Title
Probabilistic evolutionary models of cancer

Permalink
https://escholarship.org/uc/item/2f2567kr

Author
Kelly, Michael

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Probabilistic Evolutionary Models of Cancer

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Michael Kelly

Committee in charge:

  Jason Schweinsberg, Chair
  Vineet Bafna
  Lin Chao
  Bruce Driver
  Ruth Williams

2012

The dissertation of Michael Kelly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
                                          Chair

University of California, San Diego

2012

iii

DEDICATION

To my grandparents, mother, father, sister and brother.

EPIGRAPH

There are problems to whose solution I would attach an infinitely greater
importance than to those of mathematics, for example touching ethics, or our
relation to God, or concerning our destiny and our future; but their solution lies
wholly beyond us and completely outside the province of science.

—Karl Friedrich Gauss

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Jason Schweinsberg, for the hundreds of hours he has spent over the last few years teaching me, talking with me and guiding me. I would also like to thank Bruce Driver and Ruth Williams for what they have taught me, their guidance and their availability over the last five years.

For sparking my interest in math and encouraging me I would like to thank Leticia Barchini, Robert Myers, David Ullrich, David Wright and Roger Zierau from Oklahoma State University. As is common with most students, I was not sure what area of math I wanted to study when I entered graduate school. In my first two years at UCSD, Lei Ni taught challenging qualifying courses that piqued my interest in the analytic branches of mathematics.

There have been many peers that I have learned from over the years. I have had many conversations about mathematics and deeper topics as well as a lot of fun times with my roommates and office mates, Timothy Banham, Chris Deotte, Gordon Honerkamp-Smith, Jesus Oliver and Bo Yang. I would also like to thank my fellow probability students, Erik Carl, Tom Laetsch and David Lipshutz.

There have been many people that have influenced me inside and outside of academia and it is impossible to thank everyone who has in some way helped me to acquire my doctorate degree. As my family and friends know, it has been a bumpy ride. I would especially like to thank my family and the Church for their support and prayers throughout the years.

Chapter 2, in full, has been accepted for publication in the Annals of Applied Probability. The dissertation author was the sole author of this paper.

Chapter 3, in full, has been accepted for publication in the Journal of Applied Probability. The dissertation author was the sole author of this paper.

ABSTRACT OF THE DISSERTATION

Probabilistic Evolutionary Models of Cancer

by

Michael Kelly

Doctor of Philosophy in Mathematics

University of California, San Diego, 2012

Jason Schweinsberg, Chair

Cancer is currently viewed as an evolutionary process. In an organ there is a population of cells that give birth, die and mutate according to population dynamics that are determined by the types of cells under consideration. If certain cell mutations are acquired then the cells can become cancerous. In this manuscript we consider two evolutionary models that may each be viewed as a model of cancer.

The first model we consider is a Moran-type model. Individuals each have an integer valued fitness. Individuals with a higher fitness value are more likely to give birth and individuals with a lower fitness value are more likely to die. We fix the mutation rate and consider the limiting rate of adaptation as the population size tends to infinity. Similar models have been used to model cancers in liquids such as leukemia.

The second model we consider is a hierarchical model which differentiates between stem cells and progenitor cells. It has been proposed as a model of colorectal cancer. We find the limiting distribution for the time it takes for a cell to acquire two mutations as the population size tends to infinity and the mutation rates tend to 0. There are several different results depending how the mutation rates tend to 0 as a function of the population size. The two mutations represent the loss of two tumor suppressor genes. We also determine whether or not the mutations occur on stem cells or progenitor cells.

Chapter 1

Introduction

We discuss cancer as an evolutionary process. There is a population of cells that give birth, die and mutate according to population dynamics that depend on the types of cells under consideration. Cancer is only able to form after a certain set of cell mutations are acquired. For this reason, we may consider an evolutionary model of a population of asexually reproducing individuals that acquire mutations as a model of cancer.

There are many different types of cancer and many different conjectures for the number of mutations required for the formation of cancer. In 1951, Muller [28] first conjectured that cancer is the result of more than one cell mutation. In 1954 Armitage and Doll [1] proposed that cancer is the result of 6 or 7 cell mutations. In two papers in the 1970's, Knudson [21], [22] conjectured that retinoblastoma was the result of only two mutations. In [4], Calabrese et. al. estimate cancer to be the result of 4 to 9 mutations. As recently as 2008, Parsons [31] and The Cancer Genome Atlas Research Network [5] found an average of 47 mutations present in glioblastomas and Jones [19] found an average of 63 mutations present in pancreatic cancer. This gives reason to study models which view cancer as the result of many or few cell mutations. In the first model addressed in this manuscript we view cancer as a result of many mutations while in the second model we view cancer as the result of only two mutations.

The first model we study is a general evolutionary model. It is a Moran type model with mutations and selection. This model is similar to previously

studied models of cancer in liquids such as leukemia. We prove a theorem related to the mean rate of adaptation of the individuals. No assumptions are made on the number of mutations required to develop cancer. The purpose of this work is to determine how quickly a population acquires mutations.

The second model we study is a model of colorectal cancer. The model represents a colorectal crypt in which stem cells live near the base and progenitor cells migrate upwards until they reach the top and are removed. In particular, stem cells are differentiated from progenitor cells in this model. It is assumed that cancer is the result of two mutations. The two mutations represent the loss of two tumor suppressor genes. We determine the limiting distribution for the time it takes for cancer to form as well as whether or not the mutations leading to cancer occur on the stem cells.

Chapter 2

Upper Bound on the Rate of Adaptation in an Asexual Population

In a finite, asexually reproducing population with mutations it is well-known that competition among multiple individuals that get beneficial mutations can slow the rate of adaptation. This phenomenon is known as the Hill-Robertson effect, named for the authors of [17]. One may wish to consider the effect on the rate of adaptation of a population when there are many beneficial mutations present simultaneously. It is easily observed that when such a population is finite and all mutations are either neutral or deleterious the fitness of the population will decrease over time. This scenario is known as Muller's ratchet. The first rigorous results regarding Muller's ratchet were due to Haigh [15]. In an asexually reproducing population beneficial mutations are necessary to overcome Muller's ratchet. In this chapter we study a model that gives insight into both the Hill-Robertson effect and Muller's ratchet in large populations with fast mutation rates.

We study a Moran type model with mutations and selection that was first introduced by Yu, Etheridge and Cuthberson [36]. It may be considered as a general evolutionary model and is not necessarily a model of cancer. We define it as a continuous time stochastic process, $X = \{X_t : t \geq 0\}$, which has state space $\mathbb{Z}^N$. We let $X = (X^1, \ldots, X^N)$. In this model, N is the size of the population

and X_t^i represents the fitness of individual i at time t. The subscript t is usually suppressed. We fix constants $\mu > 0$, $0 < q \leq 1$ and $\gamma > 0$. The system has the following dynamics:

1. Mutation: Each individual acquires mutations at rate μ. When individual i gets a mutation it is beneficial with probability q and X^i increases by 1. With probability $1 - q$ the mutation is deleterious and X^i decreases by 1.

2. Selection: For each pair of individuals (i, j), at rate $\frac{\gamma}{N}(X^i - X^j)^+$ we set X^j equal to X^i.

3. Resampling: For each pair of individuals (i, j), at rate $1/N$ we set X^j equal to X^i.

Under the selection mechanism the event that X^j is set to equal X^i represents the more fit individual i giving birth and the less fit individual j dying. Likewise, the resampling event that causes X^j to equal X^i represents individual i giving birth and individual j dying.

We give an equivalent description of the model involving Poisson processes that may make the coupling arguments below more clear. The Poisson processes that determine the dynamics of X are as follows:

- There are N Poisson processes $\mathcal{P}^{i\uparrow}$, $1 \leq i \leq N$, on $[0, \infty)$ of rate $q\mu$. If $\mathcal{P}^{i\uparrow}$ gets a mark at t then the ith coordinate of X increases by 1 at time t.

- There are N Poisson processes $\mathcal{P}^{i\downarrow}$, $1 \leq i \leq N$, on $[0, \infty)$ of rate $(1-q)\mu$. If $\mathcal{P}^{i\downarrow}$ gets a mark at t then the ith coordinate of X decreases by 1 at time t.

- For each ordered pair of coordinates (i, j) with $i \neq j$ there is a Poisson process on $[0, \infty)$, $\mathcal{P}^{i,j}$, of rate $1/N$. If $\mathcal{P}^{i,j}$ gets a mark at t then the jth coordinate changes its value to agree with the ith coordinate at time t.

- For each ordered pair of coordinates (i, j) with $i \neq j$ there is a Poisson processes on $[0, \infty) \times [0, \infty)$, $\mathcal{P}^{i,j\uparrow}$, which has intensity $\frac{\gamma}{N}\lambda$ where λ is Lebesgue measure on $\mathbb{R}^2$. If $\mathcal{P}^{i,j\uparrow}$ gets a mark in $\{t\} \times [0, X_{t-}^i - X_{t-}^j]$ then the jth coordinate changes its value to agree with the ith coordinate at time t.

A population of homogeneously mixing cells may be represented by this model, as in Schweinsberg [33] and Durrett, Schmidt and Schweinsberg [11]. The main difference between this model and the ones found in [33] and [11] is that the rate at which cells acquire mutations does not tend to 0 in our model as it does in the others. Generally the rate of mutations is very small compared to the size of the population so it is reasonable to let the mutation rate tend to 0 in the limit. Because the mutation rate in our model is independent of the population size, we say that this population has fast mutations.

A heuristic argument in [36] shows that as N tends to infinity the mean rate of increase of the average fitness of the individuals in X is $O(\log N/(\log \log N)^2)$. Due to a typo on page 989 they state that the rate is $O(\log N/\log \log N)$. By equation (10) in [36],

$$K \log(\gamma K) = 2 \log N.$$

This implies that

$$K \approx \frac{2 \log N}{\log \log N}.$$

Plugging $2 \log N/\log \log N$ into each side of the consistency condition that they derive gives a rate of adaption of $O(\log N/(\log \log N)^2)$.

The heuristic argument is difficult to extend to a rigorous argument. Let

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X^i$$

be the continuous-time process which represents the average fitness of the individuals in the population. The rigorous results established in [36] are as follows:

- The centered process X^C, in which individual i has fitness $X^{C,i} = X^i - \overline{X}$, is ergodic and has a stationary distribution π.

- If

$$c_2 = \frac{1}{N} \sum_{i=1}^{N} (X^{C,i})^2$$

  is the variance of the centered process then

$$E^\pi[\overline{X}_t] = (\mu(2q - 1) + \gamma E^\pi[c_2])t$$

where E^π means that the initial configuration of X is chosen according to the stationary distribution π.

- For any $\delta > 0$ there exists N_0 large enough so that for all $N \geq N_0$ we have $E^\pi[\overline{X}_1] \geq \log^{1-\delta} N$.

It is difficult to say anything rigorous about $E^\pi[c_2]$ so other methods are needed to compute $E[\overline{X}_t]$. The third result of [36] shows that if there is a positive ratio of beneficial mutations then a large enough population will increase in fitness over time. A paper by Etheridge and Yu [12] provides further results pertaining to this model.

Our result is the following theorem.

Theorem 1. *Let $X_0^i = 0$ for $1 \leq i \leq N$. There exists a positive constant C which may depend on μ, q and γ such that for N large enough*

$$\frac{E[\overline{X}_t]}{t} \leq \frac{C \log N}{(\log \log N)^2}$$

for all $t \geq \log \log N$.

A difference between the result in [36] and Theorem 1 is that in [36] the initial state of the process is randomly chosen according to the stationary distribution π while we make the assumption that all of the individuals initially have fitness 0.

Other similar models can be found in the biological literature. In these models the density of the particles is assumed to act as a traveling wave in time. The bulk of the wave behaves approximately deterministically and the random noise comes from the most fit classes of individuals. One tries to determine how quickly the fittest classes advance and pull the wave forward. This traveling wave approach is used in [35] and [36] to approximate the rate of evolution as $O(\log N/(\log \log N)^2)$. For other work in this direction see Rouzine, Brunet and Wilke [32], Brunet, Rouzine and Wilke [2], Desai and Fisher [6] and Park, Simon and Krug [30]. Using non-rigorous arguments, these authors get estimates of $O(\log N)$, $O(\log N/\log \log N)$, and $O(\log N/(\log \log N)^2)$, where the differences

depend on the details of the models that they analyze. For more motivation and details concerning this model, please see the introduction in [36].

Motivated by applications to cancer development, Durrett and Mayberry have established rigorous results for a similar model in [9]. They consider two models in which all mutations are beneficial and the mutation rate tends to 0 as the population size tends to infinity. In one of their models the population size is fixed and in the other it is exponentially increasing. For the model with the fixed population size they show that the rate at which the average fitness is expected to increase is $O(\log N)$. By considering the expected number of individuals that have fitness k at time t, they establish rigorously that the density of the particles in their model will act as a traveling wave in time.

2.1   Proof of Theorem 1

We first establish some notation. Let $X_t^+ = \max\{X_t^i : 1 \le i \le N\}$ be the maximum fitness of any individual at time t and $X_t^- = \min\{X_t^i : 1 \le i \le N\}$ be the minimum fitness of any individual at time t. Define the width of the process to be $W_t = X_t^+ - X_t^-$ and define $D_t = X_t^+ - X_0^+$ be the distance the front of the process has traveled by time t. Theorem 1 states that all individuals initially have fitness 0. Therefore, a bound on $E[D_t]$ immediate yields a bound on $E[\overline{X}_t]$. The bounds we establish on D_t will depend on the width, W_t.

Let $w = w(N)$ be any positive, increasing function that satisfies

$$\lim_{N \to \infty} w(N) = \infty \text{ and } \lim_{N \to \infty} \frac{w(N)}{\log \log N} = 0.$$

Let $\mathcal{W} = \lfloor w \log N / \log \log N \rfloor$ and $\mathcal{T} = w^{-1/2} \log \log N$. Due to heuristic reasoning we conjecture that W_t is typically of size $O(\log N / \log \log N)$ so $\mathcal{W}$ is larger than the typical width of X. With probability tending to 1 selection should cause any width larger than $\mathcal{W}$ to shrink within $\mathcal{T}$ time units. Because the width is a

stochastic process we are motivated to make the following definitions.

$$t_1 = 0$$

$$s_n = \inf\{t \geq t_n : W_t \geq 2\mathcal{W}\} \text{ for } n \geq 1$$

$$t_n = \inf\{t \geq s_{n-1} : W_t < \mathcal{W}\} \text{ for } n \geq 2$$

$$Y_i = \sup_{s_i \leq t \leq t_{i+1}} D_t - D_{s_i} \text{ for } i \geq 1$$

$$N_t = \max\{i : s_i \leq t\} \text{ for } t \geq 0$$

Note that s_n and t_n exists for all $n \geq 1$ with probability 1.

Let $\mu > 0$ and $\gamma > 0$. We define branching processes $Z^{k,\uparrow}$ for $k \geq 0$ which have the following dynamics:

- Initially there are N particles of type k in $Z_0^{k,\uparrow}$.

- Each particle changes from type i to $i+1$ at rate μ.

- A particle of type i branches at rate $\gamma i + 1$ and upon branching the new particle is also type i.

Let $\overline{M}_t^{k,\uparrow}$ be the maximum type of any particle in $Z_t^{k,\uparrow}$ and let $M_t^{k,\uparrow} = \overline{M}_t^{k,\uparrow} - k$, so that $M_0^{k,\uparrow} = 0$.

We define a stochastic process X' that will be coupled with X as described in the proof of Proposition 2 for reasons that will become clear shortly. Let $\{\mathcal{Z}^n\}_{n=0}^{\infty}$ be an i.i.d. sequence of continuous-time stochastic processes which each have the same distribution as $Z^{\mathcal{W},\uparrow}$. Let $\overline{\mathcal{M}}_t^n$ be the maximum type of any particle in $\mathcal{Z}_t^n$ and let $\mathcal{M}_t^n = \overline{\mathcal{M}}_t^n - \mathcal{W}$ so that $\mathcal{M}_0^n = 0$ for all n. Define

$$X_t' = \begin{cases} X_0^+ + \mathcal{M}_t^0 & \text{if } t \in [0, \mathcal{T}] \\ X_{i\mathcal{T}}' + \mathcal{M}_{t-i\mathcal{T}}^i & \text{if } t \in (i\mathcal{T}, (i+1)\mathcal{T}] \text{ for any integer } i \geq 1 \end{cases}$$

and $D_t' = X_t' - X_0^+$. The idea is that D_t' is the maximum type of any particle in a branching process X' that has the same distribution as $Z^{\mathcal{W},\uparrow}$ except that at each time $i\mathcal{T}$ we restart the branching process so that there are once again N particles

of type $\mathcal{W}$. For each integer $i \geq 0$ at time $i\mathcal{T}$ the N particles initially have type D'_t which is the maximum type achieved by any particle in X'_t up to time t.

Now we are able to state the four propositions used to prove Theorem 1. Proposition 2 is a result of the coupling of X and X'.

Proposition 2. *Let $X^i_0 = 0$ for $1 \leq i \leq N$. Then*

$$D_t \leq D'_t + \sum_{i=1}^{N_t} Y_i$$

for all times $t \geq 0$.

Proposition 3. *Let $X^i_0 = 0$ for $1 \leq i \leq N$. For N large enough we have*

$$\sup_{t \in [\mathcal{T}, \infty)} \frac{E[D'_t]}{t} \leq \frac{6\mathcal{W}}{\mathcal{T}}.$$

With the initial condition $X^i_0 = 0$ for $1 \leq i \leq N$, we let $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ be the natural filtration associated with X.

Proposition 4. *Let $X^i_0 = 0$ for $1 \leq i \leq N$. For N large enough we have $E[Y_i | \mathcal{F}_{s_i}] \leq 5\mathcal{W}$ for all $i \geq 1$.*

Proposition 5. *Let $X^i_0 = 0$ for $1 \leq i \leq N$. For N large enough,*

$$\sup_{s \in [0, \infty)} \frac{1}{s} E[N_s] \leq \frac{1}{\mathcal{T}}$$

Proof of Theorem 1. Choose N large enough so that Propositions 3, 4 and 5 hold.

Fix $t \geq \log\log N$. It follows by definition of $\mathcal{T}$ that $t > \mathcal{T}$. Therefore,

$$
\begin{aligned}
E\left[\frac{D_t}{t}\right] &\leq E\left[\frac{D'_t + \sum_{i=1}^{N_t} Y_i}{t}\right] \text{ by Proposition 2} \\
&= E\left[\frac{D'_t}{t}\right] + E\left[\frac{\sum_{i=1}^{N_t} Y_i}{t}\right] \\
&\leq \frac{6\mathcal{W}}{\mathcal{T}} + \frac{1}{t}E\left[\sum_{i=1}^{N_t} Y_i\right] \text{ by Proposition 3} \\
&= \frac{6\mathcal{W}}{\mathcal{T}} + \frac{1}{t}\sum_{i=1}^{\infty} E[Y_i 1_{\{N_t \geq i\}}] \\
&= \frac{6\mathcal{W}}{\mathcal{T}} + \frac{1}{t}\sum_{i=1}^{\infty} E[E[Y_i 1_{\{N_t \geq i\}}|\mathcal{F}_{s_i}]] \\
&= \frac{6\mathcal{W}}{\mathcal{T}} + \frac{1}{t}\sum_{i=1}^{\infty} E[1_{\{N_t \geq i\}} E[Y_i|\mathcal{F}_{s_i}]] \\
&\leq \frac{6\mathcal{W}}{\mathcal{T}} + \frac{5\mathcal{W}}{t}\sum_{i=1}^{\infty} E[1_{\{N_t \geq i\}}] \text{ by Proposition 4} \\
&= \frac{6\mathcal{W}}{\mathcal{T}} + \frac{5\mathcal{W}}{t}E[N_t] \\
&\leq \frac{6\mathcal{W}}{\mathcal{T}} + \frac{5\mathcal{W}}{\mathcal{T}} \text{ by Proposition 5} \\
&= \frac{11w^{1/2}\log N}{(\log\log N)^2}.
\end{aligned}
$$

Since w may go to infinity arbitrarily slowly with N there must exist a constant C such that

$$
\frac{E[D_t]}{t} \leq \frac{C\log N}{(\log\log N)^2}
$$

for all $t \geq \log\log N$. This immediately gives a bound on $E[\overline{X}_t]/t$. $\qquad\square$

2.2 Bounding the rate when the width is small

Through the use of branching processes we establish a bound on D_t that depends on the width. We will make use of the strong Markov property of X at the times s_n and t_n for $n \geq 1$. For this reason, many of the statements we prove below will include conditions for which $W_0 > 0$ even though according to the conditions

of Theorem 1 we have $W_0 = 0$. In this section we establish a small upper bound for D_t on the time intervals $[t_n, s_n)$.

The following proofs will involve coupling X with various branching processes. For clarity we refer to individuals in branching processes as particles to distinguish them from the individuals in X. Also, while the individuals in X each have an integer valued fitness, the particles in a branching process will each be given an integer value that we refer to as the type of the particle.

Let $Z^C = \{Z_t^C\}_{t \geq 0}$ be a multi-type Yule process in which there are initially N particles of type 0. Particles increase from type i to type $i+1$ at rate $\mu > 0$ and branch at rate $C > 0$. When a particle of type i branches the new particle is also type i. Let M_t^C be the maximum type of any particle at time t.

The following lemma is a basic result about Yule processes.

Lemma 6. *For any $N \geq 0$, time $t \geq 0$, and natural number l,*

$$P(M_t^C \geq l) \leq \frac{N(t\mu)^l e^{Ct}}{l!}.$$

Proof. Consider a Yule process Z which is the same as Z^C except there is only one particle at time 0. It is well known that the number of particles in Z_t has mean e^{Ct}. Let M_t' be the maximum type of any particle at time t. When there are k particles in the population, we let $B_1, \ldots, B_k$ denote the types of the particles, where the numbering is independent of the mutations. For any $l \geq 0$,

$$
\begin{aligned}
P(M_t' \geq l) &= \sum_{k=1}^{\infty} P(M_t' \geq l | Z_t = k) P(Z_t = k) \\
&= \sum_{k=1}^{\infty} P(\{B_1 \geq l\} \cup \cdots \cup \{B_k \geq l\} | Z_t = k) P(Z_t = k) \\
&\leq \sum_{k=1}^{\infty} k P(B_1 \geq l) P(Z_t = k) \\
&= E[Z_t] P(B_1 \geq l) \\
&= e^{Ct} \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!} e^{-\mu t}.
\end{aligned}
$$

By Lemma 12 it follows that

$$P(M_t' \geq l) \leq \frac{(t\mu)^l e^{Ct}}{l!}.$$

Now consider Z^C. At time 0 label the particles $1, 2, \ldots, N$ and let $M'_{i,t}$ be the maximum type of any particle among the progeny of particle i at time t. Then

$$
\begin{aligned}
P(M_t^C \geq l) &= P(\{M'_{1,t} \geq l\} \cup \cdots \cup \{M'_{N,t} \geq l\}) \\
&\leq N P(M'_{1,t} \geq l) \\
&\leq \frac{N(t\mu)^l e^{Ct}}{l!}.
\end{aligned}
$$

$\square$

The next proposition will give a lower bound on the fitness of any individual up to time t given that we know the least fitness at time 0 is X_0^-. We do this by establishing an upper bound on the amount that any individual will decrease in fitness. Let

$$
S_t = \sup_{0 \leq s \leq t} (X_0^- - X_s^-).
$$

Proposition 7. *For any population size N, initial configuration X_0, time $t \geq 0$, and natural number l,*

$$
P(S_t \geq l) \leq \frac{N(t\mu)^l e^t}{l!}.
$$

Proof. By Lemma 6 we have

$$
P(M_t^1 \geq l) \leq \frac{N(t\mu)^l e^t}{l!}
$$

for any population size N, time $t \geq 0$ and natural number l. Note that from our notation above Z^1 is a Yule process with branching rate 1. To complete the proof we establish a coupling between X and Z^1 such that for any population size N and time $t \geq 0$ we have $M_t^1 \geq S_t$.

At all times every individual in X will be paired with one particle in Z^1. The coupling is as follows:

- We initially have a one-to-one pairing of each individual i in X_0 with each particle i in Z_0^1.

- The particle in Z^1 that is paired with individual i will increase in type by 1 only when individual i gets a mutation.

- For each individual i in X, at rate $(N-1)/N$ individuals $j \neq i$ are replaced by individual i due to resampling events. If individual i replaces individual j due to resampling, then the particle labeled i in Z^1 branches. If particle i has a higher type than particle j then the new particle is paired with individual j. The particle that was paired with individual j before the branching event is no longer paired with any individual in X. If particle i has a lower type than particle j then the particle that was paired with individual j remains paired with individual j and the new particle is not paired with any individual in X.

- The particle paired with individual i in Z^1 branches at rate $1/N$ and these branching events are independent of any of the events in X. When the particle paired with individual i branches due to these events the new particle is not paired with any individual in X.

- Any particles in Z^1 that are not paired with a individual in X branch and acquire mutations independently of X. The selection events in X are independent of any events in Z^1.

See Figure 2.1 for an illustration of the coupling.

Let R^i be the type of the particle in Z^1 that is paired with individual i and let

$$S_s^i = \sup_{0 \leq r \leq s} (X_0^- - X_r^i).$$

To show $M_t^1 \geq S_t$ it is enough to show $R_t^i \geq S_t^i$ for all i. Initially $S_0^i \leq R_0^i = 0$ for all i. Note that both $s \mapsto S_s^i$ and $s \mapsto R_s^i$ are increasing functions and that increases in these functions correspond to decreases in X^i.
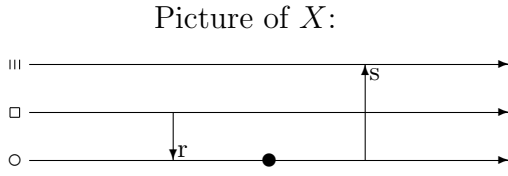
Picture of X:

Time goes from left to right.
● denotes mutations in each model.
⦀ □ ○ are used to indicate which individual in X is coupled with which particle in the branching processes.
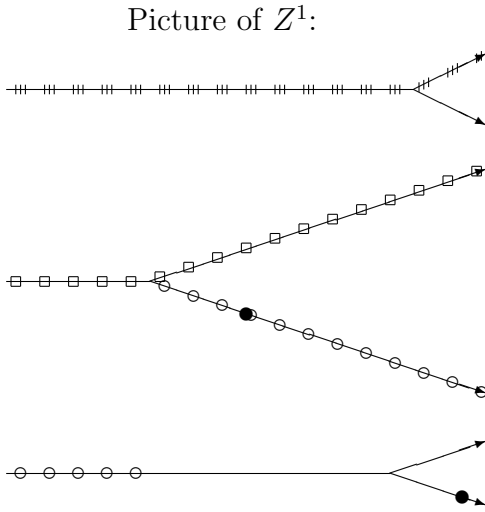
Picture of Z^1:

In the picture of X an arrow with an 'r' denotes a resampling event and an arrow with an 's' denotes a selection event.

A selection event in X does not correspond to a branching event in Z^1.

The times at which the particles are not marked indicate that the particles are not coupled with any individual in X and therefore the branching and mutation events on the unmarked particles are independent of any of the events in X.

Figure 2.1: Picture of the coupling of X with Z^1 when $N = 3$

When individual i gets a mutation, R^i increases by 1. However, if individual i gets a mutation at time s then S^i will only increase by 1 if $S^i_{s-} = X^-_0 - X^i_{s-}$ and the mutation is deleterious. Therefore, if individual i gets a mutation at time s and $S^i_{s-} \leq R^i_{s-}$ then

$$S^i_s \leq S^i_{s-} + 1 \leq R^i_{s-} + 1 = R^i_s.$$

Suppose individual j is replaced by individual i due to a resampling event at time s and that both $S^j_{s-} \leq R^j_{s-}$ and $S^i_{s-} \leq R^i_{s-}$ hold. With probability 1 we have $S^i_s = S^i_{s-}$ and $R^i_s = R^i_{s-}$. If $X^-_0 - X^i_s \leq S^j_{s-}$ then $S^j_{s-} = S^j_s$. From this it follows that $S^j_s \leq R^j_s$. If $X^-_0 - X^i_s > S^j_{s-}$ then $S^j_s = X^-_0 - X^i_s \leq S^i_s \leq R^i_s$. If $R^i_s \geq R^j_{s-}$, then by the definition of the coupling, $R^j_s = R^i_s$. If $R^i_s < R^j_{s-}$, then by definition of the coupling, $R^j_s = R^j_{s-}$. Therefore, $R^j_s \geq R^i_s$ which gives us $S^j_s \leq R^j_s$.

Selection events will never increase S^i and since S^i and R^i are increasing in time a selection event at time s will preserve the inequality $S^i_s \leq R^i_s$. This shows that any event that occurs at time s which may change the fitness of a individual i in X will preserve the inequality $S^i_s \leq R^i_s$. Since the result holds for each individual i we have $S_t \leq M^1_t$. $\qquad\square$

The next lemma is a basic result about the maximum type of any particle in $Z^{k,\uparrow}_t$.

Lemma 8. *For any time $t \geq 0$ and any integers $k \geq 0$ and $l \geq 0$ we have*

$$P(M^{k,\uparrow}_t > l) \leq \frac{N(t\mu)^l e^{(\gamma(k+l)+1)t}}{l!}.$$

Proof. While all of the particles in $Z^{k,\uparrow}_t$ have type less than $k + l$ they branch at a rate which is less than or equal to $\gamma(k + l) + 1$. Using the notation related to Z^C_t above, it follows that $P(M^{k,\uparrow}_t > l) \leq P(M^{\gamma(k+l)+1}_t > l)$. By Lemma 6 we have

$$P(M^{\gamma(k+l)+1}_t > l) \leq \frac{N(t\mu)^l e^{(\gamma(k+l)+1)t}}{l!}.$$

$\qquad\square$

We now wish to bound the distance the front of the wave moves as a function of the initial width. Recall that W_0 is the width of X at time 0.

Proposition 9. *For any initial configuration X_0, fixed time $t \geq 0$ and any integer $l \geq 0$ we have*

$$P(\sup_{0 \leq s \leq t} D_s > l) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0 + 2l) + \mu + 1)t}}{(l-1)!}.$$

Proof. We first establish a coupling between X and $Z^{W_0 + k, \uparrow}$ for each integer $k \geq 0$.

Let $T^k = \inf\{t : S_t > k\}$ for $k \geq 1$. Every individual in X will be paired with one particle in $Z^{W_0 + k, \uparrow}$ until time T^k. We couple $Z^{W_0 + k, \uparrow}$ with X for all times $t \in [0, T^k)$ as follows:

- We initially have a one-to-one pairing of each individual i in X_0 with each particle i in $Z_0^{W_0 + k, \uparrow}$. When a particle in $Z_t^{W_0 + k, \uparrow}$ is coupled with individual i, we refer to the particle as particle i.

- Particle i increase in type by 1 only when individual i gets a mutation.

- For each individual i in X, at rate $(N-1)/N$ individuals $j \neq i$ are replaced by individual i due to resampling events. If individual i replaces individual j due to resampling, then particle i branches. If particle i has a higher type than particle j then the new particle is paired with individual j. The particle that was paired with individual j before the branching event is no longer paired with any individual in X. If particle i has a lower type than particle j then the particle that was paired with individual j remains paired with individual j and the new particle is not paired with any individual in X.

- Additionally, particle i branches at rate $1/N$ and these branching events are independent of any of the events in X. When particle i branches due to these events the new particle is not paired with any individual in X.

- In X there is a time dependent rate γU_s^i at which individuals $j \neq i$ are replaced by individual i due to selection events. Namely,

$$U_s^i = \frac{1}{N} \sum_{j=1}^{N} (X_s^i - X_s^j)^+.$$

If individual j is replaced by individual i in X due to a selection event then particle i branches. If particle i has a higher type than particle j then the new particle is paired with individual j. The particle that was paired with individual j before the branching event is no longer paired with any individual in X. If particle i has a lower type than particle j then the particle that was paired with individual j remains paired with individual j. The new particle is not paired with any individual in X.

- Additionally, particle i branches at a time dependent rate $\gamma(R_t^{i,k} - U_t^i)$ where $R_t^{i,k}$ is the type of particle i. These branching events are independent of any of the events in X. When such a branching event occurs, the new particle is not paired with any individual in X.

- Any particles in $Z^{W_0+k,\uparrow}$ that are not paired with a individual in X branch and change type independently of X.

See Figure 2.2 for an illustration of the coupling.

Fix $k \geq 1$. For the above coupling between X and $Z^{W_0+k,\uparrow}$ to be well defined until time T^k, we need $R_t^{i,k} - U_t^i \geq 0$ for all $i \in \{1, \ldots, N\}$ and for all times $t \in [0, T^k)$. Let $\overline{T}^{k,i} = \inf\{t : R_t^{i,k} - U_t^i < 0\}$. The coupling between X and $Z^{W_0+k,\uparrow}$ is well-defined until time $\overline{T}^k = \min\{\overline{T}^{k,i} : 1 \leq i \leq N\}$. We will show that $T^k \leq \overline{T}^k$.

Let

$$\overline{S}_t^i = \sup_{0 \leq s \leq t} (X_s^i - X_0^+) \quad \text{and} \quad \overline{R}_t^{i,k} = R_0^{i,k} - W_0 - k.$$

Initially $\overline{S}_0^i \leq \overline{R}_0^{i,k} = 0$ for all i. Note that both $t \mapsto \overline{S}_t^i$ and $t \mapsto R_t^{i,k}$ are increasing functions, from which it follows that $t \mapsto \overline{R}_t^{i,k}$ is also an increasing function.

When individual i gets a mutation, $\overline{R}^{i,k}$ increases by 1. However, if individual i gets a mutation at time s then $\overline{S}^i$ will only increase by 1 if $\overline{S}_{s-}^i = X_{s-}^i - X_0^+$ and the mutation is beneficial. Therefore, if individual i gets a mutation at time s and $\overline{S}_{s-}^i \leq \overline{R}_{s-}^{i,k}$ then

$$\overline{S}_s^i \leq \overline{S}_{s-}^i + 1 \leq \overline{R}_{s-}^{i,k} + 1 = \overline{R}_s^{i,k}.$$
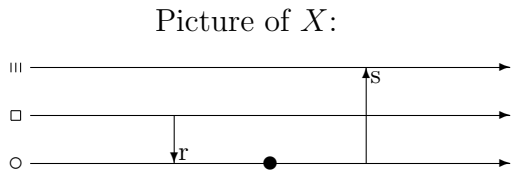
Picture of X:

Time goes from left to right.
● denotes mutations in each model.
⫲ □ ○ are used to indicate which
individual in X is coupled with which
particle in the branching processes.

Picture of $Z^{k,\uparrow}$:
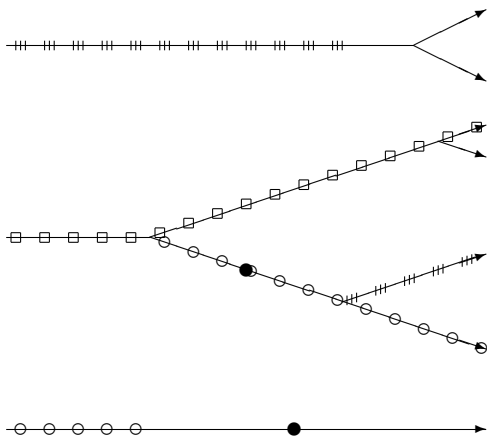
In the picture of X an arrow with an
'r' denotes a resampling event and an
arrow with an 's' denotes a selection
event.

A selection event in X corresponds
to a branching event in $Z^{k,\uparrow}$.

The times at which the particles
are not marked indicate that the
particles are not coupled with any
individual in X and therefore the
branching and mutation events on the
unmarked particles are independent of
any of the events in X.

Figure 2.2: Picture of the coupling of X with $Z^{k,\uparrow}$ when $N = 3$

Suppose individual j is replaced by individual i due to a resampling or selection event at time s and that both $\overline{S}^j_{s-} \leq \overline{R}^{j,k}_{s-}$ and $\overline{S}^i_s = \overline{S}^i_{s-} \leq \overline{R}^{i,k}_{s-} = \overline{R}^{i,k}_s$ hold. If $X^i_s - X^+_0 \leq \overline{S}^j_{s-}$ then $\overline{S}^j_{s-} = \overline{S}^j_s$. It follows that $\overline{S}^j_s \leq \overline{R}^{j,k}_s$. If $X^i_s - X^+_0 > \overline{S}^j_{s-}$ then $\overline{S}^j_s = X^-_0 - X^i_s \leq \overline{S}^i_s \leq \overline{R}^{i,k}_s$. If $\overline{R}^{i,k}_s \geq \overline{R}^{j,k}_{s-}$, then by the definition of the coupling, $\overline{R}^{j,k}_s = \overline{R}^{i,k}_s$. If $\overline{R}^{i,k}_s < \overline{R}^{j,k}_{s-}$, then by definition of the coupling, $\overline{R}^{j,k}_s = \overline{R}^{j,k}_{s-}$. Therefore, $\overline{R}^{j,k}_s \geq \overline{R}^{i,k}_s$ which gives us $\overline{S}^j_s \leq \overline{R}^{j,k}_s$.

For any time $s < T^k$ we have

$$R^{i,k}_s \geq \overline{S}^i_s + W_0 + k \geq X^i_s - X^+_0 + W_0 + k = X^i_s - X^-_0 + k.$$

If there were N individuals with fitness $X^-_0 - k$ at time $s \in [0, \overline{T}^{k,i})$ then the rate at which individual i replaces these N individuals due to selection is $\gamma(X^i_s - X^-_0 + k)$. However, for any time $s < T^k$ there are fewer than N individuals being replaced by individual i due to selection and they will all have fitnesses at least as large as $X^-_0 - k$. This gives us a bound on the rate at which resampling events occur on individual i before time T^k, namely $U^i_s \leq X^i_s - X^-_0 + k \leq R^{i,k}_s$ for all $s \in [0, T^k)$. This shows that $T^k \leq \overline{T}^{k,i}$ for all i. Hence, $T^k \leq \overline{T}^k$ and the coupling is well-defined until time T^k.

We have shown that any event that occurs at time $s \in [0, T^k)$ which may change the fitness of a individual i in X will preserve the inequality $\overline{S}^i_s \leq \overline{R}^{i,k}_s$. Since the result holds for each individual i, for any $s \in [0, T^k)$ we have

$$\sup_{0 \leq r \leq s} D_r = \sup_{1 \leq i \leq N} \overline{S}^i_s \leq \sup_{1 \leq i \leq N} \overline{R}^{i,k}_s \leq M^{W_0+k,\uparrow}_s.$$

Note that if $\sup_{0 \leq s \leq t}(X^-_0 - X^-_s) \leq k$ then $t < T^k$. If $\sup_{0 \leq s \leq t}(X^-_0 - X^-_s) \leq k$ we have $M^{W_0+k,\uparrow}_t \geq \sup_{0 \leq s \leq t} D_s$. This allows us to do the following computation:

$$P(\sup_{0 \le s \le t} D_s > l) = \sum_{i=0}^{\infty} P(\{\sup_{0 \le s \le t} D_s > l\} \cap \{\sup_{0 \le s \le t} (X_0^- - X_s^-) = i\})$$

$$\le \sum_{i=0}^{\infty} P(\{M_t^{W_0+i,\uparrow} > l\} \cap \{\sup_{0 \le s \le t} (X_0^- - X_s^-) = i\})$$

$$\le \sum_{i=0}^{\infty} P(\{M_t^{W_0+i,\uparrow} > l\} \cap \{\sup_{0 \le s \le t} (X_0^- - X_s^-) \ge i\})$$

$$\le \sum_{i=0}^{\infty} P(M_t^{W_0+i,\uparrow} > l) \wedge P(\sup_{0 \le s \le t} (X_0^- - X_s^-) \ge i)$$

$$\le \sum_{i=0}^{\infty} P(M_t^{W_0+i,\uparrow} > l) \wedge \left(\frac{N(t\mu)^i e^t}{i!}\right) \text{ by Proposition 7}$$

$$\le \sum_{i=0}^{\infty} \left(\frac{N(t\mu)^l e^{(\gamma(W_0+i+l)+1)t}}{l!}\right) \wedge \left(\frac{N(t\mu)^i e^t}{i!}\right) \text{ by Lemma 8}$$

$$\le \frac{N(t\mu)^l e^{(\gamma(W_0+l)+1)t}}{l!} \sum_{i=0}^{l-1} e^{i\gamma t} + Ne^t \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!}$$

$$\le \frac{N(t\mu)^l e^{(\gamma(W_0+l)+1)t}}{l!} \cdot l e^{l\gamma t} + Ne^t \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!}$$

$$\le \frac{N(t\mu)^l e^{(\gamma(W_0+2l)+1)t}}{(l-1)!} + \frac{N(t\mu)^l e^{(\mu+1)t}}{l!} \text{ by Lemma 12}$$

$$\le \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}. \tag{2.1}$$

$$\square$$

We now extend the bound we got on the least fit individuals in Proposition 7 to a slightly stronger result.

Definition 10. *Let $x \in \mathbb{Z}$ and let $\mathcal{S}_t^x \subset \{1, 2, \ldots, N\}$ correspond to a collection of individuals at time t which is determined by the following dynamics:*

- *Initially $\mathcal{S}_0^x$ consists of all individuals whose fitness lies in the interval (x, ∞).*

- *If a resampling or selection event occurs at time t and a individual not in $\mathcal{S}_{t-}^x$ is replaced by a individual in $\mathcal{S}_{t-}^x$ then it is added to $\mathcal{S}_t^x$.*

- *If a beneficial mutation occurs at time t on a individual not in $\mathcal{S}_{t-}^x$ that causes its fitness to increase from x to $x+1$ it is added to $\mathcal{S}_t^x$.*

- *If a resampling event occurs at time t to a individual in $\mathcal{S}^x_{t-}$ and it is replaced by a individual not in $\mathcal{S}^x_{t-}$ then it is removed from $\mathcal{S}^x_t$.*

Mutation and selection events do not cause individuals to be lost from $\mathcal{S}^x$. We now prove the following corollary to Proposition 9.

Corollary 11. *Let $A^{x,l}_t$ be the event that a individual in $\mathcal{S}^x_s$ has fitness in $(-\infty, x-l]$ for some time $s \in [0,t]$. For any initial configuration X_0, time $t \geq 0$ and any integer l,*

$$P(A^{x,l}_t) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}.$$

Note that we cannot use the bound found in Proposition 7 because individuals not in $\mathcal{S}^x_t$ may move to $\mathcal{S}^x_t$ due to selection events. In the proof of Proposition 7 the number of individuals with the least fitness cannot increase due to selection events. However, the number of individuals with the least fitness in $\mathcal{S}^x_t$ may increase due to selection events involving individuals not in $\mathcal{S}^x_t$.

Proof of Corollary 11. For $k \geq 1$ let X be coupled with $Z^{W_0+k,\uparrow}$ as in the proof of Proposition 9. Let T^k, $R^{i,k}_t$ and $\overline{R}^{i,k}_t$ be defined as they were in the proof of Proposition 9. Define $\overline{T}^i_s = \{r \in [0,s] : i \in \mathcal{S}^x_r\}$ and let

$$S^i_s = \begin{cases} \sup_{r \in \overline{T}^i_s} (x - X^i_r) & \text{if } \overline{T}^i_s \neq \emptyset \\ -\infty & \text{if } \overline{T}^i_s = \emptyset \end{cases}.$$

The goal is to show that for all $s \in [0, T^k)$ we have

$$\sup_{1 \leq i \leq N} S^i_s \leq \sup_{1 \leq i \leq N} \overline{R}^{i,k}_s \leq M^{W_0+k,\uparrow}_s.$$

Note that we can only consider the coupling of X with $Z^{W_0+k,\uparrow}$ until time T^k because after this time the coupling is not well-defined.

Initially all of the individuals in $\mathcal{S}^x_0$ have fitness in (x, ∞). Therefore, if $i \in \mathcal{S}^x_0$ then $S^i_0 \leq 0 = \overline{R}^{i,k}_0$. If $i \notin \mathcal{S}^x_0$ then $S^i_0 = -\infty < \overline{R}^{i,k}_0$.

Suppose individual i gets a mutation at time s and for any time $s' \in [0, s-)$ we have $S^i_{s'} \leq \overline{R}^{i,k}_{s'}$. Then $\overline{R}^{i,k}$ increases by 1. If $i \in \overline{\mathcal{S}}^x_{s-}$ then S^i_s will only increase by 1 if $S^i_{s-} = x - X^i_s$ and the mutation is deleterious. If $i \notin \mathcal{S}^x_{s-}$ and the mutation

does not cause the fitness of individual i to change from x to $x+1$ then $S_s^i = S_{s-}^i$. If $i \notin \mathcal{S}_{s-}^x$ and the mutation does cause the fitness of individual i to change from x to $x+1$ then $S_s^i = S_{s-}^i \vee 0$. In any of these three cases, $S_s^i \le \overline{R}_s^{i,k}$.

Suppose individual j is replaced by individual i due to a resampling or selection event at time s and that $S_{s-}^j \le \overline{R}_{s-}^{j,k}$ and $S_{s-}^i \le \overline{R}_{s-}^{i,k}$. If $i \notin \overline{S}_{s-}^x$ then $S_{s-}^j = S_s^j \le \overline{R}_{s-}^{j,k}$. Suppose $i \in \mathcal{S}_{s-}^x$. If $x - X_s^i \le S_{s-}^j$ then $S_{s-}^j = S_s^j$. From this it follows that $S_s^j \le \overline{R}_s^j$. If $x - X_s^i > S_{s-}^j$ then $S_s^j = x - X_s^i \le S_s^i \le \overline{R}_s^i$. If $\overline{R}_s^i \ge \overline{R}_{s-}^j$, then by the definition of the coupling, $\overline{R}_s^j = \overline{R}_s^i$. If $\overline{R}_s^i < \overline{R}_{s-}^j$, then by definition of the coupling, $\overline{R}_s^j = \overline{R}_{s-}^j$. Therefore, $\overline{R}_s^j \ge \overline{R}_s^i$ which gives us $S_s^j \le \overline{R}_s^j$.

Note that if $\sup_{0 \le s \le t}(X_0^- - X_s^-) \le k$ then $t < T^k$. Therefore, on the event $\{\sup_{0 \le s \le t}(X_0^- - X_s^-) \le k\}$ we have $M_t^{W_0+k,\uparrow} \ge \sup_{1 \le i \le N} S_s^i$. This allows us to do the following computation:

$$
\begin{aligned}
P(\sup_{0 \le s \le t} \sup_{1 \le i \le N} S_s^i > l) &= \sum_{i=0}^{\infty} P(\{\sup_{0 \le s \le t} \sup_{1 \le i \le N} S_s^i > l\} \cap \{\sup_{0 \le s \le t}(X_0^- - X_s^-) = i\}) \\
&\le \sum_{i=0}^{\infty} P(\{M_t^{W_0+i,\uparrow} > l\} \cap \{\sup_{0 \le s \le t}(X_0^- - X_s^-) = i\}).
\end{aligned}
$$

This is the same bound as equation (2.1) in the proof of Proposition 9. Therefore, we have established the same bound. $\qquad\square$

Lemma 12. *Let $x \ge 0$. The tail of the exponential series satisfies*

$$
\sum_{i=k}^{\infty} \frac{x^i}{i!} \le \frac{x^k e^x}{k!}.
$$

Proof. By Taylor's Remainder Theorem we know that there exists a $\xi \in [0, x]$ such that

$$
e^x = \sum_{i=1}^{k-1} \frac{x^i}{i!} + \frac{x^k e^{\xi}}{k!}.
$$

Using the series expansion of e^x we have

$$
\sum_{i=k}^{\infty} \frac{x^i}{i!} = \frac{x^k e^{\xi}}{k!} \le \frac{x^k e^x}{k!}.
$$

$\qquad\square$

Proof of Proposition 3. By definition $D'_{\mathcal{T}}$ has the same distribution as $M_{\mathcal{T}}^{\mathcal{W},\uparrow}$ so by Lemma 8 we have

$$P(D'_{\mathcal{T}} > l) \leq \frac{N(\mathcal{T}\mu)^l e^{(\gamma(\mathcal{W}+l)+1)\mathcal{T}}}{l!}.$$

Then

$$\frac{E[D'_{\mathcal{T}}]}{2\mathcal{W}} = \frac{1}{2\mathcal{W}} \sum_{l=0}^{\infty} P(D'_{\mathcal{T}} > l)$$

$$\leq \frac{1}{2\mathcal{W}} \left[2\mathcal{W} + \sum_{l=2\mathcal{W}}^{\infty} \frac{N(\mathcal{T}\mu)^l e^{(\gamma(\mathcal{W}+l)+1)\mathcal{T}}}{l!} \right]. \tag{2.2}$$

By Lemma 12 we have

$$\sum_{l=2\mathcal{W}}^{\infty} \frac{N(\mathcal{T}\mu)^l e^{(\gamma(\mathcal{W}+l)+1)\mathcal{T}}}{l!} \leq \frac{N e^{(\gamma\mathcal{W}+1)\mathcal{T}} (\mathcal{T}\mu e^{\gamma\mathcal{T}})^{2\mathcal{W}} e^{\mathcal{T}\mu e^{\gamma\mathcal{T}}}}{(2\mathcal{W})!}. \tag{2.3}$$

Note that for any $k \geq 2$ both $D'_{k\mathcal{T}} - D'_{(k-1)\mathcal{T}}$ and $D'_{\mathcal{T}}$ have the same distribution, namely that of $M_{\mathcal{T}}^{\mathcal{W}}$. Choose $t \in [k\mathcal{T}, (k+1)\mathcal{T})$ for some $k \geq 1$. Because D'_t is increasing in t we have

$$\frac{D'_t}{t} \leq \frac{1}{k\mathcal{T}} \left(D'_{(k+1)\mathcal{T}} - D'_{k\mathcal{T}} + D'_{k\mathcal{T}} - \cdots + D'_{2\mathcal{T}} - D'_{\mathcal{T}} + D'_{\mathcal{T}} \right).$$

Therefore,

$$\frac{E[D'_t]}{t} \leq \frac{(k+1)E[D'_{\mathcal{T}}]}{k\mathcal{T}} \leq \frac{2E[D'_{\mathcal{T}}]}{\mathcal{T}}.$$

Let $t > \mathcal{T}$. Dividing both sides by $2\mathcal{W}/\mathcal{T}$ and using the bounds found in equations (2.2) and (2.3) gives us

$$\frac{\mathcal{T}E[D'_t]}{2t\mathcal{W}} \leq \frac{2E[D'_{\mathcal{T}}]}{2\mathcal{W}} \leq 2 + \frac{N e^{(\gamma\mathcal{W}+1)\mathcal{T}} (\mathcal{T}\mu e^{\gamma\mathcal{T}})^{2\mathcal{W}} e^{\mathcal{T}\mu e^{\gamma\mathcal{T}}}}{2\mathcal{W}(2\mathcal{W})!}.$$

By Stirling's formula we have

$$\frac{N e^{(\gamma\mathcal{W}+1)\mathcal{T}} (\mathcal{T}\mu e^{\gamma\mathcal{T}})^{2\mathcal{W}} e^{\mathcal{T}\mu e^{\gamma\mathcal{T}}}}{2\mathcal{W}(2\mathcal{W})!} \sim \frac{N e^{(\gamma\mathcal{W}+1)\mathcal{T}} (\mathcal{T}\mu e^{\gamma\mathcal{T}})^{2\mathcal{W}} e^{\mathcal{T}\mu e^{\gamma\mathcal{T}}+2\mathcal{W}}}{(2\mathcal{W})^{2\mathcal{W}+1}\sqrt{4\pi\mathcal{W}}} = e^x$$

where x is equal to

$$\log N + \mathcal{T}(\gamma\mathcal{W}+1+\mu e^{\gamma\mathcal{T}}) + 2\mathcal{W}(\log(\mathcal{T}\mu e^{\gamma\mathcal{T}})+1) - (2\mathcal{W}+1)\log(2\mathcal{W}) - \log(4\pi\mathcal{W})/2.$$

As $N \to \infty$ we have $x \sim -(2\mathcal{W}+1)\log(2\mathcal{W}) \sim -2w\log N$. Therefore,

$$\frac{\mathcal{T}E[D'_t]}{2t\mathcal{W}} \leq 3$$

for N large enough. $\qquad\square$

Proof of Proposition 2. We now couple X with X' by coupling X with the sequence of processes $\{\mathcal{Z}^m\}_{m=0}^{\infty}$. Let

$$I_m = (m\mathcal{T}, (m+1)\mathcal{T}] \cap \bigcup_{n=1}^{\infty} [t_n, s_n) \text{ and } J_m = (0, \mathcal{T}] \cap \bigcup_{n=1}^{\infty} [t_n - m\mathcal{T}, s_n - m\mathcal{T}).$$

For any $m \geq 0$ we couple X and $\mathcal{Z}^m$ as follows:

- The particles in $\mathcal{Z}_0^m$ are labeled $1, 2, \ldots, N$.

- For any time in I_m^C the process X behaves independently of $\mathcal{Z}^m$. For any time in J_m^C the process $\mathcal{Z}^m$ behaves independently of the process X. During the time J_m^C, if a particle labeled i in $\mathcal{Z}^m$ branches, the particle remains labeled i and the new particle is unlabeled.

- The particle in $\mathcal{Z}^m$ that is paired with individual i will increase in type by 1 at time $t \in J_m$ only when individual i gets a mutation at time $t + m\mathcal{T} \in I_m$.

- For each individual i in X, at rate $(N-1)/N$ individuals $j \neq i$ are replaced by individual i due to resampling events. If individual i replaces individual j due to resampling at time $t \in I_m$ then the particle labeled i in $\mathcal{Z}^m$ branches at time $t - m\mathcal{T} \in J_m$. If particle i has a higher type than particle j then the new particle is paired with individual j. The particle that was paired with individual j before the branching event is no longer paired with any individual in X. If particle i has a lower type than particle j then the particle that was paired with individual j remains paired with individual j and the new particle is not paired with any individual in X.

- The particle paired with individual i in $\mathcal{Z}^m$ branches at rate $1/N$ for all times $t \in J_m$ and these branching events are independent of any of the events in X. When the particle paired with individual i branches due to these events the new particle is not paired with any individual in X.

- In X there is a time dependent rate γU_s^i at which individuals $j \neq i$ are replaced by individual i due to selection events. If individual j is replaced by individual i in X due to a selection event at time $t \in I_m$ then the particle

labeled i in $\mathcal{Z}^m$ splits at time $t - m\mathcal{T} \in J_m$. If particle i has a higher type than particle j then the new particle is paired with individual j. The particle that was paired with individual j before the branching event is no longer paired with any individual in X. If particle i has a lower type than particle j then the particle that was paired with individual j remains paired with individual j. The new particle is not paired with any individual in X.

- A particle labeled i in $\mathcal{Z}^m$ splits at a time dependent rate $\gamma(R_t^{i,k} - U_t^i)$ for all times $t \in J_m$ where $R_t^{i,k}$ is the type of particle i. These branching events are independent of any of the events in X. When such a branching event occurs, the new particle is not paired with any individual in X.

- Any particles in $\mathcal{Z}^m$ that are not paired with a individual in X branch and acquire mutations independently of X.

Observe the following bound for D_t:

$$
D_t \leq \sum_{i=1}^{N_t-1} (D_{t_{i+1}} - D_{s_i})
$$

$$
+ \sum_{i=1}^{N_t} (D_{s_i} - D_{t_i}) + \sup_{s_{N_t} \leq s \leq t_{N_t+1}} (D_s - D_{s_{N_t}}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}}),
$$

where we consider the supremum over the empty set to be 0. By definition we have

$$
\sum_{i=1}^{N_t-1} (D_{t_{i+1}} - D_{s_i}) + \sup_{s_{N_t} \leq s \leq t_{N_t+1}} (D_s - D_{s_{N_t}}) \leq \sum_{i=1}^{N_t} Y_i.
$$

To finish the proof we will show

$$
\sum_{i=1}^{N_t} \sup_{t_i \leq s \leq s_i} (D_s - D_{t_i}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}}) \leq D_t'.
$$

To do this we define

$$
M_t = \sum_{i=1}^{N_t} \sup_{t_i \leq s \leq s_i} (D_s - D_{t_i}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}})
$$

for all times $t \geq 0$. Suppose $M_s \leq D_s'$ for all $s \in [0, t)$ and a mutation, resampling or selection event occurs in X at time t. If $t \in (s_i, t_{i+1})$ for some $i \geq 0$ then

$M_{t-} = M_t$ because the process M does not change on these time intervals. It is possible that D'_t changes, but D'_t can only increase. Therefore, $D'_t \geq M_t$. If $t \in [t_i, s_i] \cap (m\mathcal{T}, (m+1)\mathcal{T}]$ for some $i \geq 0$ and $m \geq 0$ then at time t the processes X and X' are coupled. More precisely, X and $\mathcal{Z}^m$ are coupled and the coupling has the same dynamics as the coupling in Proposition 9 except the time shift. The same argument used in Proposition 9 shows that $D'_t \geq M_t$ whether the individual changed fitness due to mutation, resampling or selection. Since this inequality is preserved on any event that may change M_t, it is true for all times t. $\qquad \square$

2.3 Bounding the rate when the width is large

We consider what happens when the width is large in this section. By large width we mean $W_t \geq \mathcal{W}$. The statements in this section are easier to make when we consider an initial configuration of X such that $W_0 \geq \mathcal{W}$. Although the conditions of Theorem 1 state that $W_0 = 0$ we can wait for a random time τ so that $W_\tau \geq \mathcal{W}$ and apply the Strong Markov Property.

We begin this section by showing that when the width is large enough the selection mechanism will cause the width to decrease quickly. We give a labeling to the individuals that will help us in this regard. Define the following subsets of $\mathbb{R}$:

$$
\begin{aligned}
I_1 &= (-\infty, X_0^+ - \frac{3}{16}W_0] \\
I_2 &= (X_0^+ - \frac{3}{16}W_0, X_0^+ - \frac{2}{16}W_0] \\
I_3 &= (X_0^+ - \frac{2}{16}W_0, X_0^+ - \frac{1}{16}W_0] \\
I_4 &= (X_0^+ - \frac{1}{16}W_0, \infty)
\end{aligned}
$$

We will label each individual in X_0 with two labels. For the first labeling, we use $\mathfrak{a}$ to label the individuals in $I_1 \cup I_2$, we use $\mathfrak{b}$ to label the individuals in I_3 and we use $\mathfrak{c}$ to label the individuals in I_4. For the second labeling we use $\mathfrak{a}'$ to label the individuals in I_1, we use $\mathfrak{b}'$ to label the individuals in I_2 and we use $\mathfrak{c}'$ to label the individuals in $I_3 \cup I_4$.

Let $\mathfrak{A}_t$, $\mathfrak{B}_t$ and $\mathfrak{C}_t$ denote the number of individuals labeled $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ at time t, respectively. Let $\mathfrak{A}'_t$, $\mathfrak{B}'_t$ and $\mathfrak{C}'_t$ denote the number of individuals labeled $\mathfrak{a}'$, $\mathfrak{b}'$ and $\mathfrak{c}'$ at time t, respectively.

The individuals change labels over time according to the following dynamics:

- Mutations: If the fitness of a individual labeled $\mathfrak{a}$ increases so that it is in I_3 then the individual is relabeled $\mathfrak{b}$. If the fitness of a individual labeled $\mathfrak{a}'$ increases so that it is in I_2 then the individual is relabeled $\mathfrak{b}'$. Likewise, if the fitness of a individual labeled $\mathfrak{b}$ increases so that it is in I_4 then it is relabeled $\mathfrak{c}$ and if the fitness of a individual labeled $\mathfrak{b}'$ increases so that it is in I_3 then it is relabeled $\mathfrak{c}'$. Deleterious mutations do not cause individuals to be relabeled.

- Resampling: Any resampling event in which individual i is replaced by individual j causes individual i to inherit the labels of individual j.

- Selection: If a individual labeled $\mathfrak{a}$ is replaced due to a selection event it inherits the corresponding label of the individual that replaced it. If a individual labeled $\mathfrak{a}'$ is replaced due to a selection event it inherits the corresponding label of the individual that replaced it. If a individual labeled $\mathfrak{b}$ is replaced by a individual labeled $\mathfrak{c}$ due to a selection event then the individual that was labeled $\mathfrak{b}$ is relabeled $\mathfrak{c}$. If a individual labeled $\mathfrak{b}'$ is replaced by a individual labeled $\mathfrak{c}'$ due to a selection event then the individual that was labeled $\mathfrak{b}'$ is relabeled $\mathfrak{c}'$. Any other selection events do not cause the labels of the individuals to be changed.

Let A_1 be the event that there is a individual labeled $\mathfrak{b}$ with fitness in $(-\infty, X_0^+ - \frac{5}{32}W_0)$ for some time $t \in [0, \mathcal{T}]$. Let A_2 be the event that there is a individual labeled $\mathfrak{c}$ with fitness in $(-\infty, X_0^+ - \frac{3}{32}W_0)$ for some time $t \in [0, \mathcal{T}]$. Let A'_1 be the event that there is a individual labeled $\mathfrak{b}'$ with fitness in $(-\infty, X_0^+ - \frac{7}{32}W_0)$ for some time $t \in [0, \mathcal{T}]$. Let A'_2 be the event that there is a individual labeled $\mathfrak{c}'$ with fitness in $(-\infty, X_0^+ - \frac{5}{32}W_0)$ for some time $t \in [0, \mathcal{T}]$.

Lemma 13. *Suppose $W_0 \geq \mathcal{W}$ for all N. Then*

$$P(A_1 \cup A_2 \cup A_1' \cup A_2') \to 0 \ as \ N \to \infty.$$

Proof. First we show the result for A_1. We apply Corollary 11 by setting

$$x = X_0^+ - 2W_0/16, t = t_0 \text{ and } l = W_0/32.$$

Recall that we had defined $\mathcal{S}_t^x$ in Definition 10. Because $x = X_0^+ - 2W_0/16$ we have that $\mathcal{S}_0^x$ consists of all the individuals labeled $\mathfrak{b}$ or $\mathfrak{c}$. Setting $t = \mathcal{T}$ and $l = W_0/32$ will make $A_t^{x,l}$ the event that a individual labeled $\mathfrak{b}$ or $\mathfrak{c}$ has fitness less than $X_0^+ - \frac{5}{32}W_0$ by time $\mathcal{T}$. Note that according to the relabeling dynamics, individual i being labeled $\mathfrak{b}$ or $\mathfrak{c}$ is equivalent to $i \in \mathcal{S}^x$. Therefore, $A_1 \subset A_t^{x,l}$ and we get

$$P(A_1) \leq P(A_t^l) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{\lfloor l-1 \rfloor!}.$$

Applying Stirling's formula we have

$$\frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{\lfloor l-1 \rfloor!} \sim \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t+\lfloor l-1 \rfloor}}{\lfloor l-1 \rfloor^{\lfloor l-1 \rfloor}\sqrt{2\pi\lfloor l-1 \rfloor}} = e^x$$

where x is equal to

$$\log(2N)+l\log(t\mu)+(\gamma(W_0+2l)+\mu+1)t+\lfloor l-1 \rfloor - \lfloor l-1 \rfloor\log(\lfloor l-1 \rfloor)-\log(2\pi\lfloor l-1 \rfloor)/2.$$

As $N \to \infty$ we have $x \sim -\lfloor l-1 \rfloor \log(\lfloor l-1 \rfloor) \sim -w\log N/32$. Therefore,

$$P(A_1) \to 0 \text{ as } N \to \infty.$$

We can apply Corollary 11 with $x = X_0^+ - W_0/16$, $t = \mathcal{T}$ and $l = W_0/32$ to get the same bound for $P(A_2)$. By choosing x, t and l in this way the event $A_t^{x,l}$ is the event that a individual labeled $\mathfrak{c}$ has fitness less than $X^+(0) - \frac{3}{32}W_0$ by time $\mathcal{T}$. This shows that $P(A_2)$ also tends to 0 as N tends to infinity.

Likewise, to show $P(A_1')$ tends to 0 as N goes to infinity we can apply Corollary 11 with $x = X_0^+ - \frac{3}{16}W_0$, $t = \mathcal{T}$ and $l = W_0/32$, and to show $P(A_2')$ tends to 0 as N goes to infinity we can apply Corollary 11 with $x = X_0^+ - \frac{2}{16}W_0$, $t = \mathcal{T}$ and $l = W_0/32$. $\qquad\square$

Lemma 14. *Suppose $W_0 \geq \mathcal{W}$ for all N. Let T be a stopping time whose definition may depend on N such that $\mathfrak{C}'_T \geq N/4$ for all N. Let*

$$B_T = \inf\{t \geq T : X_t^- > X_0^+ - W_0/4\}.$$

Then

$$P(B_T 1_{\{T < \frac{1}{2}\mathcal{T}\}} > \frac{1}{2}\mathcal{T}) \to 0 \ as \ N \to \infty.$$

Proof. Let A'_3 be the event that $\mathfrak{C}'_t \geq N/5$ for all times $t \in [T, T + \frac{1}{2}\mathcal{T})$. The only way for a individual labeled $\mathfrak{c}'$ to change its label is for it to be replaced by a individual labeled $\mathfrak{a}'$ or $\mathfrak{b}'$ via a resampling event. The rate at which individuals marked $\mathfrak{c}'$ undergo resampling events with individuals marked $\mathfrak{a}'$ or $\mathfrak{b}'$ at time t is

$$\frac{\mathfrak{C}'_t(N - \mathfrak{C}'_t)}{N} \leq \frac{N}{4}.$$

Let $\{U_n\}_{n=0}^\infty$ be a simple random walk with $U_0 = N/4 \leq \mathfrak{C}'_T$. Denote by $T \leq t_1 < t_2 < \ldots$ the sequence of times at which individuals labeled $\mathfrak{c}'$ are involved in resampling events with individuals that are not labeled $\mathfrak{c}'$ after time T. We couple $\{U_n\}_{n=0}^\infty$ with X so that if at time t_n a individual is labeled $\mathfrak{c}'$ due to a resampling event then $U_n = U_{n-1} + 1$. If at time t_n a individual loses the label $\mathfrak{c}'$ due to a resampling event then $U_n = U_{n-1} - 1$. To have $U_m < N/5$ for some m satisfying $0 \leq m \leq n$ we will need $\max_{0 \leq m \leq n} |U_m - U_0| \geq N/20$. It follows from the reflection principle that there exists a constant C such that $E[\max_{0 \leq m \leq n} |U_m - U_0|] \leq C\sqrt{n}$ for all $n \geq 0$. By Markov's inequality,

$$P\left(\max_{0 \leq m \leq n} |U_m - U_0| \geq N/20\right) \leq C\sqrt{n}/N$$

for some constant C.

Let R be the number of resampling events that occur in the time interval $[T, T + \frac{1}{2}\mathcal{T})$ that involve pairs of individuals such that one is labeled $\mathfrak{c}'$ and the other is not. Using Lemma 12 and the fact that the rate at which resampling events occur is bounded above by $N/4$ we have

$$P(R > k) \leq \sum_{i=k+1}^\infty \frac{(N\mathcal{T})^i e^{-N\mathcal{T}/8}}{8^i i!} \leq \frac{(N\mathcal{T})^k}{8^k k!}.$$

Then

$$P((A_3')^C) \le P(\{\max_{0 \le m \le R} |U_m - U_0| \ge N/20\} \cap \{R \le N^{3/2}\})$$

$$+ P(\{\max_{0 \le m \le R} |U_m - U_0| \ge N/20\} \cap \{R > N^{3/2}\})$$

$$\le P(\{\max_{0 \le m \le N^{3/2}} |U_m - U_0| \ge N/20\}) + P(R > N^{3/2})$$

$$\le \frac{C}{N^{1/4}} + \frac{(N\mathcal{T})^{N^{3/2}}}{8^{N^{3/2}} \lceil N^{3/2} \rceil!}$$

$$\to 0 \text{ as } N \to \infty.$$

Let A_4' be the event that $\mathfrak{A}_t' = 0$ for some time $t \in [T, T + \frac{1}{2}\mathcal{T})$. Notice that if $\mathfrak{A}_t' = 0$ then $\mathfrak{A}_s' = 0$ for $s \ge t$. Therefore, A_4' is the event that the label $\mathfrak{a}'$ is eliminated by time $T + \frac{1}{2}\mathcal{T}$. By the given dynamics $\mathfrak{A}_t'$ can only increase when individuals marked $\mathfrak{a}'$ replace individuals marked $\mathfrak{b}'$ or $\mathfrak{c}'$ via resampling events. At time t the rate at which this happens is

$$\frac{1}{2} \cdot \frac{\mathfrak{A}_t'(N - \mathfrak{A}_t')}{N} \le \mathfrak{A}_t'. \tag{2.4}$$

We define the event $\mathcal{E}$ as

$$\mathcal{E} = (A_1')^C \cap (A_2')^C \cap A_3' \cap \{T < \frac{1}{2}\mathcal{T}\}.$$

Selection will cause $\mathfrak{A}'$ to decrease. On the event $(A_2')^C$ all of the individuals marked $\mathfrak{c}'$ will have fitness at least $\frac{1}{32}W_0$ greater than any individual marked $\mathfrak{a}$ until time t_0. Thus, on the event $(A_2')^C \cap \{T < \frac{1}{2}t_0\}$ all of the individuals marked $\mathfrak{c}'$ will have fitness at least $\frac{1}{32}W_0$ greater than any individual marked $\mathfrak{a}$ for all times $t \in [T, T + \frac{1}{2}\mathcal{T})$. On the event A_3' there are at least $N/5$ individuals marked $\mathfrak{c}$ for all times $t \in [T, T + \frac{1}{2}\mathcal{T})$. Hence, on the event $\mathcal{E}$ individuals marked $\mathfrak{a}'$ will become individuals marked $\mathfrak{c}'$ by a rate of at least

$$\frac{\gamma \mathfrak{A}_t' \mathfrak{C}_t' W_0}{32N} \ge \frac{\gamma}{160} W_0 \mathfrak{A}_t' \tag{2.5}$$

for all times $t \in [T, T + \frac{1}{2}\mathcal{T})$.

Let $\{U_n'\}$ be a biased random walk which goes up with probability

$$p' = \frac{160}{160 + \gamma W_0}$$

and down with probability $1 - p'$. Let N be large enough so that $p' < 1/2$. Because the random walk is biased downward, the probability that the random walk visits a state $j < U_0'$ is 1. Once the random walk is in state j, it goes up 1 with probability p' and will eventually return to j with probability 1. The random walk will go down 1 with probability $1 - p'$ and, from basic martingale arguments, the probability that it never returns to j again is $(1 - 2p')/(1 - p')$. Therefore, once U' is in state j, the probability it never returns to state j is

$$\frac{(1 - 2p')}{1 - p'} \cdot (1 - p') = 1 - 2p'.$$

Hence the number of times U' visits a state $j < U_0'$ has the geometric distribution with mean $1/(1 - 2p')$. For more details see pages 194-196 of [8].

By equations (2.4) and (2.5) we see that on the event $\mathcal{E}$, if $\mathfrak{A}'$ changes during the time interval $[T, T + \frac{1}{2}\mathcal{T})$ it decreases with probability higher than p'. The expected number of times that $\mathfrak{A}'$ will visit state j is therefore less than or equal to $1/(1 - 2p')$ for any $j \in \{1, 2 \ldots, N-1\}$. Also, the rate at which $\mathfrak{A}_t'$ changes state is at least

$$\frac{\gamma}{160} W_0 \mathfrak{A}_t'$$

for all times $t \in [T, T + \frac{1}{2}\mathcal{T})$ by equation (2.5). Let $\overline{A} = \{t \geq T : \mathfrak{A}_t' > 0\}$ and let λ be Lebesgue measure. Then

$$E[\lambda(\overline{A}) 1_{\mathcal{E}}] \leq \frac{160}{(1 - 2p')\gamma W_0} \sum_{j=1}^{N} \frac{1}{j} \sim \frac{160 \log N}{\gamma W_0}$$

as $N \to \infty$.

Observe that

$$P(\mathcal{E} \cap (A_4')^C) = P\left(\mathcal{E} \cap \left\{\lambda(\overline{A}) \geq \frac{1}{2}\mathcal{T}\right\}\right)$$

$$= P\left(\lambda(\overline{A}) 1_{\mathcal{E}} \geq \frac{1}{2}\mathcal{T}\right)$$

$$\leq \frac{2E[\lambda(\overline{A}) 1_{\mathcal{E}}]}{\mathcal{T}} \text{ by Markov's Inequality}$$

$$\to 0 \text{ as } N \to \infty.$$

Therefore,

$$P(\mathcal{E} \cap A_4') - P\left(T < \frac{1}{2}\mathcal{T}\right) \to 0 \text{ as } N \to \infty.$$

This allows us to do the following computation:

$$
\begin{aligned}
1 &= \lim_{N\to\infty}\left(P\left(T<\tfrac{1}{2}\mathcal{T}\right)+P\left(T\geq\tfrac{1}{2}\mathcal{T}\right)\right)\\
&= \lim_{N\to\infty}\left(P(\mathcal{E}\cap A_4')+P\left(T\geq\tfrac{1}{2}\mathcal{T}\right)\right)\\
&= \lim_{N\to\infty}\left(P\left((A_1')^C\cap(A_2')^C\cap A_3'\cap A_4'\cap\left\{T<\tfrac{1}{2}\mathcal{T}\right\}\right)+P\left(T\geq\tfrac{1}{2}\mathcal{T}\right)\right)\\
&\leq \lim_{N\to\infty}\left(P\left(\left\{B_T\leq\tfrac{1}{2}\mathcal{T}\right\}\cap\left\{T<\tfrac{1}{2}\mathcal{T}\right\}\right)+P\left(T\geq\tfrac{1}{2}\mathcal{T}\right)\right)\\
&= \lim_{N\to\infty}P\left(B_T 1_{\{T<\frac{1}{2}\mathcal{T}\}}\leq\tfrac{1}{2}\mathcal{T}\right).
\end{aligned}
$$

$\square$

Let $B=\inf\{t:X_t^->X_0^+-W_0/4\}$.

Proposition 15. *Suppose $W_0\geq\mathcal{W}$ for all N. As N tends to infinity,*

$$
P(B>\mathcal{T})\to 0.
$$

Proof. First note that if $\mathfrak{B}_0+\mathfrak{C}_0\geq N/4$ then, because all of the individuals labeled $\mathfrak{b}$ or $\mathfrak{c}$ at time 0 are also labeled $\mathfrak{c}'$, we have that $\mathfrak{C}_0'\geq N/4$. The result then follows by Lemma 14 with $T=0$. On the other hand, if $\mathfrak{B}_0+\mathfrak{C}_0<N/4$ then $\mathfrak{A}_0\geq 3N/4$.

Let $T=(\inf\{t:\mathfrak{A}_t<N/4\})\wedge(\inf\{t:\mathfrak{C}_t\geq N/4\})$. Let A_5 be the event that $\mathfrak{A}_t\geq N/4$ for all times $t\in[0,\tfrac{1}{2}\mathcal{T})$. Let A_6 be the event that $\mathfrak{C}_t<N/4$ for all times $t\in[0,\tfrac{1}{2}\mathcal{T})$. Define ζ to be the infimum over all times such that a individual labeled $\mathfrak{b}$ has fitness in $(-\infty,X_0^+-\tfrac{5}{32}W_0)$, a individual labeled $\mathfrak{c}$ has fitness in $(-\infty,X_0^+-\tfrac{3}{32}W_0)$, or $\mathfrak{A}_t<N/4$. Note that $A_1^C\cap A_2^C\cap A_5\subset\{\zeta\geq\tfrac{1}{2}\mathcal{T}\}$.

On the event $\{\zeta\geq\tfrac{1}{2}\mathcal{T}\}$ the rate of increase of $\mathfrak{C}_t$ due to selection is at least

$$
\frac{\gamma\mathfrak{A}_t\mathfrak{C}_t W_0}{32N}\geq\frac{1}{128}\gamma\mathfrak{C}_t W_0 \tag{2.6}
$$

for all $t\in[0,\tfrac{1}{2}\mathcal{T})$. On the other hand, because $\mathfrak{C}_t$ can only decrease due to resampling, $\mathfrak{C}_t$ will decrease no faster than

$$
\frac{1}{2}\cdot\frac{\mathfrak{C}_t(N-\mathfrak{C}_t)}{N}\leq\mathfrak{C}_t. \tag{2.7}
$$

Let $\{U_n\}_{n=0}^{\infty}$ be a biased random walk with $U_0 = 1$ which goes up with probability

$$p = \frac{\gamma W_0}{128 + \gamma W_0}$$

and down with probability $1 - p$. Let N be large enough so that $p > 1/2$. By similar reasoning as used in the proof of Lemma 14, the number of times U_n visits a state $j \geq 1$ has the geometric distribution with mean $1/(2p - 1)$. Also, by basic martingale arguments, the probability that U_n ever reaches state 0 is

$$\frac{1 - p}{p} = \frac{128}{\gamma W_0}.$$

Note that $\mathfrak{C}_0 \geq U_0$ since the individual with the highest fitness is initially labeled $\mathfrak{c}$. On the event $\{\zeta \geq \frac{1}{2}\mathcal{T}\}$, we see from equations (2.6) and (2.7) that if $\mathfrak{C}$ changes during time $[0, \frac{1}{2}\mathcal{T})$ then it increases with a probability of at least p. Therefore, the expected number of times that $\mathfrak{C}$ visits state j is less than or equal to $1/(2p - 1)$ and the probability the $\mathfrak{C}_t$ reaches state 0 for some time $t \in [0, \frac{1}{2}\mathcal{T})$ is less than $128/(\gamma W_0)$. Let A_7 be the event that $\mathfrak{C}_t$ reaches state 0 for some time $t \in [0, \frac{1}{2}\mathcal{T})$.

By equation (2.6), the rate at which $\mathfrak{C}$ changes is at least

$$\frac{1}{128}\gamma \mathfrak{C}_t W_0$$

for all times $t \in [0, \frac{1}{2}\mathcal{T})$ on the event $\{\zeta > \frac{1}{2}\mathcal{T}\}$. Let $\overline{C} = \{t \in [0, \frac{1}{2}\mathcal{T}) : \mathfrak{C} < \frac{1}{4}N\}$ and let λ be Lebesgue measure. Then

$$E[\lambda(\overline{C})1_{\{\zeta \geq \mathcal{T}/2\}}] = E[\lambda(\overline{C})1_{\{\zeta \geq \mathcal{T}/2\}}1_{A_7}] + E[\lambda(\overline{C})1_{\{\zeta \geq \mathcal{T}/2\}}1_{A_7^c}]$$

$$\leq \frac{1}{2}\mathcal{T}P(A_7) + \frac{128}{(2p - 1)\gamma W_0}\sum_{j=1}^{\lfloor N/4 \rfloor}\frac{1}{j}$$

$$\sim \frac{128\log(N/4)}{\gamma W_0}.$$

By Markov's inequality

$$P(A_1^C \cap A_2^C \cap A_5 \cap A_6) \leq P(A_1^C \cap A_2^C \cap A_5 \cap \{\lambda(\overline{C}) \geq \frac{1}{2}\mathcal{T}\})$$

$$\leq P(\{\zeta \geq \frac{1}{2}\mathcal{T}\} \cap \{\lambda(\overline{C}) \geq \frac{1}{2}\mathcal{T}\})$$

$$= P(\lambda(\overline{C})1_{\{\zeta \geq \mathcal{T}/2\}} \geq \frac{1}{2}\mathcal{T})$$

$$\leq \frac{2E[\lambda(\overline{C})1_{\{\zeta \geq \mathcal{T}/2\}}]}{\mathcal{T}}$$

$$\leq \frac{256w^{1/4}\log(N/4)}{\mathcal{T}\gamma W_0} \text{ for } N \text{ large enough}$$

$$\to 0 \text{ as } N \to \infty.$$

Because $P(A_1^C \cap A_2^C) \to 1$ we have $P(A_5^C \cup A_6^C) \to 1$ as $N \to \infty$.

Note that $A_5^C \cup A_6^C \subset \{T < \frac{1}{2}\mathcal{T}\}$. Therefore, $P(T < \frac{1}{2}\mathcal{T}) \to 1$ as $N \to \infty$. Let $E_2 = (A_1')^C \cap (A_2')^C \cap \{T < \frac{1}{2}\mathcal{T}\}$. Then $P(E_2) \to 1$ as $N \to \infty$. To show $P(B \leq \mathcal{T}) \to 1$ we can show $P(\{B \leq \mathcal{T}\} \cap E_2) \to 1$. At time T, at least $\frac{1}{4}N$ individuals will be labeled either $\mathfrak{b}$ or $\mathfrak{c}$. According to the labeling all of these individuals are labeled $\mathfrak{c}'$ so that at time T we have $\mathfrak{C}_T \geq \frac{1}{4}N$. By Lemma 14 we have

$$P\left(B_T 1_{\{T < \frac{1}{2}\mathcal{T}\}} \leq \frac{1}{2}\mathcal{T}\right) \to 1 \text{ as } N \to \infty.$$

Note that

$$\left\{B_T 1_{\{T < \frac{1}{2}\mathcal{T}\}} \leq \frac{1}{2}\mathcal{T}\right\} = \left\{B_T \leq \frac{1}{2}\mathcal{T}\right\} \cup \left\{T \geq \frac{1}{2}\mathcal{T}\right\}.$$

Because $E_2 \subset \{T < \frac{1}{2}\mathcal{T}\}$ we have

$$\left\{B_T 1_{\{T < \frac{1}{2}\mathcal{T}\}} \leq \frac{1}{2}\mathcal{T}\right\} \cap E_2 = \left\{B_T \leq \frac{1}{2}\mathcal{T}\right\} \cap E_2.$$

It then follows that

$$P\left(\left\{B_T \leq \frac{1}{2}\mathcal{T}\right\} \cap E_2\right) \to 1 \text{ as } N \to \infty.$$

However,

$$\left\{B_T \leq \frac{1}{2}\mathcal{T}\right\} \cap E_2 \subset \left\{B_T \leq \frac{1}{2}\mathcal{T}\right\} \cap \left\{T < \frac{1}{2}\mathcal{T}\right\} \subset \{B \leq \mathcal{T}\}$$

which gives the conclusion. $\qquad\square$

Let $V_t^1 = \{i : X_t^i > X_0^+ + W_0/4\}$ and $V_t^2 = \{i : X_t^i < X_0^- - W_0/4\}$. Let $F = \inf\{t : V_t^1 \cup V_t^2 \neq \emptyset\}$. We now want to bound the time it takes for the width to increase.

Proposition 16. *Suppose $W_0 \geq \mathcal{W}$ for all N. Then*

$$\lim_{N\to\infty} P(F > \mathcal{T}) = 1.$$

Proof. By Proposition 9 with $l = W_0/4$ and $t = \mathcal{T}$ we have

$$P(\inf\{s : V_s^1 \neq \emptyset\} < t) = P\left(\sup_{0 \leq s \leq t} D_s \geq l\right)$$
$$\leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}$$
$$\to 0 \text{ as } N \to \infty.$$

By Proposition 7 with $l = W_0/4$ and $t = \mathcal{T}$ we have

$$P(\inf\{s : V_s^2 \neq \emptyset\} < t) = P\left(\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \geq l\right)$$
$$\leq \frac{N(t\mu)^l e^t}{l!}$$
$$\to 0 \text{ as } N \to \infty.$$

$\square$

Recall that $Y_i = \sup_{s_i \leq s \leq t_{i+1}} D_s - D_{s_i}$ and that $\{\mathcal{F}_t\}_{t \geq 0}$ is the natural filtration associated with X. Note that if $W_0 < 2\mathcal{W}$ then for all $n \geq 1$ the width satisfies $W_{s_n} = \lceil 2\mathcal{W} \rceil$.

Proof of Proposition 4. We consider a sequence of initial configurations X_0 depending on N such that $W_0 = \lceil 2\mathcal{W} \rceil$ for all N. Because $W_0 \geq 2\mathcal{W}$ we have $s_1 = 0$ and $Y_1 = \sup_{0 \leq s \leq t_2} D_s - D_0$. We will show that for N large enough, $E[Y_1] < 5\mathcal{W}$. The result then follows because X is a strong Markov process.

We make the following definitions:

$$V_t^1(s) = \{i : X_t^i > X_s^+ + W_s/4\} \text{ for } t \geq s \geq 0$$

$$V_t^2(s) = \{i : X_t^i < X_s^- - W_s/4\} \text{ for } t \geq s \geq 0$$

$$F_0 = B_0 = r_0 = 0$$

$$F_n = \inf\{t \geq r_{n-1} : V_t^1(r_{n-1}) \cup V_t^2(r_{n-1}) \neq \emptyset\} \text{ for } n \geq 1$$

$$B_n = \inf\{t \geq r_{n-1} : X_t^- > X_{r_{n-1}}^+ - W_{r_{n-1}}/4\} \text{ for } n \geq 1$$

$$r_n = F_n \wedge B_n \text{ for } n \geq 1$$

$$n_* = \inf\{n \geq 1 : W_{r_n} < \mathcal{W}\}.$$

Note that r_1 is the first time that the event $F \cup B$ occurs and that, conceptually, r_n acts like the first time that $F \cup B$ occurs when the process is started at time r_{n-1} for $n \geq 2$. The random variables F_n and B_n play the roles of the events F and B when the processes are started at time r_{n-1}.

On the event $n - 1 < n_*$, by Proposition 15 and the strong Markov property of X we have $P(B_n \leq r_{n-1} + \mathcal{T} | \mathcal{F}_{r_{n-1}}) \to 1$ uniformly on a set of probability 1 as $N \to \infty$. Likewise, on the event $n - 1 < n_*$, by Proposition 16 and the strong Markov property we have $P(F_n > r_{n-1} + \mathcal{T} | \mathcal{F}_{r_{n-1}}) \to 1$ uniformly on a set of probability 1 as $N \to \infty$. Therefore, on the event $n - 1 < n_*$ we have $P(B_n < F_n | \mathcal{F}_{r_{n-1}}) \to 1$ uniformly on a set of probability 1.

Because the bounds in Propositions 15 and 16 do not depend on n we can choose a sequence $p = p_N$ such that $p \to 1$ as $N \to \infty$ and almost surely

$$p 1_{\{n-1 < n_*\}} \leq P(B_n < F_n | \mathcal{F}_{r_{n-1}}) 1_{\{n-1 < n_*\}}$$

for all $n \geq 0$. Let $\{S_n\}_{n=0}^\infty$ be a random walk starting at 1 which goes down 1 with probability p and up 1 with probability $1 - p$ until it reaches 0. Once S reaches 0 it is fixed. For $n < n_*$ we couple S with X so that $2^{S_n-1}W_0 \geq W_{r_n}$. The coupling is defined as follows:

- Each step of the process S corresponds to a time r_n.

- On the event $\{F_n < B_n\}$ we have $S_n - S_{n-1} = 1$.

- On the event $\{B_n \leq F_n\}$ with probability $p/P(B_n \leq F_n)$ we have $S_n - S_{n-1} = -1$ and with probability $1 - p/P(B_n \leq F_n)$.

We will show that this coupling is well-defined and gives the necessary bound. Initially, $S_0 = 1$ and $2^{S_0-1}W_0 = W_0$. On the event that $B_n \leq F_n$, we have $W_{r_n} < \frac{1}{2}W_{r_{n-1}}$ and $\sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}} \leq \frac{1}{4}W_{r_{n-1}}$. On the event that $F_n < B_n$, we have $W_{r_n} < 2W_{r_{n-1}}$ and $\sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}} \leq \frac{1}{4}W_{r_{n-1}} + 1$. Therefore, if $2^{S_{n-1}-1}W_0 \geq W_{r_{n-1}}$ then $2^{S_n-1}W_0 \geq W_{r_n}$ by the coupling. It follows that $2^{S_n-1}W_0 \geq \sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}}$ as well. By induction, $2^{S_n-1}W_0 \geq W_{r_n}$ for all $n < n_* \wedge \inf\{m : S_m = 0\}$. If $n = \inf\{m : S_m = 0\}$ then $W_{r_n} \leq \mathcal{W}$. Therefore $n_* \leq \inf\{m : S_m = 0\}$ and the induction holds for all $n < n_*$.

We define a function d on $(\{0\} \cup \mathbb{N})^\infty$ such that if $x = (x_0, x_1, \dots)$ then

$$d(x) = \sum_{i=0}^{\infty} 1_{\{x_i > 0\}} 2^{x_i - 1} W_0.$$

Consider $S = (S_0, S_1, \dots)$ as a random element in $(\{0\} \cup \mathbb{N})^\infty$. Then

$$d((S_0, S_1, \dots, S_n, 0, 0, \dots)) \geq \sum_{i=1}^{n} \left(\sup_{r_{i-1} \leq t \leq r_i} D_t - D_{r_{i-1}} \right) \geq \sup_{0 \leq t \leq r_n} D_t$$

for all n such that $n - 1 < n_*$. By definition, n_* is the first n such that $W_{r_n} < \mathcal{W}$. Hence $d(S) \geq Y_1$.

For any $n \geq 0$ we have

$$P(S_{2n+1} = 0) = \binom{2n+1}{n}(1-p)^n p^{n+1} \leq 4^n (1-p)^n p^{n+1}.$$

If $S_{2n+1} = 0$ then

$$d(S) \leq \left(2 + 2 \sum_{i=1}^{n} 2^{i-1} \right) W_0 = 2^{n+1} W_0$$

which is obtained by taking n steps up followed by $n + 1$ steps down.

Therefore,

$$E[Y_1] \leq E[d(S)] \leq \sum_{n=0}^{\infty} [4(1-p)]^n p^{n+1} 2^{n+1} W_0 = \frac{2pW_0}{1 - 8(1-p)p} \sim 4\mathcal{W}$$

because $W_0 = \lceil 2\mathcal{W} \rceil$ and $p \to 1$ as $N \to \infty$. This shows that for N large enough we have $E[Y_1] < 5\mathcal{W}$, which gives the conclusion. $\qquad\square$

Let $l = \lfloor \mathcal{W}/2 \rfloor$. We make the following definitions for the rest of the section:

$$K_1 = \frac{2N(\mathcal{T}\mu)^l e^{(\gamma(W_0 + 2l) + \mu + 1)\mathcal{T}}}{(l-1)!}$$

$$K_2 = \frac{N(\mathcal{T}\mu)^l e^{\mathcal{T}}}{l!}$$

$$p = 1 - K_1 - K_2.$$

Lemma 17. *Suppose $W_0 \leq \mathcal{W}$ for all N. Then*

$$P\left(\sup_{0 \leq s \leq \mathcal{T}} W_s \leq 2\mathcal{W}\right) \geq 1 - K_1 - K_2.$$

Proof. By Proposition 9 we have

$$P\left(\sup_{0 \leq s \leq \mathcal{T}} D_s \geq l\right) \leq K_1.$$

By Proposition 7 we have

$$P\left(\sup_{0 \leq s \leq \mathcal{T}} (X_0^- - X_s^-) \geq l\right) \leq K_2.$$

On the event that $\sup_{0 \leq s \leq t} D_s \leq \mathcal{W}/2$ and $\sup_{0 \leq s \leq t} X_0^- - X_s^- \leq \mathcal{W}/2$ we have $\sup_{0 \leq s \leq t} W_t \leq 2\mathcal{W}$. This gives the result. $\qquad \square$

Proof of Proposition 5. Notice that

$$\{N_s \geq i\} = \{s_i \leq s\} \subset \left\{\sum_{j=1}^{i}(s_j - t_j) \leq s\right\}.$$

Therefore,

$$P(N_s \geq i) \leq P\left(\sum_{j=1}^{i}(s_j - t_j) \leq s\right).$$

Applying Lemma 17 and the strong Markov property of X we have

$$1 - K_1 - K_2 \leq P\left(s_j - t_j \geq \mathcal{T}|\mathcal{F}_{t_j}\right)$$

for all j. Taking expectations of both sides yields

$$1 - K_1 - K_2 \leq P(s_j - t_j \geq \mathcal{T})$$

for all j, so

$$1 - K_1 - K_2 \leq \inf_j P(s_j - t_j \geq \mathcal{T}).$$

Note that $p \to 1$ as $N \to \infty$. Define an i.i.d. sequence $\{V_i\}_{i=1}^{\infty}$ of random variables with distribution $P(V_i = 0) = 1 - p$ and $P(V_i = \mathcal{T}) = p$. Then

$$P\left(\sum_{j=1}^{i}(s_j - t_j) \leq s\right) \leq P\left(\sum_{j=1}^{i} V_i \leq s\right).$$

This will allow us to define a new process N_s' such that $N_s' = i$ if

$$\sum_{j=1}^{i} V_i \leq s < \sum_{j=1}^{i+1} V_i.$$

Note that $P(N_s' = 0) = p$ for $s \in [0, \mathcal{T})$ and that $P(N_s' \geq k) \geq P(N_s \geq k)$ for all k. Therefore, it is enough to bound $E[N_s']/s$.

Let $V_0 = 0$. Jumps of the process N_s' only occur at points $k\mathcal{T}$ where k is a positive integer. On the time interval $[0, \mathcal{T})$ the process N_s' is constant and has value $\max\{i \geq 0 : V_i = 0\}$. Therefore N_s' has the shifted geometric distribution for $s \in [0, \mathcal{T})$ with mean $(1-p)/p$. We can now make use of the fact that N_s' is a Markov process. If we consider values at $k\mathcal{T}$ for $k \geq 0$ we have for $s \in [(k-1)\mathcal{T}, k\mathcal{T})$ that $E[N_s'] = k(1-p)/p$. For $k \geq 2$ we then have

$$\frac{1}{s}E[N_s'] = \frac{k(1-p)}{sp} \leq \frac{k(1-p)}{(k-1)p\mathcal{T}}.$$

This gives us

$$\frac{\mathcal{T}}{s}E[N_s'] \leq \frac{k(1-p)}{(k-1)p} \to 0 \text{ as } N \to \infty.$$

On the time interval $[0, \mathcal{T})$ we have

$$\frac{\mathcal{T}}{s}E[N_s'] \leq \frac{(1-p)}{p} \to 0 \text{ as } N \to \infty.$$

$\square$

Chapter 2, in full, has been accepted for publication in the Annals of Applied Probability. The dissertation author was the sole author of this paper.

Chapter 3

A Hierarchical Probability Model of Colon Cancer

Cairns [3] first raised the question of how stem cells affect the development of cancer in 2006. We are interested in a particular model in which stem cells play a central role. Komarova [23] discusses three mathematical models which may be used to model the mutations that lead to cancer. The first is the Moran model, which may be used to model cancers in liquids such as Leukemia. This model is the same as the stochastic process $\{X_t : t \geq 0\}$ from chapter 2 except that there is no selection mechanism and a mutation on individual i only causes X^i to increase by 1. The second is a spatial model which may be used to model cancers in solid tissues. This model is similar to the Moran model except that the cells are given spatial locations and when they die they are only replaced by nearby cells. The third model, which is referred to as the hierarchical model, differentiates between stem cells and daughter cells. The hierarchical model is the focus of this chapter.

The hierarchical model was originally proposed in [23] as a model of colorectal cancer. As discussed in [23], many cells in the human body, including those in the colon, go through a three step process. It begins with a stem cell which will stay in the population for a long time and have many descendants. Some of these descendants will also be stem cells, but others will be differentiated progenitor cells. The progenitor cells, or what we shall refer to as daughter cells, will split into more daughter cells. The number of times these cells split is dependent

upon what organ of the body they are in. We will refer to the number of splits that a daughter cell has undergone as the generation of the cell. Once the cells split enough times they reach maturity and are swept out of the population in a biological process called apoptosis. The colon is lined with crypts that contain pockets of cells. The cells in the colon, as described by Komarova [26], are such that stem cells reside at the bottom of the crypt and the daughters migrate up the crypt so that the higher generation daughter cells are near the top.

In the original model proposed in [23], cancer is the result of two mutations. The reason for two mutations is that it represents the inactivation of two alleles in a tumor suppressor gene. Knudson claims that retinoblastoma is the result of two mutations in [21, 22]. For other sources on two mutation models of cancer one can refer to [18], [25] and [29]. We also model cancer as a result of two mutations.

The hierarchical model shall be referred to as H_1. This model has a fixed population of size $N = 2^l$ where l is the number of generations of daughter cells in the crypt. At all times $t \geq 0$ there is one stem cell and for $k \in \{1, 2, \ldots, l\}$ there are 2^{k-1} daughter cells of generation k. We start with a full crypt and no mutations. At each integral time unit all of the cells split in the following way:

- The stem cell splits into a stem cell and a generation 1 daughter cell.

- For each generation k with $1 \leq k \leq l - 1$, a daughter cell of generation k will split into two cells of generation $k + 1$.

- The daughter cells of generation l undergo apoptosis and are swept from the population.

Notice that the generations are constant size throughout time. The cells will accumulate mutations via Poisson processes. A cell with 0, 1 or 2 mutations is called a type-0, type-1 or type-2 cell respectively. A mutation which occurs on a type-0 or type-1 cell is called a type-1 or type-2 mutation respectively. This terminology is used so that a mutation that makes a cell type-2 is called a type-2 mutation. Once a type-2 mutation occurs the colon is assumed to have cancer. The cells will each have two Poisson processes marking them, one which will cause type-1 mutations and one which will cause type-2 mutations. The first Poisson

process that marks a cell will only cause a type-1 mutation if the cell is a type-0. If a mark of the Poisson process occurs while the cell is not a type-0 then nothing happens. Likewise, the second Poisson process only causes mutations on type-1 cells. If a mark from this Poisson process occurs on a cell while it is type-1 then the cell becomes type-2, but if the cell is not a type-1 then nothing happens. All of the Poisson processes are independent. The mutations are passed to the descendants when a cell splits. It is sometimes convenient to think of the cells as fixed in a binary tree and the mutations as traveling through the tree in a direction which takes them from the root to the leaves. Because of this we will often refer to the sequence of stem cells as the stem cell line and we fix the Poisson processes that are marking the cells on particular locations in the tree.

For our model, the rates at which stem cells acquire type-1 and type-2 mutations are u_1 and u_2 respectively. The rates at which the daughter cells acquire type-1 and type-2 mutations are v_1 and v_2 respectively. All of the rates are functions of N and will approach 0 as N approaches infinity. We will always consider what happens as N goes to infinity. All limits will be assumed as taking N to infinity unless otherwise stated.

We should mention that several other very similar models have been used to study how stem cells affect the development of cancer. In [24], Komarova and Cheng consider the effects of the development of cancer based on the quantity of stem cells in the population. In [13], Frank, Iwasa and Nowak consider a model in which the stem cells only split finitely many times.

In the hierarchical model there are three ways in which the mutations may occur. Stem cells may acquire both mutations so that cancer is a result of mutations of stem cells only. It is possible that a stem cell receives the first mutation and a daughter cell gets the second, or a daughter cell and one of its descendants will each receive mutations before they are swept from the crypt. In [23] these cases are abbreviated ss, sd and dd respectively.

A type-1 mutation to a cell is called successful if that cell or one of its descendants receives a type-2 mutation. A type-1 mutation to a stem cell is always successful and a type-1 mutation to a daughter cell is successful if the daughter cell

has a type-2 descendent before its progeny is eliminated from the population. We will call the successful type-1 mutation whose type-2 descendant is the first type-2 to occur the cancer causing type-1 mutation. Note that being the cancer causing type-1 mutation is not equivalent to being the first successful type-1 mutation.

We prove the theorem by coupling various models. This motivates us to define the following functions.

- $\tau'(A)$ is the time at which the cancer causing type-1 mutation occurs in model A.

- $\tau(A)$ is the first time that any cell gets a type-2 mutations in model A.

- $\sigma(A) := j/l$ when the cancer causing type-1 mutation occurs in generation j in model A. If the cancer causing type-1 mutation occurs on a stem cell in model A then $\sigma(A) = 0$.

- $\rho(A) := j/l$ when the first type-2 mutation occurs in generation j in model A. If the first type-2 mutation occurs on a stem cell in model A then $\rho(A) = 0$.

One of the two goals of this chapter is to find the asymptotic distribution of $\tau(H_1)$ as N approaches infinity. Similar work has been done for the Moran model by Schweinsberg [33] and Durrett, Schmidt and Schweinsberg [11], in which more general results have already been found, and for the spatial model by Durrett and Moseley [10]. In [23], Komarova makes a connection between the Moran model and the hierarchical one as follows: In the Moran model a mutation may undergo fixation, meaning it spreads throughout the entire population through the birth-death process and all of the cells are the same type. Because the last generation is always removed in the hierarchical model, the only way to get fixation is if a stem cell gets a mutation. These are the cases ss and sd. In these cases the mutation will spread throughout the population in l time units. In the Moran model it is also possible that the progeny of mutated cells undergo what is called stochastic tunneling. This is when multiple mutations are acquired before they fixate. This is analogous to daughter cells acquiring two mutations before a stem cell gets one mutation in the hierarchical model. This is the dd case and can also happen in

the sd case if the second mutation occurs before the first has time to fixate (in particular the second mutation occurs in less than l time units).

The rate at which daughter cells get successful type-1 mutations is given in [23] to be approximately

$$\sum_{i=1}^{l} v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-2)}). \tag{3.1}$$

To see this, suppose all the cells are type-0. When all of the cells in generation i are type-0 then type-1 mutations occur on generation i at rate $v_1 2^{i-1}$. Each of the cells will have $2^{l-i+1} - 2$ descendants. Every descendant lives for one time unit and gets type-2 mutations at rate v_2. This gives the probability of success of a type-1 mutation in generation i to be approximately $1 - e^{-v_2(2^{l-i+1}-2)}$. Then we sum over all generations.

Our second goal is to determine the limiting distributions of $\sigma(H_1)$ and $\rho(H_1)$. The location of the mutations can be essential to the treatment of cancer. As an example, studies of the effects of the drug imatinib on chronic myeloid leukemia have shown that leukemic stem cells will most likely not cause tumors but rather that a tumor is a result of a mutation on one of the daughter cells, see Dingli and Michor [7] and Michor [27]. Imatinib treats leukemic daughter cells but not leukemic stem cells. While using imatinib problems arising from cancer are prevented but patients cannot stop treatment because the leukemic stem cells will continue producing new leukemic daughter cells. Therefore, the location of where the mutations occur may play a pivotal role in determining how to treat the cancer.

We do not find the limiting distribution of $\tau'(H_1)$ as there seems to be no motivation to do so. We only make the definition $\tau'(A)$ because it will occasionally be useful for achieving the two goals described above.

We have established most of the notation for chapter 2 above but some more will be included here. For any real number a we define $a^+ = a \vee 0$. For functions $f(x)$ and $g(x)$ we will denote the limits $f(x)/g(x) \to 0$, $f(x)/g(x) \to 1$, and $f(x)/g(x) \to \infty$ as $x \to \infty$ by $f \ll g$, $f \sim g$ and $f \gg g$ respectively. To reduce the number of subscripts, we will use $\log x$ for $\log_2 x$. Note that with this notation

$l = \log N$. We will use $\to_d$ to denote convergence in distribution and $\to_p$ to denote convergence in probability. We make the following assumptions throughout most of chapter 3.

Assumption 1: There exist constants $\alpha, \beta > 0$ such that $v_2 \sim \beta N^{-\alpha}$.

Assumption 2: The mutation rates satisfy $u_1 \leq u_2$ and $v_1 \leq cv_2$ for some $c > 0$.

We do not allow $\alpha = 0$ so as to reduce the number of cases to be considered. As a result of Assumption 1, the probability that the cancer causing type-1 mutation occurs on a daughter cell in generation $i < l(1 - \alpha)^+$ tends to 0. According to Komarova in [24], Assumption 2 agrees with almost all of the biologically relevant cases. We let X be an exponentially distributed random variable with mean 1 and we let Y be a random variable with the Rayleigh distribution so that $P(Y \leq t) = 1 - e^{-t^2/2}$ for any $t > 0$.

The following theorem is proved in chapter 2.

Theorem 18. *Suppose Assumptions 1 and 2 hold. Recall that all limits are taken as N goes to infinity.*

1. *If $v_1 v_2 \ll 1/(N(\log N)^2)$ and $v_1 v_2 N \log N \gg u_1$ then*

$$(\alpha \wedge 1)v_1 v_2 N(\log N)\tau(H_1) \to_d X.$$

   The distribution of $\sigma(H_1)$ converges to the uniform distribution on the interval $((1 - \alpha)^+, 1]$ and $\rho(H_1)$ converges in probability to 1.

2. *If $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$ and $v_1 v_2 \gg u_1^2/N$ then*

$$\sqrt{v_1 v_2 N}\tau(H_1) \to_d Y.$$

   Both $\sigma(H_1)$ and $\rho(H_1)$ converge in probability to 1.

3. *If $v_1 v_2 \gg 1/N$ then*

$$\sqrt{v_1 v_2 N}\tau(H_1) \to_d Y.$$

   Both $\sigma(H_1)$ and $\rho(H_1)$ converge in probability to 1.

4. *If we have the following two conditions:*

   - *Either*
     $$v_1 v_2 \ll 1/(N(\log N)^2) \ \text{and} \ u_1 \gg v_1 v_2 N \log N$$

     or
     $$1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N \ \text{and} \ u_1 \gg \sqrt{v_1 v_2 N}$$

   - *Both $u_2 \ll 1/\log N$ and $u_2 \ll v_2 N$*

   then
   $$u_1 \tau(H_1) \to_d X.$$

   The probability that the first mutation occurs on the stem cell line converges to 1 and $\rho(H_1)$ converges in probability to $\alpha \wedge 1$.

5. *If we have the following two conditions:*

   - *Either*
     $$v_1 v_2 \ll 1/(N(\log N)^2) \ \text{and} \ u_1 \gg v_1 v_2 N \log N$$

     or
     $$1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N \ \text{and} \ u_1 \gg \sqrt{v_1 v_2 N}$$

   - *Either $u_2 \gg 1/\log N$ or $u_2 \gg v_2 N$*

   then the probability that both mutations occur on the stem cell line converges to 1. If $u_1 \ll u_2$ then
   $$u_1 \tau(H_1) \to_d X$$

   and if $u_1 \sim A u_2$ for some $A > 0$ then

   $$u_1 \tau(H_1) \to_d X + Z$$

   where Z is an exponentially distributed random variable with mean A which is independent of X.

There are many boundary cases and most of them are not included, where we use the term boundary case to refer to the boundary between two of the conditions. That is, if $v_1 \ll 1/N$ gives one result and $v_1 \gg 1/N$ gives another, we would consider $v_1 \sim A/N$ for some constant A to be a boundary case. If included, the boundary cases would make up the bulk of this thesis. One reason for this is that our variables $\{v_1, v_2, u_1, u_2\}$ span a four dimensional space so that the regions will have many boundaries. Moreover, sometimes three regions intersect in the same place. It does not seem that there would be any special difficulties in computing most of these boundary cases and that they could be done with methods similar to those used in chapter 3.

The first three cases of Theorem 18 are the dd regime. Case 4 is the sd regime and case 5 is the ss regime. In case 1 the condition $v_1 v_2 \ll 1/(N(\log N)^2)$ indicates that with probability tending to 1 the first successful type-1 mutation on a daughter cell will occur after $\log N$ time. The condition $v_1 v_2 N \log N \gg u_1$ indicates that a type-2 mutation will occur on a daughter cell before a type-1 mutation occurs on a stem cell with probability tending to 1. Because the amount of time that can pass between a successful type-1 mutation and a type-2 mutation is bounded by $\log N$ the time it takes for the type-2 mutation to occur is negligible in the limit. This is why the distribution of $\tau(H_1)$ converges to an exponential distribution.

There is a useful picture to keep in mind. We will graph time scaled by $1/\log N$ on the horizontal axis and generation scaled by $1/\log N$ on the vertical axis. A mutation on a cell in generation i at time t will be represented by a circle at $(t/l, i/l)$. We only represent the successful type-1 and type-2 mutations. When a successful type-1 mutation is marked, the following type-2 mutation will be connected to it by a line. Figure 3.1 is an illustration of case 1 of Theorem 18.

The distribution of $\sigma(H_1)$ arises from a balance between the large number of cells in the later generations versus the large number of descendants of cells in the earlier generations as discussed above. The reasoning used to derive equation (3.1) shows that generation i gets mutations at a rate of approximately

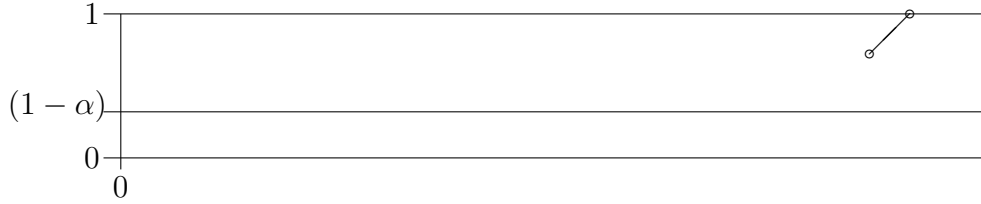$$v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1})}) \approx v_1 v_2 N.$$

Figure 3.1: Case 1 of Theorem 18

Note that the approximate rate is independent of i. This balance causes the distribution of the marks of the successful type-1 mutations to converge to a uniform Poisson process on $[0, \infty) \times ((1 - \alpha)^+, 1)$. The probability that the second mutation occurs in the later generations is just a result of the bulk of the population being concentrated in the later generations.

In case 2 the condition $1/(N(\log N)^2) \ll v_1 v_2$ indicates that a daughter cell will get a successful type-1 mutation before time $\log N$ with probability tending to 1. The condition $v_1 v_2 \ll 1/N$ indicates that the time it takes for a successful type-1 mutation to occur on a daughter cell tends to infinity. The condition $v_1 v_2 \gg u_1^2/N$ indicates that the cancer causing type-1 mutation will occur on a daughter cell with probability tending to 1. As in case 1, $\rho(H_1) \to_p 1$ because most of the cells are in the later generations. Because cells split at rate 1 it takes $O(\log N)$ time units before a significant number of an individuals progeny is realized. In this case the type-2 mutation will occur much faster than $\log N$ time with probability tending to 1. Therefore, an individuals progeny does not play such an important role. For this reason the cancer causing type-1 mutation is approximately equally likely to occur on any cell. Most of the cells are in the later generations so $\sigma(H_1)$ tends to 1. We illustrate this case in Figure 3.2.

In Figure 3.2, a type-2 mutation will occur by time t if a successful type-1 mutation has occurred in the triangle beneath time t. Note that Figure 3.2 illustrates an example in which the first successful type-1 mutation is not the
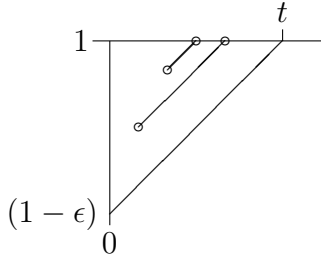
Figure 3.2: Case 2 of Theorem 18 - A magnified image of the top left corner

cancer causing type-1 mutation. Because the marks of the type-1 mutations are converging to a uniform Poisson process in the triangle, the distribution of $\tau(H_1)$ will converge to the Rayleigh distribution.

In case 3 the condition $v_1 v_2 \gg 1/N$ indicates that some cell will receive two mutations before time 1 with probability tending to 1. Any daughter cell is equally likely to get the two mutations and because $u_1 \to 0$ the probability that the stem cell gets the two mutations tends to 0. This causes $\sigma(H_1)$ and $\rho(H_1)$ to tend to 1 in probability since the bulk of the population is concentrated in the later generations. The waiting time for the first individual to get two mutations has a Rayleigh distribution, which gives the result for $\tau(H_1)$. The results hold for this case when $\alpha = 0$.

We now explain the assumptions of case 4 which ensure that the sd regime occurs with probability tending to 1. If stem cells could not mutate and $v_1 v_2 \ll 1/(N(\log N)^2)$ then according to case 1 $(\alpha \wedge 1)v_1 v_2 N \log N \tau(H_1) \to_d X$. The condition $u_1 \gg v_1 v_2 N \log N$ indicates that a type-1 mutation occurs on the stem cell line before a type-2 mutation occurs on a daughter cell when the mutation rates of the daughter cells satisfy $v_1 v_2 \ll 1/N(\log N)^2$. Likewise, if the stem cell could not mutate and $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$ then according to case 2 $\sqrt{v_1 v_2 N} \tau(H_1) \to_d Y$. The condition $u_1 \gg \sqrt{v_1 v_2 N}$ indicates that the stem cell line gets a type-1 mutation before the daughter cells get a type-2 mutation when the mutation rates of the daughter cells satisfy $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$.

The condition $u_2 \gg 1/\log N$ or $u_2 \gg v_2 N$ indicates that the first type-2 mutation occurs on a daughter cell rather than the stem cell line.

In case 4 the time at which the type-1 mutation occurs on the stem cell line is much larger than $\log N$ with probability tending to 1. Therefore, the time it takes for the first type-2 mutation to occur is negligible. This implies that the type-1 mutation that occurs on the stem cell line is the cancer causing type-1 mutation with probability tending to 1 and illustrates why $u_1 \tau(H_1)$ is converging to an exponential distribution. Once a stem cell gets a type-1 mutation the daughter cells inherit the type-1 mutation at an exponential rate. For any $\epsilon > 0$ the probability that the first type-2 mutation will occur when the type-1 mutation has spread to generation i for some $i \in ((\alpha \wedge 1 - \epsilon) \log N, (\alpha \wedge 1 + \epsilon) \log N)$ is tending to 1. This is why $\rho(H_1) \to_p (\alpha \wedge 1)$. Figure 3.3 gives an illustration of this case.
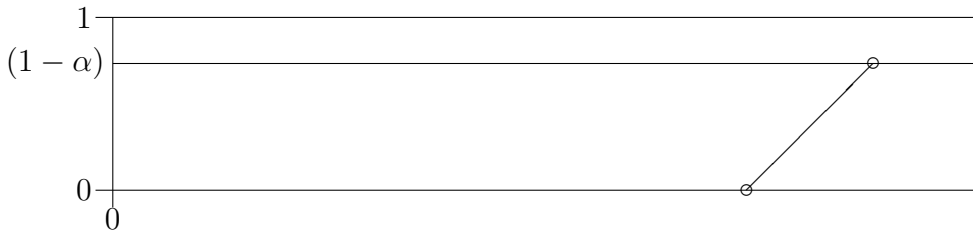

Figure 3.3: Case 4 of Theorem 18 - Stem cell mutations occur on $[0, \infty) \times \{0\}$

The first condition in case 5 is the same as the first condition in case 4. Under this condition the probability that the first successful type-1 mutation occurs on the stem cell line tends to 1. The second condition in case 5 implies that the first type-2 mutation occurs on the stem cell line with probability tending to 1.

The results for $\tau(H_1)$ are similar to the results when waiting for two mutations in the Moran model. In particular, when the mutation rates are slow in the Moran model the time until two mutations converges to the exponential distribution and when the rates are faster the waiting time converges to the Rayleigh distribution. The original results can be found in [18] and [34] and they are also a

special case of the results in [33].

When all of the mutation rates are the same we refer to H_1 as the null model. The following proposition gives the results for the null-model, including results for the boundary cases.

Proposition 19. *Let $\mu = u_1 = u_2 = v_1 = v_2$. Suppose Assumption 1 holds so that there exist constants $\beta, \alpha > 0$ such that $\mu \sim \beta N^{-\alpha}$.*

1. *If $\mu \ll 1/(N \log N)$ then*

$$\mu \tau(H_1) \to_d X.$$

   The probability that the first successful type-1 mutation occurs on the stem cell line converges to 1 and $\rho(H_1)$ converges in probability to 1.

2. *If $\mu \sim A/(N \log N)$ then*

$$(1 + A)\mu \tau(H_1) \to_d X.$$

   Let ξ be a Bernoulli random variable such that $P(\xi = 1) = A/(1 + A)$ and $P(\xi = 0) = 1/(1 + A)$. Let U be a random variable, independent of ξ, with the uniform distribution on $[0, 1]$. Then

$$\sigma(H_1) \to_d U\xi$$

   and

$$\rho(H_1) \to_p 1.$$

3. *If $1/(N \log N) \ll \mu \ll 1/(\sqrt{N} \log N)$ then*

$$(\alpha \wedge 1)\mu^2 N (\log N)\tau(H_1) \to_d X.$$

   The distribution of $\sigma(H_1)$ converges to a uniform distribution on $((1-\alpha)^+, 1]$ and $\rho(H_1)$ converges in probability to 1.

4. *If $\mu \sim A/(\sqrt{N} \log N)$ then*

$$\lim P(\tau(H_1)/\log N \le t) = (1 - e^{-A^2 t^2/2})1_{[0,1/2]}(t) + (1 - e^{-A^2 t/2 + A^2/8})1_{(1/2,\infty)}(t).$$

Let Z be a random variable with density

$$f(x) = \left(\int_{1-x}^{1/2} A^2 e^{-A^2 t^2/2} dt + 2e^{-A^2/8} \right) 1_{[1/2,1]}(x).$$

As N goes to infinity $\sigma(H_1)$ converges in distribution to Z and $\rho(H_1)$ converges in probability to 1.

5. *If $1/(\sqrt{N} \log N) \ll \mu \ll 1/\sqrt{N}$ then*

$$\mu \sqrt{N} \tau(H_1) \to_d Y.$$

Both $\sigma(H_1)$ and $\rho(H_1)$ converge in probability to 1.

6. *If $\mu \sim A/\sqrt{N}$ then for each fixed time $t > 0$ there exist constants c and C such that*

$$\liminf P(\tau(H_1) \le t) \ge c > 0 \ and \ \limsup P(\tau(H_1) \le t) \le C < 1.$$

Both $\sigma(H_1)$ and $\rho(H_1)$ converge in probability to 1.

7. *If $1/\sqrt{N} \ll \mu$ then*

$$\mu \sqrt{N} \tau(H_1) \to_d Y.$$

Both $\sigma(H_1)$ and $\rho(H_1)$ converge in probability to 1.

Cases 1, 3, 5 and 7 of Proposition 19 follow directly from Theorem 18. Cases 2, 4 and 6, the boundary cases, will be done in the last section.

In case 2 of Proposition 19 the cancer causing type-1 mutation may occur on a stem cell or a daughter cell. The event $\xi = 1$ corresponds to the cancer causing type-1 mutation occurring on a daughter cell and the event $\xi = 0$ corresponds to the cancer causing type-1 mutation occurring on the stem cell line.

In case 4 the mutations occur in $O(\log N)$ time units. Figure 3.4 is an illustration for this case.

Notice that the exponents in the limiting distribution for $\tau(H_1)$ in part 4 correspond to the area of a triangle or quadrilateral. This is because the cancer causing type-1 mutation will occur in $O(\log N)$ time units. Let t_1 and t_2 be the
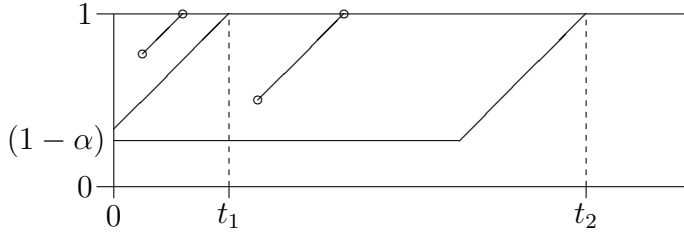
Figure 3.4: Case 4 of Proposition 19

times marked in Figure 3.4. The probability that a type-2 mutation has occurred by time $t_1/\log N$ is the probability that a mark indicating a successful type-1 mutation has occurred in the triangle associated with t_1 in Figure 3.4. Likewise, the probability that a type-2 mutation has occurred by time $t_2/\log N$ is the probability that a mark indicating a successful type-1 mutation has occurred in the quadrilateral associated with t_2 in Figure 3.4.

The main result of case 6 is that when $\mu \sim A/\sqrt{N}$ the time until two mutations is $O(1)$. The results are therefore affected by the discreteness of the model.

In the next section we introduce a new model which will be coupled with H_1. Theorem 18 will be proved with this new model in place of H_1 and the coupling will give the results for H_1. The second section of chapter 3 is devoted to getting results about the dd regime. The third section is on results about the sd and ss regimes. In section 3.4 we prove Theorem 18. The last section is a discussion of the boundary cases in the null model and a proof of Proposition 19.

3.1   A Useful Model

In this section we define a new model, H_2, which will be useful to compare with H_1. In model H_2 there is one stem cell and for each integer i there are 2^{i-1} generation i daughter cells for all times $t \geq 0$. The cells in model H_2 split at each

integral time unit in the same way that the cells in model H_1 split. Just as in model H_1, the stem cells in model H_2 receive type-1 and type-2 mutations at rates u_1 and u_2 respectively and the daughter cells receive type-1 and type-2 mutations at rates v_1 and v_2 respectively. The difference between the models is how the cells accumulate type-1 mutations. In model H_1 all type-1 mutations have the same behavior. A type-1 mutation proposed to occur on a type-1 cell in H_1 is rejected because the cell is already a type-1. In model H_2 the behavior of type-1 mutations differ depending on whether or not the mutation occurred on a stem cell. If a type-1 mutation occurs on a stem cell it has the same behavior as in model H_1. The mutation will eventually be passed to all other cells in the population and any type-1 mutation proposed to occur on a type-1 stem cell or a daughter cell that is the progeny of a type-1 stem cell is rejected. However, all type-1 daughter cells which are type-1 cells as a result of a type-1 mutation occurring on a daughter cell are able to accumulate type-1 mutations. If a type-1 mutation is proposed to occur on such a daughter cell with one type-1 mutation, then the mutation is accepted and the cell now carries two type-1 mutations. Type-1 mutations to type-0 daughter cells result in cells that are allowed to carry any number of type-1 mutations, and when a cell has k type-1 mutations it receives type-2 mutations at rate kv_2. Because the type-1 mutations on daughter cells do not change the rate at which type-1 mutations occur, equation (3.1) is more accurate for model H_2.

We now give an alternate description of model H_2 which will allow us to make a coupling between models H_1 and H_2. Consider the daughter cells as fixed in a tree and consider the mutations as moving to the higher generation daughter cells at each integral time unit in model H_1. Label the daughter cells $D_1, D_2, \ldots D_{N-1}$.

In model H_2 each daughter cell D_i has a counter C_i starting at 0 and is acted on by a sequence of Poisson processes $\{P_n^i\}_{n=1}^\infty$, each having rate v_2, which determine the type-2 mutations. All of the Poisson processes are independent of one another. When a type-1 mutation occurs on a daughter cell D_i it increases the counter C_i by 1. This is considered as a type-1 mutation. If a type-1 mutation increases the counter to n, it is the nth type-1 mutation on the cell. When the counter C_i has reached n, any type-2 mutations that would occur according

to the Poisson processes $P_1^i, P_2^i, \ldots P_n^i$ are accepted as type-2 mutations on cell D_i. Any type-2 mutations that would occur according to the Poisson processes $P_{n+1}^i, P_{n+2}^i, \ldots$ are rejected. If a type-2 mutation occurs on cell D_i as a result of the Poisson process P_n^i, then the nth type-1 mutation according to C_i is considered to be successful. If the first type-2 mutation on a cell is a result of the Poisson process P_n^i, then the nth type-1 mutation according to C_i is the cancer causing type-1 mutation. Rather than the mutations moving up the tree, at each integral time unit the daughter cells in generations $i \geq 2$ will inherit the counter number from their ancestor in the previous generation. The daughter cell in generation 1 will reset its counter to 0 at each integral time unit. However, a type-1 mutation on a stem cell does not have a counter. Once a type-1 mutation has spread from a stem cell to a daughter cell the daughter cell can no longer accumulate type-1 mutations and the model is the same as model H_1.

We couple H_1 and H_2 as follows:

- The Poisson processes that mark the stem cells are the same.

- If a daughter cell has inherited a type-1 mutation from a stem cell then the Poisson processes marking type-2 mutations on the cell are the same in each model.

- The Poisson processes marking type-1 mutations on daughter cells are the same.

- The Poisson processes marking type-2 mutations on daughter cells in model H_1 are the same as the Poisson processes P_1^i in model H_2 so long as the daughter cells did not inherit their type-1 mutations from a stem cell.

There are no analogous Poisson processes in model H_1 for the $N - 1$ sequences of Poisson processes $P_2^i, P_3^i, \ldots$ in model H_2.

Lemma 20. *Let the Poisson processes in models H_1 and H_2 be coupled as described above. Then $P(\tau(H_1) = \tau(H_2))$, $P(\rho(H_1) = \rho(H_2))$ and $P(\sigma(H_1) = \sigma(H_2))$ all converge to 1.*

Proof. A type-2 mutation which occurs in model H_2 but not in H_1 is a result of the rejection of the type-1 mutation in model H_1 that has led to the type-2 mutation in H_2. This type-1 mutation could only be rejected in model H_1 because the cell on which it was supposed to occur was already a type-1 cell. Type-1 mutations on the stem cell line will occur at the same time in both models. If we consider a type-1 mutation that occurs on a daughter cell in model H_2, the probability that it also occurs in model H_1 is the probability that the cell is a type-0. Because the differentiated cells will be removed from the population after $\log N$ time, if we propose a type-1 mutation at a time t on any cell that has not inherited a type-1 mutation from a stem cell, then the probability that the cell has a type-1 mutation is at most $1 - e^{-v_1 \log N}$. Therefore, if a type-1 mutation occurs in model H_2 at time t, with probability at least $e^{-v_1 \log N}$ it will also occur in model H_1. We show that the same will be true of the cancer causing type-1 mutation.

We number the positions of the cells $1, 2, \ldots, N$ and let 1 be the position of the stem cell line. Let $\bar{N} = \{1, 2, \ldots, N\}$ and $L = [0, l] \cup \{\infty\}$. First we note that the Poisson processes marking the daughter cells in model H_2 induce a Poisson process on the space $[0, \infty) \times \bar{N} \times L$. A point (t, i, s) is marked to indicate that a type-1 mutation occurred at time t on the cell at location i and at time $s + t$ the type-1 mutation became successful. If the type-1 mutation is not successful then $s = \infty$. One may note that this is a Poisson process by two applications of the Marking Theorem (see Kingman [20] page 55). Type-1 mutations occur according to a Poisson process on $[0, \infty)$ at rate $v_1(N-1) + u_1$. Each daughter cell has probability $v_1/(v_1(N-1) + u_1)$ of being the cell that receives the type-1 mutation and the stem cell has probability $u_1/(v_1(N-1) + u_1)$ of being the cell that receives the type-1 mutation. By a first application of the Marking Theorem this gives us a Poisson process on $[0, \infty) \times \bar{N}$. The probability that a type-1 mutation is successful can be determined from the associated point (t, i) which tells at what time and on what cell the type-1 mutation occurred. Each one of these points has an associated value s that indicates when, and if, the type-1 mutation becomes successful. This gives the Poisson process on $[0, \infty) \times \bar{N} \times L$.

Let Z be the random variable which indicates the value in $[0, \infty) \times \bar{N} \times L$

that corresponds to the time of the cancer causing type-1 mutation, the cell on which it occurred, and the time of the first type-2 mutation. If we condition on the event $Z = (t_0, i_0, s_0)$ for some i_0 in generation j which is not the stem cell line, then there can be no marks in subset

$$\{(t, i, s) : s < t_0 + s_0 - t\} \cup \{(t, i, s) : (\lceil t \rceil - i, \{1\}, s + t - \lceil t \rceil)\}$$

of $[0, \infty) \times \bar{N} \times L$. The marks that occur outside of this subset occur independently of the marks that occur within. Conditioning does not change the probability that a mark outside of this set has occurred by time t_0. This only reduces the rate at which type-1 mutations occur before time t_0. Therefore,

$$P(\tau(H_1) \neq \tau(H_2) | Z = (t_0, i_0, s_0)) \leq 1 - e^{-v_1 \log N}.$$

Let P_Z be the probability measure on $[0, \infty) \times \bar{N} \times L$ induced by Z. Then

$$P(\tau(H_1) \neq \tau(H_2)) = \int_{[0,\infty) \times \bar{N} \times L} P(\tau(H_1) \neq \tau(H_2) | Z = x) P_Z(dx)$$

$$\leq \int_{[0,\infty) \times \bar{N} \times L} (1 - e^{-v_1 \log N}) P_Z(dx)$$

$$= 1 - e^{-v_1 \log N}.$$

This shows that $P(\tau(H_1) \neq \tau(H_2)) \to 0$ if $v_1 \ll 1/\log N$. It follows from Assumption 1 that $v_2 \ll 1/\log N$ and combining this with Assumption 2 we see that $v_1 \ll 1/\log N$ as well.

On the event $\tau(H_1) = \tau(H_2)$ we have $\rho(H_1) = \rho(H_2)$ and $\sigma(H_1) = \sigma(H_2)$ with probability 1. The only way these equalities can fail is if two type-2 mutations occur simultaneously in model H_2, an event whose probability is 0. Therefore, $P(\rho(H_1) = \rho(H_2))$ and $P(\sigma(H_1) = \sigma(H_2))$ both converge to 1 as well. □

The rest of the work in proving Theorem 18 is in proving Theorem 18 with H_2 in place of H_1. Once this is done Theorem 18 follows from Lemma 20.

3.2   The dd regime

To understand the behavior in the dd regime, we consider a new model which is the same as H_2 except that mutations only occur on daughter cells. That

is, there are no Poisson processes that mark mutations on the stem cells. This new model will be called model M_1. The purpose of this section is to prove Proposition 21.

Proposition 21. *1. If $v_1 v_2 \ll 1/(N(\log N)^2)$ then*

$$(\alpha \wedge 1) v_1 v_2 N (\log N) \tau(M_1) \to_d X.$$

The distribution of $\sigma(M_1)$ converges to a uniform distribution on $((1-\alpha)^+, 1]$ and $\rho(M_1)$ converges in probability to 1.

2. If $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$ then

$$\sqrt{v_1 v_2 N} \tau(M_1) \to_d Y.$$

Both $\sigma(M_1)$ and $\rho(M_1)$ converge in probability to 1.

Lemma 22. *For any positive integer $k < l$ we have*

$$P(\rho(M_1) \geq (l-k)/l) > 1 - 1/2^k.$$

Proof. Let Z be the number of generations between the cancer causing type-1 mutation and the first type-2 mutation. Then $Z \in \{0, 1, 2, \ldots, l\}$. Because there are only l generations, if the second mutation occurs $l - k$ generations or more after the first then it must be in the last k generations. So

$$P(\rho(M_1) \geq (l-k)/l | Z \in \{l-k, l-k+1, \ldots, l\}) = 1.$$

If we condition on the event that $Z = j$ for some $j \leq l - k - 1$, then the probability that the cancer causing type-1 mutation occurs on any cell in generations $1, 2, \ldots, l - j$ is equally likely. This is because the Poisson processes marking the mutations on the descendants of the cells j generations after any generation i are independent and identically distributed. The last k of the $l - j$ generations always make up at least a fraction of $1 - 1/2^k$ cells, so we have $P(\rho(M_1) \geq (l-k)/l | Z \in \{0, 1, 2, \ldots, l-k-1\}) > 1 - 1/2^k$ where we get a strict inequality because we do not count the stem cell line. The result follows. $\square$

It is important to notice that Lemma 22 holds for any N and we do not require $N \to \infty$. Also, the rates at which v_1 and v_2 tend to 0 are irrelevant.

Corollary 23. *As N goes to infinity, $\rho(M_1)$ will converge to 1 in probability.*

Lemma 24. *Let $(\beta_1, \beta_2] \subset (0, 1]$. Let C be a positive constant and let $C' \in \{1, 2\}$. Then*

$$\sum_{i \in \mathbb{N} \cap (l\beta_1, l\beta_2]} v_1 2^{i-1}(1 - e^{-Cv_2(2^{l-i+1}-C')}) \sim C(\beta_2 - \beta_1 \vee (1-\alpha))^+ v_1 v_2 N \log N.$$

Proof. We will first define some notation for this proof for the sake of readability. Let $I \subset \mathbb{R}$. We define

$$I^* := I \cap (l\beta_1, l\beta_2] \cap \mathbb{N}.$$

First we can do the case when $\alpha \geq 1$. Using the upper bound

$$1 - e^{-Cv_2(2^{l-i+1}-C')} \leq Cv_2 2^{l-i+1}$$

we have

$$\frac{\sum_{i \in (l\beta_1, l\beta_2]^*} v_1 2^{i-1}(1 - e^{-Cv_2(2^{l-i+1}-C')})}{v_1 v_2 2^l l} \leq C(\beta_2 - \beta_1).$$

From the second order Taylor expansion we get a lower bound of

$$1 - e^{-C(2^{l-i+1}-C')} \geq Cv_2(2^{l-i+1} - C') - \frac{1}{2}C^2 v_2^2(2^{l-i+1} - C')^2.$$

We will break this sum into 5 parts,

$$2^{i-1}(1-e^{-Cv_2(2^{l-i+1}-C')})$$
$$\geq Cv_2 2^l - CC' v_2 2^i - C^2 v_2^2 2^{2l-i} + C^2 C' v_2^2 2^l - C^2(C')^2 v_2^2 2^{i-2}.$$

We get the following computations for each of the five individual sums:

$$\sum_{i \in (l\beta_1, l\beta_2]^*} Cv_2 2^l / (v_2 2^l l) \to C(\beta_2 - \beta_1).$$

$$\sum_{i \in (l\beta_1, l\beta_2]^*} CC'v_2 2^i / (v_2 2^l l) \le CC' 2^{l+1} / (2^l l) \to 0.$$

$$\sum_{i \in (l\beta_1, l\beta_2]^*} C^2 (C')^2 v_2^2 2^{i-2} / (v_2 2^l l) \le C^2 C'^2 v_2 / l \to 0.$$

$$\sum_{i \in (l\beta_1, l\beta_2]^*} C^2 C' v_2^2 2^l / (v_2 2^l l) \le C^2 C' v_2 \to 0.$$

$$\sum_{i \in (l\beta_1, l\beta_2]^*} C^2 v_2^2 2^{2l-i} / (v_2 2^l l) = C^2 v_2 2^l \left(\sum_{i=\lceil l\beta_1 \rceil}^{\beta_2} 2^{-i} \right) / l \le C^2 v_2 2^{l(\beta_2 - \beta_1)} \to 0$$

so long as $v_2 \ll 1/2^{l(\beta_2 - \beta_1)} = N^{-(\beta_2 - \beta_1)}$ which will hold since this is the case $\alpha \ge 1$.

So we have

$$\lim_{N \to \infty} \left(\frac{\sum_{i \in (l\beta_1, l\beta_2]^*} v_1 2^{i-1} \left(1 - e^{-Cv_2(2^{l-i+1} - C')}\right)}{v_1 v_2 2^l l} \right) = C(\beta_2 - \beta_1)$$

which finishes the case for $\alpha \ge 1$.

Now let $0 < \alpha < 1$ and let $\epsilon > 0$ be small enough so that $0 < 1 - \alpha - \epsilon < 1 - \alpha + \epsilon < 1$. We now break the sum into three pieces,

$$\frac{\sum_{i \in [1, l(1-\alpha-\epsilon))^* \cup [l(1-\alpha-\epsilon), l(1-\alpha+\epsilon)]^* \cup (l(1-\alpha+\epsilon), l]^*} 2^{i-1} \left(1 - e^{-Cv_2(2^{l-i+1} - C')}\right)}{v_2 2^l l}.$$

We can consider each of these three sums individually.

As for the middle sum, we only need the bound

$$0 \le \frac{\sum_{i \in [l(1-\alpha-\epsilon), l(1-\alpha+\epsilon)]^*} 2^{i-1} \left(1 - e^{-Cv_2(2^{l-i+1} - C')}\right)}{v_2 2^l l} \le 2C\epsilon$$

which follows by the upper bound $1 - e^{-Cv_2(2^{l-i+1} - C')} \le Cv_2 2^{l-i+1}$.

One can apply similar computations as in the case when $\alpha = 1$ to obtain the following:

$$\frac{\sum_{i \in (l(1-\alpha+\epsilon), l]^*} 2^{i-1} \left(1 - e^{-Cv_2(2^{l-i+1} - C')}\right)}{v_2 2^l l} \to C(\beta_2 - \beta_1 \vee (1 - \alpha + \epsilon))^+.$$

For the first sum, note that $1 - e^{-Cv_2(2^{l-i+1}-C')} \leq 1$. This gives the bound

$$0 \leq \sum_{i \in [1, l(1-\alpha-\epsilon))^*} \frac{2^{i-1}\left(1 - e^{-Cv_2(2^{l-i+1}-C')}\right)}{v_2 2^l l}$$

$$\leq \sum_{i \in [1, l(1-\alpha-\epsilon))^*} \frac{2^{i-1}}{v_2 2^l l}$$

$$\leq \frac{2^{l(1-\alpha-\epsilon)}}{v_2 2^l l} \to 0.$$

The convergence is a result of the definition of α. In particular, $v_2 \gg N^{-\alpha-\epsilon}(\log N)^{-1}$.

Combining the three sums yields

$$C(\beta_2 - \beta_1 \vee (1 - \alpha + \epsilon))^+ \leq \liminf \frac{\sum_{i \in (l\beta_1, l\beta_2]^*} v_1 2^{i-1}\left(1 - e^{-Cv_2(2^{l-i+1}-C')}\right)}{lv_1 v_2 2^l}$$

and

$$\limsup \frac{\sum_{i \in (l\beta_1, l\beta_2]^*} v_1 2^{i-1}\left(1 - e^{-Cv_2(2^{l-i+1}-C')}\right)}{lv_1 v_2 2^l} \leq C(\beta_2 - \beta_1 \vee (1 - \alpha + \epsilon))^+ + 2C\epsilon.$$

Letting ϵ approach 0 gives the result. $\square$

Corollary 25. *Let T be the time at which the first successful type-1 mutation occurs. Then $(\alpha \wedge 1)v_1 v_2 N(\log N)T \to_d X$.*

Proof. For $1 \leq i \leq l$ there are 2^{i-1} cells in generation i. Each of these cells is getting type-1 mutations at rate v_1. The cells in generation i have $2^{l-i+1} - 2$ descendants. If the cell splits as soon as it becomes a type-1, the probability that none of its descendants get a type-2 mutation is $e^{-v_2(2^{l-i+1}-2)}$. On the other hand, after a cell gets a type-1 mutation it could live for at most 1 time unit until it splits. If this is the case, then the probability that neither the cell that receives the type-1 mutation nor any of its descendants get a type-2 mutation is $e^{-v_2(2^{l-i+1}-1)}$. If we let $R(t)$ be the rate at which the successful type-1 mutations occur at time t, then for any time t we have

$$1 = \lim \frac{\sum_{i=1}^l v_1 2^{i-1}\left(1 - e^{-v_2(2^{l-i+1}-2)}\right)}{(\alpha \wedge 1)v_1 v_2 N \log N} \leq \liminf \frac{R(t)}{(\alpha \wedge 1)v_1 v_2 N \log N}$$

$$\leq \limsup \frac{R(t)}{(\alpha \wedge 1)v_1 v_2 N \log N} \leq \lim \frac{\sum_{i=1}^l v_1 2^{i-1}\left(1 - e^{-v_2(2^{l-i+1}-1)}\right)}{(\alpha \wedge 1)v_1 v_2 N \log N} = 1,$$

where the limits are results of Lemma 24.

The successful type-1 mutations occur according to a time inhomogeneous Poisson process with an intensity measure ν where $\nu([0,t]) = \int_0^t R(s)ds$. We have shown that ν satisfies

$$t\sum_{i=1}^{l} v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-2)}) \leq \nu([0,t]) \leq t\sum_{i=1}^{l} v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-1)})$$

for all $t \geq 0$ and all N. For any $t \geq 0$ we have

$$P\left(T \leq \frac{t}{(\alpha \wedge 1)v_1 v_2 N(\log N)}\right) = 1 - e^{-\nu([0,t])/((\alpha \wedge 1)v_1 v_2 N \log N)} \to 1 - e^{-t}$$

where the limiting results follow by Lemma 24. Therefore, $(\alpha \wedge 1)v_1 v_2 N(\log N)T$ is converging in distribution to an exponentially distributed random variable with parameter 1. $\square$

The next lemma states that when $v_1 v_2 \ll 1/(N(\log N)^2)$ the probability that the first successful type-1 mutation is the cancer causing type-1 mutation tends to 1.

Lemma 26. *Let T be the time at which the first successful type-1 mutation occurs in model M_1. If $v_1 v_2 \ll 1/(N(\log N)^2)$ then $P(T = \tau'(M_1)) \to 1$.*

Proof. Let $Z = \tau(M_1) - T$ be the time it takes to get the first type-2 mutation after the first successful type-1 mutation has appeared and let $\hat{T}$ be the time it takes to get the second successful type-1 mutation after the first.

By Corollary 25,

$$(\alpha \wedge 1)v_1 v_2 N(\log N)T \to_d X \text{ and } (\alpha \wedge 1)v_1 v_2 N(\log N)\hat{T} \to_d X.$$

Then because a type-2 mutation must occur within $\log N$ time after a successful type-1 mutation on a daughter cell we have

$$P(\hat{T} < Z) \leq P(\hat{T} < \log N) = P((\alpha \wedge 1)v_1 v_2 N(\log N)\hat{T} < (\alpha \wedge 1)v_1 v_2 N(\log N)^2) \to 0.$$

Moreover, $P(\hat{T} \geq Z) \leq P(T = \tau'(M_1))$ so $P(T = \tau'(M_1)) \to 1$. $\square$

Lemma 27. *If $v_1 v_2 \ll 1/(N(\log N)^2)$ then $(\alpha \wedge 1)v_1 v_2 N(\log N)\tau(M_1) \to_d X$.*

Proof. From Lemma 26 we know that the probability that the first successful type-1 mutation is the cancer causing mutation is converging to 1. Combining this with Corollary 25, $(\alpha \wedge 1)v_1 v_2 N (\log N) \tau'(M_1) \to_d X$.

Due to apoptosis $\tau(M_1) - \tau'(M_1)$ is bounded above by $\log N$ so it follows that $(\alpha \wedge 1)v_1 v_2 N (\log N)(\tau(M_1) - \tau'(M_1)) \to_p 0$. Then

$$(\alpha \wedge 1)v_1 v_2 N (\log N)\tau(M_1) = (\alpha \wedge 1)v_1 v_2 N (\log N)(\tau'(M_1) + (\tau(M_1) - \tau'(M_1)))$$

$$\to_d X.$$

$\square$

Lemma 28. *If $v_1 v_2 \ll 1/(N(\log N)^2)$ then the distribution of $\sigma(M_1)$ converges to the uniform distribution on $((1 - \alpha)^+, 1]$.*

Proof. By Lemma 26 the first successful type-1 mutation will be the cancer causing type-1 mutation with probability tending to 1. Therefore, to find the limiting results on $\sigma(M_1)$ it is enough to find the depth at which the first successful type-1 mutation occurs as N tends to infinity.

Each generation i with $1 \le i \le l$ is getting successful type-1 mutations independently at a rate bounded between $v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-2)})$ and $v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-1)})$ for any time t. Therefore, for a fixed N and i, the probability that the first successful type-1 mutation occurs on generation i is between

$$\frac{v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-2)})}{\sum_{j=1}^{l} v_1 2^{j-1}(1 - e^{-v_2(2^{l-j+1}-1)})}$$

and

$$\frac{v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-1)})}{\sum_{j=1}^{l} v_1 2^{j-1}(1 - e^{-v_2(2^{l-j+1}-2)})}.$$

Let $\beta \in [0, 1]$. Using the notation and result from Lemma 24,

$$\limsup P(\sigma(M_1) \le \beta) \le \limsup \frac{\sum_{i \in (0, l\beta]^*} v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-1)})}{\sum_{j \in (0, l]^*} v_1 2^{j-1}(1 - e^{-v_2(2^{l-j+1}-2)})}$$

$$= \frac{(\beta - (1 - \alpha)^+)^+}{\alpha \wedge 1}$$

and

$$\liminf P(\sigma(M_1) \leq \beta) \geq \liminf \frac{\sum_{i \in (0, l\beta]^*} v_1 2^{i-1}(1 - e^{-v_2(2^{l-i+1}-2)})}{\sum_{j \in (0, l]^*} v_1 2^{j-1}(1 - e^{-v_2(2^{l-j+1}-1)})}$$
$$= \frac{(\beta - (1 - \alpha)^+)^+}{\alpha \wedge 1}.$$

$\square$

Combining the results of Corollary 23 and Lemmas 27 and 28 we have part 1 of Proposition 21. For the next two proofs we note that Corollary 23 already gives us that $\rho(M_1)$ converges to 1 in probability.

Proof of part 2 of Proposition 21. For the slower mutation rates it was enough to notice that a cell in generation i has $2^{l-i+1} - 2$ descendants. Under these conditions the mutation rates are fast enough that we will need to consider how many descendants a cell in generation i has at a time before its progeny undergoes apoptosis. For each $k \in \mathbb{N} \cup \{0\}$, let $C_{i,k}$ be the collection of cells in generation i during time $[k, k+1)$. If $t \geq l - i + k$ the number of descendants of each one of the cells in $C_{i,k}$ will be $2^{i-1}(2^{l-i+1} - 2)$ and their progeny will no longer be in the population. For $k < t < l - i + k$ the number of descendants of each cell in $C_{i,k}$ will be between 2^{t-1-k} and 2^{t+1-k}. This will allow us to give upper and lower bounds on the number of cells in or descended from cells in generation i by time t. If we consider a time $t < l - i$ then the descendants of the cells in $C_{i,0}$ will not yet have undergone apoptosis. Therefore, at time $t < l - i$ the number of cells that have been in generation i and their descendants is between

$$\sum_{j=0}^{\lfloor t \rfloor} 2^{t-1-j} \geq 2^t - 1$$

and

$$\sum_{j=0}^{\lfloor t \rfloor} 2^{t+1-j} \leq 2^{t+2} - 1.$$

If $t \geq l - i$ then some of the cells that have descended from generation i cells will have undergone apoptosis. The total number of cells that have been in or

descended from generation i cells at time t, including those that have undergone apoptosis, will be between

$$\sum_{j=0}^{l-i} 2^{l-i-j-1} + (t-l+i)(2^{l-i+1} - 2) = 2^{l-i} - 1 + (t-l+i)(2^{l-i+1} - 2)$$

and

$$\sum_{j=0}^{l-i} 2^{l-i-j+1} + (t-l+i)(2^{l-i+1} - 2) = 2^{l-i+2} - 1 + (t-l+i)(2^{l-i+1} - 2).$$

Recall that there are always 2^{i-1} cells in generation i which are acquiring type-1 mutations at rate v_1. We can once again multiply the rate of type-1 mutations on generation i by the bounds on the probability that such a mutation is successful to find bounds on the rate of successful type-1 mutations in generation i. We find that successful type-1 mutations occur on generation i according to a Poisson process that has intensity measure between

$$2^{i-1}v_1(1 - e^{-v_2(2^t-1)}) \text{ and } 2^{i-1}v_1(1 - e^{-v_2(2^{t+2}-1)})$$

if $t < l - i$ and

$$2^{i-1}v_1(1 - e^{-v_2(2^{l-i}-1+(t-l+i)(2^{l-i+1}-2))}) \text{ and } 2^{i-1}v_1(1 - e^{-v_2(2^{l-i+2}-1+(t-l+i)(2^{l-i+1}-2))})$$

if $t \geq l - i$.

We now use the bounds on the rates of successful type-1 mutations in each generation i to find the limiting distribution of $\tau(M_1)$. By the hypothesis $1/(N(\log N)^2) \ll v_1 v_2$, for N large enough we will have $t < \sqrt{v_1 v_2 N} \log N$ for any real number t. Let $\theta = t/\sqrt{v_1 v_2 N}$ and let N be large enough that $\theta < l$. Then

$$P(\tau(M_1) \leq \theta) = 1 - e^{-f(N,t)}$$

where by summing over the generations and using the fact that $1 - e^{-x} \leq x$ we

obtain

$$f(N,t) \leq \sum_{0 \leq i < l-\theta} 2^{i-1} v_1 (1 - e^{-v_2(2^{\theta+2}-1)})$$

$$+ \sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 (1 - e^{-v_2(2^{l-i+2}-1+(\theta-l+i)(2^{l-i+1}-2))})$$

$$\leq \sum_{0 \leq i < l-\theta} 2^{i-1} (2^{\theta+2} - 1) v_1 v_2$$

$$+ \sum_{l-\theta \leq i \leq l} 2^{i-1} \left(2^{l-i+2} - 1 + (\theta - l + i)\left(2^{l-i+1} - 2\right)\right) v_1 v_2.$$

As for the first sum,

$$\sum_{0 \leq i < l-\theta} 2^{i-1} (2^{\theta+2} - 1) v_1 v_2 \leq \frac{1}{2} (2^{\theta+2} - 1)(2^{l-\theta+1} - 1) v_1 v_2$$

$$\leq 2^{l+2} v_1 v_2 \to 0.$$

As for the second sum, we first compute

$$\sum_{l-\theta \leq i \leq l} 2^{i-1} (2^{l-i+2} - 1) v_1 v_2 \leq 2^{l+2} v_1 v_2 \theta \to 0.$$

Lastly,

$$\sum_{l-\theta \leq i \leq l} 2^{i-1} (\theta - l + i)(2^{l-i+1} - 2) v_1 v_2 \leq 2^l v_1 v_2 \sum_{l-\theta \leq i \leq l} (\theta - l + i)$$

$$\leq \frac{2^l v_1 v_2}{2} (\theta + 1)^2$$

$$\to \frac{t^2}{2}.$$

Therefore, $\limsup P(\sqrt{v_1 v_2 N} \tau(M_1) \leq t) \leq 1 - e^{-t^2/2}$.

As for the lower bound, we have

$$f(N,t) \geq \sum_{0 \leq i < l-\theta} 2^{i-1} v_1 (1 - e^{-v_2(2^\theta - 1)})$$

$$+ \sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 (1 - e^{-v_2(2^{l-i}-1+(\theta-l+i)(2^{l-i+1}-2))})$$

$$\geq \sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 (1 - e^{-v_2(\theta-l+i)(2^{l-i+1}-2)}).$$

Using the bound $1 - e^{-x} \geq x - x^2/2$ we have

$$\sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 (1 - e^{-v_2(\theta - l + i)(2^{l-i+1} - 2)})$$

$$\geq \sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 (v_2(\theta - l + i)(2^{l-i+1} - 2) - v_2^2(\theta - l + i)^2(2^{l-i+1} - 2)^2/2).$$

First consider

$$\sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 v_2^2 (\theta - l + i)^2 \frac{(2^{l-i+1} - 2)^2}{2}.$$

This sum is bounded between 0 and $\sum_{l-\theta \leq i \leq l} v_2 t^2 2^{l-i}$. Let $0 < \epsilon < \alpha$. For N large enough we have $t < \sqrt{v_1 v_2 N} l(\alpha - \epsilon)$ which is equivalent to $l(1 - \alpha - \epsilon) < l - \theta$. So for N large enough we have

$$\sum_{l-\theta \leq i \leq l} v_2 t^2 2^{l-i} \leq \sum_{l(1-\alpha+\epsilon) \leq i \leq l} v_2 t^2 2^{l-i} \leq l v_2 N^{\alpha - \epsilon} \to 0.$$

This leaves us to show

$$\liminf \sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 v_2 (\theta - l + i)(2^{l-i+1} - 2) \geq \frac{t^2}{2}.$$

Let $j \in \mathbb{N}$ and $t > 0$. For large enough values of N we will have $j < \theta < \log N$. Notice that if $i \leq l - j$ then $2^{l-i+1} - 2 \geq (1 - 2^{-j})2^{l-i+1}$, so

$$\sum_{l-\theta \leq i \leq l} 2^{i-1} v_1 v_2 (\theta - l + i) (2^{l-i+1} - 2)$$

$$\geq \sum_{l-\theta \leq i \leq l-j} 2^{i-1} v_1 v_2 (\theta - l + i) (1 - 2^{-j})2^{l-i+1}.$$

Because j is fixed we have

$$\sum_{l-j \leq i \leq l} 2^{i-1} v_1 v_2 (\theta - l + i) (1 - 2^{-j})2^{l-i+1} \to 0$$

since each of the summands converges to 0. Therefore, we can add this sum without changing the limit. This gets us a lower bound of

$$\liminf \sum_{l-\theta \leq i \leq l} 2^l v_1 v_2 (\theta - l + i) (1 - 2^{-j}) \geq \frac{t^2}{2}(1 - 2^{-j}).$$

We chose j to be any natural number, so $\liminf P(\sqrt{v_1 v_2 N} \tau(M_1) \leq t) \geq 1 - e^{-t^2/2}$.

The above two bounds establish that $P(\sqrt{v_1 v_2 N}\tau(M_1) \le t) \to 1 - e^{-t^2/2}$ for any $t \ge 0$. This leaves us to show that $\sigma(M_1)$ converges in probability to 1. First note that for any $\epsilon > 0$ we have

$$P(\tau(M_1) \le \epsilon \log N) = P(\sqrt{Nv_1 v_2}\tau(M_1) \le \sqrt{Nv_1 v_2}\epsilon \log N) \to 1$$

which follows because the distribution of $\sqrt{Nv_1 v_2}\tau(M_1)$ is converging to the Rayleigh distribution and $\sqrt{Nv_1 v_2}\epsilon \log N$ is converging to infinity. Let $\delta > 0$. By Corollary 23 we know that $\rho(M_1)$ converges in probability to 1 so that as N goes to infinity, $P(\rho(M_1) > 1-\delta) \to 1$. If $\sigma(M_1) < 1-2\delta$ and $\rho(M_1) > 1-\delta$ then $\tau(M_1) > \delta \log N$. Because $P(\tau(M_1) > \delta \log N) \to 0$ we must also have $P(\sigma(M_1) < 1 - 2\delta) \to 0$ where $\delta > 0$ was arbitrary. Then $P(1 - \sigma(M_1) > 2\delta) \to 0$ for any $\delta > 0$ so $\sigma(M_1) \to_p 1$. $\qquad\qquad\square$

3.3   The sd and ss regimes

In this section we need two different models. The first one is the same as model H_2 except that only stem cells receive type-1 mutations and only daughter cells receive type-2 mutations. The second is the same as H_2 except that only stem cells receive mutations. These will be referred to as models M_2 and M_3 respectively.

Proposition 29.   *1. If $u_1 \ll 1/\log N$ and $u_1 \ll Nv_2$ then $u_1\tau(M_2) \to_d X$ and $\rho(M_2) \to_p (\alpha \wedge 1)$.*

2. If $u_1 \ll u_2$ then $u_1\tau(M_3) \to_d X$.

3. Let $A > 0$ and Z be an exponentially distributed random variable mean A which is independent of X. If $u_1 \sim Au_2$ then $u_1\tau(M_3) \to_d X + Z$.

The goal of this section is to prove Proposition 29. It will be shown later that the conditions used in Proposition 29 for the sd regime are the only relevant conditions.

Lemma 30. *For time $t \leq \log N$ after a stem cell receives a type-1 mutation we have*

$$e^{-2^{t+2}v_2} \leq P(\tau(M_2) - \tau'(M_2) > t) \leq e^{-(2^{t-2}-2)v_2}.$$

Proof. Let $Z = \tau(M_2) - \tau'(M_2)$. First we establish the upper bound. After the stem cell line gets the first mutation it takes at most one time unit until the mutation is passed along to the first generation daughter cell. Assuming it does take one time unit until the first generation daughter cell inherits the mutation we can get an upper bound on $P(Z > t)$. Let time $t = 0$ denote the time at which the stem cell line receives the type-1 mutation. There are no mutations being acquired by the daughter cells for time $t \in [0,1)$. For time $t \in [1,2)$ the generation 1 daughter cell is the only type-1 daughter cell. So for $t \in [1,2)$ we have $P(Z > t) = e^{-(t-1)v_2}$. For time $t \in [2,3)$ the first two generations have the mutation which is a total of 3 cells. Therefore, for $t \in [2,3)$ we have $P(Z > t) = e^{-(3(t-2)v_2+v_2)}$ where the v_2 is added because of the probability of having a mutation before time 2. Extending this inductively gives us

$$P(Z > t) \leq e^{-[(2^{\lfloor t \rfloor}-1)(t-\lfloor t \rfloor)+\sum_{i=2}^{\lfloor t \rfloor}(2^{i-1}-1)]v_2} \leq e^{(-2^{t-2}-1)v_2}$$

for any $t \leq \log N$.

For the lower bound we use the same reasoning as above except that we assume it takes 0 time for the generation 1 daughter cell to become a type-1 after the stem cell line is type-1. This gets us

$$P(Z > t) \geq e^{-[(2^{\lceil t \rceil}-1)(t-\lfloor t \rfloor)+\sum_{i=1}^{\lfloor t \rfloor}(2^i-1)]v_2} \geq e^{-2^{t+2}v_2}.$$

$\square$

Lemma 31. *The location of the second mutation satisfies $\rho(M_2) \to_p \alpha \wedge 1$.*

Proof. Let $Z = \tau(M_2) - \tau'(M_2)$. By Lemma 30 we have $P(Z > \log N) \geq e^{-4Nv_2}$. If $\alpha > 1$ then $P(Z > \log N) \to 1$ and the mutation will spread throughout the entire crypt. If this is the case then any cell is equally likely to have the second mutation. Therefore $P(\rho(M_2) \leq \beta) \leq (2^{\beta l} - 1)/(2^l - 1)$ for any $\beta \in [0,1)$ so $\rho(M_2) \to_p 1$.

Now suppose $\alpha \leq 1$. Let $\epsilon > 0$ so that $\alpha - \epsilon > 0$. Then by Lemma 30

$$P(Z > l(\alpha - \epsilon)) \geq e^{-2^{l(\alpha-\epsilon)+2}v_2}.$$

Because $4N^{\alpha-\epsilon}v_2 \to 0$ we get the convergence $P(Z > l(\alpha - \epsilon)) \to 1$. By time $l(\alpha-\epsilon)$ the mutation will have spread to the first $\lfloor l(\alpha-\epsilon) \rfloor$ generations so that for times after $l(\alpha - \epsilon)$ we know that at least $2^{\lfloor l(\alpha-\epsilon) \rfloor}$ cells have the type-1 mutation. Therefore,

$$P(\{\rho(M_2) \leq \beta\} \cap \{Z > l(\alpha - \epsilon)\}) \leq (2^{\beta l} - 1)/(2^{(\alpha-\epsilon)l-1} - 1).$$

Thus, for any $\beta < \alpha - \epsilon$,

$$P(\rho(M_2) \leq \beta) < \frac{2^{\beta l} - 1}{2^{(\alpha-\epsilon)l-1} - 1} + P(X_2 \leq l(\alpha - \epsilon)) \to 0$$

Hence $P(\rho(M_2) \geq \alpha - \epsilon) \to 1$. Because ϵ may be arbitrarily small we have finished the case when $\alpha = 1$.

Suppose $\alpha < 1$ and let $\epsilon > 0$ so that $\alpha + \epsilon \leq 1$. Then by Lemma 30

$$P(Z > l(\alpha + \epsilon)) \leq e^{-(2^{l(\alpha+\epsilon)-2}-1)v_2}.$$

Because $N^{\alpha+\epsilon}v_2/4 \to \infty$, we have $P(Z > l(\alpha + \epsilon)) \to 0$. By time $l(\alpha + \epsilon)$ the mutation has only spread to the first $l(\alpha+\epsilon)$ generations, so $P(\rho(M_2) > \alpha+\epsilon) \to 0$ where ϵ is arbitrarily small. $\qquad\square$

Lemma 32. *If $u_1 \ll 1/\log N$ and $u_1 \ll Nv_2$ then $u_1\tau(M_2) \to_d X$.*

Proof. Since the stem cell line is getting mutations according to a Poisson process at rate u_1 we have that $u_1\tau'(M_2)$ is an exponentially distributed random variable with mean 1. This leaves us to show $u_1(\tau(M_2) - \tau'(M_2)) \to_p 0$.

Suppose we consider a new model M_2' which is the same as model M_2 except that the type-2 mutations can only occur on daughter cells $\log N$ time after the stem cell line has a type-1 mutation. We can couple models M_2 and M_2' so that the same Poisson processes are marking the mutations on the cells in each model but that any proposed type-2 mutation is rejected in model M_2' until $\log N$ time after the stem cell line is type-1. Under the coupling $\tau'(M_2) = \tau'(M_2')$. Also, if we

let $Z = \tau(M_2') - \tau'(M_2')$ then $Z \geq \tau(M_2) - \tau'(M_2)$. Therefore it is enough to show that $u_1 Z \to_p 0$.

If we wait $\log N$ time after the stem cell line receives a type-1 mutation then all of the daughter cells will be type-1. Thus for any fixed N we have

$$P(Z > t) = 1_{[0, \log N]}(t) + e^{-v_2(N-1)(t - \log N)} 1_{(\log N, \infty]}(t).$$

Let $\epsilon > 0$. Then

$$P(u_1 Z > \epsilon) = 1_{[0, \log N]}\left(\frac{\epsilon}{u_1}\right) + e^{-v_2(N-1)(\epsilon/u_1 - \log N)} 1_{(\log N, \infty]}\left(\frac{\epsilon}{u_1}\right).$$

By our assumptions, $u_1 \log N \to 0$ so for N large enough this becomes

$$P(u_1 Z > \epsilon) = e^{-v_2(N-1)(\epsilon/u_1 - \log N)}.$$

Also by our assumptions, $-v_2(N-1)(\epsilon/u_1 - \log N) \sim -v_2 N \epsilon/u_1 \to -\infty$, so

$$P(u_1 Z > \epsilon) \to 0.$$

$\square$

Proof of Proposition 29. Combining Lemmas 31 and 32 we get part 1 of Proposition 29.

Notice that $u_1 \tau(M_3)$ has the exponential distribution with mean 1. To prove part 2 of Proposition 29 we need to show that $u_1(\tau(M_3) - \tau'(M_3)) \to_p 0$. Let $\epsilon > 0$. Then

$$P(u_1(\tau(M_3) - \tau'(M_3)) > \epsilon) = P((\tau(M_3) - \tau'(M_3)) > \epsilon/u_1) = e^{-\epsilon u_2/u_1}.$$

Since $u_2/u_1 \to \infty$ we have $P(u_1(\tau(M_3) - \tau'(M_3)) > \epsilon) \to 0$.

Lastly we prove part 3 of Proposition 29. In model M_3 both mutations occur on the stem cell line. In this case $u_1 \tau'(M_3)$ and $u_2(\tau(M_3) - \tau'(M_3))$ are both exponentially distributed with mean 1. We have that $u_1(\tau(M_3) - \tau'(M_3))$ is exponentially distributed with mean u_1/u_2 which can be observed by rewriting $u_1(\tau(M_3) - \tau'(M_3))$ as $(u_1/u_2)u_2(\tau(M_3) - \tau'(M_3))$. By assumption, $u_2/u_1 \to 1/A$ so $u_1(\tau(M_3) - \tau'(M_3))$ converges in distribution to Z. The random variables $\tau'(M_3)$ and $\tau(M_3) - \tau'(M_3)$ are independent for each N so

$$u_1 \tau(M_3) = u_1 \tau'(M_3) + u_1(\tau(M_3) - \tau'(M_3)) \to_d X + Z.$$

$\square$

3.4 Proof of the Theorem

Proof of part 3 of Theorem 18. We shall make use of the following well known fact:
If $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that $a_n \to a$, then

$$\lim_{n \to \infty} \left(1 - \frac{a_n}{n}\right)^{n-1} = e^{-a}.$$

Before time 1 the cells never split and there is no apoptosis. Let H_1' be
the same as model H_1 except that stem cells never receive mutations. Note that
H_1' differs from M_1 because daughter cells cannot accumulate type-1 mutations in
model H_1'. If we ignore the splitting and apoptosis and consider how long it takes
for a cell to acquire two mutations under the mutation mechanism alone then we
have $N - 1$ daughter cells acquiring mutations independently. For any individual
cell, the time it takes to acquire two mutations will have the same distribution
as the sum of two independent exponentially distributed random variables with
means $1/v_1$ and $1/v_2$. If we denote the time until cell i has a type-2 mutation by
T_i and assume $v_1 \neq v_2$ then

$$P(T_i \leq t) = 1 - \frac{v_2 e^{-v_1 t} - v_1 e^{-v_2 t}}{v_2 - v_1}.$$

There are $N - 1$ cells independently getting mutations, so for $t \leq 1$ we have

$$P(\tau(H_1') \leq t) = 1 - \left(\frac{v_2 e^{-v_1 t} - v_1 e^{-v_2 t}}{v_2 - v_1}\right)^{N-1},$$

or equivalently,

$$P(\sqrt{v_1 v_2 N}\tau(H_1') \leq t) = 1 - \left(\frac{v_2 e^{-\sqrt{v_1/(v_2 N)}t} - v_1 e^{-\sqrt{v_2/(v_1 N)}t}}{v_2 - v_1}\right)^{N-1}.$$

Notice that

$$N\sqrt{v_1^3/v_2 N^3} = v_1^2/\sqrt{v_1 v_2 N} \to 0 \text{ and } N\sqrt{v_2^3/v_1 N^3} = v_2^2/\sqrt{v_1 v_2 N} \to 0.$$

For N large enough we can apply the third degree Taylor expansion of the expo-
nential function to get the bounds

$$1 - \frac{t^2}{2N} - \sqrt{\frac{v_1^3}{v_2 N^3}}\frac{t^3}{6} \leq \frac{v_2 e^{-\sqrt{v_1/(v_2 N)}t} - v_1 e^{-\sqrt{v_2/(v_1 N)}t}}{v_2 - v_1} \leq 1 - \frac{t^2}{2N} + \sqrt{\frac{v_2^3}{v_1 N^3}}\frac{t^3}{6}.$$

For any fixed t we have

$$\left(1 - \frac{t^2}{2N} - \sqrt{\frac{v_1^3}{v_2 N^3} \frac{t^3}{6}}\right)^{N-1} \to e^{-t^2/2}$$

and

$$\left(1 - \frac{t^2}{2N} + \sqrt{\frac{v_2^3}{v_1 N^3} \frac{t^3}{6}}\right)^{N-1} \to e^{-t^2/2}.$$

If $v_1 = v_2$ and we ignore splitting and apoptosis then the probability that one cell has two mutations by time t is $1 - e^{-v_1 t} - v_1 t e^{-v_1 t}$. The probability that one of the N cells has two mutations by time t is $1 - (e^{-v_1 t} - v_1 t e^{-v_1 t})^N$. By applying the same techniques as above we get $P(\sqrt{v_1 v_2 N}\tau(H_1') \leq t) \to 1 - e^{-t^2/2}$ when $v_1 = v_2$.

Combining the two results above we have $P(\sqrt{v_1 v_2 N}\tau(H_1') \leq t) \to 1 - e^{-t^2/2}$ when ignoring splitting and apoptosis. Then

$$P(\tau(H_1') < 1) = P(\sqrt{v_1 v_2 N}\tau(H_1') < \sqrt{v_1 v_2 N}) \to 1.$$

Therefore, the probability that two mutations occur before time 1 is converging to 1 so we may ignore splitting and apoptosis in this case. This gives the desired result for $\tau(H_1')$.

Stem cells get type-1 mutations at rate $u_1 \to 0$ in model H_1. Let T be the first time the stem cell line gets a mutation in model H_1. Then $P(T < 1) \to 0$. We can couple models H_1 and H_1' so that the same Poisson processes are marking the mutations on the daughter cells. Then

$$P(\tau(H_1) = \tau(H_1')) \geq P(\{T \geq 1\} \cap \{\tau(H_1') < 1\}) \to 1$$

which gives the results for model H_1.

Because any cell is equally likely to get the two mutations, it is clear that $\sigma(H_1)$ and $\rho(H_1)$ both converge in probability to 1. $\qquad\square$

This gives the result for part 3 of Theorem 18 even if $\alpha = 0$.

In this section we will apply the following lemma several times.

Lemma 33. *Let $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\beta_n\}_{n=1}^{\infty}$ be sequences of positive numbers which converge to 0. Let $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ be sequences of random variables and let X and Y be positive random variables such that $\alpha_n X_n$ converges in distribution to X and $\beta_n Y_n$ converges in distribution to Y as $n \to \infty$. If $\alpha_n/\beta_n \to 0$ as $n \to \infty$ then $P(X_n \geq Y_n) \to 1$ as $n \to \infty$.*

Proof. Note that $\alpha_n Y_n = (\alpha_n/\beta_n)\beta_n Y_n$ and $\alpha_n/\beta_n \to 0$ so $\alpha_n Y_n$ converges in probability to 0. Let $F(t) = P(X \leq t)$. Let $\delta > 0$ and choose $\epsilon > 0$ such that $F(t)$ is continuous at ϵ and $F(\epsilon) < \delta/2$. Choose N_1 such that if $n \geq N_1$ then $P(\alpha_n X_n > \epsilon) > 1 - \delta$. Choose N_2 such that if $n \geq N_2$ then $P(\alpha_n Y_n \geq \epsilon) < \delta$. By independence, for $n \geq N_1 \vee N_2$ we have

$$
\begin{aligned}
P(X_n > Y_n) &= P(\alpha_n X_n > \alpha_n Y_n) \\
&\geq P(\{\alpha_n X_n > \epsilon\} \cap \{\epsilon > \alpha_n Y_n\}) \\
&= 1 - P(\{\alpha_n X_n \leq \epsilon\} \cup \{\alpha_n Y_n \geq \epsilon\}) \\
&\geq 1 - 2\delta.
\end{aligned}
$$

$\square$

We will couple the models H_2, M_1, M_2 and M_3 so that the Poisson processes used in models M_1, M_2 and M_3 are the appropriate subcollections of Poisson processes which are used in model H_2. Let T be the time that a type-1 mutation occurs on the stem cell line in model H_2. Note that because stem cells cannot inherit type-1 mutations the coupling implies that $T = \tau'(M_2) = \tau'(M_3)$.

Lemma 34. *Assume that $v_1 v_2 \ll 1/(N(\log N)^2)$. If $u_1 \ll v_1 v_2 N \log N$ then $P(\tau(M_1) < T) \to 1$. If $u_1 \gg v_1 v_2 N \log N$ then $P(\tau(M_3) < \tau(M_1)) \to 1$.*

Proof. By part 1 of Proposition 21 $(\alpha \wedge 1)v_1 v_2 N(\log N)\tau(M_1) \to_d X$. Mutations to the stem cell line occur at rate u_1 so $u_1 T \to_d X$. Because the Poisson processes that mark the mutations in model M_1 are independent of the Poisson process that marks the mutations on the stem cell line, if $u_1 \ll v_1 v_2 N \log N$ then $P(\tau(M_1) < T) \to 1$ by Lemma 33.

On the other hand, suppose $u_1 \gg v_1 v_2 N \log N$. We are assuming $u_1 \leq u_2$ so we could decrease $P(\tau(M_3) < \tau(M_1))$ by decreasing u_2 to u_1. Then the distribution of $u_1 \tau(M_3)$ is the distribution of the sum of two independent exponentially distributed random variables. By Lemma 33, $P(\tau(M_3) < \tau(M_1)) \to 1$. $\qquad \square$

Lemma 35. *Assume that* $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$. *If* $u_1 \ll \sqrt{v_1 v_2 N}$ *then* $P(\tau(M_1) < T) \to 1$. *If* $u_1 \gg \sqrt{v_1 v_2 N}$ *then* $P(\tau(M_3) < \tau(M_1)) \to 1$.

Proof. First suppose $u_1 \ll \sqrt{v_1 v_2 N}$. It follows by part 2 of Proposition 21 that $\sqrt{v_1 v_2 N} \tau(M_1) \to_d Y$. The stem cell line is getting mutations at rate u_1 so $u_1 T \to X$. The Poisson processes that are marking the mutations in model M_1 are independent of the Poisson process that marks mutations on the stem cell line, so the result follows by Lemma 33.

If $u_1 \gg v_1 v_2 N \log N$ then the proof follows by the same reasoning as used in Lemma 34 when considering $u_1 \gg v_1 v_2 N \log N$. $\qquad \square$

Lemma 36. *If* $u_2 \ll 1/\log N$ *and* $u_2 \ll N v_2$ *then* $P(\tau(M_2) < \tau(M_3)) \to 1$.

Proof. By the coupling $\tau'(M_2) = \tau'(M_3)$. After time $\tau'(M_2)$ the Poisson processes marking the mutations in models M_2 and M_3 are independent. Let

$$T_2 = \tau(M_2) - \tau'(M_2) \text{ and } T_3 = \tau(M_3) - \tau'(M_3).$$

Then $P(\tau(M_2) < \tau(M_3)) = P(T_2 < T_3)$.

Consider again the model M_2' that was introduced in the proof of Lemma 32 which is the same as model M_2 except that the type-2 mutations can only occur on daughter cells $\log N$ time units after the stem cell line has a type-1 mutation. We can couple models M_2 and M_2' as we did before so that the time at which the stem cell line gets a mutation is the same in models M_2 and M_2'. In particular, $\tau'(M_2') = \tau'(M_2) = \tau'(M_3)$. Let $T_2' = \tau(M_2') - \tau'(M_2')$. Then $T_2' \geq T_2$ so it is enough to show that $P(T_2' < T_3) \to 1$.

If we wait $\log N$ time after the stem cell line receives a type-1 mutation then all of the daughter cells will be type-1 and the $(N-1)$ daughter cells are getting type-2 mutations at rate v_2. Thus for any fixed N we have

$$P(T_2' > t) = 1_{[0,\log N]}(t) + e^{-v_2(N-1)(t-\log N)} 1_{(\log N, \infty]}(t).$$

Let $\epsilon > 0$. Then

$$P(T_2' < T_3) = P(T_2' < T_3 | T_3 < \log N)P(T_3 < \log N)$$
$$+ P(T_2' < T_3 | T_3 \geq \log N)P(T_3 \geq \log N).$$

Because $u_2 \ll 1/\log N$ and $u_2 T_3$ has the exponential distribution with mean 1, we have $P(T_3 \geq \log N) \to 1$. The memoryless property of the exponential distribution gives us

$$P(T_2' < T_3 | T_3 \geq \log N) = \frac{v_2(N-1)}{v_2(N-1) + u_2} \to 1$$

which completes the proof. $\qquad\square$

Lemma 37. *If $u_2 \gg 1/\log N$ or $u_2 \gg Nv_2$ then $P(\tau(M_3) < \tau(M_2)) \to 1$.*

Proof. By the coupling $\tau'(M_2) = \tau'(M_3)$. After time $\tau'(M_2)$ the Poisson processes marking the mutations in models M_2 and M_3 are independent. Let

$$T_2 = \tau(M_2) - \tau'(M_2) \text{ and } T_3 = \tau(M_3) - \tau'(M_3).$$

Then $P(\tau(M_3) < \tau(M_2)) = P(T_3 < T_2)$.

Suppose $u_2 \gg 1/\log N$. By Lemma 31 we know that $\rho(M_2) \to_p \alpha \wedge 1$. If $0 < \delta < (\alpha \wedge 1)$ then $P(\rho(M_2) > (\alpha \wedge 1) - \delta) \to 1$. If $\rho(M_2) > (\alpha \wedge 1) - \delta$ then the second mutation occurs on a generation higher than $((\alpha \wedge 1) - \delta)l$. Since only stem cells get type-1 mutations in model M_2 we have that $T_2 \geq \lfloor ((\alpha \wedge 1) - \delta)l \rfloor$ because it takes at least that much time for the type-1 mutation to spread to the generation $\lceil ((\alpha \wedge 1) - \delta)l \rceil$ daughter cells. On the other hand, in model M_3 the second mutation is occurring at rate u_2 so that $u_2 T_3$ is exponentially distributed with mean 1. Then $P(T_3 < K \log N) = P(u_2 T_3 < u_2 K \log N) \to 1$ for any positive number K since $u_2 \log N \to \infty$. Therefore $P(T_3 < T_2) \to 1$.

Suppose $u_2 \gg Nv_2$. The rate at which type-2 mutations occur in model M_2 is always bounded by $(N-1)v_2$. Suppose we consider a new model M_2'' which is the same as M_2 except that once the stem cell line has a type-1 mutation, all of the daughter cells also have a type-1 mutation instantaneously. Models M_2 and M_2'' can be coupled so that after the stem cell line gets a type-1 mutation then any type-2 mutation proposed by a Poisson process on a daughter cell is accepted

in model M_2''. Let $T_2'' = \tau(M_2'') - \tau'(M_2'')$. Then $(N-1)v_2 T_2''$ has the exponential distribution with mean 1. By Lemma 33, $P(T_3 < T_2'') \to 1$. Because $T_2 \geq T_2''$ we have the desired result. $\qquad\square$

Proof of Theorem 18. From the coupling we have $\tau(H_2) = \tau(M_1) \wedge \tau(M_2) \wedge \tau(M_3)$ because any type-2 mutation which occurs in model H_2 must occur in at least one of the models M_i for some i, and if a mutation occurs in model M_i then it will also occur in model H_2.

Suppose $P(\tau(M_1) < T) \to 1$. Before time T only stem cells are acquiring type-1 mutations in models M_2 and M_3. Therefore, models M_2 and M_3 only have type-0 cells before time T and $P(\tau(M_1) < \tau(M_2) \wedge \tau(M_3)) \to 1$.

- Suppose $v_1 v_2 \ll 1/(N(\log N)^2)$ and $u_1 \ll v_1 v_2 N \log N$. By Lemma 34 it follows that $P(\tau(M_1) < T) \to 1$ so by part 1 of Proposition 21 and the coupling of H_2 with M_1 we have $(\alpha \wedge 1)v_1 v_2 N(\log N)\tau(H_2) \to_d X$. Also by Lemma 34, the distribution of $\sigma(H_2)$ converges to a uniform distribution on $((1-\alpha)^+, 1]$ and $\rho(H_2)$ converges in distribution to 1.

- By Lemma 35 if $1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N$ and $u_1 \ll \sqrt{v_1 v_2 N}$ then $P(\tau(M_1) < T) \to 1$ so by part 2 of Proposition 21 and the coupling of H_2 with M_2 we have $\sqrt{v_1 v_2 N}\tau(H_2) \to_d Y$. Also by Lemma 35, both $\sigma(H_2)$ and $\rho(H_2)$ converge in distribution to 1.

If either

$$v_1 v_2 \ll 1/(N(\log N)^2) \text{ and } u_1 \gg v_1 v_2 N \log N$$

or

$$1/(N(\log N)^2) \ll v_1 v_2 \ll 1/N \text{ and } u_1 \gg \sqrt{v_1 v_2 N}$$

then $P(\tau(M_3) < \tau(M_1)) \to 1$ by Lemmas 34 and 35 respectively. Therefore, $P(\tau(M_2) \wedge \tau(M_3) < \tau(M_1)) \to 1$ which implies that the cancer causing type-1 mutation occurs on the stem cell line in model H_2 with probability converging to 1. Given these four conditions, we are left only to compare $\tau(M_2)$ and $\tau(M_3)$.

- By Lemma 36 if $u_2 \ll 1/\log N$ and $u_2 \ll Nv_2$ then $P(\tau(M_2) < \tau(M_3)) \to 1$. Because $u_1 \leq u_2$ the hypotheses are true for u_1 as well. Therefore, by the coupling of H_2 with M_2 and part 1 of Proposition 29 we have $u_1\tau(H_2) \to_d X$ and $\rho(H_2)$ converges in probability to $\alpha \wedge 1$.

- By Lemma 37 if $u_2 \gg 1/\log N$ or $u_2 \gg Nv_2$ then $P(\tau(M_3) < \tau(M_2)) \to 1$. If $u_1 \ll u_2$ then by the coupling of H_2 with M_3 and part 2 of Proposition 29 we have $u_1\tau(H_2) \to_d X$. If $u_1 \sim Au_2$ then by the coupling of H_2 with M_3 and part 3 of Proposition 29 we have $u_1\tau(H_2) \to_d X + Z$ where Z is an exponentially distributed random variable with mean A that is independent of X.

By Lemma 20 the results hold for model H_1 as well. $\qquad\square$

3.5   The Null Model

For this section we always have $u_1 = u_2 = v_1 = v_2 = \mu$ and we prove Proposition 19 for model H_2. Then Proposition 19 will hold for model H_1 as well by Lemma 20. We begin this section by pointing out that the conditions of part 5 of Theorem 18 always fail in the null model. The two conditions in the first conjunction become $\mu \ll 1/(N \log N)$. Of the two conditions in the second conjunction, one becomes $\sqrt{N} \ll 1$ which always fails. This reduces all of the conditions in the first bullet point to $\mu \ll 1/(N \log N)$. The conditions in the second bullet point become $\mu \gg 1/\log N$ or $1 \gg N$, so the conditions in part 5 are reduced to $\sqrt{N} \ll 1$, $1 \gg N$ or $1/\log N \ll \mu \ll 1/(N \log N)$ which all fail.

This shows that the probability that the first type-2 mutation occurs on the stem cell line converges to 0. For this reason, we will never consider model M_3 in this section.

Proof of part 2 of Proposition 19. We can couple model H_2 with models M_1 and M_2 such that the Poisson processes marking model M_1 are independent of the Poisson processes marking model M_2. Before time $\tau'(M_2)$ the Poisson processes marking model M_1 are also marking the daughter cells in model H_2 and the Poisson

process that marks the stem cell in model M_2 is also marking the stem cell line in model H_2. After time $\tau'(M_2)$, the Poisson processes marking the cells in model M_1 are only marking the daughter cells in model H_2 that have not yet inherited the type-1 mutation from the stem cell. All of the Poisson processes marking type-2 mutations on cells in model M_2, meaning that those cells have inherited the type-1 mutation from the stem cell, also mark the corresponding cells in model H_2. After time $\tau'(M_2) + \log N$, only the Poisson processes marking model M_2 are marking model H_2.

Let T be the time at which the first successful type-1 mutation occurs in model M_1 and let Z be the time at which the first successful type-1 mutation occurs in model H_2. By Corollary 25 we have $A\mu T \to_d X$. Because the stem cell is getting type-1 mutations at rate μ and every type-1 mutation on the stem cell is successful, we have $(A+1)\mu Z \to_d X$. If the first successful type-1 mutation occurs on a daughter cell, then the type-2 mutation must occur within $\log N$ time of Z since after this time the progeny of the cell will no longer be in the population. Let Y_2 be the time it takes to get the second successful type-1 mutation after the first has occurred. If the first successful type-1 mutation occurs on the stem cell then all of the cells will be type-1 within $\log N$ time. Therefore, if the first successful type-1 mutation occurs on the stem cell and there is not another successful type-1 mutation within $\log N$ time, $Y_2 = \infty$ since there can be no more type-1 mutations. We have $\limsup P((1 + A)\mu Y_2 \leq t) \leq 1 - e^{-t}$. Therefore,

$$\limsup P(Y_2 < (\tau(H_2) - Z)) \leq \limsup P(Y_2 < \log N)$$
$$= \limsup P((1 + A)\mu Y_2 < (A + 1)\mu \log N)$$
$$\leq 1 - e^{-(1+A)\mu \log N} \to 0.$$

As a similar result to the one in Lemma 26, we have $P(Z = \tau'(H_2)) \to 1$. Hence, it is enough to find the distribution of the time of the first successful type-1 mutation.

We have established $(A + 1)\mu Z \to_d X$ and $P(Z = \tau'(H_2)) \to 1$ which imply $(1 + A)\mu\tau'(H_2) \to_d X$. Let A_1 be the event that the first successful type-1 mutation occurs on a daughter cell and A_2 be the event that the first successful type-1 mutation occurs on the stem cell. If the first successful type-1 mutation

occurs on a daughter cell, then due to apoptosis $\tau(H_2) - Z$ is bounded above by $\log N$. Therefore

$$P(\{A\mu(\tau(H_2) - Z) > \epsilon\} \cap A_1) \to 0.$$

If the first successful type-1 mutation occurs on a stem cell, then in $\log N$ time all of the cells will be type-1 and type-2 mutations will occur at rate μN. Let $\hat{Z}$ be an exponentially distributed random variable with mean $1/\mu N$. Then we have

$$P(\{A\mu(\tau(H_2) - Z) > \epsilon\} \cap A_2) \le P(A\mu(\log N + \hat{Z}) > \epsilon) \to 0.$$

Since either A_1 or A_2 must occur, we have $A\mu(\tau(H_2) - Z) \to_p 0$. Then

$$(1 + A)\mu\tau(H_2) = A\mu(Z + (\tau(M_1) - Z)) \to_d X.$$

By the coupling, before time $\tau'(M_2)$ the daughter cells in model H_2 get successful type-1 mutations at the same rate as the daughter cells in model M_1. We know from the proof of Lemma 28 that each generation i with $1 \le i \le l$ is getting successful type-1 mutations independently at a rate bounded between $\mu 2^{i-1}(1 - e^{-\mu(2^{l-i+1}-2)})$ and $\mu 2^{i-1}(1 - e^{-\mu(2^{l-i+1}-1)})$ for any time t in model M_1. Therefore, these bounds also hold for the rate at which daughter cells get successful type-1 mutations in model H_2 before time $\tau'(M_2)$. Let $\beta \in [0, 1]$. Using the notation and result from Lemma 28 and the fact that the stem cell line is getting type-1 mutations at rate μ,

$$\limsup P(\sigma(H_2) \le \beta) \le \limsup \frac{\mu + \sum_{i \in (0, l\beta]^*} \mu 2^{i-1}(1 - e^{-\mu(2^{l-i+1}-1)})}{\mu + \sum_{i \in (0, l]^*} \mu 2^{j-1}(1 - e^{-\mu(2^{l-j+1}-2)})}$$
$$= \frac{1}{1 + A} + \frac{A}{1 + A}\beta$$

and

$$\liminf P(\sigma(H_2) \le \beta) \ge \liminf \frac{\mu + \sum_{i \in (0, l\beta]^*} \mu 2^{i-1}(1 - e^{-\mu(2^{l-i+1}-2)})}{\mu + \sum_{i \in (0, l]^*} \mu 2^{j-1}(1 - e^{-\mu(2^{l-j+1}-1)})}$$
$$= \frac{1}{1 + A} + \frac{A}{1 + A}\beta.$$

Lemma 20 gives the result for $\sigma(H_1)$.

Because $\rho(M_1)$ and $\rho(M_2)$ both converge in probability to 1, we will have $\rho(H_2) \to_p 1$ as well. Lemma 20 then implies $\rho(H_1) \to_p 1$. $\qquad\square$

Let $\mathcal{N}$ be the set of Radon measures ν on a Polish space $(\Psi, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-field such that $\nu(\{x\}) \in \mathbb{N} \cup \{0, \infty\}$ for all $x \in \Psi$. For the next proof we will consider a point process to be a random variable taking on elements of $\mathcal{N}$. We consider $\nu(\{x\})$ to be the number of times the point x has been marked. For a Poisson point process whose intensity measure has no atoms $\nu(\{x\})$ is 0 or 1 for all x and $\{x \in \Psi : \nu(\{x\}) > 0\}$ is discrete with probability 1.

Let $\Psi = [0, \infty) \times [0, 1]$. The Poisson point process of successful type-1 mutations in model M_1 induces a point process on Ψ where if a successful type-1 mutation occurs at time t on a cell in generation i in model M_1 then there is a point of Ψ at $(t/l, i/l)$. We will call this point process P_M.

Lemma 38. *If $\mu \sim A/(\sqrt{N} \log N)$ then the limiting distribution of P_M is a Poisson point process P_∞ which has intensity measure $\nu' = A^2(\lambda \times \lambda_{[1/2,1]})$ where λ is the Lebesgue measure and $\lambda_{[1/2,1]}$ is the measure defined by $\lambda_{[1/2,1]}(B) = \lambda(B \cap [1/2, 1])$ for any Lebesgue measurable set B.*

Proof. We let $C_C(\Psi, [-1, 0])$ be the set of continuous functions $h : \Psi \to [-1, 0]$ such that the set $\{\psi \in \Psi : h(\psi) \neq 0\}$ is precompact. Recall that a point process X has an associated generating functional $\mathfrak{F} : C_C(\Psi, [-1, 0]) \to \mathbb{R}$ defined by

$$\mathfrak{F}(h) = E[\prod_{\psi \in \Psi} (h(\psi) + 1)^{\nu(\psi)}]$$

where ν is a Radon measure on Ψ as described above. Probability generating functionals uniquely determine the distribution of point processes (see Theorem 14 of section 29.5 in [14]). Moreover, a sequence of point processes converges in distribution to a point process if and only if the corresponding sequence of generating functionals converges pointwise to a functional $\mathfrak{F}$ that satisfies the following: If h_m is in the domain of $\mathfrak{F}$ for each m, $\bigcup_{m=1}^{\infty}\{\psi : h_m(\psi) \neq 0\}$ is relatively compact, and $h_m(\psi) \to 0$ as $m \to \infty$ for each ψ, then $\mathfrak{F}(h_m) \to 1$ as $m \to \infty$. In this case $\mathfrak{F}$ is the probability generating functional of the limiting point process (see Theorem 20 of Section 29.7 in [14]).

Notice that for any N the points marked in Ψ will all have coordinates (x, y) where y takes values in $\{1/\log N, 2/\log N, \ldots, 1\}$. We know from the proof

of Lemma 28 that the rate at which mutations occur along generation i is bounded between $2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-2)})$ and $2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-1)})$. Therefore, if we look at the points that are marked in Ψ whose second coordinate is fixed at $i/\log N$, the rate at which the marking will occur will be between $(\log N)2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-2)})$ and $(\log N)2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-1)})$ where the $\log N$ appears because time is scaled by $1/\log N$. This observation will allow us to work with time homogeneous Poisson point processes.

Let $\mathfrak{F}$ denote the generating functional associated with P_M. Let $\mathfrak{F}_1$ be the generating functional associated with the Poisson process on Ψ which marks points at rate $(\log N)2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-2)})$ on $y = i/l$ and let $\mathfrak{F}_2$ be the generating functional associated with the Poisson process on Ψ which marks points at rate $(\log N)2^{i-1}\mu(1-e^{-\mu(2^{l-i+1}-1)})$ on $y = i/l$. Call the time homogeneous Poisson point processes P_1 and P_2 respectively. Because the intensity measure of P_M is always between the intensity measures of P_1 and P_2 we have the bounds $\mathfrak{F}_1 \le \mathfrak{F} \le \mathfrak{F}_2$.

Let X be a Poisson process with intensity measure ν. It is known that the probability generating functional associated with X is

$$\mathfrak{P}(h) = e^{-\int_\Psi h d\nu}.$$

To show a sequence of Poisson processes $\{X_n\}_{n=0}^\infty$ with intensity measures $\{\nu_n\}_{n=0}^\infty$ converges in distribution to a Poisson process X with intensity measure ν it is enough to show that $\{\nu_n\}_{n=0}^\infty$ converges weakly to ν. That is, for each function $h \in C_C(\Psi, [-1,0])$ we need $\int_\Psi h d\nu_n \to \int_\Psi h d\nu$ as $n \to \infty$. Let ν_N^1 be the intensity measure of P_1 when there are N cells in the population and let ν_N^2 be the intensity measure of P_2 when there are N cells in the population. The goal is to show ν_N^1 and ν_N^2 both converge weakly to ν'. Then the limiting distribution of P_M will be P_∞.

Let $R = (a, b] \times (c, d] \subset \Psi$. Then

$$\nu_N^1(R) = (b-a)(\log N) \sum_{i \in (lc, ld]} 2^i \mu(1 - e^{-\mu(2^{l-i+1}-2)})$$

$$\to A^2(d - c \vee \tfrac{1}{2})^+(b-a) = \nu'(R)$$

by Lemma 24 and the fact that $\mu^2 N \log N \sim A^2 / \log N$ which follows from the assumption that $\mu \sim A/(\sqrt{N} \log N)$. Now let O be any open subset of Ψ. We can write $O = \bigcup_{n=1}^{\infty} R_n$ where each R_n is a half open rectangle in the same form as R above and the sets $\{R_n\}_{n=1}^{\infty}$ are pairwise disjoint. Then

$$\liminf_{N \to \infty} \nu_N^1(O) = \liminf_{N \to \infty} \sum_{j=1}^{\infty} \nu_N^1(R_j) \geq \sum_{j=1}^{\infty} \nu'(R_j) = \nu'(O)$$

where the inequality follows by Fatou's lemma. The same reasoning applied to ν_N^2 implies that $\liminf \nu_N^2(O) \geq \nu'(O)$ for any open subset O of Ψ also. It follows by the Portmanteau Theorem that both ν_N^1 and ν_N^2 converge weakly to ν' as N goes to infinity. Hence the limiting distribution of P_M is P_{∞}. $\qquad \square$

The notation used in Lemma 38 will also be used in this proof.

Proof of part 4 of Proposition 19. Notice that this is the boundary between two cases that are determined by model M_1. By Corollary 23 we know $\rho(M_1) \to_p 1$ for all conditions that we are considering. Therefore, $\rho(H_1) \to_p 1$ in this case.

The strategy is to define functions g and h on the set of Radon measures that are continuous everywhere except a set of measure 0. Then we will apply the Continuous Mapping Theorem to get the desired convergence in distribution. Let D be the subset of $\mathcal{N}$ such that $\nu \in D$ if there exists $(x, y) \in \Psi$ and $t \in \mathbb{R}$ such that $\nu(x, y) > 0$ and $\nu(x + t, y + t) > 0$. For all $t \geq 0$ define sets

$$T_t = \{(x, y) : 1/2 \leq y \leq 1 \text{ and } 0 \leq x \leq y + t - 1\} \subset \Psi.$$

These sets correspond the the triangles and quadrilaterals that were shown in Figure 3.4. Let $V = \{(x, y) \in \Psi : \nu(x, y) > 0\}$ and define $t_0 = \inf\{t : V \cap T_t \neq \varnothing\}$. Define

$$g(\nu) = \limsup_{\epsilon \to 0}\{y : (x, y) \in V \cap T_{t_0 + \epsilon} \text{ for some } x\}$$

and $h(\nu) = t_0$.

Given a Poisson point process P on Ψ whose intensity has no atoms, we can project the points of P onto the line $y = -x$ in $\mathbb{R}^2$ along perpendicular angles of $\pi/4$. With probability 1 no two points of P will be mapped to the same point under the projection. That is, under the law of P, D has probability 0. Moreover, with

probability 1 there will be no limit points under the projection. Therefore, under the intensity measure $A^2(\lambda_{[1/2,1]} \times \lambda)$, there exists a unique point $(x_0, y_0) \in V \cap T_{t_0}$ and an $\epsilon > 0$ such that $V \cap T_{t_0+\epsilon} = \{(x_0, y_0)\}$ with probability 1. By definition $g(P) = y_0$. We claim that g and h are continuous at any Radon measure $\nu \in \mathcal{N}\backslash D$.

Let $\nu \in \mathcal{N}\backslash D$ and let $\{\nu_n\}_{n=1}^{\infty}$ be a sequence of Radon measures that converges weakly to ν. Let $\epsilon > 0$ and let (x_0, y_0) be the unique point of $T_{t_0+\epsilon}$ such that $\nu(x_0, y_0) > 0$. For each point $(x', y') \in \Psi$ and every natural number m define a function

$$f_{(x',y'),m}(x, y) = \begin{cases} -1 & \text{if } |(x, y) - (x', y')| < \frac{\epsilon}{m} \\ -(2 - \frac{m|(x,y)-(x',y')|}{\epsilon}) & \text{if } \frac{\epsilon}{m} \leq |(x, y) - (x', y')| \leq \frac{2\epsilon}{m} \\ 0 & \text{otherwise} \end{cases}$$

For m large enough we have $\int_{\Psi} f_{(x_0,y_0),m}(x, y)d\nu = -1$. It follows that for m large enough $\int_{\Psi} f_{(x_0,y_0),m}(x, y)d\nu_n \to -1$ as $n \to \infty$. Because we can make m arbitrarily large, there must be a sequence of points $\{(x_n, y_n)\}_{n=1}^{\infty}$ such that $\nu_n(x_n, y_n) = 1$ for all n and $(x_n, y_n) \to (x_0, y_0)$ as $n \to \infty$. Likewise, for any point $(x', y') \in T_{t_0+\epsilon}$ there exists a large enough m such that $\int_{\Psi} f_{(x',y'),m}(x, y)d\nu = 0$ so it follows that $\int_{\Psi} f_{(x',y'),m}(x, y)d\nu_n \to 0$ as $n \to \infty$. This shows that for n large enough the Radon measures ν_n will assign measure 0 to all points in a ball of radius ϵ/m about (x', y'). From this it is easy to conclude $g(\nu_n) \to g(\nu)$ and $h(\nu_n) \to h(\nu)$. Therefore, g and h are both continuous on $\mathcal{N}\backslash D$. By Lemma 38 and the Continuous Mapping Theorem $g(P_M)$ converges in distribution to $g(P_{\infty})$ and $h(P_M)$ converges in distribution to $h(P_{\infty})$.

The next goal is to show that

$$g(P_M) - \sigma(M_1) \to_p 0 \text{ and } h(P_M) - \frac{\tau(M_1)}{\log N} \to_p 0.$$

Then we will have that $\sigma(M_1) \to_d g(P_{\infty})$ and $\tau(M_1)/\log N \to_d h(P_{\infty})$. To achieve this we will first show that the probability that (x_0, y_0) corresponds to the cancer causing type-1 mutation converges in probability to 1. Suppose (x_0, y_0) does not correspond to the cancer causing type-1 mutation and let (x_1, y_1) denote the point in Ψ corresponding to the cancer causing type-1 mutation in M_1. Let $\epsilon > 0$ and suppose that $(x_1, y_1) \notin T_{t_0+\epsilon}$. The point $(x_0, y_0) \in T_{t_0}$ corresponds to a successful

type-1 mutation in model M_1, and by the way that model M_1 marks points in Ψ there will be a type-2 mutation in model M_1 that corresponds to a point in T_{t_0}. The ray starting at (x_1, y_1) with an angle of $\pi/4$ will represent all of the descendants of the cancer causing type-1 mutation. The point on this line whose first coordinate is t_0 will be (t_0, y'') where $y'' \leq 1 - \epsilon$. In this case $\rho(M_1) = y'' \leq 1 - \epsilon$.

Let E_1 be the event that (x_0, y_0) is the point in Ψ that corresponds to the cancer causing type-1 mutation and E_2 be the event that two or more points occur in $T_{t_0+\epsilon}$. On E_1^C let (x_1, y_1) be the point in Ψ corresponding to the cancer causing type-1 mutation. We know that P_M converges in distribution to P_∞ by Lemma 38 so

$$
\begin{aligned}
\limsup P(E_1^C) &= \limsup(P(E_1^C \cap \{(x_1, y_1) \in T_{t_0+\epsilon}\}) \\
&\quad + P(E_1^C \cap \{(x_1, y_1) \notin T_{t_0+\epsilon}\})) \\
&\leq \limsup P(E_2) + \limsup P(\rho(M_1) < 1 - \epsilon) \\
&\leq \frac{A^2}{2}\epsilon
\end{aligned}
$$

where the last line follows because $P(E_2) \leq P(V \cap (T_{t_0+\epsilon}\backslash T_{t_0}) \neq \varnothing)$ and because $\rho(M_1) \to_p 1$. We chose $\epsilon > 0$ arbitrarily so we have $\lim P(E_1^C) = 0$.

The above has established that $\lim P(E_1) = 1$. By definition of $\sigma(M_1)$ and $g(P_M)$ it is clear that

$$
P(\sigma(M_1) - g(P_M) = 0|E_1) = 1
$$

because $\sigma(M_1) = g(P_M) = y_0$. Conditional on the event E_1 we also know that $\tau'(M_1) = (\log N)x_0$. Let (x_0', y_0') be the point in Ψ that corresponds to the type-2 mutation in M_1, so that $\rho(M_1) = y_0'$. Let ν be the Radon measure of points in Ψ induced by M_1 and consider the fact that the descendants of the cancer causing type-1 mutation will lie on a line starting at (x_0, y_0) with angle $\pi/4$. It is clear that $h(\nu) = t_0 = x_0 + 1 - y_0$ and $\rho(M_1) = y_0 + \tau(M_1)/\log(N) - x_0$. Thus, if $h(\nu) - \tau(M_1)/\log N > \epsilon$ then $1 - \rho(M_1) > \epsilon$, or equivalently $\rho(M_1) < 1 - \epsilon$. Therefore, because $P(E_1) \to 1$,

$$
P(h(P_M) - \tau(M_1)/\log N > \epsilon|E_1) = P(\rho(M_1) < 1 - \epsilon|E_1) \to 0.
$$

Again using the fact that $P(E_1) \to 1$ we get the desired result.

Now we are left to show that $g(P_\infty)$ and $h(P_\infty)$ have the distributions that are stated in part 4 of Proposition 19. We have $P(h(P_\infty) \leq t)$ is the probability that a point of the Poisson process with intensity $A^2(\lambda_{[1/2,1]} \times \lambda)$ has been marked in T_t. For $t \leq 1/2$ this is $1 - e^{-A^2t^2/2}$ and for $t > 1/2$ this is $1 - e^{-A^2t/2+A^2/8}$. Therefore,

$$P(\tau(M_1)/\log N \leq t) \to (1 - e^{-A^2t^2/2})1_{[0,1/2]}(t) + (1 - e^{-A^2t/2+A^2/8})1_{(1/2,\infty)}(t).$$

To find the distribution of $g(P_\infty)$ we will use the joint density function of $g(P_\infty)$ and $h(P_\infty)$. From the above computation it is clear that the density of $h(P_\infty)$ is

$$f_h(t) = A^2te^{-A^2t^2/2}1_{[0,1/2]}(t) + \frac{A^2}{2}e^{-A^2t/2+A^2/8}1_{(1/2,\infty)}(t).$$

Conditioned on the event that $h(P_\infty) = t$ we know that $g(P_\infty)$ will have uniform distribution. If $t \leq 1/2$ then $g(P_\infty)$ is uniformly distributed on the interval $[1-t, 1]$. If $t > 1/2$ then $g(P_\infty)$ is uniformly distributed on $[1/2, 1]$. This gives us the conditional density function

$$f_{g|h}(s|t) = \begin{cases} \frac{1}{t} & \text{if } 1 - t \leq s \leq 1 \text{ and } 0 \leq t \leq \frac{1}{2} \\ 2 & \text{if } \frac{1}{2} \leq s \leq 1 \text{ and } t > \frac{1}{2} \end{cases}.$$

Therefore, the joint density function of $g(P_\infty)$ and $h(P_\infty)$ is

$$f(s,t) = A^2e^{-A^2t^2/2}1_{[0,1/2]}(t)1_{[1-t,1/2]}(s) + A^2e^{-A^2t/2+A^2/8}1_{(1/2,\infty)}(t)1_{[1/2,1]}(s).$$

Integrating over t we find that the density of $g(P_\infty)$ is

$$f_g(s) = \left(\int_{1-s}^{1/2} A^2e^{-A^2t^2/2}dt + 2e^{-A^2/8} \right) 1_{[1/2,1]}(s).$$

This gives the desired limiting distribution for model M_1. By the usual coupling arguments the results will hold for model H_1 as well. $\square$

Proof of part 6 of Proposition 19. Note that under these conditions both mutations occur on daughter cells with probability tending to 1. First we consider

a model M_1' so that only generation $l-1$ will get type-1 mutations and generation l will get type-2 mutations. Also, assume that only one of the daughters will keep a mutation when the cells split so that if a type-1 cell splits it has a type-0 daughter and a type-1 daughter. The rate at which the type-1 mutations occur will be $\mu N/4$ since there are $N/4$ cells in generation $l-1$. Note that $\mu N/4 \sim A\sqrt{N}/4$. The probability that a type-1 mutation will have a type-2 descendant is $1 - e^{\mu t} \sim \mu t \sim At/\sqrt{N}$. Therefore, the type-2 mutations occur according to a Poisson process whose intensity measure ν satisfies $\nu([0,t]) \geq (A\sqrt{N}/4)(At/\sqrt{N}) = A^2 t/4$. We have may have to wait up to two time units for the type-2 mutation to occur after the successful type-1 appears. For the sake of a lower bound we will always assume it takes 2 time units after a successful type-1 mutation until the type-2 mutation. By coupling model M_1' with model M_1 in the obvious way we have $\liminf P(\tau(M_1) \leq t) \geq 1 - e^{-2-A^2 t/4}$.

For the upper bound we consider a model M_1'' in which type-1 cells never undergo apoptosis. There are $N-1$ cells getting type-1 mutations so the type-1 mutations occur at rate $\mu(N-1) \sim A\sqrt{N}$. If we wait t time units after a type-1 mutation has occurred on a cell then the cell will have at most 2^t descendants. If the type-1 mutation had occurred at time 0 and all of the descendants had existed since the type-1 mutation occurred then the probability that one of the cells had acquired a type-2 mutation would be $t2^{\lfloor t \rfloor}\mu \leq t2^t\mu \sim t2^t A/\sqrt{N}$. Because the type-1 mutation may occur after time 0 and there have not been 2^t descendants with the type-1 mutation since the mutation occurred this is an upper bound on the probability that a type-2 mutation has occurred by time t. Therefore, the type-2 mutations occur according to a Poisson process with intensity measure defined by $\nu([0,t]) \leq (A\sqrt{N})(t2^t A/\sqrt{N}) = t2^t A^2$. By coupling model M_1'' with model M_1 in the obvious way we have $\limsup P(\tau(M_1) \leq t) \leq 1 - e^{-A^2 2^t t}$. This shows part 6 of Proposition 19 with $c = 1 - e^{-2-A^2 t/4}$ and $C = 1 - e^{-A^2 2^t t}$.

By Corollary 23 we know $\rho(M_1) \to 1$. By the definitions of $\sigma(M_1)$ and $\rho(M_1)$ for any $\epsilon > 0$ if $\rho(M_1) - \sigma(M_1) > \epsilon$ then $\tau(M_1) > \epsilon \log N$. Therefore,

$$P(\rho(M_1) - \sigma(M_1) > \epsilon) \leq P(\tau(M_1) > \epsilon \log N) \leq e^{-A^2 2^{\delta \log N}(\delta \log N)} \to 0.$$

Let $\epsilon > 0$ and $\delta > 0$ and choose N large enough so that $P(1 - \rho(M_1) > \epsilon/2) < \delta/2$

and $P(\rho(M_1) - \sigma(M_2) > \epsilon/2) < \delta/2$. Then

$$
\begin{aligned}
P(1 - \sigma(M_1) > \epsilon) &= P(1 - \rho(M_1) + \rho(M_1) - \sigma(M_1) > \epsilon) \\
&\leq P(1 - \rho(M_1) > \epsilon/2) + P(\rho(M_1) - \sigma(M_1) > \epsilon/2) \\
&< \delta.
\end{aligned}
$$

Therefore, $\sigma(M_1) \to_p 1$.

By the usual coupling arguments we get the same results for H_1. $\qquad\square$

Chapter 3, in full, has been accepted for publication in the Journal of Applied Probability. The dissertation author was the sole author of this paper.

Bibliography

[1] P. Armitage, R. Doll (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1-12.

[2] E. Brunet, I. Rouzine, C. Wilke (2008). The stochastic edge in adaptive evolution. *Genetics* **179**, 603-620.

[3] J. Cairns (2006). Cancer and the Immortal Strand Hypothesis. *Genetics* **174**, 1069-1072.

[4] P. Calabrese, J.P. Mecklin, J. Järvinen, L.A. Aaltonen, S. Tavaré and D. Shibata (2005). Numbers of mutations to different types of colorectal cancer. *BMC Cancer* **5**, 126.

[5] The Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068.

[6] M. Desai, D.S. Fisher, (2007). Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759-1798.

[7] D. Dingli, F. Michor (2006). Successful therapy must eradicate cancer stem cells. *Stem Cells* **24**, 2603-2610.

[8] R. Durrett. Probability Models for DNA Sequence Evolution, 2nd Ed. Springer, 2008.

[9] R. Durrett and J. Mayberry (2011). Travelling waves of selective sweeps. *Ann. Appl. Probab.* **21**, 699-744.

[10] R. Durrett, S. Moseley (2010). A Spatial Model for Tumor Growth. Preprint.

[11] R. Durrett, D. Schmidt, J. Schweinsberg (2009). A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Probab.* **19**, 676-718.

[12] A. Etheridge and F. Yu (2011). Girsanov Transformation and the Rate of Adaptation. Preprint.

[13] S.A. Frank, Y. Iwasa, M.A. Nowak (2003). Patterns of Cell Divisions and the Risk of Cancer. *Genetics* **163**, 1527-1532.

[14] B. Fristedt, L. Gray, A Modern Approach to Probability Theory, Birkhäuser Boston, 1997.

[15] J. Haigh (1978). The accumulation of deleterious genes in a population. *Theor. Popul. Biol.* **14**, 251-267.

[16] D. Hanahan, R.A. Weinberg (2000). The hallmarks of cancer. *Cell* **100**, 57-70.

[17] W.G. Hill and A. Robertson (1966). The effect of linkage on limits to artificial selection. *Genetics Research* **8**, 269-294.

[18] Y. Iwasa, F. Michor, N.L. Komarova, M.A. Nowak (2005). Population Genetics of Tumor Suppressor Genes. *J. Theor. Biol.* **233**, 15-23.

[19] S. Jones, et al. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806.

[20] J.F.C. Kingman, Poisson Processes, Oxford University Press Inc., New York, 1993.

[21] A.G. Knudson (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68**, 820-823.

[22] A.G. Knudson (1978). Retinoblastoma: A Prototypic Hereditary Neoplasm *Semin. Oncol.* **5**, 57-60.

[23] N.L. Komarova (2007). Loss- and gain-of-function mutations in cancer: mass-action, spatial and hierarchical models, *J. Stat. Phys.* **128**, 413-446.

[24] N.L. Komarova, P. Cheng (2006). Epithelial tissue architecture protects against cancer. *Math. Biosci.* **200**, 90-117.

[25] N.L. Komarova, A. Sengupta, M.A. Nowak (2003). Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J. Theor. Biol.* **233**, 433-450.

[26] N.L. Komarova, L. Wang (2004). Initiation of Colorectal Cancer. *Cell Cycle* **3:12** 1558-1565.

[27] F. Michor (2007). Chronic myeloid leukemia blast crises arises from progenitors. *Stem Cells* **25**, 1114-1118.

[28] H.J. Muller (1951). Radiation damage to the genetic material. In *Science in Progress, Seventh Series* (G.A. Baitsell, ed.), Yale University Press, pp. 93-165.

[29] M.A. Nowak, Evolutionary Dynamics: Exploring the Equations of Life, Belknap Press, 2006.

[30] S.C. Park, D. Simon, J. Krug (2010). The speed of evolution in large asexual populations. *J. Stat. Phys.* **138**, 381-410.

[31] D.W. Parsons, et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812.

[32] I. Rouzine, E. Brunet, C. Wilke (2007). The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Popul. Biol.* **73**, 24-46.

[33] J. Schweinsberg (2008). The waiting time for m mutations. *Electron. J. Probab.* **13**, 1442-1478.

[34] D. Wodarz, N.L. Komarova, Computational Biology of Cancer: Lecture Notes and Mathematical Modeling, World Scientific Publishing, 2005.

[35] F. Yu, A. Etheridge (2008), Rate of Adaptation of Large Populations in *Evolutionary Biology from Concept to Application*, Springer, 3-27.

[36] F. Yu, A. Etheridge, C. Cuthbertson (2009). Asymptotic Behaviour of the Rate of Adaptation. *Ann. Appl. Probab.* **20**, 978-1004.