ADVANTAGE OF MUTANTS IN A STOCHASTICALLY

SELECTIVELY NEUTRAL ENVIRONMENT

Weichen Zhou
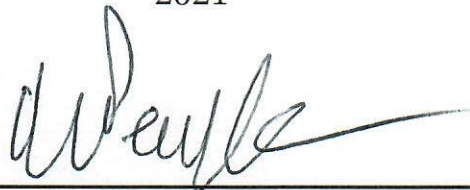
A THESIS

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial

Fulfillment of the Requirements for the Degree of Master of
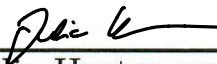
Philosophy

2021

Supervisor of Thesis

Robin Pemantle, Merriam Term Professor of Mathematics

Graduate Group Chairperson

Julia Hartmann, Professor of Mathematics

Thesis Committee:
Robin Pemantle, Merriam Term Professor of Mathematics
Bhaswar Bhattacharya, Assistant Professor of Statistics

Acknowledgments

First, I would like to thank my advisor, Prof. Robin Pemantle, for all the guidance, help and advice that he has given me throughout this project and during the years I worked in Probability Theory. I would also like to thank Prof. Julia Hartmann, Prof. Bhaswar Bhattacharya, Dr. Bharathwaj Palvannan, Prof. Jennifer Morse, Prof. Jonah Blasiak and Prof. Xiaoming Song for all their support and encouragement when I was lost and when things were rough and challenging.

The five years at Penn were wonderful. I want to take this chance to thank all my friends in the Math department for making this experience absolutely unforgettable. I owe special thanks to George Wang, Yao-Rui Yeo, and Andrea Thevis for not only being such wonderful friends but also great peer mentors for me.

Last but not the least, I would like to thank my parents for always being there for me, giving me all the love and support no matter what.

Abstract

ADVANTAGE OF MUTANTS IN A STOCHASTICALLY

SELECTIVELY NEUTRAL ENVIRONMENT

Weichen Zhou

Advisor: Robin Pemantle

Fixation probability describes the probability for an invading mutant to take over the entire population on a spatial model in a stochastically selectively neutral environment. Our main goal in this survey is to study the advantage of mutants on the specific model defined on a line with N vertices. In the case when the variation is small, we compute the precise expression of the asymptotics of the fixation probability. At the end of the survey, we discuss a few possible future working directions on fixation probability modeling over different geometries.

Contents

List of Figures

1 Introduction

1.1 Background

Evolution dynamics in random environments is an interesting problem that arose in the intersection of ecology, biology and mathematics in the past few decades and has caught the interest of many scientists. The process considers different types of cells that evolve following certain birth and death rules in a given environment. An important factor that was taken into consideration is spatial randomness, that is, the birth and death rate of the present cells are determined by their locations in the environment. It is worth noticing that these studies perform a critical role in the study of the evolution of cancerous cells. Cells located at different locations in the body have different surrounding environments that directly lead to different access of oxygen and nutrition levels that support the reproduction or dying of the cancerous cells. Gaining understanding to the evolution process therefore contributes greatly to the medical studies of finding treatments for cancers.

In the actual biological process, there are many principal factors that may affect the reproduction performance of cells and the biological interactions between two cells are very complicated processes. The studies were then brought to a more theoretical level, where the environment under consideration are graphs and there are only two types of cells in the environment: the regular cells and the mutants. Moreover, a reproduced cell can only take over a neighboring cell of different type if and only if

the two cells are connected by an edge. Those studies focused on what is known as the fixation probability, the probability that a mutant eventually takes over the entire system. Markov Chain methods, branching processes and diffusion approximations are the three main methods that were used when studying the fixation probability, starting with the work by Fisher[1], Haldane[4], and Wright[10] in the early 90's, then to Kimura[5] in mid-90's, eventually extending to the work of Mahdipour-Shirayeh, Darooneh, Long, Komarova, Kohandel, Farhang-Sardroodi and Nikbakht [8][2] in 2017.

To model genotype evolution by environmental interaction, we introduce a Moran process, a process of evolution that captures the behavior of the birth and death process of cells in the system naturally. A Moran process can be described as the following: we start with a fixed population of N cells where each cell is associated with a fitness factor that determines the probability for the cell to reproduce or die, one cell is chosen at random to die and another cell is chosen at random to reproduce, then the newly produced cell will replace the one that died keeping the total number of cells constant.

The fixation probability on a ring with N cells (one of which chosen at random to become the first mutant) was studied in [8] and the following results were found:

- When the mutant and regular cells have the same fitness at a given location, the mean fixation probability is $\frac{1}{N}$.

- Let r be fixed and σ_i be a random variable with mean 0. On site i, a regular

2

cell has fitness $r + \sigma_i$ and a mutant cell has fitness $r - \sigma_i$, then the fixation probability is greater than $\frac{1}{N}$.

- When the fitness for regular cells and mutants at each site i are either randomly sampled from the same mean zero distribution or the fitness of regular cells are shuffled given the location of the first mutant in the system, the fixation probability is greater than $\frac{1}{N}$.

Similar results were also observed when considering a complete graph with N vertices. An interesting finding was that on a complete graph, the processing being a birth-death process or death-birth process affects the fixation probability [8].

In the second result above, the fitness values of regular cells and mutants are sampled from certain probability distributions, so it is natural to ask whether this randomness accelerates or delays fixation. Further studies in [2] by Farhang-Sardroodi, Darooneh, Nikbakht, et al. shows that on complete graphs with N vertices, randomness accelerates fixation for all N, whereas on a ring with N vertices randomness delays fixation only for large enough N. And the speed of fixation depends on the distribution that the random quantities δ_i are sampled from.

In this survey, we will start with studying the fixation probability on a line with N sites indexed 1 through N. Each site is associated with a random mutant fitness and a random regular fitness. The goal is to give a rigorous expression for the asymptotics of the fixation probability when the initial mutant is placed at site 1. We will also start with some works on generalizing the line model when the initial mutant is

placed at a different site and exploring fixation probabilities on different geometries. Lastly, we will state a conjecture and several possible directions for future studies.

1.2 Line Model Setup

Let N be a positive integer. We shall start with constructing the model on a line with N sites. Let $(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)$ be the underlying probability space and fix $\delta \in (0, 1)$. For $1 \le k \le N$, we define the following quantities for the site indexed by k:

(a) Rademacher random variable B_k and B_k';

(b) Regular fitness at site k: $\mu_k = 1 + \delta B_k$;

(c) Mutant fitness at site k: $\nu_k = 1 + \delta B_k'$.

Here, we require that all Rademacher random variables $\{B_k\}_{k=1}^N$ and $\{B_k'\}_{k=1}^N$ to be independent.

Our initial configuration has only one interface, that is where an edge of the graph connects a mutant cell and a regular cell, since the initial mutant is at site 1 and all remaining sites are holding regular cells. We can see that the number of interfaces will always be 1 unless the evolution reaches one of the two absorbing states: the mutants die out, or, the mutants take over the world. And if some mutants are still present after some time of evolution, we will reach the configuration where mutants occupy the the sites $\{1, 2, \cdots, k\}$ for some k and sites $\{k+1, \cdots, N\}$ are occupied by regular cells. In this way, we can identify the configurations by $\{0, 1, 2, \cdots, N-1, N\}$, where 0 corresponds to the configuration where mutants die out, 1 corresponds to the same

4

configuration as our initial state, N corresponds to the configuration where mutants take over the world, and in particular, k corresponds to the intermediate configuration mentioned above.

Let $\xi_t^{(i,j)}$ denote a rate 1 Poisson process for $1 \leq i,j \leq N$ independent of all Rademacher random variables defined. We want to use this Poisson process to keep track of the evolution. By the nature of the geometry, cell type exchange can only take place across the edge that acts as an interface, so we also restrict that $|i - j| = 1$. It is very important to keep in mind that each site can only affect its neighboring one or two sites.

This is a continuous time process, but we can discretize it by only sampling on the time stamps where the configuration changes. Suppose that after some time into the evolution process, the interface is between site k and $k+1$ (for generality, we suppose $1 < k < N$), we can denote this time by $\xi^{(k,k+1)}$. There are only two edges connecting site k, hence in the next time stamp where the configuration changes, there are only two possible outcomes:

(a) The mutant at site k reproduce, and takes over the regular cell at site $k+1$, leaving the configuration as $\{1, \cdots, k+1\}$ being mutants and $\{k+2, \cdots N\}$ being regular cells;

(b) The regular cell at site $k+1$ reproduces, and takes over the mutant at site k, leaving the configuration as $\{1, \cdots, k-1\}$ being mutants and $\{k, \cdots, N\}$ begin regular cells.

We define the random quantity representing the relative fitness level of site k:

$$\beta_k := \frac{\mu_k}{\mu_k + \nu_k} \tag{1.1}$$

The evolution process is indeed a Markov process. At state k, the process either evolves to state $k+1$ or state $k-1$. We can explicitly define the transition probabilities:

$$\mathbb{P}(k, k+1) = \frac{\beta_{k+1}}{\beta_{k+1} + (1 - \beta_k)}, \tag{1.2}$$

$$\mathbb{P}(k, k-1) = \frac{1 - \beta_k}{(1 - \beta_k) + \beta_{k+1}}. \tag{1.3}$$

1.3   Main Theorems

We mentioned earlier that there are only two absorbing states for the evolution: (1) the mutants die out, and (2) the mutants take over the world. Let $G_1 = G_1(\delta)$ be the event that the mutant takes over all N sites on the line. The subscript 1 here is to indicate the position of the initial mutant is at site 1. Then, the fixation probability is denoted by $\mathbb{P}_N(G_1(\delta))$. Following the setup, we will prove the following theorems in this survey.

The first theorem considers the asymptotics of the the fixation probability given $N \to \infty$ and $\delta\sqrt{N} \to c$, where c is a fixed positive constant.

Theorem 1.1. *Suppose $N \to \infty$ and $\delta\sqrt{N} \to c$, for fixed $c > 0$, we have*

$$N\mathbb{P}_N(G_1(\delta)) \to g(c) \tag{1.4}$$

where

$$g(c) = \mathbb{E}\left[\frac{1}{\int_0^1 \exp(\sqrt{2}cB_s)ds}\right] \tag{1.5}$$

for a standard Brownian Motion $\{B_s\}$. The function g is continuous and strictly increasing on $(0, \infty)$ satisfying the following criteria:

$$g(c) - 1 \quad \sim \quad \frac{c^2}{24} \qquad as \quad c \downarrow 0, \tag{1.6}$$

$$g(c) \quad \sim \quad \frac{c}{\sqrt{\pi}} \qquad as \quad c \to \infty. \tag{1.7}$$

The second theorem considers the asymptotics where $\delta \gg \frac{1}{\sqrt{N}}$ and $\delta \ll (\log N)^{-\epsilon}$ for some $\epsilon > 0$.

Theorem 1.2. *Assuming $\delta\sqrt{N} \to \infty$, and suppose that there is an $\epsilon > 0$ such that $\delta(\log N)^\epsilon \to 0$, we have*

$$\mathbb{P}_N(G_1(\delta)) \sim \frac{\delta}{\sqrt{\pi N}}. \tag{1.8}$$

Theorem 1.1 and Theorem 1.2 supports the statement that mutants have an advantage in the process of evolution in a stochastic environment. The last main theorem we shall prove further supports the statement and it reveals that the overall advantage comes from the mutants taking advantage of the cases where sites favoring mutants

and sites favoring regulars cells are balanced, namely $\sum_k B_j = \sum_l B'_l$.

Theorem 1.3. *For even integer $N = 2k$, let $\tilde{\mathbb{P}}_N(G_1(\delta))$ denote the conditional probability $\mathbb{P}_N(G_1(\delta) \mid \sum_j B_j = \sum_l B'_l)$. Then for all N and all δ, we have*

$$N\tilde{\mathbb{P}}_N(G_1(\delta)) = 1. \tag{1.9}$$

2 Techniques and Proofs

In this section, we will explore the techniques and proofs for the main theorems stated in section 1.3.

2.1 Proof of Theorem 1.1 and Theorem 1.3

We start working towards of the proofs of Theorem 1.1 and Theorem 1.3 from the following result on birth and death processes.

Proposition 2.1. *For a birth and death process with transition probability*

$$p_k = p(k, k+1) \qquad and \qquad p(k, k-1) = q_k := 1 - p_k.$$

Let $\mathcal{Q}_x$ be the law of the process starting at x and let τ_a be the hitting time at state a, then

$$\mathcal{Q}_1(\tau_N < \tau_0) = \frac{1}{\sum_{k=0}^{N-1} \prod_{j=1}^{k} \frac{q_j}{p_j}}. \tag{2.1}$$

8

Proof. To prove the statement above, we use the correspondence between reversible Markov chains and electrical networks. Recall that an electrical network is a finite graph with values assigned to the edges called resistances. And the conductance on each edge is given by the reciprocal of the resistance.

Our finite graph is a line with N vertices. On edge $(k, k+1)$, assign the resistance

$$R(k, k+1) = \prod_{j=1}^{k} \frac{q_j}{p_j},$$

then the conductance on edge $(k, k+1)$ is

$$C(k, k+1) = \frac{1}{R(k, k+1)} = \prod_{j=1}^{k} \frac{p_j}{q_j}.$$

The correspondence between the birth-death process and the electrical network is given by

$$p(k, k+1) = \frac{C(k, k+1))}{C(k, k+1) + C(k, k-1)} = p_k.$$

Observe that

$$1 - p_1 = q_1 = \frac{C(1, 0)}{C(1, 0) + C(1, 2)},$$

which implies $C(1,0) = 1$. Then

$$\mathcal{Q}_1(\tau_N < \tau_0) = \frac{C(1,N)}{C(1,N) + C(1,0)} = \frac{\frac{1}{\sum\limits_{k=1}^{N-1} R(k,k+1)}}{\frac{1}{\sum\limits_{k=1}^{N-1} R(k,k+1)} + 1}$$

$$= \frac{1}{1 + \sum\limits_{k=1}^{N-1} R(k,k+1)} = \frac{1}{\sum\limits_{k=0}^{N-1} \prod\limits_{j=1}^{k} \frac{q_j}{p_j}}.$$

$\square$

Our next goal is to show that the denominator of equation (2.1) can be well approximated by a functional of a random walk, so that we can then approximate that random walk functional with a functional of a Brownian motion.

Consider the following notation:

$$X_k := \log \frac{q_k}{p_k} = \log \frac{1 - \beta_k}{\beta_{k+1}}, \qquad S_k := \sum_{j=1}^{k} X_k \qquad \text{for} \quad 1 \le k \le N - 1;$$

with this notation, the expression in equation (2.1) can be re-written as

$$\mathbb{P}_N(G_1(\delta)) = \mathbb{E}\left[\frac{1}{\sum\limits_{k=0}^{N-1} \exp(S_k)} \right]. \tag{2.2}$$

Define

$$\tilde{X}_k := \log \frac{1 - \beta_{k+1}}{\beta_{k+1}} = \log \frac{\nu_{k+1}}{\mu_{k+1}}, \qquad \tilde{S}_k := \sum_{j=1}^{k} X_j \qquad \text{for} \quad 1 \le k \le N - 1.$$

Then $\tilde{S}_k$ is a random walk on $\log \frac{1+\delta}{1-\delta}\mathbb{Z}$, with the probability to move forward is $\frac{1}{4}$, the probability to move backward is $\frac{1}{4}$ and the probability to stay is $\frac{1}{2}$. Let $\delta' = \frac{1+\delta}{1-\delta}$, also note that $\delta' = 2\delta + O(\delta^2)$.

For $\delta \le 1 - \epsilon$,

$$
\begin{aligned}
|\tilde{S}_k - S_k| = \left| \sum_{j=1}^{k} \tilde{X}_j - X_j \right| &= \left| \sum_{j=1}^{k} \log(1 - \beta_{j+1}) - \log(1 - \beta_k) \right| \\
&= |\log(1 - \beta_{k+1}) - \log(1 - \beta_1)| \\
&\le \log(\frac{1+\delta}{1-\delta})^2 = 2\delta' = C_\epsilon \delta.
\end{aligned}
\tag{2.3}
$$

Donsker's Theorem states that for a random walk $S_n = \sum_{i=1}^{n} X_n$ with $\mathbb{E}[X_i] = 0$ and $Var(X_i) = 1$, $\frac{(S_{\lfloor Nt \rfloor})_{t \in (0,1)}}{\sqrt{N}}$ converges in distribution to a standard Brownian motion in the Skorokhod topology when $n \to \infty$. Scaling our defined $\tilde{X}_k$ by $\sqrt{2}\delta$ then apply Donsker's Theorem, we get that

$$
\frac{(S_{\lfloor Nt \rfloor})_{t \in (0,1)}}{\sqrt{2}\delta\sqrt{N}} \xrightarrow{N \to \infty} B_s
$$

$$
(S_{\lfloor nt \rfloor})_{t \in (0,1)} \xrightarrow{N \to \infty} \sqrt{2} \cdot c \cdot B_s
\tag{2.4}
$$

for a standard Brownian motion B_s where $\delta\sqrt{N} \to c$. Now we prove the following lemma:

Lemma 2.2. *Allow both δ and N to vary but $\delta\sqrt{N}$ is bounded away from 0 and infinity, then the random variables $\{N\mathbb{P}_N(G_1(\delta))\}$ are uniformly integrable.*

Proof. For a simple random walk S_n, by the reflection principle we have that

$$\mathbb{P}(\min_{j \leq n} S_j \leq -r) = \mathbb{P}(S_n = -r) + 2\mathbb{P}(S_n < -r). \tag{2.5}$$

Since a simple random walk is a martingale, we can then apply Azuma's inequality:

$$\mathbb{P}(\min_{j \leq n} S_j \leq -r) = \mathbb{P}(S_n = -r) + 2\mathbb{P}(S_n < -r)$$

$$\leq 2\mathbb{P}(S_n = -r) + 2\mathbb{P}(S_n < -r)$$

$$= 2\mathbb{P}(S_n \leq -r)$$

$$\leq 2\exp\left(\frac{-2r^2}{n}\right). \tag{2.6}$$

Recall that we constructed $\tilde{S}_k$ to be a simple random walk with step size scaled by δ' and the holding probability is $\frac{1}{2}$. Adapt the analysis above to $\tilde{S}_k$, we have

$$\mathbb{P}\left(\min_{j \leq k} \tilde{S}_k < -t\right) \leq 2\exp\left(\frac{-2t^2}{(\delta')^2 k}\right) \tag{2.7}$$

for any $k \in \mathbb{Z}^+$ and $t > 0$. And from the construction of S_k and $\tilde{S}_k$, we also have that

$$\left| \frac{\sum_{j=1}^{N} \exp(\tilde{S}_j)}{\sum_{j=1}^{N} \exp(S_j)} - 1 \right| = o(1) \tag{2.8}$$

With the expression in equation (2.2), in order to show that $\{N\mathbb{P}_N(G_1(\delta))\}$ are uni-

formly integrable, we only need to show that $\left\{ \dfrac{N}{\sum\limits_{j=1}^{N} \exp(\tilde{S}_j)} \right\}$ are uniformly integrable.

By equation (2.7), we have

$$\mathbb{P}(\min_{k \leq \epsilon N} \tilde{S}_k \leq -1) \leq 2 \exp\left(\frac{-2}{(\delta')^2 \epsilon N} \right). \tag{2.9}$$

Let $B = \frac{2}{(\delta')^2 N}$, notice that this quantity depends on $\delta\sqrt{N}$ continuously. Since we

assumed that $0 < \epsilon < 1$, then $\epsilon < \epsilon N < N$, thus we have the following inequality:

$$\frac{N}{\sum\limits_{j=1}^{N} \exp(\tilde{S}_j)} \leq \frac{N}{\sum\limits_{j=1}^{\epsilon N} \exp(\tilde{S}_j)} \leq \frac{N}{\epsilon N \exp(\min\limits_{j \leq \epsilon N} \tilde{S}_j)} \leq \frac{1}{\epsilon \exp(\min\limits_{j \leq \epsilon N} \tilde{S}_j)}. \tag{2.10}$$

Then by (2.10) and (2.9), taking $\epsilon = \frac{e}{x}$, for any positive K,

$$
\begin{aligned}
\mathbb{E}\left[\frac{N}{\sum\limits_{j=1}^{N} \tilde{S}_j} \mathbb{1}\left(\frac{N}{\sum\limits_{j=1}^{N} \tilde{S}_j} > K \right) \right] &= \int\limits_{x \geq K} \mathbb{P}\left(\frac{N}{\sum\limits_{j=1}^{N} \tilde{S}_j} \geq x \right) \, dx \\
&\leq \int\limits_{x \geq K} \mathbb{P}\left(\frac{1}{\epsilon \exp\left(\min\limits_{j \leq \epsilon N} \tilde{S}_j \right)} \geq x \right) \, dx \\
&\leq \int\limits_{x \geq K} \mathbb{P}\left(\exp\left(\min\limits_{j \leq \epsilon N} \tilde{S}_j \right) \leq (\epsilon x)^{-1} \right) \, dx \\
&= \int\limits_{x \geq K} \mathbb{P}\left(\min\limits_{j \leq (eN/x)} \tilde{S}_j \leq -1 \right) \, dx \\
&\leq \int\limits_{x \geq K} 2 \exp\left(\frac{-Bx}{e} \right) \, dx.
\end{aligned}
$$

13

Notice that

$$\int_{x \geq K} 2 \exp\left(\frac{-Bx}{e}\right) \, dx \xrightarrow{K \to \infty} 0,$$

and the inequality holds for all positive integer N. Now we have that $\{N / \sum_{j=1}^{N} \exp(\tilde{S}_j)\}$ are uniformly integrable, hence the random variables $\{N\mathbb{P}_N(G_1(\delta))\}$ are uniformly integrable. $\qquad \square$

With Lemma 2.2, we can prove the following theorem.

Theorem 2.3. *Suppose* $\delta\sqrt{N} \xrightarrow{N \to \infty} c$ *where* $c > 0$ *is fixed, then for some standard Brownian motion* $\{B_s\}$, *we have that*

$$N\mathbb{P}_N(G_1(\delta)) \to \mathbb{E}\left[\frac{1}{\int_0^1 \exp(\sqrt{2}cB_s) \, ds}\right]. \tag{2.11}$$

Proof. By Lemma 2.2, and the result of convergence in distribution (2.4) from Donsker's theorem:

$$N\mathbb{P}_N(G_1(\delta)) = \mathbb{E}\left[\frac{N}{\sum_{k=0}^{N-1} \exp(S_k)}\right] \xrightarrow{N \to \infty} \mathbb{E}\left[\frac{1}{\int_0^1 \exp(\sqrt{2}cB_s \, ds)}\right].$$

$\qquad \square$

Statement (2.11) is exactly the first half of Theorem 1.1, and to prove the remaining part of Theorem 1.1, we need to evaluate the Brownian Integral $\int_0^1 \exp(\sqrt{2}Bs \, ds)$.

14

In order to do that, we first introduce the following notations:

$$A_\alpha(t) \quad := \quad \int_0^t e^{\alpha B_s} \, ds, \tag{2.12}$$

$$m_\alpha(t) \quad := \quad \mathbb{E}\left[A_\alpha(t)^{-1}\right]. \tag{2.13}$$

To finish the proof of Theorem 1.1, we need to evaluate $g(c) = m_{c\sqrt{2}}(1)$. We will use the following result from [9].

Proposition 2.4. *[9] Let $\{B_t\}_{t \geq 0}$ be a standard brownian motion and let $A(t) := \int_0^t e^{2B_s} \, ds$. Then*

$$\mathbb{E}[A_2(t)^{-1}|B_t = x] = \frac{xe^{-x}}{t \sinh(x)} \quad \text{if} \quad x \neq 0, \tag{2.14}$$

$$\mathbb{E}[A_2(t)^{-1}|B_t = 0] = t^{-1}, \tag{2.15}$$

$$m_2(t) \quad \sim \quad \sqrt{\frac{2}{\pi t}} \quad as \quad t \to \infty. \tag{2.16}$$

Proof. Equation (2.14) and (2.15) are proved by Matsumoto and Yor in Theorem 5.6 of [9], the exact equations are stated as Proposition 5.9 in [9]. Here, we'll only provide a sketch of the proof.

To prove the two equations, it takes an explicit expression of the probability density of A_t computed by Yor in [11], which is also proved as Theorem 4.1 in [9].

15

For a fixed $t > 0$, and for $u > 0$ and $x \in \mathbb{R}$, the probability density is as follows:

$$\mathbb{P}(A_t \in du, B_t \in dx) = \exp\left(-\frac{1 + e^{2x}}{2u}\right)\theta\left(\frac{e^x}{u}, t\right)\frac{du\ dx}{u} \tag{2.17}$$

where equation (2.9) in [9] defines $\theta(r, t)$ as

$$\theta(r, t) = \frac{r}{(2\pi^3 t)^{1/2}}e^{\pi^2/2t}\int_0^\infty e^{\xi^2/2t}e^{-r\cosh(\xi)}\sinh(\xi)\sin\left(\frac{\pi\xi}{t}\right)\ d\xi.$$

The proof of this probability density uses the Schrödinger operator H_λ ($\lambda > 0$) on $\mathbb{R}$ with the Liouville potential given by $H_\lambda = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}\lambda^2 e^{2x}$ for $x \in \mathbb{R}$. Then, combining the results on Green function with respect to Lebesgue measure and the heat kernel of the semigroup $\exp(-tH_\lambda), t > 0$, also with the integral representation of the product of two modified Bessel functions, we obtained the desired probability density.

Integrating both sides of (2.17) with respect to u yields (this is stated as Proposition 4.2 in [9])

$$\int_0^\infty e^{xr}\theta(r, t)\frac{dr}{r} = \frac{1}{\sqrt{2\pi t}}\exp\left(\frac{(\text{Argcosh}^2(x))}{2t}\right). \tag{2.18}$$

Take a non-negative Borel function f on $\mathbb{R}$, then (2.17) gives

$$\int_\mathbb{R} f(x)\mathbb{E}\left[\exp(-\lambda/A_t)|B_t = x\right]\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}\ dx = \mathbb{E}\left[f(B_t)\exp(-\lambda/A_t)\right].$$

16

Combine this with (2.18) we have the expression

$$\mathbb{E}\left[\exp(-\frac{\lambda}{A_t})|B_t = x\right] = \exp\left(-\frac{\text{Argcosh}^2(\lambda e^x + \cosh(x)) - x^2}{2t}\right). \qquad (2.19)$$

This equation holds for any $t > 0$ and $\lambda > 0$, and we obtain the explicit expression of

equation (2.14) and (2.15) by taking $(\lambda, x) = (1, x)$ and $(\lambda, x) = (1, 0)$ respectively.

To prove 2.16, from the definition of $A_\alpha(t)$ and $m_\alpha(t)$ above, let $\alpha = 2$, then

$$m_2(t) = \mathbb{E}\left[A_2(t)^{-1}\right] = \mathbb{E}\left[\int_0^1 e^{2B_s} \, ds\right].$$

Applying 2.14, we have

$$m_2(t) = \int \frac{xe^{-x}}{t \sinh(x)} dN(0,1)(x).$$

Then, with a change of variables $y = \frac{x}{\sqrt{t}}$, we have

$$m_2(t) = \frac{1}{\sqrt{t}} \int \frac{ye^{-\sqrt{t}y}}{\sinh(\sqrt{t}y)} dN(0,1)(y).$$

Since $\sinh(\sqrt{t}y) = \frac{e^{\sqrt{t}y} - e^{-\sqrt{t}y}}{2}$, then $\frac{ye^{-\sqrt{t}y}}{\sinh(\sqrt{t}y)} = \frac{2y}{e^{2\sqrt{t}y} - 1}$. This only converges for $y < 0$

as $t \to \infty$, and the limit is $2|y|\mathbb{1}_{y<0}$. Then taking the limit as $t \to \infty$, we have the

following result for $\sqrt{t}m_2(t)$:

$$\sqrt{t}m_2(t) = \int \frac{ye^{-\sqrt{t}y}}{\sinh(\sqrt{t}y)} dN(0,1)(y) \xrightarrow{t \to \infty} \int_{-\infty}^0 2|y| dN(0,1)(y) = \mathbb{E}[N(0,1)] = \sqrt{\frac{2}{\pi}}.$$

17

Thus,

$$m_2(t) \sim \sqrt{\frac{2}{\pi t}},$$

proving equation (2.16). □

Notice that in the previous proposition, we have $\alpha = 2$. To adapt the results above to what's needed for Theorem 1.1, we need to scale the above results:

Lemma 2.5. *For $\alpha, \nu, t > 0$,*

$$m_\alpha(t) = \frac{\alpha^2}{\nu^2} m_\nu \left(\frac{\alpha^2}{\nu^2} t \right). \tag{2.20}$$

Then we have:

$$m_\alpha(1) \sim \frac{\alpha}{\sqrt{2\pi}} \quad as \quad \alpha \to \infty. \tag{2.21}$$

Proof. Let $f_\alpha(x, t)$ be the density function of $A_\alpha(t)^{-1}$ at x, and define

$$W_t := \frac{\alpha}{\nu} B_{\frac{\nu^2}{\alpha^2} t}.$$

Since B_t is a standard Brownian motion, by the scaling property of Brownian motions,

18

W_t is also a standard Brownian motion and νW_t and αB_t equal in distribution. Hence

$$
\begin{aligned}
f_\alpha(x,t)dx &= \mathbb{P}\left(\left(\int_0^t e^{\alpha B_s \; ds}\right)^{-1} \in [x, x+dx]\right) \\
&= \mathbb{P}\left(\left(\frac{\nu^2}{\alpha^2}\int_0^{\alpha^2/\nu^2} e^{\nu W_u} \; du\right)^{-1} \in [x, x+dx]\right) \\
&= \mathbb{P}\left(\left(\int_0^{\alpha^2/\nu^2} e^{\nu W_u} \; du\right)^{-1} \in \left[\frac{\nu^2}{\alpha^2}x, \frac{\nu^2}{\alpha^2}(x+dx)\right]\right) \\
&= \frac{\nu^2}{\alpha^2}\mathbb{P}\left(\left(\int_0^{\alpha^2/\nu^2} e^{\nu W_u} \; du\right)^{-1} \in \left[\frac{\nu^2}{\alpha^2}x, \frac{\nu^2}{\alpha^2}x + \frac{\alpha^2}{\nu^2}dx\right]\right) \\
&= \frac{\nu^2}{\alpha^2}f_\nu\left(\frac{\nu^2}{\alpha^2}x, \frac{\alpha^2}{\nu^2}t\right) \; dx.
\end{aligned}
$$

Using the moment generating function, we have

$$
\begin{aligned}
m_\alpha(t) &= \int_{-\infty}^{+\infty} x \; f_\alpha(x,t)dx \\
&= \int_{-\infty}^{+\infty} \frac{\nu^2}{\alpha^2} \; x \; f_\nu\left(\frac{\nu^2}{\alpha^2}x, \frac{\alpha^2}{\nu^2}t\right) \; dx \\
&= \frac{\nu^2}{\alpha^2}\int_{-\infty}^{+\infty} x \; f_\nu\left(\frac{\nu^2}{\alpha^2}x, \frac{\alpha^2}{\nu^2}t\right) \; dx.
\end{aligned}
$$

We do a substitution by letting $y = \frac{\nu^2}{\alpha^2}x$, then $dy = \frac{\nu^2}{\alpha^2}dx$, and

$$
\begin{aligned}
m_\alpha(t) &= \frac{\nu^2}{\alpha^2} \int_{-\infty}^{+\infty} x\, f_\nu\left(\frac{\nu^2}{\alpha^2}x, \frac{\alpha^2}{\nu^2}t\right)\, dx \\
&= \frac{\nu^2}{\alpha^2} \int_{-\infty}^{+\infty} \frac{\alpha^2}{\nu^2} y f_\nu(y, \frac{\alpha^2}{\nu^2}t)\frac{\alpha^2}{\nu^2}dy \\
&= \frac{\alpha^2}{\nu^2} \int_{-\infty}^{+\infty} y f_\nu(y, \frac{\alpha^2}{\nu^2}t)dy \\
&= \frac{\alpha^2}{\nu^2} m_\nu\left(\frac{\alpha^2}{\nu^2}t\right).
\end{aligned}
$$

This proves equation (2.20). Then for $t = 1$, we have

$$
m_\alpha(1) = \frac{\alpha^2}{\nu^2} m_\nu\left(\frac{\alpha^2}{\nu^2}\right).
$$

To finish the proof, we use the result from Prop (2.4) and let $\nu = 2$:

$$
m_\alpha(1) = \frac{\alpha^2}{4} m_2(\frac{\alpha^2}{4}) \sim \frac{\alpha^2}{4}\sqrt{\frac{2}{\pi\frac{\alpha^2}{4}}} = \frac{\alpha}{\sqrt{2\pi}}.
$$

$\square$

Now we have evaluated the Brownian integral, we just need to prove the following

lemma to have all the pieces we need to prove Theorem 1.1 and Theorem 1.3.

Lemma 2.6. $\phi(x) := \frac{xe^{-x}}{\sinh(x)}$ *is a strictly convex function. And* $\int \phi(x)dN(0,t)(x)$ *is*

strictly increasing in t.

Proof. For $\phi(x) = \frac{xe^{-x}}{\sinh(x)}$, we compute the following derivatives:

$$\phi'(x) = \frac{e^{-x}\left(\sinh x - x\sinh x - x\cosh x\right)}{\sinh^2 x}.$$

$$\phi''(x) = \frac{e^{-x}}{\sinh^3 x}\left(2x + (2x - 2)(\sinh^2 x + \sinh x\cosh x)\right)$$

$$= \frac{e^{-x}}{\sinh^3 x}\left(2x + (2x - 2)\left(\frac{e^{2x}}{2} - \frac{1}{2}\right)\right)$$

$$= \frac{e^{-x}}{\sinh^3 x}\left((x-1)e^{2x} + x + 1\right).$$

To show the strict convexity of $\phi(x)$, we show that $\phi''(x)$ is strictly positive when $x \neq 0$. Observe the expression of $\phi''(x)$, notice that e^{-x} is always positive, and $\sinh(x)$ is positive on $(0, \infty)$ and negative on $(-\infty, 0)$. Consider the factor

$$f(x) = (x-1)e^{2x} + x + 1,$$

$$f'(x) = e^{2x}(2x - 1) + 1,$$

$f'(x) = 0$ has unique solution at $x = 0$, and $f'(x) > 0$ on $(-\infty, 0)$ and $f'(x) > 0$ on $(0, \infty)$. Also, $f(0) = 0$. From this, we can deduce that $f(x) < 0$ on $(-\infty, 0)$ and $f(x) > 0$ on $(0, \infty)$. Thus $f(x)$ and $\sinh(x)$ has the same sign on $(-\infty, 0)$ and $(0, \infty)$, hence $\phi''(x) > 0$ for all nonzero x. Thus, $\phi(x)$ is strictly convex. Since $\phi(x)$ is a strictly convex function, $\int \phi(x)dN(0,t)(x)$ is strictly increasing in t. $\qquad\square$

Proof of Theorem 1.1:

Proof. The first half of theorem is proved in Theorem 2.3.

By Lemma 2.6, we have that function g is continuous and strictly increasing on $(0, \infty)$.

In the proof of Lemma 2.5, we have

$$g(c) = m_{c\sqrt{2}}(1) \sim \frac{c\sqrt{2}}{\sqrt{2\pi}} = \frac{c}{\sqrt{\pi}} \qquad \text{as} \qquad c \to \infty.$$

Then, to show the symptotics as $c \downarrow 0$, analogous to the proof of Proposition 2.4, we have the following integral for $\alpha = 2$:

$$tm_2(t) = \int \frac{xe^{-x}}{\sinh(x)} dN(0,1)(x).$$

Since the integrand $\frac{xe^{-x}}{\sinh(x)}$ has Taylor Series expansion $T(x) = 1 - x + \frac{x^2}{3} + O(x^4)$, we substitute this into the integral above to get

$$tm_2(t) = \int \left(1 - x + \frac{x^2}{3} + O(x^4) \right) dN(0,t)(x)$$

$$= 1 + \frac{t}{3} + O(t^2),$$

for $t \downarrow 0$. Plug in $\alpha = c\sqrt{2}$ and $\nu = 2$ to the first statement in Lemma 2.5, we have

$$g(c) = m_{c\sqrt{2}}(1) = \frac{c^2}{2} m_2(\frac{c^2}{2}) = 1 + \frac{c^2 + O(1)}{6},$$

completing the proof of (1.6). □

The last part of this section will be proving Theorem 1.3.

Proof. We start by extending the definition of X_k. Recall that we only defined X_k for $1 \leq k \leq N - 1$, now define $X_N := \log \frac{1-\beta_N}{\beta_1}$. With this extension, the sequence $\{X_k\}_{k \geq 1}$ satisfies that $(X_1, \ldots, X_N)$ and $(X_2, \ldots, X_N, X_1)$ equals in distribution. This means that the sequence is shift invariant. Notice that given $\sum_j B_j = \sum_j B'_j$ indicates that the collection of values of β_j's are the same as the collection of values of $(1 - \beta_j)$'s. Recall that we use $\tilde{\mathbb{P}}_N$ to denote the conditional probability of $\mathbb{P}_N(G_{(\delta)})$ conditioned on the equality $\sum_j B_j = \sum_j B'_j$ and

$$S_N = \sum_{j=1}^{N} (\log(1 - \beta_j) - \log(\beta_j)) = \sum_{j=1}^{N} \log(1 - \beta_j) - \sum_{j=1}^{N} \log(\beta_j),$$

then $\tilde{\mathbb{P}}_N(S_N = 0) = 1$. The sequence $\{X_k\}_{k \geq 0}$ is now shift invariant means:

$$(X_1, \ldots, X_N) \xrightarrow{\text{k shifts}} (X_{k+1}, \ldots, X_{k+N}),$$

$$(S_1, \ldots, S_N) \xrightarrow{\text{k shifts}} (S_{k+1} - S_k, \ldots S_{k+N} - S_k).$$

This indicates that we have the following "shift" as well:

$$\frac{1}{\sum\limits_{j=0}^{N-1} \exp(S_j)} \xrightarrow{\text{k shifts}} \frac{\exp(S_k)}{\sum\limits_{j=0}^{N-1} \exp(S_j)}.$$

Another conclusion from shift invariance is that

$$\mathbb{E}[f(X_1, \ldots, X_N)] = \frac{1}{N} \mathbb{E}\left[\sum_{k=1}^{N} f(X_{k+1}, \ldots, X_{k+N}) \right].$$

Thus

$$\mathbb{E}\left[\frac{1}{\sum\limits_{j=0}^{N-1} \exp(S_j)}\right] = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\exp(S_k)}{\sum\limits_{j=0}^{N-1} \exp(S_j)} = \frac{1}{N}.$$

This completes the proof of Theorem 1.3. $\qquad\square$

2.2 Proof of Theorem 1.2

In the previous section, we proved Theorem 1.1 and Theorem 1.3. Those theorems computes the asymptotics of the fixation probability when the value of δ is small. In this section, we will prove Theorem 1.2 for bigger δ values in two cases:

(a) Medium-sized δ: $\frac{1}{\sqrt{N}} \ll \delta \ll \frac{1}{\log(N)}$.

(b) Large δ: $\delta = o(\frac{1}{(\log(N))^\epsilon})$ for some $0 < \epsilon < 1$.

The key for proving Theorem 1.2 is a technique known as KMT coupling, a coupling of a random walk and a Brownian motion. We will start with two results from KMT coupling.

2.2.1 KMT Coupling

We first state the following result referenced from [7].

Lemma 2.7. *Suppose X_k is a simple random walk with i.i.d. increments ξ with $\mathbb{E}[\xi] = 0$, $\mathbb{E}[\xi^2] = 1$ and $\mathbb{E}[e^{t|\xi|}] < \infty$ for sufficiently small t. Define $X_t = X_{\lfloor t \rfloor}$, we are extending X_k to continuous time. For all $T > e$ and $\alpha < \infty$, there is a coupling*

24

of random walk X_k and standard Brownian motion B_t and a constant C_α so that

$$\mathbb{P}\left[\sup_{t\in[0,T]} |X_t - B_t| \geq C_\alpha \log(T)\right] \leq C_\alpha T^{-\alpha}. \qquad (2.22)$$

This is stated as Theorem 7.1.1 in [7] and proved in section 7.5 of [7]. This result was originally done by J. Komlós, P. Major and Tusnády in [6] in 1975. We will not include the detailed proof here, the general idea is to prove

$$\mathbb{P}(|S_j - B_j| \geq C_\alpha \log T) \leq C_\alpha T^{-\alpha}$$

by splitting the index j into two groups: first for $|j - i_k| < C_\alpha \log T$, find a sequence $\{i_k\}$ that partitions the interval $[0, T]$, show that the inequality holds for each segment, then apply triangle inequality; then the inequality for the remaining j values is proved using results from dyadic coupling.

The next lemma is a result on asymptotics of random walk hitting probability using the Brownian motion hitting probability result and the KMT coupling above.

Lemma 2.8. *Suppose that we have a random walk $\{S_n\}$ with i.i.d increments $\{X_n\}$ satisfying the assumptions in Lemma 2.7, then for $u \in [M^\epsilon, M^{\frac{1}{2}-\epsilon}]$ where $\epsilon \in [0, \frac{1}{2}]$, we have*

$$\mathbb{P}\left(\max_{1\leq t\leq M} S_t \leq u\right) \quad \sim \quad \sqrt{\frac{2}{\pi}}\frac{u}{\sqrt{M}}. \qquad (2.23)$$

Proof. Satisfying the assumptions in Lemma 2.7 indicates that we have a coupling

standard Brownian motion B_t. Let the Brownian motion run to time M. For any $u \in$ $(0, M^{1/2-\epsilon}$, we consider $\mathbb{P}(\sup_{t \in [0,M]} B_t \leq u)$, that is, the probability that the Brownian motion does not go beyond u. By the reflection principle, we have that

$$\mathbb{P}(\sup_{1 \leq t \leq M} B_t \leq u) = 1 - 2\mathbb{P}_0(B_M),$$

and

$$\mathbb{P}(\sup_{1 \leq t \leq M} B_t \leq u) \sim \sqrt{\frac{2}{\pi}} \frac{u}{\sqrt{M}}.$$

Substitute in this asymptotics to the result of Lemma 2.7, we have

$$\sqrt{\frac{2}{\pi}} \frac{u}{\sqrt{M}} + C_\alpha \log M - M^{-\alpha} \leq \mathbb{P}(\max_{1 \leq t \leq M} S_t \leq u) \leq \sqrt{\frac{2}{\pi}} \frac{u}{\sqrt{M}} + C_\alpha \log M + M^{-\alpha}.$$

Therefore

$$\mathbb{P}\left(\max_{1 \leq t \leq M} S_t \leq u\right) \sim \sqrt{\frac{2}{\pi}} \frac{u}{\sqrt{M}}.$$

$\square$

2.2.2 Results on Two Brownian Functionals

In order to prove Theorem 1.2 in the two cases mentioned above, we want to show that we can couple our random walk functionals with Brownian motion functionals and they behave the same way asymptotically. In this section, we will state the results for the two Brownian functionals that we need for the proof. Let B_s be a standard

Brownian motion, consider the following two Brownian functionals:

$$X_M := \frac{1}{\int_0^M \exp(B_s)ds},\tag{2.24}$$

$$Y_M := \frac{\max\{-B_M, 0\}}{\int_0^M \exp(2B_s)ds}.\tag{2.25}$$

In the remaining work of this survey, we will take the notation $Z^- = \max\{-Z, 0\}$ for simpler presentation. In this notation, we write the second functional above as

$$Y_M = \frac{B_M^-}{\int_0^M \exp(2B_s)ds}.$$

The first result is the uniform integrability of X_M.

Lemma 2.9. *The family of variables* $\left\{\frac{X_M}{\mathbb{E}X_M}\right\}_{M \geq 1}$ *is uniformly integrable.*

Proof. Observe the definition of X_M, putting it in the notation we used in (2.12) and (2.13), then X_M is precisely $(A_1(m))^{-1} = \left(\int_0^M \exp(B_s)\, ds\right)^{-1}$ and $\mathbb{E}[X_M] = m_1(M)$. Thus by the rescaling in Lemma 2.5 and the third result in Proposition 2.4, we have that $\mathbb{E}[X_M] = \Theta(M^{1/2})$. Then, we want to find a bound of $\mathbb{P}(X_M \geq t)$ so that when computing the expectation of $\frac{X_M}{\mathbb{E}[X_M]}$ we get the desired result.

Notice that the event $\{X_M \geq t\}$ can be interpreted as $\{\int_0^M \exp(B_s)ds \leq t^{-1}\}$. The probability that this event happens is maximized when B_t hits $-\log(t)$ within the first unit of time during the process and comes back below $-\log(t)$ within one unit of time. By the reflection principle, we have that

$$\mathbb{P}(\inf_{t \in [0,1]} B_t \leq -\log(t)) = 2\mathbb{P}(B_1 \leq -\log(t)).$$

Since $B_1 \sim N(0, 1)$, there is constant $C > 0$ such that the right hand side quantity is bounded above by $C \exp\left(\frac{1}{2}(-\log(t))^2\right)$. Conditioning on the event that B_t hits $-\log(t)$ for $t \in [0, 1]$, the hitting time is a stopping time τ, hence the strong Markov property states that $B_{\tau+t}$ for $t \in [0, 1]$ is independent of the natural filtration given B_τ. Then by Levy's arcsine law, we have that the probability that B_t stays above $-\log(t)$ after hitting $-\log(t)$ for less than one unit of time is $\frac{2}{\pi} \arcsin(\sqrt{1/M - \tau})$. Notice that $\arcsin(\sqrt{1/M - \tau})/(1/\sqrt{M}) \xrightarrow{M \to \infty} 1$ asymptotically. Combining the two probabilities, we have

$$\mathbb{P}(X_M \geq t) \leq C \exp\left(\frac{1}{2}(-\log(t))^2\right) \frac{1}{\sqrt{M}}.$$

Taking the limit as $t \to \infty$, we can see that $\exp\left(\frac{1}{2}(-\log(t))^2\right) = O(t^{-2})$. Then, using this probability we can compute the following to finish the proof

$$\mathbb{E}\left[\frac{X_M}{\mathbb{E}X_M} \mathbb{1}_{\frac{X_M}{\mathbb{E}X_M} \geq K}\right] \leq C\sqrt{M} \int_K^\infty t^{-2} M^{-1/2} \, dt \leq \frac{C}{K} \xrightarrow{K \to \infty} 0.$$

$\square$

The next lemma is to compute the asymptotics of the first and second moment of Y_M. In the computation of the second moment, we are going to use equation (2.19) stated in the proof of Theorem 2.4 from [9]:

28

Lemma 2.10. *As $M \to \infty$, we have the following two results:*

$$\mathbb{E}Y_M \quad \to \quad 1 \qquad\qquad (2.26)$$

$$\frac{\mathbb{E}Y_M^2}{\sqrt{M}} \quad \sim \quad 4\sqrt{\frac{2}{\pi}} \qquad\qquad (2.27)$$

Proof. We first compute $\mathbb{E}Y_M$ using the first result of Proposition 2.4, the same change of variables $y = \frac{x}{\sqrt{M}}$ and the same limit of $\lim_{M\to\infty} \frac{y^2 e^{y\sqrt{M}}}{\sinh(y\sqrt{M})} = 2y^2 \mathbb{1}_{y<0}$ we used to prove (2.16):

$$
\begin{aligned}
\mathbb{E}Y_M &= \mathbb{E}\left[\frac{B_M^-}{\int_0^M \exp(2B_s)\, ds}\right] \\
&= \int_{-\infty}^{\infty} -\frac{x^2 e^{-x}\mathbb{1}_{x\le 0}}{M\sinh(x)}\, dN(0,M)(x) \\
&= \int_{-\infty}^{\infty} -\frac{y^2 e^{y\sqrt{M}}\mathbb{1}_{x\le 0}}{\sinh(y\sqrt{M})}\, dN(0,1)(y) \\
&\xrightarrow{M\to\infty} \int_{-\infty}^{\infty} 2y^2 \mathbb{1}_{y<0}\, dN(0,1)(y) \\
&= \int_{-\infty}^{0} 2y^2 dN(0,1)(y) \\
&= 1.
\end{aligned}
$$

To compute the second moment, we take the statement (2.19) and treat it as a function of λ. Then its second derivative with respect to λ is what we needed:

$$\mathbb{E}\left[\left(\int_0^M \exp(2B_2)\, ds\right)^{-2} \Big| B_M = x\right] = \frac{e^{-2x}(x^2 \sinh(x) + Mx\cosh(x) - M\sinh(x))}{M^2 \sinh^3(x)}.$$

Then following a similar computation process as we did for the first moment, substituting in $y = \frac{x}{\sqrt{M}}$ and taking the limit in the integrand as $M \to \infty$, we have

$$
\frac{\mathbb{E}[Y_M^2]}{\sqrt{M}} =
$$

$$
= \int_{-\infty}^{\infty} \frac{e^{-2x}(x^4 \sinh(x) + Mx^3 \cosh(x) - Mx^2 \sinh(x))\mathbb{1}_{x\leq 0}}{M^{5/2} \sinh^3(x)} \, dN(0, M)(x)
$$

$$
= \int_{-\infty}^{\infty} \frac{e^{-2y\sqrt{M}}(y^4 M^2 \sinh(y\sqrt{M}) + M^{5/2}y^3 \cosh(y\sqrt{M}) - M^2 y^2 \sinh(y\sqrt{M}))\mathbb{1}_{y<0}}{M^{5/2} \sinh^3(y\sqrt{M})} \, dN(0, 1)(y)
$$

$$
\xrightarrow{M\to\infty} \int_{-\infty}^{\infty} -4y^3 \mathbb{1}_{y<0} \, dN(0, 1)(y)
$$

$$
= \int_{-\infty}^{0} -4y^3 d(0, 1)(y)
$$

$$
= 4\sqrt{\frac{2}{\pi}}.
$$

As $M \to \infty$, we have $\mathbb{E}[Y_M^2] \sim 4\sqrt{\frac{2}{\pi}}\sqrt{M}$. $\qquad\square$

For the last lemma of this section, we show the following result for $\mathbb{E}Y_M$.

Lemma 2.11. *For all events E satisfying $\mathbb{P}(E) \leq M^{-1/2-\epsilon}$ for some $\epsilon > 0$, $\mathbb{E}[Y_M \mathbb{1}_E] = o(1)$ as $M \to \infty$.*

Proof. With the first and second moments we computed in Lemma 2.10, for each M, we see that Y_M satisfies to have finite mean and finite non-zero variance, hence we can apply the Chebyshev's inequality to get that for some constant c,

$$
\mathbb{P}(Y_M \geq t) \leq c\frac{\sqrt{M}}{t^2}.
$$

Then for some $\epsilon > 0$, consider the event that $\{Y_M \geq M^{1/2+\epsilon}\}$

$$\mathbb{P}(Y_M \geq M^{1/2+\epsilon}) \leq c\frac{\sqrt{M}}{M^{1+2\epsilon}} = cM^{-1/2-2\epsilon},$$

then

$$\mathbb{E}[Y_M \mathbb{1}_{Y_M > M^{1/2+\epsilon}}] = \int_{M^{1/2+\epsilon}}^{\infty} \mathbb{P}(Y_M \geq M^{1/2+\epsilon}) dt \leq \int_{M^{1/2+\epsilon}}^{\infty} cM^{-1/2-2\epsilon} dt = cM^{-\epsilon} = o(1).$$

Let E_0 denote the event $\{Y_M \geq M^{1/2+\epsilon}\}$. Then for any event E satisfying $\mathbb{P}(E) \leq M^{-1/2-\epsilon}$, as $M \to \infty$, we have

$$\mathbb{E}[Y_M \mathbb{1}_E] = \mathbb{E}[Y_M \mathbb{1}_E \mathbb{1}_{E_0}] + \mathbb{E}[Y_M \mathbb{1}_E \mathbb{1}_{E_0^C}]$$

$$\leq o(1) + M^{1/2+\epsilon}\mathbb{P}(E)$$

$$= o(1),$$

completing the proof. □

2.2.3 Proof of Theorem 1.2 in Two Cases

In this section, we will prove Theorem 1.2 in the medium δ regime and the large δ regime.

2.2.3.1 Case I: Medium δ: $\frac{1}{\sqrt{N}} \ll \delta \ll \frac{1}{\log(N)}$.

Lemma 2.12. *Let E be an event such that $\mathbb{P}(E) \leq N^{-2}$, then for $\delta \in [\epsilon, 1 - \epsilon]$, there exists a constant C' such that*

$$e^{-C'\delta \log(N)} \mathbb{E}\left[\frac{\mathbb{1}_{E^C}}{N \int_0^1 \exp\left(\sqrt{N}\delta' B_t/\sqrt{2}\right) dt}\right] - \frac{1}{N^2} \leq \mathbb{E}[\mathbb{P}_N(G_1(\delta))] \qquad (2.28)$$

$$\leq e^{C'\delta \log(N)} \mathbb{E}\left[\frac{\mathbb{1}_{E^C}}{N \int_0^1 \exp\left(\sqrt{N}\delta' B_t/\sqrt{2}\right) dt}\right] + \frac{1}{N^2}.$$

Proof. Recall the definition of S_k and $\tilde{S}_k$ in Section 2.1. We expressed $\mathbb{P}_N(G_1(\delta))$ in terms of S_k in (2.2) and we have the relationship (2.3) between S_k and $\tilde{S}_k$. From (2.3), we have

$$-C_\epsilon \delta + S_k \leq \tilde{S}_k \leq C_\epsilon \delta + S_k.$$

Taking the exponential then sum over k, we have

$$\exp(-C_\epsilon \delta)\left[\sum_{k=1}^N \exp(S_k)\right] \leq \left[\sum_{k=1}^N \exp(\tilde{S}_k)\right] \leq \exp(C_\epsilon \delta)\left[\sum_{k=1}^N \exp(S_k)\right].$$

For $t \in [0, N]$, we extend $\tilde{S}_k$ to continuous time by letting $\tilde{S}_t = \tilde{S}_{\lceil t \rceil}$, then the finite sum $\sum_{k=1}^N \exp(\tilde{S}_k)$ is extended to being $\int_0^N \exp(\tilde{S}_t)\, dt$. Hence, the inequality above can be extended to continuous times as

$$\exp(-C_\epsilon \delta) \leq \mathbb{P}_N(\delta) \cdot \left[\int_0^N \exp(\tilde{S}_t)\, dt\right] \leq \exp(C_\epsilon \delta), \qquad (2.29)$$

where $\mathbb{P}_N(\delta) = 1/\left(\sum_{k=1}^N \exp(S_k)\right)$. Note that in this notation, $\mathbb{P}_N(G_1(\delta)) = \mathbb{E}[\mathbb{P}_N(\delta)]$.

In Section 2.1, by construction, $\tilde{S}_k$ is a simple random walk with step size δ' and holding probability $\frac{1}{2}$, then the variance of the i.i.d increments $\tilde{X}_k$ is $\frac{\delta'^2}{2}$ and the mean is 0. Then, $\frac{\sqrt{2}}{\delta'}\tilde{S}_k$ is a random walk whose i.i.d increments have mean 0 and variance 1. Then by the KMT coupling in Lemma 2.7, there is a Brownian motion on the same probability space as $\tilde{S}_k$ such that for $\alpha = 2$,

$$\mathbb{P}\left(\sup_{t\in[0,N]}\left|\frac{\sqrt{2}}{\delta'}\tilde{S}_t - B_t\right| \geq C_2\log(N)\right) \leq \frac{1}{N^2}.$$

Conditioning on the event $E_N := \left\{\sup_{t\in[0,N]}\left|\frac{\sqrt{2}}{\delta'}\tilde{S}_t - B_t\right| \geq C_2\log(N)\right\}$, inequality (2.29) then becomes:

$$\exp\left(-C_\delta - \frac{\delta'}{\sqrt{2}}C_2\log(N)\right) \leq \mathbb{P}_N(\delta)\cdot\left[\int_0^N \exp\left(\frac{\delta'}{\sqrt{2}}B_t\right)\,dt\right] \leq \exp\left(C_\epsilon\delta + \frac{\delta'}{\sqrt{2}}C_2\log(N)\right).$$

Since by definition $\delta' = 2\delta + O(\delta^2)$, there exists a constant C'' combining C_ϵ and C_2 such that

$$\exp\left(-C''\delta\log(N)\right) \leq \mathbb{P}_N(\delta)\cdot\left[\int_0^N \exp\left(\frac{\delta'}{\sqrt{2}}B_t\right)\,dt\right] \leq \exp\left(C''\delta\log(N)\right).$$

Now, apply Brownian scaling, we have

$$\exp\left(-C''\delta\log(N)\right) \leq \mathbb{P}_N(\delta)\cdot\left[N\int_0^1 \exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt\right] \leq \exp\left(C''\delta\log(N)\right).$$

That is

$$\exp\left(-C''\delta\log(N)\right)\frac{1}{\left[N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt\right]}\le\mathbb{P}_N(\delta)$$

$$\le\exp\left(C''\delta\log(N)\right)\frac{1}{\left[N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt\right]}.$$

Remember that the inequality above is conditioned on the event E_N with $\mathbb{P}(E_N)\le\frac{1}{N^2}$,

removing the conditioning and taking the expectation gives the desired inequality:

$$\exp\left(-C''\delta\log(N)\right)\mathbb{E}\left[\frac{\mathbb{1}_{E_N^C}}{N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt}\right]-\frac{1}{N^2}\le\mathbb{P}_N(G_1(\delta))$$

$$\le\exp\left(C''\delta\log(N)\right)\mathbb{E}\left[\frac{\mathbb{1}_{E_C^N}}{N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt}\right]+\frac{1}{N^2}.$$

$\square$

Now we can prove Theorem 1.2 for medium δ:

Proof. Given that $\frac{1}{\sqrt{N}}\ll\delta\ll\frac{1}{\log(N)}$, this implies that $\delta\log(N)\to 0$. Then, the result

from Lemma 2.12 becomes:

$$\mathbb{E}\left[\frac{\mathbb{1}_{E_N^C}}{N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt}\right]-\frac{1}{N^2}\le\mathbb{P}_N(G_1(\delta))\le\mathbb{E}\left[\frac{\mathbb{1}_{E_C^N}}{N\int_0^1\exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right)\,dt}\right]+\frac{1}{N^2}.$$

When $\mathbb{1}_{E_N^C} = 1$, notice that

$$\mathbb{E}\left[\frac{1}{N \int_0^1 \exp\left(\frac{\sqrt{N}\delta'}{\sqrt{2}}B_t\right) \, dt}\right] = \frac{1}{N}m_{\frac{\sqrt{N}\delta'}{\sqrt{2}}}(1),$$

then by the Brownian scaling result in Lemma 2.5 with $\alpha = \frac{\sqrt{N}\delta'}{\sqrt{2}}$, we have

$$\frac{1}{N}m_{\frac{\sqrt{N}\delta'}{\sqrt{2}}}(1) \sim \frac{\delta'}{2\sqrt{\pi N}} \sim \frac{\delta}{\sqrt{\pi N}}.$$

The term $\frac{1}{N^2}$ will not affect the asympototics, thus

$$\mathbb{P}_N(G_1(\delta)) \sim \frac{\delta}{\sqrt{\pi N}}.$$

When $\mathbb{1}_{E_N^C} = 0$, we only have the $1/N^2$ term for which the asymptotics still holds. $\square$

2.2.3.2 Case II: Large δ: $\delta = o\left(\frac{1}{(\log N)^\epsilon}\right)$.

We need to introduce a few new variables and notations. Let r be a real parameter and $r > 6$. Set $T = \lceil \delta^{-r} \rceil$. Since $\delta = o\left(\frac{1}{(\log N)^\epsilon}\right)$, then $\exp(\delta^{-r}) \approx \exp(N^{\epsilon r})$ and we can assume that $\delta^{-r} = o(N^s)$ for any positive s.

Let $Z := \tilde{S}_T$, where $\tilde{S}$ follows the same definition as in Section 2.1. Let A denote the partial sum of $\tilde{S}_k$ from $k = 1$ up to $K = T$, and let $\mathcal{F}_T = \sigma(\tilde{S}_1, \ldots, \tilde{S}_T)$ denote the natural filtration up to time T. We also follow the notation that $X^+ = \max\{X, 0\}$ denotes the positive part of X and $X^- = \min\{-X, 0\}$ denotes the negative part of X.

The idea of proving Theorem 1.2 in this regime is that we first show that the target probability has the following asymptotics depending on expectation of Z and A, then we compute the expectations with KMT coupling explicitly. The following lemma gives the asymptotic result:

Lemma 2.13.

$$\mathbb{P}_N(G_1(\delta)) = (1 + o(1))\frac{2}{\delta'\sqrt{\pi N}}\left(\mathbb{E}\frac{Z^-}{A} + O\left(\mathbb{E}\left[\frac{\delta^{-s}}{A}\right]\right)\right) + O(e^{-\delta^{-st}}). \qquad (2.30)$$

Proof. We prove this equality by proving a two-sided inequality. First, we want to show that $\mathbb{P}_N(G_1(\delta)) \geq (1 + o(1))\frac{2}{\delta'\sqrt{\pi N}}\left(\mathbb{E}\left[\frac{Z^-}{A}\right] + O\left(\mathbb{E}\left[\frac{\delta^{-s}}{A}\right]\right)\right)$. By (2.3), we see that as $\delta \to 0$, we can use $\tilde{S}_k$ to approximate S_k as they get uniformly close. Therefore, we can approximate the statement (2.2) with $\mathbb{E}[A^{-1}]$ by replacing S_k with $\tilde{S}_k$. Fix $s \in (\epsilon^{-1}, \frac{r}{2} - 1)$ and consider the event $G_1' = \{\max_{T \leq m \leq N} \tilde{S}_m \leq -\delta^{-s}\}$, then for conditional probability we have

$$\mathbb{P}_N(G(\delta)|\mathcal{F}_T) = \mathbb{P}_N(G_1(\delta)|A, Z)$$

$$\geq \mathbb{P}(G_1(\delta) \cap G_1'|A, Z)$$

$$\geq \frac{P_N(G'|A, Z)}{A + Ne^{-\delta^{-s}}}. \qquad (2.31)$$

We get the denominator by splitting up the summation $\sum_{k=1}^{N} \tilde{S}_k$ to A and the tail part indexed from T to N, then bounding by the bound in G_1'.

From Section 2.1, by construction $\tilde{S}_k$ is a simple random walk with mean 0 and

variance $\frac{\delta'^2}{2}$, we rescale it to get a random walk $\frac{\sqrt{2}}{\delta'}\tilde{S}_k$ with mean 0 and variance 1, so that we can apply the KMT coupling Lemma 2.7:

$$\mathbb{P}_N(G'|A, Z) \sim \sqrt{\frac{2}{\pi(N-T+1)}} \frac{\sqrt{2}}{\delta'} (-Z - \delta^{-s})^+ \sim \frac{2(Z+\delta^{-s})^-}{\delta'\sqrt{\pi N}}.$$

Plug this asyptotic into (2.31), we get

$$\mathbb{P}_N(G_1(\delta)|\mathcal{F}_T) \geq (1+o(1))\frac{2}{\delta'\sqrt{\pi N}}\frac{(Z+\delta^{-s})}{A+Ne^{-\delta^{-2}}} \sim (1+o(1))\frac{2}{\delta'\sqrt{\pi N}}\frac{(Z+\delta^{-s})}{A}.$$

Taking the expectation on the left hand side removes the conditional, and taking the expectation on the right hand side gives what we desired.

The second inequality we want to prove is

$$\mathbb{P}_N(G_1(\delta)) \leq (1+o(1))\frac{2}{\delta'\sqrt{\pi N}}\left(\mathbb{E}\left[\frac{Z^-}{A}\right] + O\left(\mathbb{E}\left[\frac{\delta^{-2}}{A}\right]\right)\right) + O(e^{-\delta^{-s}}).$$

Analogous to the proof above, here we consider conditioning on $G_1'' = \{\max_{T \leq m \leq N} \tilde{S}_m \leq \delta^{-s}\}$ and $G_1(\delta)$ on $\mathcal{F}_T$. $\qquad\square$

In order to get the exact asymptotics, we need to explicitly compute the two expectations in Lemma 2.13.

Lemma 2.14.

$$\mathbb{E}\left[\frac{Z^-}{A}\right] \quad \sim \quad \delta^2. \tag{2.32}$$

$$\mathbb{E}\left[\frac{1}{A}\right] \quad \sim \quad \frac{2}{\sqrt{\pi T}}. \tag{2.33}$$

Proof. Using the KMT coupling Lemma 2.7, for random walk $\frac{\sqrt{2}}{\delta'}\tilde{S}_t$, there exists a Brownian motion B_t such that

$$\mathbb{P}\left(\sup_{t\in[0,T]}\left|\frac{\sqrt{2}}{\delta'}\tilde{S}_t - B_t\right| \geq C_2\log(T)\right) \leq T^{-2},$$

and denote this event involved by E.

Then E^C is the event that $\left\{\sup_{t\in[0,T]}\left|\frac{\sqrt{2}}{\delta'}\tilde{S}_t - B_t\right| \leq C_2\log(T)\right\}$. In this event, we have that for certain $t \in [0,T]$,

$$C\delta'\log(T) + \frac{\delta'}{\sqrt{2}}B_t \leq \tilde{S}_t \leq C\delta'\log(T) + \frac{\delta'}{\sqrt{2}}B_t.$$

We do the same extension of $\tilde{S}_k$ as in the proof of Lemma 2.12 by letting $\tilde{S}_t = \tilde{S}_{\lceil t\rceil}$. We can write $A = \int_0^T \exp(\tilde{S}_t)\,dt$ and conditioning on E^C, we have

$$
\begin{aligned}
A &= \int_0^T \exp(\tilde{S}_t)dt \\
&= \int_0^T \exp\left(\frac{\delta'}{\sqrt{2}}B_t + O(\delta'\log(T))\right)dt.
\end{aligned}
$$

Since we defined $T = \delta^{-r}$, then $\delta'\log(T) = O(\delta\log(\delta)) \to 0$ for the regime we are in.

Thus, when conditioning on E^C,

$$A \sim \int_0^T \exp\left(\frac{\delta'}{\sqrt{2}} B_t\right) dt.$$

By Brownian scaling, we have

$$\int_0^T \exp\left(\frac{\delta'}{\sqrt{2}} B_t\right) dt = \int_0^T \exp(2 B_{(\delta')^2 t/2^3} dt = \frac{(\delta')^2}{2^3} \int_0^{\frac{(2^3}{(\delta')^2}T} \exp(2B_t) \, dt.$$

By Lemma 2.9, $\{\frac{X_M}{\mathbb{E}X_M}\}$ are uniformly integrable, then removing the conditional on E^C will not change the asymptotics.

Now consider the following two pieces:

$$\mathbb{E}[A^{-1} \mathbb{1}_E] = O(T^{-2}),$$

and by Proposition 2.4

$$
\begin{aligned}
\mathbb{E}[A^{-1} \mathbb{1}_{E^C}] &\sim (1 + o(1)) \mathbb{E}\left[\frac{1}{\frac{(\delta')^2}{2^3} \int_0^{\frac{2^3}{(\delta')^2}T} \exp(2B_t) \, dt}\right] \\
&\sim (1 + o(1)) \left(\frac{2^3}{(\delta')^2} \sqrt{\frac{2}{\pi \frac{2^3 T}{(\delta')^2}}}\right) \\
&= (1 + o(1)) \left(\frac{2^3}{(\delta')^2} \cdot \frac{\sqrt{2}}{\sqrt{\pi T}} \cdot \frac{\delta'}{2\sqrt{2}}\right) \\
&= (1 + o(1)) \frac{4}{\delta' \sqrt{\pi T}}.
\end{aligned}
$$

Recall that $\delta' \sim 2\delta$. Combine the two results above, we have

$$\mathbb{E}\left[\frac{1}{A}\right] \mathbb{E}[A^{-1}\mathbb{1}_E] + \mathbb{E}[A^{-1}\mathbb{1}_{E^C}] \sim \frac{2}{\delta\sqrt{\pi T}}.$$

Analogous to the proof above, to compute $\mathbb{E}\left[\frac{Z^-}{A}\right]$, we consider the two pieces $\mathbb{E}\left[\frac{Z^-}{A}\mathbb{1}_E\right]$ and $\mathbb{E}\left[\frac{Z^-}{A}\mathbb{1}_{E^C}\right]$. Via the same computation for $\tilde{S}_k$ and Brownian scaling, we have

$$
\begin{aligned}
\mathbb{E}\left[\frac{Z^-}{A}\right] &= \mathbb{E}\left[\frac{Z^-}{A}\mathbb{1}_E\right] + \mathbb{E}\left[\frac{Z^-}{A}\mathbb{1}_{E^C}\right] \\
&= O(T^{-2}) + \mathbb{E}\left[\frac{\frac{\delta'}{\sqrt{2}}B_T^-\mathbb{1}_{E^C}}{\int_0^T \frac{\delta'}{\sqrt{2}}B_s\ ds}\right] + O(\delta\log(T)\mathbb{E}[m_{\delta'/\sqrt{2}}(T)]) \\
&= O(\frac{\log(T)}{\sqrt{T}}) + \frac{\delta'}{\sqrt{2}}\mathbb{E}\left[\frac{\delta'}{\sqrt{2^3}}Y_{2^3/(\delta')^2 T}\mathbb{1}_{E^C}\right] \\
&= O(\frac{\log(T)}{\sqrt{T}}) + \frac{(\delta')^2}{2}\mathbb{E}\left[Y_{2^3/(\delta')^2 T}\mathbb{1}_{E^C}\right].
\end{aligned}
$$

The event E satisfies the requirements in the statement of Lemma 2.11, hence

$$\mathbb{E}[Y_{2^3/(\delta')^2 T}\mathbb{1}_{E^C}] = \mathbb{E}[Y_M] + o(1).$$

Then, by the results on moments in Lemma 2.10, we have

$$\mathbb{E}[Y_{2^3/(\delta')^2 T}\mathbb{1}_{E^C}] = 1 + o(1).$$

Substituting this in to the expression of $\mathbb{E}\left[\frac{Z^-}{A}\right]$ while having $\delta' \sim 2\delta$,we have

$$\mathbb{E}\left[\frac{Z^-}{A}\right] = O(\frac{\log(T)}{\sqrt{T}}) + (1 + o(1))\frac{(\delta')^2}{2} \sim \delta^2.$$

$\square$

The final piece of this section is to prove Theorem 1.2 in the large δ regime:

Proof. Plugging in the result of Lemma 2.14 to the result of Lemma 2.13 yields:

$$\mathbb{P}_N(G_1(\delta)) = (1 + o(1))\frac{2}{\delta'\sqrt{\pi N}}\left(\delta^2 + O\left(\frac{2\delta^{-s}}{\delta\sqrt{\pi T}}\right)\right) + O(e^{-\delta^{-s}})$$

$$= (1 + o(1))\frac{2}{2\delta\sqrt{\pi N}}\left(\delta^2 + O\left(\frac{2\delta^{-s}}{\delta\sqrt{\pi T}}\right)\right) + O(e^{-\delta^{-s}}).$$

Recall that we picked $s \in (1/\epsilon, r/2 - 1)$, hence $s > 1$. And we assumed $r > 6$, so $T = \lceil \delta^{-r} \rceil = \Omega(\delta^{-6})$, then the equation above simplifies to

$$\mathbb{P}_N(G_1(\delta)) = (1 + o(1))\frac{1}{\delta\sqrt{\pi N}}\left(\delta^2 + O\left(\frac{1}{\delta\sqrt{T}}\right)\right) + O(e^{-\delta^{-s}})$$

$$\sim \frac{\delta}{\sqrt{\pi N}},$$

completing the proof of Theorem 1.2.

$\square$

3   Future Work

There are many things to be done along the line of finding explicit asymptotics of fixation probability for mutants in stochastically selectively neutral environments. As for directions of possible future work, one may start with one of the following three cases: (1) generalize the existing line model, (2) working with a loop with N vertices; (3) working with a complete graph with N vertices.

3.1   Going Further from the Line Model

Thinking about the evolution on a loop with N vertices after replacing one of the sites with a mutant, intuitively, we see that instead of having only one interface as the line model we mainly studied in this paper, the evolution starts with two interfaces. This nature gives the evolution process 2 degrees of freedom, as mutants can attack or being attacked from two directions. One possible way to start working on this loop model is to start with a slight modification on our line model: instead of having the initial configuration as mutant on site 1 and regular cells on the rest, we can start with the initial configuration where the first mutant is place at site 2 while site $\{1, 3, 4, 5, \cdots, N\}$ are regular cells. Ideally, if we can find the asymptotics of the fixation probability on this new line model, and with some more work identifying site 1 and N, we shall be able to find the asymptotics of the fixation probability on the loop model.

Here we start with construction the new line model. Fix $\delta \in (0, 1)$ and take the

underlying probability space to be $(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)$. We define the collection of independent Rademacher random variables $B_1, \cdot, B_N, B'_1, \cdots, B'_N$, the rate 1 Poisson process $\xi_t^{(i,j)}$, and the random quantities μ_k, ν_k and β_k in the exact same way as we did in the line model in Section 1.2.

Notice that if an mutant-regular interface is at site k, the only edges that matter are $k \to k+1$ and $k \to k-1$. A mutant at site k takes over the regular cell at site $k+1$ with probability β_{k+1} and the number of mutants increases by 1. A regular cell at site $k+1$ takes over the mutant cell at site k with probability $1 - \beta_k$ and the number of mutants decreases by 1. We also have similar statements for a regular-mutant interface.

If we start with a single mutant at site 2, after some time for evolution, we have three possible configurations: (1) all mutants die out; (2) mutants take over the world; (3) or a mixed configuration taking the pattern "R...RM...MR...R", two sequences of regular cells(R) with a string of mutant(M) in the middle, total length of sequence is N. Here, configurations (1) and (2) are the two absorbing states.

Consider the state space $\mathbb{S} = \{1, \ldots, N\} \times \{1, \ldots, N\}$. The first coordinate represents the beginning position of the sequence of mutants and the second coordinate represents the ending position of the sequence of mutants. Our initial configuration here in this case is $(2, 2)$ and our absorbing state for mutants to take over the world is $(1, N)$. For any state (k, l) with $l < k$, we claim those are also states when the mutants die out.

For $(k, l) \in \mathbb{S}$ where $2 \leq k \leq l \leq N - 1$, the only relevant positions are: $(k, l + 1)$, $(k, l - 1)$, $(k + 1, l)$ and $(k - 1, l)$. With the setup and notations above, we have the transition probabilities:

$$p((k, l), (k, l + 1)) = \frac{\beta_{l+1}}{\beta_{l+1} + (1 - \beta_l) + (1 - \beta_k) + \beta_{k-1}}, \tag{3.1}$$

$$p((k, l), (k, l - 1)) = \frac{1 - \beta_l}{\beta_{l+1} + (1 - \beta_l) + (1 - \beta_k) + \beta_{k-1}}, \tag{3.2}$$

$$p((k, l), (k + 1, l)) = \frac{1 - \beta_k}{\beta_{l+1} + (1 - \beta_l) + (1 - \beta_k) + \beta_{k-1}}, \tag{3.3}$$

$$p((k, l), (k - 1, l)) = \frac{\beta_{k-1}}{\beta_{l+1} + (1 - \beta_l) + (1 - \beta_k) + \beta_{k-1}}. \tag{3.4}$$

Further, we also need to define the distribution when one end of the sequence of mutants is absorbed. It is worth noting that if site 1 becomes a mutant, it will not be reversed to a regular cell unless site 2 becomes a regular cell and reproduce to take it over. If that happens, the mutants die out. The same scenario holds for site N: if site N ever becomes a mutant, the only way for it to become a regular cell again is when $N - 1$ is regular and reproduce to take it over, and this only happens when mutants dies out. This is due to the nature that the sequence of mutants is always going to be connected given our setup.

Suppose that some point in the evolution, the mutant cell at site 2 takes over the

regular cell at site 1, we have $k = 1$ and $3 \leq l \leq N - 1$, with

$$p((1, l), (1, l + 1)) = \frac{\beta_{l+1}}{\beta_{l+1} + (1 - \beta_l}), \tag{3.5}$$

$$p((1, l), (1, l - 1)) = \frac{1 - \beta_l}{\beta_{l+1} + (1 - \beta_l)}. \tag{3.6}$$

Suppose that at some point in the evolution, the mutant cell at site $N - 1$ takes over the regular cell at site N, we have $2 \leq k \leq N - 2$ and $l = N$, with

$$p((k, N), (k + 1, N)) = \frac{1 - \beta_k}{(1 - \beta_k) + \beta_{k-1}}, \tag{3.7}$$

$$p((k, N), (k - 1, N)) = \frac{\beta_{k-1}}{(1 - \beta_k) + \beta_{k-1}}. \tag{3.8}$$

Notice that as soon as one the left end point of the sequence of mutants reach site 1 or the right end point of the mutants reach site N, we automatically get into an intermediate configuration in the line model in section 1.2 and we should be able to use those results from there.

Let $G_2 = G_2(\delta)$ denotes the event that the mutant takes over the world given that initial mutant is placed at site 2. Let M_2 denote the event that the mutant on site 2 takes over the regular cell at site 1 before the regular cell at site 1 takes over the mutant at site 2. Claim that $\mathbb{P}_N(G_2) \leq \mathbb{P}_N(G_2|M_2)$, the conditional probability gives an upper bound of the actual fixation probability.

Since $\mathbb{P}_N(G_2|M_2) = \frac{\mathbb{P}_N(G_2 \cap M_2)}{\mathbb{P}_N(M_2)}$, in order to express the conditional probability explicitly, we compute $\mathbb{P}_N(M_2)$ and $\mathbb{P}_N(G \cap M_2)$ separately.

45

Recall that $G_1(\delta, N)$ denotes the event that mutants take over the world in the original line model in section 1.2. Let $G_1^l(\delta)$ denote the event that mutants take over the world in our original line model with mutants on sites $\{1, 2, \cdots, l\}$ as the starting configuration. Here, $\{1, 2, 3, \cdots, l\}$ is an intermediate configuration in the original line model and it is also an intermediate configuration in our new line model. Let R_l denote the event that the right end point of the sequence of mutants is at position l for $1 \leq l \leq N$. In our new setup conditioning on M_2, when R_l happens, we always have the left end point of the sequence of mutants at position 2.

Considering the event $G \cap M_2$, we can write this event as a discrete union of sub-events based on the position of the right end point of the sequence of mutants when site 2 takes over site 1.

$$
\begin{aligned}
\mathbb{P}_N(G_2 \cap M_2) &= \sum_{l=2}^{N} p((2, l), (1, l)) \mathbb{P}_N(\text{right end point of mutants at position l}) \mathbb{P}_N(G_1^l(\delta)) \\
&= \sum_{l=2}^{N} p((2, l), (1, l)) \mathbb{P}(R_l) \mathbb{P}_N(G_1^l(\delta, N)). \quad (3.9)
\end{aligned}
$$

Consider the probability $\mathbb{P}_N(M_2)$, the probability that site 2 takes over site 1 before site 1 takes over site 2. By the probability distribution we define above, we see that the probability at a step in the evolution that site 2 successfully takes over site 1 depends on l, the location of the right-hand ending point of the sequence of mutants. So we can write this event as a disjoint union based on the location of the right-hand

46

ending point of the sequence of mutants:

$$\mathbb{P}_N(M_2) = \sum_{l=2}^{N} \mathbb{P}_N(\text{right end point of mutants at l when site 2 takes over site 1})$$

$$= \sum_{l=2}^{N} \mathbb{P}_N(\text{right end point of mutants at position l})p((2,l),(1,l))$$

$$= \sum_{l=2}^{N} \mathbb{P}_N(R_l)p((2,l),(1,l)). \tag{3.10}$$

Then we have

$$\mathbb{P}_N(G_2|M_2) = \frac{\mathbb{P}_N(G_2 \cap M_2)}{\mathbb{P}_N(M_2)} = \frac{\sum_{l=2}^{N} \mathbb{P}_N(R_l)p((2,l),(1,l))\mathbb{P}_N(G_1^l(\delta))}{\sum_{l=2}^{N} \mathbb{P}(R_l)p((2,l),(1,l))}. \tag{3.11}$$

Corollary 3.1. $\mathbb{P}_N(G_2|M_2)$ *is an upper bound of* $\mathbb{P}_N(G_2)$.

To show that $\mathbb{P}_N(G_2|M_2)$ is an upper bound, it suffices to show that

$$\mathbb{P}_N(G_2|M_2) - \mathbb{P}(G_2) \geq 0.$$

Since

$$\mathbb{P}_N(G_2|M_2) = \frac{\mathbb{P}_N(G_2 \cap M_2)}{\mathbb{P}_N(M_2)},$$

then

$$\mathbb{P}_N(G_2|M_2) - \mathbb{P}_N(G_2) = \frac{\mathbb{P}_N(G_2 \cap M_2)}{\mathbb{P}_N(M_2)} - \mathbb{P}_N(G_2) = \frac{\mathbb{P}_N(G_2 \cap M_2) - \mathbb{P}_N(G_2)\mathbb{P}_N(M_2)}{\mathbb{P}_N(M_2)}.$$

Substituting in the expressions in equation (3.9) and (3.10), we have

$$\mathbb{P}_N(G_2|M_2) - \mathbb{P}_N(G_2) = \sum_{l=2}^{N} [p((2,l),(1,l))\mathbb{P}(R_l)] \left[\mathbb{P}_N(G_1^l(\delta) - \mathbb{P}_N(G_2))\right].$$

To prove Corollary 3.1, it suffices to show that $\mathbb{P}_N(G_1^l(\delta)) \geq \mathbb{P}_N(G_2)$.

Lemma 3.2. *On a line with N sites, mutants are placed in a sequence starting at site 1, then the more mutant we place, the easier for the mutants to take over the world.*

Proof. For any given position $1 \leq l < N$, starting with the first l sites being mutants, the number of mutants will increase to $l+1$ at some point in order for the mutants to take over the world. This is saying that

$$\mathbb{P}_N(G_1^l(\delta)) = \mathbb{P}_N(\text{first get } l+1 \text{ mutants then take over the world})$$

$$= \mathbb{P}_N(\text{mutant at site } l \text{ takes over site } l+1 \text{ at some point})\mathbb{P}_N(G^{l+1}(\delta))$$

$$\leq \mathbb{P}_N(G_1^{l+1}(\delta)).$$

Notice that the position of the right end point of mutants forms a Markov chain as it only depends on the previous position, and the event $\{l \text{ takes over } l+1\}$ and the conditional event $\{\text{mutant takes over the world given } l+1 \text{ mutants}\}$ are independent by the Markov property. We yield the last inequality from the fact that $\mathbb{P}_N(l \text{ take over } l+1) \leq 1$. □

This shows that as l increases, $\mathbb{P}_N(G_1^l(\delta))$ increases. Then, to show that $\mathbb{P}_N(G_1^l(\delta)) \geq \mathbb{P}_N(G_2)$, it suffices to show that $\mathbb{P}_N(G_1^2(\delta)) \geq \mathbb{P}_N(G_2)$.

Lemma 3.3. *Let $G_1^2(\delta, k)$ denote the event that the mutant takes over the first k sites when we begin with mutants at sites $\{1, 2\}$, then*

$$\mathbb{P}_N(G_1^2(\delta, k)) \geq \mathbb{P}_N(G_1^2(\delta, k+1))$$

for $2 \leq k \leq N - 1$.

Proof. Analogous to the proof of Lemma 3.2, when starting the process with site 1 and 2 being mutants, in order to take over $k + 1$ mutants in the world, the number of mutants will take over k mutants at some point before taking over $k + 1$ mutants. Hence

$$\mathbb{P}_N(G_1^2(\delta, k+1)) = \mathbb{P}_N(\text{take over the first } k \text{ sites then take over site } k+1)$$
$$= \mathbb{P}_N(\text{site } k \text{ take over site } k+1 \text{ at some point})\mathbb{P}_N(G_1^2(\delta, k))$$
$$\leq \mathbb{P}_N(G_1^2(\delta, k)).$$

$\square$

Let $\mathbb{P}_k(G_2)$ denote the probability that the mutant takes over the first k sites, then in order to take over $k + 1$ sites, we need to first take over k sites. This shows that

$$\mathbb{P}_k(G) \geq \mathbb{P}_{k+1}(G_2).$$

Lemma 3.4. *For all $2 \leq n \leq N$, $\mathbb{P}(G_1^2(\delta, n)) \geq \mathbb{P}_n(G_1)$, that is, when we place two mutants at site 1 and 2, it is easier for for the mutants to take over the world, comparing with only placing one mutant at site 2.*

Proof. The two processes in question are Markov chains, and we can prove the lemma with Markov Chain coupling.

For any state in the two processes, we can describe the state with the set of site numbers occupied by the mutants. Let $\{X_n\}$ denote the process where we start with two mutants at site 1 and 2, then $X_0 = \{1, 2\}$. Let $\{Y_n\}$ denote the process where we start with only one mutant at site 2, then $Y_0 = \{2\}$.

We define the following Markov chain coupling $X_n \times Y_n$: among the sites shared by X_k and Y_k, we uniformly pick a site and then we uniformly pick a direction to attack. For example, going from $X_0 \times Y_0$ to $X_1 \times Y_1$, only site 2 can be picked and if the attacking direction is to the right, we have $X_1 \times Y_1 = \{1, 2, 3\} \times \{2, 3\}$; or if the attacking direction is to the left, we have $X_1 \times Y_1 = \{1, 2\} \times \{1, 2\}$. By induction, we can see that following this rule of coupling $Y_k \subseteq X_k$ for all k, if we discretize by only sampling at the times where the configuration changes. Then, by the time $Y_l = \{1, 2, \cdots, N\}$, since $Y_l \subseteq X_l$, X_l must be at state $\{1, 2, 3, \cdots, N\}$ already, completing the proof. $\qquad\square$

The proof of Corollary 3.1 follows from combining the results of Lemma 3.2, 3.3 and 3.4.

The modified line model with N vertices and initial mutant at site 2 may be worth further discussion and it may eventually contribute to computing the asymptotics of the fixation probability for mutants on other geometries like the loop with N vertices. The lemmas in this section would possibly serve as the starting point for those works in the future.

3.2 Thoughts on Complete Graph

In this section, we will state some thoughts and observations on the assymptotics of the fixation probability of mutants on a complete graph with N vertices. We start with the geometry of a complete graph with N vertices. And the model is setup as follows. Fix $\delta \in (0,1)$ and let $B_1 \ldots B_n$ and $B'_1 \ldots B'_N$ be independent Rademacher random variables. At site k, the regular fitness is given by $1 + \delta B_k$ and the mutant fitness at site k $1 + \delta B'_k$. We then define three types of sites:

(a) Sites in favor of mutants: sites with mutant fitness $1 + \delta$ and regular fitness $1 - \delta$.

(b) Sites in favor of regular cells: sites with mutant fitness $1 - \delta$ and regular fitness $1 + \delta$.

(c) Sites indifferent of either type: site with mutant fitness $1 - \delta$ and regular fitness $1 - \delta$, or sites with mutant fitness $1 + \delta$ and regular fitness $1 + \delta$.

From this setup, we also know the following:

- $\mathbb{P}(\text{Mutant favoring site taken by a mutant}) = \frac{1+\delta}{2}$.

$\mathbb{P}(\text{A mutant dies at mutant favoring site}) = \frac{1-\delta}{2}$.

- $\mathbb{P}(\text{Indifferent site taken by a mutant}) = \frac{1}{2}$.

  $\mathbb{P}(\text{A mutant dies at an indifferent site}) = \frac{1}{2}$.

- $\mathbb{P}(\text{regular cell favoring site taken by a mutant}) = \frac{1-\delta}{2}$.

  $\mathbb{P}(\text{A mutant dies at regular cell favoring site}) = \frac{1+\delta}{2}$.

Once we start an evolution process, we have a fixed environment, say we have A mutant favoring sites, B indifferent sites and C regular cell favoring sites such that $A + B + C = N$. The initial state of the evolution process is that a single mutant is placed at one of sites selected uniformly random. During the evolution, a site is chosen uniformly random and tries to take over a neighboring site connect by an edge at rate 1. This is a continuous time Markov Process and we discretize it by sampling only at the times when the configuration changes. Suppose after l steps, there are still mutants in the system and we reach the state (x, y, z) in the Markov process, where $1 \leq x < A$ is the number of mutants on the mutant favoring sites, $1 \leq y < B$ is the number of mutants occupying indifferent sites and $1 \leq z < C$ is the number of mutants occupying regular cell favoring sites. We assume that this is an intermediate state, not an absorbing state, thus $1 \leq x + y + z < N$. There are only 6 possible states that could be reached from (x, y, z):

- $(x, y, z) \to (x + 1, y, z)$:

  This is when one of the $(x + y + z)$ mutants take over a mutant favoring site.

The transition probability

$$p((x, y, z), (x + 1, y, z)) = \frac{x + y + z}{N} \cdot \frac{A - x}{N} \cdot \frac{1 + \delta}{2}.$$

- $(x, y, z) \rightarrow (x - 1, y, z)$:

  This is when one of the x mutants on mutant favoring sites dies. The transition probability:

  $$p((x, y, z), (x - 1, y, z)) = \frac{N - (x + y + z)}{N} \cdot \frac{x}{N} \cdot \frac{1 - \delta}{2}.$$

- $(x, y, z) \rightarrow (x, y + 1, z)$:

  This is when one of the $(x + y + z)$ mutants take over an indifferent site. The transition probability:

  $$p((x, y, z), (x, y + 1, z)) = \frac{x + y + z}{N} \cdot \frac{B - y}{N} \cdot \frac{1}{2}.$$

- $(x, y, z) \rightarrow (x, y - 1, z)$:

  This is when one of the y mutants on indifferent sites dies. The transition probability:

  $$p((x, y, z), (x, y - 1, z)) = \frac{N - (x + y + z)}{N} \cdot \frac{y}{N} \cdot \frac{1}{2}.$$

- $(x, y, z) \rightarrow (x, y, z + 1)$:

53

This is when one of the $(x + y + z)$ mutants take over a regular cell favoring site. The transition probability:

$$p((x, y, z), (x, y, z + 1)) = \frac{x + y + z}{N} \cdot \frac{C - z}{N} \cdot \frac{1 - \delta}{2}.$$

- $(x, y, z) \to (x, y, z - 1)$:

This is when one of the z mutants on regular cell favoring sites dies. The transition probability:

$$p((x, y, z), (x, y, z - 1)) = \frac{N - (x + y + z)}{N} \cdot \frac{z}{N} \cdot \frac{1 + \delta}{2}.$$

We can recognize this process as a random walk in the rectangular parallelepiped of size $A \times B \times C$ with $A + B + C = N$. We can identify the initial configuration by $(1, 0, 0)$ or $(0, 1, 0)$ or $(0, 0, 1)$ each with probability $\frac{1}{3}$. The two absorbing states are $(0, 0, 0)$ representing the mutants die out, or (A, B, C) representing the mutants take over all N sites. Notice that there are many cases that are worth considering here: for small N; for very large N and $N \to \infty$ eventually; $A = C$, the system is overall neutral; $A > C$, the system favors mutants overall; $A < C$, the system favors regular cells overall.

At time k, we denote the current configuration by $L_k = (x_k, y_k, z_k)$ with the transition probability of going to the next configuration as given above. Denote the upcoming change by ΔL_k and it should be worth it to study the expected value of

54

the upcoming change $\mathbb{E}[\Delta L_k | L_k]$. It is reasonable to consider the rate of change at time k site L_k as

$$dL_k = \frac{1}{N}\phi(L_k)dt + \frac{1}{N}dB_t,$$

where $\phi(L_k)$ is function of x_k, y_k and z_k, hence $\frac{1}{N}\phi(X_k)dt$ is the determinist piece and $\frac{1}{N}dB_t$ is the stochastic piece. From the transition probabilities, we see that the determinist change in each direction can be computed as a weighted sum:

$$\frac{dx}{dt} = A(x_k + y_k + z_k) + Nx_k(1 - \delta) + \delta(x_k + y_k + z_k)(A - 2x_k),$$

$$\frac{dy}{dt} = B(x_k + y_k + z_k) - Ny_k,$$

$$\frac{dz}{dt} = C(x_k + y_k + z_k) - Nz_k(1 + \delta) + \delta(x_k + y_k + z_k)(-C + 2z_k).$$

We plot a few examples of this vector field with $\delta = \frac{1}{2}$:

(a) For small N = 5:
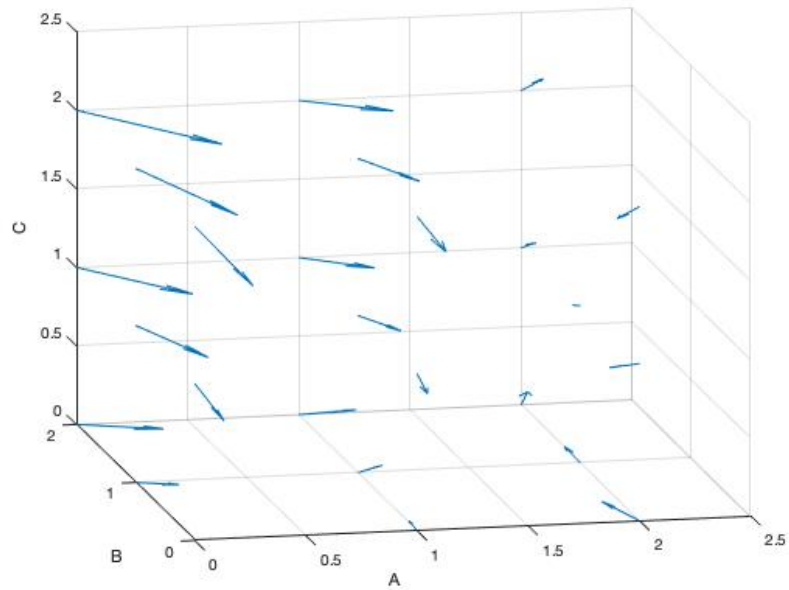
Figure 1: Vector field for small N when $A = C$. $A = 2, \ B = 1, \ C = 2$.
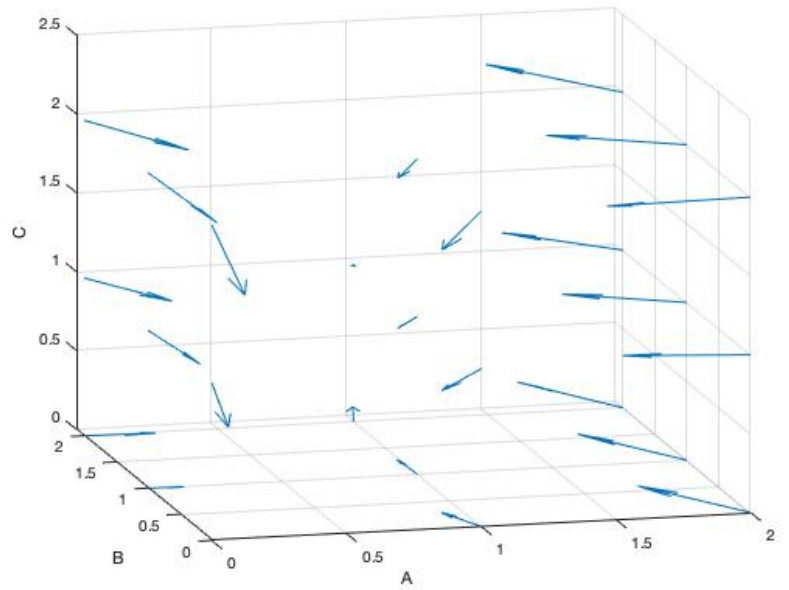
Figure 2: Vector field for small N when $A < C$. $A = 1, \ B = 2, \ C = 2$.
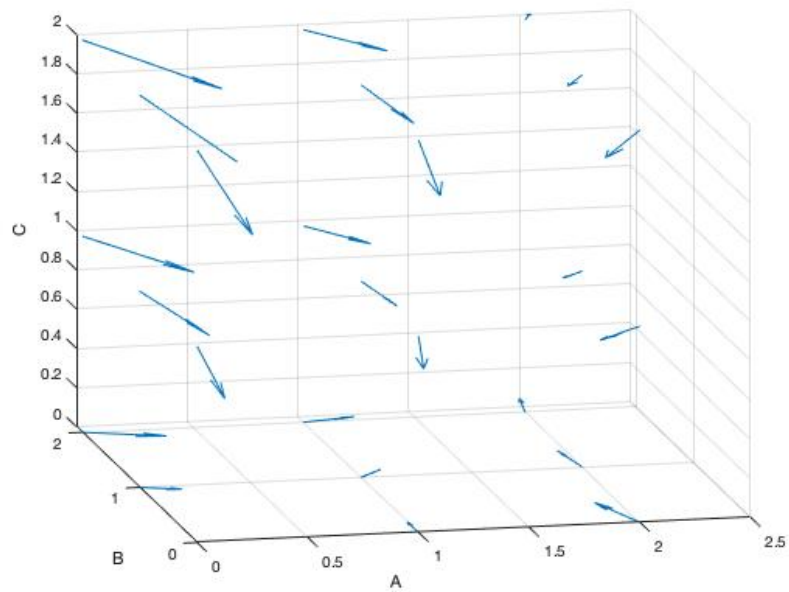
Figure 3: Vector field for small N when $A > C$. $A = 2$, $B = 2$, $C = 1$.

(b) For big $N = 200$ (figures below zoomed in to absorbing state (A, B, C))
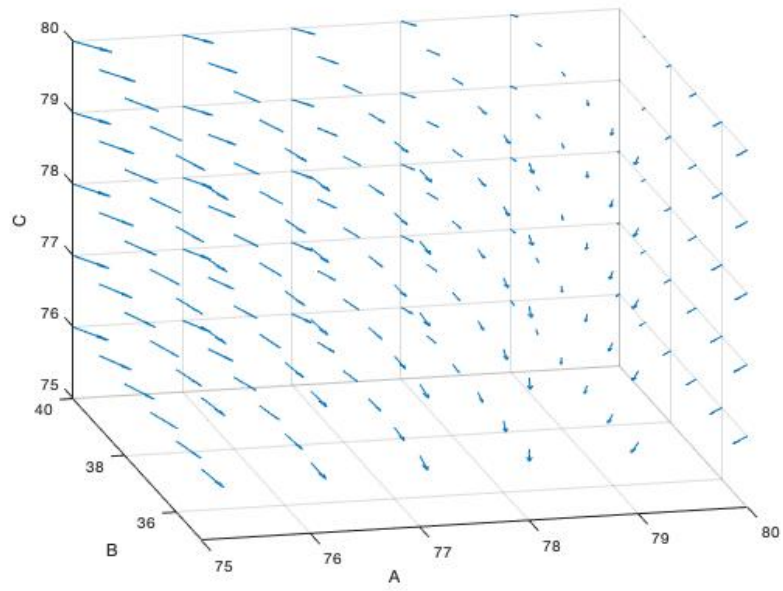
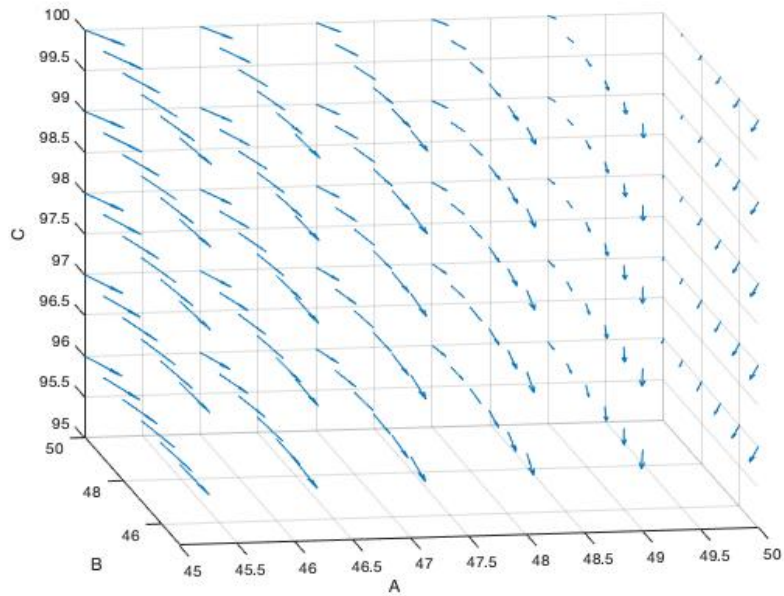Figure 4: Vector field for large N when $A = C$. $A = 80$, $B = 40$, $C = 80$.

Figure 5: Vector field for large N when $A < C$. $A = 50$, $B = 50$, $C = 100$.
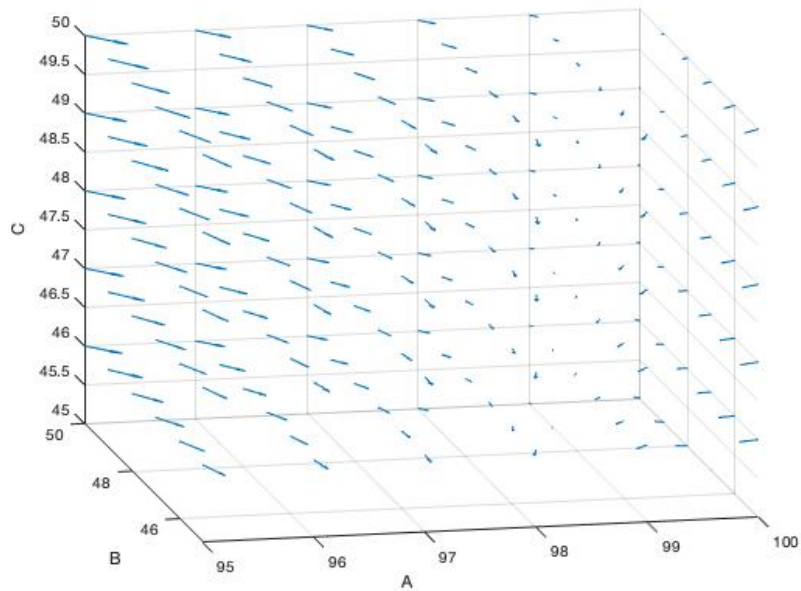
Figure 6: Vector field for large N when $A > C$. $A = 100$, $B = 50$, $C = 50$.

(c) This is a case that might arise from random fluctuation: $A = B - d$, $C = B - d$ where $d \sim \sqrt{B}$. When N is large, A, B and B will be very close, but random fluctuation indicates that since N is large, there are many steps to take in the process, and it is possible that the more steps to take, the further the process deviates from the expected result.
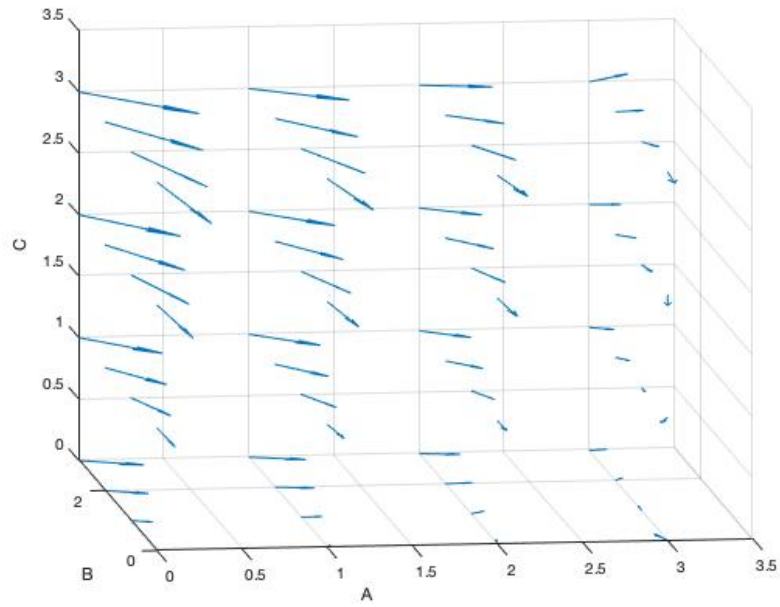
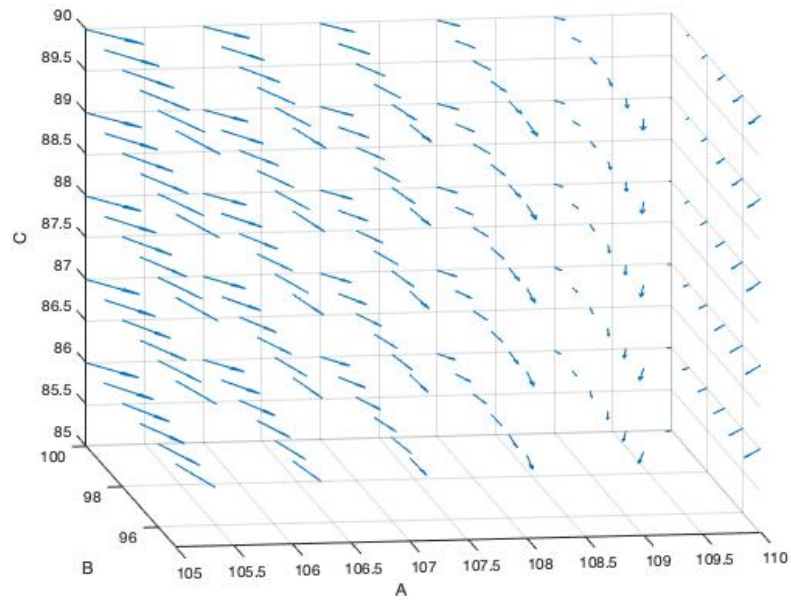Figure 7: Vector field for $N = 6$. $A = 2 + d,\ B = 2,\ C = 2 - d$ and $d \sim \sqrt{B}$.

Figure 8: Vector field for $N = 210$. $A = 70 + d,\ B = 70,\ C = 70 - d$ and $d \sim \sqrt{B}$.

Notice that when A and C are independent of B, for larger N, the vector field follows a much clearer pattern. We see that for larger N the vector field tends to the absorbing state (A, B, C) while for small N it is somewhat unclear. In the case where A and C depends on B, this is a case that corresponds to what might arise from random fluctuation. One thing might be worth considering in the future is to include the effect of the stochastic piece on the vector field: does it help with the convergences to the absorbing state or not. Another thing to consider is the effect of the prefixed value δ and whether its relationship with the size of the population N poses and effect on the vector field. A wild guess based on the result of Theorem 1.3 is that: in the case where the environment is overall neutral, that is when $A = C$, the fixation probability for the mutant to take over a complete graph of N vertices is $\frac{1}{N}$.

3.3   Questions for the Future

It is believed by the authors in [3] that the condition for $\delta = o\left((\log N)^{-\epsilon}\right)$ for some positive ϵ can be removed, leading to the following conjecture:

Conjecture 3.5. *If $\delta\sqrt{N} \to \infty$ and $\delta \to 0$, then*

$$\frac{\mathbb{P}_N(G_1(\delta))}{\delta N^{-1/2}} \to \frac{1}{\sqrt{\pi}}. \tag{3.12}$$

In any case, we shall have

$$0 < \frac{C_1 \delta}{\sqrt{N}} \leq \mathbb{P}_N(G_1(\delta)) \leq \frac{C_2 \delta}{\sqrt{N}} \tag{3.13}$$

as $N \to \infty$.

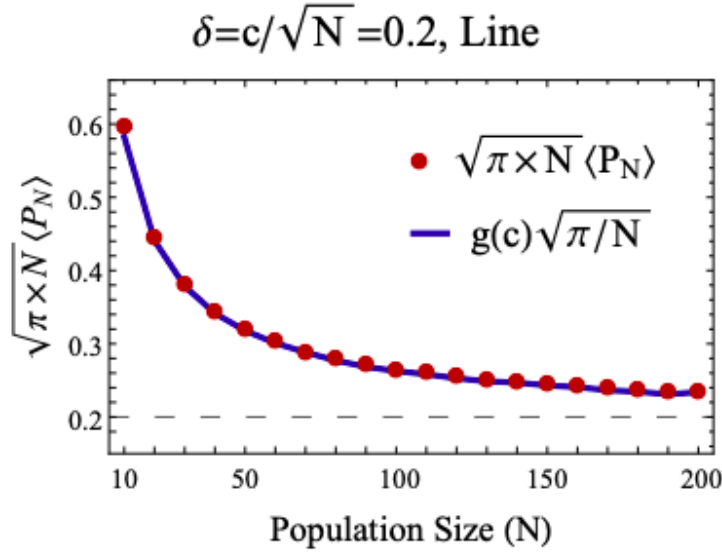Figure 9 below shows simulation of $\sqrt{\pi N}\mathbb{P}_N$ with $\delta = 0.2$ supporting this conjecture:

Figure 9: [3] Simulation of $\sqrt{\pi N}\mathbb{P}_N$ plotted with $(g(\delta\sqrt{N}))\sqrt{\pi/N}$ as functions of N, $\delta = 0.2$.

There are also many other questions that may be worth further discussions:

(a) As mentioned in section 3.1, extend the fixation probability asymptotics analysis and computation to a more general line model, where the initial mutant is not placed at an end point.

(b) Following the trail of though in section 3.1, extend the analysis to the model on

a loop with N sites. Figure 10 is the simulation showing that the asymptotics

of the loop model behaves a lot like the asymptotics of the line model that was
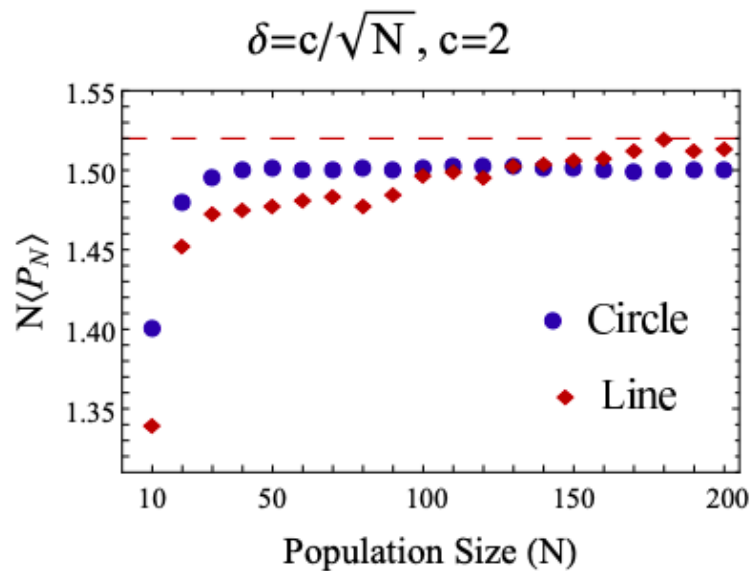
constructed in Section 1.2.

Figure 10: [3] Simulation of line model and loop model plotting $N\mathbb{P}_N$ as functions of
N, $\delta = 2\sqrt{N}$, averaging over 10^6 random simulations for each N.

(c) Extend the analysis further: work with more complicated graphs like binary

trees or general trees. For these cases, what makes the analysis difficult is

that after some time into the evolution process, the segments or sequences of

mutants could be no longer connected. For example, there can be two segments

of mutants in two different subtrees.

References

[1] Fisher, R. A. *On the Dominance Ratio.* Proceedings of the Royal Society of Edinburgh 42, 321-341 (1922).

[2] Farhang-Sardroodi, S., Darooneh, A. H., Nikbakht, M., Komarova, N. L. and Kohandel, M. *The effect of spatial randomness on the average fixation time of mutants.* PLoS Comput Biol 13(11):e1005864, (2017).

[3] Farhang-Sardroodi, S., Komarova, N. L., Michelen, M. and Pemantle, R. *Success probability for selectively neutal invading species in the line model.* arXiv, math.PR, 2005.13491. (2020).

[4] Haldane, J. B. S. *A mathematical theory of natural and artificial selection, Part V: Selection and mutation.* Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press 23 (1927).

[5] Kimura, M. *Process leading to quasi-fixation of genes in natural populations due to random fluctuations of selection intensities.* Genetics 39(3), 280 (1954).

[6] Kolmós, J., Major, P. and Tusnády, G. *An Approximation of Partial Sums of independent Rv'-s, and the sampe DF. I.* Probability Theory and related Fields, Vol. 32. (1975).

[7] Lawler, G. and Limic, V. *Random walk: a modern introduction.* Cambridge University Press, Cambridge, (2010).

[8] Mahdipour-Shirayeh, A., Darooneh, A.H., Long, A.D. Komarova, N. L. and Kohandel, M. *Genotype by random environmental interactions gives an advantage to non-favored minor alleles.* Sci Rep 7, 5193 (2017).

[9] Matsumoto, H. and Yor, M. *Exponential functionals of Brownian motion, I: Probability laws at fixed time.* Probability Surveys, 2:312-347 (2005).

[10] Wright, S. *Evolution in Mendelian populations.* Genetics 16(2),97-159 (1931).

[11] Yor, M. *On some exponential functionals of Brownian motion.* Adv. Appl. Prob., 24, 509?531.(1992).