CENTRALE NANTES

Ecole Centrale de Nantes

Projet methodes bayésiennes et modèles hiérarchiques

Seeds: Random effect logistic regression

Auteurs:
Paul Trincklin
Jiacheng Fu
Enzo Mori
Grégory Perret

Encadrant:
Mathieu Ribatet

March 24, 2022

1 Problématique

L'objectif de ce projet est de modéliser la capacité de germination de deux types de graines (O. aegyptiaco 75 et O. aegyptiaco 73) pour deux variétés différentes (haricot et concombre).

Pour cela on étudie 21 assiettes dans lesquelles ont été planté ces différents types de graines.

On note n_i le nombre total de graines plantées dans la ième assiette, r_i le nombre de graines ayant germées dans la ième assiette, et p_i la probabilité de germination d'une graine dans la ième assiette.

Nous disposons du jeu de données suivant :

INSERER TABLEAU JEUX DE DONNEES

On utilise le modèle suivant :

$r_i \sim Bin(p_i, n_i)$

avec $logit(p_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i} + bi$

et $bi \sim N(0, \tau)$ où $\tau = \frac{1}{\sigma^2}$

x_{1i} indique le type de graine (0 pour O. aegyptiaco 75 et 1 pour O. aegyptiaco 73) et x_{2i} indique la variété (0 pour haricot, 1 pour concombre) dans la ième assiette.

2 Modélisation mathématique

Les lois a priori de nos différents paramètres sont données: $\alpha_0 \sim N(0, 10^6), \alpha_1 \sim N(0, 10^6), \alpha_2 \sim N(0, 10^6), \alpha_{12} \sim N(0, 10^6), \sigma^2 \sim InverseGamma(10^{-3}, 10^{-3})$

On calcule les lois à posteriori pour les différents paramètres pour notre algorithme MCMC associé à un échantilloneur de Gibbs:

(remarque: ici on ne passe pas à l'échelle log, mais ceci est fait dans notre code afin d'améliorer les calculs)

$$\begin{aligned}
\pi(b_i \mid ...) &\propto \pi(b_i \mid \sigma^2) \times \pi(r_i \mid ...) \\
&\propto e^{-\frac{b_i^2}{2\sigma^2}} \times (p_i)^{r_i} \times (1 - p_i)^{n_i - r_i}
\end{aligned}$$

$$\begin{aligned}
\pi(\alpha_0 \mid ...) &\propto \pi(\alpha_0) \times \prod_{i=1}^{N} \pi(r_i \mid ...) \\
&\propto e^{-\frac{\alpha_0^2}{2\times 10^6}} \times \prod_{i=1}^{N} [(p_i)^{r_i} \times (1 - p_i)^{n_i - r_i}]
\end{aligned}$$

$$\begin{aligned}
\pi(\alpha_1 \mid ...) &\propto \pi(\alpha_1) \times \prod_{i=1}^{N} \pi(r_i \mid ...) \\
&\propto e^{-\frac{\alpha_1^2}{2\times 10^6}} \times \prod_{i=1}^{N} [(p_i)^{r_i} \times (1 - p_i)^{n_i - r_i}]
\end{aligned}$$

$$
\begin{aligned}
\pi(\alpha_2 \mid ...) &\propto \pi(\alpha_2) \times \prod_{i=1}^{N} \pi(r_i \mid ...) \\
&\propto e^{-\frac{\alpha_2^2}{2\times 10^6}} \times \prod_{i=1}^{N} [(p_i)^{r_i} \times (1-p_i)^{n_i - r_i}]
\end{aligned}
$$

$$
\begin{aligned}
\pi(\alpha_{12} \mid ...) &\propto \pi(\alpha_{12}) \times \prod_{i=1}^{N} \pi(r_i \mid ...) \\
&\propto e^{-\frac{\alpha_{12}^2}{2\times 10^6}} \times \prod_{i=1}^{N} [(p_i)^{r_i} \times (1-p_i)^{n_i - r_i}]
\end{aligned}
$$

$$
\begin{aligned}
\pi(\sigma^2 \mid ...) &\propto \pi(\sigma^2) \times \prod_{i=1}^{N} \pi(b_i \mid \sigma^2) \\
&\propto (\sigma^2)^{(-0.001-1)} e^{-\frac{0.001}{\sigma^2}} \times (\frac{1}{\sigma})^N e^{-\frac{\sum_{i=1}^N b_i^2}{2\sigma^2}} \\
&\propto InverseGamma(0.001 + \frac{N}{2}, 0.001 + \frac{\sum_{i=1}^N b_i^2}{2})
\end{aligned}
$$

où $p_i = sigmod(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i} + bi)$

3 Résultat obtenus et analyse

Vérifions l'allure des chaines de Markov obtenues:

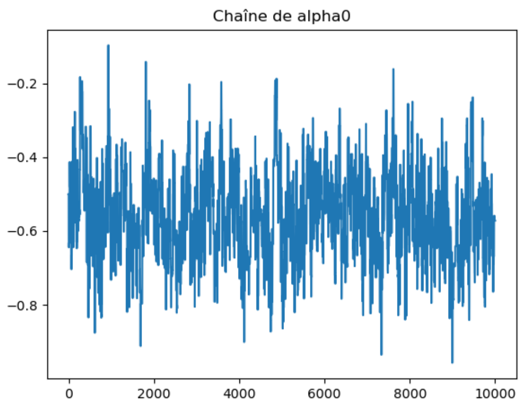

Figure 1: Chaine de Markov de α_0

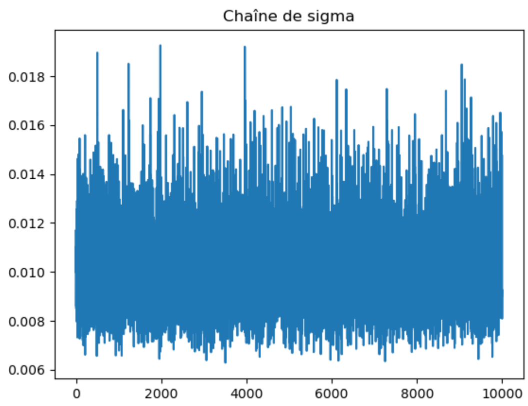

Figure 2: Chaine de Markov de σ

Nous obtenons les résultats suivants:

Résultats					
	α_0	α_1	α_2	$\alpha_1 2$	σ
Moyenne	-0,558	0,144	1,327	-0,789	0,01
Variance	0,017	0,049	0,032	0,09	$2,7.10^{-6}$

Nous obtenons des résultats similaires à ceux attendus, excepté pour α_1 et σ.

Interpretons les différents coefficients trouvé afin de voir l'influence de chaque graine sur la probabilité de germer.

En accord avec le modèle choisit, la probabilité de germer est donnée ainsi pour chaque type de graine et chaque variétée:

-graine de type O. aegyptiaco 75 , haricot : $p = sigmoide(\alpha_0) = sigmoide(-0,558) = 0,361$

-graine de type O. aegyptiaco 75 , concombre : $p = sigmoide(\alpha_0 + \alpha_2) = sigmoide(-0,558 + 1,327) = sigmoide(0,769) = 0,682$

-graine de type O. aegyptiaco 73 , haricot : $p = sigmoide(\alpha_0+\alpha_1) = sigmoide(-0,558+ 0,144) = sigmoide(-0,414) = 0,399$

-graine de type O. aegyptiaco 73 , concombre : $p = sigmoide(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_{1,2}) = sigmoide(-0,558 + 0,144 + 1,327 - 0,789) = sigmoide(0,124) = 0,532$

Probabilité de germination		
	seed O. aegyptiaco 75	O. aegyptiaco 75
Haricot	0,361	0,399
Concombre	0,682	0,532

On voit ainsi que pour les concombres, il est préférable de choisir des graines de type O. aegyptiaco 75 puisque la probabilité de germer est supérieure.

Au contraire, pour les haricots, il vaut mieux planter des graines de type O. aegyptiaco 73 qui présente une plus grande probabilité de germination.

Pour les deux types de graines, les concombres germent avec une plus grande probabilité que les haricots.