Grid Based Bayesian Inference for Stochastic Differential Equation Models

Chaitanya Joshi & Simon Wilson
School of Computer Science & Statistics,
Trinity College Dublin, Ireland.

September 22, 2011

Abstract

This paper introduces a new approach of approximate Bayesian inference for Stochastic Differential Equation (SDE) models. This approach is not MCMC based and aims to provide a more efficient option for Bayesian inference on SDE models. There are two novel aspects about this approach: the first concerns the way in which the parameter space is explored and the second concerns the evaluation of the posterior. We propose two new methods to implement this approach. These methods are the Gaussian Modified Bridge Approximation (GaMBA) and its extension GaMBA- Importance sampling (GaMBA-I). This paper provides an easy to use algorithm for both these methods, discusses their consistency properties, describes examples where these methods provide efficient inference and also discusses their limitations.

Keywords: Bayesian Inference, Stochastic Differential Equations, Diffusion Processes, Gaussian Approximation

1 Introduction

Stochastic Differential Equation (SDE) models have been widely used to model real life processes in a wide range of applications from economics and finance to engineering and biology. However, such models are often non-trivial to solve and the statistical inference

on their parameters is also not straightforward in most cases.

This paper focuses on the Bayesian inference for diffusion process parameters which is a research area that has seen a lot of activity recently. However, there remains a need to develop accurate but computationally cheaper methods. We propose a new approach to approximate Bayesian inference for SDE model parameters. This approach leads to two new algorithms named as GaMBA (for '*Ga*ussian *M*odified *B*ridge *A*pproximation') and its extension called GaMBA-I (for *GaMBA-Importance sampling*). These methods are easy to implement and computationally cheaper compared to the existing alternatives.

There are two novel aspects about this work. Firstly the parameter space is explored using a set of predetermined points using a lattice. Secondly the posterior is then approximated at these pre-determined set of parameter points. This approach is inspired from the work of Rue et al. (2009) on the Integrated Nested Laplace Approximation (INLA) for Gaussian Markov Random Field (GMRF) models.

The remainder of the paper is organised as follows. Section 2 provides a brief overview of the inference methods developed for inference on SDEs. Section 3 introduces the basic idea behind GaMBA, and describes in detail how this could be actually implemented in practice. Section 4 discusses the extension GaMBA-I, and the conditions under which the posterior obtained using GaMBA-I would have the desirable convergence properties. Section 5 provides some examples to illustrate the use of GaMBA and GaMBA-I on standard SDE models, while Section 6 discusses the computational aspect. Section 7 provides a discussion on various practical aspects concerning GaMBA and GaMBA-I including their limitations. Finally, we conclude with a succinct summary in Section 8.

2 Bayesian Inference on SDE models

The general form of a one-dimensional SDE can be expressed as:

$$dX_t = f(X_t, \mu)\, dt + g(X_t, \sigma)\, dW_t, \quad X(t_0) = x_0 \tag{1}$$

where $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the *drift* coefficient of the SDE, $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the *diffusion* coefficient of the SDE, and W_t is a one-dimensional process having independent scalar Wiener Process components. The stochastic process $\{X_t(w)\}$ defined on the probability space $(\Omega, \mathfrak{F}_t, P)$, which satisfies Equation (1) is called a (Itô) *diffusion process*. If such a process exists, then it is a continuous time Markov process with continuous sample paths *a.s.* Its transition density is governed by the parameters $\Theta = \{\mu, \sigma\}$ of the SDE. Note that μ and σ could be vector valued.

The common theme underlying many of the methods developed thus far to obtain Bayesian inference on discretely observed diffusions is to interpret the inference problem as that of a hidden Markov model where the unobserved Markov process forms the set of latent variables. These are Monte-Carlo based methods which rely on drawing large number of samples of both the latent variables and the diffusion process parameters. Many of these methods sample only the discretised paths of the unobserved diffusion. Some others can sample the continuous paths. A lot of research has focused on two main aspects: (i) developing Monte-Carlo algorithms which get around the problem of dependency between the latent variables and the diffusion parameters, and (ii) developing more efficient and more accurate proposals for the latent variables which in turn results in more accurate and efficient inference on the diffusion process parameters.

The inference set-up used by Bayesian methods based on time discretisation is as follows. The diffusion process $\{X_t\}$is observed at time points $t_0, t_1, \cdots, t_n$ and $\mathbf{Y} = \{y_0, y_1, \cdots, y_n\}$ denote these observations. Our objective is to estimate $P(\Theta|\mathbf{Y})$. If the transition density of X_t is known, then the posterior distribution $P(\Theta|\mathbf{Y})$ is given by

$$P(\Theta|\mathbf{Y}) \quad \propto \quad P(\mathbf{Y}|\Theta) \cdot P(\Theta) \propto P(y_0|\Theta) \prod_{i=0}^{n} P(y_{i+1}|y_i, \Theta) \cdot P(\Theta) \tag{2}$$

where $P(y_{i+1}|y_i, \Theta)$ is the transition density of the process X_t going from y_i to y_{i+1} in time Δt.

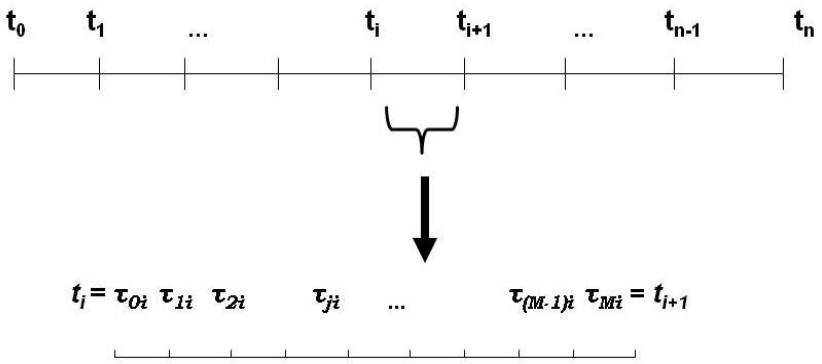

Figure 1: Latent data points between each pair of observed data.

Except in a few isolated cases, the transition densities for diffusion processes are not available in closed form, and therefore posterior inference using Equation (2) is non-trivial. The standard approach then is to approximate the true transition density using the density of some numerical method such as, the first order Euler's method. However, Δt is usually not small enough for such an approximation to be reasonably accurate.

The most intuitive approach (see Figure 1) to get around this problem is to impute $M-1$ latent variables between every pair of consecutive observations. These latent variables are simulated using some simulation mechanism (e.g. the Euler's method). This is typically achieved by partitioning each interval $[t_i, t_{i+1}]$ into M equal parts as $t_i = \tau_{0_i} < \tau_{1_i} < \cdots < \tau_{M_i} = t_{i+1}$, such that $\tau_{(j+1)_i} - \tau_{j_i} = \delta_\tau$ is small enough for the simulation mechanism to be reasonably accurate. Equal spacing is not necessary but has been used only to make the notation easier. Let $\mathbf{X^{(i)}} = X_{1_i}, X_{2_i}, \ldots, X_{(M-1)_i}$ denote the latent variables corresponding to times $\tau_{1_i} < \tau_{2_i} < \cdots < \tau_{(M-1)_i}$, between the observations y_i and y_{i+1} for $i = 0, 1, \cdots, n-1$.

These X_{j_i}'s for $i = 0, \cdots, n, \quad j = 1, \cdots, M-1$, could now be approximated by simulation using Euler's method. The paths simulated using Euler's method converge *weakly* to the underlying true diffusion as $M \to \infty$, (see, for example, Kloeden & Platen

(1992)). The transition density for Euler's approximation is given by

$$P_E(X_{(j+1)_i}|X_{j_i},\Theta) = N(\mu_{eu},\sigma^2_{eu}),\ \ j=0,1,...,(M-2) \tag{3}$$

where

$$\mu_{eu} = X_{j_i} + f(X_{j_i},\mu)\cdot\delta_\tau$$

$$\sigma_{eu} = g(X_{j_i},\sigma)\cdot\delta_\tau^{1/2}.$$

The inference problem is thus translated into a data augmentation model with parameters Θ, the latent variables $\mathbf{X} = \{X_{j_i}\}$ for $i = 0,\cdots,n,\quad j = 1,\cdots,M-1$, and the observed data $\mathbf{Y}$.

The basic idea behind the early MCMC schemes proposed for Bayesian inference on SDE models, such as for example Elerian et al. (2001) was as to use the standard Gibbs type methods where one samples alternately from $P(\mathbf{X}|\Theta,\mathbf{Y})$ and $P(\Theta|\mathbf{X},\mathbf{Y})$ in each MCMC iteration until the chains converge. The proposal density for $\mathbf{X}$ was obtained by approximating the target density at the mode using a multivariate Gaussian or a multivariate t-distribution using the Newton-Raphson method. Chib & Shephard (2002) proposed implementing this algorithm by using the density of the modified Brownian bridge (MBB) construct proposed by Durham & Gallant (2002) as a proposal distribution for the latent variables $\mathbf{X}$. It is not immediately clear why such an approach should not yield good results.

However, Roberts & Stramer (2001) pointed out that this most intuitive Bayesian imputation approach can worsen linearly with M. This is because the quadratic variation of the diffusion converges $a.s$ to a function of the diffusion coefficients, i.e,

$$\lim_{M\to\infty}\sum_{j=1}^{M}(X_{j_i} - X_{(j-1)_i})^2 = t_i\cdot g(\sigma)^2 \tag{4}$$

$w.p.$ 1. Therefore, as $M\to\infty$ the data augmentation scheme in Gibbs sampling based methods is reducible; the imputed data merely confirm the current value of the diffusion coefficient which then remains unaltered since it is in turn determined by the quadratic variation of the sample path of $\mathbf{X}$. This can result in arbitrarily slow rates of convergence

for such MCMC methods.

To avoid this dependency problem Roberts & Stramer (2001) proposed transforming the SDE to obtain an SDE with a unit diffusion coefficient and then implementing the MCMC scheme on this transformed process. Though this approach gets around the problem of dependency between $\mathbf{X}$ and the diffusion coefficients, the required transformation may not be possible for many non-linear multivariate SDEs.

Chib et al. (2006) proposed an innovative solution to eliminate the issue of dependency. They suggested sampling the Wiener process components W_t instead of the latent variables $\mathbf{X}$ and then constructing the $\mathbf{X}$'s deterministically using the Euler approximation. Since the Wiener process is independent of the diffusion parameters σ, there is no dependency and the MCMC scheme can now be used without the need of any transformation.

Golightly & Wilkinson (2007) instead used the modified Brownian bridge (MBB) construct to deterministically obtain $\mathbf{X}$'s instead of using the Euler's method. The advantage of doing this, they argue, is that unlike Euler's method, the paths generated by MBB do not miss the observations and hence this becomes a more efficient method of reparameterizing the latent variables.

Both these methods claim to in fact improve with M rather than worsen. Though these methods do get around the problem of dependency, it is at the cost of added complexity to the existing MCMC algorithms, and they are as a result are computationally expensive.

Filtering based methods have also been proposed; see e.g. Del Moral et al. (2002), Golightly & Wilkinson (2006) and Sarkka & Sottinen (2008). These methods have an advantage that the inference framework does not need to be restarted from scratch as new data becomes available. However, computation efficiency of such methods is an ongoing research problem (Golightly & Wilkinson (2006)). Also, methods which use a diffusion specific importance sampler (e.g Sarkka & Sottinen (2008)) are less widely applicable than

some of the methods described above.

Finally, it is important to note that methods which can simulate from the exact density of the diffusion have been proposed (Beskos et al. (2006), Beskos et al. (2008), and Fernhead et al. (2010)). These methods can be applied to a wide range of (though not all) diffusion processes. The advantage that they have over the time-discretisation based methods described above is that these methods are free of the discretisation error introduced by discretising the underlying continuous time process.

3 GaMBA

GaMBA uses the same data augmentation set-up described in Section 2. For such a set-up, the joint posterior can be expressed as:

$$P(\mathbf{X}, \Theta \,|\mathbf{Y}) \propto P(\mathbf{Y}, \mathbf{X} \,|\Theta) \cdot P(\Theta), \tag{5}$$

where $P(\Theta)$ is an appropriate prior distribution. The joint posterior can be factorised as:

$$P(\mathbf{X}, \Theta \,|\mathbf{Y}) = P(\mathbf{X} \,|\Theta, \mathbf{Y}) \cdot P(\Theta \,|\mathbf{Y}). \tag{6}$$

Thus, using Equations (5) and (6), the posterior $P(\Theta|\mathbf{Y})$ can be approximated, up to the proportionality constant by using the following identity:

$$P(\Theta|\mathbf{Y}) \quad \propto \quad \frac{P(\mathbf{Y}, \mathbf{X}|\Theta) \cdot P(\Theta)}{P(\mathbf{X}|\Theta, \mathbf{Y})}, \tag{7}$$

valid for any $\mathbf{X} \sim P(\mathbf{X}|\Theta, \mathbf{Y})$. Equation (7) can be used to obtain Bayesian inference on parameters Θ. However, evaluating Equation (7) analytically is usually too complicated in practice.

Let Θ^* be the (unknown) modal value of the posterior density $P(\Theta|\mathbf{Y})$, and Ξ^* be the subspace of the parameter space Ξ containing Θ^* such that Ξ^* such that $\int_{\Xi \setminus \Xi^*} P(\Theta|Y)\ d\Theta \leq \epsilon$ for some pre-determined small value $\epsilon \geq 0$. If for a given value of Θ, say Θ^o, it is possible

to compute the right hand side of Equation (7), then the posterior probability of Θ^o given the observed data $\mathbf{Y}$ can thus be obtained up to a proportionality constant. Repeating this procedure for a large sample of Θ values carefully sampled from the subspace Ξ^* would thus give an approximation to the unknown posterior density $P(\Theta|\mathbf{Y})$. One way to efficiently sample from Ξ^* is by using a grid-sampling method; i.e. by approximating the continuous Ξ^* space using a grid $\mathcal{G}_{\Xi^*}$ consisting of finite number of points. This is the basic idea behind GaMBA. This basic idea can be summarised in the following procedure.

GaMBA : Inference Procedure

1. Determine Ξ^*.
2. Define a discrete grid $\mathcal{G}_{\Xi^*}$ on Ξ^*.
3. For each point on the grid $\Theta_j \in \mathcal{G}_{\Xi^*}$,
   - Sample $\mathbf{X}$ from $P(\mathbf{X}|\Theta_j, \mathbf{Y})$,
   - For this $\mathbf{X}$ evaluate

$$P(\Theta_j|\mathbf{Y}) \propto \left. \frac{P(\mathbf{Y}, \mathbf{X}|\Theta_j) \cdot P(\Theta_j)}{P(\mathbf{X}|\Theta_j, \mathbf{Y})} \right|_{\mathbf{X}}$$

     .
4. Normalise to obtain $P(\Theta|\mathbf{Y})$ over $\mathcal{G}_{\Xi^*}$.

While trying to put this simple idea into practice though, the following practical considerations need to be made.

3.1 Identifying Ξ^*

In the implementation of GaMBA, identifying Ξ^* is a very important non-trivial step. If there is credible prior information, then Ξ^* can be determined using this information. In the absence such information, one approach to determine Ξ^* can possibly be to first find the mode of Equation (7) using some numerical method such as for example, Newton's method,

and then approximate $P(\Theta|\mathbf{Y})$ using Laplace's approximation around this mode by computing the Hessian. It is then possible to determine Ξ^* such that $\int_{\Xi \setminus \Xi^*} P(\Theta|Y)\, d\Theta \leq \epsilon$ for some pre-determined small value $\epsilon \geq 0$. Such a procedure has been used to obtain Ξ^* for latent GMRF models by Rue et al. (2009). Note that, this is only one way to determine Ξ^*; there possibly can not be just one ŕight' Ξ^*. Also note that, this method may not be suitable when $P(\Theta|\mathbf{Y})$ is likely to be multi-modal.

Identification of Ξ^* is a very important research problem in itself, but this is not the focus of this research work. The main aim of this research work is to explore how step (3) of the inference procedure mentioned above can be implemented for SDE models. The key question that this paper aims to answer is, how to evaluate $P(\mathbf{Y}, \mathbf{X}|\Theta)$ and $P(\mathbf{X}|\Theta, \mathbf{Y})$ either exactly or approximately? Further, if an approximation is possible, then how its accuracy could be determined? In what follows, some possible answers to these questions are suggested.

For the purposes of this paper we assume that credible prior information is available and that it is possible to identify Ξ^* based on this information.

3.2 Evaluating $P(\mathbf{Y}, \mathbf{X}|\Theta)$

The complete likelihood $P(\mathbf{Y}, \mathbf{X}|\Theta)$ can be factorised as

$$P(\mathbf{Y}, \mathbf{X}|\Theta) = P(y_0|\Theta) \cdot \prod_{i=1}^{n} P(y_i|X_{M-1_{i-1}}, \Theta) \cdot \prod_{i=1}^{n} P(X_{1_{i-1}}|y_{i-1}, \Theta) \cdot \prod_{i=1}^{n} \prod_{j=2}^{M-1} P(X_{j_{i-1}}|X_{j-1_{i-1}}, \Theta). \tag{8}$$

Since, y_0 is considered observed, $P(y_0|\Theta)$ can be considered as constant and thus

$$P(\mathbf{Y}, \mathbf{X}|\Theta) \propto \prod_{i=1}^{n} P(y_i|X_{M-1_{i-1}}, \Theta) \cdot \prod_{i=1}^{n} P(X_{1_{i-1}}|y_{i-1}, \Theta) \cdot \prod_{i=1}^{n} \prod_{j=2}^{M-1} P(X_{j_{i-1}}|X_{j-1_{i-1}}, \Theta). \tag{9}$$

Equation (9) can be evaluated exactly if the exact transition density is known. However in many cases, the exact transition density is not available, and in such cases, Equation (9)

can be approximated using Euler's density up to the unknown constant $P(y_0|\Theta)$ as

$$P(\mathbf{Y},\mathbf{X}|\Theta) \approx \prod_{i=1}^{n} P_E(y_i|X_{M-1_{i-1}},\Theta)\cdot \prod_{i=1}^{n} P_E(X_{1_{i-1}}|y_{i-1},\Theta)\cdot\prod_{i=1}^{n}\prod_{j=2}^{M-1} P_E(X_{j_{i-1}}|X_{j-1_{i-1}},\Theta), \quad (10)$$

where P_E is the Euler density of Equation (3).

3.3 Sampling $\mathbf{X} \sim P(\mathbf{X}|\Theta,\mathbf{Y})$

Equation (7) is valid for any $\mathbf{X}$ sampled from its full conditional distribution $P(\mathbf{X}|\Theta,\mathbf{Y})$. However, for most of the SDE models used in practice, $P(\mathbf{X}|\Theta,\mathbf{Y})$ and $P(\mathbf{Y},\mathbf{X}|\Theta)$ are not known in closed form and need to be approximated. $P(\mathbf{Y},\mathbf{X}|\Theta)$ can be approximated using the Euler approximation as described above. It can be shown that an approximation to $P(\mathbf{X}|\Theta,\mathbf{Y})$ obtained by conditioning the Euler density can be applicable to only a very limited class of SDEs and therefore a different approximation (say P_B) is required instead. Sampling an $\mathbf{X}$ randomly from $P_B(\mathbf{X}|\Theta,\mathbf{Y})$ can make the inference unreliable in the sense that repeating GaMBA on the same set of data and same Ξ^* can yield different posteriors. This is specially true when P_B and P_E do not have the same mode and also if the tails of P_B are not thicker than those of P_E. Therefore, as a practical consideration, it is desirable to identify a mechanism of sampling $\mathbf{X}$ which results in reliable inference using GaMBA.

This paper explores two possible ways in which this could be achieved. The first approach of doing this as described in Section 3.4 is to choose the $\mathbf{X}$ which maximises $P_B(\mathbf{X}|\Theta,\mathbf{Y})$ over $\mathbf{X}$, i.e. the modal value X^* of the density $P_B(\mathbf{X}|\Theta,\mathbf{Y})$. This approach is referred to as GaMBA. Not only is this approach computationally cheaper than the second approach, but it also has an additional advantage that when choosing not to 'integrate out' the $\mathbf{X}$'s, it is intuitively appealing to instead evaluate the expressions at the modal value of $P_B(\mathbf{X}|\mathbf{Y},\Theta)$ rather than at any random X. The reader is referred to Rue et al. (2009) where a similar approach has been successfully used.

The second approach considered in this paper is to sample several (say K) $\mathbf{X}$'s instead of just one and then approximate the likelihood using the average

$$\frac{1}{K}\sum_{k=1}^{K}\frac{P_E(\mathbf{Y},\mathbf{X}_k|\Theta)}{P_B(\mathbf{X}_k|\Theta,\mathbf{Y})}.$$

This approach is essentially GaMBA implemented using Importance sampling and hence is referred to as GaMBA-I. It is described in Section 4. Though this approach is computationally less efficient compared to the first one, it is found to provide more accurate inference and it is also possible to prove the consistency properties of GaMBA-I if certain conditions are met. Note that GaMBA-I is a stochastic approach unlike GaMBA which is deterministic.

3.4 Evaluating $P(\mathbf{X}|\Theta,\mathbf{Y})$

Because of the Markovian nature of a diffusion process, each observation is conditionally independent to other observations given its previous observation. Thus when concerned with the distribution of $P(\mathbf{X}|\mathbf{Y},\Theta)$, one is in fact dealing with n independent discretised diffusion bridges, each conditioned only on the corresponding pair of successive observations $\{y_i, y_{i+1}\}$. Using these discretised bridges, $P(\mathbf{X}|\Theta,\mathbf{Y})$ can be factorised as

$$P(\mathbf{X}|\mathbf{Y},\Theta)=\prod_{i=0}^{n-1}P(\mathbf{X}^{(i)}|y_i,y_{i+1},\Theta). \tag{11}$$

Since the SDE is assumed to be time-homogeneous, each of the $(\mathbf{X}^{(i)}|y_i,y_{i+1},\Theta)$ are also identically distributed. Therefore, without loss of generality and for notational ease, only the variables corresponding to the second diffusion bridge will be considered here for the deliberation on how to approximate $P(\mathbf{X}|\Theta,\mathbf{Y})$. While $\{y_1,y_2\}$ are the observations corresponding to this bridge, $\{X_1,\cdots,X_{M-1}\}$ denote the corresponding missing variables of the discretised diffusion bridge with $X_0=y_1$ and $X_M=y_2$ (see Figure 2). Thus, the question 'how to evaluate $P(\mathbf{X}|\Theta,\mathbf{Y})$ in Equation (7)?' will be answered by elaborating how possibly $P(X_1,\cdots,X_{M-1}|y_1,y_2,\Theta)$ can be evaluated.

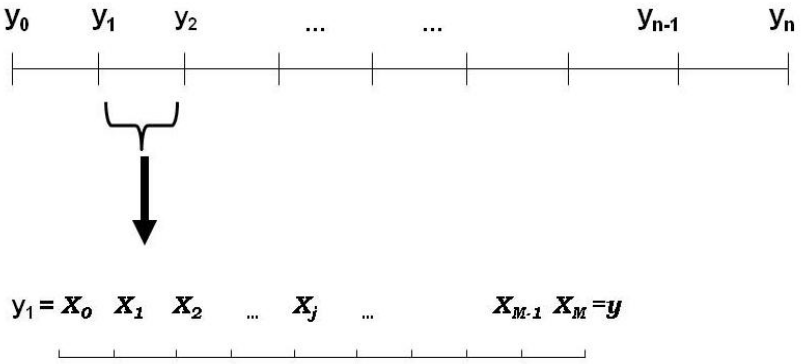

Figure 2: Second discretised diffusion bridge.

Note that using the Markovian property of the diffusion process $P(X_1, \cdots, X_{M-1}|y_1, y_2, \Theta)$ can be written as

$$P(X_1, \cdots, X_{M-1}|y_1, y_2, \Theta) = \quad P(X_1|y_1, y_2, \Theta) \cdot P(X_2|X_1, y_2, \Theta) \cdots P(X_{M-1}|X_{M-2}, y_2, \Theta). \tag{12}$$

It is important to note that the above simplification is possible because the data have been assumed to be observed without error. Such simplification may not be possible if the data were observed with errors. In that case the dependence structure between the errors and their distributional assumptions would also have to be taken in to account.

Equation (12) can be written as

$$P(X_1, \cdots, X_{M-1}|y_1, y_2, \Theta) = \prod_{j=1}^{M-1} P(X_j|X_{j-1}, X_M, \Theta) \tag{13}$$

where $X_0 = y_1$ and $X_M = y_2$. In order to implement GaMBA, $P(X_1, \cdots, X_{M-1}|y_1, y_2, \Theta)$ needs to be evaluated either exactly or approximately. However, note that this also involves sampling (simulating) X_j from $P(X_j|X_{j-1}, X_M, \Theta)$ for every j.

There might be different ways in which this could be achieved. It would be ideal to have a diffusion bridge construct, which will be easy to implement, be applicable to a wider class of diffusion processes, computationally not too expensive and be reasonably

accurate as well. The Modified Brownian Bridge, a construct proposed by Durham & Gallant (2002) is one such construct and we use it to approximate the density in Equation (13).

Using the above notation, $P(X_j|X_{j-1}, X_M, \Theta)$ can be approximated using the MBB as

$$P_{MBB}(X_j|X_{j-1}, X_M, \Theta) \approx N_{X_j}(\mu_{mbb}, \sigma^2_{mbb}), \tag{14}$$

where

$$\mu_{mbb} = X_{j-1} + \left(\frac{X_M - X_{j-1}}{\tau_M - \tau_{j-1}}\right)\delta_\tau,$$

$$\sigma_{mbb} = g(X_{j-1}, \sigma)\sqrt{\left(\frac{M-j}{M-j+1}\right)\delta_\tau}.$$

where the subscript MBB has been used to denote the density corresponding to the MBB.

Note that the MBB construct does not have the desirable asymptotic properties; i.e the approximation $P_{MBB}(X_j|X_{j-1}, X_M, \Theta)$ would not converge to the desired conditional distribution $P(X_j|X_{j-1}, X_M, \Theta)$ even as $M \to \infty$. However, it is widely applicable and has been successfully used as a proposal distribution for sampling diffusion bridges (Durham & Gallant (2002),Chib et al. (2006),and Golightly & Wilkinson (2007)). MBB can be applied to all one dimensional SDE's and extensions to multivariate SDEs is possible.

For the purposes of implementing GaMBA, MBB density can be used to approximate Equation (13) as

$$P(X_1, \cdots, X_{M-1}|y_1, y_2, \Theta) \approx \prod_{j=1}^{M-1} P_{MBB}(X_j|X_{j-1}, X_M, \Theta) \tag{15}$$

where P_{MBB} is the density of Equation (14). To ensure reliable inference (as described in Section 3.3), X_j can also be chosen to be X_j^* – the modal value of $P_{MBB}(X_j|X_{j-1}, X_M, \Theta)$.

It is important to note that because the mean of the MBB construct (i.e X_j^*) is just the linear interpolation between the two neighbouring data points, the posterior obtained

using GaMBA may not be very accurate for non-linear processes. This is a limitation of GaMBA and will be discussed later.

3.5 Implementing GaMBA

GaMBA can thus be summarised into the following procedure.

GaMBA algorithm :

Implement the GaMBA inference procedure, where step 3 is implemented as follows:

3. For each point on the grid $\Theta_j \in \mathcal{G}_{\Xi^*}$,

   (a) Sampling $\mathbf{X} \sim P_{MBB}(\mathbf{X}|\Theta_j, \mathbf{Y})$: choose $\mathbf{X}$ to be

   $$\mathbf{X}^* = \arg\max_{X}(P_{MBB}(\mathbf{X}|\Theta_j, \mathbf{Y}))$$

   as described in Section 3.4

   (b) For this $\mathbf{X}$ evaluate

   $$P_{GaMBA}(\Theta_j|\mathbf{Y}) \propto \frac{P_E(\mathbf{Y}, \mathbf{X}^*|\Theta_j) \cdot P(\Theta_j)}{P_{MBB}(\mathbf{X}^*|\Theta_j, \mathbf{Y})}$$

   .

where:

- $P_E(\mathbf{Y}, \mathbf{X}|\Theta_j)$ is the Euler's density as in Equation (10);
- $P(\Theta_j)$ a suitable prior density;
- $P_{MBB}(\mathbf{X}|\mathbf{Y}, \Theta_j)$ is the MBB density as in Equation (15).

Note that, the posterior thus obtained will be a 'discretised' approximation to the true continuous posterior. If desired, a *continuous looking* posterior can be obtained by using standard kernel *smoothing* techniques which are easily available in any statistical package.

4 Link to Importance Sampling

The approach taken by GaMBA is to approximate the likelihood as

$$P(\mathbf{Y}|\Theta) \approx \frac{P_E(\mathbf{Y}, \mathbf{X}|\Theta)}{P_{MBB}(\mathbf{X}|\mathbf{Y}, \Theta)}. \tag{16}$$

Thus an approximation to $P(\mathbf{Y}|\Theta)$ is evaluated using Equation (16) for every predetermined value of Θ, by sampling (or choosing) $\mathbf{X}$ ***once*** from $P_{MBB}(\mathbf{X}|\mathbf{Y}, \Theta)$. Instead, a better approximation, could possibly be obtained by sampling multiple (say K) values of $\mathbf{X}$ for every predetermined value of Θ, and then evaluating

$$P(\mathbf{Y}|\Theta) \approx P_{M,K}(\mathbf{Y}|\Theta)\frac{1}{K}\sum_{k=1}^{K}\frac{P_E(\mathbf{Y}, \mathbf{X}_k|\Theta)}{P_{MBB}(\mathbf{X}_k|\mathbf{Y}, \Theta)}, \tag{17}$$

where $P_{M,K}$ denotes the approximation to the likelihood obtained for given M and K.

4.1 GaMBA-I (GaMBA with Importance Sampling)

GaMBA algorithm can be extended using this modification. This extended algorithm is as follows:

GaMBA-I algorithm :

Implement the GaMBA inference procedure, where step 3 is implemented as follows:

3. For each point on the grid $\Theta_j \in \mathcal{G}_{\Xi^*}$

   (a) for $k = 1, \cdots, K$,

   Sample $\mathbf{X}_k \sim P_{MBB}(\mathbf{X}_k|\Theta_j, \mathbf{Y})$ as described in Section 3.4

   (b) Evaluate

$$P_{GaMBA-I}(\Theta_j|\mathbf{Y}) \propto \frac{1}{K}\sum_{k=1}^{K}\frac{P_E(\mathbf{Y}, \mathbf{X}_k|\Theta_j)\cdot P(\Theta_j)}{P_{MBB}(\mathbf{X}_k|\Theta_j, \mathbf{Y})}$$

   .

Again, note that the posterior thus obtained will be a 'discretised' approximation to the true continuous posterior. If desired, a *continuous looking* posterior can be obtained

by using standard kernel *smoothing* techniques.

Note that the approximation to the likelihood in Equation (17) is same as the one used in the Simulated likelihood method proposed by Durham & Gallant (2002) when the MBB density is used as the importance density. Thus, GaMBA-I (*GaMBA - Importance sampling*) can be seen as the novel implementation of the simulated likelihood method proposed by Durham & Gallant (2002) used in the Bayesian context. The novelty being that the parameter space is explored using a predetermined set of points chosen using a grid over Ξ^*. On the other hand the approach taken so far (Pedersen (1995),Santa-Clara (1995),Brandt & Santa-Clara (2002) and Durham & Gallant (2002)) has been to explore the parameter space using numerical optimisation methods such as a Newton-Raphson method, for example.

4.2 Convergence of GaMBA-I

By establishing link with the Importance sampling, GaMBA-I creates a possibility for exploring its consistency properties. Some consistency properties are already known. We make the following assumptions.

- As for all statistical methods, it is assumed that a *weak* solution to Equation (1) exists. In particular, it is assumed that the $f(\cdot)$ and $g(\cdot)$ satisfy the following:

  A 1 **:** For any $R > 0$ and all $x, y \in \mathbb{R}^p$ such that $|x| < R$, $|y| < R$, and $t \in [0, T]$, there exists a constant $K_R < \infty$ such that

  $$|f(x,t) - f(y,t)| + |g(x,t) - g(y,t)| < K_R|x - y|,$$

  A 2 **:** For all $x, y \in \mathbb{R}^p$ and $t \in [0, T]$, there exists a constant $C < \infty$ such that

  $$|f(x,t)| + |g(x,t)| < C(1 + |x|).$$

- It is also assumed that the following holds:

  A 3 : (**i**) : f and g are time homogeneous, *i.e.* $f(X_t, \Theta) = f(X, \Theta)$, and $g(X_t\Theta) = g(X, \Theta)$.

(**ii**) : $f(X,\Theta)$ and $g(X,\Theta)$ are bounded with bounded derivatives of any order.

(**iii**) : $a(X,\Theta) = g(X,\Theta)\cdot g(X,\Theta)'$ is strongly positive definite, that is there exists an $\epsilon(\Theta) > 0$ such that $a(X,\Theta) - \epsilon(\Theta)\, I_p$ is a non-negative definite for all $X \in \mathbb{R}^p$.

A 4 : For every $\Theta \in \Xi$ and for every $\mathbf{X} \sim P_{MBB}(\mathbf{X}|\Theta,\mathbf{Y})$, the following holds

$$E\left[\frac{P_E(\mathbf{Y},\mathbf{X}|\Theta)}{P_{MBB}(\mathbf{X}|\Theta,\mathbf{Y})}\right] < \infty$$

Then, under assumption **A** 4, $P_{M,K}(\mathbf{Y}|\Theta) \overset{a.s}{\to} P(\mathbf{Y}|\Theta)$ as $M \to \infty$ and $K \to \infty$ under the strong law of large numbers (Geweke (1989)) as long as the support of P_{MBB} contains the support of P_E.

Further, under the above assumptions Stramer & Yan (2007) have shown that for SDE's with unit diffusion,

1. the variability associated with the Importance sampling estimate is uniformly bounded in M, and
2. the total error in this estimation is of the order $O(1/M) + O(1/\sqrt{K})$.

Since most one dimensional SDEs with non-constant diffusion coefficients could be transformed using the Itô's formula into SDEs with unit diffusion coefficient, the result above is applicable to such SDEs as well.

Based on the results above, Stramer & Yan (2007) suggest choosing $K = M^2$ as an asymptotically optimal choice. However, they also point out that these asymptotic results depend on how closely the data are observed and in practice, may take effect only for very high values of M.

For SDEs which can not be transformed into ones with unit diffusion coefficients, a general result such as the one above does not yet exist.

5 Examples

5.1 Two parameter Ornstein-Uhlenbeck (O-U) Process

The two parameter Ornstein-Uhlenbeck process is the solution to the stochastic differential equation,

$$dX_t = -\theta_2 X_t\, dt + \theta_3\, dW_t \;\; X_0 = x_0 \tag{18}$$

with $\theta_2 \in \mathbb{R}$, and $\theta_3 \in \mathbb{R}^+$.

The transition density for this process is Gaussian and hence true posterior distribution can be determined.

$$X_t = x_0 e^{-\theta_2 t} + \theta_3 \int_0^t e^{-\theta_2 (t-u)}\, du. \tag{19}$$

A dataset of 500 observations was simulated using parameter values $\theta_2 = 0.1$, $\theta_3 = 0.25$ and starting value of $Y_0 = 0$. This dataset was simulated using Euler's method with $M = 10$ and $\delta_{tau} = 0.1$ (i.e $\Delta_t = 1$). Figure 3 shows the simulated data. The parameter space $\Xi^* = \Xi_2^* \times \Xi_3^*$ was chosen based on prior knowledge as $\Xi^* = [-0.1, 0.3] \times [0.1, 0.4]$, and a grid $\mathcal{G}_{\Xi^*}$ was considered with $\Delta\Xi_2 = 0.01$ and $\Delta\Xi_3 = 0.01$. GaMBA was implemented on the 1271 points thus sampled from Ξ^* and marginal posteriors distributions were obtained.

In addition, GaMBA-I was also implemented on the same $\mathcal{G}_{\Xi^*}$ as above for various M and K. Table 1 lists the mean squared error (MSE) obtained for each of the parameters using different methods for inference. Asymptotic properties of the GaMBA-I posterior can be seen. Figure 4 plots the marginal posteriors obtained using GaMBA-I (for [$M = 5$,$K = 10$] and [$M = 10$,$K = 25$]) along with those obtained used GaMBA and the true likelihood. It shows that for θ_2, GaMBA-I works much better than GaMBA, and also that as M and K increases posteriors obtained using GaMBA-I become more accurate.

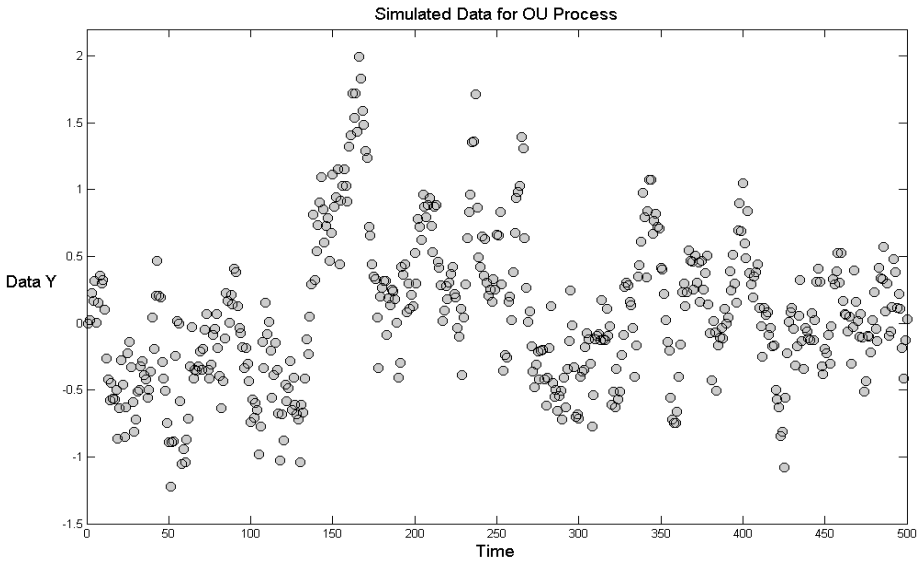

Figure 3: Simulated data for O-U process.

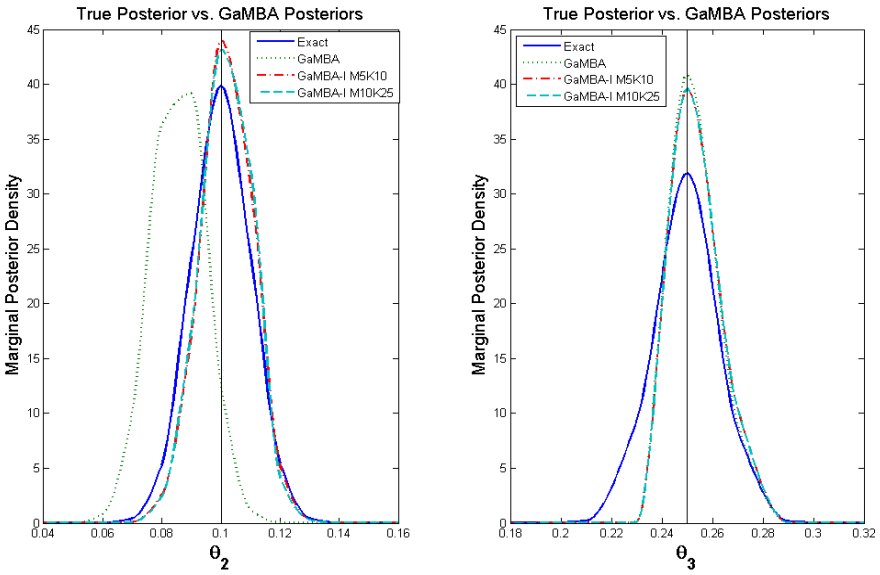

Figure 4: O-U: Marginal True posterior (line) vs. posterior using GaMBA $M = 5$ (dotted line). Vertical lines denote the true values $\theta_2 = 0.1$ and $\theta_3 = 0.25$.

Table 1: O-U: MSE obtained for θ_2 & θ_3 for the posteriors obtained using different methods

$Method$	M	K	MSE θ_2	MSE θ_3
$TrueLikelihood$	-	-	1.32×10^{-5}	1.75×10^{-5}
$GaMBA$	5	-	1.83×10^{-5}	9.02×10^{-6}
$GaMBA-I$	5	10	1.45×10^{-5}	1.09×10^{-5}
$GaMBA-I$	5	25	1.21×10^{-5}	1.03×10^{-5}
$GaMBA-I$	10	10	1.23×10^{-5}	1.03×10^{-5}
$GaMBA-I$	10	25	9.74×10^{-6}	9.78×10^{-6}

5.2 Cox-Ingersoll-Ross (CIR) Process

The Cox-Ingersoll-Ross (CIR) Process is the solution to the stochastic differential equation:

$$dX_t = (\theta_1 - \theta_2 X_t)\, dt + \theta_3 \sqrt{X_t}\, dW_t \tag{20}$$

with θ_1, θ_2, $\theta_3 \in \mathbb{R}^+$. The transition density for this process is known and is a non-central Chi-squared distribution. This process is used in financial applications to model interest rates.

A dataset of 100 observations was simulated using parameter values $\theta_1 = 1$, $\theta_2 = 0.5$, $\theta_3 = 0.2$ and starting value of $y_0 = 2.5$. This dataset was simulated using Euler's method with $M = 10$ and $\delta_{tau} = 0.1$ (i.e $\Delta_t = 1$). Figure 5 shows the simulated data. The parameter space $\Xi^* = \Xi_1^* \times \Xi_2^* \times \Xi_3^*$ was chosen based on prior knowledge as $\Xi^* = [0.1, 1.9] \times [0.1, 0.8] \times [0.1, 0.3]$, and a grid $\mathcal{G}_{\Xi^*}$ was considered with $\Delta\Xi_1 = 0.1$, $\Delta\Xi_2 = 0.05$ and $\Delta\Xi_3 = 0.02$. GaMBA was implemented on the 3135 points thus sampled from Ξ^* and marginal posteriors distributions were obtained.

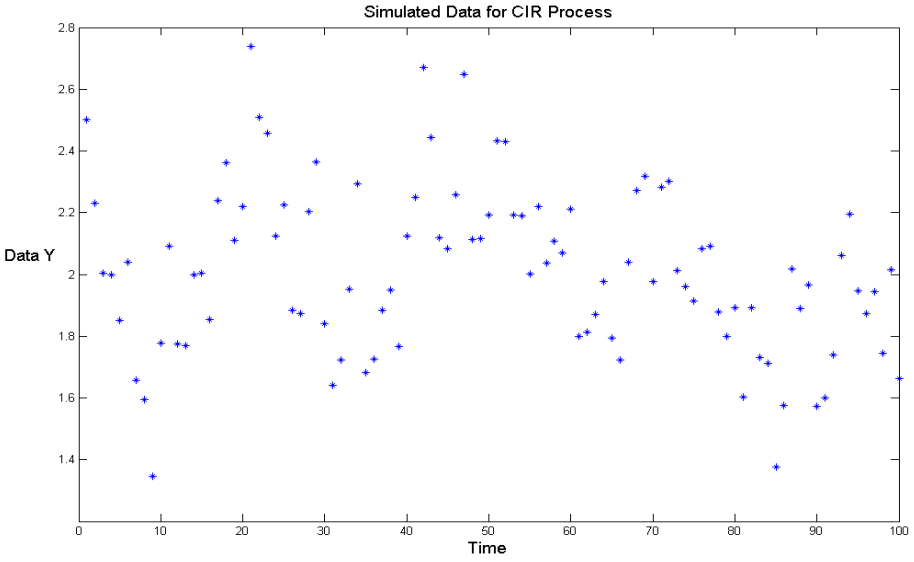

Figure 5: Simulated data for CIR process.

Table 2: CIR: MSE obtained for θ_1, θ_2 & θ_3 for the posteriors obtained using different methods

$Method$	M	K	MSE θ_1	MSE θ_2	MSE θ_3
$GaMBA$	5	-	0.11623	0.034306	0.000576
$GaMBA-I$	5	5	0.05767	0.01799	0.000575
$GaMBA-I$	5	25	0.04311	0.01342	0.0003975
$GaMBA-I$	10	5	0.0455	0.01394	0.000401
$GaMBA-I$	10	25	0.041	0.01232	0.000351

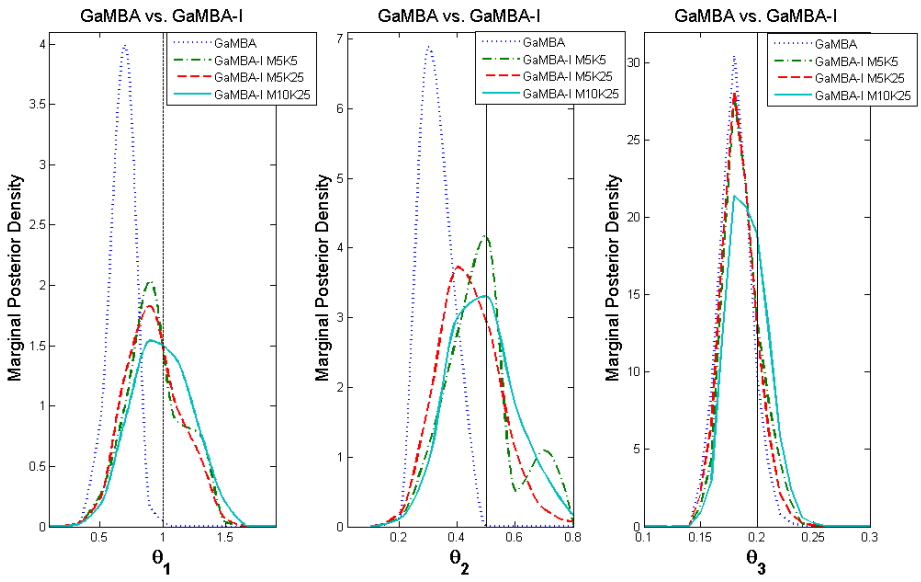

Figure 6: CIR: Marginal posterior using GaMBA for $M = 5$ vs. posteriors obtained using GaMBA-I for $[M = 5, K = 5]$, $[M = 5, K = 25]$ and $[M = 10, K = 25]$ respectively. Vertical lines denote the true values $\theta_1 = 1$, $\theta_2 = 0.5$ and $\theta_3 = 0.2$.

Table 2 lists the mean squared error (MSE) obtained for each of the parameters using different methods for inference. Asymptotic properties of the GaMBA-I posterior can be seen. Posteriors obtained using GaMBA are not accurate for the CIR process. GaMBA-I does provide better approximation and that these approximations become more accurate as M and K increase. Figure 6 plots some of these posteriors.

Example: Euro-Dollar interest rate data

We use this process to model the Euro-Dollar interest rate data. This data set of size 100 has been simulated using parameter values $\theta_1 = 0.00036$, $\theta_2 = 0.0047$, $\theta_3 = 0.012$, $M = 10$, and $Y_0 = 8$. These chosen parameter values are the posterior means obtained by Roberts & Stramer (2001) after analysing a real-life Euro-Dollar interest rate data. Figure 7 shows the simulated data.

Example: Euro-Dollar interest rate data

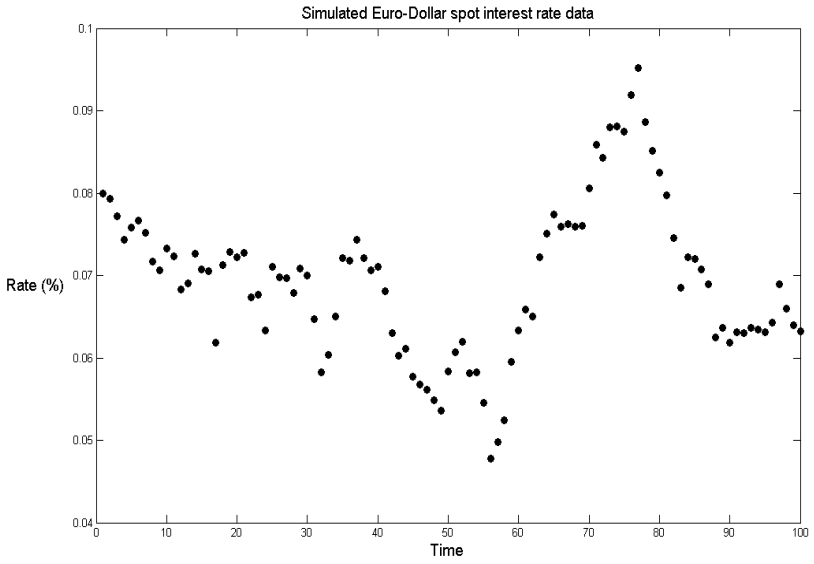

Figure 7: Simulated Euro-Dollar interest rate data.

We analyse this data using a standard MCMC based method proposed by Chib & Shephard (2002) It is a straightforward method which does not incorporate the modifications suggested in Chib et al. (2006) or Golightly & Wilkinson (2007) and this makes it simpler and computationally cheaper to implement than the modified methods. The issue of dependency between the latent variables and the diffusion coefficient can be dealt with by transforming the process as suggested by Roberts & Stramer (2001).

The CIR process can be transformed to a process $Y_t = 2\sqrt{X_t}/\theta_3$. Using Itô's formula the new SDE has a constant diffusion coefficient and is given by

$$dY_t = \left[\frac{(\theta_1 - \theta_2 X_t)\sqrt{X_t}}{\theta_3} - \frac{\theta_3 X_t^{-1/2}}{4}\right] + dW_t. \tag{21}$$

After transforming the SDE an MCMC scheme based on Chib & Shephard (2002) is implemented for a fixed M such that all the $M-1$ latent variables between every pair of consecutive observations are updated in one block, as described below. Between every pair of consecutive observations $\{y_i, y_{i+1}\}$, $M-1$ equally spaced latent variables denoted by $\mathbf{X}^{(i)} = \{X_{1_i}, \cdots, X_{(M-1)_i}\}$ are imputed. The entire set of latent variables for $n+1$

observations $\{y_0, y_1, \cdots, y_n\}$ is denoted by $\mathbf{X} = \{\mathbf{X}^{(i)}\}_{i=1,\cdots,n-1}$.

MCMC algorithm :

Step I: Initialize Θ, and generate initial $\mathbf{X}$'s using Modified Brownian Bridge density $P_{MBB}(\mathbf{X}^{(i)}|y_i, y_{i+1}, \Theta)$

Step II: For $i = 0, 1, ..., (n-1)$,

(a) Propose $\mathbf{X}^{(i)*}$ using Modified Brownian Bridge density $P_{MBB}(\mathbf{X}^{(i)*}|y_i, y_{i+1}, \Theta)$

(b) Accept $\mathbf{X}^{(i)*}$ with probability

$$\alpha = \min\left(\left[\frac{P_{MBB}(\mathbf{X}^{(i)}|y_i, y_{i+1}, \Theta) \cdot P_E(y_{i+1}|\mathbf{X}^{(i)*}, y_i)}{P_{MBB}(\mathbf{X}^{(i)*}|y_i, y_{i+1}, \Theta) \cdot P_E(y_{i+1}|\mathbf{X}^{(i)}, y_i)}\right], 1\right)$$

Step III: Propose Θ^* using a proposal density $Q(\Theta \to \Theta^*)$ and accept Θ^* using

$$\beta = \min\left(\left[\frac{P_E(\mathbf{X}, \mathbf{Y}|\Theta^*) \cdot P(\Theta^*) \cdot Q(\Theta^* \to \Theta)}{P_E(\mathbf{X}, \mathbf{Y}|\Theta) \cdot P(\Theta) \cdot Q(\Theta \to \Theta^*)}\right], 1\right)$$

Step IV: Repeat steps **II** and **III** until the chains satisfy a convergence criteria.

P_E refers to the density corresponding to the Euler approximation. The implementation of this MCMC algorithm is now illustrated by using the following toy example.

Visual check of the MCMC trace plot along with the correlograms were used to assess stationarity. In addition, the Kolmogorov-Smrinov test was used to confirm stationarity. After discarding the first $10,000$ samples as a 'burn-in' period, the next $10,000$ samples were considered to be correlated draws from the stationary distribution. Figure 8 shows the MCMC chains along with their correlograms, and Figure 9 the marginal posteriors. The vertical lines in Figure 9 indicate the true values used to simulate this data. Thus, it can be seen that, after transforming the SDE, this basic MCMC based method yields accurate results. However, this implementation (using MATLAB 7.5.0) takes about 14 minutes on a personal computer.

This data is now analysed using GaMBA-I with $M = 5$ and $K = 10$. The parameter space $\Xi^* = \Xi_1^* \times \Xi_2^* \times \Xi_3^*$ was chosen based on prior knowledge as $\Xi^* = [0.00001, 0.00401] \times$

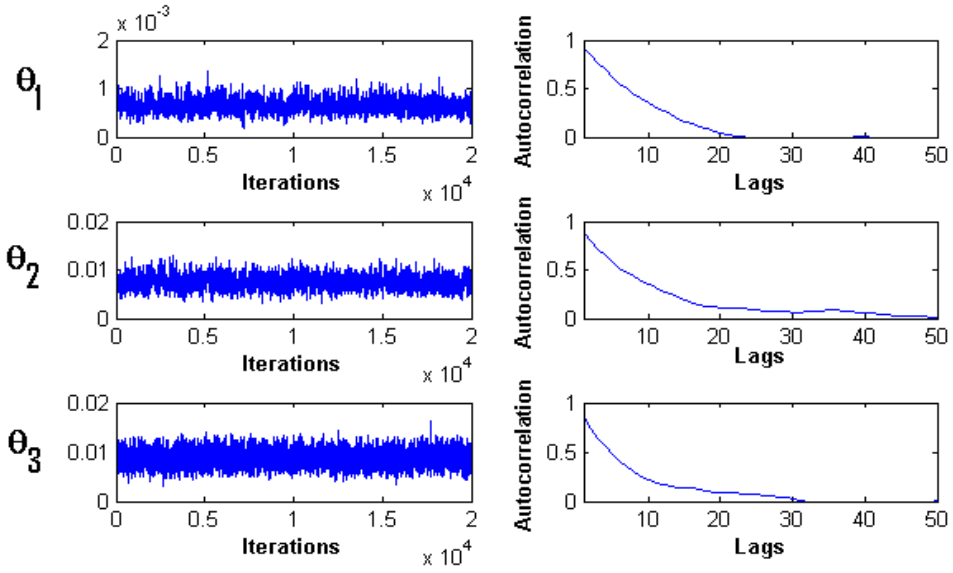

Figure 8: MCMC chains for the Euro-Dollar interest rate data along with their Correlograms.

$[0.00002, 0.02] \times [0.001, 0.0181]$, and a grid $\mathcal{G}_{\Xi^*}$ was considered with $\Delta\Xi_1 = 0.0003$, $\Delta\Xi_2 = 0.003$ and $\Delta\Xi_3 = 0.001$. GaMBA was implemented on the 1764 points thus sampled from Ξ^* and marginal posteriors distributions were obtained.

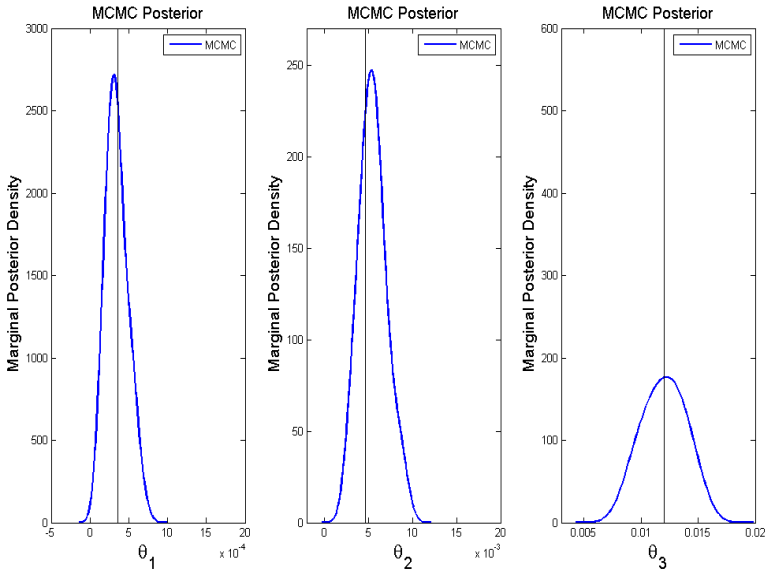

Figure 9: MCMC posteriors for the Euro-Dollar interest rate data - vertical lines indicating the true value of the parameter.

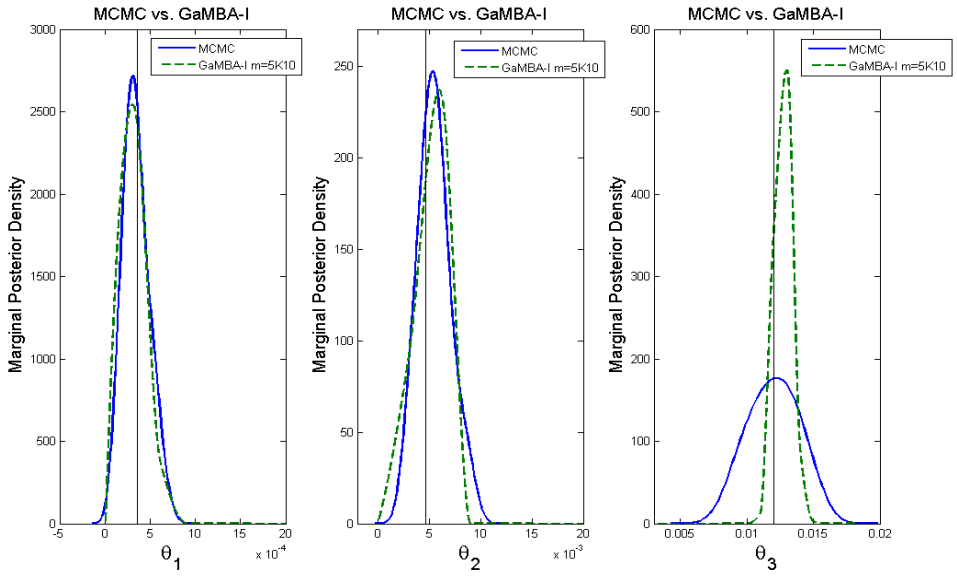

Figure 10: Euro-Dollar data: Marginal posteriors using MCMC vs. posteriors obtained using GaMBA-I for $[M = 5, K = 10]$. Vertical lines denote the true values $\theta_1 = 00036$, $\theta_2 = 0.0047$ and $\theta_3 = 0.012$.

Figure 10 shows the GaMBA-I posterior thus obtained plotted over the MCMC pos-

terior of Figure 9. It can be seen that the marginal posteriors using these two methods for θ_1 and θ_2 are quite close. For θ_3, the posterior obtained using GaMBA-I has a smaller variance than the one obtained using MCMC, but it still covers the true value of the parameters. GaMBA-I takes only 6 minutes to implement in this case, and as described in the next section, this time can be further reduced multi-fold by using multiple parallel processing computing units. Furthermore as seen in the earlier example, if desired, accuracy of GaMBA-I can be improved by increasing M and K. Also note that the MCMC results were obtained after transforming the SDE while no such pre-processing was required for implementing GaMBA-I.

6 Computational Efficiency

The main motivation behind this research is to gain computational efficiency over MCMC based methods. We discuss this aspect in detail below.

- Computational effort required for implementing the MCMC based method described above is directly proportional to M and the number (R) of the samples needed to be drawn using the Markov chains. Computational effort required for implementing GaMBA-I is directly proportional to M, the cardinality of $\mathcal{G}_{\Xi^*}$ and K.
- Computational effort required by GaMBA-I (with $K = 1$) for *one* value of Θ is only slightly smaller than the effort required by MCMC method specified. This slight advantage is due to the fact that unlike MCMC (which samples new Θ from $P(\cdot|\mathbf{X}, \mathbf{Y})$), GAMBA-I takes selects Θ from a pre-determined set of values.
- Let $\mathcal{T}_M$ and $\mathcal{T}_{GI}$ be the computational time required by the MCMC method and GaMBA-I respectively. Then, ignoring the small difference mentioned in the previous point, and assuming that same M is used for both MCMC and GaMBA-I, the computational advantage that GaMBA-I has over MCMC is

$$\mathcal{T}_M - \mathcal{T}_{GI} \propto R - |\mathcal{G}_{\Xi^*}| \times K.$$

- As seen from the Euro-Dollar data example, accurate inference using MCMC requires a large R. Further, a smaller M might be sufficient for inference using GaMBA-I, as seen in that example. Thus the computational advantage mentioned above is expected to be significant for many SDE models.
- In addition to the above computational advantage, the computational effort required for GaMBA-I can be reduced multi-fold using multiple parallel processing computing units. Because multi-core processors have now become a standard in personal computers, this does not require the user to have access to external servers. It is not straightforward to implement MCMC based methods using parallel-processing capabilities.
- Computational effort required for GaMBA is same as the computational effort required for GaMBA-I with $K = 1$.

Parallel processing has not been implemented in any of the examples mentioned in this paper. However, because this technology has a significant potential to improve the computational efficienty of GaMBA-I, presented below is the algorithm of how it could be implemented.

Algorithm for parallel processing of GaMBA-I:

Assume that P parallel processing units are available.

1. Identify Ξ^* (as described in Section 3.1.)
2. Define a discrete grid $\mathcal{G}_{\Xi^*}$ on Ξ^*.
3. Divide the total number of points on $\mathcal{G}_{\Xi^*}$ on Ξ^* into P subsets (mutually exclusive and exhaustive) and assign each subset to a different parallel processing unit.
4. For each of the subsets, say $\mathcal{G}_{\Xi^*}^{(p)}$, for $p = 1, \cdots, P$, implement the following steps $5(a)$ and $5(b)$ simultaneously (using the parallel processing units) for each of the subsets
5. For each point on the subset $\Theta_j \in \mathcal{G}_{\Xi^*}^{(p)}$
   (a) for $k = 1, \cdots, K$,
   Sample $\mathbf{X}_k \sim P_{MBB}(\mathbf{X}_k|\Theta_j, \mathbf{Y})$ as described in Section 3.4

(b) Evaluate

$$P_{GaMBA-I}(\Theta_j|\mathbf{Y}) \propto \frac{1}{K}\sum_{k=1}^{K}\frac{P_E(\mathbf{Y},\mathbf{X}_k|\Theta_j)\cdot P(\Theta_j)}{P_{MBB}(\mathbf{X}_k|\Theta_j,\mathbf{Y})}$$

.

6. Normalise to obtain $P(\Theta|\mathbf{Y})$ over $\mathcal{G}_{\Xi^*}$.

When GaMBA-I is implemented using parallel processing, it is expected that its computational advantage over the MCMC would be

$$\mathcal{T}_M - \mathcal{T}_{GI} \propto R - \frac{|\mathcal{G}_{\Xi^*}| \times K}{P}.$$

7 Discussion

Two algorithms have been introduced in this chapter to obtain computationally efficient Bayesian inference on SDE models. Various aspects of these algorithms are discussed below.

7.1 Which is more appropriate: GaMBA or GaMBA-I?

Both these methods use the MBB density to sample the latent variables $\mathbf{X}$. However while GaMBA is a deterministic approach, GaMBA-I is stochastic. There are also important differences between them regarding the computatinal efficienty, consistency and applicability of these two methods. Some guidelines are provided here to help the practitioner decide which method is more appropriate.

As described earlier, the mean of the MBB density is the linear interpolation between the two observed points. As a consequence GaMBA and GaMBA-I would yield efficient and accurate inference when the data are observed close enough so that the path taken by the process between any two consecutive observations can be considered to be approximately linear. We have noticed that when the data are more than one time unit apart, GaMBA

does not produce accurate inference whereas GaMBA-I does by using a large enough K.

Though GaMBA has a clear computational advantage over GaMBA-I, it is less widely applicable and does not provide the possibility to study the consistency properties of its posterior. Therefore, it would be advisable to use GaMBA only if data are fairly closely observed and when the computational resources are scarce. In all other circumstances GaMBA-I would be more preferable than GaMBA.

7.2 Practical Considerations

Because of its asymptotic properties, and wider applicability GaMBA-I is the preferred algorithm and is therefore the focus of the following discussion. However, unless otherwise stated, the same discussion is applicable to GaMBA as well.

Sparsely observed data: As an importance density MBB is ideally suited to SDE's with linear drift coefficients. For other SDEs (such as, for example CIR), MBB turns out to be efficient only when data are closely observed. For all the examples considered in this paper, data was observed at a unit time interval.

Choosing M : Section 5 has illustrated that reasonably accurate results could be obtained for values of M as small as 5 or 10. However, in many inference problems, the accuracy achieved in Section 5 may not be enough and higher M will have to be used to get better results.

Choosing K : Stramer & Yan (2007) prove that $K = M^2$ is the computationally optimal value for SDEs with unit diffusion coefficient. As can be seen from the examples considered in Section 5, a smaller value of K (between 5 and 20) is often enough to provide efficient inference even for SDEs with non-unit diffusion coefficients (untransformed CIR process). Our for un-transformed CIR process are in line with those obtained for Stramer & Yan (2007).

Choosing $\Delta\Xi$: $\Delta\Xi$ governs how fine (or coarse) the grid constructed on the support Ξ^* is. Because GaMBA-I produces discretised posteriors, a smaller $\Delta\Xi$ implies a finer grid and a better approximation to the true continuous posterior. As illustrated in the examples, $\Delta\Xi$ will be different for every parameter, and its value will depend on the length of the support for that parameter. As a rule-of-thumb, $\Delta\Xi$ should be small enough so that the grid selects at least 10 values from the support for each parameter. This means if there are 3 parameters, $|\mathcal{G}_{\Xi^*}| \geq 1000$.

It is very important to note that this rule-of-thumb strategy advised above is applicable only to symmetric uni-modal posterior distributions. If there are reasons to believe that the posterior distribution may be highly skewed or multi-modal in at least one of the dimensions then much finer $\mathcal{G}_{\Xi^*}$ might be required.

No dependency between X and Θ : GaMBA (and also GaMBA-I) get around the problem of dependency between the latent process $\mathbf{X}$ and the parameters Θ. This means that there is no need to re-parameterise an SDE before inference. This was illustrated by the Euro-dollar example, where the SDE had to be first transformed before using the MCMC method. Thus GaMBA can be applied to a wider class of SDE models; for example: to multivariate SDE's.

7.3 Limitations

Ξ^* is not known : In its present form, GaMBA-I assumes that the support Ξ^* is known due to prior knowledge or otherwise. This is a very important limitation, since this assumption may not be valid for many practical modeling problems. As mentioned in Section 3.1, it may be possible to develop a methodology to objectively identify Ξ^*, however this is a separate research problem in itself at present remains a open problem.

Dimensionality of the parameter space Ξ : While implementing GaMBA-I, the parameter values are sampled from the space Ξ^* by constructing a regular grid on this space. Though this may possibly be the simplest way to sample, it may not be the most efficient one. The grid sampling method works well when the parameter space is five-dimensional or less, but beyond that it very rapidly becomes computationally

too expensive. Thus, GaMBA-I may not be computationally efficient for SDE models with five or more parameters.

Sparsely observed data : For non-linear SDE's, if the data are not observed closely enough, larger M and K would be required to obtain accurate inference and GaMBA-I may no longer be computationally efficient. This clearly limits the potential of GaMBA-I to be applicable in situations where the data is sparsely observed.

8 Summary

A new approach to approximate Bayesian inference for SDE models has been proposed with the aim to gain computational efficiency over MCMC based methods. This approach leads to two methods named as GaMBA and GaMBA-I, with GaMBA-I having attractive consistency properties and more widely applicable. In its present form, GaMBA-I is available for all one-dimensional SDEs. It has been shown to provide efficient and accurate inference when the data are closely observed. This approach has some important limitations, these have been discussed in Section 7.3.

Acknowlegement

The authors would like to thank Gareth Roberts, Alexandros Beskos, Håvard Rue, Rozenn Dahyot, Nial Friel and Giorgos Sermaidis for providing valuable inputs during the development of this work. This work is supported by the STATICA project, funded by the Principal Investigator Programme of Science Foundation Ireland, contract number $08/IN.1/I1879$.

References

Beskos, A., Papaspiliopoulos, O., Roberts, G., & Fernhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *Royal Stat. Soc. B*, *68*, 333–382.

Beskos, A., Roberts, G., Stuart, A., & Voss, J. (2008). MCMC methods for diffusion bridges. *Stochastics and Dynamics*, *8*, 319–350.

Brandt, M. W. & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with application to exchange rate dynamics in an incomplete market. *Journal of Financial Economics.*, *63*, 161–210.

Chib, S., Pitt, M., & Shephard, N. (2006). Likelihood based inference for diffusion driven models. Olin School of Business, Washington University.

Chib, S. & Shephard, N. (2002). Comment on Garland B. Durham and A. Ronald Gallant 'numerical techniques for maximum likelihood estimation of continuous-time diffusion processes'. *J.Busi.& Eco. Statis.*, *20*, 325–327.

Del Moral, P., Jacod, J., & Protter, P. (2002). The monte carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields*, *120*, pp. 346 – 368.

Durham, G. B. & Gallant, R. (2002). Numerical techniques for maximum likelihood estimation of continuous time diffusion processes. *J.Busi.& Eco. Statis.*, *20*, 297–316.

Elerian, O., Chib, S., & Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, *69*, 959–993.

Fernhead, P., Papaspiliopoulos, O., Roberts, G., & Stuart, A. (2010). Random-weight particle filtering of continuous time processes. *Royal Stat. Soc. B*, *72*, 497–512.

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, *57*(6), pp. 1317–1339.

Golightly, A. & Wilkinson, D. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statist. Comput.*, *16*, 323–338.

Golightly, A. & Wilkinson, D. (2007). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, *52*, 1674–1693.

Kloeden, P. & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations* (First ed.). New York, USA: Springer.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete approximations. *Scand. J. Stat.*, *22*, 55–71.

Roberts, G. & Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, *88*, 603–621.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of Royal Statistical Society, Series B*, *71*, 1–35.

Santa-Clara, P. (1995). Simulated likelihood estimation of diffusion with application to the short term interest rate. *Anderson Graduate School of Management, University of California at Los Angeles.*

Sarkka, S. & Sottinen, T. (2008). Application of girsanov theorem to particle filtering of discretely observed continuous time non-linear systems. *Bayesian Analysis*, *3*, pp. 555–584.

Stramer, O. & Yan, J. (2007). Asymptotics of an efficient monte carlo estimation for the transition density of diffusion processes. *9*, 483–496.