[SQUEAKING]

[RUSTLING]

[CLICKING]

**ERIK DEMAINE:** All right. Welcome back to 6.1200. Today we continue our theme of probability, as we will continue to do for the next few lectures. First, a recap from last class and last recitation. A key concept we'll be using today is conditional probability. Remember this notation? It's the probability of event A occurring given that you already know that B occurred, which is like, if you imagine this entire blackboard as the sample space, all of the possibilities of what could happen, and we have events. A is some subset of that space. B is some subset of that space.

What the vertical bar B means, conditioning on B means focusing on just this red subset of where B is. And just think about the probability of A happening relative to those events, instead of thinking the probability of A happening relative to all events. And you can compute that as the probability of this intersection divided by the probability of B. So we're just focusing in-- assuming B already happened, then this becomes our entire sample space that we can restrict to.

OK, and then we can use that, for example, to compute probabilities of events. So this was the law of total probability. Probability of an event A is if you pick your favorite event, E, sort of like a proof by cases, you get to choose any event E you want. That's whatever's most convenient. Then you can compute probability of A by splitting into two cases-- one where E occurs, and one where E does not occur.

So you take the probability that E occurs, you multiply it by the probability that A occurs given E, and you do the same for not E. And because E and not E are by definition disjoint events, exactly one of them happens, this works out. You have the product rule here between these events and the sum rule between these events.

I think, easiest to see in this tree. I tend to draw my trees vertically. You can draw them horizontally, whatever. We have event E occurring or not occurring. And then separately, we have event A or not A occurring. And any of these could happen in any combination. When we write down a tree, usually we write on the edges conditional probabilities. At the beginning there's no condition, so it's just probability of E on one side and probability of not E on the other side.

But then usually on this branch, we write, what's the probability of A given that E occurred? And on this branch, what's the probability that not A occurred, given that E occurred? This whole branch is that E occurred, and then we care about-- given that we got here, what's the probability of A happening? What's the probability of not A happening? And similarly on the other side, what's the probability of A and not A happening, given that E didn't happen?

Those are all the cases. And then you can compute the probability of each of these individual situations by taking the product of these probabilities down the path. That's what we have shown in the past. And so in particular, if you want to compute the probability of A, A is these two nodes together. These are the two A branches. And so that is this product plus this product. By construction, everything here is disjoint. And so you can just add the probabilities up and you get this law. We may use it today.

Today I want to talk about a related concept which is independence. So today is Independence Day, not starring Will Smith, or any other modern versions of *Independence Day.* And the tagline for *Independence Day* back in the '90s-- I was alive back then-- it says, "The question of whether or not we are alone in the universe has been answered." Are we independent from the rest of the universe, or are we dependent? It turns out the answer is, we're independent if and only if the probability of A given B is the probability of A. That's-- sorry, spoilers for the movie, but that's what you get at the end.

So that's our mathematical definition of A is independent of B if probability of A given B, probability of A conditioned on B, is just the probability of A again. So in other words, conditioning on B does not affect the probability of A. With the condition and without the condition, you get the same thing.

We have to also say or the probability of B equals 0. For this annoying technical reason, which I forgot to mention up here, probability of A given B is not defined when probability of B is 0, because then you get a division by 0 error. We don't like those. So I don't even know what this means, when the probability of B is 0. So we need to say whether we're independent in that case, and we are. So it turns out the empty event, which never happens, is independent of everything. Also the full event, which always happens, is independent of everything.

So this is a way to measure whether two things are independent, meaning whether B happened has any impact on whether A will happen. That's intuitively what this means. And it's a useful concept. Here is-- wrong page-- an example that gives you some intuition about why the Monty Hall problem comes out the way it does. So let's do Monty Hall.

I'm going to define two events. One is door 1 has a prize. So I forget the door names that Brynmor used. So I'm just going to call them door number 1, 2, and 3, just like in the game show. And I don't have fancy colored boxes.

OK, I'm not going to repeat the whole analysis of Monty Hall, but I just want you to think about these two events. So event A is that door 1 has the prize. We know without anything else, this happens with 1/3 probability. Event B is that door 2 is revealed to not have the prize after the player chooses door 1.

OK, this is a bit of a mouthful. But let's just suppose that the player starts by choosing door number 1. In the analysis we did, we had a uniform random choice of choosing all three doors. It doesn't actually matter which door you choose, because everything else is random. It's fine if you always choose door 1. But whether it's random, whether it's deterministic, let's just suppose the player chooses door 1. And event B is, will door number 2 be revealed to not have the prize? And now my question is, is Event A and event B, are they independent from each other? And I claim that they are.

And the way we can determine that is by computing these two probabilities, probability of A and probability of A given B. Now, probability of A, we know is 1/3, right? This is just, does door 1 have the prize? So that's uniform random, so 1/3 for that. And so for this to be independent, it better be the case that probability of A given B is also 1/3.

And if you think about it, that is one of the things we computed, because what we computed was, what's the probability of the switching rule winning? So we're supposing here the player chose door 1. And so if in the switching strategy, they will switch to, in the condition on B means that the player chose door 1, and door 2 was revealed not to have the prize, and so we would switch to door 3. And we worked out that the winning probability in the switching strategy was 2/3. So that means the probability that door 1 has the prize must be one third, because the probability that door 3 has the prize is 2/3. And 1 minus 2/3 is 1/3. So indeed, these are equal, and the events are independent.

And while I don't claim this is super-intuitive yet-- I mean, it's like you stare at it for an hour, and then it's intuitive. [LAUGHS] So it's not especially intuitive, but I think while before we saw you can just compute everything and find the answer, I think this does give a little intuition. The idea is that revealing door 2 to not have the prize doesn't tell you anything about whether door 1 has the prize. And that's intuitively why switching is good, because you've reduced down to two possibilities instead of three. If you stick to A, it will be as if no doors were revealed.

Suppose you pick door number 1 at the beginning, and you decide, I'm just going to stay no matter what. Then whatever information they show you doesn't tell you anything about whether door 1 will have the prize. And that's what this is confirming, that that vague feeling that might be true. This shows that indeed, this information being revealed tells you nothing about whether A happened, I guess. The probabilities are the same either way. And so this gives you some intuition for why the Monty Hall problem came out the way it did. Still, you should compute things, but the goal by the end of this class is you'll have at least some intuition about probability, and hopefully it will be the correct one.

Here's another situation to think about. Suppose you have two events, A and B, which are disjoint. So there are two subsets of the sample space, and they don't have any elements in common. So we'd normally write this as A intersect B is the empty set. Then are they independent? Well, let's just compute these probabilities. Well, probability of A I don't know how to compute. But probability of A given B, that I know how to compute. I zoom into B, and I see how many As are in there, and 0. There's no A inside B. In other words, the probability here, probability of A intersect B is 0, because there's no intersection. So whatever the denominator is, is 0.

OK? And so this means they are not independent typically. Unless some exceptional cases happen-- namely, I guess it could be that A is also the empty set. There's nothing in A. Then probability of A equals this thing, probability of A given B. And the other situation is, probability of B is 0, because that's another situation where we just have independence by definition.

OK? So disjoint events are not independent. The intuition here is, if I tell you that B happened, you know very well A did not happen, because we assume these are disjoint events.

OK, another example is subsets. Let's suppose-- which one is easier? B is a subset of A. So A is the big thing, and B is nestled inside.

Then I want to know, what is the probability of A given B? Well, if I zoom into B, It looks in completely A. All of this is included inside A. And so this probability is 1.

And so this is another situation where A and B are not independent. Unless some very special cases happen. The namely, it could be A is everything. A always occurs. Probability of A equals 1. Or again, it could be probability of B equals 0. So if B is just nothing, then they're independent, just by definition. If A is everything, then yeah. You're restricting to B, it's still everything. Question.

**AUDIENCE:** Yeah. Why do we say not independent? Can we say dependent?

**ERIK DEMAINE:** [LAUGHS] Why don't we say dependent? That's a good question. I believe traditionally, not independent. We could just say dependent. I guess, I don't know what dependent means. I guess you could try to define dependence like amount of dependence, and maybe statisticians do that. But in our worlds, we generally just define independent. And then there's the converse, and we call it non-independent. I agree, it's weird. You could maybe think of it as dependent, but I'm pretty sure it's not standard.

All right. Let me tell you some more ways-- of course, you take two events, they're probably dependent, in some sense. Independence is kind of the special thing. And it's what we are going to focus on, is detecting when an event is independent. So let me tell you some other ways to compute it.

So we can use this definition and compute it. But it's oddly asymmetric, right? We talk about the probability of A given B. What if it's easier to compute the probability of B given A, for example, or something else? There's another nice characterization, which I'm going to call the product rule. We've already seen a product rule with inclusion-exclusion. But here's another one.

So an event A is independent of another event B if and only if probability of A intersect B is the product of the probabilities. OK, this is what makes sense, I guess in some sense. Naively, if you want to think about the probability of an intersection-- in other words, the probability of one event happening and another event happening-- naively, that would be the product. And independence is exactly when you can do this. If you're independent, this is true. If you're not independent, this is false. So this is in some sense why we care about independence.

And yeah, let me get another chalk here. And it's effectively what we're using in what I was saying. Oh, the probability of this event happening is the probability of this thing happening times the probability of this thing happening conditioned on this one. That multiplication is OK in particular because event E is independent of A given E, in some sense. I guess this is not properly an event, but you can think of it that way.

All right, let's prove this theorem by cases. So case 1, what I'd like to branch on is, is the probability of B equal to 0? Because it's just annoying to work with this definition. There's this special case of B equals 0. So that turns out to be the natural case breakdown.

Case 2 is going to be probability of B is not 0. OK. So when probability of B equals 0, we know that A is independent of B. That's part of the definition. So this part is true. So for this, if and only if to be true, we better make sure that probability of A intersect B is equal to that product.

But then this number is 0. So this product is 0. So what we really want is that the probability of A intersect B is 0. And indeed, probability of A intersect B is at most, probability of B. Why? Because this set is contained in this set, and probability is monotone. If you take a smaller set, it contains all the elements of this set. Then this probability will be less than or equal to that.

And probability of B is 0 by assumption. And so it's also equal to probability of A times probability of B. I guess I'm also using that probabilities are non-negative. And so I've sandwiched this probability between 0 and 0, which means it equals 0. It's both less than and greater than or equal to 0. So it equals 0, and so we're done with that case.

OK. Probability B not equal to 0 is the more interesting case, because that's when we get to work with this formula. Probability of A given B is equal to probability of A. So that's the definition of independence. So we're at independence if and only if probability of A given B equals probability of A.

OK. So now we have some computation to do. Let's see. We also have a formula for probability of A given B up here. That's the definition. So let's plug that in. Probability of A intersect B divided by probability of B.

OK, so this is equal to that, and this is equal to that. So these two things are equal. And now on this equation, I see probability of A here. I see probability of B on the denominator here. I'd like to multiply both sides by probability of B. And I get, this is equivalent to probability of A times probability of B equals probability of A intersect B, which is exactly what I wanted to prove. So I'm done.

Crucially here, though, I'm using-- in order to multiply both sides by a probability of B and to get an if and only if, it's crucial that this is not 0. If you multiply both sides of a formula by 0, you don't get a very interesting formula, and you don't get an if and only if. But once I assume that probability is not 0, then I can actually do the multiplication. And I get a useful statement, and this is true. So I get the cancellation that I wanted. Otherwise I get 0 over 0.

OK. So this tells us a different way to determine independence. We could compute this thing and this thing, or we can compute this thing and this thing. And if they're equal, it's the same thing as that condition being true.

Notably, this definition-- or this condition. Sometimes this is the definition of independence. You might see both. They're equivalent, so you use either one. This definition is symmetric in A and B. If I swap A and B, I get exactly the same thing, right? A intersect B is the same as B intersect A. Probability of A times probability of B is the same thing as probability of B times probability of A. So I no longer have to say A is independent of B. I can just say A and B are independent without worrying about which one's first.

So corollary, independent of is a symmetric relation. A is independent of B if and only if B is independent of A. So I'll just talk about A and B being independent from now on. This definition is not obviously symmetric, but now that you know that these are interchangeable, if it's easier to compute probability of B given A, you can do that as well.

You can do probability of B given A equals probability of B. That's just as good as checking this one. Not obvious from that definition, but through this transformation (CHUCKLING) it becomes obvious. All right. Let's do some more interesting examples.

And let's start with flipping two independent fair coins. OK, now that you know what independence means, we're going to be using it a lot, in particular just as an assumption. A lot of probability problems are not well-defined, unless you specify as part of the statement that two random events are independent from each other. You already saw this in your problem set. We wanted all of the errors that the bots made to be independent from each other, meaning whether one thing happens does not influence the outcome of the other.

So a simple example of this, where this is a reasonable assumption is you have two coins. I don't use physical money anymore, (CHUCKLING) but imagine you have two coins and you flip them. Whether this one comes out heads and whether this one comes out heads should be independent of each other. Unless you, I don't know, glue the two coins together, then they're very dependent. I used dependent. All right.

So usually when we're saying we're going to flip a bunch of coins, we need to specify, are they somehow linked together? Maybe, I don't know, they have some quantum entanglement, and whether one comes out heads influences the other, or they're physically attached, or something like that? But normally, we think of the case where the coins are all independent. OK, so we're going to use independence also as an assumption in the problem statement. And that specifies in some sense what your sample space should be. Then there's another assumption here, which is fair. That means that they come out 50% heads, 50% tails.

So the sample space is heads and tails squared. So heads and tails cross heads and tails, the set of pairs, heads tails, heads tails, or xy where x and y are in H or T. I forget whether we've squared sets before, but that's the meaning. Just like the cross product of two of them.

And so there's the first coin and the second coin. This is an ordered pair. And they're each equally likely. The assumption that they're independent tells us they should be, all of these outcomes are equally likely, or the assumption that they're independent and fair makes them all equally likely, I guess.

All right, so let's do this more precisely. Let's say A is the event that the first coin comes out heads. I'll say it is. And let's let B be the event that the second coin comes out heads.

OK. I would like to claim that these two events are independent from each other, not just by assumption, but by computing according to either of these formulas. So let's do that so we fully understand the situation. I'm going to draw a tree here. So this is the first coin, and this is the second coin. Each one can be heads or tails.

OK, and because we're assuming fair, this should be probability of 1/2, this should be probability of 1/2. Also here, the probability of each of these coin flips being heads or tails given the previous one, it doesn't matter that the first one happened. So we just write a 1/2, 1/2 on all of these edges.

OK? So now I want to say, let's say probability of B given A. This is going to be a little bit circular, because we're using independence both as an assumption and the thing we're computing. But we should be used to drawing these kinds of diagrams under this implicit independence assumption by now. We are told the number we write here is the probability of this thing happening given that this one happened. So probability of B given A, so A is this edge. But I guess this node is event A, in some sense. A is that the first coin flip was a heads.

So probability of B given A is exactly the probability written on this edge, because this would be B given A, that event. A happened, and then B happened. Or I guess maybe the edge is more, but is thought of this way, this is more like A intersect B that they both happened. But however you slice it, the probability of B given A is 1/2. That's what we defined. The probability of the second coin being a heads given that the first one is also a heads shouldn't matter is 1/2. And on the other hand, the probability of second coin being heads by itself is 1/2. So yes, it's independent. Amazing.

OK, not a very exciting example. Slightly more exciting is that let's say first coin, heads probability is p, and second coin heads probability is q. So I invest in some weighted coins. It's the probability of p that the first one comes up heads, probability of q the second one comes up heads. Maybe p and q are equal to 1/2, or maybe there's something else.

Then still, I get, for example, I don't know the probability that A and B happens is going to be p times q. [CHUCKLES] It's hard to say this without assumption that these are independent events, but this is how we've been computing things. And that's also a probability of A product with the probability of B. OK. Not very exciting.

Let's make this interesting and define some events that are not just a function of one coin, but in fact depend on the entire sample space. So I'm going to go back to this example and define S to be the event that-- I guess I should write A and B are independent here, and A and B are independent here.

OK, but now I'm going to write an event that's not so obvious. The S event is going to be that the coin flips come out the same, either both heads or both tails. This clearly involves both coin flips, so you might think that S and A, say, are dependent on each other. Intuitively they are. But according to this definition, they are independent. That's the claim. So let's compute it.

I claim the easy one is to compute probability of S given A. So in this diagram, A is that we're here, that we flipped ahead so far. And now I want to know the probability that the coins come out the same. Well, conditioned on A being a head, so that's the same as the probability of the second coin being heads. So this is just the probability of B, which is not quite what I wanted. [LAUGHS] We wanted it to be probability of S, I guess. But it is equal to the probability of B, which is 1/2, by assumption.

On the other hand, we want to know, is the probability of S the same thing? Well, the probability of S corresponds to this node and this node, right? This is the heads heads case. This is the tails tails case. Well, we know the probability of this is a quarter, because it's 1/2 times 1/2. And the probability of this is also 1/4, 1/2 times 1/2. We know all of these are equally likely at 1/4. And so probability of S is the sum of those two. This is 1/4 plus 1/4, also known as 1/2. And so these two things are equal.

So you can think of this as a betting game. Suppose I ask you to bet on whether two coins are going to come out the same or different, and they're fair, independent coins. So I'm to flip them and you get $1 if they're the same and you lose $1 if they're different, or something like that.

If I flip the first coin and it comes out heads, that doesn't tell you anything. [CHUCKLES] You're still going to have 1/2 probability of winning when the second flip happens. Doing the first flip does not tell you anything about whether the second flip is going to be the same as the first one, because it's still equally likely to be equal or not equal. That's what this is formalizing. But until you compute it, I would say this is not exactly obvious.

All right. Let's do the same thing. So here A and S are independent. Let's try the same thing with unfair coins. Suppose we have probabilities p and q, and we have S is the event that the coins come out the same. Now things are not so easy.

So what's the probability that they come out the same? Maybe I should draw this tree again, for the unfair case. So we've got heads, tails for the first coin. So this is probability p. This is probability 1 minus p. And then we've got heads and tails for the second coin, which we're assuming now is probability q and 1 minus q, independent of what happened before.

OK. And still, S is this plus this. And so we just take these products and add them up. So the probability of S, the probability of heads heads. Maybe I'll write it out. Probability of heads heads plus the probability of tails tails.

The first one is p times q. So the probability of getting the first heads is p. The probability of getting the second heads is q. And by assumption, these are independent, so I can take their product. And the other one is 1 minus p times 1 minus q. And now I just get some ugly algebra. This is like 2pq minus p minus q plus 1, I think. Check. Yes.

OK. On the other hand, let's try probability of S given A. This one is actually easy, because this is saying, OK, we conditioned-- here's A-- we condition on there being a heads in the first step. What's the probability that the same? Well, that's the same as the probability that the second coin comes up heads as well. And so that's exactly q.

So these things are going to be independent. So I guess S and A are independent, if and only if 2pq minus p minus q plus 1 equals q. In other words, 2pq minus p minus 2q plus 1 equals 0. And this happens to factor. I don't claim this as obvious. But it is the same as 2q minus 1, p minus 1 equals 0.

And so there's two ways for this to be true. One is that p equals 1. In other words, the first coin always comes up heads. I would call that a one-sided coin, or you can make them by manufacturing a coin that's heads on both sides. That's a two-sided coin, but really only one side. Or q is 1/2.

If the second coin is a fair coin, then S and A are independent of each other. And so this tells us that what we found here, which was that S and A were independent of each other, really means that the coins are fair. Fair coins and one-sided coins are the only situation where this S event here that the coins come out the same is independent of the first or second coin.

So if we were playing this gambling game with unfair coins and you knew p and q, and I told you that the first one came out heads, that actually tells you something about whether you're more or less likely for the next coin to match it, because then matching is no longer symmetric. All right. So lots of examples of coin flipping. Maybe it gives you some intuition for some examples that are independent and some that are dependent.

Why do we care about unfair coins? Well, one fun fact is that normal coins are not fair. They're more likely to come up the same orientation that you started in than they are to flip over. And there's a nice physics model by Persi Diaconis and others from back in 2007 that computes that for normal coin flips, you should get about 51% chance of staying on the same side. So if you play this game with somebody and you make money, please thank us. [CHUCKLES]

There's this thing called precession coins. When you flip them vigorously, (CHUCKLING) as they say, they tend to also spin in this way, which does not cause them to flip over. And so they spend more time, a little more time face original versus face upside down. And that gives you a slightly higher chance. On the right is a machine that always flips the coin the same number of flips. And so it always produces the same result. That's cheating.

This theoretical model was recently verified experimentally for 350,000 real coin flips by these 50 authors. This was like, an online internet collaboration, just completed late last year. And they compute maybe not 51% but 50.8% is the thing that they get, or somewhere between 50.6% and 50.9%.

So it confirms this intuition, although it does depend somewhat which currency you use and who you are. Different flippers have a different style in flipping, and so your personal probability may be different from this one. So keep that in mind. There are a zillion videos online of these like, 12-hour marathons, just for verifiable science, where you can watch people flipping coins and typing in the result. And flip a coin, type in the results. And anyway, that is-- [CHUCKLES]

[LAUGHTER]

--you can watch this for the next 12 hours, or we could go back to lecture. All right.

So some more concepts. Independence between two events is kind of cool, but what if we have more than two events? We do this all the time in this class. We talk about, I don't know, whether two vertices in a graph are connected. And then we want to know, well, what if we want the whole graph to be connected? What if we want all the vertices in the graph to be connected? For that, it was enough to do things pairwise. We said, well, if every pair of vertices there's a path, then that's a connected graph. That turned out to be the right answer for graph connectivity. With probability, it's much more annoying.

So one thing you might define is what we call pairwise independence. So I guess, let's suppose I have events E1 through En. Then I'll call them pairwise independent if for all i and j between 1 and n, with i not equal to j, Ei and Ej are independent.

Of course, Ei and Ei are not independent. That's one of the overlapping cases. So Ei is not going to be independent from itself, but it should be independent from all the others. OK, this would be the natural pairwise definition.

On the other hand, we have a stronger notion called mutual independence. This is more annoying to verify, but it's probably what you actually want. And it says, not only if I take any two events that are independent from each other, but if I take any subset of events, it's independent from any one event.

You could also compare subsets to subsets. That doesn't turn out to matter. So what matters is, if you fix any event Ei, so i is between 1 and n, and I take another set j, it's a subset of the elements from 1 to n, except it should not have i, because we don't want to compare Ei to itself, then we want Ei to be independent from the intersection of the Ejs.

So I forget whether you've seen this notation. You've seen big sigma for sums. You've seen big pi for products. This is big intersection for intersection. I guess we ran out of clever Greek letters. So this is just the intersection of Ej for all little j and big J. So we want each event Ei to be independent from the intersection of any subset of the other events.

This is a pain to verify for lots of events. Let's just think about it for three events for a little bit. So there's on the one hand, we have pairwise independence where you just compare E1 versus E2, E2 versus E3, and E1 versus E3. And here you do that, but then you also compare like, E1 versus the intersection of E2 and E3. And you compare E2 versus the intersection of E1 and E3. And you compare E3 with the intersection of E1 and E2. So twice as much work. You check this independence definition and you get your answer.

Let us do an example. Maybe first, some motivation. Let's do some forensics. Suppose you are at a murder scene and you find 9 pieces of evidence. That's a lot of evidence. (CHUCKLING) It's really not realistic, but for sake of example, say you find 9 markers that are consistent with a particular suspect being at the scene of the crime. DNA, whatever. And let's say Mi is the event that a randomly chosen human matches marker i.

Oh. Yeah, I should tell you another version of mutual independence, which was useful. Equivalently, this is the product rule form. For any j that's a subset of 1 up to n, the probability of the intersection of Ej for j and big J is equal to the product of the probabilities of the individual events for j and big J.

OK, this is the natural generalization of What we had for the product rule for independence between two events.

OK, so back to forensics. Suppose we have a bunch of these events, and for each one we compute, and we know their probability because we have taken lots of surveys or whatever. Let's say the probability of each Mi is equal to 10%, 10% chance that you match this particular marker for each marker, just for simplicity. So let's say it's the same for all of them. In reality, it would be different.

So now what we care about. Supposing that-- so this is just assuming sampling a random human, maybe in the area, I don't know. Let's say all humans because we're fair or something. We have no prior assumption. Then what we care about is the probability that all of these come out as matching our particular human that we have in mind. So we have a suspect, and now we want to take the intersection of all those events.

And suppose that human matches all of those markers. Now we want to know, what's the probability that that happened for a random human? And you would like to say, this is 1 over 10 to the 9, 1 in a billion chance. So there's only, like, six people who could match. But that's only true if these events are mutually independent. It's actually true if and only if the events are mutually independent. This is a definition of mutual independence.

Well, OK. So mutual independence is a little stronger. It says not only the intersection of all 9 events is the product. Probability is the product of the probabilities, but also for any subset of those 9 events. So that's crucial. We don't need that for this example. This example, we only care about the intersection of all 9. But if you want to prove that something is mutually independent, you need to check all of the events intersected together and also different subsets of them intersected together. So keep that in mind on problem sets.

Let's do a more interesting example where this matters. OK, back to coin flipping, the source of all examples today. Let's say we flipped 3 mutually independent coins. Fair coins. So we're going to state independence as an assumption.

This is now saying that the coins are not related in any way. They're each flipped completely separately from each other, no impact from one to the other, according to that definition, say. But now I'm going to define some events that are interesting. So A is going to be the event that the first and second coins come out the same when you flip them. Event B is going to be the event that the second and third coins come out the same. And event C is going to be the event that the third and first coins come out the same.

Are these events independent? I don't think it's obvious. So we compute.

All right. Let's start with pairwise independence. Let's think about, say, the probability of A intersect B versus the product of the probabilities of A and B. If you restrict to just A, for example-- let's start with probability of A-- probability of A only depends on the first and second coins. So this is exactly an example we saw before, where we've just flipped two independent coins. And we had this S was the event that the coins are the same. So that's the probability of S. That's going to be 1/2.

So probability of A over here-- sorry for the relabeling. This is just like probability of S before, and so it's 1/2. OK, probability of each of these events is individually 1/2. So what we would like is for the intersection of two of them to be probability 1/4, 1/2 times 1/2. So let's see if it is.

All right. Now we have to actually look at the problem. So A is the event that the first and second coins are the same. B is the event that the second and third coins are the same. So this is like, coin 1 equals coin 2. And this is like, coin 2 equals coin 3. So if both of those things happen, if C1 equals C2 and C2 equals C3, the intersection here is C1 equals C2, equals C3.

So in other words, this is the probability of heads heads heads. I guess I'll write it as a set, probably. Too many brackets. I'm going to have all three types of brackets. This is exciting. Probability gets square brackets, then set gets curly braces, then an ordered pair gets round parentheses. I'm happy they're all distinct. It makes it so much clearer. [CHUCKLES]

All right. So this braced quantity in the middle here is an event. Remember events are just sets over the set-- subsets of the sample space. So it's a set of possible samples. So A intersect B is the same thing as all the coins being the same. That means head heads heads, or tails tails tails.

So that probability-- we could rewrite this, of course, as the probability of heads heads heads. I'm going to omit some parens here just for simplicity, plus the probability of tails tails tails. Because these are different samples, they are disjoint from each other, and so we can use the sum rule. And the probability of each of those events is 1/8. It's 1/2 to the power 3. It's 1/2 times 1/2 times 1/2, because each of them is independent. So this is 1/8 plus 1/8, which is 1/4.

And that's not what we wanted. That's funny. Did I make a mistake? I did. I did. Oh, no. That's what I wanted. Right? Sorry. This probability is A. What I wanted was this to equal not probability of A, but probability of A times probability of B. These were each 1/2. And so their product is 1/4, and so that matches. So these are pairwise independent.

OK, I only computed it for A intersect B. But you should also in principle do it for B intersect C, and A intersect C, all the pairs. But by symmetry, if you think about this the right way, those will all come out the same. OK, I'll just wave my hands at that.

But what about mutual independence? So these three events are pairwise independent, but they are not mutually independent. Let's prove that, or let's compute that.

For that, we could do the probability of the intersection. I claim it's a little easier to think about this probability, the probability of C given A intersect B, because we just thought a lot about what a intersect B means. It turns out to be equivalent to saying that the 3 coins are the same. So now, what is the probability of C given that the 3 coins are the same? Proba-- C says the third coin and the first coin are the same.

So what's this probability?

**AUDIENCE:** Is it 1?

**ERIK DEMAINE:** 1, yeah. Remember, this notation means, compute the probability of this event, zooming into the special case where A intersect B is true. So the sample space has 8 possible spots. It's like a cube. I guess each axis is heads or tails. So there's also a back face here. And what we're saying here is, A intersect B is just the two extremes-- heads heads heads, tails tails tails. Maybe you're used to the cube and this is familiar, that they're opposite corners like this. Maybe not.

And so if we zoom into just those two places, and we want to evaluate, well, what are the chances that coin 2 equals coin 3? Or actually, what was it? First and third coins. (CHUCKLING) Well, it doesn't matter, right? What are the chances that the first coin and the third coin are the same in these two scenarios? Well, 100%, it's guaranteed. If you already know that all three coins are the same, which is what we figured out A intersect B means, then of course, the first and third coins are the same.

Whereas the probability of C without any condition is 1/2. Just like probability of A here. That was the S that we computed before. So these are different from each other. So not mutually independent. So this is kind of an interesting example where you have pairwise independence, but not full independence, not triple-wise independence. And there's only three events. So that's the most you could hope for.

In computer science, these things actually matter. When you do algorithms and like, hashing in particular, it matters how much independence you have. Some hash functions are pairwise independent, or maybe 3-wise independent or 5-wise independent. All these things are important concepts in hashing, for example. But we won't be talking about that here. That's the next class, 6.121.

All right. Finally, we have-- well, one or two more interesting examples. Big one is the birthday paradox. Or you might call it the birthday principle. But I like living in a contradiction, so we'll keep it as a paradox. It's not a paradox. It's just a surprising fun fact, something where intuition does not match mathematics.

So the paradox is that if you sample a bunch of people and measure their birthdays, you're a lot more likely than you might expect to have two people with the same birthday. OK? Let's call them birthday twins. If you're actually twins, then you have the same birthday almost certainly, unless you span midnight. [CHUCKLES]

Are there any twins in the audience? Like, both of you are here? OK, I didn't think so. So then I'm going to assume that your birthdays are all independent from each other. I think twins are pretty much the only way to get dependents. And I'm going to assume that you're equally likely to be on all days, and that's less true in reality. But the effect of the birthday paradox will only be stronger if you have a non-uniform distribution like we do in reality.

So given what you already know in this class, in particular, the pigeonhole principle or the Pidgey hole principle, if you're a Pokemon fan, how many people do I need to take in order to guarantee that I have a duplicate birthday? Some birthday twins.

Remember all the way back to the pigeonhole principle? Yeah.

**AUDIENCE:** 366.

**ERIK DEMAINE:** 366 for a 365-day year. So d here, for example, could be 365. 365? 325? How many days are in a year? [CHUCKLES]

[LAUGHTER]

All right. So in general, if n equals d plus 1, then we get guaranteed duplicate. And of course, if it's bigger also. And the surprising thing is that probabilistically-- so this is for worst case behavior. I mean, if I had n equals d, could be each of you has their own birthday, each person in their own birthday pigeonhole. But that's very unlikely, it turns out. And we're about to compute how likely it is.

So maybe we can do an experiment, which is, I'm going to see whether any two people have the same birthday, if you're willing to raise your hand. So please pay attention, because I want to go through days really fast. So let's start with August, the most common month. So OK, among the August people, how many on August 1? August 2? August 3? August 4? August 5? August 6? August 7? August 8? August 9? August 10? 10? Just one? August 11? One? August 12?

**AUDIENCE:** Two of us.

**ERIK DEMAINE:** Oh, two? All right. August 11, two August 11's. Yay, we did it!

[LAUGHTER]

Didn't even have to get through August. [CHUCKLES] Because we have a quite a few people here. I don't know, maybe 100. So turns out, for d equals 365, let's assume no leap years.

So we had, like, 100 people. I'm going to warm you up with 23 people. 23 people is where the probability is just barely greater than 50% that there's two people. So big difference from 365. n equals 30, the probability goes up to 70%. n equals 60, the probability goes up to 97%. Oh, 99%. 99.4%. OK, very high. n equals 100.

Any guesses how many 9s? 5? 6? Six 9s. The Holy Grail of six 9s. 99.99997%. That's a lot of 9s. OK, so the probability grows exponentially. But in particular, this break-even point is surprisingly low, 23 versus 365.

It's a little bigger than the square root. It's like, 1.2 times the square root of 365 is about 23. So that's the surprise, the paradox. Let's compute it.

It's actually similar to something we computed before in a combinatorics setting, which was the how many dollar bills are there where none of the digits in the serial number were duplicate? So we computed that way back when. I'm going to do a similar thing here. Similar trick. Let me first set up the sample space.

This is like coin flipping, but more like dice rolling, I suppose. You have n people. Each of them rolls a d-sided die. So that means the sample space is just the set of tuples. Do I want to call it d? No let's call it b for birthday, b1 up to bn, where each bi is between 1 and d, let's say. Let's number our days 1 to d. And this is a curly brace.

The events that we care about is getting duplicate birthdays. So I'm going to call this S. So what we are interested in is the set of tuples of birthdays where there are some two indices, i and j where bi equals bj. That's the duplicate birthday. That's the birthday twins.

This happens with high probability, and is very annoying to compute the probability of. But there's this one trick which you may recall from when we were looking at serial numbers on dollar bills. Let's think about the complement event. So if we look at E bar, this is the set of birthday tuples where there are no duplicates. And that's easier to figure out. So we want to say for all i not equal to j, bi does not equal bj.

OK, this is easier to think about as a process. If we think of this going back to generalized product rule, we want to know how many different ways are there to do this where the first birthday that I choose, it could be anyone. I don't care. There's d different possibilities for the first birthday that I choose. Then the next birthday I choose, there's d minus 1 options, if I want to avoid duplication.

OK, then the next birthday, there's d minus 2 options. Hopefully this is old hat by now and you're used to generalized product rule. And the hard part is knowing when to stop. This is going to be d minus n plus 1, I think. That should be the right-- sure, yeah. This is generalized product rule. This is the number of ways.

And we're assuming that each tuple of birthdays is equally likely. This was our assumption, that everything is independent and uniform. So what we need to do is take the number of ways this happens and divide it by the entire size of the sample space. Divide it by the size of S. OK, so I'm going to divide this by-- and this is going to give me the probability of E bar here.

The probability of getting no duplicate birthdays is the number of ways to get no duplicate birthdays divided by the number of different ways to get birthdays, period. This is the non generalized product rule. What is the size of S here?

**AUDIENCE:** d to the n.

**ERIK DEMAINE:** Yeah. d to the n. Thank you. It's either n to the d or d to the n. You got to think for a while. It's like yeah, d to the n. There's d ways for this, d ways for this, d ways. So it's just the product of n d's, the product of d times d times d n times. So I'm going to write that as d times d times d times d. There's exactly n terms upstairs, and there's also n terms downstairs. OK?

So now I have this crazy ratio between two products. I just need to compute it. OK, this is a bit annoying. But one big insight-- so this thing is equal to, instead of taking this product and this product separately and dividing them, I could instead take this ratio and multiply it by this ratio and multiply it by this ratio, and so on. That's why I wrote out this d to the n as a bunch of d's, so I could take them element by element.

So this is going to be equal to the product from i equals 1 to d, d of n, sorry. There's n different terms. Each of them is of the form d minus i over d. And I have an index error, so I need to start at 0 and end at n minus 1. OK, that will give me the right thing on top. I start with d minus 0. I end with d minus n minus 1.

OK, so this ugly formula actually gets pretty clean. Now is when it's hard to evaluate. But there's this great fact. So first I'm going to rewrite this as a product minus 1. This d over d is 1. And then we have an i over d in the second half. OK, this turns out to be nice, because 1 minus x is less than e to the minus x. e here is Euler's constant. 2, roughly, a little bigger than 2.

And if you look at the Taylor expansion of e to the minus x, which is something you may know from calculus, it starts with 1 minus x. And then it's like, plus x squared minus x cubed, whatever. Maybe some denominators there too. But 1 minus x is actually well-approximated by e to the minus x, and so it's approximately equal for x small. And if you think about what x is here, it's the probability. It's less than 1, so that's pretty small. And so this will actually be a pretty good approximation. We're going to replace 1 minus x with e to the minus x. And then this product will be easy to evaluate, easier.

So this is going to be e to the minus i over d. OK? Sorry, not equal. This is going to be less than. So it's a good approximation when x is small, but it's always less than, because the next term is positive, at least for x non-negative. It's going to be less than. And so if we get an upper bound-- remember, we're evaluating here the probability of E bar. E bar is like our error case, the bad case, we want to show is unlikely. So if we can upper bound the probability of the bad thing happening, that gives us a lower bound on the good thing happening, good thing being birthday twins.

OK, we're almost done with birthday paradox. So the cool thing here is we have a product of exponentials. You may recall that if you have, I don't know, e to the x times e to the y, this is equal to e to the x plus y, right? This is taking the product of two exponentials, is the exponential of the sum.

And more generally, if you take the product of n exponentials, that will be the same as the exponential of the sum of those n things. So this is equal to e to the sum. I didn't leave myself enough room. e to the sum, i equals 0 to n minus 1 of minus i over d, I guess. Maybe I should put the minus out here. It's a little cleaner.

OK. And I can parenthesize like this over d, because this d does not depend on i. So the sum of i over d is the same as the sum of i all over d. I've still got this minus out front. And this is our good friend arithmetic sum, the triangular numbers. It's indexed a little funny. Usually we think 1 up to n, which is n times n plus 1 over 2. This is 0 up to n minus 1.

0, of course, doesn't matter. So it's really 1 up to n minus 1. So this is going to be e to the minus n times n minus 1 over 2. And then the whole thing is over d. So we end up with a d in the denominator.

OK. So this was the probability that there are no duplicate birthdays. And you can see, it goes down exponentially. This is remember, e to the minus x is 1 over e to the x. So the denominator here is increasing exponentially as we increase n. And the big thing is-- and so that's why you get so many 9s very quickly out here. But if we're interested in this crossover point where we get 50/50 chance that there's a duplicate birthday versus not, which is interesting, the big thing that matters is that we have an n squared up here.

There's a minus n, but the big part is n squared. And if we invert it, this is where we're going to get a square root behavior. So let's suppose that this probability is equal to 1/2, or roughly equal to 1/2. When does that happen? We did some approximation here with this less than. So this isn't going to be exactly right, but it's close. We want this exponential to equal 2. So we want e to the n, n minus 1 over 2d. Sorry, move the e down here. This is all in the exponent.

We want this to be equal to 2. That's when 1 over it will equal 1/2. So we take logs of both sides. That's like, n times n minus 1 over 2d equals ln 2, natural log. So we maybe move this over. We get n times n minus 1 equals 2d ln 2. And at this point, I'm going to wave my hands a little bit and say, this minus 1 doesn't matter too much. And so basically this is a square root. So it's going to be n is roughly the square root of 2d ln 2. And so there's this constant term, which is square root of 2 ln 2. This is about 1.177.

So this is about 1.2 to the square root of d. This is what I claimed before. If you take 1.2 times the square root of 365, you get approximately 23. And if you compute it exactly, you get 23. So kind of neat. In general, if we have d days, you need about square root of d people to get a duplicate. This has all sorts of implications in computer science, usually in the context of hash tables. If you're randomly placing people into locations-- pigeonholes, I suppose-- it's once you have square root of n or square root of d different people, you're likely to get two that happen to cohabitate if you choose them uniformly.

So for example, if you're a cryptographer and you're signing your emails, you want to guarantee that you don't have two emails that have the same signature, because if you did, someone who's receiving them and checking the signatures can't distinguish between the two emails. So that tells you, like the set of possible signatures on your emails to be at least the square of the number of possible emails you could write. So this gives you a lower bound on how big cryptographic signatures need to be in order to be safe, just in terms of whether there are duplications.

Or you're Dropbox and you're trying to sync files from one computer to another, and you want to compute some hash on the files on both sides to see, do I need to retransfer this file? Did it change on either side? It better be that the space of possible hashes is at least the square of the number of possible files. Otherwise you're going to get accidental hash matches. All right.

A little more time. So let me tell you about one independent concept not particularly related to independence, but related to conditional probability, which is gambler's fallacy. Can't spell.

Gambler's fallacy can be phrased many ways, but let's talk about coins, because that's what we're doing. Let's say you do 10 coin flips, and they all come out heads. And I want to know, what's the probability that the next coin flip is also heads?

What's the answer? Sorry, fair coin, I should say. I'll give you a fair coin. Each flip is independent from the other flips. What are the chances that the 11th flip is heads?

**AUDIENCE:** 1/2?

**ERIK DEMAINE:** 1/2. I should hope so. [CHUCKLES] If you're a probabilist, or if you're a student in 6.1200, the answer is 1/2. If you're a gambler, probably an addicted gambler-- I mean, this is a statement about human psychology. Who knows? They say less than 1/2, because I'm due for a tails, don't you think? [CHUCKLES]

A hypothetical gambler imagines that the coin has memory. It remembers all those heads it got. And when you flip it the next time, now, finally, it's going to be tails. But no. Probability is actually 1/2. This is something that's actually false. Gambler believes, oh, finally I'm due for tails. It's very unlikely that the next coin is tails. I haven't seen an earthquake in the last hundred years. Very likely I'll see one this year.

I mean, that may be true, but it also may not be true. It depends on your prior probabilities about earthquake models. And if I tell you it's a fair coin, then you know it's not true. It has to be exactly half, every single time. It's actually pretty likely to get a bunch of heads in a row. Now, it's not that likely to get 10 coin flips in a row. That's 1 in 1,000, roughly. So a probabilist just says, wow, 1 in 1,000. Cool. Well, probably the next one is 1/2. OK?

There's a third type of person, which I'll call the Bayesian. And they would say, maybe it's bigger than 1/2. Maybe you lied to me and the coin is not fair. That's the other interpretation. And if you flipped a coin 10 flips-- maybe it's not enough, but 11 or 12-- then you start thinking, hmm. I think there's something weird about this coin. Then you want to inspect the coin and see if it's heads on both sides or whatever.

So you can formalize this a little bit using the law of total probability combined with Bayes' rule in the following way. I'm going to write this a little informally, but the probability of a head conditioned on getting 10 heads previously, you can split it into two parts and say, I'm just going to think of a simple model. Either the coin is really fair, or you're a cheater and the coin is heads on both sides. Those are the only two possibilities for this model.

Of course, you could consider a more general model, and there's a whole world about this. So I'm going to call these two scenarios cheat and fair. And hopefully this equation is relatively obvious. Just, it's a mouthful.

OK, everything here is conditioned on H to the 10. So you can ignore that part, because we're assuming throughout that we have 10 coin flips starting as heads. And now there are two scenarios. It could be you're cheating, or it could be you're fair. And these are disjoint scenarios.

So I can split things up according to law of total probability. It's the probability of cheat times the same thing conditioned on cheating in addition, or probability of fair times the probability of the same thing conditioning on fair, in addition to what I conditioned on before. So everything is conditioned on H to the 10.

And so you can write this out. And we actually know what these are. They look ugly. But the probability of heads with a cheating coin that is heads on both sides, that's 1. So it's 1 times the probability of cheating, of there being a cheat here, given H to the 10 plus this probability, the probability of heads with a fair coin, that we know is 1/2 half times probability of you're using a fair coin, given that I've seen 10 heads.

Now these two quantities, you need more modeling to understand what the chances are that I'm using a cheating coin given that I saw 10 heads, but you can compute it. And so now, you can actually quantify how much bigger than 1/2. In this particular model of cheating, what's the chances that the next coin flip will be heads? So fun example, and that's all for today. Enjoy your Independence Day.