Stochastic Ebola Model

Grzegorz A. Rempala

Division of Biostatistics
and
Mathematical Biosciences Institute
The Ohio State University

https://neyman.cph.ohio-state.edu

Bedlewo, June 8, 2015

Micro and Macro Systems in Life Sciences

Ebola Outbreak in Africa 2014

Index case occurred in December 2013. As of May 29,2015 there were 27,091 (lab confirmed 15,015) cases and 11,175 deaths.

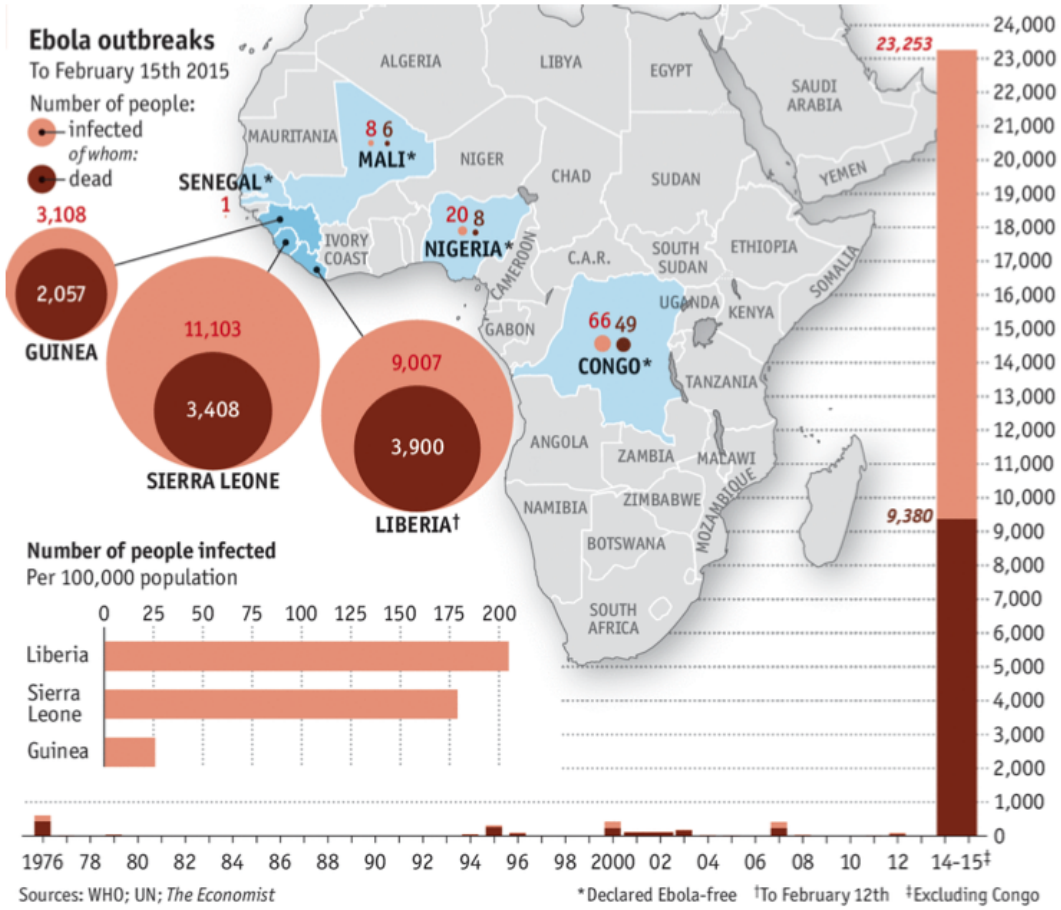
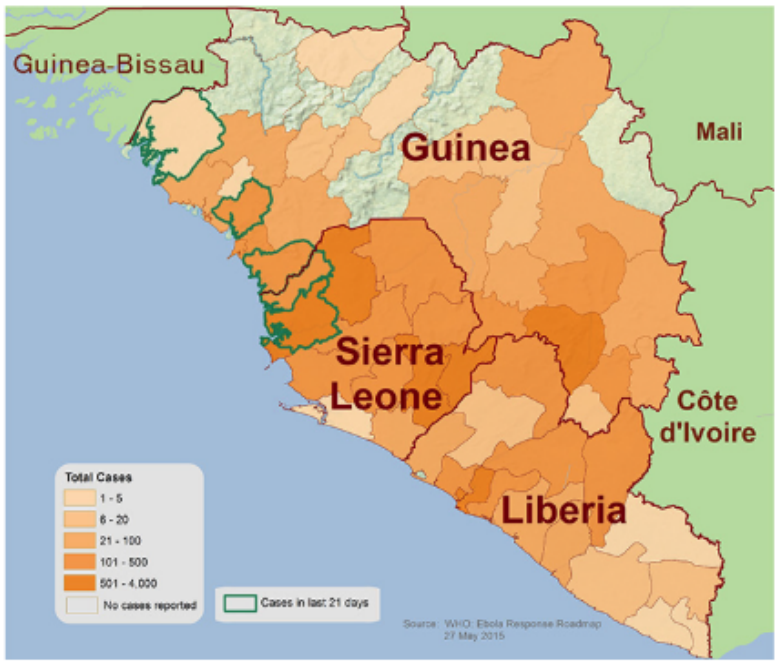


Image from http://www.economist.com/

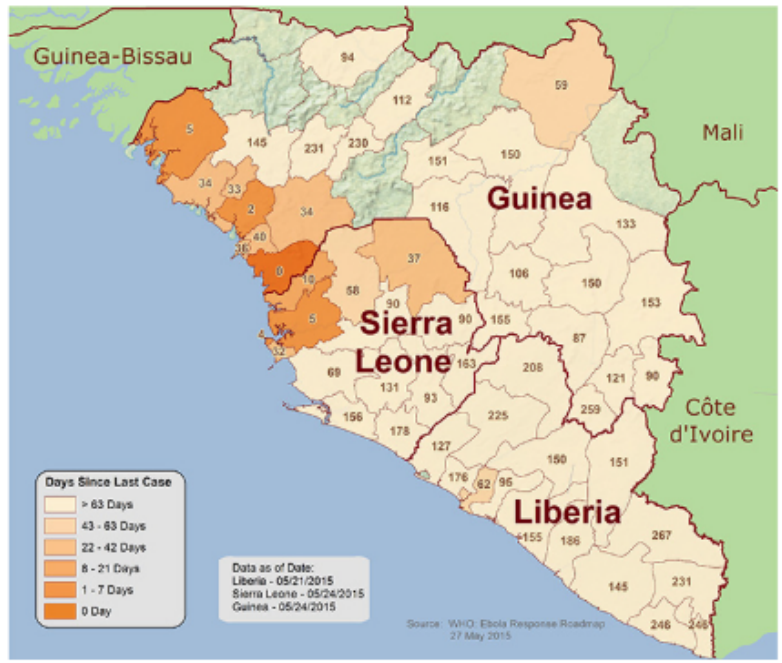
Ebola in West Africa 2014

In West Africa the epidemic was not recognized for several month. In time of model analysis (Feb /Mar 2015): confirmed cases 16,389 (11,830 lab) and 6,336 death.

Total Cases

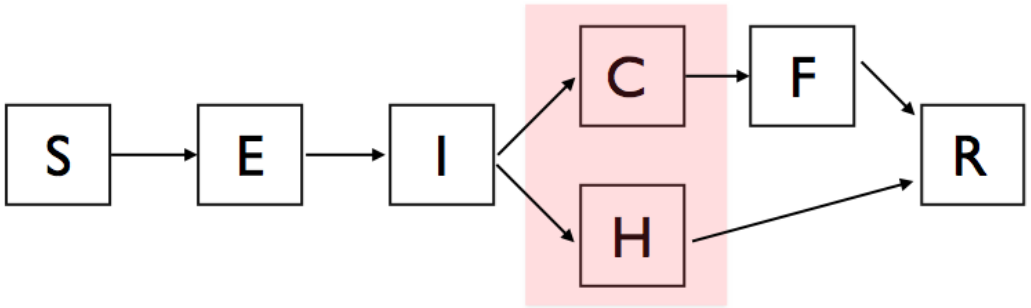


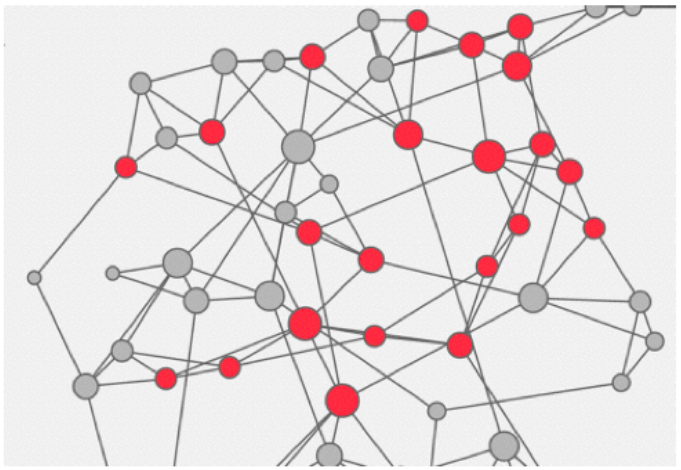
Days Since Last Case



Ebola Spread Model

The general consensus seems to be that the hospital environment and funeral arrangements are important elements. Hence the compartmental model looks like this.





- Heterogenous contact network
- Stage-dependent degree distribution
- Household vs. healthcare contacts

Markov (Complete Graph) Models

Trajectory Equation

$$\begin{aligned} X(t) &= X(0) + \sum_k R_k(t)(\nu_k' - \nu_k) \\ &= X(0) + \sum_k Y_k\left(\int_0^t \lambda_k(X(s))ds\right)(\nu_k' - \nu_k) \end{aligned}$$

- $X(t)$ state of the system at time $t > 0$
- $\nu_k' - \nu_k$ is net change due to kth interaction
- Y_k are independent unit Poisson processes with rates λ_k
- Basic assumption: the population is *uniformly mixed* (i.e., the graph is complete).

DDJP Deterministic Approximation

- If x gives the number of each species present, then $c = n^{-1}x$ gives the concentrations per unit volume
- *Density dependent jump process* (DDJP) assumes

$$\lambda_k^n(x) \approx n\alpha_k \prod_i c_i^{\nu_{ik}} \equiv n\tilde{\lambda}_k(c).$$

- The *law of large numbers* for the Poisson process implies $n^{-1}Y(nu) \approx u$,

$$C^n(t) = n^{-1}X(t) \approx C^n(0) + \sum_k \int_0^t \alpha_k \prod_i C_i^n(s)^{\nu_{ik}}(\nu'_k - \nu_k)ds,$$

when $n \to \infty$ the density process $C^n \to C$ satisfying ODE

$$\frac{dC(t)}{dt} = \sum_k \alpha_k \prod_i C_i(t)^{\nu_{ik}}(\nu'_k - \nu_k)$$

Example: SIR - Kermack and McKendrick (1927)

- Here $\theta = (\gamma, \nu)$; ODEs for susceptible (x), infectious (y), removed (z), as

$$\begin{aligned} \dot{x}(t) &= -\gamma x(t) y(t) \\ \dot{y}(t) &= \gamma x(t) y(t) - \nu y(t) \\ \dot{z}(t) &= \nu y(t) \end{aligned}$$

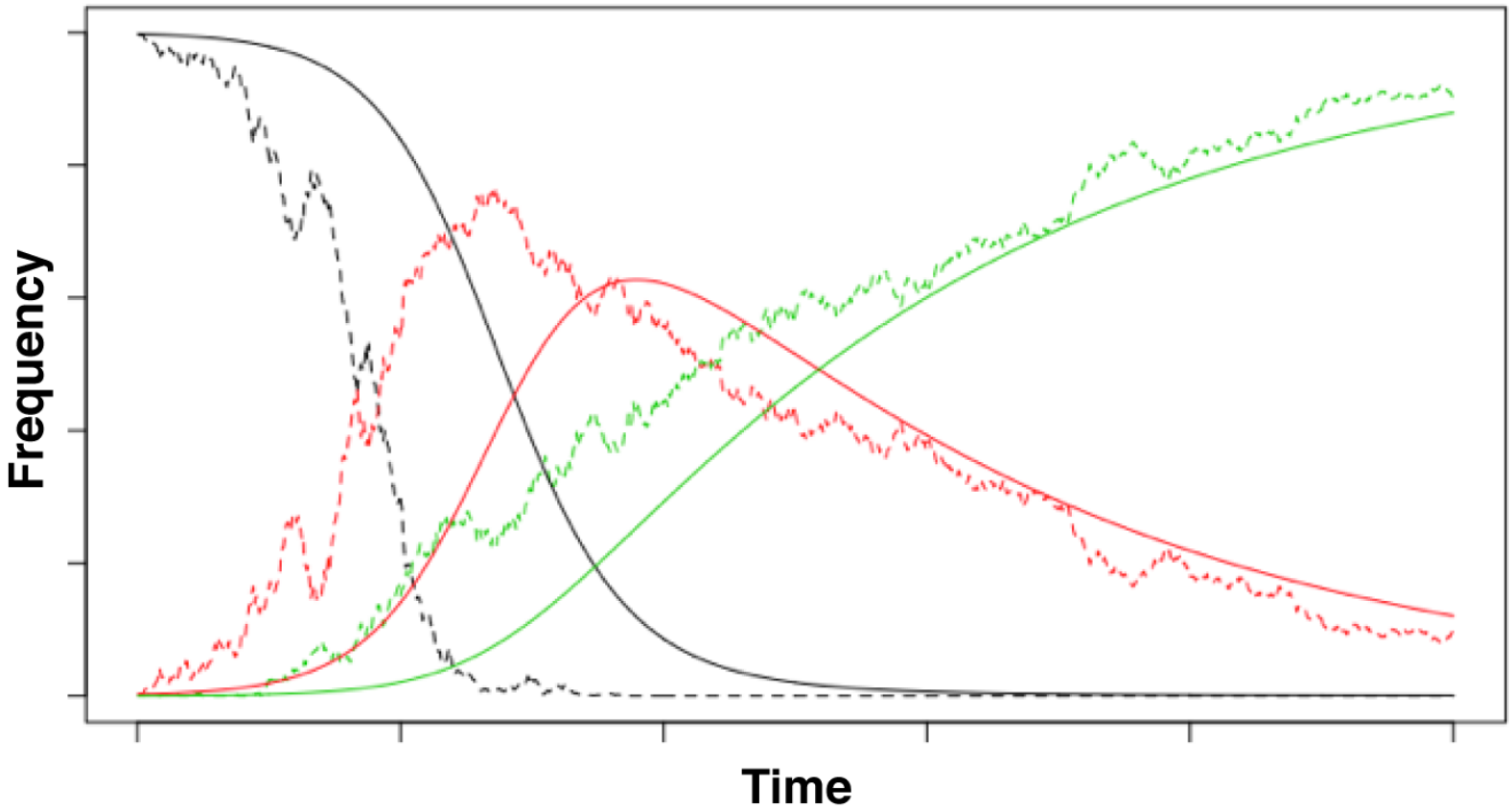
  $x(0) = 1$ and $y(0) = \tau \ll 1, z(0) = 0.$
- Basic classical model for epidemic disease spread within fixed size population $(1 + \tau)$
- Parameters: (θ, τ)

Example: SIR (cont)

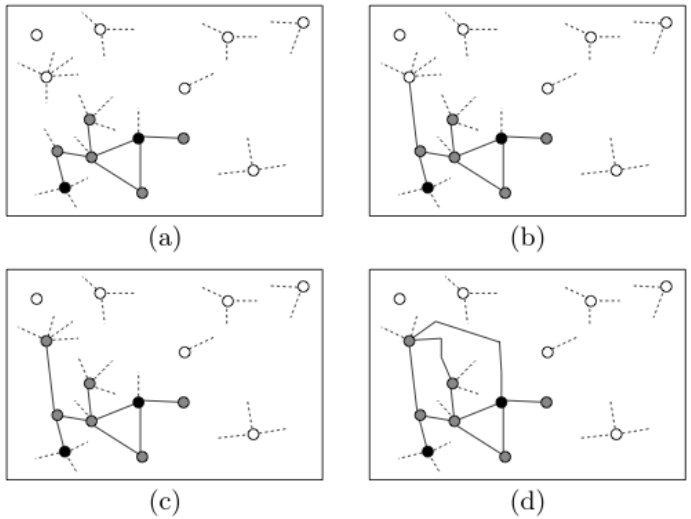
$$X + Y \xrightarrow{\lambda_1} 2Y; \quad \lambda_1 = \gamma XY \quad (\gamma = 0.5)$$

$$Y \xrightarrow{\lambda_2} Z; \quad \lambda_2 = \nu Y \quad (\nu = 0.3)$$

$$\mathcal{R}_0 = \gamma/\nu > 1 \quad P(\text{major outbreak}) > 0$$



SIR Epidemic on Configuration Model (CM) Graph



- $\mathcal{G}(q, n)$ is a CM random graph with degree distribution q (stubs (half-edges) are paired uniformly at random)
- The degree of each node is given and for each I (resp., R) we know its number of edges of type IS (resp., RS)
- A contaminating half-edge is chosen and next susceptible infected (or dropped). We then determine how many of its remaining stubs are linked to classes I, R and so on
- All events (I, R) are on independent exponential clocks

Contact Network Model for Today's Presentation

$$S \longrightarrow I \longrightarrow R$$

- Ignore (E,H,F) stages
- SIDR on CM random graph $\mathcal{G}(q, n)$ with exp. clocks:
  - γ rate of infection ($S \longrightarrow I$)
  - ν rate of recovery ($I \longrightarrow R$), but also
  - SI edges *disappear at rate* δ (decreased contacts for I's)
- Degree distribution is Poisson-type (PT): Poisson, binomial, negative binomial
- Consider a counting process for (S, I, SI)
- SI edges are needed to track I nodes connections
- SIDR on $\mathcal{G}(q, n)$ as $n \to \infty$
- IDC (infected degree cond.) $\max_{i \in I_0} \deg(i) = o(n)$

Counting Process

- Let $X(t)$ $t \geq 0$ be the counting process for (S, I, SI) starting at a non-random point $X(0) = x_0$
- $X(t)$ evolves according to the times and locations of the node events (I, R, D) occurring independently at rates (γ, ν, δ)
- NOTE: $X(t)$ is not the Markov process but $X(t) - EX(t)$ is the zero-mean *martingale*
- Let $\mathcal{A}(t) = (A_i(t))_{i=1}^n$ be the sequence of exponential clock events for (I, R, D) up to time t at the vertices of $\mathcal{G}(q, n)$;
- Then $X(t)$ satisfies the trajectory equation

$$n^{-1}X(t) = n^{-1}X(0) + \mathcal{F}(n^{-1}X(t_-), \mathcal{A}(t))$$

By the general theory $\sup_{t\in[0,T]} |X(t) - EX(t)|/n \to 0$ w.p.1

SLLN

Assume $X(0)/n = c_0 > 0$ is non random such that IDC holds. Then for any $0 < T < \infty$

$$\sup_{t\in[0,T]} \left| n^{-1}X(t) - C(t) \right| \to 0 \quad \text{as} \quad n \to \infty$$

where $C(t)$ is the unique solution of

$$C(t) = c_0 + \int_0^t \mathcal{F}(C(u), E\mathcal{A}(u))du.$$

When q is a PT distribution, we have

$$\mathcal{F}(C(u), E\mathcal{A}(u)) = \begin{pmatrix} -\gamma C_3(u) \\ \gamma C_3(u) - \nu C_2(u) \\ \kappa\gamma \frac{C_3(u)}{C_1(u)} \left(\psi'(1)C_1(u)^{2\kappa} - C_3(u)\right) - \tilde{\nu} C_3(u) \end{pmatrix}$$

where $\kappa = \psi''(1)/(\psi'(1))^2$ and $\tilde{\nu} = \gamma + \nu + \delta$.

$S(\infty)$ and $\mathcal{R}_0$

- Basic reproduction number is

$$\mathcal{R}_0 = \frac{\psi'(1)\,\kappa\,\gamma}{\gamma+\delta+\nu}$$

- Analysis of the limit of $X(t)/n$ gives the formula for the average final survivor proportion s_∞. Let $\tau = C_2(0)$, then
  - if $\kappa = 1$

$$s_\infty = \exp(-(1+\tau-s_\infty)\mathcal{R}_0)$$

  - if $\kappa > 1$ then

$$1 - s_\infty^\kappa = \mathcal{R}_0 \frac{1-\kappa}{\kappa} s_\infty^\kappa (s_\infty^{\kappa-1} - (1+\tau))$$

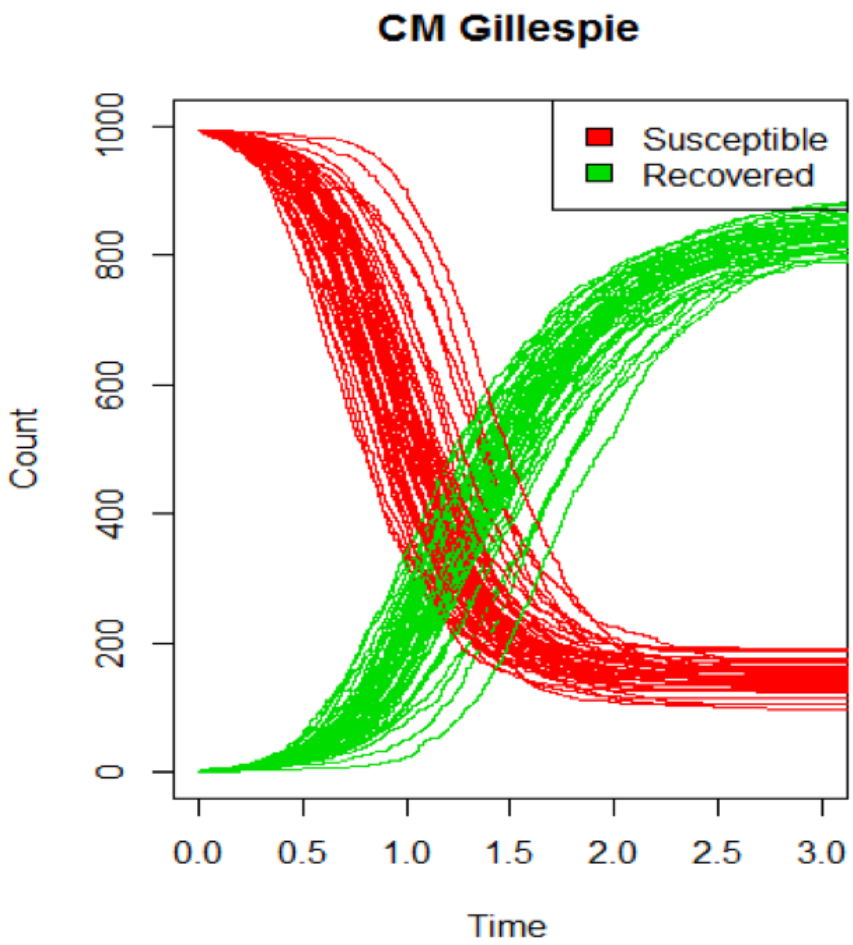
- Early stage growth: linearize around the initial condition (1,0,0) and compute max eigenvalue

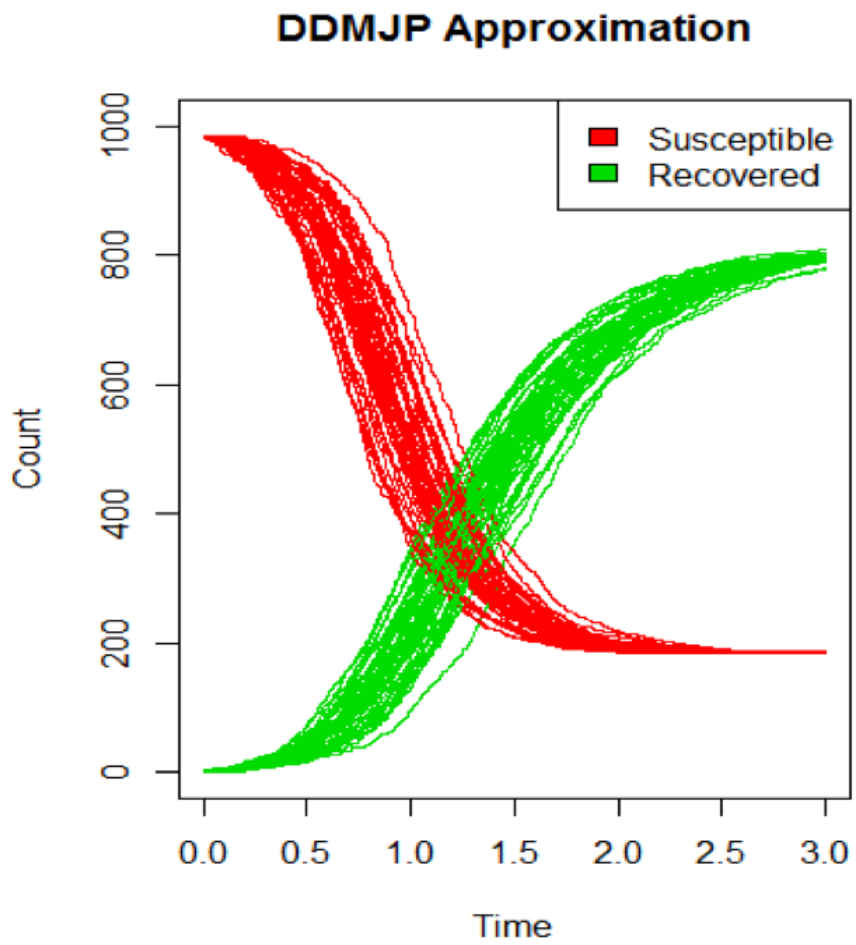
Approximate Process

- Direct analysis of the process $X(t)$ requires tracking the number of SI's
- Computationally expensive and not always feasible in practice
- Idea: approximate $X(t)$ with the space-homogenous Markov counting process $\tilde{X}(t)$ on (S, I) only
- Use the limiting equation for C_3 to update the jump rates after each new infection (e.g., with Euler's method)
- The trajectory equation is

$$\tilde{X}(t) = \tilde{X}(0) + \begin{pmatrix} -Y_1(\gamma n \int_0^t C_3(u)du) \\ Y_1(\gamma n \int_0^t C_3(u)du) - Y_2(\nu \int_0^t \tilde{X}_2(u)du) \end{pmatrix}$$

where Y_1, Y_2 are independent unit Poisson processes.


CM Gillespie
Susceptible
Recovered
Count
0
200
400
600
800
1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time


DDMJP Approximation
Susceptible
Recovered
Count
0
200
400
600
800
1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time

Likelihood Function (I)

- The approximation is "average-consistent", that is

$$\sup_{t\in[0,T]} \left| n^{-1}X(t) - \begin{pmatrix} n^{-1}\tilde{X}(t) \\ C_3(t) \end{pmatrix} \right| \to 0 \quad \text{w.p.1}$$

- Given that $\tilde{X}$ is space homogenous, it is relatively easy to write its likelihood function
- Assume we have the sequence of observed values $\boldsymbol{x}(t_i)$ $i = 1, \ldots m$ of the process $\tilde{X}$ on $[0, T]$
- NOTE: In practice, only daily counts of new infections are given
- Then ν is estimated from other sources (medical records, etc) and $\tilde{X}_2$ is imputed

Likelihood Function (II)

- Let $\theta = (\gamma, \delta, \nu, \kappa, \mu)$ (where $\mu = \psi'(1)$)
- Additional parameters estimated by *likelihood profiling*
- Let $\lambda_1(t;\theta) = \gamma n C_3(t;\theta)$ and $\lambda_2(t;\theta) = \nu \tilde{X}_2(t)$; $\lambda_0(t;\theta) = \lambda_1(t;\theta) + \lambda_2(t;\theta)$
- Let N_1, N_2 be the resp. jump count processes and $k_i \in \{1, 2\}$

$$\begin{aligned}\mathcal{L}(\theta;\boldsymbol{x}) &= \prod_{i=1}^{m} \lambda_{k_i}(t_{i-1};\theta)\exp\{-\lambda_0(t_{i-1};\theta)[t_i - t_{i-1}]\}\times \\ &\quad \times \exp\{-\lambda_0(t_m;\theta)[T - t_m]\} \\ &= \exp\left\{\sum_{i=1}^{2}\left(\int_0^T \log(\lambda_i(t-;\theta))dN_i(t) - \int_0^T \lambda_i(t-;\theta)dt\right)\right\}\end{aligned}$$

Diffusion Approximation (DA)

- Since Poisson process can be approximated by BM

$$\frac{Y(nu) - nu}{\sqrt{n}} \approx W(u),$$

- replacing $Y_k(nu)$ by $\sqrt{n}W_k(u) + nu$

$$\begin{aligned} C^n(t) &= C^n(0) + \sum_k n^{-1} Y_k\Big(\int_0^t \lambda_k(X^n(s))ds\Big)(\nu_k' - \nu_k) \\ &\approx C^n(0) + \sum_k n^{-1/2} W_k\Big(\int_0^t \tilde{\lambda}_k(C^n(s))ds\Big)(\nu_k' - \nu_k) \\ &\quad + \int_0^t \sum_k \tilde{\lambda}_k(C^n(s))(\nu_k' - \nu_k)ds, \end{aligned}$$

where $W_k(\int_0^t \tilde{\lambda}_k(C^n(s))ds) \sim \int_0^t \sqrt{\tilde{\lambda}_k(\tilde{C}^n(s))}d\tilde{W}_k(s)$ (Itô diffusion)

Model Fitting via Approximate Likelihood (LSE)

- DA allows for approximating the likd $\mathcal{L}$ by Gaussian one
- Transition kernels recovered by simulation (Linder and R 2015)
- Fitting LSE and running bootstrap simulations to get the correct SE's
- So far, tested on cumulative data until mid Feb only

Table : Initial growth rate and $\mathcal{R}_0$ estimates

Country	Growth rate	$\mathcal{R}_0$
Liberia	$0.038 - 0.04$	$1.4 - 1.84$
Guinea	$0.017 - 0.020$	$1.2 - 1.42$
Sierra Leone	$0.028 - 0.03$	$1.3 - 1.63$

Complete Cumulative Data Analysis (until mid Feb 2015)

	Liberia		Guinea	
Param	Pois	NB	Pois	NB
γ	0.174	0.226	0.145	0.201
ν	0.249	0.210	0.234	0.204
Dist	2.623	$(r,p) =$ (2.219,0.392)	2.714	$(r,p) =$ (2.16,0.398)
n/N	0.0153	0.0290	0.006	0.013
	Sierra Leone			
γ	0.129	0.169		
ν	0.219	0.190		
Dist	2.952	$(r,p) =$ (2.269,0.415)		
n/N	0.011	0.021		

Results: Liberia

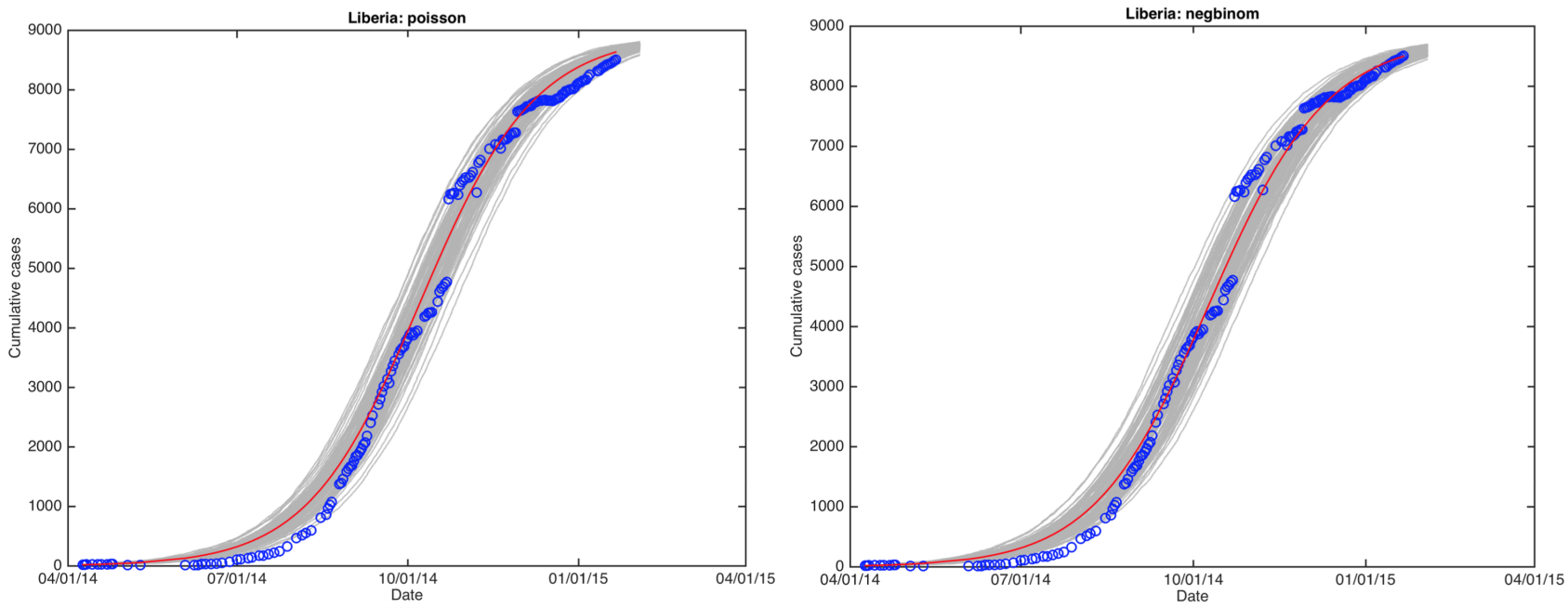


Figure : Poisson and NB Model for Liberia; $\mathcal{R}_0 = 1.7$

Results: Guinea

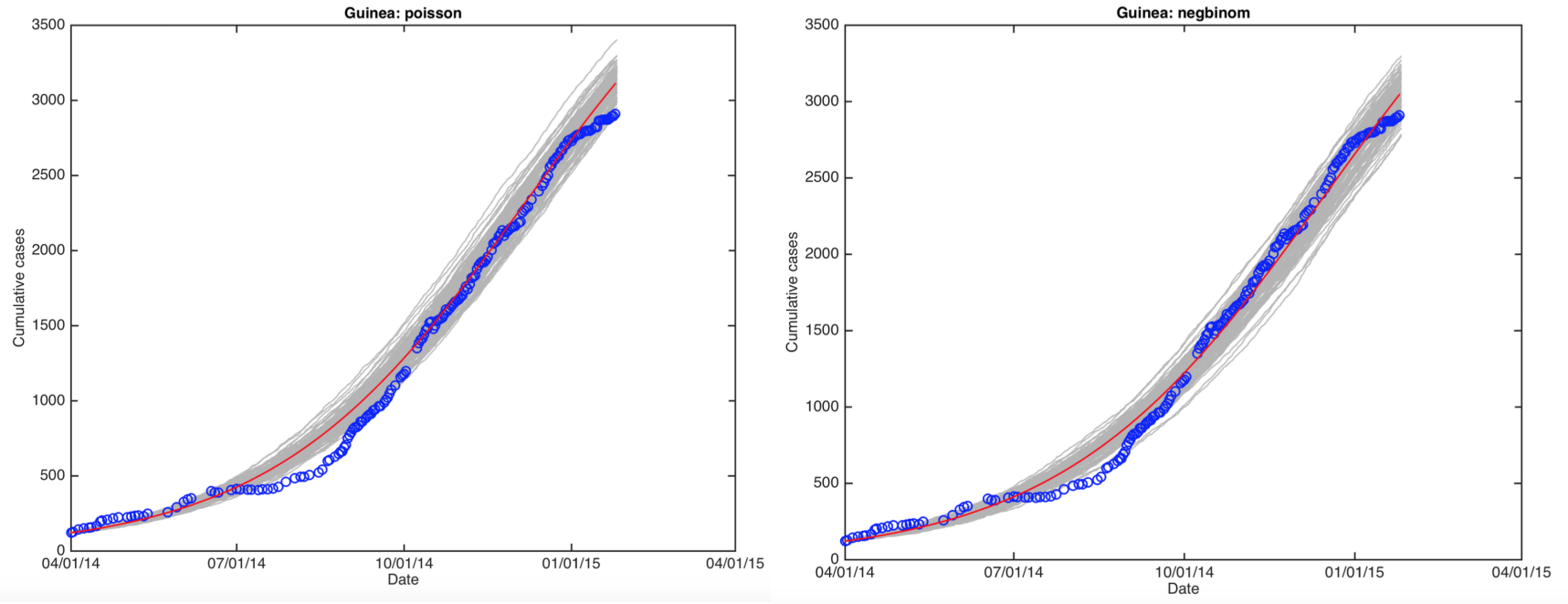


Figure : Poisson and NB Model for SL; $\mathcal{R}_0 = 1.35$

Results: Sierra Leone

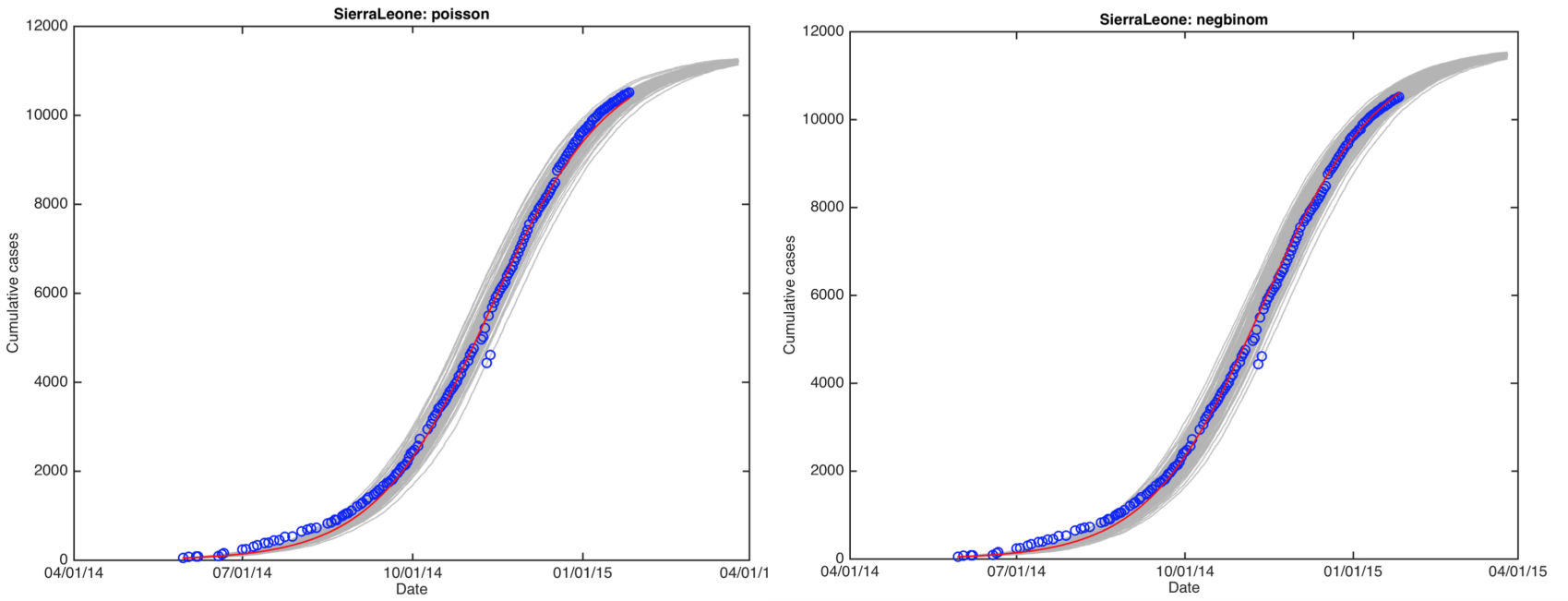
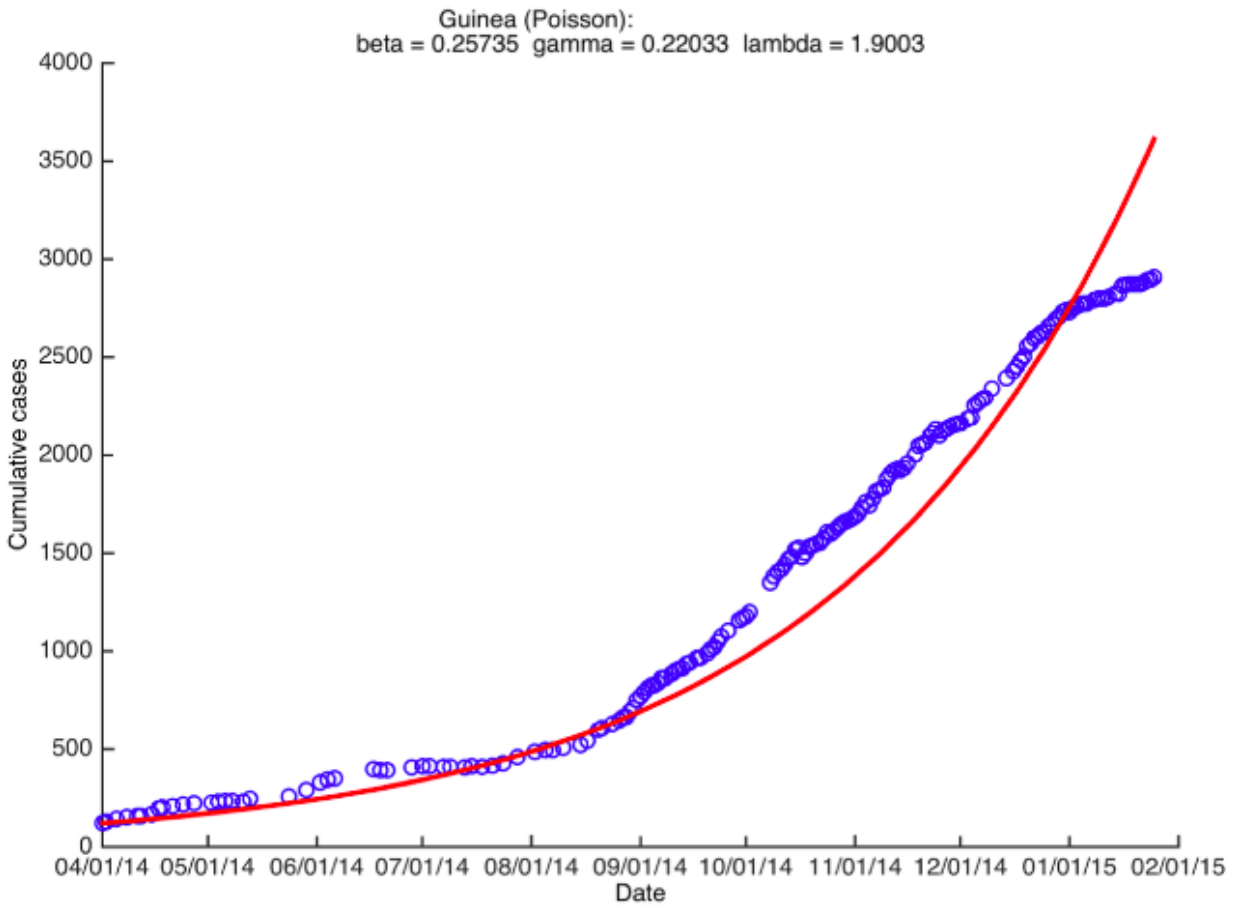


Figure : Poisson and NB Model for SL; $\mathcal{R}_0 = 1.5$

Results w/out effective size adjustment



Summary

- Contact structure is essential to capture dynamics of Ebola
- Here: contact SIDR model via a CM on a random graph for PT degree dist
- Approximated by Markovian process based on CM SLLN
- Aggregate data analysis using this model seems reasonable
- More data is coming: opportunity for analysis across scales and for incorporate E,H, and F stages.

Thanks...

- National Science Foundation (DMS-1513489)
- Mathematical Biosciences Institute at OSU
- Karly Jacobsen, Mark Burch, Joe Tien

References

Joel C Miller and Erik M Volz. (2013).
Incorporating disease and population structure into models of sir disease in contact networks.
PloS One, 8(8):e69162.

S Janson, M Luczak, P Windridge (2014).
Law of large numbers for the SIR epidemic on a random graph with given degrees.
Random Structures & Algorithms 45 (4), 724-761.

D. Linder and G. Rempala (2015).
Bootstrapping least-squares estimates in biochemical reaction networks.
Journal of Biological Dynamics 9(1): 125-46.