Communications of the Korean Statistical Society
2010, Vol. 17, No. 3, 441–450

Markov Chain Approach to Forecast in the Binomial Autoregressive Models

Hee-Young Kim[1,a], Yousung Park[b]

[a]Institute of Economics, Korea University; [b]Department of Statistics, Korea University

Abstract

In this paper we consider the problem of forecasting binomial time series, modelled by the binomial autoregressive model. This paper considers proposed by McKenzie (1985) and is extended to a higher order by Weiß (2009). Since the binomial autoregressive model is a Markov chain, we can apply the earlier work of Bu and McCabe (2008) for integer valued autoregressive(INAR) model to the binomial autoregressive model. We will discuss how to compute the h-step-ahead forecast of the conditional probabilities of X_{T+h} when T periods are used in fitting. Then we obtain the maximum likelihood estimator of binomial autoregressive model and use it to derive the maximum likelihood estimator of the h-step-ahead forecast of the conditional probabilities of X_{T+h}. The methodology is illustrated by applying it to a data set previously analyzed by Weiß (2009).

Keywords: Binomial thinning, binomial AR(p) model, Markov chain.

1. Introduction

When the observed time series are low frequency count, the autoregressive moving average(ARMA) models can not be applied to integer-valued case. The reason is that the multiplication of an integer by a real number generally gives noninteger value. There are models using thinning operations in place of scalar multiplication in ARMA models to obtain an ARMA-like autocorrelation structure. In particular, the binomial thinning operation developed by Steutel and van Harn (1979) is the most popular thinning operation which is used to model integer-valued time series data. The purpose of the binomial thinning is to ensure the integer discreteness of the process. It is defined as $\alpha \circ X = \sum_{i=1}^{X} Y_i$, where Y_i are assumed to be *i.i.d.* Bernoulli random variables with $P(Y_i = 1) = \alpha$, $P(Y_i = 0) = 1 - \alpha$, and independent of X. The first integer-valued ARMA(INARMA) model using binomial thinning operation is INAR(1) model which is introduced by McKenzie (1985) and independently by Al-Osh and Alzaid (1987). For references on recent developments on INARMA models, refer to McKenzie (2003), Weiß (2008), Kim and Park (2008) and Park *et al.* (2006).

The objective of this paper is the forecasting issue of time series of binomial counts. Even if binomial distribution is very simple, it can not be marginal distribution of INAR(1) model, and for the reason, see Weiß (2008). The binomial AR(1) model suggested by McKenzie (1985) is designed for only binomial counts, still using binomial thinning operation and Weiß (2009) proposed binomial AR(p) model. Binomial AR(p) model is useful tool to model for time series of binomial counts, because it has an AR(p)-like autocorrelation structure and the model order p can be identified with the aid of the partial autocorrelation function.

This work was supported by the Korea Research Foundation grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund) (KRF-2007-411-J03302).

[1] Corresponding author: Research Professor, Institute of Economics, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-701, Korea. E-mail: starkim@korea.ac.kr

A practical problem in forecasting in the integer valued time series models using binomial thinning operation is the use of binomial thinning operation, which makes X_t into the convolution of several components. There are numerous efforts to forecast integer valued time series model using a thinning operation to resolve theses problems and to produce coherent forecasting: see Freeland and McCabe (2004) and Jung and Tremayne (2006) and Kim and Park (2006a,b, 2010) and Bu and McCabe (2008). Of theses articles, Bu and McCabe (2008) produced forecasting in INAR(p) model by treating the model as a Markov chain.

In this paper, we apply Bu and McCabe (2008)'s idea to the binomial AR(p) model to build confidence intervals for the probabilities of all possible values taken the process in the future, since the binomial AR(p) model can be viewed as a Markov chain.

This paper is organized as follows. Section 2 reviews the binomial AR(1) and the binomial AR(p) model, and Section 3 describes a method for producing h-step-ahead forecasts of the conditional probability distribution of the binomial AR(p) process $\{X_t\}$. The h-step-ahead forecasts of conditional probability distribution is calculated by Markov chain approach and the estimation is done in maximum likelihood framework. The forecasting procedure is illustrated with an empirical example in Section 4. In addition, we consider a proportional odds model as a competitor of binomial AR model (which is a regression model for a categorical time series) and we compare the fitted results. Section 5 outlines the conclusion and results.

2. The Binomial Autoregressive Models

Time series models for binomial marginal distribution are presented here. The first order dependence is modelled by binomial AR(1) model proposed by McKenzie (1985) and the higher order dependence is modelled by binomial AR(p) model of Weiß (2009). In this section we briefly summarize the basic properties of these models, for obtaining a deeper understanding through the graphical illustration of these models, refer to Kim and Park (2010).

2.1. The Binomial AR(1) model

Definition 1. *(Binomial AR(1) model, McKenzie, 1985) Fix* $n \in \mathbb{N}$*. Let* $\pi \in (0, 1)$*, and* $\rho \in [\max(-\pi/(1-\pi), -(1-\pi)/\pi), 1]$*. Define* $\beta = \pi(1-\rho)$ *and* $\alpha = \beta + \rho$*. The process* $\{X_t\}$*,*

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}), \quad t \geq 1, \; X_0 \sim B(n, \pi) \tag{2.1}$$

is said to be a Binomial AR(1) model, where all thinnings are performed independently of each other, and the thinnings at time t are independent of $\{X_s, s < t\}$*.*

Suppose there are n units which are independent of each other, either in state 1 or state 0 and X_t is the number of units being in state 1 at time t. Then binomial AR(1) model defines the number of units being in state 1 at time t to be sum of two random variable, the number of units which are still in state 1 at time t, with individual transition probability α, and the number of units which moved from state 0 to state 1 at time t, with individual transition probability β. The former is $\alpha \circ X_{t-1}$ and the latter is $\beta \circ (n - X_{t-1})$.

The marginal distribution of binomial AR(1) model is $B(n, \pi)$, and so $E(X_t) = n\pi$, $\text{Var}(X_t) = n\pi(1-\pi)$. The autocorrelation function(ACF) of $\{X_t\}$ as defined by Equation (2.1) is given by $\rho(k) = \rho^k$ for $k = 1, 2, \ldots,$ is identical to the ACF of a usual AR(1) process, also permitting negative autocorrelation. This property is good characteristic of binomial AR(1) model compared to INAR(1) model using the same binomial thinning operation.

The conditional probability is

$$P(X_t = k|X_{t-1} = l) = \sum_{m=\max(0,k+l-n)}^{\min(k,l)} \binom{l}{m}\binom{n-l}{k-m}\alpha^m(1-\alpha)^{l-m}(1-\beta)^{n-l+m-k}. \tag{2.2}$$

And the conditional expectation and the conditional variance are respectively

$$E(X_t|X_{t-1}) = \rho X_{t-1} + n\beta \quad \text{and} \quad \mathrm{Var}(X_t|X_{t-1}) = \rho(1-\rho)(1-2\pi)X_{t-1} + n\beta(1-\beta).$$

2.2. Binomial AR(p) model

Since the introduction of first-order model, it took a long time to develop higher order model. More recently, a higher-order autoregressive model is proposed by Weiß (2009) as a tool for modelling and generating sequences of dependent binomial process.

Definition 2. *(Binomial AR(p) model, Weiß, 2009) Let* $\pi \in (0,1)$ *and* $\rho \in [\max(-\pi/(1-\pi), -(1-\pi)/\pi), 1]$. *Define* $\beta = \pi(1-\rho)$ *and* $\alpha = \beta + \rho$. *Let* $\{\mathbf{D}_t\}$ *be an i.i.d. multinomial distribution with parameters* $\mathbf{D}_t = (D_{t,1}, \ldots, D_{t,p}) \sim MULT(1; \phi_1, \ldots, \phi_p)$. *Let a process* $\{X_t\}$ *with range* $\{0, \ldots, n\}$ *follow the recursion*

$$\begin{aligned} X_t &= D_{t,1}(\alpha \circ_t X_{t-1} + \beta \circ_t (n - X_{t-1})) + \cdots + D_{t,p}(\alpha \circ_t X_{t-p} + \beta \circ_t (n - X_{t-p})) \\ &= \sum_{i=1}^{p} D_{t,1} f_i(X_{t-i}), \end{aligned} \tag{2.3}$$

where " $\circ_{t+k}$ *" indicates that the thinning is performed at time* $t+k$ *and* $f_k(X_t)$ *is a random function defined in the form* $\alpha \circ_{t+k} X_t + \beta \circ_{t+k} (n - X_t)$.

If conditions $C1 \sim C3$ *are satisfied, it is said to be a binomial AR(p) process. And in addition to* $C1 \sim C3$, *if C4 is satisfied, then it is called a binomial AR(p)-Independent thinning process.*

(C1) The thinnings at time t are performed independently of each other and of $\{\mathbf{D}_t\}$.

(C2) $\mathbf{D}_t = (D_{t,1}, \ldots, D_{t,p})$ *is independent of all* $\{X_s, s < t\}$ *and* $\{f_j(X_s), s < t, j = 1, \ldots, p\}$.

(C3) The conditional probability

$$\begin{aligned} &P\Big(f_1(X_t) = i_1, \ldots, f_p(X_t) = i_p | X_t = x_t, X_{t-k} = x_{t-k}, k \geq 1; f_j(X_{t-k}) = z_{t-k}, k \geq 1, j = 1, \ldots, p\Big) \\ &= P\Big(f_1(X_t) = i_1, \ldots, f_p(X_t) = i_p | X_t = x_t\Big). \end{aligned}$$

(C4) All thinnings $f_1(X_t), \ldots, f_p(X_t)$ *are conditionally independent, conditioned on* X_t.

Weiß (2009) showed that the ACF of binomial AR(p)-Independent thinnings process satisfies the usual form of the Yule-Walker equations for usual AR(p): $\rho(k) = \rho \sum_{i=1}^{p} \phi_i \rho(|k-i|)$, and the partial autocorrelation function(PACF) is $\phi_{kk} = 0, k > p$. The conditional of binomial AR(p)-Independent thinnings process is

$$\begin{aligned} &P\Big(X_t = x | X_{t-1} = x_{t-1}, \ldots, X_{t-p} = x_{t-p}\Big) \\ &= \sum_{i=1}^{p} \phi_i \sum_{y=0}^{x} \binom{x_{t-i}}{y} \alpha^y (1-\alpha)^{x_{t-i}-y} \binom{n - x_{t-i}}{x-y} \beta^{x-y} (1-\beta)^{n-x_{t-i}-x+y}. \end{aligned} \tag{2.4}$$

And the conditional expectation of binomial AR(p)-Independent thinnings process is

$$E(X_t|X_{t-1},\ldots,X_{t-p}) = n\pi(1-\rho) + \rho\sum_{i=1}^{p}\phi_i X_{t-i}.$$

3. Forecasting the Conditional Distribution of *h*-Step-Ahead

Subsection 3.1 presents method for computing the h-step-ahead forecast of the conditional probabilities of X_{T+h} given $X_T, X_{T-1},\ldots,X_{T-p+1}$ when the parameters of the model are known. Subsection 3.2 deals real situation of unknown parameters, so how to compute the likelihood function of binomial AR(p) model and derive the asymptotic distribution of the MLE of the h-step-ahead forecast of the conditional probabilities. Our approach is based on Bu and McCabe (2008) in INAR(p) model.

3.1. A Markov chain approach: Parameters are known

We first define a vector time series $\{\boldsymbol{Y}_t := (X_{t-p+1}, X_{t-p+2},\ldots,X_t)'\}$ from a binomial AR(p) process $\{X_t\}$. Then it follows that $\{\boldsymbol{Y}_t\}$ is a first-order Markov chain on a finite state space $\mathbb{S}^p$, where $\mathbb{S}$ is the state space of $\{X_t\}$. For example, for binomial AR(p) process $\{X_t\}$, the state space $\mathbb{S}$ is $\{0,\ldots,n\}$, and for $p=2$, $\mathbb{S}^p=\mathbb{S}^2$ is

$$\mathbb{S}^2 = \left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\ldots,\begin{pmatrix}0\\n\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix},\ldots,\begin{pmatrix}n\\n\end{pmatrix}\right\}.$$

And the transition probabilities of making a transition from one state at time $t-1$ to another state at time t are given by

$$\begin{aligned} P\left(\boldsymbol{Y}_t = (l,k)' \,|\, \boldsymbol{Y}_{t-1} = (j,i)'\right) &= P\left(X_{t-1} = l, X_t = k \,|\, X_{t-2} = j, X_{t-1} = i\right) \\ &= P\left(X_t = k, X_{t-1} = l \,|\, X_{t-1} = i, X_{t-2} = j\right) \\ &= \begin{cases} P\left(X_t = k \,|\, X_{t-1} = l, X_{t-2} = j\right), & \text{if } l = i, \\ 0, & \text{if } l \neq i, \end{cases} \end{aligned} \tag{3.1}$$

where $(l,k)'$ and $(j,i)' \in \mathbb{S}^2$. The probability of Equation (3.1) is given by Equation (2.4) with $p=2$, *i.e.*, $P(X_t = x|X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}) = \sum_{i=1}^{2}\phi_i\sum_{y=0}^{x}\binom{x_{t-i}}{y}\alpha^y(1-\alpha)^{x_{t-i}-y}\binom{n-x_{t-i}}{x-y}\beta^{x-y}(1-\beta)^{n-x_{t-i}-x+y}$.

Therefore we can make one-step transition probability matrix of process $\boldsymbol{Y}_t = (X_{t-p+1}, X_{t-p+2},\ldots, X_t)'$ using the above scheme and denote it as $\boldsymbol{Q}_{p,n}$. In the case of $p=2, n=2$, the matrix $\boldsymbol{Q}_{2,2}$ can be written as

$$\boldsymbol{Q}_{2,2} = \begin{array}{cc} & X_t \\ X_{t-2} & X_{t-1} \\ 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \end{array} \begin{array}{ccccccccc} 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \end{array} \atop \left[\begin{array}{ccccccccc} p(0|0,0) & p(1|0,0) & p(2|0,0) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p(0|1,0) & p(1|1,0) & p(2|1,0) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p(0|2,0) & p(1|2,0) & p(2|2,0) \\ p(0|0,1) & p(1|0,1) & p(2|0,1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p(0|1,1) & p(1|1,1) & p(2|1,1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p(0|2,1) & p(1|2,1) & p(2|2,1) \\ p(0|0,2) & p(1|0,2) & p(2|0,2) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p(0|1,2) & p(1|1,2) & p(2|1,2) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p(0|2,2) & p(1|2,2) & p(2|2,2) \end{array}\right],$$

where $P(i|j,k)$ is denote $P(X_t = i|X_{t-1} = j, X_{t-2} = k)$. These results may be easily extended to arbitrary p and n, so we can obtain the one-step transition probability matrix $\boldsymbol{Q}_{p,n}$ as a $(n+1)^p \times (n+1)^p$ matrix.

To represent the probabilities of finding a system in each of the $(n+1)^p$ different states at a given time t, define $(n+1)^p \times 1$ vector π_t. For example, $p = 2, n = 2$, the vector π_t is

$$\pi_t = \left(P_t^{0,0}, P_t^{0,1}, P_t^{0,2}, P_t^{1,0} P_t^{1,1}, P_t^{1,2}, P_t^{2,0}, P_t^{2,1}, P_t^{2,2}\right), \tag{3.2}$$

where $P_t^{i,j}$ is $P(X_{t-1} = i, X_t = j)$.

And for each $i \in \{0, 1, \ldots, n\}$, define $(n+1)^p \times 1$ vector v_i, which has $n+1$ entries equal to one and all others equal to zero. In the case of $p = 2$, $n = 2$, $v_0 = (1, 0, 0, 1, 0, 0, 1, 0, 0)^{'}$, $v_1 = (0, 1, 0, 0, 1, 0, 0, 1, 0)^{'}$ and $v_2 = (0, 0, 1, 0, 0, 1, 0, 0, 1)^{'}$.

To compute $P(X_{T+h} = i \,|\, X_T, \ldots, X_{T-p+1})$, $i = 0, \ldots, n$, we require the Markov chain theory: (i) the h-step transition probability matrix is equal to the h^{th} power of the one-step transition matrix, *i.e.*, $\boldsymbol{Q}^{(h)} = \boldsymbol{Q}^h$, (ii) the h-step-ahead forecast of the probability vector is equal to the current probability vector times the h-step transition probability matrix, *i.e.*, $\pi_{T+h} = \pi_T \boldsymbol{Q}^{(h)}$. For further details on the Markov chain theory, see Meyn and Tweedie (1993) and Kemeny and Snell (1976). Therefore, the h-step-ahead forecast of the conditional probability of X_{T+h} given $X_T, X_{T-1}, \ldots, X_{T-p+1}$ is written as follows

$$P\left(X_{T+h} = i \,|\, X_T, \ldots, X_{T-p+1}\right) = \pi_{T+h}^{'} v_i \tag{3.3}$$

$$= \pi_T^{'} \boldsymbol{Q}^h v_i, \tag{3.4}$$

where $i = 0, \ldots, n$.

3.2. A Markov chain approach: Parameters are unknown

Conditioning on the first p observations leads to a simple form of the likelihood $L(\phi_1, \phi_2, \ldots, \phi_{p-1}, \alpha, \beta) = \prod_{t=p+1}^{T} P(X_t \,|\, X_{t-1}, \ldots, X_{t-p+1})$ and so knowledge of the transition probabilities is sufficient for its construction. And we know the explicit formula of transition probabilities of binomial AR(p) model from Equation (2.4). Under the standard regularity conditions, the maximum likelihood estimator(MLE) of $\boldsymbol{\theta} = (\phi_1, \phi_2, \ldots, \phi_{p-1}, \alpha, \beta)^{'}$ is asymptotically normally distributed around the true parameters $\boldsymbol{\theta}_0$, that is $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{i}^{-1})$, where $\boldsymbol{i}^{-1}$ is the inverse of the Fisher information matrix.

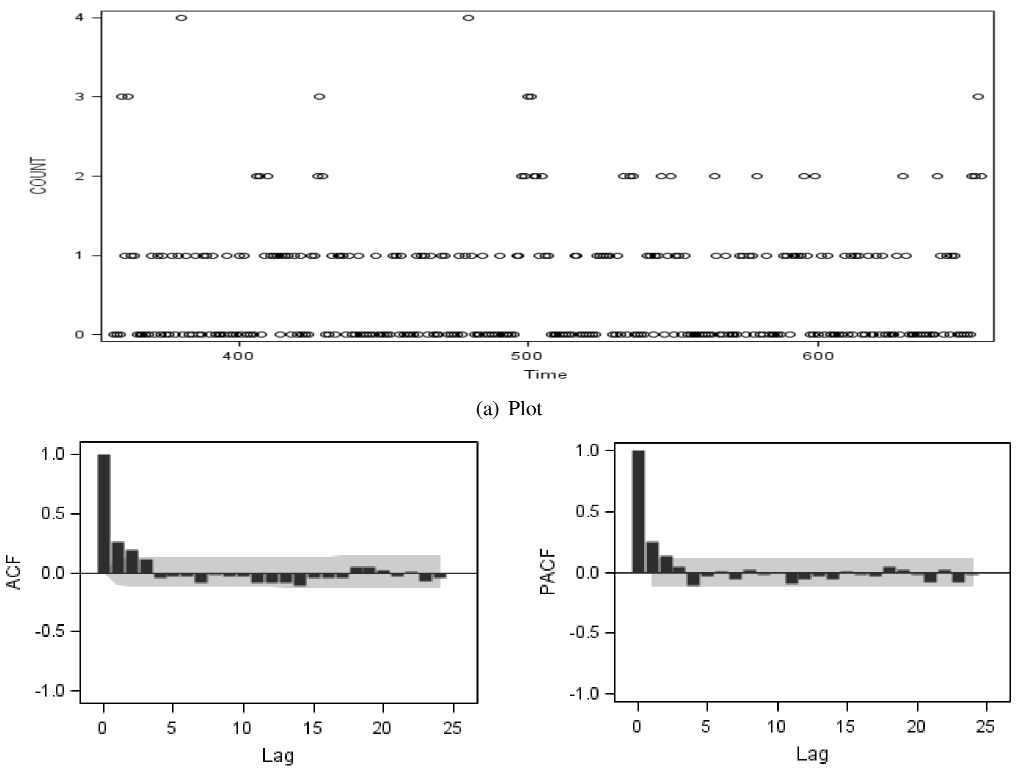
Denote the h-step-ahead forecast of the conditional probability function as

$$g_i(\boldsymbol{\theta}) := P_T(X_{T+h} = i \,|\, \boldsymbol{\theta}), \quad i = 0, \ldots, n$$

to emphasis its dependence on the parameters $\boldsymbol{\theta}$ and given observations of the process up to time T. And we already know from Equation (3.4) that $g_i(\boldsymbol{\theta})$ is given by $\pi_T^{'} \boldsymbol{Q}^h v_i$. As mentioned in the Section 3.1, $(n+1)^p \times 1$ two vectors π_T and v_i are composed of only one and zero and the elements of $(n+1)^p \times (n+1)^p$ matrix $\boldsymbol{Q}$ are given Equation (2.4).

Therefore, the asymptotic distribution of the MLE of the h-step-ahead forecast of the conditional probability function is easily obtained by the delta method as follows. Construct two $(n+1) \times 1$ vectors $\boldsymbol{g}(\hat{\boldsymbol{\theta}}) := (g_0(\hat{\boldsymbol{\theta}}), \ldots, g_n(\hat{\boldsymbol{\theta}}))^{'}$ and $\boldsymbol{g}(\boldsymbol{\theta}) := (g_0(\boldsymbol{\theta}), \ldots, g_n(\boldsymbol{\theta}))^{'}$. Then the MLE of the h-step ahead forecasts, $\boldsymbol{g}(\hat{\boldsymbol{\theta}})$, has asymptotically multivariate normal distribution

$$\sqrt{T}\left(\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})\right) = \sqrt{T}\left(\begin{pmatrix} g_0(\hat{\boldsymbol{\theta}}) \\ \vdots \\ g_n\left(\hat{\boldsymbol{\theta}}\right) \end{pmatrix} - \begin{pmatrix} g_0(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \vdots \\ g_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{D}\boldsymbol{i}^{-1}\boldsymbol{D}^{'}\right), \tag{3.5}$$



(a) Plot

(b) SACF

(c) SPACF

Figure 1: *Time series plot, sample ACF, and PACF of the access counts: (a) Time series plot; (b) sample ACF; (c) sample PACF.*

where $\boldsymbol{i}$ is the Fisher information matrix and $\boldsymbol{D} = \partial \boldsymbol{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is a matrix of partial derivatives. Therefore, the asymptotic $100(1-\alpha)\%$ confidence interval for $g_i(\boldsymbol{\theta})$, $i = 0, \ldots, n$ is given by

$$\hat{P}_T\left(X_{T+h} = i\right) \pm z_{\alpha/2}\sigma_{i+1}(\hat{\boldsymbol{\theta}}), \tag{3.6}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile point of a standard normal distribution, and $\sigma_{i+1}(\hat{\boldsymbol{\theta}})$ is the square root of $(i+1, i+1)$ element of estimated $\boldsymbol{T}^{-1}\boldsymbol{D}\boldsymbol{i}^{-1}\boldsymbol{D}'$.

4. Empirical Application

As an application of the results obtained above, we will apply the methods in Section 3 to the data used in Weiß (2009) who originally introduced binomial AR(p) model. The data are the number of access, say X_t, to the home directory of six servers of the Department of Statistics at the University of Würzburg for each minute. The analyzed data are collected on 29^{th} day of November, 2005 and consist of 661 observations. Evidently, X_t take value in $\{0, 1, \ldots, 6\}$.

We use observations $\{X_t : 357 \leq t \leq 656\}$ to estimate parameter vectors, leaving $\{X_t : 657 \leq t \leq 661\}$ to evaluate the out-of-sample forecast performance of the procedure. Therefore we use 300 observations sample size and the prediction horizon h are $h = 1, \ldots, 5$. The Figure 1(a) displays plot, sample autocorrelation function, and sample partial autocorrelation function of the 300 observations.

Note that Figure 1(b) and Figure 1(c) show a typical pattern of the AR model. Although the possible range of X_t is from 0 to 6, the observed counts are composed of 56.67% 0, 32.67% 1, 8.00% 2, 2.00% 3 and 0.67% 4. The last two observations of $\{X_t : 357 \le t \le 656\}$ are $X_{655} = 3, X_{656} = 2$ and observed value of 657–661 are $X_{657} = 1$, $X_{658} = 1$, $X_{659} = 0$, $X_{660} = 1$, $X_{661} = 0$.

Kim and Park (2010) tried to find the optimal order of binomial autoregressive model representing theses data, we accept the earlier work of them and so we estimate binomial AR(2) model which is defined via

$$X_t = D_{t,1}\Big(\alpha \circ_t X_{t-1} + \beta \circ_t (n - X_{t-1})\Big) + D_{t,2}\left(\alpha \circ_t X_{t-p} + \beta \circ_t (n - X_{t-p})\right),$$

where $(D_{t,1}, D_{t,2}) \overset{i.i.d.}{\sim} \text{MULT}(1; \phi_1, \phi_2)$.

In this article, it is more convenient to work with parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \phi_1)'$ than $(\pi, \rho, \phi_1)'$ in Weiß (2009). The model parameter are estimated by conditional ML(CML) which is obtained numerically by maximizing the conditional log-likelihood function from Equation (2.4). We utilize NLPTR optimization subroutine of SAS/IML procedure to obtain CML estimates, with Yule-Walker estimates as initial values and NLPFDD optimization subroutine of SAS/IML procedure to acquire estimated asymptotic standard error from the inverse of the Hessian. The CML estimates and the estimated asymptotic standard errors are $\hat{\alpha} = 0.3590995(0.0629169)$, $\hat{\beta} = 0.0686873(0.00821)$, $\hat{\phi}_1 = 0.5502303(0.1661743)$.

To obtain $100(1-\alpha)\%$ confidence interval for $P_{656}(X_{656+h} = i|X_{656} = 3, X_{655} = 2)$, $h = 1, \ldots, 5$ in Equation (3.6), we note the last two observations of $\{X_t : 357 \le t \le 656\}$ are $X_{655} = 3, X_{656} = 2$. This means that at $t = 656$ the system is in state (3,2) with probability 1. So we fix 49×1 vector π_T in Equation (3.2) as follows

$$\begin{aligned} \pi_{656} &= \left(P^{0,0}_{656}, \ldots, P^{3,1}_{656}, P^{3,2}_{656}, P^{3,3}_{656}, \ldots, P^{6,6}_{656}\right) \\ &= (\quad 0, \ldots, \quad 0, \quad 1, \quad 0, \ldots, \quad 0). \end{aligned}$$

Then we are ready to calculate $100(1-\alpha)\%$ confidence interval for $P_{656}(X_{656+h} = i\,|X_{656} = 3, X_{655} = 2)$, $i = 0, \ldots, 6; h = 1, \ldots, 5$. Table 1 shows point estimates for $h = 1, 2, \ldots, 5$. For $h = 1$, the largest estimated conditional probability is $P_{656}(X_{656+1} = 1) = 0.422616$ and at $t = 657$ observed value is 1. This phenomenon that the value taking the largest estimated conditional probability is identical with the observed value occurs at $h = 1, 2, 3, 5$.

Figure 2 displays the 95% confidence intervals for the one-, two, three-, four- and five-step ahead conditional probability taking the middle of the 95% confidence intervals as a point estimate of the probability in Table 1. We can notice that the confidence intervals at the different mass points are similar for $h = 4$ and 5. It is explained by the fact the Binomial AR(2) model is stationary, in stationary time series model, the forecast distributions approach to the marginal distributions as lead time h increases.

We compare the obtained results from binomial AR(2) model with the the fit using some recent studies by Fokianos and Kedem (2003) on categorical time-series to the access count data. For recent developments concerning regression theory for time series, refer to Fokianos and Kedem (2002, 2003). The cumulative logistic or proportional odds model is a model for ordinal categorical time series. Suppose we have a categorical time-series $\{X_t\}, t = 1, \ldots, T$, and let m be the number of categories. In access count data, X_t can take value in $\{0, 1, \ldots, 6\}$, but the observed counts $\{0, 1, \ldots, 4\}$ and so we set $m = 5$.

Table 1: Point estimates for the h-step ahead conditional probability for the access counts data

h	0	1	2	3	4	5	6
	$\hat{P}(X_{656+h} = i \mid X_{655} = 3, X_{656} = 2)$						
$h = 1$	0.2665665	0.4226160	0.2423446	0.0616543	0.0073097	0.0004020	0
$h = 2$	0.3878023	0.4094812	0.1680294	0.0315370	0.0030040	0.0001433	0
$h = 3$	0.4762983	0.3775630	0.1230050	0.0210403	0.0019923	0.0000991	0
$h = 4$	0.5109036	0.3635069	0.1072703	0.0167825	0.0014674	0.0000680	0
$h = 5$	0.5288321	0.3555084	0.0995128	0.0148453	0.0012448	0.0000556	0

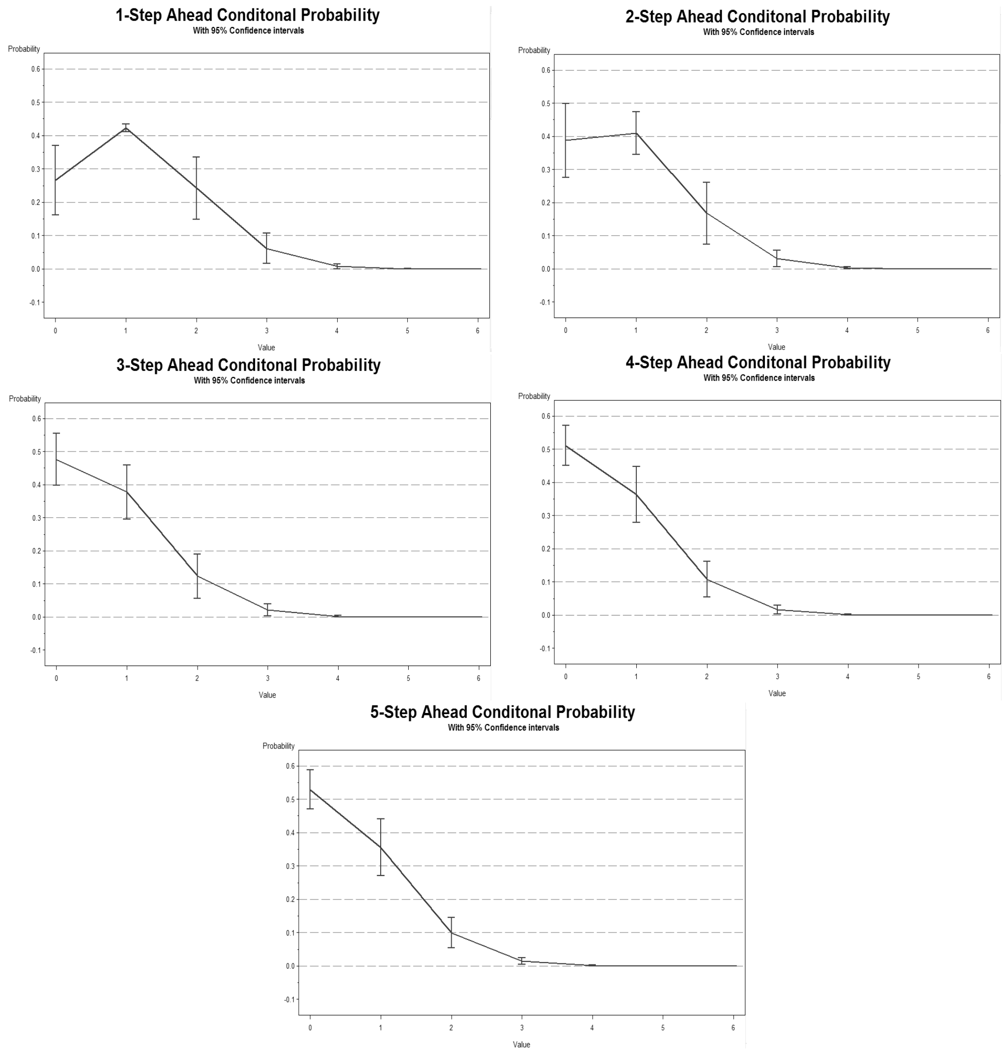


Figure 2: *The 95% confidence intervals for the one-, two-, three-, four- and five-step ahead conditional probability for the access counts data .*

We can express the t^{th} observation by the vector $\mathbf{X}_t = (X_{t0}, \ldots, X_{t3})'$ with elements

$$X_{tj} = \begin{cases} 1, & \text{if the } j^{th} \text{ category is observed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

Table 2: Results from Binomial AR(2) and Proportional odds models applied to the access counts data

Model	Number of parameters	AIC	BIC
Binomial AR(2) model	3	589.77719	600.88854
Proportional odds model of order 1	8	590.025	619.602
Proportional odds model of order 2	12	591.721	636.086

for $t = 1, \ldots, T$ and $j = 0, 1, \ldots, 3$.

The cumulative logistic or proportional odds model has the form

$$\log\left\{\frac{P(X_t \leq j\,|\,\mathcal{F}_{t-1})}{P(X_t > j\,|\,\mathcal{F}_{t-1})}\right\} = \theta_j + \gamma' \mathbf{z}_{t-1}, \quad i = 0, 1, \ldots, 3$$

in which θ_j are intercept parameters, γ are vector of parameters, $\mathbf{z}_{t-1}$ is a covariate vector of the same dimension of γ. The covariate vector $\mathbf{z}_{t-1}$ may consist of lagged values of the response process $\{\mathbf{X}_t\}$ and of any other auxiliary process known to the observed at time t. The σ-field $\mathcal{F}_{t-1}$ is generated by $\mathbf{z}_s,\ s \leq t-1$,

We fit the following two proportional odds models to the access count data: a first-order model given by

$$\log\left\{\frac{P(X_t \leq j|\mathcal{F}_{t-1})}{P(X_t > j|\mathcal{F}_{t-1})}\right\} = \theta_j + \gamma_1 X_{(t-1)0} + \gamma_2 X_{(t-1)1} + \gamma_3 X_{(t-1)2} + \gamma_4 X_{(t-1)3}, \quad j = 0, 1, \ldots, 3 \tag{4.1}$$

and a second-order model given by

$$\begin{aligned}\log\left\{\frac{P(X_t \leq j|\mathcal{F}_{t-1})}{P(X_t > j|\mathcal{F}_{t-1})}\right\} &= \theta_j + \gamma_1 X_{(t-1)0} + \gamma_2 X_{(t-1)1} + \gamma_3 X_{(t-1)2} + \gamma_4 X_{(t-1)3} \\ &\quad + \gamma_5 X_{(t-2)0} + \gamma_6 X_{(t-2)1} + \gamma_7 X_{(t-1)2} + \gamma_8 X_{(t-1)3}, \quad j = 0, 1, \ldots, 3.\end{aligned} \tag{4.2}$$

Table 2 reports the results of binomial AR(2) and two proportional odds models, where the second column lists the number of parameters in the model and the next two columns correspond to Akaike Information Criterion(AIC) and Bayesian information criterion(BIC). We see that the AIC criterion and the BIC criterion are minimized for the binomial AR(2) model. Therefore it seems reasonable to conclude that binomial AR(2) model is adequate for the access count data.

5. Conclusion

This study focused on the method of forecasting procedures in a binomial AR(p) model which is developed by Weiß (2009) to model for time series of binomial counts. Since the binomial AR(p) model can be regarded as a Markov chain, we applied the method introduced by Bu and McCabe (2008) to the binomial AR(p) model. We derived the h-step-ahead forecasts of conditional probability distribution using a Markov chin representation of the model, and obtained the MLE of those forecast mass function. We employed it to real data set which are the number of access times to the home directory of six server of the Department of Statistics of the University of Würzburg for each minute. Our analysis showed that its usefulness in binomial AR(p) model. In addition, we compared the fitted results of binomial AR(p) model and another regression time series model, *i.e.*, proportional odds model. The results indicate that binomial AR(p) model is more satisfactory to the access counts data.

References

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis,* **8**, 261–275.

Bu, R. and McCabe, B. (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based markov chain approach, *International Journal of Forecasting,* **24**, 151–162.

Fokianos, K. and Kedem, B. (2002). *Regression Models for Time Series Analysis*, Wiley,

Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series, *Statitical Science,* **18**, 357–376.

Freeland, R. K. and McCabe, B. P. M. (2004). Forecasting discrete valued low count time series, *International Journal of Forecasting,* **20**, 427–434.

Jung, R. C. and Tremayne, A. R. (2006). Coherent forecasting in integer time series models, *International Journal of Forecasting,* **22**, 223–238.

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov Chains*, Springer-Verlag, New York.

Kim, H.-Y. and Park, Y. (2006a). Bootstrap confidence intervals for the INAR(1) process, *The Korean Communications in Statistics*, **13**, 343–358.

Kim, H.-Y. and Park, Y. (2006b). Prediction mean squared error of the poisson INAR(1) process with estimated parameters, *Journal of the Korean Statistical Society,* **35**, 37–47.

Kim, H.-Y. and Park, Y. (2008). A non-stationary integer-valued autoregressive model, *Statistical Papers,* **49**, 485–502.

Kim, H.-Y. and Park, Y. (2010). Coherent forecasting in binomial AR(p) model, *The Korean Communications in Statistics,* **17**, 27–38.

McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resources Bulletin,* **21**, 645–650.

McKenzie, E. (2003). Discrete variate time series. In: Rao, C.R., Shanbhag, D.N. (eds.) Handbook of Statistics. Elsevier, Amsterdam, 573–606.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability,* Springer-Verlag, London.

Park Y, Choi, J. W. and Kim H.-Y. (2006). Forecasting cause-age specific mortality using two random processes, *Journal of the American Statistical Association,* **101**, 472–483.

Steutel, F. W. and van Harn, K. (1979). Discrete analogs of self-decomposability and stability, *Annals of Probability,* **7**, 893–899.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts-a survey, *Advances in Statistical Analysis,* **92**, 319–341.

Weiß, C. H. (2009) A new class of autoregressive models for time series of binomial count, *Communications in Statistics-Theory and Methods,* **38**, 447–460.

Received February 2010; Accepted April 2010