

Let us now put to use our understanding of the coin-tossing model and the associated binomial probabilities.

We will solve the following problem.

We have a coin, which is tossed 10 times.

And we're told that exactly three out of the 10 tosses resulted in heads.

Given this information, we would like to calculate the probability that the first two tosses were heads.

This is a question of calculating a conditional probability of one event given another.

The conditional probability of event A, namely that the first two tosses were heads, given that another event B has occurred, namely that we had exactly three heads out of the 10 tosses.

However, before we can start working towards the solution to this problem, we need to specify a probability model that we will be working with.

We need to be explicit about our assumptions.

To this effect, let us introduce the following assumptions.

We will assume that the different coin tosses are independent.

In addition, we will assume that each coin toss has a fixed probability,  $p$ , the same for each toss, that the particular toss results in heads.

These are the exact same assumptions that we made earlier when we derived the binomial probabilities.

And in particular, we have the following formula that if we have  $n$  tosses, the probability that we obtain exactly  $k$  heads is given by this expression.

So now, we have a model in place and also the tools that we can use to analyze this particular model.

Let us start working towards a solution.

Actually, we will develop two different solutions and compare them at the end.

The first approach, which is the safest one, is the following.

Since we want to calculate a conditional probability, let us just start with the definition of conditional probabilities.

The conditional probability of an event given another event is the probability that both events happen, divided by the probability of the conditioning event.

Now, let us specialize to the particular example that we're trying to solve.

So in the numerator, we're talking about the probability that event A happens and event B happens.

What does that mean?

This means that event A happens-- that is, the first two tosses resulted in heads, which I'm going to denote symbolically this way.

But in addition to that, event B happens.

And event B requires that there is a total of three heads, which means that we had one more head in the remaining tosses.

So we have one head in tosses 3 all the way to 10.

As for the denominator, let's keep it the way it is for now.

So let's continue with the numerator.

We're talking about the probability of two events happening, that the first two tosses were heads and that in tosses 3 up to 10, we had exactly one head.

Here comes the independence assumption.

Because the different tosses are independent, whatever happens in the first two tosses is independent from whatever happened in tosses 3 up to 10.

So the probability of these two events happening is the product of their individual probabilities.

So we first have the probability that the first two tosses were heads, which is  $p$  squared.

And we need to multiply it with the probability that there was exactly one head in the tosses numbered from 3 up to 10.

These are eight tosses.

The probability of one head in eight tosses is given by the binomial formula, with  $k$  equal to 1 and  $n$  equal to 8.

So this expression, this part, becomes  $8 \text{ choose } 1$ ,  $p$  to the first power times  $1 - p$  to the seventh power.

So this is the numerator.

The denominator is easier to find.

This is the probability that we had three heads in 10 tosses.

So we just use this formula.

The probability of three heads is given by:  $10 \text{ tosses choose } 3$ ,  $p$  to the third, times  $1 - p$  to the seventh power.

And here we notice that terms in the numerator and denominator cancel out, and we obtain  $8 \text{ choose } 1$  divided by  $10 \text{ choose } 3$ .

And to simplify things just a little more, what is  $8 \text{ choose } 1$ ?

This is the number of ways that we can choose one item out of eight items.

And this is just 8.

And let's leave the denominator the way it is.

So this is the answer to the question that we had.

And now let us work towards developing a second approach towards this particular answer.

In our second approach, we start first by looking at the sample space and understanding what conditioning is all about.

In our model, we have a sample space.

As usual we can denote it by  $\omega$ .

And the sample space contains a bunch of possible outcomes.

A typical outcome is going to be a sequence of length 10.

It's a sequence of heads or tails.

And it's a sequence that has length 10.

We want to calculate conditional probabilities.

And this places us in a conditional universe.

We have the conditioning event B, which is some set.

And conditional probabilities are probabilities defined inside this set B and define the probabilities, the conditional probabilities of the different outcomes.

What are the elements of the set B?

A typical element of the set B is a sequence, which is, again of length 10, but has exactly three heads.

So these are the three-head sequences.

Now, since we're conditioning on event B, we can just work with conditional probabilities.

So let us find the conditional probability law.

Recall that any three-head sequence has the same probability of occurring in the original unconditional probability model, namely as we discussed earlier, any particular three-head sequence has a probability equal to this expression.

So three-head sequences are all equally likely.

This means that the unconditional probabilities of all the elements of B are the same.

When we construct conditional probabilities given an event B, what happens is that the ratio or the relative proportions of the probabilities remain the same.

So conditional probabilities are proportional to unconditional probabilities.

These elements of B were equally likely in the original model.

Therefore, they remain equally likely in the conditional model as well.

What this means is that the conditional probability law on the set B is uniform.

Given that B occurred, all the possible outcomes now have the same probability.

Since we have a uniform probability law, this means that we can now answer probability questions by just counting.

We're interested in the probability of a certain event, A, given that B occurred.

Now, given that B occurred, this part of A cannot happen.

So we're interested in the probability of outcomes that belong in this shaded region, those outcomes that belong within the set B. To find the probability of this shaded region occurring, we just need to count how many outcomes belong to the shaded region and divide them by the number of outcomes that belong to the set B.

That is, we work inside this conditional universe.

All of the elements in this conditional universe are equally likely.

And therefore, we calculate probabilities by counting.

So the desired probability is going to be the number of elements in the shaded region, which is the intersection of A with B, divided by the number of elements that belong to the set B.

How many elements are there in the intersection of A and B?

These are the outcomes or sequences of length 10, in which the first two tosses were heads-- no choice here.

And there is one more head.

That additional head can appear in one out of eight possible places.

So there's eight possible sequences that have the desired property.

How many elements are there in the set B?

How many three-head sequences are there?

Well, the number of three-head sequences is the same as the number of ways that we can choose three elements out of a set of cardinality 10.

And this is 10 choose 3, as we also discussed earlier.

So this is the same answer as we derived before with our first approach.

So both approaches, of course, give the same solution.

This second approach is a little easier, because we never had to involve any p's in our calculation.

We go to the answer directly.

The reason that this approach worked was that the conditional universe, the event  $B$ , had a uniform probability law on it.