

Let us now revisit the subject of expectations and develop an important linearity property for the case where we're dealing with multiple random variables. We already have a linearity property. If we have a linear function of a single random variable, then expectations behave in a linear fashion.

But now, if we have multiple random variables, we have this additional property. The expected value of the sum of two random variables is equal to the sum of their expectations. Let us go through the derivation of this very important fact because it is a nice exercise in applying the expected value rule and also manipulating PMFs and joint PMFs.

We're dealing with the expected value of a function of two random variables. Which function? If we write it this way, we are dealing with the function g , which is just the sum of its two entries.

So now we can continue with the application of the expected value rule. And we obtain the sum over all possible x, y pairs. Here, we need to write to g of x, y . But in our case, the function we're dealing with is just x plus y . And then we weigh, according to the entries of the joint PMF. So this is just an application of the expected value rule to this particular function.

Now let us take this sum and break it into two pieces-- one involving only the x -term, and another piece involving only the y -term. Now, if we look at this double summation, look at the inner sum.

It's a sum over y 's. While we're adding over y 's, the value of x remains fixed. So x is a constant, as far as the sum is concerned. So x can be pulled outside this summation.

Let us just continue with this term, the first one, and see that a simplification happens. This quantity here is the sum of the probabilities of the different y 's that can go together with a particular x . So it is just equal to the probability of that particular x . It's the marginal PMF.

If we carry out a similar step for the second term, we will obtain the sum over y 's. It's just a symmetrical argument. And at this point we recognize that what we have in front of us is just the expected value of X , this is the first term, plus the expected value of Y . So this completes the derivation of this important linearity property.

Of course, we proved the linearity property for the case of the sum of two random variables. But you

can proceed in a similar way, or maybe use induction, and one can easily establish, by following the same kind of argument, that we have a linearity property when we add any finite number of random variables. The expected value of a sum is the sum of the expected values.

Just for a little bit of practice, if, for example, we're dealing with this expression, the expected value of that expression would be the expected value of $2X$ plus the expected value of $3Y$ minus the expected value of Z . And then, using the linearity property for linear functions of a single random variable, we can pull the constants out of the expectations. And this would be twice the expected value of X plus 3 times the expected value of Y minus the expected value of Z .

What we will do next is to use the linearity property of expectations to solve a problem that would otherwise be quite difficult. We will use the linearity property to find the mean of a binomial random variable.

Let X be a binomial random variable with parameters n and p . And we can interpret X as the number of successes in n independent trials where each one of the trials has a probability p of resulting in a success.

Well, we know the PMF of a binomial. And we can use the definition of expectation to obtain this expression. This is just the PMF of the binomial. And therefore, what we have here is the usual definition of the expected value.

Now, if you look at this sum, it appears quite formidable. And it would be quite hard to evaluate it. Instead, we're going to use a very useful trick. We will employ what we have called indicator variables. So let's define a random variable X_i , which is a one if the i th trial is a success, and zero otherwise.

Now if we want to count successes, what we want to count is how many of the X_i 's are equal to 1. So if we add the X_i 's, this will have a contribution of 1 from each one of the successes. So when you add them up, you obtain the total number of successes.

So we have expressed a random variable as a sum of much simpler random variables. So at this point, we can now use linearity of expectations to write that the expected value of X will be the expected value of X_1 plus all the way to the expected value of X_n .

Now what is the expected value of X_1 ? It is a Bernoulli random variable that takes the value 1 with

probability p and takes the value of 0 with probability 1 minus p . The expected value of this random variable is p . And similarly, for each one of these terms in the summation. And so the final end result is equal to n times p .

This answer, of course, makes also intuitive sense. If we have to p equal to $1/2$, and we toss a coin 100 times, the expected number, or the average number, of heads we expect to see will be $1/2$ half times 100, which is 50. The higher p is, the more successes we expect to see. And of course, if we double n , we expect to see twice as many successes.

So this is an illustration of the power of breaking up problems into simpler pieces that are easier to analyze. And the linearity of expectations is one more tool that we have in our hands for breaking up perhaps complicated random variables into simpler ones and then analyzing them separately.