Exact Likelihood Estimation and Probabilistic Forecasting in Higher-order INAR(p) Models

Lu, Yang

University of Paris 13

1 January 2018

Exact Likelihood Estimation and Probabilistic Forecasting in Higher-order INAR(p) Models

Yang Lu*

January 5, 2018

Abstract: The computation of the likelihood function and the term structure of probabilistic forecasts in higher-order INAR(p) models are qualified numerically intractable and the literature has considered various approximations. Using the notion of compound autoregressive process, we propose an *exact* and fast algorithm for both quantities. We find that existing approximation schemes induce significant errors for forecasting.

Keywords: compound autoregressive process, probabilistic forecast of counts, matrix arithmetic.

MSC code: 62-15, JEL code: C32

1 Introduction

INteger-valued AutoRegressive process (INAR) has recently received wider attention in the literature. The benchmark model, introduced by McKenzie (1985), Al-Osh and Alzaid (1987) in the first-order case (called INAR(1)) and Du and Li (1991) in the higher-order case (called INAR(p)), postulates that:

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t, \qquad \forall t, \tag{1}$$

where the binomial thinning operators $\alpha_i \circ X_{t-i}$, with i and t varying, are mutually conditionally independent given the past $\underline{X_{t-1}} = (X_{t-1}, X_{t-2}, ...)$, and have binomial distributions

*University of Paris 13, Department of Economics (CEPN). Email: luyang000278@gmail.com

$Binom(X_{t-i}, \alpha)$, whereas the innovation process (ϵ_t) is i.i.d. and Poisson $\mathcal{P}(\lambda)$ distributed.

Since these seminal works, many extensions of the above model have been proposed. However, besides the basic INAR(1), the computation of the likelihood function and/or that of the multi-step-ahead conditional probability mass function (p.m.f.) in higher-order models are documented to be intractable and various approximation methods are proposed [see e.g. Pedeli et al. (2015) for estimation and Jung and Tremayne (2006); McCabe et al. (2011) for forecasting]. These methods can induce significant approximation errors, and some are still computationally intensive. This paper solves these two difficulties using the compound autoregressive (CaR) property of the INAR(p) model. Relying on simple, matrix algebra, we obtain a fast and unified algorithm for the two aforementioned quantities, without making any approximation error. Moreover, the methodology is applicable to a very wide range of INAR(p) models beyond the benchmark case.

The paper is organized as follows. Section 2 reviews a natural link between the probability generating function (p.g.f.) and the probability mass function (p.m.f.) for a count distribution. Section 3 computes the Taylor's expansion of the one-step-ahead conditional p.g.f. and deduce the corresponding p.m.f., which allows for likelihood-based estimation. Section 4 deals with forecasting and adapts the approach to deduce multiple-step-ahead p.m.f.'s. Section 5 concludes.

2 Link between the p.g.f. and p.m.f. of a count distribution

Let X be a count variable with known p.g.f.:

$$\phi(u) = \mathbb{E}[u^X] = \sum_{n=0}^{\infty} u^n p(n),$$

where argument $u \geq 0$. The aim is to compute the p.m.f. $p(n)$, for any $n \in \mathbb{N}$. While the analogous problem for continuous distributions usually involve approximation methods [see e.g. Barndorff-Nielsen and Cox (1979); Davies (1973)], this is not necessarily the case for a count

distribution. Indeed, the Taylor-expansion (at $u = 0$) of the p.g.f. is:

$$\phi(u) = \sum_{n=0}^{\infty} \frac{\phi^{(n)}(0)}{n!} u^n.$$

thus by identification we have:

$$p(n) = \frac{\phi^{(n)}(0)}{n!}. \tag{2}$$

Hence the p.m.f. can be computed *exactly*, so long as the Taylor's expansion of ϕ is tractable.

Example 1. For instance, if X follows Poisson $\mathcal{P}(\lambda)$ distribution, then we have: $\phi(u) = e^{\lambda(u-1)} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{u^n}{n!}$. Thus we recover the p.m.f. $p(n) = e^{-\lambda} \frac{\lambda^n}{n!}$.

3 Likelihood-based estimation

Let us now explain how to adapt this idea to the context of count process. In order to conduct maximum likelihood (ML) estimation, we have to derive $p(\cdot|\underline{X_{t-1}})$, i.e. the conditional p.m.f. of X_t given the past $\underline{X_{t-1}}$. This latter distribution is the convolution of p binomial distributions $\alpha_i \circ X_{t-i}$, $i = 1, ..., p$, as well as the Poisson distribution of ϵ_t. Thus we have [see e.g. Drost et al. (2009)]:

$$p(X_t|\underline{X_{t-1}}) = \sum_{n_1+n_2+\cdots+n_p+n_{p+1}=X_t} \Big(\prod_{j=1}^{p} \mathbb{P}[\alpha_i \circ X_{t-i} = n_i] \Big) \mathbb{P}[\epsilon_t = n_{p+1}]. \tag{3}$$

The RHS involves a $p-$dimensional summation, that is a complexity of $O(X_t^p)$. This explains why ML estimation has not yet been considered for INAR(p) models[1] with $p \geq 3$. While moment based estimators are generically consistent [see e.g. Al-Osh and Alzaid (1987)], they can suffer from significant efficiency loss [see e.g. Bu et al. (2008)]. Recently, Pedeli et al. (2015) propose a saddle-point approximation of $p(\cdot|\underline{X_{t-1}})$. Its drawbacks is, first, the approximated $p(n|\underline{X_{t-1}})$ does not sum up to one when n varies across $\mathbb{N}$. Second, the saddle-point itself has to be approximated numerically, resulting in further computational complexity and approximation error. Pedeli et al. (2015) show that for certain parameter values, the relative error of the likelihood function can be as large as $2-5$ percent. As a consequence, their approximate ML estimator is not consistent

[1] Bu and McCabe (2008) and Bu et al. (2008) estimate INAR(2) models.

and simulation results (see their Table 1) show that in finite sample, the bias can be significantly larger than that of the exact ML estimator.

Our solution consists in first computing the $n-$th Taylor's expansion (at zero) of the corresponding conditional p.g.f.:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] = \exp\Big[\sum_{i=1}^{p} X_{t-i}\log(\alpha_i u + 1 - \alpha_i) + \lambda(u-1)\Big] \tag{4}$$

$$= \exp\Big[-\lambda + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i) + \lambda u + \sum_{i=1}^{p} X_{t-i}\log(1+\frac{\alpha_i}{1-\alpha_i}u)\Big]$$

$$= \exp\Big[-\lambda + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i)\Big]\exp\Big[\lambda u + \sum_{i=1}^{p} X_{t-i}\sum_{j=1}^{n}\frac{(-1)^{j-1}}{j}(\frac{\alpha_i}{1-\alpha_i})^j u^j + O(u^{n+1})\Big]$$

$$= \exp\Big[-\lambda + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i)\Big]\exp\Big[\sum_{j=1}^{n} A_j u^j\Big] + O(u^{n+1})$$

$$= \exp\Big[-\lambda + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i)\Big]\sum_{k=0}^{n}\frac{1}{k!}\Big[\sum_{j=1}^{n} A_i u^j\Big]^k + O(u^{n+1}), \tag{5}$$

where for the computation of the likelihood function, we typically take $n = X_t$, and coefficients A_i are given by:

$$A_1 = \lambda + \sum_{i=1}^{p} X_{t-i}\frac{\alpha_i}{1-\alpha_i},$$
$$A_j = \frac{(-1)^{j-1}}{j}\sum_{i=1}^{p} X_{t-i}(\frac{\alpha_i}{1-\alpha_i})^j, \qquad \forall j = 2, ..., n.$$

Thus by equation (2), the probability $p(n|\underline{X_{t-1}})$ is the coefficient in front of the term u^n in the expansion (5). Hence it suffices to compute recursively the $n+1$ first terms of the polynomial $\Big[\sum_{i=1}^{n} A_i u^i\Big]^k$ for each k. While such an expansion can be obtained using a symbolic calculation package such as Mathematica, the following proposition provides a matrix-based algorithm that is simple for statistical packages[2].

Proposition 1. *The $n+1$ first coefficients of polynomial $\Big[A_0 + \sum_{j=1}^{n} A_i u^i\Big]^k$, where $k = 0, ...n$*

[2]The link between polynomial multiplication and matrix operations is well documented in the computer science literature. The algorithm we provide here is the simplest for array programming languages such as R or Matlab and there also exist algorithms with a lower complexity [see e.g. Knuth (1997)].

are given by the column vector:

$$\begin{bmatrix} A_0 & 0 & 0 & \cdots & 0 \\ A_1 & A_0 & 0 & \cdots & 0 \\ A_2 & A_1 & A_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_n & A_{n-1} & \cdots & A_1 & A_0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}. \tag{6}$$

The proof is obvious and omitted. In our case we have $A_0 = 0$, thus the square matrix above is triangular inferior.[3] We can also remark that on the contrary to formula (3), which becomes cumbersome so long as p is mildly large, in our approach, the computational cost is essentially independent of the value of p.

To illustrate the computational gain, we implement the likelihood function using both our new method and the direct method based on (3). We consider an INAR(2) model[4] as well as an INAR(5) model[5]. For both models, we simulate trajectories with different sample sizes $T = 100, 500$ and 2500, respectively and compute the likelihood function using the two methods.[6] The following table compares the execution time in R of the two methods on a standard laptop. We do not report the likelihood values obtained from these two methods as they coincide with each other.

Model	INAR(2)		INAR(5)	
Sample size	$T = 100$	$T = 500$	$T = 100$	$T = 500$
Our method	0.002 s	0.010 s	0.002 s	0.009 s
Direct method	0.001 s	0.007 s	0.060 s	0.192 s

Table 1: Execution time of the two methods. The time unit is second.

We can see that when the order p of the INAR model is small (i.e. $p = 2$), both methods are

[3] This square matrix is known as the Toeplitz matrix in the literature.

[4] The parameters of the model are fixed as:

$$\alpha_1 = 0.8, \alpha_2 = 0.1, \lambda = 0.5, \tag{7}$$

with initial values $X_1 = X_2 = 1$.

[5] The parameters of this model are fixed as:

$$\alpha_1 = 0.3, \alpha_2 = 0.2, \alpha_3 = 0.1, \alpha_4 = 0.05, \alpha_5 = 0.01, \lambda = 1, \tag{8}$$

with initial values $X_1 = \cdots = X_5 = 1$.

[6] The R program is available upon request.

fast. This is understandable, since for this model, the computation of the conditional p.g.f. has a complexity of $O(X_t^3)$ for both models. When p increases to 5, our method becomes dominant, as its computational cost remains nearly unchanged whereas the direct method is 25 times slower.

To summarize, the simplicity of our approach is mainly due to equation (4), which says that the one-step-ahead conditional p.g.f. is an exponential linear function of past observations. Such processes (X_t) are called ($p-$th order) compound autoregressive (CaR(p)) processes in the general time series literature [see Darolles et al. (2006)]. Their analysis is generically simple using the conditional p.g.f.[7] . Here since we have a count distribution, the p.m.f. is also easily computable.

4 Multi-step forecasting

While the conditional mean $\mathbb{E}[X_{t+h}|\underline{X_{t-1}}]$ is simple to obtain for the INAR(p) model [see e.g. Du and Li (1991)], it is not informative enough to characterize the whole conditional distribution. Moreover, since the mean is generically non-integer, it is incompatible with the discrete sample space. A proper, probabilistic forecasts involves the multiple-step-ahead conditional p.m.f. of X_{t+h} given $\underline{X_{t-1}}$:

$$p_h(n|\underline{X_{t-1}}) = \sum_{X_{t+h-1}=0}^{\infty} \sum_{X_{t+h-2}=0}^{\infty} \cdots \sum_{X_t=0}^{\infty} p(n|\underline{X_{t+h-1}}) \cdots p(X_t|\underline{X_{t-1}}), \tag{9}$$

which is a $h-$dimensional *infinite* summation involving $p(X_t|\underline{X_{t-1}})$. Thus equation (9) is in practice impossible to use. Jung and Tremayne (2006) propose to conduct simulations of future trajectories to approximate $p_h(n|\underline{X_{t-1}})$, but this latter is also highly computationally intensive.

Recently, Bu and McCabe (2008); McCabe et al. (2011) propose a closed form approximation for $p_h(n|\underline{X_{t-1}})$. Their idea is to neglect the probability that of process taking values larger than a threshold, say, n. Thus the p-dimensional vector $Y_t = (X_t, X_{t-1}, ..., X_{t-p+1})$, can be regarded as a first-order, finite state Markov chain. Then $p_h(\cdot|\underline{X_{t-1}})$ satisfies a recursive formula involving the $(n+1)^p \times (n+1)^p$ transition matrix Π of chain (Y_t) Its drawbacks are, first, it is unclear whether regularity conditions are satisfied to ensure that when n goes to infinity,

[7] Darolles et al. (2006) propose to conditional Laplace transform. In our case since (X_t) takes values in $\mathbb{N}$, it is easier to work with the conditional p.g.f.

the approximated conditional p.m.f. and the marginal p.m.f.[8] converge to their theoretical counterparts [see Gibson and Seneta (1987); Tweedie (1998) for a discussion]. Secondly, the dimension of Π is ultra-high, rendering this approach impossible so long as n and/or p are moderately large.

Let us now propose a fast and exact algorithm using the Taylor's expansion approach. As in section 3, we first derive the conditional p.g.f., i.e. $\phi_h(u|\underline{X_{t-1}}) = \mathbb{E}[u^{X_{t+h}}|\underline{X_{t-1}}]$, using the CaR property of the model. The following proposition is a higher-order generalization of Corollary 1 in Darolles et al. (2006), which is focused on CaR(1) processes. It says that at higher horizon $h \geq 2$, $\phi_h(u|\underline{X_{t-1}})$ is still exponential affine in $\underline{X_{t-1}}$.

Proposition 2. *For any integer $h \geq 0$, we have:*

$$\phi_h(u|\underline{X_{t-1}}) := \mathbb{E}[u^{X_{t+h-1}}|\underline{X_{t-1}}] = \exp\Big[B(h,0)(u) + \sum_{i=1}^{p} B(h,i)(u)X_{t-i}\Big], \tag{10}$$

where functional (in u) coefficients $B(h,i)(u), i = 0, ..., p$ satisfy the recursive formula:

$$B(1,0)(u) = \lambda(u-1), \tag{11}$$

$$B(1,i)(u) = \log(\alpha_i u + 1 - \alpha_i), \qquad \forall i = 1, ..., p, \tag{12}$$

$$B(h+1,0)(u) = B(h,0)(u) + \lambda[e^{B(h,1)(u)} - 1], \tag{13}$$

$$B(h+1,p)(u) = \log(\alpha_p e^{B(h,1)(u)} + 1 - \alpha_p), \tag{14}$$

$$B(h+1,i)(u) = B(h,i+1)(u) + \log(\alpha_i e^{B(h,1)(u)} + 1 - \alpha_i), \qquad \forall i = 1, ..., p-1. \tag{15}$$

Proof. See Appendix. □

While $B(h,i)(u)$ are intractable for large h, their Taylor's expansions (at $u = 0$), and hence that of $\phi_h(u|\underline{X_{t-1}})$, can be easily obtained by recursion, thanks to the simple Taylor's expansion of the exponential/logarithmic functions involved in equations (11) to (15). More precisely,

- Suppose that the $n-$th order Taylor's expansions of $B(h,i)(u)$ with respect to u have already been obtained for each $i = 0, ..., p$, then equations (13), (14) and (15) allow to

[8] That is the conditional p.m.f. at infinite horizon.

obtain those of $B(h+1,i)(u)$. For instance, by (13) we get:

$$B(h+1,0)(u) = B(h,0)(u) + \lambda \sum_{k=1}^{n} \frac{B^k(h,1)(u)}{k!} + O(u^{n+1}),$$

whereas (14) leads to:

$$\begin{aligned} B(h+1,p)(u) &= \log(1-\alpha_p) + \log\left[1 + \frac{\alpha_p}{1-\alpha_p} e^{B(h,1)(u)}\right] \\ &= \log(1-\alpha_p) + \sum_{j=1}^{n} \frac{(-1)^{j-1}\alpha_p^j}{j(1-\alpha_p)^j} e^{jB(h,1)(u)} + O(u^{n+1}) \\ &= \log(1-\alpha_p) + \sum_{j=1}^{n} \frac{(-1)^{j-1}\alpha_p^j}{j(1-\alpha_p)^j} \sum_{k=0}^{n} \frac{j^k}{k!} B^k(h,1)(u) + O(u^{n+1}) \\ &= \log(1-\alpha_p) + \sum_{k=0}^{n} \frac{B^k(h,1)(u)}{k!} \sum_{j=1}^{n} \frac{(-1)^{j-1}\alpha_p^j}{(1-\alpha_p)^j} j^{k-1} + O(u^{n+1}) \end{aligned}$$

Then we apply Proposition 1 to get the $n-$th order Taylor's expansions of the successive powers of $B(h,1)(u)$, to get that of $B(h+1,i)(u)$, $i = 0, ..., p$.

- Then equation (10) can be re-arranged into:

$$\mathbb{E}[u^{X_{t+h+1}} | \underline{X_t}] = \exp\left[A_{h+1,0} + \sum_{j=1}^{n} A_{h+1,j} u^j + O(u^{n+1})\right] \tag{16}$$

where each $A_{h+1,j}, j = 0...n$ is linear in $X_t, ..., X_{t+1-p}$. Thus the Taylor's expansion of the RHS of equation (16) can be obtained by applying Proposition 1 and expanding the exponential function.

- Finally the values of $\mathbb{P}[X_{t+h+1} = j | \underline{X_{t-1}}]$ are obtained altogether for $j = 0, ..., n$, by a simple coefficient identification. In practice, we take n to be sufficiently large to ensure that the tail probability $\mathbb{P}[X_{t+h+1} \geq n+1 | \underline{X_{t-1}}]$ is negligible.

In terms of computational effort, the Taylor's expansions of $B(h,i)(u)$ do not depend on values of (X_t) and only need to be computed once when t changes. They necessitate a complexity of $O(n^3)$. Then for each iteration, the Taylor's expansion of (16), requires a complexity of $O(2n^3)$. This is much smaller than the Markov chain approach, which has a complexity[9] of $O(n^{(p+3)})$.

[9]Indeed, filling each row of the transition matrix has a complexity of $O(n^3)$ using the method described in

Let us now illustrate how our method fares against the Markov chain approach of McCabe et al. (2011) and the Monte-Carlo simulation approach of Jung and Tremayne (2006), using a pre-specified INAR(2) process.[10] The following table reports the execution time, as well as the conditional p.m.f. at horizons $h = 1, 5$ and 10 from the three methods, given observations up to time T.

Method	Markov chain approach			Our exact approach			Simulation approach		
Execution time	1 second			0.01 second			6 second		
horizon h	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
$p_h(0\|\underline{X_T})$	0.061	0.19	0.167	0.061	0.171	0.191	0.062	0.169	0.191
$p_h(1\|\underline{X_T})$	0.185	0.325	0.298	0.185	0.298	0.312	0.185	0.296	0.312
$p_h(2\|\underline{X_T})$	0.262	0.258	0.253	0.262	0.263	0.259	0.261	0.268	0.258
$p_h(3\|\underline{X_T})$	0.234	0.162	0.169	0.234	0.157	0.145	0.232	0.155	0.145
$p_h(8\|\underline{X_T})$	0.002	0.015	0.016	0.002	0.000	0.000	0.002	0.000	0.000
$p_h(9\|\underline{X_T})$	0.000	0.008	0.008	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: Conditional p.m.f. at different horizons obtained by different methods. For expository purpose we have only displayed their values at points $0, 1, 2, 3, 8$ and 9.

The reported execution times correspond to the calculation of $p_h(n|\underline{X_t})$ for horizons $h = 1, 2, ..., 20$ and $n = 20$. For the third approach, the number of simulated paths is equal to $N = 50000$. This spells a high computational cost, but is necessary to guarantee the forecasting precision. Indeed, in Table 2 the forecasts provided by our approach and the simulation approach are quite similar across different horizons, whereas in an unreported comparison where we take instead $N = 5000$ as in Jung and Tremayne (2006), we find that the relative error is around 4 percent at horizon 1. As for the Markov chain approximation approach, while at horizon 1, it provides reliable forecasts, this is no longer the case at higher horizons $h = 5, 10$. In other words the approximation error accumulates as h increases and this approach fails to well approximate the long-term behavior of the process, which echos the concerns we raised at the beginning of Section 4. We do not report the counterpart of this table for an INAR(5) model since *i*) for this latter model, the Markov chain approach is too costly for a PC *ii*) the comparison result between our approach and the simulation approach is similar to Table 2, with ours being both much faster and more precise.

Section 3. Since there are in total $(n+1)^p = O(n^p)$ non-zero entries, the total complexity is $O(n^{(p+3)})$.

[10] The parameter values of the model are set to be $\alpha_1 = \alpha_2 = 0.2$, $\lambda = 1$, with terminal values $X_T = 3, X_T = 5$. We are interested in the prediction of $X_{T+1}, ..., X_{T+h}$.

5 Concluding remarks

We have solved the open problem that in INAR(p) models, both the likelihood function and multi-step probabilistic forecasts "seem" to be computationally intractable. Our method is based on *i*) the simple relationship between the p.g.f. and p.m.f. for a count distribution; *ii*) the CaR property of the INAR(p) process. Our method eliminates the estimation bias due to the saddlepoint approximation of Pedeli et al. (2015), as well as the approximation (resp. simulation) error of McCabe et al. (2011) [resp. Jung and Tremayne (2006)] when it comes to forecasting.

Finally, while we have illustrated our methodology for the benchmark INAR(p) model with binomial thinning and Poisson innovation, the same technique can be applied to a large family of CaR models of the form:

$$X_t = \sum_{i=1}^{X_{t-1}} Z_{1,i,t+1} + \sum_{i=1}^{X_{t-2}} Z_{2,i,t+1} + \cdots + \sum_{i=1}^{X_{t-p}} Z_{p,i,t+1} + \epsilon_t \tag{17}$$

where $Z_{i,j,t}$ and ϵ_t are mutually independent, but not necessarily binary/Poisson distributed, respectively. In this case, for estimation (resp. forecasting), we need to replace, in eqn. (4) [resp. eqn (11)-(15)], functions $\log(\alpha_i u + 1 - \alpha_i)$, $i = 1, ..., p$ and/or $\lambda(u-1)$ by the log p.g.f. functions of $Z_{i,j,t}$, and ϵ_t. Thus for the conditional p.m.f. of model (17) to be computable, it suffices that the two new log p.g.f. are easily Taylor-expanded. For instance, Ristić et al. (2009); Gouriéroux and Lu (2017) assume $Z_{i,j,t}$ and/or ϵ_t to be negative binomial (or geometric) distributed to account for conditional over-dispersion,[11], Zhu and Joe (2010) introduce a family of INAR(1) models such that $Z_{1,t+1}$ and ϵ_t have Taylor-expandable p.g.f., Schweer and Weiß (2014) consider compound Poisson distributed innovations, whereas Drost et al. (2009) allow ϵ_t to have a flexible non-parametric distribution.

Another extention concerns bivariate INAR(p) models. The behchmark bivariate INAR(1) model [see e.g. Pedeli and Karlis (2013)] assumes that:

$$X_{1t} = \alpha_{11} \circ X_{1t-1} + \alpha_{12} \circ X_{2t-1} + \epsilon_{1t}$$

$$X_{2t} = \alpha_{11} \circ X_{1t-1} + \alpha_{12} \circ X_{2t-1} + \epsilon_{2t}$$

[11] The log p.g.f. of a negative binomial distribution with parameters $p < 1$ and $r > 0$ is equal to $r\log(1-p) - r\log(1-pu)$, which can be Taylor-expanded into $r\log(1-p) - r\sum_{i=1}^{n} \frac{1}{i} p^i u^i + O(u^{n+1})$.

where $\alpha_{i,j} \circ X_{1,t-1},\ i,j = 1,2$ are conditionally independent and binomial distributed, and $(\epsilon_{1t}, \epsilon_{2t})$ follow a bivariate Poisson distribution. They show that the conditional p.m.f. $p(X_{1t}, X_{2t}|X_{t-1})$ in this model involves a four dimensional summation, and argue that a higher-order generalization would have an intractable likelihood function. The methodology developed here seems adapted to this problem, since both the Taylor's expansion and the theory of CaR process are readily available in the bivariate case. This is left for future research.

Appendix: Proof of Proposition 2

Let us proceed by induction. Initial conditions (11) and (12) are consequences of Equation (4). Assume that (10) holds for a certain horizon h, then:

$$\begin{aligned}
&\mathbb{E}[u^{X_{t+h}}|\underline{X_{t-1}}] = \mathbb{E}\Big[\mathbb{E}[u^{X_{t+h}}|\underline{X_t}] \mid \underline{X_{t-1}}\Big] \\
&= \mathbb{E}[\exp\Big[B(h,0)(u) + \sum_{i=1}^{p} B(h,i)(u)X_{t+1-i}\Big]|\underline{X_{t-1}}] \\
&= \exp\Big[B(h,0)(u) + \sum_{i=2}^{p} B(h,i)(u)X_{t+1-i}\Big]\mathbb{E}\Big[\exp\Big(B(h,1)(u)X_t\Big)|\underline{X_{t-1}}\Big] \qquad (18) \\
&= \exp\Big[B(h,0)(u) + \sum_{i=2}^{p} B(h,i)(u)X_{t+1-i} + \lambda[e^{B(h,1)(u)} - 1] + \sum_{i=1}^{p} \log(\alpha_i e^{B(h,1)(u)} + 1 - \alpha_i)X_{t-i}\Big]. \qquad (19)
\end{aligned}$$

where in (18) we have used the conditional moment generating function $\mathbb{E}[e^{tX_t}|\underline{X_{t-1}}]$. Its expression is obtained by replacing u by e^t in (4). Thus we recover the RHS of (10) and (10) holds for any h.

References

Al-Osh, M. and Alzaid, A. A. (1987). First-order Integer-valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis*, 8(3):261–275.

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and Saddle-point Approximations with

Statistical Applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 279–312.

Bu, R. and McCabe, B. (2008). Model Selection, Estimation and Forecasting in INAR (p) Models: A Likelihood-based Markov Chain Approach. *International Journal of Forecasting*, 24(1):151–162.

Bu, R., McCabe, B., and Hadri, K. (2008). Maximum Likelihood Estimation of Higher-Order Integer-Valued Autoregressive Processes. *Journal of Time Series Analysis*, 29(6):973–994.

Darolles, S., Gourieroux, C., and Jasiak, J. (2006). Structural Laplace Transform and Compound Autoregressive Models. *Journal of Time Series Analysis*, 27(4):477–503.

Davies, R. B. (1973). Numerical Inversion of a Characteristic Function. *Biometrika*, 60(2):415–417.

Drost, F. C., Akker, R. v. d., and Werker, B. J. (2009). Efficient Estimation of Auto-Regression Parameters and Innovation Distributions for Semiparametric Integer-Valued AR (p) models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):467–485.

Du, J. and Li, Y. (1991). The Integer-Valued Autoregressive (INAR (p)) Model. *Journal of Time Series Analysis*, 12(2):129–142.

Gibson, D. and Seneta, E. (1987). Augmented Truncations of Infinite Stochastic Matrices. *Journal of Applied Probability*, 24(3):600–608.

Gouriéroux, C. and Lu, Y. (2017). Negative Binomial Autoregressive Process. *CREST DP*.

Jung, R. C. and Tremayne, A. (2006). Coherent Forecasting in Integer Time Series Models. *International Journal of Forecasting*, 22(2):223–238.

Knuth, D. E. (1997). *The Art of Computer Programming*, volume 3. Pearson Education.

McCabe, B. P., Martin, G. M., and Harris, D. (2011). Efficient Probabilistic Forecasts for Counts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):253–272.

McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Journal of the American Water Resources Association*, 21(4):645–650.

Pedeli, X., Davison, A. C., and Fokianos, K. (2015). Likelihood Estimation for the INAR (p) Model by Saddlepoint Approximation. *Journal of the American Statistical Association*, 110(511):1229–1238.

Pedeli, X. and Karlis, D. (2013). Some Properties of Multivariate INAR (1) Processes. *Computational Statistics & Data Analysis*, 67:213–225.

Ristić, M. M., Bakouch, H. S., and Nastić, A. S. (2009). A New Geometric First-order Integer-valued Autoregressive (NGINAR (1)) Process. *Journal of Statistical Planning and Inference*, 139(7):2218–2226.

Schweer, S. and Weiß, C. H. (2014). Compound Poisson INAR (1) Processes: Stochastic Properties and Testing for Overdispersion. *Computational Statistics & Data Analysis*, 77:267–284.

Tweedie, R. (1998). Truncation Approximations of Invariant Measures for Markov Chains. *Journal of Applied Probability*, 35(3):517–536.

Zhu, R. and Joe, H. (2010). Negative Binomial Time Series Models Based on Expectation Thinning Operators. *Journal of Statistical Planning and Inference*, 140(7):1874–1888.