Statistical Papers (2008) 49:485–502
DOI 10.1007/s00362-006-0028-1

REGULAR ARTICLE

A non-stationary integer-valued autoregressive model

Hee-Young Kim · Yousung Park

Received: 23 June 2006 / Revised: 18 October 2006 / Published online: 17 November 2006

Abstract It is frequent to encounter a time series of counts which are small in value and show a trend having relatively large fluctuation. To handle such a non-stationary integer-valued time series with a large dispersion, we introduce a new process called integer-valued autoregressive process of order p with signed binomial thinning (INARS(p)). This INARS(p) uniquely exists and is stationary under the same stationary condition as in the AR(p) process. We provide the properties of the INARS(p) as well as the asymptotic normality of the estimates of the model parameters. This new process includes previous integer-valued autoregressive processes as special cases. To preserve integer-valued nature of the INARS(p) and to avoid difficulty in deriving the distributional properties of the forecasts, we propose a bootstrap approach for deriving forecasts and confidence intervals. We apply the INARS(p) to the frequency of new patients diagnosed with acquired immunodeficiency syndrome (AIDS) in Baltimore, Maryland, U.S. during the period of 108 months from January 1993 to December 2001.

Keywords Non-stationarity · Integer-valued time series · Signed binomial thinning · Bootstrapping · Over-dispersion · Quasi-likelihood

H.-Y. Kim
Institute of Statistics, Korea University, Seoul 136-701, Korea
e-mail: starkim@korea.ac.kr

Y. Park (✉)
Department of Statistics, Korea University, Seoul 136-701, Korea
e-mail: yspark@korea.ac.kr

1 Introduction

In most cases, the data for the time series of counts exhibits dependency on past observations that must be appropriately modeled. For this type of non-stationary integer-valued time series, the usual real-valued ARIMA model has been used in many applications (Waiter et al. 1991; Anderson and Grenfell 1984; Zaidi et al. 1989). However, the ARIMA model may be inappropriate in the case of a rare disease. For example, the frequency of new patients per month diagnosed with acquired immunodeficiency syndrome (AIDS) in Baltimore are below 20 after 1997 and fading away thereafter.

For a time series of count data, integer-valued analogues of the stationary ARMA models have been suggested (McKenzie 1986; Al-Osh and Alzaid 1987; Alzaid and Al-Osh 1990; Jin-Guan and Yuan 1991; Aly and Bouzar 1994; Park and Oh 1997; McCormick and Park 1997). These stationary integer-valued time series models possess many features in common with the standard ARMA models. Both can express their processes in the form of difference equations and share a common behavior in their time correlations. A primary difference between them is that the integer-valued time series model uses a binomial thinning operator in place of the multiplication done in the standard ARMA model. However, these integer-valued analogues are no longer valid for integer-valued time series with a time trend.

Quasi-likelihood methods are especially useful for analysis of overdispersed count data (Zeger and Qaqish 1988; McCullagh and Nelder 1989; Breslow 1990). They defined separate mean and variance functions to avoid the overdispersion problem, in which the empirical variance is greater than the theoretical variance. The mean function usually includes lagged values of responses to reflect the dependency of the response on its past and covariates to incorporate the general trend of the response variable. The variance function is expressed as multiplication of the mean function and a constant dispersion parameter to better fit the empirical variance. However, the constant dispersion parameter is often not large enough to describe the variability of count data such as the variability of the AIDS data in Baltimore (see Sect. 5 for details). This constant dispersion parameter has been relaxed by allowing the dispersion parameter varying across observations through a regression model (Dey et al. 1997; Cordeiro and Botter 2001). However, their methods may not be directly applied to our time series since they assumed independent observations.

We introduce a new operator called a signed binomial thinning to develop a new integer-valued time series model for handling an overdispersed and non-stationary integer-valued time series. We call this model integer-valued autoregressive model of order p with signed binomial thinning (INARS(p)), which can remove the time trend and seasonality by using the differencing operator in the same way as the familiar continuous ARIMA model does. One advantage of the INARS(p) is that it can handle negative integer-valued time series, whereas the previous integer-valued time series models are only applicable to nonnegative integer-valued time series. The INARS model also allows negative autocorrelations of the general pth order to an integer-valued time series, whereas the previous integer-valued time series model can only work with positive autocorrelations.

The previous integer-valued autoregressive models with order p (INAR(p)) have been considered, among other things, by Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991). They are different in autocorrelation structure and conditional mean function. The INARS(p) persists these differences.

The most commonly used technique in constructing forecasts in time series model is conditional expectations. However, this method lacks data coherency when the time series is integer valued. To meet this integer valued nature of data, the median was used as a forecast of an INAR(1) model (Freeland and McCabe 2004) and bootstrapping approaches have been suggested for INAR(p) models (Cardinal et al. 1999; Jung and Tremayne 2006). There are several bootstrap alternatives in the literature to construct prediction intervals for real-valued autoregressive models of order p (AR(p)). Findley (1986), Masarotto (1990), and Grigoletto (1998) used bootstrap methods to estimate the density of the forecast errors including uncertainty due to parameter estimation. Thombs and Schucany (1990) used the backward representation of AR(p) models to generate bootstrap series with a known p. Recently, Pascual et al. (2004) proposed a bootstrap method which incorporates the variability arising from parameter estimation into forecast intervals without requiring the backward representation of the process. We modify the bootstrap method of Pascual et al. (2004) to preserve the integer-valued nature of data for forecasts and their confidence intervals.

The rest of this paper is divided into six sections. In Sect. 2, we investigate the autocorrelation structure and conditional expectation of INARS(2) models under two different model assumptions. Then we show the stationarity and ergodicity of the INARS(p) process under the same condition as the previous INAR(p). In Sect. 3, the strong consistency of moment estimators and asymptotic normality of the conditional least squares estimators are shown. Section 4 includes a bootstrap approach to preserve the integer-valued nature of count data. In Sect. 5, the INARS(p) model is applied to the AIDS cases in Baltimore from January 1993 to December 2001. The INARS(p) model is compared with the usual AR(p) model and the quasi-likelihood model of Zeger and Qaqish (1988). Section 6 includes some concluding remarks.

2 Integer-valued autoregressive process with signed binomial thinning

Differencing, a method commonly used to remove the time trend and seasonality from a time series, is adopted for a non-stationary integer-valued time series. By the same notation as used in the usual ARIMA process, $\{\nabla^d \nabla_s^D X_t\}$ is defined as the differenced series of X_t where $\nabla X_t = X_t - X_{t-1}$, $\nabla_s X_t = X_t - X_{t-s}$, and d and D indicate the repeated times of ∇X_t and $\nabla_s X_t$, respectively.

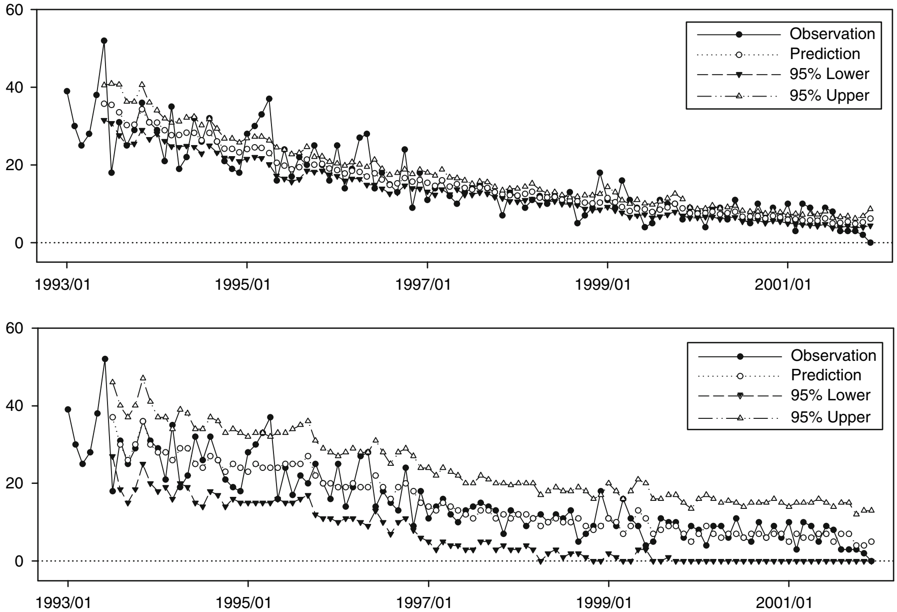

Fig. 1 Fitted values and confidence intervals using **a** Quasi-likelihood model and **b** INARS(5)

Hereafter, we denote $\nabla^d \nabla_s^D X_t$ by y_t for simple notation. This differenced series is still integer-valued but can be negative-valued. When a time series linearly decreases as illustrated in Fig. 1 for AIDS, the differenced series $X_t - X_{t-1}$ is much smaller than the original series X_t. This implies that when ARIMA model is fit to such a differenced time series, its goodness-of-fit worsens. The differenced time series can not fit to the previous INAR(p)-type of models, because the INAR(p) model applies only to a non-negative valued-time series.

To model this differenced time series with an integer-valued time series process, we introduce a new operator represented by $\odot$ and call it the "signed binomial thinning" operator. This new operator is an extension of the previous binomial thinning as shown below. Let $\{w_{tj}(\alpha)\}$ be i.i.d. Bernoulli random variables with $P(w_{tj}(\alpha) = 1) = |\alpha|$ for each given t. Define $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$. Using this notation, the signed binomial thinning is formally defined as

$$\alpha \odot y_t \equiv \mathrm{sgn}(\alpha)\mathrm{sgn}(y_t) \sum_{j=1}^{|y_t|} w_{tj}(\alpha), \tag{1}$$

where the subscript t in $w_{tj}(\alpha)$ describes the observed time of process y_t. For simple notation, the subscripts t and α are dropped from $w_{tj}(\alpha)$ if no confusion arises. When $y_t \geq 0$ and $\alpha \geq 0$, the signed binomial thinning is reduced to the binomial thinning denoted by $\alpha \circ y_t = \sum_{j=1}^{y_t} w_j$.

2.1 INARS(2) model

Before getting into a generalized INAR(p)-type of models using the signed binomial thinning, we first consider, so-called, integer-valued autoregressive model of order 2 with signed binomial thinning (INARS(2)) to illustrate its connection to the two types of the previous INAR(2) models presented by Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991). The INARS(2) model is given by

$$y_t = \alpha_1 \odot y_{t-1} + \alpha_2 \odot y_{t-2} + \epsilon_t \tag{2}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. integer-valued random variables with mean μ_ϵ and variance σ_ϵ^2 and $0 \leq |\alpha_1|, |\alpha_2| \leq 1$.

When $y_t \geq 0$ and $\epsilon_t \geq 0$ for all t and α_1 and α_2 are also non-negative, the INARS(2) of (2) becomes INAR(2) expressed by

$$y_t = \alpha_1 \circ y_{t-1} + \alpha_2 \circ y_{t-2} + \epsilon_t.$$

In this INAR(2) model, Alzaid and Al-Osh (1991) assumed that the conditional distribution of $(\alpha_1 \circ y_t, \alpha_2 \circ y_t)$ given y_t is multinomial with parameter $(\alpha_1, \alpha_2, y_t)$ and is independent of the past history of the process. These assumptions can be likewise adopted in the INARS(2) model after a slight modification: $\left(\sum_{j=1}^{|y_t|} w_j(\alpha_1), \sum_{j=1}^{|y_t|} w_j(\alpha_2)\right)$ given y_t is multinomial with parameter $(|\alpha_1|, |\alpha_2|, |y_t|)$ where $|\alpha_1| + |\alpha_2| \leq 1$ and is independent of the past history of the process. Then, we have the following results.

Proposition 1 *When y_t is a stationary INARS(2) process with the Alzaid and Al-Osh type of assumptions described above, the covariance functions are*

$$\gamma(1) = \alpha_1\gamma(0) + \frac{\alpha_1\alpha_2(\gamma(0) - E|y_t|)}{1 - \alpha_2} \quad \textit{and} \quad \gamma(k) = \alpha_1\gamma(k-1) + \alpha_2\gamma(k-2),$$

where $k \geq 2$ and $\gamma(l) = \text{cov}(y_t, y_{t-l})$ for $l = 0, 1, \ldots$.

The proof and all subsequent theoretical proofs are provided in the Appendix. These covariance are the typical forms of those from a continuous ARMA(2,1) and are the same as those of INAR(2) model of Alzaid and Al-Osh (1990). Generally, we can show that INARS(p)'s covariance function behaves like those of ARMA($p, p-1$) under Alzaid and Al-Osh's assumptions as the covariance function of their INAR(p) behaves like those of ARMA($p, p-1$). Moreover, the conditional expectation of $E(y_t|y_{t-1}, y_{t-2})$ in INARS(2) is not necessarily linear in y_{t-1} and y_{t-2} because the INAR(2) of Alzaid and Al-Osh (1990) has been shown to be non-linear and is a special case of INARS(2).

On the other hand, Jin-Guan and Yuan (1991) assumed that the counting series $w_{tj}(\alpha)$'s in the binomial thinning $\alpha \circ y_t = \sum_{j=1}^{y_t} w_{tj}(\alpha)$ are not only i.i.d. but also are independent of y_t. Thus, $\alpha_1 \circ y_t$ and $\alpha_2 \circ y_{t'}$ are conditionally independent given in y_t and $y_{t'}$ as long as their counting series are not the same. Likewise, when we impose these Jin-Guan and Yuan's assumptions on the signed binomial thinning in defining the INARS(2) model, we have the following covariance function and basic properties of signed binomial thinning operator.

Proposition 2 *When y_t is a stationary INARS(2) with the assumptions of Jin-Guan and Yuan (1991),*

1. $\gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(1)$ *and* $\gamma(k) = \alpha_1\gamma(k-1) + \alpha_2\gamma(k-2)$ *for* $k \geq 2$
2. *for integer-valued two random variables y_{1t} and y_{2t} which have the same sign and the same counting series,* $E|\alpha \odot y_{1t} - \alpha \odot y_{2t}| = |\alpha|E(|y_{1t} - y_{2t}|)$ *and* $E(\alpha \odot y_{1t} - \alpha \odot y_{2t})^2 = \alpha^2 E(y_{1t} - y_{2t})^2 + |\alpha|(1-|\alpha|)E(|y_{1t} - y_{2t}|)$.

The stationarity mentioned in Proposition 2 is not pre-requisite for the second property (2) but this property will be used to show the stationarity of INARS(p) in Sect. 2.2. The covariance function described in (1) of Proposition 2 is the same as those of both the continuous AR(2) and the Jin-Guan and Yuan's INAR(2) model. Moreover, the conditional independence of $\alpha_1 \odot y_t$ and $\alpha_2 \odot y_{t'}$ given y_t and $y_{t'}$ also implies that the conditional expectation $E(y_t|y_{t-1}, y_{t-2})$ is a linear function of y_{t-1} and y_{t-2} in INARS(2) (i.e., $E(y_t|y_{t-1}, y_{t-2}) = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \mu_\epsilon$).

Based on above discussion on the connection between INARS(2) and two types of the INAR(2) models, we can conclude that the INARS(p) under Alzaid and Al-Osh's (1990) assumption is different from that under Jin-Guan and Yuan's (1991) assumption. In particular, as discussed in Alzaid and Al-Osh (1990), the conditional expectation of y_t given past information in INAR(p) model is non-linear and is extremely complex. Thus, the conditional expectation in INARS(p) under Alzaid and Al-Osh assumptions must be even more complex. Accordingly estimating parameters and forecasting future values may be intractable. For these reasons, from now on, we focus only on INARS(p) under Jin-Guan and Yuan's assumptions.

2.2 INARS(p) model

Using the signed binomial thinning operator, we are now able to define a new process for an integer-valued time series model which allows negative integer-valued and negative correlated time series.

$$y_t = \sum_{i=1}^{p} \alpha_i \odot y_{t-i} + \epsilon_t, \quad t = 0 \pm 1, \pm 2, \ldots \tag{3}$$

where the signed binomial thinning operator $\odot$ is given in (1), $\{\epsilon_t\}$ is a sequence of i.i.d. integer-valued random variables with mean μ_ϵ, and variance σ_ϵ^2, $0 \leq |\alpha_i| \leq 1$ for $i = 1, \ldots, p$. The $\{\epsilon_t\}$ are uncorrelated with y_{t-i} for $i \geq 1$ and counting series $w_{tj}(\alpha)$ in signed binomial thinning $\alpha \odot y_t = \sum_{j=1}^{y_t} w_{tj}(\alpha)$ are i.i.d. and independent of y_t. Because the new process uses a signed binomial thinning, we call the model given in (3) an integer-valued autoregressive process of order p with signed binomial thinning (INARS(p)).

To show that the INARS(p) process is stationary and ergodic, we follow a similar approach as that of Jin-Guan and Yuan (1991). Define

$$y_{n,t} = \begin{cases} 0, & n < 0 \\ \epsilon_t, & n = 0 \\ \alpha_1 \odot y_{n-1,t-1} + \cdots \alpha_p \odot y_{n-p,t-p} + \epsilon_t, & n > 0 \end{cases} \tag{4}$$

where $\text{Cov}(y_{n,t'}, \epsilon_t) = 0$ when $t' < t$ for any n, and the signs of $y_{n,t}$ and $y_{n',t'}$ are the same when $t = t'$ for any n and n'. Then, using Proposition 2, we can construct a INARS(p) process from $y_{n,t}$ as shown below.

Theorem 1 *Suppose that all roots of the polynomial $\lambda^p - \alpha_1\lambda^{p-1} - \cdots - \alpha_{p-1}\lambda - \alpha_p = 0$ are inside the unit circle. Then, the process y_t in L_2 space uniquely satisfies*

$$y_t = \sum_{i=1}^{p} \alpha_i \odot y_{t-i} + \epsilon_t, \quad t = 0 \pm 1, \pm 2, \ldots \tag{5}$$

where $y_t = \lim_{n\to\infty} y_{n,t}$ and $\text{Cov}(y_{t'}, \epsilon_t) = 0$ for $t' < t$. Furthermore, this process y_t is stationary and ergodic.

Note that the condition of Theorem 1 is the same stationary condition as that for the usual ARMA model.

3 Estimation

The ergodicity and stationarity of the INARS(p) process for y_t ensure that, using Durrett's (1991) approach,

$$\frac{1}{n}\sum_{t=1}^{n} y_t \xrightarrow{a.s} E(y_t), \quad \frac{1}{n}\sum_{t=1}^{n} |y_t| \xrightarrow{a.s} E(|y_t|),$$
$$\text{and} \quad \frac{1}{n}\sum_{t=1}^{n} y_t y_{t-k} \xrightarrow{a.s} E(y_t y_{t-k}) \quad \text{for } k = 0, 1, 2, \ldots. \tag{6}$$

Let $\hat{\alpha}_i$ $(i = 1, 2, \ldots, p)$ be the estimator satisfying

$$\hat{\gamma}_k = \hat{\alpha}_1\hat{\gamma}_{k-1} + \cdots + \hat{\alpha}_i\hat{\gamma}_{k-i} + \cdots + \hat{\alpha}_p\hat{\gamma}_{k-p}$$

where $\hat{\gamma}_k = \hat{\gamma}_{-k}$ and $\hat{\gamma}_k = (1/(n-k))\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t-k} - \bar{y})$. Using $\hat{\alpha}_i$, define $\hat{\sigma}_\epsilon^2 = (1/n)\sum_{t=1}^{n}(\hat{\epsilon}_t - \bar{\epsilon}_n)^2 + (1/n)\sum_{t=1}^{n}|y_t| \cdot \sum_{i=1}^{p}|\hat{\alpha}_i|(1 - |\hat{\alpha}_i|)$ and $\bar{\epsilon}_n = (1/n)\sum_{t=1}^{n}\hat{\epsilon}_t$ where $\hat{\epsilon}_t = y_t - \hat{\alpha}_1 y_{t-1} - \cdots - \hat{\alpha}_p y_{t-p}$. Then, by (6), we have the following results.

Lemma 1 *Let $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_t)$ and $\mu_\epsilon = E(\epsilon_t)$. Then, $\hat{\alpha}_i$, $\hat{\sigma}_\epsilon^2$, and $\bar{\epsilon}_n$ are strong consistent estimators of respective parameters α_i, σ_ϵ^2, and μ_ϵ.*

When we assume a conventional AR(p) model, the consistent estimator of σ_ϵ^2 given in Lemma 1 becomes $(1/n)\sum_{t=1}^n(\hat{\epsilon}_t - \bar{\epsilon}_n)^2$. Thus, the additional term $(1/n)\sum_{t=1}^n |y_t| \cdot \sum_{i=1}^p |\hat{\alpha}_i|(1-|\hat{\alpha}_i|)$ in the consistent estimator of σ_ϵ^2 also distinguish the INARS(p) from the AR(p). Let $\mathcal{F}_t$ be a sigma-field generated by $\{y_1, \ldots, y_t\}$. Then, by the definition of signed binomial thinning, we have

$$E(y_t|\mathcal{F}_{t-1}) = \mu_\epsilon + \sum_{i=1}^p \alpha_i y_{t-i}$$
$$\text{and } \mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^p |\alpha_i|(1-|\alpha_i|)|y_{t-i}| + \sigma_\epsilon^2, \tag{7}$$

implying the same conditional expectation but different conditional variance compared to those of the continuous AR(p) model. Thus, the confidence interval for the one-ahead best predictor (i.e., the conditional expectation given in (7)) is wider in the INARS(p) than the AR(p). Moreover, since the variance of INARS(p) model is varying over time t, the INARS(p) is an autoregressive conditional heteroscedasticity model. These differences also produce asymptotic variances different from those of the usual AR(p) shown in the following asymptotic result.

Let $\boldsymbol{\xi} = (\mu_\epsilon, \alpha_1, \alpha_2, \ldots, \alpha_p)$. Then, the conditional least squares estimators (CLS) for $\boldsymbol{\xi}$ can be obtained by minimizing

$$Q_n(\boldsymbol{\xi}) = \sum_{t=p+1}^n [y_t - E(y_t|\mathcal{F}_{t-1})]^2. \tag{8}$$

It can be seen that all regularity conditions proposed by Klimko and Nelson (1978) are satisfied to give the following asymptotic normality.

Theorem 2 *Let $\widehat{\boldsymbol{\xi}}_n^{\mathrm{LS}}$ be the CLS minimizing (8). Then, if $E\epsilon_t^4 < \infty$, we have*

$$\sqrt{n}\left(\widehat{\boldsymbol{\xi}}_n^{\mathrm{LS}} - \boldsymbol{\xi}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1})$$

where

$$\mathbf{V} = \left[E\Big(\frac{\partial E(y_{p+1}|\mathcal{F}_p)}{\partial \xi_i}\frac{\partial E(y_{p+1}|\mathcal{F}_p)}{\partial \xi_j}\Big)\right],$$
$$\mathbf{W} = E\Big[A_p^2(\boldsymbol{\xi})\frac{\partial E(y_{p+1}|\mathcal{F}_p)}{\partial \xi_i}\frac{\partial E(y_{p+1}|\mathcal{F}_p)}{\partial \xi_j}\Big],$$
and $A_p(\boldsymbol{\xi}) = y_{p+1} - E(y_{p+1}|\mathcal{F}_p)$.

Table 1 Mean-squared errors for α_1, α_2, and λ in $y_t = \alpha_1 \odot y_{t-1} + \alpha_2 \odot y_{t-2} + \epsilon_t$ where $\alpha_2 = -0.3$, $\epsilon_t \sim Poisson(\lambda)$ with $\lambda = 1$ (the numbers in parentheses are biases)

Sample size	α_1	MSE α_1	MSE α_2	MSE λ	Sample size	α_1	MSE α_1	MSE α_2	MSE λ
	−0.7	0.019	0.019	0.041		−0.7	0.010	0.010	0.022
		(−0.002)	(0.012)	(−0.005)			(−0.001)	(0.011)	(−0.001)
50	−0.3	0.023	0.021	0.045	100	−0.3	0.011	0.009	0.021
		(−0.001)	(0.010)	(−0.005)			(0.005)	(0.006)	(−0.016)
	0.3	0.024	0.021	0.065		0.3	0.010	0.010	0.028
		(0.026)	(0.017)	(−0.036)			(0.010)	(0.005)	(−0.015)
	0.7	0.024	0.021	0.081		0.7	0.011	0.009	0.035
		(0.025)	(0.024)	(−0.072)			(0.017)	(0.005)	(−0.030)

Observe that

$$E(\alpha \odot y_{t+k}|\mathcal{F}_t) = \alpha y_{t+k} \text{ for } k \le 0 \text{ and}$$
$$E(\alpha \odot y_{t+k}|\mathcal{F}_t) = E\big(E(\alpha \odot y_{t+k}|\mathcal{F}_{t+k})\big|\mathcal{F}_t\big) = \alpha E(y_{t+k}|\mathcal{F}_t) \text{ for } k \ge 1.$$

Thus, the conditional expectation of $\alpha \odot y_{t+k}$ given $\mathcal{F}_t$ is

$$E(\alpha \odot y_{t+k}|\mathcal{F}_t) = \alpha E(y_{t+k}|\mathcal{F}_t).$$

This implies that the conditional expectation of y_{t+k} which follows a INARS(p) process is the same as that of the usual AR(p) process:

$$E(y_{t+k}|\mathcal{F}_t) = \sum_{i=1}^{p} \alpha_i E(y_{t+k-i}|\mathcal{F}_t) + \mu_\epsilon. \tag{9}$$

Hence, this $E(y_{t+k}|\mathcal{F}_t)$ is the minimum variance predictor of y_{t+k} conditioned on $\mathcal{F}_t$ and obtained by recursion as the usual AR(p).

We perform Monte Carlo simulations to explore how well the CLS works. Simulated samples are obtained from an INARS(2) process with $\alpha_1 = -0.7$, -0.3, 0.3, and 0.7 but with fixing $\alpha_2 = -0.3$ because other values of α_2 yield a similar pattern as that in Table 1. We generate ϵ_t from Poisson with mean 1. Although MSE and bias of λ slightly increase as α_1 increases, based on MSEs presented in Table 1, the CLS estimates are reliable for both sample sizes, $T = 50, 100$.

4 Bootstrap approach

The detection of unusual high incidence in AIDS is of importance for a prevention planning. Since forecasts of a time series are obtained under the assumption that the current conditions and patterns hold in the future time, forecasts and their confidence intervals are useful information to develop a statistical strategy for public health officials to detect an unusual occurrence of AIDS.

The most common procedure for constructing forecasts in time series model is to use the conditional expectations because these conditional expectations yield forecasts with minimum mean squared error as described in Sect. 3. However, this method does not preserve the integer-valued nature of the data in generating forecasts when the time series is integer-valued. Moreover, confidence intervals of forecasts in the INARS(p) model require the distribution of forecasted errors, which may be impossible to obtain mainly because of the distributional complexity accrued from the signed binomial thinning operator.

Some bootstrap approaches have been proposed as distribution free alternatives to obtain forecasts and their confidence intervals. Among others, we employ the bootstrap method proposed by Pascual et al. (2004) after some modifications to incorporate the nature of interger-valued time series as the following five steps. Since we analyze the AIDS data given in Fig. 1, we use $y_t = X_t - X_{t-1}$ where X_t is a nonnegative integer valued original series.

Step 1: Compute residuals $r_t = y_t - \sum_{i=1}^{p} \hat{\alpha}_i y_{t-i}$ for $t = p+1, \ldots, n$ where n is the sample size and $\hat{\alpha}_i$'s are the conditional estimates from (8).

Step 2: Construct the empirical distribution for modified residuals r_t^* defined by $r_t^* = [r_t]$ as in Cardinal et al. (1999), where $[\cdot]$ represents the value rounded to the nearest integer.

Step 3: Draw a i.i.d sample r_t^* from the empirical distribution and define X_t^* by the recursion: for $t = 1, 2, \ldots, n$

$$X_t^* = X_{t-1}^* + \sum_{i=1}^{p} \hat{\alpha}_i \odot y_{t-i}^* + r_t^*$$

where $X_t^* = 0$ if $X_t^* < 0$ and $y_t^* = X_t^* - X_{t-1}^*$.

Step 4: Based on $\{y_2^*, y_3^*, \ldots, y_T^*\}$, compute the conditional estimates $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_p^*$.

Step 5: Compute future bootstrap observations by recursion:

$$X_{T+h}^* = X_{T+h-1}^* + \sum_{i=1}^{p} \hat{\alpha}_i^* \odot y_{T+h-i}^* + r_t^*$$

where T denotes the last observation, $X_t^* = 0$ if $X_t^* < 0$, and $y_t^* = y_t$ for $t \leq T$.

Repeating Steps 3–5 B times, we have a bootstrap distribution function of X_{T+h}^* given by $F_{X_{T+h}}^*(x) = \frac{\#\{X_{T+h}^{*b} \leq x\}}{B}$ for $b = 1, 2, \ldots, B$. The $(1-\alpha)100\%$ confidence interval for X_{T+h} is $\left(F_{X_{T+h}}^{*-1}(\frac{\alpha}{2}), F_{X_{T+h}}^{*-1}(1 - \frac{\alpha}{2})\right)$. As suggested in Alonso et al. (2002), we generate INARS(p) resample from Step 3 with sample size equal to $n + 100$ and discard the first 100 observations.

Step 2 and the procedure $X_t^* = 0$ if $X_t^* < 0$ in Steps 3 and 5 are used to reflect the data coherency that original series X_t is non-negative integer-valued. The greatest difference from the previous bootstrap methods for a continuous time series model is the signed binomial thinning operator used in Steps 3 and 5, by which we incorporate the variability caused by the signed binomial thinning operator into forecasts and confidence intervals. Since the INARS(p) model includes the disturbance term ϵ_t with mean μ_ϵ which is not necessarily zero, a centering procedure for r_t^* is not needed in Step 2. This is another difference from the bootstrap procedure for a conventional ARIMA model.

5 Application

The US Center for Disease Control and Prevention (CDC) releases data involving statistical information on AIDS. This includes demographic factors, case-definition, date of diagnosis, and other information for AIDS. Our data analysis takes the monthly numbers of new patients diagnosed with AIDS in which the sexual orientation is taken account. We used the information of adult/adolescent homosexual male in Baltimore.

The number of AIDS patients per month shows a clear linear time trend with the numbers dropping below 20 after January 1997 as shown in Fig. 1. Thus, we consider $y_t = X_t - X_{t-1}$ for our data analysis. Since INARS(p) process has the same autocorrelation structure as the usual continuous AR(p) process from the proof of Lemma 1, the sample autocorrelation functions can be used to check for the stationarity of y_t and to select an appropriate p in the INARS(p). Using the sample autocorrelation function of $y_t = X_t - X_{t-1}$, we temporarily allowed p to equal 6. Because the variance of the error term in INARS(p) is always larger than that in the standard AR(p) as shown in Sect. 3, Table 2 shows that μ_ϵ and α_6 in INARS(6) are not significant from Theorem 2, whereas those in the standard AR(6) are significant. This means that the use of AR(6) as an approximation of the INARS(6) may lead to incorrect model specification. It is worth mentioning that CLS estimation gives the same estimates for both the standard AR(p) and INARS(p) models and that the difference lies in the asymptotic expression for the variance as given in Theorem 2. From the data analysis, the following INARS(5) model is implemented:

$$X_t = X_{t-1} - \underset{(0.163)}{0.816} \odot y_{t-1} - \underset{(0.202)}{0.603} \odot y_{t-2} - \underset{(0.163)}{0.547} \odot y_{t-3} - \underset{(0.123)}{0.535} \odot y_{t-4} - \underset{(0.087)}{0.227} \odot y_{t-5} + \epsilon_t, \tag{10}$$

where the numbers in parentheses are standard errors. All roots of the characteristic equation of (10) are $\lambda = -0.60, -0.49 \pm 0.55i$, and $0.39 \pm 0.74i$, indicating that all roots are inside the unit circle and the INARS(5) is stationary by Theorem 1. We will further discuss for the final model selection between this INARS(5) and the INARS(6) provided in Table 3, by comparing mean absolute errors (MAE) and mean-squared errors (MSE) of their fitted values.

The AIDS data are also used to demonstrate advantages of the INARS(p) model over the non-normal time series model such as a following quasi-likelihood model introduced by Zeger and Qaqish (1988) with a constant dispersion

Table 2 Parameter estimates and standard errors

		μ_ϵ	α_1	α_2	α_3	α_4	α_5	α_6
Estimates		−0.313	−0.885	−0.765	−0.712	−0.717	−0.473	−0.299
Standard error	INARS(6)	0.679	0.220	0.212	0.209	0.185	0.177	0.158
	AR(6)	0.105	0.096	0.122	0.125	0.125	0.123	0.096

Table 3 MAE and MSE for bootstrap fitted values and mean length of bootstrap 95% confidence intervals

	INARS(5)	INARS(6)	AR(5)	AR(6)	Quasi-likelihood
MAE	3.43	3.42	3.63	3.37	3.31
MSE	19.70	19.02	24.08	18.81	21.48
mean length of 95 % C.I	17.11	16.07	16.61	14.63	4.46

parameter.

$$\begin{aligned} E(X_t|\mathcal{F}_{t-1}) &= \mu_t = \exp(\beta_0 + \beta_1 t \times 10^{-3})(X_{t-1})^{\alpha_1} \cdots (X_{t-5})^{\alpha_5} \\ \mathrm{Var}(X_t|\mathcal{F}_{t-1}) &= \phi \mu_t \end{aligned} \tag{11}$$

where the 10^{-3} is used to maintain consistency of the estimated β_1 (Davis and Dunsmuir 2000) and ϕ is a dispersion parameter. Using the estimating equation defined by the two moments given in (11),

$$\begin{aligned} \log \mu_t = 4.78 - 24.2t \times 10^{-3} - 0.018 \log X_{t-1} - 0.015 \log X_{t-2} \\ - 0.088 \log X_{t-3} - 0.185 \log X_{t-4} - 0.009 \log X_{t-5} \\ \text{and} \quad \phi = 1.339 \end{aligned} \tag{12}$$

where all coefficients of X_{t-1} through X_{t-5}, except the coefficient of X_{t-4}, are not significant unlike those in the INARS(5) model.

We use the bootstrap approach described in Section 4 to select our final model between the INARS(5) and the INARS(6) and to compare INARS models with the quasi-likelihood model given in (12) and with the standard AR models which have the same CLS as the corresponding INARS models. The main reason to use the bootstrap method is to incorporate the integer-valued nature of the AIDS data and to avoid the difficulty in calculating the error of forecasts obtained from the INARS models. To produce coherent fitted values or forecasts, we use the median of the bootstrap distribution for the two INARS models, instead of the conditional expectation of the distribution, which usually used in the standard ARMA model. For a fair comparison, we also use bootstrap fitted values obtained by the conditional expectations and confidence intervals, generated from the AR(5) and the AR(6) models. That is, under AR(5) or AR(6), we apply the bootstrap procedures of Pascual et al. (2004) to the AIDS data with the constraint of non-negativeness of X_t.

Table 3 shows MAE and MSE for bootstrap fitted values and mean lengths of 95% bootstrap confidence intervals for INARS(5), INARS(6), AR(5), AR(6), and the quasi-likelihood model using the AIDS data from July 1993 to December 2001.

The MAE and MSE of INARS(5) in fitted values are little different from those of INARS(6). Thus, our final model is the INARS(5) by the parsimony policy. Contrary to this, when we use AR models as an approximation, the AR(6) is better than than AR(5) as expected in Table 2 because the coefficient of lag six α_6 of AR(6) model is significant. It is also interesting that MAE and MSE of INARS(5) are smaller than those of AR(5), which may stem from the coherent fitted values of INARS(5). Because INARS model has larger variance than the corresponding AR model as discussed in Sect. 3, the mean lengths of 95% confidence intervals of INARS(5) and INARS(6) are wider than those of AR(5) and AR(6), respectively.

Table 3 also shows that the MAE of quasi-likelihood model (12) is slightly smaller than that of INARS(5). However, the mean length of its 95% confidence intervals is too short. As a result, the quasi-likelihood model seriously underestimates the variance of its fitted values as shown in Fig. 1. The underestimated variance by the quasi-likelihood model produces a very narrow confidence interval. Most observations are located outside of the 95% confidence intervals, implying that the quasi-likelihood model does not overcome the over-dispersion problem in the AIDS data. On the other hand, the 95% confidence bounds from the INARS(5) include most observations except three observations, implying that the INARS(5) is a good alternative model that can be used to avoid the over-dispersion problem frequently arising from an integer-valued time series.

Because the time series of the AIDS data is apparently fading away, instead of forecasting future values, we split the data into two parts: the data from January 1993 to December 2000 for estimating parameters; and the data from January 2001 to December 2001 for out-of-sample forecast. Namely, we calculate bootstrap out-of-sample forecasts and their 95% confidence intervals for 12 months from January 2001 to December 2001 by re-estimating INARS(5) model using AIDS data from January 1993 to December 2000. Table 4 shows that the 95% confidence intervals contain all observations, implying that any AIDS incident during the 12 months is not an unusual event. The 95% confidence interval used in this section is, in fact, slightly wider than a true 95% confidence interval because we are using integers. The MAE of out-of-sample

Table 4 Bootstrap out-of-sample forecasts and 95% confidence intervals for year 2001

Month		1	2	3	4	5	6	7	8	9	10	11	12
Observation		10	3	10	9	5	9	8	3	3	3	2	0
Forecast		5	6	6	5	5	4	4	4	4	4	4	4
95%	L95	0	0	0	0	0	0	0	0	0	0	0	0
CI	U95	15	15	15	15	15	14	15	15	15	15	15	15

forecasts is 2.83 which is smaller than that of fitted values (i.e., 3.43 in Table 3), indicating stable forecasts of the INARS(5) model.

6 Conclusion

We developed a new integer-valued time series model for a non-stationary time series, and called INARS(p) process, as a counterpart of the usual AR(p) process. The INARS(p) can deal with negative-valued and negative-correlated count data unlike the previous integer-valued time series model. We showed that INRAS(p) process is ergodic and stationary under the same condition as that of the conventional AR(p) process.

We showed that the conditional expectation in INARS(p) process for forecasts is identical to that in AR(p). However, the conditional expectation lacks data coherency when the time series is integer-valued such as the AIDS data. To preserve the integer-valued nature of data and to avoid the difficulty in deriving the distributional properties of the forecasts of INARS(p) model, we use a bootstrap approach as a distribution free alternative. Through this bootstrap approach, we showed that INARS(p) model is more appropriate than the corresponding AR(p) model for the AIDS data. Our data analysis also showed that the confidence interval under a quasi-likelihood model is too narrow to include the corresponding observation because of serious underestimation in the variance of its fitted value. However, confidence intervals by the INARS(p) include most of observations by the property of the INARS(p) process with the signed binomial thinning. Thus, the INARS(p) is a good alternative to the previous models in resolving the over-dispersion for integer-valued time series.

A Appendix: Proof of proposition 1

Since

$$
\begin{aligned}
&E(\alpha_1 \odot y_t \cdot \alpha_2 \odot y_t) \\
&\quad = E\Big(\mathrm{sgn}(\alpha_1)\mathrm{sgn}(y_t)\sum_{j=1}^{|y_t|} w_j(\alpha_1) \cdot \mathrm{sgn}(\alpha_2)\mathrm{sgn}(y_t)\sum_{j=1}^{|y_t|} w_j(\alpha_2)\Big) \\
&\quad = \alpha_1\alpha_2 E|y_t|(|y_t| - 1),
\end{aligned}
$$

where we used $(\sum_{j=1}^{|y_t|} w_j(\alpha_1), \sum_{j=1}^{|y_t|} w_j(\alpha_2)) \sim$ multinomial$(|\alpha_1|, |\alpha_2|, |y_t|)$ conditioning on y_t. Thus, since $E(\alpha \odot y_t) = \alpha E(y_t)$, we have

$$
\mathrm{cov}(\alpha_1 \odot y_t, \alpha_2 \odot y_t) = \alpha_1\alpha_2(\mathrm{var}(y_t) - E|y_t|). \tag{13}
$$

Using the conditional independence of $\alpha_1 \odot y_t$ and $\alpha_2 \odot y_{t-1}$ given in y_t, we have $E(\alpha_1 \odot y_t \cdot \alpha_2 \odot y_{t-1}) = E\big(E(\alpha_1 \odot y_t|y_t)E(\alpha_2 \odot y_{t-1}|y_t)\big) = \alpha_1 E(y_t \cdot \alpha_2 \odot y_{t-1})$

and thus

$$\mathrm{cov}(\alpha_1 \odot y_t, \alpha_2 \odot y_{t-1}) = \alpha_1 \mathrm{cov}(y_t, \alpha_2 \odot y_{t-1}). \tag{14}$$

Now, using (13), (14), and stationarity, since $\mathrm{cov}(y_t, \alpha_2 \odot y_{t-1}) = \mathrm{cov}(\alpha_1 \odot y_{t-1} + \alpha_2 \odot y_{t-2} + \epsilon_t, \alpha_2 \odot y_{t-1}) = \alpha_1\alpha_2(\mathrm{var}(y_t) - E|y_t|) + \alpha_2\,\mathrm{cov}(y_t, \alpha_2 \odot y_{t-1})$, we have

$$\mathrm{cov}(y_t, \alpha_2 \odot y_{t-1}) = \frac{\alpha_1\alpha_2(\gamma(0) - E|y_t|)}{1 - \alpha_2}. \tag{15}$$

Finally, using conditional expectation technique, we have

$$\mathrm{cov}(\alpha \odot y_t, y_{t-k}) = \alpha \cdot \mathrm{cov}(y_t, y_{t-k}) \text{ for } k \geq 0. \tag{16}$$

From INARS(2) expressed by $y_t = \alpha_1 \odot y_{t-1} + \alpha_2 \odot y_{t-2} + \epsilon_t$, we have

$$\begin{aligned} &\mathrm{cov}(y_t, y_{t-1}) = \mathrm{cov}(\alpha_1 \odot y_{t-1}, y_{t-1}) + \mathrm{cov}(\alpha_2 \odot y_{t-2}, y_{t-1}) \\ \text{and} \quad &\mathrm{cov}(y_t, y_{t-2}) = \mathrm{cov}(\alpha_1 \odot y_{t-1}, y_{t-2}) + \mathrm{cov}(\alpha_2 \odot y_{t-2}, y_{t-2}). \end{aligned} \tag{17}$$

Plugging (15) and (16) into (17), we have the results.

B Appendix: Proof of proposition 2

1. By conditional independence of $\alpha_1 \odot y_t$ and $\alpha_2 \odot y_{t'}$ given y_t and $y_{t'}$, for $0 \leq \alpha_1, \alpha_2 \leq 1$,

$$\begin{aligned} E(\alpha_1 \odot y_t \cdot \alpha_2 \odot y_{t'}) &= E\big(E(\alpha_1 \odot y_t|y_t)E(\alpha_2 \odot y_{t'}|y_{t'})\big) \\ &= \alpha_1\alpha_2 E(y_t y_{t'}). \end{aligned} \tag{18}$$

   Thus, the covariance function follows.
2. Since y_{1t} and y_{2t} have the same sign and the same counting series,

$$E|\alpha \odot y_{1t} - \alpha \odot y_{2t}| = E|\sum_{j=1}^{|y_{1t}|} w_j - \sum_{j=1}^{|y_{2t}|} w_j|. \tag{19}$$

If $|y_{1t}| \geq |y_{2t}|$, (19) equals $E\left|\sum_{j=|y_{2t}|+1}^{|y_{1t}|} w_j\right| = E\left|\sum_{j=1}^{|y_{1t}|-|y_{2t}|} w_j\right| = E\sum_{j=1}^{|y_{1t}|-|y_{2t}|} w_j$. Similarly, if $|y_{2t}| \geq |y_{1t}|$, (19) equals $E\sum_{j=1}^{|y_{2t}|-|y_{1t}|} w_j$. Hence, we have

$$E\left|\sum_{j=1}^{|y_{1t}|} w_j - \sum_{j=1}^{|y_{2t}|} w_j\right| = E\sum_{j=1}^{||y_{1t}|-|y_{2t}||} w_j = |\alpha| E\big(\big||y_{1t}| - |y_{2t}|\big|\big). \tag{20}$$

Finally, because y_{1t} and y_{2t} have the same sign, $\big||y_{1t}| - |y_{2t}|\big| = \big|\text{sgn}(y_{1t})|y_{1t}| - \text{sgn}(y_{2t})|y_{2t}|\big| = |y_{1t} - y_{2t}|$. This shows the first claim. From (20), we have $E(\alpha \odot y_{1t} - \alpha \odot y_{2t})^2 = E\big(\sum_{j=1}^{||y_{1t}|-|y_{2t}||} w_j\big)^2$. Thus, $E\big((\sum_{j=1}^{||y_{1t}|-|y_{2t}||} w_j)^2 \big| y_{1t}, y_{2t}\big) = \alpha^2(\big||y_{1t}|-|y_{2t}|\big|)^2 + |\alpha|(1-|\alpha|)(\big||y_{1t}|-|y_{2t}|\big|)$. The relationship of $\big||y_{1t}|-|y_{2t}|\big| = |y_{1t} - y_{2t}|$ produces the result.

C Appendix: Proof of theorem 1

Since we let $y_t = \lim_{n\to\infty} y_{n,t}$, both y_t and $y_{n,t}$ have the same sign for any n. Thus, by Proposition 2, $E(\alpha \odot y_{n,t} - \alpha \odot y_t)^2 = \alpha^2 E(y_{n,t} - y_t)^2 + |\alpha|(1-|\alpha|)E|y_{n,t} - y_t|$.

It can be shown by the same approach of Jin-Guan and Yuan (1991) that $y_{n,t} \xrightarrow{L_2} y_t$. Thus, $E(\alpha \odot y_{n,t} - \alpha \odot y_t)^2$ converges to 0 as $n \to \infty$. Thus, the process $\{y_t\}$ satisfies (5).

For uniqueness of such a process $\{y_t\}$, suppose that we have another process $\{y_t^*\}$ such that $y_{n,t} \xrightarrow{L_2} y_t^*$. Then, by Hölder inequality, for some constant c and $|\lambda| < 1$,

$$E|y_t - y_t^*| \le \big(E(y_{n,t} - y_t)^2\big)^{1/2}\big(E(y_{n,t} - y_t^*)^2\big)^{1/2} = c\lambda^n.$$

Thus, $E|y_t - y_t^*| = 0$ which implies $y_t = y_t^*$ almost surely. By L_2 convergence of $y_{n,t}$ to y_t, $\lim_{n\to n} \text{Cov}(y_{n,t'}, \epsilon_t) = \text{Cov}(y_{t'}, \epsilon_t)$. Since $\text{Cov}(y_{n,t'}, \epsilon_t) = 0$ for $t' < t$, $\text{Cov}(y_{t'}, \epsilon_t) = 0$ for $t' < t$.

Since $y_{n,t} = 0$ for $n < 0$, recursively solving the Eq. (4), regardless of t, $y_{n,t}$ can be expressed as a function of $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-n}$ with the same signed binomial thinning operators which are expressed only by $\alpha_1, \alpha_2, \ldots, \alpha_p$. Hence, the distribution of $y_{n,t}$ depends only on n but not on t. This implies that for each n and T, $(y_{n,0}, y_{n,1}, \ldots, y_{n,T})$ and $(y_{n,k}, y_{n,1+k}, \ldots, y_{n,T+k})$ have the same distribution for every k. The L_2 convergence of $y_{n,t}$ to y_t as $n \to \infty$ means that $y_{n,t}$ converges to y_t in probability. Thus, $\sum_{i=0}^{T} a_i y_{n,t+i}$ converges to $\sum_{i=0}^{T} a_i y_{t+i}$ in probability for all real values of a_i's. This implies by Cramer-Wold device that

$$\begin{aligned}(y_{n,0}, y_{n,1}, \ldots, y_{n,T}) &\xrightarrow{d} (y_0, y_1, \ldots, y_T) \text{ and} \\ (y_{n,k}, y_{n,1+k}, \ldots, y_{n,T+k}) &\xrightarrow{d} (y_k, y_{1+k}, \ldots, y_{T+k}).\end{aligned} \tag{21}$$

Since $(y_{n,0}, y_{n,1}, \ldots, y_{n,T})$ and $(y_{n,k}, y_{n,1+k}, \ldots, y_{n,T+k})$ have the same distribution for every k, $(y_0, y_1, \ldots, y_T)$ and $(y_k, y_{1+k}, \ldots, y_{T+k})$ also have the same distribution for every k from (21). This show that y_t is stationary.

Let $\mathbf{w}(t)$ be all counting series in $\alpha_1 \odot y_{t-1} + \ldots + \odot\alpha_p \odot y_{t-p}$ and $\sigma(y)$ be a σ-field generated by a random variable y. Note that $\sigma(y_t, y_{t-1}, \ldots) \subset \sigma(\mathbf{w}(t), \epsilon_t, \mathbf{w}(t-1), \epsilon_{t-1}, \ldots)$. Because $\{\mathbf{w}(t), \epsilon_t\}$ are independent sequence, Kolmogorov's Zero-One law implies that any event in the tail σ-filed denoted by $\bigcap_{t=1}^{\infty} \sigma(y_t, y_{t+1}, \ldots)$, has probability 0 or 1. This shows by Durrett (1991) that y_t is ergodic.

D Appendix: Proof of lemma 1

Since the process expressed by $y_t = \sum_{i=1}^{p} \alpha_i \odot y_{t-i} + \epsilon_t$ is stationary, it is easy to see that

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{p-k} \tag{22}$$

where $\gamma_k = \mathrm{Cov}(y_t, y_{t-k})$ and $\gamma_k = \gamma_{-k}$ by stationarity. Thus $\hat{\alpha}_i$ for $i = 1, 2, \ldots, p$ is strongly consistent because $\hat{\gamma}_k \xrightarrow{\text{a.s.}} \gamma_k$ by (6).

Observe that

$$\epsilon_t = y_t - \sum_{i=1}^{p} \alpha_i \odot y_{t-i}.$$

Thus, we have

$$E(\epsilon_t) = \left(1 - \sum_{i=1}^{p} \alpha_i\right)\mu \tag{23}$$

where $E(y_t) = \mu$ by stationarity. A little calculation shows that

$$\begin{aligned} E(\epsilon_t^2) = (\gamma_0 + \mu^2)\left(1 + \sum_{i=1}^{p} \alpha_i^2\right) + \sum_{i=1}^{p} |\alpha_i|(1 - |\alpha_i|)E|y_{t-i}| \\ -2\sum_{i=1}^{p} \alpha_i(\gamma_i + \mu^2) + 2\sum_{1 \le i < j \le p} \alpha_i\alpha_j(\gamma_{|i-j|} + \mu^2). \end{aligned} \tag{24}$$

Since

$$\frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t = \frac{1}{n}(y_t - \hat{\alpha}_1 y_{t-1} - \cdots - \hat{\alpha}_p y_{t-p}),$$

Eq. (6) and strong convergence of $\hat{\alpha}$s imply that $\bar{\epsilon}_n$ converges $(1 - \sum_{i=1}^{p} \alpha_i)\mu$ almost surely. Similarly, one also can show that, by (6) and consistency estimators $\hat{\alpha}$'s,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t^2 \xrightarrow{a.s} (\gamma_0 + \mu^2)\left(1 + \sum_{i=1}^{p} \alpha_i^2\right) - 2\sum_{i=1}^{p} \alpha_i(\gamma_i + \mu^2) \\ +2\sum_{1 \le i < j \le p} \alpha_i\alpha_j(\gamma_{|i-j|} + \mu^2). \end{aligned} \tag{25}$$

Thus, (24) and (25) yield that $\hat{\sigma}_\epsilon^2$ converges to σ_ϵ^2 almost surely.

References

Alonso AM, Peña D, Romo J (2002) Forecasting time series with sieve bootstrap. J Stat Plan Infer 100:1–11

Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. J Time Ser Anal 8:261–275

Aly AA, Bouzar N (1994) On some integer-valued autoregressive moving average models. J Multivariate Anal 50:132–151

Alzaid AA, Al-Osh M (1990) An integer-valued pth-order autoregressive structure (INAR(p)) process. J Appl Probab 27:314–324

Anderson RM, Grenfell BT (1984) Oscillatory fluctuations in the incidence of infectious disease and impact of vaccination: time series analysis. J Hyg-Cambridge 93:587–608

Breslow N (1990) Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. J Am Stat Assoc 85:565–571

Cardinal M, Roy R, Lambert J (1999) On the application of integer-valued time series models for the analysis of disease incidence. Stat Med 18:2025–2039

Center for Disease Control and Prevention (CDC 2001) AIDS Public Information Data Set

Cordeiro GM, Botter DA (2001) Second-order of maximum likelihood estimates in overdispersed generalized linear models. Stat Probabil Lett 55:269–280

Davis RA, Dunsmuir WTM, Wang Y (2000) On autocorrelation in a Poisson regression model. Biometrika 87:491–505

Dey DK, Gelfand AE, Peng F (1997) Overdispersed generalized linear models. J Stat Plan Infer 64:93–107

Durrett R (1991) Probability. Wadsworth, Belmont, CA, pp 292–308

Findley DF (1986) On bootstrap estimates of forecast mean square errors for autoregressive processes. In: Allen DM (ed) Computer science and statistics: the interface NorthHolland Amsterdam, pp 11–17

Freeland RK, McCabe BPM (2004) Forecasting discrete valued low count time series. Int J Forecasting 20:427–434

Grigoletto M (1998) Bootstrap prediction intervals for autoregressions: some alternative. Int J Forecasting 14:447–456

Jin-Guan D, Yuan L (1991) The integer-valued autoregressive (INAR(p)) model. J Time Ser Anal 12:129–142

Jung RC, Tremayne AR (2006) Coherent forecasting in integer time series models. Int J Forecasting 22:223–238

Klimko LA, Nelson PI (1978) On conditional least squares estimation for stochastic processes. Ann Stat 6:629–642

Masarotto G (1990) Bootstrap prediction intervals for autoregressions. Int J Forecasting 6:229–239

McCormick WP, Park YS (1997) An analysis of Poisson moving-average processes. Probab Eng Inform Sci 11:487–507

McCullagh P, Nelder JA (1989) Generalized linear models. Chapman & Hall, London

McKenzie E (1986) Autoregressive moving-average processes with negative binomial and geometric marginal distributions. Adv Appl Probab 18:679–705

Park YS, Oh CH (1997) Some asymptotic properties in INAR(1) processes with Poisson marginals. Stat Pap 38:287–302

Pascual L, Romo J, Ruiz, E (2004) Bootstrap predictive inference for ARIMA processes. J Time Ser Anal 25:449–465

Thombs LA, Schucany WR (1990) Bootstrap prediction intervals for autoregression. J Am Stat Assoc 85:486–492

Waiter L, Richardson S, Hubert B (1991) A time series construction of an alert threshold with application to S. bovismorbificans in France. Stat Med 10:1493–1509

Zaidi AA, Schnell DJ, Reynolds G (1989) Time series analysis of sypholis surveillance data. Stat Med 8:353–362

Zeger SL, Qaqish B (1988) Markov regression models for time series: a Quasi-Likelihood approach. Biometrics 44:1019–1031