



Discrete-valued ARMA processes

Atanu Biswas^{a,*}, Peter X.-K. Song^b

^a Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India

^b Department of Biostatistics, School of Public Health, University of Michigan, United States

ARTICLE INFO

Article history:

Received 28 October 2008

Received in revised form 9 May 2009

Accepted 20 May 2009

Available online 12 June 2009

ABSTRACT

This paper presents a unified framework of stationary ARMA processes for discrete-valued time series based on Pegram's [Pegram, G.G.S., 1980. An autoregressive model for multilag markov chains. *J. Appl. Probab.* 17, 350–362] mixing operator. Such a stochastic operator appears to be more flexible than the currently popular thinning operator to construct Box and Jenkins' type stationary ARMA processes with arbitrary discrete marginal distributions. This flexibility allows us to yield an ARMA model for time series of binomial or categorical observations as a special case, which was unavailable with the extended thinning operator [Joe, H., 1996. Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* 33, 664–677] because the binomial/categorical distribution is not infinitely divisible. We also study parameter estimation and comparison with the thinning operator based method, whenever applicable. Real data examples are used to examine and illustrate the proposed method.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The theory of ARMA models (Box and Jenkins, 1976) has been playing a central role in modeling of continuous time series data arising in many practical fields. When the time series data are non-Gaussian, the ARMA models are no longer suitable. For example, a stationary AR(1) model for time series of counts is constructed via the so-called *thinning operator* (e.g. Steutel and van Harn (1979)).

Joe (1996) extended the thinning operator to a family of infinitely divisible distributions, including the Gaussian ARMA model as a special case. Also see Jørgensen and Song (1998). It is very unfortunate that the important cases of binomial and categorical processes are excluded from this class of models.

Discrete-valued stationary processes have been studied by Jacobs and Lewis (1978, 1983). However, their representation of discrete-valued ARMA processes differs from that of Box and Jenkins'. Kanter's (1975) binary stationary process is defined through a stochastic operator of "addition mod 2", and the operation of *modulo* can only be manipulated arithmetically, and hence not applicable for categorical-valued stochastic processes. In contrast, Pegram (1980) defined a class of discrete-valued AR(p) processes that resembles Box and Jenkins' ARMA processes and allows some of the serial correlations to be negative. Moreover, the resulting autocorrelation function is independent of the marginal mean parameters. It is evident that Pegram's AR models are more flexible in terms of the range of correlation and the ease of interpretation.

One main contribution of this paper is to demonstrate that Pegram's operator, originally proposed only for AR processes, can be effectively extended to define discrete-valued ARMA processes with arbitrary discrete margins, where related arithmetic manipulation takes place in marginal or conditional probabilities, rather than directly in the values of Y_t . Hence, a unified framework of discrete-valued stationary processes is established.

* Corresponding author. Tel.: +91 33 25772088; fax: +91 33 25773104.

E-mail addresses: atanu@isical.ac.in (A. Biswas), pxsong@umich.edu (P.X.-K. Song).

2. Pegram’s ARMA processes

Pegram’s operator $*$ is defined as follows. For two independent discrete random variables U and V , and for a given coefficient $\phi \in (0, 1)$, Pegram’s operator “mix” them to produce a random variable Z : $Z = (U, \phi) * (V, 1 - \phi)$, with the marginal probability function given by

$$P(Z = j) = \phi P(U = j) + (1 - \phi)P(V = j), \quad j = 0, 1, \dots \tag{1}$$

Clearly, this operation works for any discrete distribution $D(p_0, p_1, \dots)$, where $X \sim D(p_0, p_1, \dots)$ means that the domain of X is $\{0, 1, \dots\}$ and $P(X = i) = p_i, i = 0, 1, \dots$. This operator $*$ indicates a mixture of two discrete distributions, with the respective mixing weights ϕ and $1 - \phi$. Therefore, we call $*$ *the mixing operator* in the rest of this paper.

2.1. AR(p) processes

Suppose that time series Y_t is discrete and its marginal probability mass function is time-independent. Let $Y_t \sim D(p_0, p_1, \dots)$. Suppose ϵ_t ’s are also independently and identically distributed as the same $D(p_0, p_1, \dots)$. Denote $\mu = E(Y_t)$ and $\sigma^2 = \text{var}(\epsilon_t)$.

Definition 1. Let $\{Y_t\}$ be a discrete-valued stochastic process such that

$$Y_t = (I[Y_{t-1}], \phi_1) * (I[Y_{t-2}], \phi_2) * \dots * (I[Y_{t-p}], \phi_p) * (\epsilon_t, 1 - \phi_1 - \dots - \phi_p), \tag{2}$$

which is a mixture of $(p + 1)$ discrete distributions, where $I[Y_{t-1}], \dots, I[Y_{t-p}]$ are p point masses, $I[\cdot]$ being indicator variable, and $\epsilon_t \sim D(p_0, p_1, \dots)$, with respective mixing weights being $\phi_1, \phi_2, \dots, \phi_p$ and $1 - \phi_1 - \dots - \phi_p, \phi_j \in (0, 1), j = 1, \dots, p$ and $\sum_{j=1}^p \phi_j \in (0, 1)$. This implies that for every $t \in 0, \pm 1, \pm 2, \dots$, the conditional probability function takes the form

$$\begin{aligned} P(Y_t = j | Y_{t-1}, Y_{t-2}, \dots) &= P(Y_t = j | Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}) \\ &= (1 - \phi_1 - \dots - \phi_p)p_j + \phi_1 I[Y_{t-1} = j] + \dots + \phi_p I[Y_{t-p} = j], \end{aligned} \tag{3}$$

where $\phi_j, j = 1, \dots, p$, are chosen such that the polynomial equation $1 - \phi_1 z - \dots - \phi_p z^p = 0$ has roots lying outside of the unit disc.

Marginal stationarity of $\{Y_t\}$ is immediate, and the conditional expectation of Y_t given Y_{t-h} ’s is $E(Y_t | Y_{t-1}, \dots, Y_{t-p}) = (1 - \phi_1 - \dots - \phi_p)\mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$. It is easy to show that the ACVF of Y_t is

$$\gamma(h) = \text{cov}(Y_t, Y_{t-h}) = \phi_1 \gamma(h - 1) + \dots + \phi_p \gamma(h - p).$$

It follows that the autocorrelation function (ACF) is given by

$$\rho(h) = \phi_1 \rho(h - 1) + \dots + \phi_p \rho(h - p), \quad \text{for all } h. \tag{4}$$

Note that the $\rho(h)$ has the same expression as in the Box–Jenkins’ AR(p) model, and hence the ACF of the AR(1) process is $\rho(h) = \text{corr}(Y_t, Y_{t-h}) = \phi^{|h|}$.

As a Markov process, the AR(1) model characterizes the transition probabilities described as follows. First, the probability transition equation can be expressed as of the form $\mathbf{p}_t = T \mathbf{p}_{t-1}$ where marginal probabilities $\mathbf{p}_t = \mathbf{p}_{t-1} = (p_1, \dots, p_K)^\top$, and the $K \times (K + 1)$ transition matrix $T_{t|t-1} = [t_{ij}] = [P(Y_t = j | Y_{t-1} = i)]$. It is easy to see that $T = \phi I_{K+1} + (1 - \phi)P$, where I_{K+1} is the $(K + 1)$ -dimensional identity matrix and $P = (p_0 \mathbf{1}_{K+1}, \dots, p_K \mathbf{1}_{K+1}) = \mathbf{p} \otimes \mathbf{1}_{K+1}$, with $\mathbf{1}_{K+1}$ being the $(K + 1)$ -element vector of all 1’s. Here \otimes represents the Kronecker product. Also, by some simple algebra, we find the following recursive relation for the transition matrices, $T^t = \phi^t T + (1 - \phi^t)P, t = 0, 1, \dots$

2.2. Auto-odds-ratio function

As pointed out by [Jacobs and Lewis \(1983\)](#), the numerical measures of autocorrelations have no meaningful interpretation, unless the categories are ordinal. Instead, we may use the odds ratio association in the case of categorical time series for the serial dependence ([Agestri, 2002](#)). This naturally gives rise to *the auto-odds-ratio function (AORF)* that measures the serial association of a categorical process at two different time points. The lag- h odds ratio takes the form

$$\theta(h) = \frac{P(Y_t = 1, Y_{t-h} = 1)P(Y_t = 0, Y_{t-h} = 0)}{P(Y_t = 1, Y_{t-h} = 0)P(Y_t = 0, Y_{t-h} = 1)}, \quad h \geq 1. \tag{5}$$

This describes the ratio of odds that $Y_t = 1$ when $Y_{t-h} = 1$ compared to when $Y_{t-h} = 0$. We illustrate the relation between the regular ACF and the proposed AORF. This relation allows us to estimate the sample AORF by using the sample ACF and to establish the large sample properties of the sample AORF by applying the existing results of the sample ACF in the literature. In the case of stationary binary AR(1) processes, with the following joint probabilities,

$$P(Y_t = j_0, Y_{t-1} = j_1) = P(Y_t = j_0 | Y_{t-1} = j_1)P(Y_{t-1} = j_1) = \{(1 - \phi)p_{j_0} + \phi I[j_1 = j_0]\} p_{j_1},$$

we can calculate the lag-1 odds ratio as $\theta(1) = 1 + \frac{\phi}{(1-\phi)^2\sigma^2}$, where $p_j = P(Y_t = j), j = 0, 1$, and $\sigma^2 = \text{var}(Y_t) = p_1p_0$. Furthermore, by induction,

$$P(Y_t = j_0 | Y_{t-h} = j_h) = (1 - \phi^h) p_{j_0} + \phi^h I[j_h = j_0], \tag{6}$$

implies $P(Y_t = j_0 | Y_{t-h-1} = j_{h+1}) = (1 - \phi^{h+1}) p_{j_0} + \phi^{h+1} I[j_{h+1} = j_0]$. Hence we have the following theorem:

Theorem 1. *If $\{Y_t\}$ is a stationary binary AR(1) process with the ACF $\rho(h)$, then the AORF $\theta(h)$ is given by*

$$\theta(h) = 1 + \frac{\rho(h)}{\{1 - \rho(h)\}^2\sigma^2}, \quad h \geq 1.$$

Generalizing it to the case with a finite number of categories (three or larger), we define the *generalized lagged odds ratio function* (GAORF) as follows: For a pair of categories indexed by $(j_1, j_2), j_1 \neq j_2$,

$$\theta_{j_1, j_2}(h) = \frac{P(Y_t = j_1, Y_{t-h} = j_1)P(Y_t = j_2, Y_{t-h} = j_2)}{P(Y_t = j_1, Y_{t-h} = j_2)P(Y_t = j_2, Y_{t-h} = j_1)}, \quad h \geq 1.$$

This describes the ratio of odds that $Y_t = j_1$ when $Y_{t-h} = j_1$ compared to when $Y_{t-h} = j_2$. By a similar argument, we obtain the following result.

Corollary 1. *Suppose $\{Y_t\}$ is a stationary AR(1) process with marginal distribution $D(p_0, p_1, \dots, p_K)$. Then the GAORF is*

$$\theta_{j_1, j_2}(h) = 1 + \frac{\rho(h)}{[1 - \rho(h)]^2 p_{j_1} p_{j_2}}, \quad h \geq 1, j_1, j_2 = 0, 1, \dots, K.$$

The above results also hold for AR(p) processes with $h \geq p$. The related proofs are similar. For AR(1) process, the coefficient ϕ determines patterns of ACF and AORF functions, both functions decay to zero in an exponential rate. For AR(p), ACF and AORF behave similarly.

In addition, to ensure the positivity of the AORF, parameter ϕ has to satisfy $\phi/(1 - \phi)^2 > -\sigma^2 = -p_1(1 - p_1) \geq -1/4$. This lower bound is achieved when $p_0 = p_1 = 1/2$. This leads to $\phi > -1$, which is always true.

The partial autocorrelation function (PACF), denoted by $\alpha(h)$, of an AR(p) process can be defined similarly to that in Box and Jenkins' ARMA processes. First $\alpha(1) = \text{corr}(Y_2, Y_1) = \rho(1)$, and $\alpha(h) = 0$ when $|h| > p$, and $\alpha(h) = \gamma(h) - \phi_1\gamma(h - 1) - \dots - \phi_{h-1}\gamma(1)$ when $|h| \leq p$.

Similar to the ACF, the PACF may lose its interpretability in the case of nominal categorical observations. One solution is to "transfer" the PACF to the AORF through the relation between the AORF $\theta(h)$ and the ACVF $\gamma(h)$. For example, in the case of the AR(1) process, such a relation leads to $\theta(h) = 1 + \frac{\gamma(h)}{[\sigma^2 - \gamma(h-1)]^2}, h \geq 1$.

2.3. MA(q) and ARMA(p, q) processes

A discrete moving average (MA) process of order q , denoted by MA(q), for an integer $q > 0$, is defined by Pegram's mixing operator as follows:

$$Y_t = (\epsilon_t, \theta_0) * (I[\epsilon_{t-1}], \theta_1) * \dots * (I[\epsilon_{t-q}], \theta_q), \tag{7}$$

where $\theta_i \geq 0$ and $\sum_{i=0}^q \theta_i = 1$. It is easy to see that the marginal stationarity, and subsequently the ACF of the MA(q) process is $\rho(h) = \sum_{i=0}^{q-h} \theta_{i+h}\theta_i$ or 0 according to $|h| \leq q$ or $|h| > q$, which is the same as the Box-Jenkins' form, where the θ_i 's may be interpreted as the variance standardized coefficients.

Likewise, a moving average process of infinite order, MA(∞), takes the form $Y_t = (\epsilon_t, \theta_0) * \{\sum_{i=1}^{\infty} (I[\epsilon_{t-i}], \theta_i)\}$, with $\theta_i \geq 0$ and $\sum_{i=0}^{\infty} \theta_i = 1$. The limit of the process exists because its conditional probability satisfies

$$P(Y_t = j | \epsilon_t, \epsilon_{t-1}, \dots) = \theta_0 I[\epsilon_t = j] + \sum_{i=1}^{\infty} \theta_i I[\epsilon_{t-i} = j] \leq \sum_{j=0}^{\infty} \theta_i = 1.$$

Moreover, the conditional expectation of MA(∞) is given as follows. For $Y_t \sim D(p_0, p_1, \dots)$ for all $t, E(Y_t | \epsilon_t, \epsilon_{t-1}, \dots) = \sum_{i=0}^{\infty} \theta_0 \epsilon_{t-i}$. It is worth pointing out that discrete-valued AR process Y_t in (2) can be written as an MA(∞). To elucidate, let us consider an AR(1) process first. That is, the Definition 1 implies that

$$\begin{aligned} P(Y_t = j) &= \phi_1 P(Y_{t-1} = j) + (1 - \phi_1) P(\epsilon_t = j) \\ &= \phi_1^2 P(Y_{t-2} = j) + \phi_1(1 - \phi_1) P(\epsilon_{t-1} = j) + (1 - \phi_1) P(\epsilon_t = j), \end{aligned}$$

and hence by the argument of induction, we can write for a positive integer m ,

$$Y_t = (I[Y_{t-m}], \phi_1^m) * (I[\epsilon_{t-m+1}], \phi_1^{m-1}(1 - \phi_1)) * \dots * (I[\epsilon_{t-1}], \phi_1(1 - \phi_1)) * (\epsilon_t, 1 - \phi_1).$$

Let $m \rightarrow \infty$. Because the reminder series $|\sum_{i=m+1}^{\infty} \phi_1^i (1 - \phi_1) P(\epsilon_{t-i} = j)| \leq |\phi_1|^m \rightarrow 0$, AR(1) can be written as an MA(∞), $Y_t = (\epsilon_t, \theta_0) * \{*\}_{i=1}^{\infty} (I[\epsilon_{t-i}], \theta_i)$, with $\theta_i = \phi_1^i (1 - \phi_1)$, $i \geq 0$. For an AR(p) process, note that the polynomial $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = (1 - \xi_1 z) \dots (1 - \xi_p z)$, where ξ_i^{-1} , $i = 1, \dots, p$ are the roots. To convert an AR(p) into an MA(∞), we recursively execute the operation p times, where each operation deals with $(1 - \xi_i z)$, which effectively converts an AR(1) process into an MA(∞).

Finally, an ARMA(p, q) process can be constructed by combining the AR(p) and MA(q) processes as follows:

$$Y_t = (I[Y_{t-1}], \phi_1) * (I[Y_{t-2}], \phi_2) * \dots * (I[Y_{t-p}], \phi_p) * (\epsilon_t, \theta_0) * (I[\epsilon_{t-1}], \theta_1) * \dots * (I[\epsilon_{t-q}], \theta_q), \tag{8}$$

where $\phi_i \geq 0$ and $\theta_j \geq 0$ for all i and j , and $\sum_{i=1}^p \phi_i + \sum_{j=0}^q \theta_j = 1$.

3. Parameter estimation

3.1. Yule–Walker equation approach

Here we focus on the AR(p) process. Note that $\Gamma \phi = \gamma_p$, where $\Gamma = [\gamma(i - j)]$, $\phi = (\phi_1, \dots, \phi_p)^\top$, and $\gamma_p = (\gamma(1), \dots, \gamma(p))^\top$. Obviously, this equation is also valid with the ACF, where $\gamma(h)$ is replaced by the ACF $\rho(h)$. Given a time series data estimate the $\gamma(h)$ by its sample ACVF, $\hat{\gamma}(h)$. As the $\hat{\gamma}(h)$ consistently estimates the true ACVF $\gamma(h)$ (Brockwell and Davis, 1991), the resultant estimator $\hat{\phi} = \hat{\Gamma}^{-1} \hat{\gamma}_p$ is consistent provided that the matrix $\hat{\Gamma}$ is invertible. As pointed out by Brockwell and Davis (1991), the invertibility is guaranteed if the ACF $\rho(h) \rightarrow 0$ as $h \rightarrow \infty$.

For the case of AR(1), the Yule–Walker equation estimator of the autocorrelation coefficient ϕ is given by

$$\hat{\phi}_n = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\sum_{i=2}^n (Y_i - \bar{Y}_n)(Y_{i-1} - \bar{Y}_n)}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \tag{9}$$

where \bar{Y}_n is the sample mean of the time series. This is the well-known Durbin–Waston estimator. Following Brockwell and Davis (1991, Chapter 7) and Billingsley (1961), it is easy to show that under some mild moment conditions both the marginal mean estimator \bar{Y}_n and the Durbin–Waston estimator $\hat{\phi}_n$ are asymptotically normally distributed. In particular, $\bar{Y}_n \sim AN(\mu, n^{-1}v)$ with $v = \sigma^2(1 + \phi)/(1 - \phi)$, and $\hat{\phi}_n \sim AN(\phi, n^{-1}w)$ with $w = 1 - \phi^2$, obtained by Bartlett’s formula.

3.2. Maximum likelihood estimation

We illustrate the procedure of maximum likelihood estimation categorical AR(1) processes. Given a time series data $\{Y_1, Y_2, \dots, Y_n\}$ from an AR(1) process where $p_j = 0$ for $j > K$, we have $P(Y_i = j_i, i = 1, \dots, n) = P(Y_1 = j_1) \prod_{i=2}^n P(Y_i = j_i | Y_{i-1} = j_{i-1})$. So, the likelihood is $L = p_{j_1} [\prod_{i=2}^n \{(1 - \phi)p_{j_i} + \phi I(j_{i-1} = j_i)\}]$.

We did some simulation experiments to compare the MLE and Y–W estimates. The details are presented as a [web-only material](#). It is interesting to note that the Yule–Walker equation method works as well as the MLE method, and in some cases, it performs even better than the MLE. One reason might be that the Yule–Walker method is based on sufficient statistics. Similar results were found in the other simulations we ran. Detailed study on estimation is carried out by Song, Freeland and Biswas, in an unpublished article.

3.3. Comparison: Thinning model versus proposed model for Poisson AR(1)

A second simulation study was run to compare the fit of AR(1) Poisson process, between the thinning operator based model and Pegram’s operator based model. The details are presented as a [web-only material](#). We found that neither of the thinning or Pegram’s model is uniformly better than the other. Our simulation study suggests that the AIC seems to make a trustful indication as to which approach works more appropriately for the data.

4. Data analyses

4.1. Modeling of worker compensation claims

In this section, we apply the proposed models to analyze a count data time series. The data set contains monthly number of claims of short-term disability benefits made by the logging workers who had cut injury to the British Columbia Workers’ Compensation Board (WCB) during January, 1985 to December, 1994. The data were previously analyzed by Freeland (1998) and Zhu and Joe (2006) using the thinning operator based model. A difference between Pegram’s model and the thinning model for integer-valued time series lies on the modeling of transition properties. The former attempts to model transition probabilities of the process, while the latter directly models the outcome of the process. We may interpret the AR(1) model in light of conditional expectations. Given the number of injuries in the previous month, Y_{t-1} , the expected number of

injuries in the current month is predicted by a convex combination of Y_{t-1} and the average number of injuries μ within a month. Here, parameter ϕ indicates the weight for the prediction associated with the number of the previous injuries, and $(1 - \phi)$ is then the weight for the new injuries. The figures are available in the [web-only material](#). The top panel of Figure 1 displays the time series plot of the data, from which a weak seasonal trend – more claims in the summer session – seems present. However, the sample PACF plot of Figure 1, as well as the ACF plot (not shown), do not indicate the significance of the season effect. To compare with the previous analyses, we here fit the data to an AR(1) model with a Poisson margin. Note that the Poisson AR(1) process we use in the analysis is different from the thinning operator based model. The choice of the first-order autoregressive model is based on the sample PACF plot shown in the bottom panel of Figure 1. There is only the first lag significant in this plot, which means that AR(1) model is appropriate. Our Yule–Walker equation estimation for both autocorrelation coefficient $\hat{\phi}_{YW} = 0.558$ with the asymptotic standard error (*s.e.* = 0.076) and mean parameter $\hat{\mu}_{YW} = 6.133$ (*s.e.* = 0.42), comparing to the maximum likelihood estimates of these parameters, respectively, $\hat{\phi}_{MLE} = 0.431$ (*s.e.* = 0.051) and mean parameter $\hat{\mu}_{MLE} = 6.130$ (*s.e.* = 0.36), obtained in the previous analyses. Essentially, our results are in great agreement with those of the previous analyses. The reason that both models yield similar results is that they have the same marginal distribution and ACF. Again, the resulting AR(1) model may be utilized for forecasting. That is, given the number of injuries in the current month, say Y_t , the expected number of injuries to occur in the next month would be $0.56Y_t + 2.71$.

4.2. Hardware failure data

This is a data set obtained from the computer hardware defect data of the Indian Statistical Institute, Kolkata. The service provider in the Institute is supposed to take care of all the defects and problems occurred in the Institute in a day by day basis. In this data set, we consider the number of hardware defects for the 260 working days in 2005. The data is given in Figure 2. Here we observe that the frequency of cases where $Y_{t-1} = Y_t$ is considerable. It intuitively suggests that a Pegram type model might work better than the thinning model. If we consider an AR(1) model, the AIC of the thinning model is 489.398 and that of Pegram's model is 489.239. Pegram's model works slightly better.

4.3. Infant sleep data

Stoffer et al. (1988) reported a collection of 24 categorical time series of infant sleep status in an EEG study, from which we select one observed in 119 min for our analysis. During minute t , the baby's sleep status was recorded in 6 ordered categories, namely, 1 being 'quiet sleep' with trace alternant, 2 being 'quiet sleep' with high voltage, 3 being 'indeterminate sleep', 4 being 'active sleep' with low voltage, 5 being 'active sleep' with mixed voltage, and 6 being 'awake'. Figure 3 consists two panels of plots: the top one is the time series plot of the raw data and the bottom one displays the sample PACF, in which only the first lag appears to be significant. To further judge the stationarity of the series, we first compare the sample ACF and the theoretical ACF of the AR(1) process with $\phi = 0.873$ estimated from the data. Figure 4 shows that the sample ACF decays to zero faster than the AR(1)'s ACF, implying that among all AR(p) models with $p = 1, 2, \dots$, the AR(1) model seems most appropriate. This is because the AR(p) models, $p \geq 2$, have longer memory than the AR(1), so their decay rates will become even slower than that of the AR(1).

We apply the Yule–Walker equation approach to estimating the parameters of the autocorrelation coefficient ϕ and the marginal probabilities $p_j, j = 1, \dots, 6$. It follows immediately that $\hat{\phi} = 0.873$ (*s.e.* = 0.045) and $\hat{p}_j, j = 1, \dots, 6$ are, respectively, 0.403, 0.017, 0.092, 0.387, 0.101, and 0.000 (the baby was never in status 6). Clearly, the baby was not in status 6 over the whole period of the study and was more likely in status 1 and status 4 than any other status, both cases having marginal probability approximately equal to 0.4 with a total probability 0.8 for the two. Also, the estimated 95% confidence interval of the autocorrelation ϕ is (0.784, 0.960), indicating that the sleep status at the previous minute has a very strong and positive impact on the status at the current minute.

5. Concluding remarks

In this paper we presented a unified framework of stationary discrete-valued ARMA processes that are useful to model discrete-valued time series. This class of models is new and provides a valuable arsenal to enrich the theory and practice of the popular Box and Jenkins' ARMA models. As shown in the data analysis, the thinning operator based ARMA models are unavailable to analyze categorical time series data.

For count data, our model is an alternative to the existing thinning based model. As is observed in Section 3.3, neither of the thinning based model and Pegram's model is uniformly better than the other. Pegram's model outperform the thinning model (with respect to AIC) if the data generation mechanism is close to Pegram's mechanism. Intuitively if we have considerable proportions of cases $Y_{t-1} = Y_t$, Pegram's model works better, as is observed in the *Computer Hardware Failure* data example of Section 4.2.

But, the most important feature of our model and this article is that our model serves as an excellent tool to model and analyze any type of discrete time series data. This is a new contribution to the time series community. Thus, we can now provide some way out of handling data examples like the *Infant Sleep Status Data* in an elegant way. We can also handle binomial time series in a similar way. This is the prime feature of our model and this paper.

An important extension of the proposed models and methods is to incorporate covariates in the marginal probabilities via, for example, the logistic model. There have been some recent developments in this direction, including Qaqish's (2003) transition model for binary nonstationary processes, Benjamin et al.'s (2003) generalized autoregressive moving average models, regression theory for categorical time series by Fokianos and Kedem (2003), and also the regression models for nonstationary categorical time series by Kaufmann (1987).

Acknowledgments

The authors wish to thank the Editor and two anonymous referees for their constructive suggestions which led to some improvements in some earlier versions of the manuscript.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.spl.2009.05.025.

References

- Agresti, A., 2002. *Categorical Data Analysis*, 2nd Ed. Wiley, New York.
- Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. *J. Amer. Statist. Assoc.* 98, 214–223.
- Billingsley, P., 1961. Statistical methods in Markov chains. *Ann. Math. Statist.* 32, 12–40.
- Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control*, Revised edition. Holden-Day, San Francisco.
- Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*. Springer-Verlag, New York.
- Fokianos, K., Kedem, B., 2003. Regression theory for categorical time series. *Statistical Science* 18, 357–376.
- Freeland, R.K., 1998. *Statistical Analysis of Discrete Time Series with Application to the Analysis of Workers' Compensation Claims Data*. Ph.D. Thesis, Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia.
- Kaufmann, H., 1987. Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Ann. Statist.* 15, 79–98.
- Jacobs, P.A., Lewis, P.A.W., 1978. Discrete time series generated by mixtures. I: Conditional and runs properties. *J. R. Statist. Soc. B* 40, 94–105.
- Jacobs, P.A., Lewis, P.A.W., 1983. Stationary discrete autoregressive-moving average time series generated by mixtures. *J. Time Series Analysis* 4, 19–36.
- Joe, H., 1996. Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* 33, 664–677.
- Jørgensen, B., Song, P.X.-K., 1998. Stationary time-series models with exponential dispersion model margins. *J. Appl. Probab.* 35, 78–92.
- Kanter, M., 1975. Autoregression for discrete processes mod 2. *J. Appl. Probab.* 12, 371–375.
- Pegram, G.G.S., 1980. An autoregressive model for multilag markov chains. *J. Appl. Probab.* 17, 350–362.
- Qaqish, B.F., 2003. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 90, 455–463.
- Steutel, F.W., van Harn, K., 1979. Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7, 893–899.
- Stoffer, D.S., Scher, M.S., Richardson, G.A., Day, N.L., Coble, P.A., 1988. A Walsh–Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *J. Amer. Statist. Assoc.* 83, 954–963.
- Zhu, R., Joe, H., 2006. Modelling count data time series with Markov processes based on binomial thinning. *J. Time Ser. Anal.* 27, 725–738.