Module C: Understanding the Data-Generating Process

Slide Deck C3:

Poisson Distributions

The section in which we learn about a third named discrete distributions. When a distribution is encountered frequently, we name it and try to calculate formulas for its parameters. A random variable that follows a Poisson distribution will model the number of successes over a time or space.

Today's Objectives

By the end of this slidedeck, you should

1. determine what random variables follow a Poisson distribution using its definition
2. calculate probabilities from a Poisson distribution
3. calculate the expected value, variance, median, and probabilities associated with a Poisson random variable

Definition of Poisson Experiment

Definition

The **Poisson distribution** is used to model a random variable that is the count of successes over an area or a time period.

Examples

- heads flipped in an hour
- number of dents on a car
- errors on a page
- number of terrorist attacks in a year
- bacteria in a swimming pool
- influenza cases in a week
- wars in a year

Poisson Probability Mass Function

Recall that the probability mass function (pmf) provides the probability of each element of the sample space. For a Poisson random variable, there are an infinite number of possible outcomes:

$$S = \{0, 1, 2, \ldots\}$$

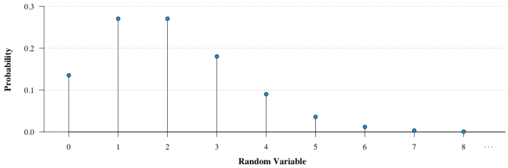
Using mathematics (Calculus II), it can be proven that the probability mass function is

$$\mathbb{P}[\ X = x\ ] = \begin{cases} \dfrac{e^{-\lambda}\ \lambda^x}{x!} & x \in S \\ 0 & \text{Otherwise} \end{cases}$$

The parameter λ (lambda) represents the average rate (successes per time period, per area, per volume, etc.).

Poisson pmf

The graph of the Poisson($\lambda = 2$) distribution looks like this:



Poisson Parameters

For a Poisson:

$$\mathbb{E}\left[X\right] = \lambda$$

$$\mathbb{V}\left[X\right] = \lambda$$

I state these without proof. If you would like the proofs, come see me. The proofs only require summations, but I am not sure they are helpful in understanding statistics.

Poisson Example 1: Fair Coins

Example

Let X be the number of heads flipped in a minute, given that I average 24 flips per minute. If the coin is fair, then it is clear that $X \sim \mathcal{Pois}\,(\lambda = 12)$.

1. What is the probability of getting no heads in that minute?
2. What is the probability of getting at most 20 heads in that minute?
3. What is the probability of getting at least one head in the first 6 seconds?

Poisson Example 1: Fair Coins

Example

1. What is the probability of getting no heads in that minute?

We are asked to calculate $\mathbb{P}[\ X = 0\ ]$. This is a simple application of the pmf:

$$\mathbb{P}[\ X = x\ ] = \frac{e^{-\lambda}\ \lambda^x}{x!}$$

$$\begin{aligned}\mathbb{P}[\ X = 0\ ] &= \frac{e^{-12}\ 12^0}{0!} \\ &= \frac{e^{-12}\ 1}{1} \qquad = 0.000\,006\,144\end{aligned}$$

Poisson Example 1: Fair Coins

Example

1. What is the probability of getting no heads in that minute?

In R, this is

`dpois(0, lambda=12)`

Poisson Example 1: Fair Coins

Example

2. What is the probability of getting at most 20 heads in that minute?

We are asked to calculate $\mathbb{P}[\ X \leq 20\ ]$. This is a simple application of the pmf:

$$\mathbb{P}[\ X \leq 20\ ] = \sum_{x=0}^{20} \frac{e^{-12}\ 12^x}{x!}$$

$$= 0.9884023$$

In R, this is `ppois(20, lambda=12)`

Poisson Example 1: Fair Coins

Example

3. What is the probability of getting at least one head in the first 6 seconds?

We are asked to calculate $\mathbb{P}[\ Y \geq 1\ ]$, where $Y \sim \mathcal{P}ois\,(\lambda = 12/10)$:

$$\begin{aligned}\mathbb{P}[\ Y \geq 1\ ] &= \sum_{x=1}^{\infty} \frac{e^{-12/10}\ (12/10)^x}{x!} \\ &= 1 - \frac{e^{-12/10}\ (12/10)^0}{0!} \\ &= 0.6988058\end{aligned}$$

In R, this is `1 - dpois(0, lambda=12/10)`

Poisson Example 2: Interstate War

Example

I would like to estimate the probability of going five years without an interstate war. To do this, we need to know the *rate* of interstate wars in five years. From the Binomial slide deck (C2), we have that the probability of a war in a single year is 0.8462141. Thus, the average rate of wars in *one* year is $1/0.8462141 = 1.181734$.

That is the average rate for a single year. The average rate of wars in *five* years is just 5 times that:

$$5 \times 1.181734 = 5.90867$$

With this information, what is the probability that we have 5 years without a war?

Poisson Example 2: Interstate War

Solution:
To emphasize the steps involved, let's write this in bullet-form:

1. Let W be the number of wars in 5 years.
2. From the problem, this means $W \sim \mathcal{P}ois\,(\lambda = 5.90867)$
3. We are asked to calculate $\mathbb{P}[\ W = 0\ ]$.
4. In R, this is just
   `dpois(0, lambda=5.90867)`
   $= 0.002715796$

Poisson Example 2: Interstate War

Conclusion:

According to this calculation, the probability of going five years without an interstate war is *very* small (0.0027). In fact, this suggests we would expect to go five years without war once every 400 years or so.

Question: How reasonable is this result?

Poisson Example 3: The Ferris Building

Example

Jonathan, a former student of mine was looking for ideas for his statistics research project. I was able to match him with a real estate developer in Galesburg who wanted to determine if it would be profitable to renovate the Ferris Building downtown into a boutique hotel.

One aspect of determining this was to estimate the average residency in the hotel. After some extensive research into minor college towns and hotel/motel occupancy, Jonathan estimated the average number of rooms occupied would be about 150 per week.

The developer calculated that it would take a minimum of 130 rooms per week to turn a profit. With this information, what proportion of weeks will *not* be profitable?

Poisson Example 3: The Ferris Building

Solution:
To emphasize the steps involved, let's write this in bullet-form:

1. Let us define C as the number of customers in a given week.
2. The problem tells us $C \sim \mathcal{Pois}\,(\lambda = 150)$.
3. We are asked to calculate $\mathbb{P}[\ C < 130\ ]$.
   - This calculation is equivalent to $\mathbb{P}[\ C \leq 129\ ]$.
4. The calculation using R is

   ```
   ppois(129, lambda=150)
   ```

   $= 0.04453316$

Poisson Example 3: The Ferris Building

Conclusion:

According to this calculation, the expected proportion of weeks in a year that the Ferris Building will not turn a profit is 4.45%.

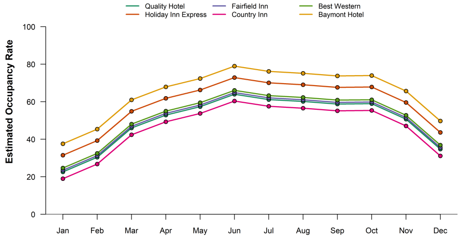
Be Aware: This estimate is based on the hotel occupancy rates over the past five years.

Statistics Question: How does this raise the question of whether the sample is representative?

Modeling Note: It is quite clear that the probability of a room being rented in a hotel is *not* independent of the time of year. There are periods where the occupancy rate is higher than others.

The next graphic illustrates this last point.

Poisson Example 3: The Ferris Building



Poisson Example 4a: Galesburg Crime

Example

In conjunction with the previous example, we would like to develop a system to determine if the crime rate in Galesburg has significantly increased. To do this, we will compare the crime rate from last *year* with the number of crimes reported over the past *week*.

According to the US Department of Justice, the number of violent crimes reported in Galesburg in 2020 was just 96.

For the sake of using this information, let us assume that Galesburg experienced 4 violent crimes last week.

Does this suggest that there is a significant increase in the violent crime rate?

Poisson Example 4a: Galesburg Crime

Solution:

If V is the number of violent crimes in a week, then we are asked to calculate $\mathbb{P}[\ V \geq 4\ ]$, given $V \sim \mathcal{Pois}\left(\lambda = 96/52\right)$. This is just

$$\begin{aligned}\mathbb{P}[\ V \geq 4\ ] &= 1 - \mathbb{P}[\ V < 4\ ] \\ &= 1 - \mathbb{P}[\ V \leq 3\ ] \\ &= \texttt{1 - ppois(3, lambda=96/52)} \\ &= 0.1162368\end{aligned}$$

Conclusion:
This probability is relatively large ($p = 0.116$). Thus, there really is not a lot of evidence of an increase in the violent crime rate. If the annual crime rate has not changed, then this result is not surprising.

Poisson Example 4b: Galesburg Crime

From a pedagogical perspective: I would like to extend this example and illustrate the importance of sample size on we understand the world around us.

Example

Let us continue the previous example. Now suppose that there were 16 violent crimes last month (the same rate, but over a longer time period).

With this new information, is there significant evidence of an uptick in the crime rate?

Note that the previous example had the same rate (4 per week) but with a shorter period of time (1 week). Here, we are measuring over a longer period of time… meaning we have more data now.

Poisson Example 4b: Galesburg Crime

Solution:

If X is the number of violent crimes in a month, then we are asked to calculate $\mathbb{P}[\ X \geq 16\ ]$, given $X \sim \mathcal{Pois}\,(\lambda = 96/12)$. This is just

$$\begin{aligned}\mathbb{P}[\ X \geq 16\ ] &= 1 - \mathbb{P}[\ X \leq 15\ ]\\ &= 1 - \sum_{x=0}^{15} \frac{e^{-96/12}\ (96/12)^x}{x!}\\ &= \texttt{1 - ppois(15, lambda=96/12)}\\ &= 0.008231011\end{aligned}$$

This probability is relatively small. Thus, there is evidence of an increase in the violent crime rate. That is, if the annual crime rate has not changed, then this result *is* surprising.

Poisson Example 4c: Galesburg Crime

Again, I would like to extend this example and illustrate the importance of sample size on we understand the world around us.

Example

Let us continue the previous example. Now suppose that there were 208 violent crimes this year (the same rate, but over a longer time period).

With this new information, is there significant evidence of an uptick in the crime rate?

Note that the previous examples had the same rate (4 per week) but with a shorter period of time (1 week and 4 weeks). Here, we are measuring over a longer period of time... meaning we have much more data now.

Poisson Example 4c: Galesburg Crime

Solution:

If X is the number of violent crimes in a year, then we are asked to calculate $\mathbb{P}[\ X \geq 208\ ]$, given $X \sim \mathcal{Pois}\,(\lambda = 96)$. This is just

$$\begin{aligned}
\mathbb{P}[\ X \geq 208\ ] &= 1 - \mathbb{P}[\ X \leq 207\ ] \\
&= 1 - \sum_{x=0}^{207} \frac{e^{-96}\ (96)^x}{x!} \\
&= \texttt{1 - ppois(207, lambda=96)} \\
&\ll 0.0001
\end{aligned}$$

This probability is definitely small. Thus, there is increible evidence of an increase in the violent crime rate. That is, if the annual crime rate has not changed, then this result *is* extremely surprising.

Today's Objectives

Now that we have concluded this lecture, you should be able to

1. determine what random variables follow a Poisson distribution using its definition
2. calculate probabilities from a Poisson distribution
3. calculate the expected value, variance, median, and probabilities associated with a Poisson random variable

Useful R Functions

In this slide deck, we covered (or hinted) on the following R functions related to the Poisson distribution:

- `dpois(x, lambda)` is the pmf, p for: $\mathbb{P}[\ X = x\ ] = p$
- `ppois(x, lambda)` is the CDF, p for: $\mathbb{P}[\ X \leq x\ ] = p$
- `qpois(q, lambda)` is the quantile function, x for: $\mathbb{P}[\ X \leq x\ ] = p$
- `rpois(n, lambda)` generates n random values from: $\mathcal{Pois}\,(\texttt{lambda=}\lambda)$

Please do not forget to access the `allProbabilities` document that provides all of the important probability functions in R.

Supplemental Activities

The following may be of interest to you in terms of today's topics:

- SCA 5a is for some discrete distributions

Note that you can access all Statistical Computing Activities here:
`https://www.kvasaheim.com/courses/stat200/sca/`

In addition to the SCA, **Laboratory Activity B** is helpful for learning how to handle discrete distributions (including the Binomial distribution). The lab actually shows the connection between sampling and discrete distributions. It uses three named distributions.
`https://www.kvasaheim.com/courses/stat200/labs/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of the Poisson distribution:

- Hawkes Learning: Section 5.3
- Intro to Modern Statistics: None
- R for Starters: Appendix A.7
- Wikipedia: Poisson Distribution