Scott Hounsell and Sam Pfeiffer
Progress Report 10/26/15

We will begin with a brief summary of the paper by section in plain terms. We also provide a summary of our progress in replicating the results for a one-dimensional process. This can be found in Section 5. I have skipped some sections containing lengthy and involved proofs.

1 Introduction

We define some d-dimensional stochastic process by

$$X_0 = x_0, \qquad \text{and} \qquad dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t.$$

The goal is to estimate

$$V_0 \coloneqq E[g(X_T)] \qquad \text{where } \; g: R^d \to R$$

Our simulation which we will use to estimate will be denoted like so

$$\widehat{X_0} = x_0, \qquad \text{and} \qquad d\widehat{X_t} = \hat{\mu}\left(t, \widehat{X_t}\right)dt + \hat{\sigma}\left(t, \widehat{X_t}\right)dW_t$$

where $\hat{\mu}: R^d \to R^d$ and $\hat{\sigma}: R^{dxd} \to R^{dxd}$ will be updated at exponential time points which I will henceforth refer to as "arrivals".

Comment [S1]: In order to compensate for the change of the coefficients of the SDE, our main representation result relies on the automatic differentiation technique induced by Elworthy's formula from Malliavin calculus, as exploited by Fourni´e et al. [10] for the simulation of the Greeks in financial applications.

Automatic differentiation technique from Elworthy's formula from Malliavin calculus.

This technique allows us to deal with gradients and Hessian's in the coefficient corrections (the Malliavin weights), avoiding the need to limit the PDE's we wish to simulate.

Also, note that for our purposes, a single dimension suffices. Thus, throughout the paper we can consider $d = 1$.

2 Regime switching diffusion representation

This section begins by discussing some brief Lipschitz and other "niceness" conditions for the coefficients and g functions.

A number of notations are defined here which I will summarize in simpler terms. T_k is the time of the k^{th} arrival (in fact the minimum of this with the final time T, to stay inside the time range). T_0 is 0. N_t is the number of arrivals strictly before t and is a Poisson process with intensity β. If we are at t, then $\left(T_{N_t}, \widehat{X_{T_{N_t}}}\right)$ are the time and location of the most recent arrival. ΔW_t^k is the change in the driving Brownian motion between T_{k-1} and T_k (unless $t < T_k$, in which case it is the change from T_{k-1} to t). The authors also define shifted variables T_k^t, the k^{th} arrival since t, and N_s^t, the number of arrivals between t and s (where $t \leq s$). Note that this is not used at all throughout the rest of the paper. Some more notation is covered but serves mainly to confuse the reader.

Next, we define the Malliavin weights $\widehat{W}^1$ and $\widehat{W}^2$ as

$$D^i E\left[\phi\left(T_1^t, \widehat{X_{T_1^t}}\right)\right] = E\left[\phi\left(T_1^t, \widehat{X_{T_1^t}}\right)\widehat{W}^i(T_1^t, \Delta W^{1,t})\right], \quad i = 1,2$$

where D^1 is the gradient, D^2 is the Hessian, and $\phi: R^d \to R$ (essentially some proxy for g).

$$\widehat{W}_k = \left(\widehat{W}^1, \widehat{W}^2\right)(T_{k+1}^t, \Delta W^{k+1,t}) \in R^d, R^{dxd}$$

$$\Delta f_k \coloneqq (\mu, a)\left(T_k^t, \hat{X}_{T_k^t}\right) - (\hat{\mu}, \hat{a})\left(T_k^t, \hat{X}_{T_k^t}\right) \in R^d, R^{dxd}$$

where

$$a(\bullet) = \frac{1}{2}\sigma\sigma^T(\bullet) \quad \text{and} \quad \hat{a}(\bullet) = \frac{1}{2}\hat{\sigma}\hat{\sigma}^T(\bullet)$$

Essentially, it seems that this notation is just saying that the parameters in the right tuple are unpacked into each function in the left tuple (i.e. $\widehat{W}^1(T_{k+1}^t, \Delta W^{k+1,t}) \in R^d$ and $\widehat{W}^2(T_{k+1}^t, \Delta W^{k+1,t}) \in R^{dxd}$). Δf

appears to be understood as error terms for our estimators of μ and σ, which are being "corrected" via $\widehat{W}^1$ and $\widehat{W}^2$.
Next we define $\hat{\psi}$, which amounts to an estimator for g in that it should have the same expectation at T.

$$\hat{\psi} := e^{\beta(T-t)}\left(g(X_T) - g\left(X_{T_{N_T^t}}\right)1_{N_T^t>0}\right)\beta^{-N_T^t}\prod_{k=1}^{N_T^t}\Delta f_k \bullet \widehat{W}_k$$

(Note that we start from t, $\prod_{k=1}^{0} = 1$ and $(p, P) \bullet (q, Q) = p \cdot q + P : Q$). In this case the colon operator is defined as $A : B = Tr(AB^T)$. Thus the term inside the product is something like $(\mu - \hat{\mu}) \cdot \widehat{W}^1 + \left(\frac{1}{2}\sigma\sigma^T - \frac{1}{2}\hat{\sigma}\hat{\sigma}^T\right) : \widehat{W}^2$. This looks like the error of each coefficient dotted with a correction weight.

It is stated and subsequently proven that $\hat{\psi}$ has the same expectation as $g(x)$. In addition there is a brief section discussing error analysis.

3 The constant diffusion coefficient case

This section offers an "algorithm" for SDE's with drift functions and a constant matrix diffusion coefficient:

$$dX_t = \mu(t, X_t)dt + \sigma_0 dW_t$$

Note that $\sigma_0 \in R^{dxd}$. We assume that g and μ are Lipschitz and "nice".

3.1 The algorithm

It is proposed that we choose $\hat{\mu} = \mu$ and $\hat{\sigma} = \sigma$. The discretized process is thus

$$\hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \mu\left(T_k, \hat{X}_{T_k}\right)\Delta T_{k+1} + \sigma_0 \Delta W_{T_{k+1}}, \quad k = 0..N_T$$

Next, the formulas for the weights are simply given, and the reader is referred to the likelihood ratio method of Broadie and Glasserman. Studying that paper, I have yet to find how the application yields the following weights, but nonetheless, they are

$$\widehat{W^1}(\delta t, \delta w) := (\sigma_0^T)^{-1}\frac{\delta w}{\delta t}$$

Note that the $\widehat{W^2}$ is given but is not needed for this constant diffusion case (no correction term necessary). Therefore, starting at 0, our estimator for g is

$$\hat{\psi} := e^{\beta T}\left(g\left(\hat{X}_T\right) - g\left(\hat{X}_{T_{N_T}}\right)1_{N_T>0}\right)\beta^{-N_T}\prod_{k=1}^{N_T}\overline{W_k^1}$$

where $\overline{W_k^1}$ apparently represents the combined error/correction product and is defined as

$$\overline{W_k^1} := \frac{\left(\mu\left(T_k, \hat{X}_{T_k}\right) - \mu\left(T_{k-1}, \hat{X}_{T_{k-1}}\right)\right) \cdot (\sigma_0^T)^{-1}\Delta W_{T_{k+1}}}{\Delta T_{k+1}}$$

Comment [S.H.2]: Note that this does not seem to correlate perfectly with the Δf "error" term. It seems that this is a difference over time rather than between the actual and estimated μ. Thus we are unsure how exactly it has been derived.

A proof of the square integrability of $\hat{\psi}$ follows, but is skipped for the purposes of this summary. Similarly the author justifies a choice for the intensity β, which we will skip for now, as well.

4 One-dimensional driftless SDE

The process we wish to simulate in this section is $dX_t = \sigma(t, X_t)dW_t$.

4.1 The algorithm

It is proposed that we choose $\hat{\mu} = \mu = 0$ and $\hat{\sigma}(s, y, t, x) = \sigma(s, y) + \partial_x\sigma(s, y)(x - y)$
Note that s and y represent the time and location at the most recent arrival. Technically all of our $\hat{\mu}$ and $\hat{\sigma}$ functions have these parameters in front, since our process is regime-switching, however, they have been irrelevant and unnecessary notation until now. Essentially, $\hat{\sigma}$ looks like a Taylor expansion

across the time between now and the last regime change (sort of our last “known” time point).

Therefore our discretized process is

$$d\widehat{X_t} = \left(\sigma\left(T_k, \widehat{X_{T_k}}\right) + \partial_x\sigma\left(T_k, \widehat{X_{T_k}}\right)\left(\hat{X}_t - \widehat{X_{T_k}}\right)\right) dW_t$$

Depending on whether or not the partial derivative is zero, there are two potential explicit solutions given. For brevity the following notations are used

$$c_1^k = \sigma\left(T_k, \widehat{X_{T_k}}\right) - \partial_x\sigma\left(T_k, \widehat{X_{T_k}}\right)\widehat{X_{T_k}}$$
$$c_2^k = \partial_x\sigma\left(T_k, \widehat{X_{T_k}}\right)$$

If $c_2^k = 0$,

$$\hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \sigma\left(T_k, \widehat{X_{T_k}}\right)\Delta W_{T_{k+1}}$$

Otherwise,

$$\hat{X}_{T_{k+1}} = -\frac{c_1^k}{c_2^k} + \frac{c_1^k}{c_2^k}\exp\left(-\frac{\left(c_2^k\right)^2}{2}\Delta T_{k+1} + c_k^2\Delta W_{T_{k+1}}\right) + \hat{X}_{T_k}\exp\left(-\frac{\left(c_2^k\right)^2}{2}\Delta T_{k+1} + c_k^2\Delta W_{T_{k+1}}\right)$$

Our estimator for g is

$$\hat{\psi} := e^{\beta T}\left(g\left(\hat{X}_T\right) - g\left(\hat{X}_{T_{N_T}}\right)1_{N_T>0}\right)\beta^{-N_T}\prod_{k=1}^{N_T}\overline{W_k^2}$$

where again $\overline{W_k^2}$ apparently represents the combined error/correction product and is defined as

$$\overline{W_k^2} := \frac{a\left(T_k, \widehat{X_{T_k}}\right) - \tilde{a}_k}{2a\left(T_k, \widehat{X_{T_k}}\right)}\left(-\partial_x\sigma\left(T_k, \widehat{X_{T_k}}\right)\frac{\Delta W_{T_{k+1}}}{\Delta T_{T_{k+1}}} + \frac{\left(\Delta W_{T_{k+1}}\right)^2 - \Delta T_{T_{k+1}}}{\left(\Delta T_{T_{k+1}}\right)^2}\right)$$

where

$$a(\cdot) := \frac{1}{2}\sigma^2(\cdot)$$
$$\tilde{a}_k(\cdot) := \frac{1}{2}\tilde{\sigma}_k^2(\cdot)$$
$$\tilde{\sigma}_k := \sigma\left(T_{k-1}, \hat{X}_{T_{k-1}}\right) + \partial_x\sigma\left(T_{k-1}, \hat{X}_{T_{k-1}}\right)\left(\hat{X}_{T_k} - \hat{X}_{T_{k-1}}\right)$$

Since $\hat{\psi}$ is of infinite variance, the author provides an antithetic variable $\hat{\psi}^-$, which we will not cover here.

The choice of Malliavan weights are proven to be satisfactory following this section. However, their derivation is not shown, and we will skip this proof.

5 Numerical examples

This section provides a numerical example of the Exact method described in the previous sections. Equation 5.1 is a driftless process with a non-constant diffusion coefficient. The implementation for this type of process is described in section 4. Equation 5.2 is a process with a non-constant drift coefficient and a constant diffusion coefficient of 1. The implementation for this type of process is described in section 3. Note that equation 5.2 is the same as in 5.1 with a Lamperti transform applied. Figure 1 on page 18 shows the implied volatility of the process in 5.1 and 5.2 using both the Euler scheme and the Exact method described in the paper.

The code we have provided implements equation 5.1 using both the Euler scheme and the Exact method. Using Euler scheme, we match the implied volatility shown in figure 1. However, our implementation of the Exact method does not converge and sometimes gives negative values. We are not sure why.

Furthermore, we did not implement equation 5.2 since it depends on equation 5.1 which is not working properly. We did however implement the Exact simulation methods for processes of the same form as equation 5.2 (using section 3). We applied it to the process

$$dX_t = -\beta(X_t - \bar{X})dt + dW_t$$

This process did not converge either using the Exact method. Again, we are not sure why. Note that the simulation of this process using Euler scheme *does* converge.

One idea that we have not yet implemented, is to code the Exact simulation method described in Exact Simulation of Diffusions by Beskos and Roberts (attached). The paper we have been referencing improves upon the Beskos and Roberts paper since it works for multi-dimensional processes. Since this does not concern us, we will try to implement the method described in Beskos and Roberts.

6 Further discussions

The following subsections discuss application to general drift and diffusion along with path-dependent cases.

6.1 The general drift and diffusion case

Here, the paper offers a formula and gives general weights. However, the author remarks that our g estimator, $\hat{\psi}$, has infinite variance and is therefore unsuitable in application in general. The remaining sections concerning path dependence are irrelevant to us.