CONTINUOUS-TIME BINARY BRANCHING PROCESSES FOR VIRUS EVOLUTION IN SERIAL PASSAGES

ALLISON YUSIM

Abstract. Serial passages experiments provide valuable insights on the dynamics of a virus's evolution over time, but are often costly and require materials that can be difficult to obtain. We created a stochastic model for such experiments to account for the probabilistic nature of virus evolution. We gathered information from previous experiments to develop realistic assumptions and used a binary branching process where the branching rate is inversely proportional to the length of the virus to build a simulation for virus replication in host cells. We then modeled the process for two special cases, one of which is a simple birth-death process modeled by a continuous-time Markov chain that converges to an ordinary differential equation. We modeled the second case with a more complex stochastic equation and simulation statistics suggest that it will converge to a partial differential equation.

1. Background and Motivation

Virus evolution is rapid, extensive, and often difficult to track. Gaining insights on how genomes change as they adapt to different host environments allows scientists to better understand how viruses work, which can lead to advancements in vaccines, medications, and cures. There have been many promising experiments done to observe the evolution trends of different viruses. In particular, serial-passage experiments have revealed interesting insights. We focus on simulating virus replication in the context of a serial-passages experiment and developing a model that will ultimately show us a deterministic outcome for virus evolution in the long run.

In a typical serial passages experiment, we begin with N virus genomes that are introduced into a culture of host cells. These genomes are $M \in \mathbb{N}$ nucleotides long. They undergo replication within the host cells, where mutations can occur. The offspring may have less (deletion), more (insertion), or different (substitution) nucleotides than the parent. If deleterious mutations occur, the offspring are unable to survive, reproduce, or exit host cells. We refer to the genotypes that cause virus particles to lose such abilities as DIPS (defective interfering particles), while the functional genotypes are referred to as WT (wild type) particles. We model this within-host cell replication process in this paper.

We only consider a particle a DIP if it loses the enzyme that allows a virus to replicate in a new host cell called the RNA-dependent RNA polymerase (RdRP). If a DIP enters a host cell alone, it will not be able to replicate, but if it enters a host cell togther with a WT particle, it will replicate normally. If a particle becomes a DIP, all of its offspring will also be DIPs, i.e. offspring cannot have the RdRP enzyme if the parent does not have it.

Date: August 13, 2024.

After a fixed amount of time, a subset of the free virions, i.e. the particles that have exited the host cell, are selected as the new initial population for the next round of passages. This subset is called the bottleneck and is often a fixed percentage of the free virions. Through repeated passages, researchers observe which genotypes become more prevalent, providing insights into the fitness of different genetic variants. It is important to note for this paper that the number of nucleotides, i.e. the length, of a genome heavily impacts its rate of replication.

Serial passage experiments have many drawbacks, including time, cost, and availibility of materials. One key drawback is that scientists are not able to see the genotypes of every virion, as doing so would destroy the particle, preventing it from being selected for the bottleneck. This is why creating mathematical models to simulate serial passages is crucial, as these models can provide insights into viral evolution and genotype dynamics without the need for extensive experimental resources. It is also important to account for the stochastic nature of virus evolution, as events such as mutations are rarely predictable and can significantly impact the evolutionary trajectory of viral populations.

To address these challenges, we have developed a stochastic model that uses a continuous-time binary branching process in order to simulate within-host cell replication in serial passages. The replication process is the most crucial part of a serial passages experiment because it is the only part that is not controlled by the researcher. Our model focuses specifically on the length of the virus particles during replication. We look at the relationship between a virus particle's replication rate and its length as well as deletions and instertions of nucleotides during the replication process. These mutations impact the length of the genomes, which in turn impacts the replication rate. All of these factors together impact the virus's overall fitness over time.

2. Model Description

In this paper, we consider a continuous-time binary branching process that models virus evolution over time.

2.1. **Within-Host Cell Model.** Here, we specifically focus on the number of nucleotides, i.e. the length, of the virions as well as the existence of the replication enzyme, which we will refer to as E. The model works as follows:

1. At time $t = 0$, we begin with one particle inside of one host cell. We assume that the inital particle has E, which is exactly one nucleotide long.
2. When this particle replicates, it gives birth to exactly one offspring. Deletion or insertion may occur during replication, so the offspring and parent can have different lengths.
3. If deletion occurs, then E may get deleted. If a particle loses E and becomes a DIP, then its offspring will also be a DIP. Even if insertion occurs, every particle with a DIP ancestor will be a DIP.
4. The replication process continues, now with two particles. A particle is born each time replication occurs.
5. Each particle can die at rate μ, which is the only way for it to stop replicating.
6. At a chosen time t, we observe the mutations that have occurred.
7. This process is repeated multiple times to represent replication in multiple host cells.

This process is a continuous time Markov chain with state space $\mathbb{S}$. We call the length of the i-th particle L_i. The i-th particle is represented by sequence of length L_i $\forall i$. The state space for an individual particle is $S = \{1, 2, \dots\}X\{\text{FS, DS}\}$, where DS is a defective sequence (i.e. a DIP particle) and FS is a functional sequence (i.e. a WT particle). The state space for the whole Markov chain is $\mathbb{S} = S^j$, where j is the number of particles (state) of the process at time t.

When a particle replicates, we first need to decide whether insertion or deletion occurs. Let $Y \sim \text{Bernoulli}(\sigma)$. Y decides whether insertion (with probability σ) or deletion (with probability $1 - \sigma$) occurs. We assume that deletions of nucleotides occur independently and are identically distributed (i.i.d.). Therefore, this event follows a binomial distribution. With deletion probability p, the distribution of deletion is $X \sim \text{Bin}(L_i, p)$. Thus, if deletion occurs, then the length of the offspring of the ith particle is $L_i - X$. Note that for the ith particle, if deletion occurs, then E is deleted with probability $\binom{X-1}{L_i-1}/\binom{X}{Li}$. If insertion occurs, then the length of the offspring of the ith particle is exactly $L_i + 1$. Note that deletion must occur in order for a particle to become a DIP.

When a virus is longer, it has a higher replication time. In this model, the replication time follows an exponential distribution where the maximum rate λ is inversely proportional to the length of the virus. For the ith particle, we have $T_{\text{birth}} \sim \text{Exp}(\frac{\lambda}{L_i})$. The death time of the virus follows an exponential distribution with rate μ, so for each particle, $T_{\text{death}} \sim \text{Exp}(\mu)$. We use t to refer to the time within each serial passage and T to refer to the time at which we choose to stop the replication process. Figure (1) depicts an example of the branching process within a host cell. Refer to Table 1 for the values we use for the variables in this paper.

Variable	Value	Description
σ	0.9	The probability that insertion will occur.
λ	0.5	The replication rate of a virus particle.
μ	0.001	The death rate of a virus particle.
p	0.03	The probability that one nucleotide will be deleted.

Table 1. Constant Variables

2.2. **Serial Passages.** To model the whole serial passages experiment, we use the within-host cell model as well as a bottleneck percentage. Suppose we have N host cells, N particles, and exactly one particles enters each host cell. Then we simulate the within-host cell model N times and choose a bottleneck percentage of free virions to use for the next serial passage. If the bottleneck percentage results in b particles being chosen, then we take b host cells (assuming each particle enters exactly one host cell) and simulate the within-host cell model b times. Then, we use the same bottleneck percentage of free virions and continue this process as many times as desired. This paper focuses on the within-host cell model, but future research may explore the serial passages model.

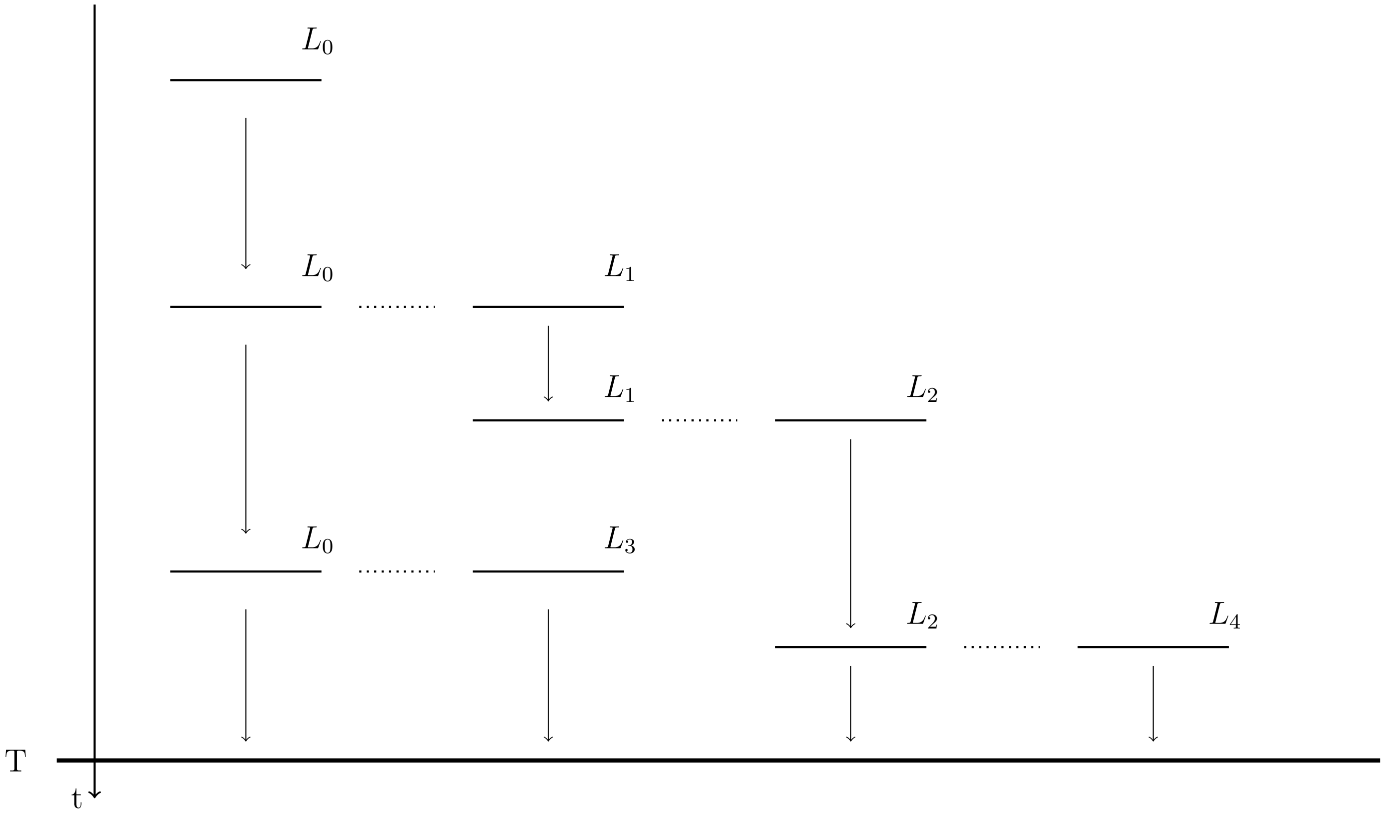

Figure 1. Particle 0 is the initial particle with initial length L_0. Each vertical arrow represents replication time with rate λ/L_i for particle i. Each dotted line represents replication and the birth of a new particle. Particle i is the ith particle born with length L_i. Notice that here, particle 1 died before the stopping time, T, so only particles 0, 2, 3, and 4 remain. Some of these particles may be longer than particle 0 and some may be shorter. If a particle is shorter than particle 0, it may be a DIP. Our model and this diagram use ideas from the continuous time binary branching process for coalescence [Lam18] [LS13].

3. Simulation

We created a program for the within-host cell model using python and simulated a special case, which is a slightly modified version of the model. In our simulation, we suppose $\sigma = 0$. This just means that insertion cannot occur. So each time a particle replicates, the length of the offspring will definitely be less than or equal to the length of the parent.

Our simulation shows us the distribution of lengths over time. We start with one particle of length 50. Over time, the average length decreases. Figure (2) depicts the results at 3 time points from 2 randomly chosen runs of our simulation.

As you can see, even though the images are not identical, the overall shape of the curves looks very similar. Note that the graphs from the simulations look similar, just not exactly at the same time. This is because we started with one particle, and its replication time is modeled with an exponential random variable. This means that the replication times of the

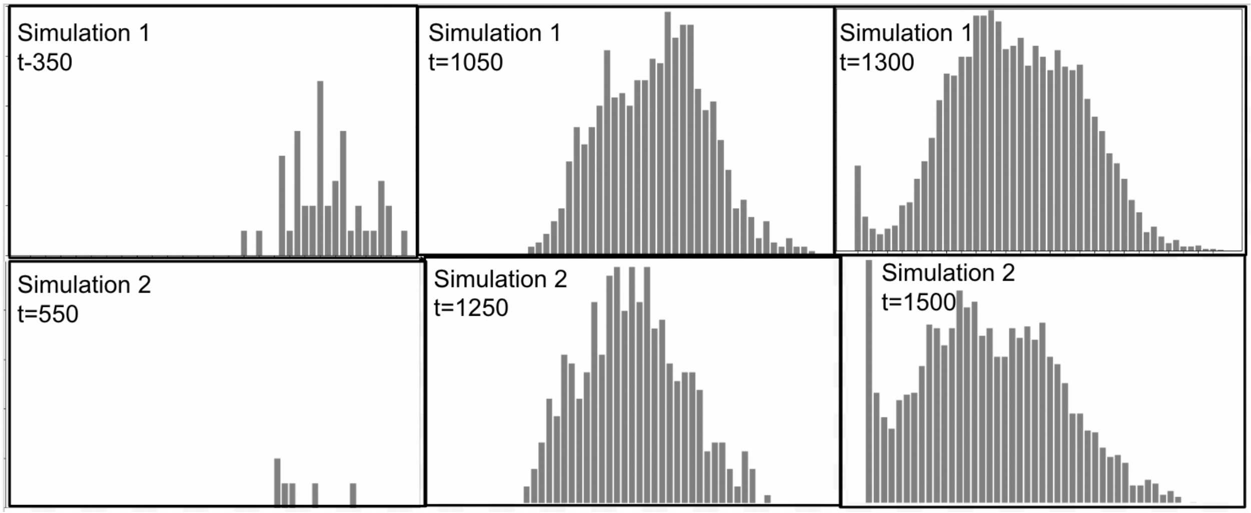

Figure 2. Here, we see 3 randomly chosen times from two randomly chosen simulations. We begin our simulation at $t = 0$ with one particle of length 50, so $M = 1$ and $l_{\max} = 50$. We use the values from table (1) for the parameters, so $\mu = 0.001, \lambda = 0.5$, and $p = 0.03$. The only difference is that insertion cannot occur, so we set $\sigma = 0$. On each graph, the x-axis represents the lengths and the y-axis represents the number of particles of each length x. The three images in the top row are snapshots of 3 different times from one simulation and the bottom row represents 3 different times from a second simulation.

initial particles from both simulations can vary greatly. However, because of the law of large numbers, we will see that for a large enough M, the expected replication time for the initial particles will be $l_{\max}/\lambda$ every time. This will eliminate the difference in times.

From these results, we predict that as our initial population M and our maximum length $l_{\max}$ approach infinity, the bars on the graph will get infinitely close together and by the law of large numbers, ultimately converge to a distribution that we can model with a deterministic equation. In order to successfully find this equation and prove that our stochastic model will converge to it, we must first find a way to mathematically quantify the distribution of lengths (number of particles of length L $\forall L \in [0, l_{\max}]$) at every time t.

4. Modeling With Stochastic Equations

4.1. **Simplified Birth-Death Process Equation.** As a starting point in finding our stochastic equation, we explored the ideas from equations that describe the following birth death-process. Suppose λ is constant. In the context of the model, this is the case where L does not change. Suppose $\lambda > \mu$. We can model the number of individuals in a host cell with the ordinary differential equation

$$\frac{dy(t)}{dt} = (\lambda - \mu)y(t),$$

where $y(t)$ is the number of individuals at time t.

If the initial number of particles is small, it is very likely that the population will go extinct very quickly. For instance, if we start with 1 individual, the probability, q_1, that the population will go extinct is

$$q_1 = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\mu+\lambda}(q_1)^2 \iff q_1 = \frac{\lambda}{\mu}.$$

We want our initial number of particles M to be large enough so that that the population does not go extinct. Let q_k be the probability that the population will go extinct if we start with k particles

Lemma 1. *$\exists M$ s.t. $\forall k \geq M$, $q_k = 0$*

Proof. If we start with 1 particle, we know that

$$q_1 = \frac{\lambda}{\mu}. \tag{1}$$

Since each round of replication is independent, we can say that

$$q_k = (q_1)^k \iff q_k = \left(\frac{\mu}{\lambda}\right)^k \tag{2}$$

Since we assumed that $\lambda > \mu$, we know that $\lim\limits_{k\to\infty} \left(\frac{\mu}{\lambda}\right)^k = 0$, so we can conclude that $\exists M > 0$ s.t. $\forall k \geq M, q_k = 0$. □

Going forward, we assume that our initial population M is large enough so that the virus does not go extinct. We consider the case where the population does not go extinct.

Let $M \in \mathbb{N}$ be the initial population large enough so that $q_M \approx 0$. Let $\left(Y_t^{(M)}\right)_{t\in\mathbb{R}_+}$ be the continuous time Markov chain that describes the within-host cell birth-death process of the virus starting at $Y_0^M = M$. The possible population at any time t, i.e. the state space $S = \{0, 1, 2, ...\}$.

Let $\mathcal{N}_\lambda$ be a Poisson Process with rate λ that models birth and let $\mathcal{N}_\mu$ be a Poisson Process with rate μ that models death. By the law of large numbers, we know that Y_t^M converges to its expected value as $M \to \infty$, so Y_t^M solves the equation

$$Y_t^M - Y_0^M = \mathcal{N}_\lambda\left(\int_0^t \lambda \cdot Y_s^{(M)} ds\right) - \mathcal{N}_\mu\left(\int_0^t \mu \cdot Y_s^{(M)} ds\right).$$

When we divide this equation by M to normalize it, we get

$$\frac{Y_t^M}{M} = \frac{Y_0^M}{M} + \frac{1}{M}\cdot\mathcal{N}_\lambda\left(\int_0^t \lambda\cdot M\cdot\frac{Y_s^{(M)}}{M} ds\right) - \frac{1}{M}\cdot\mathcal{N}_\mu\left(\int_0^t \mu\cdot M\cdot\frac{Y_s^{(M)}}{M} ds\right).$$

As $M \to \infty$, by the law of large numbers,

$$\frac{Y_t^{(M)}}{M} = \frac{Y_0^{(M)}}{M} + (\lambda-\mu)\left(\int_0^t \frac{Y_s^{(M)}}{M} ds\right).$$

So, if $\frac{Y_0^{(M)}}{M} \to 0$ as $M \to \infty$, we know that

$$\frac{Y_t^{(M)}}{M} = \frac{Y_0^{(M)}}{M} + (\lambda - \mu)\left(\int_0^t Y_s^{(M)} ds\right) \to (\lambda - \mu)\int_0^t y(s)ds + y(0) = y(t).$$

Since $(Y_t^{(M)})_{t\in\mathbb{R}_+}$ is a valid continuous time Markov chain for the within-host cell birth-death process that converges to $y(t)$, we can use the ODE $\frac{dy(t)}{dt} = (\lambda - \mu) \cdot y(t)$ to analyze the long term behavior of the virus. It is much easier to analyze this simple ODE than a simulation, so this method can give us better and more precise results.

This birth-death process is much simpler than our model, but we use a similar idea for our model, specifically the special case that we simulated in section (3). Following the same outline as for the birth-death process, we look to accomplish these steps:

(1) Find a stochastic equation that gives us the number of particles of each length at time t $\forall t \geq 0$.
(2) Find a deterministic equation (in our case, a partial differential equation) that describes the long-term behavior of the virus
(3) Prove that our equation in step (1) converges to a solution for our equation in step (2) as the initial number of particles and the initial length (maximum length) both approach infinity.

Accomplishing these steps allows us to be able to analyze the long-term behavior of the virus deterministically rather than stochastically, as well as use an equation rather than a simulation.

4.2. **Stochastic Equation for our model.** We consider the same special case of our model that we simulated in section (3). For every time t, we focus on keeping track of how many particles there are of each possible length $k \in L$, where $L = \{0, 1, \ldots, l_{\max}\}$, where $l_{\max}$ is the maximum possible length of a virus. Suppose insertion cannot occur, i.e. $\sigma = 0$. We ignore the DIPs here.

Our first task is to accomplish step 1: find a stochastic equation that gives us the number of particles of each length at time t $\forall t \geq 0$. We use the following stochastic equation to model the number of virions of each length L at time t:

$$\mu_t^{l_{\max},M} = \frac{1}{M} \cdot \sum_{k=1}^{l_{\max}} f_k(t) \cdot \delta_{\frac{k}{l_{\max}}},$$

where $\mu_t^{l_{\max},M}$ is a measure on the distribution of lengths.

Figure (3) gives us an example of the distribution of lengths for some time t. $\mu_t^{M,l_{\max}}$ sums the height of each vertical bar to give us the total number of particles at time t. In the graph,

- $x \in [0,1]$, where $x = \frac{k}{l_{\max}}$ for some length $k \in L$.
- $y = f_t(k)$, which is the number of particles of length k.

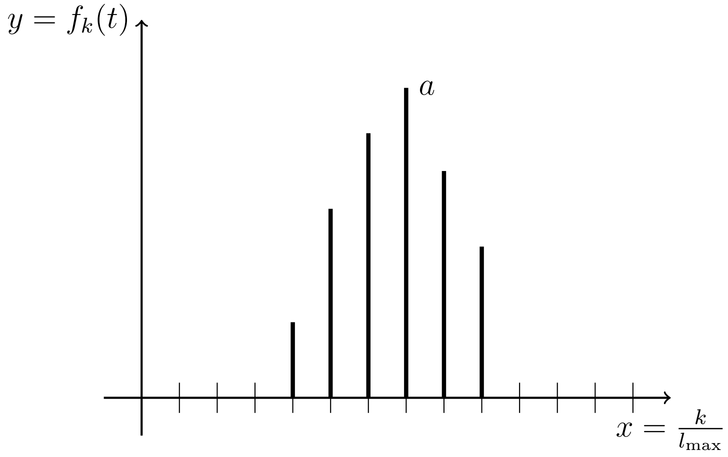

FIGURE 3. This graph depicts the length distribution at some time t. For example, node a is positioned at $x = \frac{k}{l_{\max}}$ with height $f_k(t)$. This means that there are $f_k(t)$ particles of length k at time t.

Note that $\delta_{\frac{k}{l_{\max}}}$ is a Dirac Delta measure on $[0,1]$ that excludes any k where $f(k) \leq 0$ and works as follows. Let $[a,b] \subset [0,1]$ be an interval that includes every $x \in [0,1]$ s.t. $f(x) > 0$. Then

$$\mu_t^{l_{\max},M}([a,b]) = \frac{1}{M} \cdot \sum_{k=1}^{l_{\max}} f_k(t) \cdot \delta_{\frac{k}{l_{\max}}}([a,b]),$$

where

$$\delta_{\frac{k}{l_{\max}}}([a,b]) = \begin{cases} 1 \text{ if } \frac{k}{l_{\max}} \in [a,b] \\ 0 \text{ if } \frac{k}{l_{\max}} \notin [a,b] \end{cases} = \begin{cases} 1 \text{ if } l_{\max} \cdot a \leq k \leq l_{\max} \cdot b \\ 0 \text{ otherwise} \end{cases}$$

So our equation can be rewritten as

$$\mu_t^{l_{\max},M}([a,b]) = \frac{1}{M} \cdot \sum_{a \cdot l_{\max} \leq k \leq b \cdot l_{\max}} f_k(t).$$

For every $k \in [a,b]$, $f_k(t)$ gives us the number of particles of length k at time t. Together, $\delta_{\frac{k}{l_{\max}}}$ tells us whether we have a bar of length k and $f_k(t)$ gives us the height of the bars that do exist. The measure $\mu_t^{l_{\max},M}$ adds up all of the heights and gives us the total number of particles of every length at time t

Finding $\mu_t^{l_{\max},M}$ can be tricky and multiple methods can be used to achieve it. We explored the possibility of using an infinitesimal generator.

Definition 4.1. *For a continuous time Markov process $\{X(t)\}_{t \geq 0}$ with state space S, the generator A is the operator that acts on a function f defined on S. For a state $x \in S$, A is defined as*

$$Af(x) = \lim_{t \to 0} \frac{\mathbb{E}[f(X(t))|X(0) = x] - f(x)}{t}.$$

More simply, $Af(x)$ represents the instantaneous rate of change of the expectation of f given that the process is currently at state x.

We can use this concept to find $\mu_t^{l_{\max},M}$, as shown in the following example. For this example, we go back to our original model where $\sigma = 0.9$, i.e. deletion and insertion can both occur. We do this in order to show a more complex example that can be easily simplified. Suppose that a continuous time Markov chain has a state space $\mathbb{S}$ and jumps from state i to state j with rate q_{ij} $\forall i, j \in \mathbb{S}$ where $i \neq j$. Then the generator $\mathscr{L}$ is an operator s.t. $\forall f : \mathbb{S} \to \mathbb{R}$,

$$\mathscr{L} f(i) = \sum_{j \in \mathbb{S}} (f(i) - f(j)) \cdot q_{ij}.$$

We apply this to the following state. Suppose we have one particle of length L that is a DIP. With probability σ, its offspring will have length $L+1$ and with probability $1-\sigma$, its offspring will have length $L - X$. Since the particle is already a DIP, its offspring will also be a DIP. Note that for any $x \in \{0, L\}$, $P(X = x) = \binom{L}{X} \cdot p^x \cdot (1-p)^{L-x}$. With death rate μ and birth rate $\frac{\lambda}{L}$, our generator is

$$\begin{aligned}\mathcal{L} f((L, \mathrm{DS})) =& (f(\emptyset) - f((L, \mathrm{DS})) \cdot \mu + f((L+1, \mathrm{DS})) \cdot \frac{\lambda}{L} \cdot \sigma \\ &+ \sum_{x=0}^{L} (f((L+x, \mathrm{DS})) - f((L, \mathrm{DS}))) \cdot \frac{\lambda}{L} \cdot (1-\sigma) \binom{L}{x} \cdot p^x \cdot (1-p)^{L-x}.\end{aligned}$$

This process can be done for every state and can give us $\mu_t^{l_{\max},M}$ $\forall t \geq 0$ and $\forall M, l_{\max} > 0$.

Recall that in section 4.1, the stochastic equation normalized by M converges to a simple ordinary differential equation. As mentioned in section 3, we predict that the distribution of lengths over time will converge to a deterministic equation as the initial number of particles and the initial length both increase. Connecting this prediction from the simulation to our equation in section 4.2, we think that $\lim_{t \to \infty} \mu_t^{l_{\max},M} = u(t, x)$, where $u(t, x)$ solves a partial differential equation such that x is the length distribution and t is time.

5. Future Direction

5.1. **Partial Differential Equation.** Recall that in section 4.1, we mention three steps that we need to take in order to be able to deterministically analyze the behavior of the virus. We found a stochastic equation that gives us the total number of viruses of each length at time t. We call this step (1) in section 4.1. Our next steps are to find a partial differential equation with solution $u(t, x)$, where t is time and x is the length distribution, and prove that $\mu_t^{l_{\max},M}$ converges to $u(t, x)$ as $M \to \infty$ and $l_{\max} \to \infty$. In the future, we plan use the following outline to accomplish these steps:

- Suppose that our initial condition $\mu_0^{M,l_{\max}}$ converges to $\phi(x)dx$ as $M, l_{\max} \to \infty$ for any density function $\phi(x)dx$.
- We want to show that for any time $t \in (0, \infty)$, $\mu_t^M \to u(t, x)$ as $M, l_{\max} \to \infty$, where $u(t, x)$ solves a partial differential equation with initial condition $\phi(x)$.

$u(t, x)$ is a deterministic equation, meaning there is no probability involved, so we will get the same result each time. Using $u(t, x)$ to analyze the within-host cell virus evolution is much more efficient than using the simulation, as we can make many observations about the behavior of the virus and find solutions for the equation that we can use to understand the virus better.

5.2. **Including DIPS and Improving our Model.**

We do not include the replication enzyme E in our stochastic model, as it gets complicated very quickly when it is included. In the future, we plan to modify our stochastic equation to take E into account. We also simplify the model by setting $\sigma = 0$ for our work in sections 3 and 4.2, so we plan to change the value of σ in the future, which will allow an offspring to be longer than its parent.

There are also other important genes that that allow virus particles to perform certain tasks. Particles that lack these genes are considered DIPs, so E is not the only nucleotide that controls whether a sequence is defective (DS) or functional (FS). These genes allow virus particles to exit the host cell (we call this the package gene), replicate on their own, and many other crucial tasks. In the future, we hope to incorporate these genes in our model.

Virus evolution is very complicated and is difficult to simulate realistically. Like any mathematical model, we chose multiple assumptions to use in our model and chose our parameters from past research. This model can continuously be changed and adapted to better model the replication process for an actual virus. Creating this model is an ongoing process and as we progress in our research and learn more about virus evolution, we hope to make our model as realistic as possible.

5.3. **Serial Passages Model.**

We focus specifically on within-host cell replication because it is the most complex and probabilistic component of a serial passages experiment. In the future, we plan to use our within-host cell model to simulate a full serial passages experiment. Past experimental research for plaque-to-plaque transfers, which is similar to serial passages experiments have resulted in what appears to be a stationary distribution for the number of virus particles [LEDM02]. Going forward, we will analyze the long-term behavior of a serial passages experiment using our model, which can possibly result in a stationary distribution as well.

6. Term Glossary

- Serial Passages: " In a serial-passage experiment, a cell culture or live host is inoculated by viral (or other) pathogens, usually already well adapted to different cell types or hosts. A pathogen's growth under the restrictive host environment leads to within-host selection for advantageous variants, either present as a minority in the founder population or generated from error-prone replications. After a certain amount of time (approximately days) of such growths, a small subset of the resulting pathogen population is sampled and used to inoculate a fresh new medium or host, initiating a subsequent round of the passage [WR14]."
- Wild Type (WT) Particles: "A phenotype, genotype, or gene that predominates in a natural population of organisms or strain of organisms (Merriam-Webster dictionary)." WT particles are not missing any important genes.
- RNA-dependent RNA polymerase (RdRP): The RdRP is an important enzyme that allows a virus particle to replicate on its own, without the presence of any other

virus particles. If a virus particle lacks the RdRP enzyme, it is considered a defective interfering particle [KSPS20].

- Defective Interfering Particles (DIPs): "Defective interfering particles are particles containing degenerate forms of the virus genomes that are non-replicative per se, but remain infectious by complementation with wild-type virus [RLV18]."
- Continuous Time Binary Branching Process: Each individual either gives birth to an individual with rate λ or dies with rate 1. The birth and death rates are modeled with exponential random variables with rates λ and 1, respectively. "If the rate λ exponential random variable happens before the rate 1 exponential, then the individual is replaced by two individuals. If the rate 1 exponential random variable happens before the rate λ exponential then the individual dies with no offspring. Every new individual gets two independent exponential random variables attached to it (one with rate λ and the other with rate 1) and so on [Sch14]."

7. Variable Glossary

- t: the time, where $t = 0$ is when the particle enters the host cell.
- T: the time at which the within-host cell experiment is stopped, which is chosen by the researcher.
- E: the replication enzyme.
- λ: the birth rate of a particle with length 1.
- $\frac{\lambda}{L}$: the birth rate of a particle with length L.
- T_{birth}: an exponential random variable with rate $\frac{\lambda}{L_i}$ describing the birth rate of one particle of length L_i.
- μ: the death rate of a particle.
- T_{death}: an exponential random variable with rate μ that describes the death rate of one particle.
- S: the state space of one virus particle.
- $\mathbb{S}$: the state space of many virus particles.
- σ: the probability that insertion will occur.
- Y: a Bernoulli random variable with rate σ describing the probability that either insertion or deletion will occur.
- p: the probability of deletion for each nucleotide.
- X: a binomial random variable with parameters L_i and p that describes the probability of $x \in [0, L_i]$ particles being deleted for a particle with length L_i.
- L: the number of nucleotides that make up a virus particle, which we call the length of the virus particle.
- L_0: the length of the initial particle(s).
- L_i: the length of the i-th particle born.
- l_{max}: the maximum possible length of a virus particle.
- M: the initial number of particles that enter one host cell.

Acknowledgements

I would like to thank Dr. Wai-Tong (Louis) Fan for providing the opportunity for me to be a part of this REU and being a great mentor along the way. Thank you to Daniel Rickert for working with me throughout the summer. I would also like to acknowledge and thank the Mathematics Department at Indiana University for hosting this program. I would like to thank NSF grant DMS-2051032 for making this program possible as well as everyone involved in this REU for providing such a positive research environment.

References

[KSPS20] Prashant Khare, Utkarsha Sahu, Satish Chandra Pandey, and Mukesh Samant, *Current approaches for target-specific drug discovery using natural compounds against sars-cov-2 infection*, Virus Research **290** (2020).

[Lam18] Amaury Lambert, *The coalescent of a sample from a binary branching process*, Theoretical population biology **122** (2018), 30–35.

[LEDM02] Ester Lázaro, Cristina Escarmís, Esteban Domingo, and Susanna C. Manrubia, *Modeling viral genome fitness evolution associated with serial bottleneck events: Evidence of stationary states of fitness*, Journal of Virology **76** (2002), no. 9, 4420–4428.

[LS13] Amaury Lambert and Tanja Stadler, *Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies*, Theoretical Population Biology **90** (2013), 113–128.

[RLV18] Veronica V. Rezelj, Laura I. Levi, and Marco Vignuzzi, *The defective component of viral populations*, Current Opinion in Virology **33** (2018), 74–80.

[Sch14] Rinaldo B. Schinazi, *Classical and spatial stochastic processes*, Birkhäuser, 2014.

[WR14] Hying Jun Woo and Jaques Reifman, *Quantitative modeling of virus evolutionary dynamics and adaptation in serial passages using empirically inferred fitness landscapes*, Journal of Virology (2014).

Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706 United States

Email address: ayusim@wisc.edu