Computation of Expected Epidemic Duration

Alejandro Alarcón González[a,*], Guillermo A. Pérez[a], Shrisha Rao[a]

[a] *University of Antwerp – Flanders Make, Middelheimlaan 1, 2020, Antwerp, Belgium*

Abstract

This paper discusses the mean duration of a closed epidemic modelled by a discrete-time Markov chain. We develop a methodology for the efficient computation of the quantity of interest. The Markov chain model in consideration is bivariate, and is formally handled. We derive explicit terms for the probability to transition from one state to another, and prove that the chain is absorbing. The computation of the mean duration is translated to the computation of the expected hitting times to the set of absorbing states. We use the theory of absorbing Markov chains to derive a matrix formulation that gives way to an efficient algorithm to solve for the expected hitting times. This approach is instantiated in the form of a concrete algorithm, which is further optimized by using dynamic programming. Finally, we have implemented the method and tested it against the use of simulations to estimate mean durations.

Keywords: Compartment model, Stochastic epidemic model, Markov chain, Mean hitting times

1. Introduction

Stochastic epidemic models aim to reflect the nature of the contact and transmission processes involved in the spread of disease in a population. However, their analysis tends to be difficult, which motivates the need for mathematically tractable yet realistic models. This is the context where Markovian epidemic models find their relevance. Markov models are stochastic systems for which the future depends only on the current state of the system. They rely on a matured mathematical theory (see for instance [1] and [2]) and enjoy a good deal of popularity in the biological sciences (see for instance [3] and [4]).

A substantial share of effort in mathematical epidemiology has been devoted to the computation of quantities of interest, with (perhaps) the most famous example being the basic reproduction number R_0 expressed as the leading eigenvalue of the next generation matrix [5]. The quantity we consider in this work is the duration of the stochastic epidemic. That is, the time that elapses till the extinction of the disease. For the class of Markovian epidemic models, Bailey [6] analyzed the Reed-Frost and Greenwood models and derived probability generating functions for the duration of the epidemic in those models. Keeling

*Corresponding author

Email address: `alejandrojavier.alarcongonzalez@uantwerpen.be` (Alejandro Alarcón González)

and Ross [7] in turn addressed the computation of the expected duration in a continuous-time Markovian model.

In this paper, we analyze a discrete-time Markovian epidemic model in a closed population. It is made by discretizing time in a simple deterministic epidemic model and then turning it stochastic with tools from basic probability theory. This derivation follows the work done by Abrams et al. [8], although we take as a basis the much simpler deterministic SIR model [9]. The resulting Markovian epidemic model resembles the chain-binomial constructions of Reed-Frost and Greenwood [6], but with more realistic features: e.g. the time during which an individual remains infectious is not just an instant. However, it is simple enough to allow for a clear exposition of the mathematical and algorithmic treatment herein developed.

Similarly to Black and Ross [10], we analyze the transition matrix of a Markovian epidemic model and then state our research problem in terms of a matrix formulation. Similarly as well, the method we devolop yields the exact value for the quantity of interest. However, the class of models they discuss stem from an event-driven perspective. That is, the only two valid transitions making up such models correspond to an infection and a recovery event. Such event-driven approach is frequently used (see for instance [11, 7, 9]). Instead, the model we analyze stems from a probabilistic interpretation of an Euler-forward discretization of a system of differential equations, which makes it discrete-time and with a constant time step size. Resulting from this construction the number of infectious and recovery events per time step can be very large, as it will be explained in the fourth section.

We address the problem of computing the mean duration of the epidemic in the model we consider. Interestingly, our algorithm gives us these values from all possible starting states. This problem translates to the well known calculation of the mean hitting times in a discrete-time Markov chain [1], which is a system of linear equations. We emphasize that our method can be used to analyze models with more compartments and with contact structure.

Contributions

As a first contribution, we derive an explicit formula for the probability to transition from one state to another in the Markovian model we consider. Then, based on the latter we derive a matrix formulation of the mean hitting times. By analyzing the structure of this matrix we are able to give an expression for the hitting times that already yields an efficient algorithm. Our method is further optimized by using dynamic programming, a mathematical optimization technique that resembles the use of a Pascal triangle to compute binomial coefficients.

Finally, we have implemented the method and tested it against the use of simulations to estimate mean durations.

Based on the method developed and our experiments we observe that: advancing in the Markov chain with respect to the order we defined has very low probability (transition by transition). It follows that empirically estimating the expected termination time of the epidemic using simulations of the chain model can give values that are very different from the exact mean, which our algorithm allows to compute. In practice we observe the difference can be substantial even with large sample numbers. So for important uses, have a good prior or use the exact algorithm.

2. Preliminaries

In this section we give a definition of Markov chains and recall some of their basic properties. We mostly follow the notation from [1] for stochastic processes and Markov chains.

A (finite-state discrete-time) *stochastic process* is a sequence of random variables X_t, for $t \in \mathbb{N}$, all defined on a common probability space and taking values from the same finite set Q of states. We write $(X_t)_{t\geq 0}$ for the sequence of variables. Additionally, we often make implicit use of the probability space $(\Omega, \mathcal{F}, \Pr)$ induced by the stochastic process. In particular Ω, the set of outcomes, consists of all sequences of possible values for the random variables. Then, $\mathcal{F}$ is a σ-algebra over Ω and Pr is a probability measure on $\mathcal{F}$. In this probability space, a *random variable* X is a function $X : \Omega \times \mathbb{N} \to Q$. Usually, we write $\Pr(X_t = i)$ to denote $\Pr(\{\omega \in \Omega : X(\omega, t) = i\})$.

2.1. Markov chains

A (finite-state discrete-time) *Markov chain* is a stochastic process $(X_t)_{t\geq 0}$ such that the sequence starts with an initial distribution over Q, specified by $\boldsymbol{\lambda} = (\lambda_i : i \in Q)$, and satisfies the *Markov property*. That is, for all $t \in \mathbb{N}$ and all $i_0, \ldots, i_{t+1} \in Q$ we have that the probability of X_{t+1} taking the value i_{t+1} depends only on the value of X_t. In symbols:

$$\begin{aligned}
&\Pr(X_0 = i_0) = \lambda_{i_0}, \text{ and} \\
&\Pr(X_{t+1} = i_{t+1} \mid X_t = i_t, \ldots, X_0 = i_0) = \Pr(X_{t+1} = i_{t+1} \mid X_t = i_t).
\end{aligned}$$

Note that this definition guarantees the existence of the so-called *transition matrix* $\boldsymbol{P} = (p_{ij} \mid i, j \in Q)$, where:

$$p_{ij} = \Pr(X_{t+1} = j \mid X_t = i).$$

The matrix $\boldsymbol{P}$ is *stochastic*, i.e., $\sum_{j\in S} p_{ij} = 1$, for all $i \in Q$.

We shall make use of the following characterization of the Markov property based on the transition matrix (see, e.g., [1, Theorem 1.1.1]).

Theorem 1 (Markov property). *Let $(X_t)_{t\geq 0}$ be a stochastic process. It is a Markov chain if and only if for all $n \in \mathbb{N}$ and all $i_0, \ldots, i_n \in Q$ the following holds.*

$$\Pr(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} \prod_{j=1}^{n} p_{i_{j-1} i_j}$$

Later, it will be useful to consider several probability measures depending on fixed initial states for the Markov chain. For this purpose, let $\Pr_i$ denote the probability measure obtained by setting $\boldsymbol{\lambda}$ so that $\lambda_i = 1$ and $\lambda_j = 0$ for $j \neq i$. Similarly, let $\mathbb{E}_i[\cdot]$ for the expectation operator with respect to $\Pr_i$.

2.1.1. Absorbing Markov chains

We say that state j is *reachable* from state i if and only if for some $t \in \mathbb{N}$:

$$\Pr(X_t = j \mid X_0 = i) > 0.$$

A state $i \in Q$ is *absorbing* if and only if $p_{ii} = 1$. If the set of absorbing states A of a Markov chain $(X_t)_{t\geq 0}$ is not empty and A is reachable from all states, we

say that $(X_t)_{t\geq 0}$ is an *absorbing Markov chain*. In an absorbing Markov chain, a state that is not absorbing is *transient*.

The transition matrix of an absorbing Markov chain has special properties. Consider an ordering of Q such that the t transient states are first, followed by the r absorbing states. Now, the transition matrix will have the following *canonical form*.

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{R} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$$

Above, $\boldsymbol{I}$ is a $r \times r$ identity matrix; $\boldsymbol{0}$, a $r \times t$ matrix of zeros; $\boldsymbol{R}$, a nonzero $t \times r$ matrix; and $\boldsymbol{T}$, a $t \times t$ matrix.

Theorem 2. *Consider an absorbing Markov chain $(X_t)_{t\geq 0}$. The matrix $\boldsymbol{I} - \boldsymbol{T}$ has an inverse.*

The result above is well known. It is, for example, stated and proven in [2, Theorem 11.4]. The inverse of $\boldsymbol{I} - \boldsymbol{T}$ is commonly written $\boldsymbol{N}$ and called the *fundamental matrix*.

2.1.2. Expected hitting times

Let $A \subseteq Q$ be a set of *target states*. We write τ_A to denote the first *hitting time* of a state in A. More formally, τ_A is the random variable:

$$\begin{aligned} \tau_A : \Omega &\to \mathbb{N} \cup \{\infty\} \\ \omega &\mapsto \inf\{t \in \mathbb{N} \mid X(\omega, t) \in A\}. \end{aligned}$$

Note that τ_A can take countably many values only and they all are nonnegative. Hence, its expectation, denoted $\mathbb{E}[\tau_A]$, satisfies the following (see also [1, Section 1.3]).

$$\mathbb{E}[\tau_A] = \sum_{t=0}^{\infty} t \Pr(\tau_A = t) + \infty \Pr(\tau_A = \infty) \tag{1}$$

Let $i \in Q$ be a state and write k_i^A for the value $\mathbb{E}_i[\tau_A]$. The following characterization of the expected hitting times will be useful later. The result is well known and can be found, for instance, in [1, Theorem 1.3.2].

Theorem 3. *The vector of expected hitting times $\boldsymbol{k^A} = (k_i^A : i \in Q)$ is the minimal (w.r.t. the product order) nonnegative solution to the following system.*

$$\begin{cases} k_i^A = 0 & \text{if } i \in A \\ k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A & \text{otherwise.} \end{cases} \tag{2}$$

2.2. Expected hitting times in absorbing Markov chains

For absorbing Markov chains, the probability that the process reaches an absorbing state is one. In fact, the probability of not having reached it decreases exponentially (see, e.g., [2, Proof of Theorem 11.3]).

Lemma 1. *Consider an absorbing Markov chain $(X_t)_{t\geq 0}$ with absorbing set of states $A \subseteq Q$. There exist $r \in (0, 1)$ and $m \in \mathbb{N}$ such that for all states $i \in Q$ and all $n \in \mathbb{N}$ the following holds.*

$$\Pr_i(\tau_A \geq mn) \leq r^n$$

Proof. The argument hinges on the facts that the Markov chain has a finite state space and that it is absorbing. Indeed, from each state i there is some absorbing state j that is reachable. Let m_i be the minimum number of steps needed for i to reach j. That is, m_i is the minimal $t \in \mathbb{N}$ such that $\Pr_i(X_t \in A) > 0$. Now, define $m = \min_{i \in Q} m_i$ and $r = 1 - \min_{i \in Q} r_i$ where $r_i = \Pr_i(X_{m_i} \in A)$. Observe that $0 < r < 1$. Now, using the Markov property, we get that for all $i \in Q$ the probability of not reaching an absorbing states after $2m$ steps is at most r^2. In symbols: for all $i \in Q$ we have $\Pr_i(\tau_A \geq 2m) \leq r^2$. The claim then follows by induction. □

Using the lemma above, we can prove that in absorbing Markov chains the expected hitting times are always finite.

Theorem 4. *Consider an absorbing Markov chain $(X_t)_{t \geq 0}$ with absorbing set of states $A \subseteq Q$. For all $i \in Q$ we have that $\Pr_i(\tau_A = \infty) = 0$ and $\mathbb{E}_i[\tau_A] < \infty$.*

Proof. For the first part of the claim, we define, for each $n \in \mathbb{N}$, the event $A_n = \{\omega \in \Omega : \tau_A(\omega) \geq mn\}$, where m is as in Lemma 1. Observe that the events are monotone decreasing, i.e. $A_\ell \supseteq A_n$ for all $\ell < n$, and that $\lim_{n \to \infty} A_n = \{\omega \in \Omega : \tau_A(\omega) = \infty\}$. Using σ-additivity and Lemma 1 we get, for all $i \in Q$, the following.

$$\Pr_i(\tau_A = \infty) = \lim_{n \to \infty} \Pr_i(A_n) \leq \lim_{n \to \infty} r^n$$

Since the last limit converges to 0, we are done.

For the second part of the claim, because of the first part together with Equation (1), it suffices to argue that $\sum_{t=0}^{\infty} t \Pr(\tau_A = t) < \infty$. To make use of Lemma 1, we split the series into a finite sum of subseries, one per residue class modulo m. Formally, we have that the following hold.

$$\begin{aligned}
\sum_{t=0}^{\infty} t \Pr(\tau_A = t) &= \sum_{b=0}^{m-1} \sum_{n=0}^{\infty} (mn + b) \Pr(\tau_A = mn + b) \\
&= \sum_{b=0}^{m-1} \sum_{n=0}^{\infty} (mn + b) \Pr(\tau_A \geq mn) && \text{by (finite) additivity} \\
&\leq \sum_{b=0}^{m-1} \sum_{n=0}^{\infty} (mn + b) r^n && \text{by Lemma 1} \\
&\leq \sum_{b=0}^{m-1} 2m \sum_{n=0}^{\infty} n r^n && \text{since } b < m
\end{aligned}$$

To conclude, we can use d'Alambert's test (a.k.a. the limit test) to confirm that the series $\sum_{n=0}^{\infty} n r^n$ converges. □

As a final result regarding the expected hitting time of absorbing Markov chains we observe that the above theorem can be obtained via linear algebra. In fact, a formula for the vector of expected hitting times exists in terms of the fundamental matrix (see, e.g., [2, Theorem 11.5]) and a vector $\mathbf{1}$ of all ones.

Theorem 5. *Consider an absorbing Markov chain $(X_t)_{t \geq 0}$ with absorbing set of states $A \subseteq Q$ and fundamental matrix $\boldsymbol{N}$ and let $\boldsymbol{k^A} = (k_i^A : i \in Q \setminus A)$ be the vector of expected hitting times. Then, $\boldsymbol{k^A} = \boldsymbol{N}\mathbf{1}$.*

$$\boxed{S} \xrightarrow{\beta} \boxed{I} \xrightarrow{\gamma} \boxed{R}$$

Figure 1: Overview of the flow of individuals in the SIR model: Following a disease infection, susceptible individuals (S) move to an infectious state (I) in which they can infect others. Such infectious individuals will recover over time.

In our developments below, we shall refine the above result for our specific transition and fundamental matrices.

3. SIR Models: From ODEs to Probabilities

Deterministic SIR Models are simple mathematical models of the spreading of infectious diseases. In them, a *population* of size $N \in \mathbb{N}$ is partitioned into *compartments* with labels: S for susceptible, I for infectious, and R for recovered. As illustrated in Figure 1, people may move between compartments following time-dependent dynamics which are usually prescribed by ordinary differential equations (ODEs). Hence, we write $S(t)$, $I(t)$, and $R(t)$ to highlight the fact that these values are function of time $t \geq 0$. Below, we first state the deterministic SIR model for the case of a closed population [9], i.e., births, deaths or infections resulting from contacts with individuals from outside the population are not being considered. (The values β and γ are explained in the sequel.) Then, we derive a stochastic discrete-time version thereof based on Bailey's chain binomial [6].

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\beta I(t)S(t) \\
\frac{dI(t)}{dt} &= \beta I(t)S(t) - \gamma I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t)
\end{aligned}$$

3.1. Towards a stochastic SIR model

When an infectious individual makes contact with a susceptible individual, there is some probability that such contact will lead to disease transmission. This probability times the contact rate is denoted by β^*, and we take it to be irrespective of the specific susceptible-infectious pair. Furthermore, we define the *force of infection*, denoted by $\Lambda(t)$, as the rate for a susceptible individual to become infected at time t. Assuming *homogeneous mixing* within the population yields the relation

$$\Lambda(t) = \frac{I(t)}{N}\beta^*. \tag{3}$$

To simplify notation, we let $\beta = \beta^*/N$. The formulation of the force of infection in (3) is referred to as *mass action transmission* [9]. The exposition above is extended to all susceptible individuals, leading to the continuous-time relation:

$$\frac{dS(t)}{dt} = -\Lambda(t)S(t) = -\beta I(t)S(t) \tag{4}$$

We start the discretization by fixing $h > 0$. By integrating (4) over the time interval $(t, t+h]$, the following recurrence relation is deduced.

$$S(t+h) = e^{-\int_t^{t+h} \Lambda(\tau)d\tau} S(t) \tag{5}$$

Equation (5) gives the expected number of susceptibles at time $t+h$, when we assume that there are $S(t)$ susceptibles at time t. We observe in turn that the first factor on the right hand side of (5) is the probability for a susceptible to escape from infection during the time interval $(t, t+h]$. Therefore, $1 - \exp(-\int_t^{t+h} \Lambda(\tau)d\tau)$ is the probability for a susceptible to become infected during the time interval $(t, t+h]$.

The following integral becomes the *cumulative force of infection* over $(t, t+h]$.

$$\int_t^{t+h} \Lambda(\tau)d\tau \tag{6}$$

By taking the Taylor expansion of (6) around t and considering expressions (3), (5) together, the following holds.

$$\begin{aligned} S(t+h) &= e^{-\int_t^{t+h} \beta I(\tau)d\tau} S(t) \\ &= e^{-\beta(\int_t^t I(\tau)d\tau + hI(t) + \mathcal{O}(h^2))} S(t) \\ &\approx e^{-h\beta I(t)} S(t), \text{ for } h \text{ small} \end{aligned}$$

The probability $p_1(t) = 1 - e^{-h\beta I(t)}$ will now be regarded as the success probability for the Bernoulli trial corresponding to an interaction between an infectious and a susceptible individual, and the interaction occurs within $(t, t+h]$. By recalling the assumption of homogeneous mixing, the previous discussion suggests we define a random variable I_{t+h}^{new} that follows a binomial distribution and which represents the number of newly infected individuals.

$$I_{t+h}^{new} \sim Binomial(S(t), p_1(t) = 1 - \exp(-h\beta I(t))) \tag{7}$$

To finish this modelling part, the infectious period $\Delta(t)$ is assumed to be exponentially distributed with parameter $\gamma \in \mathbb{R}_{\geq 0}$, irrespective of the individual. Accordingly, the cumulative density function $p_2 = 1 - \exp(-h\gamma)$ will represent the probability for an infectious individual to have recovered after time $h > 0$. This suggests that the corresponding number of newly recovered individuals R_{t+h}^{new} can be defined as a random variable with binomial distribution as follows.

$$R_{t+h}^{new} \sim Binomial(I(t), p_2 = 1 - \exp(-h\gamma)) \tag{8}$$

3.2. The SIR process

Let $S(0)$, $I(0)$, and $R(0)$ be fixed constants such that $N = S(0) + I(0) + R(0)$ — these are just the initial conditions for the ODE version of an SIR model — and $h > 0$. For all $t \in \mathbb{N}$, we define discrete random variables S_t, I_t, and R_t, all of which take values from $\mathbb{N}$. In particular, let $S_0 = S(0)$, $I_0 = I(0)$, and $R_0 = R(0)$. For $t \geq 0$ we base the following definition on Equations (7) and (8):

$$S_{t+1} = S_t - Y_t \tag{9}$$

$$I_{t+1} = I_t + Y_t - Z_t \tag{10}$$

$$R_{t+1} = R_t + Z_t \tag{11}$$

with

$$Y_t \sim Binomial(S_t, 1 - \exp(-h\beta I_t)) \text{ and}$$
$$Z_t \sim Binomial(I_t, 1 - \exp(-h\gamma)).$$

By definition, we have the property of *conservation of population.*

Lemma 2. *For all $t \in \mathbb{N}$ we have that $N = S_t + I_t + R_t$.*

Let $t \in \mathbb{N}$ be arbitrary and write $(S_t, I_t, R_t) = (m_1, m_2, m_3)$. We will focus on the probability mass function $p_S^{(t+1)}$ of S_{t+1}. The equations below follow directly from our definition.

$$\begin{aligned} p_S^{(t+1)}(n_1) &= \Pr(S_t = n_1) \\ &= \Pr(S_t - Y_t = n_1) \\ &= \Pr(Y_t = m_1 - n_1) \\ &= \binom{m_1}{m_1 - n_1} (\exp(-h\beta m_2))^{n_1} (1 - \exp(-h\beta m_2)^{m_1 - n_1} \end{aligned} \tag{12}$$

(Here, we adopt the convention that $0^0 = 1$ so that $p_S^{(t+1)}(m_1)$ to be 1 when $m_2 = 0$.) Similarly, for the probability mass function $p_R^{(t+1)}$ of R_{t+1}, we get the following equations.

$$\begin{aligned} p_R^{(t+1)}(n_3) &= \Pr(R_t = n_3) \\ &= \Pr(R_t + Z_t = n_3) \\ &= \Pr(Z_t = n_3 - m_3) \\ &= \binom{m_2}{n_3 - m_3} (1 - \exp(-h\gamma))^{n_3 - m_3} \exp(-h\gamma)^{m_2 - n_3 + m_3} \end{aligned} \tag{13}$$

Observe, from (12) and (13), that S_{t+1} and R_{t+1} are independent. Recall that if X and Y are independent then $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y) = p_X(x)p_Y(y)$, for x, y in the sample space. Thus, the joint probability mass function $p_{S,R}^{(t+1)}$ of (S_{t+1}, R_{t+1}) is given by $p_{S,R}^{(t+1)} = p_S^{(t+1)} p_R^{(t+1)}$. Finally, note that $I_{t+1} = N - S_{t+1} - R_{t+1}$, by Lemma 2. Therefore, the joint probability mass function $p_{S,I,R}^{(t+1)}$ of $(S_{t+1}, I_{t+1}, R_{t+1})$ is also given by $p_{S,I,R}^{(t+1)} = p_S^{(t+1)} p_R^{(t+1)}$.

Henceforth, we write X_t for (S_t, I_t, R_t) and Q for the set of $(n_1, n_2, n_3) \in \mathbb{N}^3$ such that $N = n_1 + n_2 + n_3$. For all $t \in \mathbb{N}$, X_t is a discrete random variable that takes values from Q and its probability mass function (as discussed above), for all $t \geq 1$ is given by:

$$\begin{aligned} p_{S,I,R}^{(t)}(n_1, n_2, n_3) = &\binom{m_1}{m_1 - n_1} (\exp(-h\beta m_2))^{n_1} (1 - \exp(-h\beta m_2)^{m_1 - n_1} \times \\ &\binom{m_2}{n_3 - m_3} (1 - \exp(-h\gamma))^{n_3 - m_3} \exp(-h\gamma)^{m_2 - n_3 + m_3} \end{aligned} \tag{14}$$

where $X_{t-1} = (m_1, m_2, m_3)$.

In the next section, we prove that $(X_t)_{t \geq 0}$ is indeed a stochastic process and even a Markov chain. For now, we can already define what will later become its transition patrix. First, note that the function from Equation (14) that the

dependency on t can be changed to a dependency on $\boldsymbol{m} = (m_1, m_2, m_3)$. Now, for all $\boldsymbol{m}, \boldsymbol{n} = (n_1, n_2, n_3) \in Q$, we define:

$$p_{\boldsymbol{mn}} = \begin{cases} p^{(\boldsymbol{m})}_{S,I,R}(n_1, n_2, n_3) & \text{if } n_1 \leq m_1 \text{ and } m_3 \leq n_3 \leq m_2 + m_3 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Then, the matrix $\boldsymbol{P} = (p_{\boldsymbol{mn}} \mid \boldsymbol{m}, \boldsymbol{n} \in Q)$ is clearly stochastic. For convenience, we sometimes write $\boldsymbol{P}(\boldsymbol{m}, \boldsymbol{n})$ instead of $p_{\boldsymbol{mn}}$.

To conclude this section we state the following property which follows immediately from the definition above.

Lemma 3. *Let $\boldsymbol{m} = (m_1, m_2, m_3), \boldsymbol{n} = (n_1, n_2, n_3) \in Q$ be states. Then, $\boldsymbol{P}(\boldsymbol{m}, \boldsymbol{n}) > 0$ if and only if $n_1 \leq m_1$ and $m_3 \leq n_3 \leq m_2 + m_3$.*

4. Properties of the SIR Stochastic Process

In this section we study the sequence $(X_t)_{t \geq 0}$ of random variables defined in the previous section. In particular, we show that it is indeed a stochastic process. Then, we argue that it is a Markov chain and that $\boldsymbol{P}$ is its transition matrix. To do this, we follow [12, Chapter 10] to explicitly construct the probability measure space it induces.

A *path* is a sequence $X_0 = \boldsymbol{m_0}, X_1 = \boldsymbol{m_1}, \ldots$ such that $\boldsymbol{P}(\boldsymbol{m_j}, \boldsymbol{m_{j+1}}) > 0$ for all $j \geq 0$. The set of all such sequences is referred to as *Paths*. In turn, $Paths_{fin}$ denotes the set of path prefixes $X_0 = \boldsymbol{m_0}, \ldots, X_n = \boldsymbol{m_n}$, where $n \geq 0$ and $\boldsymbol{P}(\boldsymbol{m_j}, \boldsymbol{m_{j+1}}) > 0$ for $0 \leq j < n$. The *length* of $x \in Paths_{fin}(\mathcal{M})$ is the function counting the number of elements in x, and we write $\ell(x)$ for short. Let $\pi \in Paths$, then $\hat{\pi} \in pref(\pi)$ if $\hat{\pi} \in Paths_{fin}$ and $\hat{\pi}_j = \pi_j$ for $0 \leq j < n$, where $n = \ell(\hat{\pi})$. We introduce now the so-called *cylinder set* of $\hat{\pi} \in Paths_{fin}$:

$$Cyl(\hat{\pi}) = \{\pi \in Paths \mid \hat{\pi} \in pref(\pi)\}.$$

Consider a space $(\Omega, \mathcal{F})$ with $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. We say that a family $\{E_j\}_{j \geq 1} \subseteq \mathcal{F}$ *increases* to E and write $E_n \uparrow E$ if and only if:

- $E_j \subseteq E_{j+1}$, and
- $\bigcup_{j \geq 1} E_j = E$.

We enunciate now Carathéodory's extension theorem [13].

Theorem 6. *Let $\mathcal{A}$ and $\sigma(\mathcal{A})$, be an algebra and its generated σ-algebra over a set Ω. If $\nu : \mathcal{A} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is σ-additive, there is a σ-additive function $\pi : \sigma(\mathcal{A}) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ such that $\pi|_{\mathcal{A}} = \nu$. Furthermore, if there exist $\{E_j\}_{j \geq 1} \subseteq \mathcal{A}$ such that $E_j \uparrow \Omega$ and $\nu(E_j) < \infty$ for all $j \geq 1$, then π is unique.*

The condition for ν to be unique corresponds to asking that ν be σ-finite.

We use this last theorem to construct the probability measure space associated with $(X_t)_{t \geq 0}$. Let $n \geq 0$ and let $A_n \subseteq Paths_{fin}$ be as follows.

$$A_n = \{\hat{\pi} \in Paths_{fin} \mid \ell(\hat{\pi}) = n\}$$

Take the family $\{B_n\}_{n \geq 1}$ given by $B_n = \{Cyl(\hat{\pi}) \mid \hat{\pi} \in A_n\}$. Let $\mathcal{A}(\bigcup_{n \geq 1} B_n)$ be the algebra generated by $\bigcup_{n \geq 1} B_n$.

Lemma 4. *Let $\mathcal{F}$ be a semi-algebra over a set Ω and $\mathcal{A}(\mathcal{F})$ be the algebra generated by $\mathcal{F}$. If $A \in \mathcal{A}(\mathcal{F})$, there is a pairwise disjoint family $\{E_j\}_{j=1}^n \subseteq \mathcal{F}$ such that $A = \bigcup_{j=1}^n E_j$.*

Let $E \in \mathcal{A}(\bigcup_{n\geq 1} B_n)$. There exist $\hat{\pi}_1, \ldots, \hat{\pi}_k \in Paths_{fin}$, where the lengths of $\hat{\pi}_j = (\boldsymbol{a_0^j}, \boldsymbol{a_1^j}, \ldots, \boldsymbol{a_{n_j}^j})$ and $\hat{\pi}_i = (\boldsymbol{a_0^i}, \boldsymbol{a_1^i}, \ldots, \boldsymbol{a_{n_i}^i})$ need not be the same for $1 \leq i, j \leq k$, such that $\{Cyl(\hat{\pi}_j)\}_{j=1}^k$ is a pairwise disjoint family and:

$$E = Cyl(\hat{\pi}_1) \cup \cdots \cup Cyl(\hat{\pi}_k).$$

Let us define the function $\mu : \mathcal{A}(\bigcup_{n\geq 1} B_n) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ as follows.

$$\begin{aligned}\mu(E) &= \mu(Cyl(\hat{\pi}_1) \cup \cdots \cup Cyl(\hat{\pi}_k)) \\ &:= \prod_{t=0}^{n_1-1} \boldsymbol{P}(\boldsymbol{a_t^1}, \boldsymbol{a_{t+1}^1}) + \cdots + \prod_{t=0}^{n_k-1} \boldsymbol{P}(\boldsymbol{a_t^k} \boldsymbol{a_{t+1}^k})\end{aligned}$$

Note that this definition does cover all elements of $\mathcal{A}(\bigcup_{n\geq 1} B_n)$. Indeed, consider a family $\{E_j\}_{j\geq 1}$ such that $E_j, \bigcup_{j\geq 1} E_j \in \mathcal{A}(\bigcup_{n\geq 1} B_n)$ for all $j \geq 1$ and $E_i \bigcap E_k = \emptyset$ when $i \neq k$. By Lemma 4, the sets E_j and $\bigcup_{j\geq 1} E_j$ can always be expressed as a finite union of pairwise disjoint cylinder sets. Taking this into account allows us to evaluate μ of $\bigcup_{j\geq 1} E_j$ and E_j as (the limit of) a sum, as defined above.

Let $\sigma(\bigcup_{n\geq 1} B_n)$ denote the σ-algebra generated by $\bigcup_{n\geq 1} B_n$. We use Theorem 6 to show the existence of a σ-additive function $\theta : \sigma(\bigcup_{n\geq 1} B_n) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ such that $\theta|_{\mathcal{A}(\bigcup_{n\geq 1} B_n)} = \mu$. We prove now that θ is a probability function and that μ is σ-finite. Let $\boldsymbol{m} = (S(0), I(0), R(0))$ be the initial conditions for the ODE from the previous section — that is, $X_0 = \boldsymbol{m}$ in the sequence of random variables we have defined. Note that $\Omega = Paths$ follows the next decomposition.

$$\Omega = \bigcup_{\boldsymbol{n}\in Q} Cyl(X_0 = \boldsymbol{m}, X_1 = \boldsymbol{n})$$

Therefore $\theta(\Omega) = \mu(\Omega) = \sum_{\boldsymbol{n}\in Q} \boldsymbol{P}(\boldsymbol{m}, \boldsymbol{n})$. Since $\boldsymbol{P}$ is a stochastic matrix, it follows that $\theta(\Omega) = 1$. Now, define $E_j \in \mathcal{A}(\bigcup_{n\geq 1} B_n)$ and $Q_j \subseteq Q$ for $1 \leq j \leq |Q|$ inductively as follows. We set $Q_1 = Q \backslash \{\boldsymbol{n}\}$ and $E_1 = Cyl(X_0 = \boldsymbol{m}, X_1 = \boldsymbol{n})$ for an arbitrary $\boldsymbol{n} \in Q$. For $j > 2$, let $\boldsymbol{n} \in Q_{j-1}$ and set $Q_j = Q_{j-1} \setminus \{\boldsymbol{n}\}$, $E_j = E_{j-1} \cup Cyl(X_0 = \boldsymbol{m}, X_1 = \boldsymbol{n})$. Note that $E_j \uparrow \Omega$ and $\mu(E_j) < \infty$, for $j \geq 1$. Thus, Ω is σ-finite and the extension θ of μ is unique.

Theorem 7. *The SIR process $(X_t)_{t\geq 0}$ is a stochastic process. Moreover, it is a Markov chain with state space Q transition matrix $\boldsymbol{P}$.*

Proof. By the above arguments, the SIR process $(X_t)_{t\geq 0}$ is a stochastic process with probability space $(Paths, \sigma(\bigcup_{n\geq 1} B_n), \theta)$. Since μ satisfies the Markov property by definition, θ does so too by extension. It then follows from Theorem 6 that the SIR stochastic process is also a Markov chain. To conclude, by definition of μ, we get that the transition matrix of the Markov chain is $\boldsymbol{P}$. $\square$

To keep the notation simpler and uniform, in the rest of this work we write $(\Omega, \mathcal{F}, \Pr)$ for the probability space $(Paths, \sigma(\bigcup_{n\geq 1} B_n), \theta)$ defined in this section.

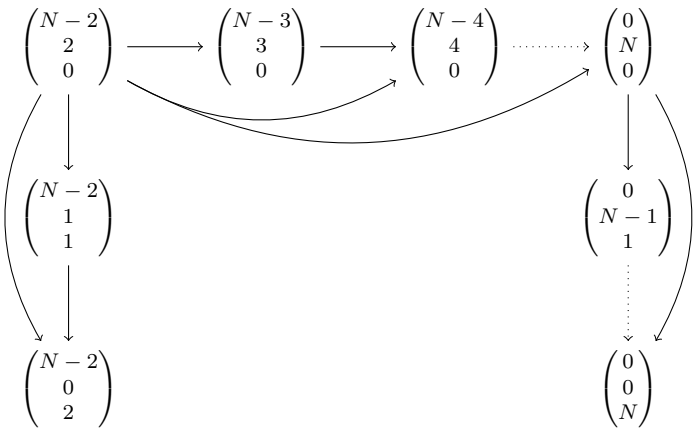


Figure 2: The SIR Markov chain has the interesting property that each state has a very large number of successor states. So much so that from any state one can reach an absorbing state in at most two transitions.

4.1. The Markovian SIR model

In this section we study the SIR Markov chain $(X_t)_{t\geq 0}$ defined in the previous sections. We take the initial distribution $\boldsymbol{\lambda}$ given by $\delta_{\boldsymbol{m}}$, where $\boldsymbol{m} = (N-1, 1, 0)$, i.e., with probability one, $(X_t)_{t\geq 0}$ starts off from the state corresponding to one infectious individual in a population of otherwise susceptible individuals.

4.2. The SIR Markov chain is absorbing

We start by arguing that the SIR Markov chain is absorbing. This will enable the results we have recalled in Section 2 for absorbing Markov chains. First note that by Lemma 8 the set:

$$A = \{(m_1, m_2, m_3) \in Q \mid m_2 = 0\}$$

is nonempty and that it is exactly the set of absorbing states of the SIR Markov chain. It remains for us to prove that A is reachable from all states.

Observe that for all $\boldsymbol{m} = (m_1, m_2, m_3) \in Q$ there exists $\boldsymbol{n} = (n_1, 0, n_3) \in A$ such that $n_1 \leq m_1$ and $m_3 \leq n_3$. Let $\boldsymbol{m'} = (n_1, m_2 + m_1 - n_1, m_3)$ and note that $\boldsymbol{m'} \in Q$. By Lemma 3, we have that $\boldsymbol{P}(\boldsymbol{m}, \boldsymbol{m'}), \boldsymbol{P}(\boldsymbol{m'}, \boldsymbol{n}) > 0$ hence:

$$\Pr(X_t = \boldsymbol{n} \mid X_0 = \boldsymbol{m}) \geq \boldsymbol{P}(\boldsymbol{m}, \boldsymbol{m'})\boldsymbol{P}(\boldsymbol{m'}, \boldsymbol{n}) > 0.$$

Lemma 5. *The SIR Markov chain $(X_t)_{t\geq 0}$ is absorbing.*

4.3. The Expected Epidemic Duration

In this section we address the first hitting time to the set of absorbing states A of the SIR Markov chain. We denote the expected hitting time to this specific set by $\mathbb{E}[\tau_{\text{end}}]$, which in turn can be computed by using the expression in Theorem 3. Instead, we take a first step towards optimizing this computation by considering Lemma 3, which gives a characterization of the nonzero terms in $\boldsymbol{P}$. In this way, for the specific case of the SIR Markov chain, the calculations from Theorem 3 are reduced to what is shown in the next lemma.

Lemma 6. *For all states $\boldsymbol{m} \in Q$, we have that $\mathbb{E}_{\boldsymbol{m}}[\tau_{\text{end}}]$ is:*

$$\begin{cases} 0 & \text{if } m_2 = 0, \\ 1 + \sum_{n_1=0}^{m_1} \sum_{n_3=m_3}^{m_2+m_3} P(\boldsymbol{m}, \boldsymbol{n})\, \mathbb{E}_{\boldsymbol{n}}[\tau_{\text{end}}] & \text{otherwise,} \end{cases}$$

where $\boldsymbol{n} = (n_1, N - n_1 - n_3, n_3)$.

In the following section we turn our attention to concrete algorithms that allow us to compute the nonzero expressions from Lemma 6. We remark that the non-linearity in the original deterministic account of the contact process (see equation (4)) is not to be seen in the formulation of our research problem (given by Lemma 6) . This will simplify the analysis, as it will be seen in the next section.

5. Algorithms for the Expected Epidemic Duration

As a first algorithm, we will specialize Theorem 5 to our SIR Markov chain leveraging Lemma 6. Later, we give pseudo-code for the algorithm described and develop a non-trivial optimization to make it faster (by a linear factor).

5.1. Linear algebra-based approach

We will rewrite the expression corresponding to the non-absorbing states in Lemma 6 using matrix algebra. In doing so, we obtain a mathematical form that already illustrates the guiding principle behind the algorithmic design we use to compute the vector of mean hitting times later. For this purpose, the states of the system are ordered according to the relation:

$$(m_1, m_2, m_3) \prec (n_1, n_2, n_3) \text{ iff } m_3 < n_3 \text{ or } m_3 = n_3 \text{ and } m_1 > n_1.$$

Let the transient states $Q \setminus A$ of the Markov chain be numbered from one to t according to the $\prec$-relation: $(\boldsymbol{m_1}, \boldsymbol{m_2}, \ldots, \boldsymbol{m_t})$. Write $\boldsymbol{k^A}$ for the vector of mean hitting times $(k^A_{\boldsymbol{s_i}} : 1 \leq i \leq t)$ corresponding to those t states in that same order. By Theorem 5, we have $\boldsymbol{k^A} = \boldsymbol{N}\boldsymbol{1}$. By definition of $\boldsymbol{N}$ we thus get the following relation.

$$(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{k^A} = \boldsymbol{1} \tag{16}$$

Now, using the order we imposed on the transient states together with Lemma 3, we can show the following.

Lemma 7. *Consider the SIR Markov chain $(X_t)_{t \geq 0}$. The matrix $\boldsymbol{I} - \boldsymbol{T}$ is upper triangular.*

This enables the computation of $\boldsymbol{k^A}$ by means of *backward substitution.*

In general, for upper triangular systems, i.e., systems of the form $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{U}$ is an invertible upper triangular matrix of size $m \times m$, the back substitution algorithm is given in [14] and it requires m^2 elementary operations: substitution and solving for a variable. Now, the size of the system (16) is $(N+1)(N+2)/2$, and therefore the application of back substitution to our system has a worst-case time complexity of $O(N^4)$, provided $\boldsymbol{P}$ has been pre-computed. In the next section, we instantiate this approach in the form of a concrete algorithm. (We thus also explicitly compute $\boldsymbol{P}$ and analyze the complexity of doing so.) Then, we give a nontrivial optimization based on dynamic programming to compute the transition matrix $\boldsymbol{P}$ efficiently.

Algorithm 1 A simple algorithm to compute $\mathbb{E}^{\boldsymbol{m}}[\tau_{\text{end}}]$ from all $\boldsymbol{m}$

1: **for** $m_1 = 1, \ldots, N$ **do** ▷ Initialize value of target states
2: $\quad$ $\text{Val}(\boldsymbol{m}) \leftarrow 0$, with $\boldsymbol{m} = (m_1, 0, N - m_1)$
3: **end for**
4: **for** $M = 1, \ldots, N$ **do** ▷ Compute value of other states
5: $\quad$ **for** $m_2 = M, \ldots, 1$ **do**
6: $\quad\quad$ $\boldsymbol{m} \leftarrow (M - m_2, m_2, N - M)$
7: $\quad\quad$ $\text{Val}(\boldsymbol{m}) \leftarrow 1$
8: $\quad\quad$ **for** $n_3 = m_3, \ldots, m_2 + m_3$ **do**
9: $\quad\quad\quad$ **for** $n_1 = 0, \ldots, m_1$ **do**
10: $\quad\quad\quad\quad$ $\boldsymbol{n} = (n_1, N - n_1 - n_3, n_3)$
11: $\quad\quad\quad\quad$ **if** $\boldsymbol{m} \neq \boldsymbol{n}$ **then** ▷ Ignore self-loops
12: $\quad\quad\quad\quad\quad$ $\text{Val}(\boldsymbol{m}) \leftarrow \text{Val}(\boldsymbol{m}) + P(\boldsymbol{m}, \boldsymbol{n})\text{Val}(\boldsymbol{n})$
13: $\quad\quad\quad\quad$ **end if**
14: $\quad\quad\quad$ **end for**
15: $\quad\quad$ **end for**
16: $\quad\quad$ $\text{Val}(\boldsymbol{m}) \leftarrow \text{Val}(\boldsymbol{m})/(1 - P(\boldsymbol{m}, \boldsymbol{m}))$ ▷ Add self-loops
17: $\quad$ **end for**
18: **end for**

5.2. Dynamic programming solution

Algorithm 1 exploits equalities from Lemma 6 in order to compute the expected hitting times from all states. Notice that, by conservation of population, two for-loops suffice to go through all source and target states. That is, in line 10 no further for-loop is needed since there is a unique value of n_2 that makes (n_1, n_2, n_3) a valid state. Further note that in line 4 we cycle through all values of $m_1 + m_2$ which is why the following for-loop runs until M. Then, similar to the case of $\boldsymbol{n}$, no further for-loop is needed since there is a unique value of m_3 such that $(M - m_2, m_2, m_3)$ is a valid state.

The next property, which follows directly from Lemma 3, will be useful in our analysis of Algorithm 1.

Lemma 8. *Let $\boldsymbol{m}, \boldsymbol{n}$ be states. If $P(\boldsymbol{m}, \boldsymbol{n}) > 0$ then:*

- *either $\boldsymbol{m} = \boldsymbol{n}$ or*
- *$m_1 \geq n_1$, $m_1 + m_2 \geq n_1 + n_2$, and one of the latter two is strict.*

Note that Algorithm 1 clearly terminates. Further observe that the algorithm has 4 nested for-loops, each ranging over at most N values each. Since $P(\boldsymbol{m}, \boldsymbol{n})$ can be calculated using $O(N)$ elementary operations[1] using Equation(14), we get the following more precise statement.

Theorem 8 (Complexity). *The worst-case time complexity of Algorithm 1 is $O(N^5)$.*

Regarding the values that the algorithm computes, we can prove that they are indeed the required expected hitting times.

[1]We take binary addition and multiplication to be elementary operations. Hence, while exponentiation can be realized using a logarithmic number of such operations via iterated squaring, we are not aware of an algorithm to compute factorials using a sub-linear number of multiplications.

Theorem 9 (Correctness). *Let* $\mathrm{Val}(\cdot)$ *be as computed by Algorithm 1. Then,* $\mathrm{Val}(\boldsymbol{m}) = \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}]$ *for all states* $\boldsymbol{m}$*.*

Proof. First, note that Lemma 8 guarantees $\mathrm{Val}(\boldsymbol{n})$ has already been computed when used in line 12. Second, we can focus on $\mathrm{Val}(\boldsymbol{m})$ such that $m_2 > 0$ since for the remaining $\boldsymbol{m}$, the claim holds already after the initialization ends in line 3.

Now, for an arbitrary $\boldsymbol{m}$ as computed in line 6 let $\mathrm{Val}_1(\boldsymbol{m})$ denote the value of $\mathrm{Val}(\boldsymbol{m})$ after line 7 is executed; $\mathrm{Val}_2(\boldsymbol{m})$ its value after line 15 is executed; and $\mathrm{Val}_3(\boldsymbol{m})$ its value after line 16. We will argue that $\mathrm{Val}_3(\boldsymbol{m}) = \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}]$, which will conclude the argument because $\mathrm{Val}(\boldsymbol{m})$ is not updated afterwards. Note that the following hold.

$$\begin{aligned}
\mathrm{Val}_2(\boldsymbol{m}) &= \mathrm{Val}_1(\boldsymbol{m}) + \sum_{n_1=0}^{m_1} \sum_{n_3=m_3}^{m_2+m_3} P(\boldsymbol{m},\boldsymbol{n})\, \mathbb{E}^{\boldsymbol{n}}[\tau_{\mathrm{end}}] - P(\boldsymbol{m},\boldsymbol{m})\, \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}] \\
&= 1 + \sum_{n_1=0}^{m_1} \sum_{n_3=m_3}^{m_2+m_3} P(\boldsymbol{m},\boldsymbol{n})\, \mathbb{E}^{\boldsymbol{n}}[\tau_{\mathrm{end}}] - P(\boldsymbol{m},\boldsymbol{m})\, \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}] \\
&= \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}] - P(\boldsymbol{m},\boldsymbol{m})\, \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}] \\
&= \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}](1 - P(\boldsymbol{m},\boldsymbol{m}))
\end{aligned}$$

Finally, since $\mathrm{Val}_3(\boldsymbol{m}) = \mathrm{Val}_2(\boldsymbol{m})/(1 - P(\boldsymbol{m},\boldsymbol{m}))$, we get that $\mathrm{Val}_3(\boldsymbol{m}) = \mathbb{E}^{\boldsymbol{m}}[\tau_{\mathrm{end}}]$. □

The question is now: can one do better? In the next section we give a positive answer to this question by improving on our algorithm using dynamic programming to efficiently pre-compute the values $P(\boldsymbol{m},\boldsymbol{n})$ for all $\boldsymbol{m}$ and $\boldsymbol{n}$.

5.3. A more efficient algorithm

Recall Equation (14) gives us that for all states $\boldsymbol{m} \neq \boldsymbol{n}$ such that $n_1 \leq m_1$ and $m_3 \leq n_3 \leq m_2 + m_3$ the probability $P(\boldsymbol{m},\boldsymbol{n})$ of transitioning from $\boldsymbol{m}$ to $\boldsymbol{n}$ is as follows.

$$\binom{m_1}{m_1 - n_1} \exp(-h\beta m_2)^{n_1} (1 - \exp(-h\beta m_2))^{m_1 - n_1} \tag{17}$$

$$\times \binom{m_2}{n_3 - m_3} (1 - \exp(-h\gamma))^{n_3 - m_3} \exp(-h\gamma)^{m_2 - n_3 + m_3} \tag{18}$$

The crux of our improvement on Algorithm 1 is the following recurrence.

Lemma 9. *Let* $\boldsymbol{m}, \boldsymbol{n}$ *be states such that* $n_1 < m_1$ *and* $m_3 < n_3 \leq m_2 + m_3$*. Then,* $P(\boldsymbol{m},\boldsymbol{n}) = \alpha(\boldsymbol{m},\boldsymbol{n}) P(m_1 - 1, m_2, m_3 + 1, \boldsymbol{n})$*, where:*

$$\alpha(\boldsymbol{m},\boldsymbol{n}) := \frac{m_1(m_2 - n_3 + m_3 + 1)(1 - \exp(-h\gamma))(1 - \exp(-h\beta m_2))}{(n_3 - m_3)(m_1 - n_1)\exp(-h\gamma)}.$$

Proof. Note that $P(\boldsymbol{m},\boldsymbol{n}), P(m_1 - 1, m_2, m_3 + 1, \boldsymbol{n}) > 0$ by Lemma 3 and our assumptions. We first consider the binomial coefficients of (17) and (18).

$$\begin{aligned}
&\binom{m_1}{m_1 - n_1}\binom{m_2}{n_3 - m_3} \\
&= \frac{m_1(m_2 - n_3 + m_3 + 1)}{(n_3 - m_3)(m_1 - n_1)} \binom{m_1 - 1}{m_1 - 1 - n_1}\binom{m_2}{n_3 - (m_3 + 1)}
\end{aligned}$$

Algorithm 2 An efficient algorithm to compute $P(\boldsymbol{m}, \boldsymbol{n})$ for all $\boldsymbol{m}$ and $\boldsymbol{n}$

```
for m_1 = 0, ..., N do
    X(m, m) = 1, with m = (m_1, 0, N − m_1)
end for
for M = 1, ..., N do
    for m_2 = 1, ..., M do
        m ← (M − m_2, m_2, N − M)
        for n_3 = m_3, ..., m_2 + m_3 do                    ▷ Fix n_1 = m_1
            X(m, n) ← P(m, n), with n = (m_1, N − m_1 − n_3, n_3)
        end for
        for n_1 = 0, ..., m_1 do                            ▷ Fix n_3 = m_3
            X(m, n) ← P(m, n), with n = (n_1, N − n_1 − m_3, m_3)
        end for
        for n_3 = m_3 + 1, ..., m_2 + m_3 do
            for n_1 = 0, ..., m_1 − 1 do
                n ← (n_1, N − n_1 − n_3, n_3)
                m′ ← (m_1 − 1, m_2, m_3 + 1)
                X(m, n) ← α(m, n)X(m′, n)
            end for
        end for
    end for
end for
```

Importantly, since we have assumed that $m_3 < n_3$ and $n_1 < m_1$, the denominator of the fraction above is not 0. On the other hand, for the terms involving "probabilities" — that is, exponential terms — we observe that the following is true for (17).

$$\begin{aligned}
&\exp(-h\beta m_2)^{n_1}(1-\exp(-h\beta m_2))^{m_1-n_1} \\
&= (1-\exp(-h\beta m_2))\exp(-h\beta m_2)^{n_1}(1-\exp(-h\beta m_2))^{m_1-1-n_1}
\end{aligned}$$

Similarly, for the probabilities in (18) we note the following.

$$\begin{aligned}
&(1-\exp(-h\gamma))^{n_3-m_3}\exp(-h\gamma)^{m_2-n_3+m_3} \\
&= \frac{1-\exp(-h\gamma)}{\exp(-h\gamma)}(1-\exp(-h\gamma))^{n_3-(m_3+1)}\exp(-h\gamma)^{m_2-n_3+m_3+1}
\end{aligned}$$

Indeed, $\alpha(\boldsymbol{m}, \boldsymbol{n})$ is exactly the product of the coefficients on the right-hand sides of the above equalities. This concludes the proof as the product of the left-hand sides is exactly $P(\boldsymbol{m}, \boldsymbol{n})$ and that of the right-hand sides (after factoring out α) is $P(m_1 - 1, m_2, m_3 + 1, \boldsymbol{n})$. □

Based on Lemma 9, we propose Algorithm 2 to compute $P(\boldsymbol{m}, \boldsymbol{n})$ for all $\boldsymbol{m}$ and $\boldsymbol{n}$. The recurrence relation allows us to compute new transition probabilities based on previously computed ones in a manner reminiscent of how Pascal's rule is used to construct Pascal's triangle.

As with Algorithm 1, we observe that Algorithm 2 clearly terminates because all for-loops are bounded and there are no jump statements in the code. Regarding the time complexity of Algorithm 2, however, we now only ever nest

at most 4 for-loops and $P(\boldsymbol{m}, \boldsymbol{n})$ is never used (i.e. computed) within a for-loop nesting of depth 4.

Theorem 10 (Complexity). *The worst-case time complexity of Algorithm 2 is $O(N^4)$.*

For correctness, we have the following statement.

Theorem 11 (Correctess). *Let $X(\cdot)$ be as computed by Algorithm 2. Then, $X(\boldsymbol{m}, \boldsymbol{n}) = P(\boldsymbol{m}, \boldsymbol{n})$ for all pairs of states $\boldsymbol{m}, \boldsymbol{n}$ such that $n_1 \leq m_1$ and $m_3 \leq n_3 \leq m_2 + m_3$.*

Proof. First, one needs to take care that the previous transition probabilities must be defined before they are being used. Note that this is only relevant in lines 13–19. It thus follows from the order in which the states $\boldsymbol{m}$ are traversed that $X(\boldsymbol{m'}, \boldsymbol{n})$ has already been computed in a previous iteration.

Now, it is easy to see that $X(\boldsymbol{m}, \boldsymbol{n}) = P(\boldsymbol{m}, \boldsymbol{n})$ if either $m_1 = n_1$ or $m_3 = n_3$ since lines 8 and 11 literally assign the right-hand side to the left-hand side of the equality. If neither equality holds, then $X(\boldsymbol{m}, \boldsymbol{n})$ is computed in the loop 13–19 and $X(\boldsymbol{m}, \boldsymbol{n}) = P(\boldsymbol{m}, \boldsymbol{n})$ by Lemma 9. Indeed, the conditions that $n_1 < m_1$ and $m_3 < n_3$ are guaranteed by values over which n_3 and n_1 range in the loops from lines 13 and 14. □

Together with the theorems regarding Algorithm 1, the above imply the following.

Theorem 12. *Algorithm 1, when pre-computing $P(\boldsymbol{m}, \boldsymbol{n})$ using Algorithm 2, computes $\mathbb{E}^{\boldsymbol{m}}[\tau_{\text{end}}]$ for all states $\boldsymbol{m}$ in time $O(N^4)$.*

6. Experiments and Conclusions

In this section we discuss the advantage of using together Algorithm 1 and Algorithm 2 over estimating the mean duration with simulations of the SIR Markov chain. All the code used to generate the experiments we present here can be fetched from `https://github.com/jalarcon-ale/MarkovChain`.

Recalling from (7) the success probabilities $p_1(t) = 1 - \exp(-h\beta I_t)$ and $p_2 = 1 - \exp(-h\gamma)$, it can be observed that using small values for h, β and γ yields very high probabilities for the Markov chain to not move forward, especially when the number of infectious individuals I_t is low. This implies that an extremely large amount of simulations is needed in order to effectively estimate the mean time elapsed from the start till the end of the epidemic.

We illustrate the difference between estimated and exact mean duration by means of simulations. In order to show the discrepancy in a relevant setting, we would (ideally) choose parameter values β and γ leading to the onset of a major outbreak. However, we do not know of a characterization of the epidemic threshold for the SIR Markov chain. Therefore, we choose for parameter values leading to a major outbreak in the corresponding deterministic SIR model, since its epidemic threshold condition is known [5] and because the SIR Markov chain stems from it.

In Figure 3 and Figure 4 the duration of 100 thousand simulations of the Markov chain is represented with *boxplots*[2]. Additionally, we include a blue dot that corresponds to the empirical mean. Algorithm 1 and Algorithm 2 were used to compute mean epidemic duration, which is represented by red dots. The population sizes considered range from ten to 100 individuals, and the duration is given in hours. Two step sizes h are used in order to illustrate how the estimation improves for increasing success probabilities $p_1(t)$ and $p_2(t)$.

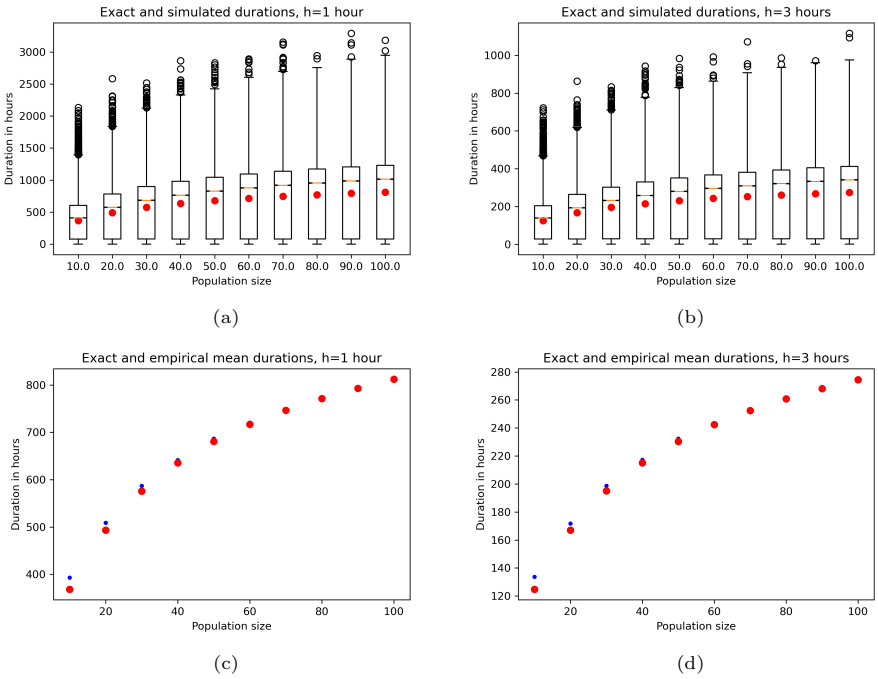
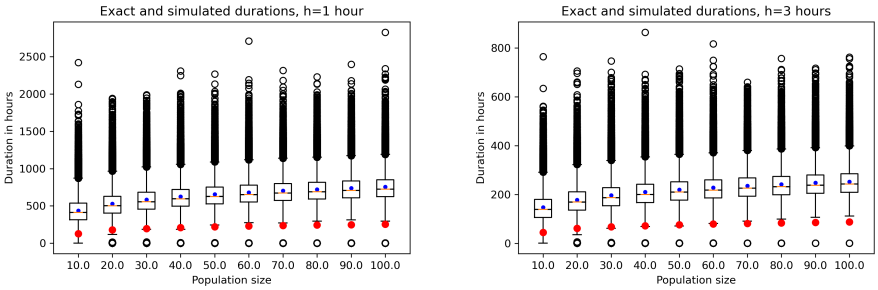


Figure 3: For a range of population sizes (N) varying from 10 to 100 individuals, Figure 3a and Figure 3b show 100 thousand simulated durations of the SIR Markov chain and mean epidemic durations according to Algorithm 1 and Algorithm 2. The parameter values used are $\beta = 1/(2N)$ and $\gamma = 1/6$, while the time step h used is one hour for Figure 3a and Figure 3c, and three hours for Figure 3b and Figure 3d. A decreasing difference between the estimated means (blue dots) and their exact counterparts (red dots) can be observed for increased size of h.

In Figure 4, unrealistically high values of the force of infection $\Lambda(t)$ have been used. This time, the difference between estimated and exact mean durations is very large. The actual value of $\Lambda(t)$ can vary greatly among diseases and contact behaviour. So for important uses, our experiments suggest one must have a good prior or use the exact algorithm.

[2]These are plots that summarize statistical properties of a data set. Concretely, a boxplot creates a box from the first quantile to the third quantile, and a (red) segment of line at the median.



(a) (b)

Figure 4: Durations of the SIR Markov chain, and mean epidemic durations according to Algorithm 1 and Algorithm 2. Again, 100 thousand simulations have been performed for each population size. The parameter values used are $\beta = 1/2$ and $\gamma = 1/6$, while the time step h has been fixed to one hour in Figure 4a and three hours in Figure 4b. Again, a decreasing difference between the estimated means (blue dots) and their exact counterparts (red dots) can be observed for increased size of h.

6.1. Future work

The methodology from this work can be extended to models with more compartments and with contact structure. However, for models with branching, i.e., models where an individual has more than one possibility to change compartments, we remark that the derivation of the transition matrix poses its own challenge. Namely, branching yields case definition and every sub-branch will add a corresponding sub-case. The chain of conditionals involved in this process carries a rather cumbersome bookkeeping effort with itself. For this kind of models, (algorithmic) automation of the transition-matrix derivation should be addressed as a first step towards the extension of our method.

Acknowledgments

We would like to thank Niel Hens for very useful feedback on early version of this work and Kasper Engelen for proof-reading the article. Financial support for this work was provided by the Flemish inter-university (iBOF) “DESCARTES” project.

References

[1] J. R. Norris, Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1997. doi:10.1017/CBO9780511810633.

[2] C. M. Grinstead, J. L. Snell, Introduction to probability, American Mathematical Soc., 1997.

[3] A. A. King, Q. Lin, E. L. Ionides, Markov genealogy processes, Theoretical Population Biology 143 (2022) 77–91. doi:https://doi.org/10.1016/j.tpb.2021.11.003.
URL `https://www.sciencedirect.com/science/article/pii/S0040580921000836`

[4] L. J. Allen, An introduction to stochastic processes with applications to biology, CRC press, 2010.

[5] O. Diekmann, J. Heesterbeek, J. Metz, On the definition and the computation of the basic reproduction ratio r0 in models for infectious-diseases in heterogeneous populations, Journal of mathematical biology 28 (1990) 365–82. doi:10.1007/BF00178324.

[6] N. T. Bailey, et al., The mathematical theory of infectious diseases and its applications, Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[7] M. Keeling, J. Ross, On methods for studying stochastic disease dynamics, Journal of the Royal Society, Interface / the Royal Society 5 (2008) 171–81. doi:10.1098/rsif.2007.1106.

[8] S. Abrams, J. Wambua, E. Santermans, L. Willem, E. Kuylen, P. Coletti, P. Libin, C. Faes, O. Petrof, S. A. Herzog, P. Beutels, N. Hens, Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories, Epidemics 35 (2021) 100449. doi:https://doi.org/10.1016/j.epidem.2021.100449.

[9] M. J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2008.
URL `http://www.jstor.org/stable/j.ctvcm4gk0`

[10] A. J. Black, J. Ross, Computation of epidemic final size distributions, Journal of Theoretical Biology 367 (2015) 159–165. doi:https://doi.org/10.1016/j.jtbi.2014.11.029.

[11] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, The journal of physical chemistry 81 (25) (1977) 2340–2361.

[12] C. Baier, J. Katoen, Principles of model checking, MIT Press, 2008.

[13] S. J. Taylor, Introduction to Measure and Integration, Cambridge University Press, 1973. doi:10.1017/CBO9780511662478.

[14] G. Golub, C. Van Loan, Matrix Computations, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 2013.
URL `https://books.google.be/books?id=X5YfsuCWpxMC`