TEST
DOI 10.1007/s11749-017-0536-4

ORIGINAL PAPER

An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion

Marcelo Bourguignon[1] · **Christian H. Weiß**[2]

Received: 6 December 2016 / Accepted: 17 April 2017

Abstract We present a novel first-order nonnegative integer-valued autoregressive model for stationary count data processes with Bernoulli-geometric marginals based on a new type of generalized thinning operator. It can be used for modeling time series of counts with equidispersion, underdispersion and overdispersion. The main properties of the model are derived, such as probability generating function, moments, transition probabilities and zero probability. The maximum likelihood method is used for estimating the model parameters. The proposed model is fitted to time series of counts of iceberg orders and of cases of family violence illustrating its capabilities in challenging cases of overdispersed and equidispersed count data.

Keywords INAR(1) process · Bernoulli distribution · Geometric distribution · Integer-valued time series · Binomial thinning · Negative binomial thinning

Mathematics Subject Classification 60J10 · 62M05 · 62M10

1 Introduction

During the last three decades, several thinning-based models have been developed for the analysis of time series of counts (Weiß 2008). The first works concentrated on the use of the Poisson distribution as an integral feature of the process. McKen-

✉ Christian H. Weiß
weissc@hsu-hh.de

Marcelo Bourguignon
m.p.bourguignon@gmail.com

[1] Departamento de Estatística, Universidade Federal Do Rio Grande Do Norte, Natal, RN, Brazil

[2] Department of Mathematics and Statistics, Helmut Schmidt University, 22008 Hamburg, Germany

Published online: 29 April 2017

zie (1985)) and, independently, Al-Osh and Alzaid (1987) introduced the first-order integer-valued autoregressive [INAR(1)] model based on the binomial thinning operator (Steutel and van Harn 1979) with Poisson marginal, called Poisson INAR(1) process. In practice, however, the Poisson distribution is not always suitable for modeling, because one commonly observes overdispersed or underdispersed counts (i. e., counts having a variance being larger/smaller than their mean). Therefore, several alternatives to models with a Poisson marginal have been proposed in the literature.

A simple approach based on the binomial thinning is to only change the innovations' distribution in such a way that the marginal distribution of the process is underdispersed or overdispersed. Following this path, Schweer and Weiß (2014) investigated an INAR(1) process with compound Poisson innovations (and, hence, compound Poisson observations), which is suitable for modeling time series with overdispersion, while the INAR(1) model considered by Weiß (2013) exhibits underdispersion. Jazi et al. (2012) discussed an INAR(1) process with zero-inflated Poisson innovations, while Bourguignon and Vasconcellos (2015) considered the case of power series innovations. Kim and Lee (2017) introduced an INAR(1) model with Katz family innovations [INARKF(1)], which can be used for modeling time series of counts with equidispersion, underdispersion and overdispersion. However, the marginal distribution of the INARKF(1) process is rather complicated and does not have closed form.

Instead of only modifying the innovations' distribution, one may also use a different type of thinning operator. A widely discussed instance is the new geometric INAR(1) model introduced by Ristić et al. (2009), which uses the negative binomial thinning operator and has geometric marginal distribution (thus overdispersion). Our aim in this paper is to introduce a novel type of INAR(1) model based on a new generalized thinning operator. The generated process is stationary with Bernoulli-geometric marginals and, thus, allows to model nonnegative integer-valued time series with equidispersion, underdispersion and overdispersion having the same autocorrelation structure as the conventional AR(1) model, i. e., the autoregressive models of order 1. To create the new type of generalized thinning operator, we use a convolution of the Bernoulli and geometric distribution such that the counting series may show equidispersion, overdispersion or underdispersion. Therefore, our thinning operator extends the binomial thinning (Steutel and van Harn 1979) and negative binomial thinning operator (Ristić et al. 2009) in such a way that it is possible to consider equidispersion, overdispersion or underdispersion simultaneously for both the counting within the thinning operator and the marginal distribution of the process.

This paper is organized as follows. Section 2 introduces the BerG distribution as a convolution of the Bernoulli and geometric distribution, and uses it to construct new type of thinning operator. Section 3 applies this thinning operator to construct a new model for AR(1)-like count data processes, which includes the popular INAR(1) models based on binomial thinning or based on negative binomial thinning as special cases. Section 4 discusses parameter estimation and presents a simulation study to investigate the performance of these approaches. Section 5 analyzes two real data examples from different areas. Finally, we conclude in Sect. 6.

2 The BerG distribution

Let $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{Z}$ denote the sets of positive integers, nonnegative integers and all integers, respectively. All random variables will be defined on the same probability space. In Sect. 2.1, we discuss the distribution of the convolution of a Bernoulli and a geometric random variable, which we refer to as the BerG distribution. This distribution is used in Sect. 2.2 to define a novel thinning operation, BiNB thinning, which contains the well-known binomial thinning and negative binomial thinning as a boundary case. Later in Sect. 3, we use this thinning operation to define a new and quite flexible model for autoregressive count data time series.

2.1 Properties of the BerG distribution

The analysis of the convolution of binomial and negative binomial random variables, say $Z^{(n)} := X^{(n)} + Y^{(n)}$ with $X^{(n)} \sim \text{Bin}(n, \pi)$ and $Y^{(n)} \sim \text{NB}\big(n, 1/(1+\mu)\big)$ with $n \in \mathbb{N}, 0 < \pi < 1$ and $\mu > 0$, dates back to Kemp (1979); we denote this distribution as $\text{BiNB}(n, \pi, \mu)$. This distribution is a special case of the generalized inverse trinomial distribution, it corresponds to the $\text{GIT}_{3,1}(n; p_1, p_2, p_3)$ model with $p_1 := \frac{1-\pi}{\mu+1}$, $p_2 := \frac{\pi}{\mu+1}$, $p_3 := \frac{\mu}{\mu+1}$. Properties of this distribution are summarized in Sect. 4.2 of Aoyama et al. (2008).

In this article, our main interest is in the case $n = 1$, i. e., in the sum $Z := X + Y$, where X and Y are independent and follow the Bernoulli distribution $\text{Ber}(\pi)$ and geometric distribution $\text{Geom}\big(1/(1+\mu)\big)$ with means $0 < \pi < 1$ and $\mu > 0$, respectively. The distribution of Z, which we refer to as $\text{BerG}(\pi, \mu)$, has support $\mathbb{N}_0$, i. e., it is a count data distribution. The stochastic properties of the $\text{BerG}(\pi, \mu)$ distribution immediately follow from the properties of the $\text{GIT}_{3,1}\big(1; \frac{1-\pi}{\mu+1}, \frac{\pi}{\mu+1}, \frac{\mu}{\mu+1}\big)$ distribution given in Sect. 4.2 of Aoyama et al. (2008). In particular, its probability mass function (PMF) is given by

$$\Pr(Z = z) = \begin{cases} \frac{1-\pi}{1+\mu}, & \text{if} \quad z = 0, \\ (\mu + \pi)\frac{\mu^{z-1}}{(1+\mu)^{z+1}}, & \text{if} \quad z = 1, 2, \ldots, \end{cases} \tag{1}$$

and the probability generating function (PGF) of Z, denoted by $\varphi_Z(s) := \text{E}[s^Z]$, is given by

$$\varphi_Z(s) = \frac{1 - \pi(1-s)}{1 + \mu(1-s)}. \tag{2}$$

The mean and variance of the BerG distribution defined in (1) are

$$\text{E}(Z) \equiv \mu_Z = \pi + \mu \quad \text{and} \quad \text{Var}(Z) \equiv \sigma_Z^2 = \pi(1-\pi) + \mu(1+\mu). \tag{3}$$

Thus, the dispersion index, which is the variance-to-mean ratio, is given by

$$I_Z := \frac{\sigma_Z^2}{\mu_Z} = 1 + \mu - \pi.$$

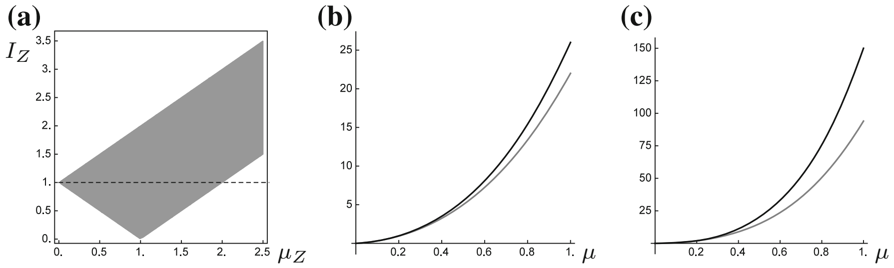

Fig. 1 **a** Possible mean-dispersion combinations (μ_Z, I_Z) for BerG distribution; *dashed line*: equidispersion. **b** $E(Z^3)$ and **c** $E(Z^4)$ against μ for equidispersed distributions BerG(μ, μ) (*black*) and Poi(2μ) (*gray*)

It follows that this distribution shows equidispersion for $\mu = \pi$ (but differs from the Poisson distribution, see below), while we have underdispersion for $\mu < \pi$, and overdispersion for $\mu > \pi$.

The attainable range of means μ_Z and dispersion indexes I_Z is illustrated by Fig. 1a, where the dashed line represents the case of equidispersion. For large mean values, only overdispersion can be achieved for the BerG distribution.

More generally, the rth factorial moment $\mu_{(r)} \equiv E\big[Z \ldots (Z - r + 1)\big]$ with $r \in \mathbb{N}$ equals

$$\mu_{(r)} = \mu^{r-1}\,(\pi + \mu)\,r!. \tag{4}$$

In particular, the third and fourth moments of Z are

$$E(Z^3) = (\pi+\mu)\,\big[1+6\mu(1+\mu)\big] \quad \text{and} \quad E(Z^4) = (\pi+\mu)\,(1+2\mu)\,\big[1+12\mu(1+\mu)\big]. \tag{5}$$

The formulae (4) and (5) show that the equidispersed BerG distribution, i. e., the BerG(μ, μ) distribution with $0 < \mu < 1$, does not coincide with its Poisson counterpart Poi(2μ). Furthermore, the pgf of the Poi(2μ) distribution is $\exp[-2\,\mu(1 - s)]$, being different from the pgf shown in (2) if $\pi = \mu$. As illustrated in Fig. 1b and c, the differences between their third and fourth moments increase with increasing μ.

Remark 1 (Relation to Zero-Modified Geometric Distribution) The *zero-modified geometric distribution* ZMG(κ, μ) with $\mu > 0$ and $\kappa \in (-1/\mu, 1)$ is defined by the PGF

$$\varphi_Z(s) = \kappa \;+\; (1 - \kappa)\,\big[1 + \mu(1 - s)\big]^{-1}$$

and the PMF

$$\Pr(Z = z) = \mathbb{I}_{\{z=0\}}\,\kappa \;+\; (1 - \kappa)\,\frac{1}{1 + \mu}\,\left(\frac{\mu}{1 + \mu}\right)^{z},$$

see Sect. 8.2.3 in Johnson et al. (2005). The moments $E[Z^n]$ are obtained from those of the geometric parent distribution Geom$\big(1/(1+\mu)\big)$, say μ_n, by computing $E[Z^n] = (1 - \kappa)\,\mu_n$. In particular,

$$\mathrm{E}(Z) = (1-\kappa)\,\mu, \quad \mathrm{Var}(Z) = (1-\kappa)\,\mu(1+\mu)\ +\ \kappa(1-\kappa)\,\mu^2.$$

If $\kappa > 0$, then the geometric's zero probability is increased (*zero-inflation*), while it is decreased for $\kappa < 0$ (*zero-deflation*). The variance-to-mean ratio is

$$\frac{\mathrm{Var}(Z)}{\mathrm{E}(Z)} = 1 + \mu\,(1+\kappa),$$

which is < 1 (underdispersion) if $\kappa < -1$, and vice versa.

Comparing with (2) and (1), it becomes clear that the BerG(π, μ) distribution is a special type of zero-deflated ZMG(κ, μ)-distribution, where $\kappa := -\pi/\mu$. This confirms that we have underdispersion for $\mu < \pi$, and vice versa.

2.2 The BiNB thinning operation

The BerG distribution discussed in Sect. 2.1 is now used to construct a new type of generalized thinning operator (Latour 1998; Weiß 2008), with the counting series consisting of independent and identically distributed (i. i. d.) BerG random variables.

Definition 1 *(BiNB thinning)* Let Z be a nonnegatively integer-valued random variable, and let $\alpha, \beta \geq 0$ be real numbers such that $\alpha + \beta \in [0, 1)$. Then the *BiNB thinning operator*, denoted by $(\alpha, \beta)\circledast$, is defined as

$$(\alpha, \beta) \circledast\ Z\ := \sum_{J=1}^{Z} W_J,$$

where $\{W_J\}_{J=1}^{Z}$ is a sequence of i. i. d. BerG(α, β) random variables with mean $\alpha+\beta$, being mutually independent of Z. Note that for $Z = 0$, the empty sum is defined as 0.

Note that the restriction "$\alpha + \beta \in [0, 1)$" in Definition 1 could be relaxed to requiring $\alpha < 1$, since the BerG(α, β) is well defined for any $\beta \geq 0$. But to make the operation "$\circledast$" a "thinning" (i.e., to guarantee that $E\big[(\alpha, \beta) \circledast\ Z\big] < E[Z]$), and since we need this restriction later anyway for obtaining a stationary solution, we already used the more restrictive requirement "$\alpha + \beta \in [0, 1)$" here.

Note that BiNB thinning is not only a generalized thinning operator in the sense of Latour (1998), Weiß (2008), it also constitutes an instance of extended thinning as defined by Zhu and Joe (2003) for some parameterizations.

By construction, the conditional distribution of $(\alpha, \beta) \circledast\ Z$ given Z is the BiNB(Z, α, β) distribution (or equivalently, the GIT$_{3,1}\big(Z;\ \frac{1-\alpha}{\beta+1}, \frac{\alpha}{\beta+1}, \frac{\beta}{\beta+1}\big)$ distribution, see Sect. 2.1), such that conditional properties of $(\alpha, \beta)\circledast Z$ given Z immediately follow from the results given in Sect. 4.2 of Aoyama et al. (2008). For instance, the conditional PGF equals

$$\varphi_{(\alpha,\beta)\circledast\ Z|Z}(s) \equiv E\big[s^{(\alpha,\beta)\circledast\ Z}\ \big|\ Z\big] = \left[\frac{1-\alpha(1-s)}{1+\beta(1-s)}\right]^{Z}. \tag{6}$$

The PMF of $(\alpha, \beta) \circledast Z | Z = z$ is given by (Aoyama et al. 2008)

$$\Pr\big((\alpha, \beta) \circledast z\big) = \sum_{i=0}^{\min\{z,k\}} \frac{z}{z+k-i} \binom{z+k-i}{z-i, i, k-i} \left(\frac{1-\alpha}{1+\beta}\right)^{z-i} \left(\frac{\alpha}{1+\beta}\right)^{i} \left(\frac{\beta}{1+\beta}\right)^{k-i}, \tag{7}$$

for $k = 0, 1, 2, \ldots$, where $\binom{x}{x_1, x_2, x_3} = x!/(x_1!x_2!x_3!)$.

Proposition 1 *Let Z be a nonnegatively integer-valued random variable, and let $\alpha, \beta \geq 0$ with $\alpha + \beta \in [0, 1)$. Then,*

$$(\alpha, \beta) \circledast Z = \sum_{j=1}^{Z} (B_j + G_j) \overset{D}{=} \sum_{j=1}^{Z} B_j + \sum_{j=1}^{Z} G_j = \alpha \circ Z + \beta * Z, \tag{8}$$

where the operator "$\circ$" is the binomial thinning *operator* (Steutel and van Harn 1979) *with the $\{B_j\}_{j=1}^{Z}$ being i. i. d.* $\text{Ber}(\alpha)$ *random variables, independent of Z, and where "$*$" denotes the* negative binomial thinning *operator* (Ristić et al. 2009) *with the $\{G_j\}_{j=1}^{Z}$ being i. i. d.* $\text{Geom}\big(1/(1+\beta)\big)$ *random variables, independent of Z. The counting series in $\alpha \circ Z$ and $\beta * Z$ are mutually independent random variables.*

Note that as discussed after Definition 1, it would be possible to relax the requirement "$\alpha + \beta \in [0, 1)$" to $\alpha < 1$.

It is interesting to note that the right side of Eq. (8) is the sum of two know thinning operators. The first term represents the binomial thinning operator proposed by Steutel and van Harn (1979), where the counting sequence has the Bernoulli distribution with mean $\alpha \in [0, 1)$, and the second term represents the negative binomial thinning operator introduced by Ristić et al. (2009), where the counting sequence has the geometric distribution (with mean $\beta \in [0, 1)$ if considering "$*$" to be a "thinning"). In particular, this implies that both binomial thinning and negative binomial thinning are just boundary cases of BiNB thinning ($\beta = 0$ and $\alpha = 0$, respectively). Nastić et al. (2016) introduced a thinning operator which is a mixture of the Bernoulli and geometric distributed random variables. However, this operator is different from the new BiNB thinning operator.

Utilizing the relation to the $\text{GIT}_{3,1}$-distribution (Aoyama et al. 2008, Sect. 4.2), it immediately follows that the conditional rth factorial moment of $(\alpha, \beta) \circledast Z$ given Z equals

$$Z \sum_{i=0}^{\min\{r,Z\}} \binom{r}{i} \frac{(Z+r-i-1)!}{(Z-i)!} \alpha^i \beta^{r-i},$$

in particular,

$$E\big[(\alpha, \beta) \circledast Z \mid Z\big] = Z\,(\alpha + \beta), \quad V\big[(\alpha, \beta) \circledast Z \mid Z\big] = Z\,\big[\alpha(1-\alpha) + \beta(1+\beta)\big]. \tag{9}$$

So by the laws of total expectation and total variance, it follows that

$$E\big[(\alpha,\beta)\circledast\ Z\big] = \mu_Z\,(\alpha+\beta),\quad V\big[(\alpha,\beta)\circledast\ Z\big]$$
$$= \sigma_Z^2\,(\alpha+\beta)^2 + \mu_Z\,\big[\alpha(1-\alpha)+\beta(1+\beta)\big],$$

also see p. 240 in Zhu and Joe (2003). If the thinned random variable Z follows itself a BerG distribution, then the following result holds.

Proposition 2 *Let* $Z \sim \mathrm{BerG}(\pi,\mu)$ *and* $\circledast$ *be the BiNB thinning operator defined above. Then,* $(\alpha,\beta)\circledast Z \sim \mathrm{ZMG}\big(\kappa,\ \beta+\mu(\alpha+\beta)\big)$ *with* $\kappa := [\beta-\pi(\alpha+\beta)]/[\beta+\mu(\alpha+\beta)]$, *which is zero-deflated for* $\pi > \beta/(\alpha+\beta)$, *and vice versa.*

Proof Using (2) and comparing with Remark 1, we obtain that

$$\varphi_{(\alpha,\beta)\circledast Z}(s) = \varphi_Z\left(\frac{1-\alpha(1-s)}{1+\beta(1-s)}\right) = \frac{\frac{1+\beta(1-s)-\beta\ \pi(1-s)-\alpha\ \pi(1-s)}{1+\beta(1-s)}}{\frac{1+\beta(1-s)+\beta\ \mu(1-s)+\alpha\ \mu(1-s)}{1+\beta(1-s)}}$$
$$= \frac{1+[\beta-\pi(\alpha+\beta)](1-s)}{1+[\beta+\mu(\alpha+\beta)](1-s)} = \kappa\ +\ \frac{1-\kappa}{1+[\beta+\mu(\alpha+\beta)](1-s)},$$

which completes the proof. □

We conclude this section by considering the boundary cases of binomial thinning ($\beta = 0$) and negative binomial thinning ($\alpha = 0$), respectively. In the case of binomial thinning, $\beta = 0$ implies that we are always concerned with a zero-*deflated* type of ZMG-distribution. In fact, Proposition 2 can be further simplified.

Corollary 3 *Let* $Z \sim \mathrm{BerG}(\pi,\mu)$ *and* $\circ$ *be binomial thinning operator defined above. Then,* $\alpha\circ Z \sim \mathrm{BerG}(\alpha\,\pi,\alpha\,\mu)$.

Proof Note that (remember (2))

$$\varphi_{\alpha\circ Z}(s) = \varphi_Z(1-\alpha(1-s)) = \frac{1-\pi\{1-[1-\alpha(1-s)]\}}{1+\mu\{1-[1-\alpha(1-s)]\}} = \frac{1-\alpha\ \pi(1-s)}{1+\alpha\ \mu(1-s)}.$$

□

Corollary 3 shows that the BerG distribution is invariant under binomial thinning.

In the case of negative binomial thinning ($\alpha = 0$), in contrast, we are always concerned with a zero-*inflated* type of ZMG-distribution (Remark 1), since $\pi < 1$ in Proposition 2.

Corollary 4 *Let* $Z \sim \mathrm{BerG}(\pi,\mu)$ *and* $*$ *be negative binomial thinning operator defined above. Then,* $\beta * Z \sim \mathrm{ZMG}\big((1-\pi)/(1+\mu),\ (1+\mu)\beta\big)$, *which is a zero-inflated geometric distribution.*

In Sect. 3, we shall use the new BiNB thinning operator "$(\alpha,\beta)\circledast$" to construct a model for count data processes having the same autocorrelation structure as the conventional AR(1) models, i. e., the autoregressive models of order 1.

3 BerG-INAR(1) processes with equi-, under- or overdispersion

In this section, we introduce a novel first-order nonnegative integer-valued autoregressive (INAR(1)) model for stationary count data processes with Bernoulli-geometric marginals. It shall turn out, see Sect. 3.2, that the proposed model contains the famous INAR(1) model by McKenzie (1985) as a special case, but also the INAR(1)-type model by Ristić et al. (2009) (for the latter, see Sect. 3.3).

3.1 INAR(1) model based on BiNB thinning

Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a count data process. We consider the case where the counts Z_t follow a BerG distribution, thus being able to describe underdispersion, equidispersion or overdispersion, see Sect. 2.1. To mimic an AR(1) serial dependence structure, we use the BiNB thinning operator "$(\alpha,\beta)\circledast$" proposed in Sect. 2.2.

Definition 2 *(BerG-INAR(1)^{BiNB} process)* A discrete-time stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ is said to be a BiNB-thinning-based first-order integer-valued autoregressive (INAR(1) BiNB) process with marginals BerG(π,μ) if it satisfies the following equation

$$Z_t = (\alpha,\beta)\circledast Z_{t-1} + \varepsilon_t, \quad t\in\mathbb{Z}, \tag{10}$$

where $\alpha,\beta\geq 0$ with $\alpha+\beta\in(0,1)$, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an innovation sequence of i. i. d. nonnegative integer-valued random variables not depending on past values of $\{Z_t\}_{t\in\mathbb{Z}}$, $\{Z_t\}_{t\in\mathbb{Z}}$ is a stationary process with BerG(π,μ) marginals, i. e., with probability mass function given by Eq. (1). It is also assumed that the counting series of $(\alpha,\beta)\circledast Z_{t-1}$ is independent of other counting series, and, moreover, independent of the innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

By construction, the process described by $Z_t=(\alpha,\beta)\circledast Z_{t-1}+\varepsilon_t$ constitutes a homogeneous Markov chain. The subsequent proposition shows (under certain conditions) that it is possible to find a distribution of the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that the observations are BerG-distributed as required by Definition 2. So a stationary solution exists under these conditions. We shall see that the distribution of the innovations satisfies $P(\varepsilon_t=k)>0$ for all $k\in\mathbb{N}_0$, so all transition probabilities (as a convolution of the BiNB distribution and the ε_t's distribution) are truly positive [closed-form formulae are derived in (15) below]. Hence, we know that the Markov chain is irreducible and aperiodic. And since we proved the existence of a stationary marginal distribution (the BerG distribution according to Definition 2), this distribution is unique, and the corresponding Markov chain is even ergodic (Feller 1968).

Proposition 5 *If* $\pi<\frac{\beta}{\alpha+\beta}$ *and* $\mu>\frac{\beta}{1-\alpha-\beta}$*, then the distribution of the innovations sequence* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is a convolution between two independent random variables with* $Y_1\sim \text{BerG}(\pi,\beta-\pi(\alpha+\beta))$ *and* $Y_2\sim\text{ZMG}\big([\beta+\mu(\alpha+\beta)]/\mu,\ \mu\big)$*. In particular,* $P(\varepsilon_t=k)>0$ *for all* $k\in\mathbb{N}_0$ *holds.*

Note that Y_2 follows a zero-*inflated* geometric distribution according to Remark 1. The conditions on the parameters given in Proposition 5 should not be violated. For example, in the boundary case $\mu = \frac{\beta}{1-\alpha-\beta}$, the distribution of Y_2 becomes ZMG$(1, \mu)$, so $Y_2 \equiv 0$ with probability 1. If even $\mu < \frac{\beta}{1-\alpha-\beta}$, then the ZMG's inflation parameter would be larger than 1, which is not possible. An analogous argumentation holds with respect to Y_1, where $\pi \geq \frac{\beta}{\alpha+\beta}$ causes the geometric parameter of the BerG distribution to be ≤ 0.

Proof Using Proposition 2, we have

$$\begin{aligned}\varphi_\varepsilon(s) &= \frac{\varphi_X(s)}{\varphi_X\left(\frac{1-\alpha(1-s)}{1+\beta(1-s)}\right)} = \frac{1-\pi(1-s)}{1+\mu(1-s)} \cdot \frac{1+[\beta+\mu(\alpha+\beta)](1-s)}{1+[\beta-\pi(\alpha+\beta)](1-s)} \\ &= \frac{1-\pi(1-s)}{1+[\beta-\pi(\alpha+\beta)](1-s)} \cdot \frac{1+[\beta+\mu(\alpha+\beta)](1-s)}{1+\mu(1-s)} = \varphi_1(s)\, \varphi_2(s),\end{aligned}$$

where, under the above condition, $\varphi_1(s) = \frac{1-\pi(1-s)}{1+[\beta-\pi(\alpha+\beta)](1-s)}$ and $\varphi_2(s) = \frac{1+[\beta+\mu(\alpha+\beta)](1-s)}{1+\mu(1-s)}$ are the PGF of Y_1 and Y_2, respectively. □

Let us now discuss some properties of the new BerG-INAR(1)BiNB process satisfying the constraints

$$\alpha, \beta \geq 0, \quad \mu, \pi, \alpha+\beta > 0, \quad \pi, \alpha+\beta < 1; \quad \pi < \tfrac{\beta}{\alpha+\beta}, \quad \mu > \tfrac{\beta}{1-\alpha-\beta}. \tag{11}$$

Using (6), the conditional PGF of Z_t given Z_{t-1} becomes

$$\varphi_{Z_t|Z_{t-1}}(s) = \left[\frac{1-\alpha(1-s)}{1+\beta(1-s)}\right]^{Z_{t-1}} \varphi_\varepsilon(s),$$

where $\varphi_\varepsilon(s)$ is the PGF of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Similarly, from (9), we obtain the conditional mean and variance as

$$\begin{aligned}E[Z_t \mid Z_{t-1}] &= (\alpha+\beta)\, Z_{t-1} \; + \; \mu_\varepsilon, \\ V[Z_t \mid Z_{t-1}] &= \big[\alpha(1-\alpha)+\beta(1+\beta)\big]\, Z_{t-1} \; + \; \sigma_\varepsilon^2.\end{aligned} \tag{12}$$

Here, the innovations have mean (see (3) and Remark 1)

$$\mathrm{E}(\varepsilon_t) \equiv \mu_\varepsilon = E[Y_1] + E[Y_2] = (\pi+\mu)\,(1-\alpha-\beta).$$

Similarly, the innovations' variance equals

$$\begin{aligned}\mathrm{Var}(\varepsilon_t) \equiv \sigma_\varepsilon^2 = V[Y_1] + V[Y_2] &= \big[\beta(1-\pi)-\pi\alpha\big]\big[1+\beta(1-\pi)-\pi\alpha\big] \\ &\quad + \pi(1-\pi) + \big(\mu-\beta(1+\mu)-\mu\alpha\big)\big[(1+\mu)(1+\beta)+\mu\alpha\big].\end{aligned}$$

According to (12), both the conditional mean and the variance are linear in the previous observation. In particular, the new BerG-INAR(1)BiNB model belongs to the class of CLAR(1) models (conditional linear autoregressive) as defined by Grunwald et al. (2000). As a consequence, the autocorrelation function (ACF) decays exponentially,

$$\mathrm{Corr}(Z_t, Z_{t-h}) \equiv \rho(h) = (\alpha+\beta)^h, \quad h \geq 0. \tag{13}$$

By using Proposition 5 together with (1) and Remark 1, the PMF of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is expressed as

$$\Pr(\varepsilon_t = 0) = \left(\frac{1-\pi}{1+\beta(1-\pi)-\pi\alpha}\right)\cdot\left(\frac{1+\beta(1+\mu)+\mu\alpha}{1+\mu}\right),$$

and

$$\begin{aligned}\Pr(\varepsilon_t = k) =& \frac{1-\pi}{1+\beta-\pi(\alpha+\beta)}\cdot\frac{[\mu-\beta-\mu(\alpha+\beta)]\mu^{k-1}}{(1+\mu)^{k+1}}\\ &+\frac{\mu-\beta-\mu(\alpha+\beta)}{\mu(1+\mu)}\cdot\frac{(1-\pi)[\beta-\pi(\alpha+\beta)]^k}{[1+\beta-\pi(\alpha+\beta)]^{k+1}}\\ &+\frac{[\beta-\pi(\alpha+\beta)+\pi][1-(\beta+\mu(\alpha+\beta))/\mu]\mu^k}{[\beta-\pi(\alpha+\beta)][1+\beta-\pi(\alpha+\beta)](1+\mu)^{k+1}}\\ &\times\frac{\left[\frac{(\pi\ \beta+\pi\ \alpha-\beta)(1+\mu)}{\mu(\pi\ \beta+\pi\ \alpha-\beta-1)}\right]^k \mu(\pi\ \beta+\pi\ \alpha-\beta-1)-(\pi\ \beta+\pi\ \alpha-\beta)(1+\mu)}{\mu+\pi\ \beta+\pi\ \alpha-\beta},\end{aligned} \tag{14}$$

for $k \geq 1$.

Remark 2 (Computation and Simulation) The marginal distribution of the observations (i. e., the BerG distribution), the conditional distribution of the BiNB thinning operation (i. e., the BiNB distribution) as well as the distribution of the BerG-INAR(1)BiNB's innovations are all convolutions of standard distributions. This makes it easy to implement these distributions in practice, i. e., to compute their PMF or to simulate them. In Sect. 4 below about parameter estimation, we used the R language (R Core Team 2016) for implementation. To simulate a BerG random variable, for instance, we simulate a Bernoulli and a geometric one with R's `rbinom` and `rgeom`, respectively, and then we take their sum. To compute the PMF as required, e. g., for maximum likelihood estimation (see Sect. 4 for details), it is not necessary to typewrite such complex formulae like (1), (7) or (14). Instead, one can just use R's `convolve(…, type="open")` together with, e. g., `dbinom` and `dgeom` in the case of the BerG distribution.

The transition probabilities $p(j|i) = \Pr(Z_t = j|Z_{t-1} = i)$ of the BerG-INAR(1)BiNB process are given by

$$\begin{aligned} p(j|0) &= \Pr(\varepsilon_t = j),\\ p(0|i) &= \left(\frac{1-\alpha}{1+\beta}\right)^i\cdot\left(\frac{1-\pi}{1+\beta(1-\pi)-\pi\alpha}\right)\cdot\left(\frac{1+\beta(1+\mu)+\mu\alpha}{1+\mu}\right), \quad i\geq 1\\ p(j|i) &= \sum_{k=0}^{j}\Pr\big((\alpha,\beta)\circledast\ Z_{t-1}=k\big|Z_{t-1}=i\big)\cdot\Pr(\varepsilon_t = j-k), \quad i,j\geq 1. \end{aligned} \tag{15}$$

Thus, we obtain that the transition probability from zero to (non-)zero equals

$$\eta := \Pr(Z_t \neq 0|Z_{t-1} = 0) = \frac{(1-\alpha)(\pi+\mu)}{(1+\mu)[1+\beta-(\alpha+\beta)\pi]}$$

and

$$1-\eta := \Pr(Z_t = 0|Z_{t-1} = 0) = \frac{(1-\pi)[1+\beta(1+\mu)+\mu\alpha]}{(1+\mu)[1+\beta(1-\pi)-\pi\alpha]},$$

respectively. The run length of zeros in the process, N, follows a geometric distribution with termination probability η, i. e., $\Pr(N = n) = \eta(1-\eta)^{n-1}, n \geq 1$. Thus, the average run length of zeros in the process is given by

$$\mathrm{E}(N) = \frac{(1+\mu)\left[1+\beta-(\alpha+\beta)\pi\right]}{(1-\alpha)(\pi+\mu)}.$$

We conclude this section by pointing out that for later analysis, the following reparameterization is advantageous:

$$\rho \in (0,1),\ \kappa \in [0,1] \quad \text{with } \rho := \alpha+\beta,\ \kappa := \tfrac{\alpha}{\alpha+\beta}, \tag{16}$$

which is solved as $\alpha = \kappa\rho$ and $\beta = (1-\kappa)\rho$. Then the parameter constraints (11) become

$$\mu > 0, \quad 0 < \pi, \rho < 1; \quad 0 \leq \kappa \leq 1, \quad \kappa < 1-\pi, \quad \mu(1-\rho) > (1-\kappa)\rho. \tag{17}$$

Note that $\rho(h) = \rho^h$, so ρ expresses the extent of autocorrelation.

3.2 INAR(1) model based on binomial thinning operator

The INAR(1)BiNB model recursion (10) reduces to the one of the well-known INAR(1) model as introduced by McKenzie (1985) if $\beta = 0$, i. e.,

$$Z_t = \alpha \circ Z_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

which corresponds to the case $\kappa = 1$ in terms of the parameterization (16). This model has been intensively studied in the literature, e. g., by Jazi et al. (2012) for count data time series showing an excessive number of zeros, by Schweer and Weiß (2014) for overdispersed counts, by Weiß (2013) for underdispersed counts, or by Moriña et al. (2011) and Bourguignon et al. (2016) for counts showing seasonality.

It has to be noted, however, that the condition given in Proposition 5 [also see (11)] cannot be satisfied if $\beta = 0$, since π cannot become smaller than 0 [an analogous contradiction follows from (17) if $\kappa = 1$]. This implies that the BerG distribution cannot be a marginal distribution (being preserved for any α) of the usual INAR(1) process. This is reasonable since the BerG distribution is not infinitely divisible (remember that it may exhibit underdispersion) and therefore not discrete self-decomposable (Steutel and van Harn 1979). But certainly, one may use an INAR(1) model together with BerG-distributed innovations.

3.3 INAR(1) model based on negative binomial thinning operator

Another special case of the INAR(1)BiNB model is obtained by setting $\alpha = 0$ in Definition 2, which corresponds to the case $\kappa = 0$ in terms of the parameterization (16). Then we obtain the INAR(1)$^{\text{NB}}$ model being based on the negative binomial thinning operator (Ristić et al. 2009). In contrast to McKenzie's INAR(1) model mentioned in Sect. 3.2 before, the INAR(1)$^{\text{NB}}$ model is able to generate BerG-distributed counts.

Definition 3 A discrete-time nonnegative integer-valued stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ is said to be an INAR(1)$^{\text{NB}}$ process with marginals BerG(π, μ) if it satisfies the following equation

$$Z_t = \beta * Z_{t-1} \; + \; \varepsilon_t, \quad t \in \mathbb{Z}.$$

where $\beta \in [0, 1)$, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an innovation sequence of i. i. d. nonnegative integer-valued random variables not depending on past values of $\{Z_t\}_{t\in\mathbb{Z}}$, $\{Z_t\}_{t\in\mathbb{Z}}$ is a stationary process with BerG(π, μ) marginals, i. e., with probability mass function given by Eq. (1). It is also assumed that the counting series of $\beta * Z_{t-1}$ is independent of other counting series, and, moreover, independent of the innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

Note that Ristić et al. (2009) used their INAR(1)$^{\text{NB}}$ model together with a geometric marginal distribution, referred to as the NGINAR(1) model by these authors, which corresponds to the boundary case $\pi = 0$ of the BerG-INAR(1)$^{\text{NB}}$ model.

We are able to adapt Proposition 5. Note that the condition given there can be restated as $\beta \in \left(\frac{\alpha\,\pi}{1-\pi}, \frac{\mu(1-\alpha)}{1+\mu}\right)$, which simplifies to $\beta \in \left(0, \frac{\mu}{1+\mu}\right)$ if $\alpha = 0$.

Corollary 6 *If $\beta < \mu/(1+\mu)$, then the distribution of the innovations sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a convolution between two independent random variables with $Y_1 \sim$ BerG$(\pi, \beta(1-\pi))$ and $Y_2 \sim$ ZMG$\big(\beta(1+\mu)/\mu,\ \mu\big)$.*

Note that Y_2 follows a zero-*inflated* geometric distribution according to Remark 1. The properties for this BerG-INAR(1)$^{\text{NB}}$ process follow from the ones given in Sect. 3.1 by setting $\alpha = 0$.

At this point, one may ask for the benefits of the full INAR(1)BiNB model (having four parameters) compared to the simplified INAR(1)$^{\text{NB}}$ model (having only three parameters since $\alpha = \kappa = 0$). Both models have the same marginal distribution, BerG(π, μ), and both are CLAR(1) models with the ACF being given by $\rho(h) = \rho^h$ (where $\rho = \alpha + \beta$ and $\rho = \beta$, respectively). So let us assume the marginal distribution to be fixed as BerG(π, μ).

Then, the extent of autocorrelation ρ for the reduced INAR(1)$^{\text{NB}}$ model is bound from above by $\mu/(\mu+1)$ (see Proposition 6), while it is only bounded by $\mu/(\mu+1-\kappa)$ with $\kappa < 1-\pi$ for the full INAR(1)BiNB model. So the full model, having the same marginal distribution as the reduced model, allows for stronger autocorrelation levels.

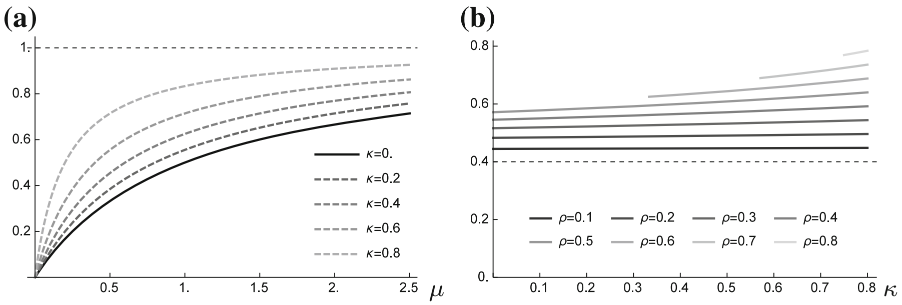

Fig. 2 **a** Upper bound for $\rho(1) = \rho$ against μ, where $\kappa = 0$ corresponds to INAR(1)$^{\text{NB}}$ model. **b** $\Pr(Z_t = 0|Z_{t-1} = 0)$ against κ for $(\mu, \pi) = (1, 0.2)$, where *dashed line* shows $\Pr(Z_t = 0) = 0.4$

This is illustrated by Fig. 2a, where the upper bound is plotted as a function of μ for different levels of κ, and where this upper bound moves toward 1 with increasing κ.

In a similar line, also the zero-zero transition probability, which determines the length of runs of zeros in the generated sample paths, is influenced by the restrictions imposed by the reduced INAR(1)$^{\text{NB}}$ model. From Eq. (15), we know that

$$\begin{aligned}\Pr(Z_t = 0|Z_{t-1} = 0) = \Pr(\varepsilon_t = 0) &= \frac{1-\pi}{1+\mu} \cdot \frac{1 + \beta(1+\mu) + \mu\alpha}{1 + \beta(1-\pi) - \pi\alpha} \\ &= \frac{1-\pi}{1+\mu} \cdot \frac{1 + \rho(1+\mu) - \kappa\rho}{1 + \rho(1-\pi) - \kappa\rho},\end{aligned}$$

where the respective first factor, $\frac{1-\pi}{1+\mu}$, is the marginal probability for observing a zero. For illustration, let us consider the model with marginal distribution BerG$(1, 0.2)$, which has marginal mean 1.2 and zero probability 0.4. Figure 2b shows the zero-zero transition probability as a function of κ for different levels of the autocorrelation parameter ρ. First note that the graphs for $\rho = 0.6, 0.7, 0.8$ are not defined in $\kappa = 0$, i. e., the reduced INAR(1)$^{\text{NB}}$ model does not even allow for such strong levels of autocorrelation. Then, it can be seen that with κ increasing toward $1 - \pi = 0.8$, the zero-zero transition probability further increases.

4 Parameter estimation

The novel BerG-INAR(1)BiNB model is able to describe count data processes having an autoregressive autocorrelation structure, and it allows the counts to show underdispersion, equidispersion or overdispersion. To make the model applicable in practice, approaches for parameter estimation are required. Such approaches are presented in Sect. 4.1, while Sect. 4.2 investigates these approaches in a small simulation study.

4.1 Approaches for parameter estimation

Let $Z_1, Z_2, \ldots, Z_n$ be a random sample of size n from a stationary BerG-INAR(1)BiNB process with PMF (1). To estimate the model parameters, a full maximum likelihood (ML) approach is easily implemented. Using that $Z_1 \sim \text{BerG}(\pi, \mu)$ and that the conditional distribution of Z_t given Z_{t-1} is computed from (15) (also remember Remark 2), the log-likelihood function is computed as

$$\ell(\mu, \pi, \alpha, \beta) = \log[\Pr(Z_1 = j)] + \sum_{t=2}^{n} \log\left[\Pr(Z_t = j | Z_{t-1} = i)\right]. \qquad (18)$$

The ML estimators $\widehat{\alpha}_{\text{ML}}$, $\widehat{\beta}_{\text{ML}}$, $\widehat{\pi}_{\text{ML}}$ and $\widehat{\mu}_{\text{ML}}$ of α, β, π and μ, respectively, are defined as the values of α, β, π and μ that maximize the log-likelihood function in (18). The resulting estimates are computed by a numerical optimization of the log-likelihood function (18) in $(\mu, \pi, \alpha, \beta)^\top$, which certainly has to consider the parameter constraints (11). To simplify these parameter constraints (11) and hence the numerical optimization, the following reparameterization is recommended:

$$\pi' := (\alpha + \beta)\, \pi, \quad \mu' := (1 - \alpha - \beta)\, \mu,$$

which is solved as $\pi = \pi'/(\alpha + \beta)$ and $\mu = \mu'/(1 - \alpha - \beta)$. Then (11) becomes

$$\alpha, \beta \geq 0, \quad \alpha + \beta < 1, \quad 0 < \pi' < \beta, \quad \mu' > \beta,$$

which are simple *linear* constraints. These can be used together with, e. g., R's `constrOptim` (R Core Team 2016).

For the numerical optimization, starting values for the model parameters $(\mu, \pi, \alpha, \beta)^\top$ (or $(\mu', \pi', \alpha, \beta)^\top$, respectively) are required. Starting values for (μ, π) might be obtained in two ways: In view of (3), the method of moments (MM) leads to

$$\widehat{\mu}_{\text{MM}} = \tfrac{1}{2}(\overline{Z} + \hat{I}_Z - 1) \quad \text{and} \quad \widehat{\pi}_{\text{MM}} = \tfrac{1}{2}(\overline{Z} - \hat{I}_Z + 1),$$

where $\overline{Z} = (1/n)\sum_{t=1}^{n} Z_t$, $\hat{I}_Z = \widehat{\gamma}(0)/\overline{Z}$ and $\widehat{\gamma}(k) = (1/n)\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t-k} - \overline{Z})$.

Proposition 7 *The method of moments yields consistent estimates.*

Proof Strong consistency of these estimators follows from the ergodicity of the process and the properties of the convergence in probability. □

Alternatively, from Eq. (1), we have that $P(Z_t > 0) = \frac{\pi+\mu}{1+\mu}$, $P(Z_t > 1) = \frac{\mu(\pi+\mu)}{(1+\mu)^2}$ and $\Pr(Z_t = 1) = \frac{\pi+\mu}{(1+\mu)^2}$. Thus, we have that $\frac{\Pr(Z_t>1)}{\Pr(Z_t=1)} = \mu$ and $\frac{[\Pr(Z_t>0)]^2 - \Pr(Z_t>1)}{\Pr(Z_t=1)} = \pi$, which implies that probability based (PB) estimators of the parameters μ and π are given by

$$\widehat{\mu}_{\text{PB}} = \frac{\sum_{t=1}^{n} \mathbb{I}_{\{Z_t>1\}}}{\sum_{t=1}^{n} \mathbb{I}_{\{Z_t=1\}}} \quad \text{and} \quad \widehat{\pi}_{\text{PB}} = \frac{(\sum_{t=1}^{n} \mathbb{I}_{\{Z_t>0\}})^2 - n\, \sum_{t=1}^{n} \mathbb{I}_{\{Z_t>1\}}}{n\, \sum_{t=1}^{n} \mathbb{I}_{\{Z_t=1\}}}.$$

The parameters α and β (or ρ and κ, respectively) can be estimated through the ACF and the transition probabilities. From Eqs. (13) and (15), we have that

$$\rho(1) = \alpha + \beta = \rho \quad \text{and} \quad p := \frac{\Pr(Z_t = 0|Z_{t-1} = 1)}{\Pr(Z_t = 0|Z_{t-1} = 0)} = \frac{1-\alpha}{1+\beta}.$$

Solving these equations for α and β, the PB estimators of α and β are defined as

$$\widehat{\alpha}_{\text{PB}} = \frac{1 - \widehat{p}(1+\widehat{\rho})}{1-\widehat{p}} \quad \text{and} \quad \widehat{\beta}_{\text{PB}} = \frac{\widehat{\rho} - 1 + \widehat{p}}{1-\widehat{p}},$$

respectively, where $\widehat{\rho} := \widehat{\rho}(1) = \widehat{\gamma}(1)/\widehat{\gamma}(0)$ and $\widehat{p} = \frac{\sum_{t=2}^{n} \mathbb{I}_{\{Z_t=0, Z_{t-1}=1\}} / \sum_{t=1}^{n} \mathbb{I}_{\{Z_t=1\}}}{\sum_{t=2}^{n} \mathbb{I}_{\{Z_t=0, Z_{t-1}=0\}} / \sum_{t=1}^{n} \mathbb{I}_{\{Z_t=0\}}}$. The idea to use the conditional probabilities to estimate the parameters of an INAR process was proposed in Nastić et al. (2017).

Proposition 8 *The estimators $\widehat{\alpha}_{PB}$, $\widehat{\beta}_{PB}$, $\widehat{\pi}_{PB}$ and $\widehat{\mu}_{PB}$ are strongly consistent for estimating α, β, π and μ, respectively.*

Proof The BerG-INAR(1)BiNB process $\{Z_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process. Thus, the statistics $\sum_{t=1}^{n} I_{\{Z_t>0\}}$, $\sum_{t=1}^{n} I_{\{Z_t=1\}}$, $\sum_{t=1}^{n} I_{\{Z_t>1\}}$ and $\widehat{\rho}(1)$ are consistent estimators of $\Pr(Z_t > 0)$, $\text{P}(Z_t = 1)$, $\Pr(Z_t > 1)$ and $\rho(1)$, respectively. Finally, the consistency of the PB estimators follows from this and the properties of the convergence in probability. □

4.2 Monte Carlo simulation study

In order to compare the performances of the proposed estimators previously discussed, in this section, we perform a small simulation study for different sample sizes and for different parameter values. All simulations were carried out using the R programming language (R Core Team 2016), also see Remark 2. The data set $Z_1, \ldots, Z_n$ was always generated according to model (10). The empirical results displayed in the table and box plots, that is, the empirical biases and mean square errors (MSE), were computed over 1000 replications. The values of the MSE are given between parentheses. The sample sizes considered were $n = 200, 400$ and 800. We considered three scenarios: $\alpha = 0.3, \beta = 0.1, \pi = 0.2, \mu = 0.2$ (equidispersed, $\rho(1) = 0.4$), $\alpha = 0.4, \beta = 0.2, \pi = 0.3, \mu = 2.0$ (overdispersed with $I_Z = 2.7$, $\rho(1) = 0.6$) and $\alpha = 0.25, \beta = 0.15, \pi = 0.35, \mu = 0.30$ (underdispersed with $I_Z = 0.95$, $\rho(1) = 0.4$).

Table 1 presents the biases and MSE of the estimators of the parameters α, β, π and μ. From Table 1, it can be seen that the ML estimator causes much smaller biases (in absolute values) and MSE than the other estimators, for all scenarios. As expected, increasing the sample size reduces substantially both bias and MSE.

The previous findings are confirmed by the box plots shown in Fig. 3, which were obtained for sample size $n = 400$ for all scenarios. Again, both biases and MSE for the ML estimators are smaller than those for the other methods. Therefore, we recommend the use $\widehat{\alpha}_{\text{ML}}$, $\widehat{\beta}_{\text{ML}}$, $\widehat{\pi}_{\text{ML}}$ and $\widehat{\mu}_{\text{ML}}$ as the estimators for the parameters α, β, π and μ

Table 1 Simulated values of biases (MSEs within parentheses) of estimators of α, β, π, μ

n	Estimator of α		Estimator of β		Estimator of π			Estimator of μ		
	$\widehat{\alpha}_{ML}$	$\widehat{\alpha}_{PB}$	$\widehat{\beta}_{ML}$	$\widehat{\beta}_{PB}$	$\widehat{\pi}_{ML}$	$\widehat{\pi}_{PB}$	$\widehat{\pi}_{MM}$	$\widehat{\mu}_{ML}$	$\widehat{\mu}_{PB}$	$\widehat{\mu}_{MM}$
$\alpha = 0.3$, $\beta = 0.1$, $\pi = 0.2$ and $\mu = 0.2$										
200	−0.0134	−0.0723	0.0010	0.0481	−0.0017	0.0015	0.0047	−0.0009	−0.0026	−0.0071
	(0.0059)	(1.3568)	(0.0011)	(1.3569)	(0.0023)	(0.0043)	(0.0051)	(0.0038)	(0.0064)	(0.0063)
400	−0.0070	−0.0118	0.0057	0.0041	−0.0010	−0.0016	0.0025	0.0032	0.0049	−0.0004
	(0.0030)	(0.0367)	(0.0007)	(0.0425)	(0.0011)	(0.0022)	(0.0024)	(0.0021)	(0.0035)	(0.0034)
800	−0.0065	−0.0140	0.0039	0.0087	0.0000	0.0000	0.0006	−0.0004	−0.0002	−0.0011
	(0.0015)	(0.0185)	(0.0005)	(0.0224)	(0.0005)	(0.0010)	(0.0014)	(0.0009)	(0.0015)	(0.0017)
$\alpha = 0.4$, $\beta = 0.2$, $\pi = 0.3$ and $\mu = 2.0$										
200	−0.0276	−0.2364	0.0319	0.2202	−0.0171	−0.0083	0.0248	0.0119	0.0676	−0.0317
	(0.0072)	(3.5783)	(0.0079)	(3.5947)	(0.0111)	(0.0228)	(0.0794)	(0.1246)	(0.3080)	(0.1583)
400	−0.0119	−0.0892	0.0141	0.0795	−0.0080	0.0002	0.0231	0.0115	0.0355	−0.0202
	(0.0028)	(0.1646)	(0.0029)	(0.1685)	(0.0050)	(0.0105)	(0.0380)	(0.0568)	(0.1368)	(0.0746)
800	−0.0086	−0.0365	0.0099	0.0315	−0.0063	−0.0010	0.0061	0.0130	0.0136	−0.0022
	(0.0015)	(0.0362)	(0.0016)	(0.0395)	(0.0022)	(0.0052)	(0.0178)	(0.0291)	(0.0665)	(0.0386)
$\alpha = 0.25$, $\beta = 0.15$, $\pi = 0.35$ and $\mu = 0.30$										
200	−0.0166	−0.1588	0.0128	0.1409	0.0020	0.0028	0.0051	0.0047	0.0031	−0.0012
	(0.0046)	(0.9818)	(0.0014)	(0.9895)	(0.0029)	(0.0039)	(0.0054)	(0.0060)	(0.0083)	(0.0084)
400	−0.0129	−0.0516	0.0089	0.0403	0.0013	0.0009	0.0043	−0.0010	0.0002	−0.0050
	(0.0024)	(0.1010)	(0.0008)	(0.1063)	(0.0013)	(0.0019)	(0.0027)	(0.0028)	(0.0040)	(0.0040)
800	−0.0081	−0.0338	0.0070	0.0292	0.0003	−0.0009	0.0008	0.0016	0.0030	0.0004
	(0.0016)	(0.0363)	(0.0005)	(0.0400)	(0.0007)	(0.0010)	(0.0014)	(0.0014)	(0.0019)	(0.0019)

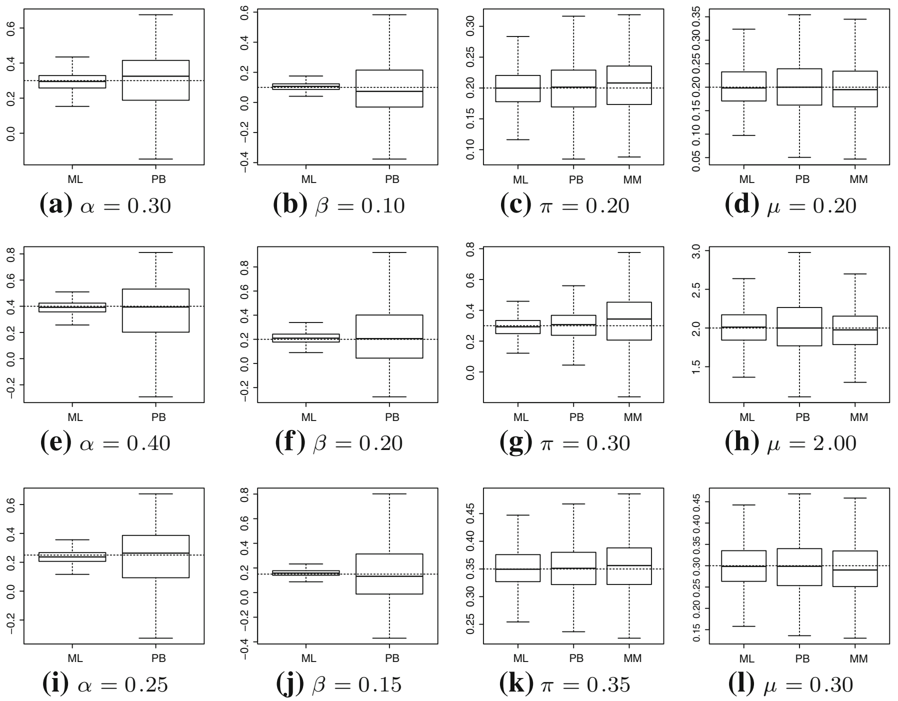

Fig. 3 *Box plots* from 1000 simulated estimates of α, β, π and μ, sample size $n = 400$. **a** $\alpha = 0.30$, **b** $\beta = 0.10$, **c** $\pi = 0.20$, **d** $\mu = 0.20$, **e** $\alpha = 0.40$, **f** $\beta = 0.20$, **g** $\pi = 0.30$, **h** $\mu = 2.00$, **i** $\alpha = 0.25$, **j** $\beta = 0.15$, **k** $\pi = 0.35$, **l** $\mu = 0.30$

of a BerG-INAR(1)BiNB process, having a good performance in terms of bias and MSE.

5 Applications to real data

Here, we conduct two applications of the BerG-INAR(1)BiNB model to real data for illustrative purposes. We estimate the unknown parameters of the fitted models by the ML method, as discussed in Sect. 4.1. We also compare the full BerG-INAR(1)BiNB process with the reduced BerG-INAR(1)$^{\text{NB}}$ process, remember the discussion in Sect. 3.3.

5.1 Iceberg order data

As a first example, let us consider a time series of counts of iceberg orders (for the ask side, per 20 min) with respect to the Deutsche Telekom shares traded in the XETRA system of Deutsche Börse. The data were collected for 32 consecutive trading days in the first quarter of 2004 (800 observations), and they were already analyzed by Jung and Tremayne (2011), Weiß (2015).

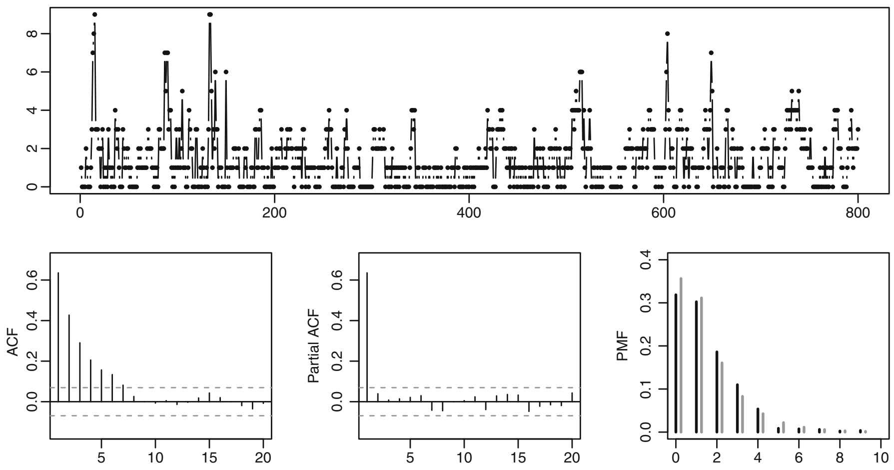

Fig. 4 Plots of time series, SACF, SPACF and PMF for iceberg counts (Sect. 5.1)

A plot of the data is shown in Fig. 4, and their sample autocorrelation and partial autocorrelation function (SACF and SPACF, respectively) indicate a first-order autoregressive autocorrelation structure with $\widehat{\rho}(1) = 0.6355$. The empirical PMF (shown in black in Fig. 4, while the gray bars refer to the fitted BerG distribution below) has mean 1.4063 and dispersion index 1.5512. The latter value is significant according to the overdispersion test by Schweer and Weiß (2014) (p value about 0). So altogether, an AR(1)-like model being able to explain overdispersion is required. Therefore, we try to fit the BerG-INAR(1)BiNB model to the data.

The ML estimates are determined as described in Sect. 4.1 by using R's `constrOptim`. The starting values are chosen based on the above moment estimates: $\widehat{\pi}_{\text{MM}} = 0.4275$, $\widehat{\mu}_{\text{MM}} = 0.9787$, and α, β are initialized by $\widehat{\rho}(1)/2$. The final ML estimates are

$$\widehat{\alpha}_{\text{ML}} = 0.4251, \quad \widehat{\beta}_{\text{ML}} = 0.1842, \quad \widehat{\pi}_{\text{ML}} = 0.2633, \quad \widehat{\mu}_{\text{ML}} = 1.0668.$$

Approximate standard errors are obtained from the inverse Hessian as 0.0453, 0.0507, 0.0411 and 0.1099, respectively. Let us analyze if this model is really adequate for the given data. First, we compare properties of the fitted model with the empirical ones as computed from the data. The fitted model's mean, $\widehat{\pi}_{\text{ML}} + \widehat{\mu}_{\text{ML}} = 1.3301$ (approximate standard error 0.1073), and dispersion index, 1.8036, are reasonably close to the empirical values 1.4063 and 1.5512, respectively. The whole BerG-PMF agrees quite well with the empirical one, see the gray resp. black bars in Fig. 4. The proportion of zeros within the fitted model equals 0.3565, the empirical one 0.3188. Finally, also the ACFs are in good agreement, e. g., with the lag-1 values being 0.6093 (model) and 0.6355 (empirical), respectively.

As a second tool for checking the model adequacy, we computed the standardized Pearson residuals using (12). These residuals do not show any significant ACF value, confirming the adequacy of the fitted model's autocorrelation structure. The mean of the residuals, 0.0261, is close to 0, and their variance, 0.8742, is close to 1. In fact, the variance is slightly smaller than 1, so the data show slightly less dispersion than being considered by the model (Harvey and Fernandes 1989), which goes along with the slight discrepancy between the dispersion index values above. But altogether, the model constitutes an adequate fit to the data.

Finally, let us discuss the question if the reduced BerG-INAR(1)$^{\mathrm{NB}}$ model from Sect. 3.3 might serve as an alternative for the data. Already the moment estimates $\widehat{\pi}_{\mathrm{MM}} = 0.4275$, $\widehat{\mu}_{\mathrm{MM}} = 0.9787$ and $\widehat{\beta}_{\mathrm{MM}} = \widehat{\rho}(1) = 0.6355$ indicate the weak point of this model: The data show too much autocorrelation with respect to the model, the upper bound for β is $\widehat{\mu}_{\mathrm{MM}}/(\widehat{\mu}_{\mathrm{MM}} + 1) = 0.4946$ and thus violated. Doing an ML estimation anyway, one ends up with $\widehat{\beta}_{\mathrm{ML}} = 0.6503$ (close to the actual autocorrelation level) and $\widehat{\pi}_{\mathrm{ML}} = 0.2957$, $\widehat{\mu}_{\mathrm{ML}} = 1.8599$. Compared to the estimates of the full model, we realize a much larger estimate for the BerG's parameter μ. As a result, the fitted model is not adequate for the data, e. g., its mean 2.1556 is much larger than the empirical mean 1.4063, and also its dispersion index (2.5642 vs. 1.5512). So the full model is clearly preferable with respect to the iceberg counts data. Analogous arguments apply to the NGINAR(1) model (Ristić et al. 2009), which constitutes a special instance of BerG-INAR(1)$^{\mathrm{NB}}$ model, see the discussion after Definition 3.

5.2 Family violence data

In the second application, we consider the series of monthly counts of family violences in the 11th police car beat in Pittsburgh, which have already been analyzed by Bakouch and Ristić (2010). The data set is obtained from the crime data section of the forecasting principles site (file `PghCarBeat.csv` at http://www.forecastingprinciples.com/index.php/crimedata), and it is also available from the authors upon request. It consists of 144 observations, starting in January 1990 and ending in December 2001.

The time series data, their SACF and SPACF as well as their PMF (again with the PMF of the fitted model in gray, see below) are displayed in Fig. 5. Analyzing Fig. 5, we conclude that a first-order autoregressive model may be appropriate for the given data series, because of the clear cutoff after lag 1 in the SPACF. Furthermore, the behavior of the series indicates that it may be stationary. The sample mean is 0.4028, the sample variance is 0.3821, and the first-order autocorrelation is 0.1770. The ratio between the sample variance and the sample mean (empirical dispersion index) is 0.9486. The application of the overdispersion test by Schweer and Weiß (2014) did not reject the null hypothesis of equidispersion (Poisson INAR(1) model), with the p value for the test being 0.6831.

The ML estimates for the fitted model are $(\widehat{\alpha}_{\mathrm{ML}}, \widehat{\beta}_{\mathrm{ML}}, \widehat{\pi}_{\mathrm{ML}}, \widehat{\mu}_{\mathrm{ML}}) = (0.0685, 0.1332, 0.2408, 0.1669)$. Since these estimates ($\widehat{\alpha}_{\mathrm{ML}}$ and $\widehat{\beta}_{\mathrm{ML}}$) are close to the boundary of parameter space as defined by (11), the approximate standard errors become rather large, given by $\approx (0.23, 0.28, 0.06, 0.07)$. The mean, variance and ACF within the estimated model are given by, respectively, $\widehat{\pi}_{\mathrm{ML}} + \widehat{\mu}_{\mathrm{ML}} = 0.4077$ (approximate standard error ≈ 0.065), $\widehat{\pi}_{\mathrm{ML}}(1 - \widehat{\pi}_{\mathrm{ML}}) + \widehat{\mu}_{\mathrm{ML}}(1 + \widehat{\mu}_{\mathrm{ML}}) = 0.3776$ and $\widehat{\alpha}_{\mathrm{ML}} + \widehat{\beta}_{\mathrm{ML}} = 0.2017$. Note that these values for the fitted model are close to the corresponding empirical values. An analogous statement holds with respect to

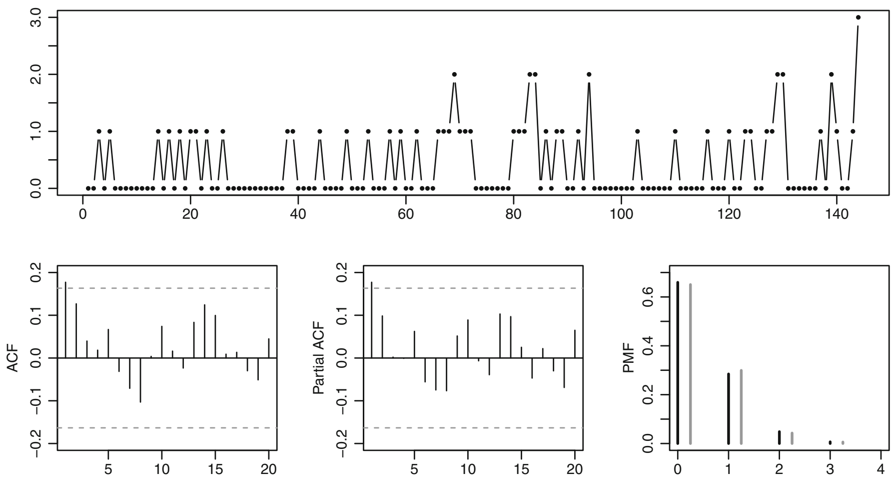

Fig. 5 Plots of time series, SACF, SPACF and PMF for family violence counts (Sect. 5.2)

the dispersion index (0.9261 within the fitted BerG-INAR(1)BiNB model, 0.9486 empirically), although we did not observe a significant violation of the equidispersion property. Our model not only captures well the dispersion index, it does also better than a Poisson model if considering the proportion of zeros. In the data set, this is 0.6597, while the proportion of zeros based on the estimated BerG-INAR(1)BiNB model is 0.6506 (a Poisson INAR(1) model would have 0.7121). Thus, the proposed model works well for capturing the proportion of zeros in this application, in contrast to the Poisson INAR(1) model. In fact, the whole PMF of the fitted model (gray) is very close to the empirical one (black), see Fig. 5. Also an analysis of the Pearson residuals did not lead to any contradictions (e. g., mean 0.0056 and variance 1.0484).

For the reduced model with $\alpha = 0$ (i. e., BerG-INAR(1)$^{\text{NB}}$ model), in contrast, we observe again that the parameter estimates are out of the parameter space (see Proposition 5): the moment estimate $\widehat{\beta}_{\text{MM}} = 0.1773$ violates the bound $\widehat{\mu}_{\text{MM}}/(1+\widehat{\mu}_{\text{MM}}) = 0.1494$, and ML estimation essentially ends up in a further model reduction, $(\widehat{\beta}_{\text{ML}}, \widehat{\pi}_{\text{ML}}, \widehat{\mu}_{\text{ML}}) = (0.6475, 0.0000, 1.8369)$. In view of the equidispersion, however, a geometric marginal distribution is not appropriate. Thus, also for this data set, the reduced model cannot be used. And also the special instance of the NGINAR(1) model (Ristić et al. 2009), see the discussion after Definition 3, cannot be used here (note again that a geometric marginal distribution is necessarily overdispersed). Finally, Bakouch and Ristić (2010) applied the (three-parameter) zero-truncated Poisson INAR(1) model to the data. They obtained a slightly lower value for the information criterion AIC (234.2462 vs. 236.5801), but the model appears a bit artificial since it can only be applied after shifting the data. We conclude by pointing out that the family violence counts might be affected by underreporting, because such type of offense might not always be presented to the authorities; a method for dealing with underreported counts is proposed by Fernández-Fontelo et al. (2016).

6 Conclusions

We discussed the distribution of the convolution of a Bernoulli and a geometric random variable and summarized some of its properties. Afterward, we introduced a stationary first-order nonnegative integer-valued autoregressive model for count data process with Bernoulli-geometric marginals based on a new thinning operator. The new thinning operator can be interpreted as the sum of two know thinning operators. The new model has several advantages: It can be used for modeling time series of counts with equidispersion, underdispersion or overdispersion, and it has simple innovation structure. The main properties of the new process are derived. Three methods for estimating the model parameters are considered. The simulation results show that the ML estimator presents much smaller biases and MSE than the other estimators. Thus, we recommend the use of the ML method to estimate the process parameters of an BerG-INAR$(1)^{\mathrm{BiNB}}$ process. Finally, we fitted the BerG-INAR(1)BiNB model to two real data sets. As part of future research, it would be of interest to extend the proposed model to autoregressive order $p > 1$ (Du and Li 1991; Latour 1998).

Acknowledgements The authors thank the two referees for very useful comments on an earlier draft of this article. The iceberg order data of Sect. 5.1 were kindly made available to the second author by the Deutsche Börse. Prof. Dr. Joachim Grammig, University of Tübingen, is to be thanked for processing of it to make it amenable to data analysis. We are also very grateful to Prof. Dr. Robert Jung, University of Hohenheim, for his kind support to get access to the data.

References

Al-Osh MA, Alzaid AA (1987) First-order integer valued autoregressive (INAR(1)) process. J Time Ser Anal 8:261–275

Aoyama K, Shimizu K, Ong SH (2008) A first-passage time random walk distribution with five transition probabilities: a generalization of the shifted inverse trinomial. Ann Inst Stat Math 60:1–20

Bakouch HS, Ristić MM (2010) A mixed thinning based geometric INAR(1) model. Metrika 72:265–280

Bourguignon M, Vasconcellos KL (2015) First order non-negative integer valued autoregressive processes with power series innovations. Braz J Probab Stat 29:71–93

Bourguignon M, Vasconcellos KL, Reisen VA, Ispány M (2016) A Poisson INAR(1) process with a seasonal structure. J Stat Comput Simul 86:373–387

Du J, Li Y (1991) The integer valued autoregressive (INAR(p)) model. J Time Ser Anal 12:129–142

Feller W (1968) An introduction to probability theory and its applications, vol I, 3rd edn. Wiley, Hoboken

Fernández-Fontelo A, Cabaña A, Puig P, Moriña D (2016) Under-reported data analysis with INAR-hidden Markov chains. Stat Med 35:4875–4890

Grunwald G, Hyndman RJ, Tedesco L, Tweedie RL (2000) Non-Gaussian conditional linear AR(1) models. Aust N Z J Stat 42:479–495

Harvey AC, Fernandes C (1989) Time series models for count or qualitative observations. J Bus Econ Stat 7:407–417

Jazi MA, Jones G, Lai C-D (2012) First-order integer valued AR processes with zero inflated poisson innovations. J Time Ser Anal 33:954–963

Johnson NL, Kemp AW, Kotz S (2005) Univariate discrete distributions, 3rd edn. Wiley, Hoboken

Jung RC, Tremayne AR (2011) Useful models for time series of counts or simply wrong ones? AStA Adv Stat Anal 95:59–91

Kemp AW (1979) Convolutions involving binomial pseudo-variables. Sankhyā Ser A 41:232–243

Kim H, Lee S (2017) On first-order integer-valued autoregressive process with Katz family innovations. J Stat Comput Simul 87:546–562

Latour A (1998) Existence and stochastic structure of a non-negative integer-valued autoregressive process. J Time Ser Anal 19:439–455

McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21:645–650
Moriña D, Puig P, Ríos J, Vilella A, Trilla A (2011) A statistical model for hospital admissions caused by seasonal diseases. Stat Med 30:3125–3136
Nastić AS, Ristić MM, Janjić AD (2016) A mixed thinning based geometric INAR(1) model. Filomat (to appear)
Nastić AS, Ristić MM, Miletić Ilić AV (2017) A geometric time-series model with an alternative dependent Bernoulli counting series. Commun Stat Theory Methods 46:770–785
R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.r-project.org
Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. J Stat Plan Inference 139:2218–2226
Schweer S, Weiß CH (2014) Compound poisson INAR(1) processes: stochastic properties and testing for overdispersion. Comput Stat Data Anal 77:267–284
Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. Ann Probab 7:893–899
Weiß CH (2008) Thinning operations for modelling time series of counts—a survey. AStA Adv Stat Anal 92:319–341
Weiß CH (2013) Integer-valued autoregressive models for counts showing underdispersion. J Appl Stat 40:1931–1948
Weiß CH (2015) A Poisson INAR(1) model with serially dependent innovations. Metrika 78:829–851
Zhu R, Joe H (2003) A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. Stoch Mod 19:235–254