

REVUE DE STATISTIQUE APPLIQUÉE

ARNAUD GUILLIN

JEAN-MICHEL MARIN

CHRISTIAN P. ROBERT

Estimation bayésienne approximative par échantillonnage préférentiel

Revue de statistique appliquée, tome 53, n° 1 (2005), p. 79-95

http://www.numdam.org/item?id=RSA_2005__53_1_79_0

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*

<http://www.numdam.org/>

ESTIMATION BAYÉSIENNE APPROXIMATIVE PAR ÉCHANTILLONNAGE PRÉFÉRENTIEL

Arnaud GUILLIN¹, Jean-Michel MARIN², Christian P. ROBERT³

¹ CEREMADE, Université Paris IX Dauphine, Paris
guillin@ceremade.dauphine.fr

² CEREMADE, Université Paris IX Dauphine, Paris
marin@ceremade.dauphine.fr

³ CEREMADE, Université Paris IX Dauphine, Paris et CREST, INSEE, Paris
xian@ceremade.dauphine.fr

RÉSUMÉ

En estimation bayésienne, lorsque le calcul explicite de la loi *a posteriori* du vecteur des paramètres à estimer est impossible, les méthodes de Monte-Carlo par chaînes de Markov (MCMC) [Robert and Casella, 1999] permettent théoriquement de fournir un échantillon approximativement distribué suivant cette loi. Il est alors possible d'utiliser ce dernier afin d'estimer l'espérance mathématique *a posteriori* de n'importe quelle fonction du vecteur paramétrique. Depuis le début des années 90, ces méthodes sont très utilisées et leurs bonnes performances sont mises en exergue. Cet engouement a peu à peu marginalisé l'utilisation des méthodes d'échantillonnage préférentiel [Rubinstein, 1981, Ripley, 1987] qui leur sont antérieures. Dans cet article, nous montrons qu'il est possible d'améliorer le schéma d'échantillonnage préférentiel classique en lui introduisant une dimension temporelle assurant son adaptativité. Il s'agit de l'algorithme Population Monte Carlo (PMC) proposé par Cappé *et al.* [2004]. Nous illustrons les bonnes propriétés d'un tel schéma sur un modèle ARMA à changement de régime markovien latent [Hamilton, 1988, 1989] pour lequel les méthodes MCMC génériques ne fournissent pas toujours des approximations satisfaisantes.

Mots-clés : Statistique bayésienne, méthodes MCMC, fonction d'importance, adaptativité, schéma PMC, chaînes de Markov cachées, modèles auto-régressifs à sauts.

ABSTRACT

For numerous models, it is impossible to conduct an exact Bayesian inference. There are many cases where the derivation of the posterior distribution leads to intractable calculations (due to the fact that this generally involves intractable integrations). The Bayesian computational literature has been dominated by the development of simulation approximations based on Markov chains, the famous Markov Chains Monte Carlo (MCMC) methods [Robert and Casella, 1999]. However, the more complicated the model, (especially if there are many parameters and/or hyperparameters), the more expensive the MCMC approaches become, in terms of time and storage. Moreover, it is difficult to monitor the convergence of Monte Carlo schemes. As an alternative to the use of the MCMC approximations, Cappé *et al.* [2004] have shown that the notion of importance sampling [Rubinstein, 1981, Ripley, 1987] can actually be greatly

generalized to encompass much more adaptive and local schemes than thought previously. This leads to the Population Monte Carlo (PMC) algorithm. We illustrate here the good properties of this scheme on a switching ARMA model [Hamilton, 1988, 1989] for which the MCMC approximations are less satisfactory.

Keywords : *Bayesian inference, MCMC approximations, importance sampling, adaptive algorithms, PMC scheme, hidden Markov chains, switching AR models.*

1. Introduction

Nous considérons un vecteur aléatoire y ayant pour densité de probabilité $f(y|\theta)$ où θ est le vecteur des paramètres inconnus. En statistique bayésienne, θ est considéré comme un vecteur aléatoire de densité $\pi(\theta)$, loi *a priori* sur θ . L'estimation bayésienne est basée sur le calcul de la loi *a posteriori* de θ , loi conditionnelle de θ sachant y dont la densité est notée $\pi(\theta|y)$. La majorité des estimateurs bayésiens sont de la forme $\mathbb{E}_\pi(h(\theta))$ où \mathbb{E}_π est l'opérateur d'espérance mathématique pour la loi $\pi(\theta|y)$.

Dans de nombreux cas, le calcul explicite de $\pi(\theta|y)$ est impossible, et celui de $\mathbb{E}_\pi(h(\theta))$, même si $\pi(\theta|y)$ est connue, peut s'avérer très difficile. Ainsi, de nombreuses méthodes d'approximation ont été proposées. Nous nous intéressons ici aux méthodes stochastiques. Il s'agit de déterminer une approximation de

$$\mathcal{I}_h = \mathbb{E}_\pi(h(\theta)) = \int h(\theta)\pi(\theta|y)d\theta. \quad (1)$$

Lorsque $\pi(\theta|y)$ est connue, la méthode de Monte-Carlo standard consiste à estimer \mathcal{I}_h par

$$\widehat{\mathcal{I}}_h = \frac{1}{n} \sum_{i=1}^n h(\theta_i) \quad (2)$$

où $\theta_1, \dots, \theta_n \stackrel{iid}{\sim} \pi(\theta|y)$. D'après la loi forte des grands nombres, lorsque n tend vers l'infini, $\widehat{\mathcal{I}}_h$ converge presque sûrement vers \mathcal{I}_h .

Lorsque $\pi(\theta|y)$ est inconnue ou trop complexe pour être simulée directement, les méthodes de Monte-Carlo par chaînes de Markov (MCMC) proposent d'estimer \mathcal{I}_h par

$$\widehat{\mathcal{I}}_h = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) \quad (3)$$

où les $\theta^{(t)}$ sont les réalisations d'une chaîne de Markov ayant pour distribution stationnaire $\pi(\theta|y)$. La convergence de $\widehat{\mathcal{I}}_h$ vers \mathcal{I}_h est assurée par l'ergodicité de la chaîne de Markov $\theta^{(t)}$.

Les méthodes d'échantillonnage préférentiel (ou d'importance), qui sont également des méthodes de Monte-Carlo, utilisent le fait que l'intégrale \mathcal{I}_h s'exprime aussi sous la forme

$$\mathcal{I}_h = \int h(\theta) \frac{\pi(\theta|y)}{\rho(\theta)} \rho(\theta) d\theta = \mathbb{E}_\rho \left[h(\theta) \frac{\pi(\theta|y)}{\rho(\theta)} \right] \quad (4)$$

pour toute densité ρ partout non nulle. Elles proposent alors d'estimer \mathcal{I}_h par

$$\widehat{\mathcal{I}}_h = \frac{1}{n} \sum_{i=1}^n \frac{\pi(\theta_i|y)}{\rho(\theta_i)} h(\theta_i) \quad (5)$$

où $\theta_1, \dots, \theta_n \stackrel{iid}{\sim} \rho(\theta)$. Cet estimateur est convergent pour tout choix de $\rho(\theta)$ tel que [Robert and Casella, 1999] :

$$\text{Supp}(\pi(\theta|y)) \subset \text{Supp}(\rho(\theta)) \quad \text{et} \quad \mathbb{E}_\pi \left[h^2(\theta) \frac{\pi(\theta|y)}{\rho(\theta)} \right] \quad \text{existe}$$

où $\text{Supp}(l(\theta))$ correspond au support de la densité de probabilité $l(\theta)$.

Depuis le début des années 90, les méthodes MCMC ont été très utilisées et ont permis des avancées notables dans le traitement bayésien de nouveaux modèles; c'est notamment le cas de nombreux modèles à données manquantes. L'algorithme de Metropolis–Hasting et l'échantillonneur de Gibbs sont les schémas de construction des chaînes ergodiques les plus répandus [Robert and Casella, 1999]. De par leur caractère « universel » et leur relative facilité d'implémentation, ces algorithmes sont couramment utilisés comme outils d'approximation stochastique et sont préférés aux méthodes d'échantillonnage préférentiel. Cependant, la convergence de la chaîne de Markov vers la loi stationnaire $\pi(\theta|y)$ peut s'avérer très lente et surtout impossible à quantifier théoriquement, rendant la qualité de l'approximation incertaine. C'est notamment le cas lorsque la dimension de θ est grande ou lorsque la variable y n'est que partiellement observée, comme dans l'exemple du modèle auto-régressif à changement de régime markovien que l'on examinera. Aussi, dans certains cas, exhibés dès le début de l'utilisation des méthodes MCMC, le temps de convergence n'est pas compatible avec les exigences de production, ce qui invalide l'emploi de ces méthodes. D'un autre côté, le nombre de simulations nécessaires à l'obtention d'une bonne approximation de \mathcal{I}_h par échantillonnage préférentiel dépend du choix de la distribution d'importance ρ . La recherche d'une bonne fonction d'importance est difficile et lorsque cette dernière n'est pas adaptée, l'efficacité de ce type de méthode est médiocre, voire nulle si l'espérance $\mathbb{E}_\pi \left[h^2(\theta) \frac{\pi(\theta|y)}{\rho(\theta)} \right]$ est infinie.

C'est ainsi qu'en tirant profit de l'expérience acquise lors du développement des méthodes MCMC tout en supprimant la difficulté fondamentale d'évaluation de la convergence, Cappé *et al.* [2004] ont proposé de généraliser la technique de l'échantillonnage préférentiel. Ils suggèrent d'introduire une dimension itérative à la production d'échantillons préférentiels, de manière à adapter la forme de la densité d'importance ρ en fonction de la loi cible $\pi(\theta|y)$. Cette adaptation à la loi $\pi(\theta|y)$ est implantée de manière séquentielle : à l'itération t de l'algorithme, n_t réalisations

sont simulées à partir d'une fonction d'importance dépendant des n_{t-1} réalisations de l'itération $t - 1$.

De manière à montrer les problèmes et avantages que présentent ces différentes méthodes de Monte-Carlo, la section 3 traite le cas d'un modèle auto-régressif à changement de régime markovien non observé. Il s'avère ainsi qu'une méthode MCMC comme l'échantillonneur de Gibbs peut nécessiter un très grand nombre de simulations avant de cerner les vrais valeurs des paramètres alors que la méthode par échantillonnage préférentiel adaptatif ou Population Monte Carlo (PMC), que nous présentons plus avant dans la section 2, ne requiert que peu de simulations.

2. L'approche Population Monte-Carlo

Nous présentons ici en détail l'approche PMC proposée initialement par Cappé *et al.* [2004] pour traiter le problème de reconstruction de signal (observé bruité) dans un canal ionique. Très récemment, Celeux *et al.* [2003] ont également mis à profit cette méthode pour étudier des modèles à données manquantes comme la volatilité stochastique.

2.1. Principes

Fondamentalement, l'approche PMC consiste en l'itération de deux étapes : tout d'abord, on simule une *population* de n variables aléatoires (parfois dites *particules*) suivant une loi choisie *a priori* (soit fonction des anciennes particules, soit selon une méthode choisie indépendamment), puis on calcule les *poids* de chaque particule suivant (4) et on retire un échantillon dans la population générée proportionnellement aux poids calculés. Soit q_{it} ($i = 1, \dots, M$ et $t = 1, \dots, T$) une distribution d'importance, la représentation algorithmique suivante correspond à ces deux étapes :

Algorithme 1 : Schéma PMC Général

- *étape 0* : Choix de $(\theta_0^{(1)}, \dots, \theta_0^{(M)})$;

- *étape t* ($t = 1, \dots, T$) :

a) pour $i = 1, \dots, M$:

$$\text{générer } \theta_t^{(i)} \sim q_{it}(\theta) \text{ et calculer } r_t^{(i)} = \frac{f(y|\theta_t^{(i)}) \pi(\theta_t^{(i)})}{q_{it}(\theta_t^{(i)})};$$

b) calculer $\omega_t^{(i)} = r_t^{(i)} / \sum_{k=1}^M r_t^{(k)}$ et rééchantillonner les $\theta_t^{(i)}$ en utilisant les poids

$$\omega_t^{(i)};$$

c) construire $q_{i(t+1)}$ à partir de l'échantillon courant (il existe de nombreuses possibilités, nous en détaillerons une dans le cas de données manquantes).

Après avoir itéré T fois l'algorithme précédent, un estimateur approximativement non biaisé de $\mathbb{E}_\pi(h(\theta))$ est donné par

$$\hat{\mathcal{I}}_h^T = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^M \omega_t^{(j)} h(\theta_t^{(j)}). \quad (6)$$

Pour chaque itération t , lorsque que le nombre de simulations M augmente, la qualité de l'estimateur précédent s'améliore. De manière évidente, d'après la loi forte des grands nombres, lorsque M tend vers l'infini,

$$\sum_{j=1}^M \omega_t^{(j)} h(\theta_t^{(j)})$$

converge presque sûrement vers $\mathbb{E}_\pi(h(\theta))$ quel que soit t . Par contre, à M fixé, lorsque T tend vers l'infini, la convergence presque sûre de

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^M \omega_t^{(j)} h(\theta_t^{(j)})$$

vers $\mathbb{E}_\pi(h(\theta))$ requiert de nombreuses hypothèses sur les distributions d'importance q_{it} et n'est pas étudiée ici.

2.2. Calibration

Le choix du temps d'arrêt T de l'algorithme est un problème important qui est en cours d'étude. À ce stade, on privilégie une approche empirique. Cette dernière consiste à étudier au cours des itérations le comportement de l'estimateur précédent et lorsque ce dernier est stabilisé, on arrête le schéma. Notons qu'il est possible d'augmenter M au cours des itérations et de donner ainsi un poids plus important aux dernières itérations. Par ailleurs, l'approximation (6) est valide quel que soit T car elle procède des mêmes arguments que les méthodes de Monte-Carlo usuelles et non pas de résultats de convergence comme les méthodes MCMC [Cappé *et al.*, 2004]. Ceci est vrai malgré la dépendance non markovienne au passé résultant de la construction de la fonction d'importance et de la sélection des particules en fonction de leur poids d'importance.

L'adaptativité de cet algorithme est ce qui justifie son caractère itératif : sans modification de la fonction d'importance entre deux itérations, il est inefficace d'employer PMC car les étapes de rééchantillonnage appauvrissent l'échantillon ainsi créé et génèrent des corrélations entre les points de l'échantillon qui augmentent la variance de $\hat{\mathcal{I}}_h^T$. Comme nous pouvons le remarquer dans l'écriture algorithmique ci-dessus, la méthode autorise également un choix de fonctions d'importance qui dépend de l'indice i dans l'échantillon produit. Cette caractéristique conduit à pouvoir utiliser des approximations par noyau de la loi cible, comme dans Warnes [2001]. La loi q_{it} correspond alors à la sélection (uniforme ou pondérée) d'une particule de la génération précédente, ξ_i , et à la génération d'une loi normale (ou autre) centrée en ξ_i . Voir Cappé *et al.* [2004] pour une illustration de cette approche dans le contexte des mélanges.

2.3. Données manquantes

Le schéma PMC très général proposé ci-dessus peut être avantageusement mis en œuvre dans le cas des modèles à structure latente [Little and Rubin, 1987, Tanner and Wong, 1987, Tanner, 1996, Van Dyk and Meng, 2001]. Il s'agit de modèles pour lesquels

$$f(y|\theta) = \int g(y, z|\theta) dz$$

où z représente le vecteur des données non observées. De tels modèles ont été à l'origine de nombreux développements méthodologiques en statistique mathématique. On pense notamment à l'algorithme EM [Dempster *et al.*, 1977, McLachlan and Krishnan, 1997] et à l'algorithme de «Data Augmentation», [Tanner and Wong, 1987, Gelfand and Smith, 1990].

Supposons que les distributions de $z|y, \theta$ et $\theta|y, z$ soient connues. On désigne par $k(z|y, \theta)$ et $\pi(\theta|y, z)$ ces deux lois. Une version spécifique de l'algorithme PMC général peut s'inspirer de l'échantillonneur de Gibbs en simulant z et θ suivant leurs lois conditionnelles respectives. On peut utiliser comme fonction d'importance q_{it} la distribution consistant à simuler $z_t^{(i)}$ suivant $k(z_t|y, \theta_{t-1}^{(i)})$ et $\theta_t^{(i)}$ suivant $\pi(\theta_t|y, z_t^{(i)})$. Le poids correspondant est alors

$$\omega_t^{(i)} \propto g(y, z_t^{(i)}|\theta_t^{(i)}) \pi(\theta_t^{(i)}) / k(z_t^{(i)}|y, \theta_{t-1}^{(i)}) \pi(\theta_t^{(i)}|y, z_t^{(i)}),$$

ce qui apporte la justification de l'algorithme suivant :

————— *Algorithme 2 : Schéma PMC pour les modèles à structure latente* —————

- *étape 0 : Choix de $(\theta_0^{(1)}, \dots, \theta_0^{(M)})$;*
- *étape t ($t = 1, \dots, T$) :*

a) Pour $i = 1, \dots, M$:

générer $z_t^{(i)} \sim k(z_t|y, \theta_{t-1}^{(i)})$;

générer $\theta_t^{(i)} \sim \pi(\theta_t|y, z_t^{(i)})$;

calculer $r_t^{(i)} = \frac{g(y, z_t^{(i)}|\theta_t^{(i)}) \pi(\theta_t^{(i)})}{k(z_t^{(i)}|y, \theta_{t-1}^{(i)}) \pi(\theta_t^{(i)}|y, z_t^{(i)})}$;

b) calculer $\omega_t^{(i)} = r_t^{(i)} / \sum_{k=1}^M r_t^{(k)}$ et rééchantillonner les $\theta_t^{(i)}$ en utilisant les poids $\omega_t^{(i)}$.

Il apparaît alors légitime de confronter l'approche de l'échantillonneur de Gibbs à l'approche PMC. C'est l'objet de la section suivante.

3. Modèle ARMA à changement de régime markovien

Dans cette section, nous comparons les comportements respectifs des schémas d'approximation MCMC et PMC pour l'estimation bayésienne des paramètres d'un modèle ARMA à changement de régime markovien latent. L'étude des propriétés de tels modèles a été initiée par Hamilton [1988]. L'idée consiste à ajouter un niveau d'incertitude supplémentaire aux modèles ARMA afin d'autoriser le traitement des séries chronologiques localement stationnaires. Ces nouveaux modèles sont en fait des modèles ARMA dont les valeurs des paramètres peuvent changer suivant un processus stochastique markovien non observé. Ils sont très utilisés en économie et en traitement du signal où l'hypothèse de stationnarité n'est en général pas vérifiée. Le lecteur peut se référer à McCulloch and Tsay [1994] pour l'étude d'exemples pratiques. Citons également Esseblar [2003] pour l'utilisation de ce type de modèle à la modélisation des flux touristiques avec conjoncture non observée.

Le problème de l'inférence bayésienne pour de tels modèles a été étudié par Albert and Chib [1993], Chib and Greenberg [1994] et Billio *et al.* [1999]. Ces différents auteurs proposent principalement d'utiliser les méthodes MCMC afin d'obtenir un échantillon approximativement distribué suivant la loi *a posteriori* des paramètres.

3.1. Présentation du modèle

Nous limitons notre étude au modèle AR d'ordre 1 (AR(1)) à sauts de moyenne markoviens. Il s'agit d'un processus stochastique dont la fonction d'autocovariance est celle d'un processus AR(1) et dont l'espérance mathématique peut prendre un nombre fini de valeurs différentes en fonction de l'évolution d'une chaîne de Markov non observée. Nous considérons ici le cas de deux espérances mathématiques différentes. Il est possible de généraliser cette étude à un nombre quelconque connu d'espérances. Nous supposons ainsi que nous observons une réalisation des variables aléatoires y_1, \dots, y_n telles que :

$$y_1 \sim \mathcal{N}(m_{\alpha_1}, \sigma^2) \text{ et pour tout } t \in \{2, \dots, n\}$$

$$y_t \mid y_{t-1}, \alpha_t, \alpha_{t-1} \sim \mathcal{N}(m_{\alpha_t} + \phi(x_{t-1} - m_{\alpha_{t-1}}), \sigma^2)$$

où $(\alpha_t)_{t \geq 0}$ est une chaîne de Markov à deux états $\{0, 1\}$ ayant pour état initial $\alpha_0 = 0$ et pour matrice de transition

$$\begin{bmatrix} p_0 & 1 - p_0 \\ 1 - p_1 & p_1 \end{bmatrix}.$$

Les paramètres $\sigma > 0$, $\phi \in]-1, 1[$, $(m_0, m_1) \in \{(m_0, m_1) \in \mathbb{R}^2 \mid m_0 < m_1\}$ et $(p_0, p_1) \in]0, 1[^2$ sont inconnus. Notons que la contrainte $m_0 < m_1$ est imposée

pour assurer l'identifiabilité du modèle. Nous les considérons comme des variables aléatoires et utilisons les lois *a priori* indépendantes peu informatives suivantes :

$$\pi(m_1, m_0) \propto \exp \left\{ -0.5 (m_0 - m_1)^2 / \zeta^2 \right\} \mathbb{I}_{]0, +\infty[}(m_1 - m_0)$$

$$\pi(\phi, \sigma^2) = 1/\sigma \mathbb{I}_{]-1, 1[}(\phi) \mathbb{I}_{]0, +\infty[}(\sigma)$$

$$\pi(p_0, p_1) = \mathbb{I}_{]0, 1[}(p_0) \mathbb{I}_{]0, 1[}(p_1)$$

où $\zeta > 0$ est un hyperparamètre connu. Le caractère non informatif de cet ensemble de lois *a priori* est due aux modélisations par mesures σ -finies des paramètres m_0, m_1, σ [Mengersen and Robert, 1996].

Le calcul explicite de la loi de probabilité *a posteriori* associée est impossible. Il s'agit donc d'un contexte naturel pour l'emploi des schémas d'approximation MCMC et PMC.

L'approche MCMC retenue est celle de l'échantillonneur de Gibbs, voir Robert and Casella [1999], fondé sur la connaissance des densités conditionnelles des paramètres connaissant les autres, qu'il est donc nécessaire d'évaluer. On peut montrer aisément [Robert and Casella, 1999] que la loi conditionnelle de ϕ sachant toutes les autres variables aléatoires du modèle est une loi normale tronquée sur $] - 1, 1[$ d'espérance

$$\sum_{t=2}^n (y_t - m_{\alpha_t}) (y_{t-1} - m_{\alpha_{t-1}}) / \sum_{t=1}^{n-1} (y_t - m_{\alpha_t})^2$$

et de variance

$$\sigma^2 / \sum_{t=1}^{n-1} (y_t - m_{\alpha_t})^2 .$$

De même, la loi conditionnelle de σ^2 sachant toutes les autres variables aléatoires est une loi Inverse Gamma de paramètres $(n + 1)/2$ et

$$\left((y_1 - m_{\alpha_1})^2 + \sum_{t=2}^n (y_t - m_{\alpha_t} - \phi[y_{t-1} - m_{\alpha_{t-1}}])^2 \right) / 2$$

Pour tout $(i, j) \in \{0, 1\}^2$, notons

$$n_{ij} = \sum_{t=1}^{n-1} \mathbb{I}_i(\alpha_t) \mathbb{I}_j(\alpha_{t+1})$$

où $\mathbb{I}_a(b)$ égale 1 si $a = b$ et 0 sinon. n_{ij} désigne alors le nombre de transitions de i vers j de la chaîne de Markov $(\alpha_t)_{t \geq 0}$.

Les lois conditionnelles de p_0 et p_1 sont les lois Betas suivantes [Robert and Casella, 1999] :

$$p_0 | \dots \sim \mathcal{B}(n_{00} + 1, n_{01} + 1), \quad p_1 | \dots \sim \mathcal{B}(n_{11} + 1, n_{10} + 1).$$

Pour tout $(i, j) \in \{0, 1\}^2$, notons

$$s_{ij} = \sum_{t=1}^{n-1} \mathbb{I}_i(\alpha_t) \mathbb{I}_j(\alpha_{t+1}) y_t \quad \text{et} \quad c_{ij} = \sum_{t=1}^{n-1} \mathbb{I}_i(\alpha_t) \mathbb{I}_j(\alpha_{t+1}) y_{t+1}$$

s_{ij} correspond alors à la somme des y_t lorsqu'en t on effectue une transition de i vers j et c_{ij} correspond à la somme des y_{t+1} lorsqu'en t on effectue une transition de i vers j .

La loi conditionnelle de m_0 [Robert and Casella, 1999] est une loi normale tronquée sur $] - \infty, m_1[$ de moyenne

$$\frac{(1-\phi)(c_{00} - \phi s_{00}) + \phi(n_{01}m_1 - c_{01} + \phi s_{01}) + \phi n_{10}m_1 + c_{10} - \phi s_{10} + \zeta^2 \sigma^2 m_1}{n_{00}(1-\phi^2) + n_{01}\phi^2 + n_{10} + \zeta^2 \sigma^2}$$

et de variance

$$\frac{\sigma^2}{n_{00}(1-\phi^2) + n_{01}\phi^2 + n_{10} + \zeta^2 \sigma^2}.$$

De même, la loi conditionnelle de m_1 est une loi normale tronquée sur $]m_0, +\infty[$ de moyenne

$$\frac{(1-\phi)(c_{11} - \phi s_{11}) + \phi(n_{10}m_0 - c_{10} + \phi s_{10}) + \phi n_{01}m_0 + c_{01} - \phi s_{01} + \zeta^2 \sigma^2 m_0}{n_{11}(1-\phi^2) + n_{10}\phi^2 + n_{01} + \zeta^2 \sigma^2}$$

et de variance

$$\frac{\sigma^2}{n_{11}(1-\phi^2) + n_{10}\phi^2 + n_{01} + \zeta^2 \sigma^2}.$$

Par ailleurs, pour tout $t \in \{2, \dots, n-1\}$ et pour $k \in \{0, 1\}$

$$\begin{aligned} \mathbb{P}(\alpha_t = k | \dots) \propto \exp\left(-0.5/\sigma^2 \left((y_t - m_k - \phi(y_{t-1} - m_{\alpha_{t-1}}))^2 + \right. \right. \\ \left. \left. (y_{t+1} - m_{\alpha_{t+1}} - \phi(y_t - m_k))^2 \right) \right) \times \end{aligned}$$

$$(p_{\alpha_{t-1}} \mathbb{I}_k(\alpha_{t-1}) + (1-p_{\alpha_{t-1}}) \mathbb{I}_k(1-\alpha_{t-1})) \times (p_k \mathbb{I}_k(\alpha_{t+1}) + (1-p_k) \mathbb{I}_k(1-\alpha_{t+1})).$$

Pour $k \in \{0, 1\}$, on a :

$$\mathbb{P}(\alpha_1 = k | \dots) \propto \exp\left(-0.5/\sigma^2 (y_1 - m_k)^2 + (y_2 - m_{\alpha_2} - \phi(y_1 - m_k))^2\right) \times \\ p_k \times (p_k \mathbb{I}_k(\alpha_2) + (1 - p_k) \mathbb{I}_k(1 - \alpha_2)).$$

Enfin, pour $k \in \{0, 1\}$, on a :

$$\mathbb{P}(\alpha_n = k | \dots) \propto \exp\left(-0.5/\sigma^2 (y_n - m_k - \phi(y_{n-1} - m_{\alpha_{n-1}}))^2\right) \times \\ (p_{\alpha_{n-1}} \mathbb{I}_k(\alpha_{n-1}) + (1 - p_{\alpha_{n-1}}) \mathbb{I}_k(1 - \alpha_{n-1})).$$

En utilisant toutes ces lois conditionnelles, il est donc possible de mettre en œuvre l'échantillonneur de Gibbs et le schéma PMC tirant profit de la structure latente.

3.2. Présentation des résultats

Deux jeux de données simulées ont été utilisés pour mettre en avant l'efficacité de l'approche PMC. Le premier, nommé (J1), est constitué d'une série de 500 réalisations d'un processus AR(1) à sauts pour lequel $\sigma^2 = 0,01$, $\phi = 0,1$, $m_0 = 0$, $m_1 = 1$, $p_0 = 0,9$ et $p_1 = 0,75$. La figure 1 présente les données simulées (J1) ainsi que la chaîne de Markov cachée (α_t). La variance σ^2 et l'autocorrélation ϕ étant faibles, ce modèle AR(1) à sauts est assez proche d'un simple modèle de changement de moyennes. Aussi, il semble relativement aisé de fournir de bonnes estimations des paramètres, tout au moins en ce qui concerne les deux moyennes m_0 et m_1 .

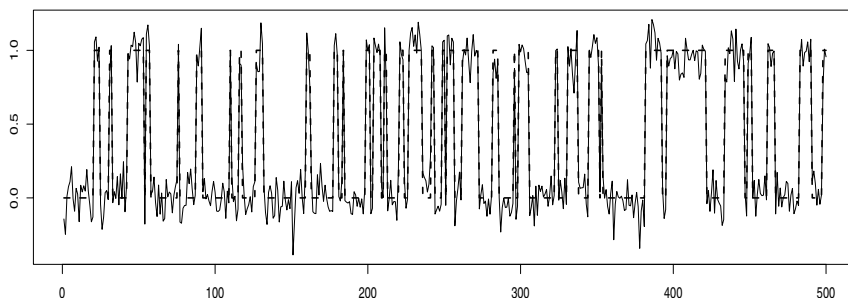


FIGURE 1

Représentation graphique des données simulées (J1) (en pointillés, on a la chaîne de Markov cachée)

Dans un premier temps, nous avons utilisé l'échantillonneur de Gibbs afin d'obtenir un échantillon approximativement distribué suivant la loi *a posteriori* des paramètres. La figure 2 présente l'évolution au cours des itérations des allocations

de l'échantillonneur de Gibbs pour les différents paramètres. La lecture de ce graphique nous laisse penser qu'il est nécessaire d'attendre environ 10 000 itérations avant que la chaîne de Markov générée ne converge vers sa mesure stationnaire, *i.e.* la loi *a posteriori* des paramètres. Une série d'histogrammes de différents échantillons simulés au-delà de la 10 000^{ème} itération confirme ce sentiment. La mesure stationnaire est bien atteinte à l'issue de 10 000 itérations. Ce temps de chauffe peut paraître relativement long dans la mesure où, comme nous l'avons mentionné ci-dessus, l'exemple traité semble facile. Cependant, à l'instar de ce que nous avons observé dans le cas d'un mélange de deux normales [Celeux *et al.*, 2003], les 6 000 premières simulations Gibbs peuvent stagner sur un mode local invisible de la distribution *a posteriori*. Il existe en effet une forme de compensation entre les distributions *a posteriori* des paramètres ϕ et σ^2 . La figure 3 présente, pour chaque paramètre, l'évolution de la moyenne empirique cumulée pour les 10 000 dernières simulations Gibbs, ces quantités estiment les différentes moyennes *a posteriori*, règle bayésienne d'estimation que nous avons choisie. Le temps de chauffe étant passé, on observe très clairement une stabilisation rapide et les estimations sont excellentes.

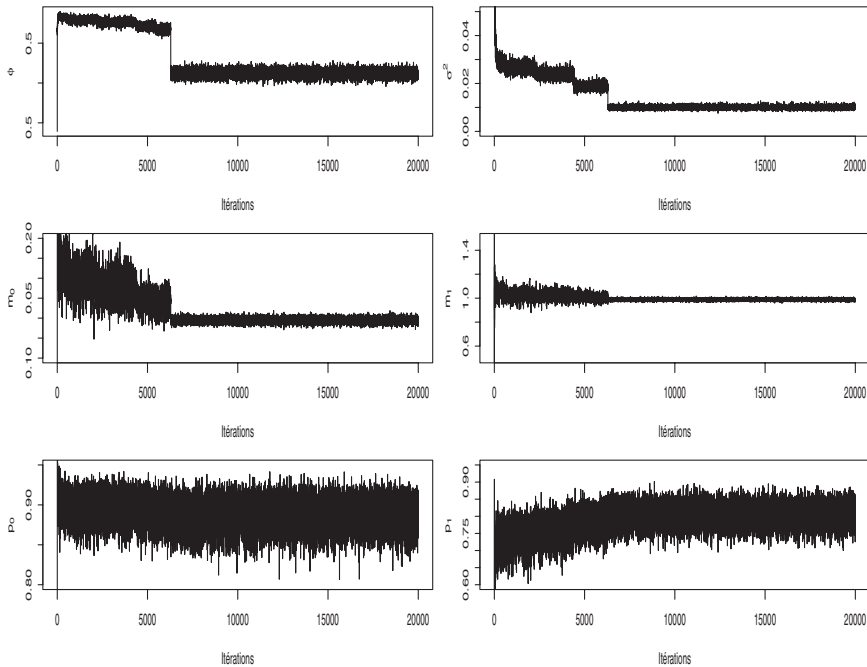


FIGURE 2

Évolution au cours des itérations des différentes allocations de l'échantillonneur de Gibbs pour les données (J1)

Nous allons maintenant comparer ces différents résultats avec ceux obtenus en utilisant le schéma PMC. Un point important dans la comparaison des deux approches est qu'elles visent évidemment toutes deux à fournir des approximations

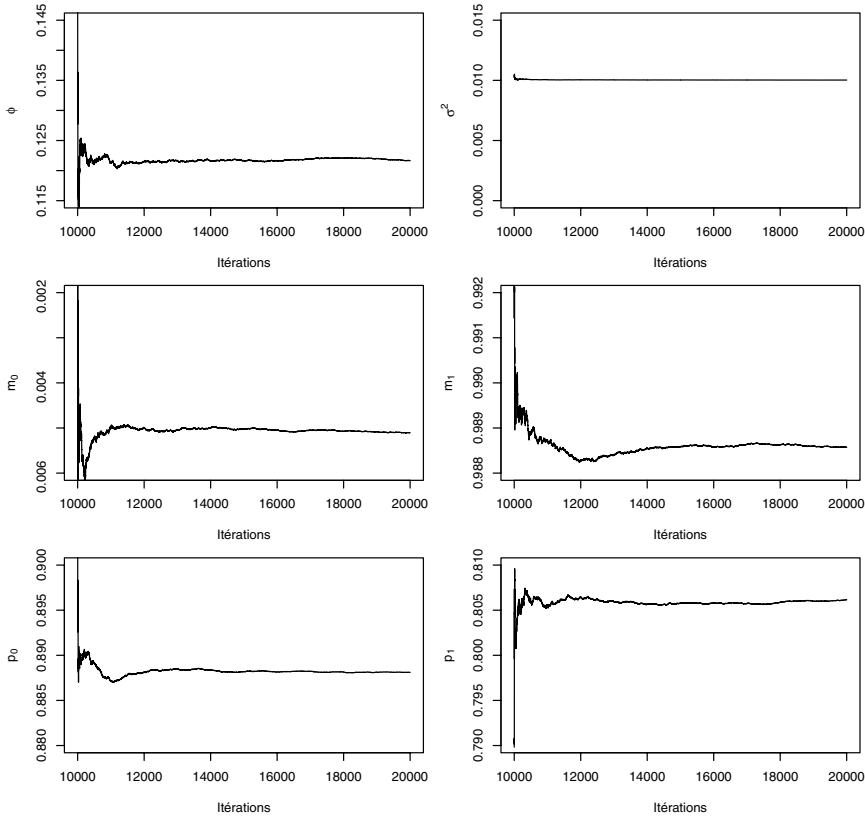


FIGURE 3

Évolution au cours des itérations des estimateurs Gibbs des moyennes a posteriori pour les données simulées (J1)

de la même loi *a posteriori*. En principe, un nombre suffisant d'itérations devrait produire des résultats similaires. Nous avons itéré 10 fois le schéma PMC en simulant 500 échantillons pour chaque itération, soit $T = 10$ et $M = 500$ avec les notations du paragraphe (2.1). La figure 4 présente après 10 itérations de l'algorithme PMC, les histogrammes pondérés des simulations pour chaque paramètre. Ces histogrammes sont des estimations des lois *a posteriori* des paramètres, ils sont très similaires à ceux obtenus en utilisant l'échantillonneur de Gibbs après le temps de chauffe (ils ne sont pas fournis ici). La figure 5 présente l'évolution au cours des 5 dernières itérations des moyennes empiriques pondérées cumulées des échantillons PMC. Ces quantités estiment les différentes moyennes *a posteriori*. On constate que ces estimations sont très stables et sont très similaires à celles obtenues en utilisant l'échantillonneur de Gibbs. En définitive, pour un nombre de simulations moindre, nous avons obtenu d'excellentes approximations avec le schéma PMC.

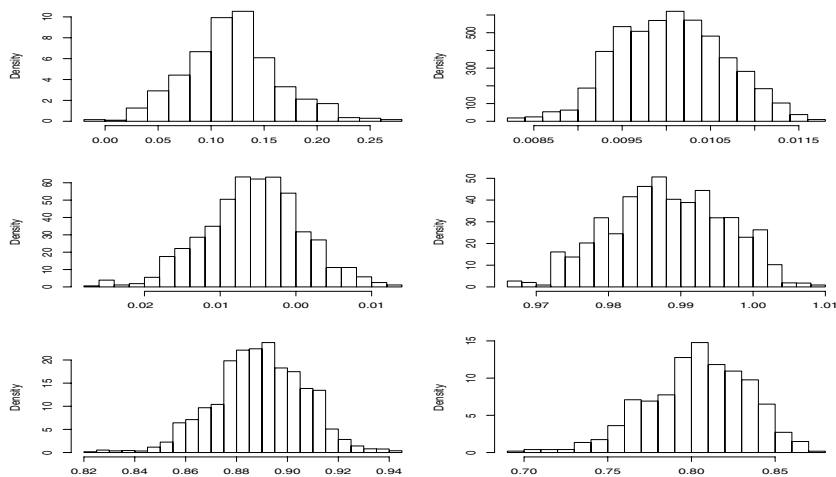


FIGURE 4

Après 10 itérations de l'algorithme, histogrammes pondérés des échantillons PMC pour les données simulées (J1), en haut pour ϕ et σ^2 (de gauche à droite), au milieu pour m_0 et m_1 et en bas pour p_0 et p_1

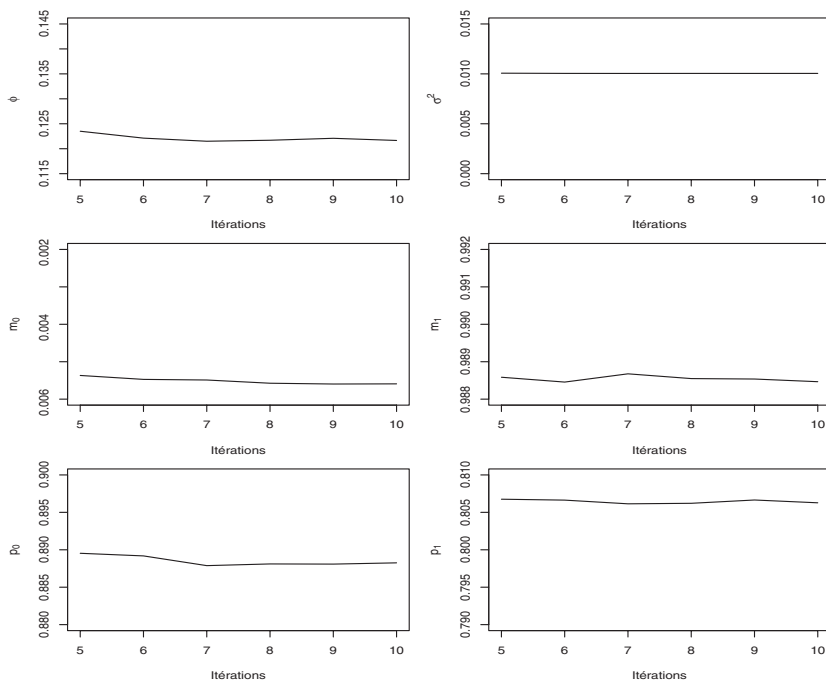


FIGURE 5

Évolution au cours des itérations des estimateurs PMC des moyennes a posteriori pour les données simulées (J1)

Le deuxième jeu de données, nommé (J2), est constitué d'une série de 500 réalisations d'un processus AR(1) à sauts pour lequel $\sigma^2 = 0.01$, $\phi = 0.8$, $m_0 = 0$, $m_1 = 0.2$, $p_0 = 0.9$ et $p_1 = 0.8$. La figure 6 présente les données simulées (J2) ainsi que la chaîne de Markov cachée (α_t). L'autocorrélation ϕ étant importante et les moyennes m_0 et m_1 très proches, il semble ici relativement difficile de fournir de bonnes estimations des paramètres.

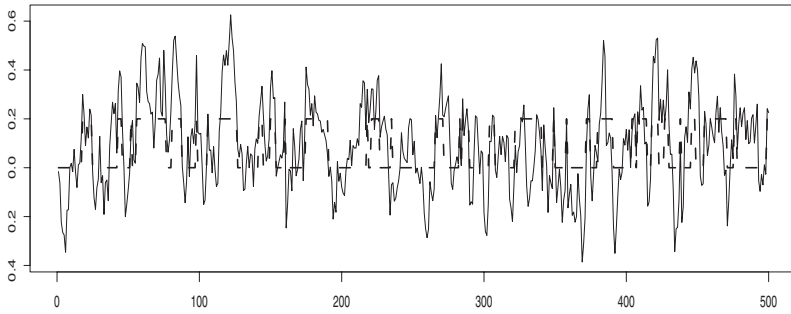


FIGURE 6

*Représentation graphique des données simulées (J2)
en pointillés, on a la chaîne de Markov cachée*

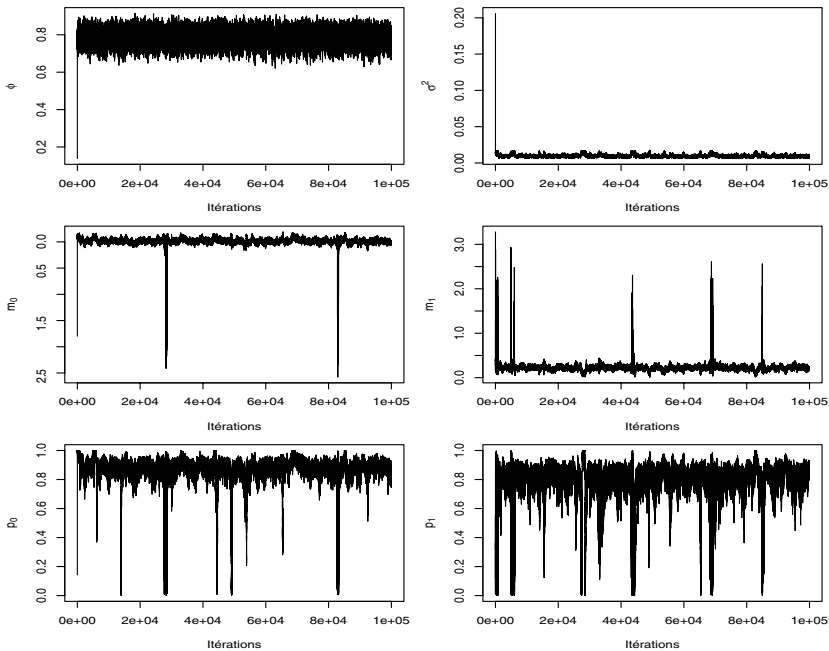


FIGURE 7

*Évolution au cours des itérations des différentes allocations
de l'échantillonneur de Gibbs pour les données simulées (J2)*

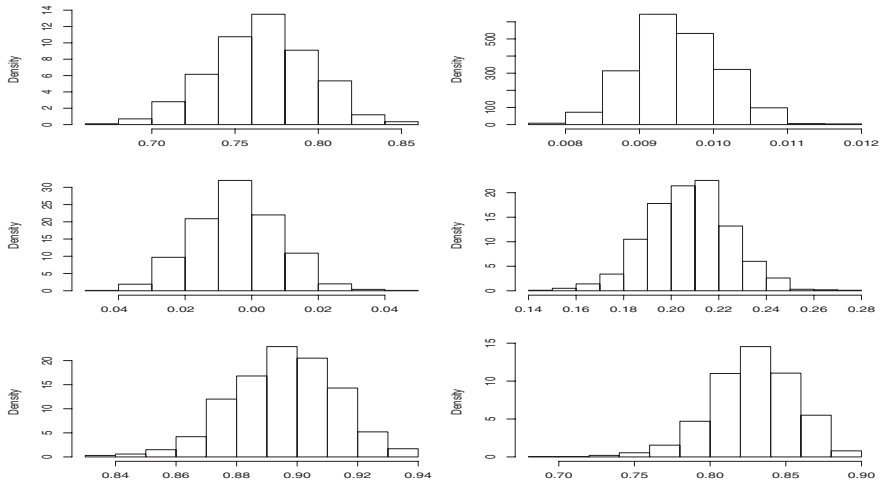


FIGURE 8

Après 10 itérations de l'algorithme, histogrammes pondérés des échantillons PMC pour les données simulées (J2), en haut pour ϕ et σ^2 (de gauche à droite), au milieu pour m_0 et m_1 et en bas pour p_0 et p_1

À travers la figure 7, on constate qu'après 100 000 itérations de l'échantillonneur de Gibbs, la chaîne de Markov générée a beaucoup de mal à cerner les valeurs de certains paramètres, le mixage n'est pas bon. La figure 8 met, quant à elle, en évidence les résultats obtenus avec l'approche PMC après seulement 10 itérations et 1 000 simulations par itération. Ils sont excellents, les vrais valeurs des paramètres sont parfaitement cernées.

4. Conclusion

Dans cet article, nous avons montré que le schéma PMC pouvait être une alternative crédible aux méthodes MCMC. Pour ce faire, nous avons utilisé le modèle AR(1) à sauts de moyenne markovien latents, pour lequel les méthodes de MCMC ne fournissent pas des approximations satisfaisantes. Pour un tel modèle, en utilisant deux jeux de données simulées, nous avons constaté, qu'après un nombre de simulations relativement réduit, le schéma PMC fournit d'excellentes approximations des espérances *a posteriori* des paramètres. Il est clair que la nature itérative du schéma PMC fait disparaître la dépendance naturelle du schéma d'échantillonnage préférentiel classique à la fonction d'importance.

Un point encore mal compris est le temps d'arrêt de l'algorithme PMC. À ce stade et faute de mieux, on privilégie une approche empirique. Une étude théorique permettant de déterminer un temps d'arrêt basé sur des critères mathématiquement objectifs est en cours et doit être approfondie.

Références

- ALBERT J. H. and CHIB S. (1993), Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics*, 1, 1–15.
- BILLIO M., MONFORT A. and ROBERT C. P. (1999), Bayesian estimation of switching ARMA models, *Journal of econometrics*, 93, 229–255.
- CAPPÉ O., GUILLIN A., MARIN J. M. and ROBERT C. P. (2004), Population Monte Carlo, *Journal of Computational and Graphical Statistics*, 13, 907-929.
- CELEUX G., MARIN J. M. and ROBERT C. P. (2003), Iterated importance sampling in missing data problems, *Cahiers du Ceremade*, Université Paris Dauphine.
- CHIB S. and GREENBERG E. (1994), Bayes inference for regression models with ARMA(p,q) errors, *Journal of Econometrics*, 64, 183–206.
- DEMPSTER A. P., LAIRD N. M. and RUBIN D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society Series B*, 39, 1–38.
- ESSEBBAR B. (2004), Sur un modèle de série temporelle à conjoncture dans le tourisme, *Revue de Statistique Appliquée*, 52, n° 4, 55–70.
- GELFAND A. E. and SMITH A. F. M. (1990), Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, 398–409.
- HAMILTON J. D. (1988), Rational-expectations econometric analysis of changes in regime : An investigation of the term structure of interest rates, *Journal of Economic Dynamics and Control*, 12, 385–423.
- HAMILTON J. D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycles, *Econometrica*, 57, 357–384.
- LITTLE R. J. A. and RUBIN D. B. (1987), *Statistical Analysis with Missing Data*, J. Wiley, New York.
- MCCULLOCH C. E. and TSAY R. (1994), Statistical analysis of macroeconomic time series via Markov switching models, *Journal of time series analysis*, 15, 523–539.
- MCLACHLAN G. J. and KRISHNAN T. (1997), *The EM Algorithm and Extensions*, J. Wiley, New York.
- MENGERSEN K.L. and ROBERT C.P. (1996), Testing for mixtures : A Bayesian entropic approach (with discussion), in J.O. Berger, J.M. Bernardo, A.P. Dawid, D.V. Lindley and A.F.M. Smith, editors, *Bayesian Statistics 5*, 255–276, Oxford University Press, Oxford.
- RIPLEY B. D. (1987), *Stochastic Simulation*, J. Wiley, New York.
- ROBERT C. P. and CASELLA G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.
- RUBINSTEIN R. Y. (1981), *Simulation and the Monte Carlo Method*, J. Wiley, New York.

- TANNER M. A. (1996), *Tools for Statistical Inference : Observed Data and Data Augmentation Methods*, 3 edition, Springer-Verlag, New York.
- TANNER M. A. and WONG W. (1987), The calculation of posterior distribution distributions by data augmentation, (with discussion), *Journal of the American Statistical Association*, 82, 528–550.
- VAN DYK A. and MENG X. L. (2001), The Art of Data Augmentation, (with discussion), *Journal of Computational and Graphical Statistics*, 10, 1–110.
- WARNES G. R. (2001), The normal kernel coupler. An adaptative Markov Chain Monte Carlo method for efficient sampling from multi-mode distributions, 395, Department of Statistics, University of Washington, Seattle.