
Dugongs - Non linear growth curve

-BAYES-Projet-2

IDRISSI KARIM, KOUTIT ABDELLAH,
MOURDI ELIAS, SELAMNIA NAJIB

Encadré par : M. Mathieu Ribatet

Présentation des données

Les informations du sujet sont données sous la forme suivante :

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

L'objectif de notre projet est d'estimer la taille des dugongs en fonction de leurs âges. Pour ce faire, nous disposons de ces deux données pour 27 dugongs, recensées dans le tableau qui suit.

On peut afficher le graphique suivant :

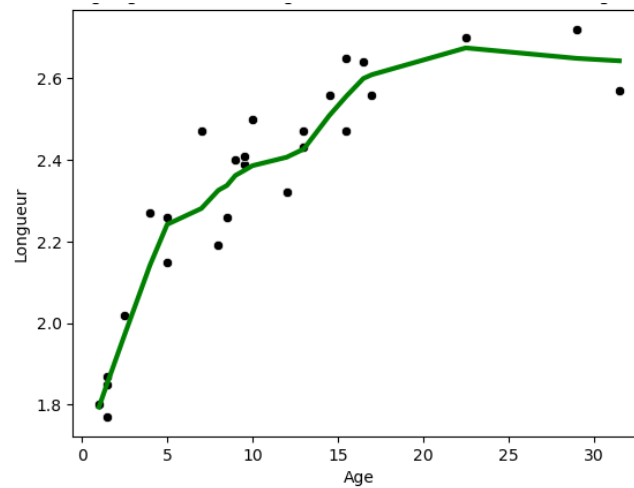


FIGURE 1 – Taille des dugongs en fonction de leur âge

La taille des dugongs ne semble pas suivre une tendance linéaire et converge vers une valeur constante. Ainsi, dans le but de décrire la relation entre l'âge et la taille, nous allons considérer un modèle non linéaire, le modèle de régression asymptotique qui nous semble approprié à l'étude.

L'analyse non paramétrique de la distribution de Y suggère que cette distribution peut être décrite comme la combinaison de plusieurs distributions normales. Nous pouvons donc considérer le modèle suivant :

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

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

Ainsi, lorsque l'âge augmente, la moyenne μ_i reste pratiquement inchangée, car γ est inférieur à 1. On a l'effet recherché puisque les dugongs les plus âgés ont une taille semblable d'environ 2,5 mètres, c'est à dire qu'on a convergence vers une taille particulière, là où les plus jeunes voient leurs tailles de plus en plus importantes avec l'âge.

Modèle mathématique

Nous allons alors calculer les différentes lois à posteriori 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	$N(0, 10^3)$	$N(0, 10^3)$	$\Gamma(10^{-3}, 10^{-3})$	$U(0.5, 1)$

TABLE 1 – Lois à 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 l'algorithme de Metropolis-Hastings avec une marche aléatoire log-normale et une loi de proposition gaussienne $N(0, \sigma_{\text{prop}}^2)$.

De plus, pour γ , on considère que la valeur proposée doit être inférieure à 1 pour être acceptée.

Lois à posteriori

Après calcul, nous obtenons :

— Pour α :

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

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

— Pour β :

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

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

— Pour τ :

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

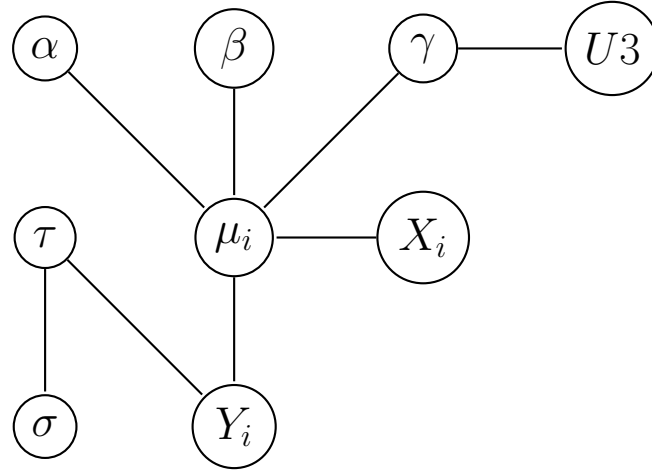
$$\pi(\tau|\dots) \propto \tau^{n/2} \times \exp\left(-\frac{\sum_{i=1}^n (Y_i - \alpha + \beta\gamma^{x_i})^2 \tau}{2}\right) \times \tau^{k-1} \exp\left(-\frac{\tau}{\theta}\right)$$

— 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 :



Implémentation algorithmique

Pour mener à bien notre projet, nous avons utilisé l'algorithme de Gibbs. 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

Les chaînes de Markov et les densités des différents paramètres que l'on souhaite estimer (α, β, γ et σ) sont représentées ci-contre. 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.

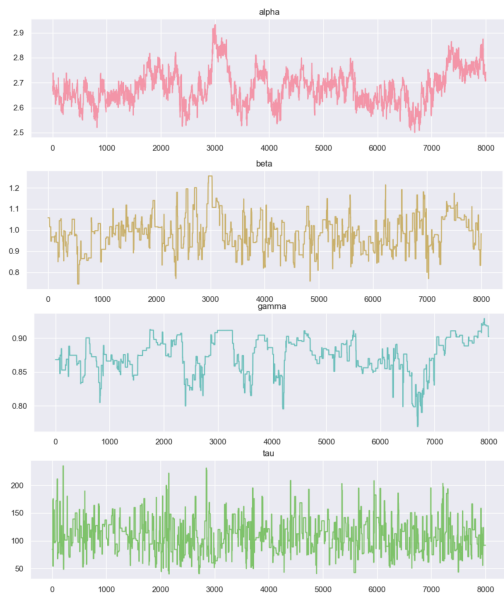


FIGURE 2 – Chaînes de Markov des paramètres estimés

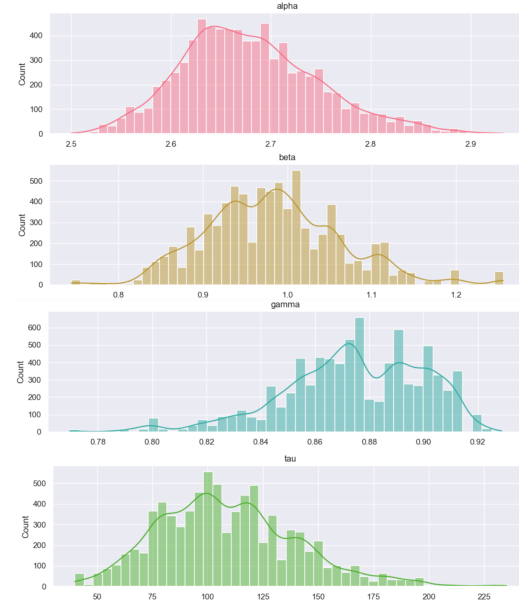


FIGURE 3 – Densité des paramètres estimés

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.

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.

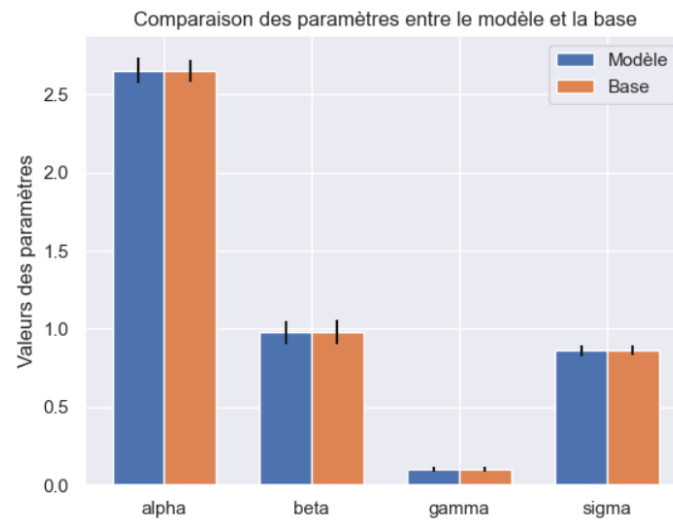


FIGURE 4 – Comparaison entre les données estimées et les données de référence

Interprétation

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 leurs moyennes 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.

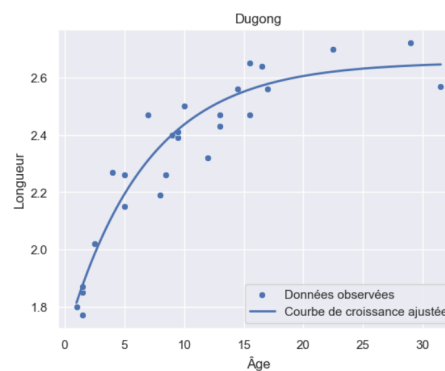


FIGURE 5 – Taille des dugongs en fonction de leur âge