Bayesian nonparametric predictions for count time series

Luisa Bisaglia and Antonio Canale

Abstract In this paper we introduce a Bayesian nonparametric methodology for producing coherent predictions of count time series using the INAR(1) process. Our predictions are based on estimates of the p-step ahead predictive mass functions assuming a nonparametric prior for the distribution of the error term having large support on the space of discrete probability mass functions. An efficient Gibbs sampler is developed for posterior computation.

Key words: INAR(1); Dirichlet process mixtures; Gibbs sampling algorithm

1 Introduction

Recently, there has been a growing interest in studying nonnegative integer-valued time series and, in particular, time series of counts. Examples are categorical time series, binary processes, birth-death models and counting series.

The most common approach to build an integer-valued autoregressive processes is using a probabilistic operation called thinning. Using binomial thinning, Al-Osh & Alzaid (1987) and McKenzie (1988) first introduced integer-valued autoregressive processes (INAR). A recent review on integer-valued AR processes can be found in Silva et al. (2005) and Jung & Tremayne (2011). While theoretical properties of INAR models have been extensively studied in the literature, relatively few contributions discuss the development of forecasting methods that are coherent, in the sense of producing only integer forecasts of the count variable. Freeland & McCabe (2004), in the context of INAR(1) process with Poisson innovations suggest some solutions that are somewhat problem-specific. Thus, McCabe & Martin (2005)

L. Bisaglia
Dept. of Statistical Science, Università di Padova, e-mail: bisaglia@stat.unip.it

A. Canale
Dept. of Economics and Statistics, Università di Torino, e-mail: antonio.canale@unito.it

consider the Bayesian point of view and present a methodology for producing coherent forecasts of low count time series that is completely general. The predictive probability mass function, defined only over the support of the discrete count variable, is a natural outcome of Bayes theorem.The results are valid for any sample size and not only asymptotically, moreover the innovations can be any arbitrary discrete distribution, within a specified finite set of distributions. In particular, the authors focus on Poisson, binomial and negative binomial distributions.

In this paper, we consider INAR(1) models with flexible specifications of the error term under a Bayesian nonparametric approach. The assumption of a nonparametric prior with large support for the innovation distribution, bypasses the need to specify a finite set of discrete distribution as in McCabe & Martin (2005). Our approach leads to two main improvements: first we overcome the specification of the predictive probability as a mixture of K predictive distributions, and second we do not rely on the usual strict parametric models. Among the different proposal made in the Bayesian nonparametric literature to model count distributions, we use that of Canale & Dunson (2011). Some possible applications are, for example, the number of clients in an Internet server by hour, the daily number of traded stocks in a firm, the daily number of guests in a hotel, the monthly incidence of a disease. Furthermore, also continuous valued series in which the observations fall in one of a small number of categories or that can be discretized, can be treated as integer-valued time series.

2 Model specification

2.1 The INAR(1) model

To introduce the class of INAR model we first recall the thinning operator, '$\circ$', defined as follows.

Definition *Let Y be a non negative integer-valued random variable, then for any* $\alpha \in [0,1]$

$$\alpha \circ Y = \sum_{i=1}^{Y} X_i$$

where X_i is a sequence of iid count random variables, independent of Y, with common mean α.

The INAR(1) process $\{Y_t; t \in \mathbf{Z}\}$ is defined by the recursion

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t \quad (1)$$

where $\alpha \in [0,1]$, and ε_t is sequence of iid discrete random variables with finite first and second moment. The components of the process $\{Y_t\}$ are the surviving elements of the process Y_{t-1} during the period $(t-1,t]$, and the number of elements which entered the system in the same interval, ε_t. Each element of Y_{t-1} survives with probability α and its survival has no effect on the survival of the other elements, nor on ε_t which is not observed and cannot be derived from the Y process in the INAR(1) model. In the next section we discuss a nonparametric prior for the distribution of the error term.

2.2 *Rounded mixture priors*

To define a nonparametric model for counts, Canale & Dunson (2011) proposed to round an underlying variable having an unknown density given a Dirichlet process mixture of Gaussians prior. Such rounded mixture of Gaussians (RMG) have been showed to be highly flexible and having excellent performance in small samples while having appealing asymptotic properties in terms of large support and strong posterior consistency.

Following Canale & Dunson (2011) we let the probability that the discrete error equals j, for $j \in \mathbf{N}$ to be

$$p(j) = g(f)[j] = \int_{a_j}^{a_{j+1}} f(y^*)dy^* \tag{2}$$

with the thresholds chosen as $a_0 = -\infty$ and $a_j = j-1$ for $j \in \{1,2,\dots\}$ and modelling the underlying f as the mixture model

$$f(y^*;P) = \int \phi(y^*;\mu,\tau^{-1})dP(\mu,\tau), \quad P \sim DP(\eta P_0). \tag{3}$$

Here, $\phi(y;\mu,\tau^{-1})$ is a Gaussian density having mean μ and precision τ and $DP(\eta P_0)$ corresponding to the Dirichlet process with P_0 chosen to be Normal-Gamma and $\eta > 0$. Equations (2)–(3) induce a prior $p \sim \Pi$ over $\mathscr{C}$, the pace of the probability mass functions on the non negative integers.

3 p-step ahead predictive probability mass function

Exploiting the birth-and-death process interpretation of the INAR(1) model, the distribution of Y_t given y_{t-1}, α and p is

$$Pr(Y_t = y_t \mid y_{t-1}, \alpha, p) = \sum_{s=0}^{\min\{y_t, y_{t-1}\}} Pr(B^{\alpha}_{y-1} = s) \times p(y_t - s) \tag{4}$$

where p is a random probability measure obtained through (2)–(3) and $B^{\pi}_{k} \sim \mathrm{Be}(k,\pi)$.

The likelihood function given $\mathbf{y} = (y_1, \ldots, y_T)$ of α and the random discrete measure p turns out to be

$$\ell(\theta \mid \mathbf{y}) \propto \prod_{t=2}^{T} \sum_{s=0}^{\min\{y_t, y_{t-1}\}} \alpha^s (1-\alpha)^{y_{t-1}-s} p(y_t - s) \tag{5}$$

where $\theta \in \Theta$ and $\Theta = \mathbf{R} \times \mathscr{C}$. The posterior distribution can be obtained as

$$\pi(\theta \mid \mathbf{y}) \propto \ell(\theta \mid \mathbf{y}) \pi(\theta) \tag{6}$$

where $\pi(\theta)$ is the prior probability. Given the nonparametric prior $p \sim \Pi$ it is sufficient to elicit a prior for $\alpha \sim \pi_\alpha$. In presence of prior information we can use a beta distribution with given mean corresponding to one's prior belief about α. Being noninformative one can assume a uniform prior distribution between zero and one. Assuming that α and p are independent a priori, the prior $\pi(\theta)$ is $\pi(\theta) = \Pi \times \pi_\alpha$.

The p-step ahead probability mass function is here defined as

$$Pr(Y_{T+p} = j \mid \mathbf{y}) = \int_\Theta Pr(Y_{T+p} = j \mid \mathbf{y}, \theta) d\pi(\theta \mid \mathbf{y}) \tag{7}$$

where $\pi(\theta \mid \mathbf{y})$ is the posterior distribution (6).

The following Gibbs sampler computes the quantity in (7) iterating the following steps

1. Data augmentation step given p and α.
   - For $t = 2, \ldots, T$, given y_{t-1}, simulate B_t from $P(B_t = j) \propto \binom{y_{t-1}}{j} \alpha^j (1-\alpha)^{y_{t-1}-j} \times p(y_t - j)$ for $j = 0, \ldots, y_t$.
   - For $t = 2, \ldots, T$, simulate $\varepsilon_t^* \sim f$ where f is as in (2)–(3) under the constraints $a_{y_t - B_t} \leq \varepsilon_t^* \leq a_{y_t - B_t + 1}$
2. Update the parameters of the RMG as in Canale & Dunson (2011)
3. Update α from its conditional posterior distribution via Metropolis-Hastings step
4. After burn in, simulate Y_{t+p} as in equation (4)

The main advantage of this approach with respect McCabe & Martin (2005) is that we does not assume any parametric model within a set of k models. Also there are advantages in terms of forecast accuracy which we show by mean of an exstensive Monte Carlo experiment.

References

Al-Osh, M. A. & Alzaid, A. A. (1987). First order integer-valued autoregressive INAR(1) process. *Journal of Time Series Analysis*, **8**(3), 261–275.

Canale, A. & Dunson, D. B. (2011). Bayesian kernel mixtures for counts. *Journal of the American Statistical Association*, **106**, 1528–1539.

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

Jung, R. C. & Tremayne, A. R. (2011). Useful models of time series of counts or simply wrong ones? *Advances in Statistical Analysis*, **95**, 59–91.

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

McKenzie, E. (1988). Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability*, **20**, 822–835.

Silva, I., Silva, M. E., Pereira, I., & Silva, N. (2005). Replicated inar(1) process. *Methodology and Computing in Applied probability*, **7**, 893–899.