libspatialSEIR: Model, Algorithm, and Implementation
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.

Persons in a population are assumed to move through these disease categories sequentially according to the disease process, though in practice these disease states are combined in different ways. For example, a simple epidemic model might use an S-I-R structure in which individuals become immediately infectious (ie, there is no latent period), and are subsequently removed from the population (ie, assuming permanent immunity). For a disease which conferrs only temporary immunity, such as influenza, many researchers have employed S-I-R-S models, which include a potential for previously recovered individuals to be reintroduced to the susceptible population. Similarly, SEIR and SEIRS models introduce a latent period during which exposed individuals are not yet, but will become, infectious. In cases where the reintroduction process is either complicated, or data is sparse, researchers can employ a "Serial SEIR" model, which simply re-sets the susceptible population at regular intervals.

This diverse family temporal structures has been generalized to allow epidemics to be modeled over space, as well as time.

1.2 TL;DR - What is libsaptialSEIR for?

libspatialSEIR was designed to provide a computationally efficient and user friendly method to fit several important epidemic models in the Bayesian spatial SEIRS family. In particular, spatial and single location models are allowed to evolve through time via one of several SEIRS-like structures, with a particular effort to allow the inclusion of linear predictors to drive important model parameters. Notable examples include:

- Single location SEIR and SEIRS models.
- Spatial SEIR and SEIRS models, employing a user specified measure of distance between spatial locations.
- Serial SEIR models, both for single locations and spatially indexed data.

The focus on linear predictors provides an additional layer of flexibility to these models, as they can accommodate temporal basis functions for epidemic shape fitting, intervention indicators for public health efforts, demographic effects on population mixing, and anything else that can be expressed as a linear combination of unknown, normally distributed parameters.

Example code for each of these models is included in the *scripts* directory, and several full tutorials are available in the tutorial subdirectory of *doc* folder.

1.3 SEIRS Model Family - Formal Development

Developed below is the most general member of the SEIRS family which can be fit with libspatialSEIR, namely the spatial SEIRS model. Important special cases are discussed in the following section.

1.3.1 Compartments and Notation

Denote the spatial locations of interest $\{s_i : i = 1, ..., n\}$
Let $d(s_i, s_l) = d_{il}$ define a measure of distance between spatial locations. Note that s_i and s_l. $d(s_i, s_i) = 0$, and $d(s_i, s_l) = d(s_l, s_i)$

Let time (in units appropriate to the data and disease process) be denoted $t_j : j = 1, ..., T$
For now, assume that all time points are equally spaced - the library allows the inclusion of an offset term to relax this assumption, but for illustrative purposes we'll keep the notation simple here. Define the following components for each s_i and t_j:

- y_{ij} is the observed data.
- $\mathbf{N_{ij}}$ is the population size
- $\mathbf{S_{ij}}$ is the count of susceptible individuals
- $\mathbf{E_{ij}}$ is the count of exposed individuals
- $\mathbf{I_{ij}}$ is the count of infectious individuals
- $\mathbf{R_{ij}}$ is the count of recovered/removed individuals
- $\mathbf{S^*_{ij}}$ is the number of newly susceptible individuals
- $\mathbf{E^*_{ij}}$ is the number of newly exposed individuals
- $\mathbf{I^*_{ij}}$ is the number of newly infectious individuals
- $\mathbf{R^*_{ij}}$ 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
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.

1.4 Data Model

An obvious first step is to specify the data model.

$$\{y_{ij} \mid I^*_{ij}\} \sim \ (indep.)\ g(I^*_{ij}, \Theta)$$

Currently, libspatialSEIR only supports the identity data model, in which:

$$g(I^*_{ij}) = I^*_{ij}$$

with probability one. This is under active development.

1.4.1 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.

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

While models of this form are often fit using deterministic systems of 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:

$$\{S_{ij}^* | \pi_j^{(RS)}, R_{ij}\} \sim \ (indep.)\ binom(R_{ij}, \pi_j^{(RS)})$$

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

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

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

1.5 Transition Probability Model

While π^{EI}, and π^{IR} can be easily parameterized, more care must be given to the development of a model for the $\left\{\pi_{ij}^{SE}\right\}$ and $\left\{\pi_j^{RS}\right\}$. The first of these describes the actual infection process and must account for predictor variables as well as the spatial structure of $\{s_i\}$. The second drives the reinfection process, which captures the effects of diminishing temporary immunity, disease strain mutation, and the introduction of new infectious agents. These important components are examined in the following sections.

1.6 Infection Process - CAR Model Motivation

Consider the process by which people become infected with a communicable disease. Namely, consider the situation in which a person 'A' has contacted another person, 'B', who is infectious (for some suitable definition of contacted). Let p be the probability that person 'A' becomes infected with the disease, and let $q = 1 - p$. Now we introduce a number of assumptions:

- Assume that the number of 'contacts' K_i between a person of interest and other individuals within a spatial unit s_i at a given time point follows a poisson distribution:

$$K_j \sim Po(\lambda_i)$$

- Assume that when individuals travel to other spatial locations, their contact behavior is well modeled by the contact behavior of that spatial unit (when in Rome).

- Contact between spatial locations is proportional to some known function $f(d_{il})$ of the chosen distance metric between the centroids of s_i and s_l

Define δ_{ij} to be the proportion of persons who are infectious in spatial unit s_i at time t_j. Then, letting $Inf(s_i, t_j)$ denote the event that a person becomes infected from contact within spatial unit s_i at time t_j, we can derive:

$$P(Inf(., t_j)) = 1 - P(!Inf(s_i, t_j)) \cdot P(!Inf(s_{-i}, t_j))$$

where

$$P(!Inf(s_i, t_j)) = E(!Inf(s_i, t_j)) = E(E(!Inf(s_i, t_j)|K_i = k_i))$$

$$= E(((1 - \delta_{ij})q)^{k_i})$$

$$= \sum_{k=0}^{\infty}((1 - \delta_{ij})q)^k(\frac{\lambda_i^k e^{-\lambda_i}}{k!})$$

$$= \sum_{k=0}^{\infty} q_{ij}^k(\frac{\lambda_i^k e^{-\lambda_i}}{k!})$$

$$= \frac{e^{-\lambda_i}}{e^{-q_{ij}\lambda_i}}(1) = e^{-\lambda_i \cdot (1-q_{ij})} = e^{-\lambda_i \cdot p_{ij}} = e^{-\lambda_i \cdot (\delta_{ij} p)}$$

Therefore, $P(Inf(s_i, t_j)) = 1 - e^{-\lambda_i \cdot (\delta_{ij} p)}$

Similarly,

$$P(!Inf(s_{-i}, t_j)) = \prod_{\{l \neq i\}} [P(!Inf(s_l, t_j))]$$

$$= \prod_{\{l \neq i\}} [E(!Inf(s_{-i}, t_j))] = \prod_{\{l \neq i\}} [E(E(!Inf(s_{-i}, t_j)|K_i = k_i))]$$

$$= \prod_{\{l \neq i\}} \left[E((1 - \delta_{lj})q)^k)\right]$$

$$= \prod_{\{l \neq i\}} \left[\sum_{k=0}^{\infty} (q_{lj}(i))^k \frac{(\lambda_l \cdot f(d_{il}))^k e^{-\lambda_l \cdot f(d_{il})}}{k!}\right] = \prod_{\{l \neq i\}} \left[\frac{e^{-\lambda_l \cdot f(d_{il})}}{e^{-q_{lj}\lambda_l f(d_{il})}}(1)\right]$$

$$= \prod_{\{l \neq i\}} \left[e^{-\lambda_l \cdot f(d_{il}) p_{lj}}\right] = \prod_{\{l \neq i\}} \left[e^{-\lambda_l \cdot f(d_{il}) \cdot (\delta_{lj} p)}\right]$$

$$= exp\left\{\sum_{\{l \neq i\}} [p\lambda_l \delta_{jl} f(d_{il})]\right\}$$

Thus, for the probabilty of infection for a person living in s_i at time t_j we have:

$$1 - \left(e^{-\lambda_i \cdot (\delta_{ij} p)}\right)\left(e^{\left\{\sum_{\{l \neq i\}} \left[p\lambda_l \delta_{jl} f(d_{il})\right]\right\}}\right)$$

$$= 1 - exp\left\{-\delta_{ij} e^{\theta_i} - \sum_{\{l \neq i\}} (f(d_{il})\delta_{il} e^{\theta_l})\right\}, \text{ where } \theta_v = log(\lambda_v p)$$

Currently, libspatialSEIR supports distance functions of the form:

$$f(d_{il}) = \rho \cdot (d_{il})^{-\frac{1}{2}}$$

While not required, the option to re-scale this distance matrix to be row stochastic is available to ensure a proper posterior distribution.

Support for other distance functions, such as the gravity model, is planned.

A few miscellaneous notes:

1. By defining $f(d_{ii})$ to be equal to 1 for all i, the above expression has a simple matrix form.

2. WLOG, we can make the mixing parameters dependent on space and time, defining λ_{v_1,v_2}, and correspondingly θ_{v_1,v_2}

3. When constructing full conditional distributions, it is important to keep in mind the constraint: $S + E + I + R = N$, as well as the fact that $p_{i,j}^{(SE)}$ does, in fact, depend on the value of I. This is easy to neglect and hard to debug.

This spatial probability structure belongs to the CAR, or Conditionally Auto-Regressive, class of spatial dependence structures, and can accommodate several different data types. For example, in the case where spatial data is indexed by discrete areal units, a neighborhood matrix may define the requisite 'distance' between spatial locations. On the opposite end of the spectrum, spatial locations may occur on a continuum (ie, latitude and longitude). In this case, a more usual distance definition of distance can be employed (with or without an explicit spatial range). libspatial-SEIR is optomized for the second case (ie, sparse matrix methods are not employed), though is perfectly capable of fitting sparser spatial (or entirely non-spatial) models.

1.7 Re-infection Process

To model the π_j^{RS}, some structure of lower than T dimensions is desireable to reduce potential identifiability issues. A covariate structure constructed to capture natural variation in this quantity, a set of trigonometric basis functions with appropriate period, for example, will do nicely. Let $X(\pi_j^{RS})$ and $\beta_{\pi^{RS}}$ denote the covariate vector and corresponding regression parameter estimates for the j'th R to S transition probability respectively. Any number of link functions might be appropriate here, however due the the ease with which it can be generalized to inconsistent temporal indices the following (generalized from Lekone and Finkenstädt (2006)) is used:

$$\pi_j^{RS} = 1 - e^{-exp(X(\pi_j^{RS})\beta_{\pi RS})}$$

2 Basic Reproductive Number

The basic reproductive number, $\mathcal{R}_0$, is an important quantity in epidemiology. While the interepretation 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.

Defining the relevant infection event for a person indexed by k to be: $I_k(s_i, s_l, t_j)$, we see that the expected number of such infections is:

$$E\big[\sum_{k=0}^{N_{i,j}}(I_k(s_i,s_l,t_j))\big]$$

$$=\sum_{k=0}^{N_{i,j}}\cdot P(I_k(s_i,s_l,t_j))=N_{i,j}\cdot P(I_k(s_i,s_l,t_j))$$

Where, as before:

$$P(I_k(s_i,s_l,t_j))=1-exp\left\{-f(d_{il})\delta_{lj}e^{\theta_l}\right\}$$

This gives:

$$G_{i,l}(t_j)=\tfrac{N_{i,j}}{I_{l,j}}\cdot\left[1-exp\left\{-f(d_{il})\delta_{lj}e^{\theta_l}\right\}\right]$$

Additionally recall the diagonal case, where $d_{ii}=0$ and $f(0)=1$:

$$G_{i,i}(t_j)=\tfrac{N_{i,j}}{I_{i,j}}\cdot\left[1-exp\left\{-\delta_{ij}e^{\theta_i}\right\}\right]$$
$$=\delta_{ij}^{-1}\cdot\left[1-exp\left\{-\delta_{ij}e^{\theta_i}\right\}\right]$$

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

3 Posterior Distribution and Full Conditionals

Bringing together the aforementioned spatio-temporal structures we can define the requisit prior distributions and deterministic relationships among parameters, and thus construct the requisit posterior distribution.

3.1 Summary of Distribution Components

$\{y_{ij}|I^*_{ij}\}\sim\ (indep.)\ g(I^*_{ij},\Theta)$

$\{S^*_{ij}|\pi^{(RS)}_j,R_{ij}\}\sim\ (indep.)\ binom(R_{ij},\pi^{(RS)}_j)$

$\{E^*_{ij}|\pi^{(SE)}_{ij},S_{ij}\}\sim(indep.)\ binom(S_{ij},\pi^{SE}_{ij})$

$\{I^*_{ij}|\pi^{(EI)},E_{ij}\}\sim\ (indep.)\ binom(E_{ij},\pi^{(EI)})$

$\{R^*_{ij}|\pi^{(IR)},I_{ij}\}\sim\ (indep.)\ binom(I_{ij},\pi^{(IR)})$

$\gamma^{(IR)} \sim \ gamma(\alpha^{(IR)}, \beta^{(IR)})$

$\gamma^{(EI)} \sim \ gamma(\alpha^{(EI)}, \beta^{(EI)})$

$\{\theta_{ij}\} \sim \mathcal{N}(\eta_{ij}, \sigma_\theta^2)$

$\{\beta\} \sim \mathcal{N}(0, \tau_\beta^2)$

$\{\beta_{\pi^{RS}}\} \sim \mathcal{N}(0, \tau_{RS}^2)$

$\sigma_\theta^2 \sim \Gamma(\alpha_\theta, \beta_\theta)$

$\rho \sim U(0, 1)$

3.2 Deterministic Functions

$S = f_S(S_0, E_0^*, S_0^*, S^*, E^*)$

$E = f_E(E_0, I_0^*, E_0^*, E^*, I^*)$

$I = f_I(I_0, R_0^*, I_0^*, I^*, R^*)$

$R = f_R(R_0, S_0^*, R_0^*, R^*, S^*)$

$log(\pi_{ij}^{(SE)}) = -\delta_{ij} e^{\theta_{ij}} - \sum_{\{l \neq i\}} d_{il} \delta_{il} e^{\theta_{il}}$

$\{\pi_{EI}\} = 1 - exp{-\gamma^{(EI)}}$

$\{\pi_{IR}\} = 1 - exp{-\gamma^{(IR)}}$

$log(\pi_j^{RS}) = -X(\pi_j^{RS}) \beta_{\pi^{RS}}$

$d_{il} = f(\rho, s_i, s_l)$

$\delta_{ij} = \frac{I_{ij}}{N_{ij}}$

$\eta_{ij} = X_{ij} \beta$

3.3 Posterior Distribution

$$
\begin{aligned}
log(p(\theta,\beta,\rho,S^*,E^*,R^*|.)) \propto \Bigg[\sum_{i=1}^{n} \Bigg\{ \sum_{j=1}^{T} \Big\{ ln(g(y_{ij}|I^*_{ij})) + \\
(S^*_{ij} log(\pi_j^{(RS)}) + (R_{ij} - S^*_{ij}) log(1-\pi_j^{(RS)})) \\
+ (E^*_{ij} log(\pi_{ij}^{(SE)})) + (S_{ij} - E^*_{ij}) log(1-\pi_{ij}^{(SE)}) \\
+ (I^*_{ij} log(\pi^{(EI)})) + (E_{ij} - I^*_{ij}) log(1-\pi^{(EI)}) \\
+ (R^*_{ij} log(\pi^{(IR)})) + (I_{ij} - R^*_{ij}) log(1-\pi^{(IR)}) \\
- \frac{1}{2} log(\sigma^2_\theta) - \frac{(\theta_{ij} - \eta_{ij})^2}{2\sigma^2_\theta} \Big\} \Bigg\} \\
+ \sum_{k=1}^{K} \Bigg\{ - \frac{\beta_k^2}{10} \Bigg\} + (log(\pi(\sigma^2_\theta))) + (log(\pi(\rho))) \\
+ \Bigg\{ - \frac{\tau^2_{RS}}{2} \|\beta_{\pi_j^{RS}}\| \Bigg\} \\
+ log(\pi(\gamma^{(EI)})) + log(\pi(\gamma^{(IR)})) \Bigg]
\end{aligned}
\tag{1}
$$

For simplicity, and because the model is sufficiently flexible without the added over-dispersion, set $\theta_{ij} = \eta_{ij}$ with probability 1. This gives the following simplified posterior distribution:

$$
\begin{aligned}
log(p(\theta,\beta,\rho,S^*,E^*,R^*|.)) \propto \\
&\Bigg[\sum_{i=1}^{n}\Bigg\{\sum_{j=1}^{T}\Big\{ln(g(y_{ij}|I^*_{ij}))\\
&+log\Bigg(\binom{R_{ij}}{S^*_{ij}}\Bigg)+(S^*_{ij}log(\pi_j^{(RS)})+(R_{ij}-S^*_{ij})log(1-\pi_j^{(RS)}))\\
&+log\Bigg(\binom{S_{ij}}{E^*_{ij}}\Bigg)+(E^*_{ij}log(\pi_{ij}^{(SE)}))+(S_{ij}-E^*_{ij})log(1-\pi_{ij}^{(SE)})\\
&+log\Bigg(\binom{E_{ij}}{I^*_{ij}}\Bigg)+(I^*_{ij}log(\pi^{(EI)}))+(E_{ij}-I^*_{ij})log(1-\pi^{(EI)})\\
&+(log\Bigg(\binom{I_{ij}}{R^*_{ij}}\Bigg)+R^*_{ij}log(\pi^{(IR)}))+(I_{ij}-R^*_{ij})log(1-\pi^{(IR)})\Big\}\Bigg\}\\
&+\sum_{k=1}^{K}\Big\{-\frac{\beta_k^2}{10}\Big\}+(log(\pi(\rho)))\\
&+\Big\{-\frac{\tau^2_{RS}}{2}\|\beta_{\pi_j^{RS}}\|\Big\}\\
&+\Big\{-\frac{\tau^2_{RS}}{2}\|\beta_{\pi_j^{RS}}\|\Big\}\\
&+log(\pi(\gamma^{(EI)}))+log(\pi(\gamma^{(IR)}))\Bigg]
\end{aligned}
\tag{2}
$$

3.4 Full Conditional Distributions

- Full Conditional For S^*

$$
\begin{aligned}
log(p(S^*|.)) \propto \sum_{j=1}^{T}\sum_{i=1}^{n}\Big\{&log\Bigg(\binom{f_R(R^*,S^*,A_0)_{ij}}{S^*_{ij}}\Bigg)\\
&+log(\pi_j^{(RS)})\{S^*_{ij}\}+log(1-\pi_j^{(RS)})\{f_R(R^*,S^*,A_0)_{ij}-S^*_{ij}\}\\
&+log\Bigg(\binom{f_S(S^*,E^*,A_0)_{ij}}{E^*_{ij}}\Bigg)+f_S(S^*,E^*,A_0)_{ij}log(1-\pi_{ij}^{(SE)})\Big\}
\end{aligned}
\tag{3}
$$

- Full Conditional for E^*

$$
\begin{aligned}
log(p(E^*|.)) \propto \sum_{i=1}^{n}\sum_{j=1}^{T} \Bigg\{ & log\Bigg(\begin{pmatrix} f_S(S^*, E^*, A_0)_{ij} \\ E^*_{ij} \end{pmatrix}\Bigg) \\
& + (E^*_{ij} log(\pi^{(SE)}_{ij})) + (f_S(S^*, E^*, A_0)_{ij} - E^*_{ij}) log(1 - \pi^{(SE)}_{ij}) \Bigg\} \\
& + log\Bigg(\begin{pmatrix} f_E(E^*, I^*, A_0)_{ij} \\ I^*_{ij} \end{pmatrix}\Bigg) + log(1 - \pi^{(EI)}) \sum_{i=1}^{n}\sum_{j=1}^{T} \{f_E(E^*, I^*, A_0)_{ij}\}
\end{aligned} \tag{4}
$$

- Full Conditional for I^*

$$
\begin{aligned}
log(p(I^*|.)) \propto \sum_{i=1}^{n}\sum_{j=1}^{T} \Bigg\{ & log(g(y_{ij}|I^*_{ij})) + log\Bigg(\begin{pmatrix} f_E(E^*, I^*, A_0)_{ij} \\ I^*_{ij} \end{pmatrix}\Bigg) \\
& + (I^*_{ij} log(\pi^{(EI)})) + (f_E(E^*, I^*, A_0)_{ij} - I^*_{ij}) log(1 - \pi^{(EI)}) \Bigg\} \\
& + log\Bigg(\begin{pmatrix} f_I(I^*, R^*, A_0)_{ij} \\ R^*_{ij} \end{pmatrix}\Bigg) + log(1 - \pi^{(IR)}) \sum_{i=1}^{n}\sum_{j=1}^{T} \{f_I(I^*, R^*, A_0)_{ij}\}
\end{aligned} \tag{5}
$$

- Full Conditional for R^*

$$
\begin{aligned}
log(p(R^*|.)) \propto \sum_{i=1}^{n}\sum_{j=1}^{T} \Bigg\{ & log\Bigg(\begin{pmatrix} f_I(I^*, R^*, A_0)_{ij} \\ R^*_{ij} \end{pmatrix}\Bigg) \\
& + log(\pi^{(IR)})\{R^*_{ij}\} + log(1 - \pi^{IR})\{f_I(I^*, R^*, A_0)_{ij} - R^*_{ij}\} \\
& + log\Bigg(\begin{pmatrix} f_R(R^*, S^*, A_0)_{ij} \\ S^*_{ij} \end{pmatrix}\Bigg) + log(1 - \pi^{(RS)}_{j})\{f_R(R^*, S^*, A_0)_{ij}\} \\
& + log(\pi^{(SE)}_{ij})\{E^*_{ij}\} + log(1 - \pi^{SE}_{ij})\{f_S(S^*, E^*, A_0)_{ij} - E^*_{ij}\} \Bigg\}
\end{aligned} \tag{6}
$$

- Full Conditional for $\{\theta\}$

$$log(p(\{\theta\}|.)) \propto \sum_{i=1}^{n}\Big\{\sum_{j=1}^{T}\Big\{(E^*_{ij}log(\pi^{(SE)}_{ij}))+(f_S(S^*,E^*,A_0)_{ij}-E^*_{ij})log(1-\pi^{(SE)}_{ij})\Big\}\Big\} \tag{7}$$

- Full Conditional for $\{\beta\}$

$$log(p(\{\beta\}|.)) \propto \sum_{i=1}^{n}\Big\{\sum_{j=1}^{T}\Big\{(E^*_{ij}log(\pi^{(SE)}_{ij}))+(f_S(S^*,E^*,A_0)_{ij}-E^*_{ij})log(1-\pi^{(SE)}_{ij})\Big\}\Big\} + \sum_{k=1}^{K}\Big\{-\frac{\beta_k^2}{10}\Big\} \tag{8}$$

- Full Conditional for ρ

$$log(p(\rho|.)) \propto \sum_{i=1}^{n}\Big\{\sum_{j=1}^{T}\Big\{(E^*_{ij}log(\pi^{(SE)}_{ij}))+(f_S(S^*,E^*,A_0)_{ij}-E^*_{ij})log(1-\pi^{(SE)}_{ij})\Big\}\Big\} + (log(\pi(\rho))) \tag{9}$$

- Full Conditional for $\{\beta_{\pi_j^{(RS)}}\}$

$$\begin{aligned} log(p(\{\pi_j^{(RS)}\}|.)) \propto \sum_{j=1}^{T}\Big\{ & log(\pi_j^{(RS)})\Big[\sum_{i=1}^{n}\{S^*_{ij}\}\Big] \\ & + log(1-\pi_j^{(RS)})\Big[\sum_{i=1}^{n}\{f_R(R^*,S^*,A_0)_{ij}-S^*_{ij}\}\Big]\Big\} \\ & + \Big\{-\frac{\tau^2_{RS}}{2}\|\beta_{\pi_j^{RS}}\|\Big\} \end{aligned} \tag{10}$$

- Full Conditional for $\gamma^{(EI)}$

$$log(p(\gamma^{(EI)})) \propto log(\pi^{(EI)})\Big[\sum_{i=1}^{n}\sum_{j=1}^{T}\{I^*_{ij}\}\Big] + log(1-\pi^{(EI)})\Big[\sum_{i=1}^{n}\sum_{j=1}^{T}\{(f_E(E^*, I^*, A_0)_{ij} - I^*_{ij})\}\Big] + log(\pi(\gamma^{(IR)})) \tag{11}$$

- Full Conditional for $\gamma^{(IR)}$

$$log(p(\gamma^{(IR)})) \propto log(\pi^{(IR)})\Big[\sum_{i=1}^{n}\sum_{j=1}^{T}\{R^*_{ij}\}\Big] + log(1-\pi^{(IR)})\Big[\sum_{i=1}^{n}\sum_{j=1}^{T}\{f_I(I^*, R^*, A_0)_{ij} - R^*_{ij}\}\Big] + log(\pi(\gamma^{(IR)})) \tag{12}$$

4 MCMC Samplers: Current and Future Work

While various combinations of slice and Metropolis-Hastings sampling have been explored, the current recommendation is to use blocked Metropolis-Hastings for all parameters. As implemented here, each of the full conditional distributions is sampled by proposing a new value centered on the current value for the entire parameter. Coupled with using the auto-tuning functions to maintain a reasonable acceptance rate, this has different effects for compartments and other parameters. For compartments (because they are integers), such an approach has the effect of choosing a random sample of time-locations to update, moving up or down by a small ammount (1 or 2). For other parameters, the entire parameter vector is either updated or not at each step.

More development is planned to deal with the considerable autocorrelation issues experienced so far.

See the tutorial documents for the most current implementation details.

5 OpenCL: Performance Tradeoffs

OpenCL is a useful way to introduce parallelism to computationally intensive problems, and can target multi-core CPU's as well as GPU's. Indeed, initial results suggest that for moderately large problems (20-30k time/location values) OpenCL can speed up the work of libspatialSEIR by around %20. This number is expected to increase as more of the library is parallelized.

On the other hand, for small problems OpenCL has enough overhead to significantly slow down run times. The Google Flu tutorial document gives some example code which is useful for checking whether parallelism is the best option for your particular problem. You can easily switch between OpenCL devices and single core CPU mode at any time during sampling.

6 R Level API Summary

Once the R level API is complete, this section will include proper documentation and example code. For now, development is moving too fast to keep a current summary. All example code and tutorial documents are kept up to date, so that's probably the best resource for now.