

## Implantation du paradigme Bayésien pour l'estimation récursive d'état

Arnaud DOUCET††, Eric BARAT† et Patrick DUVAUT‡

†LETI-CEA Technologies Avancées 91191 Gif sur Yvette  
‡ETIS - ENSEA 6, avenue du Ponceau 95014 Cergy Pontoise

### Résumé

Le problème de l'estimation Bayésienne récursive d'état a de nombreuses applications en traitement du signal. Théoriquement, la solution est simple mais n'admet pas généralement une expression analytique. Récemment, des approximations Monte Carlo sont apparues. Dans cet article, nous proposons diverses implantations du paradigme Bayésien et les couplons à des procédures de tirages aléatoires conditionnels. Une simulation sur une série temporelle non linéaire est présentée.

### 1. INTRODUCTION

L'estimation récursive Bayésienne d'état est un sujet d'un grand intérêt pratique, de très nombreux systèmes physiques pouvant se modéliser sous la forme de modèles d'état non linéaires non Gaussiens. Malheureusement, hormis quelques exceptions tel le filtre de Kalman pour les systèmes linéaires Gaussiens, la solution est de dimension infinie et on doit faire des approximations.

Récemment, l'augmentation de la puissance des calculateurs a permis l'introduction de nouvelles méthodes d'approximations de type Monte Carlo consistant à approximer la densité par un ensemble d'échantillons. Ces méthodes simulent l'équation d'évolution du système et approximent la règle de Bayes soit numériquement soit statistiquement.

Nous proposons dans cet article de nouveaux algorithmes d'approximation des équations de filtrage basés sur des implantations statistiques du paradigme Bayésien et l'utilisation de tirages conditionnels.

### 2. MODÈLE D'ÉTAT NON LINÉAIRE NON GAUSSIEN

Considérons le modèle d'état non linéaire non Gaussien suivant :

$$x(k) = f_k(x(k-1), v(k)) \quad (1)$$

$$y(k) = g_k(x(k), w(k)) \quad (2)$$

### Abstract

*Recursive Bayesian state estimation has many applications in signal processing. The theoretical solution is simple but generally does not admit an analytical expression. Recently, some Monte Carlo approximations have appeared. In this article, we propose various statistical implementations of the Bayesian paradigm and couple these algorithms to conditional random sampling. A simulation on a nonlinear time series is given.*

où  $x(k)$ ,  $v(k)$ ,  $y(k)$  et  $w(k)$  sont de dimension respec-

tive  $n$ ,  $m$ ,  $p$  et  $r$ .  $f_k$  est la fonction de transition du système,  $g_k$  est l'équation de mesure.  $v(k)$  et  $w(k)$  sont deux suites i.i.d. indépendantes l'une de l'autre de densités  $p(v(k))$  et  $p(w(k))$ .  $p(x(0))$  est supposé connue. Le but est d'estimer la densité conditionnelle  $p(x(k)|Y_k)$  où  $Y_k = \{y(1), \dots, y(k-1), y(k)\}$ . La propagation de la densité de filtrage de  $p(x(k-1)|Y_{k-1})$  à  $p(x(k)|Y_k)$  peut se décomposer en 2 parties, prédiction et correction, comme il est d'usage, ou en lissage et correction. Cette dernière décomposition, introduite par [8][2], a un grand intérêt pratique.

$$p(x(k-1)|Y_k) = \frac{p(y(k)|x(k-1))p(x(k-1)|Y_{k-1})}{\int p(y(k)|x(k-1))p(x(k-1)|Y_{k-1})dx(k-1)} \quad (3)$$

$$p(x(k)|Y_k) = \int p(x(k)|x(k-1), y(k)) p(x(k-1)|Y_k) dx(k-1) \quad (4)$$

En général, ces équations n'admettent pas de solution analytique à de rares exceptions près. Il est nécessaire de procéder à des approximations.



### 3. APPROXIMATIONS MONTE CARLO

Avec l'augmentation de la puissance de calcul, des algorithmes utilisant des procédures d'intégration numérique classiques sont apparues [6]. Parallèlement, une approche plus orientée "simulation" consistant à représenter la densité de filtrage par un ensemble important d'échantillons aléatoires et des intégrations, ainsi Monte Carlo, s'est développée [1][7][5][2]. Malheureusement, malgré de bonnes performances, ces algorithmes requièrent en pratique de nombreux réglages : choix d'un seuil, coefficient de régularisation etc.

Supposons que l'on dispose d'un ensemble d'échantillons aléatoires  $\{x_i(k); i = 1, \dots, P\}$  distribués selon  $p(x(k)|Y_k)$ . On fera ainsi l'approximation

$$p(x(k)|Y_k) \simeq \frac{1}{P} \sum_{i=1}^P \delta(x(k) - x_i(k)) \quad (5)$$

Il est ainsi facile de calculer une approximation de l'espérance conditionnelle :

$$\begin{aligned} E(x(k)|Y_k) &= \int x(k) p(x(k)|Y_k) dx(k) \\ &\simeq \frac{1}{P} \sum_{i=1}^P x_i(k) \end{aligned} \quad (6)$$

mais aussi des moments ou des intervalles de confiance. Nos algorithmes sont des approximations statistiques des équations (3)(4) qui, étant donné  $N$  échantillons distribués selon  $p(x(k-1)|Y_{k-1})$ , génèrent  $N$  échantillons distribués selon  $p(x(k)|Y_k)$ . Ces équations nécessitent une intégration numérique (4), et la règle de Bayes (3).

#### 3.1. Intégration Monte Carlo

On suppose disposer d'un ensemble d'échantillons aléatoires  $\{x_i'(k-1); i' = 1, \dots, N\}$  distribués selon  $p(x(k-1)|Y_k)$ . Pour obtenir  $\alpha(N) = \gamma N \geq N$  échantillons  $\{x_i(k); i = 1, \dots, \alpha(N)\}$  selon  $p(x(k)|Y_k)$ , on simule  $\forall i$   $\gamma$  échantillons de bruit de dynamique distribués selon  $p(v(k)|x_i'(k-1), y(k))$ . Puis on calcule l'équation d'évolution (1) pour obtenir le nouvel état. On voit ici apparaître l'intérêt des tirages conditionnels introduits par Salut et coll [2][8]. En effet, expérimentalement si l'on simule l'étape de prédiction classique, cette phase apparaît "aveugle" car elle n'est pas conditionnée par l'observation  $y(k)$ . Il peut y avoir alors une forte dispersion des échantillons dans l'espace d'état due à cette exploration *a priori* et les résultats peuvent être désastreux. Ici, le tirage est conditionnel à l'observation. Les échantillons se localisent naturellement dans une zone pertinente de l'espace d'état. En contrepartie, il faut pouvoir simuler des échantillons de bruit de dynamique distribués selon  $p(v(k)|x_i'(k-1), y(k))$ . Par la suite, on supposera  $p(v(k)|x_i'(k-1), y(k))$  et  $p(y(k)|x(k-1))$  connus. Si tel n'est pas le cas, une approximation est proposée dans la partie 3.3.

### 3.2. Implantation statistique du paradigme Bayésien

Nous proposons plusieurs méthodes statistiques d'implantation du paradigme Bayésien pour approximer (3). Celles-ci permettent l'élimination statistique des points de faible vraisemblance et limite très sérieusement les problèmes de dégénérescence. En contrepartie, l'étude, même asymptotique, du comportement de l'algorithme est quasi impossible car les points générés ne sont pas indépendants.

#### 3.2.1. Algorithme de Métropolis-Hastings

Nous décrivons brièvement l'algorithme de Métropolis-Hastings. Soit  $\pi(x)$  la densité selon laquelle on désire échantillonner. On construit une chaîne de Markov  $\{x(k), k \in N\}$  dont la distribution invariante est  $\pi(x)$ . Soit  $q(x, x')$  une probabilité génératrice d'état  $x'$  sachant  $x$ . Cette valeur est acceptée avec la probabilité  $\alpha(x, x')$  donnée par :

$$\alpha(x, x') = \begin{cases} \min \left[ \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}, 1 \right] & \text{si } \pi(x)q(x, x') > 0 \\ 1 & \text{sinon} \end{cases} \quad (7)$$

Alors, sous quelques hypothèses peu restrictives, on montre que  $\pi(x)$  est la distribution invariante vers laquelle la chaîne de Markov converge.

L'algorithme de Métropolis se restreint au cas  $q(x, x') = q(x', x)$ : Cette égalité est notamment vérifiée lorsque l'on considère  $x' = x + z$ ,  $z$  étant un accroissement aléatoire de densité symétrique  $q_a(z) = q_a(-z)$ . L'état tiré suit une marche aléatoire. Dans un cadre Bayésien, on a  $\pi(x) = p(x|y) = \frac{p(y|x)p(x)}{p(y)}$  et la vraisemblance  $p(y|x)$  est généralement connue. On a alors :

$$\alpha(x, x') = \begin{cases} \min \left[ \frac{p(y|x')p(x')}{p(y|x)p(x)}, 1 \right] & \text{si } \pi(x)q_a(x - x') > 0 \\ 1 & \text{sinon} \end{cases} \quad (8)$$

L'application de cet algorithme à l'estimation réursive est quelque peu complexe. En effet, il est nécessaire de fixer une densité d'incrément aléatoire  $q_a(z) = q(x - x') = q(x, x')$ , *a priori* quelconque. On peut utiliser une Gaussienne multivariée. De plus, le calcul de  $\alpha(x, x')$  requiert l'évaluation numérique en tout point  $x$  de  $p(y|x)p(x)$ , soit dans notre application  $p(y(k)|x(k-1))p(x(k-1)|Y_{k-1})$  alors qu'on ne dispose que d'un nombre important d'échantillons. Pour reconstruire la densité, on peut convoluer par un noyau. Cette méthode permet de traiter des cas où la nouvelle observation  $y(k)$  modifie totalement le support de la densité ou, par exemple, l'identification de paramètre constant. L'algorithme à tirages indépendants considère le cas où  $q(x, x') = q_i(x')$ . En Bayésien, si l'on choisit  $q(x') = p(x')$  alors on obtient :

$$\alpha(x, x') = \begin{cases} \min \left[ \frac{p(y|x')}{p(y|x)}, 1 \right] & \text{si } \pi(x)q(x') > 0 \\ 1 & \text{sinon} \end{cases} \quad (9)$$

Cet algorithme nécessite le calcul de la vraisemblance  $p(y|x)$ , i.e.  $p(y(k)|x(k-1))$  ici, et la possibilité d'échantillonner selon  $p(x)$ , i.e.  $p(x(k-1)|Y_{k-1})$ . Cette densité est discrète, mais on peut aussi la convoluer par un noyau.

En pratique, on itère un nombre  $p$  de fois l'algorithme jusqu'à ce que la distribution stationnaire soit supposée atteinte. Puis on garde un échantillon toutes les  $q$  itérations afin de limiter l'autocorrélation entre les échantillons. Sous des hypothèses peu restrictives, la convergence est géométrique [10], ainsi  $p$  demeure raisonnable. On peut bien sûr aussi lancer plusieurs algorithmes en parallèle, avec des initialisations différentes l'une de l'autre.

### 3.2.2. Bootstrap pondéré

Une alternative aux méthodes itératives est l'algorithme de bootstrap pondéré ou Sampling Importance Resampling [9], c'est une variante de la procédure de bootstrap classique. Supposons que l'on dispose de  $\alpha(N)$  échantillons indépendants  $\{x(i); i = 1, \dots, \alpha(N)\}$  distribués selon la densité  $g(x)$ . On désire obtenir  $N$  échantillons indépendants distribués selon la densité non normalisée  $h(x) \propto f(x)$ .

On calcule  $w_i = f(x_i)/g(x_i)$  et  $p_i = w_i / \sum_{j=1}^{\alpha(N)} w_j$  puis

on effectue  $N$  tirages  $x'(i)$  selon la distribution discrète constituée de Dirac de support  $\{x(i); i = 1, \dots, N\}$  et de masse associée  $p_i$ . Ces tirages peuvent être effectués rapidement par dichotomie. Alors pour  $\frac{\alpha(N)}{N} \rightarrow +\infty$ , sous hypothèse d'absolue continuité de  $f$  par rapport à  $g$ , ces tirages sont indépendants et tendent avec la probabilité 1 vers la densité requise [9][3]. Cette procédure est très simple et a de nombreux avantages. Dans un cadre Bayésien, on a  $h(x) = p(x|y) \propto p(y|x)p(x)$ . On simule donc des échantillons selon la loi *a priori* qu'on retire avec une probabilité proportionnelle à leur vraisemblance. La simulation de l'équation de correction classique par bootstrap pondéré est due à [5]. On peut bien sûr l'utiliser de même pour la simulation de (3) avec  $g(x) = p(x(k-1)|Y_{k-1})$  et  $h(x) = p(x(k-1)|Y_k)$ , soit  $p_i = p(y(k)|x_i(k-1))$ . Cet algorithme est une version non itérative très proche de l'algorithme M-H à tirages indépendants.

### 3.3. Approximations des tirages conditionnels

L'utilisation de tirages conditionnels améliore souvent substantiellement les résultats en pratique puisqu'elle permet d'éviter une exploration *a priori* de l'espace d'état. Malheureusement, cette technique requiert de savoir tirer des échantillons selon  $p(v(k)|x_i(k-1), y(k))$  et le calcul de  $p(y(k)|x_i(k-1))$ . Dans de très nombreux cas pratique, il est impossible d'établir ces expressions. On peut utiliser alors soit des techniques de génération classiques, soit procéder à une seconde approximation de type Monte Carlo en remarquant que d'après Bayes :

$$p(x(k)|x_i(k-1), y(k)) = \frac{p(y(k)|x(k))p(x(k)|x_i(k-1))}{p(y(k)|x_i(k-1))} \quad (10)$$

$$p(y(k)|x_i(k-1)) = \int p(y(k)|x(k))p(x(k)|x_i(k-1))dx(k) \quad (11)$$

La première de ces équations requiert donc une approximation de la règle de Bayes, la seconde une intégration Monte Carlo. On peut donc utiliser les algorithmes précédemment décrits. Cette étape permet une élimination statistique des échantillons générés *a priori* dans une partie de l'espace d'état non pertinente par rapport à  $y(k)$ . On peut ainsi appliquer l'algorithme à tirages conditionnels. Il faut noter qu'il est des cas où l'utilisation de tirages conditionnels sera vaine. Dans ces cas, il est nécessaire de lisser [3][4].

## 4. SIMULATION

Nous présentons les performances du filtre sur le modèle non linéaire introduit par Kitagawa[6]. Soit le modèle suivant :

$$\begin{aligned} x(k) &= 0.5x(k-1) + 25x(k-1)/(1+x^2(k-1)) \\ &\quad + 8\cos(1.2(k-1)) + v(k) \\ y(k) &= x^2(k)/20 + w(k) \end{aligned}$$

où  $v(k)$  et  $w(k)$  sont deux bruits blancs Gaussiens d'écart type respectivement égaux à  $\sigma_v = 10.0$  et  $\sigma_w = 1.0$ . On peut établir les expressions des densités conditionnelles mais il n'est pas possible d'effectuer des tirages selon  $p(x(k)|x(k-1), y(k))$ . Nous allons donc simuler ces densités conditionnelles en utilisant une deuxième étape Monte Carlo, comme décrit en 3.3. Les paramètres sont fixés à  $\alpha(N) = N = 1000$ , on utilise un bootstrap pondéré pour simuler la règle de Bayes.

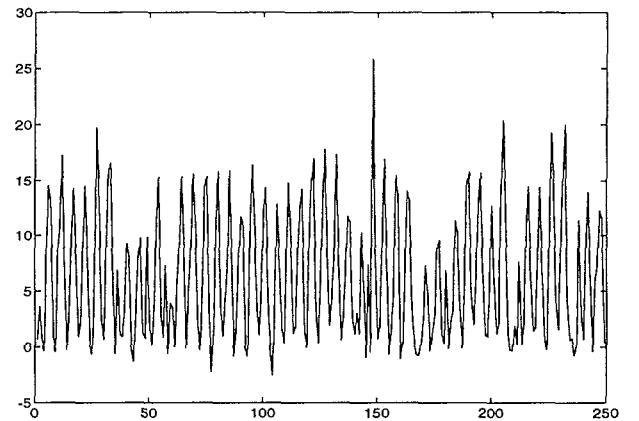


Figure 1: Observations

Le signal est donné en pointillé, le lisseur en trait plein. Pour ce problème, le lisseur de Kalman étendu donne des résultats très médiocres car il ne peut prendre en compte la structure bimodale de la densité *a posteriori*. Notre algorithme présente des résultats bien meilleurs que cet algorithme classique et que des algorithmes plus sophistiqués

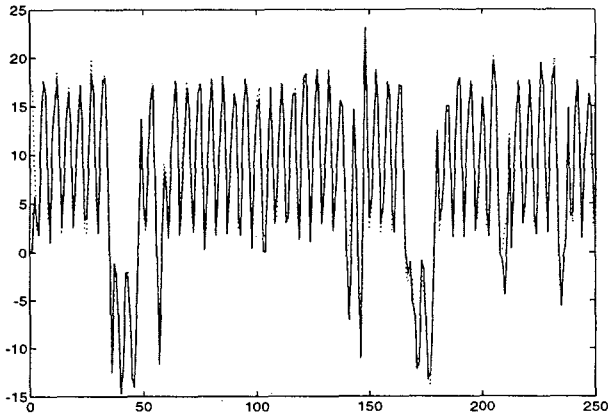


Figure 2: Lisseur à pas fixe et Signal

[6][5]. Sur 20 réalisations de 500 points, la moyenne de l'erreur du lisseur à pas fixe (pas fixé à 5) est de  $m_\varepsilon = 0.01$  et son écart type de  $\sigma_\varepsilon = 1.27$ .

## 5. CONCLUSION

Dans cet article, nous avons proposé plusieurs algorithmes d'approximation de l'estimation Bayésienne récursive d'état. Ces algorithmes sont basés sur une représentation des densités par un ensemble important d'échantillons. Ils utilisent des approximations statistiques, itératives ou non itératives, de la règle de Bayes, des intégrations Monte Carlo et des tirages conditionnels. En simulation, ces algorithmes se sont révélés très performants sur des problèmes non linéaires et/ou non Gaussiens [3][4]. Ils n'ont pas requis de procédures heuristiques pour leur bon fonctionnement expérimental. Malheureusement, leur étude, même asymptotique, est quasi irréalisable. La charge de calcul est de plus relativement prohibitive bien qu'une parallélisation triviale est réalisable.

## 6. RÉFÉRENCES

- [1] P.R. Benyon, "Monte Carlo and Other Methods for Nonlinear Non-Gaussian Estimation", *Mathematics and Computers in Simulation*, no. 32, pp. 215-220, 1990.
- [2] P. Del Moral, J.C. Noyer, G. Rigal et G. Salut, "Traitement Non-Linéaire par Réseau Particulaire", 14ème conférence Grets, Juan-les-Pins, France, pp. 399-402, 1993.
- [3] A. Doucet, "Sampling Based Approaches to Recursive Bayesian State Estimation", rapport technique, 1994.
- [4] A. Doucet, E. Barat et P. Duvaut, "A Monte Carlo Approach to Recursive Bayesian State Estimation", *Proceedings IEEE Conference on Higher Order Statistics*, Espagne, Juin 1995.

- [5] N.J. Gordon, D.J. Salmond et A.F.M. Smith, "Novel Approach to Nonlinear Non Gaussian Bayesian State Estimation", *IEE Proceedings-F*, vol. 140, no. 2, pp. 107-113, 1993.
- [6] G. Kitagawa, "Non-Gaussian State-Space Modeling of Nonstationary Time Series", *Journal of the American Statistical Association*, vol. 82, no. 400; pp. 1032-1063, 1987.
- [7] P. Müller, "Monte Carlo Integration in General Dynamic Models", *Contemporary Mathematics*, vol. 115, pp. 145-163, 1991.
- [8] G. Rigal, *Filtrage non-linéaire, résolution particulière et applications au traitement du signal*, thèse de l'Université Paul Sabatier, Toulouse, Juillet 1993.
- [9] A.F.M. Smith et A.E. Gelfand, "Bayesian Statistics without Tears: a Sampling-Resampling Perspective", *The American Statistician*, vol. 46, no. 2, pp. 84-88, 1992.
- [10] L. Tierney, "Markov Chains for Exploring Posterior Distributions", rapport technique, 1991.