Bayesian Modelling of Epidemics

Part II: Population-Averaged Models

Dr. Rob Deardon[1,2] **& Dr. Caitlin Ward**[3]

Department of Mathematics & Statistics, University of Calgary[1]
Faculty of Veterinary Medicine, University of Calgary[2]
Division of Biostatistics & Health Data Science, School of Public Health, University of Minnesota[3]

rob.deardon@ucalgary.ca
ward-c@umn.edu

ENAR, Baltimore, March 2024

UNIVERSITY OF CALGARY

UNIVERSITY OF MINNESOTA

Outline

Outline

Brief Introduction to R NIMBLE

Before we get to the stochastic models, we will briefly illustrate the software we will use for implementation throughout Part II of the workshop

`NIMBLE` is an R package for fitting statistical models with Bayesian MCMC methods

Models are defined using BUGS syntax and compiled with C++

NIMBLE offers many advantages for fitting Bayesian models:

- Computational speed
- Open-source software and BUGS language increases usability
- Flexibility - users can define their own distributions and samplers

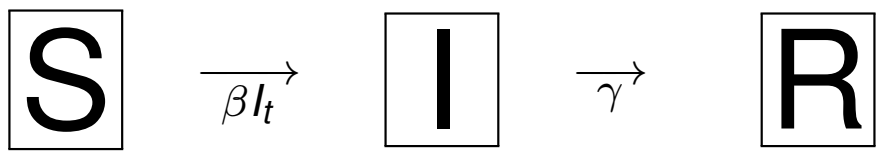
File: “RMarkdown Part2a.Rmd"

Outline

Model Structure

So far, we have considered **deterministic** ODE models

For example, the $\mathcal{SIR}$ model, where individuals are either susceptible, infectious, or removed from the population (either recovered and immune, or dead).



where:

S_t is the **proportion** of susceptible individuals in the population at time t

I_t is the **proportion** of infectious individuals in the population at time t

R_t is the **proportion** of removed individuals in the population at time t

Note: $S_t + I_t + R_t = 1$ for all t

Stochastic Population-Averaged Model

The model can be made stochastic by probabilistically modeling the transitions between compartments

In the population-averaged model, S_t, I_t, and R_t denote the **count** of individuals in the indicated compartment at time $t = 0, ..., \tau$

In discrete time, transitions between compartments are considered according to the set of difference equations

$$\begin{aligned} S_{t+1} &= S_t - I_t^* \\ I_{t+1} &= I_t + I_t^* - R_t^* \\ R_{t+1} &= R_t + R_t^* \end{aligned}$$

I_t^* and R_t^* represent the number of individuals that transition into the indicated compartment in $(t, t+1]$

$S_t + I_t + R_t = N$ for all t, where N is the population size

Compartment Membership over Time

Consider a population with $N = 100$ and initial conditions $S_0 = 99$ and $I_0 = 1$

t	$\boldsymbol{S}$	$\boldsymbol{I}^*$	$\boldsymbol{I}$	$\boldsymbol{R}^*$	$\boldsymbol{R}$
0	99	0	1	0	0
1	99	2	1	0	0
2	97	3	3	0	0
3	94	4	6	3	0
4	90	1	7	2	3
5	89	2	6	0	5
6	87	3	8	2	5
7	84	4	9	2	7
8	80	11	11	3	9
9	69	9	19	3	12
10	60	0	25	0	15

Given a set of initial conditions, the population size, and the transition vectors $\boldsymbol{I}^*$ and $\boldsymbol{R}^*$, the compartment memberships over time ($\boldsymbol{S}$, $\boldsymbol{I}$, $\boldsymbol{R}$) are fully determined

Model Structure

The transitions between compartments are described using the “chain binomial" model

$$\begin{aligned} I_t^* &\sim Bin\left(S_t, \pi_t^{(SI)}\right) \\ R_t^* &\sim Bin\left(I_t, \pi^{(IR)}\right) \end{aligned}$$

Transmission probability is derived from a Poisson contact process and homogeneous mixing. Proof is on the next slide.

$$\pi_t^{(SI)} = 1 - \exp\left(-\beta \frac{I_t}{N}\right)$$

The removal probability assumes the length of time infectious is exponentially distributed

$$\pi^{(IR)} = 1 - \exp\left(-\gamma\right)$$

Proof: let X denote the time an individual is infectious. $X \sim Exp(\gamma)$,

$$\pi^{(IR)} = P(X \leq s+1 | x > s) = P(X \leq 1) = 1 - e^{-\gamma}$$

Derivation of Transmission Probability

Let $Y_t \in \{0, 1\}$ indicate a transmission event between two individuals at time t

Person A and Person B contact each other

Person B is infectious

Person B successfully transmits the disease to Person A

Let $K \sim Pois(\lambda)$ describe the number of contacts each individual in the population has each day

At time t, the probability of contacting an infectious person is $\delta_t = I_t/N$ (under homogeneous mixing)

Assume contacts are independent, so $P(Y_t = 0|K = k) = (1 - \delta_t p)^k$, where p is the probability of transmission from an infectious person to a susceptible one given an epidemiologically significant contact

Derivation of Transmission Probability

At any time,

$$P(Y_t = 1) = 1 - P(Y_t = 0)$$

Then,

$$\begin{aligned}
P(Y_t = 0) &= E_Y[I(Y_t = 0)] = E_K\left[E_{Y|K}\left[I(Y_t = 0)|K = k\right]\right] \\
&= E_K\left[(1 - \delta_t p)^k\right] \\
&= \sum_{k=0}^{\infty}(1 - \delta_t p)^k \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \sum_{k=0}^{\infty} \frac{(\lambda(1 - \delta_t p))^k e^{-\lambda(1-\delta_t p)}}{k!} \frac{e^{-\lambda}}{e^{-\lambda(1-\delta_t p)}} \\
&= \exp(-\lambda(\delta_t p)) \\
&= \exp(-\beta I_t/N) \qquad\qquad \beta = \lambda p, \delta_t = I_t/N
\end{aligned}$$

So $P(Y_t = 1) = 1 - \exp(-\beta I_t/N)$

Reproductive Number

As with the deterministic setting, the basic reproductive number is approximately

$$R_0 = \beta/\gamma$$

In the stochastic setting, R_0 is the *expected* number of secondary infections caused by a single infectious individual in a completely susceptible population

Proof: on any given day the expected number of infections from a single infectious individuals is

$$N \times [1 - \exp(-\beta/N)] \approx \beta \qquad \text{for } x << 1, 1 - \exp(-x) \approx x$$

Infectious individuals remain infectious for an average of $1/\gamma$ days, so $R_0 = \beta/\gamma$

R_0 still has the same thresholding behavior:

- $R_0 \geq 1$ means the epidemic will continue
- $R_0 < 1$ means the epidemic will die out

Inference

Inference for stochastic models is typically made in the Bayesian framework

The Bayesian paradigm offers several advantages:

1. Inference based on a posterior distribution is highly interpretable

2. Epidemic data are often incomplete and imputation of missing data at each iteration of the Markov chain is straightforward

3. Incorporate prior knowledge about the disease process, which is often available

Likelihood and Priors

The log-likelihood for the chain binomial $\mathcal{SIR}$ model assuming infection and removal times are observed:

$$\ell(\mathbf{I}^*, \mathbf{R}^* | \Theta) = \sum_{t=0}^{\tau} \left\{ \log \binom{S_t}{I_t^*} + I_t^* \log \pi_t^{(SI)} + (S_t - I_t^*) \log(1 - \pi_t^{(SI)}) \right.$$
$$\left. + \log \binom{I_t}{R_t^*} + R_t^* \log \pi^{(IR)} + (I_t - R_t^*) \log(1 - \pi^{(IR)}) \right\}.$$

where $\Theta = \{\beta, \gamma\}$

Both β and $\gamma > 0$, so commonly used priors are gamma, half normal, or uniform

Often, informative prior information is available on γ as $1/\gamma$ is the mean length of the infectious period

File: “RMarkdown Part2b.Rmd"

Outline

Incomplete Data

The complete likelihood for the population-averaged $\mathcal{SIR}$ model assumes the counts of newly infectious and newly removed individuals over time are observed

More often, epidemic data are only partially observed

Consider the case where we observe only symptom onset dates or counts of cases detected each day, which can determine $\boldsymbol{I}^*$

Using data-augmented MCMC, $\boldsymbol{R}^*$ can be imputed at each iteration of the Markov chain

Estimating Missing Removal Times

Proposing $\boldsymbol{R}^*$ is challenging as the removal time vector is discrete and subject to constraints according to the nature of the epidemic:

- For a complete epidemic, the number of removals must equal the number of infections
- Removals must come after infections

A default sampler (e.g., Metropolis Hastings or slice sampling) would not account for this structure

Instead, a proposal which accounts for the nature of the epidemic must be used

Details of this proposal can be found in the appendix of these slides

Outline

Example: Ebola in Sierra Leone

Example:

We want to model the spread of Ebola Virus Disease (EVD) in Sierra Leone during the 2014-2015 epidemic in Western Africa.

Data provide symptom onset date for 11,903 confirmed and suspected EVD cases

We know the population size in Sierra Leone from census data.

EVD was first detected in Sierra Leone in May 2014

The United Nations Mission for Emergency Ebola Response (UNMEER) was initiated in early October 2014.

100% case isolation and safe burials were essentially achieved at the end of December 2014

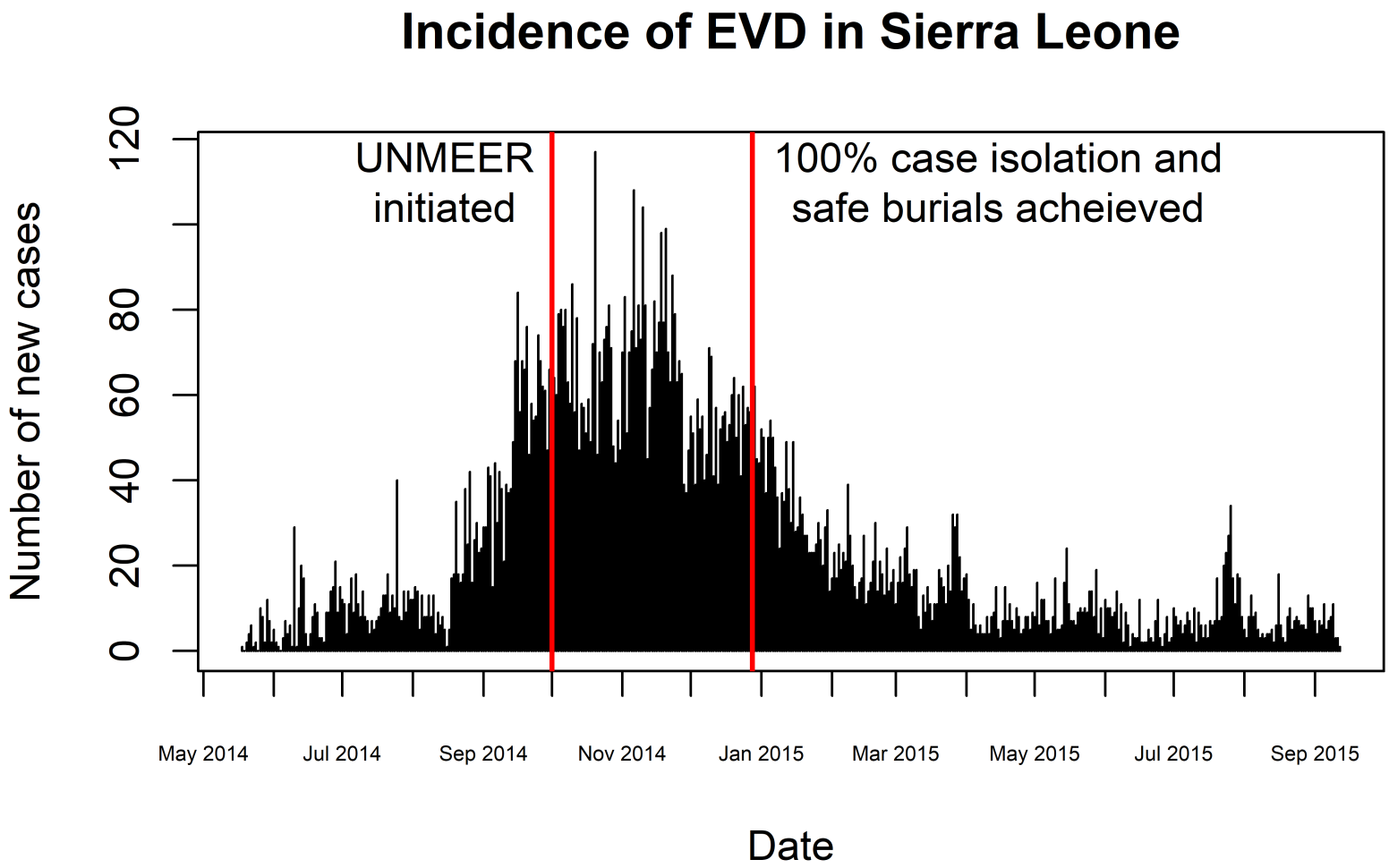
Ebola in Sierra Leone

Data available in `outbreaks` package in R and comes from Fang et al. (2016)

```
  id age sex    status date_of_onset date_of_sample district    chiefdom
1  1  20   F confirmed    2014-05-18     2014-05-23 Kailahun  Kissi Teng
2  2  42   F confirmed    2014-05-20     2014-05-25 Kailahun  Kissi Teng
3  3  45   F confirmed    2014-05-20     2014-05-25 Kailahun Kissi Tonge
4  4  15   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
5  5  19   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
6  6  55   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
```

```
      age          sex              status     date_of_onset
 Min.   : 0.0   F   :4719   confirmed:8358   Min.   :2014-05-18
 1st Qu.:16.0   M   :5109   suspected:3545   1st Qu.:2014-10-05
 Median :28.0   NA's:2075                   Median :2014-11-18
 Mean   :30.3                               Mean   :2014-12-04
 3rd Qu.:42.0                               3rd Qu.:2015-01-11
 Max.   :92.0                               Max.   :2015-09-12
 NA's   :956
 date_of_sample              district               chiefdom
 Min.   :2014-05-23   Western Urban:3165   W/Urban        :2274
 1st Qu.:2014-10-12   Port Loko    :1701   W/Rural        :1146
 Median :2014-11-24   Western Rural:1522   Freetowm       : 891
 Mean   :2014-12-10   Bombali      :1190   Bombali Sebora: 665
 3rd Qu.:2015-01-16   Kenema       : 780   Nongowa        : 527
 Max.   :2015-09-13   Bo           : 606   Waterloo       : 376
                      (Other)      :2939   (Other)        :6024
```

Ebola in Sierra Leone

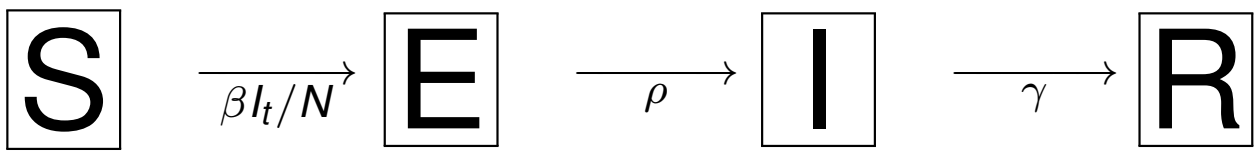


Adding Compartments

Once infected with EVD, an individual enters the latent period, which is generally between 8 to 10 days, but could be anywhere between 2 to 21 days

Individuals can only transmit EVD to others after symptoms have appeared, and remain infectious for around 4 to 10 days

To incorporate the latent period, we can an an exposed compartment and use the $\mathcal{SEIR}$ model



$$
\begin{aligned}
E_t^* &\sim Bin\left(S_t, \pi_t^{(SE)}\right), \quad \pi_t^{(SE)} = 1 - \exp\left(-\beta_t \frac{I_t}{N}\right) \\
I_t^* &\sim Bin\left(E_t, \pi^{(EI)}\right), \quad \pi^{(EI)} = 1 - \exp(-\rho) \\
R_t^* &\sim Bin\left(I_t, \pi^{(IR)}\right), \quad \pi^{(IR)} = 1 - \exp(-\gamma)
\end{aligned}
$$

Adding Compartments

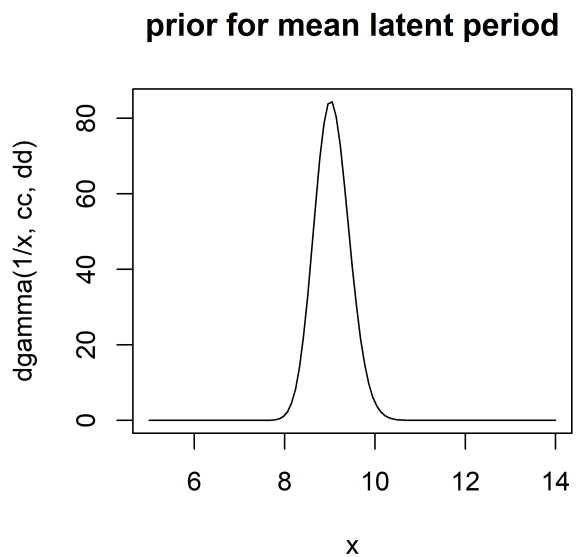
The data provides incidence over time (***I****), so ***E**** and ***R**** must be imputed

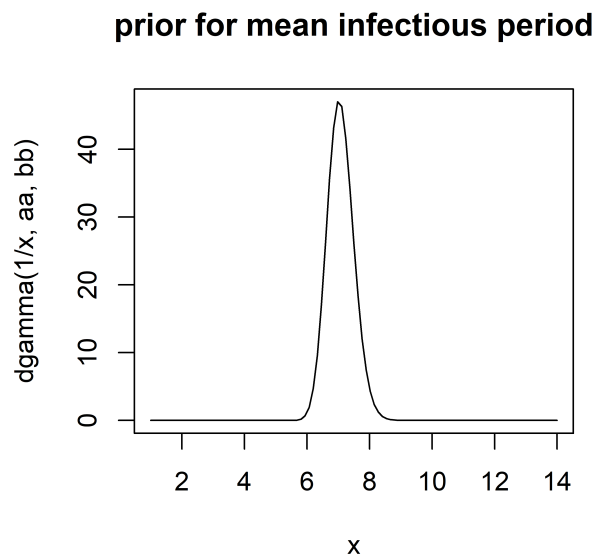
Strong priors can be used for ρ and γ using previous knowledge about the latent and infectious periods for EVD

As $\rho, \gamma > 0$, we use gamma priors

$\rho \sim \Gamma(556, 5000)$ puts 99% prior probability on the mean latent period being between 8 and 10 days

$\gamma \sim \Gamma(286, 2000)$ puts $\approx$98% prior probability on the mean infectious period being between 6 and 8 days





Incorporating Interventions

The impact of the interventions on transmission can be investigated by expanding the transmission probability

$$\pi_t^{(SE)} = 1 - \exp\left(-\beta_t \frac{I_t}{N}\right)$$

where

$$\beta_t = \begin{cases} \beta_0 & t < t_1^* \\ \beta_0 + \beta_1 & t_1^* \leq t < t_2^* \\ \beta_0 + \beta_1 + \beta_2 & t \geq t_2^* \end{cases}$$

with t_1^* and t_2^* being epidemic time when UNMEER was initiated and when 100% case isolation and safe burials were acheievd, respectively

This can be written as a linear combination of the design matrix
$\boldsymbol{X}_{\tau \times 3} = [1, I(t \geq t_1^*), I(t \geq t_2^*)]$

To ensure $\beta_t > 0$ for all t, exponentiate: $\beta_t = \exp(\boldsymbol{X}\boldsymbol{\beta})$

Flexible Transmission Modeling

Instead of using change points in transmission to interpret intervention effects, we could flexibly model β_t with cubic splines

$$\beta_t = \exp(\boldsymbol{X}_B \boldsymbol{b})$$

where $\boldsymbol{X}_B$ is the basis matrix spanning the range of time in the data and $\boldsymbol{b}$ are the associated parameters

The basis matrix is generated using two knots on the intervention dates:

October 1 - UNMEER initiated

Dec 28 - 100% case isolation and safe burials achieved

We will compare and contrast these two transmission models

Model 1 = piecewise transmission

Model 2 = flexible (spline) transmission

Using Posterior Prediction to Assess Interventions

Posterior prediction can be used to visualize the impact of an intervention on epidemic trajectory

For epidemic models, posterior prediction proceeds with two steps:

1. Draw sample of model parameters from posterior samples

2. Simulate epidemics over time according to the chain binomial model

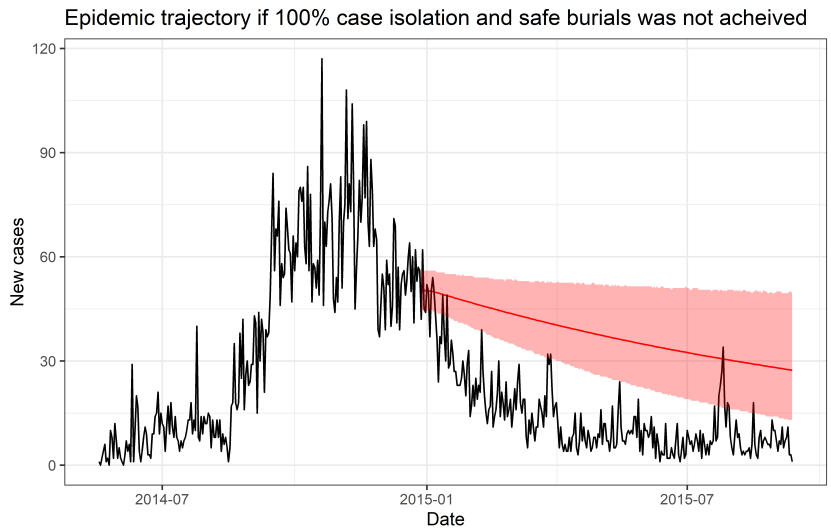
This is repeated many times to obtain a posterior predictive distribution

Using Posterior Prediction to Assess Interventions

Using the covariate structure of Model 1, we can find the potential trajectory without any control measures by setting $\beta_1 = \beta_2 = 0$

Or without control measures reaching 100% efficacy by setting $\beta_2 = 0$

Then we can predict epidemic trajectory from the start of these interventions



File: “RMarkdown Part2c.Rmd"

Outline

Spatially Stratified Transmission Modeling

The population-averaged model can be extended to incorporate spatial heterogeneity

Let $j = 1, ..., J$ denote spatial regions and let $\eta_{tj} = \beta_{tj} \frac{I_{tj}}{N_j}$ be the transmission rate in location j at time t

If $A_{jk} \in \{0, 1\}$ indicates region j and k are neighbors, the transmission probability can be defined as:

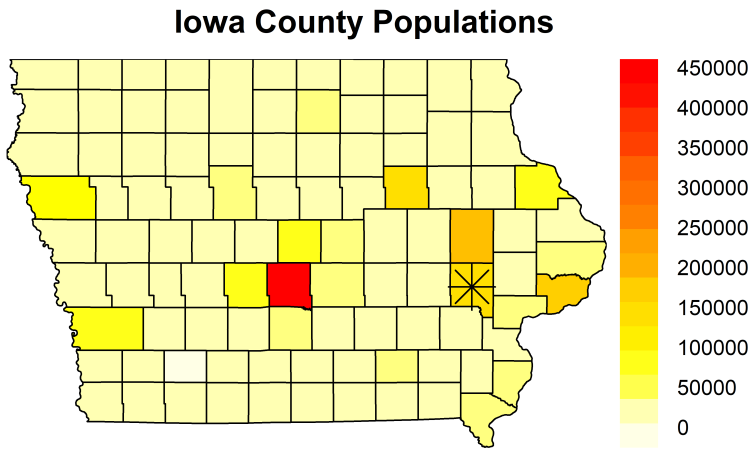
$$\pi_{tj}^{(SI)} = 1 - \exp\left[-\left(\eta_{tj} + \rho \sum_{k \neq j} A_{jk} \eta_{tk}\right)\right]$$

where $0 < \rho < 1$

Transmission to region j is a combination of transmission from within region j and the neighboring regions

Simulating Spatially Stratified Epidemics

To illustrate the spatial model, we will simulate spread of an infectious disease across counties in the state of Iowa, starting in Johnson county



Spatial structure: obtained from `raster` package

County populations: DATA/iowaCountyPopulations.csv

File: “RMarkdown Part2d.Rmd"

Selected References

Most Relevant:

P. E. Lekone & B. F. Finkenstädt (2006) “Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study" in *Biometrics*, 62(4), 1170-1177.

P. D. O’Neill & G. O. Roberts (1999) “Bayesian inference for partially observed stochastic epidemics” in *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162(1), 121-129.

Broader Background Reading:

Fang, et al. (2016) “Ebola virus disease in Sierra Leone” in *Proceedings of the National Academy of Sciences*, 113(16) 4488-4493

Appendix - Details of Data-Augmented MCMC

Proposing Removal Times

More details in O'Neill and Roberts (1999) and Lekone and Finkenstädt (2006)

A 'relatively' efficient proposal scheme can be specified as follows:

1. Initialize $\boldsymbol{R}^*$ by assuming removals occur a fixed number of days after cases are reported

2. At each iteration of the Markov chain, repeatedly propose many small perturbations of $\boldsymbol{R}^*$

3. For each perturbation, compute the MH acceptance ratio and accept/reject the candidate proposal

$$\frac{f(x')}{f(x)} \frac{g(x|x')}{g(x'|x)}$$

How can $\boldsymbol{R}^*$ be perturbed efficiently?

Proposing Removal Times

Consider three possible moves when proposing removal times:

1. Adding a removal time
2. Subtracting a removal time
3. Moving a removal time

Assume equal (1/3) probability of each move

These moves are not always symmetric, so the last thing needed is to determine the ratio $\frac{g(x|x')}{g(x'|x)}$, where x' is the candidate proposal and x is the current state

Proposing Removal Times

Removals can be added to any time point, but only subtracted from times with non-zero counts

1. Adding a removal time

$$x = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \rightarrow x' = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \text{ or } \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \text{ or } \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}, \qquad g(x'|x) = 1/3$$

2. Subtracting a removal time

$$x = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \rightarrow x' = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \text{ or } \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \qquad g(x'|x) = 1/2$$

Proposing Removal Times

3. Moving a removal time

$$x = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \rightarrow$$

If x_2 is moved: $\quad x' = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

If x_3 is moved: $\quad x' = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$

$$g(x'|x) = 1/2 \times 1/3$$

Proposing Removal Times

In summary, if n_{nz} is the number of non-zero elements of x, n'_{nz} is the number of non-zero elements of x', and τ is the total length of both x and x'

1. Adding a removal time ($x'|x$ = add, $x|x'$ = subtract)

$$\frac{g(x|x')}{g(x'|x)} = \frac{1/n'_{nz}}{1/\tau}$$

2. Subtracting a removal time ($x'|x$ = subtract, $x|x'$ = add)

$$\frac{g(x|x')}{g(x'|x)} = \frac{1/\tau}{1/n_{nz}}$$

3. Moving a removal time

$$\frac{g(x|x')}{g(x'|x)} = \frac{1/n'_{nz} \times 1/\tau}{1/n_{nz} \times 1/\tau} = \frac{1/n'_{nz}}{1/n_{nz}}$$

Note: we typically compute these on the log scale

File: “RMarkdown Part2e_appendix.Rmd"