A Comparison of Stochastic Simulation Algorithms in Infection Models

H.T. Banks

Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine
North Carolina State University

SiMCRT2011
Intl. Workshop on Simulation and Modeling
Kobe University, Kobe, Japan
November 1-3, 2011
REU efforts from Summer, 2011, jointly with

- Shuhua Hu (North Carolina State University)
- Michele Joyner (East Tennessee State University)
- Anna Broido (Boston College)
- Brandi Canter (East Tennessee State University)
- Kaitlyn Gayvert (State University of New York at Geneseo)
- Kathryn Link (Bryn Mawr College)

Two infection models and two different modeling approaches in simulation studies.
Infection Models:

- Models of the transmission of Vancomycin-Resistant Enterococcus (VRE)—a group of hospital-acquired bacterial infections (group of bacterial species of genus enterococcus that is resistant to the antibiotic vancomycin).

- Models of Human Immunodeficiency Virus (HIV) (a retrovirus that targets the CD4+ T-cells in the immune system) that can often lead to AIDS)

Modeling Approaches:

- Discrete valued continuous time Markov chain (CTMC) models as often used when dealing with low species count—in particular, for small population or sample sizes

- Deterministic ordinary differential equations (ODE) to approx large populations/sample sizes with continuum model
Discrete Valued Markov Chains for VRE Spread in Hospitals

- Patients in a hospital unit are classified by compartments or states as either (i) uncolonized $U(t)$, (ii) VRE colonized $C(t)$, or (iii) VRE colonized in isolation $J(t)$.

- Patients are admitted to the hospital unit at a rate $\Lambda$ per day and some fraction $m$ are already VRE colonized. The transition from one compartment to another follows an exponential distribution and the expected mean duration within a compartment is given by the inverse of the parameter of the exponential distribution.
• A hand-hygiene policy applied to health care workers on isolated VRE colonized patients reduces infectivity by a factor of \( \gamma \) (\( 0 < \gamma < 1 \)). It is assumed VRE colonization periods last from weeks to months and because spontaneous decolonization occurs infrequently, clearance of the bacteria is not considered in the model. VRE colonized patients are moved into isolation at a rate \( \alpha \).
Figure 1: Compartmental VRE model
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U(t)$</td>
<td>Number of uncolonized patients</td>
</tr>
<tr>
<td>$C(t)$</td>
<td>Number of VRE colonized patients</td>
</tr>
<tr>
<td>$J(t)$</td>
<td>Number of VRE colonized patients in isolations</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>Patients admission rate</td>
</tr>
<tr>
<td>$m$</td>
<td>VRE colonized patients on admission rate</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Effective contact rate</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>HCW hand hygiene compliance rate</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Patient Isolation rate</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>Uncolonized patients discharged rate</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>VRE colonized patients discharged rate</td>
</tr>
</tbody>
</table>
The dynamics of the VRE colonization of patients in a hospital unit with \( N \) beds (we assume beds always at capacity, i.e., full, so \( N \) is also total number of patients in systems) are modeled as a continuous time Markov Chain (MC) with discrete state space embedded in \( \mathbb{Z}^3 \), the set of 3-dimensional column vectors with integer components. In this case, the population of patients (a total of \( N \)) is considered discrete (i.e., VRE colonization occurs in units of whole individuals) and the timing of events is a probabilistic process. The state of the Markov chain at time \( t \) is denoted by

\[
\{U(t) = i, C(t) = j, J(t) = k\}, \quad t \geq 0 \text{ and } i, j, k \in \{0, 1, \ldots N\}.
\]

The probability during a small time interval, \( dt \), of transiting from one state to another is described by transition probabilities
Notation: $N$ is total number beds available in the hospital, and
\[ \{X^N(t), t \geq 0\} \] is a continuous time Markov chain with
\[ X^N = (X_1^N, X_2^N, X_3^N)^T = (U^N, C^N, J^N)^T, \]
defined by

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1^N(t)$</td>
<td>Number $U^N(t)$ of uncolonized patients at time $t$ in a hospital with $N$ beds</td>
</tr>
<tr>
<td>$X_2^N(t)$</td>
<td>Number $C^N(t)$ of VRE colonized patients at time $t$ in a hospital with $N$ beds</td>
</tr>
<tr>
<td>$X_3^N(t)$</td>
<td>Number $J^N(t)$ of VRE colonized patients in isolation at time $t$ in a hospital with $N$ beds</td>
</tr>
</tbody>
</table>

Table 2: State variables for stochastic VRE model.
Transition Probabilities

In any small time interval of length $\Delta t$, we assume $\{X^N(t), t \geq 0\}$ jumps from state $x^N$ to $x^N + v_j$ with probability $\lambda_j(x^N)\Delta t + o(\Delta t)$, that is,

$$\text{Prob}\{X^N(t + \Delta t) = x^N + v_j \mid X^N(t) = x^N\} = \lambda_j(x^N)\Delta t + o(\Delta t), \quad j = 1, 2, \ldots, l, \quad (1)$$

where $x^N = (x_1^N, x_2^N, x_3^N)^T \in \mathbb{Z}^3$, $v_j \in \mathbb{Z}^3$, and $\lambda_j$ is the transition rate for reaction $j$. Here and below $\mathbb{Z}^n$ is the set of $n$-dimensional column vectors with integer components.
### Transition Rates:

<table>
<thead>
<tr>
<th>Transitions</th>
<th>$\lambda_j(x^N)$</th>
<th>$v_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1^N \to X_2^N$</td>
<td>$m\mu_1 x_1^N + \beta x_1^N (x_2^N + (1 - \gamma)x_3^N)$</td>
<td>$-e_1 + e_2$</td>
</tr>
<tr>
<td>$X_3^N \to X_2^N$</td>
<td>$m\mu_2 x_3^N$</td>
<td>$e_2 - e_3$</td>
</tr>
<tr>
<td>$X_2^N \to X_1^N$</td>
<td>$(1 - m)\mu_2 x_2^N$</td>
<td>$e_1 - e_2$</td>
</tr>
<tr>
<td>$X_3^N \to X_1^N$</td>
<td>$(1 - m)\mu_2 x_3^N$</td>
<td>$e_1 - e_3$</td>
</tr>
<tr>
<td>$X_2^N \to X_3^N$</td>
<td>$\alpha x_2^N$</td>
<td>$-e_2 + e_3$</td>
</tr>
</tbody>
</table>

Table 3: Transition rates $\lambda_j(x^N)$ as well as the corresponding state changes $v_j$ for the stochastic VRE model (1), $j = 1, 2, \ldots, 5$. 
By the Kurtz Limit Theorem (SLLN) ordinary differential equations can be used to approximate a (density dependent) CTMC when the sample size $N$ is sufficiently large, i.e., averaged sample paths of $X^N(t)$ can be approximated by solutions to the system of ordinary differential equations given by

\[
\begin{align*}
\frac{d\bar{U}(t)}{dt} &= (1 - m)[\mu_1 \bar{U}(t) + \mu_2 (\bar{C}(t) + \bar{J}(t))] \\
&\quad - \beta \bar{U}(t)[\bar{C}(t) + (1 - \gamma)\bar{J}(t)] - \mu_1 \bar{U}(t) \\
\frac{d\bar{C}(t)}{dt} &= m[\mu_1 \bar{U}(t) + \mu_2 (\bar{C}(t) + \bar{J}(t))] \\
&\quad + \beta \bar{U}(t)[\bar{C}(t) + (1 - \gamma)\bar{J}(t)] - (\alpha + \mu_2)\bar{C}(t) \\
\frac{d\bar{J}(t)}{dt} &= \alpha \bar{C}(t) - \mu_2 \bar{J}(t),
\end{align*}
\]

with initial conditions $\bar{U}(0) = U_0$, $\bar{C}(0) = C_0$, and $\bar{J}(0) = J_0$. 
Figure 2: Sample of 5 stochastic realizations in comparison to numerical solution for the deterministic model; \( N = 37 \) patients.
Figure 3: Sample of 5 stochastic realizations for each compartment in comparison to the numerical solution of the deterministic model for $N = 137, 537$, $t_{stop} = 500$. 
Figure 4: Sample of 5 stochastic realizations for each compartment in comparison to the numerical solution of the deterministic model for $N = 937, 2037$, $t_{stop} = 500$. 
• Forward simulations of Markov chains can be very costly computationally (more in a moment on Gillespie stochastic simulation algorithm) so Kurtz approximation can be very useful

• How to do inverse problems?????? Neat application of Kurtz approximation: Transition probabilities in stochastic model are rate coefficients in Kurtz approximation system—Idea: View data as a sample for a large population system described by ODE approximations—use it with available methodology for corresponding inverse problems, thereby obtaining transition probabilities for stochastic model!!!
References


Stochastic Simulation Algorithm-Gillespie

Gold standard or method of choice of stochastic simulation algorithms for MC: the Gillespie algorithm—Gillespie (1976) to simulate time evolution of stochastic formulation of chemical kinetics (takes into account that molecules come in whole numbers as well as inherent degree of randomness in their dynamical behavior). Assume $X = (X_1, X_2, ..., X_n)^T$ represents state variables of system where $X_i(t) =$ number in state $X_i$ at time $t$ ($X_i$ may be the number of patients, cells, species, etc). Furthermore, it is assumed $l$ transitions (often referred to as reaction channels in biochemistry literature) are possible with associated transition rates $\lambda_i, i = 1, ..., l$. 
Step 1. Initialize the state of the system \( x_0 \);

Step 2. For the given state \( x \) of the system, calculate the transition rates \( \lambda_i(x), i = 1, ..., l \);

Step 3. Calculate the sum of all transition rates, \( \lambda = \sum_{i=1}^{l} \lambda_i(x) \);

Step 4. Simulate the time, \( \tau \), until the next transition by drawing from an exponential distribution with mean \( 1/\lambda \);

Step 5. Simulate the transition type by drawing from the discrete distribution with probability \( \text{Prob}(\text{transition} = i) = \lambda_i(x)/\lambda \). Generate a random number \( r_2 \) from a uniform distribution and choose the transition as follows: If \( 0 < r_2 < \lambda_1(x)/\lambda \), choose transition 1; if \( \lambda_1(x)/\lambda < r_2 < (\lambda_1(x) + \lambda_2(x))/\lambda \) choose transition 2, and so on;

Step 6. Update the new time \( t = t + \tau \) and the new system state;

Step 7. Iterate steps 2-6 until \( t \geq t_{\text{stop}} \).
**Tau-Leaping Methods**

Gillespie method keeps track of each transition—can be impractical to implement for certain applications due to computational time required - Gillespie proposed approximate procedure, the tau-leaping method, which accelerates computational time while only sustaining a small loss in accuracy. **Idea:** Instead of taking incremental steps in time, keeping track of $X(t)$ at each time step as in SSA method, tau-leaping method *leaps* from one subinterval to next, approximating how many transitions take place during a given subinterval. **Tacit assumption** *leap condition*: value of leap, $\tau$, is sufficiently small that there is no significant change in value of transition rates during the subinterval $[t, t + \tau]$. The tau-leaping method advantage: simulating many transitions in one *leap* while not loosing significant accuracy, resulting in speed up in computational time—we considered two tau-leaping methods: an explicit and an implicit version.
An Explicit Tau-Leaping Method

- Explicit tau-leaping method based on an explicit formulation for update in number of species $X$ at time $t + \tau$, given $X(t) = x$. The basic explicit tau-leaping method approximates $K_j$, the number of times a transition $j$ is expected to occur within the time interval $[t, t + \tau]$, by a Poisson random variable $P_j(\lambda_j(x), \tau)$ with mean (and variance) $\lambda_j(x)\tau$.

- Once number of transitions estimated, approximate number of species, known as the tau-leaping approximation, of $X$ at time $t + \tau$ is given by

$$X(t + \tau) = x + \sum_{j=1}^{l} P_j(\lambda_j(x), \tau)v_j \tag{2}$$

with $v_j = (v_{1j}, \ldots, v_{nj})^T$ where $v_{ij}$ is change in state variable $X_i$ caused by transition $\lambda_j$. 

22
Remark: The process for selecting $\tau$ is critical in the tau-leaping method. If $\tau$ is chosen too small, tau-leaping will essentially stop, leading to the standard SSA algorithm; on the other hand, if the value of $\tau$ is too large, the leap condition may not be satisfied, possibly causing significant inaccuracies in the simulation.
Figure 5: Comparison of computational times for different algorithms (SSA, Explicit Tau-Leaping and Implicit Tau-Leaping) for VRE model for an average of five typical simulation runs with $N$ varying from 37, 185, 370, 1850, 3700, 18500, 37000.
Remarks:

• From this figure, we see that the computational times for all the algorithms increase as the value $N$ increases. This is expected for the SSA as the mean time stepsize for the SSA is the inverse of the sum of all transition rates, which increases as $N$ increases (roughly proportional to $N^2$ as can be seen from the transition rates illustrated in Table 3).

• For the explicit tau-leaping method we found, for all the $N$ that we tried, the value of $\tau_1$ is often less than $10/\lambda$, which implies that the SSA is implemented most of the time as opposed to the tau-leaping method. This also explains why the SSA and the explicit tau-leaping perform similarly. The same thing is also observed for the implicit tau-leaping method when $N = 37$ and 185.

• However, when $N$ increases to 370, we found that the implicit
tau-leaping method requires significant time for implementation, and we also found that its time stepsize in this case is still not significantly larger than those of the SSA. Note that systems of nonlinear equations must be solved for the implicit tau-leaping method. Hence, the computational time in this case is expected to be larger than for the other two methods (this can be observed from this figure).

- As $N$ continues increasing, we see that the computational times for the implicit tau-leaping is similar to those of the SSA and the explicit tau-leaping method; this is because the time stepsize becomes significantly larger than those of these two methods in which solving systems of nonlinear equations comprises the main time consuming cost.
An HIV Model

- HIV is a retrovirus that targets the CD4+ T-cells in the immune system.

- Wide variety of mathematical models have been proposed to describe the various aspects of in-host HIV infection dynamics—most basic of these models typically include two or three key dynamic compartments: virus, uninfected target cells, and infected cells.

- These compartmental depictions lead to systems of linear or nonlinear ODE in terms of state variables representing the concentrations in each compartment and parameters describing viral production and clearance, cell infection and death rate, treatment efficacy, etc.
• Stochastic model we use to demonstrate the computational efficiency of the SSA, and explicit and implicit tau-leaping methods is based on the deterministic HIV model proposed and validated with clinical data by NCSU group. Data fitting results validate this model and verify that it provides reasonable fits to all 14 patients studied—Moreover it has impressive predictive capability
Deterministic HIV Model

Figure 6: Solid gray arrows indicate birth/input. PI and RTI denote protease inhibitors and reverse transcriptase inhibitors, respectively.
### State Variables

<table>
<thead>
<tr>
<th>States</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>cells/µl-blood</td>
<td>conc. of uninfected activated CD4+ T-cells</td>
</tr>
<tr>
<td>$T_1^*$</td>
<td>cells/µl-blood</td>
<td>conc. of infected activated CD4+ T-cells</td>
</tr>
<tr>
<td>$T_2$</td>
<td>cells/µl-blood</td>
<td>conc. of uninfected resting CD4+ T-cells</td>
</tr>
<tr>
<td>$T_2^*$</td>
<td>cells/µl-blood</td>
<td>conc. of infected resting CD4+ T-cells</td>
</tr>
<tr>
<td>$V_I$</td>
<td>RNA copies/µl</td>
<td>conc. of infectious free virus</td>
</tr>
<tr>
<td>$E_1$</td>
<td>cells/µl-blood</td>
<td>conc. of HIV-specific effector CD8+ T-cells</td>
</tr>
<tr>
<td>$E_2$</td>
<td>cells/µl-blood</td>
<td>conc. of HIV-specific memory CD8+ T-cells</td>
</tr>
</tbody>
</table>

Table 4: Model states for the deterministic HIV model.
Corresponding compartmental ODE model:

\[
\dot{T}_1 = -d_T T_1 - \beta_T V_I T_1 - \gamma_T T_1 + n_T \left( \frac{a_T V_I}{V_I + \kappa_V} + a_A \right) T_2,
\]

\[
\dot{T}_1^* = \beta_T V_I T_1 - \delta_V T_1^* - \delta_E E_1 T_1^* - \gamma_T T_1^* + n_T \left( \frac{a_T V_I}{V_I + \kappa_V} + a_A \right) T_2^*,
\]

\[
\dot{T}_2 = \zeta_T \frac{\kappa_s}{V_I + \kappa_s} + \gamma_T T_1 - d_T T_2 - \beta_T V_I T_2 - \left( \frac{a_T V_I}{V_I + \kappa_V} + a_A \right) T_2,
\]

\[
\dot{T}_2^* = \gamma_T T_1^* + \beta_T V_I T_2 - d_T T_2^* - \left( \frac{a_T V_I}{V_I + \kappa_V} + a_A \right) T_2^*,
\]

\[
\dot{V}_I = n_V \delta_V T_1^* - c V_I - (\beta_T T_1 + \beta_T T_2) V_I,
\]

\[
\dot{E}_1 = \zeta_E + \frac{b_E T_1^*}{T_1^* + \kappa_b} E_1 - \frac{d_E T_1^*}{T_1^* + \kappa_d} E_1 - d_E E_1 E_1 - \gamma_E \frac{T_1 + T_1^*}{T_1 + T_1^* + \kappa_\gamma} E_1
\]

\[
+ n_E \frac{a_E V_I}{V_I + \kappa_V} E_2,
\]

\[
\dot{E}_2 = \gamma_E \frac{T_1 + T_1^*}{T_1 + T_1^* + \kappa_\gamma} E_1 + \frac{b_E \kappa_b}{E_2 + \kappa_b} E_2 - d_E E_2 E_2 - \frac{a_E V_I}{V_I + \kappa_V} E_2,
\]
The Stochastic HIV Model

Corresponding stochastic HIV model based on the deterministic model. Let \( \nu \) denote the volume of blood (in units \( \mu l \)-blood), and the parameter vector \( \kappa = (\kappa_V, \kappa_s, \kappa_{b1}, \kappa_d, \kappa_\gamma, \kappa_{b2})^T \) where \( \kappa_V, \kappa_s, \kappa_{b1}, \kappa_d, \kappa_\gamma, \kappa_{b2} \) are saturation constants in the deterministic model. Then we define \( k = \nu \kappa \) with

\[
k = (k_V, k_s, k_{b1}, k_d, k_\gamma, k_{b2})^T
\]

– used in the transition rates for stochastic model. Let \( \{X^{\nu}(t), t \geq 0\} \) be a pure jump Markov process with \( X^{\nu} = (X_1^{\nu}, X_2^{\nu}, \ldots, X_7^{\nu})^T \).
Meanings of random variables $X_i^\nu$, $i = 1, 2, \ldots, 7$:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1^\nu(t)$</td>
<td>no. non-infected activated CD4+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_2^\nu(t)$</td>
<td>no. infected activated CD4+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_3^\nu(t)$</td>
<td>no. non-infected resting CD4+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_4^\nu(t)$</td>
<td>no. infected resting CD4+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_5^\nu(t)$</td>
<td>no. RNA copies of infectious free virus in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_6^\nu(t)$</td>
<td>no. HIV-specific effector CD8+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
<tr>
<td>$X_7^\nu(t)$</td>
<td>no. of HIV-specific memory CD8+ T-cells in $\nu$ $\mu$l-blood</td>
</tr>
</tbody>
</table>

Table 5: State variables for the stochastic HIV model.
In any small time interval of length $\Delta t$, we assume that there is only one event (or reaction) that occurs (e.g., a cell dies, a cell becomes infected, an activated cell is differentiated into a resting cell, a resting cell becomes activated), and the process $\{X^\nu(t), t \geq 0\}$ jumps from state $x^\nu$ to $x^\nu + v_j$ with probability $\lambda_j(x^\nu)\Delta t + o(\Delta t)$, that is,

$$\text{Prob} \{X^\nu(t + \Delta t) = x^\nu + v_j \mid X^\nu(t) = x^\nu\} = \lambda_j(x^\nu)\Delta t + o(\Delta t), \quad j = 1, 2, \ldots, l,$$

where $x^\nu = (x_1^\nu, x_2^\nu, \ldots, x_7^\nu)^T \in \mathbb{Z}^7$, $v_j \in \mathbb{Z}^7$, and $\lambda_j$ is the transition rate for the $j$th transition, $j = 1, 2, \ldots, l$. These transition rates ($l = 19$ here) are obtained from rate constants in ODE modeling, assuming a burst production for the viral load.
Figure 7: Comparison of computational times of different algorithms (SSA, Explicit Tau-Leaping and Implicit Tau-Leaping) for an average of five typical simulation runs $\nu$ varying from $10, 50, 10^2, 2 \times 10^2, 5 \times 10^2, 10^3$ for the SSA and $10, 50, 10^2, 2 \times 10^2, 5 \times 10^2, 10^3, 10^4, 10^5, 10^6, 5 \times 10^6$ for the explicit and implicit tau-leaping schemes.
• Find: computational times for the SSA increase as the value $\nu$ increases (again expected as the mean time stepsize for the SSA is the inverse of the sum of all transition rates, which increases as $\nu$ increases (roughly proportional to $\nu$, as can be seen from the transition rates. In addition, even with $\nu = 10^3$, it took the SSA more than 8000 seconds for one sample path (which is why we did not run any simulations for the SSA when $\nu$ is greater than $10^3$). Hence, impractical to implement the SSA if we want to run this HIV model for a normal person (generally having approximately $5 \times 10^6$ $\mu$l-blood). This is expected due to the large value of uninfected resting CD4+ T cells.

• Also see that computational times for explicit tau-leaping method increase as the value $\nu$ increases from 10 to 50, and decrease as $\nu$ increases to 100. Then its computational times decrease dramatically as the value of $\nu$ increases from 100 to $10^4$. 


and stabilizes somewhat for $\nu \geq 10^4$. (The increase of computational times as $\nu$ increases from 10 to 50 is because $\tau_1$ is so small that a large number of SSA steps are implemented instead of tau-leaping.)

- Also observe that computational times for implicit tau-leaping method decrease as $\nu$ increases when $\nu \leq 10^4$ and then stabilizes there for $\nu \geq 10^4$. In addition, we see computational times for implicit tau-leaping is significantly higher than those of the SSA and the explicit tau-leaping at $\nu = 10$ (because under this case the implicit tau-leaping is implemented many times (solving systems of nonlinear equations in each implicit tau-leaping step is costly) and the time stepsize is not significantly larger than those of the other two methods).
Concluding Remarks:

• Based on findings, for smaller values of $\nu$ (less than 100) the SSA is the choice due to its simplicity, accuracy and efficiency. However, for larger values the tau-leaping methods are definitely the choice with implicit tau-leaping performing better than explicit tau-leaping (expected–stiffness of the system (large variations in both parameter values and state variables).

• Why is this so important? In many infections, wish to study early acute stages (some populations are small–ODE not valid-need discrete valued stochastic models); as population grows, move to limiting ODE to investigate large populations!

• Modeling and computational methodology important in many applications: just-in-time production networks, manufacturing and delivery, logistics/supply, small/large poplns (multi-scale)