Bernoulli **12**(6), 2006, 1077–1098

Retrospective exact simulation of diffusion sample paths with applications

ALEXANDROS BESKOS*, OMIROS PAPASPILIOPOULOS** and GARETH O. ROBERTS†

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.
*E-mail: *a.beskos@lancaster.ac.uk; **o.papaspiliopoulos@lancaster.ac.uk,*
†*g.o.roberts@lancaster.ac.uk*

Keywords: conditioned diffusion processes; discretely observed diffusions; exact simulation; Monte Carlo maximum likelihood; rejection sampling

We present an algorithm for exact simulation of a class of Itô's diffusions. We demonstrate that when the algorithm is applicable, it is also straightforward to simulate diffusions conditioned to hit specific values at predetermined time instances. We also describe a method that exploits the properties of the algorithm to carry out inference on discretely observed diffusions without resorting to any kind of approximation apart from the Monte Carlo error.

1. Introduction

Applications of diffusion models are ubiquitous throughout science; a representative list might include finance, biology, physics and engineering. In many cases the evolution of some phenomenon is described by a scalar stochastic process $X = \{X_t;\, 0 \leqslant t \leqslant T\}$ determined as the solution of a stochastic differential equation (SDE) of the type

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \qquad X_0 = x \in \mathbf{R},\ t \in [0,\ T], \tag{1}$$

driven by the Brownian motion $\{B_t;\, 0 \leqslant t \leqslant T\}$. The drift b and the coefficient function σ are presumed to satisfy the regularity conditions (locally Lipschitz, with a growth bound) that guarantee the existence of a weakly unique, global solution of (1); see Kloeden and Platen (1995: Chapter 4). In fact, we will restrict our attention to SDEs of the type

$$\mathrm{d}X_t = \alpha(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 = x \in \mathbf{R},\ t \in [0,\ T], \tag{2}$$

for some drift function α, since (1) can be transformed into an SDE of unit coefficient under mild additional conditions on σ, by applying the transformation $X_t \to \eta(X_t)$, where

$$\eta(x) = \int_z^x \frac{1}{\sigma(u)}\,\mathrm{d}u, \tag{3}$$

with z some element of the state space of X.

Simulation and inference for SDEs of the form (1) generally require some kind of

1350–7265

discrete approximation (see, for example, Roberts and Stramer 2001; Kloeden and Platen 1995). To do this, an Euler approximation of the SDE might be used:

$$X_{t+\Delta} = X_t + \alpha(X_t)\Delta + N(0, \Delta).$$

Such methods are correct only in an infill asymptotic sense (i.e. as $\Delta \to 0$ with $N\Delta = T$). When applicable, our exact algorithm returns skeletons of exact paths of X.

The exact algorithm carries out rejection sampling using Brownian paths as proposals, and returns skeletons of the target SDE (2) obtained at some random time instances. The skeletons can be filled in later *independently* of X by interpolation of Brownian bridges. An initial version of the exact algorithm was given in Beskos and Roberts (2005). Its most demanding restriction was that the functional $\alpha^2 + \alpha'$ of the drift be bounded. In this paper we describe an easier and more flexible method of applying the rejection sampling algorithm proposed by Beskos and Roberts, but most importantly we substantially relax the boundedness condition on the drift. Assuming that $\alpha^2 + \alpha'$ is bounded from below, we now require that $\limsup_{u\to\infty}(\alpha^2 + \alpha')(u)$ and $\limsup_{u\to-\infty}(\alpha^2 + \alpha')(u)$ are not both $+\infty$. Thus, the method introduced in this paper can now be applied to a substantially wider class of diffusion processes.

A major appeal of the exact algorithm is that it can be adapted to provide solutions to a variety of challenging problems related to diffusion processes, and in this paper we explore some of these possibilities. Maximum likelihood inference for discretely observed diffusions is known to be particularly difficult, even when the drift and diffusion coefficients are restricted to small-dimensional parametric classes. This problem has received considerable attention in the last two decades from the statistical, applied probability and econometric community; see, for example, Sørensen (2004) for a recent review and references. All current methods are subject to approximation error, and are often either difficult to implement or subject to substantial Monte Carlo error. We show how our simulation algorithm can easily be used to provide efficient Monte Carlo maximum likelihood estimators (MLEs).

The exact algorithm can also be used to simulate diffusion processes conditioned to hit specific values at predetermined time instances (diffusion bridges). Contrary to all other current simulation methods, with our approach the conditional simulation is in fact easier than the unconditional simulation.

The structure of the paper is as follows. In Section 2 we present the Exact Algorithm. In Section 3 we describe analytically the extension noted above; it requires the stochastic calculus theory that makes possible a simulation-oriented decomposition of a Brownian path at its minimum. Section 4 presents some results on the efficiency of the algorithm. In Section 5 we apply our algorithm to simulate exactly from the otherwise intractable logistic growth model. In Section 6 we demonstrate the use of the method for the simulation of conditional diffusions and parametric inference. We finish in Section 7 with some general remarks about the exact algorithm and possible extensions in future research.

2. Retrospective rejection sampling for diffusions

Before we introduce the basic form of the exact algorithm we require the following preliminary notation. Let $C \equiv C([0, T], \mathbf{R})$ be the set of continuous mappings from $[0, T]$ to $\mathbf{R}$ and ω be a typical element of C. Consider the coordinate mappings $B_t : C \to \mathbf{R}$, $t \in [0, T]$, such that for any t, $B_t(\omega) = \omega(t)$ and the cylinder σ-algebra $\mathcal{C} = \sigma(\{B_t; 0 \leqslant t \leqslant T\})$. We denote by $W^x = \{W_t^x; 0 \leqslant t \leqslant T\}$ the Brownian motion starting at $x \in \mathbf{R}$.

Let $\mathbb{Q}$ be the probability measure induced by the solution X of (2) on $(C, \mathcal{C})$, that is, the measure with respect to which the coordinate process $B = \{B_t; 0 \leqslant t \leqslant T\}$ is distributed according to X, and $\mathbb{W}$ the corresponding probability measure for W^x. The objective is to construct a rejection sampling algorithm to draw from $\mathbb{Q}$. The Girsanov transformation of measures (see, for instance, Øksendal 1998: Chapter 8) implies that

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{W}}(\omega) = \exp\left\{\int_0^T \alpha(B_t)\,\mathrm{d}B_t - \frac{1}{2}\int_0^T \alpha^2(B_t)\,\mathrm{d}t\right\}.$$

Under the condition that α is everywhere differentiable we can eliminate the Itô integral after applying Itô's lemma to $A(B_t)$ for $A(u) := \int_0^u \alpha(y)\,\mathrm{d}y$, $u \in \mathbf{R}$. Simple calculations give

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{W}}(\omega) = \exp\left\{A(B_T) - A(x) - \frac{1}{2}\int_0^T \left(\alpha^2(B_t) + \alpha'(B_t)\right)\mathrm{d}t\right\}. \tag{4}$$

To remove the possible inconvenience of A being unbounded and in any case to simplify the Radon–Nikodym derivative of the two probability measures, we will use candidate paths from a process identical to W^x except for the distribution of its ending point. We call such a process a *biased* Brownian motion.

We thus consider the biased Brownian motion defined as $\hat{W} \stackrel{d}{=} (W^x \mid W_T^x \sim h)$ for a density h proportional to $\exp\{A(u) - (u - x)^2/2T\}$, $u \in \mathbf{R}$. It is now necessary that this function is integrable. Let $\mathbb{Z}$ be the probability measure induced on $(C, \mathcal{C})$ by this process. Note that it is straightforward to simulate a skeleton of $\omega \sim \mathbb{Z}$ just by simulating first its ending point $\omega(T) \sim h$. The measures $\mathbb{Z}$ and $\mathbb{W}$ are equivalent and their Radon–Nikodym derivative can be easily derived from the following proposition, proved in the Appendix:

Proposition 1. *Let* $M = \{M_t; 0 \leqslant t \leqslant T\}$, $N = \{N_t; 0 \leqslant t \leqslant T\}$ *be two stochastic processes on* $(C, \mathcal{C})$ *with corresponding probability measures* $\mathbb{M}$, $\mathbb{N}$. *Assume that* f_M, f_N *are the densities of the ending points* M_T *and* N_T *respectively, with identical support* $\mathbf{R}$. *If it is true that* $(M \mid M_T = \rho) \stackrel{d}{=} (N \mid N_T = \rho)$, *for all* $\rho \in \mathbf{R}$, *then*

$$\frac{\mathrm{d}\mathbb{M}}{\mathrm{d}\mathbb{N}}(\omega) = \frac{f_M}{f_N}(B_T).$$

Therefore, $\mathrm{d}\mathbb{W}(\omega)/\mathrm{d}\mathbb{Z} = \mathcal{N}_{x,T}/h(B_T) \propto \exp\{-A(B_T)\}$, where $\mathcal{N}_{x,T}$ represents the density of the normal distribution with mean x and variance T, and

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Z}}(\omega) = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{W}}(\omega)\,\frac{\mathrm{d}\mathbb{W}}{\mathrm{d}\mathbb{Z}}(\omega) \propto \exp\left\{-\int_0^T \left(\frac{1}{2}\alpha^2(B_t) + \frac{1}{2}\alpha'(B_t)\right)\mathrm{d}t\right\}.$$

Assume now that $(\alpha^2 + \alpha')/2$ is bounded below. Then we can obtain a non-negative function ϕ such that

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Z}}(\omega) \propto \exp\left\{-\int_0^T \phi(B_t)\,\mathrm{d}t\right\} \leqslant 1, \quad \mathbb{Z}\text{-almost surely.} \tag{5}$$

Analytically, ϕ is defined as

$$\phi(u) = \frac{\alpha^2(u) + \alpha'(u)}{2} - k, \qquad u \in \mathbf{R}, \text{ for a fixed } k \leqslant \inf_{u\in\mathbf{R}}(\alpha^2 + \alpha')(u)/2. \tag{6}$$

We now summarize the three conditions that allow the derivation of (5):

1. The drift function α is differentiable.
2. The function $\exp\{A(u) - (u-x)^2/2T\}$, $u \in \mathbf{R}$, for $A(u) = \int_0^u \alpha(y)\,\mathrm{d}y$, is integrable.
3. The function $(\alpha^2 + \alpha')/2$ is bounded below.

In a simulation context, were it possible to draw complete continuous paths $\omega \sim \mathbb{Z}$ on $[0, T]$ and calculate the integral involved in (5) analytically, then rejection sampling on the probability measures $\mathbb{Q}$, $\mathbb{Z}$ would be straightforward. The exact algorithm manages to circumvent these difficulties in order to carry out exact rejection sampling using only finite information about the proposed paths from $\mathbb{Z}$. Given a proposed path $\omega \sim \mathbb{Z}$, it defines an event of probability $\exp\{-\int_0^T \phi(B_t)\,\mathrm{d}t\}$. The truth or falsity of this event can be determined after unveiling ω only at a finite collection of time instances. The rejection sampling scheme is carried out in a *retrospective* way since the realization of the proposed variate (the path $\omega \sim \mathbb{Z}$) at some required instances follows that of the variates that decide for the acceptance or not of this proposed variate. The idea of retrospective sampling, in a different context, is introduced in Papaspiliopoulos and Roberts (2004), where it is applied to an MCMC algorithm for Bayesian analysis from a Dirichlet mixture model.

We now present a simple theorem that demonstrates the idea behind the method.

Theorem 1. *Let ω be any element of $C([0, T], \mathbf{R})$ and $M(\omega)$ an upper bound for the mapping $t \mapsto \phi(\omega_t)$, $t \in [0, T]$. If Φ is a homogeneous Poisson process of unit intensity on $[0, T] \times [0, M(\omega)]$ and N is the number of points of Φ found below the graph $\{(t, \phi(B_t));\ t \in [0, T]\}$, then*

$$\mathrm{P}[N = 0 \mid \omega] = \exp\left\{-\int_0^T \phi(B_t)\,\mathrm{d}t\right\}.$$

Proof: Conditionally on ω, N follows a Poisson distribution with mean $\int_0^T \phi(B_t)\,\mathrm{d}t$. □

Theorem 1 suggests that were it possible to generate complete, continuous paths $\omega \sim \mathbb{Z}$ of the biased Brownian motion $\tilde{W}$ then we could carry out rejection sampling on $\mathbb{Q}$, $\mathbb{Z}$ without

having to evaluate the integral involved in (5); we would only have to generate a realization of the Poisson process Φ and check if all the points of Φ fall above the ϕ-graph to decide on the acceptance (or rejection in any other case) of ω.

The simple observation that only finite information about some $\omega \sim \mathbb{Z}$ suffices for the algorithm to decide about it will yield a simple, valid rejection sampling scheme which involves finite computations. The technical difficulty of locating a rectangle for the realization of the Poisson process, or equivalently an upper bound for the map $t \mapsto \phi(\omega_t)$, imposes some restrictions on the applicability of the algorithm.

2.1. The case when ϕ is bounded

Let M be an upper bound of ϕ. This is the simple case and is similar to the one considered in Beskos and Roberts (2005). We now devise a different approach which is simpler and even more efficient than the one presented by Beskos and Roberts. Moreover, this method is particularly useful for the more general case of Section 3. Based on Theorem 1 and (5), we can generate a feasible rejection sampling scheme just by thinking retrospectively, that is, first realizing the Poisson process and then constructing the path $\omega \sim \mathbb{Z}$ only at the time instances required to determine N.

Exact Algorithm 1

1. *Produce a realization* $\{x_1, x_2, \ldots, x_\tau\}$, *of the Poisson process* Φ *on* $[0, T] \times [0, M]$, *where* $x_i = (x_{i,1}, x_{i,2})$, $1 \leqslant i \leqslant \tau$.
2. *Simulate a skeleton of* $\omega \sim \mathbb{Z}$ *at the time instances* $\{x_{1,1}, x_{2,1}, \ldots, x_{\tau,1}\}$.
3. *Evaluate* N.
4. *If* $N = 0$ *go to step 5, else go to step 1.*
5. *Output the currently constructed skeleton* $S(\omega)$ *of* ω.

Exact Algorithm 1(EA1) returns an exact skeleton of X and is shown (see Section 4) to run in finite time. In Section 4 we present details about the efficiency of the algorithm. When a skeleton $S(\omega)$ is accepted as a realization from $\mathbb{Q}$, we can continue constructing it according to $\mathbb{Z}$, that is, using Brownian bridges between the successive currently unveiled instances of ω. In this way, the corresponding path of $\mathbb{Q}$ can be realized at any requested time instances.

Figure 1 illustrates the retrospective idea behind EA1. Figure 1(b) emphasizes the fact that skeletons can be readily filled in using independent Brownian bridges and without further reference to the dynamics of the target process X. This is precisely the property we shall exploit in Section 6.1 to carry out maximum likelihood parametric inference.

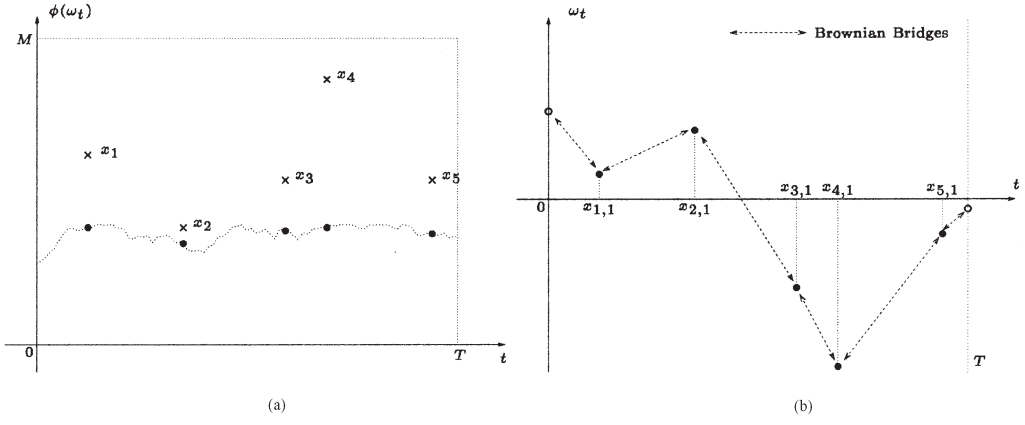

Figure 1. The retrospective idea: crosses (×) show the points from a realization $\{x_1, \ldots, x_5\}$ of Φ. The black dots in (a) show the only instances of the path $\phi(\omega_t)$ we needed to realize before accepting (since $N = 0$) the corresponding unveiled instances of ω (the black dots in (b)) as an exact skeleton of X. Intermediate points of the skeleton are obtained via Brownian bridges between successive instances of it.

3. The case when either $\limsup_{u\to\infty} \phi(u) < \infty$ or $\limsup_{u\to-\infty} \phi(u) < \infty$

It is well-known that it is possible to construct a Brownian path (or a biased Brownian path such as $\hat{W}$) on a bounded time interval after simulating first its minimum (or maximum) and then the rest of the path using Bessel(3) processes. For definitions and properties of Bessel processes, see, for instance, Karatzas and Shreve (1991) and Revuz and Yor (1991).

We will exploit this decomposition of the Brownian path to present an extension of EA1 when either $\limsup_{u\to\infty} \phi(u) < \infty$ or $\limsup_{u\to-\infty} \phi(u) < \infty$. Without loss of generality, we will consider the case when $\limsup_{u\to\infty} \phi(u) < \infty$; it is then possible to identify an upper bound $M(\omega)$ for the mapping $t \mapsto \phi(\omega_t)$, $t \in [0, T]$, after decomposing the proposed path ω at its minimum, say b, and considering

$$M(\omega) \equiv M(b) = \sup\{\phi(u);\, u \geqslant b\} < \infty. \tag{7}$$

By symmetry, the same theory covers the case when $\limsup_{u\to-\infty} \phi(u) < \infty$. To describe the decomposition of a Brownian path at its minimum we need to recall some properties of Brownian motion.

3.1. Decomposing the Brownian path at its minimum

Let $W = \{W_t;\, 0 \leqslant t \leqslant T\}$ be a Brownian motion starting at 0, $m_T = \inf\{W_t;\, 0 \leqslant t \leqslant T\}$ and $\theta_T = \sup\{t \in [0, T] : W_t = m_T\}$. Then, for any $a \in \mathbf{R}$,

$$\mathrm{P}[m_T \in \mathrm{d}b,\ \theta_T \in \mathrm{d}t \mid W_T = a] \propto \frac{b(b-a)}{\sqrt{t^3(T-t)^3}} \exp\left\{-\frac{b^2}{2t} - \frac{(b-a)^2}{2(T-t)}\right\} \mathrm{d}b\,\mathrm{d}t, \tag{8}$$

with $b \leqslant \min\{a, 0\}$ and $t \in [0, T]$. For the derivation of this distribution see, for instance, Karatzas and Shreve (1991: Chapter 2). In the following proposition (proved in the Appendix) we describe a simple algorithm for drawing from (8). We denote by Unif(0, 1) the uniform distribution on (0, 1) and by IGau(μ, λ), $\mu > 0$, $\lambda > 0$, the inverse Gaussian distribution with density

$$\mathrm{IGau}(\mu, \lambda, u) = \sqrt{\frac{\lambda}{2\pi u^3}} \exp\left\{-\frac{\lambda(u-\mu)^2}{2\mu^2 u}\right\}, \qquad u > 0.$$

A very simple algorithm for drawing from this density is described in Devroye (1986: Chapter IV). We take $\mathbb{I}[a < b]$ to be 0 if $a > b$ and 1 otherwise, for any $a, b \in \mathbb{R}$.

Proposition 2. *Let $E(1)$ be a random variable distributed according to the exponential distribution with unit mean and $Z_1 = (a - \sqrt{2\,TE(1) + a^2})/2$. If $Z_1 = b$ is a realization of Z_1, set $c_1 = (a-b)^2/2T$, $c_2 = b^2/2T$. Let $U \sim \mathrm{Unif}(0, 1)$, $I_1 \sim \mathrm{IGau}(\sqrt{c_1/c_2}, 2c_1)$ and $I_2 \sim 1/\mathrm{IGau}(\sqrt{c_2/c_1}, 2c_2)$ independently, and define*

$$V = \mathbb{I}[U < (1 + \sqrt{c_1/c_2})^{-1}] \cdot I_1 + \mathbb{I}[U \geqslant (1 + \sqrt{c_1/c_2})^{-1}] \cdot I_2.$$

Then the pair (Z_1, Z_2), for $Z_2 := T/(1+V)$, is distributed according to (8).

Recall from the definition of $\hat{W}$ that to construct a path $\omega \sim \mathbb{Z}$ it is necessary that we first draw its ending point. Thus, we actually have to decompose a Brownian bridge at its minimum which justifies the conditioning on W_T at (8). Proposition 2 provides precisely the method for drawing the minimum and the time instance when it is achieved for a path $\omega \sim \mathbb{Z}$ given its ending point; Theorem 2 below gives the way of filling in the remainder of ω; Figure 2 illustrates the construction.

We denote by $R(\delta) = \{R_t(\delta);\ 0 \leqslant t \leqslant 1\}$ a three-dimensional Bessel bridge of unit length from 0 to $\delta \geqslant 0$ and by $W^c = (W \mid m_T = b,\ \theta_T = t,\ W_T = a)$ the Brownian motion of length T starting at 0 and conditioned on obtaining its minimum b at time t and ending at a.

Theorem 2. *The processes $\{W^c_s;\ 0 \leqslant s \leqslant t\}$ and $\{W^c_s;\ t \leqslant s \leqslant T\}$ are independent with*

$$\{W^c_s;\ 0 \leqslant s \leqslant t\} \stackrel{d}{=} \sqrt{t}\{R_{(t-s)/t}(\delta_1);\ 0 \leqslant s \leqslant t\} + b,$$

$$\{W^c_s;\ t \leqslant s \leqslant T\} \stackrel{d}{=} \sqrt{T-t}\{R_{(s-t)/(T-t)}(\delta_2);\ t \leqslant s \leqslant T\} + b,$$

where $\delta_1 = -b/\sqrt{t}$ and $\delta_2 = (a-b)/\sqrt{T-t}$.

Proof: This is Proposition 2 of Asmussen *et al.* (1995) with the difference that we have rescaled the Bessel processes to obtain bridges of unit length and the Brownian path to incorporate the case where decomposition of a path of arbitrary length T is requested. □

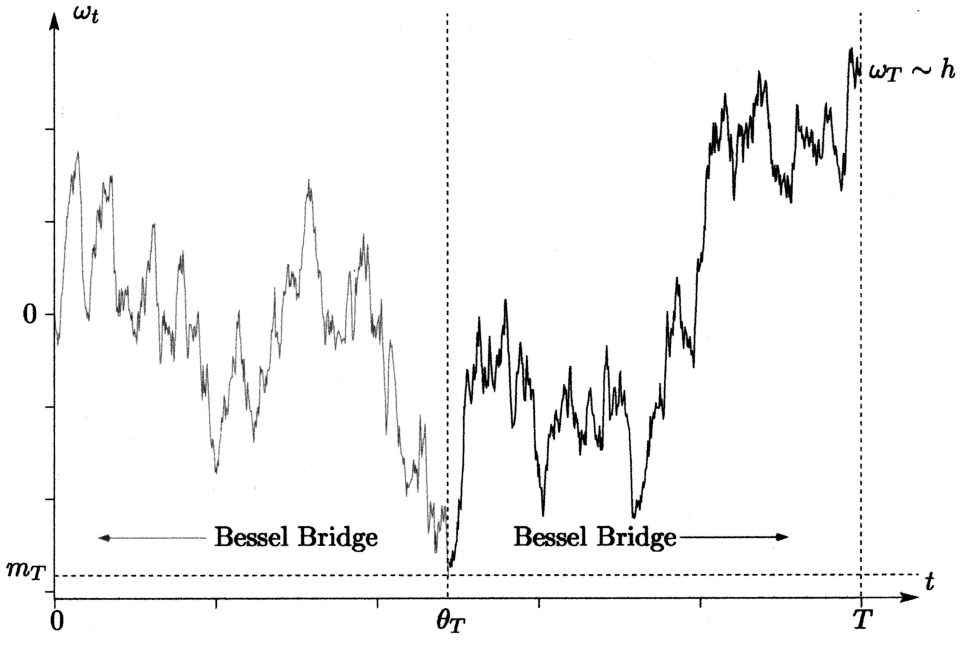

Figure 2. The decomposition of $\omega \sim \mathbb{Z}$, at its minimum. The order in the simulation of the involved random elements is ω_T, m_T, θ_T. The two independent Bessel bridges connect the already simulated instances of ω.

It is well known (see Bertoin and Pitman 1994) that if W_1^{br}, W_2^{br}, W_3^{br} are three independent standard Brownian bridges (bridges of unit length, starting and ending at 0) then we can produce a Bessel bridge $R(\delta)$ as:

$$R_t(\delta) = \sqrt{(\delta \cdot t + W_{1,t}^{br})^2 + (W_{2,t}^{br})^2 + (W_{3,t}^{br})^2}, \qquad t \in [0, 1].$$

Recall also (Bertoin and Pitman 1994) that a standard Brownian bridge W^{br} can easily be expressed in terms of an unconditional Brownian motion W starting at 0 via the transformation $W_t^{br} = W_t - t\,W_1$.

3.2. The extension of the exact algorithm

Theorem 2 makes possible the construction of a subroutine, say decompose $(T, \beta_0, \beta_T, t, b, \mathbf{s})$, that returns the location at the time instances $\mathbf{s} = \{s_1, \ldots, s_n\}$ of a Brownian path on $[0, T]$ starting at β_0, conditioned on obtaining its minimum b at time t and finishing at β_T. Furthermore, from Proposition 2, we can extend EA1 for exact simulation from (2) when the functional $\phi \geqslant 0$ considered in (5) is not necessarily bounded above but satisfies the weaker condition $\limsup_{u\to\infty} \phi(u) < \infty$. In translating the

construction to the present setting, the Poisson process Φ defined in Theorem 1 in this case is a point process on rectangles whose height depends on the current proposed path ω.

Exact Algorithm 2

1. *Initiate a path* $\omega \sim \mathbb{Z}$ *on* $[0, T]$ *by drawing* $\omega_T \sim h$.
2. *Simulate its minimum* m *and the moment* θ *when it is achieved.*
3. *Find an upper bound* $M(m)$ *for* $t \mapsto \phi(\omega_t)$, $t \in [0, T]$.
4. *Produce a realization* $\{x_1, x_2, \ldots, x_\tau\}$ *of* Φ *in* $[0, T] \times [0, M(m)]$.
5. *Call decompose() to construct* ω *at* $\{x_{1,1}, x_{2,1}, \ldots, x_{\tau,1}\}$.
6. *Evaluate* N.
7. *If* $N = 0$ *go to step 8, else go to step 1.*
8. *Output the currently constructed skeleton* $S(\omega)$ *of* ω.

Note that, in comparison with the skeleton returned by EA1, in Exact Algorithm 2 (EA2) the skeleton $S(\omega)$ is more informative about the underlying continuous path ω since it includes its minimum and the instance when it is achieved. As with the initial algorithm, we can continue filling in the accepted skeleton $S(\omega)$ at any other time instances under the two independent Bessel bridges' dynamics.

4. The efficiency of the exact algorithms

It is critical for the efficiency of the exact algorithm that we can exploit the Markov property of the target diffusion X (2) and merge skeletons of some appropriately chosen length T (or even of varying lengths) to construct a complete skeleton on some larger interval. In what follows we preserve T to denote the length of the time interval upon which the exact algorithm is applied (and which is typically smaller than the length, say l, of the sample path of interest).

Assume that we apply the exact algorithm on X over the time interval $[0, T]$ for some starting point $x \in \mathbf{R}$. Let ϵ denote the probability of accepting a proposed path and D the number of Poisson process points needed to decide whether to accept the proposed path. Let $N(T)$ be the total number of Poisson process points needed until the first accepted path. The following proposition is proved in the Appendix.

Proposition 3. *Consider EA1 with* $\phi \leqslant M$. *Then*

$$\epsilon \geqslant \exp(-M \cdot T), \qquad \mathrm{E}[D] = M \times T.$$

For EA2 with $M = M(\omega)$ *defined as in* (7),

$$\epsilon \geqslant \exp\{-\mathrm{E}[M(\hat{m})] \cdot T\}, \qquad \mathrm{E}[D] = \mathrm{E}[M(\hat{m})] \times T,$$

where $\hat{m}$ *is the minimum of a proposed path* $\omega \sim \mathbb{Z}$. *In both cases,*

$$\mathrm{E}[N(T)] \leqslant \mathrm{E}[D] / \epsilon.$$

It is natural to ask how we should implement the algorithms in order to simulate the diffusion on the time interval $[0, KT]$ for some positive integer K. For concreteness, consider EA1. Proposition 3 suggests that implementing the rejection sampling algorithm on the entire interval will incur a computational cost which is $O(KMTe^{KMT})$, which will be huge even for moderate K-values. On the other hand, the simulation problem can be broken down into K simulations performed sequentially, each on an interval of length T, and each incurring a cost which is $O(MTe^{MT})$ and therefore giving an overall cost which is $O(KMTe^{MT})$, linear in the length of the time interval required.

Clearly there are important practical problems of algorithm optimization (by choosing T) for simulating from a diffusion on a fixed time-scale. We shall not address these here, except to remark that the larger M is, the smaller T ought to be.

It is difficult to obtain explicit results on the efficiency of EA2. From Proposition 3 it is clear that to show that the algorithm runs in finite time we need to bound $\mathrm{E}[M(\hat{m})]$, which is difficult. The following proposition (proved in the Appendix) gives a general result we have obtained, and which will be found useful in the case of the logistic growth model example of Section 5.

Proposition 4. *Assume that the drift function α is bounded below and there exist $k > 0$ and b_0 such that $M(b) \leqslant \exp(-kb)$ for $b \leqslant b_0$. Then for any starting value x and any length T of the time interval under consideration, $\mathrm{E}[M(\hat{m})] < \infty$. Therefore for EA2, the expected number of proposed Poisson points needed to accept a sample path is finite.*

Note that once we have established a theoretical result about the finiteness of $\mathrm{E}[M(\hat{m})]$ we can easily proceed to a Monte Carlo estimation of its actual value.

4.1. User-impatience bias

Many non-trivial simulation schemes return results in unbounded, random times. If the experiment is designed in such a way that its running time is *not* independent of the output, then the results can be 'user-impatience' biased; the experimenter might be tempted (or even obliged) to stop a long run and restart the simulation procedure. So the sample will be biased towards the observations that require shorter runs. The case where this situation has appeared in its most concrete form is the coupling from the past (CFTP) technique for the perfect simulation of distributions using Markov chains. The discussion in Section 7 of Propp and Wilson (1998) presents analytically the problem in the CFTP case.

A similar problem can be identified for EA2. The time to decide about the acceptance or otherwise of the current proposed path ω depends on ω. If D is the number of Poisson process points to decide about a proposed path, then $\mathrm{E}[D \mid \omega] = T \times M(\hat{m}(\omega))$, for T the length of the time interval under consideration, $\hat{m}(\omega)$ the minimum of the path ω and M as defined in (7). Since, in theory, the proposed path ω can produce any small value, it is impossible to bound $\mathrm{E}[D \mid \omega]$ uniformly over all ω. For the example with the logistic growth model we present in Section 5 we had to adjust the length T of the time interval to which EA2 is applied, according to the values of the parameters. In this way we have not only

approximately optimized efficiency, but also minimized the effects of the user-impatience bias.

In line with the suggestions in Propp and Wilson (1998), the experimenter should try some preliminary executions of the algorithm to obtain some evidence about the distribution of the running time and, if necessary (and possible), redesign the algorithm so that this distribution is of small variance.

5. Application: the logistic growth model

EA2 can be applied to the stochastic analogue of the logistic growth model:

$$\mathrm{d}V_t = r\,V_t(1 - V_t/K)\,\mathrm{d}t + \beta\,V_t\,\mathrm{d}B_t, \qquad V_0 = v > 0,\ t \in [0,\ T], \tag{9}$$

for positive parameters r, β, K. This diffusion is useful for modelling the growth of populations. The instantaneous population of some species V_t grows, in the absence of any restraints, exponentially fast in t with growth rate per individual equal to r. The actual evolution of the population is cut back by the saturation-inducing term $(1 - V_t/K)$. The constant $K > 0$ is called the 'carrying capacity' of the environment and usually represents the maximum population that can be supported by the resources of the environment. The parameter β represents the effect of the noise on the dynamics of V. For an example of how (9) can be derived, see Goel and Richter-Dyn (1974: Chapter 6).

After setting $X_t = -\log(V_t)/\beta$ we obtain the SDE

$$\mathrm{d}X_t = \left\{\frac{\beta}{2} - \frac{r}{\beta} + \frac{r}{\beta K}\exp(-\beta X_t)\right\}\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 = x = -\frac{\log(v)}{\beta},\ t \in [0,\ T]. \tag{10}$$

It can be shown using the theory of scale and speed densities (see Karlin and Taylor 1981: Section 15.7), that the original process V with probability 1 does not hit 0 in finite time for any values of the parameters, so considering $\log(V_t)$ is valid. Let α be the drift function of the modified SDE (10). Then

$$(\alpha^2 + \alpha')(u) = \frac{r^2}{\beta^2 K^2}\exp(-2\beta u) - \frac{2r^2}{\beta^2 K}\exp(-\beta u) + \left(\frac{\beta}{2} - \frac{r}{\beta}\right)^2, \qquad u \in \mathbf{R}.$$

This mapping is bounded below and has the property $\limsup_{u\to+\infty}(\alpha^2 + \alpha')(u) < \infty$, so it satisfies the conditions required for the applicability of EA2.

We follow the analysis of Section 2 to construct the Radon–Nikodym derivative (5) that is appropriate for rejection sampling. The ending point of the biased Brownian motion $\hat{W}$ must be distributed according to a density h proportional to

$$\exp\left\{\left(\frac{\beta}{2} - \frac{r}{\beta}\right)u - \frac{r}{\beta^2 K}\mathrm{e}^{-\beta u} - \frac{(u-x)^2}{2T}\right\} \propto \exp\left\{-\frac{(u-g_1)^2}{2T} - g_2\mathrm{e}^{-\beta u}\right\}, \qquad u \in \mathbf{R}, \tag{11}$$

for $g_1 = x + T(\beta/2 - r/\beta)$, $g_2 = r/(\beta^2 K)$. We can draw from this density in a very efficient way via rejection sampling with envelope function

$$\exp\left\{-\frac{(u-g_1)^2}{2T} - g_2 e^{-\beta g_1}(1-\beta(u-g_1))\right\}, \qquad u \in \mathbf{R},$$

This function is proportional to an $\mathcal{N}(g_1 + T\beta g_2 e^{-\beta g_1}, T)$ density. We can now obtain (5) for $\mathbb{Z}$ the probability measure that corresponds to the biased Brownian motion $\hat{W}$ with ending point $\hat{W}_T$ distributed according to (11) and

$$\phi(u) = \frac{r^2}{2\beta^2 K^2} e^{-2\beta u} - \frac{r^2}{\beta^2 K} e^{-\beta u} + \frac{r^2}{2\beta^2} \geqslant 0, \qquad u \in \mathbf{R},$$

If b is the minimum of a proposed path $\omega \sim \mathbb{Z}$ we can use (7) to find an upper bound for the mapping $t \mapsto \phi(\omega_t)$:

$$M(\omega) \equiv M(b) = \max\{\phi(b),\ r^2/2\beta^2\}.$$

It is now easy to apply the algorithm. Empirical evidence suggests that the efficiency of EA2 for this problem is sensitive to r, β and the ratio v/K. Table 1 gives a comparison of running times as parameters vary in all simulations we have taken $K = 1000$. Since we can take advantage of the Markov property of the process V to merge exact skeletons of some appropriate length T to reach the instance $l = 10$, we have approximately optimized over the choice of T in each case. We also compare the results with the Euler scheme that approximates (9) with the discrete time process:

$$Y_0 = v, \qquad Y_{(i+1)h} = Y_{ih} + r\, Y_{ih}(1 - Y_{ih}/K)\, h + \beta\, Y_{ih}\, Z_{(i+1)h}, \quad i \geqslant 0,$$

where $\{Z_{(i+1)h}\}_{i \geqslant 0}$ are independent and identically distributed with $Z_h \sim N(0, h)$. We applied the Euler scheme for increments of the type $h = 2^{-n}$, for $n = 1, 2, \ldots$. To find the increment h that generates a good approximation to the true process, we carried out four separate Kolmogorov–Smirnov tests on pairs of samples of size 10^5 from X_t, Y_t, for $t = 1/2, 1, 5, 10$ (all eight samples were independent) and found the maximum h for which at least three of the four tests did not reject the null hypothesis of equivalence in distribution of the two samples at a 10% significance level. This ensures a significance level for the composite test of 0.052.

From the evidence presented in Table 1 and other simulations we have performed, EA2 for the logistic growth model seems to be generally computationally quicker than the Euler approximation. However, when r/β is very large, EA2 becomes rather slow. For $r = 1$ and $\beta = 0.1$, for instance, the 'sufficiently accurate' Euler approximation appears to be considerably more rapid, although of course Euler still only produces an approximation to the true distribution.

To illustrate the output obtained from the algorithm, Figure 3 shows an exact skeleton of the target diffusion $\{V_t;\ 0 \leqslant t \leqslant 1000\}$, $V_0 = 1000$, for parameter values $K = 1000$, $r = 0.5$, and $\beta = 0.5$ as generated by EA2. We applied EA2 at time intervals of length $T = 0.5$ and we show the instances of the skeleton for the times $t = 0, 1, 2, \ldots$. We will use these data in Section 6.1.

Table 1. Comparative presentation of the EA2 and the Euler scheme as applied to the logistic growth model for different parameter values. In each case, simulation was carried out over the time interval [0, 10]. $\bar{D}$ denotes the mean number of Poisson process points needed to decide about a proposed path; $\bar{I}$ denotes the mean number of proposed paths until (and including) the first successful one; and T denotes the time interval used to apply the EA2. The times given represent the overall running time to obtain a sample of 100 000 sample paths. The Euler simulations are given for sufficiently small h to satisfy the composite Kolmogorov–Smirnov test described in the main text

Parameters			Exact				Euler	
υ	r	β	T	$\bar{D}$	$\bar{I}$	Time (s)	h	Time (s)
1000	0.01	0.10	5.00	0.0245	1.0011	1.2	2^{-5}	19.4
50	0.01	0.10	5.00	0.0250	1.0228	1.9	2^{-3}	7.8
1800	0.01	0.10	5.00	0.0367	1.0173	1.4	2^{-6}	35.2
1000	1.00	1.00	0.25	0.1623	1.0652	21.6	2^{-9}	256.9
1	1.00	1.00	0.25	0.1273	1.1174	22.3	2^{-8}	137.9
3500	1.00	1.00	0.25	0.2396	1.0808	25.7	2^{-10}	558.1
1000	1.00	0.10	0.10	5.0031	1.0223	284.2	2^{-6}	35.4
750	1.00	0.10	0.10	5.0009	1.0458	291.7	2^{-7}	86.4
1250	1.00	0.10	0.10	5.0000	1.0398	297.6	2^{-8}	152.9

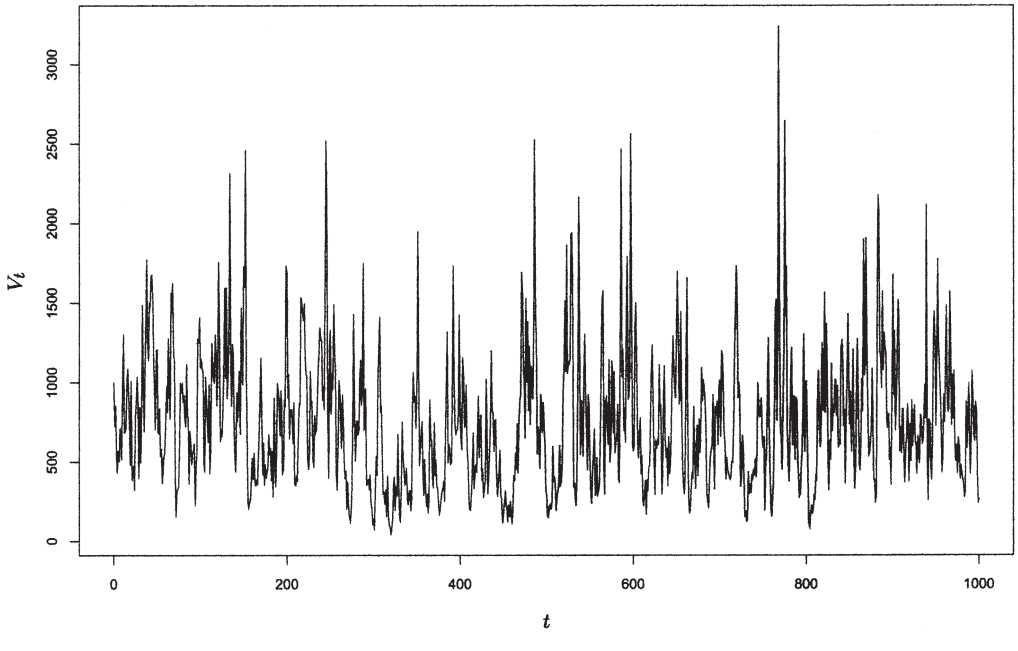

Figure 3. An exact skeleton of length 1000 of the logistic growth model (9), for $r = 0.5$, $\beta = 0.5$ and initial value $\upsilon = 1000$.

6. Applications of the exact algorithm

6.1. Monte Carlo maximum likelihood for discretely observed diffusions

An important problem in inference from diffusions involves the calculation of maximum likelihood estimators from a finite discrete set of observations of the diffusions. Assume that the drift b and the diffusion coefficient σ of the SDE in (1) have parametric forms which depend on certain unknown parameters ψ. As a running example in this section, consider the logistic growth model (9), where $\psi = (r, K, \beta)$. The goal is to find the MLE of ψ based on discrete-time observations of X, say $X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n$. It is common practice to set $t_0 = 0$ and treat the first point x_0 as fixed, although this is not essential. Let $p_\psi(t, x, y)$ be the transition density of the diffusion process, that is, the Lebesgue density of the transition measure $\mathrm{P}[X_t \in \mathrm{d}y \,|\, X_0 = x]$, which we index by ψ to denote its implicit dependence on the unknown parameters. The likelihood function is the density of the observed data $x_1, \ldots, x_n$ (conditionally on $X_{t_0} = x_0$) as a function of ψ,

$$L(\psi) = \prod_{i=1}^{n} p_\psi(t_i - t_{i-1}, x_{i-1}, x_i) = \prod_{i=1}^{n} L_i(\psi), \tag{12}$$

where we define

$$L_i(\psi) = p_\psi(t_i - t_{i-1}, x_{i-1}, x_i), \qquad i = 1, \ldots, n,$$

to be the likelihood contribution of the ith data point. Notice that the factorization in (12) is due to the Markov property. The aim is to maximize (12) as a function of ψ; however, this is a daunting task since the transition density $p_\psi(t, x, y)$ is typically not analytically available. Hence, this problem has attracted considerable attention and constitutes a very active area of research; a recent review and references can be found in Sørensen (2004).

Here we present a simple, fast and efficient Monte Carlo method for estimating $L_i(\psi)$, $i = 1, \ldots, n$, for any given value ψ. The MLE can then be found by a grid search method, although more sophisticated stochastic search algorithms can also be applied for its identification. Our approach is applicable whenever EA1 or EA2 can simulate from the diffusion process which generates the data (or at least from the process transformed to have unit diffusion coefficient). Thus, using our exact algorithm together with a variance reduction technique, we provide independent unbiased estimators $\hat{L}_1(\psi), \ldots, \hat{L}_n(\psi)$. We proceed by describing how to find $\hat{L}_1(\psi)$, since the derivation for any other $i > 1$ proceeds in the same way.

Let $\tilde{p}_\psi(t, x, y)$ be the transition density of the process (2) transformed to unit diffusion coefficient, that is, of the process obtained by the transformation $X_t \to \eta(X_t)$, where η is defined in (3). Then

$$p_\psi(t, x, y) = \tilde{p}_\psi(t, \eta(x), \eta(y)) \, |\eta'(y)|,$$

where η might also depend on ψ (as, for example, for the logistic growth model). Let S be a skeleton of the transformed process starting at $\eta(X_0) = \eta(x)$, returned by our exact algorithm on $[0, t + \gamma]$, $\gamma > 0$, for $t \equiv t_1$ the instance of the first datum. We can write

$$S = \{(u_0, \eta(X_{u_0})), \ldots, (u_l, \eta(X_{u_l}))\},$$

where $u_0 = 0$, $u_l = t + \gamma$ and $\eta(X_{u_0}) = \eta(x)$, but otherwise the time instances u_i, $0 < i < l$, will be random. We define

$$\tilde{p}_\psi(t, z, w \,|\cdot) = \mathrm{P}[\eta(X_t) \in \mathrm{d}w \,|\, \eta(X_0) = z, \cdot]/\mathrm{d}w$$

to be the density of $\eta(X_t)$ given the starting value and any other random elements. Then, by the standard property of conditional expectation, we have

$$\tilde{p}_\psi(t, \eta(x), \eta(y)) = \begin{cases} \mathrm{E}[\tilde{p}_\psi(t, \eta(x), \eta(y)|S)], & \text{when } \phi \text{ is bounded,} \\ \mathrm{E}[\tilde{p}_\psi(t, \eta(x), \eta(y)|S, m_{t+\gamma}, \theta_{t+\gamma})], & \text{when } \limsup_{u\to\infty} \phi(u) < \infty, \end{cases}$$

where $m_{t+\gamma}$ is the minimum of the diffusion path and $\theta_{t+\gamma}$ is the time when the minimum occurs. In the second case, both $m_{t+\gamma}$ and $\theta_{t+\gamma}$ are elements of S (see the discussion at the end of Section 3.2), but we write them separately to distinguish between the skeleton produced by EA1 and EA2. To avoid excessively cumbersome notation, we will drop the subscript $t + \gamma$ from both m and θ in the rest of the section.

We first consider the case of EA1, when S is simulated as in Section 2.1. This is the case when the functional ϕ is bounded. Notice that due to the Markov property,

$$\tilde{p}_\psi(t, \eta(x), \eta(y) \,|\, S) \equiv \tilde{p}_\psi(t, \eta(x), \eta(y) \,|\, \eta(X_{t_-}), \eta(X_{t_+})), \tag{13}$$

where we define

$$t_- \equiv \sup\{u_i, i = 0, \ldots, l : u_i < t\}, \qquad t_+ \equiv \inf\{u_i, i = 0, \ldots, l : u_i > t\}.$$

By construction, both t_- and t_+ are well defined, since the skeleton contains at least the instances $u_0 \equiv 0 < t$ and $u_l \equiv t + \gamma > t$. However, we have already argued (Section 2.1) that, given the skeleton, the distribution of the process at any time $s \in (0, t + \gamma)$ is given by the law of the Brownian bridge between the points of the skeleton adjacent to s. Therefore, (13) becomes

$$\frac{1}{\sqrt{2\pi \cdot \mathrm{Var}}} \exp\left\{ -\frac{1}{2 \cdot \mathrm{Var}} \left(\eta(y) - \eta(X_{t_-}) - \frac{t - t_-}{t_+ - t_-} (\eta(X_{t_+}) - \eta(X_{t_-})) \right)^2 \right\}, \tag{14}$$

for $\mathrm{Var} = (t - t_-)(t_+ - t)/(t_+ - t_-)$. Note that expression (14) does not depend on ψ.

We turn now to EA2. When S is simulated as in Section 3, where we first simulate the minimum of the diffusion paths, we have that

$$\tilde{p}_\psi(t, \eta(x), \eta(y)|S, m, \theta) \equiv \tilde{p}_\psi(t, \eta(x), \eta(y)|\eta(X_{t_-}), \eta(X_{t_+}), m, \theta). \tag{15}$$

Let $q(t, z, w)$ be the transition density of the three-dimensional Bessel process, which is known analytically and can be readily computed (see, for example, Revuz and Yor 1991: 446). Then, using Theorem 2 and working from first principles, it can be shown that (15) is

$$\frac{q\left(\frac{t-t_-}{t+\gamma-\theta}, \frac{\eta(X_{t-})-m}{\sqrt{t+\gamma-\theta}}, \frac{\eta(X_t)-m}{\sqrt{t+\gamma-\theta}}\right) \times q\left(\frac{t_+-t}{t+\gamma-\theta}, \frac{\eta(X_t)-m}{\sqrt{t+\gamma-\theta}}, \frac{\eta(X_{t+})-m}{\sqrt{t+\gamma-\theta}}\right)}{q\left(\frac{t_+-t_-}{t+\gamma-\theta}, \frac{\eta(X_{t-})-m}{\sqrt{t+\gamma-\theta}}, \frac{\eta(X_{t+})-m}{\sqrt{t+\gamma-\theta}}\right)\sqrt{t+\gamma-\theta}}, \quad \theta < t,$$

$$\frac{q\left(\frac{t_+-t}{\theta}, \frac{\eta(X_{t+})-m}{\sqrt{\theta}}, \frac{\eta(X_t)-m}{\sqrt{\theta}}\right) \times q\left(\frac{t-t_-}{\theta}, \frac{\eta(X_t)-m}{\sqrt{\theta}}, \frac{\eta(X_{t-})-m}{\sqrt{\theta}}\right)}{q\left(\frac{t_+-t_-}{\theta}, \frac{\eta(X_{t+})-m}{\sqrt{\theta}}, \frac{\eta(X_{t-})-m}{\sqrt{\theta}}\right)\sqrt{\theta}}, \quad \theta \geqslant t, \tag{16}$$

which, as in (14), does not depend on ψ.

Therefore, we can now construct a Monte Carlo algorithm for the unbiased estimation of $\hat{L}_1(\psi)$, for any given value of ψ. We describe it analytically for the more involved situation where S is simulated as in Section 3; it is clear how to modify the algorithm for the simpler case when ϕ is bounded.

Monte Carlo estimation of $L_1(\psi)$

1. *For $j = 1 : k$, repeat:*
2. *Using the exact algorithm, return a skeleton S^j on $[0, t_1 + \gamma]$ starting from $\eta(X_{t_0}) = \eta(x_0)$. Let (θ^j, m^j) be the time and value of the minimum of S^j.*
3. *Compute $\tilde{p}(t_1, \eta(x_0), \eta(x_1)|S^j, m^j, \theta^j)$ according to (16).*
4. *Go to step 1.*
5. *Output the estimate*

$$\hat{L}_1(\psi) = \frac{1}{m}\sum_{j=1}^{m} \tilde{p}(t_1, \eta(x_0), \eta(x_1)|S^j, m^j, \theta^j)\,|\eta'(x_1)|.$$

As an initial illustration of the validity of the algorithm we implemented it for the case of the Brownian motion with drift $X_t = at + \sigma B_t$ with parameters $\psi = (a, \sigma)$ when the likelihood over a set of data is explicitly known. We generated a data set $\{X_0, X_1, \ldots, X_{1000}\}$, $X_0 = 0$, after using $a = 1$ and $\sigma = 1$ and applied the algorithm that uses the Bessel process described above (although we could have used the simpler algorithm with the Brownian bridges). For this trivial example the exact algorithm will propose paths from a probability measure that coincides with the probability measure of the target process X, so they will be accepted with probability 1 and the skeletons S^j will always include precisely three points obtained at the instances $u_0 = 0$, $u_1 = \theta_{t+\gamma}$ and $u_2 = t + \gamma$. In our case, $t = 1$; also, we chose $\gamma = 0.5$. Figure 4 shows the logarithm of the actual profile likelihoods for the two parameters and their estimation as provided by our algorithm for 5000 Monte Carlo iterations and a grid (for both parameters) of increment 0.05.

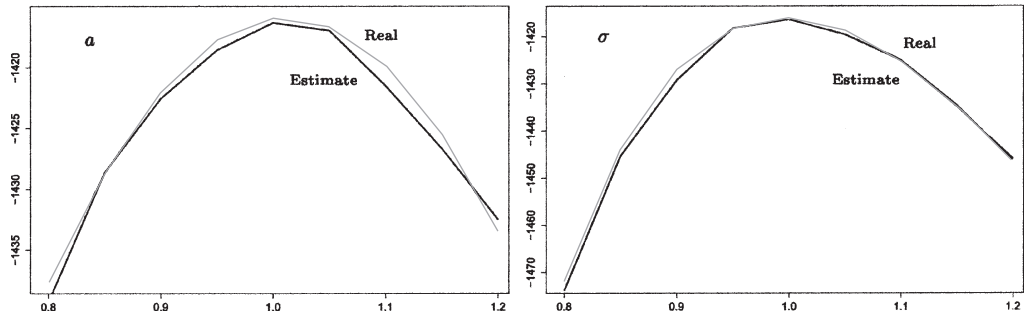

Figure 4. The actual and the estimated (based on 5000 samples) profile likelihoods for the parameters a, σ in the case of the Brownian motion with drift.

We then applied our algorithm to the logistic growth model (9) and for the data shown in Figure 3 which are generated for $K = 1000$, $r = 0.5$, $\beta = 0.5$ and initial value $v = 1000$. The data are equidistant with time increment 1 and are obtained on the time interval [0, 1000]. We considered K as known and tried to find the MLEs for r and β. Figure 5 shows the logarithm of the profile likelihoods for both parameters as estimated by our algorithm for 10 000 Monte Carlo iterations and grid increment 0.05; we used $\gamma = 0.1$.

The success of our Monte Carlo method lies in the fact that both $\tilde{p}(t, z, w|S)$ and $\tilde{p}(t, z, w|S, m, \theta)$ have known explicit forms, given in (14) and (16), respectively. Thus the only source of Monte Carlo error is on the averaging over the skeletons. The technique of obtaining closed-form expressions when conditioning on certain random variables, and then averaging over these variables using Monte Carlo, is known as *Rao–Blackwellization* (Gelfand and Smith 1990) and is a well-established variance reduction method. In this paper we have only briefly outlined the potential of this method, and it will be pursued in detail in subsequent work.

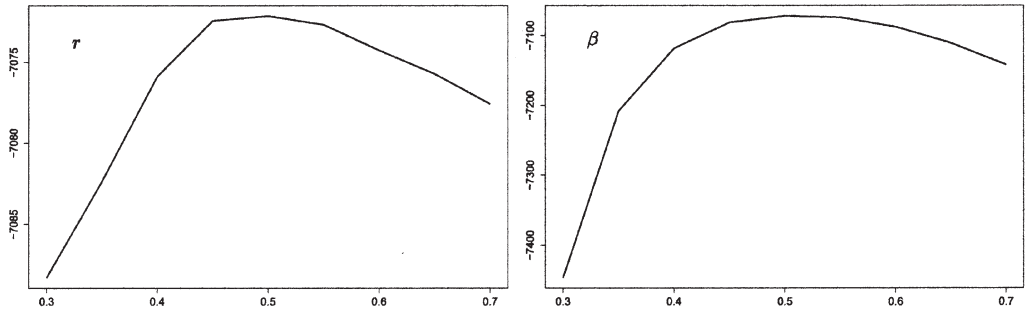

Figure 5. The estimated (based on 10 000 samples) profile likelihoods for the parameters r, β of the logistic growth model.

6.2. Exact simulation of conditioned diffusions

In this section, we address the problem of exact simulation from SDEs of type (2) conditioned on $X_T = y$ for a fixed $y \in \mathbf{R}$. This is an important problem, in particular for problems of Bayesian inference for discretely observed diffusions. Recent advances in the simulation of diffusion bridges include Roberts and Stramer (2001), Stuart *et al.* (2004) and Chib *et al.* (2004).

We shall see that our exact methods extend easily to the problem of simulating diffusion bridges. However, first we will demonstrate why conventional simulation methods are problematic for conditioned diffusions.

If Y is a diffusion with the law of (2) conditioned on $X_T = y$, then it is well known (for instance by h-transforms; see Rogers and Williams 2000: Section IV.39) that Y satisfies the SDE

$$\mathrm{d}Y_t = \mathrm{d}W_t + \left(\alpha(Y_t) + \frac{\partial \log h(T - t, Y_t)}{\partial z}\right)\mathrm{d}t, \tag{17}$$

where $h(s, z)$ is the density of the diffusion (2) starting from z at location y and time s, that is, $\mathrm{P}_z(X_s \in \mathrm{d}y)/\mathrm{d}y$. It satisfies the parabolic partial differential equation

$$\frac{\partial h(s, z)}{\partial s} = \frac{1}{2}\frac{\partial^2 h(s, z)}{\partial z^2} + \alpha(z)\frac{\partial h(s, z)}{\partial z}. \tag{18}$$

Conventional approximate methods for simulating from (17) pose great difficulties. They require a preliminary numerical calculation of h from (18) and have to deal with the fact that the drift explodes as t approaches T.

In contrast, the Radon–Nikodym derivative between the law of Y and that of a Brownian bridge from $(0, x)$ to (T, y) is proportional to

$$\exp\left\{-\int_0^T \phi(B_t)\,\mathrm{d}t\right\}$$

for ϕ as defined in (6). So, the exact algorithm can easily be adapted to conditional simulation simply by omitting the biased Brownian motion step (since B_T is now considered fixed). It will be feasible under conditions 1 and 3 given in Section 2 and the property that $\limsup_{u\to\infty} \phi(u)$ and $\limsup_{u\to-\infty} \phi(u)$ are not both $+\infty$.

7. Conclusions

This paper has introduced an algorithm for perfect simulation of diffusion processes, relaxing some of the regularity conditions imposed by the previous method of Beskos and Roberts (2005). The algorithm seems to be computationally efficient in many cases, as demonstrated by our simulation study for the logistic growth model. Currently, we know little about why the method's efficiency relative to a standard Euler-scheme alternative varies so considerably, and a systematic simulation study and theoretical analysis will be required to investigate this further.

We outline two potential application areas, both of which warrant further investigation. The Monte Carlo maximum likelihood approach capitalizes on the form of the output given by EA1 or EA2 to provide unbiased likelihood estimates. The computational efficiency of this and related techniques will rely on being able to provide likelihood estimates which are smooth as a function of the parameter, and with variances which are robust to data sample sizes.

Simulation of conditioned diffusions is a straightforward by-product of EA1 or EA2. However, there are a number of potentially important examples of it. Our primary motivation comes from the need to carry out imputation of unobserved sample paths as part of data-augmentation schemes for Bayesian inference for diffusions.

This paper has focused on the simulation of one-dimensional diffusions. Progress is certainly possible for some multidimensional diffusions at least. For instance, for diffusions which can be reduced by tranformation to unit diffusion coefficient (using a multivariate analogue of (3)), the integration-by-parts argument leading to (4) generalizes trivially so long as the diffusion drift can be written as the gradient of a suitable potential function. Details of this generalization will appear in subsequent work.

The major unsolved problem in this area involves the simulation of diffusions whose drift is not bounded in either tail. This is considerably more challenging than the case considered here, mainly due to the difficulties with obtaining simple closed-form solutions to two-sided boundary crossing problems for Brownian motion.

Appendix

Proof of Proposition 1. Recall that B_t, $t \in [0, T]$, is the coordinate mapping on $(C, \mathcal{C})$, that is, $B_t(\omega) = \omega(t)$ for any $\omega \in C$. The property $(M \mid M_T = \rho) \stackrel{d}{=} (N \mid N_T = \rho)$, for all $\rho \in \mathbf{R}$, can be expressed in a rigorous way as

$$\mathbb{M}[A \mid \sigma(B_T)] = \mathbb{N}[A \mid \sigma(B_T)], \qquad \mathbb{N}\text{-almost surely},$$

for any $A \in \mathcal{C}$. It suffices to prove that:

$$\mathbb{M}[A] = \mathrm{E}_{\mathbb{N}}\left[\mathbb{I}_A \cdot \frac{f_M}{f_N}(B_T)\right], \qquad \text{for any } A \in \mathcal{C},$$

where the index on the expectation shows the probability measure with respect to which this expectation is considered. Simple calculations give:

$$\mathrm{E}_{\mathbb{N}}\left[\mathbb{I}_A \cdot \frac{f_M}{f_N}(B_T)\right] = \mathrm{E}_{\mathbb{N}}\left[\mathrm{E}_{\mathbb{N}}\left[\mathbb{I}_A \cdot \frac{f_M}{f_N}(B_T) \mid \sigma(B_T)\right]\right] = \mathrm{E}_{\mathbb{N}}\left[\frac{f_M}{f_N}(B_T) \cdot \mathbb{N}[A \mid \sigma(B_T)]\right]$$

$$= \mathrm{E}_{\mathbb{M}}[\mathbb{N}[A \mid \sigma(B_T)]] = \mathrm{E}_{\mathbb{M}}[\mathbb{M}[A \mid \sigma(B_T)]] = \mathbb{M}[A].$$

□

Proof of Proposition 2. From Karatzas and Shreve (1991: 95) we obtain

$$\mathrm{P}[m_T \in \mathrm{d}b \mid W_T = a] = \frac{2}{T}(a-2b)\exp\left\{-\frac{(a-2b)^2}{2T} + \frac{a^2}{2T}\right\}\mathrm{d}b, \qquad b \leqslant \min\{a, 0\}.$$

We can easily derive the representation

$$[m_T \mid W_T = a] \stackrel{d}{=} (a - \sqrt{2\,TE(1) + a^2})/2 \equiv Z_1.$$

It is clear from (8) that

$$\mathrm{P}[\theta_T \in \mathrm{d}t \mid m_T = b,\, W_T = a] \propto \frac{1}{\sqrt{t^3(T-t)^3}} \exp\left\{-\frac{b^2}{2t} - \frac{(a-b)^2}{2(T-t)}\right\}\mathrm{d}t, \qquad 0 \leqslant t \leqslant T.$$

If $V := (T-\theta_T)/\theta_T$ and f denotes the density of V conditionally on $W_T = a$, $m_T = b$ then

$$f(y) \propto y^{-3/2}\exp\left(-\frac{c_1}{y} - c_2 y\right) + y^{-1/2}\exp\left(-\frac{c_1}{y} - c_2 y\right), \qquad y \geqslant 0, \tag{19}$$

for $c_1 = (a-b)^2/2T$, $c_2 = b^2/2T$. This density can be identified as the mixture of an $\mathrm{IGau}(\sqrt{c_1/c_2}, 2c_1)$ and $1/\mathrm{IGau}(\sqrt{c_2/c_1}, 2c_2)$. The ratio of the integral of the left summand in (19) to the integral of f is

$$p = \frac{1}{1+\sqrt{c_1/c_2}},$$

so with probability p we draw $I_1 \sim \mathrm{IGau}(\sqrt{c_1/c_2}, 2c_1)$, and with probability $1-p$ we draw $I_2 \sim 1/\mathrm{IGau}(\sqrt{c_2/c_1}, 2c_2)$. If we use $U \sim \mathrm{Unif}(0, 1)$ to decide on the choice between the densities of the mixture we can obtain that for the Z_2 defined in Proposition 2, $[Z_2 \mid Z_1 = b] \stackrel{d}{=} [\theta_T \mid m_T = b,\, W_T = a]$. □

Proof of Proposition 3. When $\phi \leqslant M$, it is clear from (5) that $\epsilon \geqslant \exp(-M \cdot T)$. Trivially, Theorem 1 gives that $\mathrm{E}[D] = M \times T$.

When $\limsup_{u\to\infty} \phi(u) < \infty$ and $M(b) = \sup\{\phi(u);\, u \geqslant b\}$, $b \in \mathbf{R}$, then we can bound ϵ using Jensen's inequality:

$$\epsilon = \mathrm{E}\left[\exp\left\{-\int_0^T \phi(\omega_t)\,\mathrm{d}t\right\}\right] \geqslant \mathrm{E}[\exp\{-M(\hat{\boldsymbol{m}}) \cdot T\}] \geqslant \exp\{-\mathrm{E}[M(\hat{\boldsymbol{m}})] \cdot T\}.$$

Also, $\mathrm{E}[D] = \mathrm{E}[\mathrm{E}[D \mid \hat{\boldsymbol{m}}]] = \mathrm{E}[M(\hat{\boldsymbol{m}})] \times T$. For both cases, to estimate $N(T)$, assume that $\{\omega_1, \omega_2, \ldots\}$ is a sequence of proposed paths, D_i is the number of Poisson process points needed to decide on the acceptance or rejection of the ith proposed path for any $i \geqslant 1$, and $I = \inf\{i \geqslant 1 : \omega_i \text{ is accepted}\}$. Assume also that $\mathrm{E}[D \mid A]$, $\mathrm{E}[D \mid A^c]$ are the expected number of Poisson process points conditionally on accepting and rejecting the proposed path, respectively. Then, if $\mathrm{E}[D] < \infty$,

$$\begin{aligned}\mathrm{E}[N(T)] &= \mathrm{E}[D_1 + D_2 + \ldots + D_I] = \mathrm{E}[\mathrm{E}[D_1 + D_2 + \cdots + D_I \mid I]] \\ &= \mathrm{E}[(I-1)\cdot \mathrm{E}[D \mid A^c] + \mathrm{E}[D \mid A]] = \left(\frac{1}{\epsilon} - 1\right)\mathrm{E}[D \mid A^c] + \mathrm{E}[D \mid A] \\ &= \frac{\mathrm{E}[D]}{\epsilon}.\end{aligned}$$

If $\mathrm{E}[D] = \infty$ then $\mathrm{E}[N(T)] = \infty$. □

Proof of Proposition 4. $\mathrm{E}_x[M(\hat{m})]$ is decreasing in x; the minimum of the $\mathbb{Z}$-path is stochastically increasing in x and $M(b) = \sup\{\phi(u);\ u \geqslant b\}$ is decreasing. To simplify the proof, we can choose $x < b_0$; for this x, $M(\hat{m}) \leqslant \exp(-k\hat{m})$ for any $\hat{m}$ in its domain $(-\infty, x)$. Consider a real δ such that $\alpha(u) \geqslant \delta$, $u \in \mathbf{R}$. It can be easily shown that:

$$\text{biased Brownian motion } \hat{W} \overset{\text{stoch.}}{\geqslant} \text{ Brownian motion (BM) with drift } \delta, \tag{20}$$

where both processes are considered on $[0, T]$ and for the same starting point x. This follows after representing the Brownian motion with drift δ as a biased Brownian motion with ending point distributed according to $f(u) \propto \exp\{\delta u - (u-x)^2/2T\}$. Recall that the ending point of $\hat{W}$ is distributed according to $h(u) \propto \exp\{A(u) - (u-x)^2/2T\}$. Also

$$\left\{\log\left(\frac{h}{f}\right)\right\}'(u) = \alpha(u) - \delta \geqslant 0,$$

so h/f is increasing. That indicates that the distribution with density h is stochastically larger than that with density f, which implies (20).

We denote by m_δ the minimum of the Brownian moiton with drift δ. Property (20) yields

$$\mathrm{E}[M(\hat{m})] \leqslant \mathrm{E}[M(m_\delta)] \leqslant \mathrm{E}[\exp(-k \cdot m_\delta)]. \tag{21}$$

Using Girsanov's theorem we can derive the density of m_δ (Borodin and Salminen 2002):

$$\mathrm{P}[m_\delta \in \mathrm{d}b]/\mathrm{d}b = 2\mathcal{N}_{x+\delta T, T}(b) + 2\delta \cdot \exp\{2\delta(b-x)\} \cdot \Phi\left(\frac{b - x + \delta T}{\sqrt{T}}\right), \qquad b \in (-\infty, x),$$

where $\mathcal{N}_{\mu,\sigma^2}(b)$ is the density of a normal distribution with mean μ and variance σ^2 evaluated at b and $\Phi(u) = \int_{-\infty}^{u} \mathcal{N}_{0,1}(y)\,\mathrm{d}y$, $u \in \mathbf{R}$. It is easy now to check that this density has finite exponential expectations, $\mathrm{E}[\exp(-k \cdot m_\delta)] < \infty$. □

Acknowledgements

The authors would like to thank Bruno Casella, Paul Fearnhead and the referees for stimulating comments and discussions on the subject of this paper.

References

Asmussen, S., Glynn, P. and Pitman, J. (1995) Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, **5**(4), 875–896.

Bertoin, J. and Pitman, J. (1994) Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.*, **118**(2), 147–166.

Beskos, A.P. and Roberts, G.O. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, **15**, 2422–2444.

Borodin, A.N. and Salminen, P. (2002) *Handbook of Brownian motion – Facts and Formulae*, 2nd edn. Basel: Birkhäuser.

Chib, S., Pitt, M.K. and Shephard, N. (2004) Likelihood-based inference for diffusion driven models. Unpublished manuscript. http://www.nuff.ox.ac.uk/economics/papers/2004/w20/chibpittshephard.pdf.

Devroye, L. (1986) *Nonuniform Random Variate Generation.* New York: Springer-Verlag.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.

Goel, N.S. and Richter-Dyn, N. (1974) *Stochastic Models in Biology.* New York: Academic Press.

Karatzas, I. and Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd edn. New York: Springer-Verlag.

Karlin, S. and Taylor, H.M. (1981) *A Second Course in Stochastic Processes*. New York: Academic Press.

Kloeden, P.E. and Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations.* Berlin: Springer-Verlag.

Øksendal, B.K. (1998) *Stochastic Differential Equations: An Introduction with Applications.* Berlin: Springer-Verlag.

Papaspiliopoulos, O. and Roberts, G.O. (2004) Retrospective MCMC methods for Dirichlet process hierarchical models. Manuscript. Department of Mathematics and Statistics, Lancaster University. Submitted for publication. http://www.maths.lancs.ac.uk/~papaspil/research.html. In revision.

Propp, J. and Wilson, D. (1998) Coupling from the past: a user's guide. In *Microsurveys in Discrete Probability (Princeton, NJ 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pp. 181–192. Providence, RI: American Mathematical Society.

Revuz, D. and Yor, M. (1991) *Continuous Martingales and Brownian Motion.* Berlin: Springer-Verlag.

Roberts, G.O. and Stramer, O. (2001) On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika*, **88**(3), 603–621.

Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov Processes, and Martingales. Vol. 2: Itô Calculus.* Cambridge: Cambridge University Press.

Sørensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. *Internat. Statist. Rev.*, **72**, 337–354.

Stuart, A.M., Voss, J. and Wiberg, P. (2004) Conditional path sampling of SDEs and the Langevin MCMC method. *Comm. Math. Sci.*, **2**, 685–697.

Received August 2005 and revised April 2006