Journal of Multivariate Analysis 100 (2009) 1155–1167

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Branching Markov processes and related asymptotics

S.Y. Hwang [a,*], I.V. Basawa [b]

[a] Sookmyung Women's University, Republic of Korea
[b] University of Georgia, USA

ARTICLE INFO

Article history:
Received 15 April 2008
Available online 8 November 2008

AMS 2000 subject classifications:
primary 62P10
secondary 62M99

Keywords:
Branching processes
Markov processes
Law of large numbers
Central limit theorem
Conditional exponential family
Least squares
Quasilikelihood and maximum likelihood estimation

ABSTRACT

Models for Markov processes indexed by a branching process are presented. The new class of models is referred to as the branching Markov process (BMP). The law of large numbers and a central limit theorem for the BMP are established. Bifurcating autoregressive processes (BAR) are special cases of the general BMP model discussed in the paper. Applications to parameter estimation are also presented.

1. Introduction

Branching processes have served as useful models for population growth. Galton–Watson (G–W) branching processes, in particular, have generated a vast literature on the theory and applications of models for population dynamics. See [1,2] for the classical theory and limit theorems for G–W branching processes. Guttorp [3] presents a comprehensive review of inference for branching processes.

If Z_t denotes the generation size of the tth generation, $t = 0, 1, \ldots$, with $Z_0 = 1$, the G–W branching process $\{Z_t\}$ has the representation

$$Z_t = \sum_{j=1}^{Z_{t-1}} \xi_{tj} \tag{1.1}$$

for $Z_{t-1} \geq 1$, where $\{\xi_{tj}, t = 1, 2, \ldots, j = 1, 2, \ldots\}$ are independent and identically distributed (iid) non-negative integer-valued random variables with $E(\xi_{tj}) = m$ and $\mathrm{Var}(\xi_{tj}) = \sigma^2$, where m and σ^2 are the offspring mean and variance respectively. The representation (1.1) shows that $\{Z_t\}$ is a Markov chain. Extensive literature is now available on the limiting behavior of $\{Z_t\}$ [2] and on estimation of m and σ^2 [3].

In many situations dealing with population biology, epidemiology, physics and chemistry, one may be interested in measuring some characteristics of interest on each individual of a G–W branching process such as life-time of a cell,

* Corresponding author.
E-mail address: shwang@sm.ac.kr (S.Y. Hwang).

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.10.014

presence (or absence) of a certain chemical of a protein, radio-activity level and severity of an epidemic (e.g. specific stage of AIDS), etc. Suppose $X_t(j)$ denotes the observation on the j-th individual of the t-th generation. We then have a tree-indexed process $\{Z_t, X_t(j)\}, j = 1, 2, \ldots, Z_t, t = 0, 1, 2, \ldots,$ assuming $Z_t \geq 1$. The main goal of this paper is to introduce models for the tree-indexed process which we shall refer to as a branching Markov process (BMP). For processes where each individual splits exactly into two offspring, we have $Z_t = 2^t$, and the process indexed by the deterministic binary tree, viz., $\{X_t(j)\}, j = 1, 2, \ldots, 2^t, t = 0, 1, 2, \ldots$, is known as a bifurcating process. Bifurcating autoregressive processes (BAR) for continuous data have been studied by various authors including [4–8], among others. Basawa and Zhou [9] discussed some preliminary examples of non-Gaussian bifurcating processes including models for count data.

In this paper, we present new models for the general branching–tree indexed processes which accommodate both continuous and count data for the observable characteristics $\{X_t(j)\}$. The branching Markov processes (BMP) are introduced in Section 2. Results on ergodicity and the law of large numbers are presented in Section 3. Section 4 is concerned with a central limit theorem. Some applications to parameter estimation are discussed in Section 5.

As to the statistical importance of the limit theorems derived in this paper, we note that the law of large numbers and the central limit theorem are useful in studying the asymptotic properties of the least squares, quasilikelihood, and the maximum likelihood estimators and related large sample tests. See Section 5 for details. See [10,5] for real data examples of BMP models indexed by a binary tree. For examples of the general BMP model see Section 2. Finally, it is to be noted that the branching Markov processes (BMP) discussed in this paper are quite different from (and unrelated to) the well known Markov branching processes (MBP) discussed extensively in the literature. The MBP are continuous time Markov processes with a countable state space $(0, 1, 2, \ldots)$. See [11] and the references therein for MBP and their various generalizations. On the other hand, the BMP model discussed in this paper is a tree indexed process possessing a certain type of Markov property. The state space of a BMP can be discrete or continuous.

2. Model specification and examples

Let $\{Z_t, t = 0, 1, 2, \ldots\}$ be a Galton–Watson branching process with $Z_0 = 1$. Assume that $E(Z_1) = m > 1$ and $\mathrm{Var}(Z_1) = \sigma^2$ so that $\{Z_t\}$ is super-critical. It is further assumed that $P(Z_1 = 0) = 0$, i.e., the process $\{Z_t\}$ does not become extinct. We collect some well-documented results on $\{Z_t\}$ which will be needed later.

Proposition 2.1. *Suppose $\{Z_t\}$ is a G–W branching process specified above.*

(i) *Let $W_n = Z_n/m^n$. Then there exists a random variable W such that*

$$W_n \xrightarrow{a.s.} W, \quad \text{as } n \to \infty$$

where '$\xrightarrow{a.s.}$' denotes almost sure convergence.

(ii) $P(W > 0) = 1$

(iii) $E(W) = 1$ *and* $\mathrm{Var}(W) = \sigma^2/m(m-1)$.

See, for instance, [3, p. 13]. Throughout, we will write $a_n \sim b_n$ for denoting 'asymptotic equivalence' of the two strictly positive real sequences $\{a_n\}$ and $\{b_n\}$, i.e., $a_n \sim b_n$ if a_n/b_n converges to one as $n \to \infty$. It then follows from (i) *and* (ii) *of Proposition 2.1 that*

$$Z_n \sim m^n W \quad (a.s.) \tag{2.1}$$

where and in what follows $(a.s.)$ is used for representing 'with probability one'.

Consider the process $\{X_t(j), t = 0, 1, 2, \ldots, j = 1, 2, 3, \ldots\}$ for which $X_t(j)$ denotes observation on the j-th individual in generation t. It will be assumed that $\{X_t(j)\}$ and $\{Z_t\}$ are independent of each other. The observable process $\{Z_t, X_t(j)\}$ consists of

$$\{(Z_t, X_t(j))\,;\, t = 0, 1, 2, \ldots, j = 1, 2, \ldots, Z_t\}. \tag{2.2}$$

We will refer to the process (2.2) as a branching Markov process (BMP). The subscript t will be used for denoting 'tth generation' unless indicated otherwise. Let $X_{t-1}(t(j))$ be an observation on the immediate parent of the j-th individual of generation t. Let $\mathcal{F}_t$ denote the σ-field generated by observations up to generation t, i.e.,

$$\mathcal{F}_t = \sigma\{(Z_s, X_s(1), \ldots, X_s(Z_s)), s = 0, 1, 2, \ldots, t\}, \quad t \geq 1 \tag{2.3}$$

and $\mathcal{F}_0$ is a trivial σ-field. Notice that $X_{t-1}(t(j)) \in \mathcal{F}_{t-1}$. An illustrative realization of BMP model is given in Fig. 1.

Notice that $x_2(2)$ produces three offsprings in the next generation and $x_2(3)$ gives rise to $x_3(5)$ and $x_3(6)$, i.e., $x_2(3(5)) = x_2(3)$ and $x_2(3(6)) = x_2(3)$. In particular when $\{Z_t\}$ is deterministic in such a way that $Z_t = m^t$ where m is an integer larger than 1, BMP reduces to a m-splitting model for tree-structured data where each individual in one generation gives rise to m-offsprings in the next generation. For the special case when $m = 2$ and $Z_t = 2^t, t = 0, 1, 2, \ldots$, BMP becomes a bifurcating (binary-splitting) model. See Fig. 2. Refer to [4,10,9] for a review of bifurcating models.

In order to formulate the BMP model, it will be assumed that

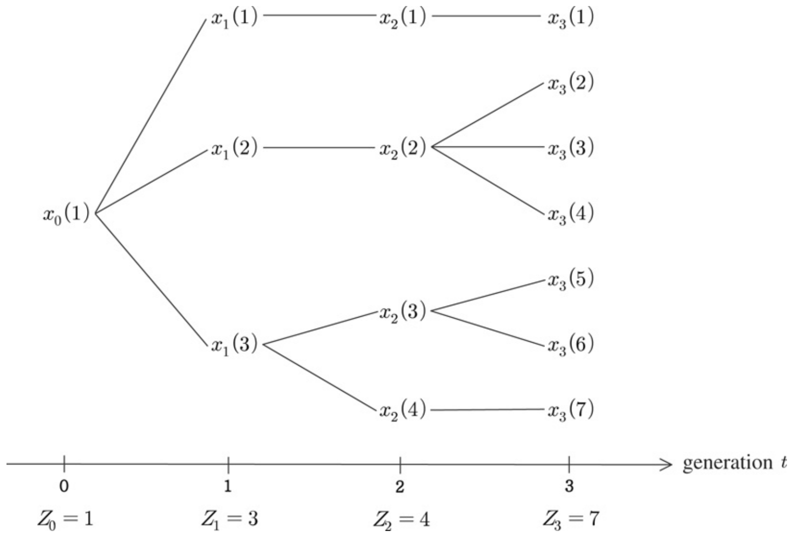

Fig. 1. A realization of BMP model.

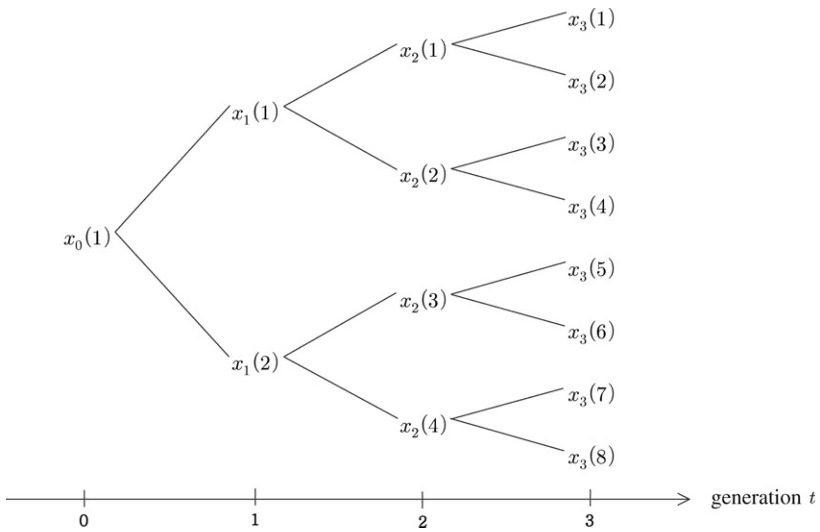

Fig. 2. Bifurcating data.

(A.1) Markovity

$$p(x_t(j) \mid \mathcal{F}_{t-1}) = p(x_t(j) \mid x_{t-1}(t(j)))$$

where $p(\cdot \mid \cdot)$ denotes a conditional density. Markovian Assumption (A.1) implies that conditional distribution of any x in tth generation given $\mathcal{F}_{t-1}$ depends only on a x in $(t-1)$th generation on the same path, and thus $p(\cdot \mid \cdot)$ can be viewed as a transition density from one generation to the next generation.

The conditional mean and conditional variance of the transition density $p(\cdot \mid \cdot)$ will be denoted by $\mu_t(j)$ and $v_t(j)$ respectively, each defined by

$$\mu_t(j) = E[X_t(j) \mid X_{t-1}(t(j))] \tag{2.4}$$

and

$$v_t(j) = \mathrm{Var}[X_t(j) \mid X_{t-1}(t(j))]. \tag{2.5}$$

It is noted that both $\mu_t(j)$ and $v_t(j)$ are $\mathcal{F}_{t-1}$-measurable as functions of $X_{t-1}(t(j))$.

(A.2) Conditional independence: Conditionally on $\mathcal{F}_{t-1}$, $\{X_t(j), j = 1, 2, \ldots, Z_t\}$ are independent.

We present some examples, as special cases of BMP models which will be discussed in Section 5. These examples satisfy (A.1) and (A.2).

Ex.1 Branching autoregressive processes (B-AR(1))
Branching AR(1) process is defined by

$$X_t(j) = \beta_0 + \beta_1 X_{t-1}(t(j)) + \varepsilon_t(j), \quad |\beta_1| < 1 \tag{2.6}$$

where $\{\varepsilon_t(j); t = 0, 1, 2, \ldots, j = 1, 2, \ldots\}$ is iid with mean zero and variance $\sigma_\varepsilon^2 > 0$. As a special case, see [4] for a standard bifurcating AR(1) (BAR(1)) model.

Ex.2 Branching conditionally linear autoregressive processes (B-CLAR(1))
This class of models is formulated by

$$\mu_t(j) = E(X_t(j) \mid X_{t-1}(t(j))) = \beta_0 + \beta_1 X_{t-1}(t(j)), \quad |\beta_1| < 1. \tag{2.7}$$

This class is rich enough to accommodate count data and non-negative data. These models do not require autoregression (AR)-structure (2.6). Refer to [12] for comprehensive discussions on CLAR(1) models. A random coefficient AR(1) model is defined by

$$X_t(j) = \beta_0 + \beta_t X_{t-1}(t(j)) + \varepsilon_t(j) \tag{2.8}$$

where $\{\beta_t\}$ represents a sequence of random variables (with mean β_1) which is assumed to be independent of $\{\varepsilon_t(j)\}$. See [13]. A binomial thinning model is specified by

$$X_t(j) = \sum_{i=1}^{X_{t-1}(t(j))} B_i + \varepsilon_t(j) \tag{2.9}$$

where $\{B_i\}$ denotes iid Bernoulli random variables and $\{\varepsilon_t(j)\}$ is an iid non-negative integer-valued random variables which are independent of $\{B_i\}$. Models (2.8) and (2.9) are examples of B-CLAR(1) class.

Ex.3 Branching-conditional exponential family
A branching-conditional exponential family is a class of models with transition densities of the form

$$p(x_t(j) \mid x_{t-1}(t(j))) = c(x_t(j)) \exp[x_t(j)\eta_t(j) - k(\eta_t(j))] \tag{2.10}$$

where $\eta_t(j)$ is a function of $x_{t-1}(t(j))$ via $\eta_t(j) = g(\mu_t(j))$ where $\mu_t(j)$ is the conditional mean in (2.4) and $g(\cdot)$ is referred to as a link function. Notice that $\eta_t(j)$ and $k(\cdot)$ belong to $\mathcal{F}_{t-1}$. For illustration, consider the following conditional exponential model with conditional mean $\mu_t(j)$.

$$p(x_t(j) \mid x_{t-1}(t(j))) = \exp[-x_t(j)/\mu_t(j) - \log \mu_t(j)], \quad x_t(j) > 0 \tag{2.11}$$

Conditional Poisson with $\mu_t(j)$ is defined by

$$p(x_t(j) \mid x_{t-1}(t(j))) = [x_t(j)!]^{-1} \exp[x_t(j) \log \mu_t(j) - \mu_t(j)], \quad x_t(j) = 0, 1, 2, \ldots \tag{2.12}$$

In order to identify the ancestral path of the observation $x_t(j)$, we use the notation

$$\{x_{t-i}(t(j)), i = 0, 1, 2, \ldots, t\} \tag{2.13}$$

where $x_{t-i}(t(j))$ represents i-th ancestor of $x_t(j)$ with the understanding that $x_t(t(j)) = x_t(j)$. Note that $x_0(t(j)) = x_0(1)$ for any $t(j)$, i.e., for any t and j. It follows from Markovity Assumption (A.1) that $\{x_{t-i}(t(j)), i = 0, 1, 2, \ldots, t\}$ constitutes a Markov process with transition density appearing in (A.1). It will be assumed that

(C.1) The Markov process $\{x_{t-i}(t(j)), i = 0, 1, 2, \ldots, t\}$ is strictly stationary and ergodic.

For sufficient conditions ensuring (C.1), refer to, among others, [14–16]. In particular, see [12] for simple but readily applicable conditions for (C.1) in the context of CLAR(1) class. Notice that the ergodic(stationary) distribution remains the same for all ancestral paths.

Lemma 2.1. *Under (C.1), for any fixed $k = 0, 1, 2, \ldots,$ and for any $t, s, j,$ and u*

$$(X_t(t(j)), X_{t-1}(t(j)), \ldots, X_{t-k}(t(j))) \stackrel{d}{=} (X_s(s(u)), X_{s-1}(s(u)), \ldots, X_{s-k}(s(u))). \tag{2.14}$$

Here '$\stackrel{d}{=}$' is used for 'equality in distribution'.

Proof. The lemma follows from the strict stationarity along with any ancestral paths with starting value $X_0(1)$. □

In particular, for $k = 0$, we have $X_t(j) \stackrel{d}{=} X_s(u)$, for any t, s, j and u. Taking $k = 1$, (2.14) reduces to

$$(X_t(t(j)), X_{t-1}(t(j))) \stackrel{d}{=} (X_s(s(u)), X_{s-1}(s(u))). \tag{2.15}$$

Consequently, Eq. (2.14) can be viewed as a joint-distributional property of the pathwise stationarity of the BMP model along with any particular ancestral path. It is however stressed that BMP may not be stationary in the usual sense. From now on, probabilistic statements such as $E(\cdot)$ and $\mathrm{Var}(\cdot)$ will be made under the pathwise stationary distribution. The covariance between the two generations t and $t-k$, $k \geq 0$, is denoted by γ_k given by $\gamma_k = \mathrm{Cov}(X_t(j), X_{t-k}(t(j)))$. Accordingly, the correlation coefficient ρ_1 between adjacent generations is obtained by

$$\rho_1 = \gamma_1/\gamma_0 \tag{2.16}$$

where $\gamma_0 = \mathrm{Var}(X_t(j))$.

In the next two sections, we will derive a strong law of large numbers (SLLN) and a central limit theorem (CLT) which will be of use in statistical inference for the BMP models.

3. A law of large numbers for BMP models

Consider a real-valued random function $f(X_t(j), X_{t-1}(t(j)))$ of $X_t(j)$ and its immediate parent $X_{t-1}(t(j))$. It will be shown that

$$\left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum_{t=1}^{n}\sum_{j=1}^{Z_t} f(X_t(j), X_{t-1}(t(j))) \xrightarrow{a.s.} \mu_f \tag{3.1}$$

where

$$\mu_f = E\left[f(X_t(j), X_{t-1}(t(j)))\right] \tag{3.2}$$

In what follow, the double (random) summation $\sum\sum$ will be used for denoting $\sum_{t=1}^{n}\sum_{j=1}^{Z_t}$ unless indicated otherwise. Denote for simplicity

$$f_t(j) = f(X_t(j), X_{t-1}(t(j))) \tag{3.3}$$

and therefore $\mu_f = Ef_t(j)$. Consider first the sample average over the n-th generation defined by

$$A_n = (Z_n)^{-1}\sum_{j=1}^{Z_n} f_n(j) \tag{3.4}$$

consisting of the random sum of Z_n-variates ; $f_n(1), \ldots, f_n(Z_n)$. Consider an arbitrary realization of $Z_1 = c_1, Z_2 = c_2, \ldots, Z_n = c_n, \ldots$. It follows from (2.1) that c_n/m^n converges to a positive constant as $n \to \infty$. We will approximate A_n by $\tilde{A_n}$ obtained by replacing Z_n by c_n, i.e.,

$$\tilde{A_n} = c_n^{-1}\sum_{j=1}^{c_n} f_n(j) \tag{3.5}$$

To proceed, we are willing to rule out the case where a single individual dominates the next generation by imposing the condition below.

(C.2) Among c_n^2-possible pairs of $(X_n(u), X_n(v))$, $u, v = 1, \ldots, c_n$, let τ_n be the number of pairs sharing the same immediate parent belonging to $(n-1)$-th generation. Assume that

$$\tau_n = O\left(c_n^{2-\delta}\right) \quad \text{for some } \delta > 0 \tag{3.6}$$

where $O(\cdot)$ denotes standard 'big O' notation, i.e., $a_n = O(b_n)$ when $\{a_n\}$ is at most of order $\{b_n\}$.

Remarks. Bifurcating model for which $c_n = 2^n$ identifies $\tau_n = c_n + c_{n-1}$. For m-splitting case where $c_n = m^n$ for which m is an integer greater than of equal to 2, one may obtain

$$\tau_n = c_n + \binom{m}{2} c_{n-1}$$

and thus one can choose $\delta = 1$ in (C.2).

Lemma 3.1. *Under* (C.1) *and* (C.2) *plus* $Ef_n^2(j) < \infty$, *we have as* $n \to \infty$,

$$\tilde{A_n} \xrightarrow{a.s.} \mu_f \tag{3.7}$$

Proof. Write

$$\tilde{A_n} = \tilde{A_{n1}} + \tilde{A_{n2}}$$

where

$$\tilde{A_{n1}} = c_n^{-1}\sum_{j=1}^{c_n}\{f_n(j) - E(f_n(j) \mid \mathcal{F}_{n-1})\}$$

and

$$\tilde{A_{n2}} = c_n^{-1}\sum_{j=1}^{c_n} E\left(f_n(j) \mid \mathcal{F}_{n-1}\right).$$

For $\tilde{A_{n1}}$, notice that $\{f_n(j) - E(f_n(j) \mid \mathcal{F}_{n-1}), j = 1, \ldots, c_n\}$ constitutes a sequence of martingale differences for each fixed n. Thus $E\tilde{A_{n1}} = 0$ and

$$\mathrm{Var}\left(\tilde{A_{n1}}\right) = c_n^{-2} \sum_{j=1}^{c_n} \mathrm{Var}\{f_n(j) - E\left(f_n(j) \mid \mathcal{F}_{n-1}\right)\}. \tag{3.8}$$

Observe also that

$$\begin{aligned} \mathrm{Var}\{f_n(j) - E(f_n(j) \mid \mathcal{F}_{n-1})\} &= E\left[\mathrm{Var}(f_n(j) \mid \mathcal{F}_{n-1})\right] \\ &\le \mathrm{Var}(f_n(j)) < \infty, \quad \text{due to } Ef_n^2(j) < \infty. \end{aligned}$$

Accordingly, $\mathrm{Var}(\tilde{A_{n1}}) = c_n^{-1}\mathrm{Var}\{f_n(1) - E(f_n(1) \mid \mathcal{F}_{n-1})\} = O(c_n^{-1})$ and in turn implies via Borel–Cantelli's lemma that

$$\tilde{A_{n1}} \xrightarrow{a.s.} 0 \tag{3.9}$$

since $c_n = \omega m^n$ for some $\omega > 0$ and $m > 1$. It then suffices to verify that

$$\tilde{A_{n2}} \xrightarrow{a.s.} \mu_f. \tag{3.10}$$

Consider

$$\begin{aligned} \tilde{A_{n2}} - \mu_f &= c_n^{-1} \sum_{j=1}^{c_n} [E(f_n(j) \mid \mathcal{F}_{n-1}) - \mu_f] \\ &= c_n^{-1} \sum_{j=1}^{c_n} \Delta_n(j), \qquad \Delta_n(j) = E(f_n(j) \mid \mathcal{F}_{n-1}) - \mu_f. \end{aligned} \tag{3.11}$$

Note that $\Delta_n(j) \in \mathcal{F}_{n-1}$ and $E(\Delta_n(j)) = 0$. Now,

$$\mathrm{Var}\left(\sum_{j=1}^{c_n} \Delta_n(j)\right) = \sum_{u=1}^{c_n} \sum_{v=1}^{c_n} E\left(\Delta_n(u)\Delta_n(v)\right). \tag{3.12}$$

Let

$$e(u, v) = E(\Delta_n(u)\Delta_n(v) \mid \mathcal{F}_{n-2}), \qquad u, v = 1, \ldots, c_n. \tag{3.13}$$

It then follows from the conditional independence assumption (A.2) that $e(u, v)$ reduces to zero unless u and v shares the same immediate parent in $\mathcal{F}_{n-1}$. It then follows from (C.2) that

$$\mathrm{Var}\left(\left(\sum_{j=1}^{c_n} \Delta_n(j)\right)\right) = O({c_n}^{2-\delta})$$

and in turn

$$\begin{aligned} \mathrm{Var}(\tilde{A_{n2}}) &= O(c_n^{-\delta}) \\ &= O(m^{-\delta n}), \quad \delta > 0. \end{aligned}$$

Thus, by noting $m > 1$, this implies (3.10) due to Borel–Cantelli's lemma, completing the proof. □

We are now in a position to present a strong law of large numbers for BMP models.

Theorem 3.1. *Under the same conditions as for Lemma 3.1, we have as $n \to \infty$,*

(i) $A_n \xrightarrow{a.s} \mu_f$
(ii) *The SLLN specified in (3.1) holds.*

Proof. Let

$$Z = \{Z_1 = c_1, Z_2 = c_2, \ldots, Z_n = c_n, \ldots\}$$

where the sequence $\{c_n\}$ of constants is an arbitrary realization of Z's appearing in $\tilde{A}_n$. Consider the conditional probability

$$P\left(A_n \to \mu_f \mid Z\right)$$

which is equivalent to

$$P\left(\tilde{A}_n \to \mu_f \mid Z\right)$$

which in turn reduces to $P\left(\tilde{A_n} \to \mu_f\right)$ due to the independence between $\{Z_t\}$ and $\{X_t(j)\}$. Consequently, Lemma 3.1 gives

$$P\left(A_n \to \mu_f \mid Z\right) = 1 \tag{3.14}$$

for any arbitrary realization of Z's. By taking expectation on both sides of (3.14), we obtain the assertion (i). For verifying (ii), write

$$\left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum\sum f_t(j) = \left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum_{t=1}^{n} Z_t A_t \tag{3.15}$$

where $A_t = (Z_t)^{-1} \sum_{j=1}^{Z_t} f_t(j)$ is a sample average over the t-th generation. Since $Z_t \xrightarrow{a.s} \infty$, as $t \to \infty$, employing Toeplitz lemma (cf. [17]), (i) implies that the term in (3.15) converges to μ_f (a.s.). This concludes (ii). □

Due to Theorem 3.1, the stationary mean(μ_X) and variance(σ_X^2) of the BMP model can be consistently estimated by their sample counterparts. The sample mean $\bar{X}$ and sample variance S_X^2 are defined respectively by

$$\bar{X} = \left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum\sum X_t(j) \tag{3.16}$$

and

$$S_X^2 = \left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum\sum \left(X_t(j) - \bar{X}\right)^2. \tag{3.17}$$

It then readily follows from Theorem 3.1 that $\bar{X} \xrightarrow{a.s} \mu_X$ and $S_X^2 \xrightarrow{a.s} \sigma_X^2$. In addition, the correlation ρ_1 between two adjacent generations (see (2.16)) can be consistently estimated by the sample correlation coefficient $\hat{\rho}_1$ defined by

$$\hat{\rho}_1 = \sum\sum \left(X_t(j) - \bar{X}\right)\left(X_{t-1}(t(j)) - \bar{X}\right)/S_X^2.$$

In the next section we will establish a CLT which will be useful in deriving asymptotic distributions of parameter estimators.

4. A central limit theorem for BMP models

Consider the random sum defined by

$$S_n = \left(\sum_{t=1}^{n} Z_t\right)^{-1/2} \sum_{t=1}^{n}\sum_{j=1}^{Z_t} D_t(j) \tag{4.1}$$

where

$$D_t(j) = f_t(j) - E\left(f_t(j) \mid \mathcal{F}_{t-1}\right), \quad t = 1, 2, \ldots, n$$

with

$$f_t(j) = f\left(X_t(j), X_{t-1}(t(j))\right)$$

defined in (3.3). To begin with, we will deal with a non-random sum version of S_n. Specifically, consider again a sequence of constants $\{c_n\}$ such that $Z_1 = c_1, Z_2 = c_2, \ldots, Z_n = c_n$. Define

$$\tilde{S_n} = \left(\sum_{t=1}^{n} c_t\right)^{-1} \sum_{t=1}^{n}\sum_{j=1}^{c_t} D_t(j) \tag{4.2}$$

where c_n/m^n converges to a positive constant. Notice that $\{S_n\}$ is regarded as a sequence of randomly selected partial sums associated with $\{\tilde{S_n}\}$. Billingsley [18, p. 143] discussed conditions under which randomly selected partial sum S_n converges in distribution to the same limiting distribution as for the non-random sum $\tilde{S_n}$. It is then natural to expect that S_n and $\tilde{S_n}$ are asymptotically equivalent (in distribution) as addressed in the following lemma. It is remarked that the case of only one random selection is discussed in [18, p. 143] whereas n random selections $Z_1, \ldots, Z_n$ are involved with S_n in (4.1).

Lemma 4.1. *If $\tilde{S_n} \xrightarrow{d} N(0, \mathrm{Var}(D_t(j)))$, as $n \to \infty$, then*

$$S_n \xrightarrow{d} N(0, \mathrm{Var}(D_t(j))) \tag{4.3}$$

where $\mathrm{Var}(D_t(j)) < \infty$ is assumed.

Proof. Notice that for each fixed $-\infty < x < \infty$

$$\begin{aligned} P(S_n \le x \mid Z_1 = c_1, \ldots, Z_n = c_n) &= P(\tilde{S_n} \le x \mid Z_1 = c_1, \ldots, Z_n = c_n) \\ &= P(\tilde{S_n} \le x) \end{aligned}$$

which is due to the independence between $\{Z_t\}$ and $\{X_t(j)\}$. Since $P(\tilde{S_n} \le x)$ converges to $\Phi(x)$ where $\Phi(x)$ denotes the distribution function of normal random variate with mean zero and variance given by $\mathrm{Var}(D_t(j))$ we have as $n \to \infty$

$$P(S_n \le x \mid Z_1 = c_1, \ldots, Z_n = c_n) \to \Phi(x), \quad -\infty < x < \infty. \tag{4.4}$$

Employing bounded convergence theorem, taking expectation on both sides of (4.4), we conclude (4.3), completing the proof. □

For analyzing $\tilde{S_n}$, we first relabel the data in a linear fashion in such a way that

$$\begin{array}{llll} X_1(1) = U_1, & X_1(2) = U_2, & \ldots, & X_1(c_1) = U_{c_1} \\ X_2(1) = U_{c_1+1}, & \ldots, & \ldots, & X_2(c_2) = U_{c_1+c_2} \\ X_n(1) = U_{c_1+\cdots+c_{n-1}+1}, & \ldots, & \ldots, & X_n(c_n) = U_{c_1+\cdots+c_n}. \end{array}$$

Notice that

$$U_l = X_t(j) \tag{4.5}$$

for which

$$l = \sum_{i=1}^{t-1} c_i + j, \quad t = 1, \ldots, n, j = 1, \ldots, c_t. \tag{4.6}$$

Denote $N = c_1 + \ldots + c_n$ so that total N observations $\{U_1, \ldots, U_N\}$ are available. Define σ-field B_l associated with $\{U_l, l = 1, \ldots, N\}$ as

$$B_l = \sigma(U_1, \ldots, U_l).$$

Now, $\tilde{S_n}$ in (4.2) can be written in terms of U's as

$$\tilde{S_n} = N^{-1/2} \sum_{l=1}^{N} D_l$$

where

$$D_l = D_t(j) = f_t(j) - E\left(f_t(j) \mid \mathcal{F}_{t-1}\right) \tag{4.7}$$

with l given by (4.6). Here $D_t(j)$ is treated as a function of U's and notice that $B_{l-1} = \mathcal{F}_{t-1}$ due to Markovity Assumption (A.1). Accordingly, $\{D_l\}$ is a martingale difference sequence with respect to $\{B_l\}$. We now present the asymptotic normality of S_n defined in (4.1).

Theorem 4.1. *Under the same conditions as for Theorem 3.1, we have*

$$S_n \xrightarrow{d} N(0, \mathrm{Var}(D_t(j))), \quad n \to \infty.$$

Proof. If suffices to verify (due to Lemma 4.1) that

$$\tilde{S_n} \xrightarrow{d} N(0, \mathrm{Var}(D_t(j))). \tag{4.8}$$

Since $n \to \infty$ is equivalent to $N \to \infty$, we write $\tilde{S_N} = \tilde{S_n}$. The sum of conditional variances associated with $\tilde{S_N}$ can be written as follows.

$$\begin{aligned} N^{-1} \sum_{l=1}^{n} \mathrm{Var}(D_l \mid B_{l-1}) = & \left(\sum_{t=1}^{n} c_t\right)^{-1} \sum_{t=1}^{n} \sum_{j=1}^{c_t} \mathrm{Var}(D_t(j) \mid \mathcal{F}_{t-1}) \\ & \xrightarrow{a.s.} E\left[\mathrm{Var}(D_t(j)) \mid \mathcal{F}_{t-1}\right] = \mathrm{Var}[D_t(j)] \end{aligned} \tag{4.9}$$

where '$\xrightarrow{a.s.}$' holds due to Theorem 3.1 (SLLN). Also, the relevant Lindeberg condition is satisfied. To see this, consider, for a given $\varepsilon > 0$

$$N^{-1} \sum_{l=1}^{N} E\left(|D_l|^2 I_{\left[|D_l| > \varepsilon\sqrt{N}\right]}\right). \tag{4.10}$$

Since $D_1, \ldots, D_N$ are identically distributed and $ED_l^2 < \infty$, we conclude that (4.10) goes to zero. Thus, together with (4.9), key conditions for a martingale CLT due to [19] are satisfied, yielding (4.8). This completes the proof. □

The random norm in S_n enables us to obtain asymptotic normality of S_n. If Z_t is replaced by its expectation $E(Z_t) = m^t$, we obtain instead a variance mixture of normals. Consider S_n^* defined by

$$S_n^* = \left(\sum_{t=1}^{n} m^t\right)^{-1/2} \sum_{t=1}^{n} \sum_{j=1}^{Z_t} D_t(j).$$

Corollary 4.1. *Under the same conditions as for Theorem 4.1, as $n \to \infty$,*

$$S_n^* \xrightarrow{d} W^{1/2} \cdot N(0, \mathrm{Var}(D_t(j))) \tag{4.11}$$

where W denotes the almost sure limit of Z_n/m^n defined in Proposition 2.1.

Proof. Note that

$$Z_1 + \cdots + Z_n \sim \left(\sum_{t=1}^{n} m^t\right) W \quad (a.s.) \tag{4.12}$$

Corollary follows by combining Theorem 4.1 and Eq. (4.12). □

In particular, for bifurcating model, note that $Z_t = 2^t$ and $W = 1$ (a.s.) and thus (4.11) reduces to

$$\left(2^{n+1}\right)^{-1/2} \sum_{t=1}^{n} \sum_{j=1}^{2^t} D_t(j) \xrightarrow{d} N(0, \mathrm{Var}(D_t(j))). \tag{4.13}$$

As an application of Theorem 4.1, consider the following innovation process $\{e_t(j); t = 1, \ldots, n, j = 1, \ldots, Z_t\}$ defined by

$$e_t(j) = [X_t(j) - \mu_t(j)] / \sqrt{v_t(j)} \tag{4.14}$$

where $\mu_t(j)$ and $v_t(j)$ are respectively conditional mean and variance of $X_t(j)$. B-AR(1) model in Ex.1 gives $\mu_t(j) = \beta_0 + \beta_1 X_{t-1}(t(j))$ and $v_t(j) = \sigma_\varepsilon^2$. For a conditional exponential model with $\mu_t(j)$ in (2.11), it is seen that $v_t(j) = [\mu_t(j)]^2$. It can be readily shown that $E(e_t(j) \mid \mathcal{F}_{t-1}) = 0$ and $\mathrm{Var}(e_t(j) \mid \mathcal{F}_{t-1}) = 1$ and thus $E(e_t(j)) = 0$ and $\mathrm{Var}(e_t(j)) = 1$. In addition, it follows from (A.2) that $\mathrm{Cov}(e_t(j), e_s(u))$ vanishes unless $t = s$ and $j = u$. Consequently $\{e_t(j)\}$ is a white noise process with mean zero and variance unity.

By setting $e_t(j) = D_t(j)$ in (4.1), Theorem 4.1 reduces to

$$\left(\sum_{t=1}^{n} Z_t\right)^{-1/2} \sum\sum e_t(j) \xrightarrow{d} N(0, 1) \tag{4.15}$$

which facilitates the usage of the standard normal distribution for approximating the distribution of $\sum\sum e_t(j)$. One may expect with confidence level 0.95 that $\sum\sum e_t(j)$ lies in the random interval

$$\left(-1.96\sqrt{\sum_{t=1}^{n} Z_t},\ 1.96\sqrt{\sum_{t=1}^{n} Z_t}\right) \tag{4.16}$$

for large n. Accordingly it will be inappropriate to use a BMP model when $\sum\sum e_t(j)$ falls outside of the interval (4.16).

5. Applications to parameter estimation

We now consider some examples to illustrate the models and limit theorems addressed in previous sections. Least squares, quasilikelihood and maximum likelihood estimation for model parameters are discussed.

5.1. Least squares estimation

Recall the branching-AR(1) process defined in Ex.1 by

$$X_t(j) = \beta_0 + \beta_1 X_{t-1}(t(j)) + \varepsilon_t(j), \quad |\beta_1| < 1. \tag{5.1}$$

Repeated application of the recursive equation (5.1) leads to the stationary solution

$$X_t(j) = \sum_{k=0}^{\infty} \beta_1^k (\beta_0 + \varepsilon_{t-k}(t(j))) \tag{5.2}$$

where $\varepsilon_{t-k}(t(j))$ represents iid errors corresponding to $X_{t-k}(t(j))$ for $k < t$, and we set $\varepsilon_{t-k}(t(j)) = \varepsilon_{t-k}(1)$ for all $k \geq t$ with the understanding that $\varepsilon_0(1), \varepsilon_{-1}(1), \varepsilon_{-2}(1), \ldots,$ are iid. It is easy to verify from (5.2) that the stationary moments are given by

$$\mu_X = \beta_0(1 - \beta_1)^{-1}, \quad \sigma_X^2 = \sigma_\varepsilon^2/(1 - \beta_1^2)$$

and

$$\gamma_k = \mathrm{Cov}(X_t(j), X_{t-k}(t(j))) = \beta_1^k \sigma_X^2, \quad k \geq 0 \tag{5.3}$$

It can further be verified that the sequence of random variables along any ancestral path $\{X_{t-k}(t(j)), k = 0, 1, 2, \ldots, t\}$ with $X_0(t(j)) = X_0(1)$ satisfies (C.1) of Section 2. As a direct consequence of Theorem 3.1, we have as $n \to \infty$, $\bar{X} \xrightarrow{a.s.} \mu$ and $S_X^2 \xrightarrow{a.s.} \sigma_X^2$ where $\bar{X}$ and S_X^2 are respectively sample mean and variance defined (3.16) and (3.17). Furthermore, it can be shown that $\hat{\gamma}_k \xrightarrow{a.s.} \gamma_k$ where $\hat{\gamma}_k$ denotes sample autocovariance of generation-lag k, defined by

$$\hat{\gamma}_k = \left(\sum_{t=1}^{n} Z_t\right)^{-1} \sum_{t=k+1}^{n} \sum_{j=1}^{Z_t} (X_t(j) - \bar{X})(X_{t-k}(t(j)) - \bar{X}), \quad k \geq 0.$$

Denote by $(\hat{\beta}_0, \hat{\beta}_1)$ the least squares (LS) estimates of (β_0, β_1) based on the sample

$$\{Z_t, X_t(j); t = 0, 1, \ldots, n, j = 1, \ldots, Z_t\}.$$

Define

$$A_n = \begin{pmatrix} \sum Z_t & \sum\sum X_{t-1}(t(j)) \\ \sum\sum X_{t-1}(t(j)) & \sum\sum X_{t-1}^2(t(j)) \end{pmatrix}, \qquad B_n = \begin{pmatrix} \sum\sum \varepsilon_t(j) \\ \sum\sum \varepsilon_t(j) X_{t-1}(t(j)) \end{pmatrix} \tag{5.4}$$

where $\varepsilon_t(j) = X_t(j) - \beta_0 - \beta_1 X_{t-1}(t(j))$. Here and in what follow $\sum$ is used for $\sum_{t=1}^n$ for simplicity of notation. Also $\sum\sum$ denotes $\sum_{t=1}^n \sum_{j=1}^{Z_t}$ unless stated otherwise. It is seen that

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = A_n^{-1} \begin{pmatrix} \sum\sum X_t(j) \\ \sum\sum X_t(j) X_{t-1}(t(j)) \end{pmatrix}. \tag{5.5}$$

Limiting distribution of LS estimators $(\hat{\beta}_0, \hat{\beta}_1)$ is identified in the next theorem.

Theorem 5.1. *As $n \to \infty$, we have*

(i) $\left(\sum Z_t\right)^{-1} A_n \xrightarrow{a.s.} A = \begin{pmatrix} 1 & \mu_X \\ \mu_X & \gamma(0) + \mu_X^2 \end{pmatrix}$

(ii) $\left(\sum Z_t\right)^{-1/2} B_n \xrightarrow{d} N(0, \sigma_\varepsilon^2 A)$

and hence

(iii)

$$\left(\sum Z_t\right)^{1/2} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \xrightarrow{d} N(0, \sigma_\varepsilon^2 A^{-1}). \tag{5.6}$$

Proof. Assertion (i) is immediate from the SLLN of Theorem 3.1. To verify (ii), let $a = (a_1, a_2)^{\mathrm{T}}$ denote non-zero vector of constants. Consider

$$a^{\mathrm{T}} B_n = \sum\sum \varepsilon_t(j)[a_1 + a_2 X_{t-1}(t(j))].$$

Choose $D_t(j) = \varepsilon_t(j)[a_1 + a_2 X_{t-1}(t(j))]$ in Theorem 4.1 (CLT) to get

$$\left(\sum Z_t\right)^{-1/2} a^{\mathrm{T}} B_n \xrightarrow{d} N(0, \mathrm{Var}(D_t(j))) \tag{5.7}$$

where

$$\mathrm{Var}(D_t(j)) = \sigma_\varepsilon^2 (a^{\mathrm{T}} A a). \tag{5.8}$$

Thus, via the Cramer–Wold device, (ii) follows from (5.7) and (5.8). Write

$$\left(\sum Z_t\right)^{1/2} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} = \left[\left(\sum Z_t\right)^{-1} A_n\right]^{-1} \left[\left(\sum Z_t\right)^{-1/2} B_n\right]$$

which implies (iii) using (i) and (ii). □

Since $\sum Z_t \sim \left(\frac{m^{n+1}}{m-1}\right) W$, we also have

$$\left(\frac{m^{n+1}}{m-1}\right)^{1/2} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \xrightarrow{d} W^{-1/2} N(0, \sigma_\varepsilon^2 A^{-1}). \tag{5.9}$$

Note that the limit distribution in (5.9) is a mixture of bivariate normals. The results in (5.6) and (5.9) give the limit distribution of the least squares estimators using a random and a non-random norm, respectively.

5.2. Quasilikelihood estimation

Consider a process $\{X_t(j)\}$ specified only by its conditional mean and conditional variance, viz.,

$$\mu_{tj}(\beta) = E\left(X_t(j) \mid X_{t-1}(t(j))\right) = \beta_0 + \beta_1 X_{t-1}(t(j))$$

and

$$v_{tj}(\beta) = \text{Var}\left(X_t(j) \mid X_{t-1}(t(j))\right)$$

where $\mu_{tj}(\beta)$ and $v_{tj}(\beta)$ are used instead of $\mu_t(j)$ and $v_t(j)$ respectively, in order to emphasize dependency on the parameter vector β. Grunwald et al. [12] have discussed several examples of models of this kind for standard Markov processes. These examples can be readily extended to the BMP models. Note that the AR-structure (5.1) is not required for this class of models. The class includes random coefficient AR(1) and binomial thinning process. See (2.8) and (2.9). In many interesting applications, the conditional variance $v_{tj}(\beta)$ is a quadratic function of the conditional mean $\mu_{tj}(\beta)$. See [12] for further details. Here we illustrate the application of our results to quasilikelihood estimation of β for the special case of the first-order B-CLAR(1) process with $\mu_{tj}(\beta) = \beta_0 + \beta_1 X_{t-1}(t(j))$. The quasilikelihood estimating equation for β is given by

$$\sum\sum v_{tj}^{-1}(\beta) \left(\frac{\partial \mu_{tj}(\beta)}{\partial \beta}\right) \left(X_t(j) - \mu_{tj}(\beta)\right) = 0. \tag{5.10}$$

See, for instance, [20] for background on quasilikelihood estimation. The least squares estimating equation for β is given by

$$\sum\sum \left(\frac{\partial \mu_{tj}(\beta)}{\partial \beta}\right) \left(X_t(j) - \mu_{tj}(\beta)\right) = 0. \tag{5.11}$$

For the B-CLAR(1) model, the least squares estimates of β_0 and β_1 are given by (5.5).

Consider a modified quasilikelihood estimating equation defined by

$$\sum\sum v_{tj}^{-1}(\hat{\beta}) \left(\frac{\partial \mu_{tj}(\beta)}{\partial \beta}\right) \left(X_t(j) - \mu_{tj}(\beta)\right) = 0 \tag{5.12}$$

where $\hat{\beta}$ denotes the least squares estimate of β given by (5.5). Then, for B-CLAR(1) class, (5.12) gives the (modified) quasilikeilhood estimates

$$\begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} \sum\sum v_{tj}^{-1}(\hat{\beta}) & \sum\sum v_{tj}^{-1}(\hat{\beta}) X_{t-1}(t(j)) \\ \sum\sum v_{tj}^{-1}(\hat{\beta}) X_{t-1}(t(j)) & \sum\sum v_{tj}^{-1}(\hat{\beta}) X_{t-1}^2(t(j)) \end{pmatrix} \begin{pmatrix} \sum\sum v_{tj}^{-1}(\hat{\beta}) X_t(j) \\ \sum\sum v_{tj}^{-1}(\hat{\beta}) X_t(j) X_{t-1}(t(j)) \end{pmatrix}. \tag{5.13}$$

Following similar arguments as for Section 5.1, one can derive the limit distribution of the (modified) quasilikelihood estimates given by

Theorem 5.2. *As $n \to \infty$, we conclude*

$$\left(\sum Z_t\right)^{1/2} \begin{pmatrix} \tilde{\beta}_0 - \beta_0 \\ \tilde{\beta}_1 - \beta_1 \end{pmatrix} \xrightarrow{d} N(0, B^{-1}) \tag{5.14}$$

where the elements b_{11}, b_{12} and b_{22} are determined by

$$\left(\sum Z_t\right)^{-1} \sum\sum v_{tj}^{-1}(\beta) \xrightarrow{a.s.} b_{11}$$

$$\left(\sum Z_t\right)^{-1} \sum\sum v_{tj}^{-1}(\beta) X_{t-1}(t(j)) \xrightarrow{a.s.} b_{12}$$

$$\left(\sum Z_t\right)^{-1} \sum\sum v_{tj}^{-1}(\beta) X_{t-1}^2(t(j)) \xrightarrow{a.s.} b_{22}.$$

It will be useful to identify b's for a specific example, say, random coefficient AR(1) process defined in (2.8) for which $E(\beta_t) = \beta_1$ and $\mathrm{Var}(\beta_t) = \sigma_\beta^2$. Then, $v_{tj}(\beta) = \sigma_\beta^2 X_{t-1}^2(t(j)) + \sigma_\varepsilon^2$ and in turn $b_{11} = E\left[v_{tj}^{-1}(\beta)\right]$, $b_{12} = E\left[v_{tj}^{-1}(\beta)X_{t-1}(t(j))\right]$ and $b_{22} = E\left[v_{tj}^{-1}(\beta)X_{t-1}^2(t(j))\right]$. In particular when β_t degenerates at β_1, i.e., $\sigma_\beta^2 = 0$, note that B reduces to $\sigma_\varepsilon^{-2}A$ appearing in Theorem 5.1. For the binomial thinning model presented in (2.9), one can use for obtaining B,

$$v_{tj}(\beta) = \beta_1(1-\beta_1)X_{t-1}^2(t(j)) + \sigma_\varepsilon^2, \quad 0 < \beta_1 < 1$$

where β_1 represents the success probability associated with the sequence of iid Bernoulli random variables.

Note that one can replace the random norm in (5.14) by the non-random norm as in (5.9) to get a mixture of bivariate normals as the limit distribution.

5.3. Maximum likelihood estimation

First, consider the general branching Markov process $\{Z_t, X_t(j)\}$ defined by (A.1) and (A.2) in Section 2. Suppose the transition density $p(x_t(j) \mid x_{t-1}(t(j)))$ depends on a parameter vector β. It follows from (A.1) and (A.2) that the likelihood function based on the sample $\{Z_t, X_t(j);\ t = 1, \ldots, n, j = 1, \ldots, Z_t\}$ is given by

$$L_n(\beta) = p(x_0(1)) \prod_{t=1}^{n} \left[p(z_t \mid z_{t-1}) \prod_{j=1}^{z_t} p_\beta(x_t(j) \mid x_{t-1}(t(j))) \right]. \tag{5.15}$$

The likelihood score function $l_n(\beta)$ is given by

$$l_n(\beta) = \partial \log L_n(\beta)/\partial\beta = \sum\sum \partial \log p_\beta(x_t(j) \mid x_{t-1}(t(j)))/\partial\beta \tag{5.16}$$

since $p(z_t \mid z_{t-1})$ does not depend on β. We now suppose that the transition density belongs to a conditional exponential family having density specified in (2.10) of Ex.3. It is noted that $\mu_t(j)$ and $v_t(j)$ can be obtained by successive differentiation of $k(\cdot)$, viz.,

$$\mu_t(j) = \partial k/\partial\eta_t(j) \quad \text{and} \quad v_t(j) = \partial^2 k/\partial\eta_t^2(j) = \partial\mu_t(j)/\partial\eta_t(j).$$

The conditional exponential family includes a large number of examples such as conditional Poisson, gamma, beta, normal, etc. Refer to [21] for a background on conditional exponential families. The likelihood score function reduces to

$$l_n(\beta) = \sum\sum v_t^{-1}(j)\left(\frac{\partial\mu_t(j)}{\partial\beta}\right)(X_t(j) - \mu_t(j)). \tag{5.17}$$

Note that the likelihood score here has the same form as the quasilikelihood score given in (5.10), due to the special structure of the conditional exponential family. It is assumed that $\mu_t(j)$ (and hence $v_t(j)$) is a function of a parameter vector β and then use the notation $\mu_{tj}(\beta)$ and $v_{tj}(\beta)$ for $\mu_t(j)$ and $v_t(j)$. The conditional information matrix $F_n(\beta)$ corresponding to $l_n(\beta)$ in (5.17) is given by

$$F_n(\beta) = \sum\sum\left(\frac{\partial\mu_{tj}(\beta)}{\partial\beta}\right)\left(\frac{\partial\mu_{tj}(\beta)}{\partial\beta}\right)^{\mathrm{T}} v_{tj}^{-1}(\beta). \tag{5.18}$$

One can readily show that by Theorem 3.1, as $n \to \infty$

$$\left(\sum Z_t\right)^{-1} F_n(\beta) \xrightarrow{a.s.} F(\beta) \tag{5.19}$$

where $F(\beta) = E\left[\left(\frac{\partial\mu_{tj}(\beta)}{\partial\beta}\right)\left(\frac{\partial\mu_{tj}(\beta)}{\partial\beta}\right)^{\mathrm{T}} v_{tj}^{-1}(\beta)\right]$ which is assumed to be positive definite. Furthermore, by Theorem 4.1, it can be verified that

$$\left(\sum Z_t\right)^{-1/2} l_n(\beta) \xrightarrow{d} N(0, F(\beta)). \tag{5.20}$$

Under regularity conditions similar to those in [22], Theorem 2.1), one can establish that with probability tending to one, there exists a consistent solution $\hat\beta_{ML}$ of the likelihood equation $l_n(\beta) = 0$. We are now in a position to address the asymptotic distribution of $\hat\beta_{ML}$.

Theorem 5.3. *As $n \to \infty$, we have*

(i) $\left(\sum Z_t\right)^{1/2}(\hat\beta_{ML} - \beta) \xrightarrow{d} N(0, F^{-1}(\beta))$

(ii) $\left(\frac{m^{n+1}}{m-1}\right)^{1/2}(\hat\beta_{ML} - \beta) \xrightarrow{d} W^{-1/2}N(0, F^{-1}(\beta))$.

Proof. Using Taylor's expansion of $l_n(\beta)$ about $\beta = \hat{\beta}_{ML}$, one can obtain

$$\left(\sum Z_t\right)^{1/2}(\hat{\beta}_{ML} - \beta) = \left[\left(\sum Z_t\right)^{-1} F_n(\beta)\right]^{-1}\left[\left(\sum Z_t\right)^{-1/2} l_n(\beta)\right] + o_p(1)$$

where $o_p(1)$ denotes a term converging to zero in probability. The result (i) follows by (5.19) and (5.20). Note that $\sum Z_t \sim \left(\frac{m^{n+1}}{m-1}\right) W$ (a.s.), leading to (ii). □

It is worth mentioning that the BMP model possesses the local asymptotic mixed normality (LAMN) in the sense that the asymptotic distribution of the maximum likelihood estimator is a mixture of normals as in (ii) of Theorem 5.3 and therefore BMP model belong to the non-ergodic class of stochastic processes due to [23]. Statistical applications on this aspect is now under investigation and will be addressed elsewhere.

References

[1] T.E. Harris, Branching Processes, Springer, New York, 1963.
[2] K.B. Athreya, P.E. Ney, Branching Processes, Berlin, Springer, 1972.
[3] P. Guttorp, Statistical Inference for Branching Processes, Wiley, 1991.
[4] R. Cowan, R.G. Staudte, The bifurcating autoregression model in cell lineage studies, Biometrics 42 (1986) 769–783.
[5] R.M. Huggins, I.V. Basawa, Extensions of the bifurcating autoregressive model for cell lineage data, J. Appl. Probab. 36 (1999) 1225–1233.
[6] R.M. Huggins, I.V. Basawa, Inference for extended bifurcating autoregressive model for cell lineage studies, Aust. N. Z. J. Stat. 42 (2000) 423–432.
[7] J. Zhou, I.V. Basawa, Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors, J. Time Ser. Anal. 26 (2005) 825–842.
[8] J. Zhou, I.V. Basawa, Least squares estimation for bifurcating autoregressive processes, Stat. Probab. Lett. 74 (2005) 77–88.
[9] I.V. Basawa, J. Zhou, Non-Gaussian bifurcating models and quasi-likelihood estimation, J. Appl. Probab. (2004) 55–64. Spec. 41A.
[10] R.M. Huggins, R.G. Staudte, Variance components models for dependent cell populations, J. Amer. Statist. Assoc. 89 (1994) 19–29.
[11] A. Chen, Ergodicity and stability of generalized Markov branching processes with resurrection, Journal of Applied Probability 39 (2002) 786–803.
[12] G.K. Grunwald, R.J. Hyndman, L. Tedesco, R.L. Tweedie, Non-Gaussian conditional linear AR(1) models, Aust. N. Z. J. Stat. 42 (2000) 479–495.
[13] D.F. Nicholls, B.G. Quinn, Random Coefficient Autoregressive Models: An Introduction, Springer, New York, 1982.
[14] R.L. Tweedie, Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space, Stochastic Process. Appl. 3 (1975) 385–403.
[15] P.D. Fiegin, R.L. Tweedie, Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments, J. Time Ser. Anal. 6 (1985) 1–14.
[16] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer, Berlin, 1993.
[17] P.G. Hall, C.C. Heyde, Martingale Limit Theory and Its Applications, Academic Press, New York, 1980.
[18] P. Billingsley, Convergence of Probability Measures, Wiley, 1968.
[19] B.M. Brown, Martingale central limit theorems, Ann. Math. Statist. 42 (1971) 59–66.
[20] C.C. Heyde, Quasi-likelihood and Its Applications, Springer, New York, 1997.
[21] U. Küchler, M. Sorensen, Exponential Families of Stochastic Processes, Springer Series in Statistics, Springer, 1997.
[22] P. Billingsley, Statistical Inference for Markov Processes, The Univ. of Chicago Press, 1961.
[23] I.V. Basawa, D.J. Scott, Asymptotic Optimal Inference for Non-Ergodic Models, Springer, New York, 1983.