Virus propagation with randomness

Benito M. Chen-Charpentier\textsuperscript{a,}\textsuperscript{*}, Dan Stanescu\textsuperscript{b}

\textsuperscript{a} Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019-0408, USA
\textsuperscript{b} Department of Mathematics, University of Wyoming, Laramie, WY 82071-3036, USA

\begin{abstract}
Viruses are organisms that need to infect a host cell in order to reproduce. The new viruses leave the infected cell and look for other susceptible cells to infect. The mathematical models for virus propagation are very similar to population and epidemic models, and involve a relatively large number of parameters. These parameters are very difficult to establish with accuracy, while variability in the cell and virus populations and measurement errors are also to be expected. To deal with this issue, we consider the parameters to be random variables with given distributions. We use a non-intrusive variant of the polynomial chaos method to obtain statistics from the differential equations of two different virus models. The equations to be solved remain the same as in the deterministic case; thus no new computer codes need to be developed. Some examples are presented.
\end{abstract}

1. Introduction

Viruses cause a large number of diseases in humans, animals, and other organisms. Among the most common are the influenza virus, HIV, rotavirus, and the tobacco mosaic virus. Viruses consist of genetic material, a protein coating, and sometimes a lipid envelope when it is outside a cell. Viruses are parasites that need to infect a host to reproduce. The virus attaches to a receptor in, and eventually gets into, a cell. It sheds its protein coating and uses the cell genetic material to reproduce its own genetic material. At the same time, new viral proteins are built that combine with the new genetic material to form new viruses that burst the cell to exit and are ready to infect new cells. There are a wide variety of mathematical models to simulate different virus processes. See, for example, [1–4]. Here, we consider two models: a simple model that includes only uninfected cells, infected cells, and free viruses, and a model that also considers CTL dynamics. Cytotoxic T lymphocytes (CTLs) form a branch of the immune system that is essential in the fight against many viral infections [4].

As is the case in many epidemic models, the transmission or infection coefficients and other parameters in virus models are difficult to determine, and they have variability due to errors in measurements, differences in population size, and other sources of uncertainty. There is also variability due to errors in virus replication and mutation. One way to incorporate these errors in mathematical models is by considering the coefficients to be random variables; thus the time-dependent solutions become stochastic processes. Such mathematical models involving differential equations in which some or all the coefficients are considered random variables or that incorporate stochastic noise terms have been used in the last few decades to deal with errors and uncertainty; see [5,6].

Monte Carlo methods [7] are a traditional way to deal with randomness because of their simplicity; however, they usually require many realizations for reasonable accuracy. Other possible choices are the method of moments [8] and polynomial chaos. In this paper, we use a non-intrusive version of the method of polynomial chaos (NIPC) to deal with the randomness in the model equations.

\textsuperscript{*} Corresponding author. Tel.: +1 817 272 3913; fax: +1 817 272 5802.
\textit{E-mail} addresses: bmchen@uta.edu, bchen@uwyo.edu (B.M. Chen-Charpentier), stanescu@uwyo.edu (D. Stanescu).
The rest of the paper is organized as follows. In Section 2, we describe the virus propagation models. Section 3 presents the non-intrusive polynomial chaos method. Numerical results are given in Section 4, and finally we draw our conclusions in Section 5.

2. Virus propagation models

Two virus population models are considered for the purpose of this work. The first model was proposed by Nowak and Bangham [9] in their study of the human T cell leukemia virus (HTLV-1) and the human immunodeficiency virus (HIV-1). It is given by the following system of differential equations:

\[ \begin{align*}
\dot{x} &= \lambda - dx - \beta xv \\
\dot{y} &= \beta xv - ay \\
\dot{v} &= ky - uv.
\end{align*} \]  

(1)

In these equations, \( x \) is the population of healthy, uninfected cells, which are constantly produced by the body at a rate \( \lambda \), die at a rate \( dx \) proportional to the current population, and become infected at a rate \( \beta xv \) proportional to their interaction with the virus population \( v \). Infected cells \( y \) die at a rate \( ay \), while new viruses are produced at a rate \( ky \) proportional to the number of infected cells, and die at a rate \( uv \). Failure to establish an infection, when \( R_0 = \frac{\lambda \beta k}{dau} < 1 \), leads to the equilibrium solution, \( x^* = \frac{\lambda}{d}, y^* = \frac{\lambda \beta k}{a}, v^* = 0 \), while successful infection, \( R_0 > 1 \), leads to the following equilibrium solution:

\[ \begin{align*}
x^* &= \frac{au}{\beta k^2}, \\
y^* &= \frac{\beta k - dau}{a \beta} \frac{\lambda}{u}, \\
v^* &= \frac{\lambda \beta k}{a}. 
\end{align*} \]  

(2)

The virus population has a much faster growth rate than the populations of uninfected and infected cells, so a common assumption is that the virus population is almost steady, \( v = ky/u \). Then system (1) simplifies to

\[ \begin{align*}
\dot{x} &= \lambda - dx - Bxy \\
\dot{y} &= Bxy - ay,
\end{align*} \]  

(3)

with \( B = \beta k/u \).

The second model considered here [9,10] incorporates CTL dynamics. To fight the infection, CTLs kill infected cells and thus try to stop the spread of the virus. Letting \( z \) be the population of CTLs, system (1) is changed to

\[ \begin{align*}
\dot{x} &= \lambda - dx - \beta xv \\
\dot{y} &= \beta xv - ay - pyz \\
\dot{z} &= cyz - bz \\
\dot{v} &= ky - uv.
\end{align*} \]  

(4)

where \( c \) is the proliferation rate of CTLs in response to antigenic stimulation, \( b \) is the death rate of CTLs in the absence of antigenic stimulation, and \( p \) is the rate at which CTLs kill infected cells. The parameter \( c \) is known as CTL responsiveness. Making the same assumption as above that the free virus population grows at much faster rates than the other populations, we again obtain \( v = ky/u \). Then Eq. (4) leads to

\[ \begin{align*}
\dot{x} &= \lambda - dx - Bxy \\
\dot{y} &= Bxy - ay - pyz \\
\dot{z} &= cyz - bz.
\end{align*} \]  

(5)

In this case, if the same condition \( R_0 > 1 \) is obeyed, and the CTL responsiveness is large enough, i.e. if \( c(\lambda/a - d/B) > b \), a sustained CTL response develops, and the system converges to the equilibrium state:

\[ \begin{align*}
x^* &= \frac{\lambda c}{dc + Bb}, \\
y^* &= \frac{b}{c}, \\
z^* &= \frac{c(\lambda B - da) - abB}{p(dc + Bb)}.
\end{align*} \]  

(6)

3. Non-intrusive polynomial chaos

In the general case, one may be interested in a quantity of interest which is usually a functional of the solution of a model that describes certain phenomena under investigation. Let this quantity be denoted by \( X \), and assume that the model \( X = M(\Lambda) \) used to compute it depends on parameters \( \Lambda \), both \( X \) and \( \Lambda \) considered to be finite dimensional. Such a model can take a number of forms, from simple algebraic relationships to complicated partial differential equations. Assume that the parameter values are random variables, i.e. that they depend on the outcome \( \omega \in \Omega \) of an experiment, with the set of all possible outcomes \( \Omega \) properly equipped with a \( \sigma \)-algebra \( \mathcal{F} \) and a probability measure \( P \) such that the triple \( (\Omega, \mathcal{F}, P) \) forms a probability space [11]. Thus, the quantity of interest also becomes a random variable \( X = M(\Lambda(\omega)) \). If
the parameter distributions are known, one way to obtain the distribution of the output \( X \) is to use the Monte Carlo (MC) method [7]. This entails generating a number of samples from the parameter distributions, running the model to obtain the value of the output for each sample, then constructing the histogram or computing the desired moments (expected value, variance) of the output. In its basic form, the MC method is extremely attractive to use, since it only needs repeated runs of the model in order to obtain the desired information. However, as is well known, the convergence of the method is proportional to the square root of the number of samples, which usually translates into excessively large numbers of runs of the model to obtain a certain accuracy. This becomes prohibitive when the cost for running the model is large.

Another approach for approximating the solution of random differential equations is the generalized polynomial chaos (GPC) approach [12–14]. In this context, the parameters, which are random quantities, and the solutions of the differential equations, which are stochastic processes, are projected on the space of polynomial chaos. These chaos expansions are performed in terms of basis functions chosen according to the probability density functions of the random variables and stochastic processes (more details are given below). The usual next step is to use a Galerkin projection taking advantage of the orthogonality of the basis functions, and then to truncate the polynomial chaos series to a finite number of terms to obtain a system of ordinary differential equations of the same type as the original one. However, to solve this system the original code used for the deterministic equations cannot be used, and new codes need to be developed. To overcome this difficulty, some non-intrusive polynomial chaos methods have been developed, which can use the deterministic codes without modification.

The sampling NIPC method is a reasonable alternative to the MC method. Like the latter, it only requires running the model (together with solving a relatively small-size linear system of equations) to obtain a probabilistic description of the output. The number of model runs may, however, be much smaller than that required for the MC method for the same level of accuracy. The NIPC method may thus be potentially more efficient than the MC method. It makes use of the Cameron–Martin theorem [15] that guarantees convergence of the chaos expansion

\[
X(\omega) = \tilde{X}_0 \Gamma_0 + \sum_{i_1=1}^{\infty} \hat{\tilde{x}}_{i_1} \Gamma'_1(\tilde{\xi}_{i_1}(\omega)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{l_1} \hat{\tilde{x}}_{i_1 i_2} \Gamma_2(\tilde{\xi}_{i_1}(\omega), \tilde{\xi}_{i_2}(\omega)) + \cdots
\]

(7)

for second-order functionals. The successive chaos \( \Gamma_i \) that appear in this expansion are polynomials in the random variables \( \tilde{\xi} \), which are all of the same type. This type, together with the corresponding polynomials, can be chosen depending on the available knowledge about the random variables for which the expansion is performed [14]. The chaos can also be constructed as tensor products of univariate monomials of order \( p = 0, 1, \ldots \) in each of the chaos variables. For computational purposes, this chaos expansion is truncated after a finite number of terms and rearranged as

\[
X(\omega) = \sum_{i=0}^{P} X_i \Phi_i(\tilde{\xi}),
\]

(8)

where \( \tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots) \) and \( \Phi_0(\tilde{\xi}) = 1 \). The same expansion can be performed for the parameters of the model. Thus, one value for the random parameter of the model \( \Lambda \) is mapped into one value for the quantity of interest \( X \) as \( X(\omega) = M(\Lambda(\tilde{\xi}(\omega))) \), where \( \Lambda(\tilde{\xi}(\omega)) = \sum_{i=0}^{P} \Lambda_i \Phi_i(\tilde{\xi}) \) and the coefficients \( \Lambda_i \) are known.

To obtain the NIPC approximation, a sample for the parameter set is first obtained by first sampling the vector \( \tilde{\xi} \) and then obtaining \( \Lambda \) from its known expansion. Let such a sample be denoted by \( \tilde{\xi}_i = \tilde{\xi}(\omega_i) \), where \( i = 1, 2, \ldots, S \) and \( S \) is the total number of samples. Then one equation is obtained for the coefficients of the output variable by running the model and requesting that \( X(\omega) = M(\Lambda(\tilde{\xi})) \). This leads to the following system of equations for the coefficients of the output variable:

\[
\begin{bmatrix}
1 & \Phi_1(\tilde{\xi}_1) & \cdots & \Phi_p(\tilde{\xi}_1) \\
1 & \Phi_1(\tilde{\xi}_2) & \cdots & \Phi_p(\tilde{\xi}_2) \\
\vdots & \vdots & \ddots & \vdots \\
1 & \Phi_1(\tilde{\xi}_S) & \cdots & \Phi_p(\tilde{\xi}_S)
\end{bmatrix}
\begin{bmatrix}
X_0 \\
X_1 \\
\vdots \\
X_p
\end{bmatrix}
= \begin{bmatrix}
M(\Lambda(\tilde{\xi}_1)) \\
M(\Lambda(\tilde{\xi}_2)) \\
\vdots \\
M(\Lambda(\tilde{\xi}_S))
\end{bmatrix}
\]

(9)

Note that the number of equations \( S \) is not necessarily related to the number of coefficients \( P + 1 \). In actual simulations, we usually construct an overdetermined system and solve it using least squares.

4. Numerical results

The first set of numerical experiments is performed on model (1). A base case with deterministic values of the parameters \( \lambda = 1, d = 0.1, \beta = 0.1, a = 0.2, k = 1, \) and \( u = 0.3 \) is considered. A plot of the evolution of the populations versus time in this case is shown in Fig. 1.

Randomness is then introduced by asking that some or all of the coefficients be uniformly distributed random variables with mean their deterministic value and spread over an interval of length 0.4 times their mean value. A first case we
Fig. 1. System time evolution for the first model in the deterministic case.

Fig. 2. Errors in higher-order statistics for $y^*$ for different monomial orders $p$ in the tensor product and a sample size $S = 27$. Three-dimensional chaos.

Table 1
Expected values (left) and variances (right) for $x^* = au/\beta k$ for various sample sizes and polynomial orders with two-dimensional NIPC for $\beta$ and $k$.

<table>
<thead>
<tr>
<th>$S$</th>
<th>$P = 2$</th>
<th>$P = 5$</th>
<th>$P = 9$</th>
<th>$P = 2$</th>
<th>$P = 5$</th>
<th>$P = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.62476</td>
<td>0.35746</td>
<td>-</td>
<td>0.00663</td>
<td>0.02336</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>0.62476</td>
<td>0.61776</td>
<td>0.56800</td>
<td>0.00993</td>
<td>0.01067</td>
<td>0.01523</td>
</tr>
<tr>
<td>16</td>
<td>0.61909</td>
<td>0.61769</td>
<td>0.61652</td>
<td>0.01098</td>
<td>0.01038</td>
<td>0.01057</td>
</tr>
<tr>
<td>25</td>
<td>0.62153</td>
<td>0.61675</td>
<td>0.61652</td>
<td>0.00894</td>
<td>0.01059</td>
<td>0.01053</td>
</tr>
</tbody>
</table>

examined was $\beta \sim U(0.08, 0.12), k \sim U(0.8, 1.2)$, while the other parameters have their deterministic value. The expected values and variances for the equilibrium states can be determined analytically in this case using the algebraic relationships in Eq. (2). The same relationships can be used for the simulation, but results reported here have been obtained by integrating the model in time with an explicit method up to non-dimensional time $t = 60$, when all transients have died out. Table 1 shows these statistics for $x^*$ for several polynomial orders and number of samples, obtained by averaging the results over ten realizations. The analytical values are $E[x^*] = 0.61650732$ and $\text{Var}[x^*] = 0.010544$, while the Monte Carlo results for 1000 samples are $E[x^*] = 0.616571$ and $\text{Var}[x^*] = 0.01047$. The severely undetermined case $S = 4, P = 9$ is not shown. Good accuracy can be noticed in the results for the higher polynomial orders as soon as $S > P + 1$.

For the three-dimensional chaos case, we also allow $a \sim U(0.16, 0.24)$. One of the advantages of NIPC versus usual polynomial chaos using a Galerkin projection is that, in the latter, higher-dimensional chaos may lead to lengthy inner products that are difficult to evaluate, whereas in NIPC the chaos dimension makes no difference other than solving a bigger matrix; the setup process of this matrix is the same. Fig. 2 displays the errors in the statistics for the equilibrium value $y^*$ at different polynomial orders for a sample size $S = 27$.

Next we consider the CTL model (5) with the same parameter values as before; hence $B = 1/3$, and $p = 0.5, c = b = 0.1$, which leads to a CTL response. The time evolution of the deterministic system is shown in Fig. 3. We focus on the ability of NIPC to obtain accurate statistics for larger chaos dimensions. To this purpose, we let all the parameters in the model be...
random, and compute the error in the expected value and variance for $z^*$, for which a seven-dimensional chaos is necessary. The multivariate polynomials in the chaos basis are constructed via tensor products of monomials with order up to $p = 3$. The errors for a fixed sample size $S = 128$ are shown in Fig. 4; they show the expected faster-than-algebraic convergence trend.

5. Conclusions

Parameters in virus models have a large variability; even for the same type of virus the parameters involved have variations, in particular the number of viral particles produced from a single infected cell. The NIPC method is an useful tool for studying the effect this variability has on the outcome of the mathematical models. Like the MC technique, it does not require the development of any special code except the one for the deterministic model itself, but it requires less sampling, and consequently much less computer time than the latter, for comparable accuracy. Another advantage of this method is that it produces a series expansion of the output process. This is particularly valuable if the output process becomes in turn an input for some other model: realizations of this output process are then readily at hand.

Since the NIPC method uses the same deterministic code for solving the model equations, it is easily used in many problems in which randomness needs to be introduced. But, as with all problems that require randomness, it may be hard or even impossible to determine the density probability functions of the random parameters involved, and estimates of the variability may have to be used.

References