Master 1 Data Science
Année 2020-2021

Travaux Encadrés de Recherche

Algorithme EM

Alexandre Belnou et Fanny Courant

Encadrés par Frédéric Proïa

Table des matières

Première partie

Définitions et propriétés de l'algorithme EM

L'algorithme EM (Expectation-Maximisation) est un algorithme itératif proposé par Dempster, Laird et Rubin. Il permet d'obtenir les estimateurs du maximum de vraisemblance lorsque cela n'est pas possible par le calcul direct, en introduisant des variables inobservables appelées variables latentes.

1 Notations

Dans cette partie, nous noterons Y le vecteur aléatoire correspondant aux données observées et incomplètes, appartenant à l'espace noté $\mathcal{Y}$ et ayant pour fonction de densité $g(y;\phi)$. $\phi = (\phi_1, \ldots, \phi_r)^T$ est un vecteur de paramètres inconnus. On notera $X = (x_1, \ldots, x_n)$ le vecteur des données complètes appartenant à l'espace $\mathcal{X}$. Ces données ne sont pas observables directement mais indirectement à travers les données complètes. C'est à dire qu'il existe une relation de $\mathcal{X}$ vers $\mathcal{Y}$.

Soit $f(x;\phi)$ la densité des données complètes où $\mathcal{X}(y)$ est un sous-espace observable de $\mathcal{X}$, c'est-à-dire :
$\mathcal{X}(y) = \{x \in \mathcal{X}; y = y(x)\} \subset \mathcal{X}$
La densité des données incomplètes est reliée à celle des données complètes par la relation :
$g(y;\phi) = \int_{\mathcal{X}(y)} f(x;\phi)dx$
L'algorithme EM a pour objectif d'estimer la valeur de ϕ maximisant la valeur de la densité g(y ;ϕ) sachant les données observées Y.

2 Les étapes E et M

Chaque itération est composée de 2 étapes :

1. **Etape E**

   Lors de cette étape, on procède à l'estimation des données inconnues sachant les données observées et la valeur du paramètre $\phi^{(p)}$ obtenue à l'étape précédente. Pour cela, on calcule :

   $$Q(\phi|\phi^{(p)}) = \mathbb{E}_{\phi^{(p)}}[\log\ f(X;\phi)\,|\,Y]$$

   $Q(\phi|\phi^{(p)})$ est l'espérance conditionnelle de la log-vraisemblance des données complètes X sachant les données incomplètes Y et le paramètre $\phi^{(p)}$.

2. **Etape M**

   Pour obtenir $\phi^{(p+1)}$, on maximise la fonction $Q(\phi|\phi^{(p)})$ obtenue à l'étape E par rapport à ϕ, c'est-à-dire :

   $$\phi^{(p+1)} = \underset{\phi}{\operatorname{argmax}}\, Q(\phi|\phi^{(p)})$$

   On maximise la vraisemblance des données complètes en utilisant l'estimation de la log-vraisemblance des données complètes X sachant les données incomplètes Y obtenue à l'étape E.

L'article de Dempster, Laird et Rubin demontre que sous les conditions évoquées dans le théorème qui suit, que nous admettons, la suite des estimateurs $\phi^{(p)}$ obtenue par l'algorithme EM converge.

Théorème 1.
Supposons que $\forall p$, $\phi^{(p)}$ *soit tel que :*
1) $\log\left(g(y;\phi^{(p)})\right)$ *soit borné*
2) $\exists \lambda > 0$ *un scalaire tel que :* $Q(\phi^{(p+1)}|\phi^{(p)}) - Q(\phi^{(p)}|\phi^{(p)}) \geq \lambda(\phi^{(p+1)} - \phi^{(p)})(\phi^{(p+1)} - \phi^{(p)})^T$
Alors la suite des $\phi^{(p)}$ *converge vers une valeur* ϕ^*.

3 Famille exponentielle régulière

On suppose ici que $f(x;\phi)$ appartient à la famille exponentielle régulière, c'est-à-dire $f(x;\phi)$ est de la forme :

$$f(x;\phi) = b(x)\ \exp(\phi\ t(x)^T)/a(\phi)$$

où ϕ est le vecteur de taille $1 \times r$ des paramètres, t(x) le vecteur de taille $1 \times r$ correspondant à une statistique exhaustive pour le vecteur X et $a(\phi)$ et $b(x)$ sont des fonctions scalaires.

Soit la fonction de log-vraisemblance :

$$\log\ f(X;\phi) = -\log(a(\phi)) + \log(b(x)) + \phi\ t(x)^T$$

On remarque que maximiser $\log\ f(X;\phi) = -\log(a(\phi)) + \log(b(x)) + \phi\ t(x)^T$ par rapport à ϕ est équivalent à maximiser $-\ \log(a(\phi)) + \phi\ t(x)^T$.

A l'étape E, on calcule $Q(\phi|\phi^{(p)})$:

$$Q(\phi|\phi^{(p)}) = \mathbb{E}_{\phi^{(p)}}[\log\ f(X;\phi)\,|\,Y] = -\ \log(a(\phi)) + \phi\,\mathbb{E}_{\phi^{(p)}}[t(X)^T|Y]$$

De plus, on a :

$$\begin{aligned}
\mathbb{E}_\phi[t(X)] &= \int_{\mathcal{X}} t(x)\ f(x;\phi)\ dx \\
&= \int_{\mathcal{X}} t(x)\ b(x)\ \exp(\phi\ t(x)^T)/a(\phi)\ dx \\
&= \frac{1}{a(\phi)} \int_{\mathcal{X}} t(x)\ b(x)\ \exp(\phi\ t(x)^T)\ dx \\
&= \frac{1}{a(\phi)} \int_{\mathcal{X}} \frac{\partial}{\partial\phi}\ b(x)\ \exp(\phi\ t(x)^T)\ dx \\
&= \frac{1}{a(\phi)}\ \frac{\partial}{\partial\phi} \int_{\mathcal{X}}\ b(x)\ \exp(\phi\ t(x)^T)\ dx \qquad \text{d'après la règle d'intégration de Leibniz}
\end{aligned}$$

Or, $f(x;\phi) = b(x)\ \exp(\phi\ t(x)^T)/a(\phi)$ est une densité. On a donc :

$$\int_{\mathcal{X}} b(x)\ \exp(\phi\ t(x)^T)/a(\phi)\ dx = 1 \iff \int_{\mathcal{X}} b(x)\ \exp(\phi\ t(x)^T)\ dx = a(\phi)$$

Ce qui implique, $\mathbb{E}_\phi[t(X)] = \frac{a'(\phi)}{a(\phi)}$

A l'étape M, on maximise la fonction $Q(\phi|\phi^{(p)})$ par rapport à ϕ :

$$\begin{aligned}\frac{\partial}{\partial\phi}Q(\phi|\phi^{(p)}) = 0 &\iff -\frac{a'(\phi)}{a(\phi)} + \mathbb{E}_{\phi^{(p)}}[t(X)|Y] = 0 \\ &\iff -\mathbb{E}_{\phi}[t(X)] + \mathbb{E}_{\phi^{(p)}}[t(X)|Y] = 0 \\ &\iff \mathbb{E}_{\phi}[t(X)] = \mathbb{E}_{\phi^{(p)}}[t(X)|Y]\end{aligned}$$

On a donc $t^{(p)} = \mathbb{E}_{\phi^{(p)}}[t(X)|Y] = \mathbb{E}_{\phi}[t(X)]$. Pour déterminer $\phi^{(p+1)}$, on doit résoudre l'équation $t^{(p)} = \mathbb{E}_{\phi}[t(X)]$. Où $t^{(p)}$ correspond à la valeur de la statistique exhaustive calculée en fonction de l'estimation des données complètes calculées à l'étape E.

Deuxième partie

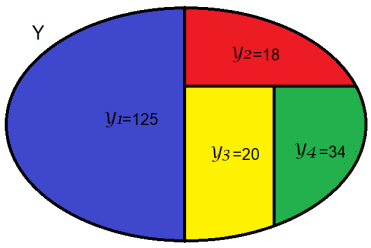
Exemples d'application de l'algorithme EM

4 Exemple classification des animaux

Dans ce premier exemple, nous traiterons le cas de la répartition d'individus d'une même espèce au sein de différentes catégories. Dans ce problème présenté par Rao en 1965, nous disposons de 197 observations réparties en quatre groupes, l'échantillon des données observées est :

$$Y = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$$

Où y_i désigne le nombre d'animaux appartenant au i-ème groupe.
On pose $N = y_1 + y_2 + y_3 + y_4 = 197$



On suppose que la distribution des individus au sein de chaque classe suit une loi multinomiale dont les paramètres sont :

$$(N, \frac{1}{2} + \frac{1}{4}\pi, \frac{1}{4} - \frac{1}{4}\pi, \frac{1}{4} - \frac{1}{4}\pi, \frac{1}{4}\pi)$$

avec $\pi \in$ [0,1] .

La fonction de densité des données observées est :

$$g(Y, \pi) = \binom{N}{y_1, y_2, y_3, y_4}(\frac{1}{2} + \frac{1}{4}\pi)^{y_1}(\frac{1}{4} - \frac{1}{4}\pi)^{y_2}(\frac{1}{4} - \frac{1}{4}\pi)^{y_3}(\frac{1}{4}\pi)^{y_4}$$

Premièrement, calculons $\widehat{\pi}$ l'estimateur de π obtenu en maximisant la vraisemblance de Y.
Soit $\ell(Y;\pi) = K_Y(\frac{1}{2} + \frac{1}{4}\pi)^{y_1}(\frac{1}{4} - \frac{1}{4}\pi)^{y_2}(\frac{1}{4} - \frac{1}{4}\pi)^{y_3}(\frac{1}{4}\pi)^{y_4}$ la vraisemblance de Y, où $K_Y = \binom{N}{y_1, y_2, y_3, y_4}$.

La log-vraisemblance de Y est :
$\ell\ell(Y;\pi) = \log(K_Y) + y_1 \log(\frac{1}{2} + \frac{1}{4}\pi) + y_2 \log(\frac{1}{4} - \frac{1}{4}\pi) + y_3 \log(\frac{1}{4} - \frac{1}{4}\pi) + y_4 \log(\frac{1}{4}\pi)$
En derivant par rapport à π on obtient :

$$\begin{aligned}\frac{\partial \ell\ell(Y;\pi)}{\partial \pi} &= y_1 \frac{1}{2+\pi} - y_2 \frac{1}{1-\pi} - y_3 \frac{1}{1-\pi} + y_4 \frac{1}{\pi} \\ &= \frac{y_1\pi(1-\pi) - (y_2+y_3)\pi(2+\pi) + y_4(2+\pi)(1-\pi)}{\pi(2+\pi)(1-\pi)} \\ &= \frac{-N\pi^2 + (y_1 - 2(y_2+y_3) - y_4)\pi + 2y_4}{\pi(2+\pi)(1-\pi)}\end{aligned}$$

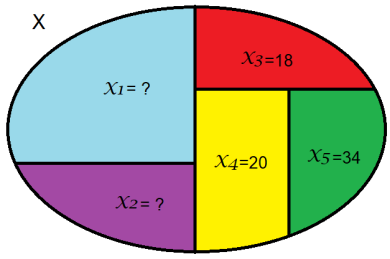
Soit l'équation de vraisemblance : $\frac{\partial \ell\ell(Y;\pi)}{\partial \pi} = 0 \iff -N\pi^2 + (y_1 - 2(y_2+y_3) - y_4)\pi + 2y_4 = 0$.
En remplaçant N,y_1, y_2, y_3 et y_4 par leur valeur numerique, on obtient l'équation : $-197\pi^2 - 15\pi + 68 = 0$, dont les racines sont $\pi_1 = \frac{15+\sqrt{53809}}{394}$ et $\pi_2 = \frac{15-\sqrt{53809}}{394}$. Or $\pi \in [0,1]$ et $\pi_2 < 0$ donc $\pi \neq \pi_2$. Ce qui implique $\widehat{\pi} = \pi_1$.

De plus : $\frac{\partial^2 \ell\ell(Y;\pi)}{\partial \pi^2} = -y_1 \frac{1}{(2+\pi)^2} - y_2 \frac{1}{(1-\pi)^2} - y_3 \frac{1}{(1-\pi)^2} - y_4 \frac{1}{\pi^2} < 0$
Donc $\widehat{\pi} = \frac{15+\sqrt{53809}}{394} \approx 0.63$ est l'unique estimateur du maximum de vraisemblance de Y.

Le nombre d'individu formant le premier groupe étant nettement supérieur à celui des autres classes, on suppose que le groupe 1 est la réunion de deux catégories d'individus distinctes.
Cette division de la population en 5 parties constitue les données complètes X telles que :

$$X = (x_1, x_2, x_3, x_4, x_5)$$

Avec $y_1 = x_1 + x_2$, $y_2 = x_3$, $y_3 = x_4$, $y_4 = x_5$.



On suppose que les données complètes suivent également une loi multinomiale dont la fonction de densité est la suivante :

$$f(X,\pi) = \binom{N}{x_1, x_2, x_3, x_4, x_5} (\frac{1}{2})^{x_1} (\frac{1}{4}\pi)^{x_2} (\frac{1}{4} - \frac{1}{4}\pi)^{x_3} (\frac{1}{4} - \frac{1}{4}\pi)^{x_4} (\frac{1}{4}\pi)^{x_5}$$

Notre objectif est de déterminer l'effectif des groupes x_1 et x_2.
On applique donc l'algorithme EM afin d'estimer la valeur du paramètre π maximisant la vraissemblance des données complètes et inobservables X sachant les données observées Y.

Notons $x_1^{(p)}$, $x_2^{(p)}$ et $\pi^{(p)}$ les estimations respectives des effectifs des classes x_1, x_2 et du paramètre π après p-itérations de l'algorithme.

Etape E :

Calculons la fonction $Q(\pi|\pi^{(p)})$.

$$\begin{aligned}Q(\pi|\pi^{(p)}) &= \mathbb{E}_{\pi^{(p)}}(\log(f(X,\pi))|Y) \\ &= \mathbb{E}_{\pi^{(p)}}(\log(K_X) + x_1 \log(\frac{1}{2}) + x_2 \log(\frac{1}{4}) + x_3(\frac{1}{4} - \frac{1}{4}\pi) + x_4 \log(\frac{1}{4} - \frac{1}{4}\pi) + x_5 \log(\frac{1}{4}\pi)|Y)\end{aligned}$$

Où $K_X = \binom{N}{x_1,x_2,x_3,x_4,x_5}$

Sachant Y, les valeurs K_X, x_3, x_4, x_5 sont des constantes.
Et par linéarité de l'espérance :

$$Q(\pi|\pi^{(p)}) = \log(K_X) + \log(\frac{1}{2})\mathbb{E}_{\pi^{(p)}}(x_1|Y) + \log(\frac{1}{4}\pi)\mathbb{E}_{\pi^{(p)}}(x_2|Y) + x_3\log(\frac{1}{4} - \frac{1}{4}\pi) + x_4\log(\frac{1}{4} - \frac{1}{4}\pi) + x_5\log(\frac{1}{4}\pi)$$

$\mathbb{E}_{\pi^{(p)}}(x_1|Y)$ et $\mathbb{E}_{\pi^{(p)}}(x_2|Y)$ étant les valeurs x_1 et x_2 estimées à la p-ième itération, c'est à dire :

$$x_1^{(p)} = \mathbb{E}_{\pi^{(p)}}(x_1|Y) \text{ et } x_2^{(p)} = \mathbb{E}_{\pi^{(p)}}(x_2|Y)$$

Pour déterminer la valeur de ces espérances, on détermine la loi de la variable conditionnelle $\{X|Y\}$. Pour cela, on calcule $k(X|Y,\pi)$ sa fonction de densité.

$$\begin{aligned} k(X|Y,\pi) &= \frac{f(X,Y,\pi)}{g(Y,\pi)} \\ &= \frac{f(X,\pi)}{g(Y,\pi)} \quad \text{car X} \subset \text{Y} \\ &= \frac{\binom{N}{x_1,x_2,x_3,x_4,x_5}(\frac{1}{2})^{x_1}(\frac{1}{4}\pi)^{x_2}(\frac{1}{4} - \frac{1}{4}\pi)^{x_3}(\frac{1}{4} - \frac{1}{4}\pi)^{x_4}(\frac{1}{4}\pi)^{x_5}}{\binom{N}{y_1,y_2,y_3,y_4}(\frac{1}{2} + \frac{1}{4}\pi)^{y_1}(\frac{1}{4} - \frac{1}{4}\pi)^{y_2}(\frac{1}{4} - \frac{1}{4}\pi)^{y_3}(\frac{1}{4}\pi)^{y_4}} \end{aligned}$$

Après simplification, on obtient :

$$\begin{aligned} k(X|Y,\pi) &= \frac{y_1!}{x_1!x_2!}\frac{(\frac{1}{2})^{x_1}(\frac{1}{4}\pi)^{x_2}}{(\frac{1}{2} + \frac{1}{4}\pi)^{y_1}} \\ &= \frac{y_1!}{x_1!(y_1 - x_1)!}\frac{(\frac{1}{2})^{x_1}(\frac{1}{4}\pi)^{y_1 - x_1}}{(\frac{1}{2} + \frac{1}{4}\pi)^{y_1}} \\ &= \binom{y_1}{x_1}(\frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\pi})^{x_1}(1 - \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\pi})^{y_1 - x_1} \end{aligned}$$

Donc {X|Y} suit la loi binomiale de paramètres $(y_1, \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\pi})$.
Ce qui implique $x_1^{(p)} = \mathbb{E}_{\pi^{(p)}}(x_1|Y) = y_1\frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\pi^{(p)}}$

Et $x_2^{(p)} = \mathbb{E}_{\pi^{(p)}}(x_2|Y) = \mathbb{E}_{\pi^{(p)}}(y_1 - x_1|Y) = \mathbb{E}_{\pi^{(p)}}(y_1|Y) - \mathbb{E}_{\pi^{(p)}}(x_1|Y) = y_1 - x_1^{(p)} = y_1\frac{\frac{1}{4}\pi^{(p)}}{\frac{1}{2} + \frac{1}{4}\pi^{(p)}}$

D'où :
$Q(\pi|\pi^{(p)}) = \log(K_X) + \log(\frac{1}{2})y_1\frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\pi^{(p)}} + \log(\frac{1}{4}\pi)y_1\frac{\frac{1}{4}\pi^{(p)}}{\frac{1}{2} + \frac{1}{4}\pi^{(p)}} + x_3\log(\frac{1}{4} - \frac{1}{4}\pi) + x_4\log(\frac{1}{4} - \frac{1}{4}\pi) + x_5\log(\frac{1}{4}\pi)$

Etape M :

Maximisons $Q(\pi|\pi^{(p)})$ en fonction de π, pour cela annulons sa dérivée par rapport à π.

$$\begin{aligned} \frac{\partial Q(\pi|\pi^{(p)})}{\partial\pi} = 0 &\iff x_2^{(p)}\frac{1}{\pi} - x_3\frac{1}{1-\pi} - x_4\frac{1}{1-\pi} + x_5\frac{1}{\pi} = 0 \\ &\iff \frac{1}{\pi}(x_2^{(p)} + x_5) = \frac{1}{1-\pi}(x_3 + x_4) \\ &\iff \pi = \frac{x_2^{(p)} + x_5}{x_2^{(p)} + x_3 + x_4 + x_5} \end{aligned}$$

On obtient $\pi^{(p+1)} = \frac{x_2^{(p)}+x_5}{x_2^{(p)}+x_3+x_4+x_5}$ l'estimateur de π après (p+1)-itérations.

Demontrons qu'il s'agit d'un maximum.

Démonstration.

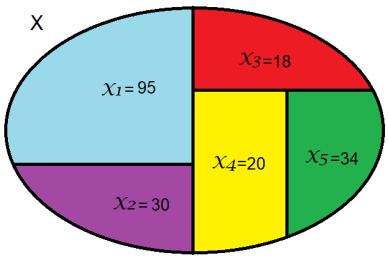
$$\begin{aligned}\frac{\partial Q(\pi|\pi^{(p)})}{\partial \pi^2} &= \frac{\partial}{\partial \pi}(x_2^{(p)}\frac{1}{\pi} - x_3\frac{1}{1-\pi} - x_4\frac{1}{1-\pi} + x_5\frac{1}{\pi}) \\ &= -x_2^{(p)}\frac{1}{\pi^2} - x_3\frac{1}{(1-\pi)^2} - x_4\frac{1}{(1-\pi)^2} - x_5\frac{1}{\pi^2} < 0\end{aligned}$$

Donc $\pi^{(p+1)}$ est la valeur de π qui maximise la fonction $Q(\pi|\pi^{(p)})$. □

Pour étudier la vitesse de convergence, on itère l'algorithme tant que l'écart entre la valeur de l'estimateur $\pi^{(p)}$ et la valeur théorique de l'estimateur du maximum de vraisemblance $\widehat{\pi}$ calculée précédemment est superieur à 10e-8 .
En choisissant l'initialisation $\pi^{(0)} = \frac{1}{2}$, on obtient les résultats présentés dans le tableau ci-dessous.

p	$\pi^{(p)}$	$\widehat{\pi} - \pi^{(p)}$	$x_1^{(p)}$	$x_2^{(p)}$
0	0.5	1.27e-1	100.0	25.0
1	0.6082474226804123	1.86e-2	95.8498023715415	29.1501976284585
2	0.6243210503692704	2.50e-3	95.26273470420941	29.7372652957905 93
3	0.6264888790796673	3.33e-4	95.18410757086511	29.81589242913489
4	0.6267773223473098	4.42e-5	95.17365551816091	29.826344481839087
5	0.6268156321100443	5.87e-6	95.17226749529516	29.82773250470484
6	0.6268207190193079	7.79e-7	95.17208319162889	29.827916808371114
7	0.6268213944559841	1.03e-7	95.17205871995539	29.827941280044612
8	0.6268214841396689	1.37e-8	95.1720554706364	29.827944529363606

La repartition des individus en classe obtenu à l'aide de l'algorithme EM est donc :

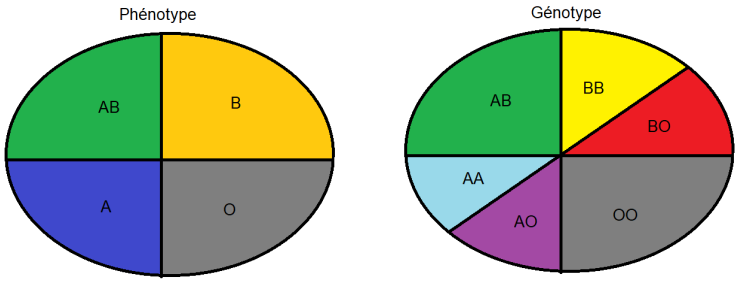


5 Exemple des groupes sanguins

Dans cet exemple, nous nous intéressons à un problème de génétique classique. Nous allons estimer les fréquences des allèles A,B et O sachant les phénotypes observés A, B, AB et O.

Tout d'abord, le phénotype d'un individu est l'ensemble des caractères apparents d'un individu. Alors que le génotype d'un individu est la composition allélique des gènes de l'individu. Découvert en 1900, le système ABO permet de classifier les différents groupes sanguins A, B, AB et O. Ces 4 groupes sanguins sont observables chez une personne. Or, par exemple, une personne du groupe A peut hériter d'un allèle A du parent 1 et d'un allèle A du parent 2 ou d'un allèle A du parent 1 et d'un allèle O du parent 2. Cette composition allélique n'est pas observable.

A l'aide de l'algorithme EM, nous allons estimer la proportion des 6 compositions alléliques possibles (AA, AO, BB, BO, AB, OO) dans la population en sachant la proportion des 4 groupes sanguins (A, B, AB, O).



Le tableau ci-dessous nous présente les différents génotypes ainsi que le phénotype correspondant et la fréquence de ces génotypes. Soit π_A la fréquence de l'allèle A, π_B la fréquence de l'allèle B et π_O la fréquence de l'allèle O. De plus, on a la contrainte $\pi_A + \pi_B + \pi_O = 1$.

Phénotype	Génotype	Fréquence du génotype
A	AA	π_A^2
A	AO	$2\,\pi_A\,\pi_O$
B	BB	π_B^2
B	BO	$2\,\pi_B\,\pi_O$
AB	AB	$2\,\pi_A\,\pi_B$
O	OO	π_O^2

Soit X l'échantillon des données complètes :

$$X = (x_{AA}, x_{AO}, x_{BB}, x_{BO}, x_{AB}, x_{OO})$$

Soit $n = x_{AA} + x_{AO} + x_{BB} + x_{BO} + x_{AB} + x_{OO}$ le nombre de personnes.

On suppose que X suit une loi multinomiale de paramètres n $\in \mathbb{N}$ et $\pi_A^2, 2\,\pi_A\,\pi_O, \pi_B^2, 2\,\pi_B\,\pi_O, 2\,\pi_A\,\pi_B, \pi_O^2 \in [0,1]^6$ avec $\pi_A^2 + 2\,\pi_A\,\pi_O + \pi_B^2 + 2\,\pi_B\,\pi_O + 2\,\pi_A\,\pi_B + \pi_O^2 = 1$.La fonction de densité des données complètes est :

$$f(X;\pi) = \binom{n}{x_{AA}, x_{AO}, x_{BB}, x_{BO}, x_{AB}, x_{OO}} (\pi_A^2)^{\,x_{AA}}\ (2\,\pi_A\,\pi_O)^{x_{AO}}\ (\pi_B^2)^{\,x_{BB}}\ (2\,\pi_B\,\pi_O)^{x_{BO}}\ (2\,\pi_A\,\pi_B)^{x_{AB}}\ (\pi_O^2)^{\,x_{OO}}$$

Soit Y l'échantillon des données observées :

$$Y = (y_A, y_B, y_{AB}, y_O) = (x_{AA} + x_{AO}, x_{BB} + x_{BO}, x_{AB}, x_{OO})$$

où y_i désigne le nombre de personnes du groupe sanguin i.

On suppose que Y suit une loi multinomiale de paramètres n $\in \mathbb{N}$ et $\pi_A^2 + 2\,\pi_A\,\pi_O, \pi_B^2 + 2\,\pi_B\,\pi_O, 2\,\pi_A\,\pi_B, \pi_O^2 \in [0,1]^4$ avec $\pi_A^2 + 2\,\pi_A\,\pi_O + \pi_B^2 + 2\,\pi_B\,\pi_O + 2\,\pi_A\,\pi_B + \pi_O^2 = 1$. La fonction de densité des données incomplètes est :

$$g(Y;\pi) = \binom{n}{y_A, y_B, y_{AB}, y_O} (\pi_A^2 + 2\,\pi_A\,\pi_O)^{y_A}\ (\pi_B^2 + 2\,\pi_B\,\pi_O)^{y_B}\ (2\,\pi_A\,\pi_B)^{y_{AB}}\ (\pi_O^2)^{\,y_O}$$

Etape E :

Calculons la fonction $Q(\pi|\pi^{(p)})$

Soit $K = \binom{n}{x_{AA},x_{AO},x_{BB},x_{BO},x_{AB},x_{OO}}$

On a :

$$\log f(X;\pi) = \log K + x_{AA}\log(\pi_A^2) + x_{AO}\log(2\,\pi_A\,\pi_O) + x_{BB}\log(\pi_B^2) + x_{BO}\log(2\,\pi_B\,\pi_O) + x_{AB}\log(2\,\pi_A\,\pi_B) + x_{OO}\log(\pi_O^2)$$

$$\begin{aligned}
Q(\pi|\pi^{(p)}) &= \mathbb{E}_{\pi^{(p)}}[\log\ f(X;\pi)\,|\,Y] \\
&= \mathbb{E}_{\pi^{(p)}}[\,\log\ K\ +\ x_{AA}\ \log(\pi_A^2) + \ x_{AO}\ \log(2\,\pi_A\,\pi_O)\ +\ x_{BB}\ \log(\pi_B^2)\ +\ x_{BO}\ \log(2\,\pi_B\,\pi_O) \\
&+\ x_{AB}\ \log(2\,\pi_A\,\pi_B)\ +\ x_{OO}\ \log(\pi_O^2)\,|\,Y] \\
&=\ \log\ K\ +\ \mathbb{E}_{\pi^{(p)}}[\,x_{AA}|\,x_{AA}\ +\ x_{AO}]\ \log(\pi_A^2)\ +\ \mathbb{E}_{\pi^{(p)}}[x_{AO}|\,x_{AA}\ +\ x_{AO}]\ \log(2\,\pi_A\,\pi_O) \\
&+\ \mathbb{E}_{\pi^{(p)}}[\,x_{BB}|\,x_{BB}\ +\ x_{BO}]\ \log(\pi_B^2)\ +\ \mathbb{E}_{\pi^{(p)}}[x_{BO}|\,x_{BB}\ +\ x_{BO}]\ \log(2\,\pi_B\,\pi_O) \\
&+\ x_{AB}\ \log(2\,\pi_A\,\pi_B)\ +\ x_{OO}\ \log(\pi_O^2)
\end{aligned}$$

Comme X suit une loi multinomiale, $\{x_{AA}|x_{AA}+x_{AO}\}$ suit une loi binomiale de paramètres $x_{AA}+x_{AO}$ et $\frac{\pi_A^2}{\pi_A^2+2\,\pi_A\,\pi_O}$ et $\{x_{BB}|x_{BB}+x_{BO}\}$ suit une loi binomiale de paramètres $x_{BB}+x_{BO}$ et $\frac{\pi_B^2}{\pi_B^2+2\,\pi_B\,\pi_O}$.

On a donc :

$$\begin{cases}
x_{AA}^{(p)} = \mathbb{E}_{\pi^{(p)}}[\,x_{AA}|\,x_{AA}\ +\ x_{AO}] = \ y_A\ \dfrac{(\pi_A^{(p)})^2}{(\pi_A^{(p)})^2\ +\ 2\,\pi_A^{(p)}\,\pi_O^{(p)}}\ =\ y_A\ \dfrac{\pi_A^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}} \\
x_{AO}^{(p)} = \mathbb{E}_{\pi^{(p)}}[\,x_{AO}|\,x_{AA}\ +\ x_{AO}] = \ y_A\ \dfrac{2\,\pi_A^{(p)}\,\pi_O^{(p)}}{(\pi_A^{(p)})^2\ +\ 2\,\pi_A^{(p)}\,\pi_O^{(p)}}\ =\ y_A\ \dfrac{2\,\pi_O^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}} \\
x_{BB}^{(p)} = \mathbb{E}_{\pi^{(p)}}[\,x_{BB}|\,x_{BB}\ +\ x_{BO}] = \ y_B\ \dfrac{(\pi_B^{(p)})^2}{(\pi_B^{(p)})^2\ +\ 2\,\pi_B^{(p)}\,\pi_O^{(p)}}\ =\ y_B\ \dfrac{\pi_B^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}} \\
x_{BO}^{(p)} = \mathbb{E}_{\pi^{(p)}}[\,x_{BO}|\,x_{BB}\ +\ x_{BO}] = \ y_B\ \dfrac{2\,\pi_B^{(p)}\,\pi_O^{(p)}}{(\pi_B^{(p)})^2\ +\ 2\,\pi_B^{(p)}\,\pi_O^{(p)}}\ =\ y_B\ \dfrac{2\,\pi_O^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}}
\end{cases}$$

Donc :

$$\begin{aligned}
Q(\pi|\pi^{(p)}) =&\ \log\ K\ +\ (x_{AA}\ +\ x_{AO})\ \frac{\pi_A^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(\pi_A^2)\ +\ (x_{AA}\ +\ x_{AO})\ \frac{2\,\pi_O^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(2\,\pi_A\,\pi_O) \\
&+\ (x_{BB}\ +\ x_{BO})\ \frac{\pi_B^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(\pi_B^2)\ +\ (x_{BB}\ +\ x_{BO})\ \frac{2\,\pi_O^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(2\,\pi_B\,\pi_O) \\
&+\ x_{AB}\ \log(2\,\pi_A\,\pi_B)\ +\ x_{OO}\ \log(\pi_O^2) \\
=&\ \log\ K\ +\ y_A\ \frac{\pi_A^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(\pi_A^2)\ +\ y_A\ \frac{2\,\pi_O^{(p)}}{\pi_A^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(2\,\pi_A\,\pi_O)\ +\ y_B\ \frac{\pi_B^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(\pi_B^2) \\
&+\ y_B\ \frac{2\,\pi_O^{(p)}}{\pi_B^{(p)}\ +\ 2\,\pi_O^{(p)}}\ \log(2\,\pi_B\,\pi_O)\ +\ y_{AB}\ \log(2\,\pi_A\,\pi_B)\ +\ y_O\ \log(\pi_O^2)
\end{aligned}$$

Etape M :

Maximisons $Q(\pi|\pi^{(p)})$ en fonction de $\pi = (\pi_A, \pi_B, \pi_O)$ et sous la contrainte $\pi_A + \pi_B + \pi_O = 1$.

Soit le lagrangien :

$$\begin{aligned}\mathcal{L} &= Q(\pi|\pi^{(p)}) \; - \lambda\,(\pi_A + \pi_B + \pi_O - 1) \\ &= \log K \; + \; x_{AA}^{(p)} \; \log(\pi_A^2) \; + \; x_{AO}^{(p)} \; \log(2\,\pi_A\,\pi_O) \; + x_{BB}^{(p)} \; \log(\pi_B^2) \; + \; x_{BO}^{(p)} \; \log(2\,\pi_B\,\pi_O) \\ &+ \; x_{AB} \; \log(2\,\pi_A\,\pi_B) \; + \; x_{OO} \; \log(\pi_O^2) - \lambda\,(\pi_A + \pi_B + \pi_O - 1)\end{aligned}$$

Déterminons les valeurs de π_A, π_B et π_O annulant les dérivées partielles du premier ordre du lagrangien.

On a :

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial\,\pi_A} = \dfrac{2\,x_{AA}^{(p)}}{\pi_A} \; + \; \dfrac{x_{AO}^{(p)}}{\pi_A} \; + \; \dfrac{x_{AB}}{\pi_A} \; - \; \lambda \; = 0 \\ \dfrac{\partial \mathcal{L}}{\partial\,\pi_B} = \dfrac{2\,x_{BB}^{(p)}}{\pi_B} \; + \; \dfrac{x_{BO}^{(p)}}{\pi_B} \; + \; \dfrac{x_{AB}}{\pi_B} \; - \; \lambda \; = 0 \\ \dfrac{\partial \mathcal{L}}{\partial\,\pi_O} = \dfrac{x_{AO}^{(p)}}{\pi_O} \; + \; \dfrac{x_{BO}^{(p)}}{\pi_O} \; + \; \dfrac{2\,x_{OO}}{\pi_O} \; - \; \lambda \; = 0 \\ \dfrac{\partial \mathcal{L}}{\partial\,\lambda} = \; \pi_A \; + \; \pi_B \; + \; \pi_O \; - \; 1 \; = 0 \end{cases}$$

On obtient :

$$\begin{cases} \lambda = \dfrac{2\,x_{AA}^{(p)}}{\pi_A} \; + \; \dfrac{x_{AO}^{(p)}}{\pi_A} \; + \; \dfrac{x_{AB}}{\pi_A} \; = \dfrac{2\,x_{BB}^{(p)}}{\pi_B} \; + \; \dfrac{x_{BO}^{(p)}}{\pi_B} \; + \; \dfrac{x_{AB}}{\pi_B} \; = \dfrac{x_{AO}^{(p)}}{\pi_O} \; + \; \dfrac{x_{BO}^{(p)}}{\pi_O} \; + \; \dfrac{2\,x_{OO}}{\pi_O} & \text{(1a)} \\ \pi_A \; + \; \pi_B \; + \; \pi_O \; = \; 1 & \text{(1b)} \end{cases}$$

$$
\begin{aligned}
(1a) \iff \pi_A &= \pi_O \frac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}} \\
\iff \pi_A &= (1 - \pi_A - \pi_B) \frac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}} \\
\iff \pi_A\,(1 + \frac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}}) &= (1 - \pi_B) \frac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}} \\
\iff \pi_A\,(2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)} + 2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}) &= (1 - \pi_B)\,(2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}) \\
\iff \pi_A\,(2\,x_{OO} + 2\,x_{AO}^{(p)} + 2\,x_{AA}^{(p)} + x_{BO}^{(p)} + x_{AB}) &= (1 - \pi_B)\,(2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}) \\
\text{Or} \quad \pi_B &= \pi_A \frac{2\,x_{BB}^{(p)} + x_{BO}^{(p)} + x_{AB}}{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}} \\
\iff \pi_A\,(2\,x_{OO} + 2\,x_{AO}^{(p)} + 2\,x_{AA}^{(p)} + x_{BO}^{(p)} + x_{AB}) &= 2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB} \\
&\quad - \pi_A\,(2\,x_{BB}^{(p)} + x_{BO}^{(p)} + x_{AB}) \\
\iff \pi_A\,(\,2\,x_{AO}^{(p)} + 2\,x_{AA}^{(p)} + 2\,x_{BB}^{(p)} + 2\,x_{BO}^{(p)} + 2\,x_{AB} + 2\,x_{OO}) &= 2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB} \\
\iff \pi_A\,2\,n &= 2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}
\end{aligned}
$$

Soient les estimations de π_A, π_B et π_O à l'itération p+1 :

$$
\begin{cases}
\pi_A^{(p+1)} = \dfrac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{2\,n} \\
\pi_B^{(p+1)} = \dfrac{2\,x_{BB}^{(p)} + x_{BO}^{(p)} + x_{AB}}{2\,n} \\
\pi_O^{(p+1)} = \dfrac{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}}{2\,n}
\end{cases}
$$

On vérifie que $\pi = (\pi_A, \pi_B, \pi_O)$ est bien un maximum. Soient les dérivées partielles du lagrangien du second ordre :

$$
\begin{cases}
\dfrac{\partial^2 \mathcal{L}}{\partial \pi_A^2} = -\dfrac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{\pi_A^2} \\
\dfrac{\partial^2 \mathcal{L}}{\partial \pi_B^2} = -\dfrac{2\,x_{BB}^{(p)} + x_{BO}^{(p)} + x_{AB}}{\pi_B^2} \\
\dfrac{\partial^2 \mathcal{L}}{\partial \pi_O^2} = -\dfrac{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}}{\pi_O^2} \\
\dfrac{\partial^2 \mathcal{L}}{\partial\,\pi_A\,\partial\,\pi_O} = \dfrac{\partial^2 \mathcal{L}}{\partial\,\pi_B\,\partial\,\pi_O} = \dfrac{\partial^2 \mathcal{L}}{\partial\,\pi_A\,\partial\,\pi_B} = 0
\end{cases}
$$

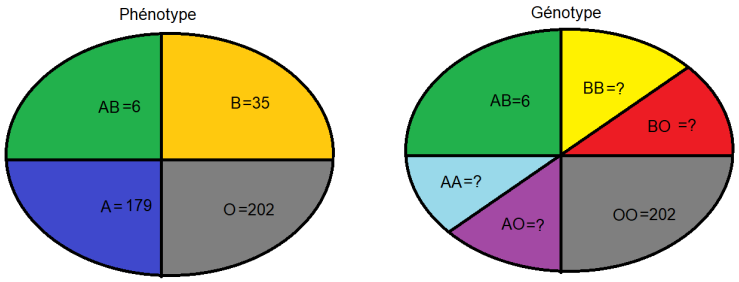
La matrice hessienne de $\mathcal{L}$ au point $(\pi_A^{(p+1)}, \pi_B^{(p+1)}, \pi_O^{(p+1)})$ est donnée par :

$$
H = \begin{pmatrix}
-\frac{2\,x_{AA}^{(p)} + x_{AO}^{(p)} + x_{AB}}{\pi_A^2} & 0 & 0 \\
0 & -\frac{2\,x_{BB}^{(p)} + x_{BO}^{(p)} + x_{AB}}{\pi_B^2} & 0 \\
0 & 0 & -\frac{2\,x_{OO} + x_{AO}^{(p)} + x_{BO}^{(p)}}{\pi_O^2}
\end{pmatrix} < 0
$$

$(\pi_A^{(p+1)}, \pi_B^{(p+1)}, \pi_O^{(p+1)})$ est donc un maximum local, de plus il est unique donc il est global.

Mise en application de l'algorithme EM

Nous allons mettre en application l'algorithme EM sur un échantillon de 422 personnes de données $Y = (179, 35, 6, 202)$.



L'initialisation de $\pi_A^{(0)}$, $\pi_B^{(0)}$ et $\pi_O^{(0)}$ joue un rôle important. En effet, une mauvaise initialisation ne permettra pas de converger vers π_A, π_B et π_O. On décide d'initialiser les valeurs initiales en prenant en compte les données observées. On initialise π_A, π_B et π_O de la manière suivante :

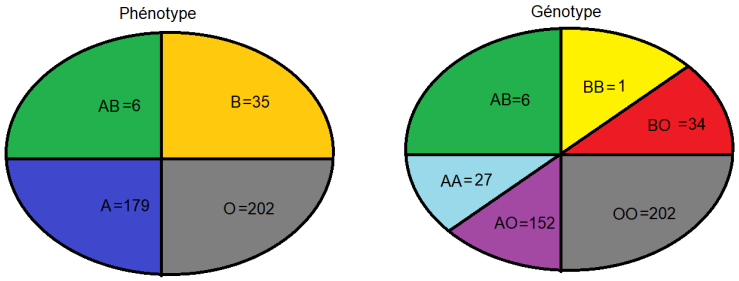
$$\begin{cases} \pi_A^{(0)} = \dfrac{y_A + \frac{y_{AB}}{2}}{n} \\ \pi_B^{(0)} = \dfrac{y_B + \frac{y_{AB}}{2}}{n} \\ \pi_O^{(0)} = 1 - \pi_A^{(0)} - \pi_B^{(0)} = \dfrac{n - y_A - y_B - y_{AB}}{n} = \dfrac{y_O}{n} \end{cases}$$

Par convergence de l'algorithme EM, on a lorsque p tend vers l'infini : $\pi* = \pi^{(p)} = \pi^{(p+1)}$. Dans cet exemple, on fixe un seuil de tolérance ϵ égale à 10e-6. On itère l'algorithme tant que :

$$\frac{\| \pi^{(p+1)} - \pi^{(p)} \|_\infty}{\| \pi^{(p)} \|_\infty} > \epsilon$$

Nous obtenons les résultats suivants :

Etape	$\frac{\| \pi^{(p+1)} - \pi^{(p)} \|_\infty}{\| \pi^{(p)} \|_\infty}$	$\pi_A^{(p)}$	$\pi_B^{(p)}$	$\pi_O^{(p)}$
0	2.089108910891089	0.4312796208530806	0.09004739336492891	0.47867298578199047
1	0.38464627121523326	0.2850638112029504	0.0521434239025541	0.6627927648944955
2	0.04575925764831685	0.25673063102275057	0.05014769918650647	0.693121669790743
3	0.006517825068568061 5	0.2523347008621683	0.05002598355215895	0.6976393155856728
4	0.0009626669474919246	0.25167552773014	0.05001356237380181	0.698310909896058
5	0.00014316113117100024	0.2515772332136072	0.05001188591056492	0.6984108808758278
6	2.131710142733683e-05	0.25156258963280526	0.05001164139578131	0.6984257689714134
7	3.174934644576525e-06	0.25156040844252847	0.05001160512988746	0.698427986427584



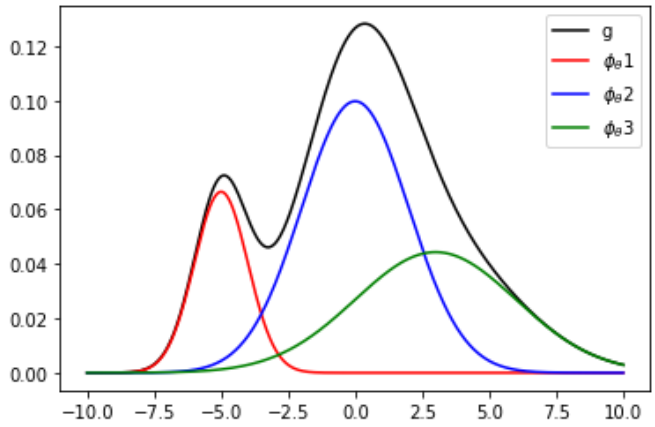
Troisième partie

Mélanges Gaussiens

Dans cette partie nous allons étudier $Y = (y_1, y_2, ..., y_n)$ un n-echantillon, n$\in \mathbb{N}^*$, de variables aléatoires indépendantes et identiquement distribuées selon une loi mélange de m-Gaussiennes, m$\in \mathbb{N}^*$, c'est à dire dont la fonction de densité est de la forme :

$$g_Y(y, \theta) = \sum_{k=1}^{m} \pi_k \phi_{\theta_k}(y)$$

Avec $\phi_{\theta_k}(y) = \frac{1}{\sqrt{(2\pi\sigma_k^2)}} \exp\left(\frac{-(y-\mu_k)^2}{2\sigma_k^2}\right)$, où $\theta_k = (\mu_k, \sigma_k^2)$, $\theta_k \in \mathbb{R} \times \mathbb{R}_+^*, \forall k \in \{1,...,m\}$.
Dans la suite nous utiliserons ϕ_{θ_k} pour désigner la fonction de densité d'une loi Gaussienne de paramètre θ_k.
Et $\pi_k \in [0,1]$ la proportion de la k-ième fonction de densité dans les mélanges, $\forall k \in \{1,...,m\}$ et tel que $\sum_{k=1}^{m} \pi_k = 1$.
Le paramètre θ est un vecteur composé de 3m-1 paramètres, $\theta = (\pi_1, ..., \pi_{m-1}, \mu_1, ..., \mu_m, \sigma_1^2, ..., \sigma_m^2)$.



Sur le graphiqueci-dessus on $g_Y(y, \theta) = \frac{1}{6}\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y+5)^2}{2}\right) + \frac{1}{2}\frac{1}{\sqrt{8\pi}} \exp\left(\frac{-y^2}{8}\right) + \frac{1}{3}\frac{1}{\sqrt{18\pi}} \exp\left(\frac{-(y-3)^2}{18}\right)$
C'est à dire : $\pi_1 = \frac{1}{6}$, $\pi_2 = \frac{1}{2}$,$\pi_3 = \frac{1}{3}$, $\theta_1 = (-5, 1)$, $\theta_2 = (0, 2^2)$ et $\theta_3 = (3, 3^2)$, soit $\theta = (\frac{1}{6}, \frac{1}{2}, -5, 0, 3, 1, 2^2, 3^2)$.

6 Calcul des estimateurs

Soit $Z = (z_1, z_2, ..., z_n)$ un n-echantillon de variables aléatoires indépendantes et identiquement distribuées selon la loi donnée par le vecteur $\pi = (\pi_1, ..., \pi_m)$, où $z_i \in [1, ...m]$ correspond au groupe d'appartenance du i-ème individu de Y, $\forall i \in \{1, ..., n\}$.
Dans le cadre des mélanges gaussiens, les données complètes sont contenues dans le vecteur $X = (x_1, x_2, ..., x_n)$ où $x_i = (y_i, z_i)$ et les données observées sont représentées par le vecteur Y. Nous utiliserons ici l'algorithme EM afin de déterminer la valeur des paramètres de chaque gaussienne de façon à maximiser l'espérance de la vraisemblance de X sachant Y.

La vraisemblance du modèle complet, notée h, est la suivante :

$$h(X) = \prod_{i=1}^{n} \sum_{j=1}^{m} \pi_j \phi_{\theta_j}(y_i) \mathbb{1}_{\{z_i=j\}}$$

On en déduit la log-vraisemblance.

$$\begin{aligned}\log(h(X)) &= \log(\prod_{i=1}^{n} \sum_{j=1}^{m} \pi_j \phi_{\theta_j}(y_i) \mathbb{1}_{\{z_i=j\}}) \\ &= \sum_{i=1}^{n} \log(\sum_{j=1}^{m} \pi_j \phi_{\theta_j}(y_i) \mathbb{1}_{\{z_i=j\}})\end{aligned}$$

Les termes de la somme étant nuls si z_i est différent de j, on a :

$$\begin{aligned}\log(h(X)) &= \sum_{i=1}^{n} \sum_{j=1}^{m} \log(\pi_j \phi_{\theta_j}(y_i)) \mathbb{1}_{\{z_i=j\}} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2}) \mathbb{1}_{\{z_i=j\}}\end{aligned}$$

Etape E

Ceci nous permet d'obtenir la fonction Q à calculer lors de l'étape E de l'algorithme.

$$\begin{aligned}Q(\theta|\theta^{(p)}) &= \mathbb{E}_{\theta^{(p)}}(\log(h(X)|Y)) \\ &= \mathbb{E}_{\theta^{(p)}}(\sum_{i=1}^{n} \sum_{j=1}^{m} (\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})) \mathbb{1}_{\{z_i=j|Y\}}) \\ &= (\sum_{i=1}^{n} \sum_{j=1}^{m} (\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})) \mathbb{E}_{\theta^{(p)}}(\mathbb{1}_{\{z_i=j|Y\}}) \\ &= (\sum_{i=1}^{n} \sum_{j=1}^{m} (\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})) \mathbb{P}_{\theta^{(p)}}(z_i = j|Y)\end{aligned}$$

Or : $\mathbb{P}_{\theta^{(p)}}(z_i = j|Y) = \frac{\pi_j^{(p)} \phi_{\theta_j^{(p)}}(y_i)}{\sum_{k=1}^{m} \pi_k^{(p)} \phi_{\theta_k^{(p)}}(y_i)}$, par application de la formule de Bayes.

Donc : $Q(\theta|\theta^{(p)}) = (\sum_{i=1}^{n} \sum_{j=1}^{m} (\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(Y_i-\mu_j)^2}{2\sigma_j^2})) \frac{\pi_j^{(p)} \phi_{\theta_j^{(p)}}(y_i)}{\sum_{k=1}^{m} \pi_k^{(p)} \phi_{\theta_k^{(p)}}(y_i)}$

Dans la suite nous noterons : $f_j^{(p)}(y_i) = \frac{\pi_j^{(p)} \phi_{\theta_j^{(p)}}(y_i)}{\sum_{k=1}^{m} \pi_k^{(p)} \phi_{\theta_k^{(p)}}(y_i)}$

Etape M

L'étape M consiste à calculer la valeur du paramètre θ maximisant la fonction $Q(\theta|\theta^{(p)})$ sous la contrainte $\sum_{k=1}^{m} \pi_k = 1$.
Pour cela introduisons le Lagrangien : $\mathcal{L}(\theta, \lambda) = Q(\theta|\theta^{(p)}) + \lambda(\sum_{k=1}^{m} \pi_k - 1)$

Estimation des π_k

Soit d∈{1,...,m}, afin d'obtenir l'estimation de π_d après (p+1)-iterations annulons la dérivée partielle du Lagrangien par rapport à cette variable.

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \pi_d} &= \frac{\partial Q(\theta|\theta^{(p)})}{\partial \pi_d} + \lambda \\
&= \frac{\partial}{\partial \pi_d}(\sum_{i=1}^{n}(\sum_{j=1}^{m}(\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) + \lambda \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \pi_d}(\sum_{j=1}^{m}(\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) + \lambda \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \pi_d}(\log(\pi_d)f_d^{(p)}(y_i))) + \lambda \\
&= \frac{1}{\pi_d}\sum_{i=1}^{n}(f_d^{(p)}(y_i)) + \lambda
\end{aligned}
$$

En annulant cette dérivée on obtient :

$\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \pi_d} = 0 \iff \frac{1}{\pi_d}\sum_{i=1}^{n}(f_d^{(p)}(y_i)) + \lambda = 0 \iff \pi_d^{(p+1)} = -\frac{\sum_{i=1}^{n}(f_d^{(p)}(y_i))}{\lambda}$

Puis par application de la contrainte sur la somme des π_k, on calcul la valeur de lambda :

$$
\begin{aligned}
\sum_{d=1}^{m}\pi_j^{(p+1)} = 1 &\iff \sum_{d=1}^{m} -\frac{\sum_{i=1}^{n} f_d^{(p)}(y_i)}{\lambda} = 1 \\
&\iff \lambda = -\sum_{d=1}^{m}\sum_{i=1}^{n} f_d^{(p)}(y_i) \\
&\iff \lambda = -\sum_{d=1}^{m}\sum_{i=1}^{n} \frac{\pi_d^{(p)}\phi_{\theta_d^{(p)}}(y_i)}{\sum_{k=1}^{m}\pi_k^{(p)}\phi_{\theta_k^{(p)}}(y_i)} \\
&\iff \lambda = -\sum_{i=1}^{n}\sum_{d=1}^{m} \frac{\pi_d^{(p)}\phi_{\theta_d^{(p)}}(y_i)}{\sum_{k=1}^{m}\pi_k^{(p)}\phi_{\theta_k^{(p)}}(y_i)} \\
&\iff \lambda = -\sum_{i=1}^{n}(\frac{\sum_{d=1}^{m}\pi_d^{(p)}\phi_{\theta_d^{(p)}}(y_i)}{\sum_{k=1}^{m}\pi_k^{(p)}\phi_{\theta_k^{(p)}}(y_i)}) \\
&\iff \lambda = -n
\end{aligned}
$$

La valeur de $\pi_d^{(p+1)}$ obtenue à l'étape M après (p+1)-itérations de l'algorithme est :

$$\pi_d^{(p+1)} = \frac{\sum_{i=1}^{n} f_d^{(p)}(y_i)}{n}$$

Estimation des μ_k

A l'aide d'un raisonnement similaire, on déduit $\mu_d^{(p+1)} = \frac{\sum_{i=1}^{n} y_i f_d^{(p)}(y_i)}{\sum_{i=1}^{n} f_d^{(p)}(y_i)}$

Démonstration.

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \mu_d} &= \frac{\partial}{\partial \mu_d}(\sum_{i=1}^{n}(\sum_{j=1}^{m}(\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \mu_d}(\sum_{j=1}^{m}(\ln(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \mu_d}(\frac{-(y_i-\mu_d)^2}{2\sigma_d^2}f_d^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2}f_d^{(p)}(y_i)
\end{aligned}$$

Puis en annulant la dérivée :

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \mu_d} = 0 &\iff \sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2}f_d^{(p)}(y_i) = 0 \\
&\iff \mu_d^{(p+1)} = \frac{\sum_{i=1}^{n} y_i f_d^{(p)}(y_i)}{\sum_{i=1}^{n} f_d^{(p)}(y_i)}
\end{aligned}$$

□

Estimation des σ_k^2.

L'estimation de $\sigma_d^{2^{(p+1)}}$ est donnée par la formule : $\sigma_d^{2^{(p+1)}} = \frac{\sum_{i=1}^{n}(y_i-\mu_d)^2 f_d^{(p)}(y_i)}{\sum_{i=1}^{n} f_d^{(p)}(y_i)}$

Démonstration.

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \sigma_d^2} &= \frac{\partial}{\partial \sigma_d^2}(\sum_{i=1}^{n}(\sum_{j=1}^{m}(\ln(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{-(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \sigma_d^2}(\sum_{j=1}^{m}(\log(\pi_j) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(y_i-\mu_j)^2}{2\sigma_j^2})f_j^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}(\frac{\partial}{\partial \sigma_d^2}((-\frac{1}{2}\ln(\sigma_d^2) - \frac{(y_i-\mu_d)^2}{2\sigma_d^2})f_d^{(p)}(y_i))) \\
&= \sum_{i=1}^{n}(-\frac{1}{2\sigma_d^2} + \frac{2(y_i-\mu_d)^2}{(2\sigma_d^2)^2})f_d^{(p)}(y_i) \\
&= \frac{1}{2\sigma_d^2}\sum_{i=1}^{n}(-1 + \frac{(y_i-\mu_d)^2}{\sigma_d^2})f_d^{(p)}(y_i)
\end{aligned}$$

En résolvant l'équation suivante on a :

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \sigma_d^2} = 0 &\iff \sum_{i=1}^{n}(-1 + \frac{(y_i-\mu_d)^2}{\sigma_d^2})f_d^{(p)}(y_i) = 0 \\
&\iff \sigma_d^{2^{(p+1)}} = \frac{\sum_{i=1}^{n}(y_i-\mu_d)^2 f_d^{(p)}(y_i)}{\sum_{i=1}^{n} f_d^{(p)}(y_i)}
\end{aligned}$$

□

Ce qui nous permet d'obtenir $\theta^{(p+1)} = (\pi_1^{(p+1)}, \ldots, \pi_{m-1}^{(p+1)}, \mu_1^{(p+1)}, \ldots, \mu_m^{(p+1)}, \sigma_1^{2(p+1)}, \ldots, \sigma_m^{2(p+1)})$.
Prouvons qu'il s'agit bien d'un estimateur du maximum de vraisemblance :

Démonstration.
$\forall (b,d) \in \{1,...m\}^2, b \neq d$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \pi_d^2} = \frac{\partial}{\partial \pi_d}(\frac{1}{\pi_d}\sum_{i=1}^{n}(f_d^{(p)}(y_i)) - n) = -\frac{1}{\pi_d^2}\sum_{i=1}^{n} f_d^{(p)}(y_i) < 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \pi_d \partial \pi_b} = \frac{\partial^2 L(\theta,\lambda)}{\partial \pi_d \partial \mu_b} = \frac{\partial^2 L(\theta,\lambda)}{\partial \pi_d \partial \sigma_b^2} = 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \mu_d^2} = \frac{\partial}{\partial \mu_d}\sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2} f_d^{(p)}(y_i)) = \frac{-1}{\sigma_d^2}\sum_{i=1}^{n} f_d^{(p)}(y_i) < 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \mu_b \partial \mu_d} = \frac{\partial}{\partial \mu_b}\sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2} f_d^{(p)}(y_i) = 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \sigma_b^2 \partial \mu_d} = \frac{\partial}{\partial \sigma_b^2}\sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2} f_d^{(p)}(y_i) = 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \sigma_b^2 \partial \sigma_d^2} = \frac{\partial}{\partial \sigma_b^2}(\frac{1}{2\sigma_d^2}\sum_{i=1}^{n}(-1+\frac{(y_i-\mu_d)^2}{\sigma_d^2}) f_d^{(p)}(y_i)) = 0$$

$$\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial \sigma_d^2 \partial \mu_d} = \frac{\partial}{\sigma_d^2}\sum_{i=1}^{n}\frac{(y_i-\mu_d)}{\sigma_d^2} f_d^{(p)}(y_i) = -\frac{1}{\sigma_d^2}\sum_{i=1}^{n}\frac{(y_i-\mu_d) f_d^{(p)}(y_i)}{\sigma_d^2}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\theta,\lambda)}{\partial (\sigma_d^2)^2} &= \frac{\partial}{\partial \sigma_d^2}(\frac{1}{2\sigma_d^2}\sum_{i=1}^{n}(-1+\frac{(y_i-\mu_d)^2}{\sigma_d^2}) f_d^{(p)}(y_i)) \\
&= \frac{-1}{(\sigma_d^2)^2}\sum_{i=1}^{n}(-\frac{1}{2}+\frac{(y_i-\mu_d)^2}{\sigma_d^2}) f_d^{(p)}(y_i)) \\
&= \frac{-1}{(\sigma_d^2)^2}\sum_{i=1}^{n}((-1+\frac{(y_i-\mu_d)^2}{\sigma_d^2}) f_d^{(p)}(y_i) + \frac{1}{2} f_d^{(p)}(y_i)) \\
&= \frac{-1}{(\sigma_d^2)^2}\sum_{i=1}^{n}(-1+\frac{(y_i-\mu_d)^2}{\sigma_d^2}) f_d^{(p)}(y_i)) - \frac{1}{2(\sigma_d^2)^2}\sum_{i=1}^{n} f_d^{(p)}(y_i)
\end{aligned}$$

En $\theta = \theta^{(p+1)}$: $\sum_{i=1}^{n}(-1+\frac{(y_i-\mu_d^{(p)})^2}{\sigma_d^{2(p)}}) f_d^{(p)}(y_i)) = 0$ et $\sum_{i=1}^{n}\frac{(y_i-\mu_d) f_d^{(p)}(y_i)}{\sigma_d^2} = 0$ par construction des estimateurs.

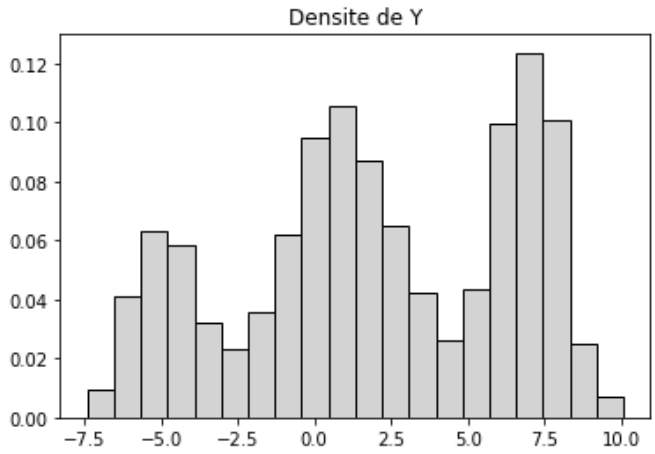
Donc : $\frac{\partial^2 \mathcal{L}(\theta^{(p)},\lambda)}{\partial \sigma_d^2 \partial \mu_d} = 0$ et $\frac{\partial^2 \mathcal{L}(\theta^{(p)},\lambda)}{\partial (\sigma_d^2)^2} = -\frac{1}{2(\sigma_d^2)^2}\sum_{i=1}^{n} f_d^{(p)}(y_i) < 0$.
La matrice hessienne de $\mathcal{L}$ évaluée en $\theta = \theta^{(p+1)}$ est définie négative.
La valeur $\theta^{(p+1)}$ est donc bien la valeur de theta qui maximise $Q(\phi|\phi^{(p+1)})$ sous la contrainte $\sum_{k=1}^{m}\pi_k = 1$. □

7 Application de l'algorithme

Afin de tester l'algorithme EM, nous simulons un échantillon Y de mille variables indépendantes et identiquement distribuées selon une loi mélange composée de trois gaussiennes dont les paramètres sont :

k	1	2	3
π	1/6	1/2	1/3
μ	-5	1	7
σ^2	1	4	1

L'histogramme de Y est alors :



Appliquons maintenant l'algorithme EM afin d'obtenir une estimation de la fonction de densité $g_Y(.,\theta)$. Il nous faut d'abord définir un critère d'arrêt de l'algorithme ainsi que les conditions concernant l'initialisation des paramètres $\pi^{(0)}$, $\mu^{(0)}$, $\sigma^{2^{(0)}}$.

7.1 Condition d'arrêt

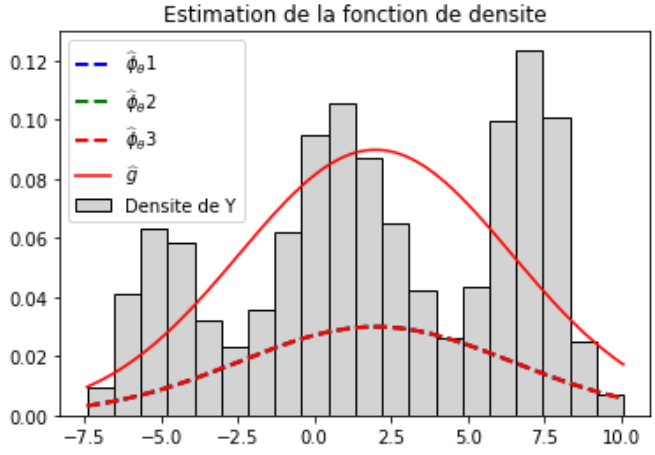
Les valeurs des paramètres estimés convergent donc nous utiliserons un critère de stabilité comme condition d'arrêt.
Fixons au préalable un seuil de tolérance noté ϵ, égale a 1.10e-3 dans l'exemple, puis on itère l'algorithme tant que :

$$\frac{\|\pi^{(p+1)}-\pi^{(p)}\|_\infty}{\|\pi^{(p)}\|_\infty} + \frac{\|\mu^{(p+1)}-\mu^{(p)}\|_\infty}{\|\mu^{(p)}\|_\infty} + \frac{\|\sigma^{2^{(p+1)}}-\sigma^{2^{(p)}}\|_\infty}{\|\sigma^{2^{(p)}}\|_\infty} > \epsilon$$

On renormalise chaque norme de différence par la norme infinie du paramètre afin de tenir compte des différents ordres de grandeur car $\pi_k \in [0,1]$, $\mu_k \in \mathbb{R}$ et $\sigma_k^2 \in \mathbb{R}_+^*$.

7.2 Initialisation des paramètres

L'initialisation des paramètres π, μ et σ^2 est importante afin d'assurer la convergence vers les valeurs estimées.
Si l'on effectue l'initialisation suivante : $\pi_k^{(0)} = \frac{1}{m}$, $\mu_k^{(0)} = 0$ et $\sigma_k^{2^{(0)}} = 1$, $\forall k \in \{1,...,m\}$.
Les valeurs des π_k, μ_k et σ_k resteront toujours égales au cours des itérations, donc les fonctions $\phi_{\widehat{\theta_k}}$ seront égales.
Graphiquement, on obtient :

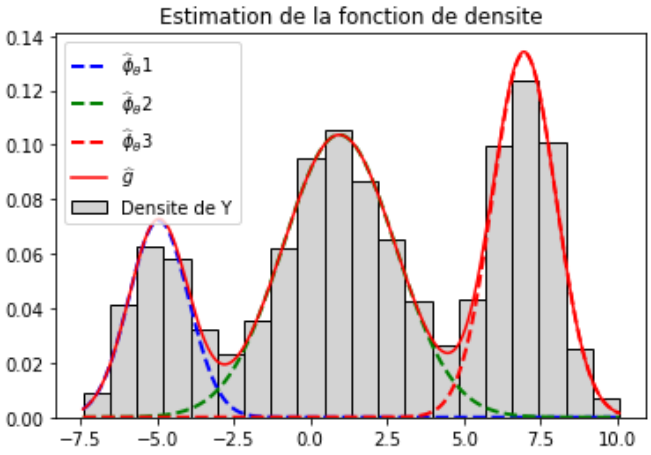


Les estimations des paramètres arrondies à 10e-2 près sont les suivantes :

k	1	2	3
π	0.33	0.33	0.33
μ	1.99	1.99	1.99
σ^2	19.74	19.74	19.74

Pour initialiser les valeurs des μ_k, nous trions Y puis nous le divisons en m groupes (avec m=3 dans cette exemple). Nous prenons ensuite pour valeur de $\mu_k^{(0)}$ la moyenne du k-ième groupe de l'échantillon trié. En appliquant cette méthode et en conservant les valeurs initiales de $\pi^{(0)}$ et $\sigma^{(0)}$, nous obtenons les résultats ci-dessous, après 62 itérations.

k	1	2	3
π	0.18	0.48	0.35
μ	-4.94	0.94	6.97
σ^2	0.95	3.36	1.07



8 Classification des individus

Les estimateurs obtenus à l'aide de l'algorithme nous permettent d'évaluer pour chaque individu i sa probabilité d'appartenance à chacun des m-groupes conditionnellement à la donnée observée y_i.
Nous savons que $\mathbb{P}(z_i = j|Y) = \frac{\pi_j \phi_{\theta_j}(y_i)}{\sum_{k=1}^{m} \pi_k \phi_{\theta_k}(y_i)}$.

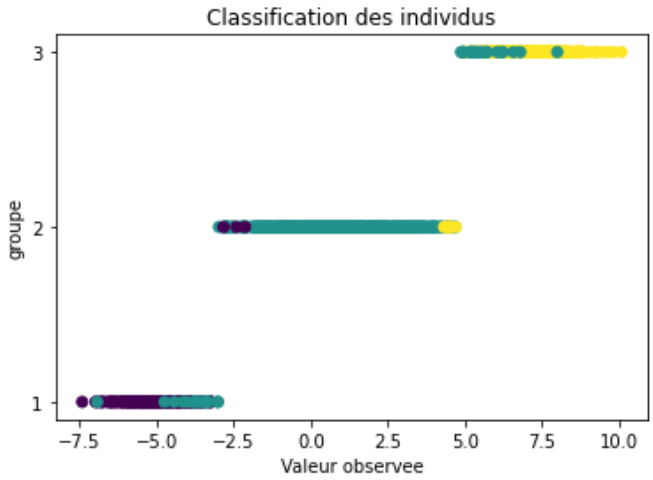
L'estimation de z_i sera alors donnée par k, avec k tel que :

$$k = max_{j\in\{1,\dots,m\}}(\mathbb{P}(z_i = j|Y)) = max_{j\in\{1,\dots,m\}}(\pi_j \phi_{\theta_j}(y_i)).$$

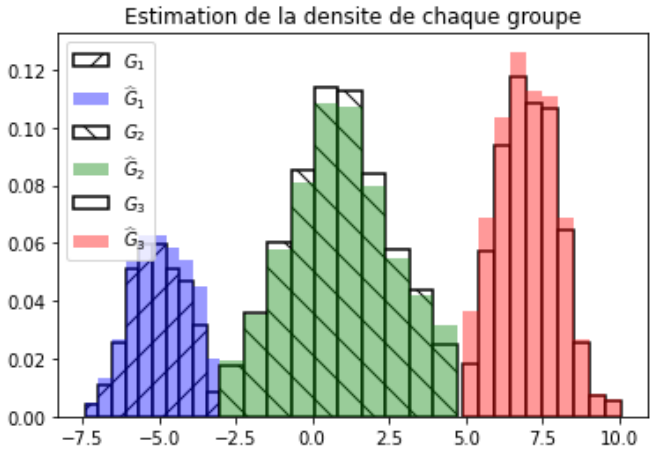
En comparant le groupe d'appartenance estimé et le groupe d'appartenance réel dans l'exemple précédent, on obtient la matrice de confusion suivante :

	$\widehat{G_1}$	$\widehat{G_2}$	$\widehat{G_3}$
G_1	164	3	0
G_2	15	463	20
G_3	0	6	329

Ici le taux d'erreur de classification est de 0.044.
Le nuage de point ci-dessus represente les individus avec en ordonnée le groupe d'appartenance réel, en abscisse la valeur observée et la couleur du point represente la classe estimée.



L'histogramme des densités pondérées par le poids π de chacun des m-groupes estimés est le suivant :



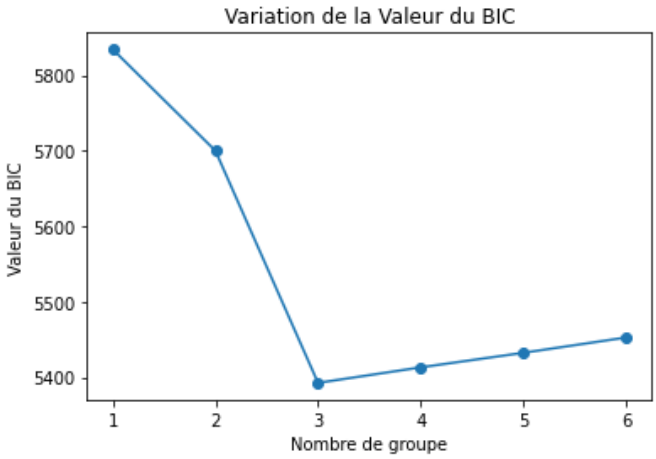
9 Choix du nombre de groupes

Dans les sections précédentes, nous estimions les valeurs des paramètres en supposant connu le nombre de groupes.
Mais dans la pratique ce nombre n'est pas toujours renseigné.
Afin de choisir et déterminer le nombre de groupe modélisant au mieux les données observées, nous appliquons la minimisation du critère BIC (Bayesian Information Criterion).
Ce critère se calcule avec la formule :

$$BIC = -2\ln(h(X)) + (3m-1)\ln(n)$$

Où h(X) correspond à la vraisemblance des données complètes, le facteur (3m-1) est dû au nombre de degré de liberté du paramètre θ que nous cherchons à estimer, et n correspond au nombre d'individus dans l'échantillon. La minimisation de ce critère permet de maximiser la vraisemblance en fonction du nombre de groupe tout en prenant en compte la taille de l'échantillon.

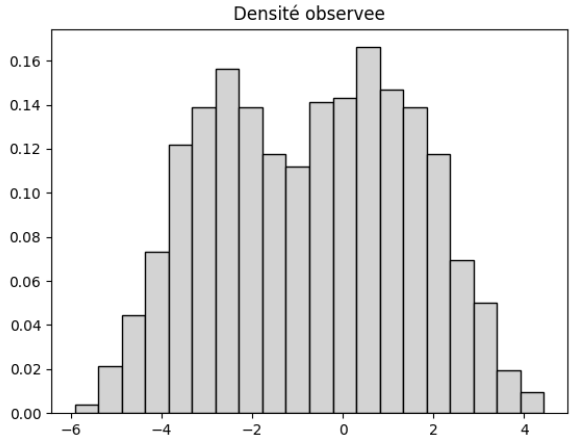
Appliquons ce critère dans le cadre de l'exemple abordé dans la parite précédente et étudions la valeur du BIC lorsque le nombre de groupe retenu dans le modèle varie entre 1 et 6.



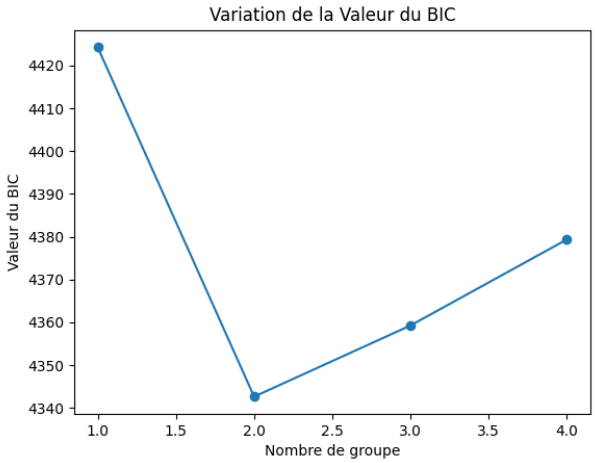
La valeur du critère BIC est minimale pour un modèle avec trois groupes, ce qui correspond au modèle original. Cela est notamment expliqué par la forte distinction entre les groupes. Plaçons nous désormais dans le cadre où la différentiation des classes est moins évidente.
Nous simulons cette fois un échantillon de taille 1000 avec les paramètres suivantes :

k	1	2	3
π	$1/3$	$1/2$	$1/6$
μ	-3	0	2
σ^2	1	2	1

L'histogramme des données observées est :



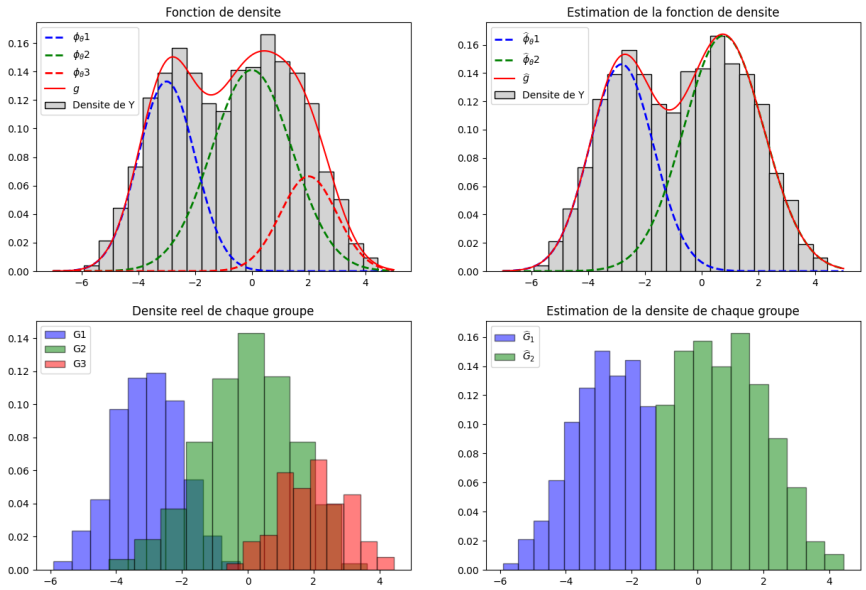
En appliquant le critère BIC à ces données, nous obtenons les résultats suivants.



La proximité des observations parmi les individus des différents groupes ne permet pas de détecter qu'ils forment trois groupes distincts à l'aide du critère BIC.
L'algorithme EM va donc estimer que les données sont distribuées selon une loi de densité mélanges composée de deux gaussiens dont les paramètres sont :

k	1	2
π	0.41	0.59
μ	-2.82	0.78
σ^2	1.27	1.98

Les fonctions de densité et la densité des groupes estimés sont alors :

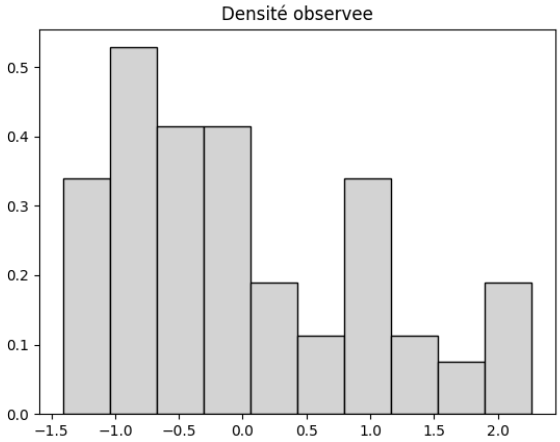


10 Application sur des données réelles

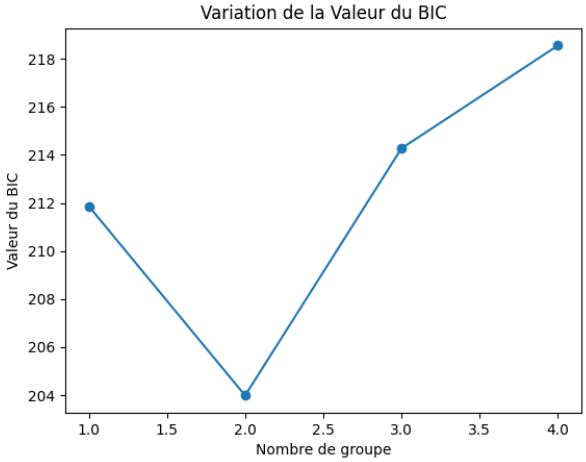
Dans cette partie, nous étudierons le jeu de données Golub qui contient l'expression de 3751 gènes mesurés chez 72 patients atteints de Leucémie.
Parmi les patients, 25 sont atteints d'une leucémie aigüe myéloblastique 'AML' et 47 d'une leucémie aiguë lympho-

blastique ‘ALL’.
Nous allons ici appliquer l’algorithme EM afin de déterminer par quelle type de leucémie est atteint le patient en connaissance uniquement l’expression d’un gène donné.

Nous prenons ici comme échantillon Y l’expression du gènes d’indice 955 dans la base de données.
Son histogramme en densité est le suivant :



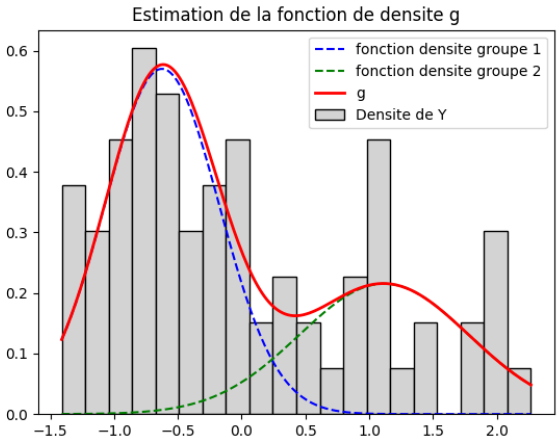
On observe ici l’expression du gène semble suivre une loi mélange composée de deux fonctions gaussiennes.
Appliquons la minimisation du critère BIC pour s’assurer de choisir un modèle avec deux gaussiennes qui maximise la vraisemblance de des données observées.



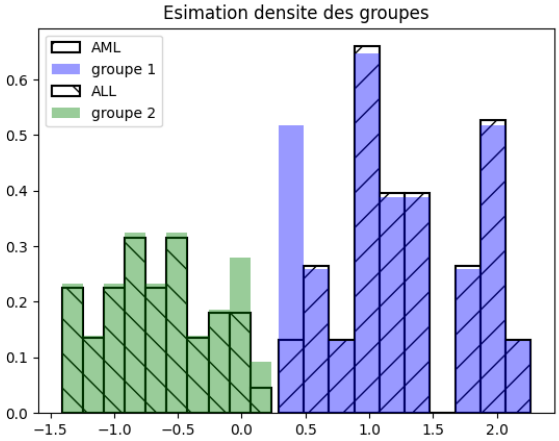
La valeur du BIC est minimale pour m=2, ce qui est cohérent avec les observations qui sont constitués d’individus de deux populations distinctes.
Les paramètres du modèle retenu sont les suivants :

k	1	2
π	0.64	0.36
μ	-0.63	1.11
σ^2	0.22	0.44

Graphiquement on obtient :

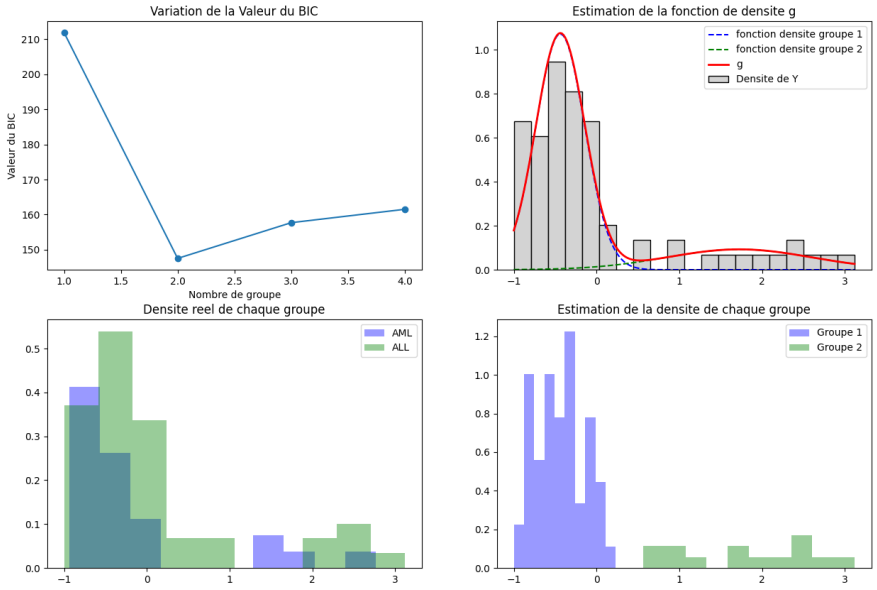


A l'aide de ces paramètres, nous pouvons calculer pour chaque individu la probabilité qu'il appartienne au groupe 1 ou groupe 2.
Comparons la densité des groupes estimés avec la densité des réels des groupes d'individus atteint de leucémie ALL et AML :



Les paramètres estimés nous permettent d'obtenir une estimation du type de leucémie dont est atteint l'individu en fonction de l'intensité de l'expression du gène numéro 955.

En revanche, l'expression de certains gènes ne permet pas à l'aide de notre algorithme d'estimer correctement de quelle leucémie est atteint un patient.
Prenons par exemple le gène d'indice 0 pour lequel nous avons les résultats ci-dessous :



La minimisation du BIC nous conduit bien à choisir un modèle à deux composantes dont les paramètres fournissent une bonne approximation de la fonction de densité g.
Cependant, la classification des individus en fonction des données calculées n'est pas bonne, ce qui est liée au faite que le gène ne permet pas une discrimination correcte des deux types de leucémie.

Références

[1] Dempster A.P., Laird N.M, Rubin D.B, *Maximum likehood from incomplete data via the EM algorithm.* J. Roy. Stat. Soc. B. 39-1 (1977), pp 1-38

[2] Jean-Louis Foulley, *Algorithme EM : théorie et application au modèle mixte.* Journal de la société française de statistique, tome 143, no 3-4 (2002), p. 57-109

[3] Isabelle Michaud, *Application de l'algorithme EM au modèle des risques concurrents avec causes de pannes masquées* Mémoire, août 2005

[4] Danho Djrobie, *Modèle de mèlange et classfication* Mémoire, 2016

[5] *Modèle de mélanges gaussiens - Définition et Explications*, Techno-Science.net, URL : https ://www.techno-science.net/definition/6349.html

[6] *Modèle de mélange gaussien*, Wikipedia, URL : https ://fr.wikipedia.org/wiki/Modèle_de_mèlange_gaussien

[7] *Leukemia data*, web stanford, URL : https ://web.stanford.edu/h̃astie/CASI_files/DATA/leukemia.html