# Lesson 7: Poisson distribution

TB sections 3.4

Meike Niederhausen and Nicky Wakim

2025-10-27

# Learning Objectives

1. Calculate probabilities for different events using a Poisson distribution

# Where are we?

## Data
- Collecting data
- Categorical vs. Numeric
- Summary statistics
- Data visualization

## Probability
- Probability rules
- Independence, conditional
- Random variables and probability distributions
- Linear combinations
- Binomial, Normal, and Poisson

## Sampling Variability, and Statistical Inference
- Sampling distributions
- Central Limit Theorem
- Confidence Intervals
- Hypothesis tests

## Inference for continuous data/outcomes
- One sample t-test
- 3+ independent samples
- Simple linear regression / correlation
- 2 sample tests: paired and independent
- Power and sample size
- Non-parametric tests

## Inference for categorical data/outcomes
- One proportion test
- Fisher’s exact test
- Non-parametric tests
- Chi-squared test
- 2 proportion test
- Power and sample size

## R
- Basics
- Reproducibility
- Quarto
- Packages
- Data visualization
- Data wrangling
- R Projects
- …

# Introduction to the Poisson distribution

- The Poisson distribution is often used to model **count data (# of successes)**, especially for **rare events**
  - It is a **discrete distribution!**

- It is used most often in settings where **events happen at a rate** $\lambda$ per unit of population and per unit time

- **Example:** historical records of hospitalizations in New York City indicate that an average of 4.4 people are hospitalized each day for an acute myocardial infarction (AMI)
  - We can plot the distribution of hospitalizations on each day

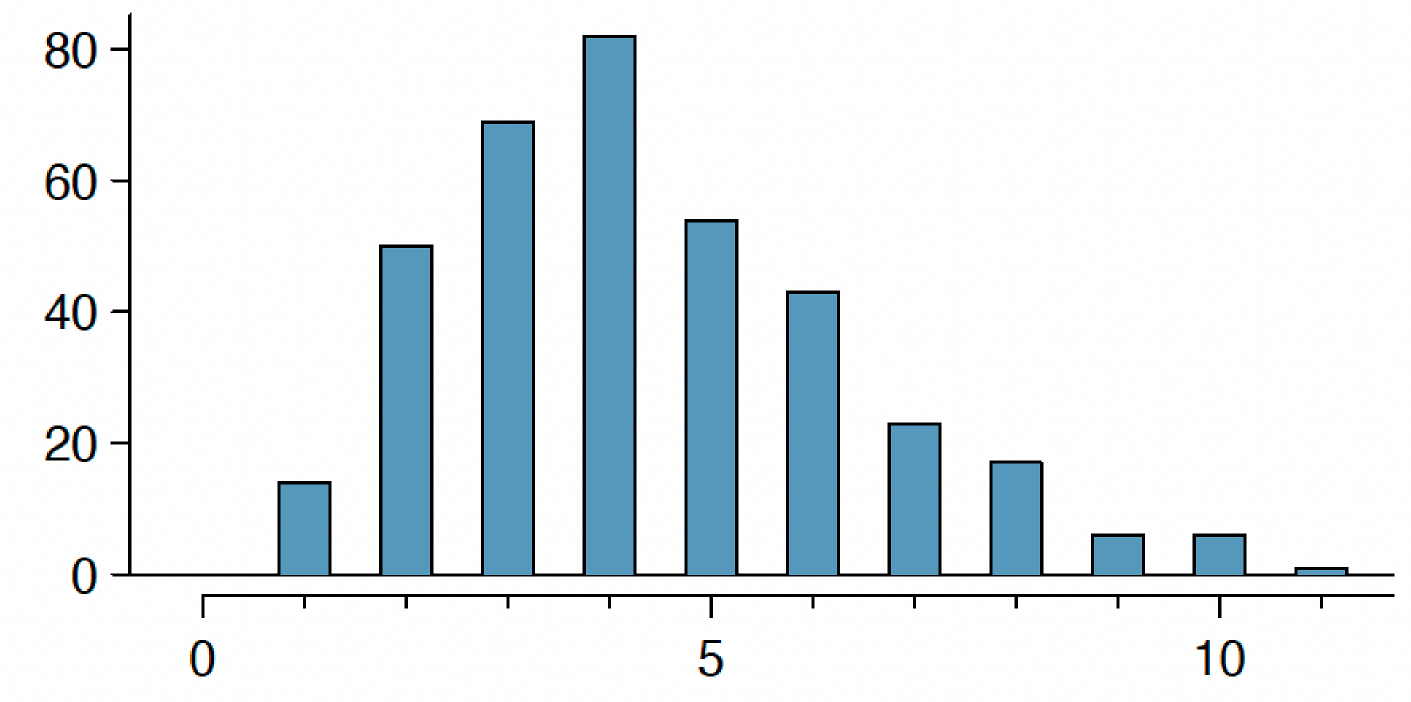



Figure 3.20: A histogram of the number of people hospitalized for an AMI on 365 days for NYC, as simulated from a Poisson distribution with mean 4.4.

# Poisson distribution

- Suppose events occur over time in such a way that
  1. The probability an event occurs in an interval is proportional to the length of the interval.
  2. Events occur independently at a rate $\lambda$ per unit of time.
- Then the probability of exactly $x$ events in one unit of time is

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!},\ \ k=0,1,2,\ldots$$

- For the Poisson distribution modeling the number of events in one unit of time:
  - The mean is $\lambda$.
  - The standard deviation is $\sqrt{\lambda}$.
- Shorthand for a random variable, $X$, that has a Poisson distribution:

$$X\sim \text{Pois}(\lambda)$$

# Poisson distribution: R commands

R commands with their input and output:

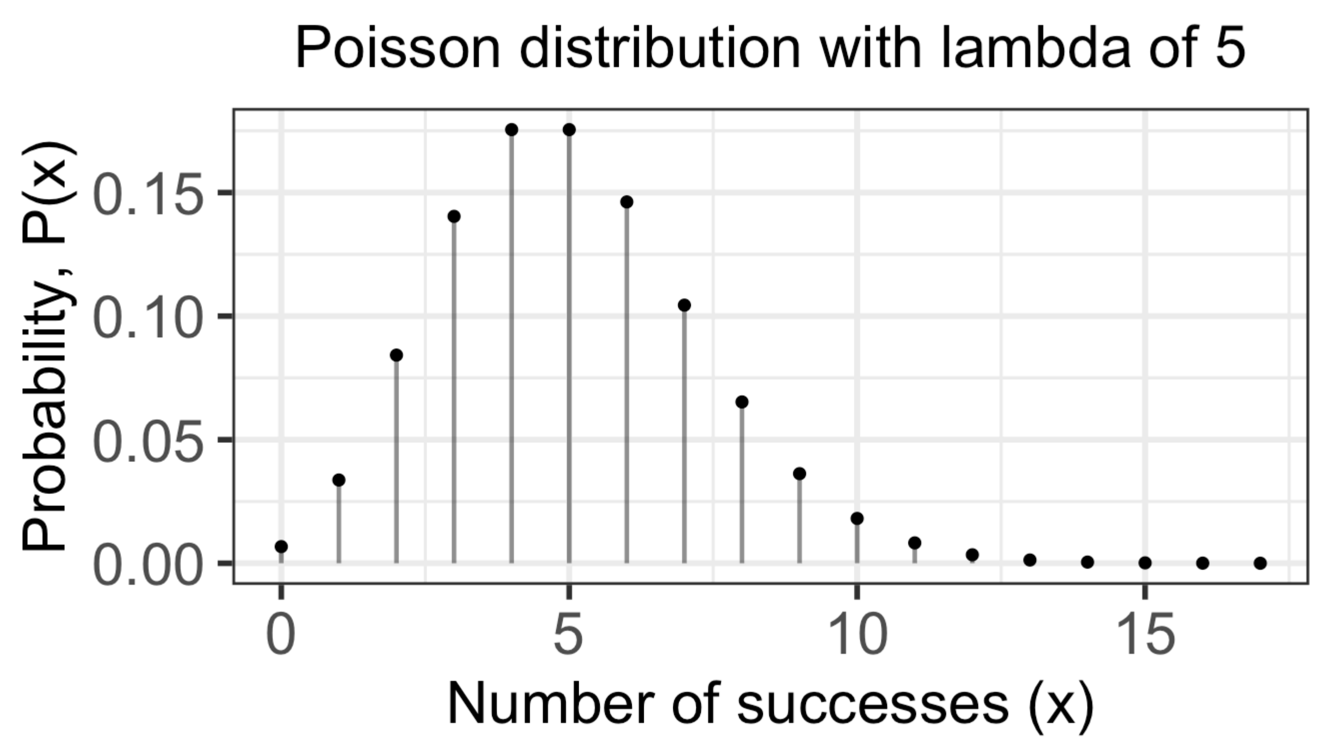
| R code | What does it return? |
|---|---|
| `rpois()` | returns sample of random variables with specified Poisson distribution |
| `dpois()` | returns value of probability density at certain point of the Poisson distribution |
| `ppois()` | returns cumulative probability of getting certain point (or less) of the Poisson distribution |
| `qpois()` | returns number of cases corresponding \| to desired quantile |

# Example: probabilities from a Poisson distribution (1/4)

## Typhoid fever

Suppose there are on average 5 deaths per year from typhoid fever over a 1-year period.

1. What is the probability of 3 deaths in a year?
2. What is the probability of 2 deaths in 0.5 years?
3. What is the probability of more than 12 deaths in 2 years?



# Example: probabilities from a Poisson distribution (2/4)

**Typhoid fever**

Suppose there are on average 5 deaths per year from typhoid fever over a 1-year period.

1. What is the probability of 3 deaths in a year?

- $\lambda = 5$ and we want $P(X = 3)$

$$P(X = 3) = \frac{e^{-5} 5^3}{3!} = 0.1404$$

```r
dpois(x = 3, lambda = 5)
```

```
[1] 0.1403739
```

# Example: probabilities from a Poisson distribution (3/4)

**Typhoid fever**

Suppose there are on average 5 deaths per year from typhoid fever over a 1-year period.

2. What is the probability of 2 deaths in 0.5 years?

$\lambda =?$ and we want $P(X = 2)$

- $\lambda = 5$ was the rate for one year. When we want the rate for half year, we need to calculate a new $\lambda$:
  - $\lambda = \frac{5 \text{ deaths}}{1 \text{ year}} \cdot \frac{1 \text{ year}}{2 \text{ half-years}} = \frac{2.5 \text{ deaths}}{1 \text{ half-year}}$

$$P(X = 2) = \frac{e^{-2.5} 2.5^2}{2!} = 0.0.2565$$

```
dpois(x = 2, lambda = 2.5)
```

```
[1] 0.2565156
```

# Example: probabilities from a Poisson distribution (4/4)

**Typhoid fever**

Suppose there are on average 5 deaths per year from typhoid fever over a 1-year period.

3. What is the probability of more than 12 deaths in 2 years?

$\lambda =?$ and we want $P(X > 12)$

- Need to calculate a new $\lambda$ for 2 years: $\lambda = \frac{5 \text{ deaths}}{1 \text{ year}} \cdot \frac{2 \text{ years}}{1 \text{ two-years}} = \frac{10 \text{ deaths}}{1 \text{ two-year}} = \frac{10 \text{ deaths}}{2 \text{ years}}$

$$P(X > 12) = 1 - P(X \leq 12) = 1 - \sum_{k=0}^{12} \frac{e^{-10} 10^k}{k!} = 0.2084$$

```r
1 - ppois(q = 12, lambda = 10)
```

```
[1] 0.2084435
```

```r
ppois(q = 12, lambda = 10,
      lower.tail = F)
```

```
[1] 0.2084435
```

# Poisson approximation of binomial distribution

- Poisson distribution can be used to approximate binomial distribution when $n$ is large and $p$ is small
  - When Normal approximation does not work
- Binomial distributions for different $n$ (columns) and $p$ (rows)