

UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE NANTES

Mars 2023

Etude des germinations par une régression logistique à effet aléatoire

MINETTE Cloée, BAKHOUCHE Sarah,
SIMON Antoine, DAVID Romain

Table des matières

1	Introduction	2
2	Modèle	2
2.1	Choix du modèle	2
2.2	Construction du modèle	3
3	Analyse et interprétation des résultats	4

1 Introduction

Le jeu de données SEEDS a été étudié par Crowder en 1978. Le but est de modéliser la germination de deux types de graines, *O. aegyptiaco 75* et *O. aegyptiaco 73* pour deux variétés différentes, concombre et haricot. Nous regardons la germination de ces graines pour 21 plaques.

<i>O. aegyptiaco 75</i>						<i>O. aegyptiaco 73</i>					
Haricot			Concombre			Haricot			Concombre		
r	n	$\frac{r}{n}$	r	n	$\frac{r}{n}$	r	n	$\frac{r}{n}$	r	n	$\frac{r}{n}$
10	39	0.26	5	6	0.83	8	16	0.5	3	12	0.25
23	62	0.37	53	74	0.72	10	30	0.33	22	41	0.54
23	81	0.28	55	72	0.76	8	28	0.29	15	30	0.5
26	51	0.51	31	51	0.63	23	45	0.51	32	51	0.63
17	39	0.44	46	79	0.58	0	4	0	3	7	0.43
			10	13	0.77						

TABLE 1 – Représentation des données

2 Modèle

2.1 Choix du modèle

Notre jeu de données compte le nombre de graines germées sur chaque plaque. Nous sommes donc soumis à un espace discret. Ces données de dénombrement sont souvent influencée par une variation des données plus grande que prévue, c'est la surdispersion. Pour cette étude, il est suggérable d'utiliser un modèle bayésien Binomial hiérarchique à effet aléatoire afin de mieux capturer la surdispersion. La loi Binomiale donne le nombre de succès d'une expérience, soit le nombre de graines germées par rapport au nombre de graines plantées, elle est donc adaptée à notre modèle. De plus, nous travaillons avec un modèle à effet aléatoire car les données sont structurées sur plusieurs niveaux. C'est à dire que des individus partagent un environnement commun qui peut affecter le comportement étudié, par exemple x_{1i} et x_{2i} .

La loi de p_i , la probabilité de germination sur la plaque i , est modélisée par une régression logistique permettant d'étudier les relations les variables. On y ajoute la variable b_i correspondant à une erreur de modélisation suivant une loi Normale de paramètre $(0, \tau)$, avec τ suivant une loi Inverse Gamma(a, b). Nous utilisons la loi Inverse Gamma car elle décroît très vite et est utilisée pour modéliser une grande variété de phénomènes, et tout particulièrement les phénomènes se déroulant au cours du temps.

On décide d'appliquer ce modèle au jeu de données SEEDS. Il est exprimé comme suit, avec r_i le nombre de graines germées sur la i -ème plaque, n_i le nombre total de graines sur la i -ème plaque, p_i la probabilité de germination sur la i -ème plaque. On dispose également de x_{1i} le type

de semence à valeur binaire, 0 correspond à *O. aegyptiaco 75* et 1 correspond à *O. aegyptiaco 75*, et x_{2i} la variété à valeur binaire, 0 correspond au haricot et 1 au concombre. Nous ajoutons à celà le terme d'interaction $\alpha_{12}x_{1i}x_{2i}$.

Pour $i \in \{1; 21\}$, le modèle est tel que :

$$r_i \sim \text{Binomiale}(p_i, n_i)$$

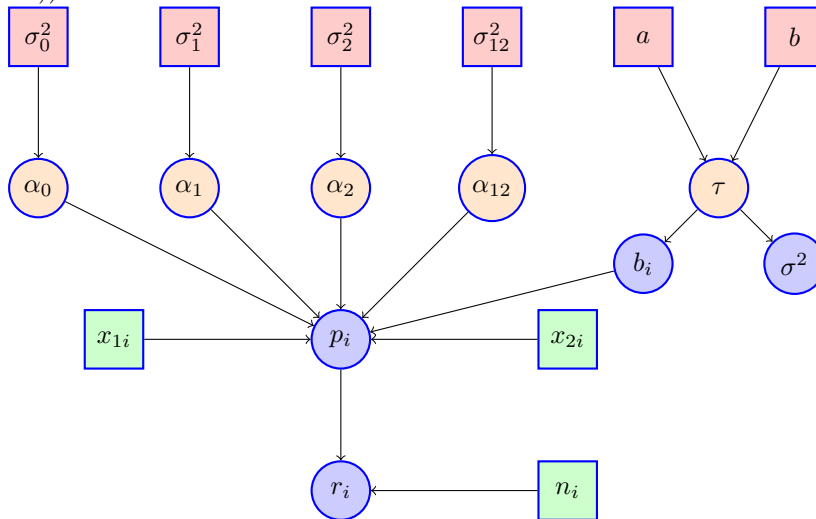
$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i} + b_i$$

$$b_i \sim \text{Normale}(0, \tau)$$

Les paramètres $\alpha_0, \alpha_1, \alpha_2, \alpha_{12}$ et τ sont indépendants et ont des lois a priori non informatives. Nous utiliserons une méthode d'échantillonnage de Gibbs pour calculer les lois de nos estimateurs.

2.2 Construction du modèle

On peut représenter notre modèle par un graphe orienté acyclique (Directed Acyclic Graphe (DAG)) :



Notons tout d'abord que :

$$\alpha_0 \sim \mathcal{N}(0, \sigma_0^2)$$

$$\alpha_1 \sim \mathcal{N}(0, \sigma_1^2)$$

$$\alpha_2 \sim \mathcal{N}(0, \sigma_2^2)$$

$$\alpha_{12} \sim \mathcal{N}(0, \sigma_{12}^2)$$

Nous obtenons les lois a posteriori suivantes :

$$\Pi(r_i|\dots) \propto \Pi(r_i|p_i, n_i) \propto (p_i)^{r_i} \times (1 - p_i)^{n_i - r_i}$$

$$\Pi(\alpha_0|\dots) \propto e^{-\frac{\alpha_0^2}{2\sigma_0^2}} \prod_i ((p_i)^{r_i} \times (1 - p_i)^{n_i - r_i})$$

$$\Pi(b_i|\dots) \propto \Pi(b_i|\sigma^2) \times \Pi(r_i|\dots) \propto e^{-\frac{b_i^2}{2\tau^2}} \prod_i ((p_i)^{r_i} \times (1 - p_i)^{n_i - r_i}) \propto e^{-\frac{b_i^2 \sigma^2}{2}} \prod_i ((p_i)^{r_i} \times (1 - p_i)^{n_i - r_i})$$

$$\Pi(\sigma^2|\dots) \propto \Pi(\sigma^2) \prod_i \Pi(b_i|\sigma^2) \propto$$

Pour α_1 , α_2 et α_{12} , les lois conditionnelles s'écrivent de la même manière que la loi conditionnelle de α_0 avec respectivement σ_1^2 , σ_2^2 et σ_{12}^2 comme variance.

3 Analyse et interprétation des résultats