Inference for Epidemic Network Models from Influenza Count Data

Razvan Romanescu & Rob Deardon

Department of Mathematics and Statistics, University of Guelph, ON Canada

Contact Information:
Razvan Romanescu
Email: rromanes@uoguelph.ca

Introduction

- Network models are an important class of models in epidemiology, which can produce realistic predictions on disease systems, for the case of heterogeneous populations.
- There exists a rich mathematical literature on the spread of disease on networks, both equation based (e.g. Volz (2007), Newman(2002)) and stochastic (e.g. Stack et al. (2012)).
- Although theoretical modeling of the subject is well developed, there seems to be a lack of statistical methods for inferring the parameters of such models from data. Some work in this sense has been undertaken in Groendyke et al. (2011) and Stack et al. (2012), but the results presented are either limited to a certain network type (in the former) or inference is simulation intensive and doesn't take advantage of the powerful analytical results in the field (in the latter).
- The method we propose uses the pgf approach developed, among others, in Volz (2007) and Newman (2002), and which traces its roots from percolation theory. We develop a week ahead prediction model which uses epidemiological quantities of interest estimated directly from data.
- We illustrate our method to both simulated data and an ILI dataset from the CDC influenza surveillance program, for various degree distributions.

Main Objectives

1. Build a prediction model for hospital visits in the next week given hospital visits observed in the current week.
2. Recover parameters of the model developed above, first from simulated epidemic and network data, then from real ILI surveillance data. Discuss suitability and performance of three commonly used degree distributions with regard to modelling and inference.

Materials and Methods

- The underlying population is assumed to be part of a network constructed according to the configuration model (Molloy and Reed (1998)). One key feature of this network model is that given the degree distribution of the nodes, contacts are then established at random.
- Assuming the pre-existing contact network, the spread of disease is set in a discrete time SIRS framework, with the following simplifying assumptions, specific to influenza:
  - The time step is one week (to match the data collection frequency).
  - Incubation and infectious periods are fixed, and cumulatively last one week.
  - Following infection, the individual is temporarily immune to the whole spectrum of ILI strains, with length of immunity distributed as Exponential with rate ν.

Define:

- K_t is the degree of a random susceptible at time t;
- p_t^k is the proportion of nodes of degree k in the entire population at time t;
- S_t is the fraction of the susceptible population, and
- S_t^* is the fraction of the susceptible population with at least one connection.

Then we have (Volz (2007)) for all $k \geq 0, p_t^k = p_0^k \theta_t^k$, so that the proportion of nodes of degree k gets "depleted" at rate θ_t^k, i.e. k times faster than nodes of degree 1. It can also be shown that

$$S_t = g(\theta_t),$$

where g is the p.g.f. of K_0. If we now denote K_t^* as the degree of a random infected, then the number of new cases arising from this individual at time t is Binomial$(n = K_t^* - 1, p = \alpha S_t)$, since, on average, only a fraction S_t of his contacts will still be susceptible. From this, we can derive the expansion factor X_t, i.e. the number of secondary cases from an infected individual, which has mean and variance

$$E(X_t) = \alpha S_t^* \theta \frac{g''(\theta)}{g'(\theta)}, \qquad \text{where} \quad \theta = g^{-1}(S_t), \tag{1}$$

and

$$Var(X_t) = \alpha S_t^* \theta \frac{g''(\theta)}{g'(\theta)} \left(\alpha S_t^* \theta \frac{g'''(\theta)}{g''(\theta)} + 1 - \alpha S_t^* \theta \frac{g''(\theta)}{g'(\theta)} \right). \tag{2}$$

Observation model and approximate likelihood calculation

- For true counts, we have $I_{t+1}|I_t = \sum_{j=1}^{I_t} X_{t,j} \sim N(\mu_t I_t, \sigma_t^2 I_t)$, where μ_t and σ_t^2 are $E(X_t)$ and $Var(X_t)$, respectively.
- Assume that we observe counts $Y_t|I_t \sim N(\rho I_t, \psi \rho I_t)$, where ψ is an overdispersion factor as in Ionides et al. (2006).
- Assume also that given the observed counts, the true are $I_t|Y_t \sim N(\rho^{-1}Y_t, \psi\rho^{-2}Y_t)$.
- Under these assumptions we can show that

$$E(Y_{t+1}|Y_t) = \mu_t Y_t$$

$$\text{and} \qquad Var(Y_{t+1}|Y_t) = \mu_t Y_t \left[\frac{\sigma_t^2}{\mu_t}\rho + \psi + \psi\mu_t \right].$$

- Given data $y_t,\ t = 0, ..., T$ on the number of doctor's visits for each discrete time period , we can compute the likelihood for parameter set $\boldsymbol{\theta}$ as

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{t=0}^{T-1} f(y_{t+1}|y_t), \text{and} \quad y_{t+1}|y_t \sim N(y_t E(X_t), \rho y_t Var(X_t),$$

where $y_{t+1}|y_t$ is normal, with mean and variance derived above.

- We still need an estimate of S_t^* (note that $S_t = S_t^* + p_0^0$). We estimate this at each time step as

$$S_{t+1}^* = S_t^* - \frac{y_t}{K} + \nu(1 - S_t^*),$$

where K is the number of visits per week if everyone in the population (with degree at least one) were infected.

Network degree distributions

- The methodology presented above works for any degree distribution. In practice, a few distributions have received special attention.
  - The Poisson degree distribution with mean λ, with

$$P(K_0 = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \geq 0$$

  - The discrete Exponential distribution with mean λ

$$P(K_0 = k) = (1 - e^{-1/\lambda})e^{-k/\lambda}, \quad k \geq 0$$

  - The power law (scale-free) distribution with an upper cutoff

$$P(K_0 = k) \propto k^{-\lambda}, \quad k = 1, 2, ..., k_{max}.$$

- For some distributions, calculations simplify and we can specify an exact distribution for the expansion factor:
  - For the Poisson distribution, $X_t \sim \text{Poisson}(\lambda' = \alpha\lambda\theta S_t^*)$
  - For the Exponential distribution, $X_t \sim \text{NegBin}(r' = 2, p' = \frac{\alpha\theta e^{-1/\lambda} S_t^*}{1+\alpha\theta e^{-1/\lambda} S_t^* - \theta e^{-1/\lambda}})$.
  - For a mass action model, $K_0 = N - 1$, and $\alpha = \frac{\alpha'}{N-1}$. When $N \to \infty$, (and α' is held constant), $E(X_t) = Var(X_t) = \alpha' S_t^*$.
- In the scale-free case, there are no simplifications, and so we use

$$g(z) = \frac{\sum_{k=1}^{k_{max}} \frac{z^k}{k^\lambda}}{\sum_{k=1}^{k_{max}} \frac{1}{k^\lambda}},$$

and invert this function numerically to obtain $\theta = g^{-1}(S_t^*)$.

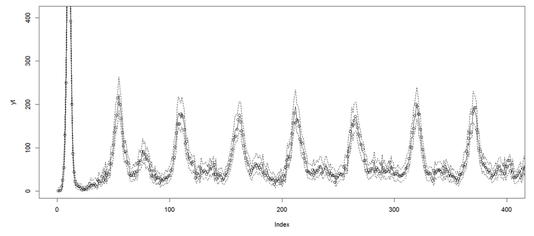

Figure 1: A typical epidemic simulation from a network with Poisson degree distribution. Points represent simulated values, the solid line is the fitted conditional mean, and dashed lines give confidence intervals for the one step ahead predictions.

Data

1. We test our approach on epidemics simulated on synthetic networks generated from the Configuration Model, for each of the three degree distributions.
   - The parameters are $\alpha_0, \lambda, \nu, \beta, K, \rho$, where
   - α_0 is the baseline transmission probability per contact, and
   - β controls the strength of seasonality, such that
   - $\alpha_t = \alpha_0 * \left(1 + \beta \sin(\frac{2\pi}{52.18}(t + \phi))\right), \quad t = 1, 2, ...T.$

   Figure 1 shows a realization of a simulated epidemic, along with the fitted conditional mean and confidence bounds.
2. We fit the different models to the outpatient surveillance data from the CDC, Region 3. This includes data aggregated from the states: Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia, from around 1998 - 2015.

Results

1. The table below shows the results of fitting the degree models to their respective simulated datasets. For reference, we include the likelihood computed using the mass action model.

Param	TRUE	Exp. m.l.e	TRUE	Power m.l.e	TRUE	Pois. m.l.e
α_0	0.2	0.124	0.2	0.366	0.2	0.183
λ	10	(set at 10)	1.5	2.06	10	(set at 10)
ν	0.05	0.065	0.05	0.055	0.05	0.057
β	0.4	0.302	0.3	0.35	0.3	0.273
K	10,000	9353	10,000	44,098	10,000	9686
log likeli.		-3684.14		-3502		-3120.15
log likeli.	(mass action)	-4178		-3142		-3134

2. Next we show the parameter estimates and log likelihood for the ILI data. The value of ρ has been set to 1.27 (derived from literature).

Param	Mass action	Exponential net	Poisson net
α_0	α_0'=1.12	0.158	0.13
λ	–	3.04	8.6
ν	0.175	0.166	0.161
β	0.151	0.142	0.143
K	1.3	2.57	1.56
ψ	19.7	19.9	20.3
ϕ	5.78	5.3	5.36
log likeli.	3406.96	3403.66	3406.47

- The exponential and Poisson degree network models fit the simulated data better than the mass action model.
- The scale-free model recovers most of its parameters well, except for K (the optimization routine overestimates susceptibe rates).
- The CDC data seem to prefer a mass action model slightly; the three models are seen to agree on many of the parameters.

Ongoing Work and Future Research

1. The likelihood function is found to be rugged, with ridges, peaks and local minima. This prevents standard errors to be estimated for the parameter m.l.e.'s in the usual way. We plan to simulate time series starting from the best fit parameters, to obtain a sampling distribution for the m.l.e.'s.
2. A uniform reporting rate ρ for the entire population is not realistic. Some age groups have more visits per infection than others. We can use the detailed CDC data to estimate a different degree distribution for each age group.

References

[1] Groendyke, C., Welch, D., & Hunter, D. R. (2011). Bayesian inference for contact networks given epidemic data. Scandinavian Journal of Statistics, 38(3), 600-616.

[2] Newman, M. E. (2002). Spread of epidemic disease on networks. Physical review E, 66(1), 016128.

[3] Stack, J. C., Bansal, S., Kumar, V. A., & Grenfell, B. (2012). Inferring population-level contact heterogeneity from common epidemic data. Journal of The Royal Society Interface, rsif20120578.

[4] Volz, E. (2008). SIR dynamics in random networks with heterogeneous connectivity. Journal of mathematical biology, 56(3), 293-310.

Acknowledgements

This research was supported by the OMAF and MRA-U of G Partnership.