Geometric Autoregressive Models Revisited

Offer Kella[*][†] Andreas Löpker[‡] Christian H. Weiß[§]

January 15, 2026

Revision of AP-2024-19509

Abstract

This paper considers a flexible extension of the binomial and negative-binomial thinning operator, referred to as a modulating operator, that is linked to a class of certain Möbius matrices. Different count processes using this operator are considered, like INAR(1)- and INMA(1)-type processes as well as minification processes. A novel connection to linear birth-and-death processes is also provided.

Keywords: count time series, minification processes, integer-valued autoregressive models, thinning operators.

1 Introduction

The traditional first-order autoregressive (AR(1)) model with model recursion $Y_n = \alpha \cdot Y_{n-1} + \epsilon_n$ cannot be applied to a count process (X_n), i.e., a stochastic process consisting of non-negative integer-valued random variables (r.v.), because the involved multiplication would not preserve the integer nature of the count r.v. For this reason, various integer substitutes of the multiplication have been proposed in the literature. In particular, so-called thinning operators are widely used for defining integer-valued autoregressive (INAR) models for count processes, see Weiß (2008), Scotto et al. (2015). The probably most well-known thinning operator is *binomial thinning* of Steutel & van Harn (1979), which is defined through a conditional binomial distribution: Given a non-negative integer valued random variable X, one defines $p \circ X = \sum_{i=1}^{X} B_i$ with independent and identically distributed (i.i.d.) Bernoulli r.v. $B_i \sim \mathrm{Ber}(p)$, being also independent of X, such that $p \circ X | X \sim \mathrm{Bin}(X; p)$. Here, $p \in [0, 1]$ denotes the 'success probability'. Binomial thinning entails a 'shrinkage' of X in the sense that $p \circ X \leq X$ almost surely (a.s.), so also $\mathbb{E}\,(p \circ X) \leq \mathbb{E}\,(X)$.

In this paper, we study an operator '$\odot$' on non-negative integer-valued random variables that generalizes binomial thinning and has several desirable properties. For a r.v. X with values in $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, we let

$$\theta \odot X = \sum_{i=1}^{X} Z_i, \tag{1}$$

[*]Department of Statistics and Data Science, the Hebrew University of Jerusalem, Jerusalem, Israel
[†]Supported in part by ISF grant 3336/24 and by the Vigevani Chair in Statistics.
[‡]Department of Computer Science and Mathematics, University of Applied Sciences, Dresden, Germany
[§]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany.
E-Mail: weissc@hsu-hh.de. ORCID: 0000-0001-8739-6631.

where the r.v. $(Z_i)_{i=1,2,\dots}$ are i.i.d. and independent of X, and where their distribution is a mixture of an atom at zero and a geometric distribution on $\mathbb{N} = \{1, 2, \dots\}$. This distribution can be equivalently expressed as that of $B \cdot V$, where V, B are independent, V is geometrically distributed on $\mathbb{N}$, and B is Bernoulli distributed. The shape of this distribution will be determined by a two dimensional parameter $\theta = (\alpha, \beta)$ with $\beta \geq 0$ and $0 \leq \alpha \leq \beta + 1$ (to be explained later in Section 2). The operator $\odot$ and the class of modified geometric distributions was also studied in Bourguignon & Weiß (2017) and coincides with the definition in Aly & Bouzar (2019), with different notations (their (m, r) is our (α, β)); we adopt their name *T-geometric distribution* for it, where the 'T' refers to the truncation at zero. Our operator $\odot$ is also, again with a different parametrization, the operator suggested in Andrews & Balakrishna (2023). This operator shares some of the convenient properties of *binomial thinning* '$\circ$' (see Steutel & van Harn (1979)), but there are also substantial differences. In particular, while a thinning operator necessarily leads to a reduction of the mean (recall the above discussion of binomial thinning), the operator defined by (1) also allows for increases in the mean of the involved random variable. For this reason, we refer to '$\odot$' according to (1) more generally as a *modulating operator* (modulation). What makes T-geometric distributions convenient to work with is the fact that their probability generating function is of a linear fractional type. Expressed in terms of the *alternate probability generating function* (apgf) $\mathscr{G}_Z(s) = \mathbb{E}((1-s)^Z)$, we have

$$1 - \mathscr{G}_Z(s) = \frac{\alpha s}{1 + \beta s}.$$

In particular, $1 - \mathscr{G}_Z(s)$ is a *Möbius transformation* and we will show how a matrix representation of such functions can be exploited to investigate properties of stochastic processes involving T-geometric distributions. While this connection with Möbius tranformations seems novel, the pleasant property that the class of linear fractional functions are closed under functional iteration has already been observed before. In particular, in the study of branching processes, the linear fractional case has been emphasized before as a tractable case, e.g., in Athreya & Ney (2012), Section I.4. (see also Kevei (2011) for an example in the INAR context).

There is a vast literature regarding generalized thinning operators related to our definition. A survey of thinning operators, including generalized versions, is given by Weiß (2008), Scotto et al. (2015). Al-Osh & Aly (1992) describe an operator $\omega * X = \sum_{i=1}^{\lambda \circ X} Z_i$, where the *counting variables* (Z_i) have a geometric distribution on $\mathbb{N}_0$ with parameter $\omega/(1+\omega)$, say $\mathrm{Geo}_{\mathbb{N}_0}(\frac{\omega}{1+\omega})$. Here, '$\circ$' is the binomial thinning operator, and $\lambda, \lambda/\omega \in [0, 1]$ has to hold. Operator '$*$' has a structure equivalent to (1) if we write $\sum_{i=1}^{\lambda \circ X} Z_i = \sum_{i=1}^{X} B_i \cdot Z_i$ with Bernoulli r.v. B_i, showing that '$*$' is covered by our modulating operator '$\odot$'. The operator we discuss is also a special case of the 'compounding multiplicative operator' in Weiß & Zhu (2024). We also note that Ristić et al. (2009) define a thinning operator '$*$' that relates to our case, but with counting variables that have a pure geometric distribution (without zero inflation). See also Latour (1998), McKenzie (1986), Zhu & Joe (2010) and Nastić & Ristić (2012) for related models and Weiß (2008), Scotto et al. (2015) for an overview over the literature regarding thinning operators.

Using the operator $\odot$, one can define the autoregressive INAR-type process

$$X_{n+1} = \theta \odot X_n + \varepsilon_n \tag{2}$$

as studied in Aly & Bouzar (2019). INAR(1)-type processes with different kinds of thinning operators have been proposed by Al-Osh & Aly (1992), Andrews & Balakrishna (2023), Latour (1998), Aly & Bouzar (2005), and many further authors. A zero-inflated geometric setting is also discussed in Barreto-Souza (2015), Bakouch et al. (2018), and Bourguignon & Weiß (2017).

In this paper, we additionally investigate a *minification process* defined by

$$X_{n+1} = (\theta \odot X_n) \wedge U_n,$$

where the addition in the INAR-type setting is replaced by a minimum. Again, the (U_n) are i.i.d. and independent of everything else. Minification processes were first studied in the continuous setting with a simple multiplication instead of '$\odot$', see Tavares (1980), Lewis & McKenzie (1991). A discrete variant was introduced by Kalamkar (1995). There, a process $X_n = K \cdot (X_{n-1} \wedge \xi_n)$ is defined, where $K \geq 1$ is some fixed integer and (ξ_n) are i.i.d. The main focus of that paper is the class of possible stationary distributions. Littlejohn (1992) defines an operator '$\rho\backslash$' such that $\rho\backslash X = X + \sum_{i=1}^{X+1} G_i$ and G_i are i.i.d. geometric random variables (on $\mathbb{N}_0$). The minification process is then $Y_{n+1} = (\rho\backslash Y_n) \wedge \eta_{n+1}$ with i.i.d. (η_n). A quite similar process is studied in Aleksić & Ristić (2021), Qian & Zhu (2022), and Stojanović et al. (2024). Here, a modified negative-binomial thinning operator $\diamond$ is used, namely $\alpha \diamond X = \sum_{i=1}^{X+1} G_i$ (note the additional '+1' in the upper summation limit), where $(G_n) \in \mathrm{Geo}_{\mathbb{N}_0}(1/(1+\alpha))$ are i.i.d. with a geometric distribution and $\mathbb{P}(G_i = 0) = \frac{1}{1+\alpha}$. In terms of our more general operator '$\odot$', this is the special case where $\alpha \diamond X = G_0 + (\alpha, \alpha) \odot X$. Barreto-Souza et al. (2023) study a process $X_n = Z_n \wedge X_{n-1} + \varepsilon_n$ with independent geometrically distributed Z_n with support on $\mathbb{N}_0$. There is also a short discussion regarding a generalization to zero-inflated Z.

The remainder of this article is organized as follows. After having introduced our modulating operator '$\odot$' in Section 2, our first main contribution is presented in Section 3. There, we show how our modulating operator can be related to the multiplication of certain Möbius matrices, and how matrix algebra can be utilized to simplify calculations. With Section 4, we then turn to the modeling of count processes. First, we consider a generalized version of the INAR(1) model according to (2), where limiting cases and forecast distributions are investigated. Afterwards in Section 5, we discuss the aforementioned minification process, which has interesting stochastic properties. In Section 6, we show seemingly new connections of the aforementioned modulation-based processes to linear birth-and-death processes. Section 7 then briefly discusses the moving-average (MA) counterparts ('INMA') to the INAR-type processes of Sections 4 and 5. Finally, we conclude in Section 8, where we also outline possible directions for future research.

2 The modulating operator

In what follows, we let

$$\mathscr{G}_X(s) = \mathbb{E}\left((1-s)^X\right), \qquad s \in (0,2),$$

denote the *alternate probability generating function* (apgf). We mainly use this variant of the probability generating function instead of the usual $\mathcal{P}_X(s) = \mathbb{E}\left(s^X\right)$ as it turns out that this definition is well suited to our needs, see e.g. the central formula (7) below. Note that $\mathscr{G}_X(s) = \mathbb{P}(X \leq \Gamma_s)$, where Γ_s has a geometric distribution on $\mathbb{N}_0$ with $\mathbb{P}(\Gamma_s = k) = s(1-s)^k$ for $k \in \mathbb{N}_0$. We have $\mathscr{G}_X(0) = 1$, $\mathscr{G}_X(1) = \mathbb{P}(X = 0)$, and $\mathscr{G}_X(s)$ is completely monotone on $(0,1)$, in particular non-increasing. The mean and variance of X can be deduced from $\mathscr{G}_X'(0) = -\mathbb{E}(X)$ and $\mathscr{G}_X''(0) = \mathbb{E}(X(X-1))$. For a sum of independent r.v. X and Y, we have the usual product rule $\mathscr{G}_{X+Y}(s) = \mathscr{G}_X(s)\mathscr{G}_Y(s)$. We refer the reader to Fristedt & Gray (1997) for more properties of probability generating functions. Since $\mathscr{G}_X(s) = \mathcal{P}_X(1-s)$, properties of either transform immediately translate to properties of the other.

The classes of distributions to be used in the sequel are summarized in Table 1. We write $\wedge$

Table 1: Notations and properties for some discrete distributions, where $p \in [0,1]$, $\lambda > 0$.

Name	Abbreviation	Support	$\mathbb{P}(X = k)$	$\mathcal{G}_X(s)$
Geometric	$\text{Geo}_{\mathbb{N}}(p)$	$\mathbb{N}$	$p(1-p)^{k-1}$	$\frac{p(1-s)}{p+(1-p)s}$
	$\text{Geo}_{\mathbb{N}_0}(p)$	$\mathbb{N}_0$	$p(1-p)^k$	$\frac{p}{p+(1-p)s}$
	$\text{BerG}(\alpha,\beta)$	$\mathbb{N}_0$	see eqn. (6)	$1 - \frac{\alpha s}{1+\beta s}$
Bernoulli	$\text{Ber}(p)$	$\{0,1\}$	p, for $k = 1$	$1 - ps$
Binomial	$\text{Bin}(n;p)$	$\{0,1,\dots,n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$(1-ps)^n$
Poisson	$\text{Pois}(\lambda)$	$\mathbb{N}_0$	$e^{-\lambda}\frac{\lambda^k}{k!}$	$e^{-\lambda s}$
Geometric Poisson	$\text{GeoP}(\lambda,p)$	$\mathbb{N}_0$	see Johnson et al. (2005)	$\exp\left(-\frac{\lambda\alpha s}{1+\beta s}\right)$

and $\vee$ for the minimum and maximum operation, respectively, and x^+ for $0 \vee x$. Empty sums are defined to be zero, empty products are one, and $0^0 = 1$ as well.

In Section 2.1, we discuss the T-geometric distribution, which serves as the basis for defining our modulating operator in Section 2.2. After a brief summary of some basic properties, our first main contribution is presented afterwards in Section 3, where we associate the modulating operator with the Möbius transform and a corresponding matrix representation.

2.1 T-geometric random variables

Let the r.v. $X \in \text{Geo}_{\mathbb{N}}(p)$ be geometrically distributed with support $\mathbb{N}$ and $\mathbb{P}(X = n) = (1-p)^{n-1}p$ for $n \geq 1$, recall Table 1. Then, its apgf is given by

$$\mathcal{G}_X(s) = 1 - \frac{s}{p + (1-p)s}. \tag{3}$$

If $Z \sim BX$ is a $(1-a,a)$-mixture of a geometric distribution with an independent Bernoulli r.v. $B \sim \text{Ber}(a)$, then clearly

$$\mathcal{G}_Z(s) = 1 - a + a\mathcal{G}_X(s) = 1 - \frac{as}{p + (1-p)s} = 1 - \frac{\frac{a}{p}s}{1 + \frac{1-p}{p}s}. \tag{4}$$

Noting that $\mathbb{P}(Z = 0) = \mathbb{P}(B = 0) = 1 - a$, it also follows that a non-negative integer-valued r.v. Z has such a distribution if and only if (iff) $\mathbb{P}(Z = n | Z > 0) = (1-p)^{n-1}p$ for $n \geq 1$ and $\mathbb{P}(Z > 0) = a$. This distribution was called *T-geometric distribution* in Aly & Bouzar (2019).

In the special case $a = 1 - p$, the distribution defined by (4) reduces to the ordinary geometric distribution on $\mathbb{N}_0$, $\text{Geo}_{\mathbb{N}_0}(p)$. Hence, $a > 1-p$ ($a < 1-p$) corresponds to inflating (deflating) the zero probability compared to $\text{Geo}_{\mathbb{N}_0}(p)$. In what follows, we use the reparametrization $\alpha = a/p$ and $\beta = (1-p)/p$, or equivalently, $p = (1+\beta)^{-1}$ and $a = \alpha(1+\beta)^{-1}$, for $\beta \geq 0$ and $0 \leq \alpha \leq 1+\beta$. Following Bourguignon & Weiß (2017) and Aly & Bouzar (2019), we write $Z \sim \text{BerG}(\theta)$ with parameter $\theta = (\alpha,\beta)$ if Z has this T-geometric distribution, denoting the parameter space by

$$\Theta = \{(\alpha,\beta) | \beta \in [0,\infty),\ \alpha \in [0, 1+\beta]\}. \tag{5}$$

With this choice of parameters, we then have (cf. (2.3) in Aly & Bouzar (2019))

$$\mathbb{P}(Z=k)=\begin{cases}1-\dfrac{\alpha}{1+\beta} & ;k=0\\ \dfrac{\alpha}{1+\beta}\dfrac{1}{1+\beta}\left(1-\dfrac{1}{1+\beta}\right)^{k-1} & ;k=1,2,3,\ldots\end{cases}, \tag{6}$$

and the apgf of Z becomes particularly nice, namely

$$\mathcal{G}_Z(s)=1-\frac{\alpha s}{1+\beta s},\qquad s\in[0,2). \tag{7}$$

It is straightforward and is stated in Aly & Bouzar (2019) that $\mathbb{E}(Z)=-\mathcal{G}_Z'(0)=\alpha$ and

$$\operatorname{Var}(Z)=\alpha(1+2\beta)-\alpha^2. \tag{8}$$

Since $\beta\geq(\alpha-1)^+$, we have $\operatorname{Var}(Z)\geq(\alpha^2-\alpha)\vee(\alpha-\alpha^2)=\alpha|1-\alpha|$. It is easy to check that equality holds iff either $a=0$ or $a=1$ or $p=1$.

We have seen that given $Z\geq 1$, Z has a conditional geometric distribution with mean $1+\beta$, and that Z has the distribution of the product $B\cdot X$, where B has a Bernoulli distribution with $\mathbb{P}(B=1)=1-\mathbb{P}(B=0)=\alpha/(1+\beta)$, X is geometric on $\mathbb{N}$ with $\mathbb{P}(X=1)=1/(1+\beta)$, and both are independent.

Note that for $\alpha=\beta$, the T-geometric distribution $\operatorname{BerG}(\alpha,\beta)$ reduces to the ordinary geometric distribution $\operatorname{Geo}_{\mathbb{N}_0}(1/(1+\beta))$ on $\mathbb{N}_0$. The case of zero *inflation* corresponds to $\alpha\leq\beta$, where we can rewrite (6) as

$$\mathbb{P}(Z=k)=\begin{cases}1-\dfrac{\alpha}{\beta}+\dfrac{\alpha}{\beta}\dfrac{1}{1+\beta} & ;k=0\\ \dfrac{\alpha}{\beta}\dfrac{1}{1+\beta}\left(1-\dfrac{1}{1+\beta}\right)^{k} & ;k=1,2,3,\ldots\end{cases}. \tag{9}$$

So $Z\sim B'\cdot X_0$, where now B' has a Bernoulli distribution with $\mathbb{P}(B'=1)=1-\mathbb{P}(B'=0)=\alpha/\beta$, X_0 is geometric on $\mathbb{N}_0$ with $\mathbb{P}(X=0)=1/(1+\beta)$, and both are again independent.

By contrast, zero *deflation* is attained for $\alpha\geq\beta$, where we have

$$\mathcal{G}_Z(s)=1-\frac{\alpha s}{1+\beta s}=\frac{1-(\alpha-\beta)s}{1+\beta s}=\big(1-(\alpha-\beta)s\big)\cdot\left(1-\frac{\beta s}{1+\beta s}\right), \tag{10}$$

such that $Z\sim B''+X$ with $\mathbb{P}(B''=1)=1-\mathbb{P}(B''=0)=\alpha-\beta$ and X as before. The maximal extent of zero deflation is attained for $\alpha=\beta+1$, where Z's distribution becomes zero-*truncated*, namely $Z\sim\operatorname{Geo}_{\mathbb{N}}(p)$.

2.2 Definition and properties of the modulating operator

For a non-negative integer-valued r.v. X and $\theta=(\alpha,\beta)\in\Theta$, Aly & Bouzar (2019) define an operator

$$\theta\odot X=\sum_{i=1}^{X}Z_i, \tag{11}$$

where the counting variables $(Z_i)_{i=1,2,\ldots}\in\operatorname{BerG}(\theta)$ are i.i.d. and independent of X and which we will call *modulating operator*. We agree that in an expression, where several modulations are performed, we can use one θ-symbol but still assume that all modulations are independent of each other. As an example, if X and Y are independent, then

$$\theta \odot (X+Y) \overset{d}{\sim} \theta \odot X + \theta \odot Y. \tag{12}$$

The following results for the mean, variance, covariance, and probability mass function can also be found in Andrews & Balakrishna (2023), see also Aly & Bouzar (2019).

Lemma 1. *Suppose that X and Y are non-negative integer-valued but not necessarily independent and that $\theta \odot X = \sum_{i=1}^{X} Z_i$ and $\xi \odot Y = \sum_{i=1}^{Y} Z'_i$, where $Z_i \sim \mathrm{BerG}(\theta)$, $\theta = (\alpha, \beta)$ and $Z'_i \sim \mathrm{BerG}(\xi)$, $\xi = (\gamma, \delta)$.*

(*a*) $\mathbb{E}(\theta \odot X) = \alpha \cdot \mathbb{E}(X)$

(*b*) $\mathrm{Var}(\theta \odot X) = \mathbb{E}(X)\left(\alpha(1+2\beta) - \alpha^2\right) + \alpha^2 \mathrm{Var}(X)$.

(*c*) $\mathrm{Cov}(\theta \odot X, \xi \odot Y) = \mathbb{E}\left(\sum_{i=1}^{X}\sum_{j=1}^{Y} \mathrm{Cov}(Z_i, Z'_j)\right) + \alpha \cdot \gamma \cdot \mathrm{Cov}(X, Y)$.

(*d*) $\mathrm{Cov}(\theta \odot X, X) = \alpha \mathrm{Var}(X)$.

(*e*) *The correlation coefficient is given by*

$$\varrho(\theta \odot X, X) = \sqrt{\frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \mathbb{E}(X)\left(\frac{1+2\beta}{\alpha} - 1\right)}}.$$

(*f*) *If we define $\binom{n}{j} = 0$ for $j > n$ then*

$$\mathbb{P}(\theta \odot X = k) = \left(\frac{\beta}{1+\beta}\right)^k \sum_{n=0}^{\infty} \mathbb{P}(X=n) \sum_{j=0}^{k} \binom{n}{j}\binom{n+k-j-1}{n-1}\left(\frac{\alpha-\beta}{\beta}\right)^j \left(\frac{1+\beta-\alpha}{1+\beta}\right)^{n-j}.$$

The formula in (f) can be found in Aly & Bouzar (2019) and Andrews & Balakrishna (2023), but it is also a consequence of the connection of the modulating operator with linear birth-and-death processes (Section 6) and well-known formulas for the latter (see e.g. Bailey (1964), p. 112).

3 Möbius transform and modulating operator

3.1 Matrix representations of Möbius transforms

Several results in this paper are based on compositions of rational functions $f : \mathbb{R} \setminus \{-d/c\} \to \mathbb{R}$ of the type

$$f(s) = \frac{as+b}{cs+d}, \qquad a, b, c, d \geq 0, \quad ad - bc \neq 0, \tag{13}$$

where $ad = bc$ is excluded as $f(s)$ is a constant function in that case. Note that under this exclusion neither (a, b) nor (c, d) is $(0, 0)$, so that both numerator and denominator are not identically zero.

Such functions are called *Möbius transformations* Behrends (2022), and a function of the type (13) is represented by the 2×2 matrix

$$M_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \det M_f \neq 0. \tag{14}$$

We observe that with the permutation matrix $P = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, PM_f is the matrix representation of

$1/f(s)$ while $M_f P$ is that of $f(1/s)$. We note that if $c \neq 0$, then

$$f(s) = \frac{1}{c}\left(a - \frac{\det M_f}{cs+d}\right). \tag{15}$$

Furthermore, for every $m \geq 1$

$$\frac{\mathrm{d}^m}{\mathrm{d}s^m} f(s) = m!\left(-\frac{c}{cs+d}\right)^{m-1} \frac{\det M_f}{(cs+d)^2}. \tag{16}$$

If $c = 0$, in turn, then the condition $\det M_f \neq 0$ implies that $c = 0 \neq d$. In this case, (16) also trivially turns out to hold for every $m \geq 1$ and s (unrestricted).

The above matrix representation is not unique, because two matrices that differ only by a scalar factor represent the same Möbius transformation. The usefulness of the representation stems from the fact that the functional composition $f \circ g$ of two Möbius transforms is represented by a matrix multiplication (see e.g. Schiff (2022), Theorem 1.38),

$$M_{f\circ g} = M_f \cdot M_g.$$

In order to distinguish important cases, we define the *discriminant* of M_f by

$$\Delta = (a-d)^2 + 4bc,$$

which is obviously non-negative in our case. In the theory of Möbius transforms f (or M_f for that matter) is called

- *hyperbolic* if $\Delta > 0$ and
- *parabolic* if $\Delta = 0$,

see e.g. Behrends (2022), p.161. Regarding the eigenvalues of M_f, there is the following result.

Lemma 2. *The matrix M_f has the two real eigenvalues $\lambda = \frac{1}{2}(a+d+\sqrt{\Delta})$ and $\mu = \frac{1}{2}(a+d-\sqrt{\Delta})$. In particular, $\mu \leq \lambda$. The eigenvalues fulfill the inequalities $\mu \leq a \wedge d$ and $\lambda \geq a \vee d$, with equalities iff $bc = 0$. Here, $\mu = d$ (equiv., $\lambda = d$) iff $a \geq d$ (equiv., $a \leq d$) and $bc = 0$.*

Proof. The characteristic polynomial of M_f is

$$(a-x)(d-x) - bc = x^2 - (a+d)x + ad - bc. \tag{17}$$

The roots are given by $\lambda = \frac{1}{2}(a+d+\sqrt{\Delta})$ and $\mu = \frac{1}{2}(a+d-\sqrt{\Delta})$. Moreover, since $\sqrt{\Delta} \geq |a-d|$,

$$\lambda = \frac{a+d+\sqrt{\Delta}}{2} \geq \frac{a+d+|a-d|}{2} = a \vee d, \quad \mu = \frac{a+d-\sqrt{\Delta}}{2} \leq \frac{a+d-|a-d|}{2} = a \wedge d,$$

so that the inequalities follow. If $bc = 0$ then $\sqrt{\Delta} = |a-d|$ and hence $\lambda = a \vee d$ and $\mu = a \wedge d$. The last assertion follows since $\mu = d$ is equivalent to $a - d = \sqrt{(a-d)^2 + 4bc}$ and $\lambda = d$ is equivalent to $d - a = \sqrt{(a-d)^2 + 4bc}$. □

Denote by I the 2×2 identity matrix and define $M_f^0 \equiv I$. Recall that we are assuming throughout that $\det M_f = \lambda\mu \neq 0$, so that neither λ nor μ is zero.

Lemma 3. *$M_f^n = u_n I + v_n M_f$ for every $n \geq 0$, where*

$$u_n = \begin{cases} -\frac{\lambda^n\mu - \lambda\mu^n}{\lambda-\mu} & ;\Delta > 0, \\ -(n-1)\lambda^n & ;\Delta = 0; \end{cases} \qquad v_n = \begin{cases} \frac{\lambda^n - \mu^n}{\lambda-\mu} & ;\Delta > 0, \\ n\lambda^{n-1} & ;\Delta = 0. \end{cases} \tag{18}$$

The n-fold composition $f^{\circ n} = f \circ f \circ \cdots \circ f$ of the Möbius transform f with itself is given by

$$f^{\circ n}(s) = \frac{(u_n + v_n a)s + v_n b}{v_n c s + u_n + v_n d}. \tag{19}$$

Proof. The decomposition $M_f^n = u_n I + v_n M_f$ can be found in Williams (1992). Equation (2) there gives

$$M_f^n = \lambda^n \frac{M_f - \mu I}{\lambda - \mu} + \mu^n \frac{M_f - \lambda I}{\mu - \lambda}$$

for $\lambda \neq \mu$ and $M_f^n = \lambda^{n-1}\left(n M_f - (n-1)\lambda I\right)$ for $\lambda = \mu$, which yields (18) after rearrangement. With (14), this yields

$$M_f^n = u_n I + v_n M_f = \begin{pmatrix} u_n + v_n a & v_n b \\ v_n c & u_n + v_n d \end{pmatrix},$$

a matrix representing the function given in (19). □

Remark 1. It is not hard to show that if M_f is not proportional to I, the decomposition according to Lemma 3 is unique for every $n \geq 0$, in the sense that no other $\tilde{u}_n, \tilde{v}_n$ exist such that $M_f^n = \tilde{u}_n I + \tilde{v}_n M_f$ holds. Also observe that, as expected, $(u_0, v_0) = (1, 0)$ and $(u_1, v_1) = (0, 1)$.

Also, we mention that the fact that, for every $n \geq 0$, M_f^n is a linear combination of I and M_f, which should not be surprising as for any $d \times d$ dimensional matrix A, Cayley–Hamilton's Theorem implies that there are constants $\gamma_0, \ldots, \gamma_{d-1}$ such that $A^d = \sum_{k=0}^{d-1} \gamma_k A^k$, which implies by induction that for every $n \geq d$, A^n is a linear combination of $I, A, \ldots, A^{d-1}$ (for $0 \leq n < d$ this is a triviality). For $d \geq 3$, the coefficients participating in this linear combination may have a substantially messier representation than the simple expressions given in (18). ◇

Remark 2. We note that the case where a, b, c, d are not necessarily nonnegative and/or $ad - bc = 0$ can also be handeled similarly, which can also result in $\Delta < 0$ and therefore λ, μ being complex conjugates of one another. It is easy to check that in this case, u_n, v_n appearing in (18) remain real-valued. Since we will not need this kind of generality, in what follows, we have assumed from the outset that $a, b, c, d \geq 0$ and $ad - bc \neq 0$. ◇

We immediately observe that in light of Lemma 3, (15) and (16), we have for $n \geq 1$ that

$$\begin{aligned} f^{\circ n}(s) &= \frac{1}{c_n}\left(a_n - \frac{(\det M_f)^n}{c_n s + d_n}\right), \\ \frac{\mathrm{d}^m}{\mathrm{d}s^m} f^{\circ n}(s) &= m!\left(-\frac{c_n}{c_n s + d_n}\right)^{m-1} \frac{(\det M_f)^n}{(c_n s + d_n)^2}, \end{aligned}$$

provided that $c_n \neq 0$ for the top equation and $c_n s + d_n \neq 0$ for the bottom equation, where $a_n = u_n + v_n a$, $c_n = v_n c$, and $d_n = u_n + v_n d$.

Let $r = \mu/\lambda$. Noting that $|r| < 1$ if $\Delta > 0$, $r = 1$ if $\Delta = 0$, and that $v_n > 0$ for all $n \geq 1$, it is easy to check that

$$\mu + \frac{u_n}{v_n} = \left\{ \begin{array}{ll} \frac{(1-r)r^n}{1-r^n}\mu & ; \Delta > 0 \\ \frac{1}{n}\mu & ; \Delta = 0 \end{array} \right\} \to 0 \quad \text{for } n \to \infty. \tag{20}$$

Via $M_f^n = u_n I + v_n M_f = v_n\big((\mu + u_n/v_n)I + M_f - \mu I\big)$, this implies that

$$\frac{1}{v_n} M_f^n \to M_f - \mu I, \tag{21}$$

from which it follows that, whenever $c \neq 0$ and $\mu \neq d$,

$$f^{\circ n}(s) = \frac{(u_n + v_n a)s + v_n b}{v_n cs + u_n + v_n d} \to \frac{(a-\mu)s + b}{cs + d - \mu} = \frac{b}{d-\mu} = \frac{a-\mu}{c} \tag{22}$$

as $n \to \infty$. Note that $M_f - \mu I$ is singular and therefore $(a - \mu, b)$ and $(c, d - \mu)$ are linearly dependent and thus the ratio $((a-\mu)s + b)/(cs + d - \mu)$ is constant in s. Letting $s = 0$ and $s \to \pm\infty$ gives the two rightmost ratios. The rightmost equation also follows from the fact that $\det(M_f - \mu I) = 0$.

We observe that when $c \neq 0$, then $\mu = d$ iff $b = 0$ and $d \leq a$. In this case, $f^{\circ n}(s) \to \frac{a-d}{c}$. Similarly, when $c = 0$, then $\mu \neq d$ iff $a < d$, in which case $\mu = a$, which gives $f^{\circ n}(s) \to \frac{b}{d-a}$.

As just observed, convergence of M_f^n is not necessary in order to have convergence of $f^{\circ n}$. Instead, since $a_n M_f^n(s)$, with (a_n) some positive sequence, represents the same Möbius function as $M_{f^{\circ n}}$, it is sufficient to have convergence of $a_n M_f^n$, provided that it does not vanish as $n \to \infty$.

Theorem 4 (Asymptotics of $f^{\circ n}$)**.**

(a) *If $c = 0$, then*

(i) *if $0 < a = d$, then $f^{\circ n}(s) = s + \frac{b}{a}n$;*

(ii) *otherwise*

$$f^{\circ n}(s) = \frac{b}{d-a}\left(1 - \left(\frac{a}{d}\right)^n\right) + \left(\frac{a}{d}\right)^n s. \tag{23}$$

(b) *If $c \neq 0$ and $\Delta > 0$, then $r = \mu/\lambda < 1$ and*

$$f^{\circ n}(s) = \begin{cases} \dfrac{b}{d-\mu} + O(r^n) & ; b \neq 0, \quad (24a) \\ \dfrac{a-d}{c} + O(r^n) & ; b = 0, a > d, \quad (24b) \\ O(r^n) & ; b = 0, a < d. \quad (24c) \end{cases}$$

(c) *If $c \neq 0$ and $\Delta = 0$, then*

$$f^{\circ n}(s) = \frac{as}{cns + a}. \tag{25}$$

For completeness: If $a = c = 0$, a case that is not allowed in the definition (13), then $f^{\circ n}(s)$ is constant.

Proof of Theorem 4. We first consider the case where $c = 0$ (and necessarily $d \neq 0$), i.e. $f(s) = \frac{a}{d}s + \frac{b}{d}$. If $0 \neq a = d$, then $f^{\circ n}(s) = s + \frac{bn}{a}$, so $f^{\circ n}(s) \to \infty$. If $0 < a \neq d$, then

$$f^{\circ n}(s) = \left(\frac{a}{d}\right)^n s + b\frac{\left(\left(\frac{a}{d}\right)^n - 1\right)}{a-d} = \frac{b}{d-a} + \left(s - \frac{b}{d-a}\right)(a/d)^n,$$

which is (23). If $a > d$, then obviously $f^{\circ n}(s) \to \infty$.

Now suppose $c \neq 0$. First note that it follows from (17) that

$$\begin{aligned} (a-\mu)(d-\mu) &= bc \quad \text{and} \quad (a-\lambda)(d-\lambda) = bc, \\ \text{as well as} \quad (a-\mu)(a-\lambda) &= -bc \quad \text{and} \quad (d-\mu)(d-\lambda) = -bc, \end{aligned} \tag{26}$$

where the second row of equalities follows by noting that $\lambda + \mu = a + d$ implies $d - \lambda = \mu - a$. Also recall from Lemma 2 that $\mu \leq a \wedge d$ and $\lambda \geq a \vee d$ with equalities if $bc = 0$.

For the case $\mu < \lambda$, it follows from Lemma 3 that

$$\frac{\lambda - \mu}{\lambda^n} M_f^n = (1 - r^n) M_f - (\mu - \lambda r^n) I = M_f - \mu I - (M_f - \lambda I)\, r^n,$$

so we can write

$$f^{\circ n}(s) = \frac{(a-\mu)s + b - \big((a-\lambda)s + b\big)\, r^n}{cs + d - \mu - (cs + d - \lambda)\, r^n}.$$

Note that by (26),

$$(d-\mu)\big((a-\mu)s + b\big) - b(cs + d - \mu) = \big((a-\mu)(d-\mu) - bc\big)s + b(d-\mu) - b(d-\mu) = 0,$$

$$\begin{aligned} b(cs + d - \lambda) - (d-\mu)\big((a-\lambda)s + b\big) &= \big(bc - (d-\mu)(a-\lambda)\big)s + b(\mu - \lambda) \\ = \big((a-\mu) - (a-\lambda)\big)(d-\mu)s + b(\mu-\lambda) &= (\lambda-\mu)\big((d-\mu)s - b\big). \end{aligned}$$

So

$$f^{\circ n}(s) - \frac{b}{d-\mu} = \frac{(\lambda-\mu)\big((d-\mu)s - b\big)\, r^n}{(d-\mu)\big(cs + d - \mu - (cs + d - \lambda)\, r^n\big)},$$

so that (24a) follows.

Suppose that $b = 0$, so $\mu = a \wedge d$ and $\lambda = a \vee d$. We can easily show by induction that

$$f^{\circ n}(s) = \frac{a^n (a-d)s}{a^n cs - d^n(d - a + cs)}. \tag{27}$$

If $a > d$, then we can write (27) as

$$f^{\circ n}(s) = \frac{(a-d)s}{cs - r^n(d - a + cs)} = \frac{a-d}{c} + O(r^n). \tag{28}$$

If $a < d$, then

$$f^{\circ n}(s) = \frac{r^n(a-d)s}{a - d - (1 - r^n)cs} = O(r^n).$$

If $\lambda = \mu$ $(= a = d)$, then we necessarily have $b = 0$ since $\Delta = 0$ and $c \neq 0$. It follows that $f(s) = \frac{as}{cs+a}$. It is not difficult to show that (25) holds.

Finally, if $a = c = 0$, then $f(s) = \frac{b}{d}$. □

3.2 Further properties of the modulating operator

By well-known properties of random sums, the probability generating function of $\theta \odot X$ is $\mathcal{P}_{\theta\odot X}(s) = \mathcal{P}_X(\mathcal{P}_Z(s))$. Consequently, keeping (7) in mind, the apgf is given by the simple expression

$$\mathcal{G}_{\theta\odot X}(s) = \mathcal{P}_{\theta\odot X}(1-s) = \mathcal{P}_X(\mathcal{P}_Z(1-s)) = \mathcal{G}_X\left(1 - \mathcal{G}_Z(s)\right) = \mathcal{G}_X\left(f_\theta(s)\right), \tag{29}$$

where the Möbius transform f_θ and its associated matrix M_θ are given by

$$f_\theta(s) = \frac{\alpha s}{1 + \beta s}, \qquad M_\theta = \begin{pmatrix} \alpha & 0 \\ \beta & 1 \end{pmatrix}. \tag{30}$$

Table 2: Correspondences between Möbius transforms and $(\Theta, \odot)$.

	Möbius transform	Associated matrix	Element of Θ
Definition	$f_\theta(s) = \frac{\alpha s}{1+\beta s}$	$M_\theta = \begin{pmatrix} \alpha & 0 \\ \beta & 1 \end{pmatrix}$	θ
Operation	$f_\theta(s) = f_{\theta_1}(f_{\theta_2}(s))$	$M_\theta = M_{\theta_1} M_{\theta_2}$	$\theta = \theta_1 \odot \theta_2$

With a slight abuse of notation, we use the simpler M_θ here instead of M_{f_θ}. As an example, if $X \sim \text{BerG}\,(\gamma, \delta)$, then

$$\mathcal{G}_{\theta \odot X}(s) = 1 - \frac{\gamma\, f_\theta(s)}{1 + \delta\, f_\theta(s)} = 1 - \frac{\gamma\alpha\, s}{1 + (\beta + \delta\alpha)\, s}, \tag{31}$$

i.e. the modulating operator preserves the zero-truncated geometric distribution, with $\theta \odot X \sim \text{BerG}\,(\gamma\alpha, \beta + \delta\alpha)$ (this is a special case of Lemma 5 below).

With respect to matrix multiplication, the set of matrices of this particular type, with $(\alpha, \beta) \in \Theta$, form a non-commutative monoid with unit I, the 2×2 identity matrix. The corresponding Möbius transforms inherit the same structure, with function composition serving as the operation. It then makes sense to define a corresponding operation '$\odot$' also on our parameter space Θ, in a way that $\theta = \theta_1 \odot \theta_2$ is defined such that $M_\theta = M_{\theta_1} M_{\theta_2}$ (cf. the operation (2.6) in Aly & Bouzar (2019) and the results following it). Then, $(\Theta, \odot)$ becomes a non-commutative monoid, too. We thus have the correspondences summarized in Table 2.

Using the above association with matrix multiplication, we easily show the following lemma.

Lemma 5. *Let* $(\theta_j)_{j=1,2,\ldots,n} = (\alpha_j, \beta_j)_{j=1,2,\ldots,n} \in \Theta$. *Then*

$$\theta_n \odot \theta_{n-1} \odot \cdots \odot \theta_1 = \left(\prod_{i=1}^{n} \alpha_i, \sum_{j=1}^{n} \beta_j \prod_{i=1}^{j-1} \alpha_i \right). \tag{32}$$

If the θ_j *are identical, then*

$$\theta^{\odot n} = \theta \odot \theta \odot \cdots \odot \theta = \begin{cases} (1, n\beta) & ; \alpha = 1 \\ \left(\alpha^n, \frac{1-\alpha^n}{1-\alpha}\beta\right) & ; \alpha \neq 1. \end{cases} \tag{33}$$

Proof. The results follow easily from multiplication of M_{θ_2} and M_{θ_1}. Then $(\alpha_2, \beta_2) \odot (\alpha_1, \beta_1) = (\alpha_2\alpha_1, \beta_2\alpha_1 + \beta_1)$. Using induction, one obtains (32), which, in turn, implies (33) as a special case. □

Remark 3. The particular case (33) can also be found in Aly & Bouzar (2019). This is basically (5) in Athreya & Ney (2012), p. 7, where generating functions of the form

$$\mathcal{P}_Z(s) = 1 - \frac{b}{1-p} + \frac{bs}{1-ps}$$

are considered, corresponding to $\text{BerG}\,(\alpha, \beta)$ distributions with $\alpha = \frac{b}{(1-p)^2}$ and $\beta = \frac{p}{1-p}$. ◇

With $\Rightarrow$ denoting convergence in distribution, we have the following.

Theorem 6. *Consider a sequence* $(\theta_j)_{j=1,2,\ldots,n}$ *as in Lemma 5 and let* $X_0 = X$ *and* $X_n = \theta_n \odot X_{n-1}$, *where we assume that all modulations are performed with independent counting variables. Let*

$A_n = \prod_{i=1}^{n} \alpha_i$ *and* $B_n = \sum_{j=1}^{n} \beta_j \prod_{i=1}^{j-1} \alpha_i$. *Then, the following is true.*

(*a*) *If the limits* $A = \lim_{n\to\infty} A_n$ *and* $B := \lim_{n\to\infty} B_n$ *both exist and are finite, then* $X_n \Rightarrow (A, B) \odot X$.

(*b*) *If* $A_n \to \infty$ (*and necessarily* $B_n \to \infty$) *and* $q^* = \lim_{n\to\infty} A_n/B_n \in [0, 1]$ *exists, then* $X_n \Rightarrow X^*$, *where* X^* *is degenerate with values in* $\{0, \infty\}$ *and* $\mathbb{P}(X^* = 0) = \mathcal{G}_X(q^*)$.

Proof. Assertion (a) is an immediate consequence of $\mathcal{G}_{X_n}(s) = \mathcal{G}_X(\frac{A_n s}{1+B_n s})$ (see (32)) and the continuity theorem for probability generating functions. To prove (b), note that $\mathcal{G}_{X_n}(s) \to \mathcal{G}_X(q^*)$ as $n \to \infty$, which is indeed the apgf of a degenerate random variable with values 0 and ∞ (a version of the continuity theorem including degenerate random variables can be found in Fristedt & Gray (1997), Theorem 22, p. 262). □

Suppose that X has a Poisson distribution with mean λ, then

$$\mathcal{G}_{\theta\odot X}(s) = \exp\left(-\frac{\lambda\alpha s}{1+\beta s}\right), \tag{34}$$

i.e., $\theta \odot X$ has a *geometric Poisson distribution*, also known as *Pólya-Aeppli distribution* (Johnson et al. 2005, Section 9.7). The following extension of Theorem 7 in Kella & Löpker (2023) is quite immediate.

Theorem 7 (Geometric Poisson approximation)**.** *Assume that for every* $n \geq 1$, $X_1^{(n)}, X_2^{(n)}, \ldots X_n^{(n)}$ *are non-negative integer-valued i.i.d., and let* $S_n = \sum_{i=1}^{n} X_i^{(n)}$. *Suppose that* $X_1^{(n)} \Rightarrow X$, $EX_1^{(n)} \to EX_1 < \infty$, $n\alpha_n \to \lambda$ *and* $\beta_n \to \kappa$ *as* $n \to \infty$. *Also let* $(\theta_j)_{j\in\mathbb{N}} = (\alpha_j, \beta_j)_{j\in\mathbb{N}} \in \Theta$ *as before. Then,* $\theta_n \odot S_n \Rightarrow X^*$, *where* X^* *has a geometric Poisson distribution with apgf* $\mathcal{G}_{X^*}(s) = \exp\left(-\frac{\lambda\mathbb{E}(X)s}{1+\kappa s}\right)$.

Proof. The proof follows from Kella & Löpker (2023) by observing that $\theta_n \odot S_n$ is a sum of $\alpha_n \circ S_n$ independent (of $\alpha_n \circ S_n$) and i.i.d. geometrically distributed random variables with parameter β_n, where $\alpha_n \circ S_n$ is an α_n-binomial thinning of S_n. It was shown there that, under the conditions specified, $\alpha_n \circ S_n$ converges in distribution to a Poisson distribution with parameter λEX, and the geometric distribution with parameter β_n obviously converges to a geometric distribution with parameter κ. It follows e.g. from Theorem 4.1.2 in Gnedenko & Korolev (2020) that if $N_n, Y_1^{(n)}, Y_2^{(n)}, \ldots$ are independent, N_n is non-negative integer valued, $Y_1^{(n)}, Y_2^{(n)}, \ldots$ are i.i.d., $N_n \Rightarrow N$ and $Y_1^{(n)} \Rightarrow Y$, then $\sum_{i=1}^{N_n} Y_i^{(n)} \Rightarrow \sum_{i=1}^{N} Y_i$, where $N, Y_1, Y_2, \ldots$ are independent and $Y_i \sim Y$ then $S_{N_n}^{(n)} \Rightarrow \sum_{i=1}^{U} Y_k$. □

3.3 Special cases of modulating operator

After having established the relation between Möbius transforms and the modulating operator in Section 3.2, we can now utilize this relation to consider some interesting special cases (cf. p. 95–96 in Aly & Bouzar (2019)).

Example 1. If $\beta = 0$, then $\alpha \in [0, 1]$, and the Z_i have a Bernoulli distribution with $\mathbb{P}(Z = 1) = \alpha$. In this case, the operator $\theta \odot X$ is the well-known *binomial thinning*, usually denoted by $\alpha \circ X$, see Steutel & van Harn (1979). As a sum of independent Bernoulli random variables, $(\alpha \circ X|X) \sim \text{Bin}(X; \alpha)$. The thinning operator results in a simple multiplication in the argument of the apgf: $\mathcal{G}_{\theta\odot X}(s) = \mathcal{G}_X(\alpha s)$. The associated matrices $M_\alpha := \left(\begin{smallmatrix}\alpha & 0\\ 0 & 1\end{smallmatrix}\right)$ are closed under multiplication, $M_{\alpha_1} M_{\alpha_2} = M_{\alpha_1\alpha_2}$, so $\alpha_1 \circ (\alpha_2 \circ X) \overset{d}{\sim} (\alpha_1\alpha_2) \circ X$. □

Example 2. If $\alpha = \beta \geq 0$, then the Z_i have a geometric distribution on $\mathbb{N}_0$ with parameter $p = 1/(1+\beta)$. This corresponds to *negative-binomial thinning* as proposed by Ristić et al. (2009). □

Example 3. For $\alpha = \beta + 1 \geq 1$, the Z_i are geometrically distributed on $\mathbb{N}$ with the same p as in Example 2. Hence, $\theta \odot X$ is truly positive unless $X = 0$. More generally, $\theta \odot X \geq X$ holds a.s., which is opposite to the binomial thinning's property $\alpha \circ X \leq X$, also see Examples 8 and 9 below. Furthermore, (33) implies that $\theta^{\odot n} = (\alpha^n, \alpha^n - 1)$, which also covers the case $\alpha = 1$. □

Example 4. If $\alpha = 2\beta$, then necessarily $\beta \leq 1$, $\alpha \leq 2$, and

$$\mathbb{P}(Z = k) = \begin{cases} \frac{1-\beta}{1+\beta} & ; k = 0 \\ \frac{2}{1+\beta}\left(\frac{\beta}{1+\beta}\right)^k & ; k = 1, 2, 3, \ldots \end{cases}.$$

In this case, Z is equidispersed (but non-Poisson), i.e. $\mathbb{E}(Z) = \mathrm{Var}(Z) = \alpha$, recall (8). Lemma 1(b) then implies $\mathrm{Var}(\theta \odot X) = \alpha \mathbb{E}(X) + \alpha^2 \mathrm{Var}(X)$. □

Example 5. In the case $\alpha = 1$, where $\theta = (1, \beta)$, the associated Möbius transform is parabolic. Let us call the associated operator *parabolic*. In this case, the modulating operator conserves expectations, i.e. $\mathbb{E}(\theta \odot X) = \mathbb{E}(X)$. The family of associated Möbius matrices $M_\theta = \left(\begin{smallmatrix} 1 & 0 \\ \beta & 1 \end{smallmatrix}\right)$ is closed w.r.t. matrix multiplication. Hence, these matrices as well as the Möbius transforms and θ parameters of the form $(1, \beta)$ result in submonoids. In fact, $(1, \beta_1) \odot (1, \beta_2) = (1, \beta_1 + \beta_2)$. □

The parabolic operator from Example 5 might be used within the following decomposition.

Proposition 8 (Decomposition)**.** *Let* $(\alpha, \beta) \in \Theta$, *and let* $\delta \in [(\alpha - 1)^+, \beta]$ *be arbitrary. Then,* (α, β) *can be decomposed into products from the left and from the right of* (α, δ) *with a parabolic parameter*:

$$(\alpha, \beta) = (\alpha, \delta) \odot (1, \beta - \delta) = \left(1, \tfrac{\beta - \delta}{\alpha}\right) \odot (\alpha, \delta). \tag{35}$$

Proof. Follows by multiplication of the associated matrices. □

For example, if one takes $\delta = 0$, then

$$(\alpha, \beta) = \left(1, \tfrac{\beta}{\alpha}\right) \odot (\alpha, 0) = (\alpha, 0) \odot (1, \beta), \tag{36}$$

i.e., the operator with (α, β) is equivalent to successively having a binomial thinning and a parabolic operator, or first a parabolic operator and then a binomial thinning. The first decomposition in (36) shares analogies to the thinning operator proposed by Al-Osh & Aly (1992), although their counting variables are not parabolic.

Furthermore, if $\alpha = 1$ (and $\delta \in [0, \beta]$), we have commutativity:

$$(1, \beta) = (1, \delta) \odot (1, \beta - \delta) = (1, \beta - \delta) \odot (1, \delta).$$

4 The geometric INAR-type process

In this section, we consider the well-studied geometric INAR(1) process

$$X_{n+1} = \theta \odot X_n + \varepsilon_n, \tag{37}$$

with $(\varepsilon_i)_{i\in\mathbb{N}_0}$ i.i.d., distributed like ε, non-negative integer valued, not a.s. zero, and independent of everything else. We assume that $\mathbb{E}(\varepsilon) < \infty$. See Andrews & Balakrishna (2023), Aly & Bouzar (2019) and the literature survey in Section 1. The process can also be seen as a branching process with immigration (Athreya & Ney (2012), VI.7), where the reproduction distribution is a T-geometric distribution and the 'innovations' ε represent additionally arriving members of the population.

Remark 4. When $\varepsilon_n \equiv 0$ and $\alpha > 1$ then by (33) $X_n \overset{d}{\sim} (\alpha^n, \beta(\alpha^n - 1)/(\alpha - 1)) \odot X_0$ and it follows easily from Theorem 6 that given $X_0 = n$, we have

$$\mathbb{P}(X^* = 0 | X_0 = n) = \left(1 - \frac{\alpha - 1}{\beta}\right)^n, \tag{38}$$

which resembles the 'ruin probability' in a simple random walk with positive drift. Indeed, if $R_n = R_{n-1} + B_n$ with $B_n \in \{-1, 1\}$ and $\mathbb{P}(B_n = 1) = p > 1/2$, then $R_n \Rightarrow R^*$ with R^* degenerate and $\mathbb{P}(R^* = 0 | R_0 = n) = ((1-p)/p)^n$ (see e.g. Feller (1968), XIV.2). ◇

As mentioned in Section 1, there is a vast literature regarding geometric INAR(1)-type models. But utilizing the relation of '⊙' to Möbius transforms, we can complement the existing results by some new closed-form expression for h-step-ahead properties. To this end, note that

$$\mathcal{G}_{X_{n+1}}(s) = \mathcal{G}_{X_n}(f_\theta(s))\,\mathcal{G}_\varepsilon(s), \tag{39}$$

and hence

$$\mathcal{G}_{X_n}(s) = \mathcal{G}_{X_0}\left(f_\theta^{\circ n}(s)\right) \prod_{i=0}^{n-1} \mathcal{G}_\varepsilon(f_\theta^{\circ i}(s)), \tag{40}$$

where $f_\theta(s) = \alpha s/(1 + \beta s)$ as before. This is in accordance with well-known results in the branching process literature, see e.g. Athreya & Ney (2012), p. 263. From there, we also cite the following limit theorem.

Theorem 9 (Athreya & Ney (2012), Theorem 1, p. 263)**.** *If $\alpha < 1$, then $X_n \Rightarrow X^*$ as $n \to \infty$, where*

$$\mathcal{G}_{X^*}(s) = \prod_{i=0}^{\infty} \mathcal{G}_\varepsilon(f_\theta^{\circ i}(s)). \tag{41}$$

In Andrews & Balakrishna (2023), the stationary solution to model (37) was studied in the case, where the marginal distribution is geometric, namely $\mathcal{G}_{X^*}(s) = (1 + \kappa s)^{-1}$ with mean parameter $\kappa > 0$. Solving (39) in $\mathcal{G}_\varepsilon(s)$, it follows that

$$\mathcal{G}_\varepsilon(s) = \frac{1 + (\beta + \alpha\kappa)s}{(1 + \beta s)(1 + \kappa s)} = \left(1 - \frac{\alpha\kappa}{\kappa - \beta}\right)\frac{1}{1 + \kappa s} + \frac{\alpha\kappa}{\kappa - \beta}\frac{1}{1 + \beta s}, \tag{42}$$

which is a mixture of two geometric distributions with mean parameters κ and β, respectively. Note that $\beta < \kappa(1 - \alpha)$ is necessary to ensure that this expression for $\mathcal{G}_\varepsilon(s)$ constitutes a valid apgf.

Using Lemma 5, we can now derive some novel closed-form expressions for h-step-ahead properties of model (37) if we assume $\alpha < 1$, i.e. the stationary case. Using $f_\theta(s)$ together with

(33), we get $f^{\circ i}(s) = \frac{(1-\alpha)\alpha^i s}{\beta(1-\alpha^i)s+1-\alpha}$. Therefore, see (40), the conditional h-step-ahead apgf equals

$$\mathscr{G}_{X_{n+h}|X_n}(s) = \left(1 - \frac{(1-\alpha)\alpha^h s}{\beta(1-\alpha^h)s + 1 - \alpha}\right)^{X_n} \prod_{i=0}^{h-1} \mathscr{G}_\varepsilon\left(\frac{(1-\alpha)\alpha^i s}{\beta(1-\alpha^i)s + 1 - \alpha}\right).$$

In the special case (42) as considered by Andrews & Balakrishna (2023), this expression further simplifies ('telescope product') to

$$\mathscr{G}_{X_{n+h}|X_n}(s) = \frac{1 - \alpha + \beta s + s\alpha^h\big((1-\alpha)\kappa - \beta\big)}{(1+\kappa s)\big(1 - \alpha + \beta(1-\alpha^h)s\big)} \left(1 - \frac{(1-\alpha)\alpha^h s}{\beta(1-\alpha^h)s + 1 - \alpha}\right)^{X_n},$$

which can be applied for time series forecasting. Note that for $h \to \infty$, we have $\alpha^h \to 0$ such that the apgf $\mathscr{G}_{X_{n+h}|X_n}(s)$ converges to $\mathscr{G}_{X^*}(s) = (1 + \kappa\, s)^{-1}$, in agreement with the ergodicity of the INAR(1) process.

5 A minification process

In this section, we consider a modification of the INAR(1) process (37), where the sum is substituted by a minimum operation. Hence, the model recursion shall be of the form $X_{n+1} = (\theta \odot X_n) \wedge U_n$, see Section 5.2 for details, which we refer to as a 'minification process'. The process might also be described as a 'gated population process'. This terminology is motivated by the following interpretation: in analogy to classical branching processes, $\theta \odot X_n$ might be understood as the number of offsprings generated by the population X_n, which try to pass a gate successively. But the maximal number of individuals allowed to pass the gate is limited by U_n such that the number of individuals that actually pass the gate is given by $(\theta \odot X_n) \wedge U_n$.

5.1 Prerequisites

In this section, we consider the minimum $Z \wedge Y$ of a variable $Z \in \mathrm{BerG}(\theta)$ with some independent non-negative integer-valued r.v. Y. We start with a useful lemma.

Lemma 10. *Let $Z \in \mathrm{BerG}(\alpha, \beta)$ and Y non-negative integer-valued, independent of Z. Let $f : \mathbb{N}_0 \to \mathbb{R}$ be such that $\mathbb{E}f(Z)$, $\mathbb{E}f(Y)$, $\mathbb{E}f(Z \wedge Y)$, and $\mathbb{E}f(Z+Y)$ are finite. Then,*

$$\mathbb{E}f(Z \wedge Y) = \mathbb{E}f(Z) + \mathbb{E}\left((f(Y) - f(Z+Y)) \cdot q^Y\right), \tag{43}$$

where we set $q = \beta/(1+\beta)$.

Proof. Let $X \sim \mathrm{Geo}_{\mathbb{N}}(p)$ and let $q = 1 - p$. Then,

$$\mathbb{E}f(X \wedge Y) = \mathbb{E}\left(f(X)\mathbb{I}_{\{X \le Y\}}\right) + \mathbb{E}\left(f(Y)\mathbb{I}_{\{X > Y\}}\right).$$

Now, $\mathbb{P}(X > y) = q^y$ for $y \in \mathbb{N}_0$, so

$$\mathbb{E}\left(f(Y)\mathbb{I}_{\{X>Y\}}\right) = \mathbb{E}\left(f(Y)q^Y\right).$$

By the memoryless property, $f(X)\mathbb{I}_{\{X>Y\}} \sim f(X'+Y)\mathbb{I}_{\{X>Y\}}$, where X, Y, X' are independent and $X' \sim X$. Therefore,

$$\mathbb{E}\left(f(X)\mathbb{I}_{\{X>Y\}}\right) = \mathbb{E}\left(f(X'+Y)\mathbb{I}_{\{X>Y\}}\right) = \mathbb{E}\left(f(X'+Y)q^Y\right) = \mathbb{E}\left(f(X+Y)q^Y\right). \tag{44}$$

This implies that

$$\mathbb{E}\left(f(X)\mathbb{I}_{\{X\le Y\}}\right) = \mathbb{E}\left(f(X)(1-\mathbb{I}_{\{X>Y\}})\right) = \mathbb{E}f(X) - \mathbb{E}\left(f(X+Y)q^Y\right) .$$

Therefore,

$$\mathbb{E}f(X\wedge Y) = \mathbb{E}f(X) + \mathbb{E}\left(f(Y)q^Y\right) - \mathbb{E}\left(f(X+Y)q^Y\right) . \tag{45}$$

Now let $B \sim \mathrm{Ber}(\alpha p)$ be a Bernoulli random variable such that X, Y, B are independent and set $Z = BX$. Then, $Z\wedge Y = B(X\wedge Y)$, and we obtain

$$\mathbb{E}\left(f(Z+Y)q^Y\right) = (1-\alpha p)\mathbb{E}\left(f(Y)q^Y\right) + \alpha p\mathbb{E}\left(f(X+Y)q^Y\right).$$

Hence, assuming for a moment that $f(0) = 0$,

$$\begin{aligned}\mathbb{E}f(Z\wedge Y) &= \alpha p\,\mathbb{E}\left(f(X\wedge Y)\right) = \alpha p\left(\mathbb{E}f(X) + \mathbb{E}\left(f(Y)q^Y\right) - \mathbb{E}\left(f(X+Y)q^Y\right)\right)\\ &= \mathbb{E}f(Z) + \alpha p\mathbb{E}\left(f(Y)q^Y\right) - \mathbb{E}\left(f(Z+Y)q^Y\right) + (1-\alpha p)\mathbb{E}\left(f(Y)q^Y\right)\\ &= \mathbb{E}f(Z) + \mathbb{E}\left((f(Y) - f(Z+Y))q^Y\right). \end{aligned}\tag{46}$$

If $f(0) = C \neq 0$, then by the above,

$$\mathbb{E}\left(f(Z\wedge Y) - C\right) = \mathbb{E}\left(f(Z) - C\right) + \mathbb{E}\left((f(Y) - f(Z+Y))q^Y\right),$$

so that also in this case, (46) holds. □

If we let $f(x) = 1-(1-s)^x$ in Lemma 10, we obtain the following interesting decomposition formula for the apgf of $Z\wedge Y$.

Corollary 11. *Let $Z \in \mathrm{BerG}\,(\alpha,\beta)$ and let the non-negative integer-valued r.v. Y be independent of Z. Then, the apgf of the minimum of Y and Z has the following factorization*:

$$1 - \mathcal{G}_{Z\wedge Y}(s) = \left(1 - \mathcal{G}_Y\left(\tfrac{1+\beta s}{1+\beta}\right)\right) \times \left(1 - \mathcal{G}_Z(s)\right). \tag{47}$$

In particular, if $X \sim \mathrm{BerG}\,(\alpha,\beta)$ and $Y \sim \mathrm{BerG}\,(\gamma,\delta)$ are independent, then

$$X\wedge Y \sim \mathrm{BerG}\left(\frac{\alpha\gamma}{1+\beta+\delta}, \frac{\beta\delta}{1+\beta+\delta}\right).$$

It follows that $\mathrm{BerG}\,(\cdot,\cdot)$ is closed w.r.t. to the minimum operation.

Letting $\alpha = \beta$, we can reproduce Proposition 1 in Barreto-Souza et al. (2023):

$$\mathcal{P}_{Z\wedge Y}(s) = \frac{1 - \alpha(1-s)\,\mathcal{P}_Y\left(\frac{\alpha s}{1+\alpha}\right)}{1+\alpha(1-s)}.$$

If we apply Lemma 10 to $f(x) = x^n$ (and assume existence of the moments), we obtain

$$\mathbb{E}\left((Z\wedge Y)^n\right) = \mathbb{E}\left(Z^n\right)\left(1 - \mathbb{E}(q^Y)\right) - \sum_{i=1}^{n-1}\binom{n}{i}\mathbb{E}\left(Z^{n-i}\right)\mathbb{E}\left(Y^i q^Y\right) \tag{48}$$

with $q = \beta/(1+\beta)$ as above.

5.2 Definition of the process

Let X_0 be non-negative integer-valued, and define for $n \geq 1$

$$X_{n+1} = (\theta \odot X_n) \wedge U_n, \tag{49}$$

where $\theta = (\alpha, \beta) \in \Theta$ and $(U_n)_{n \in \mathbb{N}_0}$ is a sequence of i.i.d. BerG (η)-distributed random variables, $\eta = (\gamma, \delta)$. In view of the interpretation of the process sketched in the beginning of Section 5, we may refer to the U_n as the 'gate thresholds'. As always, we assume that the modulations associated with the process are independent of each other. Note that, in a sense, our model is opposite to the max-INAR(1) model of Scotto et al. (2018).

It follows from (47) that

$$1 - \mathcal{G}_{X_{n+1}}(s) = \left(1 - \mathcal{G}_{\theta \odot X_n}\left(\tfrac{1+\delta s}{1+\delta}\right)\right) \times \frac{\gamma s}{1 + \delta s} = \left(1 - \mathcal{G}_{X_n}\left(\Psi(s)\right)\right) \times \frac{\gamma s}{1 + \delta s},$$

where

$$\Psi(s) = \frac{\alpha\left(\frac{1+\delta s}{1+\delta}\right)}{1 + \beta\left(\frac{1+\delta s}{1+\delta}\right)} = \frac{\alpha(1 + \delta s)}{1 + \delta + \beta(1 + \delta s)}, \qquad M_\Psi = \begin{pmatrix} \alpha\delta & \alpha \\ \beta\delta & 1 + \beta + \delta \end{pmatrix}. \tag{50}$$

The eigenvalues corresponding to M_Ψ are

$$\mu = \frac{1 + (\alpha + 1)\delta + \beta - \sqrt{\Delta}}{2}, \quad \lambda = \frac{1 + (\alpha + 1)\delta + \beta + \sqrt{\Delta}}{2}, \tag{51}$$

with discriminant $\Delta = (\delta(\alpha - 1) - \beta - 1)^2 + 4\alpha\beta\delta$. A zero discriminant would imply either (1) $\beta = 0$ (so $\alpha \leq 1$) and $\delta(\alpha - 1) = 1$, which is not possible, or (2) $\delta = 0$ and then $-\beta - 1 = 0$, which is also impossible. Hence, $\Delta > 0$.

Iterating (49) yields the finite-time apgf of the process:

$$\mathcal{G}_{X_n}(s) = 1 - \left(1 - \mathcal{G}_{X_0}\left(\Psi^{\circ n}(s)\right)\right) \times \prod_{i=0}^{n-1} \frac{\gamma \Psi^{\circ i}(s)}{1 + \delta \Psi^{\circ i}(s)}, \tag{52}$$

where as before, $\Psi^{\circ 0}(s) = s$ and $\Psi^{\circ i}(s) = \Psi\big(\Psi^{\circ(i-1)}(s)\big)$.

Applying (48) and (52), also recalling (29), i.e. $\mathcal{G}_{\theta \odot X}(s) = \mathcal{G}_X(f_\theta(s))$, we obtain

$$\begin{aligned}
\mathbb{E}\left(X_{n+1}\right) = \mathbb{E}\left((\theta \odot X_n) \wedge U_n\right) &= \mathbb{E}\left(U_n\right)\left(1 - \mathcal{G}_{\theta \odot X_n}\left(\tfrac{1}{1+\delta}\right)\right) = \mathbb{E}\left(U_n\right)\left(1 - \mathcal{G}_{X_n}\left(\frac{\alpha}{1 + \delta + \beta}\right)\right) \\
&= \mathbb{E}\left(U_n\right)\left(1 - \mathcal{G}_{X_n}(\Psi(0))\right) = \gamma\left(1 - \mathcal{G}_{X_0}\left(\Psi^{\circ n}(\Psi(0))\right)\right) \times \prod_{i=0}^{n-1} \frac{\gamma \Psi^{\circ i}(\Psi(0))}{1 + \delta \Psi^{\circ i}(\Psi(0))}.
\end{aligned}$$

It follows that for $n \geq 1$,

$$\mathbb{E}\left(X_n\right) = \gamma\left(1 - \mathcal{G}_{X_0}\left(\Psi^{\circ n}(0)\right)\right) \times \prod_{i=1}^{n-1} \frac{\gamma \Psi^{\circ i}(0)}{1 + \delta \Psi^{\circ i}(0)}. \tag{53}$$

Using Lemma 3 and (19), we can express the $\Psi^{\circ n}(0)$, $n \geq 1$, in terms of the two eigenvalues of

M_Ψ given in (51):

$$\Psi^{\circ n}(0) = \frac{v_n b}{u_n + v_n d} = \frac{\alpha}{1 + \beta + \delta - \lambda\mu \frac{\lambda^{n-1} - \mu^{n-1}}{\lambda^n - \mu^n}}.$$

Remark 5. Since $\mathbb{E}(U) = \gamma$ and $\mathbb{P}(U = 0) = 1 - \gamma/(1+\delta)$, we expect that the process will behave more and more like the simple process defined by $Y_n = \theta \odot Y_{n-1}$ as γ becomes large and γ and δ stay close. Indeed, if $\gamma \to \infty$ and $\delta = \gamma \cdot (1 + o(1))$, then the product

$$\prod_{i=0}^{n-1} \frac{\gamma \Psi^{\circ i}(s)}{1 + \delta \Psi^{\circ i}(s)} \tag{54}$$

will tend to one and $\mathscr{G}_{X_n}(s) \to \mathscr{G}_{X_0}(\Psi^{\circ n}(s))$, thus $X_n \Rightarrow Y_n$ (if $Y_0 = X_0$). ◇

Remark 6. There are at least two cases when the product (54) becomes 'telescoping' and hence can be simplified considerably. One case, when $\alpha = \beta + 1$ and $\gamma = \delta + 1$, will be discussed in the next section. In the other case, $\beta = 0$, i.e. $\alpha \le 1$, and hence $\theta \odot X = \alpha \circ X$ is just a binomial thinning. Indeed, here

$$\Psi(s) = \frac{\alpha(1 + \delta s)}{1 + \delta}, \qquad M_\Psi = \begin{pmatrix} \alpha\delta & \alpha \\ 0 & 1+\delta \end{pmatrix},$$

and it follows from Theorem 4 (a) (ii) that

$$\Psi^{\circ i}(s) = \frac{\rho}{\delta} \frac{1 - \rho^i}{1 - \rho} + \rho^i s, \quad i \ge 1,$$

where $\rho = \alpha\delta/(1+\delta)$. Consequently,

$$1 + \delta \Psi^{\circ i}(s) = 1 + \rho \frac{1 - \rho^i}{1 - \rho} + \delta \rho^i s = \frac{\delta}{\rho} \left(\frac{\rho}{\delta} \frac{1 - \rho^{i+1}}{1 - \rho} + \rho^{i+1} s \right) = \frac{\delta}{\rho} \Psi^{\circ(i+1)}(s),$$

and hence

$$\prod_{i=0}^{n-1} \frac{\gamma \Psi^{\circ i}(s)}{1 + \delta \Psi^{\circ i}(s)} = \prod_{i=0}^{n-1} \frac{\gamma \Psi^{\circ i}(s)}{\frac{\delta}{\rho} \Psi^{\circ(i+1)}(s)} = \left(\frac{\gamma\rho}{\delta} \right)^n \frac{s}{\Psi^{\circ n}(s)}. \tag{55}$$

◇

5.3 Stationarity

Suppose that either $\gamma < 1 + \delta$ and then $\mathbb{P}(U_n = 0) = 1 - \frac{\gamma}{1+\delta} > 0$, or $\alpha < 1 + \beta$ and then $\mathbb{P}(\theta \odot X_n = 0) > 0$. Then, the first time instance when the process becomes zero, $\tau = \inf\{n : X_n = 0\}$, say, is finite a.s. with a finite mean, and $X_n = 0$ a.s. for $n \ge \tau$ (in fact 0 is absorbing). Thus, only the case where $\alpha = \beta + 1$ and $\gamma = \delta + 1$ is interesting when it comes to the behavior of X_n as $n \to \infty$. This is the case when both the counting series of the operator '$\odot$' and the gate thresholds U_n have a zero-truncated distribution, namely a geometric distribution on $\mathbb{N}$.

We additionally assume that $\mathbb{P}(X_0 > 0) = 1$, so the process does not already start in zero. Hence, we assume in this section that

$$\mathbb{P}(X_0 > 0) = 1, \quad \alpha = \beta + 1, \quad \gamma = \delta + 1, \tag{$\star$}$$

In particular, we have $\alpha \ge 1$ and $\gamma \ge 1$. Whether X_n converges in distribution to a limit depends on the convergence of (52), which is now

$$\mathcal{G}_{X_n}(s) = 1 - \big(1 - \mathcal{G}_{X_0}(\Psi^{\circ n}(s))\big) \times \prod_{i=0}^{n-1} \frac{(\delta+1)\Psi^{\circ i}(s)}{1+\delta\Psi^{\circ i}(s)}. \tag{56}$$

Inserting $a = \alpha\delta$, $b = \alpha$, $c = (\alpha-1)\delta$, and $d = \alpha+\delta$ into Equation (19), we get

$$\Psi^{\circ n}(s) = \frac{\alpha + (\alpha-1)\delta s - \alpha(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^n}{\alpha + (\alpha-1)\delta s + (\alpha-1)\delta(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^n}$$

and thus

$$\frac{(1+\delta)\Psi^{\circ n}(s)}{1+\delta\Psi^{\circ n}(s)} = \frac{\alpha + (\alpha-1)\delta s - \alpha(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^n}{\alpha + (\alpha-1)\delta s - \alpha(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^{n+1}}.$$

Hence, being concerned with a kind of 'telescope product', we get

$$\prod_{i=0}^{n-1} \frac{(1+\delta)\Psi^{\circ i}(s)}{1+\delta\Psi^{\circ i}(s)} = \frac{s\big(\alpha + (\alpha-1)\delta\big)}{\alpha + (\alpha-1)\delta s - \alpha(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^n}.$$

Since $\alpha \geq 1$, it follows that $\frac{\delta}{\alpha(1+\delta)} < 1$, and hence we have $\Psi^{\circ i}(s) \to 1$. Thus, we obtain in the limit from (56)

$$\mathcal{G}_{X^*}(s) = 1 - \frac{s\big(\alpha + (\alpha-1)\delta\big)}{\alpha + (\alpha-1)\delta s} = \frac{\alpha(1-s)}{\alpha + (\alpha-1)\delta s}. \tag{57}$$

We summarize:

Theorem 12. *If* ($\star$) *holds, then* X_n *converges in distribution to a geometric random variable* $X^* \in \mathrm{Geo}_{\mathbb{N}}(p)$ *and* $\mathbb{P}(X^* = 1) = \frac{\alpha}{\alpha+(\alpha-1)\delta}$.

Remark 7. We note the following alternative justification for the convergence to a stationary distribution. The two row sums of the Möbius matrix for Ψ,

$$M_\Psi = \begin{pmatrix} \alpha\delta & \alpha \\ (\alpha-1)\delta & \alpha+\delta \end{pmatrix},$$

It is an easy excercise to show by induction that for $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$, with $p, q \in [0,1]$ and p, q not both zero one has that

$$P^n = \underbrace{\begin{pmatrix} \frac{q}{q+p} & \frac{p}{p+q} \\ \frac{q}{q+p} & \frac{p}{p+q} \end{pmatrix}}_{=:P^\infty} + (1-p-q)^n \begin{pmatrix} \frac{p}{q+p} & -\frac{p}{p+q} \\ -\frac{q}{q+p} & \frac{q}{p+q} \end{pmatrix}$$

and thus, unless $p = q = 1$ (the Markov chain is irreducible with period 2) or $p = q = 0$ (the Markov chain is reducible with two recurrent states), P^n converges geometrically fast to P^∞. This includes the case where either p or q (but not both) is zero (the reducible case where one state is recurrent and the other transient). Thus, for these cases $\Psi^n(s) \to 1$. When $p = q = 1$, P^n does not converge and $\Psi^n(s)$ alternates between s and $1/s$. Finally, when $p = q = 0$, $P^n = I$ for all $n \geq 0$, so that $\Psi^n(s) = s$ for all $n \geq 0$.

◇

Next, let us consider the h-step-ahead properties of this stationary solution to (49). From (52), we know that the conditional h-step-ahead apgf equals

$$
\begin{aligned}
1-\mathcal{G}_{X_{n+h}|X_n}(s) &= \Big(1-\big(1-\Psi^{\circ h}(s)\big)^{X_n}\Big)\times\prod_{i=0}^{h-1}\frac{(1+\delta)\Psi^{\circ i}(s)}{1+\delta\Psi^{\circ i}(s)}\\
&= \left(1-\left(\frac{\big(\alpha+(\alpha-1)\delta\big)(1-s)}{(\alpha-1)\delta(1-s)+\big(\alpha+(\alpha-1)\delta s\big)\big(\frac{\delta}{\alpha(1+\delta)}\big)^{-h}}\right)^{X_n}\right)\\
&\quad\times\frac{s\big(\alpha+(\alpha-1)\delta\big)}{\alpha+(\alpha-1)\delta s-\alpha(1-s)\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}}.
\end{aligned}
$$

From this expression, we derive the conditional mean $\mathbb{E}\left(X_{n+h}|X_n\right)=-\mathcal{G}'_{X_{n+h}|X_n}(0)$ as

$$
\mathbb{E}\left(X_{n+h}|X_n\right)=\frac{\big(\alpha+(\alpha-1)\delta\big)}{\alpha\Big(1-\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}\Big)}\left(1-\left(\frac{\big(\alpha+(\alpha-1)\delta\big)\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}}{\alpha+(\alpha-1)\delta\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}}\right)^{X_n}\right). \tag{58}
$$

Hence, the product mean $\mathbb{E}\left(X_{n+h}X_n\right)=\mathbb{E}\left(X_n\,\mathbb{E}\left(X_{n+h}|X_n\right)\right)$ takes the form

$$
\mathbb{E}\left(X_{n+h}X_n\right)=\frac{\big(\alpha+(\alpha-1)\delta\big)}{\alpha\Big(1-\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}\Big)}\left(\mathbb{E}\left(X_n\right)-A\,\mathbb{E}\left(X_n\,A^{X_n-1}\right)\right),
$$

where

$$
A:=\frac{\big(\alpha+(\alpha-1)\delta\big)\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}}{\alpha+(\alpha-1)\delta\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}}.
$$

Now, we can make use of the stationary marginal distribution in (57) with mean $\mathbb{E}\left(X_n\right)=1+\frac{(\alpha-1)\delta}{\alpha}$. For the pgf of some r.v. X, $\mathcal{P}_X(s)=\mathbb{E}(s^X)=\mathcal{G}_X(1-s)$, it holds that $\mathcal{P}'_X(s)=\mathbb{E}(X\,s^{X-1})$. Therefore, $\mathbb{E}(X_n\,A^{X_n-1})$ can be computed by taking the first derivative of $\mathcal{G}_{X^*}(1-s)$ according to (57), which leads to

$$
\mathbb{E}\left(X_n\,A^{X_n-1}\right)=\frac{\big(\alpha+(\alpha-1)\delta\big(\frac{\delta}{\alpha(1+\delta)}\big)^{h}\big)^2}{\alpha\big(\alpha+(\alpha-1)\delta\big)}
$$

and thus

$$
\mathbb{E}(X_{n+h}X_n)=\frac{\alpha+(\alpha-1)\delta}{\alpha^2}\left(\alpha+(\alpha-1)\delta+(\alpha-1)\delta\left(\frac{\delta}{\alpha(1+\delta)}\right)^{h}\right).
$$

Hence, lag-h autocovariance and autocorrelation follow as

$$
\operatorname{Cov}\left(X_{n+h},X_n\right)=\frac{(\alpha-1)\delta}{\alpha}\Big(1+\frac{(\alpha-1)\delta}{\alpha}\Big)\Big(\frac{\delta}{\alpha(1+\delta)}\Big)^{h}, \tag{59}
$$

$$
\varrho\left(X_{n+h},X_n\right)=\Big(\frac{\delta}{\alpha(1+\delta)}\Big)^{h}. \tag{60}
$$

Note that the autocorrelation function (acf) $\varrho(h)=\varrho\left(X_{n+h},X_n\right)$ is exponentially decreasing like for an ordinary AR(1) process, although the process (49) is highly nonlinear. The acf can only take non-negative values, where the largest values are achieved for $\alpha\to 1$. This boundary case corresponds to the 'random-walk like' recursion $X_{n+1} = X_n \wedge U_n$ (see also Example 6 below).

6 Connection to linear birth-and-death processes

A continuous-time non-homogeneous linear birth-and-death process, see Tavaré (2018) for a recent survey, is a time-homogeneous Markov process $(Y_t)_{t\in[0,\infty)}$ with values in $\mathbb{N}_0$ and transition probabilities depending on the rate functions $r_1(t) \geq 0$ and $r_2(t) \geq 0$ as follows:

$$\begin{aligned}
\mathbb{P}\,(Y_t = n+1 | Y_0 = n) &= r_1(t)n + o(t), \\
\mathbb{P}\,(Y_t = n-1 | Y_0 = n) &= r_2(t)n + o(t), \\
\mathbb{P}\,(Y_t = n | Y_0 = n) &= 1 - (r_1(t) + r_2(t))n + o(t),
\end{aligned}$$

as $t \to 0$, i.e. there are a.s. only jumps of absolute size one, and the jump intensities are proportional to the current state. The apgf of Y_t is given by (see Kendall (1948))

$$\mathcal{G}_{Y_t|Y_0=n}(s) = \left(1 - \frac{\alpha(t)s}{1+\beta(t)s}\right)^n, \tag{61}$$

where

$$\alpha(t) = e^{-\rho(t)}, \quad \beta(t) = e^{-\rho(t)} \int_0^t e^{\rho(s)} r_1(s)\, ds, \quad \rho(t) = \int_0^t (r_2(s) - r_1(s))\, ds.$$

Here, $\rho(t)$ is known as the Malthusian parameter. Note that Kendall (1948) uses a slightly different parametrization: $\xi_t = 1 - \frac{\alpha(t)}{1+\beta(t)}$ and $\eta_t = \frac{\beta(t)}{1+\beta(t)}$.

(61) implies that Y_t is a stochastic process with

$$Y_t \quad \overset{d}{\sim} \quad (\alpha(t), \beta(t)) \odot Y_0. \tag{62}$$

From Lemma 1, we easily obtain the well-known formulas

$$\begin{aligned}
\mathbb{E}\,(Y_t) &= \alpha(t) \cdot \mathbb{E}\,(Y_0)\,, \\
\operatorname{Var}(Y_t) &= \mathbb{E}\,(Y_0) \left(\alpha(t)\,(1+2\beta(t)) - \alpha(t)^2\right) + \alpha(t)^2 \operatorname{Var}(Y_0)\,.
\end{aligned}$$

Example 6. Relation (62) shows that the considered linear birth-and-death process is related to the INAR-type process (37) with $\varepsilon_n \equiv 0$ from Section 4, i.e. when

$$X_n = \theta \odot X_{n-1}. \tag{63}$$

It follows from (33) that the distribution of X_n is given by

$$X_n \overset{d}{\sim} \begin{cases} (\alpha^n, \beta\frac{\alpha^n-1}{\alpha-1}) \odot X_0 & ;\alpha \neq 1, \\ (1, n\beta) \odot X_0 & ;\alpha = 1. \end{cases} \tag{64}$$

The homogeneous case of our birth and death process, $r_1(t) \equiv r_1$, $r_2(t) \equiv r_2$, constitutes a kind of time-continuous version of that process. This also becomes clear when comparing (64) to the resulting expressions for $\alpha(t), \beta(t)$:

$$\alpha(t) = \begin{cases} 1 & r_1 = r_2, \\ e^{(r_1-r_2)t} & r_1 \neq r_2, \end{cases} \qquad \beta(t) = \begin{cases} r_1 t & r_1 = r_2, \\ \frac{r_1}{r_2-r_1}\left(1 - e^{(r_1-r_2)t}\right) & r_1 \neq r_2. \end{cases} \qquad \square$$

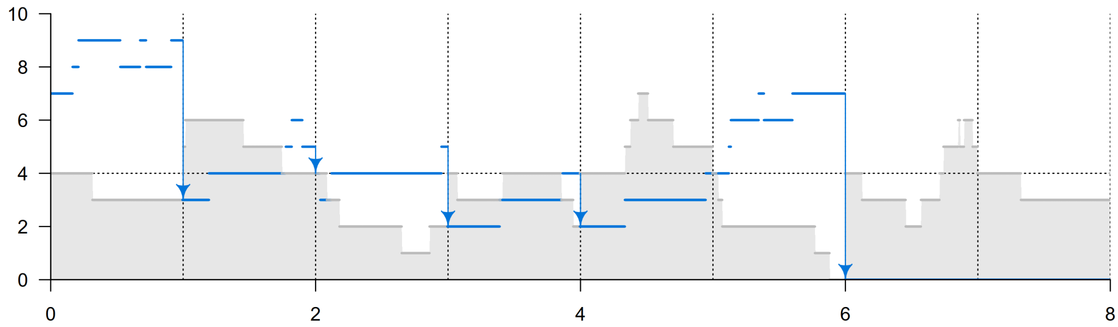

Figure 1: Simulated sample path of process (65), where U_t is shown in gray, and Y_t in blue.

Example 7. In the case where $r_1(t) = r_2(t)$, we obtain $\rho(t) = 0$ and thus $\alpha(t) = 1$, and

$$Y_t \overset{d}{\sim} \left(1, \int_0^t r_1(s)\, ds\right) \odot Y_0.$$

This corresponds to the parabolic case. □

Example 8. If Y_t is a pure death process, then $r_1(t) = 0$ and hence $\beta(t) = 0$. So Y_t is 'continuously thinned', $Y_t \overset{d}{\sim} \alpha(t) \circ Y_0$, recall Example 1. □

Example 9. If Y_t is a pure birth process ('non-homogeneous Yule process', see Belzunce et al. (2001)), then $r_2(t) = 0$ and

$$\beta(t) = e^{-\int_0^t r_1(s)\, ds} \int_0^t e^{\int_0^s r_1(u)\, du} r_1(s)\, ds = e^{-\int_0^t r_1(s)\, ds} \left(e^{\int_0^t r_1(u)\, du} - 1\right) = 1 - \alpha(t).$$

Hence, the pure birth case corresponds to the case $\alpha(t) = 1 + \beta(t)$ from Example 3. □

Finally, there is also a relation to Section 5. For $0 \leq t < 1$, let Y_t be a (left-continuous) birth-and-death process with parameter $\theta(t) = (\alpha(t), \beta(t))$ as defined above, and let U_t be a second independent (left-continuous) birth-and-death process with parameter $\eta(t) = (\gamma(t), \delta(t))$. Then, let $Y_1 = Y_{1-} \wedge U_{1-}$. For $t \in [1, 2)$, let Y_t and U_t continue as before. Continue this construction by letting $Y_k = Y_{k-} \wedge U_{k-}$ at the 'intervention times' $k \in \mathbb{N}$. Then,

$$Y_{k+1} = (\theta(1) \odot Y_k) \wedge U_k, \tag{65}$$

so the embedded discrete time process $(Y_k)_{k \in \mathbb{N}_0}$ is a minification process according to Section 5.2, and we can apply results from Section 5 to it.

An illustrative example is shown in Figure 1, where the process U_t is plotted in gray (with $U_0 = 4$), and the process Y_t in red. At time $t = 6$, the process Y_t jumps down to the value of $U_6 = 0$ and then stays in the absorbing state zero.

7 MA-type models using modulating operator

7.1 An INMA(1)-type process

The modulating operator can also be used for constructing INMA-type models, where stationarity does not require for additional parameter constraints. If defining a first-order INMA model by

$$X_n = \theta \odot \varepsilon_{n-1} + \varepsilon_n, \tag{66}$$

then the classical INMA(1) model of Al-Osh & Alzaid (1988) using binomial thinning is covered as the special case $\theta = (\alpha, 0)$. For the general setup in (66), the stationary marginal distribution of (X_n) is computed by using the independence of the innovations as well as (29):

$$\mathcal{G}_X(s) = \mathcal{G}_\varepsilon(s)\, \mathcal{G}_{\theta\odot\varepsilon}(s) = \mathcal{G}_\varepsilon(s)\, \mathcal{G}_\varepsilon\left(f_\theta(s)\right). \tag{67}$$

As an example, if the innovations are i.i.d. T-geometric as well, say $\varepsilon_n \sim \text{BerG}\,(\gamma, \delta)$, then

$$\mathcal{G}_X(s) = \left(1 - \frac{\gamma s}{1+\delta s}\right)\left(1 - \frac{\alpha\gamma s}{1+(\alpha\delta+\beta)s}\right),$$

also recall (31). Regarding the serial dependence structure, it is obvious that X_n and X_{n-h} are independent if $h \geq 2$. Hence, it suffices to study the joint distribution of (X_n, X_{n-1}), the joint apgf of which can be split as follows (using the independence of the innovations):

$$\begin{aligned}\mathcal{G}_{X_n,X_{n-1}}(s,t) &= \mathbb{E}\left((1-s)^{\theta\odot\varepsilon_{n-1}+\varepsilon_n}(1-t)^{\theta\odot\varepsilon_{n-2}+\varepsilon_{n-1}}\right)\\ &= \mathcal{G}_\varepsilon(s)\, \mathcal{G}_{\theta\odot\varepsilon}(t)\, \mathbb{E}\left((1-s)^{\theta\odot\varepsilon}(1-t)^{\varepsilon}\right).\end{aligned}$$

So compared to (67), the last term requires further attention. Note that $\mathbb{E}\left((1-s)^{\theta\odot\varepsilon}|\varepsilon\right) = \mathcal{G}_Z(s)^\varepsilon = (1-f_\theta(s))^\varepsilon$. So by the law of total expectation, conditioning on ε, we get

$$\mathbb{E}\left((1-s)^{\theta\odot\varepsilon}(1-t)^{\varepsilon}\right) = \mathbb{E}\left((1-t)^\varepsilon(1-f_\theta(s))^\varepsilon\right) = \mathcal{G}_\varepsilon\left(t + (1-t)f_\theta(s)\right),$$

and we end up with the joint apgf

$$\mathcal{G}_{X_n,X_{n-1}}(s,t) = \mathcal{G}_\varepsilon(s)\, \mathcal{G}_{\theta\odot\varepsilon}(t)\, \mathcal{G}_\varepsilon\left(t + (1-t)f_\theta(s)\right).$$

As there is only little literature on INMA-type models, we recommend a more detailed study of model (66) and its special cases for future research, see Section 8 for details.

7.2 A variant of the minification process

In analogy to the process in Section 7.1, one could also consider an MA(1)-type version of the minification process, defined via

$$X_n = (\theta \odot U_{n-1}) \wedge U_n. \tag{68}$$

An obvious advantage compared to the AR(1)-like model (49) is given by the fact that zeros do not constitute absorbing states anymore. The stationary marginal distribution of (X_n) can be immediately concluded from Corollary 11. As $\theta \odot U_{n-1} \sim \text{BerG}\,(\gamma\alpha, \beta + \delta\alpha)$ according to (31), and also independent of U_n, we have

$$X_n \sim \mathrm{BerG}\left(\frac{\alpha\gamma^2}{1+\beta+\delta+\alpha\delta}, \frac{\delta(\beta+\alpha\delta)}{1+\beta+\delta+\alpha\delta}\right). \tag{69}$$

Regarding serial properties, again the joint distribution of (X_n, X_{n-1}) is relevant, expressed in terms of its apgf, which can be computed via conditioning on U_{n-1}:

$$\mathscr{G}_{X_n,X_{n-1}}(s,t) = \mathbb{E}\Big[\mathbb{E}\left((1-s)^{(\theta\odot U_{n-1})\wedge U_n}|U_{n-1}\right)\,\mathbb{E}\left((1-t)^{(\theta\odot U_{n-2})\wedge U_{n-1}}|U_{n-1}\right)\Big].$$

Here, the first factor $(\theta\odot U_{n-1})\wedge U_n|U_{n-1}$ is the minimum of the conditionally independent r.v. $U_n \sim \mathrm{BerG}(\gamma,\delta)$ and $\theta\odot U_{n-1}|U_{n-1}$ with $\mathscr{G}_{\theta\odot U_{n-1}|U_{n-1}}(s) = (1-\frac{\alpha s}{1+\beta s})^{U_{n-1}}$, so its apgf can be computed via Corollary 11. Similarly, the second factor $(\theta\odot U_{n-2})\wedge U_{n-1}|U_{n-1}$ is the minimum of the r.v. $\theta\odot U_{n-2}\sim\mathrm{BerG}(\gamma\alpha,\beta+\delta\alpha)$ and the conditionally constant $U_{n-1}|U_{n-1}$ with $\mathscr{G}_{U_{n-1}|U_{n-1}}(t) = (1-t)^{U_{n-1}}$, so its apgf can again be computed via Corollary 11.

8 Conclusions

In this article, we considered a modulating operator based on a T-geometric counting series, covering various well-known special cases. We showed that the modulating operator can be related to the Möbius transform such that the successive application of modulating operators corresponds to the multiplication of respective Möbius matrices. We derived novel closed-form results on the n-fold application of modulating operators and Möbius transforms, respectively, as well as corresponding limiting results. Our modulating operator was first used to construct geometric INAR-type processes, where we were able to derive novel results on h-step-ahead and limiting distributions. Also a corresponding INMA-version was briefly investigated. Secondly, we proposed a minification process, which is a kind of INAR process based on a minimum operation rather than a sum. This process is generally non-stationary with absorbing state zero, but has a stationary solution if counting series and gate thresholds are (zero-truncated) geometric on $\mathbb{N}$. Again, we derived closed-form expressions for various stochastic properties and briefly considered an INMA-version of the process. Finally, we focused on the linear birth-and-death process (also with additional interventions), which is a continuous-time process that can again be expressed by using our modulating operator. We showed several connections of this process to the INAR-type processes discussed before.

There are various directions for future research. As INMA-type models have been widely neglected in the literature, we recommend to further investigate both the classical INMA(1) model $X_n = \theta\odot\varepsilon_{n-1}+\varepsilon_n$ from Section 7.1 as well as $X_n = (\theta\odot U_{n-1})\wedge U_n$ from Section 7.2 in more detail. In particular, closed-form moment expressions (marginal, conditional, and serial) appear to be relevant for a deeper understanding of these processes, which, in turn, can be derived from the apgfs presented above. But also higher-order autoregressive processes in analogy to Du & Li (1991), such as $X_n = (\sum_{i=0}^{p}\theta_i\odot X_{n-i})\wedge U_n$, or a max-version of the minification process, i.e. a generalization of the max-INAR(1) process considered by Scotto et al. (2018), deserve further attention in future research.

Acknowledgements

The authors thank the two referees for their useful comments on an earlier draft of this article.

References

Al-Osh, M. & Aly, E.-E. A. (1992), 'First order autoregressive time series with negative binomial and geometric marginals', *Communications in Statistics-Theory and Methods* **21**(9), 2483–2492.

Al-Osh, M. & Alzaid, A. (1988), 'Integer-valued moving average (INMA) process', *Statistical Papers* **29**(1), 281–300.

Aleksić, M. S. & Ristić, M. M. (2021), 'A geometric minification integer-valued autoregressive model', *Applied Mathematical Modelling* **90**, 265–280.

Aly, E.-E. A. A. & Bouzar, N. (2005), 'Stationary solutions for integer-valued autoregressive processes', *Int. J. Math. Math. Sci.* **2005**(1), 1–18.

Aly, E.-E. A. & Bouzar, N. (2019), 'Expectation thinning operators based on linear fractional probability generating functions', *Journal of the Indian Society for Probability and Statistics* **20**(1), 89–107.

Andrews, D. K. & Balakrishna, N. (2023), 'A novel geometric AR(1) model and its estimation', *Journal of Statistical Computation and Simulation* **93**(16), 2906–2935.

Athreya, K. B. & Ney, P. E. (2012), *Branching processes*, Vol. 196, Springer Science & Business Media.

Bailey, N. T. (1964), *The elements of stochastic processes with applications to the natural sciences*, John Wiley & Sons.

Bakouch, H. S., Mohammadpour, M. & Shirozhan, M. (2018), 'A zero-inflated geometric INAR(1) process with random coefficient', *Applications of Mathematics* **63**, 79–105.

Barreto-Souza, W. (2015), 'Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros', *Journal of Time Series Analysis* **36**(6), 839–852.

Barreto-Souza, W., Ndreca, S., Silva, R. B. & Silva, R. W. (2023), 'Non-linear INAR(1) processes under an alternative geometric thinning operator', *TEST* **32**(2), 695–725.

Behrends, E. (2022), *Tilings of the Plane*, Springer.

Belzunce, F., Lillo, R. E., Ruiz, J.-M. & Shaked, M. (2001), 'Stochastic comparisons of nonhomogeneous processes', *Probability in the Engineering and Informational Sciences* **15**(2), 199–224.

Bourguignon, M. & Weiß, C. H. (2017), 'An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion', *TEST* **26**(4), 847–868.

Du, J.-G. & Li, Y. (1991), 'The integer-valued autoregressive (INAR(p)) model', *Journal of Time Series Analysis* **12**(2), 129–142.

Feller, W. (1968), *An introduction to probability theory and its applications. Vol. I*, Vol. 2, John Wiley & Sons Inc, New York.

Fristedt, B. E. & Gray, L. F. (1997), *A modern approach to probability theory*, Springer Science & Business Media.

Gnedenko, B. V. & Korolev, V. Y. (2020), *Random summation: Limit theorems and applications*, CRC press.

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005), *Univariate discrete distributions*, Vol. 444, John Wiley & Sons.

Kalamkar, V. (1995), 'Minification processes with discrete marginals', *Journal of Applied Probability* **32**(3), 692–706.

Kella, O. & Löpker, A. (2023), 'On binomial thinning and mixing', *Indagationes Mathematicae* **34**(5), 1121–1145.

Kendall, D. G. (1948), 'On the generalized 'birth-and-death' process', *The Annals of Mathematical Statistics* **19**(1), 1–15.

Kevei, P. (2011), 'Asymptotics of nearly critical Galton–Watson processes with immigration', *Acta Sci. Math.* **77**, 681–702.

Latour, A. (1998), 'Existence and stochastic structure of a non-negative integer-valued autoregressive process', *J. Time Ser. Anal.* **19**(4), 439–455.

Lewis, P. A. & McKenzie, E. (1991), 'Minification processes and their transformations', *Journal of Applied Probability* **28**(1), 45–57.

Littlejohn, R.-P. (1992), 'Discrete minification processes and reversibility', *Journal of Applied Probability* **29**(1), 82–91.

McKenzie, E. (1986), 'Autoregressive moving-average processes with negative-binomial and geometric marginal distributions', *Adv. in Appl. Probab.* **18**(3), 679–705.

Nastić, A. S. & Ristić, M. M. (2012), 'Some geometric mixed integer-valued autoregressive (INAR) models', *Statist. Probab. Lett.* **82**(4), 805–811.

Qian, L. & Zhu, F. (2022), 'A new minification integer-valued autoregressive process driven by explanatory variables', *Australian & New Zealand Journal of Statistics* **64**(4), 478–494.

Ristić, M. M., Bakouch, H. S. & Nastić, A. S. (2009), 'A new geometric first-order integer-valued autoregressive (NGINAR(1)) process', *J. Statist. Plann. Inference* **139**(7), 2218–2226.

Schiff, J. L. (2022), *Topics in Complex Analysis*, Vol. 88, de Gruyter.

Scotto, M. G., Weiß, C. H. & Gouveia, S. (2015), 'Thinning-based models in the analysis of integer-valued time series: a review', *Statistical Modelling* **15**(6), 590–618.

Scotto, M., Weiß, C., Möller, T. & Gouveia, S. (2018), 'The max-INAR(1) model for count processes', *TEST* **27**(4), 850–870.

Steutel, F. & van Harn, K. (1979), 'Discrete analogues of self-decomposability and stability.', *Ann. Probab.* **7**, 893–899.

Stojanović, V. S., Bakouch, H. S., Bojičić, R., Alomair, G. & Alghamdi, S. A. (2024), 'Poisson–Lindley minification INAR process with application to financial data', *AIMS Mathematics* **9**(8), 22627–22654.

Tavaré, S. (2018), 'The linear birth–death process: an inferential retrospective', *Advances in Applied Probability* **50**(A), 253–269.

Tavares, L. V. (1980), 'An exponential Markovian stationary process', *Journal of Applied Probability* **17**(4), 1117–1120.

Weiß, C. H. (2008), 'Thinning operations for modeling time series of counts – a survey', *AStA Advances in Statistical Analysis* **92**(3), 319–341.

Weiß, C. H. & Zhu, F. (2024), 'Conditional-mean multiplicative operator models for count time series', *Computational Statistics and Data Analysis* **191**, 107885.

Williams, K. S. (1992), 'The nth power of a 2×2 matrix', *Mathematics Magazine* **65**(5), 336–336.

Zhu, R. & Joe, H. (2010), 'Negative binomial time series models based on expectation thinning operators', *J. Statist. Plann. Inference* **140**(7), 1874–1888.