Learning Stochastic Differential Equations with Bridge Sampling

Anonymous Author(s)
Affiliation
Address
email

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

Points to cover:

- data specification
- Hermite polynomial and drift function representation
- Expectation and maximization formulas assuming data is filled in
- Filling data in with Brownian bridge
- MCMC iterations of brownian bridge using girsanov likelihood
- how synthetic data is generated
- results: 1D, 2D, 3D damped duffing, 3D lorenz
- plots: error of theta vs noise, error vs amount of data (number of data points) parametric curves for noise levels, brownian bridge plots for illustration, ...
- Note: constant noise case, not inferring the gvec

1 Model Setup

Our goal here is to start with $\mathbb{R}^d$ dimensional time series data and a non-parametric model to infer the coefficients of the Hermite basis functions which would result in a parametric equation. The general form of the governing equation is:

$$\mathrm{d}X(t) = f(X(t))\,\mathrm{d}t + \Gamma\,\mathrm{d}W(t) \tag{1}$$

where Γ is a constant diagonal matrix of noise levels. For these experiments, we assume that the noise levels are known beforehand and are not inferred.

To model the general form of the drift function, we represent the drift function as an additive function of the Hermite basis polynomials of maximum degree D. The ζ parameters will be inferred. Through the combination of inferred parameters and the Hermite polynomial basis, the governing equation can be discovered.

$$f(x) = \sum_{i=0}^{M} \zeta_i\,\psi_i(x) \tag{2}$$

The orthonormal Hermite basis functions have the general form

$$H_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2} \tag{3}$$

which results in the following set of polynomials:

$$\begin{aligned}
H_0(x) &= 1/c_0, & c_0 &= \sqrt{2\pi\sqrt{0!}} \\
H_1(x) &= x/c_1, & c_1 &= \sqrt{2\pi\sqrt{1!}} \\
H_2(x) &= (x^2-1)/c_2, & c_2 &= \sqrt{2\pi\sqrt{2!}} \\
H_3(x) &= (x^3-3x)/c_3, & c_3 &= \sqrt{2\pi\sqrt{3!}}
\end{aligned}$$

We demonstrate the method for 1, 2 and 3 dimensional systems.

- For the 1-dimensional system, we use the ? oscillator:

$$\mathrm{d}X(t) = (\alpha X(t) + \beta X(t)^2 + \gamma)\,\mathrm{d}t + g\,\mathrm{d}W(t) \tag{4}$$

- For the 2-dimensional system, we use the undamped Duffing oscillator:

$$\begin{aligned}
\mathrm{d}X_1(t) &= X_2(t)\mathrm{d}t + g_1\,\mathrm{d}W_1(t) \\
\mathrm{d}X_2(t) &= (-X_1(t) - X_1^3(t))\mathrm{d}t + g_2\,\mathrm{d}W_2(t)
\end{aligned}$$

- For the 3-dimensional case, we consider 2 different form of equations. The first one is the damped Duffing oscillator, a general form of the damped oscillator considered in the 2-dimensional case:

$$\begin{aligned}
\mathrm{d}X_1(t) &= X_2(t)\,\mathrm{d}t + g_1\,\mathrm{d}W_1(t) \\
\mathrm{d}X_2(t) &= (\alpha X_1(t) - \beta X_1(t) - \delta X_2(t) + \gamma\cos(X_3(t)))\,\mathrm{d}t + g_2\,\mathrm{d}W_2(t) \\
\mathrm{d}X_3(t) &= \omega\,\mathrm{d}t + g_3\,\mathrm{d}W_3(t)
\end{aligned}$$

- Another example considered for the 3-dimensional case is the Lorenz oscillator:

$$\begin{aligned}
\mathrm{d}X_1(t) &= \sigma(X_2(t) - X_1(t))\,\mathrm{d}t + g_1\,\mathrm{d}W_1(t) \\
\mathrm{d}X_2(t) &= (X_1(t)(\rho - X_3(t)))\mathrm{d}t + g_2\,\mathrm{d}W_2(t) \\
\mathrm{d}X_3(t) &= (X_1(t)X_2(t) - \beta X_3(t))\,\mathrm{d}t + g_3\,\mathrm{d}W_3(t)
\end{aligned}$$

For the general case where $X = (x_1, \cdots, x_d) \in \mathbb{R}^d$ case, the general form of the drift function $\vec{f} = (f_1(x_1, \cdots, x_d), \cdots, f_d(x_1, \cdots, x_d))$ includes cross interactions between the dimensions of X to create an additive function of Hermite polynomial of maximum degree M:

$$f(x_1, \cdots, x_d) = \sum_{m=0}^{M} \sum_{\sum i_m = 0}^{\sum i_m = m} \zeta_{i_m}^d \psi_{i_m}^d \tag{5}$$

For simplicity, consider the example where the $X \in \mathbb{R}^2$ and the highest degree of the Hermite polynomial is three, including four Hermite polynomials:

$$\begin{aligned}
f(x_1, x_2) &= \sum_{m=0}^{2} \sum_{i+j=0}^{i+j=m} \zeta_{i,j}\,\psi_{i,j} \\
&= \sum_{d=0}^{3} \sum_{i+j=0}^{i+j=3} \zeta_{i,j} H_i(x_1) H_j(x_2) \\
&= \sum_{i+j=0} \zeta_{i,j} H_i(x_1) H_j(x_2) + \sum_{i+j=1} \zeta_{i,j} H_i(x_1) H_j(x_2) + \sum_{i+j=2} \zeta_{i,j} H_i(x_1) H_j(x_2) + \sum_{i+j=3} \zeta_{i,j} H_i(x_1) H_j(x_2) \\
&= \zeta_{0,0} H_0(x_1) H_0(x_2) + \zeta_{0,1} H_0(x_1) H_1(x_2) + \zeta_{1,0} H_1(x_1) H_0(x_2) + \zeta_{0,2} H_0(x_1) H_2(x_2) \\
&+ \zeta_{2,0} H_2(x_1) H_0(x_2) + \zeta_{1,1} H_1(x_1) H_1(x_2) + \zeta_{0,3} H_0(x_1) H_3(x_2) + \zeta_{3,0} H_3(x_1) H_0(x_2) \\
&+ \zeta_{2,1} H_2(x_1) H_1(x_2) + \zeta_{1,2} H_1(x_1) H_2(x_2)
\end{aligned}$$

2 Expectation Maximization Steps

The data provided is in the form of a time series, $X \in \mathbb{R}^d$ at regular time points $t_l, 0 \leq l \leq L$. Note: EM step in the sampling writeup.

3 Brownian bridge sampling

When the inter-observation time of the observed data X is large, the expectation step becomes less accurate. To mitigate this problem, we can fill in the observed data with a Brownian bridge. We generate many samples of the $N-$dimensional Brownian bridge and accept-reject samples using the Metropolis-Hastings algorithm. The approximation of the likelihood is obtained using the Girsanov likelihood function.

3.1 Brownian bridge

The $\mathbb{R}^N$ dimensional Brownian bridge is defined by the integral:

$$I(t) = \int_0^t \frac{1-t}{1-T} \, \mathrm{d}W(t) \tag{6}$$

3.2 Metropolis Algorithm