The Korean Communications
in Statistics Vol. 10 No. 1, 2003
pp. 203-210

Bayesian Estimation of State-Space Model Using the Hybrid Monte Carlo within Gibbs Sampler

Ilsu Choi[1)]

Abstract

In a standard Metropolis-type Monte Carlo simulation, the proposal distribution cannot be easily adapted to "local dynamics" of the target distribution. To overcome some of these difficulties, Duane et al. (1987) introduced the method of hybrid Monte Carlo(HMC) which combines the basic idea of molecular dynamics and the Metropolis acceptance-rejection rule to produce Monte Carlo samples from a given target distribution. In this paper, using the HMC within Gibbs sampler, an asymptotical estimate of the smoothing mean and a general solution to state space modeling in Bayesian framework is obtained.

Keywords : Hybrid Monte Carlo, Gibbs sampler, State-space model.

1. Introduction

For the last decade, various simulation-based nonlinear and non-Gaussian filters and smoothers have been proposed, in order to improve precision of the state estimates and reduce a computational burden. Recently, filters and smoothers have been developed by applying various sampling techniques such as Gibbs sampling, rejection sampling, and Metropolis-Hastings Algorithm within Gibbs sampling.

Carlin et al. (1992) and Carter and Kohn (1994, 1996) applied the Gibbs sampler to evaluate the smoother and the smoothing means in a Bayesian framework. Random draws of the state variables for all time periods are jointly generated, which implies that the smoothing procedure is formulated. They choose the prior densities such that random draws are easily generated. They utilize rejection sampling as well as Gibbs sampling in the case of the nonlinear system. It is known that rejection sampling is sometimes computationally inefficient. We sometimes have the case where rejection sampling does not work well, depending on the underlying assumptions on the functional form or the error terms.

Tanizaki (1996, 1999) and Tanizaki and Mariano (1998) proposed nonlinear filter and

1) Assistant professor, Department of Applied Mathematics, Yeosu National University, Yeosu, Korea.
E-mail : ischoi@yosu.ac.kr

smoother utilizing rejection sampling. When the acceptance probability is close to zero, rejection sampling takes a long time computationally. In order to avoid these computational disadvantages of the existing procedures, Geweke and Tanizaki (1999) suggested the nonlinear and non-Gaussian smoother using the Metropolis-Hastings Algorithm within Gibbs sampling, where the measurement and transition equations are specified in any general formulation and the error terms in the state-space model are not necessarily normal. They also focus on smoothing in a non-Bayesian framework. However, although this procedure might be applicable, the random walk nature of the algorithm makes it very inefficient to explore the posterior distribution.

Duane et al. (1987) introduced Hybrid Monte Carlo (HMC) which can be very effective means for exploring complex posterior distribution. Hybrid Monte Carlo(HMC) as a Markov chain Monte Carlo(MCMC) technique built upon the basic principle of Hamiltonian mechanics. Its applications in molecular simulation have attracted much interest from researchers.

Thus we are to propose that the Hybrid Monte Carlo within Gibbs sampler, dealing with any nonlinear and non-Gaussian state-space model in a Bayesian framework. We apply Gaussian state-space model in Shephard and Pitt(1997) as daily exchange rates of won/dollar.

2. Hybrid Monte Carlo

Hybrid Monte Carlo(HMC) as first introduced by Duane et al.(1987) is a Markov chain Monte Carlo(MCMC) technique built upon the basic principle of Hamiltonian mechanics. The method is designed to promote rapid mixing of the Markov chain, and is especially suited to problems involving complex densities where exploration by a random walk may be too slow. In its simplest and original form, HMC introduces a set of auxiliary momenta variables $\boldsymbol{p} = (p_1, \cdots, p_T)$ and the related Hamiltonian function $H(\boldsymbol{x}, \boldsymbol{p})$:

$$H(\boldsymbol{x}, \boldsymbol{p}) = U(x_1, \cdots, x_T) + \frac{1}{2} \sum_{t=1}^{T} p_t^2 = U(\boldsymbol{x}) + \boldsymbol{p}^2/2$$

From the Gibbs factor:

$$P(\boldsymbol{x}, \boldsymbol{p}) \propto \exp[-H(\boldsymbol{x}, \boldsymbol{p})] = \exp[-U(\boldsymbol{x})]\exp[-\boldsymbol{p}^2/2]$$

The method deduce that, from the statistical point of view, the momenta $\boldsymbol{p}$ are nothing but a set of independent, Gaussian distributed, random variables of zero mean and variance equal to the system. There is no simple closed form for the proposal probability $g(\boldsymbol{x}'|\boldsymbol{x})$, and the proposal change $\boldsymbol{x} \rightarrow \boldsymbol{x}'$ is done in the following way: first, a set of initial values for the momenta $\boldsymbol{p}$ are generated by using the Gaussian distribution $\exp[-\boldsymbol{p}^2/2]$ as suggested by the above equation; next, Hamilton's equation of motion, $\dot{x}_i = p_i$, $\dot{p}_i = F_i$, where $F_i(\boldsymbol{x}) = -\partial U(\boldsymbol{x})/\partial x_i$ is the force acting on the variable x_i, are integrated numerically using the leap-flog algorithm with a time step δt:

$$x_t' = x_t + \delta t p_t + \frac{\delta t^2}{2} F_t(\boldsymbol{x})$$
$$p_t' = p_t + \frac{\delta t}{2}[F_t(\boldsymbol{x}) + F_t(\boldsymbol{x}')], \qquad t=1,\cdots,T$$

The proposal $\boldsymbol{x}'$ is obtained after n iterations of the previous basic integration step. In other words: by numerical integration of Hamilton's equations during a time $n\delta t$. The value $\boldsymbol{x}'$ must now be accepted with a probability given by:

$$h(\boldsymbol{x}'|\boldsymbol{x}) = \min\{1, \exp[-(H(\boldsymbol{x}', \boldsymbol{p}') - H(\boldsymbol{x}, \boldsymbol{p}))]\} \qquad (*)$$

Summing up, the HMC proceeds by generating representative configurations by using a proposal obtained by some of the mappings given above. This proposal must now be accepted with a probability given by (*). In the special case where only one deterministic step used, it is called the Langevin algorithm, which is a discrete time approximation to the Langevin diffusion process.

3. Bayesian formula in State-Space Model

We consider a nonlinear and nonnormal state-space model in the following general form:

(Measurement Equation) $\quad y_t = h_t(x_t, \varepsilon_t, \gamma)$

(Transition Equation) $\quad x_t = f_t(x_{t-1}, \eta_t, \delta)$

for $t=1,2,\cdots,T$, where T denotes the sample size. Suppose we observe only y_t and the functional forms of both $h(\cdot)$ and $f(\cdot)$ are known, whereas x_t is not directly observed. Since the analytical computation of the likelihood function of γ is generally infeasible, the standard maximum likelihood estimation method cannot be applied. We overcome this difficulty by contaminated error ε_t.

Treating the problem as a missing data problem, we write the pseudo posterior distribution of α and γ as follows:

$P(x_T, \gamma | Y_T) \propto P(Y_T | x_T, \gamma) P(x_T | \gamma) P(\gamma)$

It can be shown that under mild conditions, the pseudo posterior of γ converges to its true posterior almost surely as $\sigma^2 \to 0$.

Under the setup, the density of x_T and Y_T given γ and δ is written as:

$$P(x_T, Y_T | \gamma, \delta) = P(x_T | \delta) P(Y_T | x_T, \gamma)$$

where the two densities in the right hand side are represented by:

$$P(x_T|\delta)=\begin{cases} P(x_0|\delta)\prod_{T=1}^{T}P(x_t|x_{t-1},\delta), & \text{if } x_0 \text{ is stochastic,}\\ \prod_{T=1}^{T}P(x_t|x_{t-1},\delta) & \text{otherwise}\end{cases}$$

$$P(Y_T|x_T,\gamma)=\prod_{T=1}^{T}P(y_t|x_t,\gamma)$$

where $P(x_0|\delta)$ denotes the initial density of x_0 when x_0 is assumed to be a random variable. From the Bayes theorem, the conditional distribution of x_T given Y_T, γ and δ is obtained as follows:

$$P(x_T|Y_T,\gamma,\delta)=\frac{P(x_T,Y_T|\gamma,\delta)}{\int P(x_T,Y_T|\gamma,\delta)dx_T}$$

4. HMC within Gibbs sampler

Carter and Kohn(1996) applied the Gibbs sampler to evaluate the smoothing means in a Bayesian framework. Generally, the smoothing random draws are generated as follows:

(STEP 1)

1) Take appropriate values for γ, δ and x_t, $t=1,2,\cdots,T$.
2) Generate a random draw of x_t from $P(x_t|\cdot)$ for $t=1,2,\cdots,T$.
3) Generate a random draw of γ from $P(\gamma|\cdot)$.
4) Generate a random draw of δ from $P(\delta|\cdot)$.
5) Repeat 2)-4) N times to obtain N random draws of x_T, δ and γ.

Unfortunately it is hard to tell how long it takes to reach the stationary distribution or how correlated are the values of successive iterations. In the state space models, state variable at present time has high correlation with that at past time. Therefore, convergence of the Gibbs sampler is unacceptably slow. Because It takes rejection method which takes a long time computationally when the acceptance probability is close to zero. Hence rejection method cannot be applied at state space model because of acceptance probability trouble. Generating a candidate state by randomly perturbing all weights at once does not solve this problem, since a randomly chosen direction in the high dimensional weight space is unlikely to be close to that desired. What is needed is an elaboration of the Metropolis algorithm that makes use of the gradient information provided by a candidate directions in which changes have a high probability of being accepted. Geweke and Tanizaki(1999) used the Metropolis algorithm, an attempt is made to generate random draws of x_T, γ and δ directly from $P(x_T|Y_T,\gamma,\delta)$, $P(\gamma|x_T,Y_T,\delta)$ and $P(\delta|x_T,Y_T,\gamma)$. The Metropolis algorithm within the Gibbs sampler is applied to random number generation. But Metropolis algorithm within Gibbs sampler is inefficient of random walk nature.

The Hybrid Monte Carlo(HMC) method devised by Duane, Kennedy, Pendeleton, and Roweth(1987) for use in quantum chromodynamics calculations does this. It also eliminates much of the random walk aspect of the Metropolis algorithm, further speeding exploration of the parameter space. The Hybrid Monte Carlo method are generated as follows:

(STEP 2)

1) Generate a new momentum vector x_t from Gaussian distribution $\pi(p_t) \propto \exp\{-K(p_t)\}$.

2) Run the leapfrog algorithm for L steps to reach a new configuration in the phase space $(y_t^{(L)}, p_t^{(L)})$.

3) Let $(y_{(t,n)}, p_{(t,n)}) = (y_t^{(L)}, p_t^{(L)})$ with probability

$$\min[1, \exp\{-H(y_t^{(L)}, -p_t^{(L)}) + H(y_{(t,n-1)}, p_t)\}]$$

where $H(\cdot, \cdot)$ is Hamiltonian.

It seems that the Markov chain Monte Carlo procedure is less computational than any other estimators. However, the Markov chain Monte Carlo methods in the state-space model need a lot of random draws compared with the independence Monte Carlo methods because in the Markov chain Monte Carlo methods we usually discard the first 10% - 20% random draws and a random draw is positively correlated with the next random draw in general. Moreover, it is known that convergence of the Gibbs sampler is very slow especially in the case where there is high correlation between x_t and x_{t-1}. In particular we have the case where rejection sampling does not work well, depending on the underlying assumptions on the functional form or the error terms.

We adopt Hybrid Monte Carlo within Gibbs sampler to eliminates much of the random walk. The smoothing random draws are generated as follows:

(STEP 3)

1) Given the state, sample γ and δ from their conditional distributions.(STEP 1)

2) Given γ and δ, impute the states x_t by HMC.(STEP 2)

5. Application and its conclusion

5.1 Application

Consider the following Gaussian state-space model in Shephard and Pitt(1997):

$$y_t = \mu + x_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$
$$x_t = \phi x_{t-1} + \eta_t \qquad \eta_t \sim N(0, \sigma_\eta^2) \quad x_1 \sim N(0, \sigma_\eta^2/(1-\phi^2))$$

where one observes y and is interested in sampling from the posterior distribution of x and μ. Our dataset consist of daily exchange rates of won/dollar from 1/3/1998 to 12/31/2001 (a total of $T=1188$ observations).

Let $\boldsymbol{x}=(x_1,\cdots,x_T)$ and $\boldsymbol{y}=(y_1,\cdots,y_T)$, and let prior for σ_η^2, $\mathrm{Inv}-\chi^2(\alpha_1,\beta_1)$; for μ, $N(0,\sigma_\mu^2)$; for σ_ε^2, $\mathrm{Inv}-\chi^2(\alpha_2,\beta_2)$; and for $(\phi+1)/2$, a beta prior with shape parameters α_3 and β_3. Then the following conditional distributions can be easily sampled from:

$$\sigma_\eta^2|\phi,\boldsymbol{x} \sim \mathrm{Inv}-\chi^2(T+\alpha_1, V)$$

where $V=\frac{1}{T+\alpha_1}\left\{\alpha_1\beta_1+x_1^2(1-\phi^2)+\sum_{t=2}^{T}(x_t-\phi x_{t-1})^2\right\}$

$$\phi|\sigma_\eta^2,\boldsymbol{x} \propto \exp\left\{-\frac{x_1^2(1-\phi^2)+\sum_{t=1}^{T-1}(x_{t+1}-\phi x_t)^2}{2\sigma^2}\right\}(1+\phi)^{\alpha_3-0.5}(1-\phi)^{\beta_3-0.5}$$

$$\sigma_\varepsilon^2|\boldsymbol{x},\boldsymbol{y},\mu \sim \mathrm{inv}-\chi^2\left(\alpha_2+T, \frac{1}{\alpha_2+T}\left\{\alpha_2\beta_2+\sum_{t=1}^{T}(y_t-\mu-x_t)^2\right\}\right)$$

$$\mu|\boldsymbol{y},\boldsymbol{x} \sim \left(\frac{\sigma_\mu^2\sum_{t=1}^{T}(y_t-x_t)}{2(T\sigma_\mu^2+\sigma_\varepsilon^2)}, \frac{\sigma_\varepsilon^2\sigma_\mu^2}{T\sigma_\mu^2+\sigma_\varepsilon^2}\right)$$

Once the parameter values are given, the negative log density is

$$U(\boldsymbol{x})=\sum_{t=1}^{T}\frac{(y_t-\mu-x_t)^2}{2}+\frac{x_1^2(1-\phi^2)}{2\sigma_\eta^2}+\sum_{t=1}^{T-1}\frac{(x_{t+1}-\phi x_t)^2}{2\sigma_\eta^2}$$

The posterior density of $\boldsymbol{x}$, given the parameter values, is proportional to $\exp\{-U(\boldsymbol{x})\}$.

We implemented the following iterative sampling algorithm:

(1) Given $\boldsymbol{x}$, we drew the parameters μ, σ^2, and ϕ from the above conditional distributions.

(2) Whereas given μ, σ^2, and ϕ, we drew the state variable by the HMC.

This HMC within Gibbs sampler were run for 20,000 iterations and the results from the last 15,000 iterations are reported in table 1.

Parameter	Mean	Standard deviation	Covariance		
μ	-0.0249	0.0038	0.1449e-04	0.0369e-04	-0.0447e-04
σ	0.0838	0.0020	0.0369e-04	0.0412e-04	-0.0144e-04
ϕ	0.9734	0.0060	-0.0447e-04	-0.0144e-04	0.4421e-04

table 1. Bayes estimates of the parameters in the state-space model.

5.2 Conclusion

The classical sampling methods such as resampling, rejection sampling and Markov Chain Monte Carlo methods have convergence problems to apply nonlinear and non-Gaussian methods.

The resampling procedure by Kitagawa (1997) has the disadvantage that it requires heavy computation, especially for smoothing. However, this problem will be improved in the future as computer progresses.

The disadvantages of the rejection sampling procedure (Tanizaki (1999)) are : (i) the proposal density has to be appropriately chosen by a researcher (use of the transition equation might be recommended, but not necessarily), (ii) it takes a long time computationally when the acceptance probability is small (i.e., we cannot predict how long the computer program will run), and (iii) sometimes the supremum of the ratio of the target density does not exist.

The Markov chain Monte Carlo procedure proposed by Geweke and Tanizaki (1999) has the following problems: (i) the proposal density has to be appropriately chosen by a researcher (it might be plausible to take the transition equation for the proposal density), and (ii) convergence is very slow because the Gibbs sampler and the Metropolis-Hastings are simultaneously used (remember that the random draw generated by the Markov chain Monte Carlo method is correlated with the next one).

The above classical methods cannot be applied to daily exchange rate data due to the convergence problems. Thus, we used Hybrid Monte Carlo in sampling state variables while keeping Gibbs sampler for the sampling of parameters.

Although HMC has been found useful for Bayesian computations, many important issues remain open. For example, how to choose tuning parameters in HMC, e.g., the step size and the number of the leapfrog iterations, is still difficult problem.

References

[1] Carlin, B.P., Polson, N.G. and Stoffer, D.S., (1992), A Monte Carlo Approach to Nonnormal and nonlinear State Space Modeling, *Journal of American Statistical Association*, Vol.87, No.418, pp.493-500.

[2] Carter, C.K. and Kohn, R., (1994), On Gibbs sampling for state space models, *Biometrika*, Vol81, N0.3, pp.541-553.

[3] Carter, C.K. and Kohn, R., (1996), Markov Chain Monte Carlo in Conditionally Gaussian State space models. *Biometrika*, Vol83, N0.3, pp.589-601.

[4] Duane, S., Kennedy, A.D., Pendleton, B. and Roweth, D., (1987), Hybrid Monte Carlo. *Physics Letters B*, vol. 195, 216-222.

[5] Geweke, J. and Tanizaki, H.(1999), Bayesian Estimation of State-Space Models Using the

Metropolis-Hastings Algorithm within Gibbs Sampling, Unpublished manuscript.

[6] Kitagawa, G., Non-Gaussian State-Space Modeling of Nonstationary Time Series, *Journal of the American Statistical Association*, Vol.82, pp.1032-1063, 1987.

[7] Shephard, N. and Pitt, M. Likelihood analysis of non-Gaussian measurement timeseries:, *Biometrika*, Vol84, pp.653-667, 1997.

[8] Tanizaki, H.(1996), Nonlinear Filters: Estimation and Applications (Second, rivesed and enlarged edition), Springer-Verlag.

[9] Tanizaki, H.(1999), On the Nonlinear and nonnormal filter using rejection sampling, IEEE transactions on automatic control, Vol. 44, N0. 2, pp.314-319.

[10] Tanizaki, H. and Mariano, R.S., Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Simulations, *Journal of Econometrics*, Vol.83, No.1-2, pp.263-290, 1998.

[Received December 2002, Accepted March 2003]