$$\begin{array}{ccccccc} \longrightarrow X_{k-1} & \xrightarrow{P} & X_k & \longrightarrow & X_{k+1} & \longrightarrow \cdots \\ \downarrow Q & & \downarrow Q & & \downarrow Q & \\ \cdots\ Y_{k-1} & & Y_k & & Y_{k+1} & \cdots \end{array}$$

(Chaînes de Markov cachées)

- (X_k) C.M. de matrice de transition P et loi initiale ν
- Conditionnellement à $(X_1, \dots, X_n)$, les $(Y_i)_i$ sont indép. et la loi de Y_k dépend seulement de la valeur de X_k, i.e.

$$\mathbb{P}(Y_{1:n} = y_{1:n} \mid X_{1:n} = x_{1:n}) = \prod_{k=1}^{n} \mathbb{P}(Y_k = y_k \mid X_k = x_k)$$
$$= \prod_{k=1}^{n} Q(x_k, y_k).$$

<u>Conséquences</u> :

- $(X_n, Y_n)_{n \geq 1}$ est une CM sur $E \times F$ et

$$\mathbb{P}(X_n = x_n, Y_n = y_n \mid X_{1:n-1} = x_{1:n-1}, Y_{1:n-1} = y_{1:n-1})$$
$$= P(x_{n-1}, x_n)\, Q(x_n, y_n).$$ (en particulier ne dépend pas de y_{n-1})

- $\mathbb{P}(X_n = x_n \mid X_{1:n-1} = x_{1:n-1}, Y_{1:n-1} = y_{1:n-1}) = P(x_{n-1}, x_n)$
(faire la somme sur y_n)

<u>Preuve</u> :

$$\mathbb{P}(X_n = x_n, Y_n = y_n \mid X_{1:n-1} = x_{1:n-1}, Y_{1:n-1} = y_{1:n-1}) = (*)$$
$$= \frac{\mathbb{P}(X_{1:n} = x_{1:n}, Y_{1:n} = y_{1:n})}{\mathbb{P}(X_{1:n-1} = x_{1:n-1}, Y_{1:n-1} = y_{1:n-1})}.$$

$$\mathbb{P}(X_{1:n} = x_{1:n}, Y_{1:n} = y_{1:n}) = \mathbb{P}(Y_{1:n} = y_{1:n} \mid X_{1:n} = x_{1:n}) \times \mathbb{P}(X_{1:n} = x_{1:n})$$
$$= \prod_{k=1}^{n} Q(x_k, y_k) \times \nu(x_1) \prod_{k=2}^{n} P(x_{k-1}, x_k)$$

$$\Rightarrow (*) = Q(x_n, y_n) \times P(x_{n-1}, x_n).$$

$\hookrightarrow$ cas particulier de chaîne de Markov sur $E \times F$.

Applications : finances, traitement du signal, analyse d'images, analyse de rég. génomiques

Questions :

① Estimer les paramètres :

$\theta = (\nu, P, Q)$

Q : matrice d'observation

P : matrice de transition de la chaîne cachée (X_n)

ν : loi initiale de la chaîne $(X_n)_n$

② Estimer la suite des états cachés $(X_1, \dots, X_n)$ sachant les observations $(Y_1, \dots, Y_n)$.

③ Autres Problèmes liés :

→ Prévision (prédiction) : Estimer $\mathcal{L}oi(X_n)$ sachant $Y_{1:n-1}$

→ Filtrage : Estimer $\mathcal{L}oi(X_n)$ sachant $Y_{1:n}$.

→ Lissage : Estimer $\mathcal{L}oi(X_k)$ sachant $Y_{1:n}$, avec $1 \leq k < n$.

---

① Estimation des paramètres : méthode du maximum de vraisemblance.

- La log-vraisemblance (incomplète) des observations

$$L_n(\theta) := \log \mathbb{P}_\theta(Y_{1:n}) = \log\left(\sum_{x_{1:n}} \mathbb{P}_\theta(X_{1:n} = x_{1:n}, Y_{1:n})\right)$$

$L_n(\theta ; y_{1:n}) = \log \mathbb{P}_\theta(Y_{1:n} = y_{1:n})$

$|E|^n$ termes, impossible dès que n grand.

→ modèles avec données incomplètes

l'algorithme EM ("Expectation - Maximization")

→ pour maximiser (localement) la vraisemblance dans modèles avec données incomplètes quand le calcul de la vraisemblance complète est simple.

$Y_{1:n}$: données observées, $X_{1:n}$: données manquantes

$(X_{1:n}, Y_{1:n})$: données complètes.

→ On suppose que la vraisemblance complète $\log \mathbb{P}_\theta(X_{1:n}, Y_{1:n})$ est facile à calculer.

Idée : (algorithme Baum-Welch dans le cadre des chaînes de Markov cachées)

- Démarrer avec une valeur initiale θ_0
- A la k-ème itération :

→ L'étape E ("espérance")

Calculer $Q(\theta \mid \theta_k) := \mathbb{E}_{\theta_k}\left[\log \mathbb{P}_\theta(X_{1:n}, Y_{1:n}) \mid Y_{1:n}\right]$

(l'espérance conditionnelle de la log-vraisemblance complète des données sous le paramètre courant θ_k)

↳ la meilleure info que l'on a sur la vraisemblance complète, sachant les observations

→ L'étape M ("maximization")

Trouver $\theta_{k+1} := \operatorname{Argmax}_\theta Q(\theta \mid \theta_k)$.

- S'arrêter quand $\delta := \dfrac{\|\theta_{k+1} - \theta_k\|}{\|\theta_k\|} \leq \varepsilon$ (erreur relative)

ou après un nb. max. d'itérations, ou quand
$\left|\log \mathbb{P}_{\theta_{k+1}}(Y_{1:n}) - \log \mathbb{P}_{\theta_k}(Y_{1:n})\right| < \varepsilon$.

Remarques : (*) À chaque itération, la vraisemblance des données obs. augmente

- En utilisant différents points de départ, l'algo peut trouver éventuellement le max. global (l'EMV) (estimateur de max. de vrais.)

Preuve de (*) :

On a $Q(\theta_{k+1} \mid \theta_k) \geq Q(\theta_k \mid \theta_k) \Leftrightarrow Q(\theta_{k+1} \mid \theta_k) - Q(\theta_k \mid \theta_k) \geq 0$

$$0 \leq \mathbb{E}_{\theta_k}\left[\log\left(\frac{\mathbb{P}_{\theta_{k+1}}(X_{1:n}, Y_{1:n})}{\mathbb{P}_{\theta_k}(X_{1:n}, Y_{1:n})}\right) \Bigm| Y_{1:n}\right]$$

$$\overset{\text{Jensen}}{\leq} \log \mathbb{E}_{\theta_k}\left[\frac{\mathbb{P}_{\theta_{k+1}}(X_{1:n}, Y_{1:n})}{\mathbb{P}_{\theta_k}(X_{1:n}, Y_{1:n})} \Bigm| Y_{1:n}\right]$$

$$= \log\left\{\sum_{x_{1:n}} \frac{\mathbb{P}_{\theta_{k+1}}(X_{1:n} = x_{1:n}, Y_{1:n})}{\mathbb{P}_{\theta_k}(X_{1:n} = x_{1:n}, Y_{1:n})} \mathbb{P}_{\theta_k}(X_{1:n} = x_{1:n} \mid Y_{1:n})\right\}$$

$$= \log\left(\sum_{x_{1:n} \in E^n} \frac{\mathbb{P}_{\theta_{k+1}}(X_{1:n} = x_{1:n}, Y_{1:n})}{\mathbb{P}_{\theta_k}(Y_{1:n})}\right) = \log\left(\frac{\mathbb{P}_{\theta_{k+1}}(Y_{1:n})}{\mathbb{P}_{\theta_k}(Y_{1:n})}\right).$$

Donc $\mathbb{P}_{\theta_{k+1}}(Y_{1:n}) \geq \mathbb{P}_{\theta_k}(Y_{1:n})$.

$\boxed{\text{Dans notre cas}}$:

- Calcul de la log-vraisemblance complète :

$$\log \mathbb{P}_\theta(X_{1:n}, Y_{1:n}) = \log\left(\nu(\cancel{x_1}) \prod_{k=2}^{n} p(\cancel{x_{k-1}}, \cancel{x_k}) \cdot \prod_{k=1}^{n} Q(\cancel{x_k}, y_k)\right)$$

$$= \log(\nu(X_1)) + \sum_{k=2}^{n} \log p(X_{k-1}, X_k) + \sum_{k=1}^{n} \log Q(X_k, Y_k).$$

L'Espérance conditionnelle sous le paramètre courant θ^* (θ_k à l'étape k)

$$Q(\theta|\theta^*) = \mathbb{E}_{\theta^*}\left[\log \mathbb{P}_\theta(X_{1:n}, Y_{1:n}) \mid Y_{1:n}\right]$$

$$= \sum_{x\in E} \log(\nu(x)) \times \mathbb{P}_{\theta^*}(X_1 = x \mid Y_{1:n})$$

$$+ \sum_{k=2}^{n} \sum_{x,x'\in E} \log(p(x,x')) \times \mathbb{P}_{\theta^*}(X_{k-1} = x, X_k = x' \mid Y_{1:n})$$

$$+ \sum_{k=1}^{n} \sum_{x\in E} \log Q(x, Y_k) \times \mathbb{P}_{\theta^*}(X_k = x \mid Y_{1:n}).$$

↳ On a besoin de calculer $\underline{\mathbb{P}_\theta(X_k \mid Y_{1:n})}$ et $\underline{\mathbb{P}_\theta(X_{k-1}, X_k \mid Y_{1:n})}$

(Prop.) $\boxed{\text{Prévision et Filtrage}}$: Pour $\theta = (\nu, P, Q)$, $x\in E$, $y_{1:n}\in F^n$ on a

- $$\underbrace{\mathbb{P}_\theta(X_n = x \mid Y_{1:n-1} = y_{1:n-1})}_{\text{"prévision"}} = \sum_{z\in E} \underbrace{\mathbb{P}_\theta(X_{n-1} = z \mid Y_{1:n-1} = y_{1:n-1})}_{\text{"filtrage"}} \times P(z,x)$$

- $$\mathbb{P}_\theta(X_n = x \mid Y_{1:n} = y_{1:n}) = \frac{\mathbb{P}_\theta(X_n = x \mid Y_{1:n-1} = y_{1:n-1}) \times Q(x, y_n)}{\sum_{z\in E} \mathbb{P}_\theta(X_n = z \mid Y_{1:n-1} = y_{1:n-1}) \times Q(z, y_n)}.$$

↳ algo. récursif (forward), avec valeurs initiales

$$\mathbb{P}_\theta(X_1 = x \mid Y_1 = y) = \frac{\mathbb{P}(Y_1 = y \mid X_1 = x)\mathbb{P}(X_1 = x)}{\mathbb{P}(Y_1 = y)} = \frac{\nu(x)Q(x,y)}{\sum_{z\in E} \nu(z)Q(z,y)}.$$

Preuve :

- $$\mathbb{P}(X_n = x \mid Y_{1:n-1} = y_{1:n-1}) = \sum_{z\in E} \underbrace{\mathbb{P}(X_n = x \mid X_{n-1} = z, Y_{1:n-1} = y_{1:n-1})}_{P(z,x)} \times \mathbb{P}(X_{n-1} = z \mid Y_{1:n-1} = y_{1:n-1}).$$

- $$\mathbb{P}(X_n = x \mid Y_{1:n} = y_{1:n}) = \mathbb{P}(X_n = x \mid Y_{1:n-1} = y_{1:n-1}, Y_n = y_n)$$

$$= \frac{\mathbb{P}(X_n = x, Y_n = y_n \mid Y_{1:n-1} = y_{1:n-1})}{\mathbb{P}(Y_n = y_n \mid Y_{1:n-1} = y_{1:n-1})} = \frac{\mathbb{P}(X_n = x \mid Y_{1:n-1} = y_{1:n-1}) \times \overbrace{\mathbb{P}(Y_n = y_n \mid X_n = x, Y_{1:n-1} = y_{1:n-1})}^{Q(x,y_n)\,!}}{\sum_z \mathbb{P}(X_n = z \mid Y_{1:n-1} = y_{1:n-1}) \times Q(z, y_n)}.$$

prévision(n) ← filtr(n-1) ← prév(n-1) ← filtrage(n-2) ← .... ← filtrage(1)

Remarque (Lemme) : Conditionnellement à X_n,

Y_n est indépendante de $Y_{1:n-1}$, $X_{1:n-1}$.

(et $Y_{n:n+p}$)

On va m.q. $\mathbb{P}(Y_n = y_n \mid X_n = x, \cancel{X_{1:n-1} = x_{1:n-1}}, Y_{1:n-1} = y_{1:n-1}) = \mathbb{P}(Y_n = y_n \mid X_n = x) = Q(x, y_n)$

$$\mathbb{P}(Y_n = y_n, X_n = x, Y_{1:n-1} = y_{1:n-1}) = \sum_{x_{1:n-1}} \mathbb{P}(X_{1:n-1} = x_{1:n-1}, X_n = x, Y_{1:n} = y_{1:n})$$

$$= \sum_{x_{1:n-1}} \nu(x_1) \left(\prod_{k=2}^{n-1} P(x_{k-1}, x_k) \times Q(x_k, y_k) \right) \cdot Q(x_1, y_1) Q(x, y_n) P(x_{n-1}, x)$$

$$\mathbb{P}(X_n = x_n, Y_{1:n-1} = y_{1:n-1}) = \sum_{y_n} \dots$$

$$= \sum_{x_{1:n-1}} \nu(x_1) Q(x_1, y_1) P(x_{n-1}, x) \left(\prod_{k=2}^{n-1} P(x_{k-1}, x_k) Q(x_k, y_k) \right).$$

$\hookrightarrow \mathbb{P}(Y_n = y_n \mid X_n = x, Y_{1:n-1} = y_{1:n-1}) = Q(x, y_n)$

Prop. $\boxed{(\underline{\text{Lissage}})}$ Pour $2 \le k \le n$, $y_{1:n} \in F^n$, on a

$$\bullet\ \mathbb{P}(X_{k-1} = x, X_k = x' \mid Y_{1:n} = y_{1:n}) = P(x, x') \frac{\mathbb{P}(X_{k-1} = x \mid Y_{1:k-1} = y_{1:k-1})}{\mathbb{P}(X_k = x' \mid Y_{1:k-1} = y_{1:k-1})} \times$$
$$\times \mathbb{P}(X_k = x' \mid Y_{1:n} = y_{1:n})$$

et pour $1 \le k \le n-1$,

$$\bullet\ \mathbb{P}(X_k = x' \mid Y_{1:n} = y_{1:n}) = \sum_z P(x', z) \times \frac{\mathbb{P}(X_k = x' \mid Y_{1:k} = y_{1:k})}{\mathbb{P}(X_{k+1} = z \mid Y_{1:k} = y_{1:k})} \times \mathbb{P}(X_{k+1} = z \mid Y_{1:n} = y_{1:n})$$

(formule des probas totales suivant $\bigcup_z \{X_{k+1} = z\}$)

(calcul par récurrence inverse ("arrière") ("backward")) lissage$(k) \leftarrow$ liss$(k+1) \leftarrow \dots \leftarrow$ liss(n) = filtrage(n)

Preuve : $\bullet\ \mathbb{P}(X_{k-1} = x, X_k = x' \mid Y_{1:n} = y_{1:n}) = \mathbb{P}(X_k = x' \mid Y_{1:n} = y_{1:n}) \times \mathbb{P}(X_{k-1} = x \mid X_k = x', Y_{1:n} = y_{1:n})$

$$\text{et } \mathbb{P}(X_{k-1} = x \mid X_k = x', Y_{1:n} = y_{1:n}) = \frac{\mathbb{P}(X_{k-1} = x, X_k = x', Y_{1:k-1} = y_{1:k-1}, Y_{k:n} = y_{k:n})}{\mathbb{P}(X_k = x', Y_{1:k-1} = y_{1:k-1}, Y_{k:n} = y_{k:n})}.$$

En haut : $\mathbb{P}(X_k = x' \mid X_{k-1} = x, Y_{1:k-1} = y_{1:k-1}) \times \mathbb{P}(X_{k-1} = x, Y_{1:k-1} = y_{1:k-1}) \times$

$\times \mathbb{P}(Y_{k:n} = y_{k:n} \mid X_k = x', \cancel{X_{k-1} = x, Y_{1:k-1} = y_{1:k-1}})$ (Lemme v.c. en haut)

$= P(x, x') \times \mathbb{P}(X_{k-1} = x \mid Y_{1:k-1} = y_{1:k-1}) \times \mathbb{P}(Y_{1:k-1} = y_{1:k-1}) \times \mathbb{P}(Y_{k:n} = y_{k:n} \mid X_k = x')$

En bas : $\mathbb{P}(X_k = x' \mid Y_{1:k-1} = y_{1:k-1}) \times \mathbb{P}(Y_{1:k-1} = y_{1:k-1}) \times \mathbb{P}(Y_{k:n} = y_{k:n} \mid X_k = x', \cancel{Y_{1:k-1} = y_{1:k-1}})$

d'où le résultat, en divisant ...

("avant - arrière")

↳ algo "forward - backward" pour calculer ces probas. condit. et donc $Q(\theta|\theta^*)$.

⊛ **L'étape M ("maximization")**

On cherche $\theta = (\nu, P, Q)$ qui maximise $Q(\theta|\theta')$ à $y_{1:n}$ et θ' fixés.

$$Q(\theta|\theta') = \sum_{x\in E} \log(\nu(x)) \times \mathbb{P}_{\theta'}(X_1 = x | Y_{1:n})$$
$$+ \sum_{k=2}^{n} \sum_{x,z\in E} \log(P(x,z)) \times \mathbb{P}_{\theta'}(X_{k-1} = x, X_k = z | Y_{1:n})$$
$$+ \sum_{k=1}^{n} \sum_{x\in E} \log(Q(x, Y_k)) \times \mathbb{P}_{\theta'}(X_k = x | Y_{1:n}),$$

sous les contraintes $\sum_x \underbrace{\nu(x)}_{\geq 0} = 1$, $\sum_z \underbrace{P(x,z)}_{\geq 0} = 1$, $\sum_z \underbrace{Q(x,z)}_{\geq 0} = 1$, $\forall x$

↳ maximiser séparément chacune des sommes !

- Pour Q : $\forall x \in E$ fixé, maximiser $\sum_{k=1}^{n} \log Q(x, Y_k) \times \mathbb{P}_{\theta'}(X_k = x | Y_{1:n})$ $= \sum_{y\in F} \log Q(x,y) \times \sum_{k=1}^{n} \mathbb{1}_{\{y_k = y\}} \times \mathbb{P}(X_k = x | Y_{1:n})$

Exc : m.q. la somme est maximale pour $Q(x,y) \propto \sum_{k=1}^{n} \mathbb{1}_{\{y_k = y\}} \mathbb{P}_{\theta'}(X_k = x | Y_{1:n} = y_{1:n})$

De la même façon on obtient :

$$\begin{cases} \hat{Q}(x,y) = \dfrac{\sum_{k=1}^{n} \mathbb{1}_{\{y_k = y\}} \mathbb{P}_{\theta'}(X_k = x | Y_{1:n} = y_{1:n})}{\sum_{k=1}^{n} \mathbb{P}_{\theta'}(X_k = x | Y_{1:n} = y_{1:n})} \\[2ex] \hat{P}(x,x') = \dfrac{\sum_{k=2}^{n} \mathbb{P}_{\theta'}(X_{k-1} = x, X_k = x' | Y_{1:n} = y_{1:n})}{\sum_z \sum_{k=2}^{n} \mathbb{P}_{\theta'}(X_{k-1} = x, X_k = z | Y_{1:n} = y_{1:n})} \\[2ex] \qquad\quad = \dfrac{\sum_{k=2}^{n} \mathbb{P}_{\theta'}(X_{k-1} = x, X_k = x' | Y_{1:n} = y_{1:n})}{\sum_{k=1}^{n-1} \mathbb{P}_{\theta'}(X_k = x | Y_{1:n} = y_{1:n})} \\[2ex] \hat{\nu}(x) = \mathbb{P}_{\theta'}(X_1 = x | Y_{1:n} = y_{1:n}) \end{cases}$$

intuitivement aussi.

Lemme E espace discret, $p \in \mathrm{Prob}(E)$. Alors la fonction $\mathcal{H}_p : \mathrm{Prob}(E) \to \mathbb{R}$ par $\mathcal{H}_p(p') = \sum_{x\in E} p(x) \log p'(x)$ est maximale pour $p' = p$. (max. unique)

(On suppose que $\mathcal{H}_p(p) > -\infty$. $-\mathcal{H}_p(p)$ est l'entropie de p.)

Calcul de la vraisemblance des observations pour $\hat{\theta}$ obtenu avec l'algo EM

On a vu que

$$\mathbb{P}_{\hat{\theta}}(Y_n = y_n \mid Y_{1:n-1} = y_{1:n-1}) = \sum_z \mathbb{P}_{\hat{\theta}}(X_n = z \mid Y_{1:n-1} = y_{1:n-1}) \times \hat{Q}(z, y_n)$$

$$\parallel$$

$$\frac{\mathbb{P}_{\hat{\theta}}(Y_{1:n} = y_{1:n})}{\mathbb{P}_{\hat{\theta}}(Y_{1:n-1} = y_{1:n-1})} \longrightarrow \text{calcul par récurrence, en passant au log} \ldots$$

car la vrais. très petite

En pratique : exécuter l'algo EM avec plusieurs points de départ et choisir $\hat{\theta}$ qui donne la plus grande vraisemblance !

---

② Estimer la suite des états cachés

A la fin de l'algo EM, on a accès aux probas

$\mathbb{P}_{\hat{\theta}}(X_k = x \mid Y_{1:n})$, $\forall x$, $\forall k$

Une solution possible : $\hat{X}_k = \underset{x}{\mathrm{Argmax}}\, \mathbb{P}_{\hat{\theta}}(X_k = x \mid Y_{1:n})$

(maximum a posteriori)

⊛ Une autre solution :

algo de Viterbi pour trouver

$$\hat{X}_{1:n} = \underset{x_{1:n}}{\mathrm{Argmax}}\, \mathbb{P}_{\hat{\theta}}(X_{1:n} = x_{1:n}, Y_{1:n})$$

(programmation dynamique)

algo récursif exact

Pb. : instable / n : si on enlève la dernière obs. y_n, $\hat{X}_{1:n}$ change complètement.

⊛ Variante de l'algo EM : SEM

↑ stochastic

- Étape S ("simulation") : On tire au hasard une réalisation $x_{1:n}^{(k)}$ de la suite $X_{1:n}$ suivant la loi $\mathbb{P}_{\theta_k}(X_{1:n} = \cdot \mid Y_{1:n})$
- Étape M : On choisit θ_{k+1} qui maximise

$$\theta_{k+1} = \underset{\theta}{\mathrm{Argmax}}\, \mathbb{P}_\theta(X_{1:n} = x_{1:n}^{(k)}, Y_{1:n} = y_{1:n})$$

Remarque : On a, pour $A_{k-1} \in \sigma(X_0, X_1, \dots, X_{k-1})$:

$$\mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k, A_{k-1}, Y_{1:n})$$
$$= \mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k, Y_{1:n}) .$$

Preuve (idée) :
Calculer $\mathbb{P}(X_{k+1} = x_{k+1} \cap X_k = x_k \cap A_{k-1} \cap Y_{1:n})$
et $\mathbb{P}(X_k = x_k \cap A_{k-1} \cap Y_{1:n})$
et m.q. le rapport ne dépend pas de A_{k-1}.

- Pour estimer les états cachés $X_{k:l}$, pour $1 \leq k < l \leq n$

on maximise $\mathbb{P}_{\hat{\theta}}(X_{k:l} = x_{k:l} \mid Y_{1:n})$.

En utilisant la remarque ci-dessus, on peut montrer (exo.)

que $$\mathbb{P}_{\hat{\theta}}(X_{k:l} = x_{k:l} \mid Y_{1:n}) = \mathbb{P}_{\hat{\theta}}(X_k = x_k, X_{k+1} = x_{k+1} \mid Y_{1:n}) \times$$
$$\times \frac{\mathbb{P}_{\hat{\theta}}(X_{k+1} = x_{k+1}, X_{k+2} = x_{k+2} \mid Y_{1:n})}{\mathbb{P}_{\hat{\theta}}(X_{k+1} = x_{k+1} \mid Y_{1:n})} \times \dots \times \frac{\mathbb{P}_{\hat{\theta}}(X_{l-1} = x_{l-1}, X_l = x_l \mid Y_{1:n})}{\mathbb{P}_{\hat{\theta}}(X_{l-1} = x_{l-1} \mid Y_{1:n})}$$

avec toutes ces probabilités calculées dans l'algo forward-backward.