Stochastic Processes Report

Modelling and predicting Covid19 infections using the SIR model and MCMC sampling

Data Scraping:

The data was scraped from WorldoMeters, which keeps a copy of the data published by the Ministry of Health and Family Welfare. The data is a time series, starting from Feb 15, 2020, to March 22, 2023. The data consists of the Total Daily Cases, New Cases, Active Cases, Total Deaths, Daily Deaths, and New Recoveries. From there, the data used is the Active Cases (which is I), and the cumulative sum of the New Recoveries (which is R).

The SIR Model:

The SIR model is a compartmental model consisting of:

- 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 to the infectious compartment.
- I: The number of infectious individuals. These are individuals who have been infected and are capable of infecting susceptible individuals.
- R: The number of removed (and immune) or deceased individuals. These are individuals who have been infected and have either recovered from the disease and entered the removed compartment or died. It is assumed that the number of deaths is negligible with respect to the total population.

The SIR system without vital dynamics (birth and death) can be described above can be expressed by the following system of ordinary differential equations:

$$\frac{\mathrm{dS}}{\mathrm{d}t} = -\frac{\beta I S}{N} \tag{1}$$

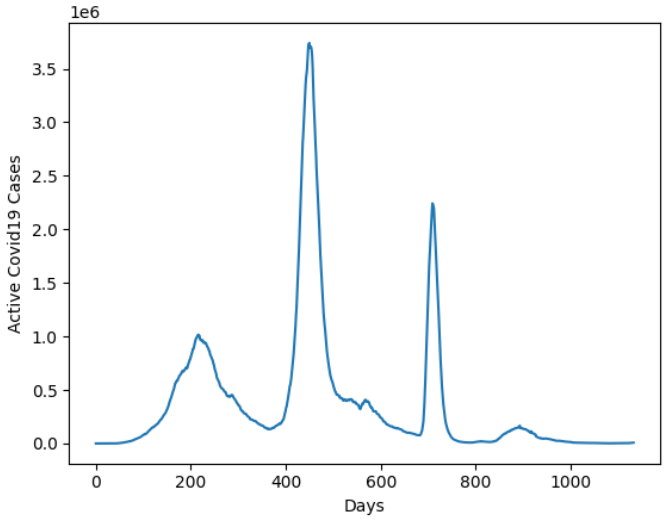
$$\frac{\mathrm{dI}}{\mathrm{d}t} = \frac{\beta I S}{N} - \gamma I \tag{2}$$

$$\frac{\mathrm{dR}}{\mathrm{d}t} = \gamma I \tag{3}$$

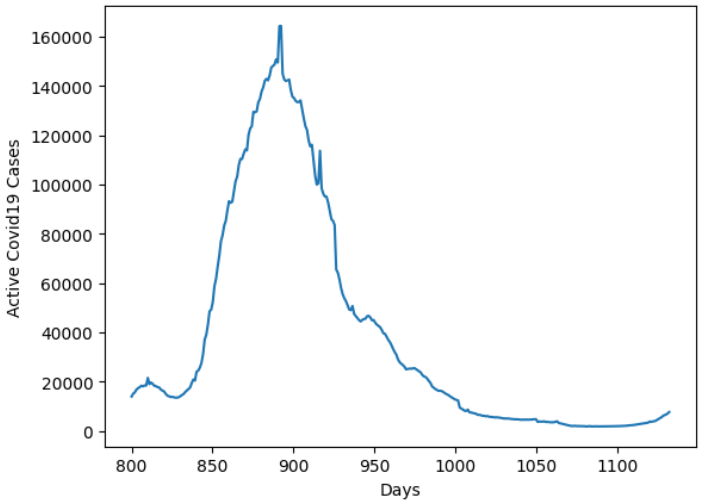
$$S(t) + I(t) + R(t) = N \tag{4}$$

$$R_0 = \frac{\beta}{\gamma} \tag{5}$$

Here, R_0 is the basic reproduction number, the expected number of cases generated by one case. In our model, the N = 1366417754, which was the population of India in 2019. We aim to determine the values of β and γ, given the data and some initial states, using a No U-Turn Sampler (NUTS).



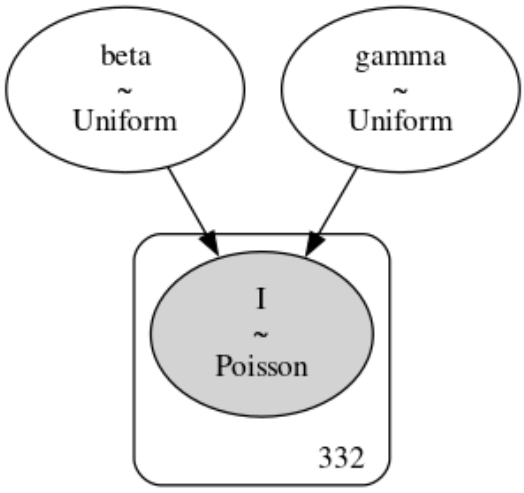
(a) Active Cases v/s Days for all waves



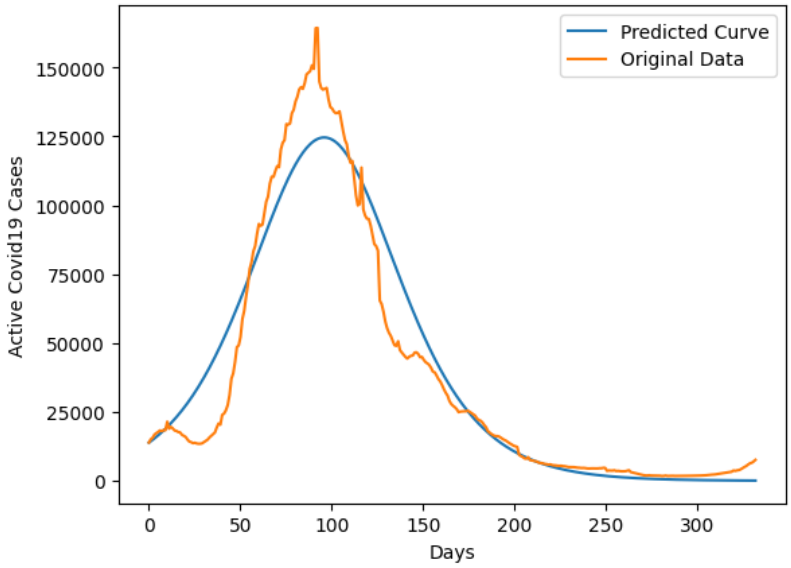
(b) Active Cases v/s Days for 3rd wave

Calculating β and γ:

The data collected is for a large period of time, and the three major covid waves are visible in it. The SIR model is meant for only a single wave, so we clip the first and second waves and only consider data gathered from day number 800 (From the start).
Similarly, the data for S and R is also truncated. This is then used to make a hierarchical bayesian model, with the number of active cases being assumed to be a Poisson process, with means being dependent on the ODE, which in turn is dependent on beta and gamma, which are sampled from a uniform prior distributions, with hyperparameters being selected based on empirical data.



(a) Bayesian Model



(b) Predicted Active Cases Curve

After creating the model, we ran the NUTS sampler and calculated posterior distributions for beta and gamma. The mean of beta and gamma was 2.773 and 2.652, respectively. Using these values, we can calculate R_0, which is 1.0456. Using this, we predict the number of active cases on March 23, 2023 (our data used to make the model was till March 22, 2023) to be 7952. Based on the actual data, it is found to be 7927. Thus, our estimate is very close to the actual data.