# MITx Video | 6.041x | Spring 2014 | Video transcript | U09: Sums of a binomial and a Poisson-distributed number of Bernoulli r.v.'s

Hello. We will now discuss a subject that we touched upon earlier in the course, and this is the distribution of the sum of a random number of independent identically distributed random variables. We will look at some special cases, where favorable results can be obtained by using ideas from the Bernoulli and the Poisson processes.

So what we have here is a number of random variables, X1, X2, X3, and so on, which are independent and identically distributed. They can be either discrete or continuous. And we are interested in Y, which is the sum of the first N of them, capital N. Now, capital N is a discrete random variable. In particular, it's going to be an integer that's non-negative. And N is going to be independent of Xi.

Now, there's an issue of notation here. What happens if N is equal to 0. If N is equal to 0, we use the convention that Y is also equal to 0. So Y is well defined as a random variable, given that the X's are random variables and N is also a random variable.

Now, we have encountered random sums of random variables earlier in the course, and we derived the expected value of Y and the variance of Y. We obtained these two formulas. The expected value of Y is obtained in terms of the expected value of N and the expected value of X. And the variance of Y is obtained, again, in terms of the expected value of N and the expected value of X. But the formula also involves the variance of X and the variance of N.

These are important formulas, and they are special cases of more general formulas that are based on the law of iterated expectations and the law of total variance.

Now what we will do in this segment is to look at a more general quantity, not just the expected value of Y and the variance of Y, but also the entire distribution of Y, which is not given by these formulas. In particular, we're going to show that if X here denotes the generic Xi-- so X has the same distribution as all these X's.

Now, suppose that X is a Bernoulli random variable, 1 or 0, with probability, q and 1 minus q. And if N-- the integer number of terms in this sum-- if N has a binomial distribution with probability p, and number of trials m, then Y is also binomial over m trials with probability p times q.

The other case is when N, the discrete random number of terms in the sum, has a Poisson distribution with parameter lambda. Then Y is also a Poisson random variable with parameter lambda q. So this is what we're going to show. And in what follows, we're going to go into the derivation and make sure that the results are correct.

As a representative example, we'll consider a store in which customers arrive according to a Bernoulli process with probability p per slot. We're going to let N be the number of customer arrivals within the first m time slots. So N is a binomial random variable. And we are going to assume that each customer buys-- that enters the store-- buys something with probability, q, and does not buy with a complementary probability, 1 minus q, independently of every other customer. And then what we will do is view Y as the number of sales within a given number of time slots, and we will verify that the distribution is binomial.

So here is the example. We have this store in which customers arrive according to a Bernoulli process with probability, p. And each customer with probability, q, will buy something and with probability, 1 minus q, will not buy. So customers arrive according to a Bernoulli process. That means that at every time slot-- and we're going to consider a fixed number, m, of time slots-- at each time slot, there is probability, p, that a customer will arrive and the complementary probability, 1 minus p, that he will not arrive.

So let's call now N the number of customers that are going to arrive in the first m time slots. And let's define Bernoulli random variables corresponding to the buy process. Let's call Xi to be 1, if the i-th arriving customer buys something, and 0, if this customer does not buy.

Then what we have, if we consider the sum of all these X's from 1 up to N, we're going to have the total number of sales. And we're going to call this Y. So this is X1 plus all the way up to XN. And this is the number of sales in m time slots.

So this Y here is a random variable, and it's the sum of Bernoulli random variables that are independent and identically distributed. And the number of this Bernoulli random variables is random, because N is random. In fact, N has a binomial distribution, because it's the number of arrivals in m slots with probability of arrival equal to p.

Now we'll have another way to calculate Y, which is Y equal to W1 all the way up to Wm. m is the number of time slots. It's different than N. N is the random number of arrivals, m is the number of time slots. And these W variables are going to be Bernoulli random variables, again, with Wi equal to 1, if the i-th time slot is a sale. In other words, at the i-th time, a customer comes and that customer buys also. So this W is a Bernoulli random variable with probability, p, for arriving, q, for buying, a total of pq. So WY is Bernoulli with probability, p times q.

So now we see that the total number of sales within the m time slots is the sum of a fixed number of Bernoulli random variables. Therefore, it is binomial as we show here with over m trials and with probability, pq. So we have shown the result for the first box.

However, it's interesting here to interpret this result in terms of a Bernoulli process and splitting it in two processes. The Bernoulli process is the arrival process. The split processes are the arrivals of sales and the arrivals of non sales.

So let's draw this in the figure. We have Bernoulli arrivals with probability, p, at each trial. And these are recorded on a time axis that starts at the first time slot, the first, second, and so on, all the way up to m. This is the time interval that we consider. And now customers arrive according to this Bernoulli process, and they are split in two, the ones that buy and the ones that do not buy.

So let's record a hypothetical sequence of arrivals here, so an arrival at this time slot. This is discrete time, so these points here are integer points. But I'm not showing that they are integers. So this is the first arrival, and let's say that this customer buys something. We record an arrival in the buying process on the top. Let's say the second one does not buy. So we record an arrival in the bottom process.

Let's say this one buys. This one buys also. So we record arrivals, again, the buying process. This one does not buy. This one buys. So now, this Bernoulli process has been split into two according probabilities, q and 1 minus q. And here, what we have is the process of customer buys or sales. And this is a process of customers that do not buy.

Now, we know from the theory of Bernoulli processes that if we split a Bernoulli process in an independent way, the resulting pieces are also Bernoulli processes with corresponding probabilities, pq and p 1 minus q. And now we reason as follows. What we have here at the top is a Bernoulli process. And we are looking for Y, the number of arrivals, the number of sales, within the first m time periods. By the theory of the Bernoulli process, this Y is a binomial random variable over m trials and with probability, pq. So this is the result, and it's consistent with this algebraic derivation that we had before.

We will now discuss a variation of the problem, where the customer arrival process is Poisson with parameter lambda instead of being Bernoulli. And in this way, we are going to derive this result here.

So what is going to change in this figure is that the customers arrive according to a continuous time process, a Poisson process, with rate parameter, lambda. And again as before, every arriving customer will buy something with probability, q, or will not buy with probability, 1 minus q. N is going to be the number of customer arrivals in one time unit-- remember now, time is continuous, so we can't talk about time slots-- so in one time unit.

And let me correct this figure. We have here a Poisson process with parameter, lambda. And we're looking at arrivals of customers from time, 0, to time, 1. And this process is split into two according to probabilities, q and 1 minus q. So m, being the number of customer arrivals in one time unit, is Poisson distributed with parameter, lambda, according to the standard processes with a Poisson process.

Xi stays the same. So Xi is still Bernoulli with probability, q and 1 minus q. We're looking at Y, the sum of the random number of customer sales within one time unit, the first time unit time from 0 to 1, so in the time interval between 0 and 1. And now unfortunately, the W's do not make sense anymore. The W's, remember, corresponded to some time slot, but there's no notion of time slot in the context of this continuous time process. So the preceding line of reasoning based on the W's does not apply. And I'm going to erase this, because it's not relevant to the Poisson case.

However this figure is still relevant. What we have here is a Poisson process that's being split into two. And the top process is the process of customer sales. And the number of arrivals in this Poisson process is Y within the interval, 0 and 1. And we know from theory of splitting Poisson processes that the portions are also Poisson with the corresponding rates modified by q and 1 minus q. So this process is Poisson with parameter, lambda q. And this one is Poisson with parameter, lambda 1 minus q.

And now we focus on the top process. And Y is the number of arrivals within the interval from 0 to 1, so we know that Y is Poisson with parameter, lambda q. And this is exactly what we wanted to prove, is this result here.

OK. So let's try now to make a few observations and ask a couple of questions. First, given the binomial version of the result, does the Poisson version make sense. Well, the answer to that is that it does make sense, because the Poisson process can be viewed as the limit of the Bernoulli process. So it is natural that in the limit, the binomial number of arrivals becomes Poisson.

Another question. Here we have some general formulas for the expected value, Y, and for the variance of Y. We have obtained-- with specific assumptions for the character of X and N-- we have obtained the distribution of Y. So if we would take this distribution and calculate its mean and variance, we should match these formulas, these more general formulas. We should obtain special cases of these formulas. So let's do the calculations and make sure that our results are consistent.

So we will work for the Bernoulli case, where N is a binomial random variable over m trials with probability, p. So the expected value of N is m times p. X is a Bernoulli random variable with probability of 1 equal to q. So this expected value is equal to q. So altogether, we obtain that the expected value, Y, is m times pq. But this matches this formula here, because the mean of the binomial random variable is the number of trials times the probability of success at each trial and matches this. So this result is consistent with what we have obtained. And now let's turn to the variance.

Again, the expected value of N is mp, as before. The variance of X is the variance of the Bernoulli random variable with probability, q. So it is q times 1 minus q. The expected value of the Bernoulli is q, so this gives us q squared. And the variance of N is the variance of the binomial random variable. It is m times p 1 minus p.

OK. So now let's do a calculation. We'll take out the common factor, mpq. And what's left inside is 1 minus q plus q times p 1 minus p. And if you calculate this out, collect the terms, and eliminate as necessary, you get mpq times 1 minus pq. Now this formula here is precisely the formula for the variance of the binomial over m trials and with probability, pq.

So we have verified indeed that what we obtained from these results matches and is consistent with the general formulas for the mean and the variance. We checked our result for the case of the binomial, but the same calculation applies for the case of the Poisson.