Exact Bayesian inference for discretely observed diffusions

Inferenza Bayesiana esatta per processi di diffusione discretamente osservati

Isadora Antoniano-Villalobos and Stephen G. Walker

Abstract Inference for discretely observed real-valued diffusion processes is commonly based on approximations. The only methods available for exact inference profit from a retrospective rejection sampler for exact simulation of diffusion paths which provides Maximum Likelihood estimates of the model parameters. In the context of Bayesian analysis, however, the current approach resorts to MCMC estimation for which the complexity of the algorithm grows linearly with the size of the data set. We propose a reinterpretation of the exact simulation algorithm, in terms of a set of latent variables which transform the Bayesian parametric diffusion model into a Bayesian non or semi parametric model. We then propose an estimation method, involving trans-dimensional MCMC methods, which allows exact inference for a specific family of diffusions.

Abstract *L'inferenza per processi di diffusione, a valori reali discretamente osservati, é comunemente basata su criteri di approssimazione. Gli unici metodi disponibili per fare inferenza esatta sfruttano un algoritmo di accettazione/riuto retrospettivo al fine di generare uno stimatore di massima verosimiglianza basato sulla simulazione esatta della traiettoria. Nel contesto Bayesiano, comunque, il presente approccio richiama una stima MCMC in cui la complessitá cresce linearmente con la numerositá dei dati. La nostra proposta é la reinterpretazione dellalgoritmo di simulazione esatta in termini di variabili latenti che trasformano il modello di diffusione parametrico in un modello non o semi parametrico. Si propone, dunque, un metodo di stima che coinvolge metodi MCMC trans-dimensionali che permettono di fare inferenza esatta per una famiglia specica di diffusioni.*

Key words: Exact simulation, Trans-dimensional MCMC, Latent model

---

Isadora Antoniano-Villalobos
Universita Bocconi, Milano, Italia e-mail: isadora.antoniano@unibocconi.it

Stephen G. Walker
University of Texas at Austin, USA e-mail: s.g.walker@math.utexas.edu

1 Introduction

Diffusion processes have been widely studied in the context of probability theory and in many other areas, ranging from the natural sciences like biology or genetics to the realms of economics and finance. One of the main features of this family of stochastic processes, the continuity of their paths, makes them attractive models for several phenomena. Unfortunately, it also makes statistical inference a highly challenging task, due to the intractability of the transition densities involved in the likelihood function of a discretely observed path.

Consider a sample of n observations $y_{1:n} = (y_{t_1}, \ldots, y_{t_2})$, for fixed, known times $0 < t_1 < \ldots < t_n < \infty$, corresponding to a discretely observed path from a diffusion process $Y = \{Y_t : t \geq 0\}$, defined as the unique solution to an SDE,

$$\mathrm{d}Y_t = \alpha_\theta(Y_t)\mathrm{d}t + \mathrm{d}W_t; \quad Y_0 = y_0, \tag{1}$$

where $W = \{W_t : t \geq 0\}$ denotes a standard scalar Brownian Motion. For simplicity, we consider a unitary diffusion coefficient, but the results can be extended under adequate conditions. Assume that the parametric form of the drift coefficient α_θ is known and the process is well defined for every value of θ in some parameter space Θ. The likelihood for the sample, given by

$$f(y_{1:n}|\theta) = \prod_{i=1}^{n} f_{\Delta_i}(y_i|y_{i-1}, \theta), \quad \Delta_i = t_i - t_{i-1}, \tag{2}$$

depends on the transition densities for the process, f_t which, under some standard assumptions, can be proved to exist. In most cases, however, such transitions do not have an analytic form, thus resulting in an intractable likelihood.

Inference for the unknown parameter θ, both in the frequentist and Bayesian setups, is usually based on some form of approximation or interpolation technique, such as approximate simulation, analytic approximations of the transition density or the complete likelihood functions or direct approximation of the maximum likelihood estimator. For a review of these methods see Sorensen (2004).

An important breakthrough was brought about by the definition of an exact simulation algorithm for diffusion paths at arbitrary time points within a closed time interval $[0,t]$, with no approximation error (see Beskos and Roberts, 2005; Beskos et al., 2006). It is a retrospective rejection sampler which exploits the factorization of the diffusion path in terms of a finite set of points, known as the skeleton, connected by independent Brownian bridges. Beskos et al. (2009) propose a method for Monte Carlo Maximum Likelihood estimation based on this algorithm. In the context of Bayesian analysis, however, the application of such method requires independent exact simulation of the skeleton between each pair of consecutive data points, so that the number of simulation steps involved in each Markov Chain iteration grows linearly with n. We propose an alternative Markov chain method based on a reinterpretation of the representation involved in the exact simulation algorithm, in terms of a latent model, which allows for both simulation and Bayesian

inference. Since the model is the same and no approximation is used, apart from the usual Monte Carlo error, the results obtained in this manner are equivalent to those obtained via the original exact simulation method. Our algorithm, however, is not based on the simultaneous acceptance or rejection of complete sample paths and is therefore equally applicable, regardless of the length of the time interval $[0,t]$ under consideration.

2 Retrospective Rejection Sampler

Following Beskos et al. (2006), we assume that the drift coefficient of the SDE (1) is such that, for every $\theta \in \Theta$ we can write

$$l(\theta) \leq \inf_{u\in\mathbb{R}} \left\{ \left[\alpha_\theta^2(u) + \alpha_\theta'(u)\right]/2 \right\}; \quad r(\theta) \geq \sup_{u\in\mathbb{R}} \left\{ \left[\alpha_\theta^2(u) + \alpha_\theta'(u)\right]/2 - l(\theta) \right\},$$

for some $l : \Theta \to \mathbb{R}$ and $r : \Theta \to (0,\infty)$. It is to be noted that the results of Beskos et al. (2006) and Beskos et al. (2009) work under less restrictive assumptions, which cover most of the diffusion process commonly used for statistical modelling. In the present work, however, we focus on the simplest version of the algorithm as it suffices for illustrative purposes.

The exact simulation algorithm is based on the observation that, under the above conditions, the transition density of the diffusion (with respect to Lebesgue measure) can be written as

$$f_t(y_t|y_0,\theta) = N(y_t|y_0,t)\exp\{A_\theta(y_t) - A_\theta(y_0) - l(\theta)\}\mathbb{E}_{\mathbb{W}^{y_0}}\left[\exp\left\{-r(\theta)\int_0^t \varphi_\theta(y_s)\mathrm{d}s\right\}\Big|y_t\right],$$

where $N(\cdot|\mu,\sigma^2)$ denotes the normal density function,

$$A_\theta(u) = \int \alpha_\theta(u)\mathrm{d}u; \qquad \varphi_\theta(u) = \frac{1}{r(\theta)}\left(\frac{\alpha_\theta^2(u) + \alpha_\theta'(u)}{2} - l(\theta)\right),$$

and the expectation is taken with respect to the Weiner measure of a Brownian motion started ad y_0. Such expectation is intractable but, given y_t, it coincides with the expected value of an indicator variable which takes the value one when a realization of a homogeneous Poisson process on $[0,t]\times[0,1]$ with intensity $r(\theta)$ has no points below the graph $s \mapsto \varphi(\omega_s;\theta)$, where ω_s is a Brownian bridge path started at $\omega_0 = y_0$ and ended at $\omega_t = y_t$.

Therefore the exact simulation algorithm performs rejection sampling by means of an auxiliary marked Poisson process. There is no need to simulate the complete proposed path ω. All the information required to calculate the acceptance indicator is contained in a finite number k of points, given by the marked Poisson Process and necessary instances of the proposed Brownian Bridge can be obtained retrospectively given k. The set $S_k = \{(\tau_1, y_{t_1}), \ldots, (\tau_k, y_{t_k})\}$ of accepted points is called the Skeleton of the path and for any $s \in (0,t)$, y_s can be simulated exactly using Brownian Bridge interpolation between consecutive skeleton points.

3 MCMC exact simulation and Bayesian inference

The diffusion model, together with the Brownian bridge factorization of the observed path on the complete time interval $[0,t]$, for $t > t_n$ can be reinterpreted as a latent model in the following way.

The likelihood function for the complete sample can be expressed as,

$$f(y_{0:n}|\theta) = \exp\left\{A_\theta(y_t) - A_\theta(y_0) - t\left[l(\theta) + r(\theta)\right]\right\} \prod_{i=1}^{n} N(y_{t_i}|y_{t_{i-1}}, \Delta_i)\mathbb{E}_{\mathbb{W}^{y_{t_{i-1}}}}\left[\exp\left\{r(\theta)\int_{t_{i-1}}^{t_i}\left[1-\varphi_\theta(y_s)\right]\mathrm{d}s\right\}\Big|y_{t_i}\right].$$

We then observe that the retrospective rejection criterion described in the above section is equivalent to the introduction of an appropriate set of latent variables obtained through a series expansion of the exponential function inside the expectation (for details, see Antoniano-Villalobos, 2012). This results in the latent expression,

$$f(y_{1:n}, k_{1:n}, S_k|\theta) = g(y_{1:n},\theta)\prod_{i=1}^{n}\frac{\left[r(\theta)\right]^{k_i}}{k_i!}\prod_{l=1}^{k_i} N\left(x_{i,l}|x_{i,l-1}, \tau_{i,l}-\tau_{i,l-1}\right)\left[1-\varphi_\theta(x_{i,l})\right],$$

where S_k is the set of skeleton points for the complete path on $[0,t]$, that is $(x_{i,l}, \tau_{i,l})$ for $i = 1,\ldots,n$ and $l = 1,\ldots,k_i$; and we denote $x_{i,0} = y_i$. It can be checked that, by integrating over the latent variables $k_{1:n}, S_k$, the original likelihood is recovered.

Notice that the latent likelihood can be interpreted in terms of a Bayesian nonparametric (or semi-parametric) model, that is, the joint distribution of the observations and an infinite-dimensional parameter (S,k). The likelihood for such model is the product of the densities of independent brownian bridges between k consecutive skeleton points. From the Bayesian perspective, the joint prior over (k,S) is induced by the choice of the functional form of the diffusion coefficient and should be complemented with a prior $\pi(\theta)$ over the parameter space Θ.

An MCMC algorithm may be used to simulate both from the posterior distributions of (k,S) and θ, enabling Bayesian inference. Rather than accepting or rejecting a complete skeleton over the whole interval $[0,t]$, or splitting the interval between observation times, our algorithm treats each skeleton point over the complete time interval, individually, leaving the decision of the adequate k to the MCMC scheme. The implementation of the full Gibbs Sampler applies a Metropolis-Hastings version of a reversible jump step, proposed by Godsill (2001), in order to simulate from the full conditional distribution for k. The choice of $t > t_n$ makes prediction possible, since any unobserved point can be sampled from fully defined Brownian bridges, given the posterior sample of skeleton points.

4 Illustration

As an illustration, we consider the diffusion process defined by the SDE

$$dY_t = \sin(Y_t - \theta)dt + dW_t; \qquad \theta \in \Theta = [0, 2\pi) \tag{3}$$

Using the retrospective rejection sampler of (EA1 in Beskos et al., 2006), we generate a single skeleton for the sine diffusion in the time interval $[0, 100]$ and use Brownian bridge interpolation to simulate $10,000$ equally spaced data points, i.e. 100 observations per time unit (see Figure 1).

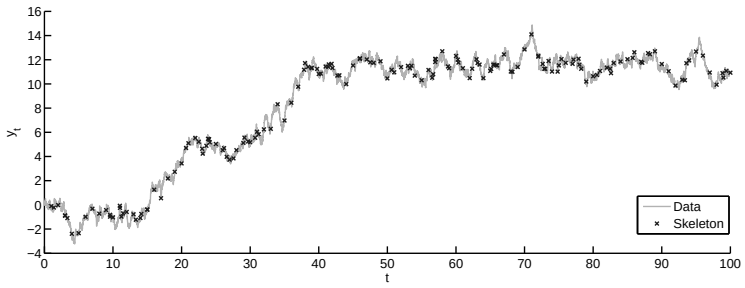
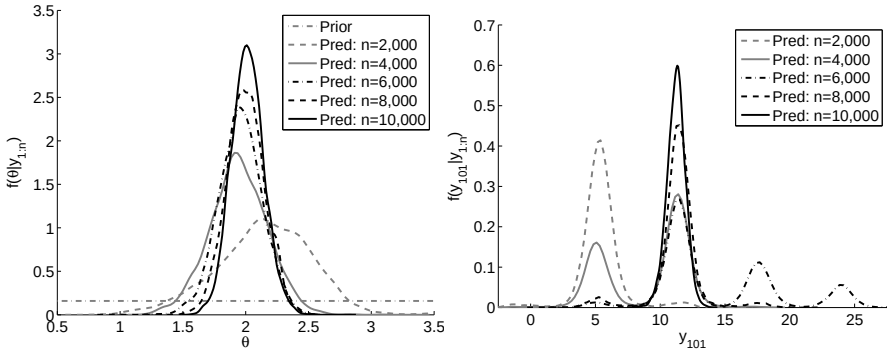


Fig. 1: $10,000$ data points from the sine diffusion in the time interval $[0, 100]$, with parameter $\theta = 2$ and initial point $y_0 = 0$.

We define a uniform prior on Θ and use the MCMC algorithm to produce a sample from the posterior distribution $\pi^n(\theta)$, for increasing sample sizes of high frequency data. Specifically, we consider the data set consisting of the first $n = 2$ thousand data points, in the time interval $[0, 20]$ and produce a posterior sample of size $N = 10,000$ from the MCMC algorithm, with a burning period of $10,000$ iterations and a thinning of 1 every 10 iterations for the sample. We repeat the analysis for $n = 4, 6, 8, 10$ thousand. The estimated posterior densities for the parameter are shown on the left hand side of Figure 2. We can see that the posterior mass seems to accumulate around the true value $\theta_0 = 2$ as the sample size increases.

The right panel of Figure 2 shows the estimated predictive densities for the process Y_t at time $t = 101$, for each of the sample sizes. The sine diffusion does not have a stationary density, therefore we don't expect to recover a fixed marginal behaviour. However, as the interval of observations approaches the time of prediction, we can observe the evolution of the predictive distribution. As expected from a regular diffusion process, the variance decreases towards the end, as the point y_{101} is highly correlated to y_{100}, the last data point.



(a) Posterior density (b) Predictive density

Fig. 2: Estimated posterior density for the parameter of the sine diffusion (left) and predictive density for the observation at time $t = 101$ (right).

References

Isadora Antoniano-Villalobos. *Bayesian Inference for Models with Infinite-dimensionally Generated Intractable Components*. PhD thesis, SMSAS, University of Kent, Canterbury, UK, 2012.

A. Beskos and G. O. Roberts. Exact Simulation of Diffusions. *Annals of Applied Probability*, 15(4):2422–2444, 2005.

A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead. Exact and Computationally Efficient Likelihood-Based Estimation for Discretely Observed Diffusion Processes (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 68(3):333–382, 2006.

A. Beskos, O. Papaspiliopoulos, and G. Roberts. Monte Carlo Maximum Likelihood Estimation for Discretely Observed Diffusion Processes. *The Annals of Statistics*, 37(1):223–245, 2009.

S. J. Godsill. On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty. *Journal of Computational and Graphical Statistics*, 10 (2):230–248, 2001.

H. Sorensen. Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey. *International Statistical Review*, 72(3):337–354, 2004.