Noyaux marginalisés

Cours Master 2004/05

Jean-Philippe Vert

Jean-Philippe.Vert@mines.org

Plan

- Noyaux marginalisés
- Noyaux pour séquences par chaînes de Markov cachées
- Noyaux pour graphes

Noyaux marginalisés

Motivations

- On se place dans le cas ou les données $\mathbf{x} \in \mathcal{X}$ ne sont que la *partie "visible" d'objets plus complexes* $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{Z} := \mathcal{X} \cup \mathcal{Y}$.
- $\mathbf{y} \in \mathcal{Y}$ n'est *pas observé*, mais contient beaucoup d'*information pertinente* sur $\mathbf{x}$ (ex: $\mathbf{x}$ est une phrase, $\mathbf{y}$ sa structure grammatical)
- On a un noyau naturel sur $\mathcal{Z}$ (*les données completes*).
- Comment en déduire un noyau sur $\mathcal{X}$ (*les données observées*)?

Cadre probabiliste

- Etant donné une observation $\mathbf{x} \in \mathcal{X}$, la donnée cachée correspondante $\mathbf{y} \in \mathcal{Y}$ ne peut pas être déduite de maniere certaine.
- On se place donc dans un *cadre probabiliste*: $(\mathcal{Y}, \mathcal{B})$ est un espace mesurable, et on suppose que pour chaque $\mathbf{x} \in \mathcal{X}$, on a une probabilité "conditionnelle" $P_{\mathbf{x}} \in \mathcal{M}_1^+ (\mathcal{Y}, \mathcal{B})$ (une mesure positive de masse 1).

Noyau marginalisé

Définition 1 *Si $K_{\mathcal{Z}}$ est une fonction sur $\mathcal{Z}^2 = (\mathcal{X} \times \mathcal{Y})^2$, mesurable en tant que fonction de $\mathbf{y}$ pour tout $\mathbf{x}$, l'opération de* marginalisation *consiste a définir la fonction $K_{\mathcal{X}}$ sur $\mathcal{X}^2$ par:*

$$K_{\mathcal{X}}\left(\mathbf{x}, \mathbf{x}'\right) := E_{P_{\mathbf{x}}(d\mathbf{y}) \times P_{\mathbf{x}'}(d\mathbf{y}')} K_{\mathcal{Z}}\left(\mathbf{z}, \mathbf{z}'\right)$$
$$= \int \int K_{\mathcal{Z}}\left(\left(\mathbf{x}, \mathbf{y}\right), \left(\mathbf{x}', \mathbf{y}'\right)\right) P_{\mathbf{x}}\left(d\mathbf{y}\right) P_{\mathbf{x}'}\left(d\mathbf{y}'\right)$$

Théorème 2 *Si $K_{\mathcal{Z}}$ est un noyau défini positif sur $\mathcal{Z}$, alors $K_{\mathcal{X}}$ est un noyau défini positif sur $\mathcal{X}$ appelé* noyau marginalisé*.*

Preuve du théoreme 1

Si $K_{\mathcal{Z}}$ est un n.d.p. sur $\mathcal{Z}$, alors il existe un espace de Hilbert $\mathcal{H}$ et $\Phi_{\mathcal{Z}} : \mathcal{Z} \to \mathcal{H}$ tel que

$$K_{\mathcal{Z}}\left(\mathbf{z}, \mathbf{z}'\right) = \left\langle \Phi_{\mathcal{Z}}\left(\mathbf{z}\right), \Phi_{\mathcal{Z}}\left(\mathbf{z}'\right)\right\rangle_{\mathcal{H}}.$$

La marginalisation donne donc:

$$\begin{aligned} K_{\mathcal{X}}\left(\mathbf{x}, \mathbf{x}'\right) &= E_{P_{\mathbf{x}}(d\mathbf{y})\times P_{\mathbf{x}'}(d\mathbf{y}')} K_{\mathcal{Z}}\left(\mathbf{z}, \mathbf{z}'\right) \\ &= E_{P_{\mathbf{x}}(d\mathbf{y})\times P_{\mathbf{x}'}(d\mathbf{y}')} \left\langle \Phi_{\mathcal{Z}}\left(\mathbf{z}\right), \Phi_{\mathcal{Z}}\left(\mathbf{z}'\right)\right\rangle_{\mathcal{H}} \\ &= \left\langle E_{P_{\mathbf{x}}(d\mathbf{y})} \Phi_{\mathcal{Z}}\left(\mathbf{z}\right), E_{P_{\mathbf{x}}(d\mathbf{y}')} \Phi_{\mathcal{Z}}\left(\mathbf{z}'\right)\right\rangle_{\mathcal{H}} \end{aligned}$$

donc $K_{\mathcal{X}}$ est un n.d.p. sur $\mathcal{X}$. $\square$

Noyaux pour séquences par chaînes de Markov cachées

Motivations

- Les chaînes de Markov cachées (*HMM*: hidden Markov models) sont *extrememement utilisées* pour modéliser des séquences biologiques
- Les états cachés (voir les slides suivants) ont en général un *sens biologique*
- La marginalisation va permettre de déduire un n.d.p. basé sur cette information biologique codée dans le modele probabiliste
- Cette applications n'est pas restreinte aux séquences biologiques: les HMM sont tres utilisées dans de nombreux autres domaines (reconnaissance vocale, traitements du signal,...).

Chaîne de Markov

Soit $\mathcal{S}$ un ensemble fini, comprenant un élément S (Start) et un élément E (End). Une chaíne de Markov est une distribution de probabilité P sur $\mathcal{Y} = \mathcal{S}^*$ définie par:

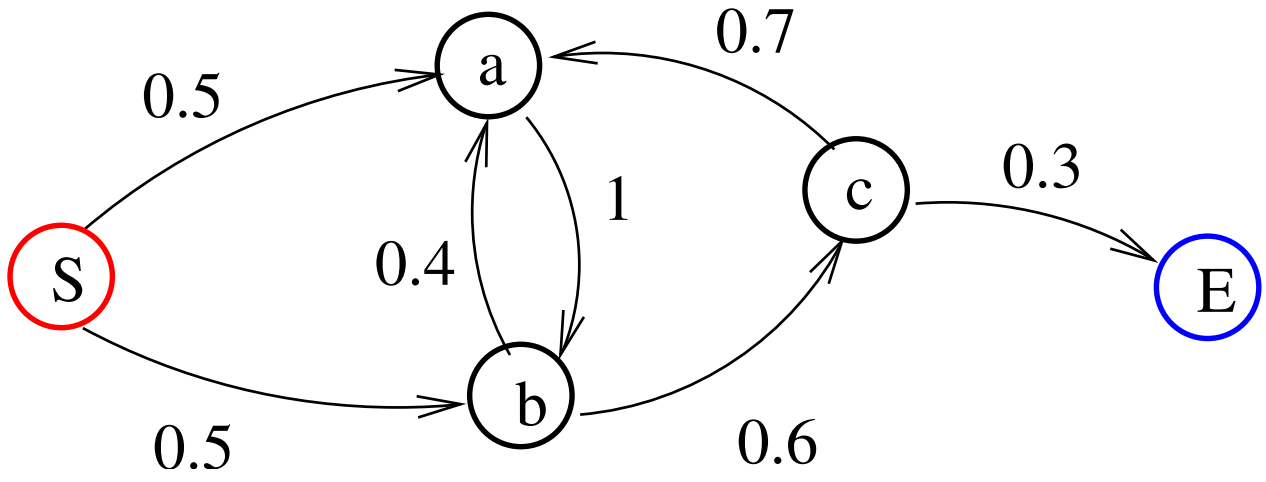
$$P\left(y_0 y_1 \dots y_n\right) =$$

$$\begin{cases} 0 & \text{si } y_0 \neq S \text{ ou } y_n \neq E \text{ ou } y_i = E \text{ pour } i < n, \\ \prod_{i=1}^{n} p\left(y_{i+1} | y_i\right) & \text{sinon,} \end{cases}$$

avec pour tout $a, b \in \mathcal{Y}$

$$\begin{cases} p(a|b) \geq 0 \\ \sum_{c \in \mathcal{S}} p(c|a) = 1. \end{cases}$$

Exemple



Calcul de la probabilité d'une séquence sous ce modele:

$$P\left(SababcE\right) = 0.5 \times 1 \times 0.4 \times 1 \times 0.6 \times 0.3 = 0.0036$$

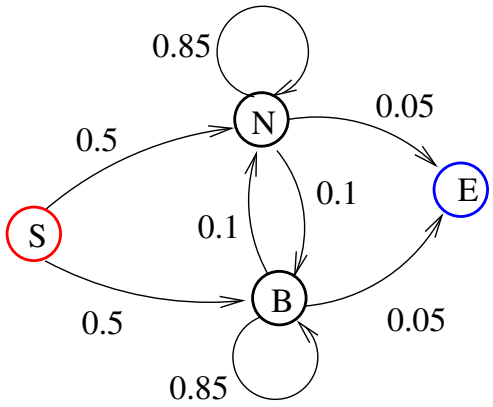
Chaîne de Markov cachée (HMM)

On n'observe pas la suite des états, qui forment un chaîne de Markov. Par contre, on suppose que chaque état $y \in \mathcal{S}$ (sauf S et E) émet, indépendamment des autres, une variable $x \in \mathcal{A}$ (fini) que l'on observe.
On obtient donc une probabilité sur les paires $(\mathbf{x}, \mathbf{y}) \in \mathcal{A}^* \times \mathcal{S}^*$ définie par:

$$P\left(x_1 \dots x_n, y_0 \dots y_{n+1}\right) = p(y_0 \dots y_{n+1}) \times \prod_{i=1}^{n} \pi(x_i|y_i),$$

ou π est la probabilité conditionnelle d'émission ($\sum_{x \in \mathcal{A}} \pi(x|y) = 1$ pour $y \in \mathcal{S}$).

Exemple: pile ou face biaisé



- Une piece normale N, une biaisée B (non observé)

- On observe pile (0) ou face (1) avec probabilités:

$$\begin{cases} \pi(0|N) = 1 - \pi(1|N) = 0.5, \\ \pi(0|B) = 1 - \pi(1|B) = 0.8. \end{cases}$$

- Exemple de réalisation:

```
NNNNNBBBBBBBBBNNNNNNNNNNNNBBBBBB
10010111011111010010111001111011
```

Noyau pour données observées

- Si on n'observe que $\mathbf{x} \in \mathcal{A}^*$, le n.d.p. le plus simple est le 1-spectral:

$$K\left(\mathbf{x}, \mathbf{x}'\right) = \sum_{a \in \mathcal{A}} n_a\left(\mathbf{x}\right) n_a\left(\mathbf{x}\right),$$

ou $n_a\left(\mathbf{x}\right)$ est le nombre d'occurrences de a dans $\mathbf{x}$.

- Exemple:

$$\begin{aligned}\mathbf{x} &= 1001011101111010010111001111011,\\ \mathbf{x}' &= 0011010110011111011010111101100101,\end{aligned}$$

$$\begin{aligned}k\left(\mathbf{x}, \mathbf{x}'\right) &= n_0\left(\mathbf{x}\right) n_0\left(\mathbf{x}'\right) + n_1\left(\mathbf{x}\right) n_1\left(\mathbf{x}'\right)\\ &= 11 \times 13 + 20 \times 21 = 563.\end{aligned}$$

Noyaux pour données totales

- Si on observait les données totales $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{A}^* \times \mathcal{S}^*$, on pourrait utilise le noyau:

$$K_{\mathcal{Z}}\left(\mathbf{z}, \mathbf{z}'\right) = \sum_{(a,s) \in \mathcal{A} \times \mathcal{S}} n_{a,s}\left(\mathbf{z}\right) n_{a,s}\left(\mathbf{z}\right),$$

ou $n_{a,s}\left(\mathbf{x}, \mathbf{y}\right)$ est le nombre d'occurrences de s dans $\mathbf{y}$ qui émettent a dans $\mathbf{x}$.

- Exemple:

$$\begin{aligned} \mathbf{z} &= 1001011101111010010111001111011, \\ \mathbf{z}' &= 0011010110011111011010111101100101, \end{aligned}$$

$$\begin{aligned} k\left(\mathbf{z}, \mathbf{z}'\right) &= n_0\left(\mathbf{z}\right) n_0\left(\mathbf{z}'\right) + n_0\left(\mathbf{z}\right) n_0\left(\mathbf{z}'\right) + n_1\left(\mathbf{z}\right) n_1\left(\mathbf{z}'\right) + n_1\left(\mathbf{z}\right) \\ &= 7 \times 15 + 9 \times 12 + 13 \times 6 + 2 \times 1 = 293. \end{aligned}$$

Noyau marginalisé

Le noyau marginalisé pour données observées est défini par:

$$K_{\mathcal{X}}\left(\mathbf{x},\mathbf{x}'\right) = \sum_{\mathbf{y},\mathbf{y}'\in\mathcal{S}^*} K_{\mathcal{Z}}\left(\left(\mathbf{x},\mathbf{y}\right),\left(\mathbf{x},\mathbf{y}\right)\right) P\left(\mathbf{y}|\mathbf{x}\right) P\left(\mathbf{y}'|\mathbf{x}'\right)$$

$$= \sum_{(a,s)\in\mathcal{A}\times\mathcal{S}} \Phi_{a,s}\left(\mathbf{x}\right)\Phi_{a,s}\left(\mathbf{x}'\right),$$

avec

$$\Phi_{a,s}\left(\mathbf{x}\right) = \sum_{\mathbf{y}\in\mathcal{S}^*} P\left(\mathbf{y}|\mathbf{x}\right) n_{a,s}\left(\mathbf{x},\mathbf{y}\right)$$

Calcul du noyau marginalisé

Nous allons calculer explicitement $\Phi_{a,s}(\mathbf{x})$, pour tout $(a,s) \in \mathcal{A} \times \mathcal{S}$. Supposons

$$\mathbf{x} = x_1 \dots x_n \in \mathcal{A}^n.$$

Alors $P(\mathbf{y}|\mathbf{x}) > 0$ implique $\mathbf{y}$ soit de la forme:

$$\mathbf{y} = S y_1 \dots y_n E \in \mathcal{S}^{n+2}.$$

On a alors, en notant δ est le symbole de Kronecker ($\delta(u,v) = 1$ si $u = v$, 0 sinon):

$$n_{a,s}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \delta(x_i, a)\, \delta(y_i, b),$$

Calcul du noyau marginalisé

On en déduit:

$$\begin{aligned}
\Phi_{a,s}(\mathbf{x}) &= \sum_{\mathbf{y}\in\mathcal{S}^*} P(\mathbf{y}|\mathbf{x})\, n_{a,s}(\mathbf{x},\mathbf{y}) \\
&= \sum_{\mathbf{y}\in\mathcal{S}^*} P(\mathbf{y}|\mathbf{x}) \left\{ \sum_{i=1}^{n} \delta(x_i, a)\, \delta(y_i, s) \right\} \\
&= \sum_{i=1}^{n} \delta(x_i, a) \left\{ \sum_{\mathbf{y}\in\mathcal{S}^*} P(\mathbf{y}|\mathbf{x})\, \delta(y_i, s) \right\} \\
&= \sum_{i=1}^{n} \delta(x_i, a)\, P(y_i = s|\mathbf{x})\,.
\end{aligned}$$

Calcul de $P\left(y_i = s|\mathbf{x}\right)$

Il existe une méthode classique pour calculer $P\left(y_i = s|\mathbf{x}\right)$ de maniere efficace: l'algorithme *Forward-Backward*. Soient:

$$\forall\,(s,i) \in \mathcal{S} \times [1,n], \quad \begin{cases} f_s(i) := P\left(x_1 \dots x_i, y_i = s\right), \\ b_s(i) := P\left(x_{i+1} \dots x_n | y_i = s\right) \end{cases}$$

Le calcul de $P\left(y_i = s|\mathbf{x}\right)$ se ramene a celui de $f_s(i)$ et $b_s(i)$, car:

$$\begin{aligned} P\left(y_i = s|\mathbf{x}\right) &= \frac{P\left(x_1 \dots x_n, y_i = s\right)}{P\left(\mathbf{x}\right)} \\ &= \frac{f_s(i) b_s(i)}{\sum_{s' \in \mathcal{S}} f_{s'}(i) b_{s'}(i)} \end{aligned}$$

Calcul de $f_s(i)$

$f_s(i)$ (forward) se calcule *récursivement* pour par:

$$\forall t \in \mathcal{S}, \quad f_t(1) = P(x_1, y_1 = t) = p(t|S)\pi(x_1|t),$$

et pour $t \in \mathcal{S}$ et $j = 1, \ldots, i$:

$$\begin{aligned} f_t(j) &= P(x_1 \ldots x_j, y_j = t) \\ &= \sum_{u \in \mathcal{S}} P(x_1 \ldots x_j, y_{j-1} = u, y_j = t) \\ &= \sum_{u \in \mathcal{S}} P(x_1 \ldots x_{j-1}, y_{j-1} = u) P(x_j, y_j = t | y_{j-1} = u) \\ &= \sum_{u \in \mathcal{S}} f_u(j-1) \times p(t|u)\pi(x_j|u). \end{aligned}$$

Calcul de $b_s(i)$

De même $b_s(i)$ (backward) se calcule par:

$$\forall t \in \mathcal{S}, \quad b_t(n) = 1,$$

et pour $t \in \mathcal{S}$ et $j = n, n-1, \ldots, i$:

$$\begin{aligned} b_t(j) &= P\left(x_{j+1} \ldots x_n | y_j = t\right) \\ &= \sum_{u \in \mathcal{S}} P\left(x_{j+1} \ldots x_n, y_{j+1} = u | y_j = t\right) \\ &= \sum_{u \in \mathcal{S}} P\left(x_{j+1}, y_{j+1} = u | y_j = t\right) P\left(x_{j+2} \ldots x_n | y_{j+1} = u\right) \\ &= \sum_{u \in \mathcal{S}} b_u(j+1) \times p(u|t) \pi(x_{j+1}|u). \end{aligned}$$

Remarque

Les algorithmes forward et backward sont les algorithmes de bases des HMM. On peut en particulier calculer la probabilité d'une séquence $\mathbf{x} = x_1 \dots x_n$ par:

$$P(\mathbf{x}) = \sum_{s \in \mathcal{S}} P(x_1 \dots x_n, y_n = s) = \sum_{s \in \mathcal{S}} f_s(n).$$

Extension

- Plutot que le noyau 1-spectral, peut-on marginalisér des noyaux spectraux d'ordre supérieur?
- Par exemple:

$$K_{\mathcal{Z}}\left(\mathbf{z},\mathbf{z}'\right)=\sum_{(a,b,s,t)\in\mathcal{A}^2\times\mathcal{S}^2} n_{a,b,s,t}\left(\mathbf{z}\right)n_{a,b,s,t}\left(\mathbf{z}\right),$$

ou $n_{a,b,s,t}\left(\mathbf{z}\right)$ est le nombre d'occurrences du 2-mer st dans $\mathbf{y}$ qui émettent ab.

- Intérêt: encoder plus d'information dans le noyau

Extension (exercice)

Le noyau 2-spectral peut se marginalisér en:

$$K_2\left(\mathbf{x},\mathbf{x}'\right) = \sum_{(a,b,s,t)\in\mathcal{A}^2\times\mathcal{S}^2} \Phi_{a,b,s,t}\left(\mathbf{x}\right)\Phi_{a,b,s,t}\left(\mathbf{x}\right),$$

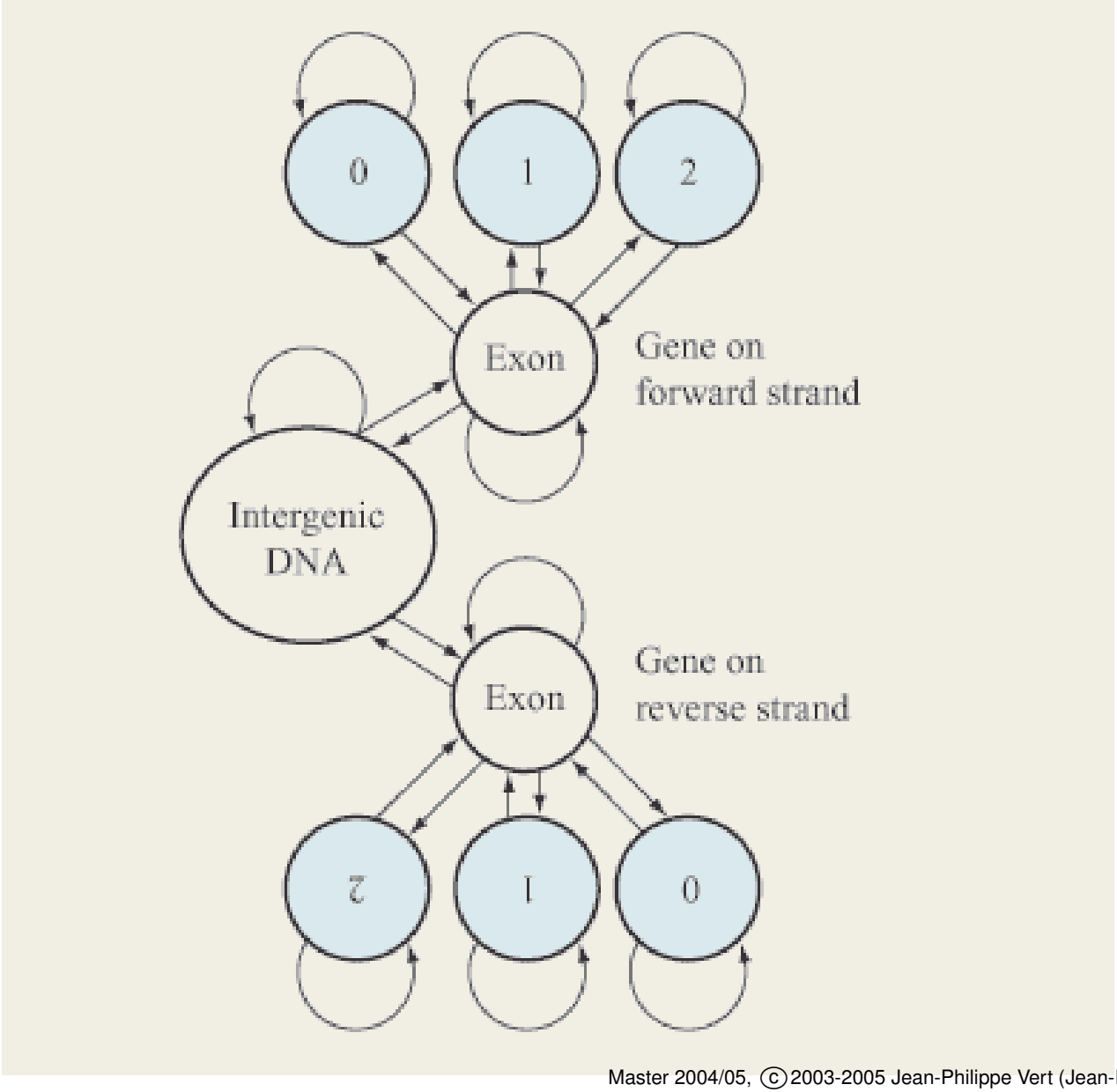
avec:

$$\Phi_{a,b,s,t}\left(\mathbf{x}\right) = \sum_{i=1}^{n-1} \delta\left(x_i,a\right)\delta\left(x_{i+1},b\right)P\left(y_i=s,y_{i+1}=t|\mathbf{x}\right),$$

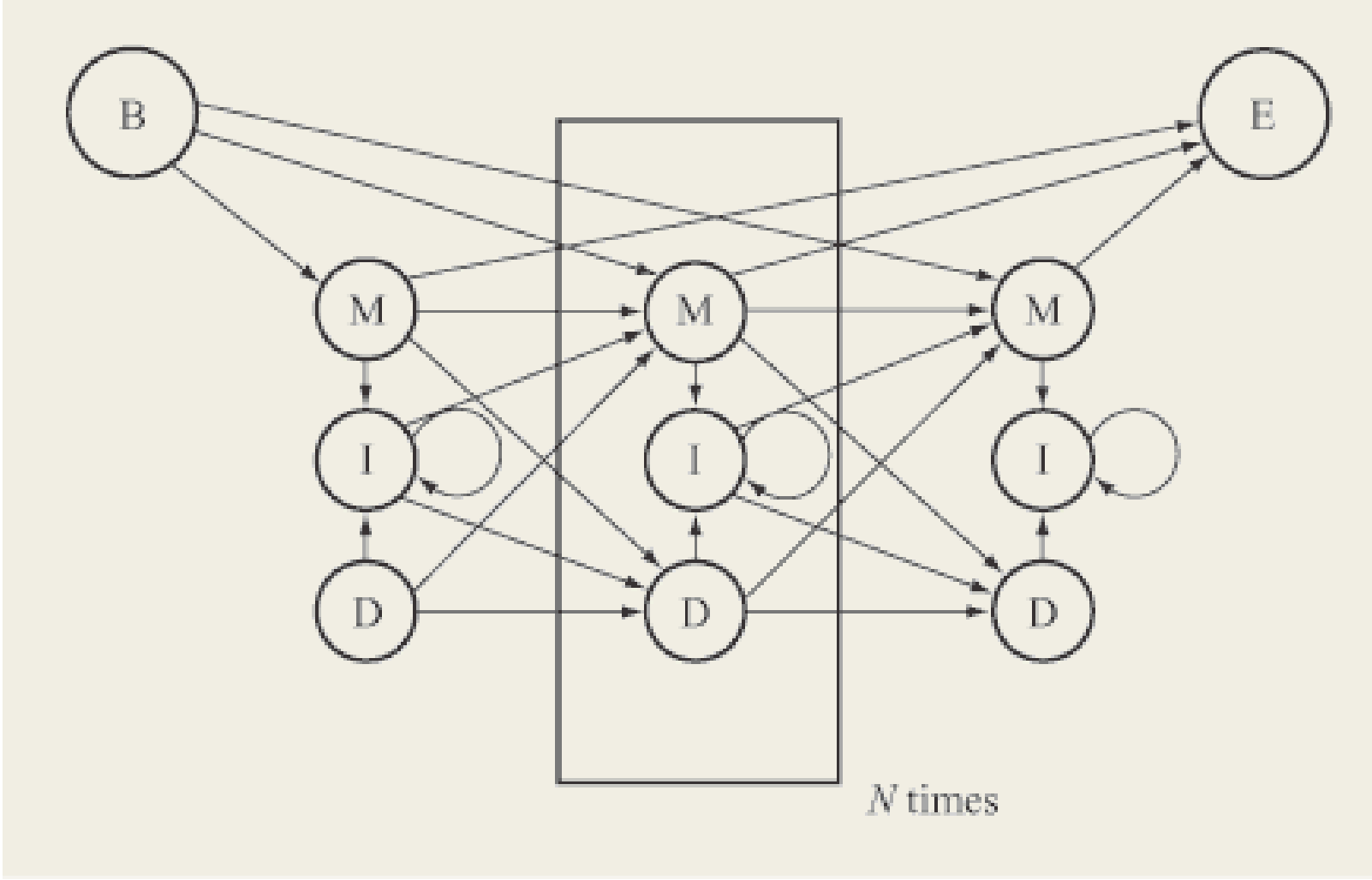
calculable par:

$$P\left(y_i=s,y_{i+1}=t|\mathbf{x}\right) = \frac{p(t|s)\pi(x_{i+1}|t)f_s(i)b_t(i+1)}{P\left(\mathbf{x}\right)}.$$

Exemple de HMM (DNA)



Exemple de HMM (protein)



Noyau marginalisé pour graphe