

L'Analyse en Composantes Principales selon l'approche
bayésienne: All you need is a prior, Battacharya & Dunson's
prior

Eric Parent, Pierre Gloaguen

AppliBUGS, La Rochelle, 20 Juin 2024

Plan

- ▶ Le Joint Species Distribution Distribution Model, un LogNormalPoisson matriciel pour l'écologie
 - Données Bornéo de Giacomo Sellan,
 - Structure hiérarchique du JSJM d'Oskainen,
 - Examen des résultats de l'analyse des arbres de Bornéo,
 - Problèmes d'inférence.
- ▶ Comment réduire une matrice de variance-covariance creuse grâce au prior de Battacharya?
 - Rappels sur la décomposition d'un tableau en valeurs singulières,
 - Utilisation du SVD pour Analyse en Composantes Principales Probabiliste,
 - Construction du prior de Battacharya & Dunson.
- ▶ Conclusions, Discussion et perspectives.

Le dispositif forêt de Bornéo de Giacomo Sellan

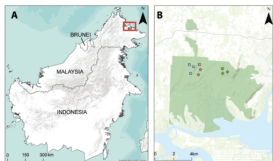
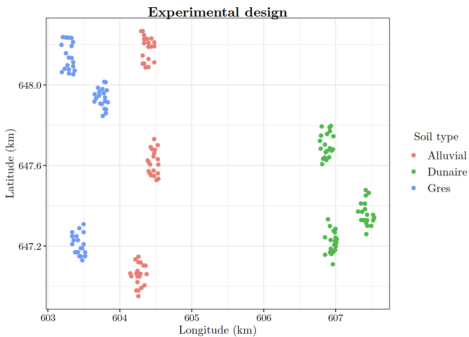


Figure 1 Study sites: (A) the island of Borneo with the approximate position of the Sabah-Sarawak Forest Reserve in Sabah, Malaysia highlighted in red; (B) the nine study plots within the KSRF (dark green); blue squares represent the sandstone plots, red squares represent the alluvial plots and green squares represent the heath forest plots

Figure 1: Lieu de récolte des données, d'après (Sellan 2021)



Le jeu de données Bornéo de Giacomo Sellan



Y: matrice de comptages
n=900 sites, p=500 espèces

	<i>Dehaasia_cassia</i>	<i>Polyalthia_canangioides</i>	<i>Dipterocarpus_acutangulus</i>
292-34_W	7	0	0
508-34_E	7	0	7
508-42_Y	7	0	0
508-64_D	7	0	0
292-21_E	6	0	0

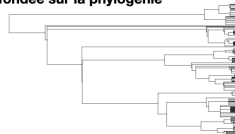
X: matrice de covariables
n=900 sites, k=23 effets potentiels

Site	Sh1	pH	Eno	An-P	C	N	NS3	NS4	Eno.de	Eno.A1	Eno.Ch	Eno.Pg	Eno.K	Eno.Sa	Eno.Culicosa	NS6Base
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1 292-11_A A12a	1.4	3.84	4.87	0.58	0.09	4.87	2.62	9.59	4.45	0.17	0.31	0.16	0.06	5.66	10.3	
2 292-11_B A12a	4.83	3.77	1.84	0.61	0.1	10.8	9.78	6.23	5.72	0.35	0.86	0.22	0.06	7.61	18.1	
3 292-11_C A12a	4.74	4.58	2.29	0.67	0.11	9.27	2.46	6.80	6.46	0.07	1.69	0.17	0.07	9.31	25.7	
4 292-11_D A12a	4.68	4.34	2.85	0.67	0.09	6.22	3.61	7.21	6.64	0.21	0.92	0.14	0.07	6.80	15.4	
5 292-11_E A12a	3.98	4.57	2.28	0.78	0.11	7.82	1.61	7.48	6.30	0.43	1.98	0.2	0.07	8.91	22.8	
6 292-11_F A12a	4.85	3.21	0.88	0.11	6.81	5.81	8.88	5.66	5.22	1.47	0.18	0.05	0.05	6.28	27.7	
7 292-11_G A12a	4.34	5.1	2.81	0.88	0.15	5.86	3.19	7.01	6.48	0.90	3.64	0.41	0.07	12.0	41.8	
8 292-12_A A12a	4.64	4.48	2.42	0.78	0.11	9.41	1.6	8.91	5.36	0.23	1.79	0.16	0.05	6.03	27.7	
9 292-12_B A12a	4.53	3.41	2.69	0.68	0.1	6.80	1.65	5.15	4.68	0.07	0.23	0.14	0.04	5.94	9.7	
10 292-12_C A12a	4.53	5.47	2.58	0.63	0.14	7.41	0.76	7.2	4.48	1.9	1.44	0.2	0.06	11.1	34.2	

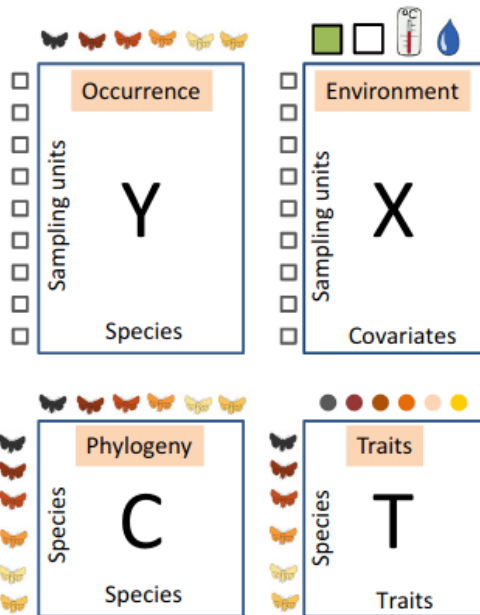
T: 3 traits phénotypiques pour chaque espèce

Espece	TsCroissance	Densite	Hauteur
<chr>	<chr>	<chr>	<chr>
1 <i>Styrobolus_borneensis</i>	0.00769	0.750	18.7
2 <i>Dysoxylum_indet</i>	0.0269	0.185	8.59
3 <i>Momoxylum_indet</i>	0.0127	0.783	8.49
4 <i>Cratogeomys_cochinchinensis</i>	0.0248	0.87	9.89
5 <i>Sterealia_stipitata</i>	0.0269	0.365	10.1
6 <i>Lepisanthes_indet</i>	0.0115	0.62	9.40

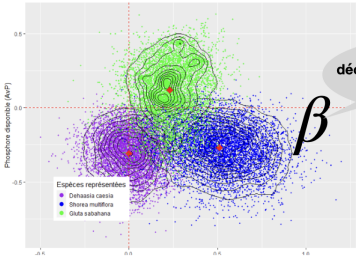
C: matrice de corrélation inter-espèces fondée sur la phylogénie



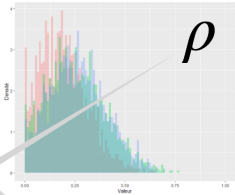
Joint species distribution modeling (Ingrédients)



Jsdm sur arbres de Borneo

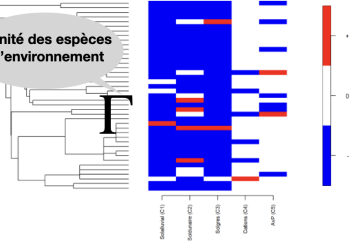


La loi a posteriori décrit les niches de 3 espèces avec leurs incertitudes

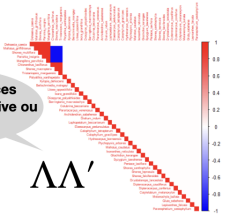


Un quart de la corrélation résiduelle semble héritée de la phylogénie

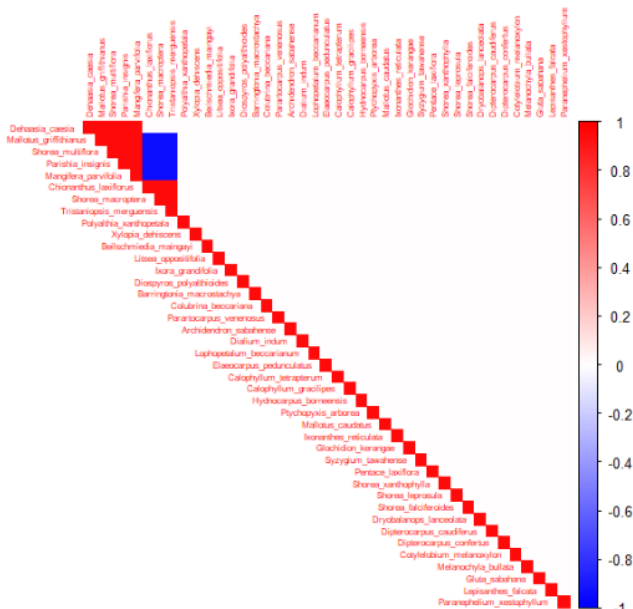
Affinité des espèces à l'environnement



Des blocs d'espèces en interactions positive ou répulsive



Mais comment obtien-t-on les co-occurrences résiduelles?



Comment faire l'inférence du JSMD?

- ▶ Réécriture avec k effets aléatoires :

$$\mathbf{L}_{n \times p} = \mathbf{X}\beta + \eta_{n \times k} \Lambda^{(p \times k)T} + \epsilon_{n \times p}$$

avec $\Sigma = \Lambda \mathbb{E}(\eta^T \eta) \Lambda^T + \mathbb{E}(\epsilon^T \epsilon) = \Lambda \Lambda^T + \text{diag}(\sigma_{j=1:p}^2)$

- ▶ A la fréquentiste: $\theta = \Sigma, \rho, \Gamma$, avec le package **PLN** chercher

$$\max_{\theta} \mathbb{P}_{\theta}(Y) - c_{\text{penalty}} \times |\Sigma|$$

- ▶ si besoin EM pour les v.a. latentes, puis $\Lambda \approx \sqrt{\Sigma}$
 - ▶ introduction de pénalisations sur la dimension k du modèle et sur la régularisation de la covariance.
- ▶ Selon l'approche bayésienne : focus sur le **multiplicative gamma process shrinkage prior** de *Bhattacharya* pour Σ .



Décomposition en valeurs singulières (SVD)

- ▶ On veut étudier la structure d'une base de données massives, un tableau \mathbf{M} (n individus $\times p$ variables). Par exemple, en écologie, les individus seraient des unités de recueil de données $i = 1 : n$ et les variables les espèces d'intérêt $j = 1 : p$ dont on mesure une caractéristique $M_{i,j}$.

- ▶ Théorème:

$$\mathbf{M} = \mathbf{U} \mathbf{D} \mathbf{V}'$$

$n \times p$ $(n \times n)(n \times p)(p \times p)$

- \mathbf{U} matrice unitaire (rotation) $n \times n$,
- \mathbf{D} une matrice $n \times p$ dont les coefficients diagonaux sont des réels positifs ou nuls
- \mathbf{V}' transposée de \mathbf{V} , matrice unitaire $p \times p$.

Décomposition en valeurs singulières (SVD)

► Principe de la démo:

- Trouver un premier jeu de deux vecteurs u et v tels que
 $1 \times n$

$$\max u' M v - \frac{\sigma_u}{2} u' u - \frac{\sigma_v}{2} v' v$$

La maximisation donne $M' u = \sigma_v v$ et $M v = \sigma_u u$ soit

$$M M' u = \sigma_u \sigma_v u$$

$$M' M v = \sigma_u \sigma_v v$$

u est vecteur propre de $M M'$, v est vecteur propre de $M M'$, on montre aussi $\sigma_u = \sigma_v = \sigma$.

- Comme pour l'ACP, on cherche ensuite deux vecteurs unitaires orthogonaux aux premiers obtenus maximisant $u' M v$, etc., et on range les valeurs propres par ordre de magnitude décroissante. On construit ainsi les "rotations" U et V d'où

$$M = \underset{n \times p}{U} \underset{(n \times n)}{D} \underset{(n \times p)}{V} \underset{(p \times p)}{'}$$

Interprétation algébrique de la SVD

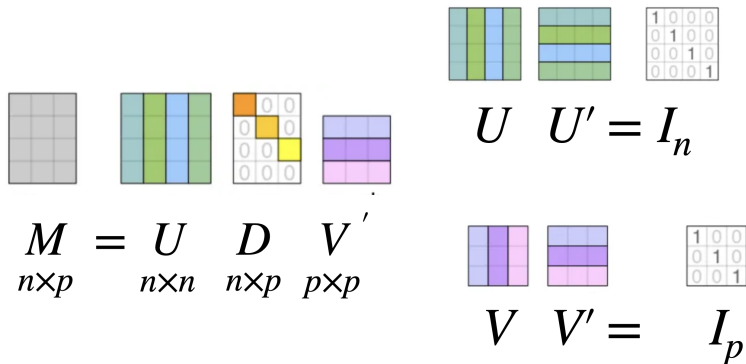
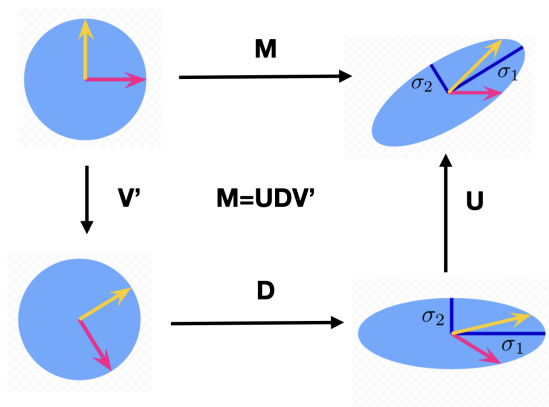


Figure 2: Interprétation algébrique

Interprétation géométrique de la SVD



Pour toute application linéaire de $\mathbb{R}^p \rightarrow \mathbb{R}^n$, on peut trouver une base orthonormale pour \mathbb{R}^p et une base orthonormale pour \mathbb{R}^n telles que l'on associe au i -ème vecteur de base de \mathbb{R}^p un multiple positif du i -ème vecteur de base de \mathbb{R}^n , les vecteurs restants ayant pour image 0. Dans ces bases, l'application est donc ainsi représentée par une matrice diagonale dont les coefficients sont des réels positifs.

SVD : un exemple par simulation

```
p <- 20; k <- 2; n <- 15
Etroit <- matrix(runif(n*k), nrow = n, ncol = k)
Long <- matrix(runif(p*k), nrow = p, ncol = k)
residus = matrix(rnorm(n*p,0,0.01), nrow = n, ncol = p)
M <- Etroit%*%t(Long)+residus; svdM = svd(M); str(svdM)
```

```
List of 3
 $ d: num [1:15] 9.1345 1.6582 0.0719 0.069 0.0608 ...
 $ u: num [1:15, 1:15] -0.154 -0.005 -0.335 -0.416 -0.309 ...
 $ v: num [1:20, 1:15] -0.146 -0.119 -0.128 -0.135 -0.249 ...
```

```
print(sum(abs(M-svdM$u%*%diag(svdM$d)%*%t(svdM$v))^2))
```

```
[1] 5.129562e-29
```

```
diag(t(svdM$u)%*%svdM$u)
```

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
diag(t(svdM$v)%*%svdM$v)
```

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
#print((eigen(svdM$v%*%t(svdM$v), only.values = TRUE))$values[1:min(30,
```

SVD : l' exemple d'une image pixellisée

Analysons la structure du dossier image *mystère* `img1`

```
library(FactoMineR)
dim(img1)
```

```
[1] 500 500 4
```

```
res.pca <- PCA(img1[, ,1], graph=FALSE)
head(res.pca$eig[,1])
```

```
comp 1    comp 2    comp 3    comp 4    comp 5    comp 6
181.11540  84.00610  72.64358  54.79642  18.71580  15.71998
```

```
head(exp(diff(log(res.pca$eig[,1]))),15)
```

```
comp 2    comp 3    comp 4    comp 5    comp 6    comp 7    comp 8    comp 9    comp 10    comp 11    comp 12    comp 13
0.4638264 0.8647417 0.7543189 0.3415515 0.8399311 0.7554233 0.7575143 0.7701226 0.6576041 0.8502850 0.7387635 0.9752632
comp 14    comp 15    comp 16
0.9021925 0.8553352 0.8959694
```

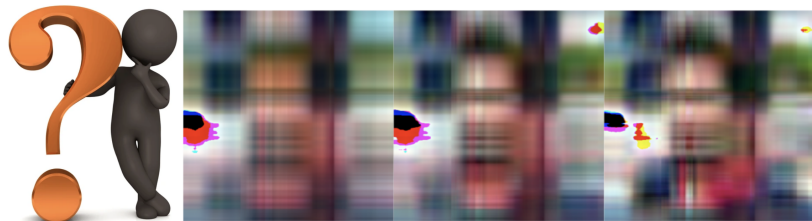
SVD : l' exemple d'une image pixellisée

```
barplot((res.pca$eig[,1]), main = "Eigenvalues",  
        names.arg = paste("Dim", 1:nrow(res.pca$eig), sep = ""))
```

Eigenvalues



Reconstitution de la Compression SVD sur 2,3 et 4 axes principaux



L'image originelle fait 500×500 pixels.

On a gardé seulement 2, 3 ou 4 dimensions principales

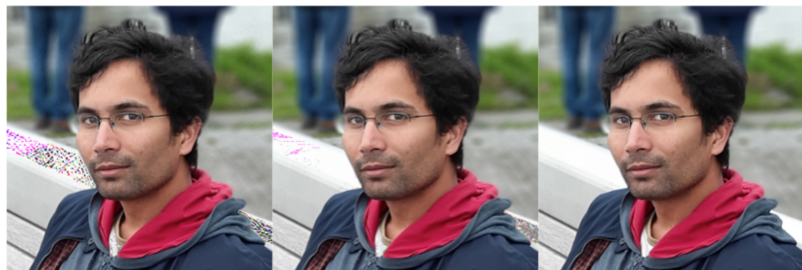
Compression SVD 15,20 et 50 axes principaux



L'image originelle faisait 500×500 pixels.

On a gardé 15,20 et 50 dimensions principales

Compression SVD 100,250 et 500 axes principaux



L'image originelle faisait 500×500 pixels.

On a gardé 100,250 dimensions principales et l'image originelle.

“Stochastisation” de la SVD

$$\eta_{i,\dots}^u \sim \mathcal{N}_k \left(\mathbf{0}, \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{pmatrix} \right)$$



$$\mathbb{E}(\eta^u \eta'^u) = kI_k$$



$$\mathbb{E}(M' M)$$

$$\mathbb{E}(\Lambda \Lambda')$$

$$\Lambda' = \begin{matrix} \sqrt{k} & * & \text{blue} & \text{blue} & \text{blue} \\ \sqrt{k} & * & \text{purple} & \text{purple} & \text{purple} \\ \sqrt{k} & * & \text{pink} & \text{pink} & \text{pink} \end{matrix}$$

Figure 3: Stochastisation

Donner un rang faible à la matrice de variance

$$Y_i \approx X_i \beta + \begin{bmatrix} u_i^1 & u_i^2 & \dots & u_i^n \end{bmatrix} \begin{bmatrix} d_1^2 & 0 & \dots & 0 & 0 \\ 0 & \ddots & \vdots & 0 & 0 \\ \vdots & 0 & d_k^2 & 0 & 0 \\ 0 & 0 & \dots & d_n^2 & 0 \end{bmatrix} \begin{bmatrix} v_1^1 & v_1^2 & \dots & v_1^p \\ \vdots & \vdots & \vdots & \vdots \\ v_p^1 & v_p^2 & \dots & v_p^p \end{bmatrix}$$

v. essentielles
1 k
SVD et mise à 0 au delà de k

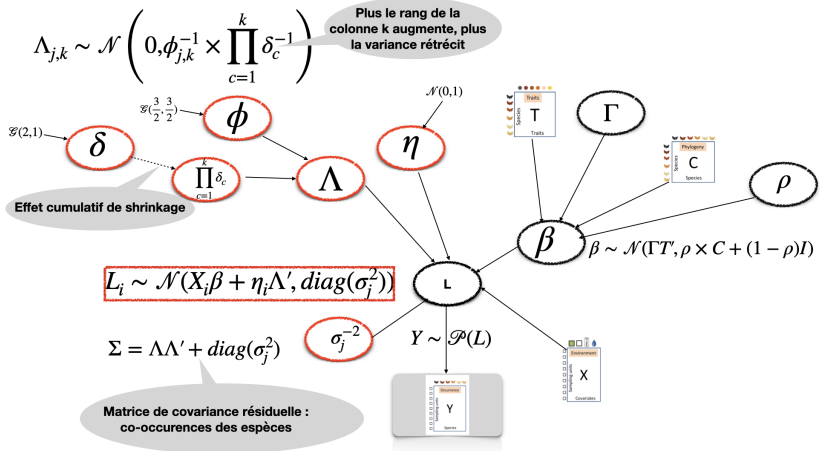
$$Y_i = X_i \beta + \begin{bmatrix} \eta_i^1 & \eta_i^2 & \dots & \eta_i^k \end{bmatrix} \begin{bmatrix} d_1^2 \times (v_1^1 & v_1^2 & \dots & v_1^p) \\ \vdots & \vdots & \vdots & \vdots \\ d_k^2 \times (v_k^1 & v_k^2 & \dots & v_k^p) \end{bmatrix} + \epsilon_i$$

$\sim \mathcal{N}(0,1)$

$$\epsilon_i \sim \mathcal{N}_p \left(\mathbf{0}, \begin{pmatrix} \left(\begin{matrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_p^2 \end{matrix} \right) \right)$$

Λ^T

Un DAG pour l'ACP Probabiliste



Priors de Bhattacharya (en écriture maths)

- ▶ **Variance des résidus** sur $\text{diag}(\sigma_j^2)$, $1 \leq j \leq p$ Inverse Gamma: **Standard**
- ▶ **priors des latentes** sur η_j , $1 \leq j \leq p$: $\mathcal{N}_p(0, I_p)$ **Standard**
- ▶ **priors des composantes** on Λ : *multiplicative gamma process shrinkage prior* avec l'idée de *pénaliser* les dernières colonnes afin de les emmener au fur et à mesure vers un poids nul;
- ▶ Pour $1 \leq j \leq p$ et $1 \leq h \leq q$, $\phi_{j,h} \stackrel{\text{ind}}{\sim} \text{Gamma}\left(\frac{3}{2}, \frac{3}{2}\right)$.
- ▶ Pour $1 \leq h \leq q$, $\delta_h \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha, 1)$ tel que $\alpha > 1$ (donc $\mathbb{E}[\delta_h] > 1$); par ex $\alpha = 2$
- ▶ Puis le prior sur Λ :

$$\Lambda_{j,h} | \phi_{j,h}, \delta_{1:h} \stackrel{\text{ind}}{\sim} \mathcal{N}\left(0, \phi_{j,h}^{-1} \prod_{\ell=1}^h \delta_{\ell}^{-1}\right).$$

- ▶ When h increases (for last columns), the variance tends to collapse to 0.
- ▶ Reste le prior de dernier niveau sur α si on ne le fixe pas : non informatif, de support $\alpha \geq 1$.

Echantillonnage du posterior

Gibbs sampling

- ▶ La loi a posteriori non explicite;
- ▶ On a recours à l'échantillonneur de Gibbs sampling, facile car les distributions conditionnelles ressortent de la conjugaison Normal-Gamma;
- ▶ Toutes sauf si on a laissé le paramètre α libre \Rightarrow Metropolis-Hastings;

Inférence variationnelle

- ▶ Beaucoup d'inconnues quand n, p grands! par ex: (900 sites \times 500 espèces)! On cherche alors à approximer le posterior par un produit de distributions marginales issues d'une famille bien choisie (evidemment Normal-Gamma);
- ▶ Si on utilise la divergence de Kullback Leibler pour mesurer les écarts de distribution, les indépendances conditionnelles font apparaitre une somme de log-densités à optimiser.
- ▶ L'inférence variationnelle peut être résolue terme à terme dans le cas conjugué et de façon itérative, analogue au Gibbs sampling;
- ▶ Et l'algorithme d'optimisation itère entre les caractéristiques à obtenir selon une descente de gradient explicite (Coordinate ascent variational inference);

Conclusions

- ▶ Les propriétés de décomposition en valeurs singulières d'un tableau permettent d'interpréter l'algorithme du **multiplicative gamma process shrinkage prior** de *Bhattacharya* et *Dunson* pour l'ACP probabiliste.
- ▶ Le Log Normal Poisson avec ACP latent est un outil puissant pour modéliser les co-occurrences en écologie;
- ▶ Les priors standards du "baby-bayes" restent tout à fait efficaces pour l'inférence;

Discussions

- ▶ Quid de l'orthodoxie Bayésienne?
 - Est-ce vraiment une approche bayésienne? Pas d'élicitation, Pas de fonction de coût explicite, pas de décision formalisée
 - utiliser le prior de *Bhattacharya & Dunson's* s'emboîte à la perfection dans le Lego land de l'inférence bayésienne de la modélisation hiérarchique et de l'inférence de type Gibbs.
 - Ce module révèle automatiquement la dimension du modèle!

Perspectives: les promesses de l'approche

- ▶ Passage en variable latente pour des GLMs type $Y = \mathcal{P}(e^Z)$; employé aussi dans le cadre classique PLN de Chiquet, Robin et Mariadassou (voir leur package). Extensions au zero-inflaté attendues.
- ▶ Passage à l'échelle n ou p grands: recours au *Variational* Bayes pour une inférence a posteriori type champ moyen. (cf Hoffman et Blei, 2013). On continue à tirer parti du caractère modulaire et conjugué. Le gain computationnel suivant demande une optimisation fonctionnelle grâce aux réseaux de neurones (cf Kingma et Welling, 2014).
- ▶ Codes (à vos risques et périls!) see https://github.com/papayoun/VI_fo_Poisson_PCA

Ouf!

Merci pour votre intérêt!



PS:Pas de jaloux. . .

References

- Bhattacharya, Anirban, and David B Dunson. 2011. "Sparse Bayesian Infinite Factor Models." *Biometrika*, 291–306.
- Chiquet, Julien, Mahendra Mariadassou, and Stéphane Robin. 2018. "Variational Inference for Probabilistic Poisson PCA." *The Annals of Applied Statistics* 12 (4): 2674–98.
- Ovaskainen, Otso, and Nerea Abrego. 2020. *Joint Species Distribution Modelling: With Applications in R*. Cambridge University Press.
- Ovaskainen, Otso, Gleb Tikhonov, Anna Norberg, F Guillaume Blanchet, Leo Duan, David Dunson, Tomas Roslin, and Nerea Abrego. 2017. "How to Make More Out of Community Data? A Conceptual Framework and Its Implementation as Models and Software." *Ecology Letters* 20 (5): 561–76.

Ecriture du modèle en Jags : juste quelques lignes

```
modelString = "  
model{  
a[1] <- 3; a[2] <- 3; deminu <- 3/2; a_sigma <- 3; b_sigma <- 2;  
  
#shrinkage prior  
          delta[1] ~ dgamma(a[1],1)  
for( h in 2:q){delta[h] ~ dgamma(a[2],1)}  
for( h in 1:q){ tau[h] <- prod(delta[1:h])}  
for( j in 1:p){ for( h in 1:q){ phi[j,h] ~ dgamma(deminu,deminu)  
          precilambda[j,h] <- phi[j,h]*tau[h]  
          lambda[j,h]~ dnorm(0,precilambda[j,h])}}}  
for(i in 1:n){for (h in 1:q){eta[i,h] ~ dnorm(0,1)}}  
muY <- eta %*% t(lambda)  
for(j in 1:p){ preciE[j] ~ dgamma(a_sigma,b_sigma)  
          for (i in 1:n){Y[i,j] ~ dnorm(muY[i,j],preciE[j])}}}  
}"
```

Lancer le code Jags

```
library(rjags); library(ggmcmc); Y=M
data_for_JAGS <- list( Y=Y, n = dim(Y)[1], p = dim(Y)[2], q = 5)
n.adapt = 300; burnin = 300; n.iter = burnin * 3; thin = 1
jm <- jags.model(file = textConnection(modelString),
  data = data_for_JAGS,
  n.chains = 3 ,
  n.adapt = n.adapt)
```

```
Compiling model graph
  Resolving undeclared variables
  Allocating nodes
Graph information:
  Observed stochastic nodes: 300
  Unobserved stochastic nodes: 300
  Total graph size: 1021

Initializing model
```

```
update(jm, burnin)
jsamples <- coda.samples(model = jm,
  variable.names = c("lambda", "preciE", "eta", "delta", "phi", "tau"),
  n.iter = n.iter,
  thin = thin)

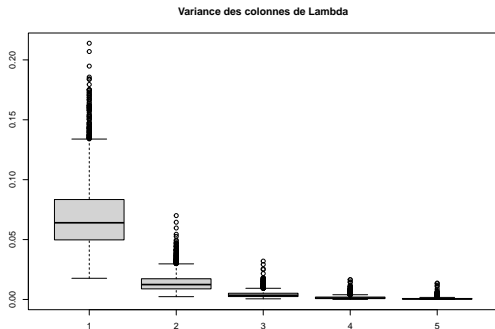
result_jags <- ggs(jsamples)
```

Extraire les variables d'intérêt

```
library(tidyverse)
n_samples <- n.iter*3
Lambdas_jags <- filter(result_jags, str_detect(Parameter, "lambda")) %>%
  group_by(Chain, Iteration) %>%
  group_map(function(x, g){
    pull(x, value) %>%
      matrix(nrow = ncol(Y), byrow = TRUE)
  }) %>%
  {.[1:n_samples]}
Lambdas_jags_var <- Lambdas_jags %>%
  sapply(function(x) apply(x, 2, var)) %>%
  t()
```

Représenter les sorties

```
Lambdas_jags_var %>% boxplot()  
title(main = paste ("Variance des colonnes de", expression(Lambda)))
```



Représenter les sorties

```
filter(result_jags, str_detect(Parameter, "tau")) %>%  
  group_by(Chain, Iteration) %>%  
  group_map(function(x, g){pull(x, value) }) %>%  
  {.[1:n_samples]} %>% sapply(function(x) 1/x) %>% t() %>% boxplot()  
title(main = expression(tau^-2))
```

