

# Projet BAYES

Théo Duquesne et Stéphane Courtault

Avril 2025

## 1 Introduction

Nous avons ici l'étude du pic de sensibilité des yeux d'une population de singes. L'énoncé du problème nous stipule que l'on a un mélange gaussien. Nous vérifions l'hypothèse avec un algorithme M.E.

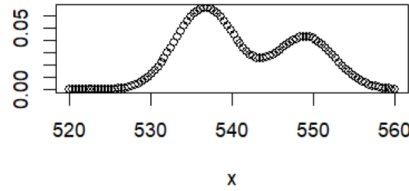


Figure 1: Résultat M.E

On a bien quelque chose qui se rapproche de notre décomposition par noyaux gaussiens.

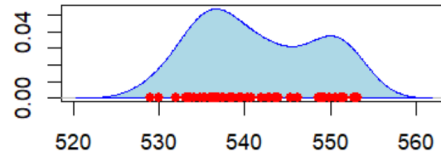


Figure 2: Résultat noyaux gaussiens

Nous allons exprimer les lois marginales de chacune des composantes de notre problème.

### 1.1 $\lambda_1$

$$\pi(\lambda_1 \mid -) \propto \mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \cdot \mathcal{L}(\lambda_2 \mid \theta, \lambda_1) \cdot \pi(\lambda_1)$$

$\mu$  est déterministe en fonction de  $T$ ,  $\lambda_1$  et  $\lambda_2$ . Ainsi :

$$\mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \propto \prod_i \mathbb{1}_{\mu_i = \lambda_{T_i}}$$

De plus,  $\lambda_2$  est déterministe en fonction de  $\lambda_1$  et  $\theta$ . Donc :

$$\mathcal{L}(\lambda_2 \mid \theta, \lambda_1) \propto \mathbb{1}_{\lambda_2 = \lambda_1 + \theta}$$

$\lambda_1$  suit une loi a priori normale centrée, d'écart-type  $10^6$ . On en conclut que  $\lambda_1$  suit une loi déterministe en  $\lambda_2 - \theta$ .

## 1.2 $\lambda_2$

$$\pi(\lambda_2 \mid -) \propto \mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \cdot \pi(\lambda_2 \mid \theta, \lambda_1)$$

$\mu$  est déterministe en fonction de  $T$ ,  $\lambda_1$  et  $\lambda_2$  :

$$\mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \propto \prod_i \mathbb{1}_{\mu_i = \lambda_{T_i}}$$

De plus :

$$\pi(\lambda_2 \mid \theta, \lambda_1) \propto \mathbb{1}_{\lambda_2 = \lambda_1 + \theta}$$

On en conclut que :

$$\pi(\lambda_2 \mid -) \propto \prod_i \mathbb{1}_{\mu_i = \lambda_{T_i}} \cdot \mathbb{1}_{\lambda_2 = \lambda_1 + \theta}$$

## 1.3 $P$

$$\pi(P \mid -) \propto \mathcal{L}(T \mid P) \cdot \pi(P \mid \alpha)$$

$P$  suit une loi de Dirichlet de paramètre  $\alpha$ .  $T$  est un vecteur dont les composantes suivent des lois de Bernoulli iid de paramètre  $P$ .

Ces deux lois sont conjuguées. D'après Wikipédia, la loi a posteriori de  $P$  est une Dirichlet de paramètres proportionnels à  $(\alpha + \#2, 1 - \alpha + \#1)$ , où  $\#i$  est le nombre d'éléments attribués au groupe  $i$ .

## 1.4 $T$

$$\pi(T \mid -) \propto \mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \cdot \pi(T \mid P)$$

$\mu$  est déterministe en fonction de  $T$ ,  $\lambda_1$  et  $\lambda_2$  :

$$\mathcal{L}(\mu \mid T, \lambda_1, \lambda_2) \propto \prod_i \mathbb{1}_{\mu_i = \lambda_{T_i}}$$

Ainsi, comme  $T$  est parfaitement défini par les valeurs prises par  $\mu$ , il n'est même pas nécessaire de s'intéresser à  $\pi(T \mid P)$ .

Cependant, si on le fait :

$$\pi(T | P) \propto \prod_i (\mathbb{1}_{T_i=1} \cdot (1 - P) + \mathbb{1}_{T_i=2} \cdot P)$$

Donc, on retrouve encore :

$$\pi(T | \cdot) \propto \prod_i \mathbb{1}_{\mu_i=\lambda_{T_i}} \cdot \mathbb{1}_{\lambda_2=\lambda_1+\theta}$$

### 1.5 $\mu$

$$\pi(\mu | \cdot) \propto \mathcal{L}(y | \mu, \tau) \cdot \pi(\mu | T, \lambda_1, \lambda_2)$$

Pour tout  $i$ ,  $y_i$  suit une loi  $\mathcal{N}(0, 1)$ . Donc :

$$\mathcal{L}(y | \mu, \tau) \propto \prod_i \phi(y_i, \mu_i, \frac{1}{\tau})$$

De plus,  $\mu_i$  est déterministe sachant  $T$ ,  $\lambda_1$ , et  $\lambda_2$ . Donc on ne peut tirer les valeurs que parmi une unique configuration de  $\mu$ .

### 1.6 $\tau$

$$\pi(\tau | \cdot) \propto \mathcal{L}(y | \mu, \tau) \cdot \pi(\tau)$$

Or,  $\tau \sim \Gamma(0.001, 0.001)$

$y | \mu, \tau \sim \mathcal{N}(\mu, \frac{1}{\tau})$

Donc on a avec les lois conjuguées et on a :

$$\tau | \cdot \sim \Gamma(\alpha + \frac{n}{2}; \beta + \frac{\sum_i (x_i - \lambda_1)^2}{2})$$

Nous allons exprimer les lois marginales de chacune des composantes de notre problème. Le problème c'est qu'on a du déterminisme ainsi nous allons schématiser un nouveau DAG.

## 2 Loi marginale de $\lambda_1$

$$\pi(\lambda_1 | \cdot) \propto \mathcal{L}(y | T, \lambda_1, \tau, \theta) \cdot \pi(\lambda_1)$$

Or,

$$\pi(y | \cdot) \propto \prod_i \phi(y_i, \lambda_1 + \mathbb{1}_{T_i=2}\theta, \frac{1}{\tau})$$

$$\pi(y | \cdot) \propto \prod_i \phi(y_i - \mathbb{1}_{T_i=2}\theta, \lambda_1, \frac{1}{\tau})$$

Ce sont donc des lois normales de même variance et moyenne.

$$\pi(\lambda_1) \propto \phi(\lambda_1, 0, 10^6)$$

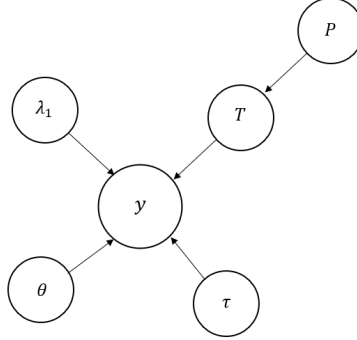


Figure 3: Nouveau DAG

Les deux se conjuguent et nous donnent :

$$\lambda_1 \mid - \propto \mathcal{N} \left( \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \left( \frac{\mu_0}{\sigma_0^2} + \frac{\sum_i x_i}{\sigma^2} \right); \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \right)$$

avec  $\sigma_0^2 = 10^6$ ,  $\sigma^2 = \frac{1}{\tau}$  et  $x_i = y_i - \mathbb{1}_{T_i=2}\theta$

theta va suivre une loi demi gaussienne qui est très pratique à simuler une loi avec une loi gaussienne

### 3 Loi marginale de $\theta$

$$\pi(\lambda_1 \mid -) \propto \mathcal{L}(y \mid T, \lambda_1, \tau, \theta) \cdot \pi(\lambda_1)$$

Or,

$$\pi(y \mid -) \propto \prod_{i, T_i=2} \phi(y_i - \lambda_1, \theta, \frac{1}{\tau})$$

Et,

$$\pi(\theta) \propto \mathbb{1}_{\theta \geq 0} \cdot \phi(\theta, 0, 10^6)$$

Cela peut se conjuguer :

$$\theta \mid - \propto \mathcal{N} \left( \frac{1}{\frac{1}{\sigma_0^2} + \frac{k}{\sigma^2}} \left( \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i, T_i=2} \tilde{x}_i}{\sigma^2} \right); \frac{1}{\frac{1}{\sigma_0^2} + \frac{k}{\sigma^2}} \right)$$

avec  $\sigma_0^2 = 10^6$ ,  $\sigma^2 = \frac{1}{\tau}$ ,  $\tilde{x}_i = y_i - \lambda_1$ ,  $k = \text{Card} \{i, T_i = 2\}$  et  $\mu_0 = 0$

## 4 Loi marginale de P

$$P \sim \text{Dirichlet}(\alpha + k, \alpha + n - k)$$

## 5 Loi marginale de $\tau$

$$\pi(\tau \mid -) \propto \mathcal{L}(y \mid -) \cdot \pi(\tau)$$

Or,  $\tau \sim \Gamma(\alpha, \beta)$

Ainsi, en conjugant les lois on retrouve :

$$\tau \mid - \propto \Gamma\left(\alpha + \frac{n}{2}; \beta + \frac{\sum_i x_i}{2}\right)$$

## 6 Loi marginale de T

On voit les  $T_i$  comme des variables latentes

$$\pi(T_i \mid -) \propto \mathcal{L}(y_i \mid -) \cdot \pi(T_i \mid P)$$

Or,  $\pi(T_i \mid P) = P$  si  $T_i = 2$  et  $1 - P$  sinon

$$T \mid - \propto \phi(y_i, \lambda_1 + \mathbb{1}_{T_1=2}\theta, \frac{1}{\tau}) \cdot (P \mathbb{1}_{T_i=2} + (1 - P) \mathbb{1}_{T_i=1})$$

## 7 Simulation

Nous allons simuler nos variables à l'aide de l'algorithme de Gibbs

## 8 Résultats

On obtient systématiquement un résultat faux. Alors que la probabilité d'appartenance au groupe 2 aurait dû tendre vers 0.4, elle tendait systématiquement vers 0.5 . Que ce soit une cause où une conséquence, nous avons obtenu dans les meilleurs cas un  $\lambda$  autour de 40 et un  $\theta$  autour de 5 ou 6, loin des valeurs de 36 et 12 respectivement que nous étions sensés trouver.

Lorsque l'on regarde les chaînes en détail, on remarque que  $\theta$  reste souvent proche de 0, et parfois prend des valeurs extrêmement grandes.

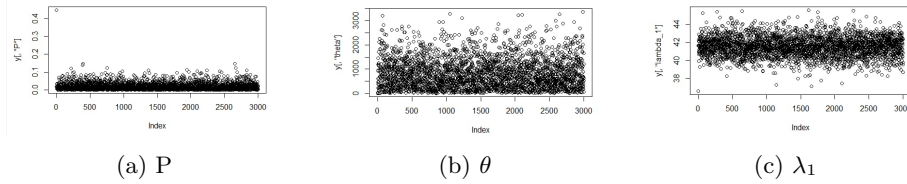


Figure 4: Échantillonnage sans tentatives d'interférence de notre part

Nous avons compris que cela arrive lorsque la variable latente  $T$  (qui attribue leurs labels aux points) est un vecteur uniforme en 1. À ce moment, la vraisemblance n'a donc plus aucun impact sur la loi a posteriori de  $\theta$ , et celle-ci se retrouve alors être la loi a priori, avec une très forte variance

Notre meilleure hypothèse explicative est la suivante :

Nous sommes coincés dans un cercle vicieux car :

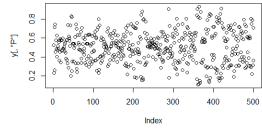
- $\tau$  est très petit
- les vraisemblances des  $y$  sachant leur label n'ont pas beaucoup d'influence sur les attributions des labels (car la vraisemblance est normale avec une faible precision)
- seule  $P$  compte dans l'attribution des labels, donc ne fait rien pour augmenter  $\tau$  (on ne s'attend pas vraiment à ce qu'un vecteur de bernouilli donne les bons labels)
- on est dans une boucle de retroaction positive où  $P$  tend très rapidement vers 0 ou 1.
- on reste coincés avec un vecteur  $T$  constant jusqu'à ce que par hasard on sorte de cette situation en inversant les rôles des deux groupes.
- comme tous les  $y_i$  sont dans le même groupe, la précision reste très faible

Pour régler ce problème nous avons essayé plusieurs choses:

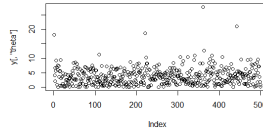
Nous avons forcé quelques labels extrémaux afin d'empêcher de faire entièrement disparaître une catégorie. Cela a conséquemment augmenté la stabilité, mais, pour que le résultat soit un minimum concluant, il faut en fixer une quantité très importante.

Nous avons essayé de jouer sur les paramètres initiaux, en particulier de tau et de alpha, sans grand succès. On a fait du thinning pour retirer les valeurs aberrantes.

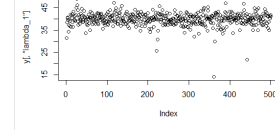
Le thinning a permis de retirer la majeure partie des valeurs aberrantes, mais les moyennes des paramètres demeurent néanmoins fausses.



(a)  $P$



(b)  $\theta$



(c)  $\lambda_1$

Figure 5: Échantillonnage les rectifications évoquées ci-dessus