Most probable transition pathway of biological SDEs

Stanley Nicholson

Illinois Institute of Technology

π day, 2021

Why Stochastics?

- All physical systems are "noisy"

Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion

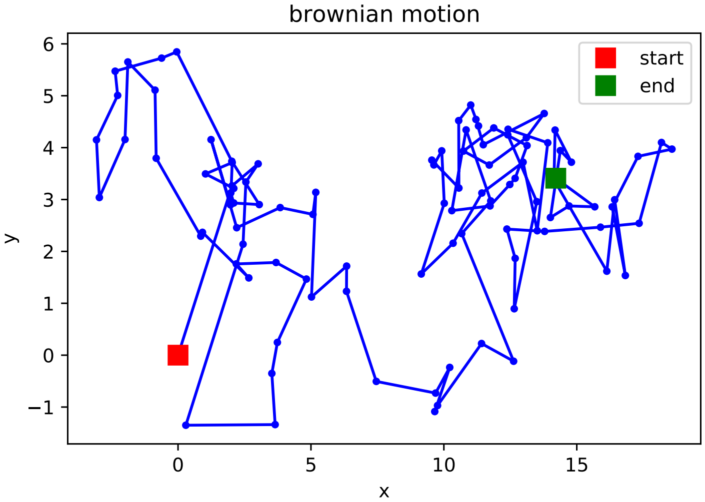
Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion
- Accounts for sudden changes in system

Why Stochastics?

- All physical systems are "noisy"
- Noise: think of Brownian motion
- Accounts for sudden changes in system
- Many biological processes are stochastic: mutation, gene regulation, etc.

Brownian Motion



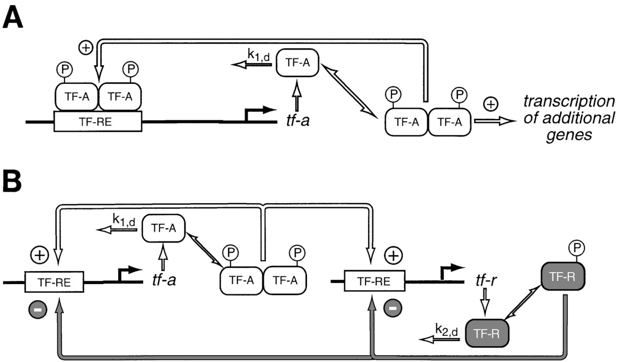
What's an SDE?

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

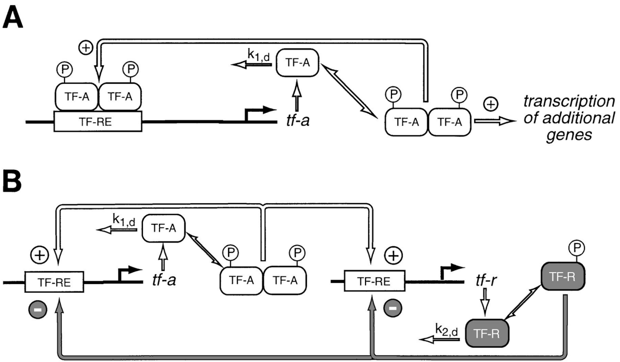
$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t \tag{1}$$

Think: f determines direction, σ denotes noise intensity

The Biology (Smolen et al.)

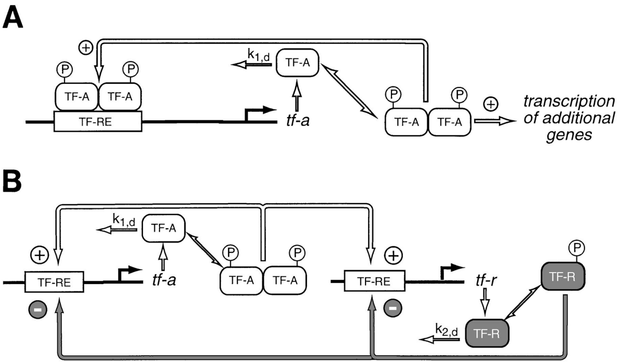


The Biology (Smolen et al.)



$$\frac{d[\text{TF A}]}{dt} = \frac{k_{1,f}[\text{TF A}]^2}{[\text{TF A}]^2 + K_{1,d}} - k_{1,d}[\text{TF A}] + r_{1,\text{bas}} \qquad (2)$$

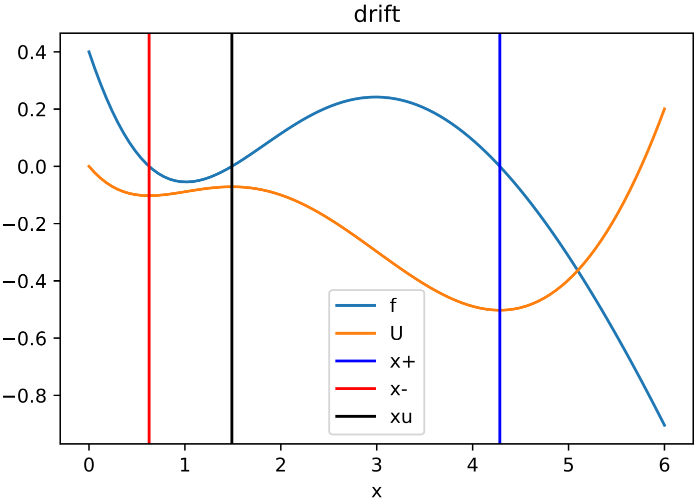
The Biology (Smolen et al.)



$$\frac{d[\text{TF A}]}{dt} = \frac{k_{1,f}[\text{TF A}]^2}{[\text{TF A}]^2 + K_{1,d}} - k_{1,d}[\text{TF A}] + r_{1,\text{bas}} \qquad (2)$$

Let this be our f and we use constant noise $\sigma = 0.10$.

Our Model

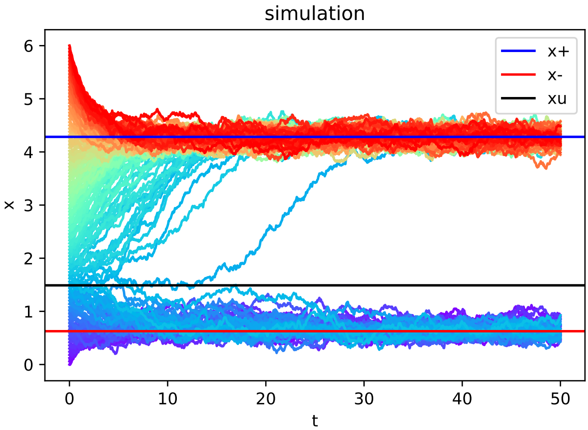


Simulation

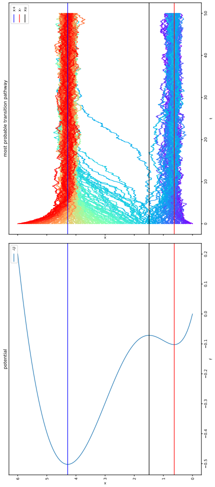
Given f and σ, we simulate the SDE with the Euler-Maruyama method.

Simulation

Given f and σ, we simulate the SDE with the Euler-Maruyama method.



Model against Potential



Learn our Model

Given this simulated data, we wish to find learn the underlying drift and diffusion

Learn our Model

Given this simulated data, we wish to find learn the underlying drift and diffusion

$$f(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{X_{\Delta t} - X_0}{\Delta t} \middle| X_0 = x \right) \tag{3}$$

$$\sigma^2(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{(X_{\Delta t} - X_0)^2}{\Delta t} \middle| X_0 = x \right) \tag{4}$$

Learn our Model

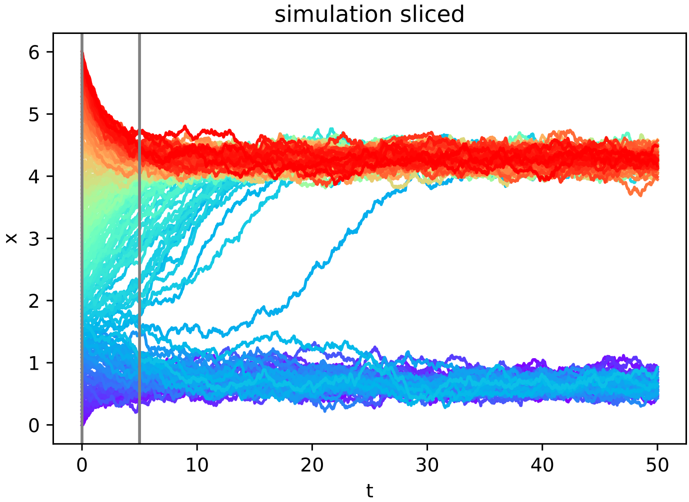
Given this simulated data, we wish to find learn the underlying drift and diffusion

$$f(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{X_{\Delta t} - X_0}{\Delta t} \middle| X_0 = x \right) \tag{3}$$

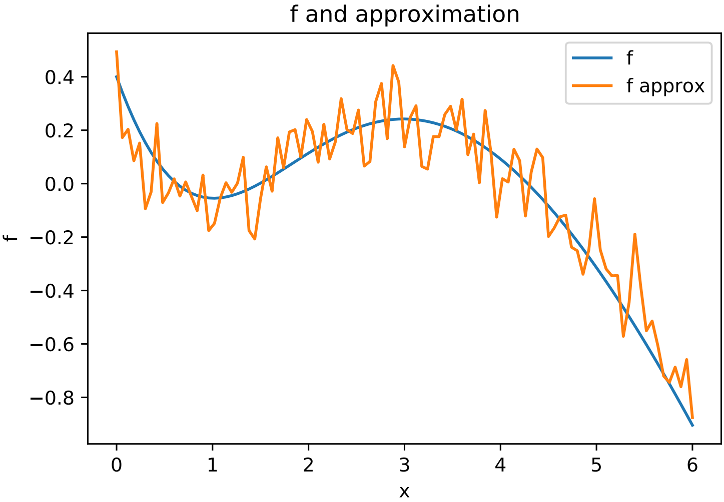
$$\sigma^2(x) = \lim_{\Delta t \to 0} \mathbb{E}\left(\frac{(X_{\Delta t} - X_0)^2}{\Delta t} \middle| X_0 = x \right) \tag{4}$$

found by Dai et al. (2020). Proof involves stochastic version of Taylor expansion.

Simulation Zoom-In



Learned f



Most Probable Pathway (Chen et al.)

Starting at the stable concentration x_- of transcription factor, can we jump to x_+?

Most Probable Pathway (Chen et al.)

Starting at the stable concentration x_- of transcription factor, can we jump to x_+?
What is the most probable path that will be taken?

Most Probable Pathway (Chen et al.)

Starting at the stable concentration x_- of transcription factor, can we jump to x_+?
What is the most probable path that will be taken?

$$OM(\dot{z}, z) = \left(\frac{f(z) - \dot{z}}{\epsilon}\right) + f'(z). \tag{5}$$

This must satisfy the Euler-Lagrange equation, giving us that our $\ddot{z}$ satisfies the differential equation

Most Probable Pathway (Chen et al.)

Starting at the stable concentration x_- of transcription factor, can we jump to x_+?
What is the most probable path that will be taken?

$$OM(\dot{z}, z) = \left(\frac{f(z) - \dot{z}}{\epsilon}\right) + f'(z). \tag{5}$$

This must satisfy the Euler-Lagrange equation, giving us that our $\ddot{z}$ satisfies the differential equation

$$\ddot{z} = \frac{\epsilon^2}{2} f''(z) + f'(z) f(z). \tag{6}$$

with the condition that $z(t_0) = x_-$ and $z(t_f) = x_+$.

Most Probable Pathway (Chen et al.)

Starting at the stable concentration x_- of transcription factor, can we jump to x_+?
What is the most probable path that will be taken?

$$OM(\dot{z}, z) = \left(\frac{f(z) - \dot{z}}{\epsilon} \right) + f'(z). \tag{5}$$

This must satisfy the Euler-Lagrange equation, giving us that our $\ddot{z}$ satisfies the differential equation

$$\ddot{z} = \frac{\epsilon^2}{2} f''(z) + f'(z) f(z). \tag{6}$$

with the condition that $z(t_0) = x_-$ and $z(t_f) = x_+$.

- Solving analytically is very difficult
- The shooting method lets us find the "velocity" that minimizes the "loss"

The Shooting Method

We rewrite our second order ordinary differential equation into a first order:

$$\begin{cases} v = \dot{z} \\ \dot{v} = \frac{\epsilon^2}{2} f''(v) + f'(v) f(v) \end{cases} \tag{7}$$

The Shooting Method

We rewrite our second order ordinary differential equation into a first order:

$$\begin{cases} v = \dot{z} \\ \dot{v} = \frac{\epsilon^2}{2} f''(v) + f'(v) f(v) \end{cases} \tag{7}$$

Now we have an initial value problem where we are trying to find the root of $z(t_f) - x_+$ given the initial conditions $z(t_0) = x_-$ and $\dot{v}(t_0) = v_0$ (for some v_0 we vary).

The Shooting Method

We rewrite our second order ordinary differential equation into a first order:

$$\begin{cases} v = \dot{z} \\ \dot{v} = \frac{\epsilon^2}{2} f''(v) + f'(v) f(v) \end{cases} \tag{7}$$

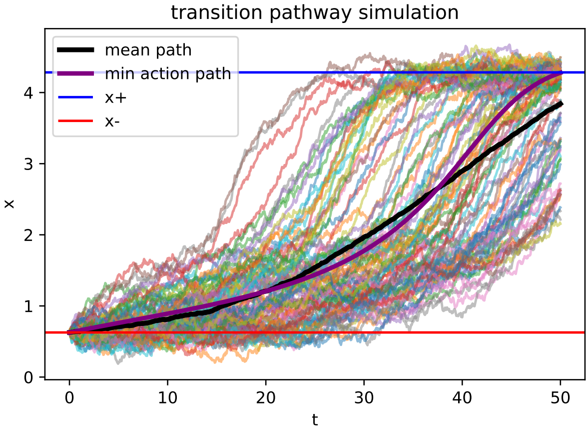
Now we have an initial value problem where we are trying to find the root of $z(t_f) - x_+$ given the initial conditions $z(t_0) = x_-$ and $\dot{v}(t_0) = v_0$ (for some v_0 we vary). We find $v_0 = 0.0285$.

Mean Path vs Learned Path

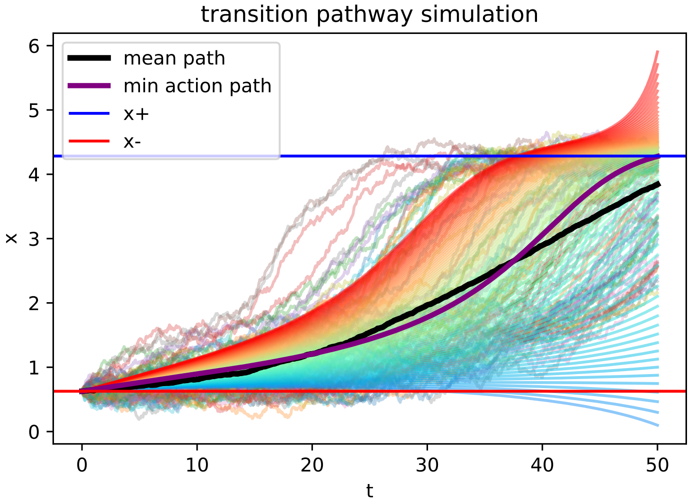
Running the simulation over $50,000$ times we obtain 93 runs that jump from x_- to x_+. That is only 0.186% of the runs.

Mean Path vs Learned Path

Running the simulation over $50,000$ times we obtain 93 runs that jump from x_- to x_+. That is only 0.186% of the runs.



Learned Probable Pathway



Future Work

- Fine tune/optimize specific data extraction and fitting
- Other stochastic information: mean exit time
- Consider higher dimensional data: consider the second transcription factor TF-R
- Applying machine learning techniques to estimating pathway distribution (needed for higher dimensions)
- Applications outside of biology

Bibliography

- 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.
- 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.
- 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.