Parametric inference for stochastic differential equations by path integration

Christian Skaug
CNR, Istituto per le Applicazioni del Calcolo, Bari, Italy

Abstract

When we use stochastic differential equations as models of financial data that appear as time series, we have to estimate the equation parameters. For complex models this is not straightforward. Approximate maximum likelihood methods are useful tools for this purpose. We suggest the following approach: The likelihood function given by the time series and the parameters is estimated for fixed values of the parameter vector. We apply a standard optimization method that repetedly calls the estimation procedure, with different parameter vectors as arguments, until the optimization converges. The likelihood function is the product of the transition probability densities given by the data. By solving the Fokker-Planck equation associated with the stochastic differential equation, one can obtain these probability densities. Exact, analytical solutions to the Fokker-Planck equation can rarely be found. We therefore apply a path integral method to find approximate solutions. This path integral method is based on the fact that solutions to stochastic differential equations are Markov processes. The time intervals between all the pairs of consecutive data are split in smaller partitions, so that the Euler-Maruyama method is fairly accurate. For all the pairs of data, the total probability law is then applied recursively with the delta distribution given by the first data point as the initial density. The propagating density is represented numerically on a finite, adaptive grid, and the Euler-Maruyama method provides an approximation to the conditional probability density that appears in the total probability law. The method is tested on artificially generated time series with known parameter vectors. It is seen that it yields satisfying parameter estimates for different models in quite reasonable CPU-time, even with thousands of data. It also compares favorably with other methods.
Keywords: stochastic differential equations, parameter estimation, maximum likelihood, Fokker-Planck, path integration.

1 Introduction

Stochastic differential equations (SDEs) have a wide range of applications in science, engineering and economics. They are essential for pricing and hedging financial derivatives and also appear in interest rates modeling. Numerical simulation of SDE sample paths can be done in a large number of ways and there exists a well established theory of order, convergence and stability of these methods (see e. g. [1]).

Whenever an SDE is used as a model, we have to use the empirical data to estimate its parameters. This is an inverse problem: We observe what we assume is the solution at discrete points in time, and then try to figure out the parameters of the equation. It is probably fair to say that the theory of this inverse problem is a bit less developed. For some simple models there exist feasible analytical solutions to the estimation problem based on the maximum likelihood principle. But these models more often than not fail to describe financial data adequately. Realistic models thus require approximate inference methods. These methods are getting increased attention and research seems to develop in many different directions:

- Approximate maximum likelihood estimation
- Martingale estimating functions
- Bayesian analysis (Markov chain Monte Carlo)
- Moment methods
- Kalman filtering
- Neural networks

For details on some of the above methods, see [2], [3], [4] and [5]. The approach to be presented here belongs to the class of approximate maximum likelihood methods for one-dimensional SDEs whose solutions are ergodic stochastic processes. We shall not be concerned with the challenging inference problem for stochastic volatility models, i. e. vector SDEs where some vector components are unobserved.

2 Model and assumptions

Consider an Itô stochastic differential equation

$$dX_t = a(t, X_t;\, \boldsymbol{p})dt + b(t, X_t;\, \boldsymbol{p})dB_t, \tag{1}$$

where B_t is a standard brownian motion and $\boldsymbol{p}$ is a parameter vector in an n-dimensional space. We shall assume sufficient smoothness of the functions a and b, i. e. the problem is well posed and there will always be a solution to eqn. (1). A time series of financial data

$$\{t_0, X_0\}, \{t_1, X_1\}, \ldots, \{t_N, X_N\} \tag{2}$$

is given. We assume that the series (2) is a discrete-time observation of a sample path given by eqn. (1) and that the time step $\Delta t = t_{i+1} - t_i$ is constant. Our aim is to develop a numerical method for obtaining the most likely value of the full

parameter vector $\boldsymbol{p}$ for general functions a and b, and for any vector dimension n. We make no special assumptions that exclude large classes of interesting models.

3 The maximum likelihood principle

Let

$$f(x_{i+1}|x_i;\,\boldsymbol{p}) = \frac{dP\,(X_{i+1} \leq x_{i+1}|X_i = x_i;\,\boldsymbol{p})}{dx_{i+1}} \tag{3}$$

denote the transition probability density associated with the stochastic process X_t. The maximum likelihood method requires that we find the global maximum of

$$L(\boldsymbol{p}) = \prod_{i=0}^{N-1} f(x_{i+1}|x_i;\boldsymbol{p}). \tag{4}$$

For implementation purposes, we seek the global minimum of $-\log(L(\boldsymbol{p}))$ in all our numerical experiments, which is an equivalent problem. If we have a tractable, exact representation of the transition probability density, the optimal $\boldsymbol{p}$ is easily obtained. But exact transition probabilities are generally unknown. They can only be found in the simplest cases. Approximations are thus called for.

4 Approximate maximum likelihood

There exist two schools of thought in approximate maximum likelihood methods. The analytical approach is to approximate the transition probability density by a closed-form expression that shares important properties with the exact distribution. The numerical approach is to approximate the exact distribution on a finite number of grid points. Here we shall subscribe to the latter philosophy.

An idea that comes up immediately is to substitute the exact transition probability density with the closed-form approximation derived from the forward Euler-Maruyama scheme [1]:

$$X_{i+1} = X_i + a(t_i, X_i;\,\boldsymbol{p})\Delta t_i + b(t_i, X_i;\,\boldsymbol{p})\Delta B_i. \tag{5}$$

The approximate maximum likelihood function is thus

$$\prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi b^2(t_i, X_i;\,\boldsymbol{p})\Delta t_i}} \exp\left(\frac{(X_{i+1} - (X_i + a(t_i, X_i;\,\boldsymbol{p})\Delta t_i))^2}{2b^2(t_i, X_i;\,\boldsymbol{p})\Delta t_i}\right). \tag{6}$$

Finding the global maximum of (6) is an easy, quickly solvable problem. But that maximum will be a poor approximation to the true value of $\boldsymbol{p}$ unless the time-steps Δt_i are very small. In practice, samples of X_t are taken less frequently than required for the Euler scheme to provide a sharp estimate of $\boldsymbol{p}$. The relative error with respect to the true value can be high, even orders of magnitude. Introducing some higher order method like the Milstein scheme can improve the situation, but still it will not yield adequate approximations to $\boldsymbol{p}$.

Another idea, suggested by Poulsen [6], is based on the fundamental result from stochastic calculus that the transition probability density is the solution to the Fokker-Planck equation

$$\begin{aligned}\frac{\partial f(x_{i+1}|x_i;\,\boldsymbol{p})}{\partial t} = &-\frac{\partial}{\partial x_{i+1}}\left(a(t_{i+1},x_{i+1};\,\boldsymbol{p})f(x_{i+1}|x_i;\,\boldsymbol{p})\right)\\ &+\frac{1}{2}\frac{\partial^2}{\partial^2 x_{i+1}}\left(b^2(t_{i+1},x_{i+1};\,\boldsymbol{p})f(x_{i+1}|x_i;\,\boldsymbol{p})\right).\end{aligned} \tag{7}$$

In our case, the initial distribution at t_i is the delta "function" centered at X_i. Eqn. (7) is also known as the forward Kolmogorov equation. By solving this equation N times, for the steps between the $N+1$ observations, we obtain the likelihood function $L(\boldsymbol{p})$. Since the Fokker-Planck equation is hard to solve in the general case, we apply a numerical method to find approximations to $L(\boldsymbol{p})$ at *fixed* values of $\boldsymbol{p}$. An approximation to the optimal $\boldsymbol{p}$ can then be found by optimisation. This method has the advantage that it is asymptotically as accurate as the exact maximum likelihood estimator.

5 Optimisation aspects

In practice, the optimisation problem is not trivial to solve efficiently, even if the likelihood function is often well-behaved and smooth without local maxima. When using algorithms like conjugate gradient or Newton's method "blindly", we are likely to encounter problems with different parameter scales and lack of quality of the initial guess for $\boldsymbol{p}$. Without any a priori knowledge of the approximate scales, finite difference approximations to the gradient may be inaccurate. Many fast optimisation methods assume that the object function is similar to a quadratic form near the maximum. While that may be true in a neighbourhood sufficiently close to the maximum, it might not be the case elsewhere. A direct optimisation method may thus be required to localise the maximum's neighbourhood. However, scale differences can remain a problem even as we get closer to the maximum. The optimization algorithm takes tentative steps in different search directions. Without approximate knowledge of the scales, we risk stepping out of the interesting area, or spend a lot of time back-stepping. The likelihood function can also be ill-conditioned. These potential problems, as well as our ambitions of making a robust method, constitute an argument in favour of using a direct optimisation method until convergence, even if a more sophisticated scheme may sometimes be a lot faster. We are then forced to calculate approximations to the likelihood function quite a few times, so it is essential to keep the computational cost of each evaluation down. See [7] for a detailed discussion on optimisation matters.

6 Solving Fokker-Planck by the path integral method

We would like to solve the Fokker-Planck equation numerically, but how? Poulsen [6] uses the Crank-Nicholson finite difference method and obtains satisfying results.

However, numerical experience suggests that finite difference methods often fail if the equation is significantly nonlinear, especially in the deterministic component. The same goes for finite elements. We shall take on an approach that has been successfully applied to Fokker-Planck equations with nonlinear drift terms in stochastic mechanics [8], although we do not test it on financial models with nonlinear drift terms in the present work (that is the topic of ongoing research). Our approach exploits the fact that solutions to SDEs are Markov processes. A Markov process obeys the total probability law

$$f(x_{i+1}|x_i;\boldsymbol{p}) = \int_{-\infty}^{\infty} f(x_{i+1}|\hat{x};\boldsymbol{p}) f(\hat{x}|x_i;\boldsymbol{p}) d\hat{x} \tag{8}$$

for an intermediate $\hat{x}$ at $\hat{t} \in [t_i, t_{i+1}]$. By introducing $m-1$ intermediate points

$$t_i = \hat{t}_0 < \hat{t}_1 < \ldots < \hat{t}_{m-1} < \hat{t}_m = t_{i+1} \tag{9}$$

in time, the right hand side of the total probability law (8) is rewritten as an m-dimensional path integral (the dependence of $\boldsymbol{p}$ is omitted for clarity):

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_{i+1}|\hat{x}_{m-1}) \ldots f(\hat{x}_1|\hat{x}_0) f(\hat{x}_0|x_i) d\hat{x}_0 \ldots d\hat{x}_{m-1} \tag{10}$$

And then the naive Euler approach suddenly becomes useful. If m is large enough to render the Euler-Maruyama scheme sufficiently accurate, it provides analytic approximations to the transition probability densities required by the path integral. The path integral (10) and the Euler method (5) are the core elements of our method. The rest is implementation details, however important they may be.

7 Implementation

We assume that the time steps of the partition (9) are all equal to $\delta t = \Delta t/m$. The first integrand is just a transition density times a delta distribution, so the corresponding integral is carried out analytically. The resulting distribution is represented numerically on a finite grid with 31 equally spaced points. The grid is on an interval whose endpoints are defined by the deterministic drift component of the SDE plus-minus four times the standard deviation given by the stochastic component at X_i. The grid points move adaptively according to the same criteria as the probability density propagates by means of the $m-1$ remaining integrations. These are carried out consecutively for all grid points using Simpson's method, limiting the numerical quadrature to the intervals where the transition probability density as it would have been provided by the backward Euler scheme is greater than 10^{-4}. The number of quadrature points is also equal to 31. As Simpson's method calls for numerical values of the probability density outside the grid, an interpolated value is provided by natural cubic splines [9].

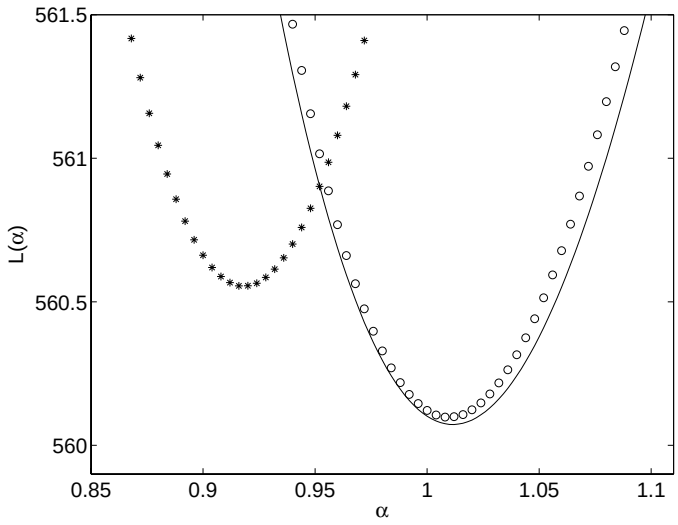

Figure 1: Exact likelihood and the two approximations

8 A simple numerical experiment

Consider the Ornstein-Uhlenbeck (OU) process

$$dX_t = -\alpha X_t dt + \sqrt{2\alpha}\, dB_t. \tag{11}$$

This very simple SDE has an analytical solution to the parameter estimation problem, as the transition probability density is given by

$$f(x_{i+1}|x_i;\, \alpha) = \frac{1}{\sqrt{2\pi(1-e^{-2\alpha\Delta t})}} \exp\left(-\frac{(x_{i+1}-x_i e^{-\alpha\Delta t})^2}{2(1-e^{-2\alpha\Delta t})}\right) \tag{12}$$

Using the exact solution and a random number generator, we provide 20 realizations of eqn. (11) in the interval $t \in [0, 100]$ at 1000 discrete, equally spaced sample points for $\alpha = 1$. The likelihood function is then evaluated on a fine grid at the interval $[0.8, 1.2]$ for all the 20 experiments in the following three ways:

- Exact maximum likelihood
- Path integration with $m = 10$ (ten subintervals)
- Naive Euler method as described above

The mean and standard deviation of the three sets of corresponding parameter estimates are seen in Table 1. For one of the realizations, we have plotted the three corresponding negative log-likelihood functions in Figure 1. The full line corresponds to the exact likelihood, the circles to the approximate likelihood provided by the path integral, and the asterisks to the one provided by the Euler method.

Table 1: Statistics of OU parameter estimates from 20 experiments.

Method	α mean	α standard deviation
Exact maximum likelihood	1.002	0.045
Path integration	1.012	0.051
Naive Euler	0.919	0.039

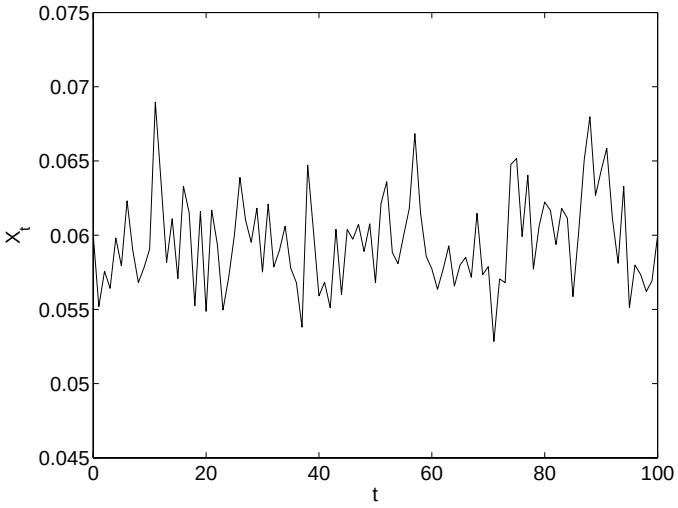

Figure 2: Part of a sample of the CIR process

It is seen that path integration provides almost as sharp an estimate as exact maximum likelihood. That was perhaps to be expected with such a high data density, but note that the Euler method is biased even for such a small time step.

9 An experiment with the CIR process

We now try out the full estimation method on the so-called Cox-Ingersoll-Ross (CIR) model, given by

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t. \tag{13}$$

Eqn. (13) has been suggested as an interest rates model [10]. We have simulated 100 different realizations of the CIR model on the interval $[0, 2000]$ using the Milstein scheme and a time step equal to 0.01 with the parameter vector $(\alpha, \kappa, \sigma) = (0.06, 2, 0.025)$. From this realization we have sampled 2000 data, i. e. the time step between the "observed" data is 1.0. Part of one of these samples is seen in Figure 2. For all the samples, we have fed the data into the path

Table 2: Statistics of CIR parameter estimates from 100 experiments.

Parameter	Mean	Standard deviation
α	0.05998	0.00013
κ	1.929	0.15
σ	0.02386	0.00084

integration procedure to be called by the Simplex direct optimisation method [9]. First we run the Simplex method with initial parameter vector $(1, 1, 1)$ to convergence using a path integral method with two subintervals. The result is then used as the initial value in another call to Simplex, this time with four path integration subintervals. The subintervals are doubled until Simplex provides two consecutive estimates that differ by less than five percent in the L_2 metric (this usually happens at sixteen subintervals). The resulting parameter estimate statistics are shown in Table 2. For one of the samples, we have evaluated the approximate likelihood function with sixteen partitions on a 3D parameter grid to check by brute force that Simplex actually comes up with the global minimum. A contour plot of the likelihood for α fixed at 0.06 is shown in Figure 3.

It is not surprising that α is estimated sharply, as it is very close to the time series' mean value. The estimate of the diffusion parameter σ is also close to the mark. It turns out that the most difficult parameter to estimate is the speed of mean reversion κ. The mean estimate of κ is between three and four percent biased and it has the largest relative standard deviation (about twice that of σ). These results are of slightly inferior quality with respect to the ones provided by Sørensen [11], who develops a special Martingale method for small diffusion terms. In a similar experiment with the parameter vector $(5, 2, 0.01)$, his mean estimate for κ is 2.0183 with standard deviation 0.1152. But the time step between his "observed" data is 0.1, that is ten times denser. Keeping in mind that our method is completely general, it provides a very good result. Mean CPU time for our calculations is about ten minutes on a Dell PC with a 2.6GHz processor, which is a decent computational cost considering the accuracy of the results.

10 Conclusion and suggestions for further research

We have studied the parameter estimation problem for one-dimensional stochastic differential equations that are observed at discrete points in time. It is seen that even with a large number of data points, a relatively sharp estimate of the unknown parameter vector can be found in quite reasonable CPU time by repeatedly solving the Fokker-Planck equation for the transition probability densities of the maximum likelihood function by the path integral method at fixed parameter values, applying standard direct optimisation methodology to find the optimal parameter vector. The suggested algorithm should be tested on more complicated financial models,

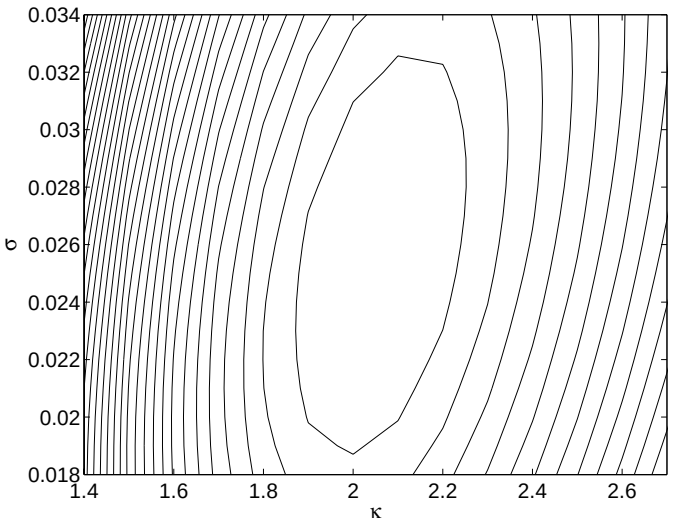

Figure 3: Contour plot of the approximate likelihood as a function of κ and σ with α fixed at 0.06.

also models with nonlinear drift terms, to assess more thoroughly its potential as a general parameter estimation method for stochastic differential equations. It should also be possible to reduce CPU time, as very little has so far been done to optimise the implementation or speed up the code. Perhaps a higher order approximation for the transition probability density can also save time.

References

[1] Kloeden, P.E. & Platen, E., *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag: Berlin and New York, 1992.

[2] Pedersen, A.R., A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, **22**, pp. 55–71, 1995.

[3] Bibby, B.M. & Soerensen, M., Martingale estimating functions for discretely observed diffusion processes. *Bernoulli*, **1**, pp. 17–39, 1995.

[4] Eraker, B., Mcmc analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, **19(2)**, pp. 177–191, 2001.

[5] Gallant, A.R. & Long, J.R., Estimating stochastic differential equations efficiently by minimum chi-square. *Biometrika*, **84**, pp. 125–142, 1997.

[6] Poulsen, R., Approximate maximum likelihood estimation of discretely observed diffusion processes. *Working Paper Series No 29, Centre for Analytical Finance, University of Århus*, 1999.

[7] Kolda, T.G., Lewis, R.M. & Torczon, V., Optimization by direct search: New

perspectives on some classical and modern methods. *SIAM Review*, **45(3)**, pp. 385–482, 2003.

[8] Naess, A., Predictions of extreme response of nonlinear oscillators subjected to random loading using the path integral solution technique. *Journal of Research of the National Institute of Standards and Technology*, **99(4)**, 1994.

[9] Press, W.H., *Numerical recipes in Fortran 77: The art of scientific computing*. Cambridge University Press: Cambridge, 1996.

[10] Cox, J.C., Ingersoll, J.E. & Ross, S.A., A theory of the term structure of interest rates. *Econometrica*, **53**, pp. 385–407, 1985.

[11] Soerensen, M., Small dispersion asymptotics for diffusion martingale estimating functions. *Preprint 2000–2, Department of Applied Mathematics and Statistics, University of Copenhagen*, 2000.