approx 1.62$.

Note: The fact that we found 2 consecutive table values lying on either side of .9470 means, of course, (since $F(z)$ is an increasing function) that there's no need to keep looking for this value in the table, because it won't be there.

Math 1228A/B Online

Lecture 30:

Other Normal Random Variables

(text reference: Section 4.3)

4.3 Normal Random Variables

There are a whole family of normally distributed r.v.'s, with different means and standard deviations. The mean can be any real number, and the standard deviation can be any positive number. In the previous section, we talked about a particular normal r.v., the Standard Normal random variable, which has mean $\mu = 0$ and standard deviation $\sigma = 1$.

Recall that the density function of any normal r.v. is a bell shaped curve. For a particular normal distribution, the value of the mean μ determines the line about which the bell-shaped curve is symmetric, i.e. at what value of the r.v. the peak of the bell occurs (see text p. 163, Figure 4.8). The standard deviation determines the height and thickness of the bell-shape. Of course, the total area under the curve is always 1. A small value of σ means that the distribution is less spread out, so more of this area (i.e. probability) is closer to the mean, causing the bell to be tall and skinny. A larger value of σ means the distribution is more spread out, so more of the area is further away from the mean, causing the bell to be more short and squat in shape (see text p. 163, Figure 4.9).

Normal random variables are the most commonly used among all continuous random variables. This is because:

1. normal r.v.'s occur naturally in many applications,
2. many r.v.'s which aren't actually normally distributed are *approximately* normally distributed, including many discrete r.v.'s,
3. normal r.v.'s are easy to use.

The thing that makes normal r.v.'s easy to use is the fact that any normally distributed random variable can be expressed in terms of the standard normal r.v. Z. This means that we can find probabilities involving *any* normal r.v. using the Z-table.

Fact: If X is normally distributed with mean μ and standard deviation σ, then

$$X = \sigma Z + \mu \quad \text{so that} \quad Z = \frac{X - \mu}{\sigma}$$

Example 4.9. *If X is a normal random variable, with $\mu = -2$ and $\sigma = .5$, express X in terms of the standard normal random variable Z.*

We have $\sigma = .5$ and $\mu = -2$, and we know that $X = \sigma Z + \mu$, so $X = .5Z - 2$.

Example 4.10. *If Y is a normal random variable, with $\mu = 500$ and $\sigma = 10$, express the standard normal random variable Z in terms of Y.*

We have:

$$Z = \frac{Y - \mu}{\sigma} = \frac{Y - 500}{10}$$

If we want to find probabilities involving a normal r.v., we simply convert the events to events in terms of Z and use the Z-table. If X is normally distributed with mean μ and standard deviation σ, then the event $(X < k)$ can be 'standardized' as follows:

$$(X < k) = (\sigma Z + \mu < k) = (\sigma Z < k - \mu) = \left(Z < \frac{k - \mu}{\sigma}\right)$$

(Remember, σ is always positive.) Therefore we see that for any normal r.v. X with mean μ and standard deviation σ,

$$Pr[X < k] = Pr\left[Z < \frac{k-\mu}{\sigma}\right]$$

Similarly, we have $Pr[X > k] = Pr\left[Z > \frac{k-\mu}{\sigma}\right]$. Taking these both together, we get

$$Pr[a < X < b] = Pr\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right]$$

Example 4.11. *X is a normal random variable with mean $\mu = 25$ and standard deviation $\sigma = 5$. (a) Find $Pr[X < 40]$.*

$$Pr[X < 40] = Pr\left[Z < \frac{40-\mu}{\sigma}\right] = Pr\left[Z < \frac{40-25}{5}\right] = Pr\left[Z < \frac{15}{5}\right] = Pr[Z < 3] = .9987$$

Example 4.11. *(b) Find $Pr[X < 20]$.*

$$\begin{aligned} Pr[X < 20] &= Pr\left[Z < \tfrac{20-25}{5}\right] && = Pr\left[Z < \tfrac{-5}{5}\right] \\ &= Pr[Z < -1] && = Pr[Z > 1] \quad \text{(By symmetry)} \\ &= 1 - Pr[Z < 1] && \quad \text{(Use the complement)} \\ &= 1 - .8413 && = .1587 \end{aligned}$$

Example 4.11. *(c) Find k such that $Pr[X < k] = .9990$.*

From the Z-table, we see that .9990 = ... uh-oh. It's there several times. We can use any of the z-values that give this probability. We would typically want to know the smallest value of k which provides the desired level of probability, so let's use the smallest one. We get $.9990 = Pr[Z < 3.08]$.

Of course, we know that $Pr[X < k] = Pr\left[Z < \frac{k-25}{5}\right]$. So we must have $\frac{k-25}{5} = 3.08$. So then $k - 25 = 5(3.08) = 15.4$ and we get $k = 15.4 + 25 = 40.4$. (That is, since $X = \sigma Z + \mu$, then $Z < 3.08 \Rightarrow X < 3.08\sigma + \mu \Rightarrow X < 5(3.08) + 25 = 40.4$.)

(*Check:* $Pr[X < 40.4] = Pr\left[Z < \frac{40.4-25}{5}\right] = Pr[Z < 3.08] = .9990$)

Example 4.11. *(d) Find $Pr[24 < X < 27]$.*

$$\begin{aligned} Pr[24 < X < 27] &= Pr\left[\frac{24-25}{5} < Z < \frac{27-25}{5}\right] \\ &= Pr\left[-\frac{1}{5} < Z < \frac{2}{5}\right] \\ &= Pr[-.2 < Z < .4] = Pr[Z < .4] - Pr[Z < -.2] \\ &= Pr[Z < .4] - Pr[Z > .2] \\ &= Pr[Z < .4] - (1 - Pr[Z < .2]) \\ &= Pr[Z < 0.40] + Pr[Z < 0.20] - 1 \\ &= .6554 + .5793 - 1 = .2347 \end{aligned}$$

Example 4.11. *(e) Find a probability expression in terms of* X *which corresponds to* $Pr[1 < Z < 2]$.

Since $Z = \frac{X-25}{5}$, then if $Z > 1$, we have

$$\frac{X-25}{5} > 1 \Rightarrow X - 25 > 5 \Rightarrow X > 30$$

(That is, since $= \sigma Z + \mu$, then $Z > 1 \Rightarrow X > 5(1) + 25 = 30$.)
Similarly, if $Z < 2$, we have

$$\frac{X-25}{5} < 2 \Rightarrow X - 25 < 10 \Rightarrow X < 35$$

(i.e. $Z < 2 \Rightarrow X < 2\sigma + \mu = 2(5) + 25 = 35$)
Thus we see that $(1 < Z < 2) = (30 < X < 35)$, so that $Pr[1 < Z < 2] = Pr[30 < X < 35]$. That is, we have:

$$\begin{aligned} Pr[1 < Z < 2] &= Pr\left[1 < \frac{X-25}{5} < 2\right] \\ &= Pr[5 < X - 25 < 10] \qquad \text{(multiply all through by 5)} \\ &= Pr[30 < X < 35] \qquad \text{(add 25 to all parts)} \end{aligned}$$

or $Pr[1 < Z < 2] = Pr[1\sigma + \mu < X < 2\sigma + \mu] = Pr[5 + 25 < X < 10 + 25] = Pr[30 < X < 35]$

Example 4.11. *(f) Find* a *such that* $Pr[22 < X < a] = .4514$.

$$\begin{aligned} Pr[22 < X < a] &= Pr\left[\frac{22-25}{5} < Z < \frac{a-25}{5}\right] = Pr\left[-\frac{3}{5} < Z < \frac{a-25}{5}\right] \\ &= Pr\left[Z < \frac{a-25}{5}\right] - Pr[Z < -0.60] = Pr\left[Z < \frac{a-25}{5}\right] - Pr[Z > 0.60] \\ &= Pr\left[Z < \frac{a-25}{5}\right] - (1 - Pr[Z < 0.60]) = Pr\left[Z < \frac{a-25}{5}\right] - (1 - .7257) \\ &= Pr\left[Z < \frac{a-25}{5}\right] - .2743 \end{aligned}$$

So we have $Pr\left[Z < \frac{a-25}{5}\right] - .2743 = .4514$, which gives

$$Pr\left[Z < \frac{a-25}{5}\right] = .4514 + .2743 = .7257$$

But we already used the fact that $.7257 = Pr[Z < 0.60]$, so we get

$$\begin{aligned} & Pr\left[Z < \frac{a-25}{5}\right] = Pr[Z < 0.60] \\ \Rightarrow \quad & \frac{a-25}{5} = 0.60 \\ \Rightarrow \quad & a - 25 = 5(0.60) = 3.0 \\ \Rightarrow \quad & a = 3 + 25 = 28 \end{aligned}$$

(That is, $Z < .60 \Rightarrow X < .6\sigma + \mu = (.6)(5) + 25 = 3 + 25 = 28$.)

Check: Find $Pr[22 < X < a]$ when $a = 28$:

$$\begin{aligned}
Pr[22 < X < 28] &= Pr\left[\frac{22-25}{5} < Z < \frac{28-25}{5}\right] = Pr\left[-\frac{3}{5} < Z < \frac{3}{5}\right] \\
&= 2Pr\left[Z < \frac{3}{5}\right] - 1 \qquad \text{(using the useful formula from Lecture 29, pg. 176)} \\
&= 2Pr[Z < 0.60) - 1 = 2(.7257) - 1 = .4514
\end{aligned}$$

We see that using $a = 28$ does give the required value for $Pr[22 < X < a]$.

Example 4.12. *The weight of ball bearings made by a certain manufacturing process is normally distributed with mean 75 mg and standard deviation 4 mg. What percentage of ball bearings weigh more than 80 mg?*

Let X be the weight of a ballbearing made by this process. Then X is a normal r.v. with $\mu = 75$ and $\sigma = 4$. The percentage of ball bearings that weigh more than 80 mg is simply the probability that $X > 80$. We get

$$\begin{aligned}
Pr[X > 80] &= Pr\left[Z > \frac{80-\mu}{\sigma}\right] \\
&= Pr\left[Z > \frac{80-75}{4}\right] \\
&= Pr\left[Z > \frac{5}{4}\right] \\
&= Pr[Z > 1.25] \\
&= 1 - Pr[Z < 1.25] \\
&= 1 - .8944 \\
&= .1056
\end{aligned}$$

We see that 10.56% of ball bearings made by this process weigh more than 80 mg.

Example 4.13. *The Thread-Easy Company manufactures spools of thread. The actual length of the thread on a spool labelled '200 metres' is normally distributed, with standard deviation .5 metres. What should the mean thread length be in order to ensure that less than 1% of all '200 m' spools actually hold less than 200 metres of thread?*

Let X be the length of the thread on a '200 m' spool. Then X is normally distributed with standard deviation $\sigma = .5$ and unknown mean μ. We wish to ensure than less than 1% of all such spools hold less than 200 m of thread, i.e. that

$$Pr[X < 200] < .01$$

Standardizing, we see that we need

$$Pr\left[Z < \frac{200-\mu}{.5}\right] < .01$$

Since this probability is less than .5, we see that $\frac{200-\mu}{.5}$ is negative, i.e. that μ is bigger than 200. (Of course it is. We know, because of symmetry, that $Pr[X < \mu] = .5$, so if μ was smaller than 200, then more than half of the spools produced would hold less than 200 m of thread.)

Now that we realize that $\frac{200-\mu}{.5}$ is negative, we use symmetry to see that we need

$$Pr\left[Z > -\left(\frac{200-\mu}{.5}\right)\right] < .01 \qquad \Rightarrow \qquad Pr\left[Z > \frac{\mu - 200}{.5}\right] < .01$$

Now, looking at the complementary event, we have

$$Pr\left[Z < \frac{\mu - 200}{.5}\right] > 1 - .01 = .99$$

From the Z-table, we see that $Pr[Z < k] > .99$ when $k \geq 2.33$ (i.e. the smallest value of k for which $Pr[Z < k] > .99$ is $k = 2.33$), so we have

$$\frac{\mu - 200}{.5} \geq 2.33 \qquad \Rightarrow \qquad \mu - 200 \geq 2.33(.5) = 1.165 \qquad \Rightarrow \qquad \mu \geq 201.165$$

Therefore the mean length of the thread should be at least 201.165 metres.

Check: Let $\mu = 201.165$ and find $Pr[X < 200]$ to check that it is less than .01

$$\begin{aligned} Pr[X < 200] &= Pr\left[Z < \frac{200 - 201.165}{.5}\right] = Pr[Z < -2.33] \\ &= Pr[Z > 2.33] = 1 - Pr[Z < 2.33] = 1 - .9901 = .0099 \end{aligned}$$

Thus if the mean thread length is 201.165 m, the requirement that less than 1% of all spools contain less than 200 m is satisfied.

Math 1228A/B Online

Lecture 31:

Discrete Random Variables Which Are Approximately Normal

(text reference: Section 4.4, pgs. 173 - 174)

4.4 Approximately Normal Discrete Random Variables

(*Note:* That's not the title the text uses for this section.)

Many discrete r.v.'s are approximately normal. What does that mean?

> *Definition:* If discrete r.v. X, with mean μ and standard deviation σ, is **approximately normal**, then the continuous random variable Y which is normally distributed, with the same mean and standard deviation as X, is a good approximation for X.

Of course, if we are approximating a discrete r.v. X by a continuous r.v. Y, then we must use the approximation techniques we learned in section 4.1, and apply a continuity correction. (*Notice:* in order for a discrete random variable X to be approximately normal, it must be true that the possible values of X are evenly spaced, since that is required in order to be able to approximate a discrete r.v. using a continuous r.v.) Therefore, finding probabilities of events involving a discrete r.v. X which is approximately normal involves 2 steps.

Given: Discrete r.v. X which is approximately normal with mean μ and standard deviation σ

Step 1: Apply the continuity correction to approximate X by Y, the normal r.v. with the same mean and standard deviation as X

Step 2: Standardize, i.e. convert events in terms of normal r.v. Y into events in terms of the Standard Normal r.v. Z

For instance, to find $Pr[X \leq a]$, where X is a discrete r.v. which is approximately normally distributed with mean μ and standard deviation σ, whose possible values are evenly spaced k units apart, we have

$$Pr[X \leq a] = Pr[Y < a + \frac{k}{2}]$$

(where Y is the normal r.v. with mean μ and standard deviation σ)

$$= Pr\left[Z < \frac{\left(a + \frac{k}{2}\right) - \mu}{\sigma}\right]$$

(where Z is the standard normal random variable)

Example 4.14. *The number of candies in a large bag is approximately normal with mean 100 and standard deviation 5. What is the probability that a bag contains fewer than 90 candies?*

Let X be the number of candies in a bag. Then the possible values of X are consecutive integers, i.e. we're counting the candies, and so the possible values are integers, i.e. are $k = 1$ unit apart (so X is a discrete r.v.). Thus we know that X is a discrete r.v. which (we are told) is approximately normal, with mean $\mu = 100$ and standard deviation $\sigma = 5$. We are looking for $Pr[X < 90] = Pr[X \leq 89]$.

Notice: Any time that X is defined to be "the number of" something, X can only have integer values and **must** be a discrete random variable. (This is not the same as measuring the *amount* of something.)

Let Y be the normal r.v. which has the same mean and standard deviation as X, i.e. with $\mu = 100$ and $\sigma = 5$. Then, since X is approximately normal, we know that Y is a good approximation for

X. We approximate the event $(X \le 89)$ by the event $\left(Y < 89 + \frac{k}{2}\right) = \left(Y < 89 + \frac{1}{2}\right)$, so we have

$$\begin{aligned}
Pr[X < 90] &= Pr[X \le 89] \\
&= Pr[Y < 89.5] \\
&= Pr\left[Z < \frac{89.5 - 100}{5}\right] \\
&= Pr\left[Z < -\frac{10.5}{5}\right] \\
&= Pr[Z < -2.10] \\
&= Pr[Z > 2.10] \\
&= 1 - Pr[Z < 2.10] \\
&= 1 - .9821 \\
&= .0179
\end{aligned}$$

We see that only about 1.8% of all bags contain fewer than 90 candies.

Example 4.15. *A variety store sells a daily newspaper, for which the demand is approximately normal with mean 20.1 copies and standard deviation 3.9 . If the store stocks 22 copies each day, what is the probability that at least one customer will be disappointed on any given day because the store has run out of papers?*

We need to find a probability having to do with 'at least one customer will be disappointed'. What does that mean? A customer is disappointed if he or she goes to the store to buy a newspaper but cannot, because the store has run out. We are told that 'demand' for the papers is approximately normal. The 'demand' for newspapers is not the number that customers *actually* buy, it's the number that customers *wish to* buy. There is at least one disappointed customer on any day on which demand exceeds supply, i.e. customers wish to buy more newspapers than the store has in stock.

Let X be the number of newspapers which customers wish to buy at the store on a particular day, i.e. the demand for newspapers that day. Of course, the number of newspapers which customers wish to buy *must be an integer*, so X is a discrete random variable. We are told that demand is approximately normal, so X is a discrete r.v. which is approximately normal, with mean $\mu = 20.1$ and standard deviation $\sigma = 3.9$.

We need to determine the probability that at least 1 customer will be disappointed (because they come into the store to buy the newpaper but the store has run out), on a day when the store has stocked 22 copies of the paper. That is, we want to find the probability that the demand is *more than 22* (i.e. more than 22 customers come in to buy the paper).

Therefore we want to find $Pr[X > 22] = Pr[X \ge 23]$. Of course, we could also view this as $Pr[X > 22] = Pr[(X \le 22)^c]$, so we are looking for $Pr[X > 22] = 1 - Pr[X \le 22]$.

Since X is approximately normal, let Y be the normal r.v. with the same mean and standard deviation as X. Then Y has mean $\mu = 20.1$ and standard deviation $\sigma = 3.9$ and Y is a good approximation for X. Since X is counting the number of customers who want to buy the paper, then the possible values of X are consecutive integers, i.e. are evenly spaced $k = 1$ unit apart, so the continuity correction we need to apply is $\frac{k}{2} = \frac{1}{2}$ and we have

$$(X \le 22) \approx (Y < 22 + 1/2) = (Y < 22.5)$$

so we get:

$$\begin{aligned}
Pr[X > 22] &= 1 - Pr[X \le 22] \\
&= 1 - Pr[Y < 22.5] \quad \text{(apply continuity correction)} \\
&= 1 - Pr\left[Z < \frac{22.5 - 20.1}{3.9}\right] \quad \text{(standardize)} \\
&= 1 - Pr\left[Z < \frac{2.4}{3.9}\right] \\
&= 1 - Pr[Z < 0.62] \quad \text{(round to 2 decimal places)} \\
&= 1 - .7324 \quad \text{(from } Z\text{-table)} \\
&= .2676
\end{aligned}$$

That is, the probability of having at least one disappointed customer on any given day is about .2676, when the store stocks 22 newspapers .

Example 4.16. *How many copies of the newspaper should the variety store in the Example 4.15 stock each day if the owner wishes to have no disappointed customers at least 90% of the time?*

Recall what we had in Example 4.15: demand for newspapers is approximately normal with mean $\mu = 20.1$ and standard deviation $\sigma = 3.9$. As before, we define X to be the number of newspapers demanded on a particular day, and Y to be the normal r.v. with mean $\mu = 20.1$ and standard deviation $\sigma = 3.9$ so that the continuous r.v. Y is a good approximation for the discrete r.v. X.

In order to have no disappointed customers on a given day, the store must have enough newspapers in stock to satisfy all demand. We need to determine how many newspapers the store should stock in order for this to be true at least 90% of the time. That is, we need to find the value a such that the probability that no more than a newspapers are demanded on any day is at least .9, so we want to find a such that $Pr[X \le a] \ge .9$.

As before, since the possible values of X are consecutive integers, the continuity correction we require is 1/2, so we get

$$(X \le a) \approx (Y < a + .5) = \left(Z < \frac{(a + .5) - 20.1}{3.9}\right)$$

So we need $Pr\left[Z < \frac{(a+.5)-20.1}{3.9}\right] \ge .9$. From the Z-table, we see that the first Z-value for which the cumulative probability is at least .9 is 1.29. That is, the first number at least as big as .9 that we find in the table is $.9015 = Pr[Z < 1.29]$. Therefore, we need the z-value, $\frac{a+.5-20.1}{3.9}$, to be at least 1.29. We have

$$\begin{aligned}
\frac{(a + .5) - 20.1}{3.9} \ge 1.29 \quad &\Rightarrow \quad a + .5 - 20.1 \ge (1.29)(3.9) \\
&\Rightarrow \quad a - 19.6 \ge 5.031 \\
&\Rightarrow \quad a \ge 5.031 + 19.6 = 24.631
\end{aligned}$$

Of course, the store must stock an integer number of newspapers, so we need to round this up to the next whole number. We see that the store should stock at least 25 newspapers in order to have no disappointed customers at least 90% of the time.

Notice: We would *always* round up in this situation, never down, even if the lower bound on a was closer to the rounded down value. (For instance, even if we found that we required something like $a \ge 24.012$, we would conclude that a must be at least 25, because 24 is *not* greater or equal to

24.012.) We can check that we've rounded the right way by making sure that the probability of having no disappointed customers really is at least 90%. We have

$$\begin{aligned} Pr[X \leq 25] &\approx Pr[Y < 25.5] \\ &= Pr\left[Z < \frac{25.5 - 20.1}{3.9}\right] \\ &= Pr[Z < 1.38] \\ &= .9162 \end{aligned}$$

Math 1228A/B Online

Lecture 32:

$B(n, p)$ is often Approximately Normal

(text reference: Section 4.4, pgs. 169 - 173)

Approximating the Binomial Distribution

Recall: A random variable X which counts the number of successes in a series of n Bernoulli trials in which the probability of success is p is a **Binomial r.v.**, $B(n, p)$, with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$.

Of course, we have $Pr[B(n,p) = k] = \binom{n}{k}p^k(1-p)^{n-k}$. And to find the cdf of $B(n,p)$, we have to add up a bunch of these. That is, if $X = B(n,p)$, then $Pr[X \leq x]$ is found by summing $Pr[X = k]$ for all $k \leq x$. When n is small, we can find values of the cdf fairly easily. But as n gets large, these probabilities become very obnoxious to calculate. Fortunately, as n becomes larger, the binomial distribution becomes more symmetric, with the following result:

Fact: If n is ***sufficiently large*** and p is ***not too close to 0 or 1***, then the Binomial r.v. $B(n,p)$ is approximately normal.

How large is "sufficiently large" for the number of trials? How close to 0 or 1 is "too close" for the probability of success? It's hard to say precisely, because n and p interact. However, we have a guideline that we can use to determine whether or not a particular binomial distribution is approximately normal:

Rule of Thumb: If $np > 5$ and also $n(1-p) > 5$, then $B(n,p)$ is approximately normal.

That is, as long as both the expected number of successes and the expected number of failures are bigger than 5, then the normal r.v. with mean np and standard deviation $\sqrt{np(1-p)}$ is a good approximation for $B(n,p)$.

Notice: Since the possible values of $B(n,p)$ are (all) the integers from 0 to n, so they are always evenly spaced 1 unit apart, approximating $B(n,p)$ by a normal r.v. always requires a continuity correction of $\frac{1}{2}$.

Example 4.17. *A pair of dice is tossed 180 times.*
(a) What is the probability that the sum on the two dice is 7 exactly 30 times?
(b) What is the probability that the sum is 7 fewer than 25 times?
(c) What is the probability that the sum is 7 at least 25 times, and not more than 35 times?

We are performing the experiment 'Toss a pair of dice 180 times', so we are performing $n = 180$ Bernoulli trials. We are interested in occurrences of the event 'the sum is 7'. Defining this event to be a success, the probability of success is $p = \frac{6}{36} = \frac{1}{6}$. (Recall: Considering the sample space S to be the set of all pairs of numbers which could be tossed, S is an equiprobable sample space with $n(S) = 36$. The event 'sum is 7' contains the 6 sample points (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1).)

Letting X be the number of times that the sum is 7, X is counting the number of successes in these 180 Bernoulli trials, so X is a Binomial r.v., $B(180, 1/6)$, which has mean $\mu = (180)(1/6) = 30$ and standard deviation $\sigma = \sqrt{(180)(1/6)(5/6)} = \sqrt{(30)(5/6)} = \sqrt{25} = 5$.

(a) We need to find the probability of observing exactly 30 successes. We can find $Pr[X = 30]$ using the Bernoulli formula. The probability that $B(180, 1/6)$ is exactly 30 is given by

$$Pr[X = 30] = \binom{180}{30}\left(\frac{1}{6}\right)^{30}\left(\frac{5}{6}\right)^{150} = .07955977$$

(b) Now, we need to find the probability of observing fewer than 25 successes. In order to find $Pr[X < 25] = Pr[X \leq 24]$ using the Binomial distribution itself, we would have to use the Bernoulli formula 25 times, for $k = 0, 1, ..., 24$ and add them all up. However, we have $np = (180)(1/6) = 30 > 5$ and $n(1-p) = (180)(5/6) = 150 > 5$, so according to our rule of thumb, X is approximately normal. Therefore we can approximate probabilities involving X using the normal r.v. Y with mean $\mu = 30$ and standard deviation $\sigma = 5$.

Before we do the calculation required for part (b), let's see just how good an approximation we get, by seeing what answer a normal approximation gives us for part (a).

(a) revisited:
We want to approximate $Pr[X = 30]$ using the normal r.v. Y with $\mu = 30$ and $\sigma = 5$. Of course, since X is a discrete r.v. and we are going to approximate it by a continuous r.v., we must apply a continuity correction. As we observed previously, for any binomial r.v., the possible values are consecutive integers, so we have $k = 1$ and $k/2 = 1/2$. We get:

$$\begin{aligned} Pr[X = 30] &= Pr[30 - 1/2 < Y < 30 + 1/2] = Pr[29.5 < Y < 30.5] \\ &= Pr\left[\frac{29.5 - 30}{5} < Z < \frac{30.5 - 30}{5}\right] = Pr\left[-\frac{.5}{5} < Z < \frac{.5}{5}\right] \\ &= Pr[-0.10 < Z < 0.10] = 2 \times Pr[Z < 0.10] - 1 = 2(.5398) - 1 = .0796 \end{aligned}$$

We see that the approximation is very very good. (If we round the precise value of $Pr[X = 30]$ we found to 4 digits, we get the same value as the approximation gives.)

Back to (b):
To find $Pr[X < 25]$, we have

$$\begin{aligned} Pr[X < 25] &= Pr[X \leq 24] = Pr[Y < 24 + 1/2] = Pr\left[Z < \frac{24.5 - 30}{5}\right] \\ &= Pr\left[Z < -\frac{5.5}{5}\right] = Pr[Z < -1.1] \\ &= 1 - Pr[Z < 1.1] = 1 - .8643 = .1357 \end{aligned}$$

(c) We need to find the probability that the sum is 7 at least 25 times and not more than 35 times, so we need $X \geq 25$ and also $X \leq 35$. We get:

$$\begin{aligned} Pr[25 \leq X \leq 35] &= Pr[24.5 < Y < 35.5] \\ &\quad \text{(extend the interval by 1/2 in each direction)} \\ &= Pr\left[\frac{24.5 - 30}{5} < Z < \frac{35.5 - 30}{5}\right] = Pr\left[-\frac{5.5}{5} < Z < \frac{5.5}{5}\right] \\ &= Pr[-1.1 < Z < 1.1] = 2(Pr[Z < 1.1]) - 1 = 2(.8643) - 1 = .7286 \end{aligned}$$

Let's look at one more example.

Example 4.18. *A certain manufacturing process produces items which have a probability of .05 of being defective.*
(a) What is the probability that a batch of 1000 items contains more than 55 defective items?
(b) What is the probability that the batch contains at least 35 defective items?

Each item produced by this process has probability .05 of being defective. Therefore, producing a batch of 1000 items corresponds to performing $n = 1000$ independent trials in which the probability of success, i.e. having the item be defective, is $p = .05$. Defining X to be the number of defective items in the batch, we see that X is a Binomial r.v., $B(1000, .05)$. The mean is $\mu = (1000)(.05) = 50$ and the standard deviation is $\sigma = \sqrt{(1000)(.05)(.95)} = \sqrt{(50)(.95)} = \sqrt{47.5} \approx 6.892$. Notice that we have $np = 50$ and $n(1-p) = (1000)(.95) = 950$, so X is approximately normal. Let Y be the normal r.v. with the same mean and standard deviation as X. As always with binomial r.v.'s, we have the possible values of X being consecutive integers, so the continuity correction is 1/2.

(a) We need to find the probability that there are more than 55 defective items in the batch, i.e. $Pr[X > 55]$. We get:

$$\begin{aligned} Pr[X > 55] &= 1 - Pr[X \leq 55] = 1 - Pr[Y < 55.5] \\ &= 1 - Pr\left[Z < \frac{55.5 - 50}{\sqrt{47.5}}\right] \approx 1 - Pr[Z < 0.80] \\ &= 1 - .7881 = .2119 \end{aligned}$$

We see that there is about a 21% chance that there will be more than 55 defective items in the batch.

(b) We want to know the probability that the batch will contain at least 35 defective items, i.e. $Pr[X \geq 35]$. We get:

$$\begin{aligned} Pr[X \geq 35] &= Pr[35 \leq X] = Pr[34.5 < Y] \\ &\quad \text{(i.e. push the lower bound to the left by 1/2)} \qquad * \\ &= Pr[Y > 34.5] = Pr\left[Z > \frac{34.5 - 50}{6.892}\right] \\ &= Pr[Z > -2.25] = Pr[Z < 2.25] \\ &= .9878 \end{aligned}$$

We see that almost 99% of batches of 1000 items will contain at least 35 defective items.

* We can also get the approximating event in the following way:
We need

$$(X \geq 35) = (X < 35)^c$$

so we have

$$Pr[X \geq 35] = 1 - Pr[X < 35] = 1 - Pr[X \leq 34]$$

and this is approximated by

$$1 - Pr[Y < 34.5] = Pr[Y > 34.5]$$