Parameter estimation for multivariate diffusion systems
Melvin M. Varughese
Department of Statistical Sciences, University of Cape Town, Rondebosch, Cape Town 7707, South Africa

A R T I C L E   I N F O
Article history:
Received 27 October 2011
Received in revised form 9 July 2012
Accepted 9 July 2012
Available online 16 July 2012

Keywords:
Diffusion process
Fokker–Planck equation
Cumulant truncation procedure
Saddlepoint approximation
Markov Chain Monte Carlo

A B S T R A C T
Diffusion processes are widely used for modelling real-world phenomena. Except for select cases however, analytical expressions do not exist for a diffusion process' transitional probabilities. It is proposed that the cumulant truncation procedure can be applied to predict the evolution of the cumulants of the system. These predictions may be subsequently used within the saddlepoint procedure to approximate the transitional probabilities. An approximation to the likelihood of the diffusion system is then easily derived. The method is applicable for a wide range of diffusion systems — including multivariate, reducible diffusion systems that existing estimation schemes struggle with. Not only is the accuracy of the saddlepoint comparable with the Hermite expansion — a popular approximation to a diffusion system's transitional density — it also appears to be less susceptible to increasing lags between successive samplings of the diffusion process. Furthermore, the saddlepoint is more stable in regions of the parameter space that are far from the maximum likelihood estimates. Hence, the saddlepoint method can be naturally incorporated within a Markov Chain Monte Carlo (MCMC) routine in order to provide reliable estimates and credibility intervals of the diffusion model’s parameters. The method is applied to fit the Heston model to daily observations of the S&P 500 and VIX indices from December 2009 to November 2010.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Diffusion processes are continuous-time, continuous-space stochastic processes that have proven to be natural modelling frameworks for many real world phenomena. Over an infinitesimal interval \( dt \), the evolution of a multivariate diffusion process \( \phi_t \) is represented by the following, possibly time inhomogeneous, stochastic differential equation (SDE):

\[
d\phi_t^i = \mu (\phi_t^i, t; \theta) \, dt + \sigma (\phi_t^i, t; \theta) \, dB_t,
\]

where \( \phi_t^i = (\phi_t^i)_{i=1,...,m}; \theta = (\theta_i)_{i=1,...,p} \) is the parameter vector; \( \mu (\phi_t^i, t; \theta) = (\mu_i)_{i=1,...,m}; \sigma^2 (\phi_t^i, t; \theta) = (\sigma_{ij})_{i,j=1,...,m} \) with \( \sigma^2 (\phi_t^i, t; \theta) = \sigma (\phi_t^i, t; \theta)^T \sigma (\phi_t^i, t; \theta) \) and \( B_t \) is an \( m \)-dimensional vector of independent Brownian motions. The \( m \)-dimensional vector \( \phi_t^i \) represents a set of state variables which characterizes the diffusion system through time. The assumption that the \( m \) Brownian motions are independent does not lead to any loss of generality since allowance is made for off-diagonal terms within the diffusion matrix \( \sigma^2 (\phi_t^i, t; \theta) \). Within this framework, the primary focus is on estimating the parameter vector \( \theta \) from discretely sampled data.

The drift vector \( \mu (\phi_t^i, t; \theta) \) and the diffusion matrix \( \sigma^2 (\phi_t^i, t; \theta) \) characterize the evolution of \( \phi_t^i \). Their individual elements are defined as:

\[
\mu_i = \lim_{\Delta t \to 0} \frac{E[\Delta \phi_i | \phi_t^i]}{\Delta t}, \quad \sigma_{ij} = \lim_{\Delta t \to 0} \frac{\text{Cov}[\Delta \phi_i, \Delta \phi_j | \phi_t^i]}{\Delta t}.
\]

E-mail addresses: melvin.varughese@gmail.com, melvin.varughese@uct.ac.za.

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.010
Over an infinitesimally small interval, the diffusion system is distributed as follows:

\[ \phi_t^{\ast} - \phi_i^{\ast} \sim \text{Normal}\left(\mu\left(\phi_i^{\ast}, t; \theta\right) \Delta t, \sigma^2\left(\phi_i^{\ast}, t; \theta\right) \Delta t\right) . \]  

(2)

That is, the diffusion system has a multivariate-normal distribution characterized by the drift vector and diffusion matrix over any infinitesimal interval. Since diffusion processes are Markovian, Eq. (2) may be used to derive the likelihood for continuously sampled diffusion paths. However with discretely sampled diffusion paths, statistical inference is considerably more challenging. This is because the distribution of the diffusion increments over discretely sampled time points is often unknown.

Instead of representing a multivariate diffusion system \( \phi_t^{\ast} \) as a stochastic differential equation, one may instead focus on the Kolmogorov forward equation, which dictates the evolution of its probability density function \( p(\phi_t^{\ast}) \). This is given by:

\[ \frac{\partial p(\phi_t^{\ast})}{\partial t} = -\sum_{i=1}^{m} \frac{\partial}{\partial \phi_i} \left[ \mu(\phi_t^{\ast}, t; \theta) p(\phi_t^{\ast}) \right] + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2}{\partial \phi_i \partial \phi_j} \left[ \sigma_{ij}(\phi_t^{\ast}, t; \theta) p(\phi_t^{\ast}) \right] . \]  

(3)

This is also known as the Fokker–Planck equation. Except where required, the dependence on the drift vector and diffusion tensor on the parameter vector shall be suppressed within the notation. Since a diffusion process is Markovian, the likelihood of a diffusion system sampled at discrete time points \((t_1, t_2, \ldots, t_N)\) is given by:

\[ L(\theta) = p(\phi_t^{\ast}) \prod_{i=2}^{N} p\left(\phi_i^{\ast} | \phi_{i-1}^{\ast}\right) . \]  

(4)

Asymptotically, for large \( N \), the term \( p(\phi_t^{\ast}) \) may be ignored whilst the transitional probability distribution \( p(\phi_t^{\ast} | \phi_{t-1}^{\ast}) \) is the solution to Eq. (3) at time \( t \) with the boundary condition that \( p(\phi_t^{\ast}) \) is given by the Dirac delta function centered around \( \phi_t^{\ast} \) at time \( t_{i-1} \). The likelihood function is central to many inference procedures: it enables us to derive parameter estimates, confidence intervals and to conduct hypothesis tests. Unfortunately, except for a few special cases, Eq. (3) (and hence also the likelihood) is analytically intractable.

The inability to solve Eq. (3) is an impediment to statistical inference. This may be circumvented by attempting to match, by choice of parameters, characteristics of the sampled path with characteristics of the diffusion model. For example, one may choose to estimate the instantaneous means and variances using the corresponding sample moments of the differenced data (Gallant and Long, 1997; Ragwitz and Kantz, 2001). Alternatively one may employ Bayesian imputation to augment the observed data so that the diffusion increments are approximately normally distributed (Roberts and Stramer, 2001).

Since likelihood based methods tend to give more precise parameter estimates than method of moments estimators (Hurn et al., 2007), we instead seek an approximation to the transitional probability distribution of the diffusion process. Monte Carlo methods may be used to approximate the likelihood (Kleinhaus and Friedrich, 2007; Durham and Gallant, 2002). Alternatively, Eq. (3) could be solved numerically. Wojtkiewicz and Bergman (2000) discretized the spatial domain and solved the partial differential equation numerically at each point on the lattice. The finite-difference method discretizes the time domain, taking advantage of the fact that over an infinitesimally small time period, the diffusion process is normally distributed (Wehner and Wolfer, 1987). Huang (2012) developed a quasi-maximum likelihood estimator that approximates the first two conditional moments using a Wagner–Platen approximation. The resulting normal distribution can be used to approximate the transitional probability.

Another possibility is to approximate the transitional probability distribution by a closed form analytic function; for example, a Hermite polynomial expansion (Ait-Sahalia, 2002). This method was shown by Ait-Sahalia to be superior to many of the competing methods—both in terms of the accuracy of the transitional distribution approximation as well as the speed of the algorithm. Stramer et al. (2010) created an MCMC procedure based on the Hermite approximation which could allow for measurement errors in the diffusion process.

It must be stressed that the Hermite approximation is only applicable for reducible diffusion processes. A diffusion process \( X \) is reducible if there exists a one-to-one transformation \( Y = h(X, \theta) \) such that the covariance function of \( Y \) is the identity matrix. Though all univariate diffusion processes are reducible, only some multivariate diffusions share this property. Ait-Sahalia (2008) extended the method to irreducible, multivariate diffusions, but not only is the procedure more difficult to implement, there is also a reduction in the accuracy of the closed-form approximation to the transitional density. Furthermore, the Hermite approximation does not in general integrate to one. Indeed, for parameter values far from the maximum likelihood estimates, the normalizer can be very far from one. This often prevents convergence when applying the Hermite approximation within an MCMC setting (Stramer et al., 2010). Consequently, without modification, the resulting MCMC credibility intervals often suffer from considerable undercoverage.

It has been proposed that the transitional probability may rather be estimated by a saddlepoint approximation (Daniels, 1954). The saddlepoint approximation is an algebraic expression based on a random variable’s cumulant generation function (CGF). In cases where the first few moments of a random variable are known but the corresponding probability density is difficult to obtain, the saddlepoint approximation to the density can be calculated. The tails of a saddlepoint approximation are more accurate than those of a Edgeworth-expansion (Barndorff-Nielsen and Klüppelberg, 1999). Saddlepoint methods have already been used to approximate the transition densities of diffusions (Ait-Sahalia and Yu, 2006; Preston and Wood,
2.1. The saddlepoint approximation

In cases where the distribution of a random vector \( X^T = (X_1, X_2, \ldots, X_m) \) is unknown, the saddlepoint method provides an algebraic approximation based on \( X \)’s cumulant generating function. Let \( \Lambda^T = (\nu_1, \ldots, \nu_m) \) be the parameters for \( X \)’s moment generating function (MGF) \( M(\Lambda) = E[\exp[\nu_1X_1 + \cdots + \nu_mX_m]] \). Also, let \( K(\Lambda) = \ln M(\Lambda) \) denote the cumulant generating function (CGF). Assuming the MGF exists in an open neighbourhood around the origin (Daniels, 1954), the leading order \( m \)-dimensional saddlepoint approximation is given by:

\[
f(x) = (2\pi)^{-m/2} |\nabla^2 K(\Lambda)|^{-1/2} \exp \left[ K(\Lambda) - \Lambda^T x \right].
\]  

(5)

where \( x \) is an \( m \)-dimensional vector and \( \nabla^2 K(\Lambda) \) represents the \( m \times m \) Hessian matrix for \( K(\Lambda) \) and

\[
\nabla K(\Lambda) = \mathbf{x}.
\]

Except for some select cases such as the Normal and the Gamma distribution, substituting the true CGF within Eq. (5) does not simplify to the true distribution, but rather to an approximation thereof.

In many cases, the CGF \( K(\Lambda) \) is unknown, but it may be approximated. Given the first \( n \) cumulants \( (\kappa_1, \kappa_2, \ldots, \kappa_n) \) for a univariate random variable \( X \), \( K(\theta) = E[\exp[\theta X]] \) may be approximated by:

\[
K(\theta) \approx \sum_{i=1}^{n} \frac{\theta^n}{i!} \kappa_i.
\]

(6)

In the univariate case, the saddlepoint is consequently given by:

\[
f_0(x) = \left( \frac{2\pi}{\sum_{i=0}^{n-2} \frac{\kappa_{i+2}\theta^i}{i!}} \right)^{-\frac{1}{2}} \exp \left[ \sum_{i=1}^{n} \frac{\kappa_i\theta^i}{i!} - \theta_0 x \right],
\]

(7)

where: \( x = \sum_{i=0}^{n-1} \frac{\kappa_{i+1}\theta^i}{i!} \).

The approximation within (7) is two-fold as not only is the saddlepoint an approximation to the true density, but the saddlepoint is now based on a truncated approximation to the true CGF. For a diffusion process, the approximate evolution of the cumulants (and hence also the approximation evolution of Eq. (6)) can be calculated through the cumulant truncation procedure.

2.2. Applying the cumulant truncation procedure to a diffusion process

Consider a diffusion process \( \phi_i^* = (\phi_i)_{i=1,\ldots,m} \) with MGF \( M(\Lambda, t) = E[\exp[\nu_1\phi_1(t) + \nu_2\phi_2(t) + \cdots + \nu_m\phi_m(t)]] \) and CGF \( K(\Lambda, t) = \ln M(\Lambda, t) \). In order to calculate the likelihood given in (4), we approximate the transitional density \( p(\phi_i^* | \phi_{i-1}^*) \) by a saddlepoint \( f(x) \). This necessitates the calculation of an approximation to \( K(\Lambda, t) \) at time \( t_i \) given that the diffusion process is equal to \( \phi_{i-1}^* \) at time \( t_{i-1} \).
Let $\Lambda^* = (\partial / \partial \nu_1, \ldots, \partial / \partial \nu_m)$. Like the partial differential equation for the probability density (given in (3)), there is a corresponding partial differential equation for the MGF $M(\Lambda, t)$, the solution of which enables us to predict the evolution of the system’s cumulants. The MGF obeys the following relation (see Appendix):

$$\frac{\partial M(\Lambda, t)}{\partial t} = \left[ \sum_{i=1}^{m} \nu_i \mu_i(\Lambda^*, t) + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \nu_i \nu_j \sigma_{ij}(\Lambda^*, t) \right] M(\Lambda, t). \tag{8}$$

Though the terms $\mu_i(\Lambda, t)$ and $\sigma_{ij}(\Lambda, t)$ denote functions of $\Lambda$ and $t$, the terms $\mu_i(\Lambda^*, t)$ and $\sigma_{ij}(\Lambda^*, t)$ represent differential operators that act on the MGF $M = M(\Lambda, t)$. So, for example, if we have a 2-dimensional diffusion system with drift term $\mu_1(\phi_1, \phi_2) = a \phi_1 - b \phi_1^2 + c \phi_1 \phi_2$, then $\mu_1 \left( \frac{\partial}{\partial \nu_1}, \frac{\partial}{\partial \nu_2} \right) M = a \frac{\partial M}{\partial \nu_1} - b \frac{\partial^2 M}{\partial \nu_1^2} + c \frac{\partial^2 M}{\partial \nu_1 \partial \nu_2}$ is a differential operator and $\mu_1 \left( \frac{\partial}{\partial \nu_1}, \frac{\partial}{\partial \nu_2} \right) M = a \frac{\partial M}{\partial \nu_1} - b \frac{\partial M}{\partial \nu_1^2} + c \frac{\partial^2 M}{\partial \nu_1 \partial \nu_2}$ represents the action of the operator on the MGF $M(\Lambda, t)$. The MGF may be used to characterize the evolution of the diffusion system through time. Unfortunately, as with Eq. (3), Eq. (8) is generally analytically intractable.

The cumulant truncation procedure enables us to convert the partial differential equation for the MGF into a system of ordinary differential equations that also depends on the system parameters $\theta$. The resulting set of equations proves to be far easier to solve numerically than the original partial differential equation. Though, in principle, one can derive a system of ordinary differential equations directly from Eq. (8), the cumulant truncation procedure is instead based on the corresponding partial differential equation for the CGF $K(\Lambda, t) = \ln M(\Lambda, t)$. This is because $K(\Lambda, t)$ can be expressed as a series expansion of the diffusion system’s cumulants.

$$K = \sum_{r_1 \geq 0} \sum_{r_2 \geq 0} \cdots \sum_{r_m \geq 0} \left( \prod_{i=1}^{m} \nu_i^{r_i} \right) \kappa_{r_1, r_2, \ldots, r_m}(t), \tag{9}$$

where:

$$\kappa_{r_1, \ldots, r_m}(t) = \begin{cases} E[\phi_1] & (r_i = 1, r_j = 0; j \neq i) \\ E \left[ \prod_{i=1}^{m} (\phi_1 - E[\phi_1])^{r_i} \right] & (4 > \sum_{j=1}^{m} r_j > 1). \end{cases}$$

In order to be able to predict the evolution of the diffusion system’s cumulants, we need to derive a partial differential equation governing the evolution of $K(\Lambda, t)$. Since $K(\Lambda, t) = \ln M(\Lambda, t)$, we have:

$$\frac{1}{M} \frac{\partial M}{\partial x} = \frac{\partial K}{\partial x}, \tag{10}$$

where $M$ and $K$ depend on some variable $x$. By repeatedly applying the above relation, it is possible to derive the corresponding partial differential equation for the CGF from the partial differential equation for the MGF given by Eq. (8).

The general form of this equation cannot be shown, but rather must be independently derived for each system of interest. The CGF partial differential equation is derived for selected examples within Sections 4 and 5.

By substituting the series expansion in (9), truncated to order $n$, within the partial differential equation for $K(\Lambda, t)$ and subsequently matching the cumulant coefficients $\left\{ \left( \prod_{i=1}^{m} \nu_i^{r_i} / \prod_{i=1}^{m} \nu_i^{r_i} \right) : r_i = 0, 1, 2, \ldots; 1 \leq i \leq m \right\}$, it is possible to derive a system of ordinary differential equations for the centered moments or cumulants. Consequently, by solving this system of differential equations, it is possible to predict the approximate evolution of the cumulants for each proposed parameter vector $\theta$. Throughout the paper this analysis was performed using MATHEMATICA v6.0. These predictions are accurate across a wide range of parameter values (Varughese, 2009). This is key, as we can subsequently use the predicted cumulants within the saddlepoint to approximate $p(\phi^*|\phi_{0!*})$ and hence derive the approximate likelihood at $\theta$. The parameter estimates are taken to be the values that maximize this approximate likelihood, which may be estimated by a Markov Chain Monte Carlo (MCMC) procedure.

3. A modified MCMC parameter estimation algorithm

In this section, a parameter estimation algorithm for a diffusion system is proposed. The algorithm uses the saddlepoint approximation introduced in Section 2 to recreate the transitional probability distribution. This is subsequently used to approximate the likelihood function of the diffusion system. The approximate likelihoods enable us to explore the parameter space of the diffusion system with a modified MCMC procedure. The algorithm is described below.

1. Apply the cumulant truncation procedure to the diffusion system thus deriving a system of ordinary differential equations that describe the cumulants’ evolution. (Note these differential equations will depend on the model parameters $\theta$.)
2. Choose an arbitrary starting set of parameter values $\theta_0$. Set $\theta_{\text{old}} = \theta_0$.
3. Propose a jump from the old set of parameter values $\theta_{\text{old}}$ to a new set of parameter values, $\theta_{\text{new}} = \theta_{\text{old}} + \Delta \theta$ using a suitably chosen proposal distribution $q(\theta_{\text{new}}|\theta_{\text{old}})$.
4. Calculate the likelihoods $L(\theta_{\text{old}})$ and $L(\theta_{\text{new}})$. $L(\theta)$ is calculated as follows:
   i. Set the likelihood, $L(\theta) = 1$.
   ii. Set $i$ to 1.
   iii. Use the system of ordinary differential equations from Step 1 together with the data values observed at time $t_i$, $\phi_{t_i} = (\phi_1, \phi_2, \ldots, \phi_m)$ to predict the values of the cumulants at time $t_{i+1}$.
   iv. Given the data $\phi_{t_i}^*$ at time $t_i$, the transitional probability distribution at time $t_{i+1}$ can be approximated by a saddlepoint approximation $f_{\theta}(\phi_{t_{i+1}}^* | \phi_{t_i}^*)$ after substituting for the cumulants derived from step iii.
   v. Set the likelihood $L(\theta) = L(\theta) \times f_{\theta}(\phi_{t_{i+1}}^* | \phi_{t_i}^*)$. This is an implementation of Eq. (4).
   vi. Set $i = i + 1$ and if $i < N$ (where $N$ is the number of data points) go to step iii.
5. Calculate the acceptance ratio $R$. This is given by:

$$R = \frac{L(\theta_{\text{new}}) \pi(\theta_{\text{new}}) q(\theta_{\text{old}}|\theta_{\text{new}})}{L(\theta_{\text{old}}) \pi(\theta_{\text{old}}) q(\theta_{\text{new}}|\theta_{\text{old}})}$$

where $\pi(\theta)$ is the prior density. We then accept the proposed parameter values $\theta_{\text{new}}$ with probability $\min(1, R)$. That is, set $\theta_{\text{new}} = \theta_{\text{old}}$ with probability $1 - \min(1, R)$. Otherwise leave $\theta_{\text{new}}$ unchanged.
6. Go back to step 3.

After a suitably chosen burn-in period, the MCMC chains sample from the posterior distribution of the diffusion parameters. Hence, in addition to using the medians of the chains as estimates of the diffusion parameters, the $\alpha/2$-th and $(1 - \alpha/2)$-th percentile may be used to construct $(100 - \alpha)$% credibility intervals.

4. A comparative study of the saddlepoint and Hermite methods

Since the Hermite approximation (Ait-Sahalia, 2002) is a popular method for estimating the parameters of a diffusion system, we compare its performance against the saddlepoint. In order to judge the relative accuracies of the saddlepoint and Hermite procedures, the true transitional probability distribution is required as a baseline. Unfortunately, the diffusion models for which the transitional distribution is known are few. Amongst diffusion processes with analytical, non-normal transitional distributions, the Cox–Ingersoll–Ross (CIR) process and Geometric Brownian motion are arguably the most well known. Neither the saddlepoint nor the Hermite approximation in general integrate to one and the approximations are not normalized when performing the comparisons in this section.

Within this section, the CIR process is analyzed didactically: the analysis is meant to illustrate the general derivation of the saddlepoint approximation to a transitional distribution. This is followed by a study of the relative accuracy, for the CIR process, of the Hermite and saddlepoint approximations to the transitional distribution. The section concludes by comparing MCMC implementations for the Hermite and saddlepoint procedures for both the CIR process as well as Geometric Brownian motion.

4.1. Example: deriving the saddlepoint approximation for the CIR process

The CIR process is commonly used to model financial data such as short-term interest rates. The model may be represented as follows:

$$dX_t = b(\mu - X_t)dt + \sigma \sqrt{X_t}dB_t. \quad (11)$$

The CIR process is mean-reverting. Furthermore, provided $2b\mu > \sigma^2$, the CIR process is always positive. This is advantageous as many real-world phenomena are positive and display mean reversion. The probability distribution of the process obeys the following partial differential equation:

$$\frac{\partial p(X_t)}{\partial t} = -\frac{\partial}{\partial X_t} \left[ b(\mu - X_t)p(X_t) \right] + \frac{1}{2} \frac{\partial^2}{\partial X_t^2} \left[ \sigma^2 X_t p(X_t) \right]. \quad (12)$$

An analytical solution to Eq. (12) exists (Cox et al., 1985). Hence it is possible to derive an analytical expression for the transitional distribution which may be subsequently compared with the Hermite and the saddlepoint approximations. This enables us to compare the relative accuracies of the two approximation schemes.

The first step of the cumulant truncation procedure is to derive the evolution of the moment generating function through time. This requires the drift and diffusion terms of the process. These are given by:

$$\mu(X_t, t) = b(\mu - X_t); \quad \sigma(X_t, t) = \sigma \sqrt{X_t}. \quad (13)$$
By substituting the drift and diffusion terms within Eq. (8), we derive the partial differential equation for the MGF $M(\lambda, t) = E[\exp(\lambda X_t)]$:

$$\frac{\partial M}{\partial t} = \lambda b \mu M - \left[ \lambda b - \frac{1}{2} \lambda^2 \sigma^2 \right] \frac{\partial M}{\partial \lambda}. \tag{13}$$

By dividing both sides of Eq. (13) by $M(\lambda, t)$ and applying Eq. (10), it is possible to derive the differential equation for the CGF:

$$\frac{\partial K}{\partial t} = \lambda b \mu - \left[ \lambda b - \frac{1}{2} \lambda^2 \sigma^2 \right] \frac{\partial K}{\partial \lambda}. \tag{14}$$

In most cases the partial differential equation for the CGF will be analytically intractable. Under such scenarios, one may instead substitute a truncated expansion of the CGF in place of the full CGF. As the order of the expansion increases, the accuracy of the approximation tends to increase (Varughese, 2009). Suppose we approximate instead substitute a truncated expansion of the CGF in place of the full CGF. As the order of the expansion increases, the approximation's accuracy tends to increase. Instead, we approximate $K(\lambda, t)$ by a fourth-order expansion of the CGF:

$$K(\lambda, t) \approx 1 + \lambda \kappa_1(t) + \frac{\lambda^2}{2} \kappa_2(t) + \frac{\lambda^3}{3!} \kappa_3(t) + \frac{\lambda^4}{4!} \kappa_4(t). \tag{15}$$

By substituting the above expansion within Eq. (14) and subsequently equating the various coefficients of $\{\lambda^i : i = 1, \ldots, 4\}$, it is possible to derive a system of ordinary differential equations that describe the evolution of the cumulants:

$$\dot{\kappa}_1(t) = b [\mu - \kappa_1(t)]$$
$$\dot{\kappa}_2(t) = \sigma^2 \kappa_1(t) - 2b \kappa_2(t)$$
$$\dot{\kappa}_3(t) = -3b \kappa_3(t) + 3\sigma^2 \kappa_2(t)$$
$$\dot{\kappa}_4(t) = -4b \kappa_4(t) + 6\sigma^2 \kappa_3(t).$$

Note that there is a trade-off as the order of the approximation to the CGF increases. Not only does the approximation's accuracy increase, but the complexity of the system of equations also increases, which will lead to a rise in computing time. A fourth order approximation is chosen as it seems to be a good balance between accuracy and speed.

Boundary conditions must be specified in order to make the solution of the above system of equations unique. When predicting the cumulants at time $t_i$, the boundary conditions are taken to be: $\kappa_1(t_{i-1}) = \theta_0$, $\kappa_2(t_{i-1}) = 0$, $\kappa_3(t_{i-1}) = 0$, $\kappa_4(t_{i-1}) = 0$. That is, in predicting the cumulants at any future time $t_i$, we condition on the observed value of the process at time $t_{i-1}$. This enables the system of ordinary differential equations to be solved numerically. The resulting predictions may be inserted into Eq. (7) to give us the saddlepoint approximation to the transitional probability distribution $p(X_{t_i}|X_{t_{i-1}})$.

Since the CGF is truncated after the fourth cumulant, we shall approximate the transitional probability distribution by $f_4(x)$. From Eq. (7) it follows that:

$$f_4(x) = \frac{\theta_0^2}{2} \kappa_2 + \frac{\theta_0^3}{3} \kappa_3 + \frac{\theta_0^4}{8} \kappa_4,$$

where:

$$x = \kappa_1 + \theta_0 \kappa_2 + \frac{\theta_0^2}{2} \kappa_3 + \frac{\theta_0^3}{6} \kappa_4.$$
Fig. 1. The relative accuracies of the saddlepoint and Hermite approximations for the CIR process at $b = 1.5$, $\mu = 58$ and $\sigma^2 = 15$. The right panel is a plot of three curves: the true transitional probability together with the saddlepoint and Hermite approximations. The saddlepoint and Hermite approximations follow the true transitional probability distribution (a non-central Chi-squared distribution) very closely. The left panel highlights the absolute difference between the true transitional density and the two approximations. Overall, the saddlepoint approximation is more accurate.

Fig. 2. The relative accuracies of the saddlepoint and Hermite approximations to the CIR process for varying values of $b$ (left panel) and $\sigma^2$ (right panel). Unless a parameter is varying, the values are fixed at $b = 1.5$, $\mu = 58$ and $\sigma^2 = 15$.

Fig. 3. The relative accuracies of the saddlepoint and Hermite approximations for varying $t_i - t_{i-1}$ for the CIR process at $b = 1.5$, $\mu = 58$ and $\sigma^2 = 15$. The Hermite approximation is far less accurate for larger time scales.

As a final point, note that unlike the saddlepoint approximation, the Hermite approximation is a small-time expansion. Consequently, the larger $t_i - t_{i-1}$ becomes, the less accurate the Hermite approximation will be. Fig. 3 shows how the Integrated Error for both the saddlepoint and the Hermite approximations changes with $t_i - t_{i-1}$. The Hermite approximation loses accuracy more quickly for larger time scales.

This investigation has been restricted to a subset of the parameter space for the CIR model. In some sense however, the accuracy of the transitional distribution is only important in as far as it enables us to obtain precise parameter estimates for the diffusion model. With multiple parameters, the computational efficiency of the MCMC procedure makes it suitable for parameter estimation. This motivates a comparison of the modified saddlepoint MCMC algorithm described in Section 3 with modified MCMC algorithms that uses the Hermite approximation.

4.3. A comparison of MCMC implementations of the two approximations

Both the saddlepoint MCMC and the Hermite MCMC are judged by comparing the coverage of their resulting credibility intervals with the corresponding coverage obtained from the true MCMC. The traditional Hermite MCMC is known to fail since the Hermite approximation can explode to infinity for parameter values that are far from the maximum likelihood estimate (Stramer et al., 2010). We thus run two versions of the Hermite MCMC: the traditional version that is known to fail and a modified algorithm developed by Stramer et al. (2010) that avoids computation of the posterior’s normalizing constant. To make the analysis more robust, the MCMC implementations are compared for both the CIR process as well as for Geometric Brownian motion.
The empirical coverage of the credibility intervals is calculated from 100 simulated time series; each of length 40. For each of the time series, the four variants of the MCMC procedure (the saddlepoint MCMC, the two Hermite MCMCs and the true MCMC) are run and a credibility interval is calculated. We use improper, uniform priors and a normal proposal density for the parameter values when running the MCMC chains. If a parameter must be positive, any negative value sampled from the proposal distribution is thrown out and another sample is taken. The coverage of each variant of the MCMC procedure is estimated as the proportion of the 100 credibility intervals that contain the true parameter values, i.e. the parameter values used to simulate the time series. In calculating the credibility intervals, an MCMC chain of length 20,000 was run and the first 10,000 steps were trimmed as part of the burn-in period. 90% credibility intervals will correspond to the 5th and 95th percentiles of the resulting chains. It is desirable that the credibility intervals of a proposed MCMC implementation match as closely as possible the observed coverages of the true MCMC.

Despite the computational efficiency of both the saddlepoint and the two versions of the Hermite procedures, a study of the coverage of their corresponding implementations is computationally intensive. 100 MCMC chains, each of length 20,000, must be run for the three MCMC implementations. This restricts us to comparing the coverage of their credibility intervals for a single set of parameter values.

In order to make our analysis as robust as possible, we chose to compare the coverage of the credibility intervals at points in the parameter space where the saddlepoint approximation performs particularly badly in comparison to the Hermite approximation. We compare the relative accuracies of the two approximations using the following statistic:

$$B = \frac{\int_0^\infty |f(x) - \text{Saddle}(x)| \, dx}{\int_0^\infty |f(x) - \text{Hermite}(x)| \, dx}$$

(17)

where \(f(x)\) represents the true distribution and \(\text{Hermite}(x)\) and \(\text{Saddle}(x)\) represent the Hermite approximation and saddlepoint approximation respectively. If the saddlepoint approximation is more accurate than the Hermite approximation, one would expect \(B\) to be less than 1.

First, for the CIR process, we choose to simulate 100 time series with the parameter values:

- \(b = 0.12, \quad \mu = 0.05, \quad \sigma = 0.02, \quad x_0 = 0.049, \quad \Delta t = 1/52\).

For the above parameter values, the relative efficiency statistic, \(B = 27.96\). Second, for Geometric Brownian motion, we choose to simulate 100 time series with the parameter values:

- \(\mu = 0.12, \quad \sigma = 0.2, \quad x_0 = 0.049, \quad \Delta t = 1/12\).

For the above parameter values, \(B = 2.99\). For both diffusion processes, the saddlepoint approximation to the transitional distribution is considerably less accurate than the Hermite approximation.

Table 1 shows the respective coverages of the saddlepoint MCMC, the two Hermite MCMCs (in the table, the traditional algorithm is referred to as Hermite whilst the modified algorithm developed by Stramer et al. (2010) is referred to as Hermite* and the true MCMC for both the CIR process as well as Geometric Brownian Motion. A Euler–Maruyama scheme was used to determine the proposal density for the external variates of the Hermite* procedure. Despite running the comparison for both models at points where the saddlepoint is less accurate than the Hermite approximation, the credibility intervals of the saddlepoint MCMC are closer to the true credibility intervals. The results indicate that the saddlepoint MCMC may be reliably used both to estimate the parameters of a diffusion model as well as to derive their corresponding credibility intervals. This suggests that the stability of the approximations throughout the parameter space is more important than their accuracy at the true parameter values. The credibility intervals produced by the traditional Hermite MCMC have severe undercoverage. This is particularly true for the Geometric Brownian motion since the MCMC chains were started far from the true parameter values (\(\mu_0 = 0.3, \sigma_0 = 0.1\)) and the chains often became stuck. For the CIR process, all the chains were started at the true parameter values (\(b_0 = 0.12, \mu_0 = 0.05, \sigma_0 = 0.02\)). This serves to highlight the convergence problems that the Hermite MCMC chains suffer from. Though, the modified Hermite* MCMC improves the coverage of the credibility intervals, there is still evidence of undercoverage for the Geometric Brownian motion. This seems to be due to the modified MCMC sometimes taking longer than the burn-in period to converge rather than the chains getting stuck. This is a topic for future research.

Table 2 shows the time taken to study the coverage properties of the credibility intervals as well as the average acceptance ratios for the proposed MCMC steps. There are indications that the acceptance ratios for the Hermite procedures are lower. Both the saddlepoint and the Hermite approximations schemes take considerably longer to run than the case where the true distribution is known. The Saddle MCMC takes 70% longer to run than the traditional Hermite for the CIR process and 217%
longer for the Geometric Brownian motion. However, the modifications to the traditional algorithm, Hermite* MCMC take considerably longer to run. Hermite* MCMC takes 81% longer than the Saddle MCMC for the CIR process and 6% longer for the Geometric Brownian motion.

### 4.4. Coverage for non-standard diffusion processes

We extend our study of the coverage of the saddlepoint procedure to a nonlinear, multivariate diffusion process with unknown transitional distribution. Consider a bivariate-system that behaves as follows:

\[
\begin{align*}
    d\phi_1 &= \left( a\phi_1\phi_2 - b\phi_1^2 \right) dt + c\phi_2 dB_t^{(1)} \\
    d\phi_2 &= g \left( \phi^* - \phi_2 \right) + \sigma dB_t^{(2)},
\end{align*}
\]

(18)

where the Brownian motions are independent. Both processes are mean-reverting; \( \phi_2 \) is an Ornstein–Uhlenbeck process and for a fixed value of \( \phi_2 \), the process \( \phi_1 \) will have long-term mean \( a\phi_2/b \). As \( \phi_2 \) increases, the instantaneous mean and variance of \( \phi_1 \) also increases. From Eqs. (8) and (10), the corresponding partial differential for \( K(v_1, v_2, t) = \ln E[\exp\{v_1\phi_1 + v_2\phi_2\}] \) is given by:

\[
\frac{\partial K}{\partial t} = av_1 \left[ \frac{\partial K}{\partial v_1} \frac{\partial K}{\partial v_2} + \frac{\partial^2 K}{\partial v_1 \partial v_2} \right] - b v_1 \left[ \left( \frac{\partial K}{\partial v_1} \right)^2 + \frac{\partial^2 K}{\partial v_1^2} \right] + \frac{c v_1^2}{2} - g v_2 \frac{\partial K}{\partial v_2} + g\phi^* v_2 + \frac{\sigma^2}{2} v_2^2. \tag{19}
\]

A third order cumulant truncation of \( K(v_1, v_2, t) \) yields:

\[
K(v_1, v_2, t) \approx 1 + v_1\kappa_{10}(t) + v_2\kappa_{01}(t) + v_1v_2\kappa_{11}(t) + \frac{v_1^2}{2!}\kappa_{20}(t) + \frac{v_2^2}{2!}\kappa_{02}(t). \tag{20}
\]

(20) may be substituted within (19) to give a system of ordinary differential equations for the cumulants that can subsequently be solved and used within the saddlepoint approximation.

The diffusion system has six parameters: \( a, b, c, g, \phi^* \) and \( \sigma \). Suppose from (18) we simulate time series, each of length 100, for \( \phi_1 \) and \( \phi_2 \). As in the previous subsection, an investigation into the coverage of the credibility intervals is computationally expensive and hence is only feasible for a single set of parameter values. We simulate 100 time series using the parameter values:

\[
a = 0.1, \quad b = 0.02, \quad c = 1.8, \quad g = 0.5, \quad \phi^* = 5, \quad \sigma = 1.
\]

For each time series, the saddlepoint procedure is used to construct 90% credibility intervals for the six parameters. The observed coverages for the parameters are:

\[
a : \frac{80}{100}; \quad b : \frac{82}{100}; \quad c : \frac{83}{100}; \quad g : \frac{87}{100}; \quad \phi^* : \frac{82}{100}; \quad \sigma : \frac{81}{100}.
\]

These coverages are close to the advertised values, suggesting that the procedure works well in this nonlinear, multivariate example.

### 5. Application to financial data

We study the S&P 500 and its relation to the VIX index (a popular measure of implied volatility) over the period December 2009–November 2010. Fig. 4 shows the two indices over this period. The Heston model is fitted to the dataset. It uses proxies for market volatility – such as the VIX index – to account for market movements. The key feature of this model is that the volatility of the asset price movements are assumed to be stochastic.

\[
d \begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} r S_t \\ \delta (\theta - V_t) \end{bmatrix} dt + \begin{bmatrix} S_t \sqrt{(1 - \rho^2)} V_t \\ \rho S_t \sqrt{V_t} \end{bmatrix} d \begin{bmatrix} B_t^{(1)} \\ B_t^{(2)} \end{bmatrix}, \tag{21}
\]

where \( S_t \) represents the asset price, \( V_t \) is the underlying volatility that drives the asset movements and \( r, \delta, \theta, \rho \) and \( \sigma \) are parameters. The Heston model is a multivariate, irreducible diffusion processes for which many of the competing methods would not be applicable.
Fig. 4. S&P 500 and VIX indices from December 2009 to November 2010. The left y-axis is for the S&P and the right y-axis is for the VIX index. The VIX index has been scaled to ensure that the Heston model may mimic the variance of the S&P.

Table 3
Parameter estimates for the S&P and scaled VIX indices over the period December 2009–November 2010.

| Par. | Estimate | 95% C.I.       |
|------|----------|----------------|
| r    | 0.118    | (0.008; 0.430) |
| δ    | 9.863    | (1.410; 13.464)|
| θ    | 0.034    | (0.020; 0.058) |
| ρ    | −0.855   | (−0.880; −0.824)|
| σ    | 0.250    | (0.234; 0.266) |

We wish to run the saddle MCMC in order to obtain both estimates and credibility intervals for the parameters for the Heston model. To run the saddle MCMC, the diffusion tensor $σ_{ij}(S_t, V_t; θ)$ is needed. This is given by:

\[
\begin{bmatrix}
S_t \sqrt{(1-ρ^2) V_t} & ρ S_t \sqrt{V_t} \\
0 & \sigma \sqrt{V_t}
\end{bmatrix}^T \begin{bmatrix}
S_t \sqrt{(1-ρ^2) V_t} & ρ S_t \sqrt{V_t} \\
0 & \sigma \sqrt{V_t}
\end{bmatrix} = \begin{bmatrix}
S_t^2 V_t & ρσ \sqrt{V_t S_t} \\
ρσ \sqrt{V_t S_t} & σ^2 V_t
\end{bmatrix}.
\]

By applying Eqs. (8) and (10) to the above diffusion process, we obtain the following partial differential equation for $K(v_1, v_2, t) = \ln E[\exp\{v_1 S_t + v_2 V_t\}]$:

\[
\frac{∂K(v_1, v_2, t)}{∂t} = v_1 \frac{∂K}{∂v_1} + v_2 δθ - v_2 δ + v_2 \left( \frac{∂^2 K}{∂v_1^2} \right) + v_1^2 \left( \frac{∂^2 K}{∂v_2^2} \right) + v_1 v_2 σ^2 \left( \frac{∂^2 K}{∂v_1 \partial v_2} \right) + v_1 v_2 ρσ \left( \frac{∂^2 K}{∂v_1^2 v_2} \right).
\]

A third order cumulant truncation is performed to yield a system of ordinary differential equations for the cumulants. These may be subsequently used to approximate the likelihood of the Heston model. The modified MCMC (described in Section 3) may then be run on the S&P and scaled VIX dataset.

The resulting parameter estimates and 95% credibility intervals are shown in Table 3. One may see that there is strong negative correlation between the two Brownian motions driving $S_t$ and $V_t$. The parameters $r$ and $δ$ have wide credibility intervals, suggesting that they are not strongly constrained by the data.

6. Conclusion

A general statistical framework for estimating the parameters of a diffusion system is proposed. It is assumed that the elements of both the mean vector and the diffusion matrix may be represented by finite order polynomials. Apart from this, the method may be readily applied to multivariate diffusion systems. Many existing parameter estimation methods are only applicable for univariate diffusion processes or, if applicable to multivariate systems, require the diffusion system to be reducible (Ait-Sahalia, 2008; Beskos et al., 2006).

The transitional distribution of a diffusion system is approximated with a saddlepoint. We show how the cumulants of the system may be predicted and subsequently used to approximate the true transitional distribution of the system. The proposed method is compared with the Hermite approximation proposed by Ait-Sahalia (2002). For both the Cox–Ingersoll Ross model and Geometric Brownian motion, the accuracy of the saddlepoint is comparable to the Hermite approximation at the true parameter values. Furthermore, since the Hermite approximation is a small-time expansion, the Hermite approximation rapidly becomes inaccurate as the time interval between successive observations increases.
The saddlepoint approximations may be easily incorporated within an MCMC algorithm. Since, in regions far from the maximum likelihood estimates, the saddlepoint approximation to the transitional density is better behaved than the Hermite approximation, the saddlepoint MCMC chains have better convergence properties. Consequently, the coverage of the credibility intervals for the saddlepoint MCMC is close to those obtained from the true MCMC. In contrast, the credibility intervals for the unmodified Hermite MCMC suffer from undercoverage. This undercoverage is especially severe when the chains are not started at the true parameter values since the chains often become stuck. This is mitigated by the modified algorithm developed by Stramer et al. (2010), which does not require the computation of the posterior’s normalizing constant. However, the modified chains are not only slower than the saddlepoint MCMC, but they take longer to converge, which caused undercoverage for the Geometric Brownian motion example studied in Section 4 since the chains had not converged by the end of the burn-in period.

The comparison between the saddlepoint and Hermite approximations was only performed for univariate diffusion processes and hence caution must be exercised in making claims for multivariate diffusion processes. However, the good coverage observed for the credibility intervals of the saddlepoint MCMC for the non-linear, multivariate example suggests that the procedure still works well for multivariate diffusion processes. Also, for irreducible multivariate processes, the Hermite procedure requires a two-fold approximation, which leads to a further drop in its accuracy (Ait-Sahalia, 2008).

The saddlepoint MCMC is used to fit the Heston model to the S&P 500 and the VIX indices over the period December 2009–November 2010. Since the Heston model is a multivariate, irreducible diffusion process, the estimation of the model parameters presents a formidable challenge, but the saddlepoint method provides reliable estimates and credibility intervals for the model parameters. As expected, there was significant negative correlation between the two Brownian motions driving the asset prices and the volatility process.

The proposed estimation algorithm has several virtues: it is fast; applicable for a wide range of diffusion systems and gives parameter estimates close to the true values. Furthermore, the corresponding credibility intervals for these estimates have coverage close to their advertised values. This makes the algorithm suitable for estimating the parameters of many diffusion systems.

Acknowledgments

This material is based upon work supported financially by the National Research Foundation of South Africa. The author is grateful to two referees for their constructive comments which have greatly improved the manuscript. Thanks are also due to Drs Trevor Hastie, Leonard Stefanski and Eric Renshaw for helpful discussions. Part of this work was performed whilst hosted by the Department of Mathematics and Statistics at the University of Melbourne.

Appendix. Derivation of Eq. (8)

The MGF of a diffusion system behaves according to the following partial differential equation:

\[ \frac{\partial M(\Lambda, t)}{\partial t} = \left[ \sum_{i=1}^{m} \nu_i \mu_i(\Lambda^*, t) + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \nu_i \nu_j \sigma_{ij}(\Lambda^*, t) \right] M(\Lambda, t). \]  

(A.1)

Proof. Consider the multivariate diffusion system \( \phi_t^* \) with instantaneous mean vector \( \mu(\phi_t^*, t) \) and diffusion matrix \( \sigma(\phi_t^*, t) \). We denote the transition probability by:

\[ \Pr[\Delta \phi_t^* = \Delta x|\phi_t^*] = g(\Delta x|\phi_t^*). \]

Let \( \Lambda^T = (\nu_1, \ldots, \nu_m) \) be the parameters of the MGF and let \( \Lambda^* = (\partial/\partial \nu_1, \ldots, \partial/\partial \nu_m) \). The MGF, \( M(\Lambda, t) = E[\exp[\nu_1 \phi_1(t) + \cdots + \nu_m \phi_m(t)]] \), obeys the following differential equation (Barbour, 1972):

\[ \frac{\partial M(\Lambda, t)}{\partial t} = \sum_{\Delta x} e^{\Lambda^T \Delta x} - 1 \) \( g(\Delta x|\Lambda^*) M(\Lambda, t). \]  

(A.2)

For small \( \Delta x \), we have:

\[ e^{\Lambda^T \Delta x} - 1 = e^{\sum_i \nu_i \Delta x_i} - 1 = \sum_{i=1}^{m} \nu_i \Delta x_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \nu_i \nu_j \Delta x_i \Delta x_j. \]  

(A.3)

For diffusion processes the higher order terms are negligible. We thus only require the joint and marginal transitional rates of the diffusion system to characterize the evolution of the MGF. We denote the joint transition probabilities for \( \phi_i \) and \( \phi_j \) by:

\[ \Pr[\Delta \phi_i = x_i, \Delta \phi_j = x_j|\phi_t^*] = g_{ij}(x_i, x_j|\phi_t^*). \]
The marginal transition rate is denoted by:
\[ Pr[\Delta \phi_i = x_i | \phi_i^*] = g_i(x_i | \phi_i^*) .\]

Note that the instantaneous mean \( \mu_i(\phi_i^*, t) \) and covariances \( \sigma_{ij}(\phi_i^*, t) \) can be represented in terms of the joint and marginal transitional rates:
\[
\begin{align*}
\mu_i(\phi_i^*, t) &= \lim_{\Delta t, \Delta x \to 0} \frac{\Delta x_i}{\Delta t} \left( g_i(\Delta x_i | \phi_i^*) - g_i(-\Delta x_i | \phi_i^*) \right). \\
\sigma_{ij}(\phi_i^*, t) &= \lim_{\Delta t, \Delta x_i, \Delta x_j \to 0} \frac{(\Delta x_i \Delta x_j)}{\Delta t} \left( g_{ij}(\Delta x_i, \Delta x_i | \phi_i^*) + g_{ij}(-\Delta x_i, -\Delta x_j | \phi_i^*) - g_{ij}(\Delta x_i, -\Delta x_j | \phi_i^*) - g_{ij}(-\Delta x_i, \Delta x_j | \phi_i^*) \right). 
\end{align*}
\]

By substituting (A.3) within (A.2), we obtain:
\[
\frac{\partial M(\Lambda, t)}{\partial t} = \sum_{i=1}^{m} \Delta x_i \left[ \frac{g_i(\Delta x_i | \Lambda^*)}{\Delta t} + \frac{g_i(-\Delta x_i | \Lambda^*)}{\Delta t} \right] u_i M \\
+ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \Delta x_i \Delta x_j \left[ \frac{g_{ij}(\Delta x_i, \Delta x_j | \Lambda^*) + g_{ij}(-\Delta x_i, -\Delta x_j | \Lambda^*) - g_{ij}(\Delta x_i, -\Delta x_j | \Lambda^*) - g_{ij}(-\Delta x_i, \Delta x_j | \Lambda^*)}{\Delta t} \right].
\]

The expressions for the instantaneous mean and variance may be substituted within (A.6). This gives us:
\[
\frac{\partial M(\Lambda, t)}{\partial t} = \sum_{i=1}^{m} u_i \mu_i(\Lambda^*, t) + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} u_i u_j \sigma_{ij}(\Lambda^*, t) M(\Lambda, t). \quad \Box
\]

References

Ait-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrika 70, 223–262. Ait-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Ann. Statist. 36, 906–937. Ait-Sahalia, Y., Yu, J., 2006. Saddlepoint approximations for continuous-time Markov processes. J. Econom. 134, 507–551. Barbour, A., 1972. The principle of the diffusion of arbitrary constants. J. Appl. Probab. 9, 519–541. Barndorff-Nielsen, O., Kluppelberg, C., 1999. Tail exactness of multivariate saddlepoint approximations. Scand. J. Stat. 26, 253–264. Beskos, A., Papaspiliopoulos, O., Roberts, G., Fearnhead, P., 2006. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. J. Roy. Statist. Soc. Ser. B 68, 333–382. Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrika 53, 385–407. Daniels, H., 1954. Saddlepoint approximations in statistics. Ann. Math. Stat. 25, 631–650. Durham, C., Gallant, A., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). J. Bus. Econom. Statist. 20, 297–338. Gallant, A., Long, J., 1997. Estimating stochastic differential equations efficiently by minimum chi-squared. Biometrika 84, 125–141. Gillespie, C., Renshaw, E., 2007. An improved saddlepoint approximation. Math. Biosci. 208, 359–374. Huang, X., 2012. Quasi-maximum likelihood estimation of multivariate diffusions. Stud. Nonlinear Dyn. Econom. 5, 390–455. Hurn, A., Jeisman, J., Lindsay, K., 2007. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. J. Financ. Econom. 5, 390–455. Kleinhaus, D., Friedrich, R., 2007. Maximum likelihood estimation of drift and diffusion functions. Phys. Lett. A 368, 194–198. Preston, S., Wood, A., 2012. Approximation of transition densities of stochastic differential equations by saddlepoint methods applied to small-time Itô–Taylor sample-path expansions. Stat. Comput. 22, 205–217. Ragwitz, M., Kantz, H., 2001. Indispensable finite time corrections for Fokker–Planck equations from time series data. Phys. Rev. D 87, 25–29. Roberts, G., Stramer, O., 2001. On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. Biometrika 88, 603–621. Stramer, O., Bognar, M., Schneider, P., 2010. Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. J. Financ. Econom. 8, 450–480. Varugheese, M., 2009. On the accuracy of a diffusion approximation to a discrete state-space Markovian model of a population. Theor. Popul. Biol. 76, 241–247. Varugheese, M., 2011. A framework for modelling ecological communities and their interactions with the environment. Ecol. Complex. 8, 105–112. Varugheese, M., Fatti, L., 2008. Incorporating environmental stochasticity within a biological population model. Theor. Popul. Biol. Wehner, M., Wolfe, W., 1987. Numerical evaluation of path-integral solutions to Fokker–Planck equations. Phys. Rev. A 35, 1795–1801. Whittle, P., 1957. On the use of the normal approximation in the treatment of stochastic processes, J. Roy. Statist. Soc. Ser. B 19, 268–281. Wójcikiewicz, S., Bergman, L., 2000. Numerical solution of high dimensional Fokker–Planck equations. In: Proceedings of the 8th Specialty Conference on Probabilistic Mechanics and Structural Reliability. Paper PMC2000-167. Zhang, S., Song, F., Shi, D., Zhou, Q., 2012. Information ratio test for model misspecification on parametric structures in stochastic diffusion models. Comput. Statist. Data Anal. 56, 3975–3987.