

# Arbres de Markov couple

Wojciech Pieczynski

Institut National des Télécommunications, Département CITI, 9, rue Charles Fourier, 91000 Evry, France

Reçu le 14 novembre 2001 ; accepté après révision le 29 avril 2002

Note présentée par Paul Deheuvels.

---

## Résumé

Les modélisations par chaînes de Markov cachées (CMC), dont la possibilité de calcul explicite de la loi Markovienne du processus caché conditionnelle aux observations est le principal intérêt, trouvent de très nombreuses applications dans les domaines les plus divers. Ce modèle a été récemment généralisé aux chaînes de Markov « Couple », présentant les mêmes avantages que les CMC au niveau des traitements et proposant un pouvoir modélisant plus important. L'objet de cette Note est préciser comment les Arbres de Markov cachés, qui sont des extensions des CMC, peuvent également être généralisés aux modèles originaux appelés « Arbres de Markov couple ». *Pour citer cet article : W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 79–82.* © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

## Pairwise Markov trees

## Abstract

The Hidden Markov Chain (HMC) models are widely applied in various problems. This success is mainly due to the fact that the hidden model distribution conditional on observations remains a Markov chain distribution, and thus different processings, like Bayesian restorations, are handleable. These models have been recently generalized to “Pairwise” Markov chains, which admit the same processing power and a better modeling one. The aim of this Note is to show that the Hidden Markov trees, which can be seen as extensions of the HMC models, can also be generalized to “Pairwise” Markov trees, which present the same processing advantages and better modelling power. *To cite this article: W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 79–82.* © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

---

## 1. Introduction

Dans les modèles de Markov cachés on considère un processus stochastique de Markov inobservable  $X$  et un processus observé  $Y$ , dont on se donne les lois conditionnelles à  $X$ . Le problème est de « restaurer »  $Y$ , à savoir estimer la réalisation de  $X$  à partir de  $Y$ , ce dernier pouvant être perçu comme étant une version « bruitée » de  $X$ . On se heurte alors à une limitation suivante : il est possible d'appliquer diverses méthodes générales d'estimation, dont les méthodes Bayésiennes, lorsque la loi de  $X$  conditionnelle à  $Y$ , dite « a posteriori », reste Markovienne. Cette dernière propriété a lieu lorsque les lois de  $Y$  conditionnelles à  $X$  ne sont pas trop compliquées, ce qui limite le pouvoir modélisant des phénomènes réels par ce type de modèle. En particulier, ce phénomène est visible lorsque l'on utilise les champs de Markov cachés en traitements d'images : garder la Markovianité de  $X$  a posteriori rend difficile une modélisation correcte des textures et des approximations du modèle doivent être effectuées [3]. Une solution à ce problème a été récemment proposée en considérant les modèles plus généraux, dits champs de Markov couple, dans lesquels on considère la Markovianité du couple  $Z = (X, Y)$  [6]. Cette dernière assure la Markovianité de  $X$  a posteriori, ce qui permet les différents traitements, et la Markovianité de  $Y$  conditionnelle à  $X$ , ce qui

---

Adresse e-mail : Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

permet de modéliser correctement les textures. Cependant, il ne s’agit plus nécessairement d’un champs de Markov caché car  $X$  n’est plus nécessairement un champ de Markov. Un modèle analogue dit chaînes de Markov Couple (CM Couple) généralisant les chaînes de Markov cachées [1] a ensuite été proposé [7] et il a été démontré qu’une telle généralisation augmente strictement le pouvoir modélisant de ces dernières sans perdre leurs pouvoir de traitements [8]. L’objet de cette Note est de montrer que les modèles par Arbres de Markov cachés (AM Cachés [2,4,5]), qui peuvent être vues comme des extensions des CM Cachées, sont également généralisables aux arbres de Markov «Couple» (AM Couple). En particulier, nous montrons que la généralisation est stricte dans le sens où le processus caché n’est plus nécessairement de Markov (une CNS pour qu’il en soit ainsi est précisée).

Comme perspectives, nous pouvons mentionner l’extension des résultats présentés à des réseaux stochastiques plus généraux [9].

**2. Chaînes de Markov couples**

Soit  $X = (X_1, \dots, X_n)$  et  $Y = (Y_1, \dots, Y_n)$  processus stochastiques, où chaque  $Z_i = (X_i, Y_i)$  prend ses valeurs dans  $\Omega \times \mathbb{R}$ , avec  $\Omega = \{\omega_1, \dots, \omega_k\}$  et  $\mathbb{R}$  l’ensemble des nombres réels. Le processus  $Z = (Z_1, \dots, Z_n)$  est appelé le processus «Couple» associé à  $X$  et  $Y$ . Afin de simplifier, nous noterons les différentes lois avec la même lettre  $p$ . Ainsi  $p(x), p(x_i), p(x_i, x_{i+1}), p(z_i) = p(x_i, y_i)$  sont respectivement des densités des lois de  $X, X_i, (X_i, X_{i+1}),$  et  $Z_i = (X_i, Y_i)$ . De même,  $p(x_{i+1}|x_i)$  désignera la loi de  $X_{i+1}$  conditionnelle à  $X_i = x_i, p(y|x)$  sera la densité de la loi de  $Y$  conditionnelle à  $X = x$ , etc. Par ailleurs, afin de réduire l’écriture, on écrira dans la suite  $z$  à la place de  $(x, y)$  à chaque fois que cela sera possible, même si les deux écritures doivent être présentes dans une même équation.

Le processus  $Z$  est dit une chaîne de Markov Couple (CM Couple) stationnaire si sa loi est donnée par  $p(z) = p(z_1)p(z_2|z_1) \dots p(z_n|z_{n-1})$ , et si les  $p(z_i, z_{i+1})$  sont indépendantes de  $i$ . La loi de  $Z$  stationnaire est ainsi déterminée par une loi de probabilité  $p(z_i, z_{i+1})$  sur  $(\Omega \times \mathbb{R})^2$ . On vérifie les propriétés suivantes [8] : (1)  $p(y|x)$  et  $p(x|y)$  sont des chaînes de Markov ; (2) en prenant  $p(z_i, z_{i+1}) = p(x_i, x_{i+1})p(y_i|x_i)p(y_{i+1}|x_{i+1})$ , on retrouve une CM Cachée classique, avec les  $(Y_i)$  indépendantes conditionnellement à  $X$  ; (3)  $p(y_i|x_i, x_{i+1}) = p(y_i|x_i)$  implique que  $X$  est une chaîne de Markov, la réciproque étant vraie dès que  $x_1 = x_3$  implique  $p(y_2|x_1, x_2) = p(y_2|x_2, x_3)$ . Par ailleurs, on vérifie que les probabilités «forward»  $\alpha(x_i) = p(y_1, \dots, y_{i-1}, z_i)$  et «backward»  $\beta(x_i) = p(y_{i+1}, \dots, y_n|z_i)$  sont calculables de proche en proche par formules (2.1), (2.2), et que les transitions  $p^y(x_{i+1}|x_i)$  et les marginales  $p^y(x_i)$  (qui permettent le calcul de certaines méthodes Bayésienne de restauration) de la loi de  $X$  conditionnelle à  $Y = y$  sont données par les formules (2.3).

$$\alpha(x_1) = p(z_1), \quad \text{et} \quad \alpha(x_{i+1}) = \sum_{x_i \in \Omega} \alpha(x_i) p(z_{i+1}|z_i) \quad \text{pour } 2 \leq i \leq n, \tag{2.1}$$

$$\beta(x_n) = 1, \quad \text{et} \quad \beta(x_i) = \sum_{x_{i+1} \in \Omega} \beta(x_{i+1}) p(z_{i+1}|z_i) \quad \text{pour } 1 \leq i \leq n - 1, \tag{2.2}$$

$$p^y(x_{i+1}|x_i) = \frac{p(z_{i+1}|z_i)\beta(x_{i+1})}{\beta(x_i)}, \quad p^y(x_i) = \frac{\alpha(x_i)\beta(x_i)}{\sum_{\omega \in \Omega} \alpha(\omega)\beta(\omega)}. \tag{2.3}$$

Lorsque  $p(z_1) = p(x_1)p(y_1|x_1)$  et  $p(z_i|z_{i-1}) = p(x_i|x_{i-1})p(y_i|x_i)$  (CM Cachée classique), les formules (2.1)–(2.3) donnent les formules classiques.

**3. Arbres de Markov couples**

Soit  $S$  un ensemble fini d’indices,  $X = (X_s)_{s \in S}, Y = (Y_s)_{s \in S}$ , deux processus, et  $Z = (Z_s)_{s \in S}$  processus donné par  $Z_s = (X_s, Y_s)$ . Comme précédemment, chaque  $(X_s, Y_s)$  prend ses valeurs  $\Omega \times \mathbb{R}$ . Soit  $S^1, \dots, S^n$  une partition de  $S$ . Posons  $N = \text{Card}(S)$  et, pour  $1 \leq i \leq n, N^i = \text{Card}(S^i)$ . Supposons que  $N^1 = 1 \leq N^2 \leq \dots \leq N^n$  et pour tout  $1 \leq i \leq n - 1$ , associons à chaque  $s \in S^i$  un sous-ensemble  $s^+$  de  $S^{i+1}$ , appelé «enfants de  $s$ », de manière à ce que  $(s^+)_{s \in S^i}$  soit une partition de  $S^{i+1}$  (l’unique élément de  $N^1$  est appelé

« racine »). Par ailleurs,  $s^{++}$  désignera l'ensemble des descendants de  $s$  et pour tout  $s \in S - S^1$ , son unique « père » sera noté  $s^-$ .

DÉFINITION 3.1. – Le processus  $Z$  est appelé Arbre de Markov Couple (AM Couple) si

$$p(z) = p(z_1) \prod_{i=2}^n \prod_{s \in S^i} p(z_s | z_{s^-}). \quad (3.1)$$

Il est appelé Arbre de Markov Caché (AM Caché) si sa loi marginale par rapport à  $X$ , dite aussi loi « a priori » de  $X$ , est de forme (3.1).

Notons que lorsque  $p(z_s | z_{s^-}) = p(x_s | x_{s^-}) p(y_s | x_s)$ , on retrouve les AM Cachés utilisés dans [4].

Nous allons montrer deux résultats suivants : (i) AM Couple jouit des mêmes propriétés intéressantes que AM Caché : la loi de  $X$  a posteriori est une loi d'un AM dont on peut calculer les transitions et les marginales ; (ii) les AM Couples généralisent strictement les AM Cachés dans le sens qu'il existe des AM Couples  $Z = (X, Y)$  tels que  $X$  n'est pas un AM (une CNS pour qu'un AM Couple soit un AM Caché, portant uniquement sur les transitions, est précisée). Notons que le premier résultat n'est pas entièrement original car le AM Couple considéré peut être vu comme un cas particulier des modèles partiellement observés à structure triangulée [9] ; cependant, son établissement direct dans le cadre considéré clarifiera, de manière utile pour un praticien, les récurrences nécessaires aux calculs des diverses quantités d'intérêt.

PROPOSITION 3.1. – Soit  $Z$  un AM Couple défini par (3.1). Alors la loi  $p^y$  de  $X$  conditionnelle à  $Y = y$  admet l'écriture (3.1) avec des  $x$  à la place des  $z$ . De plus, l'égalité (3.2) donne les transitions  $p^y(x_s | x_{s^-})$  en fonction des probabilités « backward »  $\beta(x_s) = p(y_{s^{++}} | z_s)$ , qui peuvent être calculées récursivement par (3.3).

$$p^y(x_s | x_{s^-}) = \frac{\beta(x_s) p(z_s | z_{s^-})}{\sum_{\omega_s \in \Omega} \beta(\omega_s) p(\omega_s, y_s | z_{s^-})}, \quad (3.2)$$

$$\beta(x_s) = 1 \text{ pour } s \in S^n, \text{ et } \beta(x_s) = \prod_{t \in S^+} \left( \sum_{x_t \in \Omega} \beta(x_t) p(z_t | z_s) \right) \text{ pour } s \in S - S^n. \quad (3.3)$$

Démonstration. – Nous avons  $p(x_s | x_{s^-}, y) = p(x_s | z_{s^-}, y_s, y_{s^{++}}, y_{S-s^{++}-\{s, s^-\}}) \propto p(z_s, y_{s^{++}} | z_{s^-}, y_{S-s^{++}-\{s, s^-\}}) = p(z_s, y_{s^{++}} | z_{s^-}) = \beta(x_s) p(z_s | z_{s^-})$ , d'où (3.2) (la deuxième égalité est due à l'indépendance des  $(Y_t)_{t \in S^{++}}$  et  $(Y_t)_{t \in S-s^{++}-\{s, s^-\}}$  conditionnelle à  $Z_{s^-}$ , qui vient de l'indépendance des  $(Z_t)_{t \in S^{++}}$  et  $(Z_t)_{t \in S-s^{++}-\{s, s^-\}}$  conditionnelle à  $Z_{s^-}$ , propriété classique des arbres de Markov).

Par ailleurs,

$$\begin{aligned} \beta(x_s) &= p(y_{s^{++}} | z_s) = \sum_{x_t \in \Omega} \prod_{t \in S^+} p(x_t, y_{s^{++}} | z_s) = \sum_{x_t \in \Omega} \prod_{t \in S^+} p(x_t, y_t, y_{t^{++}} | z_s) = \sum_{x_t \in \Omega} \prod_{t \in S^+} p(z_t, y_{t^{++}} | z_s) \\ &= \sum_{x_t \in \Omega} \prod_{t \in S^+} p(y_{t^{++}} | z_t, z_s) p(z_t | z_s) = \sum_{x_t \in \Omega} \prod_{t \in S^+} p(y_{t^{++}} | z_t) p(z_t | z_s) = \prod_{t \in S^+} \left( \sum_{x_t \in \Omega} \beta(x_t) p(z_t | z_s) \right), \end{aligned}$$

d'où (3.3).

Toutes les transitions  $p^y(x_s | x_{s^-})$  ayant été calculées dans une « passe ascendante », on peut en déduire les marginales a posteriori. En effet, pour  $s \in S - S^1$  quelconque il existe un chemin unique  $c = (s_1, \dots, s_{i-1}, s)$  le reliant au sommet  $s_1$ , et la restriction de  $X$  à ce chemin est, pour sa loi conditionnelle à  $Y = y$ , une chaîne de Markov.  $p^y(x_s)$  se calcule alors classiquement de proche en proche par  $p^y(x_{s_j}) = \sum_{x_{s_{j-1}} \in \Omega} p^y(x_{s_{j-1}}) p^y(x_{s_j} | x_{s_{j-1}})$ , avec  $p^y(x_{s_1})$  calculée dans la passe ascendante.  $\square$

En résumé, toutes les transitions « père-fils » a posteriori sont calculées dans une passe ascendante, en utilisant (3.2) et (3.3). Ensuite, les calculs des lois marginales a posteriori se font dans une passe descendante, en utilisant la chaîne de Markov définie sur les sites allant du sommet au site considéré.

PROPOSITION 3.2. – Soit  $Z$  un Arbre de Markov Couple vérifiant (3.1). Supposons que chaque  $s \in S - S^n$  admet au moins deux enfants et que les lois des variables indicées par les enfants, conditionnelles

au père, sont égales. Alors  $Z$  est un Arbre de Markov Caché avec  $X$  vérifiant (3.1) si et seulement si on a la propriété (P) suivante :

(P) pour tout  $s \in S - S^1$ ,  $x_{s^-}, x_s \in \Omega$ , et  $y_{s^-} \in R$ , on a  $p(x_s | x_{s^-}, y_{s^-}) = p(x_s | x_{s^-})$ .

Démonstration. – (P) est une CS. Nous avons  $p(x, y_{S-S^n}) = p(z^1) \prod_{s \notin S^n} p(z_s | z_{s^-}) \prod_{s \in S^n} \int_R p(x_s, y_s | z_{s^-}) dy_s = p(z^1) \prod_{s \notin S^n} p(z_s | z_{s^-}) \prod_{s \in S^n} p(x_s | x_{s^-}, y_{s^-}) = p(z^1) \prod_{s \notin S^n} p(z_s | z_{s^-}) \prod_{s \in S^n} p(x_s | x_{s^-})$ , la dernière égalité ayant lieu en vertu de (P). En procédant de proche en proche, on obtient  $p(x) = p(x^1) \prod_{s \in S-S^1} p(x_s | x_{s^-})$ .

(P) est une CN. Soit  $s \in S - S^n$ , et  $t, u$  deux fils de  $s$ .  $X$  et  $Z$  étant des Arbres de Markov, les lois de  $(X_s, X_t, X_u)$  et  $(Z_s, Z_t, Z_u)$  s'écrivent  $p(x_s, x_t, x_u) = p(x_s, x_t)p(x_s, x_u)/p(x_s)$  et  $p(z_s, z_t, z_u) = p(z_s, z_t)p(z_s, z_u)/p(z_s)$ . Par ailleurs  $p(z_u, z_s)p(z_s, z_t)/p(z_s) = [p(x_u, x_s)p(x_s, x_t)/p(x_s)][p(y_t, y_s | x_t, x_s)p(y_s, y_u | x_s, x_u)/p(y_s | x_s)]$ . Il en résulte que  $p(y_t, y_s | x_t, x_s)p(y_s, y_u | x_s, x_u)/p(y_s | x_s)$  est la loi de  $(Y_s, Y_t, Y_u)$  conditionnelle à  $(X_s, X_t, X_u)$ , ce qui implique que son intégrale par rapport à  $(y_s, y_t, y_u)$  vaut 1. En l'intégrant par rapport à  $(y_u, y_t)$  il vient

$$\int_{\mathbb{R}} \frac{p(y_s | x_t, x_s)p(y_s | x_s, x_u)}{p(y_s | x_s)} dy_s = 1. \tag{3.4}$$

À  $x_s$  fixé, considérons  $g(y_s) = 1/p(y_s | x_s)$  et le produit scalaire défini par  $\langle f, h \rangle = \int_{\mathbb{R}} f(y)h(y)g(y) dy$  (notons que c'est bien un produit scalaire car, si  $p(y_s | x_s) = \sum_{\omega \in \Omega} p(x_t = \omega)p(y_s | x_t = \omega, x_s) = 0$ , alors tous les  $p(y_s | x_t = \omega, x_s)$  sont nuls). Posons  $f_i(y_s) = p(y_s | x_t = \omega_i, x_s) = p(y_s | x_s, x_u = \omega_i)$ . (3.4) étant vrai pour tous  $x_t, x_u$  dans  $\Omega = \{\omega_1, \dots, \omega_k\}$ , nous avons une famille de vecteurs  $f_1, \dots, f_k$  vérifiant  $\langle f_i, f_j \rangle = 1$  pour tous  $1 \leq i, j \leq k$ . Donc pour tous  $1 \leq i, j \leq k$ ,  $\|f_i - f_j\|^2 = \|f_i\|^2 + \|f_j\|^2 - 2\langle f_i, f_j \rangle = 0$ , d'où l'égalité des vecteurs  $f_1, \dots, f_k$ . L'indépendance des  $f_i$  de  $i$  équivaut à l'indépendance des  $p(y_s | x_t = \omega_i, x_s) = p(y_s | x_s, x_u = \omega_i)$  de  $i$ , ce qui achève la démonstration.  $\square$

Notons que dans le cadre de l'utilisation des arbres cachés en traitement du signal ou d'images les sites ont souvent plusieurs enfants avec les lois conditionnelles au père identiques ; les hypothèses de la Proposition 3.2 n'apportent alors pas de contrainte supplémentaire par rapport à l'utilisation des modèles classiques. Par ailleurs, remarquons que le modèle par AM Caché utilisé par Laferté, Pérez et Heitz [4] consiste à prendre  $p(z^1) = p(x_1)f_{x_1}(y_1)$ ,  $p(z_s | z_{s^-}) = p(x_s | x_{s^-})p(y_s | x_s)$  : on constate que (P) est bien vérifiée. Notons également que l'on peut prendre des conditions moins restrictives tout en restant dans le cadre des AM Cachés. En effet, si l'on remplace  $p(y_s | x_s)$  par  $p(y_s | x_s, y_{s^-})$ , une densité qui dépend également de  $y_{s^-}$ , la propriété (P) reste vérifiée et donc le modèle demeure un AM Caché. Cela signifie, grosso modo, que l'hypothèse de « bruit blanc » habituellement retenue n'est pas nécessaire pour obtenir un AM Caché et des bruits « colorés » peuvent être considérés dans ce même cadre.

**Remerciements.** Je remercie Françoise Desbouvies (INT, Evry) et Patrick Pérez (Microsoft, Cambridge) dont les suggestions ont permis d'améliorer sensiblement la présentation de cette Note.

### Références bibliographiques

[1] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statist.* 41 (1970) 164–171.  
 [2] K. Daoudi, A. Frakt, A. Willsky, Multiscale autoregressive models and wavelets, *IEEE Trans. Inform. Theory* 45 (3) (1999) 828–845.  
 [3] H. Derin, H. Elliot, Modelling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Trans. PAMI* 9 (1) (1987) 39–55.  
 [4] J.-M. Laferté, P. Pérez, F. Heitz, Discrete Markov image modeling and inference on the quadtree, *IEEE Trans. Image Processing* 9 (3) (2000) 390–404.  
 [5] M. Luetgen, W. Karl, A. Willsky, Efficient multiscale regularization with applications to the computation of optical flow, *IEEE Trans. SP* 41 (12) (1993) 3377–3396.  
 [6] W. Pieczynski, A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Machine Graphics & Vision* 9 (3) (2000) 705–718.  
 [7] W. Pieczynski, Pairwise Markov chains and Bayesian unsupervised fusion, in: *Proceedings of 3rd ICIF, Vol. 1, FUSION 2000, July 10–13, Paris, France, 2000*, pp. MoD4-24–MoD4-31.  
 [8] W. Pieczynski, Pairwise Markov Chains, *IEEE Trans. PAMI*, submitted.  
 [9] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, Wiley Ser. Probab. Math. Statist., 1996.