

# Inference Experiments on some Stochastic Differential Equations (miscellaneous toy models)

Waleed Alhaddad

May 22, 2019

## Toy models

We generate paths from an SDE with given parameters and then fit the model to check the retrievability of the parameters. We consider the following cases:

- 1: Constant parameter Wiener process: finite and infinite horizon.
- 2: Controlled Wiener process within an interval.
- 3: Controlled Wiener process centered around a given function.

# Stochastic Differential Equation

Given the SDE

$$\begin{aligned}dX_t &= a(X_t; \boldsymbol{\theta})dt + b(X_t; \boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= x_0\end{aligned}\tag{1}$$

Consider a set of  $N$  observations,  $X = \{X_{t_0^N}, X_{t_1^N}, \dots, X_{t_N^N}\}$  observed within intervals  $\Delta_N$ .

- ▶  $\Delta_N$ : is any function of the number of samples,  $N$ . For example,  $\Delta_N = \frac{1}{N}$ .
- ▶  $a(X_t; \boldsymbol{\theta})$  is the drift function (add technical conditions here)
- ▶  $b(X_t; \boldsymbol{\theta})$  is the diffusion function (add technical conditions here)
- ▶  $\boldsymbol{\theta}$  is a vector of constant parameters

## likelihood function

We have the Likelihood function of the transition density

$$\mathcal{L}(\theta; X) = \prod_{i=1}^n \rho(X_{i+1}|X_i, \theta) \rho(X_0) \quad (2)$$

Consider a Gaussian transition density,

$$\mathcal{L}(\theta; X) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi b(X_i, \theta)^2 dt}} \exp\left(\frac{-((X_{i+1} - X_i) - a(X_i, \theta)dt)^2}{2b(X_i, \theta)^2 dt}\right) \rho(X_0)$$

We take the improper prior  $\rho(X_0) = 1$ .

## Case 1: Constant Parameters

For a fixed  $b(X_t; \boldsymbol{\theta}) = \sigma$  and a fixed  $a(X_t; \boldsymbol{\theta}) = \mu$ . We observe that varying the sampling and discretization parameters affects the convergence behavior.

We consider the following three situations:

- 1.1 Infinite Horizon expansion caused by increasing samples  $N$   
( increasing  $N$  and  $\Delta_N$  is fixed).
- 1.2 Infinite Horizon expansion caused by increasing intervals  $\Delta_N$   
( fixed  $N$  and increasing decreasing  $\Delta_N$ ).
- 1.3 Finite Horizon expansion with sample compression, i.e where  $\Delta_N = \frac{T}{N}$   
(increasing  $N$ , fixed  $T$ ).

## Case 1: Implementation

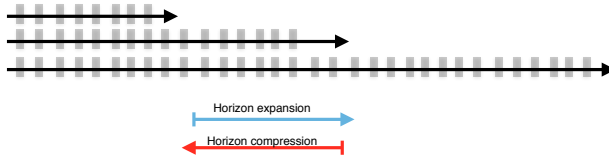
In the following, we use L-BFGS-B algorithm to optimize the likelihood with bounds on  $\sigma$  being positive. And we use  $\mu = 100$  and  $\sigma = 10$  for the path generation. We denote the estimators of  $\mu$  and  $\sigma$  by  $\hat{\mu}$  and  $\hat{\sigma}$  where,

$$\hat{\mu} = \arg \max_{\mu \in \mathbb{R}} \mathcal{L}(\boldsymbol{\theta}; X) \qquad \hat{\sigma} = \arg \max_{\sigma \in \mathbb{R}^+} \mathcal{L}(\boldsymbol{\theta}; X) \qquad (3)$$

The above cases can be seen clearly in the illustration below.

### Infinite Horizon

Fixed interval  $dt$  and increasing number of samples  $N$  (case 1.1)

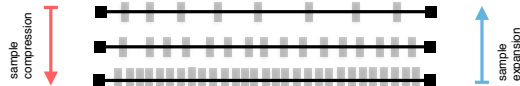


Fixed number of samples  $N$  and increasing interval  $dt$  (case 1.2)



### Finite Horizon

Increasing number of samples  $N$  and fixed time horizon  $T$  (case 1.3)



## Case 1.1: Constant Parameters

### Infinite Horizon expansion caused by increasing samples $N$

An increase of the number of samples  $N$  in a growing time interval  $T$ . In this case, we obtain convergence of the order  $\frac{1}{\sqrt{N}}$  as expected.

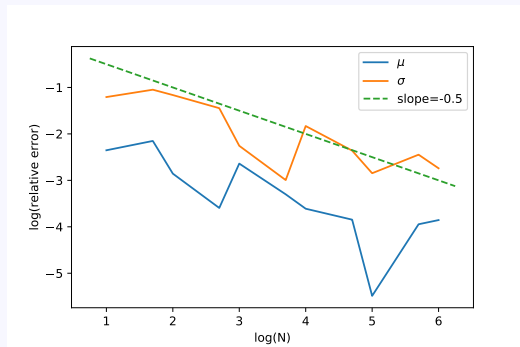


Figure 1: convergence in the case of infinite horizon expansion caused by increasing samples  $N$



## Case 1.2: Constant Parameters

### Infinite Horizon expansion caused by increasing intervals $dt$

In this case, the estimator  $\hat{\sigma}$  of the diffusion coefficient is inconsistent. This is because the time interval between the samples is increasing and so we do not have as much information about the variation of the process. However, the drift estimator  $\hat{\mu}$  is still consistent. We have not seen any theoretical proof of this situation.

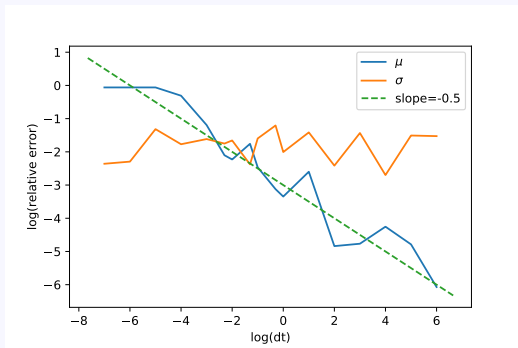


Figure 2: convergence rate in the case of infinite horizon expansion caused by increasing intervals  $\Delta_N$

## Case 1.3: Constant Parameters

### Finite Horizon expansion with sample compression

In this case, we observe increasing number of samples in a fixed time interval (sample compression). In this case the drift estimator  $\mu$  is inconsistent while the diffusion estimator  $\sigma$  is consistent. This situation matches the theoretical discussion in [1].

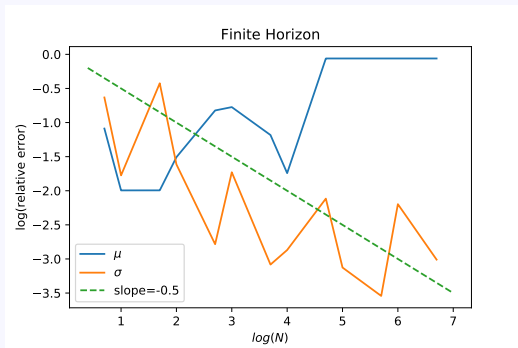


Figure 3: Convergence rate in the case of finite horizon expansion with sample compression

## Remarks

1. It is easier to obtain convergence in the diffusion coefficient. Because the estimator of the diffusion has access to information from the infinite variation of the Wiener process while the drift does not. We can show this theoretically from the efficiency condition of the estimators.
2. The convergence generally follows Monte Carlo's convergence rate in terms of sampling as expected. This statement is wrong in a high frequency setting, as we have different convergence rate for estimator of the diffusion and than for the estimator of the drift as we will see next. The estimator of the drift converges slower than the Monte Carlo rate given that both are consistent.

## Motivation of high frequency SDE data

High frequency situations are not uncommon and extreme but may happen for data that looks reasonable. For instance, in some cases weekly data is treated as high frequency data. A situation is considered high frequency with respect to the amount of variation in the drift term [1] (or in regards to the speed at which it reverts to the mean).

This also important for building confidence intervals as the limiting distribution of our estimator may not be Gaussian. For example, in the setting of high frequency data and fixed time  $T$  (case 1.3 above) the limiting distribution of the estimator is path dependent and is given by a normal variance-mixture distribution which can be heavy tailed with no finite moments [1]. However, in this case, a modification of the estimator can help in achieving a Gaussian limiting distribution of the estimator.

Please note that in the setting of high frequency SDE data, we have two possibilities:

- ▶ Finite Horizon (fixed time  $T$ ): this corresponds to case 1.3 above.
- ▶ Infinite Horizon: this corresponds to other two cases.

## Convergence of the drift estimator for high frequency SDE (Case 1.3 / Finite Horizon)

In a high frequency setting with finite horizon (case 1.3), there are **no** consistent estimators of the drift term  $a(X_t, \theta)$  (see [1]) as we break the following condition :

Given the following SDE,

$$\begin{aligned} dX_t &= a(X_t; \alpha)dt + b(X_t; \beta)dW_t \quad t > 0 \\ X(0) &= x_0 \end{aligned} \tag{4}$$

the drift estimator  $\hat{\alpha}$  of the parameter set  $\alpha$  of the system is consistent if and only if

$$n\Delta_n \rightarrow \infty \tag{5}$$

And this condition contradicts having a finite horizon as we trivially have  $T = N\Delta_N \rightarrow \infty$ . If the condition (5) is not satisfied, we cannot have a consistent estimator  $\hat{\alpha}$  of the drift.

However, the estimator  $\hat{\beta}$  of the diffusion term  $b(X_t; \beta)$  is still consistent regardless.

## Convergence of the drift estimator for high frequency SDE (Case 1.3 / Finite Horizon)

To show this result numerically, we pick any  $\Delta_n$  such that  $T = n\Delta_n \rightarrow \infty$ . In this example, we modify the non-consistent drift parameter estimator (case 1.3) such that

$$dt = \frac{T}{N} \quad \rightarrow \quad dt = \frac{T}{\sqrt{N}} \quad (6)$$

Then we recover back the convergence of the other cases as this replacement turns the finite horizon to an infinite horizon.

## Numerical Convergence for high frequency SDE

In this numerical example, we take  $\Delta_N = \frac{T}{\sqrt{N}}$  as an example.

$$\text{Finite Horizon: } dt = \frac{T}{N} \quad \text{versus} \quad \text{Infinite Horizon: } dt = \frac{T}{\sqrt{N}} \quad (7)$$

We confirm condition (5) numerically. It is clear that in the finite horizon case the estimator of the diffusion  $\hat{\sigma}$  converges while the estimator of the drift  $\hat{\mu}$  is not consistent. Whereas, in the infinite horizon case, we obtain the expected convergence in both parameters.

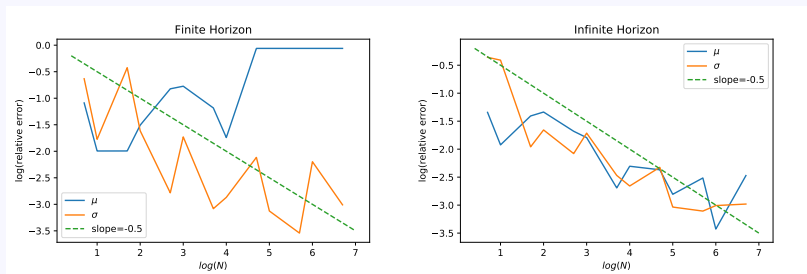


Figure 4: Right side: Finite horizon with  $dt = \frac{T}{N}$ , Left side: Infinite horizon with  $dt = \frac{T}{\sqrt{N}}$

## Convergence for high frequency SDE data

We can show that the score function of the likelihood is an efficient martingale estimating function and that the obtained estimator is optimal<sup>1</sup>. Then we have the following convergence rates [3]:

For a finite Horizon (fixed  $T$ )

$$\hat{\alpha} \text{ is not consistent} \quad \text{and} \quad \text{relative error}(\hat{\beta}) \propto \frac{1}{\sqrt{N}} \quad (8)$$

For an infinite Horizon ( $T \rightarrow \infty$ )

$$\text{relative error}(\hat{\alpha}) \propto \frac{1}{\sqrt{N\Delta_N}} \quad \text{and} \quad \text{relative error}(\hat{\beta}) \propto \frac{1}{\sqrt{N}} \quad (9)$$

Note that the estimator  $\hat{\alpha}$  of the drift is affected by the time discretization while the estimator  $\hat{\beta}$  of the diffusion is independent of the time discretization.

---

<sup>1</sup>If we consider an approximate martingale estimator function such as the pseudo likelihood, then we require an extra condition that  $n\Delta_n^{2\kappa-1} \rightarrow 0$  where  $\kappa$  is the order of  $\Delta_N$  in RHS of the expectation (11)



## High Frequency SDE General Convergence

Consider a set of  $N$  observations,  $X = \{X_{t_0^N}, X_{t_1^N}, \dots, X_{t_N^N}\}$  and the following martingale estimator function,

$$G_n(\gamma) = \sum_{i=1}^N g(\Delta_N, X_{t_i^N}, X_{t_{i-1}^N}; \gamma) \quad (10)$$

where  $t_i^n = i\Delta_n$ ,  $\gamma = (\alpha, \beta)$  and  $g = (g_1, g_2)$ . In our martingale case<sup>2</sup>, we must have,

$$E_\gamma \left( g(\Delta_N, X_{t_i^N}, X_{t_{i-1}^N}; \gamma) \middle| X_{t_{i-1}^N} \right) = 0 \quad (11)$$

where the  $\hat{\beta}$  estimator is given by solving  $G_n(\gamma) = 0$ .

---

<sup>2</sup>The results can also be adapted to a non martingale estimating function such as a pseudo likelihood

## References

- [1] Jakobsen, N. M., & Sørensen, M. (2017). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. *Bernoulli*, 23(3), 1874-1910.  
doi:10.3150/15-bej799
- [2] Sørensen, M. (2007). Efficient Estimation for Ergodic Diffusions Sampled at High Frequency. *SSRN Electronic Journal*. doi:10.2139/ssrn.1150694
- [3] Kessler, M., Lindner, A., & Sørensen, M. (2012). *Statistical methods for stochastic differential equations*. Boca Raton, FL: Chapman & Hall/CRC.

## Paths as IID samples

We have the Likelihood function of the transition density

$$\mathcal{L}(\theta; X) = \prod_{j=1}^M \prod_{i=1}^N \rho(X_{j,i+1} | X_{j,i}, \theta) \rho(X_0) \quad (12)$$

Consider a Gaussian transition density,

$$\mathcal{L}(\theta; X) = \prod_{j=1}^M \prod_{i=1}^N \frac{1}{\sqrt{2\pi b(X_{j,i}; \theta)^2 dt}} \exp\left(\frac{-((X_{j,i+1} - X_{j,i}) - a(X_{j,i}; \theta)dt)^2}{2b(X_{j,i}; \theta)^2 dt}\right) \rho(X_0)$$

We take the improper prior  $\rho(X_0) = 1$ .

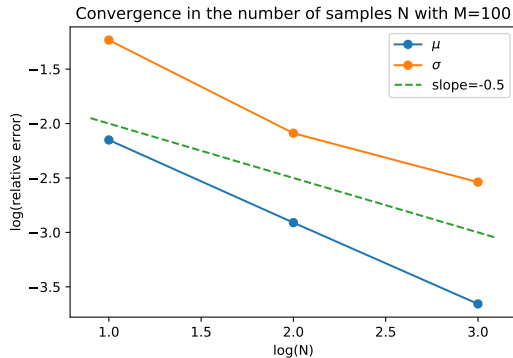
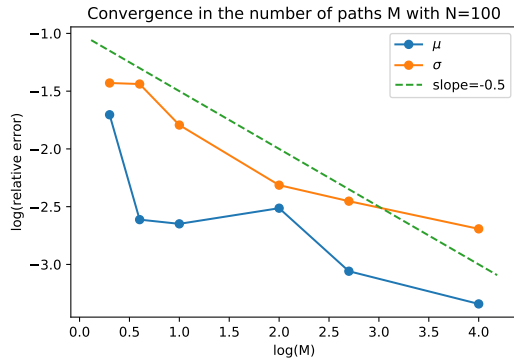
## Paths as IID samples - Implementation

As before, we use L-BFGS-B algorithm to optimize the likelihood with bounds on  $\sigma$  being positive. We denote the estimators of  $\mu$  and  $\sigma$  by  $\hat{\mu}$  and  $\hat{\sigma}$  where,

$$\hat{\mu} = \arg \max_{\mu \in \mathbb{R}} \mathcal{L}(\boldsymbol{\theta}; X) \qquad \hat{\sigma} = \arg \max_{\sigma \in \mathbb{R}^+} \mathcal{L}(\boldsymbol{\theta}; X) \qquad (13)$$

## Paths as IID samples - Convergence

We have Monte Carlo convergence in terms of samples as expected.



**Figure 5:** Optimized using L-BFGS-B algorithm. True parameters used are  $\mu = 100$  and  $\sigma = 10$ .  $M$  is the number of paths and  $N$  is the number of samples.

## Case 2: Controlled Wiener process in a unit box

We have the SDE,

$$\begin{aligned}dX_t &= a(X_t; \boldsymbol{\theta})dt + b(X_t; \boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= 0\end{aligned}\tag{14}$$

Consider a set of  $N$  observations,  $X = \{X_{t_0^N}, X_{t_1^N}, \dots, X_{t_N^N}\}$  on intervals of  $\Delta_N$ .

We choose,

$$a_{i+1}(X_i; \boldsymbol{\theta}) = \mu X_i \left(\frac{1}{2} - X_i\right) \quad b_{i+1}(X_i; \boldsymbol{\theta}) = \sigma X_i(1 - X_i)\tag{15}$$

We choose  $0 < \sigma \ll 1$  and  $\mu \in (0, 1]$ . Then, we have contained the Wiener process in the interval  $[0, 1]$  with high probability.

## Case 2: Controlled Wiener process in a unit box

We see that the process is contained within the unit box as we kill the diffusion coefficient if we touch the boundary. Additionally, the drift coefficient is always directed towards the center line  $\frac{1}{2}$  of the unit box.

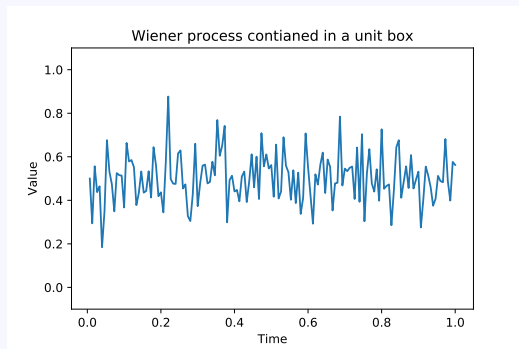
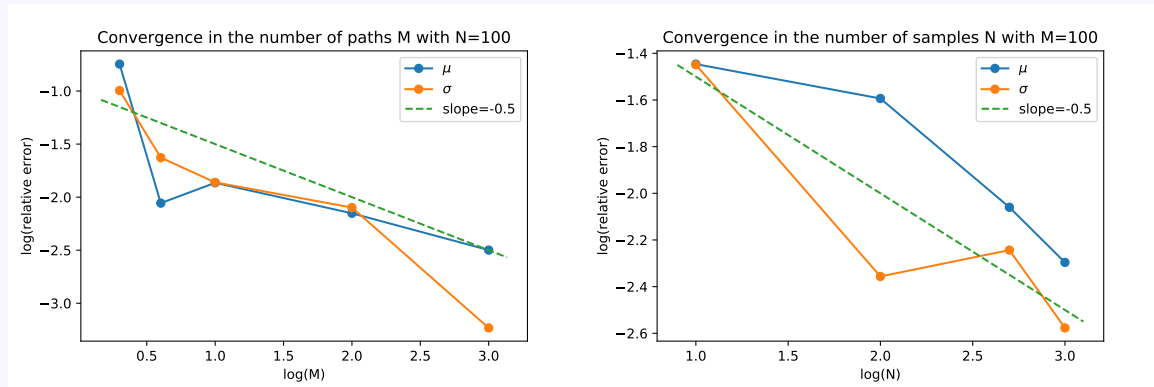


Figure 6: With high probability the wiener process stays inside the box as  $\mu \in (0,1]$  and  $\sigma \ll 1$

## Case 2: Controlled Wiener process in a unit box - Parameter Inference

We obtain the same convergence as before,



**Figure 7:** Optimized using L-BFGS-B algorithm. True parameters used are  $\mu = 0.5$  and  $\sigma = 0.1$ . M is the number of paths and N is the number of samples.



### Case 3: Controlled Wiener process around a given function

We have the SDE,

$$\begin{aligned}dX_t &= a(X_t; \boldsymbol{\theta})dt + b(X_t; \boldsymbol{\theta})dW_t \quad t > 0 \\ X_0 &= 0\end{aligned}\tag{16}$$

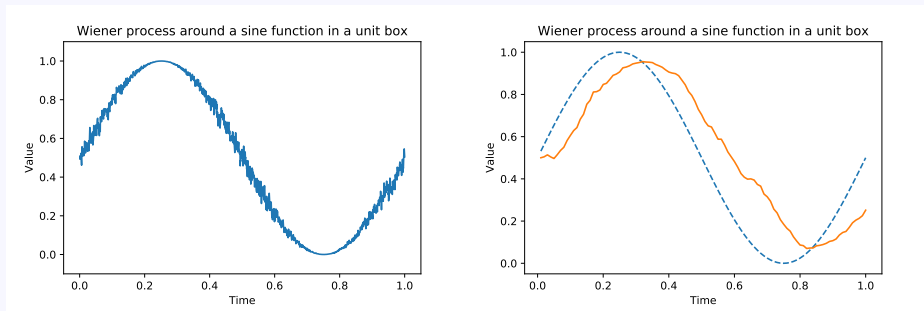
Consider a set of  $N$  observations,  $X = \{X_{t_0^N}, X_{t_1^N}, \dots, X_{t_N^N}\}$  on intervals of  $\Delta_N$ . We choose,

$$a_{i+1}(X_i; \boldsymbol{\theta}) = \mu X_i \left( \frac{1}{2}(1 + p_i) - X_i \right) \quad b_{i+1}(X_i; \boldsymbol{\theta}) = \sigma X_i (1 - X_i) \tag{17}$$

We choose  $0 < \sigma \ll 1$  and  $\mu \in (0, 1]$ . Then, we have contained the Wiener process in the interval  $[0, 1]$  with high probability. Additionally, we add mean reversion to  $\frac{1}{2}(1 + p_i)$  where  $p_i = \sin(\frac{i}{N}2\pi)$ .

### Case 3: Controlled Wiener process following a given function

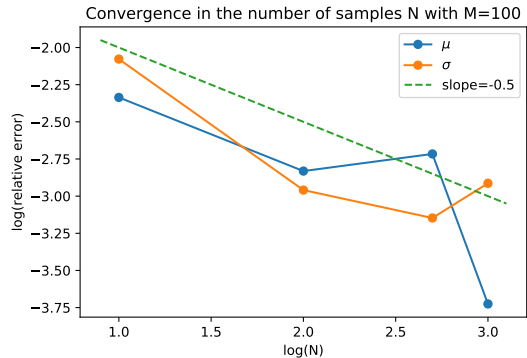
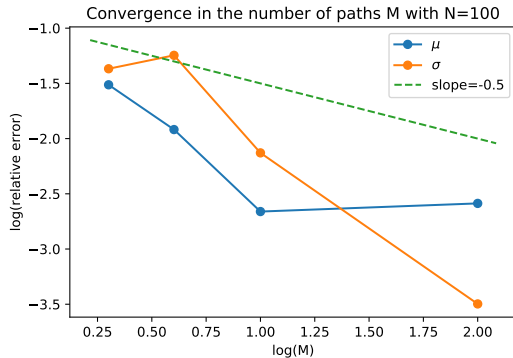
We see that the process follows the sine wave closely when the drift parameter  $\mu \approx 1$  and that we have a shift artifact when the drift parameter  $\mu \ll 1$ . Note that the process does not escape the unit box as the diffusion is controlled and the drift is directed towards the function which is always inside the unit box.



**Figure 8:** Illustrative examples: On the right, we see the controlled wiener process responding to the boundary by killing the diffusion at the peaks of the sine wave. On the right we observe an unintended shift artifact when the drift parameter  $\mu \ll 1$ .

## Case 3: Controlled Wiener process around a function - Parameter Inference

We obtain the same convergence as before,



**Figure 9:** Optimized using L-BFGS-B algorithm. True parameters used are  $\mu = 100$  and  $\sigma = 10$ .  $M$  is the number of paths and  $N$  is the number of samples.