STAT 515 – Sections 4.5 Supplement

Brian Habing – University of South Carolina
Last Updated: September 9, 2016

S4.5 – The Poisson Random Variable

Section 4.5 introduces the Poisson random variable with 4 basic characteristics. We say that something that follows these rules follows a Poisson process.

1) The experiment consists of counting the number of times (X) that a certain event occurs during some given amount of time (or area, volume, weight, distance, or any other unit of measurement).

2) The probability that an event occurs in any unit of measurement is the same for all units of that size.

3) The number of events that occur in one unit of measurement is independent of the number that occur in other units.

4) The mean number of events that occur in one unit of measurement is denoted by the Greek letter lambda (λ).

As shown on page 215 the probability mass function for a Poisson random variable is $P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}$ for x=0, 1, 2, …. .

S4.5.1 – Where does the Poisson probability mass function come from?

One obvious question is, where do the Poisson probability mass function, mean, and variance come from?

Say we wanted to make a simple model for the number of events that we would observe in one unit of time, where the average number of occurrences in one unit of time is λ (between 0 and 1). The simplest approach might be to liken it to flipping a coin with probability λ of heads – heads equals observing 1 event and tails equals not observing an event.

Model an event occurring in one unit of time as a single coin flip, where a success is seeing the event and the probability of a success is λ.

This gives a binomial random variable with p=λ and n=1. So the probability mass function is $P[X = x] = \lambda^x(1-\lambda)^{1-x}$ for x= 0 or 1 and it has mean λ and variance λ(1-λ). This seems good at first because it has property four just like we want. The problem is, we can only have one occurrence in a given unit of time. So, for example, if you were modeling the number of groups of customers who could show up during a 5 minute span at a restaurant, this model would say you could never have more than one group arriving.

One way to make this a little better is to split the time in half and imagine that we flipped two independent coins, where each has the probability of success cut in half.

|________________|________________| Model an event occurring in each 1/2 unit of time as a coin flip, where a success is seeing the event and the probability of a success is λ/2.

This would give a binomial random variable with p=λ/2 and n=2. So the probability mass function is $P[X=x]=\binom{2}{x}\frac{\lambda}{2}^{x}\left(1-\frac{\lambda}{2}\right)^{2-x}$ for x= 0, 1, or 2 and it has mean 2(λ/2) = λ and variance 2(λ/2) (1-λ/2). This has the right mean still (property 4), uses independence within different time units (property 2), and has the chance in each time unit being the same (property 3). And it's better because now we could have two groups of customers show up in that 5 minute span. But, why only two? You could keep getting a better model by splitting the amount of time represented by each flip in half.

|________|________|_______|_________| Model an event occurring in each 1/4 unit of time as a coin flip, where a success is seeing the event and the probability of a success is λ/4.

This gives $P[X=x]=\binom{4}{x}\frac{\lambda}{4}^{x}\left(1-\frac{\lambda}{4}\right)^{4-x}$ for x= 0, 1, … 4 with μ=4(λ/4) = λ and σ^2 = 4(λ/4) (1-λ/4).

|___|___|___|___|___|___|___|___|___| Model an event occurring in each 1/8 unit of time as a coin flip, where a success is seeing the event and the probability of a success is λ/8.

This gives $P[X=x]=\binom{8}{x}\frac{\lambda}{8}^{x}\left(1-\frac{\lambda}{8}\right)^{8-x}$ for x= 0, 1, … 8 with μ=8(λ/8) = λ and σ^2 = 8(λ/8) (1-λ/8).

In general, if you split it into n pieces, you would have each flip representing 1/n units of time and each flip would have probability of success λ/n. This would give $P[X=x]=\binom{n}{x}\frac{\lambda}{n}^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$ for x= 0, 1, … n with μ=n(λ/n) = λ and σ^2 = n (λ/n) (1-λ/n). If you simply picked a very large n, this seems like a reasonable model for the number of events observed in a particular amount of time.

If you've had calculus, you could take the limit of these as n goes to infinity. The mean is clearly always λ and you can show (by canceling the first n and then letting the other n go to infinity) that the variance would also be λ, just like the Poisson (see the box on page 215). It's a bit more work but you can also show that the binomial probability mass function actually turns into the Poisson probability mass function (see, for example http://www.oxfordmathcenter.com/drupal7/node/297).

S4.5.2 – Changing the units of time

Consider Example 4.14 on page 215-217. The number of reported sightings of blue whales is assumed to be approximately Poisson with λ=2.6 per week. Part b of the example wants to know the probability of having exactly 5 sightings in the one week. So that would be solved using the probability mass function.

$$P[X = 5] = \frac{2.6^5 e^{-2.6}}{5!} \approx 0.0735 .$$

Using R, you could find that with `dpois(5,2.6)`. For part c, they want $P[X < 2]$, the probability of 2 or fewer sightings. Noting that $P[X < 2] = P[X \leq 1]$ we could be found using the cumulative distribution function in R to get `ppois(1,2.6)` $\approx$ 0.2674

What happens if we know λ=2.6 per week, but they want to know about the number of occurrences in the entire month of September (30 days)? The easiest way to solve this is to be careful with our notation, and just adjust the units like you would in chemistry or physics. Putting all of the work in, this would give:

$$\lambda = \frac{2.6}{(week)} = \frac{2.6}{(7\ days)} = \frac{2.6}{7\ (days)} = \frac{0.37143}{(days)} = \frac{30}{30} \cdot \frac{0.37143}{(days)}$$

$$= \frac{11.1429}{(30\ days)} = \frac{11.1429}{(month)}$$

So we would use λ=11.1429 for problems involving months. This makes sense because the number of days in September is 30/7≈4.2857 times the number of days in a week, and the expected number of whales seen in a month (11.1429) is 4.286 times the expected number of whales seen in a week (2.6).

Say we want to know the probability of seeking exactly 10 whales in a month. That would be using the probability mass function $P[X = 10]$ using the λ=11.1429/month .

$$P[X = 10] = \frac{11.1429^{10} e^{-11.1429}}{10!} \approx 0.1177 .$$

This could be checked with R using `dpois(10,11.1429)`.

S4.5.3 – Thinking about the assumptions

Example 4.14 concerned the number of Whales that would be sighted in a week, and said it approximately followed a Poisson process with λ=2.6/week. In order to expand this to a month like we did in S4.5.2, we need assumption (2) to hold – the probability within any unit of time must be the same in any other unit of time. So, for example, if the temperature change from early to late September affects how many whales would be seen, then a single Poisson processs for the entire month is unreasonable. Another example would be the number of customers who arrive at a campus coffee shop. Even if the average over the entire day is 15/hour, that number is probably a lot higher early in the morning or between classes, and a lot lower at other times. So a Poisson process wouldn't work.

Assumption (3) says that the number of occurrences in any unit of time are independent of those showing up in any other unit. If students (or whales) travel in groups, then the Poisson process wouldn't apply to any situation where we were counting the number of individuals. It might be reasonable for counting groups showing up at restaurants (until we get an app warning us in advance about how crowded they are).