Most probable transition pathway of biology-modeling stochastic differential equations

Stanley Nicholson
Advisor: Dr. Jinqiao Duan

1 Introduction

Stochastics processes govern virtually all physical phenomena such as gene regulation, climate, and particle motion. Stochastic differential equations (SDE) push the boundaries of ordinary differential equations to gain greater insight into the modeling of nature. This paper investigates the problem of extracting a model given noisy model data. Utilizing Dr. Jinqiao Duan's developments in theory such as the Kramers-Moyal formula and "most probable transition pathway," we apply them to a simple biological model related to transcription factor regulation.

2 Methods

The definition of an SDE given a stochastic process X_t, drift function $f(X_t, t)$, and diffusion function $\sigma(X_t, t)$:

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dW_t \tag{2.1}$$

where dW_t is Brownian motion.

Furthermore, similar to ODEs, we have the Euler-Maruyama method to numerically simulate the SDE for

$$X_{n+1} = X_n + f(X_t, t)\Delta t + \sigma(X_t, t)W_{\Delta t} \tag{2.2}$$

where $W_{\Delta t} \sim \mathcal{N}(0, \sqrt{\Delta t})$.

Henceforth we consider the biological model as described by Chen et al. [1] where diffusion function $\sigma(X_t, t) = 0.01$ hereby denoted as ϵ. And the drift function is defined as

$$f(x) = \frac{k_f x^2}{x^2 + K_d} - k_d + R_{bas} \tag{2.3}$$

where $k_f = 6\,\text{min}^{-1}$, $K_d = 10$, $k_d = 1\,\text{min}^{-1}$, and $R_{bas} = 0.4\,\text{min}^{-1}$ as defined by Smolen et al. [3].

When analytically solving these equations, there exist two "stable" points and one "unstable" point which we call $x_+ = 4.28343$, $x_- = 0.62685$, and $x_u = 1.48971$. The nature of these points will become clear when looking at 3.

An important note is that both the drift and diffusion functions are time independent or autonomous. They are functions of the state of the system and not time; however, they nonetheless evolve with time.

As shown by Dai et al. [2] in equations (2) and (3), given a stochastic process X_t and simulated data, we can infer the drift and diffusion equations from the Kramers-Moyal formula,

$$f(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{X_{\Delta t} - X_0}{\Delta t} \middle| X_0 = x \right) \tag{2.4}$$

$$\sigma(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{(X_{\Delta t} - X_0)^2}{\Delta t} \middle| X_0 = x \right) \tag{2.5}$$

Lastly, an important result we utilize is given by Chen et al. [1] is determining the most probable transition pathway as given in equations (7) and (8). That is, the function which represents the most likely path for our model to travel. We combine this function with the "shooting method" to better understand how possible our biological model will jump between the stable points. Namely, given f, we determine the most probable transition pathway z as

$$\ddot{z} = \frac{\epsilon^2}{2} f''(z) + f'(z) f(z). \tag{2.6}$$

with boundaries $z(t_0) = x_-$ and $z(t_f) = x_+$.

3 Results

We began by simulating our data with parameters $t_0 = 0, t_f = 50$, $\Delta t = 0.05$, $N = 1000$ steps over a range of $x_i \in [0, 6]$ with $\Delta x = 0.06$. Using the Euler-Maruyama method as given by equation 2.2 and coloring based on the initial x a sample simulation of our data is in figure 3.

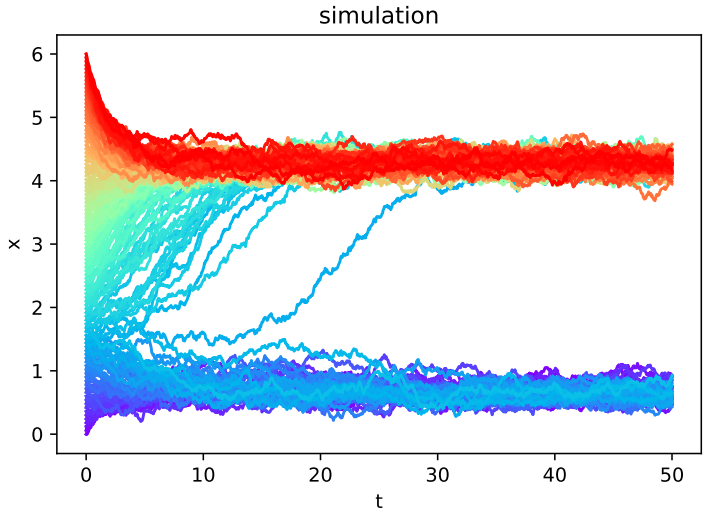

Figure 1: Simulated data given our biological model f and σ using Euler-Maruyama

Now given these data we wish estimate the drift function f to actually determine the model by which our biological system acts. After applying equation 2.4 and 2.5 we arrive at the data as shown in figure 2. Furthermore we now interpolate the data with polynomials and obtain the approximations

$$f(x) \approx p(x) = -0.02101x^3 + 0.1244x^2 - 0.1575x + 0.0909 \tag{3.1}$$

$$\sigma(x) \approx 0.09573 \tag{3.2}$$

for $x \in [0, 6]$.

Given our polynomial approximation of our drift function f which we call p, we now wish to determine the most probable transition pathway. Applying equation [**?**], we derive that

$$\ddot{z} = 0.001324x^5 - 0.01306x^4 + 0.04417x^3 - 0.06448x^2 + 0.04741x - 0.01431 \tag{3.3}$$

with the necessary boundary conditions. We convert out second order ordinary differential equation problem in a single order and solve using the

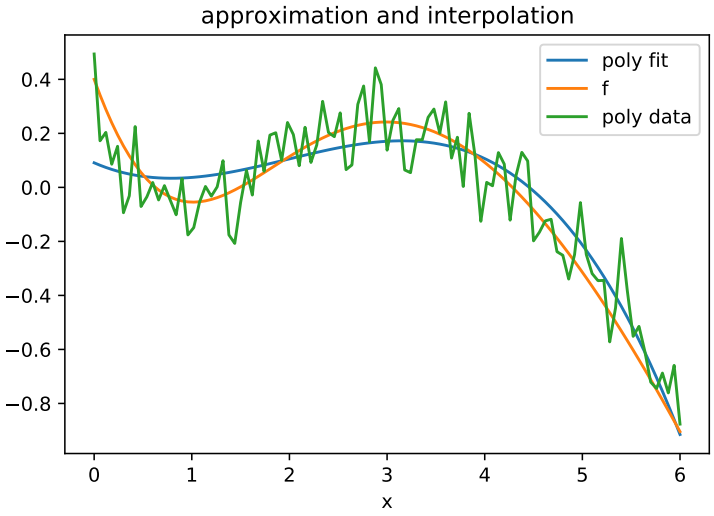

Figure 2: Application of the formula derived by Mai et al. as given in equations 2.4 and 2.5

shooting method in order to solve the boundary conditions. That is, we wish to find the path that would occur if we were to start at the stable concentration of x_- and the necessary "velocity" by which the concentration would need to change to push it up to the other stable concentration of x_+. We find that a least squared loss occurs when the initial velocity $v_0 = 0.029$. Through figure 3, we color based on the initial velocity with the velocity increasing as the color moves closer to red.

4 Discussion

In science, we wish to use deterministic answers because they provide a consistent and understandable method to physical phenomena. Instead of reporting the probability distribution of temperature we give a mean temperature. This analogy similarly applies to the most probable transition pathway. It is the most likely path that will occur if the system were to jump between its two stable points. Nonetheless, understanding the distribution of pathways provides a greater breadth to predicting changes between states of a system. Although we only have explored the one dimensional case

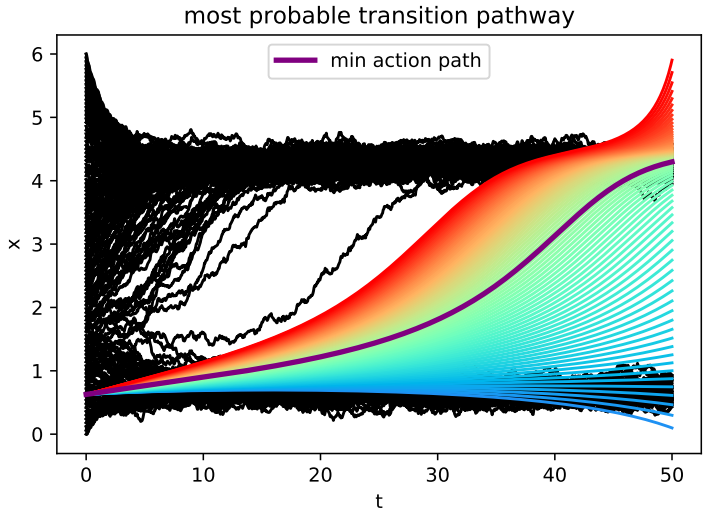

Figure 3: The most probable pathway occurs with the initial concentration of x_- occurs with an initial velocity of $v_0 = 0.029$.

here, virtually all biological systems are high dimensional (consider the various transcription factors, mRNAs, and molecules). This project will begin to work towards higher dimensional cases utilizing machine learning (specifically, normalizing flows) to estimating path distribution.

References

[1] X. Chen, F. Wu, J. Duan, J. Kurths, X. Li. Most probable dynamics of a genetic regulatory network under stable Lévy noise. Applied Mathematics and Computation, 348:425-436, 2018.

[2] M. Dai, J. Duan, J. Liao, X. Wang. Maximum Likelihood Estimation of Stochastic Differential Equations with Random Effects Driven by Fractional Brownian Motion. arXiv, 2001.01412, 2020.

[3] P. Smolen, D. A. Baxter, and J. H. Byrne. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. American Journal of Physiology-Cell Physiology, 274:531-542, 1998.

[4] J. Duan, T. Gao, G. He. Quantifying Model Uncertainties in the Space of Probability Measures. arXiv preprint arXiv:1204.0855v1, (2012).