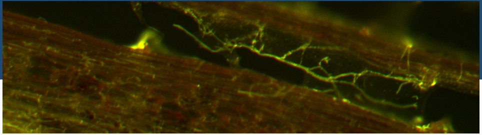
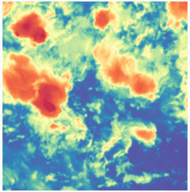
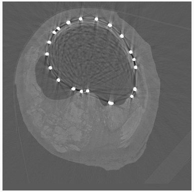
Outils bayésiens en traitement d'images

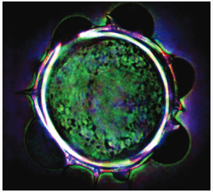
Jean-Baptiste Courbot
jean-baptiste.courbot@uha.fr

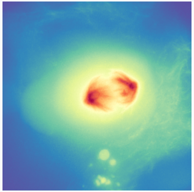
Présentation du cours











Contexte du traitement d'image :

- complexité
- nombre
- nouveauté

Problématiques :

- comment inférer ?
- comment modéliser ?
- comment faire sans exemple ?

Organisation du cours

Organisation :

- 4 séances de 3h45
- 3 rapports de TP à rendre au fil de l'eau pour une note finale
- 2/3 de la note du cours "outils avancés en traitement d'image"

Organisation du cours

Organisation :

- 4 séances de 3h45
- 3 rapports de TP à rendre au fil de l'eau pour une note finale
- 2/3 de la note du cours "outils avancés en traitement d'image"

Moyens :

- Moodle pour les ressources (slides, énoncés) et dépôt de travail.
- Google Colaboratory pour les notebooks Jupyter

Généralités

Notations

Nous travaillerons avec :

- des variables aléatoires A et leur réalisations a
- des vecteurs aléatoires $\boldsymbol{A}$ de réalisation $\boldsymbol{a}$
- la notation « générale » $p(A = a)$ est abrégée en $p(a)$ lorsqu'il n'y a pas confusion
- des paramètres $\boldsymbol{\Theta}$ qui interviennent dans une loi ; on les fait apparaître sous la forme de $p_{\boldsymbol{\Theta}}(a)$

Pourquoi ?

On considère un couple $(\boldsymbol{x}, \boldsymbol{y})$ avec :

- $\boldsymbol{y}$ un vecteur à valeurs dans $\mathbb{R}$
- $\boldsymbol{x}$ un vecteur « caché »

$\boldsymbol{y}$ est l'image observée et on souhaite estimer $\boldsymbol{x}$ qui peut prendre :

- des valeurs discrètes : on parle de **segmentation** d'images
- des valeurs réelles : **restauration** d'images

Le cadre bayésien

Pour estimer $\boldsymbol{x}$ à partir de $\boldsymbol{y}$ il faut **modéliser** le problème.

Lien déterministe : $\boldsymbol{x} = f(\boldsymbol{y})$ i.e. f tient compte de tous les paramètres

Lien probabiliste :

- $\boldsymbol{y}$ est la réalisation **observable** d'une v.a. $\boldsymbol{Y}$
- $\boldsymbol{x}$ est la réalisation **invisible** d'une v.a. $\boldsymbol{X}$

Il faut donc modéliser $p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y})$!

Le cadre bayésien (II)

La loi de Bayes nous permet d'écrire :

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

Comme $p(\boldsymbol{y})$ est constante (l'image est connue, fixée) on note souvent

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$$

Le cadre bayésien (II)

La loi de Bayes nous permet d'écrire :

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

Comme $p(\boldsymbol{y})$ est constante (l'image est connue, fixée) on note souvent

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$$

Tous ces termes portent un nom :

- $p(\boldsymbol{x}|\boldsymbol{y})$ est la probabilité a posteriori
- $p(\boldsymbol{x})$ est la probabilité a priori
- $p(\boldsymbol{y}|\boldsymbol{x})$ est la densité conditionnelle de $\boldsymbol{Y}$ sachant $\boldsymbol{X} = \boldsymbol{x}$ $\rightarrow$ la vraisemblance
- $p(\boldsymbol{x},\boldsymbol{y})$ est la densité du couple = la loi jointe

Démarche générale

On dispose de données $\boldsymbol{Y} = \boldsymbol{y}$; et on veut déduire une segmentation $\boldsymbol{X} = \boldsymbol{x}$.

1. Établir le **modèle a priori** pour $\boldsymbol{X}$: quelle est sa structure ?

Démarche générale

On dispose de données $\boldsymbol{Y} = \boldsymbol{y}$; et on veut déduire une segmentation $\boldsymbol{X} = \boldsymbol{x}$.

1. Établir le **modèle a priori** pour $\boldsymbol{X}$: quelle est sa structure ?
2. Établir un **modèle d'observation** pour lier $\boldsymbol{Y}$ à $\boldsymbol{X}$.

Démarche générale

On dispose de données $\boldsymbol{Y} = \boldsymbol{y}$; et on veut déduire une segmentation $\boldsymbol{X} = \boldsymbol{x}$.

1. Établir le **modèle a priori** pour $\boldsymbol{X}$: quelle est sa structure ?
2. Établir un **modèle d'observation** pour lier $\boldsymbol{Y}$ à $\boldsymbol{X}$.
3. Les paramètres du modèle sont-ils connus ?
    - si oui : on les note $\boldsymbol{\Theta}$
    - sinon : il faudra estimer un $\hat{\boldsymbol{\Theta}}$

Démarche générale

On dispose de données $\boldsymbol{Y} = \boldsymbol{y}$; et on veut déduire une segmentation $\boldsymbol{X} = \boldsymbol{x}$.

1. Établir le **modèle a priori** pour $\boldsymbol{X}$: quelle est sa structure ?
2. Établir un **modèle d'observation** pour lier $\boldsymbol{Y}$ à $\boldsymbol{X}$.
3. Les paramètres du modèle sont-ils connus ?
   - si oui : on les note $\boldsymbol{\Theta}$
   - sinon : il faudra estimer un $\hat{\boldsymbol{\Theta}}$
4. Estimer $\boldsymbol{X} = \boldsymbol{x}$ à l'aide de $p_{\boldsymbol{\Theta}}(\boldsymbol{X}|\boldsymbol{Y})$ ou $p_{\hat{\boldsymbol{\Theta}}}(\boldsymbol{X}|\boldsymbol{Y})$.

Positionnement du cours

Problématique principale : **segmentation d'image**.

Nous allons étudier trois modèles *a priori* :

- sites (pixels) IID
- chaîne de Markov
- champs de Markov

... dans un contexte :

- supervisé : $\boldsymbol{\Theta}$ connu
- non supervisé : $\boldsymbol{\Theta}$ inconnu

Le cas indépendant

Le cas indépendant

Cadre

Modèle a priori

Le modèle le plus simple : tous les pixels sont indépendants.

Donc sur une image $\boldsymbol{x} = \{x_s\}_{s \in \mathcal{S}}$:

▷ $\mathcal{S}$ = ensemble des pixels (sites)

$$p(\boldsymbol{X} = \boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(X_s = x_s)$$

Modèle a priori

Le modèle le plus simple : tous les pixels sont indépendants.

Donc sur une image $\boldsymbol{x} = \{x_s\}_{s \in \mathcal{S}}$: ▷ $\mathcal{S}$ = ensemble des pixels (sites)

$$p(\boldsymbol{X} = \boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(X_s = x_s)$$

Les classes sont discrètes : $\boldsymbol{x}_s \in \Omega \stackrel{\text{def.}}{=} \{\omega_1, \omega_2, \dots, \omega_K\}$

Les lois a priori sont basées sur $p(X_s = \omega_k) \stackrel{\text{def.}}{=} \pi_k$. ▷ $\sum_k \pi_k = 1$

Modèle d'observation

On se donne une loi qui lie $\boldsymbol{Y}$ et $\boldsymbol{X}$.

Hypothèse du bruit indépendant : (Conservée sur le cours)

$$p(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(Y_s = y_s | X_s = x_s)$$

Modèle d'observation

On se donne une loi qui lie $\boldsymbol{Y}$ et $\boldsymbol{X}$.

Hypothèse du bruit indépendant : (Conservée sur le cours)

$$p(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(Y_s = y_s | X_s = x_s)$$

On suppose $\boldsymbol{y}_s \in \mathbb{R}$. Souvent (et pour ce cours) on utilise la loi normale :

$$\forall s, p(Y_s = y_s | X_s = x_s) \sim \mathcal{N}\left(\mu(x_s), \sigma(x_s)\right)$$

En faisant le lien avec Ω on note plutôt :

$$\forall s, p(Y_s = y_s | X_s = \omega_k) \sim \mathcal{N}(\mu_k, \sigma_k)$$

Lois jointe et postérieure

Finalement :

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \\
&= \prod_{s\in\mathcal{S}} p(Y_s = y_s|X_s = x_s) \prod_{s\in\mathcal{S}} p(X_s = x_s) \\
&= \prod_{s\in\mathcal{S}} p(Y_s = y_s|X_s = \omega_k)p(X_s = \omega_k) \\
&= \prod_{s\in\mathcal{S}} \mathcal{N}(y_s; \mu_k, \sigma_k)p(X_s = \omega_k)
\end{aligned}$$

Lois jointe et postérieure

Finalement :

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \\
&= \prod_{s\in\mathcal{S}} p(Y_s = y_s|X_s = x_s) \prod_{s\in\mathcal{S}} p(X_s = x_s) \\
&= \prod_{s\in\mathcal{S}} p(Y_s = y_s|X_s = \omega_k)p(X_s = \omega_k) \\
&= \prod_{s\in\mathcal{S}} \mathcal{N}(y_s; \mu_k, \sigma_k)p(X_s = \omega_k)
\end{aligned}$$

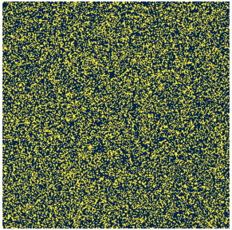
Paramètres du modèle :

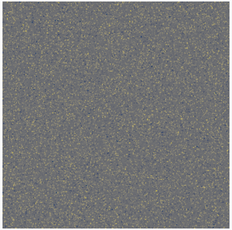
$$\boldsymbol{\Theta} = \{\pi_1, \ldots, \pi_k, \mu_1, \ldots, \mu_k, \sigma_1, \ldots, \sigma_k\}$$

$\boldsymbol{\Theta}$ est supposé connu pour l'instant.

Simulation

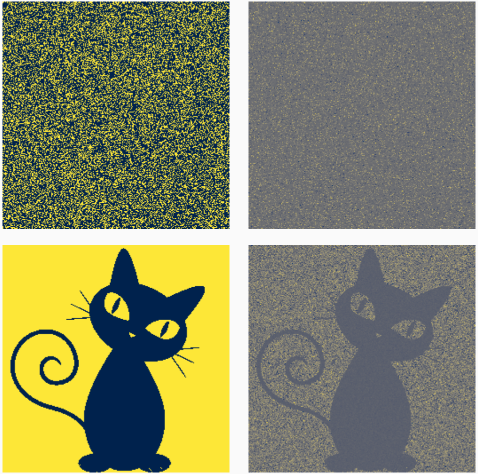
1. Simuler $\boldsymbol{X} = \boldsymbol{x}$ à l'aide des π_k : on fait un tirage aléatoire en chaque site s
2. Simuler $\boldsymbol{Y} = \boldsymbol{y}$ conditionnellement à $\boldsymbol{X} = \boldsymbol{x}$: idem.





Simulation

1. Simuler $\boldsymbol{X} = \boldsymbol{x}$ à l'aide des π_k : on fait un tirage aléatoire en chaque site s
2. Simuler $\boldsymbol{Y} = \boldsymbol{y}$ conditionnellement à $\boldsymbol{X} = \boldsymbol{x}$: idem.



Le cas indépendant

Segmentation

Segmentation

Objectif : estimer $\boldsymbol{X} = \boldsymbol{x}$ à partir de $\boldsymbol{Y} = \boldsymbol{y}$.

Pour cela il faut :

- connaître la loi à posteriori $p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y})$ (vu à l'instant)
- connaître les paramètres $\boldsymbol{\Theta}$ (supposés connus pour l'instant)
- établir une fonction de perte de décision

Stratégie bayésienne, fonctions de perte

Une **stratégie** $\hat{\eta} : \mathbb{R}^S \to \Omega^S$ associe à une observation $\boldsymbol{y}$ une estimée $\boldsymbol{x}$.

$\triangleleft$ $S = |\mathcal{S}|$ le nombre de sites

- $\hat{\eta}$ est déterministe
- $\hat{\eta}$ peut se tromper

Stratégie bayésienne, fonctions de perte

Une **stratégie** $\hat{\eta} : \mathbb{R}^S \to \Omega^S$ associe à une observation $\boldsymbol{y}$ une estimée $\boldsymbol{x}$.

$\triangleleft$ $S = |\mathcal{S}|$ le nombre de sites

- $\hat{\eta}$ est déterministe
- $\hat{\eta}$ peut se tromper

Une **fonction de perte** donne un coût aux erreurs commises :

$$L(\omega_i, \omega_j) = \left| \begin{array}{l} 0 \text{ si } \omega_i = \omega_j \\ \lambda_{ij} \text{ sinon} \end{array} \right.$$

On peut définir une perte moyenne (ou risque moyen) comme $\mathbb{E}\left[L(\hat{\eta}(Y), X\right]$

Stratégie bayésienne, fonctions de perte (II)

La **stratégie bayésienne** $\hat{\eta}_{\mathrm{Bayes}}$ est celle qui minimise le risque moyen :

$$\mathbb{E}\left[L(\hat{\eta}_{\mathrm{Bayes}}(Y), X\right] = \min_{\hat{\eta}} \mathbb{E}\left[L(\hat{\eta}(Y), X\right]$$

Stratégie bayésienne, fonctions de perte (II)

La **stratégie bayésienne** $\hat{\eta}_{\mathrm{Bayes}}$ est celle qui minimise le risque moyen :

$$\mathbb{E}\left[L(\hat{\eta}_{\mathrm{Bayes}}(Y), X\right] = \min_{\hat{\eta}} \mathbb{E}\left[L(\hat{\eta}(Y), X\right]$$

On peut montrer que $\hat{\eta}_{\mathrm{Bayes}}$ est définie par :

$$\hat{\eta}_{\mathrm{Bayes}}(y) = \omega_m \Leftrightarrow \sum_{i=1}^{k} \lambda_{m,i} p(x = \omega_i | y) \leq \sum_{i=1}^{k} \lambda_{j,i} p(x = \omega_i | y) \forall j \in \{1, \ldots, k\}$$

Interprétation : on minimise le coût pondéré par la probabilité a posteriori.

Stratégie bayésienne, fonctions de perte (III)

Deux fonctions de pertes sont très répandues. Soient deux images $\boldsymbol{x}^1$ et $\boldsymbol{x}^2$:

- La fonction de perte 0-1 sur toute l'image :

$$L(\boldsymbol{x}^1, \boldsymbol{x}^2) = \left| \begin{array}{l} 0 \text{ si } \boldsymbol{x}^1 = \boldsymbol{x}^2 \\ 1 \text{ sinon} \end{array} \right.$$

Stratégie bayésienne, fonctions de perte (III)

Deux fonctions de pertes sont très répandues. Soient deux images $\boldsymbol{x}^1$ et $\boldsymbol{x}^2$:

- La fonction de perte 0-1 sur toute l'image :

$$L(\boldsymbol{x}^1, \boldsymbol{x}^2) = \left| \begin{array}{l} 0 \text{ si } \boldsymbol{x}^1 = \boldsymbol{x}^2 \\ 1 \text{ sinon} \end{array} \right.$$

- La fonction de perte 0-1 sur chaque pixel avec somme des erreurs :

$$L(\boldsymbol{x}^1, \boldsymbol{x}^2) = \sum_{s \in \mathcal{S}} \mathbb{1}_{\{x_s^1 \neq x_s^2\}}$$

Stratégie bayésienne, fonctions de perte (IV)

Ces deux fonctions se traduisent par deux critères bayésiens classiques :

Le **maximum a posteriori** (MAP) :

$$\hat{\boldsymbol{x}}^{\mathrm{MAP}} = \arg\max_{\boldsymbol{x} \in \Omega^S} p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y})$$

Stratégie bayésienne, fonctions de perte (IV)

Ces deux fonctions se traduisent par deux critères bayésiens classiques :

Le **maximum a posteriori** (MAP) :

$$\hat{\boldsymbol{x}}^{\mathrm{MAP}} = \underset{\boldsymbol{x} \in \Omega^S}{\arg\max}\, p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y})$$

Le **maximum posterior mode** (MPM), où $\forall s \in \mathcal{S}$:

$$\hat{x}_s^{\mathrm{MPM}} = \underset{x_s \in \Omega}{\arg\max}\, p(X_s = x_s | \boldsymbol{Y} = \boldsymbol{y})$$

Dans le modèle indépendant, on montre facilement que $\hat{\boldsymbol{x}}^{\mathrm{MAP}} = \hat{\boldsymbol{x}}^{\mathrm{MPM}}$.

En effet, $p(\boldsymbol{x}|\boldsymbol{y}) = \prod_{s \in \mathcal{S}} p(x_s|y_s)$.

Stratégie bayésienne, fonctions de perte (V)

Pour résumer, segmenter (MAP ou MPM) dans le modèle indépendant c'est :

- calculer $p(X_s = \omega_k | y_s)$ pour tout site s et classe k
- pour tout site s : choisir la classe la plus probable.

Remarque : on a supposé $\boldsymbol{\Theta}$ connu !

Le cas indépendant

Segmentation non supervisée

Segmentation non supervisée

Contexte. On ne connaît pas toujours les paramètres $\boldsymbol{\Theta}$.

Objectif : estimer conjointement $\boldsymbol{x}$ et $\boldsymbol{\Theta}$ à partir de $\boldsymbol{Y} = \boldsymbol{y}$ uniquement.

Optimisation alternée avec les algorithmes SEM et EM :

- estimer $\hat{\boldsymbol{x}}$ à partir de $\boldsymbol{y}, \hat{\boldsymbol{\Theta}}$ ◁ vu précédemment
- estimer $\hat{\boldsymbol{\Theta}}$ à partir de $\hat{\boldsymbol{x}}, \boldsymbol{y}$

... et répéter jusqu'à convergence.

Estimer $\hat{\Theta}$ à partir de (x, y) : estimateurs du max. de vraisemblance (EMV)

Avec les données complètes $(\boldsymbol{x}, \boldsymbol{y})$, et pour toutes les classes $\omega_k \in \Omega$, on peut estimer :

- la proportion :

$$\hat{\pi}_k = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}$$

Estimer $\hat{\Theta}$ à partir de (x, y) : estimateurs du max. de vraisemblance (EMV)

Avec les données complètes $(\boldsymbol{x}, \boldsymbol{y})$, et pour toutes les classes $\omega_k \in \Omega$, on peut estimer :

- la proportion :

$$\hat{\pi}_k = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}$$

- la moyenne :

$$\hat{m}_k = \frac{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}} y_s}{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}}$$

Estimer $\hat{\Theta}$ à partir de (x, y) : estimateurs du max. de vraisemblance (EMV)

Avec les données complètes $(\boldsymbol{x}, \boldsymbol{y})$, et pour toutes les classes $\omega_k \in \Omega$, on peut estimer :

- la proportion :

$$\hat{\pi}_k = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}$$

- la moyenne :

$$\hat{m}_k = \frac{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}} y_s}{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}}$$

- la variance :

$$\hat{\sigma}_k^2 = \frac{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}} (y_s - \hat{m}_k)^2}{\sum\limits_{s \in \mathcal{S}} \mathbb{1}_{\{x_s = \omega_k\}}}$$

Algorithme SEM

Algorithme **Stochastic Expectation-Maximization** (SEM) :

- On se donne une estimation initiale $\hat{\boldsymbol{x}}^{(0)}$.
- On répète ensuite à chaque itération q :
  - Estimer $\hat{\boldsymbol{\Theta}}^{(q)}$ avec $(\boldsymbol{x}^{(q-1)}, \boldsymbol{y})$
  - Simuler $\boldsymbol{x}^{(q)}$ avec $(\hat{\boldsymbol{\Theta}}^{(q)}, \boldsymbol{y})$ *i.e.* avec $p_{\hat{\boldsymbol{\Theta}}^{(q)}}(\boldsymbol{x}|\boldsymbol{y})$
- Jusqu'à observer une « stabilité » des $\hat{\boldsymbol{\Theta}}^{(q)}$.

Algorithme SEM

Algorithme **Stochastic Expectation-Maximization** (SEM) :

- On se donne une estimation initiale $\hat{\boldsymbol{x}}^{(0)}$.
- On répète ensuite à chaque itération q :
  - Estimer $\hat{\boldsymbol{\Theta}}^{(q)}$ avec $(\boldsymbol{x}^{(q-1)}, \boldsymbol{y})$
  - Simuler $\boldsymbol{x}^{(q)}$ avec $(\hat{\boldsymbol{\Theta}}^{(q)}, \boldsymbol{y})$ *i.e.* avec $p_{\hat{\boldsymbol{\Theta}}^{(q)}}(\boldsymbol{x}|\boldsymbol{y})$
- Jusqu'à observer une « stabilité » des $\hat{\boldsymbol{\Theta}}^{(q)}$.

$\rightarrow$ Intérêt : simple avec les outils déjà vus

$\rightarrow$ Inconvénient : stochastique

Algorithme EM

Pour éviter l'aspect stochastique : revenir à l'objectif = estimer $\boldsymbol{\Theta}$.

Avec $\boldsymbol{x}$ connu, les EMV résolvent :

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} p_{\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{y})$$

Algorithme EM

Pour éviter l'aspect stochastique : revenir à l'objectif = estimer $\boldsymbol{\Theta}$.

Avec $\boldsymbol{x}$ connu, les EMV résolvent :

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} p_{\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{y})$$

À partir d'un $\boldsymbol{\Theta}^t$ connu : calcul de **l'espérance conditionnelle** de la vraisemblance de $\boldsymbol{\Theta}$, étant donné $p_{\boldsymbol{\Theta}^t}(\boldsymbol{x}|\boldsymbol{y})$.

$$Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^t) = \mathbb{E}_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Theta}^t}\left[p_{\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{y})\right]$$

Algorithme EM (II)

On alterne à une itération t :

- Calcul de l'espérance conditionnelle $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^t)$ ◁ Expectation
- Choix de $\boldsymbol{\Theta}^{t+1} = \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^t)$ ◁ Maximisation

Algorithme EM (II)

On alterne à une itération t :

- Calcul de l'espérance conditionnelle $Q(\mathbf{\Theta}|\mathbf{\Theta}^t)$ ◁ Expectation
- Choix de $\mathbf{\Theta}^{t+1} = \arg\max_{\mathbf{\Theta}} Q(\mathbf{\Theta}|\mathbf{\Theta}^t)$ ◁ Maximisation

Dans le cas indépendant :

- Calcul de tous les $p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)$, pour tout s et k
- Estimation de $\mathbf{\Theta}^{t+1}$ par les **estimateurs empiriques pondérés**

Algorithme EM (III)

Estimation de $\mathbf{\Theta}^{(t+1)}$ par les **estimateurs empiriques pondérés** avec $p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)$ et $\boldsymbol{y}$.

$$\hat{\pi}_k^{(t+1)} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)$$

Algorithme EM (III)

Estimation de $\mathbf{\Theta}^{(t+1)}$ par les **estimateurs empiriques pondérés** avec $p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)$ et $\boldsymbol{y}$.

$$\hat{\pi}_k^{(t+1)} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)$$

$$\hat{m}_k^{(t+1)} = \frac{\sum\limits_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \mathbf{\Theta}^t) y_s}{\sum\limits_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \mathbf{\Theta}^t)}$$

Algorithme EM (III)

Estimation de $\boldsymbol{\Theta}^{(t+1)}$ par les **estimateurs empiriques pondérés** avec $p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)$ et $\boldsymbol{y}$.

$$\hat{\pi}_k^{(t+1)} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)$$

$$\hat{m}_k^{(t+1)} = \frac{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t) y_s}{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)}$$

$$(\hat{\sigma}_k^2)^{(t+1)} = \frac{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t) \left(y_s - \hat{m}_k^{(t+1)} \right)^2}{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)}$$

Algorithme EM (III)

Estimation de $\boldsymbol{\Theta}^{(t+1)}$ par les **estimateurs empiriques pondérés** avec $p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)$ et $\boldsymbol{y}$.

$$\hat{\pi}_k^{(t+1)} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)$$

$$\hat{m}_k^{(t+1)} = \frac{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t) y_s}{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)}$$

$$(\hat{\sigma}_k^2)^{(t+1)} = \frac{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t) \left(y_s - \hat{m}_k^{(t+1)} \right)^2}{\sum_{s \in \mathcal{S}} p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)}$$

Remarque : similarité avec les EMV ; en remplaçant $\mathbb{1}_{\{x_s = \omega_k\}}$ par $p(x_s = \omega_k | y_s, \boldsymbol{\Theta}^t)$.

Le cas indépendant, pour résumer

Nous avons vu :

- les hypothèses du modèle indépendant et la loi jointe
- les fonctions de pertes et les critères du MAP, du MPM ▷ généralisables
- l'estimation des paramètres dans le cas indépendant ▷ réutilisables
- EM et SEM ▷ généralisables