



Research article

Time-dependent non-homogeneous stochastic epidemic model of SIR type

Mireia Besalú¹ and Giulia Binotto^{2,*}

¹ Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Avinguda Diagonal 643, 08028-Barcelona, Spain

² Department de Matemàtiques, Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain

* **Correspondence:** Email: giulia.binotto@uab.cat.

Abstract: To better describe the spread of a disease, we extend a discrete time stochastic SIR-type epidemic model of Tuckwell and Williams. We assume the dependence on time of the number of daily encounters and include a parameter to represent a possible quarantine of the infectious individuals. We provide an analytic description of this Markovian model and investigate its dynamics. Both a diffusion approximation and the basic reproduction number are derived. Through several simulations, we show how the evolution of a disease is affected by the distribution of the number of daily encounters and its dependence on time. Finally, we show how the appropriate choice of this parameter allows a suitable application of our model to two real diseases.

Keywords: SIR model; basic reproduction number; epidemic model; stochastic differential equations; Markov chain

Mathematics Subject Classification: 60H10, 60J10, 92D30

1. Introduction

The interest of mathematical models to describe the spread of a disease has a long history. It dates back to the eighteenth-century study of Bernoulli on smallpox inoculation [1]. Both deterministic and stochastic models have been considered and many factors have been taken into account: infectious agents, mode of transmission, latent periods, temporary or partial immunity, quarantine periods, etc. (see, for example, [2–7]). The main advantage of the deterministic approach lies in its more manageable (even if not necessarily simple) analysis. However, the most natural way to describe the spread of a disease is stochastic. This is due to several facts. On one hand, some phenomena are genuinely stochastic or present random features. On the other hand, the infective agent is introduced into the population only through a few hosts. Deterministic models cannot capture this component, given that they apply only when a small fraction (not a small number) of the large population is

infected. To solve this contrast, demographic stochasticity needs to be taken into account (see [4] for an exhaustive explanation).

An important role in mathematical epidemiology is played by the class of models that divide the population into different categories and study the transitions of state between them. One of the most classical are the SIR models, where the letters stand for Susceptible, Infectious and Removed, respectively. The first SIR stochastic model was proposed by McKendrick [8] as a continuous-time version of the deterministic model of Kermack and McKendrick [9], although more attention was given to the chain-binomial model of Reed and Frost (later published by Abbey [10]). From that moment on, many other texts and models have been developed.

Stochastic models can be divided into three major categories: discrete time models, continuous time Markov chain models and diffusion models (see [11]). The literature that deals with these three categories is extensive. To name a few, Tuckwell and Williams [12] consider a simple discrete time Markovian model in which the total population is constant and individuals meet a random number of other individuals at each time step. Ferrante et al. [13] generalize this model by adding two new classes. They assume that the number of encounters at each time step is constant for all individuals and time. Their results are more suitable for diseases with an initial latency period and the presence of asymptomatic individuals. On the other hand, Gray et al. [14] directly propose a system of equations to model the dynamics of the population. Discrete time models are mathematically less complex than those employed in the other two categories. However, they grant a simplified vision of the evolution of a disease which highlights its main characteristics. They also provide a good guide for constructing more complex models. Besides, they do not generally present specific constraints or assumptions that could be in contrast with empirical evidence as do, for example, continuous time Markov chain models.

In the present paper, we consider the SIR-type model proposed by Tuckwell and Williams [12]. We expand their results by weakening the homogeneity in time, that is, assuming the dependence on time in some parameters and by including factors that have not been taken into account. We deal with a discrete time, discrete state space stochastic model built on a generation basis. Time unit is of one-day length. The population is assumed to be closed, homogeneous and homogeneously mixing and it is divided into the three classes defined above. We assume that every day any susceptible individual meets a certain number of different individuals. If one of them is infective, the disease is transmitted with a given probability. While in [12] the number of daily encounters is time-homogeneous, in our case it can change at any time. Moreover, for the main computations and almost all simulations they fix this quantity for simplicity while we maintain it random all over the paper. With simulations and applications to real diseases, we will show that this randomness and dependence on time play an important role and can generate different scenarios. After being infected, an individual remains infectious for a fixed number of consecutive epochs. After this period, he or she recovers and becomes immune. We also add a parameter to the model to consider the case of a possible quarantine for infectious individuals.

We describe this type of model analytically using the methodologies of [12] and [13]. We derive an explicit structure for the underlying discrete time Markov chain and deduce the probability that a susceptible will become infected at the following epoch. This is the content of Section 2. Section 3 is dedicated to describing the dynamics of the model. Two possible situations are taken into account: when the duration of the disease is constant and when infectious individuals remain infectious throughout their lives. For both cases, we describe the distribution of new infected and derive a

diffusion approximation and the basic reproduction number R_0 . This is the expected number of secondary cases produced by an infectious individual during its infectious period in a virgin population. This value is used to measure the potential for transmission of a disease. Section 4 contains some simulations of the SIR-type model we have described. After fixing some parameters, we compare the evolution of the disease for different distributions of the number of daily contacts. We see how the different behaviors and the dependence on time of this parameter affect the course of the disease. Section 5 is devoted to the application of our model to real diseases. We consider the cases of influenza and meningococcal infection. The appropriate choice of the distribution of the number of daily encounters, both random and time-dependent, ensures that our model adjusts remarkably well to real data.

Throughout the paper, we use the term *infective* to refer to an individual who has contracted the disease and is infectious (he or she belongs to class I) while we connote with *infected* an individual who is either infectious or removed, i.e., no longer susceptible (he or she belongs to class I or R).

2. Description of the model

This section is dedicated to describing the basic properties of the SIR model we consider. It is an extension of the one proposed by Tuckwell and Williams [12]. The main difference regards the number of daily contacts. While Tuckwell and Williams assume that this quantity is time-homogeneous, in our case it can change at any time step allowing us to differentiate the number of contacts, for example, between weekdays and weekends, between different seasons, etc. Moreover, they assume it is a random variable but they fix it to be a constant in the main computations. On the contrary, we maintain the random character of this parameter throughout the paper. We also add a parameter to the model to consider the case that the infective individuals adopt some kind of quarantine. This can range from a perfect quarantine (there are no contacts) to no quarantine at all (the infective individuals have the same contacts as a susceptible one).

We assume:

- (1) Total population size: It is fixed at n . The population is closed.
- (2) Time: Time is discrete. In epidemics, the natural unit for the duration of an epoch is one day, although in some applications the time step is bigger (see, for example, [15]).
- (3) Definition of a sick individual: Given any individual i , with $i = 1, \dots, n$, we define a stochastic process $Y^i = \{Y^i(t), t = 0, 1, 2, \dots\}$ such that

$$Y^i(t) = \begin{cases} 1 & \text{if the individual } i \text{ is infective (and infectious) at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Then, the total number of infective and hence infectious individuals at time t is

$$Y(t) = \sum_{i=1}^n Y^i(t).$$

- (4) Daily encounters: Over $(t, t + 1]$ each individual i meets a random number $N^i(t)$ of other individuals. For all i and t , the variables $N^i(t)$ are mutually independent and independent of the

state of the population. Furthermore, for all i and t , $N^i(t)$ can take only a finite number of values n_k , each with probability $p_k(t)$. The values are fixed, that is, they depend neither on individuals nor on time while their probability varies with respect to time but not to individuals. This is, for $k = 1, \dots, m$,

$$P(N^i(t) = n_k) = p_k(t) \quad \forall i, t$$

with $\sum_{k=1}^m p_k(t) = 1$ for all t . Observe that the set of all possible values of $N^i(t)$ that we define $M = \{n_1, \dots, n_m\}$ is an ordered subset of $\mathbb{N}_{\geq 0}$. As we consider models where, for a fixed t , the distribution of the number of daily encounters is the same for all individuals, throughout the paper we will use the simplified notation $N(t)$ instead of $N^i(t)$ when appropriate.

- (5) Duration of the disease: Any individual remains infectious for r consecutive epochs where r is a positive integer. After this period, the individual recovers and becomes immune. The case without recovery, that is when $r = \infty$, is also considered.
- (6) Contagion probability: If an individual who has never been diseased up to and including time t meets an individual in $(t, t + 1]$ who is diseased at time t , then independently of the results of other encounters, this encounter results in transmission of the disease with probability p . Thus, if this individual becomes infected, it will be counted as such at epoch $t + 1$.
- (7) Encounter probability: The population is homogeneously distributed. This means that, given $Y(t) = y$, the probability that a randomly chosen individual is infectious at time t is given by y/n . We will multiply this probability by a constant $\lambda \in [0, 1]$ to characterize a possible quarantine of the infectious individuals. If $\lambda = 0$ the infectious individuals do a rigorous quarantine and for values of λ close to 1 they will do a mild or non-existent quarantine.

As mentioned before, the variables $N^i(t)$ are independent for all i and t . Assuming that for a fixed t they are also identically distributed, the model can be seen as an $(r + 1)$ -dimensional Markov chain. Observe that only if $N^i(t)$ are identically distributed with respect to time, this model is a homogeneous Markov chain.

Let

- $Y_\ell(t)$ be the number of individuals who are infective at time t and have been infective for exactly ℓ time units, with $\ell = 0, 1, \dots, r - 1$;
- $X(t)$ be the number of susceptible individuals at time t ;
- $Z(t)$ be the number of individuals who were previously infective and are recovered at time t .

We assume that all of the individuals who are infective at $t = 0$ have just become infected so that $Y(0) = Y_0(0)$ and $Y_\ell(0) = 0$ for $\ell = 1, \dots, r - 1$. Furthermore, there are no recovered individuals at $t = 0$, so that $Z(0) = 0$ and the population is made up only of susceptible and just infected individuals. So, $Y_0(0) + X(0) = n$.

Regardless of the initial conditions, the process

$$M(t) = (X(t), Y_0(t), Y_1(t), \dots, Y_{r-1}(t)) \quad t = 0, 1, 2, \dots$$

is a Markov chain with state space

$$S(n, r) = \left\{ (x, y_0, \dots, y_{r-1}) \in \mathbb{N}^{r+1} \text{ s.t. } x + \sum_{\ell=0}^{r-1} y_\ell \leq n \right\}.$$

The cardinality of $S(n, r)$ is $\binom{n+r+1}{n}$. Observe that the values of $Z(t)$ are determined if all the components of $M(t)$ are known.

The total number of infectives at time t is given by

$$Y(t) = \sum_{\ell=0}^{r-1} Y_{\ell}(t),$$

so the set $(X(t), Y(t), Z(t))$ gives the traditional SIR description.

In addition to the process $Y^i = \{Y^i(t), t = 0, 1, 2, \dots\}$, we can define in a similar manner the process $X^i = \{X^i(t), t = 0, 1, 2, \dots\}$, for $i = 1, \dots, n$, which indicates whether individual i is susceptible or not and the variable $Z^i(t) = 1 - X^i(t) - Y^i(t)$ which indicates if the individual i has recovered from the disease and is no longer infectious. For all t , we get

$$X(t) = \sum_{i=1}^n X^i(t), \quad Y(t) = \sum_{i=1}^n Y^i(t) \quad \text{and} \quad Z(t) = \sum_{i=1}^n Z^i(t).$$

Furthermore, for $i = 1, \dots, n$, we can consider the processes $Y_{\ell}^i = \{Y_{\ell}^i(t)\}_{t \geq 0}$ for $\ell = 0, \dots, r-1$ where

$$Y_{\ell}^i(t) = \begin{cases} 1 & \text{if the individual } i \text{ at time } t \text{ is infective for } \ell \text{ days} \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$Y^i(t) = \sum_{\ell=0}^{r-1} Y_{\ell}^i(t) \quad \text{and} \quad Y_{\ell}(t) = \sum_{i=1}^n Y_{\ell}^i(t).$$

Then, we can consider another non-homogeneous Markov chain

$$\tilde{M}(t) = (X^i(t), Y_0^i(t), Y_1^i(t), \dots, Y_{r-1}^i(t), i = 1, \dots, n) \quad t = 0, 1, 2, \dots$$

with state space

$$\tilde{S}(n, r) = \left\{ (x^1, y_0^1, \dots, y_{r-1}^1, \dots, x^n, y_0^n, \dots, y_{r-1}^n) \in \{0, 1\}^{n(r+1)} \text{ s.t.} \right. \\ \left. \alpha^i = x^i + \sum_{\ell=0}^{r-1} y_{\ell}^i \leq 1 \text{ for } i = 1, \dots, n \text{ and } \sum_{i=1}^n \alpha^i \leq n \right\}.$$

The cardinality of $\tilde{S}(n, r)$ is equal to $(r+2)^n$. Even if the cardinality is bigger than the one of $S(n, r)$, this Markovian model is more simple for computation purposes.

2.1. Transition probabilities

For a fixed individual i , consider the process

$$\tilde{M}^i(t) = (X^i(t), Y_0^i(t), Y_1^i(t), \dots, Y_{r-1}^i(t)) \quad t = 0, 1, 2, \dots$$

If one of the variables $Y_0^i(t), Y_1^i(t), \dots, Y_{r-1}^i(t)$ is equal to 1 then the process at times bigger than t is certain since the transition in this case is sure. The only interesting case is when the individual is susceptible at time t , that is, $X^i(t) = 1$.

Now we are interested in calculating the probability that an individual i susceptible at t becomes infected for the first time at $t + 1$. This probability depends on the total number $Y(t) = y$ of diseased individuals together with the probability p of transmission per contact and the parameter $\lambda \in [0, 1]$ which characterizes a possible quarantine. It is calculated taking into account all possible values of $N^i(t)$ and its probabilities. Recall that the variables $N^i(t)$ are identically distributed for all individuals but not with respect to time, so we will use the notion $N(t)$ instead. Assuming n is much greater than $N(t)$, so that the binomial approximation may be used, the probability of meeting exactly j infectives if $N(t) = n_k$ individuals are met is

$$p_j^i(y, n_k; n) \approx \binom{n_k}{j} \left(\frac{\lambda y}{n-1} \right)^j \left(1 - \frac{\lambda y}{n-1} \right)^{n_k-j}, \quad (2.1)$$

when $y < n - 1$, while $p_{n_k}^i(y, n_k; n) = 1$ and $p_j^i(y, n_k; n) = 0$ for $j = 0, \dots, n_k - 1$ when $y = n - 1$. The probability p_j of becoming infected if j infectives are met is

$$p_j = 1 - (1 - p)^j. \quad (2.2)$$

Then, using (2.1) as an equality the probability that an individual i susceptible at t becomes infected for the first time at $t + 1$ is

$$\begin{aligned} p(t, y) &= P(Y_0^i(t+1) = 1 | X^i(t) = 1, Y(t) = y) \\ &= \sum_{k=1}^m P(Y_0^i(t+1) = 1 | X^i(t) = 1, Y(t) = y, N(t) = n_k) P(N(t) = n_k) \\ &= \sum_{k=1}^m \sum_{j=0}^{n_k} p_j^i(y, n_k; n) p_j p_k(t) \\ &= \sum_{k=1}^m \sum_{j=0}^{n_k} \binom{n_k}{j} \left(\frac{\lambda y}{n-1} \right)^j \left(1 - \frac{\lambda y}{n-1} \right)^{n_k-j} [1 - (1 - p)^j] p_k(t) \\ &= \sum_{k=1}^m p_k(t) \sum_{j=0}^{n_k} \binom{n_k}{j} \left(\frac{\lambda y}{n-1} \right)^j \left(1 - \frac{\lambda y}{n-1} \right)^{n_k-j} \\ &\quad + \sum_{k=1}^m p_k(t) \sum_{j=0}^{n_k} \binom{n_k}{j} \left(\frac{(1-p)\lambda y}{n-1} \right)^j \left(1 - \frac{\lambda y}{n-1} \right)^{n_k-j} \\ &= 1 - \sum_{k=1}^m \left(1 - \frac{p\lambda y}{n-1} \right)^{n_k} p_k(t). \end{aligned} \quad (2.3)$$

when $y < n - 1$, while $p(t, n - 1) = 1 - \sum_{k=1}^m (1 - p)^{n_k} p_k(t)$.

Note that, as commonly used, this model contains a simplification regarding the encounters between individuals: the meeting relationship is not symmetric because if the group randomly chosen to meet individual i contains individual j , the group chosen to meet individual j does not necessarily contain individual i .

3. Calculus on the model

After presenting our model, we are interested in describing its dynamics. We study the distribution of new infective individuals and the basic reproduction number. Then, we prove that our model can

approximate a diffusion process. We consider two cases. The first is a general one where any individual remains infectious for $r < \infty$ consecutive epochs. Here r is a fixed positive constant. We refer to this model as the one with recovery. The second is the particular case without recovery that is when $r = \infty$.

3.1. The model with recovery ($r < \infty$)

Recall that the processes $X(t)$, $Y(t)$ and $Z(t)$ represent the number of susceptible, infective and (previously infected and) recovered individuals at time t , respectively. As their sum is fixed and corresponds to the size of the population, to study the number of new infective individuals it is sufficient to know just two of these quantities. We define $V(t)$ as the number of non susceptible individuals at time t , that is $V(t) = Y(t) + Z(t)$. It is useful to observe that the processes $Y(t)$ and $Z(t)$ can be expressed in terms of $V(t)$:

$$Y(t) = V(t) - V(t - r) \quad \text{and} \quad Z(t) = V(t - r).$$

The number of new infectives at time $t + 1$ is given by $V(t + 1) - V(t)$. So, the total number of infectives at time $t + 1$ is given by the number of new infectives and the number of infectives at t who have not yet recovered at $t + 1$.

Then, following the previous computations we have

$$\begin{aligned} P(V(t + 1) = w + y + z | Y(t) = y, Z(t) = z) \\ &= P(V(t + 1) = w + y + z | V(t) - V(t - r) = y, V(t - r) = z) \\ &= \binom{n - y - z}{w} p(t, y)^w (1 - p(t, y))^{n - y - z - w}. \end{aligned}$$

Therefore, the distribution of the increment in the number of infectives follows a binomial law:

$$V(t + 1) - V(t) | Y(t) = y, Z(t) = z \sim \text{Binom}(n - y - z, p(t, y)).$$

When $y < n - 1$, its mean and variance are

$$\mathbb{E}[V(t + 1) - V(t) | Y(t) = y, Z(t) = z] = (n - y - z) \left[1 - \sum_{k=1}^m \left(1 - \frac{p\lambda y}{n - 1} \right)^{n_k} p_k(t) \right]$$

and

$$\begin{aligned} \text{Var}[V(t + 1) - V(t) | Y(t) = y, Z(t) = z] \\ &= (n - y - z) \left[1 - \sum_{k=1}^m \left(1 - \frac{p\lambda y}{n - 1} \right)^{n_k} p_k(t) \right] \left[\sum_{k=1}^m \left(1 - \frac{p\lambda y}{n - 1} \right)^{n_k} p_k(t) \right]. \end{aligned}$$

In the extreme case when $y = n - 1$, z can take only two possible values. If $z = 1$, it means that there are no more susceptible individuals in the population and the number of new infectives is constantly zero. If $z = 0$, all individuals are infective except one. Then, the distribution of the increment in the number of infectives corresponds to the infection of the unique susceptible and follows a Bernoulli law with parameter $p(t, n - 1)$. Its means and variance are easily deduced.

3.1.1. The basic reproduction number R_0

The basic reproduction number R_0 is the expected number of secondary cases produced by an infective individual during its period of infectiousness in a population where all individuals are susceptible to infection. R_0 excludes new cases produced by the secondary cases.

From a theoretical point of view, it plays a vital role in the analysis of infectious disease models. Given a population in a stable demographic state with no history of a given infection, it provides insight as to whether the introduction of the infectious agent will cause an outbreak. In this sense, the basic reproduction number is used to measure the transmission potential of a disease as an epidemic occurs in a susceptible population only if $R_0 > 1$ (see [4] for additional details).

Our idea consists of estimating the basic reproduction number by observing the role of the infective individual and determining how many people he or she infects at any time step. A different approach has been used, for example, in [13] where the estimation is based on the probability of a susceptible individual being infected by the tagged one. In other cases, it can be estimated computing other related metrics such as the effective reproductive number (see, for example, [16]). We first consider the case without quarantine and then deduce the basic reproduction number when a quarantine for infectious individuals has been established.

In accordance with the definition of the basic reproduction number, we assume that at time $t = 0$ the number of susceptible individuals is $X(0) = n - 1$ and that the number of infected ones is $Y(0) = 1$. Clearly, $Z(0) = 0$.

Suppose first that there is no quarantine, that is $\lambda = 1$. We denote by $I(t)$ the number of individuals infected by our tagged individual during only the t -th period and by $S(t)$ the number of susceptible individuals met by one person at time t . We assume that a susceptible individual, met and infected by another (for example, the tagged one) at time t , is considered infective since time $t + 1$. Since our tagged individual remains infectious for r consecutive days (from $t = 0$ to $t = r - 1$), the basic reproduction number is given by

$$R_0 = \sum_{t=1}^r \mathbb{E}[I(t)].$$

For all t , $I(t)$ depends on the number $S(t - 1)$ of susceptible individuals met by the tagged one at time $t - 1$ and the probability of transmission p .

Since at time $t = 0$ all individuals except the tagged one are susceptible, $S(0) = N(0)$ and the distribution of $I(1)$ is quite simple to determine:

$$I(1)|N(0) = n_k \sim \text{Binom}(n_k, p).$$

Consequently,

$$\mathbb{E}[I(1)] = \mathbb{E}[\mathbb{E}[I(1)|N(0)]] = p \mathbb{E}[N(0)].$$

For all $t \in \{1, \dots, r - 1\}$, the variable $S(t)$ may not coincide with $N(t)$ and the distribution of $I(t + 1)$ is given by

$$I(t + 1)|S(t) = l \sim \text{Binom}(l, p). \quad (3.1)$$

Moreover, for all $t \in \{1, \dots, r - 1\}$ $S(t)$ depends on the number of individuals met by a single person and on the number of infective in the total population. Observe that during this period the number of

removed individuals is always zero since our tagged one will be the first removed individual at time $t = r$. $S(t)$ follows a hypergeometric distribution:

$$S(t)|N(t) = n_k, Y(t) = y \sim HGeom(n - 1, n - y, n_k).$$

Observe that we have the following parameters: $n - 1$ is the size of the population without counting the tagged individual; $n - y$ represents the number of susceptible individuals in the population; and n_k is the number of daily contacts for one individual. As n is supposed to be much greater than $N(t)$, we use the binomial approximation

$$S(t)|N(t) = n_k, Y(t) = y \approx Binom(n_k, p(y)), \quad (3.2)$$

where

$$p(y) = 1 - \frac{y - 1}{n - 1} \quad (3.3)$$

is the proportion of susceptible individuals in the population (without counting the tagged individual).

Observe that the total number of infectives at a time step is given by the total number of infectives of the previous generation and the new infectives they have produced. That leads to:

$$\begin{aligned} \mathbb{E}[Y(t)] &= \mathbb{E}[\mathbb{E}[Y(t)|Y(t-1), S(t-1)]] \\ &= \mathbb{E}[Y(t-1) + pS(t-1)Y(t-1)] \\ &= \mathbb{E}[(1 + pS(t-1))Y(t-1)]. \end{aligned}$$

With inductive arguments, we obtain

$$\mathbb{E}[Y(t)] = \mathbb{E}\left[\prod_{s=0}^{t-1} (1 + pS(s))\right]. \quad (3.4)$$

We can find another expression for the last term that is more manageable for our purpose. Before giving it explicitly, we present a basic example to better understand the underlying idea. With the notation employed in previous sections, using that, for $t \in \{1, \dots, r-1\}$, $Y(t) = \sum_{\ell=0}^{t-1} Y_\ell(t) = \sum_{\ell=0}^t Y_\ell(t)$ since $Y_\ell(t) = 0$ for $\ell > t$ and also that $Y_t(t) = 1$ refers to the tagged individual, for $t = 3$ we have

$$\begin{aligned} \mathbb{E}[Y(3)] &= \mathbb{E}[Y_3(3) + Y_2(3) + Y_1(3) + Y_0(3)] \\ &= \mathbb{E}[1 + pS(0) + pS(1)(1 + pS(0)) + pS(2)(1 + pS(0) + pS(1)(1 + pS(0)))] \\ &= \mathbb{E}\left[1 + pS(0) + pS(1) + pS(2) + p^2S(0)S(1) + p^2S(0)S(2) + p^2S(1)S(2) \right. \\ &\quad \left. + p^3S(0)S(1)S(2)\right] \end{aligned}$$

that can be shortly expressed as

$$\mathbb{E}[Y(t)] = \mathbb{E}\left[\sum_{\substack{j_s \in \{0,1\} \\ s=0,1,2}} p^{j_0+j_1+j_2} \prod_{s=0}^2 S(s)^{j_s}\right].$$

Induction on the general case leads to the following expression equivalent to (3.4):

$$\mathbb{E}[Y(t)] = \mathbb{E} \left[\sum_{\substack{j_s \in \{0,1\} \\ s=0,\dots,t-1}} p^{j_0+\dots+j_{t-1}} \prod_{s=0}^{t-1} S(s)^{j_s} \right]. \quad (3.5)$$

If we define $\psi(t) = \prod_{s=0}^t (1 + pS(s))$, the induction is based on the decomposition

$$\psi(t) = \psi(t-1) (1 + pS(t))$$

and the fact that the last term can be written as

$$(1 + pS(t)) = \sum_{j_t \in \{0,1\}} p^{j_t} S(s)^{j_t}.$$

Equations (3.3) and (3.5) lead to

$$\mathbb{E}[p(Y(t))] = 1 + O\left(\frac{1}{n-1}\right). \quad (3.6)$$

Then, from (3.1), (3.2) and (3.6) for all $t \in \{1, \dots, r-1\}$ we have

$$\begin{aligned} \mathbb{E}[I(t+1)] &= \mathbb{E}[\mathbb{E}[I(t+1)|S(t)]] = p \mathbb{E}[S(t)] \\ &= p \mathbb{E}[\mathbb{E}[S(t)|N(t), Y(t)]] \\ &= p \mathbb{E}[N(t) \cdot p(Y(t))] \\ &= p \mathbb{E}[N(t)] + O\left(\frac{1}{n-1}\right). \end{aligned}$$

In the last step, we use the assumption that the number of daily contacts does not depend on the state of the population which implies that $N(t)$ is independent of $Y(t)$.

Finally, we get

$$R_0 = \sum_{t=1}^r \mathbb{E}[I(t)] = p \sum_{t=0}^{r-1} \mathbb{E}[N(t)] + O\left(\frac{1}{n-1}\right).$$

If we assume that there is a quarantine, recall that the probability that a randomly chosen individual is infectious is multiplied by a parameter $\lambda \in [0, 1]$. Thinking about the role of this parameter, we see that it also affects the contagiousness of an infective individual which is also multiplied by the same factor. This means that the basic reproduction number is given by

$$R_0 = \lambda \sum_{t=1}^r \mathbb{E}[I(t)].$$

The same calculations used for the case without quarantine lead to the following expression for R_0 :

$$R_0 = \lambda p \sum_{t=0}^{r-1} \mathbb{E}[N(t)] + O\left(\frac{1}{n-1}\right). \quad (3.7)$$

As the number of susceptibles met by the tagged individual is always less or equal to the number of his or her encounters, we can deduce an upper bound for the expected value of R_0 :

$$R_0 \leq \lambda p \sum_{t=0}^{r-1} \mathbb{E}[N(t)].$$

From (3.7) it follows that, since $R_0 \approx 1$ for $\lambda p \sum_{t=0}^{r-1} \mathbb{E}[N(t)] = 1$, this is a threshold for the epidemic. This means that it grows for larger values and dies soon for smaller ones.

Finally, note that R_0 is closely related to the number of daily encounters that occur during the first r days of the spread of the disease. Since this parameter depends on time, its value could change significantly if the disease arises in a different period as will be shown in Section 4.1.1.

3.1.2. A diffusion approximation

Following the ideas of Tuckwell and Williams [12], the study of the mean and variance of the one-step increments of V indicates that for a large population size n and a small probability transition p such that $np\mathbb{E}[N([nt])]$ is of moderate size for all t , we can approximate a rescaled version of V by a diffusion process.

More accurately, if we speed up time and rescale state we can define a process

$$\widehat{V}^n(t) = \frac{V([nt])}{n} \quad \text{for all } t \geq 0,$$

where $[\cdot]$ denotes the greatest integer part. $\widehat{V}^n(t)$ can be interpreted as the fraction of the population that has been infected by the time $[nt]$ in the original time scale of V . Then, for n large and p small such that $\theta(t) = np\mathbb{E}[N([nt])]$ is of moderate size for all t , we see that with $\Delta t = \frac{1}{n}$ and $t = 0, \frac{1}{n}, \frac{2}{n}, \dots$

$$\begin{aligned} & \mathbb{E} \left[\widehat{V}^n(t + \Delta t) - \widehat{V}^n(t) \mid \widehat{V}^n(t) - \widehat{V}^n \left(t - \frac{r}{n} \right) = \hat{y}, \widehat{V}^n \left(t - \frac{r}{n} \right) = \hat{z} \right] \\ &= \frac{1}{n} \mathbb{E} \left[V([nt] + 1) - V([nt]) \mid V([nt]) - V([nt] - r) = n\hat{y}, V([nt] - r) = n\hat{z} \right] \\ &= \frac{1}{n} (n - n\hat{y} - n\hat{z}) \left[1 - \sum_{k=1}^m \left(1 - \frac{\lambda np \hat{y}}{n-1} \right)^{n_k} p_k([nt]) \right] \\ &\approx (1 - \hat{y} - \hat{z}) \frac{\lambda np \hat{y}}{n-1} \mathbb{E}[N([nt])] \\ &\approx \lambda \theta(t) \hat{y} (1 - \hat{y} - \hat{z}) \Delta t \end{aligned}$$

and

$$\begin{aligned} & \text{Var} \left[\widehat{V}^n(t + \Delta t) - \widehat{V}^n(t) \mid \widehat{V}^n(t) - \widehat{V}^n \left(t - \frac{r}{n} \right) = \hat{y}, \widehat{V}^n \left(t - \frac{r}{n} \right) = \hat{z} \right] \\ &= \frac{1}{n^2} \text{Var} \left[V([nt] + 1) - V([nt]) \mid V([nt]) - V([nt] - r) = n\hat{y}, V([nt] - r) = n\hat{z} \right] \\ &= \frac{1}{n^2} (n - n\hat{y} - n\hat{z}) \left[1 - \sum_{k=1}^m \left(1 - \frac{\lambda np \hat{y}}{n-1} \right)^{n_k} p_k([nt]) \right] \left[\sum_{k=1}^m \left(1 - \frac{\lambda np \hat{y}}{n-1} \right)^{n_k} p_k([nt]) \right] \end{aligned}$$

$$\begin{aligned}
&\approx \frac{1}{n}(1 - \hat{y} - \hat{z}) \left[\frac{\lambda n p \hat{y}}{n-1} \mathbb{E}[N([nt])] \right] \left[1 - \frac{\lambda n p \hat{y}}{n-1} \mathbb{E}[N([nt])] \right] \\
&\approx \frac{1}{n}(1 - \hat{y} - \hat{z}) \frac{\lambda n p \hat{y}}{n-1} \mathbb{E}[N([nt])] \\
&\approx \frac{\theta(t)\lambda}{n} \hat{y}(1 - \hat{y} - \hat{z}) \Delta t.
\end{aligned}$$

In both calculation we use the approximation $1 - (1 - x)^a \approx ax$ for small x . In the calculation of the variance we also use the approximation $x(1 - x) \approx x$ for small x .

Moreover, we recall that $\mathbb{E}[N([nt])]$ depends on the values n_1, n_2, \dots, n_m and its probabilities $p_k(t)$, $k = 1, \dots, m$. In fact, the time dependence on $\mathbb{E}[N([nt])]$ rely on $p_k(t) \in [0, 1]$, $k = 1, \dots, m$. So, we can assume for all n and t that $\mathbb{E}[N([nt])]$ is of order of a constant N , so

$$\theta(t) = np\mathbb{E}[N([nt])] \approx npN := \theta.$$

With the previous results and approximation methods for continuous time Markov chains using diffusion processes (similar results to [13] and [12]), we can approximate \widehat{V}^n by a diffusion process \widehat{V} in $[0, 1]$ that satisfies the stochastic delay differential equation (SDDE)

$$d\widehat{V}(t) = \lambda\theta(\widehat{V}(t) - \widehat{V}(t - \tau))(1 - \widehat{V}(t))dt + \left[\frac{\lambda\theta}{n} (\widehat{V}(t) - \widehat{V}(t - \tau))(1 - \widehat{V}(t)) \right]^{\frac{1}{2}} dW(t), \quad (3.8)$$

where $\tau = \frac{r}{n}$ and $W = \{W(t), t \geq 0\}$ is a standard Brownian motion. This equation has a unique solution for a given initial condition \widehat{V}_0 which is a non-decreasing continuous function $\widehat{V}_0 : [-\tau, 0] \rightarrow [0, 1]$ (see [17] for more references in SDDE).

Moreover, if we can assume that the variability is small then the noise term in (3.8) has little effect. So, it seems natural to conjecture that the mean of $\widehat{V}(t)$ can be approximated by $\widehat{m}(t)$ where \widehat{m} satisfies the deterministic equation

$$d\widehat{m}(t) = \lambda\theta(\widehat{m}(t) - \widehat{m}(t - \tau))(1 - \widehat{m}(t)) dt.$$

We can interpret the solution of the equation $\widehat{m}(t)$ as the proportion of the number of infective and recovered individuals. The general solution of this equation cannot be obtained using the classic methods of delay equations. However, it is easy to check that all constants $c \in [0, 1]$ are solutions. Also, this equation presents two fixed points. One is $\widehat{m}(t) = 1$, when the whole population is already infective or recovered. The other is $\widehat{m}(t) = \widehat{m}(t - \tau)$, meaning that we do not have new infected in an interval of length τ . Returning to the initial process without rescaling, it infers that for a period of length r the number of infective and recovered remains constant. That implies there will be no more infective individuals and the distribution of the population in the three classes will be invariant.

3.2. The model without recovery ($r = \infty$)

Consider now the case in which the infectious individuals remain infectious throughout the course of the epidemic. Such a situation can arise when a disease causing agent has a long life, as with tuberculosis in deer [18], or when life-prolonging drug therapies have been found, as with HIV in humans [19]. As recovered individuals are not presented in this case, the model reduces to one of SI type rather than SIR.

Here, only two classes of individuals are present: the susceptibles and the infectives. Their amount at time t is given by the variables $X(t)$ and $Y(t)$, respectively.

The distribution of new infective individuals only depends on the number of infectives of the previous generation. Assuming that at time t there are $Y(t) = y$ infectives implies that there are $n - y$ susceptibles since the population size is n and there are no recovered individuals. Recall that the distribution of daily encounters at time t , given by the variables $N(t)$, is the same for all individuals. Moreover, $N(t)$ are independent for all i and t . In this case, λ parameter will be fixed $\lambda = 1$ (no quarantine).

To study the distribution of new infectives, we follow the same ideas of the case with recovery with the difference that here it is not necessary to define a new variable that represents the number of non-susceptible individuals, as this is given by $Y(t)$. Given $Y(t) = y$, the number of new infectives at $t + 1$ follows a binomial distribution whose parameters are the number of susceptibles at t and the probability that an individual susceptible at t becomes infected for the first time at $t + 1$. This is

$$Y(t + 1) - Y(t) | Y(t) = y \sim \text{Binom}(n - y, p(t, y)),$$

where

$$p(t, y) = 1 - \sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)$$

is given by (2.3). This implies that the distribution of the increment in the number of infectives is given by

$$\begin{aligned} P(Y(t + 1) = y + w | Y(t) = y) \\ = \binom{n-y}{w} \left[1 - \sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)\right]^w \left[\sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)\right]^{n-y-w} \end{aligned} \quad (3.9)$$

for $w = 0, \dots, n - y$. Its mean is

$$\mathbb{E}[Y(t + 1) - Y(t) | Y(t) = y] = (n - y) \left[1 - \sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)\right]$$

and its variance is

$$\begin{aligned} \text{Var}[Y(t + 1) - Y(t) | Y(t) = y] \\ = (n - y) \left[1 - \sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)\right] \left[\sum_{k=1}^m \left(1 - \frac{py}{n-1}\right)^{n_k} p_k(t)\right]. \end{aligned}$$

3.2.1. The basic reproduction number R_0

In this section we calculate the value of the basic reproduction number in the case without recovery which we denote by $R_0^{(\infty)}$. The computation for the case with recovery can be adapted to this model.

Here, our tagged individual is infectious all his or her life but we can suppose that he or she lives only for a certain amount T of days after being infected. Thus, the calculation of the basic reproduction

number is the same as that of the case with recovery, changing the number of terms from r to T :

$$R_0^{(\infty)} = \sum_{t=1}^T \mathbb{E}[I(t)] = p \sum_{t=0}^{T-1} \mathbb{E}[N(t)] + O\left(\frac{1}{n-1}\right).$$

Recall that $I(t)$ denotes the number of individuals infected by our tagged individual during only the t -th period and observe that in this case no quarantine is established for infective individuals.

The considerations about an upper bound for $R_0^{(\infty)}$ and a threshold for the epidemic can be done also in the case without recovery. On one hand, as the number of susceptibles met by the tagged one is always less or equal to the number of his or her encounters, an upper bound for $R_0^{(\infty)}$ is:

$$R_0^{(\infty)} \leq p \sum_{t=1}^{T-1} \mathbb{E}[N(t)].$$

On the other hand, since $R_0^{(\infty)} \approx 1$ for $p \sum_{t=1}^{T-1} \mathbb{E}[N(t)] = 1$, this is a threshold for the epidemic.

3.2.2. A diffusion approximation

Following the ideas of the case with recovery, the study of the mean and variance of the one-step increments of Y indicates that for a large population size n and a small probability transition p such that $np\mathbb{E}[N([nt])]$ is of moderate size for all t fixed, we can approximate a rescaled version of Y by a diffusion process.

More accurately, if we speed up time and rescale state we can define a process

$$\widehat{Y}^n(t) = \frac{Y([nt])}{n} \quad \text{for all } t \geq 0$$

where $[\cdot]$ denotes the greatest integer part. Then, we can interpret $\widehat{Y}^n(t)$ as the fraction of the population that has been infected by the time $[nt]$ in the original time scale of Y . Then, for n large and p small such that $\theta(t) = np\mathbb{E}[N([nt])]$ is of moderate size for all t we see that with $\Delta t = \frac{1}{n}$ and $t = 0, \frac{1}{n}, \frac{2}{n}, \dots$

$$\begin{aligned} \mathbb{E}[\widehat{Y}^n(t + \Delta t) - \widehat{Y}^n(t) | \widehat{Y}^n(t) = \hat{y}] &= \frac{1}{n}(n - n\hat{y}) \left[1 - \sum_{k=1}^m \left(1 - \frac{np\hat{y}}{n-1} \right)^{n_k} p_k(t) \right] \\ &\approx \frac{1}{n}(n - n\hat{y}) \frac{np\hat{y}}{n-1} \mathbb{E}[N([nt])] \\ &\approx p\hat{y}(1 - \hat{y}) \mathbb{E}[N([nt])] = \theta(t)\hat{y}(1 - \hat{y})\Delta t \end{aligned}$$

using the approximation $1 - (1 - x)^a \approx ax$ for small x . With the same procedure, we have

$$\begin{aligned} \text{Var}[\widehat{Y}^n(t + \Delta t) - \widehat{Y}^n(t) | \widehat{Y}^n(t) = \hat{y}] &= \frac{1}{n^2} \text{Var}[Y([nt] + 1) - Y([nt]) | Y([nt]) = n\hat{y}] \\ &= \frac{1}{n^2}(n - n\hat{y}) \left[1 - \sum_{k=1}^m \left(1 - \frac{np\hat{y}}{n-1} \right)^{n_k} p_k(t) \right] \left[\sum_{k=1}^m \left(1 - \frac{np\hat{y}}{n-1} \right)^{n_k} p_k(t) \right] \\ &\approx \frac{1}{n}(1 - \hat{y}) \left[1 - \frac{np\hat{y}}{n-1} \mathbb{E}[N([nt])] \right] \frac{np\hat{y}}{n-1} \mathbb{E}[N([nt])] \end{aligned}$$

$$\approx \frac{1}{n}p(1 - \hat{y})\hat{y}\mathbb{E}[N([nt])] = \frac{\theta(t)}{n}\hat{y}(1 - \hat{y})\Delta t.$$

Using as before the approximation $1 - (1 - x)^a \approx ax$ for small x and $p(1 - p) \approx p$ for small p . As in the case with recovery ($r < \infty$), following the same steps we can assume

$$\theta(t) = np\mathbb{E}[N([nt])] \approx npN := \theta$$

for a constant N .

With this results and approximation methods for continuous time Markov chains using diffusion processes we can approximate \widehat{Y}^n by a diffusion process \widehat{Y} in $[0, 1]$ that satisfies the stochastic differential equation

$$d\widehat{Y}(t) = \theta\widehat{Y}(t)(1 - \widehat{Y}(t))dt + \sqrt{\frac{\theta}{n}\widehat{Y}(t)(1 - \widehat{Y}(t))}dW(t) \quad (3.10)$$

where $W = \{W(t), t \geq 0\}$ is a standard Brownian motion. The approximation of diffusion processes by Markovian chains is well explained in [20] (Chapter 10).

If the variability is assumed to be small then the noise term of the stochastic differential equation (3.10) will have little effect. In that case, we can presume that the mean of $\widehat{Y}(t)$ can be approximated by $\widehat{m}(t)$ where \widehat{m} satisfies the deterministic equation

$$d\widehat{m}(t) = \theta\widehat{m}(t)(1 - \widehat{m}(t))dt$$

with explicit solution

$$\widehat{m}(t) = \frac{1}{1 + \frac{1 - \widehat{m}_0}{\widehat{m}_0}e^{-\theta t}}, \quad t \geq 0$$

where $\widehat{m}_0 = \mathbb{E}[\widehat{Y}^n(0)] = \frac{1}{n}\mathbb{E}[Y(0)]$.

This suggests that for $t = 0, 1, 2, \dots$ (with attendant scaling up of error terms),

$$\mathbb{E}[Y(t)] = n\mathbb{E}\left[\widehat{Y}^n\left(\frac{t}{n}\right)\right] \approx n\widehat{m}\left(\frac{t}{n}\right) = \frac{n}{1 + \frac{n - y_0}{y_0}e^{-pNt}}$$

where $y_0 = \mathbb{E}[Y(0)]$.

4. Simulations

The aim of this chapter is to simulate the SIR epidemic model we have presented in the previous sections, comparing the evolution of the disease for different distributions of the number of daily contacts $N(t)$. Particularly, we study the cases when $N(t)$ follows two distributions: binomial and truncated Poisson. We will compare the results for different values of their parameters.

Since we are interested in showing the different evolution of the SIR epidemic model as the number of daily contacts changes, we fix the other parameters. We consider a population of $n = 10000$ individuals with a unique initial infective individual: $Y(0) = 1$. The probability of transmission derived from an encounter is $p = 0.1$ and the duration of the disease is $r = 5$. We assume the quarantine is nonexistent, that is, $\lambda = 1$.

For each case, we present the average of the duration of the disease (Total time), the total proportion of the infected population when the disease ends (Population infected) and the number of individuals infected by the first infective (\widetilde{R}_0). Also, $R_0^* = \lambda p \sum_{t=0}^{t-1} \mathbb{E}(N(t))$ is computed. The theoretic R_0 will be R_0^* with an error of $O\left(\frac{1}{n-1}\right)$.

The simulations are made using R programming environment. All of them are over 10,000 trials and run for a maximum of 365 days.

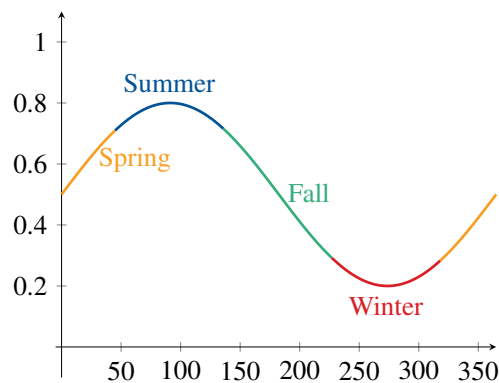
4.1. Binomial case

Here, the number of daily contacts $N(t)$ has a binomial distribution with probability depending on time. This is $N(t) \sim \text{Binom}(N, p(t))$. With this distribution, the set of all possible values of $N(t)$ is $M = \{0, \dots, N\}$.

We consider three cases. In all of them, the function $p(t)$ is periodic with yearly, monthly and weekly periods, respectively. In the first case, we fix a value for N and compare the spread of the disease for four different functions of $p(t)$. In the other two cases, we choose a function for $p(t)$ and observe how the disease evolves when the maximum number of daily contacts changes.

4.1.1. Yearly period

It is a natural assumption that the number of daily contacts varies all over the year and changes according to the season. We suppose that the mean of daily encounters arrives at its maximum in summer, decreases in fall until arriving at its minimum in winter and then increases in spring and summer again. Moreover, we assume that $p(t)$ has a sinusoidal shape where the peak coincides with summer and the valley with winter and we analyze the spread of the disease depending on the epoch of the year it starts. The nomenclature we use refers to this epoch. The following plot suggests the idea:



The four functions we consider are:

$$\begin{aligned} \text{Spring:} \quad p(t) &= 0.3 \cdot \sin\left(\frac{2\pi t}{365}\right) + 0.5 \\ \text{Summer:} \quad p(t) &= 0.3 \cdot \cos\left(\frac{2\pi t}{365}\right) + 0.5 \\ \text{Fall:} \quad p(t) &= -0.3 \cdot \sin\left(\frac{2\pi t}{365}\right) + 0.5 \end{aligned}$$

$$\text{Winter:} \quad p(t) = -0.3 \cdot \cos\left(\frac{2\pi t}{365}\right) + 0.5$$

All of them oscillate between 0.2 and 0.8. We fix the maximum number of daily contacts N at 10.

Figure 1 shows the evolution of the mean number of infectives per day. The graphic has been truncated at 110 days, as the number of cases of the following days is almost irrelevant. We have a fast spread of the disease in the *Summer* case which coincides with the time when the number of encounters is higher. By reducing this number the disease presents a lower mean number of infectives per day but it lasts longer. Observe that in *Spring* and *Fall* $p(t)$ has the same initial value $p(0) = 0.5$, but the model evolves differently depending on whether this probability increases or decreases.

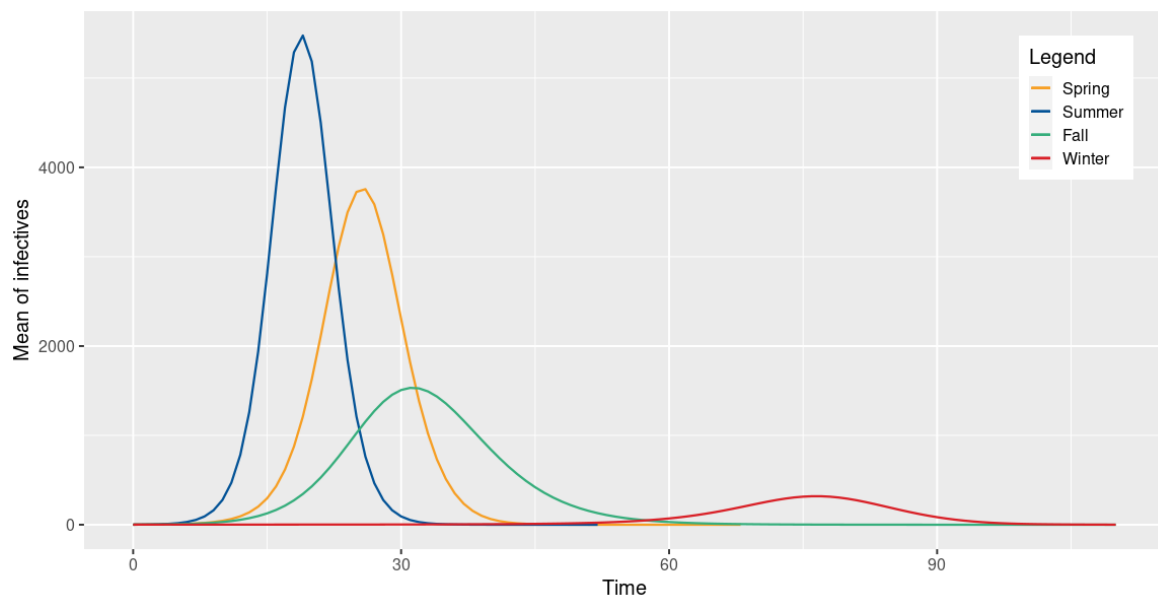


Figure 1. SIR epidemic model depending on the season the disease starts. Evolution of the mean number of infectives. $p(t)$ is year-periodic and varies for each case while N is fixed at $N = 10$.

All these facts are underlined in Figure 2 which represents the histogram of the duration of the disease (Left) and the histogram of the total number of infected in the population when the disease ends (Right). Moreover, it is interesting to notice that, while in *Spring* and *Summer* almost the entire population gets infected in most trials, in *Winter* the epidemic develops only in a limited number of cases.

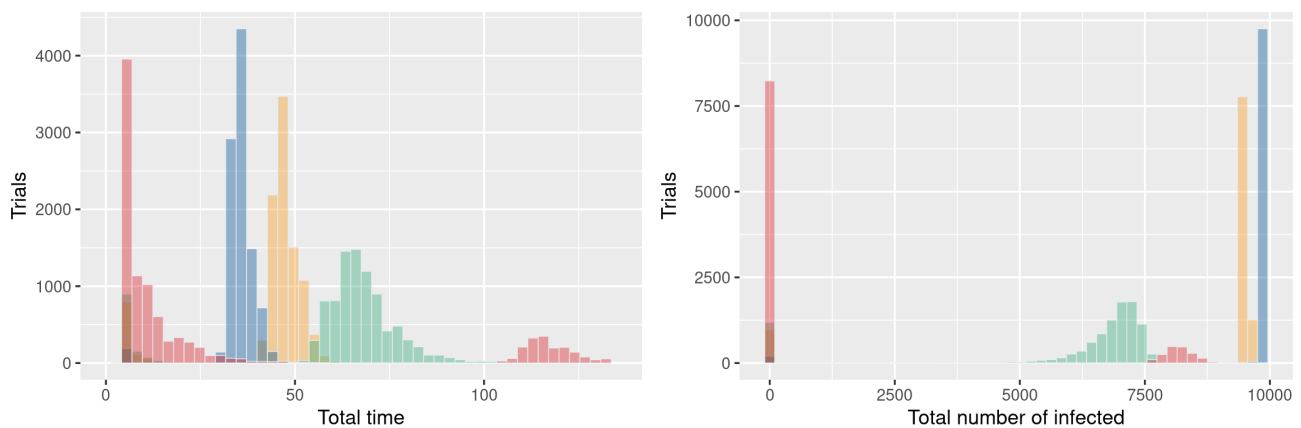


Figure 2. SIR epidemic model depending on the season the disease starts. $p(t)$ is year-periodic and varies for each case while N is fixed at $N = 10$. Top: Evolution of the mean number of infectives. Left: Histogram of the duration of the disease in days. Right: Histogram of the total number of infected in the population when the disease ends.

The first two rows of Table 1 help to better understand Figure 2 (Left and Right). The course of the disease is on average longer in *Fall* but it has a big variation and it is on average shorter in *Winter* with an even bigger dispersion. This is evident in Figure 2 (Left) where in *Winter* we notice two different behaviors: in most of the trials the disease lasts for a short time and does not exceed 50 days but in the other cases it lasts for more than 100 days. As remarked before, almost the entire population is infected in *Spring* and *Summer* but this proportion decreases drastically in *Fall* and *Winter*.

Table 1. SIR epidemic model depending on the season the disease starts. $p(t)$ is year-periodic and varies for each case while N is fixed at $N = 10$. Mean (standard deviation) of three variables: duration of the disease in days, total proportion of infecteds when the disease ends and number of individuals infected by the first infective (\widetilde{R}_0). R_0^* is computed for all the cases.

	<i>Spring</i>	<i>Summer</i>	<i>Fall</i>	<i>Winter</i>
Total time	43.54 (12.77)	35.33 (5.05)	60.83 (21.92)	29.50 (42.08)
Population infected (%)	86% (28%)	96% (14%)	61% (23%)	14% (31%)
\widetilde{R}_0	2.53 (1.26)	3.97 (1.43)	2.44 (1.25)	1.00 (0.90)
R_0^*	2.55	4.00	2.45	1.00

Table 1 also contains information regarding the number of individuals infected by the first infective (\widetilde{R}_0). If compared with R_0^* (last row) very similar values are observed. Recall that R_0^* is an approximation of the theoretical value R_0 . In the first three cases, \widetilde{R}_0 exceeds the threshold value 1 implying that the disease turns into an epidemic. In the last case, the mean of \widetilde{R}_0 coincides with the threshold itself but its variability makes it unclear if an epidemic occurs on average or not. Moreover,

observe that \widetilde{R}_0 is very similar in *Spring* and *Fall*.

Finally, notice that the dispersion is smaller in *Summer* and bigger in *Winter* when we refer to the duration of the disease and the total proportion of infected but it is reversed when we consider \widetilde{R}_0 .

4.1.2. Monthly period

We assume that the number of daily contacts changes periodically over the course of the month. This behavior could be due to economic reasons. For example, it might catch the tendency to go out more when a salary is perceived. The periodicity is reflected in $p(t)$, that is the sinusoidal function

$$p(t) = 0.3 \cdot \sin\left(\frac{2\pi t}{30}\right) + 0.5.$$

This probability oscillates between 0.2 and 0.8. We simulate and then compare the SIR epidemic model for three different values of the maximum number of daily contacts: $N = 7$, $N = 10$ and $N = 12$.

Figure 3 (Left) shows the evolution of the mean number of infectives per day in the three cases. The graphic has been truncated at 100 days. We see that the evolution of the disease behaves differently with respect to the maximum number of daily encounters. Particularly, the number and position of the peaks vary. When $N = 7$, there are three peaks, the central one big enough and the others very small. When the maximum number of daily contacts is set to $N = 10$, two similar peaks are displayed. Finally, when this parameter is increased to $N = 12$ a tall peak is followed by a small one.

It might be interesting to investigate how the mean number of infectives presents two peaks in the case $N = 10$. Figure 3 (Right) shows the evolution of the number of infectives per day for some selected simulations. Observe that the trials behave quite differently: in some cases there are two (similar or not) peaks and in others only one. However, it is also important to point out that all peaks occur at similar times as those of the mean.

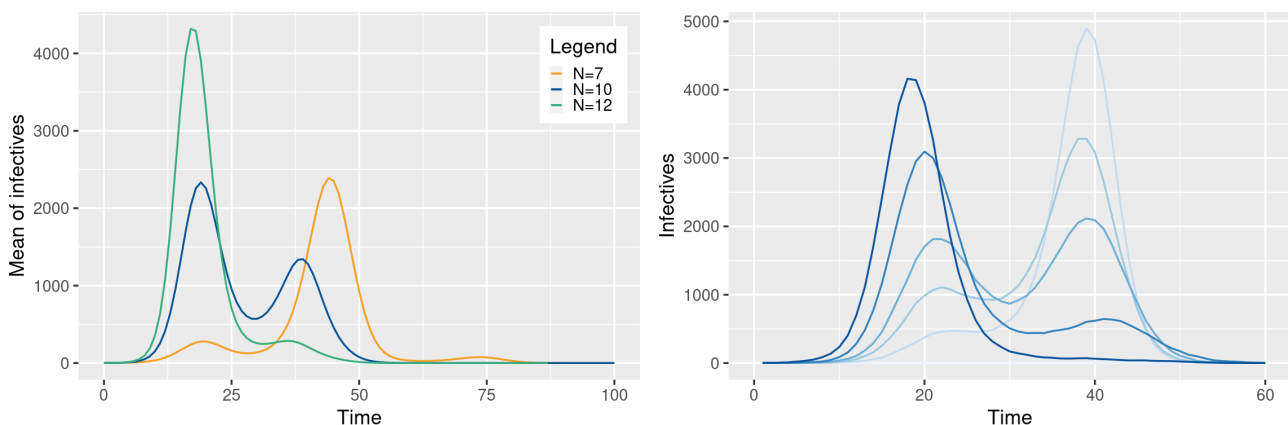


Figure 3. SIR epidemic model with $p(t)$ monthly-periodic. Left: Evolution of the mean number of infectives for three different values of N . Right: Evolution of the number of infectives in some simulations for $N = 10$.

It appears that on average the disease lasts longer when the number of daily contacts is smaller. This is noted in Figure 4 (Left) and Table 2. In addition, here we observe a bigger variation for smaller

values of N . For example, the standard deviation of the case $N = 7$ is more than double that of the case $N = 10$. This is also due to the fact that the disease does not spread in more simulations of the case $N = 7$ than of the others, as pointed out in Figure 4 (Left).

Figure 4 (Right) shows the distribution of the total number of infected over all simulations. When the number of contacts increases, this quantity is on average bigger but its variation decreases, as indicated in Table 2. Note that the cases in which the total population is infected are very few.

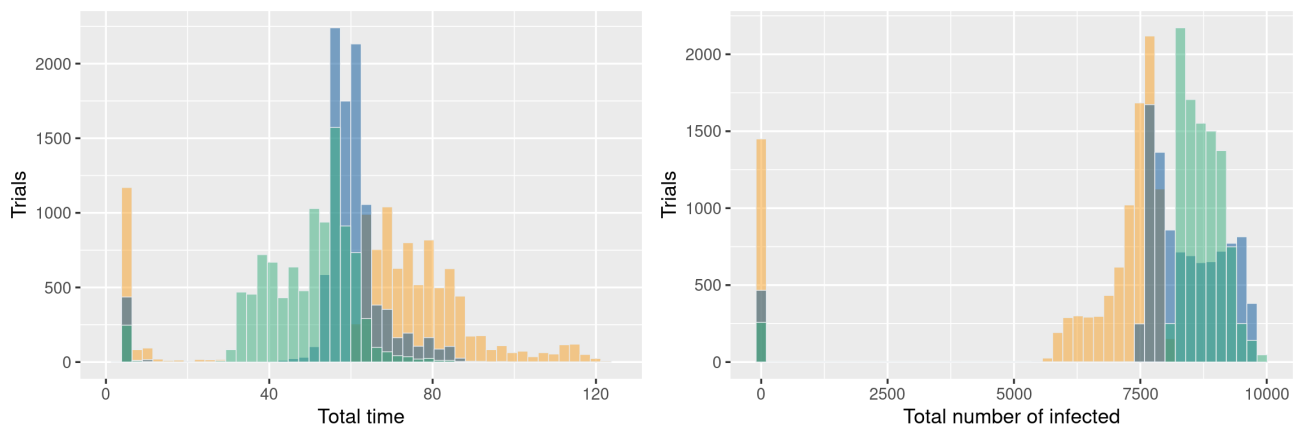


Figure 4. SIR epidemic model with $p(t)$ monthly-periodic. Comparison for $N = 7$, $N = 10$ and $N = 12$ (same colors of Figure 3). Left: Histogram of the duration of the disease in days. Right: Histogram of the total number of infected in the population when the disease ends.

Table 2. SIR epidemic model with $p(t)$ monthly-periodic and three different values of N . Mean (standard deviation) of three variables: duration of the disease in days, total proportion of infected when the disease ends and number of individuals infected by the first infective (\widetilde{R}_0). R_0^* is computed for all the cases.

	$N = 7$	$N = 10$	$N = 12$
Total time	67.14 (27.62)	58.40 (13.41)	49.18 (11.84)
Population infected (%)	63% (26%)	81% (19%)	85% (14%)
\widetilde{R}_0	2.17 (1.20)	3.07 (1.33)	3.70 (1.40)
R_0^*	2.16	3.08	3.70

Finally, Table 2 reveals that the mean number of individuals infected by the first infective increases as the maximum number of daily encounters does. Its variation follows the same tendency although the differences are quite small.

4.1.3. Weekly period

Now, we assume that the number of daily contacts changes periodically with a week period. For example, this could mark the different behavior of the population at the weekend with respect to the

weekdays. The periodicity is reflected in $p(t)$, that is the sinusoidal function

$$p(t) = 0.3 \cdot \cos\left(\frac{2\pi t}{7}\right) + 0.5.$$

This probability oscillates between 0.2 and 0.8. As in the monthly-periodic study, we simulate and then compare the SIR epidemic model for three different values of the maximum number of daily contacts: $N = 7$, $N = 10$ and $N = 12$.

Figure 5 (Left) plots the evolution of the mean number of infectives in the three cases. The graphic has been truncated at 100 days. One could observe that the results are very different depending on the values of N . When $N = 7$, the mean number of infectives oscillates many times presenting several peaks but none of them is particularly high. Some selected simulations are displayed in Figure 5 (Right). They seem to reflect the same tendency as the mean. When $N = 10$ and $N = 12$, only one peak is displayed but the curves are not similar to those of a normal distribution as they were for the yearly and monthly periodic studies. Furthermore, the different shapes the two of them presented make their behavior quite distinct.

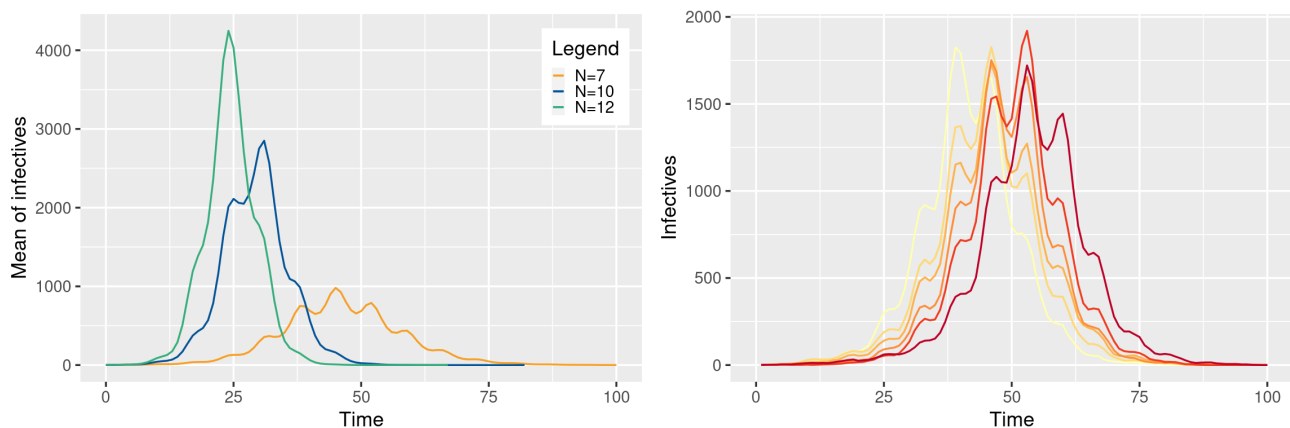


Figure 5. SIR epidemic model with $p(t)$ weekly-periodic. Left: Evolution of the mean number of infectives for three different values of N . Right: Evolution of the number of infectives in some simulations for $N = 7$.

The histograms in Figure 6 (Left and Right) also highlight the differences that the change in the number of daily contacts generates. We observe two features also reported in the monthly-periodic study. On one hand, the disease lasts longer on average when the number of daily contacts is smaller. On the other hand, the magnitude of the spread, intended as the part of the total population infected, decreases. Furthermore, note that when $N = 7$ the number of trials in which the disease does not spread is much higher than in the other two cases. This increases the variability of the results, as can be observed in Table 3.

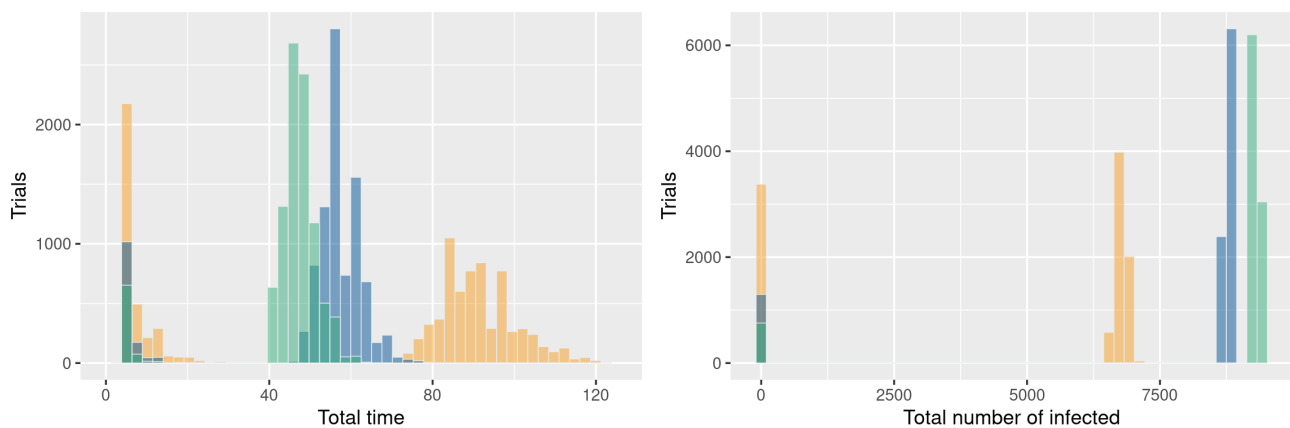


Figure 6. SIR epidemic model with $p(t)$ weekly-periodic. Comparison for $N = 7$, $N = 10$ and $N = 12$ (same colors of Figure 5). Left: Histogram of the duration of the disease in days. Right: Histogram of the total number of infected in the population when the disease ends.

Table 3. SIR epidemic model with $p(t)$ weekly-periodic and three different values of N . Mean (standard deviation) of three variables: duration of the disease in days, total proportion of infecteds when the disease ends and number of individuals infected by the first infective (\tilde{R}_0). R_0^* is computed for all the cases.

	$N = 7$	$N = 10$	$N = 12$
Total time	63.01 (40.58)	50.71 (17.87)	44.45 (11.70)
Population infected (%)	45% (32%)	76% (29%)	86% (25%)
\tilde{R}_0	1.67 (1.13)	2.39 (1.28)	2.85 (1.36)
R_0^*	1.67	2.38	2.86

The dispersion of the mean number of individuals infected by the first infective is also very high in all cases (see Table 3). Particularly, when $N = 7$ and $N = 10$ it is difficult to know if the spread of the disease will result in an epidemic or not.

4.2. Truncated Poisson case

Now, we assume that the number of daily contacts $N(t)$ follows a truncated Poisson distribution. This is the distribution of a Poisson random variable conditional on the fact that it takes values only in an interval. In our case, we fix a maximum number of daily contacts. We set this upper bound at 15 encounters.

We consider three different functions for the parameter λ depending on t :

$$\lambda_1(t) = \frac{10}{1 + \log(t + 1)}$$

$$\lambda_2(t) = \frac{50}{5 + 2\sqrt{t}}$$

$$\lambda_3(t) = -5 \sin\left(\frac{2\pi t}{15} + k\right) + 7 \quad \text{with} \quad k = \arcsin\left(-\frac{3}{5}\right)$$

where k is an adjustment constant so that all functions start at 10 when $t = 0$. Over the course of a year (365 days), $\lambda_1(t)$ takes values between 1.45 and 10, $\lambda_2(t)$ between 1.16 and 10 and $\lambda_3(t)$ between 2 and 12 (all these values are rounded to two decimals). We can see in Figure 7 (Right) the behavior of the three functions we have considered.

Figure 7 (Left) shows the evolution of the mean number of infectives per day. We notice important differences between the first two cases although the two parameters $\lambda_1(t)$ and $\lambda_2(t)$ present a similar evolution in time. The third case goes fast to zero despite the periodic behavior of the parameter $\lambda_3(t)$.

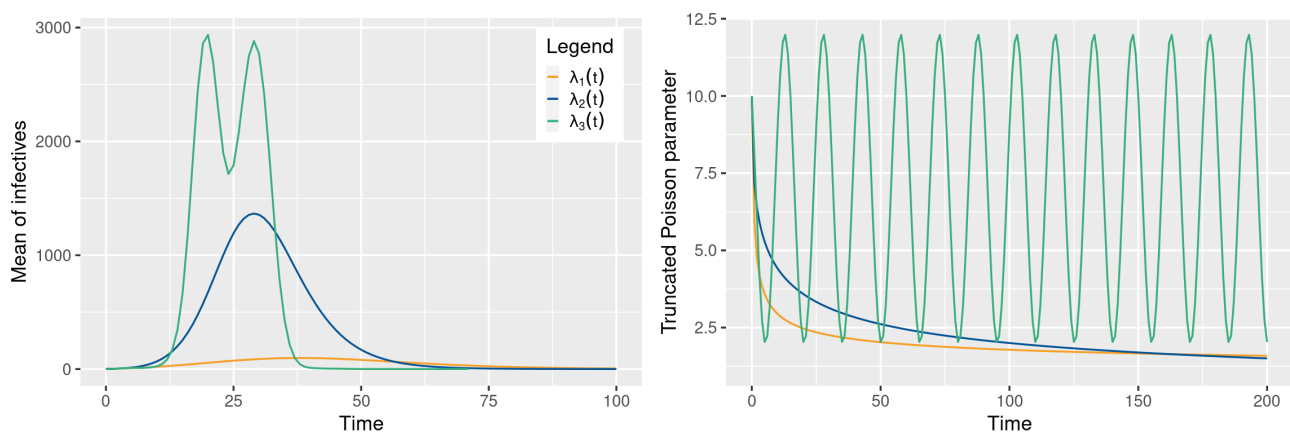


Figure 7. SIR epidemic model with number of contacts distributed as a truncated Poisson (upper bound 15). Left: Evolution of the mean number of infective. Right: Different $\lambda_i(t)$ considered.

From Figure 8 (Left) and Table 4 we can see that for the case $\lambda_1(t)$ there is a lot of dispersion of the total duration of the disease which takes values from 0 to 125 days. For the other two cases, the dispersion is much smaller and the duration of the case $\lambda_1(t)$ is the smallest one. Concerning the total number of infected, the biggest dispersion is presented by the case $\lambda_3(t)$ since there are almost 1,000 simulations that end without infecteds and much more with almost the entire population infected. The case $\lambda_1(t)$ is the one that presents a less infected population and also the one with less dispersion. In all three cases, \tilde{R}_0 exceeds the threshold value 1 so the disease turns into an epidemic. Moreover, its values are similar to the expected ones given by R_0^* .

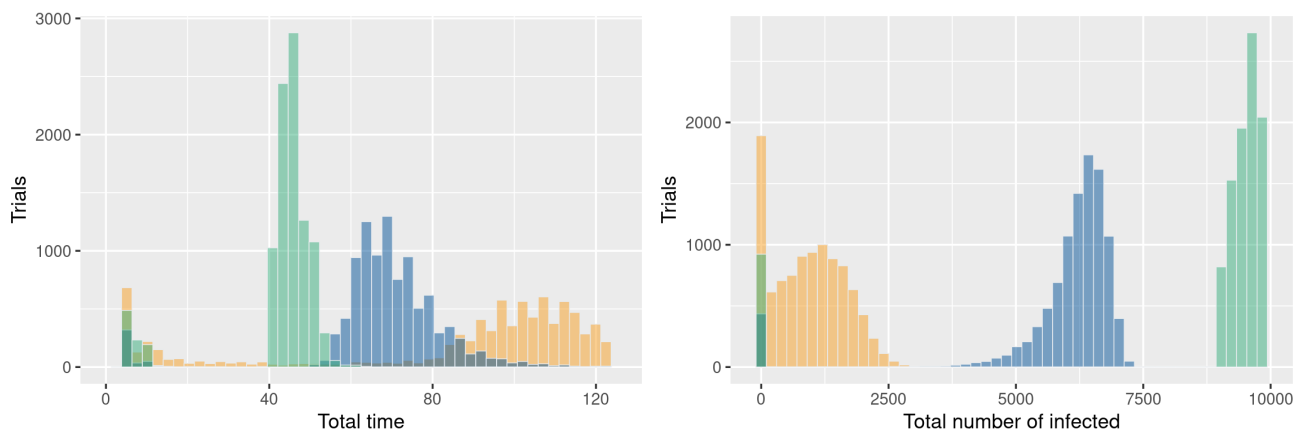


Figure 8. SIR epidemic model with number of contacts distributed as a truncated Poisson (upper bound 15). The colors refer to those used in Figure 7. Left: Histogram of the duration of the disease in days. Right: Histogram of the total number of infected in the population when the disease ends.

Table 4. SIR epidemic model with $p(t)$ weekly-periodic and three different values of N . Mean (standard deviation) of three variables: duration of the disease in days, total proportion of infecteds when the disease ends and number of individuals infected by the first infective (\tilde{R}_0). R_0^* is computed for all the cases.

	$\lambda_1(t)$	$\lambda_2(t)$	$\lambda_3(t)$
Total time	91.71 (39.92)	68.57 (16.87)	42.45 (11.85)
Population infected (%)	9% (7%)	59% (14%)	86% (28%)
\tilde{R}_0	2.83 (1.36)	3.45 (1.42)	3.05 (1.36)
R_0^*	2.83	3.46	3.05

5. Application to real diseases

In this chapter, we investigate how the SIR model we have proposed in the previous sections can approximate the spread of a real disease. For this purpose, we consider two diseases: influenza and meningococcal infection. The real data we study are part of the data frame *InfluMen* contained in the package *surveillance* of R programming, focused on modeling and monitoring epidemic phenomena. The data frame contains the cases observed in Germany between 2001 and 2006. We intentionally analyze only one year for each disease (2001 for influenza and 2006 for meningococcal infection), as the methods we employ can be applied to other periods. In the case of influenza, the development of the disease follows a similar pattern each year. The curve presents the same shape, so an adjusted model can be obtained by slightly varying the parameters. In the case of meningococcal disease, the behavior of the number of infectives is quite rough. The model has to be adjusted more carefully, but

equally good results can be achieved with a suitable choice of the number of daily encounters.

5.1. Influenza

We consider the cases of influenza registered in Germany in 2001. We compare this data with our SIR model. In that year, the German population was 82.35 million, so we fix n accordingly. The duration of the disease is estimated between 3 and 7 days, thus we choose $r = 5$. The probability of transmission is fixed at $p = 0.062$. Since we do not know the level of quarantine, we set $\lambda = 1$. We assume that the number of daily encounters follows a truncated Poisson distribution with a parameter $\lambda(t)$ that is a slight variation of the function $\lambda_2(t)$ used in Section 4.2:

$$\lambda(t) = \frac{20}{1 + \sqrt{t}}.$$

We run our simulation for a year (365 days). The results obtained are contained in Figure 9.

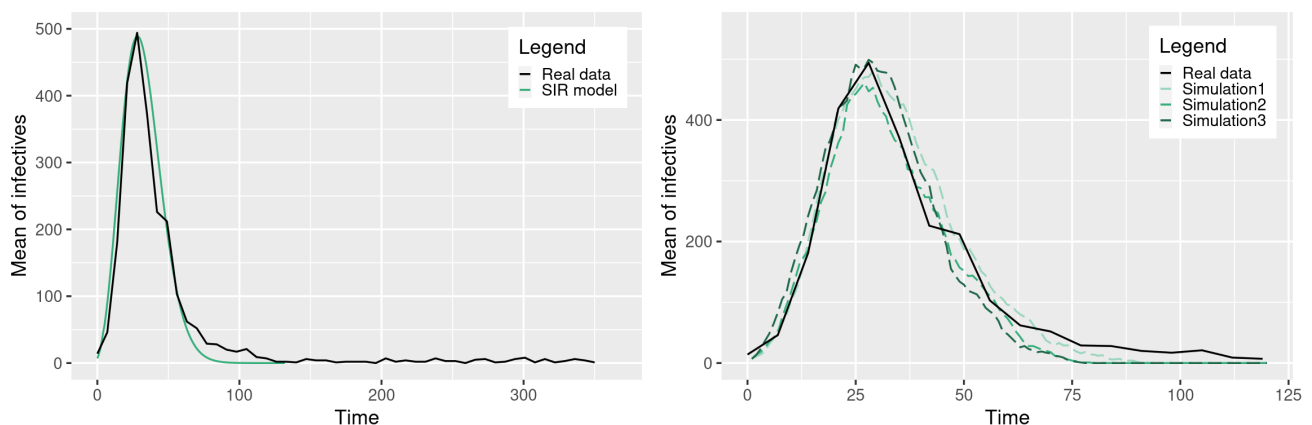


Figure 9. Comparison of real data and SIR model. Real data refers to the cases of influenza registered in Germany in 2001. Left: Evolution of the mean number of infectives of the SIR model. Right: Plot of three single simulations.

In Figure 9 (Left) real data are compared to the mean number of infectives per day. In Figure 9 (Right) three single simulations are plotted. In both cases, our model adjusts remarkably well to real data.

Using the parameters set for the model, we obtain the approximation of the basic reproduction number $R_0^* = \lambda p \sum_{t=0}^{r-1} \mathbb{E}(N(t)) = 2.8$. This value suggests that an outbreak could occur. However, remember that it is computed only considering the number of daily encounters during the first r days which means that the subsequent development of the disease may be different than expected.

5.2. Meningococcal disease

Meningococcal disease refers to any illness caused by bacteria called *Neisseria meningitidis*. About 10% of people have these bacteria without being ill. Sometimes the bacteria invade the body and cause certain illnesses known as meningococcal disease. Meningococcal bacteria are spread by respiratory and throat secretions. Generally, it takes close or prolonged contact to spread them.

Fortunately, they are not as contagious as the germs that cause the common cold or flu. This disease is often severe and can be deadly.

We consider the cases of meningococcal disease registered in Germany in 2006. These data are compared with our SIR model. In 2006, the German population was 82.38 million so n is fixed to this quantity. The duration of meningococcal infection is very variable. Even if it is typically 3 to 4 days, it can last a range of 1 to 10 days. Under these assumptions, we find it opportune to set $r = 5$. As the probability of transmission is quite low, we set $p = 0.02$. Since we do not know the level of quarantine, we set $\lambda = 1$. In this case, we assume that the number of daily encounters follows a binomial distribution with parameters N and $p(t)$. N is fixed to 12 and $p(t)$ is a step function that varies from 0.55 to 1. We run our simulation for a year (365 days). The results obtained are shown in Figure 10.

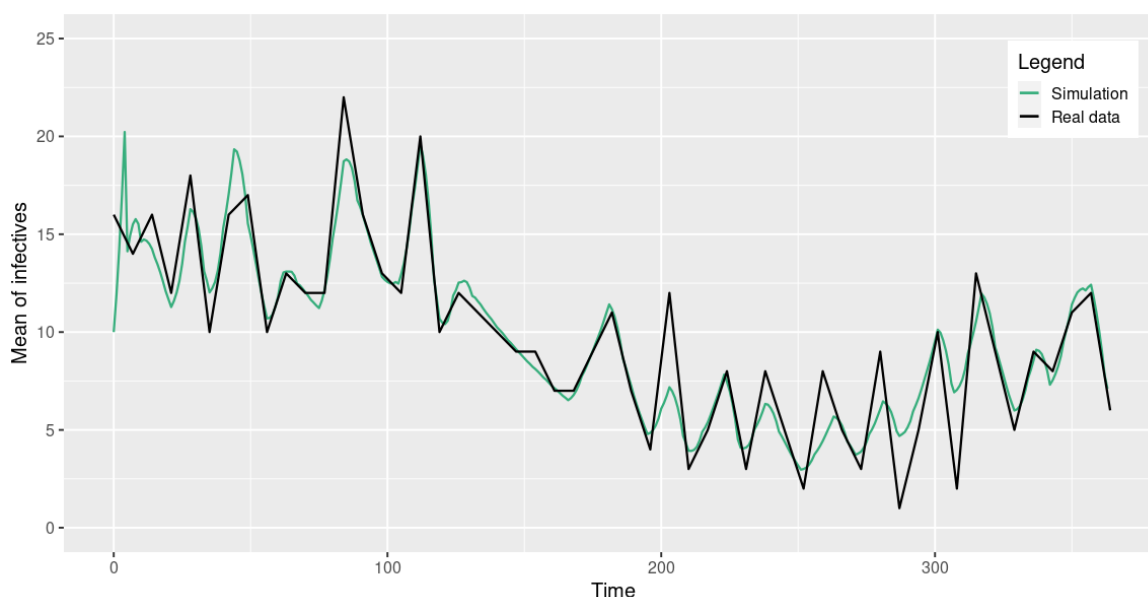


Figure 10. Comparison of real data and SIR model. The real data refers to the cases of meningococcal disease registered in Germany in 2006. As regards the SIR model, the evolution of the mean number of infective is shown.

The figure compares real data with the mean number of infectives per day simulated using our SIR model. In the early days, the simulation does not adjust much to the real data. This can be explained because contrary to our model not all infected at time $t = 0$ belong to the class $Y_0(0)$. Apart from that, our model adjusts rather well to real data. This result could not have been achieved without taking into account the dependence on time of the number of daily encounters.

As for the influenza, we can estimate the basic reproduction number by computing its approximation $R_0^* = \lambda p \sum_{t=0}^{r-1} \mathbb{E}(N(t)) = 0.96$. However, its value is closely related to the number of daily encounters in the first r days which corresponds to the time when our simulation does not fit the real data accurately, as explained above.

6. Discussion and conclusions

In this paper, we study a stochastic SIR-type epidemic model that is an extension of the one proposed by Tuckwell and Williams. We assume that the number of daily encounters of each individual depends on time and we add a parameter to control a possible quarantine of the infectious individuals. Two cases are taken into consideration: when the duration of the disease is constant and when infectious individuals remain infectious throughout their lives. In both situations, we describe analytically the underlying model and its dynamics, deriving a diffusion process and the basic reproduction number.

Several simulations are made to show how differently the disease evolves with respect to the distribution of the number of daily encounters. We consider two laws for this parameter: binomial and truncated Poisson. They will both be used in the application of our model to real diseases. We show how not only the variation of distribution but also the change of their parameters could generate a great variability of results. In particular, the evolution of the mean number of infectives can present different shapes and peaks. Also, the total duration of the disease (intended as the total time before the disease ends) and the total proportion of population infected behave differently. Moreover, we check that in all simulations the mean number of individuals infected by the first infective \widetilde{R}_0 adjusts to R_0^* , which is an approximation of the basic reproduction number.

The importance of time dependence stands out when we apply our model to real diseases. We consider two cases: influenza and meningococcal infection. In both situations, the flexibility given by the time dependence allows our model to adjust exceptionally well to real data.

These results can be the springboard for constructing more complex models to investigate diseases that cannot be adequately described by only the three classes S, I and R. For example, diseases with an initial latency period or the presence of asymptomatic individuals. In these cases, it could be interesting to explore how different distributions and the dependence on time of some parameters such as the number of daily encounters can affect the evolution of the disease.

Another future work could be to consider other parameters as random quantities. In a real-life situation, the duration of the disease r is generally not fixed and can vary from one individual to another. Therefore, an interesting case would be to study a model that deals with this parameter as a random variable and investigate how the disease behaves when its distribution varies.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Mireia Besalú is supported by Departament de Recerca i Universitats de la Generalitat de Catalunya (Spain) with the grant 2021 SGR 01421 (GRBIO).

Giulia Binotto is supported by Ministerio de Ciencia e Innovación (Spain) with the grant PID2021-123733NB-I00.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'Inoculation pour la prévenir, *Hist. Acad. R. Sci. Paris*, 1760, 1–45.
2. N. T. Bailey, *The mathematical theory of infectious diseases and its applications*, London: Charles Griffin & Company Limited, 1975.
3. R. M. Anderson, R. M. May, *Infectious diseases of humans. Dynamics and control*, Oxford: Oxford University Press, 1991. <https://doi.org/10.1002/hep.1840150131>
4. O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical tools for understanding infectious disease dynamics*, Princeton: Princeton University Press, 2013.
5. Y. Cai, Y. Kang, W. Wang, A stochastic SIRS epidemic model with nonlinear incidence rate, *Appl. Math. Comput.*, **305** (2017), 221–240. <https://doi.org/10.1016/j.amc.2017.02.003>
6. X. Bardina, M. Ferrante, C. Rovira, A stochastic epidemic model of COVID-19 disease, *AIMS Math.*, **5** (2020), 7661–7677. <https://doi.org/10.3934/math.2020490>
7. F. Flandoli, E. La Fauci, M. Riva, Individual-based Markov model of virus diffusion: Comparison with COVID-19 incubation period, serial interval and regional time series, *Math. Mod. Meth. Appl. Sci.*, **31** (2021), 907–939. <https://doi.org/10.1142/S0218202521500226>
8. A. McKendrick, Application of mathematics to medical problems, *Proc. Edinburgh Math. Soc.*, **14** (1926), 98–130. <https://doi.org/10.1017/S0013091500034428>
9. W. Kermack, A. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond.*, **115** (1927), 700–721. <https://doi.org/10.1098/rspa.1927.0118>
10. H. Abbey, An examination of the Reed-Frost theory of epidemics, *Hum. Biol.*, **24** (1952), 201–233.
11. L. J. S. Allen, An introduction to stochastic epidemic models, In: *Mathematical Epidemiology*, Berlin, Heidelberg: Springer-Verlag, 2008, 81–130. <https://doi.org/10.1007/978-3-540-78911-6>
12. H. C. Tuckwell, R. J. Williams, Some properties of a simple stochastic epidemic model of SIR type, *Math. Biosci.*, **208** (2007), 76–97. <https://doi.org/10.1016/j.mbs.2006.09.018>
13. M. Ferrante, E. Ferraris, C. Rovira, On a stochastic epidemic SEIHR model and its diffusion approximation, *Test*, **25** (2016), 482–502. <https://doi.org/10.1007/s11749-015-0465-z>
14. A. Gray, D. Greenhalgh, L. Hu, X. Mao, J. Pan, A stochastic differential equation SIS epidemic model, *SIAM J. Appl. Math.*, **71** (2011), 876–902. <https://doi.org/10.1137/10081856X>
15. T. Tsutsui, N. Minamib, M. Koiwai, T. Hamaokaa, I. Yamanea, K. Shimura, A stochastic-modeling evaluation of the foot-and-mouth-disease survey conducted after the outbreak in Miyazaki, Japan in 2000, *Prev. Vet. Med.*, **61** (2003), 45–58. [https://doi.org/10.1016/s0167-5877\(03\)00160-0](https://doi.org/10.1016/s0167-5877(03)00160-0)
16. S. Z. Huang, A new SEIR epidemic model with applications to the theory of eradication and control of diseases, and to the calculation of R_0 , *Math. Biosci.*, **215** (2008), 84–104. <https://doi.org/10.1016/j.mbs.2008.06.005>

17. S. E. A. Mohammed, Stochastic differential systems with memory: Theory, examples and applications, In: *Stochastic Analysis and related topics VI*, Boston: Birkhäuser, 1998. https://doi.org/10.1007/978-1-4612-2022-0_1
18. H. Wahlstrom, L. Englund, T. Carpenter, U. Emanuelson, A. Engvall, I. Vagsholm, A Reed-Frost model of the spread of tuberculosis within seven Swedish extensive farmed fallow deer herds, *Prev. Vet. Med.*, **35** (1998), 181–193. [https://doi.org/10.1016/s0167-5877\(98\)00061-0](https://doi.org/10.1016/s0167-5877(98)00061-0)
19. J. Ng, E. J. Orav, A generalized chain-binomial model with application to HIV infection, *Math. Biosci.*, **101** (1990), 99–119. [https://doi.org/10.1016/0025-5564\(90\)90104-7](https://doi.org/10.1016/0025-5564(90)90104-7)
20. O. C. Ibe, *Markov processes for stochastic modeling*, Elsevier, 2013.



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)