



HAL
open science

On quasi Pólya thinning operator

Jean Peyhardi

► **To cite this version:**

| Jean Peyhardi. On quasi Pólya thinning operator. 2024. hal-04691901

HAL Id: hal-04691901

<https://hal.science/hal-04691901v1>

Preprint submitted on 9 Sep 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

On quasi Pólya thinning operator

Jean Peyhardi^{1,a},

¹IMAG, University of Montpellier, CNRS, Montpellier, France, ^ajean.peyhardi@umontpellier.fr

Abstract. Thinning operation is a stochastic operation that shrinks a random count variable into another one. This kind of random operation has been intensively studied during the seventies to characterize some count distributions, such as the Poisson distribution using the binomial thinning operator (also named binomial damage model). Then, the closure under thinning operator has been studied in order to define some classes of integer valued autoregressive (INAR) models for count time series. These two properties will be studied in this paper for the new class of quasi Pólya thinning operators. Classical results concerning the binomial thinning operator are recovered as a special case. The quasi Pólya thinning operator is related to the new class of quasi Pólya splitting distributions, defined for multivariate count data. The probabilistic graphical model (PGM) of these multivariate distributions is characterized. Finally a general class of integer valued autoregressive models is introduced, including the usual cases of Poisson marginal or generalized Poisson as a special cases and the generalized negative binomial as a new case.

1 Introduction

A splitting distribution of dimension $J = 2$ can be summarized by a random variable N that is split into two random variables Y_1 and Y_2 , i.e., such that $N = Y_1 + Y_2$. The conditional distribution of Y_1 given $N = n$ is called thinning operator, e.g., the binomial distribution. The variable Y_1 can be viewed as damaged version of the original variable N . Two properties of such a model have been intensively studied: the closure under thinning operation and the characterization of the original distribution of N through the condition of independence between Y_1 and Y_2 .

The first property consists for a given thinning operator, to find a distribution for N such that Y_1 follows the same distribution (with modified parameter). It was initiated by Rao (1965) which showed that binomial, Poisson and negative binomial distribution are closed under the binomial thinning operation. It was generalized in some way by Janardan and Rao (1986) which demonstrated the closure of generalized Pólya Eggenberger distributions (including the Poisson distribution as a special case) under quasi Pólya thinning operation (including the binomial thinning operator as a special case). Then Joe (1996) related the closure under thinning operation to the closure under convolution. They used this property to define some integer valued autoregressive models for count time series. Indeed the thinning operator ensures that the degradation of X_{t-1} is still an integer and thus allow to correctly build X_t ; see (Scotto, Weiß and Gouveia, 2015; Davis et al., 2021) for a review. In this framework, the binomial thinning operator is often referred to (Steutel and van Harn, 1979) that used this operation to define the discrete analogue of self decomposability. More recently Puig and Valero (2007) characterized the distributions that are closed under the binomial thinning operation. Then Peyhardi, Fernique and Durand (2021) highlighted some distributions that are closed respectively under the hypergeometric, binomial or beta binomial thinning operation.

The second property consists in the characterization of the distribution for N such that Y_1 and Y_2 are independent, for a given thinning operator. It was initiated by Moran (1952) that

Keywords and phrases. Quasi Pólya distribution, Binomial thinning operator, Probabilistic graphical model.

characterized the Poisson distribution for N through the independence between Y_1 and Y_2 , given a binomial thinning operator. It was generalized by Bol'shev (1965) which characterized the Poisson (resp. binomial and negative binomial) distribution through independence between Y_1 and Y_2 given a binomial (resp. hypergeometric and negative hypergeometric) thinning operator. Finally Rao and Janardan (1984) characterized the generalized negative binomial distribution through the quasi hypergeometric thinning operator. Then, for the binomial thinning operator, Rao and Rubin (1964) proposed to replace the independence assumption by a weaker condition, named the Rao-Rubin condition. Other authors followed this way for the beta binomial thinning operator; see (Patil and Ratnaparkhi, 1975) for a review. But it was shown by Shanbhag and Panaretos (1979) that the Rao-Rubin condition is not sufficient to obtain the characterization for an hypergeometric thinning operator, because of its constrained support. More recently Peyhardi and Fernique (2017) characterized the probabilistic graphical model of the multivariate version of splitting distribution (i.e., the random vector (Y_1, \dots, Y_J) such that $Y_1 + \dots + Y_J = N$) for the multinomial and Dirichlet multinomial splitting distributions. It should be noted that the thinning operator was used, during the seventies, as a tool for characterizing some univariate distributions whereas the present approach focuses on the resulting multivariate distribution for count data, as in recent papers of Castañer et al. (2015); Jones and Marchand (2019); Peyhardi, Fernique and Durand (2021).

The paper is organized as follows. Section 2 introduces preliminary notations and definitions used all along the paper. Three classes of count distributions are introduced: convolution, (additive) modified power series and inverse convolution distributions. They are all characterized by a parametric sequence $\{a_\theta(n)\}_{n \in \mathbb{N}, \theta \in \Theta}$ with $\Theta = \mathbb{R}_+^*$ or $\Theta = \mathbb{N}$. We highlight all the assumptions concerning this parametric sequence that are necessary to obtain the different theorems of the next section. Section 3 introduces the class of convolution thinning operators. It is shown that the three classes of distributions, previously introduced, are closed under the convolution thinning operation (Theorem 1). Then it is shown that the modified power series is the only one distribution for N such that Y_1, \dots, Y_J are independent, more precisely the minimal PGM of the multivariate distribution is characterized (Theorem 2). Then some applications of these two properties are presented, in the frameworks of multivariate count data and count time series. Finally the continuous analogue of such distributions is presented, verifying that both theorems hold.

2 Preliminary notations and definitions

All definitions and properties of the present paper will be based on the existence of a non-negative parametric sequence $a_\theta = \{a_\theta(n)\}_{n \in \mathbb{N}, \theta \in \Theta}$ with $\Theta = \mathbb{R}_+^*$ or $\Theta = \mathbb{N}$. All along the paper, the four following assumptions will be made on a_θ .

(A0) The support:

$$\text{Supp } a_\theta = \{0, 1, \dots, K_\theta\}$$

for some $K_\theta \in \mathbb{N} \cup \{\infty\}$.

(A1) Additive closure under convolution:

$$a_\theta * a_\gamma = a_{\theta+\gamma}$$

for all $(\theta, \gamma) \in \Theta^2$

(A2) Power series convergence:

There exists a function $g(\alpha)$ such that

$$\sum_{k=0}^{K_\theta} a_\theta(k) \{g(\alpha)\}^k = h_\theta(\alpha)$$

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

(A3) Inverse series convergence:

$$\sum_{k=0}^{K_\theta} \frac{r}{k+r} \frac{a_\theta(k)}{a_{\theta+\gamma}(k+r)} = \frac{1}{a_\gamma(r)}$$

for all $(\theta, \gamma) \in \Theta^2$ and $r \in \{0, \dots, K_\gamma\}$.

In the following we will see that assumptions (A1), (A2) and (A3) are necessary conditions to demonstrate the different theorems whereas assumption (A0) could certainly be relaxed in some way. However the six examples of such parametric sequence a_θ , presented in the following, share the four assumptions. Note that the support of a_θ is bounded for only one example ($K_\theta = \theta$), whereas it is equal to \mathbb{N} for the five others ($K_\theta = \infty$). Let now introduce three families of univariate count distributions based on a_θ .

Definition 1. A discrete non-negative variable X is said to follow a convolution distribution if there exists a parametric sequence a_θ such that (A1) holds and

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

for some $(\theta, \gamma) \in \Theta^2$ and $n \in \mathbb{N}$. This distribution is denoted by $\mathcal{C}_n(\theta, \gamma)$.

Assumption (A1) ensures that we obtain a proper distribution. This class of distributions includes for instance the hypergeometric distribution if $a_\theta(n) = \binom{\theta}{n}$, or the binomial distribution if $a_\theta(n) = \theta^n/n!$. More generally it include the family of quasi Pólya distributions that was introduced by [Janardan \(1973\)](#) as urn model with predetermined strategies; see Section 1 of the Supplementary Materials for detailed examples ([Peyhardi, 2023](#)). This family of distributions will be used in next section to introduce the general class of convolution thinning operators.

Definition 2. A discrete non-negative variable X is said to follow an additive modified power series distribution if there exist a parametric sequence a_θ such that (A1), (A2) hold and a function g such that

$$P(X = k) = \frac{a_\theta(k) \{g(\alpha)\}^k}{h_\theta(\alpha)}$$

for some $\theta \in \Theta$ and $\alpha \in (0, R)$. This distribution is denoted by $\mathcal{AP}\mathcal{S}(\theta, g(\alpha))$.

Assumption (A2) ensures that we obtain a proper distribution. This is a subclass of the modified power series distributions introduced by [Gupta \(1974\)](#) since the assumption (A1) is added. It includes for instance the binomial distribution if $a_\theta(n) = \binom{\theta}{n}$, or the Poisson distribution if $a_\theta(n) = \theta^n/n!$, but not the logarithmic series distribution since $a(n) = 1/n$ do not share the convolution identity (A1). More generally it includes the family of generalized Pólya Eggenberger distributions studied by [Janardan and Rao \(1982\)](#); see Section 2 of the Supplementary Materials for detailed examples ([Peyhardi, 2023](#)).

Proposition 1. *An additive modified power series distribution is additively closed under convolution, i.e., we have*

$$\mathcal{AP}\mathcal{S}(\theta, g(\alpha)) * \mathcal{AP}\mathcal{S}(\gamma, g(\alpha)) = \mathcal{AP}\mathcal{S}(\theta + \gamma, g(\alpha)),$$

for all $(\theta, \gamma) \in \Theta^2$ and $\alpha \in (0, R)$.

All proofs of the present paper are given in Appendices. The notion of closure under convolution, introduced by [Teicher \(1954\)](#), is close to those of infinite divisibility, presented by [Steutel et al. \(1979\)](#), but is different. For instance the binomial distribution is additively closed under convolution ($\mathcal{B}_\theta(p) * \mathcal{B}_\gamma(p) = \mathcal{B}_{\theta+\gamma}(p)$) but not infinite divisible since parameters are integers ($\Theta = \mathbb{N}$).

Definition 3. A discrete non-negative variable X is said to follow an inverse convolution distribution if there exists a parametric sequence a_θ such that (A1), (A3) hold and

$$P(X = k) = \frac{r}{k+r} \frac{a_\theta(k)a_\gamma(r)}{a_{\theta+\gamma}(k+r)}$$

for some $(\theta, \gamma) \in \Theta^2$, $r \in \{0, \dots, K_\gamma\}$. This distribution is denoted by $\mathcal{IC}(r; \theta, \gamma)$.

Assumption (A3) ensures that we obtain a proper distribution. This class of distributions includes for instance the negative hypergeometric distribution if $a_\theta(n) = \binom{\theta}{n}$, or the negative binomial distribution if $a_\theta(n) = \theta^n/n!$. More generally it include the family of quasi inverse Pólya distributions introduced by [Janardan \(1975\)](#); see Section 3 of the Supplementary Materials for detailed examples ([Peyhardi, 2023](#)).

Proposition 2. *The three classes of distributions (definitions 1, 2, 3) are characterized by the parametric sequence a_θ .*

The reader can easily check that the specific sequence $a_\theta(n) = \theta^n/n!$ leads to the binomial, Poisson and negative binomial distributions, following respectively the definitions 1, 2 and 3. More generally, there exists six examples of parametric sequence a_θ that are encompassed by the general form

$$a_\theta^{[c,d]}(n) = \frac{\theta(\theta + dn)^{(n;c)}}{(\theta + dn)n!} \quad (1)$$

where $n \in \mathbb{N}$, $\theta \in \Theta$ (with $\Theta = \mathbb{N}$ when $c < 0$, $d = 0$ and $\Theta = \mathbb{R}_+^*$ otherwise), $c \in \mathbb{R}$, $d \in \mathbb{R}_+$ and $m^{(n;c)} := \prod_{k=0}^{n-1} (m + ck)$. One can distinguish six equivalence classes according to (c, d) with representative elements:

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

In fact only five representative elements are defined since $a_\theta^{[-1,d+1]}(n) = a_\theta^{[1,d]}(n)$ for all $d \geq -1$; see Section 4 of the Supplementary Material for details about equivalences between some distributions, related to this equality ([Peyhardi, 2023](#)). It could be shown that the parametric sequence $a_\theta^{[c,d]}$ respects the four assumptions (A0), (A1), (A2) and (A3), in the six

	Definition 1	Definition 2	Definition 3
$c = -1, d \geq 0$	quasi hypergeometric	generalized binomial	quasi beta binomial
$c = 0, d \geq 0$	quasi binomial	generalized Poisson	quasi negative binomial
$c = 1, d \geq -1$	quasi beta binomial	generalized negative binomial	quasi beta negative binomial

Table 1 The three distributions obtained by definitions 1, 2, 3 (columns) according to the different values of (c, d) (rows) for the parametric sequence $a_\theta^{[c,d]}(n)$ described in equation (1). The terms "quasi" and "generalized" disappears when $d = 0$.

cases. For instance details are given for the case $a_\theta(n) = \binom{n+\theta-1}{n}$ in Section 5 of the Supplementary Materials (Peyhardi, 2023). Therefore the convolution distribution (definition 1), the additive modified power series distribution (definition 2) and the inverse convolution distribution (definition 3) are well defined; see Sections 1, 2, 3 of Supplementary Material for detailed examples (Peyhardi, 2023). A summary of these distributions is given in Table 1.

Note that for the continuous case, only one example emerges from the parametric function

$$a_\theta(x) = \frac{x^{\theta-1}}{\Gamma(\theta)},$$

defined for $x \in \mathbb{R}_+^*$ and $\theta \in \mathbb{R}_+^*$. It is shown in Section 5 that this parametric function also respects the four assumptions if integers are replaced by non-negative real values and sums by integrals, with the exception that zero does not belong to the support of a_θ .

3 Thinning operator and splitting operator

A thinning operation is a probabilistic operation that shrinks a random count variable into another one. If the original count represents the size of a population, then the thinning operator randomly selects some individuals of this population. If focus is made on the number of selected individuals then we deal with the thinned distribution. If focus is made on both the number of selected and unselected individuals then we deal with the (bivariate) splitting distribution. Note that the splitting distribution can be defined in the multivariate case using a multivariate thinning operation. The first subsection studies the closure under thinning operation, i.e., is looking for original distributions that are preserved after the thinning operation. The second subsection studies the characterization of independence between the selected and unselected parts according to the original distribution.

3.1 Closure under convolution thinning operation

The most usual example of thinning operation is the binomial one. The binomial thinning operation is usually defined as follows. Let X denote a random count variable with distribution \mathcal{L} and $p \in (0, 1)$ a given parameter, then $X' \stackrel{d}{=} p \circ X$ denotes the thinned version of X . It is defined as the random sum $X' = \sum_{i=1}^X Z_i$ where $Z_i \sim \mathcal{B}(p)$ are independent random Bernoulli variables with parameter p . The binomial thinning operator has firstly been used by Sprott (1965) in ecological context, to model the number of adults X' produced by an insect, assuming that X is the number of eggs and p the probability for each egg to hatch (assuming independence between eggs). It is named binomial destructive process or binomial damage model in this framework. It has also been used in the framework of count time series to define INAR model; see (Weiß, 2008) for a review.

Remarking that the thinned distribution of X' is a binomial distribution mixed by the original distribution of X , we propose to use the notation

$$X' \sim \mathcal{B}_n(p) \wedge \mathcal{L},$$

where \wedge denotes the mixing operator. This notation is more appropriate to deal with properties of distributions. It is natural to replace the binomial distribution by a more general family that share necessary properties to define a thinning operation. We therefore introduce the convolution thinned distribution as the mixed distribution

$$\mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{L},$$

that includes the binomial thinned distribution as a special case ($a_\theta(n) = \theta^n/n!$). If $X' \sim \mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{L}$ where $(p_n)_{n \in \mathbb{N}}$ denotes the pmf of the original distribution \mathcal{L} then the pmf of the convolution thinned distribution is given by

$$P(X' = k) = \sum_{n \geq k} \frac{a_\theta(k) a_\gamma(n-k)}{a_{\theta+\gamma}(n)} p_n.$$

The most general approach of such a thinning operation was proposed by [Janardan and Rao \(1986\)](#) that studied the quasi Pólya thinned distribution $q\mathcal{P}_n^{[c,d]}(\theta, \gamma) \wedge_n \mathcal{L}$. They showed, in their Theorem 5.1, that generalized Pólya Eggenberger distributions are closed under the quasi Pólya thinning operation. This result is recovered by the second point of the following theorem.

Theorem 1. *The convolution, the modified power series and the inverse convolution distributions are closed under the convolution thinning operation when their parameter respects the additive constraint. More precisely we have the following distributions equalities:*

1. *Convolution*
 $\mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{C}_m(\theta + \gamma, \lambda) = \mathcal{C}_m(\theta, \gamma + \lambda),$
2. *Additive modified power series*
 $\mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{AP}\mathcal{S}(\theta + \gamma, g(\alpha)) = \mathcal{AP}\mathcal{S}(\theta, g(\alpha)),$
3. *Inverse convolution*
 $\mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{IC}(r; \theta + \gamma, \lambda) = \mathcal{IC}(r; \theta, \lambda),$

for all $(\theta, \gamma, \lambda) \in \Theta^3$, $\alpha \in (0, R)$ and $r \in (0, \infty)$.

As corollary of Theorem 1, based on specific case $a_\theta(n) = \theta^n/n!$, the result of [Rao \(1965\)](#) is recovered.

Corollary 1 ([\(Rao, 1965\)](#)). *The binomial, Poisson and negative binomial distribution are closed under the binomial thinning operation.*

1. *binomial*
 $\mathcal{B}_n(\pi) \wedge_n \mathcal{B}_m(p) = \mathcal{B}_m(p')$ where $p' = \pi p \in (0, 1),$
2. *Poisson*
 $\mathcal{B}_n(\pi) \wedge_n \mathcal{P}(\lambda) = \mathcal{P}(\lambda')$ where $\lambda' = \pi \lambda \in (0, \infty),$
3. *negative binomial*
 $\mathcal{B}_n(\pi) \wedge_n \mathcal{NB}(r, p) = \mathcal{NB}(r, p')$ where $p' := \frac{\pi p}{\pi p + 1 - p} \in (0, 1).$

Remark that the additive constraint on parameters does not appear in this example since the sum $\theta + \gamma$ is not identifiable and only the proportion $p = \theta/(\theta + \gamma)$ is identifiable. The same kind of corollary holds for the five other examples characterized by $a_\theta^{[c,d]}$ defined in equation (1). For instance, based on $a_\theta(n) = \binom{n+\theta-1}{n}$, we obtain the following result.

Corollary 2. *The beta-binomial, the negative binomial and the beta negative binomial distributions are closed under the beta binomial thinning operation.*

1. *beta binomial*

$$\beta\mathcal{B}_n(\theta, \gamma) \underset{n}{\wedge} \beta\mathcal{B}_m(\theta + \gamma, \lambda) = \beta\mathcal{B}_m(\theta, \gamma + \lambda),$$

2. *negative binomial*

$$\beta\mathcal{B}_n(\theta, \gamma) \underset{n}{\wedge} \mathcal{NB}(\theta + \gamma, p) = \mathcal{NB}(\theta, p),$$

3. *beta negative binomial*

$$\beta\mathcal{B}_n(\theta, \gamma) \underset{n}{\wedge} \beta\mathcal{NB}(r; \theta + \gamma, \lambda) = \beta\mathcal{NB}(r; \theta, \lambda),$$

for all $(\theta, \gamma, \lambda) \in \mathbb{R}_+^{*3}$, $p \in (0, 1)$ and $r \in (0, \infty)$.

Remark. Puig and Valero (2007) showed that a distribution \mathcal{L} is closed under binomial thinning operation if and only if it satisfies one of the following assertions

- $\mathcal{L} = \mathcal{P}(\lambda) \underset{\lambda}{\wedge} \mathcal{L}'$ where \mathcal{L}' denotes a distribution supported on $(0, \infty)$,
- $\mathcal{L} = \mathcal{B}_n(p) \underset{n}{\wedge} \mathcal{L}^*$ where \mathcal{L}^* is not closed under the binomial thinning operation.

A conjecture would be to generalize this result to any convolution thinning operation and state that a distribution \mathcal{L} is closed under any convolution thinning operation if and only if it satisfies one of the following assertions

- $\mathcal{L} = \mathcal{APS}(\theta + \gamma, g(\alpha)) \underset{\alpha}{\wedge} \mathcal{L}'$ where \mathcal{L}' denotes a distribution supported on $(0, R)$,
- $\mathcal{L} = \mathcal{C}_n(\theta + \gamma, \lambda) \underset{n}{\wedge} \mathcal{L}^*$ where \mathcal{L}^* is not closed under the convolution thinning operation.

This conjecture is consistent with Theorem 1, remarking that the inverse quasi Pólya distribution turns out to be a mixed modified power series distribution (for instance $\mathcal{NB}(r, p) = \mathcal{P}(\lambda) \underset{\lambda}{\wedge} \Gamma(r, \frac{p}{1-p})$).

3.2 Characterization of independence for convolution splitting distributions

In this subsection, the original observation n is not thinned but split into $J \geq 2$ variables Y_1, \dots, Y_J , i.e., such that $Y_1 + \dots + Y_J = n$. In the following the random vector (Y_1, \dots, Y_J) will be denoted by \mathbf{Y} and its sum by $|\mathbf{Y}|$. The convolution distribution (definition 1) has to be defined in a multivariate way, as introduced by Peyhardi and Fernique (2017).

Definition 4. The random vector \mathbf{Y} is said to follow the convolution distribution denoted by $\mathcal{C}_{\Delta_n}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J) \in \Theta^J$ if its conditional pmf, given the sum $|\mathbf{Y}| = n$, has the form

$$P_{|\mathbf{Y}|=n}(\mathbf{Y} = \mathbf{y}) = \frac{\prod_{j=1}^J a_{\theta_j}(y_j)}{a_{|\boldsymbol{\theta}|}(n)}$$

for all $n \in \mathbb{N}$ and $\mathbf{y} \in \Delta_n = \{\mathbf{y} \in \mathbb{N}^J : |\mathbf{y}| = n\}$ (Δ_n denotes the discrete simplex).

All known examples of such distributions are multivariate quasi Pólya distributions. They are obtained from the parametric sequence $a_{\boldsymbol{\theta}}^{[c,d]}(n)$ of equation (1) and their characteristics and notations are summarized in Section 6 of the Supplementary Materials (Peyhardi, 2023). Such a distribution is not a *sensu stricto* multivariate distribution since the dimension of its support is $J - 1$ instead of J ; see (Peyhardi, Fernique and Durand, 2021) for details. This problem disappears if the parameter n is assumed to be a random variable. In this case we obtain a mixed distribution denoted by

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$$

which is called convolution splitting distribution. The pmf of such a distribution is given by

$$P(\mathbf{Y} = \mathbf{y}) = \frac{p(|\mathbf{y}|)}{a_{|\boldsymbol{\theta}|}(|\mathbf{y}|)} \prod_{j=1}^J a_{\theta_j}(y_j), \quad (2)$$

for all $\mathbf{y} \in \text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$ and zero otherwise, where $p(n)$ denotes the pmf of the distribution \mathcal{L} . Note that thinning and splitting operations are closely related but not similar since the first reduces the original quantity whereas the second conserves it in the different components. It will be shown that the choice of the original distribution \mathcal{L} implies specific properties about the resulting mixed distribution.

Remark. Using the specific parametric sequence $a_{\theta}(n) = \binom{n+\theta-1}{n}$ we obtain the Dirichlet multinomial splitting distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{L}$. The specific Dirichlet multinomial splitting negative binomial distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{NB}(r, p)$ has recently been studied in depth by [Jones and Marchand \(2019\)](#).

First recall the definition of the PGM for a given multivariate distribution. According to [Lauritzen \(1996\)](#), the global Markov properties enable to derive the independence assertions that hold in a multivariate distribution by simply examining a graph (undirected, directed or mixed graph). A PGM is defined by a distribution and a graph such that all independence assertions, derived from the graph using the global Markov properties, hold in the distribution. A PGM is said to be minimal if any edge removal in the graph induces an independence assertion that is not held in the distribution. Before to characterize the minimal PGM of a convolution splitting distribution $\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{L}$, it is easier to show that a specific choice for the sum distribution \mathcal{L} implies independence between Y_1, \dots, Y_J . The demonstrations of following lemma and theorem are given in Appendices [D](#) and [E](#).

Lemma 1. *Let $\mathbf{Y} \sim \mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{APS}(|\boldsymbol{\theta}|, g(\alpha))$ for some $\alpha \in (0, R)$, assuming (A1), (A2) for the associated parametric sequence a_{θ} . Then all variables Y_j are independent and $Y_j \sim \mathcal{APS}(\theta_j, g(\alpha))$. This property can be summarized by the formula*

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{APS}(|\boldsymbol{\theta}|, \alpha) = \bigotimes_{j=1}^J \mathcal{APS}(\theta_j, g(\alpha)).$$

where the notation $\mathbf{Y} \sim \bigotimes_{j=1}^J \mathcal{L}_j$ indicates that $Y_j \sim \mathcal{L}_j$ for all $j = 1, \dots, J$ with mutual independence between Y_1, \dots, Y_J .

The independence between all variables corresponds to the empty PGM. Now we are interesting in the uniqueness of the sum distribution \mathcal{L} such that the PGM is empty. The following theorem characterizes the PGM of a convolution splitting distribution according to the sum distribution.

Theorem 2. *Let $\mathbf{Y} \sim \mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{L}$, assuming (A0), (A1), (A2) for the associated parametric sequence a_{θ} . Then the minimal PGM for \mathbf{Y} is*

- empty if $\mathcal{L} = \mathcal{APS}(|\boldsymbol{\theta}|, g(\alpha))$ for some $\alpha \in (0, R)$,
- complete otherwise.

This theorem generalizes the result of [Bol'shev \(1965\)](#) and those of [Rao and Janardan \(1984\)](#) that concerns the bivariate case ($J = 2$) only for some thinning operators. It also includes results obtained by [Peyhardi and Fernique \(2017\)](#) that characterized the PGM only for two multivariate operators: multinomial and Dirichlet multinomial. Indeed this theorem holds

for any convolution thinning operator assuming (A0), (A1), (A2) and (A3), i.e., for the five (multivariate) quasi Pólya thinning operators (hypergeometric, multinomial, Dirichlet multinomial, quasi-hypergeometric and quasi-multinomial) and the demonstration is unique; see Appendix E.

Remark. The fact that zero belongs to the support of a_θ is included in assumption (A0) and used in the proof of Theorem 2. The connectivity of the support, also included in assumption (A0), may be relaxed in some way. However, all examples of parametric sequence a_θ given in equation (1) have a connected support. For the continuous case it is different since zero does not belong to the support. An additional assumption on a_θ is needed to demonstrate Theorem 2; see Appendix E for details.

Otherwise, for extension of Theorem 2 using the weaker Rao-Rubin condition instead of independence, it is necessary to modify assumption (A0) into (A0') $\text{Supp } a_\theta = \mathbb{N}$ (or $(0, \infty)$ for the continuous case). Such an extension holds for all the quasi-Pólya splitting models except the hypergeometric splitting distribution with $a_\theta(n) = \binom{\theta}{n}$ such that $\text{Supp } a_\theta = \llbracket 0, \theta \rrbracket$ is bounded. Let us note that the Rao-Rubin condition can be viewed as a context specific conditional independence, defined by Koller and Friedman (2009), allowing the determination of local PGM but not necessary to determine the global properties of PGM.

4 Applications of Theorems 1 and 2

4.1 Applications of Theorems 1

Since the eighties, the closure property under thinning operation is studied to demonstrate the stationary of INAR models for count time series; see (Scotto, Weiß and Gouveia, 2015) for a review. This property is here used to studied the closure under marginalization of some multivariate distributions. Using Theorem 1 it is easy to define the natural multivariate extension of the three classes of distributions (definitions 1, 2, 3), using the convolution splitting as an operator from the space of univariate distributions to the space of multivariate distributions.

Corollary 3. *The following distributions are closed under marginalization.*

1. *Multivariate convolution (non-singular)*

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \mathcal{C}_m(|\boldsymbol{\theta}|, \gamma) = \mathcal{C}_{\blacktriangle_m}(\boldsymbol{\theta}, \gamma)$$
2. *Multivariate additive modified power series*

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha)) = \bigotimes_{j=1}^J \mathcal{AP}\mathcal{S}(\theta_j, g(\alpha))$$
3. *Multivariate inverse convolution*

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \mathcal{IC}(r; |\boldsymbol{\theta}|, \gamma) = \mathcal{MIC}(r; \boldsymbol{\theta}, \gamma)$$

Note that the three previous distributions can be viewed as natural multivariate extension of their univariate sum distribution. Let us give some details about these multivariate extensions:

1. The vector \mathbf{Y} is said to follow a non-singular convolution distribution, denoted by $\mathcal{C}_{\blacktriangle_m}(\boldsymbol{\theta}, \gamma)$, if the completed vector $(\mathbf{Y}, m - |\mathbf{Y}|)$ follows the singular version $\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}, \gamma)$. This non-singular version is supported on \blacktriangle_m of dimension J instead of Δ_n of dimension $J - 1$. As explained by Peyhardi, Fernique and Durand (2021), the natural multivariate extension of the univariate convolution distribution $\mathcal{C}_m(\theta, \gamma)$ turns out to be the non-singular version. Moreover we have the following distributions equality

$$\mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \mathcal{C}_m(|\boldsymbol{\theta}|, \gamma) = \mathcal{C}_{\blacktriangle_m}(\boldsymbol{\theta}, \gamma).$$

For instance, the multinomial distribution $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ is not the multivariate extension of the binomial distribution $\mathcal{B}_n(\pi)$, as usually presented. The natural multivariate extension is the non-singular version

$$\mathcal{M}_{\blacktriangleleft_n}(p \cdot \boldsymbol{\pi}) = \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \wedge_n \mathcal{B}_m(p).$$

The non-singular version is detailed for the five other quasi Pólya distributions are summarized in Section 7 of the Supplementary Materials (Peyhardi, 2023).

2. The second point of Corollary 3 allows us to define multivariate extension of the additive modified power series $\mathcal{AP}\mathcal{S}(\theta, g(\alpha))$ but with independence between components (see Lemma 1). In this sense they are not viewed as *sensu stricto* multivariate distributions but as J univariate distributions. For instance, using $a_\theta(n) = \binom{n+\theta-1}{n}$ we obtain

$$\mathcal{DM}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \mathcal{NB}(|\boldsymbol{\theta}|, p) = \bigotimes_{j=1}^J \mathcal{NB}(\theta_j, p).$$

In this case all marginals are independent and follow negative binomial distributions. The five other cases of independence are summarized in Section 7 of the Supplementary Materials (Peyhardi, 2023).

3. The third point allows us to define a *sensu stricto* multivariate extension of the inverse convolution distribution since the PGM is complete according to Theorem 2. For instance the negative multinomial distribution is obtained as

$$\mathcal{NM}(r, p \cdot \boldsymbol{\pi}) := \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \wedge_n \mathcal{NB}(r, p),$$

or the multivariate generalized waring distribution, introduced by Xekalaki (1986), can also be obtained as

$$\mathcal{MGWD}(r, \boldsymbol{\theta}, \gamma) := \mathcal{DM}_{\Delta_n}(\boldsymbol{\theta}) \wedge_n \beta \mathcal{NB}(r, |\boldsymbol{\theta}|, \gamma).$$

In the same way it is thus possible to define two new multivariate extensions. The multivariate extension of the quasi negative binomial distribution is given by

$$q\mathcal{NM}^{[d]}(r, p \cdot \boldsymbol{\pi}) := q\mathcal{M}_{\Delta_n}^{[d]}(\boldsymbol{\pi}) \wedge_n q\mathcal{NB}^{[d]}(r, p),$$

and is named the quasi negative multinomial distribution. The multivariate extension of the quasi beta negative binomial distribution is given by

$$q\mathcal{MGWD}^{[d]}(r, \boldsymbol{\theta}, \gamma) := q\mathcal{DM}_{\Delta_n}^{[d]}(\boldsymbol{\theta}) \wedge_n q\beta \mathcal{NB}^{[d]}(r, |\boldsymbol{\theta}|, \gamma),$$

and is named the quasi multivariate generalized waring distribution. Other examples are given in Section 7 of the Supplementary Materials (Peyhardi, 2023).

Moreover the parameters of these distributions are easily estimated since the log-likelihood can be decomposed in two parts: the log-likelihood of the convolution splitting distribution and the log-likelihood of the sum distribution (Peyhardi, Fernique and Durand, 2021). Otherwise, in the case of Pólya splitting distributions ($d = 0$ and $c \in \{-1, 0, 1\}$), Corollary 3 leads to nine remarkable multivariate distributions; see Section 7 of the Supplementary Materials for details (Peyhardi, 2023). Peyhardi, Laroche and Mortier (2022) showed that these distributions are stationary solutions of multivariate birth-death processes under extended neutral theory with simple parametric birth-death rates. These new models contribute to the class of joint species distribution models (JSDMs) that are among the most important statistical tools in community ecology; see (Ovaskainen and Abrego, 2020) for a review.

4.2 Application of Theorem 2

Theorem 2 characterizes the independence between variables through the original distribution of the sum. It is thus possible to write a test of independence as a test of distribution adequacy since

$$H_0 : N_1, \dots, N_J \text{ are independent} \Leftrightarrow H_0 : \mathcal{L} = \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha)) \text{ for some } \alpha \in (0, R).$$

Moreover if the sum distribution is assumed to belong to the family of additive modified power series distribution, i.e., $\mathcal{L} = \mathcal{AP}\mathcal{S}(\gamma, g(\alpha))$ for some $\gamma \in \Theta$ and $\alpha \in (0, R)$ then H_0 becomes an equality of parameters

$$H_0 : \gamma = |\boldsymbol{\theta}|.$$

The null distribution is thus nested into the alternative distribution and the likelihood ratio test can be applied.

This result holds for generalized negative binomial marginal distributions. Assume for instance that the vector (Y_1, Y_2, Y_3) follows a quasi Dirichlet multinomial splitting generalized negative binomial $q\mathcal{DM}_{\Delta_n}^{[d]}(\theta_1, \theta_2, \theta_3) \wedge_n \mathcal{GNB}^{[d]}(r, p)$. Then Y_1, Y_2 and Y_3 are mutually independent if and only if $r = \theta_1 + \theta_2 + \theta_3$. The alternative distribution is easier to estimate since parameters $(\theta_1, \theta_2, \theta_3)$ of the convolution distribution and (r, p) of the sum distribution are not related. They can thus be estimated by separately maximizing the likelihood; see (Peyhardi, Fernique and Durand, 2021) for details. For the null distribution these parameters are constrained and thus likelihood could be maximized iteratively. Finally the likelihood ratio test asymptotically follows, under the null hypothesis, a chi square distribution with one degree of freedom. Note that this result is true for $d > -1$ and in particular for $d = 0$, i.e., we obtain the case of independent negative binomial distributed variables, under the null hypothesis.

4.3 Application of Theorems 1 and 2

4.3.1 *Application in multivariate count data modelling* As consequence of Theorems 1 and 2 we can build multivariate count distribution with non-canonical PGM. Indeed Theorem 2 states that the PGM of a convolution splitting distribution is necessary empty or complete, the two most simple structures (see Figure 1 (A) and (B)). To obtain more interesting structure it would be possible to use convolution splitting models with a recursive structure. For instance with $J = 3$ components, assume that $(y_1 + y_2, y_3) \sim \mathcal{M}_{\Delta_n}(\pi, 1 - \pi) \wedge_n \mathcal{NB}(r, p)$ and $(y_1, y_2) | y_1 + y_2 = m \sim \mathcal{DM}_{\Delta_m}(\theta_1, \theta_2)$. The multivariate distribution of the vector (y_1, y_2, y_3) is then fully specified. It could be shown, according to Theorems 1 and 2, that the pmf can be factorized as follows

$$p(y_1, y_2, y_3) = p(y_3 | y_1 + y_2) p(y_1) p(y_2),$$

if $r = \theta_1 + \theta_2$ and thus we obtain a V-structure as in Figure 1 (C); see (Lauritzen, 1996) for details about PGM deduction from the pmf factorization. To understand this result, first note that $y_1 + y_2 \sim \mathcal{B}_n(\pi) \wedge_n \mathcal{NB}(r, p)$, i.e., $y_1 + y_2 \sim \mathcal{NB}(r, p')$ with $p' = \pi p / (\pi p + 1 - p)$ according to the third point of Theorem 1 using parametric sequence $a_\theta(n) = \theta^n / n!$. Then remark that $(y_1, y_2) \sim \mathcal{DM}_{\Delta_m}(\theta_1, \theta_2) \wedge_m \mathcal{NB}(r, p')$, i.e., $(y_1, y_2) \sim \bigotimes_{j=1}^2 \mathcal{NB}(\theta_j, p')$ if and only if $r = \theta_1 + \theta_2$ according to Theorem 2 with the parametric sequence $a_\theta(n) = \binom{n+\theta-1}{n}$. To conclude this paragraph, closure under convolution thinning operation (Theorem 1) and characterization of independence (Theorem 2) can be used as building blocks to construct complex PGM for discrete data. Note that study of PGM for count variables stays an open issue. Moreover, recall that such a model for multivariate count data could easily been extended to the regression framework using the decomposition of the log-likelihood; see (Peyhardi, Fernique and Durand, 2021) for details.

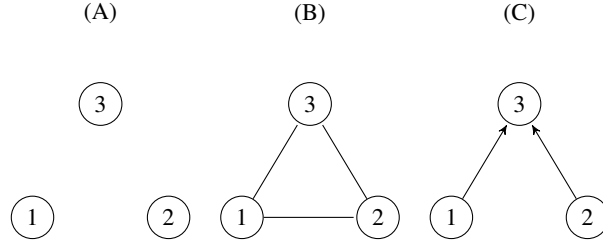


Figure 1 Minimal PGM of (A) a convolution splitting distribution such that $\mathcal{L} = \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha))$ (B) a convolution splitting distribution such that $\mathcal{L} \neq \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha))$ (C) a specific hierarchical convolution splitting distribution.

4.3.2 *Application in count time series* Let $(X_t)_{t \in \mathbb{N}}$ denote a discrete time process with discrete state space. A quasi Pólya autoregressive INAR(1) model is defined as follows:

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

where $\varepsilon_t \sim \mathcal{L}_\varepsilon$ are independent and identically distributed random variables and $\rho \circ$ denotes a quasi Pólya thinning operation independent of $(\varepsilon_t)_{t \in \mathbb{N}}$. Denoting by \mathcal{L}_t the distribution of X_t we have

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

As consequence of Theorems 1 and 2 we obtain the stationary distribution of such a process.

Theorem 3. *Let $(X_t)_{t \in \mathbb{N}}$ follow a quasi Pólya INAR(1) model with residual distribution $\mathcal{L}_\varepsilon = \mathcal{AP}\mathcal{S}^{[c,d]}(\gamma, g(\alpha))$ for some $\alpha \in (0, R)$. Then the process $(X_t)_{t \in \mathbb{N}}$ is an ergodic time reversible Markov chain with stationary distribution $\mathcal{AP}\mathcal{S}^{[c,d]}(\theta + \gamma, g(\alpha))$.*

This class of INAR(1) models includes several known examples. We have seen that there is five distinct quasi-Pólya thinning operators and thus five classes of INAR(1) processes. They can be distinguished by their marginal distribution:

1. The binomial INAR(1) model introduced by [Al-Osh and Alzaid \(1991\)](#) corresponds to the case $c = -1$ and $d = 0$.
2. The Poisson INAR(1) model introduced by [McKenzie \(1985\)](#) corresponds to the case $c = 0$ and $d = 0$.
3. The negative binomial INAR(1) model presented by [Joe \(1996\)](#) corresponds to the case $c = 1$ and $d = 0$.
4. The generalized Poisson INAR(1) model introduced by [Alzaid and Al-Osh \(1993\)](#) corresponds to the case $c = 0$ and $d > 0$.
5. The generalized negative binomial INAR(1) model, corresponding to the case $c = 1$ and $d > -1$, is here introduced.

As member of the family of dispersion models ([Jorgensen, 1997](#)), the unity variance functions of the additive modified power series have closed form; see Table 2. As remarked by [Jourdan and Kokonendji \(2002\)](#), these unity variance functions are strictly ordered as follows

$$V_B < V_P < V_{NB} < V_{GP} < V_{GNB}.$$

According to the level of dispersion observed in the data, the autoregressive model with the appropriate marginal distribution has to be used.

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

Table 2 Unity variance function of the five additive modified power series distributions

5 The continuous analogue

Convolution distributions can also be defined for non negative real values $\mathbf{y} \in \mathbb{R}_+^J$. The discrete simplex Δ_n becomes the continuous simplex $\Delta_x = \{\mathbf{y} \in \mathbb{R}_+^J : |\mathbf{y}| = x\}$ defined for $x \in \mathbb{R}_+$ (by convention $\Delta = \Delta_1$). A distribution supported on Δ_x , is said to be a continuous convolution distribution if its probability density function (pdf) has the form:

$$f(\mathbf{y}) = \frac{\prod_{j=1}^J a_{\theta_j}(y_j)}{a_{|\theta|}(x)},$$

where a_θ respects the continuous analogue of assumption (A1), (A2) and (A3). The assumption (A0) becomes $\text{Supp } a_\theta = (0, K_\theta)$ for some $K_\theta \in \mathbb{R}_+ \cup \{\infty\}$ and note that $0 \notin \text{Supp } a_\theta$. To our knowledge, only one such a parametric function exists:

$$a_\theta(y) = \frac{y^{\theta-1}}{\Gamma(\theta)},$$

for which the four assumptions hold:

(A0) $\text{Supp } a_\theta = (0, \infty)$

(A1) convolution identity

$$\begin{aligned} (a_\theta * a_\gamma)(x) &= \int_0^x a_\theta(y)a_\gamma(x-y)dy, \\ &= \frac{1}{\Gamma(\theta)\Gamma(\gamma)} \int_0^x y^{\theta-1}(x-y)^{\gamma-1}dy, \\ &= \frac{x^{\theta+\gamma-2}}{\Gamma(\theta)\Gamma(\gamma)} \int_0^x \left(\frac{y}{x}\right)^{\theta-1} \left(\frac{x-y}{x}\right)^{\gamma-1} dy, \\ &= \frac{x^{\theta+\gamma-1}}{\Gamma(\theta)\Gamma(\gamma)} \underbrace{\int_0^1 t^{\theta-1}(1-t)^{\gamma-1}dt}_{=\frac{\Gamma(\theta)\Gamma(\gamma)}{\Gamma(\theta+\gamma)}}, \\ &= \frac{x^{\theta+\gamma-1}}{\Gamma(\theta+\gamma)}, \\ (a_\theta * a_\gamma)(x) &= a_{\theta+\gamma}(x). \end{aligned}$$

with integration by substitution $t = y/x$.

(A2) Laplace transform (Euler integral of the second kind)

$$\begin{aligned} \int_0^\infty a_\theta(t)\alpha^t dt &= \frac{1}{\Gamma(\theta)} \int_0^\infty t^{\theta-1}e^{-bt} dt \\ &= \frac{1}{\Gamma(\theta)} \int_0^\infty \left(\frac{t}{b}\right)^{\theta-1} e^{-t} \frac{1}{b} dt \end{aligned}$$

$$= \frac{1}{b^\theta} \frac{1}{\Gamma(\theta)} \underbrace{\int_0^\infty t^{\theta-1} e^{-t} dt}_{=\Gamma(\theta)}$$

$$\int_0^\infty a_\theta(t) \alpha^t = \frac{1}{b^\theta}$$

where $b = -\ln \alpha$.

(A3) Euler integral of the first kind

$$\begin{aligned} \int_0^\infty \frac{r}{t+r} \frac{a_\theta(t)}{a_{\theta+\gamma}(t+r)} dt &= \int_0^\infty \frac{r}{t+r} \frac{t^{\theta-1}}{(t+r)^{\theta+\gamma-1}} \frac{\Gamma(\theta+\gamma)}{\Gamma(\theta)} dt \\ &= r^{1-\gamma} \frac{\Gamma(\theta+\gamma)}{\Gamma(\theta)} \underbrace{\int_0^\infty x^{\theta-1} (1+x)^{\theta+\gamma-1} dx}_{=B(\theta,\gamma)} \\ &= \frac{\Gamma(\gamma)}{r^{\gamma-1}} \\ \int_0^\infty \frac{r}{t+r} \frac{a_\theta(t)}{a_{\theta+\gamma}(t+r)} dt &= \frac{1}{a_\gamma(r)} \end{aligned}$$

with integration by substitution $x = t/r$.

Following the continuous analogue of definitions 1, 2 and 3, this parametric function allows us to define three distributions.

1. The convolution distribution $\mathcal{C}_x(\theta, \gamma)$ is the beta distribution $\beta_x(\theta, \gamma)$ (dilated by the parameter x) with pdf

$$f(y) = \frac{1}{B(\theta, \gamma)} \frac{y^{\theta-1} (x-y)^{\gamma-1}}{x^{\theta+\gamma-1}},$$

with $(x, \theta, \gamma) \in \mathbb{R}_+^{*3}$.

2. The additive power series (or member of exponential family) distribution $\mathcal{APS}(\theta, g(\alpha))$ is the gamma distribution $\Gamma(\theta, b)$ with $b = -\ln \alpha$ and pdf

$$f(y) = \frac{b^\theta}{\Gamma(\theta)} y^{\theta-1} e^{-by}$$

with $(\theta, b) \in \mathbb{R}_+^{*2}$.

3. The inverse convolution distribution $\mathcal{IC}(r; \theta, \gamma)$ is the inverse beta (or beta prime) distribution $\mathcal{IB}(\theta, \gamma)$ with pdf

$$f(y) = \frac{1}{B(\theta, \gamma)} \frac{y^{\theta-1} r^\gamma}{(y+r)^{\theta+\gamma}},$$

with $(r, \theta, \gamma) \in \mathbb{R}_+^{*3}$.

As corollary of Theorem 1 we obtain the closure property under the beta thinning operation.

Corollary 4. *The beta, gamma and inverse beta distributions are closed under the beta thinning operation, more precisely we have*

1. $\beta_x(\theta, \gamma) \wedge_x \beta_z(\theta + \gamma, \lambda) = \beta_z(\theta, \gamma + \lambda)$,
2. $\beta_x(\theta, \gamma) \wedge_x \Gamma(\theta + \gamma, b) = \Gamma(\theta, b)$,

$$3. \beta_x(\theta, \gamma) \wedge_x \mathcal{I}\beta(r; \theta + \gamma, \lambda) = \mathcal{I}\beta(r; \theta, \lambda).$$

Note that points 1 and 3 of Corollary 4 have respectively been obtained by Rao (1949) and Jambunathan (1954). As consequence of this corollary, we obtain the three natural multivariate extensions:

1. The non-singular Dirichlet distribution

$$\mathcal{D}_{\Delta_x}(\boldsymbol{\theta}) \wedge_x \beta_z(|\boldsymbol{\theta}|, \gamma) = \mathcal{D}_{\mathbf{\Delta}_z}(\boldsymbol{\theta}, \gamma)$$

which is supported on $\mathbf{\Delta}_z := \{\mathbf{y} \in (0, 1)^J : |\mathbf{y}| < z\}$, the interior of the simplex $\Delta_z = \{\mathbf{y} \in (0, 1)^J : |\mathbf{y}| = z\}$ (that has dimension J instead of $J - 1$), where $\boldsymbol{\theta} \in \mathbb{R}_+^{*J}$ and $\gamma \in \mathbb{R}_+^*$.

2. The multivariate (independent) gamma distribution

$$\mathcal{D}_{\Delta_x}(\boldsymbol{\theta}) \wedge_x \Gamma(|\boldsymbol{\theta}|, b) = \bigotimes_{j=1}^J \Gamma(\theta_j, b)$$

with $\boldsymbol{\theta} \in \mathbb{R}_+^{*J}$ and $b \in \mathbb{R}_+^*$. This distribution is supported on \mathbb{R}_+^{*J} .

3. The inverse Dirichlet distribution

$$\mathcal{D}_{\Delta_x}(\boldsymbol{\theta}) \wedge_x \mathcal{I}\beta(|\boldsymbol{\theta}|, b) = \mathcal{ID}(\boldsymbol{\theta}, b)$$

with $\boldsymbol{\theta} \in \mathbb{R}_+^{*J}$ and $b \in \mathbb{R}_+^*$. This distribution is supported on \mathbb{R}_+^{*J} .

Note that Δ_x is a dilatation since $\Delta_x = x \cdot \Delta = \{x \cdot \boldsymbol{\pi} : \boldsymbol{\pi} \in \Delta\}$. The Dirichlet thinning operation $\mathcal{D}_{\Delta_x}(\boldsymbol{\theta}) \wedge_x \mathcal{L}$ can thus be viewed as an external product of independent variables $X \cdot \boldsymbol{\pi}$ where $X \sim \mathcal{L}$ and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J) \sim \mathcal{D}_\Delta$. This property is used in demonstration of Theorem 2 by adding the following assumption.

(A*) Dilatation:

$$\frac{a_\theta(x)a_\gamma(t-x)}{a_{\theta+\gamma}(t)} = \frac{a_\theta(x/t)a_\gamma(1-x/t)}{ta_{\theta+\gamma}(1)}$$

for all $t \in \mathbb{R}_+^*$ and $0 < x < t$.

This assumption is equivalent to state that the convolution distribution $\mathcal{C}_x(\boldsymbol{\theta}, \gamma)$ is a dilatation of $\mathcal{C}_1(\boldsymbol{\theta}, \gamma)$, i.e., $\mathcal{C}_x(\boldsymbol{\theta}, \gamma) = x \cdot \mathcal{C}_1(\boldsymbol{\theta}, \gamma)$. In this case the continuous analogue of Theorem 2 turns out to be equivalent to the Lukacs's proportion sum independence theorem (Lukacs, 1955); see the proof in Appendix E for details. Finally, note that the dilatation property implies that the parameter r of the inverse beta $\mathcal{I}\beta(r; \boldsymbol{\theta}, \gamma)$ becomes non-identifiable and is replaced by the representative element $r = 1$ resulting in the inverted beta distribution $\mathcal{I}\beta(\boldsymbol{\theta}, \gamma)$; see Section 3 of the Supplementary Materials for details on this distribution (Peyhardi, 2023).

Remark. Let us note that assumption (A*) is may be not necessary to demonstrate Theorem 2 but, to our knowledge, the only one parametric function a_θ verifying the continuous analogue of assumptions (A1), (A2), (A3) is $a_\theta(x) = x^{\theta-1}/\Gamma(\theta)$ and also verifies the additional assumption (A*).

Discussion

The present paper introduced the class of quasi Pólya thinning operator that encompasses the hypergeometric, binomial, beta binomial, quasi binomial and quasi beta binomial thinning

operators. Additionally the two classes of (additive) modified power series and inverse quasi Pólya distributions are introduced. It is shown that these three classes of distributions are characterized by a unique parametric sequence. It allows us to highlight four assumptions about this sequence that are sufficient to obtain the two main theorems. The first Theorem states that the three classes of distributions are closed under the quasi Pólya thinning operation. The second theorem states that the (additive) modified power series is the unique distribution such that the PGM of the resulting quasi Pólya splitting distribution is empty.

Several applications of these two theorems have been discussed. Multivariate extensions of several count distributions have been easily obtained, using the quasi Pólya splitting operator as an operator from the space of univariate distribution to the space of multivariate distributions. They constitute natural multivariate extensions since they are closed under marginalization. Otherwise, the characterization of independence could be used to propose some tests of independence. Finally using both theorems it is possible to build PGM with more complex structure than the basic empty or complete graph.

In a modeling approach of multivariate count data, the family of quasi Pólya splitting distributions offers several perspectives. They can be easily extended to the regression framework by assuming a univariate regression model for the sum distribution (original distribution) and another one for the quasi Pólya distribution. Since the pmf is a product of the two models (see equation (2)), the log-likelihood is split into two parts that can be separately maximized; see (Peyhardi, Fernique and Durand, 2021) for details. The family of quasi Pólya splitting distributions allows the generalization of modeling approach recently proposed by Wang and Zhao (2017) or Tang and Chen (2019) that used Dirichlet multinomial regression to study microbiome data. Finally the quasi Pólya thinning operator could naturally be used to define new INAR(p) and INMA(q) models for count time series.

Declarations

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



Appendix A: Proof of Proposition 1

Let $X \sim \mathcal{APS}(\theta, g(\alpha))$ and $Y \sim \mathcal{APS}(\gamma, g(\alpha))$ be independent variables, then we have

$$\begin{aligned}
 P(X + Y = n) &= \sum_{k=0}^n P(X = k)P(Y = n - k) \\
 &= \sum_{k=0}^n \frac{a_{\theta}(k)\{g(\alpha)\}^k}{h_{\theta}(\alpha)} \frac{a_{\gamma}(n - k)\{g(\alpha)\}^{n-k}}{h_{\gamma}(\alpha)} \\
 &= \frac{\{g(\alpha)\}^n}{h_{\theta}(\alpha)h_{\gamma}(\alpha)} \underbrace{\sum_{k=0}^n a_{\theta}(k)a_{\gamma}(n - k)}_{=(a_{\theta} * a_{\gamma})(n)}
 \end{aligned}$$

$$P(X + Y = n) = \frac{a_{\theta+\gamma}(n)\{g(\alpha)\}^n}{h_\theta(\alpha)h_\gamma(\alpha)}$$

Otherwise, using the Cauchy product we obtain

$$\begin{aligned} h_\theta(\alpha)h_\gamma(\alpha) &= \left(\sum_{n \in \mathbb{N}} a_\theta(n)\{g(\alpha)\}^n \right) \left(\sum_{n \in \mathbb{N}} a_\gamma(n)\{g(\alpha)\}^n \right) \\ &= \sum_{n \in \mathbb{N}} \sum_{k=0}^n a_\theta(k)\{g(\alpha)\}^k a_\gamma(n-k)\{g(\alpha)\}^{n-k} \\ &= \sum_{n \in \mathbb{N}} \{g(\alpha)\}^n \underbrace{\sum_{k=0}^n a_\theta(k)a_\gamma(n-k)}_{=(a_\theta * a_\gamma)(n)} \\ &= \sum_{n \in \mathbb{N}} a_{\theta+\gamma}(n)\{g(\alpha)\}^n \\ h_\theta(\alpha)h_\gamma(\alpha) &= h_{\theta+\gamma}(\alpha), \end{aligned}$$

and thus the desired result.

Appendix B: Proof of Proposition 2

Assume that there exists another parametric sequence b_θ such that (A1), (A2), (A3) hold and lead to the same three distributions. Based on assumption (A2) it means that

$$\frac{a_\theta(k)\{g(\alpha)\}^k}{h_\theta(\alpha)} = \frac{b_\theta(k)\{g(\alpha)\}^k}{f_\theta(\alpha)},$$

for all $k \in \mathbb{N}$, where h_θ and f_θ are normalizing constant. Therefore there exists $c > 0$ such that $h_\theta = c \cdot f_\theta$ and thus $a_\theta = c \cdot b_\theta$. Replace a_θ by $c \cdot b_\theta$ in (A1) and obtain

$$\frac{c^2 \cdot b_\theta(n)b_\gamma(n-k)}{c \cdot b_{\theta+\gamma}(n)} = \frac{b_\theta(n)b_\gamma(n-k)}{b_{\theta+\gamma}(n)},$$

for all $k = 0, \dots, n$ and $n \in \text{Supp } b_{\theta+\gamma}$. It means that $c = 1$ and thus $b_\theta = a_\theta$.

Appendix C: Proof of Theorem 1

The proof is given point by point.

1. Assume $X \sim \mathcal{C}_n(\theta, \gamma) \wedge \mathcal{C}_m(\theta + \gamma, \lambda)$. We have for $k \in \llbracket 0; m \rrbracket$

$$\begin{aligned} P(X = k) &= a_\theta(k) \sum_{n=0}^{m-k} \frac{a_\gamma(n)}{a_{\theta+\gamma}(n+k)} \frac{a_{\theta+\gamma}(n+k)a_\lambda(m-n-k)}{a_{\theta+\gamma+\lambda}(m)} \\ P(X = k) &= \frac{a_\theta(k)}{a_{\theta+\gamma+\lambda}(m)} \underbrace{\sum_{n=0}^{m-k} a_\gamma(n)a_\lambda(m-k-n)}_{=a_{\gamma+\lambda}(m-k)} \end{aligned}$$

and zero otherwise, i.e., $X \sim \mathcal{C}_m(\theta, \gamma + \lambda)$. The last equality is due to assumption (A1).

2. Assume $X \sim \mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{AP}\mathcal{S}(\theta + \gamma, g(\alpha))$. We have for $k \in \mathbb{N}$

$$P(X = k) = a_\theta(k) \sum_{n \geq 0} \frac{a_\gamma(n)}{a_{\theta+\gamma}(n+k)} \frac{a_{\theta+\gamma}(n+k) \{g(\alpha)\}^{n+k}}{h_{\theta+\gamma}(\alpha)}$$

$$P(X = k) = \frac{a_\theta(k)}{h_{\theta+\gamma}(\alpha)} \{g(\alpha)\}^k \underbrace{\sum_{n \geq 0} a_\gamma(n) \{g(\alpha)\}^n}_{=h_\gamma(\alpha)},$$

The last equality is due to assumption (A2). As demonstrated in proof of Proposition 1 we have $h_{\theta+\gamma} = h_\theta h_\gamma$ and thus $X \sim \mathcal{AP}\mathcal{S}(\theta, g(\alpha))$.

3. Assume $X \sim \mathcal{C}_n(\theta, \gamma) \wedge_n \mathcal{IC}(\theta + \gamma, \lambda; r)$. We have for $k \in \mathbb{N}$

$$P(X = k) = a_\theta(k) \sum_{n \geq 0} \frac{a_\gamma(n)}{a_{\theta+\gamma}(n+k)} \frac{r}{r+n+k} \frac{a_{\theta+\gamma}(n+k) a_\lambda(r)}{a_{\theta+\gamma+\lambda}(n+k+r)}$$

$$P(X = k) = \frac{r}{r+k} a_\theta(k) a_\lambda(r) \underbrace{\sum_{n \geq 0} \frac{r+k}{r+k+n} \frac{a_\gamma(n)}{a_{\theta+\gamma+\lambda}(n+k+r)}}_{=\frac{1}{a_{\theta+\lambda}(k+r)}},$$

i.e., $X \sim \mathcal{IC}(r; \theta, \lambda)$. The last equality is due to assumption (A3).

Appendix D: Proof of Lemma 1

According to equation (2) we have for $\mathbf{y} \in \text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \frac{a_{|\theta|}(|\mathbf{y}|) \{g(\alpha)\}^{|\mathbf{y}|}}{h_{|\theta|}(\alpha)} \frac{1}{a_{|\theta|}(|\mathbf{y}|)} \prod_{j=1}^J a_{\theta_j}(y_j)$$

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \frac{1}{h_{|\theta|}(\alpha)} \prod_{j=1}^J a_{\theta_j}(y_j) \{g(\alpha)\}^{y_j}$$

As demonstrated in proof of Proposition 1 we have $h_{\theta_1+\theta_2} = h_{\theta_1} h_{\theta_2}$. By induction on J we obtain $h_{|\theta|}(\alpha) = \prod_{j=1}^J h_{\theta_j}(\alpha)$ and thus

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \prod_{j=1}^J \frac{a_{\theta_j}(y_j) \{g(\alpha)\}^{y_j}}{h_{\theta_j}(\alpha)},$$

i.e., the desired result.

Appendix E: Proof of Theorem 2

To demonstrate Theorem 2 several results are necessary. Let first remark that the pmf (2) can be written as

$$P(\mathbf{N} = \mathbf{n}) = b(n_1 + \cdots + n_J) \prod_{j=1}^J a_{\theta_j}(n_j), \quad (3)$$

where $b(n) := \frac{p(n)}{a_{|\theta|}(n)}$ for all $n \in \text{Supp } a_{|\theta|}$. Since the support of $\mathcal{C}_{\Delta_n}(\theta) \wedge_n \mathcal{L}$ is included in the Cartesian product $\text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$, the possible factorization of the pmf is studied only on this set. Indeed, outside of this set, the factorization of $P(\mathbf{N} = \mathbf{n})$ into $\prod_{j=1}^J P(N_j = n_j)$ holds since for all $\mathbf{n} \notin \text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$ we have on one hand $P(\mathbf{N} = \mathbf{n}) = 0$ by definition and on the other hand at least one component j_0 such that $P(N_{j_0} = n_{j_0}) = 0$. In the following the pmf will thus be studied only on the Cartesian product $\text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$.

Lemma 2. *The minimal PGM of a convolution splitting distribution is necessary empty or complete.*

Proof. Since the right part of the pmf given by (3) is already totally factorized, study the factorization of $P(\mathbf{N} = \mathbf{n})$ is equivalent to study the factorization of $b(n_1 + \cdots + n_J)$. However, the sum being commutative, the function b is invariant under any permutation of values n_1, \dots, n_J . Therefore if there exists a factorization of b then it necessary concerns all the variables; i.e., there exist functions b_j such that $b(n_1 + \cdots + n_J) = \prod_{j=1}^J b_j(n_j)$. In this case the minimal PGM is empty, otherwise it is complete. \square

Using this lemma, it is now sufficient to show the equivalence

$$\mathcal{L} = \mathcal{APS}(|\theta|, g(\alpha)) \text{ for some } \alpha \in (0, R) \Leftrightarrow \text{empty PGM} \quad (4)$$

to obtain the following equivalence

$$\mathcal{L} \neq \mathcal{APS}(|\theta|, g(\alpha)) \text{ for all } \alpha \in (0, R) \Leftrightarrow \text{complete PGM}$$

Existence of additive power series distribution is a consequence of assumption (A2). The first implication in (4) is the easier to show, as a consequence of assumption (A1); see Lemma 1.

Now focus on the second implication of (4). Assume that the PGM is empty, i.e., the variables N_1, \dots, N_J are all mutually independent. We need the following lemma about the support of the distribution.

Lemma 3. *Let $\mathbf{N} \sim \mathcal{C}_{\Delta_n}(\theta) \wedge_n \mathcal{L}$ follow an additive convolution splitting distribution. Assume that the minimal PGM is empty. Then the support of the sum distribution \mathcal{L} is $\text{Supp } a_{|\theta|}$ and the support of the convolution splitting distribution is the entire Cartesian product $\text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$.*

Proof. Without additional assumption on the convolution splitting distribution, we can see that the support of the sum distribution $\text{Supp } \mathcal{L}$ is included in $\text{Supp } a_{|\theta|}$ by truncation. Remark that $\text{Supp } a_{|\theta|} = \text{Supp } a_{\theta_1} + \cdots + \text{Supp } a_{\theta_J} = \{\mathbf{n} : \mathbf{n} \in \text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}\}$. Now if we assume that there exists a non-empty set \mathcal{N} such that $\text{Supp } \mathcal{L} = \text{Supp } a_{|\theta|} \setminus \mathcal{N}$ then we will see that the independence assumption between N_1, \dots, N_J cannot hold. Indeed the support of the multivariate distribution becomes the Cartesian product $\text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$ minus the union of simplex $\cup_{n \in \mathcal{N}} \Delta_n$. This cannot take a form of a Cartesian product (except if $\mathcal{N} = \text{Supp } a_{|\theta|}$, i.e., $\text{Supp } \mathcal{L} = \emptyset$) and thus we obtain a contradiction. Recall that for any multivariate distribution, a necessary condition to have an empty PGM is that its support must be equal to the Cartesian product of their marginal supports. \square

Now the empty PGM assumption is equivalent to the factorization of the non-negative pmf among $\text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$. In other words, there exists J functions b_1, \dots, b_J such that

$$P(\mathbf{N} = \mathbf{n}) = \prod_{j=1}^J b_j(n_j) a_{\theta_j}(n_j),$$

for all $\mathbf{n} \in \text{Supp } a_{\theta_1} \times \cdots \times \text{Supp } a_{\theta_J}$, where each b_j is a positive function on $\text{Supp } a_{\theta_j}$. This is equivalent to

$$b(n_1 + \cdots + n_J) = \prod_{j=1}^J b_j(n_j), \quad (5)$$

since $a_{\theta_j}(n_j) > 0$. To simplify the proof, we will need this lemma about the distribution of merged vector $(N_1 + N_2, N_3, \dots, N_J)$.

Lemma 4. *Let $\mathbf{N} \sim \mathcal{C}_{\Delta_n}(\boldsymbol{\theta}) \wedge \mathcal{L}$ follow an additive convolution splitting distribution with at least $J \geq 3$ variables. Then if the first two variables are merged, the distribution is stable, i.e., we have*

$$(N_1 + N_2, N_3, \dots, N_J) \sim \mathcal{C}_{\Delta_n}(\theta_1 + \theta_2, \theta_3, \dots, \theta_J) \wedge \mathcal{L}.$$

Proof. This property is a direct consequence of assumption (A1). First note that the vectors \mathbf{N} and $(N_1 + N_2, N_3, \dots, N_J)$ have the same sum. Therefore it is sufficient to reason conditionally on the sum. Let $J = 3$ for convenience but the demonstration holds for any $J \geq 3$. Assume that $(N_1, N_2, N_3) | N_1 + N_2 + N_3 = n + m \sim \mathcal{C}_{\Delta_{n+m}}(\theta_1, \theta_2, \theta_3)$ for some $n \in \text{Supp } a_{\theta_1 + \theta_2}$ and $m \in \text{Supp } a_{\theta_3}$, then we have

$$\begin{aligned} \mathbb{P}_{|\mathbf{N}|=n+m}(N_1 + N_2 = n, N_3 = m) &= \sum_{(n_1+n_2) \in \Delta_n} \mathbb{P}(N_1 = n_1, N_2 = n_2, N_3 = m) \\ &= \sum_{(n_1+n_2) \in \Delta_n} \frac{a_{\theta_1}(n_1) a_{\theta_2}(n_2) a_{\theta_3}(m)}{a_{\theta_1+\theta_2+\theta_3}(n+m)} \\ &= \frac{a_{\theta_3}(m)}{a_{\theta_1+\theta_2+\theta_3}(n+m)} \underbrace{\sum_{(n_1+n_2) \in \Delta_n} a_{\theta_1}(n_1) a_{\theta_2}(n_2)}_{=a_{\theta_1+\theta_2}(n)} \\ \mathbb{P}_{|\mathbf{N}|=n+m}(N_1 + N_2 = n, N_3 = m) &= \frac{a_{\theta_1+\theta_2}(n) a_{\theta_3}(m)}{a_{\theta_1+\theta_2+\theta_3}(n+m)} \end{aligned}$$

□

If N_1, \dots, N_J are mutually independent then N_1 is independent of the partial sum $N_2 + \cdots + N_J$. According to lemma 4 the vector of two variables $(N_1, N_2 + \cdots + N_J)$ follows the additive convolution splitting distribution $\mathcal{C}_{\Delta_n}(\theta_1, \theta_2 + \cdots + \theta_J) \wedge \mathcal{L}$. Therefore if the result (that we want to demonstrate) is true for $J = 2$ then $\mathcal{L} = \mathcal{AP}\mathcal{S}(\theta_1 + (\theta_2 + \cdots + \theta_J), g(\alpha)) = \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha))$ and the result stays true for any number of variables $J > 2$. Now let split the demonstration in two cases: discrete and continuous case.

The discrete case

In the binary case $J = 2$, the equation (5) becomes:

$$b(n+m) = c(n)d(m), \quad (6)$$

for all $n \in \text{Supp } a_{\theta_1}$, $m \in \text{Supp } a_{\theta_2}$, where c and d are positive for these values. Taking $n = 0$ we obtain $b(m) = c(0)d(m)$ for all $m \in \text{Supp } a_{\theta_2}$. Since c is positive we have $d(m) =$

$c(0)^{-1}b(m)$ for all $m \in \text{Supp } a_{\theta_2}$. With similar argument we obtain $c(n) = d(0)^{-1}b(n)$ for all $n \in \text{Supp } a_{\theta_1}$. Replacing in the first equation we obtain

$$b(n+m) = \{c(0)d(0)\}^{-1}b(n)b(m),$$

for all $n \in \text{Supp } a_{\theta_1}$ and $m \in \text{Supp } a_{\theta_2}$. Using assumption (A1) we know that $0 \in \text{Supp } a_{\theta_1}$ and $0 \in \text{Supp } a_{\theta_2}$, thus we have $b(0) = c(0)d(0) > 0$. We then multiply by $b(0)^{-1}$ on each side and obtain

$$B(n+m) = B(n)B(m), \quad (7)$$

for all $(n, m) \in \text{Supp } a_{\theta_1} \times \text{Supp } a_{\theta_2}$, where $B(n) := b(0)^{-1}b(n)$ for all $n \in \text{Supp } a_{\theta_1+\theta_2}$. This is the Cauchy exponential equation over a subset of non-negative integers of the form $\{0, 1, \dots, k\}$ for some $k \in \mathbb{N}^* \cup \{\infty\}$, according to assumption (A1). Taking $m = 1$ we obtain $B(n+1) = B(n)B(1)$ for all $n \in \{0, 1, \dots, k-1\}$ and thus recursively $B(n) = B(1)^n$ for all $n \in \{0, 1, \dots, k\}$. Finally $b(n) = b(0)x^n$ for all $n \in \{0, 1, \dots, k\}$, with $x = b(1)/b(0) > 0$. Back to the sum distribution we obtain $p(n) = a_{|\theta|}(n)b(n)$, i.e., $p(n)$ is proportional to $a_{|\theta|}(n)x^n$. Therefore this is an additive modified power series distribution $\mathcal{AP}\mathcal{S}(|\theta|, g(\alpha))$ for some $\alpha \in (0, R)$ (if x is independent of $a_{\theta}(n)$ then this is an additive power series distribution).

The continuous case

In the continuous case we cannot make the assumption (A0) that $\text{Supp } a_{\theta} = [0, w)$ for some $w \in \mathbb{R}^* \cup \{\infty\}$. Indeed the only known example of parametric function a_{θ} is such that $\text{Supp } a_{\theta} = (0, \infty)$. Since $0 \notin \text{Supp } a_{\theta}$ we cannot obtain the Cauchy exponential equation (7) from the equation (6). We thus use another way to demonstrate the desired result. Assume that $\mathbf{X} = (X_1, \dots, X_J) \sim \mathcal{C}_{\Delta_x}(\boldsymbol{\theta}) \wedge \mathcal{L}$ and that X_1, \dots, X_J are mutually independent. We aim at showing that $\mathcal{L} = \mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha))$ for some $\alpha > 0$ (note that power series distributions are also known as member of exponential family of distributions). We need to add an assumption (A*) on the parametric function a_{θ}

$$\frac{a_{\theta}(y)a_{\gamma}(t-y)}{a_{\theta+\gamma}(t)} = \frac{a_{\theta}(y/t)a_{\gamma}(1-y/t)}{ta_{\theta+\gamma}(1)}$$

which is equivalent to state that $\mathcal{C}_x(\boldsymbol{\theta}, \gamma)$ can be viewed as a dilatation of $\mathcal{C}_1(\boldsymbol{\theta}, \gamma)$, i.e., $Y \sim \mathcal{C}_1(\boldsymbol{\theta}, \gamma)$ is equivalent to $x \cdot Y \sim \mathcal{C}_x(\boldsymbol{\theta}, \gamma)$ for all $x \in \mathbb{R}_+^*$. In this case the convolution splitting distribution can be viewed as an external independent product

$$\mathcal{C}_{\Delta_x}(\boldsymbol{\theta}) \wedge \mathcal{L} = \mathcal{C}_{\Delta}(\boldsymbol{\theta}) \otimes \mathcal{L},$$

i.e., $\mathbf{X} \sim \mathcal{C}_{\Delta_x}(\boldsymbol{\theta}) \wedge \mathcal{L}$ is equivalent to $\mathbf{X}/|\mathbf{X}| \sim \mathcal{C}_{\Delta}(\boldsymbol{\theta})$ and $|\mathbf{X}| \sim \mathcal{L}$ independently. Now we can use the Lukacs's proportion sum independence theorem (more precisely its generalization from $J = 2$ to $J > 2$).

Theorem 4 ((Lukacs, 1955)). *Let $J \geq 2$ and X_1, \dots, X_J be mutually independent non-negative continuous random variables. Then the random sum $|\mathbf{X}|$ and the random normalized vector $\mathbf{X}/|\mathbf{X}|$ are independent if and only if there exists some $(\theta_1, \dots, \theta_J) \in \mathbb{R}_+^{*J}$ and $b \in \mathbb{R}_+^*$ such that $X_j \sim \Gamma(\theta_j, b)$ for all $j = 1, \dots, J$.*

In such a situation it is well known that $|\mathbf{X}| \sim \Gamma(|\boldsymbol{\theta}|, b)$. We have already seen that this distribution is exactly an additive power series distribution $\mathcal{AP}\mathcal{S}(|\boldsymbol{\theta}|, g(\alpha))$ with $a_{\theta}(x) =$

$x^{\theta-1}/\Gamma(\theta)$ and $\alpha = e^{-b} \in (0, 1)$. We have also already checked that assumptions (A1), (A2) and (A3) hold. Then, Lemma 1 implies that in the situation of Lukacs's Theorem we have

$$\mathbf{X} \sim \mathcal{D}_{\Delta_x}(\boldsymbol{\theta}) \wedge_x \Gamma(|\boldsymbol{\theta}|, b) = \bigotimes_{j=1}^J \Gamma(\theta_j, b).$$

Finally, the uniqueness of the sum distribution (such that the PGM of \mathbf{X} is empty) is obtained by the Lukacs's Theorem, replacing the assumption (A0) by (A*). We also obtain that the only one parametric function $a_\theta(x)$ that respects assumption (A*), (A1), (A2) and (A3) is $a_\theta(x) = x^{\theta-1}/\Gamma(\theta)$.

Appendix F: Proof of Theorem 3

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

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

The support of $p(\cdot|x)$ is related to the support of $a_\theta^{[c,d]}(\cdot)$. For the special case $(c, d) = (-1, 0)$, we have $\text{Supp } a_\theta^{[-1,0]} = \{0, \dots, \theta\}$. In this case, let show that $p(x+1|x) > 0$ for all $x \in \{0, \dots, \theta + \gamma\}$. It is sufficient to show that there exists $k \leq x$ such that $k \leq \theta$ and $k + \gamma \geq x + 1$. Remark that $k = \min(x, \theta)$ is a good candidate. In the same way it can be shown that $p(x|x+1) > 0$. For other cases of (c, d) values the demonstration is straightforward since $\text{Supp } a_\theta^{[c,d]} = \mathbb{N}$. The Markov chain $(X_t)_{t \in \mathbb{N}}$ is thus ergodic.

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

$$\begin{aligned} \mathcal{L}_{t+1} &= \left\{ q\mathcal{P}_n^{[c,d]}(\theta, \gamma) \wedge_n \mathcal{AP}\mathcal{S}^{[c,d]}(\theta + \gamma, g(\alpha)) \right\} * \mathcal{AP}\mathcal{S}^{[c,d]}(\gamma, g(\alpha)) \\ &= \mathcal{AP}\mathcal{S}^{[c,d]}(\theta, g(\alpha)) * \mathcal{AP}\mathcal{S}^{[c,d]}(\gamma, g(\alpha)) \\ \mathcal{L}_{t+1} &= \mathcal{AP}\mathcal{S}^{[c,d]}(\theta + \gamma, g(\alpha)) \end{aligned}$$

where the second and third equalities are obtain according to Theorem 1 and Proposition 1 respectively.

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

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

which is symmetric.

Funding

This research was supported by the GAMBAS project funded by the French National Research Agency (ANR-18-CE02-0025).

Supplementary Material

Supplementary materials for article entitled "On quasi Pólya thinning operator"

This supplementary files includes seven sections:

1. Examples of convolution distributions
2. Examples of additive modified power series distributions
3. Examples of inverse convolution distributions
4. Equivalence between distributions
5. Details about the specific parametric sequence $a_\theta(n) = \binom{n+\theta-1}{n}$
6. Characteristics of multivariate convolution distributions
7. Multivariate count distributions closed under marginalization

References

- AL-OSH, M. and ALZAID, A. (1991). Binomial autoregressive moving average models. *Stochastic Models* **7** 261–282.
- ALZAID, A. A. and AL-OSH, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* **45** 223–232.
- BOL'SHEV, L. N. (1965). On a characterization of the Poisson distribution and its statistical applications. *Theory of Probability & Its Applications* **10** 446–456.
- CASTAÑER, A., CLARAMUNT, M. M., LEFÈVRE, C. and LOISEL, S. (2015). Discrete Schur-constant models. *Journal of Multivariate Analysis* **140** 343–362.
- DAVIS, R. A., FOKIANOS, K., HOLAN, S. H., JOE, H., LIVSEY, J., LUND, R., PIPIRAS, V. and RAVISHANKER, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association* **116** 1533–1547.
- GUPTA, R. C. (1974). Modified power series distribution and some of its applications. *Sankhyā: The Indian Journal of Statistics, Series B* 288–298.
- JAMBUNATHAN, M. (1954). Some properties of beta and gamma distributions. *The annals of mathematical statistics* 401–405.
- JANARDAN, K. (1973). A new four urn model with predetermined strategy. In *Technical Report 37-I* Sangamon State University Springfield IL.
- JANARDAN, K. (1975). Markov-Polya urn-model with pre-determined strategies. *Gujarat Statistical Review* **2** 17–32.
- JANARDAN, K. and RAO, R. B. (1982). Characterization of generalized Markov-Polya and generalized Polya-Eggenberger distributions. *Communications in Statistics-Theory and Methods* **11** 2113–2124.
- JANARDAN, K. and RAO, R. B. (1986). Identifiability of a generalized markov-polya damage model. *Statistics: A Journal of Theoretical and Applied Statistics* **17** 303–310.
- JOE, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* **33** 664–677.
- JONES, M. and MARCHAND, É. (2019). Multivariate discrete distributions via sums and shares. *Journal of Multivariate Analysis* **171** 83–93.
- JORGENSEN, B. (1997). *The theory of dispersion models*. CRC Press.
- JOURDAN, A. and KOKONENDJI, C. C. (2002). Surdispersion et modèle binomial négatif généralisé. *Revue de statistique appliquée* **50** 73–86.
- KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- LAURITZEN, S. L. (1996). *Graphical models* **17**. Clarendon Press.
- LUKACS, E. (1955). A characterization of the gamma distribution. *The Annals of Mathematical Statistics* **26** 319–324.
- MCKENZIE, E. (1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association* **21** 645–650.
- MORAN, P. (1952). A characteristic property of the Poisson distribution. In *Mathematical Proceedings of the Cambridge Philosophical Society* **48** 206–207. Cambridge University Press.
- OVASKAINEN, O. and ABREGO, N. (2020). *Joint species distribution modelling: With applications in R*. Cambridge University Press.
- PATIL, G. and RATNAPARKHI, M. (1975). Problems of damaged random variables and related characterizations. In *A Modern Course on Statistical Distributions in Scientific Work* 255–270. Springer.

- PEYHARDI, J. (2023). Supplement to "On quasi Pólya thinning operator".
- PEYHARDI, J. and FERNIQUE, P. (2017). Characterization of convolution splitting graphical models. *Statistics & Probability Letters* **126** 59–64.
- PEYHARDI, J., FERNIQUE, P. and DURAND, J.-B. (2021). Splitting models for multivariate count data. *Journal of Multivariate Analysis* **181** 104677.
- PEYHARDI, J., LAROCHE, F. and MORTIER, F. (2022). Pólya-splitting distributions as stationary solutions of multivariate birth-death processes under extended neutral theory. Available at SSRN 4412745.
- PUIG, P. and VALERO, J. (2007). Characterization of count data distributions involving additivity and binomial subsampling. *Bernoulli* 544–555.
- RAO, C. R. (1949). On some problems arising out of discrimination with multiple characters. *Sankhyā: The Indian Journal of Statistics* 343–366.
- RAO, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A* 311–324.
- RAO, B. R. and JANARDAN, K. (1984). The Use of the Generalized Markov-Polya Distribution as a Random Damage Model and Its Identifiability. *Sankhyā: The Indian Journal of Statistics, Series A* 458–462.
- RAO, C. R. and RUBIN, H. (1964). On a characterization of the Poisson distribution. *Sankhyā: The Indian Journal of Statistics, Series A* 295–298.
- SCOTTO, M. G., WEISS, C. H. and GOUVEIA, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling* **15** 590–618.
- SHANBHAG, D. and PANARETOS, J. (1979). Some results related to the Rao-Rubin characterization of the Poisson distribution. *Austral. J. Statist* **21** 78–83.
- SPROTT, D. (1965). A class of contagious distributions and maximum likelihood estimation. *Sankhyā: The Indian Journal of Statistics, Series A* 369–382.
- STEUTEL, F. W. and VAN HARN, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability* 893–899.
- STEUTEL, F. W., KENT, J. T., BONDESSON, L. and BARNDORFF-NIELSEN, O. (1979). Infinite divisibility in theory and practice [with discussion and reply]. *Scandinavian Journal of Statistics* 57–64.
- TANG, Z.-Z. and CHEN, G. (2019). Zero-inflated generalized Dirichlet multinomial regression model for microbiome compositional data analysis. *Biostatistics* **20** 698–713.
- TEICHER, H. (1954). On the convolution of distributions. *The Annals of Mathematical Statistics* 775–778.
- WANG, T. and ZHAO, H. (2017). A Dirichlet-tree multinomial regression model for associating dietary nutrients with gut microorganisms. *Biometrics* **73** 792–801.
- WEISS, C. H. (2008). Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis* **92** 319–341.
- XEKALAKI, E. (1986). The multivariate generalized Waring distribution. *Communications in Statistics - Theory and Methods* **15** 1047–1064.