

3 - Variational inference for species abundances and network models

S. Robin

INRAE / AgroParisTech / univ. Paris-Saclay
Muséum National d'Histoire Naturelle

Winter School on Mathematical Statistics, Luxembourg, Dec'20

Outline

- 1 – Models with latent variables in ecology (statistical ecology)
- 2 – Variational inference for incomplete data models (statistics)
- 3 – Variational inference for species abundances and network models (statistical ecology)
- 4 – Beyond variational inference (statistics)

Part 3

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

To summarize

Outline

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

To summarize

Poisson log-normal model for species abundances

Data:

- ▶ n sites, p species, d covariates
- ▶ Y_{ij} = abundance of species j in site i
- ▶ x_i = vector of descriptors for site i

Poisson log-normal model for species abundances

Data:

- ▶ n sites, p species, d covariates
- ▶ Y_{ij} = abundance of species j in site i
- ▶ x_i = vector of descriptors for site i

Abundance table Y

<i>Hi.pl</i>	<i>An.lu</i>	<i>Me.ae</i>	...
31	0	108	
4	0	110	
27	0	788	
.	.	.	

Environmental covariates X

Lat.	Long.	Depth	Temp.
71.10	22.43	349	3.95
71.32	23.68	382	3.75
71.60	24.90	294	3.45
.	.	.	.

Poisson log-normal model for species abundances

Data:

- ▶ n sites, p species, d covariates
- ▶ Y_{ij} = abundance of species j in site i
- ▶ x_i = vector of descriptors for site i

Abundance table Y

<i>Hi.pl</i>	<i>An.lu</i>	<i>Me.ae</i>	...
31	0	108	
4	0	110	
27	0	788	
.	.	.	

Environmental covariates X

Lat.	Long.	Depth	Temp.
71.10	22.43	349	3.95
71.32	23.68	382	3.75
71.60	24.90	294	3.45
.	.	.	.

Poisson log-normal model.

- ▶ Latent vectors

$$Z_i \sim \mathcal{N}(0, \Sigma)$$

- ▶ Observed species counts

$$Y_{ij} \sim \mathcal{P}(\exp(x_i^\top \beta_j + Z_{ij}))$$

- ▶ Parameters

$$\theta = (\beta, \Sigma)$$

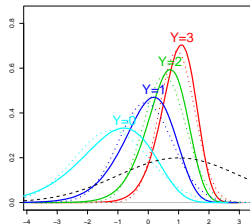
Variational inference

Conditional distribution.

- Because of the independance between sites

$$p_{\theta}(Z | Y) = \prod_i p_{\theta}(Z_i | Y_i)$$

- But $p_{\theta}(Z_i | Y_i)$ has no close form



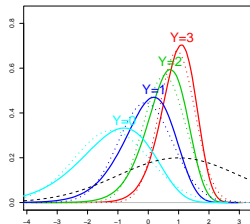
Variational inference

Conditional distribution.

- ▶ Because of the independance between sites

$$p_{\theta}(Z | Y) = \prod_i p_{\theta}(Z_i | Y_i)$$

- ▶ But $p_{\theta}(Z_i | Y_i)$ has no close form



Variational approximation. Use a Gaussian approximate distribution

$$\mathcal{Q} = \{q : q(Z) = \underbrace{\prod_i q_i(Z_i)}_{\text{no approx.}}, \quad q_i(Z_i) = \mathcal{N}(Z_i; m_i, S_i)\}$$

- ▶ Variational parameters: $m_i \simeq \mathbb{E}(Z_i | Y_i), \quad S_i \simeq \mathbb{V}(Z_i | Y_i)$

Variational EM

Variational EM algorithm. `PLNmodels` R package [CMR18]

Variational EM

Variational EM algorithm. `PLNmodels` R package [CMR18]

- VE step: update the variational parameters m_i, S_i

$$(m_i^{h+1}, S_i^{h+1}) = \arg \min_{m, S} KL[\mathcal{N}(Z_i; m, S) \| p_{\theta^h}(Z_i | Y_i)]$$

Variational EM

Variational EM algorithm. `PLNmodels` R package [CMR18]

- ▶ VE step: update the variational parameters m_i, S_i

$$(m_i^{h+1}, S_i^{h+1}) = \arg \min_{m, S} KL[\mathcal{N}(Z_i; m, S) \| p_{\theta^h}(Z_i | Y_i)]$$

→ Convex problem: doable via gradient descent

Variational EM

Variational EM algorithm. `PLNmodels` R package [CMR18]

- ▶ **VE step:** update the **variational parameters** m_i, S_i

$$(m_i^{h+1}, S_i^{h+1}) = \arg \min_{m, S} KL[\mathcal{N}(Z_i; m, S) \| p_{\theta^h}(Z_i | Y_i)]$$

→ Convex problem: doable via gradient descent

- ▶ **M step:** update the **model parameters** Σ, β

$$\theta^{h+1} = \arg \max_{\theta} \mathbb{E}_{q^{h+1}} \log p_{\theta}(Y, Z)$$

Variational EM

Variational EM algorithm. `PLNmodels` R package [CMR18]

- ▶ **VE step:** update the **variational parameters** m_i, S_i

$$(m_i^{h+1}, S_i^{h+1}) = \arg \min_{m, S} KL[\mathcal{N}(Z_i; m, S) \| p_{\theta^h}(Z_i | Y_i)]$$

→ Convex problem: doable via gradient descent

- ▶ **M step:** update the **model parameters** Σ, β

$$\theta^{h+1} = \arg \max_{\theta} \mathbb{E}_{q^{h+1}} \log p_{\theta}(Y, Z)$$

→ Σ^{h+1} : explicit formula

→ β^{h+1} : similar to Poisson regression (generalized linear model)

A first illustration: Abiotic vs biotic interactions

Barents fishes: Full model

$$Y_{ij} \sim \mathcal{P}(\exp(x_i^T \beta_j + Z_{ij}))$$

x_i = all covariates

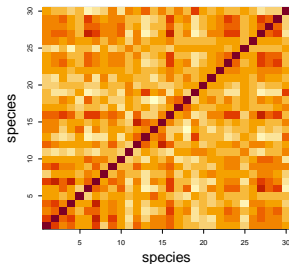
A first illustration: Abiotic vs biotic interactions

Barents fishes: Full model

$$Y_{ij} \sim \mathcal{P}(\exp(x_i^T \beta_j + Z_{ij}))$$

x_i = all covariates

inferred correlations $\hat{\Sigma}_{\text{full}}$



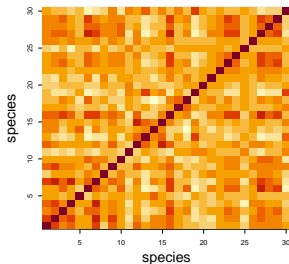
A first illustration: Abiotic vs biotic interactions

Barents fishes: Full model

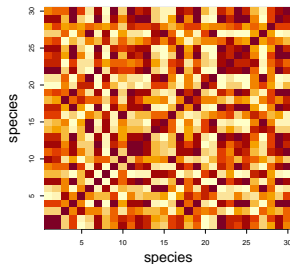
$$Y_{ij} \sim \mathcal{P}(\exp(x_i^T \beta_j + Z_{ij}))$$

x_i = all covariates

inferred correlations $\widehat{\Sigma}_{\text{full}}$



correlations between
predictions: $x_i^T \widehat{\beta}_j$



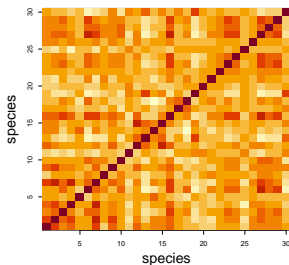
A first illustration: Abiotic vs biotic interactions

Barents fishes: Full model

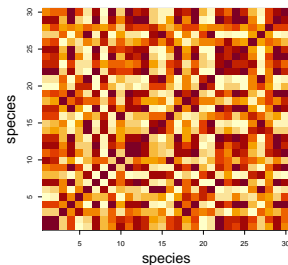
$$Y_{ij} \sim \mathcal{P}(\exp(x_i^T \beta_j + Z_{ij}))$$

x_i = all covariates

inferred correlations $\widehat{\Sigma}_{\text{full}}$



correlations between
predictions: $x_i^T \widehat{\beta}_j$



Null model

$$Y_{ij} \sim \mathcal{P}(\exp(\mu_j + Z_{ij}))$$

no covariate

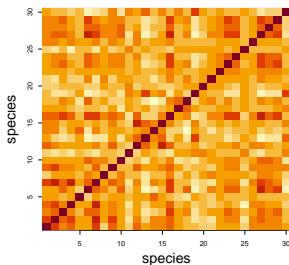
A first illustration: Abiotic vs biotic interactions

Barents fishes: Full model

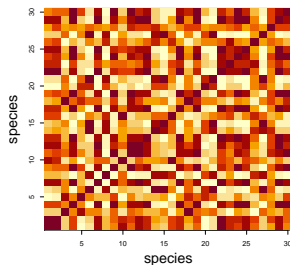
$$Y_{ij} \sim \mathcal{P}(\exp(x_i^T \beta_j + Z_{ij}))$$

x_i = all covariates

inferred correlations $\widehat{\Sigma}_{\text{full}}$



correlations between
predictions: $x_i^T \widehat{\beta}_j$

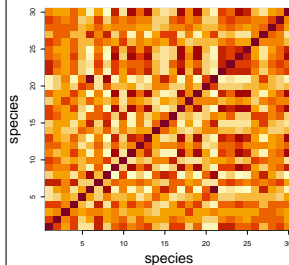


Null model

$$Y_{ij} \sim \mathcal{P}(\exp(\mu_j + Z_{ij}))$$

no covariate

inferred correlations $\widehat{\Sigma}_{\text{null}}$



Outline

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

To summarize

Dimension reduction

Typical context.

- ▶ Microbial ecology: $p = 10^2, 10^3, 10^4$ species
- ▶ 'Abundance' = 'read' count = number of genomic sequences associated with each species sampled via high-throughput sequencing ('metagenomic')

Dimension reduction

Typical context.

- ▶ Microbial ecology: $p = 10^2, 10^3, 10^4$ species
- ▶ 'Abundance' = 'read' count = number of genomic sequences associated with each species sampled via high-throughput sequencing ('metagenomic')

Aim.

- ▶ Dimension reduction (visualization)
- ▶ Accounting for major known effects

Dimension reduction

Typical context.

- ▶ Microbial ecology: $p = 10^2, 10^3, 10^4$ species
- ▶ 'Abundance' = 'read' count = number of genomic sequences associated with each species sampled via high-throughput sequencing ('metagenomic')

Aim.

- ▶ Dimension reduction (visualization)
- ▶ Accounting for major known effects

Probabilistic principal component analysis. Gaussian setting [TB99]:

$$\Sigma = \underbrace{BB^T}_{\text{low rank}} + \sigma^2 I_p, \quad \text{where } B(p \times r)$$

(PLN-)probabilistic PCA

PLN-PCA model. [CMR18]

(PLN-)probabilistic PCA

PLN-PCA model. [CMR18]

- ▶ Low dimension latent vector

$$W_i \sim \mathcal{N}_r(0, I), \quad \text{where } r \ll p$$

(PLN-)probabilistic PCA

PLN-PCA model. [CMR18]

- ▶ Low dimension latent vector

$$W_i \sim \mathcal{N}_r(0, I), \quad \text{where } r \ll p$$

- ▶ p -dimensional latent vector

$$Z_i = BW_i \quad \text{where } B(p \times r) = \text{loading matrix}$$

(PLN-)probabilistic PCA

PLN-PCA model. [CMR18]

- ▶ Low dimension latent vector

$$W_i \sim \mathcal{N}_r(0, I), \quad \text{where } r \ll p$$

- ▶ p -dimensional latent vector

$$Z_i = BW_i \quad \text{where } B(p \times r) = \text{loading matrix}$$

- ▶ Observed counts

$$Y_{ij} \sim \mathcal{P}(\exp(o_{ij} + x_i^T \beta + Z_{ij}))$$

o_{ij} = known 'offset' coefficient, accounting for the sampling effort

(PLN-)probabilistic PCA

PLN-PCA model. [CMR18]

- ▶ Low dimension latent vector

$$W_i \sim \mathcal{N}_r(0, I), \quad \text{where } r \ll p$$

- ▶ p -dimensional latent vector

$$Z_i = BW_i \quad \text{where } B(p \times r) = \text{loading matrix}$$

- ▶ Observed counts

$$Y_{ij} \sim \mathcal{P}(\exp(o_{ij} + x_i^T \beta + Z_{ij}))$$

o_{ij} = known 'offset' coefficient, accounting for the sampling effort

- ▶ Parameters

$$\theta = (\text{loading matrix } B, \text{ regression coefficient } \beta) \quad (+\text{rank } r)$$

Variational inference for PLN-PCA

VEM algorithm.

Variational inference for PLN-PCA

VEM algorithm.

- ▶ **VE step:** update the variational parameters $m_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$ and $S_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$
 - Similar to the VE step of regular PLN

Variational inference for PLN-PCA

VEM algorithm.

- ▶ **VE step:** update the variational parameters $m_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$ and $S_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$
 - Similar to the VE step of regular PLN
- ▶ **M step:** update the model parameters B^{h+1} and β^{h+1}
 - no close form, but still convex problem (gradient descent)

Model selection.

Variational inference for PLN-PCA

VEM algorithm.

- ▶ **VE step:** update the variational parameters $m_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$ and $S_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i^2)$
 → Similar to the VE step of regular PLN
- ▶ **M step:** update the model parameters B^{h+1} and β^{h+1}
 → no close form, but still convex problem (gradient descent)

Model selection.

- ▶ BIC penalty [Sch78] (Laplace approximation): $\text{pen}_{BIC}(\theta) = \underbrace{(p d)}_{\beta} + \underbrace{(p r)}_B \log n/2$

Variational inference for PLN-PCA

VEM algorithm.

- ▶ **VE step:** update the variational parameters $m_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$ and $S_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i^2)$
 → Similar to the VE step of regular PLN
- ▶ **M step:** update the model parameters B^{h+1} and β^{h+1}
 → no close form, but still convex problem (gradient descent)

Model selection.

- ▶ BIC penalty [Sch78] (Laplace approximation): $\text{pen}_{BIC}(\theta) = (\underbrace{pd}_{\beta} + \underbrace{pr}_{B}) \log n/2$
- ▶ Heuristic adaptation (replace $\log p_{\theta}(Y)$ with $J_{\theta,q}(Y)$)

$$vBIC = J_{\theta,q}(Y) - \text{pen}_{BIC}(\theta)$$

Variational inference for PLN-PCA

VEM algorithm.

- ▶ **VE step:** update the variational parameters $m_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i)$ and $S_i^{h+1} = \mathbb{E}_{q_i^{h+1}}(W_i^2)$
 → Similar to the VE step of regular PLN
- ▶ **M step:** update the model parameters B^{h+1} and β^{h+1}
 → no close form, but still convex problem (gradient descent)

Model selection.

- ▶ BIC penalty [Sch78] (Laplace approximation): $\text{pen}_{BIC}(\theta) = \underbrace{(p d)}_{\beta} + \underbrace{(p r)}_B \log n/2$
- ▶ Heuristic adaptation (replace $\log p_{\theta}(Y)$ with $J_{\theta,q}(Y)$)

$$vBIC = J_{\theta,q}(Y) - \text{pen}_{BIC}(\theta)$$

- ▶ Inspired from [BCG00] (additional penalty for the conditional entropy the W_i 's)

$$vICL = J_{\theta,q}(Y) - \text{pen}_{BIC}(\theta) - \mathcal{H}(q) = \mathbb{E}_q \log p_{\theta}(Y, Z) - \text{pen}_{BIC}(\theta)$$

Oak powdery mildew

Metabarcoding data [JFS⁺16]

- ▶ $p = 114$ OTUs
(66 bacteria and 48 fungi)
- ▶ $n = 116$ leaves
- ▶ collected on 3 trees
 - ▶ resistant
 - ▶ intermediate
 - ▶ susceptibleto oak powdery mildew;
- ▶ different protocols for bacteria and fungi
 o_{ij} = sequencing depth

Oak powdery mildew

Metabarcoding data [JFS⁺16]

- ▶ $p = 114$ OTUs
(66 bacteria and 48 fungi)

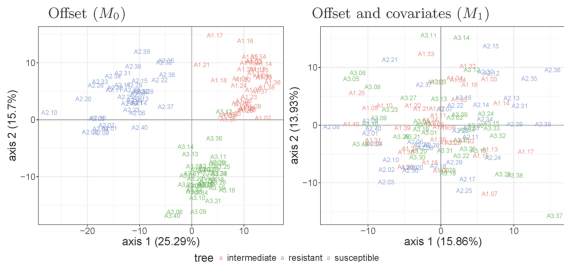
- ▶ $n = 116$ leaves

- ▶ collected on 3 trees

- ▶ resistant
- ▶ intermediate
- ▶ susceptible

to oak powdery mildew;

- ▶ different protocols for bacteria and fungi
 o_{ij} = sequencing depth



Oak powdery mildew

Metabarcoding data [JFS⁺16]

- ▶ $p = 114$ OTUs
(66 bacteria and 48 fungi)

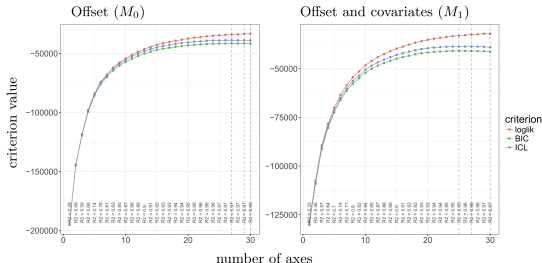
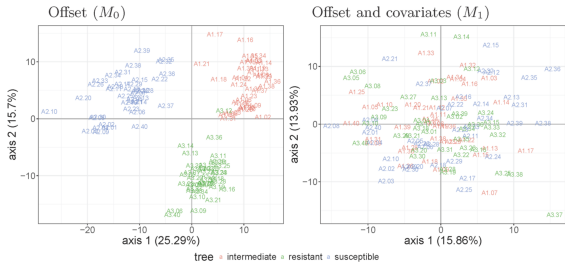
- ▶ $n = 116$ leaves

- ▶ collected on 3 trees

- ▶ resistant
- ▶ intermediate
- ▶ susceptible

to oak powdery mildew;

- ▶ different protocols for bacteria and fungi
 $o_{ij} =$ sequencing depth



Network inference

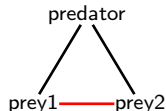
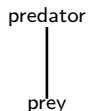
Species interaction networks.

- ▶ Aim: Understand how species from a same community interact
- ▶ Network representation = draw an edge between interacting pairs of species

Network inference

Species interaction networks.

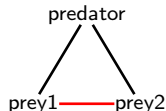
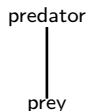
- ▶ Aim: Understand how species from a same community interact
- ▶ Network representation = draw an edge between interacting pairs of species
- ▶ Main issue: Distinguish **direct interactions** (predator-prey) from simple **associations** (two preys of a same predator)



Network inference

Species interaction networks.

- ▶ Aim: Understand how species from a same community interact
- ▶ Network representation = draw an edge between interacting pairs of species
- ▶ Main issue: Distinguish **direct interactions** (predator-prey) from simple **associations** (two preys of a same predator)

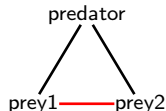
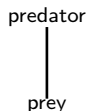


→ Obviously, analyses based on co-occurrences or correlations are not sufficient [PWT⁺19]

Network inference

Species interaction networks.

- ▶ Aim: Understand how species from a same community interact
- ▶ Network representation = draw an edge between interacting pairs of species
- ▶ Main issue: Distinguish **direct interactions** (predator-prey) from simple **associations** (two preys of a same predator)



→ Obviously, analyses based on co-occurrences or correlations are not sufficient [PWT⁺19]

Probabilistic translation.

association = marginal dependance

direct interaction = conditional dependance

Undirected graphical models

Definition. $p(U_1, \dots, U_k)$ is *faithful* to the (chordal) graph $G = ([k], E)$ iff

$$p(U_1, \dots, U_k) \propto \prod_{C \in \mathcal{C}} \psi_C(U_C)$$

where $\mathcal{C} = \{\text{cliques of } G\}$ and $U_C = (Y_j)_{j \in C}$.

Undirected graphical models

Definition. $p(U_1, \dots, U_k)$ is *faithful* to the (chordal) graph $G = ([k], E)$ iff

$$p(U_1, \dots, U_k) \propto \prod_{C \in \mathcal{C}} \psi_C(U_C)$$

where $\mathcal{C} = \{\text{cliques of } G\}$ and $U_C = (Y_j)_{j \in C}$.

Property.

separation \Leftrightarrow conditional independence

Undirected graphical models

Definition. $p(U_1, \dots, U_k)$ is *faithful* to the (chordal) graph $G = ([k], E)$ iff

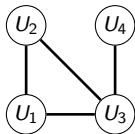
$$p(U_1, \dots, U_k) \propto \prod_{C \in \mathcal{C}} \psi_C(U_C)$$

where $\mathcal{C} = \{\text{cliques of } G\}$ and $U_C = (Y_j)_{j \in C}$.

Property.

separation \Leftrightarrow conditional independence

Example.



$$C_1 = \{1, 2, 3\}, C_2 = \{3, 4\}$$

Undirected graphical models

Definition. $p(U_1, \dots, U_k)$ is *faithful* to the (chordal) graph $G = ([k], E)$ iff

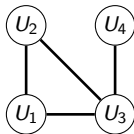
$$p(U_1, \dots, U_k) \propto \prod_{C \in \mathcal{C}} \psi_C(U_C)$$

where $\mathcal{C} = \{\text{cliques of } G\}$ and $U_C = (Y_j)_{j \in C}$.

Property.

separation \Leftrightarrow conditional independence

Example.



$C_1 = \{1, 2, 3\}, C_2 = \{3, 4\}$

$$p(U_1, U_2, U_3, U_4) \propto \psi_1(U_1, U_2, U_3) \psi_2(U_3, U_4)$$

- ▶ (U_1, U_2, U_3, U_4) all dependent
- ▶ $U_1 \not\perp U_2 \mid (U_3, U_4)$
- ▶ $U_4 \not\perp U_1 \mid U_2$
- ▶ $U_4 \perp (U_1, U_2) \mid U_3$

Gaussian graphical models

Gaussian graphical models

Suppose $Z \sim \mathcal{N}(0, \Sigma)$ and denote by $\Omega = [\omega_{jk}] = \Sigma^{-1}$ the *precision* matrix:

$$\sigma_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \quad (\text{'correlation'})$$

$$\omega_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \mid (Z_h)_{h \neq j, k} \quad (\text{'partial correlation'})$$

Gaussian graphical models

Suppose $Z \sim \mathcal{N}(0, \Sigma)$ and denote by $\Omega = [\omega_{jk}] = \Sigma^{-1}$ the *precision* matrix:

$$\sigma_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \quad (\text{'correlation'})$$

$$\omega_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \mid (Z_h)_{h \neq j, k} \quad (\text{'partial correlation'})$$

→ Ω only refers to 'direct' dependencies $\Rightarrow G$ given by the **support of Ω**

Gaussian graphical models

Suppose $Z \sim \mathcal{N}(0, \Sigma)$ and denote by $\Omega = [\omega_{jk}] = \Sigma^{-1}$ the *precision* matrix:

$$\sigma_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \quad (\text{'correlation'})$$

$$\omega_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \mid (Z_h)_{h \neq j, k} \quad (\text{'partial correlation'})$$

→ Ω only refers to 'direct' dependencies $\Rightarrow G$ given by the **support of Ω**

Graphical lasso. [FHT08]

► Common assumption: few species are in direct interaction

$\Rightarrow \Omega$ should be **sparse** (many 0's)

Gaussian graphical models

Suppose $Z \sim \mathcal{N}(0, \Sigma)$ and denote by $\Omega = [\omega_{jk}] = \Sigma^{-1}$ the *precision* matrix:

$$\sigma_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \quad (\text{'correlation'})$$

$$\omega_{jk} = 0 \Leftrightarrow (Z_j, Z_k) \text{ independent} \mid (Z_h)_{h \neq j, k} \quad (\text{'partial correlation'})$$

→ Ω only refers to 'direct' dependencies $\Rightarrow G$ given by the **support of Ω**

Graphical lasso. [FHT08]

- ▶ Common assumption: few species are in direct interaction

$\Rightarrow \Omega$ should be **sparse** (many 0's)

- ▶ Sparsity-inducing penalty (graphical lasso)

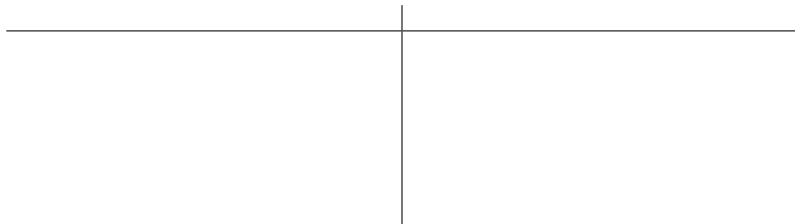
$$\max_{\Omega} \log p(Z; \Omega) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

Poisson log-normal model for network inference

PLN-network. PLN model with graphical lasso penalty [CMR19]

$$\arg \max_{\beta, \Omega, q \in \mathcal{Q}} J(\beta, \Omega, q) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

→ Convex problem for both the VE and the M step



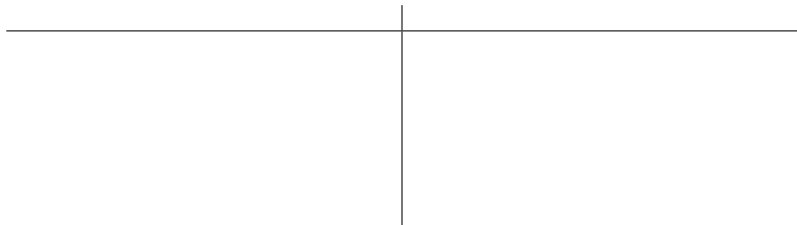
Poisson log-normal model for network inference

PLN-network. PLN model with graphical lasso penalty [CMR19]

$$\arg \max_{\beta, \Omega, q \in \mathcal{Q}} J(\beta, \Omega, q) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

→ Convex problem for both the VE and the M step

Inferring the *latent* dependency structure, not the abundance one



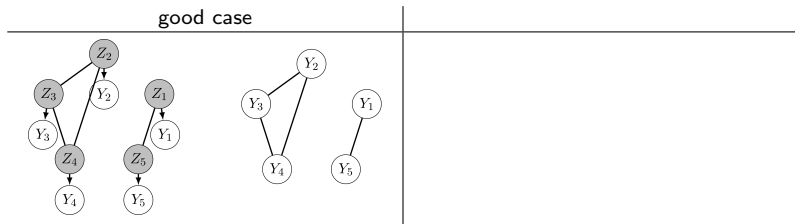
Poisson log-normal model for network inference

PLN-network. PLN model with graphical lasso penalty [CMR19]

$$\arg \max_{\beta, \Omega, q \in \mathcal{Q}} J(\beta, \Omega, q) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

→ Convex problem for both the VE and the M step

Inferring the *latent* dependency structure, not the abundance one



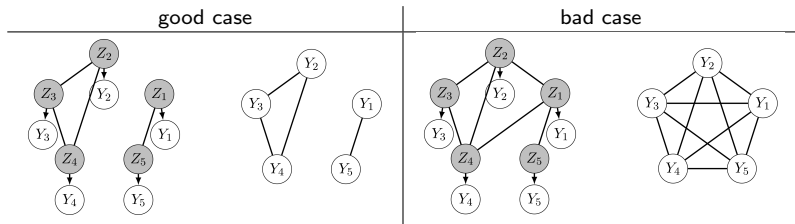
Poisson log-normal model for network inference

PLN-network. PLN model with graphical lasso penalty [CMR19]

$$\arg \max_{\beta, \Omega, q \in \mathcal{Q}} J(\beta, \Omega, q) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

→ Convex problem for both the VE and the M step

Inferring the *latent* dependency structure, not the abundance one



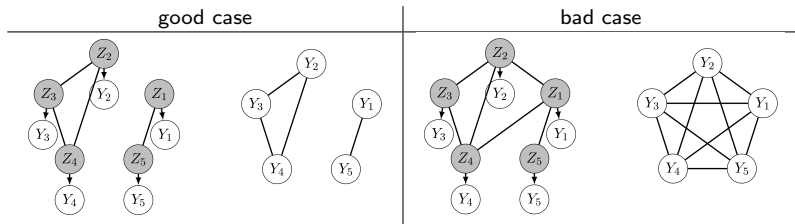
Poisson log-normal model for network inference

PLN-network. PLN model with graphical lasso penalty [CMR19]

$$\arg \max_{\beta, \Omega, q \in \mathcal{Q}} J(\beta, \Omega, q) - \lambda \underbrace{\sum_{j \neq k} |\omega_{jk}|}_{\ell_1 \text{ penalty}}$$

→ Convex problem for both the VE and the M step

Inferring the *latent* dependency structure, not the abundance one

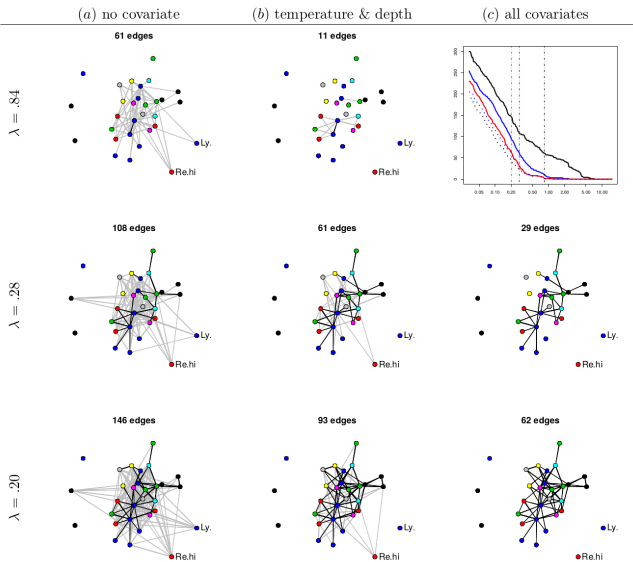


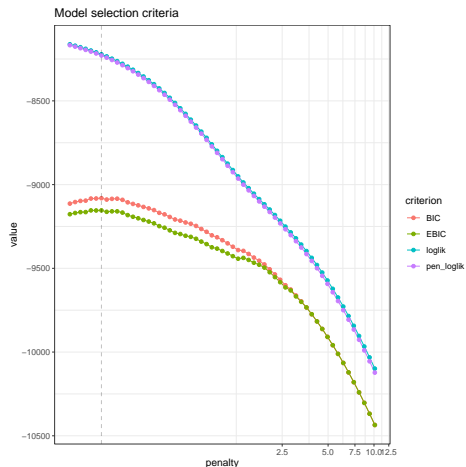
→ Similar setting for most approaches in statistical ecology [WBO⁺15, KMM⁺15, FHZD17, PHW18]

Barents' fish species

Data:

- ▶ $n = 89$ sites
- ▶ $p = 30$ species
- ▶ $d = 4$ covariates
 - ▶ latitude
 - ▶ longitude
 - ▶ temperature
 - ▶ depth



Barents' fish species: choosing λ 

Alternatively.

Use resampling and select edges based on selection frequency

[LRW10]

Outline

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

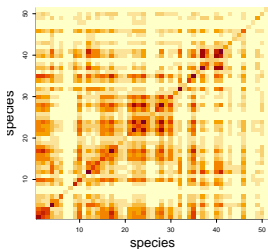
To summarize

Stochastic block-model for ecological networks

Data:

- ▶ n species
- ▶ Y_{ij} = 'intensity' (e.g. count) of the link between species i and j

Adjacency matrix.

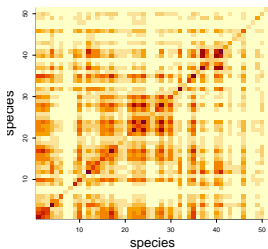


Stochastic block-model for ecological networks

Data:

- ▶ n species
- ▶ Y_{ij} = 'intensity' (e.g. count) of the link between species i and j

Adjacency matrix.



Stochastic block-model.

- ▶ K groups
- ▶ Latent group membership

$$Z_i \sim \mathcal{M}(1, (\pi_1, \dots, \pi_K))$$

- ▶ Observed count

$$Y_{ij} \sim \mathcal{P}(\exp(\alpha_{Z_i, Z_j}))$$

- ▶ Parameters

$$\theta = (\pi, \alpha)$$

+K

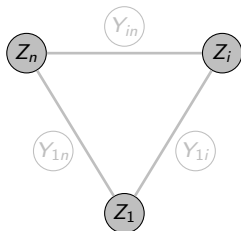
Variational inference

Conditional distribution.

- ▶ Group memberships:

$$Z_i \perp Z_j \quad \text{but} \quad Z_i \not\perp Z_j \mid Y_{ij}$$

- ▶ $p_\theta(Z \mid Y)$ is intractable



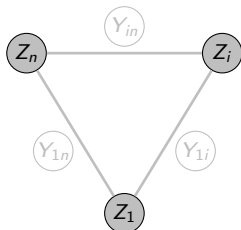
Variational inference

Conditional distribution.

- ▶ Group memberships:

$$Z_i \perp Z_j \quad \text{but} \quad Z_i \not\perp Z_j \mid Y_{ij}$$

- ▶ $p_\theta(Z \mid Y)$ is intractable



Variational approximation. Use a factorable approximate distribution

$$\mathcal{Q} = \left\{ q : q(Z) = \prod_i q_i(Z_i), \underbrace{q_i(Z_i) = \mathcal{M}(Z_i; 1, \tau_i)}_{\text{no approximation}} \right\}$$

- ▶ Variational parameters: $\tau_{ik} \simeq \Pr(Z_i = k \mid Y)$

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

- VE step: update the variational parameters τ_i

$$\tau_{ik}^{h+1} \propto \pi_k^h \prod_{j \neq i} \prod_{\ell} p_{\theta^h}(Y_{ij} \mid Z_i = k, Z_j = \ell)^{\tau_{j\ell}^{h+1}}$$

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

- VE step: update the variational parameters τ_i

$$\tau_{ik}^{h+1} \propto \pi_k^h \prod_{j \neq i} \prod_{\ell} p_{\theta^h}(Y_{ij} \mid Z_i = k, Z_j = \ell)^{\tau_{j\ell}^{h+1}}$$

→ Fix-point algorithm

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

- VE step: update the variational parameters τ_i

$$\tau_{ik}^{h+1} \propto \pi_k^h \prod_{j \neq i} \prod_{\ell} p_{\theta^h}(Y_{ij} \mid Z_i = k, Z_j = \ell)^{\tau_{j\ell}^{h+1}}$$

→ Fix-point algorithm

- M step: update the model parameters π, α

$$\theta^{h+1} = \arg \max_{\theta} \mathbb{E}_{q^{h+1}} \log p_{\theta}(Y, Z)$$

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

- VE step: update the **variational parameters** τ_i

$$\tau_{ik}^{h+1} \propto \pi_k^h \prod_{j \neq i} \prod_{\ell} p_{\theta^h}(Y_{ij} \mid Z_i = k, Z_j = \ell)^{\tau_{j\ell}^{h+1}}$$

→ Fix-point algorithm

- M step: update the **model parameters** π, α

$$\theta^{h+1} = \arg \max_{\theta} \mathbb{E}_{q^{h+1}} \log p_{\theta}(Y, Z)$$

→ Close form for both π^{h+1} and α^{h+1}

Variational EM

Variational EM algorithm. `blockmodels` R package [Lég16]

- VE step: update the variational parameters τ_i

$$\tau_{ik}^{h+1} \propto \pi_k^h \prod_{j \neq i} \prod_{\ell} p_{\theta^h}(Y_{ij} | Z_i = k, Z_j = \ell)^{\tau_{j\ell}^{h+1}}$$

→ Fix-point algorithm

- M step: update the model parameters π, α

$$\theta^{h+1} = \arg \max_{\theta} \mathbb{E}_{q^{h+1}} \log p_{\theta}(Y, Z)$$

→ Close form for both π^{h+1} and α^{h+1}

Model selection. To choose the number of groups K : $vBIC$ or $vICL$ with penalty

$$\text{pen}_{BIC}(\theta) = \underbrace{(K-1) \frac{\log n}{2}}_{\text{node memberships}} + \underbrace{\frac{K(K+1)}{2} \frac{\log(n-1)}{2}}_{\text{node links}}$$

A first illustration: Tree species network

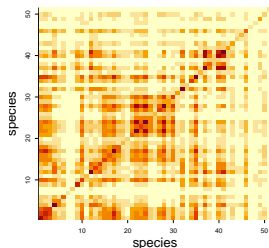
Simple model: No covariate

$$Y_{ij} \sim \mathcal{P}(\exp(\alpha_{Z_i Z_j}))$$

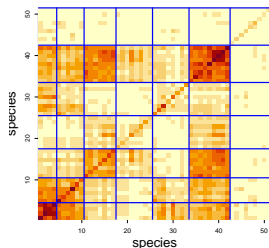
Y_{ij} = number of shared fungal parasites

$$\hat{K}_{ICL} = 7$$

adjacency matrix Y



clustered matrix



A first illustration: Tree species network

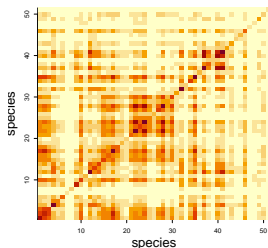
Simple model: No covariate

$$Y_{ij} \sim \mathcal{P}(\exp(\alpha_{Z_i Z_j}))$$

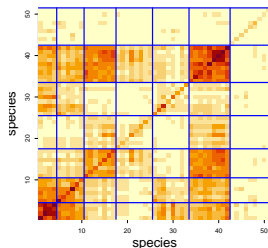
Y_{ij} = number of shared fungal parasites

$$\hat{K}_{ICL} = 7$$

adjacency matrix Y



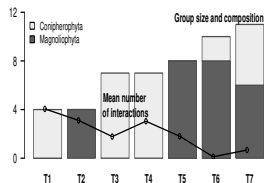
clustered matrix



'Validation'

comparison with the
phylogenetic classification

(coniferophyta vs
magnoliophyta)



Outline

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

To summarize

Accounting for covariates

Adding a regression term.

- ▶ Information about similarity or dissimilarity between species is often available
→ taxonomic, phylogenetic or geographic distance

- ▶ Obvious generalization of the stochastic block-model [MRV10]:

$$Y_{ij} \sim \mathcal{P}(\exp(\alpha_{Z_i Z_j} + x_{ij}^T \beta))$$

→ x_{ij} = vector of covariates for the pair (i, j)

- ▶ Parameters: $\theta = (\pi, \alpha, \beta)$

Accounting for covariates

Adding a regression term.

- ▶ Information about similarity or dissimilarity between species is often available
→ taxonomic, phylogenetic or geographic distance

- ▶ Obvious generalization of the stochastic block-model [MRV10]:

$$Y_{ij} \sim \mathcal{P}(\exp(\alpha_{Z_i Z_j} + \mathbf{x}_{ij}^T \beta))$$

→ \mathbf{x}_{ij} = vector of covariates for the pair (i, j)

- ▶ Parameters: $\theta = (\pi, \alpha, \beta)$

Variational EM algorithm. [MRV10]

- ▶ Very similar to SBM without covariates
- ▶ Estimation of β via weighted generalized linear model

Tree species network

Covariate:

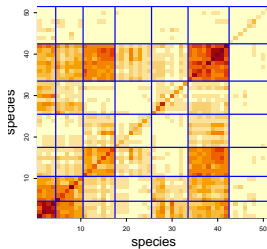
x_{ij} = taxonomic distance

Estimates:

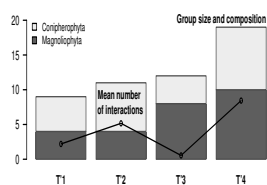
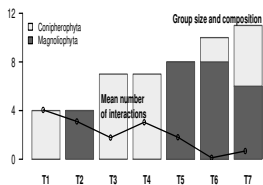
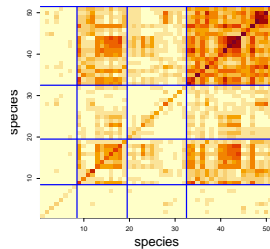
$$\hat{K}_{ICL} = 4$$

$$\hat{\beta} = -.317$$

No covariate: $\hat{K}_{ICL} = 7$



Taxonomic dist.: $\hat{K}_{ICL} = 4$



Tree species network

Covariate:

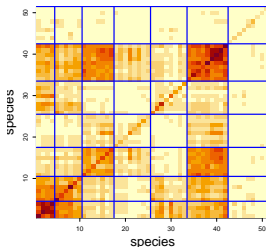
x_{ij} = taxonomic distance

Estimates:

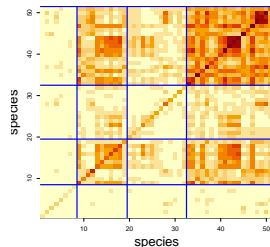
$$\widehat{K}_{ICL} = 4$$

$$\widehat{\beta} = -.317$$

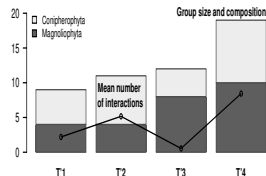
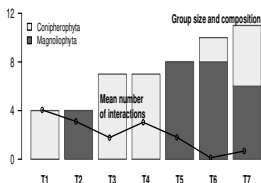
No covariate: $\widehat{K}_{ICL} = 7$



Taxonomic dist.: $\widehat{K}_{ICL} = 4$



- Taxonomy (partially) explains the links (smaller \widehat{K})
- Distant species share less parasites ($\widehat{\beta} < 0$)
- The remaining structure is not related to taxonomy



Animal behavior

Data: [RSF⁺15]

- ▶ Consider n individuals (animals) along T times (days, weeks)

- ▶ At each time, observe

Y_{ij}^t = intensity of the social interaction between individuals i and j at time t

Animal behavior

Data: [RSF⁺15]

- ▶ Consider n individuals (animals) along T times (days, weeks)
- ▶ At each time, observe

Y_{ij}^t = intensity of the social interaction between individuals i and j at time t

Questions:

- ▶ Do the individuals play different roles in the social network
- ▶ Do these roles change over time

Dynamic SBM

Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$

Dynamic SBM

Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$

$$Z_1^1$$

$$Z_i^1$$

$$Z_n^1$$

Dynamic SBM

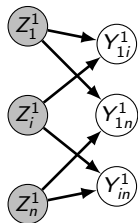
Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$



Dynamic SBM

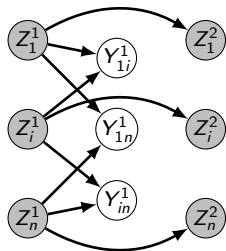
Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$



Dynamic SBM

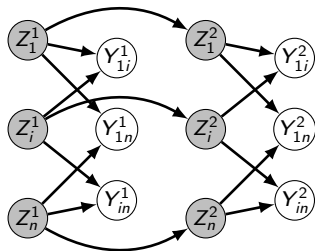
Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$



Dynamic SBM

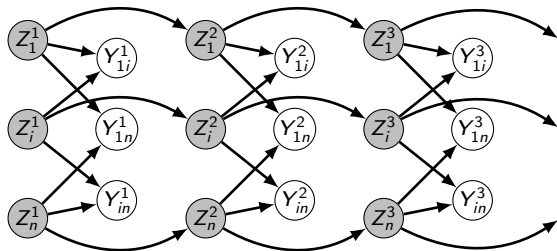
Dynamic stochastic block-model. [MM17]

- ▶ Assume that individuals belong to K clusters ('roles')
- ▶ Denote by Z_i^t the (latent) role of individual i at time t
- ▶ The successive roles of each individuals are independent Markov chains

$$Z_i = \{Z_i^t\}_{1 \leq t \leq T} \sim MC(\nu_1, \pi)$$

- ▶ Social interactions are conditionally independent

$$\{Y_{ij}^t\}_{i,j,t} \text{ independent} \mid \{Z_i^t\}_{i,t}, \quad Y_{ij}^t \mid Z_i^t, Z_j^t \sim F(\cdot; \gamma_{Z_i^t, Z_j^t})$$

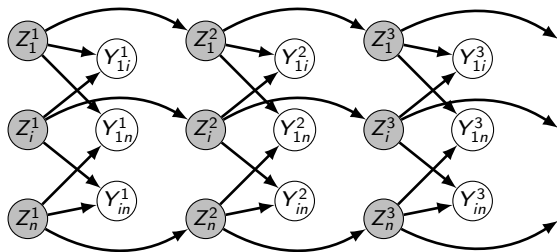


Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states

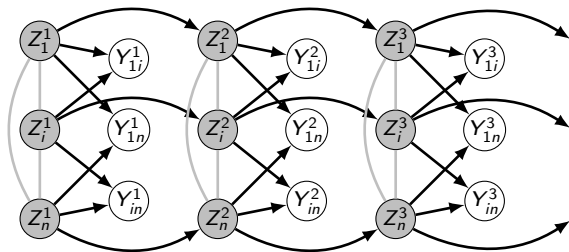
Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states



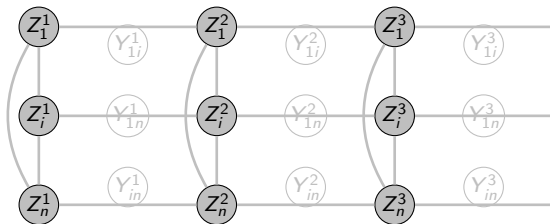
Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states



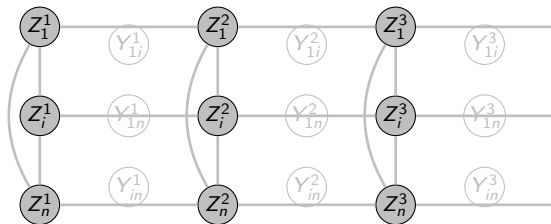
Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states



Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states

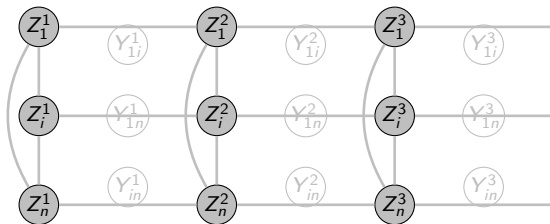


Approximation classe. $p_\theta(Z | Y) \simeq q(Z) =$ product of independent Markov chains (partial factorization)

$$\mathcal{Q} = \left\{ q : q(Z) = \prod_i q_i(Z_i), \quad q_i(Z_i) = q_i(Z_i^1) \prod_{t>1} q_i(Z_i^t | Z_i^{t-1}) \right\}$$

Variational EM

Intractable EM. Denoting $Z^t = (Z_1^t, \dots, Z_n^t)$, $(Z^t | Y)_{t \geq 1}$ is a Markov chain ... with K^n states



Approximation classe. $p_\theta(Z | Y) \simeq q(Z) =$ product of independent Markov chains (partial factorization)

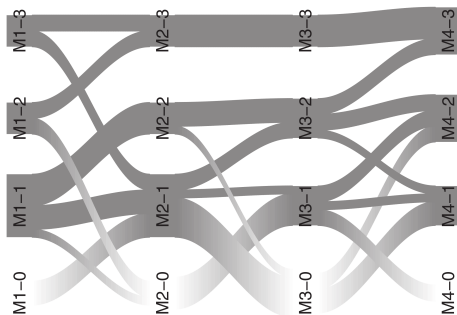
$$\mathcal{Q} = \left\{ q : q(Z) = \prod_i q_i(Z_i), \quad q_i(Z_i) = q_i(Z_i^1) \prod_{t>1} q_i(Z_i^t | Z_i^{t-1}) \right\}$$

VEM algorithm.

- VE step = running n forward-backward recursions

Onager social network

Data from [RSF⁺15]. $n = 23$ onagers, observations gathered into $T = 4$ time periods in [MM17].



- ▶ 4 groups (= 'roles') are found, from isolated to highly central
- ▶ A fraction of individuals do change role from one period to another

Latent block-model for comparative genomics

Comparative metagenomics.

- ▶ n samples (soil surrounding the root of a plant – *rhizosphere* – with given genotype), p bacterial species (*Operational Taxonomy Units* = OTUs),
- ▶ Y_{ij} = number of reads from species j in sample i

$${}^1 Y \sim \text{NegBin} \quad \Leftrightarrow \quad Y \sim \mathcal{P}(\lambda U) \quad \text{with } U \sim \text{Gamma}.$$

Latent block-model for comparative genomics

Comparative metagenomics.

- ▶ n samples (soil surrounding the root of a plant – *rhizosphere* – with given genotype), p bacterial species (*Operational Taxonomy Units* = OTUs),
- ▶ Y_{ij} = number of reads from species j in sample i
- ▶ **Question:** Do preferential (or negative) associations exist between groups of genotypes and groups of bacteria?
- ▶ **Over-dispersion:** Due to technological variability, counts are over-dispersed wrt Poisson
→ Negative-binomial (= Poisson-Gamma¹) distribution for the count

¹ $Y \sim \text{NegBin} \Leftrightarrow Y \sim \mathcal{P}(\lambda U)$ with $U \sim \text{Gamma}$.

Latent block-model for comparative genomics

Model.

²The higher, the less dispersed.

Latent block-model for comparative genomics

Model.

- ▶ $\{Z_i\}_{1 \leq i \leq n}$ sample memberships (among K groups) $\pi =$ proportions of sample groups

$$Z_i \sim \mathcal{M}(1, \pi)$$

²The higher, the less dispersed.

Latent block-model for comparative genomics

Model.

- ▶ $\{Z_i\}_{1 \leq i \leq n}$ sample memberships (among K groups) $\pi =$ proportions of sample groups

$$Z_i \sim \mathcal{M}(\mathbf{1}, \pi)$$

- ▶ $\{W_j\}_{1 \leq j \leq p}$ species memberships (among L groups) $\rho =$ proportions of species groups

$$W_j \sim \mathcal{M}(\mathbf{1}, \rho)$$

²The higher, the less dispersed.

Latent block-model for comparative genomics

Model.

- ▶ $\{Z_i\}_{1 \leq i \leq n}$ sample memberships (among K groups) $\pi =$ proportions of sample groups

$$Z_i \sim \mathcal{M}(1, \pi)$$

- ▶ $\{W_j\}_{1 \leq j \leq p}$ species memberships (among L groups) $\rho =$ proportions of species groups

$$W_j \sim \mathcal{M}(1, \rho)$$

- ▶ $\{U_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ random effects $a =$ overdispersion parameter²

$$U_{ij} \sim \mathcal{G}(\text{Gamma}(a, a))$$

²The higher, the less dispersed.

Latent block-model for comparative genomics

Model.

- ▶ $\{Z_i\}_{1 \leq i \leq n}$ sample memberships (among K groups) $\pi =$ proportions of sample groups

$$Z_i \sim \mathcal{M}(1, \pi)$$

- ▶ $\{W_j\}_{1 \leq j \leq p}$ species memberships (among L groups) $\rho =$ proportions of species groups

$$W_j \sim \mathcal{M}(1, \rho)$$

- ▶ $\{U_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ random effects $a =$ overdispersion parameter²

$$U_{ij} \sim \text{Gamma}(a, a)$$

- ▶ $\{Y_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ observed counts $\mu_j =$ mean (log-)abundance of species j

$$Y_{ij} \sim \mathcal{P}(\exp(o_i + \mu_j + \alpha_{Z_i} W_j + \log U_{ij}))$$

$o_{ij} =$ known sampling effort for species j in sample i

²The higher, the less dispersed.

Latent block-model for comparative genomics

Model.

- ▶ $\{Z_i\}_{1 \leq i \leq n}$ sample memberships (among K groups) $\pi =$ proportions of sample groups

$$Z_i \sim \mathcal{M}(\mathbf{1}, \pi)$$

- ▶ $\{W_j\}_{1 \leq j \leq p}$ species memberships (among L groups) $\rho =$ proportions of species groups

$$W_j \sim \mathcal{M}(\mathbf{1}, \rho)$$

- ▶ $\{U_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ random effects $a =$ overdispersion parameter²

$$U_{ij} \sim \mathcal{G}(\text{amma}(a, a)$$

- ▶ $\{Y_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ observed counts $\mu_j =$ mean (log-)abundance of species j

$$Y_{ij} \sim \mathcal{P}(\exp(o_i + \mu_j + \alpha_{Z_i} W_j + \log U_{ij}))$$

$o_{ij} =$ known sampling effort for species j in sample i

Parameters.

$$\theta = (\pi, \rho, a, \alpha, \mu) \quad + (K, L)$$

²The higher, the less dispersed.

Rhizosphere clustering

Variational EM. Using

$$q(Z, W, U) = q_Z(Z) q_W(W) q_U(U)$$

Model selection with $vICL$ including $\mathcal{H}(q_Z)$ and $\mathcal{H}(q_W)$

Rhizosphere clustering

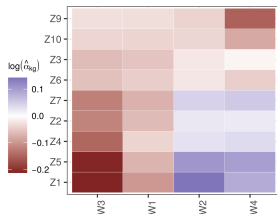
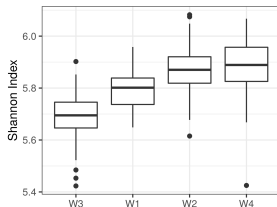
Variational EM. Using

$$q(Z, W, U) = q_Z(Z) q_W(W) q_U(U)$$

Model selection with $vICL$ including $\mathcal{H}(q_Z)$ and $\mathcal{H}(q_W)$

Results.

- ▶ $\hat{K} = 4$ sample groups, $\hat{L} = 10$ bacteria groups
- ▶ Contrasted interactions: $\alpha_{kg} \in [-.5, 1.2]$
- ▶ Sample groups display different biodiversity (Shannon index)



$^0(Z, W)$ inverted in the figure

Outline

Poisson log-normal model

Illustration

Extensions of the Poisson log-normal model

Dimension reduction

Network inference

Block-models

Illustration

Extensions of block-models

Covariates

Dynamic SBM

Metagenomics

To summarize

To summarize

VEM for latent models.

- ▶ Latent variable models: flexible and explicit framework for modelling
- ▶ Variational approximation: efficient approach for their inference
 - Mostly rely on the choice of the approximation class

To summarize

VEM for latent models.

- ▶ Latent variable models: flexible and explicit framework for modelling
- ▶ Variational approximation: efficient approach for their inference
 - Mostly rely on the choice of the approximation class

Many other problems/models.

- ▶ Account for a spatial structure, fundamental niche vs realized niche, looking for some structured in an inferred network, ...

To summarize

VEM for latent models.

- ▶ Latent variable models: flexible and explicit framework for modelling
- ▶ Variational approximation: efficient approach for their inference
 - Mostly rely on the choice of the approximation class

Many other problems/models.

- ▶ Account for a spatial structure, fundamental niche vs realized niche, looking for some structured in an inferred network, ...







Statistical guarantees.

- ▶ General properties of variational estimates?
- ▶ Combining VEM with other inference methods

References I

- C. Bernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Machine Intel.*, 22(7):719–25, 2000.
- J. Chiquet, M. Mariadassou, and S. Robin. Variational inference for probabilistic Poisson PCA. *The Annals of Applied Statistics*, 12(4):2674–2698, 2018.
- J. Chiquet, M. Mariadassou, and S. Robin. Variational inference for sparse network reconstruction from count data. In *International Conference on Machine Learning*, pages 1162–1171, 2019.
- J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- H. Fang, C. Huang, H. Zhao, and M. Deng. gCoda: conditional dependence network inference for compositional data. *Journal of Computational Biology*, 24(7):699–708, 2017.
- E. Frenkel, V. Fievet, L. Schwaller, T. Fort, C. Robin, and C. Vacher. Deciphering the pathobiome: Intra-and interkingdom interactions involving the pathogen *Erysiphe alphitoides*. *Microbial ecology*, pages 1–11, 2016.
- Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, and R. A. Bonneau. Sparse and compositionally robust inference of microbial ecological networks. *PLoS computational biology*, 11(5):e1004226, 2015.
- J. Léger. Blockmodels: A R-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. Technical report, arXiv:1602.07587, 2016.
- H. Liu, K. Roeder, and L. Wasserman. Stability approach to regularization selection (StARS) for high dimensional graphical models. In *Advances in neural information processing systems*, pages 1432–1440, 2010.
- C. Matias and V. Miele. Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1119–1141, 2017.
- M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, pages 715–742, 2010.

References II

-  Popovic, F. KC Hui, and D. I Warton. A general algorithm for covariance modeling of discrete data. *Journal of Multivariate Analysis*, 165:86–100, 2018.
-  Popovic, D. I. Warton, F. J. Thomson, F. K. C. Hui, and A. T. Moles. Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, 10(9):1571–1583, 2019.
-  Rubenstein, S. R Sundaesan, I. R Fischhoff, C. Tantipathananandh, and T. Y Berger-Wolf. Similar but different: dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and Grevy's zebra. *PLoS one*, 10(10):e0138645, 2015.
-  Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.
-  Tipping and C. M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, 61(3):611–622, 1999.
-  Warton, F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C Walker, and F. KC. Hui. So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution*, 30(12):766–779, 2015.