Numerical Integration of the Master Equation in Some Models of Stochastic Epidemiology

Garrett Jenkinson, John Goutsias*

Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, Maryland, United States of America

Abstract

The processes by which disease spreads in a population of individuals are inherently stochastic. The master equation has proven to be a useful tool for modeling such processes. Unfortunately, solving the master equation analytically is possible only in limited cases (e.g., when the model is linear), and thus numerical procedures or approximation methods must be employed. Available approximation methods, such as the system size expansion method of van Kampen, may fail to provide reliable solutions, whereas current numerical approaches can induce appreciable computational cost. In this paper, we propose a new numerical technique for solving the master equation. Our method is based on a more informative stochastic process than the population process commonly used in the literature. By exploiting the structure of the master equation governing this process, we develop a novel technique for calculating the *exact* solution of the master equation – up to a desired precision – in certain models of stochastic epidemiology. We demonstrate the potential of our method by solving the master equation associated with the stochastic SIR epidemic model. MATLAB software that implements the methods discussed in this paper is freely available as Supporting Information S1.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: goutsias@jhu.edu

Introduction

Stochasticity can play an important role when studying a disease that spreads through a population of individuals [1–3]. A common approach to modeling this problem is by means of a Markov process, whose probability distribution satisfies a deterministic differential equation known as the master equation. Solving the master equation analytically however is not in general possible and Monte Carlo sampling, based on the Gillespie algorithm [4], is often used to accomplish this goal. Unfortunately, accurate evaluation of the probability distribution of a Markov process requires a prohibitively large number of Monte Carlo samples for most systems of interest. As a consequence, Monte Carlo sampling is mostly used to estimate statistical summaries of the underlying stochastic population dynamics, such as means and variances.

To evaluate the solution of the master equation, a number of approximation techniques have been proposed in the literature, such as the system-size expansion method of van Kampen [1,5,6]. While approximations may work well in certain circumstances, they often fail when the underlying assumptions are not satisfied. The system-size expansion method for example can only produce a normal approximation to the solution of the master equation. Therefore, if the probability distribution of the population process is bimodal, then this method will produce erroneous results.

Some effort has recently shifted away from Monte Carlo sampling and approximation techniques and has focused on exploiting the linear structure of the master equation associated with the population process. This results in a numerical solution to the master equation through matrix exponentiation; e.g., see [2,7–10]. A popular

technique along these lines employs a Krylov subspace approximation (KSA) method [7,8] that dramatically reduces the size of matrix exponentiation and results in an attractive iterative algorithm for solving the master equation. However, the KSA technique is based on several approximations, whose cumulative effect may appreciably affect the method's accuracy, numerical stability, and computational efficiency.

There are two main issues that can affect performance of the KSA method. One is choosing the dimension of the approximating Krylov subspace used. If the dimension is chosen too small, the method may produce an inaccurate solution to the master equation, whereas, a value that is too large can result in an appreciable decrease of computational efficiency. Unfortunately, there is no rigorous way to optimally determine an appropriate value for this parameter, which is chosen manually, even in advanced implementations such as Expokit [7]. Another issue is the fact that, at each step, the KSA method may not necessarily produce a probability vector (i.e., a vector composed of nonnegative elements that sum to one). This problem can be addressed by using a sufficiently small step-size, but this may seriously affect the method's computational efficiency. In practice, the KSA method is equipped with a heuristic step that zeros-out all negative values and re-normalizes the positive values so that they sum to one. This step however introduces its own errors, which may affect the quality of the approximation in a manner that is not easy to predict.

Instead of using the population process, we can describe the stochastic spread of a disease by a more informative stochastic process known as the degree-of-advancement (DA). Exploiting the structure of the master equation governing this process results in a novel

numerical algorithm for calculating the *exact* solution of the master equation, up to a desired precision, which we refer to as the implicit Euler (IE) method. This technique enjoys several advantages over the KSA method: its global error is of first-order with respect to the step-size, it is numerically stable regardless of the step-size used, and always produces a solution whose elements are nonnegative and sum to one. As we will discuss in this paper, the IE method shows great promise for solving certain problems in stochastic epidemiology in which the sample space associated with the DA process is reasonably sized. It is not however meant to replace the KSA method, which is still the best numerical method available for solving the master equation in problems where implementation of the IE method is not computationally attractive or possible. To illustrate the potential of the proposed IE method, we calculate the solution of the master equation associated with the stochastic SIR epidemic model and use this solution to study some important properties of this model.

Methods

Disease dynamics

The classical SIR epidemic model (without births, deaths, or imports of disease) is one of the simplest models in epidemiology [11]. Here, each individual in a population is either susceptible to a disease, infected, or recovered. If we denote by S, I, and R the susceptible, infected and recovered individuals, respectively, and by $S(t)$, $I(t)$ and $R(t)$ their corresponding (and possibly random) population numbers, we can characterize the state of the SIR model at time t by using the 3×1 vector $[S(t)\ I(t)\ R(t)]^T$, where T denotes vector transpose. The state depends on time due to the (possibly random) occurrences of the following two reactions:

$$S + I \rightarrow 2I \quad \text{and} \quad I \rightarrow R, \tag{1}$$
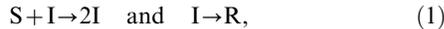
which model infection of a susceptible individual (first reaction) as well as recovery of an infected individual (second reaction).

We can model a complex epidemiological system in more general terms by using the following reactions:

$$\sum_{n \in \mathcal{N}} v_{nm} X_n \rightarrow \sum_{n \in \mathcal{N}} v'_{nm} X_n, \quad m \in \mathcal{M}, \tag{2}$$
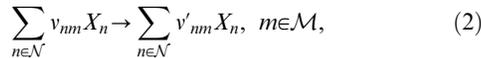
where $\mathcal{N} := \{1, 2, \ldots, N\}$ and $\mathcal{M} := \{1, 2, \ldots, M\}$. This model congregates individuals into N different groups, $X_1, X_2, \ldots, X_N$, which interact through M coupled reactions. Parameters $v_{nm} \geq 0$ and $v'_{nm} \geq 0$ are the *stoichiometry coefficients* of the mth reaction. These parameters tell us how individuals interact with each other as well as their status after occurrence of a particular reaction. For example, in the aforementioned SIR model, we may set $X_1 = S$, $X_2 = I$, $X_3 = R$, resulting in $v_{11} = v_{21} = v_{22} = v'_{32} = 1, v'_{21} = 2$, with the remaining coefficients being zero.

The usual way to characterize an epidemiological system is by means of the $N \times 1$ random vector $\mathbf{X}(t)$ with elements $X_n(t), n \in \mathcal{N}$, where $X_n(t)$ denotes the population of the nth group of individuals present in the system at time $t \geq 0$. By convention, we set $\mathbf{X}(0) = \mathbf{x}(0)$, for some known value $\mathbf{x}(0)$ (i.e., we assume that we know the initial population numbers at time $t = 0$). We refer to the multivariate stochastic process $\{\mathbf{X}(t), t > 0\}$ as the *population process*.

Let $Z_m(t)$ be the (possibly random) number of times that the mth reaction occurs during the time interval $[0, t)$ and $\mathbf{Z}(t)$ be the $M \times 1$ random vector with elements $Z_m(t), m \in \mathcal{M}$. Then, $\{Z_m(t), t > 0\}$ is a counting process, known as the *degree of advancement* (DA) of the mth reaction [5]. We set $Z_m(0) = 0$ and refer to the multivariate stochastic process $\{\mathbf{Z}(t), t \geq 0\}$ as the *DA process*. Note that

$$\mathbf{X}(t) = \mathbf{x}(0) + \mathbb{S}\mathbf{Z}(t), \quad \text{for } t \geq 0, \tag{3}$$

where $\mathbb{S}$ is the $N \times M$ *net* stoichiometry matrix of the system with elements $s_{nm} := v'_{nm} - v_{nm}$. Therefore, and for a given initial population vector $\mathbf{x}(0)$, equation (3) allows us to uniquely determine the population process $\mathbf{X}(t)$ from the DA process $\mathbf{Z}(t)$. However, we cannot in general determine the DA process from the population process. This can only be done when the nullity of $\mathbb{S}$ is zero, in which case $\mathbf{Z}(t) = (\mathbb{S}^T \mathbb{S})^{-1} \mathbb{S}^T [\mathbf{X}(t) - \mathbf{x}(0)]$. As a consequence, the DA process is more informative than the population process.

The master equation

To model an epidemiological system governed by the reactions in (2), we must specify, for each $m \in \mathcal{M}$, the probability that one reaction m will occur within an infinitesimally small time interval $[t, t + dt)$. For most systems of interest, this probability is given by $\pi_m(\mathbf{x})dt + o(dt)$, where $o(dt)$ is a term that goes to zero faster than dt and $\pi_m(\mathbf{x})$ is a (usually nonlinear) function of individual populations at time t, known as the *propensity function* of the mth reaction [12]. It turns out that the DA process $\{Z_m(t), t \geq 0\}$ is a Markovian counting process with intensity $\pi_m(\mathbf{X}(t))$. As a result, it can be shown that the probability mass function $p_{\mathbf{Z}}(\mathbf{z}; t)$ of the DA process satisfies the following master equation [13,14]:

$$\frac{\partial p_z(\mathbf{z}; t)}{\partial t} = \sum_{m \in \mathcal{M}} \{\alpha_m(\mathbf{z} - \mathbf{e}_m) p_{\mathbf{Z}}(\mathbf{z} - \mathbf{e}_m; t) - \alpha_m(\mathbf{z}) p_{\mathbf{Z}}(\mathbf{z}; t)\}, \tag{4}$$

for $t > 0$, initialized by $p_z(\mathbf{z}; 0) = \delta(\mathbf{z})$, with $\delta(\mathbf{z})$ being the Kronecker delta function, where

$$\alpha_m(\mathbf{z}) := \begin{cases} \pi_m(\mathbf{x}(0) + \mathbb{S}\mathbf{z}), & \text{if } \mathbf{z} \geq 0 \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

and $\mathbf{e}_m$ is the mth column of the $M \times M$ identity matrix. In the theory of Markov processes, the master equation is a special case of the more general forward Kolmogorov equation [5].

We can use the solution $p_z(\mathbf{z}; t)$ of the previous master equation to calculate the probability mass function $p_{\mathbf{X}}(\mathbf{x}; t)$ of the population process. Since we are dealing with discrete random variables, this calculation is straightforward:

$$p_{\mathbf{X}}(\mathbf{x}; t) = \sum_{\mathbf{z} \in \mathcal{B}(\mathbf{x})} p_z(\mathbf{z}; t), \tag{6}$$

where $\mathcal{B}(\mathbf{x}) := \{\mathbf{z} | \mathbf{x} = \mathbf{x}(0) + \mathbb{S}\mathbf{z}\}$. Therefore, by solving the master equation (4) [i.e., by calculating the probability mass function $p_z(\mathbf{z}; t)$, for $t > 0$], we can completely specify the dynamic properties of a Markovian model. However, and for most cases of interest, this is a notoriously difficult task, both analytically and computationally. In the following, we propose a promising numerical method to address this problem and illustrate its potential using a simple example based on the stochastic SIR epidemic model.

Exploiting structure

Most available algorithms for solving the master equation focus on the population process instead of the DA process. It turns out that, by using the DA process, we may reap some benefits that can lead to a simple numerical solver for the general master equation (4).

In the following, we assume that statistical analysis of an epidemiological model of interest is limited within a finite time interval $\mathcal{T} := [0, t_{max}]$, where t_{max} is the maximum time for which the DA

process is almost surely contained within an M-dimensional discrete and finite sample space $\mathcal{Z}$; i.e.,

$$\sum_{z\in\mathcal{Z}} p_z(z;t) = 1, \text{ for every } t \in \mathcal{T}.$$

We index the elements in $\mathcal{Z}$ by z_k, $k=1,2,\ldots,K$, where K is the cardinality of $\mathcal{Z}$ (i.e., the total number of elements in $\mathcal{Z}$). We can then define the $K \times 1$ vector $\boldsymbol{\phi}(t)$ with elements $\phi_k(t) = p_z(z_k;t)$, for $k=1,2,\ldots,K$. Clearly, $\boldsymbol{\phi}(t)$ specifies the probability mass function $p_z(z;t)$. It can be seen from (4) that $\boldsymbol{\phi}(t)$ can be calculated by solving the following system of K *linear* ordinary differential equations (ODEs):

$$\frac{d\boldsymbol{\phi}(t)}{dt} = \mathbb{A}\boldsymbol{\phi}(t), \quad t\in\mathcal{T}, \tag{7}$$

where $\mathbb{A}$ is a $K \times K$ matrix that can be directly constructed from the master equation. In the theory of Markov processes, $\mathbb{A}$ is known as the *generator matrix*. Note that the kth column of $\mathbb{A}$ contains zeros in most places except for the kth element that takes value $-\sum_{m\in\mathcal{M}} \alpha_m(z_k)$ ≤ 0 and M off-diagonal elements that take values $\alpha_m(z_k) \geq 0$, $m\in\mathcal{M}$. Therefore, the elements of each column of $\mathbb{A}$ add to zero. Note also that equation (7) is initialized by a vector $\boldsymbol{\phi}(0)$ whose first element equals 1 (assuming that $z_1 = 0$), whereas, the remaining elements are all zero.

The main advantage of using the DA process $\boldsymbol{Z}(t)$ is that, under an appropriate ordering of the elements in $\mathcal{Z}$, the generator matrix $\mathbb{A}$ will be *lower triangular*. We will shortly demonstrate that this can result in substantial simplification of the numerical algorithm used to solve (7).

To obtain a matrix $\mathbb{A}$ that is lower triangular, we must order the points z_k in the sample space $\mathcal{Z}$ *lexicographically*, such that $z_k \prec z_{k+1}$, for $k=1,2,\ldots,K-1$, where $\prec$ denotes that one variable is lexicographically smaller than another [e.g., $(z_1,z_2)\prec(z'_1,z'_2)$ if and only if $z_1 < z'_1$ or $z_1 = z'_1$ and $z_2 < z'_2$]. Because a reaction can only *increase* (by one) the value of a single element of z, it is not possible for probability mass to be transferred from $z_{k'}$ to z_k when $z_k \prec z_{k'}$. Such monotonic transfer of probability does not generally occur when the population process $\mathbf{X}(t)$ is used. Therefore, when the points z_k, $k=1,2,\ldots,K$, in $\mathcal{Z}$ are ordered lexicographically, the (k,k') element of matrix $\mathbb{A}$ will be zero when $k' > k$ and, therefore, $\mathbb{A}$ will be lower triangular. An example is provided in Supporting Information S2.

Numerical solver

We now proceed by exploiting the three key structural characteristics of matrix $\mathbb{A}$: its stability, triangularity, and sparsity. We have noted that the diagonal elements of $\mathbb{A}$ are non-positive. However, since $\mathbb{A}$ is triangular, its diagonal elements will be the eigenvalues of $\mathbb{A}$. Thus, the linear constant coefficient system of ODEs given by (7) is *stable*, ensuring the efficacy of implicit ODE solvers [15]. As a consequence, we can use the implicit Euler method to estimate $\boldsymbol{\phi}(t)$ at discrete time points $t_j := j\tau$, $j=1,2,\ldots$, for a given time step τ. Then, given an estimate $\hat{\boldsymbol{\phi}}(t_{j-1})$ of $\boldsymbol{\phi}(t_{j-1})$, we can obtain an estimate $\hat{\boldsymbol{\phi}}(t_j)$ of $\boldsymbol{\phi}(t_j)$ by solving the following system of linear equations:

$$(\mathbb{I} - \tau\mathbb{A})\hat{\boldsymbol{\phi}}(t_j) = \hat{\boldsymbol{\phi}}(t_{j-1}), \tag{8}$$

where $\mathbb{I}$ is the $K \times K$ identity matrix. By initializing this computation with $\hat{\boldsymbol{\phi}}(0) = \boldsymbol{\phi}(0)$, we can therefore recursively calculate the values of the probability mass function $p_z(z;t)$ of the DA process at the discrete time points t_j, $j=1,2,\ldots$. In Supporting Information

S2, we show that solving the previous system is always possible, for any $\tau > 0$, due to the invertibility of matrix $\mathbb{I} - \tau\mathbb{A}$. We also show that this procedure always returns a probability vector for any step-size $\tau \geq 0$. Moreover, we demonstrate that the resulting method is a first-order solver, since the *global* error $\|\boldsymbol{\phi}(t_j) - \hat{\boldsymbol{\phi}}(t_j)\|_1$ is of $\mathcal{O}(\tau)$ (i.e., the global error is proportional to the step-size τ). Finally, since the implicit Euler step is always stable for *any* choice of τ [15], the errors from previous iterations will not be amplified in later stages, regardless of the step-size used. Therefore, a desired error can be achieved by simply reducing the value of the step-size τ. We refer to the resulting technique for solving the master equation based on (8) as the *implicit Euler* (IE) method.

In general, solving (8) would require $\mathcal{O}(K^3)$ computations, where K is the cardinality of the sample space $\mathcal{Z}$, which will be prohibitive. However, since $\mathbb{A}$ is a *triangular* matrix, we can use forward substitution whose cost is usually of $\mathcal{O}(K^2)$. But since $\mathbb{A}$ is a *sparse* matrix, with each column having only $M+1$ non-zero elements, forward substitution can be done at a cost of $\mathcal{O}(MK)$ [16], where M is the number of reactions. In addition, calculating the probability mass function at time t_j requires storage of $\mathcal{O}(MK)$ nonzero numbers. In particular, we need to store MK nonzero elements of matrix $\mathbb{I} - \tau\mathbb{A}$ as well as $2(K-1)$ elements of vectors $\hat{\boldsymbol{\phi}}(t_j)$ and $\hat{\boldsymbol{\phi}}(t_{j-1})$ [note that the elements of each column of matrix $\mathbb{I} - \tau\mathbb{A}$ and the elements of each of the two vectors $\hat{\boldsymbol{\phi}}(t_j)$ and $\hat{\boldsymbol{\phi}}(t_{j-1})$ sum to one]. Since $K \gg M$, the computational and memory requirements of the IE method will be $\mathcal{O}(K)$, which grow *linearly* in terms of K.

Matrix exponentiation

Instead of the previous approach, we may attempt to solve the master equation governing the population process $\mathbf{X}(t)$ by a matrix exponentiation method [2,9]. Let $\mathcal{X}$ be an N-dimensional discrete and finite sample space such that

$$\sum_{\mathbf{x}\in\mathcal{X}} p_\mathbf{X}(\mathbf{x};t) = 1, \text{ for every } t \in\mathcal{T}.$$

If we index the elements in $\mathcal{X}$ by $\mathbf{x}_l$, $l=1,2,\ldots,L$, where L is the cardinality of $\mathcal{X}$, then we can define the $L \times 1$ vector $\boldsymbol{\theta}(t)$ with elements $\theta_l(t) = p_\mathbf{X}(\mathbf{x}_l;t)$, for $l=1,2,\ldots,L$. In this case, the probability mass function $\boldsymbol{\theta}(t)$ can be calculated by solving the following system of L *linear* ODEs:

$$\frac{d\boldsymbol{\theta}(t)}{dt} = \mathbb{Q}\boldsymbol{\theta}(t), \quad t\in\mathcal{T}, \tag{9}$$

where $\mathbb{Q}$ is a sparse $L \times L$ matrix whose structure can be inferred directly from the master equation [9]. Note that (9) is initialized by a vector $\boldsymbol{\theta}(0)$ whose first element equals 1 [assuming that $\mathbf{x}_1 = \mathbf{x}(0)$], whereas, the remaining elements are all zero.

We can obtain estimates $\hat{\boldsymbol{\theta}}(t_j)$ of $\boldsymbol{\theta}(t_j)$, for $j=1,2,\ldots$, by using the following recursion:

$$\hat{\boldsymbol{\theta}}(t_j) = \exp(\tau\mathbb{Q})\hat{\boldsymbol{\theta}}(t_{j-1}), \tag{10}$$

initialized with $\hat{\boldsymbol{\theta}}(0) = \boldsymbol{\theta}(0)$. The main issue with this equation however is the need to evaluate the matrix exponential $\exp(\tau\mathbb{Q})$. Although many techniques are currently available to do this job, they are not very satisfactory due to issues related to stability, accuracy, and computational efficiency [17].

The best method currently available to compute (10) is based on a *Krylov subspace approximation* (KSA) method [7,8,18]. In its simplest form, the method approximates $\boldsymbol{\theta}(t+\tau) = \exp(\tau\mathbb{Q})\boldsymbol{\theta}(t)$, for a small $\tau > 0$, by $\hat{\boldsymbol{\theta}}(t+\tau) = \|\boldsymbol{\theta}(t)\|_2 \mathbb{V}(t)\exp\{\tau\mathbb{H}(t)\}\mathbf{e}_1$, where $\mathbb{V}(t)$ is an

$L \times L_0$ matrix whose columns form an orthonormal basis for the L_0-dimensional Krylov subspace $\mathcal{K}(t) := \mathrm{span}\{\boldsymbol{\theta}(t), \tau\mathbb{Q}\boldsymbol{\theta}(t), \ldots, (\tau\mathbb{Q})^{L_0-1}\boldsymbol{\theta}(t)\}$, $\mathbb{H}(t)$ is an $L_0 \times L_0$ matrix computed during the well-known Arnoldi procedure used to calculate $\mathbb{V}(t)$, and $\mathbf{e}_1$ is the first column of the $L_0 \times L_0$ identity matrix. Then, the value of $\boldsymbol{\theta}(t_j)$ is recursively approximated by

$$\hat{\boldsymbol{\theta}}(t_j) = \|\hat{\boldsymbol{\theta}}(t_{j-1})\|_2 \, \mathbb{V}(t_{j-1}) \exp\{\tau\mathbb{H}(t_{j-1})\}\mathbf{e}_1,$$

for $j=1,2,\ldots$, with $\hat{\boldsymbol{\theta}}(0) = \boldsymbol{\theta}(0)$. The KSA method reduces the problem of calculating the exponential of the large and sparse $L \times L$ matrix $\mathbb{Q}$ to the problem of calculating the exponential of the much smaller and *dense* $L_0 \times L_0$ matrix $\mathbb{H}$ (note that $L_0 \ll L$, with $L_0 = 30$–50 being sufficient for most applications). Computation of the reduced size problem can be done by standard methods, such as Chebyshev or Padé approximation [7,8,17].

As we mentioned in the Introduction, there are several disadvantages of using the KSA method: possible error accumulation that may lead to instabilities, inability to produce, at each step, a probability vector without heuristically modifying the calculated values, and a need to specify an appropriate dimensionality for the Krylov subspace without appreciably affecting computational efficiency while achieving acceptable numerical accuracy. These issues are nicely circumvented by the IE method, but with a price: the proposed method can be applied to a smaller set of problems than the KSA method.

Practical considerations

In general, the computational and memory requirements of matrix exponentiation grow *quadratically* in terms of the cardinality L of the sample space $\mathcal{X}$, and can quickly become prohibitive for large values of L. The KSA method however can greatly reduce this expense to $\mathcal{O}(L_0(M+L_0)L)$ computations and $\mathcal{O}((M+L_0)L)$ memory locations, where L_0 is the dimension of the approximating Krylov subspace used and M is the number of reactions; see our discussion in Supporting Information S2. Thus, the relative efficiency of the IE method, which requires $\mathcal{O}(MK)$ computation and storage cost, to the KSA approach will depend on the relative values of the cardinalities K and L of the sample spaces $\mathcal{Z}$ and $\mathcal{X}$, respectively.

As we mentioned before, if the nullity of the net stoichiometry matrix $\mathbb{S}$ is zero, then there is a one-to-one correspondence between $\mathbf{x} = \mathbf{x}(0) + \mathbb{S}\mathbf{z}$ and $\mathbf{z}$. As a consequence of (6), the cardinalities of $\mathcal{X}$ and $\mathcal{Z}$ will be the same, in which case $K=L$. Under these circumstances, the IE method will outperform the KSA method. This is a consequence of the fact that $MK = ML < L_0(M+L_0)L$ and $MK = ML < (M+L_0)L$ in this case. We can easily verify that, for the simple SI model ($S+I \rightarrow 2I$), the SIR epidemic model characterized by (1), and the SEIR model ($S+I \rightarrow E+I$, $E \rightarrow I$, $I \rightarrow R$, where E denotes a group of individuals exposed to disease but not yet infectious), the nullity of $\mathbb{S}$ is indeed zero and, therefore, the IE method will be superior to the KSA method.

In general, the IE method will be computationally superior to the KSA method, provided that the cardinality of the sample space $\mathcal{Z}$ is not appreciably larger than $L_0(M+L_0)/M$ times the cardinality of the sample space $\mathcal{X}$ [or not much larger than $(M+L_0)/M$ times the cardinality of the sample space $\mathcal{X}$, if we also consider memory requirements]. Of course, in situations where the nullity of $\mathbb{S}$ is large, the sample space $\mathcal{Z}$ can become appreciably larger than $\mathcal{X}$, in which case the KSA method will be more preferable. Note that there are cases in which $\mathcal{Z}$ and $\mathcal{X}$ can become infinite (e.g., suppose an influx of people at some constant rate $\varnothing \rightarrow X_n$, in which case both sample spaces will be unbounded). In these situations, the use of a finite state projection approach [9] is required to reduce the sample spaces, and

the relative efficiency of the two methods will depend on the sizes of the resulting subspaces.

For a given step-size τ, the IE method described so far generates a sequence of probability vectors $\hat{\boldsymbol{\phi}}(t_j)$, $j=1,2,\ldots$. Assuming that the true solution $\boldsymbol{\phi}(t_{j-1})$ is known at time t_{j-1}, we can show [see equation (S.11) in Supporting Information S2] that the *local* error $\|\boldsymbol{\phi}(t_j) - \hat{\boldsymbol{\phi}}(t_j|t_{j-1})\|_1$ is of $\mathcal{O}(\tau^2)$, where $\hat{\boldsymbol{\phi}}(t_j|t_{j-1})$ is the approximation of $\boldsymbol{\phi}(t_j)$ obtained by the IE method for a given value of $\boldsymbol{\phi}(t_{j-1})$. We can further improve this result by employing a powerful computational tool known as Richardson extrapolation [19].

We have shown in the Supporting Information S2 that, if $\hat{\boldsymbol{\phi}}_\tau(t_j|t_{j-1})$ and $\hat{\boldsymbol{\phi}}_{\tau/2}(t_j|t_{j-1})$ are the approximations of $\boldsymbol{\phi}(t_j)$ obtained from $\boldsymbol{\phi}(t_{j-1})$ by the IE method with step-sizes τ and $\tau/2$, respectively, then $\hat{\boldsymbol{\phi}}_*(t_j|t_{j-1}) := 2\hat{\boldsymbol{\phi}}_{\tau/2}(t_j|t_{j-1}) - \hat{\boldsymbol{\phi}}_\tau(t_j|t_{j-1})$ also approximates $\boldsymbol{\phi}(t_j)$, but with a local error of $\mathcal{O}(\tau^3)$. We therefore expect that $\hat{\boldsymbol{\phi}}_*(t_j|t_{j-1})$ is a better approximation to $\boldsymbol{\phi}(t_j)$ than $\hat{\boldsymbol{\phi}}_\tau(t_j|t_{j-1})$ [or even $\hat{\boldsymbol{\phi}}_{\tau/2}(t_j|t_{j-1})$; see Supporting Information S2] for a sufficiently small step-size τ. This suggests that we can use a valuable modification of the IE method to obtain a better approximation to the solution of the master equation than the original technique. This modification combines two runs of the IE method, with time steps τ and $\tau/2$, and produces a solution $\hat{\boldsymbol{\phi}}_*(t_j)$, given by

$$\hat{\boldsymbol{\phi}}_*(t_j) = \begin{cases} 2\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j), & \text{if } [2\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j)]_{\min} \geq 0 \\ \hat{\boldsymbol{\phi}}_{\tau/2}(t_j), & \text{otherwise,} \end{cases} \quad (11)$$

where $[\mathbf{x}]_{\min}$ denotes the minimum value of the elements of vector $\mathbf{x}$. In this case, $\hat{\boldsymbol{\phi}}_*(t_j)$ is given by the 'improved' vector $2\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j)$ only when all elements of that vector are non-negative. Otherwise, $\hat{\boldsymbol{\phi}}_*(t_j)$ is given by the vector $\hat{\boldsymbol{\phi}}_{\tau/2}(t_j)$ calculated by the IE method with the smaller step-size $\tau/2$. This assures that $\hat{\boldsymbol{\phi}}_*(t_j)$ is always a probability vector. We will be referring to the resulting technique as the Richardson-based implicit Euler (RIE) method. We illustrate one step of this method in Figure 1.

Many ODE solvers, including the KSA method, adjust the step-size at each iteration to assure that the local error **ERR** is less than a pre-specified error tolerance **TOL** while minimizing the computational effort required to accomplish this goal. We can also modify the RIE method to accommodate variable step-sizes. By following our analysis in the Supporting Information S2, we can approximately calculate the local error ERR_j at step j by [see equation (S.16)]

$$\mathrm{ERR}_j = 1.1 \times \|\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j)\|_1,$$

where we use a factor of 1.1 to compensate for the possibility that the true (but unknown) local error is larger (by 10%) than the actual error calculated by $\|\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j)\|_1$. If $\mathrm{ERR}_j < \mathrm{TOL}$, then we consider the step successful and increase the step-size from τ to τ^*, where [see equation (S.17) in Supporting Information S2]

$$\tau^* = \tau\sqrt{\frac{\mathrm{TOL}}{\mathrm{ERR}_j}} = 0.95\,\tau\sqrt{\frac{\mathrm{TOL}}{\|\hat{\boldsymbol{\phi}}_{\tau/2}(t_j) - \hat{\boldsymbol{\phi}}_\tau(t_j)\|_1}}. \quad (12)$$

However, if $\mathrm{ERR}_j > \mathrm{TOL}$, then the step is unsuccessful. In this case, we decrease the step-size from τ to τ^* by using (12) and redo the RIE step.

We finally note that some readers might be concerned with precision loss in the forward substitution step of the IE and RIE methods. To address this issue, we could employ the standard
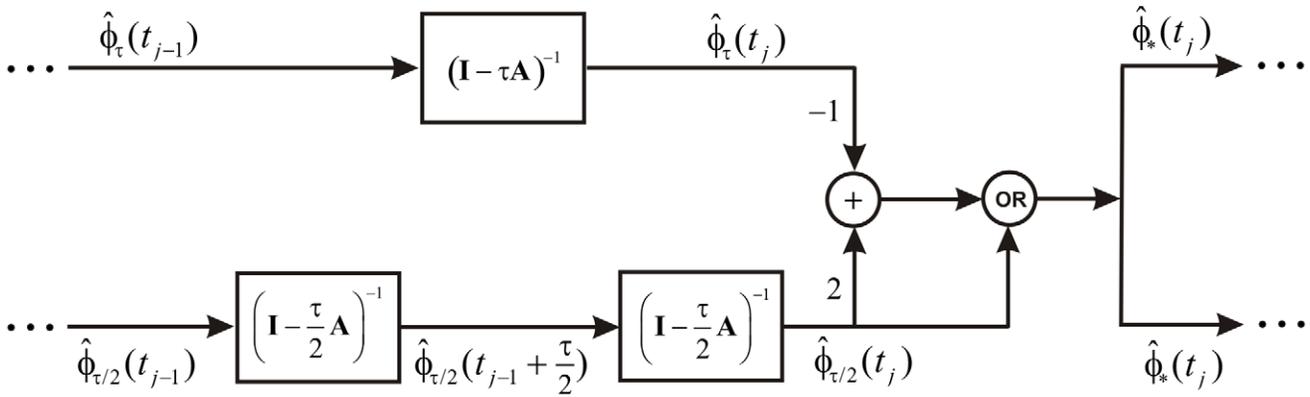
Figure 1. One step of the RIE method. The upper branch implements the standard IE method with step-size τ, whereas, the lower branch implements the IE method with step-size $\tau/2$. "OR" implements equation (11).
doi:10.1371/journal.pone.0036160.g001

numerical technique of *iterative improvement* [15] with moderate additional computational cost. However, we show in the Supporting Information S2 that the matrix $\mathbb{I}-\tau\mathbb{A}$ being inverted is never singular. Moreover, this matrix is far from being singular (i.e., its eigenvalues are not numerically close to zero) as $\tau \rightarrow 0$. We therefore suggest that a preferable method of combating precision loss is to reduce τ, since the step-size tightly regulates the global error as well (see Supporting Information S2). Although in the subsequent example we did not perform iterative improvement, the results indicate that any precision loss is negligible, despite the large dynamic range of probability values involved in the solution.

Results

To demonstrate the efficacy of our method, we tackle the problem of modeling a well-documented 1978 influenza epidemic in an English boarding school [20]. A deterministic SIR model was originally developed to analyze these data [21]. Subsequently, the model was extended to the stochastic case and approximately solved using van Kampen's system-size expansion method [1]. In the following, we use the IE method to compute the *exact* solution of the underlying master equation up to a desired precision.

There are three classes of individuals, S, I and R, representing $Q = 763$ susceptible, infected and recovered pupils. Spreading of the epidemic is governed by the reactions in (1) with propensity functions

$$\pi_1(S(t),I(t),R(t)) = k_1 S(t)I(t) \quad \text{and} \quad \pi_2(S(t),I(t),R(t)) = k_2 I(t),$$

where $k_1 = 0.00218/\text{day}$ and $k_2 = 0.44036/\text{day}$ are the rate constants of infection and recovery, respectively [1]. The initial conditions are given by

$$S(0) = 762, \quad I(0) = 1, \quad R(0) = 0,$$

reflecting the fact that only one pupil is infected at the start of the epidemic. We take the sample space $\mathcal{Z}$ to be the rectangular region in the z plane that begins at $(0,0)$ and extends to include the maximal point $(762,763)$. This is due to the fact that the first reaction can occur at most 762 times, after which all pupils will have been infected, whereas, the second reaction can occur at most 763 times, after which all pupils will have recovered from the infection. As a consequence, the sample space $\mathcal{Z}$ contains $K = 763 \times 764 = 582{,}932$ points.

Numerically solving the master equation over a period of 25 days by means of the KSA method using Expokit [7] took 72 minutes of CPU time on a 2.20 GHz Intel Mobile Core 2 Duo T7500 processor running Matlab 7.7. The resulting solution produces an L2 error $\|\boldsymbol{\theta}(25)-\hat{\boldsymbol{\theta}}(25)\|_2 = 1.48 \times 10^{-3}$, where $\boldsymbol{\theta}$ is a solution of the master equation obtained by a stringent run of Expokit, which we consider to be the 'true' solution. The required TOL value (used to determine a desired error tolerance for the KSA method and for the RIE method with variable step-size) was set to 1×10^{-3}. We obtained the Expokit solution by using a Krylov subspace approximation with dimension $L_0 = 65$. This value was determined by starting with the default value of $L_0 = 30$ and successively increasing it by 5 until the resulting Expokit error estimate was less than $\text{TOL} = 1 \times 10^{-3}$. The reported L2 errors were calculated using a solution obtained by a computationally more expensive Expokit run with $L_0 = 70$ and $\text{TOL} = 1 \times 10^{-4}$, which we consider it to be the 'true' solution. This is based on the premise that Expokit will produce the true solution for sufficiently large L_0 and small TOL.

To be compatible with Expokit, we report here the L2 error. Note however that the error analysis of our method, provided in the Supporting Information S2, is based on the L1 error. On the other hand, using equation (8) with $\tau = 0.01$ days, the IE method took a mere 53 seconds of CPU time, achieving a smaller (by a factor of 2.8) final L2 error of 5.35×10^{-4}. We can achieve a further reduction of the L2 error by using the RIE method with fixed step-size. This is clear from the results summarized in Table 1. This performance can be achieved however at the expense of increasing the CPU time required to calculate the solution. Note that we may be able to decrease the CPU time by using the RIE method with variable step-size (see Table 1). This method however results in a noticeable decrease of accuracy (at least for the example considered here), with an L2 error that is 2.8

Table 1. The L2 error and the CPU time associated with the four numerical methods discussed in this paper.

Numerical Method	L2 Error	CPU Time
KSA	1.48×10^{-3}	4328 seconds
IE	5.35×10^{-4}	52 seconds
RIE (fixed step-size)	1.11×10^{-4}	189 seconds
RIE (variable step-size)	4.06×10^{-3}	124 seconds

doi:10.1371/journal.pone.0036160.t001

times larger than the one obtained with the KSA method.

Since $R(t) = Q - S(t) - I(t)$, it suffices to focus on the joint probability mass function $\Pr[S(t),I(t)]$ of susceptible and infected pupils. It turns out however that the epidemic-free state occurs with high probability $\Pr[S(t),I(t) = 0]$, a situation that visually obscures the values of $\Pr[S(t),I(t)]$. For this reason, instead of $\Pr[S(t),I(t)]$, we depict in Figure 2 a snapshot of the calculated joint conditional probability mass function $\Pr[S(t),I(t)|I(t) > 0]$ of the susceptible and infected pupils at the end of the 6th day, given that at least one pupil is infected. The Supporting Information S3 contains a .gif movie that depicts the dynamic evolutions of $\Pr[S(t),I(t)|I(t) > 0]$ and $\Pr[R(t)|I(t) > 0]$ during the 25 day period. We have obtained these and all subsequent results by exclusively using the basic IE method.

In Figure 3, we depict the dynamic profiles of the mean numbers of susceptible, infected and recovered pupils (solid green lines) as well as the dynamic profiles of the ± 1 standard deviations (dashed red lines), computed directly from the joint probability mass function $\Pr[S(t),I(t),R(t)]$. We also depict the observed data (blue circles) obtained from the literature [20]. These results are identical to the results obtained by Monte Carlo estimation based on 1,000 trajectories sampled from the master equation using the Gillespie algorithm (only data related to the infected pupils are shown), and assures that the IE method produces the correct results. Unfortunately, we cannot employ the Gillespie algorithm to accurately estimate the joint probability mass function $\Pr[S(t),I(t),R(t)]$ in a

reasonable time, due to the prohibitively large number of samples required by this method.

The bimodal nature of the probability mass function depicted in Figure 2 clearly demonstrates that the system-size expansion method used previously in [1] is not appropriate for this model, since the method leads to a unimodal Gaussian approximation. As a matter of fact, the results depicted in Figure 3 are different than the mean and standard deviation profiles depicted in Figures 3–4 of [1]. Because of the Gaussian nature of the system-size expansion method, the results reported in [1] over-estimate the means and under-estimate the standard deviations, since this technique is blind to the bimodal nature of the probability distribution. As a matter of fact, using the means and standard deviations to characterize the stochastic properties of individual classes in the SIR model is not appropriate. This is also evident by the fact that the ± 1 standard deviations can take negative values as well as values greater than 763. In Figure 3, we have truncated these misleading values.

We can use the calculated joint probability mass functions $\Pr[S(t),I(t),R(t)]$ to study a number of dynamic properties of the SIR model in a stochastic setting. In Figure 4(a), for example, we depict the evolution of the expected number of recovered pupils (solid green line), as well as the ± 1 standard deviations (dashed red lines), given that at least one pupil is always infected. During the first few days, few infections occur, and the expected number of recovered pupils will almost be zero. Subsequently, this number increases monotonically to 763, following a near sigmoidal curve.
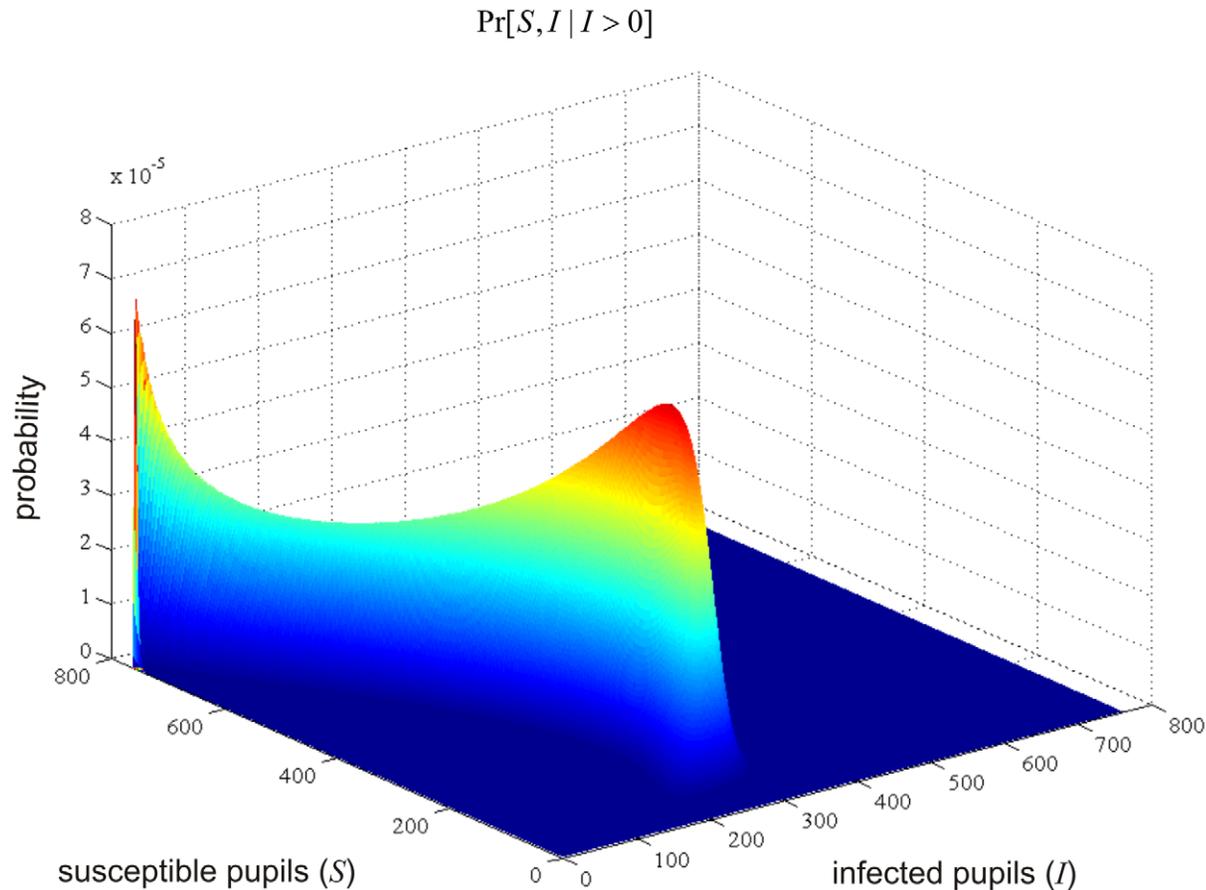
Figure 2. A snapshot of the calculated probability mass function. Joint conditional probability mass function $\Pr[S(t),I(t)|I(t) > 0]$ of susceptible and infected pupils at the end of the 6th day of the influenza epidemic.
doi:10.1371/journal.pone.0036160.g002
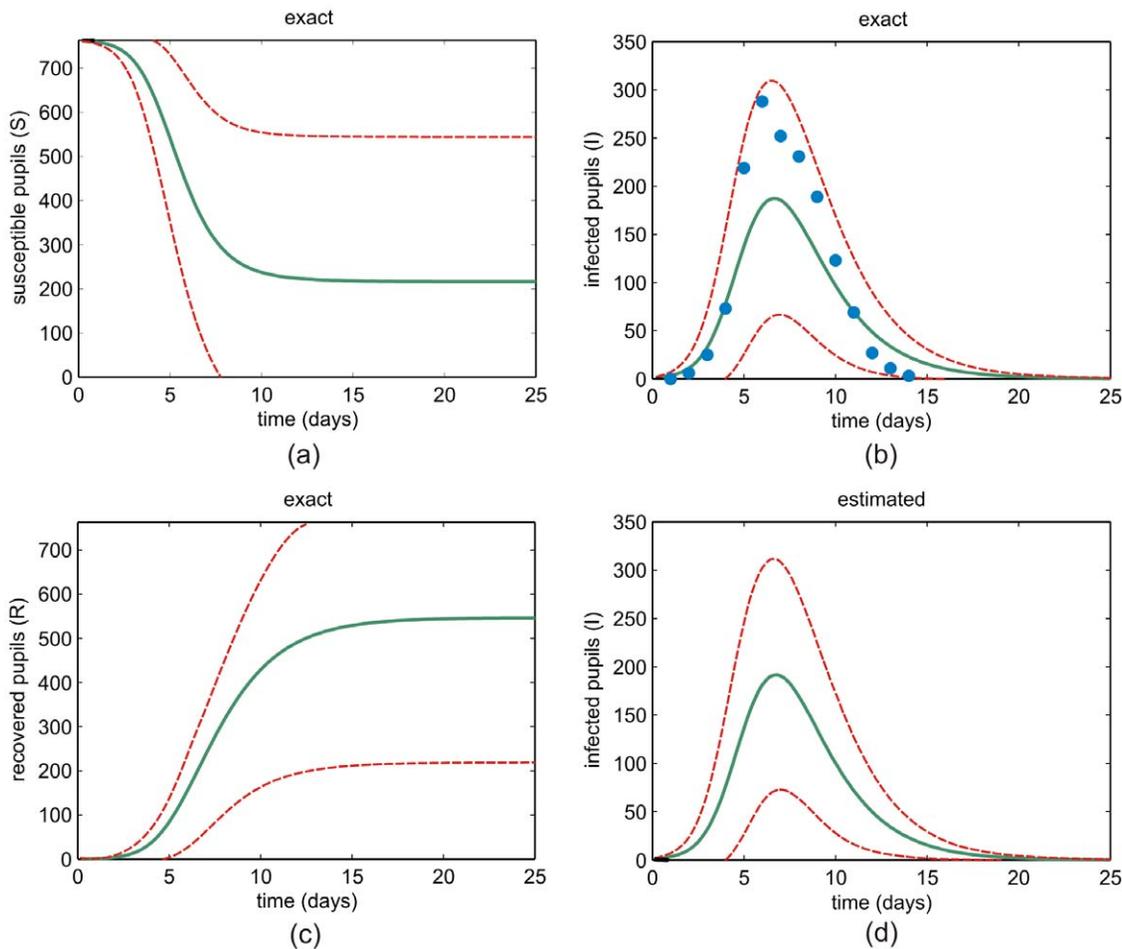
Figure 3. Dynamic mean and standard deviation profiles. The mean profiles (solid green lines) and the ± 1 standard deviation profiles (dashed red lines) of: (a) susceptible, (b) infected, and (c) recovered pupils. Monte Carlo estimates of the mean and standard deviation profiles of the infected pupils are depicted in (d). Blue circles in (b) mark available data [20].
doi:10.1371/journal.pone.0036160.g003

The ± 1 standard deviation curves and the evolution of the Fano factor (variance/mean) depicted in Figure 4(b) indicate that there is appreciable fluctuation in the number of recovered pupils during days 3–10, after which most pupils recover from the infection. According to the results depicted in Figure 4(b), the maximum fluctuation in the number of recovered pupils occurs during the 6th day.

In Figure 4(c), we depict the dynamic evolution of the calculated probability of extinction $\Pr[I(t)=0]$, $t>0$, during a period of 50 days. This evolution is characterized by four phases. During phase I (days 1–4), the probability of extinction increases rapidly from 0% to about 26%, due to the small number of infectious pupils. During phase II (days 5–17), the probability remains relatively constant to about 26%. During this period of time, the epidemic takes its natural course, increasingly infecting susceptible individuals, who eventually recover from the disease. As a consequence, we do not expect the probability of extinction to increase during this phase. On the other hand, during phase III (days 18–40), the number of infected pupils monotonically decreases to zero. It is therefore expected that, during this phase, the probability of extinction will monotonically increase to its maximum value of one. Finally, during phase IV (days 40–50), there is no infectious pupils present. As a result, the influenza virus cannot be

transmitted to the remaining susceptible pupils and the epidemic ceases to exist.

When studying an epidemic model with extinction, a task of practical interest is to calculate the number of individuals that escape infection. This is usually done by evaluating the expected number $\bar{e}$ of individuals that escape infection (or the average number of susceptible individuals that remain after extinction) as the mean value of the stationary probability mass function $\Pr[S(\infty),I(\infty)=0]$ [2]. In Figure 4(d), we depict the joint probability $\Pr[S(50),I(50)=0]$ at time $t=50$ days, which we assume to be a very close approximation to the stationary probability mass function $\Pr[S(\infty),I(\infty)=0]$. By using this probability, we compute $\bar{e} \simeq 547$. Note however that, due to the bimodal nature of $\Pr[S(50),I(50)=0]$, calculating $\bar{e}$ is misleading. On the other hand, by using the result depicted in Figure 4(d), we can confirm that there is a 73.35% chance that 40 pupils or less, and a 26.53% chance that 753 pupils or more, escape infection. Clearly, these 'confidence intervals' provide a more accurate statistical assessment of the number of individuals that escape infection than $\bar{e}$. Interestingly, there is only 0.12% chance that the number of pupils escaping infection is within the range [41,752], which includes the value of $\bar{e}$.
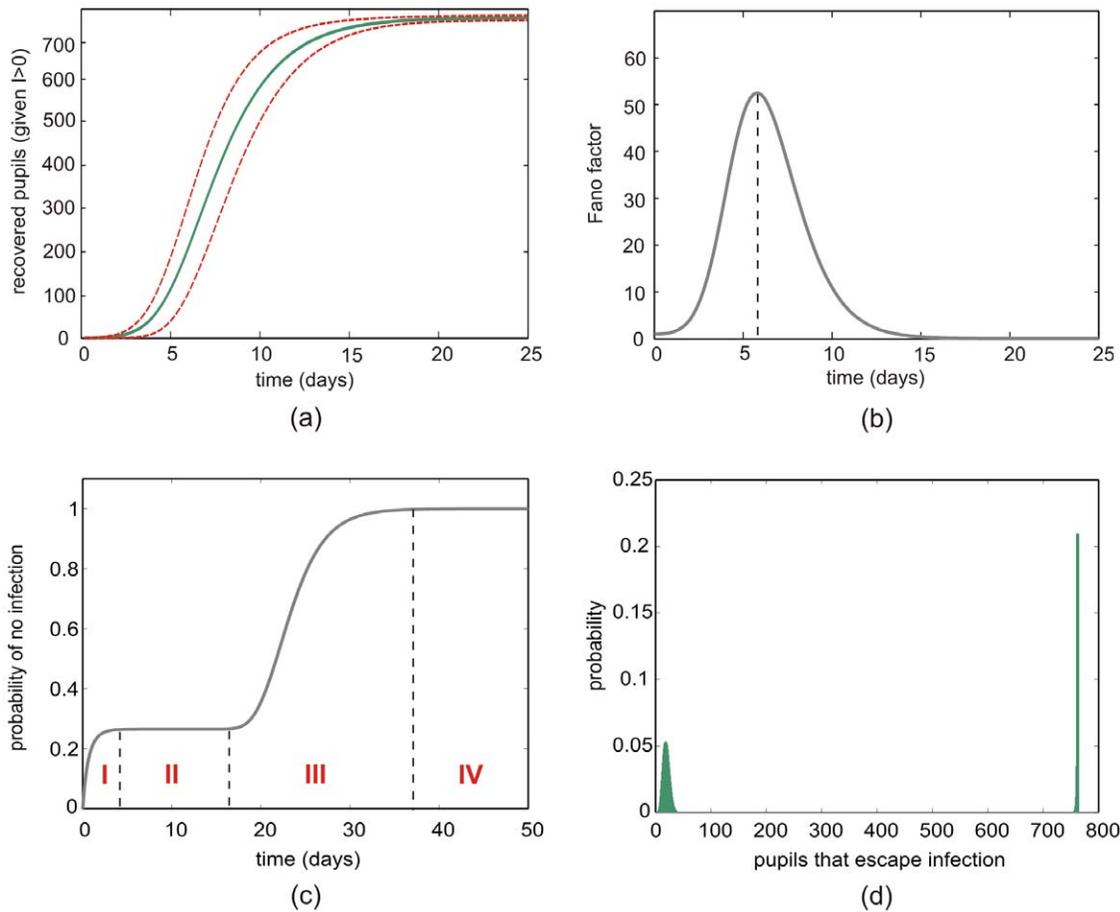
Figure 4. Dynamic properties of the SIR model. (a) Evolution of the expected number of recovered pupils (solid green line) and the ± 1 standard deviations (dashed red lines), given that at least one pupil is always infected. (b) The Fano factor (variance/mean) associated with the results in (a) as a function of time. (c) Dynamic evolution of the probability of extinction $\Pr[I(t)=0]$, $t>0$. (d) The (approximately stationary) probability mass function $\Pr[S(50), I(50)=0]$ at $t=50$ days.
doi:10.1371/journal.pone.0036160.g004

Discussion

Modeling the stochastic dynamics of a disease that spreads through a small and well-mixed population of individuals is an increasingly important subject of modern epidemiology. Unfortunately, even for the simplest model, calculating the underlying probability distribution is a daunting task.

In an effort to address this problem, we have introduced in this paper a new approach to numerically compute the probability mass function of a Markovian population process governed by the master equation. Implementation of this approach is feasible when the number of possible states is not prohibitively large. In this case, the proposed method can lead to *exact* statistical analysis – up to a desired precision – of certain stochastic epidemiological models of interest, such as the SIR epidemic model.

The method introduced in this paper is linear – both in terms of memory and computational requirements – with respect to the cardinality K of the sample space $\mathcal{Z}$ of the degrees of advancement of the underlying reactions. As a consequence, the method is feasible any time $\mathcal{Z}$ is relatively small. In general, however, the cardinality of $\mathcal{Z}$ may grow arbitrarily large, making implementation of the method impossible without an appropriate FSP approximation [9]. Thus, the proposed technique is only applicable to models that constrain the number of reaction events, such as the SIR epidemic model considered in this paper, or models for which the number of reaction

events is sufficiently small during a time period of interest (i.e., models without 'fast' reactions). Moreover, due to the well-known problem of the 'curse of dimensionality,' K grows exponentially with respect to the number of reactions M. Hence, models with many reactions cannot be solved by the proposed method.

An effort is currently underway to reduce the size of the sample space $\mathcal{Z}$, without compromising accuracy. A plausible way to accomplish this goal is to reduce the number of reactions involved by removing 'fast' reactions using a multi-scale approximation technique, such as one of the techniques introduced for biochemical reaction systems [13,14,22], and to adaptively update $\mathcal{Z}$ at each time point t by confining it to the smallest possible subspace $\mathcal{Z}(t)$ of $\mathcal{Z}$. Because of the lower-triangular and sparse nature of matrix $\mathbb{A}$ in (8), it is also plausible that we employ optimized algorithms developed for solving sparse triangular systems of linear equations on parallel and distributed memory computer architectures [23], indicating that future efforts towards solving the master equation could potentially focus on using high-performance computing systems.

Finally, it was brought to our attention by one of the reviewers that, in an earlier work, K. N. Crank proposed a method to map a general Markovian population process on a countable sample space to an augmented Markovian process with triangular generator matrix [24] by appropriately ordering that space. Crank's technique can be easily combined with our implicit Euler

method to construct an alternative algorithm for numerically solving the master equation with focus on the population process instead of the DA process. However, we cannot find any advantage of using Crank's approach over ours. We believe that an approach for numerically solving the master equation based on the DA process is more preferable than an approach based on the population process. The former can provide the probability distributions of both the DA and population processes, whereas, the latter can only produce the probability distribution of the population process. Moreover, the IE method based on the DA process is easier to implement, due primarily to a faster and more natural implementation of the lexicographic ordering used by this approach as opposed to the more complex ordering of the population sample space proposed by Crank. For more details on this issue, see our discussion in Supporting Information S2.

Supporting Information

Supporting Information S1 This file contains the MATLAB code used to generate the results presented in the paper.
(ZIP)

Supporting Information S2 This file contains additional information and proofs that elucidate various mathematical and numerical aspects of the work presented in the paper. It also provides a brief discussion of an alternative method for solving the master equation using the implicit Euler method, based on ordering the population sample space instead of the DA sample space.
(PDF)

Supporting Information S3 This file contains a video of the dynamic evolution of the joint conditional probability mass function of susceptible and infected pupils in an influenza epidemic predicted by the SIR model.
(GIF)

Author Contributions

Conceived and designed the experiments: GJ JG. Performed the experiments: GJ JG. Analyzed the data: GJ JG. Contributed reagents/materials/analysis tools: GJ JG. Wrote the paper: GJ JG.

References

1. Chen WY, Bokka S (2005) Stochastic modeling of nonlinear epidemiology. J Theor Biol 234: 455–470.
2. Keeling MJ, Ross JV (2008) On methods for studying stochastic disease dynamics. J R Soc Interface 5: 171–181.
3. Youssef M, Scoglio C (2011) An individual-based approach to SIR epidemics in contact networks. J Theor Biol 283: 136–144.
4. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22: 403–434.
5. van Kampen NG (2007) Stochastic Processes in Physics and Chemistry. Amsterdam: Elsevier, 3rd edition.
6. Black AJ, McKane AJ (2010) Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough. J R Soc Interface 7: 1219–1227.
7. Sidje RB (1998) Expokit: A software package for computing matrix exponentials. ACM T Math Software 24: 130–156.
8. Sidje RB, Stewart WJ (1999) A numerical study of large sparse matrix exponentials arizing in Markov chains. Comput Stat Data An 29: 345–368.
9. Munsky B, Khammash M (2006) The finite projection algorithm for the solution of the chemical master equation. J Chem Phys 124: 044104.
10. Keeling MJ, Ross JV (2009) Efficient methods for studying stochastic disease and population dynamics. Theor Popul Biol 75: 133–141.
11. Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42: 599–653.
12. Gillespie DT (2000) The chemical Langevin equation. J Chem Phys 113: 297–306.
13. Haseltine EL, Rawlings JB (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J Chem Phys 117: 6959–6969.
14. Goutsias J (2005) Quasiequilibrim approximation of fast reaction kinetics in stochastic biochemical systems. J Chem Phys 122: 184102.
15. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical Recipes: The Art of Scientific Computing. New York, NY: Cambridge University Press, 3rd edition.
16. Fox GC, Williams RD, Messina PC (1994) Parallel computing works!. San Francisco, CA: Morgan Kaufmann Publishers, Inc.
17. Moler C, Van Loan C (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev 45: 3–49.
18. Philippe B, Sidje RB (1995) Transient solutions of Markov processes by Krylov subspaces. In: Stewart W, ed. 2nd International Workshop on the Numerical Solution of Markov Chains. Raleigh, NC: Kluwer Academic Publishers. pp 95–119.
19. Faragó I, Havasi Á, Zlatev Z (2010) Efficient implementation of stable Richardson extrapolation algorithms. Comput Math Appl 60: 2309–2325.
20. CDSC & CDU (1978) Influenza in a boarding school. Brit Med J 1: 587.
21. Murray JD (2001) Mathematical Biology. I. An Introduction. New York, NY: Springer-Verlag, 3rd edition.
22. Haseltine EL, Rawlings JB (2005) On the origins of approximations for stochastic chemical kinetics. J Chem Phys 123: 164115.
23. Mayer J (2009) Parallel algorithms for solving linear systems with sparse triangular matrices. Computing 86: 291–312.
24. Crank KN (1988) A method for approximating the probability functions of a Markov chain. J Appl Prob 25: 808–814.