Forecasting in INAR(1) model

Nélia Silva[1,3], Isabel Pereira[1,3], M. Eduarda Silva[2,3]

(nelia@mat.ua.pt, isabelp@mat.ua.pt, mesilva@fc.up.pt)

[1] Departamento de Matemática, Universidade de Aveiro, Portugal

[2] Departamento de Matemática Aplicada, Faculdade de Ciências, Universidade do Porto, Portugal

[3] UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal

Abstract

In this work we consider the problem of forecasting integer-valued time series, modelled by the INAR(1) process introduced by McKenzie (1988) and Al-Osh and Alzaid (1987). The theoretical properties and practical applications of INAR and related processes have been discussed extensively in the literature but there is still some discussion on the problem of producing coherent, ie, integer predictions. Here Bayesian methodology is used to obtain point predictions as well as confidence intervals for future values of the process. The predictions thus obtained are compared with their classic counterparts. The proposed approaches are illustrated with a simulation study and a real example.

Keywords: INAR models, Bayesian prediction, integer prediction, Markov Chain Monte Carlo algorithm

1 Introduction

In applications we are frequently faced with time series whose characteristics are not compatible with a continuous modelling approach. Discrete variate time series occur in many contexts, often as counts of events or individuals in consecutive intervals or at consecutive points in time. Examples of these are the number of costumers waiting to be served, the daily number of absent workers in a firm, the number of busy lines in a telephone network noted every hour, the number of accidents in a manufacturing plant each month, etc. Several models that take the discreteness of the data explicitly into account have been developed in the literature. Cox (1981) proposed dividing them into two categories: observation-driven and parameter-driven models. MacDonald and Zucchini (1997), Cameron and Trivedi (1998) and the review by McKenzie (2003) provide an excellent overview of the literature in this area.

In this work we are interested in a special class of observation-driven models, the so-called integer-valued autoregressive (INAR) process introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). The theoretical properties and practical applications of INAR and related processes have been discussed extensively in the literature. Silva *et al.* (2005) consider independent replications of count time series modelled by INAR(1) and proposed several estimation methods using the classical and Bayesian approaches in time and frequency domains. Nevertheless, there is still little consensus on which processes or model classes are best used in practice in contrast to the role played by the Box-Jenkins Gaussian ARMA methodology for continuous variables. This is partly due to the lack of reliable techniques for estimation, testing and prediction. In particular, the lack of forecasting methods that are coherent in the sense of producing only integer predictions, seems to render useless the effort of using more complex models.

Usually the forecast values are obtained from the conditional expectations, which have the optimality property but rarely will generate integer values. In order to produce coherent forecasts Freeland and McCabe (2004) use the median of the k-step-ahead conditional distribution to emphasize the intention of preserving the integer structure of the data in generating the forecasts. McCabe and Martin (2005) develop a general methodology for producing coherent predictions of low count data. In contrast to the usual applications of the model INAR(1), in which the arrival process is usually Poisson, they allow the arrivals to follow any distribution in the integer class. The forecasts are based on an estimate of the k-step–ahead predictive probability mass function. To eliminate unwanted values from the conditioning set of the predictive function, Bayesian methods are used. Jung and Tremayne (2006) extend some of the ideas used by Freeland and McCabe in higher order dependence structure by proposing a computer intensive method for generating coherent, integer out-of-sample predictions, particularly obtaining the h-step-ahead predictor for the INAR(2). They use a Monte Carlo approach using estimated sampling distributions from the bootstrap methodology as a means of generating one and multi-step ahead forecasts which respects the integer structure of the data.

The purpose of this paper is to obtain coherent forecasts for the Poisson INAR(1) process. Bayesian methodology is used to obtain point predictions as well as confidence intervals for future values of the process which are compared with their classic counterpart.

The remainder of the paper is divided into four main sections. Section 2 provides the theoretical results in order to obtain the point forecasts. Section 3 presents methods for producing confidence intervals or highest posterior predictive density intervals for forecasts. In Section 4 we conduct a simulation study to compare the performance of the classical and Bayesian approaches, considering point and interval predictions. Section 5 gives an example of forecasting a count data series using PoINAR(1) model. The data are the number of claimants receiving wage loss benefits due to injuries from burns, supplied by the Workers Compensation Board of the Province of British Columbia, Canada. The proposed methodology presented in this work is applied to this data set and compared with classical inference and forecast procedures made by Freeland (1998).

2 Point Prediction

Consider a non negative integer-valued random variable X and $\alpha \in [0, 1]$, the generalized thinning operation, hereafter denoted by '$\circ$', is defined as

$$\alpha \circ X = \sum_{j=1}^{X} Y_j, \tag{1}$$

where $\{Y_j\}, \quad j = 1, \dots, X$, is a sequence of independent and identically distributed non-negative integer-valued random variables, independent of X, with finite mean α and variance σ^2. This sequence is called the counting series of $\alpha \circ X$. When $\{Y_j\}$ is a sequence of Bernoulli random variables, the

thinning operation is called binomial thinning operation and was defined by Steutel and van Harn (1979).
The well-known INAR(1) process $\{X_t; t = 0, \pm 1, \pm 2, \ldots\}$ is defined on the discrete support $\mathbb{N}_0$ by the equation

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

where $0 < \alpha < 1$, $\{\epsilon_t\}$ is a sequence of independent and identically distributed integer-valued random variables, with $E[e_t] = \mu_e$ and $Var[e_t] = \sigma_e^2$.
In this paper we consider only Poisson INAR(1) process, i.e., $\{\epsilon_t\}$ is a sequence of independent Poisson distributed variables with parameter λ, independent of all counting series $\{Y_j\}$. Note that, assuming $\epsilon_t \frown Po(\lambda)$ it is straightforward to show that $X_t \frown Po(\lambda/(1-\alpha))$. The Poisson INAR(1) process will henceforth be denoted PoINAR(1), and is a natural first candidate for modelling the data partly because its marginal distribution appears to be equidispersed.
Given that we have observed the series up through time n, i.e., $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$ is known, the most common procedure for constructing predictions in time series models is to use conditional expectations. In order to find predictions for future values, the theorem below is very important,

Theorem 1 (Freeland, 1998, pp. 30) *The moment generation function of X_{n+h} given X_n is*

$$\varphi_{X_{n+h}|X_n}(s) = \left[\alpha^h e^s + \left(1 - \alpha^h\right)\right]^{X_n} \exp\left[\lambda \frac{1-\alpha^h}{1-\alpha}(e^s - 1)\right]. \tag{2}$$

Expression (2) shows that the distribution of $X_{n+h}|X_n$ is a convolution of a binomial distribution with parameters α^h and X_n and a Poisson distribution with parameter $\lambda(1-\alpha^h)/(1-\alpha)$. That is, the probability function of $X_{n+h}|X_n$ is given by

$$\begin{aligned} f(x_{n+h}|x_n) &= P(X_{n+h} = x_{n+h}|X_n = x_n) \\ &= \exp\left\{-\lambda \tfrac{1-\alpha^h}{1-\alpha}\right\} \textstyle\sum_{i=0}^{M_h} \frac{1}{(x_{n+h}-i)!} \times \\ &\times \left(\lambda \tfrac{1-\alpha^h}{1-\alpha}\right)^{x_{n+h}-i} \tbinom{x_n}{i} (\alpha^h)^i (1-\alpha^h)^{x_n - i}, \;\; x_{n+h} = 0, 1, \ldots, \end{aligned} \tag{3}$$

where $M_h = \min(X_{n+h}, X_n)$.
Consequently, we have the following corollary,

Corollary 1 *The INAR(1) model satisfies the properties*

a) $E[X_{n+h}|X_n] = \alpha^h \left[X_n - \frac{\lambda}{1-\alpha}\right] + \frac{\lambda}{1-\alpha}$, $\quad h = 1, 2, 3, \ldots,$

b) $Var[X_{n+h}|X_n] = \alpha^h(1-\alpha^h)X_n + \lambda \frac{1-\alpha^h}{1-\alpha}$, $\quad h = 1, 2, 3, \ldots,$

c) *As $h \longrightarrow +\infty$, $X_{n+h}|X_n$ is a Poisson distribution with parameter $\lambda/(1-\alpha)$.*

So, we can conclude that for α constant, $\lim_{h\to+\infty} E[X_{n+h}|X_n] = \lim_{h\to+\infty} Var[X_{n+h}|X_n] = \lambda/(1-\alpha)$, i.e., as $h \longrightarrow \infty$ and $0 < \alpha < 1$, the mean and the variance of $X_{n+h}|X_n$ remain equal and approach the mean of the process.

2.1 Classical Methodology

The h-step-ahead predictor based on the conditional expectation of INAR(1),

$$\hat{X}_{n+h}|\mathbf{x}_n = E[X_{n+h}|X_n] = \alpha^h \left[X_n - \frac{\lambda}{1-\alpha} \right] + \frac{\lambda}{1-\alpha}, \; h = 1, 2, 3, \ldots \tag{4}$$

was obtained by Brännäs (1994) and Freeland and McCabe (2003), but it will hardly produce integer valued forecasts. In order to obtain coherent predictions for X_{n+h} Freeland and McCabe (2003) suggest using the value which minimizes the expected absolute error given the sample, i.e., the value that minimizes $E[|X_{n+h} - \hat{X}_{n+h}| \Big| X_n]$. So, they concluded that $\hat{X}_{n+h} = \hat{m}_{n+h}$ is the median of the h-step-ahead conditional distribution $f(x_{n+h}|x_n)$.

2.2 Bayesian Methodology

The Bayesian predictive probability function is very simple to understand, it is based on, both, the future observation, X_{n+h} and the vector of unknown parameters $\boldsymbol{\theta} = (\alpha, \lambda)$ to be random. As we know the information about $\boldsymbol{\theta}$ is given by the observed sample $\mathbf{x}_n$ and quantified in the posterior predictive, $\pi(\boldsymbol{\theta}|\mathbf{x}_n)$.

Definition 1 *Let $\boldsymbol{\theta} \in \Theta$ be the vector of unknown parameters. The h-step-ahead Bayesian posterior predictive distribution is given by*

$$\begin{aligned} f(x_{n+h}|\mathbf{x}_n) &= \int_\Theta f(x_{n+h}; \boldsymbol{\theta}|\mathbf{x}_n) d\boldsymbol{\theta} \\ &= \int_\Theta f(x_{n+h}|\mathbf{x}_n; \boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{x}_n) d\boldsymbol{\theta}, \end{aligned} \tag{5}$$

where $\pi(\boldsymbol{\theta}|\mathbf{x}_n)$ is the posterior probability function of $\boldsymbol{\theta}$ and $f(x_{n+h}|\mathbf{x}_n; \boldsymbol{\theta})$ is the predictive distribution (classical) given by (3).

The h-step-ahead predictive distribution of $X_{n+h}|\mathbf{x}_n$ given by expression (5) can be viewed as having all information about the future values. Once $f(x_{n+h}|\mathbf{x}_n)$ is obtained, the Bayesian h-step-ahead predictor can be given by the expected valued, the median or the mode of X_{n+h} given $\mathbf{x}_n$.
Since beta and gamma are conjugate of binomial and Poisson distributions, respectively, we use them for prior distributions of the parameters to INAR(1) model, $\alpha \frown Beta(a, b)$, $a, b > 0$ and $\lambda \frown Gamma(c, d)$, $c, d > 0$. Considering independence between α and λ, the prior distribution of (α, λ) is proportional to

$$p(\alpha, \lambda) \propto \lambda^{c-1} \exp(-d\lambda) \alpha^{a-1} (1-\alpha)^{b-1}, \quad \lambda > 0, 0 < \alpha < 1, \tag{6}$$

where a, b, c and d are known parameters. Note that, as $a \to 0$, $b \to 0$, $c \to 0$ and $d \to 0$ we have a vague prior distribution.
The posterior distribution of (α, λ) can be written as

$$\begin{aligned} p(\alpha, \lambda|\mathbf{x}_n) \propto \;& L(\mathbf{x}_n, \alpha, \lambda|x_1)\, p(\lambda, \alpha) \\ = \;& \exp\left[-(d + (n-1))\lambda\right] \lambda^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \\ & \prod_{t=2}^{n} \sum_{i=0}^{M_t} \frac{\lambda^{x_t - i}}{(x_t - i)!} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i}, \end{aligned}$$

where $L(\mathbf{x}_n|x_1)$ is the conditional likelihood function and $M_t = \min(X_t, X_{t-1})$.
Consequently for the PoINAR(1) model, the Bayesian predictive function of X_{n+h} given $\mathbf{x}_n$ is given by,

$$
\begin{aligned}
f(x_{n+h}|\mathbf{x}_n) \propto \quad & \int_\alpha \int_\lambda \sum_{i=0}^{M_h} \binom{x_n}{i} (\alpha^h)^i (1-\alpha^h)^{x_n-i} \frac{1}{(x_{n+h}-i)!} \times \\
& \times \exp\left(-\lambda \tfrac{1-\alpha^h}{1-\alpha}\right) \left(\lambda \tfrac{1-\alpha^h}{1-\alpha}\right)^{x_{n+h}-i} \exp\left[-(d+n)\lambda\right] \lambda^{c-1} \qquad (7) \\
& \times \alpha^{a-1}(1-\alpha)^{b-1} \prod_{t=2}^{n} \sum_{i=0}^{M_t} \frac{\lambda^{x_t-i}}{(x_t-i)!} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} d\alpha d\lambda \, .
\end{aligned}
$$

The complexity of $f(x_{n+h}|\mathbf{x}_n)$ does not allow us to work with it directly. In order to estimate X_{n+h}, we can adapt to the integer case the Tanner composition method. That is, to sample $(X_{n+h,1}, X_{n+h,2}, \ldots, X_{n+h,m})$, we can use the following algorithm:

Algorithm 1 *1. from the sample $(X_1, X_2, \ldots, X_n)$, calculate (through the classical method) a starting estimative to α, let α_0;*

2. using the adaptive rejection Metropolis sampling (ARMS) within Gibbs methodology, calculate from the full conditional distributions of parameters α and λ, a sample $(\alpha_1, \lambda_1), (\alpha_2, \lambda_2), \ldots, (\alpha_m, \lambda_m)$;

3. For each i $(i = 1, \ldots, m)$ sample $X_{n+h,i}$ from $f(x_{n+h}|x_n, \alpha_i, \lambda_i)$, using the inverse transform method adapted to integer variables, that is,

(a) sample u from uniform $U(0,1)$,

(b) calculate the least integer valued s: $\sum_{i=0}^{s} f(x_{n+h}|x_n, \alpha_i, \lambda_i) \geq u$,

(c) consider $X_{n+h,i} = s$.

After sampling $X_{n+h,1}, X_{n+h,2}, \ldots, X_{n+h,m}$, the h-step-ahead predictor $\hat{X}_{n+h}$, can be calculated from sample mean $(\tilde{X}_{n+h})$, median $(\tilde{m}_{n+h})$ or mode $(\tilde{mo}_{n+h})$.
But we can also calculate $E(X_{n+h}|\mathbf{x}_n)$ using an appropriate property of mathematical expectation. As we know $E[g(X_{n+h})|\mathbf{x}_n] = E[E[g(X_{n+h})|\mathbf{x}_n, \boldsymbol{\theta}]|\mathbf{x}_n]$; thus,

$$
\begin{aligned}
E(X_{n+h}|\mathbf{x}_n) &= E[E(X_{n+h}|\boldsymbol{\theta}, \mathbf{x}_n)|\mathbf{x}_n] \\
&= E\left[\alpha^h X_n + \tfrac{1-\alpha^h}{1-\alpha}\lambda \Big| \mathbf{x}_n\right] \\
&= X_n E\left[\alpha^h|\mathbf{x}_n\right] + E\left[\tfrac{1-\alpha^h}{1-\alpha}\lambda \,\Big|\, \mathbf{x}_n\right].
\end{aligned}
$$

These expected values can be estimated through Markov Chain Monte Carlo (MCMC) algorithms. We perform Metropolis algorithm in conjunction with Adaptive Rejection Sampling Method (ARMS) in order to sample values from full conditional distributions of α and λ; let them be noted by

$(\alpha_1, \alpha_2, \ldots, \alpha_m)$, $(\lambda_1, \lambda_2, \ldots, \lambda_m)$, respectively (see Silva *et al.*, 2005). We have,

$$\begin{aligned} \hat{E}\left[\alpha^h|\mathbf{x}_n\right] &= \tfrac{1}{m}\textstyle\sum_{i=1}^m \alpha_i^h, \\ \hat{E}\left[\tfrac{1-\alpha^h}{1-\alpha}\lambda\Big|\mathbf{x}_n\right] &= \tfrac{1}{m}\textstyle\sum_{i=1}^m \tfrac{1-\alpha_i^h}{1-\alpha_i}\lambda_i. \end{aligned}$$

Consequently the predictor can be written as

$$\hat{X}_{n+h} = X_n\left(\frac{1}{m}\sum_{i=1}^m \alpha_i^h\right) + \left(\frac{1}{m}\sum_{i=1}^m \frac{1-\alpha_i^h}{1-\alpha_i}\lambda_i\right). \tag{8}$$

3 Interval Prediction

3.1 Classical Methodology

A confidence interval for the predictor $\hat{X}_{n+h}$, can be calculated through the probability function of the h-step-ahead prediction error, given by

$$e_{n+h}|\mathbf{x}_n = X_{n+h}|\mathbf{x}_n - \hat{X}_{n+h}|\mathbf{x}_n.$$

Replacing $\hat{X}_{n+h}|\mathbf{x}_n$ given by (4), we obtain

$$e_{n+h}|\mathbf{x}_n = X_{n+h} - \alpha^h x_n - \lambda\frac{1-\alpha^h}{1-\alpha}.$$

Since $e_{n+h}|\mathbf{x}_n$ is a function of discrete random variable X_{n+h}, we have

$$e_{n+h}|\mathbf{x}_n = k - \alpha^h x_n - \lambda\frac{1-\alpha^h}{1-\alpha}, \qquad k = 0, 1, 2, \ldots.$$

So,

$$\begin{aligned} &P(e_{n+h} = k - \alpha^h x_n - \lambda\tfrac{1-\alpha^h}{1-\alpha}|\mathbf{x}_n) = P(X_{n+h} = k|X_n = x_n) = \\ &\exp\left\{-\lambda\frac{1-\alpha^h}{1-\alpha}\right\}\sum_{i=0}^{M_h}\frac{1}{(k-i)!}\left(\lambda\frac{1-\alpha^h}{1-\alpha}\right)^{k-i}\binom{x_n}{i}(\alpha^h)^i(1-\alpha^h)^{x_n-i}. \end{aligned} \tag{9}$$

From the expression (9) we obtain a γ level confidence interval for X_{n+h}

$$(\hat{X}_{n+h} + e_{t_1}, \hat{X}_{n+h} + e_{t_2}), \tag{10}$$

where $\hat{X}_{n+h}$ is given by (4), e_{t_1} is the largest value of e_{n+h} such as $\mathrm{P}(e_{n+h} \leq e_{t_1}) \leq (1-\gamma)/2$ and e_{t_2} is the smallest value of e_{n+h} such as $\mathrm{P}(e_{n+h} \leq e_{t_2}) \geq (1+\gamma)/2$.

3.2 Bayesian Methodology

The procedure used in this section is an adaptive generalization of the method used to obtain HPD (*Highest Posterior Density*) intervals of the model parameters, in which we consider the predictive distribution instead of the posterior.

Definition 2 *A 100γ% predictive interval for X_{n+h} is given by*

$$P(X_L \leq X_{n+h} \leq X_R) = \sum_{x_{n+h}=X_L}^{X_R} f(x_{n+h}|\mathbf{x}_n).$$

However, since $f(x_{n+h}|\mathbf{x}_n)$ is not always symmetric[1], the intervals with a maximum posterior predictive probability are more desirable than predictive intervals (Chen *et al.*, 2000).

Definition 3 *$R(\gamma) = (X_L, X_R)$ is a 100γ% HPD interval for X_{n+h} if*

$$P(X_L \leq X_{n+h} \leq X_R) = \sum_{x_{n+h}=X_L}^{X_R} f(x_{n+h}|\mathbf{x}_n) \geq K_\gamma, \tag{11}$$

where K_γ is the largest constant such that $P[X_{n+h} \in R(\gamma)] \geq \gamma$.

Due to complexity of the predictive probability function given by (7) it is not possible to calculate the exact HPD interval for X_{n+h}; we can give an approximation for $R(\gamma)$ by using the Chen and Shao (1999) algorithm, because this method does not require the knowledge of the closed form of $f(x_{n+h}|\mathbf{x}_n)$. The Chen and Shao algorithm, can be described as,

Algorithm 2

1. *Obtain an MCMC sample $(X_{n+h,1}, X_{n+h,2}, \ldots, X_{n+h,m})$ (Algorithm 1);*
2. *consider $(X_{(n+h,1)} \leq X_{(n+h,2)} \leq \ldots \leq X_{(n+h,m)})$;*
3. *compute the $100\gamma\%$ credible intervals*

$$R_i(\gamma) = (X_{(n+h,i)}, X_{(n+h,i+[m\gamma])}), \;\; 1 \leq i \leq m - [m\gamma],$$

   where $[m\gamma]$ is integer part of $m\gamma$;

4. *the $100\gamma\%$ HPD interval to X_{n+h} is the one, denoted by $\hat{R}(\gamma)$, with the smallest amplitude among all credible intervals.*

Under certain regularity conditions, $\hat{R}(\gamma) \to R(\gamma)$ a.s. as $n \to \infty$, where $R(\gamma)$ is defined in (11) (Chen *et al.*, 2000).
Sometimes we obtain more than one interval. For this situation, we consider for $\hat{R}(\gamma)$ the interval with greater absolute frequency, among the smaller intervals width. When the interval is still not unique we take the one with the smallest lower limit of the interval.

4 A simulation Study

For the simulation study we consider samples with size $n = 40, 90, 190$ generated by INAR(1) models with the parameters values $\alpha = 0.2, 0.5, 0.8$ and $\lambda = 1, 3$.
Point predictions for 10 steps ahead given the last observation are displayed in Table 1. The table includes the h-step ahead simulated and predicted values, the squared and the absolute deviations between x_{190} and $x_{190+h}, h = 1, \ldots, 10$. The last line contains the classical limiting distribution. Independently of prediction methodology used, the forecasts performance depends on two basic aspects: one is the difference between x_n and $x_{n+h}, h > 1$ (see Figure 1); the other is the approximation between x_n and $\hat{\lambda}/(1-\hat{\alpha})$, in particular with the increase in h (note that $\hat{E}(X_{n+h}|X_n) \to \hat{\lambda}/(1-\hat{\alpha}), h \to \infty$).

From the various simulated samples we conclude that large values of α and λ are related with high dispersion values. Consequently the increase in α and λ provides large values of $|x_{n+h} - x_n|, h > 1$.
To confront classical and Bayesian methodologies we use the mean square error (MSE) to compare means, the mean absolute deviation (MAD) to compare medians and the "everything or nothing" lost function (FPTN), given by $1/n \sum I(x_{n+h})$ where

$$I(x_{n+h}) = \begin{cases} 1 & \text{if } |\hat{x}_{n+h} - x_{n+h}| > \delta \\ 0 & \text{if } |\hat{x}_{n+h} - x_{n+h}| \leq \delta \end{cases},$$

to compare modes. In this situation we consider $\delta = 1$ since we have integer values.
Table 2 shows the MSE, MAD and FPTN values from 10 predictions h-step-ahead. Values of MSE($\hat{X}_{n+h}$) and MSE($\tilde{X}_{n+h}$) are obtained considering the Bayesian predictors given by (8) and Algorithm 1, respectively. Values of MAD and FPTN were calculated, respectively, through medians and modes obtained by Algorithm 1. As we can see, when $\alpha = 0.8$ Bayesian methodology provides smaller values than classical methodology, so the Bayesian predictions seems to have a better performance than classical predictions.
In order to study and compare the estimates given by the sample mean, sample median and sample mode we used the minimum absolute percentual error (MAPE), given by

$$1/H \sum_{h=1}^{H} |\hat{X}_{n+h} - X_{n+h}|/X_{n+h},$$

where H represents the number of predictions realized. This criteria does not benefit any measure in particular. The results are presented in Table 3 for three samples with sizes 40, 90 and 190 of the model $x_t = \alpha \circ x_{t-1} + \epsilon_t$, $\epsilon_t \frown P(3)$. As we can see, neither of them (mean, median or mode) is the best. Both tables are presented considering $\lambda = 3$, but for $\lambda = 1$ the conclusions are similar.
Prediction intervals for future observations were calculated using expression (10) for classical methodology and Chen and Shao algorithm for Bayesian methodology. The simulation results for the case $\lambda = 3, n = 90, \gamma = 0.95$, which are typical, are presented in Tables 4 and 5 which show, respectively, classical and Bayesian situations.
Tables 4 and 5 indicate that the prediction interval amplitude increases not only with the number of steps ahead but also with the increase in α (see also Figures 2 and 3). Independently of the value of λ, when α is small the prediction interval converges more quickly than when α is large. Since we are working with discrete variables the confidence level is not always attained in the classical methodology.

[1] We made a previous study with some samples from PoINAR(1) and we verified that many were neither symmetric nor unimodal.

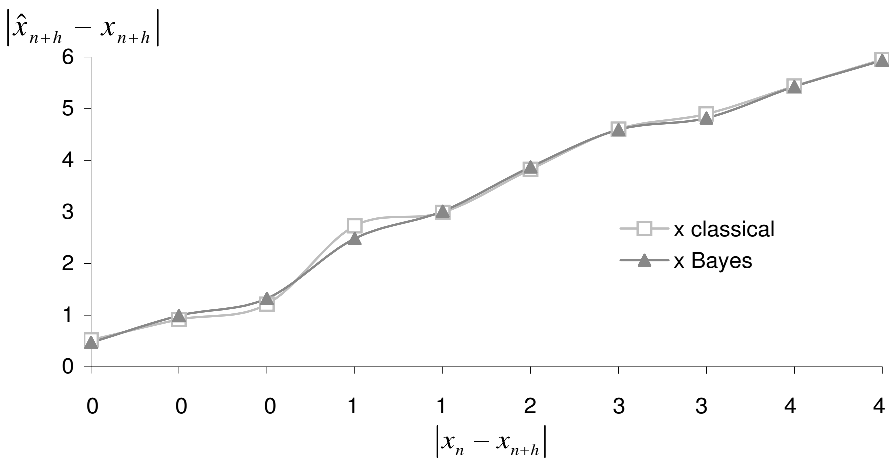

Figure 1: Values of $|\hat{x}_{n+h} - x_{n+h}|$ for a PoINAR(1) sample with $\alpha = 0.8$, $\lambda = 3$, $n = 190$ and $h = 1, 2, \ldots, 10$.

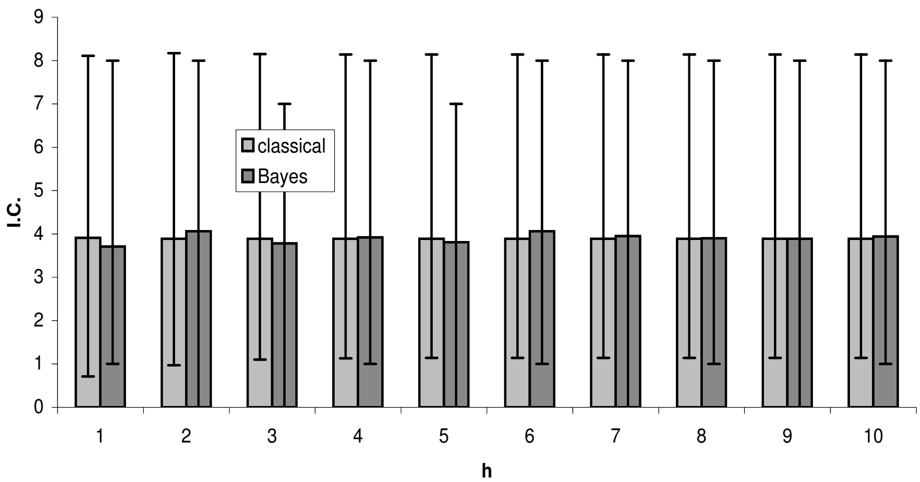

Figure 2: Pontual and intervalar prediction h-step-ahead for PoINAR(1) model with $\alpha = 0.2$, $\lambda = 3$ and $n = 190$.

Table 1: Point predictions considering two samples of size n=190 with parameters ($\lambda = 1, \alpha = 0.2, x_{190} = 0$) and ($\lambda = 3, \alpha = 0.8, x_{190} = 16$), respectively.

($\lambda = 1, \alpha = 0.2; x_{190} = 0$)						
			classical approach		Bayesian approach	
h	x_{190+h}	jump	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{n+h})^2$	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{n+h})^2$
1	2	2	1.068	0.869	1.090	0.828
2	0	0	1.247	1.555	1.292	1.670
3	0	0	1.277	1.631	1.340	1.796
4	0	0	1.282	1.643	1.342	1.801
5	5	5	1.283	13.816	1.288	13.779
6	0	0	1.283	1.646	1.302	1.695
7	1	1	1.283	0.080	1.210	0.044
8	1	1	1.283	0.080	1.348	0.121
9	1	1	1.283	0.080	1.248	0.061
10	2	2	1.283	0.514	1.298	0.493
∞			1.283			

($\lambda = 3, \alpha = 0.8; x_{190} = 16$))						
			classical approach		Bayesian approach	
h	x_{190+h}	jump	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{n+h})^2$	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{n+h})^2$
1	16	0	15.477	0.274	15.530	0.221
2	16	0	15.084	0.839	15.010	0.980
3	16	0	14.787	1.471	14.678	1.748
4	20	4	14.564	29.550	14.574	29.441
5	19	3	14.396	21.197	14.408	21.086
6	17	1	14.270	7.453	14.516	6.170
7	18	2	14.175	14.631	14.128	14.992
8	19	3	14.103	23.981	14.182	23.213
9	20	4	14.049	35.414	14.066	35.212
10	17	1	14.008	8.952	13.986	9.084
∞			13.884			

Table 2: Mean square error, mean absolute deviation and FPTN values of h-step-ahead ($h = 1, 2, \ldots, 10$), considering samples with sizes 40, 90 and 190 simulated from PoINAR(1) model, $x_t = \alpha \circ x_{t-1} + \epsilon_t,\ \epsilon_t \frown$ P(3) The indices "C" or "B" indicate which methodology is used (classical or Bayesian, respectively).

	α	0.2			0.8		
	n	40	90	190	40	90	190
MSE	$\hat{X}_{n+h,C}$	6.75	2.01	5.64	15.16	4.11	14.54
	$\hat{X}_{n+h,B}$	6.72	2.03	5.61	13.02	4.66	13.74
	$\tilde{X}_{n+h,B}$	6.76	1.87	5.69	15.70	4.01	14.26
MAD	$\hat{m}_{n+h,C}$	2.18	1.18	2.00	3.36	1.73	3.55
	$\tilde{m}_{n+h,B}$	2.46	1.18	2.09	3.27	1.77	3.46
FPTN	$\hat{mo}_{n+h,C}$	0.45	0.45	0.45	0.64	0.55	0.82
	$\tilde{mo}_{n+h,B}$	0.64	0.45	0.64	0.64	0.73	0.73

Table 3: Values of MAPE considering 10 one-step-ahead predictions for the model $x_t = \alpha \circ x_{t-1} + \epsilon_t,\ \epsilon_t \frown$ P(3) and sample sizes 40, 90 and 190. The indices "C" or "B" indicate which methodology is used (classical or Bayesian, respectively).

α	0.2			0.8		
n	40	90	190	40	90	190
$\hat{X}_{n+h,C}$	0.714	0.573	0.870	0.707	0.564	0.919
$\tilde{X}_{n+h,B}$	**0.178**	0.137	**0.260**	0.110	0.109	**0.0811**
$\hat{m}_{n+h,C}$	0.652	0.588	**0.631**	**0.606**	0.561	**0.831**
$\tilde{m}_{n+h,B}$	0.187	**0.120**	0.209	0.115	**0.103**	0.091
$\hat{mo}_{n+h,C}$	**0.619**	**0.464**	0.929	0.625	**0.506**	**0.831**
$\tilde{mo}_{n+h,B}$	0.187	0.127	0.231	**0.086**	0.125	0.098

Table 4: 95% confidence intervals for the h-step-ahead future values for INAR(1) model ($n = 90, \lambda = 3$).

	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
h	x_{n+h}	$\hat{x}_{n+h}$	int.	x_{n+h}	$\hat{x}_{n+h}$	int	x_{n+h}	$\hat{x}_{n+h}$	int
1	2	3.91	(0.11,8.11)	4	5.39	(0.88,9.89)	16	13.96	(8.76,18.77)
2	4	3.89	(0.13,8.13)	3	5.57	(0.82,10.82)	16	13.92	(7.56,19.56)
3	1	3.89	(0.14,8.14)	3	5.66	(0.79,10.79)	13	13.90	(6.41,20.41)
4	2	3.89	(0.14,8.14)	7	5.71	(0.77,10.77)	14	13.88	(6.29,21.29)
5	5	3.89	(0.14,8.14)	6	5.73	(0.76,10.76)	11	13.86	(6.19,21.19)
6	5	3.89	(0.14,8.14)	4	5.74	(0.75,10.75)	12	13.85	(6.11,21.11)
7	5	3.89	(0.14,8.14)	6	5.74	(0.75,10.75)	14	13.84	(6.05,22.05)
8	3	3.89	(0.14,8.14)	7	5.74	(0.75,10.75)	14	13.83	(5.99,21.99)
9	3	3.89	(0.14,8.14)	6	5.74	(0.75,10.75)	11	13.83	(5.96,21.96)
10	4	3.89	(0.14,8.14)	5	5.74	(0.75,10.75)	11	13.82	(5.93,21.93)
∞		3.89	(0,8)		5.75	(1,11)		13.80	(6,22)

Table 5: 95% HPD intervals or the h-step-ahead future values for INAR(1) model ($n = 90, \lambda = 3$) based on 500 replications.

	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
h	x_{n+h}	$\tilde{x}_{n+h}$	int.	x_{n+h}	$\tilde{x}_{n+h}$	int	x_{n+h}	$\tilde{x}_{n+h}$	int
1	2	3.712	(1,8)	4	5.346	(2,10)	16	13.990	(10,19)
2	4	4.064	(0,8)	3	5.626	(2,10)	16	14.092	(9,20)
3	1	3.776	(0,7)	3	5.754	(2,11)	13	13.956	(8,20)
4	2	3.924	(1,8)	7	5.662	(1,10)	14	13.638	(7,20)
5	5	3.808	(0,7)	6	5.708	(1,10)	11	13.930	(8,22)
6	5	4.060	(1,8)	4	5.866	(2,11)	12	13.906	(7,20)
7	5	3.952	(0,8)	6	5.828	(2,11)	14	13.870	(6,21)
8	3	3.902	(1,8)	7	5.774	(2,11)	14	13.906	(7,22)
9	3	3.892	(0,8)	6	5.910	(1,10)	11	13.904	(6,21)
10	4	3.942	(1,8)	5	5.728	(2,11)	11	13.792	(7,21)

5 Analysis of burn claims data

We apply the proposed methodology to a data set analysed by Freeland (1998) comprising 120 monthly counts of workers collecting Wage Loss Benefits (WLB) for burn injuries received. All the claimants are male, between the ages of 35 and 54, work in the logging industry and reported their claim to the Richmond service delivery location. Clearly these data may be considered as a birth and death process. Let X_t be the number of workers collecting WLB at time t. This number can be viewed as the sum of the number of claimants from time $t-1$ and the number of new claims at time t. All the descriptive details of the data set can be found in Freeland and McCabe (2004) which conclude PoINAR(1) is a plausible choice of modelling the data.

In order to evaluate and compare the different prediction methodologies, h-step ahead forecasts ($h = 1, 2, 3, 4, 5, 6$) are produced for the time period from July to December 1994, for which we know the observed values. The point forecasts based on the mean, median and mode and the observed values are presented in Tables 6 and 7. The mean squared error of classic point predictions is smaller than that of Bayesian predictions. This result is expected in view of the simulation results presented in the last section since the estimated value for alpha is 0.4. Also, as expected, one step-ahead predictions present smaller errors (Table 6).

Interval predictions for the period July to December 1994 are obtained using the two approaches proposed given by equation (10) and Algorithm 2. The intervals obtained, presented in Table 8, are analogous, although the Bayesian have smaller width.

Table 9 presents pontual predictions for the first semester of 1995, given observations up to December 1994, obtained using Bayesian and classical (Freeland, 1998) methodologies. The analysis of the table indicates that Bayesian predictions based on means are slightly higher than the classical ones, whilst those based on the median coincide. However, point predictions based on the mode differ between the two methodologies. The intervals obtained, presented in Table 10, are the same considering each approach and $h > 3$. However, as happened before, the HPD interval has smaller width.

6 Final remarks

Forecasting low integer values of time series of counts remains an open problem. Conditional means do not preserve coherently the integer nature of the data. Here we use a Bayesian approach which allows the estimation of both point and interval predictions. Simulations indicate that the performance of the different methods depend on the value of the model parameter.

References

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

Brännš, K. (1994). Estimation and testing in integer-valued AR(1) models. Umeå Economic Studies 335.

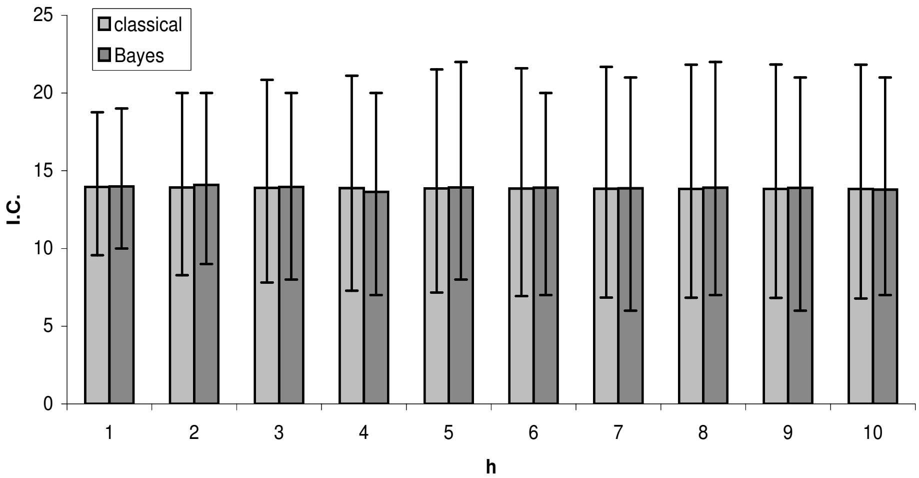

Figure 3: Pontual and intervalar prediction h-step-ahead for PoINAR(1) model with $\alpha = 0.8$, $\lambda = 3$ and $n = 190$.

Table 6: Point prediction one step ahead of monthly claims count from July to December 1994.

prediction one step ahead (h=1)							
		classical predictions			Bayesian predictions		
year/month	claims	mean	median	mode	mean	median	mode
94/07	11	7.89	8	7	7.67	7	7
94/08	12	9.50	9	9	9.50	9	9
94/09	11	9.94	10	10	10.05	10	10
94/10	12	9.57	9	9	9.75	10	11
94/11	7	10.02	10	10	10.06	10	10
94/12	11	7.95	8	8	7.78	8	7
		EQM=	DAM=	FPTN=	EQM=	DAM=	FPTN=
		6.89	2.67	0.83	7.17	2.67	0.67

Table 7: h-step ahead predictions of monthly claims count from July to December 1994.

h-step ahead predictions								
			classical predictions			Bayesian predictions		
h	year/month	claims of	mean	median	mode	mean	median	mode
1	94/07	11	7.89	8	7	7.67	7	7
2	94/08	12	8.24	8	8	8.01	8	7
3	94/09	11	8.38	8	8	8.14	8	7
4	94/10	12	8.44	8	8	8.21	8	7
5	94/11	7	8.46	8	8	8.22	8	7
6	94/12	11	8.47	8	8	8.22	8	7
			EQM=	DAM=	FPTN=	EQM=	DAM=	FPTN=
			8.67	3.00	0.83	9.79	3.17	0.83

Cameron, A.C. e Trivedi, P.K. (1998). *Regression analysis of count data.* Cambridge University Press: Cambridge.

Chen, M-H. e Shao, Q-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Jounal of Computational and Graphical Statistics*, 8, 69-92.

Chen, M-H., Shao, Q-M e Ibrahim, J.G. (2000). *Monte Carlo methods in Bayesian computation.* Springer Series in Statistics.

Cox, D.R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8, 93-115.

Freeland, R.K. (1998). *Statistical analysis of discrete time series with application to the analysis of Workert's Compensation Claims Data.* Ph.D. thesis. The University of British Columbia, Canada.

Freeland, R.K. e McCabe, B.P.M. (2003). Forecasting discrete valued low count series. *International Journal of Forecasting*, 20, 427-434.

Freeland, R.K. e McCabe, B.P.M. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of the Time Series Analysis*, 25, 701-722.

Freeland, R.K. e McCabe, B.P.M. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*, 73, 147-153.

Jung, R.C. and Tremayne, A.R (2005). Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22, 223-238.

MacDonald, I.L. e Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series.* Chapman and Hall:London.

McCabe, B.P.M. e Martin, G.M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting* , 21, 315-330.

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

McKenzie, E. (1988). Some ARMA for dependent sequences of Poisson counts. *Advanced in Applied Probability*, 20, 822-835.

McKenzie, E. (2003). Discrete variate time series. *Handbook of Statistics*, 21, 576-606.

Silva, I., M.E. Silva, I. Pereira e Silva N. (2005). *Replicated INAR(1) Process. Methodology and Computing in Applied Probability*, 7, 517-542.

Steutel, F.W. e van Harn K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 5, p. 893-899.

Table 8: 95% confidence and HPD intervals for h-step ahead of monthly claims count from July to December 1994.

h	h=1	h=2	h=3	h=4	h=5	h=6
classical	(2.04,13.04)	(2.03,14.03)	(2.02,14.02)	(2.01,15.01)	(2.00,15.00)	(2.00,15.00)
Bayesian	(3.00,13.00)	(3.00,14.00)	(3.00,14.00)	(3.00,14.00)	(3.00,14.00)	(3.00,14.00)

Table 9: h-step ahead predictions of monthly claims count from January to June 1995.

		classical predictions (Freeland)			Bayesian predictions		
h	year/month	mean	median	mode	mean	median	mode
1	95/1	9.60	9	9	9.73	10	11
2	95/2	9.04	9	9	9.19	9	7
3	95/3	8.82	9	8	8.99	9	7
4	95/4	8.73	9	8	8.91	9	9
5	95/5	8.69	9	8	8.89	9	10
6	95/6	8.68	9	8	8.86	9	10

Table 10: 95% confidence and HPD intervals for h-step ahead of monthly claims count from January to June 1995.

h	h=1	h=2	h=3	h=4	h=5	h=6
classical	(3.04,15.04)	(3.09,15.09)	(2.12,15.12)	(2.14,15.14)	(2.15,15.15)	(2.16,15.16)
Bayesian	(5,16)	(4,16)	(3,15)	(3,15)	(3,15)	(3,15)