Spring 5-5-2020

The T, T*, VI, VNI Model for Human Immunodeficiency Virus Type 1 (HIV-1) Dynamics

Amy Creel

The T, T^*, V_I, V_{NI} Model for Human Immunodeficiency Virus Type 1 (HIV-1) Dynamics

Amy Creel

submitted in partial fulfillment of the requirements for Honors in Mathematics at the University of Mary Washington

Fredericksburg, Virginia

April 2020

This thesis by **Amy R. Creel** is accepted in its present form as satisfying the thesis requirement for Honors in Mathematics.

Date	Approved
_____	_____
	Jangwoon (Leo) Lee, Ph.D. thesis advisor
_____	_____
	Julius N. Esunge, Ph.D. committee member
_____	_____
	Suzanne Sumner, Ph.D. committee member

Contents

Abstract

In this research project, I investigated deterministic and stochastic versions of the T, T^*, V_I, V_{NI} model for Human Immunodeficiency Virus Type 1 (HIV-1) dynamics. First, an analytical solution to a simplified version of the deterministic model is found. Then, numerical techniques are used to obtain an approximate solution to the deterministic model. Finally, a stochastic version of the model is discussed, and numerical methods are used to find an approximate solution to the stochastic system. These results demonstrate the behavior of HIV-1 in an infected patient under the effects of reverse transcriptase and protease inhibitors, and illustrate how the addition of randomness to the constants in the model impact the development of HIV-1 in a given patient.

1 Model for HIV-1 Dynamics

The major target of HIV infection is a class of lymphocytes known as CD4+ T-cells. Once infected, these cells produce new HIV virus particles. These virus particles are infectious, but they can be made noninfectious by the actions of protease inhibitors. The model examined in this project is one that has been a basis for many studies of HIV-1 Dynamics [3] and considers four populations: uninfected CD4+ T-cells that are susceptible to infection (denoted T), productively infected CD4+ T-cells (denoted T^*), infectious virus particles (denoted V_I), and noninfectious virus particles (denoted V_{NI}).

The following equations represent the rates of change of these populations under the effects of reverse transcriptase and protease inhibitors:

$$\frac{dT}{dt} = \lambda - dT - (1-\kappa)kV_I T \tag{1}$$

$$\frac{dT^*}{dt} = (1-\kappa)kV_I T - \delta T^* \tag{2}$$

$$\frac{dV_I}{dt} = (1-\eta)N_T \delta T^* - cV_I \tag{3}$$

$$\frac{dV_{NI}}{dt} = \eta N_T \delta T^* - cV_{NI} \tag{4}$$

Each of the constants in this model (λ, d, k, δ, N_T, c, κ, and η) are positive values. The following table contains the values used for the parameters of this model obtained from [11] along with the units of each constant and its interpretation in the model.

Parameter	Value	Units	Interpretation
λ	0.1089	cells per day	Healthy T-cell birth rate.
d	0.01089	1 / day	Healthy T-cell death rate.
k	1.179×10^{-3}	1 / (virions $\cdot$ day)	Infection rate of susceptible T-cells.
δ	0.366	1 / day	Infected T-cell death rate.
N_T	4246.4	virions / cell	Virus production rate.
c	3.074	1 / day	Viral clearance rate.

Table 1: Values of model parameters and interpretations.

Note that the parameters κ and η in our model are not included in the table above. The parameters κ and η represent the efficacy of reverse transcriptase inhibitors and protease inhibitors,

respectively. Reverse transcriptase (RT) inhibitors are a treatment method which prevent HIV RNA from being converted to DNA. As a result, the infectiousness of the virus is reduced by $(1-\kappa)$, where $0 \leq \kappa \leq 1$ [3]. Protease inhibitors, on the other hand, do not have a direct impact on the infectiousness of the virus. Rather, they cause the production of noninfectious virus particles. The value η is then defined to be the fraction of the total virus produced that is noninfectious [3], with $(1-\eta)$ being the fraction of the total virus produced that is infectious. Therefore, production of infectious virus is multiplied by the constant $(1-\eta)$, while production of noninfectious virus is multiplied by η, where $0 \leq \eta \leq 1$.

Appropriate values for κ and η are addressed in Chapter 3 within the discussion of patient data. From the patient data, it seems that the most appropriate estimates for κ and η are 0.6, and therefore $\kappa = 0.6$ and $\eta = 0.6$ are used in solving the deterministic model in Chapter 2.

1.1 Existence and Uniqueness

To show that a unique solution to our initial value problem exists, we make use of the Picard-Lindelöf Theorem. A proof of this theorem can be found in [5].

Theorem 1. *(Picard Lindelöf Theorem) Consider the ordinary differential equation*

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^m,$$

with initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$. *Let* $U = \bar{B}(\mathbf{x}_0, b)$ *and* $J = [t_0 - a, t_0 + a]$, *where* $f : U \to \mathbb{R}^m$ *is Lipschitz with Lipschitz constant* K, *and* $|f(\mathbf{x})| \leq M$ *for all* $\mathbf{x} \in U$. *Then the initial value problem has a unique solution* $x \in C^0(J, U)$ *as long as the time interval is chosen with a satisfying* $0 < a < \min\left\{\frac{1}{K}, \frac{b}{M}\right\}$.

To apply the Picard-Lindelöf Theorem to our system, let $\mathbf{x}$ and$f(\mathbf{x})$ be defined by

$$\mathbf{x} = \begin{bmatrix} T \\ T^* \\ V_I \\ V_{NI} \end{bmatrix} \quad \text{and} \quad f(\mathbf{x}) = \begin{bmatrix} \lambda - dT - (1-\kappa)kV_I T \\ (1-\kappa)kV_I T - \delta T^* \\ (1-\eta)N_T \delta T^* - cV_I \\ \eta N_T \delta T^* - cV_{NI} \end{bmatrix}$$

The Jacobian matrix of our system is then given by:

$$\begin{bmatrix} -d - (1-\kappa)kV_I & 0 & -(1-\kappa)kT & 0 \\ (1-\kappa)kV_I & -\delta & (1-\kappa)kT & 0 \\ 0 & (1-\eta)N_T\delta & -c & 0 \\ 0 & \eta N_T \delta & 0 & -c \end{bmatrix}$$

Note that the partial derivatives of f exist and are continuous, which implies that f is Lipschitz continuous. Therefore, there exists a unique solution to our system on some interval $[0, t^*]$.

1.2 Equilibria

A set of values $(T_0, T_0^*, V_{I0}, V_{NI0})$ is an equilibrium of our system if $\frac{dT}{dt} = \frac{dT^*}{dt} = \frac{dV_I}{dt} = \frac{dV_{NI}}{dt} = 0$ when we set $T = T_0$, $T^* = T_0^*$, $V_I = V_{I0}$, and $V_{NI} = V_{NI0}$. At the equilibrium, the values of T, T^*,

V_I, and V_{NI} do not change. To find the equilibria for the system, consider the following equations:

$$\frac{dT}{dt} = \lambda - dT_0 - (1-\kappa)kV_{I0}T_0 = 0 \tag{5}$$

$$\frac{dT^*}{dt} = (1-\kappa)kV_{I0}T_0 - \delta T_0^* = 0 \tag{6}$$

$$\frac{dV_I}{dt} = (1-\eta)N_T\delta T_0^* - cV_{I0} = 0 \tag{7}$$

$$\frac{dV_{NI}}{dt} = \eta N_T\delta T_0^* - cV_{NI0} = 0 \tag{8}$$

Solving equations (7) and (8) for T_0^* gives:

$$T_0^* = \frac{cV_{I0}}{(1-\eta)N_T\delta} \text{ and } T_0^* = \frac{cV_{NI0}}{\eta N_T\delta}.$$

Then, setting these two equations equal to each other, we have:

$$\frac{cV_{I0}}{(1-\eta)N_T\delta} = \frac{cV_{NI0}}{\eta N_T\delta} \implies \eta V_{I0} = (1-\eta)V_{NI0}.$$

From this result, we solve for V_{I0} in terms of V_{NI0} as well as V_{NI0} in terms of V_{I0}, and obtain

$$V_{I0} = \frac{(1-\eta)}{\eta}V_{NI0} \text{ and } V_{NI0} = \frac{\eta}{(1-\eta)}V_{I0}.$$

Substituting T_0^* in terms of V_{I0} into equation (6) then yields

$$(1-\kappa)kV_{I0}T_0 - \delta\left[\frac{cV_{I0}}{(1-\eta)N_T\delta}\right] = 0.$$

Solving this equation for the value of T_0, we obtain

$$T_0 = \frac{c}{(1-\kappa)(1-\eta)kN_T}.$$

Then, applying our steady state solution for T_0 to equation (5), we have

$$\lambda - d\left[\frac{c}{(1-\kappa)(1-\eta)kN_T}\right] - (1-\kappa)k\left[\frac{c}{(1-\kappa)(1-\eta)kN_T}\right]V_{I0} = 0.$$

Solving this equation for V_{I0} yields

$$V_{I0} = \frac{(1-\eta)N_T\lambda}{c} - \frac{d}{(1-\kappa)k}.$$

Then, applying this substitution to our equation for V_{NI0} in terms of V_{I0}, we have:

$$V_{NI0} = \frac{\eta}{1-\eta}\left[\frac{(1-\eta)N_T\lambda}{c} - \frac{d}{(1-\kappa)k}\right] = \frac{\eta N_T\lambda}{c} - \frac{d\eta}{(1-\eta)(1-\kappa)k}.$$

Finally, applying this substitution for V_{NI0} to our equation for T_0^* in terms of V_{NI0}, we obtain

$$T_0^* = \frac{c}{\eta N_T\delta}\left[\frac{\eta N_T\lambda}{c} - \frac{d\eta}{(1-\eta)(1-\kappa)k}\right] = \frac{\lambda}{\delta} - \frac{cd}{(1-\eta)(1-\kappa)N_T\delta k}.$$

Therefore, our equilibrium is given by the following values:

$$T_0 = \frac{c}{(1-\kappa)(1-\eta)kN_T}, \qquad T_0^* = \frac{\lambda}{\delta} - \frac{cd}{(1-\eta)(1-\kappa)N_T\delta k},$$

$$V_{I0} = \frac{(1-\eta)N_T\lambda}{c} - \frac{d}{(1-\kappa)k}, \text{ and } V_{NI0} = \frac{\eta N_T\lambda}{c} - \frac{d\eta}{(1-\eta)(1-\kappa)k}.$$

2 Solutions to the Deterministic Model for HIV-1 Dynamics

2.1 Introduction to the Simplified Model

One of the goals of this project was to find the analytical solution to the T, T^*, V_I, V_{NI} model. Since it is difficult to compute an analytical solution to the full system, we use a simplified version of the model which includes terms involving cell production rates, death rates of infected cells, virus production, and drug efficacy of both protease and reverse transcriptase inhibitors. This simplified model is given by the following system of equations:

$$\frac{dT}{dt} = -(1-\kappa)kV_I T \tag{9}$$

$$\frac{dT^*}{dt} = (1-\kappa)kV_I T - \delta T^* \tag{10}$$

$$\frac{dV_I}{dt} = (1-\eta)N_T \delta T^* \tag{11}$$

$$\frac{dV_{NI}}{dt} = \eta N_T \delta T^* \tag{12}$$

Note that the removal of λ, the constant representing production of susceptible cells, results in no positive terms on the right hand side of equation (9), and therefore the number of susceptible CD4+ T-cells is always decreasing. The term $(1-\kappa)kV_I T$ represents the rate of infection of susceptible cells under the effects of protease inhibitors. The rate of infection is proportional to V_I and T, with proportionality constants k and $(1-\kappa)$, which are the infection rate constant and the efficacy of the protease inhibitors, respectively.

In equation (10), the term δT^* represents the death rate of infected cells. This rate is proportional to the population of infected cells with proportionality constant δ.

Removal of the terms including the viral clearance rate c in equations (11) and (12) mean that the right hand side of these equations will always be positive. Therefore, the amount of virus present in the body will always be increasing. As mentioned in Chapter 1, the term $N_T \delta T^*$ represents the rate of production of virus particles. This production rate is proportional to T^*, with proportionality constant $N_T \delta$. To account for the efficacy of RT inhibitors (which determines the proportion of virus produced as noninfectious), this rate of production is multiplied by the constant $(1-\eta)$ in equation (11) and η in equation (12).

2.2 Analytical Solution to the Simplified Model

In this section, we will find an analytical solution to the simplified model given by:

$$\frac{dT}{dt} = -(1-\kappa)kV_I T \tag{13}$$

$$\frac{dT*}{dt} = (1-\kappa)kV_I T - \delta T^* \tag{14}$$

$$\frac{dV_I}{dt} = (1-\eta)N_T \delta T^* \tag{15}$$

$$\frac{dV_{NI}}{dt} = \eta N_T \delta T^* \tag{16}$$

The initial conditions of this model are:

$$T(0) = T_0, \quad T^*(0) = T_0^*, \quad V_I(0) = V_{I0}, \quad \text{and} \quad V_{NI}(0) = V_{NI0}.$$

Note that

$$\frac{d}{dt}\left[T + T^* + \frac{1}{N_T}\left(V_I + V_{NI}\right)\right] = 0 \quad \text{and} \quad \frac{d}{dt}\left[V_I + \left(\frac{\eta - 1}{\eta}\right) V_{NI}\right] = 0.$$

Therefore,

$$T + T^* + \frac{1}{N_T}(V_I + V_{NI}) = P \quad \text{and} \quad V_I + \left(\frac{\eta - 1}{\eta}\right) V_{NI} = P_V,$$

where P and P_V are constants, with P representing total population of white blood cells and virus in the body and P_V representing the total viral population in the body. Then from these equations,

$$T = P - T^* - \frac{1}{N_T}(V_I + V_{NI}) \quad \text{and} \quad T^* = P - T - \frac{1}{N_T}(V_I + V_{NI}),$$
$$V_I = P_V - \left(\frac{\eta - 1}{\eta}\right) V_{NI} \quad \text{and} \quad V_{NI} = \left(\frac{\eta}{\eta - 1}\right)\left(P_V - V_I\right).$$

Additionally, with the initial conditions defined above, we have:

$$P = T_0 + T_0^* + \frac{1}{N_T}(V_{I0} + V_{NI0}) \quad \text{and} \quad P_V = V_{I0} + \left(\frac{\eta - 1}{\eta}\right) V_{NI0}.$$

To obtain an analytical solution for this simplified model, we will first use an iterative technique to obtain an implicit solution for V_I, and use this solution to solve for V_{NI}, T, and T^* in the simplified model.

We begin by differentiating $\dfrac{dV_I}{dt}$ and applying substitutions for T and V_{NI} to obtain the second derivative of V_I with respect to time. For the sake of simplicity, in the calculations below, let $\alpha = (1 - \eta)\delta(1 - \kappa)k$ and $\omega = \alpha N_T = (1 - \eta)N_T\delta(1 - \kappa)k$.

$$\begin{aligned}
\frac{d^2V_I}{dt^2} &= (1 - \eta)N_T\delta\big[(1 - \kappa)kV_IT - \delta T^*\big] \\
&= (1 - \eta)N_T\delta(1 - \kappa)kV_IT - \delta\big[(1 - \eta)N_T\delta T^*\big] \\
&= (1 - \eta)N_T\delta(1 - \kappa)kV_IT - \delta\frac{dV_I}{dt} \\
&= (1 - \eta)N_T\delta(1 - \kappa)kV_I\big[P - T^* - \frac{1}{N_T}\big(V_I + V_{NI}\big)\big] - \delta\frac{dV_I}{dt} \\
&= \omega V_I\left[P - \frac{1}{N_T}\big(V_I + V_{NI}\big)\right] - (1 - \eta)N_T\delta(1 - \kappa)kV_IT^* - \delta\frac{dV_I}{dt} \\
&= \omega V_I\left[P - \frac{1}{N_T}\big(V_I + V_{NI}\big)\right] - (1 - \kappa)kV_I\frac{dV_I}{dt} - \delta\frac{dV_I}{dt} \\
&= \omega PV_I - \alpha V_I^2 - \alpha V_{NI}V_I - (1 - \kappa)kV_I\frac{dV_I}{dt} - \delta\frac{dV_I}{dt} \\
&= \omega PV_I - \alpha V_I^2 - \alpha V_I\left(\frac{\eta}{\eta - 1}\right)(P_V - V_I) - (1 - \kappa)kV_I\frac{dV_I}{dt} - \delta\frac{dV_I}{dt} \\
&= \omega PV_I - \alpha V_I^2 - \alpha\left(\frac{\eta}{\eta - 1}\right)P_VV_I + \alpha\left(\frac{\eta}{\eta - 1}\right)V_I^2 - (1 - \kappa)kV_I\frac{dV_I}{dt} - \delta\frac{dV_I}{dt}
\end{aligned}$$

Then the second derivative of V_I with respect to time is:

$$\begin{aligned} V_I'' &= \omega P V_I - \alpha V_I^2 - \alpha\left(\frac{\eta}{\eta-1}\right)P_V V_I + \alpha\left(\frac{\eta}{\eta-1}\right)V_I^2 - (1-\kappa)kV_I V_I' - \delta V_I' \\ &= \left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I - (1-\kappa)kV_I V_I' - \delta V_I' \end{aligned}$$

Let $u = \dfrac{dt}{dV_I} = \dfrac{1}{V_I'}$. Then:

$$\begin{aligned} \frac{du}{dV_I} &= \left[\frac{du}{dt}\right]\left[\frac{dt}{dV_I}\right] = [-(V_I')^{-2}V_I'](V_I')^{-1} = -(V_I')^{-3}(V_I'') \\ &= -\left[\frac{1}{V_I'}\right]^3\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I - (1-\kappa)kV_I V_I' - \delta V_I'\right] \\ &= \left[-\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 - \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]u^3 + \left[(1-\kappa)kV_I + \delta\right]u^2 \end{aligned}$$

Note that the initial condition for u is

$$u(V_{I0}) = \frac{1}{(1-\eta)N_T\delta T_0^*}$$

Let $\phi = \ln(u)$. Then $u = e^{\phi}$. Then:

$$\begin{aligned} \frac{d\phi}{dV_I} &= \left[\frac{d\phi}{du}\right]\left[\frac{du}{dV_I}\right] = \frac{1}{e^{\phi}}\left[\frac{du}{dV_I}\right] \\ &= \frac{1}{e^{\phi}}\left[\left[(1-\kappa)kV_I + \delta\right]u^2 - \left(\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right)u^3\right] \\ &= \frac{1}{e^{\phi}}\left[\left[(1-\kappa)kV_I + \delta\right]e^{2\phi} - \left(\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right)e^{3\phi}\right] \\ &= \left[(1-\kappa)kV_I + \delta\right]e^{\phi} - \left[\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]e^{2\phi} \end{aligned}$$

The initial condition for ϕ is then given by

$$\phi(V_{I0}) = -\ln\left((1-\eta)N_T\delta T_0^*\right).$$

We can use the Taylor series expansion of e^{ϕ} to rewrite this equation as

$$\begin{aligned} \frac{d\phi}{dV_I} &= \left[(1-\kappa)kV_I + \delta\right]\left[\sum_{n=0}^{\infty}\frac{\phi^n}{n!}\right] - \left[\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\left[\sum_{n=0}^{\infty}\frac{(2\phi)^n}{n!}\right] \\ &= \left[(1-\kappa)kV_I + \delta\right]\left[1 + \phi + \frac{\phi^2}{2} + \cdots\right] \\ &\quad - \left[\left[\alpha\left(\frac{\eta}{\eta-1}\right) - \alpha\right]V_I^2 + \left[\omega P - \alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\left[1 + 2\phi + \frac{(2\phi)^n}{2} + \cdots\right] \end{aligned}$$

For a first approximation, namely $\phi_1(V_I)$, use the first two terms of the Taylor series:

$$\begin{aligned}\frac{d\phi_1}{dV_I} &= \big[(1-\kappa)kV_I+\delta\big]\Big[1+\phi_1\Big]-\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\Big[1+2\phi_1\Big]\\ &= \left((1-\kappa)kV_I+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right)\\ &\quad+\left((1-\kappa)kV_I+\delta-2\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\right)\phi_1\end{aligned}$$

This result is a first order linear differential equation of the form $\dfrac{d\phi_1}{dV_I}=Q(V_I)+R(V_I)\phi_1$, where

$$Q(V_I)=(1-\kappa)kV_I+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I,$$

$$R(V_I)=(1-\kappa)kV_I+\delta-2\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right].$$

We can solve this differential equation using an integrating factor, $F(V_I)=e^{-\int R(V_I)dV_I}$. The integrating factor that will be used is given by:

$$F(V_I)=e^{\frac{2}{3}\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^3+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I^2-\frac{1}{2}(1-\kappa)kV_I^2-\delta V_I}$$

Applying this integrating factor to solve the differential equation above, we have:

$$\int_{V_{I0}}^{V_I}\frac{d}{d\xi}\big[F(\xi)\phi_1(\xi)\big]d\xi=\int_{V_{I0}}^{V_I}F(\xi)\left[(1-\kappa)k\xi+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]d\xi$$

We then obtain the solution to the ϕ_1, the first approximation for ϕ:

$$\phi_1(V_I)=\frac{1}{F(V_I)}\left[F(V_{I0})\phi_1(V_{I0})+\int_{V_{I0}}^{V_I}F(\xi)\left[(1-\kappa)k\xi+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]d\xi\right]$$

Note that $\dfrac{d\phi}{dV_I}$ can be rewritten in the following manner:

$$\begin{aligned}\frac{d\phi}{dV_I} &= \Big[(1-\kappa)kV_I+\delta\Big]\left[1+\phi+\sum_{n=2}^{\infty}\frac{\phi^n}{n!}\right]\\ &\quad-\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\left[1+2\phi+\sum_{n=2}^{\infty}\frac{(2\phi)^n}{n!}\right]\end{aligned}$$

We can then obtain a second approximation to ϕ, ϕ_2, by plugging ϕ_1 (our first approximation) into the infinite sum expression above:

$$\begin{aligned}\frac{d\phi_2}{dV_I} &= \left[(1-\kappa)kV_I+\delta\right]\left[1+\phi_2+\sum_{n=2}^{\infty}\frac{(\phi_1)^n}{n!}\right] \\ &\quad - \left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\left[1+2\phi_2+\sum_{n=2}^{\infty}\frac{(2\phi_1)^n}{n!}\right] \\ &= \left[(1-\kappa)kV_I+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right] \\ &\quad + \left[(1-\kappa)kV_I+\delta-2\left(\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right)\right]\phi_2 \\ &\quad + \sum_{n=2}^{\infty}\left[(1-\kappa)kV_I+\delta\right]\left[\frac{(\phi_1)^n}{n!}\right]-\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]V_I^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]V_I\right]\left[\frac{(2\phi_1)^n}{n!}\right]\end{aligned}$$

This result is another first order differential equation, which can again be solved using an integrating factor. Note that the coefficient for ϕ_2 in this equation is the same as the coefficient for ϕ_1 in the differential equation solved earlier, and therefore we will use the same integrating factor, $F(V_I)$. Using this integrating factor, we obtain the following solution for $\phi_2(V_I)$:

$$\begin{aligned}\phi_2(V_I) &= \frac{1}{F(V_I)}\left[F(V_{I0})\phi_2(V_{I0})\right] \\ &\quad + \frac{1}{F(V_I)}\left[\int_{V_{I0}}^{V_I}F(\xi)\left[(1-\kappa)k\xi+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]d\xi\right] \\ &+\frac{1}{F(V_I)}\int_{V_{I0}}^{V_I}F(\xi)\sum_{n=2}^{\infty}\left[(1-\kappa)k\xi+\delta\right]\left[\frac{(\phi_1)^n}{n!}\right]-\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]\left[\frac{(2\phi_1)^n}{n!}\right]d\xi\end{aligned}$$

We can continue this process to get successively more accurate approximations to ϕ. In general, for any integer $m>1$, we have:

$$\begin{aligned}\phi_{m+1}(V_I) &= \frac{1}{F(V_I)}\left[F(V_{I0})\phi_{m+1}(V_{I0})\right] \\ &\quad + \frac{1}{F(V_I)}\left[\int_{V_{I0}}^{V_I}F(\xi)\left[(1-\kappa)k\xi+\delta-\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2-\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]d\xi\right] \\ &+\frac{1}{F(V_I)}\int_{V_{I0}}^{V_I}F(\xi)\sum_{n=2}^{\infty}\left[(1-\kappa)k\xi+\delta\right]\left[\frac{(\phi_m)^n}{n!}\right]-\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]\left[\frac{(2\phi_m)^n}{n!}\right]d\xi\end{aligned}$$

This term can be written more compactly as

$$\begin{aligned}\phi_{m+1}(V_I) &= \phi_1(V_I)+\frac{1}{F(V_I)}\int_{V_{I0}}^{V_I}F(\xi)\left[(1-\kappa)k\xi+\delta\right]\left[e^{\phi_m(\xi)}-1-\phi_m(\xi)\right]d\xi \\ &\quad - \frac{1}{F(V_I)}\int_{V_{I0}}^{V_I}F(\xi)\left[\left[\alpha\left(\frac{\eta}{\eta-1}\right)-\alpha\right]\xi^2+\left[\omega P-\alpha\left(\frac{\eta}{\eta-1}\right)P_V\right]\xi\right]\left[e^{2\phi_m(\xi)}-1-2\phi_m(\xi)\right]d\xi\end{aligned}$$

Each successive approximation of $\phi(V_I)$ becomes closer and closer to the actual solution of $\phi(V_I)$. In other words,

$$\lim_{m\to\infty} \phi_m(V_I) = \phi(V_I).$$

With this limit, we have obtained a solution to $\phi(V_I)$, and we can now work backwards to obtain the solution of V_I. Recall that

$$\frac{dt}{dV_I} = e^{\phi(V_I)}.$$

Separating variables and integrating both sides, we have

$$\int_{t_0}^{t} d\xi = \int_{V_{I0}}^{V_I} e^{\phi(\xi)} d\xi.$$

Then, taking $t_0 = 0$, we have $V_I(t)$ given implicitly by:

$$t = \int_{V_{I0}}^{V_I} e^{\phi(\xi)} d\xi.$$

Now that we have obtained an implicit solution for V_I, we can use this solution to find solutions for T, T^*, and V_{NI}. First, the solution for V_{NI} can be found using the following equation:

$$V_{NI} = \left(\frac{\eta}{\eta - 1}\right)(P_V - V_I).$$

We will now obtain a solution for T. Recall that

$$\frac{dT}{dt} = -(1-\kappa)kV_I T.$$

Separating variables and integrating,

$$\int_{T_0}^{T} \frac{1}{\xi} d\xi = -(1-\kappa)k \int_{t_0}^{t} V_I(\xi) d\xi.$$

Taking $t_0 = 0$,

$$T = T_0 e^{-(1-\kappa)k \int_0^t V_I(\xi) d\xi}.$$

Finally, we can obtain a solution for T^*. Recall that

$$T^* = P - T - \frac{1}{N_T}\Big(V_I + V_{NI}\Big)$$

Therefore, applying the solutions for T^*, V_I, and V_{NI} obtained above, we have:

$$T^* = P - T_0 e^{-(1-\kappa)k \int_0^t V_I(\xi) d\xi} - \frac{1}{N_T}\left[V_I + \left(\frac{\eta}{\eta - 1}\right)(P_V - V_I)\right]$$

2.3 Numerical Solution to the Simplified Model

The following figure depicts a numerical solution to this simplified model, with initial conditions $T_0 = 10$, $T_0^* = 0$, $V_{I0} = 0.1$, and $V_{NI0} = 0$, and parameter values $k = 1.179 \times 10^{-3}$, $\delta = 0.366$, $N_T = 4246.4$, $\kappa = 0.6$, and $\eta = 0.6$.

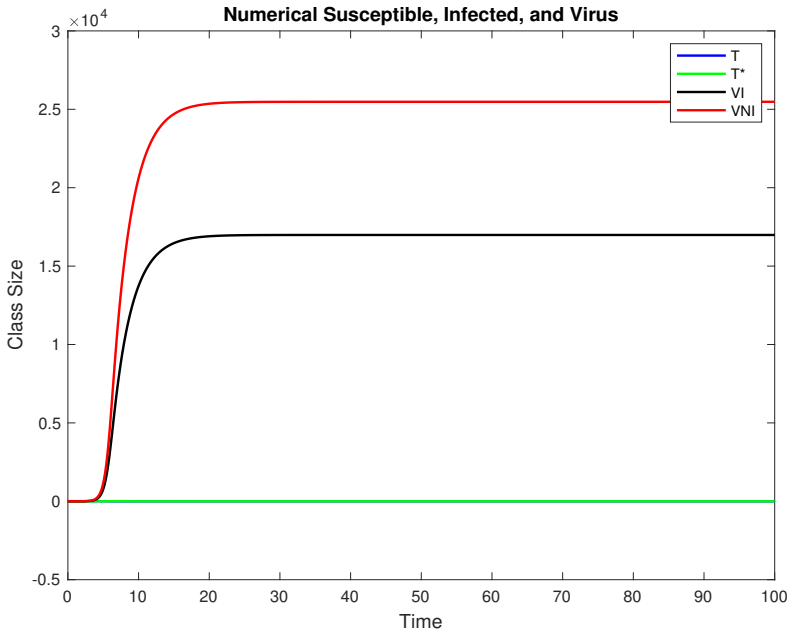


Figure 1: Approximate solution for T, T^*, V_I, and V_{NI} in the simplified model.

The differences in the class sizes of T and T^* compared to V_I and V_{NI} make it difficult to understand the behavior of each class in the above figure. Consider Figure 2 below, which shows the approximate solutions of T and T^* separate from the approximate solutions of V_I and V_{NI} to make it easier to observe the information this simplified model gives about the behaviors of each of these classes over time.

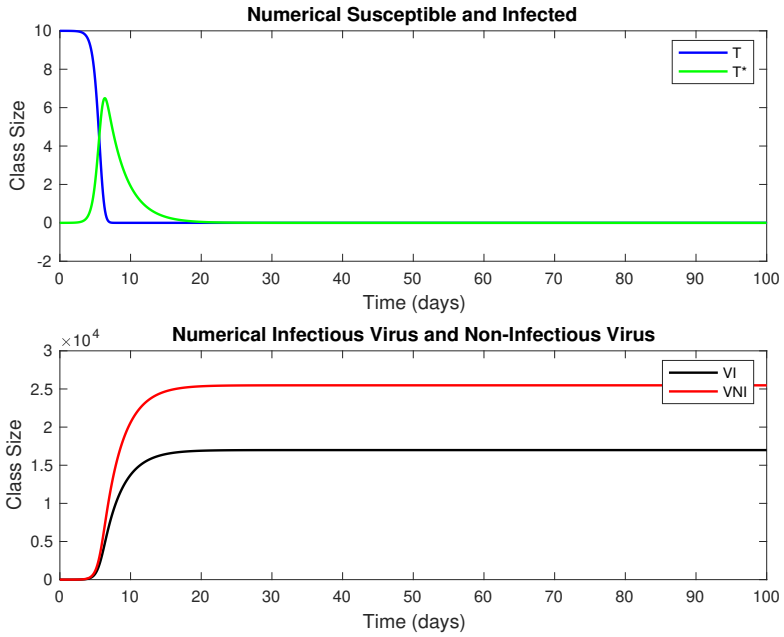


Figure 2: Approximate solution for T, T^*, V_I, and V_{NI} in the simplified model.

Note that because of the removal of the susceptible cell birth rate λ from the model, the amount of susceptible cells T is always decreasing, until it reaches its steady state. Additionally, because of the removal of the viral clearance rate, the class sizes of V_I and V_{NI} are always increasing until they reach their steady states. In the next section, an approximate solution will be found to the full deterministic model, which will produce a more biologically sound model for the study of HIV-1 Dynamics.

2.4 Numerical Solution to the Full Deterministic Model

To produce a numerical approximation to our model, a predictor-corrector method was used that combines the explicit four-step Adams-Bashforth Method and the implicit three-step Adams-Moulton Method, and uses the Runge-Kutta Method of order four to obtain its starting values.

With initial conditions $T_0 = 10$, $T_0^* = 0$, $V_{I0} = 0.1$, and $V_{NI0} = 0$, and parameter values $\lambda = 0.1089$, $d = 0.01089$, $k = 1.179 \times 10^{-3}$, $\delta = 0.366$, $N_T = 4246.4$, $c = 3.074$, $\kappa = 0.6$, and $\eta = 0.6$, the following approximation was obtained:

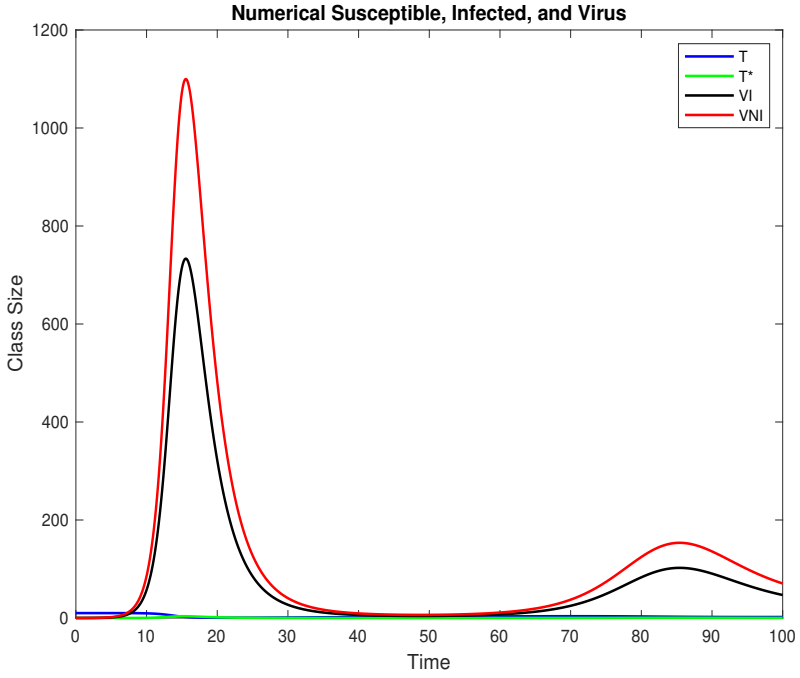


Figure 3: Approximate solution for T, T^*, V_I, and V_{NI} obtained by the Multistep Method.

In the figure above, it is difficult to see how T and T^* progress over time in the model, due to the discrepancy in the class sizes between virus particles (V_I and V_{NI}) and CD4 T-cells (T and T^*). To get a better understanding of the behavior of each class as time progresses, it helps to plot V_I and V_{NI} separately from T and T^*, as in Figure 4 below.

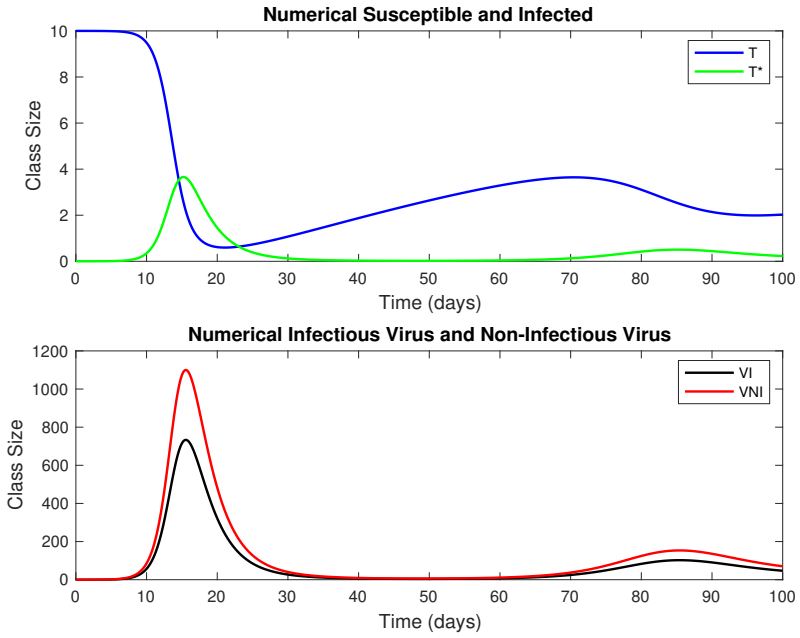


Figure 4: Top: Approximate solution for T and T^* obtained by the Multistep Method. Bottom: Approximate solution for V_I and V_{NI} obtained by the Multistep Method.

From Figures 3 and 4, we can see that around $t = 10$ days, the amount of infectious and non-infectious virus present in the body quickly increases, and as the viral concentration increases, the amount of infected CD4 T-cells increases and the amount of uninfected (susceptible) CD4 T-cells decreases. Once the virus concentration hits its peak and begins to decrease, the number of infected cells in the body also decrease, and the number of susceptible cells slowly increases. This pattern continues as time progresses, until each class reaches its steady state, as seen in Figure 5.

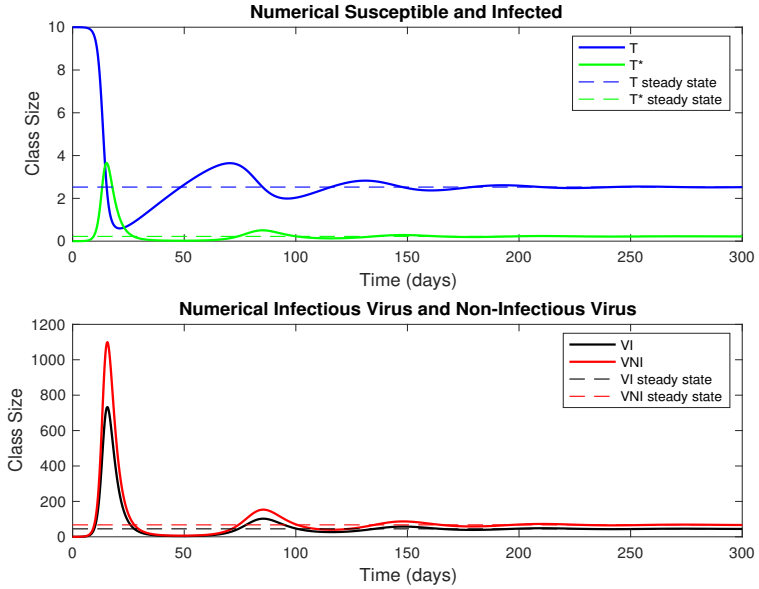


Figure 5: Approximate solutions for T, T^*, V_I, and V_{NI}, along with their steady state solutions.

3 Solutions to the Stochastic Model for HIV-1 Dynamics

3.1 Preliminary Concepts

Definition 1. *Let Ω be the set of all possible outcomes of a probabilistic experiment. A random variable X is a function from the set of outcomes Ω to the set of real numbers [12].*

Example 1. *Consider an experiment in which a fair coin is tossed twice. The outcome of each coin toss is either heads (H) or tails (T), so the set of all possible outcomes is $\Omega = \{HH, HT, TH, TT\}$. Let X be the number of heads that occur in one trial of this experiment. Then X is a random variable, and the possible values of X are 0, 1, and 2.*

Definition 2. *Let X be a random variable. Then $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$ is called the probability distribution function of X [5].*

Definition 3. *A random variable is defined to be discrete if it has a finite or countably infinite number of outcomes [12].*

Definition 4. *A continuous random variable is a variable that can take on a continuum of values rather than a finite or countably infinite number [10].*

The random variable X in Example 1 is an example of a discrete random variable. Some examples of continuous random variables include measurements of distance and speed, and passage of time.

Definition 5. *The probability mass function of a discrete random variable X, denoted pmf(x), is given by $p(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$*

Definition 6. *The probability density function of a continuous random variable X, denoted pdf(x), is the function $p(x)$ such that $F_X(x) = \int_{-\infty}^{x} p(s)ds$.*

The probability mass function and probability density functions are functions that describe the probability properties of a random variable X [10]. The domain of these functions is the possible values of the random variable X, and the range is the interval $[0, 1]$.

Example 2. *Suppose we toss a fair coin three times and observe the number of heads. Say we run this experiment 30 times, with the following results: we observe one head in 12 trials, no heads in 10 trials, and two heads in 8 trials.*

Here, pdf$(0) = Pr(X = 0)$ is the probability of observing no heads when a coin is tossed three times. Thus pdf$(0) = \frac{10}{30} = 0.333$. The values of pdf$(1)$ and pdf(2) can be calculated in the same way, and the resulting probability distribution function for the random variable X is given by:

$$pdf(x) = \begin{cases} \frac{10}{30} = 0.333 & when\ X = 0 \\ \frac{12}{30} = 0.4 & when\ X = 1 \\ \frac{8}{30} = 0.267 & when\ X = 2 \end{cases}$$

Note that the domain of the probability distribution function are the possible values of X, domain$(\text{pdf}(x)) = \{0, 1, 2\}$, and the resulting value is a number in the interval $[0, 1]$.

Definition 7. *A measure of location is a measure of the center of a set of numbers [10].*

The most commonly used measure of location for a data set is the arithmetic mean, often denoted by $\bar{x}$, which is the sum of the observed values in a data set divided by the number of observations in the set.

Example 3. *Consider the data set* $S = \{42, 36, 89, 22, 10, 56\}$. *The arithmetic mean of* S *is:*

$$\bar{x} = \frac{42 + 36 + 89 + 22 + 10 + 56}{6} = \frac{255}{6} = 42.5.$$

Another measure of location for a random variable X is the median, which is a measure of the center of all observed outcomes of X. For an odd sample size, the median is the middle value of the ordered observations. For an even sample size, the median is the average of the two values in the middle of the ordered observations [10].

Example 4. *Consider the data set* S *in Example 2, which has 6 observations. To find the median, we put these six observations in numerical order and take the average of the two observations in the middle of the set:*

$$S = \{10, 22, 36, 42, 56, 89\} \rightarrow median(S) = \frac{36 + 42}{2} = 39.$$

The median is sometimes used as a measure of the center (rather than the arithmetic mean) because it is robust, meaning it is insensitive to extremely large or small values in the data [10].

Example 5. *To understand this idea of robustness, suppose we perform ten trials of an experiment, where we toss a fair coin 400 times and record the number of heads observed. Consider two possible outcomes of this experiment,* S_1 *and* S_2 *as defined by:*

$$S_1 = \{232, 174, 259, 180, 183, 186, 400, 226, 172, 215\}$$
$$S_2 = \{232, 174, 259, 180, 183, 186, 276, 226, 172, 215\}$$

Note that the only difference between these two sets of outcomes is the observation of the seventh trial, in which we observed 400 heads in S_1 *and 276 heads in* S_2. *To find the median of each set, we place the observed values in order and take the average of the middle two values:*

$$S_1 = \{172, 174, 180, 183, 186, 215, 226, 232, 259, 400\} \rightarrow median(S_1) = \frac{186 + 215}{2} = 200.5$$
$$S_2 = \{172, 174, 180, 183, 186, 215, 226, 232, 259, 276\} \rightarrow median(S_2) = \frac{186 + 215}{2} = 200.5$$

Note that the median of S_1 *and the median of* S_2 *are the same, because the positioning of the numbers in the ordered data is not affected by the exact value of the largest number in the data set. Now, to find the arithmetic mean of each set, we simply take the sum of the data values and divide by the number of observations:*

$$mean(S_1) = \frac{232 + 174 + 259 + 180 + 183 + 186 + 400 + 226 + 172 + 215}{10} = \frac{2227}{10} = 222.7$$
$$mean(S_2) = \frac{232 + 174 + 259 + 180 + 183 + 186 + 276 + 226 + 172 + 215}{10} = \frac{2103}{10} = 210.3$$

We can see that the mean of S_1 *is larger than the mean of* S_2*; this difference is because each data value is used to compute the mean, and therefore extremely large or small data values will skew the mean in the direction of the extreme value. In this case, the set of observations* S_1 *had a large value of 400, and therefore its mean was larger than that of* S_2.

3.2 Patient Data

Data for 176 HIV-infected persons was retrospectively collected from medical records of Severance Hospital, South Korea. The patient data consists of observed values for the following variables: S_0, I_0, V_0, λ, d, k, μ, N_T, c, η, and ϵ.

The variables of this data set correspond with the initial conditions and parameters of our system in the following manner:

Our System	Patient Data	Interpretation
T_0	S_0	Initial number of susceptible CD4 T-cells present in the body.
T_0^*	I_0	Initial number of infected CD4 T-cells present in the body.
V_{I0}	V_0	Initial amount of virus present in the body.
V_{NI0}	-	Initial amount of non-infectious virus present in the body.
λ	λ	Production rate of CD4-T cells.
d	d	Death rate of susceptible CD4-T cells.
k	k	Viral infection rate.
δ	$d + \mu$	Death rate of infected CD4 T-cells.
N_T	N_T	Viral production rate.
c	c	Viral clearance rate.
κ	η	Reverse transcriptase inhibitor efficacy.
η	ϵ	Protease inhibitor efficacy.

Table 2: Relation of the variables in the patient data set to the initial conditions and parameters of the model, along with their interpretations.

Note that no variable in the patient data provided corresponds directly to the initial amount of non-infectious virus present in the body, which is the initial condition V_{NI0} in our model. For the purposes of producing an approximate solution, we will assume that $V_{NI0} = 0$ for each patient.

The arithmetic mean, or average, of each of the variables in the patient data can be computed by adding all observed values of the variable, and dividing by the sample size, which is the number of patients in the data set. Because we have an even sample size, the median of each variable can be computed by taking the average of the two values in the middle of the ordered observations. Average and median values for each variable in the patient data can be seen in the table below.

Variable	Mean Value	Median Value
T_0	254.995	238.428
T_0^*	9.820	2.965
V_{I0}	9.8832	0.09127
λ	8.3978	2.388
d	0.0138	0.00447
k	0.15739	0.00264
δ	0.4888	0.01076
N_T	6039.344	44.938
c	553.618	8.968
κ	0.69886	0.6
η	0.3034	0.6

Table 3: Mean and median values for each variable in the patient data set.

Note that for some variables in the patient data, the differences between the average and median values are very large in relation to the values the variables take on themselves. For example, the mean value of the viral clearance rate c is 553.618, while its median value is 8.968.

The mean of a random variable is not robust because it is sensitive to outliers, which are observations of a random variable that are far from most of the observed values of that same variable in the data set [10].

Whether or not an observation is an outlier can be determined computationally using the interquartile range (IQR) of the data, which is a measure of its dispersion. The IQR can be computed by finding the difference between the 25^{th} and 75^{th} percentiles of the data.

Let Q3 denote the 75^{th} percentile of the data, and Q1 denote the 25^{th} percentile of the data. Then the interquartile range is $\text{IQR} = \text{Q3} - \text{Q1}$. An observation x is defined to be an outlier if $x < \text{Q1} - (1.5 \times \text{IQR})$ or if $x > (1.5 \times \text{IQR}) + \text{Q3}$. In the former case, x is referred to as a low outlier, and in the latter case, x is referred to as a high outlier.

Boxplots are a convenient way to visually describe the data and determine which points in a data set are outliers, and how much these points impact the mean. A boxplot is created using the median and interquartile range of the data. As an example, consider Figure 6, a boxplot of the initial condition T_0 in the patient data.

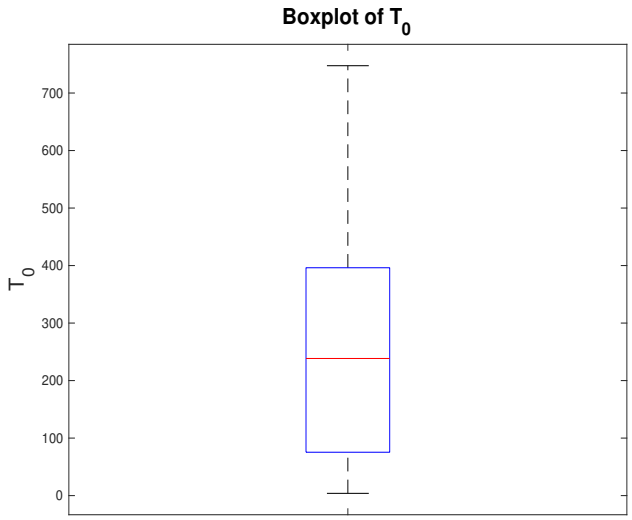


Figure 6: Boxplot of initial condition T_0 in patient data.

The blue box represents the bounds of the interquartile range of T_0. The top line of the blue box is the 75^{th} percentile of the data, and the bottom line of the box is the 25^{th} percentile of the data. The dotted black line represents the spread of the entire data set, with the maximum value of the data corresponding to the top black line and the minimum value of the data set corresponding to the bottom black line.

To get a better understanding of how outliers can be identified through the use of boxplots, consider the Figure 7 below, a boxplot of the variable λ in the patient data.

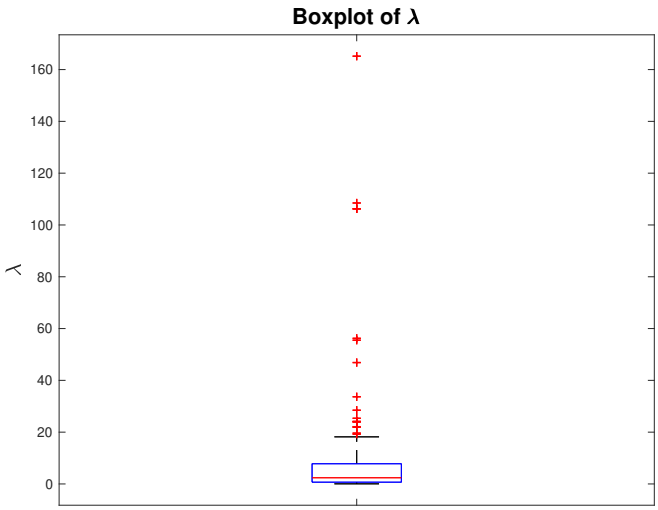


Figure 7: Boxplot of random variable λ in patient data.

Again, the blue box represents the interquartile range of λ, with the red line signifying the median of the data set, and the bottom and top lines of the blue box corresponding to the 25^{th} and 75^{th} percentiles of the data, respectively. Each red cross above the blue box represents a high outlier of the data set.

Note that there are many high outliers in the set of observations of λ. Because of these outliers, it is difficult to get a good understanding of the distribution of λ from this plot. A boxplot of the random variable λ after removal of outliers can be seen below.

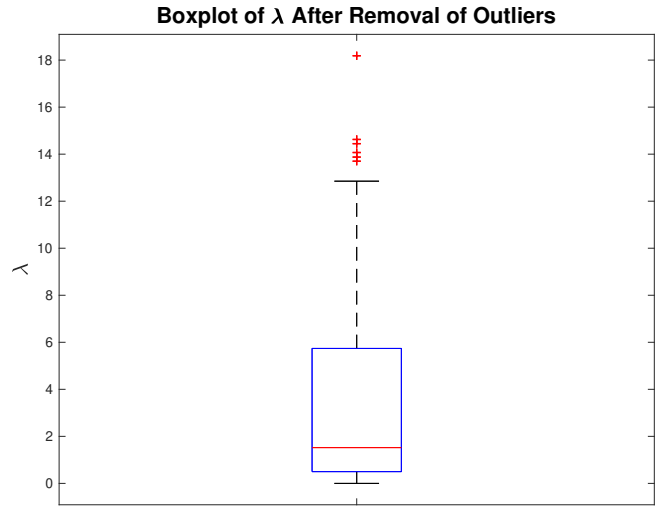


Figure 8: Boxplot of random variable λ in patient data after removal of outliers.

Note that even after removal of the original outliers, there are still outliers present in our data set. In this case, removal of outliers impacted the values of the 25^{th} and 75^{th} percentiles of the data, and therefore the criteria for a data point being an outlier changed as well. However, we can still see that after removal of the initial outliers the underlying distribution of λ is much clearer.

Generally, when trying to identify underlying trends in the distributions of data sets, it is not advised to completely remove outliers and make firm statistical conclusions with the new data set. However, in our case, the data is being used to identify underlying distributions of each random variable, and therefore it is beneficial to temporarily remove the outliers from the data to determine the possible distributions of each variable.

The impact of outliers on the average value of a random variable is further illustrated in the table below, which shows mean and median values for each variable in the patient data before the removal of outliers, as well as the mean and median values after the outliers were removed for each individual variable. Boxplots for each variable before and after removal of outliers are also included.

Variable	Before Removal of Outliers		After Removal of Outliers	
	Mean Value	Median Value	Mean Value	Median Value
T_0	254.995	238.428	254.995	238.428
T_0^*	9.820	2.965	4.144	12.296
V_{I0}	9.8832	0.09127	0.1355	0.06209
λ	8.3978	2.388	3.716	1.899
d	0.0138	0.00447	0.0074	0.00371
k	0.15739	0.00264	0.00405	0.001515
δ	0.4888	0.01076	0.01933	0.00723
N_T	6039.344	44.938	49.6379	35.6186
c	553.618	8.968	16.655	7.059
κ	0.69886	0.6	0.69886	0.6
η	0.3034	0.6	0.3034	0.6

Table 4: Mean and median values of each variable before and after the removal of outliers.

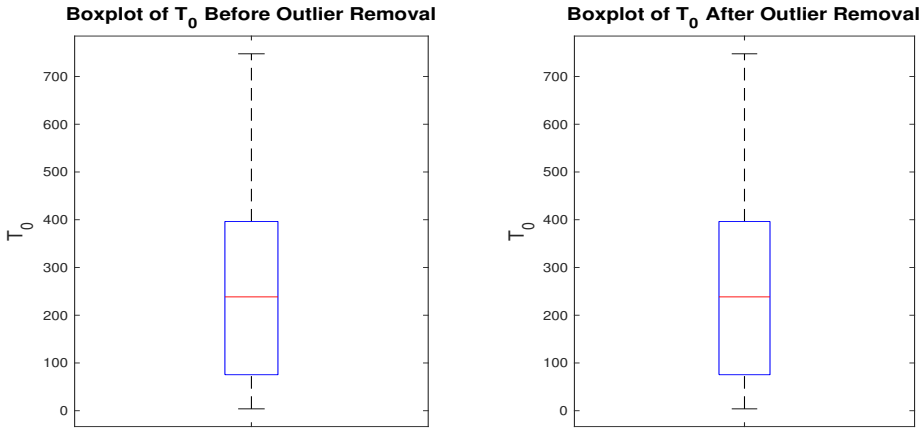


Figure 9: Comparison of T_0 boxplot before and after outlier removal.

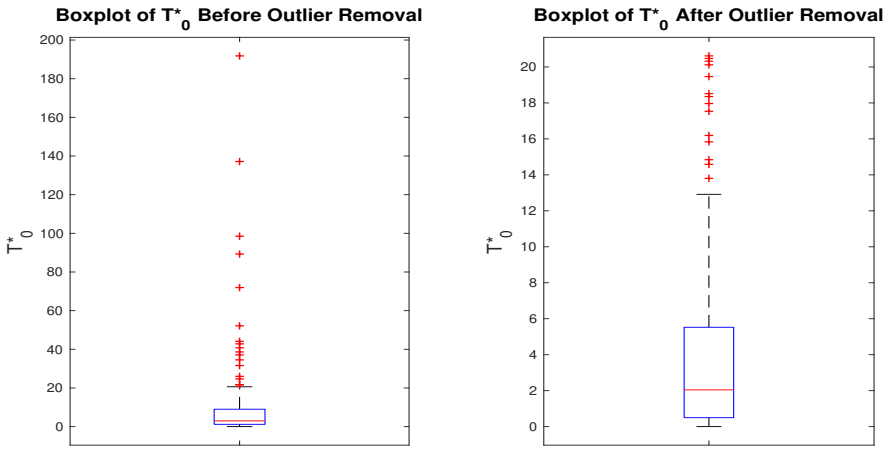


Figure 10: Comparison of T_0^* boxplot before and after outlier removal.

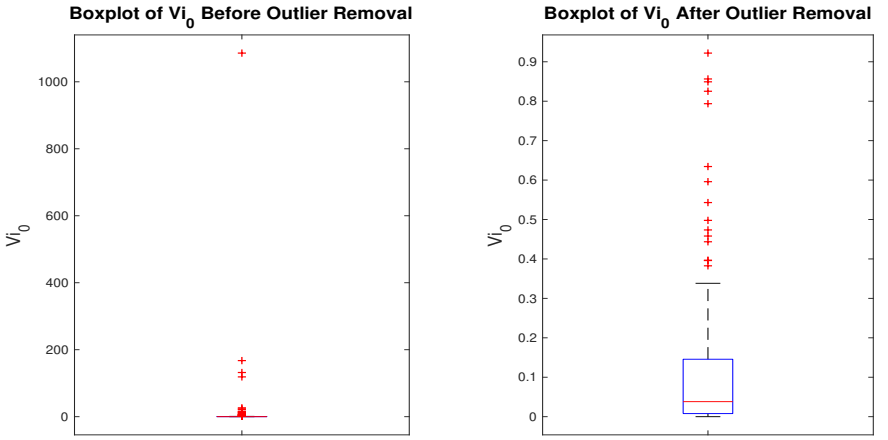


Figure 11: Comparison of V_{I0} boxplot before and after outlier removal.

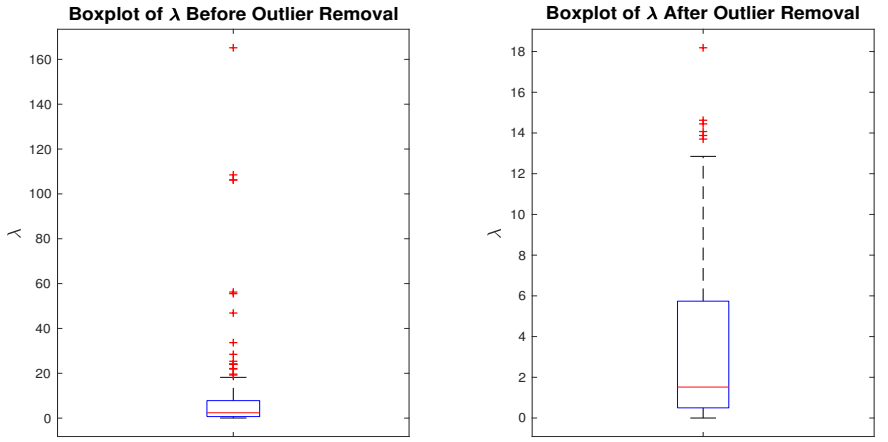


Figure 12: Comparison of λ boxplot before and after outlier removal.

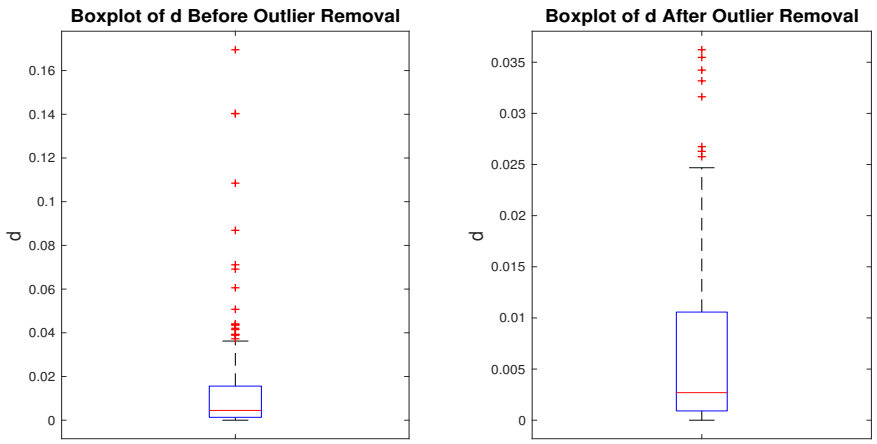


Figure 13: Comparison of d boxplot before and after outlier removal.

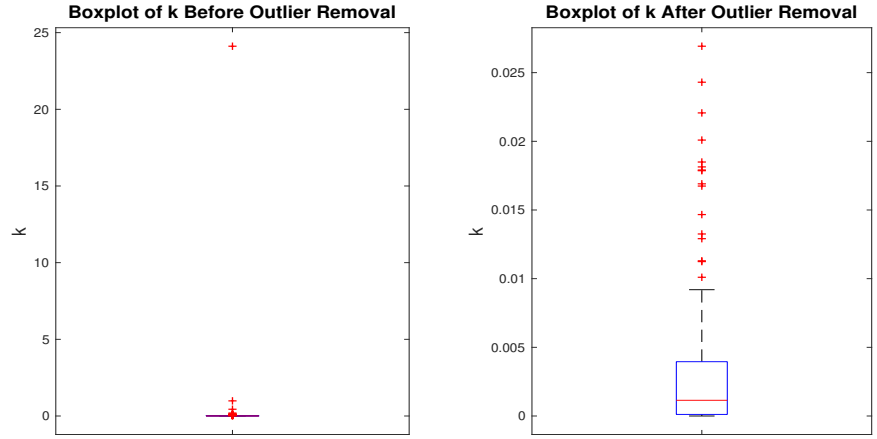


Figure 14: Comparison of k boxplot before and after outlier removal.

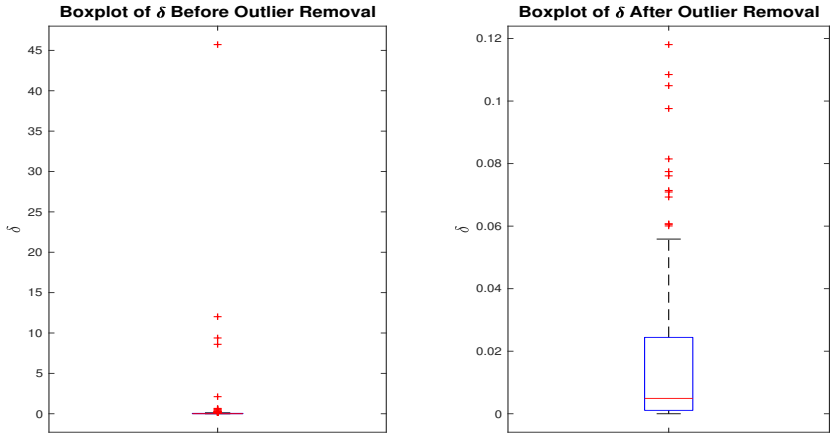


Figure 15: Comparison of δ boxplot before and after outlier removal.

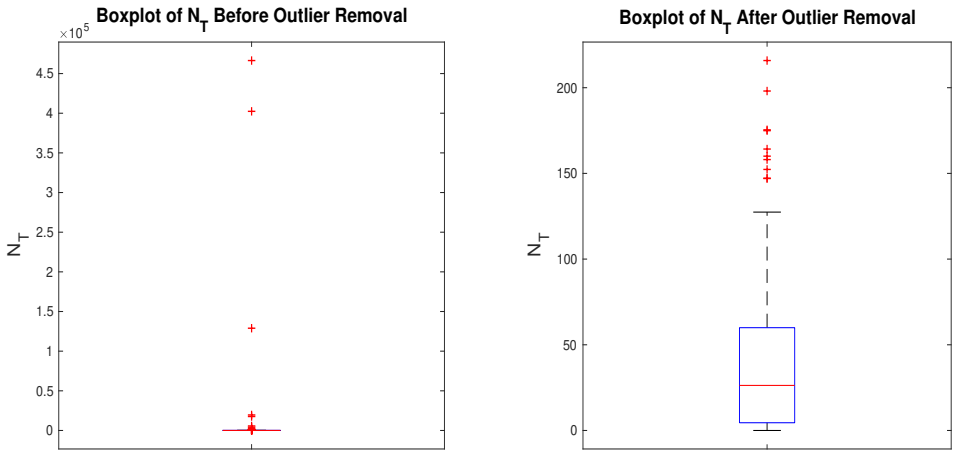


Figure 16: Comparison of N_T boxplot before and after outlier removal.

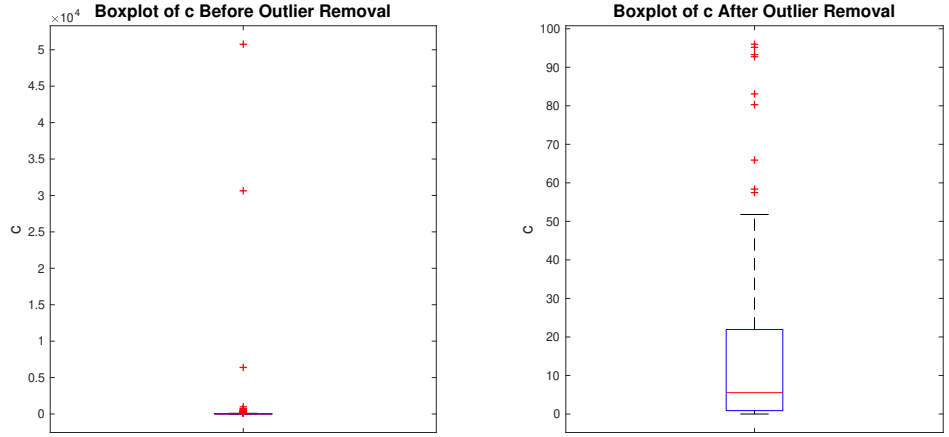


Figure 17: Comparison of c boxplot before and after outlier removal.

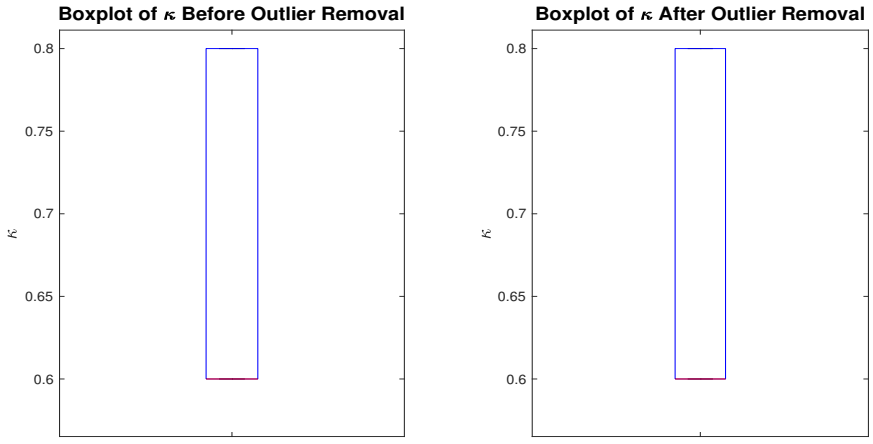


Figure 18: Comparison of κ boxplot before and after outlier removal.

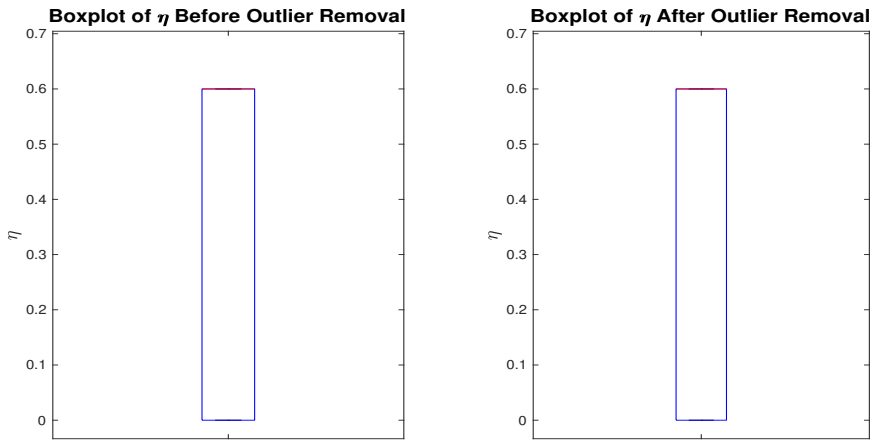


Figure 19: Comparison of η boxplot before and after outlier removal.

3.3 The Monte Carlo Method

The Monte Carlo Method is a technique used to analyze phenomena demonstrated by stochastic systems of equations by use of the generation of random numbers [12]. In order to understand the underlying process of the Monte Carlo Method, it is crucial to understand the concepts of the probability function and expected value of a random variable.

Recall from Definitions 5 and 6 in Section 3.1 that the probability mass function and probability density function are functions that describe the probability properties of a random variable X.

Definition 8. *The expected value of a random variable X, often referred to as the mean or weighted average of X, denoted $\mathbb{E}(X)$, is the average of the values we expect to obtain when multiple trials of an experiment are observed [12].*

For a discrete random variable, we use the formula $\mathbb{E}(X) = \sum_{\Omega} x \text{ pdf}(x)$, and for a continuous random variable, $\mathbb{E}(X) = \int_{-\infty}^{\infty} x \text{ pdf}(x)dx$.

Example 6. *Recall the 30 trial coin-toss experiment from Example 2 in Section 3.1. Since X is a discrete random variable, $\mathbb{E}(X) = \sum_{\Omega} x \ pdf(x)$, and we have*

$$\begin{aligned}\mathbb{E}(X) &= (0 \cdot pdf(0)) + (1 \cdot pdf(1)) + (2 \cdot pdf(2)) \\ &= (0 \cdot 0.333) + (1 \cdot 0.4) + (2 \cdot 0.267) = 0.934.\end{aligned}$$

Therefore, based on the information that has been gathered from this experiment, on average, we expect to observe heads 0.934 *times when a fair coin is tossed three times.*

The Monte Carlo Method essentially works in the following manner: we will generate independent random observations $X_1, X_2, \ldots, X_n$ for each random variable in our model and use these observations to obtain a solution to our model. We then find the expected value of these solutions, and the result of this calculation is the solution to our stochastic model by the Monte Carlo Method.

To use this method on our stochastic system, we need to identify the probability distributions of each random variable in the model. One way to determine the possible underlying probability distributions of data sets is to create a histogram of the data to visualize its distribution. The figures below show histograms for each random variable in the patient data set.

Because the purpose of creating these histograms is to get an idea of the distribution of each data set, the outliers of each data set were removed before the histograms were created.

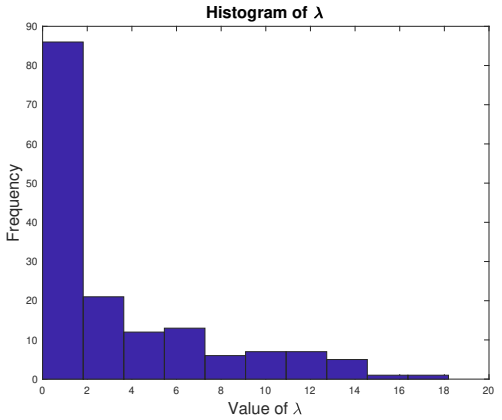


Figure 20: CD4 T-cell production rate.

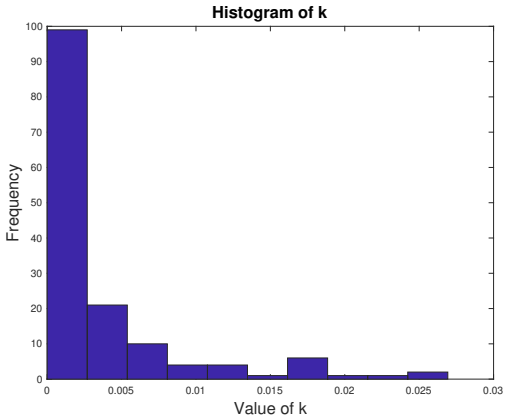


Figure 21: CD4 T-cell infection rate.

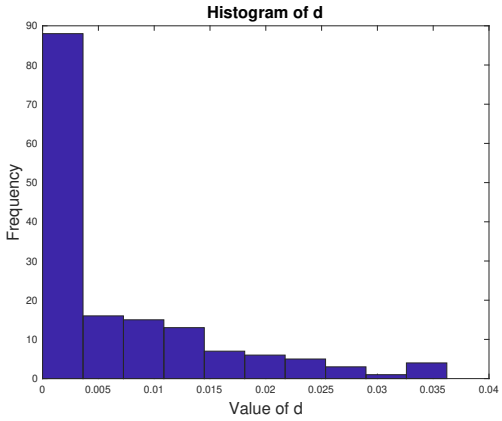


Figure 22: Susceptible CD4 T-cell death rate.

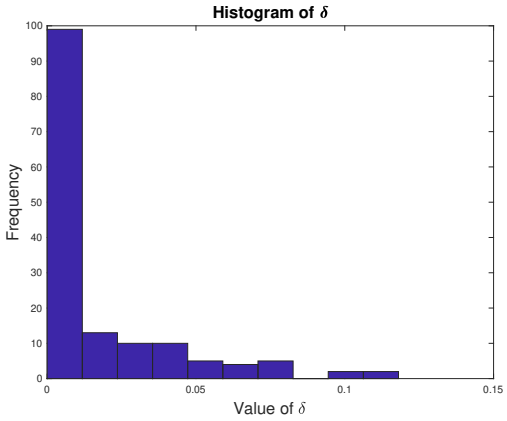


Figure 23: Infected CD4 T-cell death rate.

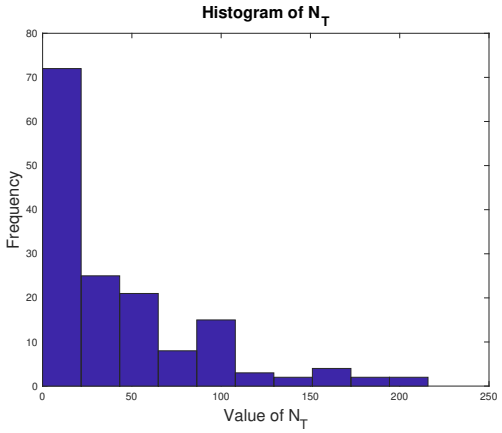


Figure 24: Virion production rate.

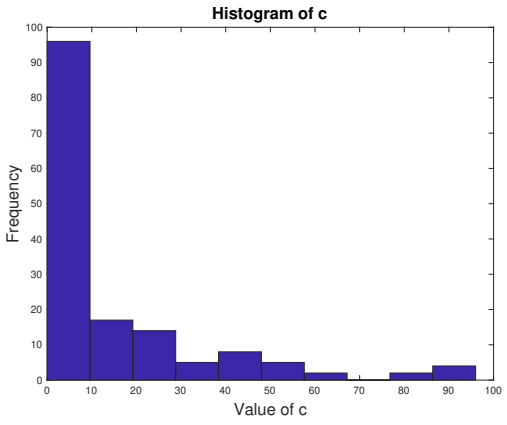


Figure 25: Viral clearance rate.

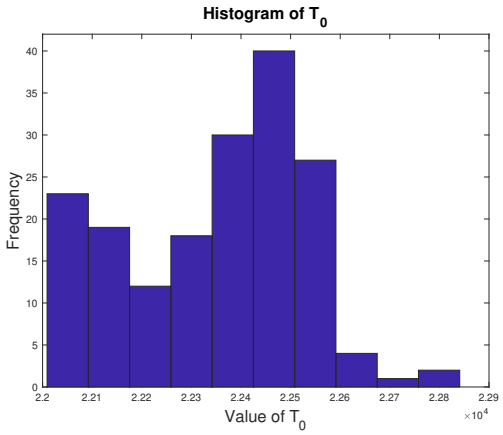


Figure 26: Initial CD4 T-cell count.

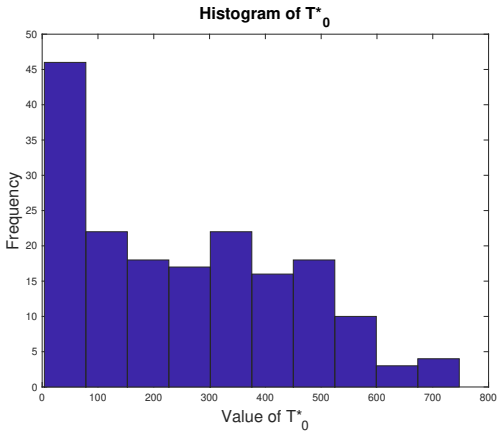


Figure 27: Initial infected T-cell count.

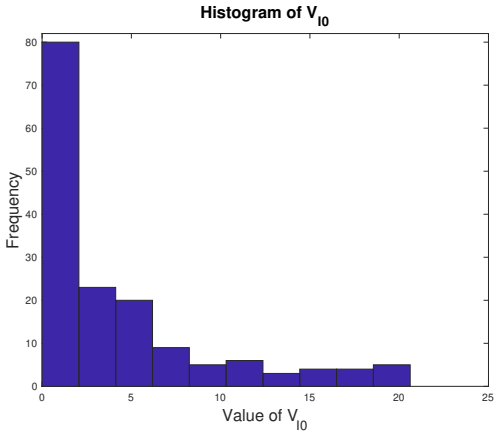


Figure 28: Initial concentration of infectious virus.

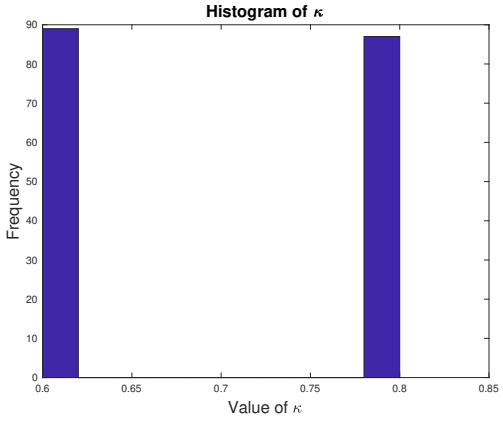


Figure 29: RT inhibitor efficacy.

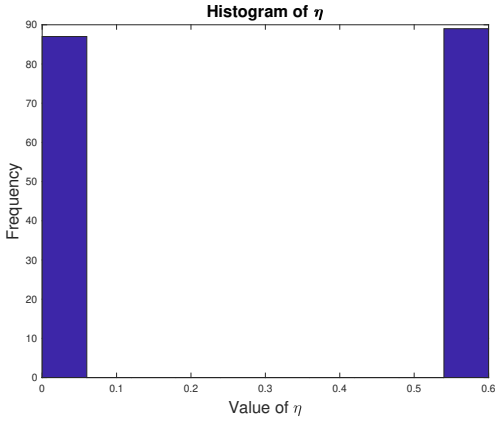


Figure 30: Protease inhibitor efficacy.

The histograms of λ, k, d, δ, N_T, c, T_0, T_0^*, and V_{I0} all seem to resemble the shape of either an exponential or gamma distribution. The probability density function of a gamma random variable is given by the function

$$g(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t},$$

where $t \geq 0$, the parameters $\alpha, \beta > 0$, and $\Gamma(\alpha) = (\alpha - 1)!$ [10].

The exponential distribution is a special case of the gamma distribution with $\alpha = 1$, and therefore its probability density function is given by

$$f(x) = \beta e^{-\beta x},$$

where $x \geq 0$ and the parameter $\beta > 0$ [10].

The histograms of κ and η, however, do not resemble either an exponential or a gamma distribution. The shape of their histograms are very different from the rest, and the distribution they most closely resemble is that of a uniform distribution. According to [10], the probability density function of a continuous uniform distribution on the general interval $[\alpha, \beta]$ is given by

$$f(x) = \frac{1}{\beta - \alpha}, \quad \alpha \leq x \leq \beta.$$

In order to use these distributions to generate random samples for each of these random variables in the model, we need to estimate the parameters of each distribution based on the patient data. The parameter estimates for each of these distributions was obtained using the Method of Moments as described in [10], with the following results:

Random Variable	**Estimated Distribution and Parameter Values**
T_0	Gamma ($\alpha = 1.802, \beta = 141.4863$)
T_0^*	Exponential ($\beta = 0.2413$)
V_{I0}	Gamma ($\alpha = 0.013898, \beta = 0.26095$)
λ	Exponential ($\beta = 0.2691$)
d	Exponential ($\beta = 135.3297$)
k	Gamma ($\alpha = 0.5086, \beta = 0.00796$)
δ	Exponential ($\beta = 51.7362$)
N_T	Exponential ($\beta = 0.02015$)
c	Gamma ($\alpha = 0.5675, \beta = 29.3465$)
κ	Uniform ($\alpha = 0.6, \beta = 0.8$)
η	Uniform ($\alpha = 0, \beta = 0.6$)

Table 5: Parameter estimates for the distributions of each random variable in the model.

3.4 Numerical Solution to the Stochastic Model

The distributions in Table 5 were used to generate random values for each variable using MATLAB random generators. The expected values of the solutions to T, T^*, V_I, and V_{NI} were then computed. Figure 31, seen below, is an example of the output produced by the Monte Carlo Method.

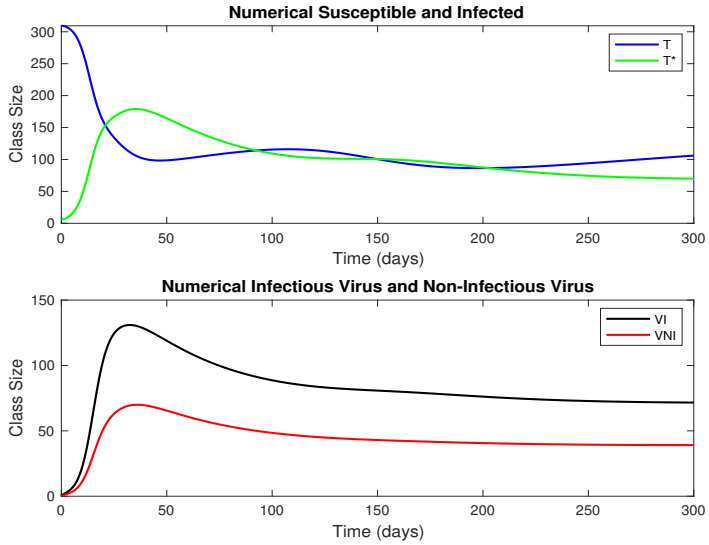


Figure 31: Approximate solution to the stochastic model obtained using the Monte Carlo Method.

A solution to the model was also computed using simply the average of the random values obtained for each variable in the model. The output of this solution is seen in the figure below.

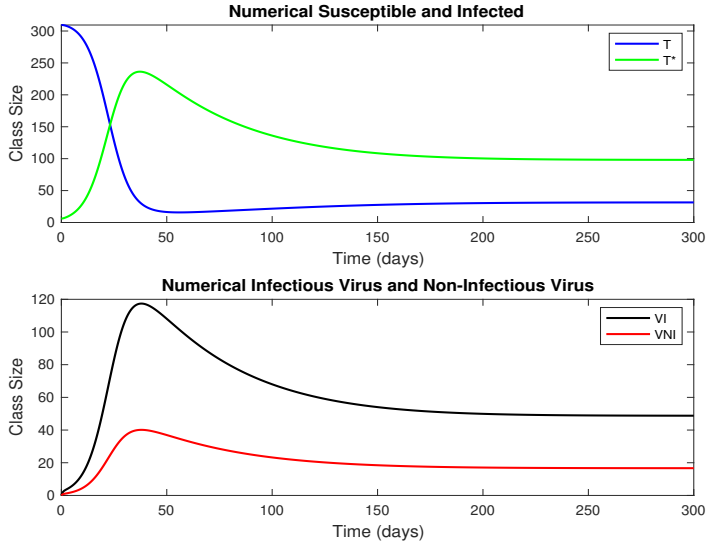
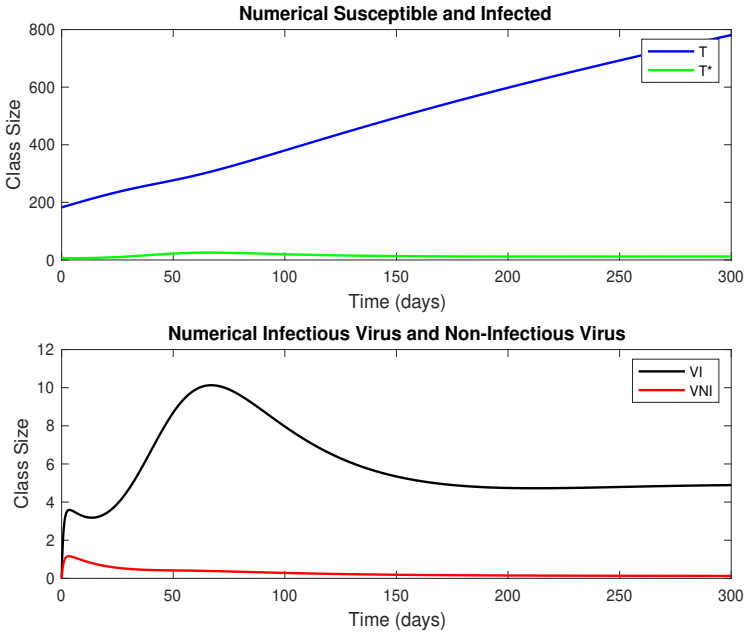


Figure 32: Approximate solution to the stochastic model obtained using the average values of the random samples for each random variable in the model.

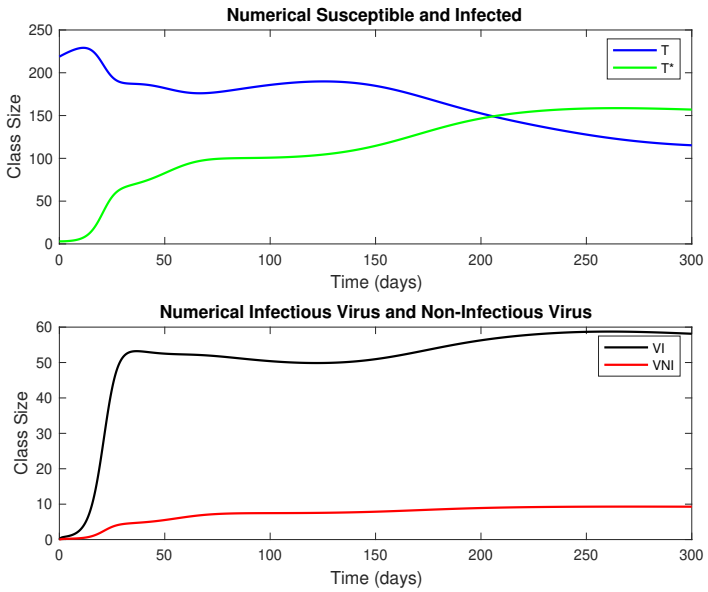
The output seen in Figure 31 is similar to the behavior we would expect in this model based on the approximation to the deterministic solution obtained in Chapter 2. However, because of the randomness involved in the Monte Carlo Method, it is interesting to see how the differences in parameters and initial conditions change the behavior of the model. Consider Figures 33 through 36 below.



Parameter	Value
T_0	88.0686
T_0^*	4.5488
V_{I0}	4.4013×10^{-25}
λ	2.0098
d	0.031
k	2.771×10^{-5}
δ	0.0069
N_T	59.9326
c	2.5783
κ	0.7078
η	0.4189

Figure 33: Example of an approximate solution obtained using Monte Carlo Method.

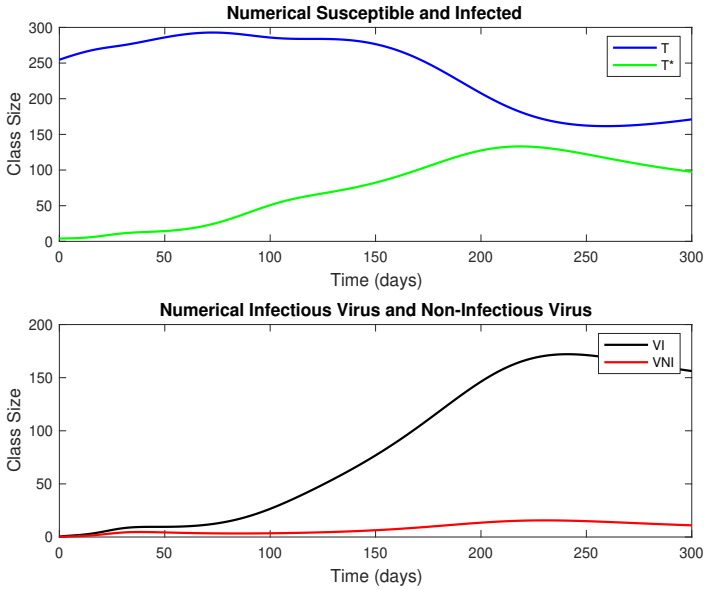
In Figure 33, we see an example of a patient whose susceptible cell count is steadily increasing, seemingly without bound. This increase is likely due to the high value of the parameter λ relative to the constants d and k, as well as the high initial CD4+ T-cell count, contrasted by the low initial virus concentration.



Parameter	Value
T_0	154.0674
T_0^*	0.6735
V_{I0}	1.4679×10^{-12}
λ	4.1360
d	0.0100
k	0.0036
δ	0.0142
N_T	21.2045
c	1.4111
κ	0.7172
η	0.1573

Figure 34: Example of an approximate solution obtained using Monte Carlo Method.

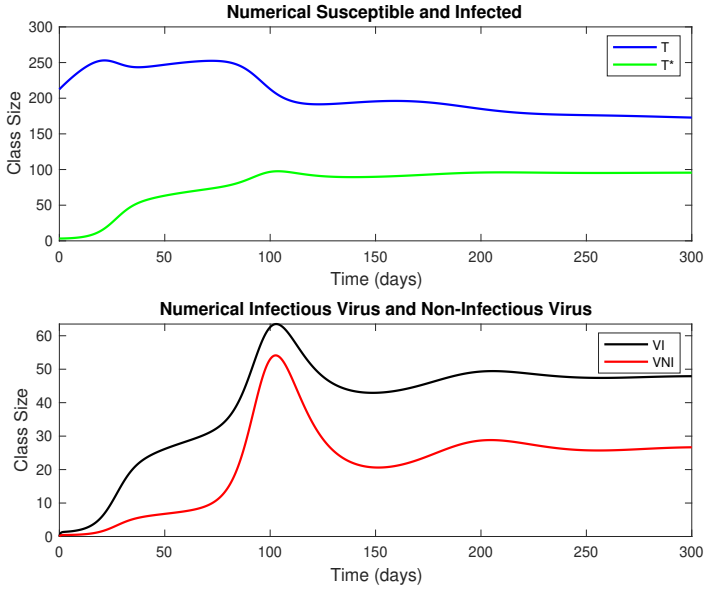
In Figure 34, we see an example of a patient whose virus concentration initially grows in what appears to be an exponential fashion, and then slowly begins to level off. This high increase in virus concentration is likely due to the low viral clearance rate (c) combined with a low level of efficacy for the RT inhibitor (η).



Parameter	Value
T_0	47.7870
T_0^*	2.1338
V_{I0}	3.4754×10^{-12}
λ	0.9618
d	0.0034
k	0.0099
δ	0.0428
N_T	96.3058
c	2.5466
κ	0.6636
η	0.3652

Figure 35: Example of an approximate solution obtained using Monte Carlo Method.

In Figure 35, we see that virus concentration slowly increases (in contrast to Figure 34 above, in which we saw virus concentration increasing extremely quickly). In this scenario, it appears that we have a fairly high viral production rate due to the values of N_T and δ being relatively high compared to the previous scenarios.



Parameter	Value
T_0	82.2732
T_0^*	0.1557
V_{I0}	8.6507×10^{-136}
λ	8.7534
d	0.0154
k	0.0016
δ	0.1639
N_T	99.1055
c	3.8313
κ	0.7089
η	0.59862

Figure 36: Example of an approximate solution obtained using Monte Carlo Method.

The patient in Figure 36 has an extremely small amount of infectious virus initially present in the body. However, the high viral production rate ($N_T\delta$) allows the virus concentration to grow fairly quickly early on. This high viral production rate is then contrasted by the relatively high viral clearance rate, which causes the virus population to hit a peak and begin to decrease. The behavior of V_I and V_{NI} in this model is similar to the behavior we would expect based on the analysis of the numerical approximation to the deterministic model in Chapter 2.

Use of the Monte Carlo Method with the probability distributions of each constant and initial condition allows us to account for the randomness that occurs biologically within the model. Obtaining a solution to the stochastic model gives us a better understanding of how HIV-1 Dynamics may change in a given patient due to differences in model parameters.

4 Conclusion

This research concludes our analysis of HIV-1 dynamics using the T, T^*, V_I, V_{NI} model. In this project, we obtained an analytical solution to a simplified version of the model that included only healthy cell production rate, death rates, viral production rate, and efficacy of treatment methods. We also used a Multistep Method to obtain a numerical solution to the full deterministic system. We then discussed a stochastic version of the system, and used the Monte Carlo Method to obtain a numerical approximation to the solution of our stochastic system. The results of this project display the development of HIV-1 in an infected patient under the effect of reverse transcriptase and protease inhibitors, and these results are made more practical with the inclusion of an element of randomness in the stochastic system of equations.

References

[1] Richard L. Burden and Dougles J. Faires, *Numerical analysis*, 7 ed., Thomson Learning, Inc., 2001.

[2] Richard L. Burden, Dougles J. Faires, and Annette M. Burden, *Numerical analysis*, 10 ed., Cengage Learning, 2016.

[3] Duncan S. Callaway and Alan S. Perelson, *HIV-1 infection and low steady state viral loads*, Bulletin of Mathematical Biologu **64** (2002), no. 1, 29-64.

[4] O. Diekmann and J.A.P. Heesterbeek, *Mathematical epidemiology of infectious diseases*, Wiley Series in Mathematical and Computational Biology, John Wiley and Sons Ltd, 2000.

[5] Emily MacIndoe, *Analysis of deterministic and stochastic HIV models*, Undergraduate honors thesis, University of Mary Washington, 2019.

[6] J.D. Murray, *Mathematical biology I: An introduction*, 3 ed., Springer, 2001.

[7] A.S. Perelson, Avidan U. Neumann, Martin Markowitz, John M. Leonard, and David D. Ho, *HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time*, Science **271** (1996), 1582-1586.

[8] Alan S. Perelson, Denise E. Kirschner, and Rob De Boer, *Dynamics of HIV infection of CD4+ T cells*, Mathematical Biosciences **114** (1993), 81-125.

[9] Alan S. Perelson and Patrick W. Nelson, *Mathematical analysis of HIV-1 dynamics in vivo*, SIAM **41** (1999), no. 1, 3-44.

[10] John A. Rice, *Mathematical statistics and data analysis*, 3 ed., Brooks/Cole, Cengage Learning, 2007.

[11] Peter A. Roemer, *Stochastic modeling of the persistence of HIV: Early population dynamics*, Master's thesis, United States Naval Academy, 2013.

[12] Ronald W. Shonkwiler and Franklin Mendivil, *Exploration in monte carlo methods*, Springer Science+Business Media, LLC, 2009.

[13] Henry C. Tuckwell and Emmanulle Le Corfec, *A stochastic model for early HIV-1 population dynamics*, J. Theor. Biol. **195** (198), 451-463.