Shyane Siriwardena
ss2032@cam.ac.uk

Basic Probabilities

The probabilities that we'll be learning about build from the set theory that we learned last class, only this time, the sets are specifically sets of *events*.

What are "events"?

Roughly, ***events*** are ways-things-can-go. So, if I'm tossing a coin, there are two ways-things-can-go: namely, it can land heads or it can land tails. The total list of possible ways-things-can-go with respect to a given occurrence is known as the ***event space***. In other words, the event space is equal to the *set of all ways-things-can-go*. So, in the case of a single coin toss, this would be equal to the set {Heads, Tails}. I like to visualise the event space as a table (the shaded area is the entire event space):

Toss 1

Heads	Tails
H	T

Fig.1

I emphasised 'single' above to raise the following point: events are not just identical with their outcome--it matters *how* the events came about. To illustrate, imagine I'm performing two tosses of a single coin. It is tempting to think that the event space should look like this {(Heads, Heads), (Tails, Tails), (Heads, Tails)} (i.e. that there are only three possible ways-things-can-go).

2 Heads	2 Tails	Heads/Tails
(H, H)	(T, T)	(H, T)

Fig.2

After all, what difference does the *order* in which the results land matter--isn't it just the same if the coin lands heads first as if it lands tails first?

In a word, ***no***. *It's not at all the same, where probabilities are concerned!* What is important is that there are *two* different ways in which the (Heads, Tails) outcome can come about. This is why I called events 'ways-things-can-go'--this includes how they get there. Thus, the event

Shyane Siriwardena
ss2032@cam.ac.uk

space for two coin tosses is accurately as follows: {(Heads, Heads), (Heads, Tails), (Tails, Heads), (Tails, Tails)}

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig.3

Another way to visualise the event space that might help make the previous point clear is in terms of branching possibilities. Consider a single coin toss. In the following, the vertical axis marks the progression of time. Imagining yourself at a time prior to Toss, we have two possible results:

```
Time              TOSS
 |               /    \
 |              /      \
 ↓          HEADS       TAILS
```

Fig. 4

Now, let’s imagine ourselves at a time prior to Toss again, but this time take into consideration the second toss. We don’t yet know how the first toss will go, so we can’t eliminate either of those branches yet. We have to reason as follows: “supposing the first toss lands heads, the second could land either heads or tails; or, supposing the first toss lands tails, the second could land either heads or tails.” Visually, we represent this as follows:

```
Time              TOSS
 |               /    \
 |              /      \
 |          HEADS       TAILS
 |           /   \       /   \
 ↓      HEADS   TAILS  HEADS   TAILS
```

Fig. 5

Shyane Siriwardena
ss2032@cam.ac.uk

Using either the table or the branching chart can help to make calculating probabilities much easier. (NB: the branching chart will not be feasible for more cumbersome probabilities like "Monday Girl".) In general, when we ask for the probability of some A here, we are asking for the ratio of the number events where A comes about to the number of events in the event space; i.e.

# of events where A occurs

# of events in the event space

OR

# of ways A can come about

# of ways-things-can-go

Fig. 6

So, for instance, in the case of the coin tossed twice, the probability of a coin's landing heads at least once can be calculated by looking at all of the events in our event space, and tallying up the number of instances in which there is at least one heads. Recall Fig. 3, where green represented the total event space:

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig.3

Now, consider Fig. 7, where orange represents all of the events with which we're concerned (i.e. all those that satisfy "heads lands at least once"):

Shyane Siriwardena
ss2032@cam.ac.uk

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig.7

So, the probability of heads landing at least once is going to be equal to:

# of orange squares (i.e. # of events satisfying the description)

# of green squares (i.e. # of events in the event space)

which equals…

3/4

Conditional Probabilities

Conditional probabilities are really no different to basic probabilities, for our purposes; the only thing that changes is the reference class, or in more practical terms, the denominator of our fractions from above. In these cases, instead of asking about the probability of A given *all* the ways-things-can-go, we are concerned about the probability of A given *some (subset) of* the ways-things-can-go. In practice, this means counting a different set of events for our denominator (usually one that is a proper subset of, or some selection of, the event space).

So, imagine we want to know the probability of heads landing at least once given that tails lands at least once. We begin, as before, by returning to Fig. 3:

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig.3

Shyane Siriwardena
ss2032@cam.ac.uk

But this time, we need to check the number of events that satisfy the “given that” clause first! In this example, we want to know the probability of something *given that tails lands at least once*. So let’s find all of the events in which tails lands at least once:

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig.8

I’ve shaded these boxes in a lighter green deliberately. If you like, with respect to the conditional probability we are trying to calculate, this is our new, restricted event space. We’re now really only concerned with the following:

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	**(H, T)**
	Tails	**(T, H)**	**(T, T)**

Fig. 8.1

Given this new restricted ‘given that’ space, we now want to know how many of *these* events satisfy the description “heads lands at least once”. And this time, our answer is **2** (rather than 3, as in the first example).

		Toss 2	
		Heads	Tails
Toss 1	Heads	(H, H)	(H, T)
	Tails	(T, H)	(T, T)

Fig. 9

Shyane Siriwardena
ss2032@cam.ac.uk

So, the conditional probability of heads landing at least once given that tails lands at least once is equal to

$$\frac{\text{\# of orange squares (i.e. \# of events satisfying the description)}}{\text{\# of green squares (i.e. \# of events in ‘given that’ space)}}$$

which equals…

2/3

Conditional Probabilities, Pt. 2 -- The Formula

The formula you were given for conditional probabilities looks like this:

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

But why? Let’s try to make sense of this using more diagrams.

Imagine the following is our event space:

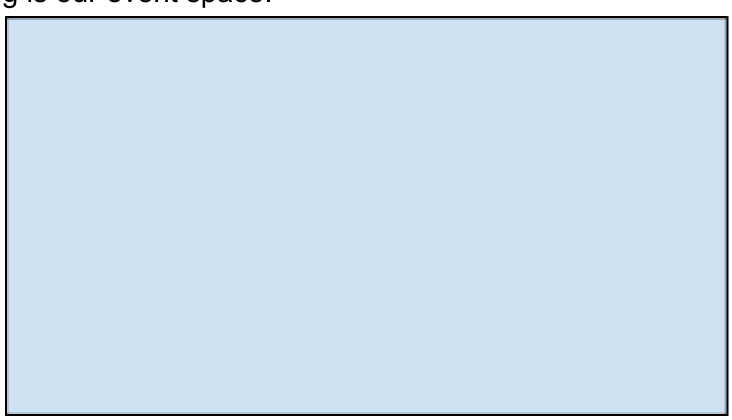

Fig. 10

Shyane Siriwardena
ss2032@cam.ac.uk

When we ask for the probability of some A, we want to know the ratio of the red area below (i.e. the set of all A events) to the entire rectangle area (i.e. the set of all possible events). This represents the ratio of the number of events satisfying A to the number of events in the event space:

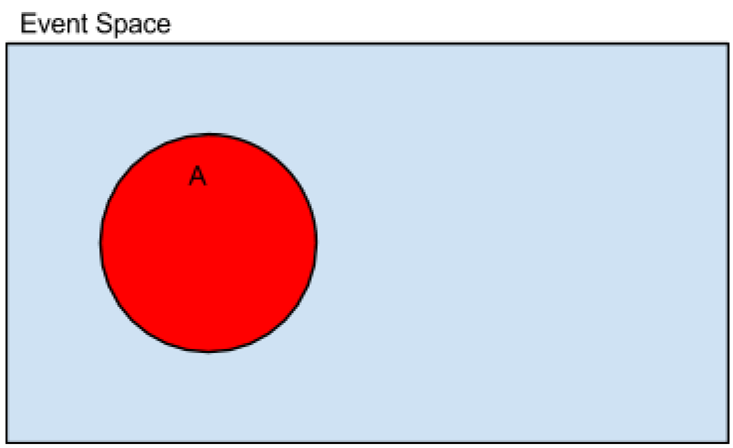

Fig. 11

Now let's supposed we're also concerned with another kind of event B:

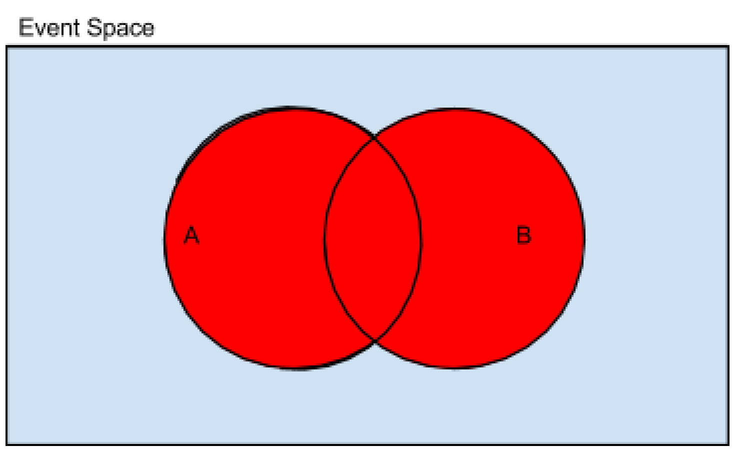

Fig. 12

If each of the shaded areas represents a set of events, then the area shaded red represents (A ∪ B), i.e. all the events that are either A or B.

Remember what was said before about conditional probabilities and restricting the event space?

Shyane Siriwardena
ss2032@cam.ac.uk

Here's how we can represent that using venn diagrams. Suppose we are interested in the conditional probability of B given that A (i.e. Pr(B|A)). It was the 'given that' clause that constituted the restricted 'given that' space--that is to say, it was the 'given that' clause that described what figures as the denominator in our fraction (see p.6). Thus, the denominator in our formula for conditional probabilities is the probability of the 'given that' event--in this case A. Visually, then, the denominator is constituted by the area shaded in YELLOW below:

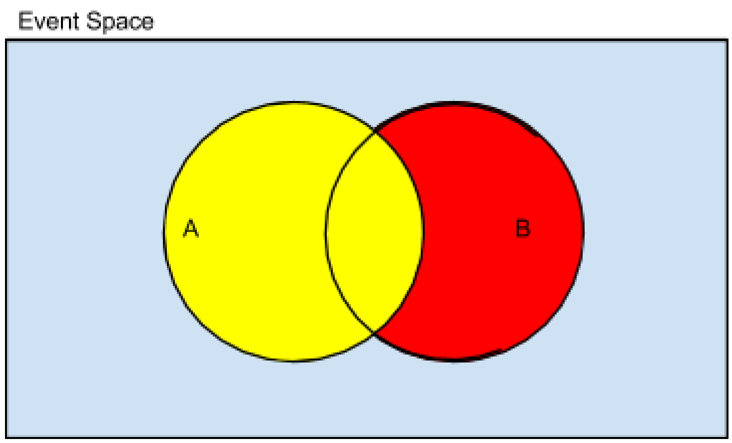

Fig. 13

Now, return to the original formula for conditional probabilities:

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)}$$

We've just figured out why the denominator is such as it is. Now let's figure out why the numerator consists of Pr(A ∩ B).

Remember, in a conditional probability, we wanted to know how many events in the restricted 'given that' space (in this case, the A space in yellow) satisfy the description--in this case, B. In other words, we need first to figure out how many A's are also B; then we can calculate the ratio of the # of A's-that-are-B to the # of A's. In our venn diagrams, the shaded areas represent sets of events. If we want to know the number of events that are A's **and** B's, then we want to know which events fall into **both** sets. In other words, we are interested in the **intersection** of A and B. Hence Pr(A ∩ B)! Visually, we are concerned with the area in orange below:

Shyane Siriwardena
ss2032@cam.ac.uk

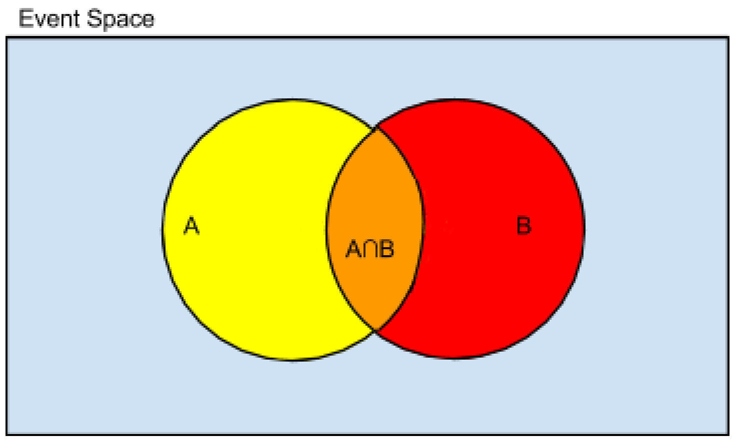

Fig. 14

Bayes' Theorem

Bayes' Theorem is the formula we use when we want to know about certain kinds of conditional probabilities. First we'll go through how to get the simple version of the theorem by building on what we just learned about conditional probabilities.

First, note that Bayes' Theorem is also concerned with Pr(B|A).

Now, recall the formula for conditional probabilities from above:

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

In words, the probability of B given that A is equal to the probability of A and B (i.e. A intersection B), divided by the probability of A.

First, let's multiply both sides by Pr(A)...

$$\Pr(B|A) \cdot \Pr(A) = \frac{\Pr(A \cap B)}{\Pr(A)} \cdot \Pr(A)$$

Shyane Siriwardena
ss2032@cam.ac.uk

which equals

$$\Pr(B|A) \cdot \Pr(A) = \Pr(A \cap B)$$

Using the original formula for conditional probability, we can also say that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

In words, the probability of A given that B is equal to the probability of A and B, divided by the probability of B.

Once again, we can multiply both sides by the denominator:

$$\Pr(A|B) \cdot \Pr(B) = \frac{\Pr(A \cap B)}{\Pr(B)} \cdot \Pr(B)$$

which equals

$$\Pr(A|B) \cdot \Pr(B) = \Pr(A \cap B)$$

Notice that we now have two different equations that are equal to Pr(A ∩ B):

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B) = \Pr(B|A) \cdot \Pr(A)$$

Shyane Siriwardena
ss2032@cam.ac.uk

Now, recall again that Bayes' Theorem concerns Pr(B|A). Given this, let's rearrange the above to isolate Pr(B|A) on one side of the equal sign.

Focus on this part:

$$\Pr(A|B) \cdot \Pr(B) = \Pr(B|A) \cdot \Pr(A)$$

And now dividing both sides by Pr(A), we're left with...

$$\Pr(B|A) = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A)}$$

This is the basic formulation of Bayes' Theorem.

Why would we ever use this, instead of the original formula for conditional probabilities? In this version of the theorem, it's not immediately obvious, since we need at least as much information to solve for Pr(B|A) here as we did with our original formula for conditional probabilities--that is to say, in both cases, we need to know Pr(A) and Pr(B). Compare both formulas:

Bayes' Theorem: $$\Pr(B|A) = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A)}$$

Cond. Prob.: $$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$