STAT*2060DE
Summary Notes for Unit 2: Probability

An understanding of basic probability is required in order to carry out statistical inference techniques and properly interpret the results. We also need a knowledge of basic probability in order to make good decisions in other areas. For example, suppose a supplier claims that on average only 2% of his parts have substantial manufacturing defects. If 8 out of the first 10 parts supplied have substantial manufacturing defects we may legitimately have reason to doubt his claim. We may ask ourselves, "how likely is that to occur if his claim is true?" In some situations we'll use probability to give us a measure of the strength of the evidence for or against a claim.

Many students find the probability section of the course tough. Most students do not acquire a lot of information simply by reading the material; students should do many practice questions to engrain some of the main concepts.

Two basic probability examples:

Suppose we toss two coins. What is the probability both come up heads?
Suppose two dice are rolled. What is the probability they are both an even number, but not equal?

These are two examples of **experiments** (in probability theory, which is a little different from our earlier experiment and observational study discussion)**.**

Some definitions:

In a **probability experiment**, individual outcomes cannot be predicted with certainty, but over a large number of trials the outcomes will follow a distribution.

Depending on one's philosophy, there are different interpretations of probability. One common one is the *frequentist* perspective, in which the **probability** of an outcome is the proportion of times that outcome would occur in a very long (infinite) series of trials. We'll typically look at probability from this perspective, but it is not without its problems. For instance, we are not about to have an infinite number of trials in any experiment. And suppose I were to ask the question, “what is the probability that a meteorite falls on your head in the next 10 seconds?” It's tough to view this type of probability question from a frequentist perspective, as that period of time will never be repeated under those exact conditions again.

The **Sample Space** of an experiment is the set of all possible outcomes.
Simple events (also known as **sample points**) are the most basic outcome of the experiment.

Examples:

If we roll an ordinary six-sided die, the sample space is $S = \{1,2,3,4,5,6\}$. These 6 possibilities are the simple events.

If we draw a card from a 52 card deck, the sample space would consist of the 52 cards. (52 simple events)

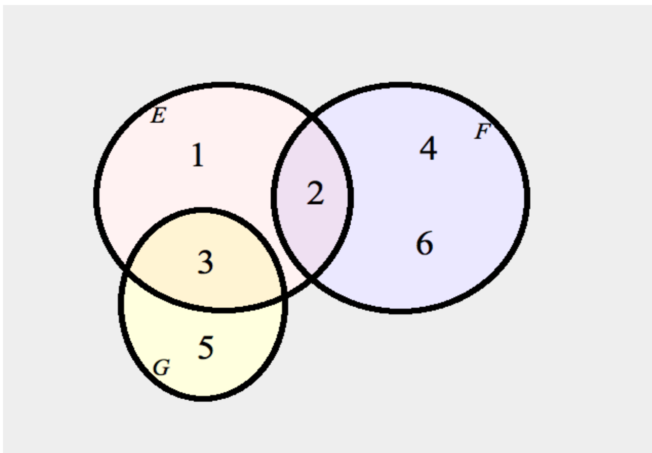
If we toss a coin twice, the sample space is S = {TT, TH, HT, HH}   (4 simple events)
(where TH represents tails on the first toss, heads on the second).

An **event** is a subset of the sample space. We could alternatively define an event as a collection of simple events.

For example, consider the following 3 events for the roll of an ordinary six-sided die:

Rolling a one, two, or three: $E = \{1,2,3\}$
Rolling an even number. $F = \{2,4,6\}$
Rolling a 3 or a 5. $G = \{3,5\}$

To help with visualizing the various relationships, it is often convenient to draw a Venn diagram:

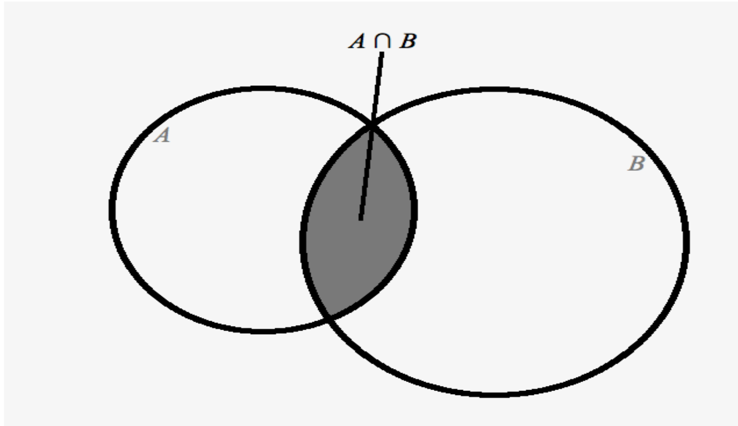


If the die is perfectly balanced, then each of these six outcomes is equally likely to occur. Of course it's not likely that a die is *perfectly* balanced, but it provides a reasonable model for us in many situations. If all of the simple events are equally likely, then the probability of an event A is:

$$P(A) = \frac{\text{Number of simple events that make up } A}{\text{Total number of simple events}}$$

For example, in the die example above Event E is made up of 3 simple events (1, 2, 3), Event F is made up of 3 simple events (2, 4, 6), and Event G is made up of 2 simple events (3, 5). Since all of the simple events are equally likely: P(E) = 3/6 = ½, P(F) = 3/6 = ½, P(G) = 2/6 = 1/3

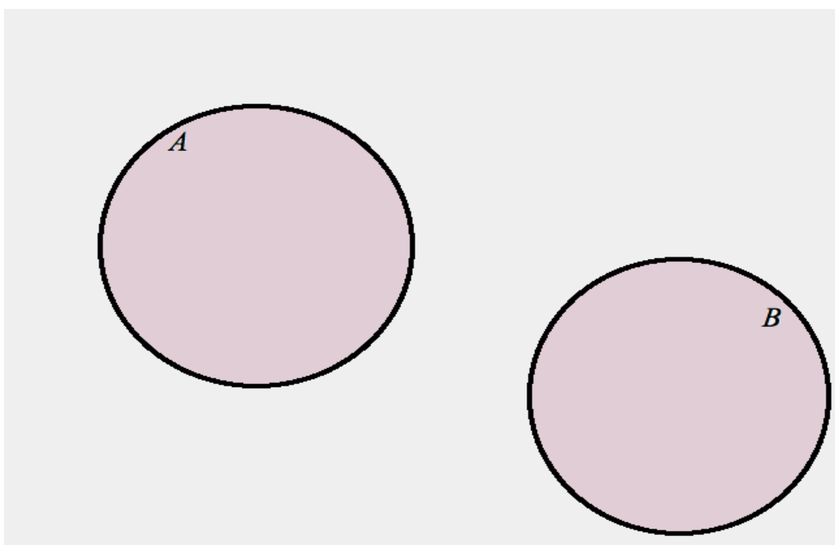
The **intersection** of events A and B is the event that both A and B occur.
This is denoted by $A \cap B$, A and B, or simply AB.



$E \cap F$ = {2} (Only the number 2 is contained in both E and F) $P(E \cap F) = 1/6$.
$E \cap G$ = {3} (Only the number 3 is contained in both E and G) $P(E \cap G) = 1/6$.
$F \cap G$ = {} (The empty set – there are no observations in both both F and G) $P(F \cap G) = 0$.

Events are **mutually exclusive (disjoint)** if they have no events in common. In other words, they cannot both occur on the same trial of an experiment.

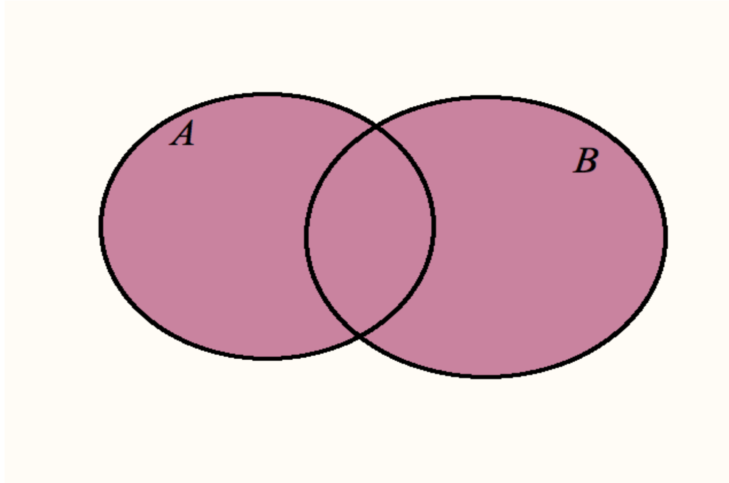
Two events A and B are mutually exclusive if $P(A \cap B) = 0$.



Here A and B are mutually exclusive, since they have no intersection.

In the die example above, events F and G are mutually exclusive; they cannot both occur on the same trial of the experiment.

The **union** of events A and B is the event that either A or B or both occur.
The union is denoted by $A \cup B$, or sometimes simply *A or B*.



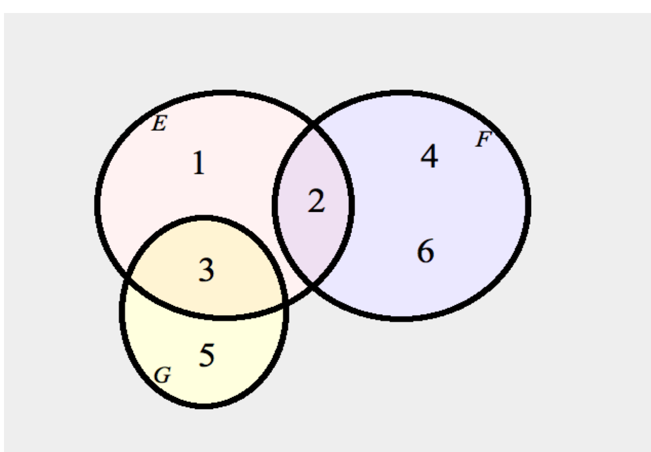
For example, consider the die example above. $E = \{1,2,3\}$, $F = \{2,4,6\}$. The union of E and F is the set of all simple events that occur in E, F, or both. Since the numbers 1, 2, 3, 4, 6 all occur in E or F or both, $E \cup F = 1,2,3,4,6$, and $P(E \cup F) = \frac{5}{6}$.

When we want to find the probability of the union of two events, we can use the addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Note that if we simply added $P(A) + P(B)$, we'd be adding the intersection twice. Since we wish to include the intersection only once, it needs to be subtracted.

Returning to the example, $P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$, which is what we obtained above using a little thought.

The **complement** of an event A, denoted by A^c, is the event that A does not occur. $P(A^c) = 1 - P(A)$

Sometimes we have additional information available to us, and we wish to use this information to update our original probabilities. In **conditional probability**, we know (or are assuming) that a certain event has occurred, and are interested in the probability of another event occurring.
The conditional probability of A, given B, is $P(A|B) = \frac{P(A \cap B)}{P(B)}$, provided P(B) > 0 .
Let's return to the roll of a single die and look at some examples.



Suppose that we are betting $1 on the outcome of the die roll. If we are betting on outcome *E*, then we want a 1, 2 or 3 to occur, and *P(E)* = 3/6 = 1/2. But suppose we are given additional information. Suppose a friend of ours has seen the die, and he remembers only that the die has even number on the up face. Given this information, we know that the number is a 2, 4, or 6. What is the probability that event *E* has occurred, given this information? Well, the only one of these 3 events that result in *E* is the number 2. So, P(E|F) = 1/3 (one of the 3 outcomes result in *E*). This is the *conditional probability* of *E*, given that *F* has occurred.

To find this conditional probability, we could also have used our conditional probability formula: $P(E|F)=\frac{P(E\cap F)}{P(F)}=\frac{1/6}{1/2}=\frac{1}{3}$. The conditional probability formula gives us the same answer we found above using our logic skills.

If we are interested in the conditional probability of event *E*, given event *G*: $P(E|G)=\frac{P(E\cap G)}{P(G)}=\frac{1/6}{1/3}=\frac{1}{2}$. This brings us to a very important concept. Note that P(E|G) = P(E). In other words, knowing that Event *G* has occurred has not changed the probability of event *E*. We say that events *E* and *G* are **independent**. The concept of independence is an extremely important concept in probability and statistics.

If the occurrence or nonoccurrence of *A* does not change the probability of *B*, we say that *A* and *B* are **independent.** More formally, *A* and *B* are independent if and only if

P(A|B) = P(A)
P(B|A) = P(B)
P(A∩B) = P(A)P(B)

Each of these conditions implies the others, so it is only necessary to check if one of them is true. They can be used as a check for independence, or as a tool in calculating probabilities if we know two events are independent. THEY ARE ONLY TRUE IF THE TWO EVENTS ARE INDEPENDENT.

In the die example, P(*E*|*G*) = P(*E*) and thus *E* and *G* are independent events. If I'm betting on *E* occurring (the number is a 1, 2, or 3), and you tell me that you know the number is a 3 or a 5, then this is meaningless to me. My chances of winning are still the same, ½. However, if you tell me that event *F* has occurred (the number is a 2, 4, or 6), then *P(E|F) = 1/3,* which does not equal P(*E*) = ½. I am now less likely to win my bet. My original probability of ½ has changed, and thus events *E* and *F* are **not independent** (which we could, of course, also call **dependent**).

When we need to a find the probability of the intersection of two events, we can rearrange our conditional probability formula, leading to the **multiplication** rule:

$$P(A\cap B)=P(A)P(B|A)=P(B)P(A|B)$$

N.B. At some point you are going to have to use your logic skills, think a little bit, and find a probability based on the information that you are given. If you are throwing numbers haphazardly into different formulas, you will end up going around in circles.

Example

For two events A and B, P(A) = .70, P(B) =.30, and $P(A \cup B) = .8$. What is the probability of A, given that B has occurred?

The conditional probability formula tells us that $P(A \cap B) = \frac{P(A \cap B)}{P(B)}$. We require P(B), which is given, and $A \cap B$, which is not given directly.

Tempting as it may be to some, we cannot say that P(A∩B) = P(A)P(B), as this is only true of independent events. We are not told that A and B are independent, nor do we have any reason to believe that A and B are independent. We must find the probability of the intersection in another way. There is no way of finding the probability of the intersection from the individual probabilities (unless the two events are independent).

If we use the addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
This implies that .8 = .7 + .3 - P(A∩B), which implies that P(A∩B) = 0.2. Now we're home free:

$$P(A \cap B) = \frac{P(A \cap B)}{P(B)} = \frac{.2}{.3} = \frac{2}{3}$$

Note that this does not equal P(A), and thus A and B are not independent.

Example

The following table is based reasonably accurately on an August 2009 poll in the U.S., in which 1000 randomly selected adults were asked if they approved or disapproved of the way Barack Obama is handling his job as President.

	Republicans	Democrats	Independents
Approve	60	361	139
Disapprove	270	49	121

If a person is randomly selected from the above group, what is the probability that they are Republican? There are 60+70 = 330 Republicans in the sample of 1000 individuals, and thus P(Republican) = 330/1000 = .330.

What is the probability that they approve of the way Obama is handling his job?
P(Approve) = (60+361+139)/1000 = .560.

Given a randomly selected person is Republican, what is the probability they approve?

We can use the conditional probability formula for this of course, but sometimes we can answer questions like this quickly using a little bit of logic. Here we are given that the individual is a

Republican, which means we are dealing only with the 330 Republicans in the sample. Of the 330 Republicans, 60 approve, and so P(Approve | Republican) = 60/330. We can always use the conditional probability formula is we desire:

$$P(Approve|Republican)=\frac{P(Approve\cap Republican)}{P(Republican)}=\frac{60/1000}{330/1000}=\frac{60}{330}$$

, which is what we found using a little logic.

Note that P(Approve) ≠ P(Approve|Republican). This implies the events: *The person is a Republican*, and *They approve of the way Obama is handling his job as President* are not independent (which should make sense, if you follow US politics at all).

For the following pairs of events, choose whether P(B|A) = P(B), P(B|A) > P(B), or P(B|A) < P(B). These examples do not have probabilities given, they just require a little thought.

A: It rains tomorrow. *B*: It rains the day after tomorrow.

P(B|A) > P(B). A and B are not independent. Weather systems can be large, and take several days to pass through. If it rains tomorrow, it is more likely to rain the day after than if there is bright sunshine tomorrow.

C: It rains Sunday. *D:* It rains 3650 days from Sunday.
P(C|D) = P(C) (essentially). These events are independent. Knowing the weather on Sunday will tell us nothing about the weather 10 years from now.

For a randomly picked adult:
E: The person did not finish high school. *F*: The person has been to prison.

These events are not independent. People who don't finish high school are more likely to go to prison.
P(F|E) > P(F)

Example

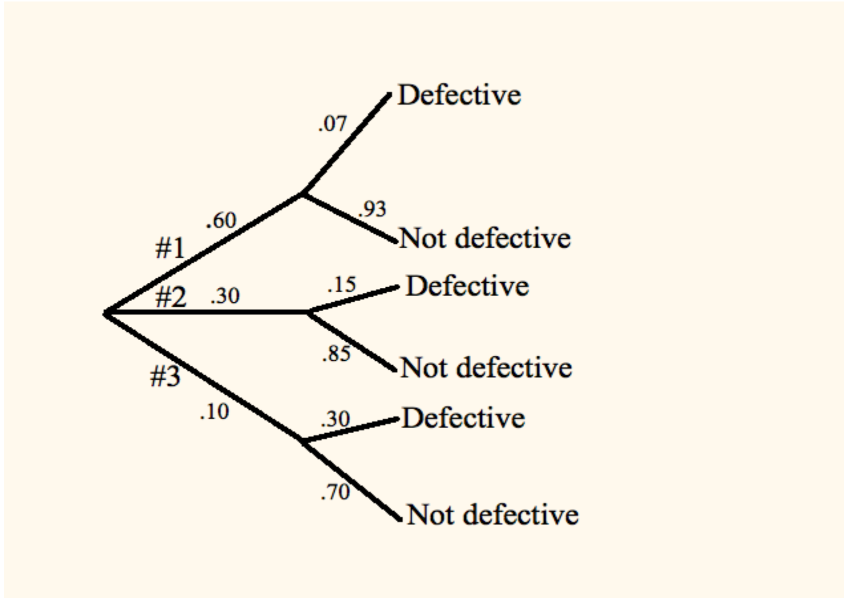
Three machines make parts at a factory.

Machine #1 makes 60% of the parts
Machine #2 makes 30% of the parts
Machine #3 makes 10% of the parts

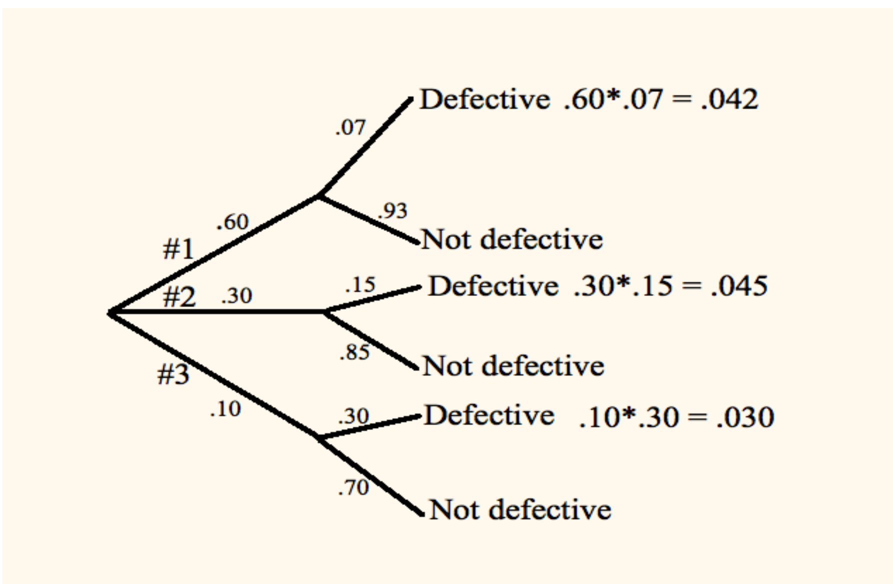
Of the parts Machine #1 makes, 7% are defective
Of the parts Machine #2 makes, 15% are defective
Of the parts Machine #3 makes, 30% are defective

If a part is randomly selected, find the probability it is defective.

These types of problems, are often best illustrated using a **tree diagram:**



Since we are interested in the probability a part is defective, we find the probabilities of the branches that result in a defective part, and add them up:



Reading off the tree diagram, P(Defective) = .042 + .045 + .030 = 0.117.

P(Defective) = P(Machine #1 makes the part and it is defective) +
P(Machine #2 makes the part and it is defective) +
P(Machine #3 makes the part and it is defective)
= .60(.07) + .30(.15) + .10(.30)= 0.117.

If a randomly selected part is defective, what is the probability that it came from Machine #1?

Looking again at the tree diagram, since we are given the part is defective, we know that it came from one of the 3 branches that resulted in defectives. If we know that it came from one of the 3 branches that resulted in defectives, then we can read from the tree diagram:

P(#1 | Defective) = P(#1 ∩ Defective)/P(Defective) = .042/0.117 = 0.3589744.

This is an example of Bayes' Theorem. Our original probability the part came from Machine #1 is .60, but based on the additional information that the part is defective, the probability now is only .359. You are not responsible for the *formula* for Bayes' Theorem, but you may be asked questions that can be answered using a simple tree diagram set up. The formula for Bayes' Theorem simply walks through what we did above using a tree diagram, but it can be intimidating for some students at first glance.

Example.
Two cards are drawn without replacement from a standard deck. What is the probability both cards are red?

P(Both are red) = P(First card is red)P(Second card is red | first card is red) = (26/52)(25/51) = 0.245.

Note that this is a little bit less than ¼. If the first card is red, that makes it a little less likely that the second card is red (since we are drawing *without replacement*)

What is the probability at least one is a king?

P(at least 1 king) = 1- P(no kings) Since the only possibilities are 0, 1, and 2.
= 1 - (48/52)(47/51)

This is an often used "trick" in probability calculations. Very frequently, we are interested in the probability of *at least one* event. For example, what is the probability that at least one bomb goes off on a certain plane? Since one bomb is enough to do the plane in, we simply need to know the probability there is at least one. As a lighter example, if we are nervous about being late for a job interview, we may set several alarm clocks in an effort to ensure that *at least one* goes off. Since in these cases we are dealing with count data, the only possibility other than *at least one* is *none*, and thus we often use P(at least one) = 1 – P(None), since P(None) is simple to calculate in many situations.

A Few More Probability Examples

Try working through these on your own. Solutions are available below.

Example #1.
A certain rare disease affects 0.5% of the population. A test for this disease is available.
If a person has the disease the test will be positive with probability .94.
If the person does not have the disease, the test will be positive with probability .02.

A) A randomly selected person is tested for the disease. What is the probability they test positive?
B) A randomly selected person tests positive for the disease. What is the probability that they have it?

Example #2.
An important satellite guidance system relies on a component that has a .4 probability of failure.

As this probability is high, designers put in 5 independent back-up components (each has the same probability of failure). The guidance system works as long as at least one of the 6 components works.

A) What is the probability all six components work?
B) What is the probability the guidance system works?

Example #3.
A set of old Christmas tree lights are connected in series; if one of the 20 bulbs fails, the string of lights does not work. Assume that the probability each individual bulb is not working is .03, and the individual bulbs can be considered independent. What is the probability the set of lights works?

Example #4.
You go on 20 job interviews in two days. You feel that they all go astoundingly well, and that there is an 80% chance you will be offered each individual job. You are turned down on the first 19 jobs. What is the probability you get a job offer on the 20^{th} one?

<u>Answer to the Examples Above</u>

Example #1. This type of thing is best illustrated with a tree diagram. Try drawing it out.

A) P(Test positive) = P(Disease ∩ test positive) + P(No disease ∩ test positive)
= .005(.94) + (1-.005)(.02) = 0.0246

B) P(Disease|Test Positive) = P(Disease ∩ Test positive)/P(Test positive) = .005(.94)/.0246 = .191.

This is an example of Bayes' Theorem. You are not expected to know the Bayes Theorem formula as given in the text, but you may have to answer basic probability questions that can be answered with a simple tree diagram.

Example #2.

A) P(All 6 work) = P(All six do not fail) = $(1-.4)^6$ = 0.046656. But this probability is not that interesting. We do not much care about all of the devices working, as we simply need at least one of them to function for the guidance system to work.

B) P(Guidance system works) = P(At least one component works) = 1- P(All components fail) = $1-.4^6$ = 1 - .004096 = 0.995904.

Even though the individual probability of failure is large, when we only need one of 6 components to work, the overall probability of success is large.

N.B. It is easy to *state* independence (just say "they are independent"), but it is much harder to achieve. In practical cases, it may be difficult to have these components working independently. For example, they may all have to be connected to the same power supply, or may be in close proximity where one cause of failure, a fire, say, could take them all out at once.

Example #3. This is different from Example #2 in that here the probability that *all* bulbs work is important to us. P(all 20 work) = $(1-.03)^{20}$ = 0.5437943.

Example #4. Probably pretty low. If your original assessment of 80% on each job is true, and the jobs were independent, and there was no time effect, then you'd have an 80% chance. However, since your probability of rejection on the first 19 was miniscule under your original assessment, you probably vastly overestimated your chances of getting a job in the first place.

Combinations and Permutations

Example.
You need to assign drivers to 3 different bus routes. If there are 20 bus drivers available, how many different ways can this be achieved?

The first route can be assigned to any of the 20 drivers. But after the first driver is selected, there are only 19 possibilities left for the second route. Once the first two drivers are selected, there are only 18 possibilities left for the 3rd route, and thus the number of ways we can assign the drivers is:
20*19*18 =6840.

Example. How many different 4 letter "words" can be created if any letter can only be used once?

We could list all the possibilities: *abcd*, *abce*, *abcf*, ... This would obviously be very time consuming, to put it mildly. Alternatively, we could realize that the first letter could be one of 26 options, then the second could be one of 25, etc. Leading to 26*25*24*23 = 358,800 possible orderings of 4 letters.

Generalizing this idea, suppose we have n distinct objects, and we are interested in the number of possible orderings of x objects from the n total.

The number of **permutations** of x objects from the n is $nPx=\frac{n!}{(n-x)!}$,
where $n! = n(n-1)...(2)(1)$. For example, 5! = 5*4*3*2*1 = 120. By definition 0! = 1.

For the letter example, $P_x^n=\frac{n!}{(n-x)!}=\frac{26!}{(26-4)!}=358{,}800$, which is the value we obtained using our logic skills.

If the *ordering* of selection is not meaningful, then the situation is changes. For example, if the lettering *abcd* is considered the same as *dacb*, and *dbac*, then we can use the *combinations* formula.

The number of ways of choosing x objects from n distinct objects is: $nCx=\binom{n}{x}=\frac{n!}{x!(n-x)!}$
Note that this is similar to the permutations formula above. The difference is the $x!$ in the denominator. This represents the possible orderings of the x items that are selected.

Example. If you buy 3 tickets to Lotto 6/49, what is the probability you win the grand prize? In Lotto 6/49 the order of selection is not important. If your 6 numbers come out of that drum in any order, you win the grand prize.

There are $\frac{49!}{6!(49-6)!}$ = 13983816 possible ways to select 6 numbers from 49.
The probability for a single ticket is 1/ 13983816. Your probability on winning on one of your 3 tickets is thus 3/13983816 (assuming that you don't pick the same set of numbers twice of course, unlike Mary Wollens).

N.B. Many calculators have pre-programmed functions that will automatically calculate combinations and permutations. Check your calculator's manual. Look for a *nCr* key (or something to that effect).

Simple Random Sampling

When we sample from a population, we want to obtain a sample that is *representative* of the population. A commonly used sampling method (that was outlined in Unit #1) is **simple random sampling**, which our text simply calls **random sampling**. I don't like the use of the term "random sampling" in the text, as random sampling has many different possible meanings, only one of which is simple random sampling.

In a simple random sample, every possibly sample of size *n* is equally likely to be selected. One implication of this is that every individual in the population has the same chance of being selected in the population.

Suppose we want to randomly choose a sample of size 5 from a population of size 57. How many different samples are possible?

The order of selection is not important here. We just want to know how many different possible samples there are. The order in which the individuals were drawn is not important.

$$\binom{57}{5}=\frac{57!}{5!(57-5)!}=4187106$$

We want to randomly pick one of these possible samples.

How? The usual way is to get a computer program to randomly pick 5 cases from a list of 57.

Old school method: Use a random number table.

Example: Pick a random sample of size 5 from a population of 57 individuals.

Number the members from 01-57.
Pick a random point in the random number table.
Let's say we start of row 18 of our textbook random number table: 01011 54092 33362

Since we require only two-digit numbers, we can split these up into 2 digit numbers:

01 01 15 40 92 33 36

Since the first observation is 01, the first individual in the list is in our sample. The second number is also 01, but we ignore it because the fist individual is already in. Then #15 and #40 enter the sample. We ignore the number 92 since we do not have anybody numbered that high (our population is numbered from 01-57). Then 33 and 36 enter, and we have our sample of size 5.

This leaves us with individuals 01, 15, 40, 33, and 36 as our simple random sample of size 5 from our population of 57 individuals. Many of our inference procedures we use later on in the course will assume a simple random sample from the population. This will not always be the case, but we hope that our sample behaves like a simple random sample.

Random Variables

Example. Randomly select a Fortune 500 company.

Let X represent the number of employees in the company.
Let Y represent the total weight of all employees in the company.

X and Y are **random variables**.

A **random variable** is a variable that takes on numerical values according to the outcome of an experiment.

We cannot predict a value of the random variable with certainty, but the values of the random variable follow a **probability distribution**.

Random variables can either be **discrete** or **continuous**.

Discrete random variables can take on a *countable* number of possible values. Discrete random variables do not necessarily *represent* a count. For example, if the random variable X can take on only the values -3.2 and 4.7, then there are two possible values. I can count to two. Since X takes on only two possible values, X is a discrete random variable.

Examples:

Let X represent *the number of broken eggs in a carton of eggs.*
The possible values are 0, 1, 2, ..., 12. There are 13 possible values. This is a discrete random variable.

Toss a coin until the first time heads appears. Let X be *the number of the toss on which heads first appears*. The possible values are 0, 1, 2, ... Although X can take on an infinite number of possible values, it is what we call *countably infinite*. X is a discrete random variable.

Bet $.37 on red on the spin of a Roulette wheel (in American casinos, a bet on Red is paid at even money, but has a 18/38 chance of winning). Let X represent your profit on one spin. Your profit will be either +.37 dollars, or -.37 dollars. Your profit X is a discrete random variable (it takes on one of two possible values).

Continuous random variables can take on an infinite number of possible values, corresponding to all values in an interval.

Examples:
Let X represent the length of a randomly selected cucumber in a grocery store.
(A randomly selected cucumber has to have a length that is greater than 0. Although we know that the cucumber cannot be 237173498274874298 cm long, say, there is no natural upper bound, and we can say that the range of possible values is $X > 0$)

Let Y represent the weight of cheese on a randomly selected *Pizza Pizza* pizza.
Let Z represent the time it takes for a randomly selected Harvard MBA grad to get a job.

Discrete Probability Distributions

Example.

According to Statistics Canada, 80% of Canadians live in urban centres.
Suppose we are about to randomly sample 2 Canadians. Let *X* represent the number that live in an urban centre. Write out all possible values of *X*, and their probabilities of occurring.

There are 4 possible outcomes of our sampling experiment:
Outcome: UU, UR, RU, RR
Value of *X*: 2 1 1 0

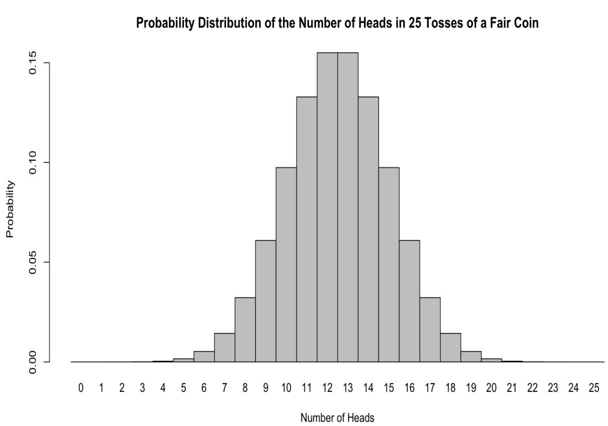
Value of *X*	0	1	2
Probability	.2*.2 = .04	.8*.2 + .2*.8 = .32	.8*.8 = .64

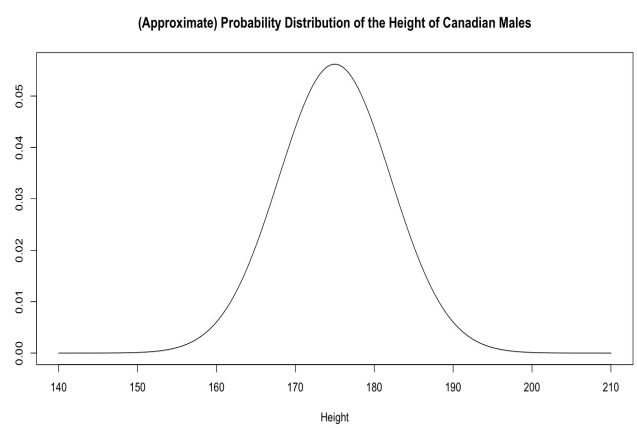
This is the **probability distribution** of the number that live in urban centres, when 2 Canadians are randomly sampled.

The **probability distribution** of a discrete random variable *X* is a listing of all possible values of *X,* and their probabilities of occurring. This can be illustrated using a table, histogram, or formula.

To be a valid discrete probability distribution, two conditions must be satisfied:
1) $0 \leq p(x) \leq 1$
2) $\sum p(x) = 1$

Consider the following two plots.





The plot on the left gives the probability distribution of the number of heads that occur when a fair coin is tossed 25 times. This is a *discrete* random variable that takes on 26 possible values.

The plot on the right is a rough approximation of the distribution of the height of Canadian males. This is a *continuous* random variable. We will have to make a few adjustments when we deal with continuous random variables.

Note that for discrete random variables there are discrete jumps between the values of the variable, whereas continuous random variables are represented by a smooth curve.

Expected Value

The **expected value** of a random variable is the average value of that variable if the experiment was repeated a very large (infinite) number of times.

The mean of a random variable is also called its expected value: $E(X)=\mu_X=\sum x\,p(x)$

Note that this is the theoretical mean of the random variable, sometimes called the mean of the probability distribution, and it is not the sample mean of a few observations. Thus the expected value of a random variable is a parameter, not a statistic.

The variance of a discrete random variable is: $\sigma_X^2=E[(x-\mu)^2]=\sum (x-\mu_X)^2\,p(x)$

Again, this is the theoretical value for the probability distribution, and is thus a parameter, not a statistic.

Note that although this may look different (and *is* different) from the formula for the sample variance, it is similar in spirit and can still be thought of as the *average squared distance from the mean*.

As usual, the standard deviation is simply the square root of the variance. The population standard deviation is represented by: $\sigma_X=\sqrt{\sigma_X^2}$. We usually leave out the subscript when we are dealing with only a single random variable.

Example.
The probability that a shipment of goods is lost in transport is 1/5000. If the goods are worth $20,000, what is the expected value and variance of the loss?

Let the random variable *X* represent the loss. Then the distribution of *X* is:

X	0	20 000
Probability	4999/5000	1/5000

$E(X)=\mu=\sum x\,p(x)=0(4999/5000)+20000(1/5000)=4$. On average, we will lose $4.

$\sigma^2=\sum (x-\mu)^2\,p(x)=(0-4)^2(4999/5000)+(20000-4)^2(1/5000)=79984.$
$\sigma=\sqrt{79984}.$

Suppose a lottery ticket has the following payouts. Let *X* represent the payout of a ticket.

x	0	1	10	100
p(x)	.689	.30	.010	.001

What is the expected value of the payout?
E(*X*) = 0(.689) + 1(.30) + 10(.01) + 100(.001) = 0.50. We are only getting back $.50 of every dollar we spend, which is pretty typical of lottery tickets.

What is the variance of the payout?

$$\sigma^2 = (0-.5)^2(.689) + (1-.5)^2(.30) + (10-.5)^2(.010) + (100-.5)^2(.001) = 11.05$$

What is the lottery's expected profit if they sell 25,000 of these tickets at $1 each?
They make an average of $1 - $.50 per ticket, so 25000*.5 = $12,500 profit on average.

Example.

Suppose you run a pharmaceutical company have a promising new drug in development.
You have already spent $15 million on research for this drug. Your researchers estimate you need to invest another $25 million before there is a chance of approval by Health Canada or the FDA in the US.

The drug may be approved for use in Canada, the US, both, or neither (ignore the global possibilities in this grossly simplified example). The following table gives estimated probabilities for approval and estimated profit (M) in each case.

Approval	Neither county	Canada	US	Both US and Canada
Estimated profit	-25	-5	130	200
Estimated probability	.20	.15	.10	.55

You are offered $100M for all rights to this drug. Should you accept?

This is of course a complicated question, with a lot of factors involved that are ignored here. However, a reasonable place to start is to calculate our expectation if we continue with the research. Well, we don't know our true expectation, but we can estimate it given our estimates of profits and probability:

-25(.20) -5(.15) + 130(.10) + 200(.55) = 117.25.

We estimate we'll make 117.25 M in profit if we continue, and we have been offered $100M. So from our estimates, we'd be better off taking our chances and continuing with the development. But there's certainly something to be said for locking up the $100M, thereby reducing our time commitment, and also the variability in the bottom line.

The Effect of a Linear Transformation on the Mean and Variance.

Example.

Suppose a certain machine costs $500 to operate per day, plus $200 to fix for every breakdown.If the number of daily breakdowns has a mean of 1.1, and a standard deviation of 0.5, what is the mean and standard deviation of the daily operating cost? This is an example of a linear transformation:
$Y = a + bX$.

For constants a and b:

$E(a + bX) = a + bE(X)$. Equivalently, we could write this as: $\mu_{a+bX} = a + b\mu_X$

Recall that the additive constant does not affect the variance: $\sigma^2_{a+bX} = b^2\sigma^2_X$, $\sigma_{a+bX} = |b|\sigma_X$

Cost = 500 + 200X. $\mu_{cost} = \mu_{500+200X} = 500 + 200\mu_X = 500 + 200(1.1) = 720.$

$\sigma_{cost} = \sigma_{500+200X} = 200\sigma_X = 200(0.5) = 100$

Example.

A friend of yours has a business opportunity in which he has an expected profit of $2M, with a standard deviation of $.8M.
(Although these values are nearly impossible to calculate in practical situations, they can often be estimated. For the purposes of this question assume they are the true values).

In order to reduce his risk, he will sell you ¼ of the profits (and possible losses) for a fee of $.4M. If you take the deal, what is your expected profit?

Let X represent your friend's profit on his original deal. Then $\mu = 2$, and $\sigma = 0.8$.
Your profit = $-.4 + ¼ X$. $\mu_{profit} = \mu_{-.4+.25X} = -.4 + .25\mu_X = -.4 + .25(2) = .1$.

What is the standard deviation of your profit? $\sigma_{profit} = \sigma_{-.5+.25X} = .25\sigma_X = .25(0.8) = .2$

If X and Y are two random variables:

X and Y are **independent** if knowing the value of one variable gives no information about the value of the other. There are formal mathematical definitions for independence of two random variables that are analogous to the definition of independent events in probability.

$E(X + Y) = E(X) + E(Y)$, or the alternative notation, $\mu_{X+Y} = \mu_X + \mu_Y$

The above result (the mean of the sum is the sum of the means) is true regardless of whether X and Y are independent.

$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\text{covariance}(X, Y)$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\text{covariance}(X, Y)$$

The covariance is a measure of the relationship between *X* and *Y*. We won't get too much into this at this point, because we will mainly deal with the situation in which *X* and *Y* are independent. If *X* and *Y* are independent, their covariance is equal to 0 and things simplify a great deal.

If *X* and *Y* are independent: $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$, $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$

Example

A painting process consists of the application of two coats, with a two hour drying time in between.
Coat 1: $\mu_X = 10$ minutes, $\sigma_X = 2$ minutes.
Coat 2: $\mu_Y = 15$ minutes, $\sigma_Y = 3$ minutes.

What is the mean and standard deviation of the process? (Assume *X* and *Y* are independent, even though this assumption might be a little sketchy in many scenarios)

Let *Z* represent the process time. $Z = X+Y+120$.
$E(Z) = E(X+Y+120) = E(X) + E(Y) + 120 = 10 + 15 + 120 = 145$ minutes.

$$\sigma^2_Z = \sigma^2_{X+Y+120} = \sigma_X + \sigma_Y = 2^2 + 3^2 = 13.$$

$$\sigma = \sqrt{13}$$

Note that the standard deviation does NOT equal the sum of the two standard deviations. When events are independent, we add their **variances** to get the variance of their sum.

There are infinitely many possible discrete probability distributions. But there are certain types of distributions that we tend to see repeatedly in common situations. Let's talk about a few of the more common discrete probability distributions that we frequently run into in business, science, and elsewhere.

Binomial Distribution

The **binomial distribution** is a very common discrete probability distribution. It can help us solve some basic probability problems, and it provides the base for some of our statistical inference later on in the course.

Examples:
A certain telemarketer knows he makes a sale on approximately 1% of his calls.
How many sales is he likely to make in 100 calls?

A family has 8 children. What is the probability they have at least 4 boys?

If we have
- n identical and independent trials
- each trial can be classified into one of two outcomes (labelled success and failure)
- P(Success) = p and is the same from trial to trial, P(Failure) = q = 1 - p

If we let X represent the number of successes in n trials, then X is a random variable that follows a binomial distribution:

$$p(x)=\binom{n}{x} p^x q^{n-x}$$

A binomial random variable X can take on the possible values 0, 1, 2, ..., n.

The mean and variance of a binomial random variable: $E(X) = \mu_X = np$, $\sigma^2 = npq$

Binomial or not?

A die is rolled 1000 times. Let X represent the number of times that it comes up 6.

Binomial. Even though there are 6 possibilities when we roll a die, we are only interested in the number of sixes. This gives us two labels: “six”, “not a six”. It is also reasonable to assume independence between rolls of a die, so the binomial conditions are pretty much perfectly justified here.

A car manufacturer calls a random sample of 200 purchasers of one of their models. Let X be the number that say they are satisfied.

Sure, binomial with n= 20, and unknown p.

A room contains 5 males and 5 females. Pick a random sample of size n = 4 (without replacement). Let X represent the number of males in the sample.

Not binomial. Here the probability of a success changes from trial to trial. For example, the first person picked has a probability of 5/10 = 0.5 of being male. But if the first person is male, the probability the next one is male is only 4/9. Since this probability changes from trial to trial, it's not a binomial random variable. If we sampled **with** replacement, it would be.

The number of free throws a specific NBA player makes out of 100.

Debatable. The binomial model would likely be reasonable, but not perfectly true. We do have a fixed number of trials and we are counting up the number of successes. It is the independence assumption that is a little troubling. If the player misses 3 in a row, will this not affect him at all? Will he be just as likely to make the fourth as ever? That is unlikely, as human beings are not robots. There has been some academic work investigating this, and if my memory serves the results show little to no evidence of dependence. Overall the binomial model would be reasonable here, though not perfectly true.

Example.
Suppose there are 10 customers in line at the breakfast/lunch changeover time at McDonald's. Suppose from past experience it is known that in these situations that 60% of the customers want to order breakfast.

What is the probability that exactly 3 customers want to order breakfast?(Assume independence between customers). Although the independence assumption might be a bit sketchy here, under the premise as described above, this is a straightforward binomial problem. We have n = 10 independent trials, and the probability of a success (wanting breakfast) is .60.

$$P(X=3)=\binom{10}{3}.6^3(1-.6)^{10-3}=0.04246733.$$

What is the probability that at least one customer wants to order breakfast?

$$P(X\geq 1)=P(X=1)+P(X=2)+...+P(X=10)=1-P(X=0)=1-\binom{10}{0}.6^0(1-.6)^{10-0}$$

= 1-0.0001048576 = 0.9998951.

What is the mean number of customers that want breakfast?
The mean of a binomial random variable is $\mu = np$ = 10(.6) = 6.

What is the variance? The variance of a binomial random variable is $\sigma^2 = npq$ = 10(.6)(1-.6) = 2.4.

The Hypergeometric Distribution

I won't make you calculate probabilities using the hypergeometric distribution. This section simply illustrates how it relates to the binomial, and why we may or may not be able to use the binomial distribution as an approximation in certain cases.

Example. A room contains 12 males and 8 females. If 10 people are randomly selected without replacement, what is the probability that exactly 5 are female?

This is NOT a binomial problem. Since people are being removed without replacement, the probability of success (that is, getting a female) changes from trial to trial. For example, the first person selected has a probability of 8/20 =.40 of being female. If the first person is female, the probability the second is female is 7/19 = 0.3684211. The probability is changing from trial to trial, and is thus this is not a binomial problem. To calculate this probability we would have to use the **hypergeometric distribution**:

$$\frac{\binom{12}{5}\binom{8}{5}}{\binom{20}{10}}=0.2400572$$

If we incorrectly used the binomial distribution with n = 10, p = 8/20 = .40, we'd get an answer of 0.2006581, which is a substantial difference.

A small arena contains 1200 males and 800 females. If 10 people are randomly selected without replacement, what is the probability exactly 5 are female?

This has the same problem as above, but for one fundamental difference. Here the number selected represents only a small part of the whole (10 selected from 1200), whereas above the number selected was a large part of the whole (10 from 20). This is still not a binomial problem, but we can approximate it as such. If we use the binomial formula with *n = 10, p = 800/1200 =.4 ,* we find that:

$$P(X=5)=\binom{10}{5}.4^5(1-.4)^5=0.2006581$$

Which is very close to the exact value based on the hypergeometric distribution:

$$\frac{\binom{1200}{5}\binom{800}{5}}{\binom{2000}{10}}=0.2009932$$

This illustrates that in situations where our sample only represents a small proportion of the total number of objects, the binomial distribution can provide us with a reasonable approximation. As a rough guideline, the binomial distribution will provide a reasonable approximation in these situations if we are not sampling more than 5% of the total number of objects.

<u>Example (a statistical inference problem)</u>

In the coming weeks we will be discussing *statistical inference*. This example gives a hint about the thought processes we will use there.

A certain company produces a weight loss pill. They claim that only 10% of purchasers of their product say they are dissatisfied with the effect of the pill. In a random sample of 100 of customers, 23 said they were dissatisfied with the product. If the company's claim is true, what is the probability that in a sample of 100 customers, 23 or more would say they are dissatisfied?

If the company's claim is true, the number that say they are dissatisfied in a sample of 100 should follow a binomial distribution with $n = 100$, $p = 0.1$.

$$P(X \geq 23) = 1 - P(X \leq 22)$$

This would be a HUGE hassle to do by hand, but it would be trivially easy to do in S-Plus, Excel, R, or any other statistical package. Using a computer, I find:

$$P(X \geq 23) = 1 - P(X \leq 22) = 1 - .99988 = .0001$$

The probability that at least 23 customers say they are dissatisfied if the company's claim is true is .0001. Since we actually observed 23, and it is a very unlikely occurrence if their claim is true, we have strong evidence that their claim is not true. We will use this type of argument frequently in the coming weeks.

For additional reading, see the text: Chapter 3 (3.1-3.7), Chapter 4 (4.1 – 4.4).

Suggested textbook questions.

Chapter 3. Probability.
3.1, 3.3, 3.5, 3.7, 3.9, 3.13, 3.23, 3.25, 3.27, 3.29, 3.31, 3.35, 3.37, 3.41, 3.43, 3.45, 3.47, 3.49, 3.51, 3.57, 3.59, 3.69 (and give some thought to part b – this part cannot be answered, despite what the text says), 3.73, 3.81, 3.87, 3.91, 3.93, 3.105, 3.109, 3.115, 3.117, 3.119, 3.121.

Chapter 4.
Section 4.1-4.2. Discrete Random Variables.
4.1, 4.3, 4.5, 4.11, 4.13, 4.15, 4.17, 4.19, 4.21, 4.25, 4.29, 4.33, 4.35

Section 4.3. The Binomial Distribution
4.37, 4.39, 4.41, 4.43, 4.45, 4.47, 4.49, 4.51, 4.53.

Suggested Course Manual Questions: All of Unit #2.