Modeling Rates

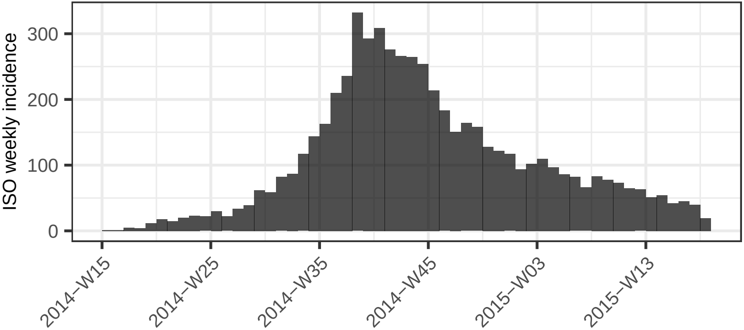
STAT 244NF: Infectious Disease Modeling

Marie Ozanne

9/14/2021

Motivating Example

Simulated Ebola Data set (`ebola_sim` from *outbreaks* package):
new cases per week in a single large city



Motivating Example

- What is the structure of the data we are considering? Continuous? Discrete?
- Why are the data entries considered rates?

Where we left off in Lab 1

In Lab 1, we spent some time working with the *incidence* package, which facilitated visualizing outbreaks through epidemic curves (think of the overall shape of the histogram as a curve). At the end of the lab, we briefly introduced a regression-based model.

Where we left off in Lab 1

The *incidence* package allows us to fit what they refer to as a log-linear model (more on this later).

```
epidemic_first_20 <- fit(ebola_incidence_weekly[1:20])
epidemic_first_20
```

```
## <incidence_fit object>
##
## $model: regression of log-incidence over time
##
## $info: list containing the following items:
##   $r (daily growth rate):
## [1] 0.03175771
##
##   $r.conf (confidence interval):
##           2.5 %     97.5 %
## [1,] 0.02596229 0.03755314
##
##   $doubling (doubling time in days):
## [1] 21.8261
##
##   $doubling.conf (confidence interval):
##         2.5 %   97.5 %
## [1,] 18.45777 26.69823
##
##   $pred: data.frame of incidence predictions (20 rows, 5 columns)
```

Epidemic

- "The increase and subsequent decrease in incidence following the (re)introduction of an infection in a population."
- "...the occurrence of cases in a given locality at a frequency which greatly exceeds what is expected."

Growth Rate (of an epidemic)

- The rate at which the prevalence of infectious people increases
- Typically calculated at the early stages of an epidemic

```
epidemic_first_20$info$r
```

```
## [1] 0.03175771
```

```
epidemic_first_20$info$r.conf
```

```
##            2.5 %     97.5 %
## [1,] 0.02596229 0.03755314
```

- How do we interpret this?

Doubling Time

- The time it takes for the number of infectious individuals to double

```
epidemic_first_20$info$doubling
```

```
## [1] 21.8261
```

```
epidemic_first_20$info$doubling.conf
```

```
##          2.5 %   97.5 %
## [1,] 18.45777 26.69823
```

- How do we interpret this?

Back to the "log-linear" model

- The *incidence* package fits a model that looks like:

$$\log(y) = r \times t + b$$

where y is a the number of new cases at a particular time, r is the (daily) rate of growth, t is the number of days since the start of an outbreak, and b is an intercept.

Back to the “log-linear” model

- What is this model? It is misleading to call is a “log-linear” model.
- How would we fit this model without the *incidence* package?

Poisson regression and the log-linear model

- In statistics, the log-linear model typically refers to a Poisson regression model, which is a generalized linear model.
- The simplest form of the model is:

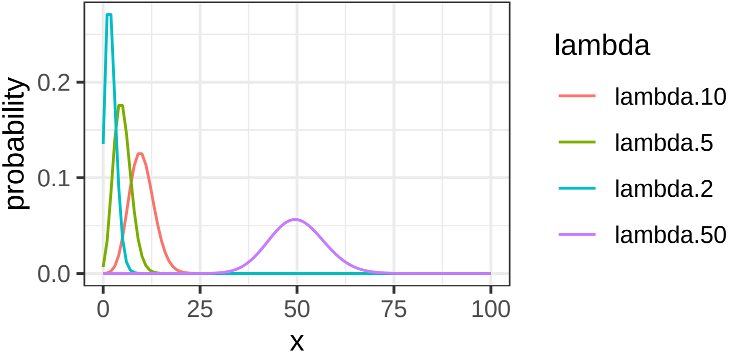
$$\log(\lambda) = \beta_0 + x \times \beta_1$$

where λ is the mean number of new cases (which we do not know, but can estimate with data), x is the time since the outbreak started, and β_1 is the growth rate.

We will spend more time with this model next class. For now, we need to understand the underlying distribution - the Poisson distribution.

Poisson Distribution

- Used to model discrete events that occur infrequently over time and space



Poisson Distribution

- X is a random variable representing the number of occurrences of some event of interest over a given interval.
- It can take on values between 0 and ∞.
- It has one parameter, called the rate, which we represent with λ (pronounced lambda). This rate also represents the mean number of occurrences of the event of interest in the interval of interest.
- The probability of a particular number of events happening in an interval is:

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Poisson Distribution

Assumptions

1. The probability that a single event occurs within an interval (of time, in our case) is proportional to the length of the interval.
2. Within an interval, theoretically infinitely many events could occur. There is no restriction to a fixed number of trials (as in the binomial distribution).
3. Events occur independently, both within the same interval, and between consecutive intervals.

Poisson Distribution

Calculating Poisson probabilities in R: `dpois`

- $P(X = 5|\lambda = 1) =$ `dpois(x=5, lambda=1)`

```r
dpois(x=5, lambda=1)
```

```
## [1] 0.003065662
```

Poisson Distribution

Calculating Poisson probabilities in R: `ppois`

- $P(X \leq 5|\lambda = 1) =$ `ppois(x=5, lambda=1)`

```
ppois(q=5, lambda=1, lower.tail=TRUE)
```

```
## [1] 0.9994058
```

(or equivalently)

```
1-ppois(q=5,lambda=1, lower.tail=FALSE)
```

```
## [1] 0.9994058
```

References

- E. Vynnycky and R. White.
  An Introduction to Infectious Disease Modelling. Oxford University Press. 2010.