ABSTRACT

SELECTED PROBLEMS OF INFERENCE ON BRANCHING PROCESSES AND POISSON SHOCK MODEL

by
Satrajit Roychoudhury

This dissertation explores the development of statistical methodology for some problems of branching processes and poisson shock model.

Branching process methods have become extremely popular in recent days. This dissertation mainly explores two fundamental inference problems of Galton-Watson processes. The first problem is concerned with statistical inference regarding the nature of the process. Two methodologies have been developed to develop a statistical test for the null hypothesis that the process is supercritical versus an alternative hypothesis that the process is non-supercritical. Another problem we investigate involves the estimation of the '*age*' of a Galton-Watson Process. Three different methods are discussed to estimate the '*age*' with suitable numerical illustrations. Computational aspects of these methods have also been explored.

The literature regarding non-parametric aging properties is quite extensive. Bhattacharjee (2005) recently introduced a new notion of non-parametric aging property known as Strong decreasing Failure rate (SDFR). This dissertation explores necessary and sufficient conditions for which this nonparametric aging property is preserved under Essary-Marshall-Proschan shock model. It has been proved that the discrete SDFR property is transmitted to continuous version of SDFR under a shock model operation. A counter example has been constructed to show that the converse is false.

SELECTED PROBLEMS OF INFERENCE ON BRANCHING PROCESSES AND POISSON SHOCK MODEL

by
Satrajit Roychoudhury

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

August 2006

APPROVAL PAGE

SELECTED PROBLEMS OF INFERENCE ON BRANCHING PROCESSES AND POISSON SHOCK MODEL

Satrajit Roychoudhury

Manish C. Bhattacharjee, Ph.D., Dissertation Advisor — Date
Professor of Mathematical Sciences, NJIT

Sunil K. Dhar, Ph.D., Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Denis Blackmore, Ph.D., Committee Member — Date
Professor of Mathematical Sciences, NJIT

David Mendonca, Ph.D., Committee Member — Date
Assistant Professor of Information System , NJIT

Thomas Spencer, Ph.D., Committee Member — Date
Professor of Applied Management and Decision Sciences, Walden University

BIOGRAPHICAL SKETCH

Author: Satrajit Roychoudhury

Degree: Doctor of Philosophy

Date: August 2006

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2006
- Master of Science in Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2005
- Master of Science in Statistics (MStat),
  Indian Statistical Institute, India, 2002
- Bachelor of Science in Statistics,
  University of Calcutta, Calcutta, India, 2000

Major: Mathematical Sciences

Presentations and Publications:

S. Roychoudhury, M.C. Bhattacharjee, "A Family of Probability Generating Functions Induced by Shock Models," *Center for Applied Mathematics and Statistics (CAMS) Publications*, Report Number 40, May 2006.

S. Roychoudhury, M.C. Bhattacharjee, "A Family of Probability Generating Functions Induced by Shock Models," *Contributed Speaker*, Joint Statistical Meeting (JSM), Minneapolis, Minnesota, August 2005.

S. Roychoudhury, M.C. Bhattacharjee, "A Family of Probability Generating Functions Induced by Shock Models," *Poster Presentation*, Frontiers in Applied and Computational Mathematics, Newark, NJ, May 2005.

S. Roychoudhury, M.C. Bhattacharjee, "A Family of Probability Generating Functions Induced by Shock Models," *Poster Presentation*, Einstein in the City, City College of New York, New York, NY, March 2005.

To My Loving Parents and Sister

ACKNOWLEDGMENT

It is my pleasure and honor to express my sincere gratitude to my advisor, Dr. Manish Bhattacharjee for his valuable guidance. Without his generous encouragement and help, I would not have been able to complete this work. I also want to thank all members of my dissertation committee. Their advice and patience are highly appreciated. I am also specially indebted to Prof. P. K. Sen of the University of North Carolina Chapel Hill for his constructive suggestions, which have been invaluable for completing this research.

My earnest thanks go to Dr. Sunil K. Dhar and Dr. Thomas Spencer, who greatly enriched my knowledge with their exceptional insights into different areas of modern statistics.

I am also grateful to Dr. Rose Dios, Dr. Aridaman Jain and Dr. Kenneth Johnson for their constant support, encouragement, and advice.

The nice company of my class-fellows and friends always proved to be a source of inspiration and encouragement. I would like to say thanks to all my class-fellows. I especially want to thank Soumi Lahiri for her appreciation, moral support and encouragement during my stay here.

I would also like to thank the Department of Mathematical Sciences at NJIT for providing me financial support for this research.

This study would have been impossible without the prayers, love and affection of my family. I wish to express my appreciation and deep sense of gratitude to my parents, loving sister and grandmother, who's hands always arise in prayers for my success.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Motivation and Background of Branching Process

The theory of branching processes is an area of Applied Probability that describe situations in which an entity exists for a time and then may be replaced by one, two, or more entities of a similar or different type. It is a well-developed and active area of research with theoretical interests and practical applications.

The theory of branching processes has made important contributions to biology and medicine since Francis Galton originally considered the extinction of family names among the British peerage in the nineteenth century. More recently, branching processes have been successfully used to illuminate problems in the areas of molecular biology, cell biology, developmental biology, immunology, evolution, ecology, medicine, and others. For the experimentalist and clinician, branching processes have helped in the understanding of observations that seem counterintuitive, have been used to develop develop new experiments and clinical protocols, and have provided predictions which have been tested in real-life situations. For an applied probabilist and statistician, the challenge of understanding new biological and clinical observations has motivated the development of new methods in the field of branching processes.

1.2 An Overview

This dissertation explores two different inference problems regarding Galton-Watson process.The first problem deals with estimation of *'age'* of a Galton-Watson process and the second one is constructing a statistical testing procedure for the mean value of the first generation of a Galton-Watson process.

In this dissertation different nonparametric estimators of the age of Galton-Watson Process have been developed, which substantially improves upon earlier work

in this case, that were either concerned with some specific parametric families of offspring distribution (such as geometric and Poisson families) or in the nonparametric case where restricted to single observation. Another contribution is the development of large sample nonparametric tests for the explosion vs. non-explosion hypothesis. This is also a significant improvement over the past literature, that were mostly for parametric cases. In this dissertation a significant work has been done to see whether a new nonparametric aging notion, introduced by Bhattacharjee (2005), is preserved under shock model operation.

There are many situations where it is required to estimate the age of a process. This situation arises when somebody is interested in estimating the length of time a specific species has existed in its present form, without knowing much past information. This kind of problems arises often in Anthropology, social studies etc. This problem also arises in genetics when somebody wants to know the age of mutations of an allele. There are two ways a statistician can deal with such a problem. The first method considers the *'age'* as a parameter and then attempts to estimate it by using classical or Bayesian estimation procedures. The technique of second method is to treat the *'age'* as a random variable. In this technique the probability structure of the process is used to find the distribution of *'age'*. Examples of such a method is *'stopping time'* of a Markov process. In this dissertation, the chosen emphasis is on developing methodology in the context of the first method. In Chapter 3, three different methods of estimation of *'age'* of a Galton-Watson process is discussed along with examples. An EM (Expectation Maximization) algorithm is developed to estimate the parameters of offspring distribution. All the estimators have been developed in the nonparametric setup. This means no analytical parametric form of the offspring distribution is assumed. This chapter concludes an estimation procedure for the probability of extinction of a Galton-Watson process.

In Chapter 4, a fundamental question regarding the nature of Galton-Watson process is discussed. Given data of several consecutive generations of a Galton-Watson process, the first question that comes into mind is about the extinction of the process. The extinction of the Galton-Watson process has a relation with the mean of first generation. All these relationships are discussed in Chapter 2. In Chapter 4, two large sample test statistics have been developed along with their asymptotic properties to deal with such testing problems.

The body of concepts, tools and methods collectively known as the statistical theory of reliability owe their genesis to problems dealing with "lifetimes" of hardware component and systems. Originally, interest in such problems were driven by a need to successfully model and predict the probability of a complex system of interconnected components to operate successfully, allowing for possibility of component failures. Over time, it was realized however that many of these ideas whose development were first motivated by problems in hardware reliability had parallels in other fields. For example, the notions of failure intensity and hazard functions are also known to and used by demographers and actuaries as the "force of mortality". Similarly, various notions of "aging" to model degradation of performance as developed by reliability theorists were found to have interconnections with appropriate notions of various forms of stochastic partial orderings.A specific problem posed in Chapter 5 is one of investigating a certain class of probability generating functions, and was in fact motivated by the connection between a strong anti-aging (nonparametric) property and failure distributions which have a shock model representation driven by a Poisson process.

CHAPTER 2

PRELIMINARIES OF GALTON-WATSON PROCESS

This chapter discusses some basic preliminaries of Galton-Watson processes that we will need in Chapter 3 and 4. The definitions and results stated in this chapter can be found in by Athreya and Ney (1972). These are used to prove our results in the next two chapters of this dissertation.

2.1 Galton-Watson Process

A Galton-Watson process is a Markov chain $\{Z_n; n = 0, 1, 2, \cdots\}$ on the nonnegative integers. Its transition function is defined in terms of a given probability distribution $\{p_k; k = 0, 1, 2, \cdots\}, p_k \geq 0, \Sigma p_k = 1$, by

$$P(i,j) \;=\; P\{Z_{n+1} = j | Z_n = 1\} = \begin{cases} p_j^{*i} & \text{if } i \geq 1,\ j \geq 0, \\ \delta_{0j} & \text{if } i = 0,\ j \geq 0, \end{cases} \tag{2.1}$$

δ_{ij} being the Kronecker delta and $\{p_k^{*i}; k = 0, 1, 2, \cdots\}$ being the i-fold convolution of $\{p_k; k = 0, 1, 2, \cdots\}$.

The probability function $\{p_k\}$ is the total datum of the problem. The process can be thought of as representing an evolving population of particles. It starts at time 0 with Z_0 particles, each of which splits independently of the others into a random number of offsprings according to the probability law $\{p_k\}$. The total number Z_1 of particles thus produced is the sum of Z_0 random variables, each with probability function $\{p_k\}$. It constitutes the first generation. These go on to produce a second generation of Z_2 particles, and so on. The number of "offspring" produced by a single "parent" particle at any given time is independent of the history of the process, and of other particles existing at present. The number of particles in th n-th generation is a random variable Z_n. The Equation (2.1) tells us that if $Z_n = 0$, then $Z_{n+k} = 0$

for all $k \geq 0$. Thus 0 is an absorbing state, and reaching 0 is the same as the process becoming extinct.

The branching process with i initial particles is denoted by $\{Z_n{}^i, n = 0, 1, 2, \cdots\}$. Since most of the time we will be assuming $Z_0 = 1$, it will be convenient to write, $Z_n^1 = Z_n$.

2.2 Generating Functions

An important tool in the analysis of the process is the generating function

$$f(s) = \sum_{k=0}^{\infty} p_k s^k, \quad |s| \leq 1, \tag{2.2}$$

and its iterates

$f_0(s) := s$, $f_1(s) = f(s)$, $f_{n+1}(s) = f[f_n(s)]$, for $n \geq 1$,

where s is assumed to be real. It can be observed that

$$\sum_j P(1,j)s^j = f(s); \quad \sum_j P(i,j)s^j = [f(s)]^i, \quad i \geq 1. \tag{2.3}$$

Also, letting $p_n(i,j)$ be the n-step transition probabilities, and using the Chapman-Kolmogorov equation, we get

$$\begin{aligned} \sum_j P_{n+1}(1,j)s^j &= \sum_j \sum_k P_n(1,k)P(k,j)s^j \\ &= \sum_k P_n(1,k) \sum_j P(k,j)s^j \\ &= \sum_k P_n(1,k)[f(s)]^k. \end{aligned}$$

Thus if $\sum_j P_n(1,j)s^j = f_{(n)}(s)$, then it can be shown that

$$f_{(n+1)}(s) = f_{(n)}[f(s)].$$

Hence, it follows by induction that

$$f_{(n)}(s) = f_n(s), \tag{2.4}$$

a crucial formula.

From Equation (2.3) and (2.4) it can be deduced

$$\sum_{j=0}^{\infty} P_n(i,j)s^j = [f_n(s)]^i. \tag{2.5}$$

2.3 Additive Property

The process $\{Z_n^i; n = 0, 1, 2, \cdots\}$ is the sum of i independent copies of the branching process $\{Z_n; n = 0, 1, 2, \cdots\}$. In other words, if P_i denotes the measure on $\mathcal{F}$ (the minimal σ-field induced by the process)corresponding to the initial measure $P\{Z_0 = i\} = 1$, then P_i is the i-fold convolution of P_1. Thus the joint distribution of $(Z_{n_1}^{(i)}, \cdots, Z_{n_k}^{(i)})$, for integers $1 \leq n_1 \leq \cdots \leq n_k$, is the i-fold convolution of the distribution $(Z_{n_1}, \cdots, Z_{n_k})$.

In order to avoid trivialities, throughout it is assumed that

$$p_0 + p_1 < 1,$$

and

$$p_j \neq 1 \quad \text{for any } j.$$

2.4 Moments

The moments of the process, when they exist can be expressed in terms of the derivatives of $f(s)$ at $s = 1$. For the mean we have

$$E(Z_1) = \sum P(1,j)\, j = f'(1) \equiv m \quad \text{(say)},$$

and from the chain rule

$$E(Z_n) = \sum_j P(1,j)\, j = f_n'(1) = f_{n-1}'(1)f'(1) = \cdots = [f'(1)]^n = m^n.$$

Similarly, using the fact that

$$f_{n+1}''(1) = f''(1)\,[f_n'(1)]^2 + f'(1)f_n''(1),$$

it can be shown that

$$f_n''(1) = f''(1)[m^{2n-2} + m^{2n-3} + \cdots + m^{n-1}],$$

and hence, letting $\sigma^2 =$ variance Z_1, it can be concluded that

$$\text{var } Z_n = \begin{cases} \dfrac{\sigma^2 m^{n-1}(m^n - 1}{m-1}, & \text{if } m \neq 1, \\ n\sigma^2, & \text{if } m = 1. \end{cases}$$

Higher moments can be derived similarly.

2.5 Elementary Properties of Generating Functions

All the properties of the transition functions $P_n(i,j)$ are contained in the generating functions $f_n(s)$. In particular, the asymptotic behavior of $\{f_n(s)\}$ can be translated into limit theorems about the $\{Z_n\}$ process, which are discussed in the next section. The simple properties are as follows.

Let t be real. From the definition of f as a power series with non-negative coefficients $\{p_k\}$ adding to 1, and with $p_0 + p_1 < 1$,

(i) f is strictly convex and increasing in $[0,1]$;

(ii) $f(0) = p_0; f(1) = 1$;

(iii) if $m \leq 1$ then $f(t) \geq t$ for $t \in [0,1)$;

(iv) if $m > 1$ then $f(t) = t$ has a unique root in $[0, 1)$.

Let q be the smallest root of $f(t) = t$ for $t \in [0,1]$. Then (i)-(iv) imply that there is such a root and furthermore:

Lemma 2.5.1. *If $m \leq 1$ then $q = 1$; if $m > 1$ then $q < 1$.*

Lemma 2.5.2. *If $t \in [0,q)$ then $f_n(t) \uparrow q$ as $n \to \infty$.*

If $t \in (q,1)$ then $f_n(t) \downarrow q$ as $n \to \infty$.

If $t = q$ or 1 then $f_n(t) = t$ for all n.

Lemma 2.5.3. *The functions $f_n(s)$ are differentiable and converge on $[0,1)$. Moreover for all $s \in [q,1)$, $f_n'(s) \leq (f'(s))^n$ and for all $s \in [0,q)$, $f_n'(s) \geq (f'(s))^n$. This suggests that $f_n'(s)$ has a geometric rate of decay.*

2.6 Extinction Probability

As a special case of Lemma 2.1.2 it can be noted that $f_n(0) \uparrow q$. But $\lim_{n\to\infty} f_n(0) = \lim_n P\{Z_n = 0\} = \lim_n P\{Z_i = 0 \text{ for some } 1 \leq i \leq n\} = P\{Z_i = 0 \text{ for some } i \geq 1\}$, which is by definition the probability that the process eventually becomes extinct. Applying Lemma 2.1.1, the classical extinction probability theorems are obtained.

Theorem 2.6.1. *The extinction probability of the $\{Z_n\}$ process is the smallest non-negative root (q) of the equation $t = f(t)$. It is 1 if $m \leq 1$ and < 1 if $m > 1$.*

Theorem 2.6.2. $\lim_{n\to\infty} P\{Z_n = k\} = 0$ *for $k \geq 1$. Furthermore,*

$$P\{\lim_n Z_n = 0\} \;=\; 1 - P\{\lim_n Z_n = \infty\} = q.$$

2.7 Limit Theorems

Different limit theorems describing the divergence nature of Z_n are stated in this section.

The stochastic process $\{Z_n^i; n = 0, 1, 2, \cdots\}$ is the sum of i independent copies of the process $\{Z_0 \equiv 1, Z_1, Z_2, \cdots\}$. Using the Markov property

$$\begin{aligned} &E(Z_{n+k}|Z_n = i_n, Z_{n-1} = i_{n-1}, \cdots, Z_1 = i_1, Z_1 = i_1) \\ &= E(Z_{n+k}|Z_n = i_n) = i_n E(Z_k|Z_0 = 1) = i_n m^k. \end{aligned}$$

Hence if we set

$$W_n \;\equiv\; Z_n\, m^{-n},$$

then

$$E(W_{n+k}|W_0, W_1, \cdots, W_n) \quad = \quad W_n \quad \text{a.s.}$$

so that $\{W_n, n \geq 0\}$ is a Martingale (see Appendix-B). The following results are well known (Atreya and Ney (1972)).

Theorem 2.7.1. *If* $0 < m \equiv f'(1-) < \infty$, $W_n = Z_n\ m^{-n}$, *and* $\boldsymbol{F_n}$ *is the* σ-*algebra generated by* $Z_0, Z_1, \cdots, Z_n$, *then the sequence* $\{W_n, \boldsymbol{F_n}; n = 0, 1, 2, \cdots\}$ *is a martingale. Furthermore, since* $W_n \geq 0$, *there exists a random variable* W *such that*

$$\lim_{n\to\infty} W_n \quad = \quad W \quad a.s.$$

Theorem 2.7.2. *If* $m > 1$, $\sigma^2 < \infty$, *and* $Z_0 \equiv 1$, *then*

(i) $\lim_{n\to\infty} E(W_n - W)^2 = 0; \quad i.e., W_n \xrightarrow{L^2} W$

(i) $E(W) = 1, \quad var\ (W) = \sigma^2/(m^2 - m)$

(iii) $P(W = 0) = q \equiv P(Z_n = 0\ \ for\ some\ n)$.

CHAPTER 3

ESTIMATION OF AGE OF A GALTON WATSON PROCESS

There are several situations in which we might want to estimate the age of a markov process. For example, we might know, at least approximately, the number of plants or animals of a certain species in existence. Now suppose we are interested in estimating the length of time the species has been in existence in its present form, without having much historic information such as fossils for carbon dating. Another source of applications would be genetics. In a genetic context, the usual problem is to find the age of an allele, given its current frequency. In other words, we are interested in estimating how long ago a mutation took place. For a discrete time branching process (Galton Watson Process) $\{Z_j : j = 0, 1, 2, \cdots\}$, its *age* is the generation label n that corresponds to our earliest observed value Z_n of this process. There are essentially two different approaches to the problem of estimating the 'age' parameter n for stochastic process. one could adopt a statistical approach by forming a likelihood based on our observations, and then estimate the age by, for instance, maximum likelihood *paradigm.* For examples of this method, see Stigler (1970), Thompson (1976). Alternatively, one can define the age in terms of some random variable, and find its distribution. For example of this method see Levikson, B. (1977). In this chapter the first method is used to deal with the problem.

3.1 Background Work

Let $Z_0 = 1, Z_1, Z_2, \ldots$ denote the sizes of successive generations in a Galton-Watson process, starting with a single ancestor, and with a probability generating function (pgf) of the *progeny* distribution with mean m. Without loss of generality, $Z_0 = 1$ is assumed as is customary; since for the case of multiple ancestors ($Z_0 > 1$), the process $\{Z_j, j \geq 0\}$ is equivalent to Z_0 statistically identical copie of a Galton Watson

Process with the same offspring distribution and starting with a single ancestor. The problem of estimating the age of Galton-Watson process was first addressed by Stigler (1970). He assumed the generating function is known and further that it has fractional linear form (i.e. a geometric distribution with modified zero term) and the process is supercritical (i.e., $m > 1$) . So in this case conditioning on non-extinction of the process the likelihood of generation 'n' based on one observation Z_n, where n is unknown, is

$$L(n) = P(Z_n = k) = \frac{g_n^{(k)}(0)}{k!(1 - g_n(0))}$$

where g_n (defined by $g_1 := g, g_n := g(g_{n-1}), n \geq 2$) is the generating function of Z_n. For the fractional linear case g_n is of the form

$$g_n(s) = 1 - \frac{b_n}{1 - c_n} - \frac{b_n s}{1 - c_n s},$$

which gives $P(Z_n = k) = b_n c_n^{k-1}$ and $g_0 = (1 - b_n)/(1 - c_n)$. Thus the likelihood becomes

$$L(n) = (1 - c_n)c_n^{k-1}$$

where $c_n = (m^n - 1)/(m^n - q)$, q is the probability of extinction of the process. Using this likelihood, the maximum likelihood (MLE) of n can be obtained as

$$\hat{n} = \frac{\ln(k(1 - q) - q)}{\ln m}$$

Stigler (1970) also proved the consistency and asymptotic efficiency of this estimator. Later, Crump and Howe (1972) studied the case where n is estimated from data containing several generations of a Galton-Watson process viz, $Z_l, Z_{l+1}, \ldots Z_n$. They have explored the Markov structure of $\{Z_n\}$ to estimate the generation. Since Z_k's form a Markov chain with stationary transition probabilities, the likelihood with

respect to n is proportional to the marginal distribution of Z_l. Hence the mle of n is $\hat{l} + d$ where $\hat{l}$ is obtained by the formula given by Stigler(1970) and d is the number of observation taken (i.e., $d = n - l$). They have used a non-parametric MLE of m, which is

$$\hat{m} = \frac{\sum_{k=l+1}^{n} Z_k}{\sum_{k=l}^{n-1} Z_k},$$

obtained by Harris (1963). Crump and Howe (*ibid*) also proved the assymptotic properties of the estimator. In both of the previous cases the underlying assumption is that the generating function is fractional linear generating function. It is possible to find the MLE for some cases other than the fractional linear generating function. Adès et al. (1982) developed an algorithm to obtain the MLE for several other parametric families of offspring distributions such as Poisson and negative binomial.

3.2 General Case

Let the observed sample from a Galton-Watson process, be over $(r+1)$ consecutive generations, denoted by $(Z_n, Z_{n+1}, \ldots, Z_{n+r})$, where $r \in \{0, 1, 2, \ldots\}$, and the age parameter n is unknown. Throughout this chapter no restrictive assumptions are made about the form of the offspring distribution's pgf $g(s)$. The only assumption will be $m > 1$, since otherwise the population would become extinct with probability one, in which case the estimation problem would most likely not arise at all. It should also be pointed out that the assumption that $Z_0 = 1$ is not testable. Suppose the offspring distribution has a finite support $\{0, 1, \ldots, M\}$. This assumption of a maximum number of offsprings is not unrealistic in applied contexts. Since we have observations of $(r+1)$ generations; we may choose for example,$M = \max(Z_n, Z_{n+1}, \ldots, Z_{n+r})$. Let $\boldsymbol{p} = (p_0, p_1, \ldots, p_M)$ be the corresponding offspring distribution (i.e. p_j = probability of j offsprings). Now to estimate n, first it is required to find the MLE of m. Finding the MLE of m is equivalent to finding the MLE of $\boldsymbol{p}$, since $m = \sum_j jp_j$. The

nonparametric likelihood of $\boldsymbol{p}$ is given by,

$$\begin{aligned} \mathrm{L}(\boldsymbol{p}) &= \prod_{j=0}^{r-1} P(Z_{n+j+1} = k_{n+j+1} | Z_{n+j} = k_{n+j}) \\ &= \prod_{j=0}^{r-1} \left\{ \frac{k_{n+j+1}!}{\prod_{l=0}^{M} k_{n+j}^{(l)}!} \prod_{l=0}^{M} p_l^{k_{n+j}^{(l)}} \right\}, \end{aligned}$$

where $k_{n+j}^{(l)}$ denotes the number of individuals in $(n+j)$-th generation who gave rise to exactly l offsprings in the next generation. This likelihood can be further simplified as

$$\begin{aligned} L(\boldsymbol{p}) &= \left\{ \prod_{j=0}^{r-1} \frac{k_{n+j+1}!}{\prod_{l=0}^{M} k_{n+j}^{(l)}!} \right\} \prod_{l=1}^{M} p_l^{\sum_{j=0}^{r} k_{n+j}^{(l)}} \\ &= \left\{ \prod_{j=0}^{r-1} \frac{k_{n+j+1}!}{\prod_{l=0}^{M} k_{n+j}^{(l)}!} \right\} \prod_{l=1}^{M} p_l^{k^{(l)}}, \end{aligned} \tag{3.1}$$

where $k^l = \sum_{j=0}^{r-1} k_{n+j}^{(l)}$ denote the total number of individuals in the observed r generations who gave birth to exactly l offspring in the next generation. Note, $k^{(l)}$ is not observable. Now taking logarithm in both sides of Equation (3.1) the equation becomes

$$\ln L(\mathbf{p}) := \mathrm{l}(\boldsymbol{p}) = C - \sum_{j=0}^{r} \sum_{l=0}^{M} \ln(k_{n+j}^{(l)}!) + \sum_{l=0}^{l=M} k^{(l)} \ln p_l,$$

Using Sterling's Approximation formula for factorials,

$$l(\boldsymbol{p}) \approx C' - \sum_{j=0}^{r} \sum_{l=0}^{M} \{-k_{n+j}^{(l)} + (k_{n+j}^{(l)} + 0.5) \ln k_{n+j}^{(l)}\} + \sum_{l=0}^{M} k^{(l)} \ln p_l, \tag{3.2}$$

where C and C' involve terms independent of $\boldsymbol{p}$ and $k_{n+j}^{(l)}$.

The objective is to maximize the $l(\boldsymbol{p})$ subject to the following set of constraints

$$\begin{aligned}
\textbf{a)}\ \sum_{l=0}^{M} p_l &= 1 : (\text{“honest” distribution condition}) \\
\textbf{b)}\ \sum_{l=0}^{M} l p_l &> 1 : (\text{“supercriticality” assumption}) \\
\textbf{c)}\ 0 \leq p_l &\leq 1 \ \ \forall \ \ l = 0, 1, \ldots, M.
\end{aligned}$$

The problem posed above involves both equality and inequality constraints. So maximization is not possible directly. There is another issue involved here. It is clear that the likelihood is related to the data by the factors $k_{n+j}^{(l)}$. So the optimal value of $\boldsymbol{p}$ will be a function of $k_{n+j}^{(l)}$ if equation (3.2) is maximized directly. But the values of $k_{n+j}^{(l)}$ are unknown. One method to deal with such a situation is described by Dion et. al.(1982), where they have considered all the processes for which the given data can be obtained. But that method is very much tideous, specially when M is large. The second problem for that method is, the generation lables must be known. An alternative method based on EM algorithm is described below which has been implemented to solve those problems. A brief overview of EM algorithm is given in Appendix A.

3.2.1 Solution via EM Algorithm

The likelihood function can be written as

$$l(\boldsymbol{p}) = l(k_{n+j}^{(l)} : j = 0, \ldots, r, l = 0, \ldots, M | \boldsymbol{p})$$

Here $\{k_{n+j}^{(l)} : j = 0, \ldots, r, l = 0, \ldots, M\}$ is the unobserved part of the data and $\{Z_k : k = n, n+1, \ldots, n+r\}$ is the observed part and $\boldsymbol{p}$ is the parameter to estimate. Using EM algorithm, the objective surrogate function becomes

$$Q(\boldsymbol{p}|\boldsymbol{p}^{(d)}) = E\{l(k_{n+j}^{(l)} : j = 0, \ldots, r, l = 0, \ldots, M | \boldsymbol{p}) | Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}\}$$

$$Q(\boldsymbol{p}|\boldsymbol{p}^{(d)}) = E\{C' - \sum_{j=0}^{r}\sum_{l=0}^{M}\{-k_{n+j}^{(l)} + (k_{n+j}^{(l)} + 0.5)\ln k_{n+j}^{(l)}\} + \sum_{l=1}^{M} k^{(l)} \ln p_l | Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}\}$$

$$Q(\boldsymbol{p}|\boldsymbol{p}^{(d)}) = C' - E\{\sum_{j=0}^{r}\sum_{l=0}^{M}\{-k_{n+j}^{(l)} + (k_{n+j}^{(l)} + 0.5)\ln k_{n+j}^{(l)}\}|Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}\} + \sum_{l=1}^{l=M} E\{k^{(l)}|Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}\}\ln p_l^{(d)}, \tag{3.3}$$

where C' is the term independent of $\boldsymbol{p}$ and $\boldsymbol{p}^{(d)}$ denotes the value of $\mathbf{p}$ at d-th iteration.

The problem now reduces to find $\boldsymbol{p}$ which maximizes Q, instead of l, under the constraints a), b) and c) following (3.2). The surrogate function Q can be further simplified. The definition of k^l in (3.1) is

$$k^{(l)} = k_n^{(l)} + k_{n+1}^{(l)} + \cdots\cdots + k_{n+r}^{(l)}.$$

Here an interesting observation is that $k_{n+j}^{(l)}$ follows binomial distribution with parameters Z_{n+j} and p_l. This of course yields

$$E(k_{n+j}^{(l)}|Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}) = Z_{n+j}p_l^{(d)},$$

which in turn implies,

$$E(k^{(l)}|Z_n, Z_{n+1}, \ldots, Z_{n+r}, \boldsymbol{p}^{(d)}) = (Z_n + Z_{n+1} + \cdots\cdots + Z_{n+r})\, p_l^{(d)} \tag{3.4}$$

Using equation (3.4) in equation (3.3) the final form of Q becomes

$$Q(\boldsymbol{p}|\boldsymbol{p}^{(d)}) = C' - \sum_{j=0}^{r}\sum_{l=0}^{M}\sum_{t=0}^{Z_{n+j}}\{-t + (t+0.5)\ln t\; B(Z_{n+j}; t; p_l) + N\sum_{l=1}^{M} p_l^{(d)} \ln p_l^{(d)} \tag{3.5}$$

where $N = \sum_{j=0}^{r} Z_{n+j}$ and $B(Z_{n+j}, t, p_l) = P(k_{n+j}^{(l)} = t) = \begin{pmatrix} Z_{n+j} \\ t \end{pmatrix} p_l^t (1 - p_l)^{Z_{n+j}-t}$.

Now the problem reduces to maximizing (3.5) under the constraints a), b) and c). A closed form exact analytical solution for a "Nonparametric maximum likelihood estimator" (NPMLE) of $\boldsymbol{p}$ is not possible. An approximate solution can be obtained by implementing a suitable numerical optimization scheme using MATLAB, or other suitable softwares.

Example 1:

The following simulated data is generated form a Galton-Watson Process with probability generating function $g(s) = 0.1 + 0.3s + 0.6s^2$. Four hundred (400) samples of $r = 10$ generations are generated from this precess and EM method is used to estimate $\mathbf{p} = (p_0, p_1)'$ of offspring distribution in each case. The following table is showing the estimated values of $\boldsymbol{p}$ starting with 10 different initial values. Convergence is achieved in all cases. where t denotes the number of steps needed for convergence and

Table 3.1 Estimation of p_0, p_1 of Offspring Distribution by EM Algorithm

Initial $\boldsymbol{p}$	$\hat{\boldsymbol{p}}$	t
(0.3,0.3)	(0.1563,0.3213)	12
(0.3,0.2)	(0.1127,0.3010)	19
(0.3,0.1)	(0.1701,0.2947)	23
(0.0,0.1)	(0.1007,0.3211)	45
(0.6,0.2)	(0.1057,0.3119)	42
(0.4,0.3)	(0.1207,0.3562)	33
(0.5,0.4)	(0.2003,0.3829)	26
(0.8,0.2)	(0.1443,0.3004)	42
(0.2,0.0)	(0.1021,0.3425)	32
(0.6,0.2)	(0.1652,0.2851)	35

Example 2:

The following data is simulated from a Galton-Watson Process with probability generating function $g(s) = 0.01 + 0.1s + 0.3s^2 + 0.25s^3 + 0.2s^4 + 0.14s^5$ of the offspring distribution. 400 samples of $r = 10$ generations are generated and EM method is method used to estimate the offspring distribution $\mathbf{p} = (p_0, p_1, p_2, p_3, p_4)'$. The following table shows the estimated values pf $\mathbf{p}$ for 10 different initial values. Convergence is achieved in all cases.

Table 3.2 Estimation of p_0, p_1, p_2, p_3, p_4 of Offspring Distribution by EM Algorithm

Initial **p**	estimated **p**	t
(0.1,0.1,0.1,0.1,0.1)	(0.0051,0.122,0.266,0.276,0.196)	67
(0.3,0.2,0.1,0.1,0.1)	(0.0142,0.113,0.333,0.237,0.192)	78
(0.1,0.1,0.1,0.3,0.1)	(0.0076,0.101,0.309,0.279,0.201)	83
(0.2,0.3,0.1,0.0,0.1)	(0.0032,0.117,0.287,0.251,0.217)	65
(0.0,0.1,0.2,0.2,0.2)	(0.0112,0.088,0.292,0.239,0.208)	79
(0.0,0.0,0.0,0.3,0.3)	(0.0137,0.129,0.312,0.261,0.187)	81
(0.5,0.1,0.1,0.1,0.1)	(0.0091,0.092,0.320,0.259,0.215)	66
(0.04,0.2,0.1,0.0,0.0)	(0.0099,0.167,0.273,0.242,0.206)	73
(0.2,0.0,0.0,0.0,0.0)	(0.0082,0.083,0.296,0.261,0.203)	69
(0.2,0.2,0.0,0.0,0.3)	(0.0126,0.1310,0.331,0.248,0.198)	74

Here t denotes the number of steps needed for convergence.

3.3 Method of Moment Estimator

Suppose $Z_n, Z_{n+1}, \ldots, Z_{n+r}$ are observations of $(r+1)$ successive generations of a Galton-Watson process with p.g.f. $g(s)$ of the offspring distribution. Now $g(s)$ can be estimated by the method discussed in Section 3.2. From the basic results of branching process it is known that,

$$E(Z_{n+j}) = m^{n+j}$$
$$\Rightarrow n + j = \frac{\ln E(Z_{n+j})}{\ln m}$$

The method of moment estimator for n can be proposed as,

$$\hat{n}_1 = (r+1)^{-1} \frac{\sum_{j=0}^{r} \ln Z_{n+j}}{\ln \hat{m}} - \frac{r}{2},$$

where $\hat{m} = \sum_{l=0}^{M} l\hat{p}_l$ is a plug in estimator of the mean progeny size $m = \sum_{l=0}^{M} l\, p_l$; $\hat{p}_l$ being the estimator of p_l, obtained by maximizing (3.6) subject to constraints a), b) and c).

Now let us look into the properties of the estimator $\hat{n}$. The following theorem shows that $\hat{n}_1$ has good large sample propeties.

Theorem 3.3.1. $\hat{n}_1$ is a consistent estimator of n in the explosion set (i.e., on the set $\{Z_n \to \infty\}$).

Proof. $\hat{n}_1$ can be re-expressed via the equation,

$$r\,\hat{n}_1 \;=\; \sum_{j=0}^{r} N_j, \qquad \text{where} \qquad N_j = \frac{\ln Z_{n+j}}{\ln \hat{m}} - j.$$

To show $\hat{n}_1$ is consistent, it obviously enough to prove that

$$N_j \xrightarrow{P} n \qquad \text{as} \quad Z_{n+j} \to \infty \quad \text{a.s.}$$

Now, $Z_{n+j} \xrightarrow{a.s.} \infty \quad \Rightarrow \quad P(Z_{n+j} = 0) := p_{(n+j),0} \longrightarrow 0$ as $Z_n \to \infty$.

From basic results of branching processes, standard computations yield

$$E(Z_{n+j}|Z_{n+j} > 0) = \frac{m^{n+j}}{1 - p_{(n+j),0}}$$

As $\ln x$ is a concave function of x; Jensen inequality gives,

$$\begin{aligned} E(\ln Z_{n+j}|Z_{n+j} > 0) &\leq \ln E(Z_{n+j}|Z_{n+j} > 0) \\ &= (n+j)\ln m - \ln(1 - p_{(n+j),0}) \end{aligned}$$

Also, $\ln(1 - p_{(n+j),0}) = -p_{(n+j),0} - p^2_{(n+j),0} - p^3_{(n+j),0} - \cdots\cdots \to 0$ as $Z_{n+j} \xrightarrow{a.s} \infty$.

$$\begin{aligned} \Rightarrow\; (n+j)^{-1}E(\ln Z_{n+j}|Z_{n+j} > 0) &\leq \ln m \;-\; (n+j)^{-1}\ln(1 - p_{(n+j),0}) \\ &\longrightarrow \ln m \quad \text{as} \quad Z_{n+j} \xrightarrow{a.s} \infty. \end{aligned} \tag{3.6}$$

Now, applying Markov inequality

$$\begin{aligned} P(|\frac{\ln Z_{n+j}}{n+j} - \ln m| > \epsilon | Z_{n+j} > 0) &\leq \frac{E(|\ln m - \frac{\ln Z_{n+j}}{n+j}| | Z_{n+j} > 0)}{\epsilon} \\ &\leq \frac{E(\frac{\ln Z_{n+j}}{n+j} - \ln m | Z_{n+j} > 0)}{\epsilon} \\ &\to 0 \quad \text{as} \quad Z_{n+j} \overset{a.s}{\to} \infty, \quad \text{using (3.6).} \end{aligned}$$

$$\Rightarrow \frac{\ln Z_{n+j}}{\ln m} \overset{P}{\to} n+j \quad \text{as} \quad Z_{n+j} \overset{a.s.}{\to} \infty. \tag{3.7}$$

As all the regularity conditions are satisfied; so using properties of MLE,

$$\frac{\ln \hat{m}}{\ln m} \overset{P}{\to} 1 \quad \text{as} \quad Z_n \overset{a.s.}{\to} \infty. \tag{3.8}$$

Now (3.7) and (3.8) together imply $\hat{n}_j \overset{P}{\to} n$ as $Z_n \overset{a.s.}{\to} \infty$. Finally, applying Slutsky's lemma the proof follows. □

There are some interesting observations in this context. Firstly from the last theorem it is clear that this estimate will work better when the process is exploding. In other words this estimator will work better when probability of extinction is close to zero. Secondly in case of single observation (i.e., $r = 0$) this estimator is a special case of Stigler's (1970) estimator when q=0. So this leads to another problem regarding the probability of extinction q.

3.4 Maximum Likelihood Estimation of the Generation

This method is mainly motivated by a theorem by Adès *et al.* (1982), which can be stated as follows.

Theorem 3.4.1. *Suppose that p_0 =0. If a is such that for a given value of $Z_a \equiv k$, $P(Z_{a+1} = j) \leq P(Z_a = j)$ $\forall j \leq k$, then $P(Z_{a+2} = j) < P(Z_{a+1} = j)$ for all $j \leq k$.*

Theorem 3.4.1 guarantees the unimodality of the coefficients of the generating functions in successive generations and suggests the following algorithm to find the

MLE of generation label n, corresponding to the first observed values of a Galton-Watson process.

3.4.1 Algorithm

Let $Z_n, Z_{n+1}, \ldots, Z_{n+r}$, $n \geq 0$ be the observed $(r+1)$ consecutive generations of a Galton-Watson process with generating function $g(s)$. We assume $p_0 = 0$ so that the process is necessarily supercritical. There is no assumption about the from of the offspring distribution except that the distribution has finite support.

- Estimate $\hat{p}_l$, $l = 0, 1, \ldots M$, by using the EM algorithm discussed earlier.
- Suppose the observed value of Z_n is k.
- Compute the generating functions of Z_n, which is $g_n(s)$ the n-th composition of $g(s)$, for $n = 0, 1, 2, \ldots$.
- Estimate $g_n(s)$ by using the estimated $\hat{\boldsymbol{p}}$.
- Collect the coefficient of s^j $\forall$ $j \leq k$ from $\hat{g}_n(s)$, estimate of $g_n(s)$ for each n.
- Compare the coefficients. If n_1 is the smallest number for which coefficient of s^j in $g_{n_1}(s)$ is less than coefficient of s^j in $g_{n_1+1}(s)$ for all $j \leq k$ or in other words if $n_1 = \min\{n :$ coefficient of s^j in $g_{n+1}(s) <$ coefficient of s^j in $g_n(s)$ $\forall\, j \leq k\}$ then, using theorem 2.4.1 the MLE of n is n_1.

So the MLE of n can be written as

$$\hat{n}_2 = \min\{n : \text{ coefficient of } s^j \text{ in } g_{n+1}(s) < \text{coefficient of } s^j \text{ in } g_n(s) \ \ \forall\, j \leq k\}$$

The method discussed above is intuitively appealing and easy to comprehend and implement for moderate values of Z_n, but has few drawbacks. However the method is computationally intense when Z_n is large. Also computation of g_n for large n is really a

difficult job. If Z_n is large, then a large number of comparison have to be made to find the MLE. To apply this method when the generating function has $p_0 > 0$, unimodality condition like Theorem 3.4.1 needs to be proved for the generating function. Next we explore another method to estimate n by exploiting the Markov structure of Z_n.

3.5 Using Martingale Approach

For both method of moment estimate and MLE we have some constraints about the generating function. For method of moment estimate to work well we need a zero probability of extinction ($q = 0$) and to ensure a global maximum of likelihood for the existence of the MLE of n, we require the unimodality condition of Adès *et al.* (1982), which generally requires p_0=0. But the martingale method described below does not require any such restrictive assumption.

3.5.1 Exploring the Submartingale Structure

Again let $Z_n, Z_{n+1}, \ldots, Z_{n+r}$ be our observations of consecutive generations of a Galton-Watson process with progeny generating function $g(s)$. The supercriticality assumption ($m > 1$) is still needed to ensure that the process does not become extinct, so that the estimation problem of n is still well defined. The offspring distribution can be estimated by using the EM algorithm discussed earlier. Also here we additionally assume that the offspring distribution has a finite second moment which them implies $\mathrm{Var} Z_n < \infty$, $\forall n$. Define $M_n := (Z_n - m^n)$. Then M_{n+j} can be written as

$$\begin{aligned} Z_{n+j} - m^{n+j} = (Z_{n+j} - mZ_{n+j-1}) + m(Z_{n+j-1} - mZ_{n+j-2}) + m^2(Z_{n+j-2} - mZ_{n+j-3}) \\ + \cdots + m^{j-1}(Z_{n+1} - mZ_n) + m^j(Z_n - m^n) \end{aligned}$$

This clearly implies,

$$M_{n+j} = (T_{n+j} + mT_{n+j-1} + m^2T_{n+j-2} + \cdots + m^{j-1}T_{n+1}) + m^jM_n, \tag{3.9}$$

where $T_n := Z_n - mZ_{n-1}$.

Suppose $\Im_{n+j}$ is the minimal σ-field generated by $(Z_0, Z_1, \ldots, Z_{n+j})$. Then

$$E(M_{n+j}|\Im_{n+j-1}) = E(T_{n+j}|\Im_{n+j-1}) + mE(T_{n+j-1}|\Im_{n+j-1}) + \cdots$$
$$\cdots + m^{j-1}E(T_{n+1}|\Im_{n+j-1}) + m^j E(M_n|\Im_{n+j-1}) \tag{3.10}$$

Now, given $\Im_{n+j-1}$ with $j \geq 1$, the quantities M_n and $\{T_k,\ k = n+1, n+2, \ldots, n+j-1\}$ are constants. Thus to evaluate (3.10), only the first term in the right hand side needs to be computed. This is,

$$\begin{aligned} E(T_{n+j}|\Im_{n+j-1}) &= E(Z_{n+j} - mZ_{n+j-1}|\Im_{n+j-1}) \\ &= E(Z_{n+j}|\Im_{n+j}) - mE(Z_{n+j-1}|\Im_{n+j-1}) \\ &= mZ_{n+j-1} - mZ_{n+j-1} \\ &= 0 \end{aligned}$$

Thus,

$$\begin{aligned} \Rightarrow E(M_{n+j}|\Im_{n+j-1}) &= 0 + mT_{n+j-1} + m^2T_{n+j-2} + \cdots + m^{j-1}T_{n+1} + m^j M_n \\ &= m(T_{n+j-1} + mT_{n+j-2} + \cdots + m^{j-2}T_{n+1} + m^{j-1}M_n) \\ &= mM_{n+j-1} \\ &> M_{n+j-1} \qquad \text{since } m > 1 \end{aligned} \tag{3.11}$$

Hence, $\{M_n, \Im_n\}$ is a sub-martingale (See Appendix B, for a brief basics of the martingale theory). Now as x^2 is a convex function of x, $\{M_n^2,\ \Im_n\}$ is also a submartingale.

3.5.2 Estimator of the generation n (Age)

The underlying justification of the estimator, which will be proposed in this section, will lie in the following inequality, (See Sen and Singer (1993))

Theorem 3.5.1. (Hájek-Rènyi-Chow Inequality) If $\{X_n, \Im_n\}$ be a submartingale and let $\{c_n^*, n \geq 1\}$ be a nonincreasing sequence of positive numbers. Let $X_n^+ = \max\{X_n, 0\}$, and assume that EX_n^+ exists for every $n \geq 1$. Then, for every $\epsilon > 0$,

$$P\{\max_{1\leq k\leq n} c_k^* X_k^+ > \epsilon\} \leq \epsilon^{-1}\left\{c_n^* EX_1^+ + \sum_{k=2}^{n} c_k^* E(X_k^+ - X_{k-1}^+)\right\}$$

As we have shown that $\{M_n, \Im_n\}$ is positive valued sub-martingale, the above inequality can be used. Using Theorem (3.5.1) and choosing $c_n^* = \frac{1}{m^{\delta(r)}}$, $\forall n$, it follows that

$$P\{\max_{0\leq j\leq r} \frac{M_{n+j}^2}{m^{\delta(r)}} > \epsilon\} \leq \epsilon^{-1}\left\{\frac{EM_n^2}{m^{\delta(r)}} + \sum_{j=2}^{n} \frac{E(M_{n+j}^2 - M_{n+j-1}^2)}{m^{\delta(r)}}\right\}, \tag{3.12}$$

where $\delta(r) > 0$ is a arbitrary function of r with the property that $\dfrac{m^{\delta(r)}}{m^{2r}} \rightarrow \infty$ as $r \rightarrow \infty$. For example $\delta(r)$ can be chosen as $4r$.

For a Galton-Watson Process, it is well known that

$$\begin{aligned} EZ_n &= m^{n+j} \\ Var(Z_n) &= \frac{\sigma^2 m^{n-1}(m^n - 1)}{m-1}, \end{aligned}$$

where $\sigma^2 = \mathrm{Var}Z_1$. These properties can be used to simplify equation (3.12). Note that,

$$\begin{aligned} EM_{n+j}^2 &= E(Z_{n+j} - m^{n+j})^2 \\ &= \mathrm{Var}Z_{n+j} \\ &= \sigma^2 \frac{m^{n+j-1}(m^{n+j} - 1)}{m-1}, \quad j \geq 0 \end{aligned}$$

Using this, the inequality (3.12) can be rewritten as

$$P\{\max_{0\le j\le r}\frac{(Z_{n+j}-m^{n+j})^2}{m^{\delta(r)}}>\epsilon\}\le(\epsilon m^{\delta(r)})^{-1}[\sigma^2\frac{m^{n-1}(m^n-1)}{(m-1)}$$
$$+\frac{\sigma^2}{(m-1)}\sum_{j=0}^{r}\{m^{n+j-1}(m^{n+j}-1)-$$
$$m^{n+j-2}(m^{n+j-1}-1)\}] \tag{3.13}$$

Again,

$$\sigma^2((m-1)\epsilon m^{\delta(r)})^{-1}[m^{n-1}(m^n-1)+\sum_{j=0}^{r}\{m^{n+j-1}(m^{n+j}-1)-m^{n+j-2}(m^{n+j-1}-1)\}]$$
$$=\frac{\sigma^2}{(m-1)\epsilon m^{\delta(r)}}m^{n+r-1}(m^{n+r}-1)$$

$\Rightarrow$ RHS of equation (3.12) $\rightarrow 0$ as $r\rightarrow\infty$.

$$\Rightarrow P\{\max_{0\le j\le r}\frac{(Z_{n+j}-m^{n+j})^2}{m^{\delta(r)}}>\epsilon\}\ \rightarrow\ 0 \text{ as } r\rightarrow\infty.$$
$$\Rightarrow\ \max_{0\le j\le r}\frac{(Z_{n+j}-m^{n+j})^2}{m^{\delta(r)}}\xrightarrow{P}0 \text{ as } r\rightarrow\infty. \tag{3.14}$$

The convergence result in Equation (3.14) motivates us to propose the following:

$$\hat{n}_3'=\min_{n\ge1,\,m>1}\max_{0\le j\le r}\frac{(Z_{n+j}-m^{n+j})^2}{m^{\delta(r)}} \tag{3.15}$$

Here the optimization needs to be carried out with respect to two variables n and m simultaneously, which is computationally little bit difficult and getting a feasible solution may not be possible in many cases . Hence we propose to replace m in (3.14) by its plug in estimator $\hat{m}$. Thus the final form of the estimator is

$$\hat{n}_3=\min_{n\ge1}\max_{0\le j\le r}\frac{(Z_{n+j}-\hat{m}^{n+j})^2}{\hat{m}^{\delta(r)}}$$

The justification of the estimator comes from (3.14), which shows $\hat{n}_3$ has good large sample behavior. Note, computationally it is easier to implement than the MLE. The

most important feature of this estimator is that it does not depend on the specific properties of the offspring distribution.

3.6 Examples

Example 1:

The following data is generated from $g(s) = 0.1 + 0.3s + 0.6s^2$.

Table 3.3 Observations of 11 Consecutive Generations

k	Z_k
4	4
5	3
6	6
7	7
8	10
9	14
10	23
11	37
12	59
13	101
14	157

The EM algorithm is used to calculate the offspring distribution. The table gives you the comparison between estimates. The fractions are rounded off by taking least integer cantained in it.

Table 3.4 Estimation of Generation by Method of Moments and Martigale Method and MLE

Method	Estimated Value of Generation
Method of Moments	5
Martingale Method	13
MLE	8

Example 2:

The following data is generated from $g(s) = 0.01+0.1s+0.3s^2+0.25s^3+0.2s^4+0.14s^5$. Here clearly r=6 (since there are 7 observations). All three methods are applied to

Table 3.5 Observations

k	Z_k
2	5
3	13
4	41
5	135
6	403
7	1209
8	3508

estimate the age n (the generation label of the first observed value 5). The EM algorithm is used to calculate the offspring distribution. The table gives you the comparison between estimates.

Table 3.6 Estimation of Generation by Method of Moments, Martigale Method and MLE

Method	Estimated Value of Generation
Method of Moments	7
Martingale Method	9
MLE	X

3.6.1 Discussion

There are some interesting observations in this context. In Example 1, q=0. That is the reason the method of moment estimate is giving much better estimate than the other estimates. Another interesting feature in this example is though $p_0 \neq 0$ but this offpring distribution has a unimodal property. So the MLE can estimated and it is giving a pretty good result. But for the secong example q=0.011153, this detoriates the performance of method of moment estimate. Also the Martingale based estimator in this example does not appear to be very good. The main reason is sample size. From the assymptotic property of the estimate it can be intuitively said that for this estimator to work well, it is preferable to have a relatively large number of observation (large r). To verify this more data are drawn from the p.g.f. of Example 1, $g(s) = 0.1 + 0.3s + 0.6s^2$. The following table verifies the correctness of such intuition- based on computations, analogous to those in Example 1, carried out in progressively larger sample;

Table 3.7 Change in Martingale Method Estimate with r

r	n_3
10	13
20	14
30	10
40	11
50	9
60	9
70	7

3.7 Estimation of Probability of Extinction

As discussed earlier, one of the important factor of a Galton-Watson process is the probability of extintion. Again assume $Z_n, Z_2, \ldots, Z_{n+r}$ are $r+1$ generations of a Galton-Watson process with generating function $g(s)$. The assumptions $m > 1$ and $Z_0 = 1$ is still valid here. Because without supercriticality assumption the estimation of probability of extinction does not make any sense. Stigler(1971) obtained an estimator of the probability of extinction by estimating the offspring probabilities in parametric set up. Later Keiding(1976) used a martingale approach to find the probability of extinction of the whooping crane population of North America. He assumed the underlying generating function is negative binomial. Also Guttrop (1991) and Pakes (1975) studied non-parametric testing procedures for estimating the probability of extinction by exploring the martingale structure of $X_n = k^{(l)} - p_k N$. Pkes(1975) also discussed its asymptotic properties. But here the offspring probabilities are estimated by the considering all possible trees which can evolve the given data. Now as discussed in Chapter 2, probability of extinction(q) is the minimum root of the equation $g(s) = s$. This idea helped to propose a nonparametic

estimator for the probability of extinction. The estimator can be propose as;

$$\hat{q} = \inf\{s : \hat{g}(s) = s\}$$

where $\hat{g}(s)$ is the plug-in estimate of $g(s)$. The offspring probabilities are estimated by the EM algorithm described earlier. The asymptotic properties of the estimator can be established by using the following theorem proved by

Theorem 3.7.1. *(Guttrop 1991)* Let $\mathbf{p}_0$ be the true offspring distribution. If 1. $m > 1$, $(\partial/\partial s)g(s, \mathbf{p})$ is jointly continuous in s and $\mathbf{p}$, and $a_i = (\partial/\partial p_i)g(s, \mathbf{p}_0)$ exist.2. $N^{1/2}(\widetilde{\mathbf{p}} - \mathbf{p}_0) \xrightarrow{d} N(\mathbf{0}, \Sigma)$; Here, $\widetilde{\mathbf{p}}$ is an estimator of $\mathbf{p}$.

Now if an estimator of probability of extinction defined by,

$$\widetilde{q} = \inf\{s : g(s, \widetilde{\mathbf{p}}) = s\}$$

$$\text{Then,} \quad N^{1/2}(\widetilde{q} - q) \xrightarrow{d} N(0, h^{-2}\mathbf{a}'\Sigma\mathbf{a})$$

where $h = 1 - (\partial/\partial s)g'(s, \mathbf{p})$

These theorem ensures the asymptotic normality of $\hat{q}$.

CHAPTER 4

HYPOTHESIS TESTING IN GALTON-WATSON PROCESS

In this chapter a fundamental problem regarding Galton-Watson process has been explored. Suppose the evolution of a population follows a simple Galton-Watson process but the offspring distribution is not known. What can be said about the nature of the process by observing first few generations? In other words, from few observations is it possible to statistically infer whether that the process is going to extinct or explode in future? How does one conclude that the process is *'sub-critical'*, *'critical'* or *'super-critical'*. These questions can be mathematically formulated as the problem of testing an explosion vs extinction hypothesis; in other words as the problem of testing

$$H_0 \ : \ m > 1 \qquad vs \qquad H_1 \ : \ m \leq 1 \qquad\qquad (A_1)$$

where m is the mean of offspring distribution. The problem is challenging mainly for two reasons. First, as the observed values in a Galton-Watson process are from a Markov chain, they are not independently and identically distributed (i.i.d.). Thus common statistical testing procedures based on i.i.d. observations are not applicable in this case, second difficulty is regarding the estimation of the model parameters. As discussed in the previous chapter, an EM method is required to estimate the parameters. But in this method it is not possible to obtain a closed form estimate of the parameters. Also, due to dependence structure, one must be careful in applying the standard limit theorems to construct large sample tests.

4.1 Background Work

There is some literature regarding the testing problem mentioned above. But in most of the cases, such tests has been developed only for parametric family of

generating functions. Basawa and Scott (1976) has developed a procedure for testing such hypothesis under the assumption that the offspring distribution has a "power series" p.m.f. Later Basawa (1981) developed an conditional testing procedure to test the above hypothesis. Also this problem was dealt by Basawa and Scott (1987) by exploring the process structure and Sweeting(1978). In the next two sections a nontraditional approach has been taken to deal with such hypothesis, in a nonparametric set up.

4.2 Test based on Conditional Fisher Information

As explained in Chapter 3 (section 3.1), without loss of generality, we again assume that is $Z_0 = 1$ throughout this chapter. The first technique that has been used, is based on the asymptotic behavior of the of the maximum likelihood estimates. As discussed earlier there are some literature in branching processes where the testing problem has been explored for parametric cases, but no work has been for general nonparametric set up. In this chapter a methodology has been developed for testing the hypothesis in very general setup. Suppose $Z_1, Z_2, \ldots, Z_n$ are first n generations of a Galton-Watson process. We donot assume that the observation start at the first generation. They can start at any generation. The same methodology will be applicable on that case. The objective is to test the hypothesis stated in (A_1). Here also the only underlying assumption is that the offspring distribution has a finite support. If the offspring distribution has support $\{0, 1, 2, \ldots, M\}$ and $p_l, \quad l = 0, \ldots, M$ are the corresponding probabilities, then the nonparametric likelihood is given by,

$$L_n(\mathbf{p}) = \prod_{j=0}^{n} P(Z_{j+1} = k_{j+1} | Z_j = k_j) \tag{4.1}$$

where, $\mathbf{p} = (p_0, p_1, \ldots, p_{M-1})'$ and

$$P(Z_{j+1} = k_{j+1}|Z_j = k_j) = \frac{k_j!}{k_j^{(0)}!k_j^{(1)}!\ldots k_j^{(M)}!} p_0^{k_j^{(0)}} p_1^{k_j^{(1)}} \ldots p_M^{k_j^{(M)}}$$

where $k^{(l)}$'s are same as in equation (3.1) in Chapter 3. Suppose that $L_n(\mathbf{p})$is differentiable with respect to $\mathbf{p}$ and $E_{\mathbf{p}}(\frac{d \ln L_n(\mathbf{p})}{d\mathbf{p}})^2 < \infty$ for each n. Also suppose that $\Im_n$ is the minimal σ-field generated by $Z_1, Z_2, \ldots, Z_n$ and L_0=1.Let $\hat{\mathbf{p}}$ be the MLE of $\mathbf{p}$.

Looking into the hypotheses (A_1), it is clear that they constitute separable family of hypotheses, since under H_0 the underlying process of $Z_1, Z_2, \ldots$ being supercritical and thus exploding with positive probability, is completely different from the underlying process of $Z_1, Z_2, \ldots$ under H_1 which faces extinction with certainty. Let $\hat{\mathbf{p}}_{H_0}$ and $\hat{p}_{H_1}$ be the MLE of $\hat{\mathbf{p}}$ under H_0 and H_1 respectively. A test statistic defined a test statistic which can deal with such a situation, as defined by defined by Cox(1961), is

$$T_H = l_{H_0}(\hat{\mathbf{p}}_{H_0}) - E_{\hat{\mathbf{p}}_{H_0}}[l_{H_0}(\hat{\mathbf{p}}_{H_0})] - l_{H_1}(\hat{\mathbf{p}}_{H_1}) - E_{\hat{\mathbf{p}}_{H_1}}[l_{H_1}(\hat{\mathbf{p}}_{H_1})]$$

where $l_{H_i}(\mathbf{p}) = \ln L_n(\mathbf{p}_{H_i}|H_i)$. Taking log of Equation (4.1) ;

$$l(\mathbf{p}) = C + \sum_{l=0}^{M} k^{(l)} \ln p_l \tag{4.2}$$

where $k^{(l)}$ is the number of element gave birth to exactly l offspring in n generations and C is the term independent of $\mathbf{p}$. The log-likelihood can be approximately written as;

$$l_{H_i}(\mathbf{p}) = l_{H_i}(\hat{\mathbf{p}}_{H_i}) + \frac{1}{2}(\hat{\mathbf{p}}_{H_i} - \mathbf{p})' \left[\frac{\partial^2 l_{H_i}(\mathbf{p})}{\partial \mathbf{p}' \partial \mathbf{p}}\right]_{\mathbf{p}=\hat{\mathbf{p}}_{H_i}} (\hat{\mathbf{p}}_{H_i} - \mathbf{p}). \quad i = 0, 1$$

As $\frac{\partial l_{H_i}(\hat{\mathbf{p}}_{H_i})}{\partial \mathbf{p}}$ =0.

The conditional Fisher information is defined as

$$I_n(\mathbf{p}) := -E(\frac{\partial^2 l(\mathbf{p})}{\partial \mathbf{p}' \partial \mathbf{p}} | \Im_{n-1})$$

Then the unconditional Fisher information can be obtained by,

$$\xi_n(\mathbf{p}) = E_{H_i}(I_n(\mathbf{p}))$$

Cox(1982) has proved that if $I_n(\mathbf{p})[\xi_n(\mathbf{p})]^{-1} \xrightarrow{P} 0$ then T_h has a asymptotic normal distribution. . Here the two following theorem will show that Cox's method is not applicable in the current context.

Theorem 4.2.1. If $Z_1, Z_2, \ldots, Z_n$ are first n generations of a Galton-Watson process with generating function $g(s)$, which has a finite support. Then,

$$I_n(\mathbf{p})[\xi_n(\mathbf{p})]^{-1} \xrightarrow{a.s.} \begin{cases} \frac{1-m}{m} \sum_{k=0}^{T} Z_k \mathbf{I}_{M-1\times M-1}, & \text{under } H_1. \\ W\mathbf{I}_{M-1\times M-1}, & \text{under } H_0. \end{cases}$$

W is defined in Chapter 2 and T is the time to extinction of the process under the alternative hypothesis H_1.

To prove Theorem 4.2.1 the following Lemma is need to stated.

Lemma 4.2.2. *(Toeplitz Lemma)(Hall and Heyde 1980)* Let a_{ni}, $1 \leq i \leq k_n$, $n \geq 1$, and x_i, $i \geq 1$, be real numbers such that for every fixed i, $a_{ni} \to 0$ and for all n, $\sum_i |a_{ni}| \leq C < \infty$. If $x_n \to 0$, then $\sum_i a_{ni}x_i \to 0$, and if $\sum_i a_{ni} \to 1$, then $x_n \to x$ ensures that $\sum_i a_{ni}x_i \to x$. In particular, if a_i, $i \geq 1$, are positive numbers and $b_n = \sum_{i=1}^n a_i \uparrow \infty$, then $x_n \to x$ ensures that $b_n^{-1} \sum_{i=1}^n a_i x_i \to x$.

Proof of Theorem 4.2.1

Equation(4.2) can be rewritten as ;

$$l(\mathbf{p}) = c + \sum_{l=0}^{M-1} k^{(l)} \ln p_l + (N - \sum_{l=0}^{M-1} k^{(l)}) \ln(1 - \sum_{l=0}^{M-1} p_l) \tag{4.3}$$

with $N = \sum_{j=1}^{n} Z_j$. Now using Equation (4.3) one can obtain the following;

$$-\frac{\partial^2 l(\mathbf{p})}{\partial \mathbf{p}' \partial \mathbf{p}} = \begin{pmatrix} \frac{k^{(0)}}{p_0^2} + \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \cdots & \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} \\ \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \frac{k^{(1)}}{p_0^2} + \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \cdots & \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} & \cdots & \frac{k^{(0)}}{p_{M-1}^2} + \frac{N-\sum_{l=0}^{M-1} k^{(l)}}{(1-\sum_{l=0}^{M-1} p_l)^2} \end{pmatrix}$$

$$-\frac{\partial^2 l(\mathbf{p})}{\partial \mathbf{p}' \partial \mathbf{p}} = \mathrm{Diag}\{\frac{k^{(0)}}{p_0^2}, \frac{k^{(1)}}{p_1^2}, \cdots, \frac{k^{(M-1)}}{p_{M-1}^2}\} + \frac{N - \sum_{l=0}^{M-1} k^{(l)}}{(1 - \sum_{l=0}^{M-1} p_l)^2} \mathbf{1}'\mathbf{1}$$

Thus,

$$I_n(\mathbf{p}) := -E(\frac{\partial^2 l(\mathbf{p})}{\partial \mathbf{p}' \partial \mathbf{p}} | \Im_{n-1})| = \mathrm{Diag}\{\frac{E(k^{(0)}|\Im_n)}{p_0^2}, \frac{E(k^{(1)}|\Im_n)}{p_1^2}, \cdots, \frac{E(k^{(0)}|\Im_n)}{p_{M-1}^2}\} + \frac{E(N - \sum_{l=0}^{M-1} k^{(l)}|\Im_n)}{(1 - \sum_{l=0}^{M-1} p_l)^2} \mathbf{1}'\mathbf{1}$$

The factor $k^{(l)}$ can be written as

$$k^{(l)} = k_1^{(l)} + k_2^{(l)} + \cdots + k_n^{(l)} \tag{4.4}$$

where $k_j^{(l)}$ denotes the number of element in the j-th generation that gave birth to exactly l offsprings. From the offspring structure it is clear that for given $\{Z_1, Z_2, \ldots, Z_n\}$, $k_j^{(l)} \sim \mathrm{Bin}(Z_j, p_l)$. $\Rightarrow$ $E(k^{(l)}|\Im_n) = Np_l$.Thus $I_n(\mathbf{p})$ becomes;

$$I_n(\mathbf{p}) = N[\mathrm{Diag}\{\frac{1}{p_0}, \frac{1}{p_1}, \cdots, \frac{1}{p_{M-1}}\} + \frac{1}{1 - \sum_{l=0}^{M-1} p_l} \mathbf{1}'\mathbf{1}]_{\hat{\mathbf{p}}_{H_i}}$$

Thus,

$$\xi_n(\mathbf{p}) = E(N)[\text{Diag}\{\frac{1}{p_0}, \frac{1}{p_1}, \cdots, \frac{1}{p_{M-1}}\} + \frac{1}{1-\sum_{l=0}^{M-1} p_l}\mathbf{1}'\mathbf{1}]$$

Now $E(N) = \sum_{j=1}^{n} E(Z_j) = \sum_{j=1}^{n} m^j$. That implies,

$$I_n(\mathbf{p})[\xi_n(\mathbf{p})]^{-1} = N \times [E(N)]^{-1} \times I_{M-1\times M-1}$$

<u>Case I</u>: $m < 1$, ie, under H_1

As $m < 1$,

$$\begin{aligned} E(N) = \sum_{j=1}^{n} m^j &= \frac{m(1-m^n)}{1-m} \\ &\rightarrow \frac{m}{1-m} \qquad \text{as } n \rightarrow \infty \end{aligned}$$

Again under H_1 the process has certain extinction. So $\sum_{j=1}^{n} Z_j \overset{a.s}{\rightarrow} \sum_{j=1}^{T} Z_j$ as $n \rightarrow \infty$.
$\Rightarrow \quad N/E(N) \overset{a.s.}{\rightarrow} \frac{(1-m)\sum_{j=1}^{T} Z_j}{m}$ as $n \rightarrow \infty$.

<u>Case II</u>: $m > 1$, ie, under H_0

In Toeplitz lemma if $a_k = m^k$ then $b_n = \sum_{k=1}^{n} a_k = \sum_{k=1}^{n} m^k \uparrow \infty$ as $n \uparrow \infty$. Then

$$\frac{N}{E(N)} = \frac{\sum_{j=1}^{n} Z_j}{\sum_{j=1}^{n} m^j} = \frac{\sum_{j=1}^{n} \frac{Z_j}{m^j} m^j}{\sum_{j=1}^{n} m^j} \tag{4.5}$$

now $\frac{Z_j}{m^j} \longrightarrow W$ a.s as $j \rightarrow \infty$. So applying Topelitz lemma to Equation(4.4),
$\frac{N}{E(N)} \overset{a.s}{\rightarrow} W$ as $n \rightarrow \infty$.
Hence the proof.

This theorem proves the contiguity that is need for both null and alternative hypothesis is missing here. Under null hypothesis the ratio is converging to a nontrivial

positive valued random variable. So Cox method will not work in this case. So a alternative test can be suggested by the following way;

$$T_{n1} = \mathbf{1}'[\{[I_n(\mathbf{p})[\xi_n(\mathbf{p})]^{-1}]_{H_0}\}\{[I_n(\mathbf{p})[\xi_n(\mathbf{p})]^{-1}]_{H_1}\}^{-1}]\mathbf{1}$$

now using Theorem (4.2.1), the asymptotic null distribution of $T_{n1} \xrightarrow{a.s} T_1 = (mW/\{(1-m)\sum_{k=0}^{T} Z_k\})$ as $n \to \infty$. So this test statistic can be used to construct a asymptotic test for testing H_0 against H_1. But the problem is find the exact distribution for T_1 is really difficult. So bootstrap technique can be used to solve that problem.The outline of the method is as follows;

- from the given data a considerable number of bootstrap samples are chosen.
- For each bootstrap sample T_{n1} is calculated.
- This gives an empirical distribution of T and that is used in taking the decision.

4.2.1 Example

Consider Table 3.3. If T_{n1} calculated for this data and T_{n1}= 0.47. 95-th percentile of T_{n1} calculated with 5000 bootstrap sample is 3.56. So H_0 is not rejected at 5% level of significance. So the process is probably Supercritical. This shows that test procedure is working.

4.3 Using Least Favorable Setup

The method described in previous section is intuitively very appealing and also easy to implement. But one another problem is as it is not possible to find the exact distribution of T_1 for most of the cases there is a possibility of significant reduction in the power. In this section another method is discussed based on the *'least favorable null hypothesis'*.This is still an ongoing process. For simplicity assume, M=2. That means the offspring distribution has support in $\{0, 1, 2\}$ and $p_0, p_1, p2$

are corresponding probabilities. On that consider the following hypothesis;

$$H_0 \ : \ m \geq 1 \qquad vs \qquad H_1 \ : \ m < 1 \qquad\qquad (A_1')$$

This hypothesis can be equivalently stated as,

$$H_0 \ : \ p_2 > p_0 \qquad vs \qquad H_1 \ : \ p_2 < p_0 \qquad\qquad (A_1')$$

The generalization of this method for $M > 2$ is discussed at the end of this section.Before describing the method it is required to give definitions of some important notions;

4.3.1 Least Favorable Null Hypothesis

Suppose $X_1, X_2, \ldots, X_n$ be iid observations from a distribution with p.d.f. $f(\theta)$. Consider the following testing problem,

$$H_0' \ : \ \theta \in \mathcal{C} \qquad vs \qquad H_1' \ : \ \theta \notin \mathcal{C} \qquad\qquad (A_2)$$

So here both the null and alternative hypothesis are composite.Let L_{A_2} be the likelihood ratio test(LRT) for testing H_0' and H_1'. If l_{A_2} is the observed value of L_{A_2}. It is required to find the p-value of the test for conclusion. Now,

$$p - \text{value} := P(L_{A_1} \geq l_{A_1} | H_0)$$

As the null hypothesis is composite this probability depends on particular null value of θ, which anywhere in the null parameter space. Thus,$P_\theta(L_{A_1} \geq l_{A_1} | \theta \in H_0)$ is not just a fixed number on the null parameter space, but a function of θ, and hence does not define a p-value. In this case, a reasonable approach to overcome this difficulty appears to be not to reject H_0' if there is at least one value in the $\mathcal{C}$ with which the data are consistent; or equivalently reject H_0' if the data are inconsistent for all $\theta \in \mathcal{C}$. In this situations the usual procedure to define the p-value is as follows;

$$p - \text{value} := \sup_{\theta \in H_0'} P(L_{A_1} \geq l_{A_1}) \qquad (4.6)$$

Suppose the supremum is achieved at $\theta = \theta_0$. So it can be said that the strength of evidence against H_0' and in favor of H_1' based on $L_{A_2} = l_{A_2}$ depends on the assumed true value of $\theta \in \mathcal{C}$ and that is *'least'* when $\theta = \theta_0$. That is the reason $H_2' : \theta = \theta_0$ is called the *least favorable null value* for L_{A_2} (also called the *least favorable null configuration* of L_{A_2}. and the distribution of L_{A_2} is called *least favorable null distribution* of L_{A_2}. Further discussion on *least favorable null configuration* is available on Lehman (1994) and Sivapulle & Sen (2005).

It is time now to define the following theorem;

Theorem 4.3.1. Suppose $Z_1, Z_2, \ldots, Z_n$ are n-generations of a Galton-Watson process with generating function $g(s) = p_0 + p_1 s + p_2 s^2$. Then the testing problem A_1 or equivalently A_1' the *the least favorable null hypothesis* is $H_2 : m = 1$ or equivalently $H_2 : p_2 = p_0$

Proof. Using Equation(4.2) the log-likelihood can be written as;

$$l(p_0, p_1, p_2) = C_2 + k^{(0)} \ln p_0 + k^{(2)} \ln p_2 + (N - k^{(0)} - k^{(2)} \ln(1 - p_0 - p_2) \qquad (4.7)$$

Define; $\delta = p_2 - p_0 \quad \Rightarrow \delta > 0$ under H_0.

Now differentiating Equation(4.8) with respect to θ,

$$\begin{aligned}
\frac{\partial l(p_0, p_1, p_2)}{\partial \delta} &= \frac{\partial l(p_0, p_1, p_2)}{\partial p_0}\frac{\partial p_0}{\partial \delta} + \frac{\partial l(p_0, p_1, p_2)}{\partial p_2}\frac{\partial p_2}{\partial \delta} \\
&= -(\frac{k^{(0)}}{p_0} - \frac{N - k^{(0)} - k^{(2)}}{1 - p_0 - p_2}) + (\frac{k^{(2)}}{p_2} - \frac{N - k^{(0)} - k^{(2)}}{1 - p_0 - p_2}) \\
&= \frac{k^{(2)}}{p_2} - \frac{k^{(0)}}{p_0} \\
&= \frac{k^{(2)}}{p_0 + \delta} - \frac{k^{(0)}}{p_0} \downarrow \delta
\end{aligned}$$

This means the likelihood function is monotonically decreasing in δ. From here it is clear that the supremum in Equation(4.7) will be achieved when δ=0 $\Leftrightarrow p_2 = p0$. Hence the proof.

Theorem (4.3.1) implies that the least favorable set up for testing,

$$H_0 \ : \ p_2 > p_0 \qquad vs \qquad H_1 \ : \ p_2 < p_0 \qquad (A_1')$$

$$\text{is } H_2 \ : \ p_2 = p_0 \qquad vs \qquad H_1 \ : \ p_2 < p_0 \qquad (A_1')$$

For performing such a testing for the given set up there are other tolls required. One of them is $\bar{\chi}^2$. In next subsection a brief introduction for $\bar{\chi}^2$ is illustrated.

4.3.2 $\bar{\chi}^2$ Distribution

Let $\mathcal{C} \subset R^p$ ((C) is a closed convex cone) and let $\mathbf{Z}_{p\times 1} \sim N(0, \mathbf{V})$, where V is a positive definite matrix . The testing problem in interest is;

$$\text{is } H_0 \ : \ \theta = 0 \qquad vs \qquad H_1 \ : \ \theta \in \mathcal{C} \qquad (A_3)$$

Then $\bar{\chi}^2(\mathbf{V}, \mathcal{C})$ can be defined as the random variable, which is defined as,

$$\bar{\chi}^2(\mathbf{V}, \mathcal{C}) = \mathbf{Z}'\mathbf{V}\mathbf{Z} - \min_{\theta \in \mathcal{C}} (\mathbf{Z} - \theta)'\mathbf{V}(\mathbf{Z} - \theta) \tag{4.8}$$

Or in other words the null distribution of the LRT for testing hypothesis of type A_3 is called the $\bar{\chi}^2$ distribution. The formula for computing $\bar{\chi}^2$ is given in the next theorem;

Theorem 4.3.2. *(Gourieroux et. al.(1982))* Let $\mathcal{C}$ be a closed convex cone in R^p and $\mathbf{V}$ be a $p \times p$ positive definite matrix. Then the distribution of $\bar{\chi}^2(\mathbf{V}, \mathcal{C})$ is given by,

$$P(\bar{\chi}^2(\mathbf{V}, \mathcal{C}) \leq c) = \sum_{i=0}^{p} w_i(p, \mathbf{V}, \mathcal{C}) P(\chi_i^2 \leq c)$$

where $w_i(p, \mathbf{V}, \mathcal{C}),\ i = 0 \ldots p$ are some non-negetive numbers and $\sum_{i=0}^{p} w_i(p, \mathbf{V}, \mathcal{C}) = 1$.

The method for determining $w_i(p, \mathbf{V}, \mathcal{C})$ discussed in details in Sen and Silvapulle (2005).

4.3.3 Test Statistics

Using the theorem above and applying related techniques from Sen and Silvepulle (2005)the following the likelihood ratio test(LRT) statistic can be derived as to test A_3.

$$LRT = 2(l(\hat{\mathbf{p}}) - l(\widetilde{\mathbf{p}}))$$

where $\hat{\mathbf{p}}$ is the MLE of $\mathbf{p} = (p_0, p_2)$ under H_1 and $\widetilde{\mathbf{p}}$ is MLE under H_2. This is a work in progress to find the asymptotic distribution of LRT by using the results of $\bar{\chi}^2$ distribution. This is still an ongoing process. The analogy is same as hypothesis testing of mean of normal distribution under order constraint (see Sen & Sivapulle (2005)) . This notion can be expanded for $M > 2$ with just little bit modification of hypothesis.

CHAPTER 5

A FAMILY OF PROBABILITY GENERATING FUNCTIONS INDUCED BY SHOCK MODELS

5.1 The Problem

The question we want to investigate can be simply posed as follows. If $\mathbf{Q}$ is a probability measure on the half line, under what conditions, is the function defined by

$$\int_0^\infty \frac{yz}{1-z+yz}\, \boldsymbol{Q}(dy) \tag{5.1}$$

a *probability generating function* (p.g.f) of a positive integer valued random variable N?

For any y in $(0,1)$, recognizing the integrand to be the p.g.f. of a geometric distribution over the positive integers; the answer is clearly affirmative if the support of the mixing distribution $\boldsymbol{Q}$ is no larger than (0,1]. The case $y = 1$ corresponds to a mixing distribution degenerate at 1. For $y \in (0,1]$, we can think of N as conditionally geometric given y so that Equation 5.1 is the unconditional p.g.f of N, when the parameter y is randomized over the unit interval. In other words, if $\boldsymbol{Q}(0,1] = 1$, then the function defined by 5.1 is a Bayesian's view of the p.g.f of N when y has a prior $\boldsymbol{Q}$. In fact if $\{X_1, X_2, \ldots\}$ is a sequence of binary exchangeable random variables, then using the fact that for any integers n, k such that $1 \le k \le n$, any $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$

$$\begin{aligned} P\{X_{i_1} = X_{i_2} = \cdots = X_{i_k} = 1,\ X_j = 0 \text{ for } j \in \{1, 2, \cdots, n\} \setminus \{i_1, \cdots, i_k\}\} \\ = \int_0^1 y^k (1-y)^{n-k}\, \boldsymbol{Q}(dy). \end{aligned}$$

For some probability measure $\boldsymbol{Q}$ in (0,1], by the classic result of DeFinetti, Feller (1961), it easily follows that the random variable

$$N := \inf\{n \geq 1 \ : \ X_n = 1\}$$

indeed has the p.g.f.;

$$\int_0^1 \frac{yz}{1 - z + yz} \, \boldsymbol{Q}(dy)$$

for some unique measure $\boldsymbol{Q}$ supported by the unit interval.

However, the answer of our question is not clear, since the integrand in Equation (5.1) is not a p.g.f. for $y > 1$. This leads us to ask: if the support of $\boldsymbol{Q}$ on $\Re_+$ extends beyond $[0, 1]$, i.e., if $\boldsymbol{Q}(1, \infty) > 0$, can (5.1) still be the p.g.f. of a positive valued random variable N?

5.2 Motivation and Main Results

The motivation for this problem comes from the following observations. Let

$$\bar{S}(t) = \sum_{k=0}^{\infty} \bar{P}_k e^{-\lambda t} \frac{(\lambda t)^k}{k!} \tag{5.2}$$

be the standard Essary Marshall and Proschan (henceforth abbreviated as EMP) shock model (1973) survival probability, where failure is caused by shocks arising over time according a homogeneous Poisson process $\{N(t); t \geq 0\}$ with intensity $\lambda > 0$ and the distribution of $J :=$ the number of shocks to failure, has tail $\bar{P}_k := P(J > k)$. EMP(1973) proved that all the standard non-parametric positive and negative aging properties of J in discrete time are preserved by the survival probability $\bar{S}$ in continuous time.

Our problem stated in Section 5.1, was motivated by an apparently surprising connection, via geometric distributions, between the structure of the Laplace-Stieltje's

transform of EMP shock model distributions and a new class of negatively aging nonparametric life distributions recently introduced and studied by Bhattacharjee (2005). The next four results (Lemma 5.2.1, Theorems 3.2.2-3.2.4) makes this connection clear and puts our motivation perspective.

Definition 5.2.1. A non-discrete lifetime $X \backsim F$ has the SDFR (Strongly Decreasing Failure Rate) property, if the tail (ie, the reliability) function of X ; $\bar{F}(t) := P(X > t)$ is a *Completely Monotone Function* (see Feller 1939) on $[0, \infty)$.

Definition 5.2.2. A non-negative integer valued discrete lifetime X has the discrete SDFR property if its tail probabilities $\overline{P}_k := P(X > k)$, $k = 0, 1, 2, \cdots$, is a *Completely Monotone Sequence* (Feller 1939).

Lemma 5.2.1. The Laplace-Stieltjes transformation of S in (2) is given by

$$L(s) = \phi\left(\frac{\lambda}{\lambda + s}\right), \quad s > 0,$$

where $\phi(\cdot)$ is the p.g.f. of the random number J of shocks to failure.

Proof. Simply note that

$$\begin{aligned} s^{-1}\{1 - L(s)\} &= \int_0^\infty \exp^{-st} \overline{S}(t)dt = \sum_{k=0}^{\infty} \overline{P}_k \int_0^\infty \exp^{-(\lambda+s)t} \frac{(\lambda t)^k}{k!} \\ &= (\lambda + s)^{-1} \sum_{k=0}^{\infty} \overline{P}_k (\frac{\lambda}{\lambda + s})^k \\ &= (\lambda + s)^{-1} \frac{1 - \phi(\frac{\lambda}{\lambda+s})}{1 - \frac{\lambda}{\lambda+s}} \\ &= s^{-1}\{1 - \phi(\frac{\lambda}{\lambda + s})\}. \end{aligned}$$

□

Bhattacharjee (2005) has proved the characterization of SDFR by the following theorem.

Theorem 5.2.2. (i) $X \sim F$ is non discrete SDFR iff it has a representation

$$F(t) = \int_0^\infty (1 - e^{-\lambda t}) G(dt)$$

with a unique mixing distribution G continuous at zero.

$$\text{i.e. } X \text{ is SDFR} \Leftrightarrow X \stackrel{d}{=} \frac{Y}{Z},$$

where $\stackrel{d}{=}$ denoted equality in distribution, Y and $Z > 0$ are independent random variables, and $Y \sim$ Exp(mean=1).

(ii) A discrete non-negative integer valued random variable X is discrete SDFR iff its tail probabilities $\overline{P}_k := P(X > k)$, $k = 0, 1, 2, \cdots$, has a unique representation

$$\overline{P}_k = \int_0^1 p^k Q(dp),$$

for some probability measure Q on $[0, 1]$, with $Q\{0\} < 1$.

These above results motivate us to investigate closure properties of the EMP Shock Models with respect to the new non-parametric aging notion SDFR. We have proved the following results.

5.2.1 Main Results

Theorem 5.2.3. If $\bar{P}_k$ is SDFR $\Rightarrow$ S is SDFR.

Proof. $\bar{P}_k$ is SDFR $\Leftrightarrow$ $\bar{P}_k = \int_0^1 p^k \; \boldsymbol{Q}(dp)$,

by Theorem 5.2.2 (ii). hence for any $\lambda > 0$, the corresponding EMP shock model survival probability $\overline{S}$ with discrete SDFR shock resistance probabilities $\overline{P}_k$, can be

expresses using (5.2), as

$$\begin{aligned}\bar{S}(t) &= \sum_{k=0}^{\infty}\left(\int_0^1 p^k\ \boldsymbol{Q}(dp)\right)\ e^{-\lambda t}\frac{(\lambda t)^k}{k!}\\ &= \int_0^1 e^{-\lambda t(1-p)}\ \boldsymbol{Q}(dp)\\ &= \int_0^{\lambda} e^{-\theta t}\boldsymbol{G}(d\theta),\end{aligned}$$

for some distribution G on $[0,\lambda], \lambda > 0$. Theorem 5.2.2 (i) now implies $\bar{S}$ is non-discrete SDFR. □

Is the converse is true? The following theorem gives a necessary and sufficient condition for the converse to be true, which led us to the question posed in Section 5.1.

Theorem 5.2.4. The EMP Shock Model distribution function S is SDFR iff the number of shocks to failure has a probability generating function $\phi(z) = EZ^J$ with a unique representation

$$\phi(z) = \int_0^{\infty} \frac{zy}{(1-z)+zy}\ \boldsymbol{Q}(dy) \tag{5.3}$$

for some mixing distribution function $\boldsymbol{Q}$ with support in the half line$[0,\infty)$.

Proof. By Lemma 5.2.1 for $s > 0$, we have

$$\begin{aligned}\phi\left(\frac{\lambda}{\lambda+s}\right) &= E(e^{-sT}) \text{ where } T \text{ has tail } \bar{S}(t) \text{ in (5.2)}\\ &= \int_0^{\infty} e^{-st}S(dt)\\ &= \int_0^{\infty}\int_0^{\infty} \theta e^{-(\theta+s)}\ G(d\theta)\ dt, \text{(using Theorem 5.2.2 (i))}\\ &= \int_0^{\infty} \frac{\theta}{\theta+s}\ G(d\theta), \text{ (using Fubini's Theorem).}\end{aligned}$$

Setting $z = \dfrac{\lambda}{\lambda+s} \in (0,1)$ as $s \in (0,\infty)$, with $\lambda > 0$, this yields

$$\begin{aligned}
\phi(z) &= \int_0^\infty \frac{\theta/\lambda}{\theta + \frac{1}{z} - 1}\, G(d\theta) \\
&= \int_0^\infty \frac{\theta}{\frac{\theta}{\lambda} + \lambda(\frac{1}{z} - 1)}\, G(d\theta) \text{ with } G\{0\} = 0, \text{ by Theorem 5.2.2 (i)} \\
&\equiv \int_0^\infty \frac{zy}{(1-z) + zy}\, \mathbf{Q}(dy), \text{ with} \\
&= E\left(\frac{zY}{1 - z + Y}\right)
\end{aligned}$$

when $\dfrac{\theta}{\lambda} := Y \sim \boldsymbol{Q}$. Note, since G, the distribution of θ, is continuous at zero, so is $\mathbf{Q}$. □

Thus, if the EMP shock model probability $\bar{S}$ is to be a completely monotone function (i.e., S is to be SDFR) then the mixing distribution $\mathbf{Q}$ must have support in $[0, \infty)$; i.e., $\{y : 0 < \mathbf{Q}(y) < 1\} \subset [0, \infty)$.

At this point a obvious question is; what more can we say about $\boldsymbol{Q}$? In particular, what should be the support of $\boldsymbol{Q}$? Are there any necessary and sufficient conditions on $\boldsymbol{Q}$ such that the right hand side of (5.1) is always a probability generating function? In search of an answer to this question we have found an apparently surprising necessary condition (Theorem 5.2.6). As a preliminary, we need the following lemma.

Lemma 5.2.5. For any non-negative integer valued random variable N with distribution,

$$q_n := P(N = n); \qquad n = 0, 1, 2, \cdots$$

we have

$$E(1-z)^N = \sum_{k=0}^{\infty} c_k z^k \qquad 0 < z < 1,$$

where $$c_k = (-1)^k \textstyle\sum_{n=k}^{\infty} \cdot \begin{pmatrix} n \\ k \end{pmatrix} q_n, \quad k = 0, 1, 2, \cdots$$

Proof. Simply note,

$$\begin{aligned} E(1-z)^N &= \sum_{n=0}^{\infty}(1-z)^N q_n \\ &= \sum_{n=0}^{\infty}\{\sum_{k=0}^{n}\binom{n}{k}(-z)^k\}q_n \\ &= \sum_{k=0}^{\infty}\{\sum_{n=k}^{\infty}\binom{n}{k}q_n\}(-z)^k \end{aligned}$$

This leads us to the following result. □

Theorem 5.2.6. If the right hand side of (5.1)is a probability generating function of a nonnegative random variable N then $\boldsymbol{Q}[2,\infty) = 0$ i.e., $\boldsymbol{Q}$ cannot have support beyond (0,2).

Proof. Rewriting the integral representation (5.3) for $\phi(Z)$ as,

$$\begin{aligned} \phi(z) &= E\left(\frac{zY}{1-z+Y}\right) \qquad 0 \le Y \sim \boldsymbol{Q} \\ &= \int_0^1 \psi_y(z)\,\boldsymbol{Q}(dy) + \int_1^{\infty}\{1-\psi_{\frac{1}{y}}(1-z)\}\,\boldsymbol{Q}(dy), \end{aligned}$$

where $\psi_y = Ez^{N_y}$, $N_y \sim Geometric(y)$, $0 \le y \le 1$, havig the geometric distribution. Therefore, $P(N_y = k) = y(1-y)^{k-1}$, $k = 1, 2, \cdots$, implies

$$\phi(z) \equiv \int_0^1 Ez^{N_y}\boldsymbol{Q}(dy) + \int_1^{\infty}\{1-\ E(1-z)^{N_{\frac{1}{y}}}\}\boldsymbol{Q}(dy) \tag{5.4}$$

Hence, for the EMP Shock Model (5.2), the shock resistant probabilities $\bar{P}_k = P(J > k)$, have generating function,

$$\begin{aligned} M(z) &\equiv \sum_{k=0}^{\infty}\bar{P}_k z^k \\ &= \frac{1-\phi(z)}{1-z}, \qquad 0 < z < 1 \\ &= \int_0^1 \frac{1-Ez^{N_y}}{1-z}\boldsymbol{Q}(dy) + \int_1^{\infty} E(1-z)^{N_y^*}\boldsymbol{Q}(dy), \end{aligned} \tag{5.5}$$

where $N_y^* := N_{\frac{1}{y}} - 1$, for $y > 1$.

Applying Lemma 5.2.5 to the integrand in the second term of (5.5) with

$$\begin{aligned} q_n &= P(N_y^* = n) \qquad n = 0, 1, 2, \cdots \\ &= P(N_{\widetilde{\Pi}} = n+1), \text{ where } \widetilde{\Pi} = \frac{1}{y} < 1 \\ &= \widetilde{\Pi}(1-\widetilde{\Pi})^n, \end{aligned}$$

we get

$$\begin{aligned} c_k &= (-1)^k \sum_{n \geq k} \binom{n}{k} \widetilde{\Pi}(1-\widetilde{\Pi}) \\ &= (-1)^k \sum_{j=0}^{\infty} \binom{k+j}{k} (1-\widetilde{\Pi})^{k+j} \\ &= (-1)^k \frac{\widetilde{\Pi}(1-\widetilde{\Pi})^k}{\widetilde{\Pi}^{k+1}} [\sum_{j=0}^{\infty} \binom{k+j}{k} \widetilde{\Pi}^k (1-\widetilde{\Pi})^j \widetilde{\Pi}\,] \\ &= (-1)^k \frac{\widetilde{\Pi}(1-\widetilde{\Pi})^k}{\widetilde{\Pi}^{k+1}} \sum_{j=0}^{\infty} P\{(k+1)\text{th success in}(k+j+1)\text{th trial}\} \\ &= (-1)^k \left(\frac{1-\widetilde{\Pi}}{\widetilde{\Pi}}\right)^k \\ &= (1-y)^k, \end{aligned} \tag{5.6}$$

where $\widetilde{\Pi} = \frac{1}{y} < 1$. Note, the right hand side of (5.6) is positive for k even and negative for k odd.

Using this in (5.5) we have

$$\sum_{k=0}^{\infty} \bar{P}_k z^k = \int_0^1 \left\{ \sum_{k=0}^{\infty} y^k z^k \right\} \boldsymbol{Q}(dy) + \int_1^{\infty} \left\{ \sum_{k=0}^{\infty} (-1)^k (y-1)^k z^k \right\} \boldsymbol{Q}(dy) \tag{5.7}$$

In the first term the right hand side of (5.7), the series is absolutely convergent so the integral and the summation can be interchanged. But the series in the integrand of the second term converges iff $y < 2$. So if $\boldsymbol{Q}[2,\infty) > 0$ then the right hand side of

(5.7) diverges. Thus, we must have $\boldsymbol{Q}[2,\infty) = 0$, for the generating function of the shock resistance probabilities in (5.7) to converge for all $0 < z < 1$. □

Note, for all $z \in (0,1)$, we can thus write,

$$\sum_{k=0}^{\infty} \bar{P}_k z^k = \sum_{k=0}^{\infty} \left\{ \int_0^2 a^k(y) \boldsymbol{Q}(dy) \right\} z^k, \tag{5.8}$$

where

$$a(y) = \begin{cases} y, & \text{for } 0 \leq y < 1 \\ 1-y, & \text{for } 1 < y \leq 2. \end{cases}$$

Equating coefficients of z^k on both sides of (5.8), we get

$$\bar{P}_k = \int_0^2 a^k(y) \boldsymbol{Q}(dy), \qquad k = 0, 1, 2, \cdots \tag{5.9}$$

Theorem 5.2.5 implies for the function $\phi(z)$ defined by (5.1) to be a p.g.f., we must have $\boldsymbol{Q}[2,\infty)=0$. Is this the sharpest possible result? Or is there a sharper necessary condition? Also, are there interesting/nontrivial sufficient condition, other than $\boldsymbol{Q}$ to be supported by the unit interval, to ensure $\phi(.)$ to be p.g.f.?

The question remains : for what conditions on $\boldsymbol{Q}$, is $\bar{P}_k$ is a tail probability of a discrete non-negative random variable? i.e., what conditions on $\boldsymbol{Q}$ would ensure that, $\bar{P}_k$ non-negative and monotonically non-increasing? We have $\bar{P}_k \to 0$ directly from (5.9) the expression, so that proving $\overline{P}_k \downarrow$ guarantees that $\overline{P}_k$ represents the tail probabilities of an honest distribution.

As remarked, if $\boldsymbol{Q}(0,1]=1$, then the function in (5.3) is trivially a p.g.f, representing a mixture of geometric distributions. Contrary to crude intuition the function ϕ defined via the integral in (5.3) can be a p.g.f and the mixing distribution $\boldsymbol{Q}$ can have positive mass in the interval (1,2) as the following example shows

Counter example Define the c.d.f. Q by

$$\boldsymbol{Q}(y) = \begin{cases} (1-\beta)y & \text{if } 0 < y < 1 \\ (1-\beta) + \frac{\beta}{\alpha}(y-1) & \text{if } 1 \leq y < (1+\alpha) \\ 1 & \text{if } y \geq 1+\alpha \end{cases}$$

and choose the parameters α, β to satisfy $0 < \alpha < \frac{2}{7}$ and $\beta > \frac{1}{3}$ Using this $\boldsymbol{Q}$, and (5.9), we can easily compute

$$\bar{P}_k = \frac{1-\beta}{k+1} + (-1)^k \frac{\beta\alpha^k}{k+1}, \quad k = 0, 1, 2, \cdots$$

which is a tail probability. Here the support of $\boldsymbol{Q}$ exceeds the unit interval. Choosing $\beta = \frac{2}{3}$ and $\alpha = \frac{1}{7}$, we have

$$\Delta^2 \bar{P}_1 = -0.00523$$

$\Rightarrow$ $\bar{P}_k$ is not Completely Monotone sequence $\Rightarrow \bar{P}_k$ is not SDFR, although the corresponding EMP shock model survival probability $\overline{S}$ in (5.2) non-discrete SDFR, with $\overline{P}_k$ as chosen above.

So a necessary and sufficient condition for (5.1) to be a p.g.f is;

Theorem 5.2.7. *The necessary a sufficient condition for the converse of to hold is if $\exists$ some non-negative random variable $Y \backsim \boldsymbol{Q}$ such that*

$$EY^k(1-Y)^{n-k} \geq 0 \quad \forall n,\ k \in \mathcal{Z}_+$$

CHAPTER 6

CONCLUSION

In this dissertation, some specific inference problems of Branching Process and Poisson shock model are discussed. Methodologies has been developed to deal with such situations.

6.1 Inference of Branching Processes and Future Work

This dissertation explores two very basic problems of statistical inference of Galton-Watson Process. In most of the literature of Galton-Watson process, different parametric assumptions has been made about the offspring distribution. But in this dissertation no parametric assumption is assumed for the offspring distribution. On that sense, here a more general setup has been considered.

In Chapter 2, most of the definitions and important results of branching process has been stated. Also all the notations are introduced in this chapter. Throughout the dissertation same notations has been used.

Chapter 3 deals with a specific problem of Galton-Watson process where statistical methodologies is developed to estimate the *'age'* of Galton-Watson process. In this section the first challenge is to find a good method of estimating the offspring distribution. The method existed are either for parametric families or not computationally convenient. An EM algorithm is developed for estimating the offspring distribution. The efficiency of the method in estimating the offspring distribution is discussed with two illustrated examples. Both cases show that the method is working pretty fine. A method of moment estimate has been proposed to estimate the generation from given r succesive generation size. This method is generalization of Stigler's (1970). Asymptotic properties of the estimator are proved. In next section we have derived a

algorithm for finding maximum likelihood of the generation. The method is inspired by the results of Adés et al. The computational aspect of the method is also discussed. A third method is developed my exploring the Markov structure of Galton-Watson peocess. The limit theorem regarding this method has been established. In this chapter two numerical illustrations has been described to show the efficiency of the proposed estimators in estimating *'age'*. The initial assumption of the process is the process is supercritical. Which implies that the process has a non-trivial probability of extinction. A method of estimating the probability of extinction is proposed at the end of this chapter. It is also proved that under the assumption of supercriticallity the estimator has a asymptotic normal distribution which can be used for testing purpose.

One fundamental challenge regarding Galton-Watson process is to identify the nature of the process. That means from a given data, is it possible to conclude that the process is 'subcritical'or 'supercritical'? A statistical testing procedure is required to draw such a conclusion. All older works regarding this context is based on the parametric structure of the offspring distribution. The main problem here is to estimate the random variable W. Except few parametric cases it is very difficult to find the distribution of W. A bootstrap technique has been developed in Chapter 4 to construct a asymptotic statistical testing procedure to test such type of hypothesis. Also a method based on least favorable set up is discuss in this chapter.

6.1.1 Future Work

There are several directions of future work:

- completing the test statistics related to least favorable null hypothesis.
- Exploring and extending the use of such methods for Markov Branching Processes, and more generally for Age-dependent Branching Processes.

- Another interesting future work is inference regarding the time to extinction of a Galton-Watson Process. Bhattacharjee (1987) has proved that T has a log-convex density. Is that result can be used to find a estimator for the mean time to extinction? Density estimation and constrained likelihood methods may be handy in this context.

- Bayesian estimation for Branching processes.

- Developing software fault count models using the ideas of Branching Processes.

6.2 Poisson Shock Model and Future Work

A closure property of *SDFR* under Poisson shock model is discussed in Chapter 5. It has been shown that the discrete *SDFR* property transmitted to continuous *SDFR* under Poisson shock model decomposition. An example is constructed to show that the converse is not true. A necessary and sufficient condition for converse to be true has been developed at the end of this chapter.

6.2.1 Future Work

- Constructing a statistical procedure to test whether a given data has a underlying distribution function which has a *SDFR* property.

- Applying such a method for modeling biological phenomenons like *Cori Cycle* activity in human body, which is believed to have a log-convex density.

APPENDIX A

AN OVERVIEW OF EM ALGORITHM

Maximum likelihood estimators (MLE) are very popular and useful in estimating the parameters of statistical models, since they have *good* asymptotic properties. But in real life, it is often virtually impossible to find the MLE of the parameters by direct maximization of the likelihood function due to their complex structure. In such situations, parameters are estimated using iterative methods. The EM algorithm is one of the most effective algorithms for local maximization since it iteratively transfers a complex function to a highly stable simple one. This algorithm also overcomes the drawbacks of the Newton's method and the Fisher scoring method. Newton's method requires calculation of complicated second derivatives and the Fisher scoring method involves calculation of the expected information matrix. For problems with large number of parameters, both algorithms involve large matrix inversions and this is computationally very intrusive. In this situation the EM algorithm is useful since it is based on an optimization transfer principle that replaces a complex optimization problem by a sequence of simple ones. This method is called the EM method because the alternating steps involve an expectation and a maximization.

This method was described and analyzed by Dempster, Laird, and Rubin (1977), although the method had been used much earlier, by Hartley (1958), for example. Many additional details and alternatives are discussed by McLachlan and Krishnan (1997).

The EM methods can be explained most easily in terms of a random sample that consists of two components, one observed, while the other part is unobserved or missing. The missing data can be missing observations on the same random variable that yields the observed sample, or the missing data can be from a different random

variable that is related somehow to the random variable observed. Though many common applications of EM methods do involve missing data problems, this is not necessary. Often, an EM method can be constructed based on an artificial "missing" random variable to supplement the observed data.

Description Consider the data $\boldsymbol{U} = (\boldsymbol{Y}, \boldsymbol{Z})$, where $\boldsymbol{Y}$ is the observed part and $\boldsymbol{Z}$ is the unobserved part of the data. Our objective is to estimate the parameter vector $\boldsymbol{\theta}$, which are involved in the distribution of both components of $\boldsymbol{U}$. The EM algorithm like all maximum likelihood algorithms, seeks to maximize the log-likelihood $\boldsymbol{L}(\boldsymbol{\theta})$ of the observed data with respect unknown parameters $\boldsymbol{\theta}$. If $f(\boldsymbol{U}|\boldsymbol{\theta})$ denotes the density function (likelihood) of the complete data, then the EM algorithm maximizes the surrogate function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = E[\ln f(\boldsymbol{U}|\boldsymbol{\theta})|\boldsymbol{Y}, \boldsymbol{\theta}^{(n)}]$$

with respect to $\boldsymbol{\theta}$. This optimization is done iteratively beginning with some initial values of $\boldsymbol{\theta}$ and then update it to maximize Q.

So the EM approach to maximizing $\ln f(\boldsymbol{Y}|\boldsymbol{\theta})$ has two alternating steps:

- **E** step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$.
- **M** step: Determine $\boldsymbol{\theta}$ iteratively to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$.

The EM method can be slow to converge, however, Wu (1983) has discussed about the convergence criteria of EM algorithm.

APPENDIX B

BASICS OF MARTINGALE

Basic definitions and results of martingales are discussed in this appendix.

B.1 Definitions

Definition B.1.1. Martingale and submartingale

Suppose $E|X_n| < \infty$ for all n. Then the process $\{X_n, \Im_s\}_{n\in I}$ is called

i) a *martingale* (mg) if

$$E(X_n|\Im_s) \;=\; X_s \;\; a.s. \text{ for each pair } \; s \leq n \; \text{ in } \; I.$$

ii) a *submartingale* (submg) if

$$E(X_n|\Im_s) \;\geq\; X_s \;\; a.s. \text{ for each pair } \; s \leq n \; \text{ in } \; I.$$

iii) a *supermartingale* if

$$E(X_n|\Im_s) \;\leq\; X_s \;\; a.s. \text{ for each pair } \; s \leq n \; \text{ in } \; I.$$

Definition B.1.2. Integrability

The process $\{X_n, \Im_n\}_{n\in I}$ is integrable if $\sup\{E|X_n| : n \in I < \infty$

If $\{X_n^2, \Im_n\}_{n\in I}$ is integrable, then $\{X_n, \Im_n\}_{n\in I}$ is called *square-integrable.*

B.2 Propositions

Proposition B.2.1. *Equivalence*

The process $\{X_n, \Im_n\}_{n\in \mathrm{I}}$ *is a submg if and only if for every pair* $s \leq n$, *we have*

$$\int_{\Im}(X_n - X_s)dP \;\; \geq 0 \;\; \textit{for all} \;\; \Im \in \Im_s. \tag{B.1}$$

Similarly, $\{X_n, \Im_n\}_{n\in \mathrm{I}}$ *is a mg if and only if equality holds in the above equation.*

Proposition B.2.2. *Let* $\phi : (R, \mathcal{B}) \to (R, \mathcal{B})$ *be the Borel real line, and suppose* $E|\phi(X_n| < \infty$ *for all* $n \in \mathrm{I}$.

(a) If ϕ *is convex and* $\{X_n, \Im_n\}_{n\in \mathrm{I}}$ *is a mg, then* $\{\phi(X_n), \Im_n\}_{n\in \mathrm{I}}$ *is a submg.*

(b) If ϕ *is convex and* $\nearrow$ *and* $\{X_n, \Im_n\}_{n\in \mathrm{I}}$ *is a submg, then* $\{\phi(X_n), \Im_n\}_{n\in \mathrm{I}}$ *is a submg.*

Proof. Clearly, $\phi(X_n)$ is adapted to $\Im_n$. Let $s \leq t$.

(a) First note that $E[\phi(X_n)|\Im_s] \geq \phi(E(X_n|\Im_s))$, by the conditional Jensen's inequality

Then, for for the mg case, note

(b) $E[\phi(X_n)|\Im_s] = \phi(X_s)$ a.s.

For the submg case,

(c) $E[\phi(X_n|\Im_s] \geq \phi(E(X_n|\Im_s)$, by the conditional Jenson's inequality.

(d) $\geq \phi(X_s)$ a.s.

since ϕ is $\nearrow$ and $E(X_n|\Im_s) \geq X_s$ a.s. $\square$

APPENDIX C

R CODES

The R codes used for simulation studies and modeling are given in this appendix.

C.1 R Code for Generating Samples From a Galton- Watson Process

```
offspring ← function(n,p)
z ← 0
d ← 0
z ← cumsum(p)
s ← length(p)
u1 ← sort(runif(n,0,1))
u ← 0
for(j in 1:s)
for(i in 1:n)
if(u1[i]← z[j])
d[i] ← i;
if(u1[i]>=z[j])
d[i] ← 0;

u[j]← length(d[d!= 0])

u2 ← diff(c(0,u))
h ← 0
h[1] ← 0
for(i in 2:s)
```

```
h[i] ← u2[i]*(i-1)

k ← sum(h)
k

galtonwatson ← function(nmgen,intsize,pvec)
popsize ← 0
popsize[1] ← intsize
for(k in 1:nmgen)
if(popsize[k]>0)
popsize[k+1]← offspring(popsize[k],pvec);

if(popsize[k]==0)
popsize[k+1]← 0;


popsize

tx ← 0
for(i in 1:400)
tx[i] ← galtonwatson(20,1,c(0.5,0.4,0.1))
```

C.2 Matlab Code for Estimating Parameters of Offspring Distribution by EM Algorithm

```
function f=profit(x)
y=[0.2,0.3,0.5];
sampl= [2;2;4;7;11;19;32;53;83];
N=sum(sampl);
k= length(sampl);
g = length(y);
factterm1 = 0;
firstterm=zeros(g,k);
secondterm1 =0;
secondterm = 0;
for d=1:g
for j=1:k
for t=1:sampl(j)
factterm1(t) = (-t + (t+0.5)*log(t))*binopdf(t,sampl(j),x(d));
end
firstterm(d,j) = sum(factterm1);
end secondterm1(d) = x(d)* log(x(d));
secondterm = sum(secondterm1); end f=-(sum(firstterm)+ N*secondterm);
function [cin,ceq]=confun(x)
cin=[-x(1);-x(2);-x(3);x(2)-1;x(3)-1;-(x(2)+2*x(3)-1)];
ceq=[x(1)+x(2)+x(3)-1];
clear all
x0=[0.1,5,19];
options=optimset('Largescale','off');
[x,fval]=fmincon('profit',x0,[],[],[],[],[],[],'confun')
```

REFERENCES

[1] Adés, M., J.P.Dion, G. Labelle, and K. Nanthi(1982). "Recurrence formula and the maximum likelihood estimatéion of the age in a simple branching process". *Journal of Applied Probability.* **19**:776-784.

[2] Attreya, K.B. and Ney, P.E.(1972).*Branching Processes* , Springer-Verlag, **New York**.

[3] Basawa, I.V., and D. J. Scott (1976). "Efficient tests for branching processes." *Biometrika*, **63**:531-536.

[4] Basawa, I.V. (1981). "Efficient conditional test for mixture experiments with applications to birth and branching processes." *Biometrika*, **68**:153-164.

[5] Basawa, I.V.(1987). "Asymptotic distributions for prediction errors and related tests of fit for nonstationary processes." *Annals of Statistics*, **15**:46-58.

[6] Bhattacharjee, M. C.(1987). The time to extinction of branching processes and log-convexity *Probability in Engineering and Informational Sciences,* **1**:265-278.

[7] Bhattacharjee, M. C.(2005). "Strong Version of the DFR Property." *CAMS Research Report 0405-18, 2005.*

[8] Cox, D. R. (1961). "Tests of separate families of hypotheses." *Proc. 4th Berkeley Symp. Math.Stat. Prob.*, **1**:105-123.

[9] Cox, D. R. (1962). "Further results on tests of separate families of hypotheses." *J. Roy. Statist.Soc.B*, **24**: 406-424.

[10] Crump, K. S., and R. B. Howe (1972): "Nonparametric estimation of the age of a Galton-Watson branching process." *Biometrika*, **59**:533-538.

[11] Dempster, A., Laird, N. & Rubin D. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, **39(1)**: 1-38.

[12] Dion, J.-P., G. Labelle, and A. Latour (1982). "Small sample results for maximum likelihood estimation in branching process." *Canadian Journal of Statistics*, **10**:271-276.

[13] Esary, J. D. and Marshall, A. W.(1973). "Shock Models and Wear Processes." *Ann. Proab*, **1(4)**:624-649.

[14] Feller, W.(1939). "Completely Monotone Functions and Sequences." *Duke Math Journal*, **5**:661-674.

[15] Feller, W.(1991).*An Introduction to Probability Theory and Its Application, Vol-II* , Wiley: **New York**.

[16] Guttorp, Peter (1991).*Statistical Inference for Branching Processes*, Wiley: **New York**.

[17] Gourieroux, C. and Monfort, A. (1982). "Likehihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." *Econometrica*, **50**:63-80.

[18] Hall, Peter; and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*, Academic Press: **New York**.

[19] Harris, T. E.(1963). *The Theory of Branching Process*, Dover Phonix: **New York**.

[20] Keiding, N.(1976). "Population Growth and Branching Process in Random Enviroment." *In Proc. 9th Int. Biometric Conf.*. :149-165, Boston

[21] Lehman, E. L.(1994).*Testing Statistical Hypothesis*, Chapman and Hall: **New York**.

[22] Pakes, A.G.(1975). "Non-parametric estimation in Galton-Watson process." *Mathematical Biosciences*; **26**:1-18.

[23] McLachlan, G. J. & Krishnan T. (1997). *The EM algorithm and extensions*, John Wiley and Sons, Inc: **New York**.

[24] Sen, P.K. and Singer, L.M. (1995). *Large Sample Methods in Statistics, An Introduction with Applications.* Chapman and Hall.: **New York**.

[25] Sen, P. K. and Sivapulle M. J.(2005). *Constained Statistical Inference, Inequality, Order and Shape Restriction*, Wiley.: **New York**.

[26] Stigler, S. M. (1970)."Estimating the age of a Galton-Watson process." *Biometrika*, **58**:499-508.

[27] Stigler, S. M. (1971). "The Estimation of the probability of extinction and other parameters associated with branching process." *Biometrika*, **58**:499-508.

[28] Sweeting, T. J. (1978). "On efficient tests for branching processes." *Biometrika*', **65**:123-128.

[29] Thompson, E. A. (1976)."Estimation of Age and rate of increase of rare variant". *American Journal of Human Genetics*, **28**:442-452.

[30] Wu, C. F. J. (1983). "On the convergence properties of the EM algorithm." *Annals of Statistics*, **11**: 95-103.