Bayesian Data Augmentation for Partially Observed Stochastic Compartmental Models

Shuying Wang
Department of Statistics and Data Sciences, University of Texas at Austin
and
Stephen G. Walker
Department of Statistics and Data Sciences, University of Texas at Austin

Abstract

Deterministic compartmental models are predominantly used in the modeling of infectious diseases, though stochastic models are considered more realistic, yet are complicated to estimate due to missing data. In this paper we present a novel algorithm for estimating the stochastic SIR/SEIR epidemic model within a Bayesian framework, which can be readily extended to more complex stochastic compartmental models. Specifically, based on the infinitesimal conditional independence properties of the model, we are able to find a proposal distribution for a Metropolis algorithm which is very close to the correct posterior distribution. As a consequence, rather than perform a Metropolis step updating one missing data point at a time, as in the current benchmark Markov chain Monte Carlo (MCMC) algorithm, we are able to extend our proposal to the entire set of missing observations. This improves the MCMC methods dramatically and makes the stochastic models now a viable modeling option. A number of real data illustrations and the necessary mathematical theory supporting our results are presented.

Keywords: Data Augmentation; SIR & SEIR Models; Markov chain Monte Carlo.

1 Introduction

Compartmental models, at any given point in time, partition a population of individuals into different compartments or states. The aim is to model the transitions of individuals between compartments. The deterministic compartmental models are predominantly used in analyzing outbreaks of infectious diseases, see for example Dehning et al. (2020) used the deterministic SIR model to fit the COVID-19 outbreak in Germany assuming that there is random noise in the real-world observations but the underlying epidemic process is deterministic. The deterministic
models are usually simpler to handle because of the existence of analytical solutions to the differential equation system, but the stochastic SIR or SEIR models are considered more realistic due to the nature of the epidemic processes (Roberts et al., 2015). Further, stochastic models are more flexible when modeling the time-varying transition rates (Roberts et al., 2015). Also, the deterministic models are only appropriate for a sufficiently large population size, because the number of individuals are considered continuous variables in the differential equations while they are actually discrete (Brauer, 2008). Therefore, “stochastic models remain preferable when their analysis is possible” (Ho et al., 2018). However, the current benchmark Bayesian MCMC approaches using data augmentation with the partially observed epidemic processes (Gibson and Renshaw, 1998; O’Neill and Roberts, 1999) are not up to the task, because they are “very challenging and time-consuming, and for large systems with many hidden states, they can become computationally infeasible”, which limits its use in practical applications (Swallow et al., 2022). Other methods include Approximate Bayesian Computation (ABC) (McKinley et al., 2009; Blum and Tran, 2010; Neal, 2012), and sequential Monte Carlo (SMC) (Dukic et al., 2012; King et al., 2016) suffer from the same problem. In this paper, we develop a new MCMC algorithm with a novel data augmentation method that used the infinitesimal independence property of stochastic compartmental models to solve the missing data problem.

With the general stochastic epidemic SIR model (Bailey, 1975), the population is divided into three compartments; $S =$ susceptible, $I =$ infected and $R =$ recovered, with two types of transition; from susceptible to infectious and from infectious to recovery. It is assumed not possible to return from recovered to either susceptible or infected. More complex epidemic models incorporate additional compartments and types of transition. For example, the SEIR model includes the incubation period of the disease in a compartment called “exposed”, labeled as $E$.

The foundational framework of stochastic compartmental models lie in non–homogeneous Poisson processes which are assumed to have the Markov property. Interest focuses on the transition rates between compartments. For the SIR model, the parameters are the infection rate $\beta$ and the recovery rate $\gamma$, which yield the reproduction number $\beta/\gamma$ representing the expected number of new infections from a single infected individual from the population. The major difficulty for likelihood–based statistical inference is that the epidemic process is often, if
not always, partially observed; see Bailey (1975). However, the likelihood function can only be tractable with the complete process; it being infeasible to integrate out the unobserved part of the process. There are two kinds of scenarios here: Firstly, in a given time period, the number of infections and recoveries are known, but the exact time points when these events occur are unknown. For example, there can be a daily report of diagnosed and recovered cases. Secondly, either the infection or recovery process can be completely unobserved, which means the number of infections or recoveries are also unknown. There can, for example, be a daily report of the diagnosed cases but with no information about recoveries. In this case, the problems are more difficult to handle because the unobserved cases also comes with an unknown dimension.

Numerous methods have been developed to solve these two problems. Martingale methods and the EM algorithm have been used to obtain maximum likelihood estimators of the infection rate with a completely observed recovery process and a partially observed infection process; see Becker (1993); Becker and Hasofer (1997). For Bayesian approaches using MCMC methods, sampling the unobserved part of the epidemic process as latent variables, and therefore obtaining the full likelihood and posterior distribution of the parameters when the infection or removal process is completely unobserved, has been described in Gibson and Renshaw (1998); O’Neill and Roberts (1999); Gibson and Renshaw (2001). Rose et al. (2020) adopted this MCMC data augmentation approach to fit a stochastic compartmental model for infestation data to help locate infested homes in urban areas. Another recent application of this method is by Pooley et al. (2020), who introduced a software tool called SIRE to estimate genetic and non-genetic effects in epidemic processes. Further, Cauchemez and Ferguson (2008) used a diffusion process to approximate the number of infected in the SIR epidemic process for data augmentation, but this kind of approximation is less accurate with small populations. Keeling and Ross (2008) proposed a computational method using the Kolmogorov forward equation, in which they were able to compute the likelihood value of the SIR model with incomplete data; however, the high computational demand can be a problem because this method involves the calculation of a matrix exponential. Another computational method reparameterizes the stochastic compartmental process into a multivariate birth process and estimates the transition probabilities using a Laplace transformation, which is more computationally efficient than the matrix exponential method; see Crawford and Suchard (2012); Crawford et al. (2018); Ho et al. (2018). However, these two com-
putational methods are only applicable when both infection and removal processes are discretely observed, so they cannot deal with the situation when the number of infections or recoveries are unknown.

This paper focuses on the classic MCMC algorithm proposed by O’Neill and Roberts (1999) who developed a proposal distribution for the missing observations. The limitation of the existing MCMC algorithm is the convergence speed; when sampling the unobserved infection process with a reversible jump algorithm, the step size of the Metropolis–Hastings sampler is small because only one time point is updated at each iteration, hence it is extremely slow to explore the parameter space, especially with large data sets (Swallow et al., 2022). Our contribution is to introduce a new proposal idea which is developed to sample the unobserved infection or recovery process with larger step sizes at each iteration, and as we shall see, greatly improving the convergence speed, to such an extent that we believe the Bayesian MCMC could become the algorithm of choice for estimating SIR and SEIR models.

Before proceeding, we briefly mention some further relevant papers. Ball and Donnelly (1995) showed that a general branching process; i.e. a linear birth–death process can be used to approximate the general epidemic process in the early stage for a large population. Lekone and Finkenstädt (2006) built a discrete-time SEIR model to approximate the underlying continuous-time process for an outbreak of Ebola by using binomial distributions to model the number of transitions in each time interval. Jewell et al. (2009) used an SIR type model for emerging infectious diseases in UK farms. They used a non–centered parameterization for the unobserved infection times and also used a reversible jump sampler to deal with the unknown number of infections. Dukic et al. (2012) fit a state–space extension of the discrete–time deterministic SEIR model with the Google flu trends data. They used a discretized time–step of one week and assumed that the epidemic process was observed with error. Neal and Xiang (2017) used a non–centered parameterization to construct a collapsing MCMC algorithm with the infection rate parameter integrated out when sampling other variables in the chain. Finally, Fintzi et al. (2022) used linear noise approximation to approximate the transition density of SIR process between discrete time points with a Gaussian density and implement it into a data augmentation MCMC framework.

The layout of the paper is as follows. In Section 2 we review the current standard MCMC
algorithm, and in Section 3 we introduce our algorithm based on the infinitesimal conditional independence structure of the model. Section 4 details necessary mathematical background and Section 5 contains illustrations and examples. Section 6 is a conclusion section and the Appendix provides a proof to one of the results established in Section 4. R code for the illustrations and examples is provided in the GitHub link.

2 Model and Benchmark MCMC Algorithm

2.1 The SIR Model

The general SIR model divides the population into three compartments, namely susceptible, infected, and recovered/removed. For notation we write \((S_t, I_t, R_t)\) to denote the number of individuals in each compartment at time \(t \in [0, T]\). It is assumed that the population is closed without any natural births, deaths or immigration, and the recovered individuals are immune from further infection. Based on these two assumptions, there are two types of transition between the three compartments. Infection is the transition from susceptible to infected and recovery is the transition from infected to recovered. In this paper we look at the two transition processes, the infection process \(X = \{X_t\}_{0 \leq t \leq T}\), and the recovery process \(Y = \{Y_t\}_{0 \leq t \leq T}\). Therefore, with the population size fixed at \(N\) and \(X_0 = Y_0 = 0\), the state \((S_t, I_t, R_t)\) can be fully represented by \((X_t, Y_t)\) at a given time point \(t\), with \(S_t = S_0 - X_t\), \(I_t = I_0 + X_t - Y_t\), and \(R_t = N - S_t - I_t\). The infection rate is denoted by \(\beta\), and the recovery rate is denoted by \(\gamma\). With the Markov assumption, the time intervals between transitions are exponentially distributed. In a small time interval \((t, t + dt)\), the probability of an infection is \(\beta S_t I_t dt / N\), and the probability of a recovery is \(\gamma I_t dt\). The infinitesimal probabilities are given by

\[
P(X_{t+dt} = X_t + 1 \mid S_t, I_t) = \beta S_t I_t dt / N + o(dt) \quad \text{and} \quad P(Y_{t+dt} = Y_t + 1 \mid S_t, I_t) = \gamma I_t dt + o(dt),
\]

with \(t\) restricted to the time interval \([0, T]\). Denote the total number of infections and recoveries during this time period by \((n_x, n_y)\), and denote the time points for the occurrence of transitions by \((\{t^x_i\}_{i=1}^{n_x}, \{t^y_i\}_{i=1}^{n_y})\). Let \((x, y)\) denote the right continuous sample paths of \(\{X_t, Y_t\}_{0 \leq t \leq T}\),

\[1\text{https://github.com/ShuyingWang/SIR-SEIR-Model-MCMC}\]
characterized by \((n_x, \{t^x_i\}^n_x_{i=1})\) and \((n_y, \{t^y_i\}^n_y_{i=1})\). The path functions are given by,

\[
x(t) = \sum_{i=1}^{n_x} 1(t^x_i \leq t) \quad \text{and} \quad y(t) = \sum_{i=1}^{n_y} 1(t^y_i \leq t), \quad \forall \ t \in [0, T].
\]

(1)

Now let \(S(t)\) and \(I(t)\) denote the path functions of \(S_t\) and \(I_t\) represented by \(x\) and \(y\), and let \(S(t^x_i -) = S_0 - (i - 1)\) and \(I(t^x_i -) = I_0 + (i - 1) - Y_t^i\) be the state of \((S_t, I_t)\) just before the \(i\)th infection event, with \(t^x_i -\) being the left limit of \(t^x_i\), and similarly for \(I(t^y_i -)\).

Assume the initial state \((S_0, I_0, R_0)\) is known, then the likelihood function of the process \((X,Y)\) between times 0 and \(T\) is given by,

\[
p(x,y | \beta, \gamma) = \exp \left\{ -\int_0^T \beta \frac{S(t)I(t)}{N} dt \prod_{i=1}^{n_x} \beta \frac{S(t^x_i -)I(t^x_i -)}{N} \right\} \exp \left\{ -\int_0^T \gamma I(t) dt \prod_{i=1}^{n_y} \gamma I(t^y_i -) \right\}.
\]

(2)

Introducing conjugate Gamma priors \(\pi_0(\beta) \equiv \text{Gamma}(a_\beta, b_\beta)\) and \(\pi_0(\gamma) \equiv \text{Gamma}(a_\gamma, b_\gamma)\), it is straightforward to sample from the full conditional posterior distributions of the parameters \((\beta, \gamma)\), given by

\[
p(\beta | -) \equiv \text{Gamma} \left( n_x + a_\beta, N^{-1} \int_0^T S(t)I(t) dt + b_\beta \right)
\]

\[
p(\gamma | -) \equiv \text{Gamma} \left( n_y + a_\gamma, \int_0^T I(t) dt + b_\gamma \right),
\]

(3)

where the \(-\) denotes all other variables. However, as has been previously mentioned, this simplicity with these distributions is only the case when the process is completely observed.

2.2 Current Benchmark MCMC Algorithm

Although stochastic compartmental models have a closed form likelihood function, it is hard to integrate out unobserved states when the data is incomplete. A standard Bayesian approach is to implement a data augmentation scheme by using an MCMC algorithm to sample the unobserved part of the process as latent variables; see Gibson and Renshaw (1998); O’Neill and Roberts (1999); Gibson and Renshaw (2001). The problem setting is typically that one of the infection and recovery processes is completely unobserved and the other partially or completely observed. In the following we assume that the initial state is known, and there is a single observation
of the infection process at time point $T$, and the recovery process is completely unobserved. The method described can easily be extended to multiple observations at different discrete time points.

Consequently, the observed data consists of the initial state of the process, i.e. $(S_0, I_0, R_0)$, and a single observation $X_T = n_x$. The unobserved latent variables include the time point of infections $(t^x_i)_{i=1}^{nx}$, and the whole recovery process $(Y_t)_{0 \leq t \leq T}$. The target is to estimate the parameters $(\beta, \gamma)$ as well as the number of recoveries $n_y$ during the time interval $[0, T]$.

With the conjugate Gamma priors, the parameters $(\beta, \gamma)$ can be directly sampled from their full conditional distributions via (3). It is also straightforward to sample the infection time points $(t^x_i)_{i=1}^{nx}$ one at a time using a Metropolis–Hastings step within a Gibbs framework,

$$p(t^x_i \mid -) \propto S(t^x_i -)I(t^x_i -) \prod_{j=1}^{n_y} I(t^y_j -) \exp \left\{ \int_0^T \beta S(t)I(t) \frac{N}{N} + \gamma I(t)dt \right\}.$$  

However, the latent recovery process $(Y_t)_{0 \leq t \leq T}$ cannot be updated in this way because the number of time points $n_y$ is unknown. The classic MCMC algorithm samples this kind of latent process using a reversible jump component within the MCMC algorithm. There are three possible types of move at each iteration. A proposal to add one more time point, delete an existing time point, or move an existing time point. When adding a new time point, it is sampled uniformly in $[0, T]$. When removing a time point, it is uniformly selected from all the existing time points. When moving a time point, a combination of the above two steps is used. Succinctly, and with acceptance probabilities attached,

- **Add a new time point**: $t' \sim U(0, T)$, $n'_y = n_y + 1$, $\alpha(y, y') = \min \left\{ 1, \frac{p(y' \mid -)T}{p(y \mid -)(n_y + 1)} \right\}$.

- **Remove an existing time point**: $n'_y = n_y - 1$, $\alpha(y, y') = \min \left\{ 1, \frac{p(y' \mid -)n_y}{p(y \mid -)T} \right\}$.

- **Move an existing time point**: $n'_y = n_y$, $\alpha(y, y') = \min \left\{ 1, \frac{p(y' \mid -)}{p(y \mid -)} \right\}$.

There are some limitations with the reversible jump part. The step size is very small, because only one change can occur at each iteration, so convergence and mixing can be slow. There is also high auto–correlation between iterations, which means a run of the algorithm requires a large number of iterations to get suitable posterior samples.
3 A New MCMC Algorithm

3.1 A Likelihood Factorization

The Metropolis–Hastings procedure relies heavily on a good proposal distribution; as close to the target distribution as possible, while being able to sample easily from the proposal.

The target distribution for sampling the recovery process \((Y_t)_{0 \leq t \leq T}\) is given by

\[
p(y \mid -) \propto \prod_{i=1}^{n_y} I(t^y_{i-}) \exp\left\{ - \int_0^T \frac{\beta \, S(t) I(t)}{N} \, dt \right\} \prod_{i=1}^{n_y} \gamma I(t^y_i) \exp\left\{ - \int_0^T \gamma I(t) \, dt \right\}.
\]

Our aim is to find a proposal distribution which approximates this very well.

To this end, consider a time–inhomogeneous birth process \((Y'_t)_{0 \leq t \leq T}\), with intensity function \(\lambda^y_k(t) = \max\{0, \gamma(I_0 + x(t) - k)\}\), which is the birth rate for a birth process at time point \(t\) when \(Y'_t = k\). Here \(x(t)\) is considered to be a deterministic function of \(t\), as defined in (1). With the number and time points of births denoted by \(n_y\) and \(\{t^y_i\}_{i=1}^{n_y}\), the likelihood function of this time–inhomogeneous birth process \(Y'\) is given by

\[
f_{Y' \mid X}(y \mid x) = \exp\left\{ - \int_0^T \gamma I(t) \, dt \right\} \prod_{i=1}^{n_y} \gamma I(t^y_i),
\]

where \(I(t^y_i) = I_0 + x(t^y_i) - (i - 1)\). We will use this time-inhomogeneous birth process as the proposal distribution for process \(Y\). Likewise, consider the time–inhomogeneous birth process \(X'\) with intensity function \(\lambda^x_k(t) = \max\{0, \beta(S_0 - k)(I_0 + k - y(t))/N\}\), and likelihood function

\[
f_{X' \mid Y}(x \mid y) = \exp\left\{ - \int_0^T \frac{\beta S(t) I(t)}{N} \, dt \right\} \prod_{i=1}^{n_x} \beta S(t^x_i) I(t^x_i).
\]

It is not surprising to find that the likelihood function of the general epidemic process in (2) can be factorized as the product of the likelihoods of these two time–inhomogeneous birth processes in (5) and (6). For notation simplicity, for the following, we are going to write the likelihood using a factorization of the form

\[
p(x, y) = f_{X' \mid Y}(x \mid y) f_{Y' \mid X}(y \mid x),
\]

where \(f_{X' \mid Y}\) and \(f_{Y' \mid X}\) are given by (6) and (5), respectively. Note these are not the conditional distribution from \(p(x, y)\).
This factorization property can be easily extended to other stochastic compartmental models. For a compartmental model with three types of transitions, such as the epidemic SEIR model, the likelihood can be expressed as the product of three time-inhomogeneous birth process densities.

### 3.2 The New MCMC Algorithm

For the new MCMC algorithm we introduce, the same approach as the current MCMC algorithm is used to update the parameters $\beta$, $\gamma$ and the partially observed infection process $X$ as in Section 2, while a different approach is used to update the completely unobserved recovery process $Y$. The current MCMC algorithm uses a reversible jump method to update $Y$; the new MCMC algorithm takes advantage of the likelihood factorization (7) to develop a new proposal distribution for the $Y$ process.

It is not difficult to generate a time-inhomogeneous birth process when the intensity rate function is known, and it is even easier when the time-dependent term $x(t)$ is piece-wise constant. Therefore, the proposal distribution we use for the missing $Y$ process is $f_Y(y \mid x)$, i.e. (5). The acceptance probability is given by

$$
\alpha(y, y') = \min \left\{ 1, \frac{f_X(x \mid y') f_Y(y' \mid x) f_Y(y \mid x)}{f_X(x \mid y) f_Y(y \mid x) f_Y(y' \mid x)} \right\} = \min \left\{ 1, \frac{f_X(x \mid y')}{f_X(x \mid y)} \right\},
$$

where $Y'$ represents the proposed value. With this proposal distribution, the whole recovery process $\{Y_t\}_{0 \leq t \leq T}$ will be updated, so the step size becomes large, while the acceptance rate remains acceptable. In fact, as will be shown, the proposal distribution is close to the target distribution; see Section 4.

Sampling from the time-inhomogeneous birth process is similar to sampling from a time-inhomogeneous Poisson process. The only difference between these two is that the intensity of the time-inhomogeneous birth process will depend on both time and the current state of the process, yet the intensity of a Poisson process will only depend on time. Since the current state of the process will stay unchanged between each jump, we can treat the birth process as a piece-wise Poisson process.

The time-inhomogeneous Poisson process can be simulated using a standard Poisson process; (Kingman, 1992). To be more specific, for a time-inhomogeneous Poisson process with intensity function $\lambda(t)$, the integral of the intensity function is denoted by $\Lambda(t) = \int_0^t \lambda(s)ds$, which is the
expected number of jumps between time 0 and \( t \). The density of the waiting time before the first jump is given by \( f_T(t) = \lambda(t)e^{-\Lambda(t)} \), and can be sampled from using the transformation of a standard exponential distribution. Take a standard exponential random variable sample \( u \sim \text{exp}(1) \), then the waiting time before the first jump is given by \( t = \Lambda^{-1}(u) \). The transformation of the distribution is given by \( f_T(t) = p(u) \frac{du}{dt} = \lambda(t)e^{-\Lambda(t)} \). Similarly, for a time–inhomogeneous birth process \( Y' \) with intensity function \( \lambda_{i}^{y}(t) \), where \( i \) represent the state of \( Y' \) at time \( t \), the waiting time before the next jump, namely the time interval between the \( i \) th and \( (i+1) \) th jump, can be sampled by \( t_{i+1}^{y} - t_{i}^{y} = \Lambda_{i}^{-1}(u_{i}) \), where \( \Lambda_{i}(t) = \int_{t_{i}^{y}}^{t_{i+1}^{y} + t} \lambda_{i}^{y}(s) \, ds \) and \( u_{i} \sim \text{exp}(1) \). Hence, the distribution of \( t_{i+1}^{y} - t_{i}^{y} \) conditional on \( t_{i}^{y} \) can be understood from

\[
p(t_{i+1}^{y} - t_{i}^{y} \mid t_{i}^{y}) = \frac{p(u_{i})du_{i}}{d(t_{i+1}^{y} - t_{i}^{y})} = \lambda_{i}^{y}(t_{i+1}^{y}) \exp\{-\Lambda_{i}(t_{i+1}^{y} - t_{i}^{y})\}.
\]

We start by sampling the first waiting time \( t_{1}^{y} = \Lambda_{0}^{-1}(u_{0}) \) with \( u_{0} \sim \text{exp}(1) \), and then sample \( t_{2}^{y} - t_{1}^{y} \) to get \( t_{2}^{y} \). Keep sampling until \( t_{i+1}^{y} > T \) and then take \( n_{y} = i \). Therefore, the proposal sample \( Y' \) can be written as a deterministic function of the i.i.d. standard exponential random variables \( u = (u_{0}, u_{1}, \ldots, u_{n_{y}+1}) \) and the proposal distribution can be written as

\[
f_{Y}(y') = p(t_{1}^{y}) \cdots p(t_{n_{y}}^{y} - t_{n_{y}-1}^{y} \mid t_{n_{y}-1}^{y}) \, \text{P}(t_{n_{y}+1}^{y} \geq T \mid t_{n_{y}}^{y})
\]

\[
= \prod_{i=0}^{n_{y}-1} \frac{p(u_{i})du_{i}}{d(t_{i+1}^{y} - t_{i}^{y})} \, \text{P} \left( u_{n_{y}} \geq \Lambda_{n_{y}}(T - t_{n_{y}}^{y}) \right) = p(u) \frac{\partial u}{\partial y'}.
\]

This way of sampling a time-inhomogeneous birth process can also be used to adjust the step size of the proposal. The step size can be adjusted by only updating a randomly selected subset of \( u \) at each iteration. The only problem is that the deterministic relationship between \( u \) and \( Y \) will be changed at each iteration since \( (\beta, \gamma) \) and \( X \) were also updated, so we cannot directly use the \( u \) samples from the previous iteration. Instead, we need to recalculate \( u \) at each iteration. Here we detail the algorithm:

1. Compute \( u \) according to the current state of \( (\beta, \gamma, X, Y) \), by the illustrated deterministic relationship between \( Y \) and \( u \).

2. Randomly select a subset of \( u \). First, round \( c(n_{y} + 1) \) into an integer \( m \), where \( c \) represent the proportion of \( u \) to be updated. Then, pick the index of the elements to be updated by
uniformly sampling \( m \) integers from \((0, 1, 2, \ldots, n_y)\). Denote the subset to be updated by \( u_* \) and denote its compliment by \( u_{-*} \), so \( u = (u_*, u_{-*}) \).

3. Update \( u_* \) by sampling \( u_*' \overset{iid}{\sim} \exp(1) \). The rest part of \( u \) will stay the same, so set \( u'_{-*} = u_{-*} \), \( u' = (u'_*, u'_{-*}) \).

4. Compute \( Y' \) using \( u' \) and sample more \( u_*' \overset{iid}{\sim} \exp(1) \) as needed when \( n'_y > n_y \).

In this way, the proposal distribution will become \( q(y') = p(u'_*) \partial u' / \partial Y' \), and the acceptance probability will stay as \( \alpha(y, y') = \min \{ 1, f_X(x \mid y') / f_X(x \mid y) \} \), since

\[
\frac{f_X(x \mid y') f_Y(y' \mid x) q(y)}{f_X(x \mid y) f_Y(y \mid x) q(y')} = \frac{f_X(x \mid y') p(u'_*) p(u_{-*}) \partial u' / \partial Y'}{f_X(x \mid y) p(u_*) p(u_{-*}) \partial u' / \partial Y'} = \frac{f_X(x \mid y')}{f_X(x \mid y)}
\]

With \( u \) only partially updated, the step size will be smaller, so that the general acceptance rate can be improved. In this way, the acceptance rate can be adjusted to an ideal level by changing the size of \( u_* \).

4 Mathematical Theory

The aim in this section is to put down the details on the high acceptance probability for the proposal distribution.

The acceptance probability of the Metropolis–Hastings procedure can be understood as a distance measure between the target distribution in (4) and proposal in (5):

\[
d(p(y \mid x), f_Y(y \mid x)) = 1 - \int \int \min \{ p(y' \mid x) f_Y(y \mid x), p(y \mid x) f_Y(y' \mid x) \} dydy'
\]

\[
= 1 - \int \int \min \left\{ \frac{p(y' \mid x) f_Y(y \mid x)}{p(y \mid x) f_Y(y' \mid x)}, 1 \right\} p(y \mid x) f_Y(y' \mid x) dydy'
\]

\[
= 1 - \mathbb{E}[\alpha(Y, Y')]
\]

This can be considered as an \( L_1 \) distance, which can be upper bounded by the Hellinger distance (Kraft 1955),

\[
d(p(y \mid x), f_Y(y \mid x)) = \frac{1}{2} \int \int |p(y' \mid x) f_Y(y \mid x) - p(y \mid x) f_Y(y' \mid x)| dydy'
\]

\[
= \frac{1}{2} \| p(y' \mid x) f_Y(y \mid x), p(y \mid x) f_Y(y' \mid x) \|_{L_1}
\]

\[
\leq \sqrt{2} \ H(p(y' \mid x) f_Y(y \mid x), p(y \mid x) f_Y(y' \mid x)).
\]
The Hellinger distance can be computed as

\[
H^2(p(y' \mid x)f_Y(y \mid x), \ p(y \mid x)f_Y(y' \mid x))
\]

\[
= 1 - \int \int \frac{\sqrt{p(y' \mid x)f_Y(y \mid x)p(y \mid x)f_Y(y' \mid x)}}{dy \ dy'}
\]

\[
= 1 - \left\{ \int f_Y(y \mid x)\sqrt{f_X(x \mid y)dy} \right\}^2 /p(x) = 1 - \tilde{p}(x)^2/p(x),
\]  

(8)

where \( p(x) = \int f_Y(y \mid x)f_X(x \mid y)dy \) and \( \tilde{p}(x) = \int f_Y(y \mid x)\sqrt{f_X(x \mid y)}dy \). To illustrate clearly why \( \tilde{p}(x)^2/p(x) \) is close to 1, we look at a simplified case. Consider a Poisson process \( \{Y_t\}_{0 \leq t \leq T} \) with intensity \( \gamma \), and a time–inhomogeneous Poisson process \( \{X_t\}_{0 \leq t \leq T} \) that depends on \( Y \), with intensity function \( \beta(t) = \beta_0 + \epsilon Y_t \), so the infinitesimal probabilities of \( (X, Y) \) are given by

\[
P(X_{t+h} = X_t + 1 \mid X_t, Y_t) = \beta(t)h + o(h), \quad \text{and} \quad P(Y_{t+h} = Y_t + 1 \mid X_t, Y_t) = \gamma h + o(h).
\]

In a small time interval \( h \), \((X_{t+h}, Y_{t+h})\) can be considered conditionally independent given the current state \((X_t, Y_t)\), i.e.

\[
P(x(t+h), \ y(t+h) \mid x(t), \ y(t)) = P(x(t+h) \mid x(t), \ y(t)) \ P(y(t+h) \mid x(t), \ y(t)).
\]

The discrete–time analogue of a conditionally independent bivariate Markov chain \( \{X_i, Y_i\}_{i=1}^n \) would have

\[
P(x(n+1), \ y(n+1) \mid x(n), \ y(n)) = P(x(n+1) \mid x(n), \ y(n)) \ P(y(n+1) \mid x(n), \ y(n))
\]

It is then clear that the joint distribution of \( \{X_i, Y_i\}_{i=1}^n \) can be expressed as

\[
P(x, \ y) = \prod_{k=1}^n P(x(k), y(k) \mid x(k-1), y(k-1))
\]

\[
= \prod_{k=1}^n P(x(k) \mid x(k-1), y(k-1))P(y(k) \mid x(k-1), y(k-1)) = P_X(x \mid y) \ P_Y(y \mid x),
\]

where \( P_X(x \mid y) = \prod_{k=1}^n P(x(k) \mid x(k-1), y(k-1)) \) itself can be considered as the distribution of a univariate time–inhomogeneous Markov process when \( y \) is fixed; similarly for \( P_Y(y \mid x) \).

Therefore, we can discretize the process by dividing the time interval \([0, \ T] \) into \( n \) small time intervals of length \( h = T/n \), and consider \((X_{t+h} - X_t, \ Y_{t+h} - Y_t)\) as two Bernoulli random
variables that are independent to each other conditional on \((X_t, Y_t)\). This discretization enables us to perform the integrals involving \(p(x)\) and \(\tilde{p}(x)\) by summing over all the possible values of \(\{Y_h, Y_{2h}, \ldots, Y_{nh}\}\), and then let \(n \to \infty\). When \(X\) has no jumps between time 0 and \(T\), we are able to compute the exact expression of the Hellinger distance:

\[
H^2((p(y' | x)f_Y(y | x), p(y | x)f_Y(y' | x)) = 1 - \exp\left[-\gamma \left\{T - e^{-\epsilon T} - 4e^{-\frac{1}{2}\epsilon T} + 3\right\}\right], \quad (9)
\]

which goes to 0 as \(\epsilon \to 0\). See the Appendix A of the Supplementary Materials for the mathematical details.

Returning to the likelihood factorization in (7), when we sample \(Y\) from \(f_Y(y | x)\), we are only using the information in \(f_Y(y | x)\), and ignore the information in \(f_X(x|y)\). However, this ignored information is negligible when the value of \(\epsilon\) is small and recall that \(\epsilon\) is the change in the intensity of the \(X\) process when \(Y\) changes by a single unit. Hence, our proposal is close to the target when the intensity of \(X\) remains almost constant over a change of a single unit in \(Y\).

## 5 Illustrations

### 5.1 A First Illustration using Simulated Data

We start with a simple illustration where we assume that the recovery process \(Y\) is completely observed and the parameters \(\beta, \gamma\) are fixed and known, so only the sampling of the the unobserved infection process \(X\) is required. We evaluated and compared the performance of the current and new MCMC algorithm by looking at the effective sample size of the random variable \(n_x\), i.e. the total number of infections.

The population size was taken to be \(N = 1,000,000\), with \(I_0 = 100\) initial infections and \(R_0 = 0\) initial recoveries. The epidemic process was simulated with parameters \(\beta = 0.2\) and \(\gamma = 0.2\) between time 0 and \(T = 10\); this specifically yielded \(n_x = 195\) infections and \(n_y = 194\) recoveries. In both MCMC algorithms, i.e. ours and O’Neill and Roberts (1999), the infection process \(X\) was initialized by uniformly sampling \(n_x\) from 100 to 300, with sampling each \(t^*_i \sim U(0, 10)\).

For our MCMC algorithm, it took less than 5 seconds to get 1000 samples from the posterior, with an acceptance rate of 0.245, and an effective sample size of 119. This was computed using
the R package “mcmcse” by Flegal et al. (2021). A trace plot of the output, demonstrating a very quick convergence is presented in Figure 1. The auto-correlation was low, with the samples almost independent at lag 10; see Figure 1.

On the other hand, for the O’Neill and Roberts (1999) MCMC algorithm, with which we are making a comparison, it took more than 1 minute to get 10,000 samples, with an acceptance rate of 0.936. However, the effective sample size was only 8 and convergence was slow, around 2000 iterations, see Figure 1. Although the acceptance rate was high, a consequence of the small proposals, it is not surprising that the auto-correlation between iterations was also high because of the small step size at each iteration; see Figure 1. See the Supplementary Material for the relevant R code.

5.2 A Second Illustration using Simulated Data

In the second illustration, we work on the opposite problem by assuming that the recovery process $Y$ is completely unobserved and the infection process $X$ is completely observed, with the initial
states \((S_0, I_0, R_0)\) known. At the same time, we assume that the parameters \((\beta, \gamma)\) are unknown.

The population size was taken as \(N = 1,000,000\), with \(I_0 = 100\) initial infections and \(R_0 = 0\) initial recoveries. The epidemic process was simulated with parameters \(\beta = 0.25\) and \(\gamma = 0.15\) between time 0 and \(T = 10\), yielding \(n_x = 353\) infections and \(n_y = 231\) recoveries. We used the priors with parameters \(a_\beta = b_\beta = a_\gamma = b_\gamma = 0.1\) and initialized the unobserved recovery process \(Y\) by uniformly sampling \(n_y\) from 200 to 450 and sampling each \(t_i^Y \sim U(0, 10)\).

For our MCMC algorithm it took around 30 seconds to run 3000 iterations. It took about 300 iterations to converge and after a burn in at the 300th iteration, we obtained the posterior sample means \(\bar{\beta} = 0.238\) and \(\bar{\gamma} = 0.165\). For the corresponding trace plots, see Figures 2.

In comparison, when using the O’Neill and Roberts (1999) MCMC algorithm, it took around 450 seconds to run 30,000 iterations. It took about 8000 iterations to converge and after a burn in at the 8000th iteration, we obtained the posterior sample means as \(\bar{\beta} = 0.237\) and \(\bar{\gamma} = 0.165\). For the trace plots, see Figures 2. We can see the convergence speed of the O’Neill and Roberts (1999) MCMC algorithm is much slower than our MCMC algorithm. See the Supplementary Material for R code.

### 5.3 SIR Model and Smallpox Data

Our illustration with real data for the SIR model comes from the smallpox cases reported in Abakaliki, Nigeria, in 1967; see Thompson and Foege (1968). The dataset is given by the removal times of 32 smallpox cases in a closed community with 120 individuals. O’Neill and Roberts (1999) have illustrated their algorithm with the same dataset, assuming the total number of infections was known. We were able to replicate their results with our MCMC algorithm. However, to better illustrate our method, we assume that the infection process was completely unobserved with unknown total number of infections.

Different from the previous two illustrations with the simulated datasets, for the real smallpox dataset, we need to sample an extra initial infection time \(t_0^x\), which is negative when the first removal time is indexed at 0. With prior \(\pi_0(t_0^x) = \theta_0 \exp(\theta_0 t_0^x)1(t_0^x < 0)\), the full conditional posterior of \(t_0^x\) is given by \(p(t_0^x | -) = \theta e^{\theta(t_0^x - t_1^x)}1(t_0^x < t_1^x),\quad \theta = \theta_0 + \gamma + \beta S_0/N\). We took the prior hyper-parameters \(a_\beta = 10, b_\beta = 100, a_\gamma = 10, b_\gamma = 100\) and \(\theta_0 = 0.1\). We used the method
elaborated in Section 3 to adjust the step size of our MCMC algorithm by only updating a subset of \( u \) at each iteration to reach a moderate acceptance rate of 0.15. It only took 7 seconds to run 5000 iterations with the new MCMC algorithm. The posterior sample means and variances of the parameters are given by \( \bar{\beta} = 0.105 \), \( \text{var}(\beta) = 0.0003 \), with \( \bar{\gamma} = 0.078 \) and \( \text{var}(\gamma) = 0.0002 \). For associated trace plots and histograms, see Figures 3 and Figures 4.

In comparison with the O’Neill and Roberts (1999) MCMC, to achieve a similar effective sample size, the O’Neill and Roberts (1999) algorithm needed to run for 20,000 iterations, taking around 30 seconds, with an acceptance rate of 0.59. The posterior sample means and variances of the parameters are given by \( \bar{\beta} = 0.102 \), \( \text{var}(\beta) = 0.0004 \), and \( \bar{\gamma} = 0.078 \) and \( \text{var}(\gamma) = 0.0004 \). For corresponding trace plots and histograms, see Figures 3 and Figures 4. See the Supplementary Material for R code.
Figure 3: Trace plots for the smallpox data illustration

Figure 4: Histograms for the smallpox data illustration
5.4 SEIR Model and COVID-19 Data

The SEIR model divides the population into four compartments; $S =$ susceptible, $E =$ exposed, $I =$ infected and $R =$ recovered, with three types of transition; from susceptible to exposed, denoted by $X$, from exposed to infectious, denoted by $Y$ and from infectious to recovery, denoted by $Z$. The exposed compartment represents the incubation period of the disease and will eventually become infectious. Similar to the SIR model, it is assumed not possible to return from recovered to either susceptible or infected, and the recovery process actually includes both recoveries and deaths. Except for the infection rate $\beta$ and recovery rate $\gamma$, another parameter is introduced in the SEIR model, which is the transition rate from exposed to infectious, denoted by $\alpha$. The likelihood of the process between time 0 and $T$ is given by

$$
p(x, y, z \mid \beta, \alpha, \gamma) = \exp \left\{ - \int_0^T \beta \frac{S(t)I(t)}{N} dt \right\} \prod_{i=1}^{n_x} \beta \frac{S(t_i^x)I(t_i^x)}{N} \exp \left\{ - \int_0^T \alpha E(t) dt \right\} \prod_{i=1}^{n_y} \alpha E(t_i^y) \exp \left\{ - \int_0^T \gamma I(t) dt \right\} \prod_{i=1}^{n_z} \gamma I(t_i^z).
$$

As elaborated in Section 3, a likelihood factorization is given by

$$
p(x, y, z) = f_{X \mid Y, Z}(x \mid y, z) f_{Y \mid X, Z}(y \mid x, z) f_{Z \mid X, Y}(z \mid x, y),
$$

where the three terms are conditional distributions but not those from the joint $p(x, y, z)$, though are close to them. The second real data illustration comes from the COVID-19 cases reported in French overseas department Mayotte in March and April, 2020. The dataset is well documented in Manou-Abi and Balicchi (2020), given by the daily reports of confirmed and removed cases from March 13 to April 17 2020 in Mayotte, so the $Y$ and $Z$ processes were discretely observed, while the $X$ process is completely unobserved. We set March 13 as $t_0$ and April 17 as $T$, and used the same initial states as Manou-Abi and Balicchi (2020) with $I_0 = 3$ and $E_0 = 15$. A control measure was introduced in March 29 2020, so we considered day 16 as a change point for the infection rate $\beta$ and modeled two different values of the parameter $\beta = (\beta_1, \beta_2)$, for the two different stages of the epidemic process, i.e. before and after March 29. Manou-Abi and Balicchi (2020) used a deterministic SEIR model and estimated the infection rate $\beta$ with the other two transition rates fixed. We used a stochastic SEIR model and estimated all three transition rates. According to the prior information given in Manou-Abi and Balicchi (2020), the latent period
is around 6 days and the infectious period is around 10 days in average, so we took informative priors for $\alpha$ and $\gamma$ with $a_\alpha = 100$, $b_\alpha = 600$, $a_\alpha = 50$, $b_\alpha = 500$, and weak priors for $\beta = (\beta_1, \beta_2)$ with $a_\beta = 5$, $b_\beta = 50$.

For our MCMC algorithm it took 3000 iterations to get a well mixed posterior sample, which converges within 200 iterations. The posterior sample means of the parameters are given by $\bar{\beta}_1 = 0.402$, $\bar{\beta}_2 = 0.058$, $\bar{\alpha} = 0.152$ and $\bar{\gamma} = 0.054$. In comparison, the O’Neill and Roberts (1999) algorithm needed to run for 20,000 iterations to get a well mixed posterior sample and it took at least 2000 iterations to converge. The posterior sample means of the parameters are given by $\bar{\beta}_1 = 0.388$, $\bar{\beta}_2 = 0.064$, $\bar{\alpha} = 0.157$ and $\bar{\gamma} = 0.054$. For the associated trace plots, see Figures 5 and Figures 6.

The results were verified by the R0 package (Obadia et al., 2012) with the incidences data and the distribution of the generation time as inputs. We used 11 days as the mean generation time which is computed from the mean latent period plus a half of the mean infectious period (Svensson, 2007) with the prior information given in Manou-Abi and Balicchi (2020). By using the R0 package, we got the Basic reproduction number $r_0 = 7.03$ with the exponential growth method, and $r_0 = 8.64$ with the maximum likelihood method before the change point, compared to $r_0 = \beta_1/\gamma = 7.44$ with our posterior means. After the change point, we got $r_0 = 0.79$ with the exponential growth method, and $r_0 = 1.97$ with the maximum likelihood method, compared to $r_0 = \beta_2/\gamma = 1.07$ with our posterior means. For all three methods, we can see a big drop of the basic reproduction number through the change point, but the $r_0$ computed by R0 package before the change point is closer to our results than that after the change point, which can be explained by the bias caused by the missing information of the initial value of $I$, since the incidences data alone cannot imply the current number of infectious individuals at any time point. See the Supplementary Material for R code.

6 Conclusions

In this paper we have developed and demonstrated a new Bayesian MCMC algorithm estimating compartmental models. In particular, we have provided a strategy for sampling a completely unobserved infection or recovery process in the general SIR/SEIR model. It is important to note
Figure 5: Trace plots of $n_x$ and $\alpha$ for the COVID-19 data illustration

Figure 6: Trace plots of $\beta_1$ and $\beta_2$ for the COVID-19 data illustration
that our method can easily be extended to any kind of stochastic compartmental model.

Current samplers for the Bayesian MCMC framework are highly problematic. A Metropolis–Hastings algorithm is more efficient when the proposal distribution is close to the target distribution and we have found a proposal that approximates the target distribution very well using a time-inhomogeneous birth process. Compared to current MCMC algorithms, which use a reversible jump procedure, using the time-inhomogeneous birth process as a proposal is advantageous because it does not depend on the sample from the previous iteration. Therefore, the auto-correlation between iterations will be low and the convergence speed will be fast. This class of proposal distribution now makes Bayesian MCMC analysis of stochastic compartmental models completely viable.

References

Bailey, N. T. J. (1975). *Mathematical Theory of Infectious Diseases and Its Application*. Griffin.

Ball, F. and Donnelly, P. (1995). Strong approximation for epidemic models. *Stochastic Processes and their Applications*, 55:1–21.

Becker, N. G. (1993). Parametric inference for epidemic models. *Mathematical Biosciences*, 117:239–251.

Becker, N. G. and Hasofer, A. M. (1997). Estimation in epidemics with incomplete observations. *Journal of the Royal Statistical Society*, 59(2):415–429.

Blum, M. G. B. and Tran, V. C. (2010). HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics*, 11(4):644–660.

Brauer, F. (2008). Compartmental models in epidemiology. *Mathematical Epidemiology*, 1945:19–79.

Cauchemez, S. and Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. *Journal of the Royal Society Interface*, 5:885–897.
Crawford, F. W., Ho, L. S. T., and Suchard, M. A. (2018). Computational methods for birth-death process. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10:1423.

Crawford, F. W. and Suchard, M. A. (2012). Transition probabilities for general birth-death process with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology*, 65:553–580.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., and Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*.

Dukic, V., Lopes, H. F., and Polson, N. G. (2012). Tracking epidemics with Google flu trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107(500):1410–1426.

Fintzi, J., Wakefield, J., and Minin, V. N. (2022). A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *To appear in Biometrics*.

Flegal, J. M., Hughes, J., Vats, D., Dai, N., Gupta, K., and Maji, U. (2021). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, and Kanpur, India. R package version 1.5-0.

Gibson, G. J. and Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *IMA Journal of Mathematics Applied in Medicine & Biology*, 15:19–40.

Gibson, G. J. and Renshaw, E. (2001). Likelihood estimation for stochastic compartmental models using markov chain methods. *Statistics and Computing*, 11:347–358.

Ho, L. S. T., Crawford, F. W., and Suchard, M. A. (2018). Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. *The Annals of Applied Statistics*, 12(3):1993–2021.

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(4):465–496.
Keeling, M. J. and Ross, J. V. (2008). On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface*, 5:171–181.

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov process via the R package “Pomp”. *Journal of Statistical Software*, 69:1–43.

Kingman, J. (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press.

Kraft, C. H. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *University of California Publications in Statistics, Vol. 2*.

Lekone, P. E. and Finkenstädt, B. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62:1170–1177.

Manou-Abi, S. M. and Balicchi, J. (2020). Analysis of the COVID-19 epidemic in french overseas department Mayotte based on a modified deterministic and stochastic SEIR model. *medRxiv*.

McKinley, T., Cook, A. R., and Deardon, R. (2009). Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1):24.

Neal, P. (2012). Efficient likelihood-free Bayesian computation for household epidemics. *Statistics and Computing*, 22:1239–1256.

Neal, P. and Xiang, F. (2017). Collapsing of non-centred parameterized MCMC algorithms with applications to epidemic models. *Scandinavian Journal of Statistics*, 44:81–96.

Obadia, T., Haneef, R., and Böelle, P.-Y. (2012). The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Medical Informatics and Decision Making*, 12:147–147.

O’Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society*, 162(1):121–129.

Pooley, C. M., Marion, G., Bishop, S. C., Bailey, R. I., and Doeschl-Wilson, A. B. (2020). Estimating individuals’ genetic and non-genetic effects underlying infectious disease transmission from temporal epidemic data. *PLoS Computational Biology*, 16:1008447.
Equation (9) in the main paper is an exact expression of the Hellinger distance for a simplified case. Recall that the Hellinger distance can be used as an upper bound of the $L_1$ distance between the target and proposal densities. A detailed proof of equation (9) is included in this section.

6.1 Overview

To compute the Hellinger distance given by equation (8) in the main paper, we need to do two integrals, namely $p(x) = \int f_Y(y|x)f_X(x|y)dy$ and $\tilde{p}(x) = \int f_Y(y|x)\sqrt{f_X(x|y)}dy$. In this appendix we will focus on solving the integral $p(x)$ by computing the expectation of $f_X(x | Y)$ with respect to $Y \sim f_Y(y | x)$ and the procedure will be similar for $\tilde{p}(x)$. Consider the bivariate,
continuous-time point process \( \{X_t, Y_t\}_{0 \leq t \leq T} \) with infinitesimal probabilities

\[
P(X_{t+h} = X_t + 1 \mid X_t, Y_t) = \beta(X_t, Y_t)h + o(h)
\]

\[
P(Y_{t+h} = Y_t + 1 \mid X_t, Y_t) = \gamma(X_t, Y_t)h + o(h),
\]

where \( \beta(X_t, Y_t) \) and \( \gamma(X_t, Y_t) \) are the intensity functions, which only depend on the current state of the process. As defined in equation (1) in the main paper, the right continuous sample paths \((x, y)\) can be characterized by the number and location of jumps, denoted by \((n_x, n_y)\), \(\{t^x_i\}_{i=1}^{n_x}\) and \(\{t^y_i\}_{i=1}^{n_y}\). As explained in Section 3 of the main paper, the likelihood of \((X, Y)\) can be factorized as the product of likelihoods for two time-inhomogeneous birth processes, i.e.

\[
p(x, y) = f_X(x \mid y)f_Y(y \mid x),
\]

\[
f_X(x \mid y) = \prod_{i=1}^{n_x} \beta(x(t^x_i - ), y(t^x_i - )) \exp \left\{ - \int_0^T \beta(x(t), y(t))dt \right\}
\]

\[
f_Y(y \mid x) = \prod_{i=1}^{n_y} \gamma(x(t^y_i - ), y(t^y_i - )) \exp \left\{ - \int_0^T \gamma(x(t), y(t))dt \right\}.
\]

In the following part of the Appendix, we will consider \(Y\) as the time-inhomogeneous birth process depend on \(x\), with intensity function \(\gamma(x(t), Y_t)\) and likelihood \(f_Y(y \mid x)\) and consider \(x\) as a fixed sample path. Therefore, back to the Hellinger distance, the two integrals can be expressed as the expectations with respect to \(Y\),

\[
p(x) = E_Y[f_X(x \mid Y)] \quad \text{and} \quad \tilde{p}(x) = E_Y\left[\sqrt{f_X(x \mid Y)}\right].
\]

It is hard to find the above expectations directly since \(Y\) is a continuous-time process. Therefore, we try to do a discretization for both \(x\) and \(Y\), and compute the expectation \(E_{Y^n}[f_X(x_n \mid Y^n)]\) with a discrete-time count path \(x_n\) and discrete-time count process \(Y^n\). Then take the limit of the expectation as \(n \to \infty\) to recover the desired expectations. The convergence of expectations can be shown by the convergence of \(x_n \to x\) in Skorokhod topology and the convergence of \(Y^n \to Y\) in distribution. In the following section, we will illustrate the convergence of expectations with the function \(f_X(x \mid Y)\) and the proof is the same for \(\sqrt{f_X(x \mid Y)}\).
6.2 Discretization of the Process

Without loss of generality, instead of using the time interval \([0, T]\), we will consider \(D[0, 1]\) (\(\mathbb{R}\)\), which is the space of real functions on \([0, 1]\) that are right-continuous with left-hand limits. Furthermore, let \(D_c\) be the set of count paths in \(D\).

6.2.1 Weak Convergence

Suppose \(Y\) is a time-inhomogeneous birth process in \(D_c\), with intensity function \(\gamma(x, Y)\), where \(x \in D_c\) is a fixed count path in \(D_c\). Let \(Y^n\) be a point process that can only jump at discrete time points \(t \in T_n = \{0, 1/n, 2/n, \ldots, 1\}\), with probability \(\gamma(x(t), Y^n_t)\) \(\frac{1}{n}\), so we have \(P(Y^n_{t+1/n} = Y^n_t + 1) = \gamma(x(t), Y^n_t)\) \(\frac{1}{n}\), \(\forall \ t \in T_n, \ t \neq 1\). Here we show that \(Y^n \overset{d}{\to} Y\) by considering the convergence of the finite dimensional distributions; see \(\text{Kingman, 1992}\).

Firstly, we will show that all the finite dimensional distributions of \(Y_n\) converges weakly to the corresponding finite dimensional distributions of \(Y\). For a given time point \(t \in [0, 1]\), let \(S_t\) and \(S^n_t\) denote the waiting time until the next jump for \(Y\) and \(Y^n\), i.e. \(S_t = \inf\{s \in [0, 1-t] : Y_{t+s} = Y_{t} + 1\}\) and \(S^n_t = \inf\{s \in [0, 1-t] : Y^n_{t+s} = Y^n_{t} + 1\}\). It’s sufficient to show, for any \(t \in [0, 1]\), given \(Y_t = Y^n_t = y(t)\), that \(S^n_t\) converges in distribution to \(S_t\), since both \(Y\) and \(Y^n\) are Markov processes.

The cumulative density functions are denoted by, \(F_t(s) = P(S_t \leq s \mid Y_t = y(t))\), and \(F^n_t(s) = P(S^n_t \leq s \mid Y^n_t = y(t))\), so we want to show that \(F^n_t(s) \to F_t(s)\) for any \(t \in [0, 1]\) and \(s \in [0, 1-t]\). For \(Y\), the cumulative density function of waiting time is similar to the in-homogeneous Poisson process (\(\text{Kingman, 1992}\)),

\[
F_t(s) = 1 - \exp\left\{- \int_t^{t+s} \gamma(x(u), y(t)) \, du\right\}.
\]

For \(Y^n\), we can consider the probability that no jumps happened in \([t, t+s]\),

\[
F^n_t(s) = 1 - \prod_{u \in T_n \cap [t, t+s]} \left\{1 - \gamma(x(u), y(t)) \frac{1}{n}\right\}.
\]

As \(n \to \infty\), using product integrals (\(\text{Kingman, 1992}\)), the limit is given by \(\lim_{n \to \infty} F^n_t(s) = F_t(s)\). Therefore, conditional on the same value of current state at time \(t\), the waiting time until the next jump of \(Y^n\) converges in distribution to that of \(Y\), i.e. \(S^n_t \overset{d}{\to} S_t\) given \(Y^n_t = Y_t\), which completes the
proof. From Theorem 12.6, the convergence of finite dimensional distribution in $D_c$ implies weak convergence, so we have $Y^n \overset{d}{\to} Y$.

Now, consider $f_X(x \mid y)$ in equation (10) as a function of $y$ that maps the count path $y$ from $D_c$ to a real number. With any fixed path $x \in D_c$, this function is bounded since $n_x$ is finite and the intensity function $\beta(x, y)$ is positive and bounded; continuous almost everywhere with $\mu_Y(\{y : f_X \text{ is discontinuous at } y\}) = 0$ for any $x \in D_c$ since the discontinuities happen only when $Y$ jumps at $t \in \{t_i^n\}_{i=1}^{n_x}$.

Now using the continuous mapping theorem (see Theorem 2.7 in ?), with $Y^n \overset{d}{\to} Y$, for the bounded and continuous a.e. function $f_X$, we have the convergence of expectations, $E_{Y^n}[f_X(x \mid Y^n)] \to E_Y[f_X(x \mid Y)]$.

### 6.2.2 Convergence in Skorohod Topology

Suppose $x$ is a count path in $D_c$ and $x_n$ agrees with $x$ at discrete time points in $T_n := \{0, 1/n, 2/n, \ldots, 1\}$, and $x_n$ stays constant in each time interval $[(k-1)/n, k/n)$ for $k = 1, 2, \ldots, n$, so $x_n$ is a right continuous count path in $D_c$, but can only jump at $t \in T_n$. Define the function $\lambda_n : [0, 1] \to T_n$ as a reformation of time such that, $\lambda_n(t) := \max\{s \in T_n : s \leq t\}$, thus $x_n(t) = x(\lambda_n(t)) \forall t \in [0, 1]$, with $|\lambda_n(t) - t| < 1/n$ for any $t \in [0, 1]$, so $\lambda_n(t) \to t$ uniformly and $x_n(t) \to x(t)$ for all the continuity points $t$ of $x$. Note that $x \in D_c$ only has finitely many discontinuities, so it follows $x_n \to x$ in Skorohod Topology (?).

Denote the jump times of $x$ and $x_n$ as $\{t_i^n\}_{i=1}^{n_x}$ and $\{t_i^n\}_{i=1}^{n_x}$. Then we have $t_i^n = \min\{s \in T_n : s \geq t_i^x\}$ for $i = 1, 2, \ldots, n_x$, with $x_n(t_i^{x_n} -) = x(t_i^x -) = i - 1$ if $x_n(0) = x(0) = 0$. Note that two jumps can happen at the same time for $x_n$ when $n$ is small, but this is not a problem for large $n$ and as $n \to \infty$, since $T_n$ will become fine enough. With $|t_i^{x_n} - t_i^x| < 1/n$, we have $t_i^{x_n} \to t_i^x$ and $y(t_i^{x_n} -) \to y(t_i^x -)$ for any $y \in D_c$ that are continuous at $\{t_i^x\}_{i=1}^{n_x}$.

Now, consider the likelihood function $f_X(x_n \mid y)$. We have $x_n(t_i^{x_n} -) = x(t_i^x -)$, $x_n \to x$ in Skorohod topology, and $y(t_i^{x_n} -) \to y(t_i^x -)$ for any $y \in D_c$ that are continuous at $\{t_i^x\}_{i=1}^{n_x}$. For a continuous time count process $Y$, the probability of having a jump at any exact time point is 0, i.e. $\mu_Y(\{y : y \text{ is discontinuous at } \{t_i^x\}_{i=1}^{n_x}\}) = 0$, so we can conclude that the convergence of $x_n \to x$ in Skorohod topology leads to the convergence of expectations given by, $E_Y[f_X(x_n \mid Y)] \to E_Y[f_X(x \mid Y)]$.
6.2.3 Convergence of Expectations

We want to show the convergence of expectations \( E_{\gamma^n}[f_X(x_n \mid Y^n)] \rightarrow E_Y[f_X(x \mid Y)] \). By the triangle inequality,

\[
|E_{\gamma^n}[f_X(x_n \mid Y^n)] - E_Y[f_X(x \mid Y)]| \\
\leq |E_{\gamma^n}[f_X(x_n \mid Y^n)] - E_Y[f_X(x_n \mid Y)]| + |E_Y[f_X(x_n \mid Y)] - E_Y[f_X(x \mid Y)]|.
\]

From Section 6.2.1, we have \( E_{\gamma^n}[f_X(x \mid Y^n)] \rightarrow E_Y[f_X(x \mid Y)] \). Note that, the distribution of \( Y^n \) with \( x_n \) is exactly the same as that with \( x \), since \( x_n(t) = x(t) \) for any \( t \in T_n \). Therefore, \( (Y^n \mid x) \overset{d}{\rightarrow} (Y \mid x) \) is equivalent to \( (Y^n \mid x_n) \overset{d}{\rightarrow} (Y \mid x) \) and it follows \( E_{\gamma^n}[f_X(x_n \mid Y^n)] \rightarrow E_Y[f_X(x_n \mid Y)] \). From Section 6.2.2, we have \( E_Y[f_X(x_n \mid Y)] \rightarrow E_Y[f_X(x \mid Y)] \). Therefore, by putting together section 6.2.1 and 6.2.2, we have \( E_{\gamma^n}[f_X(x_n \mid Y^n)] \rightarrow E_Y[f_X(x \mid Y)] \).

6.3 Computing the Expectations

Here we compute the expectation \( E_{\gamma^n}[f_X(x_n \mid Y^n)] \) by using the discretization in section 6.2 and then take the limit for \( n \rightarrow \infty \) to get the desired expectation \( E_Y[f_X(x \mid Y)] \). For notation simplification, in the following computation, we will write \( x_n(k/n) \) and \( y_n(k/n) \) as \( x(k) \) and \( y(k) \), and let \( h = 1/n \).

Let \( Y^n \) be the discrete-time point process as defined in Section 6.2.1 with probability function given by \( P_{Y^n}(y_n \mid x_n) = \prod_{k=0}^{n-1} P(y(k+1) \mid x(k), y(k)) \), where \( P(y(k+1) \mid x(k), y(k)) = \gamma(x(k), y(k)) \) for \( y(k+1) = y(k) + 1 \), and \( P(y(k+1) \mid x(k), y(k)) = 1 - \gamma(x(k), y(k))h \) for \( y(k+1) = y(k) \). The function \( f_X(x_n \mid y_n) \) can be written in a similar product form,

\[
f_X(x_n \mid y_n) = \prod_{k=0}^{n-1} f(x(k+1) \mid x(k), y(k)),
\]

where \( f(x(k+1) \mid x(k), y(k)) = \beta(x(k), y(k)) \exp\{ -\beta(x(k), y(k))h \} \) for \( x(k+1) = x(k) + 1 \) and \( f(x(k+1) \mid x(k), y(k)) = \exp\{ -\beta(x(k), y(k))h \} \) for \( x(k+1) = x(k) \). Let \( \mathcal{Y}^n \) denote the support...
set of $Y^n$. We can compute $E_{Y^n}[f_X(x_n \mid Y^n)]$ by summing over all $y_n \in Y^n$.

$$E_{Y^n}[f_X(x_n \mid Y^n)] = \sum_{y \in Y^n} \prod_{k=0}^{n-1} f(x(k + 1) \mid x(k), y(k)) \prod_{k=0}^{n-1} P(y(k + 1) \mid x(k), y(k))$$

$$= f(x(1) \mid x(0), y(0)) \sum_{y(1)} f(x(2) \mid x(1), y(1))P(y(1) \mid x(0), y(0)) \ldots$$

$$\sum_{y(n-1)} f(x(n) \mid x(n-1), y(n-1))P(y(n-1) \mid x(n-2), y(n-2)).$$

To further simplify the notations, define

$$w_k := f(x(k + 1) \mid x(k), y(k)) P(y(k) \mid x(k - 1), y(k - 1)).$$

We can first compute $\sum_{y(n-1)} w_{n-1}$, and then $\sum_{y(n-2)} w_{n-2} \sum_{y(n-1)} w_{n-1}$, etc, so that

$$E_{Y^n}[f_X(x_n \mid Y^n)] = f(x(1) \mid x(0), y(0)) \sum_{y(1)} w_1 \ldots \sum_{y(n-1)} w_{n-1},$$

can be computed recursively from the end towards the beginning.

Here we derive $E_{Y^n}[f_X(x_n \mid Y^n)]$, by recursion, for a special case. Assume $x(0) = y(0) = 0$, $\gamma(x, y) = \gamma$ and $\beta(x, y) = \beta_0 + ey$, so $\beta(x, y + 1) - \beta(x, y) = \epsilon$ and we assume $x$ has no jumps in the time interval $[0, 1]$, so $x(k + 1) = x(k)$ for each $k$, and we have

$$w_k = \exp[-\{\beta_0 + \epsilon y(k)\}h] P(y(k) \mid x(k - 1), y(k - 1)).$$

The recursion starts from

$$\sum_{y(n-1)} w_{n-1} = \exp[-\{\beta_0 + \epsilon y(n-2)\}h] \{1 - \gamma(1 - e^{-\epsilon h})h\}.$$ 

The next step is given by,

$$\sum_{y(n-2)} \sum_{y(n-1)} w_{n-2} w_{n-1} = \{1 - \gamma(1 - e^{-\epsilon h})h\} \{1 - \gamma(1 - e^{-2\epsilon h})h\} \exp[-2\{\beta_0 + \epsilon y(n-3)\}h].$$

Keep repeating this, to observe the pattern. At the end of the recursion we will get

$$E_{Y^n}[f_X(x_n \mid Y^n)] = \exp[-n \{\beta_0 + \epsilon y(0)\}h] \prod_{k=1}^{n-1} \{1 - \gamma(1 - e^{-k\epsilon h})h\}.$$ 

Now take the limit as $n \to \infty$ and by product integral,

$$E_Y[f_X(x \mid Y)] = \lim_{n \to \infty} E_{Y^n}[f_X(x_n \mid Y^n)] = \exp \left\{-\beta_0 - \int_0^1 \gamma \left(1 - e^{-\alpha t}\right)dt \right\}.$$ 

29
Clearly, $E_Y \left[ \sqrt{f_X(x \mid Y)} \right]$ can be derived in a similar way, simply by replacing all the $\exp \left[ - \{\beta_0 + \epsilon y(k)\} \right]$ with $\exp \left[ -\frac{1}{2} \{\beta_0 + \epsilon y(k)\} h \right]$, so

$$E_Y \left[ \sqrt{f_X(x \mid Y)} \right] = \exp \left\{ -\frac{1}{2} \beta_0 - \int_0^1 \gamma \left( 1 - e^{-\frac{1}{2} \epsilon t} \right) dt \right\}.$$ 

Finally, we change the time interval $[0, 1]$ back to $[0, T]$ and get equation (9), i.e.

$$H^2 (p(y' \mid x) f_Y(y \mid x), p(y \mid x) f_Y(y' \mid x)) = 1 - \exp \left[ -\gamma \left\{ T - \epsilon^{-1} \left( e^{-\epsilon T} - 4e^{-\frac{1}{2} \epsilon T} + 3 \right) \right\} \right].$$