Lecture 17: The Poisson distribution

March 3, 2020

Today's objectives

- We have spent the last two lectures talking about how to handle a binary 0/1 outcome
- So far we've covered only the binomial distribution.
- Ch. 12 covers another distribution for counts: the Poisson distribution.
- The main distinction between Binomial and Poisson is that Poisson random variables have no upper bound, whereas the upper bound of a binomial random variable X was n, the size of the sample.
- Most commonly, the Poisson distribution is used to model rare events. However this is not the only case where we would use Poisson.

The Poisson distribution

- A Poisson distribution describes the count X of occurrences of a defined event in fixed, finite intervals of time or space when:

1. Occurrences are all independent (that is, knowing that one event has occurred does not change the probability that another event may occur- this was also true for our binomial distribution), and,
2. The probability of an occurrence is the same over all possible intervals of the same size.

Examples of the Poisson distribution

Rare, but infectious diseases. For example, the number of deaths X attributed to typhoid fever over a long period of time, say 1 year, follows a Poisson distribution if:

1. The probability of a new death from typhoid fever in any one day is very small.
2. The number of cases reported in any two distinct periods of time are independent random variables.

citation: https://ani.stat.fsu.edu/~debdeep/p4_s14.pdf

Examples of the Poisson distribution

Rare events occurring on a surface area. The distribution of number of bacterial colonies growing on an agar plate. The number of bacterial colonies over the entire agar plate follows a Poisson distribution if:

1. The probability of finding any bacterial colonies in a small area is very small.
2. The events of finding bacterial colonies in any two areas are independent.

citation: https://ani.stat.fsu.edu/~debdeep/p4_s14.pdf

Poisson probabilities

If X has the Poisson distribution with a mean number of occurrences per interval of μ, the possible values of X are 0, 1, 2, and so on. If k is any one of these values, then

$$P(X = k) = \frac{e^{-\mu}\mu^k}{k!}$$

- The above formula is the probability distribution function for a Poisson distribution.
- For example,

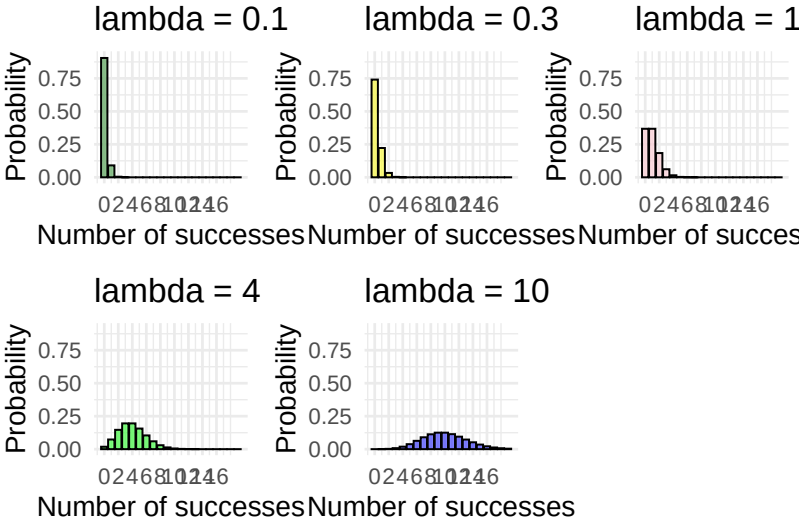
$$P(X = 2) = \frac{e^{-\mu}\mu^2}{2!}$$

will calculate the probability of observing two events for $X \sim Pois(\mu)$

Mean and SD of a Poisson random variable

- The mean of a Poisson random variable is equal to μ.
- The variance is also equal to μ, and thus the SD is equal to $\sqrt{\mu}$.
- When the mean is large, so is the SD, and this makes for a flat and wide probability distribution.
- Poisson distributions are most commonly used to describe rare, random events (or random events examined over small time intervals).
- In R, the function to calculate $P(X = x)$ for a binomial `dpois(x=?, lambda=?)`, where `lambda` (λ) is equal to the average μ (which is the book's notation).

Probability distribution of a Poisson random variable



Example: Mumps

In Iowa, the average monthly number of reported cases of mumps per year is 0.1. If we assume that cases of mumps are random and independent, the number X of monthly mumps cases in Iowa has approximately a Poisson distribution with $\mu = 0.1$. The probability that in a given month there is no more than 1 mumps case is:

$P(X \leq 1) = P(X = 0) + P(X = 1)$

$= \frac{e^{-0.1}0.1^0}{0!} + \frac{e^{-0.1}0.1^1}{1!}$ (note that $0! = 1$, by definition, and x^0=1, for any value of x.)

$= 0.9048 + 0.0905 = 0.9953$

Thus, we expect to only see 0 cases 90.5% of the months, and 1 case 9.05% of the time.

Example: Mumps calculated using R using `ppois()` and `dpois()`


```r
ppois(q = 1, lambda = 0.1) # notice that lambda is the parameter
```

```
## [1] 0.9953212
```

```r
# or,
dpois(x = 0, lambda = 0.1) + dpois(x = 1, lambda = 0.1)
```

```
## [1] 0.9953212
```

Example: Mumps, continued

Suppose you saw 4 cases of Mumps in a given month. What are the chances of seeing 4 or more cases in any given month?

```
1 - ppois(q = 3, lambda = 0.1) #careful, we used q = 3 here, why 3 and not 4?
```

```
## [1] 3.846834e-06
```

Could you have performed this calculation using `dpois()`?

If you saw 4 or more cases in any given month, this is very unlikely under this model. This suggests a substantial departure from the model, suggesting a contagious outbreak (no longer independent)

Example: Polydactyly

In the US, 1 in every 500 babies is born with an extra finger or toe. These events are random and independent. Suppose that the local hospital delivers an average of 268 babies per month. This means that for each month we expect to see 0.536 babies born with an extra finger or toe at that hospital (how do you calculate 0.536 here?). Let X be the count of babies born with an extra finger or toe in a month at that hospital.

1. What values can X take?
2. What distribution might X follow?
3. Give the mean and standard deviation of X.

Example: Polydactyly, continued

To get a sense of what the data might look like, use R to simulate data across five years (60 months) for this hospital.

```
rpois(n = 12*5, lambda = 0.536)
```

```
##  [1] 1 0 2 1 1 0 0 1 0 1 0 1 0 2 0 0 0 1 0 1 0 0 0 1 1 0 0 0 0 1 0 2 3 0 0
## [36] 1 1 1 0 3 1 0 1 1 0 0 2 0 1 0 2 0 0 0 1 1 1 0 1 1
```

More random number generation

Examining a stream of Poisson-distributed random numbers helps us get a sense of what these data look like. Can you think of a variable that might be Poisson-distributed according to one of these distributions?

```
rpois(100, lambda = 0.1)
```

```
##   [1] 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
##  [71] 0 1 1 0 0 1 0 0 1 0 0 2 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
```

```
rpois(100, lambda = 0.5)
```

```
##   [1] 2 0 0 0 0 0 0 0 0 0 1 2 0 1 0 2 0 0 1 0 0 0 0 1 0 1 2 1 0 1 0 0 1 1
##  [36] 2 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 1 0 0 1 1 1 2 0 0 1 0 0 0 0
##  [71] 0 1 0 0 1 0 0 0 1 1 1 1 0 0 1 0 0 0 2 0 0 0 0 0 2 0 2 2 2 0
```

```
rpois(100, lambda = 1)
```

Comic Relief