Some Estimation and Forecasting Procedures in Possion-Lindley INAR(1) Process

Yu Wang and Haixiang Zhang*

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

Abstract We study the estimation and forecasting in first-order integer-valued autoregressive process with Poisson-Lindley (PLINAR(1)) marginal distribution (Mohammadpour et al., 2018). Quasi-likelihood estimators are proposed for the parameters of interest and their asymptotic properties are derived. Two methods for coherent point prediction are given and the prediction intervals for future data are constructed. We present some simulations to verify rationality of the proposed estimation and prediction methods. An application to a real data about animal's anorexia is also provided.

Keywords: Forecasting; INAR models; Poisson-Lindley distribution; Prediction interval; Quasi-likelihood.

1 Introduction

Recently, integer-valued time series has been widely used in many fields, such as the number of daily transactions in the stock market (Brännäs and Quoreshi, 2010), the weekly number of patients in a hospital caused by influenza (Moriña et al., 2011), and the annual severe hurricane counts in the North Atlantic (Livsey et al., 2018), etc. The main feature of these data is their integer-valued structure, so many traditional autoregressive-type models can not fit this kind of data well. To tackle with this count time series data, the thinning operator-based technique (Steutel and Van Harn, 1979) is very popular. For example, Al-Osh and Alzaid (1987) proposed a first-order integer-valued autoregressive (INAR(1)) process, which has layed the foundation for thinning-based INAR models. Later on, a large number of new INAR models are developed. Du and Li (1991) presented a pth-order integer-valued autoregressive model. Zheng et al. (2007) proposed a first-order random coefficient integer-valued autoregressive process. Ristić and Bakouch (2009) defined

*Corresponding author: haixiang.zhang@tju.edu.cn (H. Zhang)

a new INAR(1) process with geometric distribution based on negative binomial thinning operator. Bakouch and Ristić (2010) proposed an INAR(1) process with zero truncated Poisson marginal distribution. Zhang et al. (2010) gave a p-th order integer-valued autoregressive process with signed generalized power series thinning operator. There was a significant new breakthrough in INAR modeling which happened recently by introducing a random environment in integer-valued autoregressive process by Nastić et al. (2016). Mohammadpour et al. (2018) proposed a first-order integer-valued autoregressive with Poisson-Lindley marginal distribution. More related research can be referred to Weiß (2008) and Scotto et al. (2015).

As we know, estimation and forecasting are two important topics for the analysis of time series data. Many results on parameter estimation for INAR models have appeared in the literature. Zheng et al. (2006) used the maximum likelihood, conditional least squares, modified quasi-likelihood and generalized method of moments to estimate the parameters of interest in pth-order random coefficient INAR processes. Bu et al. (2008) developed a general framework for maximum likelihood estimation of higher-order integer-valued autoregressive processes. Drost, et al. (2009) considered the semiparametric efficient estimation for INAR(p) models. Zhang et al. (2011) studied the empirical likelihood based inference for the random coefficient integer-valued autoregressive processes. Martin et al. (2014) introduced an efficient method of moment estimator for integer-valued time series. Pedeli et al. (2015) proposed the likelihood estimation for INAR(p) model by saddlepoint approximation. Zhang et al. (2017) proposed the regularized estimation method for INAR(p) processes via penalty-based approach. To the best of our knowledge, few articles are about predictions for count time series. McCabe et al. (2011) gave an efficient probabilistic forecast method for count time series data. Maiti and Biswas (2015) and Awale et al. (2017) considered the coherent forecasting for INAR(1) process with geometric marginals. Maiti et al. (2016) studied the forecasting for count time series using Box-Jenkins's AR(p) model. The aim of this work is to study both estimation and forecasting for PLINAR(1) process (Mohammadpour et al., 2018).

The remainder of this article is organized as follows. In Section 2, we introduce the motivation and provide the definition and some statistical properties of the PLINAR(1) process. In Section 3, we propose the quasi-likelihood estimator of model parameter and derive its asymptotic properties. In Section 4, several forecast methods for PLINAR(1) process are presented. Some simulation results and an application are reported in Sections 5 and 6, respectively. The concluding remarks are given in Section 7.

2 Motivation and model definition

2.1 Motivation

Time series of counts are encountered in many context, and usually dynamic in nature with significant overdispersion relative to the means. For example, the monthly number of animal's anorexia from January 2003 to December 2009 in a region of New Zealand (see Figure 1), where the mean and variance are 0.8214 and 2.8954, respectively. To model this kind of data, Mohammadpour et al. (2018) proposed an INAR(1) model with Poisson-Lindley (PL) marginal distribution, which can effectively describe the over-dispersed property of count data. Here a random variable X is said to have a PL distribution (Sankaran, 1970) if its probability mass function can be written as

$$P(X = x) = \frac{\theta^2(x + \theta + 2)}{(\theta + 1)^{x+3}},\ x = 0, 1, 2, \cdots ; \theta > 0. \tag{2.1}$$

Since $E(X) = (\theta + 2)/\{\theta(\theta + 1)\}$, and $Var(X) = (\theta^3 + 4\theta^2 + 6\theta + 2)/\{\theta^2(\theta + 1)^2\}$, we know that $Var(X)/E(X) > 1$, i.e., the PL distribution is overdispersed.

Note that the main advantages of PL distribution are as follows: First, the PL distribution belongs to compound Poisson family similar to negative binomial distribution. It has some common properties like unimodality, over-dispersion, and infinite divisibility (Ghitany and Al-Mutairi, 2009). Second, the PL distribution (2.1) can be regraded as mixture of geometric distribution with parameter $1/(1+\theta)$ and negative binomial distribution with parameters 2 and $1/(1+\theta)$ with mixing weights $\theta/(1 + \theta)$ and $1/(1 + \theta)$, respectively. Thus, the PL distribution is more flexible to fit the data in practice. Third, the skewness and kurtosis of the PL distribution are much smaller than the negative binomial distribution (Ghitany and Al-Mutairi, 2009). Based on the above advantages of PL distribution, the INAR(1) process with PL marginal may have a better fit in certain practical situations (Mohammadpour et al., 2018). From this point of view, we focus on some estimation and forecasting topics for the PLINAR(1) process in the remainder of this article.

2.2 Definition and some basic properties

Here we first review the definition of PLINAR(1) process, which was proposed by Mohammadpour et al. (2018) as follows:

Definition 2.1 (Mohammadpour et al., 2018) *The* PLINAR(1) *process is defined by the following recursive equation*

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, t \geq 1, \tag{2.2}$$

where

(i) the thinning operator "$\circ$" is defined as

$$\alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} B_{i,t},$$

where $\{B_{i,t}\}$ is an i.i.d. Bernoulli sequence with $P(B_{i,t}=1) = 1 - P(B_{i,t}=0) = \alpha \in (0,1)$.

(ii) ϵ_t is a non-negative integer-valued random variable with probability mass function (PMF)

$$f_\epsilon(x) = \alpha h(x) + (1-\alpha) g(x),$$

where h(x) is a degenerate distribution at zero, and g(x) is a probability mass function defined by

$$\begin{aligned} g(x) =& \frac{\theta^2(1-\alpha)^2 + \theta(1-\alpha^2) + 2\alpha}{(\theta(1-\alpha)+1)^2} \frac{\theta}{1+\theta}\Big(1 - \frac{\theta}{1+\theta}\Big)^x \\ &+ \frac{(1-\alpha)}{\theta(1-\alpha)+1}(x+1)\Big(\frac{\theta}{1+\theta}\Big)^2\Big(1-\frac{\theta}{1+\theta}\Big)^x \\ &- \frac{\alpha}{(\theta(1-\alpha)+1)^2} \frac{\theta+1}{\theta+1+\alpha}\Big(1 - \frac{\theta+1}{\theta+1+\alpha}\Big)^x \end{aligned}$$

with $\theta > 0$. Moreover, ϵ_t is independent of $B_{i,t}$ and X_m, for all $m \leq t$. The probability generating function (PGF) of ϵ_t is

$$\Phi_\epsilon(s) = \frac{2+\theta-s}{(1+\theta-s)^2} \frac{\big(\theta+\alpha(1-s)\big)^2}{1+\theta+\alpha(1-s)}. \tag{2.3}$$

Mohammadpour et al. (2018) derived some basic properties of model (2.2), which are summarized as the following two remarks.

Remark 2.1 *The $\{X_t\}$ is a Markov Chain on $\{0,1,2,...\}$ with one-step transition probability*

$$P(X_t = i | X_{t-1} = j) = \sum_{k=0}^{\min(i,j)} \binom{j}{k} \alpha^k (1-\alpha)^{j-k} P(\epsilon_t = i-k).$$

Remark 2.2 *Let $\{X_t\}$ be a stationary process from (2.2), then for $t \geq 1$*
(a) $E(X_t) = \frac{\theta+2}{\theta(\theta+2)} \triangleq \mu$;
(b) $E(X_{t+h-1}|X_{t-1}) = \alpha^h X_{t-1} + (1-\alpha^h)\mu$, for $h \geq 1$;
(c) $Var(X_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + \sigma_\epsilon^2$, where $\sigma_\epsilon^2 = \frac{(1-\alpha)(\theta^3+4\theta^2+6\theta+2+\alpha(\theta^2+4\theta+2))}{\theta^2(\theta+1)^2}$.

The h-step transition probability of $\{X_t\}$ plays an important role in the forecasting topic. Moreover, the probability generating function (PGF) is a classical technique in the process of deriving the h-step transition probability. Thus we are required to present the following result.

Theorem 2.1 *Given X_t, the h-step probability generating function of X_{t+h} is*

$$\Phi_{X_{t+h}|X_t}(s) = (1-(1-s)\alpha^h)^{X_t} \frac{2+\theta-s}{(1+\theta-s)^2} \frac{(\theta+\alpha^h(1-s))^2}{1+\theta+\alpha^h(1-s)}.$$

Proof. By the definition of X_t in (2.2) and recursive iteration procedure, we can derive that

$$\begin{aligned} X_{t+h} &= \alpha \circ X_{t+h-1} + \epsilon_{t+h} \\ &= \alpha^{[h]} \circ X_t + \sum_{i=0}^{h-1} \alpha^{[i]} \circ \epsilon_{t+h-i}, \end{aligned}$$

where $\alpha^{[k]} = \underbrace{\alpha \circ \alpha \circ \cdots \circ \alpha}_{k\ times}$ with $k = 1, \ldots, h$. The PGF of X_{t+h} given X_t is

$$\begin{aligned} \Phi_{X_{t+h}|X_t}(s) &= E(s^{X_{t+h}}|X_t) \\ &= E(s^{\alpha^{[h]}\circ X_t}|X_t) \prod_{i=0}^{h-1} E(s^{\alpha^{[i]}\circ \epsilon_{t+h-i}}). \end{aligned}$$

Let $\alpha^{[k-1]} \circ X_t = Y_{k-1}$, then we have

$$\begin{aligned} E(s^{\alpha\circ Y_{k-1}}|Y_{k-1}) &= \sum_{j=0}^{Y_{k-1}} s^j P(\alpha \circ Y_{k-1} = j|Y_{k-1}) \\ &= (1-(1-s)\alpha)^{Y_{k-1}}, k = 1, 2, \ldots. \end{aligned}$$

Thus,

$$\begin{aligned} E(s^{\alpha^{[h]}\circ X_t}|X_t) &= E\{(1-(1-s)\alpha)^{Y_{h-1}}|X_t\} \\ &= (1-(1-s)\alpha^h)^{X_t}. \end{aligned}$$

Since $E(s^{\alpha^{[h]}\circ X_t}|X_t) = (1-(1-s)\alpha^h)^{X_t} = E(s^{\alpha^h\circ X_t}|X_t)$, it can be concluded that $\alpha^{[h]} \circ X_t|X_t$ follows the binomial distribution with parameters (X_t, α^h). Note that ϵ_t is an i.i.d. sequence, by (2.3) we can obtain that

$$\begin{aligned} \prod_{i=0}^{h-1} E(s^{\alpha^{[i]}\circ \epsilon_{t+h-i}}) &= \prod_{i=0}^{h-1} E(s^{\alpha^{[i]}\circ \epsilon_t}) \\ &= \frac{2+\theta-s}{(1+\theta-s)^2} \frac{(\theta+\alpha^h(1-s))^2}{1+\theta+\alpha^h(1-s)}, \end{aligned}$$

where $\Phi_\epsilon(\cdot)$ is defined in (2.3). This ends the proof. $\square$

Based on the h-step probability generating function in Theorem 2.1, we can derive the h-step transition probability of $\{X_t\}$.

Theorem 2.2 *Given X_t, the h-step transition probability function of X_{t+h} is*

$$P(X_{t+h}=i|X_t=j)=\sum_{k=0}^{\min(i,j)}\binom{j}{k}\alpha^{hk}(1-\alpha^h)^{j-k}P(W_{t+h}=i-k), \tag{2.4}$$

where the probability mass function of W_{t+h} is given as

$$P(W_{t+h}=i)\begin{cases}\alpha^h+(1-\alpha^h)\Big(A_h\frac{\theta}{1+\theta}+B_h(\frac{\theta}{1+\theta})^2+C_h\frac{1+\theta}{1+\theta+\alpha^h}\Big), & i=0,\\ \\ (1-\alpha^h)\Big(A_h\frac{\theta}{1+\theta}(\frac{1}{1+\theta})^i+B_h(i+1)(\frac{\theta}{1+\theta})^2(\frac{1}{1+\theta})^i \\ \qquad +C_h\frac{1+\theta}{1+\theta+\alpha^h}(\frac{\alpha^h}{1+\theta+\alpha^h})^i\Big), & i=1,2,\ldots,\end{cases} \tag{2.5}$$

and

$$A_h=\frac{\theta^2(1-\alpha^h)^2+\theta(1-\alpha^{2h})+2\alpha^h}{(\theta(1-\alpha^h)+1)^2},$$
$$B_h=\frac{1-\alpha^h}{\theta(1-\alpha^h)+1},$$
$$C_h=\frac{-\alpha^h}{(\theta(1-\alpha^h)+1)^2}.$$

Proof. Suppose (2.4) is the PGF of W_{t+h}, which can be rewritten as

$$\begin{aligned}\Phi_{W_{n+h}}(s) &= \frac{2+\theta-s}{(1+\theta-s)^2}\frac{(\theta+\alpha^h(1-s))^2}{1+\theta+\alpha^h(1-s)}\\ &= \alpha^h+(1-\alpha^h)\Big(\frac{s^2(\theta\alpha^h-\alpha^h)+s(-\theta^2(1+\alpha^h)-2\theta\alpha^h+2\alpha^h)+\theta^3+\theta^2(\alpha^h+2)}{(1+\theta-s)^2(1+\theta+\alpha^h(1-s))}\\ &\quad+\frac{\theta\alpha^h-\alpha^h}{(1+\theta-s)^2(1+\theta+\alpha^h(1-s))}\Big)\\ &= \alpha^h+(1-\alpha^h)\Big(A_h\frac{\theta}{1+\theta-s}+B_h\frac{\theta^2}{(1+\theta-s)^2}+C_h\frac{(1+\theta)}{1+\theta+\alpha^h(1-s)}\Big),\end{aligned} \tag{2.6}$$

where

$$A_h=\frac{\theta^2(1-\alpha^h)^2+\theta(1-\alpha^{2h})+2\alpha^h}{(\theta(1-\alpha^h)+1)^2},$$
$$B_h=\frac{1-\alpha^h}{\theta(1-\alpha^h)+1},$$
$$C_h=\frac{-\alpha^h}{(\theta(1-\alpha^h)+1)^2}.$$

By (2.6), the PMF of W_{t+h} is a mixture of the geometric distribution with parameter $\frac{\theta}{1+\theta}$, negative binomial distribution with parameters $(2,\frac{\theta}{1+\theta})$, geometric distribution with parameter $\frac{1+\theta}{1+\theta+\alpha^h}$ and a degenerate distribution at zero, respectively. Then we can obtain the PMF of W_{n+h} in (2.5) with $\sum_{i=0}^{\infty}P(W_{t+h}=i)=1$. This completes the proof. □

3 Quasi-likelihood estimation

In this section, we will consider the quasi-likelihood (QL; Heyde,1997) estimators for the parameters α and θ, respectively. Note that $\mu_\epsilon = (1-\alpha)(\theta+2)/\{\theta(\theta+1)\}$, we need to estimate the parameter of interest with $\beta = (\alpha, \mu_\epsilon)'$. Assume that $\{X_1, \ldots, X_n\}$ is a sample from model (2.2). Let $\gamma = (\lambda, \sigma_\epsilon^2)'$, where $\lambda = \alpha(1-\alpha)$. Recall from Remark 2.2, the conditional mean and one-step conditional variance of X_t are $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu_\epsilon$ and $V_\gamma(X_t|X_{t-1}) \triangleq Var(X_t|X_{t-1}) = \lambda X_{t-1} + \sigma_\epsilon^2$, respectively. Given γ, a set of standard quasi-likelihood estimating equations has the form

$$\begin{aligned} \sum_{t=2}^{n} V_\gamma^{-1}(X_t|X_{t-1}) X_{t-1}(X_t - \alpha X_{t-1} - \mu_\epsilon) = 0, \\ \sum_{t=2}^{n} V_\gamma^{-1}(X_t|X_{t-1})(X_t - \alpha X_{t-1} - \mu_\epsilon) = 0. \end{aligned} \tag{3.1}$$

By solving (3.1), we can get the quasi-likelihood estimator of β with an explicit solution,

$$\hat{\beta} = Q_n^{-1} q_n, \tag{3.2}$$

where

$$Q_n = \begin{pmatrix} \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) X_{t-1}^2 & \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) X_{t-1} \\ \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) X_{t-1} & \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) \end{pmatrix},$$

$$q_n = \begin{pmatrix} \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) X_t X_{t-1} \\ \sum_{t=2}^n V_\gamma^{-1}(X_t|X_{t-1}) X_t \end{pmatrix}.$$

The following theorem establishes the asymptotic property of $\hat{\beta}$. The proof is similar to Theorem 3.2 in Zheng et al. (2007), so we omit the details here.

Theorem 3.1 *For the quasi-likelihood estimator $\hat{\beta}$ given by (3.2), as $n \to \infty$, we have*

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, T^{-1}(\gamma)),$$

where

$$T(\gamma) = \begin{pmatrix} T_1(\gamma) & T_3(\gamma) \\ T_3(\gamma) & T_2(\gamma) \end{pmatrix}, T_1(\gamma) = E(V_\gamma^{-1}(X_1|X_0) X_0^2),$$
$$T_2(\gamma) = E(V_\gamma^{-1}(X_1|X_0)), T_3(\gamma) = E(V_\gamma^{-1}(X_1|X_0) X_0).$$

In practice, we need to give the consistent estimator of parameter γ. Let

$$S(\gamma) = \sum_{t=2}^{n} \left[(X_t - \hat{\alpha}_{cls} X_{t-1} - \hat{\mu}_{cls}(1 - \hat{\alpha}_{cls}))^2 - (\lambda X_{t-1} + \sigma_\epsilon^2) \right]^2,$$

where $\hat{\alpha}_{cls}$ and $\hat{\mu}_{cls}$ represent the CLS estimators of α and μ, respectively (Mohammadpour et al., 2018). Then we can get an estimator of γ as $\hat{\gamma}_{cls} = \arg\min_{\gamma} S(\gamma)$.

4 Forecasting in the PLINAR(1) process

The point and interval forecasts of future values from model (2.2) are of great interest in practice. Freeland and McCabe (2004) introduced the concept of coherent forecasting in the context of integer-valued time series data. Here the coherent forecasting means that forecasting value is integer. For the h-step point forecast, a simple forecasting method is based on the conditional mean with $E(X_{n+h}|X_n) = \alpha^h X_n + (1-\alpha^h)\mu$, for $h \geq 1$. Although this forecast approach can be easily figured out, the $E(X_{n+h}|X_n)$ is continuous and can not describe the discrete characteristic of count time series. Thus, conditional mean-based predictor is not coherent. To deal with this problem, Maiti et al. (2015) adopted the rounded conditional mean $\hat{X}_{n+h} = \langle \alpha^h X_n + (1-\alpha^h)\mu \rangle$ as the h-step predicted value, where $\langle \cdot \rangle$ is the rounding operator (Kachour and Yao, 2009), α and μ can be consistently estimated via the CLS or QL method.

Another coherent forecasting method is the probabilistic forecasts by estimating the forecasting distribution (Freeland and McCabe, 2004; McCabe, et al., 2011). Specifically, the h-step predicted value $\hat{X}_{n+h}$ is the median of the transition probability function $P(X_{n+h}|X_n)$, which is defined in (2.4). The coherence of the median is given by the fact that it almost lies in the support of the distribution when the variable is discrete and the cardinality of the support is small (Freeland and McCabe, 2004). Moreover, the median is not sensitive to outliers compared with the mean, so the median-based approach is a more robust and coherent point prediction method.

For the h-step interval forecasts, the standard interval in autoregression models is constructed on the basis of asymptotic normality of $\tilde{X}_{n+h} = E(X_{n+h}|X_n)$ (Bhansali, 1974). However, the forecasting distribution $P(X_{n+h}|X_n)$ is positively skewed and unimodal as shown in Figure 2 and Figure 3. So the standard approach is invalid in the aspect of interval forecasts. As suggested by Maiti and Biswas (2015), we can use the $100(1-\gamma)\%$ highest predicted probability (HPP) interval with $\gamma \in (0,1)$. Specially, the $100(1-\gamma)\%$ HPP interval of X_{n+h} given X_n, denoted by $C_h = (X_L, X_U)$, is $C_h = \left\{ i : p_h(i|j) \geq K_\gamma \right\}$ with K_γ is the largest number such that

$$P(X_L \leq X_{n+h} \leq X_U | X_n = j) = \sum_{i=X_L}^{X_U} p_h(i|j) \geq (1-\gamma), \tag{4.1}$$

where $p_h(i|j) = P(X_{n+h} = i | X_n = j)$ is defined in (2.4). Based on the definition of HPP interval and the unimodality of the forecasting distribution, we propose the following algorithm to obtain the HPP interval.

Algorithm 1

Step 1. Set $K_\gamma = \delta$, and $k = 1$, where δ is a small positive number (e.g. δ=0.01)

Step 2. Let X_{L_k} is the smallest integer such that $P(X_{n+h} = X_{L_k}|X_n = j) \geq K_\gamma$, and X_{U_k} is the largest integer such that $P(X_{n+h} = X_{U_k}|X_n = j) \geq K_\gamma$. Calculate $S_k = \sum_{i=X_{L_k}}^{X_{U_k}} p_h(i|j)$.

Step 3. If $S_k \geq (1-\gamma)$, we set $K_\gamma = K_\gamma + \delta$ and $k = k+1$.

Step 4. Repeat Steps 2 and 3 until there exists an integer k_0 such that $S_{k_0} \geq (1-\gamma)$ and $S_{(k_0+1)} < (1-\gamma)$. Then, the HPP interval is $[X_{L_{k_0}}, X_{U_{k_0}}]$.

5 Numerical simulation

In this section, we conduct some simulation studies to evaluate the performance of our proposed methods with the help of R software. We generate a number of observations from the PLINAR(1) model (2.1) with four different sets of parameter combination, (i)$\alpha = 0.1, \theta = 1$, (ii)$\alpha = 0.1, \theta = 1.5$, (iii)$\alpha = 0.3, \theta = 1$ and (iv)$\alpha = 0.3, \theta = 1.5$. In Figure 4, we present some sample paths of the simulated data. All the experiments are repeated $B = 10000$ times, where the sample size n =100, 300, 500, respectively.

To check the efficiency of our proposed quasi-likelihood (QL) estimate, we compare it with the Yule-Walker (YW) and conditional least squares (CLS) methods (Mohammadpour et al., 2018). In Table 1, we report the estimated bias (BIAS) and sample standard error (SSE) of the corresponding estimate using the format (BIAS, SSE). For example, (-0.015442 0.009903) means that BIAS is -0.015442, and SSE is 0.009903. It can be seen from the results that BIAS and SSE of all the three estimates are decreasing as sample size n becoming larger. Overall, the BIAS and SSE of QL are smallest among the three methods, which indicate that the QL method produces a better estimation.

The second simulation aims to compare the rounded conditional mean-based predictor (RCM) with the median of transition probability-based predictor (MTP) towards h-step prediction. Here we set the forward step as $h = 1, ..., 4$. We generate $n+4$ observations from the PLINAR(1) process, where the first n observations are used to estimate the parameters, and the remaining observations are used to calculate the prediction mean absolute error (PMAE). The calculation formula is

$$\text{PMAE}(h) = \frac{1}{B}\sum_{k=1}^{B} |X_{n+h}^{(k)} - \hat{X}_{n+h}^{(k)}|,$$

where $X_{n+h}^{(k)}$ is the $(n+h)$th observed data, $\hat{X}_{n+h}^{(k)}$ is the corresponding h-step predication, and k is the repetition times. We give the results on PMAE in Table 2, which indicate that the MTP procedure is better than RCM method.

Finally, we study the HPP's performance for interval prediction of future observation, where the data are generated as the second simulation. In Table 3, we present the h-step predication interval of X_{n+h} in the form (X_L, X_U), where $X_L = \frac{1}{B}\sum_{i=1}^{B} X_L^{(k)}$ with $X_L^{(k)}$ being the left interval value of kth repetition simulation, $k = 1, \cdots, B$, and X_U is defined similarly. The 95% coverage probability (CP) and length of prediction interval (LPI) are also provided. From the results, we conclude that CP is close to 0.95 and LPI is decreasing as n becoming large, which indicate that the HPP method can produce reliable predication interval for the PLINAR(1) process.

6 Real data analysis

Now we apply our proposed method to a data set of animal health symptom, which records the monthly number of animals anorexia from January 2003 to December 2009 in a region of New Zealand (Mohammadpour, et al., 2018). The sample path, autocorrelation function (ACF) and partial autocorrelation function (PACF) are presented in Figure 1, which shows that AR(1)-type process is appropriate for modeling this data set. Of note this data is over-dispersed with mean and variance equal to 0.8214 and 2.8954, respectively. By Mohammadpour et al. (2018), we can adopt the PLINAR(1) process to fit this anorexia data. To check the performance of our prediction method, we use the first 80 observations to estimate the parameters, and predict the last 4 observations. The QL estimators for α and θ are $\hat{\alpha}_{QL} = 0.1845$ and $\hat{\theta}_{QL} = 1.5315$, respectively. In Table 4, we report the point predictions for the last 4 observations using the rounded conditional mean-based predictor (RCM) and the median of the transition probability-based predictor (MTP). It can be seen that the 95% HPP intervals seem to be acceptable, and the MTP is much better than RCM since MTP-based predictors are exactly the same as observed data. Thus, we suggest to use MTP-based predictor for data prediction in the PLINAR(1) process.

7 Concluding remarks

In this paper, we have considered the quasi-likelihood (QL) estimators for the parameters of interest in PLINAR(1) process. The rounded conditional mean-based predictor and the median of transition probability-based predictor were proposed to forecast future data. Furthermore, highest predicted probability interval was also presented. Simulation studies and real data application indicated that our proposed methods work well.

Acknowledgements

The authors would like to thank the Editor, the Associate Editor and the reviewer for their constructive and insightful comments and suggestions that greatly improved the manuscript. This work is supported by Science Foundation of Tianjin University (No. 2018XRG-0038).

References

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

[2] Awale, M., Ramanathan, T. and Kale, M. (2017). Coherent forecasting in integer-valued AR(1) models with geometric marginals. *Journal of Data Science*, **15**, 95-114.

[3] Bakouch, H. and Ristić, M. (2010). Zero truncated Poisson integer-valued AR(1) model. *Metrika*, **72**, 265-280.

[4] Bhansali, R. (1974). Asymptotic mean-square error of predicting more than one-step ahead using the regression method. *Journal of the Royal Statistical Society: Series C*, **23**, 35-42.

[5] Brännäs, K. and Quoreshi, A. (2010). Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics*, **20**, 1429-1440.

[6] Bu, R., McCabe, B. and Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis*, **29**, 973-994.

[7] Drost, F. , Akker, R. and Werker, B. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semi-parametric integer-valued AR(p) models. *Journal of the Royal Statistical Society: Series B*, **71**, 467-485.

[8] Du, J. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, **12**, 129-142.

[9] Freeland, R. and McCabe, B. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, **20**, 427-434.

[10] Ghitany, M. and Al-Mutairi, D. (2009). Estimation methods for the discrete Poisson-Lindley distribution. *Journal of Statistical Computation and Simulation*, **79**, 1-9.

[11] Heyde, C. (1997). Quasi-likelihood and its application: a general approach to optimal parameter estimation. Academic Press: New York.

[12] Kachour, M. and Yao, J. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis*, **30**, 417-448.

[13] Livsey, J., Lund, R., Kechagias, S. and Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, **12**, 408-431.

[14] Maiti, R. and Biswas, A. (2015). Coherent forecasting for over-dispersed time series of count data. *Brazilian Journal of Probability and Statistics*, **29**, 747-766.

[15] Maiti, R., Biswas, A. and Das, S. (2016). Coherent forecasting for count time series using Box-Jenkins's AR(p) model. *Statistica Neerlandica*, **70**, 123-145.

[16] Martin, V., Tremayne, A. and Jung, R. (2014). Efficient method of moments estimators for integer time series models. *Journal of Time Series Analysis*, **35**, 491-516.

[17] McCabe, B., Martin, G. and Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society: Series B*, **73**, 253-272.

[18] Mohammadpour, M., Bakouch, H. and Shirozhan, M. (2018). Poisson-Lindley INAR (1) model with applications. *Brazilian Journal of Probability and Statistics*, **32**, 262-280.

[19] Moriña, D., Puig, P., Ríos, J., Vilella, A. and Trilla, A. (2011). A statistical model for hospital admissions caused by seasonal diseases. *Statistics in Medicine*, **30**, 3125-3136.

[20] Nastić, A., Laketa, P. and Ristić, M. (2016). Random environment integer-valued autoregressive process. *Journal of Time Series Analysis*, **37**, 267-287.

[21] Pedeli, X., Davison, A. and Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, **110**, 1229-1238.

[22] Ristić, M., Bakouch, H. and Nastić, A. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, **139**, 2218-2226.

[23] Sankaran, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics*, **26**, 145-149.

[24] Scotto, M., Weiss, C. and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, **15**, 590-618.

[25] Steutel, F. and Van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, **7**, 893-899.

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

[27] Zhang, H., Wang, D. and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, **140**, 667-683.

[28] Zhang, H., Wang, D. and Zhu, F. (2011). Empirical likelihood inference for random coefficient INAR(p) process. *Journal of Time Series Analysis*, **32**, 195-203.

[29] Zhang, H., Wang, D. and Sun, L. (2017). Regularized estimation in GINAR(p) process. *Journal of the Korean Statistical Society*, **46**, 502-517.

[30] Zheng, H., Basawa, I. and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis*, **27**, 411-440.

[31] Zheng, H., Basawa, I. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, **137**, 212-229.

Table 1. Bias and SE of the estimators for the PLINAR(1) process.

Sample size	Parameters	YW	CLS	QL
$\alpha = 0.1, \theta = 1$				
$n = 100$	$\hat{\alpha}$	(−0.015442 0.009903)	(−0.014562 0.010006)	(−0.014187 0.009956)
	$\hat{\theta}$	(0.014088 0.011799)	(0.014265 0.011877)	(−0.003855 0.003421)
$n = 300$	$\hat{\alpha}$	(−0.004122 0.003424)	(−0.003790 0.003436)	(−0.002025 0.003137)
	$\hat{\theta}$	(0.004880 0.003833)	(0.004863 0.003843)	(0.004863 0.003843)
$n = 500$	$\hat{\alpha}$	(−0.001274 0.001968)	(−0.001073 0.001972)	(−0.001105 0.001966)
	$\hat{\theta}$	(0.002418 0.002291)	(0.002387 0.002293)	(0.002387 0.002293)
$\alpha = 0.1, \theta = 1.2$				
$n = 100$	$\hat{\alpha}$	(−0.016144 0.009994)	(−0.015313 0.010093)	(−0.014976 0.010025)
	$\hat{\theta}$	(0.022236 0.015362)	(0.022450 0.015477)	(0.022448 0.015477)
$n = 300$	$\hat{\alpha}$	(−0.005214 0.003425)	(−0.004899 0.003436)	(−0.004839 0.003419)
	$\hat{\theta}$	(0.006646 0.004892)	(0.006576 0.004900)	(0.006576 0.004900)
$n = 500$	$\hat{\alpha}$	(−0.001642 0.001982)	(−0.001447 0.001986)	(−0.001473 0.001981)
	$\hat{\theta}$	(0.003396 0.002933)	(0.003396 0.002938)	(0.003395 0.002938)
$\alpha = 0.3, \theta = 1$				
$n = 100$	$\hat{\alpha}$	(−0.022655 0.010198)	(−0.019809 0.010287)	(−0.019166 0.010297)
	$\hat{\theta}$	(0.024098 0.014764)	(0.024324 0.014910)	(0.024320 0.014910)
$n = 300$	$\hat{\alpha}$	(−0.007440 0.003461)	(−0.006448 0.003471)	(−0.006235 0.003451)
	$\hat{\theta}$	(0.005123 0.004762)	(0.005113 0.004774)	(0.005113 0.004774)
$n = 500$	$\hat{\alpha}$	(−0.004120 0.002070)	(−0.003523 0.002072)	(−0.003385 0.002051)
	$\hat{\theta}$	(0.002730 0.002810)	(0.002689 0.002817)	(0.002689 0.002817)
$\alpha = 0.3, \theta = 1.2$				
$n = 100$	$\hat{\alpha}$	(−0.023430 0.010167)	(−0.020590 0.010254)	(−0.019680 0.0102830)
	$\hat{\theta}$	(0.027463 0.018812)	(0.028283 0.018985)	(0.028262 0.018983)
$n = 300$	$\hat{\alpha}$	(−0.007831 0.003513)	(−0.006866 0.003522)	(−0.006764 0.003477)
	$\hat{\theta}$	(0.008256 0.006078)	(0.008347 0.006097)	(0.008344 0.006097)
$n = 500$	$\hat{\alpha}$	(−0.005575 0.002121)	(−0.004948 0.002124)	(−0.004787 0.002096)
	$\hat{\theta}$	(0.005804 0.003668)	(0.005790 0.003677)	(0.005791 0.003677)

Table 2. Values of PMAE for simulated PLINAR(1) process.‡

h	RCM	MTP	RCM	MTP	RCM	MTP	RCM	MTP
$n = 100$	$\alpha = 0.1, \theta = 1$		$\alpha = 0.1, \theta = 1.2$		$\alpha = 0.3, \theta = 1$		$\alpha = 0.3, \theta = 1.2$	
1	1.3507	1.2761	1.0959	1.0842	1.2541	1.1835	1.0719	0.9748
2	1.4088	1.3025	1.1044	1.0924	1.3481	1.2593	1.1085	1.0843
3	1.3816	1.2497	1.1088	1.0996	1.3697	1.2533	1.1166	1.0862
4	1.3732	1.2399	1.1070	1.0929	1.3845	1.2534	1.1215	1.0951
$n = 300$	$\alpha = 0.1, \theta = 1$		$\alpha = 0.1, \theta = 1.2$		$\alpha = 0.3, \theta = 1$		$\alpha = 0.3, \theta = 1.2$	
1	1.3157	1.2367	1.0734	1.0573	1.2435	1.1783	1.0632	0.9648
2	1.3885	1.2643	1.0848	1.0830	1.3397	1.2537	1.0955	1.0873
3	1.3752	1.2528	1.1003	1.0986	1.3664	1.2577	1.0778	1.0735
4	1.3610	1.2382	1.0721	1.0713	1.3732	1.2423	1.0798	1.0775
$n = 500$	$\alpha = 0.1, \theta = 1$		$\alpha = 0.1, \theta = 1.2$		$\alpha = 0.3, \theta = 1$		$\alpha = 0.3, \theta = 1.2$	
1	1.3023	1.2135	1.0614	1.0608	1.2432	1.1636	1.0628	0.9607
2	1.3702	1.2523	1.0727	1.0726	1.3279	1.2445	1.0874	1.0804
3	1.3640	1.2412	1.0761	1.0756	1.3474	1.2454	1.0777	1.0749
4	1.3789	1.2580	1.0644	1.0643	1.3678	1.2321	1.0740	1.0729

‡ RCM denotes rounded conditional mean-based predictor; MTP denotes median of transition probability-based predictor.

Table 3. 95% HPP intervals for the PLINAR(1) simulated data.‡

h-step	$(X_L,\, X_U)$	CP(%)	LPI	$(X_L,\, X_U)$	CP(%)	LPI
$n = 100$	$\alpha = 0.1, \theta = 1$			$\alpha = 0.1, \theta = 1.2$		
1	[0.0034, 5.0545)	96.05	5.0511	[0.0016, 4.2676)	96.10	4.2660
2	[0.0001, 5.0855)	95.95	5.0854	[0, 4.2951)	96.15	4.2951
3	[0, 5.0889)	96.25	5.0889	[0, 4.2960)	96.10	4.2960
4	[0, 5.0892)	95.60	5.0892	[0, 4.2963)	96.37	4.2963
	$\alpha = 0.3, \theta = 1$			$\alpha = 0.3, \theta = 1.2$		
1	[0.0248, 4.9299)	95.86	4.9051	[0.0141, 4.1468)	96.25	4.1327
2	[0.0012, 5.0669)	96.04	5.0657	[0.0008, 4.2677)	96.10	4.2669
3	[0.0001, 5.0795)	95.99	5.0794	[0.0002, 4.2841)	95.79	4.2839
4	[0, 5.0800)	96.15	5.0800	[0.0001, 4.2868)	96.49	4.2867
$n = 300$	$\alpha = 0.1, \theta = 1$			$\alpha = 0.1, \theta = 1.2$		
1	[0.0004, 5.0608)	95.92	5.0604	[0, 4.2245)	96.21	4.2245
2	[0, 5.0648)	96.33	5.0648	[0, 4.2257)	95.80	4.2257
3	[0, 5.0652)	96.07	5.0652	[0, 4.2267)	96.24	4.2267
4	[0, 5.0654)	96.18	5.0654	[0, 4.2271)	96.19	4.2271
	$\alpha = 0.3, \theta = 1$			$\alpha = 0.3, \theta = 1.2$		
1	[0.0210, 4.9162)	96.17	4.8952	[0.0083, 4.1556)	96.17	4.1473
2	[0.0001, 5.0587)	96.14	5.0586	[0, 4.2470)	96.14	4.2470
3	[0, 5.0658)	95.98	5.0658	[0, 4.2564)	96.05	4.2564
4	[0, 5.0662)	95.52	5.0662	[0, 4.2572)	96.17	4.2572
$n = 500$	$\alpha = 0.1, \theta = 1$			$\alpha = 0.1, \theta = 1.2$		
1	[0, 4.1674)	96.15	4.1674	[0, 4.1620)	95.82	4.1620
2	[0, 4.1680)	96.13	4.1680	[0, 4.1630)	95.97	4.1630
3	[0, 4.1717)	95.73	4.1717	[0, 4.1662)	95.80	4.1662
4	[0, 4.2034)	96.21	4.2034	[0, 4.1986)	96.09	4.1986
	$\alpha = 0.3, \theta = 1$			$\alpha = 0.3, \theta = 1.2$		
1	[0.0205, 4.9069)	95.87	4.8864	[0.0101, 4.1715)	96.29	4.1614
2	[0.0001, 5.0714)	96.18	5.0713	[0, 4.2145)	96.18	4.2145
3	[0, 5.0599)	96.06	5.0599	[0, 4.2058)	96.05	4.2058
4	[0, 5.0584)	96.47	5.0584	[0, 4.2057)	96.05	4.2057

‡ CP denotes coverage probability; LPI denotes length of prediction interval.

Table 4. Predictive analysis of Anorexia data.‡

Observed value	0	0	0	0
RCM	1	1	1	1
MTP	0	0	0	0
Lower limit	0	0	0	0
Upper limit	3	3	3	3

‡ RCM denotes rounded conditional mean-based predictor; MTP denotes median of transition probability-based predictor.

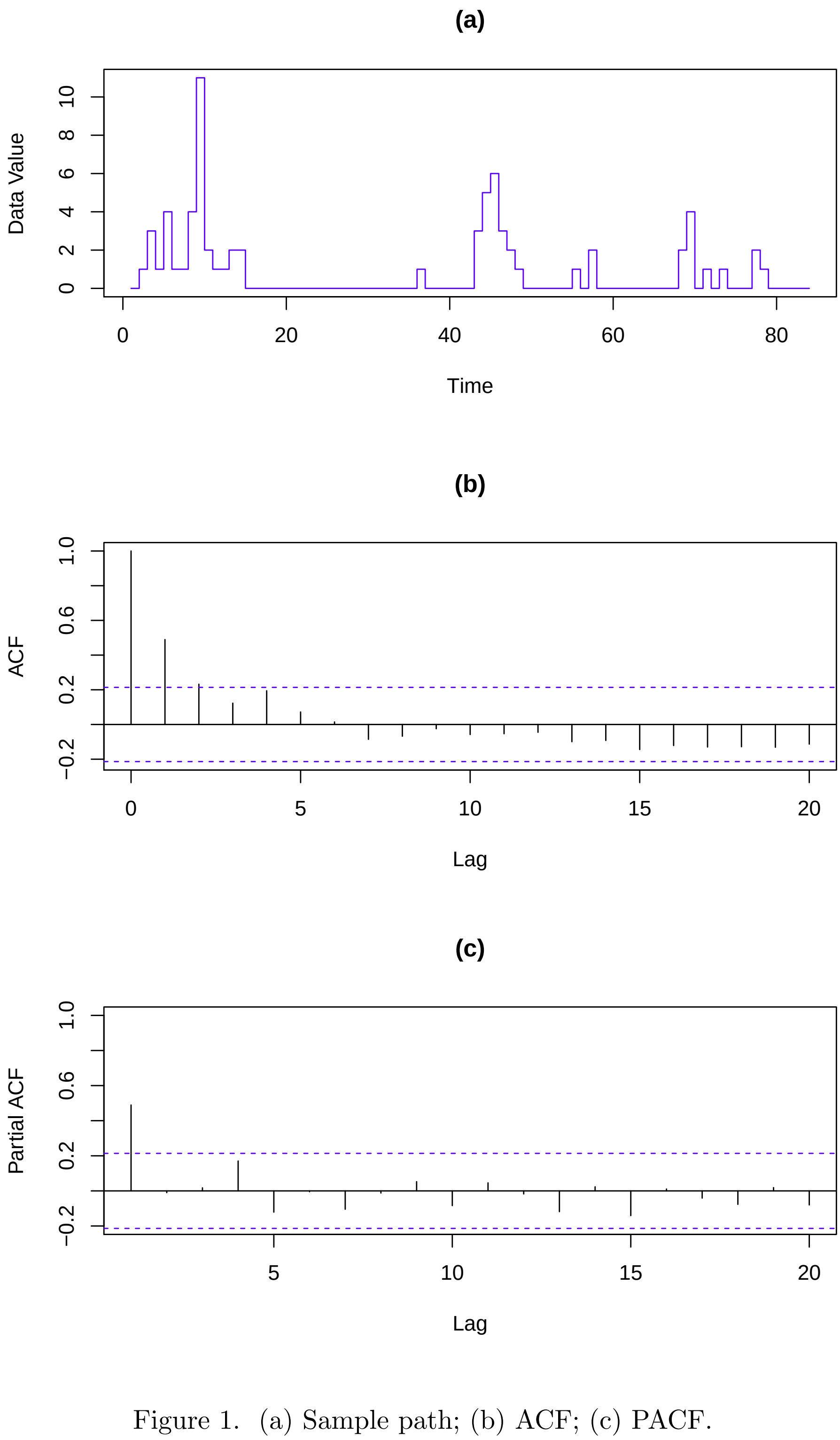

Figure 1. (a) Sample path; (b) ACF; (c) PACF.

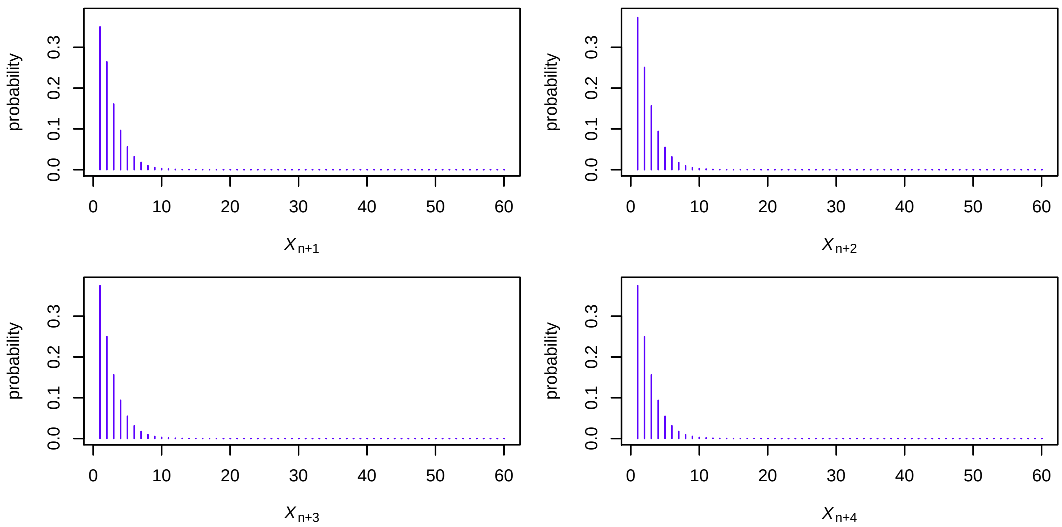

Figure 2. Forecasting distributions for model (2.2) with $\alpha = 0.3, \theta = 1$, conditional on $X_n = 2$.

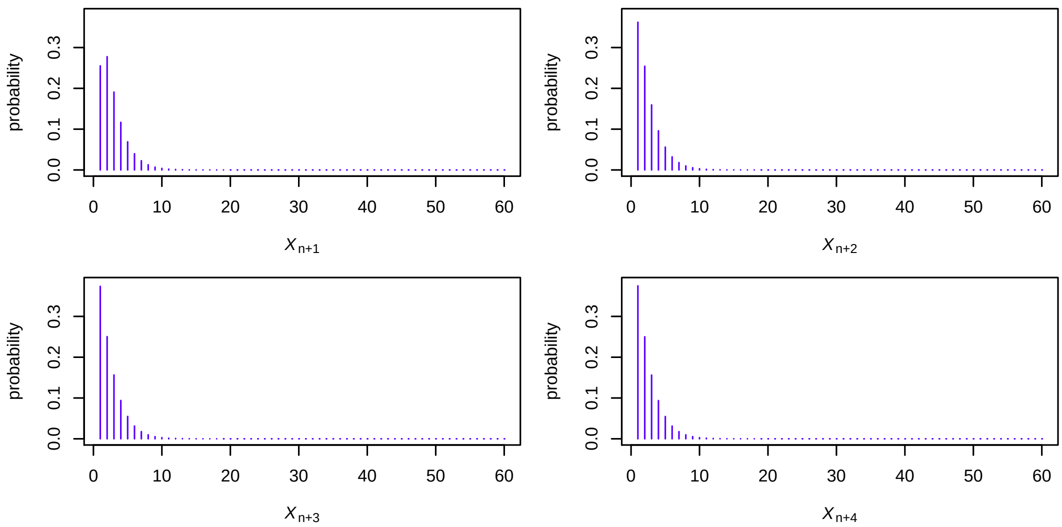

Figure 3. Forecasting distributions for model (2.2) with $\alpha = 0.3, \theta = 1$, conditional on $X_n = 5$.

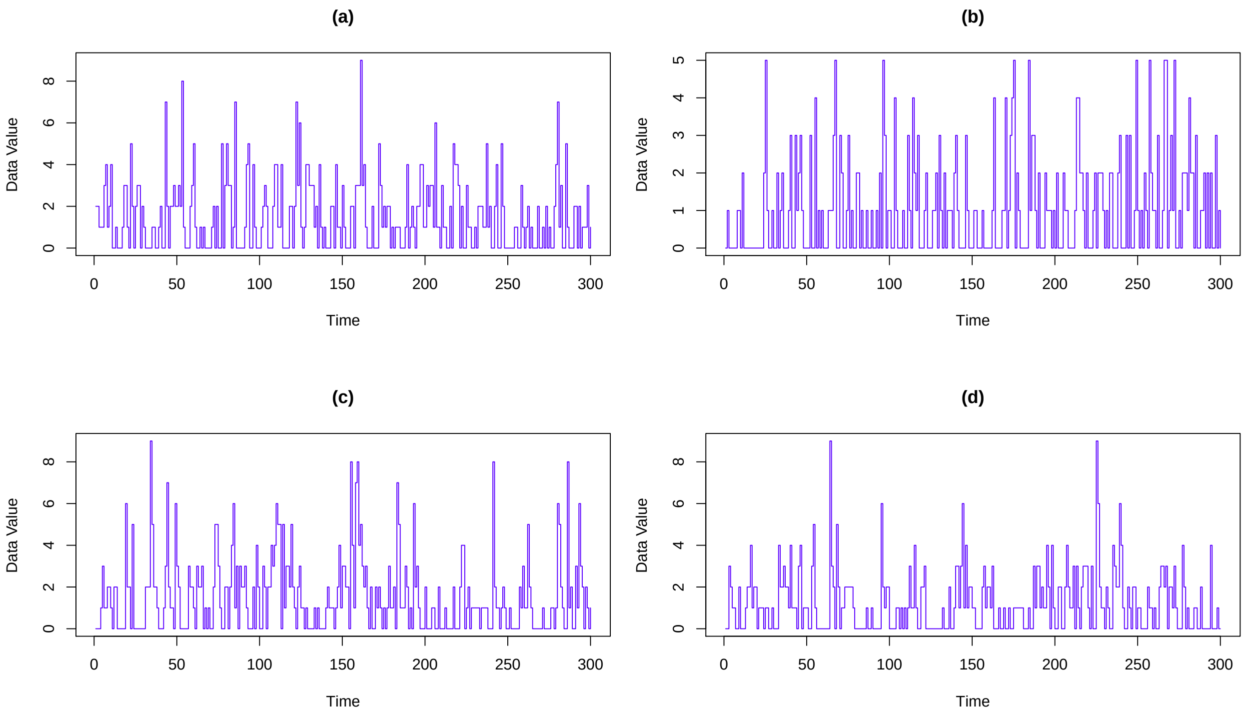

Figure 4. Sample paths of model (2.2) with sample size $n = 300$:
(a) $\alpha = 0.1, \theta$=1; (b)$\alpha = 0.1, \theta$=1.2; (c) $\alpha = 0.3, \theta$=1; (d)$\alpha = 0.3, \theta$=1.2.