

# The SIR Model

A Markov Chain model for disease spread

---

Jackson Curtis

December 1, 2017

Brigham Young University

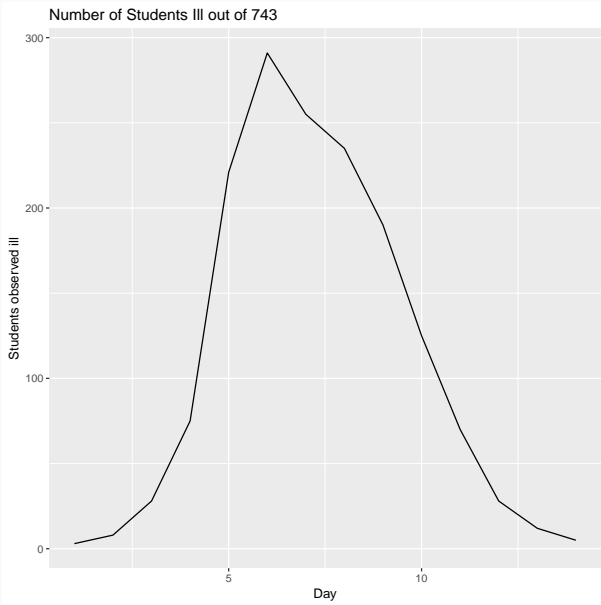
# The Model

The Susceptible-Infected-Recovered model describes how populations respond to disease.

$$\begin{bmatrix} p_{S,t} \\ p_{I,t} \\ p_{R,t} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & 1 \end{bmatrix} \begin{bmatrix} p_{S,t-1} \\ p_{I,t-1} \\ p_{R,t-1} \end{bmatrix}$$

The matrix in the middle is the transition matrix. The number in column  $i$ , row  $j$  is the probability of moving from state  $i$  to state  $j$  at each transition.

# Boarding School Data



# Deriving the Likelihood

Our fourteen observations are not independent, so we can write the likelihood as:

$$f(x_1, x_2, \dots, x_n | a_{12}, a_{32}) = f(x_1 | a_{12}, a_{32}) f(x_2 | x_1, a_{12}, a_{32}) \dots f(x_n | x_{n-1}, a_{12}, a_{32})$$

# Deriving the Likelihood

To address the fact that we didn't observe all states, we will sum over all possibilities. The R.V.  $R_i$  will represent the number of people recovered on day  $i$ .

$$f(x_i | x_{i-1}, a_{12}, a_{32}) = \sum_{j=0}^n f(x_i | x_{i-1}, a_{12}, a_{32}, R_i = j) \cdot Pr(R_i = j)$$

# Deriving the Likelihood

We will iteratively calculate  $R_i$  as we move through the algorithm:

$$Pr(R_i = r + h) = \sum_{j=0}^{x_{i-1}} Pr(H_i = j | R_{i-1} = r - j) \cdot Pr(R_{i-1} = r - j)$$

# Deriving the Likelihood

Now that we have  $R$ , we can calculate the probability of  $x_i$  given  $R_i$

$$f(x_i | x_{i-1}, a_{12}, a_{32}, r) = \sum_{h=0}^{x_{i-1}} \binom{x_{i-1}}{h} a_{32}^h (1-a_{32})^{x_{i-1}-h} \binom{n-x_{i-1}-r}{x_i-(x_{i-1}-h)} a_{12}^{x_i-(x_{i-1}-h)} (1-a_{12})^{n-r-x_i} I(x)$$

where  $I(x) = \begin{cases} 1 & \text{for } x_{i-1} - h \leq x_i \leq n - r \\ 0 & \text{otherwise} \end{cases}$

# Maximizing the Likelihood

The resulting likelihood involves a triple summation over time periods, possible values of  $R$ , and possible values of  $H$  (people healed), and two binomial densities. Needless to say, I didn't calculate the Hessian.

- Nelder-Mead constrained optimization (constraining both parameters between 0 and 1)
- Coded likelihood in C++ for faster evaluation

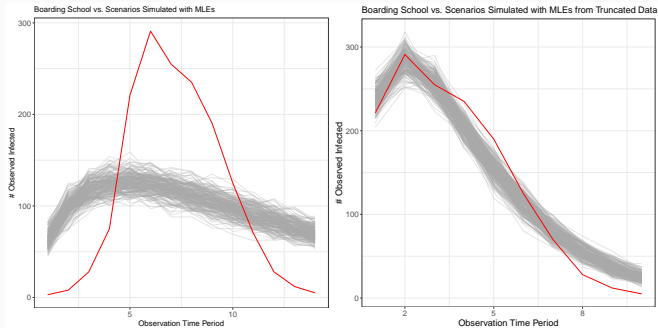


## So...Did it work?

500 Simulations of each scenario

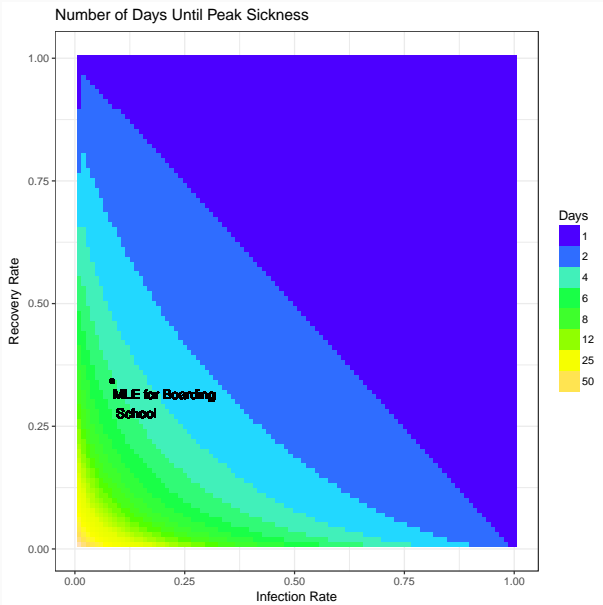
Pop.	No obs.	$a_{12}$	$a_{32}$	95% CI on $\hat{a}_{12}$	95% CI on $\hat{a}_{32}$
743	14	0.0848	0.3420	(0.0845, 0.0857)	(0.341, 0.345)
500	14	0.02	0.3	(0.020, 0.021)	(0.305, 0.319)
473	14	0.0848	0.3420	(0.084, 0.086)	(0.340, 0.345)
200	6	0.1	.3	(0.101, 0.104)	(0.301, 0.314)
100	14	0.2	0.7	(0.202, 0.206)	(0.700, 0.708)
100	14	0.7	0.2	(0.700, 0.707)	(0.222, 0.225)
100	14	0.5	0.5	(0.505, 0.513)	(0.534, 0.537)
50	20	0.1	0.3	(0.099, 0.104)	(0.282, 0.291)
20	8	0.4	0.3	(0.405, 0.422)	(0.304, 0.315)

# Lack of Fit

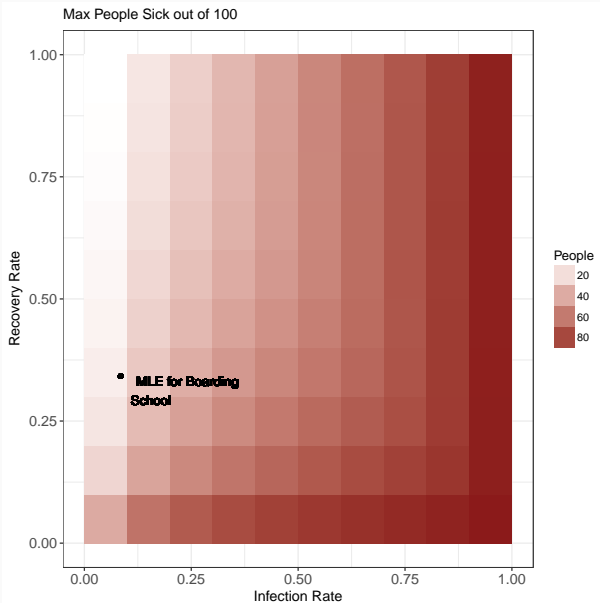


**Figure 1:** Actual data vs. 200 simulations using MLE as parameters for full dataset and dataset without first four days

# Inference



# Inference



The end