Stat Comput (2012) 22:615–623
DOI 10.1007/s11222-011-9255-y

A regularized bridge sampler for sparsely sampled diffusions

Erik Lindström

Received: 29 September 2010 / Accepted: 27 April 2011 / Published online: 15 May 2011

Abstract Sparsely sampled diffusion processes, in this paper interpreted as data sampled sparsely in time relative to the time constant, is a challenging statistical problem. Most approximations of the transition kernel are derived under the assumption that data is frequently sampled and these approximations are often severely biased for sparsely sampled data. Monte Carlo methods can be used for this problem as the transition density can be estimated with arbitrary accuracy regardless of the sampling frequency, but this is computationally expensive or even prohibited unless effective variance reduction is applied.

The state of art variance reduction technique for diffusion processes is the Durham-Gallant (modified) bridge sampler. Their importance sampler is derived using a linearized, Gaussian approximation of the dynamics, and have proved successful for frequently sampled data. However, the approximation is often not valid for sparsely sampled data.

We present a flexible, alternative derivation of the modified bridge sampler for multivariate, discretely observed diffusion models and modify it by taking uncertainty into account. The resulting sampler can be viewed as a combination of the basic sampler and the Durham-Gallant sampler, using the sampler that is most appropriate for the given problem, while still being computationally efficient. Our sampler is providing effective variance reduction for frequently and sparsely sampled data.

Keywords Bridge sampler · Time series · Diffusion processes · Monte Carlo methods · MCMC

E. Lindström (✉)
Division of Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Box 118, 22100 Lund, Sweden
e-mail: erikl@maths.lth.se

Diffusion processes, or Brownian motion driven stochastic differential equations is a class of stochastic processes defined as the solution to the stochastic differential equation defined on the space X

$$\mathrm{d}\mathbf{X}_t = f(t, \mathbf{X}_t)\mathrm{d}t + g(t, \mathbf{X}_t)\mathrm{d}\mathbf{W}_t, \tag{1}$$

where the drift $f(t, \mathbf{X}_t)$ and diffusion $g(t, \mathbf{X}_t)$ are known functions that are assumed to be sufficiently regular (Lipschitz, bounded growth) for existence and uniqueness, cf. Øksendal (2000). Diffusion processes are the natural extension of ordinary differential equations and are being used for a variety of applications, including finance, see e.g. Heston (1993), Lindström (2010), biology, see Philipsen et al. (2010), hydrology, see Jonsdottir et al. (2006) and engineering, see Madsen and Holst (1995).

Estimating parameters when data is frequently sampled is comparably easy. It is well known that non-parametric methods such as realized volatility and variations thereof, see e.g. Andersen et al. (2003), can be used to estimate quantities related to the diffusion term. These estimates can later be used to estimate parameters in the drift term, see Phillips and Yu (2009).

Another approach is to approximate the diffusion process with a discrete model, using e.g. the Euler-Maruyama scheme, to obtain approximations of Maximum Likelihood estimates. The estimates generated from this approximation are known as Discrete Maximum Likelihood (DML) estimates. We will throughout the paper assume that the time distance between two consecutive observations is significant relative to the time constant of the system. This makes approximations such as the DML inaccurate, often leading to bias and inefficiency.

There exist a large number of methods for improving the DML, see Hurn et al. (2007). The purpose of these is to find

alternative approximations that are unbiased or at least less biased. These include series expansions, numerical solution of the Fokker-Planck equation and Monte Carlo methods.

Series expansions of the transition kernel for univariate diffusions was introduced in Aït-Sahalia (2002) and extended to multivariate diffusions in Aït-Sahalia (2008). The class of processes for which the transition density can be expanded is rather large but does have some restrictions, especially for multivariate models. The performance is often excellent for densely sampled data but can be quite poor for sparsely sampled data, as the leading error term is a power function of the sampling distance.

Numerical approximation of the solution to the Fokker-Planck equation was studied in Lo (1988) and improved by Lindström (2007). This approach works for practically any sampling distance and is often very accurate for univariate models, cf. Lindström (2007). However, generalizing the methodology to multivariate diffusions is still an open research problem due to the curse of dimensionality.

Monte Carlo methods such as Simulated Maximum Likelihood (SML) was introduced in Pedersen (1995a, 1995b) and dramatically improved by Durham and Gallant (2002). It uses Monte Carlo simulation and integration to approximate the transition kernel. The efficient sampler in Durham and Gallant (2002) was extended to multivariate models in Golightly and Wilkinson (2006) and partially observed multivariate models (with additive Gaussian noise) in Golightly and Wilkinson (2008). Simulated Maximum Likelihood estimators are often biased, and this can be eliminated by using exact simulation or multi-level approximations. Exact simulation uses rejection sampling to generate exact draws from a class of diffusion processes, see Beskos et al. (2006, 2009). The efficiency of the method is related to the acceptance probabilities for the draws which may be very small for sparsely sampled data.

Series expansions and PDE-methods will not be efficient for sparsely sampled, high-dimensional models which is why we focus on simulation based techniques in this paper. These methods should in theory be able to handle high-dimensional, infrequently sampled data, but this has not been the case for many practical problems. In fact, it was shown by Fearnhead (2008) that the state of art Durham-Gallant sampler is less efficient than the naive sampler for sparsely sampled data. The inefficiency of the Durham-Gallant sampler can be traced to the poor approximation of the transition kernel. It is known, see Glasserman (2003), that importance sampling can increase the variance of the estimate if the proposal density is badly chosen.

Sparse sampling of diffusions is therefore associated with two related problems, as the comparably powerful drift term not only decreases the efficiency of the Durham-Gallant sampler but also leads to more bias. The bias can be compensated by additional computational efforts. It was shown in Stramer and Yan (2007) how to balance the bias and the variance to achieve optimal results for bridge samplers, and they find that finer discretization must be balanced with larger Monte Carlo samples for optimal convergence. The complexity is even worse for the naive sampler. Some robust variance reduction method (e.g. a bridge sampler) is therefore needed to prevent additional computational demands.

MCMC algorithms for diffusions should update blocks rather than single components, as the latter was shown to cause the mixing to break down, cf. Eraker (2001) and Elerian et al. (2001) for random block sampling. Bridge samplers are useful for generating these blocks, making this paper applicable not only to discretely observed models, but to models where MCMC techniques are needed.

This paper is organized as follows. Section 1 introduces the simulated maximum likelihood method, including the Durham-Gallant framework while Sect. 2 introduces a regularization that makes the bridge sampler more robust against sparsely sampled data. Different samplers are compared in Sects. 3 and 4 concludes the paper.

1 Simulated maximum likelihood

The Simulated Maximum Likelihood (SML) method was introduced in Pedersen (1995b), and uses Monte Carlo simulations to decrease bias.

Assume that data is sampled (not necessarily equidistant) at time points $t_1, \ldots, t_N$. The idea is to introduce a partition $t_n = \tau_0 < \tau_1 < \cdots < \tau_{M-1} < \tau_M = t_{n+1}$ and to use the law of total probability and the Markov property to obtain

$$\begin{aligned} p(\mathbf{x}_{t_{n+1}}|\mathbf{x}_{t_n}) &= \int p(\mathbf{x}_{\tau_M}, \mathbf{x}_{\tau_{M-1}}, \ldots, \mathbf{x}_{\tau_1}|\mathbf{x}_{\tau_0})\mathrm{d}\mathbf{x}_{\tau_1:\tau_{M-1}} \\ &= \int \prod_{m=1}^{M} p(\mathbf{x}_m|\mathbf{x}_{m-1})\mathrm{d}\mathbf{x}_{\tau_1:\tau_{M-1}} \\ &= \mathbb{E}[p(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_{M-1}})|\mathbf{x}_{\tau_0}] \\ &\approx \frac{1}{K}\sum_{k=1}^{K} p(\mathbf{x}_{\tau_M}|\xi_{\tau_{M-1},k}) \end{aligned} \tag{2}$$

where $\xi_{\tau_{M-1},k}$ are random variables drawn from $\xi_{\tau_{M-1},k} \sim p(\mathbf{x}_{\tau_{M-1}}|\mathbf{x}_{\tau_0})\mathrm{d}\mathbf{x}_{\tau_{M-1}}$. The third line provides a Monte Carlo estimate of the transition probability density.

This approximation is easy to implement but is often computationally inefficient, cf. Durham and Gallant (2002). More efficient samplers such as the Durham-Gallant sampler have now replaced the Pedersen sampler as the preferred approximation.

1.1 Durham-Gallant sampler

The Durham-Gallant sampler, Durham and Gallant (2002), is derived by treating the approximation of the transition

density as a Monte Carlo estimate of an expectation. Variance is reduced by applying a near optimal importance sampler based on the *linearized* dynamics

$$p(\mathbf{x}_{t_{n+1}}|\mathbf{x}_{t_n}) = \int \frac{p(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_{M-1}})\prod_{m=1}^{M-1} p(\mathbf{x}_m|\mathbf{x}_{m-1})}{q(\mathbf{x}_{\tau_1:\tau_{M-1}})} \times q(\mathbf{x}_{\tau_1:\tau_{M-1}})\mathrm{d}\mathbf{x}_{\tau_1:\tau_{M-1}}. \quad (3)$$

We present an alternative derivation of the near optimal importance sampler, cf. Durham and Gallant (2002) for univariate models and Golightly and Wilkinson (2006) for multivariate models that will be general enough for the extension we are presenting in Sect. 2.

It is well known that the optimal importance sampler is proportional to the integrand multiplied with the integrator, here given by

$$q_{Opt}(\mathbf{x}_{\tau_{M-1}:\tau_1}) \propto p(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_{M-1}}) \prod_{m=1}^{M-1} p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}). \quad (4)$$

It is straightforward to show that the optimal importance sampler is given by

$$q(\mathbf{x}_{\tau_{M-1}:\tau_1}) \propto \prod_{m=1}^{M-1} p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_M}, \mathbf{x}_{\tau_{m-1}}). \quad (5)$$

The distribution of the sampler can be found using Bayes' theorem in logarithmic form, leading to

$$\log p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_M}, \mathbf{x}_{\tau_{m-1}}) \propto \log p(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_m}) + \log p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}). \quad (6)$$

Assuming that these transition densities are Gaussian and approximating the dynamics as

$$p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}) = \phi(\mathbf{x}_{\tau_m}; \mathbf{a}, \mathrm{P}), \quad (7)$$

$$p(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_m}) = \phi(\mathbf{x}_{\tau_M}; \mathbf{x}_{\tau_m} + \mathbf{b}, \mathrm{Q}), \quad (8)$$

where $\phi(x, \mu, \Sigma)$ is the density of a multivariate Gaussian random variable with mean μ and covariance Σ makes it possible to compute the mean and covariance for the resulting Gaussian distribution.

The Durham-Gallant paper uses several discretization schemes to approximate the transition densities, and the most commonly used approximation is the (non-Markovian) Euler-Maruyama approximation

$$\mathbf{a} = \mathbf{x}_{\tau_{m-1}} + f(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})(\tau_m - \tau_{m-1}) \quad (9)$$

$$\mathbf{b} = f(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})(\tau_M - \tau_m) \quad (10)$$

$$\mathrm{P} = g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})^T(\tau_m - \tau_{m-1}) \quad (11)$$

$$\mathrm{Q} = g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})^T(\tau_M - \tau_m) \quad (12)$$

It is possible to compute the near-optimal importance sampler using (7) and (8). This is done by matching terms in (6), resulting in Gaussian distribution for $p(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_M}, \mathbf{x}_{\tau_{m-1}}) = \phi(\mathbf{x}_{\tau_m}; \mu, \Sigma)$ with parameters

$$\mu = \mathbf{a} + K_0 t(\mathbf{x}_{\tau_M} - \mathbf{a} - \mathbf{b}), \quad (13)$$

$$\Sigma = (\mathrm{P}^{-1} + \mathrm{Q}^{-1})^{-1} = (I - K_0)\mathrm{P}, \quad (14)$$

where $K_0 = \mathrm{P}(\mathrm{P} + \mathrm{Q})^{-1}$. The matrix K_0 (sometimes known as the Kalman gain in the system identification community) is balancing the prior information with the information obtained from $\mathbf{x}_{\tau_M}$.

Using the transition kernels specified by (9)–(12) leads to the well known modified Brownian Bridge sampler

$$\mu = \mathbf{x}_{\tau_{m-1}} + \frac{\mathbf{x}_{\tau_M} - \mathbf{x}_{\tau_{m-1}}}{\tau_M - \tau_{m-1}}(\tau_m - \tau_{m-1}), \quad (15)$$

$$\Sigma = \frac{\tau_M - \tau_m}{\tau_M - \tau_{m-1}} g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})g(\tau_{m-1}, \mathbf{x}_{\tau_{m-1}})^T \times (\tau_m - \tau_{m-1}), \quad (16)$$

see Durham and Gallant (2002).

1.2 Fearnhead sampler

It was suggested in Fearnhead (2008) to replace the standard approximation of the transition density for a geometrically mixing, ergodic diffusion processes with a combination of the Euler-Maruyama approximation and the stationary distribution, π weighted according to the mixing rate ρ. The approximation is then given by a mixture of the naive dynamics and the stationary distribution

$$\hat{p}(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_m}) = \exp(-\rho(\tau_M - \tau_m))p_{Euler}(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_m}) + (1 - \exp(-\rho(\tau_M - \tau_m)))\pi(\mathbf{x}_{\tau_M}). \quad (17)$$

Inserting this approximation into (6) leads to

$$q_F(\mathbf{x}_{\tau_m}) \propto (\exp(-\rho(\tau_M - \tau_m))p_{Euler}(\mathbf{x}_{\tau_M}|\mathbf{x}_{\tau_m}) + (1 - \exp(-\rho(\tau_M - \tau_m)))\pi(\mathbf{x}_{\tau_M})) \times p_{Euler}(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}) \quad (18)$$

where the first term is proportional to the Durham-Gallant sampler and the second term is proportional to the naive dynamics. Simplifying this expression further results in sampler of the form

$$q_F(\mathbf{x}_{\tau_m}) \propto \exp(-\rho(\tau_M - \tau_m))q_{DG}(\mathbf{x}_{\tau_m}) + B(1 - \exp(-\rho(\tau_M - \tau_m)))p_{Euler}(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}), \quad (19)$$

where q_{DG} is the Durham-Gallant sampler, p_{Euler} the naive (Euler-Maruyama) dynamics, ρ is the mixing coefficient and B is assumed to be constant. This sampler was shown to work well for a one-dimensional model (CIR) in Fearnhead (2008), and we refer to it as the Fearnhead sampler.

2 A regularized sampler

A limitation with the Durham-Gallant sampler is that it provides very poor approximations when the dynamics of the model is dominated by the drift term. This can be seen from the modified bridge sampler

$$\mathbf{x}_{\tau_m} = \mathbf{x}_{\tau_{m-1}} + \frac{\mathbf{x}_{\tau_M} - \mathbf{x}_{\tau_{m-1}}}{\tau_M - \tau_{m-1}}(\tau_m - \tau_{m-1}) + \Sigma^{1/2} z_{m+1}, \quad (20)$$

where $\Sigma^{1/2}$ is the Cholesky decomposition of Σ and z is a standard normal random vector. The sampler is independent of the drift which often is problematic for sparsely sampled and/or high-dimensional models.

Example 1 We use the well-known chaotic Lorenz 63 model as a test model, see Lorenz (1963). The stochastic version of the Lorenz 63 model, see Bengtsson et al. (2003) is given by

$$\begin{pmatrix} \mathrm{d}X_t^{(1)} \\ \mathrm{d}X_t^{(2)} \\ \mathrm{d}X_t^{(3)} \end{pmatrix} = \begin{pmatrix} s(X_t^{(2)} - X_t^{(1)}) \\ rX_t^{(1)} - X_t^{(2)} - X_t^{(1)} X_t^{(3)} \\ X_t^{(1)} X_t^{(2)} - bX_t^{(3)} \end{pmatrix} \mathrm{d}t + \sigma \begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \\ \mathrm{d}W_t^{(3)} \end{pmatrix} \quad (21)$$

Commonly used parameters for this model are $s = 10$, $r = 28$, $b = 8/3$ and $\sigma = 2$, see e.g. Bengtsson et al. (2003).

The expected path computed for the deterministic system and the Durham-Gallant sampler is presented in Fig. 1. It can be seen that the sampler ignores the dynamics of the model, and generates paths far from the actual realization. This results in a poor approximation of the expectation, i.e. (2).

The limitation of the Durham-Gallant sampler is that it approximates the dynamics for all time horizons with a simple Euler-Maruyama discretization. It is well known that the Euler-Maruyama scheme is a weak order 1.0 scheme, meaning that the bias is proportional to the time interval used. In other words: the bias in the approximation is substantial for longer time horizons, and it is likely that the bias is causing most of the problem.

It is well-known that, cf. Geweke (1989), Koopman et al. (2009) that it is worse to have an importance density that is too light-tailed than to have a too heavy-tailed density. We will use this insight to suggest a modification of the sampler that will simultaneously eliminate most of the path generation problem.

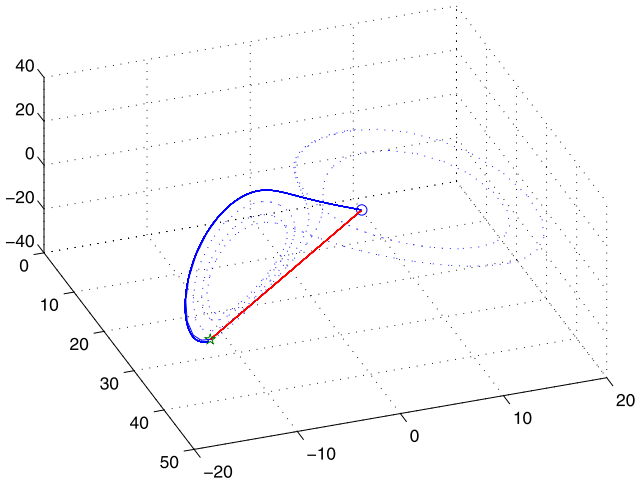

Fig. 1 (Color online) A deterministic trajectory in the Lorenz model (*blue solid line*) and the expected path generated from Durham-Gallant sampler (*red dashed line*). The initial point is displayed by the *circle* and the end point by a *pentagon*

The solution is found by studying the conditional mean squared error (MSE), i.e. the sum of the variance and squared bias. It is known for the Euler scheme that the variance and bias are given by

$$\mathrm{Var}(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t) \approx g(t, \mathbf{x}_t) g(t, \mathbf{x}_t)^T \Delta t \quad (22)$$

$$\mathrm{Bias}(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t) \approx c\Delta t \quad (23)$$

where c is some column vector that we may or may not be able to compute. The MSE is therefore given by

$$\mathrm{MSE}(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t) = g(t, \mathbf{x}_t) g(t, \mathbf{x}_t)^T \Delta t + cc^T (\Delta t)^2. \quad (24)$$

A first suggestion is to replace the conditional variances (in (7) and (8)) with the conditional MSE.

$$\tilde{\mathrm{P}} = \mathrm{P} + cc^T (\tau_m - \tau_{m-1})^2 \quad (25)$$

$$\tilde{\mathrm{Q}} = \mathrm{Q} + cc^T (\tau_M - \tau_m)^2 \quad (26)$$

This does not work well for large values of cc^T, as it can be shown that the resulting conditional variance Σ can take any value between the Durham-Gallant conditional variance and infinity.

The naive sampler is sufficiently heavy tailed and we found that it was more robust to use

$$\hat{\mathrm{P}} = \mathrm{P} \quad (27)$$

$$\hat{\mathrm{Q}} = \mathrm{Q} + cc^T (\tau_M - \tau_m)^2 \quad (28)$$

This suggestion will embed the Pedersen sampler ($c = \infty$) and the Durham-Gallant sampler ($c = 0$) as special cases of

our sampler. It will also prevent the sampler from breaking down if c is badly misspecified.

Higher order bias terms can also be included leading to

$$\bar{\mathrm{P}} = \mathrm{P} \tag{29}$$

$$\bar{\mathrm{Q}} = \mathrm{Q} + \gamma_2(\tau_M - \tau_m)^2 + \gamma_4(\tau_M - \tau_m)^4 \tag{30}$$

where γ_2 and γ_4 are suitable matrices.

We refer to these modifications as 1st ((27) and (28)) and 2nd ((29) and (30)) order regularized covariances. The corresponding samplers will be referred to as 1st and 2nd order regularized Durham-Gallant samplers.

A convenient property with our derivation of the sampler is that these modifications can be used in (13) and (14) to compute the conditional mean and covariance after replacing Q with $\hat{\mathrm{Q}}$ or $\bar{\mathrm{Q}}$.

2.1 A heuristic algorithm for finding cc^T, γ_2 and γ_4

It was shown in Stramer and Yan (2007) that it is computationally expensive to reduce bias, and it is therefore likely that $\tau_m - \tau_{m-1}$ is chosen such that the variance dominates the squared bias but some bias remains. We therefore claim as a first approximation that the size of the squared bias is a fraction α of the size of the variance. A pragmatic approach that would ensure that the MSE is a positive definite matrix is to approximate the MSE by

$$\mathrm{MSE}(\mathbf{x}_{\tau_m}|\mathbf{x}_{\tau_{m-1}}) \approx (1+\alpha)\mathrm{P} = \mathrm{P} + cc^T_{Heur}(\tau_m - \tau_{m-1})^2, \tag{31}$$

where $\alpha \in [0.01,\ 1]$ have proved successful in our simulations. Large values of α would adjust for the fact that the bias often is non-linear, and may grow faster than a linear term for longer time horizons (the bias is approximately linear only for short time horizons). Solving equation (31) yields

$$cc^T_{Heur} = \alpha \frac{\mathrm{P}}{(\tau_m - \tau_{m-1})^2}. \tag{32}$$

Similar heuristics can be used to find reasonable values for γ_2 and γ_4 by choosing γ_2 as cc^T_{Heur} (using $\alpha = 0.01$) and γ_4 such that the first bias term dominates the second for a short time span $\tau_m - \tau_{m-1}$, i.e. solving $\gamma_4(\tau_m - \tau_{m-1})^4 = \varepsilon\gamma_2(\tau_m - \tau_{m-1})^2$ for some small constant ε. This ensures that the first and second order MSE and the covariance are similar for short time horizons.

The primary effects caused by this heuristic choice is that the Kalman gain, K_0, is decreased

$$\hat{K}_0 = \hat{\mathrm{P}}(\hat{\mathrm{P}} + \hat{\mathrm{Q}})^{-1} = \mathrm{P}(\mathrm{P} + \hat{\mathrm{Q}})^{-1} \tag{33}$$

$$\bar{K}_0 = \bar{\mathrm{P}}(\bar{\mathrm{P}} + \bar{\mathrm{Q}})^{-1} = \mathrm{P}(\mathrm{P} + \bar{\mathrm{Q}})^{-1} \tag{34}$$

It can be seen that both $\bar{K}_0 \leq K_0$ and $\hat{K}_0 \leq K_0$, meaning that less importance is given to the information provided by the forthcoming observation when $\tau_M - \tau_m$ is large compared to $\tau_m - \tau_{m-1}$. Remembering from (11) and (12) that the matrix Q can be written as $Q = \frac{\tau_M - \tau_m}{\tau_m - \tau_{m-1}} P$ makes the proof straightforward

$$\begin{aligned}\hat{K}_0 &= P(P + Q + cc^T_{Heur}(\tau_M - \tau_m)^2)^{-1} \\ &= P\Bigg(P + \frac{\tau_M - \tau_m}{\tau_m - \tau_{m-1}}P \\ &\quad + \alpha\frac{\mathrm{P}}{(\tau_m - \tau_{m-1})^2}(\tau_M - \tau_m)^2\Bigg)^{-1} \\ &= I\frac{\tau_m - \tau_{m-1}}{\tau_m - \tau_{m-1} + \tau_M - \tau_m + \alpha\frac{(\tau_M - \tau_m)^2}{\tau_m - \tau_{m-1}}}.\end{aligned} \tag{35}$$

Choosing α close to zero leads to a sampler that is similar to the Durham-Gallant sampler while a large α would result in a sampler similar to the Pedersen sampler as the Kalman gain would tend to zero, meaning that no information obtained from the forthcoming observation is used.

The regularized sampler shrinks the variance of the Bridge less than the Durham-Gallant sampler would shrink it, cf. (14) which often is a good thing, cf. Geweke (1989). Having too light tails on the importance density may lead to slower than $\sqrt{N}$ convergence, cf. Koopman et al. (2009) while having too heavy tails only leads to inefficiency.

2.1.1 *Explicit form of the sampler*

It is possible to compute the explicit expression for the proposed sampler under certain conditions. We present the results for the 1st order regularized Durham-Gallant sampler as the results for the 2nd order sampler regularized Durham-Gallant sampler are similar. Assuming that an Euler-Maruyama discretization is being used (as for the modified Brownian Bridge sampler, and that the heuristic choice of cc^T_{Heur} is used) leads to

$$\hat{\mathrm{P}} = g(\cdot)g(\cdot)^T(\tau_m - \tau_{m-1}) \tag{36}$$

$$\hat{\mathrm{Q}} = g(\cdot)g(\cdot)^T\left(\tau_M - \tau_m + \alpha\frac{(\tau_M - \tau_m)^2}{\tau_m - \tau_{m-1}}\right), \tag{37}$$

resulting in the following equation for the Kalman gain

$$\hat{K}_0 = I\frac{\tau_m - \tau_{m-1}}{\tau_M - \tau_{m-1} + \alpha\frac{(\tau_M - \tau_m)^2}{\tau_m - \tau_{m-1}}} \tag{38}$$

It is obvious that the Kalman gain is an invertible matrix for any value of α.

It follows from (14) that the conditional covariance of the sampler is given by

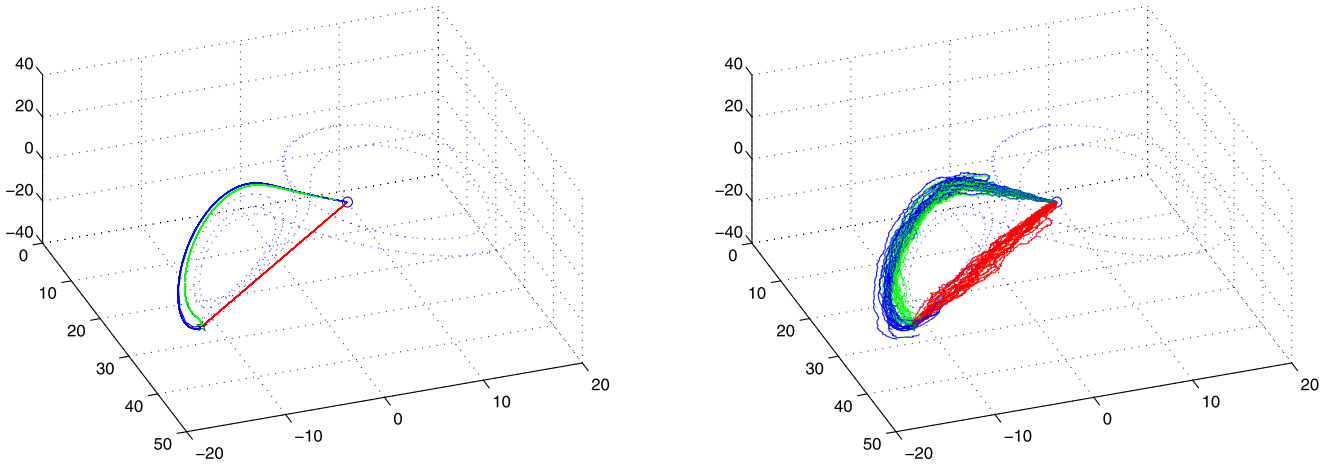

Fig. 2 (Color online) A deterministic trajectory in the Lorenz model (*blue solid line*) and the expected path generated from the Durham-Gallant sampler (*red dashed line*) and the proposed sampler using $\alpha = 0.1$ (*green dash-dotted line*). The initial point is displayed by the *circle* and the end point by a *pentagon* (*left*). Simulated paths from the Pedersen sampler (*blue lines*), the Durham-Gallant sampler (*red lines*) and the proposed sampler (*green lines*). The initial point is displayed by the *circle* and the end point by a *pentagon* (*right*)

$$\begin{aligned}\Sigma &= (I - \hat{K}_0)P = (I - \hat{K}_0K_0^{-1} + \hat{K}_0K_0^{-1} - \hat{K}_0K_0^{-1}K_0)P \\ &= (I - \hat{K}_0K_0^{-1}) \underbrace{P}_{\text{Pedersen}} + \hat{K}_0K_0^{-1} \underbrace{(I - K_0)P}_{\text{Durham-Gallant}}\end{aligned} \tag{39}$$

while the conditional mean is given by (13)

$$\begin{aligned}\mu &= \mathbf{a} + \hat{K}_0(\mathbf{x}_{\tau_M} - \mathbf{a} - \mathbf{b}) \\ &= (I - \hat{K}_0K_0^{-1})\mathbf{a} + \hat{K}_0K_0^{-1}\mathbf{a} + \hat{K}_0K_0^{-1}K_0(\mathbf{x}_{\tau_M} - \mathbf{a} - \mathbf{b}) \\ &= (I - \hat{K}_0K_0^{-1}) \underbrace{\mathbf{a}}_{\text{Pedersen}} + \hat{K}_0K_0^{-1} \underbrace{(\mathbf{a} + K_0(\mathbf{x}_{\tau_M} - \mathbf{a} - \mathbf{b}))}_{\text{Durham-Gallant}}\end{aligned} \tag{40}$$

It can now be seen that the regularized sampler is a combination of the Durham-Gallant sampler and the Pedersen sampler, much like the Fearnhead sampler.

The Pedersen term is dominating when $\tau_M - \tau_m \gg \tau_m - \tau_{m-1}$ and the Durham-Gallant term dominating when $\tau_M - \tau_m \approx \tau_m - \tau_{m-1}$. This can be seen from the ratio $\hat{K}_0K_0^{-1}$ that corresponds to the weight given to the Durham-Gallant sampler

$$\begin{aligned}\hat{K}_0K_0^{-1} &= I\frac{\tau_m - \tau_{m-1}}{\tau_M - \tau_{m-1} + \alpha\frac{(\tau_M - \tau_m)^2}{\tau_m - \tau_{m-1}}}\left(I\frac{\tau_m - \tau_{m-1}}{\tau_M - \tau_{m-1}}\right)^{-1} \\ &= I\frac{\tau_M - \tau_{m-1}}{\tau_M - \tau_{m-1} + \alpha\frac{(\tau_M - \tau_m)^2}{\tau_m - \tau_{m-1}}}\end{aligned} \tag{41}$$

which approaches the identity matrix as $\tau_m \to \tau_M$.

This is an ideal structure for the sampler as the initially dominating Pedersen sampler will generate trajectories into the desired region (and resulting in low variability of the importance sampling weights), while the Durham-Gallant that dominates the sampler at the later stages will force the trajectories to end close to $\mathbf{x}_{\tau_M}$. This is important as most trajectories will contribute to the expectation, cf. (3).

The bridge type properties of the sampler is also important from a theoretical perspective as the results in Stramer and Yan (2007) are only valid for bridge-type samplers. Samplers that are not bridge type samplers (such as the Pedersen sampler) are having even slower convergence rates.

Example 2 The 1st order proposed sampler is applied to the same Lorenz model as before, using $\alpha = 0.1$, cf. (31). The result is presented in Fig. 2(left). It can be seen that the expected path of the proposed sampler is similar to the actual path. Finally, we plot $N = 20$ different stochastic trajectories simulated using the Pedersen sampler, the Durham-Gallant sampler and the proposed 1st order sampler. The results are presented in Fig. 2(right).

It can be seen that the proposed sampler does a better job at approximating the paths of the Lorenz system than the standard Durham-Gallant sampler and that all the simulated trajectories are closer to the end point (this is crucial for the Monte Carlo approximation) than the trajectories drawn by the Pedersen sampler. This indicates that the sampler is robust against sparsely sampled data.

3 Simulations

We compare the Pedersen sampler, see Pedersen (1995b), the Durham-Gallant sampler, see Durham and Gallant (2002), the Fearnhead sampler, see Fearnhead (2008) and the two proposed samplers using the efficient sample size (ESS). The efficient sample size (ESS) provides an estimate

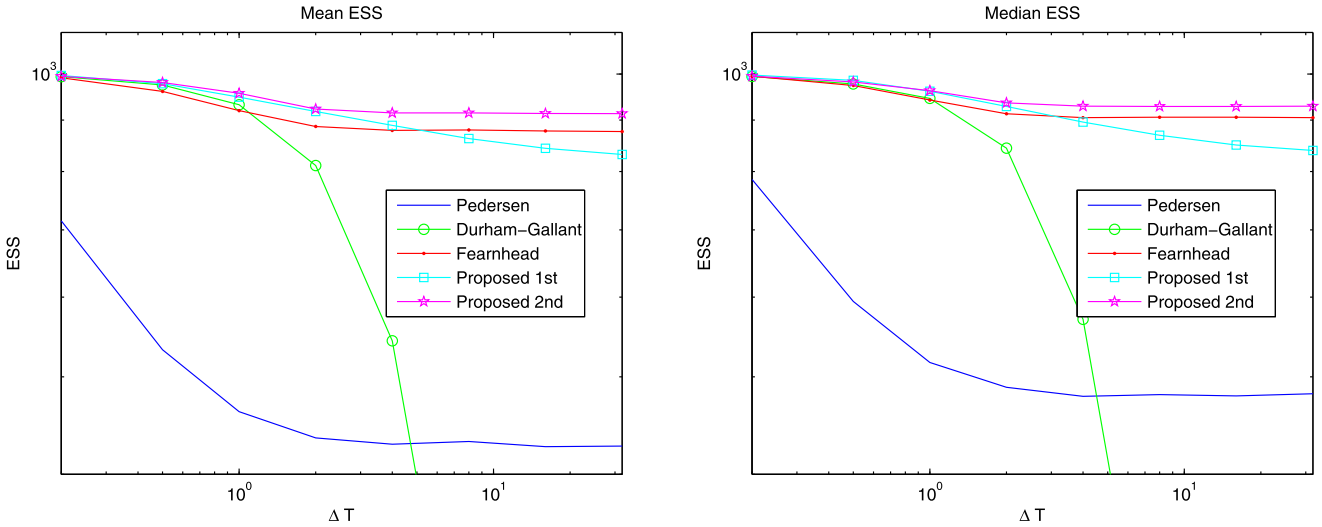

Fig. 3 Efficient sample size when using $K = 1\,000$ particles for the samplers applied to the CIR model

of the effective number of samples in a Monte Carlo estimate. A large ESS is therefore preferred to a lower ESS. The ESS is defined for a weighted Monte Carlo sample

$$\sum w_i \delta(x - x_i) \tag{42}$$

as

$$\mathrm{ESS} = \frac{(\sum w_i)^2}{\sum w_i^2}. \tag{43}$$

We compare three different models of increasing complexity.

3.1 Cox-Ingersoll-Ross model

The first model we use to evaluate the proposed samplers is the CIR model in Fearnhead (2008). The model is given by

$$\mathrm{d}X_t = (5 - X_t)\mathrm{d}t + \sqrt{X_t}\mathrm{d}W_t. \tag{44}$$

The Pedersen, Durham-Gallant, Fearnhead (using $\rho = 1$, $B = 1$ as in Fearnhead 2008), 1st order Regularized Durham-Gallant (using $\alpha = 0.1$) and 2nd order Regularized Durham-Gallant sampler (using $\alpha = 0.01, \varepsilon = 0.1$) were used to approximate the transition density. Data was generated by sampling a long time series ($T = 10\,000$) from the model, ensuring that the approximation of $p(x_{t_n}|x_{t_{n-1}})$ is tested for all relevant combinations of x_{t_n} and $x_{t_{n-1}}$, cf. Durham and Gallant (2002). The simulations used $\tau_m - \tau_{m-1} = 0.1$ while the sampling distance $t_n - t_{n-1}$ was varied from $t_n - t_{n-1} = 0.2$ to $t_n - t_{n-1} = 32$. All samplers used $K = 1\,000$ particles.

The results, reported as mean and median of pair wise efficient sample size for different sampling frequencies, are presented in Fig. 3. Both proposed samplers are performing similarly to the Fearnhead sampler and outperforming the Durham-Gallant sampler for sparsely sampled data, with the 2nd order proposed sampler is consistently performing at the top regardless of the sparsity of the data.

3.2 Double Well model

The double well model is a well known benchmark model that switches between two modes. The model is given by

$$\mathrm{d}X_t = 4X_t(1 - X_t^2)\mathrm{d}t + \mathrm{d}W_t. \tag{45}$$

Data was generated by sampling a time series ($T = 50$) from the model, using $\tau_m - \tau_{m-1} = 0.01$ while the sampling distance was varied from $t_n - t_{n-1} = 0.1$ to $t_n - t_{n-1} = 1$. The samplers use the same parameters as for the CIR model, and all samplers used $K = 1\,000$ particles. The results, reported as mean and median of pair wise efficient sample size for different sampling frequencies, are presented in Fig. 4. The difference is now pronounced, with the two proposed samplers being the most efficient. The Fearnhead sampler is less efficient and we suspect that this can be attributed to non-optimal values for ρ and B.

3.3 Lorenz model

The Lorenz model is more complex than the scalar models previously studied. This is attributed to the fact that the model is multivariate and that the deterministic version of the model is chaotic, and nearly periodic, cf. Fig. 1. Data was generated by sampling a time series ($T = 10$) from the model, using $\tau_m - \tau_{m-1} = 0.005$ to preserve numerical stability while the sampling distance was varied from

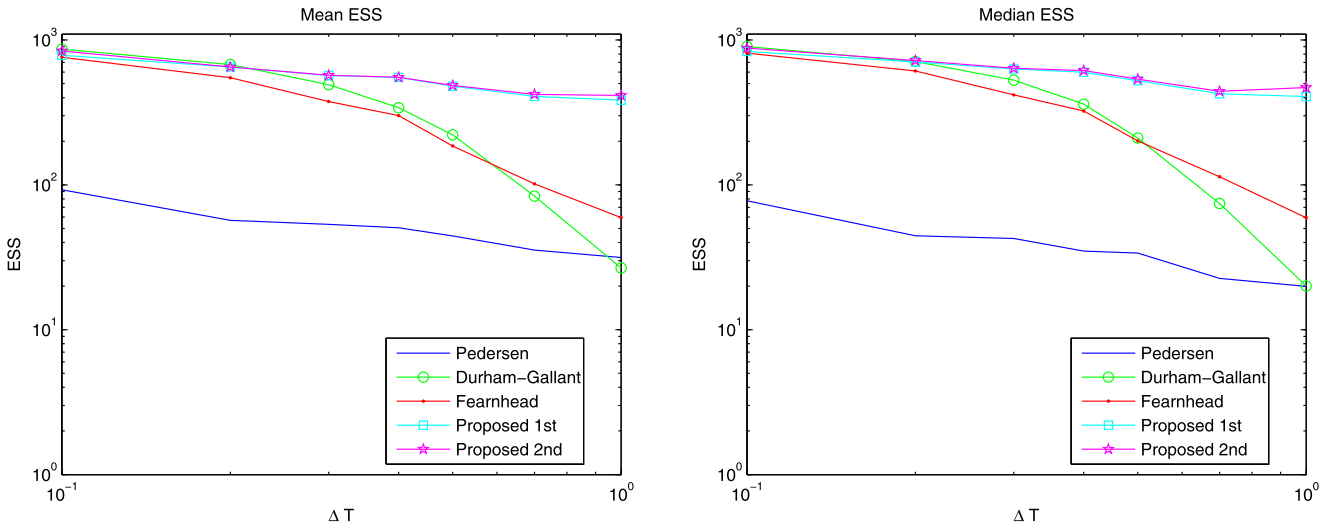

Fig. 4 Efficient sample size when using $K = 1\,000$ particles for the samplers applied to the double well model

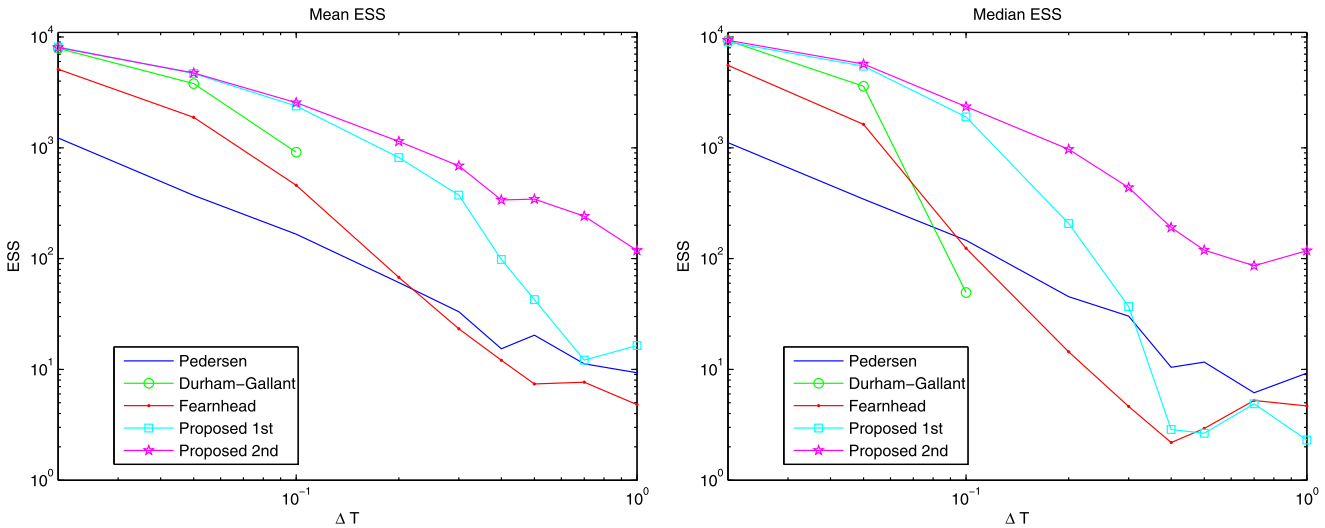

Fig. 5 Efficient sample size when using $K = 10\,000$ particles for the samplers applied to the stochastic Lorenz model

$t_n - t_{n-1} = 0.02$ to $t_n - t_{n-1} = 1$. The increased complexity was balanced by an increased number of particles, using $K = 10\,000$ particles.

The samplers used the same parameters as for the CIR and double well models expect for the Fearnhead model where we experimented with the ρ parameter. We found that increasing the ρ parameter to $\rho = 10$ gave the best results when data is sparsely sampled. It is non-trivial to find a proper value for ρ as the model is nearly periodic, as this would indicate that a small value for ρ would be appropriate. However, the local properties of the model, examined through local linearization, are very different for different parts of the state space. This indicates that a large value for ρ should be used, even if this goes beyond the theoretical justification for the sampler.

The results, reported as mean and median of pair wise efficient sample size for different sampling frequencies, are presented in Fig. 5. The Durham-Gallant sampler breaks down when $t_n - t_{n-1} > 0.2$, while the Fearnhead sampler is comparable to the Pedersen sampler for sparsely sampled data (using $\rho = 1$ resulted in a sampler similar to the Durham-Gallant sampler). However, both the proposed samplers do not degenerate, with the second order sampler being most efficient.

4 Summary

We have introduced a regularization term in the derivation of the Durham-Gallant sampler, leading to a sampler that is

a combination of the Durham-Gallant sampler and the Pedersen sampler. This makes the sampler more robust so that it can be used for generally sampled (densely or sparsely sampled), multivariate, discretely observed diffusion processes.

The samplers depends on few (one or two) parameters and we provide a heuristic methodology for choosing these, making the samplers easier to apply than the Fearnhead sampler. We have shown that the proposed samplers are almost as efficient as the Durham-Gallant sampler for frequently sampled data and far more efficient for sparsely sampled data. It is more efficient (on all time scales) than the Pedersen sampler for the problems we have examined and we expect this to hold for a very large class of problems due to the construction of the sampler.

References

Aït-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica **70**, 223–262 (2002)

Aït-Sahalia, Y.: Closed-form likelihood expansions for multivariate diffusions. Ann. Stat. **36**(2), 906–937 (2008)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. Econometrica **71**, 579–625 (2003)

Bengtsson, T., Snyder, C., Nychka, D.: Towards a nonlinear ensemble filter for high-dimensional systems. J. Geophys. Res. **108**(D24) (2003)

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

Beskos, A., Papaspiliopoulos, O., Roberts, G.O.: Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. Ann. Stat. **37**(1) (2009)

Durham, G.B., Gallant, A.R.: Numerical techniques for maximum likelihood estimation of continuous time processes. J. Bus. Econ. Stat. **20**(3), 297–316 (2002)

Elerian, O., Chib, S., Shephard, N.: Likelihood inference for discretely observed non-linear diffusions. Econometrica **69**, 959–993 (2001)

Eraker, B.: MCMC analysis of diffusion models with application to finance. J. Bus. Econ. Stat. **19**(2), 177–191 (2001)

Fearnhead, P.: Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Stat. Comput. **18**(2), 151–171 (2008)

Geweke, J.: Bayesian inference in econometric models using Monte Carlo integration. Econometrica **57**(6), 1317–1339 (1989)

Glasserman, P.: Monte Carlo Methods in Financial Engineering. Stochastic Modelling and Applied Probability, vol. 53. Springer, Berlin (2003)

Golightly, A., Wilkinson, D.J.: Bayesian sequential inference for nonlinear multivariate diffusions. Stat. Comput. **16**(4), 323–338 (2006)

Golightly, A., Wilkinson, D.J.: Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput. Stat. Data Anal. **52**(3), 1674–1693 (2008)

Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**, 327–343 (1993)

Hurn, A.S., Jeisman, J.I., Lindsay, K.A.: Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. J. Financ. Econom. **5**(3), 390 (2007)

Jonsdottir, H., Madsen, H., Palsson, O.P.: Parameter estimation in stochastic rainfall-runoff models. J. Hydrol. **326**(1–4), 379–393 (2006). ISSN 0022-1694

Koopman, S.J., Shephard, N., Creal, D.: Testing the assumptions behind importance sampling. J. Econom. **149**(1), 2–11 (2009)

Lindström, E.: Estimating parameters in diffusion processes using an approximate maximum likelihood approach. Ann. Oper. Res. **151**, 269–288 (2007). doi:10.1007/s10479-006-0126-4

Lindström, E.: Implications of Parameter Uncertainty on Option Prices. Advances in Decision Sciences, vol. 2010 (2010). ISSN 2090-3359

Lo, A.W.: Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econom. Theory **4**(2), 231–247 (1988)

Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**(2), 130–141 (1963)

Madsen, H., Holst, J.: Estimation of continuous-time models for the heat dynamics of a building. Energy Build. **22**(1), 67–79 (1995). ISSN 0378-7788

Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 5th edn. Springer, Berlin (2000)

Pedersen, A.R.: Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli **1**(3), 257–279 (1995a)

Pedersen, A.R.: A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Stat. **22**, 55–71 (1995b)

Philipsen, K.R., Christiansen, L.E., Hasman, H., Madsen, H.: Modelling conjugation with stochastic differential equations. J. Theor. Biol. **263**(1), 134–142 (2010). ISSN 0022-5193

Phillips, P.C.B., Yu, J.: A two-stage realized volatility approach to estimation of diffusion processes with discrete data. J. Econom. **150**(2), 139–150 (2009). ISSN 0304-4076

Stramer, O., Yan, J.: On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. J. Comput. Graph. Stat. **16**(3), 672–691 (2007)