Chapter 4

The Poisson Distribution

4.1 The Fish Distribution?

The Poisson distribution is named after Simeon-Denis Poisson (1781–1840). In addition, *poisson* is French for fish.

In this chapter we will study a family of probability distributions for a countably infinite sample space, each member of which is called a **Poisson Distribution**. Recall that a binomial distribution is characterized by the values of two parameters: n and p. A Poisson distribution is simpler in that it has only one parameter, which we denote by θ, pronounced *theta*. (Many books and websites use λ, pronounced lambda, instead of θ. We save λ for a related purpose.) The parameter θ must be positive: $\theta > 0$. Below is the formula for computing probabilities for the Poisson.

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, \text{ for } x = 0, 1, 2, 3, \ldots. \tag{4.1}$$

In this equation, e is the famous number from calculus,

$$e = \lim_{n \to \infty} (1 + 1/n)^n = 2.71828\ldots.$$

You might recall from the study of infinite series in calculus, that

$$\sum_{x=0}^{\infty} b^x/x! = e^b,$$

for any real number b. Thus,

$$\sum_{x=0}^{\infty} P(X = x) = e^{-\theta} \sum_{x=0}^{\infty} \theta^x/x! = 1.$$

Thus, we see that Formula 4.1 is a mathematically valid way to assign probabilities to the nonnegative integers.

The mean of the Poisson is its parameter θ; i.e. $\mu = \theta$. This can be proven using calculus and a similar argument shows that the variance of a Poisson is also equal to θ; i.e. $\sigma^2 = \theta$ and $\sigma = \sqrt{\theta}$.

When I write $X \sim$ Poisson(θ) I mean that X is a random variable with its probability distribution given by the Poisson with parameter value θ.

I ask you for patience. I am going to delay my explanation of why the Poisson distribution is important in science.

Poisson probabilities can be computed by hand with a scientific calculator. Alternatively, you can go to the following website, which is linked to our course webpage:

   http://stattrek.com/Tables/Poisson.aspx

I will give an example to illustrate the use of this site.

Let $X \sim$ Poisson(θ). The website calculates five probabilities for you:

$$P(X = x); P(X < x); P(X \leq x); P(X > x); \text{ and } P(X \geq x).$$

You must give as input your value of θ and a value of x. Suppose that I have $X \sim$ Poisson(10) and I am interested in $P(X = 8)$. I go to the site and type '8' in the box labeled 'Poisson random variable,' and I type '10' in the box labeled 'Average rate of success.' I click on the 'Calculate' box and the site gives me the following answers:

$$P(X = 8) = 0.1126; P(X < 8) = 0.2202; P(X \leq 8) = 0.3328; P(X > 8) = 0.6672;$$

$$\text{and } P(X \geq 8) = 0.7798.$$

(There is, of course, a great deal of redundancy in these five answers b/c two pairs of events are complements of each other.)

It can be shown that for the Poisson, if $\theta \leq 5$ then its probability histogram is markedly asymmetrical, but if $\theta \geq 25$ its probability histogram is approximately symmetric and bell-shaped. This last statement suggests that we might use a normal curve to compute approximate probabilities for the Poisson, provided θ is large.

For example, suppose that $X \sim$ Poisson(25) and I want to calculate $P(X \geq 30)$. We will use a modification of the method we learned for the binomial.

First, we note that $\mu = 25$ and $\sigma = \sqrt{25} = 5$. Thus, our approximating curve will be the normal curve with these values for its mean and standard deviation. Using the continuity correction, we replace $P(X \geq 30)$ with $P(X \geq 29.5)$. Next, going to the normal curve website, we find that the area above (to the right of) 29.5 is 0.1841. From the Poisson website, I find that the exact probability is 0.1821.

To summarize: To approximate $P(X \geq x)$ for $X \sim$ Poisson(θ),

- Use the normal curve with mean equal to θ and standard deviation equal to $\sqrt{\theta}$.

- Find the area under the normal curve above (to the right of) $(x - 0.5)$.

If θ is unknown we can use the value of X to estimate it. The point estimate is x and, following the presentation for the binomial, we can use the snc to obtain an approximate confidence interval for θ. The result is:

$$x \pm z\sqrt{x}.$$

42

Here is an example of its use.

Ralph assumes that X has a Poisson distribution, but does not know the value of θ. He observes $x = 30$. His point estimate of the mean is 30 and his 95% confidence interval is

$$30 \pm 1.96\sqrt{30} = 30 \pm 10.7 = [19.3, 40.7].$$

We will now investigate the accuracy of the snc approximation. Suppose that, in fact, $\theta = 40$. The 95% confidence interval will be correct if, and only if,

$$X - 1.96\sqrt{X} \leq 40 \leq X + 1.96\sqrt{X}.$$

After algebra, this becomes $(30 \leq X \leq 54)$. The probability of this event, from the website, is 0.9428, which is pretty close to the desired 0.9500.

I repeated this analysis (calculating the exact probability that the CI is correct) for several values of θ; my results are below.

θ:	30	35	40	50	100
Exact Prob. of Correct Interval	0.9308	0.9368	0.9428	0.9487	0.9450

In my opinion, the snc approximation works adequately for $\theta \geq 40$. If you believe that θ might be smaller than 40 (and evidence of this would be if X was smaller than 40), then you might want to use an exact method, as I illustrated for the binomial. In fact, the website that gives us exact CI's for the binomial also gives exact CI's for the Poisson.

Bart assumes that $X \sim \text{Poisson}(\theta)$ but does not know the value of θ. He observes $X = 3$ and wants to obtain:

- The two-sided 95% CI for θ; and

- The upper one-sided 95% CI for θ.

I will use the website to find Bart's CI's. I type '3' (the value of X) into the 'Observed Events:' box and click on compute. (I don't need to specify the confidence level b/c the 95% two-sided CI is the default answer for this site.) I get $[0.6187, 8.7673]$ as the exact two-sided 95% CI for θ.

For the one-sided CI, I scroll down and type '5' in the 'upper tail' box and '0' in the 'lower tail' box. Then I scroll up and hit compute. I get the CI: $[0.0008, 7.7537]$. This is clearly a computer error—round-off error—b/c the lower bound must be 0. So, the answer is that 7.7537 is the 95% upper bound for θ.

4.2   Poisson Approximation to the Binomial

Earlier I promised that I would provide some motivation for studying the Poisson distribution.

We have seen that for the binomial, if n is moderately large and p is not too close to 0 (remember, we don't worry about p being close to 1) then a normal curve gives good approximations to binomial probabilities. In this section we will see that if p is close to 0 and n is large, the Poisson

can be used to approximate the binomial. Thus, the Poisson provides an approximate method in one of the situations in which the normal curve approximation is poor.

I will show you the derivation of this fact below. If you have not studied calculus and limits, you might find it to be too difficult to follow. This proof will not be on any exam in this course. Remember, if $X \sim \text{Bin}(n, p)$, then for a fixed value of x,

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}.$$

Now, replace p in this formula by θ/n. In my 'limit' argument below, as n grows, θ will remain fixed which means that $p = \theta/n$ will become smaller. We get:

$$P(X = x) = \frac{n!}{x!(n-x)!} (\theta/n)^x (1 - \theta/n)^{n-x} =$$

$$\frac{\theta^x}{x!} \left[\frac{n!}{(n-x)! n^x (1 - \theta/n)^x} \right] (1 - \theta/n)^n.$$

Now the term in the square brackets:

$$\frac{n!}{(n-x)! n^x (1 - \theta/n)^x},$$

for x fixed, converges (i.e. gets closer and closer) to 1 as $n \to \infty$; thus, it can be ignored for large n.

As shown in calculus, as $n \to \infty$,

$$(1 - \theta/n)^n$$

converges to $e^{-\theta}$. The result follows.

In the *old days* this result was very useful. For very large n and small p and computations performed by hand, the Poisson might be preferred to working with the binomial. *Nowadays*, as we will see below, this result is important mostly b/c it gives us greater insight into problems.

Next, we will consider estimation. Suppose that we have $n = 10{,}000$ BT and there are $x = 10$ successes observed. The website for the exact binomial confidence interval gives $[0.0005, 0.0018]$ for the 95% two-sided confidence interval for p. Alternatively, we can approximate the distribution of X by the Poisson with parameter $\theta = 10000p$. Using the observed $x = 10$, the exact 95% two-sided confidence interval for θ is $[4.7954, 18.3904]$. The CI is an assertion that the following inequality is true:

$$4.7954 \leq \theta \leq 18.3904.$$

Now we substitute $\theta = 10000p$ and this becomes

$$4.7954 \leq 10000p \leq 18.3904.$$

Dividing thru by 10000, we get the following CI for p:

$$0.0005 \leq p \leq 0.0018,$$

the same answer we had when we used the binomial distribution.

Now, I would understand if you feel, "Why should we learn to do the confidence interval for p two ways?" Fair enough; but computers ideally do more than just give us answers to specific questions; they let us learn about patterns in answers.

For example, suppose $X \sim \text{Poisson}(\theta)$ and we observe $X = 0$. From the website, the 95% one-sided confidence interval for θ is $[0, 2.9957]$. Why is this interesting?

Well, I have said that we don't care about cases where $p = 0$. But sometimes we might hope for $p = 0$. Borrowing from the movie, *Armageddon*, let every day be a trial and the day is a 'success' if the Earth is hit by a asteroid/meteor that destroys all human life. Obviously, throughout human habitation of this planet there have been no successes. Given 0 successes in n trials, the above answer indicates that we are 95% confident that $p \leq 2.9957/n$. Just don't ask me exactly what n equals. Or how I know that the trials are i.i.d.

4.3   The Poisson Process

The binomial distribution is appropriate for counting successes in n i.i.d. trials. For p small and n large, the binomial can be well approximated by the Poisson. Thus, it is not too surprising to learn that the Poisson is also a model for counting successes.

Consider a process evolving in time in which at 'random times' successes occur. What does this possibly mean? Perhaps the following picture will help.
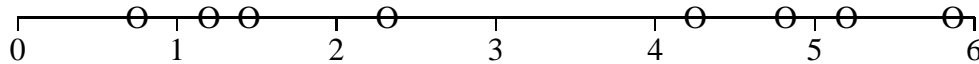
In this picture, observation begins at time $t = 0$ and time passing is denoted by moving to the right on the number line. At certain times, a success will occur, denoted by the letter 'O' placed on the number line. Here are some examples of such processes.

1. A 'target' is placed near radioactive material and whenever a radioactive particle hits the target we have a success.

2. An intersection is observed. A success is the occurrence of an accident.

3. A hockey (soccer) game is watched. A success occurs whenever a goal is scored.

4. On a remote stretch of highway, a success occurs when a vehicle passes.

The idea is that the times of occurrences of successes cannot be predicted with certainty. We would like, however, to be able to calculate probabilities. To do this, we need a mathematical model, much like our mathematical model for BT.

Our model is called the **Poisson Process**. A careful mathematical presentation and derivation is beyond the goals of this course. Here are the basic ideas:

1. The number of successes in disjoint intervals are independent of each other.

   For example, in a Poisson Process, the number of successes in the interval $[0, 3]$ is independent of the number of successes in the interval $[5, 6]$.

2. The probability distribution of the number of successes counted in any time interval depends only on the length of the interval.

   For example, the probability of getting exactly five successes is the same for interval $[0, 2.5]$ as it is for interval $[3.5, 6.0]$.

3. Successes cannot be simultaneous.

With these assumptions, it turns out that the probability distribution of the number of successes in *any* interval of time is the Poisson distribution with parameter θ, where $\theta = \lambda \times w$, where $w > 0$ is the length of the interval and $\lambda > 0$ is a feature of the process, often called its **rate**.

I have presented the Poisson Process as occurring in one dimension—time. It also can be applied if the one dimension is, say, distance. For example, a researcher could be walking along a path and at unpredictable places find successes. Also, the Poisson Process can be extended to two or three dimensions. For example, in two dimensions a researcher could be searching a field for a certain plant or animal that is deemed a success. In three dimensions a researcher could be searching a volume of air, water or dirt looking for something of interest.

The modification needed for two or three dimensions is quite simple: the process still has a rate, again called λ, and now the number of successes in an area or volume has a Poisson distribution with θ equal to the rate multiplied by the area or volume, whichever is appropriate.

4.4 Independent Poissons

Earlier we learned that if $X_1, X_2, \ldots, X_n$ are i.i.d. dichotomous outcomes (success or failure), then we can calculate probabilities for the sum of these guys X:

$$X = X_1 + X_2 + \ldots X_n.$$

Probabilities for X are given by the binomial distribution. There is a similar result for the Poisson, but the conditions are actually weaker. The interested reader can think about how the following fact is implied by the Poisson Process.

Suppose that for $i = 1, 2, 3, \ldots, n$, the random variable $X_i \sim \text{Poisson}(\theta_i)$ and that the sequence of X_i's are independent. (If all of the θ_i's are the same, then we have i.i.d. The point is that we don't need the i.d., just the independence.) Define $\theta_+ = \sum_{i=1}^{n} \theta_i$. The result is that $X \sim \text{Poisson}(\theta_+)$.

B/c of this result we will often (as I have done above), but not always, pretend that we have *one* Poisson random variable, even if, in reality, we have a sum of independent Poisson random variable. I will illustrate what I mean with an estimation example.

Suppose that Cathy observes 10 i.i.d. Poisson random variables, each with parameter θ. She summarizes the ten values she obtains by computing their total, X, remembering that $X \sim \text{Poisson}(10\theta)$. Cathy can then calculate a CI for 10θ and convert it to a CI for θ.

For example, suppose that Cathy observes a total of 92 when she totals her 10 values. B/c 92 is so large, I will use the snc to obtain a two-sided 95% CI for 10θ. It is:

$$92 \pm 1.96\sqrt{92} = 92 \pm 18.800 = [73.200, 110.800].$$

Thus, the two-sided 95% CI for θ is $[7.320, 11.080]$. BTW, the exact CI for 10θ is $[74.165, 112.83]$. This is typically what happens; the exact CI for a Poisson is shifted to the right of the approximate CI.

4.5   Why Bother with the Poisson?

Suppose that we plan to observe an i.i.d. sequence of random variables and that each random variable has for possible values: $0, 1, 2, \ldots$. (This scenario frequently occurs in science.) In this chapter I have suggested that we assume that each random variable has a Poisson distribution. But why? What do we gain? Why not just do the following? Define

$$p_0 = P(X = 0), p_1 = P(X = 1), p_2 = P(X = 2), \ldots,$$

where there is now a sequence of probabilities known only to nature. As a researcher we can try to estimate this sequence.

This question is an example of *a*, if not *the*, fundamental question a researcher always considers: How much math structure should we impose on a problem? Certainly, the Poisson leads to values for $p_0, p_1, p_2, \ldots$. The difference is that with the Poisson we impose a structure on these probabilities, whereas in the 'general case' we do not impose a structure.

As with many things in human experience, many people are too extreme on this issue. Some people put too much faith in the Poisson (or other assumed structures) and cling to it even when the data make its continued assumption ridiculous; others claim the moral high ground and proclaim: "I don't make unnecessary assumptions." I cannot give you any rules for how to behave; instead, I will give you an extended example of how answers change when we change assumptions.

Let us consider a Poisson Process in two dimensions. For concreteness, imagine you are in a field searching for a plant/insect that you don't particularly like; i.e. you will be happiest if there are none. Thus, you might want to know the numerical value of $P(X = 0)$. Of course, $P(X = 0)$ is what we call p_0 and for the Poisson it is equal to $e^{-\theta}$.

Suppose it is true (i.e. this is what Nature knows) that $X \sim$ Poisson(0.6931) which makes

$$P(X = 0) = e^{-0.6931} = 0.500.$$

Suppose further that we have two researchers:

- Researcher **A** assumes Poisson with unknown θ.

- Researcher **B** assumes no parametric structure; i.e. **B** wants to know p_0.

Note that both researchers want to get an estimate of 0.500 for $P(X = 0)$.

Suppose that the two researchers observe the same data, namely $n = 10$ trials. Who will do better? Well, we answer this question by simulating the data. I used my computer to simulate $n = 10$ i.i.d. trials from the Poisson(0.6931) and obtained the following data:

1, 0, 1, 0, 3, 1, 2, 0, 0, 4.

Researcher **B** counts four occurrences of '0' in the sample and estimates $P(X = 0)$ to be $4/10 = 0.4$. Researcher **A** estimates θ by the mean of the 10 numbers: $12/10 = 1.2$ and then estimates $P(X = 0)$ by $e^{-1.2} = 0.3012$. In this one simulated data set, each researcher's estimate is too low and Researcher **B** does better than **A**.

One data set, however, is not conclusive. So, I simulated 999 more data sets of size $n = 10$ to obtain a total of 1000 simulated data sets. In this simulation, sometimes **A** did better, sometimes **B** did better. Statisticians try to decide which does better overall.

First, we look at how each researcher did on average. If you average the 1,000 estimates for **A** you get 0.5226 and for **B** you get 0.5066. Surprisingly, **B**, who makes fewer assumptions, is, on average, closer to the truth. When we find a result in a simulation study that seems surprising we should wonder whether it is a false alarm caused by the approximate nature of simulation answers. While I cannot explain *why* at this point, I will simply say that this is *not a false alarm*. A consequence of assuming Poisson is that, especially for small values of n, there can be some **bias** in the mean performance of an estimate. By contrast, the fact that the mean of the estimates by **B** exceeds 0.5 is not meaningful; i.e. **B**'s method does not possess bias.

I will still conclude that **A** is better than **B**, despite the bias; I will now describe the basis for this conclusion.

From the point-of-view of Nature, who knows the truth, every estimate value has an error: $e = $ estimate minus truth. In this simulation the error e is the estimate minus 0.5. Now errors can be positive or negative. Also, trying to make sense of 1000 errors is too difficult; we need a way to summarize them. Statisticians advocate averaging the errors after making sure that the negatives and positives don't cancel. We have two preferred ways of doing this:

- Convert each error to an **absolute error** by taking its absolute value.

- Convert each error to a **squared error** by squaring it.

For my simulation study, the **mean absolute error** is 0.1064 for **A** and 0.1240 for **B**. B/c there is a minimum theoretical value of 0 for the mean absolute error, it makes sense to summarize this difference by saying that the mean absolute error for **A** is 14.2% smaller than it is for **B**. This 14.2% is my preferred measure and why I conclude that **A** is better than **B**.

As we will see, statisticians like to square errors, although justifying this in an intuitive way is a bit difficult. I will just remark that for this simulation study, the mean squared error for **A** is 0.001853 and for **B** it is 0.002574. (B/c all of the absolute errors are 0.5 or smaller, squaring the errors make them smaller.)

To revisit the issue of bias, I repeated the above simulation study, but now with $n = 100$. The mean of the estimates for **A** is 0.5006 and for **B** is 0.5007. These discrepancies from 0.5 are not meaningful; i.e. there is no bias.