sides may appear more or less frequently and we could estimate the probability of each side by $P(E) = \frac{frequency\ of\ E}{number\ of\ trials}$.

If the side with three spots was observed 55 times in our 600 rolls then we would estimate $P(3) = \frac{55}{600} \approx 0.0917$. Of course, if we were to roll the die another 600 times we would likely get a slightly different result but as long as we are running hundreds–or thousands–of iterations we are going to get some pretty good estimates.

To get started, we are going to need some new functions in Excel.

Excel functions:

=rand() will return a random number between 0 and 1

Why would that be useful? All probabilities are between 0 and 1–such as P(Heads) = 0.5–then we can simulate flipping a fair coin and recording whether we get Heads by using IF statements along the lines of "if a random number is less than 1/2, then we will call that Heads (success) otherwise it is Tails (failure)". For convenience, let 1 =success and 0 = failure. You will see why that is convenient soon enough.

=randbetween(bottom, top) will return an integer value from the bottom value through the top value.

This is a useful tool for simulating rolling a die, by using =randbetween (1,6)

=if(test, value if true, value if false)is a familiar function. One way we will use it in probability is to return values of 1 for success and 0 for false

Example: =IF(A3<2, 1,0) For this IF statement, if the value in cell A2 is less than 2 the result will be 1, otherwise the result is a 0. The reason that using 1 and 0 to indicate "success" and "failure" is convenient? If every "success" is 1 and failure "0" summing up the column of values will count the number of successes.

Other tools

F9 is a function key on your keyboard. Pressing the F9 key will cause all of your rand() and randbetween() function to regenerate...which you will find disturbing at first. You'll get used to it.

Example (Genetics)
In a study to determine frequency and dependence of color blindness relative to males and females, 1000 people were chosen at random and the following numbers were observed:

	FEMALE (F)	MALE (M)	Totals
Color blind (*C*)	2	24	26
Normal (*N*)	518	456	974
Totals	520	480	1000

A person is randomly selected from the study.
Find:
P(male)= $\frac{480}{1000}$ P(color blind)= $\frac{26}{1000}$ P(male **and** color blind)= $\frac{24}{1000}$

Are gender and color blindness independent? Recall our definition of independence: If two events E_1 and E_2 are **independent** then: $P(E_1\ and\ E_2) = P(E_1) \cdot P(E_2)$. We can turn this around: if $P(E_1\ and\ E_2) \neq P(E_1) \cdot P(E_2)$ then E_1 and E_2 are **dependent.**

$\frac{24}{1000} \neq \frac{480}{1000} \cdot \frac{26}{1000}$

Does $P(male\ \textbf{and}\ color\ blind) = P(male) \cdot P(color\ blind)$?
No, $0.024 \neq 0.48 \cdot 0.026$. So the likelihood that someone is color blind is affected by the person's gender.

One more: What is the probability that a colorblind person is male? $\frac{24}{26}$

Note: The *boys/girls in 3-child family* example and the *color blindness* example show two different types of probability.

1. **Theoretical probability** (boys/girls)
2. **Empirical probability** (color blindness)

→ EXPERIMENTAL

Empirical probabilities are probabilities based on observation.
For this class we will be running simulations of anywhere from 100 to 5000 observations to collect our data.

Say we were to take a fair die, drill a hole in one of the spots, fill the hole with lead and then paint over the lead to make the die appear untampered with. For a fair die the probability of any one side appearing is 1/6, but we do not know the probabilities for our tampered die. If we were to roll a fair die 600 times we would expect to observe each side about 100 times. With our tampered die some

Probability Rule #2: If two events E_1 and E_2 are **independent** (occurrence of the first doesn't affect the second), then

$$P(E_1 \text{ and } E_2) = P(E_1) \cdot P(E_2)$$

I remember this one as "you can multiply ANDs if independent"

Example: Roll two dice, find:

$P(\text{first roll is even } \textbf{AND} \text{ second is even})$
$= P(\text{first roll is even}) \cdot P(\text{second roll is even})$

$\frac{3}{6} \cdot \frac{3}{6} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ or $\frac{9}{36}$

Complementary Events:

Recall that IF events ARE disjoint, then the sum of the probabilities of **all** possible outcomes is equal to 1. An event will either occur or it will not occur, correct?

$P(\text{event occurs}) + P(\text{event does not occur}) = 1.$

So P$(\text{event occurs}) = 1 - P(\text{event does not occur})$

This can be a real timesaver.

Roll a single die, $P(1) = {}^{1}/_{6}, P(\text{roll is greater than } 1) = 1 - {}^{1}/_{6} = {}^{5}/_{6}$

Example Experiment: Observe gender distribution of children in 3-child families.

Find the following probabilities: P(all girls), P(no boys), P(at least one boy), P(two girls), P(at most 2 girls)

First, list the sample space:

GGG GGB GBB BBB
GBG BGB
BGG BBG

P(all girls)= P(GGG) = $\frac{1}{8}$ P(no boys)=P(GGG)= $\frac{1}{8}$

P(at least one boy)= 1 - P(no boys) = $1 - \frac{1}{8} = \frac{7}{8}$

P(two girls) = P(GGB or GBG or BGG) = $\frac{3}{8}$

P(at most 2 girls) = 1 – P(more than 2 girls) = 1 – P(GGG) = $1 - \frac{1}{8} = \frac{7}{8}$

0, 1, 2 3

Disjoint

Events are **disjoint** if you have defined them in such a way that there is no overlap.

When rolling a single die, the event of "number < 3" overlaps the event of "even" (the number 2 is both < 3 and even),

While the events of "1" and "4" or "even" and "odd" are disjoint.

Probability Rule #1: If two events E_1 and E_2 are **<u>disjoint</u>** (non-overlapping), then

$$P(E_1 \textbf{ or } E_2) = P(E_1) + P(E_2)$$

(I remember this one as "you can add ORs if there is no overlap")

A corollary to this rule is that:

IF events ARE disjoint, then the sum of the probabilities of ***all*** possible outcomes is equal to 1.

Roll a single die:

$$P(even) + P(odd) = 1$$

$$\frac{3}{6} + \frac{3}{6} = \frac{6}{6}$$

(Think of it like this: $P(even\ or\ odd) =$)

$$P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$$

x	P(x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Independence

Example: Roll **<u>two dice</u>** (recall the sample space). Find the following probabilities.

A. $P(First\ roll\ is\ 2\ \boldsymbol{OR}\ second\ is\ 5) = \frac{11}{36}$

How does that compare to:

B. $P(first\ roll\ is\ 2\ \boldsymbol{AND}\ second\ is\ 5) = \frac{1}{36}$

P(2) AND P(5)

$$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

"OR" and "AND" are two very different words and you will need to be careful.

Let's talk about **independence**. Two events are independent if the outcome of one does not affect the likelihood of the other occurring. The outcome of a roll of a die has no effect on the outcome if I roll it a second time. Each flip of a coin is independent of the next flip.

Write the **event** (subset of the sample space) corresponding to

E_1: A sum of 6 is rolled: (5, 1) (4, 2) (3, 3) (2, 4) (1, 5)

E_2. A double is rolled: (1, 1) (2, 2) (3, 3) (4, 4) (5, 5) (6, 6)

Having the sample space written out makes it easy to list out the ways specific events can occur.

Probability of an Event – the measure of the degree of certainty of an event occurring.
The probability of an event, E is defined by

$$P(E) = \frac{\text{number of ways for E to happen}}{\text{total number of } \textbf{equally likely} \text{ outcomes}}$$

$$0 <= P(E) <= \ 1(or\ 0\% <= P(E) <= 100\%)$$

Example: Find the **probabilities** of the these events when rolling two dice:

E_1: $P(sum\ is\ 6)\ =$ $\frac{5}{36}$

E_2: $P(double)\ =$ $\frac{6}{36} = \frac{1}{6} = .16\overline{6} \approx .167,\ 16.7\%$

Example: Roll a **single** fair die; find the following probabilities:
(Do we need to list the sample space?) 1, 2, 3, 4, 5, 6

$P(1)$ = $\frac{1}{6}$ $P(4)$ = $\frac{1}{6}$ $P(1\ or\ 4)$ = $\frac{2}{6}$

1, 2
$P(number\ is\ less\ than\ 3)$
$\frac{2}{6}$

2, 4, 6
$P(even)$
$\frac{3}{6}$

1, 2, 4, 6
$P(even\ or\ <\ 3)$
$\frac{4}{6}$

Defining Sample Spaces

Example: Consider the experiment of observing the gender composition of two-child families.

A. What is an appropriate sample space if we want to record the gender of each child in age-order in the family? BB, BG, GB, GG

Are these outcomes equally likely? yes

B. What is an appropriate sample space if we are only interested in the number of boys in the family? 0, 1, 2

Are these events equally likely? NO

Example: Give sample spaces for the following experiments

A. Flip a single coin once: H, T

B. Flip a coin twice and observe the outcomes in order: HH HT TT TH

C. Flip two coins and observe the number of heads that appear: 0, 1, 2

Some experiments have a large sample space:

Sample Space for the Experiment "Rolling Two Dice"

Second Die / First Die	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

There are 36 possible outcomes, are these equally likely? yes

Probability –the likelihood that a random event will occur. Probability assigns a value $p\%$ to the likelihood of an event occurring, if an experiment is repeated a large number of times we would expect to observe the event occurring $p\%$ of the time.

Events and Sample Space

Example: Roll a single 6-sided die once ("the experiment")
What are the possible observations? These are called ***outcomes***
The set of all possible *simple outcomes* is called **the *sample space S***.

For this experiment, the sample space is the set: 1, 2, 3, 4, 5, 6

An **event** is a subset of the sample space (including the empty set and *S* itself). The event is the specific outcome we are interested in.

Example : Consider the die rolling experiment. What outcomes of the sample space would be part of the following events?

E_1: Die comes up 3: 3

E_2: Die comes up even: 2 4 6

E_3: Die comes up less than three: 1, 2

If you start to get confused in a probability problem, back up and consider what the sample space of the experiment is.