

Exact Bayesian Inference for Stochastic Epidemic Models via Data Augmentation

Raphaël Morsomme, supervised by Prof. Jason Xu

December 15, 2021

Overview

- ▶ The stochastic SIR process
- ▶ Inference with (in)complete data
- ▶ A novel DA-MCMC algorithm
- ▶ Simulation study and analysis of Ebola pandemic
- ▶ Future directions

The SIR Model

- ▶ A compartmental model ($S \rightarrow I \rightarrow R$).
- ▶ Describes the spread of a contagious disease through a population.
- ▶ Deterministic or **stochastic**; in discrete or **continuous** time.

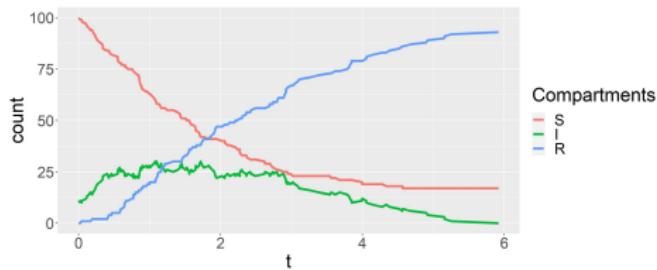


Figure: Compartment trajectories of a stochastic SIR process.

Stochastic SIR Model

Consider the agent-based epidemic process $\{X(t), t > 0\}$,

$$X(t) = (X_1(t), \dots, X_n(t)) \in \{s, i, r\}^n$$

where the agent-level subprocess

$$X_j(t) = \begin{cases} s, & t \in (0, \tau_j^I) \\ i, & t \in (\tau_j^I, \tau_j^R) , \quad i = 1, \dots, n \\ r, & t \in (\tau_j^R, \infty) \end{cases}$$

denotes the compartment of individual j over time.

Stochastic SIR Model

- ▶ Three common assumptions:
 - ▶ Homogeneously mixing population: contacts between pairs of individuals follow independent Poisson processes with rate β ,
 - ▶ Markovian assumption: iid exponentially distributed infectious periods with rate γ .
 - ▶ Closed population.
- ▶ These assumptions yield a *Markov* process characterized by the transition rates $\lambda_{x,x'} =$
$$\begin{cases} \beta I(t), & x \text{ and } x' \text{ only differ at position } j \text{ with } x_j = s \text{ and } x'_j = i, \\ \gamma, & x \text{ and } x' \text{ only differ at position } j \text{ with } x_j = i \text{ and } x'_j = r, \\ 0, & \text{otherwise.} \end{cases}$$

SIR – Population-based Formulation

The agent-based formulation is equivalent to the more typical population-based formulation with

$$W(t) = (S(t), I(t), R(t)) \in \{0, 1, \dots, n\}^3$$

and

$$\lambda_{(S, I, R), (S', I', R')} = \begin{cases} \beta S(t) I(t), & (S', I', R') = (S - 1, I + 1, R) \\ \gamma I(t), & (S', I', R') = (S, I - 1, R + 1) \\ 0, & \text{otherwise.} \end{cases}$$

The agent-based formulation will allow us to

- ▶ leave the Markovian framework,
- ▶ jointly propose latent data in our DA-MCMC.

Inference

The stochastic SIR process has a closed-form likelihood

$$\begin{aligned} L(\theta; X) &= \prod_{j \in \mathcal{I}} \beta I(\tau_j^I) \prod_{k \in \mathcal{R}} \gamma \exp \left\{ - \int_0^{t_{end}} \beta S(t) I(t) + \gamma I(t) dt \right\} \\ &= \beta^{n_I} \gamma^{n_R} \prod_{j \in \mathcal{I}} I(\tau_j^I) \exp \left\{ - \beta \int_0^{t_{end}} S(t) I(t) - \gamma \int_0^{t_{end}} I(t) dt \right\} \end{aligned}$$

where

- ▶ $\mathcal{I} = \{j : \tau_j^I \in (0, t_{end}]\}$ and $\mathcal{R} = \{j : \tau_j^R \in (0, t_{end}]\}$,
- ▶ $n_I = |\mathcal{I}|$ and $n_R = |\mathcal{R}|$

Inference

Closed-form MLE:

$$\hat{\beta} = \frac{n_I}{\int_0^{t_{end}} S(t)I(t)dt}, \quad \hat{\gamma} = \frac{n_R}{\int_0^{t_{end}} I(t)dt},$$

- ▶ the integrals correspond to finite sums ($I(t)$ and $S(t)$ are constant between event times),

and gamma conjugacy:

$$\beta \sim G(a_\beta, b_\beta), \quad \gamma \sim G(a_\gamma, b_\gamma),$$

$$\beta | X \sim G\left(a_\beta + n_I, b_\beta + \int_0^{t_{end}} I(t)S(t)dt\right),$$

$$\gamma | X \sim G\left(a_\gamma + n_R, b_\gamma + \int_0^{t_{end}} I(t)dt\right)$$

Inference - Reparameterization

The SIR likelihood can also be parameterized in terms of $\tilde{\theta} = (\beta, R_0)$ with $R_0 = S(0)\beta/\gamma$ which has the semi-conjugate prior

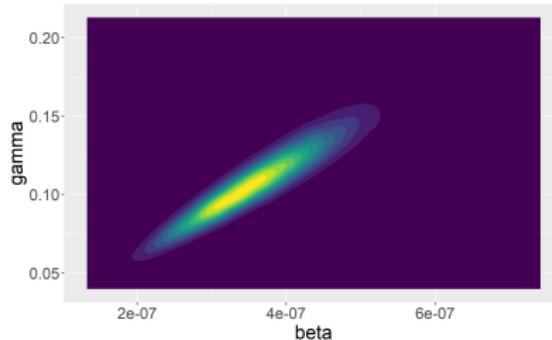
$$\beta \sim G(a_\beta, b_\beta), \quad R_0 \sim IG(a_R, b_R)$$

$$\beta | X, R_0 \sim G\left(a_\beta + n_I + n_R, b_\beta + \int_0^{t_{end}} S(t)I(t)dt + \frac{S(0)}{R_0} \int_0^{t_{end}} I(t)dt\right)$$

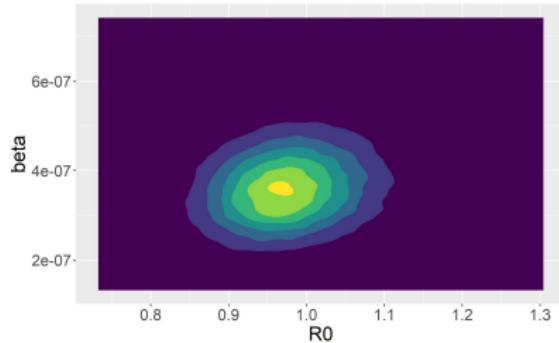
$$R_0 | X, \beta \sim IG\left(a_R + n_R, b_R + \beta S(0) \int_0^{t_{end}} I(t)dt\right)$$

\Rightarrow Gibbs sampler to explore $\pi(\tilde{\theta}|X)$.

Inference - Reparameterization



(a) $\theta = (\beta, \gamma)$



(b) $\tilde{\theta} = (\beta, R_0)$

Figure: Joint posterior given partial data under different parameterizations.

Partial Data - Incidence Data for Infection

- ▶ In practice, the process is never completely observed.
- ▶ In my research, I consider *incidence data for infections*
- ▶ Given times $t_{0:K}$, the observed data are $\mathbf{Y} = I_{1:K}$ where

$$I_k = \#\{\tau_j^I \in (t_{k-1}, t_k]\}$$

is the number of infections during the k th interval.

- ▶ E.g. weekly infection counts.
- ▶ Motivated by the 2013-2016 Ebola pandemic in Western Africa.

Intractable Partial Likelihood

The partial data likelihood

$$L(\theta; Y) = \pi(\theta) \int_{\chi_x} L(\theta; x) \delta_Y(x) dx,$$

where $\delta_Y(x) = 1$ if Y is compatible with the infection times in x and 0 otherwise, is intractable.

Prior Work

- ▶ Direct method to compute $L(\theta; Y)$ (Ho et al., 2018)
 - ▶ Prohibitively slow.
- ▶ Approximate likelihood (Fintzi et al., 2020)
 - ▶ Assumptions questionable in small populations.
- ▶ Particle filtering (King - 2015)
 - ▶ Often degenerates.
- ▶ DA-MCMC with RJ (Gibson and Renshaw - 1998)
 - ▶ Single-site update MH step.
- ▶ DA-MCMC with MH (Fintzi et al., 2017)
 - ▶ Single-site update MH step.
- ▶ DA-MCMC with Gibbs (Touloupou et al., 2020)
 - ▶ Discrete time, single-site update Gibbs step.
 - ▶ No likelihood evaluation (expensive step), but FFBS.

DA-MCMC

The closed-form likelihood $L(\theta; X)$ and its (semi) conjugacy suggests using Data-Augmentation MCMC:

- ▶ Construct latent $Z = \left\{ (z_j^I, z_j^R) \right\}_{j=1}^n$ and explore $\pi(\theta, Z|Y)$.
- ▶ Naive approach: Gibbs sampler
 - ▶ $\theta|Y, Z$ – (semi) conjugacy
 - ▶ $Z|Y, \theta$ – prohibitively difficult

Existing DA-MCMC approaches substitute the difficult Gibbs step $Z|Y, \theta$ with a simpler single-site-update MH step.

- ▶ The resulting Markov chain is very sticky.
- ▶ Limited to populations with $n < 1000$ individuals.

We propose a **DA-MCMC algorithm in which the entire latent data Z are jointly proposed in a MH step.**

Piece-wise decoupled SIR

Main idea of our DA-MCMC: use a *surrogate* process to generate latent data Z^* compatible with the observed data Y and accept/reject Z^* in a MH step.

To generate Z^* compatible with $Y = I_{1:K}$, proceed one interval at a time.

- ▶ Generate the I_k infection times in (t_{k-1}, t_k) using some procedure, e.g. iid $U(t_{k-1}, t_k)$.
- ▶ Generate the removal times of these individuals by adding an exponential RV to their infection times.

We want to generate the infection times using a procedure that is as close as possible to the SIR.

Piece-wise decoupled SIR

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

1. Let $\tilde{\mu}_{pop}$ be constant over each interval $[t_{k-1}, t_k]$:

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

The I_k infection times follow a Poisson process (iid uniform).

2. 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

$$\tilde{\mu}(t) := \frac{\tilde{\mu}_{pop}(t)}{S(t)} = \beta I(t_{k-1}) = \mu_k \text{ is piece-wise constant.}$$

- The compartment $S(t)$ follows a linear death process (LDP) where infections correspond to deaths.

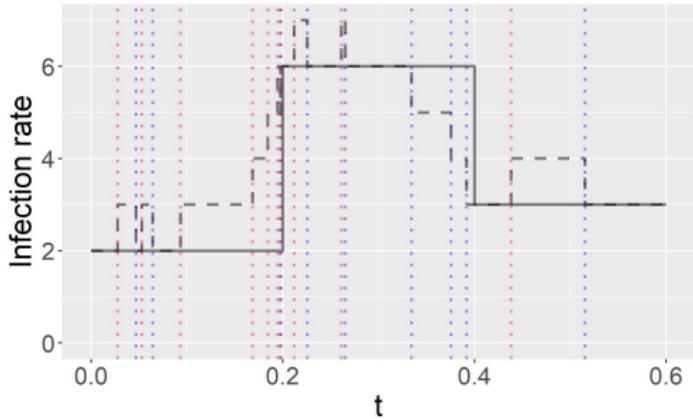
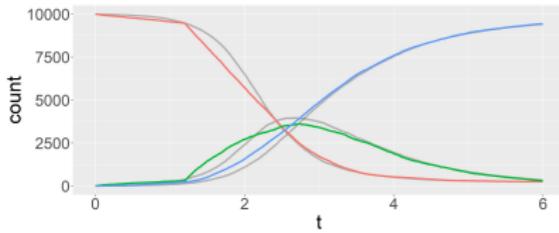
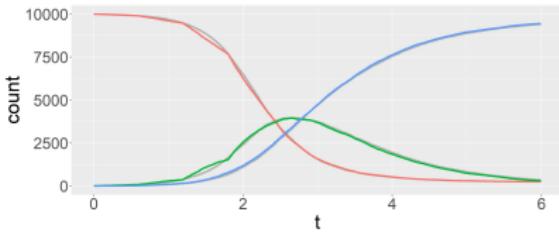


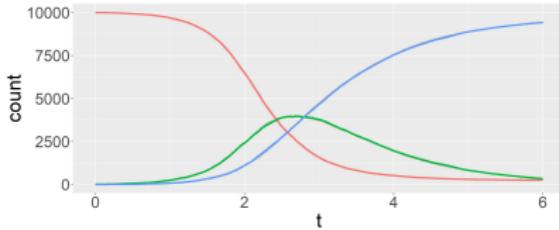
Figure: Individual infection rate under the SIR (dashed) and PD-SIR (solid) processes. Small population ($n = 10$) and $t_{0:3} = (0, 0.2, 0.4, 0.6)$.



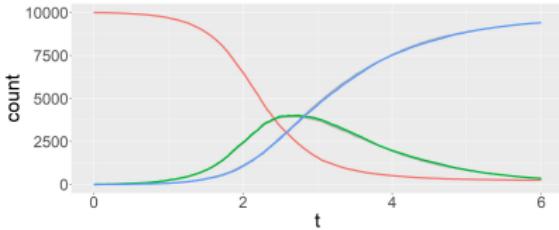
(a) $K = 5$



(b) $K = 10$



(c) $K = 50$



(d) $K = 1000$

Figure: SIR (in color) and PD-SIR process (in grey).

A Useful Theorem

Theorem

Consider a LDP with death rate μ and let $\tau_{1:N} \in (t_l, t_u]$ be the times of the N deaths occurring between times t_l and t_u . Then

$$\tau_j \stackrel{d}{=} X_{(j)}, \quad j = 1, \dots, N$$

where $X_{(j)}$ is the j^{th} order statistics of N iid random variables following a truncated exponential distribution with rate μ , lower bound t_l and upper bound t_u .

⇒ In the PD-SIR process, $z_{\mathcal{J}_k}^I | I_k \sim \text{TruncExp}(\mu_k; t_{k-1}, t_k)$ iid.

Simulating the PD-SIR – pseudo-algorithm

Algorithm 1: Simulating a PD-SIR process conditionally on Y

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

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

 Compute the infection rate:

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

 Jointly generate the infection times:

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

 Generate the removal times:

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

end

A few Remarks on the PD-SIR

- ▶ The DA-MCMC can be initialized with $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)})$ only.
- ▶ Gibbs-like scheme: $Z^* \perp Z^{(k-1)} | \theta^{(k)}$.
 - ▶ Uniform ergodicity
- ▶ Efficient:
 - ▶ Z^* is compatible with $Y = I_{1:K}$ by construction.
 - ▶ High MH acceptance rate due to close resemblance between the PD-SIR and SIR processes.
- ▶ Fast:
 - ▶ Only requires uniform RVs (inverse CDF),
 - ▶ Vectorized implementation (rate μ_k is only updated K times).

Uniform Ergodicity

A Markov chain is *uniformly ergodic* if for some finite M and positive constant $r < 1$, the n -step transition kernel P^n satisfies

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq Mr^n, \quad \forall x \in \chi$$

- ▶ Uniform ergodicity ensures a CLT (Jones, 2004).

We show that the state space $\chi = \chi_\theta \times \chi_z$ is a *small set*,

$$P(x, \cdot)^m \geq \epsilon \nu(\cdot), \quad \forall x \in \chi$$

for some measure ν , positive integer m ($m = 1$) and constant $\epsilon > 0$, which holds iff P is uniformly ergodic (Tierney, 1994).

Uniform Ergodicity – Gist of Proof

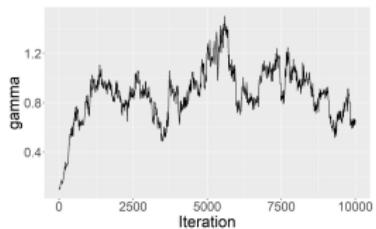
- ▶ $P = P_\theta P_z$ is composite:

$$x_1 = (\theta_1, z_1) \rightarrow (\theta_2, z_1) \rightarrow (\theta_2, z_2) = x_2.$$

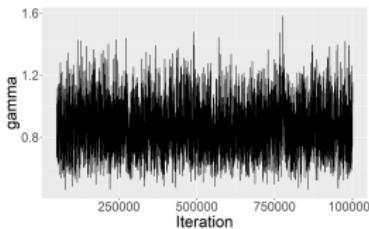
- ▶ Minorize P_θ and P_z separately.
 - ▶ P_θ : $q_\theta(\theta_2|z_1)$ is a product of two gammas with bounded parameters.
 - ▶ P_z : $q_z(z_2|\theta_2)$ is a product of positive factors (truncated exponentials and tail probabilities).
- ▶ Interesting result, since the requirement that the whole space be small is typically not satisfied for models with unbounded spaces e.g. $\chi_\theta = \mathbb{R}^+ \times \mathbb{R}^+$.

Simulation Study – Proof of Concept

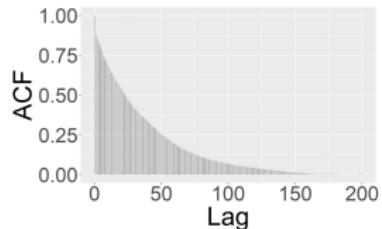
- ▶ $n = 1000; (\beta, \lambda) = (0.0025, 1)$ [$R_0 = 2.5$]
- ▶ $I_{1:10} = (13, 31, 55, 71, 103, 111, 135, 103, 74, 42)$
- ▶ $(\beta_0, \gamma_0) = (\beta/10, \gamma/10)$ (low density region)
- ▶ Stationarity is reached within 1500 iterations
- ▶ $1e6$ iterations take < 30 minutes on a laptop.



(a) Transient phase



(b) Recurrent phase



(c) ACF

Figure: Performance of our DA-MCMC in a medium-sized population.

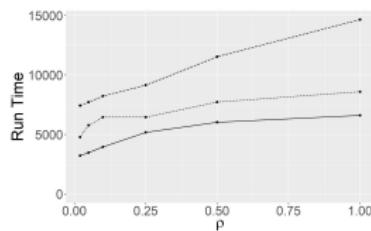
Simulation Study – Coverage

- ▶ Repeat previous experiment 2000 times.
- ▶ Coverage of 90% credible intervals.
- ▶ Weakly informative priors: $\beta \sim Ga(0.001, 1)$, $R_0 \sim IG(2, 2)$.

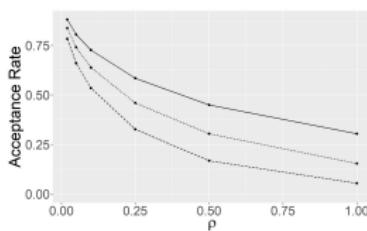
Parameter	Observed coverage rate	Average of the posterior means	SD of the posterior means
β (0.0025)	0.895	0.0025	0.000255
γ (1)	0.902	1.00	0.142
R_0 (2.5)	0.910	2.54	0.194

Simulation Study – Tuning Parameter ρ

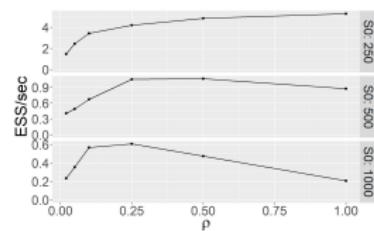
- ▶ For large n , the MH acceptance rate can be too low.
- ▶ To improve mixing, only update the trajectories of $\lceil \rho n \rceil$ randomly chosen individuals ($0 < \rho \leq 1$) per iteration.
- ▶ $n \in (250, 500, 1000)$, $\rho \in (0.02, 0.05, 0.1, 0.25, 0.5, 0.75, 1)$.



(a) Run time in seconds



(b) Acceptance rate



(c) ESS/sec for R_0

Figure: Impact of ρ on the performance of the DA-MCMC in a population of size 250 (solid), 500 (dotted) and 1000 (dashed).

Simulation Study – Comparison with SSU DA-MCMC

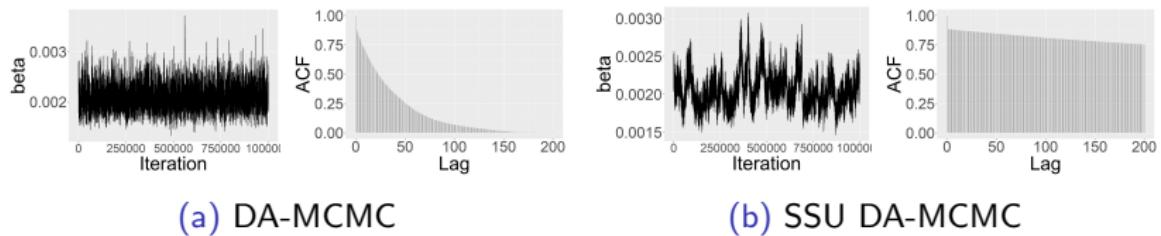


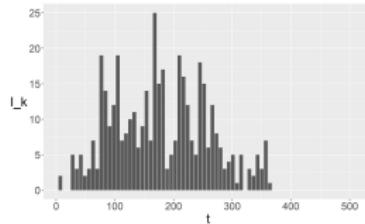
Figure: Traceplot and ACF of our DA-MCMC and a SSU DA-MCMC for β .

Parameter	DA-MCMC	SSU DA-MCMC
β	0.20	0.01
γ	0.19	0.01
R_0	0.38	0.05

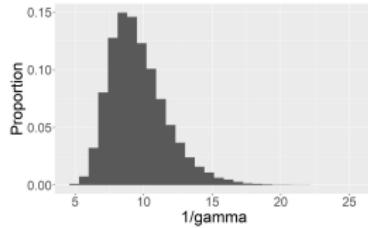
Table: ESS/sec for our DA-MCMC and the SSU DA-MCMC.

2013-2016 Ebola Pandemic in Western Africa

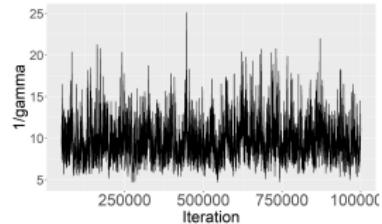
- ▶ Highly contagious virus, > 10000 victims during the 2013-2016 pandemic in Western Africa.
- ▶ Weekly infection counts per province (Gueckedou).
- ▶ $n = 290,000$; $\rho = 0.1$ (acceptance rate: 0.21)
- ▶ 1e6 iterations, 470 ESS.



(a) Observed data



(b) Posterior distribution



(c) Traceplot

Figure: Observed data for the Guéckédou prefecture and posterior distribution and traceplot of the expected infection length (γ^{-1}).

Extensions

We have extended the PD-SIR to

- ▶ non-Markovian SIR,
- ▶ SIR with heterogeneously mixing population.

I am mentoring a student who works on

- ▶ SIR with time varying infection rate.

I plan to work on the following problems

- ▶ under-reporting e.g. $\tilde{I}_k \sim \text{Bin}(I_k, \kappa)$,
- ▶ IR approximation to the SIR (small pandemic in large population, e.g. Ebola) $\Rightarrow I(t)$ is a linear birth-death process,
- ▶ applying our DA-MCMC to other processes where we observe incidence data (as opposed to prevalence data).

Conclusion

Our DA-MCMC is an efficient algorithm for exact Bayesian inference for the stochastic SIR with infection incidence data.

- ▶ carefully designed surrogate process: efficient and fast.

The combination of (i) the agent-based formulation and (ii) the PD-SIR provides a unified inferential framework for the SIR with

- ▶ non-Markovian dynamics,
- ▶ a heterogeneously mixing population,
- ▶ a time varying infection rate.

Extension – Arbitrarily Distributed Infection Periods

- ▶ Weibull distribution is more plausible for infectious periods.
- ▶ Dropping the assumptions of exponentially distributed infectious periods makes the process *Non-Markovian*.
- ▶ Yet, the likelihood still has a closed form!
- ▶ PD-SIR with

$$z_j^R | z_j^I \sim z_j^I + \text{Weibull}(.)$$

Extension – Heterogeneously Mixing Population

- ▶ Severo (1969) proposes to substitute the population-level infection rate of the SIR

$$\mu_{pop}(t) = \beta S(t)I(t)$$

with

$$\bar{\mu}_{pop}(t) = \beta S(t)^{1-b}I(t), \quad b \in [0, 1]$$

to model a heterogeneously mixing population.

- ▶ The corresponding individual-level infection rate is

$$\bar{\mu}(t) = \frac{\bar{\mu}_{pop}(t)}{S(t)} = \beta S(t)^{-b}I(t)$$

- ▶ PD-SIR with

$$\mu_k = \tilde{\mu}(t_{k-1}) = \beta S(t_{k-1})^{-b}I(t_{k-1}).$$

Extension – Time-varying Infection Rate $\beta(t)$

- ▶ $\beta \sim GP(.)$ (Kypraios, 2018)
 - ▶ expensive: invert a matrix of order $n \times n$ each iteration
- ▶ Random walk: $\Delta(\beta)_k \sim N(0, \sigma^2)$
 - ▶ multivariate normal prior \Rightarrow elliptical slice sampler
- ▶ Locally adaptive: $\Delta(\beta)_k$ follows a Laplace or HS distribution (Faulkner & Minin, 2018)
 - ▶ accommodates sudden variations in β .

Uniform Ergodicity – Minorizing P_θ

The density of the Gibbs kernel P_z corresponds to the product of the two gamma densities

$$Ga\left(\beta; a_\beta + n_I, b_\beta + \int S(t)I(t)dt\right) Ga\left(\gamma; a_\beta + n_R, b_\beta + \int I(t)dt\right)$$

Since the sufficient statistics of Z are bounded:

$$0 \leq \int S(t)I(t)dt \leq n^2 t_{end}, \quad 0 \leq n_R \leq n, \quad 0 \leq \int I(t)dt \leq nt_{end},$$

each density possesses a positive minorization whose closed form we have derived.

Uniform Ergodicity – Minorizing P_z

The MH kernel P_z depends on the current latent data z only through the ratio $q(z|\theta)/\pi(z|\theta)$.

- ▶ q and π can be bounded above and below away from 0.

It is illuminating to observe that

- ▶ the contribution of the removal times in the numerator and the denominator cancel each other (generated from identical dynamics),
- ▶ the ratio of infections times' contribution

$$\prod_{k=1}^K \prod_{j \in \mathcal{I}_k} \frac{\beta I(t_{k-1}) \exp\{-\beta I(t_{k-1})(z_j^I - t_{k-1})\}}{\beta I(z_j^I) \exp\{-\beta \int_{t_{k-1}}^{z_j^I} I(t) dt\}}$$

will typically be close to 1.