

Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes

Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O. Roberts and Paul Fearnhead*

July 6, 2006

Abstract

The objective of this paper is to present a novel methodology for likelihood-based inference for discretely observed diffusions. We propose Monte Carlo methods, which build on recent advances on the exact simulation of diffusions, for performing maximum likelihood and Bayesian estimation.

Keywords : Retrospective sampling, EM algorithm, Monte Carlo maximum likelihood, Markov chain Monte Carlo, CIR model, graphical models.

1 Introduction

Diffusion processes are extensively used for modelling continuous-time phenomena in many scientific areas; an incomplete list with some indicative references includes economics (Black and Scholes, 1973; Chan et al., 1992; Cox et al., 1985; Merton, 1971), biology (McAdams and Arkin, 1997), genetics (Kimura and Ohta, 1971; Shiga, 1985), chemistry (Gillespie, 1976, 1977), physics (Obuhov, 1959) and engineering (Pardoux and Pignol, 1984). Their appeal lies in the fact that the model is built by specifying the instantaneous mean and variance of the process through a stochastic differential equation (SDE). Specifically, a diffusion process V is defined as the solution of an SDE of the type:

$$dV_s = b(V_s; \theta) ds + \sigma(V_s; \theta) dB_s, \quad (1)$$

driven by the scalar Brownian motion B . The functionals $b(\cdot; \theta)$ and $\sigma(\cdot; \theta)$ are called the *drift* and the *diffusion coefficient* respectively and are allowed to depend on some parameters $\theta \in \Theta$. They are presumed to satisfy the regularity conditions (locally Lipschitz, with a linear growth bound) that guarantee the existence of a weakly unique, global solution of (1), see Chapter 4 of Kloeden and Platen (1995). In this paper we will consider only one-dimensional diffusions, although multivariate extensions are possible.

*Department of Mathematics and Statistics, Lancaster University, U.K., email: a.beskos@lancaster.ac.uk, o.papaspiliopoulos@lancaster.ac.uk, g.o.roberts@lancaster.ac.uk

For sufficiently small time increment dt and under certain regularity conditions (see Kloeden and Platen, 1995), $V_{t+dt} - V_t$ is *approximately* Gaussian with mean and variance given by the so-called *Euler* (or Euler-Maruyama) approximation:

$$V_{t+dt} \approx V_t + b(V_t; \theta)dt + \sigma(V_t; \theta)\sqrt{dt} \cdot Y, \text{ where } Y \sim N(0, 1),$$

though higher order approximations are also available. The *exact* dynamics of the diffusion process are governed by its transition density:

$$p_t(v, w; \theta) = P[V_t \in dw \mid V_0 = v; \theta] / dw, \quad t > 0, w, v \in \mathbf{R}. \quad (2)$$

We shall assume that the process is observed without error at a given collection of time instances,

$$\mathbf{v} = \{V_{t_0}, V_{t_1}, \dots, V_{t_n}\}, \quad 0 = t_0 < t_1 < \dots < t_n;$$

this justifies the notion of a *discretely observed diffusion process*. The time increments between consecutive observations will be denoted $\Delta t_i = t_i - t_{i-1}$ for $1 \leq i \leq n$.

The log-likelihood of the data set \mathbf{v} is:

$$\ell(\theta \mid \mathbf{v}) = \sum_{i=1}^n \ell_i(\theta), \quad \ell_i(\theta) := \log p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta).$$

Unfortunately, in all but a few special cases the transition density of the diffusion process and thus its likelihood are not analytically available. Therefore, it is already well documented that deriving maximum likelihood estimates (MLEs) for discretely observed diffusion processes is a very challenging problem. Nonetheless, theoretical properties of such MLEs are now well known in particular under ergodicity assumptions, see for example Kessler (1997); Gobet (2002).

Inference for discretely observed diffusions has been pursued in three main directions. One direction considers estimators alternative to the MLE. Established methods within this paradigm include techniques based on estimating functions (Bibby et al., 2002), indirect inference (Gourieroux et al., 1993) and efficient method of moments (Gallant and Long, 1997). Another direction involves numerical approximations to the unknown likelihood function. Ait-Sahalia (2002) advocated the use of closed-form analytic approximations to the unknown transition density; see Ait-Sahalia (2004) for multidimensional extensions. An alternative strategy has been to estimate an approximation to the likelihood using Monte Carlo (MC) methods. The approximation is given by Euler-type discretisation schemes, and the estimate is obtained using importance sampling. The strategy was put forward by Pedersen (1995); Santa-Clara (1995) and was considerably refined by Durham and Gallant (2002). The third direction employs Bayesian imputation methods. The idea is to augment the observed data with values at additional time points so that a satisfactory complete-data likelihood approximation can be written down and to use the Gibbs sampler or alternative Markov chain Monte Carlo (MCMC) schemes; see Roberts and Stramer (2001); Elerian et al. (2001); Eraker (2001). An excellent review of several methods of inference for discretely observed diffusions is given in Sørensen (2004).

The approach introduced in this paper follows a different direction, which exploits recent advances in simulation methodology for diffusions. Exact simulation of diffusion sample paths has become feasible since the introduction of the Exact Algorithm (EA) in Beskos et al. (2004b). The algorithm is reviewed in Section 2 and relies on a technique called *retrospective sampling* developed originally in Papaspiliopoulos and Roberts (2004). To date there are two versions of the algorithm, EA1 which can be applied to a rather limited class of diffusion processes, which we call \mathcal{D}_1 , and EA2 which is applicable to the much more general \mathcal{D}_2 class; all definitions are given in Section 2. The greater applicability of EA2 over EA1 comes at the cost of higher mathematical sophistication in its derivation, since certain results and techniques from stochastic analysis are required. However, its computer implementation is similar to that of EA1.

In this paper we show how to use EA in order to produce a variety of methods that can be used for maximum likelihood and Bayesian inference. We first discuss three unbiased MC estimators of the transition density (2) for a fixed value of θ : the bridge method (Section 4, first proposed in Beskos et al., 2004b), the acceptance method (Section 5) and the Poisson estimator (Section 6, first proposed in Wagner, 1988b). The last two estimators are evolved in Sections 5.1 and 6 respectively to yield unbiased estimators of the transition density *simultaneously* for all $\theta \in \Theta$. Thus, the simultaneous estimators can readily be used in conjunction with numerical optimisation routines to estimate the MLE and other features of the likelihood surface.

We proceed by introducing a Monte Carlo EM (MCEM) algorithm in Section 8. The construction of the algorithm crucially depends on whether there are unknown parameters in the diffusion coefficient σ . The simpler case where only drift parameters are to be estimated is treated in Section 8.1, whereas the general case necessitates the path transformations of Roberts and Stramer (2001) and it is handled in Section 8.2.

Section 9 presents a Markov chain Monte Carlo algorithm which samples from the joint posterior distribution of the parameters and of appropriately chosen latent variables. Unlike currently favoured methods, our algorithm is not based on imputation of diffusion paths but instead on what we call a *hierarchical simulation model*. In that way, our MCMC method circumvents computing the likelihood function.

Therefore, all our methodology is simulation-based, but it has advantages over existing methods of this type for two reasons.

1. The methods are *exact* in the sense that no discretisation error exists, and the MC estimation provides the only source of error in our calculations. Specifically, as the number of MC samples increases, the estimated MLE converge to the true MLE, and as the number of iterations in our MCMC algorithm increases, the samples converge to the true posterior distribution of the parameters.
2. Our methods are computationally efficient. Whilst approximate methods require rather fine discretisations (and consequently a number of imputed values which greatly exceeds the observed data size) to guarantee sufficient accuracy, our methodology suffers from no such restrictions.

A limitation of the methods introduced here is that their applicability is generally attached

to that of EA. However, ongoing advances on EA itself (Beskos et al., 2005b) will weaken further the required regularity conditions so that a much larger class of diffusions than \mathcal{D}_2 can be effectively simulated. It is expected that these enhanced simulation algorithms will be of immediate use to the methods presented in this paper.

Our methods are illustrated on three different diffusion models. The first one is the periodic drift model, which belongs to \mathcal{D}_1 and, although it is interesting enough in its own right since its transition density is unavailable, it is used primarily for exposition. However, we also consider two more substantial and well-known applications, the logistic diffusion model for population growth, and the Cox-Ingersoll-Ross (CIR) model for interest rates. The former belongs to the \mathcal{D}_2 class, whereas the latter is a diffusion process outside the \mathcal{D}_2 class, and it is used to illustrate how our exact methods can be extended for processes for which EA2 is not applicable. Moreover, since we can calculate analytically the likelihood for this model, we have a benchmark to test the success of our approach. We fit the CIR model to a well studied data set, which contains Eurodollar rates (recorded every ten days) between 1973 and 1995, to allow for comparisons with existing methods.

All the algorithms presented in this paper are coded in **C** and have been executed on a Pentium IV 2.6 GHz processor. We note that our methods are not computationally demanding according to modern statistical computing standards, and in the examples we have considered the computing times (reported explicitly in the following sections) were in the magnitude of seconds, or at worst minutes.

The structure of the paper is as follows. Section 2 reviews EA. Section 3 sets up the context of transition density estimation, Sections 4-6 present the three different estimators and Section 7 compares them theoretically and empirically. Section 8 introduces the MCEM algorithm and Section 9 the MCMC algorithm. We finish with some general conclusions and directions for further research in Section 10. Background material and proofs are collected in a brief Appendix.

2 Retrospective Exact Sampling of Diffusions

Understanding the statistical methodology to be presented in later sections presupposes the introduction of the simulation techniques central to our approaches. In particular, we need to understand the form of the output provided by EA. The main references for the material of this section are Beskos et al. (2004b) and Beskos and Roberts (2005). For the purposes of this paper it suffices to analyse EA for simulating diffusion paths conditional on their ending point (also known as diffusion bridges). In particular we will show how to simulate a diffusion path starting from $V_0 = v$ and ending at $V_t = w$, for any $t > 0, v, w \in \mathbf{R}$. Simulation of unconditioned paths follows in the same lines and is sketched in Section 2.3. When necessary we characterise EA as *conditional* or *unconditional* to emphasise the type of diffusion simulation it is used for.

EA performs rejection sampling by proposing paths from processes we can simulate and accepting them according to appropriate probability density ratios. The novelty lies in the fact that the proposed paths are unveiled only at finite (but random) time instances and the decision whether to accept or not the path can be easily taken.

It is essential that we first transform the diffusion process (1) into an SDE of unit diffusion coefficient by applying the 1–1 transformation $V_s \rightarrow \eta(V_s; \theta) =: X_s$, where

$$\eta(u; \theta) = \int^u \frac{1}{\sigma(z; \theta)} dz, \quad (3)$$

is any anti-derivative of $\sigma^{-1}(\cdot; \theta)$. Assuming that $\sigma(\cdot; \theta)$ is continuously differentiable, we apply Itô's rule to find that the SDE of the transformed process writes as:

$$dX_s = \alpha(X_s; \theta) ds + dB_s, \quad X_0 = x = \eta(V_0; \theta), \quad s \in [0, t], \quad (4)$$

where

$$\alpha(u; \theta) = \frac{b\{\eta^{-1}(u; \theta); \theta\}}{\sigma\{\eta^{-1}(u; \theta); \theta\}} - \sigma'\{\eta^{-1}(u; \theta); \theta\} / 2, \quad u \in \mathbf{R};$$

η^{-1} denotes the inverse transformation and σ' denotes the derivative w.r.t. the space variable. In the sequel we will make the following standard assumptions for any $\theta \in \Theta$:

1. $\alpha(\cdot; \theta)$ is continuously differentiable.
2. $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded below.
3. Girsanov's formula for X given in (32) is a martingale w.r.t. Wiener measure.

We define

$$A(u; \theta) := \int^u \alpha(z; \theta) dz,$$

to be any anti-derivative of α . The transition density of X is defined as:

$$\tilde{p}_t(x, y; \theta) = \mathbb{P}[X_t \in dy \mid X_0 = x; \theta] / dy, \quad t > 0, \quad x, y \in \mathbf{R}. \quad (5)$$

Before proceeding, we require the following preliminary notation. Let $C \equiv C([0, t], \mathbf{R})$ be the set of continuous mappings from $[0, t]$ to \mathbf{R} , \mathcal{C} the corresponding cylinder σ -algebra and $\omega = (\omega_s, s \in [0, t])$ a typical element of C . Let $\mathbb{Q}_\theta^{(t,x,y)}$ denote the distribution of the process X conditioned to start at $X_0 = x$ and finish at $X_t = y$, for some fixed x, y , and $\mathbb{W}^{(t,x,y)}$ be the probability measure for the corresponding Brownian bridge (BB). The notation highlights the dependence of the measure $\mathbb{Q}_\theta^{(t,x,y)}$ on θ .

The objective is to construct a rejection sampling algorithm to draw from $\mathbb{Q}_\theta^{(t,x,y)}$. The following lemma proved in Appendix is central to the methodology. By $\mathcal{N}_t(u)$ we denote the density of the normal distribution with mean 0 and variance t evaluated at $u \in \mathbf{R}$.

Lemma 1. *Under conditions 1–3 above, $\mathbb{Q}_\theta^{(t,x,y)}$ is absolutely continuous with respect to $\mathbb{W}^{(t,x,y)}$ with density:*

$$\frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(\omega) = \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)} \exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\}.$$

Let

$$l(\theta) \leq \inf_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta)/2\}$$

$$r(\omega, \theta) \geq \sup_{s \in [0, t]} \{(\alpha^2 + \alpha')(\omega_s; \theta)/2 - l(\theta)\}, \quad \omega \in C.$$

In general r is a positive random variable. However in special cases it can be chosen independently of ω , for example if $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above. We now define the non-negative function $0 \leq \phi \leq 1$ which transforms a given path ω as follows:

$$\phi(\omega_s; \theta) = \frac{1}{r(\omega, \theta)} \left\{ \frac{(\alpha^2 + \alpha')(\omega_s; \theta)}{2} - l(\theta) \right\}, \quad s \in [0, t]. \quad (6)$$

It is now clear that:

$$\frac{dQ_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(\omega) \propto \exp \left\{ -r(\omega, \theta) \int_0^t \phi(\omega_s; \theta) ds \right\} \leq 1, \quad \mathbb{W}^{(t,x,y)} - \text{a.s.} \quad (7)$$

Thus we have managed in (7) to bound the density ratio. The key to proceed to a feasible rejection sampler is to recognise (7) as a specific Poisson process probability.

Theorem 1. *Let $\omega \in C$, Φ be a homogeneous Poisson process of intensity $r(\omega, \theta)$ on $[0, t] \times [0, 1]$ and N be the number of points of Φ below the graph $s \mapsto \phi(\omega_s; \theta)$. Then*

$$\mathbb{P}[N = 0 | \omega] = \exp \left\{ -r(\omega, \theta) \int_0^t \phi(\omega_s; \theta) ds \right\}.$$

Theorem 1 suggests rejection sampling by means of an auxiliary Poisson process:

1. Simulate a sample path $\omega \sim \mathbb{W}^{(t,x,y)}$.
2. Calculate $r(\omega, \theta)$; generate a marked Poisson process $\Phi = \{\Psi, \Upsilon\}$, with points $\Psi = \{\psi_1, \dots, \psi_\kappa\}$ uniformly distributed on $[0, t]$ and marks $\Upsilon = \{v_1, \dots, v_\kappa\}$ uniformly distributed on $[0, 1]$, where $\kappa \sim \text{Po}(r(\omega, \theta)t)$.
3. Compute the acceptance indicator:

$$I := \prod_{j=1}^{\kappa} \mathbb{I}[\phi(\omega_{\psi_j}; \theta) < v_j]. \quad (8)$$

4. If $I = 1$, i.e. $\{N = 0\}$ has occurred, then accept ω , otherwise return to 1.

Unfortunately, this ‘‘algorithm’’ is impossible to implement since it requires the simulation of complete BBs on $[0, t]$. However, notice that it might be possible to determine I based only on partial information about the proposed path. For instance, when r is only a function of θ (see Section 2.1) we can actually reverse the order in which steps 1 and 2 are carried out. Specifically, we would first simulate Φ , and afterwards, *retrospectively*, we would realise ω at

the time instances determined by Ψ , since this is sufficient for determining I . The technique of exchanging the order of simulation to implement in finite time simulation of infinite dimensional random variables has been termed *retrospective sampling* in Papaspiliopoulos and Roberts (2004).

The general framework under which EA operates assumes that it is possible to write:

$$r(\omega, \theta) = r(k(\omega), \theta)$$

where $k(\omega)$ has the following properties:

1. $k(\omega)$ is finite dimensional.
2. The law of $k(\omega)$ can be simulated (under $\mathbb{W}^{(t,x,y)}$).
3. The finite-dimensional distributions of $\mathbb{W}^{(t,x,y)}$ given $k(\omega)$ can be simulated.

Under these conditions, the following retrospective implementation of the rejection sampler can be carried out in finite time.

Exact Algorithm (EA)

1. Simulate $k(\omega)$.
 2. Generate a realisation of the marked Poisson process $\Phi = \{\Psi, \Upsilon\}$ of rate $r(k(\omega), \theta)$.
 3. Simulate the skeleton $\{\omega_{\psi_1}, \dots, \omega_{\psi_\kappa}\}$, conditionally upon $k(\omega)$.
 4. Compute the acceptance indicator I .
 5. If $I = 1$, then accept the proposed skeleton, and return $k(\omega)$ and $S(\omega) := \{(0, x), (\psi_1, \omega_{\psi_1}), \dots, (\psi_\kappa, \omega_{\psi_\kappa}), (t, y)\}$; otherwise return to 1.
-

$S(\omega)$ is an *exact* draw from a finite-dimensional distribution of $\mathbb{Q}_\theta^{(t,x,y)}$, which can be filled in at any required times afterwards using simply BB interpolation (see Section 2.3). The technical difficulty of finding or simulating $k(\omega)$, and consequently simulating the process at some time points given $k(\omega)$, imposes some restrictions on the applicability of EA. We now describe the two cases where the algorithm can be easily applied.

2.1 Exact Algorithm 1 (EA1)

Implementation of EA is straightforward when r does not depend on ω . This will be true within the following diffusion class.

Definition 1. We say that a diffusion process V with SDE (1) belongs to \mathcal{D}_1 , and write $V \in \mathcal{D}_1$, if the drift of the transformed process $X_s = \eta(V_s; \theta)$, $s \in [0, t]$, satisfies conditions 1–3 on pg.5, and $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above.

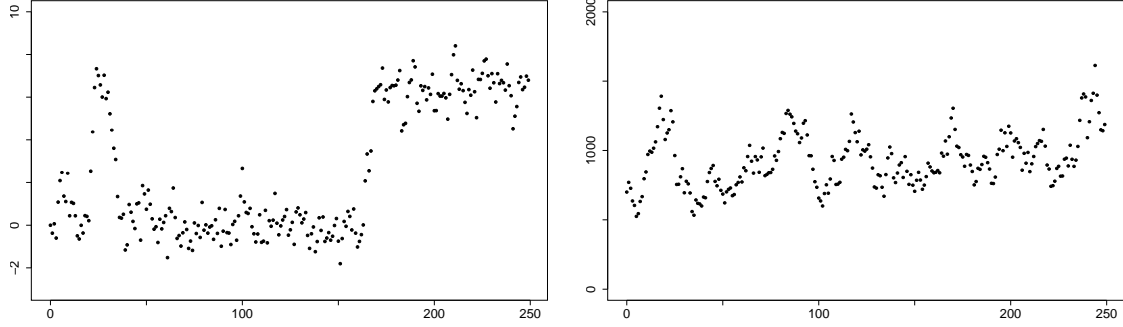


Figure 1: First 250 values of the SINE (left - using EA1) and the LOG – GROWTH (right - using EA2) data sets defined in Example 1, pg. 8, and Example 2, pg. 9, respectively.

In this case

$$r(\omega, \theta) \equiv r(\theta) = \sup_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\}.$$

Within \mathcal{D}_1 , Step 1 of EA is unnecessary, since no information about the path is *a priori* needed for determining the Poisson rate. Moreover, Step 3 entails simulation from a finite-dimensional distribution of BB (see Appendix A).

Example 1 (Periodic drift). Consider the following SDE:

$$dX_s = \sin(X_s - \theta)ds + dB_s.$$

Though apparently simple, the SDE cannot be solved analytically. However, EA1 can be applied since $X \in \mathcal{D}_1$ with $l(\theta) = -1/2$ and $r(\theta) = 9/8$. Our proposed methods will be tested on the SINE data set simulated from X with the unconditional version of EA1 (Beskos and Roberts, 2005) under the specifications $n = 1000$, $\Delta t_i = 1$, $X_0 = 0$, $\theta = \pi$; see also Figure 1. When θ is to be estimated, we take $\Theta = [0, 2\pi]$ for identifiability reasons.

2.2 Exact Algorithm 2 (EA2)

EA2 applies to the wider class of diffusion processes defined below.

Definition 2. We say that a diffusion process V with SDE (1) belongs to the class \mathcal{D}_2 , and write $V \in \mathcal{D}_2$, if the drift of the transformed process $X_s = \eta(V_s; \theta)$, $s \in [0, t]$, satisfies conditions 1–3 on pg.5, and

$$\limsup_{u \rightarrow \infty} (\alpha^2 + \alpha')(u; \theta) < \infty \quad (9)$$

or $\limsup_{u \rightarrow -\infty} (\alpha^2 + \alpha')(u; \theta) < \infty$.

Due to symmetry, we will study only the case (9). We define the following elements of the path space $C = C([0, t], \mathbf{R})$:

$$m = \inf\{\omega_s; s \in [0, t]\}, \quad \tau = \sup\{s \in [0, t] : \omega_s = m\},$$

thus m is the minimum value of a path ω , and τ is the time that the minimum is attained. Within \mathcal{D}_2 , $k(\omega) = m(\omega)$ and,

$$r(\omega, \theta) \equiv r(m, \theta) = \sup_{u \in [m, \infty]} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\} < \infty, \quad \omega \in C.$$

As shown in Beskos et al. (2004b) and summarised in Appendix A, m satisfies the three requirements in pg.7. Simulation of (τ, m) can be done using simple transformations of elementary random elements. It is known (Asmussen et al., 1995) that BB conditionally on (τ, m) , can be derived in terms of two independent *Bessel Bridges*, each operating on either side of (τ, m) . The Bessel bridge is defined as BB constrained to be positive and its simulation can be carried out by means of independent BBs; see Appendix A for details.

Example 2 (The logistic-growth model). A popular model for describing the dynamics of a population which grows at a geometric rate in an environment with limited feeding resources is given by the following SDE:

$$dV_s = R V_s (1 - V_s/\Lambda) ds + \sigma V_s dB_s, \quad \theta = (R, \Lambda, \sigma).$$

R is the growth rate per individual, Λ the maximal population supported by the resources of the environment and σ a noise parameter; for more details see ch.6 of Goel and Richter-Dyn (1974). Related models have been investigated in the context of financial economics, see e.g. Gouriéroux and Jasiak (2003). The transition density of V is not known analytically. The modified process $X_s = -\log(V_s)/\sigma$ solves the SDE:

$$dX_s = \left\{ \frac{\sigma}{2} - \frac{R}{\sigma} + \frac{R}{\sigma\Lambda} \exp(-\sigma X_s) \right\} ds + dB_s.$$

It can be verified that $V \in \mathcal{D}_2$ and that EA2 is applicable with $l(\theta) = \sigma^2/8 - R/2$ and $r(m, \theta) = \max\{(\alpha^2 + \alpha')(m; \theta)/2 - l(\theta), R^2/(2\sigma^2)\}$. Our proposed methods will be tested on the LOG – GROWTH data set simulated from V by first simulating from X using the unconditional EA2 and then transforming $X_s \rightarrow V_s$ (Beskos et al., 2004b). We took $n = 1000$, $\Delta t_i = 1$, $V_0 = 700$ and $(R, \Lambda, \sigma) = (0.1, 1000, 0.1)$; see Figure 1.

2.3 Unconditional EA and path reconstruction

EA can be applied in a similar fashion when the diffusion is conditioned only on its initial point $X_0 = x$. The main difference lies in that the final value of the proposed path is distributed according to the density $h(u) \propto \exp\{A(u; \theta) - (u - x)^2/(2t)\}$ (which is assumed to be integrable). Simulation from h can be done efficiently using an adaptive rejection sampling approach. Conditionally on this final point, the rest of the path is a BB, and EA proceeds as already described.

The output of EA is a skeleton $S(\omega)$ and possibly the collection $k(\omega)$ of variables related to the path. However, we can afterwards fill in this finite representation of an accepted path ω according to the dynamics of the proposed BB and without further reference to the target process. For EA1, the path reconstruction requires the simulation of the BBs that connect the currently unveiled instances of an accepted path ω . For EA2, the extra conditioning of the proposed BB on its minimum m implies that the filling in will be driven by two independent Bessel bridges; see Appendix A.

3 Unbiased transition density estimators – preliminaries

As already discussed in the introduction one of our primary aims is to provide unbiased MC estimators of the transition density (2). The first step is to express (2) in terms of the density (5) of the transformed process X . A change of variables argument yields

$$p_t(v, w; \theta) = \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta) \times |\eta'(w; \theta)|, \quad \text{for all } t, v, w, \theta.$$

Thus, we will demonstrate how to estimate $\tilde{p}_t(x, y; \theta)$ for arbitrary x, y, θ , using the *unconditional* EA (bridge method) and the *conditional* EA (acceptance method, Poisson estimator). Automatically, our methods yield unbiased estimators of the likelihood $\exp\{\ell(\theta | \mathbf{v})\}$, for any fixed $\theta \in \Theta$, so coupled with grid or more elaborate stochastic search algorithms they can be used for locating the MLE. However, this *pointwise* exploration of the likelihood surface is expected to be computationally inefficient due to the introduction of independent MC error at each likelihood evaluation, and it is not guaranteed to provide consistent MC estimators of the MLE. We overcome these drawbacks by modifying the two methods based on the conditional EA to give estimates of the likelihood function *simultaneously* for all $\theta \in \Theta$. The simultaneous methods achieve estimation of the complete function

$$\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta), \quad \text{for any fixed data points } v, w, \quad (10)$$

using a single stream of random elements which are independent of θ . Therefore, numerically efficient optimisation algorithms can be used to estimate features of the likelihood surface, such as the MLE and level sets. The simultaneous acceptance method possesses appealing consistency properties, as we discuss in Section 5.1.

A fundamental result which the acceptance method and the Poisson estimator are based on is the following corollary to Lemma 1. This important result is also contained in Lemma 1 of Dacunha-Castelle and Florens-Zmirou (1986).

Corollary 1. *Let X be the diffusion process in (4) with transition density $\tilde{p}_t(x, y; \theta)$. Then:*

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_t(y-x) \mathbb{E}_{\mathbb{W}(t,x,y)} \left[\exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\} \right]. \quad (11)$$

4 The bridge method

The bridge method introduced in Beskos et al. (2004b) uses explicitly the output of the *unconditional* EA to produce unbiased estimates of $\tilde{p}_t(x, y; \theta)$ for any fixed x, y, θ . We enrich the notation for the transition density to allow for conditioning upon further random elements: let $\tilde{p}_t(x, y | \cdot; \theta) = \mathbb{P}[X_t \in dy | X_0 = x, \cdot; \theta] / dy$, be the conditional density of X_t given the starting value and any other random elements.

Consider some $\delta > 0$. Suppose that $S(\omega), k(\omega)$ are the output of the unconditional EA for simulating X started at $X_0 = x$ on the extended time interval $[0, t + \delta]$. By conditional expectation properties, we have that:

$$\tilde{p}_t(x, y; \theta) = \mathbb{E}_{S(\omega), k(\omega)} [\tilde{p}_t(x, y | S(\omega), k(\omega); \theta)]. \quad (12)$$

The joint distribution of $S(\omega)$, $k(\omega)$ is intractable but can be easily simulated using EA on $[0, t + \delta]$. Recall that the value of the path at times other than those specified by $S(\omega)$ can be obtained according to BB dynamics (in EA1) or Bessel bridge dynamics (in EA2). As a result, the conditional distribution of X_t given $S(\omega), k(\omega)$, thus density on the right side of (12) can be easily identified. By construction $S(\omega)$ contains at least one time point on either side of t , since $(0, x), (t + \delta, \omega_{t+\delta}) \in S(\omega)$, and let t_- and t_+ denote the adjacent time points, $t_- < t < t_+$. By the Markov property $\tilde{p}_t(x, y | S(\omega), k(\omega); \theta)$ depends only on the points $(t_-, \omega_{t_-}), (t_+, \omega_{t_+})$ of $S(\omega)$ and it will be equal either to a Brownian or a Bessel bridge density, both of which are analytically available and can be computed explicitly.

Thus, the main attraction of the method is that the density of X_t given the simulated variables (the skeleton) is explicitly known, unlike for example the MC method by Pedersen (1995), where it is approximated.

5 The acceptance method (AM)

In this section we show how to use a simple identity related to the *conditional* EA to derive an unbiased estimator of the transition density $\tilde{p}_t(x, y; \theta)$. Let $a(x, y, \theta)$ be the acceptance probability of EA for simulating from $\mathbb{Q}_\theta^{(t,x,y)}$. Directly, (7) implies that

$$a(x, y, \theta) = \mathbb{E}_{\mathbb{W}^{(t,x,y)}} \left[\exp \left\{ -r(\omega, \theta) \int_0^t \phi(\omega_s; \theta) ds \right\} \right]. \quad (13)$$

Substituting this expression in (11) and rearranging terms according to (6), yields the following identity which relates the acceptance probability of EA to the transition density of the diffusion:

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_t(y - x) \exp\{A(y; \theta) - A(x; \theta) - l(\theta)t\} a(x, y, \theta). \quad (14)$$

Recall the definition of the acceptance indicator I in (8), which we will now rewrite as $I(x, y, \theta, \Phi, \omega)$ to emphasise the elements which determine its value: the Poisson process Φ , and the proposed bridge $\omega \sim \mathbb{W}^{(t,x,y)}$, unveiled at the times determined by Φ . By definition,

$$a(x, y, \theta) = \mathbb{E} [I(x, y, \theta, \Phi, \omega)], \quad (15)$$

where the expectation is taken with respect to the joint distribution of Φ and ω . Thus, a simple unbiased estimator of $a(x, y, \theta)$ for any fixed θ, x, y , can be obtained by recording the number of times EA accepts a proposed skeleton in, K say, number of trials.

Estimation of the transition density through (14) and (15) is referred to as the acceptance method (AM). An essential feature of the resulting estimator is that it is a.s. bounded, since $I \leq 1$. As a result, all the moments of the estimator are finite.

5.1 The simultaneous acceptance method (SAM)

In this section we upgrade AM to yield an estimator of the complete map in (10). The method is now referred to as the simultaneous AM (SAM). We emphasise that (13)-(15)

are applied for $x = x(\theta) = \eta(v; \theta)$ and $y = y(\theta) = \eta(w; \theta)$. The simultaneous estimator is obtained by expressing $a(x, y, \theta)$ as an expectation of a function of θ, x, y , and random variables none of which depends on θ, x and y . The method is presented for $V \in \mathcal{D}_1$, since the derivation of SAM for $V \in \mathcal{D}_2$ is technically much harder and it is contained in Beskos et al. (2005a), along with more theoretical results about the methods.

In AM, the Poisson process Φ and the proposed path ω both depend on θ , since the former is of rate $r(\theta)$ and the latter is a BB from $X_0 = x$ to $X_t = y$. As a result $a(x, y, \theta)$ can be estimated only by running a separate MC experiment for each $\theta \in \Theta$. Below we show how the thinning property of the Poisson process (see for example Section 5.1 of Kingman, 1993) and the relocation-invariance property of BB (see Appendix A) can be exploited to decouple the dependence between Φ, ω and θ .

The relocation-invariance property of the BB suggests that we can rewrite the acceptance indicator in terms of a standard BB, $\omega \sim \mathbb{W}^{(t,0,0)}$, as follows:

$$I(x, y, \theta, \Phi, \omega) = \prod_{j=1}^{\kappa} \mathbb{I} \left[\phi \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) < v_j \right]. \quad (16)$$

Suppose we can find an $r_{\max} < \infty$, such that

$$r(\theta) \leq r_{\max}, \quad \text{for all } \theta \in \Theta.$$

Let $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be a marked Poisson process on $[0, t] \times [0, 1]$ with rate r_{\max} and κ number of points, $\kappa \sim \text{Po}(r_{\max}t)$. The thinning property for Poisson process implies that the process obtained by deleting each point of Φ_{\max} with probability $1 - r(\theta)/r_{\max}$ is a Poisson process with rate $r(\theta)$. Conditionally on κ , let $U = (U_1, \dots, U_{\kappa})$ be a collection of iid $\text{Un}[0, 1]$ selection random variables. The thinning property implies that we can rewrite (16) in terms of Φ_{\max} as:

$$I(x, y, \theta, \Phi_{\max}, \omega, U) = \prod_{j=1}^{\kappa} \mathbb{I} \left[\mathbb{I} [U_j > r(\theta)/r_{\max}] \phi \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) < v_j \right];$$

notice that only Poisson points for which $U_j < r(\theta)/r_{\max}$, determine the value of I .

Theorem 2. *Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be an independent marked Poisson process of rate r_{\max} with κ number of points. Then*

$$\begin{aligned} \mathbb{E} [I(x, y, \theta, \Phi_{\max}, \omega, U) \mid \Psi_{\max}, \omega] &= \prod_{j=1}^{\kappa} \left[1 - \frac{r(\theta)}{r_{\max}} \phi \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) \right] \\ &= (r_{\max})^{-\kappa} \prod_{j=1}^{\kappa} \left[r_{\max} + l(\theta) - \frac{1}{2}(\alpha^2 + \alpha') \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) \right]. \end{aligned} \quad (17)$$

Notice that $a(x, y, \theta)$ is the expected value of (17) with respect to the joint distribution of Ψ_{\max} and $\omega \sim \mathbb{W}^{(t,0,0)}$, thus (17) together with (14) determine an unbiased *simultaneous* estimator of $\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta)$, for any fixed data points v, w . An MC estimator of $a(x, y, \theta)$ is obtained by averaging over independent realisations of (17).

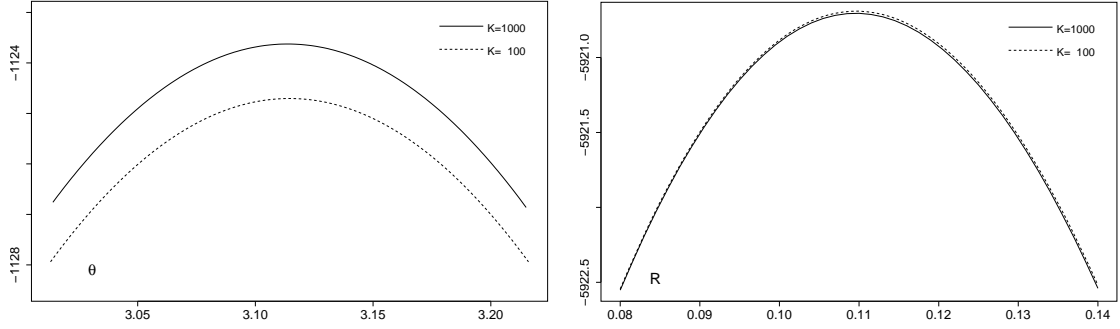


Figure 2: Simultaneous estimate of the log-likelihood for the **SINE** data set (left) and the profile log-likelihood R for the **LOG – GROWTH** data set (right), using SAM with $K = 100$ (dashed) and $K = 1000$ (solid) MC samples.

Example: Periodic drift. We applied SAM to the **SINE** data set choosing $r_{\max} = 9/8$. For two different MC sample sizes, $K = 100$ and $K = 1000$, we used Brent’s optimisation algorithm (which combines parabolic interpolation with the golden section search algorithm, see Section 10.2 of Press et al., 1992) for finding the maximum of the log-likelihood. The MLE was estimated as 3.116, when $K = 100$ (in 6 seconds), and 3.112 when $K = 1000$ (in 57 seconds). Using numerical differentiation, the standard error was estimated as 0.04 both when $K = 100$ and $K = 1000$. The left panel of Figure 2 shows the estimate of the log-likelihood function based on the two different MC sample sizes.

When the assumed model is in \mathcal{D}_2 , it is mathematically much harder to derive a formula analogous to (17) which allows for simultaneous estimation of the likelihood function. In EA2 the minimum m and the time τ when the minimum occurs both depend on θ . Since the Poisson rate r is a function of m as well as θ , and ω is simulated conditionally on (τ, m) , the dependence of ω, Φ on θ is considerably harder than in EA1 to decouple and the method requires novel couplings and results from stochastic analysis. Such constructions are beyond the scope of this paper, and are presented in Beskos et al. (2005a), where several other issues related with SAM are also tackled. In particular it is proved that the simultaneous estimator for diffusions in \mathcal{D}_1 and \mathcal{D}_2 converges uniformly in θ to the likelihood function as the number of MC samples increases. Under standard assumptions, this guarantees consistency of the MC estimate of the MLE.

Example: Logistic growth. We used SAM in conjunction with the downhill simplex optimisation method (Section 10.4 of Press et al., 1992) to locate the MLE of the **LOG – GROWTH** data set, for MC sample sizes $K = 100$ and $K = 1000$. The estimates corresponding to (R, Λ, σ) were $(0.1098, 1014.89, 0.10057)$ for $K = 100$ (in 2 minutes) and $(0.1097, 1014.78, 0.10057)$ for $K = 1000$ (in 20 minutes). Using numerical differentiation the curvature of the estimated log-likelihood at the estimated MLE was found:

$$\begin{pmatrix} 4567 & 0.072 & -9601 \\ & 0.011 & -0.05 \\ & & 198464 \end{pmatrix}$$

The corresponding standard errors are (0.016, 9.5, 0.002). The plot at the right panel of Figure 2 shows the estimates of the profile log-likelihood of R for the two MC sample sizes.

6 The Poisson estimator

Corollary 1 relates the transition density of the diffusion to an expectation over the BB measure. Thus, any unbiased estimator of the expectation on the right-side of (11) corresponds to an unbiased estimator of $\tilde{p}_t(x, y; \theta)$. Generalizing, suppose that the expectation

$$\mathbb{E}_{\mathbb{P}^{(t,x,y)}} \left[\exp \left\{ - \int_0^t f(\omega_s) ds \right\} \right] < \infty, \quad (18)$$

is to be estimated for arbitrary continuous function f and diffusion bridge measure $\mathbb{P}^{(t,x,y)}$. For any $c \in \mathbf{R}$, $\lambda > 0$, and path ω , we write:

$$\exp \left\{ - \int_0^t f(\omega_s) ds \right\} = e^{-ct} \sum_{j=0}^{\infty} \frac{1}{j!} \left\{ \lambda t \int_0^t \frac{c - f(\omega_s)}{\lambda t} ds \right\}^j = e^{(\lambda-c)t} \mathbb{E} \left[\left\{ \int_0^t \frac{c - f(\omega_s)}{\lambda t} ds \right\}^{\kappa} \right],$$

where the expectation is taken with respect to $\kappa \sim \text{Poisson}(\lambda t)$, conditionally on ω . Notice that if $\psi \sim \text{Un}[0, t]$, then $(c - f(\omega_\psi))/\lambda$ is an unbiased estimator of $\int_0^t (c - f(\omega_s))/(\lambda t) ds$. Let $\Psi = \{\psi_1, \dots, \psi_\kappa\}$ be a Poisson process of rate λ on $[0, t]$, and $\omega \sim \mathbb{P}^{(t,x,y)}$. Then, we obtain the following simple unbiased estimator of (18), which we call the *Poisson estimator*:

$$e^{(\lambda-c)t} \lambda^{-\kappa} \prod_{j=1}^{\kappa} [c - f(\omega_{\psi_j})]. \quad (19)$$

This estimator was introduced in the context of statistical physics by Wagner (1988a, 1989). It can be derived from first principles that its second moment is

$$e^{(\lambda-2c)t} \mathbb{E}_{\mathbb{P}^{(t,x,y)}} \left[\exp \left\{ \frac{1}{\lambda} \int_0^t [c - f(\omega_s)]^2 ds \right\} \right], \quad (20)$$

which is not guaranteed to be finite. Choice of c and λ with view of improving the efficiency of the algorithm is discussed in Section 7.

Taking $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, and $f = (\alpha^2 + \alpha')/2$, the Poisson estimator can be used with (11) to estimate $\tilde{p}_t(x, y; \theta)$. This transition density estimator will also be referred to as the Poisson estimator. A simultaneous estimation of the complete map $\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta)$ for any data points v, w , merely requires decoupling of ω in (19) from x and y , since Ψ clearly is independent of θ . Since in the current context $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, it suffices to exploit the relocation-invariance property of BB and rewrite (19) as:

$$e^{(\lambda-c)t} \lambda^{-\kappa} \prod_{j=1}^{\kappa} \left[c - f \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y \right) \right], \quad \omega \sim \mathbb{W}^{(t,0,0)}. \quad (21)$$

This estimator will be referred to as the *simultaneous* Poisson estimator. For more general diffusion bridge measures, decoupling of ω from x, y can be cumbersome, since the intuitive relocation-invariance property does not hold in general. Such a case is treated in the following section.

6.1 Inference for the CIR model

We apply the Poisson estimator to infer about a diffusion process outside \mathcal{D}_2 , the Cox-Ingersoll-Ross (CIR) diffusion model, which solves

$$dV_s = \rho(\mu - V_s) ds + \sigma\sqrt{V_s} dB_s, \quad \theta = (\rho, \mu, \sigma).$$

It is assumed that all parameters are positive and $2\rho\mu > \sigma^2$, which guarantees that V does not hit zero (see pg.391 of Cox et al., 1985). We fit the CIR model to a well studied data set (used amongst others by Aït-Sahalia, 1996; Elerian et al., 2001; Roberts and Stramer, 2001) which contains daily Eurodollar rates between 1973 and 1995, to allow for comparisons with existing methods. We take a sub-sample of 550 values, corresponding to time intervals of 10 days, since the diffusion model seems more appropriate on that scale. We call this sub-sample the **EURODOLLAR** data set and plot it in the left panel of Figure 3.

In this case $\eta(u; \theta) = 2\sqrt{u}/\sigma$, thus the transformation $X_s = \eta(V_s; \theta)$ solves the SDE

$$dX_s = \left(\left[\frac{2\rho\mu}{\sigma^2} - \frac{1}{2} \right] \frac{1}{X_s} - \frac{\rho}{2} X_s \right) ds + dB_s.$$

Setting $\alpha(u; \theta) = (2\rho\mu/\sigma^2 - 0.5)/u - \rho u/2$, it is easy to check that the CIR model is outside \mathcal{D}_2 . X is a.s. positive and its distribution is absolutely continuous w.r.t. that of the Bessel process. Let $\mathbb{Q}_\theta^{(t,x,y)}$ and $\mathbb{R}^{(t,x,y)}$ be the bridge measures of X and the Bessel process respectively, and $\tilde{p}_t(x, y; \theta)$ and $q_t(x, y)$ be the corresponding transition densities. Direct application of Lemma 1 yields an analogue of (11) for positive diffusions:

$$\tilde{p}_t(x, y; \theta) = \frac{y}{x} q_t(x, y) \mathbb{E}_{\mathbb{R}^{(t,x,y)}} \left[\exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\} \right]. \quad (22)$$

Thus, we can readily use the Poisson estimator to estimate the transition density of CIR for a fixed $\theta \in \Theta$. However, simultaneous estimation for all $\theta \in \Theta$ is non-trivial, since the decoupling of $\omega \sim \mathbb{R}^{(t,x,y)}$ from x, y is not straightforward. The appropriate construction is in fact a variation of the approach we have devised for SAM for diffusions in \mathcal{D}_2 and is contained in Beskos et al. (2005a). This stems from the characterisation of the Bessel bridge as a BB conditioned to remain positive. As in EA2, we consider first the minimum m of the Bessel bridge and subsequently reconstruct the path given m . The distribution of m is the restriction on $(0, \infty)$ of the distribution of the minimum of the corresponding BB. Given m , the path can be reconstructed exactly as if it were BB with known minimum m .

We applied the simultaneous Poisson estimator to the **EURODOLLAR** data set taking, after some pilot tuning, $\lambda = c = 1$. At the right panel of Figure 3 we show the real and the estimated profile log-likelihood for σ using the downhill simplex optimisation algorithm. Notice that even with 100 MC samples the estimated curve coincides with the true one, although it can be shown that the variance of the estimator is infinite in this case.

7 Comparison of different transition density estimators

We have introduced three different methods which can be used to estimate the diffusion transition density. Whereas AM (and SAM) are devised solely for this purpose, it is impor-

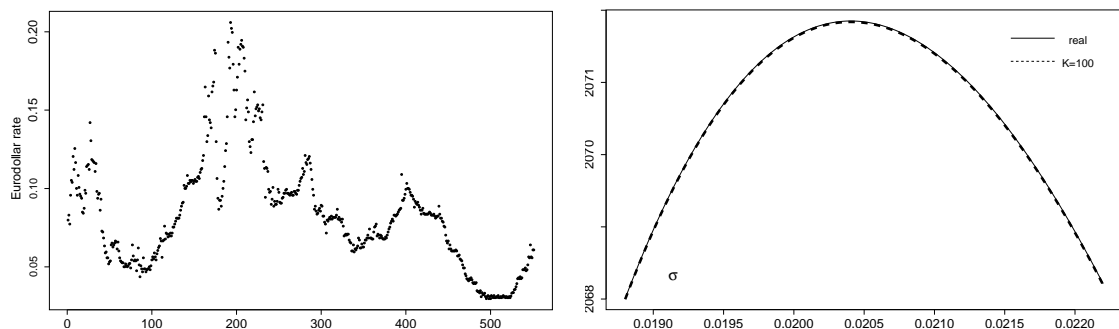


Figure 3: Left: The EURODOLLAR data set. Right: The real (solid line) profile log-likelihood of σ and its estimation (dashed line) using the simultaneous Poisson estimator with $\lambda = c = 1$ and $K = 100$ MC samples.

tant to recognise that the other two methods have greater scope. The Poisson estimator can estimate expectations of general diffusion exponential functionals (18). The bridge method is based on (12) which is a form of *Rao-Blackwellisation* that can be used to estimate other conditional probabilities for diffusions, e.g. hitting probabilities.

In the context of transition density estimation, the weakness of AM and the bridge method over the Poisson estimator is that their applicability is determined by that of EA. On the contrary, the Poisson estimator can be used without assuming that the drift α is suitably bounded.

A significant advantage of AM over the competitors is that it is guaranteed to have finite polynomial moments. For the bridge method little is known since we have not as yet derived analytic expressions for its variance. Checking whether the variance of the Poisson estimator is finite is in general tedious. A notable exception is when the diffusion is in \mathcal{D}_1 , where, since $f = (\alpha^2 + \alpha')/2$ is bounded, it is easy to see from (20) that the variance will be finite for any finite c, λ . Two examples outside \mathcal{D}_1 for which we know that this will not be the case are the CIR and the logistic growth models.

A feature of the bridge method which is likely to make it generally less efficient than the competitors is that the simulated paths ignore the final data point, since the method is founded on the *unconditional* EA.

When the aim is to explore features of the likelihood surface, it is imperative that the estimators yield estimates of the transition density simultaneously for all $\theta \in \Theta$. The bridge method is the only for which we have been unable to achieve that.

Although derived from very different perspectives, SAM and the simultaneous Poisson estimator are related. Taking $\lambda = r_{\max}$, $c = \lambda + l(\theta)$, and contrasting (17) with (21), reveals that in \mathcal{D}_1 , SAM is a special case of the simultaneous Poisson estimator. Moreover, there are certain optimality properties of this choice. For any c such that $c > r(\theta) + l(\theta)$, (20) is bounded above by $\exp\{(-2c + \lambda + (c - l(\theta))^2/\lambda)t\}$. This quantity is minimised by any pair (λ, c) such that $c = \lambda + l(\theta)$, where $\lambda \geq r(\theta)$. Requiring that the Poisson estimator yields estimates simultaneously for all $\theta \in \Theta$, the computationally most efficient bound on

Parameter		Min.	1 st Quart.	Median	3 rd Quart.	Max.
R	SAM	0.1051	0.1083	0.1092	0.1100	0.1130
	SimP	0.1030	0.1072	0.1087	0.1099	0.1139
Λ	SAM	1010.13	1013.91	1014.89	1016.07	1021.07
	SimP	1008.88	1013.48	1014.93	1016.50	1022.80
σ	SAM	0.10045	0.10053	0.10056	0.10058	0.10065
	SimP	0.10040	0.10050	0.10054	0.10058	0.10069

Table 1: Summary of 500 independent estimations of the MLE of the LOG – GROWTH data set. Each estimation is based on $K = 10$ MC samples. For each parameter, the first row corresponds to SAM and the second to the simultaneous Poisson estimator (SimP), with $\lambda = 1, c = l(\theta) + \lambda$. The average time for each experiment was around five times higher in SAM compared to the simultaneous Poisson estimator.

the variance is achieved by the choice $\lambda = r_{\max}$, $c = \lambda + l(\theta)$, under which the Poisson estimator and SAM coincide. Notice that in this choice, λ is the range and c the maximum of the functional $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}, \theta \in \Theta$. It is not obvious whether choosing $c > r(\theta) + l(\theta)$ is optimal. The connection between the two methods is less transparent outside \mathcal{D}_1 . The rate of the Poisson process used in SAM for diffusions in \mathcal{D}_2 , will depend on the minimum of the proposed path, and it is precisely due to this dependence why the estimator is a.s. bounded. The Poisson process and the proposed path in the Poisson estimator are inherently independent. Guided by our findings for \mathcal{D}_1 , we propose to choose λ according to an estimate of the range of $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}, \theta \in \Theta$, and take $c = \lambda + l(\theta)$. This choice has proved successful empirically.

A brief study of the performance of SAM and the simultaneous Poisson estimator in estimating the MLE of the LOG – GROWTH data set is summarised in Table 1. The parameters λ and c were chosen as suggested above. Taking into account its smaller computational cost, the simultaneous Poisson estimator is more efficient in this example. A feature not depicted in the table is the sensitivity of the simultaneous Poisson estimator on the choice of c, λ . The method can produce unreliable estimates, especially with few MC samples, for certain choices of the tuning parameters. On the contrary, SAM is fully automatic.

8 A Monte Carlo EM (MCEM) approach

The EM algorithm (Dempster et al., 1977) is suitable for locating the MLE when the likelihood of the observed data is intractable but the joint likelihood of the observed and some appropriately defined missing data is of a simple form. At the expense of some additional effort the EM can also be used to obtain an estimate of the observed information matrix. To fix ideas, let $\mathbf{v}_{obs} = \{V_0 = v, V_t = w\}$ be the observed data, which consist of discrete-time observations from (1) (without loss of generality we assume two data points). It is assumed throughout this section that $V \in \mathcal{D}_2$. When the diffusion coefficient does not depend on unknown parameters, that is when $\sigma(\cdot; \theta) = \sigma(\cdot)$, we can treat the paths between

consecutive observations as missing data and use the Girsanov's formula (32) to form the *complete likelihood*. This naturally leads to an EM algorithm whose implementation is described in the next section. On the other hand, when the diffusion coefficient depends on θ , the missing paths should be appropriately transformed, as shown in Section 8.2.

known diffusion coefficient

8.1 MCEM for diffusions with known diffusion coefficient

We first transform the process $V_s \rightarrow X_s = \eta(V_s)$, where η is defined in (3) and since by assumption $\sigma(\cdot; \theta) = \sigma(\cdot)$, η does not depend on θ . Thus, we define $\mathbf{x}_{obs} = \{X_0 = x, X_t = y\}$, $x = \eta(v)$, $y = \eta(w)$; \mathbf{x}_{obs} contains discrete-time observations from the diffusion X with SDE (4). The observed log-likelihood is $\ell(\theta | \mathbf{x}_{obs}) = \log \tilde{p}_t(x, y; \theta)$. Let \mathbf{x}_{mis} denote the missing path, $(X_s, s \in [0, t])$, and $\mathbf{x}_{com} = \{\mathbf{x}_{obs}, \mathbf{x}_{mis}\}$ denote the complete data, which is the continuous diffusion path starting from $X_0 = x$ and finishing at $X_t = y$. The complete log-likelihood is given by Girsanov's formula (32)

$$\ell(\theta | \mathbf{x}_{com}) = A(y; \theta) - A(x; \theta) - \frac{1}{2} \int_0^t (\alpha^2 + \alpha')(X_s; \theta) ds, \quad (23)$$

and

$$\mathbf{x}_{mis} | \mathbf{x}_{obs}; \theta \sim \mathbb{Q}_\theta^{(t, x, y)}.$$

The E-step of the EM algorithm requires analytic evaluation of

$$Q(\theta, \theta') = \mathbb{E}_{\mathbf{x}_{mis} | \mathbf{x}_{obs}; \theta'} [\ell(\theta | \mathbf{x}_{com})], \quad \theta \in \Theta.$$

However, notice that we can introduce the random variable $U \sim \text{Un}[0, t]$, which is independent of both \mathbf{x}_{mis} and \mathbf{x}_{obs} , and re-write

$$Q(\theta, \theta') = A(y; \theta) - A(x; \theta) - \frac{t}{2} \cdot \mathbb{E}_{(\mathbf{x}_{mis}, U) | \mathbf{x}_{obs}; \theta'} [(\alpha^2 + \alpha')(X_U; \theta)]. \quad (24)$$

We are unable to perform analytically the expectation above. However, since we can simulate U , and X_u for any u using EA, we adopt an MC implementation of the EM algorithm. The MCEM algorithm was introduced in Wei and Tanner (1990); convergence and implementation issues are tackled in Chan and Ledolter (1995); Sherman et al. (1999); Fort and Moulines (2003). It is well documented (see e.g. Fort and Moulines, 2003, and references therein) that the number of MC samples used to approximate the expectation should increase with the EM iterations. We have followed this approach in our examples. Several methods have been proposed for using the EM algorithm to estimate the observed information matrix; see for example Louis (1982); Jamshidian and Jennrich (2000); Meng and Rubin (1991). We have used the method suggested by Louis (1982), where we use MC estimations of the required expectations, as in Wei and Tanner (1990).

Example: Periodic drift. We applied MCEM to the SINE data set; the EM iterations are presented in the first column of Table 2. The M-step was implemented using Brent's optimisation method. The convergence is very rapid, essentially in one iteration, and the estimated MLE is in perfect agreement with the one obtained with SAM in Section 5.1. The standard error was estimated to be 0.04.

8.2 The general case: missing information and a path transformation

An important result from stochastic analysis, which has found numerous applications in statistical inference for diffusion processes (see for instance Barndorff-Nielsen and Shephard, 2002), is that a continuous-time diffusion path on an interval of time $[0, t]$ can be used to estimate perfectly the parameters involved in $\sigma(\cdot; \theta)$. A similar result does not hold for parameters in the drift $b(\cdot; \theta)$, where perfect estimation from continuous-time data typically holds asymptotically as $t \rightarrow \infty$. On the other hand, a finite number of discrete-time observations can contain only finite information about any of the parameters. The computational implication is that we cannot construct an EM algorithm as in Section 8.1 where the paths between the observed data are treated as missing data. According to the EM terminology, this augmentation scheme leads to a fraction of missing information equal to 1. This problem was first encountered in an MCMC framework by Roberts and Stramer (2001) and Elerian (1999).

One remedy to this type of problems is to find a suitable invertible transformation of the missing data in order to reduce the augmented information. In particular, we want to transform the process so that the diffusion coefficient is independent of θ , exploiting the result that continuous-time paths contain only finite information about the drift parameters. Our approach is in the spirit of Roberts and Stramer (2001). We start by transforming the process, $V_s \rightarrow X_s = \eta(V_s; \theta)$, where η is the transformation in (3), therefore $X_0 = x(\theta) = \eta(v; \theta)$ and $X_t = y(\theta) = \eta(w; \theta)$. Notice that unlike Section 8.1, $(X_s, s \in [0, t])$ is not directly observed, instead it is a function of the unknown parameters. The second level of path transformation is from $X_s \rightarrow \dot{X}_s$ where:

$$\dot{X}_s := X_s - \left(1 - \frac{s}{t}\right)x(\theta) - \frac{s}{t}y(\theta), \quad s \in [0, t]. \quad (25)$$

\dot{X} is a diffusion bridge starting from $\dot{X}_0 = 0$ and finishing at $\dot{X}_t = 0$. Its dynamics depend on θ and are typically intractable, nevertheless it is easy to simulate \dot{X}_s at any time $s \in [0, t]$, conditionally upon \mathbf{v}_{obs} and a specific value of θ , following the procedure:

1. Find $x = x(\theta)$, $y = y(\theta)$.
2. Simulate X_s from the bridge diffusion measure $\mathbb{Q}_\theta^{(t,x,y)}$ using EA.
3. Apply the transformation in (25).

The inverse transformation from \dot{X} to X is:

$$g_\theta(\dot{X}_s) := \dot{X}_s + \left(1 - \frac{s}{t}\right)x(\theta) + \frac{s}{t}y(\theta), \quad s \in [0, t].$$

We propose an augmentation scheme where $\mathbf{v}_{mis} = (\dot{X}_s, s \in [0, t])$ and $\mathbf{v}_{com} = \{\mathbf{v}_{obs}, \mathbf{v}_{mis}\}$. The following lemma, proved in Appendix B, gives the complete log-likelihood.

Lemma 2. *The log-likelihood of the complete data \mathbf{v}_{com} is given by*

$$\begin{aligned} \ell(\theta | \mathbf{v}_{com}) &= \log |\eta'(w; \theta)| + \log \mathcal{N}_t(y(\theta) - x(\theta)) \\ &+ A(y(\theta); \theta) - A(x(\theta); \theta) - \int_0^t \frac{1}{2}(\alpha^2 + \alpha')(g_\theta(\dot{X}_s); \theta) ds. \end{aligned}$$

Iteration j	$\theta^{(j)}$	$R^{(j)}$	$\Lambda^{(j)}$	$\sigma^{(j)}$
0	0.500	0.7000	1500.00	0.7000
1	3.119	0.1099	1015.32	0.1006
2	3.114	0.1097	1014.80	0.1006
3	3.115	0.1097	1014.83	0.1006
4	3.115	0.1097	1014.51	0.1006
5	3.114	0.1097	1014.66	0.1006
6	3.113	0.1097	1014.76	0.1006
7	3.113	0.1097	1014.67	0.1006
8	3.113	0.1097	1014.75	0.1006
9	3.113	0.1097	1014.71	0.1006
10	3.113	0.1095	1014.78	0.1006

Table 2: MCEM iterations for the SINE (first column) and the LOG – GROWTH (last three columns) data sets. The first five iterations were performed using 200 MC samples and the last five using 2000 samples. The execution times were 177 seconds for SINE and 18 minutes for LOG – GROWTH.

Arguing as in Section 8.1, we have that

$$\begin{aligned}
Q(\theta, \theta') &= \log |\eta'(w; \theta)| + \log \mathcal{N}_t(y(\theta) - x(\theta)) \\
&+ A(y(\theta); \theta) - A(x(\theta); \theta) - \frac{t}{2} \mathbb{E}_{(\mathbf{v}_{mis}, U) | \mathbf{v}_{obs}; \theta'} [(\alpha^2 + \alpha')(g_\theta(\dot{X}_U); \theta)].
\end{aligned}$$

We can now apply MCEM in a similar fashion as in Section 8.1.

Example: Logistic growth. We used MCEM to estimate the MLE and the observed information for the LOG – GROWTH data set. The results in Table 2 indicate rapid convergence. We used the Fletcher-Reeves version of the conjugate gradient method for the M-step of the algorithm (see Section 10.5 of Press et al., 1992). The standard errors are estimated using $K = 2000$ as $(0.015, 9.5, 0.002)$.

9 Hierarchical simulation models and inference using MCMC

In this section we develop an MCMC algorithm for Bayesian inference for discretely observed diffusions. The stationary distribution of our MCMC is the *exact* posterior distribution of the parameters. However, it is not only the *exactness* which distinguishes our approach from competitive existing MCMC algorithms. Existing methods follow a path augmentation approach, as we did in the EM algorithm in Section 8. The missing continuous paths are approximated by a fine discrete-time Markov chain, whose transition is assumed to follow an Euler-type approximation to the true diffusion transition. The joint distribution of the observed and missing data is given by an appropriate approximation of Girsanov’s formula. Then, a Gibbs sampler (or more general component-wise updating MCMC) is used to sample from the approximate posterior distribution of the parameters and the missing paths. Often, for any two data points several points need to be imputed

in between them to obtain a good approximation to the true posterior of θ . However the performance of basic MCMC schemes can severely deteriorate as the amount of imputation increases; see Roberts and Stramer (2001) for details.

The approach introduced here is not based on augmentation of paths. Instead, we construct a graphical model which involves the variables used in EA, and we show that the posterior distribution of the parameters is obtained as a marginal in this graphical model. Thus, we use an appropriate Metropolis-Hastings algorithm to sample from the joint posterior of all variables in the graph. One of the steps of the sampler involves running the *conditional* EA. The state-space of our MCMC typically has much smaller dimension than the competing augmentation methods and as a result it can be computationally more efficient. However, comparison between alternative MCMC methods is not carried out in this paper, and will be reported elsewhere.

We describe in detail the MCMC algorithm when it is assumed that $V \in \mathcal{D}_1$. The derivation of the algorithm when $V \in \mathcal{D}_2$ is technically much harder, and it can be found in Beskos et al. (2004a). Nevertheless, we present simulation results for both cases. An essential ingredient of the algorithm, is the following lemma which derives the density of the output of the conditional EA1.

Lemma 3. *Consider any two fixed points x and y . Let $\Phi = \{\Psi, \Upsilon\}$ be the marked Poisson process on $[0, t] \times [0, 1]$ with rate $r(\theta)$ and number of points $\kappa \sim Po(r(\theta)t)$, which is used in EA1 for simulating from $\mathbb{Q}_\theta^{(t,x,y)}$. Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and I be the acceptance indicator (16) which decides whether $(\omega_s + (1 - s/t)x + (s/t)y, s \in [0, t])$ is accepted as a path from $\mathbb{Q}_\theta^{(t,x,y)}$. Then, the conditional density of ω and Φ given $\{I = 1\}$, $\pi(\omega, \Phi \mid \theta, x, y, I = 1)$, is*

$$\frac{\exp\{t[1 - r(\theta)]\} r(\theta)^\kappa}{a(x, y, \theta)} \prod_{j=1}^{\kappa} \mathbb{I} \left[\phi \left(\omega_{\psi_j} + \left(1 - \frac{\psi_j}{t}\right)x + \frac{\psi_j}{t}y; \theta \right) < v_j \right], \quad (26)$$

w.r.t. the product measure $\Phi \times \mathbb{W}^{(t,0,0)}$, where Φ is the measure of a unit-rate Poisson process on $[0, t] \times [0, 1]$, and $a(x, y, \theta)$ is the acceptance probability of EA1.

To clarify the notation, in the remainder of the section (ω^*, Φ^*) represent the accepted random elements in EA1, with density $\pi(\omega^*, \Phi^* \mid \theta, x, y)$ given in (26); Φ^* , and ω^* at any collection of times, can be easily sampled using the conditional EA1.

Let $\pi(\theta)$ denote an appropriately specified prior density for θ . The aim is to sample from the posterior distribution of θ given diffusion observations $\mathbf{v} = \{V_{t_0}, \dots, V_{t_n}\}$, say $\pi(\theta \mid \mathbf{v})$, where $V \in \mathcal{D}_1$. It is convenient to define

$$x_i = x_i(\theta) := \eta(V_{t_i}; \theta), \quad i = 0, \dots, n, \quad (27)$$

where η is the transformation in (3). For a given θ , the x_i s are discrete-time samples from SDE (4). For $1 \leq i \leq n$, let $\omega_i^* = (\omega_{i,s}^*, s \in [t_{i-1}, t_i])$ and $\Phi_i^* = \{\Psi_i^*, \Upsilon_i^*\}$ be the accepted elements of the conditional EA1 applied on the time interval $[t_{i-1}, t_i]$; $S(\omega_i^*)$ will denote the corresponding output skeleton. We construct the following hierarchical model, defined

through the densities:

$$\begin{aligned}\theta &\sim \pi(\theta) \\ V_{t_i} | V_{t_{i-1}}, \theta &\sim p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta) \\ (\omega_i^*, \Phi_i^*) | \theta, x_{i-1}(\theta), x_i(\theta) &\sim \pi(\omega_i^*, \Phi_i^* | \theta, x_{i-1}(\theta), x_i(\theta)).\end{aligned}\quad (28)$$

Notice that in this hierarchical model the observed data are in the middle of the hierarchy. We term (28) a *hierarchical simulation model*, to emphasise that it represents the order in which variables have to be simulated to ensure that the output of EA1, ω_i^* , is indeed from $\mathbb{Q}_\theta^{(\Delta t_i, x_{i-1}, x_i)}$. The posterior density of interest, $\pi(\theta | \mathbf{v})$, is a marginal of the joint posterior density of θ and the latent variables, $\pi(\theta, \{\omega_i^*, \Phi_i^*, 1 \leq i \leq n\} | \mathbf{v})$. We aim at sampling from this density via the Gibbs sampler. The conditional density of the latent variables is,

$$\pi(\{\omega_i^*, \Phi_i^*, 1 \leq i \leq n\} | \theta, \mathbf{v}) = \prod_{i=1}^n \pi(\omega_i^*, \Phi_i^* | \theta, x_{i-1}(\theta), x_i(\theta)), \quad (29)$$

thus the pairs (ω_i^*, Φ_i^*) are conditionally independent with density given in (26). The key property of this hierarchical simulation model is that θ is independent of $\{(\omega_i^*, \Phi_i^*), 1 \leq i \leq n\}$ given the collection of skeletons $\{S(\omega_i^*), 1 \leq i \leq n\}$. Specifically, the conditional density of θ given the observed data and the latent variables is given in the following theorem.

Theorem 3. θ is conditionally independent of $\{(\omega_i^*, \Phi_i^*), 1 \leq i \leq n\}$ given \mathbf{v} and $\{S(\omega_i^*), 1 \leq i \leq n\}$, with density $\pi(\theta | \{S(\omega_i^*), 1 \leq i \leq n\}, \mathbf{v})$ proportional to:

$$\begin{aligned}\pi(\theta) r(\theta)^{\sum \kappa_i^*} \exp \left\{ A(x_n(\theta); \theta) - A(x_0(\theta); \theta) - [l(\theta) + r(\theta)] \sum_{i=1}^n \Delta t_i \right\} \times \\ \prod_{i=1}^n |\eta'(x_i(\theta); \theta)| \mathcal{N}_{\Delta t_i}(x_i(\theta) - x_{i-1}(\theta)) \times \\ \prod_{i=1}^n \left\{ \prod_{j=1}^{\kappa_i^*} \left[1 - \phi \left(\omega_{i, \psi_{i,j}^*}^* + (1 - \psi_{i,j}^*/\Delta t_i) x_{i-1}(\theta) + (\psi_{i,j}^*/\Delta t_i) x_i(\theta); \theta \right) \right] \right\}, \quad \theta \in \Theta.\end{aligned}\quad (30)$$

Notice that (30) is given in a simple computable form, and the intractable terms $a(x_{i-1}, x_i, \theta)$ are not involved when conditioning upon $\{S(\omega_i^*), 1 \leq i \leq n\}$. Our proposed MCMC algorithm is a component-wise updating algorithm, simulating iteratively the skeletons $\{S(\omega_i^*), 1 \leq i \leq n\}$ and θ from their posterior conditional distributions. The skeletons $\{S(\omega_i^*), 1 \leq i \leq n\}$ are conditionally independent given θ and can be easily simulated using EA1. In some cases it might be possible to sample from (30) directly, otherwise a Metropolis-Hastings step could be used. When a random walk Metropolis is being used, it will generally make sense to scale the proposal variance to be proportional to n^{-1} .

Example: Periodic drift. We implemented the algorithm for the SINE data set. We used a uniform prior on $[0, 2\pi]$. It took 47 seconds to run the algorithm for 10,000 iterations, and some summaries are shown in the rightmost column of Figure 4. The posterior mean

is estimated as 3.1127 and the posterior standard deviation as 0.04. We used a Metropolis step to update θ which had acceptance probability 0.49. The algorithm mixes very rapidly, and essentially the autocorrelation in the θ series is due to the fact that a Metropolis step is used rather than direct simulation from its conditional distribution. The dependence between θ and the latent variables is very weak.

When the assumed model is in \mathcal{D}_2 , the construction of the MCMC algorithm is more complicated. In particular, derivation of the joint density of $(\theta, \omega^*, \Phi^*)$ is challenging due to the more complex structure of EA2. This is done in Beskos et al. (2004a), where also other important issues related with the implementation of the MCMC are tackled. The algorithm necessitates recent non-centring reparametrisation methodology for hierarchical models as described in Roberts et al. (2004); Papaspiliopoulos et al. (2003).

Example: Logistic growth. Using the MCMC algorithm described in Beskos et al. (2004a), we obtained 50,000 samples from the posterior distribution of the parameters (R, Λ, σ) for the LOG – GROWTH data set. The computing time was 20 minutes, and summary of the results is given in Figure 4. The posterior means were estimated as (0.1075, 1017.4, 0.1007), the posterior precision matrix as

$$\begin{pmatrix} 4589 & 0.065 & -9408 \\ & 0.001 & 0.116 \\ & & 196060 \end{pmatrix}$$

from which the posterior standard deviations read as (0.015, 31.13, 0.002).

10 Conclusions

In this paper we have introduced a variety of methods which can be used for likelihood-based inference for discretely observed diffusions. The methods rely on recent advances in the exact simulation of diffusions. The computational efficiency of the methods was illustrated in a collection of examples. However, an exhaustive simulation study which tests the relative performance of our methods and existing approaches, under various model specifications and parameter settings, is not given here. Such a detailed empirical investigation is currently taking place. However, some qualitative remarks are made below.

In general, the computing time required for our methods critically depends on the rate of the Poisson process. *Ceteris paribus*, this rate is a function of the time increment Δt between the observations. The performance (measured either in computing time or MC error) of our methods will be very strong for small Δt (high frequency data). However, the performance of the methods as presented here, deteriorates for sparser data sets (Δt large). Specifically, the problem stems from the following two characteristics of the conditional EA: i) the Poisson rate is linear in Δt (in EA2 it can increase even faster), ii) the acceptance rate typically decreases exponentially to 0 in Δt . The computing time required for AM and SAM increases due to i) and for MCEM and MCMC due to i) and ii). In AM and SAM, the variance will also increase with Δt , since MC averages would be heavily dominated by the terms (of very small probability) corresponding to Poisson configurations with very



Figure 4: Summary of MCMC results: trace plots (top row), autocorrelation plots (middle) and posterior density estimates (bottom). Last column: 10,000 iterations of MCMC for the **SINE** data set (first 500 samples removed as burn-in for autocorrelation and density estimation). First three columns: 50,000 iterations of MCMC for the **LOG – GROWTH** data set (burn-in of 2,000 iterations).

few points. Similarly, (20) suggests that the variance of the Poisson estimator increases exponentially with Δt . In MCMC, the computing cost can be transformed to be linear in Δt . This is achieved by augmenting any two observed data with additional points in between. This is implemented by applying the conditional EA on intervals of length a fraction of Δt . Of course, the additional augmentation will affect the MCMC mixing.

A number of extensions of our methodology are possible. One important direction is the extension of our methods for diffusions outside the \mathcal{D}_2 class. As we have shown, the Poisson estimator can be readily used for likelihood inference for more general diffusion processes. On the other hand, current progress on EA itself, which attempts to remove the boundedness conditions on the drift (Beskos et al., 2005b) are expected to broaden the applicability of the estimation methods. It is worth noticing, that MCEM and AM remain unaltered irrespective of which version of EA is being used, thus they can readily accommodate extensions in EA. The other methods use explicitly the structure of EA and will have to be modified appropriately.

We have concentrated here on the case where the diffusion is observed without error at a finite set of times. However, data often occurs in different forms. For instance, data might be subject to observation error. All the methods for evaluating MLEs are difficult to extend to this case. However, the MCMC approach can be extended in a straightforward manner. An alternative interesting data form is when we observe a one-dimensional component of a higher dimensional diffusion, as for instance in continuous-time filtering models. We are currently working on a collection of such filtering problems, and

we have found that our methods can be extended to this case, although there are significant additional implementation challenges in this approach.

Our methodology extends some way to time-inhomogeneous and multivariate diffusions. In carrying out these extensions, there are two steps in the arguments used in this paper that need to be generalised. Firstly, it is necessary to generalise the transformation (3) which eliminates the diffusion coefficient. This is straightforward under mild smoothness conditions on σ in the time-inhomogeneous extension. For multivariate extensions, however, the generalised version of (3) involves the solution of an appropriate vector differential equation which is often intractable or insolvable (see e.g. Ait-Sahalia, 2004). This imposes restrictions on the class of multivariate diffusions to which the methods presented here can currently be applied. Secondly, we need to eliminate the stochastic integral from Girsanov's formula (32) to derive Lemma 1. Again this is fairly routine in the time-inhomogeneous case, while an extra condition is needed in the multivariate case, requiring that the multivariate drift be the gradient of a suitable potential function. This is a well-known condition in stochastic analysis and for ergodic, unit-diffusion coefficient diffusions corresponds to reversibility.

It is natural to ask whether the ideas of this paper extend to SDEs driven by Lévy processes. However, the Cameron-Martin-Girsanov formula, providing a closed form likelihood function, is critical to all our methodology. Unfortunately, there does not exist an analogous expression for the Radon-Nikodym derivative of an infinite activity Lévy driven SDE with respect to a tractable and easy to simulate from measure. Nevertheless, it is straightforward to extend our methods to incorporate SDEs with jumps according to a finite activity Lévy process (as for example considered in Roberts et al., 2004).

We hope that the collection of techniques described in this paper will have further applications. We are currently working on MC estimation of derivative hedge ratios in finance. The methodology builds on Rao-Blackwellisation techniques such as those devised for the bridge method.

Whilst we recognise that the mathematical details of this work are complex for those without a working knowledge of diffusion theory, we firmly believe that our methods have the potential to influence applied statistical work. This is because the algorithms we use are relatively simple and easy to code, the methods are not computationally demanding and are capable of handling long time-series. In addition, motivated by the desire to make our work as accessible as possible, we have started developing generic software for the implementation of some of our methods, beginning with EA.

Acknowledgements

The authors would like to thank Mike Pitt and Neil Shephard for helpful suggestions. We are grateful to all referees for their comments which contributed to the improvement of the paper. The first author acknowledges financial support from the Greek State Scholarships Foundation and the second author is funded by EPSRC grant GR/S61577/01.

References

- Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385–426.
- (2002) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, 223–262.
- (2004) Closed-form likelihood expansions for multivariate diffusions. Working paper, available from <http://www.princeton.edu/~yacine/research.htm>.
- Asmussen, S., Glynn, P. and Pitman, J. (1995) Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, **5**, 875–896.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**, 253–280.
- Bertoin, J. and Pitman, J. (1994) Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.*, **118**, 147–166.
- Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2004a) Markov chain Monte Carlo for exact inference for discretely observed diffusions. In progress.
- (2004b) Retrospective exact simulation of diffusion sample paths with applications. Submitted, available from <http://www.maths.lancs.ac.uk/~papaspil/research.html>.
- (2005a) Monte carlo maximum likelihood estimation for discretely observed diffusion processes. Submitted.
- (2005b) A new factorisation of diffusion measure with view towards simulation. In progress.
- Beskos, A. and Roberts, G. O. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, **15**. To appear.
- Bibby, B. M., Jacobsen, M. and Sørensen, M. (2002) Estimating functions for discretely sampled diffusion-type models. In *Handbook of financial Econometrics*. Amsterdam: North-Holland.
- Black, F. and Scholes, M. S. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–659.
- Chan, K., Karolyi, A. G., Longstaff, F. A. and Sanders, A. B. (1992) An empirical comparison of alternative models of the short-term interest rate. *J. Finance*, **47**, 1209–1227.
- Chan, K. S. and Ledolter, J. (1995) Monte Carlo EM estimation for time series models involving counts. *J. Amer. Statist. Assoc.*, **90**, 242–252.

- Cox, J. C., Ingersoll, Jr., J. E. and Ross, S. A. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.
- Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38. With discussion.
- Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, **20**, 297–338. With comments and a reply by the authors.
- Elerian, O. (1999) *Simulation estimation of continuous time series models with applications to finance*. Ph.D. thesis, Nuffield College, Oxford.
- Elerian, O., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, **69**, 959–993.
- Eraker, B. (2001) MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, **19**, 177–191.
- Fort, G. and Moulines, E. (2003) Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, **31**, 1220–1259.
- Gallant, A. R. and Long, J. R. (1997) Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, **84**, 125–141.
- Gillespie, D. T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Computational Phys.*, **22**, 403–434.
- (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Gobet, E. (2002) LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, **38**, 711–737.
- Goel, N. S. and Richter-Dyn, N. (1974) *Stochastic models in biology*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London.
- Gourieroux, C. and Jasiak, J. (2003) Jacobi process and smooth transitions. Working paper.
- Gourieroux, C., Monfort, A. and Renault, E. (1993) Indirect inference. *J. Applied Econometrics*, **8**, 85–118.
- Jamshidian, M. and Jennrich, R. I. (2000) Standard errors for EM estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **62**, 257–270.

- Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.
- Kimura, M. and Ohta, T. (1971) *Theoretical aspects of population genetics*. Princeton University Press.
- Kingman, J. F. C. (1993) *Poisson processes*, vol. 3 of *Oxford Studies in Probability*. New York: The Clarendon Press Oxford University Press. Oxford Science Publications.
- Kloeden, P. and Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag.
- Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **44**, 226–233.
- McAdams, H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, **94**, 814–819.
- Meng, X.-L. and Rubin, D. B. (1991) Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, **86**, 899–909.
- Merton, R. C. (1971) Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, **3**, 373–413.
- Obuhov, A. M. (1959) Description of turbulence in terms of Lagrangian variables. In *Advances in Geophysics, Vol. 6 (Symposium on Atmospheric Diffusion and Air Pollution, Oxford, 1958)*, 113–116. New York: Academic Press.
- Papaspiliopoulos, O. and Roberts, G. O. (2004) Retrospective MCMC methods for Dirichlet process hierarchical models. Submitted.
- Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003) Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian statistics, 7 (Tenerife, 2002)*, 307–326. New York: Oxford Univ. Press. With a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors.
- Pardoux, É. and Pignol, M. (1984) Étude de la stabilité de la solution d'une EDS bilinéaire à coefficients périodiques. Application au mouvement des pales d'hélicoptère. In *Analysis and optimization of systems, Part 2 (Nice, 1984)*, vol. 63 of *Lecture Notes in Control and Inform. Sci.*, 92–103. Berlin: Springer.
- Pedersen, A. R. (1995) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, **1**, 257–279.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) *Numerical recipes in C*. Cambridge: Cambridge University Press, second edn. The art of scientific computing.

- Revuz, D. and Yor, M. (1994) *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Berlin: Springer-Verlag, second edn.
- Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P. (2004) Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66**, 369–393.
- Roberts, G. O. and Stramer, O. (2001) On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603–621.
- Santa-Clara, P. (1995) Simulated likelihood estimation of diffusions with an application to the short-term interest rate. Working paper.
- Sherman, R. P., Ho, Y.-Y. K. and Dalal, S. R. (1999) Conditions for convergence of Monte Carlo EM sequences with an application to product diffusion modeling. *Econom. J.*, **2**, 248–267.
- Shiga, T. (1985) Mathematical results on the stepping stone model in population genetics. In *Population genetics and molecular evolution (Mishima, 1984)*, 267–279. Tokyo: Japan Sci. Soc. Press.
- Sørensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. To appear.
- Wagner, W. (1988a) Monte Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples. *Stochastic Anal. Appl.*, **6**, 447–468.
- (1988b) Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.*, **79**, 336–352.
- (1989) Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stochastics Stochastics Rep.*, **28**, 1–20.
- Wei, G. C. G. and Tanner, M. A. (1990) A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Appendix A: Background Material

1. Brownian motion and Brownian bridge

A standard Brownian motion $\{B_s; 0 \leq s \leq t\}$, $B_0 = 0$, can be simulated on a collection of time instances $0 = s_0 < s_1 < \dots < s_k \leq t$ in the following recursive way:

$$B_{s_j} = B_{s_{j-1}} + Y_j, \quad Y_j \sim N(0, s_j - s_{j-1}), \quad 1 \leq j \leq k.$$

A standard BB with distribution $\mathbb{W}^{(t,0,0)}$ can be obtained as a transformation of B :

$$B_s - \frac{s}{t}B_t, \quad s \in [0, t].$$

An arbitrary BB with distribution $\mathbb{W}^{(t,x,y)}$ can be simulated via the relocation-invariance property of BB: if $\omega \sim \mathbb{W}^{(t,0,0)}$, then $\omega_s + (1 - s/t)x + (s/t)y$, is a path from $\mathbb{W}^{(t,x,y)}$.

2. Decomposition of a Brownian path at its minimum

The following theorem, stated in Beskos et al. (2004b), shows how to simulate a BB path together with its minimum m and the time τ when the minimum is attained.

Theorem 4. *A path $\omega \sim \mathbb{W}^{(t,x,y)}$ can be simulated in the following way:*

1. Simulate m, τ according to their joint density:

$$\pi(m, \tau) = \sqrt{2\pi} t \frac{(y-m)(x-m)}{\sqrt{\tau^3(t-\tau)^3}} \exp \left\{ -\frac{(x-m)^2}{2\tau} - \frac{(y-m)^2}{2(t-\tau)} + \frac{(y-x)^2}{2t} \right\}, \quad (31)$$

for $m \leq \min\{x, y\}$ and $\tau \in [0, t]$.

2. The rest of the path is a transformation of 6 independent standard BBs, $b_j \sim \mathbb{W}^{(1,0,0)}$, $j = 1, \dots, 6$. Analytically, for $0 \leq s \leq \tau$,

$$\omega_s = m + \sqrt{\tau} \cdot \sqrt{\left(\frac{(x-m)(\tau-s)}{\tau^{3/2}} + b_{1,(1-s/\tau)} \right)^2 + b_{2,(1-s/\tau)}^2 + b_{3,(1-s/\tau)}^2},$$

and for $\tau < s \leq t$,

$$\omega_s = m + \sqrt{t-\tau} \cdot \sqrt{\left(\frac{(y-m)(s-\tau)}{(t-\tau)^{3/2}} + b_{4,(s-\tau)/(t-\tau)} \right)^2 + b_{5,(s-\tau)/(t-\tau)}^2 + b_{6,(s-\tau)/(t-\tau)}^2}.$$

In Step 1, m is drawn from its marginal distribution, which is a transformed Rayleigh distribution, and conditionally on m , τ is drawn using the inverse CDF method.

3. Bessel process and Bessel bridge

The (3-dimensional) Bessel process/bridge can be defined as a scalar Brownian motion/bridge constrained to be positive; for main results see Chapter 11 of Revuz and Yor (1994). Recall that an arbitrary Bessel bridge measure is denoted by $\mathbb{R}^{(t,x,y)}$. Simulation of $\omega \sim \mathbb{R}^{(t,0,y)}$ can be carried out by means of three independent BBs (see e.g. Bertoin and Pitman, 1994):

$$\omega_s = \sqrt{(ys + b_{1,s})^2 + b_{2,s}^2 + b_{3,s}^2}, \quad 0 \leq s \leq t, \quad b_j \sim \mathbb{W}^{(t,0,0)}.$$

Simulation of $\omega \sim \mathbb{R}^{(t,x,y)}$ is more involved and follows the lines of Theorem 4. We first simulate its minimum m , whose distribution is the restriction on $(0, \infty)$ of the distribution of the minimum of the corresponding BB. Given m , the path can be reconstructed exactly as if it were BB with known minimum m .

Appendix B: Proofs of main results

Proof of Lemma 1

Girsanov's formula, which gives the density of the law of X w.r.t. Wiener measure is

$$\exp \left\{ \int_0^t \alpha(\omega_s; \theta) d\omega_s - \frac{1}{2} \int_0^t \alpha^2(\omega_s; \theta) ds \right\}. \quad (32)$$

Thus, using Itô's Lemma

$$\begin{aligned} \frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(\omega) &= \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)} \exp \left\{ \int_0^t \alpha(\omega_s; \theta) d\omega_s - \frac{1}{2} \int_0^t \alpha^2(\omega_s; \theta) ds \right\} \\ &= \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)} \exp \left\{ A(y;\theta) - A(x;\theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\}. \end{aligned}$$

□

Proof of Lemma 2

We will find the density $\pi(V_t, \dot{X} | \theta)$ of the law of \mathbf{v}_{com} with respect to $\text{Leb} \times \mathbb{W}^{(t,0,0)}$, where Leb denotes the Lebesgue measure on \mathbf{R} . Clearly,

$$\pi(V_t, \dot{X} | \theta) = p_t(V_0, V_t; \theta) \times \pi(\dot{X} | V_t, \theta),$$

The law of the 1-1 transformation $g_\theta(\dot{X})$ given V_t is just $\mathbb{Q}_\theta^{(t,x,y)}$, where $x = \eta(V_0; \theta)$ and $y = \eta(V_t; \theta)$. Thus,

$$\pi(\dot{X} | V_t, \theta) = \frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(g_\theta(\dot{X}))$$

Use of Lemma 1 concludes the proof. □

Proof of Lemma 3

Notice that the Radon-Nikodym derivative of a homogeneous Poisson process with rate $r(\theta)$ on $[0, t] \times [0, 1]$ w.r.t. Φ is:

$$\exp\{-r(\theta)t\} \frac{r(\theta)^\kappa t^\kappa}{\kappa!} \times \left[\exp\{-t\} \frac{t^\kappa}{\kappa!} \right]^{-1},$$

which simplifies to $\exp\{t[1 - r(\theta)]\} r(\theta)^\kappa$. □

Proof of Theorem 3

The joint density of θ and the latent variables conditionally on \mathbf{v} can be decomposed as

$$\pi(\theta, \{\omega_i^*, \Phi_i^*, 1 \leq i \leq n\} | \mathbf{v}) = \pi(\theta | \mathbf{v}) \times \pi(\{\omega_i^*, \Phi_i^*, 1 \leq i \leq n\} | \theta, \mathbf{v})$$

To simplify the formulae, we write x_i instead of $x_i(\theta)$. Clearly,

$$\pi(\theta | \mathbf{v}) \propto \pi(\theta) \prod_{i=1}^n \tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta) |\eta'(x_i; \theta)|$$

Using (14), the acceptance probability of EA1 $a(x_{i-1}, x_i, \theta)$, can be expressed as:

$$\tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta) \cdot [\mathcal{N}_{\Delta t_i}(x_i - x_{i-1}) \exp\{A(x_i; \theta) - A(x_{i-1}; \theta) - l(\theta)\Delta t_i\}]^{-1}$$

Combining this expression with (29) and (26) we find that $\pi(\theta, \{\omega_i^*, \Phi_i^*, 1 \leq i \leq n\} | \mathbf{v})$ is proportional to:

$$\begin{aligned} & \pi(\theta) r(\theta)^{\sum \kappa_i^*} \exp \left\{ A(x_n; \theta) - A(x_0; \theta) - [l(\theta) + r(\theta)] \sum_{i=1}^n \Delta t_i \right\} \times \\ & \prod_{i=1}^n |\eta'(x_i; \theta)| \mathcal{N}_{\Delta t_i}(x_i - x_{i-1}) \times \\ & \prod_{i=1}^n \left\{ \prod_{j=1}^{\kappa_i^*} \mathbb{I} \left[\phi \left(\omega_{i, \psi_{i,j}^*}^* + (1 - \psi_{i,j}^*/\Delta t_i)x_{i-1} + (\psi_{i,j}^*/\Delta t_i)x_i; \theta \right) < v_{i,j}^* \right] \right\} \end{aligned}$$

It is now straightforward to integrate out $\Upsilon_i^* = \{v_{i,j}, 1 \leq j \leq \kappa_i^*\}$ and obtain (30). \square