STOCHASTIC EPIDEMIC MODELS

BY GERARDO PALAFOX[1],

[1]*Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, gerardo.palafoxcstl@uanl.edu.mx*

In this work, a stochastic model for the spread of infectious diseases is presented. Theoretical results and simulations are shown. Additionally, data from the ongoing pandemic caused by the SARS-CoV-2 virus is incorporated to discuss the implications of the model.

1. Introduction. The goal of this work is to present one of the most common stochastic epidemic models, alongside computational simulations of this. The terminology and theoretical preliminaries for later sections are detailed in Section 2. In Section 2.1, related work is presented. Models and variations are later presented in Sections 3, 4 and 5. Finally, data from the current COVID-19 pandemic, caused by the SARS-CoV-2 virus, is discussed in Section 6.

2. Preliminaries. The probability of an event A will be denoted by $\mathbb{P}\{A\}$. If X is a random variable, the expected value of X and its variance are denoted by $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ respectively. A sequence of random variables $X_1, X_2, \ldots$, converges in probability to a random variable X if, for every $\varepsilon > 0$,

$$(1) \qquad \lim_{n \to \infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0 \text{ or, equivalently, } \lim_{n \to \infty} \mathbb{P}\{|X_n - X| < \varepsilon\} = 1.$$

A sequence of random variables $X_1, X_2, \ldots$, converges in distribution to a random variable X if

$$(2) \qquad \lim_{n \to \infty} F_{X_n}(x) = F_X(x),$$

at all points x where $F_X(x)$ is continuous, where F_X is the distribution of the random variable X. A stochastic process is a collection $\{X(t) : t \in J\}$ of random variables. Index t is often interpreted as time, and $X(t)$ is referred to as the state of the process at time t. The set J is called the index set of the process, and determines whether the process is discrete-time or continuous-time, depending on whether J is countable or an interval of the real line. A Markov chain is a stochastic process where the random variables take a finite or countable number of possible values, and with the property that

$$(3) \qquad \mathbb{P}\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \mathbb{P}\{X_{n+1} = j \mid X_n = i\},$$

for any such states i_k and any $n > 0$.

2.1. *Related Work.* The mathematical study of epidemics has been long studied, with early models focusing on specific diseases [5, 23]. More general studies have been performed since then, being Kermack and McKendrick [16] some of the first in doing so, and Bailey [2], Frauenthal [13], Brauer and Castillo-Chávez [6] more recently. This work will focus on models with stochastic components [1, 7], but deterministic models have also been studied [14]. Models where the disease spreads over a network have been studied thoroughly, both in their deterministic [20] and stochastic [8, 26] form. Agent-based simulations have also been used to model epidemics [15].

3. Reed-Frost model. One of the simplest stochastic models of disease spreading is the Reed-Frost model, which is included in this work for reference. This model is a discrete-time Markov chain, were an individual in a finite population is classified as susceptible, infected or removed; these are called SIR models. The model in Section 4 is also an SIR model. At time t, there are S_t individuals susceptible to infection, and I_t currently infectious. The incubation period of the disease, as well as the recovery time, is assumed to occur between time jumps. That is, susceptible people exposed at time t will be infectious at time $t + 1$, and people infectious at time t will be removed from the process at time $t + 1$. Any two individuals have a probability $p = 1 - q$ of coming in contact at time t, with encounters being independent of each other. As such, the distribution of S_{t+1} is binomial $\mathrm{Bin}(s, q^i)$. Observe that $I_{t+1} = S_t - S_{t+1}$, since in the step from t to $t + 1$, all the infectious from time t are removed, and the decrease in susceptible individuals is precisely the amount of new infectious people. Thus, $I_{t+1} \sim \mathrm{Bin}(s, 1 - q^i)$. More of this process and its asymptotic properties can be read in a work by Von Bahr and Martin-Löf [27].

4. Standard stochastic epidemic model. The construction and details of this model can be found in the book of Andersson and Britton [1]. A description of the properties presented in Subsection 4.1, among more, can be found in a survey by Britton [7]. A fixed population of size n will be considered. It is assumed there are no births, deaths or migration; this is what is known as a closed population. Additionally, population is assumed well-mixed and homogeneous: this means any two elements of the population can be in contact with each other, and everyone is affected by the disease in the same way. Each individual will be either susceptible (S), infectious (I), or recovered (R, also called removed). Let $S(t), I(t), R(t)$ be the number of susceptible, infectious and recovered individuals at time t. At $t = 0$, one has $S(0) = n - m, I(0) = m, R(0) = 0$. The contacts an infectious individual has with others follows a Poisson process with intensity λ [1, 18]. Each such contact is with an individual selected uniformly at random from the population, with contacts of different infectious individuals being mutually independent. When the contact occurs between an infectious individual and a susceptible individual, the susceptible individual is assumed to be infected, and it is immediately infectious. Every infectious individual remains so for a random time $I \sim F_I$, called the *infectious period*, where F_I has mean $1/\gamma$. These infectious periods are independent and identically distributed, and are independent from the contact process. Epidemic starts at time 0, and goes until a time T where $I(T) = 0$. Thus the final state of the epidemic is given by $S(T) = n - R(T), I(T) = 0$ and $R(T) = m + Z$; i.e., Z people where infected in the course of the epidemic. While, regardless of the distribution F_I, there is no closed form expression for the time dynamics of the process, it is possible to derive a recursive formula for the final size of the epidemic, see the survey of Britton [7]. However, this becomes difficult for even moderately large n [7].

When $F_I \sim \mathrm{Exp}(\gamma)$, the model becomes a continuous-time Markov chain, known as the *stochastic general epidemic model*, which was first considered by Bartlett [4]. There is no epidemiological reason to assume the infectious period follows an exponential distribution, but doing so is of mathematical convenience, as it allows diffusion approximations valid for large values of n [1]. If the infectious period is not random but constant, the resulting model is a continuous-time version of the Reed-Frost model [1] (see Section 3 for a definition of this model).

4.1. *Properties of the model.* The aim of this section is to explore key questions revolving around an epidemic process, such as the probability of having a major outbreak, or how many people will be infected in such an outbreak. An approximation of the early stages of the epidemic process is also considered. Everything presented will be independent of the distribution F_I, i.e. it will not rely on the Markovian case.

Britton [7] details how, when n is sufficiently large, the initial phase of the epidemic can be approximated by a homogeneous branching process with birth-rate λ and life-distribution $I \sim F_I$ [18] having Laplace transform $\phi(s)$. The epidemic and branching process will agree at least until there has been k contacts, with k small in relation to $\sqrt{n}$. The offspring distribution of the branching process has mean $\lambda \mathbb{E}[I] = \lambda/\gamma$ [7], a quantity denoted by R_0 from here on. Also mentioned by Britton [7] is the fact that when $R_0 \le 1$ the final size of the epidemic is bounded in probability, and it is not when $R_0 > 1$. Additionally, if $R_0 > 1$, the epidemic will be minor with probability q^m and mayor with probability $\rho = 1 - q^m$, where q is the smallest solution to

$$(4) \qquad q = \phi(\lambda(1-q)).$$

A more detailed treatment of the properties just stated can be found in the work of Ball [3]. The number R_0 is known in the literature as the *basic reproduction number*, and, as it is seen, plays an important role in whether the epidemic dies out "rapidly"or there is a major outbreak.

A rigorous study of the final size of the epidemic can be found in the work of Scalia-Tomba [24, 25]. Britton [7] gives a balance equation to find the fraction of people not getting infected. Particularly, one can find in his work [7] that if z is the fraction of people infected, by the law of large numbers the limiting fraction infected in case of a major outbreak is a solution to the equation

$$(5) \qquad 1 - z = e^{-R_0 z}.$$

The value $z = 0$, which corresponds to a minor outbreak, is always a solution to this equation. Also, when $R_0 > 1$, there exists an additional unique solution z^* in the unit interval. Furthermore, denoting by Z_n the final number of infected individuals for a population of size n, one has that if $R_0 \le 1$, then $\bar{Z}_n := Z_n/n \to 0$ in probability as $n \to \infty$ [7]. If, however, $R_0 > 1$, then $\bar{Z}_n \to \zeta$, where ζ is a two point distribution defined by $\mathbb{P}\{\zeta = 0\} = q^m$, $\mathbb{P}\{\zeta = z^*\} = 1 - q^m$ [7], with q as seen in Equation 4 and z^* being the positive solution to Equation 5 previously mentioned. Defining $r^2 = \operatorname{Var}(I)/(\mathbb{E}[I])^2$, when $\bar{Z}_n \to z^*$ one has

$$(6) \qquad \sqrt{n}(\bar{Z}_n - z^*) \to \mathcal{N}\left(0, \frac{z^*(1-z^*)(1 + r^2(1-z^*)R_0^2)}{(1 - (1-z^*)R_0)^2}\right),$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution of mean μ and variance σ^2.

4.2. *Simulations.* The code for the simulations in this and further sections is publicly available in a GitHub repository[1]. The simulations were performed using Gillespie's direct method algorithm [11]. In Figure 1, two trajectories of the stochastic general epidemic model are shown. Both consist of a population of 1000 with 10 initial infected. The trajectory in Figure 1a has an $R_0 = .5$ and the one in Figure 1b has an $R_0 = 1.5$. In Figure 2, final size distributions of these models are displayed; Figure 2a shows the distribution of the final size of the model with $R_0 = .5$, while Figure 2b shows the distribution of the final size of the model with $R_0 = 1.5$.
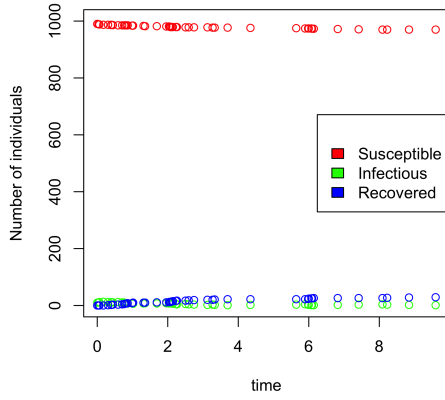
---

[1]https://github.com/palafox794/AppliedProbabilityModels/tree/master/FinalProject
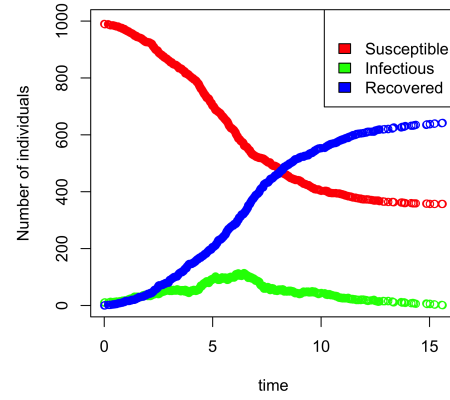
(a) Model with $R_0 = 0.5$.

(b) Model with $R_0 = 1.5$.

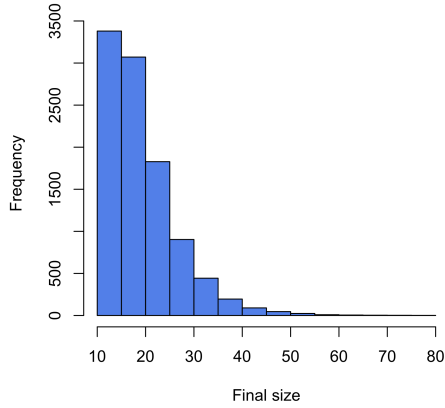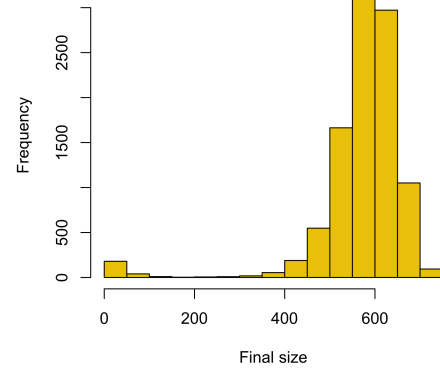Fig 1: Trajectories of a stochastic general epidemic model.

(a) Final size distribution, $R_0 = 0.5$.

(b) Final size distribution, $R_0 = 1.5$.

Fig 2: Final size distribution of 10,000 simulations in a general stochastic epidemic model.

5. Network models. One of the assumptions of the standard epidemic model, is that any two individuals have an equal probability of making contact. While this makes it simpler, it is of course less realistic. In reality, some people are more prone to being exposed (hospital or school workers, students), and more likely to infect those who are in close contact with them (coworkers, household members). A way to model this situations is to consider the members of the population as nodes in a network (see the book of Newman [21] for basic network definitions). Such epidemic models abound in the scientific literature [10, 12, 17, 22]. As an example, we define two such models, as they appear in the article of Britton [8]. The interested reader is invited to read the survey by Pastor-Satorras et al. [22] or the work of Britton [8].

DEFINITION 5.1 (Discrete time Reed-Frost model on a network). This model is a network generalization of the one described in Section 3. Having an arbitrary network, a randomly chosen node is set as infectious at time $t = 0$, while the rest are susceptible. At time t, infectious individuals will infect susceptible neighbors (i.e., adjacent nodes) independently with probability p. The susceptible individuals that where infected at time t will be infectious at time $t + 1$, and infectious individuals at time t will be removed at time $t + 1$. This goes on until a time T when there are no new infections.

DEFINITION 5.2 (Continuous-time Markov chain model on a network). This model is a network generalization of the stochastic general epidemic model described in Section 4. Having an arbitrary network, a randomly chosen node is set as infectious at time $t = 0$, and the rest are susceptible. Infectious individuals have contacts with each susceptible neighbor (i.e. adjacent nodes) randomly in time according to an independent Poisson processes with rate λ [18]. Each infected individual remains infectious for a period $I \sim \text{Exp}(\gamma)$, after which it is removed from the epidemic (considered recovered and immune). Both infectious periods and contact processes are defined independently. This goes on until a time T when there are no new infectious nodes.

As with the standard epidemic model in Section 4, the distribution of the infectious periods can be chosen arbitrarily, at the expense of the Markovian property. Similarly, setting a constant infectious period gives rise to a continuous-time Reed-Frost model [8].

6. COVID-19 pandemic. At the moment this work is being written, a pandemic caused by the SARS-CoV-2, colloquially known as COVID-19, is widespread in most countries of the world. Various authors have tried to estimate this disease's R_0; a review by Liu et al. [19] has found that the mean R_0 estimated is of 3.28, with a median of 2.79. In the Mexican northern state of Nuevo León, the first cases were reported in March 2020. In Figure 3, two graphics are shown regarding the spread of the epidemic process in Nuevo León from April 1st, 2020 to November 30th, 2020, with information provided by the Mexican government [9]. In particular, a graph of the cumulative cases in the state is shown in Figure 3a. A 7-day rolling average of cases per day is seen in Figure 3b. In a stochastic general epidemic model, as the presented in Section 4, an R_0 of 3 would give a limiting fraction of .94 infected in the population, by solving Equation 5. This can be seen in the trajectory in Figure 4, a simulated trajectory of an epidemic in a population of 1000 which has $R_0 = 3$. This of course should not be interpreted as a suggestion that the current pandemic will infect 94% of the population, since the model is quite weak in its assumptions. It is, however, a simplified example of how quickly these processes can grow out of control when there is nothing stopping its spread.

7. Conclusions. In this work, two common epidemic models were presented, as well as generalizations to network models. In particular, the standard stochastic epidemic model was studied in more detail. Simulations were performed and displayed. Finally, some discussion of how these relate to a current, ongoing pandemic was presented. The standard stochastic epidemic model can be generalized beyond its SIR presentation. Models for further study include those were there is an incubation period after infection, before becoming infectious (SEIR) or those with waning immunity (SIRS). Models with dynamic population, taking into account births and deaths or migration, were also lacking in this work. While the network models in Section 5 remove the well-mixed assumption of the population, individuals are still homogeneous. Multi-type models [1] address this limitations, and can be studied in further work.
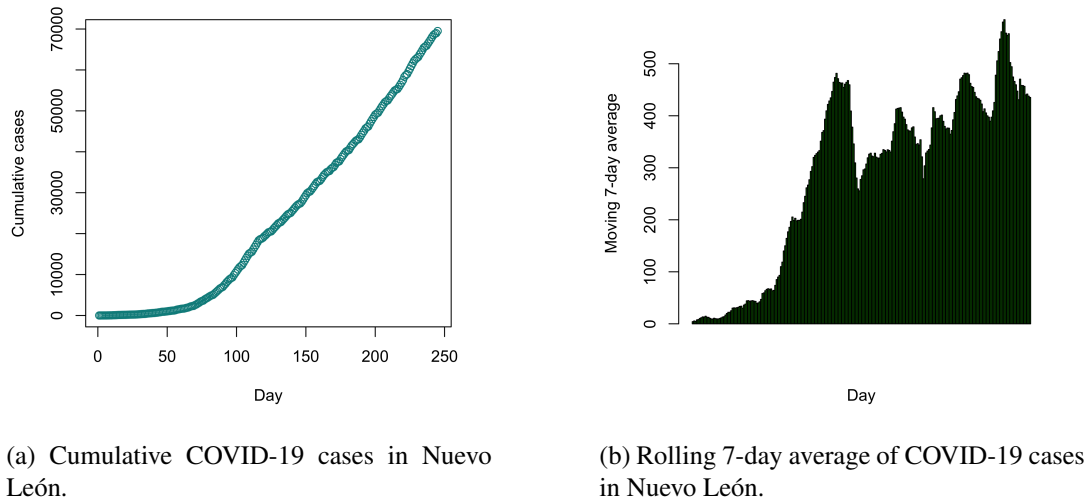
(a) Cumulative COVID-19 cases in Nuevo León.

(b) Rolling 7-day average of COVID-19 cases in Nuevo León.

Fig 3: COVID-19 cases in the state of Nuevo León from 2020-04-01 to 2020-11-30.
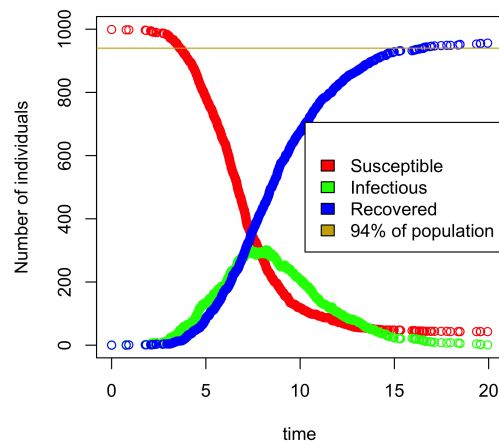
Fig 4: General stochastic model with $R_0 = 3$.

REFERENCES

[1] ANDERSSON, H. and BRITTON, T. (2000). *Stochastic epidemic models and their statistical analysis. Lecture notes in statistics*. Springer.

[2] BAILEY, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*, 2nd ed. Griffin.

[3] BALL, F. (1983). The threshold behaviour of epidemic models. *Journal of Applied Probability* **20** 227–241.

[4] BARTLETT, M. S. (1949). Some Evolutionary Stochastic Processes. *Journal of the Royal Statistical Society: Series B (Methodological)* **11** 211–229.

[5] BERNOULLI, D. (1760). Essai d'une nouvelle analyse de la mortalite causee par la petite verole, et des avantages de l'inoculation pour la prevenir. *Histoire de l'Acad., Roy. Sci. (Paris) avec Mem* 1–45.

[6] BRAUER, F. and CASTILLO-CHÁVEZ, C. (2012). *Mathematical models in population biology and epidemiology*, 2nd ed. *Texts in applied mathematics*. Springer.

[7] BRITTON, T. (2010). Stochastic epidemic models: A survey. *Mathematical Biosciences* **225** 24–35.

[8] BRITTON, T. (2019). Epidemic models on social networks – with inference. arXiv:1908.05517.

[9] GOBIERNO DE MÉXICO Covid-19 México. https://datos.covid-19.conacyt.mx/#DownZCSV.

[10] DEIJFEN, M. (2011). Epidemics and vaccination on weighted graphs. arXiv:1101.4154.

[11] DRAKE, J. M. and ROHANI, P. Stochastic Models. https://daphnia.ecology.uga.edu/drakelab/wp-content/uploads/2016/07/sismid-stochastic-lecture.pdf.

[12] FRANSSON, C. and TRAPMAN, P. (2019). SIR epidemics and vaccination on random graphs with clustering. *Journal of Mathematical Biology* **78** 2369–2398.

[13] FRAUENTHAL, J. C. (1980). *Mathematical modeling in epidemiology*. *Universitext*. Springer.

[14] HETHCOTE, H. W. (2000). The Mathematics of Infectious Diseases. *SIAM Review* **42** 599–653.

[15] HOERTEL, N., BLACHIER, M., BLANCO, C., OLFSON, M., MASSETTI, M., RICO, M. S., LIMOSIN, F. and LELEU, H. (2020). A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine* **26** 1417–1421.

[16] KERMACK, W. O. and MCKENDRICK, A. G. (1927). A contribution to the mathematical theory of epidemics. **115** 700–721.

[17] KISS, I. Z., MILLER, J. C. and SIMON, P. L. (2017). *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, 1st ed. *Interdisciplinary Applied Mathematics*. Springer International.

[18] LAWLER, G. F. (2006). *Introduction to Stochastic Processes*, 2nd ed. Taylor and Francis/CRC Press.

[19] LIU, Y., GAYLE, A. A., WILDER-SMITH, A. and ROCKLÖV, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine* **27**.

[20] MEI, W., MOHAGHEGHI, S., ZAMPIERI, S. and BULLO, F. (2017). On the dynamics of deterministic epidemic propagation over networks. *Annual Reviews in Control* **44** 116–128.

[21] NEWMAN, M. (2018). *Networks* **1**. Oxford University Press.

[22] PASTOR-SATORRAS, R., CASTELLANO, C., VAN MIEGHEM, P. and VESPIGNANI, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics* **87** 925–979.

[23] ROSS, R. (1910). *The Prevention of Malaria*. J. Murray.

[24] SCALIA-TOMBA, G. (1985). Asymptotic final-size distribution for some chain-binomial processes. *Advances in Applied Probability* **17** 477–495.

[25] SCALIA-TOMBA, G. (1990). On the asymptotic final size distribution of epidemics in heterogeneous populations. In *Stochastic Processes in Epidemic Theory* (J.-P. GABRIEL, C. LEFÈVRE and P. PICARD, eds.) 189–196. Springer Berlin Heidelberg.

[26] TRAPMAN, P. and BOOTSMA, M. C. J. (2009). A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection. *Mathematical Biosciences* **219** 15–22.

[27] VON BAHR, B. and MARTIN-LÖF, A. (1980). Threshold limit theorems for some epidemic processes. *Advances in Applied Probability* **12** 319–349.