Part IV:MC-based likelihood inference for discretely-observed diffusions with known constant diffusivity

- Missing data problems and formal data augmentation (DA)
- DA for discretely-observed diffusions with known constant diffusion coefficient
- Conditional distribution of the missing data: diffusion bridges
- Likelihood ratios for diffusion bridges, transition density identities, connections to literature
- An MCMC scheme for parameter estimation

To avoid excessive notation we focus on time-homogeneous diffusions, although this is only for convenience

We have seen that pseudo-likelihood approaches are typically inconsistent considering outfill asymptotics. We need consistent (and hopefully efficient) statistical procedures

We will use the tools we have developed (discretizations, likelihood ratios on the path space, exact simulation etc) to construct MC methods for this end

We start by considering a simpler setting: target process has known constant diffusion coefficient. Without loss of generality, we take it to be 1. Therefore, we assume observed data $\mathbf{x} = (x_0, x_1, \ldots, x_n)$ from (34) at times $t_0 = 0 < t_1 < \cdots < t_n = T$, where the **non-linear drift** $\alpha(x; \theta)$ depends on unknown parameters θ, which we wish to estimate

A lot of the results given in this unit remain true when the diffusion coefficient is any time function, even in non-elliptic cases

Why known diffusivity?

Our first approach (as it happened in the literature) will be to try an **off-the-shelf** computational tool to solve the problem for diffusions; precisely the DA we have already discussed. In the case of known diffusivity it turns out that this is possible.

There is still a major challenge in this context, which has to do with diffusion bridges. We will study this topic in this thematic unit

Nevertheless, applying the standard DA when diffusion coefficient is unknown leads to a **disaster**, and we need to think more carefully (both for diffusions, and for general DA)

Discretely-observed diffusions as missing data problem

Our problem has a structure very common in many statistical applications (e.g random effect models, analysis of surveys, inverse problems)

Inference for the complete dataset which includes both the observed data $\mathbf{x}$, and the paths in-between observations is rather straighforward using the continuous-time likelihood approach we have already seen. Let $X^c = (X_s, s \in [0, T])$ be the complete data. Then, we have the **complete log-likelihood**:

$$\log L^c(\theta \mid X^c) = \int_0^T \alpha^*(X_s, \theta)\mathrm{d}X_s - \frac{1}{2}\int_0^T [\alpha^*\alpha](X_s, \theta)\mathrm{d}s \quad (40)$$

A bit more formal

Complete data $X^c = (\mathbf{x}, X^m) = (X_s, s \in [0, T])$ according to probability measure $\mathbb{P}^{(x_0, T)}$ which has density w.r.t to a parameter-independent dominating measure $\mathbb{W}^{(x_0, T)}$, whose logarithm is given in (40)

Nevertheless, X^c is not available to us, but only $\mathbf{x}$

(Formal) Bayesian DA for missing data problems

General setup: observed data $\mathbf{x}$, unknown parameters θ

missing data X^m, complete data X^c, probability model for complete data $\mathbb{P}$, with tractable density w.r.t θ-independent dominating measure $\mathbb{W}$, $L^c(\theta \mid X^c)$, which we call complete likelihood. Without loss of generality we take $\mathbb{W}$ to be a probability measure

Observed likelihood is obtained by integrating out the missing data, formally

$$L(\theta \mid \mathbf{x}) = \mathbb{E}_{\mathbb{W}}[L^c(\theta \mid X^c) \mid \mathbf{x}] \quad (41)$$

i.e the expectation is w.r.t to the **conditional distribution** of X^m given $\mathbf{x}$ under $\mathbb{W}$, and (41) is density w.r.t to the marginal measure for $\mathbf{x}$ under $\mathbb{W}$.

DA is not usually presented in so general terms, see however [Dembo and Zeitouni, 1986] for similar approach

To see why (41) holds, note that

$$\begin{aligned} \mathbb{P}[x \in A] &= \mathbb{E}_{\mathbb{P}}[1[X^o \in A]] = \mathbb{E}_{\mathbb{W}}[1[X^o \in A] L^c(X^o, X^m)] \\ &= \mathbb{E}_{\mathbb{W}}[1[X^o \in A] \mathbb{E}_W[L^c(X^o, X^m) \mid X^o]] \end{aligned}$$

We elicit a prior distribution for θ, trying to incorporate reasonable prior knowledge. Thus, θ is also treated as random variable in a **Bayesian approach**. If the priors are relatively weak in the area where data is most informative, then the analysis would be similar to a likelihood analysis.

Then, the DA treats (θ, X^m) as two-component random vector, with joint density (w.r.t to a suitable **product** - under this construction -**measure**)

$$\pi(\theta, X^m \mid \mathbf{x}) \propto \pi(\theta) L^c(\theta \mid \mathbf{x}, X^m) \tag{42}$$

Simulate from this distribution; marginal θ draws are from the observed-data posterior, marginal for X^m allows for full probabilistic inference for the unobserved data given observations, integrating out parameter uncertainty

DA and the Gibbs sampler

Construct a Markov chain with $\pi(\theta, Y_{mis}|Y)$ as its stationary distribution by iterating

1. simulated from $\pi(\theta|X^c)$;
2. simulated from $\pi(X^m|\mathbf{x}, \theta)$.

The marginal invariant distribution of θ for this algorithm is obviously $\pi(\theta|\mathbf{x})$

However this algorithm requires two simulation steps which are often impossible in interesting examples.

More typically have to do the more general:

1. Simulate one step from a Markov chain invariant w.r.t $\pi(\theta|X^c)$ (e.g. by a standard MCMC procedure);
2. simulate one step from a Markov chain on the space of possible missing value sets, with invariant distribution $\pi(X^m|\mathbf{x}, \theta)$.

Close links to EM algorithm, Monte Carlo EM algorithm, Monte Carlo Maximum Likelihood.

DA for diffusions

Following the above paradigm we readily have $\pi(\theta, X^m \mid \mathbf{x})$ as the product of the prior $\pi(\theta)$ and the complete likelihood in (40).

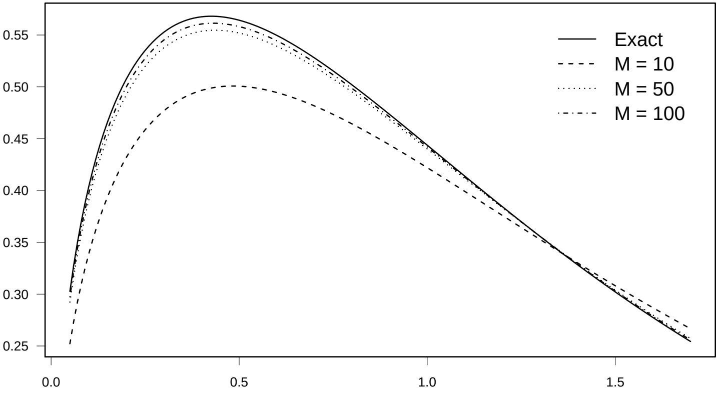
In this context, X^m is the collection of **interpolating paths** in-between observed data, $\mathbb{P}$ the law of the whole path and $\mathbb{W}$ the **Wiener measure**

Therefore, we can appeal to the generic algorithm in 125 or its more usual implementation in 126.

For computer implementation, and again without thinking too much, we would impute a fine discretization on the missing paths, say based on M intermediate points. Therefore, an approximating algorithm works with $X^{m,M}$ with bias which disappears for increasing M. $M = 0$ corresponds to the pseudo-likelihood approaches we have seen before.

Step 1 of the DA in 126 is rather straightforward

Approximation bias



Conditional distribution of missing data

Considerable sophistication underlies the 2nd step in the algorithm, when applied to diffusions. It is most clear to understand when $M = \infty$, and treat finite M as a finite-dimensional approximation of the distribution on paths

In this context, X^m consists of interpolating paths. By the Markov property, these are **conditionally independent**. Therefore, it suffices to study the path between x_i and x_{i+1} separately. Without loss of generality we consider the path between x and y for times 0 and T. The path corresponds to a diffusion conditioned on its end-points, this is what is called a diffusion bridge

Conditioned diffusions: h-transform

We have a diffusion

$$dX_t = \alpha(X_t)dt + dB_t$$

with (say) $X_0 = x$, and wish to consider its dynamics conditioned on the event that $X_T = y$.

It still Markov, and the theory of *h-transforms*, see for example Chapter IV.39 [Rogers and Williams, 2000], allows us to derive its SDE. Let $p_t(u, v)$ be the transition density, and recall that for v fixed as a function of u it satisfies the KBE (26). Let $\dot{p}$ and p' denote respectively derivatives w.r.t t and u.

The conditioned diffusion X (use the same letter for economy) satisfies an SDE

$$\mathrm{d}X_t = (\alpha(X_t) + \nabla \log p_{T-t}(X_t, y))\,\mathrm{d}t + \mathrm{d}B_t, \quad 0 \leq t \leq T\,, \quad (43)$$

where B is not the same BM as the one driving the unconditioned process. Note the behaviour of the drift near T (intuition)

h-transform proof (1-d for simplicity of notation)

p and in particular p' are badly behaved for t near to T. However it is enough to consider the dynamics of X_t up to some time $T - \epsilon$ for sufficiently small ϵ. The argument is based on a decomposition of the law of the conditioned path.

Let $\mathbb{P}_{T-\epsilon}^{(T,x,y)}$ denote the conditioned probability measure up to time $T - \epsilon$. Then note that:

$$
\begin{aligned}
\mathbb{P}_{T-\epsilon}^{(T,x,y)} &= \mathbb{P}_{T-\epsilon}^{(T-\epsilon,x,u)} \otimes p(u \mid x, y)\mathrm{d}u \\
&= \frac{p_{T-\epsilon}(x,u)p_\epsilon(u,y)}{p_T(x,y)}\,\mathrm{d}u \otimes \mathbb{P}_{T-\epsilon}^{(T-\epsilon,x,u)} \\
&= \frac{p_\epsilon(X_{T-\epsilon},y)}{p_T(x,y)}\,\mathbb{P}_{T-\epsilon}^{(T-\epsilon,x)}
\end{aligned}
\tag{44}
$$

$$\frac{\mathrm{d}\mathbb{P}_{T-\epsilon}^{(T,x,y)}}{\mathrm{d}\mathbb{W}_{T-\epsilon}^{(T-\epsilon,x)}}(X_{[0,T-\epsilon]}) = G(X_{[0,T-\epsilon]}) \times \frac{p_\epsilon(X_{T-\epsilon}, y)}{p_T(x,y)} \tag{45}$$

where G is the standard Girsanov between the unconditioned measures.

We aim to prove that the diffusion bridge has this Radon-Nikodym derivative with respect to Wiener measure.

h-transform proof

We can define a Girsanov formula for X in the usual way. In the sequel, the arguments of p and p' is throughout $(T-s, X_s, y)$

$$\log G(X) =$$

$$\int_0^{T-\epsilon} \left(\alpha(X_s) + \frac{p'}{p}\right) \mathrm{d}X_s - \frac{1}{2}\int_0^{T-\epsilon} \left(\alpha(X_s) + \frac{p'}{p}\right)^2 \mathrm{d}s$$

$$= \log G(X) + \int_0^{T-\epsilon} \frac{p'}{p}\mathrm{d}X_s - \int_0^{T-\epsilon} \left(\alpha(X_s)\frac{p'}{p} + \frac{1}{2}\left(\frac{p'}{p}\right)^2\right) \mathrm{d}s$$

To complete the proof we need to show that $\log \frac{p(\epsilon, X_{T-\epsilon}, y)}{p(T, x, y)}$ equals the last 3 terms in the expression.

We will use Itô in conjunction with KBE (26)

In the sequel, the arguments of p and p' is throughout $(T-s, X_s, y)$. $Y_t = \log p(T-t, X_t, y)$ then

$$\mathrm{d}Y_t = -\frac{\dot{p}}{p}\mathrm{d}t + \frac{p'}{p}\mathrm{d}X_t + \frac{(p''p - p'^2)}{2p^2}\mathrm{d}t$$

so that

$$\begin{aligned} Y_{T-\epsilon} - Y_0 &= \log p_\epsilon(X_{T-\epsilon}, y) - \log p_T(x, y) \\ &= \int_0^{T-\epsilon} \frac{p'}{p}\mathrm{d}X_s - \int_0^{T-\epsilon} \left(\frac{\dot{p}}{p} - \frac{(p''p - p'^2)}{2p^2}\right)\mathrm{d}s \\ &= \int_0^{T-\epsilon} \frac{p'}{p}\mathrm{d}X_s - \int_0^{T-\epsilon} \left(\alpha(X_s)\frac{p'}{p} + \frac{1}{2}\left(\frac{p'}{p}\right)^2\right)\mathrm{d}s \end{aligned}$$

Remark and warning

The argument based on (45) **does not** imply that RHS is the density of $\mathbb{P}_T^{(T,x,y)}$ w.r.t $\mathbb{W}_T^{(T-\epsilon,x)}$. It does, of course, for $\mathbb{P}_{T-\epsilon}^{(T,x,y)}$ and $\mathbb{W}_{T-\epsilon}^{(T-\epsilon,x)}$.

In the limit $\epsilon \to 0$, $\mathbb{P}_T^{(T,x,y)}$ concentrates all its mass on a path space which has 0 mass under $\mathbb{W}_T^{(T-\epsilon,x)}$. In fact, the RHS converes $\mathbb{W}_T^{(T-\epsilon,x)}$-a.s to 0.

We will revisit this point when trying to identify measures that **are** a.c w.r.t $\mathbb{P}_T^{(T,x,y)}$ in that limit.

Conditioned diffusion dynamics

Therefore, we have proved that the diffusion process conditioned on its end-points $X_0 = x, X_T = y$ follows a **time in-homogeneous** diffusion with SDE

$$\begin{aligned} \mathrm{d}X_s &= \tilde{b}(s, X_s)\,\mathrm{d}s + \mathrm{d}B_s, \quad s \in [0, T]\,, X_0 = x; \\ \tilde{b}(s, u) &= b(s, u) + \nabla_u \log p_{s,T}(u, y) \end{aligned} \tag{46}$$

Therefore we have a neat probabilistic description of the distribution of the missing data in the DA for diffusions, i.e we have characterised the distribution from which we should sample in Step 2 of 126. Is it helpful though?

The SDE of the conditioned process is intractable due to the presence of the transition density. Note that for linear SDEs with additive noise, the transition density is Gaussian, its logarithm quadratic, thus (46) a tractable linear SDE.

It is worth verifying that the Brownian bridge SDE (30) is the special case of (46) when $b = 0$.

Note that the intractability of (46) is even more fundamental from the intractability of the discrete-time dynamics of typical SDEs: here we do not even know the coefficients.

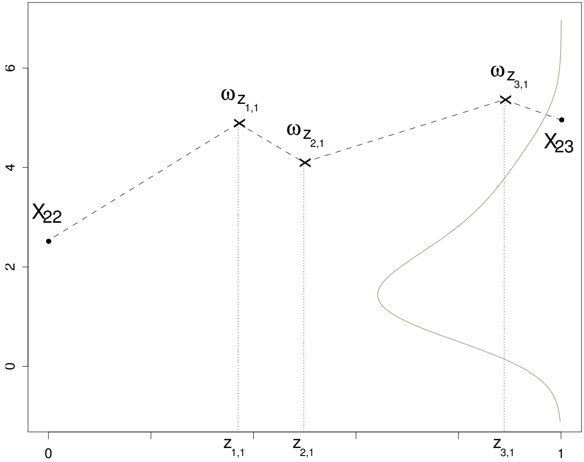
So, what can we do for carrying out the imputation step in DA for diffusions? This also has generated active and interesting research

First solution: Exact simulation of diffusion bridges

This is in fact a perfect solution, only compromised by the fact that the EA we presented earlier is not applicable for every diffusion, but requires certain assumptions on the coefficients.

EA for diffusion bridges is in fact even easier than for the unconditioned process

Trivial for the EA: instead of simulating last point, take $\omega_t = y$ and proceed as in the unconditional case in the algorithm in 93



It turns out that if EA can be applied, much more intelligent DA schemes can be devised, which go in different direction.

But at least, in this context it is easy to obtain a fine skeleton

What about SDEs for which EA cannot be applied (this is especially for multi-d case)? Ideas?

Importance sampling for conditioned diffusions

The idea is to propose from a **tractable diffusion measure** which is absolutely continuous w.r.t $\mathbb{P}^{(T,x,y)}$, with tractable likelihood ratio. In the context of MC integration, the samples should be then weighted by the LR for computing expectations.

In the context of DA, such samples can be used within an independence MH algorithm, whereby we propose candidates from this measure, and accept them with the appropriate probability

Lessons learnt from representation (46)

- the **local characteristics** of the unconditioned and conditioned processes are the same
- the drift of the conditioned process includes an extra term which forces the process to hit y at time T
- (46) is typically intractable since the drift is expressed in terms of the transition density

The main argument

Let, as before, $\mathbb{P}^{(T,x)}$ be the law of the diffusion X on $[0, T]$ with $X_0 = u$, $\mathbb{W}^{(T,x)}$ denote the Wiener law, $\mathbb{P}^{(T,x,y)}$ and $\mathbb{W}^{(T,x,y)}$ denote the laws of the corresponding diffusion bridges conditioned on $X_T = y$. Let $\mathcal{G}_{T,x}(y,)$ be the Wiener transition density

Crucially, note that the conditioned driftless process is also a linear SDE. In this setting it is just the Brownian bridge. BB is easy to simulate, thus an intresting candidate process. We show that is a valid one, i.e we obtain the likelihood ratio

Consider the following heuristic argument for deriving $\mathrm{d}\mathbb{P}^{(T,x,y)}/\mathrm{d}\mathbb{W}^{(T,x,y)}$. Consider the decomposition of the laws $\mathbb{P}^{(T,x,y)}, \mathbb{W}^{(T,x,y)}$ into the marginal distributions at time T and the diffusion bridge laws conditioned on X_T (ala the way we worked with the h-transform)

Then by a marginal-conditional decomposition we have that for a path X with $X_0 = x$,

$$\frac{\mathrm{d}\mathbb{P}^{(T,x)}}{\mathrm{d}\mathbb{W}^{(T,x)}}(X)\, 1[X_T = y] = \frac{p_{0,T}(x,y)}{\mathcal{G}_{0,T}(x,y)} \frac{\mathrm{d}\mathbb{P}^{(T,x,y)}}{\mathrm{d}\mathbb{W}^{(T,x,y)}}(X)\,. \tag{47}$$

The term on the left-hand side is given by the Cameron-Martin-Girsanov theorem. By re-arrangement:

$$\frac{\mathrm{d}\mathbb{P}^{(T,x,y)}}{\mathrm{d}\mathbb{W}^{(T,x,y)}}(X) = \frac{\mathcal{G}_{0,T}(x,y)}{p_{0,T}(x,y)} G(X)$$
$$\frac{\mathcal{G}_{0,T}(x,y)}{p_{0,T}(x,y)} \exp\left\{\int_0^T \alpha(s,X_s)^* \mathrm{d}B_s - \frac{1}{2}\int_0^T [\alpha^*\alpha](s,V_s)\mathrm{d}s\right\}\,, \tag{48}$$

Note, however, that we had effectively already obtained this result in (45). Following the same argument as in (44) (nd again for simplicity taking $d = 1$ to maintain correspondence with previous argument), we have the corresponding decomposition of Wiener measure

$$\mathbb{W}_{T-\epsilon}^{(T,x,y)} = \frac{\mathcal{G}_\epsilon(X_{T-\epsilon}, y)}{\mathcal{G}_T(x,y)} \; \mathbb{W}_{T-\epsilon}^{(T-\epsilon,x)}$$

and sustituting this into (45) we have

$$\frac{\mathrm{d}\mathbb{P}_{T-\epsilon}^{(T,x,y)}}{\mathrm{d}\mathbb{W}_{T-\epsilon}^{(T,x,y)}}(X_{[0,T-\epsilon]}) = G(X_{[0,T-\epsilon]}) \times \frac{\mathcal{G}_T(x,y)}{p_T(x,y)} \frac{p_\epsilon(X_{T-\epsilon},y)}{\mathcal{G}_\epsilon(X_{T-\epsilon},y)}$$

Note that when $\epsilon \approx 0$, by the Euler approximation

$$p_\epsilon(X_{T-\epsilon}, y) \approx \frac{1}{\sqrt{2\pi\epsilon}} \exp\left\{-\frac{1}{2\epsilon}(y - X_{T-\epsilon} - \alpha(X_{T-\epsilon})\epsilon)^2\right\}$$

therefore

$$\frac{p_\epsilon(X_{T-\epsilon}, y)}{\mathcal{G}_\epsilon(X_{T-\epsilon}, y)} \to 1 \quad \mathbb{W}_{T-\epsilon}^{(T-\epsilon,x,y)} - a.s.$$

Additional structure on α can lead to further simplifications of (48). For example, when the diffusion is time-homogenous and of gradient-type, i.e. there exists a field A such that $\alpha(v) = \nabla_v A(v)$ then using integration by parts in the exponent of (48) to eliminate the stochastic integral, we obtain

$$\frac{\mathrm{d}\mathbb{P}^{(T,x,y)}}{\mathrm{d}\mathbb{W}^{(T,x,y)}}(X) = \frac{\mathcal{G}_{0,T}(x,y)}{p_{0,T}(x,y)} \times \\ \exp\left\{A(y) - A(x) - \frac{1}{2}\int_0^T \left(||\alpha(X_s)||^2 + \nabla^2 A(X_s)\right) \mathrm{d}s\right\} . \tag{49}$$

(48) forms the basis for importance a particle approximation of the law of $\mathbb{P}^{(T,x,y)}$ using proposals from $\mathbb{W}^{(T,x,y)}$.

However, the weights are known only up to a normalizing constant due to the presence of $p_{0,T}(x, y)$.

This poses no serious complication in the application of IS (including RS), or independence MH. Note that $\mathcal{G}_{0,T}(x, y)$ is a Gaussian density which can be computed and be included explicitly in the weights, although this is not necessary for the IS.

Therefore, we can cary out the imputation step using an independence MH algorithm

Practically, we will have to simulate the proposed bridge at a finite collection of M times in $[0, T]$ and approximate the integrals in the weights by sums. This is an instance of the simulation-projection strategy.

It introduces a bias which is eliminated as $M \to \infty$. It is a subtle and largely unresolved issue how to distribute a fixed computational effort between M and N, the amount of MC replications, in order to minimize the MC variance of estimates of expectations of a class of test functions. However, a qualitative and asymptotic result is given in [Stramer and Yan, 2007] according to which one should choose $N = \mathcal{O}(M^2)$.

(A heuristic: the statistical error is $\mathcal{O}(1/\sqrt{N})$ and bias is $\mathcal{O}(1/M)$, thus to match them $M = \sqrt{N}$.)

A MC transition density identity

Re-arranging once more (48) and taking expectations on both sides we get:

$$p_{0,T}(u,v) = \mathcal{G}_{0,T}(u,v)\times$$
$$\mathbb{E}_{\mathbb{W}^{(T,x,y)}}\left[\exp\left\{\int_0^T h(s,V_s)^* \mathrm{d}B_s - \frac{1}{2}\int_0^T [h^* h](s,V_s)\mathrm{d}s\right\}\right] \tag{50}$$

RHS might be simplified further given appropriate structure (e.g σ invertible, α of gradient form, etc)
It is at this stage where the explicit computation of the Gaussian density becomes indispensable: if it were unknown we could only estimate the ratio of the two transition densities, but not $p_{0,T}(u,v)$.

A note on derivation of bridge density

We derived (50) from (48). This result can be also anticipated from the general IS framework (36), since the transition density is the probability of the conditioning event (evidence, partition function, marginal likelihood)

But it also follows from the basic principles of conditional expectation. In particular let $(\Omega, \mathcal{F})$ be a measurable space, $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on the space with Radon-Nikodym derivative $\xi = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Then, the derivative $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ restricted to $\mathcal{G}$ is $\mathsf{E}[\xi \mid \mathcal{G}]$, which follows from the definition of conditional expectation and the tower property. To obtain (47) we specify $\mathcal{G}$ as the σ-algebra generated by X_t.

Actually, note that it also follows directly from our generic missing data framework (41)

Historical Development

The expressions (48) and (50) have been derived several times in the literature with different motives. Remarkably, there is almost no cross-referencing. To our best knowledge, the expressions appear for the first time for scalar diffusions in the proof of Theorem 1 of [Rogers, 1985]. The context of the Theorem is to establish smoothness of the transition density. Again for scalar diffusions the expressions appear in the proofs of Lemma 1 of [Dacunha-Castelle and Florens-Zmirou, 1986]. The context of that paper is a quantification of the error in parameter estimates obtained using approximations of the transition density. Since both papers deal with scalar diffusions, they apply the integration by parts to get the simplified expression (49).

The [Durham and Gallant, 2002] IS estimator (see later) derived from different arguments, in the case of constant diffusion coefficient is a discretizations of (48) and (50). The context here is MC estimation of diffusion models. Since the authors work in a time-discretized framework from the beginning, the possibility to perform integration by parts when possible, is not at all considered.
[Nicolau, 2002] uses the
[Dacunha-Castelle and Florens-Zmirou, 1986] expression for the transition density as a basis for MC estimation using approximation of the weights based on M intermediate points.
[Beskos et al., 2006b] used (49) as a starting point for the exact simulation of diffusions and (50) as a basis for unbiased estimation of the transition density. Finally, [Delyon and Hu, 2006] state (48) as Theorem 2 and prove it for mutivariate processes.

Intringuingly, [Aït-Sahalia, 2002] in his analytic approximations for 1-d time homogeneous diffusions with unit diffusion coefficient, starts by writing

$$p_T(x, y) = \mathcal{G}_T(x, y) \exp\{A(y) - A(x)\}\psi(T, y)$$

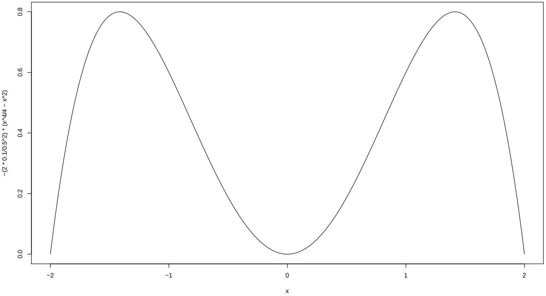
and tries to identify ψ. The connection with (49) and (50) are investigated.

Some results

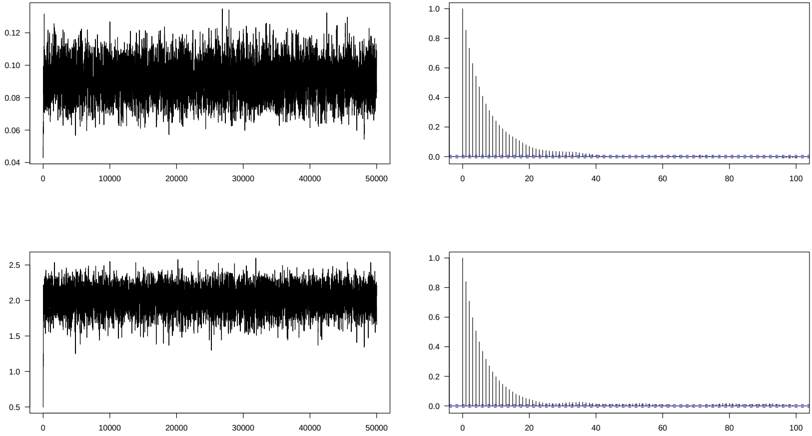
We apply the methods to the so-called double well potential model:

$$\mathrm{d}X_s = -\rho(X_s^3 - \mu X_s)\mathrm{d}s + \sigma \mathrm{d}B_s$$

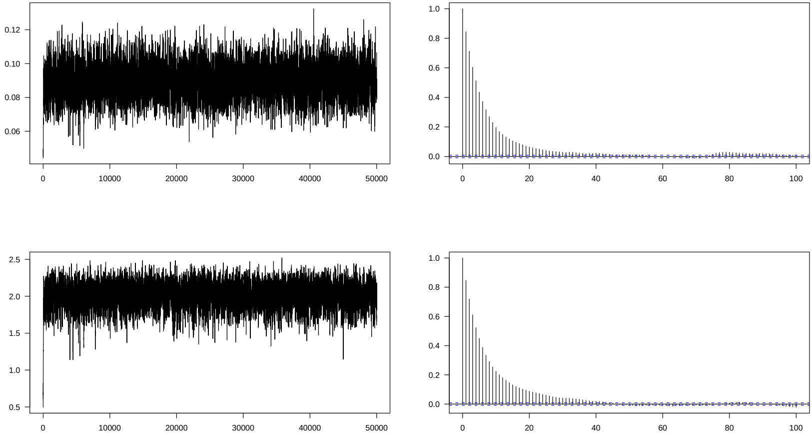
which has ergodic log-density given by $-(2\rho/\sigma^2)(x^4/4 - \mu x^2/2)$. We've simulated 1000 data with interobservation times 1, and $(\rho, \mu, \sigma) = (0.1, 2, 0.5)$



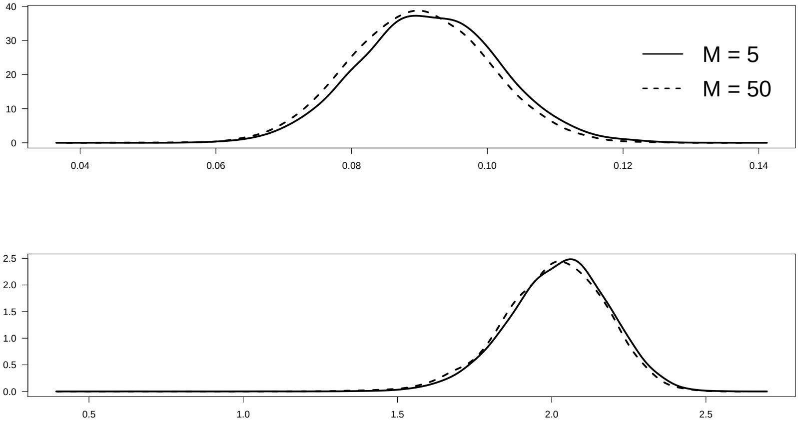
MCMC summaries $M = 5$



MCMC summaries $M = 50$



Posterior densities



Summary

We have fully addressed likelihood-based inference for discretey-observed diffusions when the diffusion coefficient is constant and known.

- Phrased problem as missing data
- Formulated a generic DA
- Probabilistically represented the distribution of missing data
- Developed MC methods for simulating efficiently from the missing data distribution

We can try to export this methodology to the general case. Before this, we address the efficiency of the diffusion bridge sampling methodology.