Week 5

10 June 2021

23 Multinomial Distribution

What if we have many different events happening all at once? We can easily take our idea of joint distributions and expand it to multiple random variables.

Through this we can talk about a generalization of the binomial distribution. Remember that the binomial distribution had two potential outcomes "yes" and "no" with probability p and $(1-p)$. What happens if we allow multiple possible outcomes with their own probabilities?

Recall that for the binomial distribution what we did was we made a tree where at every node, the tree split into two parts with probability p and $(1-p)$ at each part. We knew the coefficient at any given node because it was given by the binomial coefficient $\binom{n}{k}$. So if we did n trials and wanted to see if k came back, then we looked at

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Wikipedia: Pascal's pyramid

Binomial distribution is the multinomial distribution with $\ell = 2$.

To construct the multinomial distribution, we construct the same tree as before, but at every node we split the tree into ℓ different parts. If $\ell = 2$ this is called the binomial distribution and we saw that this tree is just Pascal's triangle. If $\ell = 3$ this is called the *trinomial distribution* and this tree is given by Pascal's pyramid. For an arbitrary tree, we would have ℓ edges coming out of each node and the probability on each edge would be labelled $p_1, p_2, \ldots, p_\ell$ where ________________. If we then perform n total trials and wanted to see if k_1 of them were for the first option, k_2 for the second, etc. then we would get:

but what is our constant here? It turns out, it's something called the "multinomial coefficient" which come from the multinomial theorem:

Wikipedia: Multinomial theorem

which is basically what we have above! So if we plug in p_i for each x_i we get

Example 23.1 Say that we're rolling a dice and we keep track of three different variables:

(1) Even numbers

(2) The number one

(3) Everything else (aka the numbers three and five)

If we roll the dice 10 times what's the probability that we get 6 even numbers and the number one 3 times?

24 Symmetry

We now look at the histograms of a given distribution and we ask when is our probability distribution symmetric. You might remember the word "symmetric" through calculus where you discussed that an even function is symmetric about the y-axis (aka you can reflect it over the y-axis and get the same function) and that an odd function is symmetric about the origin (aka you can rotate it by half a circle around the origin and get the same function). For us, we'll say that a distribution of X is *symmetric about v* if

Wikipedia: Symmetric distribution

When $v = 0$ we get symmetry about the y-axis. Equivalently, we can look at things from an inequality perspective:

Now, say that we take the sum of n random variables:

$$S_n = X_1 + X_2 + \ldots + X_n$$

where each X_i is symmetric about 0, *i.e.*, ____________________
Then we know that X and $-X$ have the same distribution since:

If we're doing independent trials, we know that this implies that S_n and $-S_n = \sum -X_i$ have the same distribution. The way I think about this is: each X_i has the same distribution as $-X_i$; and since each random variable is not dependent on the others, the distribution of S_n (aka looking down the diagonal in our tables) is the same as $-S_n$ (looking down the diagonal, but with -1). In other words ____________________
Therefore S_n is symmetric about 0.

We can do a similar thing above if the (independent) X_i are symmetric about v_i. In this case, we must do a change of variable. We let $Y_i = X_i - v_i$ and in this case each Y_i is symmetric about 0. Letting $S'_n = Y_1 + Y_2 + \cdots + Y_n$ then from above we know $-S'_n$ and S'_n have the same distribution. Replacing the variables we have:

And so

Then

In other words S_n is symmetric about $\sum v_i$.

Example 24.1 Suppose that we take 101 (independent) random numbers from the set $\{0, 1, 2, \ldots, 9\}$. Find the probability that the sum of the numbers is less than 455.

25 Expected Value

We saw in the sections on the binomial distribution something called the "expected value" which most of you have probably seen as the mean since it's calculating the average probability. This was defined as $\mu = np$ where n was the number of trials which had a probability of p of success. This makes sense since normally when we think of the mean, we think about the average value of something:

$$\frac{x_1 + x_2 + \ldots + x_n}{n}$$

We can actually think of this in a probabilistic sense and look at how many times each x_i occurs and group them together. For example:

We can rewrite this as

We can use this formula to calculate the "mean" for arbitrary distributions The *expected value* of a random variable X is

Wikipedia: Expected value

Example 25.1 Suppose that we roll a fair six-sided die randomly. What is the expected value?

25.1 Indicator function

If you think about it, the binomial distribution was kind of weird in that we weren't talking about the expected value of just *one* trial, we were asking the expected value over *multiple* trials. So let's see how this kind of works.

Let's look at a binomial distribution where we run n trials each with a probability of success p. If we focus on just one trial, what do we get?

First we need to focus on how to represent this. Since our trial either succeeds or fails we can create a random variable where $X = 1$ if we're successful and $X = 0$ if we're not. This is called an *indicator function* of an event since it *indicates* whether or not an event succeeded. Then

Wikipedia: Indicator function

the expected value of the indicator function is given by:

What happens if we run this trial n times now? What is the expected number of successes? So if we let X_i be the indicator function that the ith trial was successful, then what we're asking is after n trials, for how many X_i do we have $X_i = 1$. We can calculate this using the following expression:

This is known as *linearity for expectation*. Note that it *does not* depend on whether the random variables are independent or not! So what this means is that, in the binomial distribution case we have:

The book calls this linearity the "addition rule for expected value"

since each $E(X_i) = p$ as they are all independent and each X_i is successful with probability p.

The cool thing about the addition rule is that we can alter our probabilities and look at the success of multiple events! So for example, if I have n different events and I let X_i be the indicator function for each event, *i.e.*, $X_i = 1$ if the ith event is successful. Then let p_i be the probability that the ith event is successful. By everything above, we have:

Example 25.2 Suppose that we do a binomial distribution where we draw a card 20 times with replacement, always looking for the ace of hearts. What is the expected number of successes?

Example 25.3 Suppose that we draw 5 cards from a deck of 52 cards and we end up with X number of twos. What is the expected number of twos among the five cards?