Parametric estimation of discretely observed diffusions using the EM algorithm

Omiros Papaspiliopoulos, Giorgos Sermaidis
Department of Statistics, University of Warwick, UK

Keywords: EM algorithm, diffusion processes, random weights
AMS: 60J80

Abstract

In this paper we report ongoing work on parametric estimation for diffusions using Monte Carlo EM algorithms. The work presented here has already been extended to high dimensional problems and models with observation errors.

1. Introduction

Diffusion processes have become an extensively used tool to model continuous time phenomena in many scientific areas. A diffusion process V is defined as the solution of a SDE of the form

$$\mathrm{d}V_s = b(V_s;\theta)\,\mathrm{d}s + \sigma(V_s;\theta)\,\mathrm{d}B_s, \tag{1}$$

driven by the scalar Brownian motion B. The functionals $b(\,\cdot\,;\theta)$ and $\sigma(\,\cdot\,;\theta)$ are called the *drift* and the *diffusion coefficient* respectively and are allowed to depend on some parameters $\theta \in \Theta$. These functionals are presumed to satisfy the regularity conditions that guarantee a weakly unique, global solution of equation (1); see chapter 4 of [1] for further details. In this paper we consider only one-dimensional diffusions, although multivariate extensions are possible.

Altough the process is defined in continuous time, the available data are always sampled in discrete time. We assume that the process is observed without error at a given collection of time instances

$$\mathbf{v} = \{V_{t_0}, V_{t_1}, \ldots, V_{t_n}\}, \quad 0 = t_0 < t_1 < \cdots < t_n$$

and denote the time increments between consecutive observations by $\Delta t_i = t_i - t_{i-1}$ for $1 \le i \le n$.

The log-likelihood of the dataset $\mathbf{v}$ is given by

$$l(\theta \mid \mathbf{v}) = \sum_{i=1}^{n} \log\{p_{\Delta_{t_i}}(V_{t_i}, V_{t_{i-1}};\theta)\},$$

where $p_t(v, w;\theta)$ is the transition density:

$$p_t(v, w;\theta) = P(V_t \in \mathrm{d}w \mid V_0 = v;\theta)/\mathrm{d}w, \quad t > 0,\ w, v \in \mathbf{R}.$$

Estimation of the parameters of the diffusion process via maximum likelihood (ML) is hard since the transition density is typically not analytically available. On the other hand, ML estimation for continuous observed $\mathbf{v}$ is straightforward since a likelihood can be easily derived (see (3)). This suggests treating the paths between consecutive observations as missing data and proceed to parameter estimation using the Expectation-Maximization algorithm.

2. An Expectation-Maximization approach

The EM algorithm is a general purpose algorithm for maximum likelihood estimation in a wide variety of situations where the likelihood of the observed data is intractable but the joint likelihood of the observed and missing data has a simple form (see [2]). The algorithm works for any augmentation scheme however appropriately defined missing data can lead to easier and more efficient implementation. Without loss of generality we assume two data points and denote the observed data by $\mathbf{v}_{obs} = \{V_0 = u, V_t = w\}$.

The aim is to appropriately transform the missing paths so that the complete likelihood can be written in an explicit way. This is achieved by applying two transformations. First, we transform $V_s \to \eta(V_s) =: X_s$, where

$$\eta(u;\theta) = \int^{u} \frac{1}{\sigma(z;\theta)}\, dz\ .$$

Applying Itô's rule to the transformed process we get that

$$\mathrm{d}X_s = \alpha(X_s;\theta)\,\mathrm{d}s + \mathrm{d}B_s, \quad X_0 = x,\ s \in [0,t], \tag{2}$$

where

$$\alpha(u;\theta) = \frac{b\{\eta^{-1}(u;\theta);\theta\}}{\sigma\{\eta^{-1}(u;\theta);\theta\}} - \sigma^{'}\{\eta^{-1}(u;\theta);\theta\}\,/\,2, \quad u \in \mathbf{R};$$

η^{-1} denotes the inverse transformation and $\sigma^{'}$ denotes the derivative w.r.t. the space variable. Note that the starting and ending point are functions of θ as $x = \eta(u;\theta)$ and $y = \eta(w;\theta)$. Let

$$A(u;\theta) := \int^{u} \alpha(z;\theta)\mathrm{d}z$$

be any anti-derivative of α and $\tilde{p}_t(x,y;\theta)$ be the transition density of X. Second, we apply the transformation $X_s \to \dot{X}_s$ where:

$$\dot{X}_s := X_s - (1-\frac{s}{t})\,x(\theta) - \frac{s}{t}\,y(\theta), \quad s \in [0,t].$$

$\dot{X}$ is a diffusion bridge starting from $\dot{X}_0 = 0$ and finishing at $\dot{X}_t = 0$ and its dynamics depend on θ. The inverse transformation from $\dot{X}$ to X is:

$$g_\theta(\dot{X}_s) := \dot{X}_s + (1-\frac{s}{t})\,x(\theta) + \frac{s}{t}\,y(\theta), \quad s \in [0,t].$$

The augmentation scheme for the EM algorithm is $\mathbf{v}_{mis} = (\dot{X}_s, s \in [0,t])$ for the missing data and $\mathbf{v}_{com} = \{\mathbf{v}_{obs}, \mathbf{v}_{mis}\}$ for the complete data.

Let $\mathbb{W}^{(t,x,y)}$ be the standard Brownian Bridge (BB) measure and $\mathbb{Q}_\theta^{(t,x,y)}$ the corresponding measure induced by the diffusion process (2). The conditional density of $\dot{X}$ on the observed data (w.r.t. the standard BB measure) is derived using Lemma 2 in [3] as:

$$\pi(\dot{X} \mid \mathbf{v}_{obs}, \theta) = \frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}} \left\{ g_\theta(\dot{X}_s) \right\} = G_\theta \left\{ g_\theta(\dot{X}_s) \right\}$$

$$G_\theta(\omega) = \frac{\mathcal{N}_t(y-x)}{\tilde{p}_t(x,y;\theta)} \exp\left\{ A(y;\theta) - A(x;\theta) - \int_0^t \frac{1}{2}(\alpha^2 + \alpha^{'})(\omega_s;\theta)\mathrm{d}s \right\}, \quad (3)$$

where ω is any path starting at x at time 0 and finishing at y at time t and $\mathcal{N}_t(u)$ is the density of the normal distribution with mean 0 and variance t evaluated at $u \in \mathbf{R}$. The complete log-likelihood can now be written as

$$\begin{aligned} l(\theta \,|\, \mathbf{v}_{com}) \quad = \quad & \log|\eta'(w;\theta)| + \log[\mathcal{N}_t\{y(\theta) - x(\theta)\}] + A\{y(\theta);\theta\} - A\{x(\theta);\theta\} \\ & - \int_0^t \frac{1}{2}(\alpha^2 + \alpha')\{g_\theta(\dot{X}_s);\theta\}\,\mathrm{d}s. \end{aligned}$$

It is helpful to introduce the random variable $U \sim \mathrm{Un}[0,t]$, which is independent of $\mathbf{v}_{com}$. The E-step requires analytic evaluation of

$$\begin{aligned} Q(\theta,\theta') = \mathbb{E}_{\mathbf{v}_{mis}|\mathbf{v}_{obs};\theta'}\left[\, l(\theta \,|\, \mathbf{v}_{com}) \,\right] = \log|\eta'(w;\theta)| + \log\mathcal{N}_t\{y(\theta) - x(\theta)\} \\ +A\{y(\theta);\theta\} - A\{x(\theta);\theta\} - \frac{t}{2}\mathbb{E}_{(\mathbf{v}_{mis},U)|\mathbf{v}_{obs};\theta'}\left[\,(\alpha^2 + \alpha^{'})(g_\theta(\dot{X}_U);\theta)\,\right]. \quad (4) \end{aligned}$$

The expression above cannot be evaluated analytically so we will estimate it by Monte Carlo. This suggests maximizing the estimator $\tilde{Q}(\theta,\theta')$ of $Q(\theta,\theta')$ which leads to a Monte Carlo Expectation Maximization algorithm (MCEM).

2.1. Exact Simulation

One solution to the problem would be given by simulating M samples from $\dot{X}_U$ and estimating the intractable expectation with its sample average. Recent advances in simulation methodology have made exact simulation (ES) of diffusion processes feasible. The technique is called Retrospective Sampling and the resulting algorithm Exact Algorithm (EA). The simulation is based on rejection sampling; details are not presented in this paper but the reader can see [3] for further details. So the $\theta^{(s+1)}$ estimate of the MCEM algorithm is found as follows:

1. `Simulate` $u_i \sim$ `Un`$[0,t]$, `for` $i = 1,2,\ldots,M$.
2. `Simulate` $x_{u_i} \sim \mathbb{Q}_{\theta^{(s)}}^{(t,x,y)}$ `using the EA.`
3. `Apply the transformation in (2).`

4. `Get` $\theta^{(s+1)}$ `by maximizing`

$$\begin{aligned}\tilde{Q}(\theta,\theta^{(s)}) &= \log|\eta'(w;\theta)| + \log\mathcal{N}_t\{y(\theta)-x(\theta)\} + A\{y(\theta);\theta\} \\ &\quad -A\{x(\theta);\theta\} - \frac{t}{2M}\sum_{i=1}^{M}(\alpha^2+\alpha')\{g_\theta(\dot{x}_{u_i});\theta\}.\end{aligned}$$

2.2. Importance Sampling

A drawback of estimating $Q(\theta,\theta')$ with the EA is that the latter is based on rejection sampling which implies the simulation of a larger number of samples than M that we actually use. This observation suggests to proceed to an Importance Sampling (IS) estimation for the unknown expectation in (4). The idea is to simulate M samples from a standard Brownian Bridge and weight them by the likelihood ratio (3). This expression cannot be calculated analytically. However, following the approach by [4] we only need unbiased estimates of the weights which can be derived using the Poisson Estimator (see [3]). Let

$$f(\dot{X}_s;\theta) = \frac{1}{2}(\alpha^2+\alpha^{'})\{g_\theta(\dot{X}_s);\theta\}.$$

Let $c\in\mathbb{R},\ \lambda>0$ be user-specified constans, and $\Psi=\{\psi_1,...\psi_\kappa\}$ be a Poisson process of rate λ on $[0,t]$. Then it can be checked that

$$\exp\left\{-\int_0^t f(\dot{X}_s)ds\right\} = \exp\{(\lambda-c)t\}\mathbb{E}\left(w_\theta \mid \dot{X}\right), w_\theta = \prod_{j=1}^{\kappa}\frac{c-f(\dot{X}_{\psi_j};\theta)}{\lambda}.$$

So, the $\theta^{(s+1)}$ estimate of the MCEM algorithm using IS is derived by:

1. `Simulate` $\kappa_i \sim Po(\lambda t),$ `for` $i=1,2,\ldots,M$.
2. `Simulate` $\psi_j^{(i)} \sim \mathtt{Un}[0,t],$ `for` $j=1,2,\ldots,\kappa_i+1$.
3. `Simulate` $\dot{x}_{\psi_j^{(i)}} \sim \mathbb{W}^{(t,0,0)}$.
4. `For every` i `use first` κ_i `samples to estimate the weight` $w^{(i)}_{\theta^{(s)}}$.
5. `Get` $\theta^{(s+1)}$ `by maximizing`

$$\begin{aligned}\tilde{Q}(\theta,\theta^{(s)}) &= \log|\eta'(w;\theta)| + \log\mathcal{N}_t\{y(\theta)-x(\theta)\} + A\{y(\theta);\theta\} \\ & -A\{x(\theta);\theta\} - \frac{t\sum_{i=1}^{M}(\alpha^2+\alpha')\left\{g_\theta\left(\dot{x}_{\psi^{(i)}_{\kappa_i+1}}\right);\theta\right\}w^{(i)}_{\theta^{(s)}}}{2\sum_{i=1}^{M}w^{(i)}_{\theta_{(s)}}}.\end{aligned}$$

3. An illustrative example

The algorithms are applied to an Ornstein-Uhlenbeck (OU) process for which $b(v;\theta) = -\theta v,\ \sigma(v;\theta) = 1.0,\ \theta \in (0, +\infty)$. This process has been chosen for illustration since the likelihood is available in this case. A sample size of 1000 observations was simulated from the process using $\theta = 2.0$ and $V_0 = 0.0$. The algorithms were performed using $M = 100$ for the first five iterations and $M = 1000$ for the last five. The initial value was $\theta^{(0)} = 1.0$, the MLE was 2.00274 and $\theta^{(10)}$ was 2.00107 and 1.99918 for the ES and IS implementations respectively. The results are shown in Figure 1.

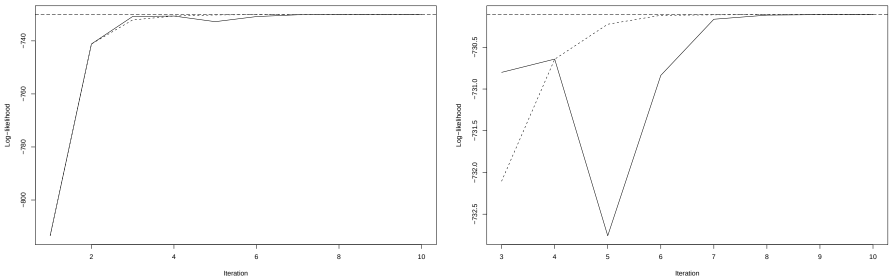

Figure 1: The log-likelihood evaluated at every EM iteration. The dotted line represents the ES method, the solid line the IS method and the dashed line the log-likelihood evaluated at the true MLE. The right panel zooms at the last seven iterations.

4. Bibliography

[1] Kloeden, P. and Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations*. New York: Springer.

[2] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1997) *Maximum likelihood from incomplete data via the em algorithm (with discussion)*. J. R. Statist. Soc. B, 39, 1–38.

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

[4] Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2006) *Particle filters for partially observed diffusions*. In revision.