Adding noise to Markov cohort state-transition models

Rowan Iskandar, PhD, MA
Department of Health Services, Policy, & Practice
Brown University School of Public Health
rowan_iskandar@brown.edu

This study received no financial support from any institutions or grants.

Abstract

Following its introduction over thirty years ago, the Markov cohort model state-transition has been used extensively to model population trajectories over time (Markov trace) in decision modeling and cost-effectiveness studies. We recently showed that a cohort model represents the average of a continuous-time stochastic process on a multidimensional integer lattice governed by a master equation (ME), which represents the time-evolution of the probability function of an integer-valued random vector. Leveraging this theoretical connection, this study introduces an alternative modeling method using a stochastic differential equation (SDE) approach which captures not only the mean behavior but also the variance of the process. The SDE method constitutes the time-evolution of a random vector of population counts. We show the derivation of an SDE model from first principles by postulating the expected change in the distribution of population across states in a small time step. We then describe an algorithm to construct an SDE and solve the SDE via simulation for use in practice. We show the applications of SDE in two case studies. The first example demonstrates that the population trajectories, and their mean and variance, from the SDE and other commonly-used methods match. The second example shows that users can readily apply the SDE method in their existing works without the need for additional inputs beyond those required for constructing a conventional cohort model. In addition, the second example demonstrates that the SDE model is superior to a microsimulation model in terms of computational speed. In summary, an SDE model provides an alternative modeling framework which includes information on variance, can accommodate for time-varying parameters, and is computationally less expensive than microsimulation for a typical modeling problem in cost-effectiveness and decision analyses.

1 Introduction

Decision models have been used in various applications from health technology assessment to clinical guideline development. In decision modeling, a Markov cohort state-transition model is often used to simulate the prognosis of a patient or a group of patients following an intervention. Beck and Pauker [1] introduced the modeling method over 30 years ago with the aim to provide a simple tool for prognostic modeling and for practical use in medical decision making. In principle, a Markov cohort state-transition model [1–3] is a recursive algebraic formula that calculates the *average* number of individuals in each state (*state-configuration*) in the next time step as the sum of current state-configuration and the change in the state-configuration during the time step. This calculation produces a deterministic quantity, i.e. the average of the process, and thus does not completely specify a stochastic representation of the modeled system.

We recently explicated the stochastic process underlying a cohort model, defined as a continuous-time integer-valued stochastic process, in the context where population counts are of interest.[3] The stochastic process is governed by a well-known time-evolution equation of a probability function, known as a master equation. Given this knowledge of the evolution equation associated with a cohort model, we can identify the relationship between a cohort model and other known methodologies in stochastic processes literature.[4] In particular, if we relaxed the whole-number (integer) constraint on the state space to allow for fractional (non-integer) numbers, a cohort model can be immediately extended to represent a change in the *random* state-configuration by adding the variance of the process to the average of the process. This representation of a change in a random vector as the sum of its average and noise is known as the stochastic differential equation (SDE) model.[5] Information on the variance of a process is useful for quantifying the inherent variability of the population trajectories[6] and the extent of accuracy in the mean prediction as well as for deciding whether it is worth collecting more information about the decision problem.[7]

This paper aims to introduce an SDE model as an alternative modeling framework by capitulating the relationship between the stochastic process underlying a cohort model and the SDE. For this paper, we adopt a continuous-time perspective, i.e. for a discrete-time process, a corresponding continuous-time process exists (*embeddability assumption*). In addition, we assume that the process of interest is Markovian, i.e. knowledge of the current state conveys all information that is relevant for forecasting future states. This paper is organized into three main parts. First, we formulate a stochastic cohort model under two representations. We first reintroduce the stochastic process underlying a Markov cohort state-transition model in the form of a master equation. Then, we postulate an SDE representation of the stochastic cohort process which captures the average and the variance of the process. Secondly, we provide a step-by-step guidance for formulating an SDE model given a state-space in the modeling problem, a set of transition rules, and the intensity of each transition. Lastly, we demonstrate two applications of an SDE approach in decision modeling. The first application shows the equivalence between existing methods[3] and an SDE. In the second application, we develop an SDE model for a recently published decision model evaluating the effect of chemotherapy-induced toxicity among breast cancer patients.[8] In this example, we demonstrate the computational efficiency of an SDE model vs. a microsimulation approach in a setting where the transition rates are time-variant. Throughout this exposition, we try to strike a balance between mathematical rigor and accessibility to practitioners.

2 Stochastic cohort model

A stochastic cohort model where each individual follows a Markov chain on a finite set of mutually exclusive and completely exhaustive health states: $\mathbb{S} = \{S_1, S_2, \ldots, S_s\}$, is defined as a stochastic process, $\{\mathbf{N}(t)\}_{t \geq t_0}$ where s is the number of states and t is time ($t \geq t_0$ for an initial time t_0). $\mathbf{N}(t)$ represents a random vector of population counts (*state-configuration*) :

$$\mathbf{N}(t) = [N_1(t)\ N_2(t)\ \ldots\ N_s(t)], \tag{1}$$

in which $\mathbf{N}(t) = \mathbf{n} \in \mathbb{N}^s$ (where $\mathbb{N}$ is the set of non-negative integers or counting numbers) with the probability of observing a particular state-configuration, $\mathbf{N}(t) = \mathbf{n} \in \mathbb{N}^s$ at time t or $P(\mathbf{N}(t) = \mathbf{n}) = P(\mathbf{n}, t)$. In this section, we first present the master equation which governs the stochastic process underlying a cohort model: a stochastic process on $\mathbb{N}^s$. [3] Then, we introduce an SDE formulation of the stochastic cohort model as a stochastic process on $\mathbb{R}^s$ (where $\mathbb{R}$ is the set of real numbers).

2.1 Master equation

An individual transitioning from S_k to S_l is denoted by the transition R_i where $k, l \in \{1 \ldots, s\}$ and $i \in \{1 \ldots, r\}$ (r is the total number of transitions and s^2 is the maximum number of possible transitions or $r \leq s^2$). Since the occurrence of each transition alters the state-configuration, each R_i is associated with a vector representing the change in the current state-configuration, δ_i, and is defined as $\delta_i = e_l - e_k$ where e_j is the j-th column of an identity matrix of size s. We associate each transition with a *propensity function*: $\nu_i(\mathbf{n}, t) = c_i(t) h_i(\mathbf{n})$, where $c_i(t)$ is the transition rate for transition R_i, and $\mathbf{N}(t) = \mathbf{n}$. The function $h_i(\mathbf{n})$ determines the number of ways R_i can occur. In the case where an interaction between individuals in two or more different states has no effect, $h_i(\mathbf{n})$ is equal to the number of at-risk individuals, e.g. for S_k to S_l transition, $h_i(\mathbf{n}) = n_k$. A model for an infectious disease will have more terms in h_i for capturing the rate of contacts between individuals in different disease states. The propensity function denotes the rate of a (any) transition occurring in continuous time as a function of the state-configuration and rate constant. The probability of observing *one particular* transition R_i in a small time interval τ is assumed to be linear in time: $c_i(t)\tau + o(\tau)$, where $o(\tau)$ represents other terms that decay to zero faster than τ. Therefore, the probability of a (any) transition from S_k to S_l in time step τ is given by $n_k c_i(t)\tau + o(\tau)$.[3] A transition R_i occurring with a probability $\nu_i(\mathbf{n}, t)\tau$ in a time step τ alters the current state $\mathbf{N}(t) = \mathbf{n}$ to:

$$\mathbf{N}(t + \tau) = \mathbf{n} + \delta_i \tag{2}$$

The master equation is obtained by enumerating all possible transitions contributing to a change in the probability of observing a particular state-configuration in an infinitesimal time step [3] and is given by:

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{i=1}^{r} \left(\nu_i(\mathbf{n} - s_i, t) P(\mathbf{n} - \delta_i, t) - \nu_i(\mathbf{n}, t) P(\mathbf{n}, t) \right) \tag{3}$$

The terms, $\nu_i(\mathbf{n}, t) P(\mathbf{n}, t)$, represent the probability which is transferred from one to another state-configuration as a result of a transition R_i. In principle, a master equation expresses the change per unit of time of the probability of observing a state-configuration $\mathbf{n}$ as the sum of two opposite effects: the probability flow into and out of $\mathbf{n}$, and exactly represents the dynamics of the stochastic process underlying a cohort model.

2.2 Stochastic differential equation

In this section, we seek an alternative representation of a stochastic cohort model. Our aim is to approximate a Master equation by assuming that the change in the state-configuration is allowed to be fractions or non-integer, i.e. $\mathbf{N}(t)$ has a continuous sample path or $N_i(t+\tau) - N_i(t) \to 0$ as $\tau \to 0$ for $i \in \{1, 2, \ldots, s\}$. In addition, instead working with the evolution equation for the $P(\mathbf{n}, t)$, we consider the evolution of a random vector, $\mathbf{N}(t)$. Our approach of deriving an SDE model is a heuristics one.

Given a state space $\mathbb{S}$ and a set of transitions R_i where $i \in \{1 \ldots, r\}$ (as in Section 2.1), the time evolution of a random vector, $\mathbf{N}(t)$, can be determined by enumerating all possible changes to $\mathbf{N}(t)$ in a small time step, τ, which is governed by the transition type (R_i), its associated change vector (δ_i), and its associated probabilities ($\nu_i(\mathbf{n}, t)\tau$). From Equation 2, we write the change in the state-configuration in τ time step as $\Delta\mathbf{N}(t) = \mathbf{N}(t+\tau) - \mathbf{N}(t)$. The $1 \times s$ vector of the expected change in the state-configurations at time t (*drift* vector) is given by:

$$E[\Delta\mathbf{N}(t)] = \sum_{i=1}^{r} \delta_i \nu_i(\mathbf{n}, t)\tau$$

or in matrix form:

$$E[\Delta\mathbf{N}(t)] = \mathbf{v}(\mathbf{n}, t)\mathbf{d}\tau = \mathbf{A}(\mathbf{n}, t)\tau \tag{4}$$

where $\mathbf{v}(\mathbf{n}, t)$ is a $1 \times r$ vector of propensities where the i-th column corresponds to the propensity function for the transition i ($\nu_i(\mathbf{n}, t)$), $\mathbf{d}$ is an $r \times s$ matrix of δs where the i-th row corresponds to the vector of changes associated with the transition i, and $\mathbf{A}(\mathbf{n}, t)$ is a $1 \times s$ vector in which the k-th element corresponds to the rate of change in the number of individuals in state S_k. The $\mathbf{A}(\mathbf{n}, t)$ is essentially the right-hand side of the ordinary differential equation-equivalence to a cohort model.[3, 9] Equation 4 is the average of the change in current state-configuration across possible transitions weighted by their probabilities of occurring. Similarly the covariance matrix of the change in the state-configurations in τ step (*diffusion* matrix) can be calculated via the definition of the covariance matrix for a random vector:

$$\begin{aligned}
\sum(\Delta\mathbf{N}(t)) &= E[\Delta\mathbf{N}(t)\Delta\mathbf{N}(t)'] - E[\Delta\mathbf{N}(t)]E[\Delta\mathbf{N}(t)]' \\
&\approx \begin{pmatrix}
\sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i1}^2 & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i1}d_{i2} & \ldots & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i1}d_{is} \\
\sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i2}^2 & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i2}^2 & \ldots & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{i2}d_{is} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{is}^2 & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{is}d_{i2} & \ldots & \sum_{i=1}^{r} \nu_i(\mathbf{n}, t)d_{is}d_{is}
\end{pmatrix} \\
&= \mathbf{B}(\mathbf{n}, t)\tau
\end{aligned} \tag{5}$$

if the terms of order higher than τ vanish. Let $\xi(t)$ be a s-dimensional standard Wiener process, i.e. $\xi(t)$ follows a standard normal distribution or $\xi(t) \sim \mathcal{N}(0, \mathbf{I})$, where $\mathbf{I}$ is an $s \times s$ identity matrix. Using a Normal approximation: $\Delta\mathbf{N}(t) = \mathbf{A}(\mathbf{n}, t)\tau + \mathbf{S}(\mathbf{n}, t)\sqrt{\tau}\xi(t) \sim \mathcal{N}(\mathbf{A}(\mathbf{n}, t), \mathbf{B}(\mathbf{n}, t)\tau)$,

the evolution of a stochastic cohort model can be represented by the following updating formula:

$$\begin{aligned}\mathbf{N}(t+\tau) &= \mathbf{N}(t) + \Delta\mathbf{N}(t) \\ &= \mathbf{N}(t) + \mathbf{A}(\mathbf{n},t)\tau + \mathbf{S}(\mathbf{n},t)\sqrt{\tau}\xi(t)\end{aligned} \tag{6}$$

where $\mathbf{S}(\mathbf{n},t) = \sqrt{\mathbf{B}(\mathbf{n},t)}$ (see Appendix A for a more detailed explanation). If τ approaches 0, then we have $\xi_i(t)\sqrt(\tau) \to W_i$ and arrive at the *Itô* SDE:

$$d\mathbf{N}(t) = \mathbf{A}(\mathbf{n},t)dt + \mathbf{B}(\mathbf{n},t)d\mathbf{W}(t) \tag{7}$$

where $\mathbf{W}(t)$ be a s-dimensional Wiener process. A more theoretical discussion on stochastic differential equations can be found in [5] and [10].

3 Developing an SDE

In this section, we describe an algorithm for developing an SDE model for applications in decision modeling. The algorithm involves the following steps:

S1 determine the state space ($\mathbb{S}$), all possible transitions (R_i) among the states in $\mathbb{S}$, and their corresponding propensities, ν_i, for $i \in \{1,\ldots,r\}$;

S2 construct the $1 \times r$ vector of propensities, $\mathbf{v}(\mathbf{n},t)$, and the $r \times s$ matrix of changes, $\mathbf{d}$;

S3 calculate the drift vector (Equation 4) and the diffusion matrix (Equation 5);

S4 choose a numerical method for approximately solving the SDE by generating a sample path (also known as *Markov trace*): $\hat{\mathbf{N}}(t)^{\tau}_{[t_0,T],k}$ where τ, $[t_0,T]$, and k, refer to the numerical method time step, the time interval of analysis, and the index of the samples, respectively;

S5 and estimate a quantity of interest as a function of the sample path(s), i.e., $q = f(\hat{\mathbf{N}}(t)^{\tau}_{[t_0,T],k})$, where: q is the quantity of interest and f is a mapping from the sample paths to the quantity of interest.

The first two steps are identical to specifying a deterministic cohort model. To calculate the drift vector and the diffusion matrix, we use Equations 4 and 5, respectively. Standard methods to diagonalize a positive semi-definite matrix [11] can be utilized to calculate the square root of the diffusion matrix.

For solving an SDE model, a numerical method, such as the Euler-Maruyama method [12], is often used since in most applications, particularly for models with time-varying parameters (e.g. age-specific probability of dying to other causes), an analytical solution to an SDE would not be feasible. A typical numerical method is based on the idea of dividing the time interval of interest, e.g. $[t_0,T]$, to n discrete time points, which may not necessarily be uniformly-spaced: $t_0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_{n-1} < \tau_n = T$, and estimating the values of $\mathbf{N}(t)$ at these time points sequentially. For example, the Euler-Maruyama approximation, $\hat{\mathbf{N}}(t)$, of an SDE model, $d\mathbf{N}(t)$, is given by the following recursive equation:

$$\hat{\mathbf{N}}(\tau_{i+1}) = \hat{\mathbf{N}}(\tau_i) + A\left(\hat{\mathbf{N}}(\tau_i),\tau_i\right)\Delta\tau_i + \mathbf{B}\left(\hat{\mathbf{N}}(\tau_i),\tau_i\right)d\mathbf{W}_{\tau_i} \tag{8}$$

where $\Delta\tau_i = \tau_{i+1} - \tau_i$ and $d\mathbf{W}_{\tau_i} = \mathbf{W}(\tau_{i+1}) - \mathbf{W}(\tau_i)$ denotes the increment of the Wiener process in the time interval $[\tau_i, \tau_{i+1}]$ and is represented by independent Gaussian random variables with zero mean and variance of $\Delta\tau_i = \tau_{i+1} - \tau_i$. In practice, the $d\mathbf{W}_{\tau_i}$ is approximated by $\sqrt{\tau_{i+1} - \tau_i}\mathcal{N}(0, \mathbf{I})$. An application of the numerical method will yield a sample path: $\hat{\mathbf{N}}(t)^{\tau}_{[t_0,T],k}$.

Quantities of interest, such as survival, prevalence or lifetime costs, and their statistics (mean and variance) can be estimated empirically by generating a *sufficiently* large number of sample paths. For example, if q represents the life expectancy of a cohort, then its estimator, $\hat{q}$, can be estimated by simulating L number of sample paths:

$$\hat{q} = \frac{1}{L}\sum_{k=1}^{L} f\left(\hat{\mathbf{N}}(t)^{\tau}_{[t_0,T],k}\right)$$

where the mapping f is defined as the ratio of the total sum of individuals across non-death states in $[t_0, T]$ to the initial population size. In general, the accuracy of the estimator depends on the choice of the Euler-Maruyama's size, τs,and the number of sample paths, L, which are problem-dependent.

4 Applications of SDE

We conduct two case studies to demonstrate: (1) the equivalence of an SDE model to other commonly-used methods in decision modeling, and (2) how to develop a stochastic cohort model using an SDE approach. In the second case study, we also compare the computational time between the SDE and microsimulation approach since both approaches produce estimate of the variance and easily accommodate for time-varying rates. In both case studies, the models are coded in R [13] and Julia [14] as open-source programs have emerged as the lingua franca for practitioners in decision modeling. [15, 16] The comparison of both methods' code execution time is conducted in Julia. The codes for the second example are available under a GNU GPL license and can be found at https://github.com/rowaniskandar/SDE.

4.1 Case study: 4-state Markov model

For the first example, we emulate the example in [3] to show the connection between SDE and other modeling methods: (1) a cohort model with τ = 1-year, (2) a microsimulation model, and (3) an SDE, which is solved with the Euler-Maruyama method and τ = 1-year. We consider a generic four-states stochastic process ($s = 4$), i.e. $\mathbb{S} = \{S_1, S_2, S_3, S_4\}$ (Figure 1), with rate constants of $c_1 = 0.05$, $c_2 = 0.01$, $c_3 = 0.001$, $c_4 = 0.1$, $c_5 = 0.05$, and $c_6 = 2$. We simulate a cohort of 1000 individuals, who start in S_1, i.e. $N_1(0) = 1000$, for a 50-years duration. To calculate the mean and the variances of $\mathbf{N}(t)$ based on the SDE model, we generate 10 and 10000 sample paths. The benchmark for comparing variances is based on 10 and 10000 Monte Carlo samples of a *microsimulation population* with τ = 1-year. Each microsimulation population is conceived as a realization of 1000 Markov chains (or individual trajectories). For the cohort model, the calculation of the 1-year transition probability matrix from the transition rates is based on the formulas from Welton and Ades. [17] The results of the population trajectories in state S_3 are given in Figure (Figure 2). We observe no difference in the trajectories between the empirical estimates from the SDE, cohort model and microsimulation. Also, the estimates of the standard deviations from both SDE and microsimulation approaches are approximately equal (Figure 2).

4.2 Case study: Breast cancer toxicity model

For the second example, we develop an SDE model for a published Markov decision model which estimates the life expectancy associated with receiving chemotherapy among 55 year-old breast cancer patients.[8] In developing the SDE model, we follow the steps in Section 3.

S1 The published (deterministic) cohort model has seven states ($s = 7$): healthy (S_H), healthy with treatment toxicity (S_T), metastasis (S_M), metastasis with treatment toxicity (S_{TM}), death due to treatment toxicity, death due to metastasis (S_{DM}), and death due to other causes (S_{DO}). There are 12 possible transitions among the states (Table 1).

S2 We construct the 1×12 vector of propensities $\mathbf{v}(\mathbf{n}, t)$. The elements of the propensity vector $\mathbf{v}$ is constructed in the following way. For example, an event of developing a metastasis from a healthy state ($S_H \to S_M$) can occur with a propensity equal to the transition rates times the at-risk individuals in S_H: $c_M N_H(t)$. We then specify the 12×7 matrix of changes, $\mathbf{d}$. For example, the transition $S_H \to S_M$ is associated with the vector $[-1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]$. The ordering of transitions (indexing of rows) in $\mathbf{d}$ should be consistent with the ordering in $\mathbf{v}(\mathbf{n}, t)$ (indexing of columns).

S3 Using the formula (Equation 4), we derive the 1×7 drift vector:

$$\mathbf{A}(\mathbf{n}, t) = \begin{pmatrix} -(c_T + c_M + c_{DO}(t))N_H(t) \\ c_T N_H(t) - (c_M + c_T + c_{DO}(t))N_T(t) \\ c_M(N_H(t) + N_T(t)) - (c_T + c_{DM} + c_{DO}(t))N_M(t) \\ c_M(N_T(t) + N_M(t)) - (c_{DT} + c_{DM} + c_{DO}(t))N_{TM}(t) \\ c_{DT}(N_T(t) + N_{TM}(t))) \\ c_{DM}(N_M(t) + N_{TM}(t)) \\ c_{DO}(t)(N_H(t) + N_T(t) + N_M(t) + N_{TM}(t)) \end{pmatrix}$$

The 7×7 diffusion matrix can be similarly constructed by using Equation 5. For example, the first and the fourth columns of $\mathbf{B}(\mathbf{n}, t)$ are equal to:

$$B_{:1}(\mathbf{n}, t) = \begin{pmatrix} (c_T + c_M + c_{DO}(t))N_H(t) \\ -c_T N_H(t) \\ -c_M N_H(t) \\ 0 \\ 0 \\ 0 \\ -C_{DO}(t)N_H(t) \end{pmatrix} \text{ and } B_{:4}(\mathbf{n}, t) = \begin{pmatrix} 0 \\ -c_M N_T(t) \\ -c_T N_M(t) \\ c_M N_T(t) + c_T N_M(t) + (c_{DT} + c_{DM} + c_{DO}(t))N_{TM}(t) \\ -c_{DT} N_{TM}(t) \\ -c_{DM} N_{TM}(t) \\ -C_{DO}(t)N_{TM}(t) \end{pmatrix},$$

respectively; where each $c_{\cdot}$ denotes a transition rate and the subscript corresponds to the (destination) state name. In the model, the rate of death due to other causes varies by age. The description and estimate of each parameter are given in [8].

S4 Given the drift vector and the diffusion matrix, we apply the Euler-Maruyama method to generate 10000 sample paths of a cohort of 1000 55 year-old women, using a 1-year time step. We use the R package YUIMA [18] and develop a Euler-Maruyama-based solver for Julia.

S5 From these sample paths, we estimate the life expectancy by summing the number of individuals across non-death health states for all times and divide the sum by the population size.

As a comparison, we also develop a microsimulation model with the following generator matrix:

$$Q(t) = \begin{pmatrix} -(c_T + c_M + c_{DO}(t)) & c_T & c_M & 0 & 0 & 0 & c_{DO}(t) \\ 0 & -(c_M + c_{DT} + c_{DO}(t)) & c_M & 0 & c_{DT} & 0 & c_{DO}(t) \\ 0 & 0 & -(c_T + c_{DM} + c_{DO}(t)) & c_T & 0 & c_{DM} & c_{DO}(t) \\ 0 & 0 & 0 & -(c_{DT} + c_{DM} + c_{DO}(t)) & c_{DT} & c_{DM} & c_{DO}(t) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C0_M & 0 & 0 & 0 & 0 \end{pmatrix},$$

and generate 10000 sample paths with 1-year time step. The life expectancies for the chemotherapy arm are estimated to be 16.9(standard deviation: 0.42) and 17.2(standard deviation: 0.41) for the SDE and microsimulation methods, respectively. Using a benchmarking tool in Julia,[14] the SDE model requires a considerably less computational time than the microsimulation model to generate the 10000 sample paths: 1.5 vs. 51.9 minutes.

5 Discussions

In this study, we introduce an SDE method which models not only the average behavior but also the variance of the stochastic process and can readily accommodate for time-varying parameters. In general, one can model the behavior of a stochastic dynamics system by considering either the time-evolution of the random vector (exemplified by SDE) or of the probability function of the random vector (exemplified by master equation). Although the probability function approach potentially provides a way to derive an explicit form of the probability function (i.e. the complete knowledge of the stochastic system) without relying on numerous numerical realizations of the system, the random vector approach is more applicable in the following context. In most applications, an analytical solution to the probability function method is almost inaccessible; thereby rendering numerical methods as the common solution method. In particular, some of the transition rates in most decision models are time-varying, for example: rate of death due to other causes is a time-dependent parameter. In this time-variant setting, practitioners often use Monte Carlo simulation (microsimulation) to estimate decision-relevant quantities from a cohort model. However, this approach, albeit flexible, is computationally expensive. To estimate the variance of the population counts across states, we need to replicate the microsimulation population a number of times *in addition to* the individual Monte Carlo runs within each population, which is a potentially computationally intensive task and is similar to applying uncertainty analysis to a microsimulation model.[19] In particular, if some of the events in the model are rare and/or the expected differences in the effect size among strategies are small, we expect the required number of Monte Carlo runs to be high. Although an SDE model also requires a Monte Carlo solution approach, the sample path of a cohort is generated more efficiently compared to microsimulation: SDE simulates a cohort *in a batch vs. individually.* Indeed, the cohort size would affect the computational speed of a microsimulation but not that of an SDE. The computational efficiency of the SDE vs. microsimulation approach as shown in the breast toxicity model example with time-varying parameter supports the aforementioned assertion.

SDE models have been applied in various fields such as finance[20], infectious disease [21], physics[22], and biology[23]. To date, no study has examined the potential application of an SDE model, as an alternative to a Markov cohort model, in decision modeling. The stochastic process underlying the commonly-used cohort model, defined as a continuous-time discrete-state stochastic process, has only been recently demonstrated.[3] Hence, the theoretical connection between a cohort model and an SDE was unknown prior to [3]. In this study, we provide a guidance to walk users through the process of developing an SDE model. Developers of Markov cohort models should be familiar with the first step which involves: enumerating the possible health states, assigning the allowable transitions to all pairs of states, and specifying the intensity of each transition (which can be time-invariant). One unique step to developing an SDE requires the user to specify the vector of changes for each transition, i.e. a vector denoting changes in the two populations involved in a transition: the initial state (-1) and the destination states ($+1$). After providing these vectors, the drift vector and the diffusion matrix can be calculated, which then completely specifies an SDE model. The

users are then tasked with choosing a solver and conducting a Monte Carlo simulation to estimate decision-relevant quantities. The implementation of the solver and the Monte Carlo method depends on the user's choice of programming language. In this study, the SDE models are implemented in R and Julia, and we use an R package with a comprehensive set of SDE-related functionalities to solve SDEs.[18]

In this study, the derivation of an SDE model follows a heuristic approach. A more rigorous method seeks to prove that the form of the equation $\Delta\mathbf{N}(t)$ satisfies (Equation 7) is a direct consequence of the following assumptions: $\Delta\mathbf{N}(t)$ is Markovian, smooth in its argument, and has finite mean and variance. Readers seeking a more detailed derivation should consult [24]. In addition, we do not show the relationship between the Fokker-Planck equation (FPE), which governs the time evolution of the probability function of a real-valued random vector (cf. Equation 3) and an SDE. In theory, the probability function of the solution to an SDE model satisfies an FPE. Moreover, an FPE is indeed a continuous approximation of a master equation thereby establishing a theoretical linkage between a cohort model (an evolution of a deterministic vector) and its stochastic counterpart, an SDE (an evolution of a stochastic vector). This fundamental fact can be shown through the machinery of stochastic calculus which is beyond the scope of this study. For a rigorous derivation, we refer readers to [25]. In summary, we opt for a simpler approach so potential users can readily appreciate the relationship between different representations of a cohort model and apply the methods in their works.

This study focuses on continuous-time cohort models and assumes that the discrete-time transition probability matrix is embeddable to a continuous-time model. Discrete-time Markov models have been the preferred method in many applications. However, as the dynamic process of interest often involves a disease progression component, a continuous-time model would be more appropriate in most applications.[26] Nevertheless, in some cases, an individual-level model naturally follows a discrete-time process, e.g. the transitions among living arrangements for a dementia patient occur in discrete time steps.[27] However, if we consider a sufficiently large cohort of individuals, the probability of one individual transitioning in a small time interval would very likely be nonzero. Nevertheless, future studies should investigate under what conditions a typical Markov model in decision modeling is embeddable.

6 Conclusions

Stochastic models can serve as powerful tools for aiding decision-makers in estimating decision-relevant outcomes and quantifying the impact of uncertainty on decision-making. Cohort model provides an easy-to-implement method for modeling recurrent events over time and has been used extensively in many applications ranging from clinical decision making to policy evaluations. The stochastic differential equation method expands the functionality of cohort models in practice by providing information on the variance of the process, decreasing computational time of Monte Carlo methods, and without necessitating additional data beyond what is usually required for cohort models.

Figure

Figure 1. A state-transition model diagram used in case study 1

Figure 2. Population trajectories for state S_3 in case study 1 by different methods and Monte Carlo sample sizes [CM: cohort model, MSM: microsimulation, SDE: stochastic differential equation, sd: standard deviation. Dashed lines indicate standard deviations and are not applicable for CM. Plots for 10000 samples are on top of each other]

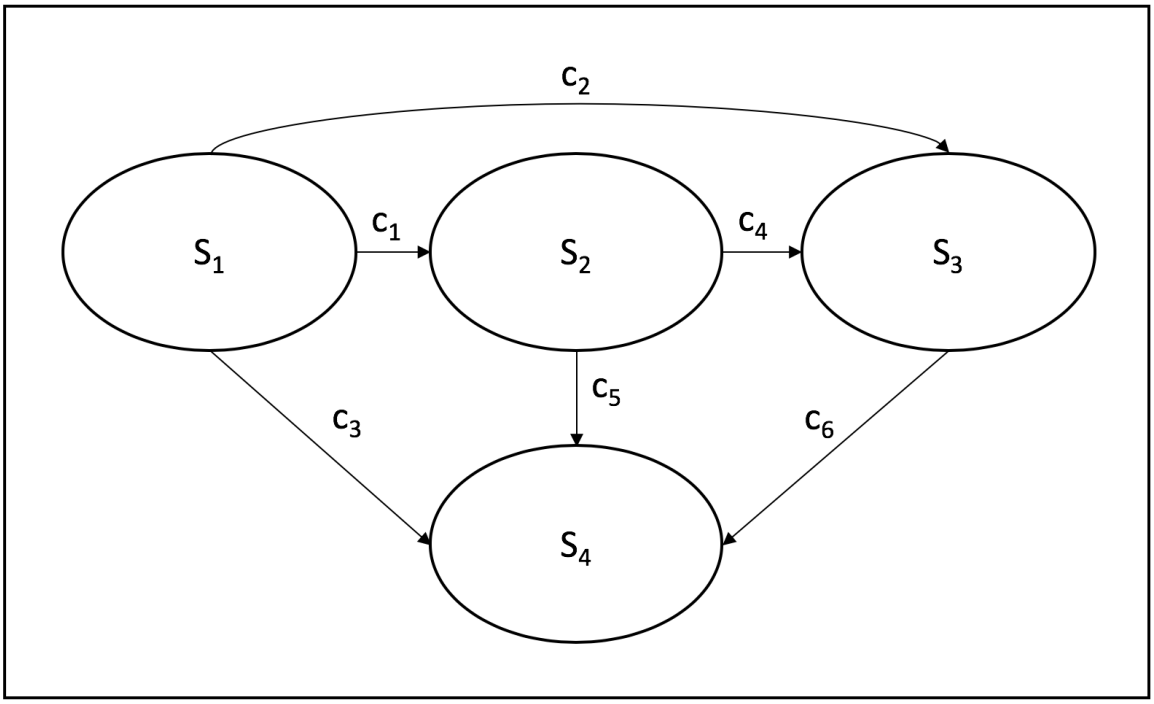

Figure 1: A state-transition model diagram used in case study 1

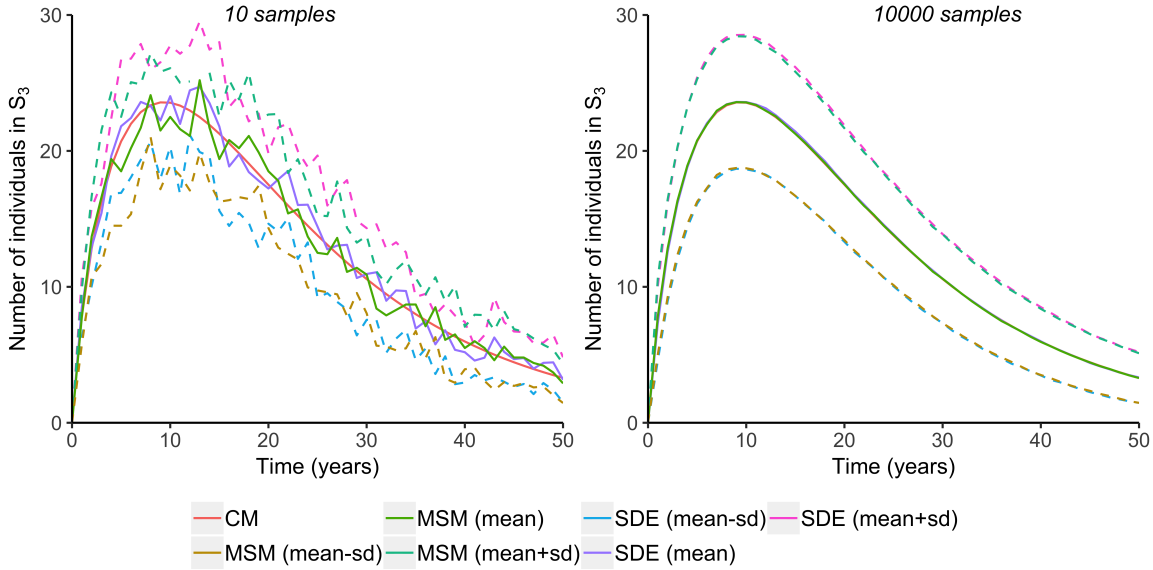

Figure 2: Population trajectories for state S_3 in case study 1 by different methods and Monte Carlo sample sizes [CM: cohort model, MSM: microsimulation, SDE: stochastic differential equation, sd: standard deviation. Dashed lines indicate standard deviations and are not applicable for CM. Plots for 10000 samples are on top of each other]

Table

Table 1. Propensities and vectors of changes for breast cancer toxicity model [$S_{\cdot}$: states; $c_{\cdot}$: transition rates; $N_{\cdot}$: number of individuals; $_{\cdot H}$: healthy; $_{\cdot T}$: treatment toxicity; $_{\cdot M}$: metastasis; $_{\cdot TM}$: metastasis with treatment toxicity; $_{\cdot DT}$: death due to treatment toxicity; $_{\cdot DM}$: death due to metastasis; $_{\cdot DO}$: death due to other causes]

Transitions	Propensities (elements of vector $\mathbf{v}$)	Vectors of changes (elements of matrix $\mathbf{d}$)
$S_H \rightarrow S_T$	$c_T N_H(t)$	$[-1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$
$S_H \rightarrow S_M$	$c_M N_H(t)$	$[-1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]$
$S_H \rightarrow S_{DO}$	$c_{DO} N_H(t)$	$[-1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]$
$S_T \rightarrow S_M$	$c_M N_T(t)$	$[0 \quad -1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0]$
$S_T \rightarrow S_{DT}$	$c_{DT} N_T(t)$	$[0 \quad -1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]$
$S_T \rightarrow S_{DO}$	$c_{DO} N_T(t)$	$[0 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]$
$S_M \rightarrow S_{TM}$	$c_T N_M(t)$	$[0 \quad 0 \quad -1 \quad 1 \quad 0 \quad 0 \quad 0]$
$S_M \rightarrow S_{DM}$	$c_{DM} N_M(t)$	$[0 \quad 0 \quad -1 \quad 0 \quad 0 \quad 1 \quad 0]$
$S_M \rightarrow S_{DO}$	$c_{DO} N_M(t)$	$[0 \quad 0 \quad -1 \quad 0 \quad 0 \quad 0 \quad 1]$
$S_{TM} \rightarrow S_{DT}$	$c_{DT} N_{TM}(t)$	$[0 \quad 0 \quad 0 \quad -1 \quad 1 \quad 0 \quad 0]$
$S_{TM} \rightarrow S_{DM}$	$c_{DM} N_{TM}(t)$	$[0 \quad 0 \quad 0 \quad -1 \quad 0 \quad 1 \quad 0]$
$S_{TM} \rightarrow S_{DO}$	$c_{DO} N_{TM}(t)$	$[0 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0 \quad 1]$

Table 1: Propensities and vectors of changes for breast cancer toxicity model in case study 2 [$S_{\cdot}$: states; $c_{\cdot}$: transition rates; $N_{\cdot}$: number of individuals; $_{\cdot H}$: healthy; $_{\cdot T}$: treatment toxicity; $_{\cdot M}$: metastasis; $_{\cdot TM}$: metastasis with treatment toxicity; $_{\cdot DT}$: death due to treatment toxicity; $_{\cdot DM}$: death due to metastasis; $_{\cdot DO}$: death due to other causes]

A Derivation of SDE

This section provides more details on the heuristic approach for deriving a stochastic differential equation (SDE) model. We use the same notations and definitions as described in the main text. Let $\mathbf{A}(\mathbf{n}, t)$ and $\mathbf{B}(\mathbf{n}, t)$ be the drift vector and the diffusion matrix, respectively. From the definition, we note that $\mathbf{B}(\mathbf{n}, t)$ is an $s \times s$ positive semi-definite matrix and hence its eigenvalues are non-negative. Another consequence of positive semi-definiteness is that $\Sigma(\Delta\mathbf{N}(t))$ has a unique positive semi-definite square root, i.e. we can find a matrix $\mathbf{S}$ such that $\mathbf{S}^2 = \frac{\Sigma(\Delta\mathbf{N}(t))}{\tau}$. One can show that $S = \mathbf{P}'\sqrt{\mathbf{H}}\mathbf{P}$,[11] where $\mathbf{P}$ is an orthogonal matrix (i.e. $\mathbf{P}\mathbf{P}^{'} = \mathbf{I}$ where $\mathbf{I}$ is an identity matrix) and $\mathbf{H}$ is a diagonal matrix of eigenvalues of $\mathbf{B}$.

For a sufficiently large $\mathbf{N}(t)$ and using the Central Limit Theorem, $\Delta\mathbf{N}(t)$ is approximately normal with mean $\mathbf{A}(\mathbf{n}, t)\tau$ and variance $\mathbf{B}(\mathbf{n}, t)\tau$. Let $\xi(t)$ be a s-dimensional standard Wiener process, i.e. $\xi(t) \sim \mathcal{N}(0, \mathbf{I})$.[5] Therefore we can express $\Delta\mathbf{N}(t)$ as a function of the drift vector and the diffusion matrix, i.e. $\Delta\mathbf{N}(t) = \mathbf{A}(\mathbf{n}, t)\tau + \mathbf{S}\sqrt{\tau}\xi(t) \sim \mathcal{N}(\mathbf{A}(\mathbf{n}, t), \mathbf{B}\tau)$. Then, the evolution of a real-valued stochastic cohort model can be represented by the following updating formula:

$$\begin{aligned}\mathbf{N}(t+\tau) &= \mathbf{N}(t) + \Delta\mathbf{N}(t) \\ &= \mathbf{N}(t) + \mathbf{A}(\mathbf{n}, t)\tau + \mathbf{S}\sqrt{\tau}\xi(t)\end{aligned} \tag{A.1}$$

If τ approaches 0, then we have $\xi_i(t)\sqrt{(}\tau)$ approaching $dW_i(t)$ at the limit ($W(t)$ is a Wiener process) and arrive at the *Itô* SDE:

$$d\mathbf{N}(t) = \mathbf{A}(\mathbf{n}, t)dt + \mathbf{B}(\mathbf{n}, t)d\mathbf{W}(t) \tag{A.2}$$

where $\mathbf{W}(t)$ be a s-dimensional Wiener process. if for all t and t_0, the following integral equation holds:

$$\mathbf{N}(t) - \mathbf{N}(t_0) = \int_{t_0}^{t} \mathbf{A}(\mathbf{n}, s)ds + \int_{t_0}^{t} \mathbf{B}(\mathbf{n}, s)d\mathbf{W}(s) \tag{A.3}$$

The state-wise SDE is:

$$dN_i(t) = A_i(\mathbf{N}(t))dt + \sum_{j=1}^{s} B_{ij}(\mathbf{N}(t))dW_j$$

with the Ito integral equation:

$$N_i(t) - N_i(t_0) = \int_{t_0}^{t} A_i(\mathbf{N}(s), s)du + \int_{t_0}^{t} \sum_{j=1}^{s} B_{ij}(\mathbf{N}(s), s)d\mathbf{W}_j(s)$$

for $i = 1, 2, \ldots, s$.

References

[1] J Robert Beck and Stephen G Pauker. The markov process in medical prognosis. *Med Decis Making*, 3(4): 419–458, 1983.

[2] MG Myriam Hunink, Milton C Weinstein, Eve Wittenberg, Michael F Drummond, Joseph S Pliskin, John B Wong, and Paul P Glasziou. *Decision making in health and medicine: integrating evidence and values*. Cambridge University Press, 2014.

[3] Rowan Iskandar. A theoretical foundation for state-transition cohort models in health decision analysis. *PloS one*, 13(12):e0205543, 2018.

[4] Mukhtar Ullah and Olaf Wolkenhauer. Family tree of markov models in systems biology. *IET systems biology*, 1(4):247–254, 2007.

[5] Thomas C Gard. *Introduction to Stochastic Differential Equations. Monographs and Text-books in pure and applied mathematics*. Dekker, Inc, 1988.

[6] Aaron A Stinnett and John Mullahy. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical decision making*, 18(2_suppl):S68–S80, 1998.

[7] Karl Claxton. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of health economics*, 18(3):341–364, 1999.

[8] Fernando Alarid-Escudero, Anne H Blaes, and Karen M Kuntz. Trade-offs between efficacy and cardiac toxicity of adjuvant chemotherapy in early-stage breast cancer patients: Do competing risks matter? *The breast journal*, 23(4):401–409, 2017.

[9] Gerardus WJ Frederix, Johan GC van Hasselt, Johan L Severens, Anke M Hövels, Alwin DR Huitema, Jan AM Raaijmakers, and Jan HM Schellens. Development of a framework for cohort simulation in cost-effectiveness analyses using a multistep ordinary differential equation solver algorithm in r. *Medical Decision Making*, 33(6): 780–792, 2013.

[10] Edward Allen. *Modeling with Itô stochastic differential equations*, volume 22. Springer Science & Business Media, 2007.

[11] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[12] Gisiro Maruyama. Continuous markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48–90, 1955.

[13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

[14] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. URL `https://doi.org/10.1137/141000671`.

[15] Decision Analysis in R for Technologies in Health (DARTH) workgroup. Decision Analysis in R for Technologies in Health, 2019.

[16] Hawre Jalal, Petros Pechlivanoglou, Eline Krijkamp, Fernando Alarid-Escudero, Eva Enns, and MG Myriam Hunink. An overview of r in health decision sciences. *Medical Decision Making*, 37(7):735–746, 2017.

[17] Nicky J Welton and AE Ades. Estimation of markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making*, 25(6):633–645, 2005.

[18] Alexandre Brouste, Masaaki Fukasawa, Hideitsu Hino, Stefano M. Iacus, Kengo Kamatani, Yuta Koike, Hiroki Masuda, Ryosuke Nomura, Teppei Ogihara, Yasutaka Shimuzu, Masayuki Uchida, and Nakahiro Yoshida. The yuima project: A computational framework for simulation and inference of stochastic differential equations. *Journal of Statistical Software*, 57(4):1–51, 2014. doi: 10.18637/jss.v057.i04. URL `http://www.jstatsoft.org/v57/i04/`.

[19] Anthony O'hagan, Matt Stevenson, and Jason Madan. Monte carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using anova. *Health economics*, 16(10):1009–1023, 2007.

[20] Nicole El Karoui, Shige Peng, and Marie Claire Quenez. Backward stochastic differential equations in finance. *Mathematical finance*, 7(1):1–71, 1997.

[21] Alison Gray, David Greenhalgh, Liangjian Hu, Xuerong Mao, and Jiafeng Pan. A stochastic differential equation sis epidemic model. *SIAM Journal on Applied Mathematics*, 71(3):876–902, 2011.

[22] Emil B Iversen, Juan M Morales, Jan K Møller, and Henrik Madsen. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting*, 32(3):981–990, 2016.

[23] Kuang-Chi Chen, Tse-Yi Wang, Huei-Hun Tseng, Chi-Ying F Huang, and Cheng-Yan Kao. A stochastic differential equation model for quantifying transcriptional regulatory network in saccharomyces cerevisiae. *Bioinformatics*, 21(12):2883–2890, 2005.

[24] Daniel T Gillespie. The multivariate langevin and fokker–planck equations. *American Journal of Physics*, 64 (10):1246–1257, 1996.

[25] Iosif Ilyich Gikhman and Anatoli Vladimirovich Skorokhod. Stochastic differential equations. In *The Theory of Stochastic Processes III*, pages 113–219. Springer, 2007.

[26] Marta O Soares and Luísa Canto e Castro. Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics*, 30(12):1101–1117, 2012.

[27] Eric Jutkowitz, Robert L Kane, Joseph E Gaugler, Richard F MacLehose, Bryan Dowd, and Karen M Kuntz. Societal and family lifetime cost of dementia: implications for policy. *Journal of the American Geriatrics Society*, 65(10):2169–2175, 2017.