



Université de Montpellier

Projet M1 SSD

Un modèle pour les nids de mouettes

Rédigé par

CARVAILLO Thomas

CÔME Olivier

PRALON Nicolas

Encadrante : Elodie BRUNEL-PICCININI

31 mars 2022

Table des matières

Introduction	2
1 Un peu de théorie	3
1.1 Modélisation du problème	3
1.2 Un cas élémentaire	4
1.3 Le cas réel	6
2 L'algorithme EM	8
2.1 Quelques preuves	8
2.2 L'algo	9
Bibliographie	10
A Annexe	11

Introduction

Chapitre 1

Un peu de théorie

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 linéaire de diverses densités. C'est-à-dire que si l'on se donne n variables aléatoires X_1, \dots, X_n de densité respective $f_1(x), \dots, f_n(x)$, alors est appelée loi de mélange toute variable aléatoire X dont la densité f s'exprime sous la forme

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

Continuons, afin de modéliser commodément le problème, par introduire les variables aléatoires suivantes :

- ✂ La variable aléatoire X , modélisant la taille des nids
- ✂ La variable aléatoire Y , décrivant la taille du nid d'une espèce donnée
- ✂ Et Z , la variable aléatoire représentant l'espèce de mouette qui a construit le nid

Hypothèse 1. Nous supposons que la taille des nids d'une espèce j (*i.e.* X conditionnellement à $(Z = j)$) suit une loi normale $\mathcal{N}(\mu_j, v_j)$.

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 $\sum_{j=1}^J \alpha(j) = 1$.

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

$$f(x) = \sum_{j=1}^J \alpha(j) \gamma_{\mu_j, v_j}(x)$$

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 :

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



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

Hypothèse 2. Soit $\Theta := \{\theta = (\alpha_j, \mu_j, v_j)_{1 \leq j \leq J} \text{ tels que } \alpha_j > 0 \forall j \in \llbracket 1, J \rrbracket \text{ et } \sum_{j=1}^J \alpha_j = 1\}$. Soient X_1, \dots, 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.

Notation 1 (Densités). Le vecteur θ ayant été dûment introduit, nous noterons

1. $g_\theta(z) = \alpha(z)$ la densité de la variable aléatoire Z
2. $f_\theta(x|Z = j) := \gamma_{\mu_j, v_j}(x)$ la densité de la loi de X sachant Z , i.e. de la variable aléatoire Y

qui sont, respectivement, contre la mesure de comptage sur \mathbb{N} , et par rapport à la mesure de Lebesgue sur \mathbb{R} ,

Introduisons deux dernières densités, qui nous seront fort utile quant à l'expression des Log-vraisemblances conditionnelles :

Proposition 3 (Densité de Z sachant X). La densité de la loi de Z sachant X est donnée par

$$g_\theta(z|X = x) = \frac{\alpha(z)\gamma_{\mu_z, v_z}(x)}{\sum_{j=1}^J \alpha(j)\gamma_{\mu_j, v_j}(x)}, \text{ pour } z \in \{1, \dots, J\}$$

Proposition 4 (Densité du vecteur (X, Z)). Soit le vecteur aléatoire (X, Z) ; sa densité nous est donnée par

$$h_\theta(x, z) := \alpha(z)\gamma_{\mu_z, v_z}(x)$$

Démonstration. En effet, si on dénote par F la fonction de répartition de ce vecteur aléatoire, on obtient

$$\begin{aligned} F(x, z) &= \mathbb{P}(X \leq x, Z \leq z) \\ &= \sum_{\zeta \in \llbracket 1, z \rrbracket} \mathbb{P}(X \leq x | Z = \zeta) \times \mathbb{P}(Z = \zeta) \\ &= \sum_{\zeta=1}^z \int_{-\infty}^x \gamma_{\mu_\zeta, v_\zeta}(t) dt \times \alpha(\zeta) \\ &= \int_{-\infty}^x \int_{\mathbb{N}} \alpha(\zeta) \mathbb{1}_{\llbracket 1, z \rrbracket}(\zeta) \gamma_{\mu_\zeta, v_\zeta}(t) dt d\delta_\zeta \end{aligned}$$

La densité de (X, Z) s'ensuit.



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

1.2 Un cas 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 supposons 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 couple (X_i, Z_i) , $i \in \llbracket 1, n \rrbracket$. On considérera dès lors la fonction de densité $h_\theta(x, z)$.

L'estimation des divers paramètres est alors élémentaire, en témoigne les propositions suivante :

Proposition 5 (Fonction de Log-vraisemblance). La Log-vraisemblance du modèle s'écrit

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

où les A_j sont définis par $A_j := \{i \in \llbracket 1, n \rrbracket \text{ tels que } Z_i = j\}$ i.e. $\bigcup_{j=1}^J A_j = \llbracket 1, n \rrbracket$

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

$$\begin{aligned}\mathcal{L}_\theta(X_1, \dots, X_n, z_1, \dots, z_n) &= \ln \left(\prod_{i=1}^n \alpha(z_i) \gamma_{\mu_j, v_j}(x_i) \right) \\ &= \sum_{i=1}^n \ln(\alpha(z_i)) + \ln(\gamma_{\mu_j, v_j}(x_i))\end{aligned}$$

z_i est à valeur dans $\llbracket 1, J \rrbracket$, on partitionne donc $I := \llbracket 1, n \rrbracket$ comme $I = \bigcup_{j=1}^J A_j$ pour obtenir

$$\begin{aligned}\mathcal{L}_\theta(X_1, \dots, X_n, Z_1, \dots, Z_n) &= \sum_{j=1}^J \sum_{i \in A_j} \ln(\alpha(z_i)) + \sum_{j=1}^J \sum_{i \in A_j} \ln(\gamma_{\mu_j, v_j}(x_i)) \\ &= \sum_{j=1}^J \sum_{i \in A_j} \ln(\alpha(j)) + \sum_{j=1}^J \sum_{i \in A_j} \ln(\gamma_{\mu_j, v_j}(x_i)) \\ &= \sum_{j=1}^J \#A_j \ln(\alpha(j)) + \sum_{j=1}^J \sum_{i \in A_j} \ln(\gamma_{\mu_j, v_j}(x_i))\end{aligned}$$

2

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

Proposition 6 (Estimateurs). *Les estimateurs du maximum de vraisemblance $\hat{\alpha}(j)$ (resp. $\hat{\mu}_j$, et \hat{v}_j) de $\alpha(j)$ (resp. μ_j et v_j) sont donnés par*

$$\begin{aligned}\hat{\alpha}(j) &= \frac{\#A_j}{n} \\ \hat{\mu}_j &= \frac{\sum_{i \in A_j} X_i}{\#A_j} \\ \hat{v}_j &= \frac{\sum_{i \in A_j} (X_i - \hat{\mu}_j)^2}{\#A_j}\end{aligned}$$

Démonstration. Soit $\theta = (\alpha(j), \mu_j, v_j)_{j \in \llbracket 1, J \rrbracket}$. Il s'agit de déterminer

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

Nous avons donc à résoudre un programme de minimisation d'une fonction convexe sur un convexe avec une contrainte 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 $\hat{\theta} := (\hat{\alpha}(j), \hat{\mu}_j, \hat{v}_j)_{j \in \llbracket 1, J \rrbracket}$ solution du programme.

$$\begin{cases} \frac{\#A_j}{\hat{\alpha}(j)} - \lambda & = 0 \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} (x_i - \hat{\mu}_j) / \hat{v}_j & = 0 \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} \frac{-0.5 \times 2 \times \pi}{2\pi \hat{v}_j} + \frac{(x_i - \hat{\mu}_j)^2}{2\hat{v}_j^2} & = 0 \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{j=1}^J \hat{\alpha}(j) & = 1 \end{cases}$$

Ceci équivaut à

$$\begin{cases} \frac{\#A_j}{\hat{\alpha}(j)} & = \lambda \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} x_i & = \sum_{i \in A_j} \hat{\mu}_j \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} (x_i - \hat{\mu}_j)^2 & = \sum_{i \in A_j} \hat{v}_j \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{j=1}^J \hat{\alpha}(j) & = 1 \end{cases} \Leftrightarrow \begin{cases} \frac{\#A_j}{\hat{\alpha}(j)} & = \lambda \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} \frac{x_i}{\#A_j} & = \hat{\mu}_j \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{i \in A_j} \frac{(x_i - \hat{\mu}_j)^2}{\#A_j} & = \hat{v}_j \quad \forall j \in \llbracket 1, J \rrbracket \\ \sum_{j=1}^J \hat{\alpha}(j) & = 1 \end{cases}$$

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

$$\begin{cases} \hat{\alpha}(j) & = \frac{\#A_j}{n} \quad \forall j \in \llbracket 1, J \rrbracket \\ \hat{\mu}_j & = \sum_{i \in A_j} \frac{x_i}{\#A_j} \quad \forall j \in \llbracket 1, J \rrbracket \\ \hat{v}_j & = \sum_{i \in A_j} \frac{(x_i - \hat{\mu}_j)^2}{\#A_j} \quad \forall j \in \llbracket 1, J \rrbracket \end{cases}$$

1.3 Le cas réel

Nous nous placerons désormais dans un contexte tout autre que celui du paragraphe précédent, un contexte correspondant davantage à la réalité. Dans ce qui suit, nous supposons 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 $X := (X_1, \dots, X_n)$ de même loi que la variable X comme définie ci-dessus.

On définit \mathcal{L}_{obs} la log-vraisemblance des observations, nous obtenons ainsi

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

$$\mathcal{L}_{obs}(\theta, X) := \ln \left(\prod_{i=1}^n f_{\theta}(X_i) \right) = \sum_{i=1}^n \ln \left(\sum_{j=1}^J \alpha(j) \gamma_{\mu_j, v_j}(X_i) \right)$$

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 $X = (X_1, \dots, X_n)$.

Proposition 7 (log-vraisemblance conditionnelle). *On définit la log-vraisemblance $\mathcal{L}_c(\theta, \tilde{\theta}, X)$ conditionnelle par*

$$\mathcal{L}_c(\theta, \tilde{\theta}, X) = \mathbb{E}_{\tilde{\theta}}[\mathcal{L}(\theta, X, Z)|X]$$

Proposition 8. *On a*

$$\mathcal{L}_c(\theta, \tilde{\theta}, X) = \sum_{i=1}^n \sum_{j=1}^J \ln(h_{\theta, j}) g_{\tilde{\theta}}(j|X_i)$$

Démonstration. En effet,

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

Or, les couples (X_i, Z_i) sont indépendants et de même loi, donc

$$\mathcal{L}_c(\theta, \tilde{\theta}, X) = \sum_{i=1}^n \mathbb{E}_{\tilde{\theta}}[\ln(h_{\theta}(X_i, Z_i)|X_i)]$$

Puis, par définition de l'espérance conditionnelle, on a $\mathbb{E}[X|X] = X$, donc

$$\begin{aligned} \mathcal{L}_c(\theta, \tilde{\theta}, X) &= \sum_{i=1}^n \int_{\mathbb{N}} \ln(h_{\theta}(X_i, z)) g_{\tilde{\theta}}(z|X_i) \delta_z \\ &= \sum_{i=1}^n \sum_{j=1}^J \ln(h_{\theta}(X_i, j)) g_{\tilde{\theta}}(j|X_i) \end{aligned}$$

Car $g_{\tilde{\theta}}(z|X = x)$ est définie pour $z \in \{1, \dots, J\}$. 🐼

Proposition 9 (Expression log-vraisemblance). *Hop gros pecs à toi de jouer*

Démonstration. hop 🐼

Proposition 10 (Estimateurs). *Hop gros pecs à toi de jouer*

Démonstration. hop 🐼

Chapitre 2

L'algorithme EM

A REVOIR

2.1 Quelques preuves

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

$$\mathcal{L}_{obs}(\theta_{k+1}, X) \geq \mathcal{L}_{obs}(\theta_k, X)$$

NOTA BENE : Nous dénoterons par $d\mathbb{P}_{\theta_k, Z|X}$ la loi de probabilité de Z sachant X , i.e.

$$d\mathbb{P}_{\theta_k, Z|X} := g_{\theta_k}(z|X = x)\delta_z$$

Démonstration. Nous allons commencer cette preuve en donnant une autre forme de la log-vraisemblance, dépendant de $\mathcal{L}_{obs}(\theta, X)$ et d'un terme $\mathcal{K}_{\theta, \theta_k}$. Nous avons :

$$\begin{aligned} \mathcal{L}_c(\theta, \theta_k, X) &= \sum_{i=1}^n \sum_{j=1}^J \ln(h_{\theta}(X_i, j)) g_{\theta_k}(j|X_i) \\ &= \sum_{i=1}^n \sum_{j=1}^J \ln[f_{\theta}(X_i) \times g_{\theta}(j|X_i = x_i)] g_{\theta_k}(j|X_i) \\ &= \sum_{i=1}^n \sum_{j=1}^J \ln(f_{\theta}(X_i)) g_{\theta_k}(j|X_i) + \sum_{i=1}^n \sum_{j=1}^J \ln(g_{\theta}(j|X_i = x_i)) g_{\theta_k}(j|X_i) \\ &= \sum_{i=1}^n \ln(f_{\theta}(X_i)) \times \underbrace{\sum_{j=1}^J g_{\theta_k}(j|X_i)}_{=1} + \sum_{i=1}^n \mathbb{E}_{\theta_k}[\ln(g_{\theta}(Z_i|X_i = x_i))] \\ &= \sum_{i=1}^n \ln(f_{\theta}(X_i)) + \mathbb{E}_{\theta_k}[\ln(g_{\theta}(Z|X = x))] \\ &= \mathcal{L}_{obs}(\theta, X) + \mathcal{K}_{\theta, \theta_k} \end{aligned}$$

Dès lors, on obtient

$$\mathcal{L}_{obs}(\theta_{k+1}, X) - \mathcal{L}_{obs}(\theta_k, X) = \mathcal{L}_c(\theta_{k+1}, \theta_k, X) - \kappa_{\theta_{k+1}, \theta_k} - \mathcal{L}_c(\theta_k, \theta_k, X) + \kappa_{\theta_k, \theta_k}$$

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

$$\mathcal{L}_c(\theta_{k+1}, \theta_k, X) - \mathcal{L}_c(\theta_k, \theta_k, X) \geq 0$$

Il reste donc à prouver que

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

En effet, on a

$$\begin{aligned}
\kappa_{\theta_k, \theta_k} - \kappa_{\theta_{k+1}, \theta_k} &= \mathbb{E}_{\theta_k} [\ln(g_{\theta_k}(Z|X=x))] - \mathbb{E}_{\theta_k} [\ln(g_{\theta_{k+1}}(Z|X=x))] \\
&= \mathbb{E}_{\theta_k} \left[\ln \left(\frac{g_{\theta_k}(Z|X=x)}{g_{\theta_{k+1}}(Z|X=x)} \right) \right] \\
&= \int \ln \left(\frac{g_{\theta_k}(Z|X=x)}{g_{\theta_{k+1}}(Z|X=x)} \right) d\mathbb{P}_{\theta_k, Z|X} \\
&= \int \ln \left(\frac{g_{\theta_k}(z|X=x)}{g_{\theta_{k+1}}(z|X=x)} \right) g_{\theta_k}(z|X=x) \delta_z \\
&= - \int \ln \left(\frac{g_{\theta_{k+1}}(z|X=x)}{g_{\theta_k}(z|X=x)} \right) g_{\theta_k}(z|X=x) \delta_z \\
&\geq - \ln \left(\int \frac{g_{\theta_{k+1}}(z|X=x)}{g_{\theta_k}(z|X=x)} g_{\theta_k}(z|X=x) \delta_z \right) \text{ (Inégalité de Jensen)} \\
&= - \ln \left(\int g_{\theta_{k+1}}(z|X=x) \delta_z \right) \\
&= - \ln(1) \\
&= 0
\end{aligned}$$

On a ainsi

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

et

$$\mathcal{L}_c(\theta_{k+1}, \theta_k, X) - \mathcal{L}_c(\theta_k, \theta_k, X) \geq 0$$

Et finalement

$$\mathcal{L}_{obs}(\theta_{k+1}, X) \geq \mathcal{L}_{obs}(\theta_k, X)$$

2

Théorème 2 (Non monotonie de la vraisemblance).

2.2 L'algo

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>

Annexe A

Annexe