Computational Statistics and Data Analysis 57 (2013) 417–428

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

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, irreducible 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.

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 ϕ_t^* is represented by the following, possibly time inhomogeneous, stochastic differential equation (SDE):

$$d\phi_t^* = \mu\left(\phi_t^*, t; \theta\right) dt + \sigma\left(\phi_t^*, t; \theta\right) dB_t, \tag{1}$$

where $\phi_t^* = (\phi_i)_{i=1,\dots,m}$; $\theta = (\theta_i)_{i=1,\dots,p}$ is the parameter vector; $\mu\left(\phi_t^*, t; \theta\right) = (\mu_i)_{i=1,\dots,m}$; $\sigma^2\left(\phi_t^*, t; \theta\right) = \left(\sigma_{ij}\right)_{i,j=1,\dots,m}$ with $\sigma^2\left(\phi_t^*, t; \theta\right) = \sigma\left(\phi_t^*, t; \theta\right)^T \sigma\left(\phi_t^*, t; \theta\right)$ and B_t is an m-dimensional vector of independent Brownian motions. The m-dimensional vector ϕ_t^* 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 $\sigma^2\left(\phi_t^*, t; \theta\right)$. Within this framework, the primary focus is on estimating the parameter vector θ from discretely sampled data.

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

$$\mu_i = \lim_{\Delta t \to 0} \frac{E[\Delta\phi_i|\phi_t^*]}{\Delta t}, \qquad \sigma_{ij} = \lim_{\Delta t \to 0} \frac{\mathrm{Cov}[\Delta\phi_i, \Delta\phi_j|\phi_t^*]}{\Delta t}.$$

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:

$$\phi^*_{t+\Delta t} - \phi^*_t \sim \text{Normal}\left(\mu\left(\phi^*_t, t; \theta\right)\Delta t, \sigma^2\left(\phi^*_t, t; \theta\right)\Delta t\right). \tag{2}$$

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 ϕ^*_t as a stochastic differential equation, one may instead focus on the Kolmogorov forward equation, which dictates the evolution of its probability density function $p(\phi^*_t)$. This is given by:

$$\frac{\partial p(\phi^*_t)}{\partial t} = -\sum_{i=1}^{m} \frac{\partial}{\partial \phi_i}\left[\mu\left(\phi^*_t, t; \theta\right)p(\phi^*_t)\right] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2}{\partial \phi_i \partial \phi_j}\left[\sigma_{ij}(\phi^*_t, t; \theta)p(\phi^*_t)\right]. \tag{3}$$

This is also known as the Fokker–Planck equation. Except where required, the dependence of 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:

$$L(\theta) = p(\phi^*_{t_1}) \prod_{i=2}^{N} p\left(\phi^*_{t_i} | \phi^*_{t_{i-1}}\right). \tag{4}$$

Asymptotically, for large N, the term $p(\phi^*_{t_1})$ may be ignored whilst the transitional probability distribution $p(\phi^*_{t_i}|\phi^*_{t_{i-1}})$ is the solution to Eq. (3) at time t_i with the boundary condition that $p(\phi^*_t)$ is given by the Dirac delta function centered around $\phi^*_{t_{i-1}}$ 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 (Kleinhans 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,