Estimating the Parameters of Stochastic Differential Equations Using a Criterion Function Based on the Kolmogorov-Smirnov Statistic

A. David McDonald
CSIRO Division of Marine Research
GPO Box 1538
Hobart, Tasmania 7001
Australia
Email: David.McDonald@marine.csiro.au

and

Leif K. Sandal
Department of Mathematics and Statistics
Norwegian School of Economics and Business Administration
Helleveien 30
N-5035 Bergen-Sandviken
Norway
Email: leif@hamilton.nhh.no

Abstract

Estimation of parameters in the drift and diffusion terms of stochastic differential equations involves simulation and generally requires substantial data sets. We examine a method that can be applied when available time series are limited to less than 20 observations per replication. We compare and contrast parameter estimation for linear and nonlinear first-order stochastic differential equations using two criterion functions: one based on a Chi-square statistic, put forward by Hurn and Lindsay (1997), and one based on the Kolmogorov-Smirnov statistic. The estimates generated appear to be precise for all models examined, especially when using the Kolmogorov-Smirnov criterion function.

1. INTRODUCTION

Stochastic differential equations (SDEs) are used in the modelling of many physical, biological and economic systems. Estimating the parameters of both the drift and diffusion terms of these linear and nonlinear models is an interesting topic in itself, especially when assessing the relative importance of deterministic and stochastic components of the process under study. We focus on estimation methods that are particularly suitable for applications when few data points are available, such as when one attempts to characterise the dynamics of fish populations, although these methods are also appropriate for large data sets. We first make use of a criterion function based on the Chi-square statistic and then examine the usefulness of a criterion function based on the Kolmogorov-Smirnov statistic.

2. THE SDE MODELS EXAMINED

The SDEs we examine take the general form

$$dx(t) = f(x,t)dt + g(x,t)dw(t) \tag{1}$$

where $x(t)$ is the state variable of interest $f(\cdot)$ and $g(\cdot)$ are arbitrary functions and $w(t)$ is a standard Wiener process. A standard Wiener process is continuous and Gaussian with independent increments such that $w(0) = 0, E[w(t)] = 0$ and $var\,[w(t) - w(s)] = t - s$ for $0 \le s \le t$. The SDE is, therefore, analagous to an ordinary differential equation perturbed by white noise and has a solution incorporating stochastic integrals. Kloeder *et al*. (1994) provide a very useful introduction to SDEs and their numerical solution.

Using subscripts to denote the time index, the specific equations that we examine are

$$dx_t = \alpha x_t d\mathrm{t} + \beta x_t dw_t \tag{2}$$

$$dx_t = x_t(1 - x_t)dt + \beta x_t^2 dw_t \tag{3}$$

$$dx_t = \alpha x_t(1 - x_t)dt + \beta x_t dw_t \tag{4}$$

$$dx_t = \alpha x_t(1 - x_t)dt + 0.2\beta x_t(1 - x_t)dw_t \tag{5}$$

$$dx_t = x_t(1 - x_t)^{\alpha} dt + 0.5 x_t^{2\beta} dw_t \tag{6}$$

where a and b are constant coefficients to be estimated as parameters.

These equations were chosen because they are of the form used frequently for modelling renewable resource systems. Obtaining point and interval estimates of the drift and diffusion parameters of such models is important in applied work because these estimates provide a means of testing hypotheses about the state of the system and the relative importance of stochastic influences on it.

Observed data for stochastic processes are recorded for discrete time intervals, regardless of whether the system is described best by a continuous or discrete model. One advantage of using a continuous model is that its solution can, in principle, be used for any time interval, without altering the meaning or interpretation of the model parameters.

Estimation of the SDE parameters requires the solution or an approximation to it. Given the difficulty associated with finding closed-form solutions for many nonlinear SDE's, we concentrate on two methods for finding discrete-time approximations to the solution of Equations 2-6; namely the strong Euler scheme that attains convergence of order 0.5 (Kloeden *et al.* 1994, pp. 140-2) and the strong Taylor scheme that attains convergence of order 1.5 (Kloeden *et al.* 1994, pp. 162-3).

2.1 Parameter Estimation Using a Chi-square Criterion Function

Hurn and Lindsay (1997) put forward a method, based on the Chi-square statistic, for estimating both the drift and diffusion coefficients of linear SDE's. This method relies on the existence of observed replicated time-series data and replicated simulation of time series realisations from a specified SDE. Parameter-estimation is then possible by optimising the fit between the observed and simulated data over a plausible continuous interval for each of the

SDE parameters. The parameter estimates are chosen to optimise the goodness of fit between the observed and simulated data. Classical least squares and maximum likelihood are inadequate for simultaneous estimation of the drift and diffusion parameters, so Hurn and Lindsay adopted a criterion function based on the Chi-square statistic for testing the null hypothesis that the observed and simulated data are drawn from the same distribution.

The method proposed involves placing the *m* observed data points in each time period into bins, as one would for a standard goodness-of-fit test. Then, using the model proposed for the stochastic process and initial parameter values, simulate *n* realisations of the process and allocate these data to the bins specified in each time period for the observed data. For each time period, *t*, one enumerates the statistic

$$\chi_t^2 = \sum_{j=1}^{r+1} \frac{(m_j - n_j)_t^2}{(m_j + n_j)_t} \qquad (7)$$

where m_j is the number of observed data points in bin *j*, n_j is the (expected) number of simulated data points in bin *j* and there are *r*+1 bins in total. This statistic is assumed to be distributed as $\chi^2_{(r)}$ where *r* is the number of degrees of freedom. A small value of this statistic lends support to the null hypothesis that the observed data are drawn from the same distribution as are the simulated data. Using the gamma function $\Gamma(\cdot)$ and the product operator Π, and assuming timewise independence, this leads to the criterion function

$$\varphi = \prod_{t=1}^{T} \frac{1}{\Gamma(r/2)} \int_{\frac{1}{2}\chi_t^2}^{\infty} y^{\frac{r}{2}-1} e^{-y} \, dy \qquad (8)$$

which is maximised with respect to the SDE drift and diffusion parameters.

2.2 Parameter Estimation Using a Criterion Function Based on the Kolmogorov-Smirnov Statistic

The Kolmogorov-Smirnov statistic adapted to a two-sample problem provides the basis for another goodness-of-fit method for SDE parameter estimation. As with the χ^2 test the two-sample Kolmogorov-Smirnov goodness-of-fit test is used to compare the empirical distribution functions of two samples. In the present paper one of these samples is generated as if observed from a fully-specified SDE and the other is generated from the same SDE but with the assumption that the coefficients are unknown. Parameter estimation in practice requires one of the samples to be observed and the other to be generated by the SDE that is used to model the data-generating process.

Following Gibbons (1985, pp. 127-31), the empirical distribution functions, denoted by $S_m(n)$ and $S_n(n)$, are defined as

$$S_m(x) = \begin{cases} 0 & if \ x < x_{(1)} \\ k/m & if \ x_{(k)} \le x < x_{(k+1)} \ \ for \ k = 1, 2, \cdots, m-1 \\ 1 & if \ x \ge x_{(m)} \end{cases}$$

$$S_n(x) = \begin{cases} 0 & if \ x < y_{(1)} \\ k/n & if \ y_{(k)} \le x \le y_{(k+1)} \ \ for \ k = 1, 2, \cdots, n-1 \\ 1 & if \ x \ge y_{(n)} \end{cases}$$

The Kolmogorov-Smirnov two-sample test statistic

$$D^*_{m,n} = \max_x |S_m(x) - S_n(x)| \tag{9}$$

is the maximum absolute difference between the two empirical distributions. This statistic can be used to test the (null) hypothesis that the population distributions are identical and, therefore, that both samples have been drawn from the same population. The two-sample Kolmogorov-Smirnov statistic has an asymptotic null distribution given by

$$\lim_{m,n \to \infty} P(\sqrt{\frac{mn}{m+n}} D_{m,n} \le D) = L(D) \tag{10}$$

where $L(D) = 1 - 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 D^2}$.

A large value of *D*, and therefore a small value of *L(D)*, indicates that the null hypothesis is unlikely to be true, whereas small *D* values support the null hypothesis.

In the present paper we are concerned with replicated time series data. This provides the opportunity to evaluate the Kolmogorov-Smirnov statistic for each time period, D_t. Because we are dealing with a stochastic process that we assume to be modelled adequately by an equation of the general form of Equation 1, we must estimate the drift and diffusion parameters so that the entire time series is taken into account. We do this in the first instance by analogy with maximum-likelihood estimation, taking as our criterion function the product of Kolmogorov-Smirnov statistics computed at each time step. Using the asymptotic null distribution, this yields

$$\Phi = \prod_{t=1}^{T} L(D_t) \tag{11}$$

Given a set of observations or simulated observations giving rise to $S_m(x)$, this criterion function is maximised with respect to the drift and diffusion parameters of an SDE that is used in the evaluation of $S_n(x)$[1].

3. EMPIRICAL RESULTS

Estimation of the drift and diffusion parameters of Equations 2-6 was conducted using the above criterion functions. Fifty realisations of eleven 'observed' data points (i.e. $m = 50,\ T = 11$) were generated for each model using the initial value $x(0) = 0.5$ and true parameters $\alpha = 1.0$ and $\beta = 0.5$. The derivative-free simplex method of Nelder and Mead (1965) was then used to estimate α and β using start values of $\hat{\alpha} = 1.3$ and $\hat{\beta} = 0.4$, initial value $x(0) = 0.1$, number of simulated replications $n = 50$, and time-series length $T = 11$. Both the Euler and Taylor SDE solver schemes referred to above were used. Parameter

[1] It is worth pointing out that the likelihood analogy should not be taken too far. We are not assuming independence here, as is the case with the χ^2-based criterion function. Rather, we are formulating a distance measure that allows weight to be given to each of the Kolmogorov-Smirnov statistics that we

estimation was carried out 500 times for each equation and solver scheme. The mean and standard deviation of parameter estimates are reported in Table 1.

evaluate in each time period. As discussed below, we also evaluate an alternate criterion function based on the Kolmogorov-Smirnov statistic.

Table 1: Mean and standard deviation of 500 parameter estimates

Criterion function	Mean value	Standard deviation
Linear model: $dx = x[\alpha dt + \beta d\omega]$ Equation 2		
Based on χ^2 (Taylor)	$\alpha = 1.0468$	$\sigma_\alpha = 0.0407$
	$\beta = 0.5466$	$\sigma_\beta = 0.0374$
(Euler)	$\alpha = 1.0420$	$\sigma_\alpha = 0.0371$
	$\beta = 0.5412$	$\sigma_\beta = 0.0345$
Based on Kol. Smir. (Taylor)	$\alpha = 1.0189$	$\sigma_\alpha = 0.0335$
	$\beta = 0.5224$	$\sigma_\beta = 0.0324$
(Euler)	$\alpha = 1.0152$	$\sigma_\alpha = 0.0313$
	$\beta = 0.5192$	$\sigma_\beta = 0.0299$
Nonlinear model 1: $dx = x(1-x)dt + \beta x^2 d\omega$ Equation 3		
Based on χ^2 (Taylor)	$\beta = 0.5130$	$\sigma_\beta = 0.0279$
Based on Kol. Smir. (Taylor)	$\beta = 0.5189$	$\sigma_\beta = 0.0211$
Nonlinear model 2: $dx = \alpha x(1-x)dt + \beta x d\omega$ Equation 4		
Based on χ^2 (Taylor)	$\alpha = 1.1257$	$\sigma_\alpha = 0.0933$
	$\beta = 0.5832$	$\sigma_\beta = 0.0564$
Based on Kol. Smir. (Taylor)	$\alpha = 1.0552$	$\sigma_\alpha = 0.0447$
	$\beta = 0.5353$	$\sigma_\beta = 0.0270$
Nonlinear model 3: $dx = x(1-x)[\alpha dt + \beta d\omega]$ Equation 5		
Based on χ^2 (Taylor)	$\alpha = 1.0237$	$\sigma_\alpha = 0.0267$
	$\beta = 0.5615$	$\sigma_\beta = 0.0404$
Based on Kol. Smir. (Taylor)	$\alpha = 0.9906$	$\sigma_\alpha = 0.0103$
	$\beta = 0.5155$	$\sigma_\beta = 0.0219$
Nonlinear model 4: $dx = x(1-x)^\alpha dt + 0.5x^{2\beta} d\omega$ Equation 6		
Based on χ^2 (Taylor)	$\alpha = 1.0897$	$\sigma_\alpha = 0.0847$
	$\beta = 0.5150$	$\sigma_\beta = 0.0170$
(Euler)	$\alpha = 1.0823$	$\sigma_\alpha = 0.0873$
	$\beta = 0.5125$	$\sigma_\beta = 0.0168$
Based on Kol. Smir. (Taylor)	$\alpha = 1.0319$	$\sigma_\alpha = 0.0455$
	$\beta = 0.5099$	$\sigma_\beta = 0.0145$
(Euler)	$\alpha = 1.0290$	$\sigma_\alpha = 0.0455$
	$\beta = 0.5087$	$\sigma_\beta = 0.0152$

True parameter values are α=1 and β=0.5.

It should be noted that, because of the heavy computational burden associated with enumerating *L(D)*, in Equation 10, we used the approximation

$$L(D) \approx 1 - e^{-2D^2} \qquad (12)$$

which is the asymptotic distribution of the one-sided two-sample test outlined by Gibbons (1985, pp. 130-1). This approximation halved the computation time taken when using either the formulation of *L(D)* from Equation 10 or the χ^2-based criterion function given by Equation 8.

The results presented in Table 1 reveal a small bias in the estimates of α and β over the 500 repeats of parameter estimation. In all cases, however, the mean of the point estimates is well within 1½ standard deviations of the true value. Interestingly the lower order of convergence Euler scheme is associated with smaller bias and standard deviation estimates for Equations 2 and 6. This is likely to be a result of the use of intermediate time steps with the equation solver to improve the SDE simulations. Also of note is the systematically smaller bias and standard error estimates associated with the Kolmogorov-Smirnov criterion function, as compared to the χ^2-based criterion function, which could be due in large part to the need for binning the data for the χ^2 method.

Further results, not reported in the present draft, relate to alternative Kolmogorov-Smirnov criterion functions and the use of the form of *L(D)* of Equation 10. A criterion function made up of the sum of D_t over the time series (as opposed to the product given in Equation 11) and used for estimating the parameters of Equation 6, gave mean and standard deviation of 1.011 and 0.057 for $\hat{\alpha}$, and of 0.505 and 0.015 for $\hat{\beta}$ using the Euler scheme. This represents a reduction in the bias. The ‘exact’ form of *L(D)* from Equation 10 applied to estimation of the parameters of Equation 6 yielded mean and standard deviation of 0.998 and 0.038 for $\hat{\alpha}$ and of 0.499 and 0.013 for $\hat{\beta}$: an apparent further reduction in bias, as well as an improvement in precision. On the basis of these results there is room for further investigation of the form of criterion function used for parameter estimation.

4. CONCLUSION

Estimation of the parameters of five linear and nonlinear SDE's using criterion functions based on Chi-square and Kolmogorov-Smirnov statistics has been examined. Although the methods used are demanding computationally, the results are satisfactory with respect to both point and interval estimation. Given the bias evident from 500 repeated estimations and the potential improvements evident with the Kolmogorov-Smirnov criterion function, further work using the methods outlined above is both justified and desirable.

REFERENCES

Gibbons, J.D. (1985). *Nonparametric Statistical Inference* (2nd Ed.), Marcel Dekker, New York.

Hurn, A.S. and Lindsay, K.A. (1997). "Estimating the Parameters of Stochastic Differential Equations by Monte Carlo Methods". **Mathematics and Computers in Simulation**, 43(3-6), 495-501.

Kloeden, P.E., Platen, E. and Schurz, H. (1994). *Numerical Solution of SDE Through Computer Experiments*, Spring-Verlag, Berlin.

Nelder, J.A. and Mead, R. (1965). " A simplex method for function minimisation". **The Computer Journal**, 7, 308-313.