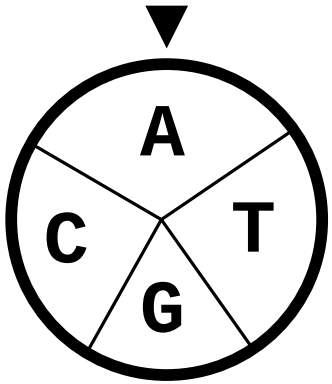
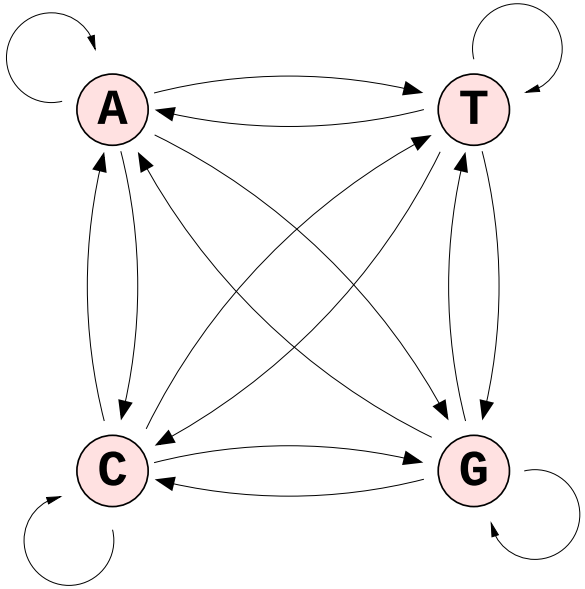
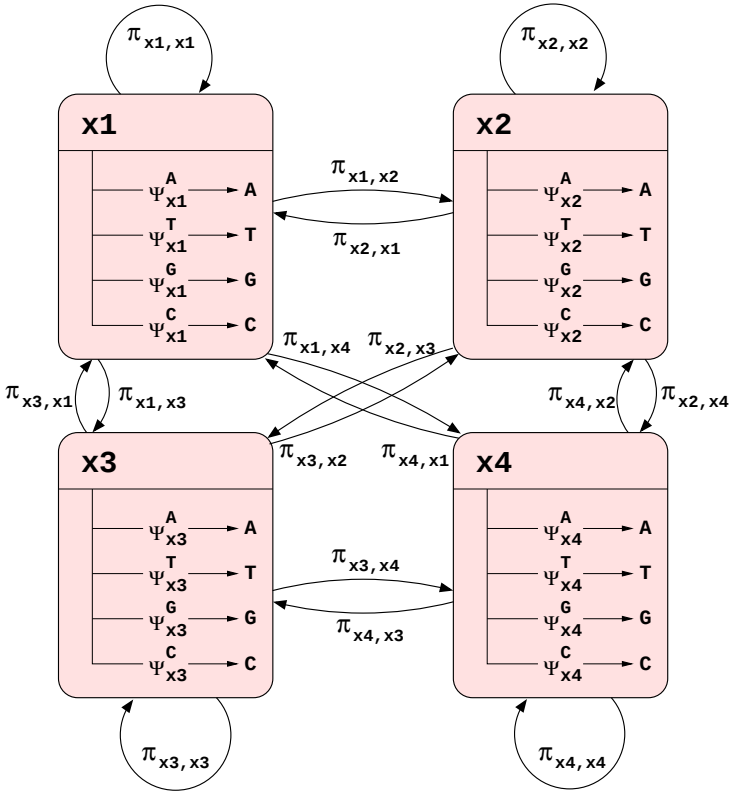
Modèles de Markov Cachés



fabien campillo



M1



M1MO

modèle de Markov *vs* modèle de Markov caché à 4 états cachés qui émettent chacun un des quatre symboles `A`, `C`, `G`, `T` selon des probabilités différentes

➜ pouvoir modéliser des séquences *"moins homogènes"* que les chaînes de Markov

étant données des observations $\mathsf{y}_{1:N}$, un modèle de Markov caché MℓMk suppose que :

- on dispose d'une séquence cachée d'état $\mathsf{X}_{1:N}$ qui est une chaîne de Markov d'ordre ℓ
- d'une séquence observée $\mathsf{Y}_{1:N}$ qui — conditionnellement à la séquence cachée $\mathsf{X}_{1:N}$ - est une chaîne de Markov d'ordre k

présenté autrement cela signifie que $\mathsf{Y}_{1:N}$ est une chaîne de Markov dont la probabilité de transition dépend d'un paramètre inconnu à valeurs dans $\mathcal{X}$ et que la suite cachée $\mathsf{X}_{1:N}$ de ces paramètres est une chaîne de Markov d'ordre ℓ

dans le transparent précédent, le schéma de gauche présente un modèle M1M0

$\mathsf{Y}_{1:N}$ est Markov d'ordre k dont la matrice de transition $\psi^{\mathsf{x}_n}_{\mathsf{y}_{n-k:n-1},\mathsf{y}_n}$ dépend de la valeur d'un processus caché $\mathsf{X}_{1:N} = \mathsf{x}_{1:N}$ qui est Markov d'ordre ℓ de matrice de transition $\pi_{\mathsf{x}_{n-\ell:n-1},\mathsf{x}_n}$, c–à–d :

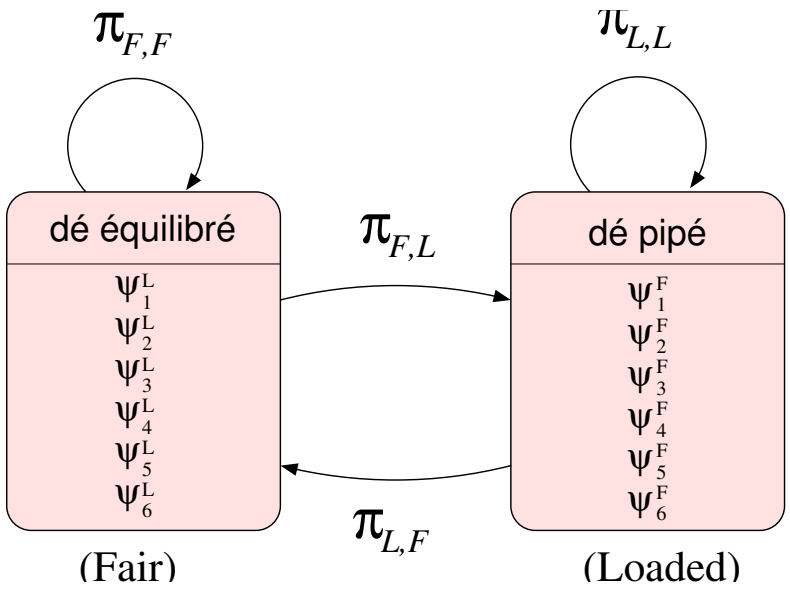
$$\begin{aligned}
&\mathbb{P}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}) \\
&= \prod_{n=1:N} \underbrace{\mathbb{P}(\mathsf{X}_n = \mathsf{x}_n | \mathsf{X}_{(n-\ell)\vee 1:n-1} = \mathsf{x}_{(n-\ell)\vee 1:n-1})}_{\text{matrice de transition d'ordre } \ell} = \prod_{n=1:N} \pi_{\mathsf{x}_{(n-\ell)\vee 1:n-1},\mathsf{x}_n}
\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N} | \mathsf{X}_{1:N} = \mathsf{x}_{1:N}) \\
&= \prod_{n=1:N} \underbrace{\mathbb{P}(\mathsf{Y}_n = \mathsf{y}_n | \mathsf{X}_n = \mathsf{x}_n, \mathsf{Y}_{(n-k)\vee 1:n-1} = \mathsf{y}_{(n-k)\vee 1:n-1})}_{\text{matrice de transition d'ordre } k \text{ : dépend d'un état caché } \mathsf{X}_n} = \prod_{n=1:N} \psi^{\mathsf{x}_n}_{\mathsf{y}_{(n-k)\vee 1:n-1},\mathsf{y}_n}
\end{aligned}$$

avec la notation/convention : $n \vee m = \inf\{n, m\}$ et si $n > m$ alors $\{n : m\} = \emptyset$ et $\{\mathsf{Y}_{n:m} = \mathsf{y}_{n:m}\} = \emptyset$ (mémo : $\mathbb{P}(B|\emptyset) = \mathbb{P}(B)$)

le modèle M1M0 est le plus simple des HMM, c'est celui que nous détaillons maintenant

♦ un croupier lance N fois un dé, on observe $\mathsf{y}_{1:N} = 623213453 \cdots 13$
♦ on le soupçonne de parfois utiliser un dé pipé : comment déterminer quand, au cours de ces lancés, a–t–il utilisé le dé pipé ?



la probabilité ψ associée au dé est inconnue et dépend d'un paramètre à valeur dans $\{F, L\}$ dont l'évolution est décrite par une chaîne de Markov de probabilité de transition π (ces paramètres sont inconnus)

➜ déterminer ces paramètres à partir de la séquence observée $\mathsf{y}_{1:N}$ permet de répondre au problème posé

● LE MODÈLE (M1M0)

$\boxed{\text{état non observé}}$ une chaîne de Markov $\mathsf{X}_{1:N}$ à valeurs dans $\mathcal{X}$ (fini)

$$\mathbb{P}(X_{1:n} = x_{1:n}) = \underbrace{\mathbb{P}(\mathsf{X}_1 = \mathsf{x}_1)}_{\substack{\text{probabilité initiale} \\ =:\nu_{\mathsf{x}_1}}} \times \prod_{n=1:N} \underbrace{\mathbb{P}(\mathsf{X}_n = \mathsf{x}_n | \mathsf{X}_{n-1} = \mathsf{x}_{n-1})}_{\substack{\text{probabilité de transition} \\ =:\pi_{\mathsf{x}_{n-1},\mathsf{x}_n}}}$$

$\boxed{\text{observation}}$ $\mathsf{Y}_{1:N}$ à valeurs dans $\mathcal{Y}$ (fini), une suite indépendante *conditionnellement* à $\mathsf{X}_{1:N}$, c–a–d

$$\mathbb{P}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N} | \mathsf{X}_{1:N} = \mathsf{x}_{1:N}) = \prod_{n=1:N} \underbrace{\mathbb{P}(\mathsf{Y}_n = \mathsf{y}_n | \mathsf{X}_n = \mathsf{x}_n)}_{\substack{\text{probabilité d'émission} \\ =:\psi_{\mathsf{y}_n}^{\mathsf{x}_n}}}$$

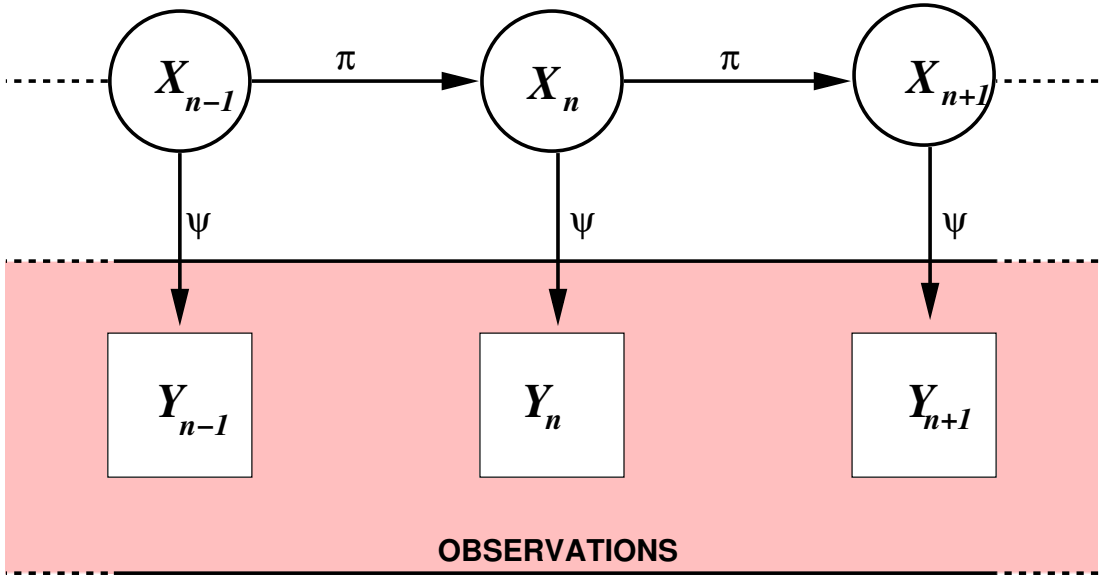
– l'occurrence de l'état $\mathsf{X}_n = x$ produit l'émission $\mathsf{Y}_n = y$ avec probabilité ψ_y^x

– les observations sont recueillies à travers un canal *sans mémoire*

● MODÈLE DE MARKOV CACHÉ

la loi de $(\mathsf{X}_{1:N}, \mathsf{Y}_{1:N})$ est entièrement déterminée par le modèle $\mathsf{M} = (\nu, \pi, \psi)$

- *loi initiale* $\nu_x = \mathbb{P}_{\mathsf{M}}(\mathsf{X}_1 = x)$
- *matrice de transition* $\pi_{x,x'} = \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{n+1} = x' | \mathsf{X}_n = x)$
- *matrice des probabilités d'observation* $\psi_y^x = \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_n = y | \mathsf{X}_n = x)$



$$\mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) = \nu_{\mathsf{x}_1}\, \pi_{\mathsf{x}_1,\mathsf{x}_2} \cdots \pi_{\mathsf{x}_{N-1},\mathsf{x}_N}\, \psi_{\mathsf{y}_1}^{\mathsf{x}_1} \cdots \psi_{\mathsf{y}_N}^{\mathsf{x}_N}\,,$$

ici $\mathsf{M} = (\pi, \psi)$: π décrit la transition du mode "honnête" au mode "malhonnête", p^F est la probabilité associée au dé équilibré (a priori c'est la probabilité uniforme), p^L est celle associée au dé pipé (la loi initiale est secondaire)

- ♦ le problème — noté ③ — est d'estimer M à partir des observations $\mathsf{y}_{1:N}$
- ♦ avant on va s'intéresser au problème — noté ② — d'estimation de l'état caché en fonction des observations : à chaque jet observé y_n on veut déterminer si le dé utilisé est pipé (L) ou non (F)
- ♦ le premier problème à se poser est de savoir comment quantifier l'adéquation entre un modèle M et les observations $\mathsf{y}_{1:N}$, il s'agit du problème ①

(ces problèmes sont ici dans l'ordre décroissant de difficulté)

● 3 PROBLÈMES

① *Évaluer le modèle* M *:* M donné ➜ calculer *efficacement*

$$\mathcal{L}(\mathsf{M}) = \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$$

② *Estimer l'état de la chaîne :* $\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}$ et M donnés ➜ calculer (de façon récursive) l'état présent X_n, ou bien intermédiaire X_ℓ pour $\ell = 1, \ldots, N$, ou la suite d'états $\mathsf{X}_{1:N}$

②–a $(n, x) \mapsto \mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:n} = \mathsf{y}_{1:n})$ (filtrage)

②–b $(n, x) \mapsto \mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$ (lissage)

②–c $\mathsf{x}_{1:N} \mapsto \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$ (trajectoire optimale)

③ *Identifier le modèle* M *:* $\mathsf{Y}_{1:N} = \mathsf{y}_{1:n}$ donnés ➜ calculer le "meilleur" M

$$\text{maximiser (localement) } \mathsf{M} \mapsto \mathcal{L}(\mathsf{M})$$

● Algorithmique HMM

- ♦ problème ① : équation *progressive* de Baum ➜ ①
- ♦ problèmes ② :
  - ✓ équation *progressive* de Baum ➜ ②–a
  - ✓ équations *progressive/rétrograde* de Baum ➜ ②–b
  - ✓ *algorithme de Viterbi* (programmation dynamique) avec les équations *progressive/rétrograde* de Baum ➜ ②–c
- ♦ problème ③ : *formules de réestimation* de Baum–Welsh (algorithme EM) avec les équations *progressive/rétrograde* de Baum) ➜ ③

♦ variables progressive α et rétrograde β de Baum

$$\begin{aligned}\alpha_x^n[\mathsf{y}_{1:n}] &= \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}, \mathsf{X}_n = x)\\ \beta_x^n[\mathsf{y}_{n+1:N}] &= \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{n+1:N} = \mathsf{y}_{n+1:N} | \mathsf{X}_n = x)\end{aligned}$$

$$\underbrace{\mathsf{y}_1\ \mathsf{y}_2 \cdots \mathsf{y}_n}_{\alpha_x^{n+1}[\mathsf{y}_{1:n}]}\underbrace{\mathsf{y}_{n+1} \cdots \mathsf{y}_{N-1}\ \mathsf{y}_N}_{\beta_x^{n+1}[\mathsf{y}_{n+1:N}]}$$

♦ équations progressive/rétrograde de Baum

$$\begin{aligned}\alpha_x^{n+1}[\mathsf{y}_{1:n+1}] &\text{ comme fonction de } \alpha_x^n[\mathsf{y}_{1:n}]\\ \beta_x^n[\mathsf{y}_{n+1:N}] &\text{ comme fonction de } \beta_x^{n+1}[\mathsf{y}_{n+2:N}]\end{aligned}$$

$$\mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) = \underbrace{\sum_{\mathsf{x}_{1:N} \in \mathcal{X}^N} \nu_{\mathsf{x}_1}\, \pi_{\mathsf{x}_1,\mathsf{x}_2} \cdots \pi_{\mathsf{x}_{N-1},\mathsf{x}_N}\, \psi_{\mathsf{y}_1}^{\mathsf{x}_1} \cdots \psi_{\mathsf{y}_N}^{\mathsf{x}_N}}_{(2N-1)\,|\mathcal{X}|^N \text{ opérations} \rightarrow \text{ trop lourd (exp. en } N)}$$

- *variable progressive* $\boxed{\alpha_x^n[\mathsf{y}_{1:n}] = \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}, \mathsf{X}_n = x)}$

  récurrence

$$\alpha_{x'}^{n+1}[\mathsf{y}_{1:n+1}] \leftarrow \psi_{\mathsf{y}_{n+1}}^{x'} \sum_{x \in \mathcal{X}} \pi_{x,x'}\, \alpha_x^n[\mathsf{y}_{1:n}] \quad \forall x' \in \mathcal{X}$$

  condition initiale

$$\alpha_x^1[\mathsf{y}_1] \leftarrow \nu_x\, \psi_{\mathsf{y}_1}^x \quad \forall x \in \mathcal{X}$$

- *résultat* $\boxed{\mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}) \leftarrow \sum_{x \in \mathcal{X}} \alpha_x^n[\mathsf{y}_{1:n}]}$
- *nombre d'opérations* $N\,|\mathcal{X}|\,(2|\mathcal{X}|+1) + (|\mathcal{X}|-1)$ (linéaire en N)

algorithme de Baum progressif

données:

observations $\mathsf{y}_{1:N}$

modèle $\mathsf{M} = (\nu, \pi, \psi)$

initialisation:

pour x dans $\mathcal{X}$ faire $\alpha_x^1 \leftarrow \nu_x \, \psi_{\mathsf{y}_1}^x$

itérations:

pour $n = 2$ à $N - 1$ faire

pour x dans $\mathcal{X}$ faire

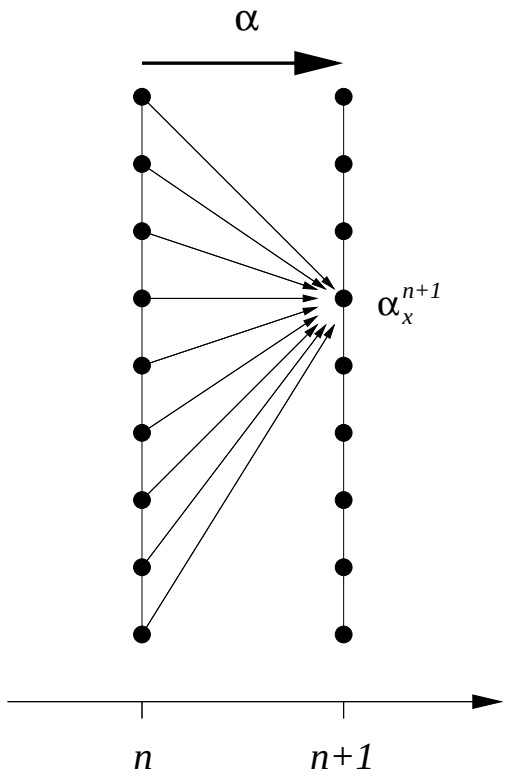
$\alpha_x^n \leftarrow \psi_x^{\mathsf{y}_n} \sum_{x' \in \mathcal{X}} \pi_{x',x} \, \alpha_{x'}^{n-1}$

fin faire

fin faire

calcul de la loi:

$\mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \leftarrow \sum_{x \in \mathcal{X}} \alpha_x^N$



● Variable Progressive Et Filtrage

♦ *loi de* X_n *sachant* $\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}$

$$x \mapsto \mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:n} = \mathsf{y}_{1:n}) = \frac{\alpha_x^n[\mathsf{y}_{1:n}]}{\sum_{x' \in \mathcal{X}} \alpha_{x'}^n[\mathsf{y}_{1:n}]} =: \bar{\alpha}_x^n[\mathsf{y}_{1:n}]$$

♦ *estimateur du maximum a posteriori (MAP)* de X_n sachant $\mathsf{Y}_{1:n} = \mathsf{y}_{1:n}$

$$\hat{\mathsf{X}}_n^{\mathsf{MAP}} \in \operatorname{Arg}\max_{x \in \mathcal{X}} \bar{\alpha}_{x'}^n[\mathsf{y}_{1:n}] = \operatorname{Arg}\max_{x \in \mathcal{X}} \alpha_{x'}^n[\mathsf{y}_{1:n}]$$

- étant donnée l'observation $\mathsf{y}_{1:N} = 623213453\cdots 13$ et un modèle $\mathsf{M} = (\nu, \pi, \psi)$ on calcule (efficacement) $\mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:N} = 623213453\cdots 13)$ à l'aide de l'algorithme de Baum progressif.
- On peut également calculer un estimation de l'état caché à l'aide du filtrage :

$$\begin{array}{rcccccccccc} \text{séquence observée} & 6 & 2 & 3 & 2 & 1 & 3 & 4 & 5 & 3 & \cdots \\ \hline & & & & & & & & & \downarrow & \cdots \\ \text{état estimé} & F & F & F & F & F & F & L & L & L & \cdots \end{array}$$

ceci se fait de manière *récursive* : pour calculer l'estimée à l'indice n il suffit de l'observation y_n et de l'estimée à l'indice $n-1$

- *variable rétrograde* $\boxed{\beta_x^n[\mathsf{y}_{n+1:N}] = \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{n+1:N} = \mathsf{y}_{n+1:N} | \mathsf{X}_n = x)}$

  récurrence (rétrograde)

$$\beta_x^n[\mathsf{y}_{n+1:N}] \leftarrow \sum_{x' \in \mathcal{X}} \pi_{x,x'} \, \psi_{x'}^{\mathsf{y}_{n+1}} \, \beta_{x'}^{n+1}[\mathsf{y}_{n+2:N}] \qquad (\forall x \in \mathcal{X})$$

  condition finale

$$\beta_x^N \leftarrow 1 \qquad (\forall x \in \mathcal{X})$$

- *dualité* $\sum_{x \in \mathcal{X}} \alpha_x^n[\mathsf{y}_{1:n}] \, \beta_x^n[\mathsf{y}_{n+1:N}]$ ne dépend pas de n
- *résultat*

$$\mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}, \mathsf{X}_n = x) \;=\; \boxed{\alpha_x^n[\mathsf{y}_{1:n}] \, \beta_x^n[\mathsf{y}_{n+1:N}] =: \gamma_x^n[\mathsf{y}_{1:N}]}$$

$$\mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \;=\; \frac{\gamma_x^n[\mathsf{y}_{1:N}]}{\sum_{x' \in \mathcal{X}} \gamma_{x'}^n[\mathsf{y}_{1:N}]} =: \bar{\gamma}_x^n[\mathsf{y}_{1:N}]$$

algorithme de Baum progressif/rétrograde

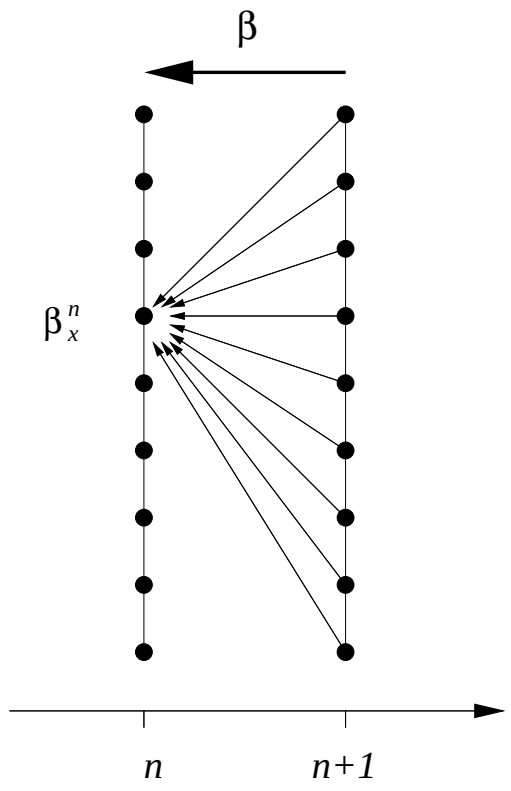
données:
- un instant $1 \leq n < N$
- observations $\mathsf{y}_{1:N}$
- modèle $\mathsf{M} = (\nu, \pi, \psi)$

on utilise l'algorithme progressif

initialisation:
- pour x dans $\mathcal{X}$ faire $\beta_x^N \leftarrow 1$

itérations:
- pour $\ell = N-1$ à $n+1$ faire
  - pour x dans $\mathcal{X}$ faire
    - $\beta_x^n \leftarrow \sum_{x' \in \mathcal{X}} \pi_{x,x'}\, \psi_{x'}^{\mathsf{y}_{\ell+1}}\, \beta_{x'}^{\ell+1}$
  - fin faire
- fin faire

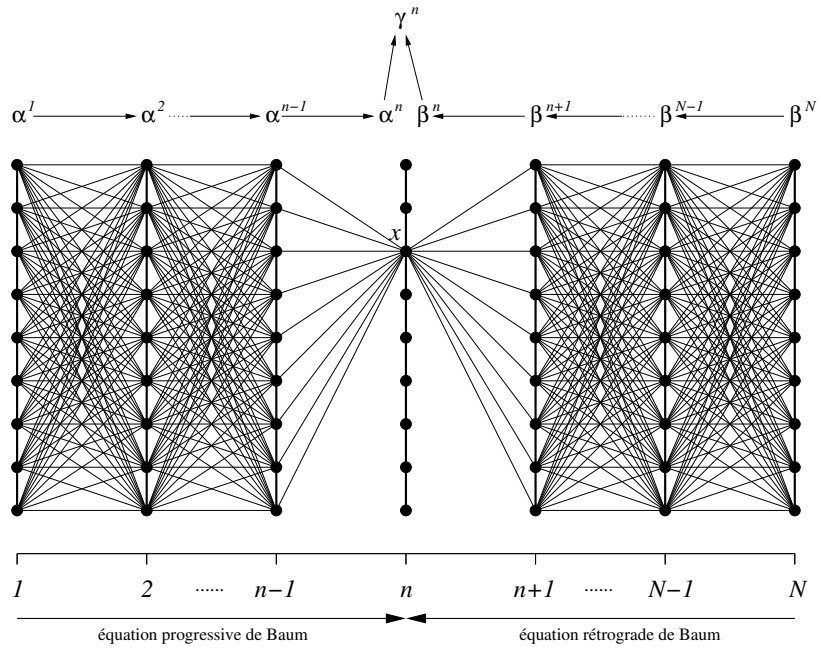
calcul de la loi:
- $C \leftarrow \sum_{x \in \mathcal{X}} \alpha_x^n\, \beta_x^{n+1}$
- $\mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \leftarrow C^{-1}\, \alpha_x^n\, \beta_x^{n+1}$



● Variables Progressive/Rétrograde et Lissage

✦ *loi de* X_n *sachant* $\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}$

$$x \mapsto \mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) = \bar{\gamma}_x^n[\mathsf{y}_{1:N}]$$



On peut calculer une nouvelle estimation de l'état caché à l'aide du lissage :

séquence observée	6	2	3	2	1	3	4	5	3	$\cdots$	1	3
état estimé par filtrage (Baum progressif)	F	F	F	F	F	F	L	L	L	$\cdots$	F	F
						$\downarrow$				$\cdots$		
état estimé par lissage (Baum rétrograde)					$\leftarrow$	F	L	L	L	$\cdots$	F	F

ceci ne peut se faire de manière récursive : l'horizon N est fixé, on effectue un "aller" avec l'équation progressive jusqu'à l'indice N et un "retour" avec l'équation rétrograde

- *estimateur local du maximum a posteriori (LMAP)*

$$\mathsf{X}_n^{\mathsf{LMAP}}(\mathsf{y}_{1:N}) \in \mathrm{Arg}\max_{x\in\mathcal{X}} \gamma_x^n[\mathsf{y}_{1:N}]$$

- *trajectoire "incohérente" :* $\mathsf{X}_{1:N}^{\mathsf{LMAP}}$ peut être incohérente avec le modèle
  on peut avoir $\mathsf{X}_n^{\mathsf{LMAP}} = x$ et $\mathsf{X}_{n+1}^{\mathsf{LMAP}} = x'$ alors que $\pi_{x,x'} = 0$ (la transition de x vers x' est *impossible* pour le modèle M)

- *estimateur du maximum a posteriori (MAP)*

$$\mathsf{X}_{1:N}^{\mathsf{MAP}}(\mathsf{y}_{1:N}) = \mathrm{Arg}\max_{\mathsf{x}_{1:N}\in\mathcal{X}^N} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$$

➜ *algorithme de Viterbi :* calcul "efficace" de $\mathsf{X}_{1:N}^{\mathsf{MAP}}(\mathsf{y}_{1:N})$

♦ formule de Bayes ⇒

$$\mathrm{Arg}\max_{\mathsf{x}_{1:N}} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) = \mathrm{Arg}\max_{\mathsf{x}_{1:N}} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$$

♦ *fonction valeur* $\delta_x^n[\mathsf{y}_{1:n}] := \max_{\mathsf{x}_{1:n-1}} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:n-1} = \mathsf{x}_{1:n-1}, X_n = x, \mathsf{Y}_{1:n} = \mathsf{y}_{1:n})$

♦ *programmation dynamique*

$$\delta_{x'}^{n+1}[\mathsf{y}_{1:n+1}] \leftarrow \psi_{x'}^{\mathsf{y}_{n+1}} \times \max_{x \in \mathcal{X}} \pi_{x,x'}\, \delta_x^n[\mathsf{y}_{1:n}] \quad , \qquad \delta_x^1[\mathsf{y}_1] = \nu_x\, \psi_x^{\mathsf{y}_1}$$

on pose

$$I_n(x') = \mathrm{Arg}\max_{x \in \mathcal{X}} \{\pi_{x,x'}\, \delta_x^n[\mathsf{y}_{1:n}]\}$$

– parmi toutes les trajectoires qui aboutissent dans l'état x' à l'instant $n+1$, la trajectoire de plus grande probabilité est passée dans l'état $I_n(x')$ à l'instant précédent n

– la transition de l'état $I_n(x')$ vers l'état x' est possible pour le modèle : $\pi_{I_n(x'),x'} > 0$

♦ *trajectoire optimale*

$$\begin{aligned}\max_{\mathsf{x}_{1:N}} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) &= \\ &= \max_{x} \max_{\mathsf{x}_{1:N-1}} \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N-1} = \mathsf{x}_{1:N-1}, \mathsf{X}_N = x, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \\ &= \max_{x} \delta_x^N[\mathsf{y}_{1:N}] \,.\end{aligned}$$

donc

$$\mathsf{X}_N^{\mathsf{MAP}} = \operatorname{Arg}\max_{x \in \mathcal{X}} \delta_x^N[\mathsf{y}_{1:N}]$$

de proche en proche, on en déduit que la trajectoire optimale est passée dans l'état

$$\mathsf{X}_n^{\mathsf{MAP}} = I_n(\mathsf{X}_{n+1}^{\mathsf{MAP}})$$

à l'instant $1 \leq n < N$

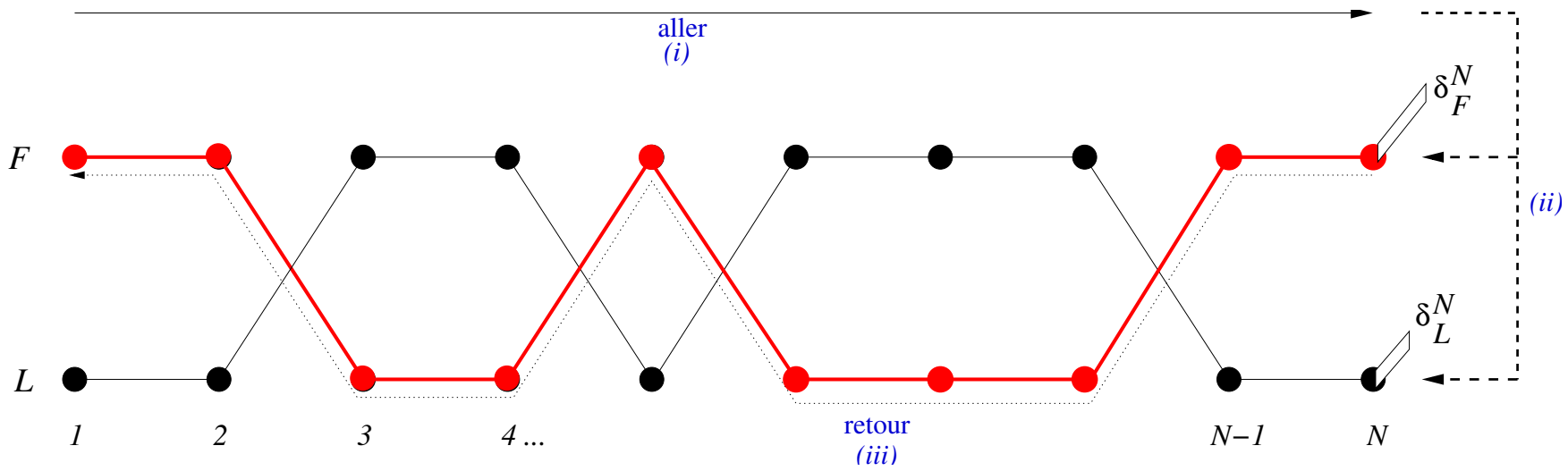
algorithme de Viterbi

données:
- observations $\mathsf{y}_{1:N}$
- modèle $\mathsf{M} = (\nu, \pi, \psi)$

initialisation:
- pour x dans $\mathcal{X}$ faire $\delta_x^1 \leftarrow \nu_x \, \psi_x^{\mathsf{y}_1}$

itérations:
- pour $n = 1$ à $N - 1$ faire
  - pour x' dans $\mathcal{X}$ faire
    - $I_n(x') \leftarrow \mathrm{Arg\,max}_{x \in \mathcal{X}} \{\pi_{x,x'} \, \delta_x^n\}$
    - $\delta_{x'}^{n+1} \leftarrow \psi_{x'}^{\mathsf{y}_{n+1}} \, \pi_{I_n(x'),x'} \, \delta_{I_n(x')}^n$
  - fin faire
- fin faire

calcul de la trajectoire optimale:
- $\mathsf{X}_N^{\mathsf{MAP}} \leftarrow \mathrm{Arg\,max}_{x \in \mathcal{X}} \, \delta_x^N$
- pour $n = N - 1$ à 1 faire
  - $\mathsf{X}_n^{\mathsf{MAP}} \leftarrow I_n(\mathsf{X}_{n+1}^{\mathsf{MAP}})$
- fin faire



(i) dans la phase aller : pour $x \in \{F, L\}$ on calcule $I_n(x)$ et δ_x^{n+1} $(n = 1, 2, \ldots, N)$

(ii) à la fin de la séquence : on calcule δ_F^N et δ_L^N le premier des deux est le plus grand

(iii) dans la phase retour : partant de F de proche en proche et à reculons on calcule la trajectoire optimale $(n = N-1, N-2, \ldots, 1)$

● Maximum De Vraisemblance

M qui maximise (localement)

$$\mathsf{M} \mapsto \mathcal{L}(\mathsf{M}) \triangleq \mathbb{P}_{\mathsf{M}}(\mathsf{Y}_{1:n} = \mathsf{y}_{1:n})$$

➜ méthode de réestimation de Baum–Welsh (forme de l'algorithme EM)

$$\xi^n_{x,x'}[\mathsf{y}_{1:N}] := \mathbb{P}_{\mathsf{M}}(\mathsf{X}_n = x, \mathsf{X}_{n+1} = x' | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$$

(se calcule en fonction de α et β)

$$\begin{aligned}
\sum_{n=1:N} \sum_{x' \in \mathcal{X}} \xi^n_{x,x'}[\mathsf{y}_{1:N}] &= \sum_{n=1:N} \bar{\gamma}^n_x[\mathsf{y}_{1:N}] = \mathbb{E}_{\mathsf{M}}\Big(\sum_{n=1:N} \mathbf{1}_{\mathsf{X}_n = x} \Big| \mathsf{Y}_{1:N} = \mathsf{y}_{1:N} \Big) \\
\sum_{n=1:N} \sum_{x' \in \mathcal{X}} \xi^n_{x,x'}[\mathsf{y}_{1:N}]\, \mathbf{1}_{\mathsf{y}_n = y} &= \sum_{n=1:N} \bar{\gamma}^n_x[\mathsf{y}_{1:N}]\, \mathbf{1}_{\mathsf{y}_n = y} = \mathbb{E}_{\mathsf{M}}\Big(\sum_{n=1:N} \mathbf{1}_{\mathsf{X}_n = x, \mathsf{Y}_n = y} \Big| \mathsf{Y}_{1:N} = \mathsf{y}_{1:N} \Big) \\
\sum_{n=1:N-1} \xi^n_{x,x'}[\mathsf{y}_{1:N}] &= \mathbb{E}_{\mathsf{M}}\Big(\sum_{n=1:N-1} \mathbf{1}_{\{\mathsf{X}_n = x, \mathsf{X}_{n+1} = x'\}} \Big| \mathsf{Y}_{1:N} = \mathsf{y}_{1:N} \Big)
\end{aligned}$$

● Le Point Clef de l'Algo EM

$$\boxed{\log \mathcal{L}(\mathsf{M}) - \log \mathcal{L}(\mathsf{M}_0) = \frac{\mathcal{Q}(\mathsf{M}_0, \mathsf{M}) - \mathcal{Q}(\mathsf{M}_0, \mathsf{M}_0)}{\mathbb{P}_{\mathsf{M}_0}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N})} + \mathcal{H}(\mathsf{M}_0, \mathsf{M})}$$

où

$$\begin{aligned}
\mathcal{Q}(\mathsf{M}_0, \mathsf{M}) &= \sum_{\mathsf{x}_{1:N} \in \mathcal{X}^N} \mathbb{P}_{\mathsf{M}_0}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \log \mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N}, \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \\
\mathcal{H}(\mathsf{M}_0, \mathsf{M}) &= \sum_{\mathsf{x}_{1:N} \in \mathcal{X}^N} \mathbb{P}_{\mathsf{M}_0}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) \log \frac{\mathbb{P}_{\mathsf{M}_0}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}{\mathbb{P}_{\mathsf{M}}(\mathsf{X}_{1:N} = \mathsf{x}_{1:N} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}
\end{aligned}$$

$\mathcal{H}(\mathsf{M}_0, \mathsf{M})$ est *l'entropie relative (ou "distance" de Kullback–Leibler)* des lois conditionnelles de $\mathsf{X}_{1:N}$ sachant $\mathsf{Y}_{1:N}$ sous le modèle M_0 et sous le modèle M (c'est positif et nul ssi les lois sont identiques)

$$\mathcal{Q}(\mathsf{M}_0, \mathsf{M}) \geq \mathcal{Q}(\mathsf{M}_0, \mathsf{M}_0) \quad \Rightarrow \quad \mathcal{L}(\mathsf{M}) \geq \mathcal{L}(\mathsf{M}_0)$$

● l'Algo EM

algo itératif de maximisation (locale) de la vraisemblance :

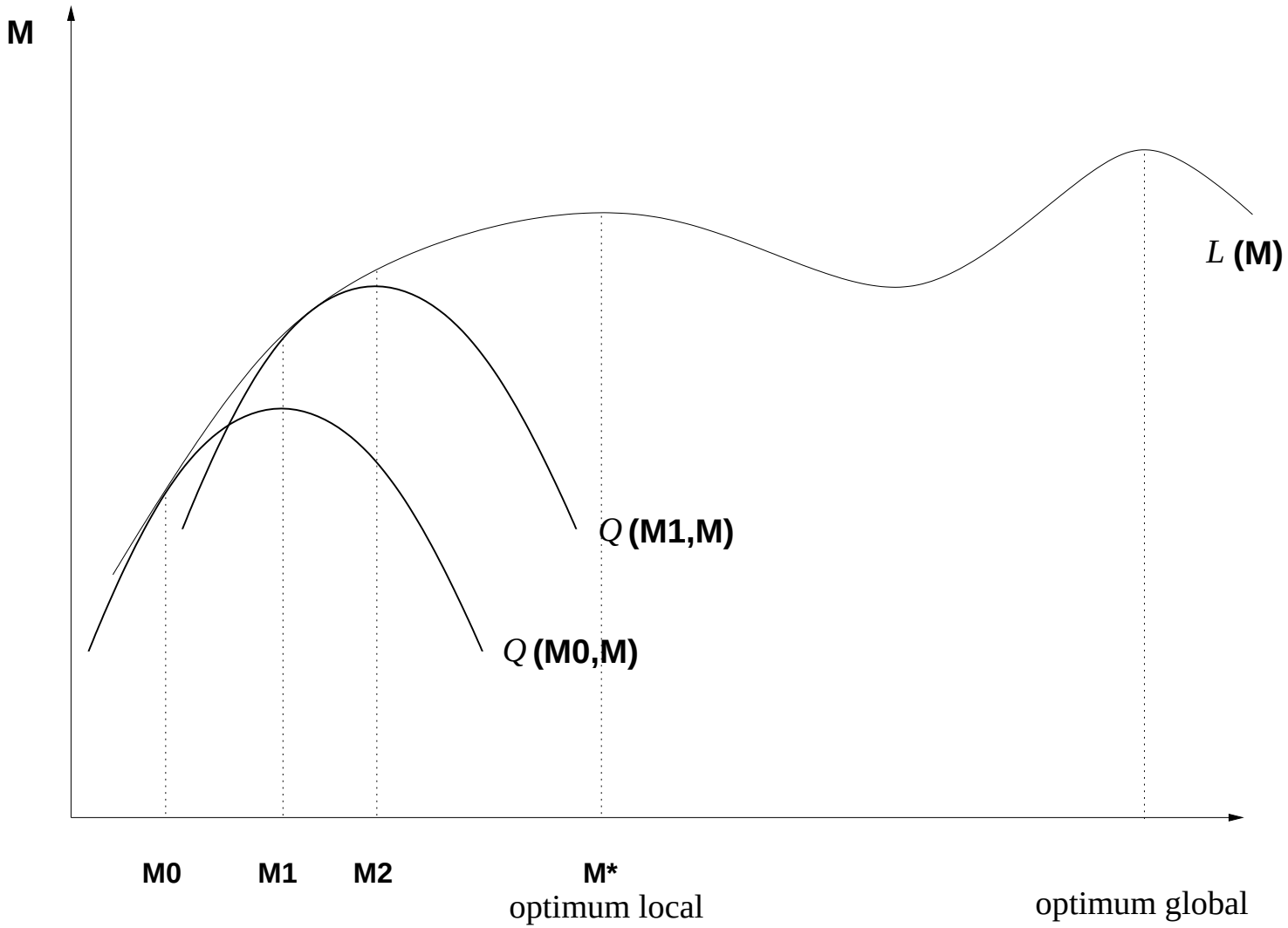
$$\mathsf{M}^{(k+1)} \leftarrow \operatorname{Arg}\max_{\mathsf{M}} \mathcal{Q}(\mathsf{M}^{(k)}, \mathsf{M})$$

l'itération $(k) \to (k+1)$ se fait en deux étapes

E (expectation) calcul de la fonction $\mathsf{M} \mapsto \mathcal{Q}(\mathsf{M}_0, \mathsf{M})$

M (maximization) maximisation de la fonction $\mathsf{M} \mapsto \mathcal{Q}(\mathsf{M}_0, \mathsf{M})$

n'a d'intérêt que si ces étapes sont explicites ou simples (ici l'étape (E) est simple, l'étape (M) est explicite)


M
L (M)
Q (M1,M)
Q (M0,M)
M0
M1
M2
M*
optimum local
optimum global

réestimation $\mathsf{M}^{(k)} = (\nu, \pi, \psi) \to (\bar{\nu}, \bar{\pi}, \bar{\psi}) = \mathsf{M}^{(k+1)}$

$$\begin{aligned}
\bar{\nu}_x[\mathsf{y}_{1:N}] &= \xi^1_x[\mathsf{y}_{1:N}] \\
\bar{\pi}_{x,x'}[\mathsf{y}_{1:N}] &= \frac{\sum_{n=1:N-1} \xi^n_{x,x'}[\mathsf{y}_{1:N}]}{\sum_{n=1:N-1} \bar{\gamma}^n_x[\mathsf{y}_{1:N}]} \\
\bar{\psi}^y_x[\mathsf{y}_{1:N}] &= \frac{\sum_{n=1:N-1} \bar{\gamma}^n_x[\mathsf{y}_{1:N}]\, \mathbf{1}_{\mathsf{y}_n=y}}{\sum_{n=1:N-1} \bar{\gamma}^n_x[\mathsf{y}_{1:N}]}
\end{aligned}$$

interprétation

$$\begin{aligned}
\bar{\nu}_x[\mathsf{y}_{1:N}] &= \frac{\mathbb{E}_{\mathsf{M}^{(k)}}(\mathbf{1}_{\mathsf{X}_1=x} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}{|\mathcal{X}|} \\
\bar{\pi}_{x,x'}[\mathsf{y}_{1:N}] &= \frac{\mathbb{E}_{\mathsf{M}^{(k)}}(\sum_{n=1:N-1} \mathbf{1}_{\mathsf{X}_n=x}\, \mathbf{1}_{\mathsf{X}_{n+1}=x'} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}{\mathbb{E}_{\mathsf{M}^{(k)}}(\sum_{n=1:N-1} \mathbf{1}_{\mathsf{X}_n=x} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})} \\
\bar{\psi}^y_x[\mathsf{y}_{1:N}] &= \frac{\mathbb{E}_{\mathsf{M}^{(k)}}(\sum_{n=1:N-1} \mathbf{1}_{\mathsf{X}_n=x}\, \mathbf{1}_{\mathsf{Y}_n=y} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}{\mathbb{E}_{\mathsf{M}^{(k)}}(\sum_{n=1:N-1} \mathbf{1}_{\mathsf{X}_n=x} | \mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}
\end{aligned}$$

● Convergence De l'Algorithme EM

– soit $\mathsf{M}^{(k)}$ est un point critique de la fonction de vraisemblance

$$\nabla_{\mathsf{M}}\mathcal{L}(\mathsf{M})|_{\mathsf{M}=\mathsf{M}^{(k)}} = 0$$

et alors $\mathsf{M}^{(k+1)} = \mathsf{M}^{(k)}$

– soit le modèle $\mathsf{M}^{(k+1)}$ est (strictement) plus vraisemblable au sens où

$$\mathcal{L}(\bar{\mathsf{M}}^{(k+1)}) > \mathcal{L}(\mathsf{M}^{(k)})$$

donc $\lim_{k\uparrow\infty} \mathsf{M}^{(k)} = \hat{\mathsf{M}}$ maximum local de la fonction de vraisemblance

➜ on passe tout au log !

- M1M1 Les Îlots C+G

	A	T	G	C
A	0.30	0.21	0.28	0.21
T	0.17	0.30	0.30	0.23
G	0.25	0.20	0.30	0.25
C	0.32	0.23	**0.08**	0.30

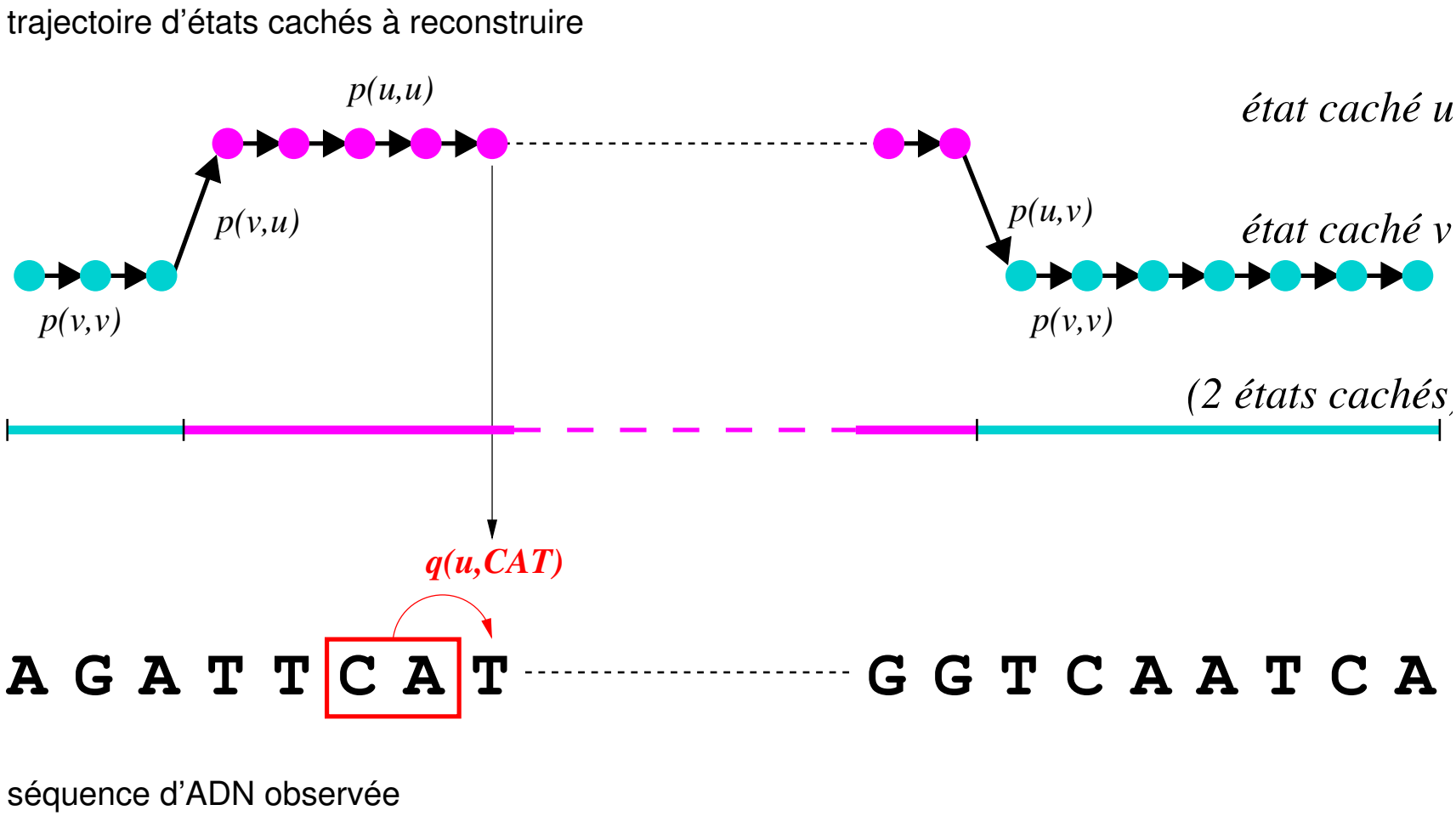
zone sans îlot : dans le nucléotide CG le C mute souvent en T (méthylation)

$\downarrow\uparrow$

	A	T	G	C
A	0.18	0.12	0.43	0.27
T	0.08	0.18	0.38	0.36
G	0.16	0.13	0.37	0.34
C	0.17	0.19	**0.27**	0.37

zone à îlots : dans les zones du génome précédant un gène, le phénomène de méthylation disparaît, la proportion de dinucléotides CG est donc plus importante (notation C+G ou CpG)

● M1M2 identifier des zones homogènes dans une séquence de trinucléotides ("à la Florence Muri" → RHOM)



- dans une séquence d'ADN un "motif" détermine une fonction

séq. 1	A C A A T G
séq. 2	T C A A C T A T C
séq. 3	A C A C A G C
séq. 4	A G A A T C
séq. 5	A C C G A T C

- Alignement Multiple

séq. 1	A C A - - - A T G
séq. 2	T C A A C T A T C
séq. 3	A C A C - - A G C
séq. 4	A G A - - - A T C
séq. 5	A C C G - - A T C

● EXPRESSION RÉGULIÈRE

site	1	2	3	insertion			4	5	6
séq. 1	A	C	A	-	-	-	A	T	G
séq. 2	T	C	A	A	C	T	A	T	C
séq. 3	A	C	A	C	-	-	A	G	C
séq. 4	A	G	A	-	-	-	A	T	C
séq. 5	A	C	C	G	-	-	A	T	C

donne *l'expression régulière* $\underbrace{\texttt{[AT]}}_{\text{site 1}}\ \underbrace{\texttt{[GC]}}_{\text{site 2}}\ \underbrace{\texttt{[AC]}}_{\text{site 3}}\ \underbrace{\texttt{[ATGC]*}}_{\text{site d'insertion}}\ \underbrace{\texttt{A}}_{\text{site 4}}\ \underbrace{\texttt{[TG]}}_{\text{site 5}}\ \underbrace{\texttt{[GC]}}_{\text{site 6}}$

- ♦ `[AT]` est `A` ou `T`
- ♦ `[ATGC]*` est n'importe quelle lettre, n'importe quelle longueur
- ♦ de l'expression régulière on tire :
  - le consensus : `A C A C - - A T C`
  - l'exceptionnelle : `T G C T - - A G G`

● LES ETATS

A	C	A	-	-	-	A	T	G	pas d'insertion
T	C	A	A	C	T	A	T	C	3 insertions
A	C	A	C	-	-	A	G	C	1 insertion
A	G	A	-	-	-	A	T	C	pas d'insertion
A	C	C	G	-	-	A	T	C	1 insertion
e_1	e_2	e_3	e_4			e_5	e_6	e_7	*7 états : e_4 représente l'état "insertion"*

● LES PROBABILITÉS DE TRANSITION

♦ les transitions $e_1 \to e_2$, $e_2 \to e_3$, $e_5 \to e_6$, $e_6 \to e_7$ se font avec probabilité 1

♦ dans les séq. 1 et 4 on passe directement de e_3 à e_5 dans les autres on passe de e_3 à e_4 :

$$\mathbb{P}(\mathsf{X}_4 = e_5 | \mathsf{X}_3 = e_3) = \tfrac{2}{5} = 1 - \mathbb{P}(\mathsf{X}_4 = e_4 | \mathsf{X}_3 = e_3)$$

♦ quand on est en "zone d'insertion" (état e_4), on a deux possibilités de rester en "insertion" contre 3 d'en sortir. Donc 5 possibilités en tout :

$$\mathbb{P}(\mathsf{X}_5 = e_4 | \mathsf{X}_4 = e_4) = \tfrac{2}{5} = 1 - \mathbb{P}(\mathsf{X}_5 = e_5 | \mathsf{X}_4 = e_4)$$

- Les Probabilités d'Emission

A	C	A	-	-	-	A	T	G
T	C	A	A	C	T	A	T	C
A	C	A	C	-	-	A	G	C
A	G	A	-	-	-	A	T	C
A	C	C	G	-	-	A	T	C

$\downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow$

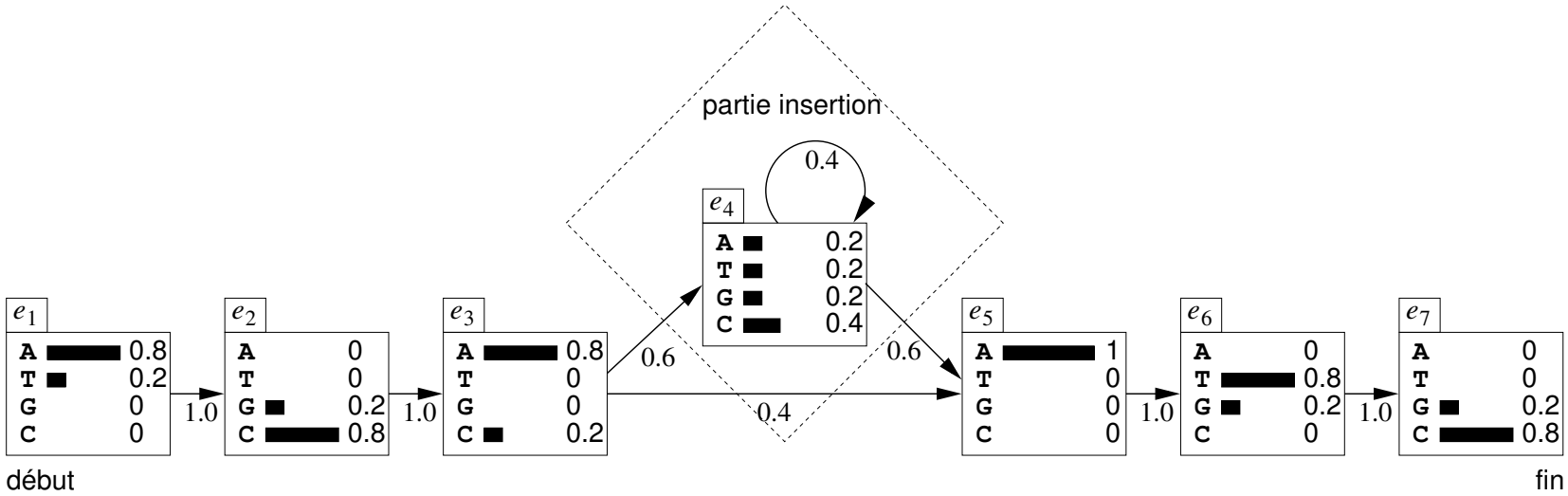
4	0	4	.	.	.	5	0	0
1	0	0	.	.	.	0	4	0
0	1	0	.	.	.	0	1	1
0	4	1	.	.	.	0	0	4

$\downarrow$

1
1
1
2

Dans l'état d'insertion e_4 on compte chacune des lettres dans toute la zone (ici 1, 1, 1, 2), pour les autres états on comptes les lettres dans la colonne. Ensuite on normalise.

● DE L'EXPRESSION RÉGULIÈRE À L'HMM



$$
\begin{aligned}
\mathbb{P}(\texttt{ACACATC}) &= \mathbb{P}(\mathsf{Y}_1 = \texttt{A}, \mathsf{Y}_2 = \texttt{C}, \mathsf{Y}_3 = \texttt{A}, \mathsf{Y}_4 = \texttt{C}, \mathsf{Y}_5 = \texttt{A}, \mathsf{Y}_6 = \texttt{T}, \mathsf{Y}_7 = \texttt{C}) \\
&= \mathbb{P}(\mathsf{X}_1 = e_1)\,\mathbb{P}(\mathsf{X}_2 = e_2|\mathsf{X}_1 = e_1)\,\mathbb{P}(\mathsf{X}_3 = e_3|\mathsf{X}_2 = e_2)\,\mathbb{P}(\mathsf{X}_4 = e_4|\mathsf{X}_3 = e_3)\,\mathbb{P}(\mathsf{X}_5 = e_5|\mathsf{X}_4 = e_4) \\
&\qquad \mathbb{P}(\mathsf{X}_6 = e_6|\mathsf{X}_5 = e_5)\,\mathbb{P}(\mathsf{X}_7 = e_7|\mathsf{X}_6 = e_6) \\
&\quad \mathbb{P}(\mathsf{Y}_1 = \texttt{A}|\mathsf{X}_1 = 1)\,\mathbb{P}(\mathsf{Y}_2 = \texttt{C}|\mathsf{X}_2 = 2)\,\mathbb{P}(\mathsf{Y}_3 = \texttt{A}|\mathsf{X}_3 = 3)\,\mathbb{P}(\mathsf{Y}_4 = \texttt{C}|\mathsf{X}_4 = 4) \\
&\qquad \mathbb{P}(\mathsf{Y}_5 = \texttt{A}|\mathsf{X}_5 = 5)\,\mathbb{P}(\mathsf{Y}_6 = \texttt{T}|\mathsf{X}_6 = 6)\,\mathbb{P}(\mathsf{Y}_6 = \texttt{T}|\mathsf{X}_7 = 7) \\
&= (1.0 \times 1.0 \times 1.0 \times 0.6 \times 0.6 \times 1.0 \times 1.0) \times (0.8 \times 0.8 \times 0.8 \times 0.4 \times 1.0 \times 0.8 \times 0.8) \simeq 0.047 \\
\mathbb{P}(\texttt{TGCTAGG}) &\simeq 0.000023
\end{aligned}
$$

➜ dépend de la longueur de la suite (et sont petits)

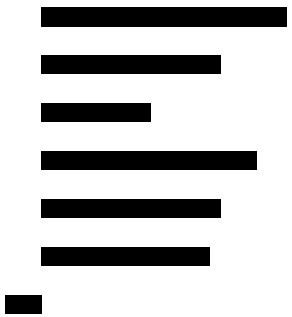
● SCORE LOG–ODDS

$$\text{score}(\mathsf{y}_{1:N}) = \log \frac{\mathbb{P}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N})}{\mathbb{P}_0(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N})} = \log \mathbb{P}(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) - \log \mathbb{P}_0(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N})$$

$\mathbb{P}$: probabilité sous le modèle en cours

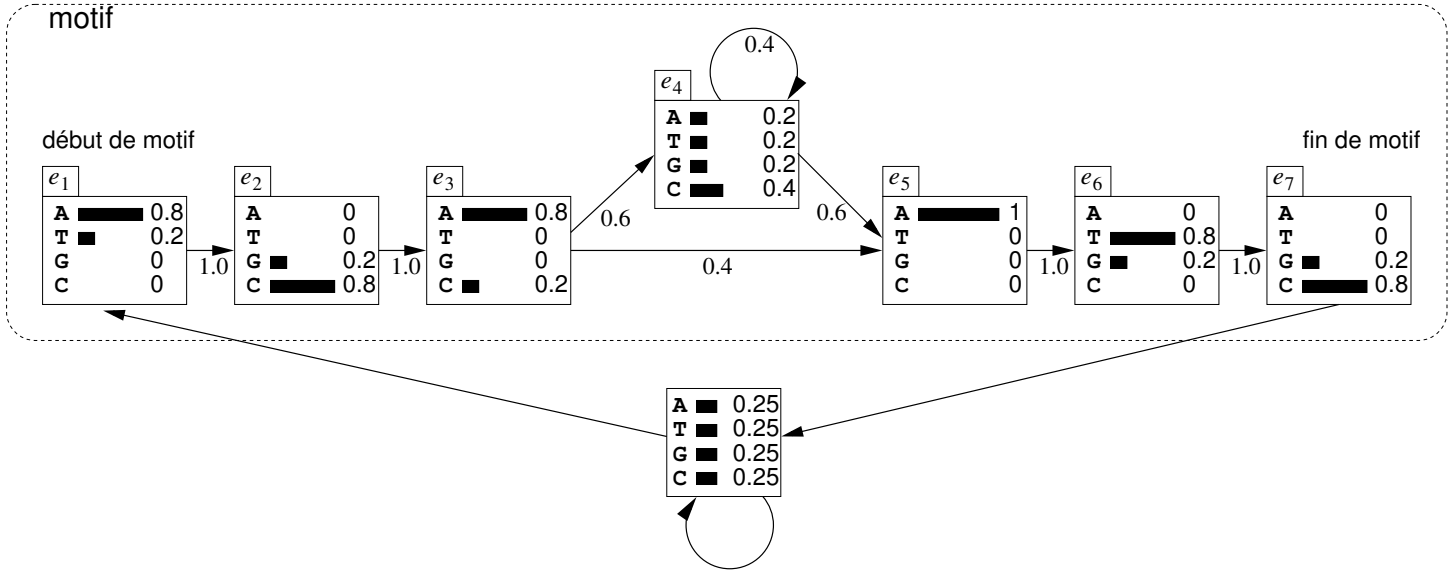
$\mathbb{P}_0$: probabilité sous le modèle nul (i.i.d $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$) c–à–d quelque soit la séquence observée $\log \mathbb{P}_0(\mathsf{Y}_{1:N} = \mathsf{y}_{1:N}) = N \log \frac{1}{4}$ (le modèle nul est plus souvent la fréquence moyenne des nucléotides dans toute la séquence)

	séquences	probabilités	score log–odds
consensus	A C A C - - A T C	0.047	6.7
séquences	A C A - - - A T G	0.033	4.9
d'origine	T C A A C T A T C	0.00075	3.0
	A C A C - - A G C	0.012	5.9
	A G A - - - A T C	0.033	4.9
	A C C G - - A T C	0.0059	4.6
exceptionnelle	T G C T - - A G G	0.000023	-0.97



● Localisation du Motif dans des séquences quelconques

- ♦ en dehors du motif la distribution des nucléotides est uniforme et indépendante
- ♦ les motifs ne se chevauchent pas



- ♦ on utilise alors l'algo de Viterbi

● LES BASES DE HMM autour de Leonard Baum

1966 — Baum, Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.

1967 — Baum, Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the AMS*, 73:360–363.

1968 — Baum, Sell. Growth transformations for functions on manifolds. *Pacific Journal of Mathematics*, 27(2):211–227.

1970 — Baum, Petrie, Soules, Weiss. A maximization technique in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.

1971 — Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In Oved Shisha (ed.), *Inequalities III, Los Angeles 1969*, pages 1–8. Academic Press.

● LA RÉFÉRENCE autour de Lawrence Rabiner

1993 — Rabiner, Juang. *Fundamentals of Speech Recognition*. Prentice–Hall, Englewood Cliffs.

Articles De Revue Et Ouvrages

1988 — Poritz. Hidden Markov models: a guided tour. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1988*, pages 7–13. IEEE–SPS.

1989 — Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

1995 — Elliott, Aggoun, Moore. *Hidden Markov Models: Estimation and Control*. Springer–Verlag.

1995 — Guttorp. *Stochastic modeling of scientific data*. Chapman & Hall.

1997 — MacDonald, Zucchini. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman & Hall.

HMM En Génomique

1998 — Durbin, Eddy, Krogh, Mitchison. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press.

Algorithme EM

1977 — Dempster, Laird, Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

1997 — McLachlan, Krishnan. *The EM algorithm and extensions*. John Wiley,