

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

Alexandre Cachera

1 Introduction et Données

Cette étude s'intéresse à la croissance de 30 rats dont le poids a été mesuré à plusieurs moments dans le temps. Chaque rat a été pesé à 5 âges différents : 8, 15, 22, 29 et 36 jours. L'objectif est de modéliser la croissance des rats et de s'intéresser à la variation propre à chaque rat tout en exploitant l'information commune entre eux.

Concernant les données, on dispose de la matrice Y contenant les poids des différents rats aux différents temps de mesure, X le vecteur des temps de mesure, ainsi que de N et T le nombre de rats et le nombre de mesures effectuées.

2 Modèle Bayésien Hiérarchique

On utilise donc un modèle hiérarchique bayésien, pertinent ici, car chaque rat suit une croissance à peu près linéaire, mais avec des variations individuelles. On suppose donc que les paramètres de croissance (intercept et pente) varient entre les rats selon une distribution normale commune à tous.

Le modèle hiérarchique s'écrit :

$$\begin{aligned} Y_{ij} &\sim \mathcal{N}(\alpha_i + \beta_i(x_j - x_{bar}), \frac{1}{\tau_c^2}) \\ \alpha_i &\sim \mathcal{N}(\alpha_c, \tau_\alpha) \\ \beta_i &\sim \mathcal{N}(\beta_c, \tau_\beta) \end{aligned}$$

On a donc un modèle de croissance linéaire randomisé propre à chaque rat, ce qui est en accord à première vue avec le jeu de données car en traçant les courbes propres à chaque rat, on observe des courbes croissantes qui semblent linéaires. De plus, bien que chaque rat dispose de ses propres paramètres, ces paramètres proviennent d'une même distribution et donc on modélise bien l'appartenance à une même population. On remarque également que ces paramètres ne sont pas corrélés dans ce modèle, c'est un choix arbitraire, qui sera rediscuté dans le deuxième modèle.

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

$$\begin{aligned} \mu_\alpha, \alpha_c &\sim \mathcal{N}(0, 10^6) \\ \tau_\alpha, \tau_\beta, \tau_c &\sim \mathcal{G}(0.001, 0.001) \\ \alpha_0 &= \alpha_c - \beta_c \tau_c \end{aligned}$$

On s'intéressera particulièrement à l'intercept au temps zéro α_0 .
En résumé :

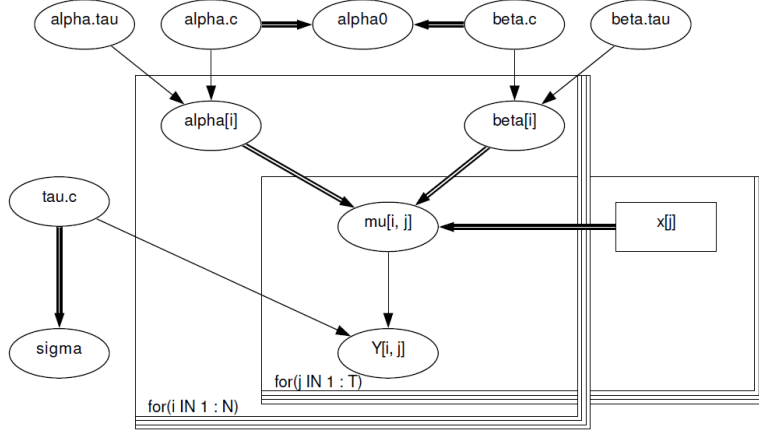


FIGURE 1 – Modèle hiérarchique

3 Implémentation de l'Échantillonneur de Gibbs

Pour estimer nos différents paramètres et mener notre étude, nous allons utiliser un échantillonneur de Gibbs, qui est ici facile à mettre en oeuvre en tenant des lois des différents paramètres, pour avoir un échantillon des lois que suivent chaque paramètre de notre modèle.

Pour sa mise en oeuvre, on définit d'abord des valeurs initiales pour nos différents paramètres (typiquement moyenne de la loi choisit par le modèle). Puis on met à jour à chaque itération en tirant selon la loi à posteriori conditionnellement aux autres variables de la variable considéré. Pour cela, on utilise les relations hiérarchiques entre chaque paramètre et les lois des différents paramètres. Voici les résultats des calculs de chaque loi à posteriori :

$$\begin{aligned}\alpha_i &\sim \mathcal{N}\left(\frac{\alpha_c * \tau_\alpha + \tau_c \sum_{i=1}^N (Y_{ij} - \beta_i(x_j - x_{bar}))}{\tau_c * T + \tau_\alpha}, \frac{1}{\tau_c * T + \tau_\alpha}\right) \\ \beta_i &\sim \mathcal{N}\left(\frac{\beta_c * \tau_\beta + \tau_c \sum_{i=1}^N (Y_{ij} - \alpha_i)(x_j - x_{bar})}{\tau_\beta + \tau_c \sum_{i=1}^N (x_j - x_{bar})^2}, \frac{1}{\tau_\beta + \tau_c \sum_{i=1}^N (x_j - x_{bar})^2}\right) \\ \alpha_c &\sim \mathcal{N}\left(\frac{\tau_\alpha \sum_{i=1}^N \alpha_i}{N\tau_\alpha + \tau}, \frac{1}{N\tau_\alpha + \tau}\right) \\ \beta_c &\sim \mathcal{N}\left(\frac{\tau_\beta \sum_{i=1}^N \beta_i}{N\tau_\beta + \tau}, \frac{1}{N\tau_\beta + \tau}\right) \\ \tau_\alpha &\sim \mathcal{G}(\alpha_\tau + N/2, \beta_\tau + \frac{1}{2} \sum_{i=1}^N (\alpha_i - \alpha_c)^2) \\ \tau_\beta &\sim \mathcal{G}(\alpha_\tau + N/2, \beta_\tau + \frac{1}{2} \sum_{i=1}^N (\beta_i - \beta_c)^2)\end{aligned}$$

$$\tau_\alpha \sim \mathcal{G}(\alpha_\tau + NT/2, \beta_\tau + \frac{1}{2} \sum_{i,j} (Y_{ij} - \alpha_i - \beta_i x_j)^2)$$

avec $\tau = 10^{-6}$, $\alpha_\tau = \beta_\tau = 10^{-3}$

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 :

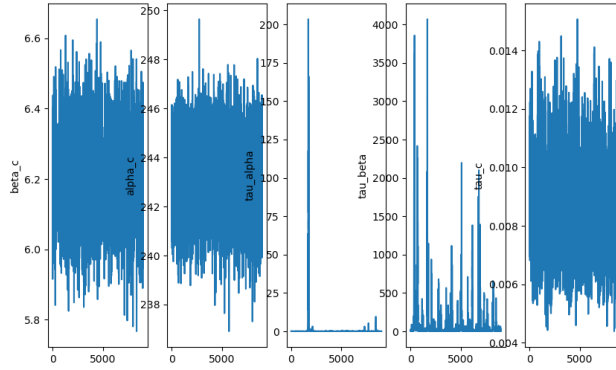


FIGURE 2 – Mixage des paramètres

Paramètre	Moyenne	Ecart-type
β_c	6.21	0.13
α_0	106.64	2.71
σ	3.56	4.09

TABLE 1 – Valeurs obtenues des paramètres

Les résultats obtenus 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 reflète 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

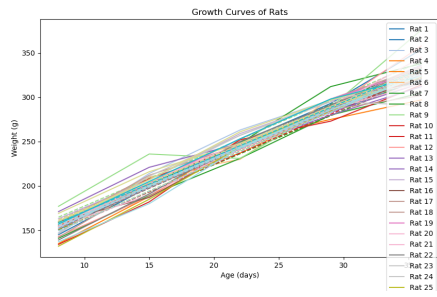


FIGURE 3 – Approximations comparées à la réalité

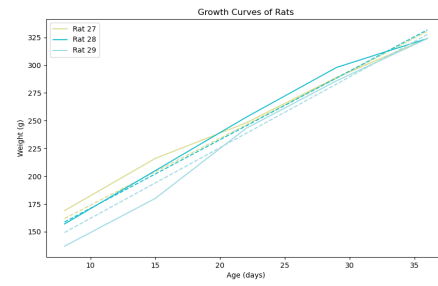


FIGURE 4 – Zoom sur 3 rats

précision. Pour améliorer encore ce modèle, une extension non-linéaire pourrait peut-être mieux s'adapter aux courbes de croissance.