INFERENCE FOR AGE-DEPENDENT BRANCHING PROCESS AND THEIR APPLICATIONS

by

Xin Cao

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

_____	Dr.Anand N. Vidyashankar, Dissertation Director
_____	Dr. Jie Xu, Co-advisor and Committee Member
_____	Dr. Guoqing Diao, Committee Member
_____	Dr. Daniel Carr, Committee Member
_____	Dr. William F. Rosenberger, Department Chair
_____	Dr. Kenneth S. Ball, Dean, The Volgenau School of Information Technology and Engineering

Date: _____ Spring Semester 2015
George Mason University
Fairfax, VA

Inference for Age-dependent Branching Process and Their Applications

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Xin Cao
Master of Science
Rutgers University, 2011
Bachelor of Science
Beihang University, 2009

Director: Dr.Anand N. Vidyashankar, Associate Professor
Department of Statistics

Spring Semester 2015
George Mason University
Fairfax, VA

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Anand N. Vidyashankar for his excellent guidance, care, patience, and providing me with an excellent atmosphere for doing research. His guidance helped me in various times during the past four years. Besides, he encouraged me to not only grow as a statistician, but also as an independent thinker.

I would like to thank my co-advisor, Dr. Jie Xu, my committee members, Dr. Guoqing Diao and Dr. Daniel Carr for their guidance, valuable discussions and accessibility.

I would like to thank my wife Yayan. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past four years of my life have been built. She was always around to cheer me up and stood by me through the good and bad times. I would like to thank my parents, Zhiliang and Xiaoping, for their faith in me. They were always supporting me and encouraging me with their best wishes.

Table of Contents

List of Tables

List of Figures

Abstract

INFERENCE FOR AGE-DEPENDENT BRANCHING PROCESS AND THEIR APPLICATIONS

Xin Cao, PhD

George Mason University, 2015

Dissertation Director: Dr.Anand N. Vidyashankar

Age-dependent branching processes arise in multiple areas of scientific applications such as cell biology, epidemic spread, and medical science. Researchers have shown that the size of the process grows exponentially fast with some rate depending on the offspring distribution and the life-time distribution of the particles, but the inference for such a rate has not been well-studied in the literature.

In this thesis, we provided a new inferential methodology for age-dependent branching process under different data structures available in cell biology and other fields. We investigated and proved the consistency and asymptotic normality of the Malthusian parameter estimator. Additionally, we provided inferential methodology for the mean of offspring distribution and the limiting age distribution, which incorporate most information of an age-dependent branching process.

Under data structure with age chart information, where traditional maximum likelihood method is not applicable for estimating the parameters of the life-time distribution, we provided an alternative method based on Hellinger distance and a computational algorithm for obtaining the related parameters. The advantage of using the minimum Hellinger distance

estimator (MHDE) is the robustness when data contaminations or model mis-specifications are present.

Chapter 1: Introduction and Literature Review

1.1 Discrete Branching Process

A discrete time discrete state branching process (see Athreya and Ney (1972)) is a Markov process that models a population in which each individual in generation n produces a random number of individuals which form generation $n+1$. It is recursively defined using the following equation

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}, \quad n = 1, 2, \ldots, \tag{1.1}$$

where Z_n is the population size of the n^{th} generation, $X_{n,i}$ is the number of offspring produced by the i^{th} particle in the n^{th} generation. It is assumed that $\{X_{n,i} : i \geq 1, n \geq 1\}$ is a sequence of independent and identical distributed (i.i.d.) non-negative, integer-valued random variables with $E(X_{n,i}) = m$ and $Var(X_{n,i}) = \sigma^2$. The branching process is referred to be supercritical if $m > 1$, critical if $m = 1$, and subcritical if $m < 1$.

The process can be represented as an evolving population of particles. At time 0, there is an initial population of Z_0 particles, each of which lives one unit of time. At the time of their death, the particles split independently of the others into a random number of offsprings according to the probability function $\{p_k; k = 0, 1, 2, \ldots, p_k \geq 0, \sum p_k = 1\}$, where p_k is the probability of producing k particles. The total number of offspring particles produced by Z_0 forms the first generation Z_1. The process repeats, yielding $Z_2, Z_3, \ldots$.

In the process, the number of offsprings produced by a single particle is independent of all other particles at the present, and it is also independent with history of the process.

Note that 0 is an extinct state, since if $Z_n = 0$, then $Z_{n+k} = 0$ for all $k \geq 0$. The

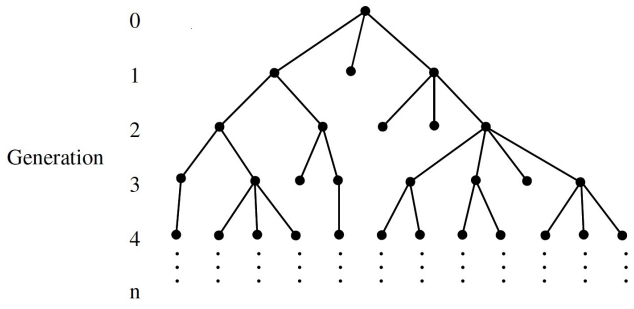


Figure 1.1: Sample Structure for Discrete Time Branching Process

extinction probability is defined to be $q = P(Z_n = 0 \text{ for some } n \geq 1)$. It is well-known that (see Harris (1963) and Athreya and Ney (1972)) the branching process dies out if $m \leq 1$ (unless $p_1 \equiv 1$) and explodes with positive probability if $m > 1$. The moment of $\{Z_n : n \geq 1\}$ can be obtained recursively using (1.1). In fact,

$$EZ_n = m^n, \tag{1.2}$$

$$Var(Z_n) = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{(m-1)} & \text{if } m \neq 1, \\ n\sigma^2 & \text{if } n = 1. \end{cases} \tag{1.3}$$

1.2 Continuous Time Markov Branching Process

In the discrete time branching process, the lifetime of each particle was one unit of time so that the process is only defined at discrete time points. Naturally, the process can be extended to continuous framework if we allow allow these life-time to be random variables. In this case, we now consider the process $\{Z(t); t \geq 0\}$, where $Z(t)$ denotes the number of particles at time t. Assuming that the life-times are independent and exponentially distributed random variables, the process is a Markov Process in continuous time and is referred to as continuous time Markov branching process. More precisely, it is defined as follows (see Athreya and Ney (1972)),

Definition 1.1. *A stochastic process* $\{Z(t,\omega); t \geq 0\}$ *on a probability space* $(\Omega, \mathcal{F}, P)$ *is called a one dimensional continuous time Markov branching process if:*

(i) its state space is the set of non-negative integers;

(ii) it is a stationary Markov chain with respect to the σ*-fields* $\mathcal{F}_t = \sigma\{Z(s,\omega); s \leq t\}$*, (for any collection* D *of real-valued, Borel measurable random variables on* $(\Omega, \mathcal{F}, P)$*,* $\sigma(D)$ *denotes the* σ*-algebra of* $\mathcal{F}$ *generated by* D*);*

(iii) the transition probability $P_{ij}(t)$ *satisfy*

$$\sum_{j=0}^{\infty} P_{ij}(t)s^j = [\sum_{j=0}^{\infty} P_{1,j}(t)s^j]^i \tag{1.4}$$

for all $i \geq 1$ *and* $|s| \leq 1$.

Properties (i) and (ii) say that $Z(t)$ is a continuous time Markov process on the integers; while (iii) characterizes the basic branching property.

One possible approach to the construction of the Markov branching process is to start with the transition probabilities

$$P_{ij}(s, s+t) = P\{Z(s+t) = j | Z(s) = i\}, \tag{1.5}$$

which, due to the assumption of time homogeneity, satisfy

$$P_{ij}(s, s+t) = P_{ij}(t), \quad t \geq 0. \tag{1.6}$$

These transition function are determined by the infinitesimal probabilities

$$0 \leq p_k, \sum_{i=0}^{\infty} p_k = 1, \tag{1.7}$$

as solutions to the Kolmogorov's forward and backward equations (see Karlin (1966))

$$\frac{d}{dt}P_{ij}(t) = -j\lambda P_{ij}(t) + a\sum_{k=1}^{j+1} kP_{ik}(t)p_{j-k+1} \quad (forward), \tag{1.8}$$

and

$$\frac{d}{dt}P_{ij}(t) = -i\lambda P_{ij}(t) + ia\sum_{k=i-1}^{\infty} p_{k-i+1}P_{kj}(t) \quad (backward), \tag{1.9}$$

with boundary conditions $P_{ij}(0+) = \delta_{ij} = 1$ for $i = j$, 0 for $i \neq j$ and $0 < \lambda < \infty$.

The generating function plays an important role in the analysis of the continuous time process. It is defined as (see Athreya and Ney (1972)):

$$f(s) = \sum_{k=0}^{\infty} p_k s^k, |s| \leq 1. \tag{1.10}$$

Setting

$$u(s) = \lambda[f(s) - s], \tag{1.11}$$

note that $m = f'(1)$ and $\alpha = u'(1) = \lambda(m-1)$. The number $0 < \lambda < \infty$, and the probability function $\{p_k; k = 0, 1, 2, \ldots\}$ contain all the information about $\{Z(t); t \geq 0\}$. Define

$$F(s,t) = \sum_{j=0}^{\infty} P\{Z(t) = j | Z(0) = 1\}s^j = \sum_{j=0}^{\infty} P_{1j}(t)s^j = Es^{Z(t)}. \tag{1.12}$$

From the Kolmogorov equations (1.8), (1.9) we can see that F(s,t) satisfies

$$\frac{\partial}{\partial t}F(s,t) = u(s)\frac{\partial}{\partial s}F(s,t), \tag{1.13}$$

and

$$\frac{\partial}{\partial t}F(s,t) = u[F(s,t)], \tag{1.14}$$

with the boundary condition

$$F(s,0) = s. \tag{1.15}$$

It can be shown that there is a unique solution to (1.13), (1.14), (1.15), yielding a generating function which uniquely determine the Markov process (see Athreya and Ney (1972)). In addition, the mean and the variance of the process can also be obtained from Kolmogorov equation. Differentiating both sides of the backward equation (1.14) with respect to s and then letting s go to 1, we obtain

$$EZ(t) = e^{\alpha t}. \tag{1.16}$$

Now under the assumption that $\sum_{k=1}^{\infty} k^2 p_k < \infty$, it follows that $m_2(t) = E(Z^2(t))$ is finite for all t. By using the backward equation we see that $m_2(t)$ satisfies the differential equation

$$\frac{d}{dt}m_2(t) = u''(1)e^{2\alpha t} + \alpha m_2(t), \tag{1.17}$$

with boundary condition $m_2(0) = 1$. The variance is then given by

$$Var(Z(t)) = \begin{cases} u''(1)\alpha^{-1}(e^{2\alpha t} - e^{\alpha t}) - e^{2\alpha t} & \text{if } \alpha \neq 0, \\ u''(1)t - e^{2\alpha t} & \text{if } \alpha = 0. \end{cases} \tag{1.18}$$

1.2.1 Birth-death Process

The birth-death process (see Karlin and Taylor (1975)) is a special case of the continuous-time Markov process where the state transitions are of only two types: "births", which increase the population by one and "deaths", which decrease the population by one. This process is specified by birth rates $\{\lambda_i\}_{i=0,\ldots\infty}$ and death rates $\{\mu_i\}_{i=0,\ldots\infty}$.

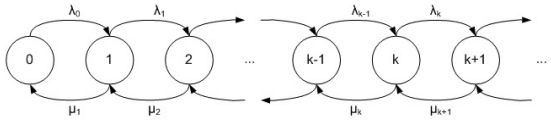


Figure 1.2: Birth-death Process

The process can be described as follows: each particle lives a random length of time (distributed as exponential distribution with parameter $\lambda + \mu$.). At the end of its life, it produces two offsprings with probability $\frac{\lambda}{\lambda+\mu}$ or produce 0 offsprings with probability $\frac{\mu}{\lambda+\mu}$. The expected number of offspring produced by the parent, namely m, equals $\frac{2\lambda}{\lambda+\mu}$. As a result, if the birth rate is greater than the death rate, the process is supercritical while if the birth rate is less than the death rate it is a subcritical process. And if they are equal, the process is critical. This model is widely used to study the evolution of cells, the number of people with a disease within a population, or the number of customers in line at the supermarket.

1.3 Age-Dependent Process

1.3.1 Single Type Age-dependent Process

In the above description, if the particles have general lifetime distribution, then the process $\{Z(t); t \geq 0\}$ is called an age-dependent branching process which was first studied by Bellman and Harris (1952). The population size process $\{Z(t); t > 0\}$ is therefor not a

Markov process since a distribution $G(\cdot)$ on $[0,\infty)$ describing the life length of the particle will also be considered. Similar to the discrete time branching process, the age-dependent branching process can be represented as follows: at time 0, there is an initial population of Z_0 particles, each of which lives a random length of time according to $G(\cdot)$. At the time of their death, the particles split independently of the others into a random number of offsprings according to the probability function $\{p_k : k = 0, 1, 2, \ldots, p_k \geq 0, \sum p_k = 1\}$, where p_k is the probability of producing k offspring particles. These offsprings again continue the same process according to life-time distribution $G(\cdot)$ and offspring distribution $\{p_k\}$. It is assumed that (see Athreya (1972)) particle productions are independent of the present state or past history of the process; and the life-time distribution are also independent of the offspring distribution.

As before, we now describe how to evaluate the generation function. To this end, consider a process starting with one particle at initial time and resulting in k particles at time t, i.e. $Z(t) = k$. A decomposition of the sample space according to the life-time and number of offspring of the initial particle suggests the following relation: the initial particle dies at time $y \quad (y < t)$, and produces $j \quad (j > 0)$ offsprings; these offsprings then produce a total of k offsprings during the rest of time $t - y$. Thus

$$F(s,t) = s[1 - G(t)] + \int_0^t f[F(s, t-y)]dG(y), |s| \leq 1. \tag{1.19}$$

This is the basic integral equation for an age-dependent branching process. Throughout the thesis, we assume $G(0+) = 0$; i.e that the probability of instantaneous death is 0. Together with the assumption of finite mean offspring distribution, this guarantees the a.s. finiteness and uniqueness of the process.

Theorem 1.1. *If* $G(0+) = 0$ *and* $m < \infty$ *then* $P\{Z(t) < \infty\} = 1$ *for each* $t \geq 0$.

The next theorem is concerned with the solution to (1.19) (see Athreya and Ney (1972)).

Theorem 1.2. *Let* f *be a probability generating function and* G *a distribution on* $[0,\infty)$

with $G(0+) = 0$. Then (1.19) has a solution $F(s,t)$, which is a generating function for each t, and which is the unique bounded solution.

One of the main interest for age-dependent branching process is the mean number of particles at time t, which is given by (see Athreya and Ney (1972))

$$\mu(t) = \frac{\partial F(s,t)}{\partial s}|_{s=1} \equiv EZ(t).$$

It is well-known that $EZ(t)$ grows exponentially fast with some growth rate (see Bellman and Harris (1952), Athreya (1972)), which is called Malthusian parameter and it is defined as follows:

Definition 1.2. *The Malthusian parameter for γ and G is the root, provided it exists, of the equation*

$$\gamma \int_0^\infty e^{-\alpha t} dG(t) = 1 \tag{1.20}$$

We denote it by $\alpha = \alpha(\gamma, G)$.

Traditionally, the term "Malthusian parameter (MP)" has been applied only to the solution $\alpha(m, G)$, where $m = f'(1) > 1$ and it is directly related to the expected number of particles at time t. As a result, the MP is very useful in describing the growth of particles under age-dependent branching process models. The following Theorem describes the relationship precisely (see Athreya and Ney (1972)).

Theorem 1.3. *If $m = 1$ then $\mu(t) \equiv 1$. If $m > 1$ and G is non-lattice, then*

$$\mu(t) \sim c' e^{\alpha t}, t \to \infty, \tag{1.21}$$

where $\alpha > 0$ is the Malthusian parameter for (m, G) and

$$c' = \frac{\int_0^\infty e^{-\alpha y}[1 - G(y)]dy}{m \int_0^\infty y e^{-\alpha y} dG(y)} = \frac{m-1}{\alpha m^2 \int_0^\infty y e^{-\alpha y} dG(y)}. \tag{1.22}$$

Through out this thesis, we will focus only on supercritical case with $m > 1$. The following results (see Athreya and Ney (1972)) show that $Z(t)$ normalized by $\mu(t)$ converges almost sure (a.s.) to a random variable W.

Theorem 1.4. *Assume that* $m > 1$.

(i) if $\sum p_k logk = \infty$ *then* $W(t) \equiv \frac{Z(t)}{c' e^{\alpha t}} \to 0$ *in probability.*

(ii) if $\sum p_k logk < \infty$ *then* $W(t)$ *converges in distribution to a non-negative random variable W having the following properties:*

(a) EW=1.

(b) $\varphi(u) = Ee^{-uW}$, $u \geq 0$, *is the unique solution of the equation*

$$\varphi(u) = \int_0^\infty f[\varphi(ue^{-\alpha y})]dG(y)$$

(c) $P(W = 0) = q \equiv P\{Z(t) = 0$ *for some* $t\}$.

(d) The distribution of W is absolutely continuous on $(0, \infty)$.

Note that $q = 0$ if and only if $p_0 = 0$. Hence, under the assumption $p_0 = 0$, $W > 0$ with probability 1. In chapter 2, we propose estimates for the Malthusian parameter α and m and establish their asymptotic properties.

Another important and useful aspect of age-dependent branching processes is the limiting behavior of the age distribution. For any realization ω of the process, let $Z(t, \omega)$ denote the total number of individuals alive at time t, $Z(x, t, \omega)$ denote the number of individuals of age less than x, and $A(\cdot, t, \omega)$ defined by $A(\cdot, t, \omega) = \frac{Z(x,t,\omega)}{Z(t,\omega)}$ denote the empirical age distribution of those alive at time t.

There has been considerable interest in the limiting behavior of the age distribution as $t \to \infty$. Harris (1952) has showed that if the offspring distribution has a second moment and the life distribution satisfies certain regularity condition, then $A(\cdot, t, \omega)$ converges to $A(\cdot)$ in distribution and referred the limiting quartile to as stable age distribution. Jagers (1975) improved this result by dropping all of the regularity assumptions on $G(\cdot)$ but still

required a finite second moment. Athreya and Kaplan (1976) showed the validity of the above result assuming that $\sum(k \log k) \cdot p_k$ is finite. Finally, Kuczek (1982) proved the almost sure convergence under the condition $1 < m < \infty$.

Theorem 1.5. *If* $1 < m < \infty$,*then*

$$lim_{t\to\infty} sup_{x\geq 0} |A(x,t,\omega) - A(x)| \stackrel{a.s.}{=} 0.$$

where

$$A(x) = \frac{\int_0^x e^{-\alpha t}(1 - G(t))dt}{\int_0^\infty e^{-\alpha t}(1 - G(t))dt}. \tag{1.23}$$

1.3.2 Multi-type Branching Process

The processes we have mentioned until now have all consisted of indistinguishable particles. Another natural generalization is to allow a number of distinguishable particles having different probabilistic behavior. In this case, to define the particle production of a p-type process, we need p generating functions, each in p variables. The i^{th} generating function, $f^{(i)}$, will determine the distribution of the number of offspring of various types to be produced by a type i particle. Thus, we set

$$f^{(i)}(s_1, \ldots, s_p) = \sum_{j_1,\ldots,j_p \geq 0} p^{(i)}(j_1, \ldots, j_p) s_1^{j_1} \cdots s_p^{j_p}, 0 \leq s_\alpha \leq 1, \alpha = 1, \ldots, p, \tag{1.24}$$

where $p^{(i)}(j_1, \ldots, j_p)$ is the probability that a type i parent produces j_1 particles of type 1, j_2 particles of type 2,..., j_p of type p. When the lifetimes in the multi-type branching process are defined to be one unit of time, exponential, and arbitrary distribution, the resulting process is referred as multi-type branching process in discrete time, multi-type Markov branching process, and multi-type age-dependent branching process respectively.

1.4 Applications for Age Dependent Branching Process

1.4.1 Dynamic Drug Resistance Model in Anti-Cancer Therapies

Cancer recurrence (or tumor recurrence) is defined as the return of cancer after treatment and after a period of time during which the cancer cannot be detected. Such recurrences are due to drug resistance, which is commonly caused by random point mutations in the genomic sequence of cells (see Baker (2005)). Although the cancer recurrence due to acquired resistance is inevitable for many therapies, the timing of recurrence varies significantly between patients (Demicheli and et al (2008) and (2012)). The source of this variability, namely, why some patients experience shorter (or longer) disease-free periods than others, is largely unknown.

A typical solid tumor has a density between 10^7 and 10^9 cancer cells per cubic centimeter. There are typically two types of cancer cells inside the tumor, namely, sensitive cells and resistant cells. They are both non-functional cells but sensitive cells can be cured by the treatment while resistant cells cannot be cured. When treatment is applied to the patient, the population of the sensitive cells will be reduced and meanwhile, the sensitive cells could mutate to the resistance cells with small probability. This mutation is typically not reversible. The tumor size is determined by the total number of these two cells. In particular, we consider the following scenario in which a population of drug-sensitive cancer cells is placed under therapy, leading to a overall decline in tumor size. However, during each replication of a sensitive cell, a mutation could arise with small probability, conferring drug-resistance (and a net positive growth rate) in cells. If such a mutation arises prior to extinction of the original population and forms a viable, growing subpopulation then the population has 'escaped' extinction. These types of escape events due to the acquired resistance cause the failure of many drugs including antibiotics, cancer therapies, and anti-viral therapies.

There has been a substantial previous work in modeling the dynamics of drug resistance in cancer. For example, Coldman and Goldie (1979) studied the emergence of resistance to

one or two drugs using stochastic processes with a differentiation hierarchy to represent the sensitive and resistant cells of a tumor . Harnevo and Agur (1992) studied drug resistance emerging due to oncogene amplification using a branching process model. Iwasa and collaborators (2003) have used multi-type branching process models to study the probability of resistance emerging due to point mutations in a variety of situations; Komarova and Wodarz (2005, 2006) also utilized a multi-type branching model to investigate the general situation in which k mutations are required to confer resistance against k drugs. However, most of these works focused on calculations of the eventual probability of developing resistance and the size of the resistant population, rather than the variations in cancer recurrence timing. Most recently, Michor and Foo (2010) evaluate the development of drug resistance and a dosing strategy in continuous and pulsed treatment schedules; the continuous treatment schedules involve administering drugs continuously at sufficiently low doses without treatment breaks and the pulse treatment schedules involve administering drugs at a higher dose in short pulses followed by a break period for toxicity recovery. The work by Foo and Michor (2010) does not take into account variability between individuals. This may cause underestimation of the variance of the predicted timing, yielding inaccurate confidence intervals or wrong dosing strategy. Our work can be extended to cancer models which yields more precise estimations.

1.4.2 Carboxy-fluorescein Diacetate Succinimidyl Ester (CFSE) Labeling Experiment

Over the past decade, carboxy-fluorescein diacetate succinimidyl ester (CFSE) labeling has become a widely used assay to study the proliferation of cell populations, lymphocytes in particular (see Hyrien (2008)). When conducted in vitro, such experiments begin by labeling several thousand cells with the fluorescent dye CFSE, which binds indiscriminately to intracellular proteins. Next, these labeled cells (ancestor cells) are placed in culture flasks and stimulated (activated) to undergo multiple successive rounds of division over several days by triggering surface receptors. During the experiment, $10^4 \sim 10^6$ individual cells are

sampled in the resulting populations, and the CFSE fluorescence intensity of thousands of these cells is measured by flow cytometry. The fluorescence intensity of any single cell is roughly proportional to the total number of CFSE molecules bound to proteins within that cell. When a cell divides, the CFSE molecules that it bears are distributed in approximately equal measure between its two daughter cells, so that the CFSE fluorescence intensities of these daughter cells are half of that of the parent cell and almost identical. Age dependent branching process is typically used to develop statistical inference in CFSE fluorescence intensities.

1.5 Structure of the Dissertation

This chapter provided basic background of branching process and literature review of related applications like cancer model, CFSE labeling experiment. Chapter 2 will focus on inferential methods for the estimation of Malthusian parameter, expected number of offspring distribution, and limiting age distribution under one-dimensional age dependent process. In chapter 3, we provide a robust method using Hellinger Distance for parameter estimation using age-chart data. Chapter 4 contains concluding remarks.

Chapter 2: Inference for Age-dependent Branching Process

Age-dependent branching processes are widely used in vitro cell experiments studies, especially in the context of cell proliferation (see Cowan (1986), Hyrien (2008), Halon and Vidyashankar (2011)) in biology and medical science (see Michor and Foo (2010)). As described in chapter 1, the growth rate of an age-dependent branching process is determined by the Malthusian Parameter (MP), which is defined as

$$m \int_0^\infty e^{-\alpha y} dG(y) = 1, \tag{2.1}$$

where m is the mean of the offspring distribution and $G(y)$ is the life-time distribution. In this chapter, we assume that $G(\cdot)$ has a density $g(\cdot)$ with respect to the Lebesgue measure, belonging to the parametric family $\mathcal{G} = \{g(\cdot;\theta); \theta \in \Theta\}$, where Θ is the parametric space and $\Theta \subset \mathbb{R}^d$, for some integer d. The MP is denoted by $\alpha = \alpha(m, G)$. The expected value of $Z(t)$ is given by

$$E(Z(t)) = c' e^{\alpha t}, \tag{2.2}$$

where

$$c' = \frac{\int_0^\infty e^{-\alpha y}[1 - G(y;\theta)]dy}{m \int_0^\infty y e^{-\alpha y} dG(y;\theta)} = \frac{m-1}{\alpha m^2 \int_0^\infty y e^{-\alpha y} g(y;\theta) dy}. \tag{2.3}$$

Thus, the estimation of α is important not only to understand the growth rate of $Z(t)$ but also to obtain an expression for the $E(Z(t))$. However, inference of the Malthusian parameter α has not been well-studied in the literature. Athreya and Keiding (1977) showed that explicit likelihood based results exist only for the offspring distribution while the estimation of MP is difficult for age-dependent branching process. Cowan (1985), Ridout

(2006) and Palmer (2008) provided approximate methods for estimating the Malthusian parameter based on the moments of life-time distributions which is not applicable when the life-times are not observable. In this chapter, we investigate methods for estimation of the MP under a variety of data structures obtainable from an age-dependent branching process.

2.1 Data Structure

Ideally, if we can observe the life-time of each particle and the number of offsprings produced in each split at all times, the age-dependent process model is fully identifiable. All the information like life-time distribution, generating function, and age distribution can be explicitly determined. In applications, however, such a detailed information is not available due to cost reasons. Based on applications in carboxy-fluorescein diacetate succinimidyl ester (CFSE) experiments (see Hyrien (2008)) and cell lineage studies (see Powell (1955), Cowan (1986) and Rohn (1932)), we can identify four different types of data structures, that are available for statistical analysis.

(D_1): The entire family tree of the age-dependent branching process is available until time point t. For each particle in the population up to time t, we denote the information about the life-time, the split-status until time t, and the number of offsprings produced by the triplet $(\tau_{i,j}(t), S_{i,j}(t), \xi_{i,j}(t))$, where $\tau_{i,j}(t)$ denotes the lifetime of the j^{th} particle in the i^{th} generation and $\xi_{i,j}(t)$ denotes the number of offsprings of the j^{th} particle in the i^{th} generation. The random variable $S_{i,j}(t)$ is defined to be

$$S_{i,j}(t) = \begin{cases} 1, & \text{if the } j^{th} \text{ particle in the } i^{th} \text{ generation splits by time t;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

Note that, $\xi_{i,j}(t)$ is defined only when $S_{i,j}(t) = 1$. Additionally, $\tau_{i,j}(t)$ is censored if $S_{i,j}(t) = 0$. In this thesis, we do not consider any censoring related statistical methods. With this

data structure, the parameters of the life-time distribution can be estimated. The generating function together with the mean number of offspring distribution m can also be estimated using the observed offspring distribution. The MP α can be then estimated using (2.1).

(D_2): The total number of particles and the life-time distribution of a subset of the members of the family tree are observable. As before, let $Z(t)$ denote the population size at time t. There are three possible sub-structures:

(i) end point assay data $Z(t)$;

(ii) paired data ($Z(t)$,$Z(t+s)$);

(iii) population sizes at time $t_1 < t_2 < \cdots < t_k$, namely $Z(t_1), Z(t_2), \ldots, Z(t_k)$.

In addition, a subset of particles from the tree are selected and their life-times are measured. In this case, the life-times can be denoted by $y_1, y_2, \ldots, y_D$, where D denotes the total number of deaths observed via the experiment. Using this data set, one can estimate the MP α, the life-time distribution $G(\cdot)$, and the mean of the offspring distribution m.

(D_3): The total number of particles and the age distribution at discrete time points are observable. In addition to the total number of particles $Z(t)$, we have the age chart $\boldsymbol{A}_t = (a_1^t, a_2^t, \ldots, a_{Z(t)}^t)$ at the time point t, where a_j^t denote the age of the j^{th} particle at time t. Then we define the empirical age distribution as

$$A(x,t) = \frac{1}{Z(t)} \sum_{j=1}^{Z(t)} \mathbf{I}_{\{a_j^t < x\}}. \tag{2.5}$$

Recall that in chapter 1, the limiting stable age distribution is defined as

$$A(x) = \frac{\int_0^x e^{-\alpha y}(1-G(y))dy}{\int_0^\infty e^{-\alpha y}(1-G(y))dy}. \tag{2.6}$$

and by Theorem 1.5,

$$lim_{t\to\infty} sup_{x\geq 0} |A(x,t) - A(x)| \overset{a.s.}{=} 0. \tag{2.7}$$

In this case, we can estimate the MP α and carryout inference concerning the parameters of life-time distribution $G(\cdot)$ and the mean number of offspring distribution m.

(D_4): Only the total number of particles at discrete time points are observable. In this case, only the MP can be estimated using the asymptotic property of age-dependent processes. Since there is no information about the life-time or age distribution, inference concerning the other parameters of the age-dependent process can not be obtained.

In this chapter, we focus on inference for the parameter (α, m, θ), where α is the MP, m is the mean of the offspring distribution and θ is the parameter associated with the life-time distribution $G(\cdot;\theta)$. We only consider three distribution for $G(\cdot;\theta)$; namely, exponential(λ), gamma(a,b), and log-normal(μ, σ^2). We emphasize here that the offspring distribution satisfies: $p_0 = 0$.

2.2 Inference for Fully Observable Family Tree Data

We begin with the description of inference using the data structure D_1. As explained previously, our data set can be represented using the triplet $(\tau_{i,j}(t), S_{i,j}(t), \xi_{i,j}(t))$, where $\tau_{i,j}(t)$ denotes the lifetime of the j^{th} particle in the i^{th} generation and $\xi_{i,j}(t)$ denotes the number of offsprings of the j^{th} particle in the i^{th} generation. The random variable $S_{i,j}(t)$ is defined to be

$$S_{i,j}(t) = \begin{cases} 1, & \text{if the } j^{th} \text{ particle in the } i^{th} \text{ generation splits by time t;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Note that, $\xi_{i,j}(t)$ is defined only when $S_{i,j}(t) = 1$. Additionally, $\tau_{i,j}(t)$ is censored if $S_{i,j}(t) = 0$. In addition, we denote the last generation before time t when a split occurred by $M(t)$ and it is defined to be

$$M(t) = \max_{\{i:B_i \neq \emptyset\}} i, \tag{2.9}$$

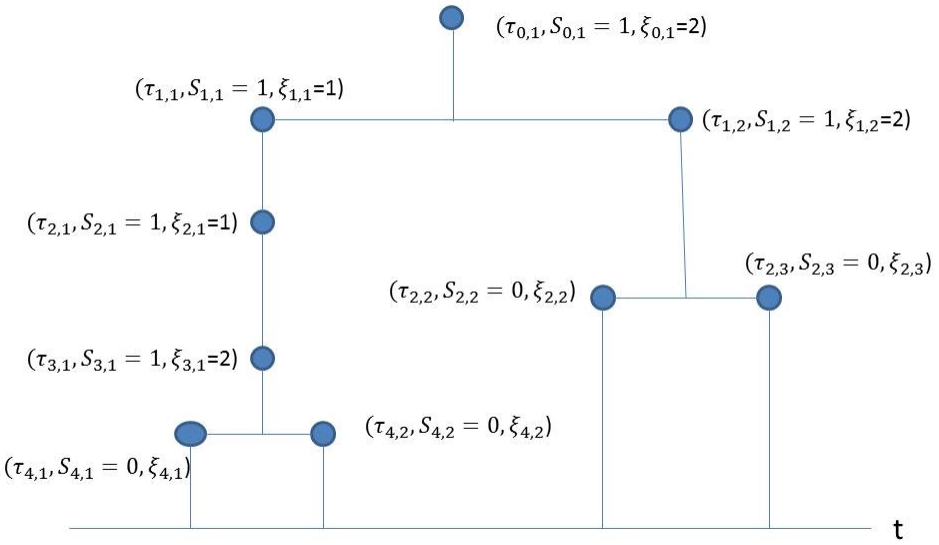


Figure 2.1: Sample data structure for fully observable family tree

where $B_i = \underset{j \geq 1}{\cup} A_{i,j}(t)$ and $A_{i,j}(t) = \{S_{i,j}(t) = 1\}$. The total number of particles in the i^{th} generation $Z_i(t)$ can then be represented as

$$Z_i(t) = \sum_{j \in C} \xi_{i-1,j}(t), \tag{2.10}$$

where $C = \{j : S_{i-1,j} = 1, 1 \leq j \leq Z_{i-1}\}$. In this case, the method of moments estimates of the offspring distribution $\{p_k, k \geq 1\}$ is given by

$$\hat{p}_k(t) = \frac{\sum_{i=0}^{M(t)} \sum_{j \in C} \mathbf{I}_{\{\xi_{i,j}(t)=k\}}}{\sum_{i=0}^{M(t)} \sum_{j=1}^{Z_i(t)} S_{i,j}(t)}. \tag{2.11}$$

Based on the above estimate of $\{p_k, k \geq 1\}$, one can estimate m using

$$\hat{m}(t) = \sum_{k=1}^{\infty} k \hat{p}_k(t) \tag{2.12}$$

We now turn to the estimation of the parameters of the life-time distribution. Recall that $g(y;\theta)$ denote the density of $G(\cdot;\theta)$. In this case, since τ_{ij}'s are observable and are i.i.d., the likelihood is given by

$$\prod_{i=1}^{M(t)} \prod_{j=1}^{Z_i(t)} g(\tau_{ij}(t);\theta) \tag{2.13}$$

Here we ignore any censoring that can possibly occur at small time t. However, at large time t, since $Z(t)$ is growing exponentially fast, the effect of censoring is typically minimized. This is especially the case for the life-time distribution under consideration. However, if the life-time distribution possess heavy-tails (for example like a Pareto distribution), the use of survival analysis based techniques may become necessary. Let $\hat{\theta}(t)$ denote the MLE of θ. Then the Malthusian parameter α is the root of

$$\int_0^\infty e^{-\alpha y} g(y;\hat{\theta}(t))dy = \frac{1}{\hat{m}(t)} \tag{2.14}$$

For exponential distribution with $\theta = \lambda$, the expression, after simplification, reduces to

$$\hat{\alpha}(t) = \hat{\lambda}(\hat{m}(t) - 1) \tag{2.15}$$

For gamma distribution with $\theta = (a, b)$, the expression reduces to

$$\hat{\alpha}(t) = \frac{\hat{m}(t)^{\frac{1}{\hat{a}}} - 1}{\hat{b}} \tag{2.16}$$

For the log-normal distribution, evaluating

$$\int_0^\infty e^{-\alpha y} g(y;\hat{\mu},\hat{\sigma}^2)dy \tag{2.17}$$

involves further complications as it is known that the Laplace transform of a log-normal distribution is a challenging problem (see Asmussen, Jensen and Rojas (2014)). We will return to this issue in section 2.4.

2.3 Inference for Partially Observable Family Tree Data

2.3.1 Estimation of Malthusian Parameter α

For estimating α, since only the total population size is available, we adopt the following methods of moments estimator: for end-point assay data $Z(t)$, we propose the following estimate for the MP α; namely

$$\hat{\alpha}_E(y) = \frac{1}{t} \log Z(t). \tag{2.18}$$

For paired-data ($Z(t)$, $Z(t+s)$), we propose the following estimate, which is similar to the Nagaev estimator (see Dion (1974)) in discrete branching process

$$\hat{\alpha}_N(t,s) = \frac{1}{s} \log \left(\frac{Z(t+s)}{Z(t)} \right). \tag{2.19}$$

For multiple time point data with same time interval t up to endpoint nt, $Z(t_1), Z(t_2), \ldots, Z(t_n) = nt$,

$$\hat{\alpha}_M(t_n) = \frac{1}{t} \log \left(\frac{\sum_{i=1}^{n} Z(t_i)}{\sum_{i=1}^{n-1} Z(t_i)} \right). \tag{2.20}$$

These are the estimators for data structure D_2, D_3 and D_4 for all life-time distributions.

Properties of the Estimators

Theorem 2.1. *Assume that the age-dependent branching process satisfies the following condition:*

(A_1) $1 < m < \infty$.

(A_2) $p_0 = 0$.

(A_3) *The life-time distribution G is non-lattice.*

Then with probability (w.p.) 1 the following hold: (i) $\lim_{t\to\infty} \hat{\alpha}_E(t) = \alpha$. *(ii)* $\lim_{t\to\infty} \hat{\alpha}_N(t,s) = \alpha$. *(iii)* $\lim_{n\to\infty} \hat{\alpha}_M(t_n) = \alpha$.

Before we start our proof of Theorem 2.1, we first focus on the estimator $\hat{\alpha}_M(t_n)$, where we have

$$t_0 = 0 < t_1 = t < t_2 < \cdots < t_n = nt, \quad t_i = it. \tag{2.21}$$

To study the behavior of the estimator

$$\hat{\alpha}_M(t_n) = \frac{1}{t} \log \left(\frac{\sum_{i=1}^{n} Z(t_i)}{\sum_{i=1}^{n-1} Z(t_i)} \right), \tag{2.22}$$

we need first investigate the behavior of $\sum_{i=1}^{n} Z(t_i)$. Our next lemma is concerned with this aspect.

Lemma 2.1. *Under the condition of Theorem 2.1:*

(A_1) $1 < m < \infty$.

(A_2) $p_0 = 0$.

(A_3) *The life-time distribution G is non-lattice.*

The following holds,

$$\lim_{t\to\infty} \frac{1}{c'} e^{-\alpha t_n} \sum_{i=1}^{n} Z(t_i) = \frac{W}{1 - e^{-\alpha t}}, \qquad \textit{w.p. 1.} \tag{2.23}$$

Proof: Note that

$$\begin{aligned}\frac{1}{c'}e^{-\alpha t_n}\sum_{i=1}^{n}Z(t_i) &= \sum_{i=1}^{n}e^{-\alpha(t_n-t_i)}\frac{Z(t_i)}{c'e^{\alpha t_i}}\\ &= \sum_{i=1}^{n}e^{-\alpha(n-i)t}W(it)\\ &= \sum_{i=0}^{n-1}W[(n-i)t]e^{-\alpha it}.\end{aligned}$$

Now as $n\to\infty$, $W[(n-i)t]\to W$. Hence, given $\epsilon>0$, there exist an n_0 such that

$$|W(nt)-W|<\epsilon, \quad \text{for all} \quad n\geq n_0. \tag{2.24}$$

Hence,

$$\begin{aligned}\sum_{i=0}^{n-1}W[(n-i)t]e^{-\alpha it} &= \sum_{i=0}^{n-n_0-1}W[(n-i)t]e^{-\alpha it}+\sum_{n-n_0}^{n}W[(n-i)t]e^{-\alpha it}\\ &= \sum_{i=0}^{n-n_0-1}\{W[(n-i)t]-W\}e^{-\alpha it}+W\sum_{i=0}^{n-n_0-1}e^{-\alpha it}\\ &+ \sum_{i=n-n_0}^{n}W[(n-i)t]e^{-\alpha it}\\ &= J_n(1)+J_n(2)+J_n(3).\end{aligned}$$

For $J_n(1)$, by (2.24), we have

$$|J_n(1)|\leq\epsilon\sum_{i=0}^{n}e^{-\alpha it}\leq\frac{\epsilon}{1-e^{-\alpha t}}. \tag{2.25}$$

Since ϵ is arbitrary, it follows

$$\lim_{n\to\infty} J_n(1) = 0, \quad \text{w.p. 1.} \tag{2.26}$$

For $J_n(3)$, we have

$$\begin{aligned} J_n(3) &= \sum_{i=n-n_0}^{n} W[(n-i)t]e^{-\alpha it} && (2.27)\\ &= \sum_{j=0}^{n_0} W(jt)e^{-\alpha(n-j)t} \quad \text{by a change of variable.} && (2.28) \end{aligned}$$

Since n_0 is fixed,

$$\lim_{n\to\infty} J_n(3) = \sum_{j=0}^{n_0} W(jt) \lim_{n\to\infty} e^{-\alpha(n-j)t} = 0. \tag{2.29}$$

Finally, we consider

$$J_n(2) = W \sum_{i=0}^{n-n_0-1} e^{-\alpha it}. \tag{2.30}$$

Thus,

$$\lim_{n\to\infty} J_n(2) = W\left[\frac{1}{1-e^{-\alpha t}}\right] = \frac{W}{1-e^{-\alpha t}}. \tag{2.31}$$

As a result,

$$\lim_{t\to\infty} \frac{1}{c'} e^{-\alpha t_n} \sum_{i=1}^{n} Z(t_i) = \frac{W}{1-e^{-\alpha t}}, \quad \text{w.p. 1.} \tag{2.32}$$

Now, we can move to the proof of Theorem 2.1.

Proof: First, we consider the convergence of $\hat{\alpha}_E(t)$.

$$\hat{\alpha}_E(t) = \frac{1}{t}\log Z(t) \tag{2.33}$$

$$= \frac{1}{t}\log\frac{Z(t)}{c'e^{\alpha t}} + \frac{1}{t}\log(c'e^{\alpha t}) \tag{2.34}$$

$$= \frac{1}{t}\log\left(W(t)\right) + \frac{1}{t}\log(c'e^{\alpha t}). \tag{2.35}$$

By Theorem 1.4, under assumptions A_1 and A_2, $\lim_{t\to\infty} W(t) = W$ w.p. 1. Hence,

$$\lim_{t\to\infty}\hat{\alpha}_E(t) = \lim_{t\to\infty}\frac{1}{t}\log\left(W(t)\right) + \lim_{t\to\infty}\frac{1}{t}\log c' + \alpha \tag{2.36}$$

$$= \alpha \qquad \text{w.p.1.} \tag{2.37}$$

Next, we consider the convergence of $\hat{\alpha}_N(t,s)$.

$$\hat{\alpha}_N(t,s) = \frac{1}{s}\log\left[\frac{Z(t+s)}{Z(t)}\right] \tag{2.38}$$

$$= \frac{1}{s}\{\log[Z(t+s)] - \log[Z(t)] + s\alpha\}. \tag{2.39}$$

Again, by Theorem 1.4,

$$\lim_{t\to\infty}\log(W(t+s)) \longrightarrow \log W, \qquad \text{w.p. 1.} \tag{2.40}$$

$$\lim_{t\to\infty}\log(W(t)) \longrightarrow \log W, \qquad \text{w.p. 1.} \tag{2.41}$$

Also $P(|\log W| < \infty) = 1$, since we assume that $p_0 = 0$. Thus taking limits as $t \to \infty$ in (2.39), we have

$$\lim_{t\to\infty}\hat{\alpha}_N(t,s) = \alpha, \qquad \text{w.p. 1.} \tag{2.42}$$

Finally, we consider the convergence of $\hat{\alpha}_M(t_n)$.

$$\begin{aligned}
\lim_{n\to\infty} \hat{\alpha}_M(t_n) &= \frac{1}{t}\log\left(\frac{\sum_{i=1}^{n} Z(t_i)}{\sum_{i=1}^{n-1} Z(t_i)}\right) && (2.43)\\
&= \lim_{n\to\infty}\frac{1}{t}\log\left(\frac{\frac{1}{c'}e^{-\alpha t_n}\sum_{i=1}^{n} Z(t_i)}{\frac{1}{c'}e^{-\alpha t_{n-1}}\sum_{i=1}^{n-1} Z(t_i)}e^{\alpha t}\right) && (2.44)\\
&= \frac{1}{t}log\left(\frac{\lim_{n\to\infty}\frac{1}{c'}e^{-\alpha t_n}\sum_{i=1}^{n} Z(t_i)}{\lim_{n\to\infty}\frac{1}{c'}e^{-\alpha t_{n-1}}\sum_{i=1}^{n-1} Z(t_i)}e^{\alpha t}\right) && (2.45)\\
&= \alpha. && (2.46)
\end{aligned}$$

Next, we investigate the asymptotic normality of the estimator $\hat{\alpha}_E(t)$.

Theorem 2.2. *Assume that the age-dependent branching process satisfies the following condition:*

(A_1) $1 < m < \infty$.

(A_2) $p_0 = 0$.

(A_3) *The life-time distribution G is non-lattice.*

$$\sqrt{Z(t)}\left[t(\alpha_E - \alpha) - \log c' - \log W - \left(c' - \frac{1}{Z(t)}\right)\right] \xrightarrow{d} N(0, c'^2 Var(W)). \qquad (2.47)$$

To prove this theorem, we start with the following lemma:

Lemma 2.2. *Assume that the age-dependent branching process satisfies the following condition:*

(A_1) $1 < m < \infty$.

(A_2) $p_0 = 0$.

(A_3) *The life-time distribution G is non-lattice.*

$$\lim_{t\to\infty}\left(\frac{\log\left(\frac{W(t)}{W}\right)}{1-\frac{W(t)}{W}}\right)=-1. \tag{2.48}$$

Proof: By Theorem 1.4, $\lim_{t\to\infty} W(t) = W$ with probability 1. Hence,

$$\lim_{t\to\infty}\frac{W(t)}{W} = 1 \tag{2.49}$$

$$\lim_{t\to\infty}\log\left(\frac{W(t)}{W}\right) = 0 \tag{2.50}$$

$$\lim_{t\to\infty}(1-\frac{W(t)}{W}) = 0. \tag{2.51}$$

By L'Hospital's rule,

$$\lim_{t\to\infty}\left(\frac{\log\left(\frac{W(t)}{W}\right)}{1-\frac{W(t)}{W}}\right)=-1. \tag{2.52}$$

Now we move to the proof of Theorem 2.2.

Proof: Let $Z(t)$ be the total number of particles at time t, and α_E is the endpoint estimate for α.

$$\begin{aligned}
\sqrt{Z(t)}\left[t(\hat{\alpha}_E-\alpha)-\log c'-\log W\right] &= \sqrt{Z(t)}\left[\log\left(\frac{Z(t)}{c'e^{\alpha t}}c'e^{\alpha t}\right)-\alpha t-\log c'-\log W\right] \\
&= \sqrt{Z(t)}\left(\log W(t)-\log W\right) \\
&= \sqrt{Z(t)}\left(\log\frac{W(t)}{W}+\log W-\log W\right) \\
&= \sqrt{Z(t)}\left[\left(\frac{\log\left(\frac{W(t)}{W}\right)}{1-\frac{W(t)}{W}}\right)\left(1-\frac{W(t)}{W}\right)\right]
\end{aligned}$$

By Lemma 2.2, $\lim_{t\to\infty}\left(\frac{\log\left(\frac{W(t)}{W}\right)}{1-\frac{W(t)}{W}}\right)=-1$, hence, we only need to evaluating $\lim_{t\to\infty}\sqrt{Z(t)}\left(1-\frac{W(t)}{W}\right)$.

Let $a_i^t, i=1,2,\ldots,Z(t)$ denote the age chart at time t. The total number of particles at time (t+s), namely Z(t+s), is the sum of all population generated by particles at time t.

$$Z(t+s)=\sum_{i=1}^{Z(t)} Z_{a_i^t}(s), \tag{2.53}$$

where $Z_{a_i^t}(s)$ is the population size generated during time interval $(t,t+s)$ by an ancestor cell with age a_i^t at time t. Let $W=\lim_{s\to\infty}W(t+s)$, we have

$$\begin{aligned} W &= \lim_{s\to\infty} W(t+s) && (2.54)\\ &= \lim_{s\to\infty}\frac{Z(t+s)}{c'e^{\alpha(t+s)}} && (2.55)\\ &= e^{-\alpha t}\lim_{s\to\infty}\sum_{i=1}^{Z(t)}\frac{Z_{a_i^t}(s)}{c'e^{\alpha s}} && (2.56)\\ &= e^{-\alpha t}\sum_{i=1}^{Z(t)} W_i, && (2.57) \end{aligned}$$

where $W_i(s)$ are i.i.d. and with same distribution as W. Then

$$\begin{aligned} \lim_{t\to\infty}\sqrt{Z(t)}\left(1-\frac{W(t)}{W}\right) &= \frac{\sqrt{Z(t)}}{W}e^{-\alpha t}\lim_{t\to\infty}\left(\sum_{i=1}^{Z(t)}(W_{x_i}(s)-\frac{1}{c'})\right)\\ &= c'\sqrt{Z(t)}\left(\left(\frac{1}{Z(t)}\sum_{i=1}^{Z(t)}W\right)-1\right)+c'\sqrt{Z(t)}(1-\frac{1}{c'Z(t)}). \end{aligned}$$

By Theorem 1.4, we have E(W)=1. Hence,

$$c'\sqrt{Z(t)}\left(\left(\frac{1}{Z(t)}\sum_{i=1}^{Z(t)}W\right)-1\right)\xrightarrow{d}N(0,c'^2Var(W)). \tag{2.58}$$

Subtracting the bias term, we have

$$\sqrt{Z(t)}\left[t(\alpha_E-\alpha)-\log c'-\log W-(c'-\frac{1}{Z(t)})\right]\xrightarrow{d}N(0,c'^2Var(W)), \tag{2.59}$$

which completes our proof.

2.3.2 Estimation for the Mean of Offspring Distribution

Under the data structure D_2, once the Malthusian parameter α is calculated, we can use the definition

$$m\int_0^\infty e^{-\alpha y}dG(y)=1 \tag{2.60}$$

to obtain the estimator of m. Now let $Y\sim G(\cdot)$, it is same as solving m by

$$E(e^{-\alpha Y})=\frac{1}{m}. \tag{2.61}$$

Exponential Distribution

When $Y\sim exp(\lambda)$, we have

$$E(e^{-\alpha Y})=\frac{\lambda}{\lambda+\alpha}=\frac{1}{m}. \tag{2.62}$$

Thus,

$$m=1+\frac{\alpha}{\lambda}. \tag{2.63}$$

The estimator of α is proposed in previous section, we now describe the estimator of λ.

Assume that we can observe the lifetime $y_1, y_2, \ldots, y_n$ in the experiment, where n is the total number of observed life-times. The likelihood function is

$$L(\lambda; y_1, y_2, \ldots, y_n) = \lambda^n exp(-\lambda \sum_{j=1}^{n} y_j). \tag{2.64}$$

The log-likelihood function is

$$l(\lambda; y_1, y_2, \ldots, y_n) = nlog(\lambda) - \lambda \sum_{j=1}^{n} y_j. \tag{2.65}$$

As a result, the maximum likelihood estimator of λ is given by

$$\hat{\lambda} = \frac{n}{\sum_{j=1}^{n} y_j}. \tag{2.66}$$

With the estimates for α and λ, we set

$$\hat{m} = 1 + \frac{\hat{\alpha}}{\hat{\lambda}}. \tag{2.67}$$

Gamma Distribution

For gamma distribution with $\theta = (a, b)$, where

$$f(y|a, b) = \frac{x^{a-1}}{\Gamma(a) b^a} exp(-\frac{x}{b}), \tag{2.68}$$

we have

$$E(e^{-\alpha Y}) = (1+\alpha b)^{-a} = \frac{1}{m} \tag{2.69}$$

$$m = (1+b\alpha)^{a}. \tag{2.70}$$

The estimator of α is proposed in previous section, we now describe the estimator of (a, b).

We assume that we can observe the lifetime $y_1, y_2, \ldots, y_n$ in the experiment, where n is the total number of observed life-time. The log-likelihood function is

$$l(a, b; y_1, y_2, \ldots, y_n) = (a-1)\sum_{j=1}^{n} \log y_j - n \log \Gamma(a) - na \log b - \frac{1}{b}\sum_{j=1}^{n} y_j.$$

The maximum likelihood estimator of b is given by

$$\hat{b} = \frac{\sum_{j=1}^{n} y_j}{na}.$$

The maximum of a is at

$$\Psi(\hat{a}) = \frac{\sum_{j=1}^{n} logy_j}{n} - log(\frac{\sum_{j=1}^{n} y_j}{n}) + loga_0,$$

where Ψ is the digamma function $\Psi(a) = \frac{\Gamma'(a)}{\Gamma(a)}$. The iteration proceeds by setting a_0 to the current $\hat{a}$, then inverting the Ψ function to get a new $\hat{a}$. Because the likelihood is concave, this iteration must converge to the (unique) global maximum. A good starting point for

the iteration is obtained via the approximation

$$\begin{aligned}
\log\Gamma(a) &\approx a\log(a) - a - \frac{1}{2}\log a + const \\
\Psi(a) &\approx \log(a) - \frac{1}{2a} \\
\hat{a} &\approx \frac{0.5}{\log\bar{y} - \overline{\log y}}.
\end{aligned}$$

With the estimates for α and (a, b), we set

$$\hat{m} = (1 + \hat{b}\hat{\alpha})^{\hat{a}} \tag{2.71}$$

2.3.3 Estimation of the Limiting Age Distribution

Recall that for any realization ω of the process, let $Z(t,\omega)$ denote the total number of individuals alive at time t, $Z(x,t,\omega)$ denote the number of individuals of age less than x, and $A(\cdot,t,\omega)$ defined by $A(\cdot,t,\omega) = \frac{Z(x,t,\omega)}{Z(t,\omega)}$ denote the empirical age distribution of those alive at time t. By Theorem 1.6,

$$lim_{t\to\infty} sup_{x\geq 0} |A(x,t,\omega) - A(x)| \stackrel{a.s.}{=} 0. \tag{2.72}$$

The limiting stable distribution $A(x)$ is defined as

$$A(x) = \frac{\int_0^x e^{-\alpha y}(1 - G(y))dy}{\int_0^\infty e^{-\alpha y}(1 - G(y))dy}. \tag{2.73}$$

Exponential Distribution

Assume that the life-time distribution is exponential, $Y \sim exp(\lambda)$. We have

$$G(y) = 1 - e^{-\lambda y}. \tag{2.74}$$

The limiting age distribution can be explicitly expressed by

$$\begin{aligned} A(x) &= \frac{\int_0^x e^{-\alpha y}(1 - G(y))dy}{\int_0^\infty e^{-\alpha y}(1 - G(y))dy} \\ &= \frac{\int_0^x e^{-(\alpha+\lambda)y}dy}{\int_0^\infty e^{-(\alpha+\lambda)y}dy} \\ &= 1 - e^{-(\alpha+\lambda)x}. \end{aligned}$$

which is an exponential distribution with parameter $\alpha + \lambda$.

Gamma Distribution

Assume that the life-time distribution is gamma, $Y \sim Gamma(a, b)$, where

$$g(y|a, b) = \frac{x^{a-1}}{\Gamma(a)b^a} exp(-\frac{x}{b}). \tag{2.75}$$

The cumulative probability distribution $G(y)$ is given by

$$G(y) = \frac{1}{\Gamma(a)}\gamma(a, \frac{y}{b}). \tag{2.76}$$

Hence, there is no closed-form solution of the limiting age distribution

$$A(x) = \frac{\int_0^x e^{-\alpha y}[1 - G(y)]dy}{\int_0^\infty e^{-\alpha y}[1 - G(y)]dy}.$$

As a result, we provided a numerical approximation for $A(x)$. For the denominator, let $v'(y) = e^{-\alpha y}$ and $u(y) = 1-G(y) = P(Y \geq y)$, we have $v(y) = -\frac{1}{\alpha}e^{-\alpha y}, u'(y) = dP(Y \geq y)$. Integral by parts,we have

$$\begin{aligned}
\int_0^\infty e^{-\alpha y}[1 - G(y)]dy &= \int_0^\infty u(y)dv(y) \\
&= u(y)v(y)|_0^\infty - \int_0^\infty v(y)du(y) \\
&= u(y)v(y)|_0^\infty - \int_0^\infty v(y)(-g(y))dy \\
&= \frac{1}{\alpha} - \frac{1}{\alpha}\int_0^\infty e^{-\alpha y}g(y)dy \\
&= \frac{1}{\alpha}(1 - (1 + \alpha b)^{-a}).
\end{aligned}$$

For the numerator, we have

$$\begin{aligned}
\int_0^x e^{-\alpha y}[1 - G(y)]dy &= \int_0^x u(y)dv(y) \\
&= u(y)v(y)|_0^x - \int_0^x v(y)du(y) \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1 - G(x)) + \frac{1}{\alpha} + \int_0^x v(y)g(y)dy \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1 - G(x)) + \frac{1}{\alpha} - \frac{1}{\alpha}\int_0^x e^{-\alpha y}g(y)dy \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1 - G(x)) + \frac{1}{\alpha} - \frac{1}{\alpha}E(e^{-\alpha Y}\mathbf{1}_{Y \leq x}).
\end{aligned}$$

Thus the stable age distribution of a gamma life-time with parameter (a, b) is given by

$$A(x) = \frac{1 - e^{-\alpha x}(1 - G(x)) - E(e^{-\alpha Y}\mathbf{1}_{Y \leq x})}{1 - (1 + \alpha b)^{-a}}. \tag{2.77}$$

2.4 Inference for Partially Observable Family Tree Data under Log-normal Life-time Distribution

When the lifetime distribution is log-normal, there is no closed form expression of (2.61) and hence m can not be expressed explicitly. Recall that a random variable X has a log-normal distribution if $X = e^Y$, where Y is a normal distribution with mean μ and variance σ^2. The density of X is

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} exp\{-\frac{(logx - \mu)^2}{2\sigma^2}\}. \tag{2.78}$$

The Laplace transform of X is

$$E(e^{-\alpha X}) = E(e^{-\alpha e^Y}) = e^{-\alpha\Delta}E(e^{-\alpha e^{Y_0}}), Y_0 \sim N(0, \sigma^2). \tag{2.79}$$

where Δ can be expressed by

$$\Delta = -\frac{1}{\alpha} log \frac{E(e^{-\alpha e^{Y_0} e^{\mu}})}{E(e^{-\alpha e^{Y_0}})}. \tag{2.80}$$

To see this, note that

$$E(e^{-\alpha e^{Y_0 + \mu}}) = e^{-\alpha\Delta}E(e^{-\alpha e^{Y_0}}) \tag{2.81}$$

$$\Delta = -\frac{1}{\alpha} log \frac{E(e^{-\alpha e^{Y_0} e^{\mu}})}{E(e^{-\alpha e^{Y_0}})}. \tag{2.82}$$

Because of (2.86), for any log-normal distribution with parameter (μ, σ^2), it is sufficient to consider the evaluation of $L(\alpha) = E(e^{-\alpha e^{Y_0}})$.

This equation has no closed form expression. The basic method to compute $L(\alpha)$ is known as Crude Monte Carlo, consists in simulating a sequence $y_1, y_2, \ldots, y_R$ of i.i.d. random variables with common distribution $LN(0, \sigma^2)$, then applying the transformation $y \to e^{-\alpha y}$ to each random variable and finally returning the arithmetic average of the transformed sequence as an estimator. The Law of Large Numbers (see Durrett (2010)) ensures unbiasedness of this estimator while the Central Limit Theorem implies the error can be made arbitrarily small by choosing R large enough. However, it is shown in Asmussen, Jensen and Rojas (2014) that

Proposition 2.1. *Let* $X \sim LN(0, \sigma^2)$. *Then* $lim_{\alpha \to \infty} \frac{Var(e^{-\alpha X})}{L(\alpha)^2} = \infty$.

This result implies that the crude Monte Carlo estimation faces the problem of a relative error that goes to infinity so that a huge value of R is required if α is large which is exactly the same issue as that arising in rare-event simulation. For this reason, we consider a second estimator which is based on the alternative representation of $L(\alpha)$ described in Asmussen, Jensen and Rojas (2014). The alternative version of Laplace transform is

$$\tilde{L}(\alpha) = \frac{1}{\sqrt{1 + \varpi(\alpha\sigma^2)}} exp\{\frac{1}{2\sigma^2}\varpi(\alpha\sigma^2)^2 + \frac{1}{\sigma^2}\varpi(\alpha\sigma^2)\}. \tag{2.83}$$

where $\varpi(\cdot)$ is the Lambert W function described as the branches of the inverse relation of the function $z = \varpi(z)e^{\varpi(z)}$. And the second estimator based on the alternative representation of $L(\alpha)$ is given by

$$\hat{L}_{IS}(\alpha) = exp\{-\frac{\varpi(\alpha\sigma^2)^2}{2\sigma^2} - \frac{\varpi(\alpha\sigma^2)}{\sigma^2}) \cdot \vartheta(Y, \alpha), \tag{2.84}$$

where $Y \sim N(0, \sigma^2)$ and $\vartheta(Y, \alpha) = exp\{-\frac{\varpi(\alpha)}{\sigma^2}(e^Y - 1 - Y)\}$.

The algorithm to compute $\hat{L}_{IS}$ for a given α is described as follows:

Algorithm 1 $\hat{L}_{IS}$ estimator

1. Simulate $Y_1, Y_2, \ldots, Y_M$ from distribution $N(0, \sigma^2)$.
2. Compute $\vartheta(Y_i, \alpha) = exp\{-\frac{\varpi(alpha\sigma^2)}{\sigma^2}(e^{Y_i} - 1 - Y_i)\}$.
3. Return $\hat{L}_{IS}(\alpha) = exp\{-\frac{\varpi(\alpha\sigma^2)^2}{2\sigma^2} - \frac{\varpi(\alpha\sigma^2)}{\sigma^2}\}\frac{1}{M}\sum_{i=1}^{M}\vartheta(Y_i, \alpha)$.

The property of $\hat{L}_{IS}(\alpha)$ has been studied in Asmussen, Jensen and Rojas (2014) and is given below:

Theorem 2.3. *The estimator $\hat{L}_{IS}(\alpha)$ is an unbiased estimator of the Laplace transform of the log-normal distribution $LN(0, \sigma^2)$ and it has bounded relative error.*

Since α is unknown, we replace α by $\hat{\alpha}$ using the methods proposed in the previous sections. Hence, the estimator of m under log-normal distribution is given by

$$\hat{m} = \frac{1}{e^{-\hat{\alpha}\Delta}E(e^{-\hat{\alpha}e^{Y_0}})} = \frac{1}{e^{-\hat{\alpha}\hat{\Delta}}\hat{L}_{IS}(\hat{\alpha})}. \tag{2.85}$$

We emphasize that Δ involves estimation using $\hat{L}_{IS}$ as well.

We now focus on the estimtors of μ and σ^2. We assume that we can observe the lifetime $y_1, y_2, \ldots, y_n$ in the experiment, where n is the total number of observed life-time. Since $Y \sim LN(\mu, \sigma^2)$, let $X \sim logY$, then $X \sim N(\mu, \sigma^2)$. We take logarithm for each y to get a set of data $x_1, x_2, \ldots, x_n$. The log-likelihood is

$$l(\mu, \sigma^2; x_1, x_2, \ldots, x_n) = (2\pi\sigma^2)^{-n/2}exp(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2). \tag{2.86}$$

The maximum likelihood estimators for μ and σ^2 are

$$\hat{\mu} = \frac{\sum_{j=1}^{n} x_j}{n}, \tag{2.87}$$

$$\hat{\sigma^2} = \frac{\sum_{j=1}^{n}(x_j - \hat{\mu})^2}{n}. \tag{2.88}$$

Assume that the life-time distribution is log-normal, $Y \sim LN(\mu, \sigma^2)$, where

$$g(y|\mu, \sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} exp(-\frac{(logy - \mu)^2}{2\sigma^2}). \tag{2.89}$$

The cumulative probability distribution $G(y)$ is given by

$$G(y) = \frac{1}{2} + \frac{1}{2} erf[\frac{logy - \mu}{\sqrt{2}\sigma}], \tag{2.90}$$

where

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \tag{2.91}$$

Hence, there is no closed form expression for the limiting age distribution

$$A(x) = \frac{\int_0^x e^{-\alpha y}[1 - G(y)]dy}{\int_0^\infty e^{-\alpha y}[1 - G(y)]dy}.$$

As a result, we provided a numerical approximation for $A(x)$. For the denominator, let $v'(y) = e^{-\alpha y}$ and $u(y) = 1 - G(y) = P(Y \geq y)$, we have $v(y) = -\frac{1}{\alpha} e^{-\alpha y}$, $u'(y) = dP(Y \geq y)$.

Integrating by parts, we have

$$\begin{aligned}
\int_0^\infty e^{-\alpha y}[1-G(y)]dy &= \int_0^\infty u(y)dv(y) \\
&= u(y)v(y)|_0^\infty - \int_0^\infty v(y)du(y) \\
&= u(y)v(y)|_0^\infty - \int_0^\infty v(y)(-g(y))dy \\
&= \frac{1}{\alpha} - \frac{1}{\alpha}\int_0^\infty e^{-\alpha y}g(y)dy \\
&= \frac{1}{\alpha}(1-E(e^{-\alpha Y})).
\end{aligned}$$

For the numerator, we have

$$\begin{aligned}
\int_0^x e^{-\alpha y}[1-G(y)]dy &= \int_0^x u(y)dv(y) \\
&= u(y)v(y)|_0^x - \int_0^x v(y)du(y) \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1-G(x)) + \frac{1}{\alpha} + \int_0^x v(y)g(y)dy \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1-G(x)) + \frac{1}{\alpha} - \frac{1}{\alpha}\int_0^x e^{-\alpha y}g(y)dy \\
&= -\frac{1}{\alpha}e^{-\alpha x}(1-G(x)) + \frac{1}{\alpha} - \frac{1}{\alpha}E(e^{-\alpha Y}\mathbf{1}_{Y\le x}).
\end{aligned}$$

Hence, the true $A(x)$ in log-normal distribution is given by

$$A(x) = \frac{1-e^{-\alpha x}(1-G(x)) - E(e^{-\alpha Y}\mathbf{1}_{Y\le x})}{1-E(e^{-\alpha Y})}. \tag{2.92}$$

2.5 Simulation Results

2.5.1 Simulation Results for Malthusian Parameter α

All our simulation results are based on generating 5000 family trees of different sizes. For each replication, we start with one single cell and record the population size at time point $2, 4, 6, 8, \ldots, 30$. For log-normal distribution, we obtain the data until time 40 since the population size grows slower. The life-time distributions are chosen to be $Exp(1)$, $Gamma(1, 1)$ and $LN(0.5, 1)$. The offspring distribution is set to be $p_2 = p_1 = 0.5$ for simplicity. The true mean of the offspring distribution is then 1.5. For α_E estimate, we only use the single data at different endpoint t. For α_N estimate, we use paired data with different combinations of endpoint t and interval s. For α_M estimate, we use different endpoint t and number of intervals n. The mean estimate and standard deviation based on 5000 replicates are given in table 2.1 to 2.9.

From the simulation results, we can see that the estimate of α_E is more biased compared to α_N and α_M. Both α_N and α_M estimates yields estimates with less bias when t exceeds 10. The advantage of α_N estimate is that it only requires a pair of data set while the α_M estimate required several time points data.

Table 2.1: α_E estimator for exponential distribution

t	Mean(standard deviation)	True α
2	0.4782(0.1612)	0.5
4	0.4857(0.1103)	0.5
6	0.4896(0.0982)	0.5
8	0.4943(0.0674)	0.5
10	0.4963(0.0435)	0.5
20	0.4986(0.0298)	0.5
30	0.5009(0.0132)	0.5

Table 2.2: α_N estimator for exponential distribution

	s=2	s=4	s=6	True α
t=2	0.4714(0.1212)	0.4736(0.1247)	0.4752(0.1138)	0.5
t=4	0.4837(0.0756)	0.4822(0.0745)	0.4834(0.0782)	0.5
t=6	0.4956(0.0229)	0.4951(0.0224)	0.4964(0.0203)	0.5
t=8	0.5032(0.0157)	0.5023(0.0145)	0.4989(0.0162)	0.5
t=10	0.5001(0.0101)	0.4998(0.0103)	0.5000(0.0102)	0.5

Table 2.3: α_M estimator for exponential distribution

	n=5	n=10	True α
t=2	0.4995(0.0212)	0.4996(0.0147)	0.5
t=4	0.4999(0.0106)		0.5
t=6	0.5000(0.0091)		0.5

Table 2.4: α_E estimator for gamma distribution

t	Mean(standard deviation)	True α
2	0.4779(0.1642)	0.5
4	0.4851(0.1124)	0.5
6	0.4899(0.0975)	0.5
8	0.4941(0.0671)	0.5
10	0.4963(0.0444)	0.5
20	0.4988(0.0299)	0.5
30	0.5007(0.0132)	0.5

Table 2.5: α_N estimator for gamma distribution

	s=2	s=4	s=8	True α
t=2	0.4721(0.1232)	0.4738(0.1239)	0.4749(0.1138)	0.5
t=4	0.4837(0.0755)	0.4831(0.0749)	0.4831(0.0780)	0.5
t=6	0.4957(0.0224)	0.4952(0.0222)	0.4964(0.0205)	0.5
t=8	0.5033(0.0151)	0.5022(0.0145)	0.4990(0.0168)	0.5
t=10	0.5000(0.0102)	0.4999(0.0102)	0.5001(0.0103)	0.5

Table 2.6: α_M estimator for gamma distribution

	n=5	n=10	True α
t=2	0.4995(0.0213)	0.4997(0.0147)	0.5
t=4	0.4998(0.0101)		0.5
t=6	0.5000(0.0089)		0.5

Table 2.7: α_E estimator for Log-normal distribution

t	Mean(standard deviation)	True α
2	0.1511(0.0612)	0.192
4	0.1723(0.0488)	0.192
6	0.1872(0.0309)	0.192
8	0.1885(0.0206)	0.192
10	0.1895(0.0147)	0.192
20	0.1902(0.0098)	0.192
30	0.1927(0.0074)	0.192
40	0.1921(0.0059)	0.192

Table 2.8: α_N estimator for Log-normal distribution

	s=2	s=4	s=8	True α
t=2	0.1632(0.0398)	0.1782(0.0392)	0.2474(0.0391)	0.192
t=4	0.1711(0.0276)	0.1762(0.0254)	0.2341(0.0243)	0.192
t=6	0.1822(0.0213)	0.1863(0.0201)	0.2083(0.0208)	0.192
t=8	0.1885(0.0187)	0.1884(0.0186)	0.1967(0.0192)	0.192
t=10	0.1898(0.0165)	0.1900(0.0160)	0.1929(0.0158)	0.192
t=16	0.1901(0.0102)	0.1908(0.0105)	0.1917(0.0101)	0.192
t=20	0.1918(0.0087)	0.1920(0.0081)	0.1918(0.0086)	0.192
t=24	0.1919(0.0075)	0.1919(0.0073)	0.1921(0.0074)	0.192
t=28	0.1918(0.0060)	0.1921(0.0059)	0.1922(0.0057)	0.192

Table 2.9: α_M estimator for Log-normal distribution

	n=5	n=10	True α
t=2	0.1882(0.0275)	0.1901(0.0277)	0.192
t=4	0.1912(0.0211)	0.1919(0.0209)	0.192
t=6	0.1919(0.0129)		0.192

2.5.2 Simulation Results for the Mean of Offspring Distribution

First of all, the offspring distribution is set to be $p_2 = p_1 = 0.5$. The true mean of the offspring distribution is then 1.5. The parameters estimation for the life-time distributions are based on maximum likelihood method described before. The Malthusian parameter are given by α_N estimate. Then the estimates for m under different distributions are given by (2.67), (2.71) and (2.85), respectively. Secondly, we adjust p_1 and p_2 to obtain different true m, which ranges from 1.1 to 1.9. The Malthusian parameter are given by α_E estimate.

In our estimation, the MLEs for parameters of life-time distribution are accurate when

the model is correctly specified. In such case, the accuracy of the estimate m is mainly depends on the accuracy of Malthusian parameter α. For log-normal, it will take longer time to get a good estimate of m compared with exponential and gamma life-time distribution, which is corresponding to the previous results in Malthusian parameter estimation. One reason is that the average splitting time is greater in log-normal life-time distribution and the other reason is that we are using Laplace transformation to obtain the estimate while there are close form solutions in both exponential and gamma life-time distributions.

Table 2.10: Mean of offspring distribution estimator for exponential distribution

	s=2	s=4	s=8	True m
t=2	1.4769(0.3211)	1.4758(0.3314)	1.4726(0.3912)	1.5
t=4	1.4873(0.1762)	1.4831(0.1712)	1.4898(0.1746)	1.5
t=6	1.4959(0.1423)	1.4961(0.1321)	1.4957(0.1248)	1.5
t=8	1.5034(0.0745)	1.5024(0.0623)	1.4981(0.0503)	1.5
t=10	1.5001(0.0356)	1.4998(0.0231)	1.5000(0.0301)	1.5

Table 2.11: Mean of offspring distribution estimator for gamma distribution

	s=2	s=4	s=8	True m
t=2	1.4769(0.3944)	1.4768(0.3984)	1.4725(0.4211)	1.5
t=4	1.4873(0.1632)	1.4832(0.1292)	1.4894(0.1566)	1.5
t=6	1.4960(0.1341)	1.4961(0.1291)	1.4958(0.1428)	1.5
t=8	1.5033(0.0866)	1.5021(0.0891)	1.4991(0.0803)	1.5
t=10	1.5002(0.0449)	1.4999(0.0341)	1.5001(0.0408)	1.5

Table 2.12: Mean of offspring distribution estimator for Log-normal distribution

	s=2	s=4	s=8	True m
t=2	1.2311(0.3921)	1.3654(0.3954)	1.7234(0.4211)	1.5
t=4	1.3842(0.2312)	1.4355(0.2411)	1.5921(0.2721)	1.5
t=8	1.4323(0.1103)	1.4887(0.1186)	1.5304(0.1098)	1.5
t=16	1.4923(0.0532)	1.4955(0.0522)	1.4979(0.0510)	1.5
t=20	1.4981(0.0311)	1.4999(0.0298)	1.4981(0.0281)	1.5
t=24	1.4976(0.0234)	1.4992(0.0221)	1.5021(0.0210)	1.5
t=28	1.4991(0.0174)	1.5011(0.0169)	1.5012(0.0171)	1.5

Table 2.13: Mean of offspring distribution estimator for exponential distribution

True m	t=10	t=20	t=30
1.1	1.0634(0.1644)	1.0787(0.0984)	1.0837(0.0411)
1.3	1.2673(0.1532)	1.2879(0.0892)	1.2998(0.0366)
1.5	1.4839(0.1241)	1.4978(0.0793)	1.4996(0.0288)
1.7	1.6938(0.1166)	1.6982(0.0602)	1.6997(0.0273)
1.9	1.8962(0.0989)	1.8989(0.0441)	1.9001(0.0258)

Table 2.14: Mean of offspring distribution estimator for gamma distribution

True m	t=10	t=20	t=30
1.1	1.0534(0.1714)	1.0801(0.1023)	1.0912(0.0723)
1.3	1.2811(0.1449)	1.2921(0.0882)	1.3007(0.0465)
1.5	1.4907(0.1132)	1.4969(0.0775)	1.5002(0.0312)
1.7	1.6945(0.0969)	1.6987(0.0622)	1.6999(0.0289)
1.9	1.8959(0.0871)	1.8988(0.0482)	1.9001(0.0271)

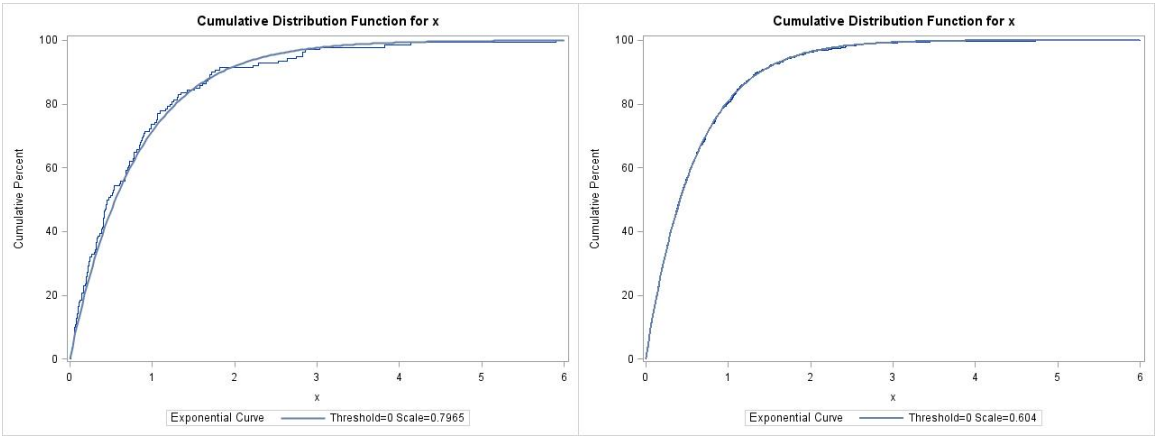
Table 2.15: Mean of offspring distribution estimator for log-normal distribution

True m	t=10	t=20	t=30	t=40
1.1	1.0417(0.1812)	1.0621(0.1201)	1.0837(0.0831)	1.0935(0.0541)
1.3	1.2521(0.1611)	1.2799(0.1008)	1.2884(0.0755)	1.2865(0.0490)
1.5	1.4754(0.1541)	1.4872(0.0952)	1.4934(0.0705)	1.4987(0.0466)
1.7	1.6812(0.1418)	1.6942(0.0887)	1.6990(0.0680)	1.6998(0.0452)
1.9	1.8931(0.1261)	1.8989(0.0852)	1.8995(0.0602)	1.9000(0.0401)

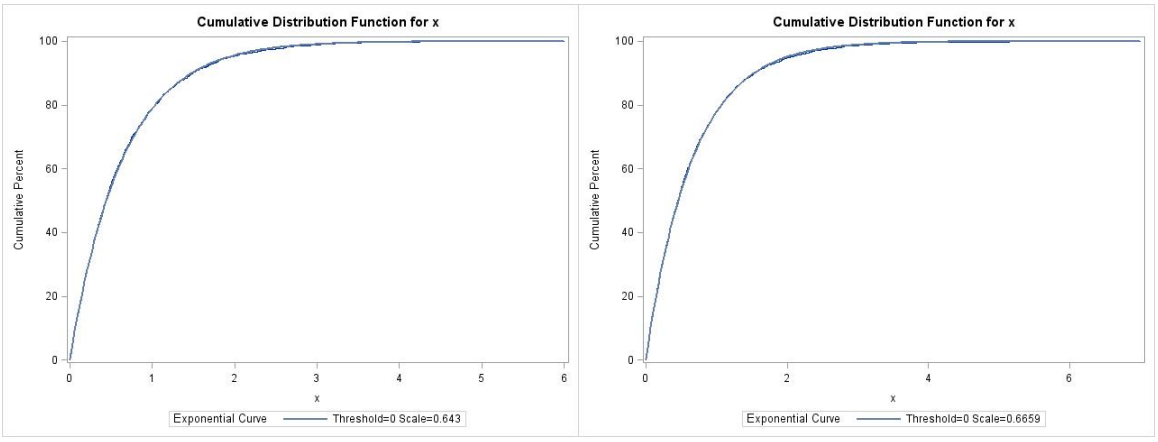
2.5.3 Simulation Results for Limiting Age Distribution

To investigate the convergence of empirical age distribution to the the stable age distribution, we obtain the age distribution data at $t = 10, 20, \ldots, 60$. For exponential life-time distribution, the true age distribution is also exponentially distributed so that we can plot both stable age distribution and empirical age distribution within the definition region. The graphs are shown in Figure 2.2 to 2.3. For gamma and log-normal life-time distribution, the stable age distributions are depending on parameters from life-time distribution but can only be obtained by numerical method. As a result, we provide the comparison between cumulative age distribution ($A(x) = P(X \leq x)$) and stable age distribution at several age point x in table 2.16 and 2.17.

For exponential life-time distribution, we fit an exponential curve for the cumulative age distribution. The true scale under stable age distribution is $\frac{1}{\lambda+\alpha} = \frac{2}{3}$. From the simulation results we provide, we can state that the empirical age distribution converges to the stable distribution when t is large. The conclusions are similar in both gamma and log-normal life-time distributions.



(a) t=10 (b) t=20



(c) t=30 (d) t=40

Figure 2.2: Empirical Age Distribution at Different Time Points

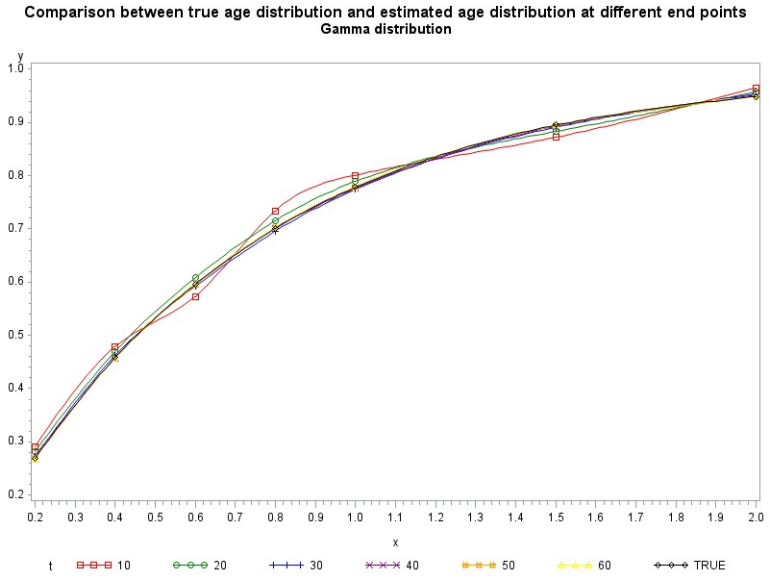


Figure 2.3: Comparison between true age distribution and empirical age distribution under gamma distribution

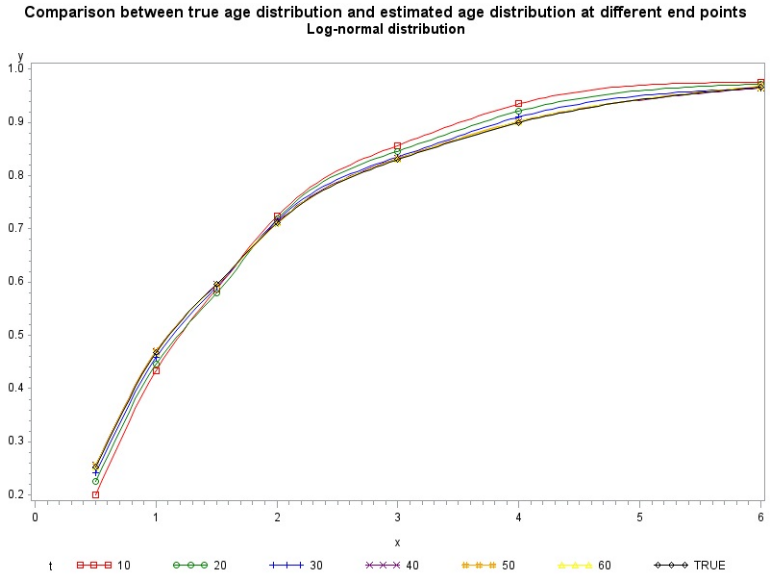


Figure 2.4: Comparison between true age distribution and empirical age distribution under log-normal distribution

Table 2.16: Comparison between true age distribution and estimated age distribution at different time point in Gamma distribution

x	Stable $A(x)$	$\hat{A}_{10}(x)$	$\hat{A}_{20}(x)$	$\hat{A}_{30}(x)$	$\hat{A}_{40}(x)$	$\hat{A}_{50}(x)$	$\hat{A}_{60}(x)$
0.2	0.269	0.291	0.282	0.273	0.274	0.270	0.268
0.4	0.458	0.479	0.470	0.462	0.457	0.457	0.458
0.6	0.596	0.573	0.608	0.592	0.597	0.594	0.597
0.8	0.701	0.734	0.716	0.696	0.701	0.702	0.702
1	0.778	0.801	0.789	0.774	0.776	0.779	0.779
1.5	0.895	0.872	0.883	0.891	0.893	0.894	0.895
2	0.949	0.965	0.957	0.953	0.951	0.949	0.950

Table 2.17: Comparison between true age distribution and estimated age distribution at different time point in Log-normal distribution

x	Stable $A(x)$	$\hat{A}_{10}(x)$	$\hat{A}_{20}(x)$	$\hat{A}_{30}(x)$	$\hat{A}_{40}(x)$	$\hat{A}_{50}(x)$	$\hat{A}_{60}(x)$
0.5	0.252	0.201	0.225	0.241	0.256	0.256	0.253
1	0.468	0.434	0.446	0.459	0.470	0.471	0.467
1.5	0.596	0.587	0.579	0.591	0.596	0.595	0.596
2	0.711	0.724	0.718	0.716	0.713	0.710	0.712
3	0.830	0.851	0.846	0.835	0.832	0.832	0.831
4	0.900	0.935	0.921	0.910	0.901	0.902	0.901
6	0.967	0.975	0.972	0.965	0.965	0.968	0.967

Chapter 3: Robust Estimation for Age-dependent Branching Process using Hellinger Distance

In the previous chapters, we provided inferential methods for data structure D_1, D_2 and D_4. Recall that the data structure D_3 in addition to the total number of particles $Z(t)$ also contains information on the age chart $\boldsymbol{A}_t = (a_1^t, a_2^t, \ldots, a_{Z(t)}^t)$ at the time point t, where a_j^t denote the age of the j^{th} particle at time t. The previously proposed methods are not applicable since the life-time distribution is not observable. Other approximation methods (see Cowan (1985), Ridout (2006) and Palmer (2008)) are not applicable either since they are based on the moments of the life-times distributions. Hence, we introduce an alternative estimation methods based on Hellinger distance to deal with the age chart data.

3.1 Minimum Hellinger distance Estimators(MHDE)

Definition: Let $g(y)$ and $h(y)$ be any two densities; the Hellinger distance between $g(y)$ and $h(y)$ is defined as the L_2-norm of the difference between square root of density functions,

$$HD^2(g,h) = ||\sqrt{g} - \sqrt{h}||_2^2 = \int \left(\sqrt{g(y)} - \sqrt{h(y)}\right)^2 dy. \tag{3.1}$$

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. real valued random variables with density belonging to a specific parametric family $\{g_\theta; \theta \in \Theta\}$ and h_n be a non-parametric estimator of the density. The Hellinger distance between g_θ and h_n is given by

$$HD_n^2(g_\theta, h_n) = ||\sqrt{g_\theta(y)} - \sqrt{h_n(y)}||_2^2. \tag{3.2}$$

The minimum Hellinger distance estimator of θ is defined to be the value $\hat{\theta}_n$, if it exists, that minimize $HD_n^2(g_\theta, h_n)$. Note that,

$$\begin{aligned} HD_n^2(g_\theta, h_n) &= \int \left(\sqrt{g_\theta(y)} - \sqrt{h_n(y)}\right)^2 dy && (3.3) \\ &= \int g_\theta(y)dy + \int h_n(y)dy - 2\int \sqrt{g_\theta(y)h_n(y)}dy && (3.4) \\ &= 2 - 2\gamma_n(\theta), && (3.5) \end{aligned}$$

where

$$\gamma_n(\theta) = \int \sqrt{g_\theta(y)h_n(y)}dy. \tag{3.6}$$

Hence, finding the minimum Hellinger distance estimator is therefore equivalent to finding the $\hat{\theta}_n$ that maximize $\gamma_n(\theta)$. In this thesis, we choose

$$h_n(y) = \frac{1}{nc_n}\sum_{j=1}^{n} K\{\frac{y - Y_j}{c_n}\}, \tag{3.7}$$

where $K(\cdot)$ is the kernel density. Then it is well known that (see Devroye (1987)) as $c_n \to 0$, $h_n(\cdot) \xrightarrow{L_1} g_\theta(\cdot)$. This implies that $HD_n^2(g_\theta, h_n) \to 0$. This argument suggest investigating estimators that minimize the Hellinger distance between the nonparametric density estimator and the proposed parametric density.

Beran (1977) have shown that the MHDE is more robust than maximum likelihood estimator when data contamination are present. Further more MHDE is known to be asymptotically efficient (see Beran (1977), Cheng and Vidyashankar (2004)) under a specified parametric family of densities and is minimax robust in a Hellinger metric neighborhood of the given family.

3.2 Hellinger Distance Estimator for Limiting Age distribution

3.2.1 Exponential Life-time Distribution

Assume that the lifetime distribution is $exp(\lambda)$ and we observe the total number of particles $Z(t)$ and age chart $a_1^t, a_2^t, \ldots, a_{Z(t)}^t$ at time t, which is the data structure D_3. The initial malthusian parameter α is obtained by the endpoint estimate we proposed last chapter. Now we consider inference for the life-time distribution parameter λ using minimum Hellinger distance. First, recall that the limiting stable age distribution $A(x)$ is given by

$$A(x) = \frac{\int_0^x e^{-\alpha y}(1 - G(y))dy}{\int_0^\infty e^{-\alpha y}(1 - G(y))dy}, \tag{3.8}$$

and the stable density function $A'(x)$ is

$$A'(x) = \frac{e^{-\alpha x}(1 - G(x))}{\int_0^\infty e^{-\alpha y}(1 - G(y))dy}. \tag{3.9}$$

The kernel density $h_{Z(t)}(x)$ for the stable density $A'(x;\lambda)$ is given by

$$h_{Z(t)}(x) = \frac{1}{Z(t)c_n}\sum_{j=1}^{Z(t)} K\{\frac{x - a_j^t}{c_n}\}. \tag{3.10}$$

The probability density function $A'(x;\lambda)$ can be obtained by differentiating A(x) and is given by

$$A'(x;\lambda) = \frac{d}{dx}(\frac{\int_0^x e^{-\alpha y}[1 - G(y)]dy}{\int_0^\infty e^{-\alpha y}[1 - G(y)]dy}) \tag{3.11}$$

$$= (\lambda + \alpha)e^{-(\lambda+\alpha)x}. \tag{3.12}$$

Recall that finding the MHDE of (λ, α) is equivalent to finding (λ, α) that maximizes the following:

$$\gamma(\lambda, \alpha) = \int \frac{(A'(x|\alpha, \lambda))^{1/2}}{(h_{Z(t)}(x))^{1/2}} (h_{Z(t)}(x))dx. \tag{3.13}$$

By strong law of large numbers (see Durrett (2010)), the above integral can be approximated by

$$\frac{1}{M} \sum_{i=1}^{M} \frac{(A'(x_{t,i}|\alpha, \lambda))^{1/2}}{(h_{Z(t)}(x_{t,i}))^{1/2}}, \tag{3.14}$$

where M is the number of the Monte Carlo samples and $x_{t,i} \sim h_{Z(t)}(\cdot)$. We use

$$O(\alpha, \lambda) = \frac{1}{M} \sum_{i=1}^{M} \omega_{t,i} (\lambda + \alpha)^{\frac{1}{2}} e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}}, \text{where} \quad \omega_{t,i} = \frac{1}{\sqrt{h_{Z(t)}(x_{t,i})}}, \tag{3.15}$$

as the objective function as in Cheng and Vidyashankar (2004). Taking the first derivative with respect to λ and α, we get

$$O'_{\lambda}(\lambda, \alpha) = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{\omega_{t,i}}{2} (\lambda + \alpha)^{-\frac{1}{2}} e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} - \frac{\omega_{t,i}}{2} x_{t,i} (\lambda + \alpha)^{\frac{1}{2}} e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} \right), \tag{3.16}$$

and

$$O'_{\alpha}(\lambda, \alpha) = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{\omega_{t,i}}{2} (\lambda + \alpha)^{-\frac{1}{2}} e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} - \frac{\omega_{t,i}}{2} x_{t,i} (\lambda + \alpha)^{\frac{1}{2}} e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} \right). \tag{3.17}$$

Since there is no closed form expression for (λ, α) from the above equation, we use Newton-Raphson method to solve for (λ, α). Taking the second derivative with respect to λ and α,

we obtain

$$\begin{aligned} O''_{\lambda}(\lambda,\alpha) &= \frac{1}{M}\sum_{i=1}^{M}[-\frac{\omega_{t,i}}{4}(\lambda+\alpha)^{-\frac{3}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} - \frac{\omega_{t,i}x_{t,i}}{4}(\lambda+\alpha)^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} \\ &- \frac{\omega_{t,i}x_{t,i}}{4}(\lambda+\alpha)^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} + \frac{\omega_i x_{t,i}^2}{4}(\lambda+\alpha)^{\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}})], \end{aligned}$$

and

$$\begin{aligned} O''_{\alpha}(\lambda,\alpha) &= \frac{1}{M}\sum_{i=1}^{M}[-\frac{\omega_{t,i}}{4}(\lambda+\alpha)^{-\frac{3}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} - \frac{\omega_{t,i}x_{t,i}}{4}(\lambda+\alpha)^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} \\ &- \frac{\omega_{t,i}x_{t,i}}{4}(\lambda+\alpha)^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}} + \frac{\omega_i x_{t,i}^2}{4}(\lambda+\alpha)^{\frac{1}{2}}e^{-\frac{1}{2}(\lambda+\alpha)x_{t,i}})]. \end{aligned}$$

Hence,

$$\begin{aligned} \lambda_{n+1} &= \lambda_n - \frac{O'_{\lambda}(\lambda,\alpha)}{O''_{\lambda}(\lambda,\alpha)}, \\ \alpha_{n+1} &= \alpha_n - \frac{O'_{\alpha}(\lambda,\alpha)}{O''_{\alpha}(\lambda,\alpha)}. \end{aligned}$$

The initial value α_0 is chosen to be $\hat{\alpha}_E$ which is described in chapter 2. For obtaining the initial value for λ_0, we use the stochastic search method in Chan and Vidyashankar (2008). We generate $\lambda_0^{(1)}, \lambda_0^{(2)}, \ldots, \lambda_0^{(N)}$ from exponential distribution and pair it with $\hat{\alpha}_E$. The initial λ is defined as

$$\lambda_0 = \underset{1\le i\le N}{\operatorname{argmax}} O(\lambda_0^{(i)}, \hat{\alpha}_E). \tag{3.18}$$

Thus, the initial estimate is $(\lambda_0, \hat{\alpha}_E)$. The algorithm can be described as follows:

Algorithm 2

1. Generate one random variable from the kernel density. Repeat M times. Set up the initial value $(\lambda_0, \hat{\alpha}_E)$.
2. Obtain the updates (λ, α) using Newton Raphson iteration.
3. When $|\lambda_{n+1} - \lambda_n| < 10^{-3}$ and $|\alpha_{n+1} - \alpha_n| < 10^{-3}$ then stop; else go to step 2.
4. Return (λ, α) as our MHDE.

We can also use the stochastic search method for further iteration rather than the Newton-Raphson to obtain the MHDE. The idea is that we start with an initial pair (λ_0, α_0). Then we generate N random combinations of $\lambda_1^{(i)}, \alpha_1^{(i)}, 1 \leq i \leq \infty$ in the neighborhood of the initial pair, for example, $(\lambda_0 - \epsilon_1, \lambda_0 + \epsilon_1)$ and $(\alpha_0 - \epsilon_2, \alpha_0 - \epsilon_2)$. The updated estimate is given by

$$(\lambda_1, \alpha_1) = \underset{1 \leq i \leq N}{\operatorname{argmax}} \quad O(\lambda_1^{(i)}, \alpha_1^{(i)}). \tag{3.19}$$

The algorithm to find MHDE using stochastic search can be described as follows:

Algorithm 3

1. Generate one random variable from the kernel density. Repeat M times. Set up the initial value $(\lambda_0, \hat{\alpha}_E)$.
2. Obtain the updates (λ, α) using stochastic search algorithm.
3. When $\frac{|O(\lambda_{n+1}, \alpha_{n+1}) - O(\lambda_n, \alpha_n)|}{|O(\lambda_n, \alpha_n)|} < 10^{-6}$ then stop; else go to step 2.
4. Return (λ, α) as our MHDE for gamma distribution.

The results of both implementations are shown in section 3.4.

3.2.2 Gamma Life-time Distribution

We now turn to the gamma life-time distribution with parameter (a, b). In this case, the limiting stable age density function is given by

$$\begin{aligned} A'(x|\alpha, a, b) &= \frac{e^{-\alpha x} S(x)}{\frac{1}{\alpha}(1 - (1 + \alpha b)^{-a})} \\ &= \frac{\alpha e^{-\alpha x} S(x)}{1 - (1 + \alpha b)^{-a}}. \end{aligned}$$

Recalling that finding the MHDE of (α, a, b) is equivalent to finding (α, a, b) that maximizes the following:

$$\frac{1}{M}\sum_{i=1}^{M}\frac{(A'(x_{t,i}|a,b))^{1/2}}{(h_{Z(t)}(x_{t,i}))^{1/2}}, \tag{3.20}$$

where M is the number of the Monte Carlo samples and $x_{t,i} \sim h_{Z(t)}(\cdot)$. The objective function in the gamma case becomes

$$O(\alpha, a, b) = \frac{1}{M}\sum_{i=1}^{M}\omega_{t,i}\alpha^{\frac{1}{2}}e^{-\frac{1}{2}\alpha x_{t,i}}S(x_{t,i})^{\frac{1}{2}}(1-(1+\alpha b)^{-a})^{-\frac{1}{2}}, \tag{3.21}$$

where $\omega_{t,i} = \frac{1}{\sqrt{h_{Z(t)}(x_{t,i})}}$ and $S(x_{t,i}) = 1 - \frac{\gamma(a,\frac{x_{t,i}}{b})}{\gamma(a)} = 1 - \frac{\int_0^{x_{t,i}/b} t^{a-1}e^{-t}dt}{\int_0^{\infty} t^{a-1}e^{-t}dt}$.

Taking the first derivative with respect to a. Set $u(a,b) = S(x_{t,i})^{\frac{1}{2}}$ and $v(a,b) = (1-(1+\alpha b)^{-a})^{-\frac{1}{2}}$. Then

$$\frac{du(a,b)}{da} = -\frac{1}{2}S(x_{t,i})^{-\frac{1}{2}}\left(\frac{\int_0^{x_{t,i}/b} t^{a-1}e^{-t}dt \cdot \int_0^{\infty} log(a-1)t^{a-1}e^{-t}dt}{(\int_0^{\infty} t^{a-1}e^{-t}dt)^2}\right) \tag{3.22}$$

$$+ \frac{1}{2}S(x_{t,i})^{-\frac{1}{2}}\left(\frac{\int_0^{x_{t,i}/b} log(a-1)t^{a-1}e^{-t}dt \cdot \int_0^{\infty} t^{a-1}e^{-t}dt}{(\int_0^{\infty} t^{a-1}e^{-t}dt)^2}\right), \tag{3.23}$$

$$\frac{dv(a,b)}{da} = -\frac{1}{2}(1-(1+\alpha b)^{a})^{-\frac{3}{2}}(-log(1+\alpha b)(1+\alpha b)^{-a}), \tag{3.24}$$

$$O'_a(\alpha, a, b) = \frac{1}{M}\sum_{i=1}^{M}\omega_{t,i}\alpha^{\frac{1}{2}}e^{-\frac{1}{2}\alpha x_{t,i}}\left(\frac{du(a,b)}{da}\cdot v(a,b) + \frac{dv(a,b)}{da}\cdot u(a,b)\right). \tag{3.25}$$

Since (α, a, b) can not be obtained from the above equation, and the Newton-Raphson method requires the second derivative which are too complex to compute. As a result, we use the stochastic search method to find the MHDE.

As before, the initial α_0 is chosen to be $\hat{\alpha}_E$. And we generate $(a_0^{(1)}, a_0^{(2)}, \ldots, a_0^{(N)}$, $b_0^{(1)}, b_0^{(2)}, \ldots, b_0^{(N)})$ from exponential distribution and pair them with $\hat{\alpha}_E$. The initial pair (a_0, b_0) is defined as

$$(a_0, b_0) = \underset{1 \le i \le N}{\operatorname{argmax}} \quad O(\hat{\alpha}_E, a_0^{(i)}, b_0^{(i)}). \tag{3.26}$$

Once the initial pair is defined, we again generate random combinations in the neighborhood of the initial pair and select (α, a, b) to maximize $O(\alpha, a, b)$.

The algorithm can be described as follows:

Algorithm 4

1. Generate one random variable from the kernel density. Repeat M times. Set up the initial value (α_0, a_0, b_0).
2. Obtain the updates (α, a, b) using stochastic search algorithm.
3. When $\frac{|O(\alpha_{n+1}, a_{n+1}, b_{n+1}) - O(\alpha_n, a_n, b_n)|}{|O(\alpha_n, a_n, b_n)|} < 10^{-6}$ then stop; else go to step 2.
4. Return (α, a, b) as our MHDE for gamma distribution.

The result of MHDE using stochastic search method is shown in section 3.4.

3.2.3 Log-normal Life-time Distribution

We now consider the log-normal life-time distribution with parameter (μ, σ^2). The limiting stable age density is given by

$$A'(x|\alpha, \mu, \sigma^2) = \frac{e^{-\alpha x} S(x)}{\frac{1}{\alpha}(1 - E(e^{-\alpha Y}))}. \tag{3.27}$$

The objective function in this case is given by

$$O(\alpha, \mu, \sigma^2) = \frac{1}{M} \sum_{i=1}^{M} \omega_{x,i} \alpha^{\frac{1}{2}} e^{-\frac{1}{2}\alpha x_{i,t}} S(x_{i,t})^{\frac{1}{2}} (1 - E(e^{-\alpha y}))^{\frac{1}{2}}, \tag{3.28}$$

where $\omega_{x,i} = \frac{1}{\sqrt{h_{Z(t)}(x_{i,t})}}$ and $S(x_{i,t}) = 1 - \Phi\left(\frac{\log x_{i,t} - \mu}{\sigma}\right)$.

Similar to the gamma distribution, (α, μ, σ^2) can not be obtained from the above equation, and the Newton-Raphson method requires the second derivative which are too complex to compute. As a result, we use the stochastic search method as described in Chan and Vidyashankar (2008) to find the MHDE. The initial α_0 is chosen to be $\hat{\alpha}_E$. And we generate $\mu_0^{(1)}, \mu_0^{(2)}, \ldots, \mu_0^{(N)}$ from normal distribution and $\sigma_0^{2,(1)}, \sigma_0^{2,(2)}, \ldots, \sigma_0^{2,(n)}$ from exponential distribution then pair them with $\hat{\alpha}_E$. The initial pair (μ_0, σ_0^2) is defined as

$$(\mu_0, \sigma_0^2) = \underset{1 \le i \le N}{\operatorname{argmax}} \quad O(\hat{\alpha}_E, \mu_0^i, \sigma_0^{2,(i)})). \tag{3.29}$$

Once the initial pair is defined, we again generate random combinations in the neighborhood of the initial pair and select (α, μ, σ^2) to maximize $O(\alpha, \mu, \sigma^2)$. The algorithm can be described as follows:

Algorithm 5

1. Generate one random variable from the kernel density. Repeat M times. Set up the initial value $(\alpha_0, \mu_0, \sigma_0^2)$.
2. Obtain the updates (α, μ, σ^2) using stochastic search algorithm.
3. When $\frac{|O(\alpha_{n+1}, \mu_{n+1}, \sigma_{n+1}^2) - O(\alpha_n, \mu_n, \sigma_n^2)|}{|O(\alpha_n, \mu_n, \sigma_n^2)|} < 10^{-6}$ then stop; else go to step 2.
4. Return (α, μ, σ^2) as our MHDE for log-normal distribution.

3.3 Robustness Estimator when Data Contamination and Model Mis-specification are Present

When the limiting stable age distribution is parametric, for example, exponential distribution, we could use maximum likelihood method to estimate the parameters. However, the maximum likelihood yields more bias when data contaminations are present. In such case, our proposed method based on Hellinger distance outperforms the maximum likelihood method due to the robustness.

Consider the following life-time models with contamination,

$$G^*(y|\theta) = (1-\beta)G(y|\theta) + \beta U, \tag{3.30}$$

where $G(y|\theta)$ is the true life-time distribution, U is a constant which is away from the the true distribution and β is the proportion of contamination.

Assume that we observe the total number of particles $Z(t)$ and age chart $a_1^t, a_2^t, \ldots, a_{Z(t)}^t$ at time t. Since the age distribution is exponential distributed with parameter (λ, α), the likelihood function is

$$L(\lambda, \alpha; a_1^t, a_2^t, \ldots, a_{Z(t)}^t) = (\lambda+\alpha)^n exp\left(-(\lambda+\alpha)\sum_{j=1}^{Z(t)} a_j^t\right). \tag{3.31}$$

The log-likelihood function is

$$l(\lambda, \alpha; a_1^t, a_2^t, \ldots, a_{Z(t)}^t) = nlog(\lambda+\alpha) - (\lambda+\alpha)\sum_{j=1}^{Z(t)} a_j^t. \tag{3.32}$$

Hence, the maximum likelihood estimator of $\lambda+\alpha$ is given by

$$\hat{\lambda}+\hat{\alpha} = \frac{n}{\sum_{j=1}^{Z(t)} a_j^t}. \tag{3.33}$$

In addition, the estimator for α can be obtained by α_E from last chapter. Hence, the estimators for α and λ based on maximum likelihood methods are given by

$$\hat{\alpha} = \frac{1}{t}\log Z(t), \tag{3.34}$$

$$\hat{\lambda} = \frac{n}{\sum_{j=1}^{Z(t)} a_j^t} - \hat{\alpha}. \tag{3.35}$$

The algorithms for MHDE under contamination model are exactly the same as we described in previous section. For exponential life-time distribution, we perform simulations to compare all three estimators, MLE, MHDE using Newton method and MHDE using stochastic search. For gamma and log-normal life-time distribution, we only provide the MHDE using stochastic search since MLE can not be obtained.

Other than the contamination model, the parametric model of the life-time distribution could be mis-specified. For example, the true life-time distribution is gamma while we assume it to be exponential. In such case, we compare the empirical age distribution from the gamma life-time distribution and fitted age distribution under exponential model to validate the robustness of MHDE.

3.4 Simulation Results

All our simulation results are based on generating 1000 family trees of different sizes. The life-time distributions are chosen to be $Exp(1)$, $Gamma(1,1)$ and $LN(0.5,1)$. The offspring distribution is set to be $p_2 = p_1 = 0.5$. For each replication, we start with single ancestor particle and obtain the age chart together with the total number of particles at time point $10, 20$, and 30. For log-normal distribution, we obtain the data until time 40. Then the kernel density function is estimated from the age chart data and the Malthusian parameters are estimated using the α_E estimator. Once the kernel density is built, 1000 observations $x_1, x_2, \ldots, x_{1000}$ are generated based on the density function.

For exponential life-time distribution, we can use both Newton method and stochastic search method. We start with generating $\lambda_0^{(1)}, \lambda_0^{(2)}, \ldots, \lambda_0^{(1000)}$ from exponential distribution with parameter 2 and pair it with $\hat{\alpha}_E$. The initial λ is defined as

$$\lambda_0 = \underset{1 \le i \le 1000}{\operatorname{argmax}} O(\lambda_0^{(i)}, \hat{\alpha}_E). \tag{3.36}$$

Then we use algorithm 2 and 3 to find MHDE. The neighborhood range is set to be 0.1. The mean and standard deviation for MHDE of both λ and α are provided in Table 3.1 and 3.2.

Table 3.1: MHDE for α using Newton method and stochastic search method under exponential life-time distribution

	Newton	Stochastic Search	True α
t=10	0.5001(0.0016)	0.5003(0.0107)	0.5
t=20	0.5000(0.0008)	0.5001(0.0099)	0.5
t=30	0.5000(0.0005)	0.5001(0.0070)	0.5

Table 3.2: MHDE for λ using Newton method and stochastic search method under exponential life-time distribution

	Newton	Stochastic Search	True λ
t=10	1.0000(0.0010)	1.0008(0.0098)	1
t=20	1.0000(0.0005)	0.9996(0.0076)	1
t=30	1.0000(0.0003)	0.9999(0.0051)	1

3.4.1 Gamma Life-time Distribution

For gamma life-time distribution, we only use stochastic search method. We start with generating $a_0^{(1)}, a_0^{(2)}, \ldots, a_0^{(N)}$, $b_0^{(1)}, b_0^{(2)}, \ldots, b_0^{(N)}$ from exponential distribution with parameter 2 and pair them with $\hat{\alpha}_E$. The initial pair (a_0, b_0) is defined as

$$(a_0, b_0) = \underset{1 \le i \le N}{\operatorname{argmax}} O(\hat{\alpha}_E, a_0^{(i)}, b_0^{(i)}). \tag{3.37}$$

Then we use algorithm 4 to find MHDE. The neighborhood range is set to be 0.1. The mean and standard deviation for MHDE are provided in table 3.3.

Table 3.3: MHDE using stochastic search method under gamma life-time distribution

	a	b	α	True value
t=10	1.0010(0.0101)	0.9988(0.0109)	0.5002(0.0121)	$(a = 1, b = 1, \alpha = 0.5)$
t=20	1.0008(0.0085)	1.0002(0.0091)	0.5001(0.0099)	$(a = 1, b = 1, \alpha = 0.5)$
t=30	1.0004(0.0073)	0.9999(0.0074)	0.5000(0.0078)	$(a = 1, b = 1, \alpha = 0.5)$

3.4.2 Log-normal Life-time Distribution

The initial α_0 is chosen to be $\hat{\alpha}_E$. We generate $\mu_0^{(1)}, \mu_0^{(2)}, \ldots, \mu_0^{(N)}$ from normal distribution $N(0, 1)$ and $\sigma_0^{2,(1)}, \sigma_0^{2,(2)}, \ldots, \sigma_0^{2,(N)}$ from exponential distribution with parameter 2. The initial pair of (μ_0, σ_0^2) is given by

$$(\mu_0, \sigma_0^2) = \underset{1 \leq i \leq N}{\operatorname{argmax}} O(\hat{\alpha}_E, \mu_0^i, \sigma_0^{2,(i)})). \tag{3.38}$$

Then we use algorithm 5 to find MHDE. The neighborhood range is set to be 0.1. The mean and standard deviation for MHDE are provided in table 3.4.

Table 3.4: MHDE using stochastic search method under log-normal life-time distribution

	μ	σ^2	α	True value
t=10	0.4173(0.1101)	0.9881(0.1013)	0.2028(0.0612)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
t=20	0.4531(0.0798)	1.0139(0.0844)	0.2001(0.0387)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
t=30	0.4963(0.0565)	0.9961(0.0521)	0.1910(0.0204)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
t=40	0.5015(0.0351)	1.0012(0.0330)	0.1917(0.0138)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$

3.4.3 Model Contamination

In this case, in addition to our previous simulation settings, the contaminated life-times are set to be twice of the mean life-time distribution. To be specific, 2 for exponential and gamma distribution, and $2e$ for log-normal distribution. The proportion of contamination ranges from 5% to 20%. The comparison between MLE and MHDE are provided in Table 3.5 to 3.6 for exponential life-time distribution. For gamma and log-normal distribution, only MHDE estimates are provided.

Table 3.5: α estimator comparison under exponential contamination models

	MLE	Newton	Stochastic Search	True α
$\beta = 5\%$	0.4876(0.0226)	0.4956(0.0102)	0.4966(0.0141)	0.5
$\beta = 10\%$	0.4623(0.0218)	0.4922(0.0098)	0.4921(0.0147)	0.5
$\beta = 15\%$	0.4417(0.0198)	0.4901(0.0102)	0.4899(0.0139)	0.5
$\beta = 20\%$	0.4208(0.0251)	0.4868(0.0110)	0.4870(0.0140)	0.5

Table 3.6: λ estimator comparison under exponential contamination models

	MLE	Newton	Stochastic Search	True λ
$\beta = 5\%$	0.9960(0.0322)	0.9982(0.0153)	0.9987(0.0201)	1
$\beta = 10\%$	0.9602(0.0311)	0.9913(0.0177)	0.9915(0.0217)	1
$\beta = 15\%$	0.9300(0.0319)	0.9833(0.0182)	0.9854(0.0231)	1
$\beta = 20\%$	0.8910(0.0352)	0.9668(0.0185)	0.9681(0.0225)	1

Table 3.7: MHDE under gamma contamination model

	a	b	α	True value
$\beta = 5\%$	0.9990(0.0191)	1.0001(0.0199)	0.4967(0.0141)	$(a = 1, b = 1, \alpha = 0.5)$
$\beta = 10\%$	0.9976(0.0185)	0.9987(0.0162)	0.4921(0.0148)	$(a = 1, b = 1, \alpha = 0.5)$
$\beta = 15\%$	0.9878(0.0175)	0.9867(0.0174)	0.4899(0.0139)	$(a = 1, b = 1, \alpha = 0.5)$
$\beta = 20\%$	0.9765(0.0179)	0.9752(0.0181)	0.4870(0.0141)	$(a = 1, b = 1, \alpha = 0.5)$

Table 3.8: MHDE under log-normal contamination model

	μ	σ^2	α	True value
$\beta = 5\%$	0.4427(0.0801)	0.9958(0.0819)	0.1973(0.0349)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
$\beta = 10\%$	0.4397(0.0785)	0.9910(0.0791)	0.1926(0.0351)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
$\beta = 15\%$	0.4336(0.0790)	0.9846(0.0794)	0.1876(0.0340)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$
$\beta = 20\%$	0.4289(0.0802)	0.9786(0.0801)	0.1803(0.0328)	$(\mu = 0.5, \sigma^2 = 1, \alpha = 0.192)$

3.4.4 Model Mis-specification

In this case, we assume that the true life-time distribution is Gamma(0.5,1). However, during the inference, we mis-specifies it as exponential distribution. With the given age-chart data coming from the gamma life-time distribution, we use the proposed algorithm for exponential life-time distribution to obtain the estimators of parameters. The empirical age distribution and fitted age distribution are given in figure 3.1.

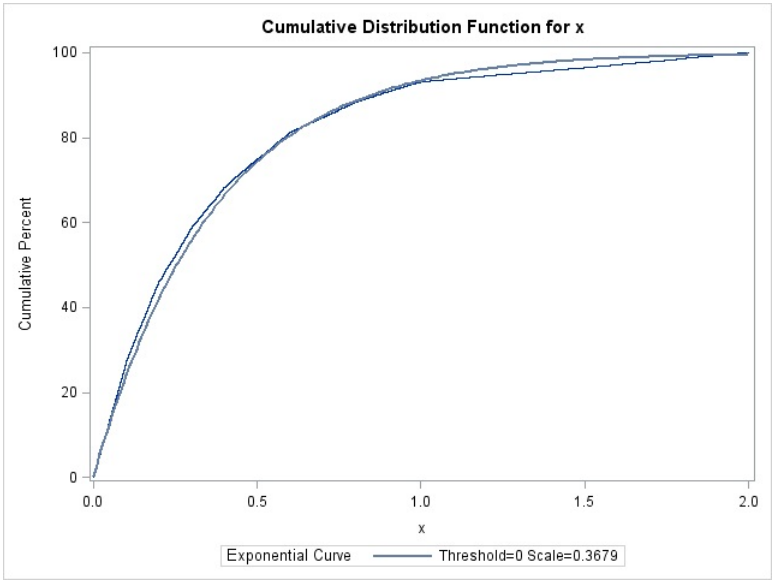


Figure 3.1: Empirical age distribution v/s Fitted age distribution

Chapter 4: Discussion and Concluding Remark

In this thesis, we provided a new inferential methodology for age-dependent branching process under different data structures available in cell biology and other fields. We investigated and proved the consistency and asymptotic normality of the Malthusian parameter estimator. Additionally, we provided inferential methodology for the mean of offspring distribution and the limiting age distribution, which incorporate most information of an age-dependent branching process. Generally speaking, the estimators are less biased in exponential and gamma life-time distribution compared to log-normal distribution. One reason is that, it will take longer time to obtain comparable population size using log-normal distribution. The second reason is that Laplace transform of log-normal distribution is harder to compute while there are explicit expressions for other distributions considered in the thesis.

Under data structure D_3 with age chart information, where traditional maximum likelihood method is not applicable for estimation of life-time distribution, we provided alternative statistical method based on Hellinger distance and computational algorithm for obtaining related parameters. The advantage of using the minimum Hellinger distance estimator (MHDE) is the robustness when data contaminations or model mis-specifications are present. MHDE can be also applied to data structure D_2.

Our inferential methodology for single-type age-dependent branching processes can be extended to multi-type framework, for example, in cancer models with resistant cells and sensitive cells. We assume that each sensitive cell or resistant cell lives a random length of time T. At the time of death, the cell is replaced by either 0 or 2 cells of same type. That is to say, the offspring distribution has all of its mass at that points 0 and 2. The existing model assumes that all patients have identical exponential life-time distribution of cells (see Foo and Michor (2010)). Under this assumption, the variations between patients may be underestimated. In addition, the reason why some patients experience shorter (or

longer) disease-free periods than others are not explained. In a typical clinical trial, the recurrence time can vary from 6 month to 10 years for patients with same cancer type and similar cancer progression (see Baker (2005)). Hence, the assumption that the life-times are identically distributed may not be true.

To account for the variability due to the differences between patients, one could assume that the parameter of the exponential distribution are not constant but randomly chosen from a distribution. This distribution is referred to as the random effect distribution. Once the random effect is incorporated in the model, the Markovian properties are lost and the resulting process turns out to be an multi-type age-dependent branching process. In such case, extension of our results to this setting will facilitate more accurate estimators of the recurrence time distribution.

There are also limitations to this research. First of all, we only applied the methodology to three commonly used continuous non-negative life-time distributions while there are several others available. Secondly, we did not consider the dependence between sibling particles and parents-children particles. We hope to address these issues in our future work.

Bibliography

[1] Athreya, K.B. and Ney, P. (1972). *Branching processes,* Springer, New York.

[2] Athreya, K.B. and Kaplan, N. (1976). *Convergence of the age distributin in the one-dimentionnal supercritical age dependent branching process, Annals of Probability,* 4: 38-50.

[3] Athreya, K.B. and Keiding, N. (1977). *Estimation theory for continuous-time branching processes, Indian Journal of Statistics,* 39: 101-123.

[4] Asmussen, S., Jensen, J.L. and Rojas, N.L. (2014). *On the laplace yransform of the lognormal distribution, Methodology and Computing in Applied Probability,* 1: 1-18.

[5] Baker, F., Denniston, M., Smith, T. and West, M. (2005). *Adult cancer survivors: how are they faring, Cancer,* 104: 2565-2576.

[6] Baranyi, J., George S.M. and Kutalik, Z. (2005). *Parameter estimation for the distribution of single cell lag times, Journal of Theoretical Biology,* 259: 24-30.

[7] Bellman, R. and Harris, T.E. (1952). *On age-dependent binary branching processes, Annals of Mathematics*, 55: 280-296.

[8] Beran, R. (1977). *Minimum Hellinger distance estimates for parametric models, Annals of Statistics*, 5: 445-463.

[9] Chan, S.S. (2008). *Robust and efficient inference for linear mixed models using skew-normal distribution, Thesis under the guidance of A.N. Vidyashankar, Cornell University.*

[10] Cheng, A.L. (2004). *Robust inference for randomized play the winner design, Thesis under the guidance of A.N. Vidyashankar, University of Georgia.*

[11] Cowan, R. and Staudte, R. (1986). *The bifurcating autoregerssion model in cell lineage studies, Biometrika*, 15: 769-783.

[12] Cowan, R., Culpin, D. and Morris, V. (1981). *A method for the measurement of variability in cell lifetimes, Mathematical Bioscience*, 54: 249-263.

[13] Cowan, R. (1985). *Branching process results in terms of the moment of the generation-time distribution, Biometrics*, 41: 681-689.

[14] Coldman, A.J. and Goldie, J.H. (1986). *A stochastic model for the origin and treatment of tumors containing drug-resistant cells, Bull. Math. Biol,* 48: 279-292.

[15] Demicheli, R. and et al. (2008). *Recurrence dynamics does not depend on the recurrence site, Breast Cancer Research,* 10: 83-93.

[16] Demicheli, R. and et al. (2012). *Recurrence dynamics for mon-small cell lung cancer: effect of surgery on the development of metastases, Journal of Thoracic Oncology,* 7: 723-730.

[17] Dion, J.P. (1974). *Estimation of the mean and the initial probabilities of a branching process, Journal of Applied Probability,* 11: 687-694.

[18] Devroye, L. (1987). *A course in density estimation,* Birkhauser, Boston.

[19] Durrett, R. (2010). *Probability: theory and examples,* Cambridge university press.

[20] Foo, J. and Michor, F. (2009). *Evolution of resistance to targeted anti-cancer therapies during continuous and pulsed Administration Strategies, PLoS Comput Biol,* 5: 11-23.

[21] Foo, J. and Michor, F. (2010). *Evolution of resistance to targeted anti-cancer therapies during general dosing schedules, Journal of Theoretical Biology,* 263: 179-188.

[22] Gardiner, C.W. (1985). *Stochastic methods,* Springer, Verlag.

[23] Goldie, J.H. and Coldman, A.J. (1979). *A mathematical model for relating the drug sensitivity of tumours to their spontaneous mutation rate, Cancer Treat Rep,* 63: 1727-1733.

[24] Goldie, J.H. and Coldman, A.J. (1983). *Quantitative model for multiple levels of drug resistance in clinical tumors, Cancer Treat Rep,* 67: 923-931.

[25] Guttorp, P. (1991). *Statistical inference for branching processes,* Wiley, Interscience.

[26] Harris, T.E. (1963). *The theory of branching processes,* Springer, Berlin.

[27] Harnevo, L. and Agur, Z. (1992). *Drug resistance as a dynamic process in a model for multistep gene amplification under various levels of selection stringency, Cancer Chemotherapy Pharmacology,* 30: 469-476.

[28] Halon, B. and Vidyashankar, A.N. (2011). *Inference for quantitation parameters in polymerase chain reactions via branching process with random effects, Journal of the American Statistical Association* 106: 525-533.

[29] Harnevo, L. and Agur, Z. (1992). *The dynamics of gene amplilification described as a multi type compartmental model and as a branching process, Mathematical Biosciences,* 103: 115-138.

[30] Hazewinkel, M. (2001). *Law of large numbers, Encyclopedia of Mathematics*, Springer.

[31] Huggins R. (1996). *Rubust inference for variance components models for single tree of cell lineage data, Annals of Statistics*, 12: 1145-1160.

[32] Hyrien, O. and Martin, S.Z. (2008). *A mixture model with dependent observations for the analysis of CSFE labeling experiments, Journal of the American Statistical Association,* 103: 222-239.

[33] Iwasa, Y., Michor, F., and Nowak, M.A. (2003). *Evolutionary dynamics of escape from biomedical intervention, Proc. Roy. Soc. Lond,* 270: 2572-2578.

[34] Jagers, P. (1975). *Branching processes with biological applications, Annals of Probability*, 6: 701-704.

[35] Karlin, S. (1966). *A first course in stochastic processes,* Academic press, New York.

[36] Komarova, N.L. and Wodarz, D. (2005). *Drug resistance in cancer: principles of emergence and prevention, PNAS,* 102: 9714-9719.

[37] Komarova, N.L. (2006). *Stochastic modeling of drug resistance in cancer, Journal of Theoretical Biology,* 239: 351-366.

[38] Kolmogorov, A.N. and Dmitriev, N.A. (1947). *Branching stochastic processes, Dokl. Akad. Nauk USSR,* 56: 5-8.

[39] Kuczek, T. (1982). *On the comvergence of the empiric age distribution for one dimensional supercritical age dependet branching processes, Annals of Probability,* 10: 252-258.

[40] Michaelson, J., Halpern, E. and Kopans, D. (1999). *Breast cancer: computer simulation method for estimating optimal intervals for screening, Radiology,* 212: 551-560.

[41] Michor, F., Nowak, M.A., and, Iwasa Y. (2006). *Evolution of resistance to cancer therapy, Curr Pharm Design,* 12: 261-271.

[42] Palmer, K.J., Ridout, M.S. and Morgan, J.T. (2008). *Modelling cell generation times by using the tempered stable distribution, Journal of the Royal Statistical Society*, 57: 379-397.

[43] Powell, E. (1955). *Some features of the generation times of individual bacteria, Biometrika*, 16: 257-277.

[44] Rahn, O. (1932). *A chemical explanation of the variability of the growth rate, Journal of General Physiology*, 15: 257-277.

[45] Ridout, M.S., Cole, D.J. and et al. (2006). *New approximations to the Malthusian parameter, Biometrics*, 62: 1216-1223.

[46] Sriram, T.N. and Vidyashankar, A.N. (2000). *Minimum Hellinger distance estimation for supercritical galton-watson processes, Statistics & Probability Letters*, 50: 331-342.

Curriculum Vitae

Xin Cao grew up in Jiangsu, China. He then attended the Beihang University, where he received his Bachelor of Science in Mathematics in 2009. He went on to receive his Master of Science in Statistics from Rutgers University in 2011. He then received his Doctorate in Statistics from George Mason University in 2015.