

Inférence dans les chaînes semi-markoviennes cachées M -stationnaires

Dans ce chapitre, nous étudions les méthodes d'estimation et de segmentation vues au chapitre précédent dans le cadre de deux modèles généralisant les chaînes de Markov cachées classiques. Ces deux modèles font partie de la famille générale de chaînes de Markov “triplets”. Nous commençons par rappeler la famille des modèles de Markov “couples”, qui généralisent les modèles de Markov cachés. Nous introduisons ensuite les chaînes “triplets” [95, 102] et nous rappelons un de leurs intérêts, important pour les applications pratiques et développé récemment dans la thèse de P. Lanchantin [65], qui réside dans leur aptitude à traiter des données cachées non stationnaires. Dans un second temps, nous précisons les deux modèles particuliers introduits dans cette thèse. Le premier modèle est celui des chaînes semi-markoviennes, qui peuvent être vues comme des CMT particuliers. Nous présentons des résultats théoriques montrant que la semi-markovianité classique, où le temps aléatoire de séjour dans un état est à valeurs dans \mathbb{N} , peut également se modéliser par un temps de séjour “minimal”. Nous définissons ainsi une sous-famille des modèles semi-markoviens cachés classiques et montrons leur intérêt, au niveau du temps de calcul, par rapport aux démarches fondées sur les chaînes semi-markoviennes classiques. Ensuite, nous étendons ce modèle au cas non stationnaire. L'intérêt des nouveaux modèles est validé par des expérimentations.

3.1 Chaînes de Markov couples et chaînes de Markov cachées

Soit $Z = (X_n, Y_n)_{1 \leq n \leq N}$ un processus où chaque X_n prend ses valeurs dans un ensemble fini $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ et chaque Y_n prend ses valeurs dans un \mathbb{R} -espace vectoriel de dimension finie \mathcal{Y} . Notons $p(z_{1:N})$ la densité de Z par rapport à la mesure produit $(\nu \otimes \lambda_{\mathcal{Y}})^N$, où ν est la mesure de décompte sur \mathcal{X} et $\lambda_{\mathcal{Y}}$ est la mesure de Lebesgue sur \mathcal{Y} . Supposons que Z est une chaîne de Markov couple. Nous avons :

$$p(z_{1:N}) = p(z_1) \prod_{n=1}^{N-1} p(z_{n+1}|z_n) = \frac{p(z_1, z_2)p(z_2, z_3) \dots p(z_{N-1}, z_N)}{p(z_2)p(z_3) \dots p(z_{N-1})}. \quad (3.1)$$

D'après la factorisation de la loi de Z et l'équivalence entre factorisation et markovianité vue au chapitre 2, on en déduit que $p(x_{1:N}|y_{1:N})$ et $p(y_{1:N}|x_{1:N})$ sont des distributions markoviennes. La chaîne sera dite stationnaire si les densités $p(z_n, z_{n+1})$ ne dépendent pas de n . La proposition suivante donne des conditions pour que la chaîne cachée X soit une chaîne de Markov :

Proposition 3.1.1. *Soit $Z = (X_n, Y_n)_{1 \leq n \leq N}$ une chaîne de Markov couple vérifiant :*

1. $p(z_n, z_{n+1})$ ne dépend pas de $n \in \{1, \dots, N-1\}$;
2. $p(z_n = a, z_{n+1} = b) = p(z_n = b, z_{n+1} = a)$ pour tout $n \in \{1, \dots, N-1\}$ et pour tout a et b .

Alors les trois conditions :

- X est une chaîne de Markov ;
- pour tout $n \in \{2, \dots, N\}$, $p(y_n|x_n, x_{n-1}) = p(y_n|x_n)$;
- pour tout $n \in \{1, \dots, N\}$, $p(y_n|x_{1:N}) = p(y_n|x_n)$

sont équivalentes.

Preuve. Voir [99]. □

Nous constatons ainsi que le modèle des chaînes de Markov couples est plus général que celui des chaînes de Markov cachées dans lequel le processus caché est une chaîne de Markov. Il permet ainsi de modéliser des bruits complexes et de considérer des processus cachés non markoviens.

3.2 Chaînes de Markov cachées M -stationnaires

3.2.1 Le modèle

Dans cette section, chaque X_n prend ses valeurs dans l'ensemble fini $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$, chaque U_n prend ses valeurs dans l'ensemble fini $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ et chaque Y_n dans un espace vectoriel de dimension finie \mathcal{Y} . Le processus (X, U, Y) est une chaîne M -stationnaire cachée si sa distribution est de la forme :

$$p(x_{1:N}, u_{1:N}, y_{1:N}) = p(u_1)p(x_1|u_1)p(y_1|x_1) \prod_{n=1}^{N-1} p(u_{n+1}|u_n)p(x_{n+1}|x_n, u_{n+1})p(y_{n+1}|x_{n+1}). \quad (3.2)$$

Dans ce modèle X n'est pas une chaîne de Markov et les processus U et Y sont indépendants conditionnellement à X . Lorsque $u_1 = \dots = u_N = \lambda_k$, $p(x_{1:N}|u_{1:N})$ est la distribution d'une chaîne de Markov de loi initiale $p(x_1|u_{1:N}) = p(x_1|u_1)$ et de transition $p(x_{n+1}|u_{1:N}, x_n) = p(x_{n+1}|u_{n+1}, x_n)$.

3.2.2 Inférence dans le modèle de chaînes de Markov cachée M -stationnaires

Nous détaillerons uniquement l'estimation des paramètres de la loi $p(x_{1:N}, u_{1:N})$, l'estimation des paramètres de $p(y_n|x_n)$ étant déjà étudiée au chapitre 2.

Soit $y_{1:N} = (y_1, \dots, y_N)$ la réalisation observée de Y . Les paramètres à estimer sont $p(u_{n+1}|u_n)$ et $p(x_{n+1}|x_n, u_{n+1})$. Les estimations de $p(u_{n+1}|u_n)$ et $p(x_{n+1}|x_n, u_{n+1})$ sont obtenues à partir de celle de $p(x_n, u_n, x_{n+1}, u_{n+1})$. Nous supposons que la chaîne (X, U) est stationnaire et réversible, soit $p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l) = p(x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k)$ et indépendant de n . Notons θ_q le vecteur paramètre obtenu à l'itération q de ICE. Afin d'estimer $p(x_n, u_n, x_{n+1}, u_{n+1})$, nous devons nous donner un estimateur à partir des données complètes $(x_{1:N}, u_{1:N}, y_{1:N})$. On choisit comme estimateur :

$$\hat{p}(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l)(x_{1:N}, u_{1:N}, y_{1:N}) = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} [I(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l) + I(x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k)].$$

L'espérance de cette quantité sachant $Y = y$ et sous le paramètre θ_q peut se calculer exactement et donne :

$$\begin{aligned} & p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l; \theta_{q+1}) = \\ & \frac{1}{2(N-1)} \sum_{n=1}^{N-1} [p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l | y_{1:N}; \theta_q) + \\ & p(x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k | y_{1:N}; \theta_q)], \end{aligned}$$

et on a :

$$\begin{aligned} p(u_{n+1} = \lambda_l | u_n = \lambda_k; \theta_{q+1}) & \propto p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l; \theta_{q+1}), \\ p(x_{n+1} = \omega_j | x_n = \omega_i, u_{n+1} = \lambda_l; \theta_{q+1}) & \propto p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l; \theta_{q+1}). \end{aligned}$$

3.3 Chaînes semi-markoviennes cachées

Dans cette section, nous présentons les chaînes semi-markoviennes cachées et nous proposons d'écrire celles-ci comme un cas particulier de chaîne de Markov triplet. Les chaînes semi-markoviennes généralisent les chaînes de Markov pour des temps de séjour non nécessairement distribués suivant une loi géométrique. Parmi les applications du modèle semi-markovien, on peut citer la génétique [18, 25, 33], la reconnaissance vocale [42, 73, 83, 108], la reconnaissance de caractères [91, 109, 121], la segmentation d'images [41] ou l'analyse de modèles graphiques [52].

Nous commençons par donner la définition générale d'une chaîne semi-markovienne cachée ainsi que certaines de ses propriétés. Une étude plus approfondie des propriétés des chaînes semi-markoviennes est disponible dans [8, 32, 75]. Dans un second temps, nous proposons un modèle semi-markovien particulier original. Nous verrons que ce nouveau modèle permet l'utilisation de l'algorithme de Baum-Welsh avec une complexité algorithmique réduite. Nous concluons le chapitre par des expérimentations utilisant le modèle original de chaînes semi-markoviennes cachées M -stationnaires.

3.3.1 Définition et propriétés d'une chaîne semi-markovienne

Nous choisissons de définir une chaîne semi-markovienne à valeurs dans un espace fini comme la marginale d'une chaîne de Markov. Une définition plus générale peut se trouver dans [74].

Définition 3.3.1 (Chaîne semi-markovienne). *Soit \mathcal{X} un ensemble fini. On définit les quantités suivantes :*

- la probabilité π sur \mathcal{X} ;
- la transition $q(\cdot|\cdot)$ sur \mathcal{X}^2 vérifiant $q(x|x) = 0$;
- pour tout $x \in \mathcal{X}$, les densités de probabilité $d(x, \cdot)$ sur \mathbb{N}^* .

Un processus $X = (X_n)_{n \geq 1}$ tel que chaque X_n prend ses valeurs dans \mathcal{X} est une chaîne semi-markovienne de loi initiale π , de transition q et de loi de durée d s'il existe un processus $U = (U_n)_{n \geq 1}$, où chaque U_n prend ses valeurs dans \mathbb{N}^* , tel que (X, U) soit une chaîne de Markov homogène donnée par :

- la loi initiale :

$$p(x_1, u_1) = \pi(x_1)d(x_1, u_1) , \quad (3.3)$$

- et les transitions :

$$p(x_{n+1}, u_{n+1}|x_n, u_n) = p(x_{n+1}|x_n, u_n) \times p(u_{n+1}|x_{n+1}, u_n), \quad (3.4)$$

avec :

$$p(x_{n+1}|x_n, u_n) = \begin{cases} \delta_{x_n}(x_{n+1}) & \text{si } u_n > 1, \\ q(x_{n+1}|x_n) & \text{si } u_n = 1, \end{cases} \quad (3.5)$$

et

$$p(u_{n+1}|x_{n+1}, u_n) = \begin{cases} \delta_{u_n-1}(u_{n+1}) & \text{si } u_n > 1, \\ d(x_{n+1}, u_{n+1}) & \text{si } u_n = 1, \end{cases} \quad (3.6)$$

où $\delta_x(x) = 1$ et $\delta_x(x') = 0$ pour $x' \neq x$.

Dans cette définition, on remarque que la transition q et la loi de temps de séjour d ne dépendent pas de n , la chaîne semi-markovienne est alors qualifiée d'"homogène".

Notons que u_n représente le temps de séjour restant de la chaîne dans $X_n = x_n$, la variable aléatoire U_n est appelée "temps de récurrence avant".

Considérons la définition suivante :

Définition 3.3.2. *Soit $X = (X_n)_{n \geq 1}$ un processus à valeurs dans un ensemble fini \mathcal{X} . On définit :*

- la suite des instants de saut $V = (V_n)_{n \geq 1}$ par :

$$V_1 = 1 \text{ et } V_n = \inf \{k > V_{n-1} : X_k \neq X_{V_{n-1}}\} ;$$

- la chaîne immergée $\tilde{X} = (\tilde{X}_n)_{n \geq 1}$ par :

$$\forall n \geq 1, \tilde{X}_n = X_{V_n} ;$$

– pour tout $n \geq 1$, le temps de séjour dans \tilde{X}_n par :

$$T_n = V_{n+1} - V_n.$$

La réalisation de la chaîne immergée correspond à la suite des valeurs visitées par le processus X . Lorsque X est une chaîne semi-markovienne de transition q , on montre que sa chaîne immergée est une chaîne de Markov de transition q . Plus exactement :

Proposition 3.3.1. *Soit $X = (X_n)_{n \geq 1}$ un processus à valeurs dans un ensemble fini $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ de chaîne immergée \tilde{X} , soit $V = (V_n)_{n \geq 1}$ la suite des instants de saut de X et soit $T = (T_n)_{n \geq 1}$ le processus des temps de séjour.*

X est une chaîne semi-markovienne homogène de transition q et de loi de durée d si et seulement si les deux conditions suivantes sont satisfaites :

1. *pour tout n , \tilde{X}_{n+1} est indépendante des variables aléatoires $\tilde{X}_1, \dots, \tilde{X}_n, T_1, \dots, T_n$ conditionnellement à \tilde{X}_n et la loi de \tilde{X}_{n+1} sachant \tilde{X}_n ne dépend pas de n ;*
2. *pour tout n , T_{n+1} est indépendante des variables aléatoires $\tilde{X}_1, \dots, \tilde{X}_{n+1}, T_1, \dots, T_n$ conditionnellement à \tilde{X}_{n+1} et la loi de T_{n+1} sachant \tilde{X}_{n+1} ne dépend pas de n .*

Sous ces conditions, la transition q et la loi de durée d sont données par :

- $q(\omega_j | \omega_i) = p(\tilde{x}_{n+1} = \omega_j | \tilde{x}_n = \omega_i)$;
- $d(\omega_j, u) = p(t_{n+1} = u | \tilde{x}_{n+1} = \omega_j)$.

Preuve.

La première implication est évidente. Montrons la deuxième implication. Notons $D(\omega_j, u) = \sum_{t \geq u} d(\omega_j, t)$, le processus X est une chaîne semi-markovienne de transition q et de loi de durée d si et seulement si :

$$\begin{aligned} & p(\underbrace{x_1 = \tilde{x}_1, \dots, x_{t_1} = \tilde{x}_1}_{t_1 \text{ fois}}, \underbrace{x_{t_1+1} = \tilde{x}_2, \dots, x_{t_1+t_2} = \tilde{x}_2}_{t_2 \text{ fois}}, \\ & \dots, \underbrace{x_{t_1+\dots+t_{L-2}+1} = \tilde{x}_{L-1}, \dots, x_{t_1+\dots+t_{L-1}} = \tilde{x}_{L-1}}_{t_{L-1} \text{ fois}}, \underbrace{x_{t_1+\dots+t_{L-1}+1} = \tilde{x}_L, \dots, x_N = \tilde{x}_L}_{N - (t_1 + \dots + t_{L-1}) \text{ fois}}) = \\ & \pi(\tilde{x}_1) d(\tilde{x}_1, t_1) \times \prod_{l=1}^{L-2} q(\tilde{x}_{l+1} | \tilde{x}_l) d(\tilde{x}_{l+1}, t_{l+1}) \times q(\tilde{x}_L | \tilde{x}_{L-1}) D(\tilde{x}_L, N - (t_1 + \dots + t_{L-1})) \end{aligned} \quad (3.7)$$

Supposons que le processus X satisfasse les conditions 1. et 2. et montrons (3.7). La loi de tout processus X s'exprime en fonction de celles de T et de \tilde{X} par :

$$\begin{aligned} & p(\underbrace{x_1 = \tilde{x}_1, \dots, x_{t_1} = \tilde{x}_1}_{t_1 \text{ fois}}, \underbrace{x_{t_1+1} = \tilde{x}_2, \dots, x_{t_1+t_2} = \tilde{x}_2}_{t_2 \text{ fois}}, \\ & \dots, \underbrace{x_{t_1+\dots+t_{L-2}+1} = \tilde{x}_{L-1}, \dots, x_{t_1+\dots+t_{L-1}} = \tilde{x}_{L-1}}_{t_{L-1} \text{ fois}}, \underbrace{x_{t_1+\dots+t_{L-1}+1} = \tilde{x}_L, \dots, x_N = \tilde{x}_L}_{N - (t_1 + \dots + t_{L-1}) \text{ fois}}) = \\ & \pi(\tilde{x}_1) p(t_1 | \tilde{x}_1) \times \prod_{l=1}^{L-2} p(\tilde{x}_{l+1} | \tilde{x}_1, \dots, \tilde{x}_l, t_1, \dots, t_l) p(t_{l+1} | \tilde{x}_1, \dots, \tilde{x}_{l+1}, t_1, \dots, t_l) \\ & \times p(\tilde{x}_L | \tilde{x}_1, \dots, \tilde{x}_{L-1}, t_1, \dots, t_{L-1}) p(t_L \geq N - (t_1 + \dots + t_{L-1}) | \tilde{x}_1, \dots, \tilde{x}_L, t_1, \dots, t_{L-1}). \end{aligned}$$

En utilisant les conditions 1. et 2., on en déduit le résultat. \square

Chaînes semi-markoviennes comme généralisation des chaînes de Markov

La proposition suivante donne les conditions nécessaire et suffisante pour qu'une chaîne semi-markovienne soit une chaîne de Markov.

Proposition 3.3.2. *On reprend les notations de la définition 3.3.1. Une chaîne semi-markovienne $X = (X_n)_{n \geq 1}$ à valeurs dans un espace d'état fini \mathcal{X} est une chaîne de Markov de matrice de transition $Q(x, x') = p(x_{n+1} = x' | x_n = x)$ si et seulement si :*

$$d(x, u) = Q(x, x)^{u-1} (1 - Q(x, x)) \text{ pour tout } x \text{ et pour tout } u \geq 1.$$

De plus, si cette condition est vraie, on a :

$$q(x'|x) = \frac{Q(x, x')}{1 - Q(x, x)} \text{ pour tout } x, x' \text{ tels que } x \neq x'.$$

Conditions de stationnarité d'une chaîne semi-markovienne

Nous donnons des conditions suffisantes pour qu'une chaîne semi-markovienne à valeurs dans un espace fini soit stationnaire. Dans [74] il y est étudié la stationnarité de chaînes semi-markoviennes dans un cadre plus général. Lorsqu'un processus $Z = (Z_n)_{n \geq 1}$ est stationnaire, nous appellerons "mesure stationnaire" la mesure de probabilité commune de la loi des Z_n . Si Z est une chaîne de Markov à espace d'états fini de transition Q , sa loi initiale π définit une mesure stationnaire si et seulement si $\pi Q = \pi$, auquel cas, π est appelée "mesure invariante". Soit (X, U) la chaîne de Markov introduite dans la définition 3.3.1 avec $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$. Supposons que (X, U) admette une mesure invariante. La mesure invariante μ de la chaîne (X, U) vérifie alors :

$$\begin{aligned} \mu(x_{n+1}, u_{n+1}) &= \sum_{x_n, u_n} \mu(x_n, u_n) p(x_{n+1} | x_n, u_n) p(u_{n+1} | x_{n+1}, u_n) \\ &= \sum_{x_n \neq x_{n+1}} \mu(x_n, 1) q(x_{n+1} | x_n) d(x_{n+1}, u_{n+1}) + \mu(x_{n+1}, u_{n+1} + 1). \end{aligned} \quad (3.8)$$

La mesure μ^{semi} obtenue par marginalisation $\mu^{\text{semi}}(x) = \sum_u \mu(x, u)$ est une mesure stationnaire pour X . Celle-ci satisfait :

$$\mu^{\text{semi}}(x_{n+1}) = \sum_{x_n \neq x_{n+1}} \mu(x_n, 1) q(x_{n+1} | x_n) + (\mu^{\text{semi}}(x_{n+1}) - \mu(x_{n+1}, 1)).$$

Les $\mu_{x_n} = \mu(x_n, 1)$ sont alors solutions du système :

$$\mu(x_{n+1}, 1) = \sum_{x_n \neq x_{n+1}} \mu(x_n, 1) q(x_{n+1} | x_n). \quad (3.9)$$

Ainsi la mesure $(\mu_{x_n})_{x_n \in \mathcal{X}}$ est une mesure invariante de la chaîne immergée. On peut également exprimer à partir de la formule (3.8), la mesure stationnaire μ^{semi} en fonction de la mesure

$(\mu_{x_n})_{x_n \in \mathcal{X}}$.

D'après (3.8), on a :

$$\mu(\omega_j, 2) = \mu_{\omega_j} - \sum_{i \neq j} \mu_{\omega_i} q(\omega_j | \omega_i) d(\omega_j, 1),$$

et comme les μ_{x_n} sont solutions de (3.9), on en déduit :

$$\mu(\omega_j, 2) = \mu_{\omega_j} (1 - d(\omega_j, 1)).$$

Procédant ainsi de manière récursive en exprimant $\mu(\omega_j, k)$ en fonction de $\mu(\omega_j, k-1)$ grâce à la formule (3.8), on en déduit que pour tout $k \geq 2$,

$$\mu(\omega_j, k) = \mu_{\omega_j} \left(1 - \sum_{l=1}^{k-1} d(\omega_j, l) \right).$$

La quantité $1 - \sum_{l=1}^{k-1} d(\omega_j, l)$ est la probabilité que le temps de séjour dans l'état ω_j soit supérieur ou égal à k . Par ailleurs, soit V_j la variable aléatoire de densité $d(\omega_j, \cdot)$, on a alors :

$$\mu^{\text{semi}}(\omega_j) = \mu_{\omega_j} \sum_{k=1}^{+\infty} \mathbb{P}(V_j \geq k),$$

Remarquant que $\sum_{k=1}^{+\infty} \mathbb{P}(V_j \geq k) = \mathbb{E}(V_j)$, on a finalement la proposition suivante :

Proposition 3.3.3. *Soit la chaîne de Markov (X, U) dont la loi est définie par (3.3), (3.4), (3.5) et (3.6). Si (X, U) admet une mesure invariante μ , alors la chaîne immergée de la chaîne semi-markovienne X admet une mesure invariante $\tilde{\mu}$ donnée par :*

$$\tilde{\mu}(x) = \mu(x, 1) \text{ pour tout } x \in \mathcal{X}.$$

Définissons la mesure de probabilité π^{semi} par :

$$\pi^{\text{semi}}(x) \propto \tilde{\mu}(x) \mathbb{E}(V_x) \text{ pour tout } x \in \mathcal{X},$$

où V_x est une variable aléatoire de densité $u \in \mathbb{N}^* \rightarrow d(x, u)$. Si la loi de X_1 a pour distribution π^{semi} , alors la chaîne semi-markovienne est stationnaire.

Exemple 3.3.1. Supposons que la chaîne semi-markovienne X prenne deux valeurs ω_1 et ω_2 et supposons que (X, U) admette une mesure invariante. La chaîne immergée est alors de matrice de transition $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ et donc toute mesure invariante $\tilde{\mu}$ de la chaîne immergée \tilde{X} satisfait $\tilde{\mu}(\omega_1) = \tilde{\mu}(\omega_2)$. Notons λ_j le temps de séjour moyen dans l'état ω_j , la probabilité π^{semi} définie par :

$$\begin{aligned} \pi^{\text{semi}}(\omega_1) &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \\ \pi^{\text{semi}}(\omega_2) &= \frac{\lambda_2}{\lambda_1 + \lambda_2}, \end{aligned}$$

est une probabilité invariante pour la chaîne semi-markovienne X .

3.3.2 Etudes des chaînes semi-markoviennes à temps fini

Dans cette section, nous nous intéressons à des processus de la forme $X_{1:N} = (X_n)_{1 \leq n \leq N}$. La proposition suivante permet de considérer un nombre fini de valeurs pour chaque U_n tout en conservant la loi voulue pour le processus $X_{1:N}$.

Proposition 3.3.4. *Soit (X, U) la chaîne de Markov définie par (3.4), (3.5) et (3.6). Notons $D(x, n) = \sum_{k \geq n} d(x, k)$ la probabilité que le temps de séjour dans l'état x soit supérieure ou égale à n .*

Le processus à temps fini $X_{1:N} = (X_1, \dots, X_N)$ est marginal de la chaîne de Markov $(X_{1:N}, W_{1:N})$ à valeurs dans $\mathcal{X} \times \{1, \dots, N\}$ dont la loi est donnée par :

– *Loi initiale :*

$$p(x_1, w_1) = \pi(x_1)p(w_1|x_1), \quad (3.10)$$

avec :

$$p(w_1|x_1) = \begin{cases} d(x_1, w_1) & \text{si } 1 \leq w_1 \leq N-1, \\ D(x_1, N) & \text{si } w_1 = N. \end{cases} \quad (3.11)$$

– *Transition :*

$$p(x_{n+1}, w_{n+1}|x_n, w_n) = p(x_{n+1}|x_n, w_n) \times p(w_{n+1}|x_{n+1}, w_n), \quad (3.12)$$

avec :

$$p(x_{n+1}|x_n, w_n) = \begin{cases} \delta_{x_n}(x_{n+1}) & \text{si } w_n > 1, \\ q(x_{n+1}|x_n) & \text{si } w_n = 1, \end{cases} \quad (3.13)$$

et

$$p(w_{n+1}|x_{n+1}, w_n) = \begin{cases} \delta_{w_n-1}(w_{n+1}) & \text{si } w_n > 1, \\ d(x_{n+1}, w_{n+1}) & \text{si } w_n = 1 \text{ et } 1 \leq w_{n+1} \leq N-1, \\ D(x_{n+1}, N) & \text{si } w_n = 1 \text{ et } w_{n+1} = N, \end{cases} \quad (3.14)$$

où $\delta_x(x) = 1$ et $\delta_x(x') = 0$ pour $x' \neq x$.

Preuve de la proposition 3.3.4. Notons $(X_{1:N}^W, W_{1:N})$ le processus dont la loi est définie par (3.10), (3.11), (3.12), (3.13) et (3.14).

Tout d'abord, intéressons nous à la loi de $X_{1:N}$. On notera L le nombre de valeurs visitées par la chaîne $X_{1:N}$ et $\tilde{x}_1, \dots, \tilde{x}_L$ la suite des valeurs visitées. On a alors :

$$\begin{aligned} p(x_{1:N}) &= p(\underbrace{\tilde{x}_1, \dots, \tilde{x}_1}_{u_1 \text{ fois}}, \underbrace{\tilde{x}_2, \dots, \tilde{x}_2}_{u_2 \text{ fois}}, \dots, \underbrace{\tilde{x}_L, \dots, \tilde{x}_L}_{N-(u_1+\dots+u_{L-1}) \text{ fois}}) \\ &= \pi(\tilde{x}_1) \prod_{j=1}^{L-1} q(\tilde{x}_{j+1}|\tilde{x}_j)d(\tilde{x}_j, u_j) \times D(\tilde{x}_L, N - (u_1 + \dots + u_{L-1})). \end{aligned}$$

Concernant la loi de $X_{1:N}^W$, l'événement :

$$x_{1:N}^W = \left(\underbrace{\tilde{x}_1, \dots, \tilde{x}_1}_{u_1 \text{ fois}}, \underbrace{\tilde{x}_2, \dots, \tilde{x}_2}_{u_2 \text{ fois}}, \dots, \underbrace{\tilde{x}_L, \dots, \tilde{x}_L}_{N-(u_1+\dots+u_{L-1}) \text{ fois}} \right)$$

est la réunion des événements A_j avec $j \in \{0, \dots, u_1 + \dots + u_{L-1}\}$ suivants :

$$(x_1^W = \tilde{x}_1, w_1 = u_1, x_{u_1+1}^W = \tilde{x}_2, w_{u_1+1} = u_2, \dots, x_{u_1+\dots+u_{L-2}+1}^W = \tilde{x}_{L-1}, w_{u_1+\dots+u_{L-2}+1} = u_{L-1}, \\ \dots, x_{u_1+\dots+u_{L-1}+1}^W = \tilde{x}_L, w_{u_1+\dots+u_{L-1}+1} = N - (u_1 + \dots + u_{L-1}) + j).$$

Pour tout $j < u_1 + \dots + u_{L-1}$, la probabilité de l'événement A_j est égale à :

$$\pi(\tilde{x}_1) \prod_{l=1}^{L-1} q(\tilde{x}_l, \tilde{x}_{l+1}) d(\tilde{x}_l, u_l) \times d(\tilde{x}_L, N - (u_1 + \dots + u_{L-1}) + j),$$

et pour $j = u_1 + \dots + u_{L-1}$, est égale à :

$$\pi(\tilde{x}_1) \prod_{l=1}^{L-1} q(\tilde{x}_l, \tilde{x}_{l+1}) d(\tilde{x}_l, u_l) \times D(\tilde{x}_L, N).$$

La réunion des événements A_j est donc de probabilité :

$$\pi(\tilde{x}_1) \prod_{j=1}^{L-1} q(\tilde{x}_j, \tilde{x}_{j+1}) d(\tilde{x}_j, u_j) \times D(\tilde{x}_L, N - (u_1 + \dots + u_{L-1})).$$

On conclut que $X_{1:N}$ et $X_{1:N}^W$ ont les mêmes lois. □

De cette proposition, on conclut que toute chaîne semi-markovienne peut être considérée comme la marginale d'une chaîne de Markov à espace d'état fini, pourvu que l'on ne considère qu'une partie finie $X_{1:N} = (X_n)_{1 \leq n \leq N}$ du processus. Ainsi un modèle latent $(X_n, Y_n)_{1 \leq n \leq N}$ tel que $(X_n)_{1 \leq n \leq N}$ soit une chaîne semi-markovienne peut être considérée comme une chaîne de Markov triplet telle que chaque (X_n, U_n) prend un nombre fini de valeurs, nous appellerons un tel modèle "chaîne semi-markovienne cachée". Cependant, même avec cette considération, l'algorithme de Baum-Welsh possède une forte complexité algorithmique, étant donné que si le processus X prend K états et le processus U prend N états, l'espace d'états a un cardinal égal à $N \times K$. Nous proposons dans la sous-section suivante un modèle semi-markovien particulier qui permet de palier ces difficultés algorithmiques.

3.3.3 Un modèle semi-markovien particulier

Dans cette sous-section, nous proposons un modèle semi-markovien particulier original. Dans ce nouveau modèle, on peut restreindre le nombre de valeurs du processus auxiliaire U tout en s'autorisant des temps de séjour arbitrairement longs. Soit $(X, U) = (X_n, U_n)_{1 \leq n \leq N}$ tel que chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ et chaque U_n prend ses valeurs dans un ensemble fini Λ . On considère :

- π une distribution de probabilité sur \mathcal{X} ;

- des distributions de probabilité $\bar{d}(x, \cdot)$ sur Λ pour tout $x \in \mathcal{X}$;
- des transitions $r(\cdot|\cdot)$ sur \mathcal{X}^2 .

La distribution $p(x_{1:N}, u_{1:N})$ de (X, U) est définie par :

$$p(x_{1:N}, u_{1:N}) = \pi(x_1) \bar{d}(x_1, u_1) \prod_{n=1}^{N-1} p(x_{n+1}|x_n, u_n) p(u_{n+1}|x_{n+1}, u_n), \quad (3.15)$$

où :

$$p(x_{n+1}|x_n, u_n) = \begin{cases} \delta_{x_n}(x_{n+1}) & \text{si } u_n > 1, \\ r(x_{n+1}|x_n) & \text{si } u_n = 1, \end{cases} \quad (3.16)$$

et

$$p(u_{n+1}|x_{n+1}, u_n) = \begin{cases} \delta_{u_{n-1}}(u_{n+1}) & \text{si } u_n > 1, \\ \bar{d}(x_{n+1}, u_{n+1}) & \text{si } u_n = 1. \end{cases} \quad (3.17)$$

A la différence avec le modèle semi-markovien classique, ici la transition $r(x|x)$ peut très bien être non nulle.

Nous allons montrer que ce modèle est un cas particulier de modèle semi-markovien et nous allons exprimer la loi du temps de séjour qui nous sera utile par la suite.

Semi-markovianité de X et loi du temps de séjour

Introduisons la chaîne \check{X} construite de la manière suivante :

$$\check{X}_k = X_{n_k}, \quad \text{où} \\ n_1 = 1 \quad \text{et} \quad \forall k \geq 2, \quad n_k = \min \{n \geq n_{k-1} : U_n = 1\} + 1.$$

La chaîne \check{X} est une chaîne de Markov de transition r et possède la même chaîne immergée que le processus X ; notons \check{X} cette chaîne immergée. Soient $T = (T_n)_{n \geq 1}$ et $\check{T} = (\check{T}_n)_{n \geq 1}$ les processus respectifs de temps de séjour de X et \check{X} . Considérons également la chaîne \check{U} définie par :

$$\check{U}_k = U_{n_k}, \quad \text{où} \\ n_1 = 1 \quad \text{et} \quad \forall k \geq 2, \quad n_k = \min \{n \geq n_{k-1} : U_n = 1\} + 1.$$

L'événement $(T_n = s)$ est égal à la réunion disjointe :

$$(T_n = s) = \bigsqcup_{k=1}^s [(\check{T}_n = k) \cap (\check{U}_{\check{T}_1 + \dots + \check{T}_{n-1} + 1} + \check{U}_{\check{T}_1 + \dots + \check{T}_{n-1} + 2} + \dots + \check{U}_{\check{T}_1 + \dots + \check{T}_{n-1} + k} = s)].$$

D'après la définition du modèle, $p(\tilde{x}_{n+1}|\tilde{x}_1, \dots, \tilde{x}_n, t_1, \dots, t_n) = \frac{r(\tilde{x}_{n+1}|\tilde{x}_n)}{1 - r(\tilde{x}_n|\tilde{x}_n)}$ et la propriété 1.

de la proposition 3.3.1 est satisfaite. Les variables aléatoires $\check{U}_{\check{T}_1 + \dots + \check{T}_{n-1} + l}$ correspondent aux valeurs de U_k lorsque $U_{k-1} = 1$ et $k \in \{T_1 + \dots + T_{n-1} + 1, T_1 + \dots + T_n\}$, elles sont donc indépendantes conditionnellement à \check{X}_n et indépendantes de $\check{X}_1, \dots, \check{X}_{n-1}, T_1, \dots, T_{n-1}, \check{T}_n$ conditionnellement à \check{X}_n . De plus, la loi de $\check{U}_{\check{T}_1 + \dots + \check{T}_{n-1} + l}$ sachant $\check{X}_n = \tilde{x}_n$ est donnée par $\bar{d}(\tilde{x}_n, \cdot)$. De même, \check{T}_n est le temps de séjour de la chaîne de Markov \check{X} dans l'état \check{X}_n , ainsi

\check{T}_n est indépendante de $\check{X}_1, \dots, \check{X}_{n-1}, T_1, \dots, T_{n-1}$ conditionnellement à \check{X}_n . De plus, comme \check{X} est une chaîne de Markov de transition r , la loi de \check{T}_n conditionnellement à \check{X}_n est donnée par $p(\check{t}_n|\check{x}_n) = (1 - r(\check{x}_n|\check{x}_n))r(\check{x}_n|\check{x}_n)^{\check{t}_n-1}$. On en déduit que la propriété 2. de la proposition 3.3.1 est satisfaite. Ainsi, X est une chaîne semi-markovienne, sa transition et sa loi de durée sont données par :

$$\forall \omega_i \neq \omega_j, q(\omega_j|\omega_i) = \frac{r(\omega_j|\omega_i)}{1 - r(\omega_i|\omega_i)} ; \quad (3.18)$$

$$d(\omega_j, s) = (1 - r(\omega_j|\omega_j)) \sum_{k=1}^s r(\omega_j|\omega_j)^{k-1} \sum_{v_1 + \dots + v_k = s} \bar{d}(\omega_j, v_1) \dots \bar{d}(\omega_j, v_k) \text{ pour tout } s \geq 1. \quad (3.19)$$

L'expression ci-dessus ne permet pas de retrouver n'importe quelle loi de temps de séjour $d(\omega_j, \cdot)$, le modèle ainsi défini est donc un modèle semi-markovien particulier. Par contre, il est possible de trouver des paramètres $(\bar{d}(\omega_j, \cdot), r(\omega_j|\omega_j))$ tels que la loi du temps de séjour soit géométrique. Notre modèle reste plus général que le modèle de chaîne de Markov.

Identifiabilité du modèle

Dans ce paragraphe, nous abordons l'identifiabilité du modèle dans le cas où chaque U_n prend ses valeurs dans $\Lambda = \{1, 2\}$. Ainsi, les paramètres de notre modèle sont les $K(K-1)$ transitions $r(\omega_j|\omega_i)$ où $i \in \{1, \dots, K\}$ et $j \in \{1, \dots, K-1\}$ et les K paramètres $\bar{d}(\omega_j, 1)$. Le modèle étant inclus dans celui des chaînes semi-markoviennes, il est déterminé par la loi du temps de séjour donnée par $(d(\omega_j, \cdot))_{1 \leq j \leq K}$ et par les transitions $(q(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$. Le modèle sera identifiable si pour une loi de temps de séjour et des transitions $q(\omega_j|\omega_i)$ données, il existe un unique jeu de paramètres $(\bar{d}(\omega_j, 1))_{1 \leq j \leq K}, (r(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$. De plus, lorsque $\omega_i \neq \omega_j, q(\omega_j|\omega_i) = \frac{r(\omega_j|\omega_i)}{1 - r(\omega_i|\omega_i)}$, car X et \check{X} ont même chaîne immergée et \check{X} est une chaîne de Markov de transition $(r(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$. Ainsi, la donnée de $q(\omega_j|\omega_i)$ et de $r(\omega_i|\omega_i)$ permet de déduire la valeur de $r(\omega_j|\omega_i)$. Le modèle sera donc identifiable si pour tout j , les probabilités $r(\omega_j|\omega_j)$ et $\bar{d}(\omega_j, 1)$ sont uniques pour une loi de temps de séjour et une transition $(q(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$ données.

Soient $(q(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$ et $d(\omega_j, \cdot)_{1 \leq j \leq K}$ les paramètres définissant la loi d'une chaîne semi-markovienne. D'après (3.19), le modèle défini par les formules (3.15), (3.16) et (3.17) existera pour les paramètres $(q(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$ et $d(\omega_j, \cdot)_{1 \leq j \leq K}$ si pour tout $j \in \{1, \dots, K\}$ le système suivant :

$$S_j : \begin{cases} d(\omega_j, 1) = & (1 - r(\omega_j|\omega_j))\bar{d}(\omega_j, 1) \\ d(\omega_j, 2) = & (1 - r(\omega_j|\omega_j)) [1 - \bar{d}(\omega_j, 1) + r(\omega_j|\omega_j)\bar{d}(\omega_j, 1)^2] \\ \vdots & \\ d(\omega_j, 2m) = & (1 - r(\omega_j|\omega_j)) \left[\sum_{l=0}^m C_{m+l}^{2l} r(\omega_j|\omega_j)^{m+l-1} \bar{d}(\omega_j, 1)^{2l} (1 - \bar{d}(\omega_j, 1))^{m-l} \right] \\ d(\omega_j, 2m+1) = & (1 - r(\omega_j|\omega_j)) \left[\sum_{l=1}^{m+1} C_{m+l}^{2l-1} r(\omega_j|\omega_j)^{m+l-1} \bar{d}(\omega_j, 1)^{2l-1} (1 - \bar{d}(\omega_j, 1))^{m-l+1} \right] \\ \vdots & \end{cases}$$

a une solution. Il sera identifiable si pour tout $j \in \{1, \dots, K\}$, la solution est unique.

On supposera dans la suite que pour tout j , $r(\omega_j|\omega_j) \neq 1$; dans le cas contraire, le temps de séjour dans l'état ω_j serait infini. Ainsi, en remplaçant dans la deuxième équation $\bar{d}(\omega_j, 1)$ par

$\frac{d(\omega_j, 1)}{1 - r(\omega_j|\omega_j)}$, on prouve que toute solution du système S_j est solution du système suivant :

$$S'_j : \begin{cases} d(\omega_j, 1) = (1 - r(\omega_j|\omega_j))\bar{d}(\omega_j, 1) \\ r(\omega_j|\omega_j)^2 - (2 - d(\omega_j, 1) - d(\omega_j, 2) - d(\omega_j, 1)^2)r(\omega_j|\omega_j) + 1 - d(\omega_j, 1) - d(\omega_j, 2) = 0 \end{cases}$$

Soit f la fonction définie sur $[0, 1]$ par :

$$f(r) = r^2 - (2 - d(\omega_j, 1) - d(\omega_j, 2) - d(\omega_j, 1)^2)r + 1 - d(\omega_j, 1) - d(\omega_j, 2),$$

celle-ci est décroissante pour

$$r \leq 1 - \frac{d(\omega_j, 1) + d(\omega_j, 2) + d(\omega_j, 1)^2}{2},$$

et croissante pour

$$r \geq 1 - \frac{d(\omega_j, 1) + d(\omega_j, 2) + d(\omega_j, 1)^2}{2}.$$

Notons

$$r_{\min} = 1 - \frac{d(\omega_j, 1) + d(\omega_j, 2) + d(\omega_j, 1)^2}{2},$$

on a $r_{\min} \in [0, 1]$, $f(0) = 1 - d(\omega_j, 1) - d(\omega_j, 2) \geq 0$, $f(1) = d(\omega_j, 1)^2 \geq 0$ et

$$f(r_{\min}) = \left(\frac{3d(\omega_j, 1) + d(\omega_j, 2) + d(\omega_j, 1)^2}{2} \right) \left(\frac{d(\omega_j, 1) - d(\omega_j, 2) - d(\omega_j, 1)^2}{2} \right).$$

Nous avons trois possibilités :

- soit $d(\omega_j, 1)(1 - d(\omega_j, 1)) > d(\omega_j, 2)$, alors $f(r_{\min}) > 0$ et donc le système S'_j n'a pas de solution ; ce qui implique que le système S_j n'a pas de solution non plus ;
- soit $d(\omega_j, 1)(1 - d(\omega_j, 1)) = d(\omega_j, 2)$, alors $f(r_{\min}) = 0$. Si $f(0) > 0$ et $f(1) > 0$, le système S'_j a une unique solution. Cette solution est donnée par $r(\omega_j|\omega_j) = r_{\min} = 1 - d(\omega_j, 1)$ et $\bar{d}(\omega_j, 1) = 1$. Pour que cette solution soit également solution de S_j , nécessairement $d(\omega_j, \cdot)$ doit être une loi géométrique. Si $f(1) = 0$, alors $d(\omega_j, 1) = 0$ et sous la condition $d(\omega_j, 1)(1 - d(\omega_j, 1)) = d(\omega_j, 2)$, on a $d(\omega_j, 2) = 0$. Une solution $(r(\omega_j|\omega_j), \bar{d}(\omega_1, 1))$ de S_j , si elle existe, doit nécessairement satisfaire $r(\omega_j|\omega_j) = 1$, ce qui est exclu. Si $f(0) = 0$, alors sous la condition $d(\omega_j, 1)(1 - d(\omega_j, 1)) = d(\omega_j, 2)$, on a $d(\omega_j, 1)d(\omega_j, 2) = d(\omega_j, 2)$. Ainsi, $d(\omega_j, 1) = 1$ ou $d(\omega_j, 2) = 0$ et donc nécessairement $d(\omega_j, 2) = 0$. Alors, si S_j a une solution $(r(\omega_j|\omega_j), \bar{d}(\omega_1, 1))$ celle-ci doit satisfaire $1 - \bar{d}(\omega_j, 1) + r(\omega_j|\omega_j)\bar{d}(\omega_j, 1)^2 = 0$. Cette dernière équation implique $\bar{d}(\omega_j, 1) = 1$ et $r(\omega_j|\omega_j) = 0$ et donc $d(\omega_j, 1) = 1$. On en déduit que $r_{\min} = 0$. Ainsi le système S'_j a une unique solution donnée par $r(\omega_j|\omega_j) = r_{\min} = 0$ et $\bar{d}(\omega_j, 1) = 1$. Sous la condition $d(\omega_j, 1) = 1$, la solution de S_j existe et est l'unique solution de S'_j ;
- soit $d(\omega_j, 1)(1 - d(\omega_j, 1)) < d(\omega_j, 2)$, alors $f(r_{\min}) < 0$ et le système S'_j a deux solutions

$(r_1(\omega_j|\omega_j), \bar{d}_1(\omega_j, 1))$ et $(r_2(\omega_j|\omega_j), \bar{d}_2(\omega_j, 1))$. On prouve que toute solution du système S_j est solution du système suivant :

$$S_j'' : \begin{cases} d(\omega_j, 1) = (1 - r(\omega_j|\omega_j))\bar{d}(\omega_j, 1) \\ r(\omega_j|\omega_j)^2 - (2 - d(\omega_j, 1) - d(\omega_j, 2) - d(\omega_j, 1)^2) r(\omega_j|\omega_j) + 1 - d(\omega_j, 1) - d(\omega_j, 2) = 0 \\ d(\omega_j, 1)r(\omega_j|\omega_j)^2 + [d(\omega_j, 1)(d(\omega_j, 1) - d(\omega_j, 2) - 1) - d(\omega_j, 3)] r(\omega_j|\omega_j) + d(\omega_j, 3) = 0 \end{cases}$$

et toute solution de S_j'' est solution de S_j' . Si aucune solution de S_j' n'est solution de S_j'' , alors le système S_j n'a pas de solution. Si une seule solution de S_j' est solution de S_j'' , alors S_j'' possède une unique solution. Notons $(r^0(\omega_j|\omega_j), \bar{d}^0(\omega_j, 1))$ la solution de S_j'' . Le système S_j possède une solution si et seulement si le temps de séjour $d(\omega_j, \cdot)$ satisfait :

$$\begin{cases} d(\omega_j, 1) = & (1 - r^0(\omega_j|\omega_j))\bar{d}^0(\omega_j, 1) \\ d(\omega_j, 2) = & (1 - r^0(\omega_j|\omega_j)) [1 - \bar{d}^0(\omega_j, 1) + r^0(\omega_j|\omega_j)\bar{d}^0(\omega_j, 1)^2] \\ \vdots & \\ d(\omega_j, 2m) = & (1 - r^0(\omega_j|\omega_j)) \left[\sum_{l=0}^m C_{m+l}^{2l} r^0(\omega_j|\omega_j)^{m+l-1} \bar{d}^0(\omega_j, 1)^{2l} (1 - \bar{d}^0(\omega_j, 1))^{m-l} \right] \\ d(\omega_j, 2m+1) = & (1 - r^0(\omega_j|\omega_j)) \left[\sum_{l=1}^{m+1} C_{m+l}^{2l-1} r^0(\omega_j|\omega_j)^{m+l-1} \bar{d}^0(\omega_j, 1)^{2l-1} (1 - \bar{d}^0(\omega_j, 1))^{m-l+1} \right] \\ \vdots & \end{cases}$$

dans ce cas la solution de S_j est l'unique solution $(r^0(\omega_j|\omega_j), \bar{d}^0(\omega_j, 1))$ de S_j'' . Si les deux solutions $(r_1(\omega_j|\omega_j), \bar{d}_1(\omega_j, 1))$ et $(r_2(\omega_j|\omega_j), \bar{d}_2(\omega_j, 1))$ de S_j' sont solutions de S_j'' , alors les deux dernières équations de S_j'' d'inconnue $r(\omega_j|\omega_j)$ sont proportionnelles. Ainsi :

$$\begin{cases} d(\omega_j, 3) = d(\omega_j, 1) (1 - d(\omega_j, 1) - d(\omega_j, 2)) \\ d(\omega_j, 3) - d(\omega_j, 1) (d(\omega_j, 1) - d(\omega_j, 2) - 1) = d(\omega_j, 1) (2 - d(\omega_j, 1) - d(\omega_j, 2) - d(\omega_j, 1)^2) \end{cases}$$

Ce dernier système implique que $d(\omega_j, 1) (d(\omega_j, 1)^2 - d(\omega_j, 1) - d(\omega_j, 2)) = 0$ et donc $d(\omega_j, 1) = 0$. Ainsi, $\bar{d}(\omega_j, 1) = 0$ ou $r(\omega_j|\omega_j) = 1$. La solution de S_j'' telle que $r(\omega_j|\omega_j) = 1$ est exclue. Ainsi $\bar{d}(\omega_j, 1) = 0$, ce qui implique que le temps de séjour dans ω_j doit nécessairement être pair et le temps de séjour de la chaîne $(X_{2n+1})_{n \geq 0}$ dans ω_j doit nécessairement être une loi géométrique pour que cette solution soit également solution de S_j .

On en déduit que sous les conditions d'existence des paramètres $(\bar{d}(\omega_j, 1))_{1 \leq j \leq K}$ et $(r(\omega_j|\omega_i))_{1 \leq i \leq K, 1 \leq j \leq K}$, le modèle est identifiable.

Condition de markovianité

Nous donnons les conditions sur $(r(\omega_j|\omega_i))_{i \in \{1, \dots, K\}, j \in \{1, \dots, K\}}$ et $\bar{d}(\omega_j, \cdot)_{j \in \{1, \dots, K\}}$ pour que X soit une chaîne de Markov.

Proposition 3.3.5. *Soit (X, U) la chaîne de Markov définie par les formules (3.15), (3.16) et (3.17). La chaîne semi-markovienne X est une chaîne de Markov de matrice de transition Q si et seulement si :*

$$\begin{cases} \bar{d}(\omega_j, 1) = 1 \\ r(\omega_j|\omega_i) = Q(\omega_i, \omega_j) \end{cases} \quad (3.20)$$

Preuve. Nous allons tout d'abord démontrer que X est une chaîne de Markov si et seulement si :

$$\forall j, \forall s \geq 1, \bar{d}(\omega_j, s) = \bar{d}(\omega_j, 1)(1 - \bar{d}(\omega_j, 1))^{s-1}. \quad (3.21)$$

Supposons d'abord que X soit une chaîne de Markov de transition Q .

Nous allons démontrer (3.21) par récurrence sur s . Il est clair que (3.21) est vraie pour $s = 1$.

Supposons le vrai pour tout $k \leq s - 1$. On a alors :

$$d(\omega_j, s) = (1 - r(\omega_j|\omega_j)) \left[\bar{d}(\omega_j, s) + \sum_{k=2}^s r(\omega_j|\omega_j)^{k-1} \bar{d}(\omega_j, 1)^k \sum_{v_1+\dots+v_k=s} (1 - \bar{d}(\omega_j, 1))^{s-k} \right]. \quad (3.22)$$

Par un calcul de dénombrement classique, on montre que le nombre d'éléments (v_1, \dots, v_k) tels que $v_j \geq 1$ pour tout j et $v_1 + \dots + v_k = s$ vaut C_{s-1}^{k-1} . Ainsi la formule (3.22) devient :

$$d(\omega_j, s) = (1 - r(\omega_j|\omega_j)) \left[\bar{d}(\omega_j, s) + \sum_{k=2}^s C_{s-1}^{k-1} r(\omega_j|\omega_j)^{k-1} \bar{d}(\omega_j, 1)^k (1 - \bar{d}(\omega_j, 1))^{s-k} \right]. \quad (3.23)$$

Sous la condition de markovianité on montre que $r(\omega_j|\omega_j)\bar{d}(\omega_j, 1) = Q(\omega_j, \omega_j) - 1 + \bar{d}(\omega_j, 1)$, d'où :

$$d(\omega_j, s) = \frac{1 - Q(\omega_j, \omega_j)}{\bar{d}(\omega_j, 1)} \left[\bar{d}(\omega_j, s) + \bar{d}(\omega_j, 1) \sum_{k=1}^{s-1} C_{s-1}^k (Q(\omega_j, \omega_j) - 1 + \bar{d}(\omega_j, 1))^k (1 - \bar{d}(\omega_j, 1))^{s-1-k} \right],$$

et donc comme de plus $d(\omega_j, s) = Q(\omega_j, \omega_j)^{s-1}(1 - Q(\omega_j, \omega_j))$, alors :

$$\bar{d}(\omega_j, s) + \bar{d}(\omega_j, 1) \sum_{k=1}^{s-1} C_{s-1}^k (Q(\omega_j, \omega_j) - 1 + \bar{d}(\omega_j, 1))^k (1 - \bar{d}(\omega_j, 1))^{s-1-k} = Q(\omega_j, \omega_j)^{s-1} \bar{d}(\omega_j, 1),$$

ainsi :

$$\bar{d}(\omega_j, s) + \bar{d}(\omega_j, 1) [Q(\omega_j, \omega_j)^{s-1} - (1 - \bar{d}(\omega_j, 1))^{s-1}] = Q(\omega_j, \omega_j)^{s-1} \bar{d}(\omega_j, 1),$$

on en déduit le résultat.

Étudions maintenant la réciproque, supposons (3.21) vraie pour tout s et montrons alors que X est une chaîne de Markov. Pour cela, on introduit $Q(\omega_j, \omega_j)$ tel que :

$$d(\omega_j, 1) = (1 - r(\omega_j|\omega_j))\bar{d}(\omega_j, 1) = 1 - Q(\omega_j, \omega_j).$$

En remplaçant alors dans l'expression de $d(\omega_j, s)$, $\bar{d}(\omega_j, s)$ par $\bar{d}(\omega_j, 1)(1 - \bar{d}(\omega_j, 1))^{s-1}$,

$r(\omega_j|\omega_j)\bar{d}(\omega_j, 1)$ par $Q(\omega_j, \omega_j) - 1 + \bar{d}(\omega_j, 1)$ et $1 - r(\omega_j|\omega_j)$ par $\frac{1 - Q(\omega_j, \omega_j)}{\bar{d}(\omega_j, 1)}$, on en déduit :

$$\begin{aligned} d(\omega_j, s) &= \frac{1 - Q(\omega_j, \omega_j)}{\bar{d}(\omega_j, 1)} \left[\bar{d}(\omega_j, 1) \sum_{k=0}^{s-1} C_{s-1}^k (Q(\omega_j, \omega_j) - 1 + \bar{d}(\omega_j, 1))^k (1 - \bar{d}(\omega_j, 1))^{s-1-k} \right] \\ &= (1 - Q(\omega_j, \omega_j))Q(\omega_j, \omega_j)^{s-1}, \end{aligned}$$

d'où le résultat.

Si $\Lambda = \{1, \dots, L\}$, on a $\sum_{s=1}^L \bar{d}(\omega_j, s) = 1$. Si X est de Markov de transition Q , comme $\bar{d}(\omega_j, s) = \bar{d}(\omega_j, 1)(1 - \bar{d}(\omega_j, 1))^{s-1}$, ainsi $(1 - \bar{d}(\omega_j, 1))^L = 0$ et donc $\bar{d}(\omega_j, 1) = 1$ et par conséquent $r(\omega_j|\omega_i) = Q(\omega_i, \omega_j)$. Réciproquement, si $\bar{d}(\omega_j, 1) = 1$, X est une chaîne de Markov de transition $(r(\omega_j|\omega_i))_{1 \leq i, j \leq K}$. \square

3.3.4 Inférence dans le modèle semi-markovien

Nous présentons maintenant l'estimation des états cachés et des paramètres dans le cas des chaînes semi-markoviennes cachées. De manière analogue aux chaînes de Markov cachées, les chaînes semi-markoviennes cachées sont représentées par un couple de processus $(X, Y) = (X_n, Y_n)_{1 \leq n \leq N}$ où la réalisation $y_{1:N}$ de Y est observée et celle de X , soit $x_{1:N}$, doit être estimée. Dans le modèle de chaînes semi-markoviennes cachées, X est une chaîne semi-markovienne et donc la marginale d'une chaîne de Markov (X, U) . En raison de la complexité algorithmique nous limiterons le nombre de valeurs prises par U à 10 et nous utiliserons le modèle semi-markovien particulier défini par les formules (3.15), (3.16) et (3.17). Le processus (X, U) étant alors une chaîne de Markov à espace d'états fini, on peut alors utiliser l'algorithme de Baum-Welsh pour estimer les états cachés. Nous allons commencer par étudier les chaînes semi-markoviennes cachées à bruit indépendant, puis nous les étendrons aux chaînes semi-markoviennes M -stationnaires à bruit indépendant, ce qui suppose l'introduction d'un deuxième processus auxiliaire.

Algorithme de Baum-Welsh dans les chaînes semi-markoviennes

Soit $(X, Y) = (X_n, Y_n)_{1 \leq n \leq N}$ la chaîne de Markov définie par (3.15), (3.16) et (3.17), où chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ et chaque U_n prend ses valeurs dans $\Lambda = \{1, \dots, L\}$. La distribution d'une chaîne semi-markovienne cachée à bruit indépendant $(X, U, Y) = (X_n, U_n, Y_n)_{n \in \{1, \dots, N\}}$ s'écrit :

$$p(x_{1:N}, u_{1:N}, y_{1:N}) = \pi(x_1) \bar{d}(x_1, u_1) p(y_1|x_1) \prod_{n=1}^{N-1} p(x_{n+1}|x_n, u_n) p(u_{n+1}|x_{n+1}, u_n) p(y_{n+1}|x_{n+1}),$$

où $p(x_{n+1}|x_n, u_n)$ et $p(u_{n+1}|x_{n+1}, u_n)$ sont donnés par les formules (3.16) et (3.17). Nous détaillerons uniquement la version classique de l'algorithme de Baum-Welsh, la version conditionnelle s'en déduisant sans difficulté particulière.

Remarquons que si $u_n > 1$ la transition $p(x_{n+1}, u_{n+1}|x_n, u_n)$ est non nulle seulement lorsque $x_{n+1} = x_n$ et $u_{n+1} = u_n - 1$; l'algorithme de Baum-Welsh est ainsi de complexité algorithmique plus faible que dans le cas d'une chaîne de Markov triplet générale pour laquelle X_n prend K valeurs et U_n prend L valeurs. Plus exactement, on a :

– Etape directe :

Pour $u_{n+1} \leq L - 1$:

$$\alpha_{n+1}(x_{n+1}, u_{n+1}) =$$

$$\sum_{x_n} \alpha_n(x_n, 1) r(x_{n+1}|x_n) \bar{d}(x_{n+1}, u_{n+1}) p(y_{n+1}|x_{n+1}) + \alpha_n(x_{n+1}, u_{n+1} + 1) p(y_{n+1}|x_{n+1}),$$

soit $K \times (L - 1)$ quantités à calculer nécessitant $K + 1$ sommations.

Pour $u_{n+1} = L$:

$$\alpha_{n+1}(x_{n+1}, L) = \sum_{x_n} \alpha_n(x_n, 1) r(x_{n+1}|x_n) \bar{d}(x_{n+1}, L) p(y_{n+1}|x_{n+1}),$$

soit K quantités à calculer nécessitant K sommations.

– Etape rétrograde :

Pour $u_n > 1$:

$$\beta_n(x_n, u_n) = \beta_{n+1}(x_n, u_n - 1) p(y_{n+1}|X_{n+1} = x_n),$$

soit $K \times (L - 1)$ quantités à calculer nécessitant 1 sommation.

Pour $u_n = 1$:

$$\beta_n(x_n, 1) = \sum_{x_{n+1}, u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1}) r(x_{n+1}|x_n) \bar{d}(x_{n+1}, u_{n+1}) p(y_{n+1}|x_{n+1}),$$

soit K quantités à calculer nécessitant $K \times L$ sommations.

La complexité algorithmique de l'algorithme de Baum-Welsh est alors égale à $\mathcal{O}(N \times K^2 \times L)$ au lieu de $\mathcal{O}(N \times K^2 \times L^2)$ dans le cas des chaînes de Markov triplets générales et au lieu de $\mathcal{O}(N^2 \times K^2)$ si on avait utilisé le modèle semi-markovien caché général.

Nous avons classiquement :

$$p(x_n, u_n | y_{1:N}) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n),$$

et

$$p(x_{n+1}, u_{n+1} | x_n, u_n, y_{1:N}) = \frac{\beta_{n+1}(x_{n+1}, u_{n+1})}{\beta_n(x_n, u_n)} p(x_{n+1}|x_n, u_n) p(u_{n+1}|x_{n+1}, u_n) p(y_{n+1}|x_{n+1}).$$

Les transitions vérifient :

$$p(x_{n+1}, u_{n+1} | x_n, u_n, y_{1:N}) \propto \begin{cases} \delta_{x_n}(x_{n+1}) \delta_{u_n-1}(u_{n+1}) & \text{si } u_n > 1 ; \\ \beta_{n+1}(x_{n+1}, u_{n+1}) r(x_{n+1}|x_n) \bar{d}(x_{n+1}, u_{n+1}) p(y_{n+1}|x_{n+1}) & \text{si } u_n = 1. \end{cases}$$

On remarque que la distribution $p(x_{1:N}, u_{1:N} | y_{1:N})$ peut être donnée par les formules (3.15), (3.16) et (3.17), avec $r(x_{n+1}|x_n)$ et $\bar{d}(x_n, \cdot)$ à remplacer par :

$$r(x_{n+1}|x_n, y_{1:N}) \propto r(x_{n+1}|x_n) p(y_{n+1}|x_{n+1}) \sum_{u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1}) \bar{d}(x_{n+1}, u_{n+1}),$$

et

$$\bar{d}(x_{n+1}, u_{n+1} | y_{1:N}) \propto \beta_{n+1}(x_{n+1}, u_{n+1}) \bar{d}(x_{n+1}, u_{n+1}).$$

On en déduit que la distribution $p(x_{1:N} | y_{1:N})$ est également celle d'une chaîne semi-markovienne.

Estimation des paramètres du modèle

Dans ce paragraphe, nous abordons l'estimation des paramètres de $p(x_{1:N}, u_{1:N})$, l'estimation de la loi conditionnelle $p(y_{1:N}|x_{1:N})$ ayant été décrite dans la section 2.3 du chapitre 2. Dans les expérimentations, nous utiliserons l'algorithme ICE. La chaîne de Markov (X, U) sera stationnaire. La mesure invariante π de (X, U) sera obtenue en résolvant $\pi Q_{X,U} = \pi$, $Q_{X,U}$ désignant la transition de (X, U) . Les estimations de $\bar{d}(\omega_j, \cdot)$ et de $r(\omega_j|\omega_i)$ seront obtenues à partir de celle de $p(x_n, u_n = 1, x_{n+1}, u_{n+1})$. L'estimateur de $p(x_n, u_n = 1, x_{n+1}, u_{n+1})$ à partir des données complètes $(x_{1:N}, u_{1:N}, y_{1:N}) = (x_n, u_n, y_n)_{1 \leq n \leq N}$ est donné par :

$$\hat{p}(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u)(x_{1:N}, u_{1:N}, y_{1:N}) = \frac{1}{N-1} \sum_{n=1}^{N-1} I(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u).$$

Soit θ_q le vecteur paramètre obtenu à l'étape q de ICE, l'espérance sachant $y_{1:N}$ sous le paramètre θ_q est alors calculable et donne :

$$p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u; \theta_{q+1}) = \frac{1}{N-1} \sum_{n=1}^{N-1} p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u | y_{1:N}; \theta_q).$$

Les ré-estimées $\bar{d}^{q+1}(\omega_j, u)$ et $r^{q+1}(\omega_j|\omega_i)$ de $\bar{d}(\omega_j, u)$ et $r(\omega_j|\omega_i)$ sont données par :

$$\bar{d}^{q+1}(\omega_j, u) = \frac{\sum_{\omega_i} p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u; \theta_{q+1})}{\sum_{\omega_i, u} p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u; \theta_{q+1})},$$

$$r^{q+1}(\omega_j|\omega_i) = \frac{\sum_u p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u; \theta_{q+1})}{\sum_{\omega_j, u} p(x_n = \omega_i, u_n = 1, x_{n+1} = \omega_j, u_{n+1} = u; \theta_{q+1})}.$$

3.3.5 Expérimentations

Nous présentons dans cette sous-section quatre expériences. Dans la première expérience, les données sont issues du modèle de chaînes de Markov cachées et segmentées suivant les modèles de chaînes de Markov cachées et de chaînes semi-markoviennes cachées. Dans la seconde expérience, les données sont issues du modèle de chaînes semi-markoviennes cachées et segmentées, comme ci-dessus, par les mêmes deux méthodes. Le but de cette deuxième expérience est de savoir si négliger la semi-markovianité a des conséquences sur la qualité de la segmentation. Dans la troisième expérience, les données suivent le modèle semi-markovien M -stationnaire et nous les segmentons suivant le modèle de chaînes de Markov cachées M -stationnaires et le modèle de chaînes semi-markoviennes cachées M -stationnaires. Dans la dernière expérience, nous appliquons ces deux derniers modèles sur une image réelle. Dans toutes les expériences, la taille des processus est égale à $N = 256 \times 256$ et les processus monodimensionnels sont transformés en images bi-dimensionnelles grâce au parcours d'Hilbert-Peano figurant en annexe B. Le nombre d'itérations de l'algorithme ICE est égal à 100 et les

paramètres sont initialisés par la méthode du K-means. Les modèles présentés dans toute la sous-section seront notés :

- chaînes de Markov cachées à bruit indépendant : CMC ;
- chaînes de Markov cachées M -stationnaires : CMC-M-S ;
- chaînes semi-markoviennes cachées à bruit indépendant : CSMC ;
- chaînes semi-markoviennes cachées M -stationnaires : CSMC-M-S.

Segmentation de données issues d'un modèle de chaîne de Markov cachées à bruit indépendant

Dans ce paragraphe, les données sont issues d'un modèle de chaîne de Markov cachées à bruit indépendant. Les états cachés sont ensuite estimés en utilisant les modèles CMC et CSMC. La matrice de transition de la chaîne de Markov X est $Q = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$ et $p(y_n|x_n)$ suit une loi normale $\mathcal{N}_{\mathbb{R}}(0, 10)$ lorsque $x_n = \omega_1$ et $\mathcal{N}_{\mathbb{R}}(1, 10)$ lorsque $x_n = \omega_2$.

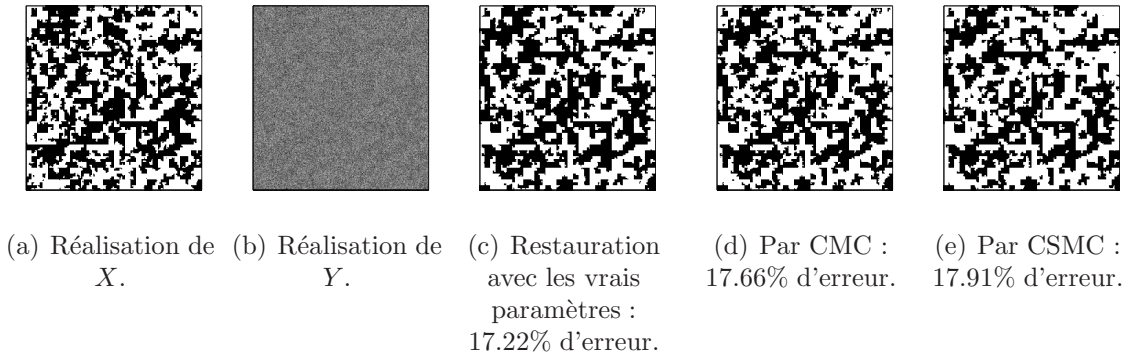


FIG. 3.1 – Simulation d'une chaîne de Markov cachée et sa restauration.

Modèle	Moyennes		Variances		Taux d'erreur
	ω_1	ω_2	ω_1	ω_2	
CMC	-0.04	1.00	9.99	9.98	17.66%
CSMC	-0.05	0.96	9.82	10.16	17.91%
Vraies valeurs	0	1	10	10	17.22%

TAB. 3.1 – Estimation des paramètres de la loi d'observation.

Concernant l'estimation de Q , le modèle CMC donne $\hat{Q}_{\text{CMC}} = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$. L'estimation de la transition $r(\omega_j|\omega_i)$ par CSMC avec $L = 10$ est donnée par la matrice $\begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix}$ et celle de \bar{d} est donnée par :

$$\begin{aligned} \hat{d}(\omega_1, \cdot) &= (0.82, 0.01, 0.0, 0.04, 0.01, 0.04, 0.03, 0.01, 0.04, 0), \\ \hat{d}(\omega_2, \cdot) &= (0.85, 0, 0.01, 0.01, 0.01, 0.02, 0.05, 0.02, 0.02, 0.01). \end{aligned}$$

Des résultats de l'estimation, nous voyons que le modèle semi-markovien caché est capable de retrouver la markovianité des données. En effet la probabilité $\hat{d}(\omega_j, u)$ est élevée pour $u = 1$ et faible pour les autres états. Par conséquent, les estimations des paramètres de la loi d'observation par le modèle CMC et par le modèle CSMC donnent des résultats comparables. Ainsi l'utilisation des CSMC à la place des CMC ne dégrade pas, dans le cadre de notre étude, la qualité des résultats en non supervisé. Notons également le degré élevé du bruitage et le très bon comportement de ICE.

Segmentation de données issues d'un modèle de chaînes semi-markoviennes cachées

Considérons une CSMC telle que chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \omega_2\}$, chaque U_n prend ses valeurs dans $\Lambda = \{1, \dots, 10\}$, la transition $r(x_{n+1}|x_n)$ est donnée par la matrice $R = \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix}$ et pour tout x_{n+1} , $\bar{d}(x_{n+1}, 10) = \frac{3}{4}$ et $\bar{d}(x_{n+1}, u_{n+1}) = \frac{1}{36}$ si $u_{n+1} \neq 10$. Les lois d'observation $p(y_n|x_n)$ sont les lois normales $\mathcal{N}_{\mathbb{R}}(0, 2)$ lorsque $x_n = \omega_1$ et $\mathcal{N}_{\mathbb{R}}(1, 2)$ lorsque $x_n = \omega_2$.

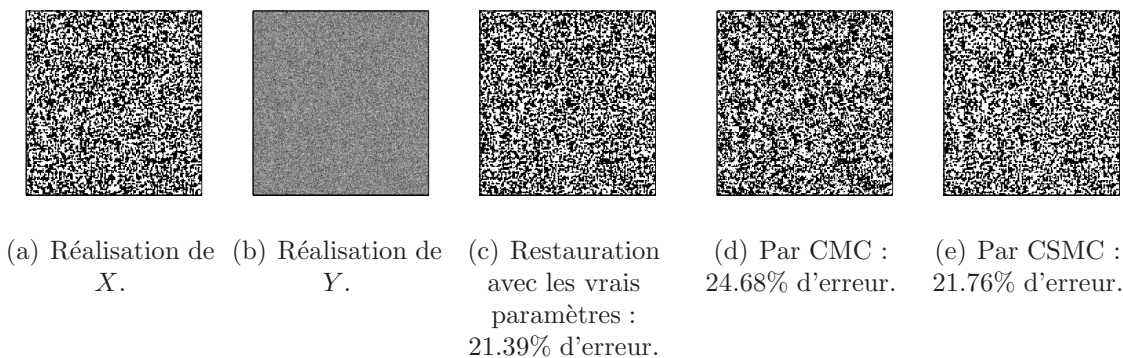


FIG. 3.2 – Simulation d'une chaîne semi-markovienne cachée et sa restauration.

Des résultats de la segmentation présentés dans la figure 3.2, nous constatons que, dans le cadre de notre étude, les CMC sont relativement robustes par rapport à la semi-markovianité des données cachées ; cependant, la différence entre les résultats n'est pas tout à fait négligeable.

Modèle	Moyennes		Variances		Taux d'erreur
	ω_1	ω_2	ω_1	ω_2	
CMC	-0.04	1.06	1.97	1.95	24.68%
CSMC	-0.03	0.96	2.02	2.00	21.76%
Vraies valeurs	0	1	2	2	21.39%

TAB. 3.2 – Estimation des paramètres de la loi d'observation.

L'estimation de R par le modèle semi-markovien caché est donnée par :

$$\hat{R}_{\text{CSMC-BI}} = \begin{pmatrix} 0.29 & 0.71 \\ 0.71 & 0.29 \end{pmatrix}, \quad (3.24)$$

et celle de $\bar{d}(x_{n+1}, u_{n+1})$ est donnée par :

$$\begin{aligned}\hat{d}(\omega_1, \cdot) &= (0.00, 0.04, 0.00, 0.05, 0.00, 0.04, 0.08, 0.00, 0.00, 0.79), \\ \hat{d}(\omega_2, \cdot) &= (0.01, 0.04, 0.00, 0.08, 0.00, 0.01, 0.03, 0.00, 0.02, 0.81).\end{aligned}$$

Nous constatons un bon comportement de l'ICE, en particulier dans l'estimation des paramètres du bruit.

Segmentation de données issues du modèle semi-markovien caché M -stationnaire

Dans le modèle étudié lors de cette expérience, nous introduisons deux processus auxiliaires U^1 et U^2 . Le processus U^1 modélise la semi-markovianité et le second processus U^2 modélise la M -stationnarité. Nous supposons que chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$, chaque U_n^1 prend ses valeurs dans $\Lambda_1 = \{1, \dots, L\}$, et chaque U_n^2 prend ses valeurs dans $\Lambda_2 = \{\lambda_1, \dots, \lambda_M\}$. Le modèle semi-markovien M -stationnaire que nous étudions est donné par :

$$\begin{aligned}p(x_{1:N}, u_{1:N}^1, u_{1:N}^2, y_{1:N}) &= p(u_1^2)p(x_1|u_1^2)p(u_1^1|x_1, u_1^2)p(y_1|x_1) \\ &\times \prod_{n=1}^{N-1} p(u_{n+1}^2|u_n^1, u_n^2)p(x_{n+1}|x_n, u_n^1, u_n^2)p(u_{n+1}^1|x_{n+1}, u_n^1, u_n^2) \\ &\times \prod_{n=1}^{N-1} p(y_{n+1}|x_{n+1}),\end{aligned}\tag{3.25}$$

où

$$p(u_{n+1}^2|u_n^1, u_n^2) = \begin{cases} \delta_{u_n^2}(u_{n+1}^2) & \text{si } u_n^1 > 1, \\ p(u_{n+1}^2|u_n^2) & \text{si } u_n^1 = 1, \end{cases}\tag{3.26}$$

$$p(x_{n+1}|x_n, u_n^1, u_n^2) = \begin{cases} \delta_{x_n}(x_{n+1}) & \text{si } u_n^1 > 1, \\ p(x_{n+1}|x_n, u_n^2) & \text{si } u_n^1 = 1, \end{cases}\tag{3.27}$$

et

$$p(u_{n+1}^1|x_{n+1}, u_n^1, u_n^2) = \begin{cases} \delta_{u_n^1-1}(u_{n+1}^1) & \text{si } u_n^1 > 1, \\ p(u_{n+1}^1|x_{n+1}, u_n^2) & \text{si } u_n^1 = 1. \end{cases}\tag{3.28}$$

Le but de cette expérience est de montrer l'importance simultanée de la M -stationnarité et de la semi-markovianité lorsque les données simulées sont issues d'un modèle semi-markovien caché M -stationnaire. Les données sont simulées selon le modèle de chaînes semi-markoviennes cachées M -stationnaires avec $K = 2$, $M = 2$ et $L = 10$ et :

– $p(u_{n+1}^2|u_n^2)$ a pour matrice de transition :

$$S = \begin{pmatrix} 0.9999 & 0.0001 \\ 0.0001 & 0.9999 \end{pmatrix};$$

– pour tout u_{n+1}^2 , $p(x_{n+1}|x_n, u_n^1 = 1, u_n^2)$ a pour matrice de transition :

$$Q = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix};$$

- si $u_{n+1}^2 = \lambda_1$, alors pour tout x_{n+1} , $p(u_{n+1}^1 | x_{n+1}, u_{n+1}^2, u_n^1 = 1) = \delta_1(u_{n+1}^1)$;
- si $u_{n+1}^2 = \lambda_2$, alors pour tout x_{n+1} , $p(u_{n+1}^1 | x_{n+1}, u_{n+1}^2, u_n^1 = 1)$ est la distribution d'une loi uniforme sur Λ_1 ;
- $p(y_n | x_n)$ est la densité d'une loi normale $\mathcal{N}_{\mathbb{R}}(0, 5)$ si $x_n = \omega_1$ et $\mathcal{N}_{\mathbb{R}}(1, 5)$ si $x_n = \omega_2$.

Nous choisissons ensuite d'estimer les paramètres par les méthodes ICE correspondantes et les états cachés par le MPM en utilisant les modèles CMC, CMC-M-S, CSMC et CSMC-M-S.

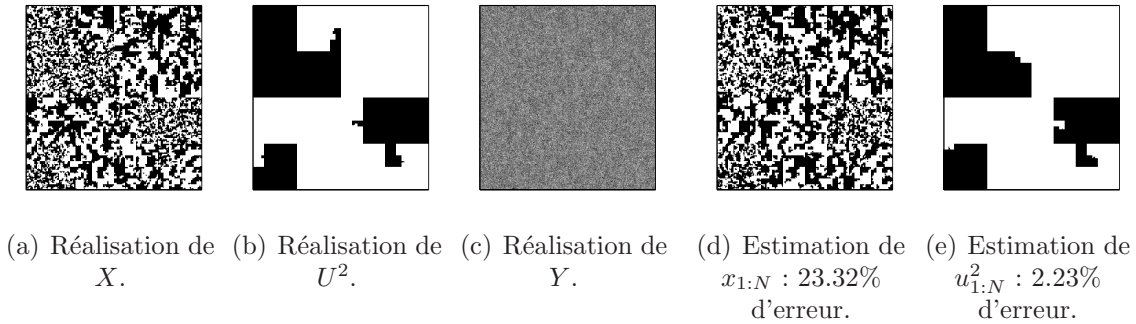


FIG. 3.3 – Simulation d'une chaîne semi-markovienne cachée M -stationnaire et sa restauration utilisant les vrais paramètres.

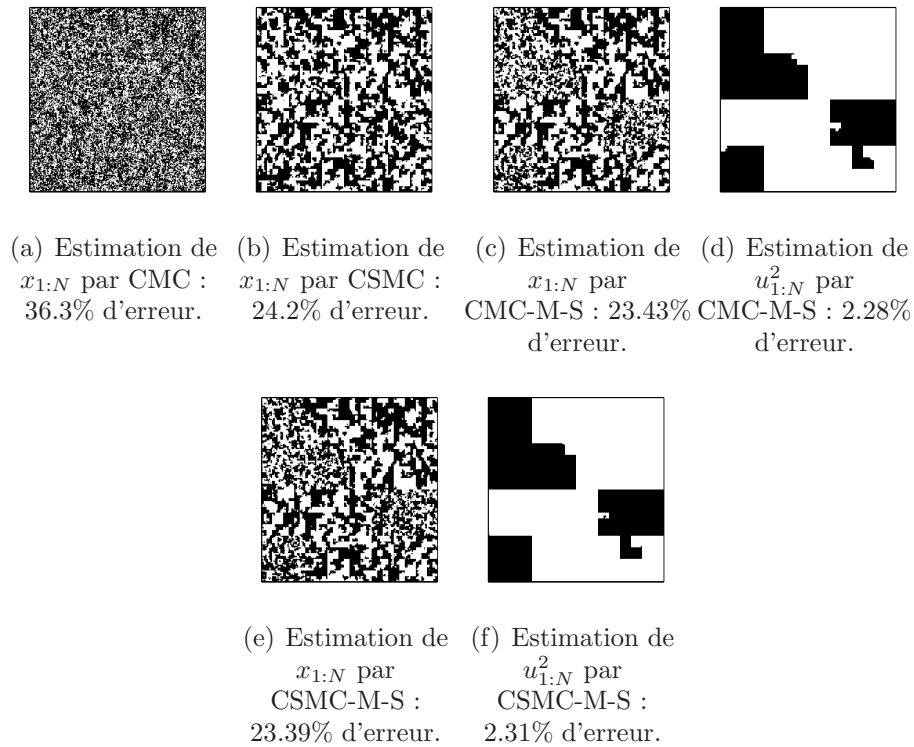


FIG. 3.4 – Restauration en utilisant les modèles CMC, CMC-M-S, CSMC et CSMC-M-S.

Modèle	Moyenne		Variance		Taux d'erreur pour X	Taux d'erreur pour U_2
	ω_1	ω_2	ω_1	ω_2		
CMC	-0.30	1.46	4.49	4.12	36.3%	-
CMC-M-S	0.00	1.00	4.92	5.03	23.43%	2.28%
CSMC	0.02	0.99	4.87	5.12	24.2%	-
CSMC-M-S	0.02	1.01	4.96	5.01	23.39%	2.31%
Vraies valeurs	0	1	5	5	-	-

TAB. 3.3 – Estimation des paramètres de la loi d'observation.

De cette expérience, nous constatons que négliger la semi-markovianité du processus caché est peu pénalisant lorsque l'on a tenu compte de la M -stationnarité. En effet les résultats obtenus en segmentant par le modèle CMC-M-S et par le modèle CSMC-M-S sont comparables. Cependant négliger l'hypothèse de M -stationnarité est plus pénalisant mais nous pouvons constater des figures 3.4,(a) et (b) que le modèle semi-markovien estime mieux les états cachés que celui des chaînes de Markov cachées.

Notons que si les taux d'erreur sont comparables dans les segmentations (b) et (c) à la figure 3.4, la (c), qui utilise le vrai modèle, est plus proche de la vraie image sur la figure 3.3. Quant à l'estimation des paramètres par ICE présentée dans le tableau 3.3, nous constatons des résultats similaires pour les 4 modèles utilisés.

Segmentation d'une image réelle

Dans cette dernière expérience, nous proposons de segmenter une image réelle SAR (Synthetic Aperture Radar). Nous comparons les modèles CMC-M-S et CSMC-M-S sur la photographie aérienne de Tokyo présentée à la Figure 3.5,(a). Dans cette expérience, on prendra $K = 3$, $M = 2$ et $L = 5$.

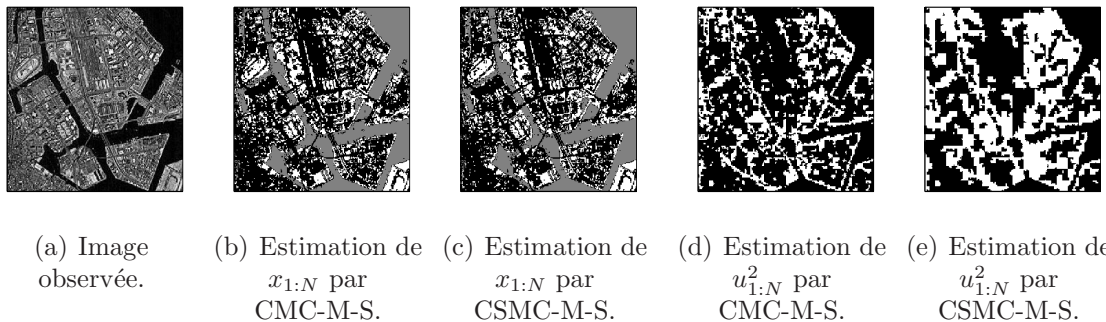


FIG. 3.5 – Segmentation d'une image réelle SAR.

De cette expérience, on voit des résultats similaires pour l'estimation de x . Cependant, l'estimation de u^2 semble être meilleure en utilisant le modèle semi-markovien caché M -stationnaire. Le modèle semi-markovien est capable de considérer des lois de temps de séjour plus générales que le modèle markovien, ce qui peut expliquer la qualité de l'estimation de

u^2 . Dans la thèse de P. Lanchantin [65], cette image a été segmentée par les modèles CMC et CMC-M-S et il y est constaté des améliorations notables apportées par le modèle CMC-M-S, par rapport au modèle CMC, au niveau des estimations des paramètres.

Conclusion

Nous avons utilisé des modèles de Markov triplets particuliers pour proposer dans ce chapitre un modèle semi-markovien caché original, permettant des gains notables de temps de calcul. Nous avons ensuite étendu le modèle des chaînes de Markov cachées M -stationnaires introduit par P. Lanchantin [65] à celui des chaînes semi-markoviennes cachées M -stationnaires. Pour cela, nous avons considéré deux processus auxiliaires ; le premier modélisant la semi-markovianité du processus caché, et le second sa non stationnarité. Le modèle proposé fait ainsi partie des chaînes triplets “multi-variées”, où la chaîne auxiliaire est composée de plus d’un processus, chacune des composantes modélisant une propriété particulière. Nous avons proposé une méthode ICE d’estimation des paramètres adéquate et les divers algorithmes de segmentation non supervisée, correspondants aux divers modèles, ont été testés via les simulations informatiques dans le contexte de bruits importants. Les résultats obtenus montrent que les deux chaînes auxiliaires sont nécessaires et l’absence d’une d’entre elles ne peut être compensée par l’autre. Ils ont également mis en évidence le très bon comportement de l’algorithme ICE, qui peut alors être utilisé dans des situations très bruitées.