Splitting models for multivariate count data

Jean Peyhardi[a,*], Pierre Fernique[b], Jean-Baptiste Durand[c]

[a] *IMAG, Université de Montpellier, place Eugène Bataillon, 34090 Montpellier, France*
[b] *CIRAD, UMR AGAP, 34098 Montpellier*
[c] *Laboratoire Jean Kuntzmann, 38058 Grenoble, France*

Abstract

We introduce the class of splitting distributions as the composition of a singular multivariate distribution and an univariate distribution. This class encompasses most common multivariate discrete distributions (multinomial, negative multinomial, multivariate hypergeometric, multivariate negative hypergeometric, ...) and contains several new ones. We highlight many probabilistic properties deriving from the compound aspect of splitting distributions and their underlying algebraic properties. These simplify the study of their characteristics, inference methods, interpretation and extensions to regression models. Parameter inference and model selection are thus reduced to two separate problems, preserving time and space complexity of the base models. In the case of multinomial splitting distributions, conditional independence and asymptotic normality properties for estimators are obtained. The use of splitting regression models is illustrated on three data sets analyzed with reproducible methodology.

Keywords: compound distribution, singular multivariate distribution, multivariate extension, probabilistic graphical model.

---

1. Introduction

The analysis of multivariate count data is a crucial issue in numerous application settings, particularly in the fields of biology [1], ecology [10] and econometrics [50]. Multivariate count data are defined as the number of items of different categories issued from sampling within a population, whose individuals are grouped. Denoting by J this number of categories, multivariate count data analysis relies on modeling the joint distribution of the discrete random vector $\boldsymbol{y} = (y_1, \ldots, y_J)$. In genomics for instance, the data obtained from sequencing technologies are often summarized by the counts of DNA or RNA fragments within a genomic interval (e.g., RNA seq data). The most usual models in this framework, are multinomial and Dirichlet multinomial regression to take account of some environmental covariate effects on these counts. In this way, Xia et al. [53] and Chen and Li [7] studied the microbiome composition (whose output are J bacterial taxa counts), while Zhang et al. [55] studied the expression count of J exon sets.

However, the multinomial and Dirichlet multinomial distributions are not appropriate for modeling the variability in the total number of counts in multivariate count data, because of their support: the discrete simplex $\Delta_n := \left\{\boldsymbol{y} \in \mathbb{N}^J : \sum_{j=1}^J y_j = n\right\}$. This particular support also induces a strong constraint in term of dependencies between the components of $\boldsymbol{y}$, since any component y_j is deterministic when the $J-1$ other components are known. This kind of distributions are said to be singular and will be denoted by $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta})$. The parameter n being related to the support, is intentionally noted as an index of the distribution, distinguishing it from other parameters $\boldsymbol{\theta}$ used to define the probability mass function (pmf). Note that initially, singular versions of some multivariate distributions have been defined by Patil [33] and Janardan and Patil [20]. However, these distinctions were unheeded until now, leading to misuse of these distributions [55]. Therefore, a distribution will be considered as a J-multivariate distribution if

1. the dimension of its support is equal to the number of variables (i.e., $\dim\{\mathrm{Supp}(\boldsymbol{y})\} = J$).

---

*Corresponding author. Email address: `jean.peyhardi@umontpellier.fr`

Another problem that occurs when defining multivariate distributions is the independence relationships between components $y_1, \dots, y_J$. For instance, the multiple Poisson distribution described by Patil and Bildikar [34], involves J mutually independent variables. Therefore, a multivariate distribution will be considered as a *sensu stricto* multivariate distribution if:

2. its probabilistic graphical model (in the sense of undirected graphs, see 26) is connected, i.e., there is a path between every pair of variables, meaning that no variable is independent of another.

Additionally, such a distribution is considered as an extension of an univariate distribution if:

3. all the univariate marginal distributions belong to the same family (extension),

4. all the multivariate marginal distributions belong to the same family (natural extension).

Even if a singular distribution is not a *sensu stricto* J-multivariate distribution, it is very versatile as soon as the parameter n is considered as a random variable. It then becomes a map between spaces of univariate and multivariate distributions. Assuming that n follows an univariate distribution $\mathcal{L}(\psi)$ (e.g., binomial, negative binomial, Poisson etc ...), the resulting compound distribution, denoted by $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}(\psi)$, is called splitting distribution. For instance, the multivariate hypergeometric - resp. multinomial and Dirichlet multinomial - splitting distribution, denoted by $\mathcal{H}_{\Delta_n}(\boldsymbol{k}) \underset{n}{\wedge} \mathcal{L}(\psi)$ - resp. by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\psi)$ and $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{L}(\psi)$ - has been introduced by Peyhardi and Fernique [36]. They studied the graphical model of independence for such distributions according to the sum distribution $\mathcal{L}(\psi)$. Jones and Marchand [23] studied the Dirichlet multinomial splitting distributions, and named it sum and Polya share distributions. They focused on the Dirichlet multinomial splitting negative binomial distribution, denoted here by $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{NB}(r, p)$.

Under mild assumptions, splitting distributions can be considered as *sensu stricto* multivariate distributions. They include all usual multivariate discrete distributions and several new ones. Many advantages derive from the compound aspect of splitting distributions. The interpretation is simply decomposed into two parts: the sum distribution (intensity of the distribution) and the singular distribution (repartition into the J components). The log-likelihood can also be decomposed according to these two parts and thus easily computed and maximized. This also facilitates the derivation of asymptotic and independence properties for maximum likelihood and Bayesian estimators. All usual characteristics (support, pmf, expectation, covariance and probability generative function (pgf)) are also easily obtained using this decomposition. Finally, the generalization to regression models is naturally achieved by compounding a singular regression by an univariate regression. This new framework eases the definition of a generalized linear models (GLMs) for multivariate count responses, taking account of the dependence between response components.

This article is organized as follows. In Section 2 notations used all along the paper are introduced. The definition of singular distributions is used as a building block to introduce splitting distributions. Positive and symmetric singular distributions are introduced, easing respectively the study of criteria 1-2 and 3-4 for resulting splitting distributions. In Section 3 the subclass of additive convolution splitting distributions are introduced in order to simplify the calculation of marginal distributions. Sections 4 and 5 focus on splitting distributions obtained with the multinomial and the Dirichlet multinomial distributions since they are both positive and additive (e.g., the multivariate hypergeometric is an additive, but non-positive convolution distribution). This leads us to precisely describe fifteen multivariate extensions (among which five are natural extensions) of usual univariate distributions giving their usual characteristics. Some detailed attention is given to maximum likelihood and Bayesian parameter estimation regarding multinomial splitting distributions. Conditional independence properties and asymptotic normality for sum and singular distribution parameters are discussed in this framework. It is then showed that multinomial splitting regression constitutes an appropriate framework to introduce a family of GLMs for multivariate count responses. In Section 6 a comparison of these regression models on two benchmark datasets and an application on a mango tree dataset are proposed.

2. Splitting distributions

2.1. Notations

All along the paper, focus will be made only on count distributions (and regression models). For notational convenience, the term count will therefore be omitted. Let $|\boldsymbol{y}| = \sum_{j=1}^{J} y_j$ denote the sum of the random vector $\boldsymbol{y}$

and assume that $|\boldsymbol{y}| \sim \mathcal{L}(\psi)$. Let $P_B(A)$ denotes the conditional probability of A given B. Let $\mathrm{E}_{|\boldsymbol{y}|}(\boldsymbol{y})$ and $\mathrm{Cov}_{|\boldsymbol{y}|}(\boldsymbol{y})$ denote respectively the conditional expectation and covariance of the random vector $\boldsymbol{y}$ given the sum $|\boldsymbol{y}|$. Let $\blacktriangle_n^J = \left\{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| \leq n\right\}$ denote the discrete corner of the hypercube. If no confusion could arise, J will be omitted in the former notations. Let $\binom{n}{\boldsymbol{y}} = n!/(n-|\boldsymbol{y}|)! \prod_{j=1}^J y_j!$ denote the multinomial coefficient defined for $\boldsymbol{y} \in \blacktriangle_n$. This notation replaces the usual notation $\binom{n}{\boldsymbol{y}} = n!/\prod_{j=1}^J y_j!$ which is defined only for $\boldsymbol{y} \in \Delta_n$. Let $(a)_n = \Gamma(a+n)/\Gamma(a)$ denote the Pochhammer symbol and $B(\boldsymbol{\alpha}) = \prod_{j=1}^J \Gamma(\alpha_j)/\Gamma(|\boldsymbol{\alpha}|)$ the multivariate beta function. Let

$$
{}_2^J F_2\{(a, a'); \boldsymbol{b}; (c, c'); \boldsymbol{s}\} = \sum_{\boldsymbol{y} \in \mathbb{N}^J} \frac{(a)_{|\boldsymbol{y}|}(a')_{|\boldsymbol{y}|} \prod_{j=1}^J (b_j)_{y_j}}{(c)_{|\boldsymbol{y}|}(c')_{|\boldsymbol{y}|}} \prod_{j=1}^J \frac{s_j^{y_j}}{y_j!}
$$

denote a multivariate hypergeometric function. Remark that $a' = c'$ lead to ${}_1^J F_1(a; \boldsymbol{b}; c; \boldsymbol{s})$ the Lauricella's type D function [27]. Moreover, if $J = 1$ then it turns out to be the usual Gauss hypergeometric function ${}_2F_1(a; b; c; s)$ or the confluent hypergeometric ${}_1F_1(b; c; s)$.

2.2. *Definitions*

As previously introduced, a distributions is said to be singular if its support is included in the simplex and will be denoted by $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta})$. The parameter n being related to the support, is intentionally noted as an index of the distribution, distinguishing it from other parameters $\boldsymbol{\theta} \in \Theta^J$ used to define the pmf. Moreover, a singular distribution is said to be:

- positive, if for any $n \in \mathbb{N}$, $p_{|\boldsymbol{y}|=n}(\boldsymbol{y}) > 0$ for all $\boldsymbol{y} \in \Delta_n$ (i.e., if its support is the whole discrete simplex),
- symmetric, if $\boldsymbol{y} \sim \mathcal{S}_{\Delta_n}(\boldsymbol{\theta}) \Rightarrow \sigma(\boldsymbol{y}) \sim \mathcal{S}_{\Delta_n}\{\sigma(\boldsymbol{\theta})\}$ for all permutation σ of $\{1, \ldots, J\}$ (i.e., if it is closed under permutation).

Remark that if the singular distribution is symmetric, then it is possible to define a non-singular extension having as support a subset of $\blacktriangle_n$ (or exactly $\blacktriangle_n$ if the singular distribution is also positive). The symmetry ensures that the choice of the last category to complete the vector, has no impact on the distribution. Such a distribution for the random vector $\boldsymbol{y}$, denoted by $\mathcal{S}_{\blacktriangle_n}(\boldsymbol{\theta}, \gamma)$, is defined such that $(\boldsymbol{y}, n - |\boldsymbol{y}|) \sim \mathcal{S}_{\Delta_n^{J+1}}(\boldsymbol{\theta}, \gamma)$.

The random vector $\boldsymbol{y}$ is said to follow a splitting distribution if there exists a singular distribution $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta})$ and an univariate distribution $\mathcal{L}(\psi)$ such that $\boldsymbol{y}$ follows the compound distribution

$$
\mathcal{S}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi}). \tag{1}
$$

It is named splitting distribution since an outcome $y \in \mathbb{N}$ of the univariate distribution $\mathcal{L}(\psi)$ is split into the J components. The pmf is then given by $p(\boldsymbol{y}) = p_{|\boldsymbol{y}|}(\boldsymbol{y})p(|\boldsymbol{y}|)$ assuming that $|\boldsymbol{y}|$ follows the univariate distribution $\mathcal{L}(\psi)$ and $\boldsymbol{y}$ given $|\boldsymbol{y}| = n$ follows the singular multivariate distribution $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta})$. Note that all univariate distributions bounded by n (denoted by $\mathcal{L}_n(\theta)$) are non-singular (univariate) distributions. The variable y is said to follow a damage distribution if there exists a bounded distribution $\mathcal{L}_n(\theta)$ and a distribution $\mathcal{L}(\psi)$ such that y follows the compound distribution $\mathcal{L}_n(\theta) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$. It is named damage distribution since an outcome $y \in \mathbb{N}$ of the distribution $\mathcal{L}(\psi)$ is damaged into a smaller value. Remark that the marginal (univariate) of any splitting distribution is a damage distribution.

Examples. Here we highlight five examples of singular distributions:

1. the multivariate hypergeometric distribution, denoted by $\mathcal{H}_{\Delta_n}(\boldsymbol{k})$ where $\boldsymbol{k} \in \mathbb{N}^J$, with pmf given by

$$
p_{|\boldsymbol{y}|=n}(\boldsymbol{y}) = \frac{\prod_{j=1}^J \binom{k_j}{y_j} \mathbf{1}_{y_j \leq k_j}}{\binom{|\boldsymbol{k}|}{n}}.
$$

2. the multinomial distribution, denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ where $\boldsymbol{\pi} \in \Delta$, with pmf given by

$$
p_{|\boldsymbol{y}|=n}(\boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \prod_{j=1}^J \pi_j^{y_j}.
$$

	Positive	Symmetric	Additive convolution	Proportional
Multivariate hypergeometric		×	×	
Multinomial	×	×	×	×
Dirichlet multinomial	×	×	×	
Generalized Dirichlet multinomial	×			
Logistic normal multinomial	×	×		

Table 1: Properties of five singular distributions.

3. The Dirichlet multinomial distribution, denoted by $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha} \in (0,\infty)^J$, with pmf given by

$$p_{|\mathbf{y}|=n}(\mathbf{y}) = \frac{\prod_{j=1}^{J} \binom{y_j+\alpha_j-1}{y_j}}{\binom{n+|\alpha|-1}{n}}.$$

4. the generalized Dirichlet multinomial distribution, denoted by $\mathcal{GDM}_{\Delta_n}(\boldsymbol{\alpha},\boldsymbol{\beta})$ where $\boldsymbol{\alpha} \in (0,\infty)^{J-1}$ and $\boldsymbol{\beta} \in (0,\infty)^{J-1}$, with pmf given by

$$p_{|\mathbf{y}|=n}(\mathbf{y}) = \binom{n}{\mathbf{y}} \prod_{j=1}^{J-1} \frac{(\alpha_j)_{(y_j)}(\beta_j)_{(\mathbf{y}_{\geq j+1})}}{(\alpha_j+\beta_j)_{(\mathbf{y}_{\geq j})}},$$

   where $\mathbf{y}_{\geq j} := y_j + \cdots + y_J$.

5. the logistic normal multinomial distribution, denoted by $\mathcal{LNM}_{\Delta_n}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ where $\boldsymbol{\mu} \in (-\infty,\infty)^{J-1}$ and $\boldsymbol{\Sigma}$ is a real square symmetric definite positive matrix of dimension $J-1$. This is a multinomial distribution mixed by a logistic normal distribution, i.e., $\mathcal{LNM}_{\Delta_n}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{\boldsymbol{\pi}}{\wedge} \mathcal{LN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. According to [2], $\boldsymbol{\pi} \sim \mathcal{LN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is equivalent to $\phi(\boldsymbol{\pi}) \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ where $\phi(\boldsymbol{\pi}) := \left(\ln \frac{\pi_j}{\pi_J}, \ldots, \ln \frac{\pi_{J-1}}{\pi_J}\right)$.

These specific singular distributions allow us to introduce five families of splitting distributions for multivariate count data, based on composition (1). Contrarily to the others, the multivariate hypergeometric distribution is not positive since its support is the intersection of the simplex Δ_n and the hyper-rectangle $\blacksquare_{\boldsymbol{k}} = \{\mathbf{y} \in \mathbb{N}^J : y_1 \leq k_1, \ldots, y_J \leq k_J\}$. Contrarily to the others, the generalized Dirichlet multinomial distribution is not symmetric. An ordering relation among components $1,\ldots,J$ is take into account. These properties are summarized in Table 1. The last three singular distributions can be viewed as multinomial distributions mixed by $\boldsymbol{\pi}$, respectively by a Dirichlet, a generalized Dirichlet [8] and a logistic normal distributions [2].

2.3. *Sensu stricto multivariate extensions*

This subsection highlights some sufficient conditions on the singular and the sum distributions to obtain a *sensu stricto* multivariate distribution (i.e., such that criteria 1 and 2 hold) or a multivariate extension (i.e., such that criteria 1, 2 and 3 hold).

Support. Firstly, let us remark that a singular distribution could be viewed as particular splitting distribution if the sum follows a Dirac distribution (denoted by $\mathbf{1}_n$), i.e. $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta}) = \mathcal{S}_{\Delta_m}(\boldsymbol{\theta}) \underset{m}{\wedge} \mathbf{1}_n$. Assume that the dimension of a set $A \subseteq \mathbb{N}^J$ is defined as the dimension of the smaller $\mathbb{R}$-vectorial space including A. The dimension of the support of a positive splitting distribution depends on the support of the sum distribution as follows:

$$\dim\left[\operatorname{Supp}\left\{\mathcal{S}_{\Delta_m}(\boldsymbol{\theta}) \underset{m}{\wedge} \mathcal{L}(\psi)\right\}\right] = \begin{cases} 0 & \text{if } \mathcal{L}(\psi) = \mathbf{1}_0, \\ J-1 & \text{if } \mathcal{L}(\psi) = \mathbf{1}_n \text{ with } n \in \mathbb{N}^*, \\ J & \text{otherwise.} \end{cases}$$

Therefore all positive splitting distributions are considered as multivariate distributions (criterion 1 holds) when the sum is not a Dirac distribution (only non-Dirac distributions will therefore be considered hereafter).

Independence. A probabilistic graphical model (or graphical model, in short) is defined by a distribution and a graph such that all independence assertions that are derived from the graph using the global Markov property hold in the distribution [26]. A graphical model is said to be minimal, if any edge removal in the graph induces an independence assertion that is not held in the distribution. A graphical model is said to be connected if there exists a path containing all its vertices (i.e., there is no pair of independent variables). This is a necessary condition (criterion 2) to obtain a *sensu stricto* multivariate distribution. Peyhardi and Fernique [36] characterized the graphical model of multinomial and Dirichlet multinomial splitting distributions according to the sum distribution. In cases where the exact graph cannot be obtained easily, it is sufficient to show that covariances are strictly positive to ensure that at least one path connects every pair of random variables in the graph. Moments can be derived using the law of total expectation

$$\mathrm{E}\,(\boldsymbol{y}) = \mathrm{E}\left\{\mathrm{E}_{|\boldsymbol{y}|}\,(\boldsymbol{y})\right\}, \tag{2}$$

and covariance

$$\mathrm{Cov}\,(\boldsymbol{y}) = \mathrm{E}\left\{\mathrm{Cov}_{|\boldsymbol{y}|}\,(\boldsymbol{y})\right\} + \mathrm{Cov}\left\{\mathrm{E}_{|\boldsymbol{y}|}\,(\boldsymbol{y})\right\}. \tag{3}$$

For instance the covariance of the multivariate hypergeometric splitting distribution $\mathcal{H}_{\Delta_n}(\boldsymbol{k}) \underset{n}{\wedge} \mathcal{L}(\psi)$ is given by

$$\mathrm{Cov}\,(\boldsymbol{y}) = \frac{1}{|\boldsymbol{k}|} \cdot \left[\{(|\boldsymbol{k}|-1)\mu_1 + \mu_2\} \cdot \mathrm{diag}(\boldsymbol{k}) + \left\{\mu_2 - \frac{|\boldsymbol{k}|-1}{|\boldsymbol{k}|}\mu_1^2\right\} \cdot \boldsymbol{k}\boldsymbol{k}^t\right],$$

where μ_i denotes the factorial moment of order i ($i = 1, 2$) for the sum distribution. To our knowledge, the binomial distribution with $|\boldsymbol{k}|$ trials, is the only one parametric distributions such that $|\boldsymbol{k}|\mu_2 = (|\boldsymbol{k}| - 1)\mu_1^2$. Therefore, any other parametric distribution can be used for the sum in order to ensure that covariances between any pair of components are positive. This method could also be used for generalized Dirichlet multinomial and logistic normal multinomial splitting distributions since their graphical models have not yet been characterized. Finally, the pgf of splitting distributions can be obtained from the pgf of the singular distribution since

$$G(\boldsymbol{s}) = \mathrm{E}\left\{\bar{G}\,(\boldsymbol{s})\right\}, \tag{4}$$

where $\boldsymbol{s} = (s_1, \ldots, s_J)$ and $\bar{G}$ denotes the pgf of $\boldsymbol{y}$ given the sum $|\boldsymbol{y}|$.

Marginals. Splitting distributions that are sensu stricto multivariate distributions (i.e., with criteria 1 and 2) are not necessarily multivariate extensions. To be considered as a multivariate extension of a specific family, the marginal distributions of y_j must belong to this family. Remark that the symmetry of the singular distribution is a sufficient condition to obtain a multivariate extension (i.e., with criteria 3). Indeed, for any $j \in \{1, \ldots, J\}$ and $y_j \in \mathbb{N}$ we have

$$\begin{aligned} p(y_j) &= \sum_{\boldsymbol{y}_{-j}} p(\boldsymbol{y}), \\ &= \sum_{n \geq y_j} p(|\boldsymbol{y}| = n) \sum_{\boldsymbol{y}_{-j}} p_{|\boldsymbol{y}|=n}(\boldsymbol{y}), \\ p(y_j) &= \sum_{n \geq y_j} p(|\boldsymbol{y}| = n) p_{|\boldsymbol{y}|=n}(y_j), \end{aligned}$$

The marginal distribution of the singular distribution, i.e., the distribution of y_j given $|\boldsymbol{y}| = n$, is a distribution bounded by n. Its parametrization has the same form $f_j(\boldsymbol{\theta})$ for all marginals y_j given $|\boldsymbol{y}| = n$ if the singular distribution is symmetric. It implies that all marginals y_j follow the damage distribution $\mathcal{L}_n\{f_j(\boldsymbol{\theta})\} \underset{n}{\wedge} \mathcal{L}(\psi)$.

Moreover, if the sum distribution is stable under the damage process, i.e., if there exists ψ'_j such that

$$\mathcal{L}_n\{f_j(\boldsymbol{\theta})\} \underset{n}{\wedge} \mathcal{L}(\psi) = \mathcal{L}(\psi'_j),$$

then the splitting distribution $\mathcal{S}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}(\psi)$ turns out to be a multivariate extension of the given distribution $\mathcal{L}(\psi)$. We will demonstrate in Section 3 that this closure property is a sufficient condition to obtain a natural multivariate extension of $\mathcal{L}(\psi)$ (i.e. with criteria 4), in the three cases of multivariate hypergeometric, multinomial and Dirichlet multinomial splitting distributions.

2.4. Log-likelihood decomposition

If the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are unrelated, the log-likelihood of the splitting distribution, denoted by $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}; \boldsymbol{y})$, can be decomposed into log-likelihoods for the singular multivariate and the sum distributions:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}; \boldsymbol{y}) &= \log\left\{p_{|\boldsymbol{y}|}(\boldsymbol{y})\right\} + \log\{p(|\boldsymbol{y}|)\}, \\
\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}; \boldsymbol{y}) &= \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) + \mathcal{L}(\boldsymbol{\psi}; |\boldsymbol{y}|).
\end{aligned} \tag{5}
$$

Therefore, the maximum likelihood estimator (MLE) of a splitting distribution with unrelated parameters can be obtained separately using respectively the MLE of the singular distribution and the MLE of the sum distribution. Hence, using similar arguments as in [23] and under usual assumptions ensuring asymptotic normality, the respective MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are asymptotically independent. Remark that usual assumptions do not include non-singular distributions $\mathcal{S}_{\blacktriangle_n}(\boldsymbol{\theta}, \gamma)$ since n is an integer parameter and is related to the support $\blacktriangle_n$ of these distributions. Moreover, with C estimators of singular distributions and L estimators of univariate distributions, one is able to estimate $C \times L$ multivariate distributions, with time complexity in $O(C + L)$. Let us remark that decomposition (5) stays true for decomposable scores such as AIC and BIC. Model selection using decomposable scores is also reduced to two separate model selection problems and has the same linear time and space complexity. The Supplementary Materials S1 gives the definition of some beta compound distributions and recall the definition of usual power series distributions. Moreover Table 1 introduces the notations of these distributions and gives some references for inference of their parameters.

2.5. Splitting regression models

Let us consider the regression framework, with the discrete multivariate response variable $\boldsymbol{y}$ and the vector of Q explanatory variables $\boldsymbol{x} = \left(x_1, \ldots, x_p\right)$. The random vector $\boldsymbol{y}$ is said to follow a splitting regression if there exists $\boldsymbol{\psi} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\Psi}$ and $\boldsymbol{\theta} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\Theta}$ such that:

- for all $n \in \mathbb{N}$, the random vector $\boldsymbol{y}$ given $|\boldsymbol{y}| = n$ and $\boldsymbol{x}$ follows the singular regression $\mathcal{S}_{\Delta_n}\{\boldsymbol{\theta}(\boldsymbol{x})\}$.
- the sum $|\boldsymbol{y}|$ given $\boldsymbol{x}$ follow the univariate regression $\mathcal{L}\{\boldsymbol{\psi}(\boldsymbol{x})\}$.

Such a compound regression model will be denoted by

$$
\boldsymbol{y} \,|\, \boldsymbol{x} \sim \mathcal{S}_{\Delta_n}\{\boldsymbol{\theta}(\boldsymbol{x})\} \underset{n}{\wedge} \mathcal{L}\{\boldsymbol{\psi}(\boldsymbol{x})\}.
$$

The decomposition of log-likelihood (5) still holds when considering explanatory variables if parametrizations of the singular distribution and the sum distribution are unrelated. Table 7 of the Supplementary material S3, gives some references for parameter inference and variable selection adapted to four singular and six univariate regression models. We thus easily obtain $4 \times 6 = 24$ appropriate regression models for multivariate count responses. Most of them are new, since usually either the modelling of the sum is forgotten either the response components y_j are considered as independent given the explanatory variables $\boldsymbol{x}$. Variable selection can be made separately on the sum and the singular distribution.

3. Convolution splitting distributions

In order to study thoroughly the graphical models and the marginals of splitting distributions, additional assumptions are necessary concerning the parametric form of the singular distribution. Convolution splitting distributions have been introduced by Shanbhag [42] for $J = 2$ and extended by Rao and Srivastava [38] for $J \geq 2$, but were only used as a tool for characterizing univariate discrete distributions $\mathcal{L}(\psi)$. We here consider convolution splitting distributions as a general family of multivariate discrete distributions, as in [36].

3.1. Definition

The random vector $\boldsymbol{y}$ given $|\boldsymbol{y}| = n$ is said to follow a convolution distribution if there exists a non-negative parametric sequence $a := \{a_\theta(y)\}_{\theta\in\Theta, y\in\mathbb{N}}$ such that for all $\boldsymbol{y} \in \Delta_n$ we have

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

where $c_{\boldsymbol{\theta}}$ denotes the normalizing constant (i.e., the convolution of $a_{\theta_1}, \ldots, a_{\theta_J}$ over the simplex Δ_n). Note that a convolution distribution is symmetric by construction. The non-singular extension denoted by $C_{\blacktriangle_n}(a; \boldsymbol{\theta}, \gamma)$ is therefore well defined, with pmf

$$p(\boldsymbol{y}) = \frac{1}{c_{\boldsymbol{\theta},\gamma}(n)} a_\gamma(n - |\boldsymbol{y}|) \prod_{j=1}^{J} a_{\theta_j}(y_j),$$

for all $\boldsymbol{y} \in \blacktriangle_n$ (this probability is null when $\boldsymbol{y} \notin \blacktriangle_n$). If the non-singular convolution distribution is univariate then it is denoted by $C_n(a; \theta, \gamma)$. A convolution distribution is said to be additive if

$$a_\theta * a_\gamma = a_{\theta+\gamma} \tag{6}$$

for all $\theta \in \Theta$ and $\gamma \in \Theta$, where the symbol $*$ denotes the convolution, i.e., $(a_\theta * a_\gamma)(n) := \sum_{y=0}^{n} a_\theta(y) a_\gamma(n-y)$. By induction on J it is shown that the normalized constant becomes $c_{\boldsymbol{\theta}}(n) = a_{|\boldsymbol{\theta}|}(n)$. An additive convolution distribution is thus fully characterized by the parametric sequence $a = \{a_\theta(y)\}_{y\in\mathbb{N}}$ and will denoted by $C_{\Delta_n}(a; \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J) \in \Theta^J$. This additivity property will be crucial in the following to demonstrate the closure property under marginalization.

Examples:. We highlight here three examples of additive convolution distributions:

1. the multivariate hypergeometric distribution with $a_\theta(y) = \binom{\theta}{y}$ and $\Theta = \mathbb{N}^*$,
2. the multinomial distribution with $a_\theta(y) = \theta^y/y!$ and $\Theta = \mathbb{R}_+^*$,
3. the Dirichlet multinomial distribution with $a_\theta(y) = \binom{y+\theta-1}{y}$ and $\Theta = \mathbb{R}_+^*$.

The additivity of these three convolution distributions, i.e., the equation (6), can be shown using respectively the binomial theorem, the Rothe-Hagen identity and the Vandermonde identity.

3.2. Properties

The following theorem expresses some closure properties of additive convolution splitting distributions under marginalization.

Theorem 1. *Let* $\mathbf{y}$ *follow an additive convolution splitting distribution* $C_{\Delta_n}(a;) \underset{n}{\wedge} \mathcal{L}(\psi)$ *then:*

1. *The marginal sum* $|\mathbf{y}_{\mathcal{I}}|$ *follows the convolution damage distribution*

$$C_n(a; |\boldsymbol{\theta}_{\mathcal{I}}|, |\boldsymbol{\theta}_{-\mathcal{I}}|) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi}).$$

2. *The subvector* $\mathbf{y}_{\mathcal{I}}$ *given* $|\mathbf{y}_{\mathcal{I}}| = n$ *follows the singular convolution distribution* $C_{\Delta_n}(a; \boldsymbol{\theta}_{\mathcal{I}})$.
3. *The subvector* $\mathbf{y}_{\mathcal{I}}$ *follows the convolution splitting damage distribution*

$$C_{\Delta_n}(a; \boldsymbol{\theta}_{\mathcal{I}}) \underset{n}{\wedge} \left\{ C_m(a; |\boldsymbol{\theta}_{\mathcal{I}}|, |\boldsymbol{\theta}_{-\mathcal{I}}|) \underset{m}{\wedge} \mathcal{L}(\boldsymbol{\psi}) \right\}.$$

4. *The subvector* $\mathbf{y}_{\mathcal{I}}$ *given* $\mathbf{y}_{-\mathcal{I}} = \mathbf{y}_{-\mathcal{I}}$ *follows the convolution splitting truncated and shifted distribution*

$$C_{\Delta_n}(a; \boldsymbol{\theta}_{\mathcal{I}}) \underset{n}{\wedge} \left[TS_{|\mathbf{y}_{-\mathcal{I}}|} \{\mathcal{L}(\boldsymbol{\psi})\} \right].$$

5. *The subvector* $\mathbf{y}_{\mathcal{I}}$ *given* $\mathbf{y}_{\mathcal{J}} = \mathbf{y}_{\mathcal{J}}$ *follows the convolution splitting truncated and shifted damage distribution*

$$C_{\Delta_n}(a;\boldsymbol{\theta}_{\mathcal{I}}) \underset{n}{\wedge} \left[TS_{|\mathbf{y}_{\mathcal{J}}|}\left\{C_m(a;|\boldsymbol{\theta}_{\mathcal{I}\cup\mathcal{J}}|,|\boldsymbol{\theta}_{-\mathcal{I}\cup\mathcal{J}}|) \underset{m}{\wedge} \mathcal{L}(\boldsymbol{\psi})\right\}\right].$$

where $\mathcal{I} \subset \{1,\ldots,J\}$, $-\mathcal{I} = \{1,\ldots,J\}\backslash\mathcal{I}$, $\mathcal{J} \subset -\mathcal{I}$, $\mathbf{y}_{\mathcal{I}}$ *(respectively* $\mathbf{y}_{-\mathcal{I}}$ *and* $\mathbf{y}_{\mathcal{J}}$*) denote the corresponding sub-vectors and* $TS_\delta\{\mathcal{L}(\psi)\}$ *denotes the truncated and shifted distribution* $\mathcal{L}(\psi)$ *with parameter* $\delta \in \mathbb{N}$ *(i.e.,* $X \sim TS_\delta\{\mathcal{L}(\psi)\}$ *means that* $P(X = x) = P_{Z\geq\delta}(Z = \delta + x)$ *with* $Z \sim \mathcal{L}(\psi)$*).*

This theorem includes results of Janardan and Patil [20], Patil [33], Xekalaki [52] as particular cases. For instance, the case of multinomial, multinomial negative, multivariate logarithmic, multivariate hypergeometric, multivariate negative hypergeometric and multivariate generalized waring distributions are specific additive convolution splitting. The third item of the theorem is the most important, implying the two following properties.

Property 1. *An additive convolution splitting distribution* $C_{\Delta_n}(a;\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ *is a natural multivariate extension of* $\mathcal{L}(\boldsymbol{\psi})$ *if the latter is stable under the convolution damage process* $C_n(a;\theta,\gamma) \underset{n}{\wedge} (\cdot)$.

For example, it can be shown that the negative binomial distribution is stable under the binomial damage process. More precisely we have:

$$\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{NB}(r,p) = \mathcal{NB}(r,p').$$

where $p' := \frac{\pi p}{\pi p+1-p}$. The multinomial splitting negative binomial distribution is therefore stable under all marginalization and can be considered as a natural multivariate extension of the negative binomial distribution. In fact this is exactly the well-known negative multinomial. More precisely we have:

$$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{NB}(r,p) = \mathcal{NM}(r,p\cdot\boldsymbol{\pi}).$$

A specific distribution that is stable under the convolution damage process is the convolution damage itself.

Property 2. *An additive convolution damage distribution is stable under itself:*

$$C_n(a;\theta,\gamma) \underset{n}{\wedge} C_m(a;\theta+\gamma,\lambda) = C_m(a;\theta,\gamma+\lambda).$$

This result can be extended to the multivariate case to obtain the particular following identity.

Property 3. *The non-singular version of an additive convolution distribution is a specific convolution splitting distribution:*

$$C_{\Delta_n}(a;\boldsymbol{\theta}) \underset{n}{\wedge} C_m(a;|\boldsymbol{\theta}|,\gamma) = C_{\blacktriangle_m}(a;\boldsymbol{\theta},\gamma).$$

4. Multinomial splitting distributions

In this section the multinomial distribution is introduced as a positive, additive and proportional convolution distribution. Then, the general case of multinomial splitting distributions (i.e., for any sum distribution $\mathcal{L}(\psi)$) is addressed. For six specific sum distributions, the usual characteristics of multinomial splitting distributions are described in Table 2 of the paper and Table 4 of Supplementary Materials S2.

4.1. Multinomial distribution

Let $a_\theta(y) = \theta^y/y!$ be the parametric sequence that characterizes the multinomial distribution as a convolution distribution. It is positive since $\theta^y/y! > 0$ for all $\theta \in \Theta = (0,\infty)$ and all $y \in \mathbb{N}$. It is additive, as a consequence from the binomial theorem: $(\theta+\gamma)^n = \sum_{y=0}^n \binom{n}{y}\theta^y\gamma^{n-y}$. It implies, by induction on n, that the normalizing constant is $c_{\boldsymbol{\theta}}(n) = a_{|\boldsymbol{\theta}|}(n) = |\boldsymbol{\theta}|^n/n!$. The pmf of the singular multinomial distribution is thus given, for $\boldsymbol{y} \in \Delta_n$, by

$$p_{|\mathbf{y}|=n}(\boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \prod_{j=1}^{J} \left(\frac{\theta_j}{|\boldsymbol{\theta}|}\right)^{y_j}, \tag{7}$$

and is denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in (0,\infty)^J$. This convolution is proportional, implying that the equivalence class of distributions $\{\mathcal{M}_{\Delta_n}(\lambda \cdot \boldsymbol{\theta}), \lambda \in (0,\infty)\}$ can be summarized by the representative element $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ where $\boldsymbol{\pi} = \frac{1}{|\boldsymbol{\theta}|} \cdot \boldsymbol{\theta}$. The parameters vector $\boldsymbol{\pi}$ lies in the continuous simplex $\Delta := \{\boldsymbol{\pi} \in (0,1)^J : |\boldsymbol{\pi}| = 1\}$ and the pmf reduces to its usual form, given by Johnson et al. [22]. The pmf of the non-singular multinomial distribution, denoted by $\mathcal{M}_{\blacktriangle_n}(\boldsymbol{\theta}, \gamma)$, is given by

$$p(\boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \left(\frac{\gamma}{|\boldsymbol{\theta}| + \gamma}\right)^{n-|\boldsymbol{y}|} \prod_{j=1}^{J} \left(\frac{\theta_j}{|\boldsymbol{\theta}| + \gamma}\right)^{y_j},$$

for $\boldsymbol{y} \in \blacktriangle_n$. In the same way there exists a representative element $\mathcal{M}_{\blacktriangle_n}(\boldsymbol{\pi}^*, \gamma^*)$ with $(\boldsymbol{\pi}^*, \gamma^*) \in (0,1)^{J+1}$ such that $|\boldsymbol{\pi}^*| + \gamma^* = 1$. Given this constraint, the last parameter $\gamma^* = 1 - |\boldsymbol{\pi}^*|$ could be let aside to ease the notation and obtain $\mathcal{M}_{\blacktriangle_n}(\boldsymbol{\pi}^*)$ where the parameters vector $\boldsymbol{\pi}^*$ lies in the continuous corner of the open hypercube $\blacktriangle = \{\boldsymbol{\pi}^* \in (0,1)^J : |\boldsymbol{\pi}^*| < 1\}$. As a particular case of the non-singular multinomial distribution (when $J = 1$), the binomial distribution is finally denoted by $\mathcal{B}_n(p)$ with $p \in (0,1)$ (which is also the representative element of its class). Even if this new definition of multinomial distributions based on equivalence classes seems somehow artificial, this is necessary to obtain all the properties that hold for convolution splitting distributions. For instance Property 3 becomes the following result (with representative element notations).

Corollary 1. *The multinomial splitting binomial distribution is exactly the non-singular multinomial distribution:*

$$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{B}_m(p) = \mathcal{M}_{\blacktriangle_m}(p \cdot \boldsymbol{\pi}).$$

We wish to highlight the following significant point regarding the difference between singular and non-singular multinomial distributions. Contrarily to the widely held view that the multinomial distribution is the extension of the binomial distribution [22], only the non-singular one should be considered as the natural extension. In fact, criterion 4 does not hold for the singular multinomial distribution (multivariate marginals follow non-singular multinomial distributions). Moreover, when confronted to multivariate counts, usual inference of multinomial distributions [22, 55] is that of singular multinomial distributions such that $\forall n \in \mathbb{N}$ the random vector $\boldsymbol{y}$ given $|\boldsymbol{y}| = n$ follows $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$. Such a point of view therefore limits the possibility of comparing these distributions to other classical discrete multivariate distributions such as the negative multinomial distribution or the multivariate Poisson distributions [24] used for modeling the joint distribution of $\boldsymbol{y}$. The singular multinomial distribution should thus not be considered as a J-multivariate distribution since criterion 1 would not hold.

4.2. Properties of multinomial splitting distributions

Let $\boldsymbol{y}$ follow a multinomial splitting distribution $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\psi)$. Criteria 1 and 3 hold, as a consequence from positivity and symmetry. The pmf is given by

$$p(\boldsymbol{y}) = p(|\boldsymbol{y}|) \binom{|\boldsymbol{y}|}{\boldsymbol{y}} \prod_{j=1}^{J} \pi_j^{y_j}, \tag{8}$$

for $\boldsymbol{y} \in \mathbb{N}^J$. According to (2) and (3), the expectation and covariance of multinomial splitting distributions are given by

$$\mathrm{E}(\boldsymbol{y}) = \mu_1 \cdot \boldsymbol{\pi}, \tag{9}$$

$$\mathrm{Cov}(\boldsymbol{y}) = \mu_1 \cdot \mathrm{diag}(\boldsymbol{\pi}) + (\mu_2 - \mu_1^2) \cdot \boldsymbol{\pi}\boldsymbol{\pi}^t, \tag{10}$$

where $\boldsymbol{\pi}^t$ denotes the transposition of the vector $\boldsymbol{\pi}$ and μ_i denotes the factorial moments of order i ($i = 1, 2$) for the sum distribution. Moreover, according to (4) we obtain the pgf of multinomial splitting distributions as

$$G(\boldsymbol{s}) = \mathrm{E}_y \left\{(\boldsymbol{\pi}^t \boldsymbol{s})^{|\boldsymbol{y}|}\right\} = G_\psi\left(\boldsymbol{\pi}^t \boldsymbol{s}\right), \tag{11}$$

where G_ψ denote the pgf of the sum distribution. The graphical model is characterized by the following property.

(a)

Distribution	$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{B}_m(p)$
Re-parametrization	$\mathcal{M}_{\blacktriangle_m}(p \cdot \boldsymbol{\pi})$
Supp($\boldsymbol{y}$)	$\blacktriangle_m$
$p(\boldsymbol{y})$	$\binom{n}{\boldsymbol{y}}(1-p)^{n-\lvert\boldsymbol{y}\rvert} \prod_{j=1}^{J}(p\pi_j)^{y_j}$
E($\boldsymbol{y}$)	$mp \cdot \boldsymbol{\pi}$
Cov($\boldsymbol{y}$)	$mp \cdot \{\text{diag}(\boldsymbol{\pi}) - p \cdot \boldsymbol{\pi}\boldsymbol{\pi}^t\}$
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$(1 - p + p\, \boldsymbol{\pi}^t \boldsymbol{s})^m$
Marginals	$y_j \sim \mathcal{B}_m(p\pi_j)$ [1]

(b)

Distribution	$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{NB}(r, p)$
Re-parametrization	$\mathcal{NM}(r, p \cdot \boldsymbol{\pi})$
Supp($\boldsymbol{y}$)	$\mathbb{N}^J$
$p(\boldsymbol{y})$	$\binom{\lvert\boldsymbol{y}\rvert+r-1}{\boldsymbol{y}}(1-p)^r \prod_{j=1}^{J}(p\pi_j)^{y_j}$
E($\boldsymbol{y}$)	$r\frac{p}{1-p} \cdot \boldsymbol{\pi}$
Cov($\boldsymbol{y}$)	$r\frac{p}{1-p} \cdot \left\{\text{diag}(\boldsymbol{\pi}) + \frac{p}{1-p} \cdot \boldsymbol{\pi}\boldsymbol{\pi}^t\right\}$
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\left(\frac{1-p}{1-p\boldsymbol{\pi}^t\boldsymbol{s}}\right)^r$
Marginals	$y_j \sim \mathcal{NB}(r, p\pi_j)$ [1]

(c)

Distribution	$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(p)$
Re-parametrization	$\mathcal{ML}(p \cdot \boldsymbol{\pi})$
Supp($\boldsymbol{y}$)	$\mathbb{N}^J \setminus (0, \ldots, 0)$
$p(\boldsymbol{y})$	$\binom{\lvert\boldsymbol{y}\rvert}{\boldsymbol{y}} \frac{-1}{\lvert\boldsymbol{y}\rvert \ln(1-p)} \prod_{j=1}^{J} \pi_j^{y_j}$
E($\boldsymbol{y}$)	$\frac{-p}{(1-p)\ln(1-p)} \cdot \boldsymbol{\pi}$
Cov($\boldsymbol{y}$)	$\frac{-p}{(1-p)\ln(1-p)} \cdot \left\{\text{diag}(\boldsymbol{\pi}) + \frac{p\{1-\ln(1-p)\}}{(1-p)\ln(1-p)} \cdot \boldsymbol{\pi}\boldsymbol{\pi}^t\right\}$
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\frac{\ln(1-p\cdot\boldsymbol{\pi}^t\boldsymbol{s})}{\ln(1-p)}$
Marginals	$y_j \sim \mathcal{L}\left(p'_j, \omega_j\right)$ [2]

Table 2: Characteristics of multinomial splitting (a) binomial, (b) negative binomial and (c) logarithmic series distribution.

Theorem 2 (Peyhardi and Fernique [36])**.** *The minimal graphical model for a multinomial splitting distribution* $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\psi)$ *is:*

- *empty if* $\mathcal{L}(\psi) = \mathcal{P}(\lambda)$ *for some* $\lambda > 0$*,*
- *complete otherwise.*

Therefore, all multinomial splitting distributions are *sensu stricto* multivariate distributions (criterion 2) except when the sum follows a Poisson distribution. As a consequence from additivity, Theorem 1 holds and yields the marginals:

Corollary 2. *Assume that* $\boldsymbol{y}$ *follows a multinomial splitting distribution* $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ *and denote by* $G_{\psi}^{(y)}$ *is the y-th derivative of the pgf* G_{ψ}*. Then, the marginals* y_j *follow the binomial damage distribution* $\mathcal{B}_n(\pi_j) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ *and*

$$p(y_j) = \frac{\pi_j^{y_j}}{y_j!} G_{\psi}^{(y_j)}(1 - \pi_j), \tag{12}$$

Using equation (12), it is easy to study the stability of power series distributions under the binomial damage process.

Property 4. *The binomial, Poisson, negative binomial and zero modified logarithmic series distributions are stable under the binomial damage process. More precisely we have*

1. $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{B}_m(p) = \mathcal{B}_m(\pi p)$
2. $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{P}(\lambda) = \mathcal{P}(\pi\lambda)$
3. $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{NB}(r, p) = \mathcal{NB}(r, p'),$ *where* $p' := \frac{\pi p}{\pi p + 1 - p}$
4. $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(p, \omega) = \mathcal{L}(p', \omega'),$ *where* $p' := \frac{\pi p}{\pi p + 1 - p}$ *and* $\omega' := \omega - \ln(\pi p + 1 - p)$*.*

Assume that $\mathcal{L}(\psi)$ is a power series distribution denoted by $PSD\{g(\alpha)\}$. It can be seen by identifiability that the resulting splitting distribution of $\boldsymbol{y}$ is exactly the multivariate sum-symmetric power series distribution (MSSPSD) introduced by Patil [33], i.e., $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} PSD\{g(\alpha)\} = \text{MSSPSD}\{\alpha \cdot \boldsymbol{\pi}\}$. The non-singular multinomial distribution, the negative multinomial distribution and the multivariate logarithmic series distribution are thereby encompassed in multinomial splitting distributions (see Table 2).

Assume now that $\mathcal{L}(\psi)$ is a standard beta compound distribution. We obtain three new multivariate distributions, which are multivariate extensions of the non-standard beta binomial, non-standard beta negative binomial and beta Poisson distributions (see Table 4 of Supplementary Material S2 for details about these three multivariate distributions and Supplementary Material S1 for definitions of the non-standard beta binomial and the non-standard beta negative binomial distributions). All the characteristics of these six multinomial splitting distributions (pmf, expectation, covariance, pgf and marginal distributions) have been calculated using equations (8), (9), (10), (11), (12) according to the sum distribution $\mathcal{L}(\psi)$.

4.3. Asymptotic and independence properties of estimators

Firstly, maximum likelihood estimation in multinomial splitting distributions (MSDs) $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ is considered. Let $\mathcal{Y} := (\boldsymbol{y}_i)_{1\leq i\leq N}$ denote an independent and identically distributed (i.i.d.) sample with size N and distribution $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ and $|\mathcal{Y}| := (|\boldsymbol{y}_i|)_{1\leq i\leq N}$ denote the corresponding i.i.d. sample of sums. As a consequence from the log-likelihood decomposition property (5), $\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\psi}; \mathcal{Y})$ writes as $\mathcal{L}(\boldsymbol{\pi}; \mathcal{Y}) + \mathcal{L}(\boldsymbol{\psi}; |\mathcal{Y}|)$. Hence, the sum distribution parameters $\boldsymbol{\psi}$ and the probability parameters $\boldsymbol{\pi}$ can be separately estimated. Computation of the MLE $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ is then equivalent to MLE computation in the statistical model associated with i.i.d. sample $|\mathcal{Y}|$ distributed according to $\mathcal{L}(\psi)$. The asymptotic properties of $\hat{\boldsymbol{\psi}}$ are inherited from the statistical model associated with $\bigotimes_{i=1}^{N} \mathcal{L}_{\psi}$, the rate of convergence being determined by the sample size N. For any $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, J\}$, let $y_{i,j}$ denote the component j in $\boldsymbol{y}_i$ and let z denote the total count $\sum_{i=1}^{N} |\boldsymbol{y}_i| = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{i,j}$. It is proved straightforwardly as in i.i.d. samples from multinomial distributions that for any $j \in \{1, \ldots, J\}$, that $\hat{\pi}_j$ has the following closed-formed expression:

$$\hat{\pi}_j = \frac{\sum_{i=1}^{N} y_{i,j}}{z}. \tag{13}$$

Moreover, $\hat{\boldsymbol{\pi}}$ satisfies the following limit central theorem [used in the proof of the Pearson chi-square test, see 35, 44]. Specifically, there exists some deterministic orthonormal family $(u_1, \ldots, u_{J-1})$ of $\mathbb{R}^J$ and i.i.d. centered, standardized random Gaussian variables $(\xi_1, \ldots, \xi_{J-1})$ such that given z,

$$\sqrt{z}\left[\frac{\hat{\pi}_1 - \pi_1}{\sqrt{\pi_1}}, \ldots, \frac{\hat{\pi}_J - \pi_J}{\sqrt{\pi_J}}\right] \xrightarrow[z\to\infty]{\mathcal{D}} \sum_{k=1}^{J-1} \xi_k u_k.$$

Moreover, the following results hold:

Theorem 3. *The estimator $\hat{\boldsymbol{\psi}}$ of sum parameters and the estimator $\hat{\boldsymbol{\pi}}$ of components proportions are two independent random vectors given the total count z.*

Secondly, Bayesian estimation in MSDs is considered. It follows from theorem 3 that $p\left(|\mathcal{Y}| \,\middle|\, \mathcal{Y}, \boldsymbol{\psi}, \boldsymbol{\pi}\right) = p\left(|\mathcal{Y}| \,\middle|\, Z, \boldsymbol{\pi}\right)$. As a consequence, Bayesian inference with independent priors for $\boldsymbol{\psi}$ and $\boldsymbol{\pi}$ leads to some factorization property of the joint posterior. Indeed, let denote $\mathcal{S} = \left(\sum_{i=1}^{n} Y_{i,j}\right)_{1\leq j\leq J}$; if $p(\boldsymbol{\psi}, \boldsymbol{\pi}) = p(\boldsymbol{\psi})p(\boldsymbol{\pi})$, then

$$\begin{aligned} p(\boldsymbol{\psi}, \boldsymbol{\pi} \,|\, \mathcal{Y}) &\propto p(\mathcal{Y} \,|\, \boldsymbol{\psi}, \boldsymbol{\pi})\, p(\boldsymbol{\psi})p(\boldsymbol{\pi}) \\ &\propto p\left(|\mathcal{Y}|, \mathcal{S} \oslash |\mathcal{Y}| \,\middle|\, \boldsymbol{\psi}, \boldsymbol{\pi}\right) p(\boldsymbol{\psi})p(\boldsymbol{\pi}), \end{aligned}$$

where $\oslash$ refers to Hadamard division and is conventionally defined as 0 if $\sum_{i=1}^{n} Y_{i,j} = 0$. Thus,

$$\begin{aligned} p(\boldsymbol{\psi}, \boldsymbol{\pi} \,|\, \mathcal{Y}) &\propto p\left(\mathcal{S} \oslash |\mathcal{Y}| \,\middle|\, |\mathcal{Y}|, \boldsymbol{\psi}, \boldsymbol{\pi}\right) p\left(|\mathcal{Y}| \,\middle|\, \boldsymbol{\psi}, \boldsymbol{\pi}\right) p(\boldsymbol{\psi})p(\boldsymbol{\pi}) \\ &\propto p\left(\mathcal{S} \oslash |\mathcal{Y}| \,\middle|\, Z, \boldsymbol{\pi}\right) p\left(|\mathcal{Y}| \,\middle|\, \boldsymbol{\psi}\right) p(\boldsymbol{\psi})p(\boldsymbol{\pi}) \end{aligned}$$

by (5) and theorem 3. Hence,

$$p\left(\boldsymbol{\psi},\boldsymbol{\pi}\,\middle|\,\mathcal{Y}\right) \propto p\left(\mathcal{S}\oslash|\mathcal{Y}|\,\middle|\,Z,\boldsymbol{\pi}\right)p(\boldsymbol{\pi})p\left(\boldsymbol{\psi}\,\middle|\,|\mathcal{Y}|\right)$$

where from the proof of Theorem 3, $\mathcal{S}\oslash|\mathcal{Y}|$ has distribution $\mathcal{M}(Z,\boldsymbol{\pi})$ given $(Z,\boldsymbol{\pi})$, up to the scaling factor Z. Note that this results extends that of lemma 1 in [45] with Poisson-distributed $(Y_{i,j})_{1\leq i\leq n;1\leq j\leq J}$.

In the particular case where $p(\boldsymbol{\pi})$ is chosen as a Dirichlet distribution $D(\alpha_1,\dots,\alpha_J)$, then the marginal posterior distribution $p\,(\boldsymbol{\pi}\,|\,\mathcal{Y})$ is Dirichlet $\mathcal{D}(\alpha_1+\mathcal{S}_0,\dots,\alpha_J+\mathcal{S}_J)$ [see 39, Chapter 3], which does not depend on Z, and

$$p\,(\boldsymbol{\psi},\boldsymbol{\pi}\,|\,\mathcal{Y}) \propto p\,(\boldsymbol{\pi}|\mathcal{S})\,p\left(\boldsymbol{\psi}\,\middle|\,|\mathcal{Y}|\right).$$

Thus, parameters $\boldsymbol{\psi}$ and $\boldsymbol{\pi}$ are independent a posteriori. Moreover, if $(\mathcal{L}_{\boldsymbol{\psi}})_{\boldsymbol{\psi}}$ is in the exponential family, its expression has the form $\mathcal{L}_{\boldsymbol{\psi}}(x)=h(x)e^{\psi x-\phi(x)}$ and a conjugate family of priors is given by

$$p(\boldsymbol{\psi}|\mu,\lambda)=\rho(\mu,\lambda)\exp\left(\psi\mu-\lambda\phi(\psi)\right),$$

where μ and $\lambda>0$ are hyperparameters and $\rho(\mu,\lambda)$ is a normalizing constant [see 39, Chapter 3]. Then the marginal posterior distribution of $\boldsymbol{\psi}$ is

$$p(\boldsymbol{\psi}\,\big|\,|\mathcal{Y}|,\mu,\lambda)=\rho(\mu+z,\lambda+n)\exp\left\{\phi\,(\mu+z)-(\lambda+n)\,\phi\,(\psi)\right\}.$$

4.4. *Generalized linear models for multivariate count responses*

Multinomial splitting distributions offer an appropriate framework for describing GLMs for multivariate count responses. Let $\boldsymbol{x}=(x_1,\dots,x_p)$ denote the vector of explanatory variables. If both singular and sum distributions are described by GLMs then the resulted splitting regression well defines a GLM for multivariate count responses. This is a consequence of the splitting decomposition of probabilities $p_{\boldsymbol{x}}(\boldsymbol{y})=p_{|\boldsymbol{y}|,\boldsymbol{x}}(\boldsymbol{y})p_{\boldsymbol{x}}(|\boldsymbol{y}|)$ and the exponential property $\exp(a+b)=\exp(a)\exp(b)$. The only known singular GLM for count response is the multinomial GLM. For the univariate case, the binomial, Poisson and negative binomial are defined in the GLM framework. Assume that $\boldsymbol{y}|\boldsymbol{x}\sim\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}\underset{n}{\wedge}\mathcal{L}\{\psi(\boldsymbol{x})\}$, where ψ is the canonical parameter of the univariate GLM. Then we have

$$\begin{aligned}
p_{\boldsymbol{x}}(\boldsymbol{y}) &= \exp\left\{\sum_{j=1}^{J}y_j\ln\pi_j+\ln\binom{|\boldsymbol{y}|}{\boldsymbol{y}}\right\}\exp\left\{\frac{|\boldsymbol{y}|\psi-b(\psi)}{\phi}+c(|\boldsymbol{y}|;\phi)\right\}\\
&= \exp\left\{\sum_{j=1}^{J}y_j\left(\ln\pi_j+\frac{\psi}{\phi}\right)-\frac{b(\psi)}{\phi}+\ln\binom{|\boldsymbol{y}|}{\boldsymbol{y}}+c(|\boldsymbol{y}|;\phi)\right\}\\
p_{\boldsymbol{x}}(\boldsymbol{y}) &= \exp\left\{\boldsymbol{y}^T\boldsymbol{\theta}-B(\boldsymbol{\theta})+C(\boldsymbol{y};\phi)\right\}
\end{aligned}$$

where $\theta_j:=\ln\pi_j+\frac{\psi}{\phi}$, $B(\boldsymbol{\theta}):=\frac{b\left[\phi\ln\left\{\sum_{j=1}^{J}\exp(\theta_j)\right\}\right]}{\phi}$ and $C(\boldsymbol{y};\phi)=\ln\binom{|\boldsymbol{y}|}{\boldsymbol{y}}+c(|\boldsymbol{y}|;\phi)$. The Poisson GLM for the sum is a particular case since the components $y_1,\dots,y_J$ are independent given the explanatory variables $\boldsymbol{x}$ in this case

$$\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}\underset{n}{\wedge}\mathcal{P}\{\lambda(\boldsymbol{x})\}=\bigotimes_{j=1}^{J}P\{\lambda(\boldsymbol{x})\pi_j(\boldsymbol{x})\}.$$

Therefore only binomial and negative binomial GLMs for the sum allow us to define multivariate GLMs with dependencies. It turns out to be respectively the non-singular multinomial GLM

$$\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}\underset{n}{\wedge}\mathcal{B}_m\{p(\boldsymbol{x})\}=\mathcal{M}_{\blacktriangle_n}\{p(\boldsymbol{x})\cdot\boldsymbol{\pi}(\boldsymbol{x})\},$$

and the negative multinomial GLM

$$\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}\underset{n}{\wedge}\mathcal{NB}\{r,p(\boldsymbol{x})\}=\mathcal{NM}\{r,p(\boldsymbol{x})\cdot\boldsymbol{\pi}(\boldsymbol{x})\}.$$

The non-singular multinomial GLM is *sensu stricto* a GLM since criterion 1 holds, contrarily to the usual (singular) multinomial GLM. Compared to the usual multinomial and negative GLMs multinomial, used for instance by Zhang and Zhou [54], our versions offer several advantages. First, estimation can be made separately on the sum and splitting. Secondly, the variety of link functions described on $\boldsymbol{\pi}(\boldsymbol{x})$ for multinomial GLMs [37, 47] can thus be used to introduce several new link functions in GLMs for multivariate count responses. It is also possible to multiply the number of models by using different link functions on $p(\boldsymbol{x})$ for negative binomial GLMs; see [17]. Note that the choice of the link function on $\boldsymbol{\pi}(\boldsymbol{x})$ is related to the symmetry of the resulting splitting GLM. Only the canonical link function (i.e., the multinomial logit link) implies the symmetry of the splitting GLM; see [37] for details about invariance properties of categorical regression models. Finally asymptotic independence between MLEs of regression parameters for $\boldsymbol{\pi}(x)$ and $p(x)$ holds under usual assumptions for GLMs described by [12].

5. Dirichlet Multinomial Splitting distributions

In this section the Dirichlet multinomial distribution is introduced as a positive and additive convolution distribution. Then, the general case of Dirichlet multinomial splitting distributions is studied. For six specific sum distributions, the usual characteristics of Dirichlet multinomial splitting distributions are described in Tables 3 and 4 of the paper and Tables 5 and 6 of Supplementary Materials S2. Finally, the canonical case of beta binomial sum distribution is detailed, with particular emphasis on parameter inference. This family corresponds to the Polya share and sum family of distributions studied by [23].

5.1. Dirichlet multinomial distribution

Let $a_\theta(y) = \binom{y+\theta-1}{y}$ be the parametric sequence that characterizes the Dirichlet multinomial distribution as a convolution distribution. It is positive since $\binom{y+\theta-1}{y} > 0$ for all $\theta \in \Theta = (0,\infty)$ and all $y \in \mathbb{N}$. It is additive, as a consequence from the convolution identity of Hagen and Rothe: $\binom{n+\theta+\gamma-1}{n} = \sum_{y=0}^{n} \binom{y+\theta-1}{y}\binom{n-y+\gamma-1}{n-y}$. It implies, by induction on n, that the normalizing constant is $c_{\boldsymbol{\theta}}(n) = a_{|\boldsymbol{\theta}|}(n) = \binom{n+|\boldsymbol{\theta}|-1}{n}$. In order to respect the usual notation, parameter $\boldsymbol{\alpha}$ will be used instead of $\boldsymbol{\theta}$, and thus the Dirichlet multinomial distribution will be denoted by $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha})$ with $n \in \mathbb{N}$ and $\boldsymbol{\alpha} \in (0,\infty)^J$. The non-singular Dirichlet multinomial distribution will be denoted by $\mathcal{DM}_{\blacktriangle_n}(\boldsymbol{\alpha}, b)$ with $b \in (0,\infty)$. The beta binomial distribution will be denoted by $\beta\mathcal{B}_n(a,b)$ with $(a,b) \in (0,\infty)^2$. Using Property 3 with $a_\theta(y) = \binom{y+\theta-1}{y}$ and $\boldsymbol{\theta} = \boldsymbol{\alpha}$ we obtain the following result.

Corollary 3. *the Dirichlet multinomial splitting beta binomial distribution with the specific constraint $a = |\boldsymbol{\alpha}|$ is exactly the non-singular Dirichlet multinomial distribution:*

$$\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \beta\mathcal{B}_m(|\boldsymbol{\alpha}|, b) = \mathcal{DM}_{\blacktriangle_m}(\boldsymbol{\alpha}, b).$$

For similar reasons as in the multinomial case, the non-singular Dirichlet multinomial distribution should be considered as the natural extension of the beta binomial distribution, rather than the singular one. Let us remark that the Dirichlet multinomial distribution turns out to be the multivariate negative hypergeometric distribution if $\boldsymbol{\alpha} \in \mathbb{N}^J$ instead of $(0,\infty)^J$.

5.2. Properties of Dirichlet multinomial splitting distributions

Let $\boldsymbol{y}$ follow a Dirichlet multinomial splitting distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{L}(\psi)$. Criteria 1 and 3 hold, as a consequence of positivity and symmetry. The pmf is given, for $\boldsymbol{y} \in \mathbb{N}^J$, by

$$p(\boldsymbol{y}) = \frac{p(|\boldsymbol{y}|)}{\binom{|\boldsymbol{y}|+|\boldsymbol{\alpha}|-1}{|\boldsymbol{y}|}} \prod_{j=1}^{J} \binom{y_j + \alpha_j - 1}{y_j}. \tag{14}$$

According to (2) and (3), the expectation and covariance of Dirichlet multinomial splitting distributions are given by

$$\mathrm{E}(\boldsymbol{y}) = \frac{\mu_1}{|\boldsymbol{\alpha}|} \cdot \boldsymbol{\alpha}, \tag{15}$$

$$\mathrm{Cov}(\boldsymbol{y}) = \frac{1}{|\boldsymbol{\alpha}|(|\boldsymbol{\alpha}|+1)} \cdot \left[\{(|\boldsymbol{\alpha}|+1)\mu_1 + \mu_2\} \cdot \mathrm{diag}(\boldsymbol{\alpha}) + \left\{\mu_2 - \frac{|\boldsymbol{\alpha}|+1}{|\boldsymbol{\alpha}|}\mu_1^2\right\} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t\right]. \tag{16}$$

(a)

Distribution	$\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \beta\mathcal{B}_m(a,b)$
Supp($\boldsymbol{y}$)	$\blacktriangle_m$
$p(\boldsymbol{y})$	$\binom{m}{\lfloor\boldsymbol{y}\rfloor}\frac{B(a+\lfloor\boldsymbol{y}\rfloor,b+m-\lfloor\boldsymbol{y}\rfloor)}{B(a,b)}\frac{\prod_{j=1}^{J}\binom{y_j+\alpha_j-1}{y_j}}{\binom{m+\lvert\boldsymbol{\alpha}\rvert-1}{m}}$
E($\boldsymbol{y}$)	$\frac{ma}{\lvert\boldsymbol{\alpha}\rvert(a+b)}\cdot\boldsymbol{\alpha}$
Cov($\boldsymbol{y}$)	$\frac{ma}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)(a+b)}\cdot\left[\left\{\frac{b(a+b+m)}{(a+b)(a+b+1)}+\frac{ma}{a+b}+\lvert\boldsymbol{\alpha}\rvert\right\}\cdot\mathrm{diag}(\boldsymbol{\alpha})+\left\{\frac{b(a+b+m)}{(a+b)(a+b+1)}-\frac{ma}{\lvert\boldsymbol{\alpha}\rvert(a+b)}-1\right\}\cdot\boldsymbol{\alpha}\boldsymbol{\alpha}^t\right]$
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\frac{(b)_m}{(a+b)_m}\,{}_2^JF_2\{(-m,a);\boldsymbol{\alpha};(-b-m+1,\lvert\boldsymbol{\alpha}\rvert);\boldsymbol{s}\}$
Marginals	$y_j \sim \beta^2\mathcal{B}_m\left(\alpha_j,\lvert\boldsymbol{\alpha}_{-j}\rvert,a,b\right)$

(b)

Distribution	$\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \beta\mathcal{B}_m(a,b)$
Constraint	$a=\lvert\boldsymbol{\alpha}\rvert$
Re-parametrization	$\mathcal{DM}_{\blacktriangle_m}(\boldsymbol{\alpha},b)$
Supp($\boldsymbol{y}$)	$\blacktriangle_m$
$p(\boldsymbol{y})$	$\binom{m-\lfloor\boldsymbol{y}\rfloor+b-1}{m-\lfloor\boldsymbol{y}\rfloor}\frac{\prod_{j=1}^{J}\binom{y_j+\alpha_j-1}{y_j}}{\binom{m+\lvert\boldsymbol{\alpha}\rvert+b-1}{m}}$
E($\boldsymbol{y}$)	$\frac{m\lvert\boldsymbol{\alpha}\rvert}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+b)}\cdot\boldsymbol{\alpha}$
Cov($\boldsymbol{y}$)	$\frac{m}{(\lvert\boldsymbol{\alpha}\rvert+1)(\lvert\boldsymbol{\alpha}\rvert+b)}\cdot\left[\left\{\frac{b(\lvert\boldsymbol{\alpha}\rvert+b+m)}{(\lvert\boldsymbol{\alpha}\rvert+b)(\lvert\boldsymbol{\alpha}\rvert+b+1)}+\frac{m\lvert\boldsymbol{\alpha}\rvert}{\lvert\boldsymbol{\alpha}\rvert+b}+\lvert\boldsymbol{\alpha}\rvert\right\}\cdot\mathrm{diag}(\boldsymbol{\alpha})+\left\{\frac{b(\lvert\boldsymbol{\alpha}\rvert+b+m)}{(\lvert\boldsymbol{\alpha}\rvert+b)(\lvert\boldsymbol{\alpha}\rvert+b+1)}-\frac{m}{\lvert\boldsymbol{\alpha}\rvert+b}-1\right\}\cdot\boldsymbol{\alpha}\boldsymbol{\alpha}^t\right]$
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\frac{(b)_m}{(\lvert\boldsymbol{\alpha}\rvert+b)_m}\,{}_1^JF_1(-m;\boldsymbol{\alpha};-b-m+1;\boldsymbol{s})$
Marginals	$y_j \sim \beta\mathcal{B}_m\left(\alpha_j,\lvert\boldsymbol{\alpha}_{-j}\rvert+b\right)$

Table 3: Usual characteristics of Dirichlet multinomial splitting standard beta binomial distribution (respectively (a) without constraint and (b) with constraint $a=\lvert\alpha\rvert$).

The pgf of a Dirichlet multinomial splitting distribution is given by

$$G(\boldsymbol{s}) = \sum_{\boldsymbol{y}\in\mathbb{N}^J}\Gamma(\lvert\boldsymbol{y}\rvert+1)\,p(\lvert\boldsymbol{y}\rvert)\,\frac{\prod_{j=1}^{J}(\alpha_j)_{y_j}}{(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}}\prod_{j=1}^{J}\frac{s_j^{y_j}}{y_j!}. \tag{17}$$

The graphical model is characterized by the following property.

Theorem 4 (Peyhardi and Fernique [36]). *The minimal graphical model for a Dirichlet multinomial splitting distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{L}(\psi)$ is:*

- *empty if $\mathcal{L}(\psi)=\mathcal{NB}(\lvert\boldsymbol{\alpha}\rvert,p)$ for some $p\in(0,1)$,*
- *complete otherwise.*

Therefore, all Dirichlet multinomial splitting distribution are *senso stricto* multivariate distributions except when the sum follows a negative binomial distribution $\mathcal{NB}(r,p)$ with the specific constraint $r=\lvert\boldsymbol{\alpha}\rvert$.

Corollary 4. *Let $\boldsymbol{y}$ follow a Dirichlet multinomial splitting distribution, $\boldsymbol{y}\sim\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})$ with $\boldsymbol{\alpha}\in(0,\infty)^J$. Then, the marginals follow the binomial damage compound by a beta distribution*

$$y_j \sim \left\{\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi})\right\} \underset{\pi}{\wedge} \beta(\alpha_j,\lvert\boldsymbol{\alpha}_{-j}\rvert). \tag{18}$$

Therefore, results previously obtained for the binomial damage distributions can be used to describe the beta-binomial damage distributions. Assume that $\mathcal{L}(\psi)$ is a standard beta compound distribution. Four new and two already known multivariate distributions are obtained or recovered. In particular, natural multivariate extensions of three beta compound distributions are described. The non-singular Dirichlet multinomial is recovered when $\mathcal{L}(\psi)=\beta\mathcal{B}_n(a,b)$ with

(a)

Distribution	$\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \beta\mathcal{NB}(r,a,b)$
Supp($\boldsymbol{y}$)	$\mathbb{N}^J$
$p(\boldsymbol{y})$	$\frac{(a)_r}{(a+b)_r} \frac{(r)_{\lvert\boldsymbol{y}\rvert}(b)_{\lvert\boldsymbol{y}\rvert}}{(r+a+b)_{\lvert\boldsymbol{y}\rvert}(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^{J} \frac{(\alpha_j)_{y_j}}{y_j!}$
E($\boldsymbol{y}$)	$\frac{rb}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)(a-1)}\boldsymbol{\alpha}$
Cov($\boldsymbol{y}$)	$\frac{rb}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)(a-1)} \cdot \left[\left\{ \frac{(r+a-1)(a+b-1)}{(a-1)(a-2)} + \frac{rb}{a-1} + \lvert\boldsymbol{\alpha}\rvert \right\} \cdot \mathrm{diag}(\boldsymbol{\alpha}) + \left\{ \frac{(r+a-1)(a+b-1)}{(a-1)(a-2)} - \frac{rb}{\lvert\boldsymbol{\alpha}\rvert(a-1)} - 1 \right\} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \right]$ [3]
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\frac{(a)_r}{(a+b)_r} {}_2^J F_2\{(r,b);\boldsymbol{\alpha};(r+a+b,\lvert\boldsymbol{\alpha}\rvert);\boldsymbol{s}\}$
Marginals	$y_j \sim \beta^2\mathcal{NB}\left(r,\alpha_j,\lvert\boldsymbol{\alpha}_{-j}\rvert,a,b\right)$

(b)

Distribution	$\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \beta\mathcal{NB}(r,a,b)$
Constraint	$r = \lvert\boldsymbol{\alpha}\rvert$
Re-parametrization	MGWD($b,\boldsymbol{\alpha},a$)
Supp($\boldsymbol{y}$)	$\mathbb{N}^J$
$p(\boldsymbol{y})$	$\frac{(a)_{\lvert\boldsymbol{\alpha}\rvert}}{(a+b)_{\lvert\boldsymbol{\alpha}\rvert}} \frac{(b)_{\lvert\boldsymbol{y}\rvert}}{(\lvert\boldsymbol{\alpha}\rvert+a+b)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^{J} \frac{(\alpha_j)_{y_j}}{y_j!}$
E($\boldsymbol{y}$)	$\frac{b}{(\lvert\boldsymbol{\alpha}\rvert+1)(a-1)}\boldsymbol{\alpha}$
Cov($\boldsymbol{y}$)	$\frac{b}{(\lvert\boldsymbol{\alpha}\rvert+1)(a-1)} \cdot \left[\left\{ \frac{(\lvert\boldsymbol{\alpha}\rvert+a-1)(a+b-1)}{(a-1)(a-2)} + \frac{\lvert\boldsymbol{\alpha}\rvert b}{a-1} + \lvert\boldsymbol{\alpha}\rvert \right\} \cdot \mathrm{diag}(\boldsymbol{\alpha}) + \left\{ \frac{(\lvert\boldsymbol{\alpha}\rvert+a-1)(a+b-1)}{(a-1)(a-2)} - \frac{b}{a-1} - 1 \right\} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \right]$ [4]
$G_{\boldsymbol{y}}(\boldsymbol{s})$	$\frac{(a)_{\lvert\boldsymbol{\alpha}\rvert}}{(a+b)_{\lvert\boldsymbol{\alpha}\rvert}} {}_1^J F_1(b;\boldsymbol{\alpha};\lvert\boldsymbol{\alpha}\rvert+a+b;\boldsymbol{s})$
Marginals	$y_j \sim \beta\mathcal{NB}\left(\alpha_j,a,b\right)$ [1]

Table 4: Usual characteristics of Dirichlet multinomial splitting standard beta negative binomial distribution (respectively (a) without constraint and (b) with constraint $r = |\boldsymbol{\alpha}|$).

the specific constraint $a = |\boldsymbol{\alpha}|$ (see Table 3). The multivariate generalized waring distribution (MGWD), introduced by Xekalaki [52], is recovered when $\mathcal{L}(\psi) = \beta\mathcal{NB}(r,a,b)$ with the specific constraint $r = |\boldsymbol{\alpha}|$ (see Table 4). Finally, a multivariate extension of the beta Poisson distribution is proposed when $\mathcal{L}(\psi) = \beta_\lambda\mathcal{P}(a,b)$ with the specific constraint $a = |\boldsymbol{\alpha}|$ (see Table 5 of Supplementary Materials S2).

Assume now that $\mathcal{L}(\psi)$ is a power series distributions leading to three new multivariate extensions. Remark that several multivariate extensions of the same univariate distribution could be defined. For instance the multinomial splitting beta binomial distribution $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \beta\mathcal{B}_m(a,b)$ and the Dirichlet multinomial splitting binomial distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{B}_m(p)$ are two multivariate extensions of the non-standard beta binomial distribution (see Tables 4 and 6 of Supplementary Materials S2). The specific case of Dirichlet multinomial splitting negative binomial distribution $\mathcal{DM}_{\Delta_n}(\boldsymbol{\alpha}) \underset{n}{\wedge} \mathcal{NB}(r,p)$ has been studied in depth by [23], named as negative binomial sum and Polya share. It is worth noticing that assumption $\boldsymbol{\alpha} = (1,\ldots,1)$ leads to the discrete Schur-constant distribution. The discrete Schur-constant distribution described by [6] can be viewed as a special case of Dirichlet multinomial splitting distribution. On one hand, according to Lefèvre and Loisel [28], the sum distribution is a binomial damage distribution mixed by a specific continous distribution : $\mathcal{B}_m(p) \underset{p}{\wedge} g \underset{m}{\wedge} \mathcal{L}$, where g is the probability distribution function (pdf) of variable $1 - U^{1/(J-1)}$ with U uniformly distributed on $(0,1)$. On the other hand, the singular distribution is given by $\mathcal{DM}_{\Delta_n}(\mathbf{1})$ where $\mathbf{1} := (1,\ldots,1)$. It should be noted that $\mathcal{DM}_{\Delta_n}(\mathbf{1}) = \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{\boldsymbol{\pi}}{\wedge} \mathcal{D}_\Delta(\mathbf{1})$ where $\mathcal{D}_\Delta(\mathbf{1})$ denotes the specific Dirichlet distribution with parameters $\mathbf{1}$, i.e., the uniform distribution on the continuous simplex Δ. In fact, $\mathcal{DM}_{\Delta_n}(\mathbf{1})$ turns out to be the uniform distribution on the discrete simplex Δ_n. Finally the discrete Schur-constant distribution can be written as $\mathcal{DM}_{\Delta_n}(\mathbf{1}) \underset{n}{\wedge} \mathcal{B}_m(p) \underset{p}{\wedge} g \underset{m}{\wedge} \mathcal{L}$.

Otherwise, note that the singular Dirichlet multinomial distribution does not belong to the exponential family. Regardless $|\boldsymbol{\alpha}|$ being fixed or not, MLEs $\widehat{\alpha_j}$ can be computed using various iterative methods [32, 43]. Finally, [3] described Bayesian estimation of $\boldsymbol{\alpha}$ and also of r and p.

5.3. *Canonical case of beta binomial sum distribution*

The case $\mathcal{L}(\psi) = \beta\mathcal{B}_n(a, b)$ is considered as the canonical case since the beta binomial distribution is the univariate version of the non-singular Dirichlet multinomial distribution. Usual characteristics of the Dirichlet multinomial splitting beta binomial distribution are derived from equations (14), (15), (16), (17) and (18) with $\mathcal{L}(\psi) = \beta\mathcal{B}_n(a, b)$. The constraint $a = |\boldsymbol{\alpha}|$ in Corollary 3, has to be taken into account in the inference procedure, either on the singular distribution or on the sum distribution. We propose to use the first alternative since the inference procedure of a constrained Dirichlet multinomial distribution (i.e., with a fixed sum $|\boldsymbol{\alpha}|$) has already been proposed by Minka [32]. The sum distribution $\beta\mathcal{B}_n\,(a, b)$ can then be estimated without constraint on parameters a or b (see Table 1 of Supplementary Materials S1). Note that, if no constraint between parameters of singular and sum distributions is assumed then the inference procedure is straightforward, since it can be separated into two independent procedures. The resulting splitting distribution is more general, including the non-singular Dirichlet multinomial distribution as a special case. As a consequence from equation (18), the marginals follow beta square binomial distributions $\beta^2\mathcal{B}_n(\alpha_j, |\boldsymbol{\alpha}_{-j}|, a, b)$ and beta binomial distributions $\beta\mathcal{B}_n(\alpha_j, |\boldsymbol{\alpha}_{-j}| + b)$ when the constraint $a = |\boldsymbol{\alpha}|$ is assumed (see Supplementary Materials S1 for definition of beta square distribution and beta square compound distributions). Identifying these two distributions we obtain a property about the product of two independent beta distributions.

Proposition 1. *For $(a, b, c) \in (0, \infty)^3$, let $X \sim \beta(a, b)$ and $Y \sim \beta(a + b, c)$ be two independent random variables, then $XY \sim \beta(a, b + c)$.*

This result can be extended by induction for a product of n independent beta distributions.

6. Empirical studies

All studies presented in this section are reproducible. (see Supplementary Materials).

6.1. *A comparison of regression models for multivariate count responses*

In order to illustrate the variety of splitting models, we considered two datasets used in the literature to illustrate models for count data. The first one consists in outcomes of football games [25] and the second one consists in simulated data mimicking data obtained from sequencing techonologies such as RNA-seq data [55]. The goal being to compare distributions and regressions models, comparisons were performed when considering all covariates or none of the covariates (see Table 5). Remark that variable selection [55, e.g., using regularization methods] is possible, but is out of the scope of this paper.

Let us first remark that the inference methodology for multinomial, Dirichlet multinomial and generalized Dirichlet multinomial regressions presented by Zhang et al. [55] and implemented by Zhang and Zhou [54] is only valid for singular versions. Their comparisons of these models against the negative multinomial is therefore invalid since the first three models focus on $\boldsymbol{y}$ given $|\boldsymbol{y}|$ and the latter focuses on $\boldsymbol{y}$. Hence, we only compared our results to their unique J-multivariate model that is the negative multinomial model and the multivariate Poisson model defined by Karlis and Meligkotsidou [24]. By limiting the number of sum models to 7 and the number of singular models to 6, we were able to propose 42 splitting models. Among those 42 models, only 4 models were not sensu stricto multivariate models since multinomial splitting Poisson models induce independent response variables.

For the first dataset, the best splitting model is a singular multinomial regression compounded by a Poisson distribution with a BIC of $508.14 + 1,130.64 = 1,638.78$. This score is inferior to the one of the best multivariate Poisson model (i.e., $1,710.05$) and the one of the best negative multinomial model (i.e., $1,705.93$). According to Property 2, the graphical model of the best model is empty; this indicates that there is no relationship between football team goals. For the second dataset, the best splitting model is a singular generalized Dirichlet multinomial regression compounded by a negative binomial regression with a BIC of $8,843.48 + 2,514.3 = 11,357.78$. This score is also inferior to the one of the best negative multinomial model (i.e., $17,657.63$).

$\boldsymbol{y}$ given $\lvert\boldsymbol{y}\rvert = n$ and $\boldsymbol{x}$	BIC$_0$	BIC$_1$
$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$	574.18	38,767.91
$\mathcal{DM}_{\Delta_n}(\alpha)$	579.49	9,969.121
$\mathcal{GDM}_{\Delta_n}(\boldsymbol{\alpha},\boldsymbol{\beta})$	579.49	9,735.45
$\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}$	**508.14**	15,145.24
$\mathcal{DM}_{\Delta_n}\{\boldsymbol{\alpha}(\boldsymbol{x})\}$	836.4	8,932.83
$\mathcal{GDM}_{\Delta_n}\{\boldsymbol{\alpha}(\boldsymbol{x}),\boldsymbol{\beta}(\boldsymbol{x})\}$	836.4	**8,843.479**

(a)

$\lvert\boldsymbol{y}\rvert$ given $\boldsymbol{x}$	BIC$_0$	BIC$_1$
$\mathcal{P}(\lambda)$	**1,130.64**	13,074.12
$\mathcal{B}_n(p)$	1,165.6	26,474.38
$\mathcal{NB}(r,p)$	1,131.85	2,678.55
$\mathcal{L}(p)$	1,370.84	3,513.92
$\mathcal{P}\{\lambda(\boldsymbol{x})\}$	1,258.65	6,353.13
$\mathcal{B}_n\{p(\boldsymbol{x})\}$	1,272.7	12,999.47
$\mathcal{NB}\{r,p(\boldsymbol{x})\}$	1,264.38	**2,514.30**

(b)

$\boldsymbol{y}$ given $\boldsymbol{x}$	BIC$_0$	BIC$_1$
$\mathcal{MP}(\lambda)$	**1,710.05**	×
$\mathcal{MP}\{\lambda(\boldsymbol{x})\}$	1,956.10	×

(c)

$\boldsymbol{y}$ given $\boldsymbol{x}$	BIC$_0$	BIC$_1$
$\mathcal{MN}(r,\boldsymbol{\pi})$	**1,705.93**	41,384.52
$\mathcal{MN}\{r,\boldsymbol{\pi}(\boldsymbol{x})\}$	2,176.3	**17,657.63**

(d)

Table 5: Bayesian Information Criteria (BIC) obtained for the first dataset [25, BIC$_0$] and the second one [55, BIC$_1$] for (a) singular models, (b) sum models, (c) Poisson and (d) negative multinomial models. Multivariate Poisson models could not be fit to the second dataset since, to our knowledge, there is no implementation available in R for more than 2 response variables [25].

6.2. *An application to mango patchiness analysis*

Recently, a statistical methodology has been proposed to characterize plant patchiness at whole plant scale [13]. However, little is known about patchiness at a whole population scale. To characterize patchiness at the plant scale, a segmentation/clustering of tree-indexed data method has been proposed in order to split an heterogeneous tree into multiple homogeneous subtrees. After the clustering step, the tree can be summarized into a multivariate count denoting the number of subtrees in each cluster (i.e., patch type). Mixture of multinomial splitting distributions can therefore be considered to recover the different types of tree patchiness that can be found in the plant population. Such a mixture model is of high interest since it enables types of tree patchiness to be discriminated according to the:

- number of patches present on trees, by fitting different sum distributions within components of the mixture model,
- distribution of these patches among types, by fitting different singular distributions within components of the mixture model.

We here consider results presented by Fernique et al. [13] to conduct our *post-hoc* analysis. Three different types of patches have been identified for mango trees: vegetative patches that contain almost only vegetative growth units (GU, plant elementary component), reproductive patches that contain almost only GUs that flowered or fructified and quiescent patches that contain GUs that neither burst, flowered nor fructified. Multinomial splitting distributions corresponding to mixture components are therefore of dimension 3, where y_1 (resp. y_2 and y_3) denotes the number of vegetative (resp. reproductive and quiescent) patches observed within a tree. Since there is at least one patch in a mango tree (i.e., the tree itself), shifted singular multinomial splitting distributions were considered with a shift equal to 1 for binomial, negative binomial and Poisson sum distributions but without shift for geometric and logarithmic distributions. Within each component the parametric form of the sum distribution was selected using BIC.

The mixture model selected using ICL [4] has two components (see Figure 1) with weights $P(L = 1) = 0.44$ and $P(L = 2) = 0.56$. In the two components $k = 1, 2$, the number of patches follow a multinomial splitting shifted negative binomial distribution $\boldsymbol{y} \,|\, L = k \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}_k) \underset{n}{\wedge} \mathcal{NB}(r_k, p_k; \delta_k)$ with estimates $\hat{\boldsymbol{\pi}}_1 = (0.21, 0.00, 0.79)$, $\hat{r}_1 = 0.16$, $\hat{p}_1 = 0.76$, $\hat{\delta}_1 = 1$ for the first component and $\hat{\boldsymbol{\pi}}_2 = (0.54, 0.17, 0.28)$, $\hat{r}_2 = 3.96$, $\hat{p}_2 = 0.40$, $\hat{\delta}_2 = 1$ for the second component. This mixture of two components indicates that the population of mango trees can be separated into two types of trees (see Figure 1):

- mango trees with a relatively low number of patches that can be either vegetative or quiescent but not reproductive (component 1),

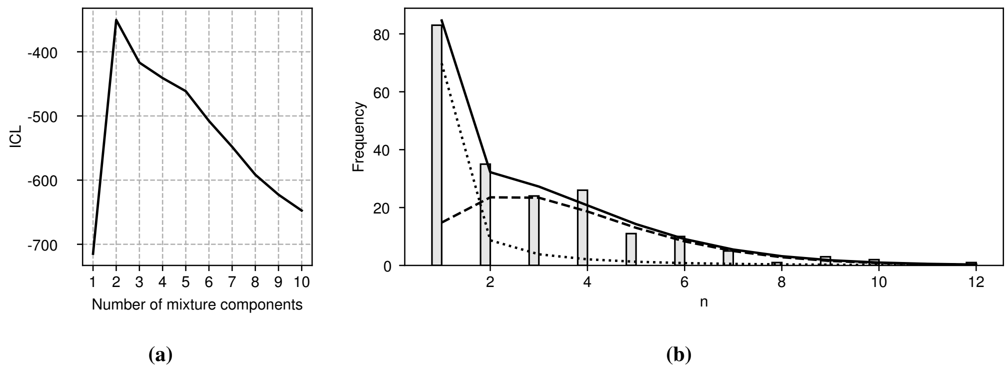

Fig. 1: (a) ICL according to the number of multinomial splitting components of mixtures. (b) Mixture of sum distributions estimated (with a solid line) confronted to data frequencies (gray bars). The sum distribution of the first (resp. second) component is represented with a dotted (resp. dashed) line.

- mango trees with a relatively high number of patches that can be of any type and in particular reproductives (component 2).

These types of trees are almost equally represented in the period considered (52% for the first component against 48%). This result tends to imply that the reproductive period of mango trees leads to an increase in patch numbers while the vegetative period leads to a decrease in patch numbers.

7. Conclusions

Convolutions splitting distributions that are positive and additive, have been studied in depth in this paper since their graphical models and marginal distributions are easily obtained. The characterization of the graphical model of hypergeometric splitting distributions stays an open issue because of the non-positivity. However, due to additivity, Theorem 1 still holds. It would be interesting to study the hypergeometric splitting distributions $\mathcal{H}_{\Delta_n} \underset{n}{\wedge} \mathcal{L}(\psi)$ for some specific univariate distributions $\mathcal{L}(\psi)$. More generally, the multivariate Polya distribution with parameters $n \in \mathbb{N}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $c \in \mathbb{R}$ encompasses the multivariate hypergeometric ($c = -1$), the multinomial ($c = 0$) and the Dirichlet multinomial ($c = 1$) distributions [19]. It would therefore be interesting to study the properties of multivariate Polya splitting distributions according to the c value. Otherwise, non-symmetric convolution distributions could be defined (including the generalized Dirichlet multinomial distribution) to ease the study of corresponding splitting distributions.

Another alternative to define new singular distributions is to consider their mixtures. To motivate such extensions of our approach, let us consider the mango tree application, in which we inferred mixtures of splitting distributions in order to characterize plant patchiness at whole plant scale. This relied on the assumption that tree patchiness is both expressed in terms of number of patches and the distribution of their types. On the one hand, if tree patchiness is only a phenomenon expressed in terms of number of patches, a mixture of sum distributions could be considered to distinguish trees. On the other hand, if tree patchiness is only a phenomenon expressed in terms of patch type distribution, singular distributions constructed using mixture of singular distributions could be of most interest. This highlights how mixture models are quite interesting to define new splitting models. Finite mixtures can be inferred using a classical expectation-maximization algorithm for multivariate distributions.

Regarding parameter estimation, properties of conditional independence of estimators for sum and singular distribution parameters have been established for MLE and Bayesian estimators in the framework of multinomial splitting distributions. Similar properties remain to be investigated for other cases of splitting (or possibly sum) distributions and regression models.

Finally, this work could be used for learning graphical models with discrete variables, which is an open issue. Although the graphical models for usual additive convolution splitting distributions are trivial (either complete or empty), they could be used as building blocks for partially directed acyclic graphical models. Therefore, the procedure of learning partially directed acyclic graphical models described by Fernique et al. [14] could be used for learning graphical models based on convolution splitting distributions and regressions. It could be used for instance to infer gene co-expression network from RNA seq dataset.

Supplementary material

Supplementary material includes the definitions univariate distributions used in the paper and references to their inference procedure (S1) tables containing the characteristics (notation, pmf, expectation, covariance and pgf) of several convolution splitting distributions (S2) and references of inference procedure for several singular and univariate regressions (S3). Moreover, the source code used for the inference of splitting distributions is available on GitHub (http://github.com/StatisKit/FPD18). Binaries can be installed using the Conda package management system (http://conda.pydata.org). Our analyses performed with *Python* and *R* packages is available in the Jupyter notebook format and can be reproduced using a Docker image [31].

Appendix

Appendix 1: proof of Theorem 1

Proof of item 1: Let $n \in \mathbb{N}$, we have

$$\begin{aligned}
P(|\boldsymbol{Y}_{\mathcal{I}}| = n) &= \sum_{\boldsymbol{y}_{\mathcal{I}} \in \Delta_n} P(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) \\
&= \sum_{\boldsymbol{y}_{\mathcal{I}} \in \Delta_n} \sum_{\boldsymbol{y}_{-\mathcal{I}}} P(\boldsymbol{Y} = \boldsymbol{y}) \\
&= \sum_{k \geq n} P(|\boldsymbol{Y}| = k) \sum_{\boldsymbol{y}_{\mathcal{I}} \in \Delta_n} \sum_{\boldsymbol{y}_{-\mathcal{I}} \in \Delta_{k-n}} P_{|\boldsymbol{Y}|=k}(\boldsymbol{Y} = \boldsymbol{y}) \\
&= \sum_{k \geq n} \frac{P(|\boldsymbol{Y}| = k)}{c_{\boldsymbol{\theta}}(k)} \sum_{\boldsymbol{y}_{\mathcal{I}} \in \Delta_n} \prod_{j \in \mathcal{I}} a_{\theta_j}(y_j) \sum_{\boldsymbol{y}_{-\mathcal{I}} \in \Delta_{k-n}} \prod_{j \in -\mathcal{I}} a_{\theta_j}(y_j) \\
P(|\boldsymbol{Y}_{\mathcal{I}}| = n) &= \sum_{k \geq n} \frac{c_{\boldsymbol{\theta}_{\mathcal{I}}}(n) c_{\boldsymbol{\theta}_{-\mathcal{I}}}(k-n)}{c_{\boldsymbol{\theta}}(k)} P(|\boldsymbol{Y}| = k)
\end{aligned}$$

where $c_{\boldsymbol{\theta}_{\mathcal{I}}}(n)$ denotes the convolution of $(a_{\theta_j})_{j \in \mathcal{I}}$ over the simplex Δ_n. Since the convolution distribution is assumed to be additive, we obtain by recursion on $j \in \mathcal{I}$ (resp. $j \in -\mathcal{I}$ and $j \in \{1, \dots, J\}$) that

$$P(|\boldsymbol{Y}_{\mathcal{I}}| = n) = \sum_{k \geq n} \frac{a_{|\boldsymbol{\theta}_{\mathcal{I}}|}(n) a_{|\boldsymbol{\theta}_{-\mathcal{I}}|}(k-n)}{a_{|\boldsymbol{\theta}|}(k)} P(|\boldsymbol{Y}| = k) \tag{19}$$

Moreover we obtain the convolution identity $\sum_{n=0}^{k} a_{|\boldsymbol{\theta}_{\mathcal{I}}|}(n) a_{|\boldsymbol{\theta}_{-\mathcal{I}}|}(k-n) = a_{|\boldsymbol{\theta}|}(k)$ and thus the last equation well defines the desired convolution damage distribution $\sim \mathcal{C}_N(a; |\boldsymbol{\theta}_{\mathcal{I}}|, |\boldsymbol{\theta}_{-\mathcal{I}}|) \underset{N}{\wedge} \mathcal{L}(\boldsymbol{\psi})$.

Proof of items 2 and 3: For $\boldsymbol{y}_{\mathcal{I}} \in \Delta_n$ we have

$$\begin{aligned}
P(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}},\ |\boldsymbol{Y}_{\mathcal{I}}| = n) &= P(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}), \\
&= \sum_{k \geq n} P(|\boldsymbol{Y}| = k) \sum_{\boldsymbol{y}_{-\mathcal{I}} \in \Delta_{k-n}} P_{|\boldsymbol{Y}|=k}(\boldsymbol{Y} = \boldsymbol{y}), \\
&= \prod_{j \in \mathcal{I}} a_{\theta_j}(y_j) \sum_{k \geq n} \frac{P(|\boldsymbol{Y}| = k)}{c_{\boldsymbol{\theta}}(k)} \sum_{\boldsymbol{y}_{-\mathcal{I}} \in \Delta_{k-n}} \prod_{j \in -\mathcal{I}} a_{\theta_j}(y_j), \\
P(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}},\ |\boldsymbol{Y}_{\mathcal{I}}| = n) &= \prod_{j \in \mathcal{I}} a_{\theta_j}(y_j) \sum_{k \geq n} \frac{a_{|\boldsymbol{\theta}_{-\mathcal{I}}|}(k-n)}{a_{|\boldsymbol{\theta}|}(k)} P(|\boldsymbol{Y}| = k).
\end{aligned}$$

Using equation (19) we obtain, for $\boldsymbol{y}_{\mathcal{I}} \in \Delta_n$, the conditional probability

$$P_{|\boldsymbol{Y}_{\mathcal{I}}|=n}(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) = \frac{1}{a_{|\boldsymbol{\theta}_{\mathcal{I}}|}(n)} \prod_{j\in\mathcal{I}} a_{\theta_j}(y_j),$$

and thus 2 holds. Let us remark that 1 and 2 imply 3 by definition of a splitting distribution.
Proof of items 4 and 5: For $\boldsymbol{y}_{\mathcal{I}} \in \mathbb{N}^I$ (where I is the cardinal of $\mathcal{I}$) we have

$$P_{\boldsymbol{Y}_{-\mathcal{I}}=\boldsymbol{y}_{-\mathcal{I}}}(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) = P_{\boldsymbol{Y}_{-\mathcal{I}}=\boldsymbol{y}_{-\mathcal{I}},\,|\boldsymbol{Y}_{\mathcal{I}}|=|\boldsymbol{y}_{\mathcal{I}}|}(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) P_{\boldsymbol{Y}_{-\mathcal{I}}=\boldsymbol{y}_{-\mathcal{I}}}(|\boldsymbol{Y}_{\mathcal{I}}| = |\boldsymbol{y}_{\mathcal{I}}|).$$

Since the sum $|\boldsymbol{Y}|$ is independent of the vector $\boldsymbol{Y}_{-\mathcal{I}}$ given its sum $|\boldsymbol{Y}_{-\mathcal{I}}|$ it can be shown that

$$P_{\boldsymbol{Y}_{-\mathcal{I}}=\boldsymbol{y}_{-\mathcal{I}}}(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) = P_{|\boldsymbol{Y}_{\mathcal{I}}|=|\boldsymbol{y}_{\mathcal{I}}|}(\boldsymbol{Y}_{\mathcal{I}} = \boldsymbol{y}_{\mathcal{I}}) P_{|\boldsymbol{Y}_{-\mathcal{I}}|=|\boldsymbol{y}_{-\mathcal{I}}|}(|\boldsymbol{Y}_{\mathcal{I}}| = |\boldsymbol{y}_{\mathcal{I}}|).$$

Thanks to the result 2, the left part of this product is given by the singular convolution distribution. Remarking that $P_{|\boldsymbol{Y}_{-\mathcal{I}}|=|\boldsymbol{y}_{-\mathcal{I}}|}(|\boldsymbol{Y}_{\mathcal{I}}| = |\boldsymbol{y}_{\mathcal{I}}|) = P_{|\boldsymbol{Y}|\geq a}(|\boldsymbol{Y}| = a + |\boldsymbol{y}_{\mathcal{I}}|)$ with $a = |\boldsymbol{y}_{-\mathcal{I}}|$ the left part is given by the truncated and shifted distribution $TS_a\{\mathcal{L}(\psi)\}$ and thus 4 holds. Let us remark that 3 and 4 imply 5.

Appendix 2: proof of Property 1

Assume that $\mathcal{L}(\boldsymbol{\psi})$ is stable under the damage process $C_N(a; |\boldsymbol{\theta}_{\mathcal{I}}|, |\boldsymbol{\theta}_{-\mathcal{I}}|) \underset{N}{\wedge} (\cdot)$ for any subset $\mathcal{I} \subset \{1, \ldots, J\}$. Thanks to the additivity of the convolution distribution, Theorem 1 can be applied. Using item 3, it is easily seen that multivariate marginals are stable. Criterion 4 holds and the convolution splitting distribution is considered as a natural multivariate extension of $\mathcal{L}(\boldsymbol{\psi})$. In particular, $\mathcal{L}(\boldsymbol{\psi})$ is stable under $C_N(a; |\theta_j|, |\boldsymbol{\theta}_{-j}|) \underset{N}{\wedge} (\cdot)$, i.e., there exists $\boldsymbol{\psi}_j \in \boldsymbol{\Psi}$ such that $Y_j \sim \mathcal{L}(\boldsymbol{\psi}_j)$.

Appendix 3: proof of Property 2

Let $y \sim C_n(a; \theta, \gamma) \underset{n}{\wedge} C_m(a; \theta + \gamma, \lambda)$. For $y \leq m$ we have

$$\begin{aligned}
p(y) &= \sum_{n=y}^{m} \frac{a_\theta(y) a_\gamma(n-y)}{a_{\theta+\gamma}(n)} p(n) \\
&= a_\theta(y) \sum_{n=y}^{m} \frac{a_\gamma(n-y)}{a_{\theta+\gamma}(n)} \frac{a_{\theta+\gamma}(n) a_\lambda(m-n)}{a_{\theta+\gamma+\lambda}(m)} \\
&= \frac{a_\theta(y)}{a_{\theta+\gamma+\lambda}(m)} \sum_{n=y}^{m} a_\gamma(n-y) a_\lambda(m-n) \\
p(y) &= \frac{a_\theta(y)}{a_{\theta+\gamma+\lambda}(m)} \underbrace{\sum_{n=0}^{m-y} a_\gamma(n) a_\lambda(m-y-n)}_{=a_{\gamma+\lambda}(m-y) \text{ by additivity}}
\end{aligned}$$

i.e., $y \sim C_m(a; \theta, \gamma + \lambda)$.

Appendix 4: proof of Property 3

Let $\boldsymbol{y}$ follow the non-singular version of an additive convolution distribution: $\boldsymbol{y} \sim C_{\blacktriangle_m}(a; \boldsymbol{\theta}, \gamma)$. It means that the completed vector $(\boldsymbol{y}, m - |\boldsymbol{y}|)$ follow the additive convolution $C_{\Delta_m^{J+1}}(a; \boldsymbol{\theta}, \gamma)$. Otherwise this singular distribution can seen as a particular splitting Dirac distribution, i.e., $C_{\Delta_m^{J+1}}(a; \boldsymbol{\theta}, \gamma) = C_{\Delta_n^{J+1}}(a; \boldsymbol{\theta}, \gamma) \underset{n}{\wedge} \mathbf{1}_m$. Thanks to the additivity, the Theorem 1 can be applied on the completed vector $(\boldsymbol{y}, n - |\boldsymbol{y}|)$ to describe the distribution of $\boldsymbol{y}$ (item 3):

$$\begin{aligned}
&\boldsymbol{y} \sim C_{\Delta_n}(a; \boldsymbol{\theta}) \underset{n}{\wedge} \left\{ C_{n'}(a; |\boldsymbol{\theta}|, \gamma) \underset{n'}{\wedge} \mathbf{1}_m \right\}, \\
\Leftrightarrow\ &\boldsymbol{y} \sim C_{\Delta_n}(a; \boldsymbol{\theta}) \underset{n}{\wedge} C_m(a; |\boldsymbol{\theta}|, \gamma).
\end{aligned}$$

Appendix 5: proof of Corollary 1

Using Property 3 with $a_\theta(y) = \theta^y/y!$ we obtain for $\boldsymbol{\theta} \in (0,\infty)^J$ and $\gamma \in (0,\infty)$

$$\mathcal{M}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(|\boldsymbol{\theta}|, \gamma) = \mathcal{M}_{\blacktriangle_m}(\boldsymbol{\theta}, \gamma)\,.$$

Denoting by $\boldsymbol{\pi} = \frac{1}{|\boldsymbol{\theta}|} \cdot \boldsymbol{\theta}$, $p = \frac{|\boldsymbol{\theta}|}{|\boldsymbol{\theta}|+\gamma}$ and $\boldsymbol{\pi}^* = \frac{1}{|\boldsymbol{\theta}|+\gamma} \cdot \boldsymbol{\theta}$ and using the proportionality we obtain equivalently

$$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{B}_m(p, 1-p) = \mathcal{M}_{\blacktriangle_m}(\boldsymbol{\pi}^*, 1 - |\boldsymbol{\pi}^*|)\,.$$

The notation of the binomial and the non-singular multinomial are then simplified by letting aside the last parameter without loss of generality, i.e. we have $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{B}_m(p) = \mathcal{M}_{\blacktriangle_m}(\boldsymbol{\pi}^*)$. Finally remarking that $\boldsymbol{\pi}^* = p \cdot \boldsymbol{\pi}$ we obtain the desired result.

Appendix 6: proof of Corollary 2

According to Theorem 1 we know that a univariate marginal of multinomial splitting distribution follows a binomial damage distribution. Let us now express the pmf of such a distribution $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(\psi)$ according to the pgf G_ψ of the sum distribution $\mathcal{L}(\psi)$:

$$\begin{aligned}
p(y) &= \sum_{n\geq y} \binom{n}{y} \pi^y (1-\pi)^{n-y} p_\psi(n),\\
&= \frac{\pi^y}{y!} \sum_{n\geq y} \frac{n!}{(n-y)!} (1-\pi)^{n-y} p_\psi(n),\\
p(y) &= \frac{\pi^y}{y!} G_\psi^{(y)}(1-\pi).
\end{aligned}$$

Appendix 7: proof of Property 4

1. As a special case of Property 2, for $\theta \in (0,\infty)$, $\gamma \in (0,\infty)$ and $\lambda \in (0,\infty)$ we have

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

   Using the representative elements $\pi := \frac{\theta}{\theta+\gamma}$ and $p = \frac{\theta+\gamma}{\theta+\gamma+\lambda}$ we obtain the desired result and the additive constraint between parameters disappears.

2. Let $y \sim \mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{P}(\lambda)$. We use the Corollary 2 which states that

   $$p(y) = \frac{\pi^y}{y!} G^{(y)}(1-\pi)$$

   Otherwise the pgf of the Poisson distribution is $G(s) = \exp\{\lambda(s-1)\}$. Recursively on y we obtain $G^{(y)}(s) = \lambda^y \exp\{\lambda(s-1)\}$ and

   $$p(y) = \exp(-\lambda\pi) \frac{(\pi\lambda)^y}{y!}.$$

3. Let $y \sim \mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{NB}(r, p)$. The pgf of the negative binomial distribution is $G(s) = \left(\frac{1-p}{1-ps}\right)^r$. Recursively on y we obtain $G^{(y)}(s) = (1-p)^r p^y \frac{(r+y-1)!}{(r-1)!} (1-ps)^{-r-y}$ and

   $$p(y) = \binom{y+r-1}{y} \left(\frac{\pi p}{1-p+\pi p}\right)^y \left(\frac{1-p}{1-p+\pi p}\right)^r.$$

4. Let $y \sim \mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(p, \omega)$

$$\begin{aligned}
p(0) &= \sum_{n\geq 0}(1-\pi)^n p(n) \\
&= \frac{\omega}{\omega - \ln(1-p)} + \sum_{n\geq 1}(1-\pi)^n \frac{p^n/n}{\omega - \ln(1-p)} \\
&= \frac{1}{\omega - \ln(1-p)}\left[\omega + \sum_{n\geq 1}\frac{\{(1-\pi)p\}^n}{n}\right] \\
p(0) &= \frac{\omega - \ln(\pi p + 1 - p)}{\omega - \ln(1-p)}
\end{aligned}$$

For $y \geq 1$ we have

$$\begin{aligned}
p(y) &= \frac{(\pi p)^y}{\omega - \ln(1-p)} \sum_{n\geq y}\binom{n}{y}\frac{\{(1-\pi)p\}^n}{n} \\
&= \frac{(\pi p)^y/y}{\omega - \ln(1-p)} \underbrace{\sum_{n\geq 0}\binom{n+y-1}{n}\{(1-\pi)p\}^n}_{=(\pi p+1-p)^{-y}} \\
p(y) &= \frac{(p')^y/y}{\omega' - \ln(1-p')}
\end{aligned}$$

where $p' := \frac{\pi p}{\pi p+1-p}$ and $\omega' := \omega - \ln(\pi p + 1 - p)$. Therefore we obtain $p(0) = \frac{\omega'}{\omega'-\ln(1-p')}$ and thus the desired result.

Appendix 8: proof of Theorem 3

From (5) and (13), the MLE $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ is a deterministic function of $|\mathcal{Y}|$. Thus, to prove that the MLE $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\psi}}$ are independent given Z, it is sufficient to prove that $\hat{\boldsymbol{\pi}}$ and $|\mathcal{Y}|$ are independent given Z. For any $(q_1, \ldots, q_{J-1}) \in \mathbb{Z}_+^{J-1}$ and for any $(n_i)_{1\leq i\leq m} \in \mathbb{N}^m$,

$$\begin{aligned}
&P\left(\hat{\pi}_1 = q_1, \ldots, \hat{\pi}_{J-1} = q_{J-1} \,\middle|\, (|\boldsymbol{Y}_i| = n_i)_{1\leq i\leq m}\right) \\
&= P\left(\sum_{i=1}^m Y_{i,1} = zq_1, \ldots, \sum_{i=1}^m Y_{i,J-1} = zq_{J-1} \,\middle|\, (|\boldsymbol{Y}_i| = n_i)_{1\leq i\leq m}\right),
\end{aligned}$$

from (13). Since $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ are independent random vectors, $\left(\sum_{i=1}^m Y_{i,1}, \ldots, \sum_{i=1}^m Y_{i,J-1}\right)$ has distribution $\mathcal{M}_{\blacktriangle_z}(\boldsymbol{\pi})$ given $(|\boldsymbol{Y}_i| = n_i)_{1\leq i\leq m}$. Thus,

$$\begin{aligned}
&P\left(\hat{\pi}_1 = q_1, \ldots, \hat{\pi}_{J-1} = q_{J-1} \,\middle|\, (|\boldsymbol{Y}_i| = n_i)_{1\leq i\leq m}\right) \\
&= P\left(\hat{\pi}_1 = q_1, \ldots, \hat{\pi}_{J-1} = q_{J-1}, | Z = z\right)
\end{aligned}$$

and $\hat{\boldsymbol{\pi}}$ and $(|\boldsymbol{Y}_i| = n_i)_{1\leq i\leq m}$ are independent given Z.

Appendix 9: proof of Corollary 4

Since the Dirichlet multinomial distribution is additive, the item 3 of Theorem 1 can be applied to describe the marginal distributions:

$$\begin{aligned}
y_j &\sim \beta\mathcal{B}_n(\alpha_j, |\boldsymbol{\alpha}_{-j}|) \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi}), \\
\Leftrightarrow y_j &\sim \left\{\mathcal{B}_n(\pi) \underset{\pi}{\wedge} \beta(\alpha_j, |\boldsymbol{\alpha}_{-j}|)\right\} \underset{n}{\wedge} \mathcal{L}(\boldsymbol{\psi}).
\end{aligned}$$

Since n and π are independent latent variables, the Fubini theorem can be applied in order to invert the sum (composition of n) and the integral (composition of π).

[1] Allen, G.I., Liu, Z., 2012. A log-linear graphical model for inferring genetic networks from high-throughput sequencing data, in: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 1–6.
[2] Atchison, J., Shen, S.M., 1980. Logistic-normal distributions: Some properties and uses. Biometrika 67, 261–272.
[3] Bhagwat, P., 2019. Models and Statistical Inference for Multivariate Count Data. Ph.D. thesis. Indian Institute of Science Education and Research Pune.
[4] Biernacki, C., Celeux, G., Govaert, G., 2000. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE transactions on pattern analysis and machine intelligence 22, 719–725.
[5] Blumenthal, S., Dahiya, R.C., 1981. Estimating the binomial parameter n. Journal of the American Statistical Association 76, 903–909.
[6] Castañer, A., Claramunt, M.M., Lefèvre, C., Loisel, S., 2015. Discrete schur-constant models. Journal of Multivariate Analysis 140, 343–362.
[7] Chen, J., Li, H., 2013. Variable selection for sparse dirichlet-multinomial regression with an application to microbiome data analysis. The Annals of Applied Statistics 7, 418–442.
[8] Connor, R.J., Mosimann, J.E., 1969. Concepts of independence for proportions with a generalization of the dirichlet distribution. Journal of the American Statistical Association 64, 194–206.
[9] Dai, H., Bao, Y., Bao, M., 2012. Maximum likelihood estimate for the dispersion parameter of the negative binomial distribution. Statistics & Probability Letters 83, 21–27.
[10] Digby, P., Kempton, R.A., 2012. Multivariate analysis of ecological communities. volume 5. Springer Science & Business Media.
[11] Dunkl, C.F., 2013. Products of Beta distributed random variables. ArXiv e-prints `arXiv:1304.6671`.
[12] Fahrmeir, L., Kaufmann, H., 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. The Annals of Statistics , 342–368.
[13] Fernique, P., Dambreville, A., Durand, J.B., Pradal, C., Lauri, P.E., Normand, F., Guédon, Y., 2016. Characterization of mango tree patchiness using a tree-segmentation/clustering approach, in: 2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), IEEE. pp. 68–74.
[14] Fernique, P., Durand, J.B., Guédon, Y., 2014. Estimation of discrete partially directed acyclic graphical models in multitype branching processes, in: COMPSTAT 2014, 21st International Conference on Computational Statistics.
[15] Forcina, A., Franconi, L., 1988. Regression analysis with the beta-binomial distribution. Rivista di Statistica Applicata 21.
[16] Gurland, J., 1958. A generalized class of contagious distributions. Biometrics 14, 229–249.
[17] Hilbe, J.M., 2011. Negative binomial regression. Cambridge University Press.
[18] Irwin, J.O., 1968. The generalized waring distribution applied to accident theory. Journal of the Royal Statistical Society. Series A (General) , 205–225.
[19] Janardan, K.G., Patil, G.P., 1970. On the multivariate polya distribution: a model of contagion for data with multiple counts. Random Count in Scientific Work 3, 143–161.
[20] Janardan, K.G., Patil, G.P., 1972. A unified approach for a class of multivariate hypergeometric models. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 34, 363–376. URL: `http://www.jstor.org/stable/25049821`.
[21] Johnson, N., Kemp, A., Kotz, S., 1993. Univariate discrete distributions. Wiley-Interscience.
[22] Johnson, N., Kotz, S., Balakrishnan, N., 1997. Discrete multivariate distributions. Wiley New York.
[23] Jones, M., Marchand, É., 2019. Multivariate discrete distributions via sums and shares. Journal of Multivariate Analysis 171, 83–93.
[24] Karlis, D., Meligkotsidou, L., 2005. Multivariate Poisson regression with covariance structure. Statistics and Computing 15, 255–265.
[25] Karlis, D., Ntzoufras, I., 2005. Bivariate poisson and diagonal inflated bivariate Poisson regression models in R. Journal of Statistical Software 14, 1–36.
[26] Koller, D., Friedman, N., 2009. Probabilistic Graphical Models. The MIT Press.
[27] Lauricella, G., 1893. Sulle funzioni ipergeometriche a piu variabili. Rendiconti del Circolo Matematico di Palermo 7, 111–158. URL: `https://doi.org/10.1007/BF03012437`, doi:`10.1007/BF03012437`.
[28] Lefèvre, C., Loisel, S., 2013. On multiply monotone distributions, continuous or discrete, with applications. Journal of Applied Probability 50, 827–847.
[29] Lesnoff, M., Lancelot, R., 2012. aod: Analysis of Overdispersed Data. URL: `http://cran.r-project.org/package=aod`. r package version 1.3.
[30] McCullagh, P., Nelder, J., 1989. Generalized linear models. Monographs on Statistics and Applied Probability 37. Chapman & Hall, London.
[31] Merkel, D., 2014. Docker: Lightweight Linux Containers for Consistent Development and Deployment. Linux Journal 2014, 2.
[32] Minka, T., 2000. Estimating a dirichlet distribution.
[33] Patil, G., 1968. On sampling with replacement from populations with multiple characters. The Indian Journal of Statistics, Series B 30, 355–366.
[34] Patil, G.P., Bildikar, S., 1967. Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties. Journal of the American Statistical Association 62, 655–674.
[35] Pearson, K., 1922. On the χ^2 test of goodness of fit. Biometrika 14, 186–191.
[36] Peyhardi, J., Fernique, P., 2017. Characterization of convolution splitting graphical models. Statistics & Probability Letters 126, 59–64.
[37] Peyhardi, J., Trottier, C., Guédon, Y., 2015. A new specification of generalized linear models for categorical responses. Biometrika 102, 889–906.
[38] Rao, C.R., Srivastava, R.C., 1979. Some characterizations based on a multivariate splitting model. Sankhyā Ser. A 41, 124–128.
[39] Robert, C.P., 2007. The Bayesian Choice. Springer Science & Business Media, 2nd Edition.
[40] Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A., Olmo-Jiménez, M., Martínez-Rodríguez, A.M., 2009. A generalized waring regression model for count data. Computational Statistics & Data Analysis 53, 3717–3725.
[41] Saez-Castillo, A.J., Vilchez-Lopez, S., Olmo-Jimenez, M.J., Rodriguez-Avi, J., Conde-Sanchez, A., Martinez-Rodriguez, A.M., 2017. GWRM: Generalized Waring Regression Model for Count Data. URL: `https://cran.r-project.org/package=GWRM`. gWRM R package version 2.1.0.3.
[42] Shanbhag, D.N., 1977. An extension of the Rao-Rubin characterization of the Poisson distribution. J. Appl. Probability 14, 640–646.

[43] Sklar, M., 2014. Fast mle computation for the dirichlet multinomial. arXiv preprint arXiv:1405.0099 .
[44] Stigler, S.M., 2008. Karl Pearson's theoretical errors and the advances they inspired. Statistical Science 23, 261–271.
[45] Stoltenberg, E., Hjort, N., 2020. Multivariate estimation of Poisson parameters. Journal of Multivariate Analysis 175, 104545.
[46] Tripathi, R.C., Gupta, R.C., Gurland, J., 1994. Estimation of parameters in the beta binomial model. Annals of the Institute of Statistical Mathematics 46, 317–331.
[47] Tutz, G., 2011. Regression for categorical data. volume 34. Cambridge University Press.
[48] Vu, T.N., Wills, Q.F., Kalari, K.R., Niu, N., Wang, L., Rantalainen, M., Pawitan, Y., 2016. Beta-poisson model for single-cell rna-seq data analyses. Bioinformatics 32, 2128–2135.
[49] Whitby, O., 1972. Estimation of parameters in the generalized beta distribution. .
[50] Winkelmann, R., 2013. Econometric analysis of count data. Springer Science & Business Media.
[51] Xekalaki, E., 1981. Chance mechanisms for the univariate generalized waring distribution and related characterizations, in: Statistical distributions in scientific work. Springer, pp. 157–171.
[52] Xekalaki, E., 1986. The multivariate generalized waring distribution. Communications in Statistics - Theory and Methods 15, 1047–1064.
[53] Xia, F., Chen, J., Fung, W.K., Li, H., 2013. A logistic normal multinomial regression model for microbiome compositional data analysis. Biometrics 69, 1053–1063.
[54] Zhang, Y., Zhou, H., 2017. MGLM: Multivariate Response Generalized Linear Models. URL: `https://CRAN.R-project.org/package=MGLM`. r package version 0.0.8.
[55] Zhang, Y., Zhou, H., Zhou, J., Sun, W., 2017. Regression models for multivariate count data. Journal of Computational and Graphical Statistics 26, 1–13.

Supplementary materials for
Splitting models for multivariate count data

1. Univariate distributions

1.1. Continuous univariate distributions

Let us recall the definition of the (generalized) beta distribution with positive real parameters c, α and b, denoted by $\beta_c(a, b)$. Its probability density function described by Whitby [49] is given, for $x \in (0, c)$, by

$$f(x) = \frac{1}{B(a,b)} \frac{x^{a-1}(c-x)^{b-1}}{c^{a+b-1}}.$$

Note that $Z = dX$ with $d \in (0, \infty)$ and $X \sim \beta_c(a, b)$ implies that $Z \sim \beta_{cd}(a, b)$. The parameter c of the beta distribution can thus be interpreted as a rescaling parameter of the standard beta distribution. By convention the standard beta distribution (i.e., defined with $c = 1$) will be denoted by $\beta(a, b)$.

Let us introduce the definition of the (generalized) beta square distribution with parameters $(a_1, b_1, a_2, b_2) \in (0, \infty)^4$ and $c \in (0, \infty)$, denoted by $\beta_c^2(a_1, b_1, a_2, b_2)$, as the product of the two independent beta distributions $\beta(a_1, b_1)$ and $\beta(a_2, b_2)$ normalized on $(0, c)$; see [11] for details. It is named the standard beta square distribution when $c = 1$ and denoted by $\beta^2(a_1, b_1, a_2, b_2)$. More generally the product of m beta distributions could be defined.

1.2. Discrete univariate distributions

1.2.1. Power series distributions:

Let $(b_y)_{y\in\mathbb{N}}$ be a non-negative real sequence such that the series $\sum_{y\geq 0} b_y\theta^y$ converges toward $g(\theta)$ for all $\theta \in D = (0, R)$, where R is the radius of convergence. The discrete random variable Z is said to follow a power series distribution if for all $y \in \mathbb{N}$

$$P(Y = y) = \frac{b_y\theta^y}{g(\theta)},$$

and is denoted by $Y \sim PSD\{g(\theta)\}$. Several usual discrete distributions fall into the family of power series distributions:

1. The Poisson distribution $\mathcal{P}(\lambda)$ with $b_y = 1/y!$, $\theta = \lambda$, $g(\theta) = e^\theta$ and $D = (0, \infty)$.
2. The binomial distribution $\mathcal{B}_n(p)$ with $b_y = \binom{n}{y}\mathbf{1}_{y\leq n}$, $\theta = p/(1-p)$, $g(\theta) = (1+\theta)^n$ and $D = (0, \infty)$.
3. The negative binomial distribution $\mathcal{NB}(r, p)$ with $b_y = \binom{r+y-1}{y}$, $\theta = p$, $g(\theta) = (1-\theta)^{-r}$ and $D = (0, 1)$.
4. The geometric distribution $\mathcal{G}(p)$ with $b_y = \mathbf{1}_{y\geq 1}$, $\theta = 1 - p$, $g(\theta) = \theta/(1-\theta)$ and $D = (0, 1)$.
5. The logarithmic series distribution $\mathcal{L}(p)$ with $b_y = \mathbf{1}_{y\geq 1}1/y$, $\theta = p$, $g(\theta) = -\ln(1-\theta)$ and $D = (0, 1)$.

Let us define the zero modified logarithmic series distribution, denoted by $\mathcal{L}(p, \omega)$ with $p \in (0, 1)$ and $\omega \in [0, \infty)$ and pmf given by

$$p(y) = \begin{cases} \frac{\omega}{\omega - \ln(1-p)} & \text{if } y = 0 \\ \frac{p^y/y}{\omega - \ln(1-p)} & \text{otherwise} \end{cases}$$

It belongs to the family of power series distributions with

$$b_y = \begin{cases} \omega & \text{if } y = 0 \\ 1/y & \text{otherwise} \end{cases}$$

where the series $\sum_{y\geq 0} b_y p^y$ converges on $D = (0, 1)$ towards $g(p) = \omega - \ln(1-p)$. Remark that $\omega = 0$ lead to the usual logarithmic series distribution $\mathcal{L}(p) = \mathcal{L}(p, 0)$ with support $\mathbb{N}^*$.

When the support is a subset of $\mathbb{N}$, the b_y values can be weighted by an indicator function as for binomial, geometric and logarithmic distributions. The b_y must be independent of θ but they may depend on other parameters as for binomial and negative binomial distributions.

1.2.2. Beta compound distributions:

Usual characteristics of the standard beta binomial [46], standard beta negative binomial - also described by Xekalaki [51] as the univariate generalized waring distribution (UGWD) - and the beta Poisson distributions [16] are first recalled in Table 2. Then we introduce these beta compound distributions in a general way, i.e. using the generalized beta distribution, described at the begining of this Section. For the Poisson case we obtain the same distribution since $\mathcal{P}(\lambda p) \underset{p}{\wedge} \beta(a,b) = \mathcal{P}(\theta) \underset{\theta}{\wedge} \beta_\lambda(a,b)$. The two other case lead us to new distributions (see Table 3 for detailed characteristics). Remark that if $\pi = 1$ then, the generalized beta binomial (resp. generalized beta negative binomial) turns out to be the standard beta binomial (resp. standard negative binomial distribution). In opposition, if $\pi < 1$, the non-standard beta binomial distribution (respectively non-standard beta negative binomial distribution) is obtained.

Generalized beta compound distributions. Let $n \in \mathbb{N}$, $a \in (0,\infty)$, $b \in (0,\infty)$ and $\pi \in (0,1)$ and consider the compound distribution $\mathcal{B}_n(p) \underset{p}{\wedge} \beta_\pi(a,b)$ denoted by $\beta_\pi\mathcal{B}_n(a,b)$. Considering π as a rescaling parameter, we have $\beta_\pi\mathcal{B}_n(a,b) = \mathcal{B}_n(\pi p) \underset{p}{\wedge} \beta(a,b)$. Moreover, using the pgf of the binomial distribution in equation (12), it can be shown that $\mathcal{B}_n(\pi p) = \mathcal{B}_N(\pi) \underset{N}{\wedge} \mathcal{B}_n(p)$. Finally, using the Fubini theorem we obtain

$$\begin{aligned}
\beta_\pi\mathcal{B}_n(a,b) &= \left\{\mathcal{B}_N(\pi) \underset{N}{\wedge} \mathcal{B}_n(p)\right\} \underset{p}{\wedge} \beta(a,b), \\
&= \mathcal{B}_N(\pi) \underset{N}{\wedge} \left\{\mathcal{B}_n(p) \underset{p}{\wedge} \beta(a,b)\right\}, \\
\beta_\pi\mathcal{B}_n(a,b) &= \mathcal{B}_N(\pi) \underset{N}{\wedge} \beta\mathcal{B}_n(a,b).
\end{aligned}$$

This is a binomial damage distribution whose the latent variable N follows a standard beta binomial distribution. The equation (12) can thus be used to compute the probability mass function. The y^{th} derivative of the pgf of the standard beta binomial distribution is thus needed

$$G_N^{(y)}(s) = \frac{(b)_n}{(a+b)_n}\frac{(-n)_y(a)_y}{(-b-n+1)_y}{}_2F_1\{(-n+y, a+y); -b-n+1+y; s\},$$

obtained by induction on $y \in \mathbb{N}$. The moments are obtained with the total law of expectation and variance given the latent variable N of the binomial damage distribution. In the same way, we obtain the pgf as $G_Y(s) = G_N(1-\pi+\pi s)$. A similar proof holds for the generalized beta negative binomial case.

Generalized beta square compound distributions. It is also possible to define the (generalized) beta square distribution, as the product of two independent beta distributions [11], and then define the (generalized) beta square compound distributions.

- The standard beta square binomial distribution is defined as $\mathcal{B}_n(p) \underset{p}{\wedge} \beta^2(a_1,b_1,a_2,b_2)$ and denoted by $\beta^2\mathcal{B}_n(a_1,b_1,a_2,b_2)$.
- The standard beta square negative binomial distribution is defined as $\mathcal{NB}(r,p) \underset{p}{\wedge} \beta^2(a_1,b_1,a_2,b_2)$ and denoted by $\beta^2\mathcal{NB}(r,a_1,b_1,a_2,b_2)$.
- The generalized beta square binomial distribution is defined as $\mathcal{B}_n(p) \underset{p}{\wedge} \beta^2_\pi(a_1,b_1,a_2,b_2)$ and denoted by $\beta^2_\pi\mathcal{B}_n(a_1,b_1,a_2,b_2)$
- The generalized beta square negative binomial distribution is defined as $\mathcal{NB}(r,p) \underset{p}{\wedge} \beta^2_\pi(a_1,b_1,a_2,b_2)$ and denoted by $\beta^2_\pi\mathcal{NB}(r,a_1,b_1,a_2,b_2)$
- The beta square Poisson distribution is defined as $\mathcal{P}(\theta) \underset{\theta}{\wedge} \beta^2_\lambda(a_1,b_1,a_2,b_2)$ and denoted by $\beta^2_\lambda\mathcal{P}(a_1,b_1,a_2,b_2)$.

Distribution	Notation	Parameter Inference
Binomial	$\mathcal{B}_n(p)$	See [5]
Negative binomial	$\mathcal{NB}(r,p)$	See [9]
Poisson	$\mathcal{P}(\lambda)$	See [21]
Logarithmic series	$\mathcal{L}(p)$	See [21]
Beta binomial	$\beta\mathcal{B}_n(a,b)$	See [29, 46] for n known
Beta negative binomial	$\beta\mathcal{NB}(r,a,b)$	See [18]
Beta Poisson	$\beta_\lambda\mathcal{P}(a,b)$	See [16, 48]

Table 1: References of parameter inference procedures for seven usual univariate discrete distributions.

Name	Standard beta binomial	Standard beta negative binomial	Beta Poisson
Definition	$\mathcal{B}_n(p) \underset{p}{\wedge} \beta(a,b)$	$\mathcal{NB}(r,p) \underset{p}{\wedge} \beta(a,b)$	$\mathcal{P}(\lambda p) \underset{p}{\wedge} \beta(a,b)$
Notation	$\beta\mathcal{B}_n(a,b)$	$\beta\mathcal{NB}(r,a,b)$	$\beta\mathcal{P}_\lambda(a,b)$
Re-parametrization		UGWD(r,b,a)	
Supp(Y)	$\{0,1,\ldots,n\}$	$\mathbb{N}$	$\mathbb{N}$
$P(Y=y)$	$\frac{(b)_n}{(a+b)_n}\frac{(-n)_y(a)_y}{(-b-n+1)_y}\frac{1}{y!}$	$\frac{(a)_b}{(a+r)_b}\frac{(r)_y(b)_y}{(r+a+b)_y}\frac{1}{y!}$ [5]	$P(Y=y)=\frac{(a)_y}{(a+b)_y}\frac{\lambda^y}{y!}\,{}_1F_1(a+y;a+b+y;-\lambda)$
E(Y)	$n\frac{a}{a+b}$	$r\frac{b}{a-1}$ [6]	$\lambda\frac{a}{a+b}$
V(Y)	$n\frac{ab(a+b+n)}{(a+b)^2(a+b+1)}$	$r\frac{b(a+r-1)(a+b-1)}{(a-1)^2(a-2)}$ [7]	$\lambda\frac{a}{a+b}\left\{1+\lambda\frac{b}{(a+b)(a+b+1)}\right\}$
$G_Y(s)$	$\frac{(b)_n}{(a+b)_n}\,{}_2F_1\{(-n,a);-b-n+1;s\}$	$\frac{(a)_r}{(a+b)_r}\,{}_2F_1\{(r,b);r+a+b;s\}$	${}_1F_1\{a;a+b;\lambda(s-1)\}$

Table 2: Usual characteristics of the standard beta compound binomial, the standard beta compound negative binomial and the beta Poisson distributions.

Name	Generalized beta binomial	Generalized beta negative binomial
Definition	$\mathcal{B}_n(p) \underset{p}{\wedge} \beta_\pi(a,b)$	$\mathcal{NB}(r,p) \underset{p}{\wedge} \beta_\pi(a,b)$
Notation	$\beta_\pi\mathcal{B}_n(a,b)$	$\beta_\pi\mathcal{NB}(r,a,b)$
Supp(Y)	$\{0,1,\ldots,n\}$	$\mathbb{N}$
$P(Y=y)$	$\frac{(b)_n}{(a+b)_n}\frac{(-n)_y(a)_y}{(-b-n+1)_y}\frac{\pi^y}{y!}\,{}_2F_1\{(-n+y,a+y);-b-n+1+y;1-\pi\}$	$\frac{(a)_r}{(a+b)_r}\frac{(r)_y(b)_y}{(r+a+b)_y}\frac{\pi^y}{y!}\,{}_2F_1\{(r+y,b+y);r+a+b+y;1-\pi\}$
E(Y)	$n\pi\frac{a}{a+b}$	$r\pi\frac{b}{a-1}$ [8]
V(Y)	$n\pi\frac{a}{a+b}\left\{\pi\frac{b(a+b+n)}{(a+b)(a+b+1)}+1-\pi\right\}$	$r\pi\frac{b}{a-1}\left\{\pi\frac{(a+r-1)(a+b-1)}{(a-2)(a-1)}+1-\pi\right\}$ [9]
$G_Y(s)$	$\frac{(b)_n}{(a+b)_n}\,{}_2F_1\{(-n,a);-b-n+1;1+\pi(s-1)\}$	$\frac{(a)_r}{(a+b)_r}\,{}_2F_1\{(r,b);r+a+b;1+\pi(s-1)\}$

Table 3: Usual characteristics of the generalized beta binomial and the generalized beta negative binomial distributions.

2. Characteristics of specific convolution splitting distributions

(a)

Distribution	$\mathcal{M}_{\Delta_N}(\boldsymbol{\pi}) \underset{N}{\wedge} \beta\mathcal{B}_n(a,b)$
Supp($\boldsymbol{Y}$)	$\blacktriangle_n$
$P(\boldsymbol{Y}=\boldsymbol{y})$	$\binom{n}{\boldsymbol{y}}\frac{B(\lvert\boldsymbol{y}\rvert+a,n-\lvert\boldsymbol{y}\rvert+b)}{B(\alpha,b)}\prod_{j=1}^{J}\pi_j^{y_j}$
E($\boldsymbol{Y}$)	$n\frac{a}{a+b}\cdot\boldsymbol{\pi}$
Cov($\boldsymbol{Y}$)	$n\frac{a}{a+b}\cdot\left\{\mathrm{diag}(\boldsymbol{\pi})+\frac{b(n-1)-a(a+b+1)}{(a+b)(a+b+1)}\cdot\boldsymbol{\pi}\boldsymbol{\pi}^t\right\}$
$G_{\boldsymbol{Y}}(\boldsymbol{s})$	${}_2F_1\{(-n,a);\alpha+b;1-\boldsymbol{\pi}^t\boldsymbol{s}\}$
Marginals	$Y_j\sim\beta_{\pi_j}\mathcal{B}_n(a,b)$

(b)

Distribution	$\mathcal{M}_{\Delta_N}(\boldsymbol{\pi}) \underset{N}{\wedge} \beta\mathcal{NB}(r,a,b)$
Supp($\boldsymbol{Y}$)	$\mathbb{N}^J$
$P(\boldsymbol{Y}=\boldsymbol{y})$	$\binom{\lvert\boldsymbol{y}\rvert+r-1}{\boldsymbol{y},r-1}\frac{B(r+a,\lvert\boldsymbol{y}\rvert+b)}{B(a,b)}\prod_{j=1}^{J}\pi_j^{y_j}$
E($\boldsymbol{Y}$)	$r\frac{b}{a-1}\cdot\boldsymbol{\pi}$ [10]
Cov($\boldsymbol{Y}$)	$r\frac{b}{a-1}\cdot\left\{\mathrm{diag}(\boldsymbol{\pi})+\frac{a(b+r+1)+r(b-1)-b-1}{(a-1)(a-2)}\cdot\boldsymbol{\pi}\boldsymbol{\pi}^t\right\}$ [11]
$G_{\boldsymbol{Y}}(\boldsymbol{s})$	$\frac{a_{(b)}}{(r+a)_{(b)}}\,{}_2F_1\left\{(r,b);r+a+b;\boldsymbol{\pi}^t\boldsymbol{s}\right\}$
Marginals	$Y_j\sim\beta_{\pi_j}\mathcal{NB}(r,a,b)$

(c)

Distribution	$\mathcal{M}_{\Delta_N}(\boldsymbol{\pi}) \underset{N}{\wedge} \beta_\lambda\mathcal{P}(a,b)$
Supp($\boldsymbol{Y}$)	$\mathbb{N}^J$
$P(\boldsymbol{Y}=\boldsymbol{y})$	$\frac{(a)_{\lvert\boldsymbol{y}\rvert}\lambda^{\lvert\boldsymbol{y}\rvert}}{(a+b)_{\lvert\boldsymbol{y}\rvert}}\,{}_1F_1(a+\lvert\boldsymbol{y}\rvert,a+b+\lvert\boldsymbol{y}\rvert;-\lambda)\prod_{j=1}^{J}\frac{\pi_j^{y_j}}{y_j!}$
E($\boldsymbol{Y}$)	$\lambda\frac{a}{a+b}\cdot\boldsymbol{\pi}$
Cov($\boldsymbol{Y}$)	$\lambda\frac{a}{a+b}\cdot\left\{\mathrm{diag}(\boldsymbol{\pi})+\lambda\frac{b}{(a+b)(a+b+1)}\cdot\boldsymbol{\pi}\boldsymbol{\pi}^t\right\}$
$G_{\boldsymbol{Y}}(\boldsymbol{s})$	${}_1F_1\left\{a;a+b;\lambda(\boldsymbol{\pi}^t\boldsymbol{s}-1)\right\}$
Marginals	$Y_j\sim\beta_{\pi_j\lambda}\mathcal{P}(a,b)$

Table 4: Characteristics of multinomial splitting (a) beta binomial, (b) beta negative binomial and (c) beta Poisson distribution.

Distribution	$\mathcal{DM}_{\Delta_N}(\boldsymbol{\alpha}) \underset{N}{\wedge} \beta_\lambda \mathcal{P}(a,b)$	
Constraint	no constraint	$a = \lvert\boldsymbol{\alpha}\rvert$
$\mathrm{Supp}(\boldsymbol{Y})$	$\mathbb{N}^J$	$\mathbb{N}^J$
$P(\boldsymbol{Y}=\boldsymbol{y})$	$\frac{(a)_{\lvert\boldsymbol{y}\rvert}\lambda^{\lvert\boldsymbol{y}\rvert}}{(a+b)_{\lvert\boldsymbol{y}\rvert}(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^J \frac{(\alpha_j)_{y_j}}{y_j!} {}_1F_1(a+\lvert\boldsymbol{y}\rvert; a+b+\lvert\boldsymbol{y}\rvert; -\lambda)$	$\frac{\lambda^{\lvert\boldsymbol{y}\rvert}}{(a+b)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^J \frac{(\alpha_j)_{y_j}}{y_j!} {}_1F_1(a+\lvert\boldsymbol{y}\rvert; a+b+\lvert\boldsymbol{y}\rvert; -\lambda)$
$\mathrm{E}(\boldsymbol{Y})$	$\frac{\lambda a}{\lvert\boldsymbol{\alpha}\rvert(a+b)} \cdot \boldsymbol{\alpha}$	$\frac{\lambda}{\lvert\boldsymbol{\alpha}\rvert+b} \cdot \boldsymbol{\alpha}$
$\mathrm{Cov}(\boldsymbol{Y})$	$\frac{\lambda a}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)(a+b)} \cdot \Big[\Big\{ \frac{\lambda b}{(a+b)(a+b+1)} + \frac{\lambda a}{a+b} + \lvert\boldsymbol{\alpha}\rvert + 1 \Big\}$ $\cdot \mathrm{diag}(\boldsymbol{\alpha}) + \Big\{ \frac{\lambda b}{(a+b)(a+b+1)} - \frac{\lambda a}{\lvert\boldsymbol{\alpha}\rvert(a+b)} \Big\} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \Big]$	$\frac{\lambda}{(\lvert\boldsymbol{\alpha}\rvert+1)(a+b)} \cdot \Big[\Big\{ \frac{\lambda b}{(\lvert\boldsymbol{\alpha}\rvert+b)(\lvert\boldsymbol{\alpha}\rvert+b+1)} + \frac{\lambda\lvert\boldsymbol{\alpha}\rvert}{\lvert\boldsymbol{\alpha}\rvert+b} + \lvert\boldsymbol{\alpha}\rvert + 1 \Big\}$ $\cdot \mathrm{diag}(\boldsymbol{\alpha}) + \Big\{ \frac{\lambda b}{(\lvert\boldsymbol{\alpha}\rvert+b)(\lvert\boldsymbol{\alpha}\rvert+b+1)} - \frac{\lambda}{a+b} \Big\} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \Big]$
$G_{\boldsymbol{Y}}(\boldsymbol{s})$	$\sum_{\boldsymbol{y}\in\mathbb{N}^J} \sum_{k\in\mathbb{N}} \frac{(a)_{\lvert\boldsymbol{y}\rvert+k} \prod_{j=1}^J (\alpha_j)_{y_j}}{(a+b)_{\lvert\boldsymbol{y}\rvert+k}(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \frac{(-\lambda)^k}{k!} \prod_{j\in\mathcal{J}} \frac{(\lambda s_j)^{y_j}}{y_j!}$	${}_0^{J+1}F_1\{(\boldsymbol{\alpha}, a+\lvert\boldsymbol{y}\rvert); a+b; (\lambda\cdot\boldsymbol{s}, -\lambda)\}$
Marginals	$Y_j \sim \beta_\lambda^2 \mathcal{P}\big(\alpha_j, \lvert\boldsymbol{\alpha}_{-j}\rvert, a, b\big)$	$Y_j \sim \beta_\lambda \mathcal{P}\big(\alpha_j, \lvert\boldsymbol{\alpha}_{-j}\rvert + b\big)$[1]

Table 5: Usual characteristics of Dirichlet multinomial splitting beta Poisson distribution respectively without constraint and with $a = \lvert\alpha\rvert$.

Distribution	$\mathcal{DM}_{\Delta_N}(\boldsymbol{\alpha}) \underset{N}{\wedge} \mathcal{L}(\psi)$		
$\mathcal{L}(\psi)$	$\mathcal{B}_n(p)$	$\mathcal{NB}(r,p)$[12]	$\mathcal{P}(\lambda)$
$\mathrm{Supp}(\boldsymbol{Y})$	$\blacktriangle_n$	$\mathbb{N}^J$	$\mathbb{N}^J$
$P(\boldsymbol{Y}=\boldsymbol{y})$	$\frac{\Gamma(n+1)p^{\lvert\boldsymbol{y}\rvert}(1-p)^{n-\lvert\boldsymbol{y}\rvert}}{\Gamma(n-\lvert\boldsymbol{y}\rvert+1)(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^J \frac{(\alpha_j)_{y_j}}{y_j!}$	$(1-p)^r \frac{(r)_{\lvert\boldsymbol{y}\rvert}p^{\lvert\boldsymbol{y}\rvert}}{(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^J \frac{(\alpha_j)_{y_j}}{y_j!}$	$e^{-\lambda} \frac{\lambda^{\lvert\boldsymbol{y}\rvert}}{(\lvert\boldsymbol{\alpha}\rvert)_{\lvert\boldsymbol{y}\rvert}} \prod_{j=1}^J \frac{(\alpha_j)_{y_j}}{y_j!}$
$\mathrm{E}(\boldsymbol{Y})$	$\frac{np}{\lvert\boldsymbol{\alpha}\rvert} \cdot \boldsymbol{\alpha}$	$\frac{rp}{\lvert\boldsymbol{\alpha}\rvert(1-p)} \cdot \boldsymbol{\alpha}$	$\frac{\lambda}{\lvert\boldsymbol{\alpha}\rvert} \cdot \boldsymbol{\alpha}$
$\mathrm{Cov}(\boldsymbol{Y})$	$\frac{np}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)} \cdot \{ \{(n-1)p + \lvert\boldsymbol{\alpha}\rvert + 1\}$ $\cdot \mathrm{diag}(\boldsymbol{\alpha}) - \frac{p(n+\lvert\boldsymbol{\alpha}\rvert)}{\lvert\boldsymbol{\alpha}\rvert} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \}$	$\frac{rp}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)(1-p)} \cdot \big[\frac{(r-\lvert\boldsymbol{\alpha}\rvert)p+\lvert\boldsymbol{\alpha}\rvert+1}{1-p}$ $\cdot \mathrm{diag}(\boldsymbol{\alpha}) + \frac{(\lvert\boldsymbol{\alpha}\rvert-r)p}{\lvert\boldsymbol{\alpha}\rvert(1-p)} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \big]$	$\frac{\lambda}{\lvert\boldsymbol{\alpha}\rvert(\lvert\boldsymbol{\alpha}\rvert+1)} \cdot \{(\lambda + \lvert\boldsymbol{\alpha}\rvert + 1) \cdot \mathrm{diag}(\boldsymbol{\alpha})$ $- \frac{\lambda}{\lvert\boldsymbol{\alpha}\rvert} \cdot \boldsymbol{\alpha}\boldsymbol{\alpha}^t \Big\}$
$G_{\boldsymbol{Y}}(\boldsymbol{s})$	$(1-p)^n \, {}_1^JF_1(-n; \boldsymbol{\alpha}; \lvert\boldsymbol{\alpha}\rvert; -\frac{p}{1-p} \cdot \boldsymbol{s})$	$(1-p)^r \, {}_1^JF_1(r; \boldsymbol{\alpha}; \lvert\boldsymbol{\alpha}\rvert; p \cdot \boldsymbol{s})$	$e^{-\lambda} \, {}_0^JF_1(\boldsymbol{\alpha}; \lvert\boldsymbol{\alpha}\rvert; \lambda \cdot \boldsymbol{s})$
Marginals	$Y_j \sim \beta_p \mathcal{B}_n(\alpha_j, \lvert\boldsymbol{\alpha}_{-j}\rvert)$	$Y_j \sim \beta_p \mathcal{NB}(r, \alpha_j, \lvert\boldsymbol{\alpha}_{-j}\rvert)$	$Y_j \sim \beta_\lambda \mathcal{P}(\alpha_j, \lvert\boldsymbol{\alpha}_{-j}\rvert)$

Table 6: Usual characteristics of Dirichlet multinomial splitting binomial, negative binomial and Poisson distribution.

3. Inference of singular and univariate regressions

(a)

Regression	Notation	Canonical link function	Inference
Multinomial	$\mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(\boldsymbol{x})\}$	$\pi_j = \frac{\exp(\boldsymbol{x}^t\boldsymbol{\beta}_j)}{1+\sum_{k=1}^{J-1}\exp(\boldsymbol{x}^t\boldsymbol{\beta}_k)},\ j = 1,\ldots,J-1$	See [37]
Dirichlet multinomial	$\mathcal{DM}_{\Delta_n}\{\boldsymbol{\alpha}(\boldsymbol{x})\}$	$\alpha_j = \exp(\boldsymbol{x}^t\boldsymbol{\beta}_j),\ j = 1,\ldots,J$	See [55]
Generalized Dirichlet multinomial	$\mathcal{GDM}_{\Delta_n}\{\boldsymbol{\alpha}(\boldsymbol{x}),\boldsymbol{\beta}(\boldsymbol{x})\}$	$a_j = \exp(\boldsymbol{x}^t\boldsymbol{\beta}_{1,j}),\ j = 1,\ldots,J-1$ $b_j = \exp(\boldsymbol{x}^t\boldsymbol{\beta}_{2,j}),\ j = 1,\ldots,J-1$	See [55]
Logistic normal multinomial	$\mathcal{LNM}_{\Delta_n}\{\boldsymbol{\mu}(\boldsymbol{x}),\Sigma\}$	$\mu_j = \boldsymbol{x}^t\boldsymbol{\beta}_j,\ j = 1,\ldots,J-1$ $\pi_j = \frac{\exp(\mu_j)}{1+\sum_{k=1}^{J-1}\exp(\mu_k)},\ j = 1,\ldots,J-1$	See [53]

(b)

Regression	Notation	Canonical link function	Parameter inference
Poisson	$\mathcal{P}\{\lambda(\boldsymbol{x})\}$	$\lambda = \exp(\boldsymbol{x}^t\boldsymbol{\beta})$	See [30]
Binomial	$\mathcal{B}_n\{p(\boldsymbol{x})\}$	$p = \frac{\exp(\boldsymbol{x}^t\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}^t\boldsymbol{\beta})}$	See [30] for n known
Negative binomial	$\mathcal{NB}\{r, p(\boldsymbol{x})\}$	$p = \exp(\boldsymbol{x}^t\boldsymbol{\beta})$	See [17]
Beta Poisson	$\beta\mathcal{P}\{a(\boldsymbol{x}), b(\boldsymbol{x})\}$	$\frac{a}{a+b} = \frac{\exp(\boldsymbol{x}^t\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}^t\boldsymbol{\beta})}$	See [48]
Beta binomial	$\beta\mathcal{B}_n\{a(\boldsymbol{x}), b(\boldsymbol{x})\}$	$\frac{a}{a+b} = \frac{\exp(\boldsymbol{x}^t\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}^t\boldsymbol{\beta})}$	See [15] and [29] for n known
Beta negative binomial	$\beta\mathcal{NB}\{r, a(\boldsymbol{x}), b(\boldsymbol{x})\}$	$\frac{a}{a+b} = \frac{\exp(\boldsymbol{x}^t\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}^t\boldsymbol{\beta})}$	See [40, 41]

Table 7: References of inference procedures for (a) singular regressions and (b) univariate regressions.