CIS 2033 Lecture 5, Spring 2017[1]

Instructor: David Dobor

Updated January 8, 2017

[1] Supplemental reading from Dekking's textbook: Chapter 3.

In this lecture, we explore the idea of independence of events in some detail.

Introduction

We now introduce and develop the concept of independence between events. The general idea is the following. Whenever we are told that a certain event A has occurred, this generally changes the probability of some other event B, and the probabilities, our beliefs about A, will have to be replaced by conditional probabilities.

But if the conditional probability turns out to be the same as the unconditional probability, then the occurrence of event A does not carry any useful information on whether event B will occur. In such a case, we say that events A and B are independent.

We will develop some intuition about the meaning of *independence* of two events and introduce an extension, the concept of *conditional independence*. We will then proceed to define the independence of a collection of more than two events. If, for any two of the events in the collection we have independence between them, we will say that we have pairwise independence.

But we will see that independence of the entire collection is something different. It involves additional conditions. Finally, we will close with an application in reliability analysis and with a nice puzzle that will serve as a word of caution about putting together probabilistic models.

Coin Tossing Example

As an introduction to the main topic of this lecture, let us go through a simple example and on the way review what we have learned so far. The example that we're going to consider involves three tosses of a biased coin. It's a coin that results in heads with probability p. (We're going to make this a little more precise in a minute.) The coin may be biased in the sense that this number p is not necessarily the same as one half.

We represent this particular probabilistic experiment in terms a tree that shows us the different stages of the experiment. Each particular branch corresponds to a sequence of possible results in the different stages, and the leaves of this tree correspond to the possible outcomes. The branches of this tree are annotated by certain

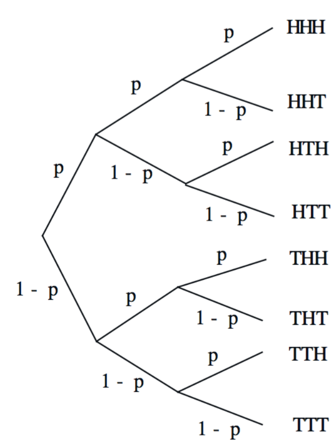

Figure 1: A model for the experiment consisting of tree coin tosses. The probability of an H or a T on any single toss is not necessarily 1/2 here.

numbers, and these numbers are to be interpreted appropriately as probabilities or conditional probabilities.

So for example, the number p that labels the branch eminating from the root of the tree is interpreted as the probability of heads in the first toss, an event that we denote as H_1 whose probability is $p = P(H_1)$, shown in red on figure 2.

Similarly, the p in the blue circle on figure 2 is to be interpreted as a *conditional* probability of obtaining heads in the second toss given that the first toss resulted in heads. And the "green p" is to be interpreted as the conditional probability of heads in the third toss, given that the first toss resulted in heads and the second toss also resulted in heads: $P(H_3 \mid H_1 \cap H_2)$. We sometimes write $P(H_3 \mid H_1 H_2)$ instead of $P(H_3 \mid H_1 \cap H_2)$ - same thing, just a slightly different notation.

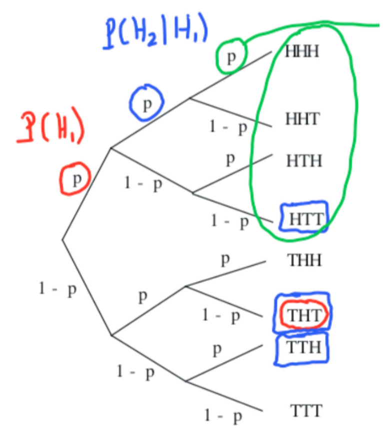

Figure 2: Same tree as in figure 1, but annotated. For example, the "blue p" is interpreted as the probability of H on the second toss, given that the first toss resulted in H as well, $P(H_2 \mid H_1)$.

Let us now continue with some calculations.

First, we're going to practice the multiplication rule, which allows us to calculate the probability of a certain outcome. In this case, the outcome of interest is tails followed by heads followed by tails, that is, the outcome THT. This outcome is the leaf circled in red on figure 2. According to the multiplication rule, to find the probability of a particular final outcome, we multiply probabilities and conditional probabilities along the path that leads to this particular outcome

$$\begin{aligned} P(THT) &= (1-p)\ p\ (1-p) \\ &= P(T_1)P(H_2 \mid T_1)P(T_3 \mid T_1 H_2) \end{aligned}$$

Let us now calculate the probability of the event that we obtain exactly one head in the three tosses. This is an event that can happen in multiple ways. In fact, there are exactly three ways in which this can happen, and these are shown in figure 2 as leaves that have blue rectangles around them.

To find the total probability of this event, we need to add the probability of the different outcomes that correspond to this event. That is, we add up the probabilities of the events (leaves in blue squares) as follows:

$$\begin{aligned} &P(\text{Got exactly one } H \text{ in three tosses of the coin}) = \\ &= P(HTT) + P(THT) + P(TTH) = \\ &= p\ (1-p)\ (1-p) + (1-p)\ p\ (1-p) + (1-p)\ (1-p)\ p \\ &= p\ (1-p)^2 + p\ (1-p)^2 + p\ (1-p)^2 \\ &= 3\ p\ (1-p)^2 \end{aligned}$$

Notice that each one of the 3 different ways that this event can happen have the same probability. So these 3 outcomes are equally likely.

Finally, let us calculate a conditional probability. Suppose that we were told that there was exactly one head. So in particular, one of the events in the blue rectangles has occurred. And we're interested in the probability that the first toss is heads, which corresponds to the three leaves circled in green in figure 3.

So we are after the following probability:

$$P(\text{first toss is } H \mid \text{exactly 1 } H \text{ occurred}) = ?$$

Can you can guess the answer? The answer should be 1/3. Why is that?

Each one of the outcomes in blue rectangles – HTT, THT, TTH – has the same probability. So when you condition on the outcome of one of these three having happened, the conditional probability of each particular outcome of the three should be 1/3. So, to reiterate, given that one of these tree happened, there's, in particular, probability 1/3 that the top one - the one that is also in the green circle that says HTT - has happened.

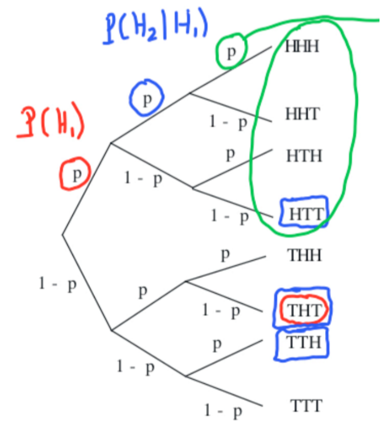

Figure 3: Same tree as in figures 1 and 2. Here, the three leaves that are circled in green have all resulted in the first toss being an H.

OK, so intuitively this probability should be 1/3. But let us see if we can derive this answer in a formal manner.

Let's use the definition of conditional probabilities:

$$\begin{aligned}&P(\text{first toss is } H \mid \text{exactly 1 } H \text{ occurred}) = \\ &\frac{P(H_1 \cap \text{exactly 1 } H \text{ occurred})}{P(\text{exactly 1 } H \text{ occurred})} = \\ &= \ldots\end{aligned}$$

Now, for the numerator, the probability of both events happening - i.e. that we have exactly one head and the first toss is heads - is the intersection of the event in the blue rectangle and the event in the green circle, which can happen only if the sequence HTT occurs, the probablity of which we calculated above. Above, we also calculated the denominator. So we continue:

$$\begin{aligned}\ldots &= \frac{p\,(1-p)^2}{3\,p\,(1-p)^2} = \\ &= 1/3\end{aligned}$$

OK, great! So we got the same answer formally as our intuition suggested - so far all is good with math. Or our intuition.

LET ME NOW MAKE A FEW COMMENTS about this particular example. This particular example is pretty *special* in the following respect.

We have that of the probability of H_2 (heads in the second toss), given H_1 (i.e. that the first one was heads), is equal to p. And the same is true for the conditional probability of heads in the second toss given that the first one was tails:

$$P(H_2 \mid H_1) = p = P(H_2 \mid T_1)$$

In other words, knowing the result of the first toss doesn't change our beliefs about what may happen, and with what probability, in the second toss.

Moreover, if you calculate the unconditional probability of heads in the second toss, what you would get using the total probability theorem would be the following:

$$P(H_2) = P(H_1)P(H_2 \mid H_1) + P(T_1)P(H_2 \mid T_1), \tag{1}$$

and if you plug in the values and do the algebra, this turns out to be equal to p again!

$$P(H_2) = p \cdot p + (1 - p) \cdot p = p$$

So the unconditional probability of heads in the second toss turns out to be the same as the conditional probabilities in equation (1). Again, knowing what happened in the first toss doesn't change your beliefs about the second toss, which were associated with this particular probability, p.

What we're going to do next is to generalize this special situation by giving a definition of independence of events, and then discuss various properties and concepts associated with independence.

Independence of Two Events

In the previous example, we had a model where the result of the first coin toss did not affect the probabilities of what might happen in the second toss. This is a phenomenon that we call *independence* which we now proceed to define.

Let us start with a first attempt at the definition. We have an event, B, that has a certain probability of occurring. We are then told that event A occurred, but suppose that this knowledge does not affect our beliefs about B in the sense that the conditional probability remains the same as the original unconditional probability. Thus, the occurrence of A provides no new information about B. In such a case, we may say that event B is independent from event A.

Thus let's write down our first - the so called "intuitive" - definition of independence:

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

If this is indeed the case, notice what is the probability that both A and B occur:

$$\begin{aligned} P(A \cap B) &= P(A)P(B \mid A) && \text{always true by multiplication rule} \\ &= P(A)P(B) && \text{also true if } B \text{ is independent of } A \end{aligned}$$

So we can find the probability of both events happening by just multiplying their individual probabilities. It turns out that this relation is a cleaner way of the defining formally the notion of independence. So we will say that two events, A and B, are independent if this relation holds:

Definition of independence: $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$

Why do we use this definition rather than the original one? This formal definition has several advantages.

First, it is *consistent* with the earlier definition. If this equality is true, then the conditional probability $P(B \mid A)$ will be equal to $P(B)$:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)\,P(B)}{P(A)} = P(B)$$

.

A MORE important reason is that this formal definition is symmetric with respect to the roles of A and B. So instead of saying that B is independent from A, based on this definition we can now say that events A and B are independent of each other.

IN ADDITION, since this definition is symmetric and since it implies the condition that $P(B \mid A) = P(B)$, it must also imply the symmetrical relation, namely, that $P(A \mid B) = P(A)$.

FINALLY, on the technical side, conditional probabilities are only defined when the conditioning event has non-zero probability. So this original definition would only make sense in those cases where the probability of the event A would be non-zero. In contrast, this new definition makes sense even when we're dealing with zero probability events. So this definition is indeed more general, and this also makes it more elegant.

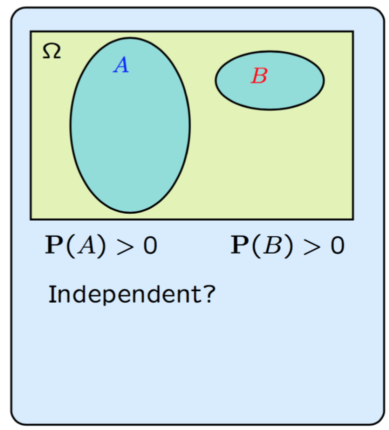

Figure 4: Two events, A and B. Are they independent?

Let us now build some understanding of what independence really is. Suppose that we have two events, A and B, both of which have positive probability. Furthermore, suppose that these two events are disjoint (i.e. they do not have any common elements). Are these two events independent?

Let us check the definition. The probability that both A and B occur is zero because the two events are disjoint. They cannot happen together. On the other hand, the probability of A times the probability of B is positive, since each one of the two terms is positive. And therefore these two events can not independent.

In fact, intuitively, these two events are as dependent as Siamese twins. If you know that A occurred, then you are sure that B did not occur. So the occurrence of A tells you a lot about the occurrence or non-occurrence of B.

So we see that *being independent is* something completely *different from being disjoint*. Independence is a relation about information. It is important to always keep in mind the intuitive meaning of independence. Two events are independent if the occurrence of one event does not change our beliefs about the other. It does not affect the probability that the other event also occurs.

When do we have independence in the real world? The typical case is when the occurrence or nonoccurrence of each of the two events A and B is determined by two physically distinct and noninteracting processes. For example, whether my coin results in heads and whether it will be snowing on New Year's Day are two events that should be modeled as independent. But I should also say that there are some cases where independence is less obvious and where it happens through a numerical accident.

Check Your Understanding: We have a peculiar coin. When tossed twice, the first toss results in Heads with probability 1/2. However, the second toss always yields the same result as the first toss. Thus, the only possible outcomes for a sequence of 2 tosses are HH and TT, and both have equal probabilities. Are the two events $A = \{\text{Heads in the first toss}\}$ and $B = \{\text{Heads in the second toss }\}$ independent?

Intuitively, the occurrence of event A gives us information on whether event B will occur, and therefore the two events are dependent.

Mathematically, $P(A) = P(B) = P(A \cap B) = 1/2$, so that $P(A \cap B) \neq P(A)P(B)$.

Check Your Understanding: Let A be an event, a subset of the sample space Ω. Are A and Ω independent?

Yes, because $P(A \cap \Omega) = P(A) = P(A) \cdot 1 = P(A)P(\Omega)$.

Intuitively, $P(A)$ represents our beliefs about the likelihood that A will occur. If we are told that Ω occurred, this does not give us any new information; we already knew that Ω is certain to occur. For this reason $P(A \mid \Omega) = P(A)$.

Check Your Understanding: When is an event A independent of itself? Choose one of the following possible answers:

(a) Always

(b) If and only if $P(A) = 0$

(c) If and only if $P(A) = 1$

(d) If and only if $P(A)$ is either 0 or 1.

Using the definition, A is independent of itself if and only if

$$P(A \cap A) = P(A) \cdot P(A).$$

Since $A \cap A = A$, we have $P(A \cap A) = P(A)$ and we obtain the equivalent condition

$$P(A) = P(A) \cdot P(A),$$

or

$$P(A) \cdot (1 - P(A)) = 0,$$

and this happens if and only if $P(A)$ is either 0 or 1.

Independence of Event Complements

Let us now discuss an interesting fact about independence that should enhance our understanding. Suppose that events A and B

are independent. Intuitively, if I tell you that A occurred, this does not change your beliefs as to the likelihood that B will occur. But in that case, this should not change your beliefs as to the likelihood that B will *not* occur. So A should be independent of B^c. In other words, the occurrence of A tells you nothing about B, and therefore tells you nothing about B^c either.

The previous paragraph gave an intuitive argument that if A and B are independent, then A and B^c are also independent.

But let us now verify this intuition through a formal proof. The formal proof goes as follows. We have the two events, A and B. Event A can be broken down into two pieces. One piece is the intersection of A with B - that's the piece shaded in red in figure 5. The second piece is the part of A which is outside B. And that piece is A intersection with the complement of B - that's the blue piece in figure 5. These red and blue pieces together comprise event A.

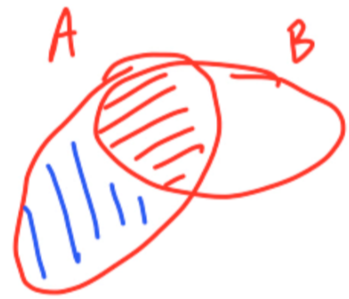

Figure 5: A is decomposed into the red and blue parts: $A = (A \cap B) \cup (A \cap B^c)$.

Now, these two pieces are disjoint from each other. And therefore, by the additivity axiom,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Using independence, the first term becomes $P(A) \cdot P(B)$, and we leave the second term as is:

$$P(A) = P(A)P(B) + P(A \cap B^c)$$

Now let us rearrange terms a little bit. From the previous two equations it follows that:

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c) \end{aligned}$$

So we proved that the probability of A and B^c occurring together is the product of their individual probabilities. And that's exactly the definition of A being independent from B^c. And this concludes the formal proof.

Check Your Understanding: Suppose that A and B are independent events. Are A^c and B^c independent?

We saw in this section that for any 2 generic events E_1 and E_2, independence of E_1 and E_2 implies independence of E_1 and E_2^c. In the case of this particular problem, we can apply this result with $E_1 = A$ and $E_2 = B$ to conclude that since A and B are assumed to be independent, then A and B^c are also independent.

Independence is symmetric, so A and B^c being independent is the same as B^c and A being independent. If we now reuse the generic result with $E_1 = B^c$ and $E_2 = A$, we can conclude that B^c and A^c are also independent, which by symmetry is the same as $A^c$and B^c being independent.

To summarize: A and B independent $\implies$ A and B^c independent $\implies$ B^c and A independent $\implies$ B^c and A^c independent $\implies$ A^c and B^c independent

Conditional Independence

Conditional probabilities are like ordinary probabilities, except that they apply to a new situation where some additional information is available. For this reason, any concept relevant to probability models has a counterpart that applies to conditional probability models. In this spirit, we can define a notion of conditional independence, which is nothing but the notion of independence applied to a conditional model.

Let us be more specific. Suppose that we have a probability model and two events, A and B. We are then told that event C occurred, and we construct a conditional model. Conditional independence is defined as ordinary independence but with respect to the conditional probabilities.

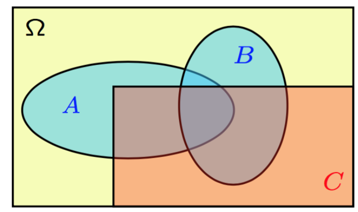

Figure 6: Are A and B independent given C?

To be more precise, remember that independence is defined in terms of this relation:

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

Now, in the conditional model we just use the same relation, but with conditional probabilities instead of ordinary probabilities:

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

So this is the definition of conditional independence.

We may now ask, is there a relation between independence and conditional independence? Does one imply the other? Let us look at an example. Suppose that we have two events and these two events are independent. We then condition on another event, C. And suppose that the picture is like the one shown in figure 7.

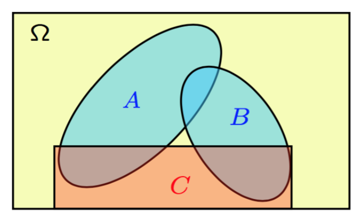

Figure 7: Are A and B independent given C?

Are A and B *conditionally* independent?

Well, in the new universe where C has happened, events A and B have no intersection. As we discussed earlier, this means that events A and B are "extremely dependent". Within C, if A occurs, this tells us that B did not occur.

THE CONCLUSION from this example is that *independence does not imply conditional independence*: in this particular example, we saw that given C, A and B are *not* independent.

Check Your Understanding: Suppose that A and B are conditionally independent given C. Suppose that $P(C) > 0$ and $P(C^c) > 0$.

1. Are A and B^c guaranteed to be conditionally independent given C?

2. Are A and B guaranteed to be conditionally independent given C^c?

1. We have seen that in any probability model, independence of A and B implies independence of A and B^c. The conditional model (given C) is just another probability model, so this property remains true.
2. This may be true in some special cases, e.g., if A and B both have zero probability. However, it is in general false. Suppose, for example, that events A and B have nonempty intersection inside C, and are conditionally independent, but have empty intersection inside C^c, which would make them dependent (given C^c).

Independence vs Conditional Independence

We have already seen an example in which we have two events that are independent but become dependent in a conditional model. Thus independence and conditional independence are not the same.

We will now see another example in which a similar situation is obtained. The example is as follows. We have two possible coins, coin A and coin B.

The top half of figure 8 shows the model of the world given that coin A has been chosen. In this conditional model, the probability of heads is 0.9. And, moreover, the probability of heads is 0.9 in the second toss no matter what happened in the first toss and so on as we continue.

$$P(H \mid \text{coin } A) = 0.9$$

A similar assumption is made in the other possible conditional universe. This is a universe in which we're dealing with coin B, the bottom half of figure 8. ,This time, the probability of heads at each toss is 0.1.

$$P(H \mid \text{coin } B) = 0.1$$

So given a particular coin, we assume that we have independent tosses. This is another way of saying that we're assuming conditional independence. Within this conditional model, that is, given that we've selected a particular coin, flips are independent.

Suppose now that we choose one of the two coins. Each coin is chosen with the same probability, 0.5. We then start flipping that chosen coin over and over.

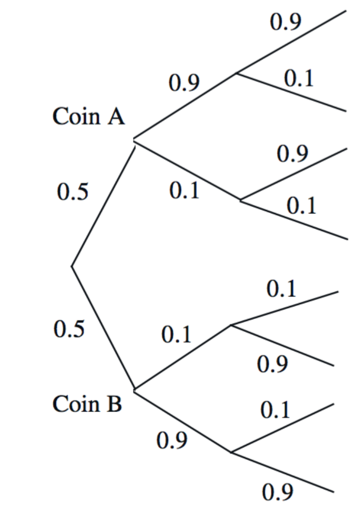

Figure 8: A model for flipping a coin, either coin A or coin B.

THE QUESTION we will try to answer is whether the coin tosses are independent. And by this we mean a question that *refers to the overall model* - in the general model, where you do not know ahead of time which of the two coins, coin A or coin B, is going to be selected and then flipped, are the different coin tosses independent?

We can approach this question by trying to compare conditional and unconditional probabilities. That's what independence is about. Independence is about certain conditional probabilities being the same as the unconditional probabilities. So this here, for example, let us compare whether the 11-th coin toss is dependent or independent from what happened in the first 10 coin tosses. I.e:

$$\begin{aligned} P(\text{ toss 11 is } H) \quad &\textit{to be compared with}: \\ P(\text{ toss 11 is } H \mid \text{ first 10 tosses are H's}) \end{aligned}$$

Let us calculate these probabilities. For the first one, we use the total probability theorem. There's a certain probability that we have coin A, and then we have the probability of heads in the 11th toss given that it was coin A. There's also a certain probablility that it's coin B and then a conditional probability that we obtain heads given that it was coin B:

$$P(\text{ toss 11 is } H) = P(A)P(H_{11} \mid A) + P(B)P(H_{11} \mid B)$$

We use the numbers that are given in this example. We have 0.5 probability of obtaining a particular coin, 0.9 probability of heads for coin A, 0.5 probability that it's coin B, and 0.1 probability of heads if it is indeed coin B.

We do the arithmetic, and we find that the answer is 0.5:

$$\begin{aligned} P(\text{ toss 11 is } H) &= 0.5 \times 0.9 + 0.5 \times 0.1 \\ &= 0.5 \end{aligned}$$

which makes perfect sense: we have coins with different biases, but the average bias is 0.5. If we do not know which coin it's going to be, the average bias is going to be 0.5. So the probability of heads in any particular toss is 0.5 when we do not know which coin it is.

Suppose now that someone told you that the first 10 tosses were heads. Will this affect your beliefs about what's going to happen in the 11th toss?

We can calculate this quantity using the definition of conditional probabilities, or the Bayes' rule, but let us instead think intuitively. *If* it is coin B, the events of 10 heads in a row is *extremely unlikely*. So if

I see 10 heads in a row, then I should conclude that there is almost certainty that I'm dealing with coin A.

The information that I'm given tells me that I'm extremely likely to be dealing with coin A in the case when I see 10 H's in a row. So we might as well condition on this equivalent information that it is coin A that I'm dealing with. But if it is coin A, then the probability of heads is going to be equal to 0.9.

$$\begin{aligned} P(\text{ toss 11 is } H \mid \text{ first 10 tosses are H's}) \approx \\ \approx P(H_{11} \mid A) = 0.9 \end{aligned}$$

So the conditional probability is quite different from the unconditional probability. The information on the first 10 tosses *affects my beliefs* about what's going to happen in the 11th toss.

Therefore, we do not have independence between the different tosses.

Independence of a Collection of Events

Suppose I have a fair coin which I toss multiple times. I want to model a situation where the results of previous flips do not affect my beliefs about the likelihood of heads in the next flip. And I would like to describe this situation by saying that the coin tosses are independent.

You may say: we already defined the notion of independent events - doesn't this notion apply? Well, not quite. We defined independence of *two* events. But here, we want to talk about independence of a collection of events. For example, we would like to say that the events, heads in the first toss, heads in the second toss, heads in the third toss, and so on, are *all* independent. What is the right definition?

Let us start with intuition. We will say that a family of events is independent if knowledge about some of the events doesn't change my beliefs, my probability model, for the remaining events.

Intuitive "definition": Information on some of the events does not change probabilities related to the remaining events

For example, if I want to say that events A_1, A_2 and so on are independent, I would like relations such as the following to be true: the probability that event A_3 happens and A_4 does not happen remains the same even if I condition on some information about some other events, like the event that A_1 happened or that both A_2 happened and A_5 did not happen:

$$\text{If } \; A_1, A_2, \dots \; \text{ are independent, then}$$

$$\Rightarrow P(A_3 \cap A_4^c) = P(A_3 \cap A_4^c \mid A_1 \cup (A_2 \cap A_5^c)) \tag{2}$$

The important thing to notice here is that the indices involved in the event of interest are distinct from the indices associated with the events on which I'm given some information. I'm given some information about the events A_1, A_2, and A_5, and this information does not affect my beliefs about something that has to do with events A_3 and A_4.

We would like all relations of the kind to be true. This could be one possible definition, just saying that the family of events are independent if and only if any relation of this type is true. But such a definition would not be aesthetically pleasing. Instead, we introduce the following definition, which mimics our earlier definition of independence of two events.

Definition: We will say that a collection of events are independent if you can calculate probabilities of intersections of these events by multiplying individual probabilities. And this should be possible for all choices of indices involved and for any number or events involved:

Definition: Events $A_1, A_2, \ldots, A_n$ are called **independent** if:
$\mathbf{P}(A_i \cap A_j \cap \cdots \cap A_m) = \mathbf{P}(A_i)\mathbf{P}(A_j)\cdots\mathbf{P}(A_m)$ for any distinct indices $i, j, \ldots, m$

Let us translate this into something concrete. Consider the case of three events, A_1, A_2, and A_3. Our definition requires that we can calculate the probability of the intersection of two events by multiplying individual probabilities.

$n = 3$:

$$\left.\begin{aligned} \mathbf{P}(A_1 \cap A_2) &= \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \\ \mathbf{P}(A_1 \cap A_3) &= \mathbf{P}(A_1) \cdot \mathbf{P}(A_3) \\ \mathbf{P}(A_2 \cap A_3) &= \mathbf{P}(A_2) \cdot \mathbf{P}(A_3) \end{aligned}\right\} \text{ pairwise independence}$$

And we would like all of these three relations to be true, because this property should be true for any choice of the indices. What do we have here? The first relation tells us that A_1 and A_2 are independent. The second one tells us that A_1 and A_3 are independent. The third - that A_2 and A_3 are independent. We call this situation *pairwise independence*.

But the definition requires *something more*. It requires that the probability of three-way intersections can also be calculated the same way by multiplying individual probabilities. And this additional condition does make a difference, as we're going to see in a later example.

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \cdot \mathbf{P}(A_3)$$

Is this the right definition? Yes.

One can prove formally that if the conditions in this definition are satisfied, then any formula of the kind shown in equation (1) above is true. In particular, we also have relations such as the following.

$$P(A_3) = P(A_3 \; A_1 \cap A_2) = P(A_3 \; A_1 \cap A_2^c) = P(A_3 \; A_1^c \cap A_2)$$

So any kind of information that I might give you about events A_1 and A_2 – which one of them occurred and which one didn't– is not going to affect my beliefs about the event A_3. The conditional probabilities are going to be the same as the unconditional probabilities. I said earlier that this definition implies that all relations of this kind are true. This can be proved. The proof is a bit tedious. And we will not go through it.

Check Your Understanding: Suppose that A, B, C and D are independent. Use intuitive reasoning (not a mathematical proof) to answer the following.

1. Is it guaranteed that $A \cap C$ is independent from $B^c \cap D$?

2. Is it guaranteed that $A \cap B^c \cap D$ is independent from $B^c \cup D^c$?

2.

1.

Independence versus Pairwise Independence

We will now consider an example that illustrates the difference between the notion of independence of a collection of events and the notion of pairwise independence within that collection.

The model is simple. We have a fair coin which we flip twice. At each flip, there is probability 1/2 of obtaining heads.

Furthermore, we assume that the two flips are independent of each other. Let H_1 be the event that the first coin toss resulted in heads, which corresponds to the upper two squares labeled HH and HT in this diagram. Let H_2 be the event that the second toss resulted in heads, which corresponds to the top left and bottom left squares – the two ways that we can have the second toss being heads.

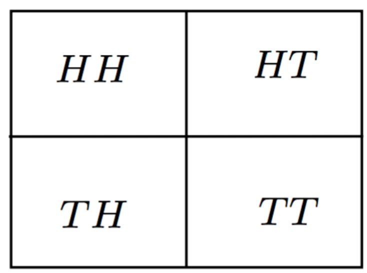

Figure 9: Two tosses of a coin.

Here's our setup:

- H_1: First toss is H - top left and top right squares in figure 9.
- H_2: Second toss is H - top left and bottom left squares in Figure 9.
- $P(H_1) = P(H_2) = 1/2$

Now, we're assuming that the tosses are independent. So the event HH has a probability which is equal to the probability that the first toss resulted in heads – that's 1/2 – times the probability that the second toss resulted in heads, which is 1/2. So we have probability 1/4 for this outcome:

$$P(HH) = P(H)P(H) = 1/2 \times 1/2 = 1/4$$

By a similar argument, we have that the probabilities of the other three events in figure 9 are:

$$P(HT) = P(TH) = P(TT) = 1/4$$

Let us now introduce a new event, call it C, which is the event that the two tosses had the same result. So this is the event that we obtain either HH or TT:

- $C = \{HH, TT\}$: two tosses had the same result – top left and bottom right squares in figure 9.

Is this event C independent from the events H_1 and H_2?

Let us first look for pairwise independence. Let's look at the probability that H_1 occurs and C occurs as well:

$$P(H_1 \cap C) = P(HH) = 1/4 \tag{3}$$

(This is true because the first toss resulted in heads, and the two tosses had the same result, which is the same as the probability of obtaining heads followed by heads.)

How about the product of the probabilities of H_1 and of C? Is it the same? Well, the probability of H_1 is 1/2. And the probability of C– what is it? Event C consists of two outcomes. Each one of these outcomes has probability 1/4. So the total is, again, 1/2.

$$P(H_1) \times P(C) = 1/2 \times 1/2 = 1/4 \tag{4}$$

So we notice that the probability of the two events happening is the same as the product of their individual probabilities, and therefore, H_1 and C *are independent events*.

By the same argument, H_2 and C are going to be independent. It's a symmetrical situation.

Similarly, you can check that H_1 and H_2 are also independent from each other. So we have all of the conditions for *pairwise* independence.

Let us now check whether we have independence. To check for independence, we need to also look into the probability of all three events happening and see whether it is equal to the product of the individual probabilities.

The probability of all three events happening – this is the probability that H_1 occurs and H_2 occurs and C occurs. What is this event?

Heads in the first toss, heads in the second toss, and the two tosses are the same– this happens if and only if the outcome is heads followed by heads. And this has probability 1/4:

$$P(H_1 \cap H_2 \cap C) = P(HH) = 1/4$$

On the other hand,

$$P(H_1) \times P(H_2) \times P(C) = 1/2 \times 1/2 \times 1/2 = 1/8$$

So $P(H_1 \cap H_2 \cap C) \neq P(H_1) \times P(H_2) \times P(C)$.

Therefore in this example, H_1, H_2, and C are pairwise independent, but they're *not* independent in the sense of an independent collection of events. How do we understand this intuitively?

If I tell you that event H_1 occurred and I ask you for the conditional probability of C given that H_1 occurred, what is this? The only way that you can have the two tosses having the same result is going to be in the second toss also resulting in heads: $P(C \mid H_1) = P(H_2 \mid H_1)$.

Furthermore, since H_2 and H_1 are independent, this is just the probability that we have H in the second toss. And this number is 1/2: $P(H_2 \mid H_1) = P(H_2) = 1/2$. And 1/2 is also the same as $P(C)$.

To recap, we have:

$$P(C \mid H_1) = P(H_2 \mid H_1) = P(H_2) = 1/2 = P(C)$$

This is another way of understanding the independence of H_1 and C: Given that the first toss resulted in H, this does not help you in any way in guessing whether the two tosses will have the same result or not. The first one was H, but the second one could be either H or T with equal probability. Thus event H_1 does not carry any useful information about the occurrence or non-occurrence of event C.

On the other hand, if I were to tell you that both events, H_1 and H_2, happened, what would the conditional probability of C be? If both H_1 and H_2 occurred, then the results of the two coin tosses were identical, so you know that C also occurred. So this probability must equal to 1:

$$P(C \mid H_1 \cap H_2) = 1$$

And this number, 1, is different from the unconditional probability of C, which is 1/2.

So we have here a situation where knowledge of H_1 having occurred does not help you in making a better guess on whether C is going to occur. H_1 by itself does not carry any useful information. But the two events together, H_1 and H_2, do carry useful information about C.

Once you know that H_1 and H_2 occurred, then C is certain to occur, so your original probability for C, which was 1/2, now gets revised to a value of 1. This means that H_1 and H_2 do carry information relevant to C. Therefore, C is not independent from these two events collectively.

And we say that events H_1, H_2, and C are *not* independent.

Reliability

//TO BE ADDED LATER

Please see your class notes for an example on reliability of components.