

# Statistical Methods in Infectious Disease Modeling

Kaitlyn Stocker

2017-04-20

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Modeling and Inference for Infectious Disease</b>	<b>1</b>
2.1	Deterministic Models . . . . .	1
2.2	Stochastic Models . . . . .	1
2.2.1	Chain Binomial . . . . .	1
2.2.2	TSIR . . . . .	1
2.2.3	Susceptible Reconstruction . . . . .	1
2.3	Spatial Models . . . . .	4
<b>3</b>	<b>Application to Measles</b>	<b>4</b>
<b>4</b>	<b>Results</b>	<b>4</b>

# 1 Introduction

## 2 Modeling and Inference for Infectious Disease

### 2.1 Deterministic Models

### 2.2 Stochastic Models

#### 2.2.1 Chain Binomial

#### 2.2.2 TSIR

#### 2.2.3 Susceptible Reconstruction

In previous exercises, I had access to perfect and complete simulated data. While having such complete data is convenient, it is not realistic. Data gathered in real-world situations is much less inference-ready than the simulated data I have been using thus far.

Real data deviates from simulated data in a number of significant ways. For one thing, real data is incomplete. Not all cases of a disease are reported, so the number of infecteds at any given time point must be estimated using the number of reported cases multiplied by the reporting rate. There is typically no information about the true number of susceptible or recovered individuals, as collecting this information would be extremely impractical and cost-prohibitive.

In order to run any kind of meaningful inference on epidemic data, it is necessary to have at minimum the infected and susceptible dynamics over time. Using the reported cases and the rate at which cases are reported, it is easy enough to construct the infected class dynamics. However, reconstructing the susceptible class dynamics is not so straight-forward.

In order to reconstruct the susceptible class dynamics, let's first define the model. I will continue with the basic SIR model, but this time I am going to add in birth dynamics. The addition of birth dynamics into the susceptible class are crucial to the susceptible reconstruction process. To do this, I will define  $B_{t-d}$  as the number of births at time  $t - d$ . Since infants are born with natural immunity from their mothers, there is a time delay (denoted by  $d$ ) between when a baby is born and when it enters the susceptible class. The length of this delay is dependent on the disease. As before, I define the size of the infected class at a given time point  $t$  to be  $I_t \in \{1, \dots, T\}$ . Similarly, I define the size of the susceptible class at a given time point  $t$  to be  $S_t \in \{1, \dots, T\}$ . Equations 12 and 13 give the model specifications.

$$I_t = \beta S_{t-1} I_{t-1} \tag{1}$$

$$S_t = B_{t-d} + S_{t-1} - I_t \tag{2}$$

In equation 14 I allow  $I_t$  to be a product of the number of reported cases,  $C_t$  and  $\rho_t$ , the reporting rate at time  $t$ . I define  $\rho$  such that when  $\rho_t = 1$ , the number of true cases has been fully reported. When  $\rho_t > 1$ , the number

of true cases has been underreported. Additionally, I assume that  $\rho_t$  follows a probability distribution with  $E(\rho_t) = \rho$ .

$$I_t = \rho_t C_t \quad (3)$$

Substituting equation 14 into equation 13, we get:

$$S_t = B_{t-d} + S_{t-1} - \rho_t C_t \quad (4)$$

If we define  $E(S_t) = \bar{S}$ , then we can define a new variable  $Z_t$  such that  $S_t = \bar{S} + Z_t$ , with  $E(Z_t) = 0$ . In this way,  $Z_t$  is the deviations from the mean of  $S_t$ .  $Z_t$  therefore follows the same recursive relationship as  $S_t$ , and can be defined as follows:

$$Z_t = B_{t-d} + Z_{t-1} - \rho_t C_t \quad (5)$$

If we allow  $Z_0$  to be the initial value of  $Z$ , we can rewrite the previous equation to look like the following:

$$Z_t = Z_0 + \sum_{i=1}^t B_{i-d} - \sum_{i=1}^t \rho_i C_i \quad (6)$$

To de-clutter this notation, allow  $Y_t = \sum_{i=1}^t B_{i-d}$  and  $X_t = \sum_{i=1}^t C_i$ . Additionally, we will assume a constant reporting rate. Now we can rewrite equation 17 as a simple linear regression equation:

$$Y_t = -Z_0 + Z_t + \rho X_t \quad (7)$$

Thus we have a linear regression equation relating cumulative births ( $Y_t$ ) to cumulative reported cases ( $X_t$ ). The susceptible dynamics  $Z_t$  are the regression remainder to equation 18, and can thus be fully reconstructed.

**Simulated Susceptible Reconstruction Example** In order to validate the accuracy of the susceptible reconstruction method taken from {citation}, I first used the method on simulated data. This way, I was able to compare the reconstructed susceptible dynamics to the true, simulated dynamics.

I first simulated an infection with transmission rate  $\beta = 1.66$  and recovery rate  $\gamma = 1/2.2$  using the previously defined TSIR model, with one modification. I added birth dynamics to allow the use of susceptible reconstruction. I simulated births using a birth rate of 12 births per 1,000 people annually {citation}.

Following the given method for reconstructing susceptible dynamics, I took the regression remainder from equation 18 and obtained  $Z_t$ . Figure 1 shows the reconstructed dynamics plotted alongside the true susceptible dynamics. The curves are equivalent, but as per definition  $Z_t$  is centered around zero.

In order to fully reconstruct the susceptible dynamics, it would be necessary to know  $\bar{S}$ . However, it is not possible to compute the mean number of susceptibles directly, as we are assuming we do not have any data about the true

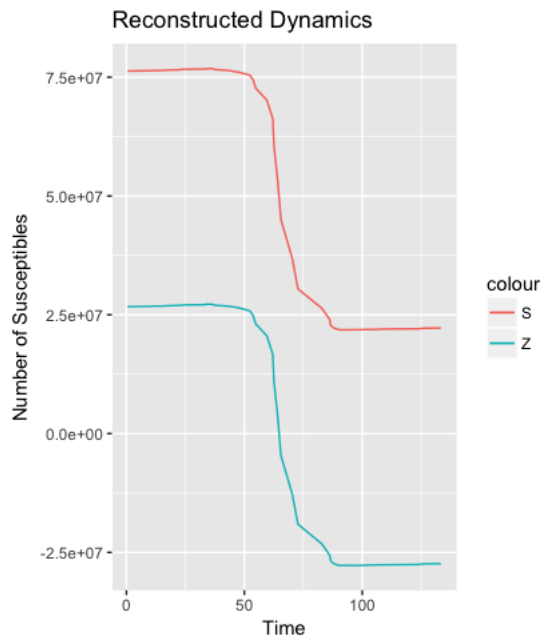


Figure 1: Comparison of reconstructed susceptible dynamics,  $Z$ , and the true susceptible dynamics  $S$  from the original simulation.

susceptible dynamics. Therefore, I had to instead infer  $\bar{S}$  along with the other model parameters ( $\beta$  and  $\gamma$ ) when I performed Bayesian Inference.

I defined the prior on  $\bar{S}$  to be normal with a mean of one fifth the initial population size and a standard deviation that was half of the mean. In other words, if we allow  $N_0$  to be the initial population size, then I defined the prior on  $\bar{S}$  as follows:  $\bar{S} \sim \text{Normal}(\frac{N_0}{5}, \frac{N_0}{10})$ . I then proceeded to replace any instance of  $S_t$  in the model definition with the equivalent expression  $\bar{S} + Z_t$ .

After running inference on my reconstructed susceptible dynamics, I took the upper and lower bounds from the 95% Credible Interval of  $\bar{S}$  and used those values to obtain upper and lower bounds on the true susceptible dynamics. 2 shows a plot of the true susceptible dynamics (from the simulation) contained within the upper and lower bounds of the 95% credible interval for the reconstructed dynamics. The true values of  $\beta$  and  $\gamma$  were also contained within their respective 95% credible intervals.

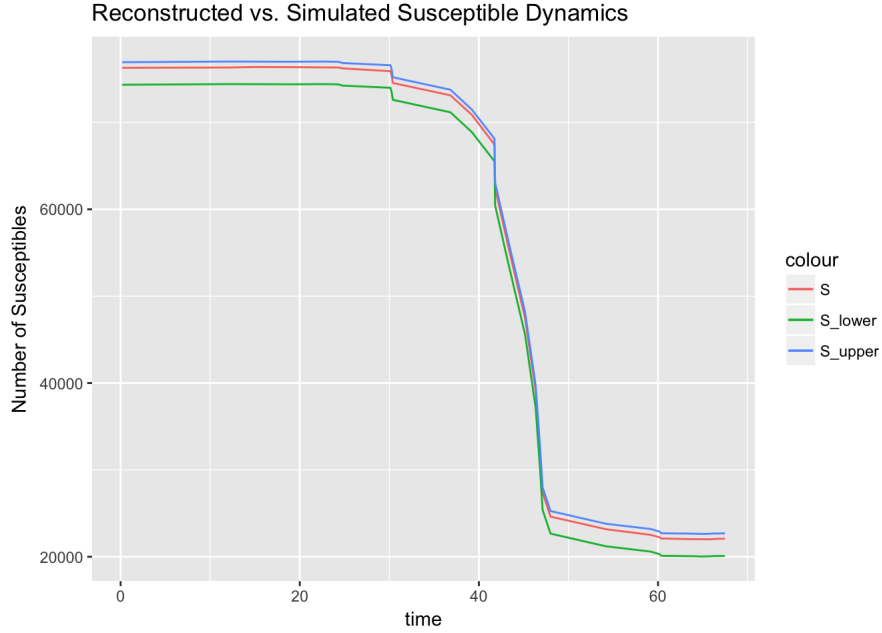


Figure 2: Plot of the true, simulated susceptible dynamics with the upper and lower estimated susceptible dynamics. The upper and lower bounds were obtained by taking the upper and lower limits of the 95% credible interval for the mean number of susceptibles, and adding each value to  $Z_t$ . The true number of susceptibles (in red) is clearly contained within the upper and lower bounds of the reconstructed susceptible dynamics.

### **2.3 Spatial Models**

## **3 Application to Measles**

## **4 Results**