Bayesian inference for compartmental models with intractable likelihood function

Cecilia Viscardi[1] and Alessio Lachi [1,2] and Michela Baccini[1]

[1] University of Florence, Department of Statistics, Computer Science, and Applications "Giuseppe Parenti" (DiSIA)
cecilia.viscardi@unifi.it,
[2] Institute of Clinical Physiology of the Italian National Research Council (IFC-CNR), Epidemiology and Health Research

Abstract. Compartmental models are widespread methods useful to understand and predict the dynamic of a phenomenon of interest, such as the spreading process of an infectious disease. Their formalization relies on systems of ordinary differential equations parameterized through some quantities of interest, such as the basic reproduction number and the duration of the infectious period. However, to account for the intrinsic uncertainty of an epidemic, we need a stochastic version of the equations based on a suitable statistical model. Unfortunately, the associated likelihood function is often intractable and requires specific methods for making inferences. In this work, we investigate three different strategies for addressing the intractability of the likelihood within a Bayesian framework. In particular, we consider Data-Augmentation and Pseudo-Marginal Markov Chain Monte Carlo algorithms, along with a likelihood-free approach based on Approximate Bayesian Computation. We describe and compare the methods at work on a simple SIR model to highlight their strengths and weaknesses.

Keywords: Compartmental models, Bayesian inference, Monte Carlo methods, Intractable likelihood.

1 Introduction

Compartmental models are a class of models used to understand, describe, and predict the dynamic of a phenomenon of interest in a given population. Their mechanistic nature is well-suited for modeling infectious diseases. The main assumption of compartmental models is that the population of interest is divided, at each point in time, into compartments – i.e. non-overlapping groups of individuals that are homogeneous for some characteristic such as the health status. Starting from an initial composition of the population, transitions between compartments are allowed according to infectious dynamics described by a system of ordinary differential equations (ODEs). The system is governed by a vector of parameters that tune the transition rates. The ODEs solution is the deterministic evolution of the size of each compartment over time. However, a stochastic

extension of the system is essential to describe and quantify the uncertainty of the phenomenon. This extension relies on the assumption that transitions between compartments follow probabilistic rules defined by a statistical model. However, the likelihood function associated with the resulting stochastic model is often intractable or computationally demanding to evaluate, and specific inference methods are needed.

This paper aims to investigate the applicability of three different Bayesian estimation strategies. In particular, we want to provide a description and a comparison of their different ways of addressing the problem of the intractability of the likelihood. We consider a Data-Augumentation Markov Chain Monte Carlo (DA-MCMC) algorithm, a Pseudo Marginal Metropolis Hastings (PMMH) algorithm, and a likelihood-free approach based on Approximate Bayesian Computation (ABC). Finally, we test the methods at work on a simple SIR model, to assess their reliability and emphasize their strengths and weaknesses.

2 A simple example: the SIR model

Let us consider the well-known Susceptible-Infected-Removed (SIR) model [5]. It considers three compartments, each one corresponding to a health status related to an infectious disease: S is the group of susceptible individuals who can potentially contract the disease when they come in contact with infectious individuals; I is the group of infectious individuals; R is the group of the so-called removed individuals who, after the infection, have become immune from the disease or have died.

The following system of difference equations is the discrete-time version of the well-known system of ODEs in [3] that describes the dynamic of the model:

$$\begin{cases} S(t) = & S(t-1) - \pi_{\mathrm{SI}}(t-1)S(t-1) \\ I(t) = & I(t-1) + \pi_{\mathrm{SI}}(t-1)S(t-1) - \pi_{\mathrm{IR}}I(t-1) \\ R(t) = & R(t-1) + \pi_{\mathrm{IR}}I(t-1). \end{cases} \tag{1}$$

In Eq (1), $\pi_{\mathrm{SI}}(t-1)$ and π_{IR} are the (time-varying) probability of being infected and the (constant in time) probability of recovering/dying during a unit time interval, respectively. We assume that the waiting times before experimenting with an event have an exponential distribution, hence $\pi_{\mathrm{SI}}(t-1) = 1 - \exp\big(-\frac{R_0}{\tau}\frac{I(t-1)}{S(0)}\big)$ and $\pi_{\mathrm{IR}} = 1 - \exp(-\frac{1}{\tau})$, where τ is the average infectious period and R_0 is the basic reproduction number – i.e. the expected number of secondary infections caused by a single infected individual at the beginning of the epidemic.

To formalize a statistical model, we assume that, at each point in time, the number of new infections and new resolutions are random variables. Let $i_{t_1:t_2} = \big(i(t_1), ..., i(t_2)\big)$ and $r_{t_1:t_2} = \big(r(t_1), ..., r(t_2)\big)$ denote the random vectors of new infections and new resolutions from t_1 until t_2 (t_1 and $t_2 \in \{0, ...T\}$, $t_2 \geq t_1$), and by $i^*_{t_1:t_2}$ and $r^*_{t_1:t_2}$ their realization. Under the assumption that

$i(t) \sim \text{Binom}\big(S(t), \pi_{\text{SI}}(t)\big)$ and $r(t) \sim \text{Binom}\big(R(t), \pi_{\text{IR}}\big)$ for each $t \in \{0, ..., T\}$, we can write down a simple analytical form for the likelihood function based on complete data – i.e. $i^*_{t_1:t_2}$ and $r^*_{t_1:t_2}$. However, when dealing with infectious diseases, there is often a problem of missing data: usually, we only observe the daily number of new infections, $i^*_{1:T}$, since recoveries are not accurately notified. If $r_{1:T}$ is a latent variable, the evaluation of the likelihood requires the following marginalization:

$$L\big(\theta \mid i_{1:T} = i^*_{1:T}\big) = \sum_{r^*_{1:T} \in \mathcal{R}} \Pr\left(i_{1:T} = i^*_{1:T}, r_{1:T} = r^*_{1:T} \mid \theta\right)$$
$$= \sum_{r^*_{1:T} \in \mathcal{R}} \prod_{t=1}^{T} \binom{S(t)}{i^*(t)} \pi_{\text{SI}}(t)^{i^*(t)} [1 - \pi_{\text{SI}}(t)]^{S^*(t) - i^*(t)} \binom{I^*(t)}{r^*(t)} \pi_{\text{IR}}^{r^*(t)} [1 - \pi_{\text{IR}}]^{I^*(t) - r^*(t)}. \tag{2}$$

In Eq (2), $S^*(t) = S^*(t-1) - i^*(t-1)$,$I^*(t) = I^*(t-1) + i^*(t-1) - r^*(t-1)$, and $\big(S(0), I(0), R(0)\big)$, the initial condition of the system, is assumed to be known. Note that $\mathcal{R}$ is the space of all possible sequences $r^*_{1:T}$ – i.e. realizations of $r_{1:T}$ – that are compatible with the observed series $i^*_{1:T}$. The structure and the cardinality of $\mathcal{R}$ often make the point-wise evaluation of the likelihood computationally intensive or infeasible.

3 Three different approaches for Bayesian inference

In Bayesian statistics, the parameters $\theta = (\tau, R_0)$ that govern the dynamic of the epidemic are considered as random variables with a prior distribution defined over the space Θ, here denoted by $\pi(\cdot)$. Thus, given the observed data $i^*_{1:T}$, our target is the posterior distribution derived through Bayes's formula:

$$\pi(\theta \mid i^*_{1:T}) \propto \pi(\theta) L(\theta \mid i^*_{1:T}) = \pi(\theta) \sum_{r^*_{1:T} \in \mathcal{R}} \Pr\left(i_{1:T} = i^*_{1:T}, r_{1:T} = r^*_{1:T} \mid \theta\right). \tag{3}$$

From Eq (3) it is apparent that the computation of the posterior probabilities requires multiple evaluations of the intractable likelihood. We propose and compare three ways of overcoming this problem to get full Bayesian estimates.

3.1 Data Augmentation Metropolis-Hastings

A possible solution is that of getting samples from the posterior distribution $\pi(\theta, r_{1:T} \mid i^*_{1:T})$ defined on the augmented space $\Theta \times \mathcal{R}$, rather than from $\pi(\theta \mid i^*_{1:T})$. To this aim, Data Augmentation Markov Chain Monte Carlo methods (DA-MCMC) [6] can be implemented using the Metropolis-Hastings (MH) scheme displayed in Alg 1. Here, the key idea is that getting samples from the joint posterior avoids the use of the intractable likelihood since, at each iteration, we impute the missing series and compute an acceptance ratio based only on the complete likelihood $L(\theta \mid i^*_{1:T}, r^*_{1:T})$. Since the produced Markov chain, has limiting distribution $\pi(\theta, r_{1:T} \mid i^*_{1:T})$, after the assessment of the convergence, samples $\theta^{(0)}, ..., \theta^{(S)}$ can be considered as drawn from the target $\pi(\theta \mid i^*_{1:T})$.

Algorithm 1 Metropolis-Hastings

Inizialize $\theta^{(0)}, r_{1:T}^{(0)}$
for s in $1:S$ **do**
  Propose missing data $r_{1:T}^{*} \sim q_r(\cdot \mid r_{1:T}^{(s-1)})$
  Propose $\theta^{*} \sim q_\theta(\cdot \mid \theta^{(s-1)})$
  Compute $\alpha = \min\left\{1, \frac{\pi(\theta^{*})L(\theta^{*} \mid r_{1:T}^{*}, i_{1:T}^{*})q_r(r_{1:T}^{(s-1)} \mid r_{1:T}^{*})q_\theta(\theta^{(s-1)} \mid \theta^{*})}{\pi(\theta^{(s-1)})L(\theta^{(s-1)} \mid r_{1:T}^{(s-1)}, i_{1:T}^{*})q_r(r_{1:T}^{*} \mid r_{1:T}^{(s-1)})q_\theta(\theta^{*} \mid \theta^{(s-1)})}\right\}$
  Sample $u \sim U(0,1)$
  if $u < \alpha$ **then**
    Set $\theta^{(s)} = \theta^{*}$ and $r_{1:T}^{(s)} = r_{1:T}^{*}$
  else
    set $\theta^{(s)} = \theta^{(s-1)}$ and $r_{1:T}^{(s)} = r_{1:T}^{(s-1)}$
  end if
end for

3.2 Pseudo Marginal Metropolis-Hastings

Another possible strategy is based on the formalization of the SIR model as a State Space Model (SSM). This approach allows the use of a Pseudo Marginal Metropolis-Hastings algorithm (PMMH) to get samples directly from the *marginal* posterior, $\pi(\theta \mid i_{i:T}^{*})$ (see [2]). It relies on an *approximation* of the likelihood obtained by running a Particle Filter (PF) algorithm at each iteration of the MH. To formalize our model as SSM, without introducing assumptions different from those used in the other two methods, we propose to consider the two following discrete-time processes: 1) the latent process $X_{1:T} = \big(S(1), R(1)\big), ..., \big(S(T), R(T)\big)$; 2) the observed process $Y_{1:T} = i_{1:T}$. According to the probabilistic rules defined in Sect 2, we can show that 1) $\Pr(X_t = x_t \mid x_{t-1}) = \mathrm{Bin}(i^{*}(t-1); \pi_{\mathrm{SI}}(t-1), S(t-1)) \cdot \mathrm{Bin}(r^{*}(t-1); \pi_{\mathrm{IR}}, I(t-1))$; 2) $\Pr(Y_t = y_t \mid x_{t-1}) = \Pr(i(t) = i^{*}(t) \mid x_t) = \mathrm{Bin}(i^{*}(t); \pi_{\mathrm{SI}}(t), S(t))$. This allows us to estimate $L(\theta \mid i_{1:T}^{*}) = \Pr(Y_{1:T} = i_{1:T}^{*})$ using a PF and use it to replace the intractable likelihood in Alg 2.

Algorithm 2 Pseudo-Marginal Metropolis-Hastings

Inizialize $\theta^{(0)}$
for s in $1:S$ **do**
  Propose $\theta^{*} \sim q_\theta(\cdot \mid \theta^{(s-1)})$
  Compute the estimate of the likelihood function $\widehat{L}(\theta^{*} \mid i_{1:T}^{*})$ running a PF
  Compute $\alpha = \min\left\{1, \frac{\pi(\theta^{*})\widehat{L}(\theta^{*} \mid i_{1:T}^{*})q_\theta(\theta^{(s-1)} \mid \theta^{*})}{\pi(\theta^{(s-1)})\widehat{L}(\theta^{(s-1)} \mid i_{1:T}^{*})q_\theta(\theta^{*} \mid \theta^{(s-1)})}\right\}$
  Sample $u \sim U(0,1)$
  if $u < \alpha$ **then**
    Set $\theta^{(s)} = \theta^{*}$
  else
    set $\theta^{(s)} = \theta^{(s-1)}$
  end if
end for

3.3 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is a likelihood-free method that dispenses with likelihood computation and only requires the availability of a

computer program, usually called "simulator", that reproduces the stochastic data generative process [7]. The underlying idea is that samples from the prior distribution can be converted into samples from the posterior through three simple steps: 1) draw S parameter proposals from $\pi(\cdot)$; 2) give each parameter proposal, $\theta^{(s)}$, as an input to the simulator $\Pr(\cdot \mid \cdot)$ to produce pseudo-data $\widehat{i}_{1:T}(\theta^{(s)})$; 3) retain only parameter proposals such that $\rho(\widehat{i}_{1:T}(\theta^{(s)}); i^*_{1:T}) < e$, where $\rho(\cdot;\cdot)$ is a suitable distance function and e is a positive tolerance threshold.

More sophisticated sampling schemes are based on a decreasing sequence of thresholds, rather than a fixed tuning parameter. Examples are the Population Monte Carlo ABC [1] and some adaptive versions inspired by it [4], such as the one summarized in Alg 3.

Algorithm 3 Adaptive Population Monte Carlo ABC

Initialize e_1
for j in $1:M$ **do**
  Simulate $\theta_j^{(1)} \sim \pi(\cdot)$ and $\widehat{i}_{1:T}(\theta_j^{(1)}) \sim \Pr(\cdot \mid \theta_j^{(1)})$ until $\rho(\widehat{i}_{1:T}(\theta_j^{(1)}); i^*_{1:T}) < e_1$
  Set $\omega_j^{(1)} = \frac{1}{M}$
end for
Select e_2 using an adaptive strategy.
for s in $2:S$ **do**
  Set Σ_s to twice the empirical covariance matrix of $\theta_1^{(s-1)}, ..., \theta_M^{(s-1)}$
  for j in $1:M$ **do**
    Pick θ_j^* from $(\theta_1^{(s-1)}, ..., \theta_M^{(s-1)})$ with probabilities $(\omega_1^{(s-1)}, ..., \omega_M^{(s-1)})$
    Generate $\theta_j^{(s)} | \theta_j^* \sim MVN(\theta_j^*, \Sigma_s)$ and $\widehat{i}_{1:T}(\theta_j^{(s)}) \sim \Pr(\cdot \mid \theta_j^{(s)})$
    Set $\omega_j^{(s)} \propto \frac{\pi(\theta_j^{(s)})}{\sum_{m=1}^{M} \omega_m^{(s-1)} \phi\{\tau_s^{-1}(\theta_m^{(s)} - \theta_m^{(s-1)})\}} \mathbb{1}\{\rho(\widehat{i}_{1:T}(\theta_j^{(s)}); i^*_{1:T}) < e_s\}$
      where ϕ represents the density of the Standard Normal distribution
  end for
  Select e_{s+1} using an adaptive strategy.
end for

The output is a sample from a target distribution different from that of the previous methods since it is an *approximate* posterior distribution:

$$\pi_{e_S}(\theta \mid i^*_{1:T}) \propto \pi(\theta) \Pr(\rho(\widehat{i}_{1:T}(\theta^{(s)}); i^*_{1:T}) < e_S).$$

Notice that, under mild conditions, $\Pr(\rho(\widehat{i}_{1:T}(\theta^{(s)}); i^*_{1:T}) < e_S) = L(\theta \mid i^*_{1:T})$ when $e_S = 0$. Thus, the greater the value of e_S, the greater the approximation w.r.t. the true posterior.

4 Results and discussion

For the sake of comparison among methods, we considered as observed data $i^*_{1:90}$ the output of a simulation from the model with parameters $R_0^{\text{true}} = 3$ and $\tau^{\text{true}} = 14$. We set uniform prior distributions: $R_0 \sim \text{Unif}(0, 6)$ and $\tau \sim \text{Unif}(7, 14)$. The proposal distribution $q_r(\cdot \mid \cdot)$ in Alg 1 is that suggested in [6], while $q_\theta(\cdot \mid \cdot)$ are Normal distributions. We gave all the algorithms a budget of time of 10

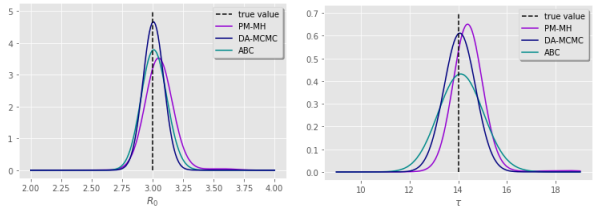

Fig. 1: Comparison among posterior distributions for R_0 (left) and τ (right).

	R_0	τ
DA-MCMC	0.0037	0.2162
PMMH	0.0114	0.4748
ABC	0.0057	0.4385

Table 1: Bayesian Mean Square Error (BMSE).

minutes. Looking at Fig 1, we can see that their outputs are quite similar in terms of approximation of the posterior distributions. This is a piece of evidence in favor of the reliability of all the methods. To compare the ability to retrieve the "true" values, we computed the Bayesian Mean Squared Errors. Looking at Tab 1 we can see that DA-MCMC outperforms the other methods. However, since it defines the chain on an augmented space, the produced Markov chain is strongly autocorrelated and a tremendous number of iterations are required to provide a valid sample. We speculate that this problem is more serious when the compartmental model is more complex than the SIR model, since the size of the latent variable space could be greater. PMMH does not suffer from this problem but it is highly sensitive to the quality of the likelihood approximation provided by PF, and this approximation could be particularly challenging in more complex models. Finally, ABC, while prone to an approximation error induced by the threshold e_S, is highly parallelizable and seems to be more promising in terms of extensions to complex cases.

References

1. M. Beaumont, J. Cornuet, J. Marin, and C. Robert. Adaptive approximate bayesian computation. *Biometrika*, 96(4):983–990, 2009.
2. N. Chopin, O. Papaspiliopoulos, et al. *An introduction to sequential Monte Carlo*, volume 4. Springer, 2020.
3. H. W. Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.
4. M. Lenormand, F. Jabot, and G. Deffuant. Adaptive approximate bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796, 2013.
5. A. Mckendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925.
6. P. O'Neill and G. Roberts. Bayesian Inference for Partially Observed Stochastic Epidemics. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 162(1):121–129, 1999. Publisher: [Wiley, Royal Statistical Society].
7. S. Sisson, Y. Fan, and M. Beaumont, editors. *Handbook of approximate Bayesian computation*. CRC Press, Taylor & Francis Group, Boca Raton, 2019.