Cours MVA 2005 - Modèles graphiques

Devoir à la maison 1 - Correction

(*rédigée à partir de la solution de Laurent Jacob et Roland Donat*)

1 Élimination pour modèles dynamiques

1.1 Premier graphe (modèle de Markov caché)

Le graphe de la figure 1 est un arbre orienté donc la largeur arborescente est égale à 1. La version moralisée est présentée sur la figure 2 (la moralisation n'ajoute pas d'arc pour un arbre orienté). La figure 3 propose un ordre d'élimination. Comme pour tous les arbres, le graphe reconstitué est identique au graphe moralisé (figure 2).

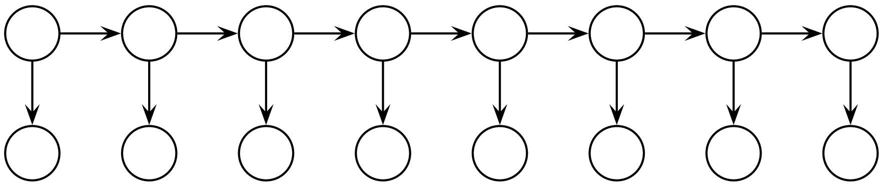

Fig. 1 – Modèle de Markov caché

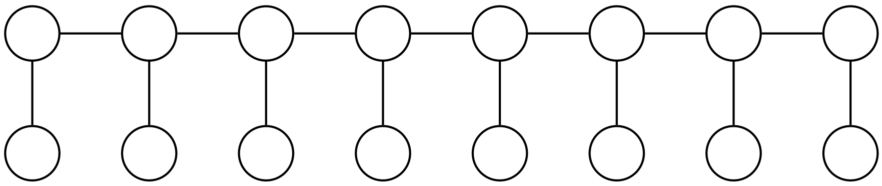

Fig. 2 – Modèle de Markov caché moralisé

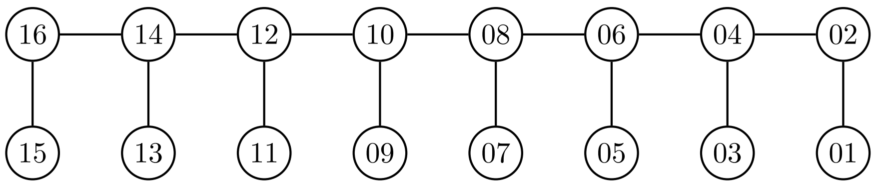

Fig. 3 – Ordre d'élimination pour le modèle de Markov caché

1.2 Deuxième graphe

Le graphe de la figure 4 comporte des structures en V. La version moralisée est présentée sur la figure 5. la largeur arborescente est égale à 2 et la figure 6 propose un ordre d'élimination. Le graphe reconstitué est identique au graphe moralisé en figure 5.

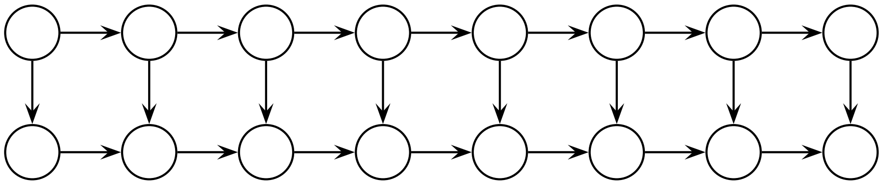

Fig. 4 – Deuxième graphe

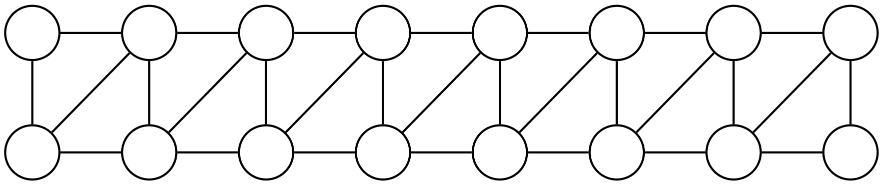

Fig. 5 – Deuxième graphe moralisé

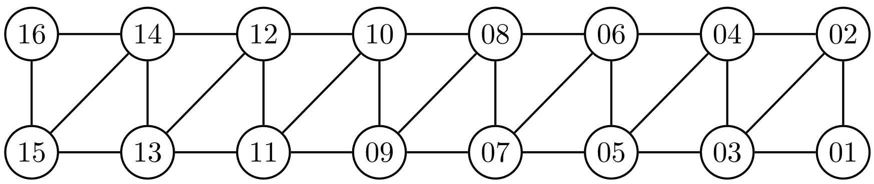

Fig. 6 – Ordre d'élimination pour le deuxième graphe

1.3 Troisième graphe

Le graphe de la figure 7 comporte des structures en V. La version moralisée est présentée sur la figure 8. la largeur arborescente est égale à 2 et la figure 9 propose un ordre d'élimination. Le graphe reconstitué est identique au graphe moralisé en figure 8.

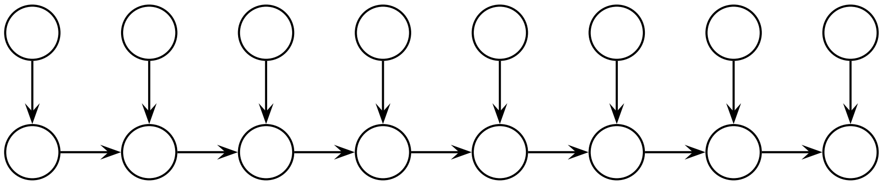

FIG. 7 – Troisième graphe

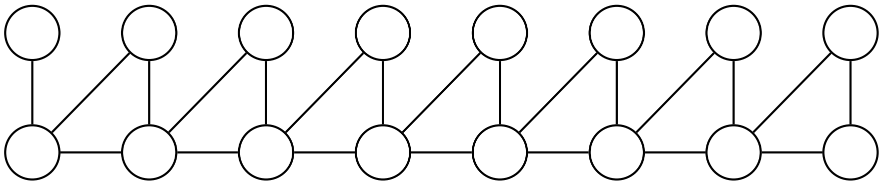

FIG. 8 – Troisième graphe moralisé

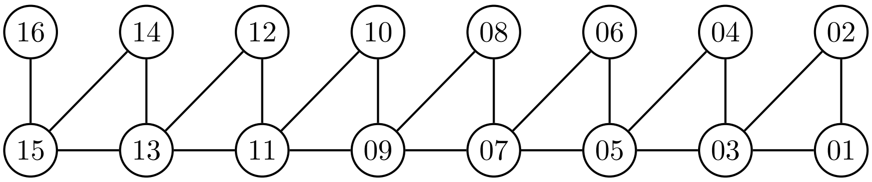

FIG. 9 – Ordre d'élimination pour le troisième graphe

1.4 Quatrième graphe (modèle de Markov caché *factoriel*)

Le graphe de la figure 10 comporte des structures en V. La version moralisée est présentée sur la figure 11. la largeur arborescente est égale à 3 et la figure 12 propose un ordre d'élimination. Le graphe reconstitué est présenté en figure 13.

Pour ces modèles factoriels il est important d'éliminer les sommets par colonnes afin d'éviter des cliques trop larges.

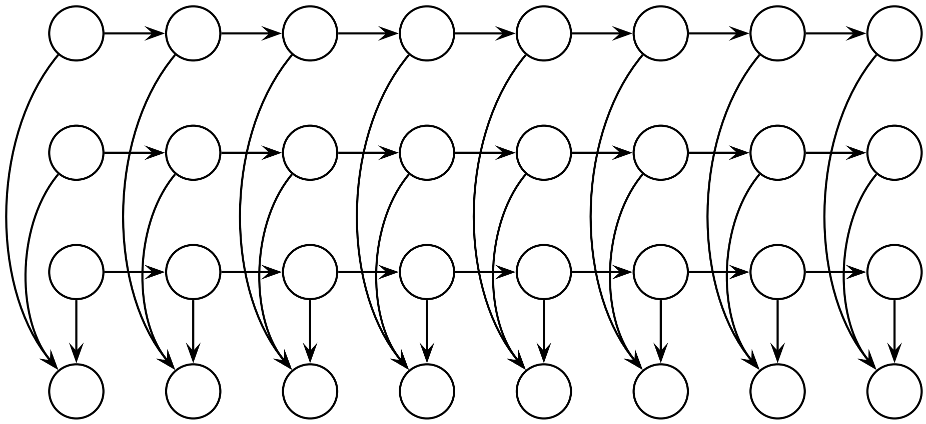

Fig. 10 – Modèle de Markov caché factoriel

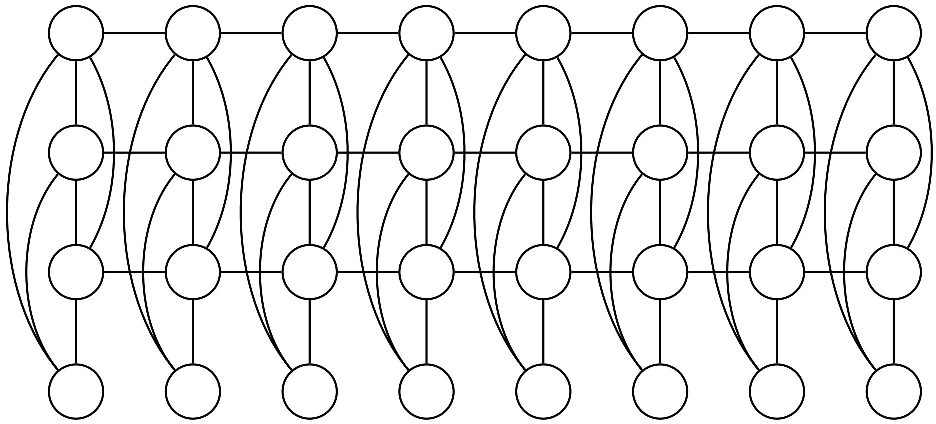

Fig. 11 – Modèle de Markov caché factoriel moralisé.

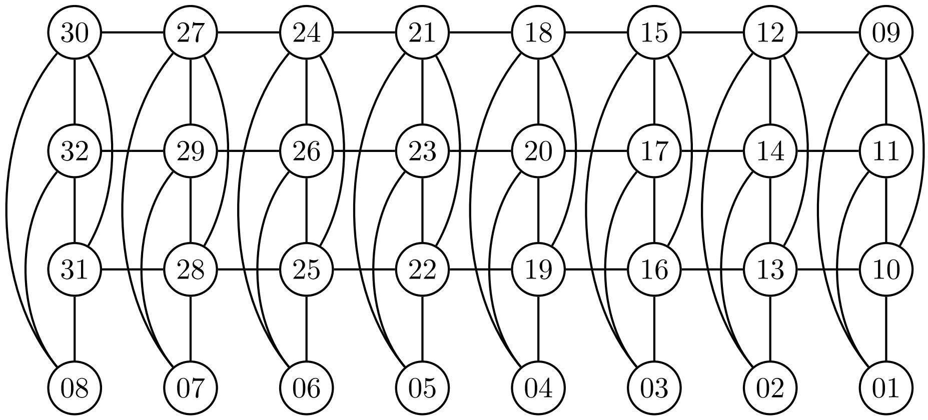

FIG. 12 – Ordre d'élimination pour le modèle de Markov caché factoriel

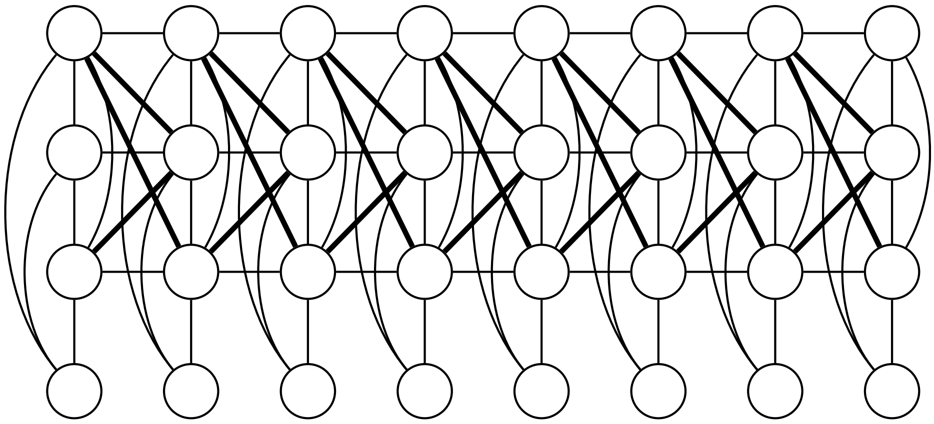

FIG. 13 – Modèle de Markov caché factoriel : graphe reconstitué correspondant à l'ordre d'élimination de la figure 12

2 Estimation de paramètres – Lois Normales

2.1 MLE d'une loi normale multivariée

On dispose de N points i.i.d. $x_i \in \mathbb{R}^p$, $i = 1, \ldots N$, issus d'une loi normale $\mathcal{N}(\mu, \Sigma)$, avec $\mu \in \mathbb{R}^p$ et $\Sigma \in \mathbb{R}^{p \times p}$, symmétrique, strictement définie positive. On a donc

$$\begin{aligned}
p(x_1, \ldots, x_N \mid \mu, \Sigma) &= \prod_{i=1}^{N} \frac{1}{c} \exp\left(-\frac{1}{2}(x_i - \mu)^\top \Sigma^{-1}(x - \mu)\right), \text{ avec } c = (2\pi)^{p/2} (\det \Sigma)^{1/2} \\
&= \frac{1}{c^N} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^\top \Sigma^{-1} x_i - x_i^\top \Sigma^{-1} \mu - \mu^\top \Sigma^{-1} x_i + \mu^\top \Sigma^{-1} \mu\right) \\
&= \frac{1}{c^N} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^\top \Sigma^{-1} x_i - 2(\Sigma^{-1} x_i)^\top \mu + \mu^\top \Sigma^{-1} \mu\right).
\end{aligned}$$

En passant au logarithme, on obtient

$$\ln p(x_1, \dots, x_N \mid \mu, \Sigma) = -N \ln c - \frac{N}{2}\mu^\top \Sigma^{-1}\mu - \frac{1}{2}\sum_{i=1}^{N} x_i^\top \Sigma^{-1} x_i - 2(\Sigma^{-1}x_i)^\top \mu.$$

Pour μ, l'estimateur de maximum de vraisemblance, noté $\hat{\mu}_{MV}$, doit vérifier

$$\begin{aligned} \tfrac{\partial}{\partial \mu} \ln p(x_1, \dots, x_N \mid \mu, \Sigma) = 0 \quad & \Longleftrightarrow \quad -N\Sigma^{-1}\mu + \Sigma^{-1}\textstyle\sum_{i=1}^N x_i = 0 \\ & \Longleftrightarrow \quad \mu = \tfrac{1}{N}\textstyle\sum_{i=1}^N x_i. \end{aligned}$$

Finalement, on a $\hat{\mu}_{MV} = \frac{1}{N}\sum_{i=1}^N x_i$.

Pour ce qui est de Σ, on rappelle que l'on a

$$p(x_1, \dots, x_N \mid \mu, \Sigma) = c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^\top \Sigma^{-1}(x-\mu)\right), \text{ avec } c = (2\pi)^{-Np/2}.$$

Il s'agit à présent de trouver les valeurs de la matrice Σ qui minimisent la vraisemblance $p(x_1, \dots, x_N \mid \mu, \Sigma)$. Remarquons tout d'abord que $(x_i - \mu)^\top \Sigma^{-1}(x-\mu) \in \mathbb{R}$, ainsi il est possible d'écrire

$$\operatorname{tr}\Big((x_i - \mu)^\top \Sigma^{-1}(x-\mu)\Big) = (x_i - \mu)^\top \Sigma^{-1}(x-\mu),$$

où $\operatorname{tr}(A)$ désigne la trace de la matrice A.

En utilisant ce qui précède, on obtient

$$\begin{aligned} p(x_1, \dots, x_N \mid \mu, \Sigma) &= c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^\top \Sigma^{-1}(x-\mu)\right) \\ &= c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\Big((x_i - \mu)^\top \Sigma^{-1}(x-\mu)\Big)\right) \\ & \quad \text{on utilise l'identité matricielle } \operatorname{tr}(AB) = \operatorname{tr}(BA) \\ &= c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\Big((x_i - \mu)(x-\mu)^\top \Sigma^{-1}\Big)\right) \\ &= c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\sum_{i=1}^{N}(x_i - \mu)(x-\mu)^\top \Sigma^{-1}\right)\right) \\ &= c\,(\det \Sigma)^{-N/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(S\Sigma^{-1}\right)\right), \end{aligned}$$

avec $S = \sum_{i=1}^N (x_i - \mu)(x-\mu)^\top \in \mathbb{R}^{p \times p}$, une matrice symétrique définie positive.

La log vraisemblance est égale à :

$$\log c - \frac{N}{2}\log\det\Sigma - \frac{1}{2}\operatorname{tr}(S\Sigma^{-1}) = \log c + \frac{N}{2}\log\det\Lambda - \frac{1}{2}\operatorname{tr}(S\Lambda)$$

où $\Lambda = \Sigma^{-1}$. la log-vraisemblance est une fonction concave en Λ et sa dérivée est égale à (voir chapitre 13 du polycopié) :

$$\frac{N}{2}\Lambda^{-1} - \frac{1}{2}S$$

Le maximum est donc atteint pour $\Lambda = NS^{-1}$.

Finalement, on en déduit l'estimateur du maximum de vraisemblance $\hat{\Sigma}_{MV}$ défini par

$$\hat{\Sigma}_{MV} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu}_{MV})(x_i - \hat{\mu}_{MV})^\top.$$

2.2 Densité a posteriori - loi normale

On dispose de N points i.i.d. $x_i \in \mathbb{R}$, $i = 1, \ldots N$ avec

$$x \mid \mu \sim \mathcal{N}(\mu, \sigma^2),\ \mu \sim \mathcal{N}(\mu_0, \tau^2),\quad \sigma^2, \tau^2 \in \mathbb{R},\ \mu_0 \in \mathbb{R}.$$

La densité *a posteriori* du paramètre μ a pour définition

$$\begin{aligned} p(\mu \mid (x_i)_{i\in\{1,\ldots,N\}}) &= \frac{p((x_i)_{i\in\{1,\ldots,N\}} \mid \mu)p(\mu)}{\int_\mu p((x_i)_{i\in\{1,\ldots,N\}} \mid \mu)p(\mu)d\mu} \\ &= C_1 \prod_{i=1}^N p(x_i \mid \mu)p(\mu),\ C_1 = \frac{1}{\int_\mu p((x_i)_{i\in\{1,\ldots,N\}} \mid \mu)p(\mu)d\mu}. \end{aligned}$$

Ainsi on obtient

$$\begin{aligned} p(\mu \mid (x_i)_{i\in\{1,\ldots,N\}}) &= C_1 \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi}\tau} \exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_0}{\tau}\right)^2\right] \\ &= C_2 \exp\left[-\frac{1}{2}\left(\sum_{i=1}^N \left(\frac{x_i-\mu}{\sigma}\right)^2 + \left(\frac{\mu-\mu_0}{\tau}\right)^2\right)\right],\ C_2 = \frac{1}{\sqrt{2\pi}^{N+1}\sigma^N\tau} C_1 \\ &= C_3 \exp\left[-\frac{1}{2}\left(\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2}\sum_{i=1}^N x_i + \frac{\mu_0}{\tau^2}\right)\mu\right)\right], \\ & \quad C_3 = \exp\left[-\frac{1}{2}\left(\frac{\sum_{i=1}^N x_i^2}{\sigma^2} + \frac{\mu_0^2}{\tau^2}\right)\right] C_2. \end{aligned}$$

En remarquant que $p(\mu \mid (x_i)_{i\in\{1,\ldots,N\}})$ est la composée d'une fonction exponentielle et d'une fonction quadratique en μ, on en déduit qu'il s'agit d'une densité gaussienne. Ainsi, en posant que $p(\mu \mid (x_i)_{i\in\{1,\ldots,N\}}) \sim \mathcal{N}(\mu_N, \sigma_N^2)$, on a :

$$p(\mu \mid (x_i)_{i\in\{1,\ldots,N\}}) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_N}{\sigma_N}\right)^2\right].$$

Par identification, il vient

$$\frac{1}{\sigma_N^2} = \frac{N}{\sigma^2} + \frac{1}{\tau^2} \iff \sigma_N^2 = \frac{\sigma^2\tau^2}{N\tau^2+\sigma^2} \geqslant 0,$$

et

$$\frac{\mu_N}{\sigma_N^2} = \frac{1}{\sigma^2}\sum_{i=1}^N x_i + \frac{\mu_0}{\tau^2} \iff \mu_N = \left(\frac{N\tau^2}{N\tau^2+\sigma^2}\right)\mu_{ML} + \left(\frac{\sigma^2}{N\tau^2+\sigma^2}\right)\mu_0$$

NB : pas besoin de calculer le terme constant ; savoir que $\log p(\mu)$ est une fonction quadratique concave de μ suffit à montrer que la loi posteriori est normale.

2.3 Densité a posteriori - loi normale multivariée

On dispose de N points i.i.d. $x_i \in \mathbb{R}^p$, $i = 1, \dots N$ avec

$$x \mid \mu \sim \mathcal{N}(\mu, \Sigma),\ \mu \sim \mathcal{N}(\mu_0, T),\quad \Sigma, T \in \mathbb{R}^{p\times p},\ \mu_0 \in \mathbb{R}^p.$$

La densité *a posteriori* du paramètre μ est donnée par

$$\begin{aligned} p(\mu \mid (x_i)_{i\in\{1,\dots,N\}}) &= C_1 \prod_{i=1}^{N} p(x_i \mid \mu) p(\mu) \\ &= C_2 \exp\left[\frac{1}{2}\left(\mu^\top (N\Sigma^{-1} + T^{-1})\mu - 2\mu^\top \left(\Sigma^{-1}\sum_{i=1}^{N} x_i + T^{-1}\mu_0\right)\right)\right]. \end{aligned}$$

Pour des raisons analogues au cas précédemment, on a $p(\mu \mid (x_i)_{i\in\{1,\dots,N\}}) \sim \mathcal{N}(\mu_N, \Sigma_N)$. Ainsi, par définition la densité est proportionnelle à

$$C_3 \exp\left[\frac{1}{2}(\mu - \mu_N)^\top \Sigma_N^{-1} (\mu - \mu_N)\right],$$

ce qui permet d'en déduire, en utilisant l'identité matricielle $(A^{-1}+B^{-1})^{-1} = B(A+B)^{-1}A$ et par identification :

$$\Sigma_N^{-1} = N\Sigma^{-1} + T^{-1} \iff \Sigma_N = \frac{1}{N} T \left(T + \frac{1}{N}\Sigma\right)^{-1} \Sigma,$$

et

$$\Sigma_N^{-1}\mu_N = N\Sigma^{-1}\sum_{i=1}^{N} x_i + T^{-1}\mu_0 \iff \mu_N = T\left(T + \frac{1}{N}\Sigma\right)^{-1} \mu_{ML} + \frac{1}{N}\Sigma\left(T + \frac{1}{N}\Sigma\right)^{-1} \mu_0.$$

NB 1 : la moyenne a posteriori est une combinaison convexe entre la moyenne a priori et la moyenne empirique.

NB 2 : lorsque $N \to \infty$, $\mu_N \to \mu_{ML}$.

NB 3 : les expressions sont cohérentes avec les résultats précédents.

3 Probabilités marginales – Algorithme Somme Produit et paires

On cherche à démontrer par récurrence sur n que l'hypothèse H_n : soit $T = (V, E)$ un arbre à n sommets et $n-1$ arêtes. Une fois les $2(n-1)$ messages de l'algorithme SOMME-PRODUIT passés en série, les marginales sont égales à

$$\forall i \in V,\ a(x_i) = \psi(x_i) \prod_{k\in\mathcal{N}(i)} m_{k,i}(x_i), \tag{1}$$

$$\forall (i,j) \in E,\ a(x_i, x_j) = \psi(x_i)\psi(x_j)\psi(x_i, x_j) \prod_{k\in\mathcal{N}(i)\setminus\{j\}} m_{k,i}(x_i) \prod_{k\in\mathcal{N}(j)\setminus\{i\}} m_{k,j}(x_j), \tag{2}$$

est vrai $\forall n \in \mathbb{N} \setminus \{0, 1\}$.

Supposons que $n = 2$.

T a deux sommets x_1 et x_2. $u(x)$ a donc pour expression

$$u(x_1, x_2) = \psi(x_1)\psi(x_2)\psi(x_1, x_2).$$

D'autre part, on a par définition

$$\begin{aligned} a(x_1) &= \sum_{x_2} \psi(x_1)\psi(x_2)\psi(x_1, x_2) \\ &= \psi(x_1) \sum_{x_2} \psi(x_2)\psi(x_1, x_2) \\ a(x_1) &= \psi(x_1) m_{2,1}(x_1), \end{aligned}$$

et de façon analogue, $a(x_2) = \psi(x_2) m_{1,2}(x_2)$. Ainsi, (1) est vérifiée.

Enfin, on a directement par définition $a(x_1, x_2) = u(x_1, x_2) = \psi(x_1)\psi(x_2)\psi(x_1, x_2)$. D'où, (2) est également vérifiée.

On a donc montré que H_2 est vrai.

Supposons maintenant que H_k est vrai $\forall k \in \{2, \ldots, n-1\}$.

Soient $T = (V, E)$ un arbre de taille n, et $I = \{1, \ldots, n\}$, un ordre respectant la structure de T. On définit $\tilde{T} = (\tilde{V}, \tilde{E})$ un arbre de taille $n - 1$ avec $\tilde{V} = V \setminus \{1\}$ et $\tilde{E} = E \setminus \{1, \pi_1\}$. Autrement dit, $\tilde{T}$ est identique a T excepté que l'on a enlevé une de ses feuilles (et comme T est un arbre, il a toujours au moins une feuille). On définit également $\tilde{I} = I \setminus \{1\}$, $\tilde{u}$, $\tilde{\psi}$, $\tilde{m}$, et $\tilde{\mathcal{N}}$, qui représentent respectivement dans $\tilde{T}$, l'ordre, la loi jointe, les potentiels, les messages et le voisinage dans $\tilde{T}$.

Par définition, on a

$$\begin{aligned} \tilde{u}(x) &= \sum_{x_1} u(x) \\ &= \sum_{x_1} \prod_{i \in V} \psi(x_i) \prod_{(i,j) \in E} \psi(x_i, x_j) \\ \tilde{u}(x) &= \prod_{i \in \tilde{V}} \psi(x_i) \prod_{(i,j) \in \tilde{E}} \psi(x_i, x_j) \underbrace{\sum_{x_1} \psi(x_1)\psi(x_1, x_{\pi_1})}_{\overset{déf}{=} m_{1,\pi_1}(x_{\pi_1})} \end{aligned}$$

Ainsi $\tilde{u}(x)$ se factorisera dans $\tilde{T}$ avec les potentiels $\tilde{\psi}$ définis par

$$\begin{array}{ll} \tilde{\psi}(x_i) = \psi(x_i), & i \in \tilde{V} \\ \tilde{\psi}(x_{\pi_1}) = \psi(x_{\pi_i}) m_{1,\pi_1}(x_{\pi_1}), & \\ & \\ \tilde{\psi}(x_i, x_j) = \psi(x_i, x_j), & (i,j) \in \tilde{E} \end{array} .$$

Par construction de $\tilde{T}$, tous les messages passés sont les mêmes dans T et $\tilde{T}$ si l'on peut montrer que les messages émis de π_1 sont les mêmes ; on a :

$$\begin{aligned}
\forall i \in \mathcal{N}(\pi_1)\backslash\{1\},\ m_{\pi_1,i}(x_i) &= \sum_{x_{\pi_1}} \psi(x_{\pi_1})\psi(x_i, x_{\pi_1}) \prod_{k\in\mathcal{N}(\pi_1)\backslash\{1\}} m_{k,\pi_1}(x_i) \\
&= \sum_{x_{\pi_1}} \psi(x_{\pi_1}) m_{1,\pi_1}(x_{\pi_1})\psi(x_i, x_{\pi_1}) \prod_{k\in\mathcal{N}(\pi_1)\backslash\{i,1\}} m_{k,\pi_1}(x_{\pi_1}) \\
&= \sum_{x_{\pi_1}} \tilde{\psi}(x_{\pi_1})\tilde{\psi}(x_i, x_{\pi_1}) \prod_{k\in\tilde{\mathcal{N}}(\pi_1)\backslash\{i\}} \tilde{m}_{k,\pi_1}(x_{\pi_1}) \\
&= \tilde{m}_{\pi_1,i}(x_i).
\end{aligned}$$

On vient donc de montrer que les messages passés dans $\tilde{T}$ sont identiques à ceux passés dans T privé du nœud 1.

Passons maintenant au calcul des marginales dans T.

Cas 1 : $i > 1$

On a

$$\begin{aligned}
\forall i \in \tilde{V},\quad a(x_i) &= \sum_{\substack{x_j \\ j\neq i}} u(x) \\
&= \sum_{\substack{x_j \\ j\neq i}} \prod_{p\in V} \psi(x_p) \prod_{(p,q)\in E} \psi(x_p, x_q) \\
&= \sum_{\substack{x_j \\ j\neq i, j\neq 1}} \prod_{p\in\tilde{V}} \psi(x_p) \prod_{(p,q)\in\tilde{E}} \psi(x_p, x_q) \sum_{x_1} \psi(x_1)\psi(x_{\pi_1}, x_1) \\
&= \sum_{\substack{x_j \\ j\neq i, j\neq 1}} \prod_{p\in\tilde{V}} \psi(x_p) \prod_{(p,q)\in\tilde{E}} \psi(x_p, x_q) m_{1,\pi_1}(x_{\pi_1}) \\
&= \sum_{\substack{x_j \\ j\neq i, j\neq 1}} \tilde{u}(x) \\
\forall i \in \tilde{V},\quad a(x_i) &= \tilde{a}(x_i).
\end{aligned}$$

Or, par hypothèse de récurrence, on a

$$\forall i \in \tilde{V},\ \tilde{a}(x_i) = \tilde{\psi}(x_i) \prod_{k\in\tilde{N}(i)} \tilde{m}_{k,i}(x_1),$$

et on a vu précédemment que l'équivalence des messages dans $\tilde{T}$ et T lorsque $i \neq 1$. Aussi, on obtient

$$\forall i \in \tilde{V},\ a(x_i) = \prod_{k\in N(i)} m_{k,i}(x_1).$$

De façon analogue, on a également

$$\begin{aligned}
\forall (i,j) \in \tilde{E},\ a(x_i, x_j) &= \sum_{\substack{x_k \\ k\neq i, k\neq j}} u(x) = \sum_{\substack{x_k \\ k\neq i, k\neq j, k\neq 1}} \tilde{u}(x) \\
\forall (i,j) \in \tilde{E},\quad a(x_i, x_j) &= \tilde{a}(x_i, x_j),
\end{aligned}$$

et par hypothèse de récurrence,

$$\forall (i,j) \in \tilde{R},\quad \tilde{a}(x_i, x_j) = \tilde{\psi(x_i)}\tilde{\psi(x_j)}\tilde{\psi(x_i, x_j)} \prod_{k\in\tilde{\mathcal{N}}(i)\backslash\{j\}} \tilde{m}_{k,i}(x_i) \prod_{k\in\tilde{\mathcal{N}}(j)\backslash\{i\}} \tilde{m}_{k,j}(x_j).$$

D'où,

$$\forall (i,j) \in \tilde{E},\ a(x_i, x_j) = \psi(x_i)\psi(x_j)\psi(x_i, x_j) \prod_{k\in\mathcal{N}(i)\backslash\{j\}} m_{k,i}(x_i) \prod_{k\in\mathcal{N}(j)\backslash\{i\}} m_{k,j}(x_j).$$

On vient de montrer que H_n est vrai pour tous les nœud de l'arbre sauf le nœud 1.

Cas 2 : $i = 1$

On a

$$\begin{aligned}
a(x_1, x_{\pi_1}) &= \sum_{\substack{x_i \\ i\neq 1, i\neq \pi_1}} u(x) \\
a(x_1, x_{\pi_1}) &= \psi(x_1)\psi(x_{\pi_1})\psi(x_1, x_{\pi_1}) \underbrace{\sum_{\substack{x_i \\ i\neq 1, i\neq \pi_1}} \frac{u(x)}{\psi(x_1)\psi(x_{\pi_1})\psi(x_1, x_{\pi_1})}}_{\alpha(x_{\pi_1})},
\end{aligned}$$

et par conséquent

$$\begin{aligned}
a(\pi_1) &= \sum_{x_1} a(x_1, x_{\pi_1}) \\
&= \left(\sum_{x_1} \psi(x_1)\psi(x_1, x_{\pi_1})\right) \psi(x_{\pi_1})\alpha(x_{\pi_1}) \\
a(\pi_1) &= m_{1,\pi_1}(x_{\pi_1})\psi(x_{\pi_1})\alpha(x_{\pi_1}).
\end{aligned}$$

On en déduit alors

$$\alpha(x_{\pi_1}) = \frac{1}{\psi(x_{\pi_1})} \frac{a(x_{\pi_1})}{m_{1,\pi_1}(x_{\pi_1})}.$$

et

$$a(x_1, x_{\pi_1}) = \psi(x_1)\psi(x_1, x_{\pi_1}) \frac{a(x_{\pi_1})}{m_{1,\pi_1}(x_{\pi_1})}.$$

En utilisant l'hypothèse de récurrence sur $a(x_{\pi_1})$, on obtient :

$$
\begin{aligned}
a(x_1, x_{\pi_1}) &= \psi(x_1)\psi(x_1, x_{\pi_1}) \frac{\psi(x_{\pi_1}) \prod_{k \in \mathcal{N}(\pi_1)} m_{k,\pi_1}(x_{\pi_1})}{m_{1,\pi_1}(x_{\pi_1})} \\
&= \psi(x_1)\psi(x_{\pi_1})\psi(x_1, x_{\pi_1}) \prod_{k \in \mathcal{N}(\pi_1)\setminus\{1\}} m_{k,\pi_1}(x_{\pi_1})
\end{aligned}
$$

ce qui montre le résultat pour la probabilité jointe sur x_1, x_{π_1}. En sommant par rapport à x_{π_1}, on obtient immédiatement le résultat pour $a(x_1)$.