

**PROFESSOR:** So we've been saving for the last the property that makes expectation calculating really easy and short circuits a lot of the ingenious methods that we've used up until now-- namely, expectation is linear.

So what that means is that if you have two random variables  $R$  and  $S$  and two constants  $a$  and  $b$ , the expectation function is linear. That is, you take a linear combination of  $R$  and  $S$ --  $aR$  plus  $bS$ , and that's equal to the corresponding linear combination of the expectations.

I'll read it again. Expectation of  $aR$  plus  $bS$  is equal to  $a$  times the expectation of  $R$  plus  $b$  times the expectation  $S$ .

Expectation is linear. OK.

That's an absolutely fundamental formula that you should be comfortable with and remember. It extends actually not only to any finite number of variables, but with some convergence conditions, it actually extends even to countable sum of variables. But let's just settle for the two random variables case for today.

Now, the crucial thing that makes it so powerful and useful is that this fact has nothing to do with independence. Whether  $R$  and  $S$  are independent or equal, it doesn't matter. This linearity holds.

The proof is not terribly informative. It's just manipulating terms and rearranging terms into sum, but let's go through the exercise. Again, something I would never do in lecture. But in a video where you can skip it or fast forward or replay it, I think it might be worth doing, And it is a proof, by the way, that I think you should be responsible for. So let's go through it.

OK. We're interested in the expectation of the random variable that you get by multiplying the random variable  $A$  by little  $a$  and the random variable  $B$  by little  $b$ . All right.

One of the definitions of expectation is that you get this expectation by taking the sum over all the outcomes of the value of this linear combination at the outcome  $\omega$  times the probability of  $\omega$ . So what's the value of the linear combination  $aA$  plus  $bB$ ? At  $\omega$ , it's simply  $a$  times  $A$  of  $\omega$  plus  $b$  times  $B$  of  $\omega$ .

OK. Now I've got a sum of these terms summing over  $\omega$ . I can split them into two groups.

I can take the sum over the  $aA$ 's at  $\omega$  times probability of  $\omega$  and  $bB$  of  $\omega$  times probability of  $\omega$ .

In other words, I'm multiplying by probability of  $\omega$  here to get a sum of two terms, and then I'm rearranging all of the capital A terms first, followed by all the capital B terms.

The result is that I wind up with the sum over  $\omega$  of the A terms times the probability of  $\omega$ -- and I factored out the little  $a$ -- plus  $b$  times the sum over all  $\omega$  of B of  $\omega$  times the probability of  $\omega$ . It's just rearranging the terms in this sum after I've multiplied through by probability of  $\omega$ .

Well, of course, this is equal by definition to  $a$  times the expectation of A. Notice this is the expectation of A, and that's the expectation of B times  $b$ . And the proof is done. Not inspiring, but routine if you use the alternative definition of expectation in terms of summing over the outcomes. It's a messier proof if you have to use the definition of the expectation, being the value times the probability that the variable takes that value. And you wind up having to convert that formula into this formula in order to carry through the proof nicely. And we're done.

OK. Let's make use of it. And in order to do that, let's make a really trivial but a very important remark about the expectation of an indicator variable. So remember  $I_A$  is three random variables that's equal to 1 if the event A occurs and 0 if the event A doesn't occur.

So what is the expectation of the indicator variable? Well, by definition, it's 1 times the probability that it equals 1 plus 0 times the probability that it equals 0. Those are the only two values it can take.

Well, we can forget this term in 0 times something. But what is the probability that  $I_A$  is equal to 1? That's exactly the probability of A, and that's the fundamental formula that we want to notice. The expectation of the indicator variable for the event A is nothing but the probability that A occurs. File that away. We're about to use it multiple times.

So let's go back to the expected number of heads in  $n$  flips which we've now seen at least two ways to do-- one by generating function argument, another by a recursive argument using total expectation. Now where we're going to knock it off very elegantly using linearity because let  $H_i$  be the indicator variable for having a head on  $i$ th flip.

So we look at the  $i$ th flip.  $H_i$  is 1 if the  $i$ th flip comes up head and  $H_i$  is 0 if the  $i$ th flip does not

come up head. Then we can make the following crucial remark, and this is a trick that we'll use regularly by expressing some quantity that we're interested in as a sum of indicator variables.

The total number of heads-- the random variable equal to the total number of heads in  $n$  flips-- is equal to the sum of the indicator variables for whether there's a head on the first flip plus whether there's a head on the second flip up through whether there's a head on the  $n$  flip.

So suddenly the random variable that I want to compute is a sum of  $n$  random variables, in fact,  $n$  indicator variables. All right.

Well, that tells me that the expectation of the number of heads is the expectation of the sum. After all, it's equal to the sum. But the expectation of the sum is going to be the sum of the expectations by linearity.

So it's simply the expectation of  $H_1$  plus the expectation of  $H_2$  out through the expectation of  $H_n$ .

But what's the expectation of getting a head on  $i$ th flip? Well, the flips are independent. It's simply the expectation of a head. So what I have is-- each of these is equal to the probability of a head, and there's  $n$  of them. So the total is  $n$  times the probability of Head, or  $np$ , which is a formula that we had derived two other ways previously. And now it really falls out very elegantly with hardly any ingenuity other than the wonderful idea of expressing the number of heads as a sum of indicators.

Let's look at one example and a very related example of asking about the probability of the expected number of hats that are returned.

When  $n$  men check their hats at a hat check and then the hats get all scrambled up by incompetent staff and then they're given out again in such a way that the probability that the  $i$ th man gets his own hat back is  $1/n$ .

What you could say is that all possible permutations of the  $n$  hats are equally likely. And we ask with all permutations equally likely, how many of them is it the case that the  $i$ th man gets his own hat back? And there's a  $1/n$  chance that the  $i$ th man is going to get his own hat back because there's  $n$  hats, and he's equally likely to get all of them.

OK.

How many men do we expect will get their hat back in this setting?

Well, let's let our  $R_i$  be the indicator variable for whether or not the  $i$ th man got his hat returned--  $R_i$  for hat returned to the  $i$ th man.

Now, notice that  $R_i$  and  $R_j$  are not independent. In the previous case, those  $H$ 's were independent because the coin flips were independent. But here, if I know, for example, that  $R_1$  got his hat back, the probability that  $R_2$  got his hat back has changed from  $1/n$  to  $1/(n-1)$  because 1 is out of the picture, and  $R_2$  is going to get his hat back among the remaining hats 2 through  $n$  and  $n$  is  $n-1$  of them. And he's got a  $1/(n-1)$  chance of getting his hat.

His probability has changed given that  $R_1$  got his hat back. So they're not independent. All right.

Nevertheless, independence doesn't matter for linearity. So I can still say that the expected number of hats returned is equal to the expectation of the sum of the indicator variables for each man getting his hat back.

And of course, the expectation of that sum is the sum of the expectations. And the expectation of each of these-- we figured out-- was  $1/n$ , and there's  $n$  of them. So it's  $n$  times  $1/n$ , or 1.

I expect when all the hats are scrambled and all permutations of the hats are equally likely, that one man is going to get his hat back. None of the others will.

OK. Now let's change the situation a little bit. And think instead of scrambling the hats in a way that all possible permutations of hats are equally likely, let's think about a Chinese banquet. So Chinese banquets are traditionally done with a table of nine in a circle, and there's a lazy Susan that spins around where there's dishes of food in front of each person.

But let's generalize it to  $n$ . Suppose that  $n$  people are sitting around a spinning table-- a lazy-Susan) with  $n$  different dishes, one dish in front of each person. And now we spin the lazy Susan randomly, and we ask how many people do we expect will wind up with the same dish that they started with after the spin?

Well, now we can let  $R_i$  indicate that the  $i$ th person got the same dish back. And now these  $R_i$ s, which are different from the previous ones about hat returns, these are  $R_i$ s are totally

dependent-- much more so than the other ones work because they're all 1 or they're all 0.

If one person gets their hat back, it means that the spinning table got back to where it used to be, and everybody has their hat back. And if one person doesn't have their hat back, then the table is shifted off where it was originally and nobody has the original dish.

I said hats. I meant the dish that they started with.

So everybody gets the same dish back, or nobody gets the same dish back. These variables are as dependent as they possibly could be, but it doesn't matter because linearity still holds. And that means that the previous argument about the expected number of dishes that get back to the person that they started with is still 1 even though all the  $R_i$ s are equal.

Well, that's so much for the lovely rule about linearity of expectation which holds regardless of assumptions about independence or not. There is a rule for products, but it requires independence.

So the independent product rule says, sure enough that the expectation of a product of two random variables,  $x$  and  $y$ , is the product of their expectations, providing that they are independent. And this extends to many variables if they're mutually independent.

Again, the proof by rearranging terms in the defining sum for the expectation of  $xy$ . Let's go through it. And again, you can fast forward or skip this part if you don't want to watch equations being manipulated.

So by definition, the expectation of the product  $xy$  is the sum over all the possible values of  $x$  and  $y$  of the value of the product  $xy$  times the probability that the first variable, capital  $X$ , equals little  $x$  and the second variable  $Y$  is equal to little  $y$ . This is by definition, the expectation of the product-- by the first definition. Not the one in terms of outcomes.

Now, using independence, this term here can be split into a product of  $X$  equals  $x$  and  $Y$  equals  $y$ . So let's do that. So this is the sum of  $xy$  times the probability that  $X$  equals  $x$  times the probability that  $Y$  equals  $y$ .

Now I'm going to do a fairly standard trick, which is I'm going to organize this sum in a clean way. Right now it's an unordered sum over all possible pairs of  $x$  and  $y$  in the range of the variables  $x$  and  $y$ , but I'm going to do the sum so I first sum over all the  $y$ 's, and then for each

y, I'm going to sum over all the x's.

This is an unordered sum really. There's no order here, but now I'm giving you an arrangement which says that I'm going to lump together the sums over all the x's, and then for each of those I'm going to sum over the y's.

Well, when I do it this way, what I've got is an interesting thing here. I've got a sum over x, and there's some y terms here that don't depend on x. I can factor them out because they don't change with x.

So if I factor out this y and probability of Y equals y, I wind up with the sum over y of this factored out term-- y times probability Y equals y-- times the sum over x's x times the probability that X equals x.

Now, this is the same term that is the coefficient of every one of these y terms that depends on y, and this term does not depend on y. So I can factor it out. And if I do that, I wind up with the sum over x of x times the probability that X equals x times the sum over y of y times the probability that Y equals y.

And guess what. This is by definition the expectation of X, and this is by definition the expectation of Y. And by that chain of equalities, I've proved, sure enough, that the expectation of XY is equal to the expectation of X times the expectation of Y QED.

So the key step here was where independence was used at the very first step when I split up the probability that X equaled x and Y equaled y into the product of the corresponding probabilities.

Now, let me just end with a warning about a couple of blunders that people make all the time. So first of all, don't forget independence as a crucial condition on the product rule for expectations.

It can hold in some cases where the variables are dependent. Independence is not a necessary condition. It's efficient, but you need some kind of a condition in order for the product rule to hold. So if you're not careful, it won't if you forget to check for independence or something that is tantamount to it. So let's just take an easy example remember what happens if independence fails.

Suppose I have a variable X-- a random variable X-- which takes positive and negative values

with equal probability. So it takes 1 and minus 1 with equal probability. It takes pi and minus pi with equal probability. I don't really care what those values are as long as it's taking some positive and negative values, and it takes a positive value at the same probability that it takes that value negated.

Well, that automatically means that the expectation of  $X$  is 0 because when I add up all these terms, the positive and negative terms cancel because they have the same probability. So any such  $X$  that's symmetric about 0 has expectations 0.

On the other hand, if I square  $X$ , then all of those positive and negative values become positive. And so I'm taking the expectation of a variable that's strictly positive-- at least with nonzero probability at a bunch of outcomes. And therefore, the expectation of  $X$  squared is positive.

So the expectation of  $X$  is 0, but the expectation of  $X$  squared is positive. Well, of course, if I multiply expectation of  $X$  times expectation of  $X$ , that's still 0.

So here's a counter example. Expectation of  $X$  times expectation of  $X$  is equal to 0, which is less than the expectation of the square of  $X$ . Of course this is about as dependent as it could possibly be because it's the same random variable, but it illustrates the failure of the product rule if you don't have some kind of a condition like independence around.

There's a second blunder that's more interesting and that people can fall in because there's a temptation to assume that if the product rule holds for independence, then so should the reciprocal rule. That is, you might think that the expectation of  $X$  over  $Y$  is equal to the expectation of  $X$  over the expectation of  $Y$  when  $X$  and  $Y$  are independent. But it's not true.

Even when they're independent, the expectation of  $X$  divided by  $Y$  is in general, not equal to the expectation of  $X$  divided by the expectation of  $Y$ . In fact, the counterexample is if  $X$  is the constant 1, the expectation of 1 over  $Y$

[AUDIO OUT]

**PROFESSOR:** [INAUDIBLE] complex instruction set code was better than risk. So I won't mention names, but prominent people have made this blunder. You shouldn't.