Approximate Bayesian Computation for Spatial SEIR Epidemic Models

Approximate Bayesian Computation (ABC) provides an attractive approach to estimation in complex Bayesian inferential problems for which evaluation of the kernel of the posterior distribution is impossible or computationally expensive. These highly parallelizable techniques have been successfully applied to many fields, particularly in cases where more traditional approaches such as Markov chain Monte Carlo (MCMC) are impractical. In this work, we demonstrate the application of approximate Bayesian inference to spatially heterogeneous Susceptible-Exposed-Infectious-Removed (SEIR) stochastic epidemic models. These models have a tractable posterior distribution, however MCMC techniques nevertheless become computationally infeasible for moderately sized problems. We have used the open source ABSEIR package for R. The performance of ABC relative to traditional MCMC methods in a small problem is explored under simulation, as well as in the spatially heterogeneous context of the Covid-19 pandemic in different states of India.

Introduction

The study of epidemics is complicated by the fact that real human populations exhibit complex structure and interact in subtle ways over both space and time. Nevertheless, in an increasingly globalized world, the ability to model pathogen outbreaks, predict ongoing spread, and evaluate interventions represents crucial abilities of public health practitioners. In this work we used an open source software, AB-SEIR through which we will estimate the SEIR model parameters.

Approximate Bayesian Computation (ABC)

Approximate Bayesian Computing is generally attributed to the work of Rubin (1980), which concerns interpretation and implementation of practical modeling techniques for applied Bayesian statisticians. Among other contributions, this work introduced one of the most commonly used algorithmic approaches to ABC: the rejection algorithm. This procedure provides an intuitive introduction to approximate Bayesian computing

techniques. We therefore begin our approach to the subject by introducing the requisite notation, and describing the basic ABC rejection algorithm.

Define a $p \times 1$ parameter vector θ with p dimensional parameter space Θ and prior distribution $\pi_\Theta(\theta)$. Further define an $N \times 1$ vector of observed data, y, with a likelihood or data generating distribution denoted by $f_Y(y|\theta)$. Finally, define a distance function (such as the Euclidean distance) between appropriately sized vectors x and y: $\rho(y, x)$. As a Bayesian sampling technique, the goal of ABC is to make inference about the posterior distribution, $f_\Theta(\theta|Y) \propto f_Y(y|\theta)\pi_\Theta(\theta)$.

The general pattern of rejection sampling ABC is quite simple. We first generate repeated samples θ_i from the prior distribution for θ. Each of these samples, indexed by i, is in turn used to generate a replicate data set x_i from the likelihood. Parameters which generate replicate data sets which are sufficiently 'close' to the observed data y, according to the distance function ρ and a tolerance ε, are retained, while the rest are discarded. Details of this procedure are shown below.

Algorithm : ABC rejection algorithm

Require : Define a tolerance $\varepsilon > 0$, and let '$\leftarrow$' denote assignment
1: for i $\leftarrow$ 1 to n do
2: d $\leftarrow \infty$
3: while d > ε do
4: draw $\theta_i \sim \pi(\Theta)$
5: draw $x_i \sim fY(y|\theta)_i$
6: d $\leftarrow \rho(y, x_i)$

It should be noted that this approach does not require the user to evaluate the potentially expensive or unavailable likelihood function, but does require the ability to draw samples from it . In its original formulation, the tolerance, ε, was taken to be zero. The key insight of the rejection approach is clear in this context: accepting only parameters which produce replicate data identical to the observed response is equivalent to conditioning on that observed data. The distribution of parameter values conditional on the observed data is the posterior distribution: our inferential target. The most commonly applied version of the algorithm, however, generally includes the aforementioned nonzero tolerance, and employs a distance measure which depends only on a set of summary statistics of x and y, thus rendering the inference 'approximate'.

Mechanistic Models

Mechanistic models are used to model and predict the spread of diseases and infections such as flu, Ebola or even coronavirus. A distinguishing feature of mechanistic models over machine learning models is their ability to encode hypothesized causal nature of disease transmission. This typically involves constructing mathematical formulations of causal mechanisms and using analytical tools to validate these hypotheses with observed data.

In short, mechanistic models are interpretable while machine learning models, more often than not, are not. Because of this, a mechanistic model which explains the phenomenon well can give insightful information about how a disease spreads and possibly how it can be dealt with.

Compartmental Models

Compartmental models are a type of mechanistic models which divide the population into groups referred to as compartments. They are based on a system of ordinary differential equations which express the dynamics between different epidemiological states of a population (aka compartments). They are useful for short and long-term forecast of the spread of a phenomenon, e.g. a disease, and can also be used to study the effect of different interventions.

SEIR model

The compartments and progression between epidemiological states in SEIR model as shown in the figure below

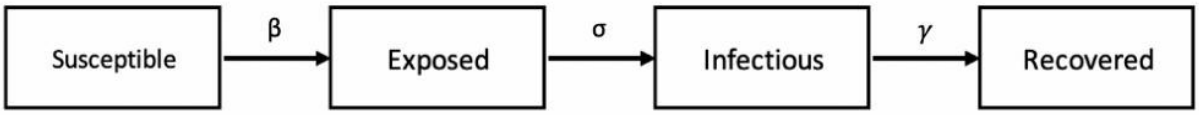

Fig. SEIR model

The SEIR model is defined by a system of ordinary differential equations (ODEs). These Equations are :

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$N = S + E + I + R$$

Where,

β is infection rate or the rate of spread

σ is the incubation rate or the rate of latent individuals becoming infectious (average duration of incubation is $1/\sigma$)

γ is the recovery rate or mortality rate. If the duration of indection is D then γ = 1/D

ABSEIR Package

The ABSEIR R package provides a user friendly interface for specifying models in the spatial SEIR(S) class. The software implements the aforementioned SMC-ABC algorithm with a variety of tunable parameters, and facilitates numerical and graphical summary of model results. Parallelism between simulations is achieved via the threading capabilities of modern C++. Implementing parallel simulations at the C++ level, as opposed to using process-level parallelism such as that provided by the 'parallel' R-package , enables ABSEIR to distribute work between multiple cores with minimal overhead. This is important for SMC-ABC models, because while the work of simulating epidemics dwarfs the rest of the algorithmic computational cost, numerous iterations may still be required. Low level parallelism is particularly beneficial for the ability of software to control how memory is accessed and copied.

Models are specified by constructing a set of model components:

- **DataModel**: describes the relationship of the observed data to the epidemic quantity of interest
- **ExposureModel**: captures the exposure covariate structure X^{SE} and specifies prior parameters for β^{SE}
- **ReinfectionModel**: determines whether a model includes a reinfection process, and if so defines X^{RS} and prior parameters for β^{RS}
- **DistanceModel**: defines, for models incorporating more than one spatial location, the set of distance matrices $\{D_z\}$ and prior distributions for the autocorrelation parameters $\{\rho_z\}$
- **TransitionPriors**: specifies a model for the E to I and I to R transitions using prior transition probabilities and associated effective sample sizes. The software also provides utility functions to assist in the creation of exponential transition models, arbitrary distribution path-specific models, and fully parameterized Weibull path-specific transition models.

- **InitialValues**: provides S_0, E_0, I_0, and R_0, vectors of compartment membership counts at the beginning of the study period
- **SamplingControl**: indicates which algorithm is to be used in fitting the model, as well as values of the requisite tuning parameters

Data

We will do parameter estimation for Covid-19 pandemic data for different states in India. Right now we have done it for West Bengal state. Daily new case data is extracted from https://data.covid19india.org/ . Daily new case count data for West Bengal state during the period 17th March, 2020 till 31st October 2021 is shown below.

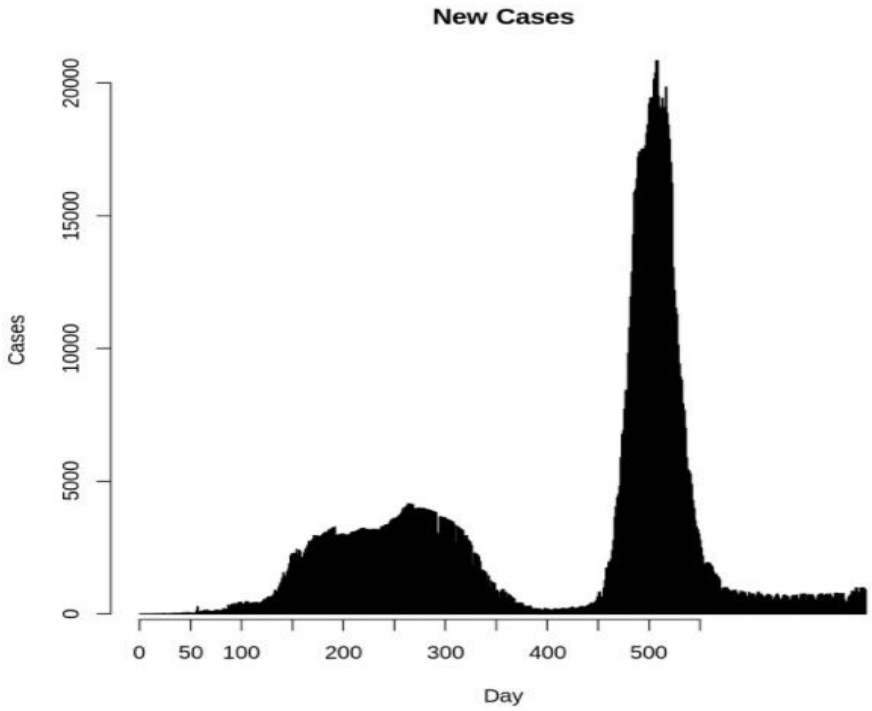

Fig. Daily new cases data for West Bengal

Model

Initial Values of SEIR : S_0 = 5.36e6, E_0 = 2, I_0 = 2 and R_0 = 0

Transition prior probabilities :

p_ei = 1 - exp(-1/5)
p_ir = 1-exp(-1/7)
p_ei_ess = 10
p_ir_ess = 10

So, using the above initial values and transition prior probabilities we will estimate the parameters of SEIR models, i.e. σ, β and γ. The ordinary differential equations(ODEs) are given by :

$$\frac{dS}{dt} = -\frac{\beta S I}{N}$$

$$\frac{dE}{dt} = \frac{\beta S I}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$N = S + E + I + R$$

Results

Parameter Estimates are :

	Mean	SD	95% LB	95% UB
Beta_SE_1 (β)	-9.82e-01	0.162	-1.283e+00	-6.63e-01
gamma_EI (σ)	5.68e-01	0.172	3.090e-01	9.66e-01
gamma_IR (γ)	3.95e-01	0.081	2.690e-01	5.75e-01