A Statistical Approach To Modeling Infectious Disease Spread (in an emergency outbreak)

Rob Deardon

Department of Mathematics & Statistics,
University of Guelph

19 Sept, 2011

Statistical modelling

- Process of developing a mathematical model **directly from data** (e.g. observed epidemic)
- Model will generally be simple
- Aim is to produce a model which **best describes the system**...
- ...but also accounts for **underlying uncertainties** we have about the data and/or system being modelled.

Statistical modelling

- Use in an emergency situation (e.g. FMD outbreak):
  - To identify key risk factors
  - To predicted future course/intensity of epidemic
  - To test competing control strategy options
- Quick fitting (and testing) of a model can be essential in an emergency outbreak situation

Sources of uncertainty

- Three sources of uncertainty:
  - unobserved covariates
    - Solution: include quantifiable stochastic element to our model (e.g. error term)
  - dates of infection / recovery
    - Solution: treat uncertain data as "nuisance" parameters
    - incorporate using data augmentation in a Bayesian Markov chain Monte Carlo (MCMC) framework
  - contact network
    - Solution: treat uncertain data as "nuisance" parameters as above; or
    - approximate the network via easy to obtain spatial information

Problem

- Data Augmented Bayesian MCMC approach is desirable
- However, it is computationally intensive!
- Sometimes, to the point of deadlock.
- This can be a major problem – especially if trying to model as outbreak unfolds.

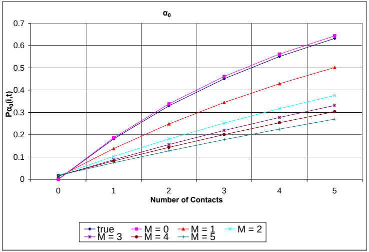
OMAFRA project goals: Model approximation

- Sometimes the model we wish to fit may take too long to fit to data (computational issues)
- It is often possible to fit ‘approximate’ models that can be fitted to data much more quickly
- Questions to be addressed:
  - how much information is lost in these ‘approximate’ models?
  - will these models still be practically useful?
  - under what circumstances?

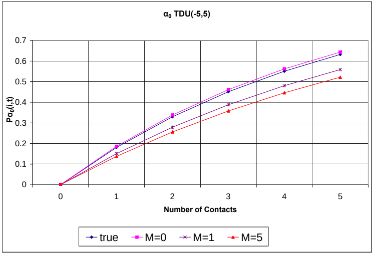
Example results: Infection Time Uncertainty

- Consider infection passing through a network
- We simulate an epidemic via a simple model through the network
- We then take this data and add noise (M) to the recorded infection times
- We then fit the original model back to this data:
  - Case 1: assuming infection times are correctly recorded
  - Case 2: incorporating infection time uncertainty

SEM Probabilities of Infection



(a) Fitting model to observed noisy data



(b) Fitting model accounting for noisy data

Acknowledgements

- Work funded by:
  - Ontario Ministry of Agriculture, Food & Rural Affairs (OMAFRA)
  - Natural Sciences & Engineering Council of Canada (NSERC)
  - Canadian Foundation for Innovation (CFI)

References / Project Publications

Deardon et al. (2010). "Inference for individual-level models of infectious diseases in large populations". *Statistica Sinica*, 20(1), p. 239–262

Gold, Deardon, & Feng. "Effects of Time–line Uncertainty upon Parameterizing Infectious Disease Models under a SIR Framework" submitted to *Computational Statistics and Data Analysis*

Gardner, Deardon & Darlington (2011) "Goodness-of-fit measures for individual-level infectious disease models in a Bayesian framework" to appear in *Spatial and Spatio-temporal Epidemiology*

Kwong & Deardon "Linearized forms of individual-level models for large-scale spatial infectious disease systems" submitted to *Bulletin of Mathematical Biology*