Stat Comput (2012) 22:205–217
DOI 10.1007/s11222-010-9218-8

Approximation of transition densities of stochastic differential equations by saddlepoint methods applied to small-time Ito–Taylor sample-path expansions

S.P. Preston · Andrew T.A. Wood

Received: 18 November 2009 / Accepted: 8 November 2010 / Published online: 2 February 2011

Abstract Likelihood-based inference for parameters of stochastic differential equation (SDE) models is challenging because for most SDEs the transition density is unknown. We propose a method for estimating the transition density that involves expanding the sample path as an Ito–Taylor series, calculating the moment generating function of the retained terms in the Ito–Taylor expansion, then employing a saddlepoint approximation. We perform a numerical comparison with two other methods similarly based on small-time expansions and discuss the pros and cons of our new method relative to other approaches.

Keywords Ito–Taylor expansion · Saddlepoint approximation · Stochastic differential equation · Transition density

1 Introduction

This paper is motivated by the problem of making inference about parameters of continuous-time (but discretely observed) time-homogeneous diffusion processes. These are specified by the Ito stochastic differential equation (SDE),

$$\mathrm{d}X_t = \mu(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\theta)\,\mathrm{d}W_t, \tag{1}$$

where W_t is a Brownian motion, θ is a vector of unknown parameters, and μ and σ are functions that respectively define the instantaneous drift and diffusion. In practice one usually specifies a model by defining μ and σ, and then calibrates the model by choosing θ in accordance with a set of discrete observations, say $\{X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}\}$, of the process. There is no problem in principle in extending our approach to time-inhomogeneous diffusion processes in which μ and σ depend on t, but for simplicity we do not consider this possibility here.

The Markov property of (1) means that the log-likelihood function for θ has the simple form,

$$l(\theta|x_{t_1},\ldots,x_{t_n}) = \sum_{i=2}^{n} \log p(x_{t_i}|x_{t_{i-1}},\theta),$$

where $p(x_{t_i}|x_{t_{i-1}},\theta)$ is the transition density, defined by

$$p(x_{t_i}|x_{t_{i-1}},\theta) = \frac{\mathrm{d}}{\mathrm{d}x_{t_i}} \mathrm{Prob}\{X_{t_i} \le x_{t_i} \mid X_{t_{i-1}} = x_{t_{i-1}},\theta\}.$$

However, one major difficulty with employing a maximum-likelihood approach in practise is that for most specifications of (1) the transition density is unknown. (There are a few exceptions, including the Ornstein–Uhlenbeck/Vasicek, Cox–Ingersoll–Ross/Feller, Bessel, and Black-Scholes processes —see Table 1 for the definitions of these processes and other common ones; see also Kloeden and Platen 1992.) When the transition density is not known explicitly, there are various methods by which it can be calculated approximately. These methods broadly fall into three categories according to whether they involve i) solving the partial differential equation that governs the transition density (e.g. Lindström 2007); ii) simulation (including the approach of Pedersen (1995) refined later by Durham and Gallant (2002), and the exact simulation methods of Beskos et al. (2006)); or iii) replacing the process with a discrete version, valid for small time intervals, for which the transition density can be approximately calculated (examples of which include the

S.P. Preston (✉) · A.T.A. Wood
Division of Statistics, School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK
e-mail: Simon.Preston@nottingham.ac.uk

A.T.A. Wood
e-mail: Andrew.Wood@nottingham.ac.uk

methods of Shoji and Ozaki (1998) and Aït-Sahalia (1999, 2002), which we discuss in Sect. 2).

The focus of this paper is on introducing a new approach that falls into iii), and on comparing this new approach with existing methods in the category. Our approach is to use an Ito–Taylor expansion of the sample path, calculate the moment generating function of the expansion, and then use saddlepoint techniques to compute the approximate transition density.

The use of saddlepoint approximations in statistics originated from work by Daniels (1954); for a recent monograph on the topic, see Butler (2007). The basic idea is to apply the method of steepest descent to the inversion integral for the moment generating function (MGF), $M(s)$, given by

$$M(s) = \exp\{K(s)\} = \int_{-\infty}^{\infty} \exp(sx) f(x)\,\mathrm{d}x,$$

where $f(x)$ is the relevant probability density function and $K(s)$ is the cumulant generating function (CGF). For a saddlepoint approximation to be possible we require that $M(s)$ be finite for all s in an open interval $\mathcal{S}$ containing zero. The inversion integral is

$$f(x) = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} \exp\{K(s) - sx\}\,\mathrm{d}s, \tag{2}$$

for any $\tau \in \mathcal{S}$. The exponent in (2), when viewed as a function of complex variable s, has a saddlepoint at the solution $\hat{s} = \hat{s}(x)$ of the so-called saddlepoint equation

$$K'(\hat{s}) = x. \tag{3}$$

Solution $\hat{s}$, called the saddlepoint, is the unique solution of (3) in $\mathcal{S}$ due to the convexity of $K(\hat{s})$ in $\mathcal{S}$ (Daniels 1954). Choosing the path of integration in (2) to pass through the saddlepoint, *i.e.*, setting $\tau = \hat{s}$, and replacing $\{K(s) - sx\}$ with its Taylor series expansion about $\hat{s}$ leads to

$$f(x) = \exp\{K(\hat{s}) - \hat{s}x\} \frac{1}{2\pi i} \int_{\hat{s}-i\infty}^{\hat{s}+i\infty} \exp\left\{\frac{1}{2}(s-\hat{s})^2 K''(\hat{s}) + O\{(s-\hat{s})^3\}\right\} \mathrm{d}s.$$

The leading-order saddlepoint approximation, $\hat{f}(x)$, to the transition density, $f(x)$, which comes from neglecting the $O(s^3)$ terms and evaluating the integral, is

$$\hat{f}(x) = \left(2\pi K''(\hat{s})\right)^{-1/2} \exp\{K(\hat{s}) - \hat{s}x\}. \tag{4}$$

A second-order approximation, which involves using a higher-order Taylor expansion and neglecting terms of order $O(s^5)$, is also possible, but in the context of our approach in this paper we found it to offer little improvement in accuracy over the leading-order approximation (4), so we do not discuss it further here.

In an independent repeated-sampling framework, in the case of (say) a sample mean, the error in the density approximation (4) is $O(n^{-1})$, and $O(n^{-2})$ in the case of the second-order approximation, where n is the sample size. However, the most remarkable property of saddlepoint approximations is that they achieve bounded relative error across the whole range of the distribution in many situations, with fixed $n = 1$. In the present setting we do not have independent repeated sampling, but as the time interval goes to zero the transition density converges to a Gaussian distribution, and the density approximation becomes exact. Hence there is good motivation for investigating saddlepoint approximations in this context.

Saddlepoint methods have previously been used in the context of estimating the transition densities of diffusions, namely by Aït-Sahalia and Yu (2006) who applied the saddlepoint approximation to the characteristic function of the transition density. Where the characteristic function is not available in closed form, Ait-Sahalia and Yu approximated it using a truncated series expansion. Their method is distinct from our method in this paper in which we apply the saddlepoint approximation to the CGF of a truncated sample-path expansion.

2 Review of small-time expansion methods

In this section we briefly summarise the methods proposed by Shoji and Ozaki (1998) and Aït-Sahalia (1999, 2002) (henceforth abbreviated by SO and AS, respectively) for estimating transition densities of (1). In common with the new methods that we introduce in this paper, both SO and AS are based on truncated small-time expansions. Both approaches make use of a change of variable that transforms (1) into a process with unit diffusion coefficient. This change of variable is

$$Y_t = \int^{X_t} \mathrm{d}u/\sigma(u;\theta) = \gamma(X_t;\theta), \tag{5}$$

from which it follows by applying Ito's formula that Y_t satisfies

$$\mathrm{d}Y_t = \mu_Y(Y_t;\theta)\,\mathrm{d}t + \mathrm{d}W_t, \tag{6}$$

where

$$\mu_Y(Y_t;\theta) = \frac{\mu(\gamma^{-1}(Y_t;\theta);\theta)}{\sigma(\gamma^{-1}(Y_t;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(Y_t;\theta);\theta),$$

and $\gamma^{-1}(\cdot;\theta)$ is the functional inverse of $\gamma(\cdot;\theta)$ for fixed θ. For models in which γ cannot be calculated in closed form, the transform can be performed numerically; see for instance Stramer and Yan (2007).

Table 1 Specifications of common stochastic processes of the form $\mathrm{d}X_t = \mu(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\theta)\,\mathrm{d}W_t$, and functions $\gamma(u;\theta)$ and $\mu_Y(u;\theta)$ such that $Y_t = \gamma(X_t;\theta)$ satisfies $\mathrm{d}Y_t = a(Y_t;\theta)\,\mathrm{d}t + \mathrm{d}W_t$

Name	$\mu(X_t;\theta)$	$\sigma(X_t;\theta)$	$\gamma(u;\theta)$	$a(u;\theta)$
Ornstein–Uhlenbeck	$\kappa(\alpha - X_t)$	β	u/β	$\kappa(\alpha\beta^{-1} - u)$
Cox–Ingersoll–Ross	$\kappa(\alpha - X_t)$	$\beta\sqrt{X_t}$	$2\beta^{-1}\sqrt{u}$	$(2\kappa\alpha\beta^{-2} - \frac{1}{2})u^{-1} - \frac{1}{2}\kappa u$
Bessel	κ	$2\sqrt{X_t}$	$\sqrt{u}$	$(\kappa - 1)(2u)^{-1}$
Black–Scholes	κX_t	βX_t	$\beta^{-1}\ln u$	$\kappa\beta^{-1} - \frac{1}{2}\beta$
Sine (periodic drift)	$\sin(X_t - \alpha)$	1	–	–
Logistic growth	$\kappa X_t(1 - X_t/\alpha)$	βX_t	$-\beta^{-1}\ln u$	$\frac{1}{2}\beta - \kappa\beta^{-1} + \kappa\alpha^{-1}\beta^{-1}\exp(-\beta u)$
Generalised CIR	$\kappa(\alpha - X_t)$	βX_t^{ρ}	$\beta^{-1}(\rho+1)^{-1}u^{\rho+1}$	$\beta^{-1}\kappa(\alpha - z)z^{\rho} - \frac{1}{2}\beta\rho z^{\rho-1}$ where $z = (\beta(\rho+1)u)^{1/(\rho+1)}$
Ginzburg–Landau	$-\kappa X_t - \alpha X_t^3$	β	u/β	$-\kappa u - \alpha u^3\beta^2$

Equation (6) is the starting point for both SO and AS. SO involves linearising the drift coefficient, via an application of Ito's lemma, to give an approximate governing equation for which the transition density can be calculated in closed form. Ito's lemma applied to the drift coefficient (which is in general allowed to depend on t, i.e., $\mu_Y = \mu_Y(y,t,\theta)$), is

$$\mathrm{d}\mu_Y = \frac{\partial\mu_Y}{\partial t}\mathrm{d}t + \frac{\partial\mu_Y}{\partial y}\mathrm{d}y + \frac{1}{2}\frac{\partial^2\mu_Y}{\partial y^2}\mathrm{d}t,$$

leading to the linearised version,

$$\begin{aligned}&\mu_Y(Y_{t_i},t_i) - \mu_Y(Y_{t_{i-1}},t_{i-1})\\ &\quad\approx \left(\frac{1}{2}\frac{\partial^2\mu_Y}{\partial y^2} + \frac{\partial\mu_Y}{\partial t}\right)\Big|_{(Y_{t_{i-1}},t_{i-1})}(t_i - t_{i-1})\\ &\qquad + \frac{\partial\mu_Y}{\partial y}\Big|_{(Y_{t_{i-1}},t_{i-1})}(Y_{t_i} - Y_{t_{i-1}}).\end{aligned}$$

Substituting this into (6), the governing equation for the linearised process is

$$\mathrm{d}Y_{t_i} = (C_1 Y_{t_i} + C_2 t_i + C_3)\mathrm{d}t_i + \mathrm{d}W_{t_i}, \tag{7}$$

where

$$\begin{aligned}C_1 &= \frac{\partial\mu_Y}{\partial y}\Big|_{(Y_{t_{i-1}},t_{i-1})},\\ C_2 &= \left(\frac{1}{2}\frac{\partial^2\mu_Y}{\partial y^2} + \frac{\partial\mu_Y}{\partial t}\right)\Big|_{(Y_{t_{i-1}},t_{i-1})},\\ C_3 &= \mu_Y(Y_{t_{i-1}},t_{i-1}) - C_2 t_{i-1} - C_1 Y_{t_{i-1}}.\end{aligned}$$

Equation (7) can be transformed to an Ornstein–Uhlenbeck process by using Girsanov's theorem and thus solved exactly to give

$$\begin{aligned}Y_{t_i} &= Y_{t_{i-1}} + \frac{\mu_Y(Y_{t_{i-1}},t_{i-1},\theta)}{C_1}(\exp\{C_1 T\} - 1)\\ &\quad + \frac{C_2}{(C_1)^2}(\exp\{C_1 T\} - 1 - C_1 T)\\ &\quad + \int_{t_{i-1}}^{t_i}\exp\{C_1(t_i - u)\}\mathrm{d}W_u,\end{aligned}$$

where $T = t_i - t_{i-1}$. The integral term is Gaussian with mean zero and variance $(\exp\{2C_1T\} - 1)/(2C_1)$.

The AS approach (Aït-Sahalia 1999, 2002) for a process Y_t satisfying (6), involves working with the transformed variable

$$Z = T^{-1/2}(Y_{t_i} - Y_{t_{i-1}}),$$

whose density in the limit $T \to 0$ is a standard normal. For finite T one can write the density as a series expansion involving Hermite polynomials, defined by $H_j(z) = \exp(z^2/2)\mathrm{d}^j/\mathrm{d}z^j\{\exp(-z^2/2)\}$, whose role is to correct for non-normality. The density of Z can be expanded as

$$p_Z(z) = \phi(z)\sum_{j=1}^{J}\eta_z^{(j)}(T, y_{t_{i-1}};\theta)H_j(z),$$

where ϕ is the standard normal density, and $\eta_z^{(j)}$ are the Hermite coefficients. Then expanding the $\eta_z^{(j)}$ as a Taylor series in T and gathering the resulting Taylor and Hermite terms in ascending powers of T leads to the series expression for the density of Y_{t_i},

$$\begin{aligned}p_{Y_{t_i}}(y_{t_i}|y_{t_{i-1}},\theta) &= T^{-1/2}\phi\left(\frac{y_{t_i} - y_{t_{i-1}}}{T^{1/2}}\right)\\ &\quad\times\exp\left\{\int_{y_{t_{i-1}}}^{y_{t_i}}\mu_Y(w,\theta)\,\mathrm{d}w\right\}\\ &\quad\times\sum_{k=0}^{K}c_k(y_{t_i}|y_{t_{i-1}},\theta)\frac{T^k}{k!},\end{aligned} \tag{8}$$

where $c_0(y_{t_i}|y_{t_{i-1}},\theta)=1$, and

$$c_k(y_{t_i}|y_{t_{i-1}},\theta)=k(y_{t_i}-y_{t_{i-1}})^{-k}\int_{y_{t_{i-1}}}^{y_{t_i}}(w-y_{t_{i-1}})^{k-1}\times\{\lambda(w,\theta)\,c_{k-1}(w|y_{t_{i-1}},\theta)+(\partial^2 c_{k-1}(w|y_{t_{i-1}},\theta)/\partial w^2)/2\}\,\mathrm{d}w, \quad (9)$$

for $k\geq 1$, where

$$\lambda(w,\theta)=-([\mu_Y(w,\theta)]^2+\partial\mu_Y(w,\theta)/\partial w)/2.$$

In Sect. 4 we compare AS and SO with the new methods that we describe in the following section.

3 Ito–Taylor expansions and discretisation schemes

The Ito–Taylor expansion of (1) is obtained (see Kloeden and Platen 1992, Ch. 5) by iteratively applying Ito's formula, which is defined on the interval $[t_0,t]$ by

$$f(X_t)=f(X_{t_0})+\int_{t_0}^t L^0 f(X_s)\,\mathrm{d}s+\int_{t_0}^t L^1 f(X_s)\,\mathrm{d}W_s, \quad (10)$$

where

$$L^0=\mu\frac{\partial}{\partial x}+\frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}\quad\text{and}\quad L^1=\sigma\frac{\partial}{\partial x}.$$

Thus applying (10) to (1) gives

$$\begin{aligned}X_t=X_{t_0}&+\mu(X_{t_0})\int_{t_0}^t\mathrm{d}s+\sigma(X_{t_0})\int_{t_0}^t\mathrm{d}W_s\\&+\Big\{\int_{t_0}^t\int_{t_0}^s L^0\mu(X_z)\,\mathrm{d}z\,\mathrm{d}s\\&+\int_{t_0}^t\int_{t_0}^s L^1\mu(X_z)\,\mathrm{d}W_z\,\mathrm{d}s\\&+\int_{t_0}^t\int_{t_0}^s L^0\sigma(X_z)\,\mathrm{d}z\,\mathrm{d}W_s\\&+\int_{t_0}^t\int_{t_0}^s L^1\sigma(X_z)\,\mathrm{d}W_z\,\mathrm{d}W_s\Big\}. \quad (11)\end{aligned}$$

Neglecting the remainder term (shown in braces) in (11) gives the simple Euler–Murayama approximation, whereas recursively applying (10) to the functions $\mu(X_z)$ and $\sigma(X_z)$ in (11) leads to higher-order approximations. These expressions involve combinations of Brownian functionals for which the density is not in general known in closed form. This motivates our use of saddlepoint techniques which we will show give very accurate approximations to the density.

The new approaches we introduce here are based on Ito–Taylor expansions of various orders. It is convenient to adopt the notation used by Kloeden and Platen (1992) of writing I_α, where α is a multi-index (vector) whose components are 0 or 1, to denote an integral over (t_0,t) of a unit integrand with respect to t (denoted by suffix 0) and/or W_t (denoted by suffix 1). Hence, for instance, in this notation, $I_{(1,0,0)}=\int_{t_0}^t\int_{t_0}^{s_3}\int_{t_0}^{s_2}\mathrm{d}W_{s_1}\mathrm{d}s_2\,\mathrm{d}s_3$. The schemes we discuss in this section involve the following integrals:

$$\begin{aligned}&I_{(0)}=T,\quad I_{(1)}=J_1,\quad I_{(0,0)}=\tfrac{1}{2}T^2,\quad I_{(1,0)}=J_2,\\&I_{(0,1)}=TJ_1-J_2,\quad I_{(1,1)}=\tfrac{1}{2}(J_1^2-T),\\&I_{(0,0,0)}=\tfrac{1}{6}T^3,\quad I_{(1,0,0)}=TJ_2-J_3,\\&I_{(0,1,0)}=2J_3-TJ_2,\quad I_{(1,1,0)}=\tfrac{1}{2}J_4-\tfrac{1}{4}T^2,\end{aligned}$$

where $T=t-t_0$ and

$$\begin{aligned}&J_1=W_T,\quad J_2=\int_0^T W_s\,\mathrm{d}s,\\&J_3=\int_0^T sW_s\,\mathrm{d}s,\quad J_4=\int_0^T W_s^2\,\mathrm{d}s.\end{aligned} \quad (12)$$

Note that the mean and covariance of (J_1,J_2,J_3,J_4) are

$$(0,0,0,T^2/2)\quad\text{and}\quad\begin{pmatrix}T & T^2/2 & T^3/3 & 0\\ & T^3/3 & 5T^4/24 & 0\\ & & 2T^5/15 & 0\\ & & & T^4/3\end{pmatrix},$$

respectively, and that J_1, J_2, J_3 are Gaussian but J_4 is non-Gaussian.

With Ito–Taylor schemes there are two senses of convergence, weak and strong (Kloeden and Platen 1992). In this paper the strong sense, defined as follows, is relevant. Let X_t denote the diffusion process of interest. Write $X_{0,t}$ for an Ito–Taylor sample path approximation to X_t, and suppose that $X_{0,t_0}=X_{t_0}$ at an initial time point t_0. We say that $X_{0,t}$ converges strongly to X_t with order γ if, for all t_0 and some constants $C>0$ and $\delta>0$,

$$\sqrt{E[\{X_{0,t}-X_t\}^2]}\leq C(t-t_0)^{\gamma/2}$$

for all $t\in(t_0,t_0+\delta)$. See, in particular, Kloeden and Platen (1992, p. 206).

Kloeden and Platen (1992, Chap. 5) give the conditions under which an Ito–Taylor expansion attains a given order of strong convergence. For each integral I_α that arises in the Ito–Taylor expansion, let $l(\alpha)$ denote the number of components in α, and let $n(\alpha)$ denote the number of zero components. In the case where X_t is one-dimensional, the $l(\alpha)-n(\alpha)$ non-zero components are all ones. Assuming that all the drift and diffusion coefficients are sufficiently smooth, a scheme attains strong convergence of order γ if it includes all terms with α satisfying $l(\alpha)+n(\alpha)\leq 2\gamma$.

Below are schemes based on various truncations of Ito–Taylor expansions. We use a and b to denote, respectively, μ and σ evaluated at the left-hand point of the interval, and primes to indicate derivatives; for example, $a' = (\partial^2\mu/\partial x^2)|_{x=X_{t_0}}$. The theoretical properties of Schemes 1–4 described below are sketched in Appendix but without giving detailed or rigorous proofs.

3.1 Scheme 1 (Euler): $b = \text{const}$, $\gamma = 1.0$; $b \neq \text{const}$, $\gamma = 0.5$

The Euler scheme is

$$X_t = X_{t_0} + aI_{(0)} + bI_{(1)},$$

for which the transition density is Gaussian with mean $X_{t_0} + aT$ and variance b^2T.

3.2 Scheme 2 (Milstein): $b = \text{const}$, $\gamma = 1.0$; $b \neq \text{const}$, $\gamma = 1.0$

The Milstein scheme is

$$X_t = X_{t_0} + aI_{(0)} + bI_{(1)} + bb'I_{(1,1)},$$

which can be written

$$X_t - X_{t_0} - c_3 = c_1J_1 + c_2J_1^2, \tag{13}$$

where

$$c_1 = b, \qquad c_2 = \tfrac{1}{2}bb', \qquad c_3 = (a - \tfrac{1}{2}bb')T.$$

The MGF of the right-hand side of (13) is

$$M(s) = (1 - 2sc_2T)^{-1/2}\exp\left\{\frac{c_1^2s^2T}{2 - 4sc_2T}\right\},$$

and the roots of the saddlepoint equation are

$$\frac{-2{c_2}^2T + {c_1}^2 + 4xc_2 \pm \sqrt{4{c_2}^4T^2 + {c_1}^4 + 4{c_1}^2xc_2}}{2c_2({c_1}^2 + 4xc_2)T}.$$

If these roots are real then the relevant root is either the negative one if $c_2 > 0$ or the positive one if $c_2 < 0$. However, the support of the density of (13) is bounded in one tail; for this reason, the Milstein scheme is a poor choice for inference, and the following scheme is preferable.

3.3 Scheme 3: $b = \text{const}$, $\gamma = 1.5$; $b \neq \text{const}$, $\gamma = 1.0$

This scheme is

$$\begin{aligned} X_t = X_{t_0} &+ aI_{(0)} + bI_{(1)} + bb'I_{(1,1)} \\ &+ \left(aa' + \frac{1}{2}b^2a''\right)I_{(0,0)} + \left(ab' + \frac{1}{2}b^2b''\right)I_{(0,1)} \\ &+ ba'I_{(1,0)}, \end{aligned}$$

or, alternatively written,

$$X_t - X_{t_0} - c_4 = c_1J_1 + c_2J_1^2 + c_3J_2, \tag{14}$$

where

$$c_1 = b + \left(ab' + \frac{1}{2}b^2b''\right)T, \qquad c_2 = \frac{1}{2}bb',$$

$$c_3 = ba' - ab' - \frac{1}{2}b^2b'',$$

$$c_4 = \left(a - \frac{1}{2}bb'\right)T + \left(aa' + \frac{1}{2}b^2a''\right)\frac{1}{2}T^2.$$

After some elementary calculations we find that the MGF of the right hand side of (14) is

$$\begin{aligned} &(1 - 2sc_2T)^{-1/2} \\ &\quad\times \exp\left\{\frac{(6{c_1}^2 + 6c_1c_3T + 2{c_3}^2T^2 - {c_3}^2T^3sc_2)Ts^2}{12(1 - 2sc_2T)}\right\}. \end{aligned}$$

Hence the saddlepoint equation is a cubic in s:

$$\begin{aligned} &\frac{1}{3}c_2^2c_3^2T^5s^3 \\ &\quad - \left(c_1c_2c_3T^3 + c_1^2c_2T^2 + \frac{7}{12}c_2c_3^2T^4 + 4xc_2^2T^2\right)s^2 \\ &\quad + \left(-2c_2^2T^2 + c_1c_3T^2 + c_1^2T + \frac{1}{3}c_3^2T^3 + 4xc_2T\right)s \\ &\quad + c_2T - x = 0. \end{aligned}$$

The saddlepoint $\hat{s}$ is the unique solution of the cubic in the interval $(-\infty, (2c_2T)^{-1})$, and is straightforward to find numerically.

When $b = \text{const}$, the transition density for this scheme is Gaussian with mean $aT + (aa' + \frac{1}{2}b^2a'')\frac{1}{2}T^2$ and variance $(b^2 + b^2a'T + \frac{1}{3}b^2(a')^2T^2)T$.

3.4 Scheme 4: $b = \text{const}$, $\gamma = 2.0$

For this scheme we limit attention to the case with $b = \text{const}$. Without loss of generality we can take $b = 1$ and write the scheme as

$$\begin{aligned} X_t = X_{t_0} &+ aI_{(0)} + I_{(1)} + \left(aa' + \frac{1}{2}a''\right)I_{(0,0)} + a'I_{(1,0)} \\ &+ \left(a\left(aa'' + (a')^2 + \frac{1}{2}a'''\right)\right. \\ &\left. + \frac{1}{2}(aa''' + 3a'a'') + \frac{1}{4}a''''\right)I_{(0,0,0)} \\ &+ \left(aa'' + (a')^2 + \frac{1}{2}a'''\right)I_{(1,0,0)} \end{aligned}$$

$$+\left(aa''+\frac{1}{2}a'''\right)I_{(0,1,0)}+a''I_{(1,1,0)},$$

or,

$$X_t - X_{t_0} - c_5 = c_1 J_1 + \cdots + c_4 J_4,$$

where

$$c_1 = 1, \qquad c_2 = a' + (a')^2 T,$$
$$c_3 = aa'' + \frac{1}{2}a''' - (a')^2 T, \qquad c_4 = \frac{1}{2}a'',$$
$$c_5 = aT + \frac{1}{2}aa'T^2 + \frac{1}{3}\left(a\left(aa'' + (a')^2 + \frac{1}{2}a'''\right) + \frac{1}{2}(aa''' + 3a'a'') + \frac{1}{4}a''''\right)T^3.$$

The MGF for this scheme, which is more complicated due to the non-Gaussian functional J_4 that arises from the $I_{(1,1,0)}$ term, can be calculated as follows. First, let us represent Brownian motion $\{W_t : t \in [0,T]\}$ (see Shepp 1982) by

$$W_t = \sum_{n=0}^{\infty} \eta_n \int_0^t \phi_n(s)\,\mathrm{d}s, \tag{15}$$

where $(\eta_n)_{n\geq 0}$ are IID $N(0,1)$ random variables and

$$\phi_0(s) = \frac{1}{\sqrt{T}}, \qquad \phi_n(s) = \sqrt{\frac{2}{T}}\cos\left(\frac{n\pi s}{T}\right) \quad \text{for } n \geq 1.$$

Then, substituting (15) into each term in (12) and integrating, we obtain

$$J_1(T) = \eta_0\sqrt{T},$$
$$J_2(T) = \frac{1}{2}T^{3/2}\eta_0 + 2\sqrt{2}\,T^{3/2}\pi^{-2}\sum_{n=1}^{\infty}\eta_{2n-1}(2n-1)^{-2},$$
$$J_3(T) = \frac{1}{3}T^{5/2}\eta_0 + \sqrt{2}\,T^{5/2}\pi^{-2}\sum_{n=1}^{\infty}\eta_n(-1)^{n+1}n^{-2},$$
$$J_4(T) = \frac{1}{3}T^2\eta_0^2 + T^2\pi^{-2}\sum_{n=1}^{\infty}\eta_n^2 n^{-2} + 2\sqrt{2}\,T^2\pi^{-2}\eta_0\sum_{n=1}^{\infty}\eta_n(-1)^{n+1}n^{-2}.$$

Before calculating the MGF $M_J(\xi) = \mathbb{E}\exp(J^{\mathrm{T}}\xi)$ (where $J = (J_1,\ldots,J_4)^{\mathrm{T}}$ and $\xi = (\xi_1,\ldots,\xi_4)^{\mathrm{T}}$), we first calculate the conditional MGF,

$$M_{J|\eta_0}(\xi) = \exp(C_1)\mathbb{E}\exp\left\{\sum_{n=1}^{\infty}\eta_{2n-1}a_n + \eta_{2n-1}^2 b_n + \sum_{n=1}^{\infty}\eta_{2n}c_n + \eta_{2n}^2 d_n\right\},$$

where

$$C_1 = \eta_0\left(\xi_1 T^{1/2} + \xi_2\frac{1}{2}T^{3/2} + \xi_3\frac{1}{3}T^{5/2}\right) + \eta_0^2\xi_4\frac{1}{3}T^2,$$
$$a_n = \pi^{-2}(2n-1)^{-2}(\xi_2 2\sqrt{2}\,T^{3/2} + \xi_3\sqrt{2}T^{5/2} + \xi_4\eta_0 2\sqrt{2}\,T^2),$$
$$b_n = \xi_4 T^2\pi^{-2}(2n-1)^{-2},$$
$$c_n = -\pi^{-2}(2n)^{-2}(\xi_3\sqrt{2}\,T^{5/2} + \xi_4\eta_0 2\sqrt{2}\,T^2),$$
$$d_n = \xi_4 T^2\pi^{-2}(2n)^{-2}.$$

Note that for $Z \sim N(0,1)$,

$$\mathbb{E}\exp(\alpha Z + \beta Z^2) = (1-2\beta)^{-1/2}\exp(\alpha^2[2(1-2\beta)]^{-1}); \tag{16}$$

thus

$$M_{J|\eta_0}(\xi) = \exp(C_1)\left(\prod_{n=1}^{\infty}(1-2b_n)^{-1/2}\right) \times \exp\left(\sum_{n=1}^{\infty}a_n^2[2(1-2b_n)]^{-1}\right) \times \left(\prod_{n=1}^{\infty}(1-2d_n)^{-1/2}\right) \times \exp\left(\sum_{n=1}^{\infty}c_n^2[2(1-2d_n)]^{-1}\right). \tag{17}$$

Furthermore, using the infinite product representation for the sine function (see for example Pennisi 1976, p. 449), we find that

$$\left(\prod_{n=1}^{\infty}(1-2b_n)^{-1/2}\right)\times\left(\prod_{n=1}^{\infty}(1-2d_n)^{-1/2}\right) = \prod_{n=1}^{\infty}(1-C_2\pi^{-2}n^{-2})^{-1/2} = (C_2^{-1/2}\sin\sqrt{C_2})^{-1/2},$$

where $C_2 = \xi_4 2T^2$. Then, using the Mittag-Leffler theorem (see for example Pennisi 1976, pp. 334–336), we obtain

$$\sum_{n=1}^{\infty}a_n^2[2(1-2b_n)]^{-1} = (\xi_2 2T^{3/2} + \xi_3 T^{5/2} + \xi_4\eta_0 2T^2)^2 C_3,$$
$$\sum_{n=1}^{\infty}c_n^2[2(1-2d_n)]^{-1} = (\xi_3 T^{5/2} + \xi_4\eta_0 2T^2)^2 C_4,$$

where

$$C_3 = \frac{1}{4}C_2^{-3/2}\tan\left(\frac{1}{2}\sqrt{C_2}\right) - \frac{1}{8}C_2^{-1},$$

$$C_4 = \frac{1}{2}C_2^{-2} - \frac{1}{24}C_2^{-1} - \frac{1}{4}C_2^{-3/2}\cot\left(\frac{1}{2}\sqrt{C_2}\right).$$

To remove the conditioning on η_0 we collect the terms in the exponent of (17) in powers of η_0 and then apply (16) again. Hence

$$\begin{aligned} M_J(\xi) &= \left(C_2^{-1/2}\sin\sqrt{C_2}\right)^{-1/2}\exp(C_5)\mathbb{E} \\ &\quad \times \exp(C_6\eta_0 + C_7\eta_0^2), \\ &= \left((1-2C_7)C_2^{-1/2}\sin\sqrt{C_2}\right)^{-1/2} \\ &\quad \times \exp\left(C_5 + C_6^2[2(1-2C_7)]^{-1}\right), \\ &= C_8^{1/2}\left(C_2^{-1/2}\sin\sqrt{C_2}\right)^{-1/2} \\ &\quad \times \exp\left(C_5 + \frac{1}{2}C_6^2C_8\right), \end{aligned}$$

where

$$C_5 = (\xi_2 2T^{3/2} + \xi_3 T^{5/2})^2 C_3 + \xi_3^2 T^5 C_4,$$

$$\begin{aligned} C_6 &= \left(\xi_1 T^{1/2} + \xi_2\frac{1}{2}T^{3/2} + \xi_3\frac{1}{3}T^{5/2}\right) \\ &\quad + \xi_4 4T^2(\xi_2 2T^{3/2} + \xi_3 T^{5/2})C_3 + \xi_3\xi_4 4T^{9/2}C_4, \end{aligned}$$

$$C_7 = \xi_4\frac{1}{3}T^2 + \xi_4^2 4T^4(C_3 + C_4),$$

$$\begin{aligned} C_8 &= (1-2C_7)^{-1} \\ &= 2C_2^{-1/2}\left(\cot\left(\frac{1}{2}\sqrt{C_2}\right) - \tan\left(\frac{1}{2}\sqrt{C_2}\right)\right)^{-1}, \end{aligned}$$

and then finally the CGF is

$$\begin{aligned} K_J(\xi) = \ln M_J(\xi) &= \frac{1}{4}\ln C_2 - \frac{1}{2}\ln\left(\sin\sqrt{C_2}\right) \\ &\quad + \frac{1}{2}\ln C_8 + C_5 + \frac{1}{2}C_6^2C_8. \end{aligned}$$

3.5 Higher-order schemes

It is possible in principle to retain more terms of the Ito–Taylor expansion. However, schemes with $\gamma > 2.0$ involve functionals cubic in Brownian motion that do not have finite MGFs. Hence it appears that for the strategy we have adopted in this paper, Scheme 4 is the practical limit.

4 Numerical comparison of methods

To examine the performance of the methods of Sect. 3 we tested them against established methods for two common benchmark models whose exact density is known, namely the Ornstein–Uhlenbeck and Cox–Ingersoll–Ross models. To measure accuracy we used the absolute error of the log density (AELD), defined as

$$\mathrm{AELD}(x_t|x_0,\theta) = |\ln\hat{p}(x_t|x_0,\theta) - \ln p(x_t|x_0,\theta)|,$$

where p is the exact density and $\hat{p}$ the corresponding estimate. In addition, we use the mean absolute error of the log density (MAELD), defined as

$$\mathrm{MAELD}(x_0,\theta) = \int \mathrm{AELD}(x_t|x_0,\theta)\, p(x_t|x_0,\theta)\,\mathrm{d}x_t.$$

Results are shown in Figs. 1–4. We use the suffix `_t` to indicate where the transform (5) is used, and `_u` to indicate where it is not. Recall that AS (indicated by AS followed by the value of K in (8) used) and SO always use this transform.

Figure 1 shows the AELD for the Ornstein–Uhlenbeck model for various approximation methods (not including SO, which happens to be exact for this model). Scheme 3 offers at least two orders of accuracy improvement over the Euler–Murayama method, and Scheme 4 offers similar improvement over Scheme 3. Scheme 4's accuracy is similar to AS with $K = 2$ for this model and set of parameters. Figure 2 shows a similar comparison for the Cox–Ingersoll–Ross model using the same parameters used by Durham and Gallant (2002) in their benchmarking. AS with $K = 2$ is strongest in these circumstances, followed by Scheme 4. SO's performance falls between Schemes 3 and 4. There is a notable improvement in Scheme 3 when transform (5) is used (cf. Scheme 3_t and Scheme 3_u). Figure 3 shows the Cox–Ingersoll–Ross model again, but this time with an extreme choice of x_0. In these circumstances AS performs poorly. The other methods remain strong with Scheme 4 performing best. Figure 4 shows MAELD for the Cox–Ingersoll–Ross model, showing how AS is strong when x_0 is close to the stationary mean, but much weaker when x_0 is in the tail of the stationary distribution.

5 Discussion and conclusions

There is no clear-cut best approach for estimating the transition density of SDEs. The user has a range of choices and must weigh up the relative importance of accuracy, speed and ease of implementation. We have presented here two novel methods: Scheme 3, which is straightforward to implement and offers a clear improvement in accuracy over Euler–Murayama; and the Scheme 4, which offers significant further accuracy improvements, but at the cost of being slower and requiring more work to implement.

Our methods are based on expansions that are valid when the time between observations is small, and in this sense

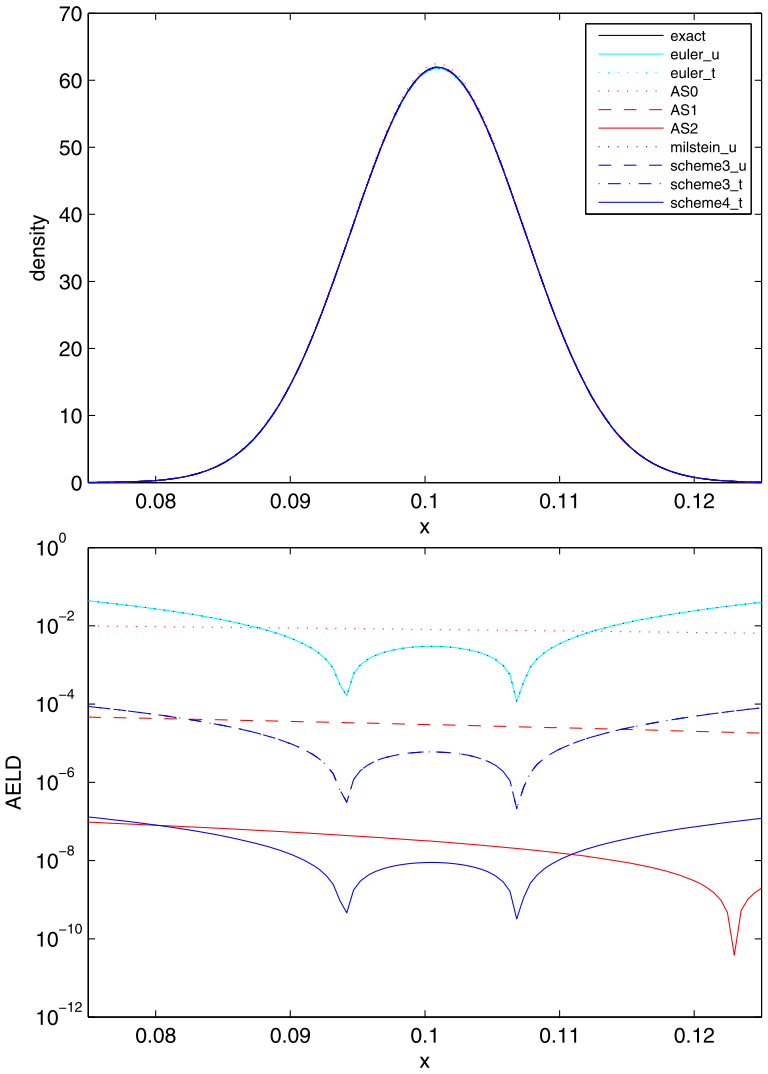

Fig. 1 (*Top*) Estimates of the Ornstein–Uhlenbeck transition density, and (*bottom*) corresponding absolute errors of the log densities. (The Shoji–Ozaki method is exact for the Ornstein–Uhlenbeck process and so is omitted from this figure.) The parameters were the same as those used by Aït-Sahalia (1999): $(\kappa, \alpha, \sigma) = (0.0717, 0.261, 0.02237)$, $x_0 = 0.1$, $t = 1/12$

they belong to the same class of approaches as SO and AS. The small-time expansion approaches have in common the advantage that they require few decisions from the user (in contrast, say, to Durham and Gallant's simulation-based approach in which the user needs to choose the number of sub-intervals, the density estimate used on sub-intervals, the number of simulations, and the type of importance sampler; and in contrast to Fokker–Planck methods where the user must take care to choose a suitable numerical scheme and step size, and a suitable way to incorporate the Dirac initial condition), and that estimates are non-stochastic, that is, they are not subject to Monte Carlo error (in contrast to simulation-based approaches). On the other hand, we do note that the Durham and Gallant simulation-based approach does have advantages not shared by small-time approaches, namely that it generalises more easily to multivariate diffusions (compared, for instance, to the rather more involved approach in Aït-Sahalia 2008) and that it can be made arbitrarily accurate by increasing the number of sub-intervals and simulations, making it a better choice when the sampling interval is not small.

With these points in mind, we focus our discussion on comparing our methods with others in the 'small-time expansion' class, that is, SO and AS. First, in terms of speed, Scheme 3_u involves finding the root of a cubic, and Scheme 4 the root of a rather complicated expression; the computational cost of these steps makes the methods much slower than Scheme 3_t, SO and AS. Typical times to find the MLE of 1000 observations from the CIR model were 25 seconds for Scheme 3_u and 150 seconds for Scheme 4, compared with 5 seconds for each of Scheme 3_t, SO and AS with $K = 2$. Despite the relative speed difference, we suspect that in many applications such run times will be perfectly acceptable.

An advantage of Scheme 3_u over the other methods is that it does not require the unit-diffusion transformation, meaning that it is straightforward to implement, especially for models in which the unit-transform cannot be performed

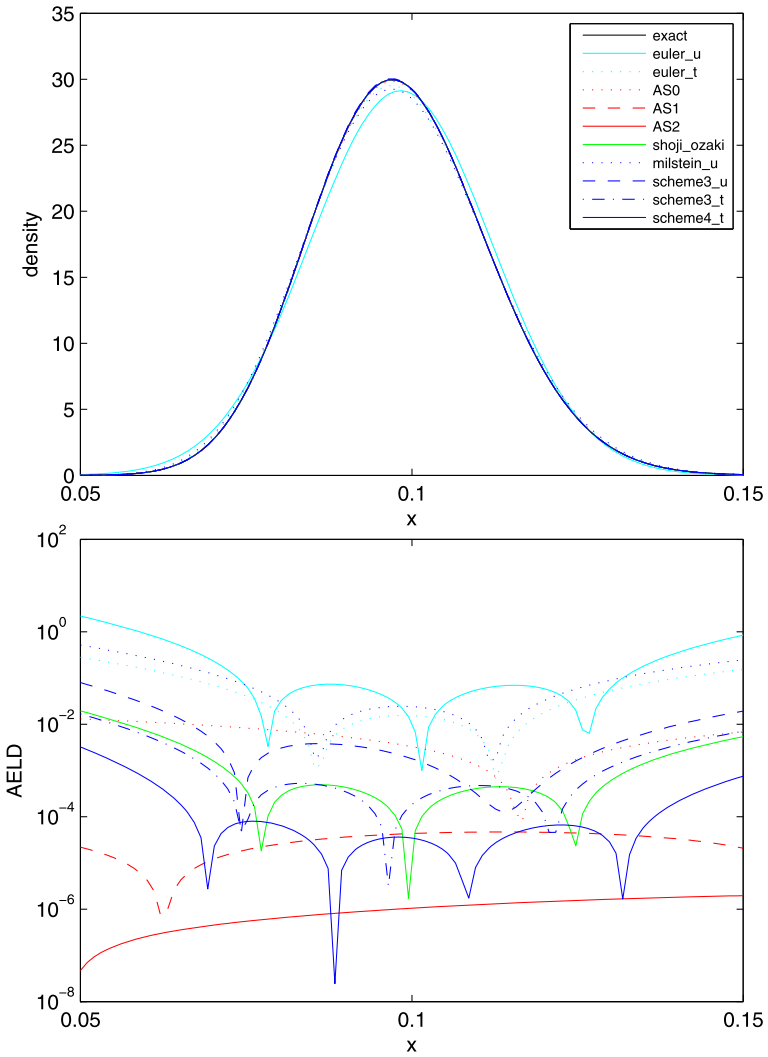

Fig. 2 (*Top*) Estimates of the Cox–Ingersoll–Ross transition density, and (*bottom*) corresponding absolute errors of the log densities. The parameters were the same as those used by Durham and Gallant (2002): $(\kappa, \alpha, \sigma) = (0.5, 0.06, 0.15)$, $x_0 = 0.1, t = 1/12$

analytically. The transform does tend to improve accuracy, however, as can be seen in Fig. 2 comparing Schemes 3_u and 3_t, and Euler_u and Euler_t; see Durham and Gallant (2002) for more discussion. Scheme 4, like SO and AS, does require an initial application of the unit-diffusion transformation, and the CGF calculations are much more involved. An important point, however, is that the CGF is non-model-specific, so applying Scheme 4 to a new model requires only that the user specifies the transformed drift coefficient and its derivatives evaluated at the initial time point. In contrast, the coefficients (9) of AS are model-specific and require rather cumbersome symbolic calculations before AS can be applied to a new model.

The comparison of accuracy between methods is difficult due to the lack of models for which the transition density is exactly known. We have focused on using the same Cox–Ingersoll–Ross model and parameters that Durham and Gallant (2002) used in their tests. Our simulations show that Scheme 4 typically outperforms SO but is less accurate than AS with $K = 2$. Even so, we found circumstances in which the behaviour of AS appears to degrade, such as when the drift term is dominant over the diffusion term—for instance when x_0 is far from the mean of the stationary density (see Figs. 3 and 4). Although a typical trajectory is unlikely to move far into the tail of the true stationary density, in practical situations (such searching over θ to maximise the likelihood) it can certainly arise that one tries to evaluate the likelihood at some θ far from its true value, for which x_0 is indeed in the tail. In very challenging cases AS density estimates can be negative, unlike for Schemes 3 and 4 and SO, a circumstance that calls for other strategies (one possibility is instead to apply an expansion to the log density, then exponentiate, as in Aït-Sahalia 2008) to avoid obvious difficulties in the MLE context. In other words, AS provides an excellent approximation to the likelihood close to the data-generating θ, but can provide a very poor approximation far away from it. These issues are not merely hypothetical: we have encountered problems in practise when using AS to perform Gauss–Newton-based MLE whereby we encounter

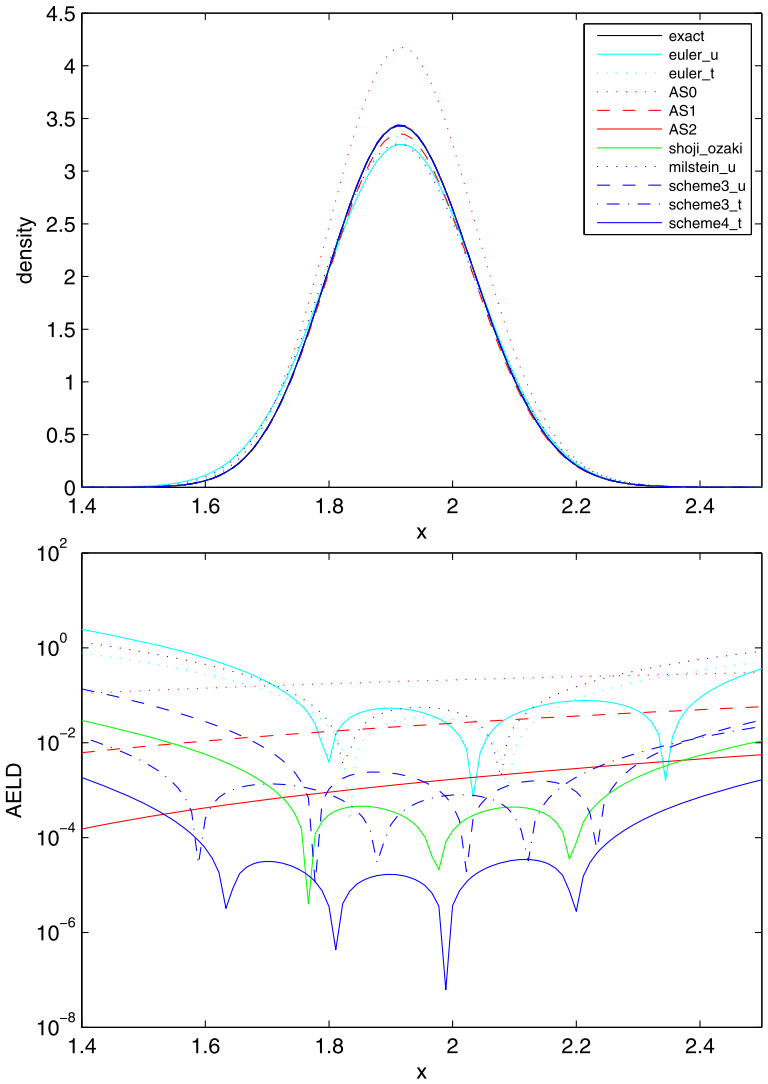

Fig. 3 (*Top*) Estimates of the Cox–Ingersoll–Ross transition density, and (*bottom*) corresponding absolute errors of the log densities, using parameters $(\kappa, \alpha, \sigma) = (1, 1, .3)$, $x_0 = 2, t = 1/12$

negative density estimates and must terminate the optimisation.

Appendix: Asymptotic results

We now consider the asymptotic properties of saddlepoint estimators of the parameter vector θ, i.e., approximate maximum likelihood estimators based on saddlepoint approximations of the transition densities. The relevant asymptotic framework is that where the time intervals between observations go to zero uniformly. There is a substantial literature on this asymptotic regime; see, for example, Prakasa-Rao (1983, 1988), Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989), Yoshida (1992), and Kessler (1997), as well as some of the references mentioned earlier in this paper. Our purpose here is to provide a brief heuristic sketch of the theoretical behaviour of the estimators proposed in this paper without providing detailed or rigorous justification.

Let $p_T(y|x, \theta)$ denote the transition density for moving from x to y in a time increment of length T. Write $p_{m,T}(y|x, \theta)$ for the probability density of the sum of the terms in the Ito–Taylor sample-path expansion consisting of precisely those terms with multi-index α satisfying $l(\alpha) + n(\alpha) \leq m$; see Sect. 3, where $l(\alpha)$ and $n(\alpha)$ are defined and further details of Ito–Taylor expansions are given.

The likelihood for parameter vector θ based on the density $p_{m,T}$ for the truncated sample path expansion, given a sample of observations $x_{t_1}, \ldots, x_{t_n}$, where $t_i = iT$, is

$$L_m(\theta) = \prod_{i=2}^{n} p_{m,T}(x_{t_i}|x_{t_{i-1}}, \theta). \tag{18}$$

The Gaussian likelihood approximation to the likelihood (18) is defined as the product of normal densities, where component normal density i, $2 \leq i \leq n$, has mean and vari-

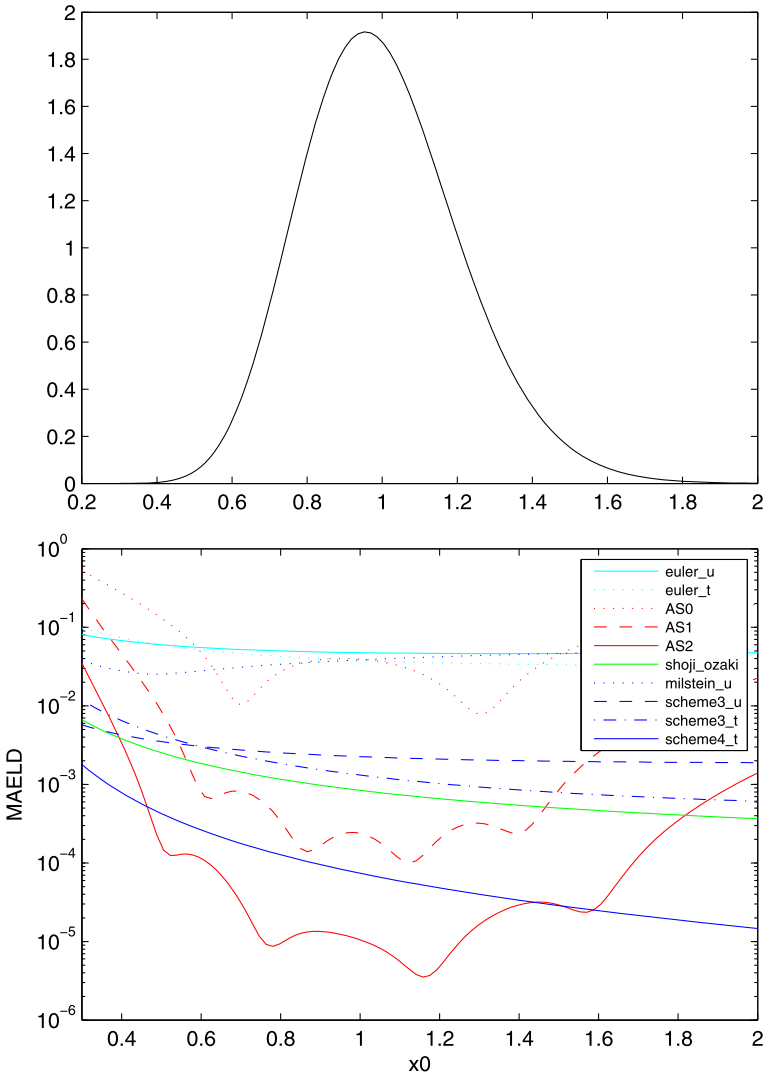

Fig. 4 (*Top*) The Cox–Ingersoll–Ross stationary density for parameters $(\kappa, \alpha, \sigma) = (1, 1, 0.3)$, and (*bottom*) the mean absolute error of the log density (MAELD) versus x_0 using these parameters with $t = 1/12$

ance given by the mean and variance of X_{t_i} conditional on $X_{t_{i-1}} = x_{t_{i-1}}$.

It is straightforward to check that the inclusion of all terms in the Ito–Taylor expansion with $l(\alpha) + n(\alpha) \leq m$ yields a conditional mean and conditional variance which differs for the corresponding conditional mean and conditional variance of the original process by $O_p(T^{c+1})$ in each case, where $c = [m/2]$ and [.] denotes integer part. For the relevant expansions of the conditional mean and conditional variance, see Kessler (1997). Bearing these points in mind, and noting formulae (1.7), (3.1) and (3.4) in Kessler (1997), we can see that $\ell_{p,n}$ in (3.5) of Kessler (1997) is, to the relevant order of approximation, the Gaussian log likelihood approximation to the log likelihood $\log L_m(\theta)$ given in (18), where p and k_0 in Kessler (1997) corresponds to m and c here. Moreover, since for each $m \geq 1$ $L_m(\theta)$ is a well-behaved and regular likelihood, we can expect that, under the conditions of Kessler's (1997) Theorem 1, the maximum likelihood estimator, $\hat{\theta}_m$ say, of θ based on $L_m(\theta)$ satisfies the conclusions of Kessler's Theorem 1. This conclusion may be formulated in the following theorem. For simplicity, we assume that $\sigma(X_t, \theta)$ in (1) does not depend on θ.

Theorem A *Let θ_0 denote the true value of θ in the stochastic differential equation* (1)*, and suppose that the conditions of Theorem* 1 *in* Kessler (1997) *are satisfied. Suppose also that as $n \to \infty$, $nT \to \infty$ and $nT^{2[m/2]+1} \to 0$, where* [.] *denotes integer part. Then*

$$(nT)^{1/2}(\hat{\theta}_m - \theta_0) \to N_d(0, \Xi),$$

where

$$\Xi = \left[\int \frac{1}{\sigma(x)^2} \frac{\partial \mu}{\partial \theta}(x, \theta_0) \frac{\partial \mu}{\partial \theta^T}(x, \theta_0) \lambda_0(dx) \right]^{-1},$$

is the Fisher information for the continuous-time diffusion, and λ_0 is the invariant measure for the (ergodic) *stochastic differential equation* (1).

We first comment on the relevance of Theorem A to Schemes 1–4 considered in Sect. 3. First of all, Scheme 1 (Euler) is not covered by the Theorem, but Florens-Zmirou (1989, Theorem 3) shows that the resulting estimator of θ is consistent, asymptotically normal and efficient if $nT^3 \to 0$. With regard to Schemes 2–4, we note the following: Scheme 2 is asymptotically equivalent to working with $p_{2,T}$, i.e. $m = 2$; Scheme 3 is equivalent to working with $m = 3$ when b is constant, and with $m = 2$ when b is not constant, because the coefficient of $I_{(1,1,1)}$ is non-zero in the latter case, yet this term is not included in Scheme 3; and Scheme 4 is equivalent to taking $m = 4$. Therefore by Theorem A, the condition for Scheme 2 and Scheme 3 to yield a consistent, asymptotically normal and efficient estimator of θ is $nT^3 \to 0$, as for Scheme 1. In contrast, for Scheme 4 Theorem A tells us that the relevant condition is $nT^5 \to 0$. However, the difficulty is that $p_{4,T}$ is not known in closed form, so some kind of approximation is needed.

We now consider what happens when $p_{m,T}$ is non-Gaussian, an explicit form for the density is not available, and we use the saddlepoint approximation $\hat{p}_{m,T}$ to approximate $p_{m,T}$. A key point is that in this asymptotic setting, the error in the approximation of $\hat{p}_{m,T}$ to $p_{m,T}$ is governed by the order of the largest non-Gaussian term in the Ito–Taylor sample path expansion. Specifically, in the case of Scheme 4, where $m = 4$, the largest stochastic term is $I_{(1)}$, which is of order $O_p(T^{1/2})$; the largest non-Gaussian term is $a''I_{(1,1,0)}$, which is of order $O_p(T^2)$; and the resulting error in the saddlepoint approximation is $O(T^{3/2})$. Suppose $y = x + T^{1/2}z$, where x, z and θ are fixed; $m = 4$; and b in Scheme 4 is constant. Then as $T \to 0$,

$$\hat{p}_{m,T}(y|x,\theta) = p_{m,T}(y|x,\theta)\{1 + O(T^{3/2})\}. \tag{19}$$

This can be established by checking that the order of the standardised third and higher odd cumulants of the truncated expansion is $O(T^{3/2})$, while the fourth and higher even cumulants are of smaller order than $O(T^{3/2})$. Specifically, if δ_{2r+1} is the joint cumulant of $T^{-1/2}a''I_{(1,1,0)}$ and $2r$ copies of $T^{-1/2}I_{(1)}$, then it is seen that $\delta_{2r+1} = O(T^{3/2})$, where the multiplication by $T^{-1/2}$ is to standardise the cumulant. Note that this is rather different from the familiar repeated sampling framework where the standardised cumulant of order $r \geq 3$ of a sample mean, say, is of size $n^{-(r-2)/2}$ as the sample size n goes to infinity.

Let us now consider what happens to the likelihood. Write

$$\hat{L}_M(\theta) = L_m(\theta)R_m(\theta),$$

where $\hat{L}_m$ is the saddlepoint approximation to $L_m(\theta)$ in (18), and

$$R_m(\theta) = \prod_{i=2}^{n} \frac{\hat{p}_{m,T}(x_{t_i}|x_{t_{i-1}},\theta)}{p_{m,T}(x_{t_i}|x_{t_{i-1}},\theta)}.$$

Provided that sufficient uniformity is present, which seems plausible in view of the smoothness and regularity of $p_{m,T}$ and $\hat{p}_{m,T}$, the error in R_m is given by

$$R_m = 1 + O_p(nT^{3/2}),$$

i.e., the order of a typical term, see (19), multiplied by the number of terms, n.

Taking logs and differentiating with respect to θ, writing $S_m(\theta) = \partial \log L_m(\theta)/\partial\theta$, $\hat{S}_m(\theta) = \partial \log \hat{L}_m(\theta)/\partial\theta$ and $G_m(\theta) = \partial \log R_m(\theta)/\partial\theta$, we obtain the score equations

$$S_m(\hat{\theta}_m) = 0 \qquad S_m(\hat{\theta}_{m,sp}) + G_m(\hat{\theta}_{m,sp}) = 0, \tag{20}$$

where $\hat{\theta}_m$ is the maximum likelihood estimator of θ based on $p_{m,T}$ and $\hat{\theta}_{m,sp}$ is the approximate maximum likelihood estimator of θ based on $\hat{p}_{m,T}$. Taylor-expanding $S_m(\hat{\theta}_{m,sp})$ about $\hat{\theta}_m$, noting that $S_m(\hat{\theta}_m) = 0$, and using the second equation in (20) and rearranging, we obtain

$$\hat{\theta}_m - \hat{\theta}_{m,sp} = I(\hat{\theta}_m^*)^{-1}G_m(\hat{\theta}_{m,sp}),$$

where $\hat{\theta}_m^*$ is a convex combination of $\hat{\theta}_m$ and $\hat{\theta}_{m,sp}$ and

$$I(\theta) = \frac{\partial^2 \log L(\theta)}{\partial\theta\,\partial\theta^{\mathrm{T}}}$$

is the information matrix based on $p_{m,T}$. The condition for $\hat{\theta}_m$ and $\hat{\theta}_{m,sp}$ to be asymptotically equivalent to first order is that $(nT)^{1/2}(\hat{\theta}_m - \hat{\theta}_{m,sp})$ converges in probability to 0; equivalently, that, in probability,

$$(nT)^{1/2}I(\hat{\theta}_m^*)^{-1}G_m(\hat{\theta}_{m,sp}) \to 0.$$

Since $I(\hat{\theta}_m^*) = O_p(nT)$, provided that there is sufficient regularity for the order of R_m to be preserved under differentiation by θ, which seems a plausible assumption, we see that the condition reduces to

$$(nT)^{1/2}(nT)^{-1}nT^{3/2} \to 0, \quad \text{i.e. } nT^2 \to 0. \tag{21}$$

This is more restrictive than the $nT^3 \to 0$ condition required by Schemes 1–3.

However, the theoretical condition (21) gives rather too pessimistic a picture compared with performance in practise. Note in particular that Scheme 4 is substantially more accurate than Schemes 1–3 in our numerical examples, which suggests that the error due to the saddlepoint approximation may typically be smaller than the error due to approximating the exact transition density p_T by the density $p_{4,T}$ of the truncated Ito–Taylor expansion, especially when T is not very small. Numerically, the saddlepoint approximation of the likelihood based on $p_{4,T}$ is extremely accurate, much more so than the size of the theoretical error suggests. Indeed, for practical purposes, the saddlepoint approximation is virtually exact for the distributions that arise here (the relevant distributions being sums of Gaussian variables and χ^2 variables).

References

Aït-Sahalia, Y.: Transition densities for interest rate and other nonlinear diffusions. J. Finance **54**, 1361–1395 (1999)

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

Aït-Sahalia, Y.: Closed-form likelihood expansions for multivariate diffusions. Ann. Stat. **36**, 906–937 (2008)

Aït-Sahalia, Y., Yu, J.: Saddlepoint approximations for continuous-time Markov processes. J. Econom. **134**, 507–551 (2006)

Beskos, A., Papaspiliopoulos, O., Roberts, G.O., Fearnhead, P.: Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). J. R. Stat. Soc. B **68**, 333–382 (2006)

Butler, R.W.: Saddlepoint Approximations with Applications. Cambridge University Press, Cambridge (2007)

Daniels, H.E.: Saddlepoint approximations in statistics. Ann. Math. Stat. **25**, 631–650 (1954)

Dacunha-Castelle, D., Florens-Zmirou, D.: Estimation of the coefficient of a diffusion from discrete observations. Stochastics **19**, 263–284 (1986)

Durham, G.B., Gallant, A.R.: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. J. Bus. Econ. Stat. **20**, 297–338 (2002)

Florens-Zmirou, D.: Approximate discrete time schemes for statistics of diffusion processes. Statistics **21**, 547–557 (1989)

Kessler, M.: Estimation of an ergodic diffusion from discrete observations. Scand. J. Stat. **24**, 211–229 (1997)

Kloeden, P., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)

Lindström, E.: Estimating parameters in diffusion processes using an approximate maximum likelihood approach. Ann. Oper. Res. **151**, 269–288 (2007)

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

Pennisi, L.L.: Elements of Complex Variables, 2nd edn. Holt, Reinhart and Winstoin, New York (1976)

Prakasa-Rao, B.L.S.: Asymptotic theory for non-linear least squares estimator for diffusion process. Math. Operationforsch. Stat. Ser. Stat. **14**, 195–2009 (1983)

Prakasa-Rao, B.L.S.: Statistical inference from sampled data for stochastic processes. Contemp. Math. **80**, 249–284 (1988)

Shepp, L.A.: On the integral of the absolute value of the pinned Wiener process. Ann. Probab. **10**, 234–239 (1982)

Shoji, I., Ozaki, T.: Estimation for nonlinear stochastic differential equations by a local linearization method. Stoch. Anal. Appl. **16**, 733–752 (1998)

Stramer, O., Yan, J.: On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. J. Comput. Graph. Stat. **16**, 672–691 (2007)

Yoshida, N.: Estimation for diffusion processes from discrete observations. J. Multivar. Anal. **41**, 220–242 (1992)