Pólya urn and multivariate birth-death processes under neutral theory of biodiversity

Jean Peyhardi [1] & Fabien Laroche [2], Frédéric Mortier [3]

[1] *IMAG, University of Montpellier, CNRS, Montpellier, France*
[2] *MR DYNAFOR, INP de Toulouse, INRAE, Auzeville Tolosane, France*
[3] *CIRAD, UPR Forêts et Sociétés, F-34398 Montpellier, France*

Résumé. La famille des modèles joints de distribution d'espèces est devenue un outil statistique utile pour analyser les données multivariées d'abondance d'espèces. La théorie neutre unifiée de la biodiversité introduite par Hubbell et al. (2001) souligne l'importance des processus stochastiques dans la structure des communautés écologiques. Sous l'hypothèse de taille de communauté fixée et l'hypothèse de neutralité (les individus de différentes espèces sont écologiquement identiques), il a montré que la distribution stationnaire d'un processus multivarié de naissance/mort est une Dirichlet multinomiale. Etienne et al. (2007) ont relaché l'hypothèse de somme fixe et ont également trouvé une distribution Dirichlet multinomiale. Mais ils ont ajouté l'hypothèse d'indépendance entre espèces. Nous proposons de généraliser cette approche, en s'affranchissant de l'hypothèse d'indépendance et en considérant la famille générale des distributions de Pólya. L'indépendance est donc une conséquence des hypothèses paramétriques faites sur les taux de natalité/mortalité et non plus une hypothèse. Nous mettons également en évidence neuf distributions remarquables de Pólya (somme aléatoire) qui sont stables par marginalisation. Nous donnons la forme paramétrique des taux de saut conduisant à ces neuf distributions, en incluant le cas particulier de la Dirichlet multinomiale (somme binomiale négative), obtenu par Etienne et al. (2007).

Mots-clés. Theorie neutre, Modèle joint de distribution d'espèces, Urnes de Pólya

Abstract. The family of joint species distribution models (JSDMs) has emerged as a useful statistical tool to analyse multivariate species abundance data. The unified neutral theory of biodiversity introduced by Hubbell et al. (2001) emphasizes the importance of stochastic processes in ecological community structure. Under the zero-sum assumption (fixed community size) and the neutrality assumption (individuals of different species are ecologically identical) he showed that the stationary distribution of a multivariate birth and death process is a Dirichlet multinomial. Etienne et al. (2007) relaxed the zero-sum assumption and also found a Dirichlet multinomial distribution. But they added the assumption of independence between species. We propose to generalize this approach in two ways, by relaxing the independence assumption and by considering the enlarged family of Pólya distributions. Independence is thus a consequence of parametric assumption made on birth/death rates and not a necessary assumption. We also highlight nine remarkable Pólya splitting distributions that are closed under all marginalizations. Then we give the parameteric form of jumping rates leading to these nine distributions, recovering the special case Dirichlet multinomial split with negative binomial sum obtained by Etienne et al. (2007).

Keywords. Neutral theory, Joint Species Distribution Model, Pólya urn

1 Biological context

Biodiversity is not only determined by the number of different entities (species in community ecology or alleles in population genetics), but also by the abundance of these entities (Magurran, 2021). Many ecological questions require the joint analysis of abundances collected simultaneously across many taxonomic groups. The family of joint species distribution models (JSDMs) has then emerged as a useful statistical tool to analyse such data. When species abundance are collected as count data (instead of presence/absence) a JSDMs turns out to be a multivariate count distribution of a random vector $\boldsymbol{N} = (N_1, \ldots, N_J)$ where J denotes the number of species. The unified neutral theory of biodiversity introduced by Hubbell et al. (2001) emphasizes the importance of stochastic processes in ecological community structure, and has challenged the traditional niche-based view of ecology. It has challenged classical theories of species diversity by showing that patterns of species diversity similar to those observed in nature can be obtained from an extremely simplified model of community dynamics where each dying individual is immediately replaced by a new individual (zero-sum) and all individuals of all species are ecologically identical (neutrality). From a mathematical point of view, focus is made on the stationary distribution of a multivariate birth and death process $\boldsymbol{N}(t) = (N_1(t), \ldots, N_J(t))$, where $N_j(t)$ is the abundance of species j at time t. Under the zero-sum assumption (i.e., fixed sum $|\boldsymbol{N}(t)| = n$) - and mild hypothesis on birth and death rates - Hubbell et al. (2001) showed that the multivariate distribution of $\boldsymbol{N}$ given $|\boldsymbol{N}| = n$ (at equilibrium) is a Dirichlet multinomial distribution $\mathcal{DM}_n(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$. Etienne et al. (2007) proposed to relax the zero-sum assumption in order to obtain a more realistic model and they also find a Dirichlet multinomial distribution at equilibrium for the conditional distribution of $\boldsymbol{N}$ given $|\boldsymbol{N}| = n$. But they added the assumption of independence between species to obtain this result since in this case it is sufficient to study the abundance distribution for each species and then calculate the product. In the present work we propose to generalize this approach in two ways, by relaxing the independence assumption and by considering the enlarged family of multivariate Pólya distributions. This work is based on both papers of Peyhardi (2023) and Peyhardi et al. (2024).

2 Pólya splitting distributions

The class of Pólya splitting distributions has been introduced by Peyhardi et al. (2021) as compound distributions $\boldsymbol{N} \sim \mathcal{P}_n^{[c]}(\boldsymbol{\theta}) \wedge_n \mathcal{L}$, meaning that $|\boldsymbol{N}|$ follows the univariate distribution $\mathcal{L}$ and $\boldsymbol{N}$ given $|\boldsymbol{N}| = n$ follows the multivariate Pólya distribution. Let us briefly recall the definition of a multivariate Pólya distribution in terms of urn models.

2.1 Pólya urn model

One urn initially contains θ_j balls of the color j for $j = 1, \ldots, J$. At each draw, one ball is drawn at random and then replaced with c additional balls of the same color, where $c \in \{-1, 0, 1\}$. This procedure is repeated n times and focus is made on the multivariate count $\boldsymbol{N} = (N_1, \ldots, N_J)$ of drawn balls for each color. Knowing the number n of draws, the conditional count distribution of $\boldsymbol{N}$ given $|\boldsymbol{N}| = n$ is known as the multivariate Pólya distribution, denoted by $\mathcal{P}_n^{[c]}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \Theta_c^J$ (where $\Theta = \mathbb{N}$ for $c = -1$ and $\Theta = \mathbb{R}_+$ otherwise). As expressed by Peyhardi (2023), if we denote

$$a_\theta^{[c]}(n) = \frac{\prod_{k=0}^{n-1}(\theta + ck)}{n!} \mathbb{1}_{\theta+cn\geq 0},$$

then the probability mass function (pmf) of a multivariate Pólya distribution $\mathcal{P}_n^{[c]}(\boldsymbol{\theta})$ takes the following form

$$P_{|\boldsymbol{N}|=n}\left(\boldsymbol{N} = \boldsymbol{n}\right) = \frac{1}{a_{|\boldsymbol{\theta}|}^{[c]}(n)} \prod_{j=1}^{J} a_{\theta_j}^{[c]}(n_j).$$

The mulyivariate Pólya distribution turns out to be the:

- multivariate hypergeometric distribution when $c = -1$ (i.e., without replacement), denoted by $\mathcal{H}_{\Delta_n}(\boldsymbol{\theta})$ with pmf

$$P_{|\boldsymbol{N}|=n}\left(\boldsymbol{N} = \boldsymbol{n}\right) = \frac{1}{\binom{|\boldsymbol{\theta}|}{n}} \prod_{j=1}^{J} \binom{\theta_j}{n_j}.$$

- multinomial distribution when $c = 0$ (i.e., with replacement meaning independent draws), denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ with pmf

$$P_{|\boldsymbol{N}|=n}\left(\boldsymbol{N} = \boldsymbol{n}\right) = \frac{1}{|\boldsymbol{\theta}|/n!} \prod_{j=1}^{J} \frac{\theta_j^{n_j}}{n_j!}.$$

- Dirichlet-multinomial distribution when $c = 1$ (i.e., with reinforcement), denoted by $\mathcal{DM}_{\Delta_n}(\boldsymbol{\theta})$ with pmf

$$P_{|\boldsymbol{N}|=n}\left(\boldsymbol{N} = \boldsymbol{n}\right) = \frac{1}{\binom{n+|\boldsymbol{\theta}|-1}{n}} \prod_{j=1}^{J} \binom{n_j + \theta_j - 1}{n_j}.$$

2.2 Remarkable Pólya splitting distributions

Properties of Pólya splitting distributions are related to the choice of the sum distribution $\mathcal{L}$. For instance the covariance between N_i and N_j (with $(i,j) \in \{1, \ldots, J\}^2$ and $i \neq j$) is given by

$$\mathrm{Cov}(N_i, N_j) = \frac{\theta_i \theta_j}{|\boldsymbol{\theta}|^2(|\boldsymbol{\theta}| + c)} \left[(\mu_2 - \mu_1^2)|\boldsymbol{\theta}| - c\mu_{(1)}^2\right], \tag{1}$$

where μ_k is the factorial moment of order k of the sum distribution $\mathcal{L}$. It implies that the sign of covariance between any pair (i,j) is driven by the moments of the sum; see Table 1 for some examples. In fact it is possible to characterize the probabilistic graphical model (PGM) of a Pólya splitting distribution according $\mathcal{L}$; see Theorem 1. We first need to introduce three remarkable choices for the sum distribution $\mathcal{L}$ (see (Peyhardi, 2023) for details):

1. The (univariate) Pólya distribution $\mathcal{P}_n^{[c]}(\theta, \gamma)$ with pmf

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

   It includes the hypergeometric ($c = -1$), binomial ($c = 0$) and beta binomial ($c = 1$).

2. The power series distribution $\mathcal{PS}^{[c]}(\theta, \alpha)$ with pmf

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

   It includes the binomial ($c = -1$), Poisson ($c = 0$) and negative binomial ($c = 1$).

3. The inverse Pólya distribution $\mathcal{IP}^{[c]}(r; \theta, \gamma)$ with pmf

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

   It includes the beta binomial ($c = -1$), negative binomial ($c = 0$) and negative beta binomial ($c = 1$).

Theorem 1 ((Peyhardi, 2023)) *The PGM of a Pólya splitting distribution* $\mathcal{P}_n^{[c]}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$ *is:*

- *empty if and only if* $\mathcal{L} = \mathcal{PS}^{[c]}(|\boldsymbol{\theta}|, \alpha)$
- *complete otherwise*

This theorem generalizes the result of Bol'shev (1965) and those of Rao and Janardan (1984) that concerns the bivariate case ($J = 2$) only for some thinning operators.

Theorem 2 ((Peyhardi, 2023)) *The Pólya, the power series and the inverse Pólya distributions are closed under the Pólya thinning operator* $\mathcal{P}_n^{[c]}(\theta, \gamma) \wedge_n (\cdot)$ *when their parameter respects the additive constraint. More precisely we have the following distributions equalities:*

1. $\mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{P}_m^{[c]}(\theta + \gamma, \lambda) = \mathcal{P}_m^{[c]}(\theta, \gamma + \lambda)$,
2. $\mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{PS}^{[c]}(\theta + \gamma, \alpha) = \mathcal{PS}^{[c]}(\theta, \alpha)$,
3. $\mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{IP}^{[c]}(r; \theta + \gamma, \lambda) = \mathcal{IP}^{[c]}(r; \theta, \lambda)$,

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

As corollary of Theorem 2, based on specific case $c = 0$, the result of Rao (1965) is recovered.

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

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

Theorem 2 allows us to build nine remarkable Pólya splitting distributions (see Table 1) that are closed under all marginalizations. According to equation (1) the sign of covariance is easily obtained in this nine examples.

Split Sum	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$	Covariance sign
Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$	negative
Power series	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,p)$	null
Inverse Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a,b)$	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a,p)$	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,a,b)$	positive

Table 1: Nine remarkable Pólya splitting distributions with different split distributions (columns) and different sum distributions (rows).

3 Multivariate birth-death processes under neutrality

We will now exhibit the birth and death rates assumptions that lead to the multivariate Pólya distribution at equibrium. The master equation describing the behaviour of the multivariate jump process $\boldsymbol{N}(t)$ is given by

$$\frac{\partial p_{\boldsymbol{n}}(t)}{\partial t} = \sum_{j=1}^{J} p_{\boldsymbol{n}-\boldsymbol{e}_j}(t) q_j^-(\boldsymbol{n}-\boldsymbol{e}_j) + p_{\boldsymbol{n}+\boldsymbol{e}_j}(t) q_j^+(\boldsymbol{n}+\boldsymbol{e}_j) - p_{\boldsymbol{n}+\boldsymbol{e}_j}(t)\{q_j^-(\boldsymbol{n}) + q_j^+(\boldsymbol{n})\}$$

where $q_j^-(\boldsymbol{n})$ (resp. $q_j^+(\boldsymbol{n})$) denotes the jumping rate from $\boldsymbol{n}$ to $\boldsymbol{n}-\boldsymbol{e}_j$ (resp. to $\boldsymbol{n}+\boldsymbol{e}_j$), $\boldsymbol{e_j}$ denotes the indicator vector of the jth element and $p_{\boldsymbol{n}}(t) = P\{\boldsymbol{N}(t) = \boldsymbol{n}\}$ (resp. $p_{\boldsymbol{n}} = P(\boldsymbol{N} = \boldsymbol{n})$) denotes the pmf at time t (resp. the pmf at stationary state). Assume that there exists some parameters $c \in \{-1,0,1\}$ and $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_J) \in \Theta_c^J$ and two non-negative functions s^+ and s^- such that the birth and death rates have the following form

$$\begin{aligned} q_j^+(\boldsymbol{n}) &= s^+(|\boldsymbol{n}|)(\theta_j + c n_j)\mathbb{1}_{\theta_j + c n_j \geq 0}, \\ q_j^-(\boldsymbol{n}) &= s^-(|\boldsymbol{n}|) n_j. \end{aligned} \tag{2}$$

The birth-deah rate $q_j(\boldsymbol{n}) := q_j^+(\boldsymbol{n})/q_j^-(\boldsymbol{n}+\boldsymbol{e_j})$ thus becomes

$$q_j(\boldsymbol{n}) = s(|\boldsymbol{n}|)\frac{\theta_j + cn_j}{n_j + 1}\mathbb{1}_{\theta_j + cn_j \geq 0} \tag{3}$$

where $s(n) = \frac{s^+(n)}{s^-(n+1)}$ for all $n \in \mathbb{N}$. It can be seen that this parametric assumption (3) respects the Kolmogorov's criterion $q_i(\boldsymbol{n})q_j(\boldsymbol{n}+\boldsymbol{e_i}) = q_j(\boldsymbol{n})q_i(\boldsymbol{n}+\boldsymbol{e_j})$, and thus leads to a reversible process. Remarking that $\prod_{k=0}^{n-1}\frac{\theta+ck}{k+1} = a_\theta^{[c]}(n)$ we add the following assumption on $s(n)$ in order to obtain a well defined stationary distribution:

$$\sum_{n\geq 0} a_\theta^{[c]}(n)\prod_{k=0}^{n-1} s(k) < \infty. \tag{4}$$

Theorem 3 ((Peyhardi et al., 2024)) *Assume that the hypothesis* (3) *and* (4) *hold then*

- *the stationary distribution of* $\boldsymbol{N}(t)$ *is the Pólya splitting distribution* $\mathcal{P}_{\Delta_n}^{[c]}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$
- $\mathcal{L}$ *is the stationary distribution of a univariate process with birth/death ratio equal to* $q(n) = s(n)r_{|\boldsymbol{\theta}|}^{[c]}(n)$*, more precisely we have* $P(|\boldsymbol{N}| = n) \propto a_{|\boldsymbol{\theta}|}^{[c]}(n)\prod_{k=0}^{n-1} s(k)$.

Theorem 1 characterizes the independence between variable $N_1, \ldots, N_J$ through the distribution of the sum $|\boldsymbol{N}|$. In terms of multivariate birth-death process the sum distribution (at equilibrium) is governed by the quantity $s(|\boldsymbol{n}|)$ that appears in jumping rates form (4). The assumption $s(|\boldsymbol{n}|) = \alpha$ thus leads to the power series distribution $|\boldsymbol{N}| \sim \mathcal{PS}^{[c]}(|\boldsymbol{\theta}|, \alpha)$ and then to independence between species. Independence is a consequence of parametric assumption made on birth and death rates and not a necessary assumption *per se*, contrary to what was posited by other authors (Etienne et al., 2007).

Theorem 2 allows us to build nine remarkable Pólya splitting distributions (see Table 1) that are closed under all marginalizations. Using our framework, the parameteric form of jumping rates leading to these distributions are easily obtained (see Table 2). Moreover it has been shown that covariance between two species abundancies $\text{Cov}(N_i, N_j)$ have the same sign for any pair of species and are negative (resp. null or positive) when the sum distribution is the Pólya (resp. the power series or the inverse Pólya).

Using Theorem 3, the parameteric form of jumping rates leading to these distributions are easily obtained (see Table 2). Overall, we advocate that Pólya-splitting distribution should become a part of the classic toolbox for the analysis of multivariate count data in ecology, providing alternative approaches to JSDM framework. Moreover, explanatory variables are easily taken into account in this framework (with a decomposable likelihood according to the compound distribution) leading to competitve statistical models compare to classical Poisson log-normal model that have not undelying process interpretation.

Split Sum	Hypergeometric $c = -1$	Multinomial $c = 0$	Dirichlet multinomial $c = 1$	Covariance sign
Pólya	$\frac{m - \lvert\boldsymbol{n}\rvert}{\gamma - m + \lvert\boldsymbol{n}\rvert + 1}$	$\frac{m - \lvert\boldsymbol{n}\rvert}{\gamma}$	$\frac{m - \lvert\boldsymbol{n}\rvert}{\gamma + m - \lvert\boldsymbol{n}\rvert - 1}$	negative
Power series	α	α	α	null
Inverse Pólya	$\frac{a + \lvert\boldsymbol{n}\rvert}{\lvert\boldsymbol{\theta}\rvert + b - \lvert\boldsymbol{n}\rvert - 1}$	$\frac{a + \lvert\boldsymbol{n}\rvert}{\lvert\boldsymbol{\theta}\rvert + b}$	$\frac{a + \lvert\boldsymbol{n}\rvert}{\lvert\boldsymbol{\theta}\rvert + b + a + \lvert\boldsymbol{n}\rvert + 1}$	positive

Table 2: Parametric form of $s(|\boldsymbol{n}|)$ and thus of jumping rates $q_j(\boldsymbol{n}) = q_j^+(\boldsymbol{n})/q_j^-(\boldsymbol{n} + \boldsymbol{e}_j)$ leading to the nine remarkable Pólya splitting distribution of Table 1 at equilibrium.

4 Discussion

Our main contribution is to connect the class of Polya-splitting distribution to the neutral theory of biodiversity in ecology, a useful null model allowing the evaluation of non-neutral processes such as adaptation or natural selection (Alonso et al., 2006). We found that for any Polya-splitting distribution, there exists a multivariate jump process of neutral species with such stationnary distribution. However, staying at the very general level for the sum distribution, the associated transition rates may not have a straightforward biological interpretation. We therefore exhibited, nine transition rates parametrization with meaningful biological interpretation leading to usual parametric distributions.

However, Pólya splitting distributions induce only two types of dependence structures: either all species are independent or fully dependent with homogenized correlation sign. To extend this binary setting towards more complex nested dependence structures between species or communities, we suggest the use of recursive application of splitting distributions along a partition tree of species.

Otherwise, the inclusion of environmental factors in Pólya splitting distributions is a natural extension. It could be achieved assuming a regression model for the sum distribution and another for the Pólya distribution (see Peyhardi et al. (2021) for more details in the multinomial splitting regression context).

Finally, combining a partition tree approach with the inclusion of environmental covariates at each node leads to propose nested multi-level inhomogenous splitting models. Such models should be interesting alternatives to classical approaches used in joint species distribution contexts mainly based on conditional Independence's (Warton et al., 2015; Ovaskainen and Abrego, 2020) and the use of the multivariate Poisson log-Normal distribution. Comparatively, our approach allows to model dependencies between species at the observation level, while the classical JSDM's model dependencies at the latent process strata.

References

Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. Trends in Ecology & Evolution 21, 451–457. doi:10.1016/j.tree.2006.03.019.

Bol'shev, L.N., 1965. On a characterization of the poisson distribution and its statistical applications. Theory of Probability & Its Applications 10, 446–456.

Etienne, R., Alonso, D., McKane, A., 2007. The zero-sum assumption in neutral biodiversity theory. Journal of theoretical biology. 248, 522–536.

Hubbell, S.P., et al., 2001. The unified neutral theory of biodiversity and biogeography. volume 32. Princeton University Press Princeton.

Magurran, A.E., 2021. Measuring biological diversity. Current Biology 31, R1174–R1177.

Ovaskainen, O., Abrego, N., 2020. Joint Species Distribution Modelling: With Applications in R. Ecology, Biodiversity and Conservation, Cambridge University Press. doi:10.1017/9781108591720.

Peyhardi, J., 2023. On quasi Pólya thinning operator. Brazilian Journal of Probability and Statistics 37, 643 – 666.

Peyhardi, J., Fernique, P., Durand, J.B., 2021. Splitting models for multivariate count data. Journal of Multivariate Analysis 181, 104677.

Peyhardi, J., Laroche, F., Mortier, F., 2024. Pólya-splitting distributions as stationary solutions of multivariate birth-death processes under extended neutral theory. Journal of Theoretical Biology .

Rao, B.R., Janardan, K., 1984. The use of the generalized markov-polya distribution as a random damage model and its identifiability. Sankhyā: The Indian Journal of Statistics, Series A , 458–462.

Rao, C.R., 1965. On discrete distributions arising out of methods of ascertainment. Sankhyā: The Indian Journal of Statistics, Series A , 311–324.

Warton, D., Blanchet, F., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S., Hui, F., 2015. So many variables: Joint modeling in community ecology. Trends in Ecology and Evolution 30, 766–779.