

# APPLICATION DU FILTRAGE PARTICULAIRE À L'INFÉRENCE STATISTIQUE DES HMM

Natacha Caylus & Arnaud Guyader & François Le Gland & Nadia Oudjane

*IRISA / Université de Rennes 1, Campus de Beaulieu, 35042 Rennes*

*Université de Rennes 2, Campus de Villejean, 35043 Rennes*

*IRISA / INRIA, Campus de Beaulieu, 35042 Rennes*

*EDF, Division R&D, 92141 Clamart*

## Résumé

On considère un modèle de Markov caché tel que les transitions de la chaîne sous-jacente dépendent d'un paramètre inconnu que l'on cherche à estimer ou à surveiller. Il n'existe pas en général d'expression analytique simple pour le noyau de transition, mais s'il est facile de simuler des trajectoires ou des transitions de la chaîne, on peut alors utiliser une méthode de Monte Carlo avec interaction, ou méthode particulière, pour calculer de manière approchée certaines fonctions de contraste et les fonctions score associées.

**Mots-clés :** modèle de Markov caché, filtrage stochastique, filtrage particulier, fonction de vraisemblance.

## Abstract

We consider an hidden Markov model (HMM) such that transitions of the underlying chain depend on an unknown parameter, which we try to estimate or to monitor. There is no simple analytical expression for the transition kernel in general, but if it is easy to simulate trajectories or transitions of the chain, then one can use a Monte Carlo method with interaction, or particle method, to compute approximately some contrast functions and the associated score functions.

**Key-words :** hidden Markov model, stochastic filtering, particle filtering, likelihood function.

## 1 Modèle de Markov caché, et fonctions de contraste

Soit  $\{X_n, n \geq 0\}$  une chaîne de Markov homogène à valeurs dans l'espace  $E$ , de noyau de transition  $Q^\theta(x, dx')$ , c'est-à-dire que

$$\mathbb{P}^\theta[X_{n+1} \in dx' \mid X_n = x] = Q^\theta(x, dx'),$$

où le noyau dépend d'un paramètre  $\theta$  qu'il s'agit d'estimer. On fait l'hypothèse qu'il est facile de simuler une v.a.  $X$  de loi  $Q^\theta(x, dx')$ , même si la forme analytique du noyau

n'est pas connue, ou est très compliquée. C'est par exemple le cas si la chaîne de Markov  $\{X_n, n \geq 0\}$  résulte de l'échantillonnage à des instants discrets  $\{t_n, n \geq 0\}$  d'un processus de diffusion  $\{X'_t, t \geq 0\}$ , c'est-à-dire si  $X_n = X'_{t_n}$ , avec

$$dX'_t = b^\theta(X'_t) dt + \sigma(X'_t) dW'_t ,$$

où  $\{W'_t, t \geq 0\}$  est un mouvement brownien. Dans ce cas, simuler une v.a. de loi  $Q^\theta(x, dx')$  revient simplement à simuler la solution à l'instant  $t_{n+1}$  de l'équation différentielle stochastique partant de la condition initiale  $X'_{t_n} = x$  à l'instant  $t_n$ . L'état  $\{X_n, n \geq 0\}$  n'est pas observé, mais on dispose d'une suite  $\{Y_n, n \geq 0\}$  d'observations à valeurs dans l'espace  $F$ , telle que conditionnellement aux états cachés, les observations sont mutuellement indépendantes, et la loi conditionnelle de  $Y_n$  ne dépend que de l'état caché  $X_n$  au même instant. On pose

$$\mathbb{P}^\theta[Y_n \in dy \mid X_n = x] = g^\theta(x, y) \lambda(dy) \quad \text{et} \quad \Psi_n^\theta(x) = g^\theta(x, Y_n) .$$

Cette hypothèse est par exemple vérifiée dans le cas où on observe l'état dans un bruit blanc additif, c'est-à-dire où l'observation  $Y_n$  est reliée à l'état caché  $X_n$  par la relation

$$Y_n = h^\theta(X_n) + V_n ,$$

où  $\{V_n, n \geq 0\}$  est un bruit blanc. Il s'agit d'estimer conjointement les états cachés et le paramètre, au vu des observations, et on introduit à cet effet

$$\mu_n^\theta(dx) = \mathbb{P}^\theta[X_n \in dx \mid Y_0, \dots, Y_n] \quad \text{et} \quad \mu_{n|n-1}^\theta(dx) = \mathbb{P}^\theta[X_n \in dx \mid Y_0, \dots, Y_{n-1}] .$$

La fonction de log-vraisemblance, dans ce modèle paramétré, peut alors s'écrire

$$\ell_n^\theta = \frac{1}{n} \sum_{k=0}^n \log \langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle ,$$

et la fonction score, c'est-à-dire sa dérivée par rapport au paramètre, s'écrit

$$\partial \ell_n^\theta = \frac{1}{n} \sum_{k=0}^n \left[ \frac{\langle w_{k|k-1}^\theta, \Psi_k^\theta \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle} + \frac{\langle \mu_{k|k-1}^\theta, S_k^\theta \Psi_k^\theta \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle} \right] ,$$

où pour tout  $k \geq 0$ , le filtre dérivé  $w_{k|k-1}^\theta$  dénote la dérivée du filtre optimal  $\mu_{k|k-1}^\theta$  par rapport au paramètre, et où  $S_k^\theta(x) = \partial \log \Psi_k^\theta(x)$ .

## 2 Filtre optimal, filtre dérivé

Dans la suite, pour alléger les notations, on ne fait plus apparaître la dépendance par rapport au paramètre. L'évolution de la suite  $\{\mu_n, n \geq 0\}$  à valeurs dans l'espace des distributions de probabilité sur l'espace  $E$ , est décrite par les étapes suivantes

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \mu_{k|k-1} = \mu_{k-1} Q \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1} = \bar{R}_k(\mu_{k-1}) ,$$

où la notation  $\cdot$  désigne le produit projectif, correspondant à la formule de Bayes, i.e.

$$\Psi_k \cdot \mu = \frac{\Psi_k \mu}{\langle \mu, \Psi_k \rangle}.$$

Pour le filtre dérivé, on fait l'hypothèse d'absolue continuité (AC) suivante [1] : le noyau dérivé  $\Gamma(x, dx')$ , c'est-à-dire la dérivée du noyau markovien  $Q(x, dx')$  par rapport au paramètre, admet une représentation probabiliste de la forme

$$\Gamma \phi(x) = \int_E \Gamma(x, dx') \phi(x') = \int_{E \times S} K(x, dx', ds') \phi(x') s' = \mathbb{E}[\phi(X_{n+1}) \Xi_{n+1} \mid X_n = x],$$

où  $\{(X_n, \Xi_n), n \geq 0\}$  est une chaîne de Markov à valeurs dans l'espace  $E \times S$ , telle que

$$\mathbb{P}[X_{n+1} \in dx', \Xi_{n+1} \in ds' \mid X_n = x, \Xi_n = s] = K(x, dx', ds').$$

On fait aussi l'hypothèse qu'il est *facile* de *simuler* une v.a.  $(X, \Xi)$  de loi  $K(x, dx', ds')$ , même si la forme analytique du noyau n'est pas connue, ou très compliquée. On vérifie alors que  $\mu \Gamma \ll \mu Q$ , et que si  $w \ll \mu$ , alors  $w Q \ll \mu Q$ . Soit

$$F_k(\mu) w = \frac{\Psi_k w}{\langle \mu, \Psi_k \rangle} - \frac{\langle w, \Psi_k \rangle}{\langle \mu, \Psi_k \rangle} \frac{\Psi_k \mu}{\langle \mu, \Psi_k \rangle}$$

la dérivée au point  $\mu$  et dans la direction  $w$ , de l'application  $\mu \mapsto \Psi_k \cdot \mu$ . De même, soit

$$G_k(\mu) = \frac{S_k \Psi_k \mu}{\langle \mu, \Psi_k \rangle} - \frac{\langle \mu, S_k \Psi_k \rangle}{\langle \mu, \Psi_k \rangle} \frac{\Psi_k \mu}{\langle \mu, \Psi_k \rangle} = F_k(\mu) (S_k \mu).$$

la dérivée de la distribution de probabilité  $\Psi_k \cdot \mu$  par rapport au paramètre. L'évolution de la suite  $\{w_n, n \geq 0\}$ , à valeurs dans l'espace des mesures signées finies sur  $E$  de masse nulle, est décrite par les étapes suivantes

$$\begin{array}{l} \text{prédiction} \\ \text{linéaire tangente} \\ w_{k-1} \longrightarrow w_{k|k-1} = w_{k-1} Q + \mu_{k-1} \Gamma \\ \\ \text{correction} \\ \text{linéaire tangente} \\ \longrightarrow w_k = F_k(\mu_{k|k-1}) (w_{k|k-1} + S_k \mu_{k|k-1}) = \bar{W}_k(\mu_{k|k-1}, w_{k|k-1}). \end{array}$$

### 3 Approximation particulière conjointe

Au vu de la relation d'absolue continuité  $w_k \ll \mu_k$ , l'idée est d'approcher conjointement le filtre optimal  $\mu_k$  et sa dérivée  $w_k$  par rapport au paramètre à l'aide de deux distributions empiriques pondérées de la forme

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N \omega_k^i \delta_{\zeta_k^i} \quad \text{et} \quad w_k \approx w_k^N = \sum_{i=1}^N \rho_k^i \omega_k^i \delta_{\zeta_k^i},$$

avec  $\sum_{i=1}^N \omega_k^i = 1$  et  $\sum_{i=1}^N \rho_k^i \omega_k^i = 0$ . Cette approximation est complètement caractérisée par la donnée du système de particules  $\Sigma_k = \{(\xi_k^1, \omega_k^1, \rho_k^1), \dots, (\xi_k^N, \omega_k^N, \rho_k^N)\}$ , et l'algorithme est décrit par le mécanisme qui permet de construire  $\Sigma_k$  à partir de  $\Sigma_{k-1}$ . On note que

$$\Psi_k(x') (\mu_{k-1}^N Q)(dx') = \sum_{i=1}^N \int_S \Psi_k(x') \underbrace{\omega_{k-1}^i K(\xi_{k-1}^i, dx', ds')}_{m_k^i(dx', ds')},$$

et

$$\begin{aligned} & \Psi_k(x') [w_{k-1}^N Q(dx') + \mu_{k-1}^N \Gamma(dx') + S_k(x') (\mu_{k-1}^N Q)(dx')] \\ &= \sum_{i=1}^N \int_S \Psi_k(x') \underbrace{[\rho_{k-1}^i + s' + S_k(x')]}_{r_k^i(x', s')} \underbrace{\omega_{k-1}^i K(\xi_{k-1}^i, dx', ds')}_{m_k^i(dx', ds')}, \end{aligned}$$

de sorte que  $\bar{R}_k(\mu_{k-1}^N)(dx')$  et  $\bar{W}_k(\mu_{k-1}^N, w_{k-1}^N)(dx')$  sont obtenues respectivement par marginalisation d'une distribution de probabilité et d'une mesure signée finie absolument continue, définies sur l'espace produit  $\{1, \dots, N\} \times E \times S$ . En suivant l'approche de [5] (i.e. échantillonnage pondéré sur l'espace produit, puis marginalisation), on obtient

$$\bar{R}_k(\mu_{k-1}^N) \approx \mu_k^N = \sum_{i=1}^N \frac{\Psi_k(\xi_k^i)}{\sum_{i=1}^N \Psi_k(\xi_k^i)} \delta_{\xi_k^i} = \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \omega_k^i = \frac{\Psi_k(\xi_k^i)}{\sum_{i=1}^N \Psi_k(\xi_k^i)},$$

et

$$\bar{W}_k(\mu_{k-1}^N, w_{k-1}^N) \approx w_k^N = \sum_{i=1}^N \frac{\Psi_k(\xi_k^i) \underbrace{[\rho_{k-1}^{\tau_k^i} + \Xi_k^i + S_k(\xi_k^i) - a_k^N]}_{r_k^{\tau_k^i}(\xi_k^i, \Xi_k^i)}}{\sum_{i=1}^N \Psi_k(\xi_k^i)} \delta_{\xi_k^i} = \sum_{i=1}^N \rho_k^i \omega_k^i \delta_{\xi_k^i},$$

avec

$$\rho_k^i = \rho_{k-1}^{\tau_k^i} + \Xi_k^i + S_k(\xi_k^i) - a_k^N,$$

où la constante  $a_k^N$  est uniquement déterminée par la condition de normalisation  $\langle w_k^N, 1 \rangle = 0$ , et où  $((\tau_k^1, \xi_k^1, \Xi_k^1), \dots, (\tau_k^N, \xi_k^N, \Xi_k^N))$  est un  $N$ -échantillon à valeurs dans l'espace produit  $\{1, \dots, N\} \times E \times S$ , de loi commune  $(m_k^1(dx', ds'), \dots, m_k^N(dx', ds'))$ , c'est-à-dire

$$\tau_{k-1}^i \sim (\omega_{k-1}^1, \dots, \omega_{k-1}^N) \quad \text{et} \quad (\xi_k^i, \Xi_k^i) \sim K(\xi_{k-1}^{\tau_{k-1}^i}, dx', ds').$$

On en déduit les approximations particulières suivantes

$$\ell_n \approx \frac{1}{n} \sum_{k=0}^n \log \left[ \sum_{i=1}^N \Psi_k(\xi_k^i) \right] \quad \text{et} \quad \partial \ell_n \approx \frac{1}{n} \sum_{k=0}^n \sum_{i=1}^N [\rho_{k-1}^{\tau_k^i} + \Xi_k^i + S_k(\xi_k^i)] \omega_k^i ,$$

pour la fonction de log-vraisemblance et pour la fonction score.

## 4 Approximation régulière

Dans le cas simple [2] où le noyau markovien vérifie  $Q^\theta(x, dx') = q^\theta(x, x') Q^{\theta_0}(x, dx')$ , avec une densité régulière par rapport au paramètre, alors l'hypothèse (AC) est vérifiée, avec  $\Xi_n^\theta = \partial \log q^\theta(X_{n-1}, X_n)$ , et il est possible d'obtenir par récurrence une approximation particulière de la forme

$$\mu_k^\theta \approx \mu_k^{N,\theta} = \sum_{i=1}^N u_k^{i,\theta} \omega_k^i \delta_{\xi_k^i} ,$$

où le système de particules  $\Sigma_k = \{(\xi_k^1, \omega_k^1), \dots, (\xi_k^N, \omega_k^N)\}$  ne dépend pas du paramètre, et où les poids  $\{u_k^{1,\theta}, \dots, u_k^{N,\theta}\}$  dépendent de manière régulière du paramètre [3] : en effet

$$\Psi_k^\theta(x') (\mu_{k-1}^{N,\theta} Q^\theta)(dx') = \sum_{i=1}^N \underbrace{u_{k-1}^{i,\theta} \Psi_k^\theta(x') q^\theta(\xi_{k-1}^i, x')}_{r_k^{i,\theta}(x')} \underbrace{\omega_{k-1}^i Q^{\theta_0}(\xi_{k-1}^i, dx')}_{m_k^i(dx')} ,$$

d'où l'approximation

$$\bar{R}_k^\theta(\mu_{k-1}^{N,\theta}) \approx \mu_k^{N,\theta} = c_k^{N,\theta} \sum_{k=1}^N \underbrace{u_{k-1}^{\tau_{k-1}^i, \theta} \Psi_k^\theta(\xi_k^i) q^\theta(\xi_{k-1}^{\tau_{k-1}^i}, \xi_k^i)}_{r_k^{\tau_{k-1}^i, \theta}(\xi_k^i)} \delta_{\xi_k^i} = \sum_{i=1}^N u_k^{i,\theta} \omega_k^i \delta_{\xi_k^i} ,$$

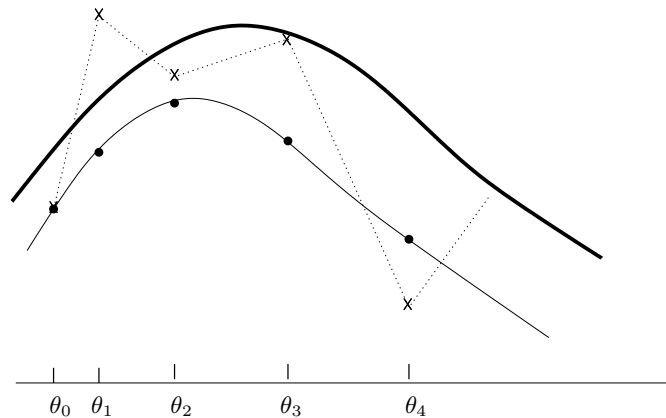
avec

$$\omega_k^{i,\theta} = c_k^{N,\theta} u_{k-1}^{\tau_{k-1}^i, \theta} \Psi_k^\theta(\xi_k^i) q^\theta(\xi_{k-1}^{\tau_{k-1}^i}, \xi_k^i) , \quad \omega_k^i = \omega_k^{i,\theta_0} \quad \text{et} \quad u_k^{i,\theta} = \frac{\omega_k^{i,\theta}}{\omega_k^{i,\theta_0}} ,$$

où  $((\tau_k^1, \xi_k^1), \dots, (\tau_k^N, \xi_k^N))$  est un  $N$ -échantillon à valeurs dans l'espace produit  $\{1, \dots, N\} \times E$ , de loi commune  $(m_k^1(dx'), \dots, m_k^N(dx'))$ , c'est-à-dire

$$\tau_{k-1}^i \sim (\omega_{k-1}^1, \dots, \omega_{k-1}^N) \quad \text{et} \quad \xi_k^i \sim Q^{\theta_0}(\xi_{k-1}^{\tau_{k-1}^i}, dx') .$$

Au lieu d'une collection d'approximations ponctuelles (croix) de la fonction de contraste (trait épais), utilisant chacune un système de particules différent, on obtient ainsi une



approximation globale (trait fin), utilisant un unique système de particules, défini pour la valeur  $\theta = \theta_0$  du paramètre (cette approximation est bien sûr d'autant moins précise que  $\theta$  est éloigné de  $\theta_0$ ). Enfin, on peut vérifier que l'approximation particulière  $w_k^N$  obtenue à la section précédente coïncide avec la dérivée en  $\theta = \theta_0$  de l'application  $\theta \mapsto \mu_k^{N,\theta}$ .

## Références bibliographiques

- [1] F. Cérou, F. Le Gland, and N. J. Newton. Stochastic particle methods for linear tangent filtering equations. In J.-L. Menaldi, E. Rofman, and A. Sulem, editors, *Optimal Control and Partial Differential Equations. In honour of professor Alain Bensoussan's 60th birthday*, pages 231–240. IOS Press, Amsterdam, 2001.
- [2] A. Doucet and V. B. Tadić. Parameter estimation in general state–space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55(2):409–422, Mar. 2003.
- [3] C. J. Geyer. Estimation and optimization of functions. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, chapter 14, pages 241–258. Chapman & Hall, London, 1996.
- [4] A. Guyader, F. Le Gland, and N. Oudjane. A particle implementation of the recursive MLE for partially observed diffusions. In *Proceedings of the 13th Symposium on System Identification (SYSID), Rotterdam*, pages 1305–1310. IFAC / IFORS, Aug. 2003.
- [5] M. K. Pitt and N. Shephard. Filtering via simulations : auxiliary particle filter. *Journal of the American Statistical Association*, 94(446):590–599, June 1999.