Modélisation Bayésienne Hiérarchique des Courbes de Croissance chez les Rats : Deuxième Modèle

Alexandre Cachera

1 Introduction et données

L'étude porte toujours sur l'analyse des courbes de croissance de 30 rats mesurés à 5 âges différents (5 semaines). Cette fois-ci, nous présentons un modèle plus élaboré que le premier. Ce modèle vise toujours à prédire les coefficents de regression pour chaque rat mais nous supposons cette fois que le coefficient de pente et d'intercept sont corrélés et sont propres à chaque rat : on utilise alors un modèle normal multivarié.

Concernant les données, elles sont identiques à celle du premier modèle, mise à part que l'on dispose maintenant en plus de la matrice d'échelle pour la distribution de Wisharp. L'intérêt est donc de voir si ce modèle prédit mieux les données que le précédent.

2 Modèle Bayésien Hiérarchique

On utilise donc toujours un modèle hiérarchique bayésien. Chaque rat dispose toujours de ses propres paramètres de croissance tirés selon la même loi pour modèliser l'appartenance à une même population

Le modèle hiérarchique s'écrit :

$$Y_{ij} \sim \mathcal{N}(\beta_{1,i} + \beta_{2,i}x_j, \frac{1}{\tau_c^2})$$

$$\beta_i \sim \mathcal{MVN}(\mu_\beta, \tau_\Omega)$$

On considère cette fois que les paramètres de la croissance linéaire sont corrélés entre eux, contrairement au premier modèle. Cela permet par exemple de modéliser si la corrélation est postive que les rats plus gros au départ gagnent du poids plus rapidement que les plus légers. La loi de Wisharp nous assure que la matrice Ω est définie postive.

Enfin pour les derniers paramètres, on fait le choix de loi à priori non informatives et indépendantes, par exemple grâce à des grandes variances :

$$\mu_{\beta_1}, \mu_{\beta_2} \sim \mathcal{N}(0, 10^6)$$

$$\tau_c \sim \mathcal{G}(0.001, 0.001)$$

$$\Omega \sim Wisharp(R, \rho)$$

avec $R = \begin{bmatrix} 200 & 0 \\ 0 & 0.2 \end{bmatrix}, \rho = 2$

La loi de Wisharp nous assure que la matrice Ω est définie postive.

En résumé :

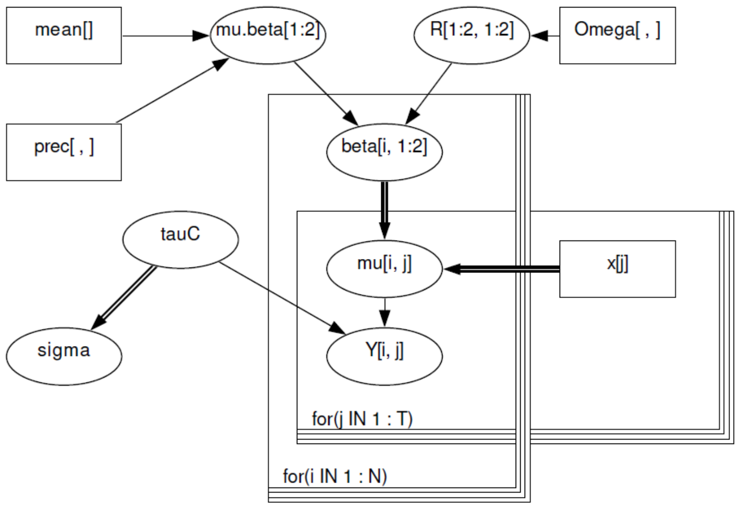

FIGURE 1 – Modèle hiérarchique

3 Implémentation de l'Échantillonneur de Gibbs

Pour estimer nos différents paramètres et mener notre étude, nous allons encore utiliser un échantilloneur de Gibbs. Pour sa mise en oeuvre, on fait comme pour le premier modèle en tenant compte que nos paramètres ont changés. Voici les résultats des calculs de chaque loi à posteriori :

$$\beta_i \mid Y, \mu, \Omega, \tau_c \sim \mathcal{MVN}((\Omega + \tau_c X^T X)^{-1}(\Omega \mu_\beta + \tau_c X^T Y_i), (\Omega + \tau_c X^T X)^{-1})$$

$$\Omega \mid \beta, \mu_\beta \sim Wisharp(N + \rho, (R^{-1} + \sum_{i=1}^{N} (\beta_i - \mu_\beta)^T (\beta_i - \mu_\beta))^{-1})$$

$$\mu_\beta \mid \beta, \Omega \sim \mathcal{MVN}((\Sigma^{-1} + N\Omega)^{-1} \Omega \sum_{i=1}^{N} \beta_i, \Sigma^{-1} + N\Omega)$$

$$\tau_c \mid Y, \beta \sim \mathcal{G}(\alpha_\tau + NT/2, \beta_\tau + \frac{1}{2} \sum_{i,j} (Y_{ij} - \beta_{1,i} - \beta_{2,i} x_j)^2)$$

Pour coller aux mesures du livre, on réalise 10000 itérations. A chaque itération, on met alors à jour chacun de ces paramètres. Puis on applique un burning de 1000 itérations comme dans le livre.

4 Résultats et inférence

Voici les résultats que nous obtenons après inférence :

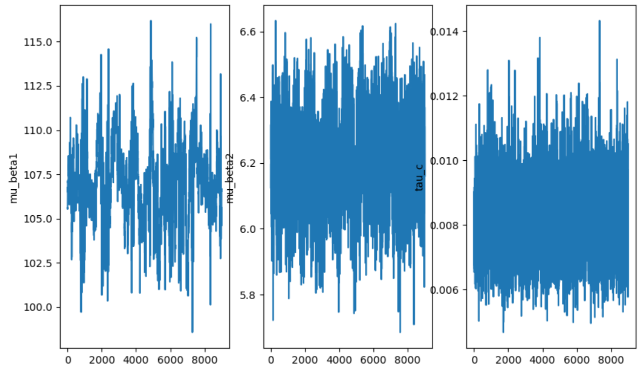

FIGURE 2 – Mixage des paramètres

Paramètre	Moyenne	Ecart-type	Intervalle de confiance
μ_{β_1}	106.92	2.15	[102.31660114, 111.02047628]
μ_{β_2}	6.20	0.12	[5.94568048 6.44858367]
σ	11.11	0.75	[9.71598805 12.68129796]

TABLE 1 – Valeurs obtenues des paramètres

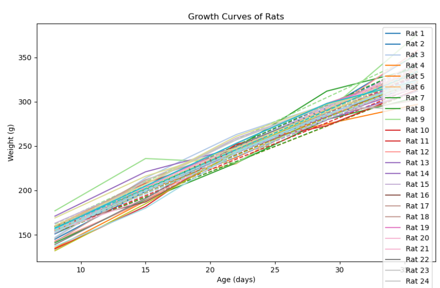

FIGURE 3 – Approximations comparées à la réalité

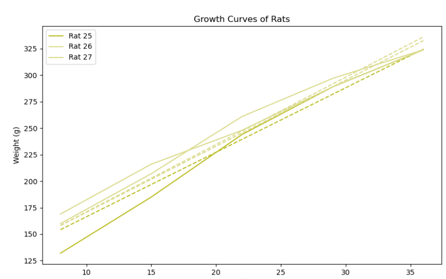

FIGURE 4 – Zoom sur 3 rats

Les résultats obetenus sont satisfaisants car les chaînes montrent un bon mixage, et l'algorithme a convergé. De plus, les valeurs obtenues sont en adéquation avec celles présentées dans le livre. Cela se réflète dans les approximations que nous obtenons pour la courbe de croissance de chaque rat : notre approximation est proche de la réalité.

5 Conclusion

Le modèle capture efficacement la variabilité propre à chaque rat tout en tenant compte de l'appartenance à une même population. Les intervalles de crédibilité sont étroits, reflétant une forte précision. Pour améliorer encore ce modèle, une extension non-linéaire pourrait peut-être mieux s'adapter aux courbes de croissance.