Integer-valued autoregressive models based on quasi Pólya thinning operator

Jean Peyhardi

▶ **To cite this version:**

Jean Peyhardi. Integer-valued autoregressive models based on quasi Pólya thinning operator. 2024. hal-04691912

HAL Id: hal-04691912
https://hal.science/hal-04691912v1

Preprint submitted on 11 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Integer-valued autoregressive models based on quasi Pólya thinning operator

Jean Peyhardi[1]

[1]IMAG, Univeristé de Montpellier, CNRS, Montpellier, 34090, France.

Contributing authors: jean.peyhardi@umontpellier.fr;

Abstract

Autoregressive models adapted to count time series have received less attention than their classical counterparts for continuous time series. The main approach is based on thinning stochastic operation that preserves the discrete nature of the variable between successive times. The binomial thinning operator is the most popular and the Poisson distribution emerges as the natural choice for the error distribution of such an autoregressive counting process. The present paper introduces the quasi Pólya thinning operator, that includes the binomial thinning operator as a special case. The family of additive modified power series distribution is defined and is shown to be the natural choice for the error distribution of such a counting process. We obtain the most general class of integer-valued autoregressive models of order 1 (INAR(1) models) with margins having analytic form and the property of closure under convolution introduced by Joe (1996). It includes the usual cases of Poisson and generalized Poisson margins, but also the less usual cases of binomial and negative binomial margins and the new case of generalized negative binomial margin. These models cover a high range of dispersion that are strictly ordered from the binomial case to the generalized negative binomial case. Asymptotic normality of the maximum likelihood estimator (MLE) for such INAR(1) models is obtained. Finally, the proposed INAR(1) models are applied on simulated and real datasets.

Keywords: Binomial thinning operator, Quasi Pólya distribution, INAR model, modified power series distribution

1 Introduction

A popular class of models for count time series is the class of INAR(1) models based on thinning operator. A thinning operator is a stochastic operator that shrinks a random count variable into a smaller one. This kind of operator is thus useful to define X_t from X_{t-1}, ensuring that X_t is still an integer. The most popular thinning operator is the binomial thinning operator, usually defined as follows. Let X denote a random count variable and $\rho \in (0,1)$ a given parameter, then $X' \overset{d}{=} \rho \circ X$ denotes the thinned version of X. It is defined as the random sum $X' = \sum_{i=1}^{X} Z_i$ where $Z_i \sim \mathcal{B}(\rho)$ are independent and identically distributed random Bernoulli variables with parameter ρ. An INAR(1) model based on binomial thinning operator is defined by the equation

$$X_t = \rho \circ X_{t-1} + \varepsilon_t, \tag{1}$$

where the residual ε_t is assumed to be independent of $(X_s)_{s \leq t-1}$ and all the residuals $(\varepsilon_t)_{t \in \mathbb{N}}$ are independent and identically distributed according to the distribution $\mathcal{L}_\varepsilon$. The most natural choice of error distribution $\mathcal{L}_\varepsilon$ associated with the binomial thinning operator is the Poisson distribution, since it is closed under this thinning operation and under convolution. It implies that equation (1) defines a stationary process with Poisson marginal distribution, as shown by McKenzie (1985). But count time series often present over dispersion and the Poisson distribution becomes inappropriate. In the case of negative binomial margin, McKenzie (1987) explicitly derived an expression for the error distribution, but for other margins it becomes more difficult.

It is possible to use other thinning operator to obtain the desired marginal distribution. Several extensions of the binomial thinning operator have been proposed in the literature. Latour (1998) introduced a generalized thinning operator by relaxing the assumption of Bernoulli distribution for the latent variables Z_i. For instance, the specific case of geometric distribution $Z_i \sim \mathcal{G}(\rho)$, has been studied by Ristić et al (2009) leading to the negative binomial operator denoted by $\rho *$. But assuming a full range for Z_i in $\mathbb{N}$ instead of $\{0,1\}$ implies that the constraint $\rho * X \leq X$ does not hold anymore, and the operation $\rho \circ$ could not been viewed as a thinning operation. Another extension of the binomial thinning operator is the random coefficient thinning operator studied by Zheng et al (2007), where the variables Z_i given ρ are assumed to follow independent Bernoulli $\mathcal{B}(\rho)$ and ρ is assumed to be a random variable. Assume for instance a beta distribution for ρ leads to the beta binomial thinning operator. It could be shown that the natural choice for the error distribution in this case is the negative binomial distribution (Joe, 1996).

The present paper proposes to view the binomial thinning operation as a compound distribution operation, i.e., $\rho \circ X$ given $X = n$ follows a binomial distribution, denoted by $\mathcal{B}_n(\rho)$. In other words, if $\mathcal{L}$ denotes the distribution of X then $\rho \circ X$ follows the compound distribution $\mathcal{B}_n(\rho) \underset{n}{\wedge} \mathcal{L}$, where the notation $\underset{n}{\wedge}\mathcal{L}$ means that the parameter n is considered as a random variable following the distribution $\mathcal{L}$. It is then possible to replace the binomial distribution by another distribution supported on $\{0, \ldots, n\}$, such as the hypergeometric distribution for instance or the quasi binomial distribution. These two examples have been studied by Al-Osh and Alzaid (1991) and Alzaid and

Al-Osh (1993) respectively. They have shown that if a binomial (resp. generalized Poisson) distribution is assumed for the residuals, then the process is stationary with the same distribution. These models, such as the binomial thinning with Poisson margin or the beta binomial thinning with negative binomial margin, fall into the class of INAR(1) models with margins in the convolution-closed infinitely divisible class described by Joe (1996). Another INAR(1) model with negative binomial margin and negative binomial error distribution has been proposed by Al-Osh and Aly (1992), using an operator based on a sum of geometrically distributed variables with a number of terms given by a binomial thinning operation.

In fact, the hypergeometric, binomial and beta binomial thinning operators have previously been studied in the framework of damage models; see (Patil and Ratnaparkhi, 1975) for a review. In this framework, a random variable N is split into two random variables Y_1 and Y_2, i.e., such that $N = Y_1+Y_2$. The conditional distribution of Y_1 given $N = n$ is called damage distribution, e.g., the binomial distribution. The variable Y_1 can be viewed as a damaged version of the original variable N. Two properties of such a model have been intensively studied: the closure under damage operation and the characterization of the original distribution of N through the condition of independence between Y_1 and Y_2. For instance, Rao (1965) showed that binomial, Poisson and negative binomial distribution are closed under the binomial thinning operation. It was generalized by Janardan and Rao (1986) which demonstrated the closure of generalized Pólya Eggenberger distributions under quasi Pólya thinning operation.

Based on this closure property, we propose a general class of INAR(1) models using the quasi Pólya distribution as thinning operator; see (Peyhardi, 2023) for a study of this stochastic operator. The present paper relates the closure properties under thinning operator to the marginal distribution of the stationary INAR(1) process. The proposed class of INAR(1) models includes the cases of binomial and quasi binomial thinning operators, already studied, the less usual cases of hypergeometric and beta binomial thinning operators and the new case of quasi beta binomial thinning operator. All properties of these models will be studied in a general way based on the class of quasi Pólya distributions. For instance, it will be shown that the natural choice for the error distribution is the class of additive modified power series distribution that lead to a stationary process with the same marginal distribution. Recently, another general approach was proposed by Sellers et al (2020) based on the Conway-Maxwell-Binomial (CMB) thinning operator with Conway-Maxwell-Poisson (CMP) as marginal distribution. It includes the three cases of hypergeometric, binomial and beta binomial thinning operators. But the first case is obtained as a limit, not really as a subcase. Moreover, the CMB and CMP distributions do not have closed form and thus need the computation of a normalizing constant. The present paper introduces the most general class of INAR(1) models with margins having analytic form and the property of closure under convolution introduced by Joe (1996).

The paper is organized as follows. Section 2 introduces some preliminary notations and definitions about the class of quasi Pólya distributions and its counterpart, the class of additive modified power series distributions. Section 3 presents the quasi Pólya thinning operator and its properties. Section 4 defines INAR(1) models based on quasi Pólya thinning operator. It focuses on the natural choice for the error distribution that

is preserved at stationary state for the marginal distribution. The five specific INAR(1) models based on quasi Pólya thinning operator are described and the standardized Pearson residuals are given in order to evaluate the model adequacy. Section 5 gives regularity conditions under which the asymptotic normality of the MLE is obtained. Finally, Section 6 presents some applied results of INAR(1) models on simulated and real datasets.

2 Preliminary notations and definitions

Let us introduce the parametric sequence

$$a_\theta^{[c,d]}(n) = \frac{\theta(\theta+dn)^{(n;c)}}{(\theta+dn)n!}, \tag{2}$$

where $n \in \mathbb{N}$, $\theta \in \Theta$ (with $\Theta = \mathbb{N}^*$ when $c = -1$, $d = 0$ and $\Theta = \mathbb{R}_+^*$ otherwise), $c \in \{-1, 0, 1\}$, $d \in \mathbb{R}_+$ and $m^{(n;c)} := \prod_{k=0}^{n-1}(m+ck)$, where $\mathbb{N}^*$ (resp. $\mathbb{R}_+^*$) denotes the set of positive integers (resp. positive reals). Let $\binom{\theta}{n}$ denote the binomial coefficient using the gamma function, i.e., $\binom{n+\theta-1}{n} = \frac{\Gamma(n+\theta)}{\Gamma(n+1)\Gamma(\theta)}$ that is well-defined for continuous values of $\theta \in \mathbb{R}_+^*$. One can distinguish five equivalence classes for the parametric sequence $a_\theta^{[c,d]}(n)$ according to the (c, d) values, with the following representative elements:

1. $a_\theta^{[-1,0]}(n) = \binom{\theta}{n}$
2. $a_\theta^{[0,0]}(n) = \theta^n/n!$
3. $a_\theta^{[1,0]}(n) = \binom{n+\theta-1}{n}$
4. $a_\theta^{[0,d]}(n) = \frac{\theta(\theta+dn)^{n-1}}{n!}$ with $d > 0$
5. $a_\theta^{[1,d]}(n) = \frac{\theta}{\theta+dn}\binom{\theta+dn+n-1}{n}$ with $d > 0$

Remark that the second case is included in the fourth case if $d = 0$ and the first and third cases are included in the fifth case if $d = -1$ and $d = 0$ respectively. The case $c = -1$ and $d \neq 0$ is omitted since $a_\theta^{[-1,d+1]}(n) = a_\theta^{[1,d]}(n)$ for all $d \geq -1$. In each case, it can be shown that the parametric sequence $a_\theta^{[c,d]}(n)$ respects the two following properties:

(P1) Additive closure under convolution:

$$a_\theta^{[c,d]} * a_\gamma^{[c,d]} = a_{\theta+\gamma}^{[c,d]}$$

(P2) Modified power series convergence: there exists a function g such that

$$\sum_{k=0}^{\infty} a_\theta^{[c,d]}(k)\{g(\alpha)\}^k = h_\theta(\alpha)$$

for all $\theta \in \Theta$ and $\alpha \in (0, R)$. Note that R is the radius of convergence of the power series when g is the identity function.

The property (P1) is exactly the convolution identity of Abel when $c = 0$ and those of Hagen-Rothe when $c \neq 0$; see (Chu, 2010) for demonstration. The symbol $*$ here denotes the convolution of sequences, i.e., $(a_\theta * a_\gamma)(n) = \sum_{k=0}^{n} a_\theta(k)a_\gamma(n-k)$.The property (P2) corresponds to usual power series when $c = 0$ and their Lagrangian expansion when $c \neq 0$: see (Consul and Famoye, 2006) for the demonstration of the case $c \neq 0$ and see Table A2 in Appendix A for details about the function g and the normalizing constant h in the five cases. These properties allow defining the following two classes of distributions.

Definition 1. *A discrete non-negative variable X is said to follow a quasi Pólya distribution if its probability mass function (pmf) is given by*

$$P(X = k) = \frac{a_\theta^{[c,d]}(k)a_\gamma^{[c,d]}(n-k)}{a_{\theta+\gamma}^{[c,d]}(n)}\mathbb{1}_{k\leq n},$$

for some $(\theta, \gamma) \in \Theta^2$ and $n \in \mathbb{N}$. This distribution will be denoted by $q\mathcal{P}_n^{[c,d]}(\theta, \gamma)$.

Property (P1) ensures that we obtain a proper distribution. This class of distributions includes for instance the hypergeometric distribution when $c = -1$ and $d = 0$, or the binomial distribution when $c = 0$ and $d = 0$. More generally, this is the family of quasi Pólya distributions, introduced by Janardan (1973) as urn model with predetermined strategies. The different examples of quasi Pólya distributions are summarized in Table 1 and more details are given in Table A1 of Appendix A. The moments of quasi Pólya distributions are recalled in the following Lemma.

Lemma 1. *Let Y follow a quasi Pólya distribution $q\mathcal{P}_n^{[c,d]}(\theta, \gamma)$ and $Z = n - Y$, then*

- $E[Y] = \dfrac{\theta}{\theta+\gamma}n$
- $V(Y) = \dfrac{\theta\gamma}{\theta+\gamma}\left[\dfrac{n^2}{\theta+\gamma} - \displaystyle\sum_{j=1}^{n-1}\frac{n^{(j+1;-1)}(d+c)^{j-1}}{\{\theta+\gamma+dx_{t-1}+(n-j+2)c\}^{(j;c)}}\right]$
- *If $d = 0$, Y follows a simple Pólya distribution $\mathcal{P}_n^{[c]}(\theta, \gamma)$, then*

$$V(Y) = \frac{\theta\gamma}{(\theta+\gamma)(\theta+\gamma+c)}n + c\frac{\theta\gamma}{(\theta+\gamma)^2(\theta+\gamma+c)}n^2$$
$$Cov(Y, Z) = -\frac{\theta\gamma}{(\theta+\gamma)(\theta+\gamma+c)}n - c\frac{\theta\gamma}{(\theta+\gamma)^2(\theta+\gamma+c)}n^2$$

The two first points of this Lemma have been shown by Janardan and Schaeffer (1977) and the third by Johnson et al (1997).

Definition 2. *A discrete non-negative variable X is said to follow an additive modified power series distribution if its pmf is given by*

$$P(X = k) = \frac{a_\theta^{[c,d]}(k)\{g(\alpha)\}^k}{h_\theta(\alpha)}$$

for some $\theta \in \Theta$ and $\alpha \in (0, R)$. This distribution will be denoted by $\mathcal{APS}^{[c,d]}(\theta, \alpha)$.

	$d=0$	$d>0$
$c<0$	hypergeometric $\mathcal{H}_n(\theta,\gamma)$	quasi hypergeometric $q\mathcal{H}_n^{[d]}(\theta,\gamma)$
$c=0$	binomial $\mathcal{B}_n(\rho)$	quasi binomial $q\mathcal{B}_n^{[d]}(\rho)$
$c>0$	beta binomial $\beta\mathcal{B}_n(\theta,\gamma)$	quasi beta binomial $q\beta\mathcal{B}_n^{[d']}(\theta,\gamma)$

Table 1 Equivalence classes of quasi Pólya distributions $q\mathcal{P}_n^{[c,d]}(\theta,\gamma)$ according to (c,d) values. The cell with quasi hypergeometric distribution is greyed out since it is equivalent to the quasi beta binomial distribution, more precisely we have $q\mathcal{H}_n^{[d]}(\theta,\gamma) = q\beta\mathcal{B}_n^{[d-1]}(\theta,\gamma)$. Note that $\rho = \frac{\theta}{\theta+\gamma}$ and $d' = \frac{d}{\theta+\gamma}$.

Property (P2) ensures that we obtain a proper distribution. This is a subclass of the modified power series distributions introduced by Gupta (1974). It includes for instance the binomial distribution when $c=-1$ and $d=0$, or the Poisson distribution when $c=0$ and $d=0$. The different examples of additive modified power series distributions are summarized in Table 2 and more details are given in Table A2 of Appendix A. The mean and variance of such distributions are expressed in the following Lemma.

Lemma 2. *Let $X \sim \mathcal{APS}^{[c,d]}(\theta,\alpha)$ then there exists two functions f_1 and f_2 such that the expectation and variance of X can be decomposed as follows:*

$$E[X] = \theta f_1(\alpha),$$
$$V(X) = \theta f_2(\alpha).$$

Proof. Let $X \sim \mathcal{APS}^{[c,d]}(\theta,\alpha)$. According to Gupta (1974) we have

$$\mu = \mathrm{E}[X] = \frac{g(\alpha)}{g'(\alpha)}\frac{h'_\theta(\alpha)}{h_\theta(\alpha)},$$
$$\mathrm{V}(X) = \frac{g(\alpha)}{g'(\alpha)}\mu'(\alpha),$$

where g and h are given in Table A2 of Appendix A. It is easily seen that

$$\frac{h'_\theta(\alpha)}{h_\theta(\alpha)} = \begin{cases} \frac{\theta}{1+\alpha} & \text{if } c=-1 \\ \theta & \text{if } c=0 \\ \frac{\theta}{1-\alpha} & \text{if } c=1 \end{cases}$$

and thus with

$$f_1(\alpha) := \begin{cases} \frac{g(\alpha)}{g'(\alpha)(1+\alpha)} & \text{if } c=-1 \\ \frac{g(\alpha)}{g'(\alpha)} & \text{if } c=0 \\ \frac{g(\alpha)}{g'(\alpha)(1-\alpha)} & \text{if } c=1 \end{cases}$$

and $f_2(\alpha) := \frac{g(\alpha)}{g'(\alpha)} f'_1(\alpha)$ we obtain $\mathrm{E}[X] = \theta f_1(\alpha)$ and $\mathrm{V}(X) = \theta f_2(\alpha)$. □

	$d=0$	$d>0$
$c<0$	binomial $\mathcal{B}_\theta(p)$	generalized binomial $\mathcal{GB}^{[d]}(\theta,p)$
$c=0$	Poisson $\mathcal{P}(\lambda)$	generalized Poisson $\mathcal{GP}^{[d]}(\lambda)$
$c>0$	negative binomial $\mathcal{NB}(\theta,\alpha)$	generalized negative binomial $\mathcal{GNB}^{[d]}(\theta,\alpha)$

Table 2 Equivalence classes of additive modified power series distributions $\mathcal{APS}^{[c,d]}(\theta,\alpha)$ according to (c,d) values. The cell generalized binomial distribution $\mathcal{GB}^{[d]}(\theta,p)$ is greyed out, since it is equivalent to the generalized negative binomial distribution $\mathcal{GNB}^{[d-1]}(\theta,p)$. Note that $p=\frac{\alpha}{1+\alpha}$ and$\lambda=\theta\alpha$.

It should be remarked that both distributions (definitions 1 and 2) are related by the same parametric sequence $a_\theta^{[c,d]}(n)$ that respects properties (P1) and (P2). This is the key point in the demonstration of the following Lemma.

Lemma 3. *An additive modified power series distribution is additively closed under convolution, more precisely we have*

$$\mathcal{APS}^{[c,d]}(\theta,\alpha) * \mathcal{APS}^{[c,d]}(\gamma,\alpha) = \mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha),$$

for all $(\theta,\gamma)\in\Theta^2$, $\alpha\in(0,R)$, $c\in\{-1,0,1\}$ *and* $d\in\mathbb{R}_+$ *(the symbol* $*$ *here denotes the convolution of distributions).*

The proof of this Lemma is given by Peyhardi (2023) in Proposition 1.

3 Quasi Pólya thinning operator

Let us first recall the definition of the classical binomial thinning operator. Let X denote a count variable and $\rho\in(0,1)$ a parameter, then $X'\overset{d}{=}\rho\circ X$ denote the thinned variable. It is usually defined as a random sum $X'=\sum_{i=1}^{X}Z_i$ where $Z_i\sim\mathcal{B}(\rho)$ are independent random Bernoulli variables with parameter ρ. Two kinds of extension have been proposed: assume that ρ is a random variable, or assume that Z_i follow another count distribution than the Bernoulli.

To propose a new extension, we need to see the binomial thinning as a binomial mixed distribution. Remark that $X'\overset{d}{=}\rho\circ X$ is equivalent to $X'\sim\mathcal{B}_n(\rho)\underset{n}{\wedge}\mathcal{L}$ where $\mathcal{L}$ denotes the distribution of X. It is thus easy to extend the binomial thinning to the quasi Pólya thinning as follows

$$X'\sim q\mathcal{P}_n^{[c,d]}(\theta,\gamma)\underset{n}{\wedge}\mathcal{L}.$$

Note that with this parametrization, ρ corresponds to the proportion $\theta/(\theta+\gamma)$. In the following, let $\rho\bullet$ denote the quasi Pólya thinning operator, i.e., $\rho\bullet X=X'$ with $X'|X=n\sim q\mathcal{P}_n^{[c,d]}(\theta,\gamma)$ for all $n\in\mathbb{N}$. This new class of thinning operators includes five distinct examples; see Table 1 for notations.

Let us briefly recall the interpretation of (quasi) Pólya distribution in terms of urn models. The Pólya urn model is generally presented in terms of n random drawings of balls from an urn, that initially contains θ white balls and γ black balls. One ball is drawn at random and then replaced with c additional balls of the same color. This procedure is repeated n times and focus is made on the count X' of drawn white balls. Knowing the number n of draws, the count distribution for X' is known as the Pólya distribution, i.e., we have $X'|X = n \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$. In this framework, the parameter $\rho = \theta/(\theta + \gamma)$ corresponds to the proportion of white balls initially contained in the urn. The parameter n being related to the support of the distribution, is intentionally noted as an index of the distribution, distinguishing it from other parameters θ and γ used to define the pmf. The Pólya distribution turns out to be the hypergeometric distribution when $c = -1$ (without replacement), the binomial distribution when $c = 0$ (with replacement meaning independent draws) and the negative hypergeometric distribution when $c = 1$ (with reinforcement). Let us note that the last two cases are defined for continuous values of θ and γ. The negative hypergeometric distribution is thereby extended to the beta binomial distribution. Finally, the parameter d is added to take into account a predetermined strategy, leading to the quasi Pólya distribution $q\mathcal{P}_n^{[c,d]}(\theta, \gamma)$; see Consul and Mittal (1975) and Janardan (1973) for more details. The quasi Pólya distribution turns out to the Pólya distribution when $d = 0$.

Proposition 1. *Let X denote an arbitrary random count variable, then*

- $E[\rho \bullet X] = \rho E[X]$
- *If $d = 0$, $\rho \bullet$ denotes the simple Pólya thinning operator $\mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{L}$, then*

$$V(\rho \bullet X) = \rho^2 V(X) + \rho(1-\rho)\frac{\theta+\gamma}{\theta+\gamma+c}E[X] + c\rho(1-\rho)\frac{1}{\theta+\gamma+c}E[X^2]$$
$$Cov(\rho \bullet X, X) = \rho V(X)$$

where $\rho = \theta/(\theta + \gamma)$.

Proof. Let $Y = \rho \bullet X$ and $Z = X - Y$. Next equalities are due to the total law of mean (variance and covariance) and moments of (quasi) Pólya given in Lemma 1.

$$\mathrm{E}[\rho \bullet X] = \mathrm{E}[\mathrm{E}[Y|X]] = \mathrm{E}[\rho X] = \rho \mathrm{E}[X]$$

Assuming $d = 0$ we have

$$\begin{aligned}
\mathrm{V}(\rho \bullet X) &= \mathrm{V}(\mathrm{E}[Y|X]) + \mathrm{E}[\mathrm{V}(Y|X)] \\
&= \mathrm{V}(\rho X) + \mathrm{E}\left[\frac{\theta\gamma}{(\theta+\gamma)(\theta+\gamma+c)}X + c\frac{\theta\gamma}{(\theta+\gamma)^2(\theta+\gamma+c)}X^2\right] \\
\mathrm{V}(\rho \bullet X) &= \rho^2 V(X) + \rho(1-\rho)\frac{\theta+\gamma}{\theta+\gamma+c}\mathrm{E}[X] + c\rho(1-\rho)\frac{1}{\theta+\gamma+c}\mathrm{E}[X^2]
\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(\rho \bullet X, X) &= \mathrm{Cov}(\mathrm{E}[Y|X], \mathrm{E}[X|X]) + \mathrm{E}\left[\mathrm{Cov}(Y, X|X)\right] \\ &= \mathrm{Cov}(\rho X, X) + \mathrm{E}\left[\mathrm{Cov}(Y, Y+Z|X)\right] \\ &= \rho \mathrm{V}(X) + \mathrm{E}\left[\mathrm{V}(Y|X) + \mathrm{Cov}(Y, Z|X)\right] \\ \mathrm{Cov}(\rho \bullet X, X) &= \rho \mathrm{V}(X)\end{aligned}$$

remarking that $\mathrm{V}(Y|X) = -\mathrm{Cov}(Y, Z|X)$ according to Lemma 1. □

These results coincide with those of Weiß (2008b) about the hypergeometric thinning operator ($c = -1$ and $d = 0$), the binomial thinning operator ($c = 0$ and $d = 0$) and the beta binomial thinning operator ($c = 1$ and $d = 0$).

Now, let us describe different properties of quasi Pólya thinning operator and additive modified power series distributions, that will be used in next sections to demonstrate the stationarity of INAR(1) model based on quasi Pólya thinning operator.

Lemma 4. *An additive modified power series distribution is closed under the corresponding quasi Pólya thinning operator if the parameters respect the sum constraint, more precisely we have*

$$q\mathcal{P}_n^{[c,d]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{APS}^{[c,d]}(\theta + \gamma, \alpha) = \mathcal{APS}^{[c,d]}(\theta, \alpha),$$

for all $(\theta, \gamma) \in \Theta^2$, $\alpha \in (0, R)$, $c \in \{-1, 0, 1\}$ *and* $d \in \mathbb{R}_+$.

The proof of this Lemma is given by Peyhardi (2023) in Theorem 1 (item 2).

Lemma 5. *Let* X, Y *and* Z *denote three random count variables, then the following assertions are equivalent*

$$\begin{cases} X \sim \mathcal{APS}^{[c,d]}(\theta + \gamma, \alpha) \\ Y = \rho \bullet X \\ Z = X - Y \end{cases} \Leftrightarrow \begin{cases} Y \sim \mathcal{APS}^{[c,d]}(\theta, \alpha) \\ Z \sim \mathcal{APS}^{[c,d]}(\gamma, \alpha) \\ Y \perp\!\!\!\perp Z \\ X = Y + Z \end{cases}$$

Proof. • Assume that the left part holds. According to Lemma 4 we have $Y \sim \mathcal{APS}^{[c,d]}(\theta, \alpha)$. Since $Y|X = n \sim q\mathcal{P}_n^{[c,d]}(\theta, \gamma)$ we have $n - Y|X = n \sim q\mathcal{P}_n^{[c,d]}(\gamma, \theta)$ and thus $Z = X - Y \sim q\mathcal{P}_n^{[c,d]}(\gamma, \theta) \underset{n}{\wedge} \mathcal{APS}^{[c,d]}(\theta + \gamma, \alpha)$. Using Lemma 4 we obtain $Z \sim \mathcal{APS}^{[c,d]}(\gamma, \alpha)$. Finally, the joint distribution of Y and Z is given by

$$\begin{aligned} P(Y = y, Z = z) &= P(Y = y, Z = z|X = y + z)P(X = y + z) \\ &= \frac{a_\theta^{[c,d]}(y) a_\gamma(z)}{a_{\theta+\gamma}^{[c,d]}(y + z)} \frac{a_{\theta+\gamma}^{[c,d]}(y + z)\{g(\alpha)\}^{y+z}}{h_{\theta+\gamma}(\alpha)} \\ P(Y = y, Z = z) &= \frac{a_\theta^{[c,d]}(y)\{g(\alpha)\}^y a_\gamma^{[c,d]}(z)\{g(\alpha)\}^z}{h_{\theta+\gamma}(\alpha)}, \end{aligned}$$

We know that $h_{\theta+\gamma} = h_\theta h_\gamma$ and thus Y and Z are independent.

- Assume that the right part holds. According to Lemma 3 we have $X \sim \mathcal{APS}^{[c,d]}(\theta+\gamma, \alpha)$. Moreover, we have

$$\begin{aligned} P(Y=y|X=n) &= \frac{P(Y=y, X=n)}{P(X=n)} \\ &= \frac{P(Y=y)P(Z=n-y)}{P(X=n)} \\ &= \frac{a_\theta^{[c,d]}(y)\{g(\alpha)\}^y}{h_\theta(\alpha)} \frac{a_\gamma^{[c,d]}(n-y)\{g(\alpha)\}^{n-y}}{h_\gamma(\alpha)} \frac{h_{\theta+\gamma}(\alpha)}{a_{\theta+\gamma}^{[c,d]}(n)\{g(\alpha)\}^n} \\ P(Y=y|X=n) &= \frac{a_\theta^{[c,d]}(y)a_\gamma^{[c,d]}(n-y)}{a_{\theta+\gamma}^{[c,d]}(n)}, \end{aligned}$$

  i.e., $Y|X=n \sim q\mathcal{P}_n^{[c,d]}(\theta,\gamma)$ or equivalently $Y = \rho \bullet X$.

□

Lemma 6. *If* $X \sim \mathcal{APS}^{[c,d]}(\theta+\gamma, \alpha)$ *then*

$$\begin{aligned} E[\rho \bullet X] &= \rho E[X], \\ V(\rho \bullet X) &= \rho V(X), \\ Cov(\rho \bullet X, X) &= \rho V(X), \end{aligned}$$

where $\rho = \theta/(\theta+\gamma)$.

Proof. According to Lemma 4 we have $\rho \bullet X \sim \mathcal{APS}^{[c,d]}(\theta, \alpha)$. Then according to Lemma 2 we have

$$\begin{aligned} \mathrm{E}[\rho \bullet X] &= \theta f_1(\alpha) \\ &= \frac{\theta}{\theta+\gamma}(\theta+\gamma) f_1(\alpha) \\ \mathrm{E}[\rho \bullet X] &= \rho \mathrm{E}(X), \end{aligned}$$

where $\rho = \frac{\theta}{\theta+\gamma}$. In the same way, we obtain $\mathrm{V}(\rho \bullet X) = \rho \mathrm{V}(X)$. Using Lemma 5 we have

$$\begin{aligned} \mathrm{Cov}(\rho \bullet X, X) &= \mathrm{Cov}(Y, Y+Z) \\ &= \mathrm{V}(Y) + \underbrace{\mathrm{Cov}(Y,Z)}_{=0} \\ \mathrm{Cov}(\rho \bullet X, X) &= \rho \mathrm{V}(X). \end{aligned}$$

□

4 Integer valued autoregressive models of order 1

Let $(X_t)_{t\in\mathbb{N}}$ denote a discrete time process with discrete state space. An INAR(1) model based on quasi Pólya thinning operator is defined as follows:

$$X_t = \rho \bullet X_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{L}_\varepsilon$ are independent and identically distributed random variables and $\rho\bullet$ denotes a quasi Pólya thinning operator (with parameters $\boldsymbol{\psi} = (c, d, \theta, \gamma)$) independent of $(\varepsilon_t)_{t\in\mathbb{N}}$. Denoting by $\mathcal{L}_t$ the distribution of X_t we have

$$\mathcal{L}_t = \left\{ q\mathcal{P}_n^{[c,d]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{L}_{t-1} \right\} * \mathcal{L}_\varepsilon,$$

where $*$ denotes the convolution of distributions.

Theorem 2. *Let $(X_t)_{t\in\mathbb{N}}$ follow an INAR(1) model based on quasi Pólya thinning operator. Assume that the error distribution is the specific additive modified power series distribution $\mathcal{L}_\varepsilon = \mathcal{APS}^{[c,d]}(\gamma,\alpha)$ for some $\alpha \in (0, R)$. Then $(X_t)_{t\in\mathbb{N}}$ is an ergodic Markov chain with stationary distribution $\mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha)$. Moreover, at stationary state we have*

$$\begin{aligned} E[X_t] &= \frac{\mu_\varepsilon}{1-\rho}, \\ V(X_t) &= \frac{\sigma_\varepsilon^2}{1-\rho}, \\ Corr(X_{t+h}, X_t) &= \rho^h, \end{aligned}$$

where $\mu_\varepsilon = E(\varepsilon_t)$, $\sigma_\varepsilon^2 = V(\varepsilon_t)$ and $\rho = \theta/(\theta+\gamma)$. Finally, the process $(X_t)_{t\in\mathbb{N}}$ is time reversible.

Proof. The conditional distribution of X_{t+1} given $X_t = x_t$ is

$$p(x_{t+1}|x_t) = \sum_{k=0}^{\min(x_{t+1},x_t)} \frac{a_\theta^{[c,d]}(k) a_\gamma^{[c,d]}(x_t-k)}{a_{\theta+\gamma}^{[c,d]}(x_t)} \frac{a_\gamma^{[c,d]}(x_{t+1}-k)\{g(\alpha)\}^{x_{t+1}-k}}{h_\gamma(\alpha)} \tag{3}$$

The support of $p(\cdot|x)$ is related to the support of $a_\theta^{[c,d]}(\cdot)$. A necessary condition for the ergodicity is to have $p(x+1|x) > 0$ and $p(x|x+1) > 0$ for all x such that x and $x+1$ belong to the support of the distribution $\mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha)$. For the four cases such that $(c,d) \neq (-1,0)$, the demonstration is straightforward since Supp $a_\theta^{[c,d]} = \mathbb{N}$. For the special case $(c,d) = (-1,0)$, we have Supp $a_\theta^{[-1,0]} = \{0,\ldots,\theta\}$. In this case, let show that $p(x+1|x) > 0$ for all $x \in \{0,\ldots,\theta+\gamma-1\}$. It is sufficient to show that there exists $k \leq x$ such that $k \leq \theta$ and $x+1-k \leq \gamma$ since at least one term in the sum of equation (3) would be positive. Remark that $k = \min(x,\theta)$ is a good candidate. A similar argument leads to $p(x|x+1) > 0$ for all $x \in \{0,\ldots,\theta+\gamma-1\}$. The Markov chain $(X_t)_{t\in\mathbb{N}}$ is thus ergodic.

Now let us show that the stationary distribution is $\mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha)$. Assume that $X_{t-1} \sim \mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha)$ and remark that the distribution of X_t is

$$\begin{aligned}
\mathcal{L}_t &= \left\{ q\mathcal{P}_n^{[c,d]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{L}_{t-1} \right\} * \mathcal{L}_\varepsilon, \\
&= \left\{ q\mathcal{P}_n^{[c,d]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha) \right\} * \mathcal{APS}^{[c,d]}(\gamma,\alpha) \\
&= \mathcal{APS}^{[c,d]}(\theta,\alpha) * \mathcal{APS}^{[c,d]}(\gamma,\alpha) \\
\mathcal{L}_t &= \mathcal{APS}^{[c,d]}(\theta+\gamma,\alpha)
\end{aligned}$$

where the third and last equalities are obtained according to lemma 4 and Lemma 3 respectively.

Using Lemma 2 it is easily seen that $\mathrm{E}[X_t] = \frac{\mu_\varepsilon}{1-\rho}$ and $\mathrm{V}(X_t) = \frac{\sigma_\varepsilon^2}{1-\rho}$. Otherwise, the covariance is given by

$$\begin{aligned}
\mathrm{Cov}(X_{t+1}, X_t) &= \mathrm{Cov}(\rho \bullet X_t + \varepsilon_{t+1}, X_t) \\
&= \mathrm{Cov}(\rho \bullet X_t, X_t) + \underbrace{\mathrm{Cov}(\varepsilon_{t+1}, X_t)}_{=0} \\
\mathrm{Cov}(X_{t+1}, X_t) &= \rho V(X_t),
\end{aligned}$$

according to Lemma 6 and thus $\mathrm{Corr}(X_{t+1}, X_t) = \rho$. By recursion, we obtain $\mathrm{Corr}(X_{t+h}, X_t) = \rho^h$.

Finally, to demonstrate the time reversibility of the process $(X_t)_{t\in\mathbb{N}}$, remark that the joint distribution is symmetric

$$\begin{aligned}
p(x_{t+1}, x_t) &= p(x_t)p(x_{t+1}|x_t) \\
&= \frac{a_{\theta+\gamma}^{[c,d]}(x_t)\{g(\alpha)\}^{x_t}}{h_{\theta+\gamma}(\alpha)} \sum_{k=0}^{\min(x_{t+1},x_t)} \frac{a_\theta^{[c,d]}(k)a_\gamma^{[c,d]}(x_t-k)}{a_{\theta+\gamma}^{[c,d]}(x_t)} \frac{a_\gamma^{[c,d]}(x_{t+1}-k)\{g(\alpha)\}^{x_{t+1}-k}}{h_\gamma(\alpha)} \\
p(x_{t+1}, x_t) &= \frac{\{g(\alpha)\}^{x_{t+1}+x_t}}{h_{\theta+2\gamma}(\alpha)} \sum_{k=0}^{\min(x_{t+1},x_t)} a_\theta^{[c,d]}(k)a_\gamma^{[c,d]}(x_t-k)a_\gamma^{[c,d]}(x_{t+1}-k)\{g(\alpha)\}^{-k}.
\end{aligned}$$

□

Models comparison

We have seen, in the previous section, that there are five distinct quasi-Pólya thinning operators. It implies that there are five distinct classes of INAR(1) models. They can be distinguished by their marginal distribution. Note that the number of parameters is varying between two and four according to these five cases:

1. Binomial marginal distribution: this is the case $c = -1$ and $d = 0$ that corresponds to the binomial INAR(1) model introduced by Al-Osh and Alzaid (1991). The vector of parameters is $\boldsymbol{\psi} = (\theta, \gamma, \alpha)$ and the space of parameters is $\Psi = \mathbb{N}^{*2} \times \mathbb{R}_+^*$. This model will be denoted by B-INAR(1).

Distribution	$V(v)$
Binomial	$v(1-v)$
Poisson	v
Negative binomial	$v(1+v)$
Generalized Poisson	$v(1+v)^2$
Generalized negative binomial	$v(1+v)\left(1+v\frac{d+1}{d}\right)$

Table 3 Unity variance function of the five additive modified power series

2. Poisson marginal distribution: this is the case $c = 0$ and $d = 0$ that corresponds to the usual Poisson INAR(1) model introduced by McKenzie (1985); Al-Osh and Alzaid (1987). The vector of parameter is $\boldsymbol{\psi} = (\rho, \alpha)$, where $\rho = \theta/(\theta+\gamma)$, and the space of parameters is $\Psi = (0,1) \times \mathbb{R}_+^*$. This model will be denoted by P-INAR(1).
3. Negative binomial marginal distribution: this is the case $c = 1$ and $d = 0$ that corresponds to the negative binomial INAR(1) model discussed by Joe (1996) as a special case of convolution-closed infinitely divisible class. The vector of parameter is $\boldsymbol{\psi} = (\theta, \gamma, \alpha)$ and the space of parameters is $\Psi = \mathbb{R}_+^{*2} \times (0,1)$. This model will be denoted by NB-INAR(1).
4. Generalized Poisson marginal distribution: this is the case $c = 0$ and $d > 0$ that corresponds to the generalized Poisson INAR(1) model introduced by Alzaid and Al-Osh (1993). The vector of parameter is $\boldsymbol{\psi} = (\rho, \alpha, d)$, where $\rho = \theta/(\theta+\gamma)$, and the space of parameters is $\Psi = (0,1) \times \mathbb{R}_+^* \times [0,1]$. Note that for the generalized Poisson distribution $\mathcal{GP}^{[d]}(\lambda)$, the parameter d must be included in interval $[0,1]$ to obtain a proper distribution (Consul and Famoye, 2006). This INAR(1) model will be denoted by GP-INAR(1). It includes the P-INAR(1) model as a special case when $d = 0$.
5. Generalized negative binomial marginal distribution: this is the case $c = 1$ and $d > -1$ that corresponds to the generalized negative binomial INAR(1) model. This is a new INAR(1) model. The vector of parameter is $\boldsymbol{\psi} = (\theta, \gamma, \alpha, d)$ and the space of parameters is $\Psi = \mathbb{R}_+^{*2} \times (0,1) \times [0, (1-\alpha)/\alpha]$. Note that for the generalized negative binomial distribution $\mathcal{GNB}^{[d]}(\theta, \alpha)$, the parameter d must be included in interval $[0, (1-\alpha)/\alpha]$ to obtain a proper distribution (Consul and Famoye, 2006). This INAR(1) model will be denoted by GNB-INAR(1). It includes the B-INAR(1) model as a special case when $d = -1$ and the NB-INAR(1) model when $d = 0$.

Since additive modified power series distributions belong to the family of dispersion models (Jorgensen, 1997), their unity variance functions have closed form; see Table 3. As remarked by Jourdan and Kokonendji (2002) these unity variance functions are strictly ordered as follows,

$$V_{\mathcal{B}} < V_{\mathcal{P}} < V_{\mathcal{NB}} < V_{\mathcal{GP}} < V_{\mathcal{GNB}}.$$

Therefore, according to the level of dispersion observed in the dataset, the autoregressive model with the appropriate marginal distribution has to be used. The different models could be compared using penalized criteria, like Akaike information criterion (AIC) as proposed by Weiß (2008b).

5 Model inference based on likelihood

Maximum likelihood estimator

The log-likelihood of a count time series $\boldsymbol{x} = (x_0, x_1, \ldots, x_T)$ for an INAR(1) model based on quasi Pólya thinning operator with parameter $\boldsymbol{\psi} = (\theta, \gamma, \alpha, d)$ is

$$L_T(\boldsymbol{\psi}) = l(x_0; \boldsymbol{\psi}) + \sum_{t=1}^{T} l(x_t | x_{t-1}; \boldsymbol{\psi}), \tag{4}$$

where $l(x_0; \boldsymbol{\psi}) = \ln p(x_0; \boldsymbol{\psi})$, $l(x_t | x_{t-1}; \boldsymbol{\psi}) = \ln p(x_t | x_{t-1}; \boldsymbol{\psi})$. Since X_0 is assumed to follow the modified power series distribution $\mathcal{APS}^{[c,d]}(\theta + \gamma, \alpha)$, then its pmf is

$$p(x_0; \boldsymbol{\psi}) = \frac{a^{[c,d]}_{\theta+\gamma}(x_0)\{g(\alpha)\}^{x_0}}{h_{\theta+\gamma}(\alpha)}.$$

Note that the log-likelihood of the first observation could be omitted, since it is asymptotically irrelevant. The right part of the log-likelihood is obtained from (3).

Let r denote the dimension of the vector of parameters $\boldsymbol{\psi}$ (note that r is varying between 2 and 4 according to the five distinct classes of INAR(1) model). Let $l_u = \partial l / \partial \boldsymbol{\psi}_u$ denote the partial derivative with respect to the u^{th} component of $\boldsymbol{\psi}$ for $u = 1, \ldots, r$. In the same way, let l_{uv} and l_{uvw} denote the partial derivatives of second and third order (same notations hold for p). Finally, $\mathrm{E}_{\boldsymbol{\psi}}$ means expectation, assuming that the true parameter value is $\boldsymbol{\psi}$ and X_0 starts with a stationary distribution. Following Joe (1997) (page 317) we introduce the following regularity conditions:

(C1) For $\boldsymbol{\psi} \in \Psi$ there exists a neighborhood $N_{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ such that for all u, v, w and x

$$\sum_{y \in \mathbb{N}} \sup_{\boldsymbol{\psi}' \in N_{\boldsymbol{\psi}}} p_u(y|x; \boldsymbol{\psi}) < \infty$$
$$\sum_{y \in \mathbb{N}} \sup_{\boldsymbol{\psi}' \in N_{\boldsymbol{\psi}}} p_{uv}(y|x; \boldsymbol{\psi}) < \infty$$
$$\mathrm{E}_{\boldsymbol{\psi}} \left[\sup_{\boldsymbol{\psi}' \in N_{\boldsymbol{\psi}}} |l_{uvw}(X_1 | X_0; \boldsymbol{\psi})| \right] < \infty$$

(C2) For $u = 1, \ldots, r$ $\mathrm{E}_{\boldsymbol{\psi}} \left[|l_u(X_1 | X_0; \boldsymbol{\psi})|^2 \right] < \infty$ and $\Sigma(\boldsymbol{\psi}) = (\sigma_{uv}(\boldsymbol{\psi}))$ is a non-singular $r \times r$ matrix where $\sigma_{uv}(\boldsymbol{\psi}) = \mathrm{E}_{\boldsymbol{\psi}} \left[l_u(X_1 | X_0; \boldsymbol{\psi}) l_v(X_1 | X_0; \boldsymbol{\psi}) \right]$.

Theorem 3. *Let $(X_t)_{t \in \mathbb{N}}$ follow an INAR(1) model based on quasi Pólya thinning operator. Assume that $d \neq 0$ or $c \geq 0$, avoiding the case of INAR(1) model based on hypergeometric thinning operator. Assume that the starting distribution is $\mathcal{L}_0 = \mathcal{APS}^{[c,d]}(\theta + \gamma, \alpha)$ for some $\alpha \in (0, R)$ and the error distribution is $\mathcal{L}_\varepsilon = \mathcal{APS}^{[c,d]}(\gamma, \alpha)$. Under the regularity conditions (C1) and (C2), the MLE $\hat{\boldsymbol{\psi}}_T$ converges*

in probability to the true parameter $\boldsymbol{\psi}$ *with asymptotic normal distribution*

$$\frac{\hat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}}{\sqrt{T}} \underset{T\to\infty}{\longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma^{-1}(\boldsymbol{\psi})).$$

Proof. The proof follows the approach of Joe (1997), that is based on the result of Billingsley (1961) about Markov chains of first order. Let us verify the following regularity conditions:

- All states of the Markov chain communicate with each other.
  In the proof of previous theorem, it has already been shown that there are no transient state since $p(x+1|x) > 0$ and $p(x|x+1) > 0$ for all $x \in \mathbb{N}$.
- The support of $p(\cdot|x;\boldsymbol{\psi})$ does not depend on $\boldsymbol{\psi}$.
  According to (3) it is easily seen that Supp $a_\theta = \mathbb{N} \Rightarrow$ Supp $p(\cdot|x;\boldsymbol{\psi}) = \mathbb{N}$. Note that this condition is not verified for the B-INAR(1) model since Supp $a_\theta = \{0, 1, \ldots, \theta\}$ in this case.
- p_u, p_{uv}, p_{uvw} $(u, v, w \in \{1, \ldots, r\})$ exist and are continuous in $\boldsymbol{\psi}$.
  According to (3), $p(\cdot|x;\boldsymbol{\psi})$ is a sum of ratio of functions a_θ, a_γ, $g(\alpha)$ and $h_\gamma(\alpha)$. Since the gamma, power and exponential functions are smooth, this condition is obtained.
- $p(\cdot|x;\boldsymbol{\psi})$ is absolutely continuous with respect to the stationary measure $\mu(\cdot;\boldsymbol{\psi})$.
  In our case, the stationary measure of the INAR(1) model is the modified power series distribution $\mathcal{APS}^{[c,d]}(\theta+\gamma, \alpha)$ for some $\alpha \in (0, R)$. Except for the binomial case ($c < 0$ and $d = 0$), the support of this distribution is the entire set of non-negative integers $\mathbb{N}$.

□

As remarked by Joe (1997) this asymptotic result implies, as in the i.i.d. case, that the negative inverse Hessian of $L_T(\boldsymbol{\psi})$ evaluated at the MLE $\hat{\boldsymbol{\psi}}$ can be used as an estimated covariance of $\hat{\boldsymbol{\psi}}$. In Section Application, the MLE is performed via the *optim* function contained in the stats R package. The optim argument, hessian=TRUE, allows determining the standard errors of MLE (via the Hessian pseudo-inverse when the Hessian is non-invertible).

Likelihood ratio test

A likelihood ratio test can also be employed, since the stationary distributions of the five different INAR(1) models are nested; see Table 2 (they belong to the family of additive modified power series distributions). The case $c < 0$ is not totally separated from the case $c > 0$ since both distribution $\mathcal{APS}^{[-1,d]}$ and $\mathcal{APS}^{[1,d-1]}$ are equivalent. Therefore, the case $c < 0$ is omitted and focus is made on the four distributions: Poisson, negative binomial, generalized Poisson, generalized negative binomial.

- The NB-INAR(1) model can be compared with the P-INAR(1) model through the null hypothesis $H_0 : c = 0$ against $H_1 : c > 0$ (assuming $d = 0$). The log-likelihood test statistic $2(\mathcal{L}_{\mathcal{NB}} - \mathcal{L}_{\mathcal{P}})$ follows a chi-square distribution $\chi^2(1)$ with one degree of freedom, under the null hypothesis.

- The GP-INAR(1) model can be compared with the P-INAR(1) through the null hypothesis $H_0' : d = 0$ against $H_1' : d > 0$ (assuming $c = 0$). The log-likelihood test statistic $2(\mathcal{L}_{\mathcal{GP}} - \mathcal{L}_{\mathcal{P}})$ follows a chi-square distribution $\chi^2(1)$ with one degree of freedom, under the null hypothesis.
- The GNB-INAR(1) model can be compared with the P-INAR(1) through the null hypothesis $H_0'' : (c, d) = (0, 0)$ against $H_1'' : (c, d) \neq (0, 0)$. The log-likelihood test statistic $2(\mathcal{L}_{\mathcal{GNB}} - \mathcal{L}_{\mathcal{P}})$ follows a chi-square distribution $\chi^2(2)$ with two degree of freedom, under the null hypothesis.

Model validation

The standardized Pearson residuals are given by

$$e_t := \frac{\mathrm{E}[X_t|x_{t-1}] - x_t}{\sqrt{\mathrm{V}(X_t|x_{t-1})}}.$$

For an adequate model, these residuals should have a null mean and a unit variance, according to Harvey and Fernandes (1989). In our context of INAR(1) model based on quasi Pólya thinning operator, if we denote $Y_t := \rho \bullet X_{t-1}$, we have $\mathrm{E}[X_t|x_{t-1}] = \mathrm{E}[Y_t|x_{t-1}] + \mathrm{E}[\varepsilon_t]$ and $\mathrm{V}(X_t|x_{t-1}) = \mathrm{V}(Y_t|x_{t-1}) + \mathrm{V}(\varepsilon_t)$ where $Y_t|x_{t-1} \sim q\mathcal{P}_{x_{t-1}}^{[c,d]}(\theta, \gamma)$ and $\varepsilon_t \sim \mathcal{APS}^{[c,d]}(\gamma, \alpha)$ are independent variables. According to the result of Janardan and Schaeffer (1977) (given in Lemma 1) we have

$$\mathrm{E}[Y_t|x_{t-1}] = \frac{\theta}{\theta + \gamma} x_{t-1},$$

$$\mathrm{V}(Y_t|x_{t-1}) = \frac{\theta\gamma}{\theta + \gamma} \left[\frac{x_{t-1}^2}{\theta + \gamma} - \sum_{j=1}^{x_{t-1}-1} \frac{x_{t-1}^{(j+1;-1)}(d + c)^{j-1}}{\{\theta + \gamma + dx_{t-1} + (x_{t-1} - j + 2)c\}^{(j;c)}} \right].$$

Moreover, according to Gupta (1974) we have

$$\mu = \mathrm{E}[\varepsilon_t] = \frac{g(\alpha)}{g'(\alpha)} \frac{h_\gamma'(\alpha)}{h_\gamma(\alpha)} \quad \text{and} \quad \mathrm{V}(\varepsilon_t) = \frac{g(\alpha)}{g'(\alpha)} \mu'(\alpha).$$

Appendix B details the computation for the five different INAR(1) models.

6 Application

In this section we will compare the five INAR(1) models based on quasi Pólya thinning operator, described in Section 4: the B-INAR(1), P-INAR(1), NB-INAR(1), GP-INAR(1) and the GNB-INAR(1) models. The MLEs have been computed using the *optim_sa* function, available in the R-package *optimization*.

6.1 Simulated data

The five kinds of INAR(1) models have been simulated with a length of time series $T = 50$ and auto correlation $\rho = 0.3$, $\rho = 0.6$ and $\rho = 0.9$. For each of the five models,

three values of parameters $\boldsymbol{\psi}$ have been simulated, $n = 1000$ times, in order to obtain different mean values: $\mu = 5$, $\mu = 20$ and $\mu = 50$. The variance has also been calibrated from under dispersion $\sigma^2 = \mu/2$ with the B-INAR(1) model to huge over dispersion $\sigma^2 = 8\mu$ with the GNB-INAR(1) model. Finally, $5 \times 3 \times 3 = 45$ scenarios have been simulated $n = 1000$ times.

For instance, focusing on the B-INAR(1) model, the parameters are given by $\theta = 2\mu\rho$, $\gamma = 2\mu(1-\rho)$ and $\alpha = 1$. Following Theorem 2 the first value is simulated according to the binomial distribution $X_0 \sim \mathcal{B}_{\theta+\gamma}(p)$ where $p = \alpha/(1+\alpha)$ and thus the process is stationary with $X_t \sim \mathcal{B}_{\theta+\gamma}(p)$ for all $t = 1, \ldots, T$. Therefore, we have $\mathrm{E}[X_t] = (\theta+\gamma)p = \mu$ and $\mathrm{V}(X_t) = (\theta+\gamma)p(1-p) = \mu/2$. For instance, with $\mu = 5$ we have $\boldsymbol{\psi} = (\theta, \gamma, \alpha) = (6, 4, 1)$. The parameters $\boldsymbol{\psi}$ of other INAR(1) models have been determined following the same way.

Then for each of the 45 scenarios, the five INAR(1) models have been inferred using the MLE on the $n = 1000$ simulated time series. The estimated marginal distributions have been compared with the true marginal distribution using the Kullback-Leibler (KL) divergence. It was computed as $\sum_i p(i) \ln\{p(i)/\hat{p}(i)\}$ for i between minimum and maximum observed values, where p (resp. $\hat{p}$) denotes the true (resp. estimated) pmf. The results have been summarized in Table 4 (resp. Tables C3 and C4 in Appendix C) for times series of length $T = 50$ with autocorrelation $\rho = 0.6$ (resp. $\rho = 0.3$ and $\rho = 0.9$). For each scenario, the mean (and standard deviations) of the $n = 1000$ KL divergences are given, and the smaller (among the five models) appears in bold.

The results can be interpreted according to the three kinds of dispersion: under dispersion (binomial), equi dispersion (Poisson) and over dispersion (negative binomial, generalized Poisson and generalized negative binomial). When the time series are simulated according to a B-INAR(1) model with small or middle auto correlation ($\rho = 0.3$ or 0.6), the smallest KL divergence is obtained with the B-INAR(1) model. The other models are more divergent except the GNB-INAR(1) model when the auto correlation is small ($\rho = 0.3$). When the auto correlation is high ($\rho = 0.9$), the smallest KL divergence is obtained with the P-INAR(1) model but the results are very close between the five models. When the time series are simulated according to a P-INAR(1) model, the KL divergence is smaller with the P-INAR(1) model but is approximately the same with the GP-INAR(1) models. When the time series are simulated according to an over dispersed INAR(1) model (with NB, GP or GNB margins), the KL divergence is huge for the B-INAR(1) model or P-INAR(1) model. On the contrary, the KL divergence is small for the three other INAR(1) models, particularly with the GP and GNB-INAR(1) models. More precisely, when the time series are simulated according to a NB-INAR(1) model, the KL divergence is smaller with the GP-INAR(1) model or GNB-INAR(1) model but never with the NB-INAR(1) model. Performances of GP-INAR(1) model and GNB-INAR(1) model are very similar. Nevertheless, it should be kept in mind that the GNB-INAR(1) model has four parameters, whereas its competitor only has three. Penalized criteria could thus be used when comparing the different INAR(1) models on real data. The results are globally similar for the different mean values ($\mu = 5$, $\mu = 20$ or $\mu = 50$) except when the mean and the variance are high together ($\mu = 50$ and $\sigma^2 = 8\mu$). In this case, the GP-INAR(1) model gives very poor results and is outperformed by both NB and GNB-INAR(1) models that are stable.

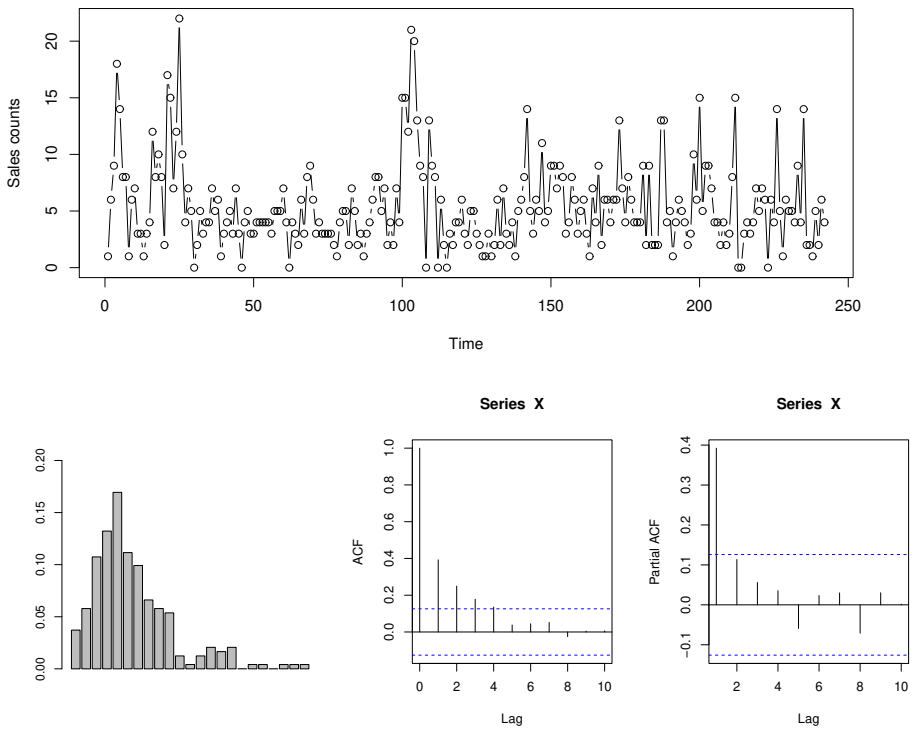

Fig. 1 Summary of soap dataset about sales counts. Plot of the count time series in top row and plots of sample probability mass function and (partial) autocorrelation function in bottom.

Finally, the results are impact by the auto correlation values. The KL divergences are globally smaller for small auto correlation ($\rho = 0.3$) and higher for high auto correlation ($\rho = 0.9$).

6.2 Real data

As an example, let us use the dataset of soap product sales presented by MacDonald and Zucchini (2016). This is a count time series of length $T = 242$ about the weekly number of sales of a soap product in a supermarket. The counts exhibit an important over dispersion, which is also remarkable from the plot of the sample pmf in Figure 1. The partial ACF (PACF) has a significant value only at lag 1 ($\hat{\rho}(1) \simeq 0.39$). It is thus natural to use INAR(1) models on this dataset. The sample mean is equal to 5.44 and the variance is equal to 15.40.

The five INAR(1) models based on quasi Pólya thinning operator have been inferred with the MLE. Summarized results are given in Table 5, allowing the comparison of these models. Using the AIC for model selection, the GP-INAR(1) model is preferred; see Figure 2 for the adjusted pmf and ACFs using this model. Note that

Marginal	Mean	Variance	B	P	NB	GP	GNB
B	$\mu = 5$	$\sigma^2 = \mu/2$	**7.73e-02** (9.28e-02)	1.15e-01 (3.10e-02)	1.22e-01 (3.66e-02)	1.27e-01 (4.09e-02)	1.48e-01 (6.16e-02)
	$\mu = 20$		**7.14e-02** (7.26e-02)	1.11e-01 (3.20e-02)	1.23e-01 (4.22e-02)	1.28e-01 (4.48e-02)	1.63e-01 (6.49e-02)
	$\mu = 50$		**7.37e-02** (1.13e-01)	1.12e-01 (3.41e-02)	1.23e-01 (3.91e-02)	1.36e-01 (6.07e-02)	1.93e-01 (9.72e-02)
P	$\mu = 5$	$\sigma^2 = \mu$	6.11e-02 (8.06e-02)	**3.87e-02** (6.04e-02)	4.20e-02 (6.15e-02)	4.14e-02 (5.84e-02)	5.80e-02 (8.04e-02)
	$\mu = 20$		6.89e-02 (1.71e-01)	**3.56e-02** (5.01e-02)	3.93e-02 (5.43e-02)	3.92e-02 (5.07e-02)	5.58e-02 (7.04e-02)
	$\mu = 50$		7.06e-02 (2.73e-01)	**3.64e-02** (4.86e-02)	3.99e-02 (5.19e-02)	4.08e-02 (5.18e-02)	5.10e-02 (6.02e-02)
NB	$\mu = 5$	$\sigma^2 = 2\mu$	2.35e-01 (1.27e-01)	2.19e-01 (1.18e-01)	7.06e-02 (8.09e-02)	**6.36e-02** (7.09e-02)	6.56e-02 (7.37e-02)
	$\mu = 20$		2.52e-01 (1.35e-01)	2.29e-01 (1.27e-01)	7.05e-02 (7.80e-02)	**6.52e-02** (7.26e-02)	6.55e-02 (7.31e-02)
	$\mu = 50$		2.66e-01 (1.48e-01)	2.34e-01 (1.26e-01)	7.01e-02 (7.97e-02)	**6.83e-02** (7.41e-02)	6.87e-02 (7.49e-02)
GP	$\mu = 5$	$\sigma^2 = 4\mu$	7.28e-01 (2.22e-01)	6.99e-01 (2.18e-01)	7.85e-02 (8.88e-02)	**6.51e-02** (7.34e-02)	6.73e-02 (7.84e-02)
	$\mu = 20$		8.29e-01 (2.13e-01)	7.91e-01 (2.07e-01)	7.12e-02 (8.95e-02)	6.45e-02 (7.95e-02)	**6.37e-02** (7.82e-02)
	$\mu = 50$		8.82e-01 (2.06e-01)	8.32e-01 (2.02e-01)	7.75e-02 (8.88e-02)	6.93e-02 (7.69e-02)	**6.91e-02** (7.62e-02)
GNB	$\mu = 5$	$\sigma^2 = 8\mu$	1.67e+00 (5.76e-01)	1.62e+00 (5.18e-01)	8.87e-02 (1.09e-01)	**6.56e-02** (8.64e-02)	6.99e-02 (9.92e-02)
	$\mu = 20$		2.14e+00 (4.52e-01)	2.07e+00 (4.36e-01)	7.76e-02 (8.44e-02)	**6.28e-02** (7.67e-02)	6.45e-02 (7.48e-02)
	$\mu = 50$		2.39e+00 (5.24e-01)	9.97e+00 (2.94e+01)	7.89e-02 (8.88e-02)	7.99e+00 (2.99e+01)	**6.98e-02** (7.92e-02)

Table 4 Means (and standard deviations) of KL divergences for different INAR(1) models based on quasi Pólya thinning operator for integer valued time series of length $T = 50$ and autocorrelation $\rho = 0.6$. The rows indicate the marginal distribution of the simulated INAR(1) model whereas the columns indicate the marginal distribution of the estimated INAR(1) model. Means of KL divergences (and standard deviations) have been computed for the 5 different estimated models, on $n = 1000$ simulations for each of the 15 scenarios. The minimal value of each scenario is written in bold.

	L	AIC	Mean	Var	ACF at lag 1	2	3	4	Pearson residuals Mean	Var
Sample			5.442	15.401	0.392	0.250	0.178	0.136		
B-INAR(1)	-682.73	1371.46	5.693	5.636	0.2604	0.0678	0.0177	0.0046	0.08653	2.448
P-INAR(1)	-680.12	1364.25	5.340	5.340	0.2300	0.0529	0.0122	0.0028	-0.02728	2.528
NB-INAR(1)	-614.69	1235.39	**5.408**	12.578	0.3235	0.1047	0.0339	0.0110	**-0.00257**	1.107
GP-INAR(1)	-614.55	**1235.10**	5.328	13.008	**0.3700**	**0.1369**	**0.0507**	**0.0187**	-0.02426	1.120
GNB-INAR(1)	**-614.26**	1236.51	5.542	**14.856**	0.3050	0.093	0.0284	0.0087	0.02626	**0.943**

Table 5 Summarized results for the five INAR(1) models based on quasi Pólya thinning operator, inferred on the soap products dataset. For each model, the log-likelihood, the AIC are given. Then the estimations of the mean, variance and autocorrelation (lag from 1 to 4) of the stationary distribution are given. Finally the mean and variance of the estimated standardized Pearson residuals are given. For each column, the best value is written in bold.

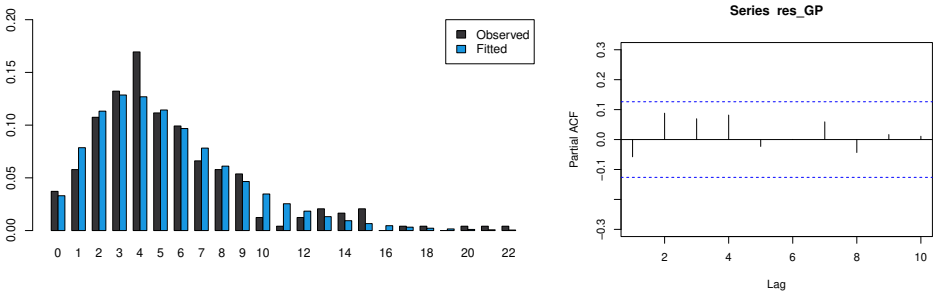

Fig. 2 Plot of the pmf and partial autocorrelation function (for Pearson residuals) for the GP-INAR(1) model.

NB, GP and GNB-INAR(1) models have very closed values of log-likelihood, contrarily to the B and P-INAR(1) models that obtained a poor fit. That could be explained by the high over dispersion observed in the dataset. The variance is clearly underestimated with these two models. On the contrary the variance is well estimated for the three other models, with ordered values $12.578 < 13.008 < 14.856$ as expected, closed to the observed variance of 15.401. The autocorrelation of first order is well captured by the GP-INAR(1) model with $\hat{\rho} = 0.37$ compared with the observed value of 0.39. For the model adequacy, we can check the mean and variance of the standardized Pearson residuals. The NB-INAR(1) model obtained the mean closest to 0 and the GNB-INAR(1) model obtained the variance closest to 1. Finally, regarding the different criteria of model adequacy, the NB, GP and GNB-INAR(1) models cannot really be distinguished.

7 Conclusion

The unifying class of INAR(1) models based on quasi Pólya thinning operator has been introduced. It includes two well-known INAR(1) models, based on binomial and

quasi binomial thinning operators. It also includes two less usual INAR(1) models, based on hypergeometric and beta binomial thinning operators. Finally, it includes a new INAR(1) model based on quasi beta binomial thinning operator. This unified framework allowed us to derive the different properties of these models in one way. For instance, the class of additive modified power series distributions emerges as the natural choice for the error distribution. The asymptotic normality of the MLE for such a class of models has been shown under usual regularity conditions. Finally, the proposed class of INAR(1) models have been applied on simulated and real datasets.

The huge range of variability of this class allows taking into account different situations: under dispersion with B-INAR(1) model, equi dispersion with P-INAR(1) model and over dispersion with NB, GP and GNB-INAR(1) models. In real situations, the data are often over dispersed and the GP-INAR(1) model and GNB-INAR(1) model well perform, regarding the simulations results. In the example of real dataset, the NB, GP and GNB-INAR(1) models give very similar results regarding the different criteria. The GNB-INAR(1) model is more flexible since it includes the NB-INAR(1) model as a special case and has four parameters. But the additional parameter d seems not sufficient to clearly distinguish it from NB or GP-INAR(1) models. It may be due to the small order of the Markov chain. Therefore, it could be interesting to compare the performances of different integer-valued autoregressive models (INAR models), with a higher order. A natural perspective is thus to define INAR(p) models based on quasi Pólya thinning operator and finally INARMA(p, q) models. Until now, INAR(p) models with closure property (i.e., same distribution for margin and residual), has been defined only for the case of Poisson margin (Alzaid and Al-Osh, 1990; Weiß, 2008a) that do not take into account under and over dispersed data. The extension to CINAR(p) models based on quasi Pólya thinning operator will use the properties of such an operator (Lemmas 4 and 3). The extension of the INAR(p) model of Alzaid and Al-Osh (1990) to non-Poisson margins will need the multivariate version of the quasi Pólya thinning operator introduced by Peyhardi (2023). The stationarity of such extensions will use the same closure properties, under thinning operator and under convolution.

Declarations

For the purpose of Open Access, a CC-BY 4.0 public copyright licence (https://creativecommons. org/licenses/by/4.0/) has been applied by the author to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

Conflict of interest

The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A Characteristics of quasi Pólya and additive modified power series distributions

Name	Notation	Space parameter	Support	$a_\theta^{[c,d]}$	pmf
quasi Pólya $(c \in \mathbb{R},\, t \in \mathbb{R})$	$q\mathcal{P}_n^{[c,d]}(\theta,\gamma)$	$n \in \mathbb{N}^*,\ (\theta,\gamma) \in \Theta_c^2$		$\frac{\theta(\theta+dn)^{(n;c)}}{(\theta+dn)n!}$	$\frac{a_\theta^{[c,d]}(k)a_\gamma^{[c,d]}(n-k)}{a_{\theta+\gamma}^{[c,d]}(n)}$
1. hypergeometric $(c=-1, d=0)$	$\mathcal{H}_n(\theta,\gamma)$	$n \in \mathbb{N}^*,\ (\theta,\gamma) \in \mathbb{N}^{*2}$	$\{0,\dots,n\} \cap \{n-\gamma,\dots,\theta\}$	$\binom{\theta}{n}$	$\frac{\binom{\theta}{k}\binom{\gamma}{n-k}}{\binom{\theta+\gamma}{n}}$
2. binomial $(c=0, d=0)$	$\mathcal{B}_n(p)$	$n \in \mathbb{N}^*,\ p \in (0,1)$ $p = \theta/(\theta+\gamma)$	$\{0,\dots,n\}$	$\frac{\theta^n}{n!}$	$\binom{n}{k}p^k(1-p)^{n-k}$
3. beta binomial $(c=1, d=0)$	$\beta\mathcal{B}_n(\theta,\gamma)$	$n \in \mathbb{N}^*,\ (\theta,\gamma) \in \mathbb{R}_+^{*2}$	$\{0,\dots,n\}$	$\binom{n+\theta-1}{n}$	$\frac{\binom{k+\theta-1}{k}\binom{n-k+\gamma-1}{n-k}}{\binom{n+\theta+\gamma-1}{n}}$
quasi hypergeometric $(c=-1, d\neq 0)$	$q\mathcal{H}_n^{[d]}(\theta,\gamma)$	$n \in \mathbb{N}^*,\ (\theta,\gamma) \in \mathbb{R}_+^{*2}$ $d \in (0,\infty)$	$\{0,\dots,n\}$	$\frac{\theta}{\theta+dn}\binom{\theta+dn}{n}$	$\frac{\frac{\theta}{\theta+dk}\binom{\theta+dk}{k}\frac{\gamma}{\gamma+d(n-k)}\binom{\gamma+d(n-k)}{n-k}}{\frac{\theta+\gamma}{\theta+\gamma+dn}\binom{\theta+\gamma+dn}{n}}$
4. quasi binomial $(c=0,\ d\neq 0)$	$q\mathcal{B}_n^{[d']}(p)$	$n \in \mathbb{N}^*$ $p = \theta/(\theta+\gamma) \in (0,1)$ $d' = d/(\theta+\gamma) \in (0,\infty)$	$\{0,\dots,n\}$	$\frac{\theta(\theta+dn)^{n-1}}{n!}$	$\binom{n}{k}\frac{p(1-p)}{1+d'n}\left(\frac{p+d'k}{1+d'n}\right)^{k-1}\left(\frac{1-p+d'(n-k)}{1+d'n}\right)^{n-k-1}$
5. quasi beta binomial $(c=1, d\neq 0)$	$q\beta\mathcal{B}_n^{[d]}(\theta,\gamma)$	$n \in \mathbb{N}^*,\ (\theta,\gamma) \in \mathbb{R}_+^{*2}$ $d \in (-1,\infty),\ d \neq 0$	$\{0,\dots,n\}$	$\frac{\theta}{\theta+dn}\binom{\theta+dn+n-1}{n}$	$\frac{\frac{\theta}{\theta+dk}\binom{\theta+dk+k-1}{k}\frac{\gamma}{\gamma+t(n-k)}\binom{\gamma+d(n-k)+n-k-1}{n-k}}{\frac{\theta+\gamma}{\theta+\gamma+dn}\binom{\theta+\gamma+dn+n-1}{n}}$

Table A1 Notations and pmf of quasi Pólya distributions. The cell with quasi hypergeometric distribution $q\mathcal{H}_n^{[d]}(\theta,\gamma)$ is greyed out since it is equivalent to the quasi beta binomial distribution $q\beta\mathcal{B}_n^{[d-1]}(\theta,\gamma)$.

Name	Notation	Parameters space	Supp	$g(\alpha)$	$h_\theta(\alpha)$	pmf
additive modified power series $(c \in \mathbb{Z},\ d \in \mathbb{R})$	$\mathcal{APS}^{[c,d]}(\theta, \alpha)$	$\theta \in \Theta_c,\ \alpha \in (0, R_c)$				$\frac{a_\theta^{[c,d]}(k)\{g(\alpha)\}^k}{h_\theta(\alpha)}$
1. binomial $(c = -1, d = 0)$	$\mathcal{B}_\theta(p)$	$\theta \in \mathbb{N}^*,\ \alpha \in (0, \infty)$ $p = \frac{\alpha}{1+\alpha} \in (0, 1)$	$\{0, \dots, \theta\}$	α	$(1+\alpha)^\theta$	$\binom{\theta}{n}(\frac{\alpha}{1+\alpha})^n(\frac{1}{1+\alpha})^{\theta-n}$
2. Poisson $(c = 0, d = 0)$	$\mathcal{P}(\alpha\theta)$	$\theta \in \mathbb{R}_+^*$ $\alpha \in (0, \infty)$	$\mathbb{N}$	α	$e^{\alpha\theta}$	$\frac{\theta^n}{n!}\alpha^n e^{-\alpha\theta}$
3. Negative binomial $(c = 1, d = 0)$	$\mathcal{NB}(\theta, p)$	$\theta \in \mathbb{R}_+^*$ $p = \alpha \in (0, 1)$	$\mathbb{N}$	α	$(1-\alpha)^{-\theta}$	$\binom{n+\theta-1}{n}\alpha^n(1-\alpha)^\theta$
generalized binomial $(c = -1,\ 1 < d \leq \frac{1+\alpha}{\alpha})$	$\mathcal{GB}^{[d]}(\theta, p)$	$\theta \in \mathbb{R}_+^*,\ \alpha \in (0, \infty)$ $p = \frac{\alpha}{1+\alpha} \in (0, 1)$	$\mathbb{N}$	$\alpha(1+\alpha)^{-d}$	$(1+\alpha)^\theta$	$\frac{\theta}{\theta+dn}\binom{\theta+dn}{n}(\frac{\alpha}{1+\alpha})^n(\frac{1}{1+\alpha})^{\theta+dn-n}$
4. generalized Poisson $(c = 0, 0 < d \leq 1)$	$\mathcal{GP}^{[d]}(\alpha\theta)$	$\theta \in \mathbb{R}_+^*,\ \alpha \in (0, \infty)$	$\mathbb{N}$	$\alpha e^{-d\alpha}$	$e^{\alpha\theta}$	$\frac{\theta(\theta+dn)^{n-1}}{n!}\frac{(\alpha e^{-d\alpha})^n}{e^{\alpha\theta}}$
5. generalized negative binomial $(c = 1,\ 0 < d \leq \frac{1-\alpha}{\alpha})$	$\mathcal{GNB}^{[d]}(\theta, p)$	$\theta \in \mathbb{R}_+^*,\ p = \alpha \in (0, 1)$	$\mathbb{N}$	$\alpha(1-\alpha)^d$	$(1-\alpha)^{-\theta}$	$\frac{\theta}{\theta+dn}\binom{\theta+dn+n-1}{n}\alpha^n(1-\alpha)^{\theta+dn}$

Table A2 Notations and pmf of additive modified power series distributions. The cell generalized binomial distribution $\mathcal{GB}^{[d]}(\theta, p)$ is greyed out since it is equivalent to the generalized negative binomial distribution $\mathcal{GNB}^{[d-1]}(\theta, p)$. After identifiability our generalized Poisson distribution $\mathcal{GP}^{[d]}(\alpha\theta)$ becomes those introduced by Consul and Jain (1973) denoted by $\mathcal{GP}^{[d^*]}(\lambda)$ with $\lambda = \theta\alpha$ and $d^* = d\alpha$.

Appendix B Standardized Pearson residuals

The standardized Pearson residuals are given by

$$e_t := \frac{\mathrm{E}[X_t|x_{t-1}] - x_t}{\sqrt{\mathrm{V}(X_t|x_{t-1})}}.$$

Let us describe the specific value of $\mathrm{E}[X_t|x_{t-1}]$ and $\mathrm{V}(X_t|x_{t-1})$ in the five cases of INAR(1) models based on quasi Pólya thinning operator.

1. For the B-INAR(1) model with parameters $\boldsymbol{\psi} = (\theta, \gamma, \alpha)$ we have

$$\mathrm{E}[X_t|x_{t-1}] = \frac{\theta}{\theta+\gamma}x_{t-1} + \gamma\frac{\alpha}{1+\alpha},$$
$$\mathrm{V}(X_t|x_{t-1}) = \frac{\theta\gamma(\theta+\gamma-x_{t-1})x_{t-1}}{(\theta+\gamma)^2(\theta+\gamma-1)} + \gamma\frac{\alpha}{(1+\alpha)^2}.$$

2. For the P-INAR(1) model with parameters $\boldsymbol{\psi} = (\rho, \alpha)$ we have

$$\mathrm{E}[X_t|x_{t-1}] = \rho x_{t-1} + (1-\rho)\lambda,$$

$$\mathrm{V}(X_t|x_{t-1}) = \rho(1-\rho)x_{t-1} + (1-\rho)\lambda.$$

3. For the NB-INAR(1) model with parameters $\boldsymbol{\psi} = (\theta, \gamma, \alpha)$ we have

$$\begin{aligned}
\mathrm{E}[X_t|x_{t-1}] &= \frac{\theta}{\theta+\gamma}x_{t-1} + \gamma\frac{\alpha}{1-\alpha}, \\
\mathrm{V}(X_t|x_{t-1}) &= \frac{\theta\gamma(\theta+\gamma+x_{t-1})x_{t-1}}{(\theta+\gamma)^2(\theta+\gamma+1)} + \gamma\frac{\alpha}{(1-\alpha)^2}.
\end{aligned}$$

4. For the GP-INAR(1) model with parameters $\boldsymbol{\psi} = (\rho, \lambda, d)$ we have

$$\begin{aligned}
\mathrm{E}[X_t|x_{t-1}] &= \rho x_{t-1} + (1-\rho)\frac{\lambda}{1-d}, \\
\mathrm{V}(X_t|x_{t-1}) &= \rho(1-\rho)\left[x_{t-1}^2 - \sum_{j=1}^{x_{t-1}-1}\frac{x_{t-1}^{(j+1;-1)}d'^{j-1}}{(1+d'x_{t-1})^j}\right] + (1-\rho)\frac{\lambda}{(1-d)^3}.
\end{aligned}$$

5. For the GNB-INAR(1) model with parameters $\boldsymbol{\psi} = (\theta, \gamma, \alpha, d)$ we have

$$\begin{aligned}
\mathrm{E}[X_t|x_{t-1}] &= \frac{\theta}{\theta+\gamma}x_{t-1} + \gamma\frac{\alpha}{1-\alpha(1+d)}, \\
\mathrm{V}(X_t|x_{t-1}) &= \frac{\theta\gamma}{\theta+\gamma}\left[\frac{x_{t-1}^2}{\theta+\gamma} - \sum_{j=1}^{x_{t-1}-1}\frac{x_{t-1}^{(j+1;-1)}(d+1)^{j-1}}{\{\theta+\gamma+(d+1)x_{t-1}-j+2\}^{(j;1)}}\right] + \gamma\frac{\alpha(1-\alpha)}{\{1-\alpha(1+d)\}^3}.
\end{aligned}$$

Appendix C Summarized results on simulated datasets

Marginal	Mean	Variance	B	P	NB	GP	GNB
B	$\mu = 5$	$\sigma^2 = \mu/2$	**4.17e-02** (4.45e-02)	1.05e-01 (1.46e-02)	1.10e-01 (1.71e-02)	1.05e-01 (1.59e-02)	6.11e-02 (6.14e-02)
	$\mu = 20$		**3.58e-02** (3.59e-02)	1.01e-01 (1.44e-02)	1.08e-01 (1.95e-02)	1.01e-01 (1.44e-02)	6.49e-02 (4.66e-02)
	$\mu = 50$		**3.47e-02** (3.77e-02)	1.00e-01 (1.47e-02)	1.10e-01 (1.90e-02)	1.00e-01 (1.48e-02)	5.86e-02 (4.32e-02)
P	$\mu = 5$	$\sigma^2 = \mu$	2.49e-02 (3.14e-02)	**1.59e-02** (2.19e-02)	1.85e-02 (2.34e-02)	1.85e-02 (2.25e-02)	2.64e-02 (3.36e-0)
	$\mu = 20$		2.73e-02 (3.62e-02)	**1.64e-02** (2.46e-02)	1.98e-02 (2.62e-02)	1.93e-02 (2.49e-02)	2.62e-02 (3.30e-02)
	$\mu = 50$		2.75e-02 (3.36e-02)	**1.64e-02** (2.23e-02)	1.89e-02 (2.34e-02)	1.89e-02 (2.28e-02)	2.62e-02 (3.48e-02)
NB	$\mu = 5$	$\sigma^2 = 2\mu$	1.99e-01 (7.52e-02)	1.86e-01 (6.53e-02)	4.08e-02 (4.32e-02)	3.82e-02 (3.85e-02)	**3.81e-02** (3.84e-02)
	$\mu = 20$		2.08e-01 (7.05e-02)	1.91e-01 (6.49e-02)	4.35e-02 (3.98e-02)	4.42e-02 (3.94e-02)	**4.28e-02** (3.78e-02)
	$\mu = 50$		2.13e-01 (6.66e-02)	1.92e-01 (6.11e-02)	4.35e-02 (3.84e-02)	4.48e-02 (3.79e-02)	**4.27e-02** (3.60e-02)
GP	$\mu = 5$	$\sigma^2 = 4\mu$	6.91e-01 (1.19e-01)	6.65e-01 (1.09e-01)	4.27e-02 (4.95e-02)	**3.58e-02** (4.09e-02)	3.83e-02 (4.16e-02)
	$\mu = 20$		7.85e-01 (1.23e-01)	7.52e-01 (1.08e-01)	3.61e-02 (3.79e-02)	3.33e-02 (3.53e-02)	**3.33e-02** (3.56e-02)
	$\mu = 50$		8.33e-01 (1.18e-01)	9.15e-01 (4.00e+00)	3.96e-02 (4.61e-02)	1.63e-01 (4.02e+00)	**3.56e-02** (3.92e-02)
GNB	$\mu = 5$	$\sigma^2 = 8\mu$	1.60e+00 (2.93e-01)	1.56e+00 (2.65e-01)	4.97e-02 (5.16e-02)	**3.61e-02** (4.28e-02)	4.00e-02 (4.46e-02)
	$\mu = 20$		2.07e+00 (2.68e-01)	2.01e+00 (2.53e-01)	3.73e-02 (3.80e-02)	**3.16e-02** (3.31e-02)	3.31e-02 (3.41e-02)
	$\mu = 50$		2.29e+00 (2.97e-01)	1.07e+01 (2.98e+01)	3.75e-02 (3.86e-02)	8.73e+00 (3.04e+01)	**3.53e-02** (3.63e-02)

Table C3 Means (and standard deviations) of KL divergences for different INAR(1) models based on quasi Pólya thinning operator for time series of length $T = 50$ and autocorrelation $\rho = 0.3$. The rows indicate the marginal distribution of the simulated INAR(1) model whereas the columns indicate the marginal distribution of the estimated INAR(1) model. Means of KL divergences (and standard deviations) have been computed for the 5 different estimated models, on $n = 1000$ simulations for each of the 15 scenarios. The minimal value of each scenario is written in bold.

Marginal	Mean	Variance	B	P	NB	GP	GNB
B	$\mu = 5$	$\sigma^2 = \mu/2$	1.30e+03 (1.31e+04)	1.70e-01 (1.38e-01)	1.84e-01 (1.50e-01)	**1.70e-01** (1.38e-01)	3.30e-01 (5.07e-01)
	$\mu = 20$		2.73e-01 (3.58e-01)	**1.75e-01** (1.66e-01)	1.87e-01 (1.74e-01)	1.77e-01 (1.64e-01)	2.10e-01 (2.68e-01)
	$\mu = 50$		3.37e-01 (5.53e-01)	**1.83e-01** (1.70e-01)	2.05e-01 (2.04e-01)	1.95e-01 (1.87e-01)	2.36e-01 (2.80e-01)
P	$\mu = 5$	$\sigma^2 = \mu$	2.19e-01 (3.12e-01)	1.40e-01 (2.03e-01)	1.47e-01 (2.12e-01)	**1.38e-01** (1.97e-01)	1.89e-01 (5.23e-01)
	$\mu = 20$		2.77e-01 (4.42e-01)	**1.26e-01** (1.89e-01)	1.33e-01 (1.96e-01)	1.27e-01 (1.82e-01)	1.89e-01 (2.99e-01)
	$\mu = 50$		2.92e-01 (4.50e-01)	**1.26e-01** (1.83e-01)	1.35e-01 (1.91e-01)	1.31e-01 (1.71e-01)	2.05e-01 (3.58e-01)
NB	$\mu = 5$	$\sigma^2 = 2\mu$	5.25e-01 (4.61e+00)	3.22e-01 (3.16e-01)	1.98e-01 (2.88e-01)	**1.62e-01** (2.31e-01)	1.71e-01 (2.44e-01)
	$\mu = 20$		4.71e-01 (6.70e-01)	3.70e-01 (3.70e-01)	2.49e-01 (2.83e-01)	**2.03e-01** (2.44e-01)	2.25e-01 (2.74e-01)
	$\mu = 50$		5.00e-01 (5.84e-01)	3.80e-01 (3.51e-01)	2.75e-01 (2.88e-01)	**2.35e-01** (2.59e-01)	2.58e-01 (2.80e-01)
GP	$\mu = 5$	$\sigma^2 = 4\mu$	5.74e+00 (1.44e+02)	8.87e-01 (6.93e-01)	2.50e-01 (3.49e-01)	**1.95e-01** (2.77e-01)	2.00e-01 (2.88e-01)
	$\mu = 20$		1.02e+00 (7.72e-01)	9.34e-01 (6.28e-01)	2.50e-01 (3.08e-01)	**1.86e-01** (2.39e-01)	1.98e-01 (3.30e-01)
	$\mu = 50$		1.10e+00 (8.67e-01)	1.13e+00 (4.93e+00)	3.44e-01 (3.32e-01)	3.95e-01 (4.93e+00)	**2.44e-01** (2.96e-01)
GNB	$\mu = 5$	$\sigma^2 = 8\mu$	5.47e+01 (7.57e+02)	1.83e+00 (1.96e+00)	6.52e+00 (1.83e+02)	**2.39e-01** (4.48e-01)	1.23e+00 (2.09e+01)
	$\mu = 20$		2.26e+00 (1.70e+00)	2.13e+00 (1.15e+00)	2.26e-01 (3.14e-01)	**1.54e-01** (2.13e-01)	1.65e-01 (2.56e-01)
	$\mu = 50$		2.40e+00 (1.11e+00)	7.10e+00 (2.96e+01)	2.72e-01 (3.16e-01)	5.18e+00 (2.99e+01)	**2.25e-01** (1.19e+00)

Table C4 Means (and standard deviations) of KL divergences for different INAR(1) models based on quasi Pólya thinning operator for time series of length $T = 50$ and autocorrelation $\rho = 0.9$. The rows indicate the marginal distribution of the simulated INAR(1) model whereas the columns indicate the marginal distribution of the estimated INAR(1) model. Means of KL divergences (and standard deviations) have been computed for the 5 different estimated models, on $n = 1000$ simulations for each of the 15 scenarios. The minimal value of each scenario is written in bold.

References

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

Al-Osh M, Alzaid A (1991) Binomial autoregressive moving average models. Stochastic Models 7(2):261–282

Al-Osh MA, Aly EEA (1992) First order autoregressive time series with negative binomial and geometric marginals. Communications in Statistics-Theory and Methods 21(9):2483–2492

Alzaid A, Al-Osh M (1990) An integer-valued pth-order autoregressive structure (inar (p)) process. Journal of Applied Probability 27(2):314–324

Alzaid A, Al-Osh M (1993) Some autoregressive moving average processes with generalized poisson marginal distributions. Annals of the Institute of Statistical Mathematics 45(2):223–232

Billingsley P (1961) Statistical methods in markov chains. The annals of Mathematical Statistics pp 12–40

Chu W (2010) Elementary proofs for convolution identities of Abel and Hagen–Rothe. The Electronic Journal of Combinatorics 24(1)

Consul P, Mittal S (1975) A new urn model with predetermined strategy. Biometrische Zeitschrift 17(2):67–75

Consul PC, Famoye F (2006) Lagrangian Probability Distributions. Springer

Consul PC, Jain GC (1973) A generalization of the poisson distribution. Technometrics 15(4):791–799

Gupta RC (1974) Modified power series distribution and some of its applications. Sankhyā: The Indian Journal of Statistics, Series B pp 288–298

Harvey AC, Fernandes C (1989) Time series models for count or qualitative observations. Journal of Business & Economic Statistics 7(4):407–417

Janardan K (1973) A new four urn model with predetermined strategy. In: Technical Report 37-I. Sangamon State University Springfield IL

Janardan K, Rao RB (1986) Identifiability of a generalized markov-polya damage model. Statistics: A Journal of Theoretical and Applied Statistics 17(2):303–310

Janardan KG, Schaeffer DJ (1977) A generalization of markov-polya distribution its extensions and applications. Biometrical Journal 19(2):87–106

Joe H (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class. Journal of Applied Probability 33(3):664–677

Joe H (1997) Multivariate Models and Multivariate Dependence Concepts. CRC press

Johnson N, Kotz S, Balakrishnan N (1997) Discrete Multivariate Distributions. Wiley Series in Probability and Statistics, Wiley, URL https://books.google.fr/books?id=B4F9QgAACAAJ

Jorgensen B (1997) The Theory of Dispersion Models. CRC Press

Jourdan A, Kokonendji CC (2002) Surdispersion et modèle binomial négatif généralisé. Revue de Statistique Appliquée 50(3):73–86

Latour A (1998) Existence and stochastic structure of a non-negative integer-valued autoregressive process. Journal of Time Series Analysis 19(4):439–455

MacDonald IL, Zucchini W (2016) Hidden markov models for discrete-valued time series. Handbook of Discrete-Valued Time Series pp 267–286

McKenzie E (1985) Some simple models for discrete variate time series 1. JAWRA Journal of the American Water Resources Association 21(4):645–650

McKenzie E (1987) Innovation distributions for gamma and negative binomial autoregressions. Scandinavian Journal of Statistics pp 79–85

Patil G, Ratnaparkhi M (1975) Problems of damaged random variables and related characterizations. In: A Modern Course on Statistical Distributions in Scientific Work. Springer, p 255–270

Peyhardi J (2023) On quasi pólya thinning operator. Brazilian Journal of Probability and Statistics

Rao CR (1965) On discrete distributions arising out of methods of ascertainment. Sankhyā: The Indian Journal of Statistics, Series A pp 311–324

Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (nginar (1)) process. Journal of Statistical Planning and Inference 139(7):2218–2226

Sellers KF, Peng SJ, Arab A (2020) A flexible univariate autoregressive time-series model for dispersed count data. Journal of Time Series Analysis 41(3):436–453

Weiß CH (2008a) The combined inar (p) models for time series of counts. Statistics & Probability Letters 78(13):1817–1822

Weiß CH (2008b) Thinning operations for modeling time series of counts—a survey. AStA Advances in Statistical Analysis 92(3):319–341

Zheng H, Basawa IV, Datta S (2007) First-order random coefficient integer-valued autoregressive processes. Journal of Statistical Planning and Inference 137(1):212–229