

Introduction to Stochastic Compartmental Models

STAT 244NF: Infectious Disease Modeling

Deterministic models

So far in this class, we have been working with deterministic models. Recall from our introduction to compartmental models, deterministic models describe what happens “on average” in a population. Parameters (e.g., λ_t , π , ρ) are fixed, and model predictions are predetermined. In other words, there is no randomness. Even though λ_t changes based on the value of I_{t-1} , it will always be the same for each value of t , no matter how many times we run our simulation. Similarly, π and ρ will always be the same.

For example, consider the simulation from the SEIR model:

```
sim_SEIR <- SEIR_simulation(N=100, E0=0, I0=1, R0=0, R_0=2, pD=3, D=2, Time=60)
head(sim_SEIR$SEIR_df, 10)
```

##	time	S	E	I	R	lambda_t
## 1	0	99.00000	0.0000000	1.0000000	0.000000	NA
## 2	1	98.01493	0.9850665	0.5000000	0.500000	0.009950166
## 3	2	97.52608	1.1455625	0.5783555	0.750000	0.004987521
## 4	3	96.96366	1.3261278	0.6710319	1.039178	0.005766862
## 5	4	96.31518	1.5325641	0.7775586	1.374694	0.006687855
## 6	5	95.56918	1.7677123	0.8996340	1.763473	0.007745434
## 7	6	94.71326	2.0343919	1.0390544	2.213290	0.008955994
## 8	7	93.73424	2.3352885	1.1976578	2.732817	0.010336749
## 9	8	92.61832	2.6727786	1.3772584	3.331646	0.011905145
## 10	9	91.35147	3.0487021	1.5795554	4.020275	0.013678176

The values above represent our predictions at the first nine days (plus the initial day), assuming a population of 100 individuals, 1 initial infectious individuals, and an infection with a basic reproductive number of 2, a pre-infectious period of 3 days, and an infectious period of 2 days.

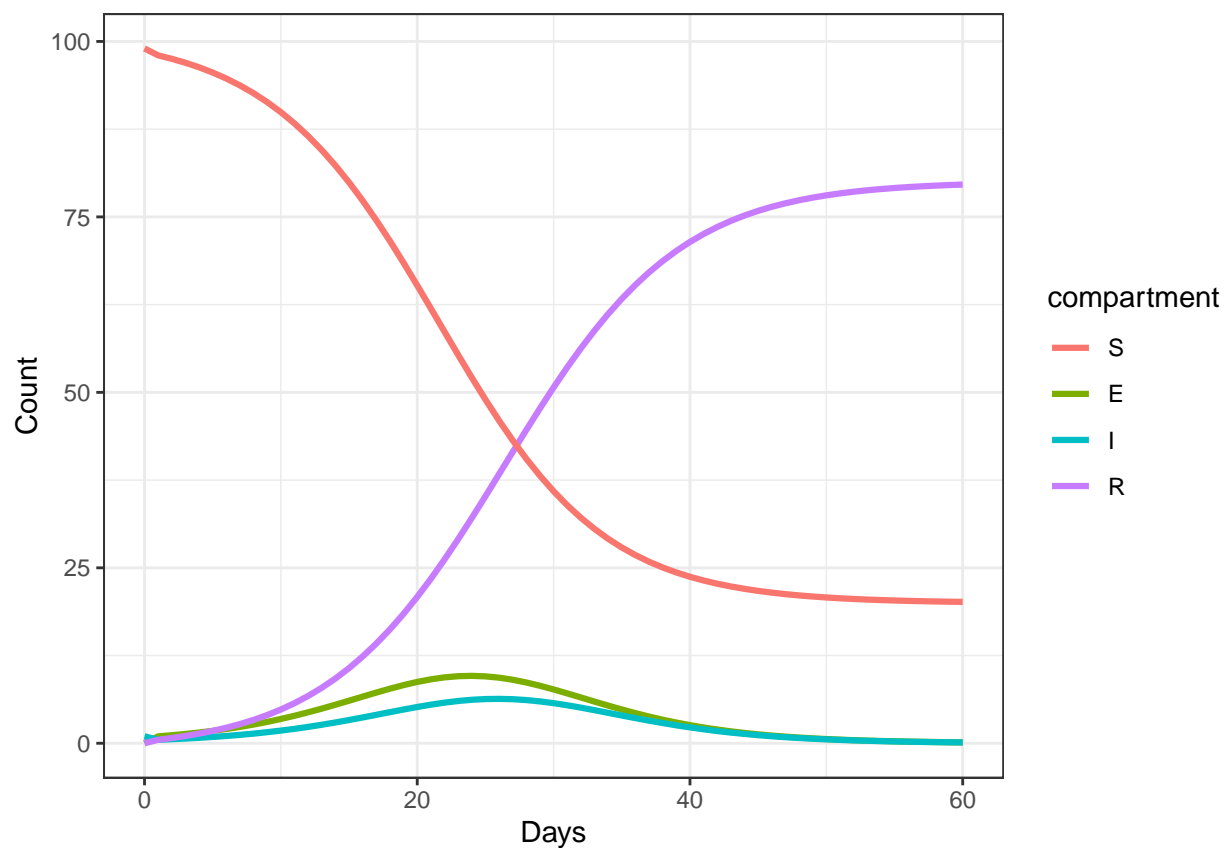
Do you think it is reasonable to predict that 1.25 individuals will be infectious at Day 7? If not, what would be more reasonable?

Do you think it is reasonable that the size of the outbreak will always be the same? If not, what would be more reasonable?

Difference Equations

$$\begin{aligned} S_t &= S_{t-1} - \underbrace{\lambda_t S_{t-1}}_{\text{not an integer}} \\ E_t &= E_{t-1} + \lambda_t S_{t-1} - \underbrace{\pi E_{t-1}}_{\text{not an integer}} \\ I_t &= I_{t-1} + \pi E_{t-1} - \underbrace{\rho I_{t-1}}_{\text{not an integer}} \\ R_t &= R_{t-1} + \rho I_{t-1} \end{aligned}$$

```
ggplot(data=sim_SEIR$df_ggplot, aes(x=time, y=count)) +  
  geom_line(aes(color=compartment), size=1.1) +  
  theme_bw() +  
  xlab("Days") +  
  ylab("Count") +  
  scale_colour_discrete(limits = c("S", "E", "I", "R"))
```



Problems with the deterministic models

1. Reasonable predictions for compartment counts should have only integer-valued - to be realistic, we have to have whole counts.

2. Random chance plays a role in outbreak size. We can conceivably have the same population and in one case we would get an outbreak, but in another we would not.

These problems can be remedied through the use of stochastic models.

Stochastic models

In stochastic models, two things are different than the deterministic models we were using. First, we assume that the number of individuals moving between compartments is a random variable that follows a particular probability distribution. Second, the number of individuals that move between compartments varies by chance. A range of outcomes are possible, which is more representative of the real world. These two qualities address the two problems listed above.

Difference Equations

$$\begin{aligned}S_t &= S_{t-1} - E_t^{(SE)} \\E_t &= E_{t-1} + E_t^{(SE)} - I_t^{(EI)} \\I_t &= I_{t-1} + I_t^{(EI)} - R_t^{(IR)} \\R_t &= R_{t-1} + R_t^{(IR)}\end{aligned}$$

Random Variables

Recall from your introductory statistics class, a *random variable* is a random process with a numerical outcome; usually represented by a capital letter like X, Y, Z.

Here we are going to be more specific in our notation for each random variable:

- $E_t^{(SE)}$: number of newly pre-infectious individuals at time t
- $I_t^{(EI)}$: number of newly infectious individuals at time t
- $R_t^{(IR)}$: number of newly recovered individuals at time t

Each random variable follows some probability distribution. What are some probability distributions that we know?

Which of these distributions will have integer outcomes?

- $E_t^{(SE)} \sim$
- $I_t^{(EI)} \sim$
- $R_t^{(IR)} \sim$

Example

Consider an SEIR model with the following:

- $N = 50$ (population is 50)
- $S(0) = 49$ (49 people are susceptible to start)
- $E(0) = 0$
- $I(0) = 1$ (1 person is infectious to start)
- $R(0) = 0$
- $R_0 = 2$ (the basic reproductive number is 2)
- $D' = 2$ days (the pre-infectious period is 2 days); $\pi = 1/D'$
- $D = 2$ days (the infectious period is 2 days); $\rho = 1/D$
- Time=15 days (total number of days we are observing this for purposes of this example)

For the values listed above, follow these steps for $t = 1, \dots, 15$:

1. Calculate the risk of infection in the next time interval using $\lambda_t = 1 - \exp\left\{-c_e \frac{I_{t-1}}{N}\right\}$.
2. Randomly sample $E_t^{(SE)} \sim \text{Binomial}(S_{t-1}, \lambda_t)$. Update your value for S_t . You can use `rbinom(n=1, size=S[t-1], prob=lambda_t)`. Add your point to the plot below.
3. Randomly sample $I_t^{(EI)} \sim \text{Binomial}(E_{t-1}, \pi)$. Update your value for E_t (you can use `rbinom(n=1, size=E[t-1], prob=pi)`). Add your point to the plot below.
4. Randomly sample $R_t^{(IR)} \sim \text{Binomial}(I_{t-1}, \rho)$. Update your values for I_t and R_t , and add the points to your plot below.

Once you are done, add your points to the plots on the blackboard.

