

Soit r un entier strictement positif. Notons \mathbb{S}_r l'ensemble des matrices de transition de taille $r \times r$, i.e. l'ensemble des matrices de taille $r \times r$ de réels positifs telles que la somme de chaque ligne soit égale à 1. Considérons un paramètre $\theta = \{P, \mu_1, \dots, \mu_r, v_1, \dots, v_r\} \in \Theta$, où $\Theta = \mathbb{S}_r \times \mathbb{R}^r \times (\mathbb{R}_+^*)^r$ et le modèle suivant paramétré par θ .

Soit $(X_k)_{0 \leq k \leq n}$ une chaîne de Markov discrète à valeurs dans $\{1, \dots, r\}$, de matrice de transition P et de loi initiale ν . Cela signifie que pour tout $1 \leq j \leq r$, $\mathbb{P}_\theta(X_0 = j) = \nu_j$ et pour tout $0 \leq k \leq n-1$ et tout $1 \leq i, j \leq r$, $\mathbb{P}_\theta(X_{k+1} = j | X_k = i) = P_{i,j}$. On considère que cette chaîne est uniquement observée au travers des variables $(Y_k)_{0 \leq k \leq n}$, indépendantes conditionnellement à $(X_k)_{0 \leq k \leq n}$ et telles que pour tout $0 \leq \ell \leq n$, la loi de Y_ℓ sachant $(X_k)_{0 \leq k \leq n}$ est une gaussienne de moyenne μ_{X_ℓ} et de variance v_{X_ℓ} .

1. Écrire la logvraisemblance jointe de $(X_{0:n}, Y_{0:n})$: $\theta \mapsto \log p_\theta(X_{0:n}, Y_{0:n})$.
2. Écrire la quantité intermédiaire de l'EM $Q(\theta, \theta')$ pour tout θ, θ' :

$$Q(\theta, \theta') = \mathbb{E}_{\theta'} [\log p_\theta(X_{0:n}, Y_{0:n}) | Y_{0:n}] .$$

3. Écrire cette quantité en faisant apparaître les probabilités

$$\omega_{k-1,k}^\theta(i, j) = \mathbb{P}_\theta(X_{k-1} = i, X_k = j | Y_{0:n}) ,$$

pour $1 \leq k \leq n$ et

$$\tilde{\omega}_k^\theta(i) = \mathbb{P}_\theta(X_k = i | Y_{0:n}) ,$$

pour $0 \leq k \leq n$.

4. À l'itération $p \geq 0$, on dispose de l'estimation $\hat{\theta}^{(p)}$. Écrire l'estimateur $\hat{\theta}^{(p+1)}$ en maximisant $\theta \mapsto Q(\theta, \hat{\theta}^{(p)})$.
5. Dans le cas où on souhaite également apprendre la loi initiale de la chaîne de Markov et que $\theta = \{P, \mu_1, \dots, \mu_r, v_1, \dots, v_r, \nu_1, \dots, \nu_r\}$, donner les équations de mise à jour de ν .
6. Calculer le gradient de la logvraisemblance des observations : $\theta \mapsto \nabla_\theta \log p_\theta(Y_{0:n})$.
7. En déduire un algorithme de mise à jour des paramètres de type "descente de gradient".
8. **Bonus:** Calcul des $\omega_{k-1,k}^\theta(i, j)$, $1 \leq k \leq n$, $1 \leq i, j \leq r$.
 - (a) Montrer que l'on peut calculer récursivement $\mathbb{P}_\theta(X_k = i | Y_{0:k})$, $0 \leq k \leq n$, $1 \leq i, j \leq r$.
 - (b) Montrer que l'on peut calculer récursivement, de $k = n$ à $k = 0$, $\mathbb{P}_\theta(X_k = i | Y_{0:n})$, $0 \leq k \leq n$, $1 \leq i \leq r$.
 - (c) Conclure.