Projet Master 1 SSD

Un modèle pour les nids de mouettes

Rédigé par

CARVAILLO Thomas

CÔME Olivier

PRALON Nicolas

Encadrante : Elodie Brunel-Piccinini

Table des matières

Introduction **2**

1 Modélisation du problème **3**
1.1 Modélisation du problème . . . 3
1.2 Une histoire de densités . . . 4
1.3 Une approche élémentaire . . . 5
1.4 Une situation concordante à la réalité . . . 6

2 L'algorithme EM **11**
2.1 Présentation laconique et pseudo-code . . . 11
2.2 Un théorème . . . 12

Bibliographie **14**

A Annexe **15**

Introduction

Chapitre 1

Modélisation du problème

1.1 Modélisation du problème

Nous allons pour commencer donner une première définition, qui est au coeur du présent projet.

Définition 1 (Loi de mélange)**.** On appelle loi de mélange toute loi dont la densité s'écrit sous la forme d'une combinaison convexe de diverses densités. C'est-à-dire que si l'on se donne J variables aléatoires $X_1, \cdots, X_J$ de densité respective $f_1(x), \cdots, f_J(x)$, alors est appellée loi de mélange toute variable aléatoire X dont la densité f s'exprime sous la forme

$$f(x) := \sum_{i=1}^{J} \alpha_i f_i(x) \ , \ \alpha_i \in \mathbb{R}$$

Afin de modéliser commodément le problème, nous introduisons les variables aléatoires suivantes :

- La variable aléatoire X, modélisant la taille des nids
- Z, la variable aléatoire représentant l'espèce de mouette qui a construit le nid

Enfin, nous nous placerons sous les hypothèses suivantes :

Hypothèse 1. *Nous supposerons que, $\forall j \in [\![1, J]\!]$, la taille des nids d'une espèce j (i.e. X conditionnellement à $(Z = j)$) suit une loi normale $\mathcal{N}(\mu_j, v_j)$. Nous dénoterons par $f_\theta(x|Z = j) := \gamma_{\mu_j,v_j}(x)$ cette densité.*

Hypothèse 2. *Soit $\Theta := \{\theta = (\alpha_j, \mu_j, v_j)_{1 \leq j \leq J}$ tels que $\alpha_j > 0 \ \forall j \in [\![1, J]\!]$ et $\sum_{j=1}^{J} \alpha_j = 1\}$. Soient $X_1, \cdots, X_n$ un échantillon de même loi que X. On supposera qu'il existe un $\theta \in \Theta$ tel que les données récoltées, ici les tailles des nids, soient la réalisation du précédent échantillon.*

Proposition 1. *La variable Z est discrète et à valeur dans un sous-ensemble fini de $\mathbb{N}$, elle suit donc une loi*

$$\sum_{j=1}^{J} \alpha_j \delta_j$$

où J représente le nombre d'espèce de mouettes considéré et les $\alpha(j)$ sont des réels, positifs stricts, représentant la proportion de nids de l'espèce j, tels que $\sum_{i=1}^{J} \alpha_j = 1$.

Il s'ensuit la proposition suivante, qui sera la racine du présent projet.

Proposition 2. *La distribution de la taille des nids de mouettes, i.e. X, admet pour densité ,au point x et par rapport à la mesure de Lebesgue sur $\mathbb{R}$, la fonction f_θ définie comme suit*

$$f_\theta(x) = \sum_{j=1}^{J} \alpha_j \gamma_{\mu_j,v_j}(x)$$

Démonstration. On vérifie que l'on obtient bien une densité de probabilité, la forme de cette dernière étant la conséquence directe de la définition de la variable aléatoire X :

$$\int_{\mathbb{R}} f_\theta(x)dx = \int_{\mathbb{R}} \sum_{j=1}^{J} \alpha_j \gamma_{\mu_j,v_j}(x)dx = \sum_{j=1}^{J} \alpha_j \int_{\mathbb{R}} \gamma_{\mu_j,v_j}(x)dx = \sum_{j=1}^{J} \alpha_j = 1$$

❧

Le but de ce projet sera d'étudier des méthodes permettant l'estimation des divers paramètres de cette densité. Nous détonerons par $\theta := (\alpha_j, \mu_j, v_j)_{1\le j\le J}$ les vecteurs des ces dits paramètres.

1.2 Une histoire de densités

Introduisons une dernière densité et une dernière probabilité, qui nous seront fort utile quant à l'expression des Log-vraisemblances conditionnelles :

Proposition 3. *Nous avons les résultats suivant :*

1. *La densité du vecteur aléatoire (X, Z) nous est donnée par :*

$$\begin{aligned} h_\theta : \mathbb{R} \times \{1, \cdots, J\} &\to \mathbb{R}_+ \\ (x, j) &\mapsto \alpha_j \times \gamma_{\mu_j,v_j}(x) \end{aligned}$$

2. *La probabilité de la loi de Z sachant $X = x$ nous est donnée par :*

$$\mathbb{P}_\theta(Z = j|X = x) = \frac{\gamma_{\mu(z),v(z)} \times \alpha_z}{\sum_{i=1}^{J} \alpha_i \times \gamma_{\mu(i),v(i)}(x)}$$

Démonstration. Par propriété des lois conditionnelles, nous avons que

$$h_\theta(x, j) = f_\theta(x|Z = j) \times \mathbb{P}_\theta(Z = j) = f_\theta(x) \times \mathbb{P}_\theta(Z = j|X = x)$$

De ceci, nous déduisons aisément la densité de la loi du vecteur aléatoire (X, Z) :

$$h_\theta(x, j) = \alpha_j \times \gamma_{\mu_j,v_j}(x)$$

Puis la densité de la loi conditionnelle de Z sachant $X = x$:

$$\mathbb{P}_\theta(Z = j|X = x) = \frac{\gamma_{\mu(z),v(z)} \times \alpha_z}{\sum_{i=1}^{J} \alpha_i \times \gamma_{\mu(i),v(i)}(x)}$$

❧

Remarque 1. Nous pouvons dès à présent noter que pour un échantillon $X_1, \cdots, X_n$ de même loi que X, nous avons

$$\forall i \in [\![1, n]\!],\ h_\theta(X_i, j) = f_\theta(X_i) \times \mathbb{P}_\theta(Z = j|X = X_i)$$

Ceci nous sera utile dans la suite.

Nous allons dès à présent nous intéresser à l'estimation de ces paramètres.

1.3 Une approche élémentaire

Regardons dans un premier temps un cas simplifié, un cas ne décrivant pas la réalité des observations mais qui a le mérite de constituer une agréable entrée en matière.
Nous supposerons ici qu'ont été relevés simultanément et les mesures des tailles des nids et l'espèce de mouette qui l'a construit. Le modèle ici considéré est donc composé des couples (X_i, Z_i), $i \in [\![1, n]\!]$. On considérera dès lors la fonction de densité $h_\theta(x, z)$.
L'estimation des divers paramètres est alors élémentaire, en témoigne les propositions suivantes :

Proposition 4 (Fonction de Log-vraisemblance)**.** *La Log-vraisemblance du modèle s'écrit*

$$\mathscr{L}_\theta(X_1, \cdots, X_n, Z_1, \cdots, Z_n) = \sum_{j=1}^{J} \#A_j ln(\alpha_j) + \sum_{j=1}^{J} \sum_{i \in A_j} ln(\gamma_{\mu_j, v_j}(X_i))$$

<u>où</u> les A_j sont définis par $A_j := \{i \in [\![1, n]\!]$ tels que $Z_i = j\}$ <u>i.e.</u> $\bigcup_{j=1}^{J} A_j = [\![1, n]\!]$

Démonstration. La Log-vraisemblance du modèle s'écrit :

$$\begin{aligned}\mathscr{L}_\theta(X_1, \cdots, X_n, Z_1, \cdots, Z_n) &= ln\left(\prod_{i=1}^{n} h_\theta(X_i, Z_i)\right) \\ &= ln\left(\prod_{i=1}^{n} \alpha_{Z_i} \gamma_{\mu_j, v_j}(X_i)\right) \\ &= \sum_{i=1}^{n} ln(\alpha_{Z_i}) + ln(\gamma_{\mu_j, v_j}(X_i))\end{aligned}$$

Z_i est à valeur dans $[\![1, J]\!]$, on partitionne donc $I := [\![1, n]\!]$ comme $I = \bigcup_{j=1}^{J} A_j$ pour obtenir

$$\begin{aligned}\mathscr{L}_\theta(X_1, \cdots, X_n, Z_1, \cdots, Z_n) &= \sum_{j=1}^{J} \sum_{i \in A_j} ln(\alpha_{Z_i}) + \sum_{j=1}^{J} \sum_{i \in A_j} ln(\gamma_{\mu_j, v_j}(X_i)) \\ &= \sum_{j=1}^{J} \sum_{i \in A_j} ln(\alpha_j) + \sum_{j=1}^{J} \sum_{i \in A_j} ln(\gamma_{\mu_j, v_j}(X_i)) \\ &= \sum_{j=1}^{J} \#A_j ln(\alpha_j) + \sum_{j=1}^{J} \sum_{i \in A_j} ln(\gamma_{\mu_j, v_j}(X_i))\end{aligned}$$

❧

Nous pouvons dès lors maximiser la log-vraisemblance afin d'obtenir les estimateurs souhaités :

Proposition 5 (Estimateurs)**.** *Les estimateurs du maximum de vraisemblance $\hat{\alpha_j}$ (resp. $\hat{\mu_j}$, et $\hat{v_j}$) de α_j (resp. μ_j et v_j) sont donnés par*

$$\hat{\alpha_j} = \frac{\#A_j}{n}$$

$$\hat{\mu_j} = \frac{\sum_{i \in A_j} X_i}{\#A_j}$$

$$\hat{v_j} = \frac{\sum_{i \in A_j} (X_i - \hat{\mu_j})^2}{\#A_j}$$

Démonstration. Soit $\theta = (\alpha_j, \mu_j, v_j)_{j \in [\![1, J]\!]}$. Il s'agit de déterminer

$$\underset{\theta\in\mathbb{R}^{3J},\sum_{j=1}^{J}\alpha_j=1}{\text{argmax}}\left(\sum_{j=1}^{J}\#A_j ln(\alpha_j)+\sum_{j=1}^{J}\sum_{i\in A_j} ln(\gamma_{\mu_j,v_j}(x_i))\right)$$

Nous avons donc à résoudre un programme de minimisation d'une fonction convexe sur un convexe avec une contraire égalité, il est ainsi naturel de faire appel au Lagrangien.
Ce dernier s'écrit

$$\begin{aligned}
L(\theta) &= \sum_{j=1}^{J}\#A_j ln(\alpha_j)+\sum_{j=1}^{J}\sum_{i\in A_j} ln(\gamma_{\mu_j,v_j}(x_i)) - \lambda\times\left(\sum_{j=1}^{J}\alpha_j - 1\right)\\
&= \sum_{j=1}^{J}\#A_j ln(\alpha_j)+\sum_{j=1}^{J}\sum_{i\in A_j} ln\left(\frac{1}{\sqrt{2\pi v_j}}\exp\left(-\frac{(x_i-\mu_j)^2}{2v_j}\right)\right) - \lambda\times\left(\sum_{j=1}^{J}\alpha_j - 1\right)\\
&= \sum_{j=1}^{J}\#A_j ln(\alpha_j)+\sum_{j=1}^{J}\sum_{i\in A_j}\left(\frac{-1}{2}ln(2\pi v_j)-\frac{(x_i-\mu_j)^2}{2v_j}\right) - \lambda\times\left(\sum_{j=1}^{J}\alpha_j - 1\right)
\end{aligned}$$

Il reste maintenant à résoudre le système suivant, afin d'obtenir le vecteur $\hat{\theta} := (\hat{\alpha_j},\hat{\mu_j},\hat{v_j})_{j\in[\![1,J]\!]}$ solution du programme.

$$\begin{cases}
\dfrac{\#A_j}{\hat{\alpha_j}} - \lambda & = 0\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}(x_i-\hat{\mu_j})/\hat{v_j} & = 0\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}\frac{-0.5\times 2\times\pi}{2\pi\hat{v_j}}+\frac{(x_i-\hat{\mu_j})^2}{2\hat{v_j}^2} & = 0\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{j=1}^{J}\hat{\alpha_j} = 1
\end{cases}$$

Ceci équivaut à

$$\begin{cases}
\dfrac{\#A_j}{\hat{\alpha_j}} & = \lambda\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}x_i & = \displaystyle\sum_{i\in A_j}\hat{\mu_j}\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}(x_i-\hat{\mu_j})^2 & = \displaystyle\sum_{i\in A_j}\hat{v_j}\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{j=1}^{J}\hat{\alpha_j} = 1
\end{cases}
\Leftrightarrow
\begin{cases}
\dfrac{\#A_j}{\hat{\alpha_j}} & = \lambda\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}\frac{x_i}{\#A_j} & = \hat{\mu_j}\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{i\in A_j}\frac{(x_i-\hat{\mu_j})^2}{\#A_j} & = \hat{v_j}\ \forall j\in[\![1,J]\!]\\
\displaystyle\sum_{j=1}^{J}\hat{\alpha_j} = 1
\end{cases}$$

En sommant les J premières lignes du système, on obtient $\displaystyle\sum_{j=1}^{J}\#A_j = \sum_{j=1}^{J}\hat{\alpha_j}\lambda$, i.e. $\lambda = n$. En injectant ceci dans le précédent système, on obtient finalement ce qui était annoncé :

$$\begin{cases}
\hat{\alpha_j} & = \dfrac{\#A_j}{n}\ \forall j\in[\![1,J]\!]\\
\hat{\mu_j} & = \displaystyle\sum_{i\in A_j}\frac{x_i}{\#A_j}\ \forall j\in[\![1,J]\!]\\
\hat{v_j} & = \displaystyle\sum_{i\in A_j}\frac{(x_i-\hat{\mu_j})^2}{\#A_j}\ \forall j\in[\![1,J]\!]
\end{cases}$$

1.4 Une situation concordante à la réalité

Nous nous placerons désormais dans un contexte tout autre que celui du paragraphe précédent, un contexte concordant davantage à la réalité. Dans ce qui suit, nous supperons que ne sont observées que les tailles des

nids, les diverses espèces de mouettes les ayant construit étant en quelques sortes des données "cachées". Nous avons donc un échantillon $X_1, \cdots, X_n$ de même loi que la variable X comme définie ci-dessus.
On définit $\mathscr{L}_{obs}$ la log-vraisemblance des observations, nous obtenons ainsi

Définition 2. La log-vraisemblance des observations s'écrit

$$\mathscr{L}_{obs}(\theta, X_1, \cdots, X_n) := ln\left(\prod_{i=1}^{n} f_\theta(X_i)\right) = \sum_{i=1}^{n} ln\left(\sum_{j=1}^{J} \alpha_j \gamma_{\mu_j, v_j}(X_i)\right)$$

Nous voyons dès lors que l'existence d'une expression analytique du maximum de la log-vraisemblance n'est pas assurée. Il est donc nécessaire de trouver un moyen d'approcher les valeurs des différents estimateurs.
Pour ce faire, on définit une log-vraisemblance des couples (X_i, Z_i) sachant le vecteurs des observations $X_1, \cdots, X_n$.

Proposition 6 (log-vraisemblance conditionnelle)**.** *On définit la log-vraisemblance $\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n)$ conditionnelle par*

$$\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n) = \mathbb{E}_{\tilde{\theta}}[\mathscr{L}_\theta(X_1, \cdots, X_n, Z_1, \cdots, Z_n)|X_1, \cdots, X_n]$$

Nous allons maintenant travailler sur l'expression de la log-vraisemblance conditionnelle et en donner une expression simplifiée, qui nous sera fort utile ultérieurement, et une expression plus substancielle, qui nous sera immédiatement utile.

Proposition 7. *Nous avons*

$$\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n) = \sum_{i=1}^{n}\sum_{j=1}^{J} ln(h_\theta(X_i, j))\mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)$$

Démonstration. En effet

$$\begin{aligned}\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n) &= \mathbb{E}_{\tilde{\theta}}[\mathscr{L}_\theta(X_1, \cdots, X_n, Z_1, \cdots, X_n)|X_1, \cdots, X_n] \\ &= \mathbb{E}_{\tilde{\theta}}\left[ln\left(\prod_{i=1}^{n} h_\theta(X_i, Z_i)\right)|X_1, \cdots X_n\right] \\ &= \sum_{i=1}^{n} \mathbb{E}_{\tilde{\theta}}[ln(h_\theta(X_i, Z_i))|X_1, \cdots, X_n]\end{aligned}$$

Or, les couples (X_i, Z_i) sont indépendants, donc

$$\begin{aligned}\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n) &= \sum_{i=1}^{n} \mathbb{E}_{\tilde{\theta}}[ln(h_\theta(X_i, Z_i))|X_i] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(h_\theta(X_i, j))\mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)\end{aligned}$$

❧

Nous nous appuierons sur l'expression suivante pour l'expression des estimateurs du maximum de vraisemblance :

Proposition 8. *La fonction $\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n)$ se réecrit sous la forme suivante :*

$$\begin{aligned}\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n) = -\frac{n}{2}log(2\pi) &+ \sum_{j=1}^{J}\left(\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)\right) log(\alpha_j) \\ &- \frac{1}{2}\sum_{j=1}^{J}\left(\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)\left(log(v_j) + \frac{(X_i - \mu_j)^2}{v_j}\right)\right)\end{aligned}$$

Démonstration. Il suffit de partir de la forme précédente de la log-vraisemblance conditionnelle, on a ainsi :

$$\begin{aligned}
\mathscr{L}_c(\theta,\tilde{\theta},X_1,\cdots,X_n) &= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(h_\theta(X_i,j))\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\alpha_j\gamma_{\mu_j,v_j}(X_i))\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} \big(ln(\alpha_j)+ln(\gamma_{\mu_j,v_j}(X_i))\big)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\alpha_j)\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)+\sum_{i=1}^{n}\sum_{j=1}^{J} ln(\gamma_{\mu_j,v_j}(X_i))\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)
\end{aligned}$$

Traitons pour commencer la double somme

$$\Delta := \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\gamma_{\mu_j,v_j}(X_i))\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)$$

Nous avons :

$$\gamma_{\mu_j,v_j}(X_i) = \tfrac{1}{\sqrt{2\pi v(j)}}e^{-\frac{1}{2}\frac{(X_i-\mu_j)^2}{v_j}}$$

$$\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i) = \frac{\alpha_j\gamma_{\mu_j,v_j}}{\displaystyle\sum_{i=1}^{J}\alpha_j\gamma_{\mu_i,v_i}}$$

La double somme devient alors

$$\begin{aligned}
\Delta &= \sum_{i=1}^{n}\sum_{j=1}^{J} ln\left(\frac{1}{\sqrt{2\pi v_j}}e^{-\frac{1}{2}\frac{(X_i-\mu_j)^2}{v_j}}\right)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln\left(\frac{1}{\sqrt{2\pi v_j}}\right)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)-\frac{1}{2}\left(\frac{(X_i-\mu_j)^2}{v_j}\right)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} -\frac{1}{2}ln(2\pi)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)-\frac{1}{2}ln(v_j)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)-\frac{1}{2}\left(\frac{(X_i-\mu_j)^2}{v_j}\right)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= -\frac{n}{2}ln(2\pi)-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(ln(v_j)+\frac{(X_i-\mu_j)^2}{v_j}\right)\times\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\\
&= -\frac{n}{2}ln(2\pi)-\frac{1}{2}\sum_{j=1}^{J}\left(\sum_{i=1}^{n}\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\times\left(ln(v_j)+\frac{(X_i-\mu_j)^2}{v_j}\right)\right)
\end{aligned}$$

On obtient de fait le résultat espéré :

$$\begin{aligned}
\mathscr{L}_c(\theta,\tilde{\theta},X_1,\cdots,X_n) = -\frac{n}{2}log(2\pi) &+ \sum_{j=1}^{J}\left(\sum_{i=1}^{n}\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\right)\times ln(\alpha_j)\\
&-\frac{1}{2}\sum_{j=1}^{J}\left(\sum_{i=1}^{n}\mathbb{P}_{\tilde{\theta}}(Z=j|X=X_i)\times\left(ln(v_j)+\frac{(X_i-\mu_j)^2}{v_j}\right)\right)
\end{aligned}$$

❧

Nous allons dès à présent énoncer une proposition vitale, celle de l'expression des estimateurs du maximum de vraisemblance de la log-vraisemblance conditionnelle. L'expression de ces derniers seront le pivot de l'algorithme EM, que nous présenterons dans le chapitre suivant.

Proposition 9. *La fonction* $\theta \mapsto \mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n)$ *admet un unique maximum* θ_M *donné par :*

$$\hat{\alpha_j} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)$$

$$\hat{\mu_j} = \frac{\sum_{i=1}^{n} X_i \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}$$

$$\hat{v_j} = \frac{\sum_{i=1}^{n}(X_i - \mu_j)^2 \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}$$

Démonstration. Soit $\theta = (\alpha_j, \mu_j, v_j)$. Il s'agit ici de maximiser la fonction $\theta \mapsto \mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n)$

Puisqu'il s'agit d'un problème d'optimisation, nous appliquons la même méthode que précédemment, en introduisant le Lagrangien du problème sous la contrainte $\sum_{i=1}^{n} \alpha(i) = 1$.

Nous reprenons ici l'écriture de $\mathscr{L}_c(\theta, \tilde{\theta}, X_1, \cdots, X_n)$ donné dans la précédente proposition, nous obtenons ainsi l'expression suivante du Lagrangien

$$\begin{aligned} L(\theta, \lambda) = &-\frac{n}{2} log(2\pi) + \sum_{j=1}^{J}\left(\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)\right) log(\alpha_j) \\ &- \frac{1}{2}\sum_{j=1}^{J}\left(\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)\left(log(v_j) + \frac{(X_i - \mu_j)^2}{v_j}\right)\right) - \lambda\left(\sum_{i=1}^{n} \alpha_i - 1\right) \end{aligned}$$

Le Lagrangien admet un maximum sous la contrainte et ce maximum θ^*vérifie le système suivant :

$$\begin{cases} \dfrac{\partial \mathscr{L}}{\partial \alpha_j}(\theta^*) = \dfrac{\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{v_j} - \lambda & = 0 \ \forall j \in [\![1, J]\!] \\ \dfrac{\partial \mathscr{L}}{\partial \mu_j}(\theta^*) = \sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)(-2X_i + 2\mu_j) & = 0 \ \forall j \in [\![1, J]\!] \\ \dfrac{\partial \mathscr{L}}{\partial v_j}(\theta^*) = -\dfrac{1}{2v_j}\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i) + \dfrac{1}{2v_j^2}\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)(X_i - \mu_j)^2 & = 0 \ \forall j \in [\![1, J]\!] \\ \dfrac{\partial \mathscr{L}}{\partial \lambda}(\theta^*) = \sum_{i=1}^{n} \alpha_i - 1 & = 0 \end{cases}$$

Sous $\tilde{\theta}$ fixé, et ce qui est bien le cas, on a $\mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)$ une constante. ???? POURQUOI ???? Le système devient alors :

$$
\begin{cases}
\alpha_j = \dfrac{\displaystyle\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)}{\lambda} & \forall j \in [\![1, J]\!] \\
\mu_j = \dfrac{\displaystyle\sum_{i=1}^{n} X_i \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\displaystyle\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)} & \forall j \in [\![1, J]\!] \\
v_j = \dfrac{\displaystyle\sum_{i=1}^{n} (X_i - \mu_j)^2 \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\displaystyle\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)} & \forall j \in [\![1, J]\!] \\
\displaystyle\sum_{i=1}^{J} \alpha_i = 1
\end{cases}
$$

Sous la contrainte $\sum_{i=1}^{n} \alpha_i = 1$ et la première équation du système précedent on obtient l'égalité suivante :

$$
\begin{aligned}
\sum_{i=1}^{J} \alpha_i &= \sum_{i=1}^{J} \left(\frac{\displaystyle\sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\lambda} \right) \\
&= \frac{\displaystyle\sum_{i=1}^{J} \sum_{i=1}^{n} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\lambda} \\
&= \frac{\displaystyle\sum_{i=1}^{n} \sum_{i=1}^{J} \mathbb{P}_{\tilde{\theta}}(Z = j|X = X_i)}{\lambda} \\
&= \frac{\displaystyle\sum_{i=1}^{n} 1}{\lambda} = 1
\end{aligned}
$$

On en déduit ainsi $\lambda = n$, ainsi que le résultat énoncé. ☙

Tous ces inesthétiques et fastidieux calculs n'ont pas été effectué en vain. Nous les avons réalisé suite à l'introduction d'une notion nouvelle, celle de la log-vraisemblance conditionnelle ; qui elle même à été introduite faute de ne pouvoir obtenir une expression analytique de la log-vraisemblance des observations. Nous allons maintenant tâcher de mettre en exergue le rapport entre ces deux log-vraisemblances.

Chapitre 2

L'algorithme EM

Dans le présent chapitre, nous nous intéresserons A REMPLIR

2.1 Présentation laconique et pseudo-code

présentation textuelle rapide de l'algo : etape E, etape M etc...

Pour l'implémentation de cet algorithme, nous nous sommes appuyés sur le pseudo-code suivant.

Algorithm 1 L'algorithme EM (Dempster et al., 1977).

Entrée(s) : $N \in \mathbb{N}$, $\widehat{\theta}_0 \in \Theta$, un jeu de données $x_1 \dots x_n$;
Initialisation ;

```
k := 1 ;
Tant que K < N + 1 faire
    ETAPE E : Calculer la fonction Q(θ; θ̂_{k-1}) = (1/n) Σ_{i=1}^{n} E_{θ̂_{k-1}}[log f(X_i, Z_i, θ)|X_i = x_i] ;
    ETAPE M : θ̂_k = argmax Q(θ; θ̂_{k-1}) ;
    k ← k + 1 ;
fin du Tant que ;
retourner θ̂_N ;
```

$$k := 1;$$
Tant que $K < N + 1$ **faire**
- **ETAPE E :** *Calculer la fonction* $Q(\theta; \widehat{\theta}_{k-1}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\widehat{\theta}_{k-1}}[log f(X_i, Z_i, \theta)|X_i = x_i]$;
- **ETAPE M :** $\widehat{\theta}_k = argmax\ Q(\theta; \widehat{\theta}_{k-1})$;
- $k \leftarrow k + 1$;

fin du Tant que ;
retourner $\widehat{\theta}_N$;

2.2 Un théorème

Dans cette concise partie, nous donnons une preuve de la croissance de la log-vraisemblance conditionnelle au fur et à mesure des itérations de l'algorithme EM.

Théorème 1. *Soit* $(\theta_k)_{k\in\mathbb{N}}$ *la suite de paramètres construite à l'aide de l'algorithme EM. La log-vraisemblance* $\mathscr{L}_{obs}$ *des observations vérifie*

$$\mathscr{L}_{obs}(\theta_{k+1}, X_1, \cdots, X_n) \geq \mathscr{L}_{obs}(\theta_k, X_1, \cdots, X_n)$$

Démonstration. Nous allons commencer cette preuve en donnant une autre forme de la log-vraisemblance, dépendant de $\mathscr{L}_{obs}(\theta, X_1, \cdots, X_n)$ et d'un terme κ_{θ,θ_k}. Nous avons :

$$\begin{aligned}
\mathscr{L}_c(\theta, \theta_k, X_1, \cdots, X_n) &= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(h_\theta(X_i, j))\mathbb{P}_{\theta_k}(Z = j|X = X_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln\big[f_\theta(X_i) \times \mathbb{P}_\theta(Z = j|X = X_i)\big]\mathbb{P}_{\theta_k}(Z = j|X = X_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(f_\theta(X_i))\mathbb{P}_{\theta_k}(Z = j|X = X_i) + \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\mathbb{P}_\theta(Z = j|X = X_i))\mathbb{P}_{\theta_k}(Z = j|X = X_i) \\
&= \sum_{i=1}^{n} ln(f_\theta(X_i)) \times \underbrace{\sum_{j=1}^{J} \mathbb{P}_{\theta_k}(Z = j|X = X_i)}_{=1} + \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\mathbb{P}_\theta(Z = j|X = X_i))\mathbb{P}_{\theta_k}(Z = j|X = X_i) \\
&= \sum_{i=1}^{n} ln(f_\theta(X_i)) + \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\mathbb{P}_\theta(Z = j|X = X_i))\mathbb{P}_{\theta_k}(Z = j|X = X_i) \\
&= \mathscr{L}_{obs}(\theta, X_1, \cdots, X_n) + \kappa_{\theta,\theta_k}
\end{aligned}$$

Dès lors, on obtient

$$\mathscr{L}_{obs}(\theta_{k+1}, X_1, \cdots, X_n) - \mathscr{L}_{obs}(\theta_k, X_1, \cdots, X_n) = \mathscr{L}_c(\theta_{k+1}, \theta_k, X_1, \cdots, X_n) - \kappa_{\theta_{k+1},\theta_k} - \mathscr{L}_c(\theta_k, \theta_k, X_1, \cdots, X_n) + \kappa_{\theta_k,\theta_k}$$

Or, la quantité $\mathscr{L}_c$ est maximisée en θ_{k+1} lors de l'étape M de l'algorithme, donc

$$\mathscr{L}_c(\theta_{k+1}, \theta_k, X_1, \cdots, X_n) - \mathscr{L}_c(\theta_k, \theta_k, X_1, \cdots, X_n) \geq 0$$

Il reste donc à prouver que

$$\kappa_{\theta_k,\theta_k} - \kappa_{\theta_{k+1},\theta_k} \geq 0$$

En effet, nous avons

$$\begin{aligned}
\kappa_{\theta_k,\theta_k} - \kappa_{\theta_{k+1},\theta_k} &= \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\mathbb{P}_{\theta_k}(Z=j|X=X_i))\mathbb{P}_{\theta_k}(Z=j|X=X_i) \\
&- \sum_{i=1}^{n}\sum_{j=1}^{J} ln(\mathbb{P}_{\theta_{k+1}}(Z=j|X=X_i))\mathbb{P}_{\theta_k}(Z=j|X=X_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} ln\left(\frac{\mathbb{P}_{\theta_k}(Z=j|X=X_i)}{\mathbb{P}_{\theta_{k+1}}(Z=j|X=X_i)}\right)\mathbb{P}_{\theta_k}(Z=j|X=X_i) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{J} ln\left(\frac{\mathbb{P}_{\theta_{k+1}}(Z=j|X=X_i)}{\mathbb{P}_{\theta_k}(Z=j|X=X_i)}\right)\mathbb{P}_{\theta_k}(Z=j|X=X_i) \\
&\geq -\sum_{i=1}^{n} ln\left(\sum_{j=1}^{J}\frac{\mathbb{P}_{\theta_{k+1}}(Z=j|X=X_i)}{\mathbb{P}_{\theta_k}(Z=j|X=X_i)}\mathbb{P}_{\theta_k}(Z=j|X=X_i)\right) \\
&\Big[\text{Cette dernière inégalité est due à la convexité du } log \text{ et au fait que } \sum_{j=1}^{J}\mathbb{P}_{\theta_k}(Z=j|X)=1\Big] \\
&= -\sum_{i=1}^{n} ln\left(\sum_{j=1}^{J}\mathbb{P}_{\theta_{k+1}}(Z=j|X=X_i)\right) \\
&= -\sum_{i=1}^{n} ln(1) \\
&= 0
\end{aligned}$$

On obtient ainsi

$$\kappa_{\theta_k,\theta_k} - \kappa_{\theta_{k+1},\theta_k} \geq 0$$

Et finalement

$$\mathscr{L}_{obs}(\theta_{k+1}, X_1, \cdots, X_n) \geq \mathscr{L}_{obs}(\theta_k, X_1, \cdots, X_n)$$

❧

Bibliographie

Liens utiles

https://www.lpsm.paris/pageperso/rebafka/BookGraphes/algorithme-em.html
https://members.loria.fr/moberger/Enseignement/AVR/Exposes/algo-em.pdf
http://faculty.washington.edu/fxia/courses/LING572/EM_collins97.pdf
https://core.ac.uk/download/pdf/155777956.pdf
http://www.cmap.polytechnique.fr/~bansaye/CoursTD6.pdf

Annexe A

Annexe