Polya urn models for multivariate species abundances: properties and applications.

Jean Peyhardi, Frédéric Mortier, Fabien Laroche,
Samuel Valiquette, Fabrice Moudjeu

June 20, 2024

Satistical context

- Object: Multivariate species abundances

  $\Leftrightarrow$ Multivariate count data

  $\boldsymbol{Y} = (Y_1, \ldots, Y_J) \in \mathbb{N}^J$ where J =number of species

Satistical context

- Object: Multivariate species abundances

  $\Leftrightarrow$ Multivariate count data

  $\boldsymbol{Y} = (Y_1, \ldots, Y_J) \in \mathbb{N}^J$ where J =number of species

- Different modeling approaches
  1. Discrete copulas
     $\rightarrow$ Slar theorem does not hold ... copula is not unique !
  2. Poisson graphical model $Y_j \sim \mathcal{P}\left\{\exp(\eta_j)\right\}$, $\eta_j = \boldsymbol{y}_{-j}^T \beta_j$
     $\rightarrow$ implies negative constraint on predictors η_j ...
  3. Poisson log-normal: $\boldsymbol{Y} \sim \mathcal{PLN}(\boldsymbol{\mu}, \Sigma)$
     $\rightarrow$ dependences on latent variables $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$
  4. Mixed multinomial $\boldsymbol{Y} \sim \mathcal{M}(n, \boldsymbol{\pi}) \underset{\boldsymbol{\pi}}{\wedge} \mathcal{D}(\boldsymbol{\theta})$
     $\rightarrow$ dependences on observed variable $\boldsymbol{Y}$

Outline

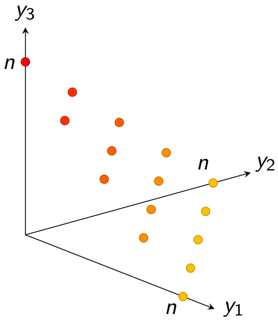
1. Polya splitting models
   - Singular distribution
   - Polya urn
   - Closure under thinning operation
   - Probabilistic graphical model
2. Neutral theory of biodiversity
   - State of the art
   - Link with Polya urn
3. Statistical point of view
   - Regression framework
   - Tree structure
   - Application on real data

Outline

Singular distributions

Definition

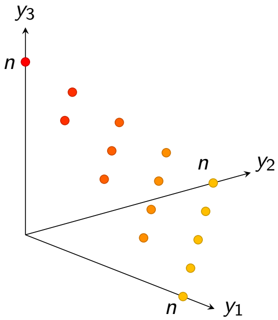
Multinomial distribution $\mathcal{M}(n, \boldsymbol{\pi})$ is supported on the discrete simplexe $\Delta_n = \{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| = n\}$.



Singular distributions

Definition

Multinomial distribution $\mathcal{M}(n, \boldsymbol{\pi})$ is supported on the discrete simplexe $\Delta_n = \{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| = n\}$.



Problem

$\dim(\Delta_n) = J - 1$

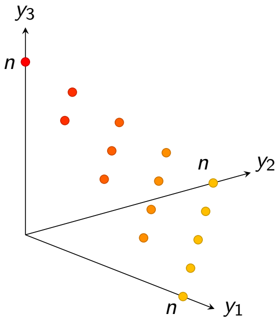
Singular distributions

Definition

Multinomial distribution $\mathcal{M}(n, \boldsymbol{\pi})$ is supported on the discrete simplexe $\Delta_n = \{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| = n\}$.
It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)$.

Problem

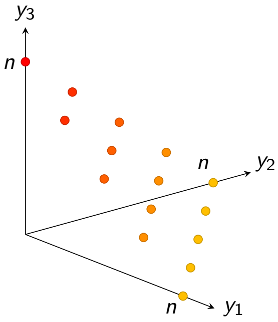
$\dim(\Delta_n) = J - 1$



Singular distributions

Definition

Multinomial distribution $\mathcal{M}(n, \boldsymbol{\pi})$ is supported on the discrete simplexe $\Delta_n = \{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| = n\}$.
It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi})$ where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$.



Problem

$\dim(\Delta_n) = J - 1$

Remark

compositional distribution	$\rightarrow$	singular distribution
continuous simplexe Δ		discrete simplexe Δ_n
$\boldsymbol{\pi} \sim \mathcal{D}$		$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{\boldsymbol{\pi}}{\wedge} \mathcal{D}$

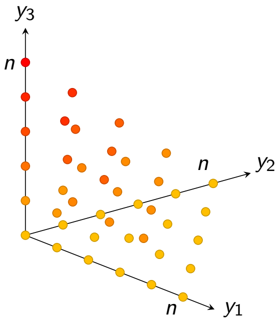
Non-singular version

Definition

$$\begin{array}{ll} & \boldsymbol{Y} \text{ follows non-singular multinomial} \\ \Leftrightarrow & (\boldsymbol{Y}, n - |\boldsymbol{Y}|) \text{ follows multinomial} \end{array}$$

Supported on the interior of Δ_n

$\blacktriangle_n = \left\{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| \leq n\right\}$.



Non-singular version

Definition

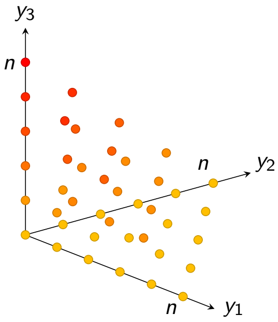
$$\begin{aligned} & \boldsymbol{Y} \text{ follows non-singular multinomial} \\ \Leftrightarrow \quad & (\boldsymbol{Y}, n - |\boldsymbol{Y}|) \text{ follows multinomial} \end{aligned}$$

Supported on the interior of Δ_n

$\blacktriangle_n = \left\{ \boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| \leq n \right\}$.

Remark

$\dim(\blacktriangle_n) = J$



Non-singular version

Definition

$$\begin{array}{ll} & \boldsymbol{Y} \text{ follows non-singular multinomial} \\ \Leftrightarrow & (\boldsymbol{Y}, n - |\boldsymbol{Y}|) \text{ follows multinomial} \end{array}$$

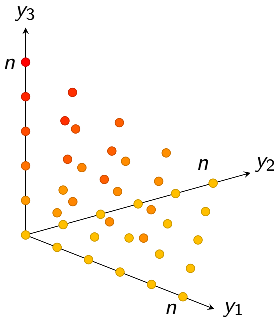
Supported on the interior of Δ_n

$\blacktriangle_n = \left\{ \boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| \leq n \right\}$.

It will be denoted by $\mathcal{M}_{\blacktriangle_n}(\boldsymbol{\pi})$.

Remark

$\dim(\blacktriangle_n) = J$



Non-singular version

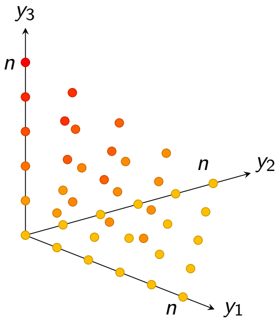
Definition

$$\begin{aligned} & \boldsymbol{Y} \text{ follows non-singular multinomial} \\ \Leftrightarrow \quad & (\boldsymbol{Y}, n - |\boldsymbol{Y}|) \text{ follows multinomial} \end{aligned}$$

Supported on the interior of Δ_n

$\blacktriangle_n = \left\{\boldsymbol{y} \in \mathbb{N}^J : |\boldsymbol{y}| \leq n\right\}$.

It will be denoted by $\mathcal{M}_{\blacktriangle_n}(\boldsymbol{\pi})$.



Remark

$\dim(\blacktriangle_n) = J$

Problem

The support $\blacktriangle_n$ is bounded

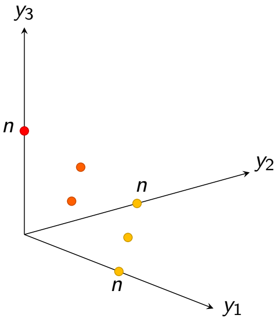
Multinomial Splitting distribution

Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.



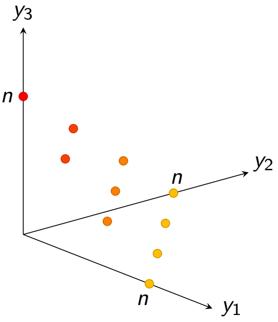
Multinomial Splitting distribution

Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.



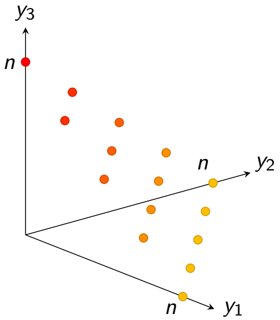
Multinomial Splitting distribution

Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.



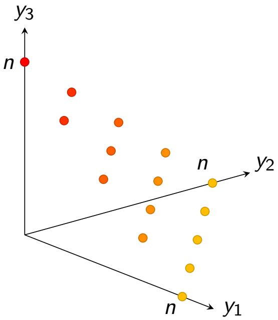
Multinomial Splitting distribution

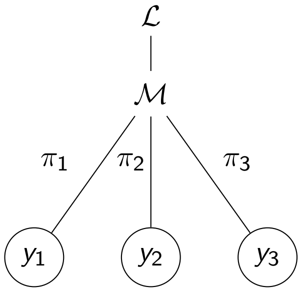
Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.





Multinomial splitting
$\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$

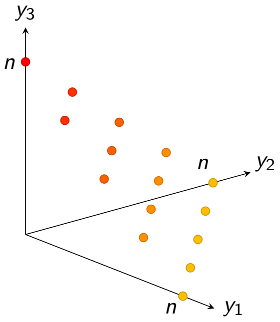
Multinomial Splitting distribution

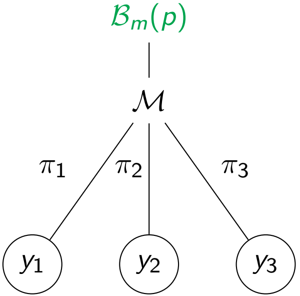
Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.





Non-singular multinomial

$\mathcal{M}_{\blacktriangle_m}(p \cdot \boldsymbol{\pi})$

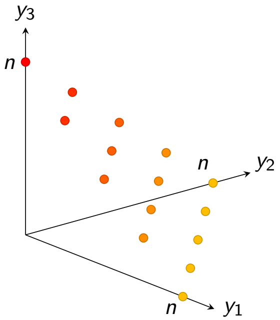
Multinomial Splitting distribution

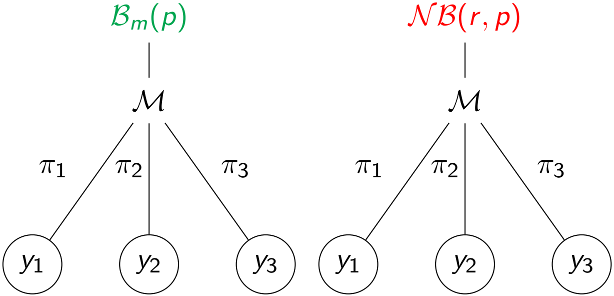
Definition

$\boldsymbol{Y}$ is said to follow a multinomial splitting distribution if

$$\begin{cases} |\boldsymbol{Y}| \sim \mathcal{L} & (sum) \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \sim \mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) & (split) \end{cases}$$

It will be denoted by $\mathcal{M}_{\Delta_n}(\boldsymbol{\pi}) \underset{n}{\wedge} \mathcal{L}$.





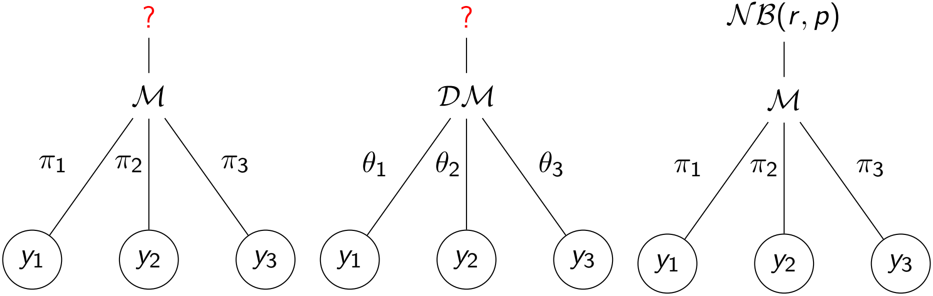
Non-singular multinomial
$\mathcal{M}_{\blacktriangle_m}(p \cdot \boldsymbol{\pi})$

Negative multinomial
$\mathcal{NM}(r, p \cdot \boldsymbol{\pi})$

Problem of models comparison . . .

Zhang et al. (2017)

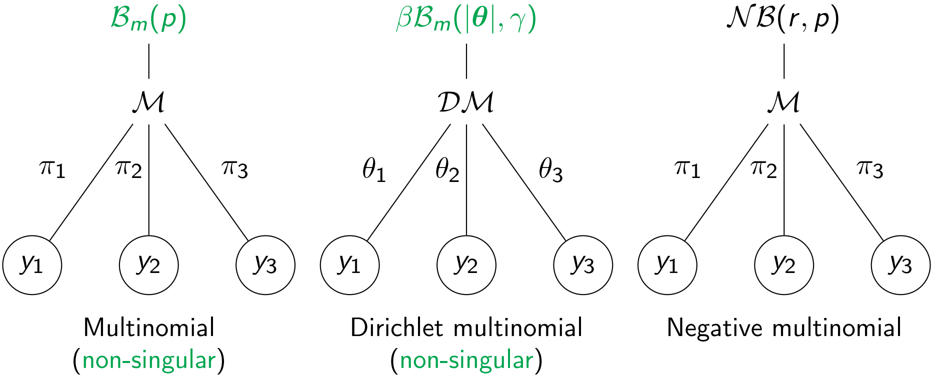
They compared the **multinomial**, the **Dirichlet multinomial** and the **negative multinomial** regression models on RNA-seq dataset.



Multinomial (singular) | Dirichlet multinomial (singular) | Negative multinomial

Zhang et al. (2017)

They compared the **multinomial**, the **Dirichlet multinomial** and the **negative multinomial** regression models on RNA-seq dataset.

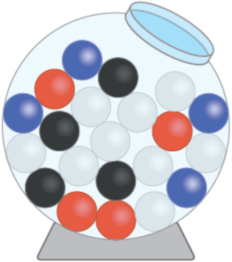


Outline

Pólya urn models

- The urn initially contains θ_j balls of the j^{th} color
- At each draw, one ball is drawn at random and then replaced with c "additional" balls of the same color ($c \in \{-1, 0, 1\}$).
- Focus on the count $\boldsymbol{y} = (y_1, \dots, y_J)$ of drawn balls after n draws

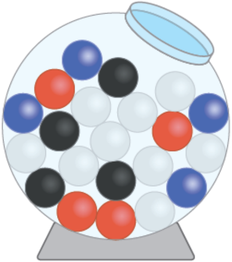
$$P_{|\boldsymbol{Y}|=n}(\boldsymbol{Y} = \boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \frac{\prod_{j=1}^{J} \theta_j^{(y_j;c)}}{|\boldsymbol{\theta}|^{(n;c)}}, \quad \text{where } \theta^{(n;c)} := \prod_{k=0}^{n-1} (\theta + ck)$$



Pólya urn models

- The urn initially contains θ_j balls of the j^{th} color
- At each draw, one ball is drawn at random and then replaced with c "additional" balls of the same color ($c \in \{-1, 0, 1\}$).
- Focus on the count $\boldsymbol{y} = (y_1, \dots, y_J)$ of drawn balls after n draws

$$P_{|\boldsymbol{Y}|=n}(\boldsymbol{Y} = \boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \frac{\prod_{j=1}^{J} \theta_j^{(y_j;c)}}{|\boldsymbol{\theta}|^{(n;c)}}, \quad \text{where } \theta^{(n;c)} := \prod_{k=0}^{n-1} (\theta + ck)$$

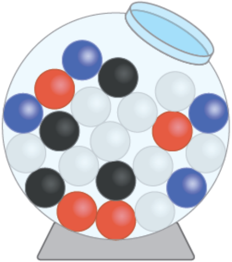


1. $c = -1 \quad \leftrightarrow \quad$ without replacement
   multivariate **hypergeometric**
2. $c = 0 \quad \leftrightarrow \quad$ with replacement (indep. draws)
   multinomial
3. $c = 1 \quad \leftrightarrow \quad$ with reinforcement
   Dirichlet multinomial

Pólya urn models

- The urn initially contains θ_j balls of the j^{th} color
- At each draw, one ball is drawn at random and then replaced with c "additional" balls of the same color ($c \in \{-1, 0, 1\}$).
- Focus on the count $\boldsymbol{y} = (y_1, \dots, y_J)$ of drawn balls after n draws

$$P_{|\boldsymbol{Y}|=n}(\boldsymbol{Y} = \boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \frac{\prod_{j=1}^{J} \theta_j^{(y_j;c)}}{|\boldsymbol{\theta}|^{(n;c)}}, \quad \text{where } \theta^{(n;c)} := \prod_{k=0}^{n-1} (\theta + ck)$$



1. $c = -1 \quad \leftrightarrow$ without replacement
   multivariate **hypergeometric**
2. $c = 0 \quad \leftrightarrow$ with replacement (indep. draws)
   multinomial
3. $c = 1 \quad \leftrightarrow$ with reinforcement
   Dirichlet multinomial

Notation: $\boldsymbol{Y} \sim \mathcal{P}_{\Delta_n}^{[c]}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$

General properties of Polya splitting

$$\boldsymbol{Y} \sim \mathcal{P}^{[c]}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$$

1. $Y_1, \ldots, Y_J$ are lineary free (random sum)
2. Log-likelihood decomposition (sum and split)

$$\mathcal{L} = \log p(|\boldsymbol{y}|) + \log p_{|\boldsymbol{y}|}(\boldsymbol{y})$$

3. Characteristics: pmf, expectation, co-variance

$$\mathrm{Cov}(Y_i, Y_j) = \frac{\theta_i \theta_j}{|\boldsymbol{\theta}|^2 (|\boldsymbol{\theta}| + c)} \left[\left(\mu_{(2)} - \mu_{(1)}^2\right) |\boldsymbol{\theta}| - c\mu_{(1)}^2 \right].$$

4. Marginal distribution (thinning operation)

$$Y_j \sim \mathcal{P}^{[c]}_{n}(\theta_j, |\boldsymbol{\theta}_{-j}|) \underset{n}{\wedge} \mathcal{L}$$

Pólya thinning operation

Let Y be an univariate count variable such that

$$Y \sim \mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{L}$$

Damage process

Pólya thinning operation

Let Y be an univariate count variable such that

$$Y \sim \mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{L}$$

Binomial thinning operation (Sprott, 1965)

Let n be the number of eggs and Y the number of **hatched** eggs

$$Y \sim \mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}$$

It is equivalent to

$$Y = \sum_{i=1}^{n} Z_i, \quad n \sim \mathcal{L}, \quad Z_i \sim \mathcal{B}(\pi) \text{ i.i.d.}$$

Damage process

Pólya thinning operation

Let Y be an univariate count variable such that

$$Y \sim \mathcal{P}_n^{[c]}(\theta, \gamma) \underset{n}{\wedge} \mathcal{L}$$

Binomial thinning operation (Sprott, 1965)

Let n be the number of eggs and Y the number of **hatched** eggs

$$Y \sim \mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}$$

It is equivalent to

$$Y = \sum_{i=1}^{n} Z_i, \quad n \sim \mathcal{L}, \quad Z_i \sim \mathcal{B}(\pi) \text{ i.i.d.}$$

→ **How to choose the sum distribution $\mathcal{L}$?**

Outline

Closure under thinning operation

Definition

A parametric distribution $\mathcal{L}(\psi)$ is said to be closed under the binomial thinning operation if there exists ψ' such that

$$\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(\psi) = \mathcal{L}(\psi')$$

Closure under thinning operation

Definition

A parametric distribution $\mathcal{L}(\psi)$ is said to be closed under the binomial thinning operation if there exists ψ' such that

$$\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{L}(\psi) = \mathcal{L}(\psi')$$

Examples (Rao, 1965)

(1) $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{B}_m(p) = \mathcal{B}_m(p')$ where $p' = \pi p$

(2) $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{P}(\lambda) = \mathcal{P}(\lambda')$ where $\lambda' = \pi\lambda$

(3) $\mathcal{B}_n(\pi) \underset{n}{\wedge} \mathcal{NB}(r, b) = \mathcal{NB}(r, b')$ where $b' = \pi b$

Pmf of univariate Polya

Let $\theta^{(n;c)} = \prod_{k=0}^{n-1}(\theta + kc)$ denote the ascendent factorial. For $k = 0, \ldots, n$

$$P(X = k) = \binom{n}{k} \frac{\theta^{(k;c)}\gamma^{(n-k;c)}}{(\theta + \gamma)^{(n;c)}}$$

Pmf of univariate Polya

Let $\theta^{(n;c)} = \prod_{k=0}^{n-1}(\theta + kc)$ denote the ascendent factorial. For $k = 0, \ldots, n$

$$P(X = k) = \binom{n}{k} \frac{\theta^{(k;c)}\gamma^{(n-k;c)}}{(\theta+\gamma)^{(n;c)}}$$

$$= \frac{\frac{\theta^{(k;c)}}{k!} \; \frac{\gamma^{(n-k;c)}}{(n-k)!}}{\frac{(\theta+\gamma)^{(n;c)}}{n!}}$$

Pmf of univariate Polya

Let $\theta^{(n;c)} = \prod_{k=0}^{n-1}(\theta + kc)$ denote the ascendent factorial. For $k = 0, \ldots, n$

$$P(X = k) = \binom{n}{k} \frac{\theta^{(k;c)}\gamma^{(n-k;c)}}{(\theta+\gamma)^{(n;c)}}$$

$$= \frac{\frac{\theta^{(k;c)}}{k!} \; \frac{\gamma^{(n-k;c)}}{(n-k)!}}{\frac{(\theta+\gamma)^{(n;c)}}{n!}}$$

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

Pmf of univariate Polya

Let $\theta^{(n;c)} = \prod_{k=0}^{n-1}(\theta + kc)$ denote the ascendent factorial. For $k = 0, \dots, n$

$$P(X = k) = \binom{n}{k} \frac{\theta^{(k;c)}\gamma^{(n-k;c)}}{(\theta+\gamma)^{(n;c)}}$$

$$= \frac{\frac{\theta^{(k;c)}}{k!} \; \frac{\gamma^{(n-k;c)}}{(n-k)!}}{\frac{(\theta+\gamma)^{(n;c)}}{n!}}$$

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

Convolution identity

$$a_\theta^{[c]} * a_\gamma^{[c]} = a_{\theta+\gamma}^{[c]}$$

$\rightsquigarrow$ Vandermonde ($c = -1$), Newton ($c = 0$), Hagen-Rothe ($c = 1$)

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\, a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \ldots, n$$

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \dots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X = k) \ \propto \ a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X=k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \dots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X=k) \ \propto \ a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(3) Inverse Polya distribution: $X \sim \mathcal{IP}_n^{[c]}(r; \theta, \gamma)$

$$P(X=k) = \frac{r}{n+r} \frac{a_\theta^{[c]}(k) a_\gamma^{[c]}(r)}{a_{\theta+\gamma}^{[c]}(k+r)}, \quad k \in \mathbb{N}$$

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \dots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X = k) \ \propto \ a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(3) Inverse Polya distribution: $X \sim \mathcal{IP}_n^{[c]}(r; \theta, \gamma)$

$$P(X = k) = \frac{r}{n+r} \frac{a_\theta^{[c]}(k) a_\gamma^{[c]}(r)}{a_{\theta+\gamma}^{[c]}(k+r)}, \quad k \in \mathbb{N}$$

	$c = -1$	$c = 0$	$c = 1$
$a_\theta^{[c]}(n)$	$\binom{\theta}{n}$	$\theta^n / n!$	$\binom{n+\theta-1}{n}$

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \dots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X = k) \propto a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(3) Inverse Polya distribution: $X \sim \mathcal{IP}_n^{[c]}(r; \theta, \gamma)$

$$P(X = k) = \frac{r}{n+r} \frac{a_\theta^{[c]}(k) a_\gamma^{[c]}(r)}{a_{\theta+\gamma}^{[c]}(k+r)}, \quad k \in \mathbb{N}$$

	$c = -1$	$c = 0$	$c = 1$
$a_\theta^{[c]}(n)$	$\binom{\theta}{n}$	$\theta^n / n!$	$\binom{n+\theta-1}{n}$
(1)	hypergeometric	binomial	beta binomial

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \dots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X = k) \ \propto \ a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(3) Inverse Polya distribution: $X \sim \mathcal{IP}_n^{[c]}(r; \theta, \gamma)$

$$P(X = k) = \frac{r}{n+r} \frac{a_\theta^{[c]}(k) a_\gamma^{[c]}(r)}{a_{\theta+\gamma}^{[c]}(k+r)}, \quad k \in \mathbb{N}$$

	$c = -1$	$c = 0$	$c = 1$
$a_\theta^{[c]}(n)$	$\binom{\theta}{n}$	$\theta^n/n!$	$\binom{n+\theta-1}{n}$
(1)	hypergeometric	binomial	beta binomial
(2)	binomial	Poisson	negative binomial

(1) Polya distribution: $X \sim \mathcal{P}_n^{[c]}(\theta, \gamma)$

$$P(X = k) = \frac{a_\theta^{[c]}(k)\ a_\gamma^{[c]}(n-k)}{a_{\theta+\gamma}^{[c]}(n)}, \quad k = 0, \ldots, n$$

(2) Power series distribution (additive): $X \sim \mathcal{PS}^{[c]}(\theta, \alpha)$

$$P(X = k) \ \propto \ a_\theta^{[c]}(k)\alpha^k, \quad k \in \mathbb{N}$$

(3) Inverse Polya distribution: $X \sim \mathcal{IP}_n^{[c]}(r; \theta, \gamma)$

$$P(X = k) = \frac{r}{n+r} \frac{a_\theta^{[c]}(k) a_\gamma^{[c]}(r)}{a_{\theta+\gamma}^{[c]}(k+r)}, \quad k \in \mathbb{N}$$

	$c = -1$	$c = 0$	$c = 1$
$a_\theta^{[c]}(n)$	$\binom{\theta}{n}$	$\theta^n / n!$	$\binom{n+\theta-1}{n}$
(1)	hypergeometric	binomial	beta binomial
(2)	binomial	Poisson	negative binomial
(3)	beta binomial	negative binomial	beta negative binomial

Closure under Pólya thinning operation

Theorem 1 (Peyhardi, 2023)

(1) Pólya

$$\mathcal{P}_n^{[c]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{P}_m^{[c]}(\boldsymbol{\theta}+\boldsymbol{\gamma},\lambda) = \mathcal{P}_m^{[c]}(\theta,\gamma+\lambda)$$

(2) Power series

$$\mathcal{P}_n^{[c]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{PS}^{[c]}(\boldsymbol{\theta}+\boldsymbol{\gamma},\alpha) = \mathcal{PS}^{[c]}(\theta,\alpha)$$

(3) Inverse Pólya

$$\mathcal{P}_n^{[c]}(\theta,\gamma) \underset{n}{\wedge} \mathcal{IP}^{[c]}(r;\boldsymbol{\theta}+\boldsymbol{\gamma},\lambda) = \mathcal{IP}^{[c]}(r;\theta,\lambda)$$

	$c=-1$	$c=0$	$c=1$
$a_\theta^{[c]}(n)$	$\binom{\theta}{n}$	$\theta^n/n!$	$\binom{n+\theta-1}{n}$
(1)	hypergeometric	binomial	beta binomial
(2)	binomial	Poisson	negative binomial
(3)	beta binomial	negative binomial	beta negative binomial

Nine **multivariate** distributions

Split / Sum	Hypergeometric $c = -1$	Multinomial $c = 0$	Dirichlet multinomial $c = 1$
(1) Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $= \blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m$)
(2) Power series	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, p)$ (support $= \mathbb{N}^J$)
(3) Inverse Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a, b)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a, p)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, a, b)$ (support $= \mathbb{N}^J$)

Nine **multivariate** distributions

Sum \ Split	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$
(1) Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $= \blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m$)
(2) Power series	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, p)$ (support $= \mathbb{N}^J$)
(3) Inverse Pólya	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a, b)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a, p)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, a, b)$ (support $= \mathbb{N}^J$)

Closure under summation

$$\left.\begin{array}{l} Y_i \sim \mathcal{L}(\theta_i) \\ Y_j \sim \mathcal{L}(\theta_j) \end{array}\right\} \quad Y_i + Y_j \sim \mathcal{L}(\theta_i + \theta_j)$$

Nine **multivariate** distributions

Sum \ Split	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$
(1) Pólya Cov <0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $= \blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert, \gamma)$ (supp $= \blacktriangle_m$)
(2) Power series Cov $=0$	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, p)$ (support $= \mathbb{N}^J$)
(3) Inverse Pólya Cov >0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a, b)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a, p)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert, a, b)$ (support $= \mathbb{N}^J$)

Closure under summation

$$\left.\begin{array}{l} Y_i \sim \mathcal{L}(\theta_i) \\ Y_j \sim \mathcal{L}(\theta_j) \end{array}\right\} \quad Y_i + Y_j \sim \mathcal{L}(\theta_i + \theta_j)$$

$$\mathrm{Cov}(Y_i, Y_j) = \frac{\theta_i \theta_j}{\lvert\boldsymbol{\theta}\rvert^2 (\lvert\boldsymbol{\theta}\rvert + c)} \left[\left(\mu_{(2)} - \mu_{(1)}^2\right) \lvert\boldsymbol{\theta}\rvert - c\mu_{(1)}^2 \right].$$

Outline

Sum \ Split	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$
(1) Pólya Cov <0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$ (supp $= \blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $= \blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$ (supp $= \blacktriangle_m$)
(2) Power series Cov $=0$	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,p)$ (support $= \mathbb{N}^J$)
(3) Inverse Pólya Cov >0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a,b)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a,p)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,a,b)$ (support $= \mathbb{N}^J$)

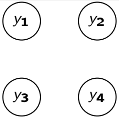
Theorem 2: characterization of PGM (Peyhardi, 2023)

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

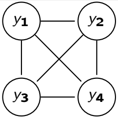
(A) empty **if and only if** $\mathcal{L} = \mathcal{PS}^{[c]}(\lvert\boldsymbol{\theta}\rvert,\alpha)$ (Power series)

(B) complete otherwise

(A)



(B)



Other applications . . .

Thm 2 $\Rightarrow$ Test of independence

Let $\boldsymbol{Y} \sim \mathcal{P}_{\Delta_n}^{[c]}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$

$$H_0 : Y_1, \dots, Y_J \text{ independent} \quad \Leftrightarrow \quad H_0 : \mathcal{L} = \mathcal{PS}^{[c]}(|\boldsymbol{\theta}|, \alpha)$$

Moreover if $\mathcal{L} = \mathcal{PS}^{[c]}(\gamma, \alpha)$ then

$$H_0 : Y_1, \dots, Y_J \text{ independent} \quad \Leftrightarrow \quad H_0 : \gamma = |\boldsymbol{\theta}|$$

Other applications . . .

Thm 2 ⇒ Test of independence

Let $\boldsymbol{Y} \sim \mathcal{P}^{[c]}_{\Delta_n}(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{L}$

$$H_0 : Y_1, \dots, Y_J \text{ independent} \quad \Leftrightarrow \quad H_0 : \mathcal{L} = \mathcal{PS}^{[c]}(|\boldsymbol{\theta}|, \alpha)$$

Moreover if $\mathcal{L} = \mathcal{PS}^{[c]}(\gamma, \alpha)$ then

$$H_0 : Y_1, \dots, Y_J \text{ independent} \quad \Leftrightarrow \quad H_0 : \gamma = |\boldsymbol{\theta}|$$

Thm 1 and Thm 2 ⇒ stationarity of INAR models (count time series)

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

- $\rho \circ X | X = n \sim \mathcal{P}_n(\theta, \gamma)$
- $\varepsilon_t \sim \mathcal{PS}(\gamma, \alpha)$
- $X_0 \sim \mathcal{PS}(\theta + \gamma, \alpha)$

$$\Rightarrow \quad X_t \sim \mathcal{PS}(\theta + \gamma, \alpha)$$

Outline

Biological modeling framework (**Neutral Theory**)

- Biodiversity is not only determined by the number J of different entities, but also by the abundance of these entities $\boldsymbol{Y} = (Y_1, \ldots, Y_J)$.

- The neutral theory of biodiversity (Caswell, 1976) assumes
  - H_0: no biological interaction beetween species
    $\rightarrow$ challenges the traditional niche-based view of ecology
  - the multivariate species abundance distribution is the **stationary distribution** of a simple stochastic process
    H_0: null model

Mathematical modeling framework (**Neutral Theory**)

Let $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_J(t))$ be a **multivariate** birth death process.

Mathematical modeling framework (**Neutral Theory**)

Let $\boldsymbol{Y}(t) = (Y_1(t), \dots, Y_J(t))$ be a **multivariate** birth death process.

Jumping rates

$$\begin{aligned} q_j^+(\boldsymbol{y}) &:= m_j(\boldsymbol{y}) + y_j b_j(\boldsymbol{y}) \\ q_j^-(\boldsymbol{y}) &:= y_j d_j(\boldsymbol{y}) \end{aligned}$$

- $m_j(\boldsymbol{y})$: **immigration** rate of species j
- $b_j(\boldsymbol{y})$: per-individual **birth** rate of species j
- $d_j(\boldsymbol{y})$: per-individual **death** rate of species j

Mathematical modeling framework (**Neutral Theory**)

Let $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_J(t))$ be a **multivariate** birth death process.

Jumping rates

$$\begin{aligned} q_j^+(\boldsymbol{y}) &:= m_j(\boldsymbol{y}) + y_j b_j(\boldsymbol{y}) \\ q_j^-(\boldsymbol{y}) &:= y_j d_j(\boldsymbol{y}) \end{aligned}$$

- $m_j(\boldsymbol{y})$: **immigration** rate of species j
- $b_j(\boldsymbol{y})$: per-individual **birth** rate of species j
- $d_j(\boldsymbol{y})$: per-individual **death** rate of species j

The master equation

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

Mathematical modeling framework (**Neutral Theory**)

Let $\boldsymbol{Y}(t) = (Y_1(t), \dots, Y_J(t))$ be a **multivariate** birth death process.

Jumping rates

$$\begin{aligned} q_j^+(\boldsymbol{y}) &:= m_j(\boldsymbol{y}) + y_j b_j(\boldsymbol{y}) \\ q_j^-(\boldsymbol{y}) &:= y_j d_j(\boldsymbol{y}) \end{aligned}$$

- $m_j(\boldsymbol{y})$: **immigration** rate of species j
- $b_j(\boldsymbol{y})$: per-individual **birth** rate of species j
- $d_j(\boldsymbol{y})$: per-individual **death** rate of species j

The master equation

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

Study the **stationary distribution** under neutral assumption on

$$q_j(\boldsymbol{y}) := \frac{q_j^+(\boldsymbol{y})}{q_j^-(\boldsymbol{y}+\boldsymbol{e}_j)}$$

Model I of Caswell (1976)

- Model
  - Speciation according to Poisson process
  - "New immigrant becomes the founder of a line of descendants generated by a linear birth-death process" ($b = d$)

Model I of Caswell (1976)

- Model
  - Speciation according to Poisson process
  - "New immigrant becomes the founder of a line of descendants generated by a linear birth-death process" ($b = d$)
- Assumptions
  1. **Neutrality** : without biological interaction
  2. Independence between species

Model I of Caswell (1976)

- Model
  - Speciation according to Poisson process
  - "New immigrant becomes the founder of a line of descendants generated by a linear birth-death process" ($b = d$)

- Assumptions
  1. **Neutrality** : without biological interaction
  2. Independence between species

- Total abundance (at time t)

  $N_t \sim \mathcal{NB}$ with expectation $E[N_t] = t$

  Problem: "the total abundance grows without bound over time"

Model of Hubbell (2001)

- Hubbell (2001) : "The total abundance increases linearly with the area: $N = \rho A$"

Model of Hubbell (2001)

- Hubbell (2001) : "The total abundance increases linearly with the area: $N = \rho A$"
- Assumptions
  1. **Neutrality** : functional equivalence among different species
  2. **Zero sum game** : **the community size is fixed over time**

Model of Hubbell (2001)

- Hubbell (2001) : "The total abundance increases linearly with the area: $N = \rho A$"
- Assumptions
  1. **Neutrality** : functional equivalence among different species
  2. **Zero sum game** : **the community size is fixed over time**
- Jumping rates

$$\begin{cases} q_j^+(\boldsymbol{y}) & = d\left(\frac{n - y_j}{n}\right)\left(\frac{y_j}{n-1}\right) \\ q_j^-(\boldsymbol{y}) & = d\left(\frac{y_j}{n}\right)\left(\frac{n - y_j}{n-1}\right) = q_j^+(\boldsymbol{y}) \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = \frac{n+1}{n-1}\frac{y_j}{y_j+1}$$

Model of Hubbell (2001)

- Hubbell (2001) : "The total abundance increases linearly with the area: $N = \rho A$"
- Assumptions
  1. **Neutrality** : functional equivalence among different species
  2. **Zero sum game** : **the community size is fixed over time**
- Jumping rates

$$\begin{cases} q_j^+(\mathbf{y}) & = d\left(\frac{n-y_j}{n}\right)\left(\frac{y_j}{n-1}\right) \\ q_j^-(\mathbf{y}) & = d\left(\frac{y_j}{n}\right)\left(\frac{n-y_j}{n-1}\right) = q_j^+(\mathbf{y}) \end{cases}$$

$$\Rightarrow q_j(\mathbf{y}) = \frac{n+1}{n-1}\frac{y_j}{y_j+1}$$

- Abundances **at equilibrium** $(t \to \infty)$

  $(y_1, \ldots, y_J) \sim \mathcal{DM}_{\Delta_n}$ (split)

Model of Haegeman and Etienne (2008)

- Assumptions
  1. **Neutrality** : functional equivalence among different species
     ~~**Zero sum game**~~ : total abundance is random !

- Jumping rates

$$\begin{cases} q_j^+(\boldsymbol{y}) & = m(|\boldsymbol{y}|)\pi_j + b(|\boldsymbol{y}|)y_j \\ q_j^-(\boldsymbol{y}) & = d(|\boldsymbol{y}|)y_j \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = \frac{b(|\boldsymbol{y}|)}{d(|\boldsymbol{y}|+1)} \frac{\theta_j + y_j}{y_j + 1}$$

where $\theta_j := I\pi_j$ and $I = \frac{m(|\boldsymbol{y}|)}{b(|\boldsymbol{y}|)}$

- Abundances **at equilibrium** $(t \to \infty)$
  - $(y_1, \ldots, y_J)$ given $|\boldsymbol{y}| = n \sim \mathcal{DM}_{\Delta_n}(\boldsymbol{\theta})$ (split)

Model of Haegeman and Etienne (2008)

- Assumptions
  1. **Neutrality** : functional equivalence among different species
     ~~**Zero sum game**~~ : total abundance is random !
  2. **Independence** between species abundances
- Jumping rates

$$\begin{cases} q_j^+(\boldsymbol{y}) & = m(|\boldsymbol{y}|)\pi_j + b(|\boldsymbol{y}|)y_j \\ q_j^-(\boldsymbol{y}) & = d(|\boldsymbol{y}|)y_j \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = \frac{b(|\boldsymbol{y}|)}{d(|\boldsymbol{y}|+1)} \frac{\theta_j + y_j}{y_j + 1}$$

where $\theta_j := I\pi_j$ and $I = \frac{m(|\boldsymbol{y}|)}{b(|\boldsymbol{y}|)}$
and $\frac{b(|\boldsymbol{y}|)}{d(|\boldsymbol{y}|+1)} = \alpha$ (constant)

- Abundances **at equilibrium** $(t \to \infty)$
  - $(y_1, \dots, y_J)$ given $|\boldsymbol{y}| = n \quad \sim \mathcal{DM}_{\Delta_n}(\boldsymbol{\theta})$ (split)
  - $|\boldsymbol{y}| \sim \mathcal{NB}(I, \alpha)$ (sum)

Outline

Model of Peyhardi et al. (2024)

- Assumptions
  1. Neutrality species are equivalent in immigration, birth and death rates
     ~~Zero sum game~~

- Jumping rates

$$\begin{cases} q_j^+(\boldsymbol{y}) & = m(|\boldsymbol{y}|)\pi_j + b(|\boldsymbol{y}|)y_j \\ q_j^-(\boldsymbol{y}) & = d(|\boldsymbol{y}|)y_j \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = s(|\boldsymbol{y}|)\frac{\theta_j + y_j}{y_j + 1}$$

where $\theta_j := I\pi_j$ and $I = \frac{m(|\boldsymbol{y}|)}{b(|\boldsymbol{y}|)}$

- Abundances **at equilibrium** $(t \to \infty)$
  - $(y_1, \ldots, y_J)$ given $|\boldsymbol{y}| = n \sim \mathcal{DM}_{\Delta_n}(\boldsymbol{\theta})$ (split)
  - $|\boldsymbol{y}| \sim \mathcal{L}$ (sum)

Model of Peyhardi et al. (2024)

- Assumptions
  1. Neutrality species are equivalent in immigration, birth and death rates
     ~~Zero sum game~~
     ~~Independence~~
- Jumping rates

$$\begin{cases} q_j^+(\boldsymbol{y}) & = m(|\boldsymbol{y}|)\pi_j + b(|\boldsymbol{y}|)y_j \\ q_j^-(\boldsymbol{y}) & = d(|\boldsymbol{y}|)y_j \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = s(|\boldsymbol{y}|)\frac{\theta_j + y_j}{y_j + 1}$$

where $\theta_j := I\pi_j$ and $I = \frac{m(|\boldsymbol{y}|)}{b(|\boldsymbol{y}|)}$

$s(|\boldsymbol{y}|) = \frac{1}{|\boldsymbol{\theta}|}\frac{m(|\boldsymbol{y}|)}{d(|\boldsymbol{y}|+1)}$ not necessary constant !

- Abundances **at equilibrium** $(t \to \infty)$
  - $(y_1, \ldots, y_J)$ given $|\boldsymbol{y}| = n \sim \mathcal{DM}_{\Delta_n}(\boldsymbol{\theta})$ (split)
  - $|\boldsymbol{y}| \sim \mathcal{L}$ (sum) $\leftrightarrow$ **any sum distribution**

Model of Peyhardi et al. (2024)

- Assumptions
  1. Neutrality species are equivalent in immigration, birth and death rates
     ~~Zero sum game~~
     ~~Independence~~
- Jumping rates (with possible density-dependent immigration)

$$\begin{cases} q_j^+(\boldsymbol{y}) & = m(|\boldsymbol{y}|)\,[\pi_j - Ky_j] + b(|\boldsymbol{y}|)y_j \\ q_j^-(\boldsymbol{y}) & = d(|\boldsymbol{y}|)y_j \end{cases}$$

$$\Rightarrow q_j(\boldsymbol{y}) = s(|\boldsymbol{y}|)\frac{\theta_j + \mathbf{c}y_j}{y_j + 1}$$

where $\theta_j := I\pi_j$ and $I = \frac{m(|\boldsymbol{y}|)}{b(|\boldsymbol{y}|)}$

$s(|\boldsymbol{y}|) = \frac{1}{|\boldsymbol{\theta}|}\frac{m(|\boldsymbol{y}|)}{d(|\boldsymbol{y}|+1)}$ not necessary constant !

- Abundances **at equilibrium** $(t \to \infty)$
  - $(y_1, \ldots, y_J)$ given $|\boldsymbol{y}| = n \sim \mathcal{P}^{[c]}_{\Delta_n}(\boldsymbol{\theta})$ (split) $\leftrightarrow$ **any Polya distribution**
  - $|\boldsymbol{y}| \sim \mathcal{L}$ (sum) $\leftrightarrow$ **any sum distribution**

Jumping rates → Polya splitting distrbutions (stationarity)

Sum \ Split	Hypergeometric $c = -1$	Multinomial $c = 0$	Dirichlet multinomial $c = 1$
(1) Pólya Cov < 0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\|\boldsymbol{\theta}\|, \gamma)$ (supp $= \blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $= \blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\|\boldsymbol{\theta}\|, \gamma)$ (supp $= \blacktriangle_m$)
(2) Power series Cov $= 0$	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\|\boldsymbol{\theta}\|}(p)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\|\boldsymbol{\theta}\|, p)$ (support $= \mathbb{N}^J$)
(3) Inverse Pólya Cov > 0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\|\boldsymbol{\theta}\|}(a, b)$ (support $= \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a, p)$ (support $= \mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\|\boldsymbol{\theta}\|, a, b)$ (support $= \mathbb{N}^J$)

Jumping rates → Polya splitting distrbutions (stationarity)

Sum \ Split	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$
(1) Pólya Cov <0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{H}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$ (supp $=\blacktriangle_m \cap \blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_m(p)$ (supp $=\blacktriangle_m$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_m(\lvert\boldsymbol{\theta}\rvert,\gamma)$ (supp $=\blacktriangle_m$)
(2) Power series Cov $=0$	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(p)$ (support $=\blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{P}(\lambda)$ (support $=\mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,p)$ (support $=\mathbb{N}^J$)
(3) Inverse Pólya Cov >0	$\mathcal{H}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{B}_{\lvert\boldsymbol{\theta}\rvert}(a,b)$ (support $=\blacksquare_{\boldsymbol{\theta}}$)	$\mathcal{M}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \mathcal{NB}(a,p)$ (support $=\mathbb{N}^J$)	$\mathcal{DM}_n(\boldsymbol{\theta}) \underset{n}{\wedge} \beta\mathcal{NB}(\lvert\boldsymbol{\theta}\rvert,a,b)$ (support $=\mathbb{N}^J$)

Sum \ Split	Hypergeometric $c=-1$	Multinomial $c=0$	Dirichlet multinomial $c=1$
(1) Pólya	$\frac{m-\lvert\mathbf{y}\rvert}{\gamma-m+\lvert\mathbf{y}\rvert+1}$	$\frac{m-\lvert\mathbf{y}\rvert}{\gamma}$	$\frac{m-\lvert\mathbf{y}\rvert}{\gamma+m-\lvert\mathbf{y}\rvert-1}$
(2) Power series	α	α	α
(3) Inverse Pólya	$\frac{a+\lvert\mathbf{y}\rvert}{\lvert\boldsymbol{\theta}\rvert+b-\lvert\mathbf{y}\rvert-1}$	$\frac{a+\lvert\mathbf{y}\rvert}{\lvert\boldsymbol{\theta}\rvert+b}$	$\frac{b+\lvert\mathbf{y}\rvert}{\lvert\boldsymbol{\theta}\rvert+a+b+\lvert\mathbf{y}\rvert}$

Table: Parametric form of $s(\lvert\mathbf{y}\rvert)$ and thus of jumping rate $q_j(\mathbf{y}) = s(\lvert\mathbf{y}\rvert)\frac{\theta_j+cy_j}{y_j+1}$.

* The grey cells corresponds to the case of Haegeman and Etienne (2008).

Outline

Polya splitting **regression**

Let $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ denotes the explanatory variables.

$$\boldsymbol{Y} \text{ given } x \sim \mathcal{P}^{[c]}_{\Delta_n}\{\boldsymbol{\theta}(x)\} \underset{n}{\wedge} \mathcal{L}\{\psi(x)\}$$

$$\Leftrightarrow \begin{cases} |\boldsymbol{Y}| \sim \mathcal{L}\{\psi(x)\} & \text{(sum regression)} \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \text{ and } x \sim \mathcal{P}^{[c]}_{\Delta_n}\{\boldsymbol{\theta}(x)\} & \text{(split regression)} \end{cases}$$

Polya splitting **regression**

Let $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ denotes the explanatory variables.

$$\boldsymbol{Y} \text{ given } x \sim \mathcal{P}^{[c]}_{\Delta_n}\{\boldsymbol{\theta}(x)\} \underset{n}{\wedge} \mathcal{L}\{\psi(x)\}$$

$$\Leftrightarrow \begin{cases} |\boldsymbol{Y}| \sim \mathcal{L}\{\psi(x)\} & \text{(sum regression)} \\ \boldsymbol{Y} \text{ given } |\boldsymbol{Y}| = n \text{ and } x \sim \mathcal{P}^{[c]}_{\Delta_n}\{\boldsymbol{\theta}(x)\} & \text{(split regression)} \end{cases}$$

Example: negative multinomial regression

$$\boldsymbol{Y} \text{ given } x \sim \mathcal{M}_{\Delta_n}\{\boldsymbol{\pi}(x)\} \underset{n}{\wedge} \mathcal{NB}\{r, p(x)\}$$

$$\pi_j = \frac{\exp(x^t\beta_j)}{\sum_{k=1}^{J}\exp(x^t\beta_k)}\; j = 1, \dots, J \quad \text{and} \quad p = \frac{\exp(x^t\delta)}{1+\exp(x^t\delta)}$$

Maximum Likelihood Estimation (MLE)

Log-likelihood decomposition

$$\mathcal{L}(\{\beta_j\}, \delta; \boldsymbol{y}) = \mathcal{L}(\{\beta_j\}; \boldsymbol{y}) + \mathcal{L}(\delta; |\boldsymbol{y}|)$$

Maximum Likelihood Estimation (MLE)

Log-likelihood decomposition

$$\mathcal{L}(\{\beta_j\}, \delta; \boldsymbol{y}) = \mathcal{L}(\{\beta_j\}; \boldsymbol{y}) + \mathcal{L}(\delta; |\boldsymbol{y}|)$$

Example: negative multinomial regression

separate estimation on sum and split

- estimate $\{\hat{\beta}_j\}$ on multivariate dataset $(\boldsymbol{y}_i, x_i)_{i=1,\ldots,n}$ (split)
- estimate $\hat{r}$ and δ on univariate dataset $(|\boldsymbol{y}_i|, x_i)_{i=1,\ldots,n}$ (sum)

Maximum Likelihood Estimation (MLE)

Log-likelihood decomposition

$$\mathcal{L}(\{\beta_j\}, \delta; \boldsymbol{y}) = \mathcal{L}(\{\beta_j\}; \boldsymbol{y}) + \mathcal{L}(\delta; |\boldsymbol{y}|)$$

Example: negative multinomial regression

separate estimation on sum and split

- estimate $\{\hat{\beta}_j\}$ on multivariate dataset $(\boldsymbol{y}_i, x_i)_{i=1,\dots,n}$ (split)
- estimate $\hat{r}$ and δ on univariate dataset $(|\boldsymbol{y}_i|, x_i)_{i=1,\dots,n}$ (sum)

GLM properties

- concavity of log-likelihood for canonical link function
- simple formulation of Fisher's scoring algorithm
- asymptotic normality of MLE

Model selection

Time complexity

$$\left.\begin{array}{l} L \text{ models for the sum} \\ M \text{ models for the split} \end{array}\right\} \Rightarrow L \times M \text{ models}$$

with $\mathcal{O}(C + L)$ time complexity.

Model selection

Time complexity

$$\left.\begin{array}{l} L \text{ models for the sum} \\ M \text{ models for the split} \end{array}\right\} \Rightarrow L \times M \text{ models}$$

with $\mathcal{O}(C + L)$ time complexity.

Existing packages

- sum:
  - Poisson
  - Negative binomial
- split:
  - multinomial
  - Dirichlet multinomial

Model selection

Time complexity

$$\left.\begin{array}{l} L \text{ models for the sum} \\ M \text{ models for the split} \end{array}\right\} \Rightarrow L \times M \text{ models}$$

with $\mathcal{O}(C + L)$ time complexity.

Existing packages

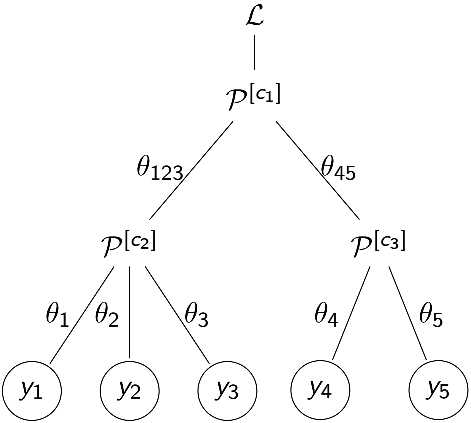
- sum:
  - Poisson
  - Negative binomial
- split:
  - multinomial
  - Dirichlet multinomial

Variables selection

LASSO, Ridge, step AIC, ...

Outline

Tree Polya splitting distribution

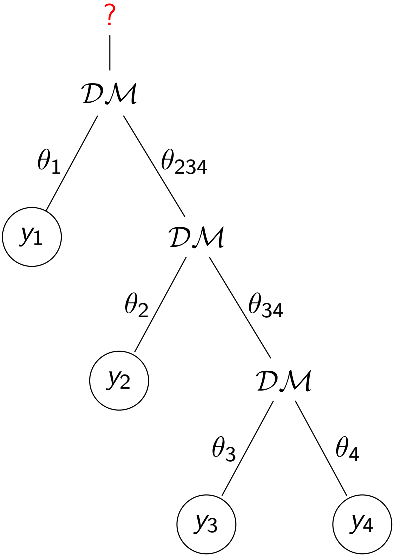


- decomposition of the likelihood

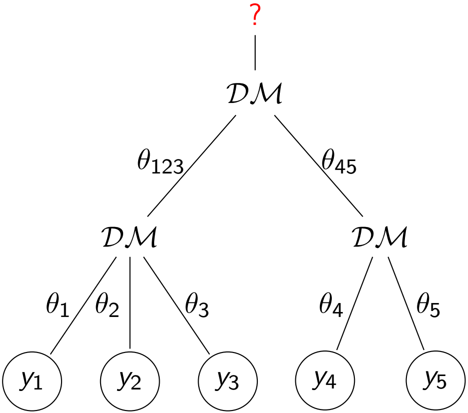
$$\mathcal{L} = \log p(|\mathbf{y}|) + \sum_{A \in \mathcal{I}} \log p_{|\mathbf{y}_A|}(\mathbf{y}_A)$$

  where $\mathcal{I}$ set of internal nodes
- more parameters $\rightarrow$ flexibility
- structured covariance
- includes DM, GDM, TDM as special cases

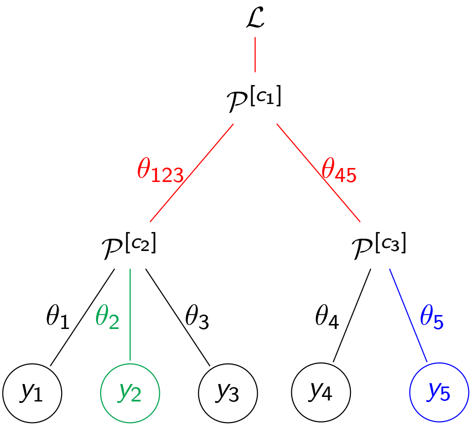
Generalization of GDM, TDM approaches



Generalized Dirichlet Multinomial

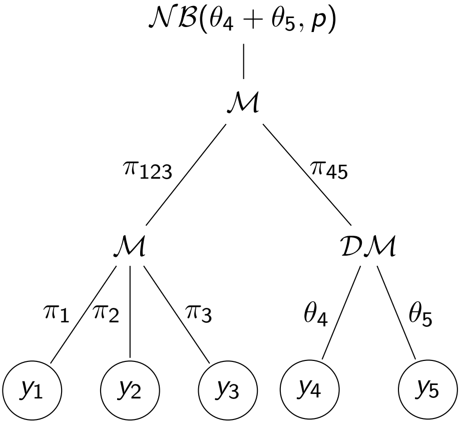


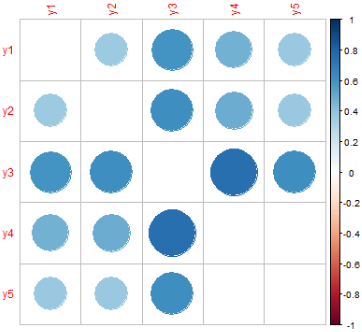
Tree Dirichlet Multinomial



$$\mathrm{Cov}(y_2, y_5) = \frac{\theta_2}{\theta_1 + \theta_2 + \theta_3} \frac{\theta_5}{\theta_4 + \theta_5} \mathrm{Cov}(y_1 + y_2 + y_3, y_4 + y_5)$$

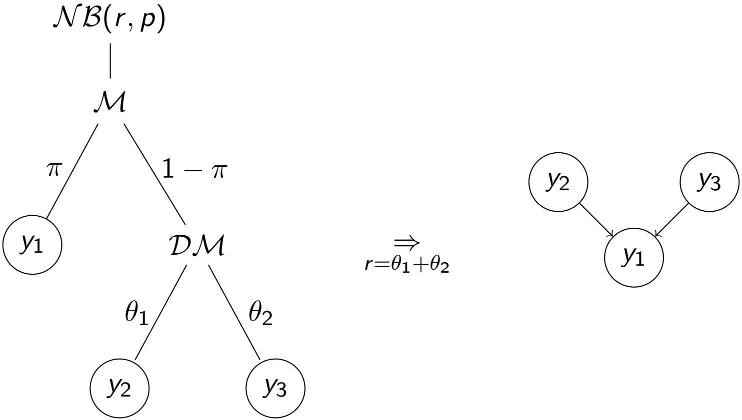
Example of structured covariance





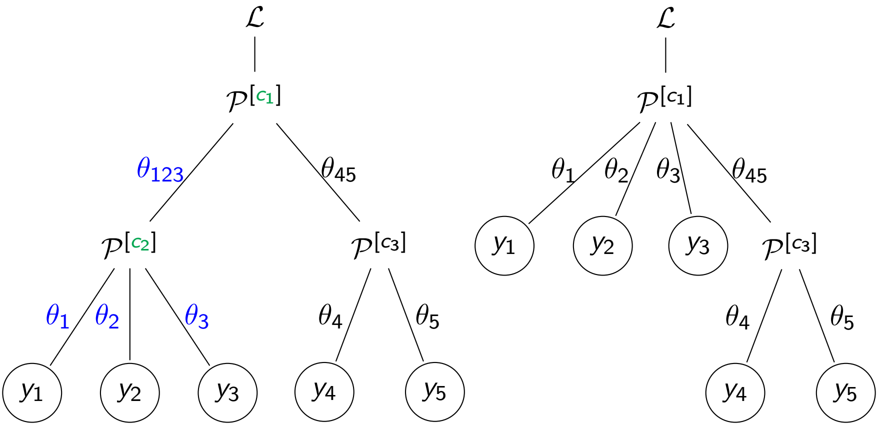
Probabilistic Graphical Model (V-structure)

If $r = \theta_1 + \theta_2$ then $p(y_1, y_2, y_3) = p(y_2)p(y_3)p(y_1|y_2 + y_3)$



- $y_2 \perp\!\!\!\perp y_3$
- $(y_2 + y_3) \nearrow \; \Rightarrow \; y_1 \nearrow$

Equivalent trees



Proposition

These two tree splitting models are equivalent **iff** $c_1 = c_2$ (same Polya) and $\theta_{123} = \theta_1 + \theta_2 + \theta_3$.

Outline

Trichoptera dataset description

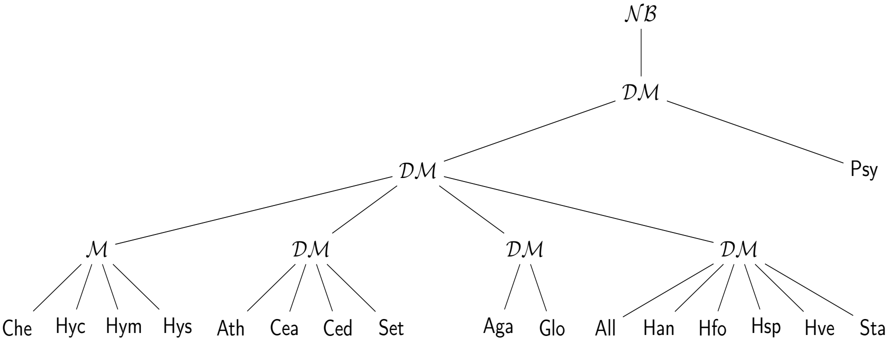
- Group of insects with aquatic larvae and terrestrial adults
- collected between 1959 and 1960 in Lyon
- available on R-package PLNmodels (Chiquet et al., 2021)

- $J = 17$ species abundances
  - Cheumatopsyche lepid (Che)
  - Hyctopsyche contuernglis (Hyc)
  - Hydropsyche mode (Hyd)

    $\vdots$
- $n = 49$ sites
- $p = 9$ covariates (meteorological factors)
  - temperature
  - Wind
  - Pressure

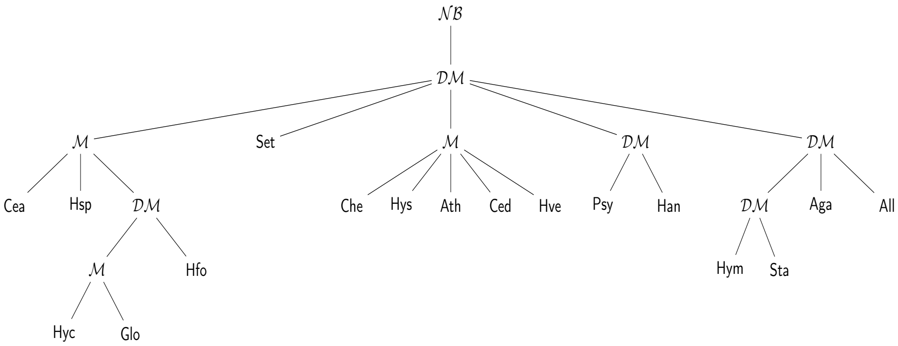
    $\vdots$

Partition tree according to families



Tree Pólya distribution (known tree)

Partition tree search



Tree Pólya distribution (unknown tree).

Models comparison (Trichoptera dataset)

Model	Log-likelihood	Nb. Param.	AIC	BIC
PLN full	-1130.05	170	2599.98	2921.71
PLN diagonal	-1187.97	34	**2443.95**	**2508.26**
PLN spherical	-1236.77	18	2509.55	2543.59

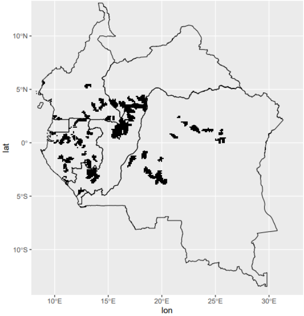
Table: Scores for different Poisson Log Normal (PLN) models

Model	Log-likelihood	Nb. Param.	AIC	BIC
DM	-1228.43	19	2494.87	2530.81
TP (known tree)	-1209.92	23	2465.85	2509.35
TP (unknown tree)	-1167.38	23	**2380.77**	**2424.28**

Table: Scores for different tree Polya (TP) models with NB sum distribution

Trees in Central Afric dataset (Fabrice Moudjeu)

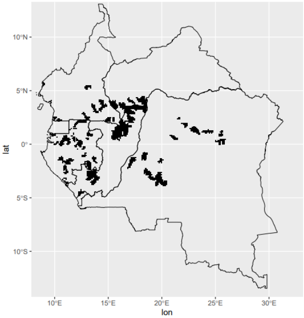
- five countries : Cameroun, Gabon, Rep. of Congo . . .
- $n = 1571$ plots

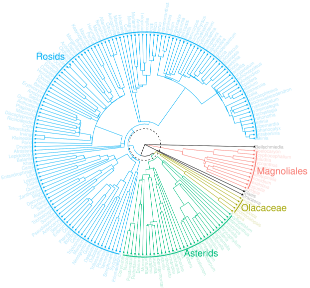


Trees in Central Afric dataset (Fabrice Moudjeu)

- five countries : Cameroun, Gabon, Rep. of Congo . . .
- $n = 1571$ plots
- $J = 180$ taxa

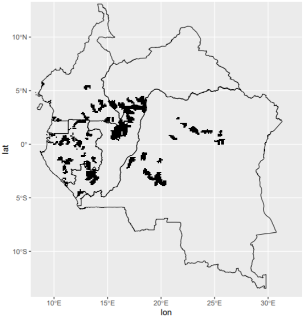
  with four families
  - Rosids
  - Asterids
  - Olacaceae
  - Magnoliales

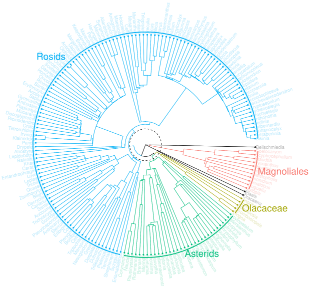




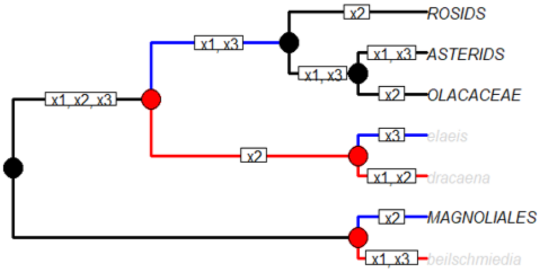
Trees in Central Afric dataset (Fabrice Moudjeu)

- five countries : Cameroun, Gabon, Rep. of Congo . . .
- $n = 1571$ plots
- $J = 180$ taxa
  with four families
  - Rosids
  - Asterids
  - Olacaceae
  - Magnoliales
- $p = 3$ covariates
  - x_1: evapotranspiration
  - x_2: seasonality
  - x_3: picks of temperature





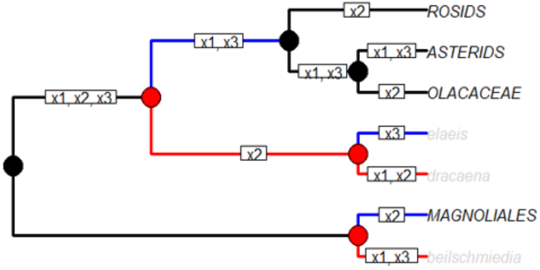
Phylogeny and Zero inflation



- Phylogeny= binary tree

$$(Y_1, Y_2) \mid Y_1 + Y_2 = n \quad \sim \quad \mathcal{DM}_{\Delta_n}(\theta_1, \theta_2)$$

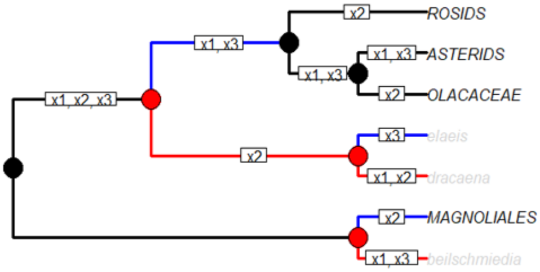
Phylogeny and Zero inflation



- Phylogeny= binary tree
- Dirchlet multinomial = beta binomial

$$
\begin{aligned}
(Y_1, Y_2) \mid Y_1 + Y_2 = n &\sim \mathcal{DM}_{\Delta_n}(\theta_1, \theta_2) \\
\Leftrightarrow \quad Y_1 \mid Y_1 + Y_2 = n &\sim \beta\mathcal{B}_n(\theta_1, \theta_2)
\end{aligned}
$$

Phylogeny and Zero inflation



- Phylogeny= binary tree
- Dirchlet multinomial $=$ beta binomial
- Zero inflation

$$
\begin{array}{rrcl}
 & (Y_1, Y_2) \mid Y_1 + Y_2 = n & \sim & \mathcal{DM}_{\Delta_n}(\theta_1, \theta_2) \\
\Leftrightarrow & Y_1 \mid Y_1 + Y_2 = n & \sim & \beta\mathcal{B}_n(\theta_1, \theta_2)
\end{array}
$$

zero inflation

$$
Y_1 \mid Y_1 + Y_2 = n \sim \pi\delta_0 + (1-\pi)\beta\mathcal{B}_n(\theta_1, \theta_2)
$$

Conlusions and perspectives

Conclusions

- Multivariate extensions of usual count distributions (hypergeometric, binomial, beta binomial, negative binomial . . .)

  $\rightsquigarrow$ similar probabilistic properties

Conlusions and perspectives

Conclusions

- Multivariate extensions of usual count distributions (hypergeometric, binomial, beta binomial, negative binomial . . .)

  $\rightsquigarrow$ similar probabilistic properties

- Link with neutral theory of biodiversity

  $\rightsquigarrow$ indepence between species abundances is a consequence of birth death rates assumptions

Conlusions and perspectives

Conclusions

- Multivariate extensions of usual count distributions (hypergeometric, binomial, beta binomial, negative binomial . . .)

  $\rightsquigarrow$ similar probabilistic properties

- Link with neutral theory of biodiversity

  $\rightsquigarrow$ indepence between species abundances is a consequence of birth death rates assumptions

- Statistical point of view
  - decomposable score (log-likelihood, AIC, BIC)
  - regression + variables selection (existing packages)
  - Tree Polya

Conlusions and perspectives

Conclusions

- Multivariate extensions of usual count distributions (hypergeometric, binomial, beta binomial, negative binomial ...)

  $\rightsquigarrow$ similar probabilistic properties

- Link with neutral theory of biodiversity

  $\rightsquigarrow$ indepence between species abundances is a consequence of birth death rates assumptions

- Statistical point of view
  - decomposable score (log-likelihood, AIC, BIC)
  - regression + variables selection (existing packages)
  - Tree Polya

Perspectives

- Tree Polya $\longleftrightarrow$ neutral theory ?
- How measure deviance from neutral theory ?
- Search algorithm $\rightsquigarrow$ the best Tree

References I

Caswell, H. (1976). Community structure: a neutral model analysis. Ecological monographs, 46(3):327–354.

Chiquet, J., Mariadassou, M., and Robin, S. (2021). The poisson-lognormal model as a versatile framework for the joint analysis of species abundances. Frontiers in Ecology and Evolution, 9:588292.

Haegeman, B. and Etienne, R. (2008). Relaxing the zero-sum assumption in neutral biodiversity theory. Journal of Theoretical Biology, 252(2):288–294.

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

Peyhardi, J. (2023). On quasi pólya thinning operator. Brazilian Journal of Probability and Statistics.

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

References II

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

Sprott, D. (1965). A class of contagious distributions and maximum likelihood estimation. Sankhyā: The Indian Journal of Statistics, Series A, pages 369–382.

Zhang, Y., Zhou, H., Zhou, J., and Sun, W. (2017). Regression models for multivariate count data. Journal of Computational and Graphical Statistics, 26(1):1–13.

Models for count time series

AR(1) model

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

$(X_t)_{t\in\mathbb{N}}$ Markovian process with **continuous** state space $\mathcal{X} = \mathbb{R}$

INAR(1) model

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

$(X_t)_{t\in\mathbb{N}}$ Markovian process with **discrete** state space $\mathcal{X} = \mathbb{N}$

The symbol $\circ$ denotes a stochastic operator that shrinks a random count variable into another one

Stationarity of Poisson INAR(1) model

Closure of Poisson distribution

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

$$\left.\begin{array}{l} \bullet\ \rho \circ X | X = n \sim \mathcal{B}_n(\rho) \\ \bullet\ \varepsilon_t \sim \mathcal{P}((1-\rho)\lambda) \\ \bullet\ X_0 \sim \mathcal{P}(\lambda) \end{array}\right\} \Rightarrow \quad X_t \sim \mathcal{P}(\lambda)$$

Stationarity of Poisson INAR(1) model

Closure of Poisson distribution

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

- $\rho \circ X | X = n \sim \mathcal{B}_n(\rho)$
- $\varepsilon_t \sim \mathcal{P}((1-\rho)\lambda)$
- $X_0 \sim \mathcal{P}(\lambda)$

$$\left.\begin{array}{l} \rho \circ X | X = n \sim \mathcal{B}_n(\rho) \\ \varepsilon_t \sim \mathcal{P}((1-\rho)\lambda) \\ X_0 \sim \mathcal{P}(\lambda) \end{array}\right\} \Rightarrow \quad X_t \sim \mathcal{P}(\lambda)$$

Closure under binomial thinning operator

$$\mathcal{B}_n(\rho) \underset{n}{\wedge} \mathcal{P}(\lambda) = \mathcal{P}(\rho\lambda)$$

Stationarity of Poisson INAR(1) model

Closure of Poisson distribution

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

- $\rho \circ X | X = n \sim \mathcal{B}_n(\rho)$
- $\varepsilon_t \sim \mathcal{P}((1-\rho)\lambda)$
- $X_0 \sim \mathcal{P}(\lambda)$

$$\left.\begin{array}{l} \rho \circ X | X = n \sim \mathcal{B}_n(\rho) \\ \varepsilon_t \sim \mathcal{P}((1-\rho)\lambda) \\ X_0 \sim \mathcal{P}(\lambda) \end{array}\right\} \Rightarrow \quad X_t \sim \mathcal{P}(\lambda)$$

Closure under binomial thinning operator

$$\mathcal{B}_n(\rho) \underset{n}{\wedge} \mathcal{P}(\lambda) = \mathcal{P}(\rho\lambda)$$

Closure under convolution

$$\mathcal{P}(\lambda) * \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu)$$

Stationarity of Power Series INAR(1) model

Closure of Power Series distribution

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

- $\rho \circ X | X = n \sim \mathcal{P}_n(\theta, \gamma)$
- $\varepsilon_t \sim \mathcal{PS}(\gamma, \alpha)$
- $X_0 \sim \mathcal{PS}(\theta + \gamma, \alpha)$

$$\left.\begin{array}{l} \rho \circ X | X = n \sim \mathcal{P}_n(\theta, \gamma) \\ \varepsilon_t \sim \mathcal{PS}(\gamma, \alpha) \\ X_0 \sim \mathcal{PS}(\theta + \gamma, \alpha) \end{array}\right\} \Rightarrow \quad X_t \sim \mathcal{PS}(\theta + \gamma, \alpha)$$

Closure of Power Series distribution

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

- $\rho \circ X | X = n \sim \mathcal{P}_n(\theta, \gamma)$
- $\varepsilon_t \sim \mathcal{PS}(\gamma, \alpha)$
- $X_0 \sim \mathcal{PS}(\theta + \gamma, \alpha)$

$$\Rightarrow \quad X_t \sim \mathcal{PS}(\theta + \gamma, \alpha)$$

Closure under Polya thinning operator

$$\mathcal{P}_n(\theta, \gamma) \underset{n}{\wedge} \mathcal{PS}(\theta + \gamma, \alpha) = \mathcal{PS}(\theta, \alpha)$$

Closure under convolution

$$\mathcal{PS}(\theta, \alpha) * \mathcal{PS}(\gamma, \alpha) = \mathcal{PS}(\theta + \gamma, \alpha)$$