Bayesian Modelling of Epidemics

Part IIIb: Inference for ILMs with the Epi-ILM R Package

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

Likelihood function

For our class of discrete-time SI and SIR models, assuming we know when infection and removals occur,

the **likelihood** is given by:

$$L(D|\theta) = \prod_t \left[\prod_{i \in S(t+1)} (1 - P(i,t)) \right] \left[\prod_{i \in I(t+1)/I(t)} P(i,t) \right]$$

where:

$S(t+1)$ is the set of all susceptible individuals at time $t+1$

$I(t+1)/I(t)$ is the set of all newly infectious individuals at time $t+1$

Statistical inference for ILMs

We will use Markov chain Monte Carlo (MCMC) techniques to sample from the posterior:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{D}) \propto L(\boldsymbol{D} \mid \boldsymbol{\theta})\ Pr(\boldsymbol{\theta})$$

Programs like NIMBLE, JAGS or BUGS are not suited to these types of models

Basically, the likelihood is too computationally cumbersome

But we can use the EpiILM R package
(Uses FORTRAN code to calculate the likelihood function)

Outline

Example 1: Spatial Disease System

Example:

Imagine we are modelling the spread of a transmissible disease through a series of farms.

We know the number of animals on each farm from a recent census

We know the spatial locations of the farms.

It seems reasonable to treat the farms themselves as individual level unit.

The extent of infection from outside the observed population of farms is unknown.

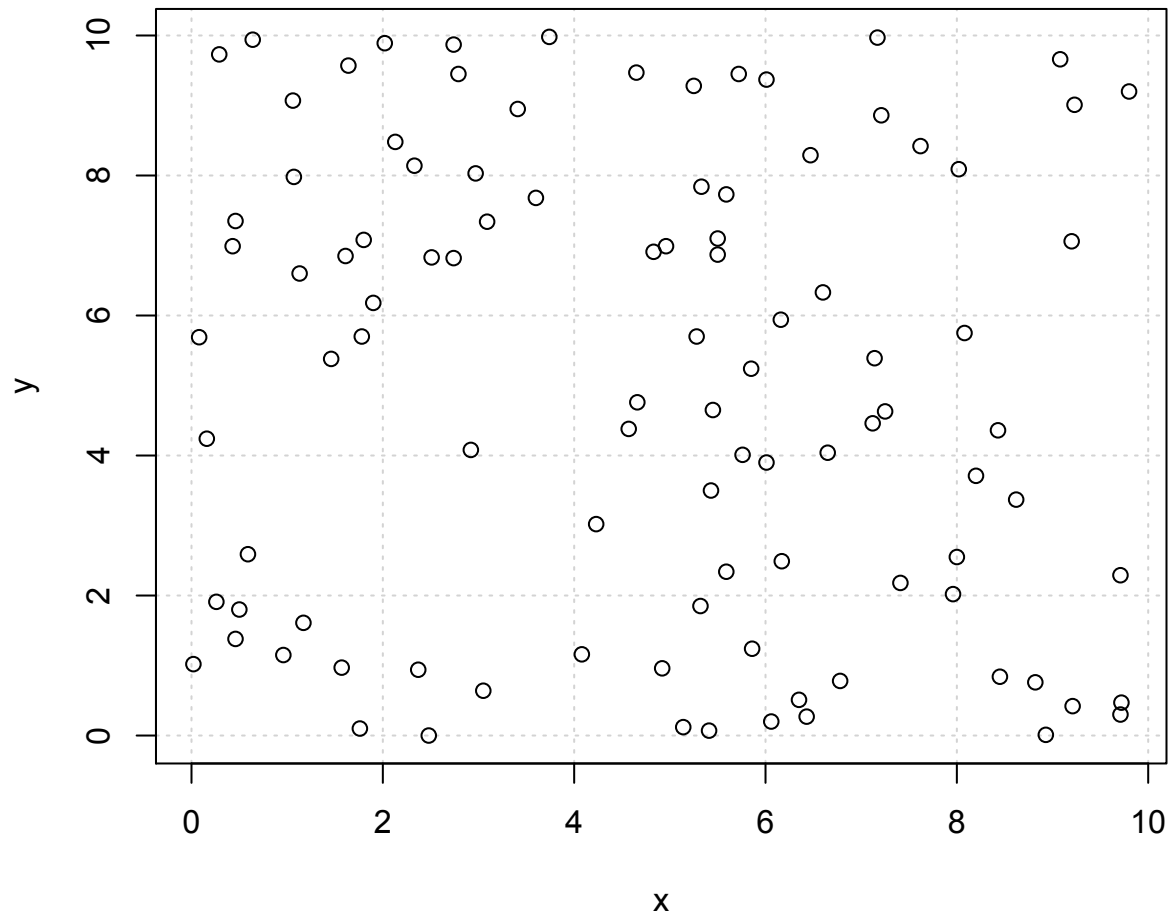
We have observed/estimated the times at which disease entered infected farms (infection times)

Outline

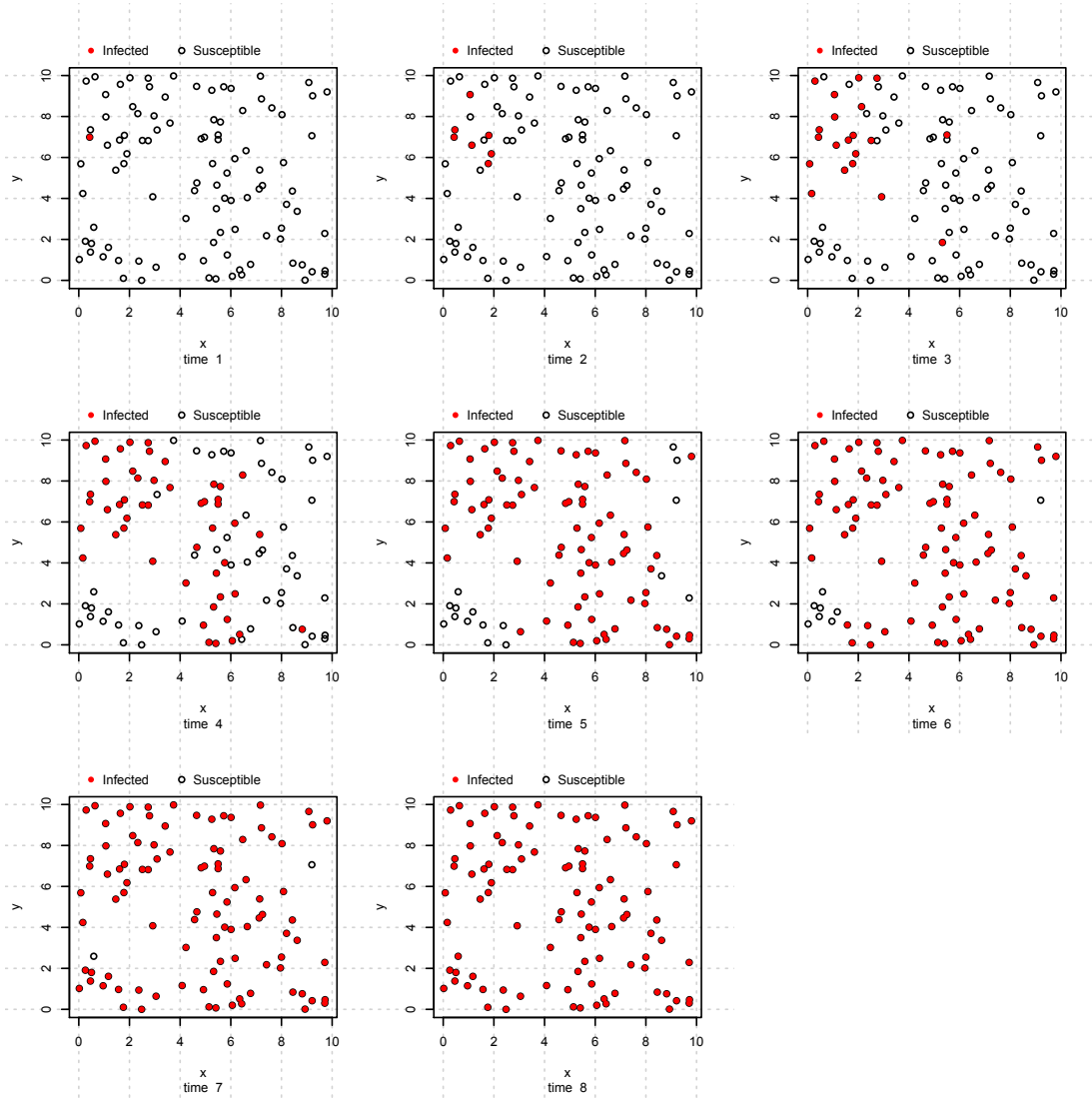
Data set: “DATA/ENAR_spatialepidemic.csv”

x	y	XA	inftime
3.11	4.92	141	2
1.07	0.61	47	2
1.54	0.93	23	2
6.21	8.53	576	4
4.16	9.94	11	5
6.61	8.37	231	5
8.45	6.6	10	5
0.52	5.08	16	4
5.68	3.36	140	4
0.49	5.21	23	4
0.29	8.05	54	5
5.34	0.14	10	5
5.96	1.51	36	4
3.91	5.92	176	3
5.35	5.89	164	4
3.09	1.92	139	2
6.24	9.7	93	5
1.9	4.39	211	3
6.49	4.28	53	4
3.14	5.57	262	3
1.72	2.85	52	2
9.4	0.81	30	5
1.03	8.24	77	5
9.12	1.48	2	6
4.21	7.16	157	3
. . .			

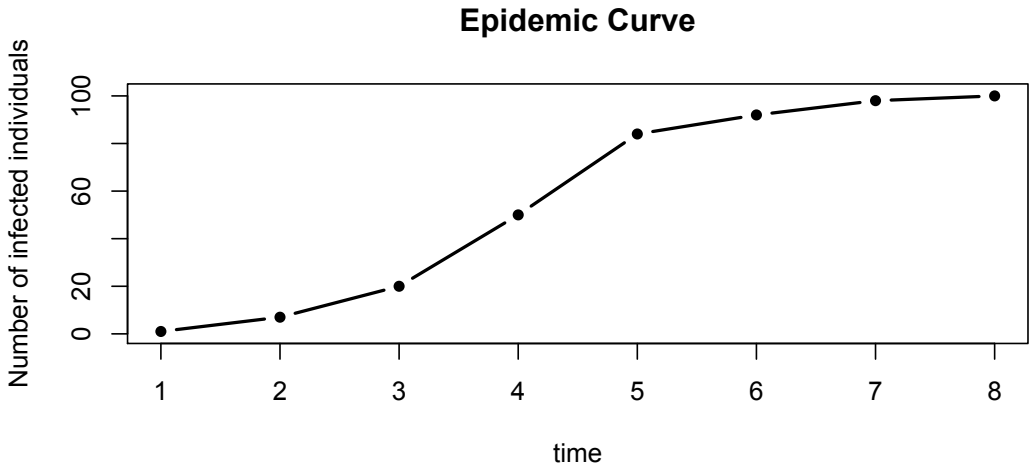
Spatial plot

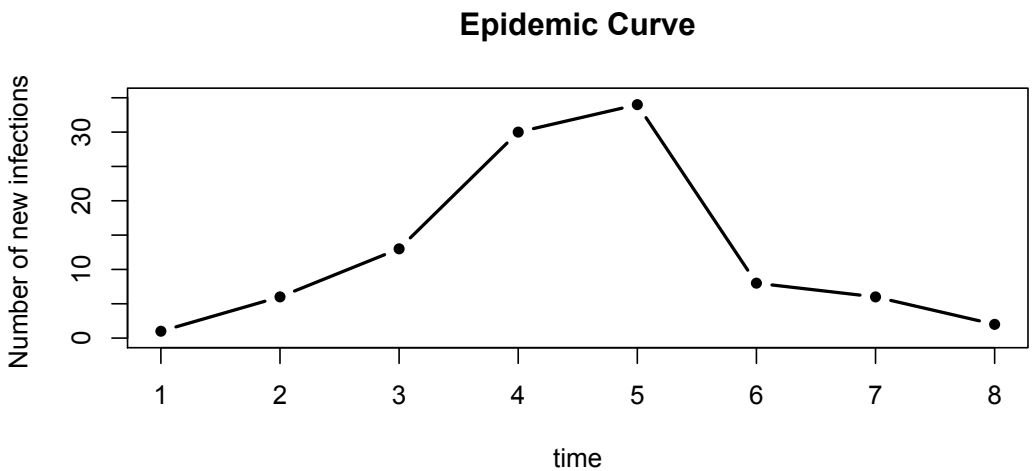


Spatial plot



Epidemic curves





Simple Spatial Model

Initially, let us:

assume an SI compartmental framework

assume negligible infection from outside of the observed population (i.e., $\epsilon = 0$)

fit a purely spatial model, ignoring any number of animals effect

Simple Spatial Model

For this model:

$$\begin{array}{lrcl} \text{Susceptibility:} & \mathcal{S}(i) & = & \alpha \\ \text{Transmissibility:} & \mathcal{T}(j) & = & 1 \\ \text{Infection Kernel:} & \kappa(i,j) & = & d_{ij}^{-\beta} \\ \text{Sparks:} & \epsilon(i,t) & = & 0 \end{array}$$

So our model is:

$$P(i,t) = 1 - \exp\left(-\alpha \sum_{j \in I(t)} d_{ij}^{-\beta}\right)$$

where:

d_{ij} is the spatial distance between individuals i and j
α is an infectivity parameter
β is a spatial parameter

Priors

Here, we have two parameters: $\boldsymbol{\theta} = (\alpha, \beta)$

Let's assume we have no substantive prior knowledge about them we wish to include in our analysis.

We could do this using any of the three priors offered by EpiILM

- half normal
- gamma
- uniform

Here, we will put uniform priors on the parameters with a wide range:

$\alpha \sim$ Uniform[0, 10000]
$\beta \sim$ Uniform[0, 10000]

File: "RMarkdown Part3b_a.Rmd"

Outline

Spatial Model With Farm Size

We will now consider a model with farm size (number of animals) included as a susceptibility covariate

For this model:

Susceptibility:	$\mathcal{S}(i)$	$=$	$\alpha_0 + \alpha_1 X_A(i)$
Transmissibility:	$\mathcal{T}(j)$	$=$	1
Infection Kernel:	$\kappa(i,j)$	$=$	$d_{ij}^{-\beta}$
Sparks:	$\epsilon(i,t)$	$=$	0

Spatial Model With Farm Size

So our model is:

$$P(i,t) \quad = \quad 1 - \exp\left(-(\alpha_0 + \alpha_1 X_A(i)) \sum_{j \in I(t)} d_{ij}^{-\beta}\right)$$

where

d_{ij} is the spatial distance between individuals i and j
α_0 is a baseline infectivity parameter
α_1 is the farm size effect
β is a spatial parameter
$X_A(i)$ is the number of animals on farm i

Priors

We now have 3 parameters to estimate: $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta)$

Once again we will assume we have no substantive prior knowledge about are the parameters we wish to include in our analysis.

Using the same idea as before, we will put uniform priors on the parameters with a wide range:

$$\begin{aligned} \alpha_0 &\sim \text{Uniform}[0, 10000] \\ \alpha_1 &\sim \text{Uniform}[0, 10000] \\ \beta &\sim \text{Uniform}[0, 10000] \end{aligned}$$

File: “RMarkdown Part3b_b.Rmd”

Outline

Spatial Model with Farm Size and Sparks term

Let us now add in a sparks term

This can be used to model infections coming in from outside the observed population

For this model:

Susceptibility:	$\mathcal{S}(i)$	$=$	$\alpha_0 + \alpha_1 X_A(i)$
Transmissibility:	$\mathcal{T}(j)$	$=$	1
Infection Kernel:	$\kappa(i,j)$	$=$	$d_{ij}^{-\beta}$
Sparks:	$\epsilon(i,t)$	$=$	ϵ

Spatial Model with Farm Size and Sparks term

So our model is:

$$P(i,t) \quad = \quad 1 - \exp\left(- \left[(\alpha_0 + \alpha_1 X_A(i)) \sum_{j \in I(t)} d_{ij}^{-\beta} + \epsilon \right] \right)$$

where

- d_{ij} is the spatial distance between individuals i and j
- α_0 is a baseline infectivity parameter
- α_1 is the farm size effect
- β is a spatial parameter
- $X_A(i)$ is the number of animals on farm i
- ϵ is a constant sparks parameter

Priors

We now have 4 parameters to estimate: $\boldsymbol{\theta} = (\alpha_0, \beta, \alpha_1, \epsilon)$

Once again we will assume we have no substantive prior knowledge about are the parameters we wish to include in our analysis.

Using the same idea as before, we will put uniform priors on the parameters with a wide range:

$$\begin{aligned} \alpha_0 &\sim \text{Uniform}[0, 10000] \\ \alpha_1 &\sim \text{Uniform}[0, 10000] \\ \beta &\sim \text{Uniform}[0, 10000] \\ \epsilon &\sim \text{Uniform}[0, 10000] \end{aligned}$$

File: “RMarkdown Part3b_c.Rmd”

Outline

Fitting SIR Models

So far we have considered fitting only SI compartmental models,
that assume infection and infectiousness continue until at least time t_{max}

If we wish to fit an SIR model we can do.

However, EpiILM does not allow us to estimate the infectious period except in a fairly limited manner.

EpiILM-CT does in the context of continuous-time models (see later)

In EpiILM, with discrete-time models we have two broad options.

Option 1: Known Removal Times

If we know the removal times for individuals, we can simply define these when carrying out our analysis.

For example, assume in our spatial disease system that we have recorded removal times.

Data set: “DATA / ENAR_SIR_fit.csv”:

SIR Simple Spatial Model

For this model:

Susceptibility:	$\mathcal{S}(i)$	$=$	α
Transmissibility:	$\mathcal{T}(j)$	$=$	1
Infection Kernel:	$\kappa(i,j)$	$=$	$d_{ij}^{-\beta}$
Sparks:	$\epsilon(i,t)$	$=$	0

So our model is:

$$P(i,t) \quad = \quad 1 - \exp\left(-\alpha \sum_{j \in I(t)} d_{ij}^{-\beta}\right)$$

where:

d_{ij} is the spatial distance between individuals i and j
α is an infectivity parameter
β is a spatial parameter

Priors

Once again, we have two parameters: $\boldsymbol{\theta} = (\alpha, \beta)$

Let's assume we have no substantive prior knowledge about them we wish to include in our analysis.

Once again, put uniform priors on the parameters with a wide range:

$\alpha \sim$ Uniform[0, 10000]
$\beta \sim$ Uniform[0, 10000]

File: "RMarkdown Part3b_d.Rmd"

Option 2: Unknown Constant Infectious Period

We can fit an SIR model in the situation that we do not know the removal times, but we assume the infectious period is the same for all individuals.

For example, say we want to test if the infectious period is 1, 3 or 5 time units long.

We can fit different models to the data in which varying assumptions are made about the infectious period, and then compare them (e.g., using the DIC)

Let us also assume that we don't know whether to include a Sparks term or not.

Then we have 6 models to fit

Comparing Models

We can then compare these models using the DIC for each fitted model:

Infectious period	Without sparks term	With sparks term
1	NA	166.4
3	**154.7**	156.1
5	161.7	163.0

So, the model without the sparks term and an infectious period of 3 time units seems to fit the data best

Note: a model with an infectious period of 1 for all indivdiuals is so inconsistent with the data that we get a likelihood of zero

Using an SIR Model

Of course, in many situations we will not know the infection times and/or removal time with certainty

Further, we may not wish to assume that the infectious period is the same for all individuals.

In such cases, we can treat the infectious periods and/or infection times as unknown "nuisance"/latent variables

Then, a so-called Bayesian **data-augmented** analysis can be carried out
(typically, using MCMC)

Unfortunately, EpiILM cannot currently carry out such analyses, but EpiILM-CT can...

Outline

Outline

Data

Example:

Imagine we are modelling the spread of a disease through a series of farms in a region.

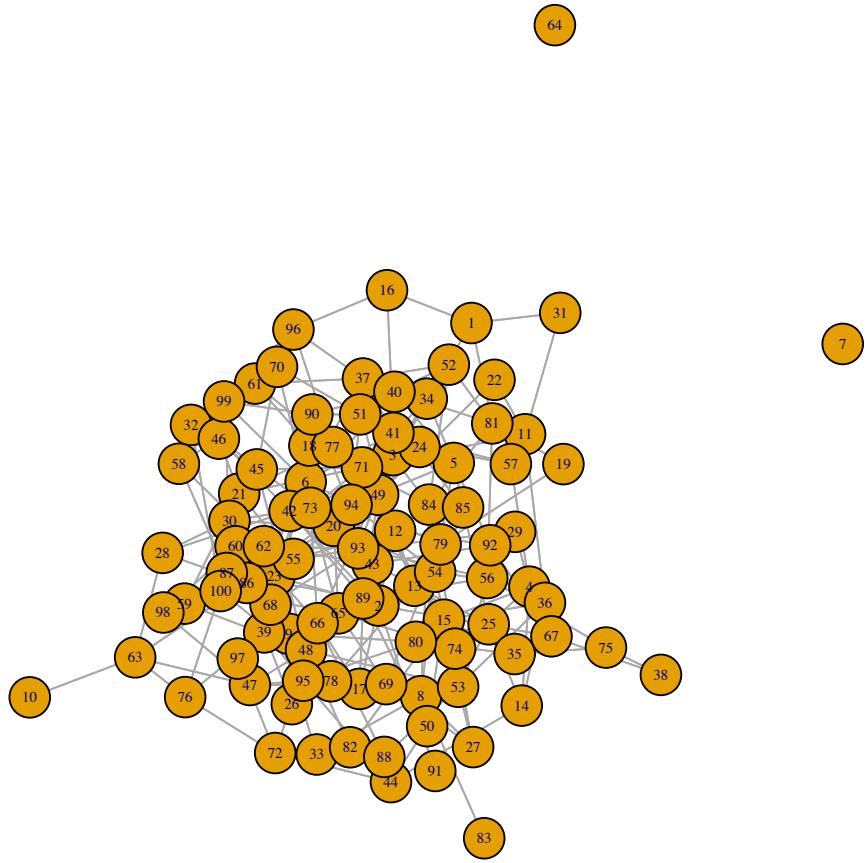
We have no individual-level covariates about the animals or farms, other than whether the farms share the same supply feed truck.

It seems reasonable to treat the farms themselves as individual level unit.

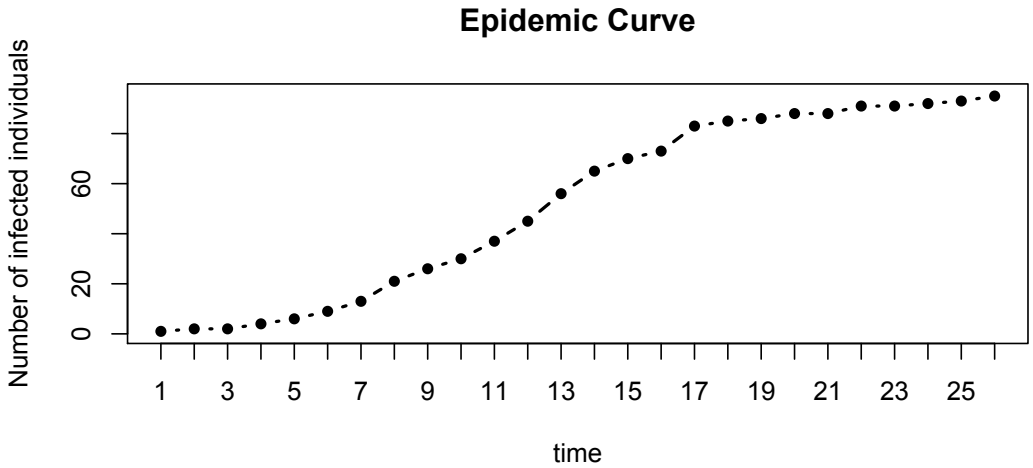
Further infection from outside the population of farms is expected to be negligible/non-existent.

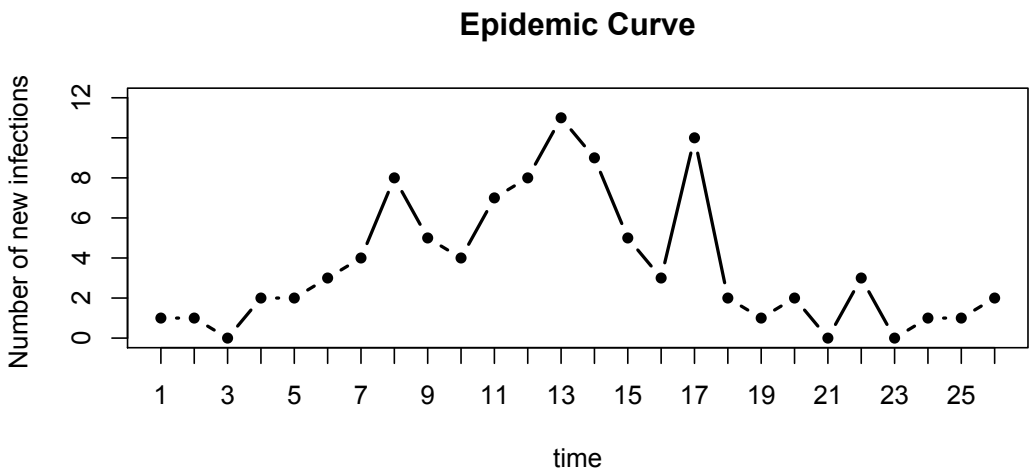
Once infected, we assume infection remains on farm.

Contact Network



Epidemic curves





Network-based model (Undirected)

We can fit the following model under SI framework

$$P(i,t) \quad = \quad 1 - \exp\left(-\alpha \sum_{j \in I(t)} c_{ij}\right)$$

where

α is an infectivity parameter

$$c_{ij} = c_{ji} = \begin{cases} 1 & \text{if } i \text{ and } j\text{share a supplier} \\ 0 & \text{otherwise} \end{cases}$$

MCMC Codes

Parameter to be estimated: α

Prior: $\alpha \sim$ Uniform$[0, 10000]$

Contact Matrix File: DATA/ENAR_network_fit.csv

Infection Times File: DATA/ENAR_network_inftime.csv

File: “RMarkdown Part3b_e.Rmd”

Outline

Posterior Predictive Forecasting

The posterior predictive distribution of new data $\tilde{\boldsymbol{D}}$, is the distribution of that data after accounting for parameter uncertainty via the posterior:

$$\pi(\tilde{\boldsymbol{D}} \mid \boldsymbol{D}) = \int_{\Theta} L(\tilde{\boldsymbol{D}} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{D}) \ d\boldsymbol{\theta}$$

Algorithmically we can produce realization of $\tilde{\boldsymbol{D}}$ very simply:

1. Sample parameter(s) $\tilde{\boldsymbol{\theta}}$ from the posterior
2. Simulate data $\tilde{\boldsymbol{D}}$ from model $\mathcal{M}(\tilde{\boldsymbol{\theta}})$
3. Repeat

Posterior Predictive Forecasting

Imagine we have observed data from a partially unfolding epidemic

We can fit our model to data we have observed so far, and then use a posterior predictive approach to simulate the range of possible outcomes we expect to see

This gives us information as to what the most likely course of an epidemic is to be (under the posterior predictive mean)

And also gives us a plausible range of outcomes after accounting for our parameter uncertainty

File: “RMarkdown Part3b_e.Rmd” (again)

Outline

Posterior Predictive Forecasting

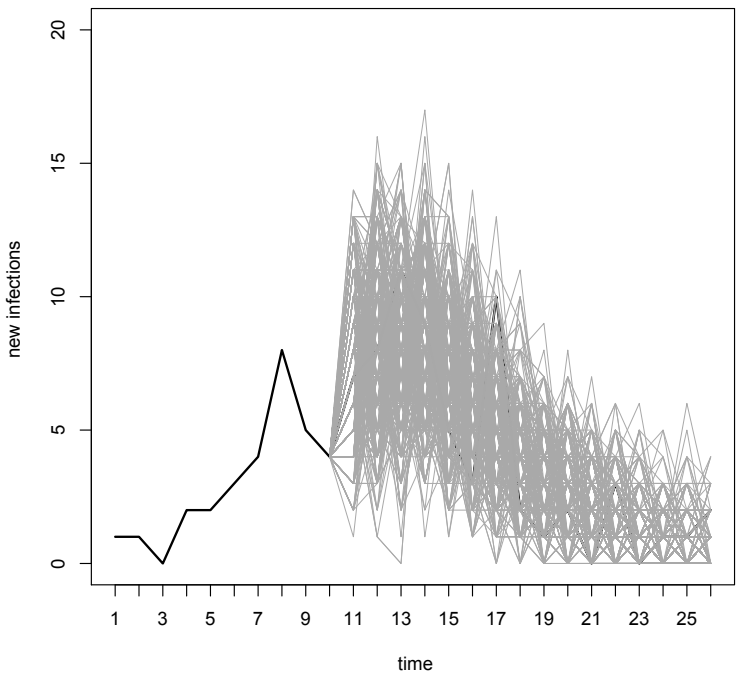
We have seen how we can fit our model to data we have observed so far, and then use a posterior predictive approach to simulate the range of possible outcomes we expect to see

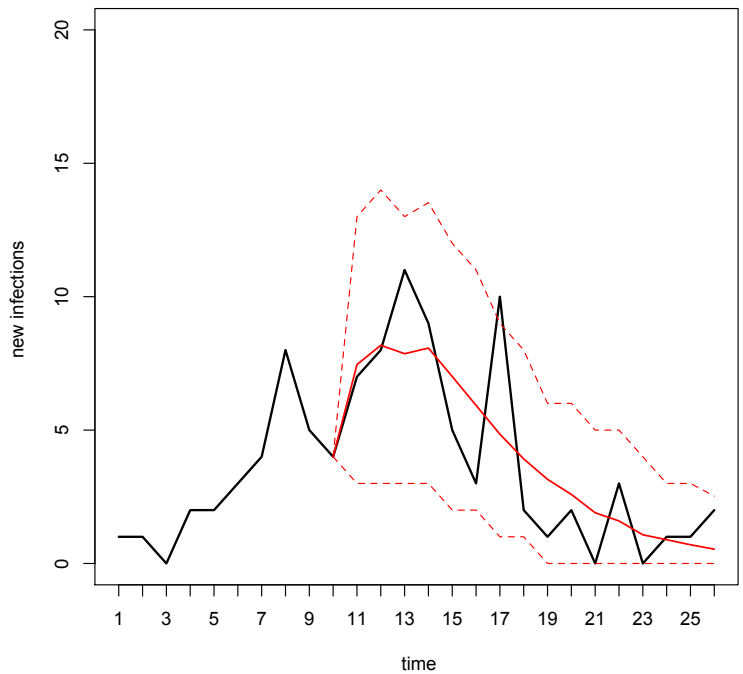
We can also carry out such simulation while imposing various control strategies upon the unfolding epidemic, in order to ascertain their likely practical usefulness/costs

Here, we test the very simple (and unrealistic) strategy of imposing quarantine/culling on a random proportion of susceptible farms at a given point in time (here, $t = 10$).

Prediction from time point 10 (No Quarantine)

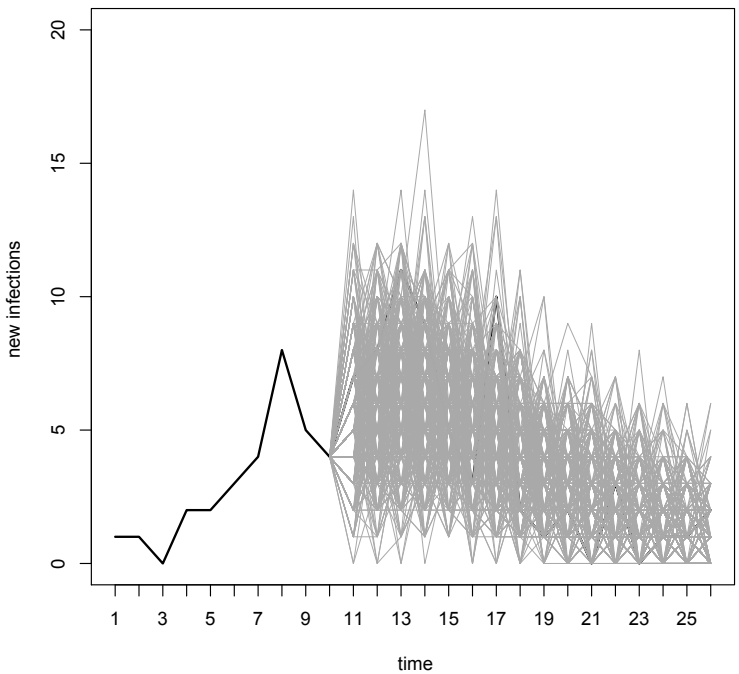
True (black) with 500 posterior predictions (gray)

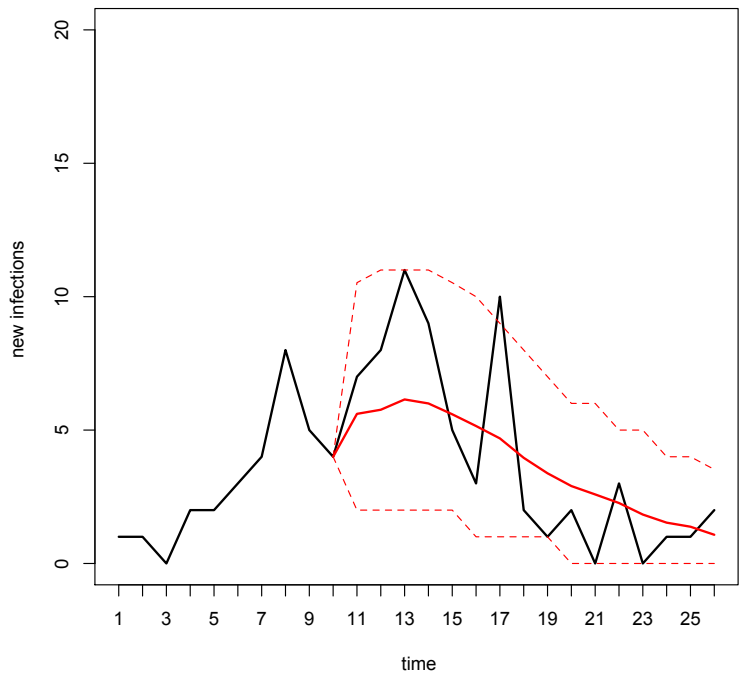




Testing Control Policy from time point 10 (25% quarantined)

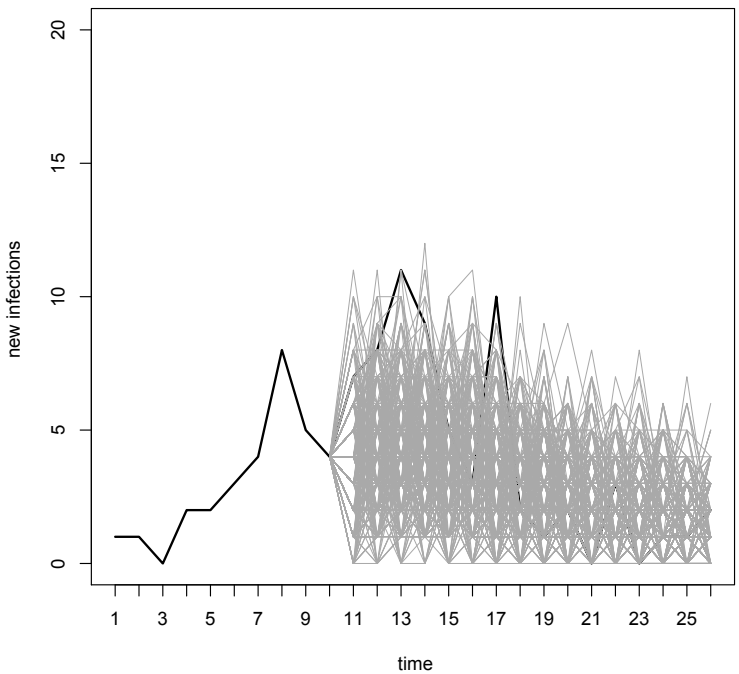
True (black) with 500 posterior predictions (gray)

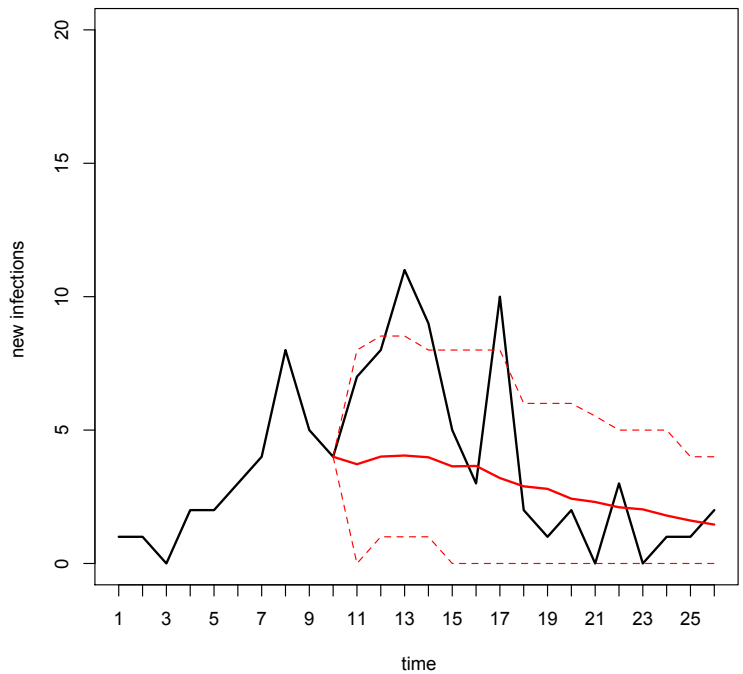




Testing Control Policy from time point 10 (50% quarantined)

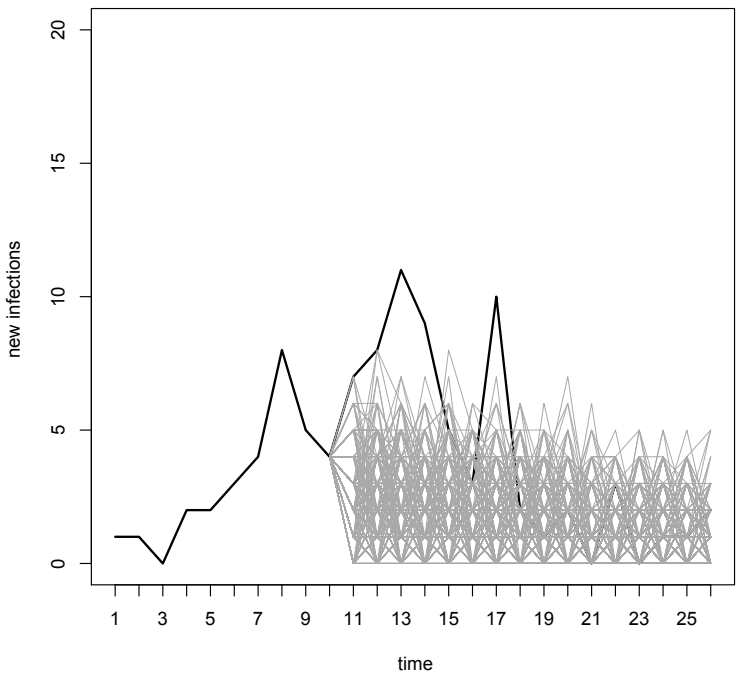
True (black) with 500 posterior predictions (gray)



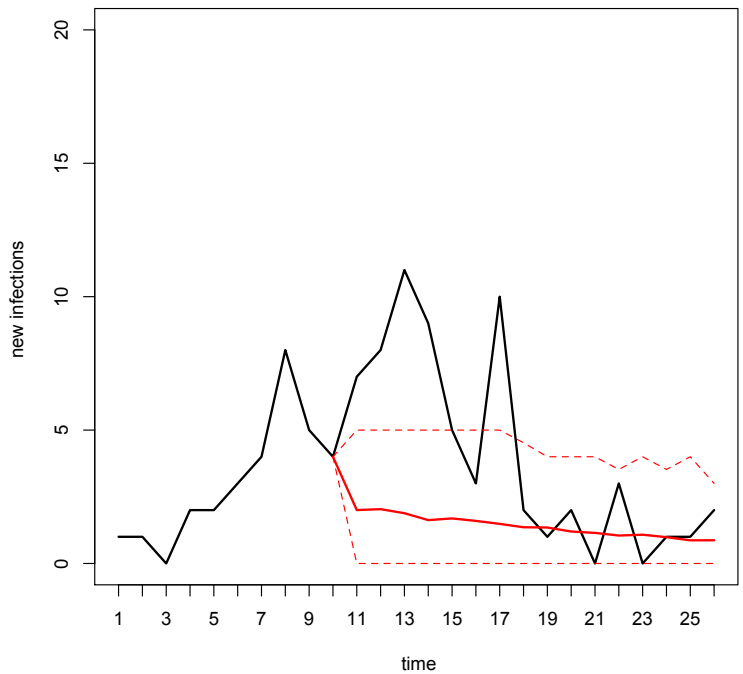


Testing Control Policy from time point 10 (75% quarantined)

True (black) with 500 posterior predictions (gray)

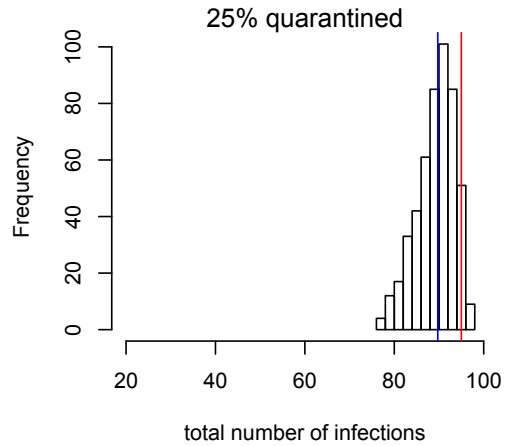


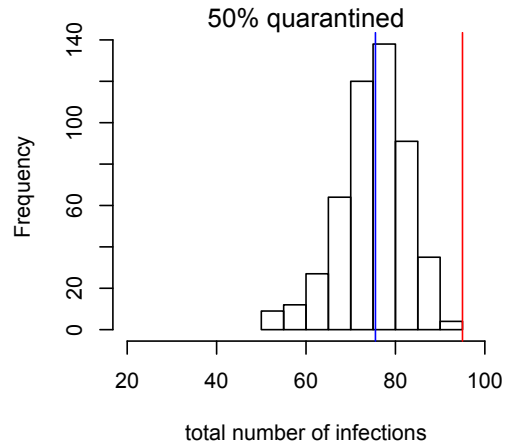
&

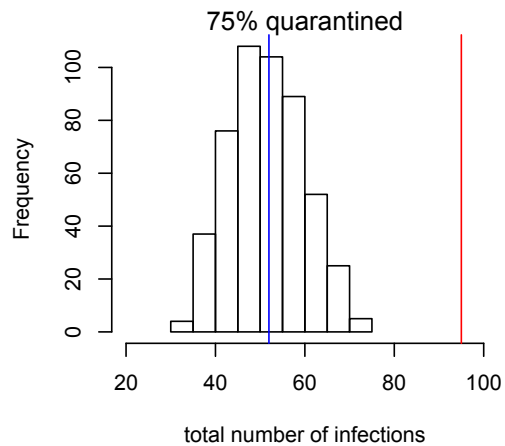


Testing Control Policy from time point 10

red:true, blue:average







Coding up quarantine/culling simulation

Coding is similar to that of the pure forecasting simulation.

Control policy can be coded as follows

For example:

We simulate an SIR rather than SI epidemic

By default, all individuals are given an infectious period greater than tmax

However, before simulation, a certain proportion (e.g., 0.25, 0.5, 0.75) of infectious periods are set equal to zero, so those individuals will enter the removed state as soon as they are infected (i.e., never infectious)

Outline

Outline

Recall: General form of the individual level model (ILM)

The **probability of susceptible individual** i **being exposed/infected at time** t is given by:

$$P(i,t) = 1 - \exp\left(-\left[\mathcal{S}(i)\left(\sum_{j \in I(t)} \mathcal{T}(j)\kappa(i,j)\right) + \epsilon(i,t)\right]\right) \tag{1}$$

where

$\mathcal{S}(i)$ is a susceptibility function for individual i
$I(t)$ is the set of infectious individuals at time t
$\mathcal{T}(j)$ is a transmissibility function for individual j
$\kappa(i,j)$ is the infection kernel
$\epsilon(i,t)$ is the random sparks function

Additionally, it is assumed that $\mathcal{S}(i), \mathcal{T}(j), \kappa(i,j)$ and $\epsilon(i,t) \geq 0$ always

Recall: General form of the individual level model (ILM)

The **probability of susceptible individual** i **being exposed/infected at time** t is given by:

$$P(i,t) = 1 - \exp\left(-\left[\mathcal{S}(i)\left(\sum_{j \in I(t)} \mathcal{T}(j)\kappa(i,j)\right) + \epsilon(i,t)\right]\right) \tag{2}$$

where

$\mathcal{S}(i)$ is a susceptibility function for individual i
$I(t)$ is the set of infectious individuals at time t
$\mathcal{T}(j)$ is a transmissibility function for individual j
$\kappa(i,j)$ is the infection kernel
$\epsilon(i,t)$ is the random sparks function

Additionally, it is assumed that $\mathcal{S}(i), \mathcal{T}(j), \kappa(i,j)$ and $\epsilon(i,t) \geq 0$ always

Positivity of Coefficients

Strictly, we have a mathematical requirement that:

$$\mathcal{S}(i)\left(\sum_{j\in I(t)}\mathcal{T}(j)\kappa(i,j)\right)+\epsilon(i,t)\geq 0$$

This requirement is necessary so that our $P(i,t)$ are actually probabilities, bounded between zero and one.

It seems reasonable to achieve this by assuming:

$$\mathcal{S}(i)\geq 0$$
$$\mathcal{T}(j)\geq 0$$
$$\kappa(i,j)\geq 0$$
$$\epsilon(i,t)\geq 0$$

Positivity of Coefficients

To help achieve this in EpiILM we require all individual model parameters/coefficients to be **positive**

So, for example, in the susceptibility function:

$$\mathcal{S}(i) = \alpha_0 + \alpha_1 X_1(i) + \alpha_2 X_2(i) + \ldots + \alpha_{n_S} X_{n_S}(i)$$

we require:

$$\alpha_0, \ldots, \alpha_{n_S} \geq 0$$

So, if our covariate data are also positive ($X_1, X_2, \ldots \geq 0$) this ensures $\mathcal{S}(i) \geq 0$

Positivity of Coefficients

Similarly:

For the spatial parameter in the spatial infection kernel:

$$\kappa(i,j) = d_{ij}^{-\beta}$$

we require:

$$\beta \geq 0$$

Positivity of Coefficients

For the network parameters in a network-based infection kernel:

$$\kappa(i,j) = c_{ij}^{(0)} + \beta_1 c_{ij}^{(1)} + \beta_2 c_{ij}^{(2)} + \dots$$

we require:

$$\beta_1, \beta_2, \dots \geq 0$$

And finally, the sparks term:

$$\epsilon(i,t) = \varepsilon$$

we require:

$$\varepsilon \geq 0$$

Positivity of Coefficients – Modelling Considerations

Example: Susceptibility

Consider an example in which we wanted to account for gender in the susceptibility function:

$$X_g(i) = \begin{cases} 0 & \text{if individual } i \text{ identifies as male} \\ 1 & \text{if individual } i \text{ identifies as female} \end{cases}$$

and let:

$$\mathcal{S}(i) = \alpha_0 + \alpha_g X_g(i)$$

then α_g is the gender effect associated with the risk of contracting this disease

Positivity of Coefficients – Modelling Considerations

Since we require that $\alpha_g \geq 0$, we are assuming that

female susceptibility cannot be lower than male susceptibility

If we wish to allow for higher male susceptibility, we need to re-code the gender indicator variable, using:

$$X_g(i) = \begin{cases} 1 & \text{if individual } i \text{ is male} \\ 0 & \text{if individual } i \text{ is female} \end{cases}$$

If we are fitting such a model to data, and do not know if male or female susceptibility would likely be higher, we can fit both models and compare

Positivity of Coefficients – Modeling Considerations

Example: Susceptibility

Consider an example in which we have a continuous covariate in our susceptibility function.

E.G., If $X_A(i)$ is the number of animals on farm i, and we let:

$$\mathcal{S}(i) = \alpha_0 + \alpha_1 X_A(i)$$

then α_1 denotes the effect of increasing $X_A(i)$ by 1 animal on the risk of farm i contracting the disease

Positivity of Coefficients – Modelling Considerations

However, we are inherently assuming a positive correlation between X_A and susceptibility

Might not be true – larger farms may tend to have better biosecurity measures.

How do we model this?

Answer: once again, recode our variable and use the recoded variable in our model

Positivity of Coefficients – Modelling Considerations

There are a number of ways of recoding our variable

One possibility is to use:

$$X_A^*(i) = M - X_A(i)$$

where

$$M \geq \max\{X_A(i)\}$$

We then use a model with susceptibility function:

$$\mathcal{S}(i) = \alpha_0 + \alpha_1 X_A^*(i)$$

Positivity of Coefficients – Modelling Considerations

Once again, if we are unsure as to whether increasing our covariate will increase or decrease susceptibility

... we can fit both versions of the model and then compare them.

Note: this means if we estimate a coefficient to be zero (e.g., α_1) it may be our covariate has no effect upon susceptibility...

... or it could be a negatively correlated relationship exists, but our modelling constraints do not allow this

Note also: we can have similar issues with contact networks since their associated coefficients are also constrained to be positive

Outline

Selected Background References

Most Relevant:

V. Warriyar, ***W. Almutiry*** & **R. Deardon**. EpiILM: Spatial and Network Based Individual Level Models for Epidemics (R Package). https://CRAN.R-project.org/package=EpiILM

V. Warriyar, ***W. Almutiry*** & **R. Deardon** (2020) "Individual level modelling of infectious disease data: EpiILM" in *The R Journal* 12(1), 199-217.

R. Deardon, S. P. Brooks, B. T. Grenfell, M. J. Keeling, M. J. Tildesley, N. J. Savill, D. J. Shaw & M. E. J. Woolhouse (2010), "Inference for individual-level models of infectious diseases in large populations" in *Statistica Sinica*, 20(1), 239-261.

R. Deardon, ***X. Fang*** & ***G.P.S. Kwong*** (2015). Statistical modelling of spatio-temporal infectious disease transmission. *Analyzing & Modeling Spatial & Temporal Dynamics of Infectious Diseases*, 211-232, John Wiley & Sons. (Ed: D. Chen, B. Moulin, J. Wu).

Some Applications :

Kwong, Poljak, **Deardon** & Dewey (2013) "Bayesian analysis of risk factors for infection with a genotype of porcine reproductive and respiratory syndrome virus in Ontario swine herds using monitoring data" in *Preventive Veterinary Medicine*, 110(3-4):405-17.

Malik, **Deardon**, ***Kwong*** & Cowling (2014) "Individual-level modeling of the spread of influenza within households" in *Journal of Applied Statistics*, 41(7), 1578-1592.

Tildesley, Savill, Shaw, **Deardon**, Brooks, Woolhouse, Grenfell & Keeling (2006). Optimal reactive vaccination strategies for an outbreak of foot-and-mouth disease in Great Britain. *Nature*, 440, 1080, 83-86.

Z. Liu, **R. Deardon**, Y. Fu, ***T. Ferdous***, T. Ware & Q. Cheng (2021)"Estimating parameters of two-level individual-level models of the COVID-19 epidemic using ensemble learning classifiers" in *Frontiers in Physics*, 8(11), Article 602722. doi: 10.3389/fphy.2020.602722

I.Vrbik*, R. Deardon**, Z. Feng, ***A. Gardner & J. Braun (2012) "Using individual-level models to model the spatio-temporal dynamics of combustion" in *Bayesian Analysis*, 7(3), 615 – 638.

Research Directions

Continuous-time models and event-time uncertainty

W. Almutiry, ***V. Warriyar*** & **R. Deardon**. EpiILMCT: Continuous Time Individual-Level Models of Infectious Disease. https://cran.r-project.org/web/packages/EpiILMCT/index.html

W. Almutiry, ***V. Warriyar*** & **R. Deardon** (2021) "Continuous-time individual-level models of infectious disease: EpiILMCT" in the *Journal of Statistical Software*, 98(10), 1-44. https://www.jstatsoft.org/article/view/v098i10

Computational Approximation

G. Pokharel & **R. Deardon** (2022) "Emulation-based inference for spatial infectious disease transmission models incorporating event time uncertainty" in the *Scandinavian Journal of Statistics*, 49(1), 455-479. http://doi.org/10.1111/sjos.12523

M. Ward, L. Deeth & **R. Deardon** (2022) "Cluster-aggretion-disaggregation methods for spatial individual level models of infectious disease transmission" in *Spatial & Spatiotemporal Epidemiology*, 41: 100497. https://doi.org/10.1016/j.sste.2022.100497

Z. Liu, **R. Deardon**, Y. Fu, ***T. Ferdous***, T. Ware & Q. Cheng (2021)"Estimating parameters of two-level individual-level models of the COVID-19 epidemic using ensemble learning classifiers" in *Frontiers in Physics*, 8(11), Article 602722. doi: 10.3389/fphy.2020.602722

C. Augusta, **R. Deardon** & G. Taylor (2019) "Deep learning for supervised classification of spatial epidemics" in *Spatial & Spatiotemporal Epidemiology*, 29, 187-198.

R. Malik, **R. Deardon** & ***G.P.S. Kwong*** (2016) "Parameterizing spatial models of infectious disease spread using sampling-based likelihood approximations" in *PLoS One*, 11(1): e0146253. doi: 10.1371/journal.pone.0146253.

G.P.S. Kwong **& R. Deardon** (2012) "Linearized forms of individual-level models for large-scale spatial infectious disease systems" in *Bulletin of Mathematical Biology*, 74(8), 1912 – 37.

Research Directions

Spatially-heterogeneous models

M. Mahsin, **R. Deardon** & P. Brown (2022) "Geographically-dependent individual-level models for infectious diseases transmission" in *Biostatistics*, 23(1), 1-17. https://doi.org/10.1093/biostatistics/kxaa009

L. Amiri, M. Torabi, **R. Deardon** & M. Pickles (2021). "Spatial modeling of individual-level infectious disease transmission: tuberculosis data in Manitoba, Canada" in *Statistics in Medicine*, 40(7), 1678-1704. https://doi.org/10.1002/sim.8863

Transmission networks and phylodynamics

J. Angevaare, Z. Feng & **R. Deardon** "Pathogen.jl: Infectious disease transmission network modelling with Julia" provisionally accepted to the *Journal of Statistical Software*. https://arxiv.org/pdf/2002.05850.pdf

J. Angevaare, Z. Feng & **R. Deardon** (2021) "Inference of latent event times and transmission network in individual level infectious disease models" in *Spatial & Spatiotemporal Epidemiology*, 37, 100410. https://doi.org/10.1016/j.sste.2021.100410

R. Cassidy, T. Kypraios & P. D. O'Neill (2020). "Modelling, Bayesian inference, and model assessment for nosocomial pathogens using whole-genome-sequence data" in *Statistics in Medicine*. 39(12), 1746-1765

M.Lau, G.Marion, G. Streftaris & G.Gibson (2015). A systematic Bayesian integration of epidemiological and genetic data. *PLOS Computational Biology*. https://doi.org/10.1371/journal.pcbi.

Time-varying and behavioural change models

C. Ward, **R. Deardon** & A. Schmidt (2023) "Bayesian modelling of dynamic behavioural change during an epidemic" *Infectious Disease Modelling*, 8(4), 947-963. https://doi.org/10.1016/j.idm.2023.08.002

M. Ward, **R. Deardon**, L. Deeth "A framework for incorporating behavioural change into individual-level spatial epidemic models" submitted to the *Canadian Journal of Statistics* (revision requested). http://arxiv.org/abs/2308.00815