NICE TO KNOW

Poisson Distribution and PopTools

Lew Coggins

recorded: February, 2011

In the Week 5 modeling challenge, you were asked to answer a question that related to how the clutch size data that was provided in that challenge may have been related to a Poisson distribution. In this recording, what I want to do is examine that question, show you how those data are related to a Poisson distribution and also give you some demos of techniques of how to generate random numbers that are drawn from a Poisson distribution.

So why would you want to do that? Well, the Poisson distribution turns out that it's very useful in modeling ecological phenomena. Because we find that, when we look at data collected out in nature on things like the distribution of animal abundance, for instance the numbers of animals that are seen in different sites distributed across the landscape, are oftentimes distributed according to a Poisson distribution.

To frame this demo, what I've done is, I've constructed a simple spreadsheet here that probably looks familiar to you, to a degree, in that I'm setting up a procedure here that is similar to, I think, the procedure that you likely set up when you addressed or when you completed the Week 5 modeling challenge.

I'm conducting an individual-based model where I have a series of individuals actually all the way down to 200. I am randomly assigning each of these individuals some sized clutch. The clutch size is being drawn, it's being informed by a uniform random variable distributed between 0 and 1. That's the so-called RAND function that we've been using a lot, and it's informed by this cumulative sample probability.

Now that cumulative sample probability is being informed by the sample distribution that you were given for the assignment. You were told 100 female rabbits were sampled to look at how big of a clutch they had, and out of those 100 animals: 14 of them exhibited a clutch size of 0; 27 of them exhibited a clutch size of 1; 27 of them a clutch size of 2; and so forth. Now from that you can set up what was the proportion of the sample that fell into each one of these categories. For instance, for a clutch size of 0, there were 14 observations of that so the proportion would be 14 divided by 100, the sample size. Because the assump to 1, we can treat this as essentially what the probability is, the observed probability, of some other animal yet to be observed, what would the probability be of that animal having clutch size of a particular size?

We're going to assume that it's according to these proportions or to these probabilities.

Then what we can do is along the same lines that you've seen in a number of spreadsheets now, we can compute what the observed cumulative-sample probability would be in this case, and it would look like this.

So that's all fine and good. That's all kind of review. How did we get these numbers?

And the answer is, we used a Poisson distribution. Before we look at that, let's go ahead and plot out the sample probability and this cumulative-sample probability.

Okay, and here it is. This is the probability mass, which is the bars, as a function of clutch size. The cumulative probability, which is the line, with clutch size. So you can see that the mode is kind of between 1 to 2, maybe 1-1/2, although we can't have an observed 1-1/2 clutches, or we can't have a clutch size that's 1-1/2 animals. So it's 1 and 2 is the mode. You can see that this is kind of a left-skewed distribution so it has this somewhat long right-hand tail. That's a pretty characteristic looking Poisson distribution when it's mean is close to zero. So what does this Poisson distribution look like, in terms of how it's expressed as a formula?

The Poisson distribution is a discrete distribution, like the binomial, meaning that observed values from the up Poisson, the y - and in this case the clutch size - can only take on integer values. In fact, they can only take on integer values greater than 0 or greater, non-negative integer values. The probability mass function is shown here. What this formula says is similar to the binomial and says the probability of y, that is, some number of observations or, in our case, a clutch size of some number, given lambda, is equal to this formula. Now one thing to note here is that this distribution is completely specified with a single parameter, and that is the lambda. What is the lambda? The lambda is the mean of the Poisson distribution so in this case the lambda is the mean clutch size.

Let's go ahead and use this formula. What I'm going to do here is, I am going to write a header here that says Poisson Probability, and I'm going to center that and color it, as we have been. I'm going to write this formula right here.

First, I have to say what lambda is. Start off with lambda as 3, and we're going to color this green because this is an input parameter. If we write this formula, we E-X-P to the minus lambda divided by y exclamation mark. So that means ‘y factorial’. A factorial of a number is a product series. If you're unfamiliar with that, just look it up on Wikipedia. But say, factorial of 3 is 3 x 2 x 1. Factorial 4 is 4 x 3 x 2 x 1. That's a factorial.

In Excel, there's a formula for a factorial, and it's F-A-C-T. It requires only a single argument, and that argument in this case would be 0. It turns out that the factorial of 0 is 1. So that's not undefined because it's in the denominator of this quotient, and that's fine. And I'll go ahead and hit Return here so it says that .05 basically percent of the probability mass is associated with a clutch size of 0 or a y equal to 0 when lambda is equal to 3. Now I can copy this down. That shows what the probability is for clutch sizes of other size, with the lambda equals to 3. I'm just going to go ahead and shrink the number of decimals here, and I'm going to plot this.

Okay. So I did two things. The first thing I did was I also calculated the cumulative Poisson probability. And I did that the same way where I computed a running sum for these Poisson Probabilities, and you can see that it gets pretty close to 1. We'd actually have to have 7 and 8 or so to get this cumulative, to get all the way to 1. And then I plotted both the Poisson probability mass, which are the bars, and the Poisson cumulative probability, which is the line.

And so you can see that it's pretty obvious that these two distributions are different so what that means is that these sample data were not drawn from a Poisson probability distribution with mean 3. What might the mean have been?

We know that we specify the Poisson with just a single parameter, which is lambda. So what if we change this to 1? Okay. Well, that's not it either. These are obviously two different distributions. What if it's 1.5? That still doesn't look quite right. All right. Well, you've probably guessed it by now. It's 2.

These data that we gave you in the assignment are distributed according to a Poisson probability distribution, with mean clutch size equal to 2. So that kind of answers the question that's in the homework. Why might this be valuable to know how to do this?

Well, the thing is that we now have the ability to generate Poisson random variables as a function of lambda so we can change this lambda and generate a set of random variables drawn from a different Poisson probability distribution. So let's do that.

If we say this is Poisson Clutch Size, and I'll go ahead and color it. I can just copy this formula over one and then make sure that it reads from the appropriate places. The first thing is, I want it to read the same random variable, and the result factor is fine, but I wanted to use a different lookup factor. I wanted to use this one, which is based on this Poisson probability mouse function. And I'll go ahead and copy that down.

If you look closely at this, you'll see that, for the most part, these numbers are the same. It could be that there comes a random variable here that comes up, that does not quite result in the same clutch size. And that's because, remember that this sample probability for a clutch size of 0 was 14 over 100. This one's 27 over 100, so these numbers here are actually exact. But, if you look closer here, you'll see that that's not quite 14. So they might be a little bit different. Actually, this is more accurate. This actually is the Poisson probability, so this would be more accurate to use if you truly wanted a Poisson distribution with the mean specified as this lambda. But this is close enough. That's a pretty nice procedure.

Now both of these, if we took the average of these clutch sizes, we should see numbers, in both cases, pretty close to 2. So let's do that, equals average. And, if we look at the clutch size across all 200 of these samples, yeah, there we go, in this case it's 1.9. See there's a little difference here, and that's because these probabilities are slightly different than these. But, if we hit F9 here and watch those two cells, they should usually be very close to 2. If instead of looking at 200 individuals, we looked at 20,000, these would almost always be right on 2. That's kind of the machinery you need to draw **Poisson Distributed Random Variables**.

Now let's one kind of by scratch, and, as you probably would guess, there is an easier way to do this.

The first way you can do it is using Excel's random-number generator, and I actually rarely use this, to tell you the truth. But the way that works is if you go to Data and then if you've added in the Data Analysis Pack, which you can do from Excel Options, Add-ins. Go here: Manage Excel Add-ins. Go and make sure that this Analysis Dual Pack is checked. Then, when you click on the Data Ribbon, you will have access to this Data Analysis Menu Option. If you click that, there is this Random Number Generation. If you say OK here, it says how many variables do you want, that's basically you can think of that as how many columns. It's one. How many random numbers? Well, we'd want 200. What's the distribution? You can choose Discrete. There's a number of distributions that you can choose, not that many, but there is Poisson. It asks for the lambda, so you can say 2. You can choose a Random Seed if you want. You don't need to; it'll pick one itself. And then you go to this Output Range, and you choose this cell and say OK. And there you go. It generated 200 random numbers from a Poisson distribution with mean 2.

But one thing you'll see here is that this is not a volatile number. If I click on this cell, it's just a number that's put in here. So when I hit F9, these don't change. If I wanted to use this in some kind of a modeling context where I wanted to draw many, many samples from a Poisson distribution, it's tedious to use this Data Analysis functionality within Excel.

There is another option. That is an Add-on for Excel that's called **PopTools**. So if you go to Google and you type PopTools, you'll see this website called ‘poptools.org’ - PopTools home.

This is a piece of software that's free, although Greg does ask you for, if you want to download the documentation, I think he asks for $5. This is really a nice piece of software that I hope he continues to maintain for Excel. One of the things it does is, it allows you to generate random numbers from a whole series of distributions. It does a lot of other things, too, but useful in modeling an estimation. But that's one thing it does.

What you will see once it's installed is, under the Add-ins, you will have this option of PopTools. Now if you click on this, this shows all the stuff that PopTools will do. And, as I said, it has lots of utility for doing different things.

One of the things it does is, it'll generate a random variable. If you click on Random Variable, this dialogue box comes up. It's got a description here, actually tells you even where the algorithms came from. And, if you click this down arrow, you can choose a Poisson distribution.

It requires as arguments where are you going to put this random variable, and I'll say right here, and, again, it wants to know what the mean is. Well, that's the lambda so we'll tell at this cell right here. Again, you can tell it to generate more than one, but I'm going to just have it generate one. So there it is, and you can see that it actually puts in a formula here.

I can copy this down 200 cells and label this PopTools Poisson, and there you have it. If I examine the mean for these, you can see that they all are pretty close to 2, but what's nice about both this one and this one is, as I change the lambda value, these change. So we start to sample from this distribution rather than this one. The Excel version, as I said, they're just fixed numbers that are inserted in there. This one and these two are volatile. Each time I push F9, we get a different sample, and these two read from whatever mean I specify.

So I think you'll find that having some familiarity with the Poisson probability distribution and how to generate random numbers from the Poisson probability distribution will really serve you well, and, hopefully, this demo will help you on your way.

< 00:18:56 END >