



ÉCOLE CENTRALE DE NANTES

BIRATS: A BIVARIATE NORMAL HIERARCHICAL MODEL

Projet Bayes

Élèves :

Meryem CHERQI
Lamia ARIFALLAH
Rim YOUSFI

Encadrants :

Mathieu RIBATET

17 avril 2024

Table des matières

1	Introduction	2
2	Présentation des Données	2
3	Modèle Mathématique	2
3.1	Structure du modèle	3
3.2	Graphe Orienté Acyclique (DAG)	3
3.3	Formules Conditionnelles	3
4	Résultats	4
5	Conclusion	5

Lien vers le code : Birats-code

1 Introduction

Ce projet analyse la croissance des rats en utilisant un modèle hiérarchique bivarié normal. Nous employons des distributions normales multivariées pour modéliser les coefficients de régression de la courbe de croissance, prenant en compte les corrélations entre l'interception et la pente. Cette étude permet d'explorer comment les caractéristiques initiales, telles que le poids, influencent la croissance ultérieure des rats, apportant des insights importants pour les études biomédicales et la gestion de la santé animale.

2 Présentation des Données

Les données utilisées dans cette étude proviennent d'une série d'expériences longitudinales sur des rats, où le poids de chaque rat a été mesuré à différents âges. Les mesures ont été effectuées à intervalles réguliers au cours du développement des rats, depuis leur naissance jusqu'à leur maturité.

Les données comprennent plusieurs variables clés :

- **Identifiant du Rat** : Un identifiant unique pour chaque rat.
- **Poids** (grammes) : Le poids du rat mesuré à chaque point temporel.
- **Âge** (jours) : L'âge du rat au moment de chaque mesure.
- **Sexe** : Le sexe du rat, car il peut influencer le taux de croissance.

3 Modèle Mathématique

Le modèle est le suivant :

$$Y_{ij} \sim \text{Normale}(m_{ij}, \tau_c)$$

$$m_{ij} = b_{1i} + b_{2i}x_j$$

$$\mathbf{b}_i \sim \text{MVN}(\mathbf{m}_b, \mathbf{W})$$

où Y_{ij} est le poids du i ème rat mesuré à l'âge x_j , et \mathbf{b}_i désigne le vecteur (b_{1i}, b_{2i}) . Nous supposons des priors Normaux univariés indépendants 'non-informatifs' pour les composantes séparées m_{b1} et m_{b2} . Un prior Wishart $\text{Wishart}(R, r)$ a été spécifié pour \mathbf{W} , la matrice de précision de population des coefficients de régression. Pour représenter une connaissance a priori vague, nous avons choisi les degrés de liberté r de cette distribution aussi petits que possible (c'est-à-dire 2, le rang de \mathbf{W}). La matrice d'échelle a été spécifiée comme suit :

$$R = \begin{pmatrix} 200 & 0 \\ 0 & 0.2 \end{pmatrix}$$

Les lois a priori des paramètres du modèle sont :

Paramètre	Loi a priori
β_i	$\text{MVN}(\mu_\beta, \Omega)$
μ_β	$\text{MVN}(\text{mean}[], \text{prec}[])$
Ω	$\text{Wishart}(R, \rho)$
τ_C	$\text{Gamma}(0.001, 0.001)$

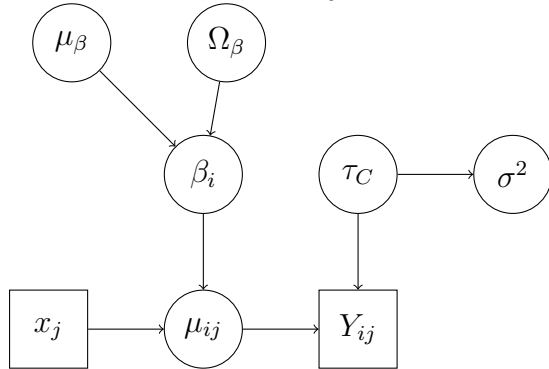
TABLE 1 – Lois a priori pour les paramètres du modèle de croissance des rats

3.1 Structure du modèle

- **Réponse** : Le poids de chaque rat i à chaque âge x_j est modélisé comme une variable aléatoire normale, $Y_{ij} \sim \text{Normal}(\mu_{ij}, \tau_c)$, où τ_c est la précision de la mesure (l'inverse de la variance).
- **Modèle linéaire** : La moyenne μ_{ij} est une fonction linéaire des âges, $\mu_{ij} = \beta_{1i} + \beta_{2i}x_j$. Ici, β_{1i} et β_{2i} représentent l'intercept et la pente de la courbe de croissance pour chaque rat, modélisés comme des tirages d'une distribution normale multivariée.
- **Effets aléatoires** : Les coefficients β_i pour chaque rat sont tirés de $\text{MVN}(\mu_\beta, \Omega)$, où μ_β sont les moyennes populationnelles des coefficients, et Ω est la matrice de précision de la population.

3.2 Graphe Orienté Acyclique (DAG)

Pour $i = 1, \dots, N$ et $j = 1, \dots, T$:



3.3 Formules Conditionnelles

- La loi conditionnelle de α est donnée par :

$$\pi(\alpha | r_1, \mu, \beta_1, \beta_2, b, \sigma^2) \propto \pi(\alpha) \prod_{i=1}^K \pi(r_{1i} | \alpha, \mu_i, \beta_1, \beta_2, b)$$

$$\pi(\alpha | r_1, \mu, \beta_1, \beta_2, b, \sigma^2) \propto \exp\left(-\frac{\alpha^2}{2 \cdot 10^6}\right) \prod_{i=1}^K p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}$$

- La loi conditionnelle de β_i :

$$\pi(\beta_i | r_1, \mu, \alpha, \beta_j, b, \sigma^2) \propto \exp\left(-\frac{\beta_i^2}{2 \cdot 10^6}\right) \prod_{i=1}^K p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}$$

— La loi conditionnelle de τ :

$$\tau|r_1, \mu, \alpha, \beta_1, \beta_2, b \sim \text{Gamma} \left(10^{-3} + \frac{K}{2}, 10^{-3} + \frac{1}{2} \sum_{i=1}^K b_i^2 \right)$$

— La loi conditionnelle de μ_i :

$$\pi(\mu_i|\dots) \propto \exp \left(-\frac{\mu_i^2}{2 \cdot 10^6} \right) \prod_{i=1}^K p_{0i}^{r_{0i}} (1 - p_{0i})^{n_{0i} - r_{0i}} p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}$$

4 Résultats

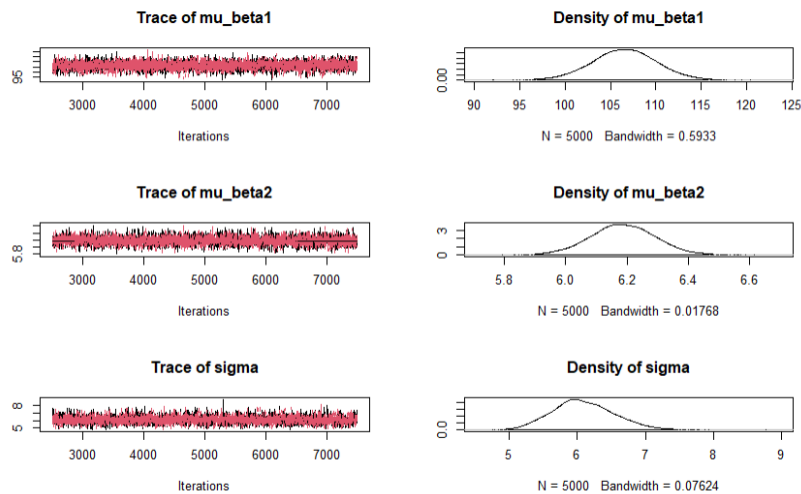
Sur le pdf, les estimations des paramètres clés sont les suivantes :

$$\begin{aligned} \mu_{\beta_1} &= 106.6 \pm 2.35, \\ \mu_{\beta_2} &= 6.185 \pm 0.1062, \\ \sigma &= 6.136 \pm 0.4781. \end{aligned}$$

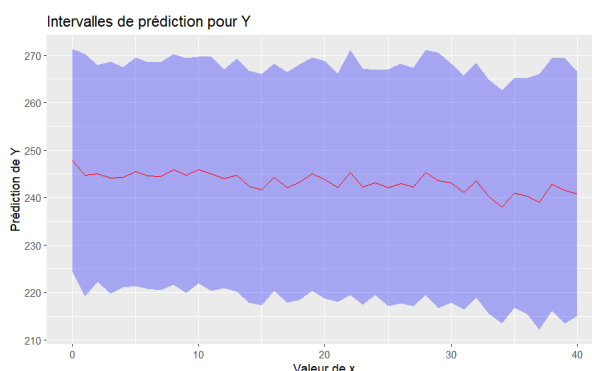
En utilisant un échantillonneur avec 10000 réalisations dont on retire les 1000 premières, on compare les moyennes et les écart-types de nos résultats à ceux de l'énoncé, on obtient des résultats quasiment pareils :

$$\begin{aligned} \mu_{\beta_1} &= 106.522 \pm 3.6083, \\ \mu_{\beta_2} &= 6.186 \pm 0.1080, \\ \sigma &= 6.092 \pm 0.4647. \end{aligned}$$

Les graphiques ci-dessous illustrent, à gauche, les chaînes de Markov générées et, à droite, leurs densités estimées. Les résultats observés sont conformes aux attentes ; les données de chaque chaîne se répartissent de manière appropriée autour des valeurs théoriques anticipées.



La stabilisation des chaînes MCMC autour de valeurs constantes sans tendance ou dérive apparente suggère une convergence réussie vers la distribution stationnaire. Les formes symétriques en cloche des densités postérieures impliquent des distributions normales, signifiant des estimations fiables et un modèle bien ajusté aux données.



Le graphique démontre la relation entre l'âge des rats (x) et leur poids (Y), avec une tendance observée à la baisse du poids moyen en fonction de l'âge. La ligne rouge représente cette tendance décroissante, tandis que l'intervalle de prédiction à 95%, marqué par la zone bleue, s'élargit avec l'avancement en âge. Cela traduit une incertitude croissante dans les estimations du poids pour les rats plus âgés.

5 Conclusion

Au fil de ce projet, nous avons dû relever des défis majeurs dus à l'évolution technologique des outils et à la complexité des données. Trois versions initiales ont révélé des divergences dans les résultats, reflétant des incompatibilités de versions logicielles et des problèmes de convergence du modèle. C'est la troisième version qui a finalement conduit à des résultats fiables et reproductibles, après des ajustements déterminants dans la configuration du modèle, la révision des priorités de distribution et l'optimisation des calculs matriciels et des procédures de convergence MCMC.