

Simple simulation of diffusion bridges with application to likelihood inference for diffusions

MOGENS BLADT¹ and MICHAEL SØRENSEN²

¹*Instituto de Investigacion en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, A.P. 20-726, 01000 Mexico, D.F., Mexico. E-mail: bladt@sigma.iimas.unam.mx*

²*Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark. E-mail: michael@math.ku.dk; url: www.math.ku.dk/~michael*

With a view to statistical inference for discretely observed diffusion models, we propose simple methods of simulating diffusion bridges, approximately and exactly. Diffusion bridge simulation plays a fundamental role in likelihood and Bayesian inference for diffusion processes. First a simple method of simulating approximate diffusion bridges is proposed and studied. Then these approximate bridges are used as proposal for an easily implemented Metropolis–Hastings algorithm that produces exact diffusion bridges. The new method utilizes time-reversibility properties of one-dimensional diffusions and is applicable to all one-dimensional diffusion processes with finite speed-measure. One advantage of the new approach is that simple simulation methods like the Milstein scheme can be applied to bridge simulation. Another advantage over previous bridge simulation methods is that the proposed method works well for diffusion bridges in long intervals because the computational complexity of the method is linear in the length of the interval. For ρ -mixing diffusions the approximate method is shown to be particularly accurate for long time intervals. In a simulation study, we investigate the accuracy and efficiency of the approximate method and compare it to exact simulation methods. In the study, our method provides a very good approximation to the distribution of a diffusion bridge for bridges that are likely to occur in applications to statistical inference. To illustrate the usefulness of the new method, we present an EM-algorithm for a discretely observed diffusion process.

Keywords: Bayesian inference; diffusion bridge; discretely sampled diffusions; EM-algorithm; likelihood inference; Milstein scheme; pseudo-marginal MCMC; time-reversion

1. Introduction

In this paper, we propose a simple general method for the simulation of a one-dimensional diffusion bridge. Our main motivation is that simulation of diffusion bridges plays a fundamental role in simulation-based likelihood inference (including Bayesian inference) for discretely sampled diffusion processes and other diffusion-type processes like stochastic volatility models.

Our approach is based on the following simple way of constructing a process that at time zero starts from a and at time T ends in b , where a and b are given numbers. One diffusion process, $X_t^{(1)}$, is started from the point a , while an independent diffusion, $X_t^{(2)}$, with the same dynamics, is started from the point b . The time of the second diffusion is reversed, so that the time starts at T and goes downwards to zero. Suppose the process $X_t^{(1)}$ hits the sample path of

the time reversed diffusion $X_{T-t}^{(2)}$, and let τ denote the first time the two paths intersect. Then the process that for $t \leq \tau$ is equal to $X_t^{(1)}$, and for $t > \tau$ equals $X_{T-t}^{(2)}$, is obviously a process in the time interval $[0, T]$ that starts at a and ends at b . Conditional on the event that the two processes intersect, we show that the process constructed in this way is indeed an approximation to a realization of a diffusion bridge between the two points. A simple rejection sampler is thus obtained by repeatedly simulating the two diffusions until they hit each other. The diffusions can be simulated by means of simple procedures like the Milstein scheme, see Kloeden and Platen [40], so the new method is easy to implement for likelihood inference for discretely sampled diffusion processes. This approximate diffusion bridge is used as proposal for a Metropolis–Hastings algorithm that has an exact diffusion bridge as its target distribution. The algorithm uses the pseudo-marginal approach of Andrieu and Roberts [3] and is easy to implement: to calculate the rejection probability, a number of independent diffusions are simulated and it is determined whether or not they intersect the proposed bridge trajectory.

Diffusion bridge simulation is a highly non-trivial problem that has been investigated actively over the last 10–15 years. A lucid exposition of the problems and the state-of-the-art can be found in Papaspiliopoulos and Roberts [44]. It was previously thought impossible to simulate diffusion bridges by means of simple procedures, because a rejection sampler that tries to hit the prescribed end-point for the bridge (or a small neighbourhood around it) will have an excessively high rejection probability. The rejection sampler presented in this paper has a quite acceptable rejection probability because what must be hit is a sample path rather than a point. The first diffusion bridge simulation methods in the literature were based on the Metropolis–Hastings algorithm with a proposal distribution given by a process that is forced to go from a to b , see, for example, Roberts and Stramer [47] or Durham and Gallant [20]. Later Beskos, Papaspiliopoulos and Roberts [6,7] developed algorithms for exact simulation of diffusion bridges. These are rejection sampling algorithms that use in a clever way a measure change and simulation of a Brownian bridge, which can easily be simulated. Under strong boundedness conditions the algorithm is relatively simple, whereas it is more complex under weaker condition. Lin, Chen and Mykland [42] proposed a sequential Monte Carlo method for simulating diffusion bridges with a resampling scheme guided by the empirical distribution of backward paths. The spirit of this approach has similarities to the methods proposed here.

An advantage of the method proposed in the present paper is that the same simple algorithm can be used for all one-dimensional diffusions with a finite speed-measure, and that it is easy to understand and to implement. It is also worth noting that the method does not require that the diffusion is transformed into one with unit diffusion coefficient. Though such a transformation exists under very general conditions for one-dimensional diffusions, the transformation is not in closed form for many interesting examples. Such a transformation is, for instance, required for the exact algorithm of Beskos, Papaspiliopoulos and Roberts [6,7]. Another, and perhaps more important, advantage is that it works particularly well for long time intervals. For ergodic diffusions, the computational complexity of our method is shown to be linear in the distance between the two end-points of the diffusion bridge. This is illustrated in a simulation study where the computer time increases linearly with the interval length, while it seems to grow at least exponentially with the interval length for the exact EA algorithms of Beskos, Papaspiliopoulos and Roberts [6]. Thus, the EA algorithm is likely not to work for long time intervals. Importantly,

it is shown that for ergodic diffusions the approximate method proposed here simulates an essentially exact diffusion bridge in long time intervals (apart from the discretization error). For exponentially mixing diffusions, which covers most diffusions used in practice, the distribution of the simulated process goes to that of a diffusion bridge exponentially fast as a function of the interval length. Thus the proposed method provides a useful supplement to previously published methods because it works particularly well for long time intervals, where the other methods tend not to work. It is worth noting that simulation-based likelihood inference for discretely sampled diffusions is mainly important for long time intervals, because for short time intervals several simpler methods provide highly efficient estimators, see the following discussion.

The main challenge to likelihood based inference for diffusion models is that the transition density, and hence the likelihood function, is not explicitly available and must therefore be approximated. When the sampling frequency is relatively high, which is often the case for financial data, rather crude approximations to the likelihood functions, like those in Ozaki [43], Bollerslev and Wooldridge [15], Bibby and Sørensen [12] and Kessler [39], give estimators with a high efficiency. This follows from results based on high frequency asymptotics in Sørensen [48]. When the interval between the observation times is relatively long, more accurate approximations to the transition density are needed. One approach is numerical approximations, either by solving the Kolmogorov PDE numerically, for example, Poulsen [46] and Hurn, Jeisman and Lindsay [33], or by expansions, for example, Ait-Sahalia [1,2] and Forman and Sørensen [25]. Alternatively, likelihood inference can be based on simulations, an approach that goes back to the seminal paper by Pedersen [45], whose method is, however, computationally costly because he did not use bridge simulation. The inference problem can be viewed as a missing data problem. If the diffusion process had been observed continuously, the likelihood functions would be explicitly given by the Girsanov formula, but the diffusion has been observed at discrete time points only, so the continuous-time paths between the observation points can be considered as missing data. This way of viewing the problem, which goes back to Dacunha-Castelle and Florens-Zmirou [17], makes it natural to apply either the EM-algorithm or the Gibbs sampler. To do so, it is necessary to simulate the missing continuous paths between the observations conditional on the observations, which is exactly simulation of diffusion bridges. It was a significant break-through when this was simultaneously realized by several authors, see Roberts and Stramer [47], Elerian, Chib and Shephard [21], Eraker [22] and Durham and Gallant [20], and approaches based on bridge simulation has since been used by several authors including Golightly and Wilkinson [27,28,30], Beskos, Papaspiliopoulos and Roberts [9], Delyon and Hu [18], Beskos, Papaspiliopoulos and Roberts [8] and Lin *et al.* [42]. To illustrate how our bridge simulation method can be used for likelihood inference, we modify an EM-algorithm in Beskos, Papaspiliopoulos, Roberts and Fearnhead [9] by using our simple simulation method.

Diffusion bridge simulation is also crucial to simulation-based inference for other types of diffusion process data than discrete time observations. Chib, Pitt and Shephard [16] presented a general approach to simulation-based Bayesian inference for diffusion models when the data are discrete time observations of rather general, and possibly random, functionals of the continuous sample path, see also Golightly and Wilkinson [29]. This approach covers for instance diffusions observed discretely with measurement error and discretely sampled stochastic volatility models. In this approach too, the underlying continuous time diffusion process must be simulated conditionally on the observations, which is done by a Metropolis–Hastings algorithm. The algorithm

mixes badly if the entire sample path is updated simultaneously, so the interval is divided into random subintervals that are updated sequentially. The sample path in a subinterval must be simulated conditionally of the values of the diffusion in the other intervals, which by the Markov property is diffusion bridge simulation given the values at the end-points. This method can be modified by using the bridge simulation method proposed here. Baltazar-Larios and Sørensen [4] presented an EM-algorithm for integrated diffusions observed discretely with measurement error based on the ideas in Chib *et al.* [16], but using the bridge simulation method of the present paper, and showed that the method worked well in simulation studies. We shall briefly review the results of this paper.

The paper is organized as follows. In Section 2, we first present the new approximate bridge simulation method and show in what sense it approximates a diffusion bridge. Then results are given about exactness and small rejection probabilities for long time intervals. Finally, the approximate bridges are used as proposal in a Metropolis–Hastings algorithm that has an exact diffusion bridge as its target distribution. In Section 3, the approximate bridge simulation method is compared to exact simulation methods in two examples, the Ornstein–Uhlenbeck process and the hyperbolic diffusion. The study indicates that our method provides a very accurate approximation to the distribution of a diffusion bridge, except for bridges that are very unlikely to occur when using the method for likelihood inference. An EM-algorithm for discretely observed diffusions based on the proposed bridge simulation method is briefly presented in Section 4. It is demonstrated how the algorithm simplifies for an exponential family of diffusions (i.e., when drift is linear in the parameters). In this case, Bayesian inference is considered too. Finally, an application of the proposed method to estimation for discretely observed integrated diffusions with measurement errors in Baltazar-Larios and Sørensen [4] is briefly reviewed. Section 5 concludes.

2. Diffusion bridge simulation

2.1. Approximate bridge simulation

Let $X = \{X_t\}_{t \geq 0}$ be a one-dimensional diffusion given by the stochastic differential equation

$$dX_t = \alpha(X_t) dt + \sigma(X_t) dW_t, \quad (2.1)$$

where W is a Wiener process, and where the coefficients α and σ are sufficiently regular to ensure that the equation has a unique weak solution that is a strong Markov process. Let a and b be given points in the state space of X . We present a method for simulating an approximation to a sample path of X such that $X_0 = a$ and $X_\Delta = b$. A solution of (2.1) in the interval $[t_1, t_2]$ such that $X_{t_1} = a$ and $X_{t_2} = b$ will be called a (t_1, a, t_2, b) -bridge. When $t_1 = 0$ and $t_2 = 1$, we sometimes simply call it an (a, b) -bridge. We will denote the transition density of X by $p_t(x, y)$. Specifically, the conditional density of X_{s+t} given $X_s = x$ is $y \mapsto p_t(x, y)$. The state space of X is denoted by (ℓ, r) where $\infty \leq \ell < r \leq \infty$.

Let W^1 and W^2 be two independent standard Wiener processes, and define X^1 and X^2 as the solutions to

$$dX_t^i = \alpha(X_t^i) dt + \sigma(X_t^i) dW_t^i, \quad i = 1, 2, X_0^1 = a \text{ and } X_0^2 = b.$$

The main idea of the paper is to realize an approximation to a $(0, a, \Delta, b)$ -bridge by simulating the process X^1 from a forward in time, and X^2 from b backward in time starting at time Δ . If the samples paths of the two processes intersect, they can be combined into a realization of a process that approximates a $(0, a, \Delta, b)$ -bridge.

Thus to simulate an approximate diffusion bridge in the interval $[0, \Delta]$, we can use any of the several methods available to simulate the diffusions X^1 and X^2 , see, for example, Kleoden and Platen [40]. Let $Y_{\delta i}^1, i = 0, 1, \dots, N$ and $Y_{\delta i}^2, i = 0, 1, \dots, N$ be (independent) simulations of X^1 and X^2 in $[0, \Delta]$ with step size $\delta = \Delta/N$. Then a simulation of an approximation to a $(0, a, \Delta, b)$ -bridge is obtained by the following rejection sampling scheme. Keep simulating Y^1 and Y^2 until the sample paths cross, that is, until there is an i such that either $Y_{\delta i}^1 \geq Y_{\delta(N-i)}^2$ and $Y_{\delta(i+1)}^1 \leq Y_{\delta(N-(i+1))}^2$ or $Y_{\delta i}^1 \leq Y_{\delta(N-i)}^2$ and $Y_{\delta(i+1)}^1 \geq Y_{\delta(N-(i+1))}^2$. Once a trajectory crossing has been obtained, define

$$B_{\delta i} = \begin{cases} Y_{\delta i}^1 & \text{for } i = 0, 1, \dots, \nu - 1, \\ Y_{\delta(N-i)}^2 & \text{for } i = \nu, \dots, N, \end{cases} \tag{2.2}$$

where $\nu = \min\{i \in \{1, \dots, N\} | Y_{\delta i}^1 \leq Y_{\delta(N-i)}^2\}$ if $Y_0^1 \geq Y_{\Delta}^2$, and $\nu = \min\{i \in \{1, \dots, N\} | Y_{\delta i}^1 \geq Y_{\delta(N-i)}^2\}$ if $Y_0^1 \leq Y_{\Delta}^2$. Then B approximates a $(0, a, \Delta, b)$ -bridge under the condition of Theorem 2.1. On top of the usual influence of the step size δ on the quality of the individual simulated trajectories, the step size also controls the probability that a trajectory crossing is not detected. Therefore, it is advisable to choose δ smaller than usual.

The rejection probability (the probability of no trajectory crossing) depends on the drift and diffusion coefficients, on the values of a and b , and on the length of the interval Δ . It is shown below that for ergodic diffusions the rejection probability is small when Δ is large (Theorem 2.4). Simulation studies in Section 3 indicate that the number of rejections is small when a and b are not very far apart. When the simulation algorithm is used to make likelihood inference for discretely observed diffusion processes (Section 4), this is the typical situation for relatively frequent sampling times, and as just noted, there are in general few rejections when an ergodic diffusion has been sampled at a low frequency.

The distribution of the process that is simulated by the algorithm above and the sense in which it is an approximation of a diffusion bridge is seen from the following theorem, where

$$m(x) = \frac{1}{\sigma^2(x)} \exp\left(2 \int_z^x \frac{\alpha(y)}{\sigma^2(y)} dy\right), \quad x \in (\ell, r) \tag{2.3}$$

is the density of the speed measure of the diffusion. Here z is an arbitrary point in the state space (ℓ, r) .

Theorem 2.1. Let $\tau = \inf\{0 \leq t \leq \Delta | X_t^1 = X_{\Delta-t}^2\}$ ($\inf \emptyset = +\infty$) and define

$$Z_t = \begin{cases} X_t^1 & \text{if } 0 \leq t \leq \tau, \\ X_{\Delta-t}^2 & \text{if } \tau < t \leq \Delta. \end{cases}$$

Assume that

$$M = \int_{\ell}^r m(x) dx < \infty. \tag{2.4}$$

Then the distribution of $\{Z_t\}_{0 \leq t \leq \Delta}$, conditional on the event $\{\tau \leq \Delta\}$, equals the distribution of a $(0, a, \Delta, b)$ -bridge, conditional on the event that the bridge is hit by an independent diffusion with stochastic differential equation (2.1) and initial distribution with density $p_{\Delta}(b, \cdot)$.

Before proving Theorem 2.1, we prove a lemma on the distribution of a time-reversed diffusion. Quite generally, the density (2.3) of the speed measure for any one-dimensional diffusion satisfies the balance equation

$$p_t(x, y)m(x) = p_t(y, x)m(y), \tag{2.5}$$

see Ito and McKean [34], page 149. Under the condition (2.4) that the speed measure is finite, an invariant probability measure exists and has the density function

$$v(x) = m(x)/M. \tag{2.6}$$

Lemma 2.2. Define the time-reversed process $\{\bar{X}_t\}$ by $\bar{X}_t = X_{\Delta-t}^2$. The process $\{\bar{X}_t\}$ and the conditional process $\{X_t\}$ given that $X_{\Delta} = b$ have the same transition densities

$$q(x, s, y, t) = \frac{p_{t-s}(x, y)p_{\Delta-t}(y, b)}{p_{\Delta-s}(x, b)} = \frac{p_{t-s}(y, x)p_{\Delta-t}(b, y)}{p_{\Delta-s}(b, x)}, \quad s < t < \Delta. \tag{2.7}$$

Assume that (2.4) holds. Then the distribution of $\{\bar{X}_t\}$ is equal to the conditional distribution of the process $\{X_t\}$ with $X_0 \sim v$ given that $X_{\Delta} = b$.

Proof. The second identity in (2.7) follows from (2.5). The first expression for q is the well-known expression for the transition density of a diffusion bridge ending in b at time 1, see Fitzsimmons, Pitman and Yor [23], page 111. It can be easily established by direct calculation. The second expression for q can similarly be obtained as the transition density of \bar{X} by direct calculation. The conditional density of \bar{X}_t given \bar{X}_s ($s < t$) is

$$p_{\bar{X}_s, \bar{X}_t}(x, y)/p_{\bar{X}_s}(x) = p_{X_{\Delta-t}^2, X_{\Delta-s}^2}(y, x)/p_{X_{\Delta-s}^2}(x) = p_{\Delta-t}(b, y)p_{t-s}(y, x)/p_{\Delta-s}(b, x).$$

Now suppose that (2.4) holds, and assume that $X_0 \sim v$. Then $X_{\Delta} \sim v$, and the joint density of (X_0, X_{Δ}) is $v(y_0)p_{\Delta}(y_0, x) = v(x)p_{\Delta}(x, y_0)$, again by (2.5). Hence, the conditional density of X_0 given $X_{\Delta} = b$ is $p_{\Delta}(b, y_0)$. Obviously, the density of $\bar{X}_0 = X_{\Delta}^2$ is $p_{\Delta}(b, y_0)$, so the process $\{\bar{X}_t\}$ and the conditional process $\{X_t\}$ given that $X_{\Delta} = b$ have the same transition densities and the same initial distribution. Therefore, they have the same distribution. \square

Remark. Note that the results of Lemma 2.2 hold for a multivariate diffusion too, provided that a function v exists such that $p_t(x, y)v(x) = p_t(y, x)v(y)$. Diffusions with this property are called v -symmetric, see the discussion in Kent [38]. The second assertion of the lemma holds provided that v is an integrable function on the state space.

Proof of Theorem 2.1. Let W^3 be a standard Wiener processes independent of W^1 , and let X^3 be the solution of

$$dX_t^3 = \alpha(X_t^3) dt + \sigma(X_t^3) dW_t^3,$$

where the distribution of X_0^3 has the density ν given by (2.6). Finally, let ρ be the first time the diffusion X^3 hits the sample path of X^1 . Define a process by

$$Y_t = \begin{cases} X_t^1 & \text{if } 0 \leq t \leq \rho, \\ X_t^3 & \text{if } \rho < t \leq \Delta \end{cases}$$

on $\{\rho \leq \Delta\}$, and $Y = X^1$ on $\{\rho = \infty\}$. By the strong Markov property Y has the same distribution as X^1 . From now on, we condition on $X_\Delta^3 = b$. Since

$$P(Y \in \cdot | X_\Delta^3 = b, \rho \leq \Delta) = P(Y \in \cdot | Y_\Delta = b, \rho \leq \Delta),$$

the theorem follows because by Lemma 2.2 the distribution of $\{X_{\Delta-t}^2\}_{0 \leq t \leq \Delta}$ equals that of $\{X_t^3\}_{0 \leq t \leq \Delta}$ conditional on $X_\Delta^3 = b$, so that $P(Y \in \cdot | X_\Delta^3 = b, \rho \leq \Delta) = P(Z \in \cdot | \tau \leq \Delta)$. The event $\{Y_\Delta = b, \rho \leq \Delta\}$ is the event that Y is a $(0, a, \Delta, b)$ -diffusion bridge and that the diffusion bridge is hit by X^3 , which under the condition $X_\Delta^3 = b$ has the initial distribution $p_\Delta(b, \cdot)$ (see the proof of Lemma 2.2). \square

By symmetry, we see that the distribution of the process \tilde{Z} defined by

$$\tilde{Z}_t = \begin{cases} X_t^1 & \text{if } 0 \leq t \leq \Delta - \tilde{\tau}, \\ X_{\Delta-t}^2 & \text{if } \Delta - \tilde{\tau} < t \leq \Delta \end{cases}$$

where $\tilde{\tau} = \inf\{0 \leq t \leq \Delta | X_{\Delta-t}^1 = X_t^2\}$, is that of a $(0, a, \Delta, b)$ -bridge conditional on the event that the bridge is hit by $\{X_{\Delta-t}^3\}$, where X^3 is an independent diffusion with stochastic differential equation (2.1) and initial distribution with density $p_\Delta(a, \cdot)$. Here we use X^1 until the last time it crosses the trajectory of $\{X_{\Delta-t}^2\}$, which happens at time $\Delta - \tilde{\tau}$. Obviously, an approximate diffusion bridge can also be simulated by using \tilde{Z} .

We can consider the diffusions, diffusion bridges and the approximate diffusion bridge Z as elements of the canonical space, C_Δ , of continuous functions defined on the time interval $[0, \Delta]$. Each of these processes induce a probability measure on the usual sigma-algebra generated by the cylinder sets. Let f_b denote the Radon–Nikodym derivative of the distribution of the $(0, a, \Delta, b)$ -diffusion bridge with respect to a dominating measure. The diffusion bridge solves a stochastic differential equation with the same diffusion coefficient as in (2.1), see, for example, (4.4) in Papaspiliopoulos and Roberts [44], so the density f_b is given by Girsanov’s theorem. Since the drift for the bridge is unbounded at the end point, one has to choose the dominating measure carefully: it must correspond to another bridge, see Papaspiliopoulos and Roberts [44], page 322, and Delyon and Hu [18]. Similarly let f_a and f_d denote the densities of the distributions of the approximate bridge Z and of a diffusion with stochastic differential equation (2.1) and

initial distribution with density $p_\Delta(b, x)$, respectively. Let us call a diffusion of the latter type a $p_\Delta(b)$ -diffusion. Finally, for any $x \in C_\Delta$, let A_x be the set of functions $y \in C_\Delta$ that intersect x . Specifically,

$$A_x = \{y \in C_\Delta \mid \text{gr}(y) \cap \text{gr}(x) \neq \emptyset\},$$

where $\text{gr}(x) = \{(t, x_t) \mid t \in [0, \Delta]\}$. With these definitions, the relation between the distribution of the approximate bridge Z and the exact $(0, a, \Delta, b)$ -diffusion bridge is

$$f_a(x) = f_b(x)\pi_\Delta(x)/\pi_\Delta. \tag{2.8}$$

Here

$$\pi_\Delta(x) = P(Y \in A_x), \quad \pi_\Delta = P((X, Y) \in A), \tag{2.9}$$

where $A = \{(x, y) \in C_\Delta^2 \mid y \in A_x\}$, X and Y are independent, X is a $(0, a, \Delta, b)$ -diffusion bridge, and Y is a $p_\Delta(b)$ -diffusion. Clearly, $\pi_\Delta(x)$ is the probability that Y hits the trajectory x , while π_Δ is probability that a $(0, a, \Delta, b)$ -bridge is hit by an independent diffusion with initial distribution $p_\Delta(b, \cdot)$. To prove equation (2.8), note that the joint density of a diffusion bridge and an independent $p_\Delta(b)$ -diffusion given that they intersect is

$$f_b(x)f_d(y)1_A(x, y)/\pi.$$

From this expression, (2.8) follows by marginalization. Equation (2.8) is important in two ways: it gives an explicit expression of the quality of our approximate simulation method, and more importantly, it can be used to improve the approximation. In Section 2.3, we will present two MCMC-algorithms, that improves the quality of the approximation. In fact, one of them gives exact diffusion bridges.

Obviously, the quality of our approximate bridge simulation scheme depends on the probability π_Δ that a $(0, a, \Delta, b)$ -bridge is hit by an independent diffusion with initial distribution $p_\Delta(b, \cdot)$. When π_Δ is close to one, the simulated process is essentially a $(0, a, \Delta, b)$ -bridge. It is important to realize that the probability π_Δ is not equal to the acceptance probability $P(\tau \leq \Delta)$. It is quite possible that $P(\tau \leq \Delta)$ is small, while π_Δ is close to one. This happens, for instance, for a diffusion with mean reversion to a level μ when $a \ll \mu \ll b$. In the next subsection, we prove that π_Δ is close to one for long time intervals, provided that the diffusion is ergodic. In Section 3, we shall investigate when π_Δ can otherwise be expected to be close to one, and when a good approximation to a diffusion bridge is obtained. Simulations indicate that also when π_Δ is not close to one (but also not close to zero), the distribution of the simulated bridge is often indistinguishable from the distribution of an exact diffusion bridge.

2.2. Long time intervals

We shall now prove that for ergodic diffusions, the probability π_Δ that a $(0, a, \Delta, b)$ -bridge is hit by an independent diffusion with initial distribution $p_\Delta(b, \cdot)$ is close to one when Δ is large. This implies that for large time intervals, the probability that the process Z defined in Theorem 2.1

is a $(0, a, \Delta, b)$ -bridge is close to one, that is, the simulated process is essentially a $(0, a, \Delta, b)$ -bridge. This is very fortunate, because the strength of the method presented in this paper is that, contrary to other methods for simulating diffusion bridges, it works well numerically for long intervals, see Section 3 and Theorem 2.4. A diffusion processes satisfying (2.4) is ergodic if

$$\int_{\ell}^z \frac{1}{\sigma^2(x)m(x)} dx = \int_z^r \frac{1}{\sigma^2(x)m(x)} dx = \infty, \tag{2.10}$$

where $m(x)$ is given by (2.3).

The convergence of the probability π_{Δ} to one is exponentially fast in Δ when the diffusion process has a spectral gap $\lambda > 0$, that is, when the infimum λ of the non-zero eigenvalues of the spectrum of the infinitesimal generator of the diffusion is strictly positive. For one-dimensional diffusions, a spectral gap is equivalent to ρ -mixing. Most ergodic diffusions used in practice have this property. Easily checked conditions on the drift and diffusion coefficients ensuring a spectral gap and hence exponential convergence of π_{Δ} can be found in Florens-Zmirou [24], Hansen and Scheinkman [31], Hansen, Scheinkman and Touzi [32] and Genon-Catalot, Jeantheau and Larédo [26].

A simple example is a diffusion with linear drift $\alpha(x) = -\beta(x - \mu)$, $\beta > 0$, which has a spectral gap $\lambda = \beta$, provided that the invariant probability measure (2.6) has finite second moment, cf. Hansen *et al.* [32]. Thus by Theorem 2.3, $1 - \pi_{\Delta} = O(e^{-\beta\Delta/2})$, so the simple method gives a good approximation when $\beta\Delta$ is moderately large, that is, when either Δ is large or when the diffusion moves fast (β large). The Pearson diffusions provide a useful broad class of diffusions with linear drift; see Forman and Sørensen [25]. Ergodic diffusions with linear drift and an arbitrary stationary distribution were given in Bibby, Skovgaard and Sørensen [11].

For a general diffusion satisfying (2.4) and (2.10), Hansen *et al.* [32] showed that if the function

$$\gamma(x) = \sigma'(x) - \frac{2\alpha(x)}{\sigma(x)}$$

has nonzero limits as $x \downarrow \ell$ and $x \uparrow r$, then the diffusion has a spectral gap $\lambda > 0$. Genon-Catalot *et al.* [26] gave an explicit lower bound for λ , which gives a useful bound on the rate of convergence of the probability π_{Δ} . To give this lower bound, we need the functions (defined on the state space of the diffusion)

$$\varphi(x) = \frac{\int_x^r m(y) dy}{\sigma(x)m(x)}, \quad \psi(x) = \frac{\int_{\ell}^x m(y) dy}{\sigma(x)m(x)},$$

and

$$C_1(x) = \sup\{\varphi^2(S^{-1}(y)): y \geq x\}, \quad C_0(x) = \sup\{\psi^2(S^{-1}(y)): y \leq x\},$$

where S is the scale function

$$S(x) = \int_z^x \frac{1}{\sigma^2(y)m(y)} dy.$$

Genon-Catalot *et al.* [26] showed that

$$\lambda \geq C = \frac{1}{8 \inf_{x \in \mathbb{R}} \max\{C_1(x), C_0(x)\}}.$$

If the limits of $\gamma(x)$ exist and are nonzero, then $C > 0$, and Theorem 2.3 implies that $1 - \pi_\Delta = O(e^{-C\Delta/2})$.

The results discussed here are summarized in the following theorem.

Theorem 2.3. *For ergodic diffusions*

$$\pi_\Delta \rightarrow 1$$

as $\Delta \rightarrow \infty$. If the diffusion has a spectral gap $\lambda > 0$, then

$$1 - \pi_\Delta = O(e^{-\lambda\Delta/2}).$$

Proof. π_Δ is the probability that a $(0, a, \Delta, b)$ -bridge X^b is hit by an independent diffusion X^d with initial distribution $p_\Delta(b, \cdot)$. Let x denote the point from which X^d starts, and assume that $x > a$. The case $x < a$ can be treated similarly.

By Karlin and McGregor [36] (page 1144) the probability that $X^d_{\Delta/2} > q$ and $X^b_{\Delta/2} \leq q$ without having been coincident in $[0, \Delta/2]$ (conditional on $X^d_0 = x$) is

$$P_{q,\Delta} = \det \left\{ \begin{array}{cc} P(X^b_{\Delta/2} \leq q) & P(X^b_{\Delta/2} > q) \\ P(X^d_{\Delta/2} \leq q | X^d_0 = x) & P(X^d_{\Delta/2} > q | X^d_0 = x) \end{array} \right\}.$$

Here q is an arbitrary real number. That the result holds for time-inhomogeneous diffusions too follows from Karlin [35]. Since the diffusion is ergodic, $X^d_{\Delta/2}$ converges weakly to the stationary distribution with density ν . Then the density of $X^b_{\Delta/2}$ (cf. (2.7)) satisfies

$$q(a, 0, y, \Delta/2) = \frac{p_{\Delta/2}(a, y)p_{\Delta/2}(y, b)}{p_\Delta(a, b)} \rightarrow \frac{\nu(y)\nu(b)}{\nu(b)} = \nu(y),$$

as $\Delta \rightarrow \infty$. Hence

$$P_{q,\Delta} \rightarrow \det \left\{ \begin{array}{cc} P_\nu((-\infty, q]) & P_\nu((q, \infty]) \\ P_\nu((-\infty, q]) & P_\nu((q, \infty]) \end{array} \right\} = 0,$$

as $\Delta \rightarrow \infty$, where P_ν denotes the stationary distribution. This implies that $\pi_\Delta \rightarrow 1$.

Now assume that $\lambda > 0$, where λ is the infimum of the nonzero eigenvalues of the spectrum of the infinitesimal generator of the diffusion. Then $p_{\Delta/2}(x, y) = \nu(y)(1 + O(e^{-\lambda\Delta/2}))$, see, for example, Karlin and Taylor [37], page 332, and it follows that the density function of $X^b_{\Delta/2}$ also satisfies that $q(a, 0, y, \Delta/2) = \nu(y)(1 + O(e^{-\lambda\Delta/2}))$. Hence,

$$P_{q,\Delta} = \det \left\{ \begin{array}{cc} P_\nu((-\infty, q]) & P_\nu((q, \infty]) \\ P_\nu((-\infty, q]) & P_\nu((q, \infty]) \end{array} \right\} + O(e^{-\lambda\Delta/2}) = O(e^{-\lambda\Delta/2}). \quad \square$$

If we replace $\Delta/2$ by the time $\gamma\Delta$ ($0 < \gamma < 1$) in the proof, it follows that $1 - \pi_\Delta = O(e^{-\gamma\lambda\Delta})$. In practice, this refinement does not make much difference.

A similar result can be proved for the rejection probability in the same way. Let $p_\Delta = P(\tau > \Delta)$ denote the rejection probability.

Theorem 2.4. *For ergodic diffusions*

$$p_\Delta \rightarrow 0$$

as $\Delta \rightarrow \infty$. If the diffusion has a spectral gap $\lambda > 0$, then

$$p_\Delta = O(e^{-\lambda\Delta/2}).$$

Theorem 2.4 implies that for ergodic diffusions the computational complexity of our method is linear in the time distance between the two end-points of the diffusion bridge, provided that the diffusion is simulated by a scheme that is linear in the interval length. This is true of simple simulation method like the Milstein scheme.

2.3. Exact bridge simulation

Since we know the relationship between the distribution of an exact diffusion bridge and the distribution of our simple approximation, cf. (2.8), it is natural to consider a Metropolis–Hastings algorithm for which the proposal is our simple simulation method and the target distribution is the distribution of an exact diffusion bridge. In each step, we draw an independent sample path $X^{(i)}$ with distribution given by f_a by means of the simple simulation method. The proposed sample path is accepted with probability $\alpha(X^{(i-1)}, X^{(i)}) = \min(1, r(X^{(i-1)}, X^{(i)}))$, where

$$r(X^{(i-1)}, X^{(i)}) = \frac{f_b(X^{(i)})f_a(X^{(i-1)})}{f_b(X^{(i-1)})f_a(X^{(i)})} = \frac{\pi_\Delta(X^{(i-1)})}{\pi_\Delta(X^{(i)})}.$$

Here $X^{(i-1)}$ is the previously accepted sample path, and $\pi_\Delta(x)$ is the probability that the sample path x is hit by an independent $p_\Delta(b)$ -diffusion, given by (2.9). The MH-algorithm produces draws of exact diffusion bridges, but as $\pi_\Delta(x)$ is not explicitly known, the algorithm cannot be used as it stands.

One possibility is to use a Monte Carlo within Metropolis algorithm. This can be done by simulating in each step N independent $p_\Delta(b)$ -diffusions, $\mathbf{Y}^{(i)} = (Y^{(i,1)}, \dots, Y^{(i,N)})$ and estimate $\pi_\Delta(x)$ consistently by

$$\tilde{\pi}_\Delta(x; \mathbf{Y}^{(i)}) = \frac{1}{N} \sum_{j=1}^N 1_{A_x}(Y^{(i,j)}).$$

If N is sufficiently large, this would certainly produce a very good approximation to a diffusion bridge, but not exact diffusion bridges. Usually, the density function $p_\Delta(b, x)$ is not explicitly known, but a $p_\Delta(b)$ -diffusion can be simulated as follows. Simulate the solution V to (2.1) in the time interval $[0, 2\Delta]$ with $V_0 = b$. Then $Y_t = V_{t+\Delta}$, $t \in [0, \Delta]$, is a $p_\Delta(b)$ -diffusion.

In order to simulate exact diffusion bridges, we propose an MCMC algorithm of the pseudo-marginal type studied in Andrieu and Roberts [3]. The basic idea of the pseudo-marginal approach is to replace the factor in the acceptance ratio, which we cannot calculate, $f_b(x)/f_a(x) = 1/\pi_\Delta(x)$ by an unbiased MCMC estimate. The beauty of the method is that by including the MCMC draws needed for the estimate of $1/\pi_\Delta(x)$ in the MH-Markov chain, the marginal equilibrium distribution of the bridge draws is exactly f_b , irrespective of the randomness of the estimate of $1/\pi_\Delta(x)$.

Define a random variable T in the following way. For a given sample path $x \in C_\Delta$, simulate a sequence of independent $p_\Delta(b)$ -diffusions $Y^{(1)}, Y^{(2)}, \dots$ until a sample path is obtained that intersects x . Then T is defined as the number of the first $Y^{(i)}$ that hits x :

$$T = \min\{i: Y^{(i)} \in A_x\}.$$

By results for the geometric distribution $E(T) = 1/\pi_\Delta(x)$, so if $\mathbf{T} = (T_1, \dots, T_N)$ is a vector of N independent draws of T , then an unbiased and consistent estimator of $1/\pi_\Delta(x)$ is

$$\hat{\rho}_\Delta(x; \mathbf{T}) = \frac{1}{N} \sum_{j=1}^N T_j.$$

Now consider the following MH-algorithm, where draws of \mathbf{T} are included in the Markov chain. In each step, we draw (independently of the previous draws) a sample paths $X^{(i)}$ by the simple algorithm, and with $x = X^{(i)}$ we draw a vector of N independent T -values, $\mathbf{T}^{(i)} = (T_1^{(i)}, \dots, T_N^{(i)})$. Note that the distribution of $\mathbf{T}^{(i)}$ depends on $X^{(i)}$. The proposed update $(X^{(i)}, \mathbf{T}^{(i)})$ is accepted with probability

$$\hat{\alpha}(X^{(i-1)}, \mathbf{T}^{(i-1)}, X^{(i)}, \mathbf{T}^{(i)}) = \min(1, \hat{r}(X^{(i-1)}, \mathbf{T}^{(i-1)}, X^{(i)}, \mathbf{T}^{(i)})),$$

where

$$\hat{r}(X^{(i-1)}, \mathbf{T}^{(i-1)}, X^{(i)}, \mathbf{T}^{(i)}) = \frac{\hat{\rho}_\Delta(X^{(i)}; \mathbf{T}^{(i)})}{\hat{\rho}_\Delta(X^{(i-1)}; \mathbf{T}^{(i-1)})}.$$

By results in Andrieu and Roberts [3], the target distribution of X is that of an exact diffusion bridge. In fact, since

$$\hat{r}(x^{(1)}, \mathbf{t}^{(1)}, x^{(2)}, \mathbf{t}^{(2)}) = \frac{f_a(x^{(2)})\mathbf{f}_g(\mathbf{t}^{(2)}|x^{(2)})\hat{\rho}_\Delta(x^{(2)}; \mathbf{t}^{(2)})f_a(x^{(1)})\mathbf{f}_g(\mathbf{t}^{(1)}|x^{(1)})}{f_a(x^{(1)})\mathbf{f}_g(\mathbf{t}^{(1)}|x^{(1)})\hat{\rho}_\Delta(x^{(1)}; \mathbf{t}^{(1)})f_a(x^{(2)})\mathbf{f}_g(\mathbf{t}^{(2)}|x^{(2)})},$$

where $\mathbf{f}_g(\mathbf{t}|x)$ is the conditional density of \mathbf{T} given $X = x$, we see that the density of the target distribution is

$$p(x, \mathbf{t}) = f_a(x)\mathbf{f}_g(\mathbf{t}|x)\hat{\rho}_\Delta(x, \mathbf{t})\pi_\Delta = f_b(x)\mathbf{f}_g(\mathbf{t}|x)\hat{\rho}_\Delta(x, \mathbf{t})\pi_\Delta(x),$$

where we have used (2.8). Since $\hat{\rho}_\Delta(x, \mathbf{t})$ is an unbiased estimator of $1/\pi_\Delta(x)$ conditionally on x , we find by marginalizing that the density of X is f_b , the density function of an exact diffusion bridge.

To produce diffusion bridges by the proposed pseudo-marginal MH-algorithm, a number of sample paths of ordinary diffusions must be simulated. If these sample paths are simulated by an approximate method, like the Milstein scheme, a small discretization error is introduced. This problem can, however, be avoided by using the methods for exactly simulating diffusions developed by Beskos, Papaspiliopoulos and Roberts [6,7]. By combining our exact MH bridge simulation algorithm with exact diffusion simulation methods, exact diffusion bridges can be efficiently simulated even in long time intervals.

For ergodic diffusions, the computational complexity of the exact algorithm is linear in the interval length Δ . In the previous subsection, we saw that simulation of the proposal is linear in Δ , and the expected number of $p_\Delta(b)$ -diffusions simulated in the i th M–H step is $N/\pi_\Delta(X^{(i)})$. For a diffusion bridge X , the expectation of $\pi_\Delta(X)$ is π_Δ , which by Theorem 2.3 converges to one as $\Delta \rightarrow \infty$. Since π_Δ is usually not small, each iteration of the M–H-algorithm producing exact diffusion bridges is not expected to require much more computing time than the approximate algorithm. Finally, since the acceptance ratio tends to one as $\Delta \rightarrow \infty$, the acceptance rate is high when Δ is large.

3. Simulation study

In this section, we investigate simulation of two examples of diffusion bridges for which our new methods can be compared to other exact algorithms, the Ornstein–Uhlenbeck process and the hyperbolic diffusion. We compare the distribution of exact bridge simulations to the distribution of the process obtained by our approximate bridge simulation method. It is found that the approximation is very accurate in most cases, including bridges that are likely to occur in applications to likelihood inference. We also compare CPU execution times, and illustrate how our exact M–H simulation method provides an exact bridge in the extreme cases where the approximate method does not provide a good approximation.

3.1. The Ornstein–Uhlenbeck bridge

First, we consider an Ornstein–Uhlenbeck bridge, which is a solution to the stochastic differential equation

$$dX_t = -\theta X_t dt + \sigma dW_t$$

conditionally on $X_0 = a$ and $X_1 = b$ for some $a, b \in \mathbb{R}$. From the well-known Gaussian transition densities of the Ornstein–Uhlenbeck process we can calculate the transition densities of the Ornstein–Uhlenbeck bridge by (2.7). Thus, we could in principle simulate the Ornstein–Uhlenbeck bridge by sampling transitions from these densities. The following well-known alternative method is, however, numerically more stable.

Lemma 3.1. *Generate $X_{t_0}, X_{t_1}, \dots, X_{t_n}, X_{t_{n+1}}$, where $0 = t_0 < t_1 < \dots < t_n < t_{n+1}$, by $X_0 = x_0$ and*

$$X_{t_i} = e^{-\theta(t_i - t_{i-1})} X_{t_{i-1}} + W_i, \quad i = 1, \dots, n + 1,$$

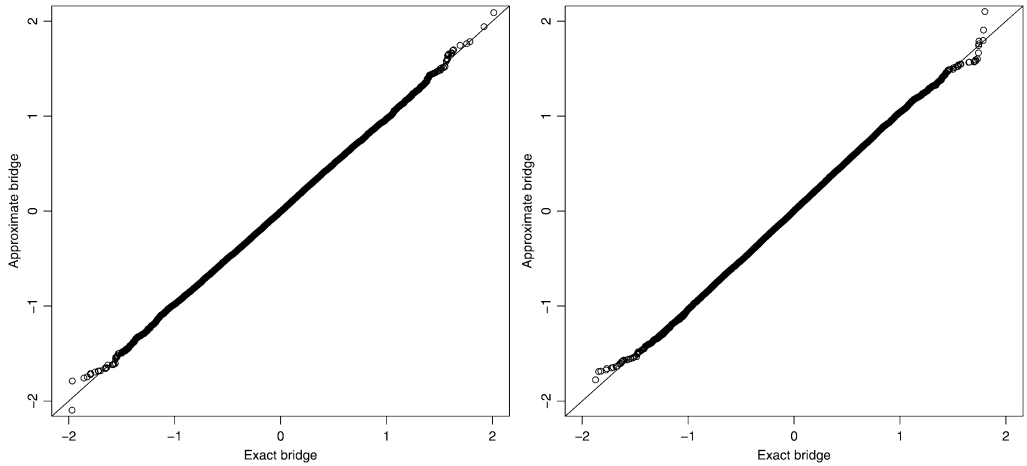


Figure 1. Q–Q plots that compare the empirical distribution at time 0.5 based on 25,000 simulated (0, 0) diffusion bridges obtained by our approximate method to that based on 25,000 exactly simulated diffusion bridges. The left plot is for the Ornstein–Uhlenbeck bridge and the right plot is for the hyperbolic diffusion bridge. Exact simulations are obtained by the method in Lemma 3.1 for the Ornstein–Uhlenbeck bridge and by the exact algorithm of Beskos, Papanioliopoulos and Roberts for the hyperbolic diffusion bridge.

where the W_i s are independent and $W_i \sim N(0, \sigma^2(1 - e^{-2\theta(t_i - t_{i-1})})/(2\theta))$. Define

$$Z_{t_i} = X_{t_i} + (x - X_{t_{n+1}}) \frac{e^{\theta t_i} - e^{-\theta t_i}}{e^{\theta t_{n+1}} - e^{-\theta t_{n+1}}}, \quad i = 0, \dots, n + 1.$$

Then $(Z_{t_0}, Z_{t_1}, \dots, Z_{t_n}, Z_{t_{n+1}})$ is distributed like an Ornstein–Uhlenbeck bridge with $Z_{t_0} = x_0$ and $Z_{t_{n+1}} = x$.

In all examples considered in the following we simulated 25,000 (or 10,000) realizations of diffusion bridges over the time interval $[0, 1]$. The Euler scheme was used with discretization level $N = 100$ (step size $\delta = 0.01$). For the Ornstein–Uhlenbeck process the Euler scheme is equal to the Milstein scheme. The methods were implemented in Fortran 90 on a Dell Precision M65 workstation (laptop).

For the Ornstein–Uhlenbeck bridge, we chose the parameter values $\theta = 0.5$ and $\sigma = 1.0$. First we considered a bridge that started at 0 and ended at 0. We compare our approximate method based on Theorem 2.1 to the exact algorithm of Lemma 3.1. To the left in Figure 1, we have plotted the quantiles of the empirical distribution at the time point 0.5 obtained by our approximate method against the quantiles of the empirical distribution obtained by the exact algorithm. The two distributions appear to be equal. Similar comparisons of quantiles at time 0.5 for our approximate method to quantiles of an exact bridge are presented in Figure 2 for (0, 1), (0, 2), (−1, 1) and (−1, 2) Ornstein–Uhlenbeck bridges. In all four cases, the two distributions seem to be essentially equal, except for a very small negative bias for the (0, 2)-bridge. Similar

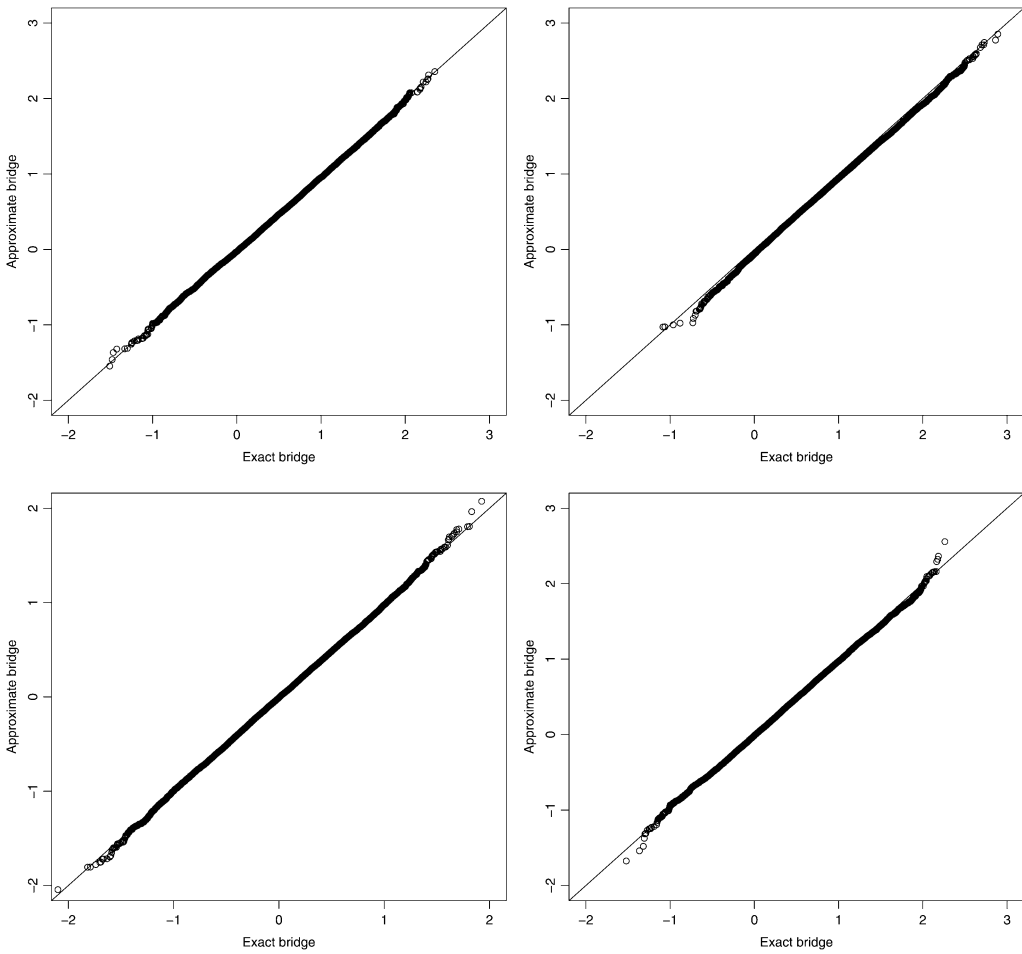


Figure 2. Q–Q plots that compare the empirical distributions at time 0.5 based on 25,000 simulated $(0, 1)$, $(0, 2)$, $(-1, 1)$ and $(-1, 2)$ Ornstein–Uhlenbeck bridges obtained by our approximate method to that based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges. Exact simulations are obtained by the method in Lemma 3.1.

results were found for several other comparisons of distributions with similar values of the start and end points, a and b .

The CPU execution time (in seconds) to simulate 10,000 Ornstein–Uhlenbeck bridges using our approximate method for the various starting points, a , and end points, b , are given in Table 1 together with estimated rejection probabilities (p_{Δ} , see Theorem 2.4). The table also gives the probabilities that an Ornstein–Uhlenbeck process moves from a to b or farther in the time interval $[0, 1]$. We see that for moves that are likely to appear in data sets, the CPU times and rejection

Table 1. The CPU execution time (in seconds) used to simulate 10,000 Ornstein–Uhlenbeck bridges using our approximate method for various starting points, a , and end points, b . Estimated rejection probabilities (p_Δ , see Theorem 2.4) and the probabilities of a move from a to b or farther are listed too. The last column gives the probability that an exact (a, b) -bridge is not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$

| $a \mapsto b$ | CPU (sec.) | Rejection prob. | Probability of move | $1 - \pi$ |
|----------------|------------|-----------------|---------------------|-----------|
| $0 \mapsto 0$ | 0.5 | 0.17 | | 0.28 |
| $0 \mapsto 1$ | 0.7 | 0.41 | 0.1 | 0.21 |
| $0 \mapsto 2$ | 1.7 | 0.77 | 0.006 | 0.08 |
| $-1 \mapsto 1$ | 1.9 | 0.80 | 0.02 | 0.16 |
| $-1 \mapsto 2$ | 11.9 | 0.97 | 0.0005 | 0.06 |

probabilities are small, and the CPU times are only slightly larger than the execution time for the exact algorithm which is about 0.5 CPU seconds. For more unlikely moves the rejection probability is quite large, but also in these cases the execution time is not a problem in applications. The last column of Table 1 gives the (estimated) probability of the event that an exact (a, b) -bridge is not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$. These probabilities were found by simulating exact Ornstein–Uhlenbeck bridges and independent Ornstein–Uhlenbeck processes with initial distribution $p_1(b, \cdot)$. If this probability were zero, our approximate method would simulate an exact diffusion bridge. The probabilities are small, but not negligible. It is remarkable that the approximate method gives a quite accurate approximation to a diffusion bridge in spite of this. The reason must be that the diffusion bridges are not hit by the independent diffusion in a systematic way for the a and b values considered here.

In order to test our approximate method in an extreme situation, we simulated 25,000 Ornstein–Uhlenbeck bridges that started from -2 and ended in 2 . The probability that an Ornstein–Uhlenbeck process with parameters $\theta = 0.5$ and $\sigma = 1.0$ moves from -2 to 2 or farther in the time interval $[0, 1]$ equals the probability that a standard normal distribution is larger than 4.04 , which is 0.00003 , so this is indeed a very extreme event. Not surprisingly that rejection rate was very high, but as appears from Figure 3 the distribution at time 0.5 fits the distribution obtained by exact simulation very well.

The only situation we have been able to find where the distribution obtained by our approximate simulation method differs appreciably from the distribution of an exact bridge is when the start and end points, a and b , have the same sign and are both far from the equilibrium point zero. This is to be expected because we simulate an exact bridge conditional on the event that it is hit by an independent diffusion with initial distribution $p_1(b, \cdot)$. When b is far from zero, most of the probability mass of $p_1(b, \cdot)$ is located considerably closer to zero than b (because of the drift towards zero). The independent diffusion will tend to move towards zero, while the (a, b) -bridge will tend to stay relatively close to a and b . Only trajectories of the bridge that move sufficiently towards zero has a reasonable chance of being hit by the independent diffusion. This creates a bias towards zero. The comparison of quantiles at time 0.5 and time 0.1 for $(-2, -2)$ and $(-3, -2)$ bridges are presented in Figure 4. As expected from the consideration above, there is a positive bias. The time point 0.5 was chosen in most simulations because it is expected that

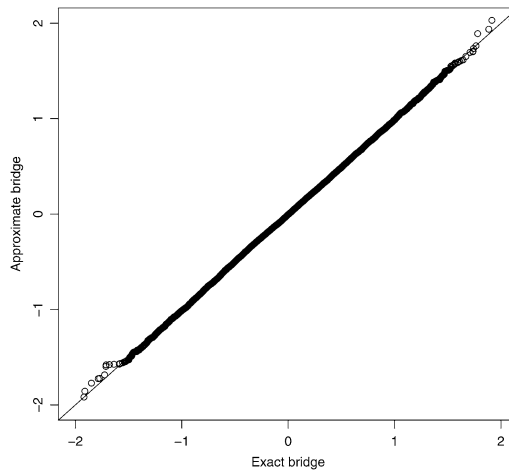


Figure 3. Q–Q plot that compares the empirical distribution at time 0.5 based on 25,000 simulated $(-2, 2)$ Ornstein–Uhlenbeck bridges obtained by our approximate method to that based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges.

this is the time point where it is most difficult to get a good approximation of the distribution of a diffusion bridge. The comparison of distributions at time 0.1 illustrate that the approximate method works better close to the end-points of the time interval, where the bias is considerably smaller. Several similar comparisons confirm that the approximate method works better close to the end-points than at time 0.5.

The first row of Figure 5 illustrates Theorem 2.1. For the $(-2, -2)$ and $(-3, -2)$ Ornstein–Uhlenbeck bridges Q–Q plots compare empirical distributions at time 0.5 of 25,000 approximate simulations to the similar empirical distributions based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges, where the simulated bridges were removed from the sample if the bridge was not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$. As expected from Theorem 2.1, the two distributions appear to be equal (there is a numerical problem in the left tail for the $(-2, -2)$ -bridge). We know from Figure 4 that this distribution differs from that of the unconditional bridge. In the second row of Figure 5, Q–Q plots compare the empirical distributions at time 0.5 based on 25,000 $(-2, -2)$ (left) and $(-3, -2)$ (right) Ornstein–Uhlenbeck bridges simulated by our MH-algorithm to the empirical distribution based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges. For the MH-algorithm the burn-in was 5000 iterations and $N = 10$ (the number of T -values simulated in each step). The algorithm worked well for a much shorter burn-in and for $N = 1$. Also these distributions appear to be equal (again there is a numerical problem in the left tail for the $(-2, -2)$ -bridge). The autocorrelations of the MH-algorithm decreases very quickly to zero. The empirical autocorrelation function of the successive values of the $(-3, -2)$ Ornstein–Uhlenbeck bridge at time 0.5 obtained by our Metropolis–Hastings algorithm is plotted in Figure 6. After less than 10 iterations the correlation is essentially zero.

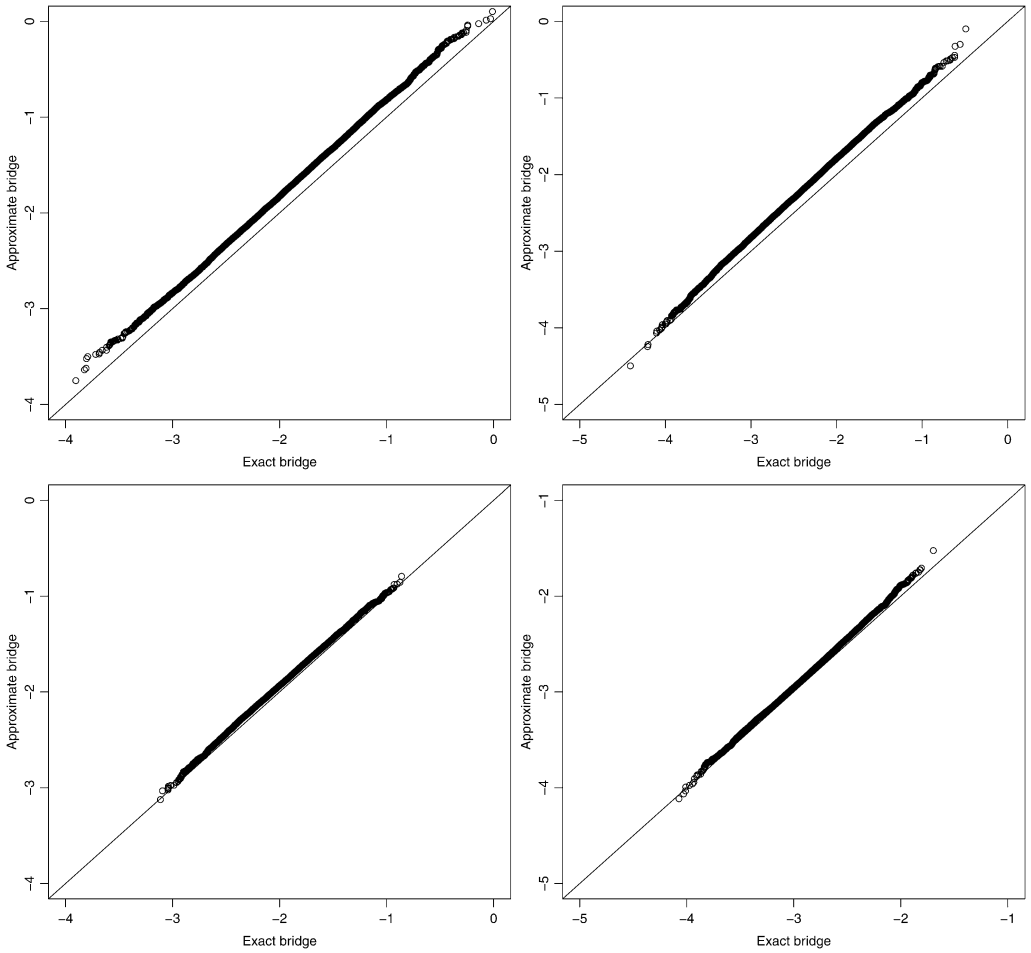


Figure 4. Q–Q plots that compare the empirical distributions at time 0.5 (first row) and time 0.1 (second row) based on 25,000 simulated $(-2, -2)$ (left) and $(-3, -2)$ (right) Ornstein–Uhlenbeck bridges obtained by our approximate method to that based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges. Exact simulations are obtained by the method in Lemma 3.1.

Table 2 gives estimated rejection probabilities (of the approximate rejection sampler) and the probability that an exact (a, b) -bridge is not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$. As expected from the discussion above, the probabilities of not being hit is quite substantial. The last column gives the probability that a stationary Ornstein–Uhlenbeck process is a distance $|a|$ or more from zero. We see that the process will only spend very little time in the parts of the state space, where the approximate method is biased.

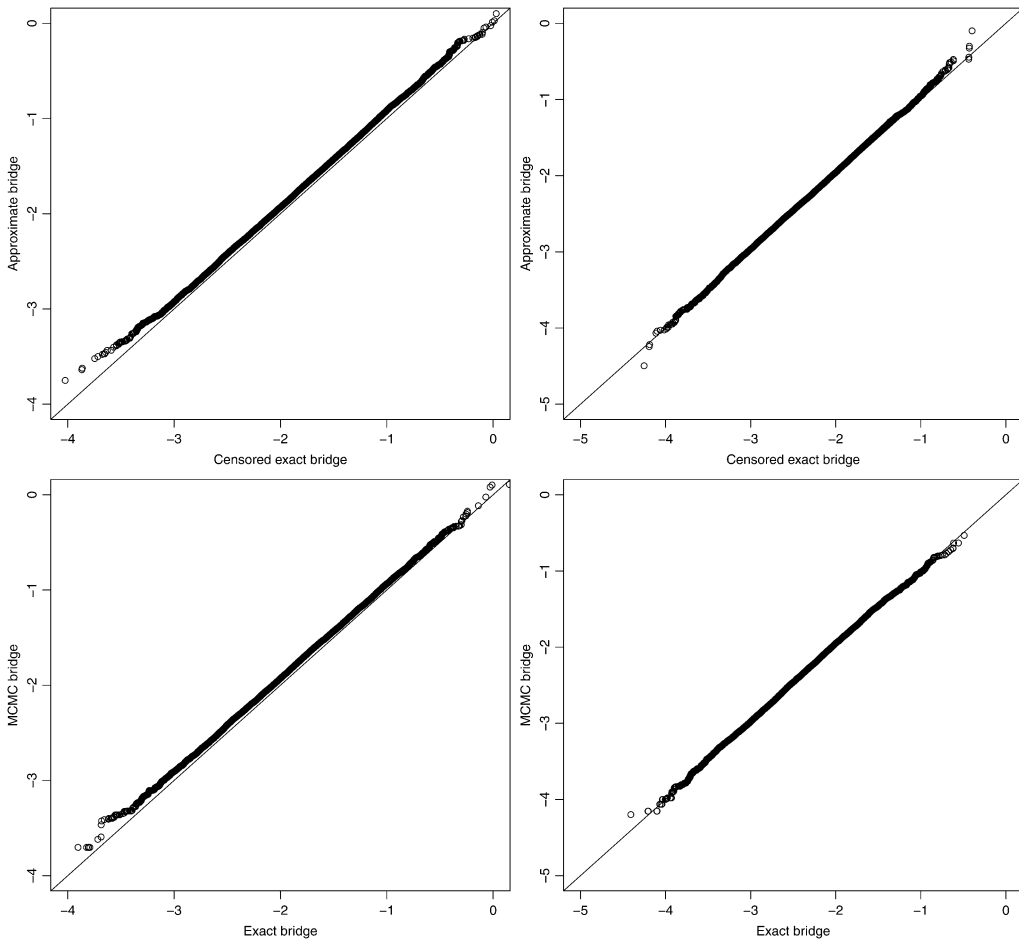


Figure 5. In the first row, Q–Q plots compare the empirical distributions at time 0.5 based on 25,000 simulated $(-2, -2)$ (left) and $(-3, -2)$ (right) Ornstein–Uhlenbeck bridges obtained by our approximate method to that based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges, where the bridges that were not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$ were removed from the sample. In the second row, Q–Q plots compare the empirical distributions at time 0.5 based on 25,000 simulated $(-2, -2)$ (left) and $(-3, -2)$ (right) Ornstein–Uhlenbeck bridges obtained by our Metropolis–Hastings algorithm (after a burn-in of 5000 iterations) to that based on 25,000 exactly simulated Ornstein–Uhlenbeck bridges.

3.2. The hyperbolic bridge

Next we consider the hyperbolic diffusion which is the solution to

$$dX_t = -\frac{\theta X_t}{\sqrt{1 + X_t^2}} dt + \sigma dW_t,$$

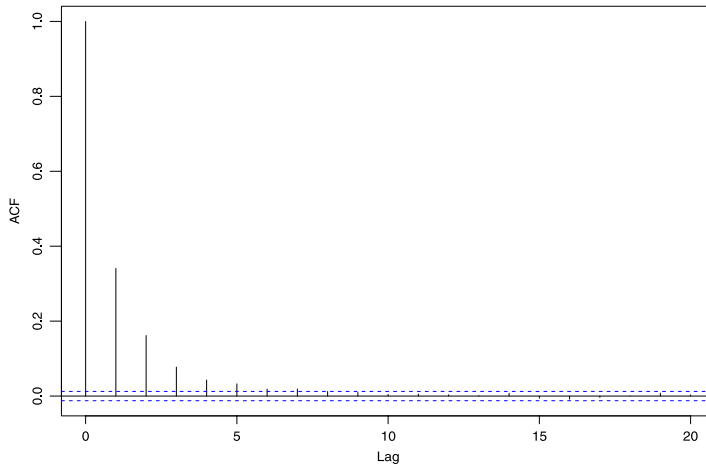


Figure 6. The empirical autocorrelation function of the successive values of the $(-3, -2)$ Ornstein-Uhlenbeck bridge at time 0.5 obtained by our Metropolis-Hastings algorithm.

with $\theta > 0$ and $\sigma > 0$. The hyperbolic diffusion was introduced by Barndorff-Nielsen [5]. It is ergodic with the standardized symmetric hyperbolic distribution as invariant distribution, see, for example, Bibby and Sørensen [10]. In this case the transition density is not explicitly known, but we can compare our method to the exact EA1 algorithm by Beskos, Papaspiliopoulos and Roberts [6]. It is applicable to diffusion processes on the form

$$dX_t = \alpha(X_t)dt + dW_t, \tag{3.1}$$

provided that α is continuously differentiable, and the function $\alpha(x)^2 + \alpha'(x)$ is bounded from above and below for all x , conditions satisfied by the hyperbolic diffusion process. The algorithm by Beskos, Papaspiliopoulos and Roberts [6] is very quick for short intervals as it essentially only requires one simulation of a Brownian bridge if it is not rejected. Rejection in the EA1 algorithm

Table 2. Estimated rejection probabilities (p_Δ , see Theorem 2.4) for the Ornstein-Uhlenbeck bridges using our approximate method for various starting points, a , and end points, b . The second column gives the probability that an exact (a, b) -bridge is not hit by an independent diffusion with initial distribution $p_1(b, \cdot)$. The last column gives the probability of finding an Ornstein-Uhlenbeck a distance $|a|$ or more from zero

| $a \mapsto b$ | Rejection prob. | $1 - \pi$ | $2P(X_t > a)$ |
|-----------------|-----------------|-----------|-----------------|
| $-2 \mapsto -2$ | 0.09 | 0.55 | 0.05 |
| $-3 \mapsto -2$ | 0.24 | 0.74 | 0.003 |

Table 3. The CPU execution time (in seconds) used to simulate 10,000 hyperbolic diffusion bridges with $\theta = \sigma = 1$ by our method for various starting points, a , and end points, b . Also estimated rejection probabilities are given (p_Δ , see Theorem 2.4)

| $a \mapsto b$ | CPU (sec.) | Rejection prob. |
|----------------|------------|-----------------|
| $0 \mapsto 0$ | 0.6 | 0.14 |
| $0 \mapsto 1$ | 0.8 | 0.36 |
| $0 \mapsto 2$ | 2.1 | 0.77 |
| $-1 \mapsto 1$ | 2.0 | 0.76 |
| $-1 \mapsto 2$ | 12.6 | 0.96 |

is not very costly computationally in our example since it is only a few points that are thrown away per rejection. Thus, we compare our algorithm to a very efficient method.

Again we simulated 25,000 bridges using the Euler scheme with a 100 points subdivision of $[0, 1]$. Also for the hyperbolic diffusion the Euler scheme equals the Milstein scheme. The parameter values were $\theta = \sigma = 1$. We start with a bridge from 0 to 0 and compare our approximate method to the exact EA1 algorithm. To the right in Figure 1, we have plotted the quantiles of the empirical distribution at the time point 0.5 obtained by our approximate method against the quantiles of the empirical distribution obtained by the exact EA1 algorithm. Also for this example the two distributions appear to be equal. Table 3 shows CPU execution times to simulate 10,000 hyperbolic diffusion bridges by our approximate method for various starting points, a , and end points, b . Also estimated rejection probabilities are given. The pattern is similar to that for the Ornstein–Uhlenbeck process. For moves that are likely to appear in data sets, the CPU times and rejection probabilities are small, and for unlikely moves the execution time is not a problem in applications, even though the rejection probability is quite large. The execution time for the EA1 algorithm was 0.3 CPU seconds, which, as expected, is faster than our method. Note that there is no reason to consider diffusions for which the EA1 algorithm does not work in order to compare our method to the more complicated simulation methods EA2 and EA3 in Beskos, Papaspiliopoulos and Roberts [6] and Beskos *et al.* [7]. The EA2 and EA3 algorithms are clearly more time consuming than EA1, while execution times for our methods can be expected to be approximately as for the two examples considered here.

Beskos, Papaspiliopoulos and Roberts [6] noted that the computing time of their exact algorithm is large for diffusion bridges over long time intervals. It is therefore of interest to compare computer time and rejection probabilities for our approximate algorithm to the EA1 algorithm. To do so, we simulated 10,000 trajectories of the $(0, 0, \Delta, 0)$ -bridge for the hyperbolic diffusion with $\theta = \sigma^2 = 4$. This was done for values of the interval length Δ ranging from 0.5 to 5. The CPU execution time (in seconds) used to simulate the 10,000 trajectories are given in Table 4. We see that for this particular diffusion the two methods use the same CPU time for an interval length of two. For smaller interval lengths the exact algorithm is somewhat faster, whereas our approximate method is much faster for long intervals. The simulations confirm that the computational complexity of the proposed method is linear in the interval length Δ as shown in Section 2, whereas the complexity appears to grow at least exponentially with Δ for the ex-

Table 4. The CPU execution time (in seconds) used to simulate 10,000 hyperbolic $(0, 0, \Delta, 0)$ -bridges with $\theta = \sigma^2 = 4$ for our approximate method and for the EA1 method in Beskos *et al.* [7] for different interval lengths Δ . Also the number of rejections while simulating the 10,000 trajectories is given

| Δ | Present paper | | Beskos <i>et al.</i> [7] EA1 | |
|----------|---------------|--------------|------------------------------|--------------|
| | CPU time | # rejections | CPU time | # rejections |
| 0.5 | 0.52 | 819 | 0.28 | 14,497 |
| 1.0 | 0.99 | 307 | 0.59 | 53,087 |
| 1.5 | 1.45 | 102 | 1.05 | 163,599 |
| 2.0 | 1.93 | 44 | 1.92 | 457,226 |
| 2.5 | 2.40 | 17 | 4.00 | 1,242,922 |
| 3.0 | 2.88 | 6 | 10.01 | 3,491,838 |
| 3.5 | 3.36 | 2 | 26.86 | 9,357,310 |
| 4.0 | 3.83 | 0 | 75.79 | 25,232,418 |
| 4.5 | 4.31 | 0 | 222.09 | 69,299,642 |
| 5.0 | 4.79 | 0 | 641.70 | 187,069,771 |

act algorithm; see Figure 7. The main reason is that for long intervals the number of rejections becomes very large for the algorithm in Beskos, Papaspiliopoulos and Roberts [6], while our approximate algorithm has a very small rejection probability for long intervals. The rapid decrease of the rejection probabilities for the approximate method as a function of Δ is expected from Theorem 2.4.

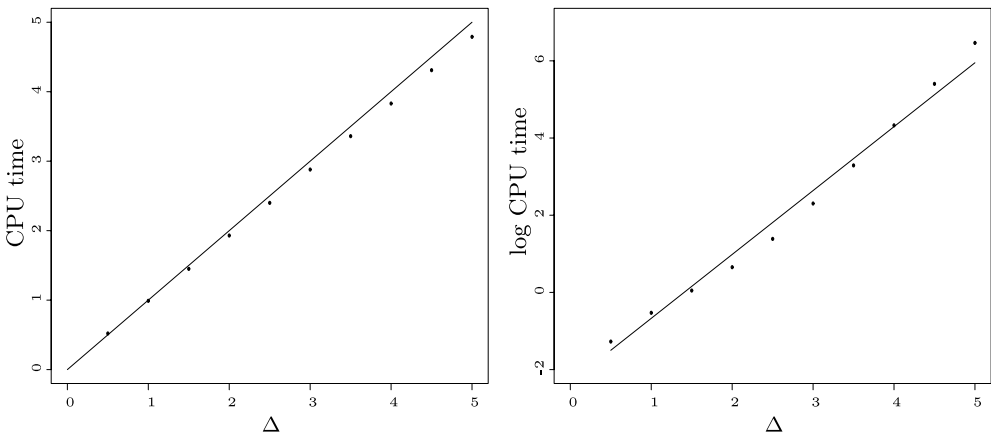


Figure 7. The CPU execution time (in seconds) used to simulate 10,000 hyperbolic $(0, 0, \Delta, 0)$ -bridges with $\theta = \sigma^2 = 4$. In the left plot the CPU time is plotted against Δ for our approximate method, while in the right plot the logarithm of the CPU time is plotted against Δ for the EA1 method in Beskos *et al.* [7].

4. Maximum likelihood estimation

The main motivation for the theory developed in this papers is the central role diffusion bridge simulation plays in simulated likelihood-based inference for processes of the diffusion type. Therefore, we end the paper by giving two examples of application of our diffusion bridge simulation method to maximum likelihood estimation first for discretely observed diffusion processes then for integrated diffusions observed with measurement error.

4.1. Discretely observed diffusions

First, we present an EM-algorithm for finding the maximum likelihood estimator for discretely observed diffusion processes. We also briefly discuss aspects of Bayesian inference.

Consider the diffusion process

$$dX_t = b_\alpha(X_t) dt + \sigma_\beta(X_t) dW_t, \quad (4.1)$$

where α and β are unknown parameters to be estimated, and W is the standard Wiener process. We assume that $\sigma_\beta(x) > 0$ for all x in the state interval. Suppose that the only data available from a realization of the diffusion process are observations at times $t_1 < t_2 < \dots < t_n$, $x_i = X_{t_i}$, $i = 1, \dots, n$.

As explained in the [Introduction](#), discrete time observation of a continuous time process can be viewed as an incomplete observation problem, so the EM-algorithm (Dempster, Laird and Rubin [19]) is a natural method for finding the maximum likelihood estimator of the parameters. Maximum likelihood estimation for discretely observed Markov jump processes was treated in this way by Bladt and Sørensen [13,14]. Unfortunately, the probability measures corresponding to complete continuous time observation of the diffusion model given by (4.1) are singular because the diffusion coefficient depends on the parameter β . It is therefore not straightforward to implement the EM-algorithm, but an approach in the spirit of Roberts and Stramer [47] was proposed by Beskos, Papaspiliopoulos and Roberts [9]. In the following, we summarize a modification of this approach using our diffusion bridge simulation technique.

The transformation

$$h_\beta(x) = \int_{x^*}^x \frac{1}{\sigma_\beta(y)} dy \quad (4.2)$$

is essential. Here x^* is some arbitrary, but appropriately chosen, point in the state interval. By Ito's formula, $Y_t = h_\beta(X_t)$ solves

$$dY_t = \mu_{\alpha,\beta}(Y_t) dt + dW_t, \quad (4.3)$$

where

$$\mu_{\alpha,\beta}(y) = \frac{b_\alpha(h_\beta^{-1}(y))}{\sigma_\beta(h_\beta^{-1}(y))} - \frac{1}{2} \sigma'_\beta(h_\beta^{-1}(y)).$$

In (4.3), the diffusion coefficient does not depend on the parameters, so the probability measures are equivalent and the likelihood function can be found. To do so the function

$$g_{\alpha,\beta}(x) = s_{\alpha,\beta}(x) \frac{1}{2} \log(\sigma_\beta(x)), \tag{4.4}$$

where

$$s_{\alpha,\beta}(x) = \int_{x^*}^x \frac{b_\alpha(z)}{\sigma_\beta^2(z)} dz, \tag{4.5}$$

is needed. Note that $\int_{y^*}^y \mu_{\alpha,\beta}(z) dz = g_{\alpha,\beta}(h_\beta^{-1}(y)) - g_{\alpha,\beta}(h_\beta^{-1}(y^*))$, and that the functions $g_{\alpha,\beta}$ and $s_{\alpha,\beta}$ are closely related to the density $\varphi_{\alpha,\beta}$ of the stationary distribution of the original diffusion model given by (4.1). Specifically, $s_{\alpha,\beta}(x)$ equals $\frac{1}{2} \log(\sigma_\beta(x)^2 \varphi_{\alpha,\beta}(x))$ apart from an additive constant. Thus, when the stationary density is known, the only problem is to find h_β and its inverse. This is for instance the case for the Pearson diffusions studied by Forman and Sørensen [25].

The problem with the transformation h_β is that it is parameter dependent, while we need to keep the original discrete time data fixed when running the EM-algorithm. To get around this problem, define

$$Y_t^*(\beta, \beta_0) = Z_t^{(i,\alpha_0,\beta_0)} + \frac{(t_i - t)(h_\beta(x_{i-1}) - h_{\beta_0}(x_{i-1})) + (t - t_{i-1})(h_\beta(x_i) - h_{\beta_0}(x_i))}{t_i - t_{i-1}}$$

for $t_{i-1} \leq t \leq t_i, i = 2, \dots, n$. Here $Z_t^{(i,\alpha_0,\beta_0)}$ denotes the $(t_{i-1}, h_{\beta_0}(x_{i-1}), t_i, h_{\beta_0}(x_i))$ -bridge for the diffusion (4.3) with parameter values α_0 and β_0 , and $Z_t^{(i,\alpha_0,\beta_0)}, i = 2, \dots, n$ are independent. Then the EM-algorithm works as follows. Let α_0, β_0 be initial values of the parameters.

- (1) (E-step) Calculate the function

$$\begin{aligned} q(\alpha, \beta) &= g_{\alpha,\beta}(x_n) - g_{\alpha,\beta}(x_1) \\ &\quad - \frac{1}{2} \sum_{i=2}^n [h_\beta(x_i) - h_\beta(x_{i-1})]^2 / (t_i - t_{i-1}) - \sum_{i=2}^n \log(\sigma_\beta(x_i)) \\ &\quad - \frac{1}{2} \sum_{i=2}^n \mathbb{E}_{Z^{(i,\alpha_0,\beta_0)}} \left(\int_{t_{i-1}}^{t_i} [\mu'_{\alpha,\beta}(Y_t^*(\beta, \beta_0)) + \mu_{\alpha,\beta}(Y_t^*(\beta, \beta_0))]^2 dt \right). \end{aligned}$$

- (2) (M-step) $(\alpha_0, \beta_0) = \arg \max_{\alpha,\beta} q(\alpha, \beta)$.
- (3) GO TO (1).

In the E-step, $\mathbb{E}_{Z^{(i,\alpha_0,\beta_0)}}$ means that the data points are fixed so that only the diffusion bridge is random, and expectation is with respect to the distribution of the diffusion bridge. Thus, the expectations in the E-step can be approximated by simulating diffusion bridges by our exact MH-method and averaging (after a burn-in period). Arguments that $q(\alpha, \beta)$ is the conditional expectation of the relevant continuous time likelihood function can be found in Roberts and Stramer [47] and Beskos, Papaspiliopoulos, Roberts and Fearnhead [9]. As pointed out in the

latter paper, the conditional expectation can also be calculated as

$$\mathbb{E}_{Z^{(i,\alpha_0,\beta_0)},U}(\mu'_{\alpha,\beta}(Y_U^*(\beta,\beta_0)) + \mu_{\alpha,\beta}(Y_U^*(\beta,\beta_0))^2),$$

where U is a uniformly distributed random variable on $[t_{i-1}, t_i]$ that is independent of $Z_t^{(i,\alpha_0,\beta_0)}$ and the data.

In the M-step the maximization of $q(\alpha, \beta)$ must in general be done by a suitable maximization algorithm. With modern software (e.g., the R-function `optim`), this is not a problem. When the drift of the original diffusion model (4.1) depends linearly on the vector of parameters α , that is, when

$$b_\alpha(x) = \alpha_1 a_1(x) + \dots + \alpha_k a_k(x), \tag{4.6}$$

where a_1, \dots, a_k are known functions, then the maximization problem is simplified somewhat. When the drift has this form, and when the diffusion parameter β is fixed, the model for continuous time observation of X as well as the transformed process Y is an exponential family of stochastic processes, see K uchler and S orenson [41], page 27. We can therefore take advantage of well known properties of exponential families of diffusions.

For the EM-algorithm the specification (4.6) implies that the function $q(\alpha, \beta)$ has the form

$$q(\alpha, \beta) = \sum_{i=1}^k \alpha_i H_{i,\beta} - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j B_{i,j,\beta} + G_\beta,$$

where

$$\begin{aligned} H_{i,\beta} &= s_{i,\beta}(x_n) - s_{i,\beta}(x_1) \\ &+ \sum_{j=2}^n \mathbb{E}_{Z^{(j,\alpha_0,\beta_0)}} \left(\int_{t_{j-1}}^{t_j} \left[a_i(h_\beta^{-1}(Y_t^*(\beta,\beta_0))) (\log \sigma_\beta)'(h_\beta^{-1}(Y_t^*(\beta,\beta_0))) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} a_i'(h_\beta^{-1}(Y_t^*(\beta,\beta_0))) \right] dt \right), \end{aligned}$$

with $s_{i,\beta}(x) = \int_{x^*}^x a_i(y) / \sigma_\beta^2(y) dy$,

$$B_{i,j,\beta} = \sum_{j=2}^n \mathbb{E}_{Z^{(j,\alpha_0,\beta_0)}} \left(\int_{t_{j-1}}^{t_j} \frac{a_i(h_\beta^{-1}(Y_t^*(\beta,\beta_0))) a_j(h_\beta^{-1}(Y_t^*(\beta,\beta_0)))}{\sigma_\beta^2(h_\beta^{-1}(Y_t^*(\beta,\beta_0)))} dt \right),$$

and

$$\begin{aligned} G_\beta &= -\frac{1}{2} \log(\sigma_\beta(x_n) / \sigma_\beta(x_1)) \\ &- \frac{1}{2} \sum_{i=2}^n [h_\beta(x_i) - h_\beta(x_{i-1})]^2 / (t_i - t_{i-1}) - \sum_{i=2}^n \log(\sigma_\beta(x_i)) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{4} \sum_{j=2}^n \mathbb{E}_{Z^{(j, \alpha_0, \beta_0)}} \left(\int_{t_{j-1}}^{t_j} \left[\sigma_\beta''(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) \sigma_\beta(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) \right. \right. \\
 & \qquad \qquad \qquad \left. \left. - \frac{1}{2} \{ \sigma_\beta'(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) \}^2 \right] dt \right).
 \end{aligned}$$

For a fixed value of β , the function $\alpha \mapsto q(\alpha, \beta)$ is maximal for

$$\hat{\alpha}(\beta) = \mathbf{B}_\beta^{-1} \mathbf{H}_\beta,$$

where $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_k)^T$, $\mathbf{H}_\beta = (H_{1,\beta}, \dots, H_{k,\beta})^T$ and $\mathbf{B}_\beta = \{B_{i,j,\beta}\}$. This is provided that \mathbf{B}_β is invertible, which it is when the functions a_i , $i = 1, \dots, k$ are linearly independent. Thus, $q(\alpha, \beta)$ attains its maximal value at $(\hat{\alpha}(\hat{\beta}), \hat{\beta})$, where $\hat{\beta}$ maximizes

$$\beta \mapsto q(\hat{\alpha}(\beta), \beta) = \frac{1}{2} H_\beta^T B_\beta^{-1} H_\beta + G_\beta.$$

The Gibbs sampler for Bayesian inference for discretely observed diffusion processes proposed by Roberts and Stramer [47] can also be modified by replacing the MCMC algorithm for simulating diffusion bridges in that paper by our diffusion bridge simulation method. We will not go into any detail for general diffusions, but will limit ourselves to pointing out that when the drift has the form (4.6), then the (continuous time) posterior distribution of α simplifies. Choose as the prior for α the conjugate prior for an exponential family of diffusions (see Küchler and Sørensen [41], page 51), which here is a multivariate normal distribution with expectation $\bar{\alpha}$ and covariance matrix Σ . Then the posterior of α (given $\beta = \beta_0$ and given simulated diffusion bridges) is a k -dimensional normal distribution with expectation $(\Sigma^{-1} + \tilde{\mathbf{B}}_{\beta_0})^{-1} (\Sigma^{-1} \bar{\alpha} + \tilde{\mathbf{H}}_{\beta_0})$ and covariance matrix $(\Sigma^{-1} + \tilde{\mathbf{B}}_{\beta_0})^{-1}$, where $\tilde{\mathbf{H}}_\beta = (\tilde{H}_{1,\beta}, \dots, \tilde{H}_{k,\beta})^T$, $\mathbf{B}_\beta = \{B_{i,j,\beta}\}$,

$$\begin{aligned}
 \tilde{H}_{i,\beta} &= s_{i,\beta}(x_n) - s_{i,\beta}(x_1) \\
 & + \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left[a_i(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) (\log \sigma_\beta)'(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) \right. \\
 & \qquad \qquad \qquad \left. - \frac{1}{2} a_i'(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) \right] dt,
 \end{aligned}$$

and

$$\tilde{B}_{i,j,\beta} = \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \frac{a_i(h_\beta^{-1}(Y_t^*(\beta, \beta_0))) a_j(h_\beta^{-1}(Y_t^*(\beta, \beta_0)))}{\sigma_\beta^2(h_\beta^{-1}(Y_t^*(\beta, \beta_0)))} dt.$$

4.2. Integrated diffusions observed with measurement error

Here we present an EM-algorithm to find the maximum likelihood estimator when an integrated diffusion is observed with measurement errors. The method was proposed and studied by Baltazar-Larios and Sørensen [4].

We consider again the diffusion process X given by (4.1), but here the data are of the form

$$V_i = \int_{t_{i-1}}^{t_i} X_s ds + Z_i, \quad i = 1, \dots, n, \quad (4.7)$$

where $Z_i \sim N(0, \tau^2)$, $i = 1, \dots, n$ are mutually independent and independent of X , and $t_0 = 0$. We assume that X is stationary and ergodic. The variance of the measurement error, τ^2 , is an extra unknown parameter, so we need to estimate the parameter $\theta = (\alpha, \beta, \tau^2)$. We can think of the data set $V = (V_1, \dots, V_n)$ as an incomplete observation of a full data set given by the sample path $X_t, t \in [0, t_n]$ and the measurement errors Z_1, \dots, Z_n , or equivalently $X_t, t \in [0, t_n]$ and $V = (V_1, \dots, V_n)$. To apply the EM-algorithm, we need to find the likelihood function for the full data set and the conditional expectation of this full log-likelihood function given the observations $V = (V_1, \dots, V_n)$.

Conditionally on the sample path of X , the observations $V_i, i = 1, \dots, n$ are independent and normal distributed with expectation $\int_{t_{i-1}}^{t_i} X_s ds$ and variance τ^2 . Again we need to apply the transformation (4.2) because the probability measures are singular. By expressing the data V_i in terms of the process Y , using the parameter-dependent transformation h_β , and by Girsanov's theorem, we find that the log-likelihood function for the full data set is

$$\begin{aligned} \ell(\theta; V_1, \dots, V_n, Y_t, t \in [0, t_n]) \\ = \sum_{i=1}^n \log \varphi \left(V_i; \int_{t_{i-1}}^{t_i} h_\beta^{-1}(Y_s) ds, \tau^2 \right) + g_{\alpha, \beta}(h_\beta^{-1}(Y_{t_n})) \\ - g_{\alpha, \beta}(h_\beta^{-1}(Y_0)) - \frac{1}{2} \int_0^{t_n} (\mu_{\alpha, \beta}(Y_t)^2 + \mu'_{\alpha, \beta}(U_t)) dt, \end{aligned} \quad (4.8)$$

where $\varphi(x; a, b)$ denotes the normal density function with mean a and variance b . The EM-algorithm works as follow. Let the initial value $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\tau}^2)$ be any value of the parameter vector θ .

(1) (E-step) Generate M sample paths of the diffusion process $X, X^{(k)}, k = 1, \dots, M$, conditional on the observations V_1, \dots, V_n using the parameter value $\hat{\theta}$, and calculate

$$g(\theta) = \frac{1}{M - M_0} \sum_{k=M_0+1}^M \ell(\theta; Y_1, \dots, Y_n, h_{\hat{\beta}}(X_t^{(k)}), t \in [0, t_n])$$

for a suitable burn-in period M_0 and M sufficiently large.

(2) (M-step) $\hat{\theta} = \operatorname{argmax} g(\theta)$.

(3) GO TO (1).

To implement this algorithm, the main issue is how to generate sample paths of X conditionally on V_1, \dots, V_n , where the relation between the V_i s and X is given by (4.7). This can be done by means of a Metropolis–Hastings algorithm. However, if the sample path in the entire time interval $[0, t_n]$ is updated in one step, the rejection probability is typically very large. Therefore,

it is more efficient to randomly divide the time interval into subintervals and update the sample path in each of the subintervals conditional on the rest of the sample path, which corresponds to simulating a diffusion bridge in each subinterval (except the end-intervals). This is a modification of the method in Chib *et al.* [16], where we use the algorithm for diffusion bridge simulation proposed in Section 2. In the following, the parameter value θ is fixed. Start by generating an initial unrestricted stationary sample path, $\{X_t^{(0)}: t \in [0, t_n]\}$, of the diffusion given by (4.1), and set $l = 1$.

(1) Generate a sample path $\{X_t^{(l)}: t \in [0, t_n]\}$ conditional on Y by updating subsets of the sample path:

- (a) Randomly split the time interval from $[0, t_n]$ into K blocks, and write these subsampling times as $0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_K = t_n$, where each τ_i is one of the end-points of the integration intervals, $t_j, j = 0, \dots, n$. Let $Y_{\{k\}}$ denote the collection of all observations Y_j for which $\tau_{k-1} < t_j \leq \tau_k$.
- (b) Draw $X_0^{(l)}$ from the stationary distribution, and simulate the conditional subpath $\{X_t^{(l)}: t \in [\tau_{k-1}, \tau_k]\} | Y_{\{k\}}, X_{\tau_{k-1}}^{(l)}, X_{\tau_k}^{(l-1)}$, for $k = 1, \dots, K - 1$. Finally, simulate $\{X_t^{(l)}: t \in [\tau_{K-1}, \tau_K]\} | Y_{\{K\}}, X_{\tau_{K-1}}^{(l)}$.
- (2) $l = l + 1$.
- (3) GO TO (1).

The random time intervals can for instance be generated by independent Poisson variables. Simulation of a $(\tau_{k-1}, a, \tau_k, b)$ -bridge conditional on $Y_{\{k\}}$, the data in $(\tau_{k-1}, \tau_k]$, can be done by a Metropolis–Hastings algorithm that uses the bridge simulation method introduced in Section 2 as proposal and accepts a proposed bridge $X^{(l)}$ with probability

$$\min \left(1, \prod_{i=1}^{n_k} \frac{\varphi(Y_{j+i}; \int_{t_{j+i-1}}^{t_{j+i}} X_s^{(l)} ds, \tau^2)}{\varphi(Y_{j+i}; \int_{t_{j+i-1}}^{t_{j+i}} X_s^{(l-1)} ds, \tau^2)} \right),$$

where the end-point τ_{k-1} is equal to t_j , n_k is the number of observations in the interval $(\tau_{k-1}, \tau_k]$ (the observations are $Y_{j+1}, \dots, Y_{j+n_k}$), and $X^{(l-1)}$ is the bridge in the previous step of the MH-algorithm.

In a simulation study by Baltazar-Larios and Sørensen [4], where the approximate algorithm of Section 2.1 was used to simulate diffusion bridges, this EM-algorithm worked well. The study considered 1500 observations of Ornstein–Uhlenbeck and CIR processes integrated over time intervals of length one. Three versions of the EM-algorithm were investigated, where the expected number of observations in each random subinterval was 11, 21 and 31, respectively.

5. Conclusion

We have presented a straightforward way of simulating an approximation to a diffusion bridge and an easily implementable Metropolis–Hastings algorithm that uses the approximate simulation as proposal and has exact diffusion bridges as the target distribution. Advantages of the

new method is that it is easy to understand and to implement, that the same simple algorithm can be used for all one-dimensional diffusion processes with finite speed measure, and most importantly, that the method works particularly well for long time intervals, where other methods tends not to work or to be very time consuming. The method allows the use of simple simulation procedures like the Milstein scheme for bridge simulation. The simulation study showed that the one-dimensional distributions obtained by the approximate method compare accurately to the results from exact simulations for bridges corresponding to data that are likely in discrete-time samples from diffusion models.

For ergodic diffusions, the computational complexity was shown to be linear in the interval length for our approximate method as well as for our exact method. The simulation study showed that the computing time for the approximate algorithm for small time intervals is of the same order of magnitude as for the exact EA1 method and for long intervals it is much faster than the EA1 method. Thus our new diffusion bridge simulation method is highly suitable for likelihood inference for discretely observed diffusions and can be used to simplify and in some cases speed up methods for likelihood inference and Bayesian inference (the EM-algorithm and the Gibbs sampler) for discretely observed diffusion processes. The method is also potentially useful for inference for more general diffusion type models like stochastic volatility models.

Acknowledgements

The authors are grateful to the referees for their insightful comments that have improved the paper very considerably. The research of Michael Sørensen was supported by the Danish Center for Accounting and Finance funded by the Danish Social Science Research Council, by the Center for Research in Econometric Analysis of Time Series funded by the Danish National Research Foundation, and by a grant from the University of Copenhagen Programme of Excellence.

References

- [1] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262. [MR1926260](#)
- [2] Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36** 906–937. [MR2396819](#)
- [3] Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. [MR2502648](#)
- [4] Baltazar-Larios, F. and Sørensen, M. (2010). Maximum likelihood estimation for integrated diffusion processes. In *Contemporary Quantitative Finance* 407–423. Berlin: Springer. [MR2732856](#)
- [5] Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.* **5** 151–157. [MR0509451](#)
- [6] Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. [MR2274855](#)
- [7] Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab.* **10** 85–104. [MR2394037](#)
- [8] Beskos, A., Papaspiliopoulos, O. and Roberts, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *Ann. Statist.* **37** 223–245. [MR2488350](#)

- [9] Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 333–382. With discussions and a reply by the authors. [MR2278331](#)
- [10] Bibby, B.M. and Sørensen, M. (2003). Hyperbolic processes in finance. In *Handbook of Heavy Tailed Distributions in Finance* (S. Rachev, ed.) 211–248. Amsterdam: Elsevier.
- [11] Bibby, B.M., Skovgaard, I.M. and Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* **11** 191–220. [MR2132002](#)
- [12] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39. [MR1354454](#)
- [13] Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 395–410. [MR2155345](#)
- [14] Bladt, M. and Sørensen, M. (2009). Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quant. Finance* **9** 147–160. [MR2512985](#)
- [15] Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Rev.* **11** 143–172. [MR1185178](#)
- [16] Chib, S., Pitt, M.K. and Shephard, N. (2010). Likelihood based inference for diffusion driven state space models. Working paper.
- [17] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19** 263–284. [MR0872464](#)
- [18] Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116** 1660–1675. [MR2269221](#)
- [19] 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. [MR0501537](#)
- [20] 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. [MR1939904](#)
- [21] Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. [MR1839375](#)
- [22] Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19** 177–191. [MR1939708](#)
- [23] Fitzsimmons, P., Pitman, J. and Yor, M. (1993). Markovian bridges: Construction, Palm interpretation, and splicing. In *Seminar on Stochastic Processes, 1992 (Seattle, WA, 1992)*. *Progress in Probability* **33** 101–134. Boston, MA: Birkhäuser. [MR1278079](#)
- [24] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20** 547–557. [MR1047222](#)
- [25] Forman, J.L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scand. J. Statist.* **35** 438–465. [MR2446729](#)
- [26] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079. [MR1809735](#)
- [27] Golightly, A. and Wilkinson, D.J. (2005). Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61** 781–788. [MR2196166](#)
- [28] Golightly, A. and Wilkinson, D.J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Stat. Comput.* **16** 323–338. [MR2297534](#)
- [29] Golightly, A. and Wilkinson, D.J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. [MR2422763](#)
- [30] Golightly, A. and Wilkinson, D.J. (2011). Bayesian parameter inference for stochastic biochemical network models using particle mcmc. *Interface Focus* **1** 807–820.

- [31] Hansen, L.P. and Scheinkman, J.A. (1995). Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* **63** 767–804. [MR1343081](#)
- [32] Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86** 1–32. [MR1650012](#)
- [33] Hurn, A., Jeisman, J.I. and Lindsay, K. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *J. Financial Econometrics* **5** 390–455.
- [34] Itô, K. and McKean, H.P. Jr. (1965). *Diffusion Processes and Their Sample Paths*. Berlin: Springer.
- [35] Karlin, S. (1988). Coincident probabilities and applications in combinatorics. *J. Appl. Probab.* **25A** 185–200. [MR0974581](#)
- [36] Karlin, S. and McGregor, J. (1959). Coincidence probabilities. *Pacific J. Math.* **9** 1141–1164. [MR0114248](#)
- [37] Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. [MR0611513](#)
- [38] Kent, J. (1978). Time-reversible diffusions. *Adv. in Appl. Probab.* **10** 819–835. [MR0509218](#)
- [39] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229. [MR1455868](#)
- [40] Kloeden, P.E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, 3rd revised printing. New York: Springer.
- [41] Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes. Springer Series in Statistics*. New York: Springer. [MR1458891](#)
- [42] Lin, M., Chen, R. and Mykland, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105** 820–838. [MR2724864](#)
- [43] Ozaki, T. (1985). Nonlinear time series models and dynamical systems. In *Time Series in the Time Domain* (E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds.). *Handbook of Statist.* **5** 25–83. Amsterdam: North-Holland. [MR0831744](#)
- [44] Papaspiliopoulos, O. and Roberts, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations* (M. Kessler, A. Lindner and M. Sørensen, eds.). *Monogr. Statist. Appl. Probab.* **124** 311–340. Boca Raton, FL: CRC Press. [MR2976985](#)
- [45] Pedersen, A.R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* **22** 55–71. [MR1334067](#)
- [46] Poulsen, R. (1999). Approximate maximum likelihood estimation of discretely observed diffusion processes. Working Paper 29, Centre for Analytical Finance, Aarhus.
- [47] Roberts, G.O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88** 603–621. [MR1859397](#)
- [48] Sørensen, M. (2010). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint, Dept. Mathematical Sciences, Univ. Copenhagen.

Received April 2011 and revised November 2012