Chapter ω

An Introduction to Inversion Methods for Stochastic Differential Equations

Outline

- Primer to Stochastic Differential Equations (SDEs)
- Numerical solution of SDEs
- Inverse methods for estimating SDE models
- Exercises in `MATLAB`: `github.com/williamjsdavis/SIO230-Stochastic`
- Further topics (time permitting)

Introduction

In the physical sciences, governing equations and physical phenomena are often described by differential equations. However, we often need to consider systems which are driven by rapid, fast fluctuations that appear to act "randomly." To make this extension, I will informally introduce stochastic differential equations. The estimation—or inversion—of SDE model equations from empirical data is a developing field, with the main pioneering paper being published in 1998 (Siegert, Friedrich, and Peinke, 1998). I will attempt to show some simple results and examples from this field.[1]

[1]While this lesson aims to provide an introduction to SDEs and SDE inversion methods, I can point the interested student to various resources for further reading. Gardiner et al. (1985) provides a superb introduction to the use of stochastic processes in science fields, and is a great book to learn from. A natural extension from there is Nicolaas Godfried Van Kampen (1992), which covers more detailed examples. For a more mathematically rigorous approach, Evans (2012) builds up the field from measure

ω.1 Ordinary Differential Equations

To introduce stochastic differential equations, I start from first-order, ordinary differential equations. Such systems are written like

$$\frac{d}{dt}X(t) = f\big(X(t)\big), \qquad (\omega.1)$$

where $X(t)$ is the variable of interest, t is time, and $f(X(t))$ is some deterministic function that models how the current state of the system effects the rate of change of $X(t)$. Additional dimension can also be added to the system, forming a *system* of ODEs.[2] An ODE of this form has a solution

$$X(t) = X_0 + \int_0^t f\big(X(s)\big)\ ds, \qquad (\omega.2)$$

with some initial condition $X(t=0) = X_0$. As long as the function $f(\circ)$ is Riemann-integrable, a unique solution of the integral will exist. When ODEs are used to describe the evolution of a system with time, it is often called a *dynamical system* (e.g., Strogatz, 2018).

Exercise 1.1: Examples of ODE applications

Draw a spider diagram of fields and applications in Earth Sciences (or other areas in physical sciences) that rely on (systems of) differential equations.

ω.2 Stochastic Differential Equations

With ODEs as a starting point, I now introduce a precursor to stochastic differential equations (SDEs).

ω.2.1 Langevin Equations

Start with previous physical motivation, with a system where the rate of change of a quantity is determined by some function of the state of the system, i.e. equation (ω.1). Consider the situation where the deterministic response compounded with an amount of some rapidly and irregularly fluctuating "random noise", $\xi(t)$. The amplitude of this fluctuating term can be set by the scalar quantity, σ. The differential equation is now written as

theoretical fundamentals, and Pavliotis (2014) continues this with more mathematical rigor. Risken (1996) is a great reference which gives a thorough treatment not just of the Fokker-Planck equation, but many other details of the field with physics based examples. For a collection of definitions, proofs, references, and inversion methods, students are referred to Rudolf Friedrich, Joachim Peinke, et al. (2011).

[2] We consider autonomous systems, although often non-autonomous systems can be made autonomous by introducing additional variables (e.g., Strogatz, 2018).

$$\frac{d}{dt}X(t) = f\big(X(t)\big) + \sigma\xi(t). \qquad (\omega.3)$$

An equation of this form is referred to as a *Langevin equation*, after Langevin's reinterpretation of Brownian motion (Langevin, 1908). They are characterised by having a fixed strength of noise fluctuations, indicated here as σ. At this point, $\xi(t)$ has not been defined yet, so it's not clear how to interpret this equation. Similarly to equation (ω.2), the integral form of (ω.3) can be written as

$$X(t) = X_0 + \int_0^t f\big(X(s)\big)\ ds + \sigma \int_0^t \xi(s)\ ds, \qquad (\omega.4)$$

with some initial condition $X(t = 0) = X_0$. To concentrate on the new part that is associated with the random component, the unscaled process can be defined as

$$Y(t) = \int_0^t \xi(s)\ ds. \qquad (\omega.5)$$

This equation is the integral of the "random noise" $\xi(t)$, which can be conceptually thought of as the the derivative of process $Y(t)$,

$$\frac{dY}{dt} = \xi(t). \qquad (\omega.6)$$

To progress further, certain choices for the form of $\xi(t)$ must be made. These choices will be based on assumptions for the "random noise" in the physical system being modeled. One common assumption is that $Y(t)$ (and as a result, $X(t)$) is a continuous function of time. This is equivalent to assuming that the process contains no discontinuous jumps. It can be shown that only the first two statistical moments of $\xi(t)$ need to be defined (Risken, 1996). The first moment (i.e. the mean) can be defined to vanish, because, without loss of generality, any systematic bias in $\xi(t)$ can be absorbed into the definition of $f(\circ)$. As such,

$$E(\xi(t)) = 0. \qquad (\omega.7)$$

For the second moment, one might want to consider the case where each instance of this noise is completely uncorrelated from all other values. Put another way, for all non-identical times, $t \neq t'$, the noise $\xi(t)$ and $\xi(t')$ are statistically independent,

$$E(\xi(t)\xi(t')) = \delta(t - t'), \qquad (\omega.8)$$

where $\delta(\circ)$ is the Dirac delta function. These choices for the form of random fluctuations represented in $\xi(t)$ are referred to as *Gaussian white noise.* The label "Gaussian" refers to the fact that $Y(t)$ is Gaussian distributed (which is a result of its assumed continuity), and "white" refers to the flat spectral content of this process, akin to the spectra of white light.

These choices give some familiar statistical properties to the integral form of this noise, $Y(t)$. Firstly, the expectation of all increments of $Y(t)$ vanish

$$E\big(Y(t) - Y(s)\big) = \int_s^t E(\xi(t'))\ dt' = 0. \tag{ω.9}$$

Furthermore, the expectation of the square of the increments gives a familiar result

$$\begin{aligned} E\big(\big[Y(t) - Y(s)\big]^2\big) &= \int_s^t dt' \int_s^t dt''\ E(\xi(t')\xi(t'')) \\ &= \int_s^t dt' \int_s^t dt''\ \delta(t' - t'') \\ &= t - s, \end{aligned} \tag{ω.10}$$

for $t > s$. It turns out that (ω.9) and (ω.10) show the same properties as a mathematical object called the Wiener process, $W(t)$. This process, $W(t)$, is defined as having the following properties:

1. Sample paths $W(t)$ are continuous,
2. Increments are Gaussian distributed, $W(t + \Delta t) - W(t) \sim \mathcal{N}(0, \Delta t)$ for all t and $\Delta t > 0$,
3. Furthermore, these increments are independent for all times, i.e. values of $W(t + \Delta t) - W(t)$ are independent from all past values of $W(t')$, where $t' \leq t$.

We shall now throw rigor joyously to the wind and define

$$W(t) = \int_0^t ds \xi(s), \tag{ω.11}$$

and

$$dW_t = \xi(t) dt. \tag{ω.12}$$

Thus, we interpret integral (ω.4) as

$$X(t) = X_0 + \int_0^t f\big(X(s)\big)\ ds + \int_0^t\ dW(s), \tag{ω.13}$$

Here, the first integral is identical to (ω.2), whereas the second integral is interpreted as a Stieltjes integral with respect to a stochastic sample function

$$dW(t) \equiv W(t + dt) - W(t) = \xi(t) dt. \tag{ω.14}$$

The theory of how to take an integral with respect to a stochastic process is developed in the field of Itô calculus (Pavliotis, 2014). However it turns out that for some simple SDEs—such as Langevin equations—solutions of $X(t)$ can be found without having to consider how this integral is interpreted.[3]

[3] Mathematicians will express stochastic integrals (ω.13) in the form $dX = f\big(X(t)\big)dt + dW$, even though objects like dX don't make much sense on their own.

ω.2.2 Stochastic Differential Equations

A more general version of the Langevin equation is the stochastic differential equation, where the amplitude of the noise is dependent on the state

$$\frac{dX}{dt} = f\big(X(t)\big) + g\big(X(t)\big)\xi(t). \qquad (\omega.15)$$

Here, $f(x)$ and $g(x)$ and are called the drift and diffusion functions, respectively (Risken, 1996). Equation (ω.15) is interpreted as a stochastic integral equation

$$X(t) = X_0 + \int_0^t f\big(X(s)\big)\ ds + \int_0^t g\big(X(s)\big)\ dW(s). \qquad (\omega.16)$$

To compute the solution to this, one must consider defining the stochastic integral

$$\int_{t_0}^t g\big(X(s)\big)\ dW(s) \qquad (\omega.17)$$

as a kind of Reimann-Stieltjes integral. Namely, the interval in time $[0, t]$ is partitioned into n subintervals

$$t_0 \le t_1 \le t_2 \le \dots t_{n-1} \le t, \qquad (\omega.18)$$

with intermediate points τ_i such that

$$t_{i-1} \le \tau_i \le t_i. \qquad (\omega.19)$$

Then the solution of the stochastic integral (ω.17) is defined as the limit of the partial sums

$$\int_{t_0}^t g\big(X(s)\big)\ dW(s) = \lim_{n\to\infty} \sum_{i=1}^n g\big(X(\tau_i)\big)\Delta W_i, \qquad (\omega.20)$$

where ΔW_i is an increment in a Wiener process

$$\Delta W_i = W(t_i) - W(t_{i-1}). \qquad (\omega.21)$$

In regular Reimann-Stieltjes integration, equations similar to (ω.20) converge to the same result that does not depend on choices of the intermediate points τ_i. However this is not the case for stochastic integrals.

ω.2.3 Itô vs. Stratonovich

The choice of evaluation points τ_i has direct consequences on the interpretation an result of (ω.20). This is generally true for stochastic integrals where the integrand function $g(\circ)$ is non-constant. In principle, any choice for the intermediate points τ_i can be taken. However there are two options that are commonly used in practice: evaluation at the start of partition points t_i; and evaluation at the midpoints of partitions. These two choices correspond to the Itô and Stratonovich interpretations, respectively (e.g., Nicolaas G Van Kampen, 1981; Mannella and McClintock, 2012). For Itô calculus (Itô, 1944; Itô, 1951), the interpretation of (ω.20) becomes

$$\int_{t_0}^{t} g\big(X(s)\big)\, dW(s) = \lim_{n\to\infty} \sum_{i=1}^{n} g\big(X(t_{i-1})\big)\big[W(t_i) - W(t_{i-1})\big]. \qquad (\omega.22)$$

The Itô interpretation provides more mathematically straightforward functionality (and easy numerical implementations), but with the complication of alterations to the chain rule. As such, it is most often used in mathematical proofs, finance, and numerical simulations.[4] In the following material, the integral is interpreted in the Itô sense, for computational ease.

Exercise 2.1: Examples of SDE applications

Extend your spider diagram from Exercise 1.1 to of fields and applications in Earth Sciences (or other areas in physical sciences) that you think SDEs would be applicable.

ω.3 Numerical Solution of Stochastic Differential Equations

Like ODEs and PDEs, the vast majority of SDEs cannot be solved by analytic methods. Instead, solutions are approximated through the application of numerical methods (Kloeden and Platen, 2013). Consider a general, scalar SDE of the form

$$\frac{dX}{dt} = f\big(X(t)\big) + g\big(X(t)\big)\xi(t), \qquad (\omega.23)$$

where $\xi(t)$ is Gaussian white noise. An initial condition at $X(t = 0)$ must be provided, which can either be deterministic or a random variable. We seek an approximate solution on the interval $[0, T]$. For this, both the sampling of the abscissa and the measure on the ordinates have to be specified.

ω.3.1 Conditions for Numerical Solutions

First, the interval $[0, T]$ is discretized to N points, specifying sample points at times t_i. This is usually accomplished with regular grid and a constant step size, Δt,

$$X(t_i) := X(i\Delta t), \qquad (\omega.24)$$

with $i = 0, 1, \dots, N$ and $T = N\Delta t$.

[4] An alternative to the Itô interpretation is to take the Stratonovich interpretation (R. Stratonovich, 1966; R. L. Stratonovich, 1967), and evaluate $g(\circ)$ at midpoints. This interpretation is more favoured in the physical sciences, as it preserves the chain rule, as well as being a more natural model for real noise. However in practice, the choice is a decision in modeling, as connecting the two interpretations is possible.

Strong Solutions

Next, a measure of "approximate solution" must be defined. This corresponds to defining the conditions of convergence. We consider pathwise convergence, also referred to as *strong* convergence, which is related to the L^1 norm. If $\hat{X}_{\Delta t}(t)$ is an approximate solution of $X(t)$ with sampling Δt, the pathwise error at some point t_i is defined as

$$\text{Error}_\text{s}\,(\Delta t) = E\left|X(t_i) - \hat{X}_{\Delta t}(t_i)\right|. \tag{ω.25}$$

From this, strong convergence is defined as[5]

$$\lim_{\Delta t \to 0} \text{Error}_\text{s}\,(\Delta t) = 0. \tag{ω.26}$$

ω.3.2 The Euler-Maruyama Method

One of the simplest numerical solutions for a SDE of the form (ω.23) an extension of the explicit Euler method (e.g., Butcher, 2016). The Euler-Maruyama (EM) method (Maruyama, 1954), can be expressed as

$$X(t_{i+1}) = X(t_i) + f\big(X(t_i)\big)\Delta t + g\big(X(t_i)\big)\Delta W(t_i), \tag{ω.27}$$

where $\Delta W(t_i) = W(t_{i+1}) - W(t_i)$ are "Brownian increments". These increments are, by the definition of the Wiener process, independent and identically distributed Gaussian random variables with mean zero and variance Δt. It is possible to rewrite (ω.27) as

$$X(t_{i+1}) = X(t_i) + f\big(X(t_i)\big)\Delta t + g\big(X(t_i)\big)\sqrt{\Delta t}\zeta_i, \tag{ω.28}$$

where $\zeta \sim \mathcal{N}(0,1)$. This method provides a numerical solution for (ω.23), given a "noise sample" of ζ_i. It can be shown that this method does strongly converge in the limit $\Delta t \to 0$. [6]

ω.3.3 Exercise: Using the Euler-Maruyama Method

Exercise 3.1: Using the Euler-Maruyama Method (`example3_1.m`)

Use the given `MATLAB` implementation of the Euler-Maruyama method to integrate the following "Ornstein-Uhlenbeck" (OU) SDE, $\frac{dX}{dt} = -X(t) + \xi(t)$. The result should look something like Fig. ω.1. Try integrating the solution again and see what happens (it should look like Fig. ω.3).

---

[5] We can also define the order of strong convergence, α, is defined when there exists a positive constant C such that $E\left|X(t_i) - \hat{X}_{\Delta t}(t_i)\right| \leq C\Delta t^{\alpha}$, for all i.

[6] The strong order of convergence $\alpha = \frac{1}{2}$. This compares with the deterministic Euler's method, which has a strong convergence of $\alpha = 1$.

Exercise 3.2: Other SDEs (`example3_2.m`)

Use the Euler-Maruyama method to integrate an SDE of your own choice. Choose something like

$$f_1(x) = 4x^3 + 2x - 100, \quad g_1(x) = 1,$$

$$f_2(x) = -2x + 5, \quad g_2(x) = \frac{1}{2} + \frac{1}{4}x^2,$$

$$f_3(x) = 3x, \quad g_3(x) = \frac{1}{5}x,$$

or come up with your own example! Congratulations, you are now qualified for a quantitative finance role.

Use the following function.[7]

```
function [t,x] = euler_maruyama(f,g,dt,tend,x0)
% Euler-Maruyama method
t = 0:dt:tend;
n = length(t);
x = zeros(size(t));
x(1) = x0;
for i = 2:n
    x(i) = x(i-1) + f(x(i-1))*dt + g(x(i-1))*sqrt(dt)*randn;
end
end
```

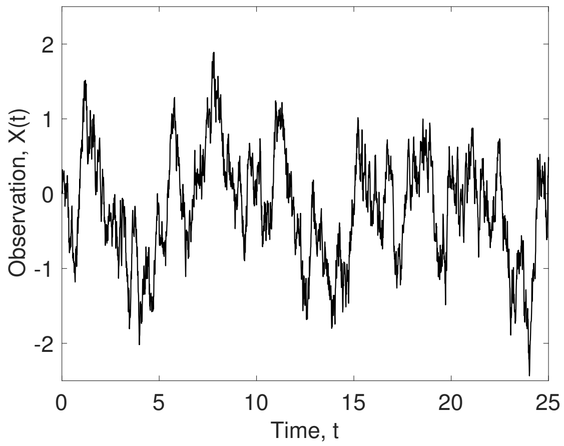

Figure ω.1: A realization of an OU process, $X(t)$, given in Example 3.1.

[7] Also found in `github.com/williamjsdavis/SIO230-Stochastic`.

ω.4 The Inverse Problem: Estimating Stochastic Differential Equations from Empirical Data

In many physical applications, it may be possible to model time time-variations of a system as a stochastic process, or more specifically, as a stochastic differential equation. From this outset, there are two options available. One option is to approach the system from first principles, and write down the fundamental mechanisms and conservation laws that will combine together and form governing equations. If one assumes that some parts of this system are modeled with fast-varying, random variables, then the governing equations will be SDEs. Such a hypotheses can then be verified against empirical data. On the other hand, if the fundamental mechanisms are not well known or inordinately complex, then a "top-down" approach can be attempted. In this case one must start with the empirical data, and attempt to infer what SDE could have plausibly produced it. This approach then becomes the process of fitting the SDE to time-series data, and can be thought of as an *inverse problem.* A schematic of this idea is shown in Fig. ω.2.

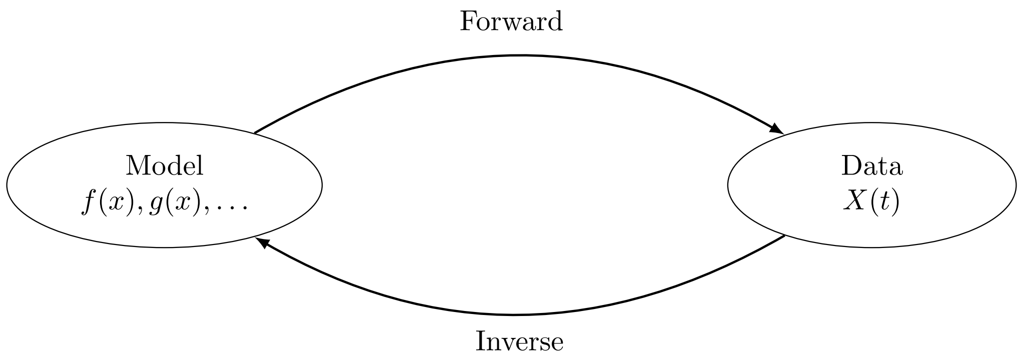

Figure ω.2: A schematic of stochastic forward and inverse modeling, showing the relationship between the model objects, (e.g. drift and noise functions $f(x)$, $g(x)$, etc...) and data objects (e.g. time-series observations, $X(t)$).

The history of this approach has largely been motivated financial modeling, as well as examples in the physical sciences. For the financial research, most inference methods are parametric, featuring rigorous proofs relating to specific financial models (Bishwal, 2007; Sørensen, 2012). For the physical sciences, non-parametric inference methods are more common, possibly due to the lack of domain prior knowledge.[8] In the statistical literature, the process of reconstructing an SDE from an SDE realization is called "estimation" or "inference," as the "true" stochastic models do actually exist. Unfortunately, this terminology is often incorrectly repeated in the physical sciences, where there is no "true"

[8] In these cases rigorous proofs for many of these estimation methods are absent. Many methods have only been demonstrated to work numerically, for a small set of empirical examples.

stochastic model. Instead, stochastic models are "fitted" to empirical data. In this section I choose to keep the "estimate" terminology, as it more accurately reflects the literature I am summarizing. However, when empirical data is concerned, I will exchange this verb for "fit."

Consider a scalar SDE of the form

$$\frac{dX}{dt} = f\big(X(t), t\big) + g\big(X(t), t\big)\xi(t), \qquad (\omega.29)$$

with an initial condition $X(t = 0) = X_0$. Here, for now, $\xi(t)$ is uncorrelated white noise, i.e. $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$. From this process, as discrete set of observations of $X(t)$ are made,

$$\{X(t)\} := \{X(t_1); X(t_2); \ldots; X(t_N)\}. \qquad (\omega.30)$$

The task is to estimate functions $f(x)$ and $g(x)$, either parametrically or non-parametrically.

An important point to illustrate is that to estimate the drift and noise functions, a "guess and check" forward modeling approach will not be successful. The key is that there is no way to reproduce the random state of $\xi(t)$. Even if one assumes the drift and noise functions have been, by chance, estimated correctly, then a realization of (ω.29) with the estimated functions will produce a solution of $X(t)$ that diverges from the data. Conceptually, this can be thought of having a different "random seed". An example of this is shown in Fig. ω.3. In probability theory terminology, such a "guess and check" method will almost surely never succeed.

In order to robustly estimate the original drift and noise functions of a SDE, or alternatively fit the drift and noise functions of a proposed model to an empirical time-series observation, a number of statistical methods have been developed. In the physical sciences, these methods can broadly be grouped into: "moment based estimation" methods; "transition probability estimation" methods; and others.

ω.5 Moments Based Estimation

The most common technique present in the literature is one based on analysis of conditional moments. The basic theory of this approach relates to something called the Fokker-Planck equation: a partial differential equation describing the time-evolution of the probability density function, $P(x, t)$, written here as

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\Big(D^{(1)}(x,t)P(x,t)\Big) + \frac{\partial^2}{\partial x^2}\Big(D^{(2)}(x,t)P(x,t)\Big). \qquad (\omega.31)$$

The coefficients $D^{(1)}(x, t)$ and $D^{(2)}(x, t)$ are the first two *Kramers-Moyal* coefficients (Kramers, 1940; Moyal, 1949; Risken, 1996). These coefficients are related back to the original SDE (ω.29) by[9]

[9] The factor of 2 in the expression for the diffusion function originates from the Taylor series expansion the Master equation (Risken, 1996). In some references, the noise $\xi(t)$ in (ω.29) is defined to have

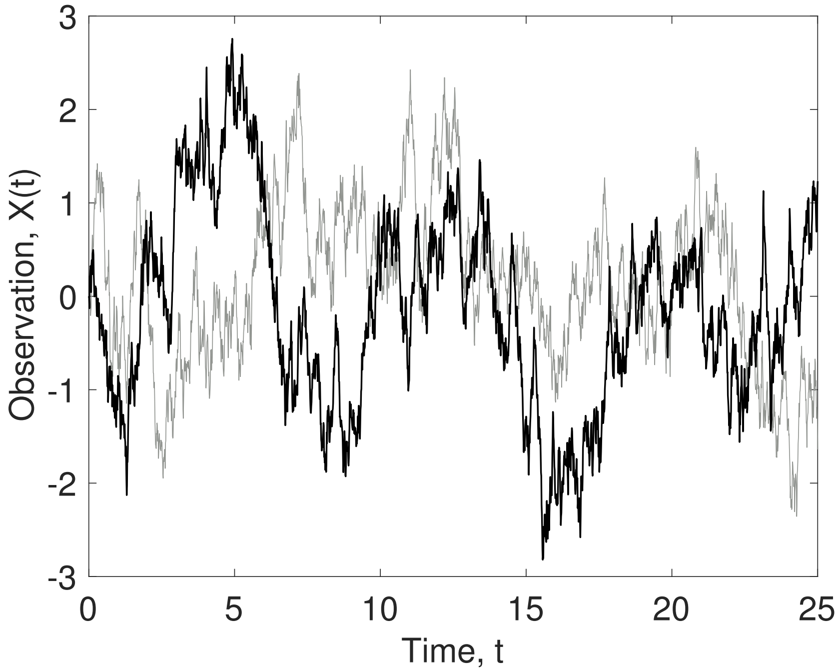

Figure ω.3: A realization of two OU processes, $X(t, \omega)$, indicated in black. This realization uses a sample from probability space $\omega \in \Omega$. Superimposed is another realization, $X(t, \tilde{\omega})$, indicated in gray. This process has the same drift and noise functions and initial condition, but with a different sample of probability space $\tilde{\omega} \in \Omega$.

$$f\big(X(t), t\big) = D^{(1)}(x, t), \tag{ω.32}$$

$$g\big(X(t), t\big) = \sqrt{2D^{(2)}(x, t)}. \tag{ω.33}$$

These Kramers-Moyal coefficients can be accessed by taking moments of transition probability densities (Risken, 1996, e.g.,)

$$D^{(n)}(x, t) = \lim_{\tau \to 0} \frac{1}{n!\tau} M^{(n)}(x, \tau), \qquad n = 1, 2, \tag{ω.34}$$

for, where the quantities

$$M^{(n)}(x, \tau) = \int_{-\infty}^{\infty} \big[x' - x\big]^n p(x', t + \tau | x, t)\; dx', \qquad n = 1, 2, \tag{ω.35}$$

are the conditional moments. Invoking (or assuming) the edgodicity of the process, the integral over possible states x' can be exchanged with a time-average. Thus, integrals (ω.35) can be interpreted as expectations

autocorrelation $\langle \xi(t)\xi(t') \rangle = 2\delta(t - t')$. This convention results in the relation $g\big(X(t), t\big) = \sqrt{D^{(2)}(x, t)}$ for the diffusion function.

$$M^{(n)}(x,\tau) = E\big(X(t+\tau) - X(t)\big|X(t) = x\big). \quad (\omega.36)$$

This expression can be readily calculated numerically, from a sample of $X(t)$, allowing for a path towards fitting drift and noise functions to empirical data. Although the drift and noise functions are being fitted, these functions are frequently referred to as "Kramers-Moyal coefficients". This is because, often, one cannot a priori assume that the process can be exclusively be defined by just the drift and noise functions: a greater number of Kramers-Moyal coefficients may be needed. This approach—coming at the drift and noise functions of a Langevin-like stochastic process from estimates of the conditional moments—is sometimes called "Kramers-Moyal analysis".

Broadly, this methodology can be thought of as an alteration to the naive "guess and check" inverse methodology of Fig. ω.2. A schematic of this is shown in Fig. ω.4. This relies on the hope that both the forward strong solution and the statistical map are in a sense injective, i.e. objects in the destination sets cannot come from many objects in the origin sets.

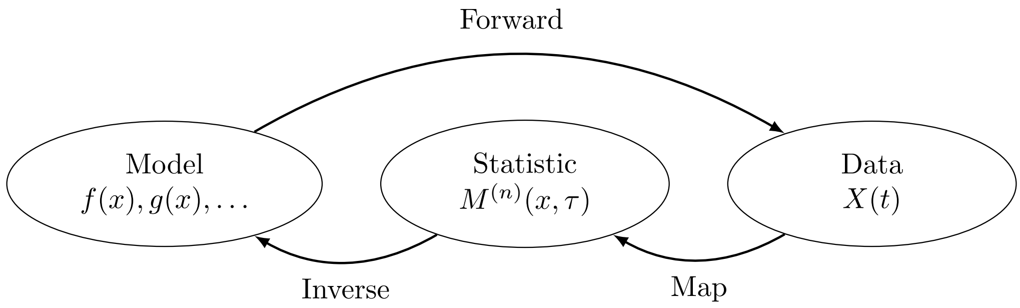

Figure ω.4: A new schematic of stochastic inference, showing the relationship between the model objects, (e.g. drift and noise functions $f(x)$, $g(x)$, etc...), intermediate statistic objects (e.g. conditional moments, $M^{(n)}(x,\tau)$), and data objects (e.g. time-series observations, $X(t)$). The line from model to data represents forward modeling, the line from data to statistics represents functional mapping, and the line from statistics model represents inverse modeling.

ω.5.1 Direct Estimation and Histogram Based Regression

The field of moment-based estimation techniques was started by Siegert, Friedrich, and Peinke (1998). In that study, the authors considered a set of numerical example SDEs of the form (ω.29), that were sampled at a regular and small time-interval. State conditioning in (ω.36) was performed with binning. At a evaluation point with state $x = x_i$ and time-shift $\tau = \tau_j$, (ω.36) is evaluated with

$$\hat{M}^{(n)}(x,\tau) = \frac{\sum_{k=1}^{N} I\big(X(t_k) \in B(x)\big)\big[X(t_k+\tau_j) - X(t_k)\big]^n}{\sum_{k=1}^{T} I\big(X(t_k) \in B(x)\big)}, \qquad (\omega.37)$$

where $I(\circ)$ is the indicator function, and binning is indicated with the half closed interval $B(x) := [x - \frac{1}{2}b_x, x + \frac{1}{2}b_x)$, where x is the desired evaluation point, and b_x is the width of the bin. This approach is similar to the regressogram method of Tukey (1961), and is referred to as histogram based regression. With the state conditioning on the conditional moments set, the limit in (ω.34) is approximated by taking a small τ expansion

$$\hat{D}^{(n)}(x) = \frac{1}{n!\tau} M^{(n)}(x,\tau) + \mathcal{O}(\tau^2), \qquad n = 1, 2. \qquad (\omega.38)$$

This approximation appears to be valid when the timestep τ is much smaller than the characteristic timescale of the system τ_{eff} (Rudolf Friedrich, Joachim Peinke, et al., 2011). From this assumption, the moments are evaluated at the smallest available timestep, $\tau = \Delta t$. As such, the estimates of the Kramers-Moyal coefficients, $\hat{D}^{(n)}(x)$, become

$$\hat{D}^{(n)}(x) \approx \frac{1}{n!\Delta t} M^{(n)}(x, \Delta t), \qquad n = 1, 2. \qquad (\omega.39)$$

This approach has been called “direct estimation”. We will use this approach in `MATLAB`.

ω.6 MATLAB examples

Exercise 6.1: Direct estimation, OU (`example6_1.m`)

Use your OU data from Example 3.1, and estimate $\hat{D}^{(n)}(x)$ at $x = -0.5$. Repeat this for a range of points in $x \in [-1, 1]$. The result should look like Fig. ω.5. Try changing Δt in the integration. How does this alter the drift and noise functions

Exercise 6.2: Direct estimation, other SDEs

Change the drift and noise functions in your Euler-Maruyama simulation to something from Example 3.2, and create a new time-series dataset. Now repeat the estimation in Example 6.1.

Exercise 6.3: Direct estimation, empirical data (`example6_3.m`)

I have included a file, `example_data.mat`, which contains an empirical data-set. Use direct estimation to estimate what likely drift and noise functions generated this data.

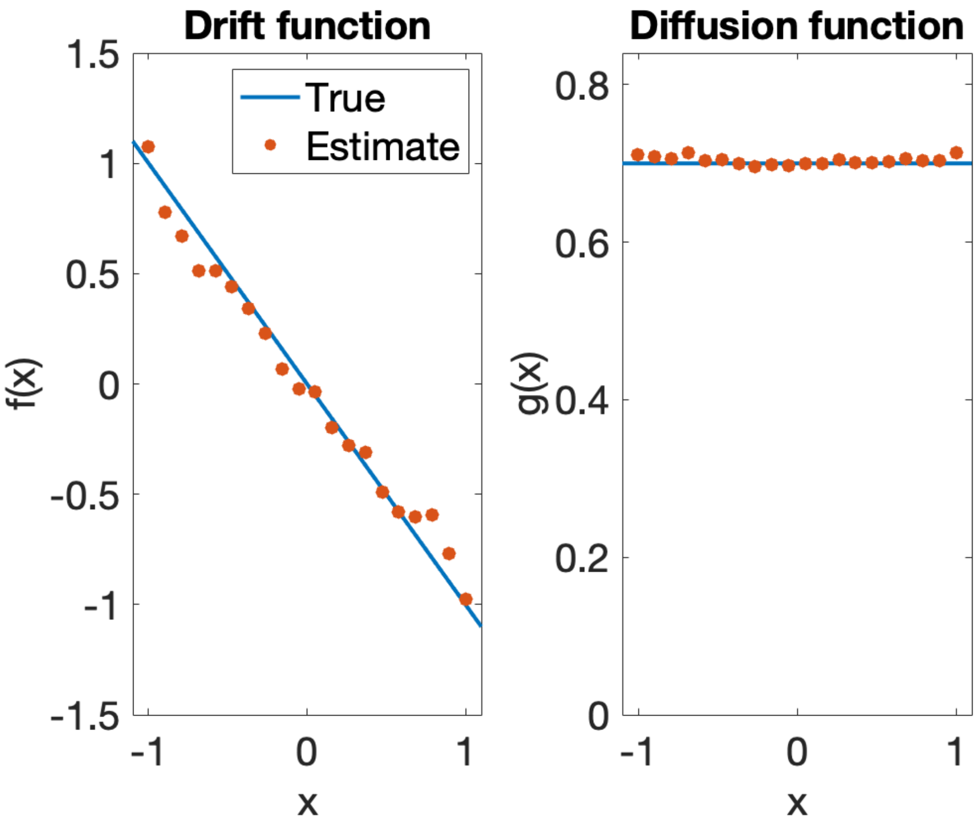

Figure ω.5: Drift and noise functions estimated in Example 6.1.

ω.7 Further Topics

Considerations of Direct Estimation

There are many considerations to make when using a direct numerical approach in moments-based estimation. These can be broadly separated into philosophical and implementational details. Philosophical details largely focus on whether the model (ω.29) is an appropriate model for the observed empirical data, $X(t)$. To list a few:

1. Is the assumption of ergodicity valid?

2. Is the observed system described by the observable dimensions, or may there be hidden slow variables that contribute to the dynamics of the system?

3. Similarly to the last point, is the system well described by a first order SDE? Would an SDE of a higher order (and therefore higher dimension) be a more appropriate model?

4. Is the system driven by white noise, or is a type of correlated noise present?

5. Can the observed empirical data be well explained by two[10] Kramers-Moyal coefficients? This implies that noise that drives the system is Gaussian, and therefore the process has no discontinuous jumps.

Independent checks may be required to test the validity of these assumptions (e.g., Kleinhans et al., 2007). For the implementation details, a number of considerations arise:

1. How is the conditional expectation in (ω.36) evaluated?

2. How is the limit $\tau \to 0$ in (ω.34) performed?

3. Is the sampling of $X(t)$ regular? I.e. $\Delta t = \Delta t_i = t_{i+1} - t_i$. If not, then it is not clear how to evaluate (ω.36). This can be thought of as conditioning in τ, where a regular sampling equates to "index-based conditioning".

When applying moment based fitting to any empirical data, a choice must be made for each of these options.

Multiple τ evaluations

Alternate methodologies can be incorporated into the direct estimation scheme. To attempt to evaluate the limit in (ω.34), the moments $\hat{M}^{(n)}(x,\tau)$ can be sampled at many τ values. An example of these moments are shown in Fig. ω.6. By using many τ samples, more robust estimates of the moments in the limit $\tau \to 0$ can be approached by considering a range of τ evaluation points, $\mathcal{T} = \{\Delta t, 2\Delta t, \ldots, \tau_{\max}\}$, and performing linear regression (e.g., Rudolf Friedrich, Siegert, et al., 2000; Gottschall and Joachim Peinke, 2008). For example for the first moment, scaled moments can be separated into a slope and intercept

$$\frac{\hat{M}^{(1)}(x_i,\tau_j)}{\tau_j} = A_i \tau_j + B_i. \qquad (\omega.40)$$

In this case, the intercepts B_i at spatial evaluation points x_i are directly the best estimates scaled moments at $\tau = 0$.

[10] Note that in this context, the Pawlua theorem states that the number of Kramers-Moyal coefficients to describe a system must be either one, two, or infinite (RF Pawula, 1967; Pawula, 1967).

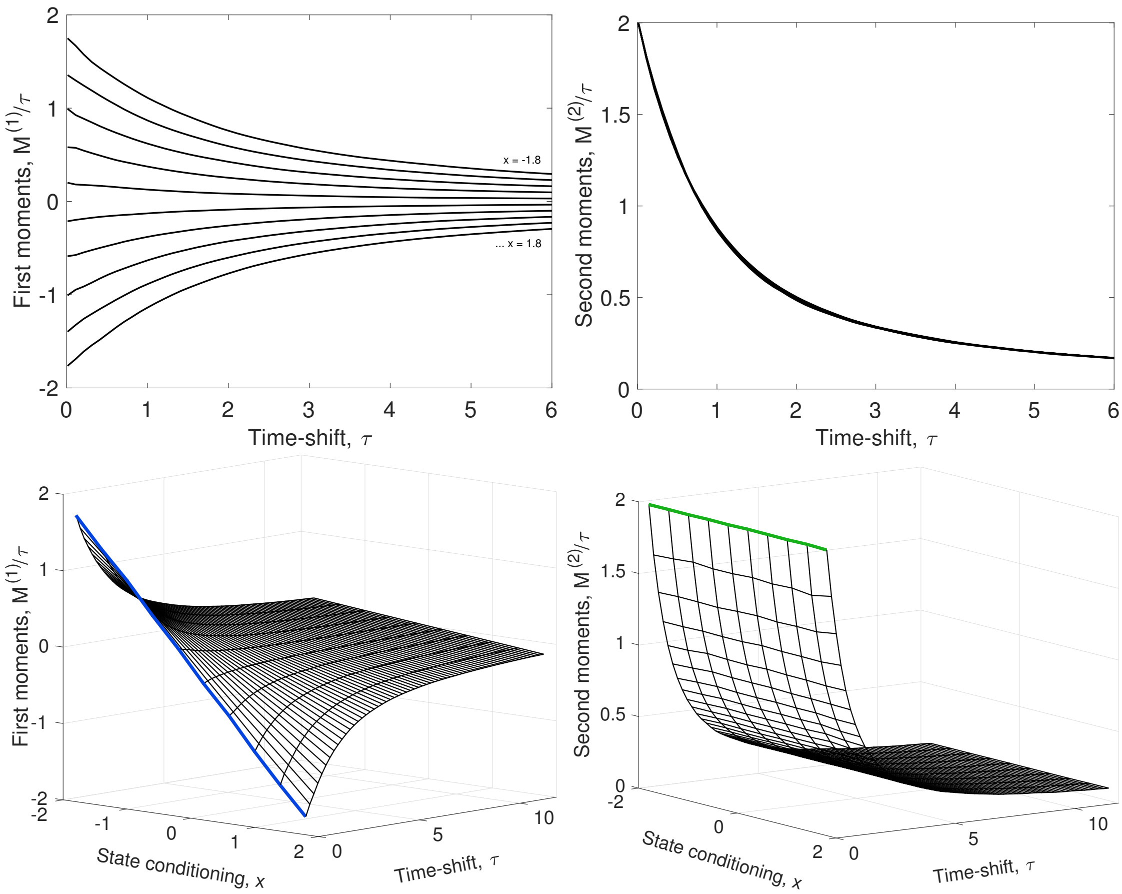

Figure ω.6: Conditional moments scaled by τ, i.e., $\hat{M}^{(n)}/\tau$. The labels in the top left plot that indicate the x conditioning is only shown for the first and last lines. The theoretical value of the moments in the limit of $\tau \to 0$, predicted by (ω.34), are shown as blue and green lines in the bottom left and bottom right plots, respectively.

Measurement Errors

Another point to consider, related to the appropriateness of the SDE model (ω.29), concerns the presence of measurement noise (or measurement errors). Consider a general SDE $Y(t)$ that is the combination of some Langevin-type process and measurement noise

$$\frac{dX}{dt} = f\big(X(t), t\big) + g\big(X(t), t\big)\xi(t), \qquad (\omega.41)$$

$$Y(t) = X(t) + \sigma\zeta(t), \qquad (\omega.42)$$

where $\zeta(t)$ is the measurement noise. To estimate Kramers-Moyal coefficients in the presence of measurement noise, Böttcher et al. (2006) introduced a method to parametrically fit drift and diffusion functions as well as the amplitude of the measurement noise. This method involved estimation and optimisation of the slopes and intercepts of moments at low τ shifts, similarly to (ω.40). This approach has been expanded in subsequent studies (e.g., Lind et al., 2010), as well as to non-white measurement noise $\zeta(t)$ (Bernd Lehle, 2011; Scholz et al., 2017), and in multiple dimensions (Bernd Lehle, 2013).

Correlated Internal Noise

A fundamental assumption of most Kramers-Moyal analysis is that the internal noise that drives fast scale variations is uncorrelated Gaussian white noise. Consider, instead of (ω.29), a more general process

$$\frac{dX}{dt} = f\big(X(t), t\big) + g\big(X(t), t\big)\eta(t), \tag{ω.43}$$

where $\eta(t)$ is correlated noise, $\langle \eta(t)\eta(t')\rangle \neq \delta(t-t')$.

Consider a simple example of a non-white process: exponentially correlated Gaussian white noise

$$\frac{d\eta}{dt} = -\frac{1}{\theta}\eta + \frac{1}{\theta}\xi(t). \tag{ω.44}$$

where θ is the characteristic timescale of the driving noise, and $\xi(t)$ is the Gaussian white internal noise. This noise process still has a vanishing mean, but has an autocorrelation of

$$\langle \eta(t)\eta(t')\rangle = \frac{1}{2\theta}e^{-|t-t'|/\theta}. \tag{ω.45}$$

The exponent in this equation indicates that (ω.44) has a finite correlation time of θ. In this case, the traditional direct-estimation method will not work.[11] Complicated estimation methods have been developed to solve this problem Lehle and Peinke (2018).[12]

References

Bishwal, Jaya PN (2007). *Parameter estimation in stochastic differential equations.* Springer.

Böttcher, Frank et al. (2006). "Reconstruction of complex dynamical systems affected by strong measurement noise". In: *Physical Review Letters* 97.9, p. 090603.

Butcher, John Charles (2016). *Numerical methods for ordinary differential equations.* John Wiley & Sons.

Evans, Lawrence C (2012). *An introduction to stochastic differential equations.* Vol. 82. American Mathematical Soc.

Friedrich, Rudolf, Joachim Peinke, et al. (2011). "Approaching complexity by stochastic methods: From biological systems to turbulence". In: *Physics Reports* 506.5, pp. 87–162.

[11] Why do you think it will fail?

[12] I have `MATLAB` and `Julia` packages implementing these methods: `github.com/williamjsdavis/ExponentialNoiseSDE`, `github.com/williamjsdavis/secn-PR`.

Friedrich, Rudolf, Silke Siegert, et al. (2000). "Extracting model equations from experimental data". In: *Physics Letters A* 271.3, pp. 217–222.

Gardiner, Crispin W et al. (1985). *Handbook of stochastic methods.* Vol. 3. springer Berlin.

Gottschall, Julia and Joachim Peinke (2008). "On the definition and handling of different drift and diffusion estimates". In: *New Journal of Physics* 10.8, p. 083034.

Itô, Kiyosi (1944). "Stochastic integral". In: *Proceedings of the Imperial Academy* 20.8, pp. 519–524.

— (1951). *On stochastic differential equations.* 4. American Mathematical Soc.

Kleinhans, David et al. (2007). "Markov properties in presence of measurement noise". In: *Physical Review E* 76.4, p. 041109.

Kloeden, Peter E and Eckhard Platen (2013). *Numerical solution of stochastic differential equations.* Vol. 23. Springer Science & Business Media.

Kramers, Hendrik Anthony (1940). "Brownian motion in a field of force and the diffusion model of chemical reactions". In: *Physica* 7.4, pp. 284–304.

Langevin, Paul (1908). "Sur la théorie du mouvement brownien". In: *Compt. Rendus* 146, pp. 530–533.

Lehle, B and J Peinke (2018). "Analyzing a stochastic process driven by Ornstein-Uhlenbeck noise". In: *Physical Review E* 97.1, p. 012113.

Lehle, Bernd (2011). "Analysis of stochastic time series in the presence of strong measurement noise". In: *Physical Review E* 83.2, p. 021113.

— (2013). "Stochastic time series with strong, correlated measurement noise: Markov analysis in N dimensions". In: *Journal of Statistical Physics* 152.6, pp. 1145–1169.

Lind, Pedro G et al. (2010). "Extracting strong measurement noise from stochastic time series: applications to empirical data". In: *Physical Review E* 81.4, p. 041125.

Mannella, Riccardo and Peter VE McClintock (2012). "Itô versus Stratonovich: 30 years later". In: *Fluctuation and Noise Letters* 11.1.

Maruyama, Gisiro (1954). "On the transition probability functions of the Markov process". In: *Nat. Sci. Rep. Ochanomizu Univ* 5, pp. 10–20.

Moyal, JE34975 (1949). "Stochastic processes and statistical physics". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 11.2, pp. 150–210.

Pavliotis, Grigorios A (2014). *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations.* Vol. 60. Springer.

Pawula, R (1967). "Generalizations and extensions of the Fokker-Planck-Kolmogorov equations". In: *IEEE Transactions on Information Theory* 13.1, pp. 33–41.

Pawula, RF (1967). "Approximation of the linear Boltzmann equation by the Fokker-Planck equation". In: *Physical review* 162.1, p. 186.

Risken, Hannes (1996). *The Fokker-Planck equation.* Springer, pp. 63–95.

Scholz, Teresa et al. (2017). "Parameter-free resolution of the superposition of stochastic signals". In: *Physics Letters A* 381.4, pp. 194–206.

Siegert, Silke, R Friedrich, and J Peinke (1998). "Analysis of data sets of stochastic systems". In: *Physics Letters A* 243.5-6, pp. 275–280.

Sørensen, Michael (2012). "Estimating functions for diffusion-type processes". In: *Statistical methods for stochastic differential equations* 124, pp. 1–107.

Stratonovich, RL196814 (1966). "A new representation for stochastic integrals and equations". In: *SIAM Journal on Control* 4.2, pp. 362–371.

Stratonovich, Rouslan L (1967). *Topics in the theory of random noise.* Vol. 2. CRC Press.

Strogatz, Steven H (2018). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* CRC press.

Tukey, John W (1961). "Curves as parameters, and touch estimation". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics.* Vol. 4. University of California Press, pp. 681–695.

Van Kampen, Nicolaas G (1981). "Itô versus stratonovich". In: *Journal of Statistical Physics* 24.1, pp. 175–187.

Van Kampen, Nicolaas Godfried (1992). *Stochastic processes in physics and chemistry.* Vol. 1. Elsevier.