2014 West African Ebola Epidemic: Model Development
Summer 2014

Grant Brown

1 The SEIR Compartmental Epidemic Model

1.1 Introduction

Compartmental epidemic modeling is a flexible and extensible method of describing epidemic behavior. Such techniques rely on the idea that individuals within a population undergoing an epidemic process can be categorized by disease state. The most common terms used to describe these disease states are:

- **Susceptible**: Individuals capable of contracting the disease of interest.
- **Exposed**: Individuals who have contracted the disease, but are not yet infectious.
- **Infectious**: Individuals who are capable of spreading the disease.
- **Recovered/Removed**: Individuals who have either recovered or been removed from the population.

The stochastic SEIR model has been successfully applied in the past to Ebola by Lekone and Finkenstädt (2006), making it a good candidate model family for the current outbreak. The development here is broadly similarly to the aforementioned paper, "Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study", with the addition of a simple conditionally autoregressive spatial correlation structure.

2 Model Development

2.1 Compartments and Notation

Denote the spatial locations of interest $\{s_i : i = 1, ..., n\}$
Let $d(s_i, s_l) = d_{il}$ be an indicator function which denotes whether two spatial locations share a border. spatial locations. Note that $d(s_i, s_i) = 0$, and that for this data analysis all locations share a border. Should this change as the epidemic progresses

(by spreading to additional nations), only small changes will be required.

Let time (in units appropriate to the data and disease process) be denoted $t_j : j = 1, ..., T$ Let each of these discrete time points be associated with an offset term which describes the relative amount of continuous time over which the discretely indexe data was aggregated. Here, this offset is the time between WHO reports. Define the following components for each s_i and t_j:

All components indexed by time and space in libspatialSEIR are layed out as arrays of double precision elements in column major order with T rows and n columns.

- N_{ij} is the population size.
- S_{ij} is the count of susceptible individuals.
- E_{ij} is the count of exposed individuals.
- I_{ij} is the count of infectious individuals.
- R_{ij} is the count of recovered/removed individuals.
- E^*_{ij} is the number of newly exposed individuals.
- I^*_{ij} is the number of newly infectious individuals. We assume that the observed data gives an accurate count of this value.
- R^*_{ij} is the number of newly recovered/removed individuals.

Let $\mathbf{N_j} = \mathbf{S_j} + \mathbf{E_j} + \mathbf{I_j} + \mathbf{R_j}$ for all j rows of the disease process compartments. As population size is not expected to change dramatically over the time scale of this epidemic, N is held fixed for each spatial location. In addition let $\mathbf{S_0}, \mathbf{E_0}, \mathbf{I_0}$, and $\mathbf{R_0}$ denote the n-vectors of unknown compartment sizes at the start of the modeling period. Here, I_0 is initialized at 86 in Guinea and 0 elsewhere, as this is the first available estimate of the number of infectious individuals. The rest of the population was considered susceptible, as is usual for these models.

2.2 Disease Evolution Process Model

Given the values of the aforementioned parameters, the disease process evolves forward in time as one would expect based on the definitions.

$$\mathbf{S_{j+1}} = \mathbf{S_j} - \mathbf{E_j^*}$$
$$\mathbf{E_{j+1}} = \mathbf{E_j} - \mathbf{I_j^*} + \mathbf{E_j^*}$$
$$\mathbf{I_{j+1}} = \mathbf{I_j} - \mathbf{R_j^*} + \mathbf{I_j^*}$$
$$\mathbf{R_{j+1}} = \mathbf{R_j} + \mathbf{R_j^*}$$

While models of this form are often fit using deterministic systems of ordinary or partial differential equations, libspatialSEIR uses a heirarchical Bayesian framework in order to adequately capture the inherent variability in the model parameters. To complete the temporal process model, specify the following chain binomial relationship:

$$\{E_{ij}^* | \pi_{ij}^{(SE)}, S_{ij}\} \overset{ind}{\sim} binom(S_{ij}, \pi_{ij}^{SE})$$

$$\{I_{ij}^* | \pi^{(EI)}, E_{ij}\} \overset{ind}{\sim} binom(E_{ij}, \pi^{(EI)})$$

$$\{R_{ij}^* | \pi^{(IR)}, I_{ij}\} \overset{ind}{\sim} binom(I_{ij}, \pi^{(IR)})$$

2.3 Transition Probability Model

While π^{EI} and π^{IR} can be easily parameterized with the usual exponential time assumption, more care must be given to the development of a model for the $\left\{\pi_{ij}^{SE}\right\}$. This set of parameters describes the actual infection process and must account for predictor variables as well as the spatial structure of $\{s_i\}$.

2.4 Infection Process - CAR Model Motivation

Define δ_{ij} to be the proportion of persons who are infectious in spatial unit s_i at time t_j. Then, letting $\rho f(d_{il})$ denote the scaled standardized neighborhood matrix component for spatial locations s_i and s_l, we can derive (with minimal further assumptions):

$$\pi_{ij}^{(SE)} = 1 - exp\left\{ -\delta_{ij} e^{\theta_i} - \rho \sum_{\{l \neq i\}} (f(d_{il}) \delta_{il} e^{\theta_l}) \right\}$$

3 Basic Reproductive Number

The basic reproductive number, $\mathcal{R}_0$, is an important quantity in epidemiology. While the interpretation must be adapted to the problem of interest, in general terms the basic reproductive number captures the expected number of secondary infections produced by a single infected individual in an entirely susceptible population.

Using the next generation matrix approach to $\mathcal{R}_0$ calculation, we first define the matrix G such that $G_{i,l}(t_j)$ is the expected number of infections in spatial location s_i caused by a single infected individual in location s_l at time t_j.

With this matrix constructed, the basic reproductive number can be immediately calculated as the dominant eigenvalue.

4 Final Notes

The sampling scheme uses variants of Metropolis sampling (not Metropolis-Hastings, as symmetric proposal distributions are employed) are used, in combination with automatic tuning capabilities which enable libspatialSEIR to choose tuning parameters with reasonable acceptance rates.