Proceedings of Nonlinear Statistical Signal Processing Workshop (NSSPW 2006), September 13-15, Cambridge, UK.

IRREDUCIBLE MARKOV CHAIN MONTE CARLO SCHEMES FOR PARTIALLY OBSERVED DIFFUSIONS

Konstantinos Kalogeropoulos, Gareth Roberts, Petros Dellaportas

University of Cambridge, University of Lancaster, Athens University of Economics and Business

ABSTRACT

This paper presents a Markov chain Monte Carlo algorithm suitable for a class of partially observed non-linear diffusions. This class is of high practical interest; it includes for instance stochastic volatility models. We use data augmentation, treating the unobserved paths as missing data. However, unless these paths are transformed, the algorithm becomes reducible. We circumvent the problem by introducing appropriate reparametrisations of the likelihood that can be used to construct irreducible data augmentation schemes.

1. INTRODUCTION

Diffusion processes constitute a natural and useful tool for modelling phenomena evolving continuously in time. They find applications in many different fields including finance, biology, physics, engineering etc. In this paper we focus on diffusions defined through stochastic differential equations (SDEs) of the following form:

$$dX_t = \mu_x(X_t, \alpha_t, \theta)dt + \sigma_x(\alpha_t, \theta)dB_t, X_0 = x, 0 \leq t \leq T \quad (1)$$
$$d\alpha_t = \mu_\alpha(\alpha_t, \theta)dt + \sigma_\alpha(\alpha_t, \theta)dW_t$$

where B and W are independent Brownian motions. The drift (μ_x, μ_α) and the volatility $diag(\sigma_x, \sigma_\alpha)$ of the diffusion should satisfy some regularity conditions (locally Lipschitz with a growth bound) to ensure that the SDE will have a weakly unique solution; see chapter 4 of [1]. Throughout this paper we will allow them to have a general but known functional form, covering cases of non linear models. Our task is to draw inferences for the parameter vector θ based on observations only on X at a discrete set of times $t_1, \ldots, t_n$, denoted by $Y = \{Y_k = X_{t_k},\ k = 1, \ldots, n\}$.

For non-linear diffusions, the likelihood is not available in closed form. As a result of this, the literature contains various methodologies that may or not be based on the likelihood; see [2] for an extensive review. Likelihood based approaches are either analytical ([3], [4]), or they use simulations ([5], [6]). They usually approximate the likelihood in a way so that the discretisation error can become arbitrarily small, although the methodology developed in [7] succeeds exact inference in the sense that it allows only for Monte Carlo error. Unfortunately, all of the above rely on the Markov property and therefore become hard to generalise to the non-Markovian case.

We proceed using Markov Chain Monte Carlo methods (MCMC) and in particular data augmentation. Next section provides the relevant details and highlights potential reducibility issues of the chain. Appropriate reparametrisations are introduced in section 3 that may be used to construct irreducible and efficient MCMC schemes. Section 4 provides some details on their implementation, section 5 presents two relevant examples and finally section 6 concludes.

2. DATA AUGMENTATION - REDUCIBILITY

A natural way to proceed is via data augmentation, a methodology introduced by [8]. The idea is based on the fact that the likelihood can always be well approximated given the entire path of the diffusion or a sufficiently fine partition of it. Therefore given θ, the unobserved paths of X (paths between observations) and of α (entire path) are treated as missing data and a finite number of points, large enough to make the approximation error arbitrarily small, is imputed. Then θ is updated conditional on the augmented path.

As noted in [9] however, there exists a strong dependence between the imputed paths and the volatility coefficients. In fact the algorithm becomes reducible as the number of imputed points increases. [9] tackle the problem for scalar diffusions by a reparametrisation on the paths of V and [10] offer an extension for some multivariate diffusions but their framework does not cover the models in (1). This paper focuses on this class and introduces novel reparametrisations of the likelihood that may serve as the basis for data augmentation schemes. Alternative approaches to this problem can be found in [11] and [12].

3. REPARAMETRISATION

[9] noted that the irreducibility is caused by the fact that the likelihood, provided by Girsanov's formula (see for instance chapter 8 of [13]) is written with respect to a dominating measure that depends on θ. The problem may be solved if we transform the diffusion appropriately.

We will consider the SDE's α and $X|\alpha$ separately. Note that if $\mu_x(X_t, \alpha_t, \theta) \equiv \mu_x(\alpha_t, \theta)$, the density of Y given α is known and given by:

$$p_\theta(Y|\alpha) = \prod_{k=1}^{n} p_\theta\left(Y_k\,|Y_{k-1}, \{\alpha_t : t_{k-1} \le t \le t_k\}\right).$$

where

$$p_\theta\left(Y_k\,|Y_{k-1}, \{\alpha_t : t_{k-1} \le t \le t_k\}\right) \sim N(\mu_k,\ \sigma_k^2)$$

with

$$\mu_k = Y_{k-1} + \int_{t_{k-1}}^{t_k} \mu_x(\alpha_s, \theta)ds,$$

$$\sigma_k^2 = \int_{t_{k-1}}^{t_k} \sigma_x(\alpha_s, \theta)^2 ds.$$

Hence in this case, there is no need to impute and consequently no need to transform the paths of X.

3.1. Transforming the paths of α

Denote by $\mathbb{P}_\theta(\alpha)$ the distribution of α and by $\mathbb{Q}_\theta(\alpha)$ that of the driftless version of α. Given the initial point α_0, Girsanov's formula provides the Radon-Nikodym derivative of $\mathbb{P}_\theta(\alpha)$ with respect to $\mathbb{Q}_\theta(\alpha)$. Clearly $\mathbb{Q}_\theta(\alpha)$ depends on θ. For this reason we introduce the following two-step transformation:

1. $\beta_t = h(\alpha_t, \theta)$, where $h(.)$ satisfies

$$\frac{\partial h(V_t, \theta)}{\partial V_t} = \{\sigma(V_t, \theta)\}^{-1}$$

2. $\gamma_t = \beta_t - \beta_0 = \beta_t - h(\alpha_0, \theta),\ \ \beta_t = \eta(\gamma_t)$

By Ito's lemma, the process γ will have unit volatility and drift:

$$\mu_\gamma(\gamma, \theta) = \frac{\mu_\alpha\left[h^{-1}(\beta_t, \theta), \theta\right]}{\sigma_\alpha\left[h^{-1}(\beta_t, \theta), \theta\right]} - \frac{1}{2}\frac{\partial \sigma_\alpha\left[h^{-1}(\beta_t, \theta), \theta\right]}{\partial h^{-1}(\beta_t, \theta)}.$$

Now we can use Girsanov's formula for the Radon-Nikodym derivative with the distribution of Brownian motion as the reference measure:

$$\frac{d\mathbb{P}_\theta(\gamma)}{d\mathbb{W}} = \exp\left(\int_0^T \mu_\gamma(\gamma_s, \theta)d\gamma_s - \frac{1}{2}\int_0^T \mu_\gamma^2(\gamma_s, \theta)ds\right).$$

See [14] for more details.

3.2. Transforming the paths of X

Given the path of α we can see $\sigma_x(\alpha_t, \theta)$ as a deterministic function of time. Hence, for each path of X between successive observations (say between times t_k and t_{k+1}), we introduce a new time scale

$$t' = \eta(t, \theta) = \int_{t_k}^{t} \sigma_x^2(\alpha_s, \theta)ds$$

and we set $U_t = X_{\eta^{-1}(t,\theta)}$. Using time change properties, we get:

$$dU_t = \left\{\frac{\mu\left(U_t, \alpha_{\eta(t)}, \theta\right)}{\sigma_x^2(\alpha_{\eta(t)}, \theta)}\right\}dt + dW_t,\ \ 0 \le t < \eta(t_{k+1}).$$

Operating on the new time scale we can use Girsanov's formula to write the likelihood with respect to the distribution of a Brownian bridge that finishes at time $T = \eta(t_{k+1}, \theta)$. To remove this dependency on θ we introduce a second time change:

$$U(t) = \nu(Z(t)) = (T-t)Z\left(\frac{t}{T(T-t)}\right) + \left(1 - \frac{t}{T}\right)y_k + \frac{t}{T}y_{k+1},$$

where $0 \le t \le T$. The SDE for Z is given by:

$$dZ_t = \left\{\frac{\mu\left(\nu(Z_t), \alpha_{g(t)}, \theta\right)}{\sigma_x^2(\alpha_{g(t)}, \theta)}\ \frac{T}{1+tT}\right\}dt + dW_t,\ \ 0 \le t < \infty,$$

where $g(t)$ denotes the initial time scale.

The transformation above stretches the ending point of the bridge to infinity, in a way so that it will still have unit volatility. As in the previous section, we ended up in a position where we can use the distribution of a Brownian motion as the reference measure. See [15] for more details.

4. MCMC IMPLEMENTATION

Based on the reparametrisations introduced in the previous section we can write down a likelihood that can be used to construct irreducible MCMC algorithms. The updates on parameters not involved in the time change are relatively straightforward and can be implemented using ordinary techniques.

To update parameters used to define the time change, we propose a random walk metropolis step. Note that each proposed value implies a different set of times of the path of X which they will not be stored in our computer as it is impossible to store the full path. However, they can be drawn retrospectively using standard Brownian bridges arguments.

To update Z, we can split the process into blocks, say the paths between successive observations, and update each one of them in turn. One may use an independence sampler with Brownian motion as the proposal distribution (reference measure of the likelihood). The fact that the time scale is defined up to $+\infty$ poses no restrictions as under the discretisation of the path we only need to simulate a finite number of points.

While it is rather clear that the path of γ should be divided into blocks, it is not straightforward how this should be done. Suppose that we observe Y at times t_k, as in section 2.2, and that we split the path of γ into n blocks $\{b_k = \gamma_s,\ t_{k-1} \le s \le t_k,\ \ k = 1, 2, \ldots, n\}$. Note that under this formulation the endpoints of the blocks are not updated at all leading to a reducible MCMC chain; an alternative blocking scheme is needed. Available strategies use overlapping [14], or random sized blocks [11].

5. EXAMPLES

The simulations performed in this section aim to demonstrate two aspects of the problem. First, we highlight the necessity of the reparametrisation introduced in section 3 by exposing the problem in the case of a rather simple model of (1). Then we check the validity of the MCMC scheme using a more realistic stochastic volatility model.

5.1. A toy example

Assume $\mu_x = \mu_\alpha = 0$, $\sigma_x = \exp(\alpha_t/2)$ and $\sigma_\alpha = \sigma$ and define Y as before ($n = 500$). Unless we apply a reparametrisation, Girsanov's formula is not useful for writing down the likelihood. Alternatively we may use the Euler-Maruyama approximation; see chapter 9 of [1]. Under our scheme, we impute the values of α that correspond to observations of X and we further impute m values between every pair of successive times with observations. For simplicity we assume that the imputed points are equidistant and denote the time interval between them by $\delta = (m+1)^{-1}$. Let $V_t = (X_t, \alpha_t)'$ and $\Sigma = diag\{\exp(\alpha_t), \sigma^2\}$. Under the Euler-Maruyama approximation and given V_0 we get:

$$\pi(Y, \sigma^2, V_t) = \prod_{t=1}^{n(m+1)+1} \pi(V_t | V_{t-1}, \sigma^2)$$

$$\pi(V_t | V_{t-1}, \sigma^2) \sim N\left(V_{t-1},\ \delta\Sigma_{t-1}\right)$$

If we assign $\pi(\sigma^2) \propto \sigma^{-2}$ as the prior for σ^2 and assume that $\alpha_0 = 0$, we get that its conditional posterior density is an Inverse-Gamma distribution with parameters:

$$a = \frac{n(m+1)}{2},\ b = \frac{(m+1)\sum_{t=1}^{n(m+1)+1}(\alpha_t - \alpha_{t-1})^2}{2}$$

We ran a MCMC chain for different numbers of imputed points ($m = 1, 10, 40, 100$), updating the paths as described in 4 and using the Gibbs step for the updates σ^2. Figure 1b shows the autocorrelation of the posterior draws of σ^2 for each value of m. Clearly, the autocorrelation increases dramatically leading to an increasingly slower chain. An alternative way to see this is to note that the variance of the conditional posterior for σ^2 goes to 0 as we increase m.

The problem can be resolved if we apply this paper's proposed reparametrisation. Following the route of 3.1, we set $\beta_t = \alpha_t/\sigma$ and $\gamma_t = \beta_t - \beta_0 = \beta_t$ and we get

$$dX_t = \exp(\sigma\gamma_t/2)dBt,$$

where β is a standard Brownian motion independent of B. The likelihood then simplifies to:

$$\pi(Y, \sigma^2, \gamma_t) = \prod_{k=1}^{n} \pi(Y_k | Y_{k-1}, \sigma^2, \gamma_t),$$

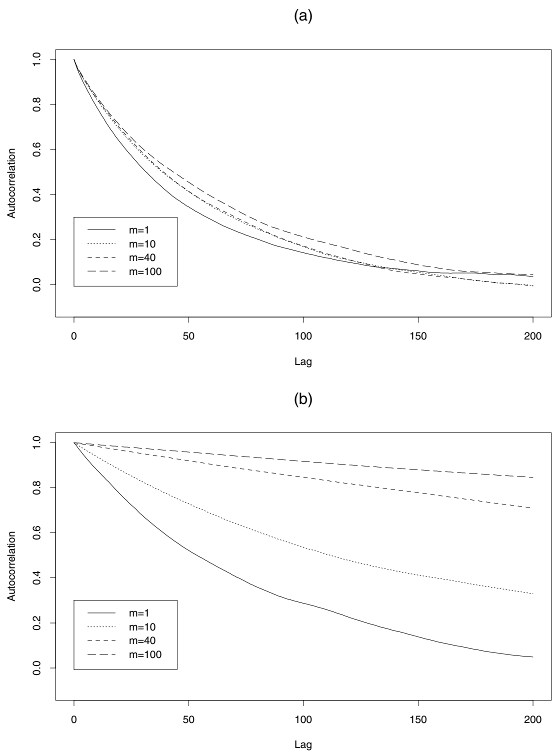

Figure 1: Autocorrelation plots of posterior draws of σ^2 for different values of imputed points between observations (m) for the simple stochastic volatility model. The draws in (a) correspond to the reparametrised scheme scheme and in (b) to the scheme without transformation.

where

$$\pi(Y_k | Y_{k-1}, \sigma^2, \gamma_t) \sim N\left\{Y_{k-1},\ \int_{t_{k-1}}^{t_k} \exp(\sigma\gamma_s)ds\right\}.$$

Figure 1a contains the corresponding autocorrelation plots of the posterior draws of σ^2 taken from the reparametrised data augmentation scheme. Unlike the previous case (figure 1b) there is clearly no increase in the autocorrelation.

5.2. Stochastic volatility

Now consider an extended version of the previous model with $\mu_x = \lambda(\nu - X_t)$ and $\mu_\alpha = \kappa(\mu - \alpha_t)$. First we need to introduce γ as before. This leads us to $\sigma_x = \exp(\sigma\gamma_t/2)$. However we also need to impute the paths of X (between observations Y) to get the likelihood for the parameters of μ_x. For an irreducible chain we may use the time change transformation of (3.2).

After implementing this MCMC scheme, we found no evidence of increasing autocorrelation of the parameter posterior draws getting similar pictures to figure 1a. Furthermore, table 1 provides the posterior means and standard deviations of the parameters. We see that these estimates are in good agreement with the values we simulated the data from.

Parameter	True value	Posterior mean	Posterior SD
λ	0.2	0.1899	0.0305
ν	0.1	0.1196	0.2274
κ	0.1	0.1944	0.0737
μ	-0.6	-0.6269	0.1554
σ	0.3	0.3131	0.1014

Table 1: Posterior means and standard deviations of the parameters versus their true values.

6. CONCLUSION

Inference on partially observed non-linear diffusions is a particularly difficult task as the likelihood is not available in closed form. Things are further complicated by the fact that the Markov property is no longer available. Data augmentation schemes provide us with a convenient framework as they allow for good approximations of likelihood, where the error may become arbitrarily small by simply increasing the number of imputed points. However, one has to be extra careful to avoid reducibility issues. In this paper we provide appropriate reparametrisations for most partially observed diffusions including stochastic volatility models.

7. ACKNOWLEDGMENTS

This work was supported by Irakleitos - Fellowships for Research of the Athens University of Economics and Business which is co-financed by the European Union and by EPSRC - grant GR/S61577/01 for funding.

8. REFERENCES

[1] P.E. Kloeden. and E. Platen, *Numerical solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1995.

[2] H. Sørensen, "Parametric inference for diffusion processes observed at discrete points in time: a survey," *International Statistical Review*, vol. 72, no. 3, pp. 337–354, 2004.

[3] Y. Aït-Sahalia, "Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach," *Econometrica*, vol. 70, pp. 223–262, 2002.

[4] Y. Aït-Sahalia, "Closed form likelihood expansions for multivariate diffusions," 2005.

[5] A. R. Pedersen, "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations," *Scand. J. Statist.*, vol. 22, no. 1, pp. 55–71, 1995.

[6] G. B. Durham and A. R. Gallant, "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes," *J. Bus. Econom. Statist.*, vol. 20, no. 3, pp. 297–316, 2002, With comments and a reply by the authors.

[7] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P Fearnhead, "Exact and efficient likelihood-based estimation for discretely observed diffusion processes," *J. R. Statist. Soc. B*, vol. 68, no. 2, pp. 1–29, 2006.

[8] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.

[9] G.O. Roberts and O. Stramer, "On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm," *Biometrika*, vol. 88, no. 3, pp. 603–621, 2001.

[10] K. Kalogeropoulos, P. Dellaportas, and G.O. Roberts, "Inference for multidimensional diffusion models using markov chain monte carlo.," 2006, In preparation.

[11] S. Chib, M. K. Pitt, and N. Shephard, "Likelihood based inference for diffusion models," 2005, In preparation.

[12] A. Golightly and D. Wilkinson, "Bayesian sequential inference for nonlinear multivariate diffusions.," 2005, In preparation.

[13] B. Øksendal, *Stochastic differential equations: An introduction with applications*, Springer, 5th edition, 2003.

[14] K. Kalogeropoulos, "Likelihood based inference for a class of multivariate diffusions with unobserved paths.," *To appear in Journal of Statistical Planning and Inference*, 2006.

[15] K. Kalogeropoulos, G.O. Roberts, and P. Dellaportas, "A unified framework for likelihood based inference on diffusion driven stochastic volatility models.," 2006, In preparation.