Likelihood-based Inference for Stochastic Epidemic Models via Data Augmentation

Jason Xu
Department of Statistical Science, Duke University

Epidemics and Information Diffusion Workshop
Simons Institute, Oct 26, 2022

Introduction: epidemic models

In many disciplines, seek to understand the spread of disease (or information, ideas, behavior) among individuals

It is desirable for models of this behavior to

- Mechanistically reflect the process of contagion
- Account for structure and rates of contacts in the population
- Allow for randomness and continuous-time dynamics

We focus on modeling and inference for stochastic epidemics

This talk: overview why inference is hard even in the simplest most prevalent models, discuss some of our contributions to overcome these challenges, and motivate new models to accommodate for modern data at higher resolutions.

Compartmental models of infection

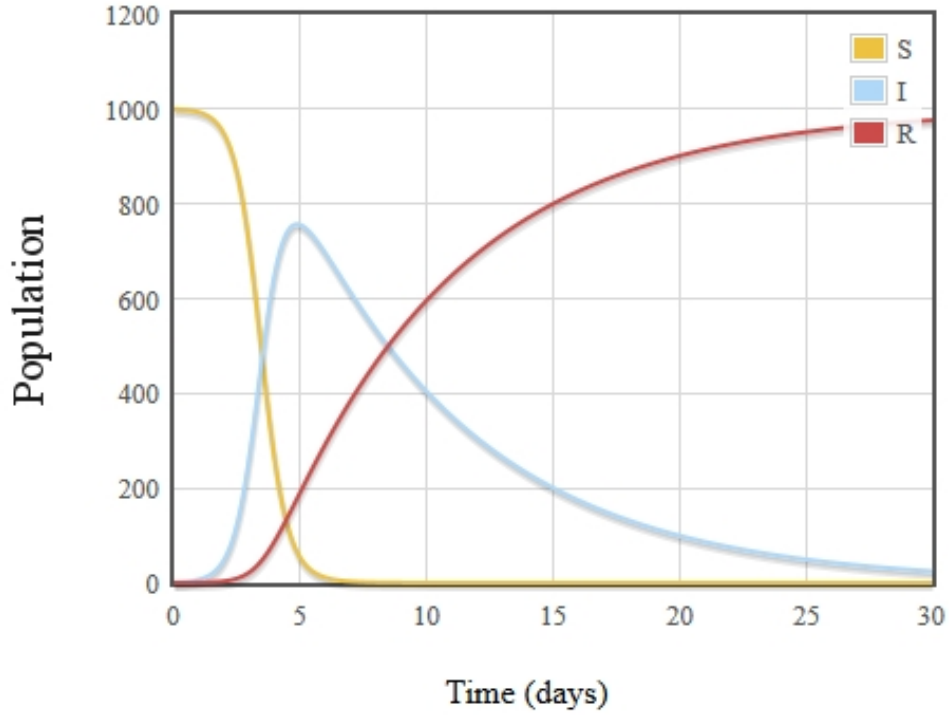
SIR model (Kermack and McKendrick 1927): describes the process of infection between groups by status within a population in terms of mechanistic but deterministic dynamics

A directed graph representation of the SIR model.

- Parameters: infection rate β, recovery rate γ
- They described a setting with quite some generality even a century ago

A simple deterministic, nonlinear system

$$\frac{dS(t)}{dt} = -\beta S(t)I(t), \quad \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \qquad \text{(SIR)}$$



$$\frac{dx}{dt} = \alpha x - \beta xy, \qquad \frac{dy}{dt} = \delta xy - \gamma y \qquad \text{(Lotka-Volterra)}$$

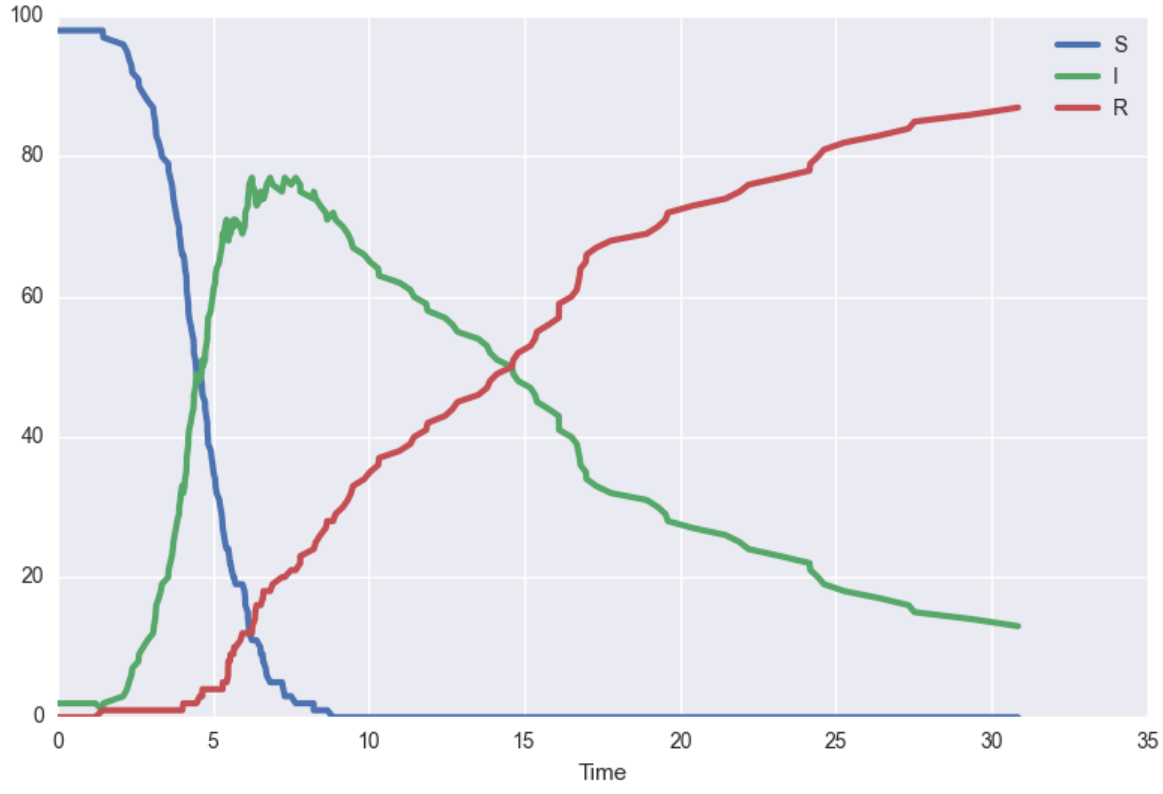
Stochastic version of the SIR model

For any small interval of time h,

$$\begin{aligned}
\Pr(\text{an infection occurs by } t+h) &= \beta S(t) I(t) h + o(h) \\
\Pr(\text{a recovery occurs by } t+h) &= \gamma I(t) h + o(h) \\
\Pr(\text{no event occurs by } t+h) &= 1 - [\beta S(t) I(t) + \gamma I(t)] h + o(h)
\end{aligned}$$

- Essentially replace rates of change in deterministic version by jump probabilities; agree for large N [Kurtz 1978]
- A continuous-time Markov chain
- Enough to track S, I if $S(t) + I(t) + R(t) = N$, i.e. closed population
- Assumes random mixing: everyone is equally likely to come in contact with one another

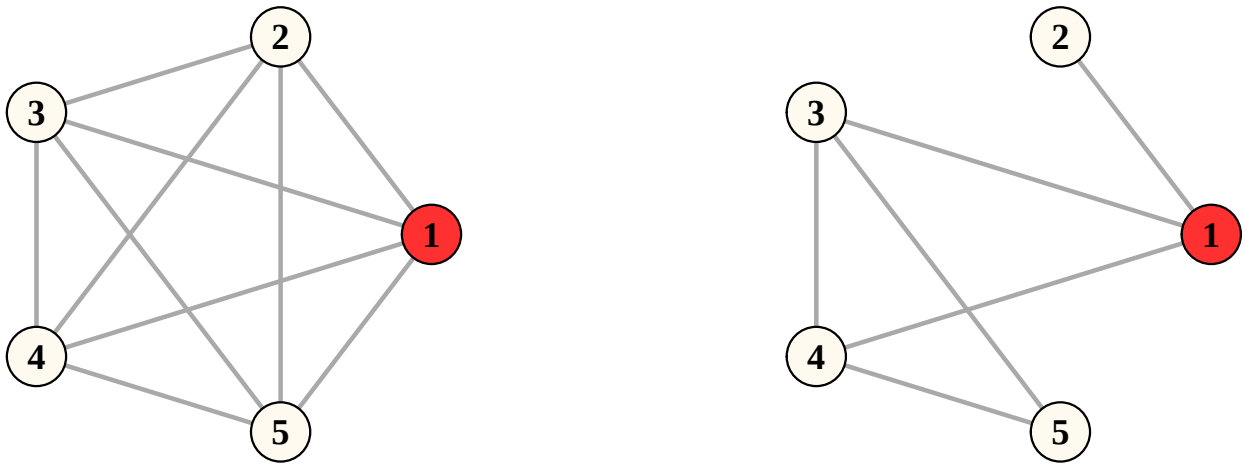
Stochastic version of the SIR model



- Basic reproductive number $R_0 = \beta N/\gamma$: "**average** number of secondary infections" caused by an infectious individual.

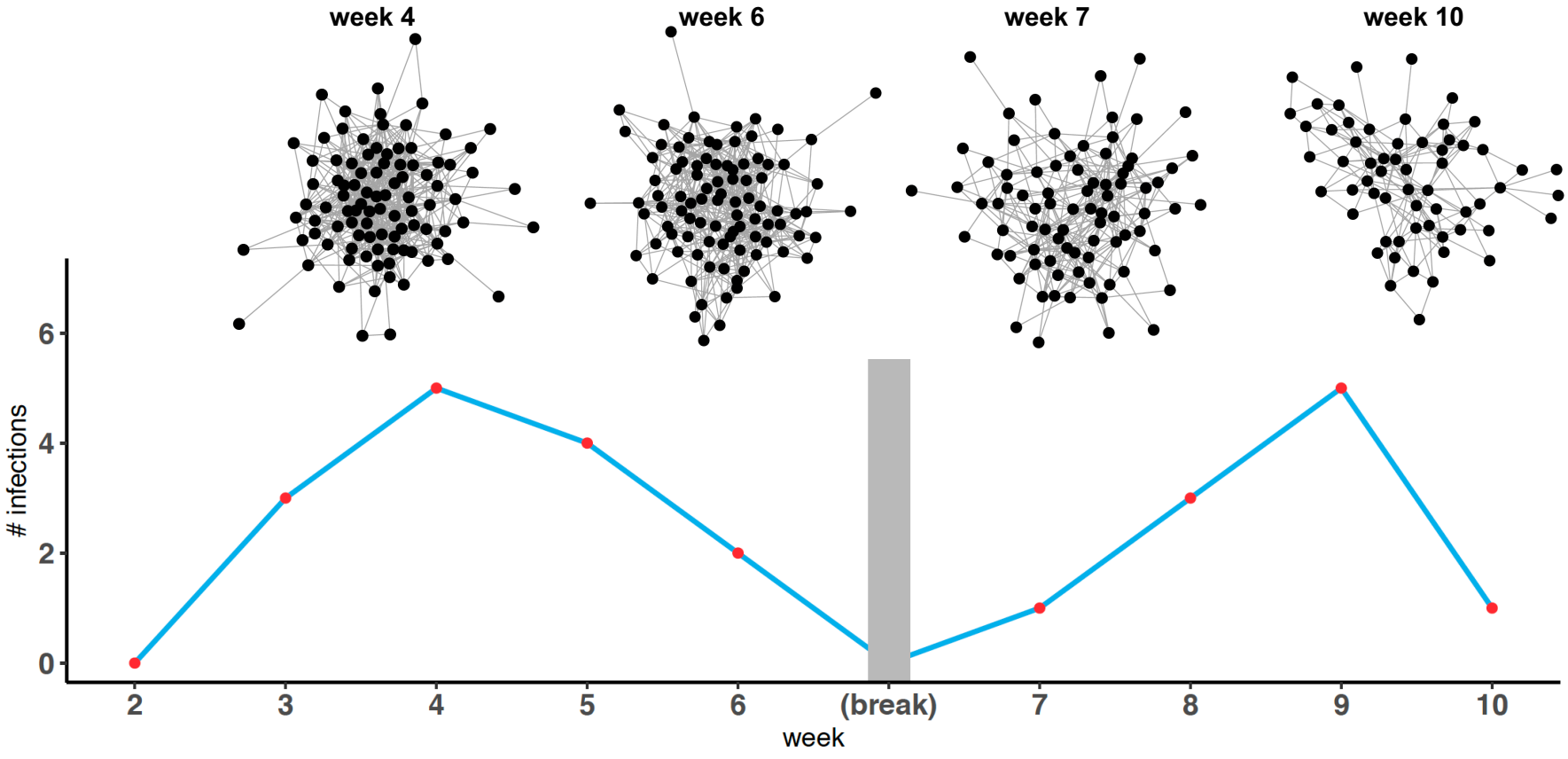
Beyond the well-mixed case

Naturally, one may be interested in transmission through **links** in a contact network (right panel).



- Recent growing interest in network setting, largely due to advances for gathering data at this resolution

SIR over a contact network: "mobile health" data



Motivating example: we will analyze eX-FLU data (Aiello et al, 2016) study; collected social contacts of 103 individuals at 5-min intervals using a Bluetooth app to study flu symptoms

Goals

We want to develop a framework that

1. models how the contact network impacts disease spread,
2. describes the influence of disease spread on network evolution,
3. accounts for randomness and uncertainty,
4. enables inference based on exact likelihood of SIR and network processes
5. accommodates partial observations (common in real data, but challenging to handle).

Even the simple stochastic SIR is difficult to fit to data

Under the baseline model, it's already difficult to infer the parameters from data without additional simplifying assumptions

- Common to approximate the model (discrete-time, branching (early), diffusion (late), locally constant, TSIR)
- Simulation methods (SMC, particle filters, ABC) are powerful and popular, often suffer poor mixing/degeneracy, nontrivial
- Ex: WHO estimates infectiousness of Ebola in recent West Africa outbreak based on *deterministic SIR* from population-level data

Instead, we pursue likelihood-based methods for fitting continuous-time, stochastic SIR models to partially observed epidemic over dynamic networks

The complete-data likelihood of any CTMC

A continuous-time Markov chain jumps between states according to exponential rates based on the generator $\mathbf{Q}$.

For instance, q_{xy} in $\mathbf{Q}$ means the time T until the process jumps between two states $x, y \in \Omega$ is distributed $T \sim \exp(q_{xy})$.

- Diagonals $-q_j := \sum_{k \in \Omega} q_{jk}$ give rate of jumping out of j.
- Likelihood is simply a product of exponentials:

$$L_c(\boldsymbol{\tau}, \mathbf{x}; \mathbf{Q}) = q_{x_0} e^{-q_{x_0}\tau_1} q_{x_1} e^{-q_{x_1}(\tau_2 - \tau_1)} \dots \cdot q_{x_{N_T-1}} e^{-q_{x_{N_T-1}}(\tau_{N_T} - \tau_{N_T-1})} e^{-q_{x_{N_T}}(T - \tau_{N_T})} \prod_{i=1}^{N_T} \frac{q_{x_{i-1}x_i}}{q_{x_{i-1}}}.$$

The complete-data likelihood

CTMC with rate matrix $\mathbf{Q}$: likelihood factorizes nicely

$$L_c(\boldsymbol{\theta}; \mathbf{X}) = \prod_{\mathbf{x} \neq \mathbf{y}} \left(q_{\mathbf{xy}}^{N(\mathbf{x},\mathbf{y})} \right) e^{\sum_{\mathbf{x}} \tau(x) q_{\mathbf{xx}}}$$

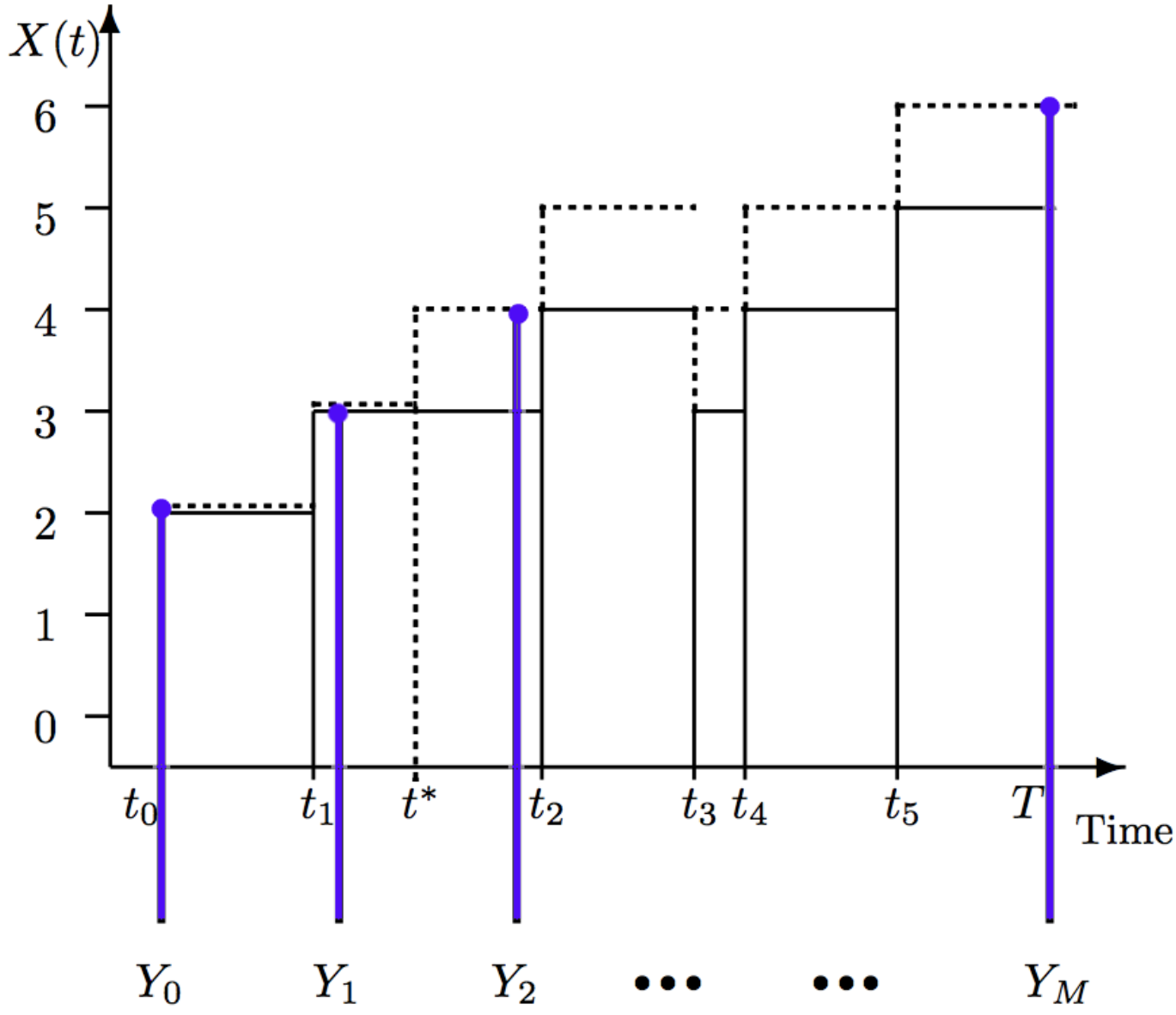
$$\ell_c(\boldsymbol{\theta}; \mathbf{X}) = \sum_{\mathbf{x}} \sum_{\mathbf{x} \neq \mathbf{y}} N(\mathbf{x}, \mathbf{y}) \ln q_{\mathbf{xy}} - \sum_{\mathbf{x}} \sum_{\mathbf{x} \neq \mathbf{y}} \tau(\mathbf{x}) q_{\mathbf{xy}}$$

$N(\mathbf{x}, \mathbf{y})$ = total $\mathbf{x} \to \mathbf{y}$ transitions, $\tau(\mathbf{x})$ = total time spent in $\mathbf{x}$

- MLEs are easy: $\hat{q}_{\mathbf{xy}} = N(\mathbf{x}, \mathbf{y}) / \tau(\mathbf{x})$
- Posteriors with conjugate Gamma(α, β) priors also easy:

$$q_{\mathbf{xy}} | \mathbf{N}, \boldsymbol{\tau} \sim \text{Gamma}(N(\mathbf{x}, \mathbf{y}) + \alpha, \tau(\mathbf{x}) + \beta)$$

Challenge: partial observations



The discretely observed data log-likelihood

$$\ell_o(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{p=1}^{m} \sum_{i=0}^{n(p)-1} \log p_{\mathbf{X}^p(t_{p,i}),\mathbf{X}^p(t_{p,i+1})}(t_{p,i+1} - t_{p,i}|\boldsymbol{\theta})$$

In particular, composed of finite-time **transition probabilities**:

$$p_{\mathbf{x},\mathbf{y}}(s) = \Pr\left(\mathbf{X}(t+s) = \mathbf{y} | \mathbf{X}(t) = \mathbf{x}\right)$$

- Marginalized over infinitely many endpoint-conditioned paths (i.e. evaluating a hard integral over all possible paths between observations)
- Classical matrix exponentiation for CTMCs is $\mathcal{O}(|\Omega|^3)$

$$\mathbf{P}(t) := \left\{p_{\mathbf{x},\mathbf{y}}(t)\right\}_{\mathbf{x},\mathbf{y}\in\Omega} = e^{\mathbf{Q}t} = \sum_{k=0}^{\infty} \frac{(\mathbf{Q}t)^k}{k!}.$$

We've taken approaches via generating functions ϕ

$$\phi_{jk}(t, s_1, s_2; \boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}\left(s_1^{X_1(t)} s_2^{X_2(t)} | X_1(0) = j, X_2(0) = k\right)$$
$$= \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} p_{(jk),(lm)}(t; \boldsymbol{\theta}) s_1^l s_2^m; \qquad |s_i| \leq 1$$

- We derive differential equations governing ϕ_{jk} using the Kolmogorov forward/backward equations
- ϕ_{jk} may have closed solutions or cheap numerical solutions

Transition probabilities are related to the PGF via differentiation:

$$p_{(jk),(lm)}(t) = \frac{1}{l!m!}\frac{\partial^l}{\partial s_1}\frac{\partial^m}{\partial s_2}\phi_{jk}(t)\Big|_{s_1=s_2=0}.$$

Repeated differentiation impractical, numerically unstable

From differentiation to integration: series inversion

- Let $s_1 = e^{2\pi i w_1}, s_2 = e^{2\pi i w_2} \Rightarrow \phi$ becomes a Fourier series:

$$\phi_{jk}(t, e^{2\pi i w_1}, e^{2\pi i w_2}) = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} p_{(jk),(lm)}(t) e^{2\pi i l w_1} e^{2\pi i m w_2}$$

$$p_{(jk),(lm)}(t) = \int_0^1 \int_0^1 \phi_{jk}(t, e^{2\pi i w_1}, e^{2\pi i w_2}) e^{-2\pi i l w_1} e^{-2\pi i m w_2} dw_1 dw_2$$

(Fourier inversion + Riemann approximation)

$$\approx \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \phi_{jk}(t, e^{2\pi i u/N}, e^{2\pi i v/N}) e^{-2\pi i l u/N} e^{-2\pi i m v/N}.$$

- *simultaneously* compute set of probabilities $\{p_{(jk),(lm)}(t)\}$ for all $l, m = 0, \ldots, N$ using Fast Fourier Transform (FFT) [Xu, Guttorp, Kato-Maeda, Minin 2015]

Nonlinear transition probabilities for marginal likelihood

Much harder to apply these ideas to non-linear processes—recall in SIR, interaction term $\beta S(t)I(t)$

Very briefly, working in the Laplace domain,

$$\phi_{ab}(s) := \mathcal{L}[p_{(jk),(ab)}(t)](s) = \int_0^\infty e^{-st} p_{(jk),(ab)}(t)dt$$

satisfies a recursion with continued fraction representation

$$\phi_{ab}^{(0)}(s) = \prod_{i=1}^{b} x_{ai} \cfrac{x_{a,b+1}}{Y_{a,b+1} + \cfrac{x_{a,b+2} Y_{ab}}{y_{a,b+2} + \cfrac{x_{a,b+3}}{y_{a,b+3} + \cfrac{x_{a,b+4}}{y_{a,b+4} + \cdots}}}}$$

- Evaluate to finite depth with bound on error

A brief outline: direct computation of SIR likelihood

"*The associated mathematical manipulations required to generate solutions can only be described as heroic.*"

— Eric Renshaw on computing the SIR marginal likelihood, Stochastic Population Processes.

- Compute continued fraction by recursion in Laplace domain
- Numerically invert Laplace transform to recover transition probabilities [Abate and Whitt]
- The marginal likelihood is simply a product of these quantities [Ho et al 2018, J. Math. Bio.]
- We've also used bivariate branching process approximations that allow for more efficient series inversion approaches via their probability generating functions

Beyond the simple SIR

These techniques to integrate over missing data directly are delicate, and do not extend easily to more complex models, i.e. those involving contact networks, non-Markovian dynamics, rates varying with covariates, etc

Nonetheless, let's define a suitable generative model and derive its corresponding complete data likelihood.

To account for missing data, we will explore the space of possible missing values via a Markov Chain, using sampling as an alternative to direct integration.

Data augmentation: integration via latent sampling

The partial data likelihood is obtained by marginalizing over missing data denoted $\mathbf{Z}$:

$$\mathcal{L}_o(\mathbf{X}|\theta) = \int \mathcal{L}_c(\mathbf{X}, \mathbf{Z}|\theta) d\mathbf{Z}$$

When difficult to compute this integral directly, instead propose possible values of $\mathbf{Z}$, and construct a Markov chain to explore the *joint posterior* distribution $\pi(\mathbf{Z}, \theta|\mathbf{X})$ [Tanner and Wong 1987].

By targeting the joint posterior, computations (i.e. acceptance ratios) require the tractable joint likelihood $\mathcal{L}_c$ above.

If we use MCMC to draw $\{(\mathbf{Z}_1, \theta_1), \ldots, (\mathbf{Z}_m, \theta_m)\} \sim \pi(\mathbf{Z}, \theta|\mathbf{X})$, marginalizing out $\mathbf{Z}$ in the posterior is trivial.

Data augmentation: integration via latent sampling

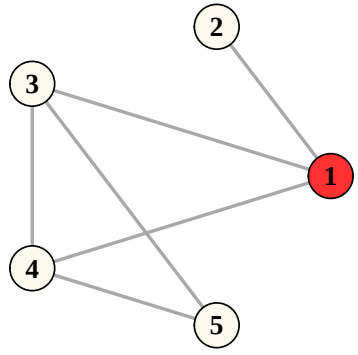
To answer original questions regarding model parameters θ given observations $\mathbf{X}$, we simply *ignore the samples of* $\mathbf{Z}_i$. That is, the remaining $\{\theta_1, \ldots, \theta_m\}$ form a sample from the posterior $\pi(\theta|\mathbf{X})$

Our strategy [Bu et al 2022, JASA]:

1. Formulate a model for co-evolution of SIR and contact network
2. Derive its complete data likelihood $\mathcal{L}_c$
3. Construct a sampler to augment the data to make use of $\mathcal{L}_c$

Notation and a very simple network process

Each individual's contacts can be represented via an adjacency matrix $\mathbf{A}$.

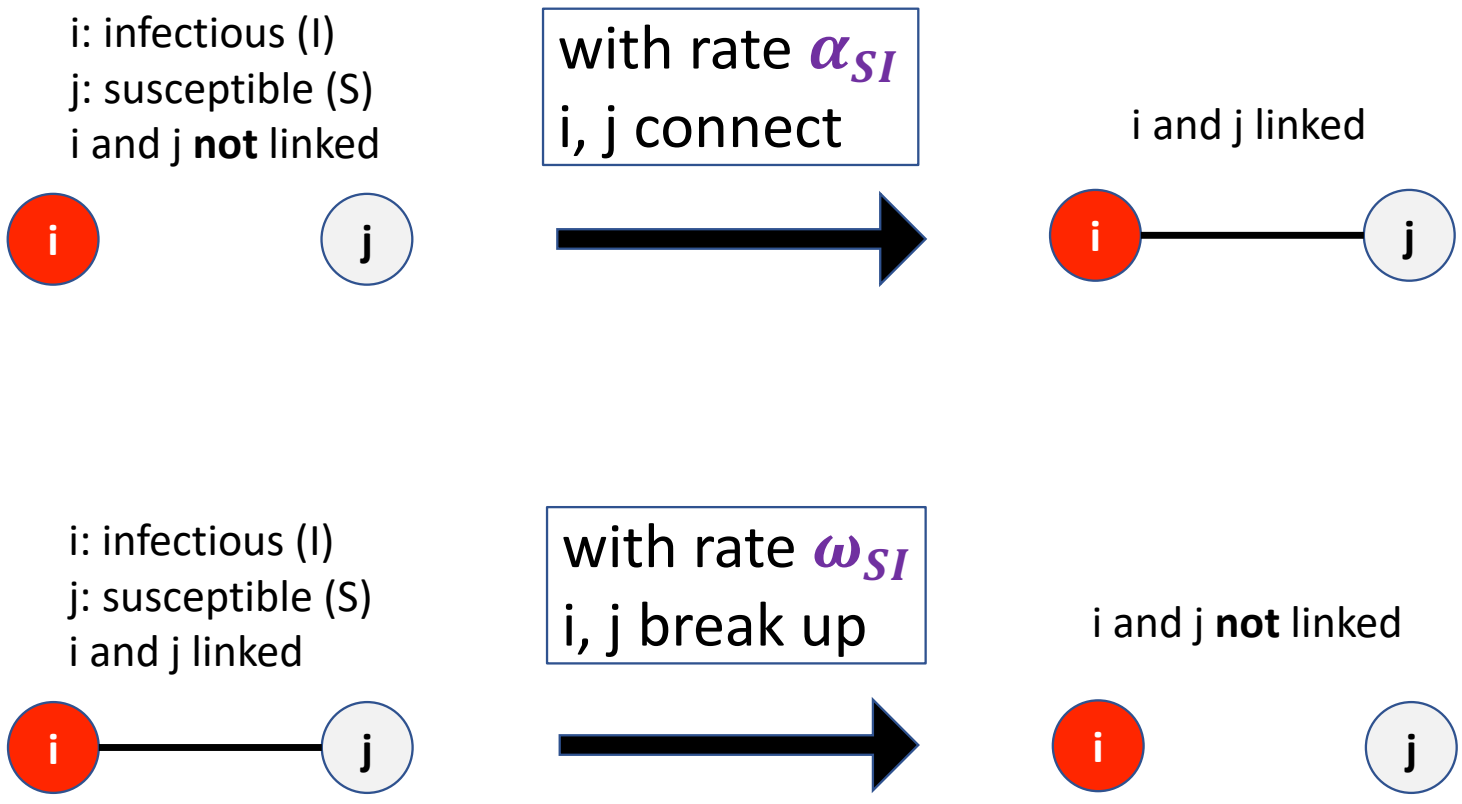


$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

To allow network to adapt, we will have each $\mathbf{A}_{ij}(t)$ evolve according to a Markov process taking values of 0 or 1.

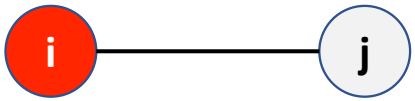
In particular, the rates of edge creation α_{AB} and deletion ω_{AB} should depend on disease status A, B of the two individuals.

Individual-level network events



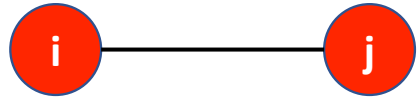
Individual-level SIR dynamics

i: infectious
j: susceptible
i and j in contact



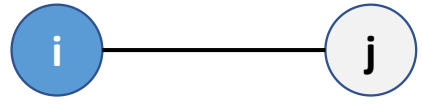
with rate β
i infects j

i and j both infected



with rate γ
i recovers

i recovered, j susceptible



Joint dynamics of the model

At the individual level, four competing Poisson processes:

1. **Infection**: an I individual infects an S neighbor with rate $\boxed{\beta}$;
2. **Recovery**: an I individual recovers with rate $\boxed{\gamma}$ independently;
3. **Link activation**: a link is formed at rate $\boxed{\alpha_{AB}}$ between a status A individual and a status B individual;
4. **Link termination**: a link is removed at rate $\boxed{\omega_{AB}}$ analogously
5. Assume recovered and susceptible individuals behave identically in terms of the network

Dynamics of the overall generative model

At the population level, the entire process sums over all individual-level rates (superposition property):

1. **Infection** occurs with rate $\boxed{\beta SI(t)}$, where $SI(t) =$ number of S-I links at time t;
2. **Recovery** occurs with rate $\boxed{\gamma I(t)}$, $I(t) =$ number of I individuals at time t;
3. **Link activation** for A-B pairs occurs with rate $\boxed{\alpha_{AB} M^d_{AB}(t)}$, $M^d_{AB}(t) =$ number of disconnected A-B pairs at time t;
4. **Link termination** for A-B pairs is dissolved with rate $\boxed{\omega_{AB} M_{AB}(t)}$, $M_{AB}(t) =$ number of connected A-B pairs at time t.

Deriving the complete-data likelihood

Given the initial network structure $\mathcal{G}_0$ initial infective at time 0,

$$\mathcal{L}(\beta, \gamma, \tilde{\alpha}, \tilde{\omega} | \mathcal{G}_0)$$

$$= \gamma^{n_R} \beta^{n_E - 1} \alpha_{SS}^{C_{HH}} \alpha_{SI}^{C_{HI}} \alpha_{II}^{C_{II}} \omega_{SS}^{D_{HH}} \omega_{SI}^{D_{HI}} \omega_{II}^{D_{II}} \prod_{j=2}^{n} \left[\tilde{M}(t_j) \left(I_{p_{j1}}(t_j) \right)^{F_j} \right]$$

$$\times \exp\left(- \int_0^{T_{\max}} \left[\beta SI(t) + \gamma I(t) + \tilde{\alpha}^T \mathbf{M}_{\max}(t) + (\tilde{\omega} - \tilde{\alpha})^T \mathbf{M}(t) \right] dt \right).$$

with parameters $\Theta = \{\beta, \gamma, \alpha_{SS}, \alpha_{SI}, \alpha_{II}, \omega_{SS}, \omega_{SI}, \omega_{II}\}$

- This takes same simple form presented in beginning of talk, just with more notation/bookkeeping (see paper for full notation!)
- Difficult to write the likelihood for very similar models, i.e. preventative rewiring [Ball & Britton 2022]

Closed form inference

Analogously to the introduction slides: when completely observed,

$$\hat{\beta} = \frac{n_E - 1}{\sum_{j=1}^{n} SI(t_j)(t_j - t_{j-1})}, \quad \hat{\gamma} = \frac{n_R}{\sum_{j=1}^{n} I(t_j)(t_j - t_{j-1})},$$

$$\hat{\alpha}_{SS} = \frac{C_{HH}}{\sum_{j=1}^{n} \left[\frac{H(t_j)(H(t_j)-1)}{2} - M_{HH}(t_j)\right](t_j - t_{j-1})}, \quad \hat{\omega}_{SS} = \frac{D_{HH}}{\sum_{j=1}^{n} M_{HH}(t_j)(t_j - t_{j-1})},$$

$$\hat{\alpha}_{SI} = \frac{C_{HI}}{\sum_{j=1}^{n} [H(t_j)I(t_j) - M_{HI}(t_j)](t_j - t_{j-1})}, \quad \hat{\omega}_{SI} = \frac{D_{HI}}{\sum_{j=1}^{n} M_{HI}(t_j)(t_j - t_{j-1})},$$

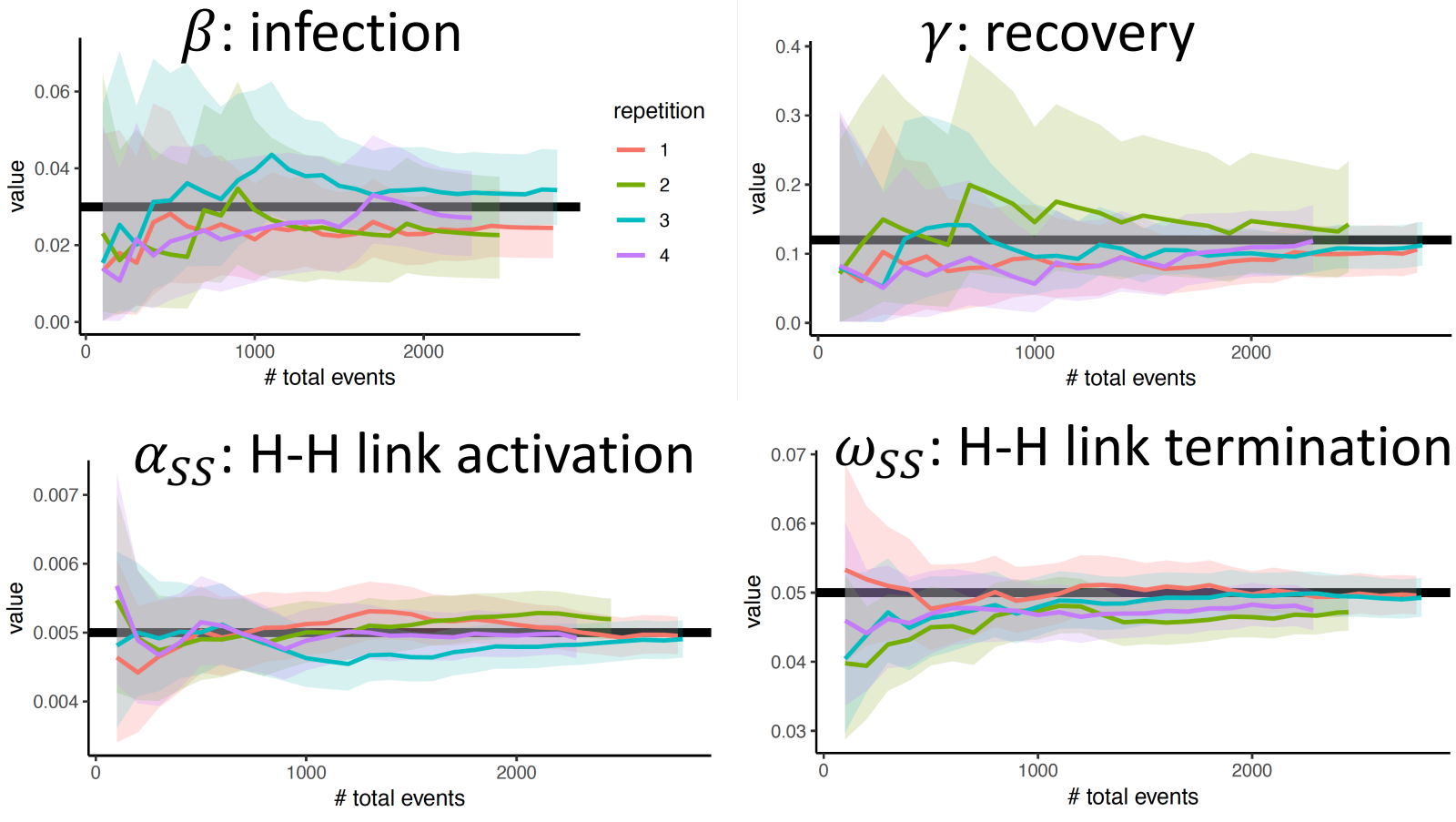
$$\hat{\alpha}_{II} = \frac{C_{II}}{\sum_{j=1}^{n} \left[\frac{I(t_j)(I(t_j)-1)}{2} - M_{II}(t_j)\right](t_j - t_{j-1})}, \quad \hat{\omega}_{II} = \frac{D_{II}}{\sum_{j=1}^{n} M_{II}(t_j)(t_j - t_{j-1})}.$$

For Bayesian approach, Gamma priors again offer conjugacy:

$$\beta \sim Ga(a_\beta, b_\beta), \quad \gamma \sim Ga(a_\gamma, b_\gamma),$$
$$\alpha_{..} \sim Ga(a_\alpha, b_\alpha), \quad \omega_{..} \sim Ga(a_\omega, b_\omega).$$

Simulation study: inference with complete data

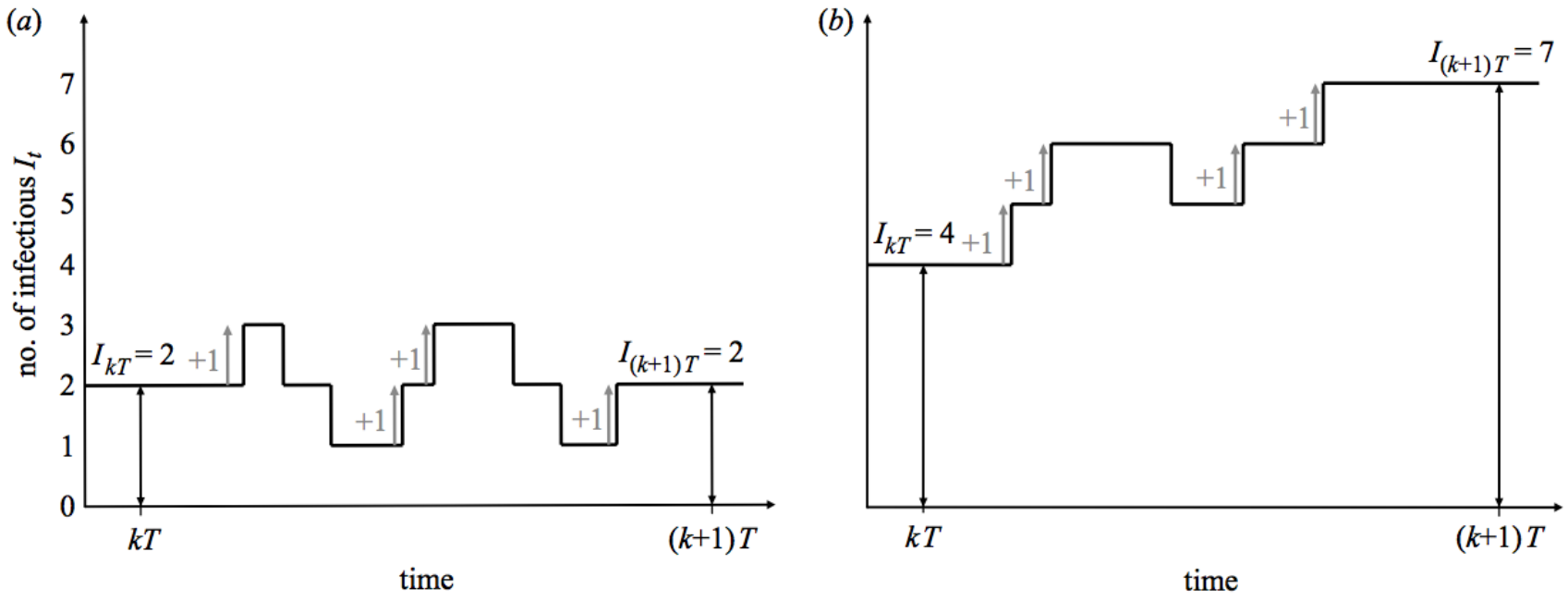
- Population size $N = 100$, 4 repetitions;
- Posterior samples concentrate around the truth (**bold line**)



Inference for partially observed case

Having validated the likelihood, how do we use it toward inference in the partially observed case?

Recall we need a way to draw valid samples of the missing events $\mathbf{Z}$, conditional on the model and observed data. In the case of the real data, $\mathbf{Z}$ is the set of unknown recovery times.



Data setting: eX-FLU

Spread of influenza on campus (Aiello et al. 2016) with weekly health status surveys and high-resolution contact network info via Bluetooth app

- $\{(u_\ell, u_{\ell+1}]\}_{\ell=1}^{L}$: intervals (weeks) during which "health status reports" are collected
- Times of social link creation/deletion are known continuously
- In the ℓth interval, there are R_ℓ new recoveries (summary data), but their event times are unknown
- Latent variables: $\mathbf{Z} = \{r_{\ell,1}, \ldots, r_{\ell,R_\ell}\}$.
- Observed data: $\mathbf{x}$ = events + endpoint statuses.

A simple algorithmic primitive

Proposed inference method: data-augmented MCMC.

for $s = 1 : \text{maxIters}$:

1. **Data augmentation**. For observation times $\ell = 1 : L$, draw recovery times $\{r_{\ell,i}^{(s)}\}_{i=1:R_\ell} = \mathbf{Z}$ from their joint conditional

$$p\left(\{r_{\ell,i}\}_{i=1:R_\ell} | \Theta^{(s-1)}, \mathbf{x}, \{r_{\ell',i}\}_{i=1:R_{\ell'}, \ell' \neq \ell}\right). \qquad (1)$$

2. **Update parameter values**. Now that $\mathbf{x}$ augmented by $\mathbf{Z}$ comprise the complete data, Gibbs sample parameters $\Theta^{(s)}$ from their (conditional) posteriors.

Conditional sampling/simulation is nontrivial

1. Simulating CTMC with fixed endpoints notoriously difficult
2. Must respect constraints imposed by the contact network structure.
3. Previous work in well-mixed case proposes individual disease histories one by one, and require intensive Metropolis-Hastings steps based on marginal likelihood

Fortunately,

- Network structure "reduces dimension" of the latent space to be explored (mechanistic information)

Data Augmentation Regulated by Contact Info (DARCI)

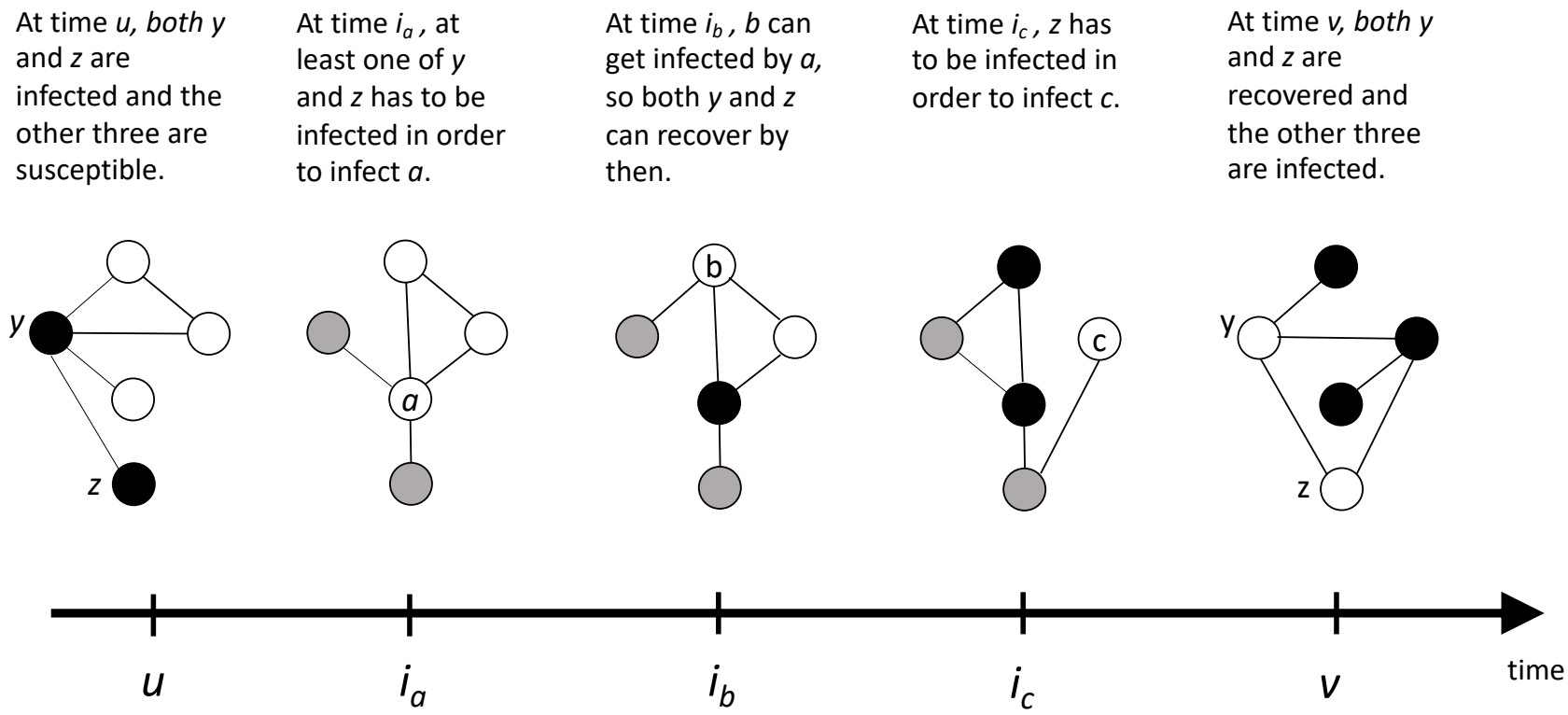


Figure: How DARCI "imputes" y and z's unknown recovery times in $(u, v]$. ● = known infected, ◉ = possibly infected, ○ = healthy.

Gibbs sampling from closed form conditionals

$$\beta|\text{Data} \sim Ga\left(a_\beta + n_E - 1, b_\beta + \sum_{j=1}^{n} SI(t_j)(t_j - t_{j-1})\right),$$

$$\gamma|\text{Data} \sim Ga\left(a_\gamma + n_R, b_\gamma + \sum_{j=1}^{n} I(t_j)(t_j - t_{j-1})\right),$$

$$\alpha_{SS}|\text{Data} \sim Ga\left(a_\alpha + C_{HH}, b_\alpha + \sum_{j=1}^{n} \left[H(t_j)(H(t_j) - 1)/2 - M_{HH}(t_j)\right](t_j - t_{j-1})\right),$$

$$\alpha_{SI}|\text{Data} \sim Ga\left(a_\alpha + C_{HI}, b_\alpha + \sum_{j=1}^{n} \left[H(t_j)I(t_j) - M_{HI}(t_j)\right](t_j - t_{j-1})\right),$$

$$\alpha_{II}|\text{Data} \sim Ga\left(a_\alpha + C_{II}, b_\alpha + \sum_{j=1}^{n} \left(I(t_j)(I(t_j) - 1)/2 - M_{II}(t_j)\right)(t_j - t_{j-1})\right),$$

$$\omega_{SS}|\text{Data} \sim Ga\left(a_\omega + D_{HH}, b_\omega + \sum_{j=1}^{n} M_{HH}(t_j)(t_j - t_{j-1})\right),$$

$$\omega_{SI}|\text{Data} \sim Ga\left(a_\omega + D_{HI}, b_\omega + \sum_{j=1}^{n} M_{HI}(t_j)(t_j - t_{j-1})\right),$$

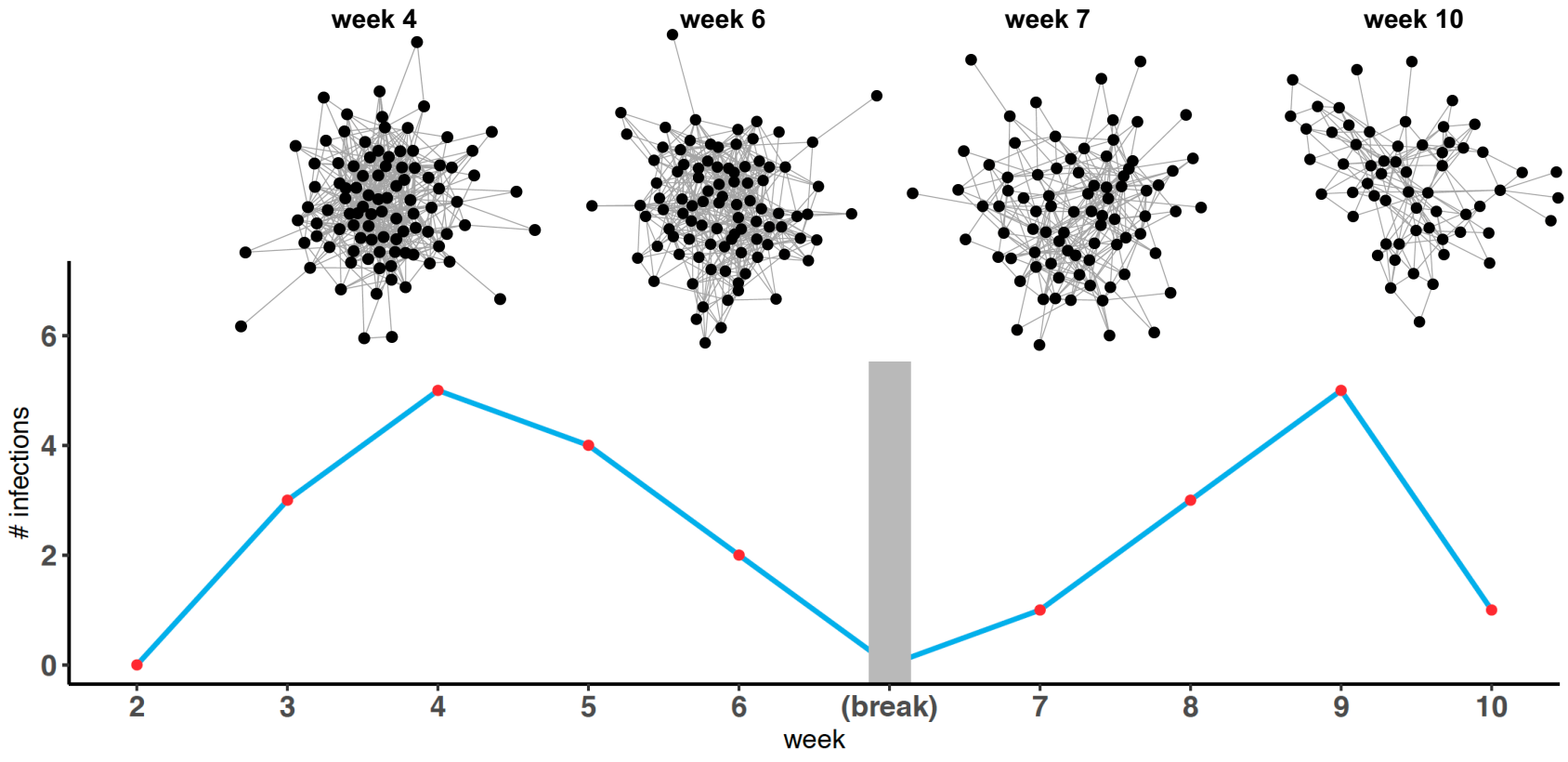
$$\omega_{II}|\text{Data} \sim Ga\left(a_\omega + D_{II}, b_\omega + \sum_{j=1}^{n} M_{II}(t_j)(t_j - t_{j-1})\right).$$

Details of case study: eX-FLU data

- 2013 study on influenza transmission among college students
- 590 students observed over 10 weeks
- Weekly surveys on influenza-like-illnesses (ILI)
- $N = 103$ participants in the sub-study, using iEpi to track social interactions via Bluetooth (5-min resolution)
- Infection: positive ILI case
- Link activation: beginning of a contact registered by iEpi
- Link termination: end of a contact registered by iEpi
- Recovery: missing

Contact information via bluetooth data

- 45,760 network events
- low network density



Results and analysis

- The data-augmented inference scheme is employed based on the exact model likelihood (posterior)
- Modification allows for external infection rate ξ
- Posterior inference gives us summaries of parameters as well as uncertainty: strong internal force of infection with slow transmission, quick recovery, short contacts.

Parameter	Mean	2.5%	97.5%
β (internal infection)	0.0695	0.0247	0.1500
ξ (external infection)	0.00331	0.00208	0.00494
γ (recovery)	0.294	0.186	0.428
α_{SS} (H-H link activation)	0.0514	0.0499	0.0529
ω_{SS} (H-H link termination)	38.26	33.55	40.62
α_{SI} (H-I link activation)	0.130	0.0785	0.194
ω_{SI} (H-I link termination)	53.5	22.5	231.7

Results and discussion

Our framework employs a generative model that

- describes the **interplay** between epidemic and network processes through time,
- accounts for the randomness of the process, and
- performs inference with partial epidemic observations based on the exact model posterior

The network model is naïve, with plenty of room for extension!

- Daunting to impute network— deterministic, behaviorally-inspired dynamics instead?
- Combine with rational economic frameworks, myopic decisions, cost functions
- Population structure, weaker "exchangeability" rather than independent edge behavior within each class?

Building upon the foundation

- Use an inhomogeneous bivariate branching process as faithful proposal density when *no event times are known*
- Accommodates general non-exponential time until recovery
- Initialized at low density regions, appears to reach stationarity after 100 iterations and yields $100,000$ posterior samples in a couple of minutes on a simulation with $S_0 = 1,000$

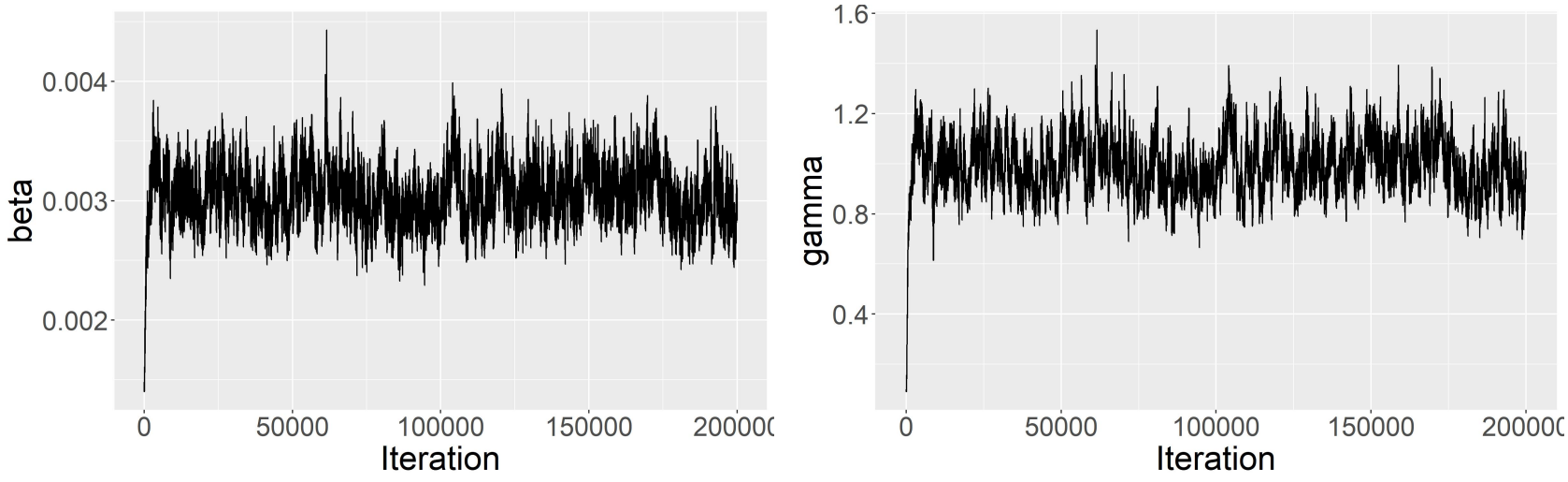


Figure: Traceplots of data-augmented MCMC with $\beta_0 = 0.003, \gamma_0 = 1$.

Piecewise-decoupled SIR

Under the SIR, the *population* infection rate $\mu_{pop}(t) = \beta I(t)S(t)$ changes after every event. Consider the approximation:

Let $\tilde{\mu}_{pop}$ be decoupled of $I(t)$ only (PD-SIR).

$$\tilde{\mu}_{pop}(t) := \beta S(t) I(t_{k-1}), \quad t \in [t_{k-1}, t_k).$$

- The corresponding *individual* infection rate is $\tilde{\mu}(t) := \dfrac{\tilde{\mu}_{pop}(t)}{S(t)} = \beta I(t_{k-1}) = \mu_k$ is piece-wise constant.
- The compartment $S(t)$ follows a linear death process (LDP) where infections correspond to deaths.
- The bivariate process is equivalent to a branching process on each interval

Simulating the PD-SIR – pseudo-algorithm

Algorithm 1: Simulating a PD-SIR process conditionally on $\mathbf{Y}$

Output: $\mathbf{Z}^\star = \{(z_i^I, z_i^R)\}_i$ compatible with $\mathbf{Y} = I_{1:K}$

for *interval* $k = 1, \ldots, K$ **do**

Compute the infection rate:

$$\mu_k \leftarrow \beta I(t_{k-1})$$

Jointly generate the infection times:

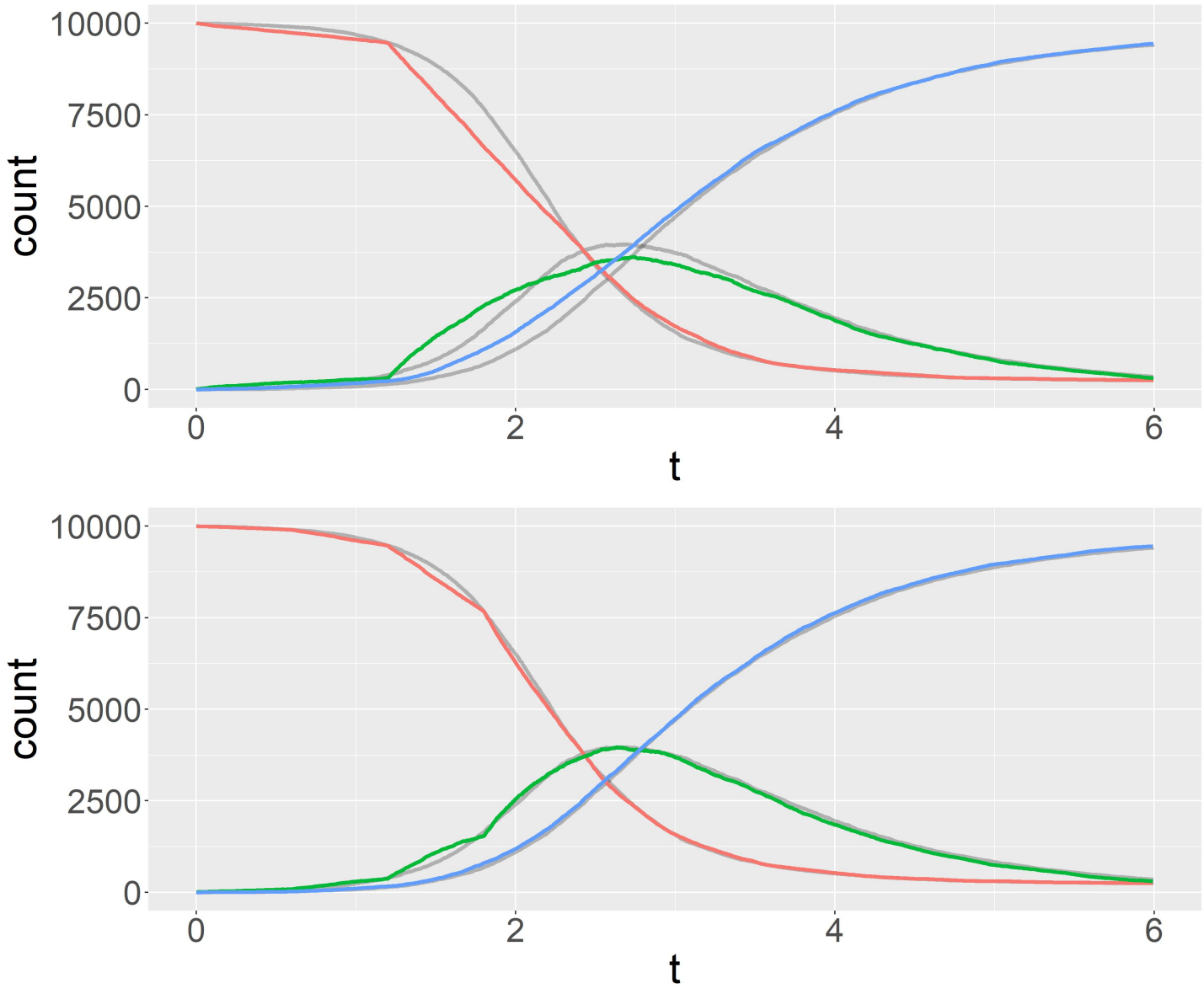
$$z_j^I \overset{iid}{\sim} \text{TruncExp}(\mu_k; t_{k-1}, t_k), \text{ for } j \in \mathcal{I}_k$$

Generate the removal times:

$$z_j^R | z_j^I \overset{indep.}{\sim} z_j^I + \textit{Exponential}(\gamma), \text{ for } j \in \mathcal{I}_k$$

end

A faithful approximation



A faithful approximation

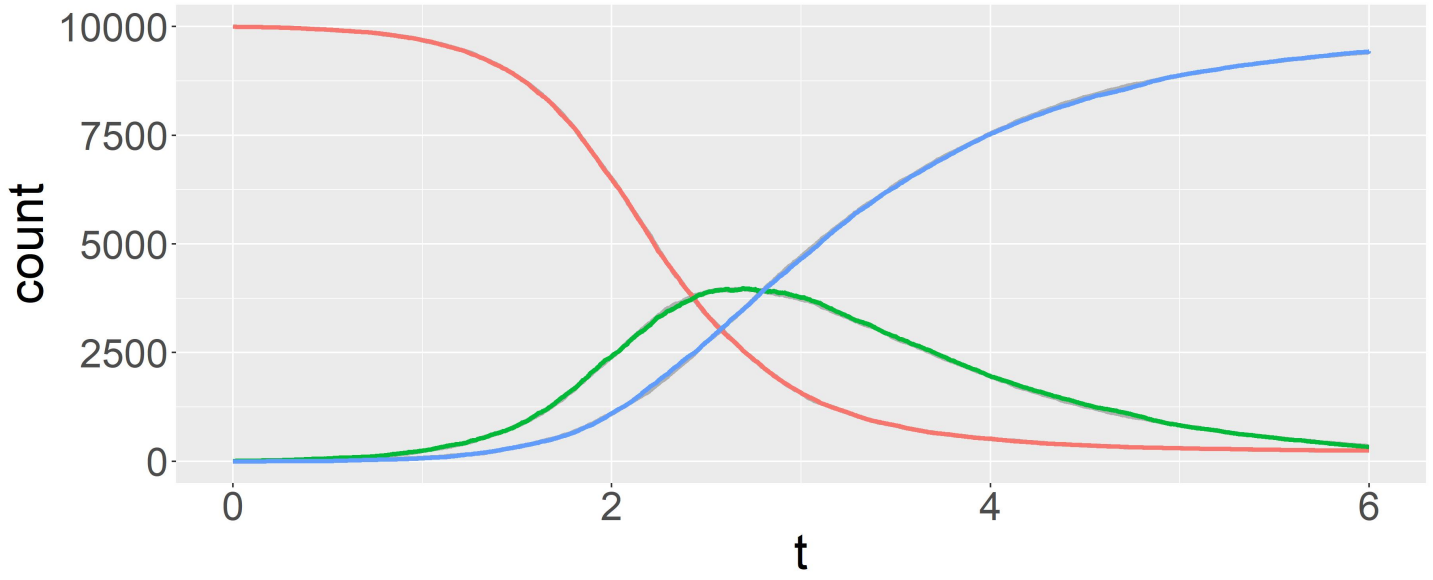


Figure: SIR (with S in red, I in green and R in blue) and a PD-SIR (in grey) trajectories under the same infection incidence data $I_{1:K}$ with $K = 5, 10, 50$ total observations.

2014-2016 Ebola Pandemic in Western Africa

- Weekly incidence data in Gueckedou, $n = 300{,}000$; $\rho = 0.1$ (acceptance rate: 0.2)
- 100,000 iterations $\approx$ 5 minutes
- Hybrid data augmentation/marginal likelihood approach for COVID resurgence data in Zhijiazhuang, China of similar scale
- Overcomes computational limitations while enabling classical MCMC ideas [Gibson & Renshaw 2001, Neal & Roberts 2005]

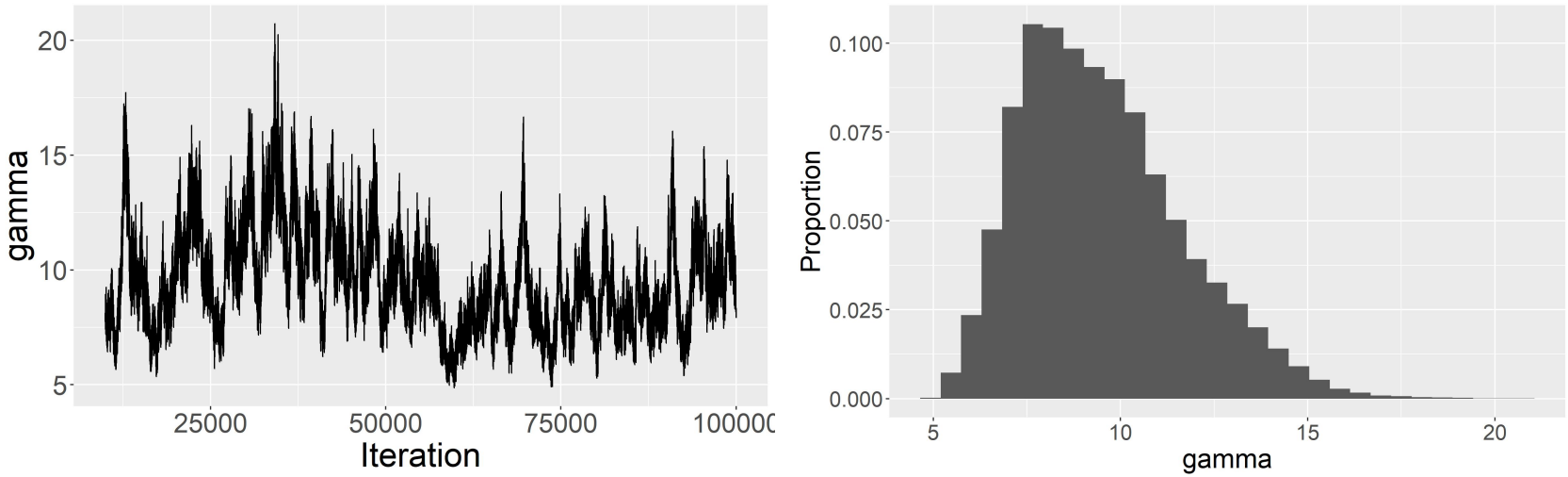


Figure: Posterior draws of $1/\gamma$, expected infection period, Ebola data

Closing thoughts

Fully stochastic modeling: interpretability, uncertainty quantification, and [eventually] reliable forecasting/decision-making

CTMC framework is extensible in many open directions and yields tractable computation, workhorse inferential procedures, and statistical guarantees

Many open directions and extensions toward model realism

- Multiple strains, compartments, covariates, reporting rates
- Network missingness/random graph models
- Dynamic force of infection $\beta(t)$ and change points
- Multi-scale hierarchies and model selection
- Hybrid differential equation models, sequential MCMC/ABC

Thank you!

Figure: Fan Bu, Raphael Morsomme, Jenny Huang

Thank you!

Likelihood-based Inference for Partially Observed Epidemics on Dynamic Networks. Bu, Aiello, Xu*, Volfovsky* 2022, Journal of the American Statistical Association.

Uniformly Ergodic Data-Augmented MCMC for Fitting the General Stochastic Epidemic Model to Incidence Data. Morsomme and Xu, 2022+, arXiv preprint.

Likelihood-based inference for partially observed stochastic epidemics with individual heterogeneity. Bu, Aiello, Volfovsky*, Xu*, 2021+, arXiv preprint.

Detecting Changes in the Transmission Rate of a Stochastic Epidemic Model. Huang, Morsomme, Dunson, Xu, 2022+, in preparation.

Birth/birth-death processes and their computable transition probabilities with biological applications. Ho, Xu, Crawford, Minin, Suchard 2018, Journal of Mathematical Biology.

Slide appendix

Simulation study: partially observed study

Forward simulate from SIR model and record the complete data

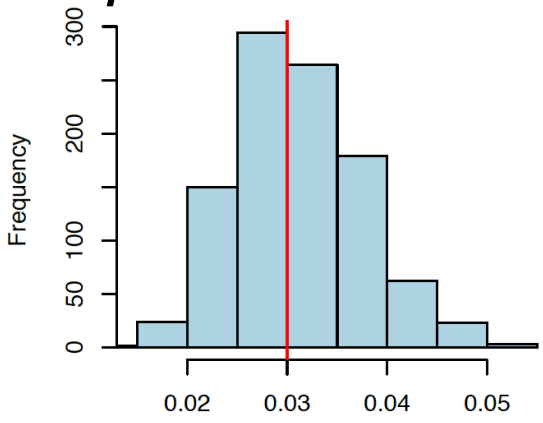
Time	**Event Type**	**Individual(s) Involved**
.....	.....	.....
95.8352	Link activation	100 & 5
95.8357	Link termination	148 & 157
95.8361	Recovery	125
95.8432	Link activation	62 & 147
95.8473	Recovery	16
95.8509	Infection	124
.....	.....	.....

Table: A sample of simulated complete data.

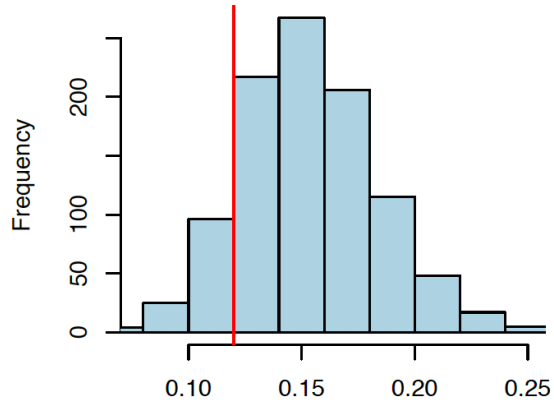
Hide all recovery times to compile "weekly health reports" to mimic discretely observed nature in real data

Accurate inference via data augmented MCMC

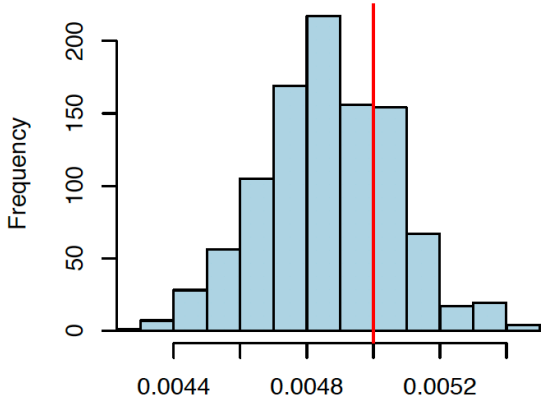
β: infection



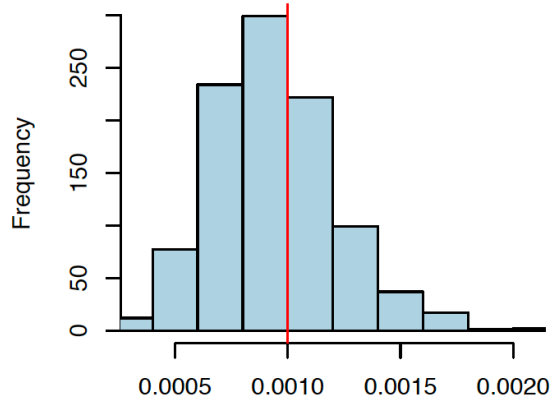
γ: recovery



α_{SS}: H-H link act.



α_{SI}: H-I link act.



Posterior for infection and recovery rates

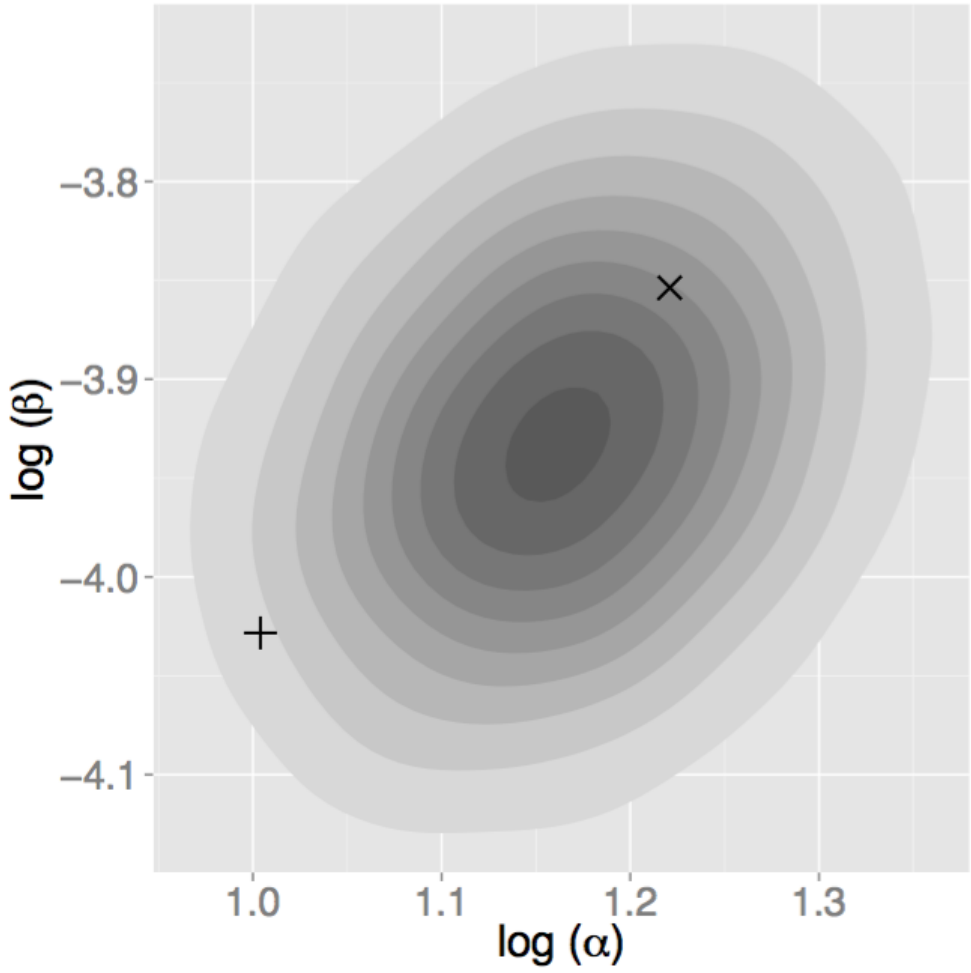


Figure: Posterior (log scale) of recovery rate α and infection rate β. The "$+$" and "$\times$" symbol are previous point estimates from deterministic and approximate models [Brauer 2008; Raggett 1982].

Relaxing the closed population assumption

- So far, closed population of size N: one can only get infected from someone **inside** the community;
- What if the population under study is a sub-population?
- ξ ("external infection rate"): constant force of infection on every susceptible.
- Likelihood:

$$\mathcal{L}(\beta, \xi, \gamma, \tilde{\alpha}, \tilde{\omega} | \mathcal{G}_0) \tag{2}$$

$$= \gamma^{n_R} \alpha_{SS}^{C_{HH}} \alpha_{SI}^{C_{HI}} \alpha_{II}^{C_{II}} \omega_{SS}^{D_{HH}} \omega_{SI}^{D_{HI}} \omega_{II}^{D_{II}} \prod_{j=2}^{n} \left[\tilde{M}(t_j) \left(\beta I_{p_{j1}}(t_j) + \xi \right)^{F_j} \right]$$

$$\times \exp\left(- \int_0^{T_{\max}} \left[\beta SI(t) + \xi S(t) + \gamma I(t) + \tilde{\alpha}^T \mathbf{M}_{\max}(t) + (\tilde{\omega} - \tilde{\alpha})^T \mathbf{M}(t) \right] dt \right).$$

- No closed-form MLEs for β and ξ, but can be numerically solved.

Slide appendix: relaxing closed population

- The $N = 103$ population is in fact a sub-population, so the closed population assumption has to be relaxed;
- ξ: external infection rate.
- Slight modification; for an infection event e_j:
  - if there are possible internal infection sources, regard it as an **internal** case (labelled as $\text{Int}_j = 1$, $\rightarrow \beta$);
  - if there are no possible internal infection sources, regard it as an **external** case (labelled as $\text{Int}_j = 0$, $\rightarrow \xi$).
- Complete data likelihood:

$$\mathcal{L}(\beta, \xi, \gamma, \tilde{\alpha}, \tilde{\omega} | \mathcal{G}_0)$$

$$= \beta^{\left(n_E^{\text{int}} - \text{Int}_1\right)} \xi^{\left(n_E^{\text{ext}} - 1 + \text{Int}_1\right)} \gamma^{n_R} \alpha_{SS}^{C_{HH}} \alpha_{SI}^{C_{HI}} \alpha_{II}^{C_{II}} \omega_{SS}^{D_{HH}} \omega_{SI}^{D_{HI}} \omega_{II}^{D_{II}} \tag{3}$$

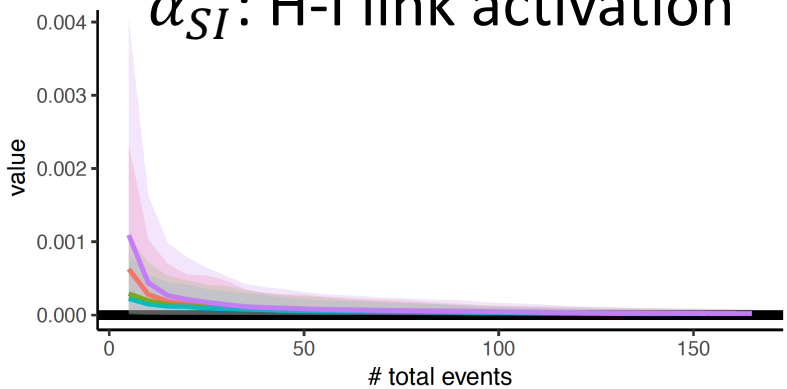
$$\times \prod_{j=2}^{n} \left[\tilde{M}(t_j) I_{p_{j1}}(t_j)^{F_j \text{Int}_j} \right]$$

$$\times \exp\left(- \int_0^{T_{\max}} \left[\beta S I(t) + \xi S(t) + \gamma I(t) + \tilde{\alpha}^T \mathbf{M}_{\max}(t) + (\tilde{\omega} - \tilde{\alpha})^T \mathbf{M}(t) \right] dt \right). \tag{4}$$

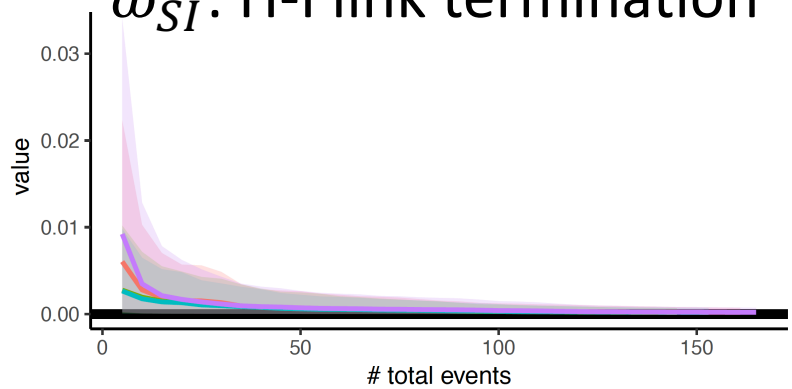
Model flexibility

- The framework generalizes:
  - static network epidemic process ($\alpha_{\cdot\cdot} = \omega_{\cdot\cdot} \equiv 0$),
  - "decoupled" network epidemic process ($\alpha_{\cdot\cdot} \equiv \alpha$, $\omega_{\cdot\cdot} \equiv \omega$).
- It can recognize those cases (e.g., static, link rates $= 0$).

α_{SI}: H-I link activation



ω_{SI}: H-I link termination



Efficiency of DARCI

Compared DARCI with two other data augmentation methods:

1. **Rejection sampling** ("**Reject**"): For $\ell = 1 : L$, keep proposing recovery times $\{r^*_{\ell,i}\}_{i=1:R_\ell} \overset{iid}{\sim} \text{TEXP}(\gamma^{(s-1)}, u_\ell, v_\ell)$ until the proposed recovery times are compatible with the observed events.
2. **Metropolis-Hastings** ("**MH**"): For $\ell = 1 : L$, propose recovery times $\{r^*_{\ell,i}\}_{i=1:R_\ell} \overset{iid}{\sim} \text{TEXP}(\gamma^{(s-1)}, u_\ell, v_\ell)$, and accept them as $\{r^{(s)}_{\ell,i}\}_{i=1:R_\ell}$ with probability

$$\min\left(1, \frac{p\left(\mathbf{x}, \{r^*_{\ell,i}\}_{i=1:R_\ell}, \{r^{(s-1)}_{\ell',i}\}_{i=1:R_{\ell'},\ell'\neq\ell} | \Theta^{(s-1)}\right) \text{pTEXP}\left(\{r^{(s-1)}_{\ell,i}\}_{i=1:R_\ell}; \gamma^{(s-1)}, u_\ell, v_\ell\right)}{p\left(\mathbf{x}, \{r^{(s-1)}_{\ell,i}\}_{i=1:R_\ell,\ell=1:L} | \Theta^{(s-1)}\right) \text{pTEXP}\left(\{r^*_{\ell,i}\}_{i=1:R_\ell}; \gamma^{(s-1)}, u_\ell, v_\ell\right)}\right).$$

Efficiency of DARCI

Take 1000 consecutive samples using each method, and evaluate

- effective sample size (ESS),
- Geweke Z-score (Geweke et al., 1991),
- two-sided p-value for the Z-score.

Statistic	β	γ	α_{II}	ω_{SI}	Method
ESS	1000.00	1000.00	1000.00	1000.00	
Z-score	-0.90	-0.20	-0.22	-0.02	**DARCI**
$Pr(>\|Z\|)$	0.37	0.84	0.82	0.99	
ESS	1000.00	1160.17	1000.00	926.63	
Z-score	0.48	-1.01	1.08	-0.16	**Reject**
$Pr(>\|Z\|)$	0.63	0.31	0.28	0.87	
ESS	566.43	1000.00	538.12	729.33	
Z-score	-1.25	-1.83	-2.09	-1.52	**MH**
$Pr(>\|Z\|)$	0.21	0.07	0.04	0.57	

Efficiency of DARCI

- Also compared **DARCI** and **Reject** in running times.
- On a dataset with 5 intervals containing various numbers of missing recovery times.

Interval	#(To recover)	Min Time		Median Time	
		Reject	**DARCI**	**Reject**	**DARCI**
1	1	227μs	224μs	484μs	245μs
2	8	285μs	287μs	563μs	319μs
3	15	163μs	161μs	279μs	181μs
4	2	138μs	138μs	153μs	156μs
5	1	133μs	133μs	146μs	147μs