Chapter 13

The Poisson Distribution

Jeanne Antoinette Poisson (1721–1764), Marquise de Pompadour, was a member of the French court and was the official chief mistress of Louis XV from 1745 until her death. The pompadour hairstyle was named for her. In addition, poisson is French for fish. The Poisson distribution, however, is named for Simeon-Denis Poisson (1781–1840), a French mathematician, geometer and physicist.

13.1 Specification of the Poisson Distribution

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.

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, \text{ for } x = 0, 1, 2, 3, \ldots. \tag{13.1}$$

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,

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

Table 13.1: A comparison of three probability distributions.

	Distribution of X is:		
	Poisson(1)	Bin$(1000, 0.001)$	Bin$(500, 0.002)$
Mean :	1	1	1
Variance :	1	0.999	0.998
x	$P(X = x)$	$P(X = x)$	$P(X = x)$
0	0.3679	0.3677	0.3675
1	0.3679	0.3681	0.3682
2	0.1839	0.1840	0.1841
3	0.0613	0.0613	0.0613
4	0.0153	0.0153	0.0153
5	0.0031	0.0030	0.0030
6	0.0005	0.0005	0.0005
≥ 7	0.0001	0.0001	0.0001
Total	1.0000	1.0000	1.0000

Thus, we see that Formula 13.1 is a mathematically valid way to assign probabilities to the nonnegative integers; i.e., all probabilities are nonnegative—indeed, they are positive—and they sum to one.

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 distribution 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.

As we will see, the Poisson distribution is closely tied to the binomial. For example, let's spend a few minutes looking at the three probability distributions presented in Table 13.1.

There is a wealth of useful information in this table. In particular,

1. If you were distressed that a Poisson random variable has an infinite number of possible values—namely, every nonnegative integer—agonize no longer! We see from the table that for $\theta = 1$, 99.99% of the Poisson probability is assigned to the event $(X \leq 6)$.

2. If you read down the three columns of probabilities, you will see that the entries are nearly identical. Certainly, any one column of probabilities provides good approximations to the entries in any other column. Thus, in some situations, a Poisson distribution can be used as an approximation to a binomial distribution.

3. What do we need for the Poisson to be a good approximation to a binomial? First, we need to have the means of the distributions match; i.e., we need to use the Poisson with $\theta = np$, as I did in Table 13.1. The variance of a binomial npq is necessarily smaller than the mean

np because $q < 1$. Thus, the variance of a binomial *cannot be made to match* the variance of the Poisson:

$$\text{Variance of binomial } = npq < np = \theta = \text{ variance of Poisson.}$$

If, however, p is very close to 0, then q is very close to one and the variances *almost match* as illustrated in Table 13.1.

I will summarize the above observations in the following result.

Result 13.1 (The Poisson approximation to the binomial.) *The Bin(n, p) distribution can be well-approximated by the Poisson(θ) distribution if the following conditions are met:*

1. *The distributions have the same mean; i.e., $\theta = np$;*

2. *The value of n is large and p is close to zero. In particular, the variance of the binomial npq should be very close to the variance of the Poisson, $\theta = np$.*

As a practical matter, we mostly use this result if $n >$ 1,000 because we can easily obtain exact binomial probabilities from a website for $n \leq$ 1,000. Also, if $np \geq 25$, our general guideline from Chapter 11 states that we may use a Normal curve to obtain a good approximation to the binomial. Thus, again as a practical matter, we mostly use this result if $\theta = np \leq 25$, allowing us some indecision as to which approximation to use at $np = 25$, Normal or Poisson.

Poisson probabilities can be computed by hand with a scientific calculator. Alternatively, the following website can be used:

`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 enter 8 in the box *Poisson random variable*, and I enter 10 in the box *Average rate of success.* I click on the *Calculate* box and the site gives me the following answers:

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

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

As with our binomial calculator, there is a great deal of redundancy in these five answers.

13.1.1 The Normal Approximation to the Poisson

Please look at the Poisson(1) probabilities in Table 13.1. We see that $P(X = 0) = P(X = 1)$ and as x increases beyond 1, $P(X = x)$ decreases. Thus, without actually drawing the probability histogram of the Poisson(1) we know that it is strongly skewed to the right; indeed, it has no left tail! For $\theta < 1$ the probability histogram is even more skewed than it is for our tabled $\theta = 1$. As the value of θ increases the amount of skewness in the probability histogram decreases, but the Poisson is never perfectly symmetric.

In this course, I advocate the general guideline that if $\theta \geq 25$, then the Poisson's probability histogram is approximately symmetric and bell-shaped. (One can quibble about my choice of 25 and I wouldn't argue about it much.) 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.

13.2 Inference for a Poisson distribution

If θ is unknown then its point estimator is X, with point estimate equal to x, the observed value of X. We have two options for obtaining a confidence interval estimate of θ: an approximate interval based on using a Normal curve approximation and an exact (conservative) confidence interval using the Poisson equivalent of the work of Clopper and Pearson.

It is possible to perform a test of hypotheses on the value of θ. The test is not widely useful in science; thus, I won't present it.

13.2.1 Approximate Confidence Interval for θ

I will very briefly sketch the rational behind the Normal curve approximation. The main ideas are pretty much exactly the ideas we had for the binomial in Chapter 12. We standardize our point estimator X to obtain

$$Z = \frac{X - \theta}{\sqrt{\theta}}.$$

Next, we replace the unknown parameter in the denominator by its point estimator, yielding

$$Z' = \frac{X - \theta}{\sqrt{X}}.$$

Slutsky's theorem applies; for θ sufficiently large, probabilities for Z' can be well-approximated by using the N(0,1) curve. With the same algebra we used in Chapter 12, we obtain the following

approximate confidence interval estimate of θ:

$$x \pm z^* \sqrt{x}, \tag{13.2}$$

where the value of z^* is determined by the choice of confidence level *in exactly the same way as it was for the binomial*. Thus, you can find the z^* you need in Table 12.1 on page 296.

I have investigated the performance of Formula 13.2 and I have concluded that the approximation is good for any $\theta \geq 40$; i.e., for any $\theta \geq 40$ the actual probability that this formula will give a correct confidence interval is close to the target reflected by the choice of z^*. As always, one can quibble with my choice of 40 as the magic threshold. It is larger than my choice, 25, for using a Normal curve to approximate Poisson probabilities in part because the confidence interval also relies on Slutsky's approximation.

In practice, of course, we estimate θ because we don't know its value. Thus, if you are concerned with having a guideline based on the value of θ, an alternative guideline is to use the approximate confidence interval if $x \geq 50$.

13.2.2 The 'Exact' (Conservative) Confidence Interval for θ

Suppose that we plan to observe a random variable X and we are willing to assume that $X \sim$ Poisson(θ). We want to use the observed value of X to obtain a confidence interval for θ, but the condition for using the approximate method of the previous subsection is not met. For example, suppose that we observe $X = 10$; what should we do?

In Chapter 12, when you learned how to use the website:

`http://statpages.org/confint.html`

you probably noticed that the website also can be used for Poisson distribution. Click on this website now and scroll down to the section **Poisson Confidence Intervals**. You will see that there is one box for data entry, called **Observed Events**; this is where you place the observed value of X. Note that the default value is 10, which, coincidentally, is the value I asked you to use! Click on the *Compute* box and the site gives you the exact—which, as in Chapter 12, really means conservative—two-sided 95% confidence interval for θ:

$$[4.7954, 18.3904].$$

If, instead, you want the two-sided 98% confidence interval for θ, then you proceed exactly as you did in Chapter 12. Scroll down to **Setting Confidence Levels**, type 98 in **Confidence Level** and click on *Compute*. Scroll back up to **Poisson Confidence Intervals** and make sure that 10 is still in the **Observed Events** box. Click on the *Compute* box and the site gives the answer:

$$[4.1302, 20.1447].$$

Suppose that I want the one-sided 90% upper confidence bound for θ, still with $x = 10$. Scroll down to **Setting Confidence Levels**, enter 10 in the **Upper Tail**, enter 0 in the **Lower Tail** and

click on *Compute*. Scroll back up to **Poisson Confidence Intervals** and make sure that 10 is still in the **Observed Events** box. Click on the *Compute* box and the site gives the answer:

$$[0.4749, 15.4066].$$

This answer is a bit strange; the lower bound in the interval should be 0, but it's not. I played around with this website a bit and here is what I learned. If $x \leq 2$ then the site gives 0 as the (correct) lower bound for the one-sided interval. If, however, $x \geq 3$, it gives a positive lower bound, which seems to be incorrect. This is not incorrect for two reasons:

1. We are free to replace the non-zero lower bound with 0 if we want; by making the interval wider, the probability of a correct interval becomes a bit larger.

2. Without examining either the programmer's code or performing a huge analysis—which I have neither the time nor interest to do—I can't be sure, but I believe that having a non-zero lower bound is part of the conservative nature of the site's intervals. Here is what I mean. If θ actually equaled the lower bound I have above for $x = 10$, which is 0.4749, then the probability of 10 or more successes is 10^{-10} (you can find this on our website for computing Poisson probabilities). Thus, if $x = 10$, values of θ smaller than 0.4749 are pretty much impossible anyways.

The next example shows why this material provides insight into some of our work in Chapter 12.

Example 13.1 (Don K. and high hopes) *Don K. was a teammate on my high school basketball team. Don wasn't very tall, but he was very quick and had a very strong throwing arm. He started his senior year as first or second player off the bench, but as the year progressed his playing time diminished. A highlight of his year was when he sank a half-court shot at the end of a quarter in a blow-out 93-40 victory. After his amazing shot, Don would spend most of his practice free time attempting very long shots. I don't remember him making many such shots, but everyone on the team noted how our coach, Mr. Pasternak— whom we affectionately dubbed* Boris *either because of his resemblance to the actor Boris Karloff or because* Doctor Zhivago *was* **the** *movie of 1965— was doing a slow boil from frustration. Finally, one day at practice, Coach could contain himself no longer and berated Don at length for not practicing a more useful basketball skill. Eight minutes later during a scrimmage as the time clock was running down to zero, Don grabbed a defensive rebound, pivoted and threw the ball 70 (?) feet, resulting in a perfect basket—swish through the net. Don ran around the court yelling, "See, Boris, I have been practicing a useful shot," while the rest of us collapsed in laughter.*

Perhaps because of my friend Don's experience, I have always been interested in situations in which successes are rare. Thus, let's look at some examples. I used the site

`http://statpages.org/confint.html`

to obtain the exact (conservative) 95% upper bound for p in each of the situations below.

- A total of $x = 0$ successes are obtained in $n = 10$ Bernoulli trials; the exact (conservative) 95% upper bound for p is: $p \leq 0.2589$.

- A total of $x = 0$ successes are obtained in $n = 100$ Bernoulli trials; the exact (conservative) 95% upper bound for p is: $p \leq 0.0295$.
- A total of $x = 0$ successes are obtained in $n = 1{,}000$ Bernoulli trials; the exact (conservative) 95% upper bound for p is: $p \leq 0.0030$.

As I have mentioned a number of times in these notes, the weakness of exact answers is that they are a black box; we can't see a pattern in the answers. There is a pattern in the above answers, as I will now demonstrate. (Indeed, you might see the pattern above, but you won't know *why* until you read on.)

Let's suppose now that our random variable X has a Poisson distribution and we observe $x = 0$. Using the same website, I can obtain an upper 95% confidence bound for θ; it is $\theta \leq 2.9957$, which, when I am feeling especially daring, I round to $\theta \leq 3.000$. Now we are going to use the fact that, under certain conditions, we can use the Poisson to approximate the binomial. Ignoring the conditions for a moment, recall that the key part of the approximation is to set θ for the Poisson equal to np from the binomial. Thus—and this is the key point—an exact confidence interval for θ is an approximate confidence interval for np. Thus, the upper bound $\theta \leq 3.000$ becomes $np \leq 3.000$ which becomes the following result.

Result 13.2 (Approximate 95% Confidence Upper Bound for p When $x = 0$.) *If $n \geq 100$,*

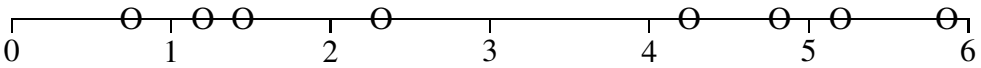
$$p \leq 3/n, \tag{13.3}$$

is a good approximation to the exact 95% confidence upper bound for p when $x = 0$. This result is sometimes referred to as the rule of 3.

13.3 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 distribution 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 the passage of time is denoted by moving to the right on the number line. At various times successes will occur, with each success 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. A road intersection is observed. A success is the occurrence of an accident.

3. A hockey (or 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 Bernoulli trials.

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. **Independence:** 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. **Identically distributed:** 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. (This assumption is needed for technical reasons that we won't discuss.)

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 occasionally finds 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 Poisson 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. Also, of course, to be a Poisson Process in two or three dimensions requires the assumptions of independence and identically distributed to be met.

13.4 Independent Poisson Random Variables

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 result is implied by the Poisson Process.

Result 13.3 (The sum of independent Poisson random variables.) *Suppose that for $i = 1, 2, 3, \ldots, n$, the random variable $X_i \sim$ Poisson(θ_i) and that the sequence of X_i's are independent. Define $\theta_+ = \sum_{i=1}^{n} \theta_i$. Then $X \sim$ Poisson(θ_+).*

Because 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 variables. 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 $X \sim$ Poisson(10θ). Cathy can then calculate a confidence interval for 10θ and convert it to a confidence interval for θ.

For example, suppose that Cathy observes a total of 92 when she sums her 10 values. Because 92 is sufficiently large (it exceeds 50), I will use the formula for the approximate two-sided 95% confidence interval for 10θ. It is:

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

The interpretation of this interval is, of course:

$$73.200 \leq 10\theta \leq 110.800.$$

If we divide through by 10, we get

$$7.3200 \leq \theta \leq 11.0800.$$

Thus, the two-sided approximate 95% confidence interval for θ is $[7.320, 11.080]$. By the way, the exact confidence interval for 10θ is $[74.165, 112.83]$. This is typically what happens; the exact confidence interval for a Poisson is shifted to the right of the approximate confidence interval because the Poisson distribution is skewed to the right.

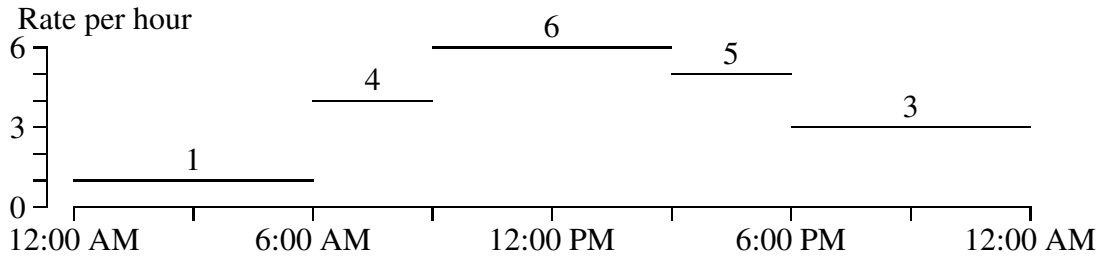
13.4.1 A Comment on the Assumption of a Poisson Process

Recall my four examples of possible Poisson Processes given on page 327. My first example, radioactive decay, was, by far, the most popular example in textbooks on probability theory, circa 1970, when I was an undergraduate student. Literally, radioactive decay involves a source of radioactive material comprised of a huge number of atoms, each of which has a very small probability of decaying in a short time period. Because atoms don't talk to each other, "Hey, Adam, I am about to decay, will you join me?" it seems extremely reasonable to believe we have a huge number of Bernoulli trials with a very small value of p. Hence, assuming a Poisson Process is simply restating the idea that the Poisson distribution approximates the binomial. All models have an implicit

expiration date; for example, if I am still shooting free throws at age 80, I definitely won't have the same p I had at age 17. For radioactive decay, if the length of observation approaches the half-life of the element then the rate will definitely decrease because—by definition—half the atoms have *decayed* at the half life. For example, uranium-232 has a half-life of 69 years and carbon-14, which is used to date fossils, has a half-life of 5,730 years.

I hope that you will agree that radioactive decay is a pretty solid example of a Poisson Process. My second and fourth examples—both involving traffic—appear, however, to be on shaky ground. Let's examine the fourth example, in which a success is the passage of a car on a remote stretch of highway. When I think of a remote highway, it is hard for me to imagine that the rate of traffic at, say, 3:00 AM is the same as it is at 3:00 PM. Thus, you might think that the assumption of a Poisson Process is reasonable only for a very limited period of time, say, 9:00 AM to 4:00 PM. You would be correct, except for what I am now going to tell you, which is the point of this subsection.

I want to make this argument very concrete. To that end, suppose that I am Nature and I know that the rate is as given in the following picture.



Let's make sure that this picture is clear. From 12:00 AM (midnight) to 6:00 AM a car passing the spot follows a Poisson Process with an average of one car per hour. From 6:00 AM to 9:00 AM the rate of the Poisson Process quadruples to four cars per hour; and so on.

If we watch the road continuously, then we do not have a Poisson Process over the 24 hours of a day because the rate is not constant. If I look at the process for *certain limited periods of time*, then I will have a Poisson Process; for example, if I observe the process over the six hour time period of 9:00 AM to 3:00 PM, I am observing a Poisson Process with rate equal to six cars per hour.

Now let's imagine, however, that we **do not** observe the process continuously at all. Instead, every day at the same time, say midnight, we are told how many cars passed the spot in the day just completed. Call this observed count x with corresponding random variable X. I will now demonstrate that X has a Poisson distribution.

We can write X as the sum of five random variables:

$$X = X_1 + X_2 + X_3 + X_4 + X_5,$$

where

- X_1 is the number of cars that pass the spot between midnight and 6:00 AM.
- X_2 is the number of cars that pass the spot between 6:00 AM and 9:00 AM.

Table 13.2: The number of homicides, by year, in Baltimore, Maryland.

Year:	2003	2004	2005	2006	2007
Number of homicide deaths:	270	276	269	276	282

- And so on, for X_3, X_4 X_5, throughout the day.

From the above picture, being Nature I know that:

- $X_1 \sim$ Poisson($6 \times 1 = 6$); $X_2 \sim$ Poisson($4 \times 3 = 12$); $X_3 \sim$ Poisson($6 \times 6 = 36$); $X_4 \sim$ Poisson($5 \times 3 = 15$); and $X_5 \sim$ Poisson($3 \times 6 = 18$).
- Also, the random variables X_1, $X_2 \ldots X_5$ are statistically independent.
- From Result 13.3, we know that X has a Poisson distribution with parameter

$$\theta_+ = 6 + 12 + 36 + 15 + 18 = 87.$$

I might even abuse language a bit and say that the number of cars passing the spot is a Poisson Process with a rate of 87 cars per day. I shouldn't say this of course, but sometimes we get a bit lazy in probability and statistics!

Of course, I am not Nature, so I would never know the exact rate. The following example with real data is illustrative of the above method.

Example 13.2 (Homicides in Baltimore.) *I recently discovered data on homicides, by year, in Baltimore, Maryland. The data are presented in Table 13.2.*

I am going to assume that the number of homicides per year is a Poisson Process with unknown rate of λ homicides per year. I will revisit this example in Chapter 14. With my assumption, I have observed the process for five units of time—five years—and counted a total of

$$270 + 276 + 269 + 276 + 282 = \text{ 1,373 successes.}$$

(Remember that whatever we are counting, no matter how tragic it might be, is called a success. Hence, a homicide death is a success.) We view 1,373 as the observed value of a random variable X with Poisson(θ) distribution. Because my observed value of X is much larger than 50, I feel comfortable using the approximate confidence interval for θ, given in Formula 13.2. For 95% confidence, we get

$$1373 \pm 1.96\sqrt{1373} = 1373 \pm 72.6 = [1300.4, 1445.6].$$

Because the process was observed for five time units, we have $\theta = 5\lambda$. Thus, the above confidence interval for θ becomes

$$1300.4 \leq 5\lambda \leq 1445.6;$$

after dividing through by 5, we get

$$260.08 \leq \lambda \leq 289.12.$$

Thus, $[260.08, 289.12]$ is my approximate 95% confidence interval for the rate of homicides per year in Baltimore during the years 2003–2007.

13.5 Summary

The Poisson is a probability distribution—see Equation 13.1—concentrated on the nonnegative integers. The Poisson distribution has a single parameter, θ, which can be any positive number. The mean and variance of a Poisson distribution both equal θ and the standard deviation equals $\sqrt{\theta}$.

Poisson probabilities can be calculated with the help of the website:

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

If $\theta \geq 25$, then the Normal curve with $\mu = \theta$ and $\sigma = \sqrt{\theta}$ will give good approximations to the Poisson(θ) distribution.

The first use for the Poisson distribution is as an approximation to the binomial distribution. In particular, suppose we have a Bin(n, p) distribution, with n large, p small and npq approximately equal to np; i.e., q is very close to one. If we set $\theta = np$, then the Poisson distribution is a good approximation to the Binomial distribution.

If $X \sim$ Poisson(θ), then X is the point estimator of θ. The standardized version of the point estimator X is

$$Z = \frac{X - \theta}{\sqrt{\theta}}.$$

As implied above, if $\theta \geq 25$, then the N(0,1) curve provides good approximate probabilities for Z. Combining the above with Slutsky's theorem, we obtain the following approximate confidence interval for θ:

$$x \pm z^* \sqrt{x},$$

where the value of z^* depends on the choice of confidence level and is given in Table 12.1 on page 296. My advice is that this interval performs as advertised provided $x \geq 50$. For smaller values of x, see the next paragraph.

There is an exact—actually conservative—confidence interval for θ, available on the website:

`http://statpages.org/confint.html`

The Poisson distribution also arises from a mathematical model for successes occurring randomly in time. In particular, the first two of the three assumptions of a Poisson Process are similar to the assumptions of Bernoulli trials. If we have a Poisson Process then the number of successes in any time interval of length w has a Poisson distribution with parameter $\theta = w\lambda$, where $\lambda > 0$ is a parameter of the process, called its rate. (If $w = 1$, then $\theta = \lambda$. Thus, the mean number of successes in one unit of time is λ; hence, the name rate.)

When I talk about a Poisson Process in general, I will speak of it evolving in time. It could, alternatively, evolve in distance. Moreover, a Poisson Process can be used for counting successes in two or three dimensions.

The Poisson distribution has the following very useful property. If the random variables X_1, X_2, $\ldots$, X_n, are independent with $X_i \sim$ Poisson(θ_i)—i.e., the X_i's need not be identically distributed—then the new random variable

$$X = X_1 + X_2 + \ldots X_n = \sum X_i,$$

has a Poisson distribution with parameter

$$\theta_+ = \theta_1 + \theta_2 + \ldots \theta_n = \sum \theta_i.$$

In words, the sum of independent Poisson random variables has a Poisson distribution; and the parameter for the sum is the sum of the parameters. This property of Poisson distributions can be very useful; I illustrate its use with data on the annual number of homicide deaths in Baltimore, Maryland.

13.6 Practice Problems

1. Suppose that $X \sim$ Poisson(20). Use the website

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

   to calculate the following probabilities.

   (a) $P(X = 20)$.
   (b) $P(X \leq 20)$.
   (c) $P(X > 20)$.
   (d) $P(16 \leq X \leq 24)$.

2. Suppose that $X \sim$ Bin(2000,0.003). I want to know $P(X \leq 4)$. Help me by calculating an approximate probability for this event.

3. Wayne Gretzky is perhaps the greatest hockey player ever. We have the following data from his NHL (National Hockey League) career.

   - During the 1981–82 season he played 80 games and scored 92 goals.
   - During the 1982–83 season he played 80 games and scored 71 goals.
   - During the 1983–84 season he played 74 games and scored 87 goals.

   Assume that Gretzky's goal scoring followed a Poisson Process with a rate of λ goals per game. Use the three seasons of data given above to obtain an approximate 98% confidence interval for λ.

4. Let $X \sim$ Poisson(θ). Given $X = 1$, find the exact 95% upper confidence bound for θ. Apply your finding to create *the rule of 4.75 when* $X = 1$.

13.7 Solutions to Practice Problems

1. For parts (a)–(c), go to the website and enter 20 for both x and the **Average rate of success**. You will obtain:

   (a) $P(X = 20) = 0.0888$.

   (b) $P(X \leq 20) = 0.5591$.

   (c) $P(X > 20) = 0.4409$.

   (d) There are several ways to get the answer. I suggest:

   $$P(16 \leq X \leq 24) = P(X \leq 24) - P(X \leq 15).$$

   I enter the website twice and obtain:

   $$P(16 \leq X \leq 24) = 0.8432 - 0.1565 = 0.6867.$$

2. Our binomial calculator website does not work for $n >$ 1,000; hence, I want an approximate answer. For the binomial, the mean is $np = 2000(0.003) = 6$. This is much smaller than 25, so I will not use the Normal curve approximation. In addition, the binomial variance is $npq = 6(0.997) = 5.982$ which is only a bit smaller than the mean. Thus, I will use the Poisson approximation. I go to the website

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

   and enter 4 for x and $\theta = np = 6$ for **Average rate of success**. The website gives me 0.2851 as its approximation of $P(X \leq 4)$.

   By the way, Minitab is able to calculate the exact probability; it is 0.2847. Thus, the Poisson approximation is very good.

3. Combining the data, we find that Gretzky scored 250 goals in 234 games. We view $x = 250$ as the observed value of a random variable X which has a Poisson distribution with parameter θ. Also, $\theta = 234\lambda$. For 98% confidence, we see from Table 12.1 that $z^* = 2.326$. Thus, the approximate 98% confidence interval for θ is

   $$250 \pm 2.326\sqrt{250} = 250 \pm 36.78 = [213.22, 286.78].$$

   Literally, we are asserting that

   $$213.22 \leq \theta \leq 286.78 \text{ or } 213.22 \leq 234\lambda \leq 286.78.$$

   Dividing through by 234, we get

   $$213.22/234 \leq \lambda \leq 286.78/234 \text{ or } 0.911 \leq \lambda \leq 1.226.$$

4. Go to the website

   `http://statpages.org/confint.html`.

   Scroll down to **Setting Confidence Levels**. Enter 5 in the **Upper** box, 0 in the **Lower** box and click on **Compute**. The site now knows that we want the 95% upper confidence bound.

   Scroll up to **Poisson Confidence Intervals**, enter 1 in the **Observed Events** box and click on **Compute**. The site gives us $[0, 4.7439]$ as the upper 95% confidence bound for θ.

   If $X \sim \text{Bin}(n, p)$ with n large and the observed value of X is 1, then 4.7439, rounded rather clumsily to 4.75, is the approximate 95% upper confidence bound for np. Thus, for n large and $X = 1$,

   $$4.75/n \text{ is the approximate } 95\% \text{ upper confidence bound for } p.$$

   As a partial check, I scrolled up to **Binomial Confidence Intervals**, entered 1 for x, entered 100 for n, and clicked on **Compute**. The site gave me 0.0466 as the exact 95% upper confidence bound for p, which is reasonably approximated by $4.75/n = 4.75/100 = 0.0475$.

Table 13.3: Traffic accident data in Madison, Wisconsin.

Year:	2005	2006	2007	2008	2009	Total
Average weekday arterial volume	26,271	25,754	25,760	24,416	24,222	126,423
Total crashes	4,577	4,605	4,779	4,578	4,753	23,292
Bike crashes	97	95	118	95	115	520
Pedestrian crashes	84	87	80	76	77	404
Fatal crashes	9	12	13	6	14	54

13.8 Homework Problems

1. Suppose that $X \sim$ Poisson(10). Use the website

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

   to calculate the following probabilities.

   (a) $P(X = 8)$.

   (b) $P(X \leq 6)$.

   (c) $P(X \leq 15)$.

   (d) $P(7 \leq X \leq 15)$.

2. Suppose that $X \sim$ Bin(5000,0.0001). According to Minitab, $P(X \leq 2) = 0.9856$. Find the Poisson approximation to this probability. Compare your approximate answer with the exact answer and comment.

3. Let $X \sim$ Poisson(θ). Given $X = 2$, find the exact 95% upper confidence bound for θ. Apply your finding to create *the rule of 6.30 when* $X = 2$.

4. The data in Table 13.3 appeared in the Wisconsin State Journal on July 13, 2010, for accidents involving autos in Madison, Wisconsin.

   In parts (a)–(c), assume that the number of crashes of interest follows a Poisson Process with unknown rate λ per year. Use the data in the *Total* column to obtain the approximate 95% confidence interval estimate of λ.

   (a) Bike crashes.

   (b) Pedestrian crashes.

   (c) Fatal crashes. Also obtain the exact confidence interval.