iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43818

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Bayesian MCMC Approach to Learning About the SIR Model

Abhik Biswas
Department of Mathematics, St. Xavier's College (Autonomous), Kolkata

Abstract: ***This project aims to study the parameters of the Deterministic SIR(Susceptible → Infected → Recovered) model of COVID-19 in a Bayesian MCMC framework. Several deterministic mathematical models are being developed everyday to forecast the spread of COVID-19 correctly. Here, I have tried to model and study the parameters of the SIR Infectious disease model using the Bayesian Framework and Markov-Chain Monte-Carlo (MCMC) techniques. I have used Bayesian Inference to predict the Basic Reproductive Rate R_t in real time using and following this, demonstrated the process of how the parameters of the SIR Model can be estimated using Bayesian Statistics and Markov-Chain Monte-Carlo Methods.***
Keywords: COVID-19, Bayesian Inference, Dynamical Systems, SIR Model, Basic Reproductive Rate, Markov-Chain Monte-Carlo(MCMC)

I. INTRODUCTION

The first COVID-19 case was detected in November 2019 and since then, the world is in a turmoil. Scientists around the globe are continuously trying to come up with multiple deterministic as well as stochastic models to understand and predict the spread of this virus. Bayesian Analysis is a method of Statistical Inference that allows us to combine prior information about a population parameter with evidence from information contained in a sample to guide statistical inference process.

In statistics, Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. By constructing a Markov chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by recording states from the chain. The more steps are included, the more closely the distribution of the sample matches the actual desired distribution. Various algorithms exist for constructing chains and in this paper, I have taken help of the Metropolis–Hastings algorithm. Taking help of the Bayesian Framework and the MCMC Techniques, I have tried to model the parameters of the SIR model for Infectious Diseases, which would assist us to correctly understand the spread of COVID-19, based on real time data.

II. THE BAYESIAN FRAMEWORK

The Bayesian Framework is set up by assuming a probability model for the observed data D given a $p \times 1$ vector of unknown parameters $\vec{\beta}$, which is $P(D|\vec{\beta})$. Then it is assumed that $\vec{\beta}$ is randomly distributed from the prior distribution $P(\vec{\beta})$. The inference for $\vec{\beta}$ is based on the Posterior distribution $P(\vec{\beta}|D)$.

Using Bayes' Theorem, we have

$$P(\vec{\beta}|D) = \frac{P(D|\vec{\beta})P(\vec{\beta})}{P(D)}$$
$$= \frac{P(D|\vec{\beta})P(\vec{\beta})}{\int_{\Omega} P(D|\vec{\beta})P(\vec{\beta})d\vec{\beta}}$$
$$\propto L(\vec{\beta})P(\vec{\beta}) = \pi(\vec{\beta}|D) \quad (1)$$

where $L(\vec{\beta})$ is the Likelihood Function and Ω is the space of the parameters contained in $\vec{\beta}$. $\int_{\Omega} P(D|\vec{\beta})P(\vec{\beta})d\vec{\beta}$ is known as the prior predictive distribution and it is a constant, which normalizes the posterior distribution $P(\vec{\beta}|D)$. $\pi(\vec{\beta}|D)$ is the unnormalized posterior distribution, $L(\vec{\beta})P(\vec{\beta})$.

III. BAYESIAN INFERENCE FOR DYNAMICAL SYSTEMS

A common method for performing parameter estimation of Dynamical Systems is to use Bayesian Inference. Some key points about Dynamical Systems:

- Assume that the dynamical system of interest can be described by the vector equation
- $x' = f(x)$ (2) , $x = (x_1, x_2, \ldots x_k)$ and $f = \big(f(x_1), f(x_2), \ldots f(x_k)\big)$, with the vector initial conditions $x^0 = (x_1^0, x_2^0, \ldots x_k^0)$.
- It is assumed that the unique solution vector $x(t)$ of (2) exists and can be obtained either explicitly or by using numerical methods.
- All the parameters in system are denoted by the vector $\vec{\beta}$.
- If the initial conditions $x_1^0, x_2^0, \ldots x_k^0$, then they are also included in the vector $\vec{\beta}$.
- An emphasis is put on the dependence of the unique solution x on the vectors $\vec{\beta}$ and time t and the unique solution vector will thus be denoted as $x(\vec{\beta}, t)$

The Data: Suppose there are m time series data sets. It is important to ensure the correct ODE model solution or combination of ODE Model solutions is fit to the j^{th} time series data set. $(j = 1,2, \ldots m)$.Sometimes a data set is scaled differently than the model solutions or the data set can be described by a summation of the ODE model solutions. In order to include these situations, we can use a linear combination of the ODE model solutions $\alpha_1^j x_1(\vec{\beta}, t) + \alpha_2^j x_2(\vec{\beta}, t) + \cdots + \alpha_k^j x_1(\vec{\beta}, t)$. If the nonzero vector of constants a^j need be estimated, then, let the nonzero vectors a^j , $j = 1,2, \ldots m$ be contained in the vector

$$\nu = \begin{bmatrix} \vec{\beta} \\ a^1 \\ a^2 \\ a^3 \\ \vdots \\ a^m \end{bmatrix}$$

In general, we will fit the function $F\big(x_1(\vec{\beta}, t_i^j), \ldots x_k(\vec{\beta}, t_i^j), a^1, \ldots a^m\big)$ to the j^{th} data set.

Distribution of Data over time: The distribution of the observations over time for each j^{th} data set must be chosen before fitting system (1) to the data. The following portion will describe the Gaussian, Poisson, Negative Binomial and other distribution options.

A. *Gaussian Distribution*

Let X be a random variable from the Gaussian Distribution with parameters μ and $\sigma^2 = \frac{1}{\tau} > 0$ and $Y \sim N(\mu, \sigma^2)$. The PDF is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathrm{R} \qquad (3)$$

The mean $\mathrm{E}(X) = \mu$ and the variance $Var(X) = \sigma^2 = \frac{1}{\tau}$. Assume that the j^{th} time series data set is given by observations $D_j = \{d_1^j, d_2^j, \ldots d_{n_j}^j\}$ with corresponding times $T_j = \{t_1^j, t_2^j, \ldots, t_{n_j}^j\}$ and the probability of observing d_i^j is

$$f\big(d_i^j\big) = \sqrt{\frac{\tau^j}{2\pi}} e^{-\tau^j \frac{\left(d_i^j - \mu_i^j\right)^2}{2}} \qquad (4)$$

where the mean μ_i^j changes depending on time t_i^j and the variance $\frac{1}{\tau^j}$ is specific to the j^{th} dataset. Given our assumption of fitting the function of the ODE model solutions and any necessary constants, $F\big(x_1(\vec{\beta}, t_i^j), \ldots, x_k(\vec{\beta}, t_i^j), a^1, a^2, \ldots a^m\big)$ to the j^{th} time series data set, we set

$$\mathrm{E}\big[D_i^j\big] = \mu_i^j = F\big(x_1(\vec{\beta}, t_i^j), \ldots, x_k(\vec{\beta}, t_i^j), a^1, a^2, \ldots a^m\big) \qquad (5)$$

Equation (4) can be thought of as a link function that transforms mean of a distribution to a linear regression model. Equation (4) equated the mean of the Gaussian Distribution to the ODE model solutions.

B. Poisson Distribution

Let X be a random variable from the Poisson Distribution with parameter $\lambda > 0, X \sim P(\lambda)$. The PMF of the discrete Poisson distribution is given by

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x \in Z \cap x \geq 0 \qquad (6)$$

The mean $E(X) = \lambda$ and variance $Var(X) = \lambda$. Assume that the j^{th} time series data is given by observations $D_j = \{d_1^j, \ldots d_{n_j}^j\}$ with the corresponding times $T_j = \{t_1^j, t_2^j, \ldots t_{n_j}^j\}$ and the probability of observing d_i^j is given by the Poisson Distribution

$$f(d_i^j) = \frac{e^{-\lambda_i^j}(\lambda_i^j)^{d_i^j}}{d_i^{j}!} \qquad (7)$$

where the mean λ_i^j changes depending on time t_i^j and hence, the variance λ_i^j also changes w.r.t time. Again, we will use equation (4) to equate the mean $E[D_i^j] = \lambda_i^j$ to the ODE model solutions.

C. Lognormal Distribution

A random variable X is said to be log-normally distributed if its natural logarithm is normally distributed. Let X be a random variable from the Lognormal distribution with parameters $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. The PDF is given by

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, \; x \in (0, \infty) \qquad (8)$$

The mean $E(X) = e^{\mu+\frac{1}{2}\sigma^2}$ and variance $Var(X) = e^{\sigma^2+2\mu}(e^{\sigma^2+2\mu} - 1)$. Assume that the j^{th} time series data is given by observations $D_j = \{d_1^j, \ldots d_{n_j}^j\}$ with the corresponding times $T_j = \{t_1^j, t_2^j, \ldots t_{n_j}^j\}$ and the probability of observing d_i^j is given by

$$f(d_i^j) = \frac{1}{d_i^j \sigma_j \sqrt{2\pi}} e^{-\frac{\left(\ln(d_i^j)-\mu_i^j\right)^2}{2\sigma_j^2}}, \; d_i^j \in (0, \infty) \qquad (9)$$

where the mean $e^{\mu_i^j+\frac{1}{2}\sigma_j^2}$ changes depending on time t_i^j and the variance $e^{\sigma_j^2+2\mu_i^j}\left(e^{\sigma_j^2+2\mu_i^j} - 1\right)$ also changes w.r.t. time. Again, we use Equation (4) to equate the mean $E[D_i^j] = \mu_i^j$ to the ODE model solutions.

Likelihood Functions: We will consider the datasets to be independent from each other. With m independent time series datasets, we will have m likelihood functions associated with them and the combined likelihood function is given by

$$L(\vec{\theta}) = CL_1(\vec{\theta})L_2(\vec{\theta}) \ldots L_m(\vec{\theta}) \qquad (10)$$

where $\vec{\theta}$ is the vector of parameters to estimate and C is any positive constant not dependent on $\vec{\theta}$, whose purpose is to simplify the likelihood function (clearing off the denominator for simplicity etc.).

D. Gaussian Model for m data sets and the Likelihood Function

Assume that for $j = 1,2, \ldots m$, the j^{th} time series data is given by $D_j = \{d_1^j, d_2^j, \ldots d_{n_j}^j\}$ with corresponding times $T_j = \{t_1^j, \ldots t_{n_j}^j\}$ and the probability density of observing d_i^j is given by Equation (3). The probability density of the observed counts $D = \{D_1, D_2, \ldots D_m\}$ given parameters $\vec{\theta}$ is given by

$$P(D|\theta) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \sqrt{\frac{\tau_j}{2\pi}} e^{-\tau_j \frac{\left(d_i^j-\mu_i^j\right)^2}{2}}$$

$$= \left(\frac{1}{2\pi}\right)^{\sum_{j=1}^{m} \frac{n_j}{2}} (\tau^1)^{\frac{n_1}{2}} \ldots (\tau^m)^{\frac{n_m}{2}} e^{-\frac{1}{2}\sum_{j=1}^{m} \tau^j \sum_{i=1}^{n_j}\left(d_i^j-\mu_i^j\right)^2} \qquad (11)$$

where Equation (4) is used to equate the mean μ_i^j to the ODE model solutions and $\vec{\theta} = \begin{bmatrix} \vec{v} \\ \tau^1 \\ \vdots \\ \tau^m \end{bmatrix}$. The Gaussian model is very beneficial for fitting since even very poor initial guesses of the vector of parameters $\vec{\theta}$ would still produce nonzero probability density. The combined likelihood function is given by

$$L(\vec{\theta}) = C\left(\frac{1}{2\pi}\right)^{\sum_{j=1}^{m}\frac{n_j}{2}} (\tau^1)^{\frac{n_1}{2}} \dots (\tau^m)^{\frac{n_m}{2}} e^{-\frac{1}{2}\sum_{j=1}^{m}\tau^j\sum_{i=1}^{n_j}\left(d_i^j-\mu_i^j\right)^2}$$

$$= (\tau^1)^{\frac{n_1}{2}} \dots (\tau^m)^{\frac{n_m}{2}} e^{-\frac{1}{2}\sum_{j=1}^{m}\tau^j\sum_{i=1}^{n_j}\left(d_i^j-\mu_i^j\right)^2} \qquad (12)$$

where $C = \left(\frac{1}{2\pi}\right)^{-\sum_{j=1}^{m}\frac{n_j}{2}}$ simplifies the likelihood function. The value of $\vec{\theta}$ that maximizes $P(D|\vec{\theta})$ also maximizes $L(\vec{\theta})$.

E. Poisson Model for m Data sets and the Likelihood Function

Assume that for $j = 1,2,\dots m$, the j^{th} time series data is given by $D_j = \{d_1^j, d_2^j, \dots d_{n_j}\}$ with corresponding times $T_j = \{t_1^j, \dots t_{n_j}\}$ and the probability density of observing d_i^j is given by Equation (6).The probability density of the observed counts $D = \{D_1, D_2, \dots D_m\}$ given parameters $\vec{\theta}$ is given by

$$P(D|\vec{\theta}) = \prod_{j=1}^{m}\prod_{i=1}^{n_j}\frac{e^{-\lambda_i^j}\left(\lambda_i^j\right)^{d_i^j}}{d_i^j!}$$

$$= \frac{1}{d_1^1!..d_{n_1}^1!}\dots\frac{1}{d_1^m!..d_{n_m}^m!} e^{-\sum_{j=1}^{m}\sum_{i=1}^{n_j}\lambda_i^j\prod_{j=1}^{m}\left(\lambda_1^j\right)^{d_1^j}} \dots\left(\lambda_{n_j}^j\right)^{d_{n_j}^j} \qquad (13)$$

where Equation (4) is used to equate mean λ_i^j to the ODE model solutions and $\vec{\theta} = \vec{v}$. The combined likelihood function $L(\vec{\theta})$ is given by

$$L(\vec{\theta}) = C\frac{1}{d_1^1!\dots d_{n_1}^1!}\dots\frac{1}{d_1^m!\dots d_{n_m}^m!} e^{-\sum_{j=1}^{m}\sum_{i=1}^{n_j}\lambda_i^j\prod_{j=1}^{m}\left(\lambda_1^j\right)^{d_1^j}} \dots\left(\lambda_{n_j}^j\right)^{d_{n_j}^j}$$

$$= e^{-\sum_{j=1}^{m}\sum_{i=1}^{n_j}\lambda_i^j\prod_{j=1}^{m}\left(\lambda_1^j\right)^{d_1^j}} \dots\left(\lambda_{n_j}^j\right)^{d_{n_j}^j} \qquad (14)$$

where $C = d_1^1!\dots d_{n_1}^1!\dots d_1^m!\dots d_{n_m}^m!$ simplifies the Likelihood function.

F. Lognormal model for m data sets and the Likelihood Function

Assume that for $j = 1,2,\dots m$, the j^{th} time series data is given by $D_j = \{d_1^j, d_2^j, \dots d_{n_j}^j\}$ with corresponding times $T_j = \{t_1^j, \dots t_{n_j}^j\}$ and the probability density of observing d_i^j is given by Equation (8).The probability density of the observed counts $D = \{D_1, D_2, \dots D_m\}$ given parameters $\vec{\theta}$ is given by

$$P(D|\vec{\theta}) = \prod_{j=1}^{m}\prod_{i=1}^{n_j}\frac{1}{d_i^j\sigma_j\sqrt{2\pi}} e^{-\frac{\left(\ln\left(d_i^j\right)-\mu_i^j\right)^2}{2\sigma_j^2}}$$

$$= \left(\frac{1}{2\pi}\right)^{\sum_{j=1}^{m}\frac{n_j}{2}}\left(\frac{1}{d_i^j\sigma_j}\right)^{n_1} \dots\left(\frac{1}{d_{n_m}^m\sigma_m}\right)^{n_m} e^{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\frac{\left(\ln\left(d_i^j\right)-\mu_i^j\right)^2}{\sigma_j^2}} \qquad (15)$$

where Equation (4) is used to equate mean μ_i^j to the ODE model solutions and $\vec{\theta} = \begin{bmatrix} \vec{v} \\ \sigma_1 \\ \vdots \\ \sigma_m \end{bmatrix}$. Lognormal distribution is useful in fitting data when it is spread over a large range.

The combined likelihood function is given by

$$L(\vec{\theta}) = C\left(\frac{1}{2\pi}\right)^{\sum_{j=1}^{m}\frac{n_j}{2}}\left(\frac{1}{d_i^j\sigma_j}\right)^{n_1}\dots\left(\frac{1}{d_{n_m}^m\sigma_m}\right)^{n_m} e^{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\frac{\left(\ln\left(d_i^j\right)-\mu_i^j\right)^2}{\sigma_j^2}}$$

$$= \left(\frac{1}{d_i^j\sigma_j}\right)^{n_1}\dots\left(\frac{1}{d_{n_m}^m\sigma_m}\right)^{n_m} e^{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\frac{\left(\ln\left(d_i^j\right)-\mu_i^j\right)^2}{\sigma_j^2}} \qquad (16)$$

where $C = \left(\frac{1}{2\pi}\right)^{-\sum_{j=1}^{m}\frac{n_j}{2}}$ which is used to simplify the likelihood function.

IV. ESTIMATING THE BASIC REPRODUCTIVE RATE (R_t) IN REAL TIME (BAYESIAN APPROACH)

In case of any infectious disease, a useful parameter to consider is R_t, which is the Effective Reproductive Number or the number of people who get infected per infectious person at time t. The most well-known version of this quantity is R_0, which is the basic Reproduction. As pandemic progresses, increasing or decreasing restrictions reduce or increase R_t respectively. If $R_t \gg 1$, then the pandemic will infect a large proportion of the population, if $R_t < 1$, then the pandemic will slow down quickly, before infecting many people. The lower the value of R_t, the more manageable is the pandemic.

Every day we learn how many more people have been infected by Covid 19. R_t today is related to R_{t-1} yesterday or R_{t-m} for any previous instance of time. By Bayes' Theorem,

$$P(R_t|K) = \frac{P(K|R_t)P(R_t)}{P(K)} \qquad (17)$$

This translates to the fact that having seen K new cases, we believe the distribution of R_t is equal to the likelihood of seeing K new cases given R_t times the prior beliefs of the value of $P(R_t)$ without data divided by the probability of seeing K cases in general. So, for a single day,

$$P(R_1|K_1) \propto P(K_1|R_1)P(R_1) \qquad (18)$$

The likelihood function is a function that estimates the probability of K new cases given a value of R_t. Since we are trying to model the number of new arrivals over time, we choose the Poisson Distribution. Given an average λ number of cases per day,

$$P(K|\lambda) = \frac{e^{-\lambda}\lambda^K}{K!} \qquad (19)$$

The visualizations of the likelihood functions with some dummy data are given in Fig. 1 and Fig. 2 .

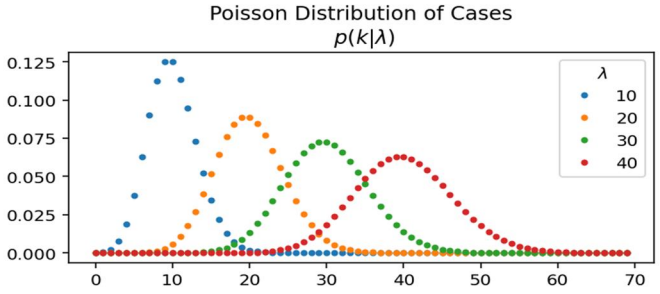


Fig. 1 The likelihood function for a series of values of K for varying λ

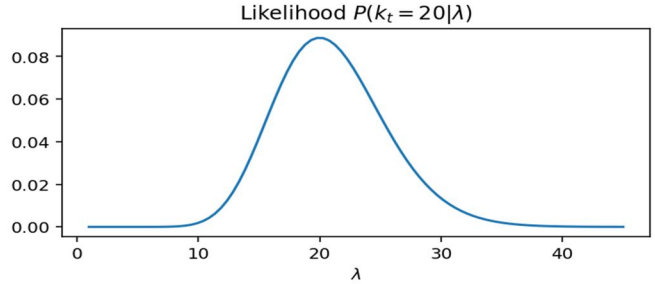


Fig. 2 Likeliness of each λ for $K = 20$

If we use the posterior probability of the previous period $P(R_{t-1}|K_{t-1})$ as the prior, $P(R_t)$ for the current period is given by

$$P(R_t|K_t) \propto P(R_{t-1}|K_{t-1})P(K_t|R_t) \qquad (20)$$

Iterating across all periods back till $t = 0$, we get

$$P(R_t|K_t) \propto P(R_0)\prod_{t=0}^{T} P(K_t|R_t) \qquad (21)$$

where $P(K|R)$ is the likelihood function. With a uniform prior $P(R_0)$, for simplicity, this reduces to

$$P(R_t|K_t) \propto \prod_{t=0}^{T} P(K_t|R_t) \qquad (22)$$

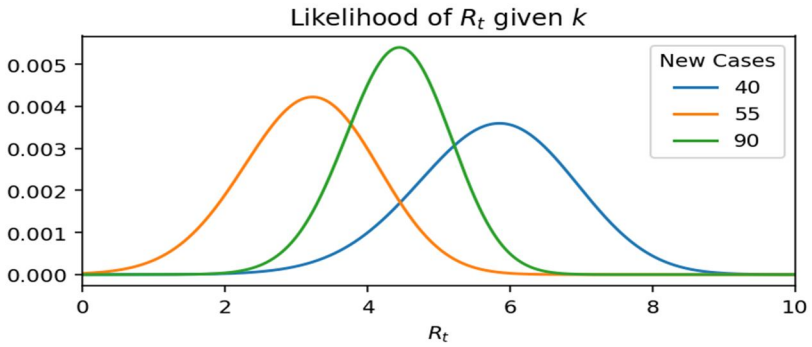


Fig. 3 The Likelihood of R_t given K

As we can see from the figure, for each day, we have independent guesses for R_t. Our goal is to combine the information we have about previous days with the current day. To perform the Bayesian update, we multiply the likelihood by the prior, which is just the previous day's likelihood(ignoring our Gaussian update for simplicity).

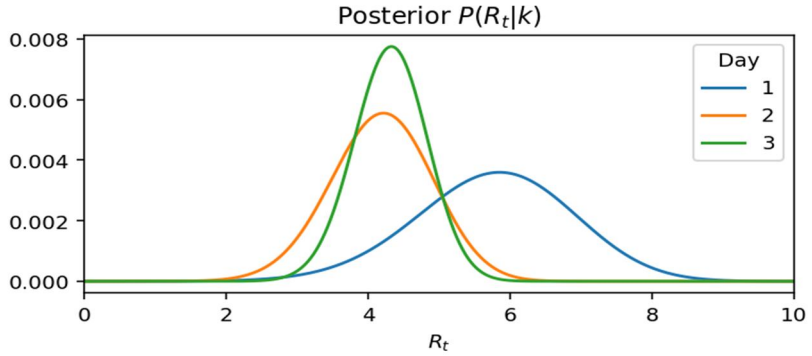


Fig. 4 For Day 1, our posterior matches Day 1's likelihood from above, as we have no prior information on this day. But when we update the prior using Day 2's information, the curve for Day 2 shifts left but not as much as above.

This is because Bayesian update uses both days' information and effectively averages out the two.

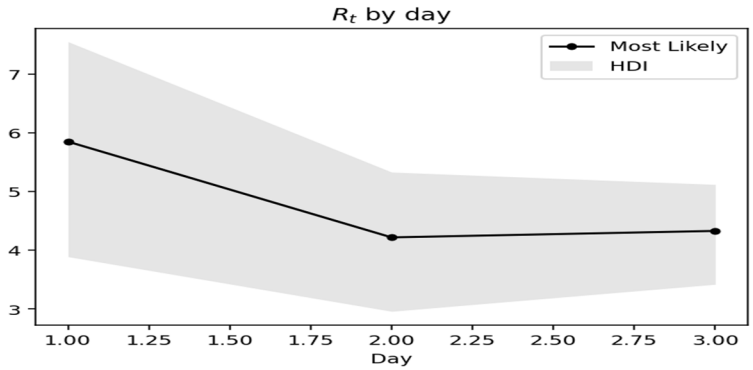


Fig. 5 As we are estimating the parameter R_t, it is important that we see error surrounding this estimation. The HDI(Highest Density Interval), indicates which points of the distribution are most credible and which cover most of the distribution.

The HDI summarises the distribution by specifying an interval that spans most of the distribution such that any point inside the interval has a higher credibility than any point outside the interval. By performing simulations in a brute force way, we obtain the most likely values of R_t surrounded by the HDI. Now, we will apply all this to the US COVID-19 Data.

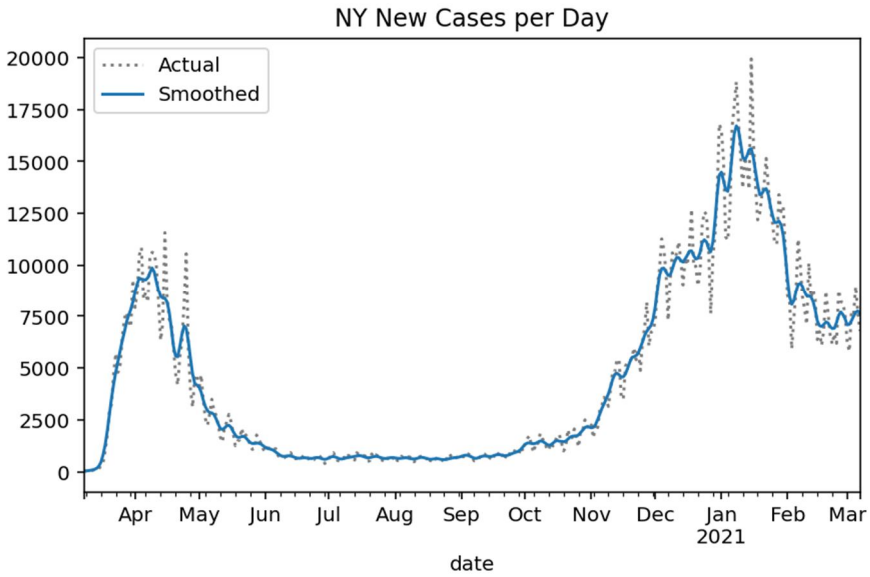


Fig. 6 New York cases per day: Actual vs Smoothed

The approach till now simply uses yesterday's posterior as today's prior. While this approach is very intuitive, it doesn't allow for the fact that the value of R_t has likely changed from that of yesterday. To allow for that change, we apply Gaussian Noise with some standard deviation σ. We are using maximum likelihood for our process of choosing the σ: We choose σ such that $P(K|\sigma)$ is maximised. Since σ is a fixed value, our approach would be to maximise K over all values of σ. Since, $P(K) = P(K_0 \cap K_1 \cap \ldots \cap K_t)$, we need to find $P(K_t)$ and this is given by the denominator from Bayes' Rule

$$P(R_t|K_t) = \frac{P(K_t|R_t)P(R_t)}{P(K_t)} \qquad (23)$$

and also, $P(K_t|R_t)P(R_t) = P(K_t \cap R_t)$, the joint distribution of K_t and R_t. We now marginalise K_t over R_t to get

$$P(K_t) = \sum_{R_t} P(K_t|R_t)P(R_t) \qquad (24)$$

Since we are looking for σ that maximizes $P(K)$, it is sufficient to maximize $\prod_{t,i} P(K_{t,i})$, where t is time and i is each state. The error can be minimised by taking the log of all the values and adding them, instead of just multiplying them, as the values are small, Hence, maximizing this log likelihood will automatically maximise the likelihood, where,

$$P(K_t|R_t) = \frac{\lambda^{k_t} e^{-\lambda}}{k_t!} \text{ and}$$

$$P(R_t) = P(R_{t-1}|K_{t-1}) + \epsilon, \epsilon \sim N(0, \sigma^2) \qquad (25)$$

and $P(K_t) = \sum_{R_t} P(K_t|R_t)\{P(R_{t-1}|K_{t-1}) + \epsilon\}$ and we will be maximizing

$$\prod_{t,i} P(K_{t,i}) \qquad (26)$$

To implement this procedure, we use the following steps:

- Calculate the value of λ, the expected arrival rate for each day.
- Calculate each day's likelihood distribution over all R_t
- Calculate the process matrix for the σ mentioned above
- We choose our initial prior as Gamma with mean 7 based on information supplied by the CDC.
- Repeat from Day 1 to the end, calculate the prior by applying Gaussian Noise to yesterday's prior, apply Bayes' rule and complete by dividing with the probability of the data.

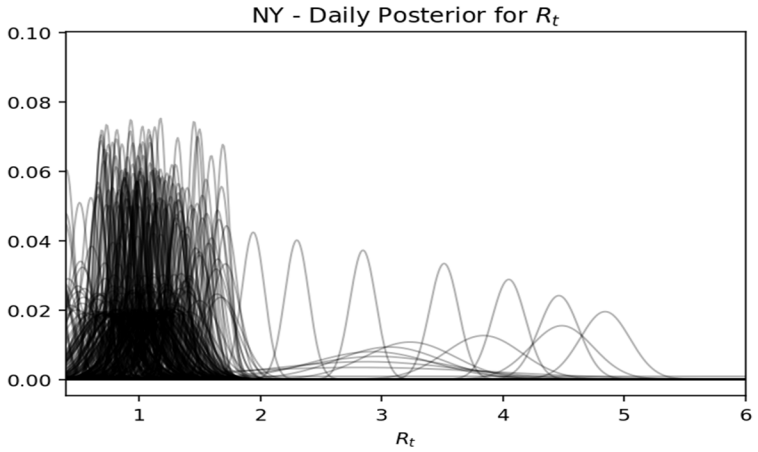


Fig. 7 Every posterior distribution corresponding to every single day corresponding to the state of New York is plotted on the figure. The posteriors start with low confidence and progressively becomes narrower(more confident) near the true value of R_t.

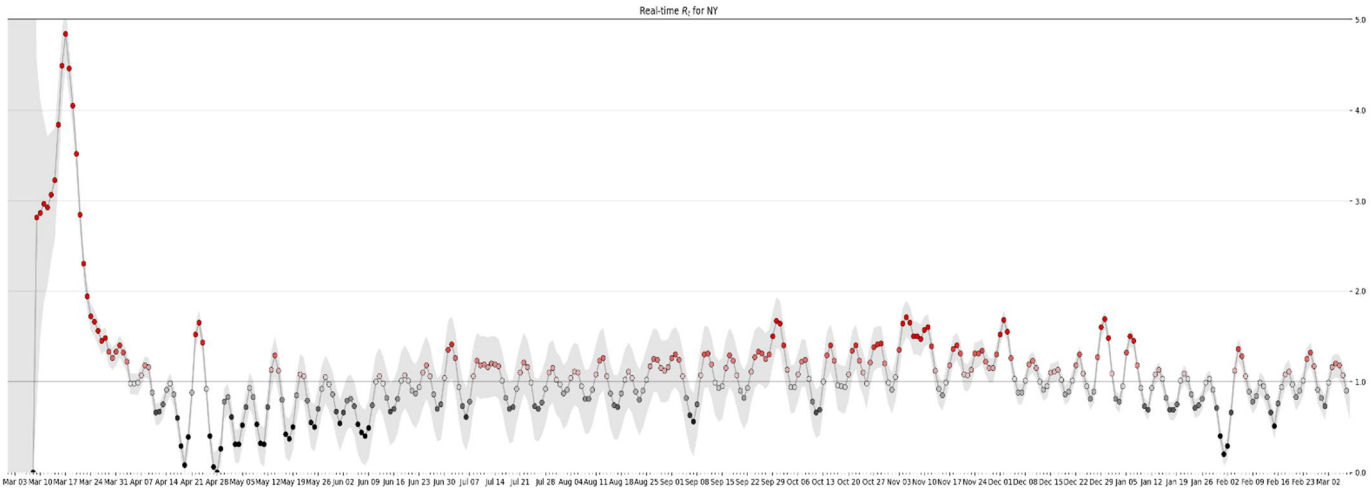


Fig. 8 Estimating the effective Reproductive Rate R_t in real time (Bayesian Approach)

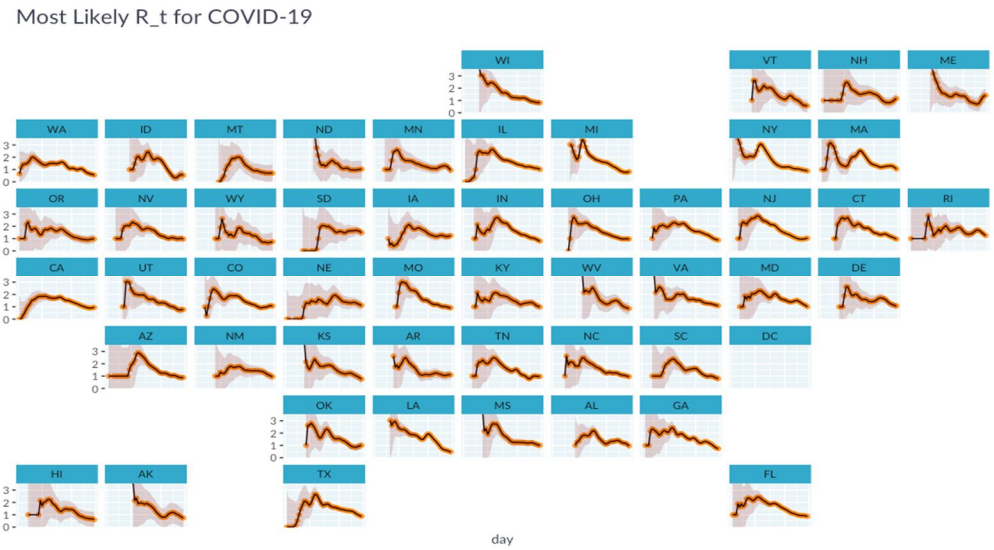


Fig. 9 Real time predicted values of R_t, with the Highest Density Interval (HDI) for every state in the US.

V. SIR MODEL

The origin of compartmental models, such as the SIR model, originated with the works of Kermack and McKendrik in 1927. The SIR Model is one of the simplest compartmental models and consists of 3 compartments:

- S: The number of Susceptible Individuals. When a susceptible and an infectious individual come into infectious} contact, the susceptible individual contracts the disease and transitions into the infectious compartment
- I: The number of Infected Individuals. The individuals are infected and are capable of spreading the infection to a susceptible group
- R: The number of deceased or recovered individuals. These individuals have either been infected and completely recovered and entered the recovered compartment, or have died

The dynamics of an epidemic are often much faster than the dynamics of birth and death, therefore, birth and death are omitted in simple compartmental models, like the SIR Model. The SIR system without the vital dynamics can be expressed by the following set of Ordinary Differential Equations:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

Thus, $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \Rightarrow S(t) + I(t) + R(t) = N = C$ expressing mathematically the constancy of the population.

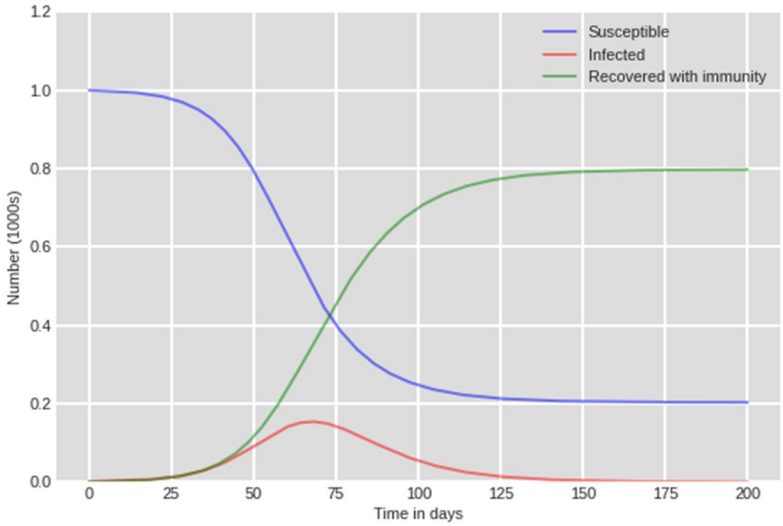


Fig. 10 A realization of the SIR Model, produced by Runge-Kutta numerical solution, with parameters $\lambda = 0.2, \frac{1}{\mu} = 10$ days, $N = 1000$ and $I(0) = 1$

Moreover, the dynamics of the infectious class depends on the basic reproduction number, defined earlier as $R_0 = \frac{\beta}{\gamma}$. If the reproduction number is too high, the probability of the pandemic is high too. If the reproduction number multiplied by the percentage of susceptible people is 1, it shows an equilibrium state and thus the number of infectious people is constant. Additionally, the recovery period is defined by $t_1 = \frac{1}{\gamma}$ and this describes the average days to recover from infection. The transmission period in the sense of the average days to transmit the infection to an un-infected person is defined to be $t_2 = \frac{1}{\beta}$. However, in a population with vital dynamics, new births can provide more susceptible individuals to a population, which may sustain the epidemic or allow new introductions to spread in the population. Taking the above vital dynamics into consideration, the SIR model can be re-formulated as

$$\frac{dS}{dt} = \mu N - \frac{\beta IS}{N} - \nu S$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \nu I$$
$$\frac{dR}{dt} = \gamma I - \nu R$$

where μ denotes the birth rate and ν denotes the death rate. To maintain a constant population, we assume μ = ν is the natural mortality rate. We assume that disease spreads from an infected person I to a susceptible person S at rate β and that an infected person I becomes a recovered person R at rate γ, i.e.

$$S + I \xrightarrow{\beta} I + I$$
$$I \xrightarrow{\gamma} R$$

VI. BAYESIAN INVERSION APPROACH FOR MODEL PARAMETER ESTIMATION

We propose Bayesian inversion methods, in which probabilities are used as a general concept to represent the uncertainty in the model parameters in order to solve the backward/inverse problem of COVID-19, i.e., the problem of accurate estimation of the epidemiological model parameters. Bayesian inference in the context of the statistical inversion theory is based on Bayes' Theorem and quantifies the uncertainty involved in the model parameters by defining a probability distribution over the possible values of the parameters and uses sample data to update this distribution. Bayesian analysis, in contrast to traditional inverse methods, is a robust inversion technique for determining parameters, yields the (a posteriori) probability distribution, and has the advantage of updating the prior knowledge about the unknown quantity using the measurement/observation data, giving confidence intervals for the unknowns instead of providing a single estimate.

As the observations or measurements are subject to noise, and the observational noise i.e., error ϵ due to modelling and measurement is unbiased and $i.i.d$, therefore it can be represented by a random variable as

$$Y = f(Q) + \epsilon \qquad (27)$$

where ϵ is a zero-mean random variable and Y is a given random variable representing observed data for which we have a model $f(Q)$ dependent on a random variable Q with realizations $q = Q(\omega)$ representing the parameters to be estimated. We now assume a probability space $(\omega, \mathcal{F}, P)$, where ω is the sample space, $\mathcal{F}$ is the $\sigma-$ algebra of events and P is a probability measure. Furthermore, we assume all random variables to be absolutely continuous. Now, Bayes' theorem in terms of probability densities can be written as

$$\pi(q|y) = \frac{\pi_0(q)\pi(y|q)}{\pi(y)} \qquad (28)$$

where,

$$\pi(y) = \int_{R^p} \pi_0(q)\pi(y|q)\,dq \qquad (29)$$

where the unknown parameters $(q_1, q_2, \ldots q_p) \in R^p$ is a realization of the random variable Q and the observed data y is the realization of the random variable Y.Also, $\pi_0(q)$ is the p.d.f of the prior, $\pi(q|y)$ is the $p.d.f$ of the posterior and $\pi(y|q)$ is the likelihood. $\pi(q|y)$ gives the posterior density and requires evaluation of the above integral.

VII. MARKOV CHAIN MONTE-CARLO METHODS

The Markov-Chain Monte-Carlo methods are a class of Monte-Carlo methods with the general idea of constructing Markov-Chains whose stationary distribution is the posterior density. The Metropolis-Hasting Algorithm is an MCMC algorithm to draw samples from a desired distribution by building a Markov-Chain of accepted values (out of proposed values) for the unknown parameter as a posteriori distribution. In this algorithm, the first state of the chain q_0 is given and the new state, q_k, $k = 1,2,\ldots N$ of the chain is constructed based on the previous state of the chain q_{k-1}. A new value q^* is proposed using the proposal density function $J(q^*|q_{k-1}) = N(q_{k-1}, \sigma_p^2)$ where σ_p is the proposal covariance. Admissibility of this proposed value is tested by means of calculating the acceptance ratio $\alpha(q^*|q_{k-1})$, defined by

$$\alpha(q^*|q_{k-1}) = min\left(1, \frac{\pi(q^*|y)}{\pi(q_{k-1}|y)} \cdot \frac{J(q_{k-1}|q^*)}{J(q^*|q_{k-1})}\right) \qquad (30)$$

where $\pi(q|y)$ is the posterior distribution and J is the proposal distribution. Applying Bayesian Inversion Approach for model parameter selection, we get α as

$$\alpha(q^*|q_{k-1}) = min\left(1, \frac{\pi(y|q^*)\pi_0(q^*)}{\pi(y|q_{k-1})\pi_0(q_{k-1})} \cdot \frac{J(q_{k-1}|q^*)}{J(q^*|q_{k-1})}\right) \qquad (31)$$

where $J(q_{k-1}|q^*) = J(q^*|q_{k-1})$ for symmetric proposal functions and $\pi_0(q)$ is a given prior distribution. Furthermore, $\pi(y|q)$ is the likelihood distribution, defined by

$$\pi(y|q) = N(y, \sigma_L^2) = \frac{1}{(2\pi\sigma_L^2)^{\frac{n}{2}}} e^{-S_q/2\sigma_L^2} \qquad (32)$$

where σ_L is the likelihood covariance, $S_q = \sum_{i=1}^{n}(y_i - f_i(q))^2$ is the sum of squares error and $f(q)$ denotes the parameter dependent model response. If the proposed value is admissible, it is accepted as q_k, otherwise the old value is kept.

VIII. DELAYED-REJECTION ADAPTIVE-METROPOLIS ALGORITHM (DRAM)

Although searching for a good proposal value can be done manually through trial and error, the problem becomes intractable in higher dimensions of the parameter space. Thus, adaptive algorithms that find optimal proposal scales automatically are advantageous. The Delayed-Rejection-Adaptive-Metropolis(DRAM) is an efficient adaptive MCMC algorithm. It is based on the idea of two powerful ideas to modify the MCMC method - Adaptive Metropolis (AM) and Delayed-Rejection(DR), which are used as global and local adaptive algorithms respectively. AM finds an optimal proposal scale and updates the proposal covariance matrix, while DR updates the proposal value when q^* is rejected. The basic idea of the DR algorithm is that, if the proposal q^* is rejected, delayed rejection provides an alternative candidate q^{**} as a second stage move rather that just retaining the previous value q_{k-1}. This step can be done for one or many stages. Furthermore, the acceptance probability of the new candidates is also calculated. Therefore, in the DR process, the previous state of the chain is updated using the optimal parameter scale or proposal covariance matrix that has been calculated via the AM algorithm. The AM algorithm is a global adaptive strategy, where a recursive relation is used to update the proposal covariance matrix. In this algorithm, we take the Gaussian Proposal centred at the current state of the chain q_k and update the chain covariance matrix at the k^{th} step using

$$V_k = s_p \mathrm{Cov}(q_0, q_1, \ldots, q_{k-1}) + \epsilon I_p \qquad (33)$$

where s_p is a design parameter and depends only on the dimension p of the parameter space. From historical references and text, this parameter $s_p = 2.38^2/p$ as the common choice for Gaussian targets and proposals, as it optimizes the mixing properties of the Metropolis-Hastings search in the case of Gaussians. I_p the p-dimensional identity matrix and $\epsilon \geq 0$ is a very small constant to ensure that V_k is not singular. In most cases, ϵ can be set to zero.
he adaptive metropolis algorithm employs the recursive relation

$$V_{k+1} = \frac{k-1}{k} V_k + \frac{s_p}{k}\left(k\overline{q}_{k-1}\overline{q}_{k-1}^T - (k+1)\overline{q}_k\overline{q}_k^T + q_k q_k^T\right) \qquad (34)$$

to update the proposal covariance matrix, where the sample mean $\overline{q}_k$ is calculated recursively by

$$\overline{q}_k = q_k + \frac{k}{k+1}(\overline{q}_{k-1} - q_k) \qquad (35)$$

A second stage candidate q^{**} is chosen using the proposal function

$$J_2(q^{**}|q_{k-1}, q6*) = N(q_{k-1}, \gamma_2^2 V_k) \qquad (36)$$

where V_k is the covariance matrix produced by the adaptive algorithm (AM) as the covariance of the first stage and $\gamma_2 < 1$ is a constant. The probability of accepting the second stage candidate, having started at q_{k-1} and rejected q_* is

$$\alpha_2(q^{**}|q_{k-1}, q^*) = min\left(1, \frac{\pi(q^{**}|y)J(q^*|q^{**})(1-\alpha(q^*|q^{**}))}{\pi(q_{k-1}|y)J(q^*|q_{k-1})(1-\alpha(q^*|q_{k-1}))}\right) \qquad (37)$$

where α is the acceptance probability in the non-adaptive approach. The acceptance probability is computed so that reversibility of the posterior Markov Chain is preserved.

IX. APPLICATION TO THE REAL-WORLD DATA (US)

The United States COVID-19 dataset has been used to perform all simulations and analysis. The dataset is available at https://covidtracking.com/api/v1/states/daily.csv . MCMC simulations were performed using this dataset and the densities of the SIR model parameters corresponding to three states of the US are demonstrated below:

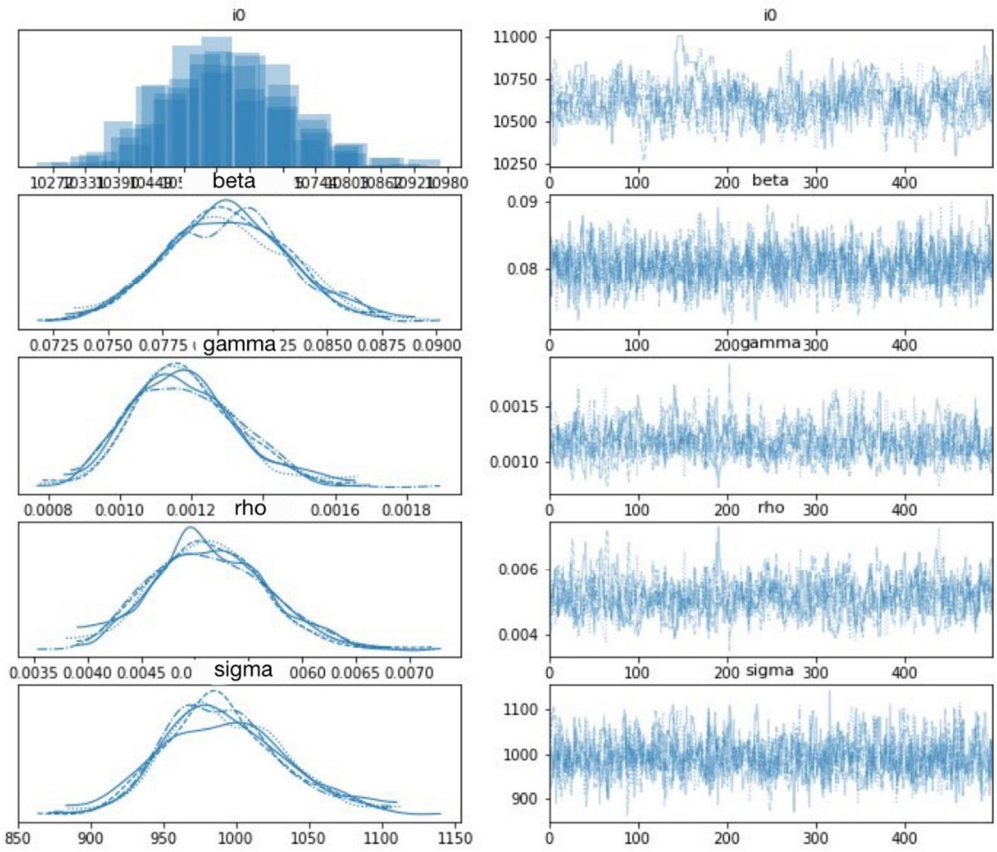


Fig. 11 The densities and the respective trace plots for the parameters of the SIR model corresponding to the state of Georgia

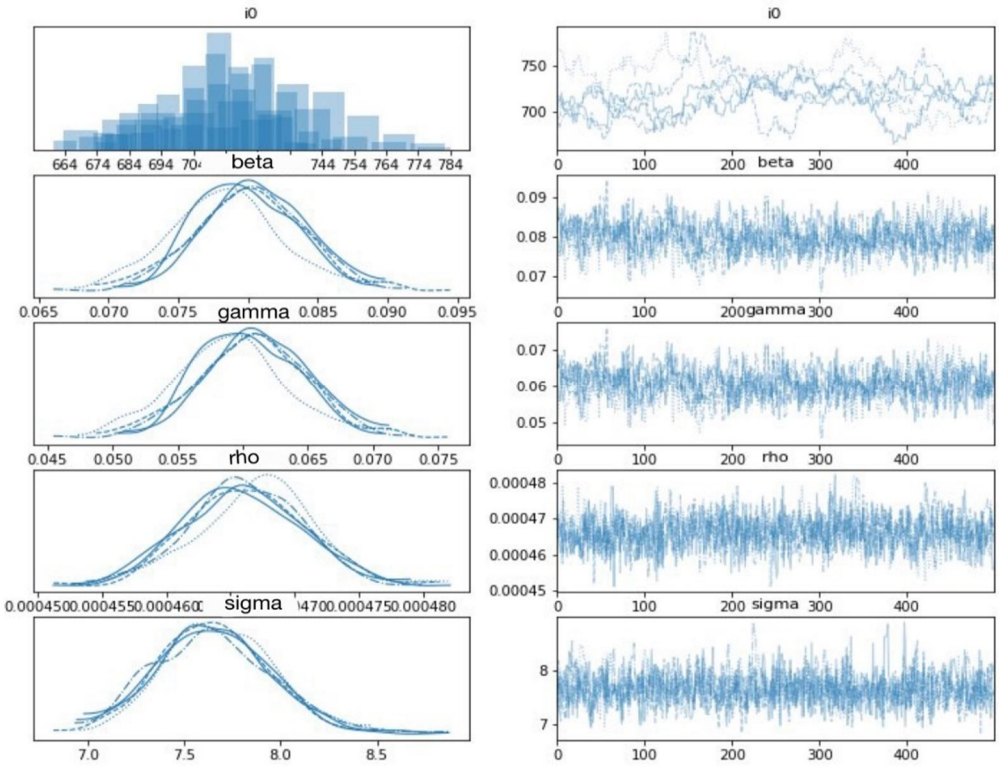


Fig. 12 The densities and the respective trace plots for the parameters of the SIR model corresponding to the state of Arkansas

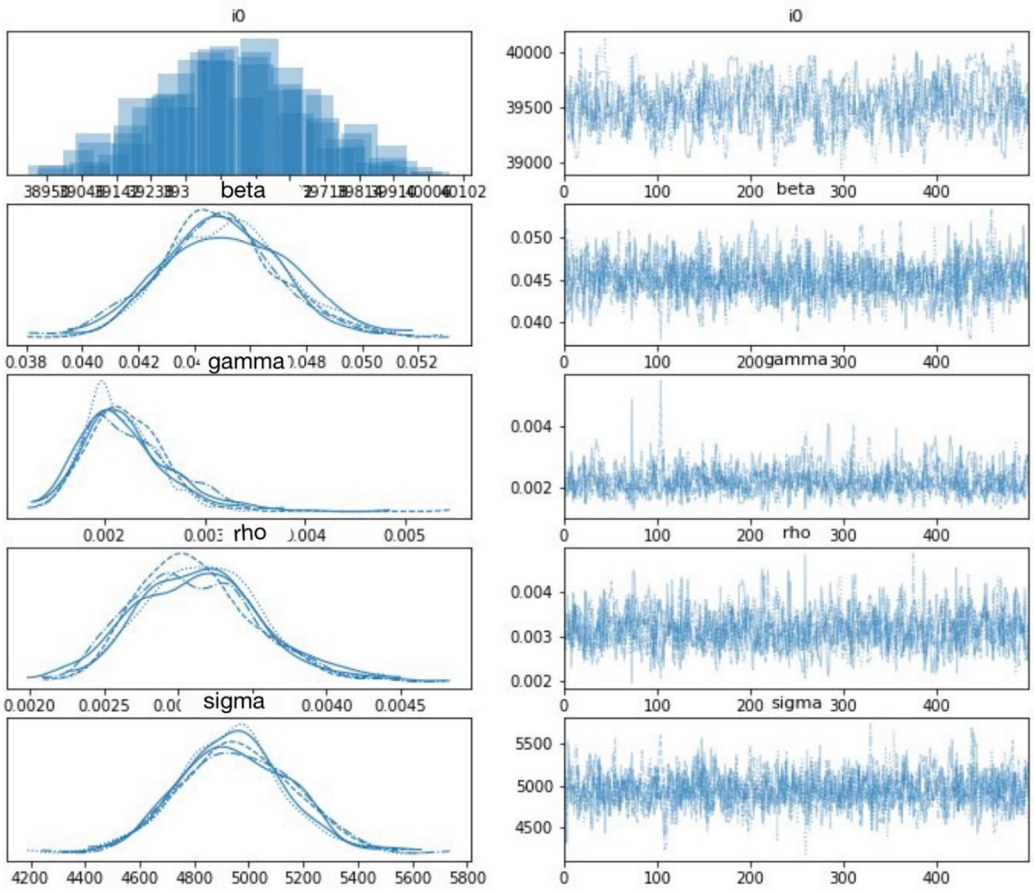


Fig. 13The densities and the respective trace plots for the parameters of the SIR model corresponding to the state of California

X. TRANSMISSION RATE β

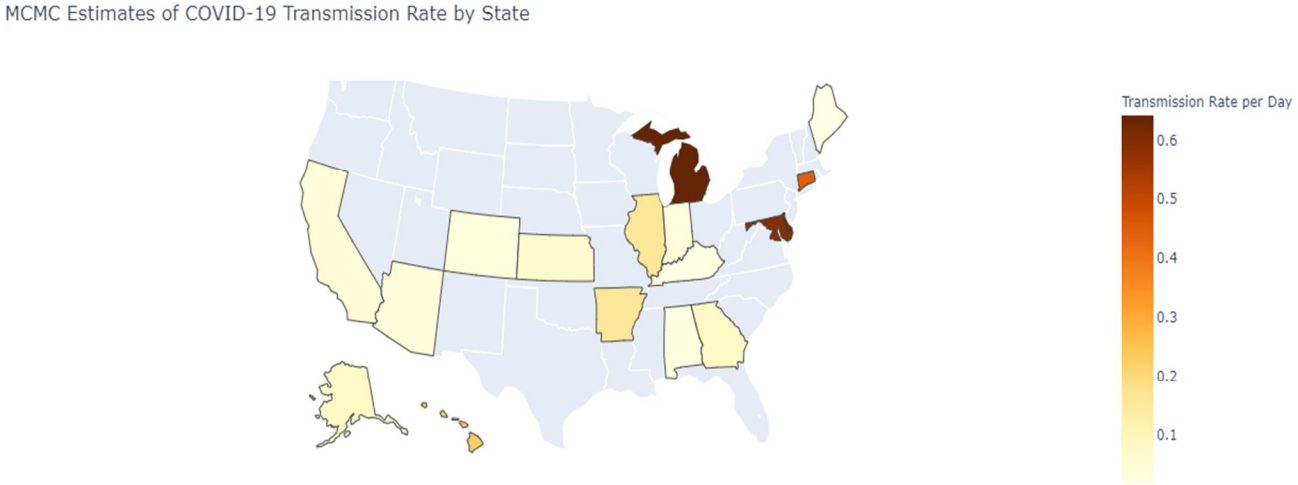


Fig. 14 MCMC Estimates of the COVID-19 Transmission Rate β of some of the states in the US: Highest is Michigan with $\beta \approx 0.64$

XI. RECOVERY RATE γ

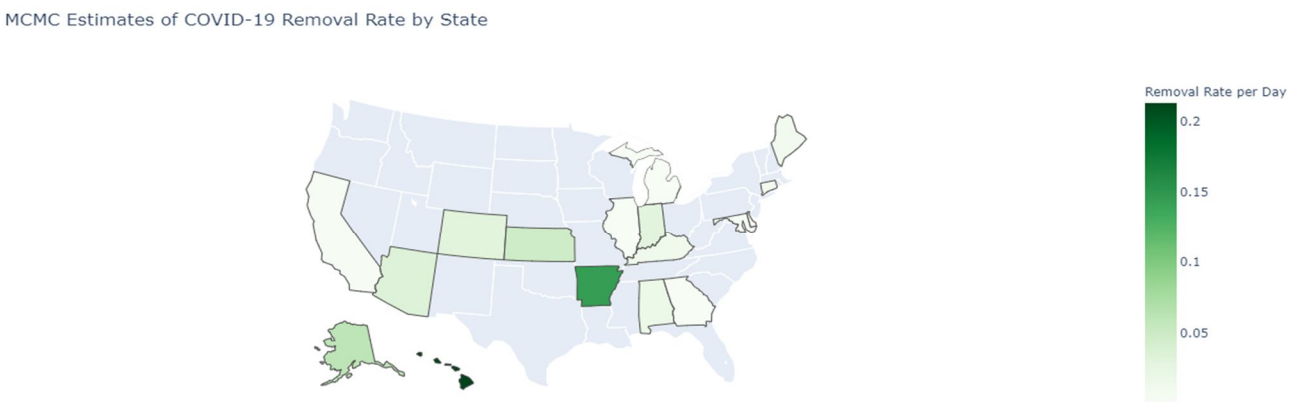


Fig. 15 MCMC Estimates of the COVID-19 Removal Rate γ of some of the states in the US: Highest is Hawaii with $\gamma \approx 0.21$

XII. CONCLUDING REMARKS

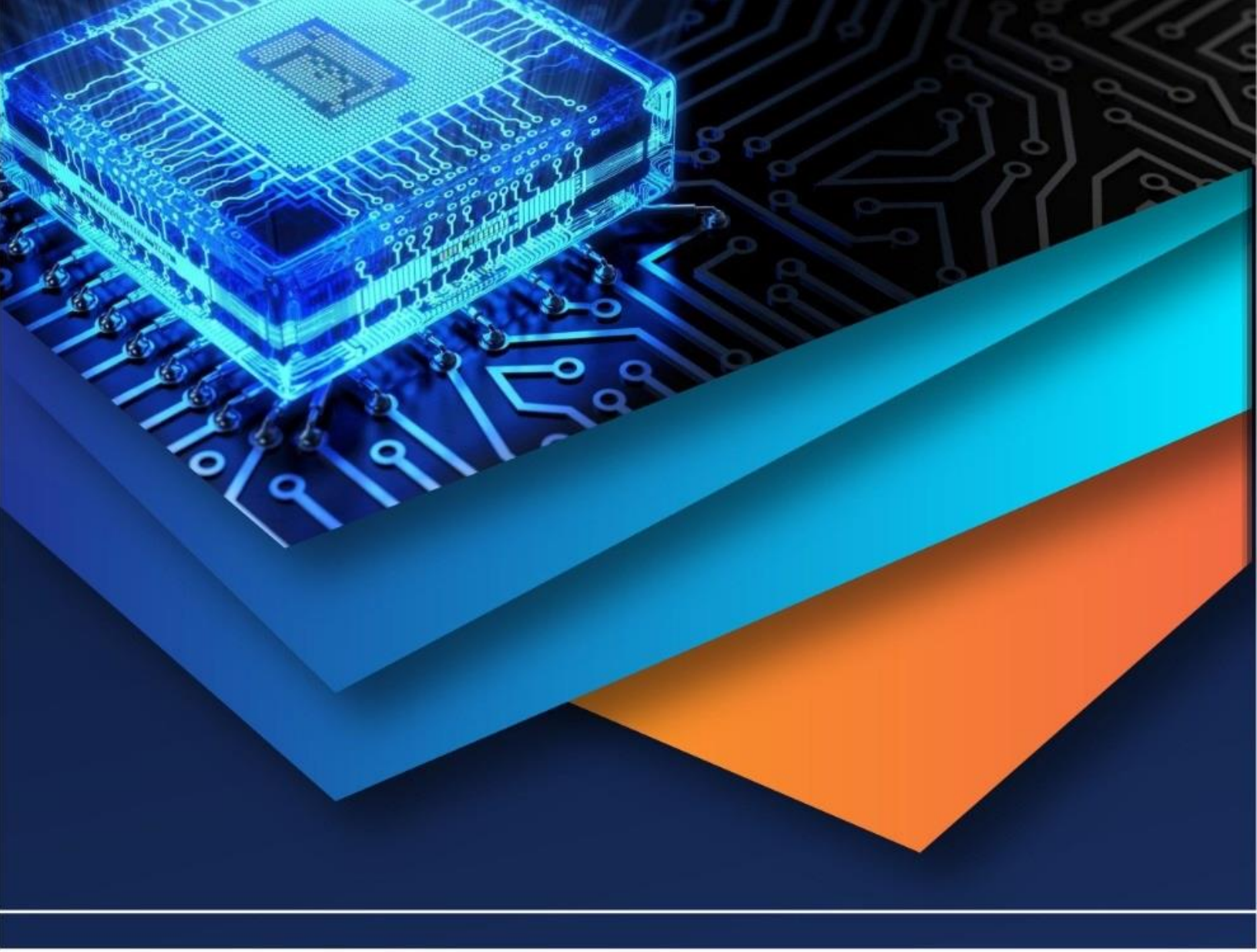
In this paper, I have tried to demonstrate how the deterministic modelling of the COVID-19 is performed based on real time data. I have used Python 3.8 with the package PyMC3 to carry out the simulations performed during the course of the project. I chose the MCMC algorithm to implement the statistical estimation and prediction because of the consideration on the prediction uncertainly. The spread dynamics of the COVID-19 virus is significantly complex and potential inaccuracy and incompleteness have also crept in in the process of collecting and storing the data. The data has been collected from the open-source project https://covidtracking.com/ . The aim of this paper was to demonstrate how the parameters of the SIR Model are distributed with respect to data obtained from a region. Similar densities and values are used when statisticians are modelling the spread of an infectious disease using the SIR Model. This paper also includes a brief Bayesian Analysis on the Basic Reproductive Rate R_t for the COVID-19 virus.

However, a thing to note is that high dimension of the parameter space might affect the convergence of Markov Chain to the posterior density. But we can always construct more efficient algorithms. Although this type of methods somehow takes us away from our original purpose which was to improve upon an existing algorithm, they still make sense within this survey in that they allow for almost automated implementations.

Based on the availability of computational power, several other Monte Carlo Techniques like the Hamiltonian Monte Carlo (HMC), scalable MCMC methods, where algorithms manage to handle large scale targets by breaking the problem into manageable or scalable pieces(divide and conquer & sub-sampling), parallelization schemes etc. can be applied to obtain better results.

BIBLIOGRAPHY

[1] Bayesian Inference for Dynamical Systems - Weston C. Roda, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada
[2] A Mathematical Model of Epidemics—A Tutorial for Students - Yutaka Okabe and Akira Shudo, Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397
[3] Bayesian Inference for Stochastic Epidemic Models using Markov chain Monte Carlo Methods - Nicholas Demiris, The University of Nottingham
[4] The Covid-19 outbreak in Spain. A simple dynamics model, some lessons, and a theoretical framework for control response - Antonio Guirao, Department of Physics, Universidad de Murcia, Ed. CIOyN, Campus de Espinardo, 30100, Murcia, Spain
[5] Uncertainty quantification in epidemiological models for the COVID-19 pandemic - Leila Taghizadeh, Ahmad Karimi, Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8–10, 1040, Vienna, Austria, Clemens Heitzinger, School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, 85287, USA
[6] A simple model for COVID-19 - Julien Arino, Department of Mathematics & Data Science NEXUS, University of Manitoba, Canada, Canadian COVID-19 Mathematical Modelling Task Force, Canada, Stephanie Portet, Department of Mathematics \& Data Science NEXUS, University of Manitoba, Canada



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY