Bayesian Inference for Discretely Sampled Diffusion Processes: A New MCMC−Based Approach to Inference

Osnat Stramer[1] and Matthew Bognar[1]

Department of Statistics and Actuarial Science, University of Iowa
241 Schaeffer Hall
Iowa City, IA 52242, USA
(e-mail: osnat-stramer@uiowa.edu, e-mail: matthew-bognar@uiowa.edu)

Abstract. For frequentist inference, the efficacy of the closed-form (CF) likelihood approximation of Aït-Sahalia (2002, 2007) in financial modeling has been widely demonstrated. Bayesian inference, however, requires the use of MCMC, and the CF likelihood can become inaccurate when the parameters θ are far from the MLE. Due to numerical stability problems, the samplers can therefore become stuck when (typically) in the tails of the posterior distribution. It may be possible to address this problem by using numerical integration to estimate the intractable normalizers in the CF likelihood, but determining the limiting distribution after using such approximations remains an open research question. Auxiliary variables have been used in conjunction with MCMC to address intractable normalizers (see Møller et al. (2006)), but choosing such variables is not trivial. We propose a MCMC algorithm (called EMCMC) that addresses the intractable normalizers in the CF likelihood which 1) is easy to implement, 2) yields a sampler with the correct limiting distribution, and 3) greatly increases the stability of the sampler compared to using the unnormalized CF likelihood in a standard Metropolis-Hastings algorithm. The efficacy of our approach is demonstrated in a simulation study of the Cox-Ingersoll-Ross (CIR) model.

Keywords. Bayesian inference, diffusion process, closed-form likelihood, EMCMC.

1 Introduction

Diffusions are used to model continuous time processes and are therefore commonly used in financial models. Data are observed at discrete points in time, and in all except a few typical cases the likelihood $\pi(\mathbf{x}|\theta)$ is not analytically available, hindering classical and Bayesian inferential techniques.

Available Bayesian MCMC techniques are mainly based for high frequency augmentation and are computationally intense due to high dependence between the parameters and missing data. A Bayesian approach that is based on the closed-form (CF) approximation of Aït-Sahalia (2002) for one-dimensional diffusions is derived in Di Pietro (2001). This approach avoids a data augmentation scheme.

For univariate diffusions, the CF approximation is an expansion based upon expanding the likelihood around a normal density using Hermite polynomials. This expansion converges to the true, unknown likelihood function as the number of terms in the Taylor-like expansion increases. In fact, the first two terms in the expansion have been shown to be sufficient in yielding very accurate approximations of the true likelihood function for many of the types of data one encounters in finance. The CF likelihood approximation, which we denote by $\pi_{CF}(\mathbf{x}|\theta)$, should be used with caution for very volatile models or sparse data-sets (see Stramer and Yan (2007)).

Unfortunately, the CF likelihood does not integrate to 1; its normalizer $Z(\theta)$ is an intractable function of the parameters θ. While the normalizer *is* very close to 1 for values of θ close to the maximum likelihood estimate (MLE), it can differ markedly from 1 when θ is far from the MLE. This is not a hindrance for practitioners simply seeking to determine the MLE, but can cause difficulties for Bayesian practitioners seeking to use MCMC techniques to sample from the posterior distribution of θ (which is proportional to the product of the likelihood $\pi(\mathbf{x}|\theta)$ and prior $\pi(\theta)$). Such samplers explore the posterior distribution of θ, and may require the evaluation of the likelihood far from the MLE. Since the CF likelihood is least accurate in such instances, stability problems can arise in the MCMC sampler; for example, the MCMC sampler may get stuck in the tails of the posterior, typically when θ is far from the MLE. In DiPietro (2001), the unknown normalizer in the CF likelihood is approximated using numerical integration techniques.

Clearly, many models of interest are multivariate. While Aït-Sahalia (2007) provides an extension for multidimensional diffusions, numerical integration methods for approximating the normalization constant can often be quite slow. However, we believe that the flexibility of our proposed algorithm, described in Section 3, may allow for its use in multivariate models.

There are numerous instances in which the likelihood of interest contains a normalizer that is an intractable function of the parameters. Different (approximate) inferential approaches have been proposed in the literature; for example, see Berthelsen and Møller et al. (2003), Heikkinen and Penttinen (1999), and Bognar (2008). A method that avoids such approximations was first proposed in Møller et al. (2006) by introducing a cleverly chosen auxiliary variable into the Metropolis-Hastings (M-H) algorithm so that the normalizing constants cancel in the M-H ratio. A simpler version that avoids having to specify an appropriate auxiliary variable, and which inspired our work, is proposed in Murray et al. (2006).

2 Bayesian model details

In what follows, assume the CF likelihood can be written as

$$\pi_{CF}(\mathbf{x}|\theta) \overset{\text{def}}{=} \frac{g(\mathbf{x}|\theta)}{Z(\theta)} \tag{1}$$

where g is the known, unnormalized CF likelihood, and Z is the intractable normalizer which depends upon the parameters. The goal is to use MCMC techniques to sample from the posterior distribution

$$\pi_{CF}(\theta|\mathbf{x}) \propto \pi_{CF}(\mathbf{x}|\theta)\pi(\theta),$$

where $\pi(\theta)$ is the prior distribution on θ, and base inference upon the sampler output. If $\pi_{CF}(\mathbf{x}|\theta)$ integrated to 1, then one could simply construct a standard M-H algorithm. However, this is not the case, and therefore additional care must be taken (otherwise the limiting distribution will not be $\pi_{CF}(\theta|\mathbf{x})$). The following MCMC algorithm is capable of dealing with the intractable (non-unity) normalizers Z by generating an external variate in such a way as to cancel the intractable normalizers present in the M-H ratio. For lack of a better term, we call the following algorithm *extended* MCMC. The EMCMC algorithm *can* be shown to satisfy detailed balance, where the limiting distribution is $\pi_{CF}(\theta|\mathbf{x})$.

3 Extended MCMC (EMCMC)

The EMCMC algorithm proceeds as follows.

1. Choose a starting value $\theta^{(t)}$ where $t = 0$.
2. Propose a new value for $\theta^{(t)}$, say θ^*, from some proposal density $q(\theta^*|\theta^{(t)})$. The proposal density may update one randomly chosen component of $\theta^{(t)}$ at a time, or the proposal density may attempt to update multiple components of $\theta^{(t)}$ simultaneously.
3. Generate $\mathbf{w} = (w_1, \ldots, w_n)$ from $\pi_{CF}(\mathbf{w}|\theta^*)$ (i.e. generate $\mathbf{w}$ from the likelihood conditional on the *proposed* parameter vector θ^*). This can be accomplished by using the Euler approximation with a M-H accept/reject step.
4. Accept θ^* (i.e. set $\theta^{(t+1)} = \theta^*$) with probability

$$\begin{aligned}\mathcal{A} &= \min\left[1, \frac{\pi_{CF}(\mathbf{x}|\theta^*)}{\pi_{CF}(\mathbf{x}|\theta^{(t)})}\frac{\pi(\theta^*)}{\pi(\theta^{(t)})}\frac{q(\theta^{(t)}|\theta^*)}{q(\theta^*|\theta^{(t)})}\frac{\pi_{CF}(\mathbf{w}|\theta^{(t)})}{\pi_{CF}(\mathbf{w}|\theta^*)}\right] \\ &= \min\left[1, \frac{g(\mathbf{x}|\theta^*)/Z(\theta^*)}{g(\mathbf{x}|\theta^{(t)})/Z(\theta^{(t)})}\frac{\pi(\theta^*)}{\pi(\theta^{(t)})}\frac{q(\theta^{(t)}|\theta^*)}{q(\theta^*|\theta^{(t)})}\frac{g(\mathbf{w}|\theta^{(t)})/Z(\theta^{(t)})}{g(\mathbf{w}|\theta^*)/Z(\theta^*)}\right] \\ &= \min\left[1, \frac{g(\mathbf{x}|\theta^*)}{g(\mathbf{x}|\theta^{(t)})}\frac{\pi(\theta^*)}{\pi(\theta^{(t)})}\frac{q(\theta^{(t)}|\theta^*)}{q(\theta^*|\theta^{(t)})}\frac{g(\mathbf{w}|\theta^{(t)})}{g(\mathbf{w}|\theta^*)}\right],\end{aligned}$$

   otherwise reject θ^* and set $\theta^{(t+1)} = \theta^{(t)}$. Generating w from $\pi_{CF}(\mathbf{w}|\theta^*)$ allows the normalizing constants $Z(\cdot)$ to cancel, leaving only *tractable* elements in $\mathcal{A}$. This modified acceptance probability does *not* upset detailed balance, thus the limiting distribution remains $\pi_{CF}(\mathbf{x}|\theta)$.
5. Repeat steps (2), (3), and (4) T times (T large).

4 CIR simulation study

A simulation study was carried out to determine the efficacy of our EMCMC algorithm in the CIR model. This model was chosen because the true transition density is known (a non-central chi-square), and hence a comparison can be made between Bayesian analyses using 1) the true non-central chi-square transition density (and, hence, true likelihood) in a standard M-H sampler, 2) the unnormalized CF likelihood in a M-H sampler, and 3) the CF likelihood with correction for the normalizer using the EMCMC algorithm. The same priors and proposal densities were used throughout.

Recall that the CIR model is characterized the stochastic differential equation

$$dX_t = \beta(\alpha - X_t)dt + \sigma\sqrt{X_t}\,dW_t$$

where α is the mean reverting level, β is the speed of the process, and σ is the volatility parameter. The true transition density is characterized by a non-central chi-square distribution; the second order ($K = 2$) (unnormalized) CF likelihood approximation can be found in Aït-Sahalia (1999).

The priors were specified according to DiPietro (2001); $\pi(\theta) = \pi(\alpha, \beta, \sigma) = I_{(0,1)}(\alpha) \cdot I_{(0,\infty)}(\beta) \cdot \sigma^{-1} I_{(0,\infty)}(\sigma)$, where I denotes the indicator function. A random-scan type algorithm was used to simulate from the posterior, where a single component of θ was randomly chosen to be updated at each iteration. All samplers had the same starting points, allowing convergence and mixing behavior to be informally assessed. Most of the chosen starting points were in the tails of the posterior distribution (for most of the chains, α was started at 0.03 or 0.30, β was started at 0.07 or 0.25, and σ was started at 0.04 or 0.10).

Using the non-central chi-square transition density, a dataset with $n = 500$ points was simulated from the CIR model with $\Delta = 1.0$ (yearly data), $\alpha = 0.07$, $\beta = 0.15$, and $\sigma = 0.07$. The exact (EXT), normalized CF (NCF), and unnormalized CF (UCF) analyses each employed 10 chains of length 10,000 (not including a 1,000 iteration burn-in period). The posterior mean from each chain was computed, and the mean and standard deviation of the 10 posterior means is summarized in the top portion of Table 1. Note that all 10 of the chains in the UCF analysis ran off into the tails of the posterior distribution and became stuck (due to numerical overflows), *regardless* of starting values. Neither the EXT or NCF analyses showed such behavior.

The standard deviations in the NCF analysis are slightly inflated compared to the EXT analysis. This is entirely expected since the EMCMC algorithm has lower acceptance rates than a true M-H algorithm. The lower acceptance rate causes inflated Monte Carlo errors (i.e. more variability in the posterior means), and leads to the inflated standard deviations seen in Table 1. In and of itself, (slightly) inflated MC errors are not problematic; longer chains would eliminate the discrepancy. Considering, however, the added complexity of the EMCMC algorithm, the difference appears rela-

		EXT	NCF	UCF
$\Delta = 1$	$\alpha\|\mathbf{x}$	0.0614 (1.9e-4); *0.288*	0.0595 (3.3e-4); *0.197*	NA (NA); *NA*
	$\beta\|\mathbf{x}$	0.1562 (1.7e-3); *0.261*	0.1579 (2.5e-3); *0.225*	NA (NA); *NA*
	$\sigma\|\mathbf{x}$	0.0704 (1.5e-4); *0.356*	0.0703 (2.0e-4); *0.251*	NA (NA); *NA*
$\Delta = 1/4$	$\alpha\|\mathbf{x}$	0.0670 (7.0e-4); *0.413*	0.0671 (1.8e-3); *0.330*	0.0672 (1.9e-3); *0.428*
	$\beta\|\mathbf{x}$	0.2231 (4.0e-3); *0.570*	0.2162 (5.6e-3); *0.446*	0.2230 (5.4e-3); *0.582*
	$\sigma\|\mathbf{x}$	0.0734 (1.3e-4); *0.367*	0.0732 (1.1e-4); *0.274*	0.0734 (8.3e-5); *0.368*
$\Delta = 1/12$	$\alpha\|\mathbf{x}$	0.1165 (8.5e-2); *0.749*	0.1103 (4.1e-2); *0.651*	0.1183 (1.0e-2); *0.763*
	$\beta\|\mathbf{x}$	0.1651 (9.5e-3); *0.662*	0.1738 (9.1e-3); *0.563*	0.1662 (1.2e-2); *0.579*
	$\sigma\|\mathbf{x}$	0.0670 (7.7e-5); *0.326*	0.0669 (1.7e-4); *0.238*	0.0670 (6.7e-5); *0.320*
$\Delta = 1/52$	$\alpha\|\mathbf{x}$	0.0670 (1.9e-3); *0.378*	0.0677 (4.1e-3); *0.247*	0.0664 (2.6e-3); *0.378*
	$\beta\|\mathbf{x}$	0.2715 (8.9e-3); *0.330*	0.2688 (1.0e-2); *0.239*	0.2759 (8.5e-3); *0.325*
	$\sigma\|\mathbf{x}$	0.0694 (5.4e-5); *0.178*	0.0694 (5.2e-5); *0.132*	0.0694 (5.1e-5); *0.175*

Table 1. Mean and standard deviation (in parentheses) of the marginal posterior means from the CIR simulation study with $\Delta = 1, 1/4, 1/12,$ and $1/52$; acceptance rates for each move-type are listed in *italics*.

tively benign. The posterior means in the NCF analysis appear to mimic the output from the EXT analysis quite well.

To determine the effect of a smaller time increment Δ, a dataset with $n = 500$ was simulated from the CIR model with $\Delta = 1/4$ (quarterly data) (α, β, and σ remained unchanged). Two of the ten chains in the UCF analysis got stuck in the tails (stuck chains were not included in the tabular summaries), suggesting that the CF likelihood has some difficulty in the tails of the posterior distribution. Again, all chains from the EXT and NCF analyses behaved well. With $n = 500$ and $\Delta = 1/12$ (monthly data), one of the ten chains in the UCF analysis got stuck in the tails, while none of the chains from the EXT and NCF analyses had stability issues. Finally, when $n = 2,000$ and $\Delta = 1/52$, three of the ten chains in the UCF analysis became stuck; all chains from the EXT and NCF analyses converged and mixed well (to attain more optimal (i.e. lower) acceptance rates (see below), this last analysis used slightly more diffuse proposal densities than employed in the other analyses). The results are summarized in Table 1.

Note that the EMCMC algorithm yields slightly lower acceptance rates (the proportion of the time that a proposed change to a parameter is accepted) than a true M-H sampler. Examination of the acceptance rates in Table 1 indicates that when $\Delta = 1$, for example, the EXT analysis accepted 28.8% of the proposed $\alpha-$candidates, while the NCF analysis accepted 19.7%. The deflation in the acceptance rates, however, is not dramatically different from a true M-H algorithm (used in the EXT analysis) for all $\Delta-$values that were considered.

5 Discussion

Using the CF likelihood for Bayesian inference poses challenges for practitioners. The CF likelihood is most accurate when the parameters θ are near the MLE, but Bayesian MCMC samplers may need to explore parts of the posterior $\pi_{CF}(\theta|\mathbf{x})$ for θ distant from the MLE. As such, using the CF likelihood in a standard M-H algorithm can cause the sampler to become stuck in the tails of the posterior distribution. Using numerical integration techniques to estimate the intractable normalizer in the CF likelihood improves convergence and mixing behavior, but the resulting limiting distribution remains unclear. Auxiliary variable techniques (Møller et al. (2006)) require the specification of an auxiliary variable, but this is not a trivial matter in many instances. Our proposed EMCMC algorithm (based on work by Murray et al. (2006)) greatly improves convergence and mixing behavior, is straightforward to implement, has good acceptance rates, and the limiting distribution can be shown to be the posterior distribution. Based upon the promising results from our simulation study, it appears as though the EMCMC algorithm may be a very effective tool for the adoption of the CF likelihood in the Bayesian inferential framework.

References

Aït-Sahalia, Y. (1999): Transition Densities for Interest Rate and Other Nonlinear Diffusions. *The Journal of Finance*, *54*, 1361–1395.

Aït-Sahalia, Y. (2002): Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica*, *70*, 223–262.

Aït-Sahalia, Y. (2007): Closed-Form Likelihood Expansions for Multivariate Diffusions. Forthcoming in the *Annals of Statistics*.

Berthelsen, K.K. and Møller, J. (2003): Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, *30*, 549–564.

Bognar, M. (2008): Bayesian modeling of continuously marked spatial point patterns. Forthcoming in *Computational Statistics*.

DiPietro, M. (2001): Bayesian Inference for Discretely Sampled Diffusion Processes with Financial Applications. Ph.D. Thesis, Department of Statistics, Carnegie Mellon University.

Heikkinen, J. and Penttinen, A. (1999): Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli*, *5*, 1119–1136.

Møller, J., Pettitt, A., Berthelsen, K. and Reeves, R. (2006): An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, *93*, 451–458.

Murray, I., Ghahramani, Z. and MacKay, D. (2006): MCMC for doubly-intractable distributions. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*.

Stramer, O. and Yan, J. (2007): On simulated likelihood of discretely observed diffusion processes and comparison to closed form approximation. *Journal of Computational and Graphical Statistics*, *23*, 672–691.