Lecture 5 - Independence

6.042 - February 20, 2003

Suppose that I flip two fair coins, a nickel and a quarter. What is the probability that both come up heads?

The obvious answer is $\frac{1}{4}$. However, if you go back and read the question above *very carefully*, you'll see that we've not explicitly ruled out the possibility that, prior to the experiment, the two coins were carefully placed side-by-side, one heads-up and one heads-down, and then wrapped in eighty-seven layers of duct tape. And— if that were to happen!— the probability of both coins coming up heads would be pretty much zero. In giving the answer $\frac{1}{4}$, we were making an implicit assumption about the coins; namely, we assumed that they behaved *independently*.

1 Independent Events

The definition below formalizes the notion of independent events.

Definition 1 *Events A and B are* independent *if:*

$$\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}$$

We can restate the original question more precisely using this notion of independence. Suppose that you flip two fair, independent coins, a nickel and a quarter. What is the probability that both come up heads? Let A be the event that the nickel comes up heads, and let B be the event that the quarter comes up heads. Then we are interested in the probability of $A \cap B$, the event that both coins come up heads. We can compute this probability as follows:

$$\begin{aligned} \Pr\{A \cap B\} &= \Pr\{A\} \cdot \Pr\{B\} \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4} \end{aligned}$$

The first step uses the fact that the events are independent, and the second uses the fact that the coins are fair.

The statement of the question is still somewhat sloppy. In particular, we said that the *coins* are independent, when actually we meant that the *events* that the nickel is heads and that the quarter is heads are independent. The difference between these statements is not so great, however, because of the following fact:

Fact 2 *If events A and B are independent, then so are events A and $\overline{B}$*

So if we assume that the nickel showing heads is independent of the quarter showing heads, then every obvious event involving only the nickel is independent of every event involving only the quarter. For example, the event that the nickel shows tails is independent of the event that the quarter shows heads. So it seems simpler just to say that the coins are independent.

The following theorem gives an equivalent way to characterize independence, which may seem more intuitive:

Theorem 3 *Events A and B are independent if and only if:*

$$\Pr\{A \mid B\} \quad = \quad \Pr\{A\} \quad \textit{(and } \Pr\{B\} > 0\textit{)}$$

or else $\Pr\{B\} = 0$.

In words, if the occurrence of event B makes event A neither more nor less likely, then the two events are independent. In the context of our two-coin example, this says that the nickel coming up heads makes the quarter neither more nor less likely to come up heads. The theorem treats $\Pr\{B\} = 0$ as a special case, because then $\Pr\{A \mid B\}$ is not defined. On that score, the theorem says that every event is independent of an event that never happens.

Let's prove the theorem.

Proof. We must show that the definition of independence implies the alternative condition stated in the theorem and vice versa.

First, suppose that events A and B are independent. We show that the alternative condition holds. There are two cases. If $\Pr\{B\} = 0$, then the condition holds immediately. Otherwise, $\Pr\{B\} > 0$ and we have:

$$\begin{aligned}
\Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\
&= \frac{\Pr\{A\} \cdot \Pr\{B\}}{\Pr\{B\}} \\
&= \Pr\{A\}
\end{aligned}$$

Again, the condition holds. The first step uses the definition of conditional probability, the second uses the definition of independence, and the last step is simplification.

Now suppose that the alternative condition holds. We show that events A and B are independent. There are two cases. If $\Pr\{B\} = 0$, then $\Pr\{A \cap B\} = 0$ by the monotonicity law and $\Pr\{A\} \cdot \Pr\{B\} = \Pr\{A\} \cdot 0 = 0$. Since both expressions are equal to zero, they are equal to each other, and so events A and B are independent. Otherwise, $\Pr\{B\} > 0$, and we can reason as follows:

$$\begin{aligned} \Pr\{A\} \cdot \Pr\{B\} &= \Pr\{A \mid B\} \cdot \Pr\{B\} \\ &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \cdot \Pr\{B\} \\ &= \Pr\{A \cap B\} \end{aligned}$$

The first step uses the condition in the theorem statement, the second uses the definition of conditional probability, and the third step uses only simplification. We conclude that events A and B are independent. $\square$

1.1 A Note on Proof Technique

Note that the proof of Theorem 3 contained two equation sequences. In each, we transformed an initial expression into another expression, carefully justifying each step. This is a good strategy in general. A common error that you should avoid in proofs is writing down one equation, modifying *both* sides until you reach a valid equation, and then claiming that the original statement holds. For example, one might reason:

$$\begin{aligned} 3 &= 4 \\ 3 \cdot 0 &= 4 \cdot 0 \\ 0 &= 0 \end{aligned}$$

The first equation implies the last, but the last equation does *not* imply the first. Be careful to avoid this pitfall!

1.2 The Independence Relation

We can regard *independence* as a relation between two events. If we do so, what properties does the independence relation have? Is it reflexive? Symmetric? Transitive?

- Independence is *not* reflexive. Let A be the event that a fair coin comes up heads. Then we have:

$$\begin{aligned} \Pr\{A \cap A\} &= \Pr\{A\} \\ &= \frac{1}{2} \end{aligned}$$

  On the other hand, we have:

$$\begin{aligned} \Pr\{A\} \cdot \Pr\{A\} &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4} \end{aligned}$$

  Since these two quantities are not equal, event A is not independent of itself.

- Independence *is* symmetric. Event A is independent of event B if:

$$\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}$$

  Commutativity of intersection and multiplication implies:

$$\Pr\{B \cap A\} = \Pr\{B\} \cdot \Pr\{A\}$$

  and so event B is independent of event A.

- Independence is *not* transitive. Let A and B be two independent events. Then A is independent of B and B is independent of A, but A need not be independent of A, as we have seen.

1.3 Which Events Are Independent?

Is astrology for real? Is your lovelife affected by the position Mars? The obvious answer is, "Of course not!" But, then again, the gravitational pull of Mars *could* have plucked an asteroid out of orbit and sent it hurtling toward the left side of row eight in 26-100, where you're preparing to put the moves on your hot date during the LSC feature presentation.

We may choose to *assume* that real-world events are independent when we are constructing a probabilistic model. But this is an assumption— much like the assumption that a coin is fair— not a mathematically verifiable fact. Of course, once we have built a probabilistic model, we may be able prove that certain events within that model are independent.

2 Mutual Independence

Independence becomes more complicated when more than two events are involved. Let's set up a situation that illustrates the problem. Suppose that we flip two fair, independent coins.

- Let A be the event that the first coin shows heads.
- Let B be the event that the second coin shows heads.
- Let C be the event that the coins agree; that is, they both show heads or both show tails.

We have already seen that events A and B are independent. Are events B and C independent? Your intuition may tempt you to say, "no"; after all, both events involve the second coin toss, and so they are clearly related to some degree. However, independence is not defined in terms of spacial or temporal disconnection. Rather, events B and C are independent if:

$$\Pr\{B \cap C\} \quad = \quad \Pr\{B\} \cdot \Pr\{C\}$$

The probability on the left is $\frac{1}{4}$, since the event $B \cap C$ contains only the outcome where both coins come up heads. The probabilities on the right are each $\frac{1}{2}$. Therefore, the equation holds, and so events B and C are independent. By similar reasoning, events A and C are independent as well.

We have shown that among the events A, B, and C, each *pair* of events is independent. On that basis, the events A, B, and C are considered *pairwise independent*. This is actually a rather weak property for a set of events to have. For example, for pairwise independent events, we can *not* conclude that:

$$\Pr\{A \cap B \cap C\} \quad = \quad \Pr\{A\} \cdot \Pr\{B\} \cdot \Pr\{C\}$$

In fact, in our example, the probability on the left is $\frac{1}{4}$, but the expression on the right is actually $\frac{1}{8}$.

Qualitatively, the pairwise independent events A, B, and C are sort of independent, but not as independent as, say, the three faces shown when three normal coins are flipped. Mathematically, the issue is that the events A, B, and C are not *mutually independent*.

Definition 4 *Events $A_1, A_2, \ldots, A_n$ are* mutually independent *if for every subset of those events, the probability of their intersection is equal to the product of their probabilities.*

In other words, events $A_1, A_2, \ldots, A_n$ are mutually independent if all of the following conditions hold:

$$\begin{aligned}
\Pr\{A_i \cap A_j\} &= \Pr\{A_i\} \cdot \Pr\{A_j\} \quad \forall \text{ distinct } i, j \in \{1, \ldots, n\} \\
\Pr\{A_i \cap A_j \cap A_k\} &= \Pr\{A_i\} \cdot \Pr\{A_j\} \cdot \Pr\{A_k\} \quad \forall \text{ distinct } i, j, k \in \{1, \ldots, n\} \\
&\ldots \\
\Pr\{A_1 \cap A_2 \cap \ldots \cap A_n\} &= \Pr\{A_1\} \cdot \Pr\{A_2\} \cdots \Pr\{A_n\}
\end{aligned}$$

The events A, B, and C in our earlier experiment are not mutually independent, because the probability of the intersection of all three events is not equal to the product of their probabilities.

Mutually independent events are usually just described as "independent", when there is no chance of confusion with, say, pairwise independence.

3 The Birthday Paradox

Suppose that there are 23 people in a room. What is the probability that two of them share the same birthday? One might expect the probability to be rather low, since there are 365 different days to choose from. Surprisingly, some two people are more likely to share a birthday than not.

Let's generalize the problem by making the specific numbers into variables. In particular, suppose there are n days in a year, and number the people in the room $1, 2, \ldots, m$. Furthermore, let's make some simplifying assumptions:

- No one was born on February 29 (leap day).
- For each person, all n possible birthdays are equally likely.
- The birthdays of the people are mutually independent; that is, the probability that a person was born on a given day is $\frac{1}{n}$ regardless of the days on which the other people were born.

Of course, none of these assumptions is completely valid. Some people really are born on leap days, an unusually large number of people are born nine months after romantic holidays, and the birthdays of twins are not independent. But, much like a back-of-the-envelope calculation, these assumptions are good enough to lead to useful insights.

There is a trick that makes this problem— and many others— easier. We do not compute the probability that two people have the same birthday directly. Rather, we compute the

probability of the complementary event, that the birthdays of the people in the room are all distinct. Our final answer is then one minus the initial answer, because for every event E:

$$\Pr\left\{\overline{E}\right\} \quad = \quad 1 - \Pr\left\{E\right\}$$

Let D_i be the event that the first i people all have different birthdays. Our goal is to determine the probability of D_m, the event that all m people have distinct birthdays. The total probability law says:

$$\Pr\left\{D_i\right\} \quad = \quad \Pr\left\{D_i \mid D_{i-1}\right\} \cdot \Pr\left\{D_{i-1}\right\} + \Pr\left\{D_i \mid \overline{D_{i-1}}\right\} \cdot \Pr\left\{\overline{D_{i-1}}\right\}$$

Of course, there is no way for the first i people to all have different birthdays if the first $i-1$ people do not. Therefore, $\Pr\left\{D_i \mid \overline{D_{i-1}}\right\} = 0$, and we get a simpler equation:

$$\Pr\left\{D_i\right\} \quad = \quad \Pr\left\{D_i \mid D_{i-1}\right\} \cdot \Pr\left\{D_{i-1}\right\}$$

Now let's tackle the term $\Pr\left\{D_i \mid D_{i-1}\right\}$. This is the probability that the first i people have different birthdays, given that the first $i-1$ people have different birthdays. This happens unless the birthday of the i-th person falls on one of those $i-1$ birthdays already taken. By our independence and uniformity assumptions, such a coincidence happens with probability $(i-1) \cdot \frac{1}{n}$. Consequently, a coincidence is avoided with probability $1 - \frac{i-1}{n}$. Substituting this into the equation above, we find:

$$\Pr\left\{D_i\right\} \quad = \quad \left(1 - \frac{i-1}{n}\right) \cdot \Pr\left\{D_{i-1}\right\}$$

Applying this formula repeatedly, we can compute the probability that all m people have different birthdays.

$$\begin{aligned}
\Pr\left\{D_m\right\} &= \left(1 - \frac{m-1}{n}\right) \cdot \Pr\left\{D_{m-1}\right\} \\
&= \left(1 - \frac{m-1}{n}\right) \cdot \left(1 - \frac{m-2}{n}\right) \Pr\left\{D_{m-2}\right\} \\
&\cdots \\
&= \left(1 - \frac{m-1}{n}\right) \cdot \left(1 - \frac{m-2}{n}\right) \cdots \left(1 - \frac{1}{n}\right) \cdot D_1 \\
&= \prod_{i=1}^{m-1} \left(1 - \frac{i}{n}\right)
\end{aligned}$$

In the last step, we use the observation that $D_1 = 1$; that is, the probability that everyone has a unique birthday in a room with only one person is 1.

We can now answer our original question by plugging numbers into the product formula above. Assuming a year with 365 days, the probability that 22 people all have different birthdays is 0.524. The probability that 23 people all have different birthdays is 0.493; that is, it is more likely than not that some two people share a birthday!

3.1 The Birthday Principle

Restating the birthday problem in more generic terms suggests why it is important: if I throw m balls uniformly and independently at random into n bins, what is the probability that some bin gets more than one ball? (Imagine that the bins are marked with days of the year, "June 7", "March 25", etc., and the balls are labeled with people's names. Then throwing the balls into the bins at random is then closely analogous to assigning random birthdays to people.) Remarkably, there are problems all across computer science— in hashing, cryptography, and coding theory— that boil down to this same question. So we'll take a closer look at the solution to the birthday problem and extract a rule of thumb that will be useful to you time and again.

Our current solution is a big product. Though you can evaluate this product with a computer, it is too complex to give much insight. For example, *roughly* how many balls can I throw at random into a collection of 1,000,000 bins before the probability that some bin contains more than one ball rises to about $\frac{1}{2}$?

To answer such questions, we can derive an approximate solution. To begin, recall the Taylor series for e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

If x is close to zero, then the first two terms on the right are much larger than all the others. This observation suggests an approximation formula:

$$1 + x \approx e^x$$

As it turns out, this is actually an inequality; $1 + x \leq e^x$ for all real x.

Let's see how this approximation simplifies the solution to the birthday problem:

$$\begin{aligned}
\Pr\{D_m\} &= \prod_{i=1}^{m-1}\left(1-\frac{i}{n}\right) \\
&\approx \prod_{i=1}^{m-1} e^{-i/n} \\
&= e^{-(1/n+2/n+\ldots+(m-1)/n)} \\
&= e^{-\frac{m(m-1)}{2n}}
\end{aligned}$$

The last step uses the summation formula:

$$1+2+\ldots+(m-1) = \frac{m(m-1)}{2}$$

Now we have an approximate expression that we can evaluate on an ordinary calculator!

Moreover, we can extract a general principle. If $m=\sqrt{n}$, then the final expression for the probability that m people have different birthdays is very nearly $e^{-1/2}$, which is about 60%. We can restate this conclusion as a rule of thumb called the *Birthday Principle*.

Fact 5 *If you throw $\sqrt{n}$ balls uniformly and independently at random into n bins, then there are about even odds that some bin contains more than one ball.*

This rule says, for example, that I can only thrown about 1000 balls at random into a million bins before some bin get two balls.

Learning to use approximations such as $1+x\approx e^x$ is important for a computer scientist. As was the case in the birthday problem, the analysis of a system often leads to formulas or equations that are too complex to handle directly. However, with a few deft approximations, the solution may fall right out!