©Copyright 2022
Ruibo Zhang

Branching process models for cancer evolution: overview and an application to colorectal cancer initiation

Ruibo Zhang

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Committee:

Ivana Bozic, Chair

Matt Lorig

Program Authorized to Offer Degree:

Applied Mathematics

University of Washington

Abstract

Branching process models for cancer evolution:
overview and an application to colorectal cancer initiation

Ruibo Zhang

Chair of the Supervisory Committee:
Professor Ivana Bozic
Department of Applied Mathematics

We study a multi-type branching process model associated with a transitional network between types. In particular, we are interested in determining the waiting time to each type in the network, employing an approximation of the process by its large time limit. We first present a literature review of results on large time limits and classify the dynamics of branching processes by their mean value matrices. For the special case when the transitional network is a single pathway, we present two approaches regarding approximating the waiting time for each type.

To apply our theory to a real world problem, we use a multi-type branching process to model the development of colorectal cancer from initially healthy tissue. The model incorporates a complex sequence of driver gene alterations, some of which result in immediate growth advantage, while others have initially neutral effects. We derive analytic estimates for the sizes of premalignant subpopulations, and use these results to compute the waiting time distributions of novel driver mutations.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First, I would like to express my deepest thanks to Professor Ivana Bozic. In my first year, I was not able to be in Seattle and had to take online courses. Despite the 5,396 mile geographical gap and the 15-hour time gap, Prof. Ivana dedicated a large amount of her time to mentorship me and provided numerous researching ideas. I cannot thank her enough for her endless patience while I'm a international student who always comes up weird sentences, and an unrealistic theorist who knows nothing about biology. Presenting this complete draft today would become an impossible event without her.

I would like to acknowledge Obinna A. Ukogu for helping me in nearly every stage of my research. While focusing on his new project, he still spent sometime to meet with me and provided numerous valuable comments. When I wrote the draft of our project, he was the one who went through the article again and again, adding countably infinite 'the', correcting almost every prepositions and rewording awkward sentence structures. Obinna is the kind of Ph.D. student that I want to be in the future.

There are numerous other names that I need to mention. I'd like to thank all other members in the cancer group. They are Trevor Crupi, Ike G. Salas, Marlin Figgins, Nathan Lee, Alanna Sholokhova, Katherine Owens, Amin Rahman, Kamran Kaveh and Konstantinos Mamis. They have provided spectacular presentations about the field. I would like to thank Jill Fulmore, Lauren Lederer and Rohan Rele for explaining me the process of doing a thesis. I would like to thank Professor Matt Lorig for coming to my thesis defense and has written great lecture notes for the amath probability series.

I would like to thank a bunch of my friends including my former roommates, Haoran Lu, Ligeng Yan and Rui Cheng, who have formed my habit of waking up at 8 in the morning - so

that I can join their game night in China. Next, I acknowledge Shey Gao and Eve Chen, my former amath 352 students, who made me feel being a TA is not completely meaningless - at least I won't forget how to speak Chinese. I would like to thank Ziyu Lu and Yiting Zhang, who sometimes play badminton with me. I hurt my right foot in the last game and eventually resulted in having plenty of time sitting down and writing my thesis. I would like to thank Ruiyu Han, an extraordinary outstanding Ph. D. student at CMU, who traveled to Seattle and brought sunshine and the heat wave. She also brought me endless happiness while I was writing this thesis. After she went back to Pittsburgh, Seattle 'burst into tears' immediately, celebrating the return of cool air. Their various types of supports play important roles in my life.

Lastly, I would like to thank my parents, Lan Mi and Yongxiang Zhang, who have proved the existence and, more importantly, the uniqueness of Ruibo. Without their love, I cannot be who I am.

DEDICATION

I dedicate this thesis to my family and friends for supporting me during my master's program.

Chapter 1

INTRODUCTION

Mathematical oncology seeks to model the evolutionary dynamics of cancer including initiation, progression, metastasis and treatment. The resulting cellular dynamics, the population change of cancer cells across a certain time period, is applicable to cancer prevention, diagnosis, and treatment.

Cancer evolution is a complex process in which cells may go through events (or steps), such as migration or mutation, and change biological states (e.g. genotype, location, etc.). Naturally, cancer cells in the process are categorized into types based on their biological states and cells of a type share a common growth rate and available events. Types with associated events form a network of all possible evolutionary pathways. For example, sensitive cancer cells become resistant to some drugs through evolution; Initially healthy cells, after accumulating driver mutations, can become malignant; Cancer at the primary site, can spread to a close lymph node through cellular migrations.

In many scenarios, we are mainly interested in one type, which we refer to the target type. Many important questions bear on the waiting time of the first target type cell, the time by which the first target type is produced. One typical case is: Consider the evolution of cells in an human organ. How long does it take for the first cancerous cell to occur? Given there are multiple sequences of driver mutations that lead to the malignant type, which sequence is the most probable pathway? One viable method to formulate these two questions is though defining a multi-type branching process on a graph, in which the population of each type is modeled as a stochastic process and the transition network is represented by the graph. Numerous aspects of cancer evolution have been modeled, including initiation [34, 30], progression [14, 7, 35, 9] , metastasis [18, 5, 13], and resistance to therapy [28, 27, 10, 8, 32].

The outline of this thesis is as follows. In Chapter 2 we construct the continuous-time branching process and present the methodology to model cancer evolution. We introduce the mean matrix, by which the dynamics of a branching process model can be determined. We show that to analyze such a branching process, one first decomposes its mean value matrix into separate blocks and analyze the dynamics in each block. We present a uniform classification of different dynamics corresponding to different kinds of blocks. Lastly, we analyze the case when the graph only contains a single pathway and approximate the waiting time distributions of each type. In Chapter 3 we discuss colorectal cancer initiation on a single evolutionary path. We apply the techniques from Chapter 2 to identify the large time behavior and the waiting time for each type. Analytic waiting time distributions, including the distribution of waiting time to malignancy, are in good agreement with results from Monte Carlo simulations of the model.

Some parts of Chapter 2 follows [4, 25, 22]. Sections 2.4 and Chapter 3 largely follow a recent preprint [43] which is the joint work by me, Obinna A. Okogu and Ivana Bozic.

Chapter 2

BRANCHING PROCESS MODELS OF CANCER EVOLUTION

Branching process models originates from the famous Galton-Walton model, which was used to investigated the inheritance of surnames among European nobles. For a rich history and theory of branching processes, one can refer to [24] and [4]. Following the theoretical success of Luria and Delbruck's model of drug resistance, branching process model shows its great potential in analyzing biological evolution.

2.1 Construction of branching processes

The name 'branching process' represents a broad class of different stochastic processes. A branching process can be constructed in the following parameter setting. There are q types of individuals. Each individual splits after living a certain amount of time, which is a random variable that we referred to as the living time. Living times of all individuals are independent. Within a same type, living times are identically distributed according to a random variable T_i. When a type-i individual splits, it is removed from the population and produces ξ_{ij} type-j individuals, in which ξ_{ij} is a non-negative integer-valued random variable. In other words, the splitting event of one individual is not allowed to cause deaths of any other individuals. We require ξ_{ij} to have a finite second moment, i.e. $\mathbb{E}[\xi_{ij}^2] < \infty$. The distribution of the random vector $\xi_i = (\xi_{i1}, \xi_{i2}, \cdots, \xi_{iq})$ is referred as the type-i offspring distribution.

In this thesis, we introduce three branching process related models that share same $\xi_{ij}, i, j \in \{1, 2, \cdots, q\}$ while have different living times.

2.1.1 Continuous time multi-type branching process

Let $N_i(t)$ be the number of type-i individuals at time $t \in \mathbb{R}, t > 0$. We define the population vector as $N(t) = (N_0(t), N_1(t), \cdots, N_q(t))$. Each type-$i$ is associated with a weight $a_i > 0$. We call $N(t)$ a **continuous time multi-type branching process**, if the waiting time for a type-i individual to split is exponentially distributed with parameter a_i, i.e. $T_i \sim \text{Exponential}(a_i)$. And when a type-i individual splits, the development of population follows ξ_i. $N(t)$ is defined to be right-continuous. Thus if an individual splits at time t, we have

$$\Delta N(t) = N(t) - N(t^-) \sim \xi_i = (\xi_{i1}.\xi_{i2}, \cdots, \xi_{iq}).$$

Here we note that since T_i is a continuous random variable, even the event of two individuals splitting at a same time is possible, the probability of that event is zero.

2.1.2 Multi-type Galton-Watson process

Let $Z_i(n)$ be the number of type-i individuals at discrete time n. n could be non-negative integers and initially $n = 0$. We define $Z(n) = (Z_0(n), Z_1(n), \cdots, Z_q(n))$. Then $Z(n)$ is called a **multi-type Galton-Watson process** if

$$Z_j(n+1) = \sum_{i=1}^{q} \sum_{k=1}^{Z_i(n)} \xi_{ij}^{(k)},$$

where $\xi_{ij}^{(k)}, k \geq 1$ are i.i.d copies of ξ_{ij}. This is equivalent to saying that each individual lives one unit of time and dies, i.e. $T_i = 1$ for all i. When a type-i dies, it splits into new individuals corresponding to ξ_i.

Let $m_{ij}^{(Z)} := \mathbb{E}[Z_j(1)|Z(0) = e_i] = \mathbb{E}[\xi_{ij}]$, where $e_i = (\delta_{i1}, \delta_{i2}, \cdots, \delta_{iq})$ and

$$\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases}$$

The mean matrix of $Z(n)$ is defined by $M_Z := (m_{ji}^{(Z)})_{q \times q}$. The subscript Z of M_Z (and superscript (Z) of $m_{ij}^{(Z)}$) is used for identifying the Galton-Watson process and distinguishing it from other mean matrices.

The construction of the multi-type Galton-Watson model is mainly borrowed from the settings in [25] and Chapter V of [4].

2.1.3 Generalized Pólya urn process

Let $Y_i(n)$ be the number of type-i individuals at discrete time $n \geq 0$. The population vector is defined as $Y(n) = (Y_0(n), Y_1(n), \cdots, Y_q(n))$. Each type-$i$ is associated with a weight $a_i \geq 0$. At each time step, only one individual dies and splits into new individuals. We say $Y(n)$ is a **Generalized Pólya urn process** if at time step $n \geq 1$, the probability of a type-i individual to split is $a_i Y_i(n-1)/\sum_j Y_j(n-1)$. Then the population changes according to the offspring distribution of the selected type, that is

$$\Delta Y(n) = Y(n) - Y(n-1) \sim (\xi_{i1}, \xi_{i2}, \cdots, \xi_{iq}) \text{ if a type-}i \text{ individual dies.}$$

We note here the living time T_i can not be simply identified since it depends on the population and the discrete time.

Let $m_{ij}^{(Y)} := a_i(\mathbb{E}[\xi_{ij}] - \delta_{ij})$. The mean matrix of $Y(n)$ is defined by $M_Y := (m_{ji}^{(Y)})$. The notations and description of a generalized Pólya urn process follows the setting in [22].

2.1.4 Connection between discrete time results and continuous time results

In this section we want to figure out how to get results of the continuous process $N(t)$ by the results of discrete processes $Z(n)$ and $Y(n)$. To transfer from a discrete time model to a continuous model, one only need to set the lifetime of each individual to an exponentially distributed random variable. Conversely, we can reduce the continuous process $N(t)$ to a discrete process by considering $N(n\delta), n \geq 0$ for fixed $\delta > 0$. Then one can show $N(n\delta)$ is actually a Galton-Watson process (similar to [4] Chapter III. Section 6. Theorem 1).

Theorem 1. *For every $\delta > 0$ the sequence*

$$Z^{(\delta)}(n) := N(n\delta)$$

is a Galton-Watson process.

The idea of applying this theorem is to first obtain a result about $N(n\delta)$ for each δ, and then go from this to the whole process $N(t)$. The translation of discrete limit theorems to continuous limit theorems relies on the following lemma (similar to [4] Chapter III. Section 7. Lemma 1).

Lemma 2. *Let $h(\cdot)$ be a continuous function on $(0,\infty)$. If $\lim_{n\to\infty} h(n\delta) = c(\delta)$ exists for every $\delta > 0$, then $\lim_{t\to\infty} h(t) = c$ exists, and thus necessarily $c(\delta) \equiv c$ for all $\delta > 0$.*

The connection between a continuous branching process and its embedded Pólya urn process is more straight forward. Let τ_n be the n-th splitting time (the time of the n-th death event) of $N(t)$, then by [4] Chapter V, Section 9.2, Theorem 2 or [22], one have the following theorem.

Theorem 3. *The stochastic processes $Y(n)$ and $N(\tau_n)$ are equivalent.*

This theorem indicates that the stopping process of a continuous branching process is its embedded Pólya urn process.

2.2 Cancer evolution models

We are mostly concerned about cancer evolution. In most scenarios, the evolution of cancer is modeled as a continuous time birth-death process, which could be treated as a special case of continuous time multi-type branching process. In such a multi-type birth-death process, the number of offspring cannot be greater than two. The type-i offspring distribution can be identified by b_i, the birth rate, d_i, the death rate, and u_{ij}, the transition rate to type-j. Another auxiliary notation is the type-i net growth rate, $\lambda_i := b_i - d_i$.

To adapt models in Section 2.1, we specify the weight and the offspring distribution of each type. The weight of each type, a_i, can be computed by

$$a_i = b_i + d_i + \sum_{j\neq i} u_{ij}.$$

The probability mass function of ξ_i is

$$\begin{aligned}\mathbb{P}(\xi_i = \mathbf{0}) &= \frac{d_i}{a_i},\\ \mathbb{P}(\xi_i = e_i + e_j) &= \frac{u_{ij}}{a_i} \quad (j \neq i),\\ \mathbb{P}(\xi_i = 2e_i) &= \frac{b_i}{a_i},\end{aligned}$$

where $\mathbf{0}$ is a zero vector. Thus, the expected value of ξ_i is

$$\begin{aligned}\mathbb{E}[\xi_i] &= \frac{d_i}{a_i}\mathbf{0} + \sum_{j\neq i}\frac{u_{ij}}{a_i}(e_i + e_j) + \frac{b_i}{a_i}2e_i\\ &= \frac{1}{a_i}(u_{i1}, \cdots, u_{i,i-1}, 2b_i + \sum_{j\neq i} u_{ij}, u_{i,i+1}, \cdots, u_{iq}).\end{aligned}$$

And notice that

$$a_i\mathbb{E}[\xi_i - e_i] = (u_{i1}, \cdots, u_{i,i-1}, b_i - d_i, u_{i,i+1}, \cdots, u_{iq})$$

Based on the outcomes of ξ_i, we can classify all events into three kinds. We call $\xi_i = 2e_i$ a birth event, $\xi_i =$ a death event, and $\xi = e_i + e_j$ a transition event. Here, the probability of transition events does not relate to the birth rate. One might model that transitions (e.g. mutations) occur in birth events (e.g. cell divisions). This simply results in replacing u_{ij} with $b_i u_{ij}$. Additionally, in a transition event from type-i to type-j, there are always two daughter individuals including a type-i individual and a type-j individual. Let (i) be a single type-i individual, than the evolution scheme (summary of all possible evolutionary events) is

$$(i) \rightarrow \begin{cases}(i)\,(i), & \text{birth at rate } b_i\\ (i),(j), & \text{transition to type-}j \text{ at rate } u_{ij}\\ \varnothing, & \text{death at rate } d_i.\end{cases}.$$

Suppose the birth-death process is embedded in a graph $G = (V, E)$, in which $V = \{1, 2, \cdots, q\}$ is the set of vertices and $E = \{(i, j)|u_{ij} \neq 0\}$ are edges. The edges represent probable transition from one type to another. We can naturally define that $u_{ij} = 0$ if $(i, j) \notin E$.

Commonly, we have the following two assumptions:

(S1) $\lambda_i \gg \sum_{j\neq i} u_{ij}$.

(S2) G is acyclic.

The first assumption comes from small mutation rates, which holds for a numerous scenarios. However, it fails to hold when there exists i such that $\lambda_i = 0$. For instance, in the CRC initiation model [34], type-0 and type-1 crypts do not follow **(S1)**. The second assumption is assuming that there are no backward flows. This is true for a majority of models in which transitions represent mutations. Since when the cell loses a copy of a certain gene, it's nearly impossible to get that copy back. However, **(S2)** fails to hold if transitions include migrations. For instance, a metastasis model which considers cancer cells that move between different sites (e.g. surfaces of organs, lymph nodes) may have a graph that contains a loop. We will see that either **(S1)** or **(S2)** leads to a situation that is easier to analyze.

In section 2.1, we've defined two mean value matrices, M_Z for the discrete Galton-Watson model and M_Y for the Pólya urn model. In terms of a birth-death model, they can be easily computed by

$$
\begin{aligned}
M_Z &= (m_{ji}^{(Z)})_{q\times q} = (\mathbb{E}[\xi_{ji}])_{q\times q} \\
&= \begin{pmatrix}
\frac{1}{a_1}(2b_1 + \sum_{j\neq 1} u_{1j}) & \frac{1}{a_2}u_{21} & \cdots & \frac{1}{a_q}u_{q1} \\
\frac{1}{a_1}u_{12} & \frac{1}{a_2}(2b_2 + \sum_{j\neq 2} u_{2j}) & \cdots & \frac{1}{a_q}u_{q2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{a_1}u_{1q} & \frac{1}{a_2}u_{2q} & \cdots & \frac{1}{a_q}(2b_q + \sum_{j\neq q} u_{qj})
\end{pmatrix}, \\
M_Y &= (m_{ji}^{(Y)})_{q\times q} = (a_j\mathbb{E}[\xi_{ji} - \delta_{ji}])_{q\times q} \\
&= \begin{pmatrix}
\lambda_1 & u_{21} & \cdots & u_{q1} \\
u_{12} & \lambda_2 & \cdots & u_{q2} \\
\vdots & \vdots & \ddots & \vdots \\
u_{1q} & u_{2q} & \cdots & \lambda_q
\end{pmatrix}.
\end{aligned}
$$

Then we note that our assumptions, **(S1)** and **(S2)**, bring us additional structural information of M_Z and M_Y.

Proposition 4. *Assumption **(S1)** (small mutation rates) implies that M_Z and M_Y are diagonalizable. Assumption **(S2)** (acyclic graph) implies that M_Z and M_Y, after a reorder of types, are lower-triangular.*

Given **(S1)**, Gershgorin circle theorem implies that the mean matrix is diagonalizable. Applying the topological sorting algorithm on the type set provides triangular mean matrices. For a detailed proof, one can refer to [32].

2.2.1 Connections between the mean value matrices in caner modeling

It seem's more natural to define M_Z and M_Y by their transposes, M_Z' and M_Y'. In fact in a majority of work by Kesten, Stigum, Atherya and Ney [25, 4], the mean matrices are indeed defined by M_Z'. Here we choose to be in line with Janson and Nicolson [22, 32] because their papers are more recent, and $Z(n), Y(n)$ could be treated as column vectors.

Eigenvalues and eigenvectors of the mean value matrix play an important. Following [22, 32] we are going to deal with M_Z,M_Y and their right eigenvectors, while we need to think about M_Z' and its left eigenvectors when reading [25, 4].

Although results are derived from M_Z and M_Y separately. We note that results that comes from one of them might be transferable to the other when some conditions hold. Let $\{\rho_i^{(Z)}\}_{i=1}^q$, $\{\rho_i^{(Y)}\}_{i=1}^q$ be the eigenvalues of matrix M_Z, M_Y respectively. In addition, let each of $\rho_1^{(Z)}$ and $\rho_1^{(Y)}$ be the eigenvalues that has largest magnitude. Notice that

$$(M_Z - I)\begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_q \end{pmatrix} = M_Y. \tag{2.1}$$

This implies the following proposition.

Proposition 5. *Suppose one of M_Z and M_Y is lower-triangular, the other also has to be lower-*

triangular. In addition

$$\rho_1^{(Z)} > 1 \Leftrightarrow \rho_1^{(Y)} > 0,$$
$$\rho_1^{(Z)} = 1 \Leftrightarrow \rho_1^{(Y)} = 0, \textit{ and}$$
$$\rho_1^{(Z)} < 1 \Leftrightarrow \rho_1^{(Y)} < 0,$$

where $p \Leftrightarrow q$ means p is equivalent to q.

Proof. We will prove the case when M_Z is lower-triangular. This implies that when $i > j$, $m_{ij}^{(Z)} = \mathbb{E}[\xi_{ji}] = 0$. Thus for $i > j$, $m_{ij}^{(Y)} = \frac{1}{a_j}\mathbb{E}[\xi_{ji}] = 0$. Hence M_Y is also lower-triangular. For a triangular matrix, we know that its eigenvalues are exactly the entries on the diagonal. Thus the eigenvalues of M_Z are $\{\frac{1}{a_i}(2b_i + \sum_{j\neq i} u_{ij})\}_{i=1}^q$ and the eigenvalues of M_Y are $\{\lambda_i\}_{i=1}^q$. By the definition of λ_i, we have

$$\lambda_i = b_i - d_i = a_i\left(\frac{1}{a_i}(2b_i + \sum_{j\neq i} u_{ij}) - 1\right)$$

Thus the statement holds true. □

We can see that some properties of M_Z and M_Y are similar. However, one should also notice that the eigenvectors of M_Z and M_Y are different. Thus it is still necessary to keep both notations.

2.3 Model classification and limit theorems

Evolutionary dynamics are closely related to the relative fitness advantages (change of growth rates) conferred by transitions. With strictly increasing clonal growth rates after mutation, Durrett and Moseley [14] modeled a pathway from healthy cells to cancerous cells, and computed distributions for clonal sizes and waiting times. Nicholson and Antal [32] studied a general framework wherein wild-type individuals have the largest fitness (growth rate), which could be applied to cases involving drug resistance. Random fitness advantages have also been investigated [19]. Based on these studies, we classify models by their mean value matrices. We will the see the general idea of dealing with each case with emphasis on obtaining limit theorems.

In the beginning we introduce the concept of irreducible matrices. We say type-i communicates with type-j, written $i \succ j$, if $(M_Y^n)_{ji} > 0$ for some $n \geq 0$. This means that starting from a single type-i individual, the probability of getting a type-j individual is nonzero after a few steps. The relation $\succ$ is transitive and reflective, i.e.

$$i \succ j, j \succ k \Rightarrow i \succ k$$
$$i \succ i \text{ for all } i \in V = \{1, 2, \cdots, q\}.$$

Thus we can classify all types (or all vertices V in the graph) into equivalence classes $\mathscr{C}_1, \mathscr{C}_2, \cdots, \mathscr{C}_h$ such that type-i and type-j belong to the same class if and only if $i \succ j$ and $j \succ i$ (in this case we also say type-i intercommunicates type-j). In addition, we say type-i is dominating if it communicates all other types, i.e. $i \succ j, \forall j \in V$. And observe that if a dominating type belongs to class $\mathscr{C}_k$, then all types in $\mathscr{C}_k$ are dominating. Hence we call class $\mathscr{C}_k$ a dominating class if there exists $i \in \mathscr{C}_k$ such that i in dominating.

A Galton-Watson (or M_Z),a Pólya urn process (or M_Y), or a continuous time branching process is irreducible if there is only one equivalent class. In other words, a branching process is irreducible if all types are dominating.

One should note that M_Z and M_Y have to be both irreducible or both reducible and they share same equivalent classes.

Proposition 6. *M_Z is irreducible is equivalent to saying M_Y is irreducible. M_Z and M_Y have same equivalent classes $\{\mathscr{C}_1, \mathscr{C}_2, \cdots, \mathscr{C}_h\}$.*

Thus, when classifying models, we choose to derive all results base on the behaviors of M_Y. However, there are some subtleties between results derived from a urn process and a Galton-Watson process. Hence in some limit theorems still involve M_Z.

2.3.1 Irreducible processes and some related cases

We are going to summarize the results of [22]. Janson et al. which studied a general case including M_Y is irreducible. Their assumptions are:

(A1) The largest eigenvalue ρ_1^Y of M_Y is positive, $\rho_1^Y > 0$.

(A2) ρ_1^Y is simple.

(A3) There exists a dominating type i and the process starts with at least one type-i individual.

(A4) ρ_1^Y belongs to the dominating class.

This guarantees that left and right eigenvectors of M_Y that corresponding to the largest eigenvalue ρ_1^Y are non-negative. We define v_1 to be the right eigenvector corresponding to ρ_1^Y. Let $a = (a_1, a_2, \cdots, a_q)$ be the weight vector. Then we can normalize v_1 by letting

$$a \cdot v_1 = 1.$$

One of their results that can be easily applied (Part of Theorem 3.8 in [22]) is stated below.

Theorem 7. *As $t \to \infty$, $e^{-\rho_1^{(Y)} t} N(t) \xrightarrow{a.s.} W v_1$, where W is a non-negative random variable.*

This indicates that the population is going to be drived by the dominating type if the dominating type has the largest eigenvalue and the process starts at that dominating type.

2.3.2 Reducible processes

When M_Y is reducible or when the process initially does not have a dominating individual, we rely on the analysis in [25]. Kesten et al. and mainly focus on M_Z instead of M_Y. So, we will follow their convention in this section and state results for M_Z. We begin with a theorem for reducible matrices.

Theorem 8. *When M_Z is reducible, the equivalent classes $\{\mathscr{C}_1, \mathscr{C}_2, \cdots, \mathscr{C}_h\}$ can be partially ordered so that the rows and columns of M_Z are be reordered and M_Z takes the form*

$$\begin{pmatrix} M(1) & & & 0 \\ M(1,2) & M(2) & & \\ \vdots & \vdots & & \\ M(1,h) & M(2,h) & \cdots & M(h) \end{pmatrix}, \tag{2.2}$$

where $M(a,b) = (m_{ij}^{(Z)})_{i\in\mathscr{C}_a, j\in\mathscr{C}_b}$ *and* $M(a) = M^{(Z)}(a,a)$*. Then if* $M(a) \neq 0$*, the subprocesses*

$$Z^a(n) = \{Z_i(n) | i \in \mathscr{C}_a\}, n \geq 0$$

are an irreducible processes.

Let ρ_a be the largest eigenvalue of block $M(a)$. We will have the following assumptions:

(B1) For all a, $M(a)$ is aperiodic.

(B2) There exists at least one block $M(a)$ that has largest eigenvalue $\rho_a > 1$.

(B3) There exists a dominating type and the process starts with at least one dominating type individual.

The strategy given by Kesten et al. is to deal each block by theorems for irreducible cases and figure out relationships between blocks. Without loss of generality, we can consider subprocess $(Z^1(n), Z^2(n))$ with the mean value matrix

$$\begin{pmatrix} M(1) & 0 \\ M(1,2) & M(2) \end{pmatrix}.$$

We have a few local assumptions:

(b1) $M(1,2) \neq 0$.

(b2) The initial individual (recall that $Z(0) = e_i$) is not of class $\mathscr{C}_2$.

(b3) $\rho_1, \rho_2 > 1$.

(b1) is just saying that a $\mathscr{C}_1$ individual is able to transition to $\mathscr{C}_2$ individual. This might happen between two vertices that on two separate pathways. **(b2)** guarantees that the populations of $\mathscr{C}_1$ will not always be 0. **(b3)** assumes that the process is supercritical, thus the process is not always going to extinct. We note that **(b3)** is a really strong condition and we are not quite sure

what's going to happen if this assumption fails to hold. Based on **(b3)** and Theorem 7, one can see that if the initial individual is of type-1, then

$$N(t)e^{-\rho_1 t} \to W v_1.$$

where v_1 is the left eigenvector of $M(1)$ corresponding to eigenvalue ρ_1.

Case 1

Suppose the initial individual is of type-1 and $\rho_1 > \rho_2$. In fact, from subsection 2.3.1, we can see that the growth of type-1 individuals is going to drive the growth of type-2 individuals. i.e

$$\frac{Z^{(2)}(n)}{\rho_1^n} \overset{w.p.1}{\to} w$$

where w is a non-negative random variable. The result can be able to transition to the continuous case.

Theorem 9.

$$N^{(2)}(t)e^{-\hat{\rho}_1 t} \to W v_2$$

where $N^{(2)}(t) = \{N_i(t) | i \in \mathscr{C}_2\}$, W is a non-negative random variable, and v_2 is a fixed vector.

Thus, we can treat the entire matrix $\begin{pmatrix} M(1) & 0 \\ M(1,2) & M(2) \end{pmatrix}$ as a Block with largest eigenvalue ρ_1. The whole proof of Theorem 9 can be found in [22].

Case 2

Suppose that the initial individual is of type-1 and $\rho_1 = \rho_2$. This case is discussed in Theorem 2.3. in [25]. For the discrete Galton-Watson process, we have

$$\frac{Z^{(2)}(n)}{n\rho_1^n} \overset{w.p.1}{\to} \left(\frac{W}{\rho_1} v_1^T M(1,2) u_2\right) v_2 \tag{2.3}$$

where v_i is the normalized left eigenvector of $M(i)$ which corresponds to eigenvalue ρ_i, and u_2 is the right eigenvector of $M(2)$ corresponding to ρ_2. And the corresponding theorem in continuous time is

Theorem 10.

$$N^{(2)}(t)\,t^{-1}e^{-\hat{\rho}_1 t} \to W v_2$$

where W is a random variable, $\hat{\rho}_1$ is the largest eigenvalue of the mean value matrix of $N^{(1)}(t)$, and v_2 is a fixed vector.

Proof. By Theorem 1, for any $\delta > 0$, process $N^{(2)}(n\delta)$ is a Galton-Watson process with mean value matrix

$$M^{(\delta)} = e^{M^{(Y)}\delta}$$

in which $M^{(Y)}$ is the mean value matrix of the process $N^{(2)}(t)$. Suppose the largest eigenvalue of $M^{(Y)}$ is $\hat{\rho}_2$. In this case, we have $\hat{\rho}_1 = \hat{\rho}_2$. By the property of matrix exponential, the largest eigenvalue of $M^{(\delta)}$ is $e^{\hat{\rho}_1\delta}$. Then following (2.3), there exists a fixed vector v such that

$$\frac{N^{(2)}(n\delta)}{ne^{\hat{\rho}_1 n\delta}} \overset{w.p.1}{\to} W_\delta v$$

This implies

$$\mathbb{P}(N^{(2)}(n\delta)\delta n e^{-\hat{\rho}_1 n\delta} \le \mathbf{x}) \to \mathbb{P}(\delta W_\delta v \le \mathbf{x}) \quad (n \to \infty)$$

for any non-negative vector $\mathbf{x}$. Finally, by Lemma 2, we have

$$\mathbb{P}(N^{(2)}(t)\,t^{-1}e^{-\hat{\rho}_1 t} \le \mathbf{x}) \to \mathbb{P}(W v_2 \le \mathbf{x}) \quad (t \to \infty)$$

where v_2 is a fixed vector. Then we conclude that

$$N^{(2)}(t)\,t^{-1}e^{-\hat{\rho}_1 t} \to W v_2 \quad \text{in distribution}$$

□

Case 3

Suppose that the initial individual is of type-1 and $\rho_1 < \rho_2$. This case is similar to the one considered by Durrett [14]. Ketsen et al. shows that (their original theorem is more complex, see Theorem 2.2. of [25])

$$\frac{Z^{(2)}(n)}{\rho_2^n} \overset{w.p.1}{\to} V \tag{2.4}$$

where V is a non-negative random vector in which all entries are independent. And our conjecture would be

Theorem 11.

$$N^{(2)}(t)e^{-\hat{\rho}_2 t} \to V$$

where V is a random vector in which all entries are independent.

Proof. By Theorem 1, for any $\delta > 0$, process $N^{(2)}(n\delta)$ is a Galton-Watson process with mean value matrix

$$M^{(\delta)} = e^{M^{(Y)}\delta}$$

in which $M^{(Y)}$ is the mean value matrix of the process $N^{(2)}(t)$. Suppose the largest eigenvalue of $M^{(Y)}$ is $\hat{\rho}_2$. By the property of matrix exponential, the largest eigenvalue of $M^{(\delta)}$ is $e^{\hat{\rho}_2\delta}$. Then following (2.4), we have

$$\frac{N^{(2)}(n\delta)}{e^{\hat{\rho}_2 n\delta}} \overset{w.p.1}{\to} V_\delta$$

This implies

$$\mathbb{P}(N^{(2)}(n\delta)e^{-\hat{\rho}_2 n\delta} \le \mathbf{x}) \to \mathbb{P}(V_\delta \le \mathbf{x}) \quad (n \to \infty)$$

for any $x \ge 0$. Finally, by Lemma 2, there exists a random vector V (indpendent of δ) such that

$$\mathbb{P}(N^{(2)}(t)e^{-\hat{\rho}_2 t} \le \mathbf{x}) \to \mathbb{P}(V \le \mathbf{x}) \quad (t \to \infty)$$

that is

$$N^{(2)}(t)e^{-\hat{\rho}_2 t} \to V \quad \text{in distribution}$$

□

We note that the main non-triviality arose by increasing eigenvalue (or increasing fitness) is that the limit consists q random variables, while in Case 1 and Case 2, the limit is a scalar random variable W multiplied by a fixed vector.

Case 4

Suppose that the initial individual is not in $\mathscr{C}_1$ or $\mathscr{C}_2$. Then Ketsen et al. introduce random variable $T_m^{p,q}$ which represents the number of descendants of type-q in the m-th generation of type-p in the $(m-1)$-th generation. Then, vector $U_r^q(m,n,a,p)$ is defined to be the number of immigrants from type-p to type-q that originated from a ancestor being the r-th individual among $T_m^{p,q}$ descendants. Then we have

$$Z^{(a)}(n) = \sum_{m=1}^{n} \sum_{q\in\mathscr{C}_a} \sum_{1\le b<a} \sum_{p\in\mathscr{C}_b} \sum_{r=1}^{T_m^{p,q}} U_r^q(m,n,a,p) \tag{2.5}$$

This is a quite tedious formula, but it clearly shows all possible transitions that could happen for between classes. They argue that all features needed to study the above formula can be covered by all involved $(Z^{(a)}(n), Z^{(b)}(n))$ pairs.

2.4 General results for a multi-type branching process on a single pathway

In previous sections, we've showed possible large time limits of a multi-type branching process associated with a complex network. In this section, we consider the special case when the network only has one pathway from the initial type to the target type. Further more, we consider the case when the death rates are zeros. To be more specific, we consider a continuous multi-type branching process in which type-i individuals can only be produced from type-$(i-1)$ individuals at rate $u_i := u_{i-1,i}$, and each newly formed type-i individual starts a new pure birth process with growth rate $\lambda_i := b_i$. Other then large time limits, we also present and extend a common approach for computing the waiting time distribution of the first type-i individual. We note that in this case, after a possible reordering of types, the corresponding mean matrix

M_Y is bi-diagonal (starting from type-0) and reducible,

$$M_Y = \begin{pmatrix} \lambda_0 & 0 & 0 & \cdots & 0 \\ u_1 & \lambda_1 & 0 & \ddots & 0 \\ 0 & u_2 & \lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & u_q & \lambda_q \end{pmatrix}.$$

Following 2.3.2, mutation from one type to the next can be treated as a two-type subprocess. From this perspective, the process containing types $1, ..., n$ can be treated as a sequence of $n-1$ two-type branching processes, with the terminal state from one two-type branching process serving as the initial state for the next. We will show that when large time limits exist for type-i populations, it suffices to analyze two-type (initial and terminal type) branching processes $(N_i(t), N_{i+1}(t))$, leading to a unified approach. When a large time limit does not exist, we introduce an different approach (see Section 2.4.4), approximating the sum (2.5) by an integration.

In previous studies, the asymptotic population dynamics of two-type branching processes have been analyzed by decomposing the population of the initial type into the product of a deterministic, time-dependent exponential function and a time-invariant random variable [12, 13, 14]. This approximation greatly simplifies the computation of the waiting time distribution by making the stochastic component of the dynamics time-invariant. However, these studies typically consider the case of a wild type with a net growth rate λ_i strictly greater than zero. In contrast, the wild-type population in our model is not supercritical and in fact (slowly) declines. This, along with allowing $\lambda_i = \lambda_{i+1}$, causes non-exponential growth of mutant populations. In this section, we extend previous approaches [12, 13, 14] so that the time-deterministic component can belong to a class of non-exponential functions, which enables us to deal with the non-exponentially growing populations (type-0 and type-1). Furthermore, we review canonical expected value formulas that will be applied in later sections. Similarly to previous approaches [14, 13] we will assume that the mutation rates from one type to the next are small.

2.4.1 Expected value and Laplace transform

Throughout Section 2.4, we will be concerned with a multi-type branching process $(N_0(t), N_1(t), ...)_{t\geq 0}$, where $N_i(t)$ denotes the number of type-i individuals at time t. The process is started at time zero, with some positive number of type-0 individuals. Type-i individuals can divide with rate $\lambda_i \geq 0$ and mutate to type-$(i+1)$ with rate $u_i > 0$.

We start with a lemma that provides the Laplace transform of $N_i(t)$ conditional on the population of its precursor, $N_{i-1}(t)$.

Lemma 12. *Let $Z_i(t)$ be the number of type-i individuals in a pure-birth process that starts with $Z_i(0) = 1$ individuals at time $t = 0$. Then*

$$\mathbb{E}\left[e^{-\theta N_i(t)}\middle| N_{i-1}(s), s \leq t\right] = \exp\left(-u_i \int_0^t N_{i-1}(s)(1-\phi_i(\theta, t-s))\right) \tag{2.6}$$

where $\phi_i(\theta, t) := \mathbb{E}\left[e^{-\theta Z_i(t)}\right]$.

This is a well-known result for multi-type branching processes. One can prove it by following the procedure of Lemma 2 in [14], replacing the start time with $s = 0$.

The expected value formula for the number of type-i individuals can be derived from Lemma 12.

Theorem 13.

$$\mathbb{E}[N_i(t)] = u_i \int_0^t \mathbb{E}[N_{i-1}(s)] e^{\lambda_i (t-s)}\, \mathrm{d}s$$

Taking the derivative of equation (2.6) and evaluating at $\theta = 0$ gives the relationship between expected values.

2.4.2 Large time limits

Lemma 14. *Consider $(N_{i-1}(t), N_i(t))_{t\geq 0}$. Suppose $N_i(t)$ is the population of type-i crypts in the multi-type branching system where the population of type-$(i-1)$ crypts is given by $N_{i-1}(t) \geq 0$, then*

$$M(t) = e^{-\lambda_i t} N_i(t) - \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s$$

is a martingale. If

$$I_i = \int_0^\infty u_i N_{i-1}(s) e^{-\lambda_i s} ds$$

has a finite expectation, then

$$e^{-\lambda_i t} N_i(t) \overset{a.s.}{\to} W_i,\ \mathbb{E}|W_i| < \infty,$$

as $t \to \infty$. Additionally, if $e^{-\lambda_i t} N_i(t)$ is uniform integrable, then

$$e^{-\lambda_i t} N_i(t) \overset{L^1}{\to} W_i.$$

This implies

$$\mathbb{E}[e^{-\lambda_i t} N_i(t)] \to \mathbb{E}[W_i] = \mathbb{E}[I_i].$$

Proof. The proof follows that of Theorem 2 in [14]. The only difference is that we have derived a new bound on the martingale. By Lemma 1 in [14],

$$M(t) = e^{-\lambda_i t} N_i(t) - \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s$$

is a martingale. If I_i has a finite expectation, then by the martingale convergence theorem (Theorem 4.2.11 in [15]), the submartingale $X(t) = -M(t)$ converges a.s. to some integrable limit X as $t \to \infty$. Since

$$I_i(t) = \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} \mathrm{d}s \overset{a.s.}{\to} I_i,$$

we also have

$$e^{-\lambda_i t} N_i(t) \overset{a.s.}{\to} W_i.$$

The martingale starts at zero (i.e. $M(0) = 0$), which implies

$$\mathbb{E}[e^{-\lambda_i t} N_i(t)] = \mathbb{E}[I_i(t)].$$

Suppose $e^{-\lambda_i t} N_i(t)$ is uniform integrable, we have (Theorem 4.6.3 in [15])

$$e^{-\lambda_i t} N_i(t) \overset{L^1}{\to} W_i,$$

which guarantees

$$\mathbb{E}[I_i(t)] = \mathbb{E}[e^{-\lambda_i t} N_i(t)] \to \mathbb{E}[W_i].$$

Thus, we have

$$\mathbb{E}[I_i(t)] \to \mathbb{E}[I_i],$$

and

$$\mathbb{E}[W_i] = \mathbb{E}[I_i].$$

□

If the first condition from the statement of the Lemma holds (i.e. if I_i has finite expectation) Lemma 14 provides a method of obtaining the long-term behavior of N_i using the limiting random variable W_i. In that case, we have $e^{-\lambda_i t} N_i(t) \to W_i$, and for large time t, $e^{\lambda_i t} W_i$ should be a good approximation of the stochastic process $N_i(t)$. The importance of $N_i(t) \approx e^{\lambda_i t} W_i$ is that it separates a stochastic process into a deterministic function $e^{\lambda_i t}$ and a time-independent random variable W_i.

If, in addition, the second condition ($e^{-\lambda_i t} N_i(t)$ is uniform integrable) holds, then the expected value of the limiting random variable W_i is obtainable. In that case, we have $E[e^{-\lambda_i t} N_i(t)] \to E[W_i]$, which makes the large time approximation $N_i(t) \approx e^{\lambda_i t} W_i$ reasonable in terms of the first moment.

2.4.3 Estimating waiting times using large time limits

Let $\tau_i, (1 \le i \le n)$ be the waiting time of the first type-i individual in a multi-type branching process. At time $s \ge 0$, the arrival rate of type-i individuals is $u_i N_{i-1}(s)$. Conditional on the trajectory of $N_{i-1}(s)$ for $0 \le s \le t$, the probability of τ_i can be written as

$$P(\tau_i > t \mid N_{i-1}(s), 0 \le s \le t) = \exp\left(-u_i \int_0^t N_{i-1}(s)\, ds\right).$$

The functional form of $N_{i-1}(s)$ is generally unknown and potentially complicated. One way of evaluating this integral is to approximate $N_{i-1}(s)$ by the product of a deterministic time-dependent growth and a time independent random variable. For example, let $N_0(t) = Z_0(t)$, a pure birth two-type branching process that starts with a single individual. It is well-known that $e^{-\lambda_0 t} Z_0(t) \to W_0 \sim$ Exponential(1) [14]. A classical approximation is $N_0(s) \approx e^{\lambda_0 t} W_0$ where the

deterministic time-dependent growth is characterized by $e^{\lambda_0 t}$ and time independent random variable is W_0. Applying this approximation yields

$$P(\tau_1 > t) \approx \mathbb{E}\left[\exp\left(-u_i \int_0^t e^{\lambda_0 s} W_0 ds\right)\right] \tag{2.7}$$

$$= \mathscr{L}_{W_0}\left(u_1 \frac{e^{\lambda_0 t} - 1}{\lambda_0}\right), \tag{2.8}$$

where $\mathscr{L}_{W_0}(\theta) = \frac{1}{1+\theta}$ is the Laplace transform of W_0. Let $f_0(t) = e^{\lambda_0 t}$ be the time deterministic function. Then the above approximation also holds if a sub-exponential term is added to f_0. In other words, we have many reasonable options for f_0. In later sections, we consider two specific deterministic functions,

$$f_{i0}(t) = e^{\lambda_i t}, \text{ and}$$
$$f_{i1}(t) = \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[W_i]}.$$

In the example mentioned above, we have

$$f_{01}(t) = \frac{\mathbb{E}[Z_0(t)]}{\mathbb{E}[W_0]} = e^{\lambda_0 t}$$
$$= f_{00}(t).$$

However, in our model when $i > 1$, $f_{i0} \neq f_{i1}$. Later we will see an approximation with f_{i1} is more precise than that with f_{i0}.

The second observation from this example is that the approximation (2.7) relies on the Laplace transform of W_0. However, one cannot obtain a closed form for the Laplace transform of $W_i, i \geq 2$. In this case, an alternative approach is to approximate W_i by another random variable V_i that has a a closed form Laplace transform. Such V_i can be found by approximating N_{i-1}, as described in Section 3.2.

Proposition 15. *Let $(N_{i-1}(t), N_i(t))$ be a two-type process such that $N_i(0) = 0$ and type-i individuals are produced by type-$(i-1)$ individuals with rate $u_i N_{i-1}(t), u_i > 0$. Suppose that $N_{i-1}(t) \approx N^*_{i-1}(t)$ such that $(f_{i-1}(t))^{-1} N^*_{i-1}(t) \to V_{i-1}$ as $t \to \infty$. Then the waiting time distribution of the*

first type-i individual can be approximated by

$$P(\tau_i > t) \approx \mathscr{L}_{V_{i-1}}\left(u_i \int_0^t f_{i-1}(s)ds\right)$$

where $\mathscr{L}_{V_{i-1}}(\theta)$ is the Laplace transform of random variable V_{i-1}.

Applying the approximations gives us

$$\begin{aligned} P(\tau_i > t) &= \mathbb{E}\left[\exp\left(-\int_0^t u_i N_{i-1}(s)ds\right)\right] \\ &\approx \mathbb{E}\left[\exp\left(-V_{i-1}u_i \int_0^t f_{i-1}(s)ds\right)\right] \\ &= \mathscr{L}_{V_{i-1}}\left(u_i \int_0^t f_{i-1}(s)ds\right). \end{aligned}$$

From Proposition 15, we see that the waiting time (τ_i) distribution depends on the Laplace transform of V_{i-1}. Here we present an iterative method for computing the Laplace transforms of V_i.

Lemma 16. *Let $\tilde{N}_i(t)$ be type-i individuals in the system with $N^*_{i-1}(t) = f_{i-1}(t)V_{i-1}$. Suppose $e^{-\lambda_i t}\tilde{N}_i(t) \xrightarrow{a.s.} V_i$, then*

$$\mathscr{L}_{V_i}(\theta) := \mathbb{E}\left[e^{-\theta V_i}\right] = \mathscr{L}_{V_{i-1}}\left(u_i \int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}}\,\mathrm{d}s\right).$$

Proof. To get the Laplace transform of V_i, we start from the star process approximation

$$N^*_{i-1}(t) = f_{i-1}(t)V_{i-1}.$$

Applying Lemma 12 to the 2-type process $\left(N^*_{i-1}(t), \tilde{N}_i(t)\right)$ yields

$$\mathbb{E}[e^{-\theta\tilde{N}_i(t)}|V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s)V_{i-1}(1-\phi_i(\theta, t-s))\mathrm{d}s\right).$$

Replacing θ with $\theta e^{-\lambda_i t}$,

$$\mathbb{E}[e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s)V_{i-1}(1-\phi_i(e^{-\lambda_i t}\theta, t-s))\mathrm{d}s\right), \tag{2.9}$$

then for each subclone of N_i, by equation (2.10), we have

$$Z_i(t-s)e^{-\lambda_i(t-s)} \to \text{Exponential}(1)\ a.s.$$

Thus, it follows

$$Z_i(t-s)e^{-\lambda_i t} \to \text{Exponential}(e^{\lambda_i s}).$$

Now consider the following limit involving terms on the right hand side of (2.9),

$$\begin{aligned}\lim_{t\to\infty} 1-\phi_i(\theta e^{-\lambda_i t}, t-s) &= 1-\int_0^\infty e^{-\theta x}e^{\lambda_i s}\exp\Big(-xe^{\lambda_i s}\Big)\mathrm{d}x \\ &= \int_0^\infty (1-e^{-\theta x})e^{\lambda_i s}\exp\Big(-xe^{\lambda_i s}\Big)\mathrm{d}x \\ &= \frac{\theta}{\theta+e^{\lambda_i s}}.\end{aligned}$$

And note that as $t\to\infty$, the left hand side of (2.9) yields

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}[e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] &= \mathbb{E}[\lim_{t\to\infty} e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] \\ &= \mathbb{E}[e^{-\theta V_i}|V_{i-1}],\end{aligned}$$

in which switching the limit and the integration is allowed by the dominated convergence theorem. Thus, we can write

$$\mathbb{E}[e^{-\theta V_i}|V_{i-1}] = \exp\left(-u_i V_{i-1}\int_0^\infty \frac{\theta f_{i-1}(s)}{\theta+e^{\lambda_i s}}\mathrm{d}s\right).$$

Taking expectation on both sides gives us

$$\mathbb{E}[e^{-\theta V_i}] = \mathscr{L}_{V_{i-1}}\left(u_i\int_0^\infty \frac{\theta f_{i-1}(s)}{\theta+e^{\lambda_i s}}\,\mathrm{d}s\right).$$

□

2.4.4 An inhomogeneous Poisson process approximation

Proposition 15 estimates the arrival time of type-i individuals by the large time limit of the previous type (V_{i-1}). However, there are situations when the random variable V_{i-1} does not exist or when $f_{i-1}V_{i-1}$ fails to provide a precise first arrival time of N_i. More generally, we may only have a good large time limit V_j for some type-j where $j<i$, so that we do not have reliable limits for type-$(j+1)$ through type-$(i-1)$. To deal with this situation, we introduce a method that uses the large time limit of each independent lineage to 'skip' V_{j+1} through V_{i-1}. First, note

that given V_j, we can approximate the birth rate of type-$(j+1)$ lineages, using the well-known fact that for each type-$(j+1)$ lineage,

$$e^{-\lambda_{j+1}t} Z_{j+1}(t) \xrightarrow{a.s.} V \sim \text{Exponential}(1). \tag{2.10}$$

Next, we want to find the likelihood of each type-$(j+1)$ lineage producing at least a single type-i individual. Recall that a type-$(j+1)$ lineage is a simple birth process initiated by a single type-$(j+1)$ individual that grows at rate λ_{j+1}. Suppose we have a type-$(j+1)$ lineage that was started by a single type-$(j+1)$ individual at time s. We define $p_0^{(j+1)\to i}(s,t)$ to be the probability that no type-i individual is produced by time t by this type-$(j+1)$ lineage.

At time s, type-$(j+1)$ crypts are producing at rate $u_{j+1}N_j(s) \approx u_{j+1}f_j(s)V_j$. Each type-$(j+1)$ lineage (present at time s) has a probability $1-p_0^{(j+1)\to i}(s,t)$ to produce at least a single type-i crypt (at time t). Finally, for fixed t, we approximate the process of producing type-i individuals as an inhomogeneous Possion process with rate $u_{j+1}N_j(s)(1-p_0^{(j+1)\to i}(s,t))$. This implies

$$\begin{aligned} P(\tau_i > t) &\approx \mathbb{E}\left[\exp\left(-V_j u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right)\right] \\ &= \mathscr{L}_{V_j}\left(u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right). \end{aligned}$$

This approach is summarized in the following proposition.

Proposition 17. *Let $(N_j(t), N_{j+1}(t), \cdots, N_i(t))$ be a $(i-j+1)$-type process such that $N_k(0)=0$, $j<k\le i$ and type-k individuals are produced by type-$(k-1)$ individuals with rate $u_kN_{k-1}(t)$, $u_k>0$. Suppose that $N_j(t)\approx N_j^*(t)$ such that $(f_j(t))^{-1}N_j^*(t)\to V_j$. Then the waiting time distribution of type-i crypts can be approximated by*

$$P(\tau_i > t) \approx \mathscr{L}_{V_j}\left(u_{j+1}\int_0^t f_j(s)\left(1-p_0^{(j+1)\to i}(s,t)\right)ds\right),$$

where $\mathscr{L}_{V_j}(\theta)$ is the Laplace transform of V_j.

Proposition 15 is consistent with Proposition 17, and one can treat Proposition 15 as a single-step version of Proposition 17.

To compute $p_0^{j\to i}(s,t)$, $j<i$, we use the iterative relationship between $p_0^{j\to i}(s,t)$ and $p_0^{(j+1)\to i}(s,t)$. This is provided by the following proposition.

Lemma 18. *For $i < j$,*

$$p_0^{j\to i}(s,t) \approx \int_0^t e^{-v} \exp\left(-v\int_s^t u_{j+1} e^{\lambda_j (r-s)} \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right) dv.$$

with $p_0^{i\to i}(s,t) = 0$.

Proof. For a type-j lineage that appears at time s, we approximate it by its large time limit. At time $r > s$, the population of this lineage would be $Z_j(r) \approx e^{\lambda_j (r-s)} V$, where $V \sim$ Exponential(1). Thus, type-$(j+1)$ individuals that mutated from this lineage are produced at rate

$$u_{j+1} Z_j(r) \approx u_{j+1} e^{\lambda_j (r-s)} V.$$

The probability for a type-$(j+1)$ individual to produce at least a type-i individual is $1 - p_0^{(j+1)\to i}(r,t)$. Thus, conditional on V, the expected number of type-i individuals that were produced by a type-j lineage that appeared at time s is

$$\begin{aligned}\Lambda^{j\to i}(s,t,V) &= \int_s^t u_{j+1} Z_j(r) \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r \\ &\approx \int_s^t u_{j+1} e^{\lambda_j (r-s)} V \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r.\end{aligned}$$

Let $X^{j\to i}(s,t)$ be the number of type-i individuals that are produced by a type-j subclone appeared at s. In the time period $[s,t]$, $X^{j\to i}(\cdot,t)$ follows an inhomogeneous Possion process with mean $\Lambda(\cdot,t,V)$. Thus the probability that no type-i crypt is produced from this particular type-$(j+1)$ crypt by time t is

$$\begin{aligned}P(X^{j\to i}(s,t) = 0 | V) &= \exp\left(-\Lambda^{j\to i}(s,t,V)\right) \\ &= \exp\left(-V\int_s^t u_{j+1} e^{\lambda_j (r-s)} \left(1 - p_0^{(j+1)\to i}(r,t)\right) \mathrm{d}r\right).\end{aligned}$$

This implies

$$\begin{aligned}p_0^{j\to i}(r,t) &:= P(X^{j\to i}(s,t) = 0) \\ &= \mathbb{E}\left[P(X^{j\to i}(s,t) = 0 | V)\right] \\ &= \int_0^t e^{-v} \exp\left(-v\int_s^t u_{j+1} e^{\lambda_j (r-s)} \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right) dv.\end{aligned}$$

Finally, for $i = j$, notice that the founding individual of a type-j lineage is of type-j. This guarantees that at any time t greater than the founding time s, the probability of having at least one type-j individual is 1. Thus $p_0^{j\to j}(s,t) = 0$. □

Chapter 3

WAITING TIMES OF COLORECTAL CANCER INITIATION ON A SINGLE PATHWAY

Cancer initiation is the biological process in which healthy wild-type cells mutate into malignant cancer cells, prior to uncontrolled proliferation. Following the seminal work by Armitage and Doll in the 1950s [2, 3], and Knudson [26] in the 1970s, cancer initiation has been identified as a multi-step process, whereby cells sequentially accumulate driver mutations required for malignant transformation. These works investigated age-dependent incidence rates for human cancers and argued that a multi-step model best explains the data across different cancers. Since the advent of genome sequencing, the molecular mechanisms of many cancer driver mutations have been described [42, 21, 41, 11, 33].

Colorectal cancer (CRC) is one of the most common cancers in the United States [37]. It has been shown that mutations in just three driver genes, usually inactivation of two tumor suppressor genes and activation of an oncogene, are sufficient to trigger the initiation of CRC [39, 17, 34]. Typically, this involves a total of five genetic alterations, as inactivation of a tumor suppressor gene (TSG) requires inactivation of both alleles, and activation of an oncogene only requires a mutation in one allele of the gene [36]. Recent work [34] studied a five-step branching process model for the initiation of colorectal cancer that involves the three most commonly mutated driver genes in colorectal cancer: tumor suppressors *APC* and *TP53* and oncogene *KRAS*. The study found that, in the majority of cases, the driver mutations accrue in a specific order, with inactivation of *APC* followed by activation of *KRAS* and inactivation of the *TP53* gene.

Following [34], we study the single most likely mutational pathway to colorectal cancer. We model the dynamics using a multi-type branching process that starts from N wild-type crypts, small tubular assemblies of cells that line the intestinal epithelium [40, 6]. As the pro-

cess evolves, individual crypts stochastically obtain driver mutations, with mutation rates determined by the genotype of the crypt and the driver gene in question. We allow growth rates of consecutive subpopulations to be equal, which leads to considerable mathematical difficulty. Furthermore, instead of starting with a growing population initiated by a single cell, we consider the case where the process is initiated from a large population of non-dividing wild-type crypts. Using results from the theory of martingales, we compute the waiting time distributions associated with each step of the process. Finally, we compare analytic results for the waiting time distributions to those from exact computer simulations of the process.

The approach presented here can be extended to other single-path branching process models in which the population growth rates of subsequent types are non-decreasing. For CRC, one can compute the waiting time distributions for other mutational pathways. The quantitative estimates provided here may also help better understand the lifetime risk of CRC.

3.1 Model

Let $N_i(t)$ be the stochastic process that counts the population of type-i crypts at time t. In a single step, a type-i crypt transforms into a type-$(i+1)$ crypt through either mutation or loss of heterozygosity (LOH) at rate u_i. Here we consider the most common mutational pathway observed (see Table 3.1) over the course of an 80-year human lifespan, i.e. $0 \leq t \leq 80$. We assume

Step	Rate	Biological process
$N_0 \to N_1$	u_1	*APC* LOH
$N_1 \to N_2$	u_2	*APC* mutation
$N_2 \to N_3$	u_3	*KRAS* mutation
$N_3 \to N_4$	u_4	*TP53* LOH
$N_4 \to N_5$	u_5	*TP53* mutation

Table 3.1: Most common pathway to CRC initiation.

(a) Non-decreasing crypt growth rates

crypts	N_0	N_1	N_2	N_3	N_4
birth rate	$\lambda_0 = 0$	$\lambda_1 = 0$	$\lambda_2 > 0$	$\lambda_3 > \lambda_2$	$\lambda_4 = \lambda_3$

(b) Biologically reasonable range of parameter values

number of wild-type crypts	N	$10^7 - 10^8$
birth rates	λ_i	$\lambda_2 = 0.2/y, \lambda_3 = 0.27/y$
transition rates	u_i	$10^{-7} - 10^{-4}/y$

Table 3.2: Parameter values for the model of CRC initiation.

that independently from mutations, type-i crypts follow a pure birth process with rate λ_i. The growth rate of a crypt is determined by its genotype. For wild-type crypts, divisions are very rare [31] so we set their growth rate to zero. It has been shown that inactivation of *APC* and/or activation of *KRAS* provides a fitness advantage to mutated crypts, leading to clonal expansion [29, 38]. In contrast, under normal conditions *TP53* inactivation alone does not provide a fitness advantage [40]. These findings are reflected in the choice of growth parameters in our model (see Table 3.2).

This branching process model can be summarized as

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \circlearrowright \text{divide at rate } \lambda_2 \; N_2(t) \xrightarrow{u_3} \circlearrowright \lambda_3 \; N_3(t) \xrightarrow{u_4} \circlearrowright \lambda_4 = \lambda_3 \; N_4(t) \xrightarrow{u_5} N_5(t)$$

The first generation of type-i crypts (all type-i crypts produced directly from type-$(i-1)$ crypts) follows an inhomogeneous Poisson process, denoted by $K_i(t)$, with instantaneous intensity $u_i N_{i-1}(t)$. The transition times are the founding times of new type-i subclones. They are defined as

$$T_{i,j} = \inf\{t \geq 0 | K_i(t) = j\}.$$

At the j-th transition time, the newly formed type-i crypt founds a linear birth-death process with birth rate λ_i and death rate u_{i+1} (transition rate to the next type). Let $Z_{i,j}(t)$ denote this branching process where t is measured from the founding time. We also call $Z_{i,j}(t)$ the j-th lineage or j-th subclone of type-i individuals. Let $Z^{(i)}$ represent a single type-i individual. The transition scheme is

$$Z^{(i)} \to \begin{cases} Z^{(i)}\, Z^{(i)}, & \text{growth rate } \lambda_i \\ Z^{(i+1)}, & \text{transition rate } u_{i+1}. \end{cases}$$

Suppose $Z_i(t)$ is a linear birth-death process initiated by a single cell and $Z_i(t) = 0$ for $t < 0$. Then measured from the starting time of CRC, $Z_{i,j}(t)$ is merely $Z_i(t)$ shifted by $T_{i,j}$. Thus

$$N_i(t) = \sum_{j=1}^{K_i(t)} Z_i\left(t - T_{i,j}\right),$$

$T_{i,j} \le t$, for $1 \le j \le K_i(t)$. Here we note that since u_{i+1} is small, we treat all lineages of type-i individuals as pure birth processes when type-i has a positive birth rate. The system initially consists N wild-type (type-0) crypts, and we seek to estimate the waiting times for the first type-i crypt which is defined by

$$\tau_i = \inf\{t \ge 0 | N_i(t) > 0\}.$$

To verify our analytic results, we developed Monte Carlo simulations of a multi-type branching process model based on the Gillespie algorithm [20]. Parameter values for our model come from Paterson et al. [34], and their typical ranges are listed in Table 3.2.

3.2 Population dynamics of type-i crypts

In this section we derive the asymptotic growth dynamics of the clonal subpopulations (or types) along the path to CRC initiation. For each type i, the population size $N_i(t)$ is a counting process indexed by $t > 0$. The first two types, $N_0(t)$ and $N_1(t)$, are well approximated by time-deterministic functions. Starting with these, we sequentially derive the large-time population dynamics for the later types, leveraging results for two-type branching processes, $(N_i, N_{i+1}), 0 \le i \le 4$. The success of this approach relies on our ability to approximate

the large-time growth dynamics by the product of a time-deterministic function and a time-invariant random variable. The existence of such a decomposition depends on the martingale properties of the counting processes $N_i(t)$.

3.2.1 Type-0

Wild-type human crypts (i.e. type-0 crypts) rarely divide [31], hence we set the growth rate of the wild-type crypts to zero ($\lambda_0 = 0$). These wild-type crypts transition into type-1 crypts at rate $u_1 N_0(t)$ after losing one copy of the *APC* gene.

Proposition 19. *$N_0(t)$ is a pure death process with death rate u_1 and initial condition $N_0(0) = N$. Its generating function at time t is*

$$G_0(s,t) = \mathbb{E}[s^{N_0(t)}] = \left(se^{-tu_1} + 1 - e^{-tu_1}\right)^N,$$

and the expectation and variance are

$$\mathbb{E}[N_0(t)] = Ne^{-u_1 t},$$
$$\mathrm{Var}[N_0(t)] = e^{-2u_1 t}N(e^{u_1 t} - 1).$$

For a detailed description of a single-type birth-death process, one can refer to chapter 3, section 5 in [4]. Typical biologically realistic parameter values are $u_1 t \sim 10^{-2} \ll 1$ and $N \sim 10^8$, which leads to

$$\frac{\sqrt{\mathrm{Var}[N_0(t)]}}{\mathbb{E}[N_0(t)]} = \sqrt{\frac{e^{tu_1} - 1}{N}} \sim 10^{-5}, \quad \mathbb{E}[N_0(t)] = Ne^{-u_1 t} \approx N.$$

Motivated by this result, we approximate the population of the wild-type crypts by its expectation

$$N_0(t) \approx Ne^{-u_1 t} \approx N,$$

in which the second approximation is justified because $u_1 t \ll 1$. Therefore, the waiting time of $N_1(t)$, $\tau_1 \sim \text{Exponential}(u_1 N)$.

3.2.2 Type-1

A type-1 crypt is produced when a healthy crypt (type-0) loses one copy of the *APC* gene. This mutation does not lead to clonal expansion [29], so the growth rate of type-1 crypts, $\lambda_1 = 0$. While being produced by wild-type crypts, type-1 crypts mutate into type-2 at rate $u_2 N_1(t)$, losing both copies of the *APC* gene. We assume that the population outflows are negligible. Initially, there are no type-1 crypts, i.e., $N_1(0) = 0$.

Proposition 20. *$N_1(t) = N - N_0(t)$ has the following expectation and variance:*

$$\mathbb{E}[N_1(t)] = N(1 - e^{-u_1 t}),$$
$$\mathrm{Var}[N_1(t)] = e^{-2u_1 t} N(e^{u_1 t} - 1).$$

This can be derived by observing $N_0(t) + N_1(t) = N$, since the outflows from type-1 to type-2 are negligible. Therefore $\mathbb{E}[N_1(t)] = \mathbb{E}[N - N_0(t)] = N - \mathbb{E}[N_0(t)], \mathrm{Var}[N_1(t)] = \mathrm{Var}[N_0(t)]$.

To approximate $N_1(t)$, we consider the case when the wild-type population is effectively constant. In this case, $N_1(t)$ can be treated as a Poisson process with intensity $u_1 N$. According to the properties of a Poisson process, the approximated expectation and variance are

$$\mathbb{E}[N_1(t)] \approx u_1 N t,$$
$$\mathrm{Var}[N_1(t)] \approx u_1 N t.$$

Notice that the approximated expected value and variance are consistent with the true expected value and variance. Since $u_1 t \sim 10^{-2} \ll 1$ is small, we have

$$\mathbb{E}[N_1(t)] = N(1 - e^{-u_1 t}) \approx u_1 N t,$$
$$\mathrm{Var}[N_1(t)] = e^{-2u_1 t} N(e^{u_1 t} - 1) = N(e^{-u_1 t} - e^{-2u_1 t}) \approx u_1 N t.$$

Additionally, observe that

$$\frac{\sqrt{\mathrm{Var}[N_1(t)]}}{\mathbb{E}[N_1(t)]} \approx \frac{1}{\sqrt{u_1 N t}} \sim 10^{-2},$$

and so we approximate $N_1(t)$ by a deterministic function

$$N_1(t) \approx u_1 N t.$$

Thus the waiting time distribution of type-2 crypts is simply estimated as

$$P(\tau_2 \le t) = 1 - \exp\left(-u_2 \int_0^t N_1(s)ds\right) \approx 1 - \exp\left(-\frac{1}{2}u_1 u_2 N t^2\right).$$

3.2.3 Type-2

After losing both copies of the *APC* genes, type-1 crypts mutate into type-2 crypts. The *APC* inactivation provides a fitness advantage to type-2 crypts [29]. Thus for type-2, we have a positive growth rate $\lambda_2 > 0$. Meanwhile, the *KRAS* oncogenes are activated at rate $u_3 N_2(t)$, mutating type-2 crypts into type-3 crypts. These outflows are ignored when considering the population of $N_2(t)$. At time $t = 0$, there are no type-2 crypts, i.e $N_2(0) = 0$. We begin with the expectation of $N_2(t)$. Using Theorem 13, we compute

$$\mathbb{E}[N_2(t)] \approx \frac{N u_1 u_2 (e^{\lambda_2 t} - \lambda_2 t - 1)}{\lambda_2^2}.$$

The approximation is obtained by noting that $u_1 \sim 10^{-4} \ll \lambda_2 \sim 10^{-1}$ and $u_1 t \sim 10^{-2} \ll 1$. The following large-time asymptotic limit exists for $N_2(t)$.

Theorem 21. *$e^{-\lambda_2 t} N_2(t) \to W_2$ a.s. and in L^1 with*

$$\mathbb{E}[W_2] = \frac{N u_1 u_2}{\lambda_2(\lambda_2 + u_1)} \approx \frac{N u_1 u_2}{\lambda_2^2}.$$

Proof. By Lemma 14, consider

$$I_2 = \int_0^\infty u_2 N_1(s) e^{-\lambda_2 s} \mathrm{d}s.$$

Since

$$\begin{aligned}
\mathbb{E}[I_2] =& \mathbb{E}\left[\int_0^\infty u_2 e^{-\lambda_2 s} N_1(s) \mathrm{d}s\right] \\
\text{(By Tonelli's theorem)} =& \int_0^\infty u_2 \mathbb{E}[N_1(s)] e^{-\lambda_2 s} \mathrm{d}s \\
\le& \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s} \mathrm{d}s \\
=& N u_1 u_2 \int_0^\infty s e^{-\lambda_2 s} \mathrm{d}s \\
=& \frac{N u_1 u_2}{\lambda_2^2} < \infty,
\end{aligned}$$

there exists W_2 s.t.

$$e^{-\lambda_2 t} N_2(t) \to W_2 \text{ a.s. as } t \to \infty.$$

Next, we show uniform integrability so that the $\mathbb{E}[W_2]$ is well-defined. We prove this for N_1, N_2 and N_3 in Lemma 25 (in Appendix A). Since L^1 convergence is guaranteed, $\mathbb{E}[W_2] = \mathbb{E}[I_2]$, so that

$$\begin{aligned}\mathbb{E}[W_2] &= \int_0^\infty u_2 \mathbb{E}[N_1(s)] e^{-\lambda_2 s} \mathrm{d}s \\ &= \int_0^\infty u_2 N(1 - e^{-u_1 s}) e^{-\lambda_2 s} \mathrm{d}s \\ &= \frac{N u_1 u_2}{\lambda_2(\lambda_2 + u_1)}.\end{aligned}$$

□

To characterize the distribution of the limiting random variable, we consider $\tilde{N}_2(t)$, the population of type-2 crypts produced by $N_1^*(t) = u_1 N t$.

Theorem 22. *$e^{-\lambda_2 t}\tilde{N}_2(t) \to V_2$ a.s. and in L^1 with*

$$\mathbb{E}[V_2] = \frac{N u_1 u_2}{\lambda_2^2}$$

and

$$\mathscr{L}_{V_2}(\theta) = \mathbb{E}\left[e^{-\theta V_2}\right] = \exp\left(\frac{N u_1 u_2 \mathrm{PolyLog}(2, -\theta)}{\lambda_2^2}\right).$$

Proof. Note that I_2^* is deterministic and has a finite expected value

$$\mathbb{E}[I_2^*] = I_2^* = \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s} \mathrm{d}s = \frac{N u_1 u_2}{\lambda_2^2} < \infty.$$

By Lemma 14, we must have $e^{-\lambda_2 t}\tilde{N}_2(t) \xrightarrow{a.s.} V_2$. Next, we show uniform integrability which guarantees L^1 convergence. It is shown in Lemma 26 that all tilde processes in this paper are uniform integrable. This implies $e^{-\lambda_2 t}\tilde{N}_2(t) \xrightarrow{L^1} V_2$ and $\mathbb{E}[V_2] = \mathbb{E}[I_2^*] = \frac{N u_1 u_2}{\lambda_2^2}$. Finally, to compute the Laplace transform of V_2, we plug $f_2(t) = e^{\lambda_2 t}$ into the formula in 16. □

Corollary 1. *$\frac{\tilde{N}_2(t)}{f_2(t)} \to V_2$ a.s. and in L^1 for all*

$$f_2(t) \in \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1\}.$$

Proof. The convergence directly follows Theorem 22 by the fact that

$$\forall\ \omega \in \{\omega : \lim_{t\to\infty} e^{-\lambda_2 t}\tilde{N}_2(\omega,t) = V_2(\omega)\},$$

$$\lim_{t\to\infty}\frac{\tilde{N}_2(\omega,t)}{f_2(t)} = \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t)\lim_{t\to\infty}\frac{\tilde{N}_2(\omega,t)}{f_2(t)} = V_2(\omega).$$

Observe that for $t>0$ sufficiently large, we have $f_2(t)>0$ and $e^{\lambda_2 t}/f_2(t) < M$. Note that

$$\sup_t \mathbb{E}\left|\frac{\tilde{N}_2(t)}{f_2(t)}\right| \le \sup_t\left(e^{\lambda_2 t}/f_2(t)\right)\sup_t E\left[\tilde{N}_2(t)e^{-\lambda_2 t}\right],$$

and by Lemma 26 $\sup_t \mathbb{E}\left[\tilde{N}_2(t)e^{-\lambda_2 t}\right]$ is bounded. Hence $\frac{\tilde{N}_2(t)}{f_2(t)}$ is uniform integrable and the convergence is in L^1. □

3.2.4 Type-3

Type-3 crypts are produced by type-2 crypts through activation of the *KRAS* oncogene, which increases the fission rate of mutated cells [38]. Thus the birth rate has a positive increment, i.e. $\lambda_3 > \lambda_2$. At a small rate $u_4 N_3(t)$, type-3 crypts lose one copy of the TSG *TP53* and mutate into type-4 crypts. The initial population is $N_3(0)=0$. Using Theorem 13, the expected value of $N_3(t)$ is

$$\mathbb{E}[N_3(t)] \approx \frac{N u_1 u_2 u_3\left(\lambda_2^2\left(e^{\lambda_3 t}-\lambda_3 t-1\right)-\lambda_3^2\left(e^{\lambda_2 t}-\lambda_2 t-1\right)\right)}{\lambda_2^2(\lambda_3-\lambda_2)\lambda_3^2},$$

where the approximation is made by observing $\lambda_i + u_1 \approx \lambda_i$ and $u_1 t \sim 10^{-2}$.

Theorem 23. *$e^{-\lambda_3 t}N_3(t) \to W_3$ a.s. and in L^1 with*

$$\mathbb{E}[W_3] = \frac{N u_1 u_2 u_3}{(\lambda_3-\lambda_2)\lambda_3(\lambda_3+u_1)} \approx \frac{N u_1 u_2 u_3}{(\lambda_3-\lambda_2)\lambda_3^2}.$$

Proof. By Lemma 14 and Lemma 25, we need to verify that I_3 has finite expectation.

$$\begin{aligned}
\mathbb{E}[I_3] &= \int_0^\infty u_3 \mathbb{E}[N_2(s)] e^{-\lambda_3 s}\mathrm{d}s \\
&\le \int_0^\infty u_3 \frac{u_1 u_2 N}{\lambda_2^2}(e^{\lambda_2 s}-\lambda_2 s-1)e^{-\lambda_3 s}\mathrm{d}s \\
&= \frac{N u_1 u_2 u_3}{\lambda_2^2}\int_0^\infty (e^{(\lambda_2-\lambda_3)s}-\lambda_2 s e^{-\lambda_3 s}-e^{-\lambda_3 s})\mathrm{d}s \\
&= \frac{N u_1 u_2 u_3}{(\lambda_3-\lambda_2)\lambda_3^2} < \infty.
\end{aligned}$$

And the expected value of W_3 is given by the expected value of I_3,

$$\begin{aligned}\mathbb{E}[W_3] &= \mathbb{E}[I_3] \\ &= \int_0^\infty u_3 \mathbb{E}[N_2(s)] e^{-\lambda_3 s} \mathrm{d}s \\ &= \int_0^\infty u_3 N u_2 \frac{u_1(e^{\lambda_2 s}-1)+\lambda_2(e^{-u_1 s}-1)}{\lambda_2(\lambda_2+u_1)} e^{-\lambda_3 s} \mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{(\lambda_3-\lambda_2)\lambda_3(\lambda_3+u_1)}.\end{aligned}$$

□

Consider a system with $N_2^*(t) := f_{20}(t)V_2 = e^{\lambda_2 t}V_2$ and let $\tilde{N}_3(t)$ denote the number of type-3 crypts in this system. The following large-time limit holds.

Theorem 24. *$e^{-\lambda_3 t}\tilde{N}_3(t) \to V_3$ a.s. and in L^1 with*

$$\mathbb{E}[V_3] = \frac{N u_1 u_2 u_3}{\lambda_2^2(\lambda_3-\lambda_2)}.$$

Proof. By Lemma 14 and Lemma 26, we compute

$$\begin{aligned}\mathbb{E}[I_3^*] &= \int_0^\infty u_3 E[N_2^*(s)] e^{-\lambda_3 s} \mathrm{d}s \\ &= \int_0^\infty u_3 \frac{u_1 u_2 N}{\lambda_2^2} e^{\lambda_2 s} e^{-\lambda_3 s} \mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{\lambda_2^2} \int_0^\infty e^{(\lambda_2-\lambda_3)s} \mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{\lambda_2^2(\lambda_3-\lambda_2)} < \infty.\end{aligned}$$

□

For other potential $f_{2,k}$'s, the Laplace transform changes accordingly.

Corollary 2. *Let $N_2^*(t) := f_2(t)V_2$ where*

$$f_2(t) \in \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1\},$$

and let $\tilde{N}_3(t)$ *denote type-3 crypts in this system. Then* $\frac{\tilde{N}_3(t)}{f_3(t)} \to V_3$ *a.s. and in* L^1 *with*

$$\mathbb{E}[V_3] = \frac{N u_1 u_2 u_3}{\lambda_2^2} \int_0^\infty f_2(t) e^{-\lambda_3 s} \mathrm{d}s,$$

and

$$\mathscr{L}_{V_3}(\theta) = \mathbb{E}[e^{-\theta V_3}] = \mathscr{L}_{V_2}\left(u_i \int_0^\infty \frac{\theta f_2(s)}{\theta + e^{\lambda_i s}} \mathrm{d}s \right),$$

for all

$$f_3(t) \in \{ f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_3 t} f_3(t) = 1 \}.$$

Proof. Since $f_2(t)e^{-\lambda_3 t}$ is integrable, the validity of the convergence and expectation follow directly from Theorem 24. Lemma 16 gives the Laplace transform of V_3. □

Note that $f_{i1}(t) = \mathbb{E}[N_i]/\mathbb{E}[W_i]$ is more precise than $f_{i0}(t) = e^{\lambda_i t}$. By Corollary 2, $\mathbb{E}[V_3] = \mathbb{E}[I_3^*] = \int_0^\infty u_3 f_2(t) \mathbb{E}[V_2] e^{-\lambda_3 s} \mathrm{d}s$, depends on $f_2(t)$ and $\mathbb{E}[V_2]$. Since $\mathbb{E}[W_3] = \int_0^\infty u_3 \mathbb{E}[N_2(s)] e^{-\lambda_3 s} \mathrm{d}s$ and $\mathbb{E}[V_2] = \mathbb{E}[W_2]$, setting $f_2(t) = \mathbb{E}[N_2(t)]/\mathbb{E}[W_2]$ equates $\mathbb{E}[V_3]$ and $\mathbb{E}[W_3]$. Indeed, when using $f_2(t) = f_{20}(t) = e^{\lambda_2 t}$, the resulting random variable V_3 has an expectation different from W_3.

3.2.5 Type-4

Since only one copy of the *TP53* gene has been lost, type-4 does not confer a growth advantage [40]. Thus the division rate of type-4 crypts $\lambda_4 = \lambda_3$. The dynamics of type-4 crypts also include negligible outflows at rate $u_5 N_4(t)$. And the starting condition is $N_4(0) = 0$. In this case, we note that using $e^{-\lambda_3 t}$ to scale the growth process fails to provide a large-time limiting random variable. Even though

$$M_t = e^{-\lambda_3 t} N_4(t) - \int_0^t u_4 N_3(s) e^{-\lambda_3 s} \mathrm{d}s,$$

is still a martingale, Lemma 14 does not apply. We have seen that $\mathbb{E}[N_3(t)] = O(e^{\lambda_3 t})$ as $t \to \infty$, and so

$$u_4 \mathbb{E}[N_3(s)] e^{-\lambda_3 s} = O(e^{\lambda_3 s}) e^{-\lambda_3 s} = O(1).$$

This implies the first condition in Lemma 14 is invalid, that is as $t \to \infty$

$$I_4(t) = \int_0^t u_4 N_3(s) e^{-\lambda_3 s} \mathrm{d}s = O(t) \to \infty.$$

Therefore we cannot derive a large time limit of type-4 individuals. Instead, we will estimate the waiting time of type-5 crypts by the inhomogeneous Poisson approximation discussed in section 2.4.4. To this end, we compute $p_0^{4\to5}(s,t)$ and $p_0^{3\to5}(s,t)$, corresponding to estimating τ_5 by V_3 and V_2 respectively,

$$p_0^{4\to5}(s,t) = \frac{1}{1+u_5\frac{\exp\{(\lambda_3(t-s))\}-1}{\lambda_3}},$$
$$p_0^{3\to5}(s,t) = \left(1+\frac{u_4u_5(\lambda_3-u_5+e^{\lambda_3(t-s)}(u_5-\lambda_3+\lambda_3\log\left(\frac{\lambda_3e^{\lambda_3(t-s)}}{\lambda_3+u_5(e^{\lambda_3(t-s)}-1)}\right)}{(\lambda_3-u_5)^2\lambda_3}\right)^{-1}.$$

3.3 Estimating waiting times for colorectal cancer initiation

In this section, we compare results obtained from exact computer simulations of the multi-type process with our analytic results. We define $p_i(t)$ as our estimate of the cumulative distribution function of τ_i, the waiting time to the first type-i crypt. For all figures in this section, parameter values are given in Table 3.3, and follow estimates from Paterson et al [34].

Crypt N_i	N_0	N_1	N_2	N_3	N_4
Initial population (crypts)	$N = 1\times10^8$	0	0	0	0
Birth rate λ_i (crypts/year)	0	0	0.2	0.27	0.27
Transition rate u_{i+1} (crypts/year)	2.86×10^{-4}	1.06×10^{-5}	9.00×10^{-7}	1.36×10^{-4}	4.56×10^{-7}

Table 3.3: Estimates of parameter values for colorectal cancer initiation from Paterson et al. [34].

3.3.1 Waiting times to type-1 and type-2

For τ_1 and τ_2, the analytic results are listed in the previous section. Recall that the population of wild-type (type-0) crypts are estimated to have a fixed population N. Thus,

$$P(\tau_1 \le t) \approx p_1(t) = 1-\exp(-u_1Nt).$$

For τ_2, recall that we approximate the population of type-1 crypts by a time deterministic function $u_1 N t$.

$$P(\tau_2 \leq t) \approx p_2(t) = 1 - \exp\left(-\frac{1}{2} u_1 u_2 N t^2\right). \tag{3.1}$$

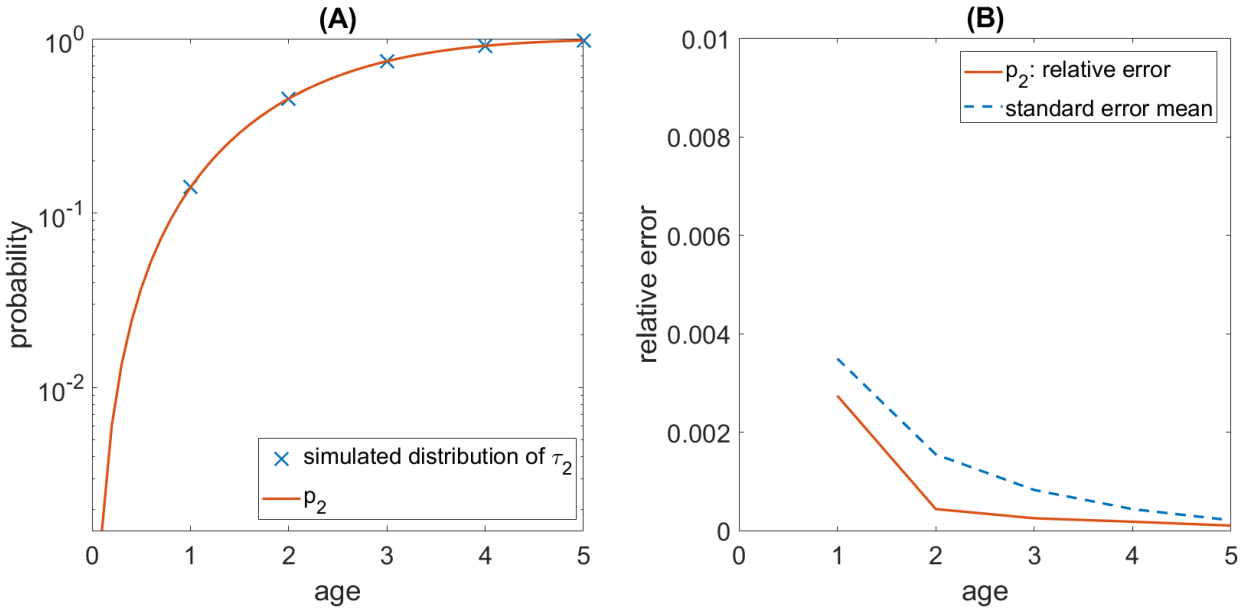

Figure 3.1: (A) Comparison of the analytic cumulative distribution function of τ_2, the waiting time to the first type-2 crypt (p_2, equation (3.1)), and the simulated distribution of τ_2 across 5×10^5 runs. (B) Dashed line shows the standard error of the mean obtained from simulations. Solid line is the relative error of the analytic result.

3.3.2 Waiting time to type-3

To compute the distribution functions of τ_3, τ_4 and τ_5, we need to apply the approximation in Proposition 15. Hence, we want to find f_2, f_3. Recall that we consider two functions for each subpopulation:

$$f_{i0}(t) := e^{\lambda_i t}, \; f_{i1}(t) := \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[V_i]}.$$

The approximation for the distribution τ_3 is

$$P(\tau_3 \le t) \approx 1 - \mathscr{L}_{V_2}\left(u_3 \int_0^t f_2(s)ds\right),$$

where $f_2(t) \in \{f \in C(\mathbb{R}) : f^{-1}(t)e^{\lambda_2 t} \to 1\}$. First, consider $f_{20}(t) = e^{\lambda_2 t}$. The corresponding distribution function is

$$p_{30}(t) = 1 - \exp\left(\frac{Nu_1u_2}{\lambda_2^2}\text{PolyLog}\left(2, -u_3\frac{e^{\lambda_2 t}-1}{\lambda_2}\right)\right). \tag{3.2}$$

In the second case,

$$f_{21}(t) = \frac{\mathbb{E}[N_2(t)]}{\mathbb{E}[V_2]} = e^{\lambda_2 t} - \lambda_2 t - 1,$$

and the corresponding distribution is

$$p_{31}(t) = 1 - \exp\left(\frac{Nu_1u_2}{\lambda_2^2}\text{PolyLog}\left(2, -u_3\left(\frac{e^{\lambda_2 t}-1}{\lambda_2} - \frac{1}{2}t^2 - t\right)\right)\right). \tag{3.3}$$

By design, the first moment of $N_{21}^*(t)$ is identical to that of $N_2(t)$,

$$\mathbb{E}[N_{21}^*(t)] = \mathbb{E}[f_{21}(t)V_i] = \mathbb{E}[N_2(t)].$$

Both (3.2) and (3.3) agree with the simulated distributions for $t > 40$ (See Fig 3.2). However, in the intermediate regime where t is small, one can observe that p_{31} is more accurate than p_{30}.

3.3.3 Waiting time to type-4

Considering Proposition 17, we have several options to compute τ_4, the waiting time to the first type-4 crypt. Specifically, we may compute the distribution of τ_4 in two ways: (i) compute the distribution directly using $\mathscr{L}_{V_3}(\theta)$; and (ii) skip V_3 and compute the distribution using $\mathscr{L}_{V_2}(\theta)$ and $p_0^{3\to 4}(s,t)$. In these approaches, by Corollary 2, the time deterministic growth functions f_2, f_3 can be any functions that have large time limits identical to the corresponding exponential functions, $e^{\lambda_2 t}$ and $e^{\lambda_3 t}$. Among different choices of the growth functions, we typically consider two specific cases, $f_i = f_{i0} = e^{\lambda_i t}$ and $f_i = f_{i1} = \mathbb{E}[N_i]/\mathbb{E}[V_i]$.

If V_3 is not skipped, $\mathscr{L}_{V_3}$ depends on the choice of f_2. Since f_{21} leads to more accurate computations of τ_2, we use f_{21} to compute $\mathscr{L}_{V_3}(\theta)$, that is

$$\mathscr{L}_{V_3}(\theta) = \mathscr{L}_{V_2}\left(u_3 \int_0^\infty \frac{f_{21}(s)\theta}{\theta + e^{\lambda_3 s}}\mathrm{d}s\right).$$

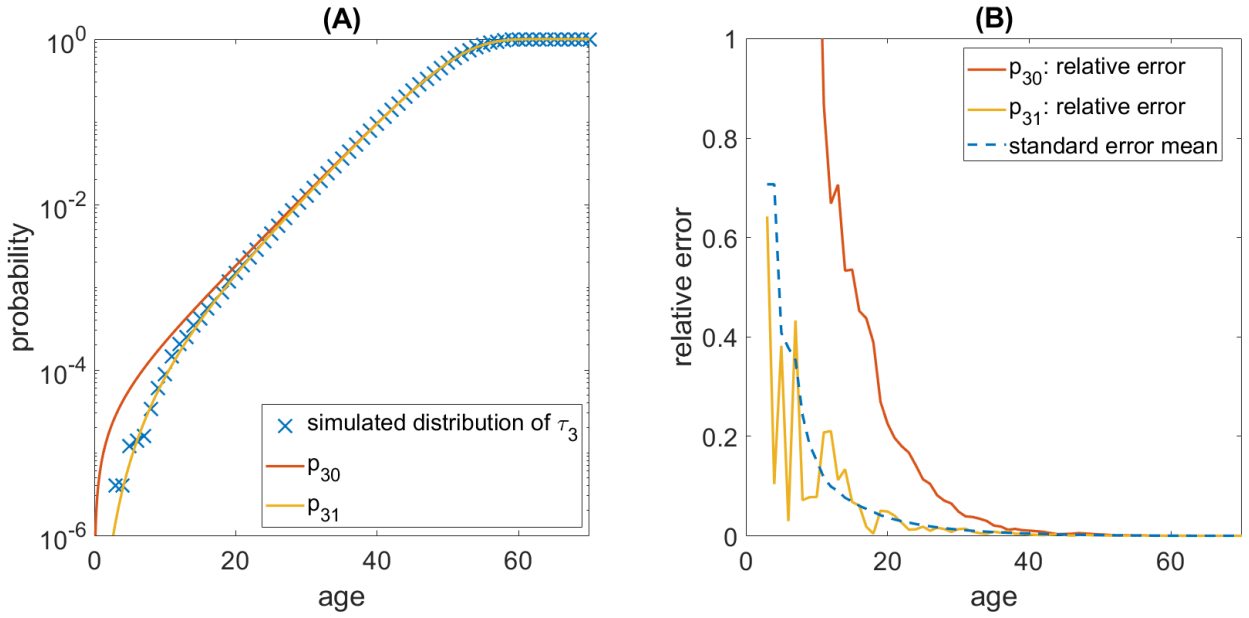

Figure 3.2: (A) Comparison of the analytic cumulative distribution functions of τ_3, the waiting time to the first type-3 crypt, and the distribution of τ_3 across 5×10^5 simulation runs. p_{30} and p_{31} are distribution functions obtained from equations (3.2) and (3.3) respectively. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

For f_3, we consider two candidate functions,

$$f_{30}(t) = e^{\lambda_3 t},$$
$$f_{31}(t) = \frac{\mathbb{E}[N_3(t)]}{\mathbb{E}[V_3]} = e^{\lambda_3 t} - \lambda_3 t - 1 - \frac{\lambda_3^2}{\lambda_2^2}(e^{\lambda_2 t} - \lambda_2 t - 1).$$

The first, $f_{30}(t)$, has a simpler form corresponding to the leading order of $\mathbb{E}[N_3(t)]$, with

$$\mathbb{E}[N_{30}^*(t)] = \mathbb{E}[f_{30}(t)V_3] \neq \mathbb{E}[N_3(t)],$$

and the second, $f_{31}(t)$, matches the expected value of $\mathbb{E}[N_3(t)]$

$$\mathbb{E}[N_{31}^*(t)] = \mathbb{E}[f_{31}(t)V_3] = \mathbb{E}[N_3(t)].$$

Next, we define

$$p_{40}(t) = 1 - \mathscr{L}_{V_3}\left(u_4 \int_0^t f_{30}(s)ds\right), \tag{3.4}$$

$$p_{41}(t) = 1 - \mathscr{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)ds\right). \tag{3.5}$$

On the other hand, if we skip V_3 and use the inhomogeneous approximation in section 2.4.4, we can consider two types of f_2's:

$$p_{40}^{s3}(t) = 1 - \mathscr{L}_{V_2}\left(u_4 \int_0^t f_{20}(s)(1 - p_0^{3\to 4}(s,t))ds\right), \tag{3.6}$$

$$p_{41}^{s3}(t) = 1 - \mathscr{L}_{V_2}\left(u_4 \int_0^t f_{21}(s)(1 - p_0^{3\to 4}(s,t))ds\right). \tag{3.7}$$

We note that p_{40}^{s3} and p_{41}^{s3} have closed form expressions, which are omitted here and listed in the Appendix. The results are shown in Fig 3.3. We observe that skipping V_3 improves the accuracy of the results for large time t, while including more terms in f_i generally improves the accuracy of the approximation.

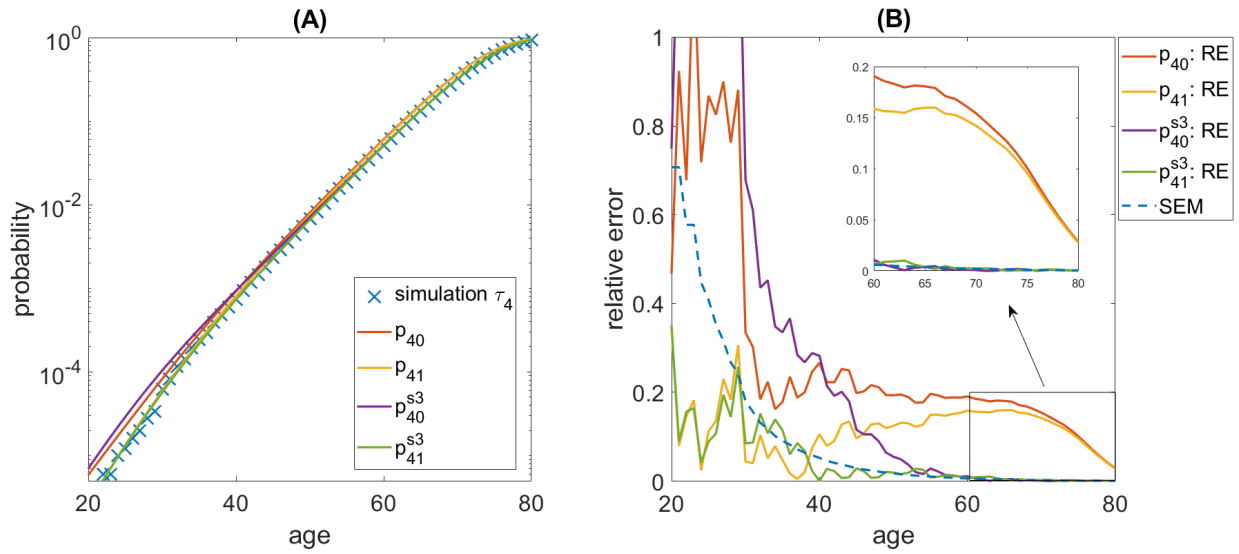

Figure 3.3: (A) Comparison of analytic cumulative distribution functions of τ_4, the waiting time to the first type-4 crypt, and the distribution of τ_4 across 5×10^5 simulation runs. p_{40} (equation (3.4)) and p_{41} (equation (3.5)) are analytic distribution functions derived from the Laplace transform V_3. p_{40}^{s3} (equation (3.6)) and p_{41}^{s3} (equation (3.7)) are analytic distribution functions derived by skipping V_3 and using the Laplace transform of V_2. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

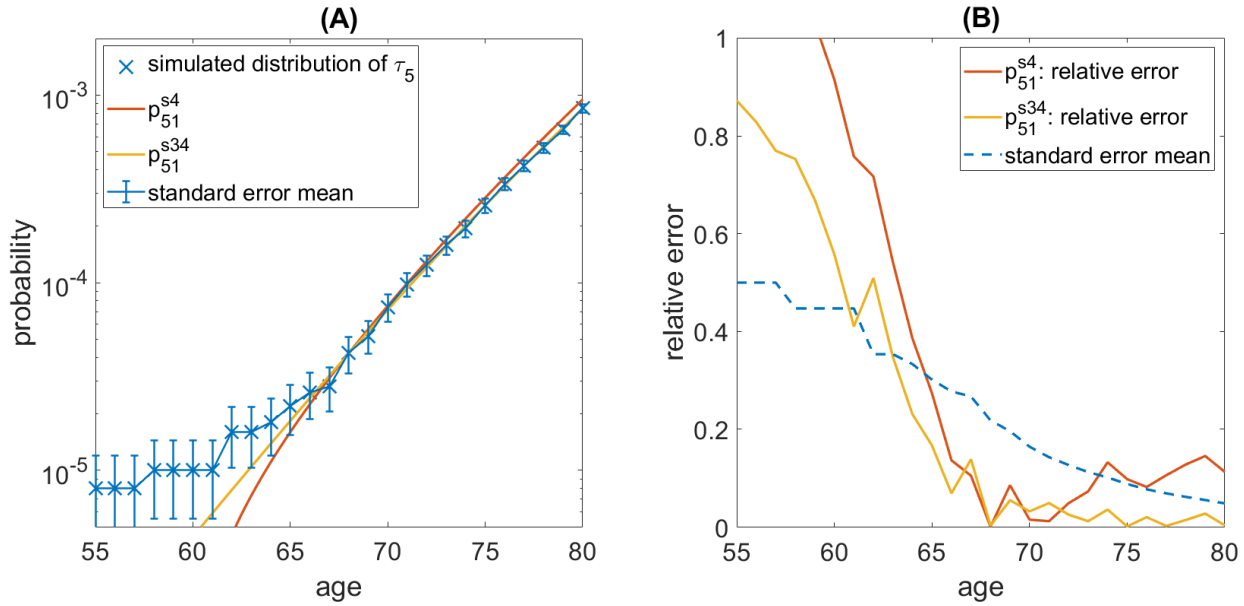

Figure 3.4: (A) Comparison of the analytic cumulative distribution functions of τ_5, the waiting time to the first type-5 crypt, and the distribution of τ_5 across 5×10^5 simulation runs. p_{51}^{s3} (equation (3.8)) is derived by skipping type-4 and using the Laplace transform of V_3. p_{51}^{s34} (equation (3.9)) is derived by using the Laplace transform of V_2 and skipping the Laplace transforms of type-4 and type-3. The error bars represent the standard error of the mean of the simulation. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results.

3.3.4 *Waiting time to type-5*

For τ_5, the waiting time to the first type-5 crypt, we consider two approaches: (i) compute the distribution using the large-time limit of type-3 crypts, $\mathscr{L}_{V_3}(\theta)$, and $p_0^{4\to5}(s,t)$ (effectively skipping the large-time limit of type-4 crypts); and (ii) skip large-time limits of both type-4 and type-3 crypts, and compute the distribution using $\mathscr{L}_{V_2}(\theta)$ and $p_0^{3\to5}(s,t)$. Note that the time-deterministic growth functions f_2 and f_3 which appear in the long-time limits of type-2 and type-3 crypts can vary across these approaches (see Corollary 2). However, the most accurate results are obtained when $f_i = f_{i1} = \mathbb{E}[N_i]/\mathbb{E}[V_i]$.

The corresponding approximations of the distribution of τ_5 resulting from approaches (i) and (ii) are

$$p_{51}^{s4}(t) = 1 - \mathscr{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)(1 - p_0^{4\to5}(s,t))ds\right), \tag{3.8}$$

$$p_{51}^{s34}(t) = 1 - \mathscr{L}_{V_2}\left(u_3 \int_0^t f_{21}(s)(1 - p_0^{3\to5}(s,t))ds\right). \tag{3.9}$$

The first result, $p_{51}^{s4}(t)$, has an explicit form given by equation (B.1), while the second result, p_{51}^{s34} (given by equation (B.3)), to the best of our knowledge, is not an elementary function or a standard special function. Comparison of these two formulas with the waiting time for the first type-5 crypt obtained from exact computer simulations is shown in Fig 3.4. We observe that p_{51}^{s34} achieves higher accuracy compared to p_{51}^{s3}, especially at later times (above age 70). In other words, compared with the result incorporating V_3, skipping this stage gives more accurate results. The intuition behind this is that approximating each subclone by its large time limit is more accurate than approximating the total population by its overall large time limit.

3.4 Discussion

Results in the previous sections show that in our case, the accuracy of different approximations is consistent for all waiting times. Firstly, given a fixed time-deterministic evolution f_i, the minimal large-time error for each waiting time is obtained by skipping as many approximated steps (identified by V_j, $j < i$) as possible. This can be observed in Fig 3.3 and Fig 3.4. For

instance, 3.3 (B) shows that at $t = 80$, p_{40}^{s3} and p_{41}^{s3} are better than other approximations. Qualitatively speaking, this is explained by the fact that in the long term, decomposing a lineage into a sequence of subclones and adding up large limits is more accurate than directly using the asymptotic limit of the lineage itself. Secondly, suppose that we start at step i (with V_i) and the number of skipped types is fixed. In this case different choices of the time-deterministic evolution (f_i) only influence the accuracy in the approximation for times near 0, which we have identified as the intermediate state. This is evident in the high accuracy of p_{41} and p_{41}^{s3} in Fig 3.3 (B) during time interval $[0,40]$, and is explained by that for fixed i, the time-deterministic functions differ only in the early time period, hence they are most influential in early time. Conversely, error is introduced by V_i as an approximation in the time period $[\tau_i, 80]$. Lastly, we note that the best time-deterministic function f_i for the approximation of N_i is $\mathbb{E}[N_i]/\mathbb{E}[V_i]$. The reason is that although all approximations are based on L^1 convergence, $f_i = \mathbb{E}[N_i]/\mathbb{E}[V_i]$ gives the approximation the same expected value for any time t. Hence the best f_i also works for short time.

Chapter 4

CONCLUSION

In this thesis, we have considered a multi-type branching process model with a transition network and have applied it to a five-step model of the main mutation path for the initiation of a subclass of colorectal cancers. The main question we want to attack is the waiting time of the first individual of each type.

A typical method of computing waiting times requires the large time limiting random variable of each type. Motivated by this requirement, we consider a classification of models by their large time limits. We have introduced mean value matrices by which we can decompose a branching process into subprocesses and obtain large time limits separately. Each subprocess has its mean matrix that is a sub-block of the the mean value matrix of the original block. Four types of blocks with different initial conditions ,structures, and large time limits have been discussed.

Among all evolutionary graphs, the simplest and most important case is when the graph is a pathway. We have mainly studied the case when the net growth rates along the path is non-decreasing, in which we allow growth rates to be same before and after a transition. Besides specifying conditions for the large time limit of each type to exist, we have also provided approaches for approximating the waiting time distributions. In addition, we have extended an approach for computing the asymptotic growth dynamics of a subpopulation in a branching process to reflect growth dynamics beyond simple exponential growth. Characterizing these dynamics is crucial for accurately computing the waiting time distributions for the mutations that mark the progression of cancer.

To verify our theory for computing waiting time distributions, we have applied our results to colorectal cancer initiation. A six-type continuous branching process has been established to

model the most probable mutational pathway to the first cancerous crypt. Compared to previous cancer models, there are two novel features: The initial type is slowly decaying; and, we allow the growth rate remain unchanged after mutation. These features result in sub-exponential population growth and the fact that one of the type does not have martingale. Despite having these difficulties, we use theorems in Chapter 2 including martingale convergence and inhomogeneous approximation to obtain the waiting time distribution for the first crypt of each type. Furthermore, we have developed Monte Carlo simulations to validate the accuracy of our analytic results.

Taken together, this thesis presents a framework that could be adapted to study other cancers and their progression, especially when the model is about a single transitional pathway. Potential future work include figuring out a more accurate approximation of the population for each type, computing the waiting time when the transitional network includes multiple paths, and finding out the most probable path among multiple paths. The irreducible case have already been studied by [22, 32], but when there is transition between two reducible processes, the resulting dynamics is still unclear at the moment.

BIBLIOGRAPHY

[1] Tibor Antal and PL Krapivsky. Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech. Theory Exp.*, 2011(08):P08018, 2011.

[2] P Armitage and R Doll. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer*, 8(1):1–12, March 1954.

[3] P Armitage and R Doll. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br. J. Cancer*, 11(2):161–169, June 1957.

[4] Krishna B Athreya and Peter E Ney. *Branching processes*. Dover Publications, New York, 2004.

[5] Stefano Avanzini and Tibor Antal. Cancer recurrence times from a branching process model. *PLoS Comput. Biol.*, 15(11):e1007423, 2019.

[6] Nick Barker, Rachel A Ridgway, Johan H van Es, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*, 457(7229):608–611, January 2009.

[7] Ivana Bozic, Tibor Antal, et al. Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.*, 107(43):18545–18550, 2010.

[8] Ivana Bozic and Martin A Nowak. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl. Acad. Sci. U.S.A.*, 111(45):15964–15968, 2014.

[9] Ivana Bozic, Chay Paterson, and Bartlomiej Waclaw. On measuring selection in cancer from subclonal mutation frequencies. *PLoS Comput. Biol.*, 15(9):1–15, 09 2019.

[10] Ivana Bozic, Johannes G Reiter, Benjamin Allen, Tibor Antal, Krishnendu Chatterjee, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife*, 2:e00747, jun 2013.

[11] Cameron W Brennan, Roel GW Verhaak, Aaron McKenna, Benito Campos, Houtan Noushmehr, et al. The somatic genomic landscape of glioblastoma. *Cell*, 155(2):462–477, 2013.

[12] David Cheek and Tibor Antal. Mutation frequencies in a birth–death branching process. *Ann. Appl. Prob.*, 28(6):3922–3947, 2018.

[13] Kaveh Danesh, Rick Durrett, Laura J Havrilesky, and Evan Myers. A branching process model of ovarian cancer. *J. Theor. Biol.*, 314:10–15, 2012.

[14] Richard Durrett and Stephen Moseley. Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Popul. Biol.*, 77(1):42–48, 2010.

[15] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge; New York, fifth edition, 2019.

[16] Rick Durrett, Jasmine Foo, Kevin Leder, John Mayberry, and Franziska Michor. Evolutionary dynamics of tumor progression with random fitness values. *Theor. Popul. Biol.*, 78(1):54–66, 2010.

[17] Eric R. Fearon. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol. Mech. Dis.*, 6(1):479–507, 2011.

[18] Jasmine Foo and Kevin Leder. Dynamics of cancer recurrence. *Ann. Appl. Probab.*, 23(4):1437–1468, 2013.

[19] Jasmine Foo, Kevin Leder, and Junfeng Zhu. Escape times for branching processes with random mutational fitness effects. *Stoch. Process. Their Appl.*, 124(11):3661–3697, 2014.

[20] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, December 1977.

[21] Eran Hodis, Ian R. Watson, Gregory V. Kryukov, Stefan T. Arold, Marcin Imielinski, et al. A landscape of driver mutations in melanoma. *Cell*, 150(2):251–263, 2012.

[22] Svante Janson. Functional limit theorems for multitype branching processes and generalized pólya urns. *Stoch. Process. Their. Appl.*, 110(2):177–245, 2004.

[23] David G. Kendall. Birth-and-death processes, and the theory of carcinogenesis. *Biometrika*, 47(1-2):13–21, 1960.

[24] David G. Kendall. Branching processes since 1873. *J. London Math. Soc.*, s1-41(1):385–406, 1966.

[25] Harry Kesten and Bernt P Stigum. Limit theorems for decomposable multi-dimensional galton-watson processes. *J. Math. Anal. Appl.*, 17(2):309–338, 1967.

[26] Alfred G Knudson. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.*, 68(4):820–823, 1971.

[27] Natalia Komarova. Stochastic modeling of drug resistance in cancer. *J. Theor. Biol.*, 239(3):351–366, 2006.

[28] Natalia L. Komarova and Dominik Wodarz. Drug resistance in cancer: Principles of emergence and prevention. *Proc. Natl. Acad. Sci. U.S.A.*, 102(27):9714–9719, 2005.

[29] Hanan Lamlum, Angeliki Papadopoulou, Mohammad Ilyas, Andrew Rowan, Cheryl Gillet, et al. APC mutations are sufficient for the growth of early colorectal adenomas. *Proc. Natl. Acad. Sci. U.S.A.*, 97(5):2225–2228, 2000.

[30] Rafael Meza, Jihyoun Jeon, Suresh H. Moolgavkar, and E. Georg Luebeck. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc. Natl. Acad. Sci. U.S.A.*, 105(42):16284–16289, 2008.

[31] Anna M. Nicholson, Cora Olpe, Alice Hoyle, Ann-Sofie Thorsen, Teja Rus, et al. Fixation and spread of somatic mutations in adult human colonic epithelium. *Cell Stem Cell*, 22(6):909–918.e8, 2018.

[32] Michael D Nicholson and Tibor Antal. Competing evolutionary paths in growing populations with applications to multidrug resistance. *PLoS Comput. Biol.*, 15(4):e1006866, 2019.

[33] Serena Nik-Zainal, Helen Davies, Johan Staaf, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534(7605):47–54, 2016.

[34] Chay Paterson, Hans Clevers, and Ivana Bozic. Mathematical model of colorectal cancer initiation. *Proc. Natl. Acad. Sci. U.S.A.*, 117(34):20681–20688, 2020.

[35] Johannes G. Reiter, Ivana Bozic, Benjamin Allen, Krishnendu Chatterjee, and Martin A. Nowak. The effect of one additional driver mutation on tumor progression. *Evol. Appl.*, 6(1):34–45, 2013.

[36] Charles J Sherr. Principles of tumor suppression. *Cell*, 116(2):235–246, 2004.

[37] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2021. *CA Cancer J. Clin.*, 71(1):7–33, 2021.

[38] Hugo J. Snippert, Arnout G. Schepers, Johan H. van Es, Benjamin D. Simons, and Hans Clevers. Biased competition between lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion. *EMBO Rep.*, 15(1):62–69, 2014.

[39] Cristian Tomasetti, Luigi Marchionni, Martin A. Nowak, Giovanni Parmigiani, and Bert Vogelstein. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.*, 112(1):118–123, 2015.

[40] Louis Vermeulen, Edward Morrissey, Maartje van der Heijden, Anna M. Nicholson, Andrea Sottoriva, et al. Defining stem cell dynamics in models of intestinal tumor initiation. *Science*, 342(6161):995–998, 2013.

[41] Kai Wang, Siu Tsan Yuen, Jiangchun Xu, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.*, 46(6):573–582, 2014.

[42] Laura D Wood, D Williams Parsons, Siân Jones, Jimmy Lin, Tobias Sjoblom, et al. The genomic landscapes of human breast and colorectal cancers. *Science*, 318(5853):1108–1113, 2007.

[43] Ruibo Zhang, Obinna A. Ukogu, and Ivana Bozic. Waiting times in a branching process model of colorectal cancer initiation. *BioRxiv*, 2022.

Appendix A

L^2 INTEGRABILITY OF THE PROCESSES IN CHAPTER 2

Lemma 25. *For $i \le 3, \sup_t \mathbb{E}\left[(e^{-\lambda_i t} N_i(t))^2\right] < \infty$.*

Proof. By Lemma 5 in [14], we know that inductively, if $\sup_t \mathbb{E}[e^{-\lambda_i t} N_i(t)]^2 < \infty$ and $\lambda_i < \lambda_{i+1}$ holds, then $\sup_t (e^{-\lambda_{i+1} t} N_{i+1})^2 < \infty$. In this case, since $0 = \lambda_0 = \lambda_1 < \lambda_2 < \lambda_3$, we only need to show that

$$\sup_t \mathbb{E}(N_0(t))^2 < \infty \text{ and, } \sup_t \mathbb{E}(N_1(t))^2 < \infty.$$

Note that by our transition scheme, one can have

$$N_0(t) + N_1(t) \le N$$

. Therefore

$$\max\left(\sup_t N_0^2(t), \sup_t N_1^2(t)\right) < N^2 < \infty.$$

□

Lemma 26. *For $i \in \{2,3\}$, $\sup_t \mathbb{E}\left[(e^{-\lambda_i t} \tilde{N}_i(t))^2\right] < \infty$.*

Proof. For $i = 2$, recall that $\tilde{N}_2(t)$ is the second type in a two-type branching process where the first type is $N_1^*(t) = u_1 N t$. And $\tilde{N}_2(t)$ is produced at rate $u_2 N_1^*(t)$. By manipulating the master equation of this two-type process, we obtain the following differential equation of $\mathbb{E}\left[\tilde{N}_2(t)^2\right]$,

$$\frac{\mathrm{d}\mathbb{E}\left[\tilde{N}_2(t)^2\right]}{dt} = 2\lambda_2 \mathbb{E}\left[\tilde{N}_2(t)^2\right] + (\lambda_2 + 2u_2 u_1 N t)\mathbb{E}[\tilde{N}_2(t)] + u_2 u_1 N t$$

subject to $\mathbb{E}\left[\tilde{N}_2(0)^2\right] = 0$. The solution is

$$\mathbb{E}\left[\tilde{N}_2(t)^2\right] = e^{2\lambda_2 t} \int_0^t e^{-2\lambda_2 s} \left((\lambda_2 + 2u_2 u_1 N s)\mathbb{E}[\tilde{N}_2(s)] + u_2 u_1 N s\right) \mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_2(t)] = \frac{Nu_1u_2(e^{\lambda_2 t} - \lambda_2 t - 1)}{\lambda_2^2} \leq \frac{u_2u_1N}{\lambda_2^2}e^{\lambda_2 t}.$$

Thus,

$$\mathbb{E}\left[\tilde{N}_2(t)^2\right] \leq e^{2\lambda_2 t}\int_0^t \left(\frac{2u_2^2u_1^2N^2}{\lambda_2^2}se^{-\lambda_2 s} + \frac{u_1u_2N}{\lambda_2}e^{-\lambda_2 s} + u_2u_1Nse^{-2\lambda_2 s}\right)\mathrm{d}s.$$

Using the fact that for $a > 0$, $\int_0^t se^{-as}\mathrm{d}s \leq 1/a^2$ and $\int_0^t e^{-as}\mathrm{d}s \leq 1/a$, we get $\sup_t \mathbb{E}\left[(e^{-\lambda_2 t}\tilde{N}_2(t))^2\right] < \infty$.

For $i = 3$, $\tilde{N}_3(t)$ is produced at rate $u_3N_2^*(t) = u_3e^{\lambda_2 t}V_2$. By the master equation we get the following differential equation:

$$\frac{\mathrm{d}\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right]}{dt} = 2\lambda_3\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right] + (\lambda_3 + 2u_3e^{\lambda_2 t}V_2)\mathbb{E}[\tilde{N}_3(t)|V_2] + u_3e^{\lambda_2 t}V_2$$

subject to $\mathbb{E}\left[\tilde{N}_3(0)^2|V_2\right] = 0$. The solution is

$$\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right] = e^{2\lambda_3 t}\int_0^t e^{-2\lambda_3 s}\left((\lambda_3 + 2u_3e^{\lambda_2 s}V_2)\mathbb{E}[\tilde{N}_3(s)|V_2] + u_3e^{\lambda_2 s}V_2\right)\mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_3(t)|V_2] = \frac{u_3V_2}{\lambda_3 - \lambda_2}(e^{\lambda_3 t} - e^{\lambda_2 t}) \leq \frac{u_3V_2}{\lambda_3 - \lambda_2}e^{\lambda_3 t}.$$

Thus we have

$$\begin{aligned} E\left[\tilde{N}_3(t)^2|V_2\right] &\leq e^{2\lambda_3 t}\int_0^t \left(\frac{2u_3^2V_2^2}{\lambda_3 - \lambda_2}e^{-(\lambda_3-\lambda_2)s} + \frac{\lambda_3u_3V_2}{\lambda_3 - \lambda_2}e^{-\lambda_3 s} + u_3e^{-(2\lambda_3-\lambda_2)s}V_2\right)\mathrm{d}s \\ &\leq e^{2\lambda_3 t}\left(\frac{2u_3^2V_2^2}{(\lambda_3 - \lambda_2)^2} + \frac{\lambda_3u_3V_2}{\lambda_3(\lambda_3 - \lambda_2)} + \frac{u_3V_2}{(2\lambda_3 - \lambda_2)}\right). \end{aligned}$$

By Lemma 16, one can compute the moments of V_2 by it's Laplace transform. Then we get

$$\mathbb{E}[V_2] = \frac{Nu_1u_2}{\lambda_2^2},\ \mathbb{E}[V_2^2] = \frac{Nu_1u_2(2Nu_1u_2 + \lambda_2^2)}{\lambda_2^4}$$

Hence we conclude that $\sup_t \mathbb{E}\left[(e^{-\lambda_3 t}\tilde{N}_3(t))^2\right] < \infty$. □

Appendix B

CLOSED FROM FORMULAS OF ANALYTIC DISTRIBUTIONS

In the main text, we have omitted a few cumbersome formulas to increase readability. Here we present their closed form expressions. We begin with the analytic probability distribution functions of τ_4.

$$p_{40}(t) = 1 - \mathscr{L}_{V_3}\left(u_4 \int_0^t f_{30}(s)ds\right) = 1 - \mathscr{L}_{V_3}\left(\frac{u_4}{\lambda_3}(e^{\lambda_3 t} - 1)\right).$$

$$\begin{aligned} p_{41}(t) &= 1 - \mathscr{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)ds\right) \\ &= 1 - \mathscr{L}_{V_3}\left(u_4\left(\frac{1}{\lambda_3}(e^{\lambda_3 t} - 1) - \frac{\lambda_3^2}{2}t^2 - t - \frac{\lambda_3^2}{\lambda_2^2}(\frac{1}{\lambda_2}(e^{\lambda_2 t} - 1) - \frac{\lambda_2}{2}t^2 - t)\right)\right). \end{aligned}$$

where

$$\begin{aligned} \mathscr{L}_{V_3}(\theta) &= \mathscr{L}_{V_2}\left(u_3 \int_0^\infty \frac{f_{21}(s)\theta}{\theta + e^{\lambda_3 s}}\mathrm{d}s\right) \\ &= \exp\left(\frac{N u_1 u_2}{\lambda_2^2}\mathrm{PolyLog}\left(2, -u_3\left(\frac{1}{\lambda_3 - \lambda_2}\theta F(1 - \frac{\lambda_2}{\lambda_3}, 1; 2 - \frac{\lambda_2}{\lambda_3}; -\theta)\right.\right.\right. \\ &\quad \left.\left.\left. -\lambda_2\frac{\mathrm{PolyLog}(2, -\theta)}{\lambda_3^2} - \frac{1}{\lambda_3}\log(1+\theta)\right)\right)\right) \end{aligned}$$

In the equation above, $F(a, b; c; z)$ is the hypergeometric function.

The two results of skipping type-4 are

$$\begin{aligned} p_{40}^{s3}(t) &= 1 - \mathscr{L}_{V_2}\left(u_4 \int_0^t f_{20}(s)(1 - p_0^{3\to 4}(s,t))ds\right) \\ &= 1 - \mathscr{L}_{V_2}\left(\frac{u_4}{\lambda_2(u_4 - \lambda_3)}\left(-u_4 + \lambda_3 F(1, \frac{\lambda_2}{\lambda_3}, \frac{\lambda_2}{\lambda_3} + 1, \frac{u_4 - \lambda_3}{u_4}e^{-\lambda_2 T})\right.\right. \\ &\quad \left.\left. + e^{\lambda_2 T}(u_4 - \lambda_3 F(1, \frac{\lambda_2}{\lambda_3}, \frac{\lambda_2}{\lambda_3} + 1, 1 - \frac{\lambda_3}{u_4}))\right)\right), \\ p_{41}^{s3}(t) &= 1 - \mathscr{L}_{V_2}\left(u_4 \int_0^t f_{21}(s)(1 - p_0^{3\to 4}(s,t))ds\right) \\ &= 1 - \mathscr{L}_{V_2}\left(u_4\left(L_{\lambda_2}^{(4)}(t) - \lambda_2 L_l^{(4)}(t) - L_c^{(4)}(t)\right)\right) \end{aligned}$$

where

$$
\begin{aligned}
p_0^{3\to4}(s,t) &= \frac{1}{1+u_4\frac{\exp\{(\lambda_3(t-s))\}-1}{\lambda_3}} \\
\mathscr{L}_{V_2}(\theta) &= \exp\left(Nu_1u_2\frac{\mathrm{PolyLog}(2,-\theta)}{\lambda_2^2}\right) \\
L_{\lambda_2}^{(4)}(t) &:= \int_0^t e^{\lambda_2 s}(1-p_0^{3\to4}(s,t))\,ds = \frac{1}{\lambda_2(u_4-\lambda_3)}\left(-u_4+\lambda_3 F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;\frac{u_4-\lambda_3}{u_4}e^{-\lambda_2 t})\right. \\
&\qquad \left. +e^{\lambda_2 t}(u_4-\lambda_3 F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;1-\frac{\lambda_3}{u_4}))\right) \\
L_l^{(4)}(t) &:= \int_0^t s(1-p_0^{3\to4}(s,t))\,ds \\
&= \frac{t^2}{2}+\frac{-\lambda_3 t\log\left(\frac{\lambda_3}{u_4}\right)+\mathrm{PolyLog}(2,\frac{-(\lambda_3-u_4)e^{-\lambda_3 t}}{u_4})-\mathrm{PolyLog}(2,1-\frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3-u_4)} \\
L_c^{(4)}(t) &:= \int_0^t (1-p_0^{3\to4}(s,t))\,ds \\
&= \frac{\log\left(u_4(e^{\lambda_3 t}-1)+\lambda_3\right)-\log(\lambda_3)-u_4 t}{\lambda_3-u_4}.
\end{aligned}
$$

For waiting time distributions of the first type-5 crypt, our estimations are $p_{51}^{s4}(t)$ and $p_{51}^{s34}(t)$. $p_{51}^{s4}(t)$ can be expressed explicitly.

$$
\begin{aligned}
p_{51}^{s4}(t) &= 1-\mathscr{L}_{V_3}\left(u_4\int_0^t f_{31}(s)(1-p_0^{4\to5}(s,t))\,ds\right) && \text{(B.1)} \\
&= \mathscr{L}_{V_3}\left(u_4\left(L_{\lambda_3}^{(5)}(t)-CL_{\lambda_2}^{(5)}(t)-(\lambda_3-C\lambda_2)L_l^{(5)}(t)+(C-1)L_c^{(5)}(t)\right)\right) && \text{(B.2)}
\end{aligned}
$$

where

$$
\begin{aligned}
L^{(5)}_{\lambda_3}(t) &:= \int_0^t e^{\lambda_3 s}(1-p_0^{4\to 5}(s,t))\,ds \\
&= \frac{u_5}{\lambda_3(\lambda_3-u_5)^2}\left(\lambda_3-u_5+e^{\lambda_3 t}\left(-\lambda_3+u_5+\lambda_3\log\left(\lambda_3 e^{\lambda_3 t}\right)-\lambda_3\log\left(\lambda_3+u_5(e^{\lambda_3 t}-1)\right)\right)\right) \\
L^{(5)}_{\lambda_2}(t) &:= \int_0^t e^{\lambda_2 s}(1-p_0^{4\to 5}(s,t))\,ds = \frac{1}{\lambda_2(u_5-\lambda_3)}\left(-u_5+\lambda_3 F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;\frac{u_5-\lambda_3}{u_5}e^{-\lambda_2 t})\right. \\
&\qquad \left. +e^{\lambda_2 t}(u_5-\lambda_3 F(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;1-\frac{\lambda_3}{u_5}))\right) \\
L^{(5)}_{l}(t) &:= \int_0^t s(1-p_0^{4\to 5}(s,t))\,ds \\
&= \frac{t^2}{2}+\frac{-\lambda_3 t\log\left(\frac{\lambda_3}{u_4}\right)+\mathrm{PolyLog}(2,\frac{-(\lambda_3-u_4)e^{-\lambda_3 t}}{u_4})-\mathrm{PolyLog}(2,1-\frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3-u_4)} \\
L^{(5)}_{c}(t) &:= \int_0^t (1-p_0^{4\to 5}(s,t))\,ds \\
&= \frac{\log\left(u_4(e^{\lambda_3 t}-1)+\lambda_3\right)-\log(\lambda_3)-u_4 t}{\lambda_3-u_4} \\
C &:= \frac{\lambda_3^2}{\lambda_2^2}.
\end{aligned}
$$

Define $I(t) =$

$$
\int_0^t (e^{\lambda_2 s}-\lambda_2 s-1)\left(1-\left(1+\frac{u_4 u_5(\lambda_3-u_5+e^{\lambda_3(t-s)}(u_5-\lambda_3+\lambda_3\log\left(\frac{\lambda_3 e^{\lambda_3(t-s)}}{\lambda_3+u_5(e^{\lambda_3(t-s)}-1)}\right)}{(\lambda_3-u_5)^2\lambda_3}\right)^{-1}\right)ds.
$$

Then we can have

$$
p_{51}^{s34}(t) = \mathscr{L}_{V_2}(u_3 I(t)) \tag{B.3}
$$

Unfortunately, we didn't find a way to solve the integral $I(t)$ explicitly. Nevertheless, we can compute this value numerically.