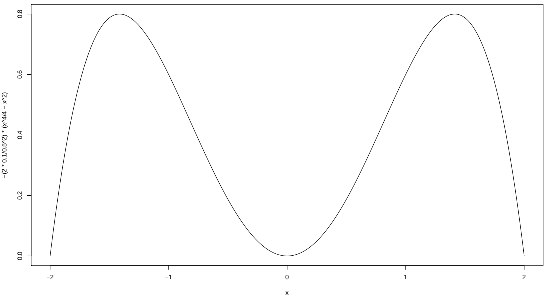
Some results

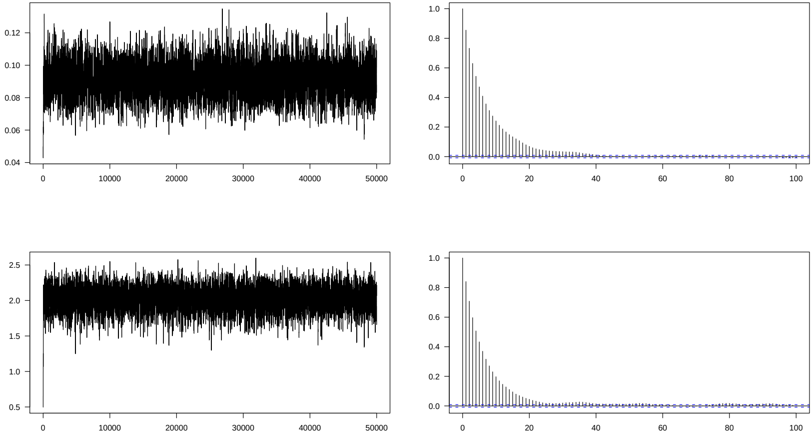
We apply the methods to the so-called double well potential model:

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

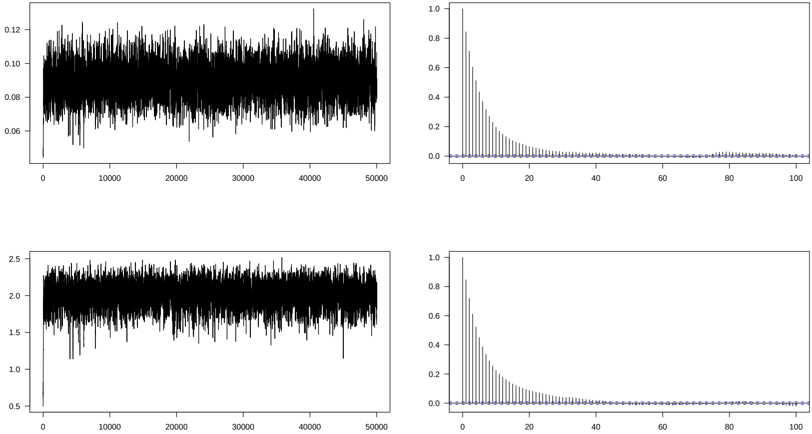
which has ergodic log-density given by $-(2\rho/\sigma^2)(x^4/4 - \mu x^2/2)$. We've simulated 1000 data with interobservation times 1, and $(\rho, \mu, \sigma) = (0.1, 2, 0.5)$



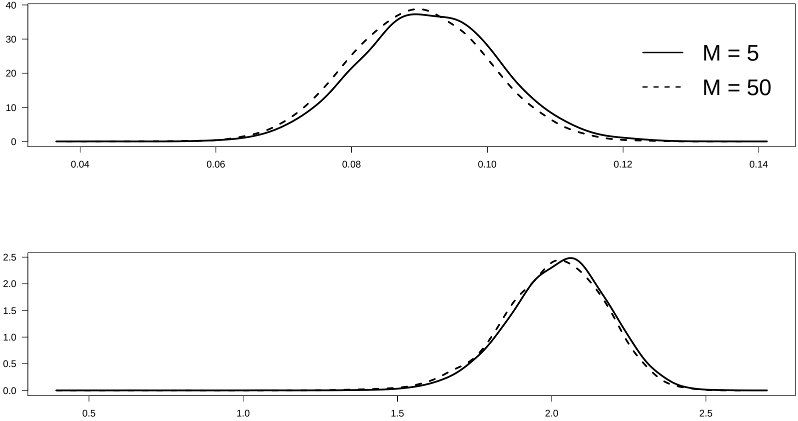
MCMC summaries $M = 5$



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.

Global sampling of diffusion bridges: problems

The approach we suggested for simulation of diffusion bridges is based on **global algorithms**, e.g IS or its MCMC variant, the independence MH algorithm.

Apriori, this is expected to scale badly with the time separation between the end points; exponentially bad typically.

We study a tractable example to get an understanding

The Ornstein-Uhlenbeck bridge

Recall the OU process (9). The corresponsding bridge process which travels from x to y at time T is given, working directly from the h-transform (44) or even from first principles since we deal with a Gaussian process, by

$$\mathrm{d}X_s = \left\{ -\alpha(X_s - \mu) + \frac{2\alpha e^{-\alpha(T-s)}}{1 - e^{-2\alpha(T-s)}}(y - e^{-\alpha(T-s)}(X_s - \mu) - \mu) \right\} \mathrm{d}s + \sigma \mathrm{d}B_s \tag{52}$$

Note that as $\alpha \to 0$ we obtain, as expected, the BB (30).

To avoid unnecessary notation, we focus on the case $\mu = 0, \sigma = 1$, $\alpha > 0$, and the ending point is at the stationary mean, $y = 0$. In this simplified setting, the bridge is given by

$$\mathrm{d}X_s = -\alpha X_s \frac{1 + e^{-2\alpha(T-s)}}{1 - e^{-2\alpha(T-s)}} \mathrm{d}s + \mathrm{d}B_s\,, \quad X_0 = x$$

Instead, we will propose from BB

$$\mathrm{d}X_s = -\frac{X_s}{T - s}\mathrm{d}s + \mathrm{d}B_s$$

and weight with

$$w(X) = \sqrt{\frac{(1 - e^{-2\alpha T})}{2\alpha T}} \exp\left\{\frac{x^2}{2}(-1/T + 2\alpha e^{-2\alpha T}/(1 - e^{-2\alpha T}))\right\}$$
$$\times \exp\left\{\frac{\alpha}{2}(x^2 + T) - \frac{\alpha^2}{2}\int_0^T X_s^2 \mathrm{d}s\right\}$$

Note the useful representation of a $BB(0, x, T, 0)$ X, in terms of a standard $BB(0, 0, 1, 0)$ Z:

$$X_s = \sqrt{T} Z_{s/T} + (1 - s/T)x \,, s \in [0, T] \tag{53}$$

thus

$$\int_0^T X_s^2 \mathrm{d}s \stackrel{\mathcal{L}}{=} T\mathcal{Z}_1 + \frac{T^2}{3}x^2 + 2\sqrt{T}x\mathcal{Z}_2$$

where the $\mathcal{Z}$'s are $\mathcal{O}(1)$ statistics,

$$\mathcal{Z}_1 = \int_0^T Z_{s/T}^2 \mathrm{d}s \quad \mathcal{Z}_2 = \int_0^T (1 - \frac{s}{T}) Z_{s/T} \mathrm{d}s$$

Thus, putting everything together we get that:
Therefore, we have

$$
\begin{aligned}
w(X) &= \sqrt{\frac{(1 - e^{-2\alpha T})}{2\alpha T}} \\
&\exp\left\{-\frac{x^2}{2}\left(\frac{1}{T} + \frac{\alpha^2 T^2}{3} - \alpha\frac{1 + e^{-2\alpha T}}{1 - e^{-2\alpha T}}\right) - \frac{\alpha T}{2}(\alpha\mathcal{Z}_1 - 1) - \alpha^2\sqrt{T}x\mathcal{Z}_2\right\} \\
&\propto \exp\left\{-\frac{\alpha T}{2}(\alpha\mathcal{Z}_1 - 1) - \alpha^2\sqrt{T}x\mathcal{Z}_2\right\}
\end{aligned}
$$

- $\alpha \to 0$, $w \to 1$.
- Starting in equilibrium, $x = 0$: for large T the weight of a path will either be exponentially small, if $\alpha\mathcal{Z}_1 > 1$ or exponentially large, if $\alpha\mathcal{Z}_1 < 1$
- Starting out of equilibrium, take $T = 1$: for large x we have a similar issue

Joint distribution of $\mathcal{Z}_1$ and $\mathcal{Z}_2$

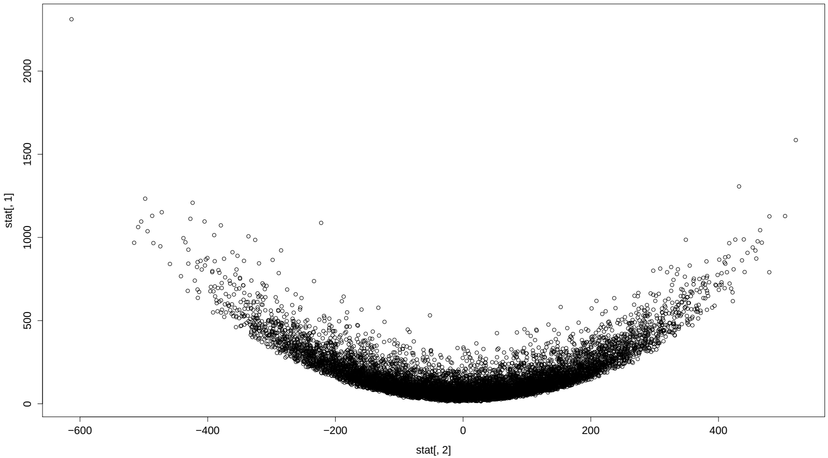
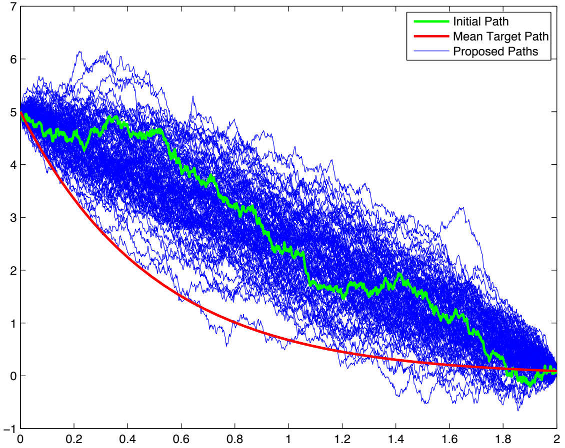


Illustration: global sampling OU bridge out of equilibrium



Local Path Sampling

We can construct MCMC algorithms which operate on the path space for simulating for path measures. Particularly relevant for bridge and other conditioned diffusion measures.

Theoretically, this has generated interesting advances in SPDE theory, see for example [Hairer et al., 2007]. For a recent article on these algorithms, see for example [Beskos et al., 2008b] (in fact the quadratic variation identity plays an important role in this field as well).

Actually, some of the algorithms can be justified in an elementary way, without resorting to SPDEs to justify them.

Random walk sampling of diffusion bridges

Recall that we wish to sample from (47), or equivalently from the law (49). Global sampling uses proposals from the dominating Brownian bridge measure $\mathbb{W}^{(T,x,y)}$

Note that, the transformation $X \to X'$, where

$$
\begin{aligned}
X'_s &= (1 - s/T)x + (s/T)y \\
&\quad + \rho\left(X_s - (1 - s/T)x - (s/T)y\right) + \sqrt{1-\rho^2}\, W_s \qquad (54) \\
&\quad s \in [0, T]\,, \quad W \sim \mathbb{W}^{(T,0,0)}\,, W \perp X\,, \quad \rho \in [0,1]
\end{aligned}
$$

is invariant w.r.t $\mathbb{W}^{(T,x,y)}$. To see this, note that it if $X \sim \mathbb{W}^{(T,x,y)}$ then X' is a Gaussian process, with the same mean and covariance as X.

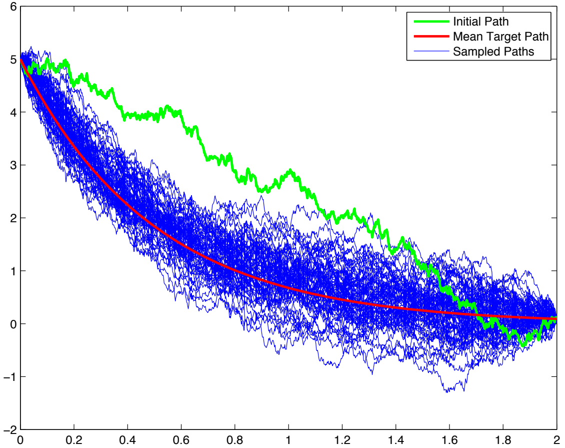
Note the special cases $\rho = 0, 1$

In fact, the transformation is also reversible w.r.t $\mathbb{W}^{(T,x,y)}$, in the sense that the joint measure on X and X' is symmetric. Then, the following algorithm yields a MH-MC on the path space which has $\mathbb{P}^{(T,x,y)}$ as a limiting distribution.

1. Start with a path X_0
2. Propose X' from (54)
3. Accept X' w.p. $G(X')/G(X)$, otherwise stay at the current path
4. Goto 2

Heuristics about why this scheme is valid

Illustration: local sampling OU bridge out of equilibrium



Part V:MC-based likelihood inference for discretely-observed reducible diffusions

- Collapse of basic DA when estimating volatility
- A toy example illustration
- Reducible diffusions
- Path transformations and efficient DA

To avoid excessive notation we focus on time-homogeneous diffusions

Framework

The problem can be appreciated even at the simplest case of unknown diffusion coefficient:

$$\mathrm{d}V_t = \alpha(V_t;\theta)\mathrm{d}t + \sigma \mathrm{d}B_t \tag{55}$$

with θ and σ unknown.

Replicating the previous approach we immediately run into a serious problem: existence of parameter-free dominating measure for DA: (42)

Therefore, we cannot design a DA which operates on path spaces. So what would happen if we tried the previous DA scheme on a discretization of the model?

A toy example

Let V_t be a Brownian motion with infinitesimal variance σ^2. Assume that $V_0 = 0$. Suppose that $V_1 = y$ is observed.

$$V_1 \sim N(0, \sigma^2)$$

Thus, given the prior $\sigma^{-2} \sim Gamma(1, 1)$,
the posterior for σ^{-2} is just

$$Gamma(3/2, 1 + y^2/2).$$

Data augmentation for the toy example

Suppose now for illustration, that the full likelihood is unavailable and data augmentation was necessary. We impute

$$V_{1/m}, V_{2/m}, \ldots, V_{(M-1)/M}\ .$$

We use the Gibbs sampling algorithm which iterates the following loop:

1. Given σ^2 impute a discretised Brownian bridge with infinitesimal variance σ^2 hitting $V_1 = y$ at time 1.
2. Given $V_0, V_{1/M}, V_{2/M}, \ldots, V_{(M-1)/M}, V_1$ draw σ^{-2} from

$$Gamma(1 + M/2, 1 + M\Sigma_V/2)$$

where Σ_V denote the quadratic variation:

$$\Sigma_V = \sum_{i=1}^{M} (V_{i/M} - V_{(i-1)/M})^2\ .$$

More on the toy example

The following result shows that for this example the convergence time of the algorithm is $\mathcal{O}(M)$ as M becomes large.

Let $\tau^{(M)}$ be the inverse variance process for the algorithm which imputes $M-1$ points. It can be shown that by speeding up $\tau^{(M)}$ by a factor of $M/4$, $\tau^{(M)}_{[tM/4]}$ converges weakly as $M \to \infty$ to a Langevin diffusion with stationary distribution given by the posterior.

Theorem

[Roberts and Stramer, 2001] Let $\mathbf{P}^{(M)}$ *be the law of* $\tau^{(M)}_{[tM/4]}$

$$\mathbf{P}^{(M)} \Rightarrow \mathbf{P}^{(\infty)}$$

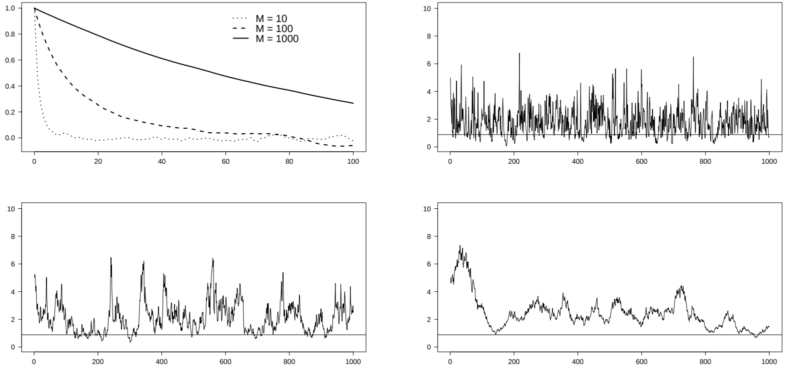
where $\mathbf{P}^{(\infty)}$ *is the law of the diffusion*

$$\mathrm{d}\xi_t = \xi_t\{5/4 - \xi_t(1/2 + X_1^2/4)\mathrm{d}t + \mathrm{d}B_t\} \ .$$

The convergence time is thus $\mathcal{O}(M)$.
ξ has stationary distribution Gamma$(3/2, 1 + X_1^2/2)$.

The fact that the algorithm is at least $\mathcal{O}(M)$ can be seen from the generic characterization of the convergence of DA in (40). Taking $h(\tau) = \tau$, we have that

$$\gamma \geq 1 - \frac{\frac{1+M/2}{(1+M\Sigma_V/2)^2}}{\frac{3/2}{1+y^2/2}} \stackrel{\text{large } M}{\approx} 1 - (4 + 2y^2)\frac{1}{3M}$$



Degeneration of the MCMC method for increasing M, but also recall plot in 128

Efficient DA using reparametrizations

An efficient DA with convergence time $\mathcal{O}(1)$ in the amount of imputation can be implemented.

In fact, it is based on a valid DA which is based on path imputation (i.e $M = \infty$).

The problem we face here arises in many other contexts, and the solution we will pursue is an instance of a general methodology, the so-called *non-centred* parameterizations, see [Papaspiliopoulos et al., 2003]

For diffusions, it will be achieved using the tools we've already used: the transformation to unit diffusion coefficient (33), and the tilting of BB paths (53), together with a general trick for obtaining laws of transformed processes.

Reducible diffusions

In multivariate setting if $\eta : R^d \to R^n$

$$\mathrm{d}\eta(V) = A\eta(V)\mathrm{d}s + \nabla\eta(V)\sigma(V)\mathrm{d}B$$

so we need η s.t:

$$\nabla\eta(V)\Gamma(V)(\nabla\eta(V))^* = I$$

A sufficient condition when $d = m = n$, which for example in [Aït-Sahalia, 2008] is given as the definition of reducible diffusions, is to obtain: $\nabla\eta(V)\sigma(V) = I$.

The conditions under which this holds are transparent when σ^{-1} exists.

It is then easy to see that $\partial \eta_k / \partial v_j = [\sigma^{-1}]_{kj}$ and since $\partial^2 \eta_k / \partial v_j \partial v_l$ should yield the same result regardless of the order of differentiation, we get the necessary condition:

$$\frac{\partial [\sigma^{-1}]_{kj}}{\partial v_l} = \frac{\partial [\sigma^{-1}]_{kl}}{\partial v_j}$$

This is also sufficient, since we can define $\eta_k = \int [\sigma^{-1}]_{kj} \mathrm{d} v_j$ (any j can be chosen). This function then solves the desired system. The conditions and proof when σ is not invertible are more intricate. [Aït-Sahalia, 2008] only proves this special case.

For intuition consider an SV model for $d = 2$ with $\sigma_{12} = \sigma_{21} = 0$.