SIR Model for Epidemic Disease Prediction

Frank Zhong

July 2024

1 Abstract

This article is aimed to provide a brief introduction to the Susceptible-Infected-Recovered (SIR) model, its characteristics, the model construction in R (a programming language for statistical computing and data visualization) and its limitation on epidemic prediction.

2 Introduction

The SIR model, introduced by Kermack and Mckendrick in the early twentieth century, is a three-compartment mathematical model describing the spread of a novel pathogen through a population [1].
Within this model, the entire population is characterised into three classes: the susceptible S, infected I, and recovered R. The recovered population are assumed to be no longer able to get infected again, (i.e. be immune, vaccinated, isolated or even died from the disease). Additionally, we assume that infected patients leave the I class with a constant rate of γ, and contact rate to be β. This model is governed by three ordinary differential equations (O.D.E.s):

$$\frac{dS}{dt} = -\beta\frac{S}{N}I,\ \frac{dI}{dt} = \beta\frac{S}{N}I - \gamma I,\ \frac{dR}{dt} = \gamma I \tag{1}$$

with the constant population constraint $S + I + R = N$.
For convenience, we nondimensionalize using N for population size and γ^{-1} for time; that is, let:

$$\hat{S} = S/N, \quad \hat{I} = I/N, \quad \hat{R} = R/N, \quad \hat{t} = \gamma t$$

and define the dimensionless basic reproductive ratio as:

$$\mathcal{R}_0 = \frac{\beta N}{\gamma}$$

Therefore, the dimensionless SIR equations are then given by:

$$\frac{d\hat{S}}{d\hat{t}} = -R_0\hat{S}\hat{I},\ \frac{d\hat{I}}{d\hat{t}} = R_0\hat{S}\hat{I} - \hat{I},\ \frac{d\hat{R}}{d\hat{t}} = \hat{I} \tag{2}$$

with dimensionless constraint $\hat{S} + \hat{I} + \hat{R} = 1$.

3 Methods

Based on SIR model, this article focuses on two questions:
(1) Under what condition does an epidemic occur?
(2) If an epidemic occurs, how much proportion of subjects within a well-mixed population are susceptible, infected or recovered with respect to time?

3.1 Epidemic Occurrence Condition

An epidemic occurs when a small number of infections introduced into a susceptible population results in an increasing number of infections. We can assume an initial population at a fixed point, perturb this fixed point by introducing a small number of infections, and determine the fixed point's stability. An epidemic occurs when the fixed point is unstable [2]. The linear stability problem may be solved by considering only the equation for $\frac{d\hat{I}}{d\hat{t}}$. With $\hat{I} \ll 1$ and $\hat{S} \approx \hat{S}_0$, we obtain:

$$\frac{d\hat{I}}{d\hat{t}} = \left(\mathcal{R}_0 \hat{S}_0 - 1\right) \hat{I}$$

so that an epidemic occurs if $\mathcal{R}_0 \hat{S}_0 - 1 > 0$. With the basic reproductive ratio $\mathcal{R}_0$ and $\hat{S}_0 = \frac{S_0}{N}$, where S_0 is the number of initial susceptible individuals, an epidemic occurs if:

$$\mathcal{R}_0 \hat{S}_0 = \frac{\beta S_0}{\gamma} > 1 \tag{3}$$

which could be anticipated. An epidemic within a population of S_0 susceptible individuals occurs if an infected individual spread the disease to more than one other subject on average. In addition, if an epidemic occurs, then initially the number of infected individuals grows exponentially with growth rate $\beta S_0 - \gamma$.

3.2 Susceptible, infected, recovered population over time

Given the original set of O.D.E.s Eq.(1) from the simplified SIR model:

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

We are able to re-parameterise the O.D.E.s with normalised variables [3]:

$$\dot{s} = \frac{ds}{dt} = -si \tag{4.1}$$

$$\dot{i} = \frac{di}{dt} = si - gi \tag{4.2}$$

$$\dot{r} = \frac{dr}{dt} = gi \tag{4.3}$$

subject to normalization of $s + i + r = \hat{N} = 1$ and time rescaling with $\tau = \beta t$, where $g = \frac{\gamma}{\beta} = \frac{1}{R_0}$.
And so on we can obtain analytical solutions for S, I, R.

3.3 Second order differential equation for the i-variable

Kudryashov et al. [4] provided an equivalent second order nonlinear differential equation *for the i-variable*:

$$\ddot{i} = -gi\dot{2} - i\dot{i} + \frac{\dot{i}^2}{i} \tag{5.1}$$

and for the SIR system:

$$\dot{\hat{I}}_t = -g\hat{I}^2 - \hat{I}\hat{I}_t + \frac{\hat{I}_t{}^2}{\hat{I}} \tag{5.2}$$

$$\dot{\hat{I}} = \frac{d\hat{I}}{dt} = \hat{I}_t \tag{5.3}$$

Proof. According to Eq.(4.1), Eq.(4.2), and Eq.(4.3), we can add them up and obtain the conservation law $\dot{s} + \dot{i} + \dot{r} = 0$. Combining Eq.(4.1), we obtain:

$$s = g + \frac{\dot{i}}{i}$$

Differentiating Eq.(4.2) and substituting Eq.(4.1):

$$\ddot{i} = -g\dot{i} + i\dot{s} + \dot{i}s = -g\dot{i} - i^2 s + \dot{i}s = -g\dot{i} + \dot{i}\left(g + \frac{\dot{i}}{i}\right) - i^2\left(g + \frac{\dot{i}}{i}\right) = -gi^2 + \frac{\dot{i}^2}{i} - i\dot{i}$$

Which is parameterised with one single variable, easier to portrait its phase.

3.4 Second order differential equation for the s-variable

The SIR system reduces to nonlinear differential equation *for the s-variable* [3]:

$$\ddot{s} = \frac{\dot{s}^2}{s} + (s - g)\,\dot{s} \tag{6.1}$$

and for the SIR system:

$$\dot{\hat{S}}_t = \frac{\hat{S}_t{}^2}{\hat{S}} + \left(\hat{S} - g\right)\hat{S}_t \tag{6.2}$$

$$\dot{\hat{S}} = \frac{d\hat{S}}{dt} = \hat{S}_t \tag{6.3}$$

Proof. Differentiating Eq.(4.1) and substituting i from Eq.(4.1), we obtain:

$$\dot{i} = \frac{di}{dt} = -\frac{\ddot{s} + i\dot{s}}{s} = \frac{\dot{s}^2}{s^2} - \frac{\ddot{s}}{s}$$

Then substituting $\dot{i}$ from Eq.(4.2) gives:

$$\frac{\dot{s}^2}{s^2} - \frac{\ddot{s}}{s} = \frac{g\dot{s}}{s} - \dot{s}$$

from which the result follows.

3.5 Second order differential equation for the r-variable

The reduced nonlinear differential equation *for the r-variable* is given by [3]:

$$\ddot{r} = g\dot{r}\left(e^{-\frac{r}{g}} - 1\right) \tag{7.1}$$

and for the SIR system:

$$\dot{\hat{R}}_t = g\hat{R}_t e^{-\frac{\hat{R}}{g}} - g\hat{R}_t \tag{7.2}$$

$$\dot{\hat{R}} = \frac{d\hat{R}}{dt} = \hat{R}_t \tag{7.3}$$

Proof. Differentiating Eq.(4.3), we have $\ddot{r} = g\left(si - gi\right) = gsi - g\dot{r} = -g\dot{s} - g\dot{r}$. In addition, dividing Eq.(4.3) by Eq.(4.1), obtaining:

$$\frac{\dot{r}}{\dot{s}} = \frac{dr}{ds} = -\frac{g}{s}$$

Therefore: $r = -g\ln s + C$. We then use the fixed-point condition $s_0 = s\left(0\right) = g$, so that:

$$s = ge^{-\frac{r}{g} + \frac{r_0}{g}}$$

Since the system is autonomous we can translate the origin to $-\infty$, where $r_0 = 0$ from which the result follows.

3.6 Determining the values for β and γ

The application of the SIR model requires primarily entries of parameters β and γ. Often, the recovery rate γ can be calculated through 1/infectious period, in which the infectious period can be found relatively easily. β will then be computed with function of $\beta = \gamma R_0$, where R_0 stands for the reproductive ratio. Various methods exists currently for the estimation of R_0 [5], for example the White & Pagano Method and the Sequential Bayes Method.

4 Results

The codes for plotting graphs are available at *https://github.com/Aflyingfrank/SIR-model-in-RStudio/tree/main*.

4.1 Case study I: Influenza A

The influenza epidemic was recorded to take place for three times in the history, the 1918–1919 Spanish flu, 1957–1958 Asian Flu and 1968–1969 Hong Kong Flu, with respective mortality rates ranging from devastating to mild [6]. However, influenza brings significant impact to the entire humanity, which calls for greater attention to it, especially for type A influenza.
In this case application, the reproductive ratio chosen for the disease was estimated to be 1/3 and the recovery rate was assumed to be 1/5 based on [5].

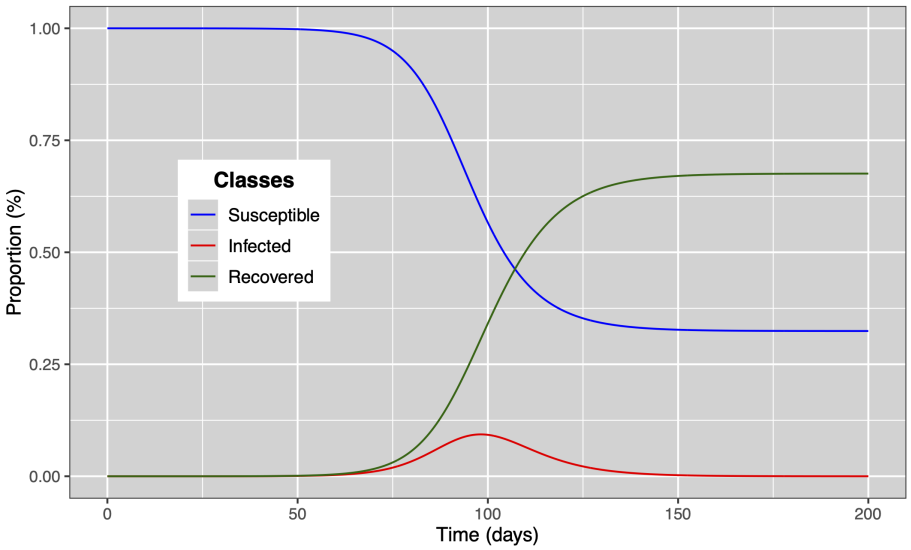

Figure 1: The proportion(%) of susceptible, infected and recovered population for influenza A over time(days) ($\beta = 1/3$, $\gamma = 1/5$).

The results showed that infections peaked at approximately day 95, and around 32% of the subjects remained uninfected throughout the epidemic (Figure 1).

4.2 Case study II: COVID-19

COVID-19 is a worldwide pandemic, and according to World Health Organization, over 769 million confirmed cases and over 6.9 million deaths have been

reported by 6 August 2023 globally [7]. Its outbreak has severely affected each and every field, such as social, scientific, industrial, transport, and medical sectors [8]. According to estimations conducted based on data in Brazil[9] and Algeria [10], we assume the values for R_0 and γ to be 1.84 and 0.1. Therefore, we obtain β to be 0.184.

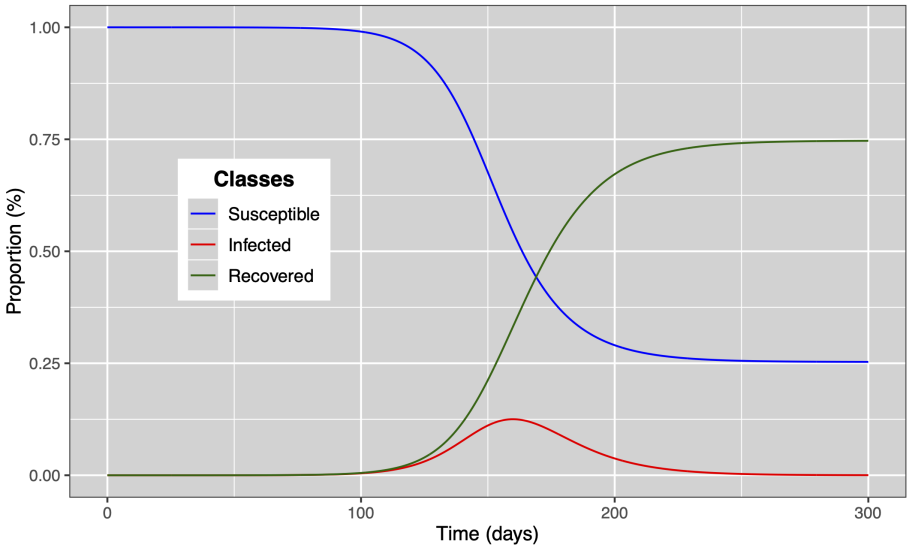

Figure 2: The proportion(%) of susceptible, infected and recovered population for COVID-19 over time(days) ($\beta = 0.184$, $\gamma = 0.1$).

The outcome turned out that infections reached maximum at approximately day 160, and around a quarter subjects escaped COVID-19 infection throughout the 300 days (Figure 2).

5 Discussion

The SIR model has some inevitable limitations in terms of epidemic prediction [1]. For example, it is generally hard to estimate the contact rate of the disease β, which depends on specific disease on social and behavioural factors [11]. Additionally, various non-medical interventions (i.e. wearing personal protective equipment, restricting interpersonal contacts, quarantine) may lead to certain changes in β. At the same time, some diseases is not applicable to the SIR model as recovered subjects may get re-infected (i.e. new variants occur, or the immune system break down again).
Another drawback of the SIR model is that the constants: contact rate β and rate of leaving infection γ, requires a particular amount of data to determine,

especially determine accurately. As a result, the SIR model cannot be applied to epidemic prediction until enough data was collected, which calls for time. However, for diseases like COVID-19, we should try to avoid the waste of time on data collection.
The Agent-Based Models (ABMs) are computer simulations used to study the interactions between people, things, places, and time. Within such models, the agents are programmed to behave and interact with others and the environment in certain ways [12]. It is not limited to observed data and can be used to model the counterfactual or experiments that may be impossible or unethical to conduct in the real world [13]. Therefore, ABMs may be used to predict the epidemic a lot earlier than the SIR model, which greatly reduces the wastes of time spent on data collection.

6 Conclusion

The SIR model provides one of the simplest pathway for epidemic prediction, with only three classes of population characterised. However, we should apply it with great care due to its underlying assumptions and limitations.

References

[1] Omar Melikechi et al. "Limits of epidemic prediction using SIR models". In: *Journal of Mathematical Biology* 85.4 (2022), p. 36.

[2] Jeffrey R. Chasnov. *The SIR Epidemic Disease Model*. URL: `https://math.libretexts.org/Bookshelves/Applied_Mathematics/Mathematical_Biology_(Chasnov)/04%3A_Infectious_Disease_Modeling/4.03%3A_The_SIR_Epidemic_Disease_Modelm` (visited on 07/01/2024).

[3] Dimiter Prodanov. "Analytical solutions and parameter estimation of the SIR epidemic model". In: *Mathematical Analysis of Infectious Diseases* (2022), pp. 163–189.

[4] Nikolay A Kudryashov, Mikhail A Chmykhov, and Michael Vigdorowitsch. "Analytical features of the SIR model and their applications to COVID-19". In: *Applied Mathematical Modelling* 90 (2021), pp. 466–473.

[5] Sawitree Boonpatcharanon, Jane M Heffernan, and Hanna Jankowski. "Estimating the basic reproduction number at the beginning of an outbreak". In: *Plos one* 17.6 (2022), e0269306.

[6] Rina Fajri Nuwarda, Abdulsalam Abdullah Alharbi, and Veysel Kayser. "An overview of influenza viruses and vaccines". In: *Vaccines* 9.9 (2021), p. 1032.

[7] World Health Organization. *Weekly epidemiological update on COVID-19 – 10 August 2023*. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2023?gad_source=1&gclid=EAIaIQobChMIpJr9h7ythAMVhcsWBR2OjQGIEAAYASABEgJER_D_BwE (visited on 07/03/2024).

[8] Saurabh Shivalkar et al. "Outbreak of COVID-19: a detailed overview and its consequences". In: *Coronavirus Therapeutics–Volume II: Clinical Management and Public Health* (2022), pp. 23–45.

[9] Hélder Seixas Lima, Unaí Tupinambás, and Frederico Gadelha Guimarães. "Estimating time-varying epidemiological parameters and underreporting of Covid-19 cases in Brazil using a mathematical model with fuzzy transitions between epidemic periods". In: *PloS one* 19.6 (2024), e0305522.

[10] Mohamed Lounis and Dilip Kumar Bagal. "Estimation of SIR model's parameters of COVID-19 in Algeria". In: *Bulletin of the National Research Centre* 44.1 (2020), p. 180.

[11] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Vol. 5. John Wiley & Sons, 2000.

[12] Donald L DeAngelis and Stephanie G Diaz. "Decision-making in agent-based modeling: A current review and future prospectus". In: *Frontiers in Ecology and Evolution* 6 (2019), p. 237.

[13] Columbia University Irving Medical Center. *Agent-Based Modeling*. https://www.publichealth.columbia.edu/research/population-health-methods/agent-based-modeling. Accessed: 2024-07-02. 2024.