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:

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

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* $\tau^{(M)}_{[tM/4]}$

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

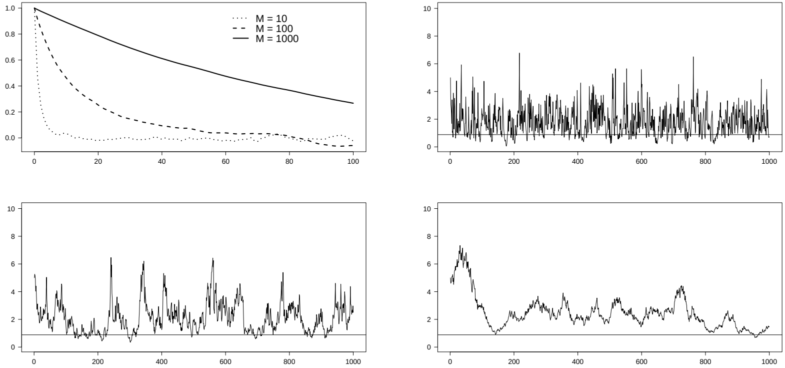
where $\mathbf{P}^{(\infty)}$ *is the law of the diffusion*

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

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

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

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



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.

In fact, we will capitalize on the work we've already done in understanding diffusion bridges

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 $\sigma_{12} = \sigma_{21} = 0$.

Setup

$$\mathrm{d}V_s = b(s, V_s; \theta)\,\mathrm{d}s + \sigma(V_s; \theta)\,\mathrm{d}B_s, \quad s \in [0, T] \; ; \qquad (56)$$

and assume that V is a reducible diffusion, i.e there exists the transformation

$$V_s \to X_s := \eta(V_s; \theta)$$

such that

$$\mathrm{d}X_s = \alpha(s, X_s; \theta)\mathrm{d}s + \mathrm{d}B_s$$

We wish to infer parameters and missing data on the basis of an observed skeleton $\mathbf{v} = \{v_0, v_1, \ldots, v_n\}$. (We focus again on discretely observed diffusions, although all these techniques carry over to all other partial observation schemes)

Recall the missing data formulation: $\mathbf{v}$, V^m, V^c

DA collapses here due to the perfect dependence of σ and V^m

Let $x_i(\theta) = \eta(v_i; \theta)$, which depends on θ via σ, and $\mathbf{x}(\theta) = \{x_0(\theta), \ldots, x_n(\theta)\}$

Transformations under an equivalent measure

A key idea is that it is enough to decouple the perfect dependence between missing data and parameters under the dominating measure (due to equivalence of measures). In particular we can do it under the proposal measure we have used in constructing the diffusion bridge.

We show how to write $V^m = g(\widetilde{V^m}, \theta)$ where $\widetilde{V^m}$ and θ are independent under the proposal measure. We also show how to obtain the joint distribution and density of the $(\widetilde{V^m}, \theta)$

The principle behind this approach is that since the measures are equivalent, they have the same almost sure events, therefore making σ and $\widetilde{V^m}$ independent under the proposal measure, will have the effect that they are not perfectly dependent under the target measure. The transformation is ideal under the dominating, it is just enough to lead to a convergent DA under the target.

The transformation

Recall that a bridge from v to w is built by first transforming to $\eta(v;\theta)$ and $\eta(w;\theta)$ and then proposing a Brownian bridge

This is not enough though since the BB depends deterministically on the parameters through the endpoints

There is a standard way to avoid this: standarize the bridge

$$V_t = \eta^{-1}(\tilde{V}_t + (1 - t/T)\eta(v;\theta) + (t/T)\eta(w;\theta)\,;\theta)$$

Note that under the proposal measure for V^m, $\widetilde{V^m} \sim \mathbb{W}^{(T,0,0)}$. Its distribution under the diffusion bridge measure is intractable. On the other hand, it is easy to simulate it under the diffusion bridge measure...

We now find the joint density of θ and $\widetilde{V^m}$ conditionally on observed data, say $\pi(\theta, \widetilde{V^m})$.

We first derive the conditional density of $\widetilde{V^m}$ given $\theta, \mathbf{v}$. A simple trick to find this, is to think of it as an importance sampling exercise: how should I weight draws $\tilde{V}$ from the proposal measure $\mathbb{W}^{(T,0,0)}$ in order to obtain diffusion bridges from v to w?

The answer is actually given in (49). Therefore, the conditional density of $\widetilde{V^m}$ w.r.t a parameter independent Brownian bridge measure is given by

$$\frac{\mathcal{G}_{0,T}(x(\theta), y(\theta))}{p_{0,T}(x(\theta), y(\theta))}$$
$$\exp\left\{\int_0^T \alpha(s, X_s(\theta, \tilde{V}))^* \mathrm{d}B_s - \frac{1}{2}\int_0^T [\alpha^*\alpha](s, X_s(\theta, \tilde{V}))\mathrm{d}s\right\}$$

On the other hand, note that the marginal posterior, in other words, the observed data posterior, is (we give it just for two successive observations at times 0, T)

$$\pi(\theta \mid v, w) \propto \pi(\theta) p_{0,T}(v, w; \theta) = \pi(\theta) p_{0,T}(x(\theta), y(\theta); \theta) J(v, \theta)$$

where J is the Jacobian transformation stemming from η.

Therefore,

$$\pi(\theta, \widetilde{V^m} \mid v, w) \propto \mathcal{G}_{0,T}(x(\theta), y(\theta))\pi(\theta)J(v,\theta)$$
$$\exp\left\{\int_0^T \alpha(s, X_s(\theta,\tilde{V}))^* \mathrm{d}B_s - \frac{1}{2}\int_0^T [\alpha^*\alpha](s, X_s(\theta,\tilde{V}))\mathrm{d}s\right\}$$

This is only given for two data points, for $n > 2$ it will be a product of such terms due to the Markov property.

From this we easily obtain $\pi(\theta \mid \widetilde{V^m}, v, w)$ up to proportionality.

The algorithm

Iterate the following steps

1. Simulate V^m according to the diffusion bridge (by independence MH, random walk on paths, etc)
2. Transform $V^m \to \widetilde{V^m}$
3. Simulate from $\pi(\theta \mid \widetilde{V^m}, v, w)$ (directly, by MH, etc)

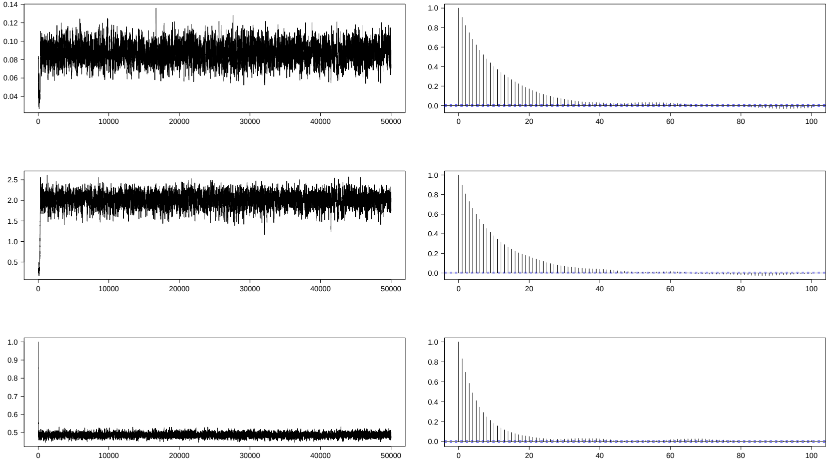
Some results

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

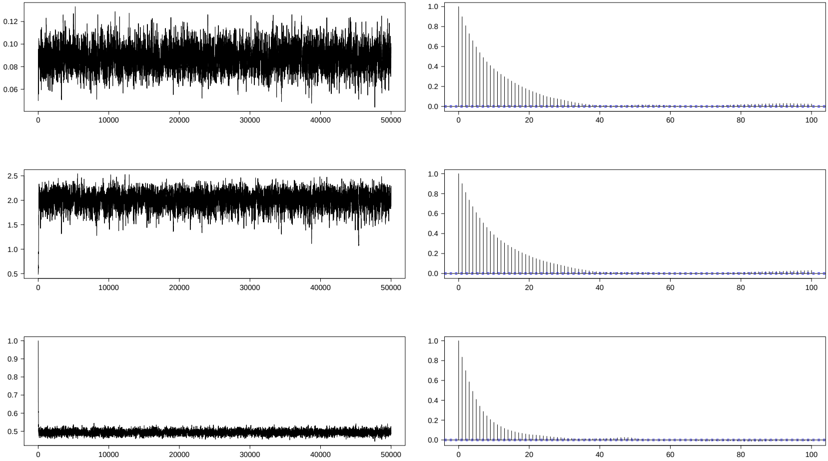
$$\mathrm{d}V_s = -\rho(V_s^3 - \mu V_s)\mathrm{d}s + \sigma \mathrm{d}B_s$$

where we've simulated 1000 data with interobservation times 1, and $(\rho, \mu, \sigma) = (0.1, 2, 0.5)$. This is the same dat as before but all parameters are treated as unknown.

MCMC summaries $M = 5$



MCMC summaries $M = 50$



Posterior densities