Two Problems in Mathematical Biology

by

Hwai-Ray Tung

Department of Mathematics
Duke University

Date: _____

Approved:

_____
Richard Durrett, Dissertation Advisor

_____
Maria-Veronica Ciocanel

_____
James H Nolen

_____
Marc Ryser

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Mathematics
in the Graduate School of Duke University
2023

Abstract

Two Problems in Mathematical Biology

by

Hwai-Ray Tung

Department of Mathematics
Duke University

Date: _____

Approved:

_____
Richard Durrett, Dissertation Advisor

_____
Maria-Veronica Ciocanel

_____
James H Nolen

_____
Marc Ryser

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mathematics in the Graduate School of Duke University
2023

Abstract

This dissertation consists of two projects in mathematical biology. The first project studies tumor heterogeneity through the site frequency spectrum, the expected number of mutations with frequency greater than f. Recent work of Sottoriva, Graham, and collaborators have led to the controversial claim that exponentially growing tumors have a site frequency spectrum that follows the $1/f$ law consistent with neutral evolution. This conclusion has been criticized based on data quality issues, statistical considerations, and simulation results. Here, we use rigorous mathematical arguments to investigate the site frequency spectrum in the two-type model of clonal evolution. If the fitnesses of the two types are $\lambda_0 < \lambda_1$, then the site frequency spectrum is c/f^α where $\alpha = \lambda_0/\lambda_1$. This deviation from the $1/f$ law is due to the advantageous mutations that produce the founders of the type 1 population; mutations within the growing type 0 and type 1 populations still follow the $1/f$ law. Our results show that, in contrast to published criticisms, neutral evolution in an exponentially growing tumor can be distinguished from the two-type model using the site frequency spectrum.

The second project considers whether three species can coexist in a resource competition model with two seasons. Investigating how temporal variation in environment affects species coexistence has been of longstanding interest. The competitive exclusion principle states that n niches can support at most n species, but what constitutes a niche is not always clear. For example, Hutchinson in 1961 drew attention

to the diversity of phytoplankton coexisting despite the small number of resources in ocean water. Hutchinson then suggested that this could be explained by a changing environment; times when different species are favored would be considered different niches. In this paper, we examine a model where three species interact with each other solely through the consumption of one resource. The growth per resource rates, death rates, resource rates, and methods of resource consumption vary periodically through time. We give a necessary and sufficient condition for the coexistence of all three species. In particular, this condition rules out coexistence for the mean field limit of a three species two seasons model studied by Chan, Durrett, and Lanchier in 2009.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisor, Rick Durrett. It is thanks to Rick that I became a mathematical biologist, and his thoughtful guidance and support were invaluable through my graduate school experience. He is a great role model, and I wish him the best as he becomes an emeritus professor.

I thank my committee, Veronica Ciocanel, Jim Nolen, and Marc Ryser, for their advice and assistance throughout my time at Duke. I thank Veronica for her outstanding mentorship during the DOmath 2021 project. I also thank the Duke professors of practice, who encouraged and cultivated my teaching skills.

I thank the professors I've had as an undergraduate, especially Kavita Ramanan at Brown and Anne Shiu at Texas A&M for wonderful research experiences that led me to aim for a PhD. I also thank the many teachers in middle and high school, especially Mr. Kittredge, Mrs. Hower, and Mr. Thibodeaux, who helped foster a love of mathematics in me.

Finally, I thank my friends and family, who remind me of the existence of life beyond mathematics research.

1

Introduction

In this dissertation, we examine the frequency of mutations in cancer tumors in Chapter 2 and the preclusion of coexistence in a seasonal ecological setting in Chapter 3. Here, we give a brief summary of our results and relate them to previous work.

In Chapter 2, we take a theoretical perspective on the frequency of mutations in a two-type branching process and compare it to the frequencies expected from neutral evolution. Our work suggests that the mutation frequencies expected from neutral evolution are not as easy to come by from other models as some may suggest. This chapter is from Tung and Durrett (2021), which has been published in *PLOS Computational Biology*.

In Chapter 3, we examine a conjecture from Chan, Durrett, and Lanchier (2009) which claimed that it was possible for three species to coexist in a two season contact process model on a lattice with long range interactions. To do so, we take a generalized ODE model and show that there cannot be coexistence if the growth rates are (almost) linearly dependent. A corollary of this result is that three species cannot coexist in the two season model. This chapter is from Tung and Durrett (2022), which has been published in *Theoretical Populations Biology*.

1.1 Tumor Evolution and Heterogeneity

Traditionally, cancers were thought to evolve through a series of selective sweeps. Over enough time, a cell in the tumor acquires a mutations that makes it fitter than the others. This allows the cell to outcompete the others in the tumor, and allows the mutation to sweep through the population. This point of view was introduced in Nowell (1976) using leukemia as an example. As noted in Noble et al. (2022), leukemia is conducive to selective sweeps because of the lack of spatial effects in leukemia. This is in contrast to spatially structured cancers like colorectal cancer, where fitter cells have difficulty sweeping because of their glandular structure.

1.1.1 Big Bang

In contrast to selective sweeps, Sottoriva et al (2015) instead proposed the Big Bang model, where tumors have all the mutations needed for growth present at the beginning of the tumor's growth. A consequence of this model is that mutations are effectively neutral, otherwise we would see sweeps.

Following up on the introduction of the Big Bang model, Sottoriva and Graham (2015) looked at the site frequency spectrum (SFS), the number of mutations with frequency $\geqslant f$, of the big bang model and described what they called "a pan-cancer signature of neutral tumor evolution:" the SFS is proportional to $1/f$. The derivation of this result is remarkably simple. They assumed that cells grow exponentially at rate λ and use $N(t)$ to be the number of cells at time t. If we assume that the mutation rate is μ, then the expected number of new mutations before time t, $M(t)$, satisfies

$$\frac{dM}{dt} = \mu\lambda N(t).$$

Solving gives

$$M(t) = \mu\lambda \int_0^t N(s)\, ds.$$

Since $N(s) = e^{\lambda s}$, we observe that a mutation that occurs at time s will have frequency $e^{-\lambda s}$ in the population. Evaluating the integral in the previous formula, we have

$$M(t) = \mu(e^{\lambda t} - 1).$$

Ignoring the -1, if we set $t_f = -(1/\lambda)\log f$ to make $N(t_f) = 1/f$ so that mutations before time t_f will have frequency $\geqslant f$, then the number of mutations with frequency $\geqslant f$ is

$$M(t_f) = \mu/f$$

Note that in this derivation, mutations occur only at birth. If we instead let mutations happen continuously throughout a cell's lifetime and call the mutation rate ν, then, as shown in Durrett (2013),

$$M(t_f) = \frac{\nu}{\lambda f}. \tag{1.1}$$

From the derivation given above, we see that the $1/f$ site frequency spectrum comes from the fact that mutations occur at a rate proportional to the size of the population and the fact that the population is growing exponentially fast.

1.1.2 Evidence and Controversy over Big Bang Prevalence

Williams et al. (2016) found that 323 of 904 samples from 14 cancer types showed excellent straight line fits ($R^2 \geqslant 0.98$) when the cumulative number of mutations of frequency $\geqslant f$ is plotted versus $1/f$. This can be seen in Figure 2B in their paper. This paper has been cited 200 times, but among these works, there are a

number of papers criticizing the result. See Noorbakhsh and Chuang (2017), Wang et al (2017), and Bozic, Paterson, and Waclaw (2019). The December 2018 issue of Nature Genetics alone contains three letters (Balaparya and De 2018, Tarabichi et al 2018, McDonald, Chakrabarti, and Michor 2018) raising objections to the conclusion. Four common criticisms are

i Inferring the allele frequency f requires accurate estimates of local copy number and ploidy. In addition, Wang et al (2017) point out that local samples may not be indicative of overall frequencies.

ii Failure to reject the null model is not the same as proving it is true. To quote McDonald, Chakrabarti, and Michor (2018) "The fact that a model of neutral evolution leads to a linear relationship between $M(f)$ (the number of mutations with frequency $\geqslant f$) and $1/f$ does not imply ... the presence of neutral evolution."

iii Tarabichi et al (2018) applied methods that look at the dN/dS ratio, which compares the number of nonsynonymous and synonymous mutations, to look for signs of selection. They claim to have found significant signs of selection in tumors that were classified as neutral. However when the analysis was repeated on publicly available pancreatic cancer data, Graham, Sottoriva et al found no values significantly different from 1.

iv Tarabichi et al (2018) say "the deterministic models of tumor growth described by Williams et al (2016) rely on strong biological assumptions. Using branching processes to simulate neutral and nonneutral growth, they show that $R^2 > 0.98$ is neither necessary nor sufficient for neutral evolution."

In this thesis we address the fourth point by examining the site frequency spectrum of a two type branching process. More specifically, the model features two cell types, type 0 and type 1. The process starts with type 0 cells reproducing at rate λ_0.

Over time, type 0 cells undergo type $1A$ mutations and generate type 1 cells, which reproduce at rate $\lambda_1 > \lambda_0$. Type 1 cells with the same type $1A$ mutation belong to the same family. Both type 0 and type 1 cells undergo neutral mutations, which we call type 0 and type 1 mutations. As time approaches infinity, the type 0 cells approach 0% of the population and the type $1A$ families partition the type 1 population. As such, the SFS consists of three mutation types - type $1A$ mutations, type 1 neutral mutations that occur in the type 1 population and are constrained to a $1A$ family, and type 0 neutral mutations that occur in the type 0 population and survive by piggy-backing off a type $1A$ mutation.

1.1.3 Durrett and Moseley (2010)

This is not the first work to look at the two-type branching process. Durrett and Moseley (2010) look at the same process without neutral mutations.

There are two results from Durrett and Moseley (2010) that we will later use. The first is that the first $1A$ mutation will happen around the same time. The second is that the relative sizes of the families can be described as the points in a poisson point process with mean measure $\rho(x, \infty) \sim x^{-\alpha}$, where $\alpha = \lambda_1/\lambda_0$. As will be discussed in Chapter 2, this second fact implies that $1A$ families follow the Poisson-Dirichlet distribution.

1.1.4 Poisson-Dirichlet Distribution

The Poisson-Dirichlet distribution, also written as $PD(\alpha, \theta)$, generates random partitions with total length 1, and has size biased order

$$(W_1, \overline{W_1}W_2, \overline{W_1 W_2}W_3, ...)$$

where $W_i \sim Beta(1-\alpha, \theta + i\alpha)$ are independent and $\overline{W_i} = 1 - W_i$.

One way to generate $PD(\alpha, \theta)$ is to use a generalized Chinese restaurant process.

We start with one person at one table. If there are n seated individuals and k tables, the next person to enter is seated at a new table with probability $(\theta + k\alpha)/(n+\theta)$ or seated at existing table i with probability $(n_i - \alpha)/(n + \theta)$, where n_i is the number of people at table i. $PD(\alpha, \theta)$ is formed by the proportion of people sitting at each table as the number of people approaches infinity.

To show that the two definitions agree, note that if we focus on the proportion of people at table 1 and group tables 2 through infinity under one table, we end up with a process where table 1 attracts people proportional to $n_1 - \alpha$ and table 2 attracts people proportional to $n_2+\alpha+\theta$. This is equivalent to the Polya Urn with initial state $1-\alpha$ balls of one color and $\theta+\alpha$ of the other color. Since the proportion of balls of one color in the Polya Urn model is well known to have distribution $Beta(1-\alpha, \alpha+\theta)$, the proportion of people at table 1 is W_1. Similarly, by conditioning on tables 1 through $i-1$, combining tables $i+1$ onward, and comparing table i with $i+1$, we use Polya Urn again to show that the proportion of people at table i is $\overline{W_1 W_2} \cdots \overline{W_{i-1}} W_i$.

The PD distribution has also been constructed using branching processes. For example, in the case that $0 \leqslant \alpha \leqslant 1$ and $\theta = 0$, consider two individual types, novel and clone. Novel individuals produce novel offspring at rate α and clonal offspring at rate $1-\alpha$. Clones produce clones at rate 1. In this model, a new novel offspring is equivalent to a person being seated at a new table, and clone offspring are people who sit at an existing table. Comparing the probabilities of whether the next offspring is novel or belongs to the family tree of novel offspring i show that this is equivalent to the generalized Chinese restaurant process.

For additional information, see Jim Pitman's book *Combinatorial Stochastic Processes*.

1.2 The Competitive Exclusion Principle

The competitive exclusion principle (CEP), sometimes called Gause's principle, states that n niches can support at most n species. In the case of Gause (1932)'s experiments with two species of Paramecium, the one niche present was clear; the species that better utilized the food Gause gave them drove the others to extinction. However, it is not always evident what can be considered a niche. George Evelyn Hutchinson (1961) drew attention to this with the "Paradox of the Plankton" - although there are at most 20 resources relevant to the growth of phytoplankton, there are hundreds of plankton species coexisting in ocean water. Hutchinson's observation ignited interest in a mathematical approach to when CEP holds.

One of the earliest mathematical models used to justify the CEP was Volterra (1928)'s model, which featured n species with populations x_i interacting through competition over a resource R.

$$\frac{1}{x_i}\frac{dx_i}{dt} = \gamma_i R - \sigma_i$$

$$R = R_{max} - F(x_1, x_2, ..., x_n)$$

There is a maximum amount of resource R_{max}. Each species grows at a rate linearly dependent on resource R and dies at rate σ_i. The amount of resource in use $F(x_1, x_2, ..., x_n)$ is a nondecreasing function of the species abundances. Volterra showed that the species with the largest $R_{max} - \sigma_i/\gamma_i$ value wins.

1.2.1 Volterra's Assumptions

In light of the "Paradox of the Plankton," many have tried to understand what assumptions in Volterra's model eliminated coexistence. Armstrong and McGehee (1980) considered many assumptions, listed below

i The dynamics can be described from species densities.

ii Species interact only through the species

iii The system is spatially homogeneous

iv The resource is uniform in quality

v The growth rates depend linearly on the quantity of resources.

vi There is no time dependence in interactions.

When one of these assumptions is violated, the competitive exclusion principle may no longer hold. (i) rules out age structure complications. (ii) rules out the possibility of predation, where one species can count as a resource for another species, and symbiosis. (iii) prevents coexistence in a patch model where the patches favor different species. (iv) prevents different resources being more useful to different species. (v) is less straightforward. Suppose dn_1/dt is a concave increasing function with respect to R and is 0 when $R = R_1$. Also suppose dn_2/dt is a linear increasing function with respect to R and is 0 when $R = R_2 > R_1$. When species 2 is at equilibrium, since $R_2 > R_1$, species 1 has a positive growth rate and won't go extinct. When species 1 is at equilibrium, the average amount of resources available has to be greater than R_1 due to Jensen's inequality. If this average is greater than or equal to R_2, then species 2 will have enough resources to grow. (vi) rules out how seasons can create additional niches. See Hening and Nguyen (2020) for the relevance of the assumptions under related SDE and piecewise deterministic Markov process models.

1.2.2 Temporal Heterogeneity

The assumption that we focus on in this paper is the one Hutchinson proposed to resolve the Paradox of the Plankton - temporal variation. By allowing the functions

γ_i, σ_i and R to be functions of time, different species can be favored at different times, which enables coexistence. Since many environments are periodic, we focus on when the functions have period T.

To demonstrate how temporal heterogeneity could encourage coexistence, Armstrong and McGehee (1976) considered a simple n season system where n species could survive on one resource. We define a season as an interval of time under which the parameters do not exhibit explicit time dependence. The system is

$$\frac{1}{n_i}\frac{dn_i}{dt} = \gamma_i R g_i(t) - \sigma_i, \qquad R = R_{max} - \sum_{i=1}^{k} s_i n_i$$

where R represents available resource, k is the number of species, and $g_i(t)$ is a function of period T that is equal to 1 on the interval $[a_i, b_i]$ and is equal to 0 on the intervals $[0, a_i]$ and $[b_i, T]$. When $g_i(t) = 1$ and there is no temporal heterogeneity, we recover Volterra (1928)'s model, one of the earliest models for justifying the competitive exclusion principle; the species with the highest $R_{max} - \sigma_i/\gamma_i$ wins. When there is temporal heterogeneity, coexistence becomes possible. Intuitively, $g_i(t)$ indicates whether species i is in a growing season or declining season. By having disjoint growing seasons, one species would quickly grow while the others would quickly shrink, preventing them from effectively competing with the currently growing species. Armstrong and McGehee then constructively proved that in their model, parameters could be found that allowed n species to coexist given n seasons. Coexistence here is an example of the storage effect proposed in Chesson (1994), which outlines how species-specific responses to the environment, covariance between environment and competition, and buffered population growth can contribute to coexistence. The name of the storage effect comes from how "storing" more benefits of advantageous times than is "spent" during disadvantageous times can enable coexistence.

1.2.3 Lotka-Volterra

The two-species system in CDL is a special case of the two-species periodic Lotka-Volterra model whose population sizes n_1 and n_2 are described by

$$\frac{1}{n_1}\frac{dn_1}{dt} = b_1(t) - a_{11}(t)n_1 - a_{12}(t)n_2$$

$$\frac{1}{n_2}\frac{dn_2}{dt} = b_2(t) - a_{21}(t)n_1 - a_{22}(t)n_2$$

where $b_i(t)$ and $a_{ij}(t)$ are periodic functions with period T. In the case that $a_{2i} = ka_{1i}$, the Lotka-Volterra model can be written as a periodic version of Volterra (1928)'s model. Cushing (1980) studied the stability of periodic solutions by generalizing the bifurcation diagrams for the constant coefficient Lotka-Volterra model, and gave an example of when there is coexistence in the periodic Lotka-Volterra model, but one of the species goes extinct when temporal variation is removed by replacing the periodic parameters with their average. Mottoni and Schiaffino (1981) study the same model using a geometric approach and, in addition to recovering some of Cushing's results, also prove that any solution approaches a solution with period T.

1.2.4 Chan, Durrett, and Lanchier (2009)

Chan, Durrett, and Lanchier (2009) considered a two season two-type contact process on a square lattice with long range interaction. In more detail, the model takes place on the grid $\mathbb{Z}^2/L$ where L is large. There are two species, 1 and 2. If species i occupies site x, then it dies at rate σ_i. If site x is empty, it is populated by species i at rate $\gamma_i(t)f_i$ where the growth rate for species i, γ_i, changes based on which of the two seasons it is, and f_i denotes the fraction of sites within distance 1 of x that is occupied by species i. The seasons both have length D and alternate. The mean

field model of their system is

$$\frac{1}{n_i}\frac{dn_i}{dt} = \gamma_i(t)R - \sigma_i \qquad R = 1 - \sum_{i=1}^{k} n_i \tag{1.2}$$

where R represents available space and k is the number of species.

There is one resource R so in the temporally homogeneous case one species will competitively exclude the others. Chan, Durrett, and Lanchier showed that for an open set of parameters, two species can coexist in a model with two seasons. The ecological explanation is that the two seasons form two niches. Defining $\overline{n_i}$ as the equilibrium solution for species i when the other species is absent, the condition for coexistence is

$$\frac{1}{2D}\int_0^{2D} \gamma_1(t)(1-\overline{n_2})dt > \sigma_1, \qquad \frac{1}{2D}\int_0^{2D} \gamma_2(t)(1-\overline{n_1})dt > \sigma_2$$

which follows from the idea of invasion. Species 1 is able to invade species 2 if the population of species 1 increases when species 1 has a near zero population and species 2 is at equilibrium. If species 1 can invade species 2, then even if species 1 approached extinction, species 2 will approach its equilibrium and the population of species 1 will rebound. The conditions can therefore be understood as whether the species can invade each other; the first integral gives the average growth rate of species 1 when species 2 is at equilibrium, and the condition checks if it is greater than the death rate σ_1.

Lastly, they conjectured that a fast dispersing species could exploit the early part of a season before losing to a superior competitor, allowing for three or more species to coexist. This is backed with simulation evidence in Figure 1 of their paper. Here, we will prove that this is not possible in the ODE.

Table 1.1: **Parameters from Figure 1 in Chan, Durrett, Lanchier (2009).** In their simulation used to support their conjecture that three species can coexist with two seasons, they looked at a system on a 400×400 lattice with interaction range $L = 200$, season lengths $D = 10$, and death rates $\sigma = 1$. The growth rates γ in each season are outlined below.

Species	Season 1 γ	Season 2 γ
1	3	1
2	1	3
3	2	2

2

A Two Type Branching Process Model of Tumor Heterogeneity

This chapter is from Tung and Durrett (2021), which has been published in *PLOS Computational Biology*. Following up on the introduction of the Big Bang model by Sottoriva et al (2015), Sottoriva and Graham (2015) described what they called "a pan-cancer signature of neutral tumor evolution:" the number of mutations with frequency $\geqslant f$ will have the form c/f. The derivation of this result is remarkably simple and is given in Section 2.4 in Methods. Williams et al. (2016) found that 323 of 904 samples from 14 cancer types showed excellent straight line fits when the cumulative number of mutations of frequency $\geqslant f$ is plotted versus $1/f$. See Figure 2B in their paper. This paper has been cited 200 times, but among these works, there are a number of papers criticizing the result - see Section 1.1.2. The criticism we focus on is the claim that other nonneutral models of evolution pass the test proposed by Williams et al.

To try to shed some light on the controversy, we will do a mathematically rigorous computation of the site frequency spectrum produced by the two-type model of clonal

evolution. We will describe the model in Section 2.1. The two-type model and its m-type generalization have been extensively studied. See Durrett (2015) for results and references. This model is relevant to the discussion of Williams et al (2016) because it appears in the criticisms of McDonald, Chakrabarti, and Michor (2018) and Bozic, Patterson, and Waclaw (2019). Before we describe the mathematical analysis, we want to make it clear that that this work only discusses the theoretical aspects of cancer genomics and is not concerned with practical problems in making inferences on cancer genomic data, which of course could hide some of the theoretical effects due to errors, bias, sampling, and other issues discussed in the criticisms listed above.

Results

2.1 A two-type model

McDonald, Chakrabarti, and Michor (2018) consider two alternative evolutionary models in order to argue that other underlying models can produce a linear relationship between $1/f$ and the cumulative number of mutations with frequency $\geqslant f$. Their second model is an infinite alleles branching process model previously studied by McDonald and Kimmel (2015). We will ignore this model, since in studying DNA sequence data the appropriate mutation scheme is the infinite sites model.

In their first model, clonal expansion begins with a single cell of the original tumor-initiating type (type 0). To make it easier to connect with previous mathematical work, we will describe their model using the notation used in Durrett (2015) and Durrett (2013). We suppose that type 0 individuals give birth at rate a_0 and die at rate b_0, so the exponential growth rate is $\lambda_0 = a_0 - b_0$. For simplicity, we will suppose that neutral mutations accumulate during the individual's life time at rate ν, instead of only at birth.

Type 0 individuals mutate to type 1 at rate u_1. Type 1 individuals give birth at rate a_1 and die at rate b_1. Their exponential growth rate is $\lambda_1 = a_1 - b_1$ where $\lambda_1 > \lambda_0$. In McDonald, Chakrabarti, and Michor (2018), different type 1 families have different increases in their growth rates that follow a normal distribution. In this section, we will assume all type 1 mutations have the same growth rate. In Section 2.2, we will consider the implications of random fitness changes for the behavior of the model.

The reader will see many complicated formulas in this paper, so it will be useful to have a concrete set of parameters to plug into these formulas. Borrowing an example from Durrett (2015), we will set

$$a_0 = a_1 = 1, \quad \lambda_0 = 0.02, \quad \lambda_1 = .04, \quad u_1 = 10^{-6}, \quad \nu = 10^{-4}. \tag{2.1}$$

We do not pretend that these parameters apply to any specific cancer, but for motivation, the reader can imagine that type 0s are colon cancer cells in which both copies of APC have been knocked out, while type 1 cells in addition have a KRAS mutation.

2.1.1 Limit theorems

As in McDonald, Chakrabarti, and Michor (2018), we will, for simplicity, restrict our attention to two types of cells. The type 0's are a simple branching process, so well-known results show that $Z_0(t)$, the population of type 0 cells at time t, follows

$$e^{-\lambda_0 t} Z_0(t) \to W_0, \tag{2.2}$$

where $W_0 = 0$ with probability b_0/a_0 and has a rate λ_0/a_0 exponential distribution with probability λ_0/a_0. For a derivation, see Athreya and Ney 1972.

The study of the second wave is simpler if we suppose that $Z_0^*(t) = V_0 e^{\lambda_0 t}$ for all $t \in (-\infty, \infty)$, where V_0 has the same distribution as $(W_0 | W_0 > 0)$, that is exponential

with rate λ_0/a_0. Mutations from type 0 to 1 occur at rate u_1. Let σ_1 be the time of the first successful type 1 mutation, i.e., one whose branching process does not die out. Durrett and Moseley (2010) showed, see (29) in Durrett (2015), that σ_1 has median

$$s^1_{1/2} = \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2 a_1}{a_0 u_1 \lambda_1}\right). \tag{2.3}$$

In the concrete example, $s^1_{1/2} = 460.51$. In colon cancer where cells divide every four days, $s^1_{1/2}$ is 1842 days or a little more than 5 years.

Durrett and Moseley were the first to rigorously prove results about the asymptotic behavior of the size of the type 1 population $Z^*_1(t)$, see Section 9 of Durrett (2015). Durrett (2013) noticed that the constants are simpler if we use a different normalization. Here we are assuming $a_0 = a_1 = 1$ to simplify the constants.

Theorem 1 (Durrett and Moseley 2010)**.** *As $t \to \infty$, $e^{-\lambda_1(t-s^1_{1/2})} Z^*_1(t) \to \bar{V}_1$ where $\bar{V}_1 = e^{\lambda_1 s^1_{1/2}} V_1$ is the sum of the points in a Poisson process with mean measure*

$$\bar{\rho}(x,\infty) = \rho(e^{-\lambda_1 s^1_{1/2}} x, \infty).$$

Using Eq (2.3), and doing some algebra

$$\bar{\rho}(x,\infty) = \alpha \lambda_0 \lambda_1^{-\alpha} \Gamma(\alpha) V_0 x^{-\alpha}.$$

In our concrete example, $\bar{\rho}(x,\infty) = 0.1772 V_0 x^{-1/2}$. Note that due to shifting time by $s^1_{1/2}$, the measure $\bar{\rho}$ does not depend on the mutation rate.

2.1.2 Site frequency spectrum

There are three classes of mutations in the two-phase model

- type 0: Neutral mutations that occur to type 0 individuals.
- type 1A: Advantageous mutations that turn type 0 individuals into type 1.

- type 1: Neutral mutations that occur to type 1 individuals.

By the argument in Section 2.4 given by Sottoriva and Graham (2015), the type 0 mutations will have a $1/f$ site frequency. The argument can also be used to prove the next result so the details are contained in Section 2.5 in Methods.

Theorem 2. *The number of type 1 mutations with frequency $\geqslant f$ with in the type 1 population will be asymptotically $\nu/(\lambda_1 f)$.*

The points in the Poisson process in Theorem 1 indicate the contributions of the various type one families to the limit $\bar{V}_1$, so if we let $x_1 > x_2 > x_3 \ldots$ be the points, then the jth largest family makes up a fraction $x_j/\bar{V}_1$ of the population. Intuitively, this implies that the number of type 1A mutations with frequency $\geqslant f$ will be asymptotically $Cf^{-\alpha}$ where $\alpha = \lambda_0/\lambda_1$. This matches with our result, whose proof is in Section 2.6 in Methods.

Theorem 3. *The site frequency spectrum of the 1A mutations is*

$$SFS_{1A}(f) = \frac{\sin(\pi\alpha)}{\pi\alpha}\left(\frac{1}{f} - 1\right)^{\alpha}. \tag{2.4}$$

When $\alpha = 1/2$, the constant is $2/\pi = 0.6366$.

Including type 0 passenger mutations in type $1A$ families does not significantly change the $f^{-\alpha}$ shape in (2.4). This is because all important 1A mutations happen soon after the first mutation, which implies that all important 1A mutations have roughly the same number of passengers. See Section 2.7 in Methods.

To illustrate the results proved above, we turn to simulations seen in Figs 2.1 and 2.2.

2.2 Random fitness increases

McDonald, Chakrabarti, and Michor (2018) considered the case in which type 1 individuals have growth rates that are normal with mean m and standard deviation

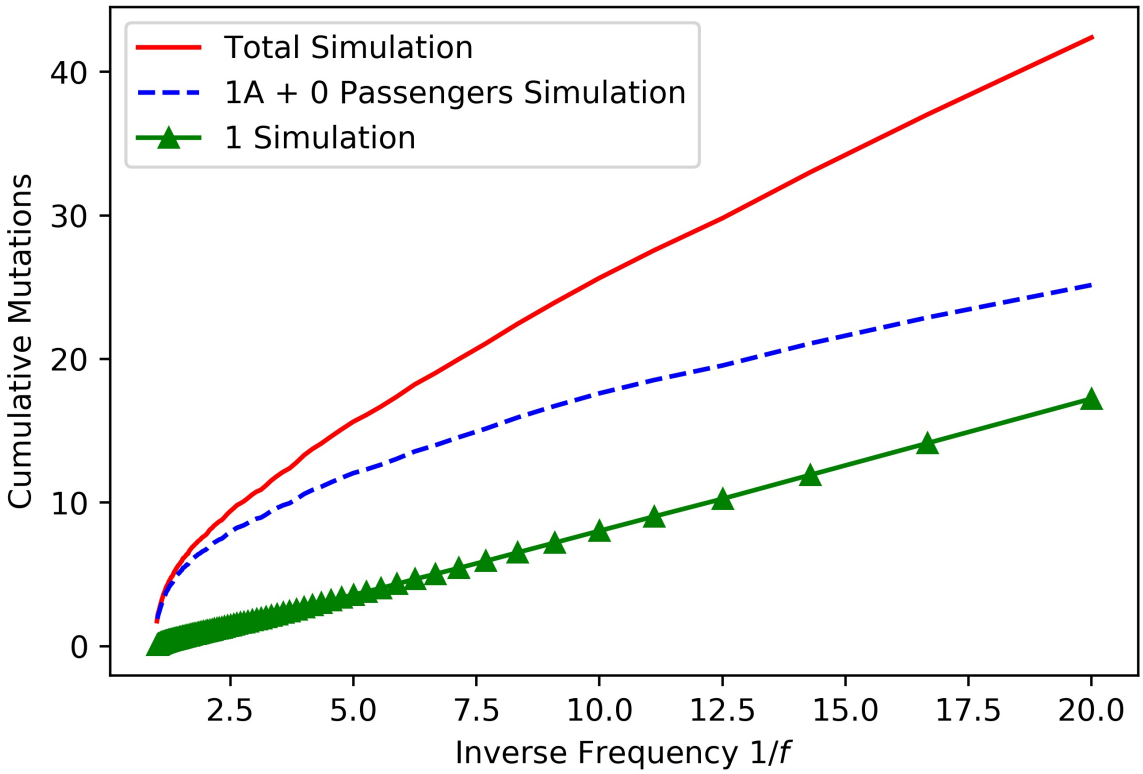

FIGURE 2.1: **Site frequency spectrum in the type 1 population.** The figure shows the contribution of the different mutation types to the site frequency spectrum. The simulation was performed with parameters $\nu = 0.02, u_1 = 2 \times 10^{-4}, \lambda_0 = 0.02, \lambda_1 = 0.04$ and $a_0 = a_1 = 1$ and is the average site frequency spectrum of 1000 runs. We simulated the $1A$ families and type 0 passenger mutations on their founders. Then, we obtained type 1 mutations for each $1A$ family by applying (2.8) in Methods. We only consider mutations present in the type 1 population because, as $t \to \infty$, the proportion of the population that is type 0 cells approaches 0. As suggested from Theorem 2, the type 1 site frequency spectrum is linear when plotted against $1/f$. The $1A + 0$ line looks similar to a power law, as suggested by (2.4).

d. Early work on models with random fitness increases in the two-type model led to very unusual behavior in the limit $t \to \infty$, see Durrett et al (2010). Results in that paper show

- If the fitness distribution was bounded then, as $t \to \infty$, individuals with fitnesses that were close to the upper limit dominated the population .
- If the distribution was unbounded, then the population could grow faster than

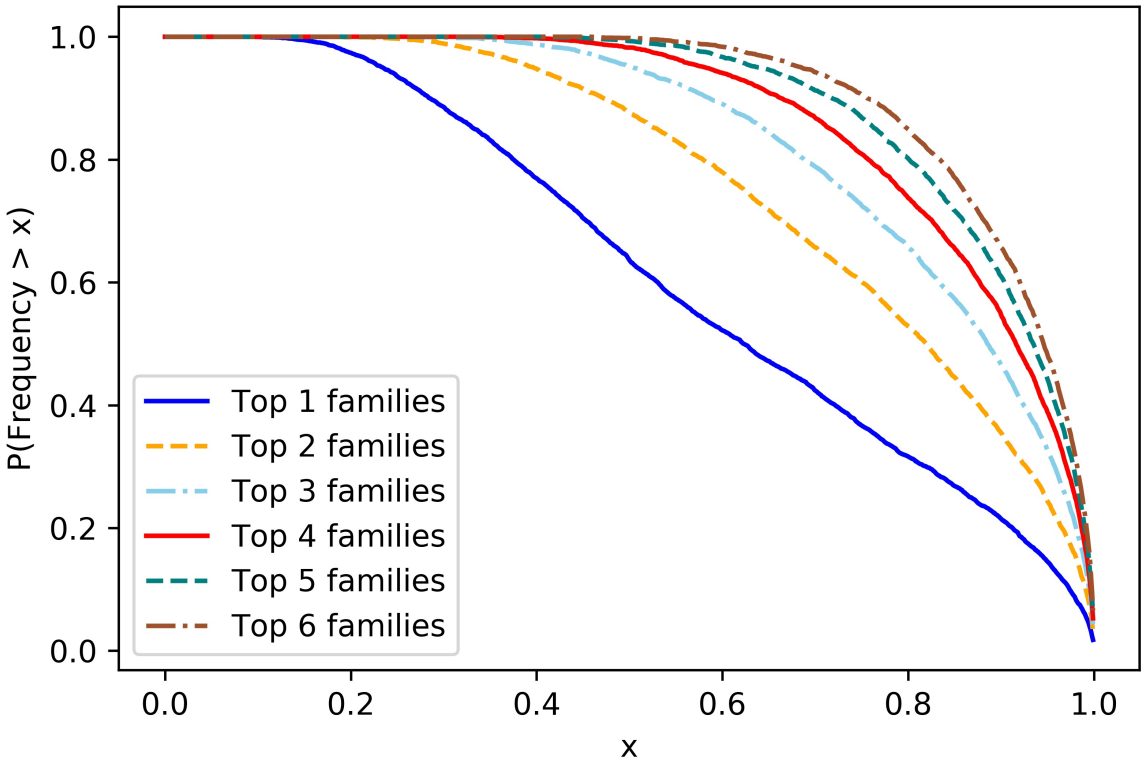

FIGURE 2.2: **Distribution of 1A family sizes in the type 1 population.** To better understand the distribution of $1A$ family sizes, we used the Poisson-Dirichlet$(\alpha, 0)$ distribution to generate the six largest families. The plot gives the probability that the number of individuals in the top i families are greater than a fraction x of the total type 1 population.

exponential.

In this section, we will modify our example from Figure 2.1 so that type 1 individuals have growth rates drawn from the normal distribution with mean $m = 0.04$ and standard deviation $d = 0.005$. We will see through simulations that in contrast to the limiting results just mentioned, random fitnesses do not substantially change the behavior.

To find the distribution of the growth rates of the mutations with the largest family sizes, we note that a mutant that occurs at time s_i and has growth rate $\lambda_{1,i}$ will grow to size $W_1 \exp(\lambda_{1,i}(1000 - s_i))$ at time 1000. The number of i that are

successful and have $\lambda_{1,i}(1000 - s_i) > x$ is Poisson with mean given by the following integral

$$10^{-6} \int_0^{1000} 50e^{0.02s} \int_{x/(1000-s)}^{\infty} \lambda\phi(\lambda)\, d\lambda ds$$

$$= 10^{-6} \int_0^{1000} 50e^{0.02s} \left[0.04 \left(1 - \Phi\left(\frac{x}{1000-s} \right) \right) + 0.005^2 \phi\left(\frac{x}{1000-s} \right) \right] ds. \quad (2.5)$$

where ϕ and Φ are the density function and distribution function, of a normal distribution with mean $m = 0.04$ and standard deviation $d = 0.005$. The first expression can be understood through the following pieces: 10^{-6} is the type $1A$ mutation rate, $50e^{0.02s}$ is the expected type 0 population at time s, the λ is the probability a type $1A$ mutation with fitness λ survives, and $\phi(\lambda)$ is the density function for the fitness. The equality follows from substituting $u = (\lambda - 0.04)^2$ for the inner integral. Figure 2.3 graphs (2.5).

The random fitnesses cause the relative sizes of the contributions of mutations to the final population to change, but as Figure 2.4 shows, the site frequency still has the form C/f^{β}, where $\beta \leqslant \alpha$ and achieves equality in the case of non-random changes, i.e. $d = 0$.

McDonald, Chakrabarti, and Michor (2018) claim that the site frequency spectrum in the two-type model is $1/f$. However, their simulation methods take the very crude approach of considering the binary split process until 1,000 or 1,000,000 cells are produced. This corresponds to 10 and 20 generations respectively. To make it possible for something to happen in this short amount of time the mutation rate for advantageous mutations is set to be 0.1 in the 1000 cell scenario, and to 0.03 when there are 1,000,000 cells. At birth, each cell acquires a Poisson mean 100 number of mutations. In contrast our simulations run for approximately 1000 generations, leading to populations of order 10^9 cells, and neutral mutations occur slowly, leading

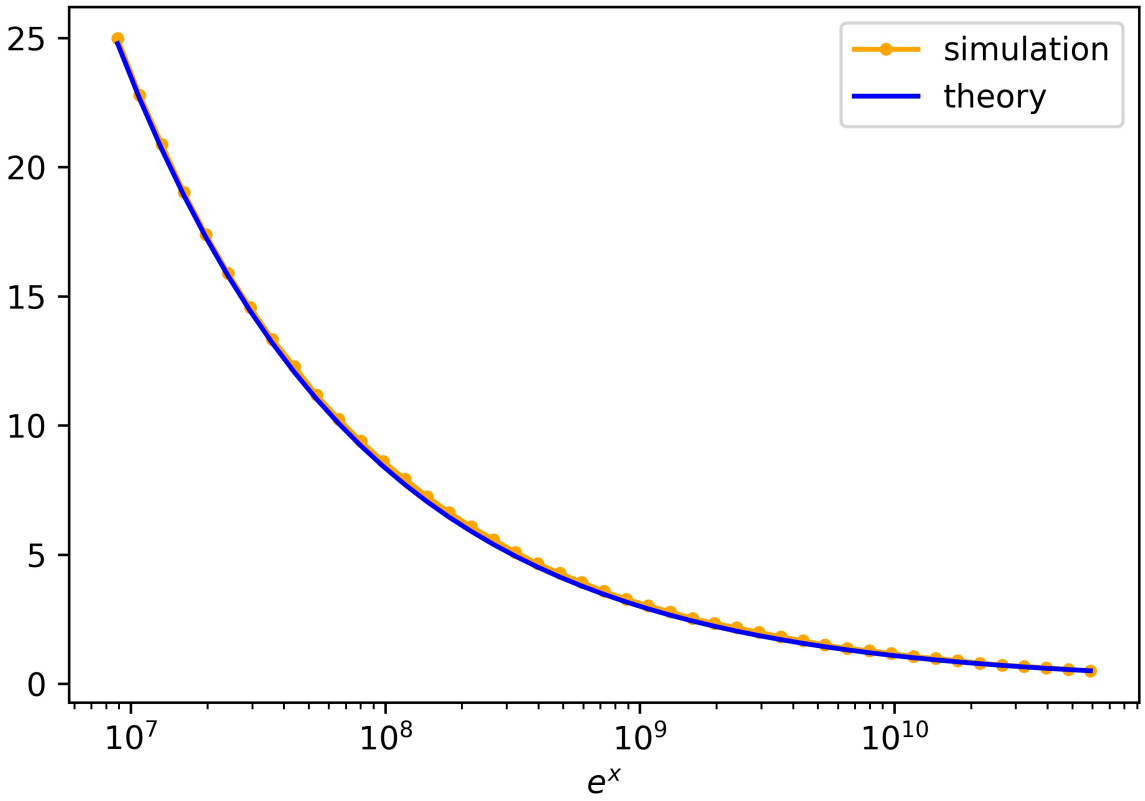

FIGURE 2.3: **Size of 1A families with random fitness.** The graph indicates the expected number of $1A$ families with $\lambda_{1,i}(1000 - s_i) > x$. The parameters are almost the same as in (2.1); rather than a single λ_1 for all type 1 families, we have a different $\lambda_{1,i}$ for each type 1A family. Each $\lambda_{1,i}$ is normally distributed with mean 0.04 and standard deviation 0.005. 500 runs were done up until time $t = 1000$. The graph shows that on average there is one family with $e^x > 10^{10}$. If the $\lambda_{1,i}$ of the largest family is within 2 standard deviations, then multiplying e^x by $1/\lambda_{1,i}$ implies a family of magnitude around 2×10^{11} or greater.

to genealogical relationships that are more like those found in growing cancer tumors.

2.3 Subclonal mutation frequencies

Bozic, Paterson, and Waclaw (2019) argue that "the fact that no subclonal driver is present at intermediate frequencies cannot be taken as proof of neutral or *effectively neutral* evolution. It can be a consequence of population dynamics which create only a short window during which the driver mutation can be detected but not fixed in the

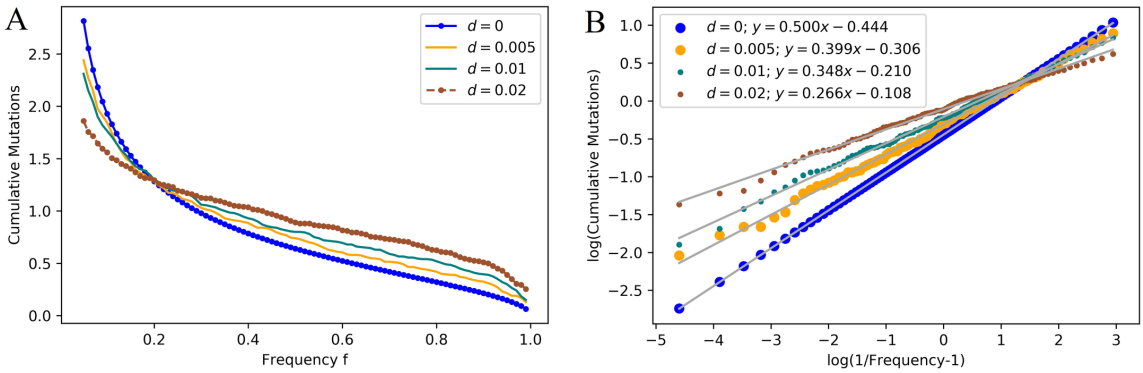

FIGURE 2.4: **Site frequency spectrum with random fitnesses.** (A) shows the site frequency spectrum for multiple values of d. The other parameters are the same as in Figure 2.3. As the contribution from neutral mutations is negligible, we will only show the contribution from $1A$ families. The line for constant, i.e., $d = 0$, is plotted from theory; the others are plotted from simulations with 200 runs. As d increases, the expected size of the frequency of the largest mutation increases. Also, fewer mutations reach above the 0.05 frequency threshold. (B) displays the same data on a log-log plot. The slopes β of the linear fits indicate that the site frequency spectrum takes the form C/f^{β}, with β decreasing as d increases.

population." In this section we will describe their results and give a simple analytic derivation.

To argue for this viewpoint, they use the two-phase model introduced in the Section 2.1 but with different notation

Table 2.1: **Notation changes between here and Bozic, Paterson, and Waclaw (2019)**

here	a_0	b_0	λ_0	a_1	b_1	λ_1	u_1
Bozic, Paterson, and Waclaw (2019)	b	d	r	b_1	d_1	r_1	u

In addition they define $c = r_1/r > 1$, and $g = c - 1$. They assume that the mutation to type 1 occurs at time 0 and run the process until the time t at which the total population size is M. Let X_0 be the population of type 0's when the mutation occurs. Since X_0 is large, $X_t \approx X_0 e^{rt}$. The type 1 population at time t is $Y_t \approx W_1 e^{rct}$, where W_1 is an exponentially distributed random variable with rate cr/b_1. Note that as

in Bozic et al (2010) the possibility of subsequent driver mutations is ignored. As Figure 2.5 shows, that change does not lead to a substantial error.

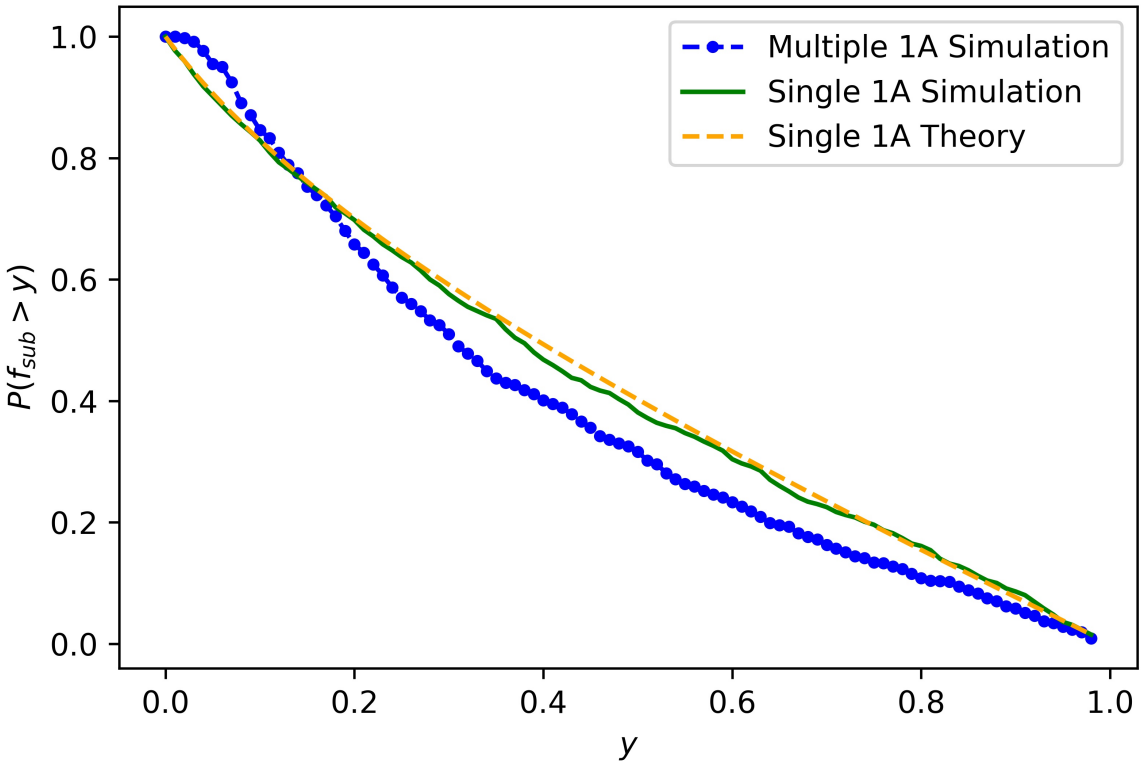

FIGURE 2.5: **Driver frequencies.** This graph gives the probability of having a driver with frequency greater than y once the tumor reaches size 10^9. The parameters used are $a_0 = a_1 = 1, \lambda_0 = 0.02, \lambda_1 = .035$ and $u_1 = 10^{-5}$ and the data was generated from 1000 runs. Single 1A refers to approach taken by Bozic et al. where there is only 1 selective mutation. Multiple 1A is our approach. The theory curve comes using a Riemann sum with interval size 500 to evaluate the integral in Eq (2.6).

Writing $f_{sub} = Y_t/(X_t + Y_t)$ they prove that when the total tumor size is $M = X_t + Y_t$ the subclonal mutation frequency has

$$P(f_{sub} \leqslant y) = \int_0^M (uc/b_1) \exp(-ucx_0/b_1) \left[1 - \exp\left(-\frac{cr}{b_1}\frac{y}{(1-y)^c} x_0^c M^{1-c}\right)\right] dx_0, \tag{2.6}$$

which is (1) in Bozic, Paterson, and Waclaw (2019). From this they can compute the probability of a subclonal driver being detectable, that is, $P(0.2 \leqslant f_{sub} \leqslant 0.8)$

To see what this complicated formula implies, the authors turn to simulation. The mutation rate to produce an additional driver is $u = 10^{-5}$. Their Figure 2A shows a moderately growing tumor $b = 0.14$, $r = 0.01$, 2B a fast growing tumor $b = 0.25$, $r = 0.07$, and 2C a slowly growing tumor $b = 0.33$, $r = 0.0013$. For moderate values of selection, e.g. $g = 30\%$, the probability that a driver mutation is in the detectable range $[0.2, 0.8]$ is $< 15\%$ for population sizes up to $M = 10^9$ cells and remain below $1/3$ for $M \leqslant 10^{11}$. For other cases considered there ($g = 70\%$ and 100%) the chance of detecting the subclonal driver is always $< 60\%$ and for a broad range of sizes is less than 30%. Panels d,e,f in their Figure 2 show the frequency of a subclonal driver in the case of moderate growth when the size $M_d = 10^7$, $M_e = 5 \cdot 10^{10}$ and $M_f = 2 \cdot 10^8$. In the three cases the frequency is near 0, near 1, and almost uniformly distributed on $[0, 1]$.

Rather than study the tumor when it reaches a fixed size, we will derive results at a fixed time by using Theorem 1. Recall that we have set $Z_0^*(t) = V_0 e^{\lambda_0 t}$ and have shown

$$e^{-\lambda_1 (t - s^1_{1/2})} Z_1^*(t) \to \bar{V}_1.$$

Combining the last two results, we see that

$$r(t) = \frac{Z_1^*(t)}{Z_0^*(t)} \approx e^{-\lambda_0 t} e^{\lambda_1 (t - s^1_{1/2})} \bar{V}_1 / V_0.$$

Inserting the values of the λ_i

$$\frac{r(t+s)}{r(t)} = e^{(\lambda_1 - \lambda_0)s} = e^{0.015 s}$$

so $Z_1^*(t)/Z_0^*(t)$ goes from $0.2/0.8 = 1/4$ to $0.8/0.2 = 4$ in time $\ln(16)/0.015 = 184$, confirming that the window in which competing subclones coexist is short.

Discussion

Work of Sottoriva and Graham (2015) and their co-authors in Williams et al (2016) has shown that in many cases an exponentially growing tumor has a $1/f$ site frequency spectrum. This result has a simple derivation but the claim has drawn a large amount of criticism. Many of these concern the quality of the data used. Here, we have performed a mathematical analysis to show that given enough sequence data the site frequency spectrum can be used to distinguish neutral evolution from one specific type of selection. This analysis provides a useful complement to studies based solely on simulation.

Here we have studied the two-type model of cancer evolution in which the exponentially growing population of type 0 cells can mutate to a fitter type 1, and all cells can experience neutral mutations. In this model there are three types of mutations that we call 0, $1A$, and 1. Type 0 mutations are neutral, occur to type 0 individuals, and have a $1/f$ site frequency spectrum. Type 1 mutations are neutral, occur to type 1 individuals, and again have a $1/f$ site frequency spectrum. Type 1A mutations are selective, occur to type 0 individuals, and result in type 1 individuals. When the two types have growth rates $\lambda_0 < \lambda_1$, where $\alpha = \lambda_0/\lambda_1$, then the site frequency spectrum has the shape $1/f^\alpha$ due to 1A mutations and the type 0 neutral mutations present in the founders of the type 1 population. These mutation types are more numerous than the others.

McDonald, Chakrabarti, and Michor (2018) have used the two-type model to suggest that models with selection can have a $1/f$ site frequency spectrum. Our results in Section 3 show this is not true when type 1 mutations all have the same fitness increase. Their model has random increases in fitness, but in Section 4 we show that this feature does not significantly change the qualitative features of the site frequency spectrum.

Bozic, Paterson, and Waclaw (2019) study the two-type model and show that it is difficult to capture a subclonal driver mutation at intermediate frequency. Their model allows only one type 1A mutation. Using our simple analytical results and computer simulations, we confirm that this prediction holds in the two type model without that restriction.

Methods

2.4 Simple derivations of the $1/f$ spectrum

Sottoriva and Graham (2015) says that "the power law signature is common to multiple tumor types and is a consequence of the effectively-neutral evolutionary dynamics that underpin the evolution of a large proportion of cancers." To explain the source of the $1/f$ curve in an exponentially growing tumor, we give the derivation of the $1/f$ frequency distribution from Williams et al (2016). They assumed that cells divide at rate λ and use $N(t)$ to be the number of cells at time t. If we assume that the mutation rate is μ (which we assume takes into account their ploidy parameter π), then the expected number of new mutations before time t, $M(t)$, satisfies

$$\frac{dM}{dt} = \mu\lambda N(t).$$

Solving gives

$$M(t) = \mu\lambda \int_0^t N(s)\, ds.$$

Since $N(s) = e^{\lambda s}$ (we have set β in Williams et al (2016) to be 1 for simplicity), we observe that a mutation that occurs at time s will have frequency $e^{-\lambda s}$ in the population. Evaluating the integral in the previous formula, we have

$$M(t) = \mu(e^{\lambda t} - 1).$$

Ignoring the -1, if we set $t_f = -(1/\lambda)\log f$ to make $N(t_f) = 1/f$ so that mutations before time t_f will have frequency $\geqslant f$, then

Theorem 4 (Sottoriva and Graham 2015)**.** *The number of mutations with frequency $\geqslant f$ is*

$$M(t_f) = \mu/f. \tag{2.7}$$

Note that in this derivation, mutations occur only at birth. If we instead let mutations happen continuously throughout a cell's lifetime and call the mutation rate ν, then Durrett (2013) has shown

$$M(t_f) = \frac{\nu}{\lambda f}. \tag{2.8}$$

From the derivation given above, we see that the $1/f$ site frequency spectrum comes from the fact that mutations occur at a rate proportional to the size of the population and the fact that the population is growing exponentially fast.

2.5 Proof of Theorem 2

Proof. We follow the derivation of Theorem 4. If we let $N(s) = Z_1^*(s)$, then the number of type 1 mutations by time t satisfies

$$\begin{aligned} M_1(t) &= \nu \int_{s^1_{1/2}}^t N(s)\,ds \approx \nu \bar{V}_1 \int_{s^1_{1/2}}^t \exp(\lambda_1(s - s^1_{1/2}))\,ds \\ &\approx \nu \bar{V}_1 \exp(\lambda_1(t - s^1_{1/2}))/\lambda_1 \end{aligned}$$

where we have again dropped the -1 that comes from the lower limit. A mutation that occurs at a time $t \leqslant t_f = s^1_{1/2} - (1/\lambda_1)\log(f\bar{V}_1)$, when there are

$$\leqslant N(t_f) \approx \bar{V}_1 \exp(\lambda_1(t_f - s^1_{1/2})) = \bar{V}_1 \exp(-\log(f\bar{V}_1)) = 1/f$$

individuals, will occur in a fraction of $\geqslant f$ of the population, so computing $M(t_f)$ gives the desired result. □

2.6 Proof of Theorem 3

Recall from Durrett and Moseley (2010) that the points in the Poisson process in Theorem 1 indicate the contributions of the various type one families to the limit $\bar{V}_1$ and that the poisson point process has mean measure $\bar{\rho}(x, \infty) \sim x^{-\alpha}$. It then follows from Pitman and Yor (1997) that the fraction of the population each 1A family contains is distributed according to the Poisson-Dirichlet distribution $PD(\alpha, 0)$. Letting $\{A_i\}_i$ be sampled from $PD(\alpha, 0)$, note that

$$SFS_{1A}(f) = \mathbb{E}\left[\sum_i \mathbb{1}_{[f,1]}(A_i)\right]$$

Next, we use a trick from Pitman and Yor (1997). Dividing and multiplying by A_i in the sum,

$$SFS_{1A}(f) = \mathbb{E}\left[\sum_i \frac{\mathbb{1}_{[f,1]}(A_i)}{A_i} A_i\right]$$

Viewing A_i as the size biased probability of picking family i and $\mathbb{1}_{[f,1]}(A_i)/A_i$ as the value obtained from picking family i, we can simplify the expression in terms of the size biased pick A^*.

$$\mathbb{E}\left[\sum_i \frac{\mathbb{1}_{[f,1]}(A_i)}{A_i} A_i\right] = \mathbb{E}\left[\frac{\mathbb{1}_{[f,1]}(A^*)}{A^*}\right]$$

Recalling that the size biased pick from $PD(\alpha, 0)$ has distribution $Beta(1-\alpha, \alpha)$ and noting that $\Gamma(\alpha)\Gamma(1-\alpha)\sin(\pi\alpha) = \pi$, we conclude that

$$
\begin{aligned}
SFS_{1A}(f) &= \mathbb{E}\left[\frac{\mathbb{1}_{[f,1]}(A^*)}{A^*}\right] \\
&= \int_0^1 \frac{1}{x}\mathbb{1}_{[f,1]}(x)\frac{x^{-\alpha}(1-x)^{\alpha-1}}{\pi/\sin(\pi\alpha)}dx \\
&= \frac{\sin(\pi\alpha)}{\pi}\int_f^1 x^{-\alpha-1}(1-x)^{\alpha-1}dx \\
&= \frac{\sin(\pi\alpha)}{\pi}\left[-\frac{1}{a}\left(\frac{1}{x}-1\right)^{\alpha}\right]_{x=f}^{x=1} \\
&= \frac{\sin(\pi\alpha)}{\pi\alpha}\left(\frac{1}{f}-1\right)^{\alpha}
\end{aligned}
$$

This concludes the proof. Note that it is possible to extend this method to determine the site frequency spectrum of the type 1 mutations. If $A_i > f$, then for a mutation in family i to have frequency $> f$, the mutation needs to have frequency $> f/A_i$ in family i. Therefore, as per (2.8), on average there are $\nu A_i/(\lambda_1 f)$ type 1 mutations in family i that reach frequency $> f$ and we get

$$
\begin{aligned}
SFS_1(f) &= \mathbb{E}\left[\frac{\nu}{\lambda_1 f}\mathbb{1}_{[f,1]}(A^*)\right] \\
&= \frac{\nu}{\lambda_1 f}(1 - I_f(1-\alpha,\alpha))
\end{aligned}
$$

where $I_f(1-\alpha,\alpha)$ is the regularized incomplete beta function, also known as the CDF of $Beta(1-\alpha,\alpha)$ evaluated at f. For the values of f that concern us, the regularized incomplete beta function is roughly linear and therefore yields a roughly $1/f$ shape for SFS_1.

2.7 Passengers do not change the shape of the SFS

To show that the important 1A mutations happen soon after the first, and that therefore all important 1A mutations have roughly the same number of passengers,

consider two successful mutations at times s_0 and s_1 which have sizes $W_0 e^{\lambda_1(t-s_0)}$ and $W_1 e^{\lambda_1(t-s_1)}$. For the second mutation to be larger, we'd need $W_0/W_1 \leqslant e^{\lambda_1(s_0-s_1)}$. Since the cdf of the quotient of two exponentials with the same rate is $P(W_0/W_1 \leqslant x) = x/(x+1)$, we find that

$$P\left(W_0/W_1 \leqslant e^{\lambda_1(s_0-s_1)}\right) = \frac{1}{e^{\lambda_1(s_1-s_0)}+1}.$$

If $s_1 = s_0 + 4/\lambda_1 = s_0 + 200$, then the probability that the second mutation is larger is $(1+e^4)^{-1} = 0.018$. Thus, in our concrete example the most significant mutants occur within 200 time units of the first successful mutation. The mean number of mutations in 200 units of time is 200ν.

3

Competitive Exclusion in a Seasonal Environment

This chapter is from Tung and Durrett (2022), which has been published in *Theoretical Populations Biology*. Understanding the conditions that allow for multiple species to coexist has been of longstanding interest. The competitive exclusion principle, sometimes called Gause's principle, states that n resources can support at most n species. For example, in Gause (1932)'s experiments with Paramecium, there was one resource, food, and the species that better utilized the food Gause gave them drove the others to extinction. However, in other situations, what constitutes a resource is not always clear. Hutchinson (1961) drew attention to this through the "Paradox of the Plankton," the enormous diversity of phytoplankton coexisting despite the small number of resources in ocean water. Many explanations for the seeming failure of the competitive exclusion principle have been explored in math models; see Armstrong and McGehee (1980) for ODE models and Hening and Nguyen (2020) for SDE and piecewise deterministic Markov process models. Hutchinson's explanation was a changing environment; times when different species are favored would be considered different niches.

Chan, Durrett, and Lanchier (2009) considered a two-type contact process on a square lattice with long range interaction and showed that for an open set of parameters, two species can coexist in a model with two seasons. Their system is a stochastic spatial analog of

$$\frac{1}{n_i}\frac{dn_i}{dt} = \gamma_i(t)R - \sigma_i \qquad R = 1 - \sum_{i=1}^{k} n_i \tag{3.1}$$

where the γ_i are periodic functions, R represents available space, and k is the number of species. There is one resource R so in the temporally homogeneous case one species will competitively exclude the others. In the case that the $\gamma_i(t)$ are constant on $[0, T_1]$, on $[T_1, T_2]$, and periodic, there are two seasons and therefore two niches; so, it is not surprising that two species can coexist. They speculated that a fast dispersing species could exploit the early part of a season before losing to a superior competitor, allowing for three or more species to coexist. Here, we will prove that this is not possible in the ODE model.

To do so, we consider a system that we call the three-species periodic Volterra model

$$\frac{1}{n_i}\frac{dn_i}{dt} = \gamma_i(t)R(n_1, n_2, n_3, t) - \sigma_i(t), \qquad i = 1, 2, 3 \tag{3.2}$$

where n_i is the population size of species i, γ_i is the growth rate gained per available resource amount for species i, R is the amount of available resource, and σ_i is the death rate of species i. R, γ_i, and σ_i are all periodic in t with period T. To prove results about this system, we suppose that

A1 R is strictly decreasing with respect to population sizes n_1, n_2, and n_3.

A2 $R \leqslant 0$ when the population size of any one species is sufficiently large.

A3 $\gamma_i(t)$ and $\sigma_i(t)$ are positive and upper bounded.

A4 R is continuous with respect to n_1, n_2, and n_3.

A5 We have existence and uniqueness of solutions.

$A1-3$ are reasonable biologically. $A1$ states that a larger population means more resource consumption, and therefore less available resource. $A2$ implies that there is a limited amount of resources that cannot support infinitely large populations. $A3$ ensures that our birth and death functions have the proper sign and do not blow up. $A4-A5$ are reasonable mathematically. The system (3.1) with three species satisfies these conditions.

We also will not consider the case that there is a nontrivial triple c_i such that

$$c_1\gamma_1 + c_2\gamma_2 + c_3\gamma_3 = \int_0^T c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3 dt = 0$$

This case is also ignored when examining the competitive exclusion principle for the Volterra model with multiple resources - see page 47 of Hofbauer and Sigmund 1998. The reason is that this case represents a degenerate case where one of the species populations can be written as a function that is increasing with respect to the other two and is periodic in t with period T. This means the system can be reduced to a two-species model. While coexistence is possible with two seasons under this case, its equilibrium lacks stability and it loses its coexistence with the slightest perturbations in γ_i or σ_i.

There are many different definitions of coexistence. For this paper, we say that the system exhibits coexistence if none of the species go extinct for any positive initial condition. A species goes extinct if $\lim_{t\to\infty} n_i(t) = 0$.

We show the following theorems.

Theorem 5. *If the growth per resource rates γ_i are linearly dependent, then the three-species periodic Volterra model does not exhibit coexistence.*

The linear dependence assumption holds in the piecewise constant three-species model of Chan, Durrett, and Lanchier and implies that the system does not exhibit coexistence. Miller and Klausmeier (2017) also come to the same conclusion, although their arguments are not rigorous.

Exact linear dependence is a strong condition, but our result is robust to slight deviations; we extend Theorem 5 to the situation in which the γ_i are nearly linearly dependent

Theorem 6. *Given c_i not all 0, σ_i, and R for the three-species periodic Volterra model, there exists an $\epsilon > 0$ such that if*

$$\int_0^T |c_1\gamma_1 + c_2\gamma_2 + c_3\gamma_3| dt < \epsilon$$

Then the model does not exhibit coexistence.

Section 3.1 gives an important lemma used to prove the two theorems. Section 3.2 proves the theorems and gives an example application.

3.1 Condition for Coexistence and Extinction

To determine if a species goes extinct, i.e., $\lim_{t\to\infty} n_i(t) = 0$, we first focus on $n_1^{c_1} n_2^{c_2} n_3^{c_3}$. This function has been used to prove results on coexistence in other models (Volterra 1928, Hofbauer 1981, Hofbauer and Sigmund 1998, Schreiber et al. 2011) and acts as an "average Lyapunov" function whose decrease implies average movement towards faces with $c_i > 0$ and away from faces with $c_i < 0$. One context this function has been used is to show competitive exclusion in the multi-resource Volterra model, where $\gamma_i R$ is replaced with $\sum_j \gamma_{ij} R_j$. To our knowledge, this is the first time this function has been used to deal with time-periodic coefficients.

Multiplying through (3.2) by c_i, summing, and setting $\boldsymbol{c}\cdot\boldsymbol{\gamma} = c_1\gamma_1 + c_2\gamma_2 + c_3\gamma_3$ and $\boldsymbol{c}\cdot\boldsymbol{\sigma} = c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3$, we get

$$\frac{1}{n_1^{c_1}n_2^{c_2}n_3^{c_3}}\frac{dn_1^{c_1}n_2^{c_2}n_3^{c_3}}{dt} = (\boldsymbol{c}\cdot\boldsymbol{\gamma})R(n_1,n_2,n_3,t) - (\boldsymbol{c}\cdot\boldsymbol{\sigma}) \tag{3.3}$$

For some systems, an appropriate choice of c_1, c_2, and c_3 will let us ignore R and show that $n_1^{c_1}n_2^{c_2}n_3^{c_3} \to 0$. Once this is established, we can use the following lemma to preclude coexistence.

Lemma 7. *The three-species periodic Volterra model* (3.2) *does not exhibit coexistence iff there exist constants* c_1, c_2, *and* c_3 *that are not all positive and*

$$\lim_{t\to\infty} n_1^{c_1}n_2^{c_2}n_3^{c_3} = 0.$$

The remainder of this section describes the ideas behind the proof of Lemma 7. We start with the easier direction. If the three-species periodic Volterra model does not exhibit coexistence, then there exists a species, which we label as species 1, whose population approaches 0. Setting $c_1 = 1$ and $c_2 = c_3 = 0$ completes this direction.

We now proceed with the other direction by considering the possible cases of the signs of c_i. Recalling that $n_i(t)$ is upper bounded by A2, if c_i is nonpositive then $n_i^{c_i}$ is bounded from below. This implies that for case 1, where all c_i are nonpositive, then $n_1^{c_1}n_2^{c_2}n_3^{c_3}$ cannot approach 0 and we can ignore this case. This also implies that for case 2, where only one of the c_i is positive, which we label as species 1, then $n_1^{c_1}n_2^{c_2}n_3^{c_3} \to 0$ implies $n_1^{c_1} \to 0$ as $t \to \infty$ and therefore that species 1 goes extinct.

Case 3, where two of the c_i are positive and one is nonpositive, is more involved. Let $c_1, c_2 > 0$ and $c_3 < 0$. Using that n_i is upper bounded once more,

$$n_1(t)^{c_1} n_2(t)^{c_2} \to 0 \tag{3.4}$$

In order to show that species 1 or 2 goes extinct, we need to rule out the possibility that species 1 and 2 take turns approaching 0, keeping $\limsup n_1 = n_1^*$ and $\limsup n_2 = n_2^*$ positive. To do so, we note that by (3.4), paths from $(n_1^*, 0)$ to $(0, n_2^*)$ must travel near the origin after some time. Then, we show that if the trajectory of (n_1, n_2) nears the origin and eventually leaves, then (n_1, n_2) will consistently leave the origin in the same direction, without loss of generality towards $(n_1^*, 0)$. This implies that $n_2^* = 0$ and therefore species 2 goes extinct. The details of the proof of case 3, as well as a visual overview of the proof, can be found in Section 3.3.

3.2 Applications

In this section, we use Lemma 7 to prove Theorems 5 and 6 and give examples of systems where we can rule out coexistence.

Proof of Theorem 5. Since the γ_i are linearly dependent, we can find c_i such that $\boldsymbol{c} \cdot \boldsymbol{\gamma} = 0$. Then by (3.3),

$$\frac{d}{dt} \left[\ln \left(n_1^{c_1} n_2^{c_2} n_3^{c_3} \right) \right] = -\boldsymbol{c} \cdot \boldsymbol{\sigma}$$

Flipping the signs of c_i if necessary, we can assume without loss of generality that $\boldsymbol{c} \cdot \boldsymbol{\sigma} > 0$. This implies $n_1^{c_1} n_2^{c_2} n_3^{c_3} \to 0$, so applying Lemma 7 completes the proof. □

Application of Theorem 5: Contact process with seasons. One notable example of a model where the γ_i are linearly dependent is the mean field limit of the three-species two-seasons model that appears in Chan, Durrett, and Lanchier (2009) - see (3.1). They showed that, in the absence of species 3, coexistence occurs when

$$\frac{1}{T}\int_0^T \gamma_1(1-\overline{n}_2)-\sigma_1 dt > 0 \text{ and } \frac{1}{T}\int_0^T \gamma_2(1-\overline{n}_1)-\sigma_2 dt > 0$$

where $\overline{n}_i$ is the nontrivial periodic solution to $\frac{1}{n_i}\frac{dn_i}{dt} = \gamma_i(t)(1-n_i)-\sigma_i$. The first integral represents the net growth rate of species 1 when n_1 has been small for a long time, giving n_2 time to converge to $\overline{n}_2$. If the growth rate is positive, then species 1 won't go extinct. Similarly, the second condition represents that species 2 has a positive growth rate when n_2 is small and n_1 is near $\overline{n}_1$. Using the ODE result, they showed that the same conditions guaranteed coexistence for the two-type contact process on the square lattice with long range interactions.

Chan, Durrett, and Lanchier also conjectured that three species could coexist with two seasons. This could be true in their stochastic model, but it does not hold in the mean field limit. To prove this, since there are only two seasons and γ_i is a function of the season, the space of possible γ_i has dimension 2. There are three species, so the γ_i are linearly dependent. Therefore, by Theorem 5 the three species cannot coexist.

Readers of Chan, Durrett, and Lanchier will note that their conjecture is supported by a numerical simulation, which seems to showcase coexistence. However, this simulation should not be taken as evidence, since the parameters are not arbitrary. In more detail, in their simulation, species 1 has growth rates of $(3,1)$, species 2 has growth rates $(1,3)$, and species 3 has growth rates $(2,2)$ for seasons 1, 2 respectively. All species have death rate 1. If we choose $\boldsymbol{c} = (1,1,-2)$, then we not only get $\boldsymbol{c}\cdot\boldsymbol{\gamma} = 0$ but also $\boldsymbol{c}\cdot\boldsymbol{\sigma} = 0$. This would make $n_1(t)^{c_1}n_2(t)^{c_2}n_3(t)^{c_3}$ an invariant, allowing for coexistence. In fact, it behaves in a periodic orbit; substituting $n_3 = C\sqrt{n_1 n_2}$ back into the differential equations reduce the system to a two-species system, which Mottoni and Schiaffino (1981) showed to always approach a periodic orbit. However, as noted earlier in the dissertation, this is not considered

for coexistence by ecologists due to its lack of robustness and stability.

Proof of Theorem 6. In the case where c_i are not all the same sign, note that $|R(n_1, n_2, n_3, t)|$ is bounded since R has monotonicity and n_i is bounded. Integrating (3.3), we get

$$\begin{aligned} \ln\left[\frac{n_1(t+T)^{c_1}n_2(t+T)^{c_2}n_3(t+T)^{c_3}}{n_1(t)^{c_1}n_2(t)^{c_2}n_3(t)^{c_3}}\right] &= \int_t^{t+T} (\boldsymbol{c}\cdot\boldsymbol{\gamma})R - (\boldsymbol{c}\cdot\boldsymbol{\sigma})ds \\ &\leqslant \int_0^T |\boldsymbol{c}\cdot\boldsymbol{\gamma}| \max|R| - (\boldsymbol{c}\cdot\boldsymbol{\sigma})ds \end{aligned} \tag{3.5}$$

Then, flipping the signs of c_i if necessary, setting

$$\epsilon < \frac{1}{\max|R(n_1, n_2, n_3, t)|}\int_0^T \boldsymbol{c}\cdot\boldsymbol{\sigma} ds$$

will force $n_1(t+T)^{c_1}n_2(t+T)^{c_2}n_3(t+T)^{c_3} < n_1(t)^{c_1}n_2(t)^{c_2}n_3(t)^{c_3}$, and therefore $n_1^{c_1}n_2^{c_2}n_3^{c_3} \to 0$. Applying Lemma 7 completes the proof.

In the case where c_i all have the same sign, we assume without loss of generality that c_i are all positive. Then,

$$\begin{aligned} \ln\left[\frac{n_1(t+T)^{c_1}}{n_1(t)^{c_1}}\right] &= \int_t^{t+T} c_1\gamma_1 R - c_1\sigma_1 ds \\ &\leqslant \int_0^T (\boldsymbol{c}\cdot\boldsymbol{\gamma}) \max|R| - c_1\sigma_1 ds \end{aligned} \tag{3.6}$$

Setting

$$\epsilon < \frac{1}{\max|R(n_1, n_2, n_3, t)|}\int_0^T c_1\sigma_1 ds$$

forces $n_1(t) \to 0$ as $t \to \infty$, which implies extinction of species 1 and no coexistence.

□

Application of Theorem 6: Numerical example. To conclude this section, we use a concrete example. Consider the system

$$\frac{1}{n_1}\frac{dn_1}{dt} = \gamma(3,5,t)R(n_1,n_2,n_3) - 1$$

$$\frac{1}{n_2}\frac{dn_2}{dt} = \gamma(4.5,3.4,t)R(n_1,n_2,n_3) - 1$$

$$\frac{1}{n_3}\frac{dn_3}{dt} = \gamma(4.1,3.78,t)R(n_1,n_2,n_3) - 1$$

$$R = 1 - n_1 - n_2 - n_3$$

$$\gamma(a,b,t) = \begin{cases} a & 0 < t \leqslant 0.6 \\ a(b/a)^{(t-0.6)/0.4} & 0.6 < t \leqslant 1 \\ b & 1 < t \leqslant 1.6 \\ b(a/b)^{(t-1.6)/0.4} & 1.6 < t \leqslant 2 \end{cases}$$

The system can be viewed as an extension of the two-season model - see Fig 3.1A; instead of making γ_i piecewise constant, γ_i now has a transition period between the two seasons, making the γ_i linearly independent. However, the γ_i are close enough to being linearly dependent that we can preclude coexistence.

To show that they cannot coexist, we mimic the proof of Theorem 6. We first bound R from above. Note that $\gamma_i \geqslant 3$. This implies that when $R > 1/3$, then $dn_i/dt > 0$ and therefore $dR/dt < 0$. Thus, after some time, $R \leqslant 1/3$. Next, we set $c_1 = -1, c_2 = -916/307$, and $c_3 = 1230/307$; these were chosen to make the γ_i linearly dependent during times $[0, 0.6] \cup [1, 1.6]$. Now, we integrate.

$$\int_0^T \boldsymbol{c}\cdot\boldsymbol{\sigma}dt = \frac{14}{307} \approx 0.046$$

$$\max|R(n_1,n_2,n_3,t)| \int_0^T |\boldsymbol{c}\cdot\boldsymbol{\gamma}|dt \leqslant 0.041$$

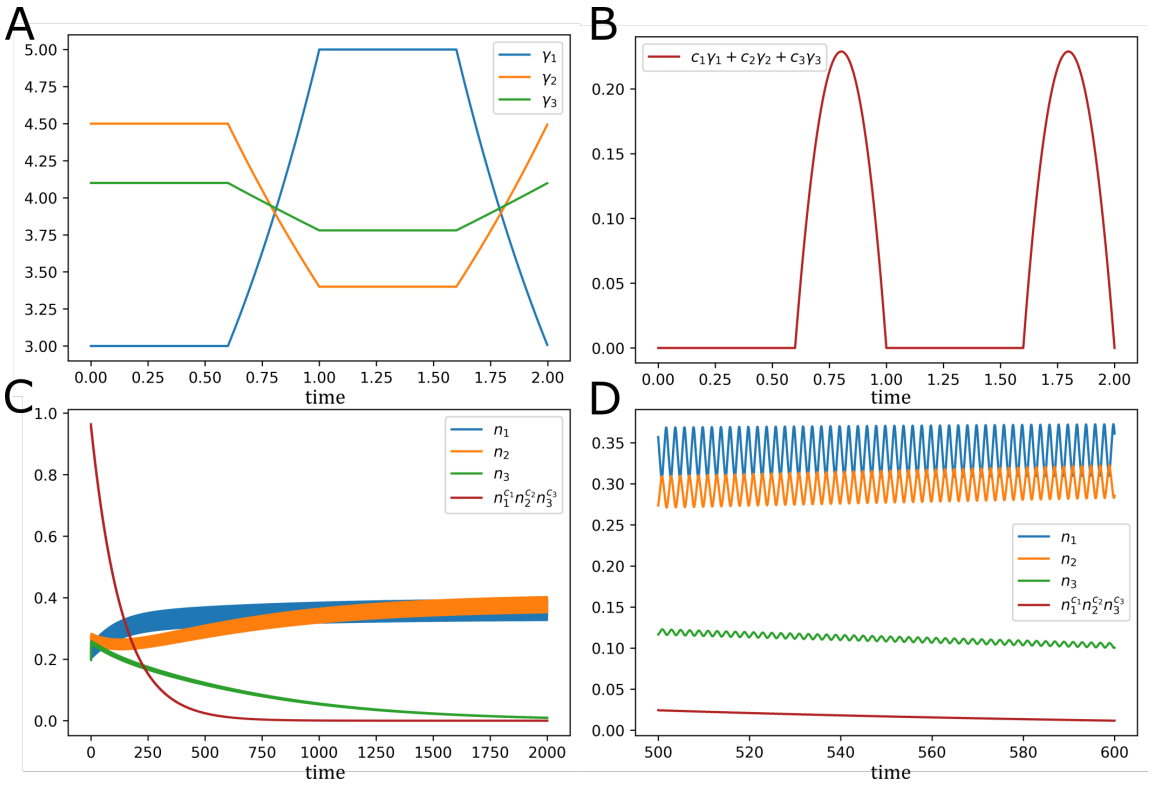

FIGURE 3.1: **Example system given in Section 3.2**. A) One period of growth functions γ_i. Instead of the two-season model considered in Theorem 5, we add transition seasons to make the γ_i continuous. B) Measure of linear dependence $\boldsymbol{c} \cdot \boldsymbol{\gamma}$ for our choice of c_1, c_2, and c_3. Since $\boldsymbol{c} \cdot \boldsymbol{\gamma}$ is sufficiently close to 0, we can preclude coexistence. C) Population dynamics. Species 3 goes extinct. D) Population dynamics zoomed. The populations oscillate over time due to changes in γ_i.

Since $0.041 < 0.046$, by (3.5), $n_1^{c_1} n_2^{c_2} n_3^{c_3} \to 0$. Applying Lemma 7 implies that one of the species goes extinct. This can be seen in Figure 3.1.

3.3 Proof of Lemma 7

Here, we give the details for case 3 in the proof of Lemma 7. For a visual overview of the proof, see Figure 3.2. We start by proving the following two lemmas.

Lemma 8. *Let $n_1(t)^{c_1} n_2(t)^{c_2} < C$. The time needed for $n_1(t)^{c_1}$ to pass through the interval $D_1 = [C/\epsilon, \epsilon]$ approaches infinity as $C \to 0$.*

Lemma 9. *If (n_1, n_2) passes through the region $D = \{(n_1, n_2) | 0 \leqslant n_1(t)^{c_1}, n_2(t)^{c_2} \leqslant$*

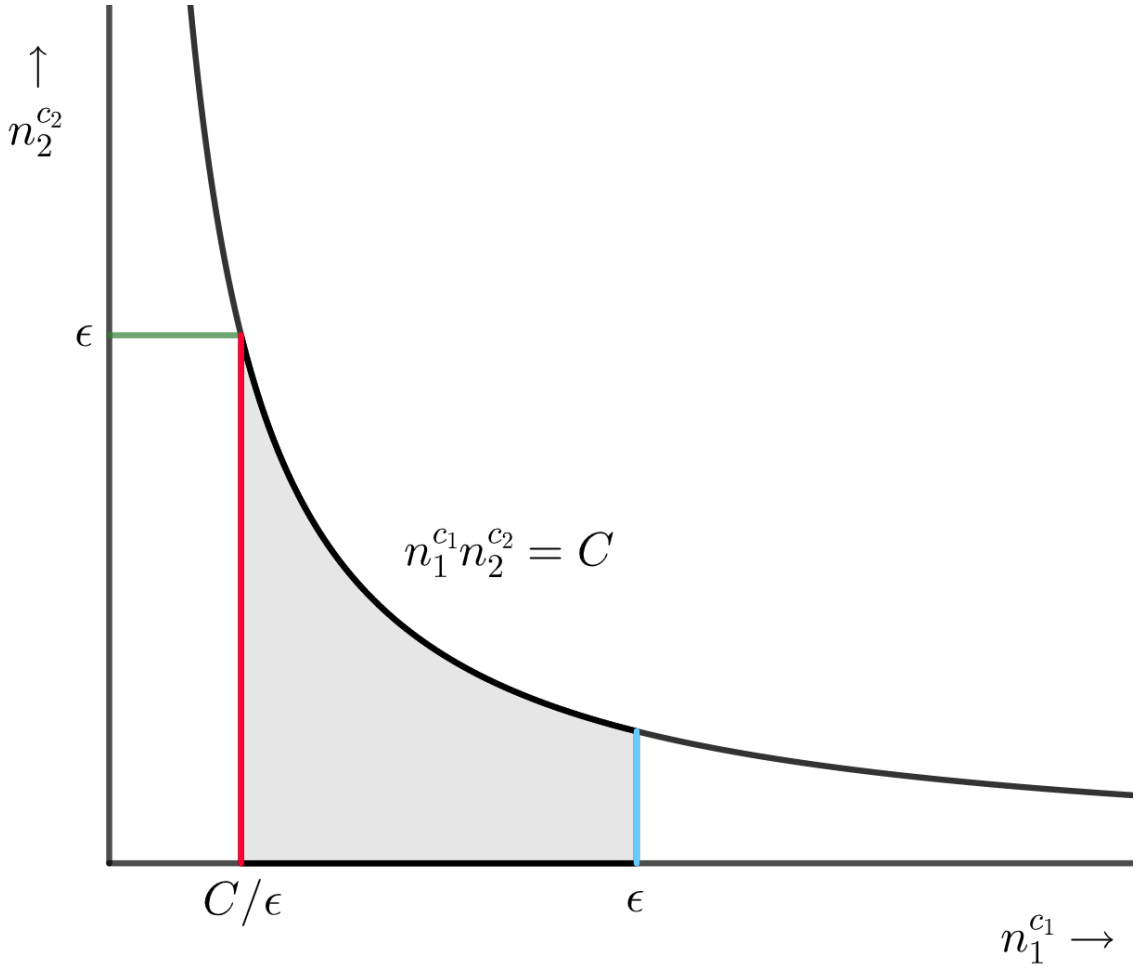

FIGURE 3.2: **Visual for the proof of case three of Lemma 7.** From (3.4), after sufficiently large time, the solution must remain below the curve $n_1^{c_1} n_2^{c_2} = C$. To disprove coexistence, we need to show that $(n_1^{c_1}, n_2^{c_2})$ cannot move between $(n_1^*, 0)$ and $(0, n_2^*)$. This is equivalent to proving that the solution cannot move from the blue line to the green line, or vice versa. Lemma 11 shows that spending sufficient time between the green and blue lines causes $n_1(t+T)/n_1(t)$ to converge to being positive or negative. Lemma 8 shows that sufficient time for convergence is eventually always achieved. We now have two cases. If the sign is positive, then the solution will not be able to move from blue to green; $n_1(t+T)/n_1(t)$ will become positive before reaching the red line, preventing the solution from reaching green. Similarly, if the sign is negative, then the solution will not be able to move from green to blue

$\epsilon\}$ *starting at some sufficiently large time* τ*, then* $n_3(\tau)$ *is bounded from below.*

We define passing through D_1 as moving from $n_1(t)^{c_1} = \epsilon$ to $n_1(t)^{c_1} = C/\epsilon$ or vice versa without leaving D_1 and passing through D as moving from $n_1(t)^{c_1} = \epsilon$ to $n_2(t)^{c_2} = \epsilon$ or vice versa without leaving D. To prove Lemma 8, note that

$$c_1 \max_t \left[\gamma_1 R(0,0,0,t) - \sigma_1\right] > \frac{d}{dt}\left[\ln\left(n_1^{c_1}\right)\right] > c_1 \min_t \left[\gamma_1 R(M,M,M,t) - \sigma_1\right]$$

Letting the RHS be p_{min} and LHS be p_{max}, the amount of time spent traveling from ϵ to C/ϵ and the other direction is lower bounded by

$$\frac{1}{p_{min}} \ln\left(C/\epsilon^2\right) \text{ and } \frac{1}{p_{max}} \ln\left(\epsilon^2/C\right)$$

respectively. As C approaches 0, both expressions, and therefore the time for pass through D_1, approach infinity.

Now, we prove Lemma 9. By Lemma 8 and (3.4), if τ is sufficiently large, $(n_1(\tau), n_2(\tau))$ cannot pass through D by time $\tau + T$. We now aim to show that if we start on D where $n_1(\tau)^{c_1} = \epsilon$, then if $n_3(\tau)$ is small, $n_1(\tau+T)^{c_1} > \epsilon$ and there is no passing through; the other side, where $n_2(\tau)^{c_2} = \epsilon$ can be proven similarly. Note that

$$\begin{aligned}
\ln\left[\frac{n_1(\tau+T)^{c_1}}{n_1(\tau)^{c_1}}\right] &= c_1 \int_\tau^{\tau+T} \gamma_1 R(n_1, n_2, n_3, t) - \sigma_1 dt \\
&\geqslant c_1 \int_\tau^{\tau+T} \gamma_1 R(\epsilon^{1/c_1}, \epsilon^{1/c_2}, n_3, t) - \sigma_1 dt
\end{aligned}$$

When ϵ is sufficiently close to 0 and $n_3 = 0$, the RHS must be positive, else it would imply that species 1 would go extinct even without competition. By continuity of R (A5), there exists some constant $a > 0$ where the RHS is still positive when $n_3(\tau) \leqslant a$, and therefore $n_1(\tau + T) > n_1(\tau)$. This would make n_1 leave D without passing

through. As such, to pass through D, there is a lower bound on the population of species 3.

In order to prove Lemma 11, we first need an understanding the dynamics of the system when only one species is present.

Lemma 10. *When $n_1 = n_2 = 0$, there exists at most 1 nontrivial periodic orbit n_3^* for species 3. If n_3^* exists and we have a nontrivial solution n_3^{**}, then $n_3^{**} \to n_3^*$.*

Proof. For simplicity of notation, we write $R(0, 0, n_3, t)$ as $R(n_3, t)$. Suppose there are two nontrivial solutions n_3^* and n_3^{**}, with n_3^* being a periodic orbit. Then

$$\frac{d\ln(n_3^*)}{dt} = \gamma_3 R(n_3^*, t) - \sigma_3$$

$$\frac{d\ln(n_3^{**})}{dt} = \gamma_3 R(n_3^{**}, t) - \sigma_3$$

Subtracting, we get

$$\frac{d\ln(n_3^*/n_3^{**})}{dt} = \gamma_3[R(n_3^*, t) - R(n_3^{**}, t)]$$

WLOG $n_3^*(0) \geqslant n_3^{**}(0)$. By the uniqueness condition, $n_3^*(t) = n_3^{**}(t)$ for any t iff $n_3^*(0) = n_3^{**}(0)$. As such, $n_3^*(t) \geqslant n_3^{**}(t)$, which implies $R(n_3^*, t) - R(n_3^{**}, t) \leqslant 0$, with equality only when $n_3^* = n_3^{**}$.

We first establish that n_3^* is a unique periodic orbit. If n_3^{**} is also a periodic orbit, then

$$0 = \int_0^T \frac{d\ln(n_3^*/n_3^{**})}{dt} dt = \int_0^T \gamma_3[R(n_3^*, t) - R(n_3^{**}, t)]dt$$

Since $\gamma_3(t) > 0$, this implies $n_3^* = n_3^{**}$.

To address the second claim, if n_3^{**} is not a periodic orbit, then note that

$$0 > \int_0^T \gamma[R(n_3^*, t) - R(n_3^{**}, t)]dt = \int_0^T \frac{d\ln(n_3^*/n_3^{**})}{dt}dt = -\ln(n_3^{**}(T)) + \ln(n_3^{**}(0))$$

As such, n_3^{**} is increasing every cycle and approaches n^*. □

Having established the existence and uniqueness of an equilibrium when only one species is present, we are now ready to prove Lemma 11.

Lemma 11. *For sufficiently small ϵ, when (n_1, n_2) is passing through D, then after finite time s,*

$$\frac{n_1(\tau + T)}{n_1(\tau)} = \int_\tau^{T+\tau} \gamma_1 R(n_1, n_2, n_3, t) - \sigma_1 dt \text{ and } \int_0^T \gamma_1 R(0, 0, n^*, t) - \sigma_1 dt$$

have the same sign, where n^* is the nontrivial equilibrium solution for n_3 in the absence of the other two species.

To prove, we first note that by monotonicity,

$$R(\epsilon^{1/c_1}, \epsilon^{1/c_2}, n_3, t) < R(n_1, n_2, n_3, t) \leqslant R(0, 0, n_3, t)$$

Let $R(0, 0, n_3, t) - R(\epsilon^{1/c_1}, \epsilon^{1/c_2}, n_3, t) < m$. Then by the ODE comparison theorem, we know that n_3 is bounded between the solutions for

$$\frac{1}{n_3}\frac{dn_3}{dt} = \gamma_3(R(0, 0, n_3, t) - m) - \sigma_3, \quad \frac{1}{n_3}\frac{dn_3}{dt} = \gamma_3 R(0, 0, n_3, t) - \sigma_3$$

Let n^* be the equilibrium solution for the upper bound and n_m the solution for the lower bound. By continuity of R we can find an ϵ that lets m be arbitrarily small. Applying Lemma 10, n_m must approach its equilibrium. Then,

$$0 = \lim_{\tau\to\infty} \int_{\tau}^{\tau+T} \gamma_3 R(0,0,n_m,t) - (\sigma_3 + \gamma_3 m)dt = \int_0^T \gamma_3 R(0,0,n^*,t) - \sigma_3$$

Rearranging the above,

$$\lim_{\tau\to\infty} \int_{\tau}^{\tau+T} \gamma_3 (R(0,0,n_m,t) - R(0,0,n^*,t))dt = m\int_0^T \gamma_3 dt$$

which approaches 0 as ϵ approaches 0. Noting that $\gamma_3 > 0$ and $R(0,0,n^*,t) < R(0,0,n_m,t)$ implies

$$\lim_{\epsilon\to 0}\lim_{\tau\to\infty} \int_{\tau}^{\tau+T} R(0,0,n_m,t) - R(0,0,n^*,t)dt = 0$$

and subsequently

$$\lim_{\epsilon\to 0}\lim_{\tau\to\infty} \int_{\tau}^{\tau+T} \gamma_i R(0,0,n_m,t) = \int_0^T \gamma_i R(0,0,n^*,t)$$

To show that the integrals have the same sign happens after time s regardless of $n_3(0)$ at time of entering D, recall from Lemma 9 that $n_3(0)$ has a nonzero lower bound $\underline{m}$ and the upper bound M. By monotonicity, n_m will take longest to reach the same sign when $n_m(0) = M$ or $\underline{m}$. Take s to be the longer time. As $n_m \leqslant n_3 \leqslant n^*$, we have our desired result.

We are now ready to prove extinction in case 3. By Lemma 11, $n_1(t+T)/n_1(t)$ will always be positive or negative after time s in D, which we know will happen from Lemma 8. If the sign is negative, then $n_1^{c_1}$ would shrink before reaching ϵ, and therefore the solution cannot move from $(0, n_2^*)$ to $(n_1^*, 0)$. If the sign is positive, then $n_1^{c_1}$ would grow before reaching C/ϵ, which implies that $n_2^{c_2} < \epsilon$ and the solution cannot move from $(n_1^*, 0)$ to $(0, n_2^*)$. As such, we have a contradiction, and the lim sup of n_1 or n_2 is 0. This concludes case 3.

4

Conclusion

In this dissertation we studied two problems in mathematical biology. In chapter 2, we studied the two-type model of cancer evolution in which the exponentially growing population of type 0 cells can mutate to a fitter type 1, and all cells can experience neutral mutations. In this model there are three types of mutations that we call 0, $1A$, and 1. Type 0 mutations are neutral, occur to type 0 individuals, and have a $1/f$ site frequency spectrum. Type 1 mutations are neutral, occur to type 1 individuals, and again have a $1/f$ site frequency spectrum. Type 1A mutations are selective, occur to type 0 individuals, and result in type 1 individuals. When the two types have growth rates $\lambda_0 < \lambda_1$, where $\alpha = \lambda_0/\lambda_1$, then the site frequency spectrum has the shape $1/f^\alpha$ due to 1A mutations and the type 0 neutral mutations present in the founders of the type 1 population. These mutation types are more numerous than the others.

As our approach focused on theory, there are many potential applications that can be explored. One is to see whether the $1/f^\alpha$ shape can be seen in data, and whether a log-log plot would serve as a better test for neutrality than the test introduced by Williams et al (2016), which involved doing linear fits of the SFS against $1/f$

and checking the resulting R^2 value. If so, then it would also be interesting to find which cancers exhibit such patterns, and what about their growth or structure makes this apparent. That being said, many factors could prevent observation of the site frequency spectrum. For example, α may be much closer to 1 than in our examples, making it hard to observe. Another example is that since our results are for when $t \to \infty$, we can expect a $1/f$ component from type 0 mutations to the site frequency spectra; the site frequency spectra of type 1 and type $1A$ are unlikely to differ much since the impactful $1A$ mutations happen at largely the same time.

In chapter 3, we studied a three species resource competition ODE model with periodic environment. We found that when the growth per resource rates are (almost) linearly dependent, then there is no coexistence. A corollary of this theorem is that three species cannot coexist when there are only two seasons. Our work also suggests that the numerical simulation used to back the conjecture in Chan, Durrett, and Lanchier (2009) should be taken with caution.

Future directions for this project include extending beyond three species, giving conditions for which species will go extinct, and adding stochasticity to the model. Unfortunately, the idea used in the proof of case 3 for Lemma 7 does not generalize beyond three species. For example, consider when c_1, c_2, c_3 are positive and c_4 is negative. We could crack this case when there were only 3 species since coexistence meant species 1 and 2 would have to exit the box of length ϵ around the origin in two different ways. In the 4 species case though, coexistence does not imply leaving a region in two different directions; one can imagine trajectory (n_1, n_2, n_3) making clockwise loops. For determining which species goes extinct, note that case 3 does not imply which species goes extinct. Work on this problem is in progress. For adding stochasticity, this can be done in a few ways. One is to revert to a spatial model as described in Chan, Durrett, and Lanchier, especially one with smaller interaction range. Another is to keep working with the ODE model but adding stochasticity for

the length of the seasons. I conjecture that when the seasons are determined by a piecewise deterministic markov chain, our results still hold; case 1 and 2 still hold, and for case 3, the increase in difficultly for exiting the ϵ box over time should allow a proof using Borel-Cantelli.

Bibliography

[1] R. A. Armstrong and R. McGehee, *Coexistence of species competing for shared resources*, Theoretical population biology, 9 (1976), pp. 317–328.

[2] ——, *Competitive exclusion*, The American Naturalist, 115 (1980), pp. 151–170.

[3] K. B. Athreya, P. E. Ney, and P. Ney, *Branching processes*, Springer, 1972.

[4] A. Balaparya and S. De, *Revisiting signatures of neutral tumor evolution in the light of complexity of cancer genomic data*, Nature genetics, 50 (2018), pp. 1626–1628.

[5] M. Borowiak, F. Ning, J. Pei, S. Zhao, H.-R. Tung, and R. Durrett, *Controlling the spread of covid-19 on college campuses*, Mathematical Biosciences and Engineering, 18 (2021), pp. 551–563.

[6] L. Boyle, S. Hletko, J. Huang, J. Lee, G. Pallod, H.-R. Tung, and R. Durrett, *Selective sweeps in sars-cov-2 variant competition*, Proceedings of the National Academy of Sciences, 119 (2022), p. e2213879119.

[7] I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak, *Accumulation of driver and passenger mutations during tumor progression*, Proceedings of the National Academy of Sciences, 107 (2010), pp. 18545–18550.

[8] I. Bozic, C. Paterson, and B. Waclaw, *On measuring selection in cancer from subclonal mutation frequencies*, PLoS computational biology, 15 (2019), p. e1007368.

[9] B. Chan, R. Durrett, and N. Lanchier, *Coexistence for a multitype contact process with seasons*, The Annals of Applied Probability, 19 (2009), pp. 1921–1943.

[10] P. Chesson, *Multispecies competition in variable environments*, Theoretical population biology, 45 (1994), pp. 227–276.

[11] J. M. Cushing, *Two species competition in a periodic environment*, Journal of Mathematical Biology, 10 (1980), pp. 385–400.

[12] M. Dawson, C. Dudley, S. Omoma, H.-R. Tung, and M.-V. Ciocanel, *Characterizing emerging features in cell dynamics using topological data analysis methods*, Mathematical Biosciences and Engineering, 20 (2023), pp. 3023–3046.

[13] P. De Mottoni and A. Schiaffino, *Competition systems with periodic coefficients: a geometric approach*, Journal of Mathematical Biology, 11 (1981), pp. 319–335.

[14] R. Durrett, *Population genetics of neutral mutations in exponentially growing cancer cell populations*, The annals of applied probability: an official journal of the Institute of Mathematical Statistics, 23 (2013), p. 230.

[15] R. Durrett, *Branching process models of cancer*, in Branching process models of cancer, Springer, 2015, pp. 1–63.

[16] R. Durrett, J. Foo, K. Leder, J. Mayberry, and F. Michor, *Evolutionary dynamics of tumor progression with random fitness values*, Theoretical population biology, 78 (2010), pp. 54–66.

[17] R. Durrett and S. Moseley, *Evolution of resistance and progression to disease during clonal expansion of cancer*, Theoretical population biology, 77 (2010), pp. 42–48.

[18] G. F. Gause, *Experimental studies on the struggle for existence: I. mixed population of two species of yeast*, Journal of experimental biology, 9 (1932), pp. 389–402.

[19] A. Hening and D. H. Nguyen, *The competitive exclusion principle in stochastic environments*, Journal of mathematical biology, 80 (2020), pp. 1323–1351.

[20] J. Hofbauer, *A general cooperation theorem for hypercycles*, Monatshefte für Mathematik, 91 (1981), pp. 233–240.

[21] J. Hofbauer, K. Sigmund, et al., *Evolutionary games and population dynamics*, Cambridge university press, 1998.

[22] G. E. Hutchinson, *The paradox of the plankton*, The American Naturalist, 95 (1961), pp. 137–145.

[23] T. O. McDonald, S. Chakrabarti, and F. Michor, *Currently available bulk sequencing data do not necessarily support a model of neutral tumor evolution*, Nature genetics, 50 (2018), pp. 1620–1623.

[24] T. O. McDonald and M. Kimmel, *A multitype infinite-allele branching process with applications to cancer evolution*, Journal of Applied Probability, 52 (2015), pp. 864–876.

[25] E. T. Miller and C. A. Klausmeier, *Evolutionary stability of coexistence due to the storage effect in a two-season model*, Theoretical Ecology, 10 (2017), pp. 91–103.

[26] R. Noble, D. Burri, C. Le Sueur, J. Lemant, Y. Viossat, J. N. Kather, and N. Beerenwinkel, *Spatial structure governs the mode of tumour evolution*, Nature ecology & evolution, 6 (2022), pp. 207–217.

[27] J. Noorbakhsh and J. H. Chuang, *Uncertainties in tumor allele frequencies limit power to infer evolutionary pressures*, Nature genetics, 49 (2017), pp. 1288–1289.

[28] P. C. Nowell, *The clonal evolution of tumor cell populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression.*, Science, 194 (1976), pp. 23–28.

[29] J. Pitman, *Combinatorial stochastic processes: Ecole d'eté de probabilités de saint-flour xxxii-2002*, Springer, 2006.

[30] J. Pitman and M. Yor, *The two-parameter poisson-dirichlet distribution derived from a stable subordinator*, The Annals of Probability, (1997), pp. 855–900.

[31] S. J. Schreiber, M. Benaïm, and K. A. Atchadé, *Persistence in fluctuating environments*, Journal of Mathematical Biology, 62 (2011), pp. 655–683.

[32] A. Sottoriva and T. A. Graham, *A pan-cancer signature of neutral tumor evolution*, bioRxiv, (2015), p. 014894.

[33] A. Sottoriva, H. Kang, Z. Ma, T. A. Graham, M. P. Salomon, J. Zhao, P. Marjoram, K. Siegmund, M. F. Press, D. Shibata, et al., *A big bang model of human colorectal tumor growth*, Nature genetics, 47 (2015), pp. 209–216.

[34] M. Tarabichi, I. Martincorena, M. Gerstung, A. M. Leroi, F. Markowetz, P. T. Spellman, Q. D. Morris, O. C. Lingjærde, D. C. Wedge, and P. Van Loo, *Neutral tumor evolution?*, Nature genetics, 50 (2018), pp. 1630–1633.

[35] H.-R. Tung, *Precluding oscillations in michaelis–menten approximations of dual-site phosphorylation systems*, Mathematical Biosciences, 306 (2018), pp. 56–59.

[36] H.-R. Tung and R. Durrett, *Signatures of neutral evolution in exponentially growing tumors: A theoretical perspective*, PLOS Computational Biology, 17 (2021), p. e1008701.

[37] ——, *Competitive exclusion in a model with seasonality: three species cannot coexist in an ecosystem with two seasons*, Theoretical Population Biology, (2022).

[38] V. Volterra, *Variations and fluctuations of the number of individuals in animal species living together*, ICES Journal of Marine Science, 3 (1928), pp. 3–51.

[39] H.-Y. Wang, Y. Chen, D. Tong, S. Ling, Z. Hu, Y. Tao, X. Lu, and C.-I. Wu, *Is the evolution in tumors darwinian or non-darwinian?*, National Science Review, 5 (2018), pp. 15–17.

[40] M. J. Williams, B. Werner, C. P. Barnes, T. A. Graham, and A. Sottoriva, *Identification of neutral tumor evolution across cancer types*, Nature genetics, 48 (2016), pp. 238–244.

Biography

Hwai-Ray (Ray) Tung attended Brown University as an undergraduate in 2014 and earned a BS in mathematics and a BS in applied mathematics-biology with honors. He worked with Anne Shiu at a Texas A&M REU to publish his first paper, Tung (2018), and worked with Kavita Ramanan to write his honors thesis. Following undergrad, Ray joined the mathematics PhD program at Duke and worked with his advisor, Rick Durrett, on problems in mathematical biology. This led to two publications, Tung and Durrett (2021) and Tung and Durrett (2022). He also often did research with undergraduates through DOmath (now Math+), a program where undergraduates work in teams with a faculty member and graduate student over the summer on research. Two of these project were under the guidance of Rick Durrett, and one was under the guidance of Veronica Ciocanel. This led to three publications, Borowiak et al (2021), Boyle et al (2022), and Dawson et al 2023. In addition to research, Ray sought to improve his teaching over his graduate career, taking part in the Certificate for College Teaching program and the Teaching on Purpose fellowship program. In 2021, Ray received the L.P. Smith Award for Teaching Excellence from the Duke math department. Ray is headed to the University of Utah for his postdoc.