Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Vladimir Minin, Chair

Jon Wakefield, Chair

M. Elizabeth Halloran

James Hughes

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

Co-Chairs of the Supervisory Committee:
Co-chair Vladimir Minin

Co-chair Jon Wakefield

An incredible abstract with all the best words will appear here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

GLOSSARY

AFSS: Automated factor slice sampler.

CLE: Chemical Langevin equation.

CP: Centered parameterization.

CTMC: Continuous–time Markov chain.

DA: Data augmentation.

ELIPTSS: Elliptical slice sampling.

ESS: Effective sample size.

ILI: Influenza–like illness.

LNA: Linear noise approximation.

MJP: Markov jump process.

NCP: Non–centered parameterization.

SDE: Stochastic differential equation.

SEM: Stochastic epidemic model.

ACKNOWLEDGMENTS

Very grateful to many people.

DEDICATION

Dedication to important people.

Chapter 1

INTRODUCTION AND DATA SETTING

1.1 Motivating examples

1.1.1 Influenza in a British boarding school

1.1.2 Ebola in West Africa

1.1.3 Pandemic A(H1N1) influenza in Finland

1.2 Organization of this dissertation

Chapter 2

BACKGROUND

2.1 Models for the Spread of Infectious Disease

2.1.1 Deterministic Representations

2.1.2 Stochastic Representations

Agent–based models

Population–level models

2.1.3 Large–Population Approximations

Diffusion approximations of Markov jump processes

Linear noise approximation

2.2 Computational Approaches to Fitting Stochastic Epidemic Models

2.3 Bayesian Computation

2.3.1 Markov Chain Monte Carlo

Slice sampling for model parameters

Elliptical slice sampling

2.3.2 Bayesian Data Augmentation

Chapter 3

AGENT–BASED DATA AUGMENTATION FOR FITTING STOCHASTIC EPIDEMIC MODELS TO PREVALENCE DATA

Chapter 4

APPROXIMATE INFERENCE FOR STOCHASTIC EPIDEMIC MODELS OF OUTBREAKS IN LARGE POPULATIONS

4.1 Overview

Surveillance and outbreak response systems often report incidence counts of new cases detected in each inter–observation time interval. Analyzing this type of time series data is challenging since we must overcome many of the same challenges that we face in modeling the transmission dynamics of infectious diseases in small population settings with prevalence data — discrete snapshots of a continuously evolving epidemic process, detecting a fraction of the new cases, and often directly observing only one aspect of the disease process. Furthermore, our task is made more difficult by the additional computational burden that results from repeated evaluation of CTMC likelihoods; the products of exponential waiting time distributions consist of polynomially increasing numbers of terms, and agent–based data augmentation (DA) MCMC algorithms become unwieldy as the numbers of subject–path proposals required to meaningfully perturb the CTMC likelihood get large [8].

In this chapter, we show how the LNA of Section 2.1.3 can be adapted to obtain approximate inference for SEMs fit to epidemic count data in large populations. Our contributions are threefold: First, we demonstrate how the SEM dynamics should be reparameterized so that the LNA can be used to approximate transition densities of the counting processes for disease state transition events. Second, we fold the LNA into a Bayesian DA framework in which latent LNA paths are sampled using the elliptical slice sampling (EliptSS) algorithm of [18]. This provides us with general machinery for jointly updating the latent paths while absolving us of the *de facto* modeling choice that the data be Gaussian in order to efficiently perform inference as in [7, 17], or the need to use computationally intensive particle filter

methods for non–Gaussian emission distributions as in [13]. Finally, we introduce a non–centered parameterization (NCP) for the LNA that massively improves the efficiency of our DA MCMC framework and makes it tractable for fitting complex models.

4.2 Fitting Stochastic Epidemic Models via the Linear Noise Approximation

For clarity, we will present the algorithm for fitting SEMs via the LNA in the context of fitting the susceptible–infected–recovered (SIR) model to Poisson distributed incidence counts. We will, however, provide notation where appropriate so that the generality of the algorithm should be apparent. The SIR model is an abstraction of the transmission dynamics of an outbreak as a closed, homogeneously mixing population of N exchangeable individuals who are either susceptible (S), infected, and hence infectious, (I), or recovered (R). It is important to note that the model compartments refer to disease states as they relate to the transmission dynamics, not the disease process. Thus, an individual is considered to be recovered when she no longer has infectious contact with other individuals in the population, not when she clears disease carriage. As another example, in the susceptible–exposed–infected–recovered (SEIR) type models that we will consider later, the latent period in which an individual is exposed, but not yet infectious, should be understood as possibly varying in population with different contact dynamics, even when the incubation period of the pathogen should arguably be consistent across groups.

4.2.1 Measurement Process and Data

Incidence data, $\mathbf{Y} = \{Y_1, \ldots, Y_L\}$, arise as increments of the numbers of new cases accumulated in a set of time intervals, $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_L : \mathcal{I}_\ell = (t_{\ell-1}, t_\ell]\}$. In outbreak or surveillance settings, we do not typically believe that every case is detected since individuals may be asymptomatic or may escape detection. Let $\mathbf{N}^c = (N^c_{SI}, N^c_{IR})$ denote the counting process for the cumulative numbers of infections ($S \to I$ transitions) and recoveries ($I \to R$ transitions), and let $\Delta\mathbf{N}^c(t_\ell) = \mathbf{N}^c(t_\ell) - \mathbf{N}^c(t_{\ell-1})$ denote the change in cumulative numbers of transitions over $\mathcal{I}_\ell$; so, $\Delta N^c_{SI}(t_\ell)$ is the incidence over $(t_{\ell-1}, t_\ell]$. We might choose to model

the number of observed cases as a Poisson sample of the true incidence with detection rate ρ. Thus,

$$Y_\ell | \Delta N^c_{SI}(t_\ell), \rho \sim \text{Pois}(\rho \Delta N^c_{SI}(t_\ell)). \tag{4.1}$$

There are two minor points that we wish to make note of before proceeding. First, we have allowed for the possibility that cases are over–reported. This is neither a necessary assumption for any of the subsequent results, nor is it unreasonable when studying outbreaks in large populations where the "fog of war" might lead to inflation of reported incidence or misclassification of individuals whose symptoms are similar to the disease of interest. This modeling choice is also not particularly problematic when the detection probability is low since the emission densities will have negligible mass above the true incidence. Second, we are also making this modeling choice with an eye on the compatibility of the measurement distribution with the eventual LNA approximation, which takes real, not integer, values. The Poisson distribution, along with the negative binomial distribution that we will use in subsequent sections, are well defined for non–integer values of the mean parameter.

4.2.2 Latent Epidemic Process

The SIR model is typically expressed in terms of compartment counts, $\mathbf{X}^c = \{S^c, I^c, R^c\}$, that evolve in continuous time on state space $\mathcal{S}^c_X = \{\mathcal{C}_{lmn} : l, m, n \in \{0, \ldots, N\},\ l + m + n = P\}$. We will make the (not particularly limiting) modeling choice to express the waiting times between disease state transitions as being exponentially distributed. Thus, $\mathbf{X}$ evolves according to a Markov jump process (MJP). If our data had consisted of prevalence counts, which arise as partial observations of infected individuals, we might have chosen to approximate transition densities of the MJP for $\mathbf{X}$ in the usual way that appears in [17, 7].

However, incidence data are discretely observed, partial realizations of the increments of counting processes that evolve continuously in time as individuals transition among disease states. The emission probabilities for incidence data, e.g., (4.1), depend on the change in N^c_{SI} over the time interval $(t_{\ell-1}, t_\ell]$, not on the change in I over the interval. It would be incorrect

to treat incidence as simply the difference in prevalence. We could easily construct a scenario where there are positive numbers of infections, but where the prevalence does not change due to an equal number of recoveries. We need to construct the LNA that approximates transition densities of $\mathbf{N}$ if we are to write down correctly specified emission probabilities.

The cumulative incidence process for infections and recoveries, $\mathbf{N}^c$, is a Markov jump process state space $\mathcal{S}_N^c = \{\mathcal{C}_{jk} : j, k \in \{0, \ldots, N\}\}$. Let β denote the per–contact infection rate, and μ denote the rate at which each infected individual recovers. The rate at which $\mathbf{N}^c$ transitions from state $\mathbf{n}$ to $\mathbf{n}'$ is

$$\boldsymbol{\lambda}_{\mathbf{n},\mathbf{n}'} = \begin{cases} \lambda_{SI} = \beta SI, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI}+1, n_{IR}),\ \text{and } n_{SI}+1 \leq P, \\ \lambda_{IR} = \mu I, & \mathbf{n} = (n_{SI}, n_{IR}),\ \mathbf{n}' = (n_{SI}, n_{IR}+1),\ \text{and } n_{IR}+1 \leq P, \\ 0, & \text{for all other } \mathbf{n} \text{ and } \mathbf{n}'. \end{cases} \tag{4.2}$$

4.2.3 Tractable Approximations for Intractable Likelihoods

We would like to make inferences about the posterior distribution of the parameters, e.g., $\boldsymbol{\theta} = (\beta, \mu, \mathbf{X}(t_0), \rho)$, that govern the latent epidemic process and sampling distribution,

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) &= \int L(\mathbf{Y}|\mathbf{N}^c, \boldsymbol{\theta})\pi(\mathbf{N}^c|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\pi(\mathbf{N}^c) \\ &= \int_{\mathcal{S}^c} \prod_{\ell=1}^{L} \Pr\left(\mathbf{Y}_\ell|\Delta\mathbf{N}_{SI}^c(t_\ell), \boldsymbol{\theta}\right) \pi\left(\mathbf{N}^c(t_\ell)|\mathbf{n}^c(t_{\ell-1}), \boldsymbol{\theta}\right) \pi(\boldsymbol{\theta})\mathrm{d}\pi(\mathbf{N}^c) \end{aligned} \tag{4.3}$$

where $\pi(\boldsymbol{\theta})$ specifies the prior density of the model parameters. However, this integral is analytically intractable and is challenging to compute numerically due to the size of the state space of $\mathbf{N}^c$. In the following subsections, we will obtain the LNA for transition densities of $\mathbf{N}^c$, turning (4.3) into an integral over a much more computationally tractable product of Gaussian densities and non–Gaussian emission probabilities. As we shall see, approximating the complete data likelihood in the posterior $\pi(\boldsymbol{\theta}, \mathbf{N}^c|\boldsymbol{\theta})$ with a Gaussian state space model will open the doors to efficient algorithms for sampling from the approximate posterior.

4.2.4 Diffusion Approximation

As outlined in Section 2.1.3, there are a variety of methods for arriving at a diffusion approximation for a Markov jump process, which under certain conditions yield equivalent results (for a comprehensive reference, see [9]). In the interest of clarity, we follow [7, 12, 13, 26] and appeal to an intuitive, though somewhat informal, construction of the CLE by matching its drift and diffusion with the approximate moments of increments of the MJP path in infinitesimal time intervals. For more detailed presentations see [9, 11, 25].

Suppose that, at the current time, the compartment counts are given by $\mathbf{X}^c(t) = \mathbf{x}_t^c$. We are interested in approximating the numbers of infections and recoveries in a small time interval, $(t, t + \mathrm{d}t]$, i.e., $\mathbf{N}^c(t + \mathrm{d}t) - \mathbf{N}(t)$. Now, suppose that we can choose $\mathrm{d}t$ such that the following two *leap* conditions hold:

1. $\mathrm{d}t$ is sufficiently *small* that the $\mathbf{X}^c$ is essentially unchanged over $(t, t + \mathrm{d}t]$, so that the rates of infections and recoveries are approximately constant:

$$\boldsymbol{\lambda}(\mathbf{X}^c(t')) \approx \boldsymbol{\lambda}(\mathbf{x}^c(t)),\ \forall t' \in (t, t + \mathrm{d}t]. \tag{4.4}$$

2. $\mathrm{d}t$ is sufficiently *large* that we can expect many disease state transitions of each type:

$$\boldsymbol{\lambda}(\mathbf{x}^c(t)) \gg \mathbf{1}. \tag{4.5}$$

Condition (4.4), which can be trivially satisfied just by choosing $\mathrm{d}t$ to be small, implies that the numbers of infections and recoveries in $(t, t + \mathrm{d}t]$ are essentially independent of one another since the rates at which they occur are approximately constant within the interval [11]. This condition also carries the stronger implication that the numbers of infections and recoveries in the interval are independent Poisson random variables with rates $\boldsymbol{\lambda}(\mathbf{x}^c(t)\mathrm{d}t)$, i.e., $N_{SI}^c(\mathrm{d}t) \sim \text{Poisson}(\beta S(t)I(t)\mathrm{d}t)$ and $N_{IR}^c(t+\mathrm{d}t) \sim \text{Poisson}(\mu I(t)\mathrm{d}t)$. Condition (4.5), which we can reasonably expect to be satisfied in large populations where transmission dynamics are near their deterministic ODE limits [25], implies that the Poisson distributed increments can be well approximated by independent Gaussian random variables.

Thus, (4.4) and (4.5) are satisfied, we can approximate the integer–valued processes, $\mathbf{X}^c$ and $\mathbf{N}^c$, with the real–valued processes, $\mathbf{X}$ and $\mathbf{N}$. For the SIR model, the state space of $\mathbf{X}$ is $\mathcal{S}_X^R = \{\mathcal{V}_{lmn} : l, m, n \in [0, N],\ l + m + n = P\}$, and the state space of $\mathbf{N}$ is $\mathcal{S}_N^R = \{\mathcal{V}_{jk} : j, k \in [0, N]\}$. More generally, the state space of $\mathbf{X}$ will be the set of compartment volumes that are non–negative and that sum to the population size, while the state space of $\mathbf{N}$ is the set of non–decreasing and non–negative incidence paths, constrained so that they do not lead to invalid prevalence paths (e.g., if at some point there are more recoveries than infections, which would lead to a negative number of infected individuals). For now, we will ignore the constraints on $\mathcal{S}_N^R$ and $\mathcal{S}_X^R$, and approximate the changes in cumulative incidence of infections and recoveries in an infinitesimal time step as

$$\mathbf{N}(t + \mathrm{d}t) - \mathbf{N}(t) \approx \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}t^{1/2}\mathbf{Z}, \tag{4.6}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\,(\boldsymbol{\lambda}(\mathbf{X}))$ and $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$. This implies the equivalent CLE,

$$\mathrm{d}\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}\mathrm{d}\mathbf{W}_t, \tag{4.7}$$

where $\mathbf{W}_t$ is a vector of independent Brownian motion and $\boldsymbol{\Lambda}(\mathbf{X}(t))^{1/2}$ denotes the matrix square root of $\boldsymbol{\Lambda}(\mathbf{X}(t))$.

Reparameterizating the CLE in terms of incidence

The LNA of (4.7) will involve derivatives of the rates, $\boldsymbol{\lambda}$, with respect to the incidence process, $\mathbf{N}$. In order to enable us to compute these derivatives, we borrow from [4, 14] a reparameterization for $\mathbf{X}(t)$ in terms of $\mathbf{N}(t)$, conditional on the initial conditions $\mathbf{X}(t) = \mathbf{x}_0$ and $\mathbf{N}(t) = \mathbf{0}$. Let $\mathbf{A}$ denote the matrix whose rows specify changes in counts of susceptible, infected, and recovered individuals corresponding to one infection or recovery event:

$$\mathbf{A} = \begin{array}{c} \\ S \to I \\ I \to R \end{array} \begin{array}{c} \begin{array}{ccc} S & I & R \end{array} \\ \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \end{array}. \tag{4.8}$$

Now, $\mathbf{X}$ is coupled to $\mathbf{N}$ via the relationship,

$$\mathbf{X}(t) = \mathbf{x}_0 + \mathbf{A}^T\mathbf{N}(t). \tag{4.9}$$

For the SIR model,

$$\begin{pmatrix} S(t) \\ I(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} S_0 - N_{SI}(t) \\ I_0 + N_{SI}(t) - N_{IR}(t) \\ R_0 + N_{IR}(t) \end{pmatrix}, \tag{4.10}$$

which enables us to rewrite (4.7) as

$$\begin{aligned} \mathrm{d}\mathbf{N}(t) &= \boldsymbol{\lambda}(\mathbf{N}(t))\mathrm{d}t + \boldsymbol{\Lambda}(\mathbf{N}(t))^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.11)\\ &= \begin{pmatrix} \beta(S_0 - N_{SI}(t))(I_0 + N_{SI}(t) - N_{IR}(t)) \\ \mu(I_0 + N_{IR}(t))\mathrm{d}t \end{pmatrix} \mathrm{d}t + \\ &\qquad \begin{pmatrix} \beta(S_0 - N_{SI}(t))(I_0 + N_{SI}(t) - N_{IR}(t)) & 0 \\ 0 & \mu(I_0 + N_{IR}(t)) \end{pmatrix}^{1/2} \mathrm{d}\mathbf{W}_t. \end{aligned}$$

Log transforming the CLE

Changes in compartment volumes affect the rates, and hence increments in the incidence process, multiplicatively. Therefore, from a scientific perspective, we would like for perturbations about the drift in (4.11) to be symmetric on a multiplicative, not an additive scale. Hence, we log transform (4.11). Let $\widetilde{\mathbf{N}} = \log(\mathbf{N} + \mathbf{1}) \implies \mathbf{N} = \exp(\widetilde{\mathbf{N}}) - \mathbf{1}$. By Itô's lemma [20], the corresponding SDE for $\widetilde{\mathbf{N}}$ is

$$\begin{aligned} \mathrm{d}\widetilde{\mathbf{N}}(t) = \mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t)) - 0.5\exp(-2\widetilde{\mathbf{N}}(t))\right)\boldsymbol{\lambda}\left(\exp(\widetilde{\mathbf{N}}(t)) - \mathbf{1}\right)\mathrm{d}t + \\ \mathrm{diag}\left(\exp(-\widetilde{\mathbf{N}}(t))\right)\boldsymbol{\Lambda}\left(\exp(\widetilde{\mathbf{N}}(t)) - 1\right)^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.12)\\ = \boldsymbol{\eta}(\widetilde{\mathbf{N}}(t))\mathrm{d}t + \boldsymbol{\Phi}(\widetilde{\mathbf{N}}(t))^{1/2}\mathrm{d}\mathbf{W}_t \qquad (4.13) \end{aligned}$$

4.2.5 Linear Noise Approximation

In Section 2.1.3, we followed [7, 12, 13] in obtaining the LNA for SDEs of the same form as (4.13). Briefly, the derivation proceeded as follows: we first decomposed $\widetilde{\mathbf{N}}$ into its

deterministic ODE limit and a stochastic residual. The SDE corresponding to (4.12) was then Taylor expanded around its deterministic limit, discarding higher order terms, to obtain a linear SDE for the residual. This linear SDE had an explicit solution as a Gaussian random variable. As noted in [25], the LNA can reasonably approximate the stochastic aspects of a density dependent MJP when conditions (4.4) and (4.5) are satisfied, at least over short time horizons. Over longer time periods the approximation may deteriorate as departures from the deterministic behavior of the system, which is determined by its initial conditions, accumulate. One solution, proposed in [7] and that we will adopt here, is to restart the LNA approximation at the beginning of each inter–observation interval.

The restarting LNA of (4.13) over a time interval, $(t_{\ell-1}, t_\ell]$, was seen to be a Gaussian approximation of the transition density of $\widetilde{\mathbf{N}}$,

$$\widetilde{\mathbf{N}}(t_\ell)|\widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta} \sim MVN\left(\boldsymbol{\mu}(t_\ell) + \mathbf{m}(\widetilde{\mathbf{n}}(t_{\ell-1}) - \boldsymbol{\mu}(t_{\ell-1})), \boldsymbol{\Sigma}(t_\ell)\right), \tag{4.14}$$

where $\boldsymbol{\mu}(\cdot)$, $\mathbf{m}(\cdot)$, and $\boldsymbol{\Sigma}(\cdot)$ are solutions to the coupled, non–autonomous system of ODEs,

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \boldsymbol{\eta}(\boldsymbol{\mu}(t)), \tag{4.15}$$

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathbf{F}(t)\mathbf{m}(t), \tag{4.16}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\mathbf{F}(t)^T + \boldsymbol{\Phi}(t), \tag{4.17}$$

with respect to initial conditions $\mathbf{N}(t_{\ell-1}) = \mathbf{0}$, $\mathbf{X}(t_{\ell-1}) = \mathbf{x}(t_{\ell-1})$, $\mathbf{m}(t_{\ell-1}) = \mathbf{0}$, and $\boldsymbol{\Sigma}(t_{\ell-1}) = \mathbf{0}$, and where $\mathbf{F}(t)$ is the Jacobian $\left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\mu}(t))}{\partial \boldsymbol{\mu}_j(t)}\right)_{i,j\in 1,\ldots,|\widetilde{\mathbf{N}}|}$ evaluated along the solution to (4.15). Note that we need never actually solve (4.16) since $\mathbf{m}(t_{\ell-1}) = \mathbf{0}$ implies that $\mathbf{m}(t_\ell) = \mathbf{0}\ \forall\ l = 0, \ldots, L-1$.

Approximating the transition densities of $\mathbf{N}$ using the LNA, (4.14), enables us to approximate the observed data likelihood in (4.3) with a Gaussian state space model. The

augmented approximate posterior is

$$\begin{aligned}\pi(\widetilde{\mathbf{N}}, \boldsymbol{\theta}|\mathbf{Y}) &\propto L(\mathbf{Y}|\widetilde{\mathbf{N}}, \boldsymbol{\theta}) \mathbb{1}_{\{\mathbf{N} \in \mathcal{S}_N^R\}} \mathbb{1}_{\{\mathbf{X} \in \mathcal{S}_X^R\}} \pi(\widetilde{\mathbf{N}}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\ &= \prod_{\ell=1}^{L} \Pr(\mathbf{Y}_\ell | \Delta\widetilde{\mathbf{N}}(t_\ell), \boldsymbol{\theta}) \mathbb{1}_{\{\mathbf{N}(t_\ell) \in \mathcal{S}_N^R\}} \mathbb{1}_{\{\mathbf{X}(t_\ell) \in \mathcal{S}_X^R\}} \pi(\widetilde{\mathbf{N}}(t_\ell) | \widetilde{\mathbf{n}}(t_{\ell-1}), \mathbf{x}(t_{\ell-1}), \boldsymbol{\theta}) \pi(\boldsymbol{\theta}).\end{aligned} \tag{4.18}$$

Note that the emission probabilities in (4.18) depend on the incidence, not the log–incidence, but that this just requires a simple reparameterization of the emission distribution. In our example, the observed incidence is a Poisson sample of the true incidence, $\Pr(\mathbf{Y}_\ell | \Delta\widetilde{\mathbf{N}}(t_\ell), \boldsymbol{\theta}) = \text{Pois}\left(\rho\left((\exp(\widetilde{N}_{SI}(t_\ell)) - 1) - (\exp(\widetilde{N}_{SI}(t_{\ell-1})) - 1)\right)\right)$. We also explicitly include indicators for whether the LNA path respects the positivity and monotonicity constraints of the original MJP. We do this for two reasons: We wish to more faithfully approximate the MJP. We also wish to avoid numerical instabilities that arise when $\mathbf{N}$ or $\mathbf{X}$ become negative and that can cause routines for numerically integrating the LNA ODEs to fail.

4.2.6 Inference via the Linear Noise Approximation

To this point, we have discussed how to approximate transition densities of a MJP via the LNA. However, this is only half the battle since we must also address the computational aspects of sampling from the augmented approximate posterior, (4.18). A central computation challenge that plagues DA MCMC is that MCMC chains may suffer from severe autocorrelation when the algorithm alternately updates the latent variables given the parameters, and parameters given the latent variables, see e.g., [3, 22, 23, 27]. As we can see in Figure, a DA MCMC algorithm that alternates between updates LNA paths and model parameters is no exception.

Non–centered Parameterization

We can improve the mixing of our MCMC chains by reparameterizing the log–incidence process as a deterministic mapping of standard normal random variables, $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$,

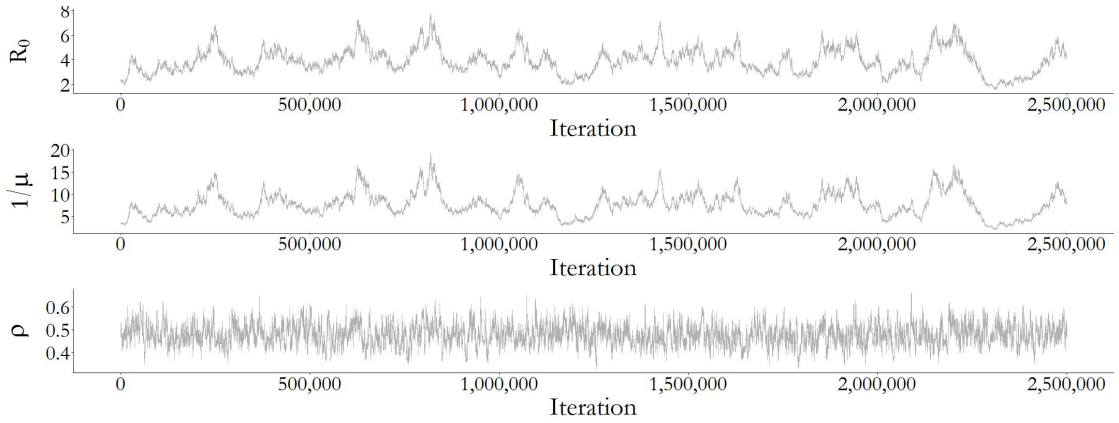

Figure 4.1: Posterior traceplots for a single MCMC chain of an SIR model fit to Poisson distributed incidence data. MCMC targeted the posterior, 4.18, alternately updating the non–restarting LNA for $\widetilde{\mathbf{N}}|\boldsymbol{\theta}, \mathbf{Y}$ via elliptical slice sampling, and $\boldsymbol{\theta}|\widetilde{\mathbf{N}}, \mathbf{Y}$ via a multivariate random walk Metropolis algorithm. $R_0 = \beta N/\mu$ is the basic reproductive number, $1/\mu$ is the mean infectious period duration, and ρ is the mean case detection rate. The true values of R_0, $1/\mu$, and ρ were 3.5, 7, and 0.5, respectively.

which are *a priori* independent of the model parameters. This NCP is carried out by noting that if $\widetilde{\mathbf{N}}(t_\ell) \sim MVN(\boldsymbol{\mu}(t\ell), \boldsymbol{\Sigma}(t_\ell)$ and $\mathbf{Z}(t_\ell) \sim MVN(\mathbf{0}, \mathbf{I})$, then $\widetilde{\mathbf{N}} \stackrel{\mathcal{L}}{=} \widetilde{\mathbf{W}}(t_\ell)$, $\widetilde{\mathbf{W}}(t_\ell) = \boldsymbol{\mu}(t_\ell) + \boldsymbol{\Sigma}(t_\ell)^{1/2}\mathbf{Z}(t_\ell)$. We now target the joint posterior of the model parameters and the non–centered LNA draws,

$$\pi(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \propto L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}))\mathbb{1}_{\{\mathbf{N}(\mathbf{Z},\boldsymbol{\theta},\mathcal{I})\in\mathcal{S}_N^R\}}\mathbb{1}_{\{\mathbf{X}(\mathbf{Z},\boldsymbol{\theta},\mathcal{I})\in\mathcal{S}_X^R\}}\pi(\mathbf{Z})\pi(\boldsymbol{\theta}). \tag{4.19}$$

We will denote by $\mathbf{N}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ and $\mathbf{X}(\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I})$ the incidence and prevalence sample paths that are output by the `doLNA` procedure. The procedure for this mapping, denoted `doLNA`, is presented in Algorithm (1).

The NCP of the log–incidence process substantially improves the mixing of MCMC chains that alternate between updates to $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ and $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$. Figure 4.1 shows traceplots of model parameters for one of MCMC chains for an SIR model fit to Poisson distributed incidence data using the centered parameterization (CP) of the LNA transition density. MCMC was run for 2.5 million iterations, following a tuning run of equal length, but each chain only manages to yield a effective sample sizes for R_0 and the infectious period duration in the

Algorithm 1 Mapping standard normal draws onto LNA sample paths.

procedure DOLNA($\mathbf{Z}, \boldsymbol{\theta}, \mathcal{I}$)
  initialize: $\mathbf{X}(t_0) \leftarrow \mathbf{x}_0,\ \mathbf{N}(t_0) \leftarrow \mathbf{0},\ \widetilde{\mathbf{N}}(t_0) \leftarrow \mathbf{0},\ \boldsymbol{\mu}(t_0) \leftarrow \mathbf{0},\ \boldsymbol{\Sigma}(t_0) \leftarrow \mathbf{0}$
  for $\ell = 1, \ldots, L$ **do**
    $\boldsymbol{\mu}(t_\ell),\ \boldsymbol{\Sigma}(t_\ell) \leftarrow$ solutions to (4.15) and (4.17) over $(t_{\ell-1}, t_\ell]$
    $\widetilde{\mathbf{N}}(t_\ell) \leftarrow \boldsymbol{\mu}(t_\ell) + \boldsymbol{\Sigma}(t_\ell)^{1/2}\mathbf{Z}(t_\ell)$ ▷ non–centered parameterization
    $\mathbf{N}(t_\ell) \leftarrow \mathbf{N}(t_{\ell-1}) + \exp(\widetilde{\mathbf{N}}(t_\ell)) - \mathbf{1}$
    restart initial conditions:
      $\mathbf{X}(t_\ell) \leftarrow \mathbf{X}(t_{\ell-1}) + \mathbf{A}^T(\mathbf{N}(t_\ell) - \mathbf{N}(t_{\ell-1})),\ \widetilde{\mathbf{N}}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\mu}(t_\ell) \leftarrow \mathbf{0},\ \boldsymbol{\Sigma}(t_\ell) \leftarrow \mathbf{0}$
return ▷ return incidence and/or prevalence sample paths
  $\mathbf{N} = \{\mathbf{N}(t_0), \mathbf{N}(t_1), \ldots, \mathbf{N}(t_L)\},\ \mathbf{X} = \{\mathbf{X}(t_0),\ \mathbf{X}(t_1), \ldots, \mathbf{X}(t_\ell)\}$

low double digits. In contrast, the NCP yields effective sample sizes per–chain of between 500–700 for each of the model parameters in only 50,000 iterations, following a short tuning run of equal length. Figure 4.2 shows the traceplot for one of the MCMC chains, which clearly mixes much better.

In each iteration of a DA MCMC algorithm, we alternate between updates to the latent path, conditional on the model parameters, and updates to the parameters, conditional on the latent path. Figure 4.3, which depicts the CP and NCP representations of an LNA path, provides some insight into why the NCP improves MCMC mixing. Under the CP (top plot), updates to $\boldsymbol{\theta}|\widetilde{\mathbf{N}}, \mathbf{Y}$ are made conditionally on a *fixed* LNA path. Therefore, proposed parameter values are accepted depending on whether they are concordant with the data *and* the current path. Even small perturbations to model parameters can result in shifts of the LNA distributions (grey densities) that would render the current path (red points) unlikely under the proposal. In contrast, perturbations to parameters implicitly perturb an LNA path defined using the NCP, even as the LNA draws, $\mathbf{Z}$, are clamped to their current values.

The NCP of the LNA also plays an important role in facilitating efficient updates of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via the elliptical slice sampling (ElliptSS) algorithm of [18], which was detailed in

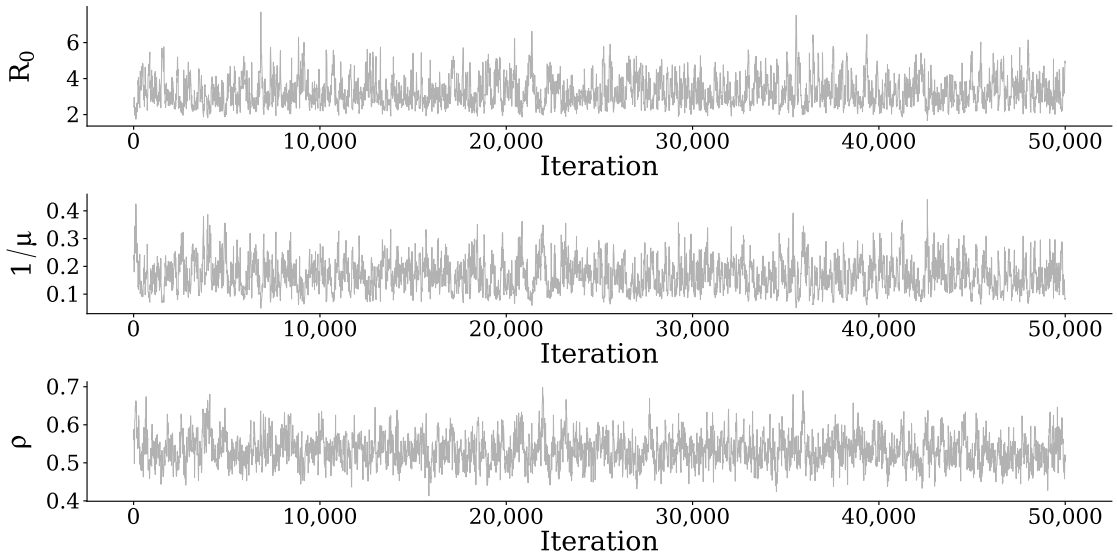

Figure 4.2: Posterior traceplots for a single MCMC chain of an SIR model fit to Poisson distributed incidence data. MCMC targeted the posterior, 4.19, alternately updating $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ via elliptical slice sampling, and $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$ via a multivariate random walk Metropolis algorithm. $R_0 = \beta N/\mu$ is the basic reproductive number, $1/\mu$ is the mean infectious period duration, and ρ is the mean case detection rate. The true values of R_0, $1/\mu$, and ρ were 3.5, 7, and 0.5, respectively.

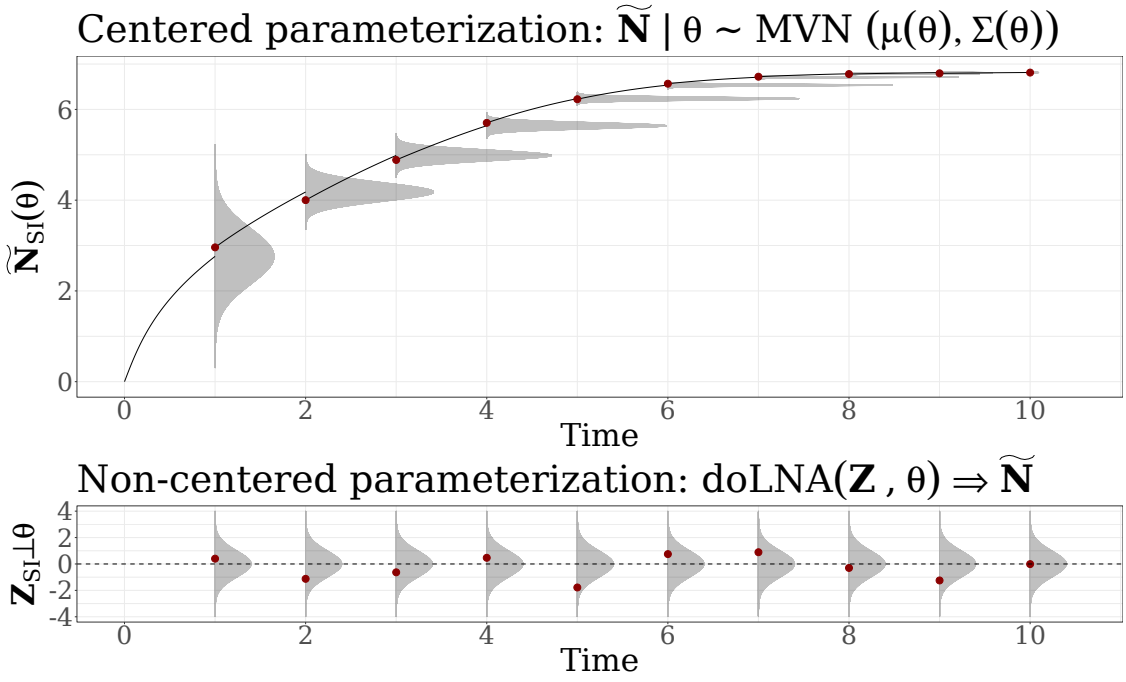

Figure 4.3: Centered (top) and non–centered (bottom) parameterizations of an LNA incidence path. In the CP, the log–incidence is normally distributed with mean and covariance obtained by solving the LNA ODEs, (4.15) and (4.17). In the NCP, the log–incidence is a draw from a standard normal distribution that is deterministically mapped to a sample path via the `doLNA` algorithm. In both the CP and NCP, state at the end of each interval determines the initial conditions of the LNA ODEs for the next interval. Plots of CP LNA transition densities are rescaled for clarity.

Section 2.3.1 and is presented in Algorithm 2. ElliptSS is an efficient and easy to implement MCMC algorithm for sampling Gaussian random variables, $\mathbf{Z}$, in models where the posterior of can be decomposed as the Gaussian prior for $\mathbf{Z}$ and an arbitrary likelihood, $L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta})$, i.e.,

$$\pi(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{Y}) \propto L(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}) MVN(\mathbf{Z}; \mu_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}). \tag{4.20}$$

The target posterior under the LNA NCP, (4.19), is of this form, regardless of whether the LNA is restarted at the beginning of each inter–observation interval, as in [7], or the non–restarting version is used as in [17]. Note that the CP cannot be expressed in the form (4.20) when we use the restarting version of the LNA. Although each transition density, (4.14), is itself Gaussian, the joint LNA path, $\mathbf{N}$, is not *a priori* Gaussian when the LNA ODEs are restarted since the mean of $\mathbf{N}(t_\ell)$ depends non–linearly on the value of $\mathbf{N}(t_{\ell-1})$. The quality of the LNA approximation is known to degenerate over long time intervals. Restarting the LNA ODEs has been established to improve the approximation when analyzing time series data of non–negligible length [7, 10]. Hence, use of the NCP is critical to enabling the use of ElliptSS for jointly updating of $\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y}$ when using the restarting version of the LNA.

Initializing the LNA draws

In simple models, reasonable parameter values will generally lead to valid LNA paths for initial $\mathbf{Z} \sim MVN(\mathbf{0}, \mathbf{I})$, i.e., paths that satisfy the monotonicity and positivity conditions, and thus have non–zero likelihood. However, this is not necessarily the case for complex models with many types of transition events, or when the time–series of incidence counts is long. One option is to include a re–sampling step after line 6 in Algorithm 1, in which $\mathbf{Z}(t_\ell)$ is redrawn in place until the conditions for a valid path over the interval are met. It is important to note that such a procedure does not sample from the correct distribution since $\mathbf{Z}$ is not actually a truncated multivariate Gaussian. To correct for this, we will "warm–up" the LNA path with an initial run of `doElliptSS`iterations in which the likelihood only consists of the indicators for whether the path is valid. Obviously, such chicanery is not always, or

Algorithm 2 Sampling LNA draws via elliptical slice sampling.

1: **procedure** `doElliptSS`($\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathbf{Y}, \mathcal{I}$)

2: Sample ellipse: $\mathbf{Z}_{prop} \sim N(\mathbf{0}, \mathbf{I})$

3: Sample threshold: $u|\mathbf{x} \sim \text{Unif}(0, L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}_{cur}, \boldsymbol{\theta}, \mathcal{I})))$

4: Initial proposal (also defining a bracket):

$$\phi \sim \text{Unif}(0, 2\pi)$$

$$(L_\psi, R_\psi) \leftarrow (\psi - 2\pi, \psi)$$

5: Set $\mathbf{Z}' \leftarrow \mathbf{Z}_{cur}\cos(\phi) + \mathbf{Z}_{prop}\sin(\phi)$.

6: **if** $L(\mathbf{Y}|\texttt{doLNA}(\mathbf{Z}', \boldsymbol{\theta}, \mathcal{I})) > u$ **then** accept $\mathbf{Z}'$

7: **return** $\mathbf{Z}'$

8: **else**

9: Shrink bracket and try a new angle.

10: **If:** $\phi < 0$ **then:** $L_\phi \leftarrow \phi$ **else:** $R_\phi \leftarrow \phi$

11: $\phi \sim \text{Unif}(L_\phi, R_\phi)$.

12: **GoTo:** 5.

even frequently necessary, and so we only resort to these shenanigans in dire circumstances. Finally, note that EllipSS, as well as any valid MCMC algorithm for updating $\boldsymbol{\theta}|\mathbf{Z}, \mathbf{Y}$, will never lead to an invalid LNA path as long as the current LNA draws and model parameters correspond to a valid path.

Parameter updates

Each MCMC iteration will consist of a number of EllipSS updates, typically one but possibly 2–3 for complex models, followed by a set of parameter updates. We will generally use the variant of the automated factor slice sampler of [24], possibly in conjunction with a hit–and–run slice sampler to preserve ergodicity when updating in only some slice directions. Details of these algorithms are presented in Section 2.3.1.

Chapter 5

DYNAMIC TRANSMISSION MODELING OF PANDEMIC A(H1N1) INFLUENZA IN FINLAND

Chapter 6

DISCUSSION AND FUTURE WORK

BIBLIOGRAPHY

[1] L.J.S. Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, New York, 2008.

[2] H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. Springer, New York, 2000.

[3] J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and D. West. Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 307. Oxford University Press, USA, 2003.

[4] C. Bretó, D. He, E.L. Ionides, and A.A. King. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, pages 319–348, 2009.

[5] T. Britton. Basic stochastic transmission models and their inference. *ArXiv e-prints*, January 2018.

[6] E. Buckingham-Jeffery, V. Isham, and T. House. Gaussian process approximations for fast inference from infectious disease data. *Mathematical biosciences*, 2018.

[7] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70:457–466, 2014.

[8] J. Fintzi, X. Cui, J. Wakefield, and V.N. Minin. Efficient data augmentation for fitting stochastic epidemic models to prevalence data. *Journal of Computational and Graphical Statistics*, 26:918–929, 2017.

[9] C. Fuchs. *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media, New York, 2013.

[10] V. Giagos. *Inference for auto-regulatory genetic networks using diffusion process approximations*. PhD dissertation, Lancaster University, 2010.

[11] D.T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113:297–306, 2000.

[12] A. Golightly and C.S. Gillespie. Simulation of stochastic kinetic models. In *In Silico Systems Biology*, pages 169–187. Springer, 2013.

[13] A. Golightly, D.A. Henderson, and C. Sherlock. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055, 2015.

[14] L.S.T. Ho, F.W. Crawford, and M.A. Suchard. Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. *arXiv preprint arXiv:1608.06769*, 2016.

[15] M.J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 2008.

[16] A.A. King, M.D. de Celles, F.M.G. Magpantay, and P. Rohani. Avoidable errors in the modeling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society, Series B*, 282:20150347, 2015.

[17] M. Komorowski, B. Finkenstädt, C.V. Harper, and D.A. Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10:343, 2009.

[18] I. Murray, R.P. Adams, and D.J.C. MacKay. Elliptical slice sampling. *JMLR: W&CP*, 9:541–548, 2010.

[19] P. Neal and G.O. Roberts. A case study in non-centering for data augmentation: stochastic epidemics. *Statistics and Computing*, 15:315–327, 2005.

[20] B. Øksendal. *Stochastic Differential Equations*. Springer, New York, 2003.

[21] P.D. O'Neill. Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in Medicine*, 29:2069–2077, 2010.

[22] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian Statistics*, 7:307–326, 2003.

[23] O. Papaspiliopoulos, G.O. Roberts, and M. Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, pages 59–73, 2007.

[24] M.M. Tibbits, C. Groendyke, M. Haran, and J.C. Liechty. Automated factor slice sampling. *Journal of Computational and Graphical Statistics*, 23:543–563, 2014.

[25] E.W.J. Wallace, D.T. Gillespie, K.R. Sanft, and L.R. Petzold. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET systems biology*, 6:102–115, 2012.

[26] D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, Boca Raton, 2011.

[27] Y. Yu and X. Meng. To center or not to center: That is not the question — An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20:531–570, 2011.

Appendix A

APPENDIX A