Example:
Suppose we toss a quarter and observe whether it falls "heads" or "tails", recording the result as "1" for "heads" and "0" for "tails".

(In Mathematical language, the result of our toss is a random variable, which we will call X, and this variable has two possible values, 0 and 1.)

Which of these two values X will actually take is going to be determined by the coin toss.

Will we find that X=0 or that X=1? We can't say. Nobody knows. What we can say is that the probability that X will be 0 is ½, and the probability that X will be 1 is also ½.

⇨ *Mathematically*, we can write the above statements in symbols:

P(coin falls heads)=P(X=1)=1/2
P(coin falls tails)=P(X=0)=1/2

A probability statement is a method of expressing our uncertainty about the outcome of a *future* event.

⇨ It does not make since to talk about the probability of an event that has already occurred (because there is no uncertainty in the outcome, i.e. the outcome has already been determined).

For example:
The question "What is the probability that the Red Sox lost the first game of the series to the Yankees?" makes no sense because we already know that the Red Sox lost that game.

⇨ However if you do not yet know the outcome, then it may make sense to talk about that probability, because it reflects your personal uncertainty about the situation.

Baseball example continued:
Now if you happened to miss the first game of the series, it would make sense to ask "What is the probability that the Red Sox lost to the Yankees in the first game of their series?" because you would still be uncertain about the game results.

Specifically, what does P(X=1)=1/2 mean?

The probability of heads represents the "tendency" of the quarter to fall heads, or the "degree of certainty" that heads will occur.

- P(X=1)=0 means no tendency at all to fall heads – there is no chance of observing heads.
- P(X=1)=1 means an overwhelming tendency to fall heads, i.e., perfect certainty – as if we were tossing a coin with a "head" on both sides.

⇨ The variable X is a special type of variable, one whose value is determined by chance. It is a <u>Random</u> variable.

This ‘experiment’ (coin toss) can be described by stating that there are two possible values for X, 0 and 1, and that their probabilities are P(X=0)=1/2 and P(X=1)=1/2.

Mathematically this description consists of three components:

(a) a random variable
(b) a set, say S={0,1}, containing all the possible outcomes of the random variable in (a)
(c) A listing of the probabilities that the random variable in (a) take each value in S.

These three components (a)-(c) are a mathematical representation (a <u>probability</u> <u>model</u>) for the process of tossing a quarter and observing the result.

Now lets consider a more complicated experiment. Suppose we toss three coins, a nickel, a dime, and a quarter. We can represent this experiment mathematically with the same three components.

This time let's use Y to represent the result. Again the result is a variable, Y, whose values is determined by chance – a random variable.

The possible outcomes for this experiment are:

(nickel, dime, quarter)
(0,0,0) (tails on all three)
(0,0,1) (nickel and dime tails, quarter heads)
(0,1,0) etc.
(0,1,1)
(1,0,0)
(1,0,1)
(1,1,0)
(1,1,1)

For this experiment, S is the set of these eight possible results:

$$S=\{(0,0,0), (0,0,1), (0,1,0),\ldots, (1,1,1)\}$$

So that result of this experiment, Y (the random variable), will have one of these eight values. Either Y=(0,0,0), Y=(0,0,1),…, or Y=(1,1,1).

The probability of each of the eight possible outcomes is 1/8:

$$P(Y=(0,0,0))=P(Y=(0,0,1))=\ldots= P(Y=(1,1,1))=1/8$$

Again we have represented the experiment in term of the three elements:

(1) a random variable, Y, representing the result
(2) a set S, called the <u>sample</u> <u>space</u>, showing the possible values of Y, and
(3) a list giving the <u>probability</u> of each value.

⇨ These three pieces, taken together, form a <u>probability</u> <u>model</u> for this experiment.

Any probability model (no matter how complicated) consists of these three components:

(1) A random variable
(2) A sample space
(3) A probability distribution over the sample space.

A probability model allows us to analyze the experiment mathematically … to answer questions like "What is the probability of getting heads on two out of the three tosses?" by exact calculations instead of by scratching our heads and guessing.

Using a probability model

Now that we have a probability model, let's examine a question we already know the answer to:

What is the probability of heads in the first toss?

In terms of our probability models, it is the probability that Y takes of the four values, (1,0,0), (1,0,1), (1,1,0), or (1,1,1).

Thus the event "Heads on the first toss" is represented in this probability model by the set of possible outcomes,

A={(1,0,0), (1,0,1), (1,1,0), (1,1,1)}

The probability of this event (or this set) is simply the sum of the probabilities of its elements:

P(A) = P({(1,0,0), (1,0,1), (1,1,0), (1,1,1)})
= P((1,0,0) or (1,0,1) or (1,1,0) or (1,1,1))
⇨ = P((1,0,0)) + P((1,0,1)) + P((1,1,0))+P((1,1,1))
= 1/8 + 1/8 + 1/8 + 1/8
= 4/8=1/2

We will see later that the important step that the little arrow points to is justified by one of the basic rules of probability. For now, we'll just accept it.

Now try a question whose answer is not so obvious:

What is the probability of heads on one of the last two tosses?

This event might be represented by the set

B={(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)}

But it might also be represent by a different set

C={(0,1,0),(0,0,1), (1,1,0),(1,0,1)}

Before we can answer the question, we must be more specific: Do we mean

"…heads on at least one of the last two tosses?" or
"… heads in exactly one of the last two tosses?"

These are two different questions, referring to two different events. The first is represented in our probability model by the set B, and the other by set C.

If we mean "…at least one" then we want the probability of event (set) B:

$$
\begin{aligned}
P(B) &= P(\{(0,1,0),(0,0,1),\ldots,(1,1,1)\}) \\
&= P((0,1,0) \text{ or } (0,0,1) \text{ or } \ldots \text{ or } (1,1,1)) \\
\Rightarrow \quad &= P((0,1,0)) + P((0,0,1)) + \ldots + P((1,1,1)) \\
&= 1/8 + 1/8 + 1/8 + 1/8 + 1/8 + 1/8 \\
&= 6/8 = 3/4
\end{aligned}
$$

If we mean "…heads on exactly one" the we want the probability of event (set) C:

$$
\begin{aligned}
P(C) &= P(\{(0,1,0),(0,0,1),(1,0,1),(1,1,0)\}) \\
&= P((0,1,0) \text{ or } (0,0,1) \text{ or } (1,0,1) \text{ or } (1,1,0)) \\
\Rightarrow \quad &= P((0,1,0)) + P((0,0,1)) + P((1,0,1)) + P((1,1,0)) \\
&= 1/8 + 1/8 + 1/8 + 1/8 \\
&= 4/8 = 1/2
\end{aligned}
$$

How about the event "Heads on the first toss <u>or</u> heads on at least one of the last two?" This event occurs if the result, Y, has any of the values that appear in either set A or set B.

A={(1,0,0),(1,0,1),(1,1,0),(1,1,1)}
B={(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)}
*   *   *

The set "A or B" consists of A plus the three checked points (0,1,0),(0,0,1), and (0,1,1), that are in B, but not in A.

Equivalently, "A or B" consists of B plus any additional points that are in A but not in B; in this example there is one such point, (1,0,0).

You find that the set "A or B" consists of the same seven points either way:

A or B =
{(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1),(1,0,0)}

And the probability of this event is

P(A or B) = 7/8.

But wait:

How do the probabilities of the events A and B relate to the probability of the event "A or B"?

$$P(A \text{ or } B) = P(A) + P(B) \quad ?$$
$$= \tfrac{1}{2} + \tfrac{3}{4} = 1\tfrac{1}{4} \quad \text{(NO)}$$

In order to answer questions like this we need some notation and techniques to help us organize our thinking about sets.

SETS

$A \cup B$: The union of A and B. The set of points that are in A or in B (or both).

$A \cap B$: The intersection of A and B. The set of points that are in both A and B.

Note: Sometimes when there is no danger of confusion we omit the "$\cap$" and write simply AB to represent this set.

$A^c = \bar{A}$: The compliment of A. The set of all points that are not in A. This set is always defined in relation to the set, S, that represents all of the points that are being considered in a particular problem. That is, A^c is the set of all point in S that are not in A.

Example (tossing three coins continued):

$A = \{(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$

$A^c = \bar{A} = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1)\}$

Because every point that is not in A^c (not in "not A") is in A, we have

$$(A^c)^c = A$$

And because every point in S is either in A or not in A (in A or in $\bar{A} = A^c$), it is always true that

$$A \cup (A^c) = S$$

The Empty Set, Ø

When two sets have not points in common, their intersection is empty. This is a perfectly good set. It has nothing in it, but it is still a set. It occurs so often that we have a special symbol for it, Ø.

Ø: The empty set. A set containing no points.

Since a set and its complement have no points in common, it is always true that

$$A \cap (A^c) = \emptyset$$

Subsets

- When every point in some set A also belongs to some other set, B, then A is a subset of B, written as $A \subset B$.

- If A contains a point that is not in B, then A is not a subset of B, written as $A \not\subset B$.

For the sets in our example,
(A is "heads in the first toss", B is "heads on at least one of the last two tosses", and C is "heads on exactly one of the last two tosses")

A={(1,0,0),(1,0,1),(1,1,0),(1,1,1)}
B={(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)}
C={(0,1,0),(0,0,1),(1,1,0),(1,0,1)}

Every point in C is also in B, so C is a subset of B, or $C \subset B$. But the set A contains a point, (1,0,0) that is not in B, so $A \not\subset B$.

Note also that every set is a subset of the sample space S. ($A \subset S$, $B \subset S$, $C \subset S$)

Examples:

$A \cap (B^c)=\{(1,0,0)\}$

This set, the intersection of A and B-compliment, or "A and not B", consists of the single point, (1,0,0) (because all of the other points are in A are also in B).

$(A^c) \cap B=\{(0,0,1),(0,1,0),(0,1,1)\}$

This is the intersection of A-compliment and B, or "not-A and B". It is composed of the points representing <u>both</u> of the events,
not-A (not "heads on the first toss", which is the same as "tails on the first toss")
and B ("heads on at least one of the last two tosses").

Now, if we add to our list the event "same result on all three tosses", which corresponds to the set

$$D=\{(0,0,0),(1,1,1)\}$$

We find that $C \cap D = \emptyset$ because the set C and D have no points in common.

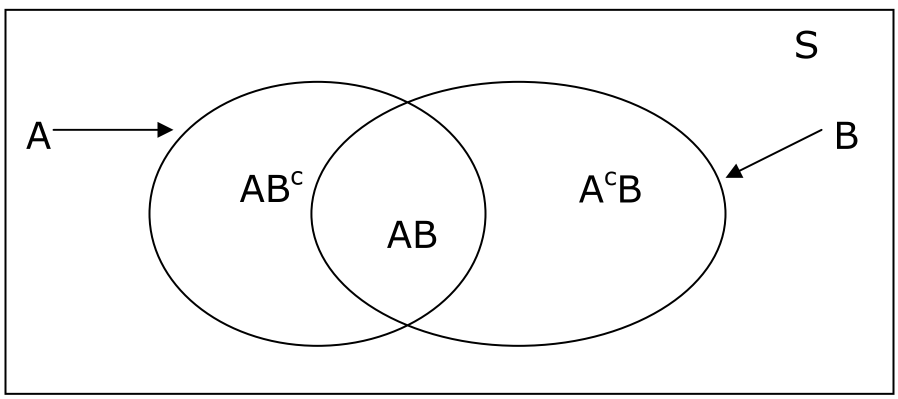
Set like this are called **mutually exclusive**, or **disjoint sets**.

※※ *Two events are <u>mutually</u> <u>exclusive</u> if their intersection is the empty set.*

Venn Diagrams

Venn diagrams are very handy tools for studying sets. Basically, a Venn diagram is just a picture of the sample space divided into its subsets. They are useful because they illustrate relations that are true of all sets.

Here is an example:



This Venn diagram displays the sets:
A, B, AB, AB^c, A^cB, S.

Note: To be perfectly clear we should have written AB^c as $A \cap (B^c)$, so that there could be no confusion with $(A \cap B)^c$. We will use the rule that unless parentheses tell us otherwise, the little "c" applies only to the set that it is attached to.

Back to Probability

Consider the following generic probability model:

(1) A random variable X
(2) A sample space S
(3) A probability distribution over the sample space.

Some *fundamental properties* of probabilities are:

⇨ (1) For any event E, $0 \leq P(E) \leq 1$

⇨ (2) For any two mutually exclusive events, E and F,
$$P(E \text{ or } F) = P(E \cup F) = P(E)+P(F)$$

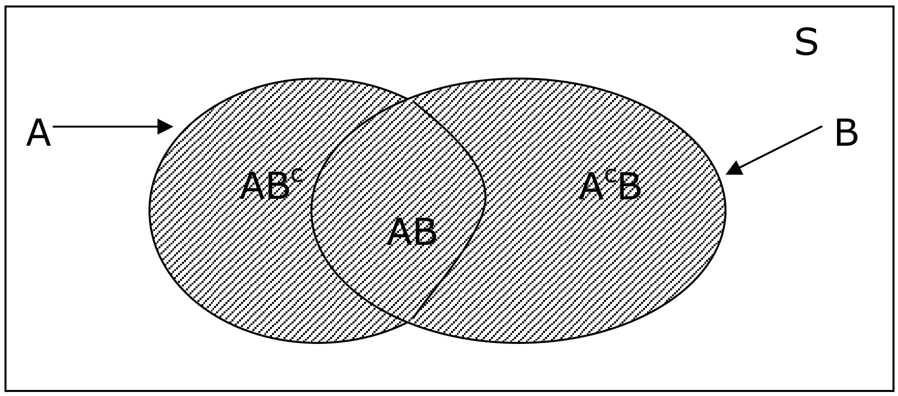
⇨ (3) $P(S) = 1$

⇨ (4) $P(\emptyset) = 0$

⇨ (5) For any event E, $P(E^c) = 1 - P(E)$

⇨ (6) For any two events $E \subset F$, $P(E) \leq P(F)$

Fundamental Properties continued:



The Venn Diagram shows us how to relate the probabilities of two events, A and B, to the probability of the event "either A or B".

⇨ (7) For any two events A and B

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

⇨ Alternatively rearranging yields:

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

Remember $P(AB) = P(A \text{ and } B) = P(A \cap B)$

To see why this must be true look at the Venn Diagram, which shows that A, B, and $A \cup B$ can all be written as unions of the three disjoint sets, AB^c, AB, and A^cB (shaded area).

$$A = AB \cup AB^c$$
$$B = AB \cup A^cB$$
$$A \cup B = AB \cup AB^c \cup A^cB \text{ (shaded area)}$$

Therefore,

$$\begin{aligned} P(A \cup B) &= P(AB) + P(AB^c) + P(A^cB) \\ &= P(AB) + P(AB^c) + P(A^cB) + P(AB) - P(AB) \\ &= P(A) + P(B) - P(AB) \end{aligned}$$

And we have proven fundamental property number (7), which is one of the most useful properties.

Conditional Probability and Independence

A key concept in probability is that of <u>independence</u>:

⇨ **Independent Events**: Two events, A and B are independent if

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

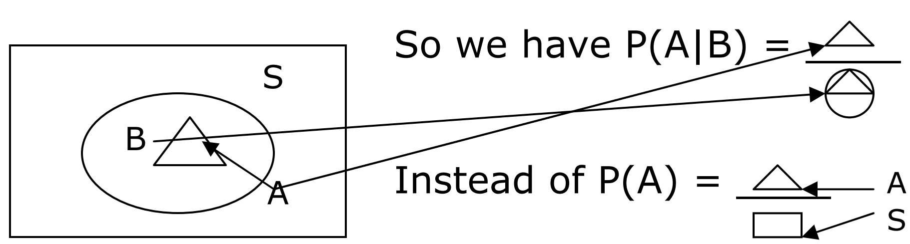
Another key concept is that of <u>conditional</u> <u>probability</u>:

⇨ The **Conditional Probability** of event A, <u>*given*</u> event B, is written $P(A|B)$ and is defined:

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

Or upon rearrangement

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



Conditional Probability is best illustrated with an example:

	Smoker	Drinker	Dieter	Total
Male	97	45	63	205
Female	80	30	72	182
Other	11	12	13	36
Total	188	87	148	423

So that we have:

P(smoker) = 188/423
P(other) = 36/423

P(smoker and other)=11/423

P(smoker or other) = 188/423 + 36/423 −11/423
= 213/423 = (97+80+11+12+13)/423

P(smoker|other) = P(smoker and other)/P(other)
= (11/423)/(36/423) = 11/36

Check independence:

$$P(\text{smoker and other}) \overset{?}{=} P(\text{smoker})P(\text{other}) = (188/423)(11/423) = 0.0116$$

But 0.0116 is not equal to P(smoker and other) = 0.026 (11/423 see above) so the events are not independent.

Another option is to check to see if P(smoker|other) = P(smoker), which is not the case.

Note: The conditional probability is conditional on the column (or marginal) total. It asks, “Out of the others, how many are smokers?” Thus we are effectively changing our sample space.

Independence and Mutually Exclusive

Mutually exclusive events are highly dependent, because if an event is not in one set, it *must* be in the other.

This is easy to see mathematically:

Suppose two events, E and F, are mutually exclusive. This implies that

(E and F) = Ø,   and that   P(E and F) = 0

For E and F to be independent events we need the following equation to hold:

P(E and F) = P(E)P(F)

But for mutually exclusive events, this would mean that

0=P(E)P(F)   so that either   P(E) = 0 or P(F) = 0
(neither of which is very interesting)

Thus at least one of the events must never occur for two events to be mutually exclusive and independent with each other.

In terms of conditional probabilities, what independence of A and B means is that A's probability of occurrence is unaffected by whether B occurs or not.

$$\begin{aligned} P(A|B) &= P(A \text{ and } B)/P(B) \\ &= P(A)P(B)/P(B) \qquad \text{by independence} \\ &= P(A) \end{aligned}$$

Likewise, B's probability of occurrence is unaffected by whether or not A occurs:

$$\begin{aligned} P(B|A) &= P(B \text{ and } A)/P(A) \\ &= P(B)P(A)/P(A) \qquad \text{by independence} \\ &= P(B) \end{aligned}$$

From the way conditional probability is defined, $P(A|B)=P(A \text{ and } B)/P(B)$, we can always write the probability of (A and B) as:

$$P(A \text{ and } B) = P(A)P(B|A) \qquad \text{Or}$$

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

When A and B are independent (and only in that case) $P(B|A)=P(B)$ and $P(A|B)=P(A)$, the above two equation reduce to $P(A \text{ and } B) = P(A)P(B)$.

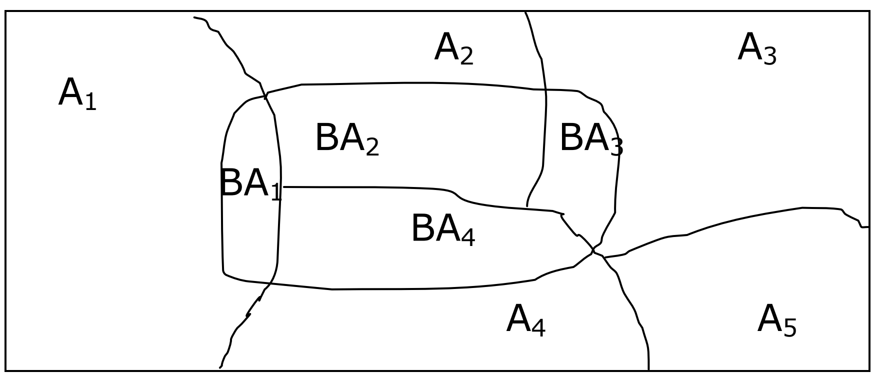
Law of Total Probability

Suppose the sample space is divided into any number of disjoint sets, say $A_1, A_2, \ldots, A_n$,

so that $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = S$.

Furthermore we can divide any set B into the disjoint subsets:

$$B = BA_1 \cup BA_2 \cup \ldots \cup BA_n$$



In this case we can write

$$\begin{aligned} P(B) &= P(BA_1 \cup BA_2 \cup \ldots \cup BA_n) \\ &= P(BA_1) + P(BA_2) + \ldots + P(BA_n) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)\,P(A_n) \end{aligned}$$

Or more generally:

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

Final note:

We said earlier that probabilities are "tendencies". Some people (like the authors of our textbook) prefer to define what probabilities are in terms of relative frequencies in very long sequences of repetitions of the experiment or process.

Still others say that probabilities are nothing more than personal "degrees of belief" (and some of those people are quite uncomfortable with the idea that a probability might have an objective "true" value that is independent of any person's beliefs).

Thus some think of probabilities as objective (and measurable) characteristics of certain parts of the world, much like temperature or viscosity, while other view probabilities as subjective measures of personal uncertainty, applicable to one's judgment and opinions.

In this course we take the position that in many important situations in makes sense to think of probabilities as objective quantities, so that it also make sense to speak of the true value of a probability, to attempt to estimate that value, and to speak of the error in an estimate.