

Inférence de réseaux à partir de mélanges d'arbres

Encadré par S. Robin¹ et C. Ambroise¹²

Raphaëlle Momal-Leisenring

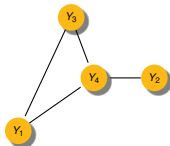
¹UMR AgroParisTech / INRA MIA-Paris

²LaMME, Evry

3 avril 2018

Réseau

- Réseau : représentation graphique de la structure de dépendance conditionnelle d'un jeu de données.
- Inférer un réseau : inférer les arêtes du graph, *i.e.* la structure de dépendance, à partir d'observations.



Les variables Y_2 et Y_1 sont indépendantes entre elles conditionnellement à la variable Y_4 .

- Tableau de données Y de dimensions $n \times d$

Exemple de réseau écologique

[Jakuschkin et al., 2016] :

- But : à partir de mesures d'abondance, identifier les liens de dépendance entre le champignon *E. alphitoïde* présent sur les feuilles du chêne, et les autres micro-organismes présents.
- Utile à la compréhension et au contrôle des maladies chez le chêne.

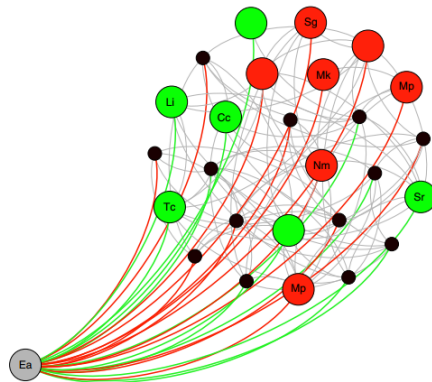


FIGURE – Model of the pathobiome *Erysiphe alphitoides* on oak leaves, source : Jakuschkin et al.

Modèle graphique

- Clique C d'un graphe G : sous-ensemble de noeuds de G qui sont tous liés entre eux.
- Clique maximale C_G : aucune autre clique de G ne la contient strictement.

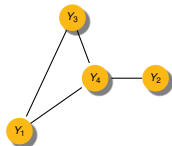
Propriété modèle graphique [Lauritzen, 1996]

Soit $Y = (Y_1, \dots, Y_q)$ et p sa densité. p se factorise selon le graphe non orienté G si :

$$p(y) \propto \prod_{C \in C_G} \Phi_C(y^C)$$

Et alors G représente la structure d'indépendance conditionnelle entre les Y_i .

Exemple : $Y = (Y_1, \dots, Y_4)$:



$$p(Y) = \phi_1(Y_2, Y_4) \times \phi_2(Y_1, Y_3, Y_4)$$

Cas gaussien (GGM, *Gaussian Graphical Models*)

Soit Y une variable gaussienne multivariée de dimension d :

$$Y = (Y_1, \dots, Y_d) \sim \mathcal{N}_d(0, \Omega^{-1}),$$
$$\Omega = (w_{ij})_{i,j}.$$

L'écriture de la gaussienne permet directement d'obtenir une factorisation :

$$p(y) \propto \exp(-y^T \Omega y / 2)$$
$$\propto \prod_{j,k, w_{jk} \neq 0} \exp(-y_j w_{jk} y_k / 2)$$

Cas gaussien (GGM, *Gaussian Graphical Models*)

Soit Y une variable gaussienne multivariée de dimension d :

$$Y = (Y_1, \dots, Y_d) \sim \mathcal{N}_d(0, \Omega^{-1}),$$

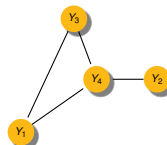
$$\Omega = (w_{ij})_{i,j}.$$

L'écriture de la gaussienne permet directement d'obtenir une factorisation :

$$p(y) \propto \exp(-y^T \Omega y / 2)$$

$$\propto \prod_{j,k, w_{jk} \neq 0} \exp(-y_j w_{jk} y_k / 2)$$

$$\Omega = \begin{pmatrix} * & 0 & * & * \\ 0 & * & 0 & * \\ * & 0 & * & * \\ * & * & * & * \end{pmatrix}$$

 \Rightarrow


Inférence de Ω : le graphical Lasso

- La log-vraisemblance de Y s'écrit :

$$L(Y, \Omega) = \frac{n}{2} \log(\det(\Omega)) - \frac{n}{2} Y^T \Omega Y + cste$$

- Estimation parcimonieuse

Le graphical-Lasso (glasso) :

Le graphical-Lasso pénalise la norme l_1 de la matrice de précision :

$$\hat{\Omega}_\lambda = \arg \min_{\Omega \in S_d^+} \left\{ L(Y, \Omega) + \lambda \sum_{i \neq j} |w_{ij}| \right\}$$

- Choix du λ ...

Arbre de dépendance

- La structure de dépendance des données s'appuie sur un arbre
- La vraisemblance de données continues se factorise sur les noeuds et les arêtes [Chow and Liu, 1968] :

$$\mathbb{P}(Y|\mathcal{T}) = \prod_{j=1}^d \mathbb{P}(Y_j) \prod_{k,l \in \mathcal{T}} \psi_{kl}(Y) \quad ,$$

Où

$$\psi_{kl}(Y) = \frac{\mathbb{P}(Y_k, Y_l)}{\mathbb{P}(Y_k) \times \mathbb{P}(Y_l)}.$$

- Cas gaussien centré réduit :

$$\log(\hat{\mathbb{P}}(Y|\mathcal{T})) = \sum_{k,l \in \mathcal{T}} \underbrace{-\frac{n}{2} \log(1 - \hat{\rho}_{kl}^2)}_{\log(\hat{\psi}_{kl})} + \sum_k -\frac{n}{2} \log(\hat{\sigma}_k^2) + cste$$

Modèle de mélange d'arbres

- 1 Un arbre est tiré dans un mélange d'arbres. En fait un mélange sur les arêtes du graphe :
 - Un poids β_{kl} est attribué à chaque arête (k, l) possible du graphe.
 - La probabilité de l'arbre de dépendance des données s'écrit alors

Loi d'un arbre T

$$\mathbb{P}(T) = \frac{1}{B} \prod_{(k,l) \in T} \beta_{kl}, \text{ avec } B = \sum_{T \in \mathcal{T}} \prod_{(k,l) \in T} \beta_{kl}$$

- Les poids sont en général fixés à l'avance, ici non ...
- 2 Les données sont simulées conditionnellement à l'arbre tiré :

$$Y|T \sim \mathcal{N}_d(0, \Omega_T^{-1})$$

L'arbre qui structure les données est traité comme une **variable latente**.

Algorithme EM : étape E

But : maximiser l'espérance conditionnelle de la log-vraisemblance complète.

$$\mathbb{P}(Y, T) = \mathbb{P}(T) \times \mathbb{P}(Y|T)$$

$$\log(\mathbb{P}(Y, T)) = \sum_{k, l \in V} \mathbb{1}_{\{(k, l) \in E_T\}} (\log(\beta_{kl}) + \log(\psi_{kl})) - \log(B) + \sum_k \log(\mathbb{P}(Y_k))$$

Espérance conditionnelle :

$$\begin{aligned} \mathbb{E}_\theta[\log(\mathbb{P}(Y, T)) | Y] &= \sum_{k, l \in V} \mathbb{P}((k, l) \in E_T | Y) (\log(\beta_{kl}) + \log(\psi_{kl})) \\ &\quad - \log(B) + \sum_k \log(\mathbb{P}(Y_k)) \end{aligned}$$

Calcul de la probabilité conditionnelle

Théorème de Kirchhoff (matrix tree, [Chaiken and Kleitman, 1978])

Pour toute matrice symétrique $W = (a_{kl})_{k,l}$, son Laplacien $Q(W)$ se définit par :

$$Q_{uv}(W) = \begin{cases} -a_{uv} & 1 \leq u < v \leq n \\ \sum_{w=1}^n a_{uw} & 1 \leq u = v \leq n. \end{cases}$$

Alors pour tout u et v :

$$|Q_{uv}^*(W)| = \sum_{T \in \mathcal{T}} \prod_{\{k,l\} \in E_T} a_{kl}$$

Calcul de la probabilité conditionnelle

Théorème de Kirchhoff (matrix tree, [Chaiken and Kleitman, 1978])

Pour toute matrice symétrique $W = (a_{kl})_{k,l}$, son Laplacien $Q(W)$ se définit par :

$$Q_{uv}(W) = \begin{cases} -a_{uv} & 1 \leq u < v \leq n \\ \sum_{w=1}^n a_{uw} & 1 \leq u = v \leq n. \end{cases}$$

Alors pour tout u et v :

$$|Q_{uv}^*(W)| = \sum_{T \in \mathcal{T}} \prod_{\{k,l\} \in E_T} a_{kl}$$

$$\begin{aligned} \mathbb{P}((k,l) \in T | Y) &= \sum_{T \in \mathcal{T}: (k,l) \in T} \mathbb{P}(T | Y) = \frac{\sum_{(k,l) \in T} \mathbb{P}(T) \mathbb{P}(Y | T)}{\sum_T \mathbb{P}(T) \mathbb{P}(Y | T)} \\ &= 1 - \frac{|Q_{uv}^*(B\Psi^{-kl})|}{|Q_{uv}^*(B\Psi)|} \\ &= \tau_{kl} \end{aligned}$$

Algorithme EM : étape M

But : optimiser les poids β_{kl} .

$$\arg \max_{\beta_{kl}} \left\{ \sum_{k,l \in V} \tau_{kl} (\log(\beta_{kl}) + \log(\psi_{kl})) - \log(B) + \sum_k \log(\mathbb{P}(Y_k)) \right\}$$

Avec :

$$B = \sum_{T \in \mathcal{T}} \prod_{k,l \in T} \beta_{kl}.$$

Comment calculer $\frac{\partial B}{\partial \beta_{kl}}$?

Mise à jour des β_{kl}

Résultat de Meila [Meila and Jordan, 2000]

En inversant un mineur du Laplacien Q , on définit la matrice symétrique M :

$$\begin{cases} M_{uv} = [Q^{*-1}]_{uu} + [Q^{*-1}]_{vv} - 2[Q^{*-1}]_{uv} & u, v < n \\ M_{nv} = M_{vn} = [Q^{*-1}]_{vv} & v < n \\ M_{vv} = 0. \end{cases}$$

On peut montrer que :

$$\frac{\partial |Q_{uv}^*(W)|}{\partial \beta_{kl}} = M_{kl} \times |Q_{uv}^*(W)|$$

$$\frac{\partial \mathbb{E}_\theta[\log(\mathbb{P}(Y, T)) | Y]}{\partial \beta_{kl}} = \frac{1}{\beta_{kl}} \tau_{kl} - \frac{1}{B} \frac{\partial B}{\partial \beta_{kl}}$$

Formule de mise à jour à l'itération $h + 1$

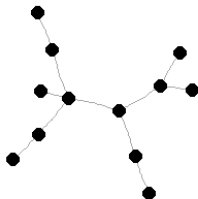
$$\hat{\beta}_{kl}^{h+1} = \frac{\tau_{kl}^h}{M_{kl}^h}$$

Plan de simulation

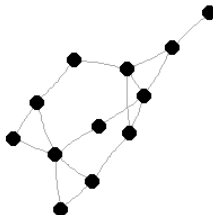
- 1 Tirer un graphe G
- 2 À partir de la matrice d'adjacence, construire Ω qui soit défini positive
- 3 Simuler les données Y à partir de Ω
- 4 Appliquer le glasso et notre EM aux données
- 5 Comparer les graphes obtenus et G

Les graphes G

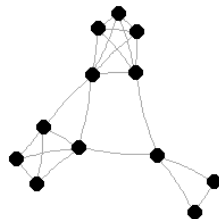
Arbre



Erdos



Cluster

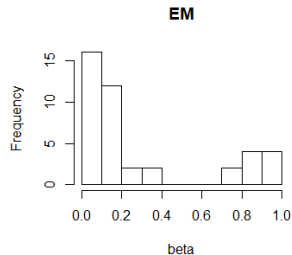
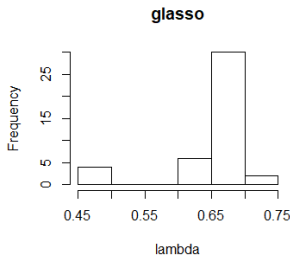
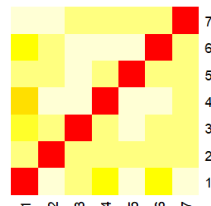


Paramètres : nombre de noeuds, densité d'arêtes...

Comparaison des graphes inférés avec G

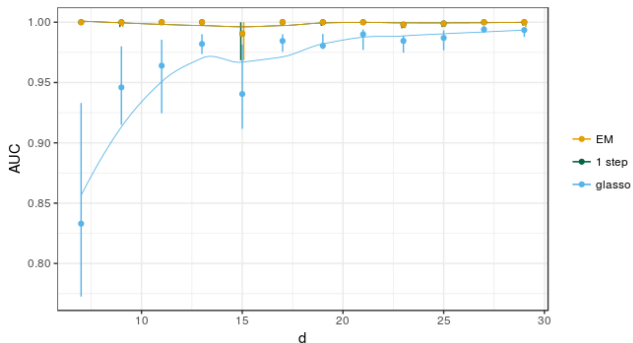
Les méthodes renvoient des matrices de scores pour chaque arête :

- Glasso : pénalités λ nécessaires pour annuler chacune des arêtes
- EM : poids β_{kl}



Résultat sur les arbres

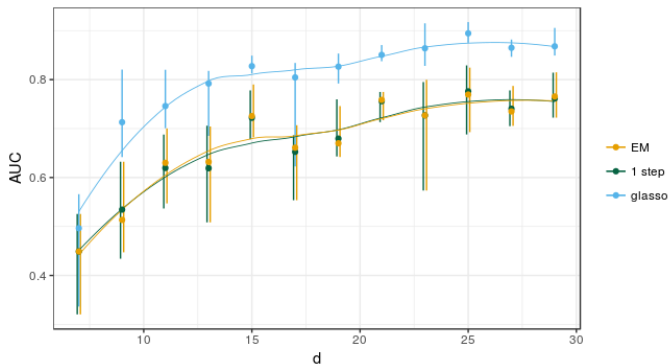
Graph of type tree: effect of d on AUC.
Cruves of medians, and 1st and 3rd quartiles.



- EM plus performant
- En particulier lorsque d est petit

Quand on s'éloigne de l'arbre...

Graph of type cluster: effect of d on AUC.
Cruves of medians, and 1st and 3rd quartiles.



Idée d'estimateur pour λ

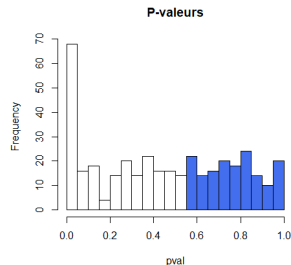
$$Y_i \sim Y_{\setminus -i}$$

- Ajuster chaque variable sur l'ensemble des autres variables
- Nouvelle interprétation des tests des coefficients en absence vs. présence d'arête : $(\mathcal{H}_0) : c_{ij} = 0$ vs $(\mathcal{H}_1) : c_{ij} \neq 0$
- Dédire des p-valeurs un estimateur du nombre de "non-arêtes"

Idée d'estimateur pour λ

$$Y_i \sim Y_{\setminus -i}$$

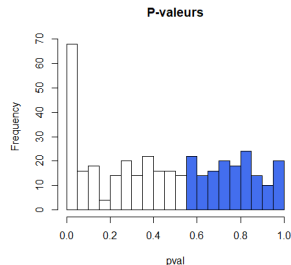
- Ajuster chaque variable sur l'ensemble des autres variables
- Nouvelle interprétation des tests des coefficients en absence vs. présence d'arête : $(\mathcal{H}_0) : c_{ij} = 0$ vs $(\mathcal{H}_1) : c_{ij} \neq 0$
- Dédire des p-valeurs un estimateur du nombre de "non-arêtes"



Idée d'estimateur pour λ

$$Y_i \sim Y_{\setminus -i}$$

- Ajuster chaque variable sur l'ensemble des autres variables
- Nouvelle interprétation des tests des coefficients en absence vs. présence d'arête : $(\mathcal{H}_0) : c_{ij} = 0$ vs $(\mathcal{H}_1) : c_{ij} \neq 0$
- Dédire des p-valeurs un estimateur du nombre de "non-arêtes"

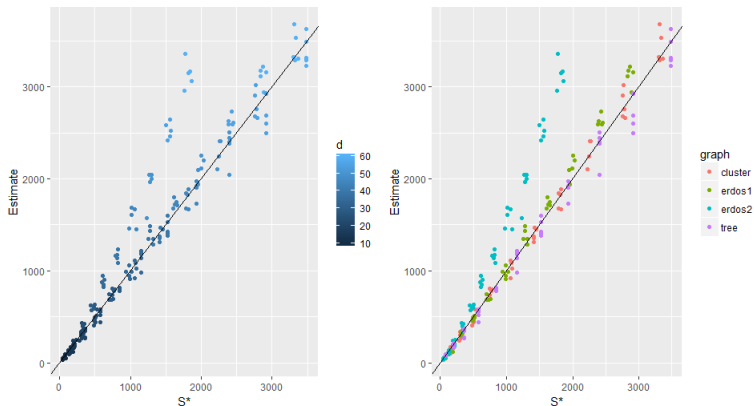


S et S^*

- $S^*(G)$: nombre d'absences d'arêtes du graphe G
- Estimateur de S^* :

$$S(Y_G) = 2 \sum \mathbb{1}\{pval(Y_G) \geq 0.5\}$$

Simulations de S



- Premiers résultats encourageants
- Semble se dégrader lorsque la probabilité d'arête augmente dans l'erdos

Estimateur de λ_{seuil}

- Seuil pour le glasso :

$$\hat{\lambda}_{seuil} = \min_{\lambda} \{|S - S_{\lambda}|\},$$

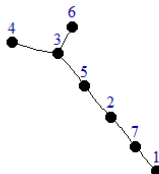
avec S_{λ} le nombre d'arêtes annulées pour le seuil λ .

- Les scores de l'EM sont des probabilités d'arêtes : le seuil est la densité estimée de G , soit :

$$\hat{\beta}_{seuil} = 1 - \frac{2S}{d(d-1)}$$

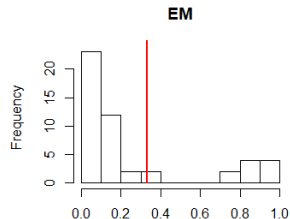
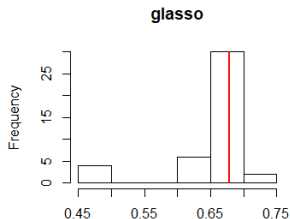
Exemple

Arbre à 7 noeuds :



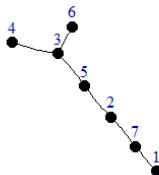
Résultats :

- $S^* = 15, S = 14$
- $\hat{\beta}_{seuil} = 0.33, \hat{\lambda}_{seuil} = 0.68$



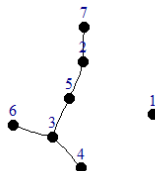
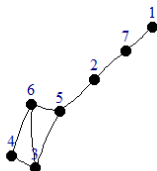
Exemple

Arbre à 7 noeuds :



Résultats :

- $S^* = 15, S = 14$
- $\hat{\beta}_{seuil} = 0.33, \hat{\lambda}_{seuil} = 0.68$



Bilan

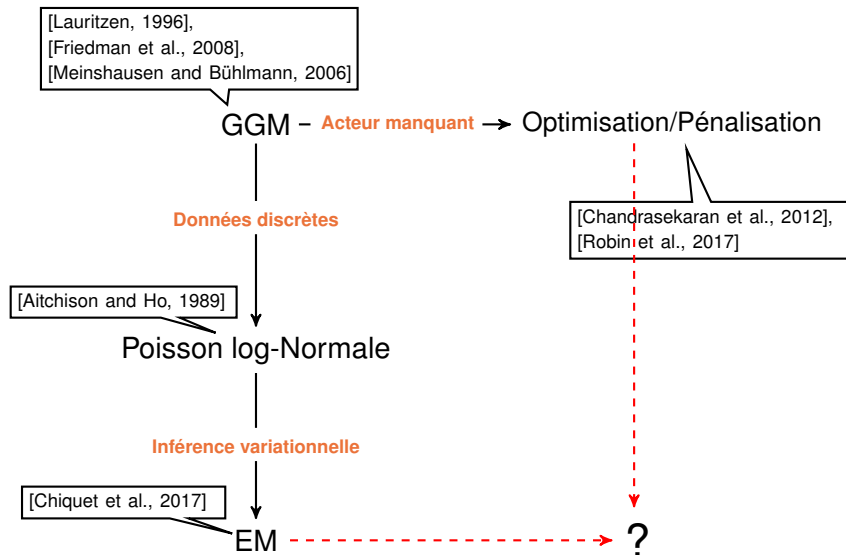
À partir d'un jeu de données :

- un a priori sur la densité du graphe,
- des seuils pour les matrices de scores.

La suite :

- Comment mesurer l'éloignement à la structure d'arbre ?

État de l'art



Avec des données de comptage et acteur manquant

Données : $(Y_{ij})_{i \in \{1, \dots, n\}, j \in \{1, \dots, p+q\}}$: i échantillons de p variables observées, on suppose q variables supplémentaires non observées.

Modèle :

La loi Poisson log-Normale

$$\left. \begin{array}{l} Z_i \text{ iid} \sim \mathcal{N}_{p+q}(\mu, \Sigma) \\ (Y_{ij})_j \perp\!\!\!\perp Z_i \\ Y_{ij}|Z_{ij} \sim \mathcal{P}(e^{Z_{ij}}) \end{array} \right\} Y \sim \mathcal{PLN}(\mu, \Sigma)$$

Méthode :

- Inclure GGM dans le modèle Poisson-log normal
- Inférence variationnelle car $p(Z|Y)$ n'est pas calculable
- Prendre en compte un acteur manquant

Autres développements envisagés :

- Prise en compte de covariables
- Adapter le modèle à des données recueillies au cours du temps

Merci pour votre attention



Aitchison, J. and Ho, C. (1989).
The multivariate Poisson-log normal distribution.
Biometrika, 76(4) :643–653.



Chaiken, S. and Kleitman, D. J. (1978).
Matrix tree theorems.
Journal of combinatorial theory, Series A, 24(3) :377–381.



Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. (2012).
Latent variable graphical model selection via convex optimization.
The Annals of Statistics, 40(4) :1935–1967.



Chiquet, J., Mariadassou, M., and Robin, S. (2017).
Variational inference for probabilistic Poisson PCA.
Technical report, arXiv :1703.06633.



Chow, C. and Liu, C. (1968).
Approximating discrete probability distributions with dependence trees.
IEEE Transactions on Information Theory, 14(3) :462–467.



Friedman, J., Hastie, T., and Tibshirani, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics, 9(3) :432–441.



Jakuschkin, B., Fievet, V., Schwaller, L., Fort, T., Robin, C., and Vacher, C. (2016).
Deciphering the pathobiome : Intra- and interkingdom interactions involving the
pathogen *erysiphe alphitoides*.
Microb Ecol, 72(4) :870–880.
doi :10.1007/s00248-016-0777-x.



Lauritzen, S. L. (1996).
Graphical Models.



Meila, M. and Jordan, M. I. (2000).
Learning with mixtures of trees.
Journal of Machine Learning Research, 1 :1–48.



Meinshausen, N. and Bühlmann, P. (2006).
High-dimensional graphs and variable selection with the lasso.
The annals of statistics, pages 1436–1462.



Robin, G., Robin, S., and Ambroise, C. (2017).
Graphical model inference with unobserved variables via latent tree aggregation.