EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR BRANCHING PROCESSES WITH IMMIGRATION

by

Pin Ren

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

Committee:

____________________ Dr. Anand Vidyashankar, Dissertation Director

____________________ Dr. Guoqing Diao, Committee Member

____________________ Dr. Jacqueline Hughes-Oliver, Committee Member

____________________ Dr. Thomas Wanner, Committee Member

____________________ Dr. William Rosenberger, Department Chair

____________________ Dr. Kenneth S. Ball, Dean, The Volgenau School of Information Technology and Engineering

Date: ________________ Spring 2014
George Mason University
Fairfax, VA

Empirical Likelihood Confidence Regions for Branching Processes With Immigration

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

By

Pin Ren
Master of Science
Shandong University, 2009
Bachelor of Science
Shandong University, 2006

Director: Dr. Anand Vidyashankar, Professor
Department of Statistics

Spring 2014
George Mason University
Fairfax, VA

Acknowledgments

I would like to thank my dissertation advisor Dr. Vidyashankar for his help, time, and guidance throughout my graduate education. I would also like to thank the members of the committee Dr. Diao, Dr. Hughes-Oliver, and Dr. Wanner for their time and efforts. Finally, I would like to thank Dr. Rosenberger, Chair of the department, for his advice and help.

Table of Contents

List of Tables

List of Figures

Abstract

EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR BRANCHING PROCESSES WITH IMMIGRATION

Pin Ren, PhD

George Mason University, 2014

Dissertation Director: Dr. Anand Vidyashankar

Branching processes arise in multiple areas of scientific applications such as social science, cell biology, epidemic spread, and finance. This dissertation is concerned with joint inference for the mean and variance parameters of branching processes with immigration. Specifically, we describe a new empirical likelihood based methodology for constructing the simultaneous confidence region for the mean and variance parameters of the branching processes. We provide a theoretical justification of the proposed methods and an algorithm for computations. Simulation and data analysis results will also be presented.

Chapter 1: Introduction and Literature Review

1.1 General Description of Branching Process

A branching process (BP), in discrete time, is a stochastic process for describing the evolution of an evolving population across time. These processes have been widely applied to model various phenomena in biology, computer science, social science, and other scientific areas. A BP can be described as follows. The process starts at time 0 with a single ancestor that lives one unit of time and produces offspring according to the probability distribution $\{p_j; j \geq 0\}$. Each of these individuals in turn live one unit of time and reproduce according to the same probability distribution independent of each other and of their ancestors. The generation sizes can then be written as the collection $\{Z_n; n = 0, 1, 2, \dots\}$.

The process $\{Z_n : n \geq 0\}$ can be recursively defined using the representation

$$Z_n = \sum_{j=1}^{Z_{n-1}} \xi_{n,j},\ n = 1, 2, 3, \dots, \tag{1.1}$$

where $\xi_{n,j}$ can be interpreted as the number of children produced by the j^{th} parent in the n^{th} generation. The probability distribution of $\xi_{n,j}$ for all n and j is

$$P(\xi_{n,j} = k) = p_k,\ k \geq 0. \tag{1.2}$$

The sequence $\{p_k : k \geq 0\}$ is referred to as the offspring distribution. Let $m = E(Z_1|Z_0 = 1)$. It is well known that (see Athreya and Ney (1972)) when $m > 1$, Z_n converges to infinity with positive probability while if $m \leq 1$, Z_n becomes extinct with probability one (unless $p_1 = 1$). We will assume throughout the thesis that $\sigma^2 = Var(Z_1|Z_0 = 1) < \infty$. Unless stated otherwise, the condition $Z_0 = 1$ will be assumed throughout the thesis.

Using the definition of branching process, it can be seen that $E(Z_n) = m^n$ (see Athreya and Ney (1972)) and

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

A useful and a realistic modification of the BP is the incorporation of an immigration component into the population from an outside source. More precisely, the branching process with immigration (BPI) is defined as

$$X_n = \sum_{j=1}^{Z_{n-1}} \xi_{n,j} + I_n, \quad n = 1, 2, ..., \tag{1.4}$$

where $\xi_{n,j}$ are same as in (1.1) and I_n is the number of immigrants entering in the n^{th} generation from an outside source. The random variables $I_1, I_2, \ldots$ are assumed to be independent and identically distributed (*i.i.d.*) and are independent of the process $\{\xi_{n,j}; j \geq 1, n \geq 1\}$. We denote their distribution by

$$P(I_1 = j) = q_j, j \geq 0. \tag{1.5}$$

Let $E(I_1) = \lambda$, $Var(I_1) = b^2$. When $m > 1$, the generation sizes of a BPI diverge to infinity with probability one whereas when $m \leq 1$, X_n can be a transient or a recurrent Markov chain.

This dissertation is concerned with inference for (m, σ^2) using empirical likelihood methodology.

1.1.1 Inference for the Mean of a Branching Process with Immigration

The problem of inference for $m, \lambda, \sigma^2, b^2$ has been extensively studied in the literature (Athreya and Ney (1972), Heyde (1976), Bhat and Adke (1981), Nagaev (1981), Wei and Winnicki (1990), Winnicki (1991)). In the case $m > 1$ and $p_0 = 0$, and $q_0 = 1$ (i.e. there is no immigration), a natural estimator for m is $\hat{m}_n = Z_{n+1}/Z_n$, which is referred to as Nagaev estimator or ratio estimator. It can be seen to be unbiased and asymptotically normal (see Athreya and Ney (1972)). The ratio estimator only uses data from generations n and $n+1$. When data from all generations, namely 1 through n, are available, Harris (1948) (see also Jagers (1975)) developed the non-parametric maximum likelihood estimator for m; namely, $\tilde{m}_n = \sum_{i=1}^{n} Z_i / \sum_{i=0}^{n-1} Z_i$. It is well-known that $\tilde{m}_n$ is consistent and asymptotically normally distributed with mean 0 and variance σ^2 (see Bhat and Adke (1981)); that is

$$\left(\sum_{i=1}^{n} Z_i \right)^{1/2} (\tilde{m}_n - m) \xrightarrow{d} N(0, \sigma^2). \tag{1.6}$$

Turning to the process with immigration, when $m < 1$, Heyde and Seneta (1972) (see also Heyde and Seneta (1974)) studied the asymptotic properties of the least squares estimators of m. Quine (1976) investigated the alternative estimators for offspring mean and immigration mean; his estimators are given by

$$\hat{m} = \frac{\sum\limits_{i=1}^{n} X_{i-1}(X_i - n^{-1}S_n)}{\sum\limits_{i=1}^{n} (X_i - n^{-1}S_n)^2}, \quad \text{and} \quad \hat{\lambda} = \left(\frac{S_n}{2n} \right) \left(\frac{\sum\limits_{i=1}^{n} (X_i - X_{i-1})^2}{\sum\limits_{i=1}^{n} (X_i - n^{-1})S_n} \right), \tag{1.7}$$

where S_n denotes $\sum\limits_{i=1}^{n} X_{i-1}$. Additionally, Quine (1976) established the asymptotic properties of the estimators of m and λ under weaker moment conditions. Venkataraman and Nanthi (1982) also investigated the properties of these estimators. Klimko and Nelson

(1978) were the first to obtain the conditional least squares estimators for general stationary stochastic processes. That is, they studied the estimation problem by minimizing $\sum_{i=1}^{n}\Big(X_i - E(X_i|\mathcal{F}_{i-1})\Big)^2$. As an application, they investigated the subcritical BPI. In this case, the estimators are

$$\hat{m}_n = \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_{i-1} - n\sum_{i=1}^{n} X_{i-1}X_i}{\left(\sum_{i=1}^{n} X_{i-1}\right)^2 - n\sum_{i=1}^{n} X_{i-1}^2}, \quad \hat{\lambda}_n = \frac{\sum_{i=1}^{n} X_{i-1}X_i \sum_{i=1}^{n} X_{i-1} - \sum_{i=1}^{n} X_{i-1}^2 \sum_{i=1}^{n} X_i}{\left(\sum_{i=1}^{n} X_{i-1}\right)^2 - n\sum_{i=1}^{n} X_{i-1}^2}. \tag{1.8}$$

Due to the heteroscedasticity of the stochastic "error" term $X_i - mX_{i-1} - \lambda$, Wei and Winnicki (1990) suggested using the conditional weighted least squares method for estimating m and λ. The objective function in this case turns out to be

$$\sum_{i=1}^{n}\left(\frac{X_i - E(X_i|\mathcal{F}_{i-1})}{\sqrt{Var(X_i|\mathcal{F}_{i-1})}}\right)^2 \tag{1.9}$$

which upon minimization yields

$$\tilde{m}_n = \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n}(1+X_{i-1})^{-1} - n\sum_{i=1}^{n} X_i(1+X_{i-1})^{-1}}{\sum_{i=1}^{n}(1+X_{i-1})\sum_{i=1}^{n}(1+X_{i-1})^{-1} - n^2} \tag{1.10}$$

and

$$\tilde{\lambda}_n = \frac{\sum_{i=1}^{n} X_{i-1} \sum_{i=1}^{n}(1+X_{i-1})^{-1}X_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_{i-1}(1+X_{i-1})^{-1}}{\sum_{i=1}^{n}(1+X_{i-1})\sum_{i=1}^{n}(1+X_{i-1})^{-1} - n^2}. \tag{1.11}$$

They derived the joint asymptotic normality of $\tilde{m}_n$ and $\tilde{\lambda}_n$ when $m < 1$, and asymptotic normality of $\tilde{m}_n$ when $m > 1$. And when $m = 1$ the asymptotic distribution of $\tilde{m}_n$ was shown to converge to a function of limiting diffusion. They also established that $\tilde{m}_n$ is consistent for all $m > 0$, while $\tilde{\lambda}_n$ is consistent when $m \leq 1$, and its asymptotic distribution is only known when $m < 1$ or $m = 1$ and $2\lambda > \sigma^2$.

1.1.2 Estimation of the Variance of a Branching Process with Immigration

Another analogous problem concerns estimation of σ^2 and b^2. In the case $m > 1$, Heyde (1976) proposed a strongly consistent and asymptotic normal estimator

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(Z_i - \hat{m}_n Z_{i-1})^2}{X_{i-1}}, \tag{1.12}$$

where $\hat{m}_n$ is the Nagaev estimator. In the subcritical case ($m < 1$) Yanev and Tchoukova-Dantcheva (1986) stated the strong consistency and asymptotic normality of $\hat{\sigma}_n^2$ and $\hat{b}_n^2$ without proof, where

$$\hat{\sigma}_n^2 = \frac{n \sum_{i=1}^{n} X_i \hat{U}_i^{\,2} - \sum_{i=1}^{n} X_{i-1} \hat{U}_i^{\,2}}{n \sum_{i=1}^{n} X_{i-1}^2 - \left(\sum_{i=1}^{n} X_{i-1} \right)^2}, \tag{1.13}$$

and

$$\hat{b}_n^2 = \frac{\sum_{i=1}^{n} \hat{U}_i^{\,2} \sum_{i=1}^{n} X_{i-1}^2 - \sum_{i=1}^{n} X_{i-1} \hat{U}_i^{\,2} \sum_{i=1}^{n} X_{i-1}}{n \sum_{i=1}^{n} X_{i-1}^2 - \left(\sum_{i=1}^{n} X_{i-1} \right)^2}, \tag{1.14}$$

$U_i = X_i - mX_{i-1} - \lambda$, and $\hat{U}_i = X_i - \hat{m}X_{i-1} - \lambda$. By observing that the minimization of

$$\sum_{i=1}^{n} \left(U_i^2 - E(U_i^2|\mathcal{F}_{i-1})\right)^2 \tag{1.15}$$

leads to conditional least squares estimators $\hat{\sigma}_n^2$ and $\hat{b}_n^2$, Winnicki (1991) studied the asymptotic properties of conditional least squares estimators and established their joint asymptotic normality in the limit when $m \leq 1$. Additionally, by considering the regression equation

$$\frac{U_i^2}{1 + X_{i-1}} = \sigma^2 + (b^2 - \sigma^2)\frac{1}{1 + X_{i-1}} + \frac{V_i}{1 + X_{i-1}},$$

and using conditional weighted least squares method the estimators of σ^2 and b^2 in this case become

$$\tilde{\sigma}_n^2 = \frac{\sum_{i=1}^{n}(1 + X_{i-1})^{-2} \sum_{i=1}^{n} U_i^2(1 + X_{i-1})^{-1} - \sum_{i=1}^{n}(1 + X_{i-1})^{-1} \sum_{i=1}^{n} U_i^2(1 + X_{i-1})^{-2}}{n \sum_{i=1}^{n}(1 + X_{i-1})^{-2} - (\sum_{i=1}^{n}(1 + X_{i-1})^{-1})^2},$$

and

$$\tilde{b}_n^2 = \frac{\sum_{i=1}^{n} X_{i-1}(1 + X_{i-1})^{-1} \sum_{i=1}^{n} U_i^2(1 + X_{i-1})^{-2} - \sum_{i=1}^{n} X_{i-1}(1 + X_{i-1})^{-2} \sum_{i=1}^{n} U_i^2(1 + X_{i-1})^{-1}}{n \sum_{i=1}^{n}(1 + X_{i-1})^{-2} - (\sum_{i=1}^{n}(1 + X_{i-1})^{-1})^2}$$

respectively. We notice here that the definition of U_i and V_i involve m and λ. Winnicki (1991) further pointed out that if we replace m and λ by their estimates, the resulting estimator of σ^2 is strongly consistent and asymptotically normal for all $m \neq 1$. Also $\tilde{b}_n^2$ is consistent when $m < 1$ or $m = 1$ and $2\lambda \leq \sigma^2$.

1.1.3 Joint Estimation of the Mean and Variance of a Branching Process with Immigration

Basawa and Vidyashankar (2003) were the first to develop quasi-likelihood estimator for all parameters $\theta = (m, \sigma^2, \lambda, b^2)$ by solving quasi-likelihood estimating equations. They considered the approximate quasi-likelihood score function, defined by, $\tilde{S}_n(\theta) = \sum_{i=1}^{n} D_i(\theta)\tilde{V_i}^{-1}(\theta)U_i(\theta)$ to yield estimating equations for branching process, where $D_i(\theta) = E\left[\partial U_i(\theta)/\partial\theta|\mathcal{F}_{i-1}\right]$, and

$$\tilde{V}_i(\theta) = \begin{pmatrix} \sigma^2 X_{i-1} + b^2 & 0 \\ 0 & 2(\sigma^2 X_{i-1} + b^2) \end{pmatrix},$$

and $U_i(\theta) = \Big(X_i - mX_{i-1} - \lambda, \;\; (X_i - mX_{i-1} - \lambda)^2 - \sigma^2 X_{i-1} + b^2\Big)$. They established the jointly asymptotic normality of the quasi-likelihood estimators $\hat{m}_{QL}$ and $\hat{\sigma}^2_{QL}$ when $m \neq 1$, and the limiting independence of $\hat{m}_{QL}$ and $\hat{\sigma}^2_{QL}$ when $m > 1$. They also demonstrated that the quasi-likelihood estimators includes the weighted conditional least squares estimators studied by Wei and Winnicki (1990) , as well as those estimators studied by Heyde and Lin (1992).

As explained previously, due to the trichotomy of BPI, the estimates of the means and the variances and their asymptotic distributions vary dramatically depending on the value m. Motivated by the properties of the joint quasi-likelihood estimators stated by Basawa and Vidyashankar (2003), in this dissertation, we will investigate the problem of estimating (m, σ^2) jointly by the empirical likelihood method. We will also study the asymptotic distribution of empirical likelihood ratio statistic for testing $H_0 : m = m_0, \sigma^2 = \sigma_0^2$.

Empirical likelihood (EL) is a non-parametric inference approach based on the data-determined likelihood ratio function. That the EL inference dose not need a specification of a parametric family of distributions for the data allows flexibility in the data analysis. EL incorporates the advantages of the likelihood methods and the robustness of non-parametric approaches. The estimates obtained and the tests constructed by EL methods have good asymptotic and small sample properties. We now describe several fundamental results concerning the empirical likelihood.

1.2 Empirical Likelihood for the Mean of *i.i.d.* Data

The EL method is a non-parametric inference approach based on the data-determined likelihood ratio function. That the EL inference does not need a specification of a parametric family of distributions for the data allows flexibility in the data analysis. EL incorporates the advantages of the likelihood methods and the robustness of non-parametric approaches. The estimates obtained and the tests constructed by EL methods have good asymptotic and small sample properties (see Owen (1988), Owen (1990), and Owen (2001).

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (*i.i.d.*) random variables with cumulative distributions function (CDF) $F(\cdot)$. Let $E(X_1) = \mu$ and consider the problem of constructing the confidence interval (CI) for μ, without making specific distributional assumption on $F(\cdot)$.

It is well-known that the empirical cumulative distribution function (ECDF) $F_n(\cdot)$, where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$$

is a non-parametric MLE of $F(\cdot)$ (see Owen (2001)). Following Owen (1988) and Owen (1990), the empirical likelihood ratio (ELR) is defined to be

$$\mathcal{R}_n(\boldsymbol{w}) = \frac{\prod_{i=1}^{n} w_i}{\prod_{i=1}^{n} \frac{1}{n}} = \prod_{i=1}^{n} n w_i. \tag{1.16}$$

where (i) $w_i \geq 0$, (ii) $\sum_{i=1}^{n} w_i = 1$ and $\boldsymbol{w} = (w_1, w_2, \cdots w_n)$. The weights $\{w_i, 1 \leq i \leq$

$n\}$ can be viewed as probabilities on the data. Thus, conditioned on the sample $\mathcal{X}_n = (X_1, X_2, \ldots, X_n)$,the likelihood is given by $\prod_{i=1}^n w_i$. Thus, $\mathcal{R}_n(\boldsymbol{w})$ can be viewed as an analogue of the likelihood ratio and hence can be used to construct the confidence interval (CI) for μ. The following theorem, due to Owen (1988), provides an asymptotic justification for this method.

Theorem 1. *Assume that* $E|X_1|^3 \leq \infty$, *and that the random variables are non-degenerate. Define*

$$\mathcal{F}_{c,n} = \{F | R(F) \geq c, F << F_n\},$$

and let

$$X_{U,n} = \sup_{F \in \mathcal{F}_{C,n}} \int x dF = \sup_{\boldsymbol{w}} \sum_{i=1}^{n} w_i X_i$$

and

$$X_{L,n} = \inf_{F \in \mathcal{F}_{C,n}} \int x dF = \inf_{\boldsymbol{w}} \sum_{i=1}^{n} w_i X_i.$$

Then

$$P(X_{L,n} \leq \mu \leq X_{U,n}) \to P(\chi_1^2 \leq -2 \ log \ c).$$

Remark 1: The above idea can be converted to develop a non-parametric likelihood ratio test for μ. To this end, let $H_0 : \mu = \mu_0$. Then the ELR is given by

$$\mathcal{R}_n(\mu) = \prod_{i=1}^{n} n w_i,$$

where w_is are such that (i) $w_i \geq 0$, (ii) $\sum_{i=1}^{n} w_i = 1$, and (iii) $\sum_{i=1}^{n} w_i X_i = \mu$. Then under H_0, $-2 \log \mathcal{R}_n(\mu_0) \xrightarrow{d} \chi_1^2$ where χ_1^2 is a chi-square distribution with one degree of freedom, and we reject H_0 if $-2 \log \mathcal{R}_n(\mu_0) \geq c$ where c is chosen from the quantiles of the χ_1^2 distribution, with appropriate conditions on the size.

1.3 Empirical Likelihood for Functionally Independent Estimating Equations

The EL framework was extended by Qin and Lawless (1994a) and Qin and Lawless (1994b) to include estimating functions. Let $X_1, X_2, \ldots, X_n$ with unknown distribution function $F_0(\cdot)$, Let θ be a r-dimensional parameter. Let $g_1(\boldsymbol{X}, \theta), g_2(\boldsymbol{X}, \theta) \ldots, g_r(\boldsymbol{X}, \theta)$ be the estimating functions such that if $\boldsymbol{g}(\boldsymbol{X}, \theta) = \big(g_1(\boldsymbol{X}, \theta), \ldots, g_r(\boldsymbol{X}, \theta)\big)'$ then $E(\boldsymbol{g}(\boldsymbol{X}, \theta)) = 0$.

As an example, suppose $\theta = (\mu, \sigma^2)$ where $\sigma^2 = var(X_1)$. Then the estimating functions are:

$$g_1(X; \theta) = X - \mu$$

$$g_2(X; \theta) = (X - \mu)^2 - \sigma^2$$

and $\boldsymbol{g}(X, \theta) = \big(g_1(X, \theta), g_2(X, \theta)\big)'$, and $E(\boldsymbol{g}(X, \theta)) = (0, 0)'$.

This gives the following EL formalism for testing

$$H_0: \ \theta = \theta_0.$$

Define

$$\mathcal{R}_n(\theta) = \prod_{i=1}^{n} n w_i,$$

where $w_i's$ are such that $w_i > 0$, $\sum_{i=1}^{n} w_i = 1$ and $\sum_{i=1}^{n} w_i \boldsymbol{g}(x_i; \theta) = 0$. Qin and Lawless (1994a) show that, under H_0,

$$-2log\mathcal{R}_n(\boldsymbol{w}) \xrightarrow{d} \chi^2_2.$$

1.4 Computation of the Empirical Likelihood

Owen (1988) discussed a contouring algorithm for a single functional of *i.i.d.* data for computing the EL. He derived an expression for the weights that give rise to the constrained confidence interval for the mean. Recalling that the objective function is given by

$$G = \sum_{i=1}^{n} w_i X_i + \lambda_1 \left(1 - \sum_{i=1}^{n} w_i\right) + \lambda_2 \left(logc - \sum_{i=1}^{n} log(nw_i)\right), \tag{1.17}$$

one can, by differentiating it with respect to w_i and setting it to 0, show that $w_i = \frac{\lambda_2}{1-\lambda_1}$, where λ_2 is a normalized constant and λ_1 can be found by either a root-finding algorithm or by interpolation. For *i.i.d.* random vectors $X_1, X_2, \cdots, X_n$ in $\mathbb{R}^p$, Owen (1990) extended the above method to computing the confidence regions for $E(X)$ by reformulating the problem of maximizing

$$\mathcal{R}(\mu) = \prod \left(1 + \lambda'(X_i - \mu)\right)^{-1}$$

subject to constrains $\int X dF = \mu$, as one of minimizing a convex function

$$f(\lambda) = -\sum log\left(1 + \lambda'(X_i - \mu)\right).$$

If μ is in the convex hull of $X_1, \cdots, X_n$, then the weights w_i can be expressed as $w_i = \dfrac{1}{n}\dfrac{1}{1 + \lambda'(X_i - \mu)}$ (see Owen (1990)), where the Lagrangian multiplier λ is a unique solution of

$$g(\lambda) = \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda'(X_i - \mu)} = 0.$$

By noticing that $-g(\lambda)$ is the gradient with respect to λ of $f(\lambda)$, and the Hessian matrix of f can be calculated as

$$H(\lambda) = \sum_{i=1}^{n} \frac{(X_i - \mu)(X_i - \mu)'}{(1 + \lambda'(X_i - \mu))^2},$$

one can use Newton Raphson iteration to solve for λ when it satisfies $\{1+\lambda'(X_i - \mu) \geq 1/n\}$ for $i \leq i \leq n$.

Hall and La Scala (1990) proposed a more general method to construct empirical likelihood confidence regions and provided several specific examples, including regions for bivariate means, covariance matrix and correlation coefficients based on a multivariate Newton algorithm. Their algorithm can be described as follows: let $\boldsymbol{x} = (x_1, \cdots, x_n)$ be a n-vector and in order to solve the equations of $f_i(x) = 0$, for $1 \leq i \leq m$, set $\boldsymbol{f} = (f_1, \cdots, f_n)$. Given an initial value of $\boldsymbol{x}^{(0)}$ as a solution of $\boldsymbol{f}(\boldsymbol{x}) = 0$, the j^{th} iteration may be written as

$$\boldsymbol{x}^{(j)} = \boldsymbol{x}^{(j-1)} - \left(J\left(\boldsymbol{x}^{(j-1)}\right)\right)^{-1} \boldsymbol{f}(\boldsymbol{x}^{(j-1)}),$$

where $J^{ij} = \frac{\partial f_i}{\partial x^{(j)}}$, and $J = J^{ij}$. If the initial solution is close enough to the true value $\boldsymbol{x}^{(0)}$, and J is continuous and nonsingular at $\boldsymbol{x}^{(0)}$, then $\boldsymbol{x}^{(j)} \to \boldsymbol{x}^{(0)}$ as $j \to \infty$ (see Boyd and Vandenberghe (2004)). When it comes to the implementation of the algorithm for the univariate mean, the confidence interval endpoints could be found as two solutions (λ_1, μ_1) and (λ_2, μ_2) of the equations

$$f_1(\lambda, \mu) = \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda(X_i - \mu)}, \quad f_2(\lambda, \mu) = \sum_{i=1}^{n} log\left(1 + \lambda(X_i - \mu)\right) - \frac{1}{2}c,$$

where c is the critical value from a χ^2 distribution. $\hat{\mu}_L$ and $\hat{\mu}_U$ are taken to be the smallest and largest value of μ_1 and μ_2 respectively.

For the bivariate case $X = (Y, Z)'$, the confidence region for the mean of $(\mu_Y, \mu_Z) = (EY, EZ)$ can be constructed by solving the following three equations for $\lambda_Y, \lambda_Z, \mu_Y, \mu_Z$; namely

$$\sum r_i(Y_i - \mu_Y) = 0, \quad \sum r_i(Z_i - \mu_Z) = 0, \text{ and } \sum log\left(r_i^{-1}\right) = \frac{1}{2}c \tag{1.18}$$

where $r_i = r_i(\lambda_Y, \lambda_Z, \mu_Y, \mu_Z) = (1 + \lambda_Y(Y_i - \mu_Y) + \lambda_Z(Z_i - \mu_Z))^{-1}$, and λ_Y, λ_Z are Lagrangian multipliers. Since the above three equations in (1.18) imply that $\sum r_i = n$, we can parameterize r_i as $r_i(u)$ for $u \in U = (0, 2\pi)$, and take $\mu_Y(u) = n^{-1}\sum r_i(u)Y_i$, $\mu_Z(u) = n^{-1}\sum r_i(u)Z_i$, $\lambda_Y = \lambda sin(u)$ and $\lambda_Z = \lambda sin(u)$.

Owen (2013) improved the computational efficiency for the empirical likelihood by using a damped Newton iteration to convex and self-concordance function. If a convex function $f(x)$ for $x \in R$ has three derivatives and $\|f'''(x)\| \leq 2f''(x)^{3/2}$, then $f(x)$ is said to be self-concordant. Since self-concordance can control the rate of changes of the second derivative of a function, it ensures the damped Newton method converges to the global solution. Instead of an unguarded Newton's method for minimizing $f(\boldsymbol{x})$ by updating $\boldsymbol{x} \leftrightarrow \boldsymbol{x} + \Delta(\boldsymbol{x})$ where $\Delta(\boldsymbol{x}) = -\left(\nabla^2 f(\boldsymbol{x})\right)^{-1} \nabla f(\boldsymbol{x})$ and $\nabla f(\boldsymbol{x})$ and $\nabla^2 f(\boldsymbol{x})$ are the gradient and Hessian matrix of f respectively, the back tracking Newton algorithm (see Boyd and Vandenberghe (2004)) is introduced to modify the update $\Delta(\boldsymbol{x})$ by a "short" vector to allow f to decrease more. The backtracking algorithm begins with $t = 1$ and update t by βt until $f(\boldsymbol{x} + t\Delta(x)) \leq f(\boldsymbol{x}) + \alpha t \nabla f(\boldsymbol{x})'\Delta(x)$ is satisfied, where $\alpha \in (0, 0.05)$ and $\beta = (0, 1)$. Newton algorithm with back-tracking combined with the self concordance property can provide a usable stopping criterion and optimize a function within a bounded number of Newton steps. For the joint empirical likelihood confidence region for BPI data, we modify the backtracking algorithm for mean and variance respectively to fit the complex properties of BPI. These are discussed in detail in Chapter 4.

1.5 Structure of the Dissertation

This dissertation is concerned with joint inference for the mean and variance parameters of branching processes with immigration. Specially, we describe a new empirical likelihood based methodology for constructing the simultaneous confidence region for the mean and variance parameters of the branching processes. We provide a theoretical justification of the proposed methods and an algorithm for computations. Simulation and data analysis results will also be presented.

The rest of the dissertation is structured as follows: Chapter 2 is concerned with asymptotic justification of the EL confidence regions for (m, σ^2) for $m \neq 1$ while Chapter 3 contains the asymptotic justification for the case $m = 1$. Chapter 4 deals with computational and implementation issues and data analysis.

Chapter 2: Empirical Likelihood Confidence Region for Non-Critical BPI

2.1 Introduction

We recall that branching process with immigration is defined by the iteration

$$X_n = \sum_{j=1}^{X_{n-1}} \xi_{n,j} + I_n, \quad n = 1, 2, ...,$$

where $E(\xi_{n,j}) = m$, $Var(\xi_{n,j}) = \sigma^2$, $E(I_n) = \lambda$, $Var(I_n) = b^2$. Let $\mathcal{F}_i$ denote the σ-field generate by $\{X_0, X_1, \ldots, X_{i-1}\}$; that is $\mathcal{F}_i = \sigma \langle X_0, X_1, \ldots, X_{i-1} \rangle$. Let $U_i = X_i - mX_{i-1} - \lambda$ and $V_i = U_i^2 - \sigma^2 X_{i-1} - b^2$; since $E(X_i|\mathcal{F}_{i-1}) = mX_{i-1} + \lambda$, and $Var(X_i|\mathcal{F}_{i-1}) = \sigma^2 X_{i-1} + b^2$, $\sum_{i=1}^{n} U_i$ and $\sum_{i=1}^{n} V_i$ are martingales with respect to $\mathcal{F}_n$. Hence, it follows that U_i and V_i are martingale differences for each $1 \le i \le n$. Our primary interest is in constructing the joint empirical likelihood confidence regions for (m, σ). To this end, we adopt the empirical likelihood approach of Owen (1988), Owen (1990) and Qin and Lawless (1994b) for the branching process problems. Studying the asymptotic properties of the resulting inferential procedure is challenging particularly since the behavior of the estimating equations is different for different values of m, σ^2.

The empirical likelihood for the supercritical case can be constructed using the following estimating functions:

$$\sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) = \left(\sum_{i=1}^{n} U_i, \sum_{i=1}^{n} V_i \right)' = \left(\sum_{i=1}^{n} (X_i - mX_{i-1} - \lambda), \sum_{i=1}^{n} (U_i^2 - \sigma^2 X_{i-1} - b^2) \right)'. \quad (2.1)$$

Hence, using (1.16) from Chapter 1, the empirical likelihood ratio is given by

$$\mathcal{R}_n(\theta) = \max \prod_{i=1}^{n} n w_i, \tag{2.2}$$

where

$$w_i \geq 0, \ \sum_{i=1}^{n} w_i = 1, \ \text{and} \ \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0.$$

We now state a Proposition that provides a useful representation formula for the ELR. Before that we need more notation. Define $N_n = \sum_{i=1}^{n} X_{i-1}$, $\bar{U}_n = \frac{1}{n}\sum_{i=1}^{n} U_i$, $\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i$, $\bar{U}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} U_i^2$, and $\bar{V}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} V_i^2$. Set

$$S_n = \frac{1}{N_n}\begin{pmatrix} \sum_{i=1}^{n} U_i^2 & \sum_{i=1}^{n} U_i V_i \\ \sum_{i=1}^{n} U_i V_i & \sum_{i=1}^{n} V_i^2 \end{pmatrix},$$

then

$$S_n^{-1} = \frac{1}{det(S_n)}\frac{1}{N_n}\begin{pmatrix} \sum_{i=1}^{n} V_i^2 & -\sum_{i=1}^{n} U_i V_i \\ -\sum_{i=1}^{n} U_i V_i & \sum_{i=1}^{n} U_i^2 \end{pmatrix},$$

and

$$det(S_n) = \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n^{(2)}\right)\left(\bar{V}_n^{(2)}\right) - \left(\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i\right)^2.$$

Proposition 1. *For a BPI, the following representation formula holds:*

$$-2log\mathcal{R}_n(\theta) = A_n - B_n + R_n, \tag{2.3}$$

where

$$\begin{aligned} A_n &= N_n \left(\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right)' S_n^{-1} \left(\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right) \\ &= N_n \left(\frac{n}{N_n} \right)^2 \left(\bar{U}_n, \bar{V}_n \right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n \right)' \end{aligned} \tag{2.4}$$

$$B_n = N_n \left(\beta_n^* \right)' S_n^{-1} \beta_n^*, \quad and \quad R_n = \frac{2}{3} \sum_{i=1}^{n} r_i^*, \tag{2.5}$$

where

$$\beta_n^* = \frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \left(\frac{r_i^2}{1 + r_i} \right)', \tag{2.6}$$

$r_i = \boldsymbol{\nu_2'} \boldsymbol{g}(X_i, \theta)$ *where* $\boldsymbol{\nu_2}$ *is Lagrangian multiplier, and* r^* *is the reminder term in the Taylor expansion, and satisfies* $P(|r_i|^* \leq C|r_i|^3, 1 \leq i \leq n) \to 1$ *as* $n \to \infty$ *for any positive and finite constant* C*.*

Proof. Taking the logarithm and using Lagrangian multipliers, the (2.2) can be reduced to maximizing the objective function

$$G = \sum_{i=1}^{n} log(nw_i) - \nu_1 \Big(\sum_{i=1}^{n} w_i - 1 \Big) - n\boldsymbol{\nu_2}' \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta)$$

where ν_1 and $\boldsymbol{\nu_2}$ are Lagrangian multipliers, and $\boldsymbol{\nu_2} = (\nu_{21}, \nu_{22})'$. Setting the derivative of

G with respective to w_i to zero, we get

$$\frac{\partial G}{\partial w_i} = \frac{1}{w_i} - \nu_1 - n\boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta) = 0$$

which implies

$$\frac{1}{w_i} = \nu_1 + n\boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta).$$

Also, since $\sum_{i=1}^{n} \frac{\partial G}{\partial w_i} w_i = 0$, we have that $\sum_{i=1}^{n} \left(\frac{1}{w_i} - \nu_1 - n\boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta) \right) w_i = n - \nu_1$, from which we get $n = \nu_1$, and for $1 \leq i \leq n$,

$$w_i = \frac{1}{\nu_1 + n\boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)} = \frac{1}{n + n\boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)} = \frac{1}{n} \frac{1}{1 + \boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)}.$$

Also by considering the derivatives of G with respect to the Lagrangian, we can show that ν_2' also satisfies $\sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0$, which implies

$$\sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = \sum_{i=1}^{n} \frac{1}{n} \frac{1}{1 + \boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)} \boldsymbol{g}(X_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\boldsymbol{g}(X_i, \theta)}{1 + \boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)}.$$

Now, write

$$g(\boldsymbol{\nu_2'}) = \frac{1}{N_n} \sum_{i=1}^{n} \frac{\boldsymbol{g}(X_i, \theta)}{1 + \boldsymbol{\nu_2'}\boldsymbol{g}(X_i, \theta)} = 0. \tag{2.7}$$

Also, let $r_i = \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)$. Using the fact that

$$\frac{1}{1 + r_i} = \frac{1 - r_i^2 + r_i^2}{1 + r_i} = \frac{(1 - r_i)(1 + r_i)}{1 + r_i} + \frac{r_i^2}{1 + r_i} = 1 - r_i + \frac{r_i^2}{1 + r_i},$$

we can write (2.7) as

$$
\begin{aligned}
g(\boldsymbol{\nu_2'}) &= \frac{1}{N_n}\sum_{i=1}^{n}\frac{\boldsymbol{g}(X_i,\theta)}{1+\boldsymbol{\nu_2'}\boldsymbol{g}(X_i,\theta)} = \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\frac{1}{1+r_i} \\
&= \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\big(1-r_i+\frac{r_i^2}{1+r_i}\big)' \\
&= \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)-\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)r_i'+\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\big(\frac{r_i^2}{1+r_i}\big)'.
\end{aligned}
$$

Furthermore, the second term on the RHS of the above equation can be simplified to be

$$
\begin{aligned}
\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)r_i' &= \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\left(\nu_2'\boldsymbol{g}(X_i,\theta)'\right) \\
&= \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\boldsymbol{g}'(X_i,\theta)\boldsymbol{\nu_2} = S_n\boldsymbol{\nu_2}.
\end{aligned}
$$

Now turning to (2.7), and using the definition of β_n^* in (2.6), we get

$$
g(\boldsymbol{\nu_2'}) = \frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)+S_n\boldsymbol{\nu_2'}+\beta_n^*.
$$

which implies that

$$
\boldsymbol{\nu_2} = S_n^{-1}\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)+S_n^{-1}\beta_n^* = \boldsymbol{\nu_2}^{(0)}+\beta_n^{(0)}, \tag{2.8}
$$

where we denote by $\boldsymbol{\nu_2}^{(0)} = S_n^{-1}\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)$, and $\beta_n^{(0)} = S_n^{-1}\beta_n^*$.

Returning to the log-likelihood ratio, it can be seen that,

$$
\begin{aligned}
-2\log\mathcal{R}_n(\theta) &= -2\log\prod_{i=1}^{n} nw_i = -2\log\prod_{i=1}^{n}\frac{1}{1+\boldsymbol{\nu_2'}\boldsymbol{g}(X_i,\theta)} \\
&= 2\sum_{i=1}^{n} r_i - \sum_{i=1}^{n} r_i^2 + \frac{2}{3}\sum_{i=1}^{n} r_i^*. \qquad (2.9)
\end{aligned}
$$

The first term can be simplified to be

$$
\begin{aligned}
\sum_{i=1}^{n} r_i &= \sum_{i=1}^{n} \nu_2'\boldsymbol{g}(X_i,\theta) \\
&= \sum_{i=1}^{n}\left(\boldsymbol{\nu_2}^{(0)} + \beta_n^{(0)}\right)'\boldsymbol{g}(X_i,\theta) \quad \text{(using 2.8)}, \\
&= \left(\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\right)' S_n^{-1}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta) + \beta_n^* S_n^{-1}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta) \quad \text{(using 2.6)}.
\end{aligned}
$$

Now the second term in (2.8) can be written as

$$
\begin{aligned}
&\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n}\nu_2'\boldsymbol{g}(X_i,\theta)\boldsymbol{g}(X_i,\theta)'\boldsymbol{\nu_2} \\
&\quad = N_n\boldsymbol{\nu_2}'^{(0)}S_n\boldsymbol{\nu_2}^{(0)} + N_n\boldsymbol{\nu_2}'^{(0)}S_n\beta_n^{(0)} + N_n\beta_n'^{(0)}S_n\boldsymbol{\nu_2}^{(0)} + N_n\beta_n'^{(0)}S_n\beta_n^{(0)},
\end{aligned}
$$

where

$$
\begin{aligned}
N_n\boldsymbol{\nu_2}'^{(0)}S_n\boldsymbol{\nu_2}^{(0)} &= N_n\left(S_n^{-1}\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\right)' S_n^{-1}\left(S_n^{-1}\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\right) \\
&= N_n\left(\frac{1}{N_n}\sum_{i=1}^{n}U_i, \frac{1}{N_n}\sum_{i=1}^{n}V_i\right) S_n^{-1}\left(\frac{1}{N_n}\sum_{i=1}^{n}U_i, \frac{1}{N_n}\sum_{i=1}^{n}V_i\right)'
\end{aligned}
$$

$$= \quad N_n \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)'$$

and

$$N_n \boldsymbol{\nu_2}'^{(0)} S_n \beta_n^{(0)} \quad = \quad N_n \left(S_n^{-1} \frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta)\right)' S_n \left(S_n^{-1} \beta_n^*\right) = N_n \left(\frac{n}{N_n}\right) \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \beta_n^*.$$

Also,

$$N_n \beta_n'^{(0)} S_n \beta_n^{(0)} = N_n \left(S_n^{-1} \beta_n^*\right)' S_n \left(S_n^{-1} \beta_n^*\right) = N_n (\beta_n^*)' S_n^{-1} \beta_n^*.$$

Thus, by simplifying, we get that

$$\sum_{i=1}^{n} r_i^2 = N_n \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)' + 2N_n \left(\frac{n}{N_n}\right) \left(\bar{U}_n, \bar{V}_n\right)' S_n^{-1} \beta_n^* + N_n (\beta_n^*)' S_n^{-1} \beta_n^*.$$

Thus,

$$\begin{aligned} -2 log \mathcal{R}_n \quad &= \quad 2 \sum_{i=1}^{n} r_i - \sum_{i=1}^{n} r_i^2 + \frac{2}{3} \sum_{i=1}^{n} r_i^* \\ &= \quad N_n \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)' - N_n (\beta_n^*)' S_n^{-1} \beta_n^* + \frac{2}{3} \sum_{i=1}^{n} r_i^*. \end{aligned}$$

This completes the proof of the Proposition. □

2.2 Empirical Likelihood Confidence Region for Supercritical BPI

Recall the estimating equations for Supercritical BPI defined in (2.1), and for given $\hat{\lambda}$ and $\hat{b}^2$, solving tese estimating equations yield

$$\hat{m}_{n,EL} = \frac{\sum_{i=1}^{n} w_i X_i - \hat{\lambda}}{\sum_{i=1}^{n} w_i X_{i-1}}, \quad \text{and} \quad \hat{\sigma}^2_{n,EL} = \frac{\sum_{i=1}^{n} w_i U_i^2 - \hat{b}^2}{\sum_{i=1}^{n} w_i X_{i-1}}.$$

This section gives the strong consistency and asymptotic normality of the estimators defined above.

Theorem 2. *Assuming that $E(X_1^4) < \infty$ and $E(I_1^4) < \infty$. Then the following hold:*

1. *If $m > 1$, $\hat{m}_{n,EL} \xrightarrow{a.s.} m$, and $\hat{\sigma}^2_{n,EL} \xrightarrow{a.s.} \sigma^2$.*

2. $\sqrt{N_n}(\hat{m}_{n,EL} - m, \hat{\sigma}^2_{n,EL} - \sigma^2) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$*, where* $\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$.

3. *Consider the testing problem $H_0 : m = m_0, \sigma^2 = \sigma_0^2$. Then under H_0,*

$$-2 \log \mathcal{R}_n(\theta) \xrightarrow{d} \chi_2^2.$$

Proof. The consistency of $\hat{m}_{n,EL}$ and $\hat{\sigma}^2_{n,EL}$ follows from the consistency and asymptotic normality of the maximum likelihood estimator for m and σ^2 (see (Heyde 1976) and by an application of Toeplitz Lemma. In order to obtain the jointly asymptotic normality of

$(\hat{m}_{n,EL}, \hat{\sigma}^2_{n,EL})$, it is sufficient to show that

$$\sum_{i=1}^{n} E\left[\frac{U_i}{\sqrt{N_n}}\frac{V_i}{\sqrt{N_n}}\bigg|\mathcal{F}_{i-1}\right] = \frac{1}{N_n}\sum_{i=1}^{n} E\left[(X_i - mX_{i-1} - \lambda)^3\bigg|\mathcal{F}_{i-1}\right] \leq \frac{C}{N_n^{\delta}}, \qquad (2.10)$$

for some $0 < \delta < \frac{1}{2}$ using Kuelbs and Vidyashankar (2011), where C is a finite constant. By (2.10) we can show that $\frac{1}{N_n}\sum_{i=1}^{n} U_iV_i = o_p(1)$.

We turn to proving 3. Now we want to show the asymptotic behavior of $-2\log\mathcal{R}_n(\theta)$. First, we need to look at the asymptotic distribution of A_n. Let $A_n = \begin{pmatrix} A_n(1,1) & A_n(1,2) \\ A_n(2,1) & A_n(2,2) \end{pmatrix}$.

Note that by using (2.10),

$$\begin{aligned} A_n(1,1) &= \frac{N_n}{det(S_n)}\left(\frac{n}{N_n}\right)^3 (\bar{U}_n)^2\left(\bar{V}_n^{(2)}\right) \\ &= N_n\left(\frac{n}{N_n}\right)\frac{(\bar{U}_n)^2}{\left(\bar{U}_n^{(2)}\right)}\left[1 - \frac{\left(\frac{1}{N_n}\sum_{i=1}^{n} U_iV_i\right)^2}{\left(\frac{n}{N_n}\right)^2\left(\bar{U}_n^{(2)}\right)\left(\bar{V}_n^{(2)}\right)}\right]^{-1} \\ &= \frac{\left(\sqrt{n}\bar{U}_n\right)^2}{\left(\bar{U}_n^{(2)}\right)}(1 - o_p(1)). \end{aligned}$$

Then by the consistency of $\bar{U}_n^{(2)}$ and the martingale central limit theorem (MCLT), $\left(\sqrt{n}\bar{U}_n\right) \xrightarrow{d} N(0,\sigma^2)$. Hence $A_n(1,1) \xrightarrow{d} \chi^2_1$. Now by similar argument one can show that

$$A_n(2,2) = \frac{N_n}{det(S_n)}\left(\frac{n}{N_n}\right)^3 (\bar{V}_n)^2\left(\bar{U}_n^{(2)}\right) \xrightarrow{d} \chi^2_1$$

by the MCLT. Following (2.10), it can be seen that $A_n(1,2) = o_p(1)$ and $A_n(2,1) = o_p(1)$. Therefore,

$$A_n = N_n \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)' \xrightarrow{d} \chi_2^2, \tag{2.11}$$

We next investigate the asymptotic behavior of B_n. Now, by the MCLT, it follows that $\frac{1}{N_n}\sum_{i=1}^{n} U_i = O_p\left(N_n^{-\frac{1}{2}}\right)$ and $\frac{1}{N_n}\sum_{i=1}^{n} V_i = O_p\left(N_n^{-\frac{1}{2}}\right)$. Also, a standard calculation shows that

$$\max_{1\le i\le n} \boldsymbol{g}(X_i,\theta) = O_p\left(N_n^{\frac{1}{2}}\right), \tag{2.12}$$

and hence

$$\|\beta_n^*\| \le \left(\max_{1\le i\le n}\left|\frac{1}{1+r_i}\right|\right)\left(\max_{1\le i\le n}\boldsymbol{g}(X_i,\theta)\right)|\boldsymbol{\nu_2}'S(n)\boldsymbol{\nu_2}| \le o_p\left(N_n^{-\frac{1}{2}}\right), \tag{2.13}$$

This implies that $N_n\beta_n'^* S_n^{-1}\beta_n^* = o_p(1)$, implying $B_n \xrightarrow{p} 0$.

Next we turn to study the asymptotic behavior of R_n. Since r_i^* is the remainder term of Taylor expansion,then for any finite and positive constant C, it can be seen that $P(|r_i^*| \le C^*|r_i^3|, 1\le i\le n) \to 1$ as $n\to\infty$. Hence,

$$\left|\sum_{i=1}^{n} r_i^*\right| \le C\|\max_{1\le i\le n}\nu_2'\boldsymbol{g}(X_i,\theta)\|N_n\|\boldsymbol{\nu_2}'S_n\boldsymbol{\nu_2}\| = o_p(1),$$

using (2.12) and (2.13). Hence $R_n \xrightarrow{p} 0$. □

2.3 Empirical Likelihood for Subcritical BPI

Now we move to investigating the asymptotic distribution of $-2\log\mathcal{R}_n(\theta)$ for the subcritical case. It is known that under finite moment conditions, $\{X_n\}$ has a unique stationary distribution (see Athreya and Ney (1972)). Since if the distribution of X_0 is the stationary distribution, then X_0 is stationary and ergodic. Due to the coupling property of recurrent Markov chains, X_n can be taken to be stationary and ergodic, regardless of the distribution of X_0 (see Wei and Winnicki (1989)).

Lemma 1. *Assuming that $E(X_1^4) < \infty$ and $E(I_1^4) < \infty$, for $\alpha = 0, 1, 2$,*

$$\frac{1}{n}\sum_{i=1}^{n} X_{i-1}^{\alpha} \xrightarrow{a.s.} EX^{\alpha}, \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_{i-1}^{\alpha} U_i^2 \xrightarrow{a.s.} E\big[X^{\alpha}(\sigma^2 X + b^2)\big]. \tag{2.14}$$

Proof. See Theorem 2.1 in Wei and Winnicki (1990) and Proposition 3.1 in Winnicki (1991). □

Lemma 2. *Under the conditions in Lemma 1,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \xrightarrow{d} N\Big(0, E(\sigma^2 X + b^2)\Big), \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \xrightarrow{d} N\Big(0, E\big(R(X)\big)\Big) \tag{2.15}$$

X is a random variable with the stationary distribution of X_n, and $R(X) = 2\sigma^4 X^2 + (a^4 + 4\sigma^2 b^2 - 3\sigma^4)$, where $a^4 = E(\xi_{i,j} - m)^4$, and $c^4 = E(I_i - \lambda)^4$.

Proof. See Theorem 2.1 in Wei and Winnicki (1990) and Theorem 3.4 in Winnicki (1991) □

Lemma 3. *Under the conditions in Lemma 1, $\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i \xrightarrow{a.s.} 0$.*

Proof. The proof follows from Lemma 2 and an application of ergordic theorem. □

Theorem 3. *Assuming that $E(X_1^4) < \infty$, $E(I_1^4) < \infty$ and $E(log^{+} I_n) < \infty$, X_n. Then the following hold:*

1. If $m < 1$, $\hat{m}_{n,EL} \xrightarrow{a.s.} m$, *and* $\hat{\sigma}^2_{n,EL} \xrightarrow{a.s.} \sigma^2$.

2. $\sqrt{N_n}(\hat{m}_{n,EL} - m, \hat{\sigma}^2_{n,EL} - \sigma^2) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}\right)$, *where* $\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$.

3. Consider the testing problem $H_0 : m = m_0, \sigma^2 = \sigma_0^2$. *Then under* H_0,

$$-2\log\mathcal{R}_n(\theta) \xrightarrow{d} \chi^2_2.$$

Proof. The consistency and asymptotic normality of $\hat{m}_{n,EL}$ and $\hat{\sigma}^2_{n,EL}$ can be demonstrated by using Lemma 1 and Lemma 2. Now we turn to prove (3). Note that A_n can be written as

$$A_n = N_n\left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)' = N_n\mathbf{1}\ \Delta_n\mathbf{1}',$$

where

$$\mathbf{1} = (1,1) \quad \text{and} \quad \Delta_n = \begin{pmatrix} \Delta_n(1,1) & \Delta_n(1,2) \\ \Delta_n(2,1) & \Delta_n(2,2) \end{pmatrix},$$

$$\Delta_n(1,1) = (det(S_n))^{-1}\left(\frac{1}{N_n}\sum_{i=1}^n U_i\right)^2 \left(\frac{1}{N_n}\sum_{i=1}^n V_i^2\right) = (det(S_n))^{-1}\left(\frac{n}{N_n}\right)^3 \bar{U}_n^2\bar{V}_n^{(2)},$$

$$\Delta_n(2,2) = (det(S_n))^{-1}\left(\frac{1}{N_n}\sum_{i=1}^n V_i\right)^2 \left(\frac{1}{N_n}\sum_{i=1}^n U_i^2\right) = (det(S_n))^{-1}\left(\frac{n}{N_n}\right)^3 \bar{V}_n^2\bar{U}_n^{(2)}.$$

Using Lemma 1 and Lemma 3, one can show that $N_n\Delta_n(1,2) = N_n\Delta_n(2,1) = o(1)$. Now,

$N_n\Delta_n(1,1)$ can be simplified to be

$$N_n\Delta_n(1,1) = \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} U_i^2} = \frac{\left(\sqrt{n}\bar{U}_n\right)^2}{\bar{U}_n^2} \xrightarrow{d} \chi_1^2, \tag{2.16}$$

where the convergence follows from the MCLT.

We next deal with $\Delta_n(2,2)$. Notice that using simplifications similar to (2.16) and using the MCLT, one can show that

$$N_n\Delta_n(2,2) = \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} V_i^2} = \frac{\left(\sqrt{n}\bar{V}_n\right)^2}{\bar{V}_n^2} \xrightarrow{d} \chi_1^2. \tag{2.17}$$

Thus, we have $A_n \xrightarrow{d} \chi_2^2$ under H_0. Hence,

$$A_n = N_n\left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)' \xrightarrow{d} \chi_2^2.$$

We now turn to the proof of $B_n \xrightarrow{p} 0$. Again, we recall that $B_n = N_n(\beta_n^*)' S_n^{-1}\beta_n^*$. To establish that $B_n \xrightarrow{p} 0$, we need to identify the rate of convergence of β_n^*. To this end, note that using $\big||x|-|y|\big| \geq |x-y|$ and algebra, it can be seen that for any $\boldsymbol{l}'$

$$\left|\boldsymbol{l}'\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\right| \geq \frac{\rho(n)}{1+\rho(n)\max\limits_{1\leq i\leq n} \boldsymbol{l}'\boldsymbol{g}(X_i,\theta)}.$$

We will show that

$$\frac{\rho(n)}{1+\rho(n)\max\limits_{1\le i\le n} l'\boldsymbol{g}(X_i,\theta)} = O_p\left(\frac{1}{\sqrt{N_n}}\right). \tag{2.18}$$

To this end, note that

$$\frac{\rho(n)}{1+\rho(n)\max\limits_{1\le i\le n} l'\boldsymbol{g}(X_i,\theta)} \le \left|\boldsymbol{l}'\begin{pmatrix}\bar{U}_n\\ \bar{V}_n\end{pmatrix}\right| = O_p\left(\frac{1}{\sqrt{N_n}}\right),$$

where the last identity follows from the MCLT. Now, expressing

$$\max_{1\le i\le n}\boldsymbol{g}(X_i,\theta) \le \left\|\begin{pmatrix}\sqrt{N_n}\sqrt{\dfrac{n}{N_n}}\sqrt{n}\bar{U}_n\\ \sqrt{N_n}\sqrt{\dfrac{n}{N_n}}\sqrt{n}\bar{V}_n\end{pmatrix}\right\|,$$

it follows that

$$\max_{1\le i\le n}\boldsymbol{g}(X_i,\theta) = O_p(\sqrt{N_n}). \tag{2.19}$$

Now, returning to β_n^*, notice that

$$\|\beta_n^*\| \le \left(\max_{1\le i\le n}\boldsymbol{g}(X_i,\theta)\right)\max_{1\le i\le n}\left(\left|\frac{1}{1+r_i}\right|\right)\|\nu_2' S_n \nu_2\|.$$

Thus, we need the rate of convergence of $\nu_2' S_n \nu_2$. First, for $\bar{U}_n^{(2)}$, since $\left\{\sum_{i=1}^{n}\left(V_i^2 - R(X_{i-1})\right)\right\}$ is a martingale, where $R(X_{i-1}) = E(V_i^2|\mathcal{F}_{i-1}) = 2\sigma^4 X_{i-1}^2 + (a^4 + 4\sigma^2 b^2 - 3\sigma^4)$, and $a^4 =$

$E(\xi_{i,j} - m)^4$, $c^4 = E(I_i - \lambda)^4$. Then,

$$
\begin{aligned}
& \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\Big[\Big(V_i^2 - R(X_{i-1})\Big)^2 |\mathcal{F}_{i-1}\Big] \\
= & \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big(E\big(V_i^4|\mathcal{F}_{i-1}\big) - 2R(X_{i-1})E\big(V_i^4|\mathcal{F}_{i-1}\big) + R(X_{i-1})^2\Big) \\
= & \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big(E\big(V_i^4|\mathcal{F}_{i-1}\big) - R^2(X_{i-1})\Big) \\
= & \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big(E\big(V_i^4|\mathcal{F}_{i-1}\big)\Big) - E[R^2(X)]
\end{aligned}
$$

Thus, under the finite fourth moment (assumptions in the Theorem 3), following the ergodic theorem one can show that

$$
\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big(E\big(V_i^4|\mathcal{F}_{i-1}\big)\Big) - E[R^2(X)] = E[K(X) - R^2(X)].
$$

Therefore by the MCLT,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(V_i^2 - R(X_{i-1})\Big) \xrightarrow{\text{d}} N\Big(0, E[K(X) - R^2(X)]\Big),
$$

then

$$
\frac{1}{N_n} \sum_{i=1}^{n} V_i^2 = \frac{1}{\sqrt{N_n}} \sqrt{\frac{n}{N_n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i^2 = O_p\bigg(\frac{1}{\sqrt{N_n}}\bigg).
$$

Now we may write

$$
\|\beta_n^*\| \quad = \quad \left\| \frac{1}{N_n} \boldsymbol{g}(X_i, \theta) \frac{r_i^2}{1 + r_i^2} \right\|
$$

$$
\begin{aligned}
&\leq \left(\max_{1\leq i\leq n}|\frac{1}{1+r_i}|\right)\left(\max_{1\leq i\leq n}\boldsymbol{g}(X_i,\theta)\right)\|\boldsymbol{\nu_2}'S_n\boldsymbol{\nu_2}\| \\
&= O_p(1)O_p(N_n^{-\frac{1}{2}})O_p(N_n^{-\frac{1}{2}}))o_p(1)O_p(N_n^{-\frac{1}{2}}) \\
&= o_p(N_n^{-\frac{1}{2}}),
\end{aligned}
\tag{2.20}
$$

and

$$
N_n\beta_n'^{*}S_n^{-1}\beta_n^{*} = N_n o_p(N_n^{-\frac{1}{2}})O_p(1)o_p(N_n^{-\frac{1}{2}}) = o_p(1). \tag{2.21}
$$

Hence, $B_n \xrightarrow{p} 0$.

We now establish the convergence of R_n to 0 in probability.

$$
\begin{aligned}
|\sum_{i=1}^n r_i^*| &\leq C\sum_{i=1}^n |r_i|^3 \\
&\leq C\|\max_{1\leq i\leq n}\nu_2'\boldsymbol{g}(X_i,\theta)\|\sum_{i=1}^n\left|\nu_2'\boldsymbol{g}(X_i,\theta)\boldsymbol{g}(X_i,\theta)'\boldsymbol{\nu_2}\right| \\
&\leq C\|\max_{1\leq i\leq n}\nu_2'\boldsymbol{g}(X_i,\theta)\|N_n\|\boldsymbol{\nu_2}'S_n\boldsymbol{\nu_2}\| \\
&= O_p(1)O(N_n)O_p(N_n^{-\frac{1}{2}})o_p(1)O_p(N_n^{-\frac{1}{2}}) \quad \text{by} \quad (2.19) \text{ and } (2.20) \\
&= o_p(1).
\end{aligned}
$$

Hence, $R_n \xrightarrow{p} 0$. This completes the proof of 3. □

Chapter 3: Empirical Likelihood Confidence Region for Critical BPI

Now we move to the critical branching process with immigration. We first summarize some known results concerning critical BPI. The weak convergence of $\frac{X_{[nt]}}{n}$ to a diffusion process makes the inferential problem more interesting and complicated. The following Lemma describes the properties of Critical BPI. The pioneering work from Seneta (1970), Pakes (1971), Pakes (1972), Kawazu and Watanabe (1971), Wei and Winnicki (1989) and Winnicki (1991) have discussed extensively the dichotomy of critical branching processes. The dichotomy can be described in the following Lemma.

Lemma 4. *Let $m = 1$, and $\tau = \frac{2\lambda}{\sigma^2}$. When $\tau > 1$, the BPI is transient; else when $\tau \leq 1$, it is null-recurrent.*

Let $D^+[0,\infty)$ be the space of nonnegative functions which are right continuous and having left limits defined on $[0,\infty)$. Let $Y_n(t) = \frac{X_{[nt]}}{n}$, then $Y_n(t)$ is a sequence of random elements in $D^+[0,\infty)$, and $Y_n(t)$ weakly converges (converges in distribution) to $Y(t)$ in $D^+[0,\infty)$, where Y is a diffusion with generator $Af(x) = \frac{1}{2}\sigma^2 x f''(x) + \lambda f'(x)$, $f \in C_c^\infty[0,\infty)$, and $Y(0) = 0$. Kawazu and Watanabe (1971) and Mellein (1982) Mellein (1983) investigated the convergence of finite dimensional distributions of $\{Y_n\}$. Lindvall (1972) was the first to provide the proof of functional convergence for branching processes given nontermination. Wei and Winnicki (1989) extended their work and demonstrated the following lemma.

Lemma 5. *When $m = 1$, and $EX^k < \infty$, then*

$$\left(\frac{X_n}{n}, \frac{1}{n^2}\sum_{i=1}^{n} X_i, \ldots, \frac{1}{n^{k+1}}\sum_{i=1}^{n} X_i^k\right) \xrightarrow{d} \left(Y(1), \int_0^1 Y(t)dt, \ldots, \int_0^1 Y^k(t)dt\right) \tag{3.1}$$

Our main result in this chapter is the following:

Theorem 4. *Assuming that $m = 1$, $E(X_1^4) < \infty$ and $E(I_1^4) < \infty$, then the following hold:*

1. *If $m = 1$, $\hat{m}_{n,EL} \xrightarrow{p} m$, $\hat{\sigma}^2_{n,EL} \xrightarrow{p} \sigma^2$.*

2. *Consider the testing problem $H_0 : m = m_0, \sigma^2 = \sigma_0^2$. Then under H_0,*

$$-2 \log \mathcal{R}_n(\theta) \xrightarrow{d} D_1^2 + D_2^2$$

when $2\lambda > \sigma^2$. D_1^2 and D_2^2 are two independent random variables which are distributed as functionals of Feller-diffusion with

$$D_1^2 = \frac{(Y(1) - \lambda)^2}{\sigma^2 \int_0^1 Y(t)dt} \quad \text{and} \quad D_2^2 = \frac{(\int_0^1 Y(t)dB(t))^2}{\int_0^1 Y^2(t)dt},$$

where $B(t)$ is a Brownian motion; and if $2\lambda \leq \sigma^2$,

$$-2 \log \mathcal{R}_n(\theta) \xrightarrow{p} 0.$$

Proposition 2. *Under the conditions of Theorem 4 and $\tau > 1$, $A_n \xrightarrow{d} D_1^2 + D_2^2$, where D_1^2 and D_2^2 are defined the same as above.*

To prove this Proposition, we need the following Lemmas.

Lemma 6. *Under the conditions of Theorem 4, $\frac{n}{N_n}\bar{U}_n^{(2)} = O_p(1)$, and $\frac{1}{N_n}\bar{V}_n^{(2)} = O_p(1)$.*

Proof. It can be seen that

$$\frac{n}{N_n}\bar{U}_n^{(2)} = \sigma^2 + \frac{1}{n}\Big(\frac{n^2}{N_n}\Big)b^2 + \frac{1}{\sqrt{n}}\Big(\frac{n^2}{N_n}\Big)\Big(\frac{1}{n^{3/2}}\sum_{i=1}^{n} V_i\Big) = O_p(1),$$

by Lemma 5 and Lemma 2.8 in Winnicki (1991). Let $R(X_{i-1}) = E(V_i^2|\mathcal{F}_{i-1}) = 2\sigma^4 X_{i-1}^2 + (a^4 + 4\sigma^2 b^2 - 3\sigma^4)$, and $a^4 = E(\xi_{i,j} - m)^4$, $c^4 = E(I_i - \lambda)^4$. Similarly, one can show that,

$$\begin{aligned}
\frac{n}{N_n}\bar{V}_n^{(2)} &= \frac{1}{N_n}\sum_{i=1}^{n}\Big(V_i^2 - R(X_{i-1})\Big) + \frac{1}{N_n}\sum_{i=1}^{n} R(X_{i-1}) \\
&= n\left[2\sigma^4 \frac{\frac{1}{n^3}\sum\limits_{i=1}^{n} X_{i-1}^2}{\frac{1}{n^2}N_n} + \frac{1}{n}(a^4 + 4\sigma^2 b^2 - 3\sigma^4) + \frac{(c^4 - b^4)}{N_n}\right] \\
&= O_p(n),
\end{aligned}$$

where the last equality follows from Lemma 5 and Lemma 2.8 in Winnicki (1991). □

Lemma 7. *Under the conditions of Theorem 4, and when* $\tau > 1$, $\frac{1}{N_n}\sum\limits_{i=1}^{n} U_i V_i = o_p(\sqrt{n})$.

Proof. It can be seen that

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i = \frac{1}{N_n}\sum_{i=1}^{n} V_i\omega_i + \frac{1}{N_n}\sum_{i=1}^{n} V_i I_i - \lambda\frac{1}{N_n}\sum_{i=1}^{n} V_i = T_1 + T_2 - T_3,$$

where $\omega_i = \sum_{j=1}^{X_{i-1}}(\xi_{i,j} - 1)$, $T_1 = \frac{1}{N_n}\sum\limits_{i=1}^{n} V_i\omega_i$, $T_2 = \frac{1}{N_n}\sum\limits_{i=1}^{n} V_i I_i$, and $T_3 = \lambda\frac{1}{N_n}\sum\limits_{i=1}^{n} V_i$. By applying Lemma 2.8 of Winnicki (1991), it follows that T_3 converges to zero in probability, that is,

$$\frac{1}{N_n}\sum_{i=1}^{n} V_i = \frac{1}{\sqrt{n}}\Big(\frac{n^2}{N_n}\Big)\Big(\frac{1}{n^{3/2}}\sum_{i=1}^{n} V_i\Big) \xrightarrow{p} 0. \tag{3.2}$$

As for T_2, by SLLN $\lim_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n} I_i = \lambda$ a.s., and $\sum\limits_{i=1}^{n} V_i = n^{3/2}\frac{1}{n^{3/2}}\sum\limits_{i=1}^{n} V_i \to \infty$ as $n \to \infty$, since $\frac{1}{\sqrt{2\sigma^2} n^{3/2}}\sum\limits_{i=1}^{n} V_i \xrightarrow{d} \int_0^1 Y(t)dB(t)$. Thus, it follows by an application of Toeplitz

Lemma, that

$$T_2 = \frac{1}{N_n}\sum_{i=1}^{n} V_i I_i = \frac{1}{\sqrt{n}}\frac{\frac{1}{n^{3/2}}\sum\limits_{i=1}^{n} X_{i-1}^2}{\frac{1}{n^2}\sum\limits_{i=1}^{n} X_{i-1}}\left(\frac{\sum\limits_{i=1}^{n} V_i I_i}{\sum\limits_{i=1}^{n} V_i}\right) \to 0. \tag{3.3}$$

We now deal with T_1,

$$T_1 = \frac{1}{N_n}\sum_{i=1}^{n} V_i\omega_i = \left(\frac{\sum\limits_{i=1}^{n} V_i X_{i-1}}{\sum\limits_{i=1}^{n} X_{i-1}}\right)\left(\frac{\sum\limits_{i=1}^{n} V_i\omega_i}{\sum\limits_{i=1}^{n} V_i X_{i-1}}\right). \tag{3.4}$$

The second term of the above equation converges to 0 using Lemma 2.14 of Wei and Winnicki (1989) and Lemma 2.8 of Winnicki (1991), and Toeplitz Lemma. Now, the first term can be expressed as

$$\sqrt{n}\left(n^{-5/2}\sum_{i=1}^{n} V_i X_{i-1}\right)\left(n^{-2}\sum_{i=1}^{n} X_{i-1}\right)^{-1}$$

Again, by Lemma 2.14 of Wei and Winnicki (1989) and Lemma 2.8 of Winnicki (1991), the above expression is $O_p(\sqrt{n})$. Using this in T_1, the lemma follows. □

Now we turn to the Proof of Proposition 2.

Proof. Recall that

$$A_n = N_n\left(\frac{n}{N_n}\right)^2\left(\bar{U}_n, \bar{V}_n\right) S_n^{-1}\left(\bar{U}_n, \bar{V}_n\right)' = N_n \mathbf{1}\Delta^{\star}(n)\mathbf{1}',$$

where $\mathbf{1} = (1,1)$ and $\Delta^\star(n) = \begin{pmatrix} \Delta_n^\star(1,1) & \Delta_n^\star(1,2) \\ \Delta_n^\star(2,1) & \Delta_n^\star(2,2) \end{pmatrix}$.

$$\Delta_n^\star(1,1) = det(S_n^{-1}) \left(\frac{n}{N_n}\right)^3 \bar{U}_n^2 \bar{V}_n^{(2)}, \quad \text{and} \quad \Delta_n^\star(2,2) = det(S_n^{-1}) \left(\frac{n}{N_n}\right)^3 \bar{V}_n^2 \bar{U}_n^{(2)}.$$

Following Lemma 5 and Lemma 6, $N_n\Delta_n^\star(1,2) = N_n\Delta_n^\star(2,1) = o_p(1)$. We now deal with $N_n\Delta_n^\star(1,1)$. After tedious algebra, and using Lemma 5, Lemma 6 and Lemma 7, it can be seen that

$$\begin{aligned} N_n\Delta_n^\star(1,1) &= \frac{N_n}{det(S_n)}\left(\frac{1}{N_n}\sum_{i=1}^n U_i\right)^2 \frac{1}{N_n}\sum_{i=1}^n V_i^2 \\ &= \frac{N_n\left(\frac{1}{N_n}\sum_{i=1}^n U_i\right)^2\left(\frac{1}{N_n}\sum_{i=1}^n V_i^2\right)}{\left(\bar{U}_{N_n}^{(2)}\right)\left(\bar{V}_{N_n}^{(2)}\right) - \left(\frac{1}{N_n}\sum_{i=1}^n U_iV_i\right)^2} \\ &= \frac{\left(\frac{1}{n}\sum_{i=1}^n U_i\right)^2}{\frac{N_n}{n^2}\sigma^2 + \frac{1}{n}b^2 + \frac{1}{\sqrt{n}}\left(\frac{1}{n^{3/2}}\sum_{i=1}^n V_i\right)} \frac{1}{1 - o_p(1)} \\ &= \frac{\bar{U}_n^2}{n^{-2}N_n\sigma^2 + n^{-1}b^2 + n^{-1}\bar{V}_n}. \end{aligned}$$

Now noticing that $\frac{1}{n}\sum_{i=1}^n U_i \xrightarrow{d} Y(1) - \lambda$ (Lemma 2.7 in Winnicki (1991)) and Lemma 5, it follows that

$$N_n\Delta_n^\star(1,1) \xrightarrow{d} \frac{\left(Y(1) - \lambda\right)^2}{\sigma^2 \int_0^1 Y(t)dt}. \tag{3.5}$$

We now turn to $N_n\Delta_n^\star(2,2)$. Using similar argument (Lemma 2.8 in Winnicki (1991)), it follows that

$$
\begin{aligned}
N_n\Delta_n^\star(1,1) &= \frac{N_n}{det(S_n)}\left(\frac{1}{N_n}\sum_{i=1}^{n}V_i\right)^2\frac{1}{N_n}\sum_{i=1}^{n}U_i^2 \\
&= \frac{2\sigma^4\left(\frac{1}{\sqrt{2}\sigma^2}\frac{1}{n^{3/2}}\sum_{i=1}^{n}V_i\right)^2}{2\sigma^4\left[\frac{1}{n^3}\sum_{i=1}^{n}X_{i-1}^2+\frac{1}{n}\frac{1}{n^2}N_n(a^4+4\sigma^2b^2-3\sigma^4)+\frac{1}{n^2}(c^4-b^4)\right]}\frac{1}{1-o_p(1)}.
\end{aligned}
$$

As before, again using Lemma 5 and Lemma 2.8 in Winnicki (1991), it follows that

$$
N_n\Delta_n^\star(2,2) \xrightarrow{d} \frac{\left(\int_0^1 Y(t)dB(t)\right)^2}{\int_0^1 Y^2(t)dt}. \tag{3.6}
$$

Therefore,

$$
A_n \xrightarrow{d} \frac{\left(Y(1)-\lambda\right)^2}{\sigma^2\int_0^1 Y(t)dt}+\frac{\left(\int_0^1 Y(t)dB(t)\right)^2}{\int_0^1 Y^2(t)dt}.
$$

Hence, when $\tau > 1$, A_n converges to sum of two independent functionals of diffusions in distribution, which can be written as

$$
A_n \xrightarrow{d} D_1^2 + D_2^2, \tag{3.7}
$$

where

$$
D_1^2 = \frac{\left(Y(1)-\lambda\right)^2}{\sigma^2\int_0^1 Y(t)dt} \text{ and } D_2^2 = \frac{\left(\int_0^1 Y(t)dB(t)\right)^2}{\int_0^1 Y^2(t)dt}.
$$

□

Proposition 3. *Under the same conditions of Theorem 4, when* $\tau \leq 1$ *and* $E(\xi_{n,j}^2 log^+ \xi_{n,j})$

$A_n \xrightarrow{p} 0.$

We begin with considering when $\tau = 1$. The following two lemmas due to Winnicki (1991) will be used repeatedly in order to prove Proposition 3.

Lemma 8. *Under the same condition of Theorem 4 and when* $\tau = 1$,

$$\sum_{i=1}^{n} \frac{U_i}{X_{i-1}} = o\left(\left(\sum_{i=1}^{n} \frac{1}{X_{i-1}}\right)^{\alpha}\right) \ \textit{a.s. for any } \alpha > 1$$

and

$$\sum_{i=1}^{n} \frac{1}{X_{i-1}} = o(n^{\beta}) \ \textit{a.s. for any } \beta > 0.$$

Proof. See Lemma 2.10 (a) and Lemma 2.12 (a) of Winnicki (1991). □

Lemma 9. *Under the same condition in Theorem 4, and when* $\tau = 1$, *for* $\beta > 0$,

$$\frac{1}{N_n} \sum_{i=1}^{n} U_i V_i = o_p(n^{\beta + \frac{1}{2}}).$$

Proof. We may write

$$\frac{1}{N_n} \sum_{i=1}^{n} U_i V_i = \left(\sqrt{n} \frac{\frac{1}{n^{5/2}} \sum_{i=1}^{n} V_i X_{i-1}}{\frac{1}{n^2} N_n}\right) \left(\sum_{i=1}^{n} \frac{1}{X_{i-1}}\right) \left(\frac{\sum_{i=1}^{n} \frac{(U_i/X_{i-1})}{\sum_{i=1}^{n} 1/X_{i-1}} V_i X_{i-1}}{\sum_{i=1}^{n} V_i X_{i-1}}\right).$$

Using Lemma 8 and Toeplitz lemma, the above equation reduces to

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i = \left(\frac{1}{n^{5/2}}\sum_{i=1}^{n} V_i X_{i-1}\right)^{-1}\left(\frac{1}{n^2}N_n\right) o(n^{\beta+\frac{1}{2}}) = o_p(n^{\beta+\frac{1}{2}}),$$

following by Lemma 5 and Lemma 2.8 in Winnicki (1991). □

Now we move to the proof of Proposition 3.

Proof. Recall that

$$A_n = N_n \begin{pmatrix} \bar{U}_n, & \bar{V}_n \end{pmatrix} S_n^{-1} \begin{pmatrix} \bar{U}_n, & \bar{V}_n \end{pmatrix}' = N_n \mathbf{1}\Delta^\star(n)\mathbf{1}',$$

where $\mathbf{1} = (1, 1)$ and $\Delta^\star(n) = \begin{pmatrix} \Delta_n^\star(1,1) & \Delta_n^\star(1,2) \\ \Delta_n^\star(2,1) & \Delta_n^\star(2,2) \end{pmatrix}$.

$$\Delta_n^\star(1,1) = det(S_n^{-1})\left(\frac{n}{N_n}\right)^3 \bar{U}_n^2 \bar{V}_n^{(2)}, \quad \text{and} \quad \Delta_n^\star(2,2) = det(S_n^{-1})\left(\frac{n}{N_n}\right)^3 \bar{V}_n^2 \bar{U}_n^{(2)}.$$

Following Lemma 5 and Lemma 6, it can be shown that $N_n\Delta_n^\star(1,2) = o_p(1)$ and $N_n\Delta_n^\star(2,1) = o_p(1)$. We first deal with the convergence rates for $N_n\Delta_n^\star(1,1)$ and $N_n\Delta_n^\star(2,2)$. For any $\beta > 0$, we may derive

$$\begin{aligned} N_n\Delta_n^\star(1,1) &= \frac{N_n}{det(S_n)}\left(\frac{1}{N_n}\sum_{i=1}^{n} U_i\right)^2 \frac{1}{N_n}\sum_{i=1}^{n} V_i^2 \\ &= \frac{N_n\left(\frac{1}{N_n}\sum_{i=1}^{n} U_i\right)^2\left(\frac{1}{N_n}\sum_{i=1}^{n} V_i^2\right)}{\left(\bar{U}_{N_n}^{(2)}\right)\left(\bar{V}_{N_n}^{(2)}\right) - \left(\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i\right)^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(\frac{n^2}{N_n}\right)^2 \left(\frac{1}{n}\sum_{i=1}^{n} U_i\right)^2}{\frac{N_n}{n^2}\sigma^2 + \frac{1}{n}b^2 + \frac{1}{\sqrt{n}}\left(\frac{1}{n^{3/2}}\sum_{i=1}^{n} V_i\right)} \left(\frac{1}{1 - o_p(n^{2\beta})}\right) \\
&= o_p(1),
\end{aligned}
$$

where the last equality follows from Lemma 5 and Lemma 9. Similarly, it follows that

$$
\begin{aligned}
N_n N_n \Delta_n^\star(2,2) &= \frac{2\sigma^4 \frac{1}{n^2} N_n \left(\frac{n^2}{N_n}\right)^2 \left(\frac{1}{\sqrt{2}\sigma^2}\frac{1}{n^{3/2}}\sum_{i=1}^{n} V_i\right)^2}{2\sigma^4\left[\frac{\frac{1}{n^3}\sum_{i=1}^{n} X_{i-1}^2}{\frac{1}{n^2}N_n} + \frac{a^4 + 4\sigma^2 b^2 - 3\sigma^4}{n} + \frac{c^4 - b^4}{N_n}\right]} \left(\frac{1}{1 - o_p(n^{2\beta})}\right). \\
&= o_p(1)
\end{aligned}
$$

Hence, this proves $A_n \xrightarrow{p} 0$ when $\tau = 1$.

Now we turn to the case when $\tau < 1$. The proof once again relies on the proof of Winnicki (1991).

Lemma 10. *Under the same condition of Theorem 4, if* $\tau < 1$, $\sum_{i=1}^{n} 1/X_{i-1} = O_p(n^{1-\tau})$.

Proof. See Lemma 2.12 (b) in Winnicki (1991). □

Using Lemma 10 and after calculation, we can obtain the following lemma.

Lemma 11. *Under the same condition of Theorem 4 and when* $\tau < 1$,

$$
\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i = o_p(n^{\frac{3}{2}-\tau})
$$

Proof. Following a similar calculation in Lemma 9, and using Lemma 5 and Lemma 10, one can show that

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i = \sqrt{n}\left(\frac{1}{n^{5/2}}\sum_{i=1}^{n} V_i X_{i-1}\right)\left(\frac{1}{n^2}N_n\right)^{-1} o(1)\left(\sum_{i=1}^{n}\frac{1}{X_{i-1}}\right) = o_p(n^{\frac{3}{2}-\tau}).$$

□

As before, using a similar argument and Lemma 10 and Lemma 11, one can show that when $N_n\Delta_n^\star(1,1) = o_p(1)$, $N_n\Delta_n^\star(2,2) = o_p(1)$, and $N_n\Delta_n^\star(1,2) = N_n\Delta_n^\star(2,1) = o_p(1)$. Hence, when $\tau < 1$, $A_n \xrightarrow{p} 0$. This completes the proof of Proposition 3. □

Proposition 4. *Under the same conditions of Theorem 4,* $B_n \xrightarrow{p} 0$ *and* $R_n \xrightarrow{p} 0$.

Proof. First we want to investigate the asymptotic behavior of B_n. By using

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i = O_p\left(\frac{1}{n}\right), \quad \frac{1}{N_n}\sum_{i=1}^{n} V_i = O_p\left(\frac{1}{\sqrt{n}}\right)$$

we have

$$\left| \boldsymbol{l}_n' \begin{pmatrix} \frac{1}{N_n}\sum_{i=1}^{n} U_i \\ \frac{1}{N_n}\sum_{i=1}^{n} V_i \end{pmatrix} \right| \geq \frac{\rho(n)}{1+\rho(n)\max_{1\leq i\leq n} l'\boldsymbol{g}(X_i,\theta)};$$

Since LHS of the above equation is $O_p(\frac{1}{n})$, it follows that

$$\frac{\rho(n)}{1+\rho(n)\max_{1\leq i\leq n} l_n'\boldsymbol{g}(X_i,\theta)} = O_p\left(\frac{1}{n}\right). \tag{3.8}$$

Also

$$\max_{1\le i\le n}\|\boldsymbol{g}(X_i,\theta)\| \le \left\|\begin{pmatrix} O_p(n) \\ O_p(n^{3/2}) \end{pmatrix}\right\|,\ \|\boldsymbol{\nu_2}\| = \left\|\left(O_p(\frac{1}{n}), O_p(\frac{1}{n^2})\right)\right\|.$$

Now we may write $\|\beta_n^*\|$ is bounded by

$$\|\beta_n^*\| = \left\|\frac{1}{N_n}\sum_{i=1}^{n}\boldsymbol{g}(X_i,\theta)\left(\frac{r_i^2}{1+r_i^2}\right)\right\| \le \left(\max_{1\le i\le n}\left|\frac{1}{1+r_i}\right|\right)\left(\max_{1\le i\le n}\boldsymbol{g}(X_i,\theta)\right)\|\nu_2' S_n \nu_2\|.$$

where the rate of convergence of S_n can be determined with the following lemma.

Lemma 12. *Under the same condition of Theorem 4, we have*

1.

$$\frac{1}{N_n}\sum_{i=1}^{n}U_iV_i = \begin{cases} n^{-\frac{1}{2}}N_n\sum_{i=1}^{n}V_iX_{i-1}o(\sqrt{n}) = o_p(\sqrt{n}), & \text{if } \tau > 1; \\ n^{-\frac{1}{2}}N_n\sum_{i=1}^{n}V_iX_{i-1}o(n^{\beta+\frac{1}{2}}) = o_p(n^{\beta+\frac{1}{2}}), & \text{if } \beta > 0 \text{ and } \tau = 1; \\ n^{-\frac{1}{2}}N_n\sum_{i=1}^{n}V_iX_{i-1}o(n^{\frac{3}{2}-\tau}) = o_p(n^{\frac{3}{2}-\tau}), & \text{if } \tau < 1; \end{cases} \tag{3.9}$$

2.

$$\frac{\sum_{i=1}^{n}U_i^3}{N_n} = \begin{cases} O_p(1) + o_p(n^{\frac{1}{2}}) = o_p(\sqrt{n}), & \text{if } \tau > 1; \\ O_p(1) + o_p(n^{\beta+\frac{1}{2}}) = o_p(n^{\beta+\frac{1}{2}}), & \text{if } \beta > 0 \text{ and } \tau = 1; \\ O_p(1) + o_p(n^{\frac{3}{2}-\tau}) = o_p(n^{\frac{3}{2}-\tau}), & \text{if } \tau < 1; \end{cases} \tag{3.10}$$

3.

$$\frac{\sum_{i=1}^{n} U_i V_i^2}{N_n} + \frac{\sum_{i=1}^{n} U_i^3}{N_n} = \begin{cases} o_p(n^{1/2}) + O_p(n) = O_p(n), & \text{if } \tau > 1; \\ o_p(n^{\beta+\frac{1}{2}}) + O_p(n) = \begin{cases} O_p(n), & \text{for } 0 < \beta \le \frac{1}{2} \\ o_p(n^{\beta+\frac{1}{2}}), & \text{for } \beta > \frac{1}{2} \end{cases} & \text{if } \tau = 1; \\ o_p(n^{\frac{3}{2}-\tau}) + O_p(n) = \begin{cases} O_p(n), & \text{for } \frac{1}{2} \le \tau < 1 \\ o_p(n^{\frac{3}{2}}), & \text{for } 0 < \tau < \frac{1}{2} \end{cases} & \text{if } \tau < 1; \end{cases} \tag{3.11}$$

Proof. The proof of this lemma can be seen by showing that

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i^3 = \frac{1}{N_n}\sum_{i=1}^{n} U_i V_i + \sigma^2 \frac{1}{N_n}\sum_{i=1}^{n} U_i X_{i-1} + b^2 \frac{1}{N_n}\sum_{i=1}^{n} U_i = \frac{1}{N_n}\sum_{i=1}^{n} U_i V_i + O_p(1)$$

and

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i^2 \approx \frac{\sum_{i=1}^{n} U_i R(X_{i-1})}{N_n} = O_p(n)$$

$$\frac{1}{N_n}\sum_{i=1}^{n} U_i^2 V_i = \sigma^2 \frac{1}{N_n}\sum_{i=1}^{n} V_i X_{i-1} + b^2 \frac{1}{N_n}\sum_{i=1}^{n} V_i + \frac{1}{N_n}\sum_{i=1}^{n} V_i^2, = O_p(n)$$

and

$$\begin{aligned} \frac{\sum_{i=1}^{n} V_i^3}{N_n} &= 2\sigma^4 \frac{\sum_{i=1}^{n} V_i X_{i-1}^2}{N_n} + (a^4 + 4\sigma^2 b^2 - 3\sigma^4)\frac{\sum_{i=1}^{n} V_i X_{i-1}}{N_n} + (c^4 - b^4)\frac{1}{N_n}\sum_{i=1}^{n} V_i \\ &= O_p(n^{3/2}) + O_p(n^{1/2}) + o_p(1) = O_p(n^{3/2}). \end{aligned}$$

□

After algebra, it can be seen that $det(S_n) = o_p(n^{3+\beta})$ fr $\beta > 0$, and $S_n = O_p(n)$ and this implies $\|\beta_n^*\| = O_p\left(\frac{1}{n}\right)$. By noticing that $S_n^{-1} = o_p(n^{-1})$, therefore, $B_n = o_p(1)$. Now we turn to study the convergence rate of R_n. It can be shown that

$$\left|\sum_{i=1}^{n} r_i^*\right| \quad \leq \quad \sum_{i=1}^{n} |r_i^*| = C\|\boldsymbol{\nu_2}\|^3 N_n \left\| \frac{1}{N_n} \sum_{i=1}^{n} \begin{pmatrix} U_i^3 + U_i V_i^2 \\ U_i^2 V_i + V_i^3 \end{pmatrix} \right\| = o_p(1).$$

This completes the proof of Proposition 4. □

Chapter 4: Numerical Methods for Empirical Likelihood

In this chapter, we provide algorithm for point and interval estimating (m, σ^2) for the supercritical case, and (m, σ^2) and (λ, b^2) for the non-supercritical case. We discuss the algorithms in two different cases, that are referred as branching processes with observed immigration and branching processes with unobserved immigration, namely, we can only observe the total population size for each generation, but neither the information for the number of offspring nor the number of immigration are not available. If the immigration can be observed, due to the independence of the offspring and immigration, and also by the independence of $\{I_i\}$ for $i \geq 1$, the algorithm for the immigration component just follows the classic EL algorithm for $i.i.d.$ data. In the partially observed BPI case, we develop the algorithm to calculate both the point and interval estimator for (m, σ^2) and (λ, b^2).

4.1 Algorithm

We recall that a BPI is defined by the iteration

$$X_n = \sum_{j=1}^{X_{n-1}} \xi_{n,j} + I_n, \quad n = 1, 2, ...,$$

where $E(\xi_{n,j}) = m,\ Var(\xi_{n,j}) = \sigma^2$, $E(I_n) = \lambda,\ Var(I_n) = b^2$. Let $\theta = (m, \sigma^2)$. The objective function to maximize is $\prod_{i=1}^{n} nw_i$, subject to $w_i \geq 0$, $\sum\limits_{i=1}^{n} w_i = 1$ and $\sum\limits_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0$, where $\boldsymbol{g}(X_i, \theta) = (X_i - mX_{i-1} - \lambda, U_i^2 - \sigma^2 X_{i-1} - b^2)$, and $U_i = X_i - mX_{i-1} - \lambda$. The weights w_i for $1 \leq i \leq n$ can be expressed as $w_i = \dfrac{1}{n}\dfrac{1}{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)}$ where $\boldsymbol{\nu_2}$ is Lagrangian multiplier satisfying

$$\sum_{i=1}^{n} \frac{\boldsymbol{g}(X_i, \theta)}{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)} = 0.$$

Then the problem of maximizing

$$\mathcal{R}(\theta) = \prod_{i=1}^{n} \frac{1}{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)}$$

can be reformulated as minimizing

$$f(\boldsymbol{\nu_2}) = -\sum_{i=1}^{n} log\left(1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)\right),$$

where the gradient and Hessian are denoted as $\nabla f(\boldsymbol{\nu_2})$ and $\nabla^2 f(\boldsymbol{\nu_2})$.

A new algorithm involves combination of Newton iteration with backtracking as discussed in Owen (2013), and modified for BPI data to find the maximum empirical likelihood estimator (MELE) for mean and variance jointly for both offspring and immigration distribution. We begin with obtaining the MELE for (m, σ^2), then we use the updated $(\hat{m}_{n,EL}, \hat{\sigma}^2_{n,EL})$ to get the MELE for (λ, b^2). Due to the independence of the offspring $\{X_n\}$ and immigration $\{I_n\}$ for $n = 1, 2, \cdots$, and the independence of I_i for $1 \le i \le n$, $(\hat{\lambda}_{n,EL}, \hat{b}^2_{n,EL})$ can be calculated based on the algorithm for *i.i.d.* data. The tolerance which is set to be $\epsilon = 10^{-4}$, is for the convenience of computation. As we discussed before, given $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$, the proposed Newton method with backtracking algorithm for minimizing $f(\boldsymbol{\nu_2})$ can be described as follows:

Step 0. Set $t = 1$, $\boldsymbol{\nu_2}' = \mathbf{0}'$, $\epsilon = 10^{-4}$, and $k = 0$. Calculate $\boldsymbol{g}(X_i, \hat{\theta}) = (\hat{U}_i, \hat{V}_i)$ where $\hat{U}_i = X_i - \hat{m}X_{i-1} - \hat{\lambda}$, and $\hat{V}_i = \hat{U}_i^{\,2} - \hat{\sigma}^2 X_{i-1} - \hat{b}^2$, and $\hat{m}, \hat{\sigma}^2, \hat{\lambda}$ and $\hat{b}^2$ can be any estimators in the literature.

Step 1. Calculate Newton step $\Delta\boldsymbol{\nu_2} = -\left(\nabla^2 f(\boldsymbol{\nu_2})\right)^{-1}\nabla f(\boldsymbol{\nu_2})$ and $\hat{m}^{(k)}_{n,EL}$. If $\left|\frac{\hat{m}^{(k)}_{n,EL}}{\hat{m}^{(k-1)}_{n,EL}} - 1\right| <$ ϵ, stops the algorithm and report $\boldsymbol{\nu_2^{(k)}}$ and $\hat{m}^{(k)}_{n,EL}$; otherwise go to Step 2.

Step 2. Backtracking Line Search: Given $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$, updates t by βt while $f\left(\boldsymbol{\nu_2^{(k)}} + t\Delta\boldsymbol{\nu_2^{(k)}}\right) \geq f\left(\boldsymbol{\nu_2^{(k)}}\right) + \alpha t \nabla f(\boldsymbol{\nu_2^{(k)}})\Delta\boldsymbol{\nu_2^{(k)}}$.

Step 3. Set $\boldsymbol{\nu_2^{(k+1)}} = \boldsymbol{\nu_2^{(k)}} + t\Delta\boldsymbol{\nu_2^{(k)}}$, $k = k + 1$. Go to Step 1.

Now we turn to the algorithm for constructing the EL confidence regions for (m, σ^2) and (λ, b^2). Based on a multivariate Newton-Raphson algorithm discussed in Hall and La Scala (1990), we modify the the algorithm by considering a multivariate Newton iteration with backtracking for BPI data. We reformulate the Lagrangian function for the endpoints of $\sum_{i=1}^n w_i \boldsymbol{g}(X_i, \theta)$ subject to $\sum_{i=1}^n w_i = 1$ for each $w_i \geq 0$ and $-2\sum_{i=1}^n log(nw_i) = c$, where c is the critical value from a chi-square distribution or a functional of Feller diffusion. We write the system of nonlinear equations as $\boldsymbol{h(s)} = (h_1(s), h_2(s), h_3(s)) = 0$, where

$$
\begin{aligned}
h_1 &= \sum_{i=1}^n \frac{\hat{U}_i}{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)} = \sum_{i=1}^n \frac{\hat{U}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i} \\
h_2 &= \sum_{i=1}^n \frac{\hat{V}_i}{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)} = \sum_{i=1}^n \frac{\hat{V}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i} \\
h_3 &= \sum_{i=1}^n \log\{1 + \boldsymbol{\nu_2}'\boldsymbol{g}(X_i, \theta)\} - \frac{1}{2}c = \sum_{i=1}^n \log\{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i\} - \frac{1}{2}c,
\end{aligned}
$$

where $\hat{U}_i$ and $\hat{V}_i$ are defined as above. Let $u \in [0, 2\pi)$. Then find the two sets of solutions $\boldsymbol{s_1} = (\nu^{(1)}, m_1, \sigma_1^2)$ and $\boldsymbol{s_2} = (\nu^{(2)}, m_2, \sigma_2^2)$ of $\boldsymbol{h(s)}$ and take $(\hat{m}_{L,EL}, \hat{\sigma}^2_{L,EL})$ and $(\hat{m}_{U,EL}, \hat{\sigma}^2_{U,EL})$ to be the smallest and largest respectively of (m_1, σ_1^2) and (m_2, σ_2^2).

Step 0. Set $t = 1$, $\epsilon = 10^{-4}$, and $k = 0$. Calculate $\boldsymbol{g}(X_i, \hat{\theta}) = (\hat{U}_i, \hat{V}_i)$ where $\hat{U}_i =$

$X_i - \hat{m}X_{i-1} - \hat{\lambda}$, and $\hat{V}_i = \hat{U}_i^{\,2} - \hat{\sigma}^2 X_{i-1} - \hat{b}^2$, and $\hat{m}, \hat{\sigma}^2, \hat{\lambda}$ and $\hat{b}^2$ can be any estimators in the literature.

Step 1. Calculate Jacobian matrix $J = (J_{ik})$ where $J_{ik} = \dfrac{\partial h_i}{\partial s_k}$ for $i = 1, 2, 3$ and $k = 1, 2, 3$, and Newton step $\Delta(\boldsymbol{s}) = -J^{-1}\boldsymbol{h(s)}$. If $\left| \frac{\hat{m}_{n,EL}^{(k)}}{\hat{m}_{n,EL}^{(k-1)}} - 1 \right| < \epsilon$, stops the algorithm and report $\Delta(\boldsymbol{s})$; otherwise go to Step 2.

Step 2. Backtracking Line Search: Updates t by $\frac{1}{2}t$ while while $\|\boldsymbol{h(s^{(k)} + t\Delta s^{(k)})}\| \geq \left(1 - \dfrac{t}{2}\right) \|\boldsymbol{h(s^{(k)})}\|$.

Step 3. Set $\boldsymbol{s_2^{(k+1)}} = \boldsymbol{s^{(k)}} + t\Delta \boldsymbol{s^{(k)}}$, $k = k + 1$. Go to Step 1.

After we obtain the confidence endpoints for (m, σ^2), we can use the updated $(\hat{m}_{L,EL}, \hat{\sigma}^2_{L,EL})$ and $(\hat{m}_{U,EL}, \hat{\sigma}^2_{U,EL})$ to obtain the confidence endpoints for (λ, b^2). Now, the problem turns to finding two sets of solutions $\boldsymbol{s_1'} = (\nu'^{(1)}, \lambda_1, b_1^2)$ and $\boldsymbol{s_2'} = (\nu'^{(2)}, \lambda_2, b_2^2)$ of $\boldsymbol{h'(s')}$, $\boldsymbol{h'(s')} = (h_1'(s'), h_2'(s'), h_3'(s')) = 0$, where

$$
\begin{aligned}
h_1' &= \sum_{i=1}^{n} \frac{\hat{U}_i}{1 + (\boldsymbol{\nu_2}')'\boldsymbol{g}(X_i, \theta)} = \sum_{i=1}^{n} \frac{\hat{U}_i}{1 + \nu' sin(u')\hat{U}_i + \nu' cos(u')\hat{V}_i} \\
h_2' &= \sum_{i=1}^{n} \frac{\hat{V}_i}{1 + (\boldsymbol{\nu_2}')'\boldsymbol{g}(X_i, \theta)} = \sum_{i=1}^{n} \frac{\hat{V}_i}{1 + \nu' sin(u')\hat{U}_i + \nu' cos(u')\hat{V}_i} \\
h_3' &= \sum_{i=1}^{n} \log\{1 + (\boldsymbol{\nu_2}')'\boldsymbol{g}(X_i, \theta)\} - \frac{1}{2}c = \sum_{i=1}^{n} \log\{1 + \nu' sin(u')\hat{U}_i + \nu' cos(u')\hat{V}_i\} - \frac{1}{2}c,
\end{aligned}
$$

where $\hat{U}_i = X_i - \hat{m}_{n,EL}X_{i-1} - \hat{\lambda}$ and $\hat{V}_i = \hat{U}_i^2 - \hat{\sigma}^2_{n,EL}X_{i-1} - \hat{b}^2$. Let $u' \in [0, 2\pi)$. Then take $(\hat{\lambda}_{L,EL}, \hat{b}^2_{L,EL})$ and $(\hat{\lambda}_{U,EL}, \hat{b}^2_{U,EL})$ to be the smallest and largest respectively of (λ_1, b_1^2) and (λ_2, b_2^2).

4.2 Simulation Results

The results in this chapter are based on 2000 simulations of 25 generations for supercritical BPI, critical BPI and subcritical BPI respectively. We generate supercritical BPI data with true offspring distribution Poisson(λ) with $\lambda = 1.6$ and true immigration distribution Poisson(λ) with $\lambda = 0.3$; critical BPI data with true offspring distribution Poisson(λ) with $\lambda = 1$ and true immigration distribution Poisson(λ) with $\lambda = 1.6$, and subcritical BPI data with true offspring distribution Poisson(λ) with $\lambda = 0.6$ and true immigration distribution Poisson(λ) with $\lambda = 2$. For supercritical BPI, we only provide the simulation results for the offspring parameters because of the dominated contribution of the population size of the offspring distribution. In the non-supercritical cases, we exhibit joint point estimation and confidence regions for mean and variance for both offspring and immigration. Tables for the point estimation give comparisons for (m, σ^2) and (λ, b^2) among the proposed empirical likelihood method and other methodologies studied in the literature. In these tables, the first column lists the names of these estimators; the second column and the third column give the averaged value and the standard error of the listed estimators based on 2000 simulations; the fourth and fifth column record the minimum and the maximum values of these estimators among the 2000 simulations. We also provide the tables for comparing the 95% confidence regions of (m, σ^2) and (λ, b^2) obtained by the empirical likelihood method and other methodologies in the literature. For the non-critical BPI, except for the EL estimator, the confidence intervals are calibrated by a normal distribution, while the interval endpoints for EL are obtained by the Newton iteration with backtracking method. In the critical BPI, because the asymptotic distribution of EL follows the functionals of Feller-diffusion, we only give the EL interval endpoints based on the Newton iteration with backtracking algorithm.

We begin with the simulation experiments for non-critical BPI, then we move to the critical BPI. For the supercritical case, due to the dominated contribution of the offspring to the population size, we only provide simulations results for the offspring distribution; while for the subcritical case, we list the simulation results for both offspring and immigration distributions.

4.2.1 Supercritical BPI

For the supercritical BPI, we compare the estimators that are discussed in Chapter 1 to the joint empirical likelihood estimator for (m, σ^2). These estimators include conditional least squares estimator (CLS, $\hat{m}_n$ in (1.8)), conditional weighed lease squares estimator (CWLS, $\tilde{m}_n$ in (1.10)), maximum likelihood estimator (MLE, $\tilde{m} = \sum_{i=1}^{n} X_i / \sum_{i=1}^{n} X_{i-1}$), ratio estimator (Ratio, $\hat{m}_n = X_n/X_{n-1}$) for offspring mean; and conditional least squares estimator (CLS, (1.13)), conditional weighed lease squares estimator (CWLS, $\tilde{\sigma}^2$ in Section 1.1.2) and Heyde's estimator (Heyde, (1.12)) for offspring variance.

Table 4.1 and Table 4.2 give the point estimator for offspring mean and variance. We compare these estimators by looking at averaged estimator (Mean), empirical standard deviation (Std Dev), mean squared error (MSE), bias, and the minimum and maximum values based on 2000 simulations. We can see that for offspring mean estimators, all the methodologies listed perform similarly. For offspring variance estimators, CLS estimator has more fluctuation compared to other estimators. Table 4.3 compare the lower bound for mean estimators (LBmean) and variance estimators (LBvar), upper bound for mean estimators (UBmean) and variance estimators (UBvar), average interval length for mean estimators (Avg.Lmean) and variance estimators (Avg.Lvar), and coverage rate for mean (CRmean) and variance (CRvar). From Table 4.3 we can see that for the supercritical BPI, EL method has advantages for obtaining interval endpoints and it has better estimated coverage rate. Compared to the asymptotic EL confidence interval for variance, the interval calculated by the backtracking algorithm from EL method has slightly lower coverage rate and tighter interval length.

Figure 4.1 gives the confidence region of offspring mean and variance for supercritical BPI. For offspring mean concentrate around its true value 1.6, whereas for offspring variance, it tends to underestimate the true value.

Supercritical BPI

Offspring Mean Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	1.5998	0.0056	0.000031	-0.0002	1.5292	1.6793
CWLS	1.5997	0.0046	0.000021	-0.0003	1.5254	1.6245
MLE	1.5995	0.0053	0.000028	-0.0005	1.5164	1.6241
Ratio	1.5996	0.0076	0.000058	-0.0004	1.4971	1.6769
EL	1.5997	0.0065	0.000042	-0.0003	1.5154	1.7124

Table 4.1: Point Estimation for m of Supercritical BPI

Supercritical BPI

Offspring Variance Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	1.2959	1.2758	1.7193	-0.3041	0.0080	12.1838
CWLS	1.5425	0.5045	0.2577	-0.0575	0.3399	4.2405
Heyde	1.6558	0.2790	0.0810	0.0558	0.4793	4.9340
EL	1.5023	0.2210	0.0584	-0.0977	0.2581	1.8796

Table 4.2: Point Estimation for σ^2 of Supercritical BPI

Supercritical BPI

95% Confidence Interval for Offspring Mean

Estimator	LBmean	UBmean	Avg.Lmean	CRmean
CLS	1.1982	2.0013	0.8031	0.990
CWLS	1.1193	2.0801	0.9608	0.990
Asy.EL	1.1211	2.0783	0.9573	0.945
EL	1.1541	2.0543	0.9001	0.929

95% Confidence Interval for Offspring Variance

Estimator	LBvar	UBvar	Avg.Lvar	CRvar
CLS	0.5775	2.0144	1.4369	0.397
CWLS	0.6874	2.3976	1.7103	0.855
Heyde	0.7379	2.5738	1.8359	0.904
Asy.EL	0.6694	2.3351	1.6656	0.945
EL	0.9654	2.5186	1.5537	0.929

Table 4.3: 95% Confidence Region for (m, σ^2) of Supercritical BPI

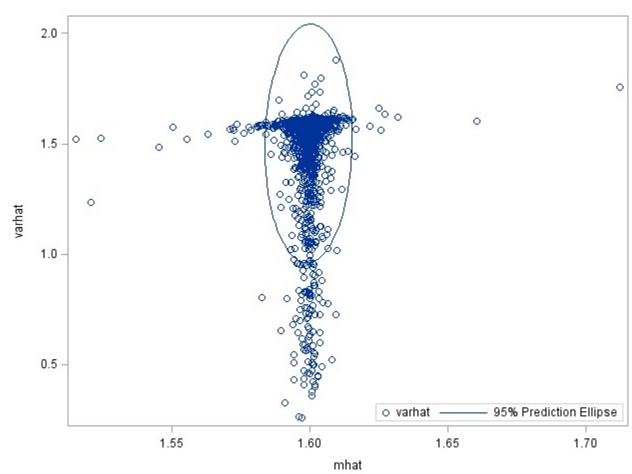

Figure 4.1: Confidence Region for (m, σ^2) of Supercritical BPI

4.2.2 Subcritical BPI

For the subcritical BPI, we obtain the joint EL estimator when we have observed immigration (EL.obs) and unobserved immigration (EL). We compare the joint empirical likelihood estimator to CLS, CWLS, MLE, Quine's estimator, Sriram et al's estimator (SBH), Sriram and Vidyashankar's estimator (SV) for offspring mean estimation, and we also compare it to CLS, CWLS, MLE, Sriram et al's estimator (SBH), Sriram and Vidyashankar's estimator (SV), and Yanev and Tchoukova-Dantcheva's estimator (YTD) for offspring variance estimation. For the immigration parameters, we compare our proposed empirical likelihood estimator to CLS, CWLS and Quine's estimator for immigration mean, and to CLS, CWLS and YTD for immigration variance. Sriram et al's estimator (SBH) and Sriram and Vidyashankar's estimator (SV) for offspring mean and variance are obtained by assuming that immigration is observable. Sriram and Vidyashankar (2000)'s variance estimator (SV) can be expressed as

$$\hat{\sigma}_n^2 = \left(\sum_{i=1}^{n} \frac{X_{i-1}}{(1+X_{i-1})^2} \right)^{-1} \left(\sum_{i=1}^{n} \frac{X_i - \hat{m}_n X_{i-1} - I_i}{(1+X_{i-1})^2} \right), \text{ where } \hat{m}_n = \frac{\sum_{i=1}^{n}(X_i - I_i)}{\sum_{i=1}^{n} X_{i-1}}.$$

Sriram and Vidyashankar's estimator is defined as

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n}(X_i - \hat{m}_n X_{i-1} - I_i)}{\sum_{i=1}^{n} X_{i-1}},$$

where $\hat{m}_n$ is same defined as above. Notice that, when there is no immigration component, namely, $I_i = 0$ for $1 \leq i \leq n$, the above $\hat{m}_n$ just reduces to the MLE. YTD's estimator for offspring and immigration variances have the same expressions as the CLS estimators for offspring and immigration variances, and in YTD's estimators, they use the estimated offspring and immigration mean defined in Heyde and Seneta (1972).

Table 4.4 - 4.5 and Table 4.6 - 4.7 provide the point estimation comparisons among

proposed estimator and all the others estimators studied in the literature, for offspring and immigration distributions respectively. In general, the estimators (EL.obs, SBH, SV) obtained under the assumption that the immigration is observed perform better than those obtained with unobserved immigration. The empirical likelihood estimator has the smallest bias in both observed immigration and unobserved immigration cases. In the unobserved immigration case, the empirical likelihood estimator has remarkable better performances than thoes immigration variance estimators, which have large variations, studied in literature. CLS, CWLS and YTD have large variations. Table 4.8 - 4.9 compare the confidence endpoints, average interval lengths and coverage rates for the listed metrologies. In both observed and unobserved immigration cases, the empirical likelihood estimator produces better results. Figure 4.2 - 4.3 show the confidence regions for offspring and immigration respectively with observed immigration, while Figure 4.4 - 4.5 are the graphs with unobserved immigration. When the immigration component is not observable, the joint estimators for both offspring and immigration tend to spread more. This is due to the limited information for the immigration that makes the inference becomes more challenging, and cause more fluctuations.

Subcritical BPI

Offspring Mean Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
MLE	0.5984	0.0932	0.0087	-0.0016	0.2727	0.9202
SBH/SV	0.5802	0.0803	0.0068	-0.0198	0.2740	0.8481
EL.obs	0.6112	0.1395	0.0196	0.0112	0.1490	1.4319
CLS	0.4012	0.1873	0.0746	-0.1988	-0.3442	0.9766
CWLS	0.4250	0.2004	0.0708	-0.1750	-0.3774	0.9409
Quine	0.5395	0.2031	0.0449	-0.0605	-0.3044	1.2783
EL	0.5692	0.0982	0.0106	-0.0308	0.0475	1.3086

Table 4.4: Point Estimation for m of Subcritical BPI

Subcritical BPI

Offspring Variance Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
SBH	0.5664	0.2201	0.0496	-0.0336	0.1848	1.8504
SV	0.5350	0.1853	0.0386	-0.0650	0.1636	1.3763
EL.obs	0.6115	0.1411	0.0200	0.0115	0.1490	1.5430
CLS	0.7153	0.7194	0.5308	0.1153	-1.3474	3.4361
CWLS	0.8138	0.4046	0.2094	0.2138	-5.1709	9.9219
YTD	0.9586	0.7335	0.6667	0.3586	-1.3288	4.0336
EL	0.5698	0.0981	0.0105	-0.0302	0.0079	3.0312

Table 4.5: Point Estimation for σ^2 of Subcritical BPI

Subcritical BPI

Immigration Mean Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	2.8716	0.8697	1.5160	0.8716	0.3750	7.4275
CWLS	2.7525	0.8826	1.3452	0.7525	0.4180	7.1719
Quine	2.5789	0.7693	0.9269	0.5789	0.6628	5.7051
EL	2.0036	0.3987	0.1590	0.0036	0.7496	5.4172
EL.obs	1.9981	0.3423	0.1172	-0.0018	0.0225	5.4748

Table 4.6: Point Estimation for λ of Subcritical BPI

Immigration Variance Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	4.0508	1.4842	6.4088	2.0508	0.7010	12.4898
CWLS	-0.5142	2.7313	13.781	-2.5142	-88.9520	67.0352
YTD	4.0868	1.5005	6.6062	2.0868	0.7110	12.7202
EL	1.9399	1.2536	1.5751	-0.0601	0.3824	6.3188
EL.obs	2.0098	0.3876	0.1503	0.0098	0.0215	5.4362

Table 4.7: Point Estimation for b^2 of Subcritical BPI

Subcritical BPI

95% Confidence Interval for Offspring Mean

Estimator	LBmean	UBmean	Avg.Lmean	CRmean
SBH/SV	0.2902	0.8702	0.5801	0.989
EL.obs	0.3913	0.9124	0.5211	0.908
Asy.EL.obs	0.3066	0.9158	0.6092	0.959
CLS	0.0626	0.7440	0.6814	0.628
CWLS	0.0769	0.7727	0.6958	0.794
EL	0.2868	3.1914	2.9045	0.957
Asy.EL	0.2745	0.8638	0.5893	0.962

95% Confidence Interval for Offspring Variance

Estimator	LBvar	UBvar	Avg.Lvar	CRvar
SBH	0.2524	0.8803	0.6279	0.804
SV	0.2384	0.8316	0.5652	0.780
EL.obs	0.2741	0.9501	0.6760	0.908
Asy.EL.obs	0.2725	0.9505	0.6780	0.959
CLS	0.3187	1.1118	0.7931	0.479
CWLS	0.3627	1.2650	0.9023	0.928
YTD	0.4272	1.4901	1.0629	0.525
EL	0.2539	0.8857	0.6318	0.957
Asy.EL	0.2604	0.9082	0.6478	0.962

Table 4.8: 95% Confidence Region for (m, σ^2) of Subcritical BPI

Subcritical BPI

95% Confidence Interval for Immigration Mean

Estimator	LBmean	UBmean	Avg.Lmean	CRmean
EL.obs	1.4436	2.5491	1.1054	0.913
Asy.EL.obs	1.4453	2.5510	1.1056	0.914
CLS	2.0950	3.6482	1.5532	0.485
CWLS	2.3187	2.6196	0.3009	0.335
EL	0.2795	2.5359	2.2565	0.751
Asy.EL	0.2636	0.8454	0.5818	0.840

95% Confidence Interval for Immigration Variance

Estimator	LBvar	UBvar	Avg.Lvar	CRvar
EL.obs	0.9613	3.1086	2.1473	0.913
Asy.EL.obs	0.8840	3.1123	2.2284	0.914
CLS	1.8052	6.2965	4.4913	0.681
CWLS	1.0751	5.0822	4.0071	0.603
YTD	1.8212	6.3524	4.5312	0.672
EL	1.4064	3.8042	2.3978	0.751
Asy.EL	1.0920	3.8091	2.7170	0.840

Table 4.9: 95% Confidence Region for (λ, b^2) of Subcritical BPI

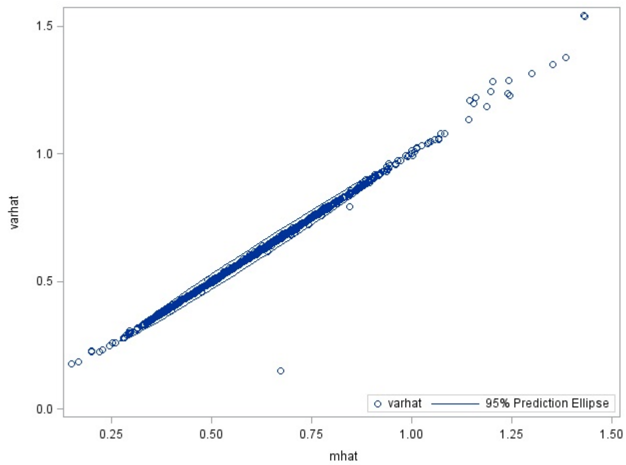

Figure 4.2: Confidence Region for (m, σ^2) with Observed Immigration

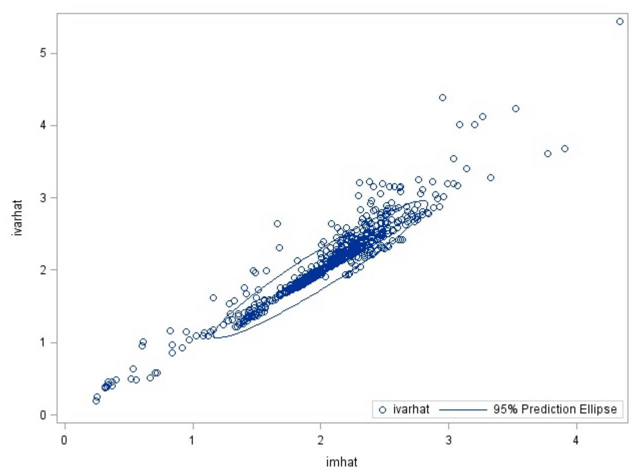

Figure 4.3: Confidence Region for (λ, b^2) with Observed Immigration

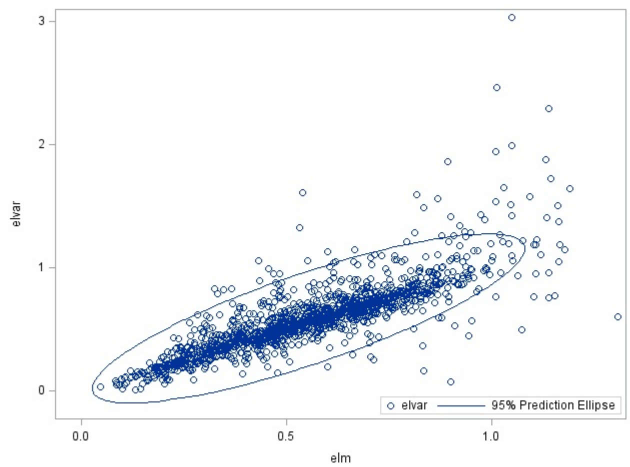

Figure 4.4: Confidence Region for (m, σ^2) with Unobserved Immigration

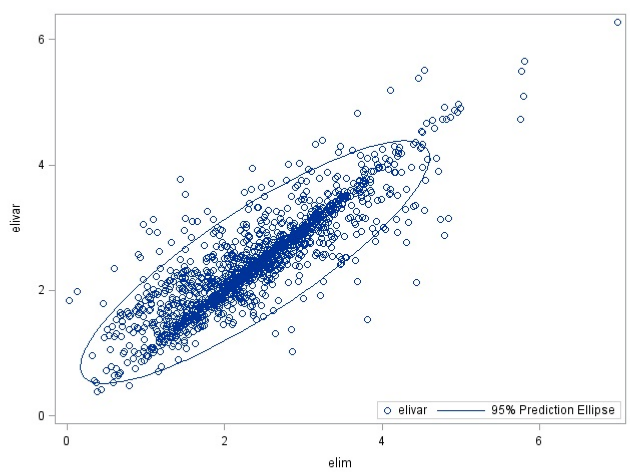

Figure 4.5: Confidence Region for (λ, b^2) with Unobserved Immigration

4.2.3 Critical BPI

Now we move the simulation studies for critical BPI. As we proved in Chapter 3, we use a functional of Feller-diffusion to calibrate the $-2log\mathcal{R}_n(\theta)$. Due to this reason, we need some extra work to generate the Feller-diffusions and find their critical value when we construct the confidence region for (m, σ^2). Recall Theorem 4 in Chapter 3, when $m = 1$ and $2\lambda > \sigma^2$,

$$-2 \log \mathcal{R}_n(\theta) \xrightarrow{d} D_1^2 + D_2^2,$$

and

$$D_1^2 = \frac{(Y(1) - \lambda)^2}{\sigma^2 \int_0^1 Y(t) dt} \quad \text{and} \quad D_2^2 = \frac{(\int_0^1 Y(t) dB(t))^2}{\int_0^1 Y^2(t) dt},$$

where $B(t)$ is a Brownian motion. We write the integral functionals of the diffusion diffusions as Riemann sum, which are

$$\begin{aligned} \int_0^1 Y(t)dt &= \lim \frac{1}{n} \sum_{i=1}^{N} X_{[nt_i]}(t_{i+1} - t_i), \\ \int_0^1 Y^2(t)dt &= \lim \frac{1}{n^2} \sum_{i=1}^{N} X_{[nt_i]}^2(t_{i+1} - t_i), \\ \int_0^1 Y(t)dB(t) &= \lim \frac{1}{n} \sum_{i=1}^{N} X_{[nt_i]}(B(t_{i+1}) - B(t_i)), \end{aligned}$$

where N is the partition, t_i's are random numbers generated from $[0, 1]$. We set N to be 30 and generate 500 generations critical BPI to calculate $D_1^2 + D_2^2$, and repeat the experiment 1000 times to find the critical value from a functional of Feller-diffusion.

For the point estimation, we compare the joint empirical likelihood estimator to CLS, CWLS, and MLE for offspring mean, and CLS and CWLS for offspring variance. In Table 4.10 - 4.11, EL and EL.obs estimators perform similarly to MLE. For offspring variance estimators, EL EL and EL.obs estimators are better than CLS and CWLS that can get negative estimates for variance. Table 4.12 - 4.13 compare the joint empirical likelihood estimator to CLS, CWLS and MLE for immigration mean, and to CLS and CWLS for immigration variance. For immigration mean estimators, EL.obs estimator produces the most accurate result. For immigration variance, we only consider EL.obs estimator due to the slow rate of convergence in the critical BPI case. From Table 4.14 we can see that EL and EL.obs have advantages in interval estimation and obtaining better coverage rates for immigration mean. Table 4.14 indicate that normal distribution can not calibrate the asymptotic distribution in critical BPI for CLS and CWLS.

Critical BPI

Offspring Mean Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	0.8885	0.1429	0.0329	-0.1115	0.1722	1.2511
CWLS	0.9192	0.1316	0.0238	-0.0808	0.1057	1.1947
MLE	0.9741	0.0687	0.0054	-0.0259	0.5977	1.1274
EL	1.0658	0.0624	0.0082	0.0658	0.4945	2.3162
EL.obs	0.9743	0.0763	0.0065	-0.0257	0.6002	1.400

Table 4.10: Point Estimation for m of Critical BPI

Critical BPI

Offspring Variance Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	0.7848	0.6015	0.4082	-0.2152	-1.8568	4.6294
CWLS	0.6180	9.7843	95.8786	-0.3814	-375.2042	114.6430
EL	0.9931	0.1422	0.0203	-0.0069	0.2377	2.9000
EL.obs	0.9738	0.0818	0.0074	-0.0262	0.3667	1.4585

Table 4.11: Point Estimation for σ^2 of Critical BPI

Critical BPI

Immigration Mean Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	3.2741	1.7826	5.9800	1.6741	-0.7371	12.5887
CWLS	2.5117	1.1682	2.1958	0.9117	0.2671	9.3386
MLE	1.6436	0.3481	0.1231	0.0436	0.3200	5.1600
EL	1.8897	0.7212	0.6040	0.2897	0.0220	7.9880
EL.obs	1.5842	0.3967	0.1576	-0.0158	0.1726	3.8353

Table 4.12: Point Estimation for λ of Critical BPI

Critical BPI

Immigration Variance Estimators

Estimator	Mean	Std Dev	MSE	Bias	Min	Max
CLS	19.4160	14.8784	538.7744	17.8160	1.8499	128.5558
CWLS	3.2293	1.8746	6.1688	1.6293	-8.9813	9.4500
EL.obs	1.5959	0.4167	0.1737	-0.0041	0.1751	4.2418

Table 4.13: Point Estimation for b^2 of Critical BPI

Critical BPI

95% Confidence Interval for Offspring Mean

Estimator	LBmean	UBmean	Avg.Lmean	CRmean
CLS	0.5558	1.2290	0.6732	0.868
CWLS	0.5255	1.30540	0.7799	0.815
EL	0.6926	1.4090	0.7165	0.857
EL.obs	0.4816	1.3480	0.8657	0.948

95% Confidence Interval for Offspring Variance

Estimator	LBvar	UBvar	Avg.Lvar	CRvar
CLS	0.3497	1.2198	0.8701	0.509
CWLS	0.2754	0.9606	0.6852	0.523
EL	0.5726	2.9239	2.3512	0.857
EL.obs	0.4372	1.5054	1.0702	0.948

Table 4.14: Confidence Region for m and σ^2 of Critical BPI

Critical BPI

95% Confidence Interval for Immigration Mean

Estimator	LBmean	UBmean	Avg.Lmean	CRmean
CLS	1.6460	4.9021	3.2561	0.642
CWLS	1.4955	3.3911	1.8956	0.618
EL	0.5012	2.7577	2.2565	0.751
EL.obs	0.6565	2.0745	1.4180	0.907

95% Confidence Interval for Immigration Variance

Estimator	LBvar	UBvar	Avg.Lvar	CRvar
CLS	8.6523	30.1796	21.5273	0.018
CWLS	1.4391	5.0196	3.5805	0.135
EL.obs	0.6798	2.4345	1.7547	0.907

Table 4.15: Confidence Region for (λ, b^2) of Critical BPI

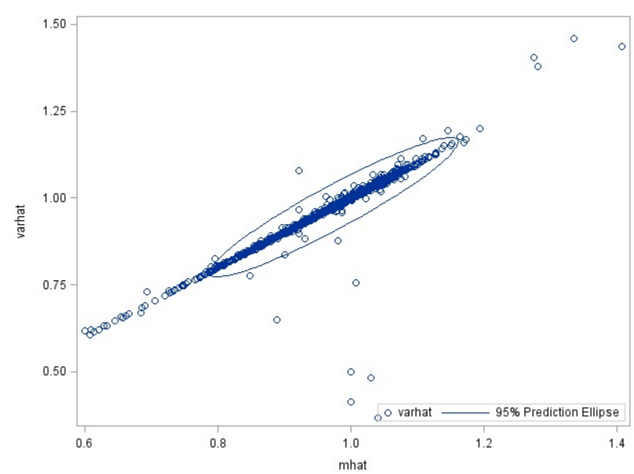

Figure 4.6: Confidence Region for Offspring with Observed Immigration

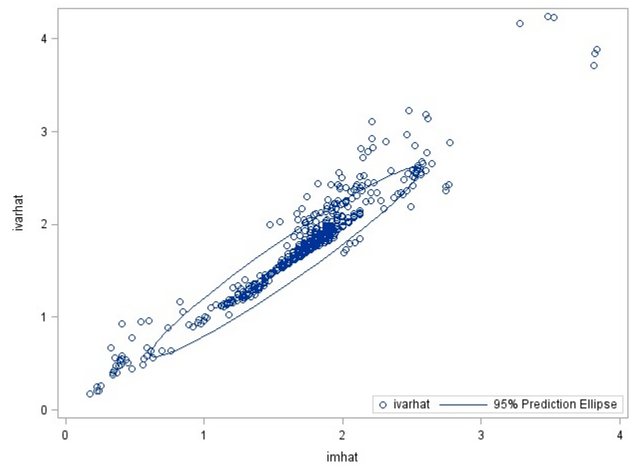

Figure 4.7: Confidence Region for Immigration with Observed Immigration

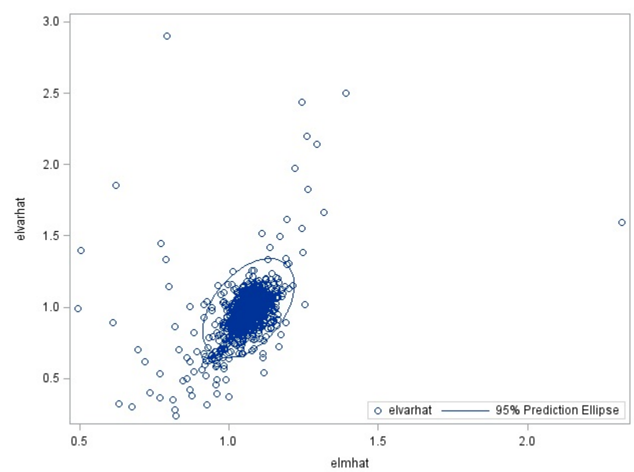

Figure 4.8: Confidence Region for Offspring with Unobserved Immigration

From the simulations, it is clear that the proposed empirical likelihood method yields joint point estimates with smaller bias. Moreover, compared to the traditional methods, the confidence regions for both offspring and immigration of branching processes are tighter and with higher coverage rates, and they are also supported by the asymptotic justification.

4.3 Data Analysis

The data analyzed in this section is from the transmission line outage data. Because of the lineage structure of the cascading failure, it allows us to use branching processes to model the propagation in the electric power system. Our goal is to apply the proposed EL method for the cascades data to obtain the joint point estimation and confidence region for the average amount of propagation of the line outages. The data is collected from a regional electric power transmission system with about 100 buses, 180 lines at 220 kV and 20 lines at 500 kV from 1997 to 2006. For the convenience of data analysis, the line outages are grouped into different cascades, then for each cascade, different stages are also grouped. The following table records the number of outages in each stage over 226 cascades. It begins with 296 outages at stage 0 of a cascade, then turn to 45 in stage 1 of a cascade and continues until stage 15 (see Ren and Dobson (2008)).

Cascades Data

Number of Outage in Each Stage

Stage Num.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Outage Num.	296	45	18	14	10	3	1	1	1	1	1	1	2	1	1	0

BP is popular in modeling cascading failures propagation in the electric power system (see Dobson, Carreras, Lynch, and Newman (2007), Dobson, Carreras, and Newman (2004a), Dobson, Carreras, and Newman (2004b), Dobson, Carreras, Lynch, Nkei, and Newman (2005), Dobson, Carreras, and Newman (2005)). A Poisson distribution with mean λ is the parametric model is assumed to describe the offspring distribution, and it has been extensively studied in Consul and Shoukri (1988) The traditional analysis method is to consider a mixture of Borel-Tanner distribution. Let Y be the total number of failures given an initial failures distribution $P(Z_0 = z_0)$, then

$$P(Y = y) = \sum_{z_0=1}^{y} P(Z_0 = z_0) z_0 m (ym)^{y-z_0-1} \frac{e^{-ym}}{(y-z_0)!},$$

where m is the average tendency for the cascade to propagate. Let $Z_j^{(k)}$ be the outage number at stage j for cascade k, then m can be estimated by MLE, which is

$$m = \frac{\sum_{k=1}^{K} Z_1^{(k)} + Z_2^{(k)} + \cdots}{\sum_{k=1}^{K} Z_0^{(k)} + Z_1^{(k)} + \cdots}.$$

Now, we apply our proposed empirical likelihood methodology to analyze the cascades data.

The above table gives the joint point estimation and confidence region for average tendency for the cascade to propagate. In this data, we do not consider the immigration component and we make inference only for the offspring distribution. The first two columns list the point estimation for mean and variance m in the cascade data. LBmean and UBmean give the lower bound and upper bound for the mean estimators of m, and LBvar and UBvar give the lower bound and upper bound for the variance estimators. Since it is a subcritical branching process, we compare our proposed join empirical likelihood estimator to other estimators for the subcritical case in the literature. The MLE for m obtained in Ren and Dobson (2008) is 0.252525, and they use the Poisson distribution to model the power outage propagation.

We apply our empirical likelihood method in this data, and compare it to CLS, CWLS, SBH and SV. For empirical likelihood, we consider the confidence region both obtained by using the Newton iteration with backtracking and normal calibration. As mentioned before, when there is no immigration, the mean estimator from SBH and SV reduce to MLE. EL produces similar mean estimation result to MLE, and its variance estimator is between SBH, CWLS, CLS and SV. EL is the only methodology produces positive confidence endpoints for both mean and variance, and that also show the advantages of the EL method which respect the range of target parameters.

Empirical Likelihood for Cascades Data

Empirical Likelihood for Cascades Data

Estimator	Mean	Var	LBmean	UBmean	LBvar	UBvar
EL	0.2540	1.4348	9.3123E-7	0.6271	0.4079	1.5992
Asy.EL	0.2540	1.4348	-0.3522	0.8601	0.4079	2.4618
CLS	0.1494	1.2795	-0.4230	0.7218	0.3638	2.1952
CWLS	0.2164	0.5520	-0.1596	0.5924	0.1569	0.9471
SBH	0.2525	0.8042	-0.2013	0.7064	0.2287	1.3798
SV	0.2525	2.6948	-0.5782	1.0833	0.7662	4.6234

Chapter 5: Discussion and Concluding Remarks

In this thesis, we proposed a new inferential methodology, namely, empirical likelihood for branching processes with immigration. We investigated the asymptotic behaviors of the joint estimator for (m, σ^2), and due to the trichotomy of branching processes with immigration, we concluded with dramatically different asymptotic distributions for non-critical branching processes with immigration and critical branching processes with immigration. We evaluated the behaviors of our methodology and demonstrated the advantages of our proposed methodology in the simulation studies and data analysis. We also provided an algorithm for obtaining the joint EL estimator and confidence regions for (m, σ^2) and (λ, b^2) based on Newton iteration with backtracking method, which has been discussed in Owen (2013) and Hall and La Scala (1990) for *i.i.d.* data.

Due to the different asymptotic distribution, this leads to a attempt of a unified empirical likelihood method for BPI by introducing a stopping time and using the sequential analysis method. Because in practice, we have no prior information about the value of m, and a unified framework would allow us to do inference without a given range of m. Also, in the subcritical BPI with unobserved immigration, the estimation and inference for (λ, b^2) could be addressed by sequential analysis method.

There are also some limitations in this research. The data for BPI are generated from Poisson distribution, there could be more distribution of BPI such as negative binomial, geometric distribution and other discrete probabilistic distributions.

Bibliography

Athreya, K. and P. Ney (1972). *Branching Process*. Wiley Online Library.

Basawa, I. and A. Vidyashankar (2003). Quasi likelihood inference for branching processes with immigration. *Journal of Indian Statistical Association*, 143–153.

Bhat, B. and S. Adke (1981). Maximum likelihood estimation for branching processes with immigration. *Advances in Applied Probability*, 498–509.

Boyd, S. P. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Consul, P. and M. Shoukri (1988). Some chance mechanisms related to a generalized poisson probability model. *American Journal of Mathematical and Management Sciences 8*(1-2), 181–202.

Dobson, I., B. A. Carreras, V. E. Lynch, and D. E. Newman (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science 17*(2), 026103.

Dobson, I., B. A. Carreras, V. E. Lynch, B. Nkei, and D. E. Newman (2005). Estimating failure propagation in models of cascading blackouts. *Probability in the Engineering and Informational Sciences 19*(4), 475.

Dobson, I., B. A. Carreras, and D. E. Newman (2004a). A branching process approximation to cascading load-dependent system failure. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pp. 10–pp. IEEE.

Dobson, I., B. A. Carreras, and D. E. Newman (2004b). Probabilistic load-dependent cascading failure with limited component interactions. In *Circuits and Systems, 2004.*

ISCAS'04. Proceedings of the 2004 International Symposium on, Volume 5, pp. V–912. IEEE.

Dobson, I., B. A. Carreras, and D. E. Newman (2005). A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences 19*(1), 15–32.

Hall, P. and B. La Scala (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review/Revue Internationale de Statistique*, 109–127.

Harris, T. E. (1948). Branching processes. *The Annals of Mathematical Statistics*, 474–494.

Heyde, C. (1976). On estimating the variance of the offspring distribution in a simple branching process. *Advances in Applied Probability*, 421–433.

Heyde, C. and Y. Lin (1992). On quasi-likelihood methods and estimation for bunching processes and heteroscedastic regression models. *Australian Journal of Statistics 34*(2), 199–206.

Heyde, C. and E. Seneta (1972). Estimation theory for growth and immigration rates in a multiplicative process. *Journal of Applied Probability*, 235–256.

Heyde, C. and E. Seneta (1974). Notes on "estimation theory for growth and immigration rates in a multiplicative process". *Journal of Applied Probability*, 572–577.

Jagers, P. (1975). *Branching processes with biological applications.* Wiley.

Kawazu, K. and S. Watanabe (1971). Branching processes with immigration and related limit theorems. *Theory of Probability & Its Applications 16*(1), 36–54.

Klimko, L. and P. Nelson (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics 6*(3), 629–642.

Kuelbs, J. and A. N. Vidyashankar (2011). Weak convergence results for multiple generations of a branching process. *Journal of Theoretical Probability 24*(2), 376–396.

Lindvall, T. (1972). Convergence of critical galton-watson branching processes. *Journal of Applied Probability*, 445–450.

Mellein, B. (1982). Diffusion limits of conditioned critical galton-watson processes. *Revista colombiana de matematicas 16*, 125–140.

Mellein, B. (1983). Kac functionals of diffusion processes approximating critical branching processes. *Boletín de la Sociedad Matemática Mexicana (2) 28*(2), 95–107.

Nagaev, S. (1981). On the asymptotic behavior of one-sided large deviation probabilities. *Theory Probab. Appl. 26*, 369–372.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

Owen, A. B. (2001). *Empirical likelihood*. CRC press.

Owen, A. B. (2013). Self-concordance for empirical likelihood. *Canadian Journal of Statistics 41*(3), 387–397.

Pakes, A. (1971). On the critical galton-watson process with immigration. *Journal of the Australian Mathematical Society 12*(04), 476–482.

Pakes, A. (1972). Further results on the critical galton-watson process with immigration. *Journal of the Australian Mathematical Society 13*(03), 277–290.

Qin, J. and J. Lawless (1994a). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 300–325.

Qin, J. and J. Lawless (1994b). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 300–325.

Quine, M. (1976). Asymptotic results for estimators in a subcritical branching process with immigration. *The Annals of Probability 4*(2), 319–325.

Ren, H. and I. Dobson (2008). Using transmission line outage data to estimate cascading failure propagation in an electric power system. *IEEE Trans. on Circuits and Systems 55*(9), 927–931.

Seneta, E. (1970). An explicit-limit theorem for the critical galton-watson process with immigration. *Journal of the Royal Statistical Society. B*, 149–152.

Sriram, T. and A. Vidyashankar (2000). Sequential point estimation for branching processes i, subcritical case: Sequential point estimation for branching processes-i. *Sequential Analysis 19*(3), 77–92.

Venkataraman, K. and K. Nanthi (1982). A limit theorem on a subcritical galton-watson process with immigration. *The Annals of Probability 10*(4), 1069–1074.

Wei, C. and J. Winnicki (1989). Some asymptotic results for the branching process with immigration. *Stochastic Processes and their Applications 31*(2), 261–282.

Wei, C. and J. Winnicki (1990). Estimation of the means in the branching process with immigration. *The Annals of Statistics*, 1757–1773.

Winnicki, J. (1991). Estimation of the variances in the branching process with immigration. *Probability Theory and Related Fields 88*(1), 77–106.

Yanev, N. and S. Tchoukova-Dantcheva (1986). Limit theorems for estimators of variances in a branching process with immigration. *Serdica Mathematical Journal*, 143–153.

Curriculum Vitae

Pin Ren graduated with a B.S. in Mathematical Statistics and a M.S. in Financial Mathematics and Financial Engineering from Shandong University, China. In 2009, she attended George Mason University to continue her PhD studies in Statistical Science.