
Projet Bayes : Oxford - Smooth Fit to Log-Odds Ratios

OUSSAMA HADDER - MARWANE FAHIM -
SALIMOU MOUSSA - ACHRAF DRISSI

OPTION MATHÉMATIQUES ET APPLICATIONS
[HTTPS://GITHUB.COM/NATSUHADDER/BAYES-PROJECT-2](https://github.com/NATSUHADDER/BAYES-PROJECT-2)

Tuteur:

Mention : MATHIEU RIBATET
Mathieu.Ribatet@ec-nantes.fr

Présentation des données

Les informations sont présentées sous forme de tableau à double entrée. La première entrée X indique l'âge des 27 dugongs, tandis que l'entrée Y représente leur taille. L'objectif du projet est de créer un modèle bayésien afin d'estimer la taille des dugongs en fonction de leur âge.

Dugongs	1	2	3	4	5	26	27
Âge (X)	1.0	1.5	1.5	1.5	2.5	29.0	31.5
Taille (Y)	1.80	1.85	1.87	1.77	2.02	2.27	2.57

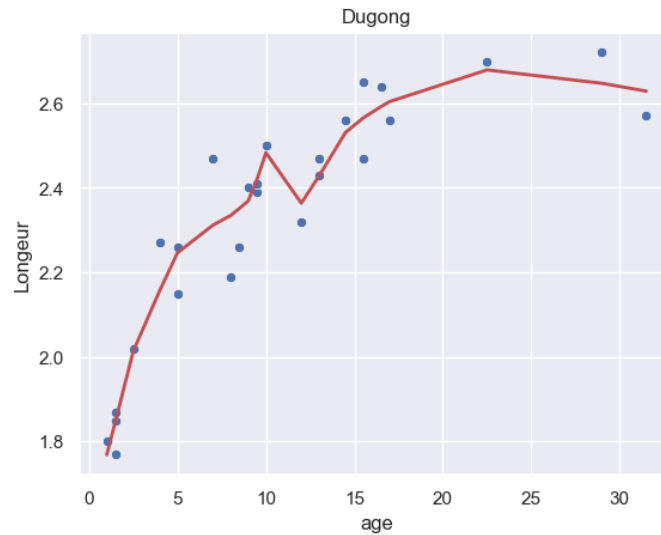


Figure 1 – Evolution de la longueur des Dugongs en fonction de leur âge.

La croissance des dugongs n'est pas linéaire et semble atteindre une valeur constante. De ce fait, nous avons choisi un modèle de régression non linéaire, le modèle de régression asymptotique, pour décrire la relation entre X (âge) et Y (taille).

D'autres modèles similaires de croissance sont disponibles dans la littérature :

- Modèle exponentiel modifié (Stevens, 1951) : $E(Y) = \alpha - \frac{\beta}{X}$
- Modèle logistique (von Bertalanffy, 1957)
- Modèle de Gompertz (Richards, 1959)

L'estimation non paramétrique de la densité des valeurs de Y (Figure 2) indique une densité suggérant une somme de densités normales. Par conséquent, nous avons opté pour le modèle suivant :

$$Y_i \sim \mathcal{N}(\mu_i, \tau) \text{ pour } i = 1, \dots, 27$$

$$\mu_i = \alpha - \beta\gamma^X \text{ avec } \alpha, \beta > 0 \text{ et } 0 < \gamma < 1$$

Lorsque X (âge) est élevé, la moyenne μ_i reste pratiquement inchangée, car γ est inférieur à 1. Cela est cohérent, car les dugongs plus âgés ont une taille semblable d'environ 2,5 mètres, tandis que les plus jeunes mesurent environ 1,7 mètre.

Modèle mathématique

Nous allons calculer les lois conditionnelles pleines pour les différents paramètres, qui seront utilisées dans notre algorithme MCMC. Les lois a priori fournies dans le Tableau 1 et le modèle DAG seront employés.

Paramètres	α	β	τ	γ
Lois a priori	$N(0, 10^3)$	$N(0, 10^3)$	$\Gamma(10^{-3}, 10^{-3})$	$\text{Uniform}(0.5, 1)$

Table 1 – Lois a priori pour nos paramètres

Les variables α , β , τ et γ doivent être strictement positives, et γ doit être compris entre 0 et 1. Nous utiliserons la chaîne de Hastings-Metropolis avec une marche aléatoire log-normale et une loi de proposition gaussienne $N(0, \sigma_{prop}^2)$.

Pour γ , nous ajoutons la condition que la valeur proposée doit être inférieure à 1 pour être acceptée.

Les lois conditionnelles

Nous obtenons les lois conditionnelles pleines suivantes :

— Pour α :

$$\begin{aligned} \pi(\alpha | \dots) &\propto \prod_{i=1}^n \pi(y_i | \alpha, \beta, \gamma, \tau) \times \pi(\alpha) \\ \pi(\alpha | \dots) &\propto \tau^{n/2} \times \exp\left(-\frac{\sum_{i=1}^n (Y_i - \alpha + \beta\gamma^{x_i})^2 \tau}{2}\right) \times \exp\left(-\frac{\alpha^2 \cdot 10^{-6}}{2}\right) \end{aligned}$$

— Pour β :

$$\begin{aligned} \pi(\beta | \dots) &\propto \prod_{i=1}^n \pi(y_i | \alpha, \beta, \gamma, \tau) \times \pi(\beta) \\ \pi(\beta | \dots) &\propto \tau^{n/2} \times \exp\left(-\frac{\sum_{i=1}^n (Y_i - \alpha + \beta\gamma^{x_i})^2 \tau}{2}\right) \times \exp\left(-\frac{\beta^2 \cdot 10^{-6}}{2}\right) \end{aligned}$$

— Pour τ :

$$\pi(\tau|\dots) \propto \prod_{i=1}^n \pi(y_i|\alpha, \beta, \gamma, \tau) \times \pi(\tau)$$

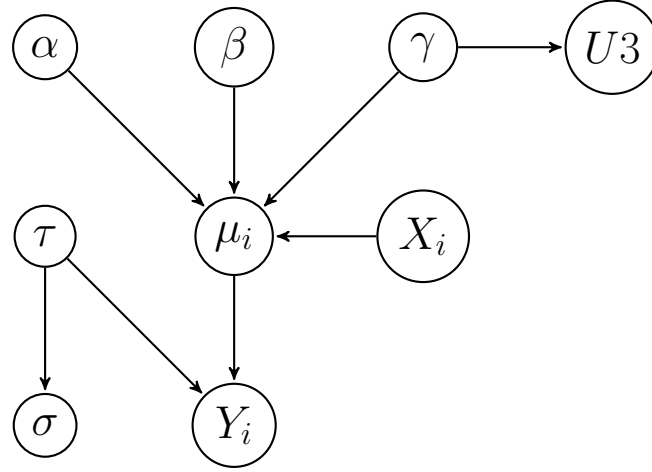
$$\pi(\tau|\dots) \propto \tau^{n/2} \times \exp\left(-\frac{\sum_{i=1}^n (Y_i - \alpha + \beta\gamma^{x_i})^2 \tau}{2}\right) \times \tau^{k-1} \exp(-\frac{\tau}{\theta})$$

— Pour γ :

$$\pi(\gamma|\dots) \propto \prod_{i=1}^n \pi(y_i|\alpha, \beta, \gamma, \tau) \times \pi(\gamma)$$

$$\pi(\gamma|\dots) \propto \tau^{n/2} \times \exp\left(-\frac{\sum_{i=1}^n (Y_i - \alpha + \beta\gamma^{x_i})^2 \tau}{2}\right) \times \mathbb{K}_{(0,1)}(\gamma)$$

Le graphe DAG peut être représenté comme suit :



Analyse des résultats

Méthode utilisée :

Pour estimer les variables aléatoires listées précédemment, nous avons utilisé pour cela l'algorithme de Gibbs. Cet algorithme est une méthode d'échantillonnage de Monte Carlo par chaînes de Markov (MCMC) qui permet de générer des échantillons à partir de la distribution a posteriori des paramètres étant donné les données. Afin d'assurer la convergence de l'algorithme, nous avons utilisé une phase de "burn-in" de 1000 échantillons. Durant cette phase, les échantillons générés par l'algorithme sont écartés et non utilisés pour estimer les paramètres. Cette étape permet de réduire l'influence des conditions initiales sur les résultats finaux.

Résultats et discussion :

Nous obtenons , après échantillonnage, les chaînes de Markov et les densités des différents paramètres que l'on souhaite estimer (α, β, γ et σ). Les chaînes de Markov présentent généralement un comportement normal, ce qui signifie que les valeurs prises par la chaîne sont distribuées autour des valeurs attendues pour les paramètres.

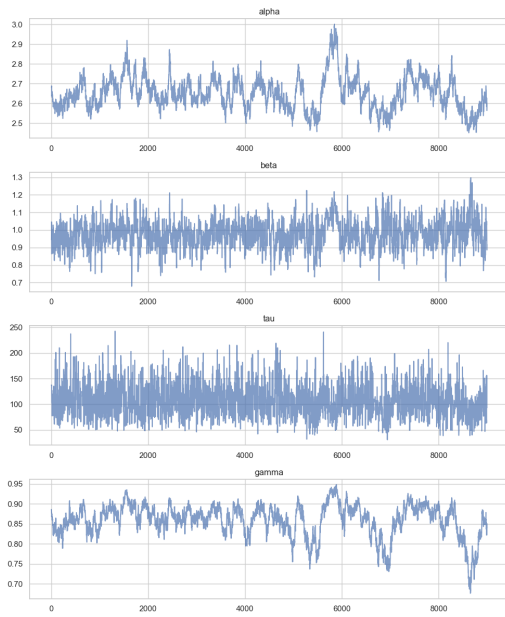


Figure 2 – Chaînes de Markov des paramètres estimés.

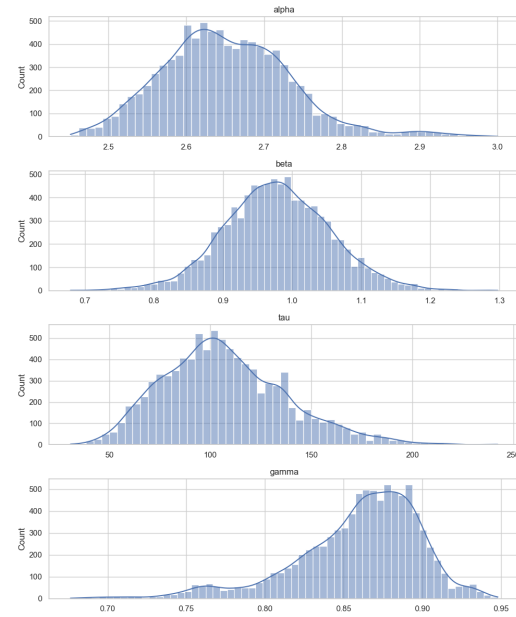


Figure 3 – Densité des paramètres estimés.

Nous pouvons donc comparer les moyennes et les écarts-types des paramètres estimés avec ceux de l'énoncé :

	Moyenne		Écart-type	
	Estimation	Référence	Estimation	Référence
α	2.651705	2.652	0.083287	0.07094
β	0.978081	0.9729	0.077506	0.07649
γ	0.099668	0.0992	0.014793	0.01496
σ	0.859035	0.8623	0.040543	0.03259

Table 2 – Résultats et énoncés des paramètres

En mettant en parallèle nos résultats avec les valeurs de référence, on peut voir que nos estimations se rapprochent de ce qu'on attendait. Ça nous montre que notre mise en place de l'algorithme de Gibbs fonctionne bien et que notre modèle est adapté pour analyser les données de cette recherche.

Il faut quand même préciser que, même si nos estimations se rapprochent des valeurs de référence, il y a toujours une incertitude qui vient avec l'utilisation de l'algorithme de Gibbs pour estimer les paramètres. Cette incertitude se voit dans les intervalles de crédibilité, qui nous donnent une idée de l'éventail où les vrais paramètres pourraient se trouver avec une certaine probabilité.

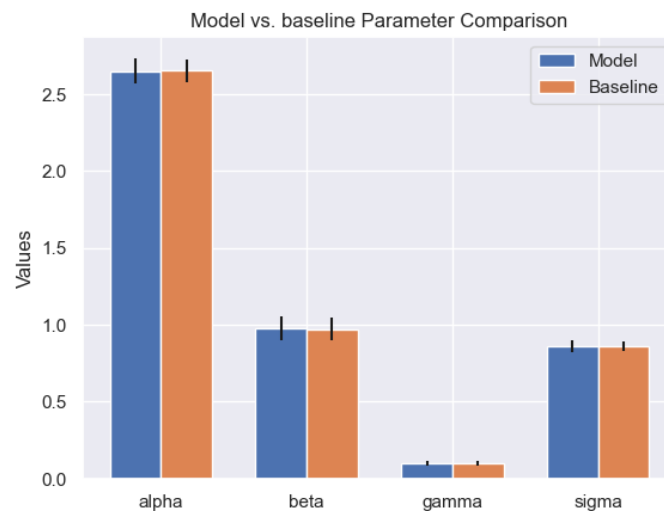


Figure 4 – Comparaison entre les données estimées et les données de référence.

Interpretation :

La valeur obtenue pour β , approchant 1, indique que notre modèle se rapproche d'un modèle exponentiel modifié. De plus, le fait que γ soit près de 0 suggère que la croissance en longueur des dugongs en fonction de leur âge est rapide. Quant à α , sa moyenne correspond bien aux données et au problème étudié, car elle est proche de la valeur maximale de Y (2,72). Enfin, la valeur de σ démontre que les écarts entre les données et leur moyenne sont minimales.

En exploitant les moyennes de nos chaînes de Markov, nous traçons la progression de la moyenne de Y_i par rapport à l'âge. La courbe confirme que la longueur moyenne des dugongs s'accroît en fonction de l'âge.

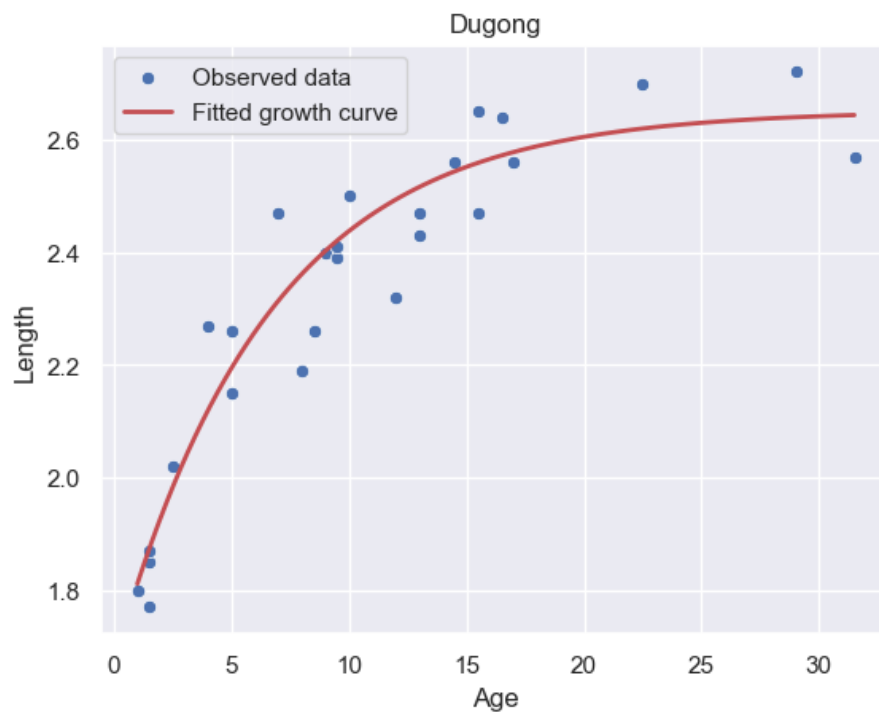


Figure 5 – Evolution de la taille estimée en fonction de l'âge.