The Canadian Journal of Statistics
Vol. 28, No. 4, 2000, Pages ???-???
La revue canadienne de statistique

Monte Carlo Kalman filter and smoothing for multivariate discrete state space models

Peter Xue-Kun SONG

Key words and phrases: Binomial time series; Kalman filter; Monte Carlo; smoothing; state space model.
AMS 1991 subject classifications: 62M10, 62E25.

ABSTRACT

The author studies state space models for multivariate binomial time series, focussing on the development of the Kalman filter and smoothing for state variables. He proposes a Monte Carlo approach employing the latent variable representation which transplants classical Kalman filter and smoothing developed for Gaussian state space models to discrete models and leads to a conceptually simple and computationally convenient approach. The method is illustrated through simulations and concrete examples.

RÉSUMÉ

L'auteur s'intéresse aux modèles à espace d'états pour des chroniques binomiales multivariées, et plus particulièrement au développement du filtre de Kalman et au lissage pour les variables d'état. Il propose une approche de Monte-Carlo exploitant la représentation des variables latentes, laquelle permet d'adapter aux modèles discrets le filtre de Kalman et le lissage classiques des modèles à espace d'états gaussiens, ce qui conduit à une approche simple, tant au plan conceptuel que calculatoire. La méthode est illustrée au moyen de simulations et d'exemples concrets.

1. INTRODUCTION

Assume $\{\mathbf{y}_t = (y_{1t}, \ldots, y_{dt})',\ t = 1, \ldots, n\}$ is a d-dimensional time series of binomial responses with the vector of probabilities of success $\boldsymbol{\pi}_t = (\pi_{1t}, \ldots, \pi_{dt})'$. The multivariate discrete state space model consists of two equations, namely the observation equation and the state equation, respectively given by

$$\pi_{it} = F\left(\mathbf{x}_{it}'\boldsymbol{\theta}_t\right),\ i = 1, \ldots, d, \quad \text{and} \tag{1}$$

$$\boldsymbol{\theta}_t = B_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\xi}_t \tag{2}$$

where $\mathbf{x}_{it}$ is $p \times 1$ vector of covariates, B_t is a $p \times p$ transition matrix and $\boldsymbol{\xi}_t$ is the p-variate Gaussian white noise with zero mean and covariance matrix Q_t. For mathematical convenience, in the paper we choose the link function F to be the standard Gaussian distribution function Φ, (1) corresponding to the probit model.

This paper focusses on the development of the Kalman filter and smoothing (KFS) based on conditional mean estimates of the states $\boldsymbol{\theta}_t$ that are minimum mean square error estimates. Such methods have gained popularity because they allow calculation of conditional mean estimates (i.e., the optimal estimates) in normal linear systems in closed form; cf., e.g., Harvey (1981). In general, the Kalman filter corresponds to a recursive procedure of estimating $\boldsymbol{\theta}_t$ given observations $\mathbf{y}_1, \ldots, \mathbf{y}_t$, whereas the smoothing involves estimating $\boldsymbol{\theta}_t$ based on all observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$. When state space models are non-linear and non-normal, exact conditional mean estimates of the states are not available. Many alternative solutions to this problem have been proposed in the literature such as the Markov chain Monte Carlo (MCMC) methods of Carlin *et al.* (1992)to obtain full posterior inference, the posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models (Fahrmeir 1992), and Kalman EM algorithm for multivariate state space models for time series of counts (Jørgensen *et al.* 1999). See also Kitagawa (1987), Naik-Nimbalkar & Rajarshi (1995), Carlin & Polson (1992) and Jørgensen *et al.* (1996).

Methods such as MCMC based on a full Bayesian analysis are now widely used, including in state space models, where the estimates of state variables are obtained from the corresponding posterior densities. It is well-known that the Bayesian approach involves calculation of high-dimensional integration which is usually intricate, and in some cases the full MCMC implementation may be unnecessarily burdensome. As an alternative to the computationally intensive Bayesian inference, this paper follows on the conventional line of the Kalman filter and smoothing (e.g., Brockwell & Davis 1987) and develops a conceptually simple and computationally convenient approach which provides point estimates of the state variables that may be as close to the minimum mean square error estimates as desired.

The idea for such development is rooted in the use of latent variable representation for probit models, which leads to linear and Gaussian state space models for which the linear Kalman filter and smoothing are available in terms of the latent variables. The Monte Carlo technique is applied to approximate the conditional mean estimates of the states by averaging sample linear Kalman filters and smoothings based on simulated latent variables. By the law of large number, the approximate estimates converge to the exact conditional mean estimates as the Monte Carlo sample size tends to infinity, since i.i.d. samples directly drawn from the given distributions are indeed used in the averaging procedure. As a matter of fact, the study of burn-in that is a necessary but very time-consuming step in MCMC methods is not needed in our method where every sample is valid.

The model specification is discussed in Section 2, and Section 3 presents the Monte Carlo Kalman filter and smoother with the corresponding mean square errors. Simulations are reported in Section 4, and two examples of data analysis are included in Section 5. Some concluding remarks are given in Section 6.

2. MODELS

2.1. Multivariate Binary Probit Model.

Consider a d-dimensional binary time series $(\mathbf{y}_t)$ with probability vector $\boldsymbol{\pi}_t$. Denote the full data vector and the full state variable vector by

$$\mathbf{Y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')' \quad \text{and} \quad \boldsymbol{\theta} = (\boldsymbol{\theta}_0', \boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_n')',$$

respectively, where $\boldsymbol{\theta}_0$ is an initializing state vector. Assume that conditional on the full vector of states $\boldsymbol{\theta}$, the dn components of $\mathbf{Y}$ are independent, and that the marginal expectation follows the probit model given by

$$\pi_{it} = P(y_{it} = 1|\boldsymbol{\theta}_t) = \Phi(\mathbf{x}_{it}'\boldsymbol{\theta}_t), \quad t = 1, \ldots, n, \tag{3}$$

where the state variables $\boldsymbol{\theta}_t$ are governed by model (2). Also assume the initial state $\boldsymbol{\theta}_0 \sim N_p(\mathbf{a}_0, Q_0)$ where both $\mathbf{a}_0$ and Q_0 are known.

Let $\{\mathbf{z}_t = (z_{1t}, \ldots, z_{dt})'\}$ be a d-dimensional latent process satisfying

$$\mathbf{z}_t = X_t'\boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, \tag{4}$$

where $X_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{dt})$ is a $p \times d$ matrix and $\boldsymbol{\varepsilon}_t$ are i.i.d. $N_d(\mathbf{0}, I)$ Gaussian innovations, and for the i-th component, define

$$y_{it} = 1 \ \text{ if and only if } \ z_{it} \geq 0, \quad i = 1, \ldots, d. \tag{5}$$

Therefore, as desired, the latent variable representation results in

$$\pi_{it} = P(z_{it} \geq 0|\boldsymbol{\theta}_t) = \Phi(\mathbf{x}_{it}'\boldsymbol{\theta}_t), \quad i = 1, \ldots, d.$$

Here we assume the same sort of conditional independence for the latent process as that for the observed process, i.e., the dn components of $\mathbf{Z} = (\mathbf{z}_1', \ldots, \mathbf{z}_n')'$ are conditionally independent given $\boldsymbol{\theta}$. In addition, we assume the sequences of $(\boldsymbol{\varepsilon}_t)$ and $(\boldsymbol{\xi}_t)$ are mutually independent, and both sequences are independent of $\boldsymbol{\theta}_0$.

Clearly models (4) and (2) together form a linear and Gaussian state space model for which the optimal linear Kalman filter and smoothing are available in terms of the latent process $(\mathbf{z}_t)$. They are the minimum mean square error estimates of the states and could be computed using the standard recursive procedure (e.g., Brockwell & Davis 1987) if the latent vectors were known. For convenience, such a state space model is called *the interim model* in the rest of the paper.

2.2. Moments.

We now give some results regarding moments for the interim models, useful for the derivation of sampling distribution in the implementation of Monte Carlo method in Section 3. Clearly $\mathbf{Z}$ follows a dn-variate Gaussian distribution with mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_n')'$ and covariance matrix $\Sigma = (\Sigma_{st})$, where $\boldsymbol{\mu}_t = (\mu_{1t}, \ldots, \mu_{dt})'$ and Σ_{st} is $d \times d$ covariance matrix of $\text{cov}(\mathbf{z}_s, \mathbf{z}_t)$, $s, t = 1, \ldots, n$. It follows immediately from the model specification that $\mu_{it} = E(z_{it}) = \mathbf{x}_{it}'E(\boldsymbol{\theta}_t)$ where $E(\boldsymbol{\theta}_t) = B_t \cdots B_1\mathbf{a}_0$. The block-diagonals of the covariance matrix Σ are equal to

$$\Sigma_{tt} = \text{var}(\mathbf{z}_t) = I + X_t'\text{var}(\boldsymbol{\theta}_t)X_t,$$

where

$$\begin{aligned}\text{var}(\boldsymbol{\theta}_t) = Q_t + B_tQ_{t-1}B_t' + (B_tB_{t-1})Q_{t-2}(B_tB_{t-1})' + \cdots +\\ (B_tB_{t-1}\cdots B_2)Q_1(B_tB_{t-1}\cdots B_2)' + (B_tB_{t-1}\cdots B_1)Q_0(B_tB_{t-1}\cdots B_1)'.\end{aligned}$$

The off-block-diagonals are given by

$$\Sigma_{t,t+s} = X_t'\text{cov}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+s})X_{t+s},$$

where

$$\text{cov}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+s}) = \text{var}(\boldsymbol{\theta}_t)(B_{t+s}\cdots B_{t+1})'.$$

Clearly, the Gaussianity of the latent process $(\mathbf{z}_t)$ and that of state variables $(\boldsymbol{\theta}_t)$ imply that the joint distribution of $(\mathbf{Z}, \boldsymbol{\theta})$ is Gaussian, and so is the marginal distribution of $\mathbf{Z}$, with its mean vector and covariance matrix given above.

2.3. Binomial Probit Model.

Let $(\mathbf{y}_t)$ be d-dimensional binomial responses with the i-th component following as binomial distribution $B(k_{it}, \pi_{it})$ and the marginal probability π_{it} following the probit model $\pi_{it} = \Phi(\mathbf{x}_{it}'\boldsymbol{\theta}_t)$. For simplicity, assume $k_{it} = k$ for all i and t. Now decompose y_{it} into an independent sum of binary variables, namely $y_{it} = y_{it1} + \cdots + y_{itk}$, where $y_{it1}, \ldots, y_{itk}$ are i.i.d. binary variables with probability π_{it}. Following Tanner & Wong (1987), draw k i.i.d. d-variates $\mathbf{z}_{tj} = (z_{1tj}, \ldots, z_{dtj})'$, $j = 1, \ldots, k$ from the normal distribution $N_d(X_t'\boldsymbol{\theta}_t, I)$ conditional on $\boldsymbol{\theta}_t$, implying $z_{itj}|\boldsymbol{\theta}_t \sim N(\mathbf{x}_{it}'\boldsymbol{\theta}_t, 1)$ where $X_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{dt})$ is a $p \times d$ matrix, and define

$$y_{itj} = 1 \ \text{ if and only if } \ z_{itj} \geq 0, \quad j = 1, \ldots, k.$$

It follows that

$$\pi_{it} = P(y_{itj} = 1|\boldsymbol{\theta}_t) = P(z_{itj} \geq 0|\boldsymbol{\theta}_t) = \Phi(\mathbf{x}_{it}'\boldsymbol{\theta}_t), \quad j = 1, \ldots, k.$$

Let $\mathbf{z}_t = (\mathbf{z}_{t1}', \ldots, \mathbf{z}_{tk}')'$ with $\mathbf{z}_{tj} = (z_{1tj}, \ldots, z_{dtj})'$.

$$\mathbf{z}_t = \mathbf{X}_t'\boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \tag{6}$$

where $\mathbf{X}_t = (X_t, \ldots, X_t)$ is a $p \times dk$ matrix and $\boldsymbol{\varepsilon}_t$ are independent innovations with $N_{dk}(\mathbf{0}, I)$. As a result, the interim state space model consisting of (6) and (2) is apparently linear and Gaussian, similarly to those obtained in Section 2.1.

3. MONTE CARLO KALMAN FILTER AND SMOOTHING

This section concerns with the development of the Kalman filter and smoothing for the binary probit models, and that for the binomial probit models can be obtained in an almost identical way and hence details are omitted.

3.1. Filter and Smoothing.

Let $\mathbf{Y}^t = (\mathbf{y}_1', \ldots, \mathbf{y}_t')'$, $\mathbf{Y} = \mathbf{Y}^n$, and $\mathbf{Z}^t = (\mathbf{z}_1', \ldots, \mathbf{z}_t')'$, $\mathbf{Z} = \mathbf{Z}^n$. The central task is to compute the two conditional mean estimates, $E(\boldsymbol{\theta}_t|\mathbf{Y}^t)$ and $E(\boldsymbol{\theta}_t|\mathbf{Y})$, respectively, given the information available up to time t and all information.

It follows from (5) that $E(\boldsymbol{\theta}_t|\mathbf{Y}^s, \mathbf{Z}^s) = E(\boldsymbol{\theta}_t|\mathbf{Z}^s)$, for all $t, s = 1, \ldots, n$. Hence we have

$$E\left(\boldsymbol{\theta}_t|\mathbf{Y}^t\right) = E\left\{E\left(\boldsymbol{\theta}_t|\mathbf{Y}^t, \mathbf{Z}^t\right)|\mathbf{Y}^t\right\} = E\left\{E\left(\boldsymbol{\theta}_t|\mathbf{Z}^t\right)|\mathbf{Y}^t\right\} = E\left(\Theta_t|\mathbf{Y}^t\right), \tag{7}$$

and similarly,

$$E(\boldsymbol{\theta}_t|\mathbf{Y}^n) = E\{E(\boldsymbol{\theta}_t|\mathbf{Z}^n)|\mathbf{Y}^n\} = E(\Theta_t^*|\mathbf{Y}^n), \tag{8}$$

where $\Theta_t = E(\boldsymbol{\theta}_t|\mathbf{Z}^t)$ and $\Theta_t^* = E(\boldsymbol{\theta}_t|\mathbf{Z}^n)$ are the conditional expectations of $\boldsymbol{\theta}_t$ with respect to the latent process $\{\mathbf{z}_t\}$, respectively.

If $\mathbf{z}_t$'s were observed, both Θ_t and Θ_t^* would be computed recursively as follows:

1. Filter prediction step

$$\Theta_{t|t-1} = E(\boldsymbol{\theta}_t|\mathbf{Z}^{t-1}) = B_t\Theta_{t-1}, \quad \Theta_0 = \mathbf{a}_0,$$

$$\Lambda_{t|t-1} = B_t\Lambda_{t-1}B_t' + Q_t, \quad \Lambda_0 = Q_0.$$

2. Filter correction step

$$\Theta_t = \Theta_{t|t-1} + \Lambda_{t|t-1}X_t\Delta_t^{-1}\left(\mathbf{z}_t - X_t'\Theta_{t|t-1}\right),$$

$$\Lambda_t = \Lambda_{t|t-1} - \Lambda_{t|t-1}X_t\Delta_t^{-1}X_t'\Lambda_{t|t-1},$$

where $\Delta_t = X_t'\Lambda_{t|t-1}X_t + I$.

3. Smoothing step

$$\Theta_t^* = E(\boldsymbol{\theta}_t|\mathbf{Z}) = \Theta_t + P_t\left(\Theta_{t+1}^* - B_{t+1}\Theta_t\right),$$

$$\Lambda_t^* = \Lambda_t + P_t\left(\Lambda_{t+1}^* - \Lambda_{t+1|t}\right)P_t'$$

where $P_t = \Lambda_t B_{t+1}'\Lambda_{t+1|t}^{-1}$, $t = n-1, \ldots, 1$. At time n, $\Theta_n^* = \Theta_n$ and $\Lambda_n^* = \Lambda_n$.

We now apply the Monte Carlo technique to approximate both conditional mean estimates of the states based on $(\mathbf{y}_t)$, using equations (7) and (8). For convenience, let $[\mathbf{u}|\mathbf{w}]$ denote the conditional distribution of $\mathbf{u}$ given $\mathbf{w}$.

Suppose $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(M)}$ are M i.i.d. samples generated from $[\mathbf{Z}|\mathbf{Y}]$. For each sample $\mathbf{Z}^{(i)} = (\mathbf{z}_1'^{(i)}, \ldots, \mathbf{z}_n'^{(i)})'$, the Kalman filter and smoothing given by above steps 1–3 produce

$$\left\{\Theta_t^{(i)}, \Lambda_t^{(i)}\right\} \quad \text{and} \quad \left\{\Theta_t^{*(i)}, \Lambda_t^{*(i)}\right\},$$

respectively, $i = 1, \ldots, M$. By the law of large number, $E(\boldsymbol{\theta}_t|\mathbf{Y})$ is then approximated by the average of M smoothers of the form

$$\mathbf{m}_t^* = \frac{1}{M}\sum_{i=1}^{M}\Theta_t^{*(i)}, \quad t = 1, \ldots, n. \tag{9}$$

Equation (9) is called *the Monte Carlo Kalman smoothing* (MCKS). The Monte Carlo approximation of $E(\boldsymbol{\theta}_t|\mathbf{Y}^t)$ needs to draw samples from $[\mathbf{Z}^t|\mathbf{Y}^t]$ for each t and the corresponding estimator may be defined in a way similar to that of MCKS. Given the availability of full samples $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(M)}$, however, a better estimator of $E(\boldsymbol{\theta}_t|\mathbf{Y}^t)$, which has smaller mean square error than the ordinary one, may be obtained by

$$\mathbf{m}_t = \frac{1}{M}\sum_{i=1}^{M}\Theta_t^{(i)}, \quad t = 1, \ldots, n. \tag{10}$$

Equation (10) is referred to as *the Monte Carlo Kalman filter* (MCKF).

Denote the mean vector of $\mathbf{Z}$ by $\boldsymbol{\mu}$ and the covariance matrix by Σ. Refer to Section 2.2 for computation of these parameters. It is known that the conditional distribution $[\mathbf{Z}|\mathbf{Y}]$ is a truncated multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, where the truncated region is a rectangular area specified by $a_j \le z_j \le b_j$, with $a_j = \log(y_j)$ and $b_j = -\log(1 - y_j)$, $j = 1, \ldots, dn$. Generating random variates from truncated multivariate normal distributions has been discussed extensively in the literature; cf., e.g., Geweke (1991) for an algorithm of the multivariate normal random generation subject to linear constraints, and Robert (1995) for the simulation of truncated normal variables via Monte Carlo Markov chain methods. See also Czado (1997) for a successive generation scheme or the sequential random generation of truncated normal. One of these algorithms may be applied to carry out the generation of truncated normal variates.

3.2. Mean Square Error.

We now give the mean square error, $E\left(\boldsymbol{\theta}_t - \mathbf{m}_t^*\right)\left(\boldsymbol{\theta}_t - \mathbf{m}_t^*\right)'$, of the Monte Carlo smoother $\mathbf{m}_t^*$. First we note that

$$\begin{aligned}
\mathrm{cov}(\boldsymbol{\theta}_t - E\boldsymbol{\theta}_t,\, E\boldsymbol{\theta}_t - \mathbf{m}_t^*) &= \frac{1}{M}\sum_{i=1}^{M}\mathrm{cov}(\boldsymbol{\theta}_t - E\boldsymbol{\theta}_t,\, E\boldsymbol{\theta}_t - \Theta_t^{*(i)}) \\
&= -\frac{1}{M}\sum_{i=1}^{M} E\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)',
\end{aligned}$$

and

$$\begin{aligned}
E(E\boldsymbol{\theta}_t - \mathbf{m}_t^*)(E\boldsymbol{\theta}_t - \mathbf{m}_t^*)' &= \frac{1}{M^2}\sum_{i=1}^{M} E\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)' \\
&\quad + \frac{1}{M^2}\sum_{i\neq j}\mathrm{cov}(E\boldsymbol{\theta}_t - \Theta_t^{*(i)},\, E\boldsymbol{\theta}_t - \Theta_t^{*(j)}) \\
&= \frac{1}{M^2}\sum_{i=1}^{M} E\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)' \\
&\quad + \frac{M-1}{M} E\left\{E\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{E\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}'
\end{aligned}$$

where the last equality is obtained from the fact that given $\mathbf{Y}$, $\mathbf{Z}^{(i)}$ and $\mathbf{Z}^{(j)}$ conditionally independent and identically distributed. It follows immediately that

$$\begin{aligned}
E(\boldsymbol{\theta}_t - \mathbf{m}_t^*)(\boldsymbol{\theta}_t - \mathbf{m}_t^*)' &= \mathrm{var}(\boldsymbol{\theta}_t) + E\left(E\boldsymbol{\theta}_t - \mathbf{m}_t^*\right)\left(E\boldsymbol{\theta}_t - \mathbf{m}_t^*\right)' \\
&\quad + 2\mathrm{cov}(\boldsymbol{\theta}_t - E\boldsymbol{\theta}_t,\, E\boldsymbol{\theta}_t - \mathbf{m}_t^*) \\
&= \mathrm{var}(\boldsymbol{\theta}_t) + \frac{M-2}{M^2}\sum_{i=1}^{M} E\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)\left(E\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)' \\
&\quad + \frac{M-1}{M} E\left\{E\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{E\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}' \\
&= \frac{2M-1}{M^2}\sum_{i=1}^{M} E\left(\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)\left(\boldsymbol{\theta}_t - \Theta_t^{*(i)}\right)' \\
&\quad - \frac{M-1}{M} E\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}' \qquad (11)
\end{aligned}$$

where the last equality is due to

$$E\left\{E(\boldsymbol{\theta}_t|\mathbf{y}) - E\boldsymbol{\theta}_t\right\}\left\{E(\boldsymbol{\theta}_t|\mathbf{y}) - E\boldsymbol{\theta}_t\right\}' = \mathrm{var}(\boldsymbol{\theta}_t) - E\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}'$$

and

$$\begin{aligned}
&E\left\{E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right) - E\boldsymbol{\theta}_t\right\}\left\{E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right) - E\boldsymbol{\theta}_t\right\}' = \\
&\qquad \mathrm{var}(\boldsymbol{\theta}_t) - E\left\{\boldsymbol{\theta}_t - E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right)\right\}\left\{\boldsymbol{\theta}_t - E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right)\right\}'
\end{aligned}$$

It is easy to prove that for each t, when $E(\boldsymbol{\theta}_t|\mathbf{Y}) < \infty$,

$$\int_A f(\mathbf{Z}|\mathbf{Y})d\mathbf{Z} \to 0 \quad \Rightarrow \quad E\left(\mathbf{m}_t^* \mathbf{m}_t^{*\prime} I_A\right) \to 0$$

uniformly for all M where f denotes the conditional density of $\mathbf{Z}$ given $\mathbf{Y}$ and I_A is the indicator of set A. Hence we have

$$E(\boldsymbol{\theta}_t - \mathbf{m}_t^*)(\boldsymbol{\theta}_t - \mathbf{m}_t^*)' \to E\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}', \quad \text{as } M \to \infty.$$

Therefore when M is large, from (11), we obtain an approximate formula given by

$$\begin{aligned} E(\boldsymbol{\theta}_t - \mathbf{m}_t^*)(\boldsymbol{\theta}_t - \mathbf{m}_t^*)' &\approx E\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}\left\{\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t|\mathbf{y})\right\}' \\ &\approx \frac{1}{M}\sum_{i=1}^{M} E\left\{\boldsymbol{\theta}_t - E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right)\right\}\left\{\boldsymbol{\theta}_t - E\left(\boldsymbol{\theta}_t|\mathbf{z}^{(i)}\right)\right\}' \\ &= \frac{1}{M}\sum_{i=1}^{M} \Lambda_t^{*(i)} \qquad (12) \end{aligned}$$

where $\Lambda_t^{*(i)}$ are the mean square errors corresponding to the Kalman smoother $\Theta_t^{*(i)}$ available in terms of M samples $\mathbf{Z}^{(i)}$, $i = 1, \ldots, M$.

3.3. Estimation of Q_t.

Assume $Q_t = Q$, independent of t. Let $\boldsymbol{\delta}_t^* = \Theta_t^* - B_t\Theta_{t-1}^*$. It is easy to show that $E\left(\boldsymbol{\delta}_t^*\right) = E(\boldsymbol{\xi}_t) = \mathbf{0}$. Note that

$$\boldsymbol{\delta}_t^* = E(\boldsymbol{\theta}_t - B_t\boldsymbol{\theta}_{t-1}|\mathbf{Z}) = E(\boldsymbol{\xi}_t|\mathbf{Z}).$$

It follows that

$$\begin{aligned} \operatorname{var}\left(\boldsymbol{\delta}_t^*\right) &= Q - \\ &\quad E\left\{(\boldsymbol{\theta}_t - \Theta_t^*) - B_t\left(\boldsymbol{\theta}_{t-1} - \Theta_{t-1}^*\right)\right\}\left\{(\boldsymbol{\theta}_t - \Theta_t^*) - B_t\left(\boldsymbol{\theta}_{t-1} - \Theta_{t-1}^*\right)\right\}' \\ &= Q - \left(\Lambda_t^* + B_t\Lambda_{t-1}^*B_t' - 2\Lambda_{t,t-1}^*B_t'\right), \end{aligned}$$

where

$$\Lambda_{t,t-1}^* = E\left\{\boldsymbol{\theta}_t - \Theta_t^*\right\}\left\{\boldsymbol{\theta}_{t-1} - \Theta_{t-1}^*\right\}' = \Lambda_{t-1}B_t'\Lambda_{t|t-1}^{-1}\Lambda_t^*, \qquad (13)$$

and the derivation of (13) is given in the appendix. Therefore we obtain

$$Q = E\boldsymbol{\delta}_t^*\boldsymbol{\delta}_t^{*\prime} + \left(\Lambda_t^* + B_t\Lambda_{t-1}^*B_t' - 2\Lambda_{t,t-1}^*B_t'\right).$$

Moreover, a consistent estimator of Q may be obtained by iteratively applying the following formula (14) until convergence, with initializing matrix being specified by, for example $Q = I$,

$$\begin{aligned} \widehat{Q} &= \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{m}_t^* - B_t\mathbf{m}_{t-1}^*\right)\left(\mathbf{m}_t^* - B_t\mathbf{m}_{t-1}^*\right)' \\ &\quad + \frac{1}{n}\sum_{t=1}^{n}\left(\overline{\Lambda}_t^* + B_t\overline{\Lambda}_{t-1}^*B_t' - 2\overline{\Lambda}_{t,t-1}^*B_t'\right), \qquad (14) \end{aligned}$$

where $\overline{\Lambda}_t^* = M^{-1}\sum_{i=1}^{M}\Lambda_t^{*(i)}$ and $\overline{\Lambda}_{t,t-1}^* = M^{-1}\sum_{i=1}^{M}\Lambda_{t-1}^{(i)}B_t'\Lambda_{t|t-1}^{(i)-1}\Lambda_t^{*(i)}$.

3.4. Estimation of Autocorrelation.

In connection with the analysis of the infant sleep data presented in Section 5.1, we now consider a one-dimensional AR(1) stationary state process with unknown autocorrelation coefficient α given by

$$\theta_t = \alpha\theta_{t-1} + \xi_t, \tag{15}$$

where the ε_t's are i.i.d. normal innovations with mean zero and variance σ^2. Note that

$$\operatorname{cov}\left(\Theta_t^*, \Theta_{t+1}^*\right) = \operatorname{cov}(\theta_t, \theta_{t+1}) - \Lambda_{t,t+1}^* = \phi\sigma^2 - \Lambda_{t,t+1}^*$$

where $\phi = \alpha/(1-\alpha^2)$. Therefore, ϕ may be consistently estimated by

$$\widehat{\phi} = \frac{1}{n\widehat{\sigma}^2}\sum_{t=1}^{n-1} m_t^* m_{t+1}^* + \frac{1}{n\widehat{\sigma}^2}\sum_{t=1}^{n-1} \overline{\Lambda}_{t,t+1}^*. \tag{16}$$

Moreover a consistent estimate of α is given by

$$\hat{\alpha} = \left(-1 + \sqrt{1+4\hat{\phi}}\right) \Big/ 2\hat{\phi}. \tag{17}$$

4. A SIMULATION EXPERIMENT

This section demonstrates some numerical practice of the Monte Carlo Kalman filter and smoothing through a simulation study based on the binary probit model. Similar to Fahrmeir (1992), consider a binary probit model given by

$$\pi_t = \Phi\left(\theta_{t1} + x_t\theta_{t2}\right), \quad \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\xi}_t, \quad t = 1, \ldots, 200,$$

with $\boldsymbol{\theta}_t = (\theta_{t1}, \theta_{t2})'$ and $\boldsymbol{\xi}_t \sim N_2(\mathbf{0}, 0.005I)$. The covariate sequence (x_t) was drawn from the standard normal and held fixed throughout the simulation study. The latent process (z_t), generated by the latent variable representation (4), determines the binary process (y_t) in light of (5). Using this true underlying latent process, which is indeed known in the simulation setting, we can obtain the "ideal" Kalman filter and smoothing for the states, denoted by $\widehat{\Theta}_t$ and $\widehat{\Theta}_t^*$, respectively. It is noted that the latter have smaller overall mean square errors than the exact conditional mean estimates on the basis of (y_t), denoted by $\widehat{\boldsymbol{\theta}}_t$ and $\widehat{\boldsymbol{\theta}}_t^*$, respectively. We actually compare the Monte Carlo Kalman filter $\mathbf{m}_t$ and smoothing $\mathbf{m}_t^*$ to $\widehat{\Theta}_t$ and $\widehat{\Theta}_t^*$, rather than to their limits $\widehat{\boldsymbol{\theta}}_t$ and $\widehat{\boldsymbol{\theta}}_t^*$.

The simulation took 150 runs, producing sequences of MC Kalman smoothing estimates $(\mathbf{m}_t^{*k})$, as well as the MC Kalman filtering estimates $(\mathbf{m}_t^k)$, based on samples (y_t^k), $k = 1, \ldots, 150$, generated from the given observation and transition models above. Therefore for a given time t, a sequence of 150 estimates of $\boldsymbol{\theta}_t$ was obtained. We first chose a run, say $k = 50$ for illustration. The two top plots in Figure 1 show, respectively, the two components of the true $\boldsymbol{\theta}_t^{50}$ processes, the ideal Kalman filters $\widehat{\Theta}_t^{50}$ and the corresponding Monte Carlo Kalman filters $(\mathbf{m}_t^{50})$. The two bottom plots in Figure 2 display, respectively, the two components of the same true processes, the ideal Kalman smoothings $(\widehat{\Theta}_t^{*50})$ and the Monte Carlo smoothings $(\mathbf{m}_t^{*50})$ with 95% upper and lower confidence bounds. Both figures clearly indicate that the Monte Carlo Kalman filter and smoothing are close to the ideal cases, and the ideal processes are well confined within the confidence bounds of the Monte Carlo smoothings. The overall mean square errors corresponding to the 50th run of the ideal smoothing and the Monte Carlo smoothing with respect to the true processes are, respectively, for the first component 0.0961 and 0.0266, indicating the Monte Carlo smoothing is slightly better, and for the second component 0.0266 and 0.0479, implying the ideal smoothing is marginally better.

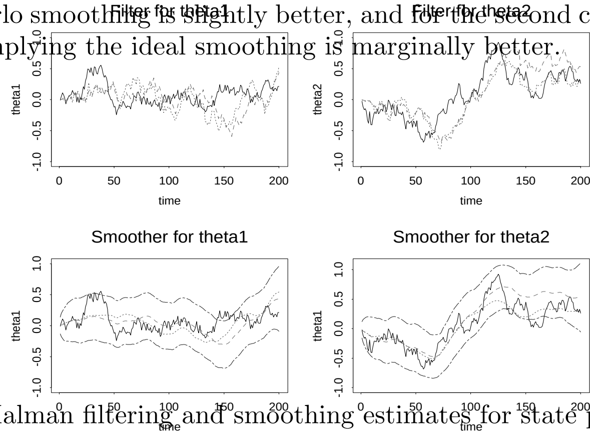

FIGURE 1: Kalman filtering and smoothing estimates for state processes at run $k = 50$. (—) represents the true processes; ($\cdots\cdot$) represents the ideal Kalman filtering (top) or smoothing (bottom) estimates; (- - -) represents the MC Kalman filtering (top) or smoothing (bottom) estimates.

According to Fahrmeir (1992), the empirical distribution functions of componentwise standardised smoothing estimates $(m_t^{*k} - \theta_t)/\sqrt{\mathrm{MSE}(m_t^{*k})}$ for a fixed t were compared to the distribution of the standard normal, shown in Figure 2 where t was chosen to be 150. We also tried the case of $t = 50$, and found the similar pattern to as shown in Figure 2. The closeness of the empirical distribution based on $\{\boldsymbol{\theta}_{150}^k, k = 1, \ldots, 150\}$ to the theoretical one indicates that our Monte-Carlo-based estimates are favourable.

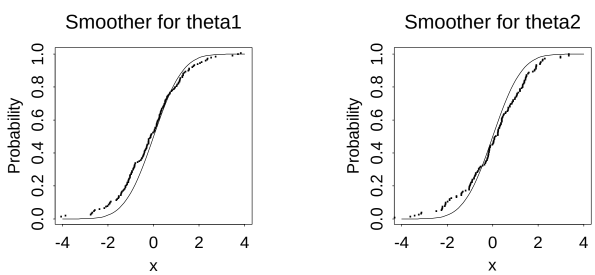

FIGURE 2: Empirical distribution of standardized smoothing estimates for $t = 150$ with 150 runs of replications. (—) represents the standard normal distribution; ($\cdots\cdot$) represents the empirical distribution.

With regard to the stopping rule for convergence of Monte Carlo method, one may terminate the iterative procedure when the difference of the empirical mean square errors of two successive Kalman smoothers is less than a prespecified accuracy. Our experience with this simulation tells that $M = 100$ is usually large enough to achieve a satisfactory approximation accuracy.

5. DATA APPLICATIONS

Two data analyses are given in the section, with the purpose of comparison to other existing methods.

5.1. Infant Sleep Data.

The binary time series of infant sleep status, reported by Stoffer *et al.* (1988), was recorded in a 120-minute EEG study where $y = 1$ if the infant was judged to be in REM sleep during minute t, $y = 0$ otherwise. The data were previously analysed by Carlin & Polson (1992) using a MCMC algorithm and we now reanalyse them by applying the Monte Carlo Kalman filter and smoothing approach developed in the paper. The discrete state space model is comprised of a simple probit observation equation $\pi_t = \Phi(\theta_t)$ and, according to Carlin & Polson (1992), a stationary AR(1) state process $\theta_t = \alpha\theta_{t-1}+\varepsilon_t$, $t = 1, \ldots, 120$, with the initializing state $\theta_0 \sim N(0, 1)$. Here, θ_t may be thought essentially of as an underlying continuous "sleep state" following a stationary Markov process of order 1 with mean zero and variance $\sigma^2/(1 - \alpha^2)$. The objective is to estimate the process θ_t and hence the probability of being in REM sleep.

The application of the MCKS algorithm, with $d = k = p = 1$, $\mathbf{x}_{it} = 1$, $B_t = \alpha$, and $Q_t = \sigma^2$ leads to the MCKS estimate of the state process θ_t shown in Figure 3, with the 95% upper and lower confidence bounds. Figure 4 shows the patterns for estimates of ϕ (the lower curve) and σ^2 (the upper curve) over 70 iterations of the MCKS for the state variables, initialised with $\phi = 0$ and $\sigma^2 = 1$. The figure clearly indicates that both estimates of ϕ and σ^2 stablise very much after iteration 20 that was hence taken as the convergence cutoff point. The corresponding estimated values at this iteration are $\hat{\phi} = 0.2252$ and $\hat{\sigma}^2 = 0.9921$, implying by (17) that $\hat{\alpha} = 0.8408$. Notice that Figure 3 shows a great deal of similarity to Figure 1 of Carlin & Polson (1992), indicating that the MCKS estimate approximates to the state variables, at least in this example, as competitively well as the MCMC estimate, but the MCKS method is much simpler conceptually and much less burdensome computationally.

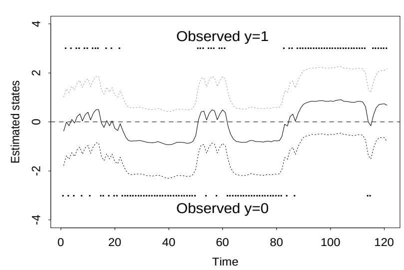

FIGURE 3: MC Kalman smoothing estimates of state process with 95% confidence bounds.

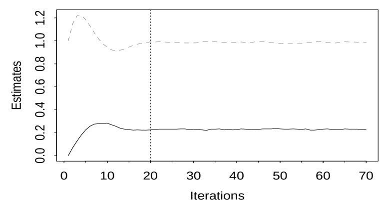

Figure 4: Patterns for estimates of hyperparameters over 70 iterations. (—) represents the estimation for ϕ^2; (- - -) represents the estimation for σ^2.

5.2. Tokyo Rainfall Data.

Consider the Tokyo rainfall data reported by Kitagawa (1987). The data are given by the daily number of occurrences of rainfall over 1 mm in Tokyo for 1983-1984. The data were previously analysed by many authors, e.g., Kitagawa (1987) and Fahrmeir (1992). To fit the data, the binomial state space model takes the form

$$\pi_t = \Phi(\theta_t), \quad \theta_t = \theta_{t-1} + \xi_t, \quad t = 1, \ldots, 366,$$

where $\xi_t \sim N(0, \sigma^2)$, with repetition number $k = 2$. Assume the initial state $\theta_0 \sim N(-1.5, 0.002)$ known from Fahrmeir & Tutz (1994, p. 282). Figure 5 shows the estimated probabilities of rain using MCKS estimate for the state process θ_t with $\hat{\sigma}^2 = 0.028$. The figure shows a similar pattern to both Figure 11 of Kitagawa (1988) and Figure 9 of Fahrmeir (1992), indicating the simple method of MCKS is fairly reliable compared to other rather sophisticated methods.

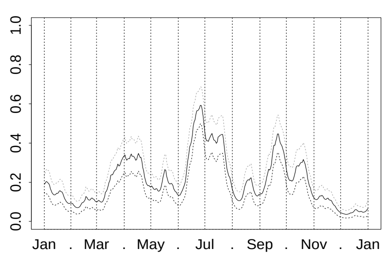

Figure 5: Monte Carlo predicted probability process with 95% confidence bounds based on Kalman smoothing.

6. CONCLUDING REMARKS

This paper proposed a conceptually simple and computationally attractive approach to computing the conditional mean estimates of the state variables through a Monte Carlo approximation technique. As shown in both the simulation study and the two examples of application, the proposed method is reliable and, to some extent, competitive with most existing methods. This idea might be generalized to other situations where an additional process similar to the latent process is available to re-formulate the original dynamic model such that the closed forms of Kalman filter and smoothing procedures are available. This direction would certainly be worth investigating in future research.

Hyperparameters such as Q_t and the autocorrelation for stationary state process involved in the dynamic probit models are estimated consistently via the method of moments. The standard errors of such estimates would possibly be given, e.g., by the estimating equation approach used by Jørgensen *et al.* (1999). This, however, was not this paper's main concern.

APPENDIX

To find equation (13), first note that $\operatorname{cov}\left(\Theta_{t-1}^*, \boldsymbol{\theta}_t - \Theta_t^*\right) = \mathbf{0}$. It is thus sufficient to determine $\operatorname{cov}\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t - \Theta_t^*\right)$. Following Jørgensen *et al.* (1999, Section 3.3), we first obtain

$$\operatorname{cov}\left(\boldsymbol{\theta}_{t-1} - \Theta_{t-1}^*, \boldsymbol{\theta}_t - \Theta_t^*\right) = \operatorname{cov}\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t - B_t\Theta_{t-1}\right) \operatorname{var}^{-1}\left(\boldsymbol{\theta}_t - B_t\Theta_{t-1}^{(i)}\right)\Lambda_t^*,$$

where $\operatorname{var}\left(\boldsymbol{\theta}_t - B_t\Theta_{t-1}\right) = \Lambda_{t|t-1} = B_t\Lambda_{t-1}B_t' + Q$ and

$$\begin{aligned}
\operatorname{cov}\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t - B_t\Theta_{t-1}^{(i)}\right) &= \operatorname{cov}\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t-1} - \Theta_{t-1}\right) B_t' \\
&= \operatorname{var}\left(\boldsymbol{\theta}_{t-1} - \Theta_{t-1}\right) B_t' \\
&= \Lambda_{t-1}B_t'.
\end{aligned}$$

Consequently, $\operatorname{cov}\left(\boldsymbol{\theta}_{t-1} - \Theta_{t-1}^*, \boldsymbol{\theta}_t - \Theta_t^*\right) = \Lambda_{t-1}B_t'\Lambda_{t|t-1}^{-1}\Lambda_t^*$.

ACKNOWLEDGEMENTS

I thank the referee and the Associate Editor for their useful comments and suggestions that led to an improvement of the paper. I also thank both Dr. D. Stoffer and Dr. B. Carlin for allowing me to use the infant sleep data. This research was supported by the National Science and Engineering Research Council of Canada.

REFERENCES

P. J. Brockwell & R. A. Davis (1987). *Time Series: Theory and Methods*. Springer-Verlag, New York.

B. P. Carlin & N. G. Polson (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics*, 4, 577–586.

B. P. Carlin, N. G. Polson & D. S. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87, 493–500.

C. Czado (1997). Multivariate probit analysis of binary time series data with missing responses. Unpublished manuscript.

L. Fahrmeir (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87, 501–509.

L. Fahrmeir & G. Tutz (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York.

J. Geweke (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface*, Seattle, WA, pp. 571–578.

A. C. Harvey (1981). *Time Series Models*. Allan, Oxford.

B. Jørgensen, S. Lundbye-Christensen, P. X.-K. Song & L. Sun (1999). A state-space model for multivariate longitudinal count data. *Biometrika*, 86, 169–181.

B. Jørgensen, S. Lundbye-Christensen, P. X.-K. Song & L. Sun (1996). State-space models for multivariate longitudinal data of mixed types. *The Canadian Journal of Statistics*, 24, 385–402.

G. Kitagawa (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, 82, 1032–1063.

U. V. Naik-Nimbalkar & M. R. Rajarshi (1995). Filtering and smoothing via estimating functions. *Journal of the American Statistical Association*, 90, 301–306.

C. P. Robert (1995). Simulate of truncated normal variables. *Statistics and Computing*, 5, 121–125.

D. S. Stoffer, M. S. Scher, G. A. Richardson, N. L. Day & P. A. Coble (1988). A Walsh-Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *Journal of the American Statistical Association*, 83, 954–963.

T. A. Tanner & W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528–547.

Received 19 May 1999
Accepted 29 December 1999

Peter Xue-Kun Song
song@pascal.math.yorku.ca
Department of Mathematics and Statistics
York University
Toronto, Ontario
Canada M3J 1P3