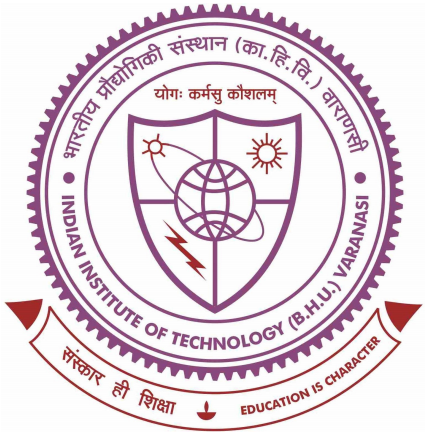
Stochastic Modelling Of Infectious Disease Spread



A Project Report for Mathematical Modelling And Simulation

CSM 321

Guide: Prof. Subir Das

By:

Akshat Agrawal
18123003
Mathematics And Computing

Palakur Eshwitha Sai
18123014
Mathematics And Computing

Dhriti Singh
18123006
Mathematics And Computing

Suyash Chhaparia
18123018
Mathematics And Computing

Harsh Sinha
18123008
Mathematics And Computing

Vishakha Tomar
18123020
Mathematics And Computing

Stochastic Modelling of Infectious Disease Spread

Abstract Why are the epidemic patterns of COVID-19 so different among different cities or countries which are similar in their populations, medical infrastructures, and people's behavior? Why are forecasts or predictions made by so-called experts often grossly wrong, concerning the number of people who get infected or die? The purpose of this study is to better understand the stochastic nature of an epidemic disease such as COVID-19, and answer the above questions. This work will stimulate mathematically inclined people to study this interesting and important field, i.e., mathematical epidemiology. The stochastic model we study in this article is based on the birth-and-death process with immigration (BDI for short). It is a special case of general birth-and-death (BD) process.

1. Introduction

Most of the mathematical models of infectious diseases seem to be based on the Kermack-McKendrick model published in 1927[4], which was proposed to explain the rapid rise and fall in the number of infected population observed in epidemics such as the plague epidemic in Bombay (1906)[1], and the great plague in London where more than 15% of the population died (1665-66); and the cholera outbreak in London caused by contamination in the Thames River (1865). The model consists of a system of three coupled nonlinear ordinary differential equations for the infected population $I(t)$, the susceptible population $S(t)$, and the recovered population $R(t)$. Kermack-McKendrick's **SIR model** is a **deterministic model**, which provides the expected values of these processes, which we denote as $\bar{S}(t)(= \mathbb{E}[S(t)])$, $\bar{I}(t)(= \mathbb{E}[I(t)])$, $\bar{R}(t)(= \mathbb{E}[R(t)])$. A majority of biological and epidemiological models[2] fall in this class of deterministic models.

Actual observed data of the infected population, for instance, is merely an instance or **a sample path** of this stochastic process $I(t)$. The process naturally deviates from the expected value $\bar{I}(t)$ obtained by a deterministic model. Thus, a deterministic model alone fails to provide any quantitative explanation when observed data differ significantly from the expected value.

In a **stochastic** (or **probabilistic**) model, on the other hand, the intrinsic stochastic nature of a process is explicitly taken into account in its model formulation. The importance of stochastic processes in relation to problems of population growth was

pointed out by W. Feller in 1939 [3]. He considered the **birth-and-death** process in which the expected birth and death rates (per person per unit time) were constants, say, λ and μ. D. G. Kendall extended Feller's birth-and-death (BD) process by considering the birth and death rates as any specified functions of the time t, $\lambda(t)$ and $\mu(t)$. The BD process is a special class of time - continuous, discrete - state Markov process, and has found applications in many scientific and engineering fields, including population biology, teletraffic and queueing theory, system modeling.

The general birth-and-death (BD) process usually defies an attempt to obtain a closed solution for the time-dependent (i.e., transient) probability distribution of the population size, etc. The BDI process, however, is among a small number of BD processes, which we can solve analytically. An important feature of the BDI process is that its probability distribution function is a generalized ***negative binomial distribution*** (NBD), with its parameter r being less than one. A NBD with small r has a long tail in its distribution form. These properties of the infection process explain why actual infection patterns exhibit enormously large variations. Furthermore, the mean value of the number infected provided by a deterministic model is far from the median of the distribution. This explains why any prediction based on a deterministic model will fail more often than not. Hence we study the stochastic modelling of disease spread further in this paper.

2. Stochastic Model for an Infectious Disease

2.1 Formulation for the time-dependent solution for the stochastic model

Let $I(t)$ represent the number of infected persons at time t, and $P_n(t)$ be the time-dependent (or transient) probability mass function (PMF) of the process $I(t)$, i.e.,

$$P_n(t) = Pr[I(t) = n], n = 0, 1, 2, ... \text{ and t} \geq 0. \tag{1}$$

We assume that each infected person is infectious, and infects susceptible persons at rate λ [persons/unit-time/person], where the time unit can be arbitrary, e.g., a second, an hour, a day, etc. Let us assume that an infected person recovers, gets removed or dies at rate μ per [unit-time]. Thus, μ^{-1} [unit-time] is the *mean infectious period*. The ratio λ/μ is equal to the *basic reproduction number*, i.e., the mean number of infections caused by an infected person during the infectious period.

We can formulate an infectious disease as a birth-and-death (BD) process, by defining the birth and death rates both of which are simple linear functions of the state n of the process *I(t)*:

$$\lambda_n = n\lambda + \nu, \ \mu_n = n\mu, \tag{2}$$

This particular state-dependent BD process is also known as the "birth-death-immigration (BDI)" process , in which the parameters λ, μ and ν represent the birth

(i.e., secondary infection), death (i.e., recovery or death) and immigration (i.e., arrival of an infected individual from outside) rates, respectively.

A few remarks are in order.

1. Actually, "recovery," "removal" and "death" are distinctly different matters. In analyzing the infection process, however, these three sources of loss from the susceptible or infected population, are mathematically equivalent in the sense they will not contribute to the infection process in the future. We assume here that those who have recovered from the disease have acquired immunity and will not be susceptible nor infectious.
   The assumption that each infected individual recovers (or is removed or dies) at rate μ is equivalent to assuming that the duration S that each sick person remains infectious is exponentially distributed with mean $1/\mu$, that is:
$$Pr[S \leq s] = 1 - e^{-\mu s},\ s \geq 0 \tag{3}$$
2. It can be shown mathematically that the results that will be obtained in terms of the probability mass function (PMF) of $I(t)$, and other related quantities are insensitive to the actual distribution of S . It is just required to set $\mu = \bar{S}^{-1}$, where $\bar{S} = \mathbb{E}[S]$.
3. Here, we assume that the population is homogeneous, and the susceptible population size remains sufficiently large. Thus, it is mathematically treated as "infinite." The parameters λ, μ and ν are assumed to be constant.

We can show that the PMF (1) of this BD process should satisfy the following set of linear differential-difference equations, a.k.a. *Kolmogorov's forward equation*:

$$\frac{dP_0(t)}{dt} = -\nu P_0(t) + \mu P_1(t) \tag{4}$$

$$\frac{dP_n(t)}{dt} = ((n-1)\lambda+\nu)P_{n-1}(t) - (n(\lambda+\mu)+\nu)P_n(t) + (n+1)\mu P_{n+1}(t),\ n = 1, 2, 3, ... \tag{5}$$

with the initial condition:

$$I(0) = I_0, \text{ i.e., } P_n(0) = \delta_{n,I_0}, n = 0, 1, 2,... \tag{6}$$

where $\delta_{m,n}$ is *Kronecker's delta.*

We transform the above set of infinitely many equations (5) into a single equation by using the probability generating function (PGF) defined by

$$G(z,t) = \mathbb{E}[z^{I(t)}] = \sum_{n=0}^{\infty} z^n P_n(t). \tag{7}$$

Multiply the set of equations (5) by z^n and sum them from n = 0 to ∞ , obtaining the following partial differential equation:

$$\frac{\partial G(z,t)}{\partial t} = (z-1)\left[(-\mu)\frac{\partial G(z,t)}{\partial z} + \nu G(z,t)\right] \tag{8}$$

with the boundary condition

$$G(Z,0) = z^{I_0} \tag{9}$$

2.2 Stochastic means of the infected process $I(t)$ and related processes

Although the process $I(t)$ is the main focus of our analysis, it will be worthwhile to introduce related processes and our assumptions.

Definition 1.

1. The process $A(t)$ is the cumulative count of external arrivals of infectious individuals from the outside. We assume that $A(t)$ is a Poisson process with rate v [persons/unit time],
2. The process $B(t)$ is the cumulative count of internally infected individuals. We assume that the birth of such persons occurs at the rate of λ [persons/unit time/infectious person].
3. The process $R(t)$ is the cumulative count of recovered / removed or dead individuals. We assume that the departure of such persons occurs at the rate of μ [persons/unit time/infected person]. Note that all infected persons are infectious persons until their recovery/removal/death.
4. The process $I(t)$ is the present number of infected persons, i.e.,

$$I(t) = I(0) + A(t) + B(t) - R(t) \tag{10}$$

The expectation and variance of the above processes will be of our interest, which we denote by

$$\overline{A}(t) = \mathbb{E}[A(t)],\ \overline{B}(t) = \mathbb{E}[B(t)],\ \overline{R}(t) = \mathbb{E}[R(t)]\ and\ \overline{I}(t) = \mathbb{E}[I(t)] \tag{11}$$

$$\sigma_A^2(t) = \mathbb{E}[(A(t) - \overline{A}(t))^2],\ etc. \tag{12}$$

Before finding the PGF $G(z,t)$ from the PDE (8), let us derive first an ordinary differential equation for $\overline{I}(t)$. By dividing both sides of (16) by $(z-1)G(z,t)$, we will have

$$(z-1)^{-1}\frac{\partial \ln G(z,t)}{\partial t} = (\lambda z - \mu)\frac{\partial \ln G(z,t)}{\partial z} + v \tag{13}$$

By setting $z = 1$, we find

$$\frac{d\overline{I}(t)}{dt} = (\lambda - \mu)\overline{I}(t) + v \tag{14}$$

where, on the LHS, we first set $z = 1$ (which corresponds to differentiation at $z = 1$), and use L'Hopital's rule, obtaining $\mathbb{E}[I(t)] = \overline{I}(t)$. The ordinary differential equation (14) can be solved, yielding

$$\overline{I}(t) = I_0 e^{at} + \frac{v}{a}(e^{at} - 1),\ t \geq 0,\ \text{where } I_0 = I(0) \text{ and } a = \lambda - \mu. \tag{15}$$

If the model parameters are set to new values, say, to λ', μ' and v' at some point $t = t_1 \geq 0$, then the solution $\overline{I}(t)$ for $t \geq t_1$ is given by

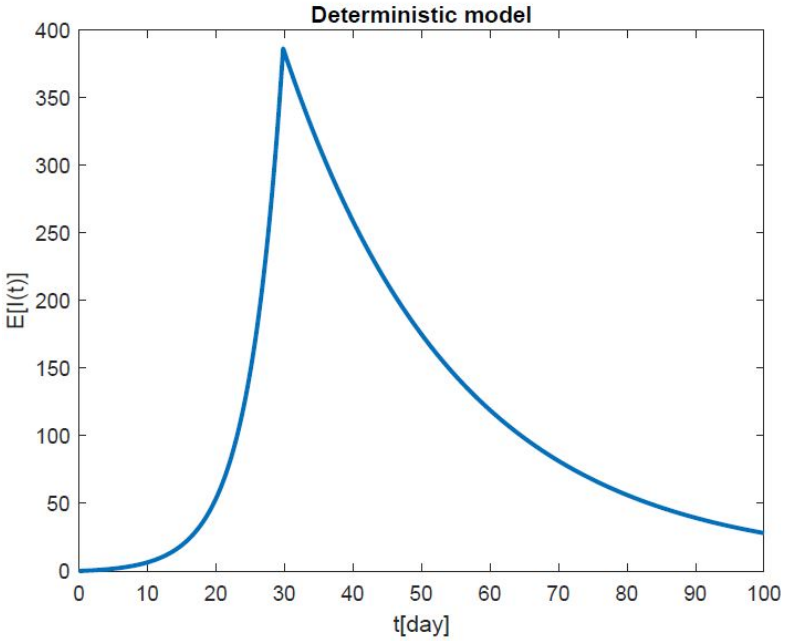
$$\overline{I}(t) = I_1 e^{a'(t-t_1)} + \frac{v'}{a'}(e^{a'(t-t_1)} - 1), \quad t \geq t_1, \text{ where } I_1 = \overline{I}(t_1),\ a' = \lambda' - \mu'. \quad (16)$$

Clearly, $\overline{I}(t)$ diverges to infinity, if $a > 0$ in (15) and converges to $\frac{v'}{|a'|}$, if $a' < 0$.

If $a = 0$, then

$$\overline{I}(t) = I_0 + vt = I_0 + A(t)\ t \geq 0, \text{ when } a = 0. \quad (17)$$

In Figure below



We show the case where $\lambda = 0.3$, $\mu = 0.1$ and $v = 0.2$ and at $t_1 = 30$, a new parameter $\lambda' = 0.06$ is set, whereas the original values of μ and v are retained.

The mean values of other processes can be easily found. $A(t)$ is a Poisson process with rate v, the differential of $\overline{B}(t)$ is given by

$$\frac{d\overline{B}(t)}{dt} = \lambda \overline{I}(t). \quad (18)$$

From this and (15) we obtain

$$\overline{B}(t) = \lambda \int_0^t \overline{I}(u)du = \frac{\lambda I_0(e^{at} - 1)}{a} + \frac{\lambda v(e^{at} - 1)}{a^2} - \frac{\lambda vt}{a}$$

$$\Rightarrow \overline{B}(t) = \frac{\lambda}{a}(\overline{I}(t) - I_0 - \overline{A}(t)),\ 0 \leq t \leq t_1. \quad (19)$$

For $t \geq t_1$, we find -

$$\begin{aligned} B(t) &= B(t_1) + \frac{\lambda' I_1 (e^{a'(t-t_1)} - 1)}{a'} - \frac{\lambda' v'(t-t_1)}{a} + \frac{\lambda' v' (e^{a(t-t_1)} - 1)}{a^2} \\ &= \frac{\lambda'}{a'}(I_1 e^{a'(t-t_1)} + \frac{v'}{a'}(e^{a'(t-t_1)} - 1) - I_0 - v't) \\ &= \frac{\lambda'}{a'}(I(t) - I_0 - \bar{A}'(t)), where : \bar{A}'(t) = v't, \end{aligned} \tag{20}$$

which takes the same form as that for $t \leq t_1$. Both the above equations can be derived from the mean value of the identity equation (10) together with the relation $\bar{R}(t) = \frac{\mu}{\lambda}\bar{B}(t)$, which is evident as shown below.
The recovery process $\bar{R}(t)$ should satisfy the following differential equation:

$$\frac{d\bar{R}(t)}{dt} = \mu \bar{I}(t). \tag{21}$$

thus, it readily follows:

$$\begin{aligned} \bar{R}(t) &= \mu(\bar{I}(t) - I_0 - \bar{A}(t))/a, 0 \leq t \leq t_1 \\ &= \mu'(\bar{I}(t) - I_0 - \bar{A}'(t))/a', t \geq t_1 \end{aligned} \tag{22}$$

The above expressions for $\bar{I}(t)$ and other processes can be viewed as our deterministic (or non - probabilistic) model for the dynamics of the BDI process. From these simple expressions, we can extract a few important characteristics concerning the mean values of I(t) and others.

1. $a = \lambda - \mu$ determines the exponential growth or decay rate of $\bar{B}(t)$, $\bar{B}(t)$ as well as $\bar{I}(t)$.
2. v is merely a linear scaling factor for $\bar{I}(t)$ and other processes, and so is I_0, the initial number of infected people.
3. If $a > 0$, then $\bar{I}(t)$ grows exponentially without bound; if $a < 0$, it decays exponentially towards $v/|a|$. If $a = 0$, $\bar{I}(t) = I_0 + vt$, i.e., $\bar{I}(t)$ grows linearly.
4. The ratio of the infection rate (or reproduction rate) λ to the recovery or removal rate μ

   $$R_0 = \frac{\lambda}{\mu} \tag{23}$$

   is called the basic reproduction number in epidemology. The term was originally defined in the context of a deterministic model called the SIS (*Susceptible - Infected - Susceptible*) epidemic model. It is the average number of persons whom an infectious person infects before his/her recovery, removal, or death. The reproduction number determines whether the infection will grow exponentially, die out, or remain constant, depending on whether $R_0 > 1$, $R_0 < 1$, or $R_0 = 1$, respectively. The exponential parameter a can be expressed in terms of R_0 and μ:

   $$a = \lambda - \mu = (R_0 - 1)\mu \tag{24}$$

5. The amount of time T that takes for $\bar{I}(t)$ or other related quantities to double, and the exponential growth parameter a are related by:

$$e^{aT} = 2 \implies aT \approx 0.693 \tag{25}$$

   Both the integral and derivates of the exponential function e^{aT} are also $\propto e^{aT}$. Thus, the above formula T equally applies, when cumulative numbers or incremental numbers are to be counted for a given a via an observed T.
6. Unless we can expect to increase the value of μ by improving the medical service or producing an effective vaccine to immunize the susceptible population, the only options we have for controlling an infectious disease is to increase μ by removing as many infectious individuals away from susceptible population as possible, and/or to decrease λ by increasing the social distances between the susceptible and the infectious.

2.3 Steady-state distribution of the *I(t)*

So far we have discussed only the mean values of the random process I(t) and other processes. Before we find the probability mass functions $P_n(t), n = 0, 1, 2,$ for any t, we obtain in this section the steady-state distribution $lim_{t\to\infty} P_n(t) = \pi_n$, if it exists. When $a > 0$, such a distribution does not exist.

Thus, the steady state distribution can possibly exist, only when $a \leq 0$. In order to find it, we set the LHS of the PDE (8) equal to zero, obtaining the following ordinary differential equation for the PGF G(z, ∞):

$$(\lambda z - \mu) + \frac{dG(z,\infty)}{dz} + vG(z,\infty) = 0 \tag{26}$$

which readily leads to

$$\frac{dlnG(z,\infty)}{dz} = -\frac{v}{\lambda z - \mu} \tag{27}$$

Integrating the above and using the boundary condition G(1, t) = 1 for any t, we find

$$G(z,\infty) = \left(\frac{1-\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}z}\right)^r, r = \frac{\lambda}{\mu} \tag{28}$$

This PGF reminds us of the ***negative binomial distribution*** (NBD). This distribution was originally introduced to express the probability of the number of failures n needed to achieve r successes in a sequence of Bernoulli trials, when the probability of failure is q.

Definition 2 (Negative binomial distribution (NBD)). Negative binomial distribution NB(r, q) is defined by

$$P_n^{NB} = \begin{pmatrix} n+r-1 \\ n \end{pmatrix} (1-q)^r q^n z^n = \frac{\Gamma(n+r)}{n!\Gamma(r)} (1-q)^r q^n z^n \text{ n = 0,1,2,...} \tag{29}$$

where the parameter r is a positive real number.

When r is a positive integer, and 0 < q < 1 the above reduces to the classical definition of the shifted negative binomial distribution, sometimes called the Pascal distribution, associated with Bernoulli trials. The Gamma function $\Gamma(x)$ is defined for a positive real number x by

$$\Gamma(x) = \int_0^\infty y^{(x-1)} e^{-y} dy, \text{ where } x > 0 \tag{30}$$

The probability generating function of NB(r, q) is given by

$$G(z) = \sum_{n=0}^{\infty} \left(\begin{array}{c} n+r-1 \\ n \end{array} \right) (1-q)^r q^n z^n = \frac{1-q}{1-qz}, |z| < q^{-1} \tag{31}$$

The mean and variance of a RV (random variable) X possessing this distribution can be readily found:

$$E[X] = \frac{qr}{1-q}, Var[X] = \frac{qr}{(1-q)^2} \tag{32}$$

From (19) and the above formula, we readily find the steady state distribution of $I(t)$ for $a < 0$is given as follows:

$$\pi_n = \lim_{t\to\infty} P_n t = \left(\begin{array}{c} n+r-1 \\ n \end{array} \right) \left(\begin{array}{c} \lambda \\ \mu \end{array} \right)^n \left(1 - \frac{\lambda}{\mu}\right)^r \tag{33}$$

3. Conclusion

Here in this project report we have focused mainly on the analysis of $I(t)$ (the infected population), But also the stochastic means of $B(t)$ (the population of the secondary infections) and $R(t)$ (the recovered/dead population) are represented as well.
We have also shown how the infection process can be controlled by changing the values of λ, μ and v at certain points in time.As known by common knowledge increasing the so-called social distance would obviously reduce the value of parameter λ.Similarly the value of μ can be increased by improving medical treatments to speed up the recovery process. Alternatively the susceptible population can be minimised by exposing the young and healthy susceptible to the disease so that they become immune to the disease

Appendix

Acknowledgements

We thank the constant efforts of our professor, Dr. Subir Das, for his constant motivation and guidance throughout this project. It was only because of his efforts that we were able to understand stochastic processes nicely and convey the ideas mentioned in this project report. We also thank him for understanding our problems during the project and helping us out of it.

References

[1] N. Bacaër, "The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality," *Journals of Mathematical Biology*, vol. 64, pp. 403–422, March 2012.

[2] R. M. Anderson and R. M. May, "Population Biology of Infectious Diseases: Part I," *Nature*, vol. 280, pp. 361–367, August 1979.

[3] M. Martcheva, *An Introduction to Mathematical Epidemiology*. Springer, 2010.

[4] W. Feller, "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoritischer Behandlung," *Acta Biotheoretica*, vol. 5, pp. 11–40, 1939.

[5] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *Proc. Royal Soc. London*, vol. A 115, pp. 700–721, 1927.