Statistical Methodology 6 (2009) 61–69

Contents lists available at ScienceDirect

Statistical Methodology

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

Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference

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

[a] *Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Republic of Korea*

[b] *Department of Statistics, University of Georgia, Athens, GA, USA*

ARTICLE INFO

Article history:
Received 31 January 2007
Received in revised form
26 January 2008
Accepted 30 March 2008

Keywords:
Bifurcating model
LAN
Martingale array
Maximum likelihood
Score test

ABSTRACT

This article is concerned with the local asymptotic normality (LAN) of the log-likelihood for the bifurcating autoregressive model (BAR) for tree structured data where each individual in one generation gives rise to two off-spring in the next generation. We derive the LAN property for the pth-order BAR model. Asymptotic optimal inference for the model parameters can be deduced as a consequence of LAN. In particular, an efficient score test is derived as an application. A simulation study is conducted to address the issue regarding how many generations are required for asymptotic results to be useful in practice.

1. Introduction

The bifurcating autoregressive model (BAR, for short) was developed to analyze binary-splitting tree structured data where each individual gives rise to two off-spring in the next generation. Cowan and Staudte [3] proposed the first-order model (BAR(1)) for cell lineage data where each line of descendants follows an AR(1) process and observations on the sister cells from the same parent are allowed to be correlated. Let X_t be an observation of some characteristics of individual t. Beginning with the starting value X_1, the BAR(1) process $\{X_t, t = 1, 2, \dots\}$ is defined as, for $t \geq 2$

$$X_t = \phi_0 + \phi_1 X_{[t/2]} + \epsilon_t, \quad |\phi_1| < 1, \tag{1.1}$$

where $[\cdot]$ denotes the greatest integer function so that one can write recursively $X_2 = \phi_0 + \phi_1 X_1 + \epsilon_2$, $X_3 = \phi_0 + \phi_1 X_1 + \epsilon_3$, $X_4 = \phi_0 + \phi_1 X_2 + \epsilon_4$, etc. For illustration, the data structure with four generations is given in Fig. 1.

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

1572-3127/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2008.03.002

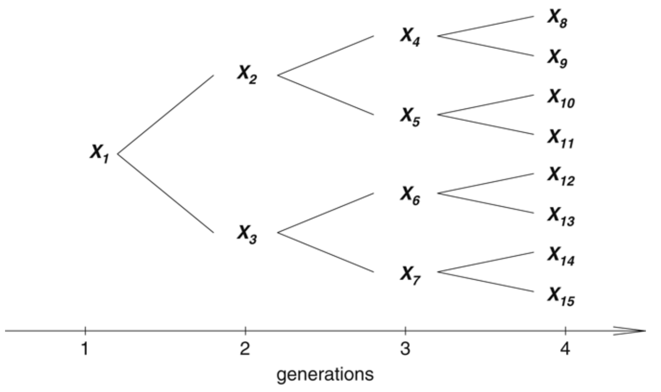

Fig. 1. Tree structured data with four generations.

It is noted that $[\log_2 t] + 1$ denotes the generation of individual t. Here $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of iid bivariate normal random vectors with common mean zero, common variance $\sigma^2 > 0$ and Corr $(\epsilon_{2t}, \epsilon_{2t+1}) = \rho$. It is noted that each line of descendants follows an AR(1) process and cousin cells within each generation are conditionally independent given the observations in the previous generations. For statistical analysis of BAR(1) model, refer to [9,1,13].

Huggins and Basawa [7] noted data sets where correlation between cousins is significantly larger than that predicted by a BAR(1) model and thus they extended the BAR(1) model to higher-order models in order to accommodate extended correlation structure. The BAR(p) process is specified by the recursive equation

$$X_t = \phi_0 + \phi_1 X_{[t/2]} + \cdots + \phi_p X_{[t/2^p]} + \epsilon_t, \quad t \geq 2^p. \tag{1.2}$$

Huggins and Basawa [8] derived the asymptotic distribution of the maximum likelihood estimators of model parameters in BAR(p) assuming *Gaussian* $\{\epsilon_t\}$. Least squares estimation for BAR(p) was recently discussed by Zhou and Basawa [14]. We are here concerned with addressing asymptotic optimal inference for BAR(p) processes. In this article $\{\epsilon_t\}$ is not necessarily Gaussian and the local asymptotic normality (LAN) of the log-likelihood will be established, from which asymptotic optimality of tests and estimators under various criteria can be deduced. Refer to [5,10,2] among others, for comprehensive discussions on LAN.

The model and preliminary results are addressed in Section 2. The main result (LAN) is presented in Section 3. In Section 4, Rao's score test for Cowan–Staudte's first-order model, for testing the composite hypothesis

$$H_0 : \phi_2 = \phi_3 = \cdots = \phi_p = 0 \tag{1.3}$$

is derived and its asymptotic efficiency is discussed as an application of the LAN property.

2. The model and preliminary results

Consider the BAR(p) process defined in (1.2). It will be assumed that $\{(\epsilon_{2t}, \epsilon_{2t+1}), t \geq 1\}$ is a sequence of iid zero mean bivariate random vectors with bivariate density $f(\epsilon_{2t}, \epsilon_{2t+1})$ with variance–covariance structure given by

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \sigma^2. \tag{2.1}$$

The bivariate density $f(\epsilon_{2t}, \epsilon_{2t+1})$ is not necessarily Gaussian. $f(\epsilon_{2t}, \epsilon_{2t+1})$ may have bivariate t-distribution, mixed normal distribution, logistic distribution (cf. [11]), bivariate Gamma and bivariate Poisson distributions with mean centered at zero. See, for instance, [1]. Notice that BAR(p) specifies