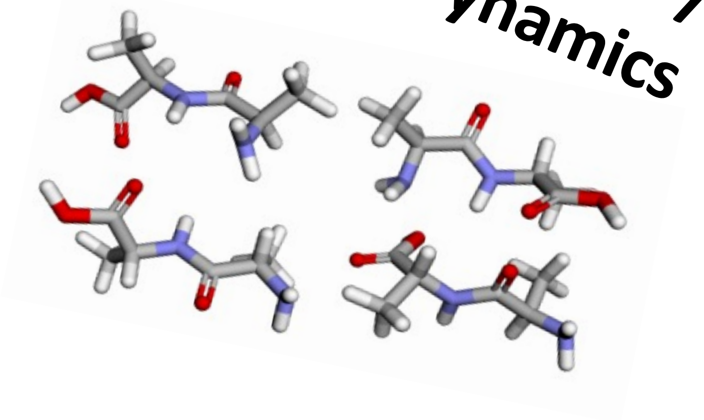
Diffusion Models, SDEs and Path Based Inference

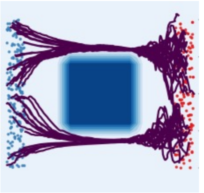
Francisco Vargas

Generative Modelling

Bayesian Inference / Molecular Dynamics



Filtering / Data Assimilation



Diffusion Models and SDEs

Lecture 1:

A very fast paced introduction to the foundations / notation.

Quick Probability Recap

Probability Space

- Sample Space

e.g $\Omega = \{0,1\}$ or $\Omega = \mathbb{R}$

$$\mathbb{P}(\Omega) = 1, \quad P(A) \geq 0$$

$$\mathbb{P}(\cup_{i\in\mathcal{I}} A_i) = \sum_{i\in\mathcal{I}} \mathbb{P}(A_i)$$

$$A_i \cap A_j = \emptyset, i \neq j \ , \quad \exists f : \mathcal{I} \longleftrightarrow \mathbb{N}$$

- Probability Measure

$$(\Omega, \Sigma, \mathbb{P})$$

- Event Space e.g $2^{\{0,1\}}$

/ Sigma Algebra: is a algebra/system of sets that are "closed" under countable # of operations $\cup, \cap, \setminus\Omega$ and $\Omega, \emptyset \in \Sigma \subseteq 2^\Omega$

Quick Probability Recap

Probability Space

- Sample Space

e.g $\Omega = \{0,1\}$ or $\Omega = \mathbb{R}$

$$\mathbb{P}(\Omega) = 1, \quad P(A) \geq 0$$

$$\mathbb{P}(\cup_{i \in \mathcal{I}} A_i) = \sum_{i \in \mathcal{I}} \mathbb{P}(A_i)$$

$$A_i \cap A_j = \emptyset, i \neq j \;, \;\; \exists f : \mathcal{I} \longleftrightarrow \mathbb{N}$$

- Probability Measure

$$(\Omega, \mathcal{B}(\Omega), \mathbb{P})$$

- Event Space e.g $2^{\{0,1\}}$

The Borel-sigma algebra is the smallest sigma algebra containing the event space (i.e. intersect all possible sigma algebra containing Omega).

Quick Probability Recap

Filtered Probability Space

- Think of a filtration as the sample space of a time series, that is a series of sample spaces:

$$\mathcal{F} = \{\mathcal{F}_t\}_{t \in [0,T]}$$

$$s \leq t \implies \mathcal{F}_s \subseteq \mathcal{F}_t$$

$$(\Omega, \mathcal{B}(\Omega), \mathcal{F}, \mathbb{P})$$

Quick Probability Recap

Stochastic Process

- Collection of Random Variables (Measurable Maps) !

$$\{X_t\}_{t\in[0,T]} \qquad X_t(\omega) : [0,T]\times\Omega \to \mathbb{R}^d$$

$$\left(C([0,T];\mathbb{R}^d), \mathcal{B}\big(C([0,T];\mathbb{R}^d)\big), \mathcal{F}, \mathbb{P}\right)$$

Quick Probability Recap

Brownian Motion

- Brownian motion is a Gaussian Process, and one of the simplest Stochastic Processes:
  - Pinned Origin: $W_0 = 0$
  - Independent increments $s, t > 0,\ W_{t+s} - W_t \perp\!\!\!\perp W_t$
  - $W_{t+s} - W_t \sim \mathcal{N}(0, s)$
  - W_t is continuous in t (almost surely)

$$W \sim \mathcal{GP}(0, \min(s, t))$$

Quick Probability Recap

Lebesgue Integral

$$\int_A \mathrm{d}\lambda = \lambda(A)$$

$$\int_\Omega \mathbb{I}_A(x)\mathrm{d}\lambda = \lambda(A)$$

$$\int_\Omega \sum_{i=1}^{n} a_i \mathbb{I}_{A_i}(x)\mathrm{d}\lambda = \sum_{i=1}^{n} a_i \lambda(A_i)$$

$$\int_A f\mathrm{d}\lambda = \sup\left\{\int s\mathrm{d}\lambda : 0 \leq s \leq f, s = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(x)\right\}$$

Quick Probability Recap

Lebesgue Measure

$$\lambda([a,b]) = |a-b|$$

$$\lambda\left(\cup_{i\in\mathcal{I}}[a_i,b_i]\right) = \sum_{i\in\mathcal{I}} \lambda\left([a_i,b_i]\right)$$

$$\lambda(A) = \inf\left\{\lambda(I) : A \subseteq I, I = \cup_{i\in\mathcal{I}}[a_i,b_i]\right\}$$

Volume/Size

- Caratheodory Extension Theorem and Criterion assert uniqueness/existence of the space

Quick Probability Recap

Lebesgue Integral

$$\int_A \mathrm{d}\lambda = \lambda(A)$$

$$\int_\Omega \mathbb{I}_A(x)\mathrm{d}\lambda = \lambda(A)$$

$$\int_\Omega \sum_{i=1}^{n} a_i \mathbb{I}_{A_i}(x)\mathrm{d}\lambda = \sum_{i=1}^{n} a_i \lambda(A_i)$$

$$\int_A f\mathrm{d}\lambda = \sup\left\{\int s\mathrm{d}\lambda : 0 \leq s \leq f, s = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(x)\right\}$$

Quick Probability Recap

Lebesgue-Stjelties Integral

$$\int_A f \mathrm{d}\lambda = \sup \left\{ \int s \mathrm{d}\lambda : 0 \leq s \leq f, s = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(x) \right\}$$

$$\int_A f(x) \mathrm{d}\lambda(x) = \int_A f(x) \mathrm{d}x = \int_A f(x) \lambda(\mathrm{d}x)$$

We can replace lambda with a probability distribution/measure yielding the familiar expectation:

$$\int_A f(x) \mathrm{d}P(x) = \mathbb{E}_P[f(X)]$$

Quick Probability Recap

Lebesgue Integral Matches Traditional Riemann Integral

Why bother with this integral formalism, isn’t Reimann enough ? Many useful theorems come for free, in particular Dominated Convergence (for integrable g):

$$f_n \overset{\text{pointwise}}{\longrightarrow} f \quad \text{and} \quad |f_n| \leq g(x)$$

$$\Downarrow$$

$$\lim_{n\to\infty} \mathbb{E}_P[f_n(X)] = \mathbb{E}_P[f(X)]$$

Quick Probability Recap

Lebesgue Integral – Exercise (Uniform Distribution)

$$P([a, b]) = |a - b|$$

$$P(\Omega) = P([0, 1]) = 1$$

$$\int_{[1/4,1/2]} \mathrm{d}P = ?$$

$$\int_{[0,1]} \mathbb{I}_{[1/e,1/(e+1)]}(x)\mathrm{d}P = ?$$

$$\int_{\Omega} x\mathrm{d}P = ?$$

Quick Probability Recap

Lebesgue Integral – Exercise (Uniform Distribution)

$$P([a, b]) = |a - b|$$

$$P(\Omega) = P([0, 1]) = 1$$

$$\int_{[1/4,1/2]} \mathrm{d}P = 1/2$$

$$\int_{[0,1]} \mathbb{I}_{[1/e,1/(e+1)]}(x)\mathrm{d}P = 1/(e(e+1))$$

$$\int_{\Omega} x\mathrm{d}P = 1/2$$

Quick Probability Recap

Radon Nikodym Theorem – Change of Measure

$$\mu << \lambda := \lambda(A) = 0 \implies \mu(A) = 0$$

$$\mu(A) = \int_A \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x)$$

$$\int_A f(x)\mathrm{d}\mu(x) = \int_A f(x)\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x)$$

Quick Probability Recap

Radon Nikodym Theorem – Probaility Density Function

$$\mathbb{P} << \lambda$$

$$\mathbb{P}(A) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x)$$

Now For sake of simplicity assume Reimann Integrability

$$\mathbb{P}(A) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}x$$

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x) = ??$$

Quick Probability Recap

Radon Nikodym Theorem – Probaility Density Function

$$\mathbb{P} << \lambda$$

$$\mathbb{P}(A) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x)$$

Now For sake of simplicity assume Reimann Integrability

$$\mathbb{P}(A) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}\lambda(x) = \int_A \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\mathrm{d}x$$

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x) = \text{Probability Density Function !}$$

Quick Probability Recap

Radon Nikodym Theorem – Importance Sampling

$$\mathbb{P} << \mathbb{Q}$$

$$\int_\Omega f(x) \mathrm{d}\mathbb{P}(x) = \int_\Omega f(x) \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(x) \mathrm{d}\mathbb{Q}(x)$$

$$\mathbb{E}_\mathbb{P}[f(X)] = \mathbb{E}_\mathbb{Q}\left[f(X)\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(X)\right]$$

$$\mathbb{E}_\mathbb{P}[f(X)] = \mathbb{E}_\mathbb{Q}\left[f(X)\frac{p(X)}{q(X)}\right]$$

Quick Probability Recap

Modes of Convergence - Exercise

- What does It mean for two random variables to be equal ? Is it as simple as saying they have the same distribution ?

Quick Probability Recap

Modes of Convergence

- What does It mean for two random variables to be equal ? Is it as simple as saying they have the same distribution ? (Law X = Distribution of X)

$$\mathbb{P}\left(|X - Y| > \epsilon\right) = 0$$

$$\mathbb{P}\left(|X - Y| = 0\right) = 1$$

$$\mathbb{E}\left[|X - Y|^p\right] = 0$$

$$\mathrm{Law}\, X = \mathrm{Law}\, Y$$

Quick Probability Recap

Modes of Convergence

- In particular we speak about modes of convergence when we consider limits

$$\lim_{n\to\infty} \mathbb{P}\left(|X - X_n| > \epsilon\right) = 0$$

$$\lim_{n\to\infty} \mathbb{P}\left(|X - X_n| = 0\right) = 1$$

$$\lim_{n\to\infty} \mathbb{E}\left[|X - X_n|^p\right] = 0$$

$$\mathrm{Law}\, X = \lim_{n\to\infty} \mathrm{Law}\, X_n$$

Quick Probability Recap

Modes of equality/convergence of r.v.s.

$$
\begin{array}{ccc}
\lim_{n\to\infty} \mathbb{P}\left(|X - X_n| > \epsilon\right) = 0 & \Longleftarrow & \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1 \\
\Uparrow & \searrow\!\!\searrow & \Downarrow \\
\lim_{n\to\infty} \mathbb{E}\left[|X - X_n|^p\right] = 0 & \Longrightarrow & \mathrm{Law}\, X = \lim_{n\to\infty} \mathrm{Law}\, X_n
\end{array}
$$

SDEs

Heuristic 1 – Discrete Time Markov Chain (Euler Maruyama Discretisation)

$$X_0 \sim \pi,$$

$$\epsilon_n \sim \mathcal{N}(0, \gamma I)$$

$$X_{n+1} = X_n + f(X_n, n)\delta t + \sqrt{\delta t}\epsilon_n,$$

SDEs

Heuristic 2 – Langevin Dynamics and White Noise

- Consider the ODE + Noise

$$X_0 \sim \pi,$$

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = f(X_t, t) + \gamma w(t),$$

$$w(\cdot) \sim \mathcal{GP}(0, \mathbb{I}_{s=t})$$

SDEs

Stochastic Integrals - Types

$$Y_t = \int_0^t X_s \mathrm{d}s$$

- Can think of this as a Reimann integral with convergence asserted in the $\mathscr{L}^p(\mathbb{P})$ sense

$$Z_t = \int_0^t Y_s \mathrm{d}X_s$$

- Now integrating against/wrt to random variable. Not so simple to define. Reimann conditions fail

SDEs

Stochastic Integrals – Counter Example

$$\mathbb{E}\left[\sum_{k=1}^{n} W_{t_k}(W_{t_{k+1}} - W_{t_k})\right] = 0$$

$$\mathbb{E}\left[\sum_{k=1}^{n} W_{t_{k+1}}(W_{t_{k+1}} - W_{t_k})\right] = t$$

- Where you evaluate the integrand (within the grid) changes the result, thus violating the conditions required to be Reimann integrable (remember upper and lower Darboux sums must much)

SDEs

Stochastic Integrals - Definition

- First partition the grid [0,t] $t_{k+1} - t_k = \frac{t}{N}$

- Now we make the following assumption

$$\lim_{n \to \infty} \mathbb{E}\left[\int_0^t |Y_t - Y_t^{(n)}|^2 \mathrm{d}s\right] = 0 \quad \text{s.t.} \quad Y^{(n)}(t) = \sum_{k=1}^{n} Y_{t_k} \mathbb{I}_{t \in [t_k, t_{k+1})}(t)$$

- Then the Ito Integral is defined as:

$$\int_0^t Y_s \mathrm{d}\overset{\sim \mathcal{N}(0,\mathrm{d}s)}{W_s} \overset{\mathscr{L}^2(\mathbb{P})}{=} \lim_{n \to \infty} \sum_{k=1}^{n} Y_{t_k}(W_{t_{k+1}} - W_{t_k})$$

Martingales

Conditional Expectation - Martingale

$$\mathbb{E}\left[X_t|\mathcal{F}_s\right] = X_s$$

$$\Downarrow$$

$$\mathbb{E}\left[X_t|X_s\right] = \mathbb{E}\left[X_t|\sigma(X_s)\right] = X_s$$

Conditional Expectation, MSE

Quick Aside (Useful Later)

The optimal predictor of X as a function of Y (Hilbert projection)

$$\underset{f-\text{is measurable}}{\arg\min} \quad \mathbb{E}\left(X - f(Y)\right)^2$$

Is given by the conditional expectation:

$$f^*(Y) = \mathbb{E}[X|Y]$$

Martingales

Martingales – Intuitive Intro

The optimal predictor of the future as a function of the past in a martingale:

$$\underset{f-\text{is measurable}}{\arg\min} \quad \mathbb{E}\left(X_{t+\delta} - f(X_t)\right)^2$$

Is given by past itself:

$$f^*(X_t) = \mathbb{E}[X_{t+\delta}|X_t] = X_t$$

SDEs

Stochastic Integrals - Martingales

$$\mathbb{E}\left[\int_0^t X_\tau \mathrm{d}W_\tau \middle| \mathcal{F}_s\right] = \int_0^s X_\tau \mathrm{d}W_\tau$$

Martingales

Stochastic Integrals - Martingales

$$\mathbb{E}\left[\int_0^t X_\tau \mathrm{d}W_\tau\right] = \mathbb{E}\left[\mathbb{E}\left[\int_0^t X_\tau \mathrm{d}W_\tau \middle| \mathcal{F}_0\right]\right]$$

$$= \mathbb{E}\left[\int_0^0 X_\tau \mathrm{d}W_\tau\right] = 0$$

SDEs

Formal Definition - Stochastic Piccard Lindeloff Theorem

- Assumptions (Lipchitz + Linear Growth):

$$|\mu(x,t) - \mu(y,s)| + |\sigma(x,t) - \sigma(y,s)| \leq L(|x-y| + |t-s|)$$

$$|\mu(x,t)| + |\sigma(x,t)| \leq C(1 + |x|)$$

- Then we have existence and uniqueness of (in $\mathscr{L}^p(\mathbb{P})$):

$$X_0 \sim \pi$$

$$X_t = X_0 + \int_0^t \mu(X_s, s)\mathrm{d}s + \int_0^t \sigma(X_s, s)\mathrm{d}W_s$$

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t$$

Solving SDEs

Brownian Motion

> Definition: A stochastic process $\mathrm{B_t}$ is a **Brownian motion** if:
>
> 1. $\mathrm{B_0} = 0$ (process starts at 0)
> 2. $\mathrm{B_t}$ is almost surely continuous
> 3. $\mathrm{B_t}$ has independent increments ($\mathrm{B_t} - \mathrm{B_s}$ is independent of $\mathrm{B_s}$)
> 4. $\mathrm{B_t} - \mathrm{B_s} \sim \mathcal{N}(0, \mathrm{t} - \mathrm{s})$ (for $0 \leq \mathrm{s} \leq \mathrm{t}$)

SDE Properties

Quadratic Variation of Brownian Motion

$$\lim_{n\to\infty} \mathbb{E}\left(t - \sum_{i=1}^{n}(W_{t_{i+1}} - W_{t_i})^2\right)^2 = 0$$

SDE Properties

Quadratic Variation of Brownian Motion

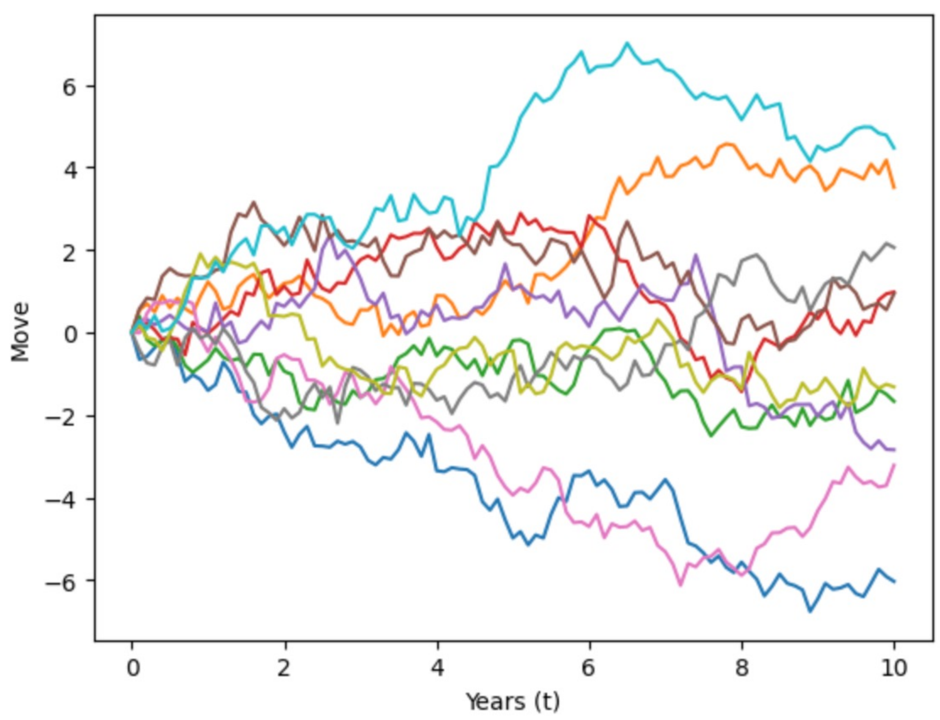
$$\lim_{n\to\infty} \mathbb{E}\left(t - \sum_{i=1}^{n}(W_{t_{i+1}} - W_{t_i})^2\right)^2 = 0$$

	$\mathrm{d}W_t$	$\mathrm{d}t$
$\mathrm{d}W_t$	$\mathrm{d}t$	0
$\mathrm{d}t$	0	0

```python
import numpy as np
import matplotlib.pyplot as plt
M = 10 # number of simulations
t = 10 # Time
n = 100 # steps we want to see
dt = t/n # time step
#simulating the brownian motion
steps = np.random.normal(0, np.sqrt(dt), size=(M, n)).T
origin = np.zeros((1,M))
bm_paths = np.concatenate([origin, steps]).cumsum(axis=0)
time = np.linspace(0,t,n+1)
tt = np.full(shape=(M, n+1), fill_value=time)

#calculate variance and quadratic variation
variance = lambda x: round(np.var(x,axis=0),3)
quadratic_variation = lambda x: round(np.square(x[:-1]-x[1:]).sum(),3)

print("Quadratic variation: ",[quadratic_variation(path)
for path in bm_paths.T[:4]])
print("Variance: ", [variance(path) for path in bm_paths[1:11]])
```

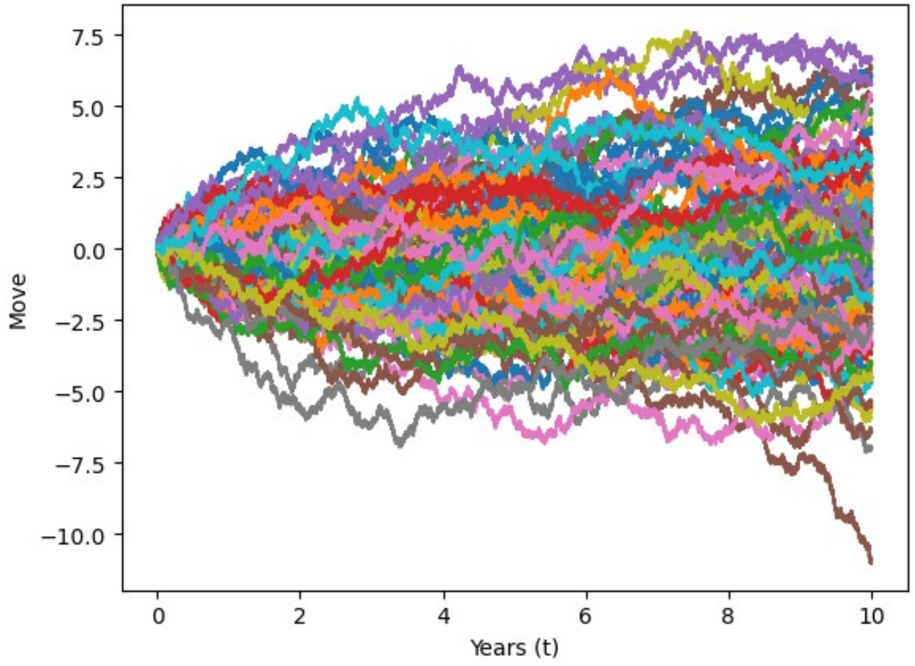
```python
import numpy as np
import matplotlib.pyplot as plt
M = 10 # number of simulations
t = 10 # Time
n = 100 # steps we want to see
dt = t/n # time step
#simulating the brownian motion
steps = np.random.normal(0, np.sqrt(dt), size=(M, n)).T
origin = np.zeros((1,M))
bm_paths = np.concatenate([origin, steps]).cumsum(axis=0)
time = np.linspace(0,t,n+1)
tt = np.full(shape=(M, n+1), fill_value=time)

#calculate variance and quadratic variation
variance = lambda x: round(np.var(x,axis=0),3)
quadratic_variation = lambda x: round(np.square(x[:-1]-x[1:]).sum(),3)

print("Quadratic variation: ",[quadratic_variation(path)
for path in bm_paths.T[:4]])
print("Variance: ", [variance(path) for path in bm_paths[1:11]])
```

```
Quadratic variation:  [9.706, 8.642, 8.719, 8.998]
Variance:  [0.143, 0.239, 0.24, 0.328, 0.277, 0.42, 0.632, 0.704, 0.875, 0.945]
```



```python
import numpy as np
import matplotlib.pyplot as plt
M = 100000 # number of simulations
t = 10 # Time
n = 100000 # steps we want to see
dt = t/n # time step
#simulating the brownian motion
steps = np.random.normal(0, np.sqrt(dt), size=(M, n)).T
origin = np.zeros((1,M))
bm_paths = np.concatenate([origin, steps]).cumsum(axis=0)
time = np.linspace(0,t,n+1)
tt = np.full(shape=(M, n+1), fill_value=time)

#calculate variance and quadratic variation
variance = lambda x: round(np.var(x,axis=0),3)
quadratic_variation = lambda x: round(np.square(x[:-1]-x[1:]).sum(),3)

print("Quadratic variation: ",[quadratic_variation(path)
for path in bm_paths.T[:4]])
print("Variance: ", [variance(path) for path in bm_paths[1:11]])
```

```
Quadratic variation:  [10.024, 10.018, 10.003, 9.839]
Variance:  [0.1, 0.2, 0.301, 0.401, 0.501, 0.602, 0.702, 0.801, 0.904, 1.001]
```



SDEs

Heuristic 2 – Langevin Dynamics and White Noise

- Consider the ODE + Noise

$$X_0 \sim \pi,$$

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = f(X_t, t) + \gamma w(t),$$

$$w(\cdot) \sim \mathcal{GP}(0, \mathbb{I}_{s=t})$$

Solve SDEs just like ODEs?

Heuristic Treatment only takes you so far

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{L}\mathbf{w}(t)$$

$$\frac{d\mathbf{x}}{dt} - \mathbf{F}\mathbf{x}(t) = \mathbf{L}\mathbf{w}(t)$$

$$\exp(-\mathbf{F}t)\frac{d\mathbf{x}}{dt} - \exp(-\mathbf{F}t)\mathbf{F}\mathbf{x}(t) = \exp(-\mathbf{F}t)\mathbf{L}\mathbf{w}(t)$$

$$\frac{d}{dt}\exp(-\mathbf{F}t)\mathbf{x}(t) = \exp(-\mathbf{F}t)\mathbf{L}\mathbf{w}(t)$$

$$\exp(-\mathbf{F}t)\mathbf{x}(t) - \exp(-\mathbf{F}t_0)\mathbf{x}(t_0) = \int_{t_0}^{t} \exp(-\mathbf{F}s)\mathbf{L}\mathbf{w}(s)ds$$

$$\mathbf{x}(t) = \exp(\mathbf{F}(t - t_0))\mathbf{x}(t_0) + \int_{t_0}^{t} \exp(\mathbf{F}(t - s))\mathbf{L}\mathbf{w}(s)ds$$

Some parts of ordinary calculus stop working!

The Chain rule bites us

$$\frac{d\mathbb{E}[\mathbf{x}(t)]}{dt} = \mathbf{F}\mathbb{E}[\mathbf{x}(t)]$$
$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T + \mathbf{L}\mathbf{Q}\mathbf{L}^T$$

$$\begin{aligned}\mathbb{E}[\frac{d\mathbf{x}(t)}{dt}] &= \mathbb{E}[\mathbf{F}\mathbf{x}(t) + \mathbf{L}\mathbf{w}(t)] \\ &= \mathbf{F}\mathbb{E}[\mathbf{x}(t)] + \mathbf{L}\mathbb{E}[\mathbf{w}(t)] \\ &= \mathbf{F}\mathbb{E}[\mathbf{x}(t)]\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T] &= \mathbf{F}\mathbb{E}[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T] + \mathbb{E}[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T]\mathbf{F}^T \\ \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T \\ &\neq \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T + \mathbf{L}\mathbf{Q}\mathbf{L}^T\end{aligned}$$

Ito's Lemma

> Definition: An **Ito process** is an adapted stochastic process X_t that can be expressed as the sum of an integral with respect to time and an integral with respect to a Brownian motion W_t: $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$

Ito's Lemma

Definition: An **Ito process** is an adapted stochastic process X_t that can be expressed as the sum of an integral with respect to time and an integral with respect to a Brownian motion W_t: $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$

Ito's lemma: Let X_t be an Ito process and $f(t, X_t)$ be a function of t and X_t that is twice continuously differentiable with respect to t and X_t. Then $f(t, X_t)$ is also an Ito process, can be denoted Y_t and we can write: $dY_t = df(t, X_t) = \frac{\partial f(t,X_t)}{\partial t}dt + \frac{\partial f(t,X_t)}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2 f(t,X_t)}{\partial X_t^2}(dX_t)^2$

SDE Properties

Given the SDE:

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t$$

Consider a function $f(t, x)$ doubly differentiable in space and admitting single derivatives in time. Then the process $Y_t = f(t, X_t)$ satisfies:

$$\mathrm{d}Y_t = \left(\partial_t f + \nabla f^\top \mu(X_t, t) + \frac{1}{2}\mathrm{tr}(\sigma(X_t, t)^\top \nabla\nabla f \sigma(X_t, t))\right)\mathrm{d}t + \nabla f^\top \sigma(X_t, t)\mathrm{d}W_t$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = ??, \quad \partial_x f = ?? \quad \partial_x^2 f = ??$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = 0, \quad \partial_x f = 1/x \quad \partial_x^2 f = -1/x^2$$

$$\mathrm{d}Y_t = \left(\frac{\mu}{X_t} \cdot X_t - \frac{\sigma^2}{2X_t^2} \cdot X_t^2\right) \mathrm{d}t - \frac{\sigma}{X_t} \cdot X_t \mathrm{d}W_t$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = 0, \quad \partial_x f = 1/x \quad \partial_x^2 f = -1/x^2$$

$$\mathrm{d}Y_t = \left(\mu - \frac{\sigma^2}{2}\right) \mathrm{d}t - \sigma \mathrm{d}W_t$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Now let us solve the SDE:

$$\mathrm{d}Y_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t - \sigma \mathrm{d}W_t$$

$$Y_t = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right)\int_0^t \mathrm{d}s - \sigma \int_0^t \mathrm{d}W_s = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t$$

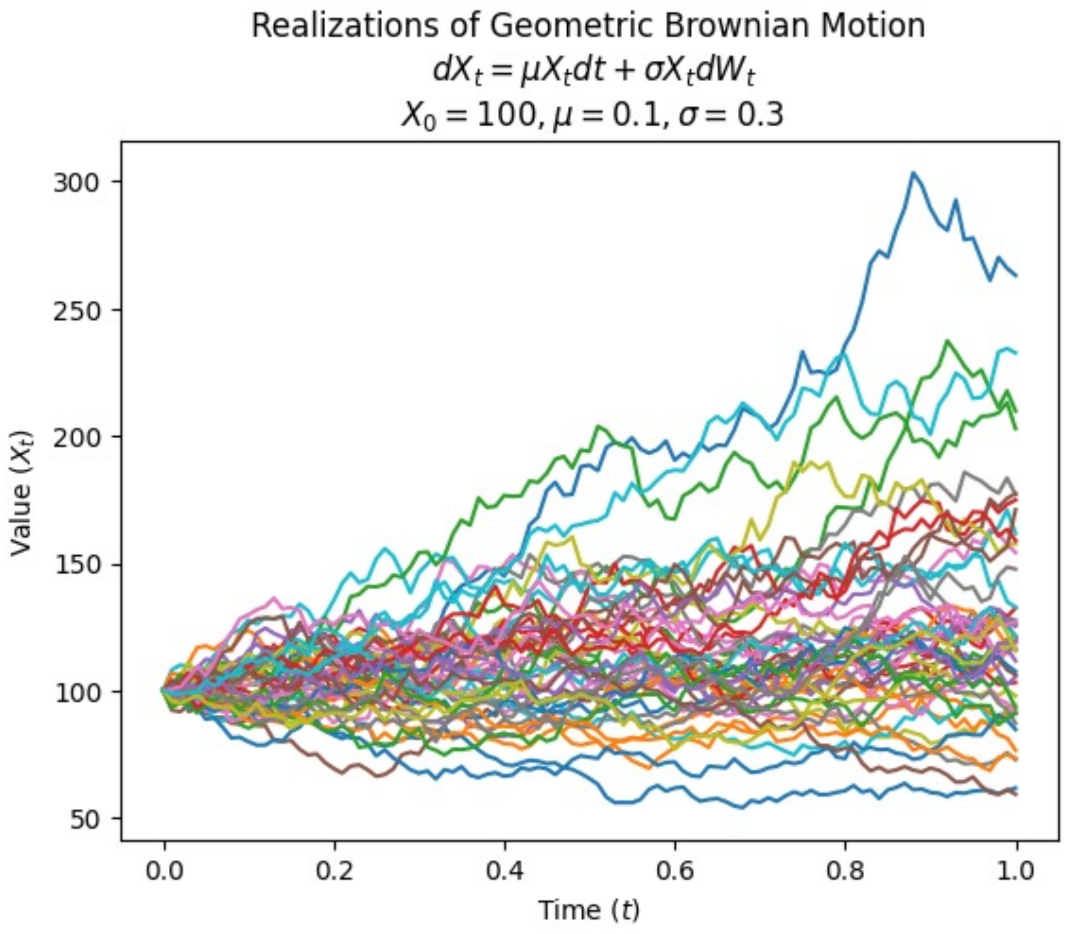
Remember $Y_t = \ln X_t$ thus:

$$X_t = e^{Y_t} = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$$

Geometric Brownian Motion - Simulation

$$X_t = e^{Y_t} = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$$

```
# simulation using numpy arrays
Xt = np.exp(
    (mu - sigma ** 2 / 2) * dt
    + sigma * np.random.normal(0, np.sqrt(dt), size=(M,n)).T
)
```



Linear SDEs

OU - Process

Mean reverting process. Reverts you back to mu.

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = \alpha(\mu - X_t)\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Linear SDEs

OU - Process

For simplicity focus on the 0-mean case.

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Linear SDEs

OU - Process

Can be solved analytically via Integrating factor + Ito's Lemma (notice how X_t looks like the DDPM kernel):

$$X_t = X_0 e^{-\alpha t} + (1 - e^{-2\alpha t})^{1/2} W_1$$

$$X_t = X_0 e^{-\alpha t} + W_{1-e^{-2\alpha t}}$$

Linear SDEs

OU - Process

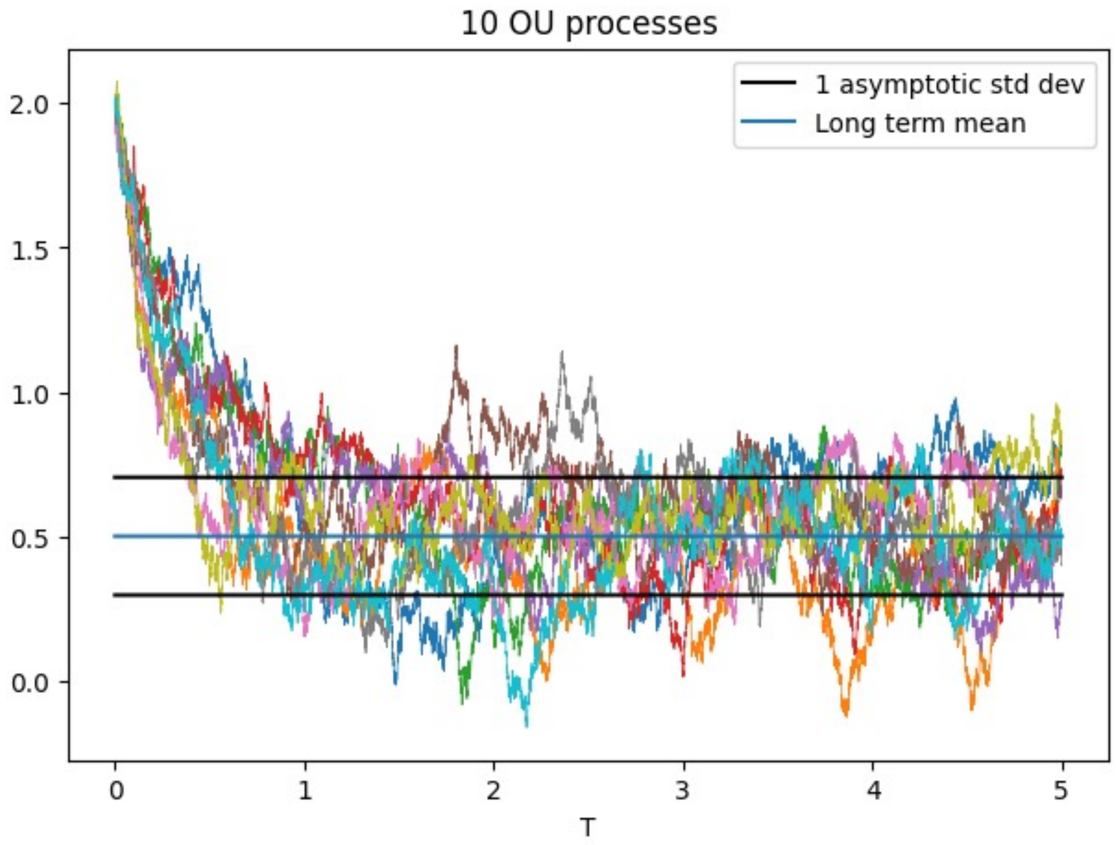
Can be solved analytically via Integrating factor + Ito's Lemma (notice how X_t looks like the DDPM kernel):

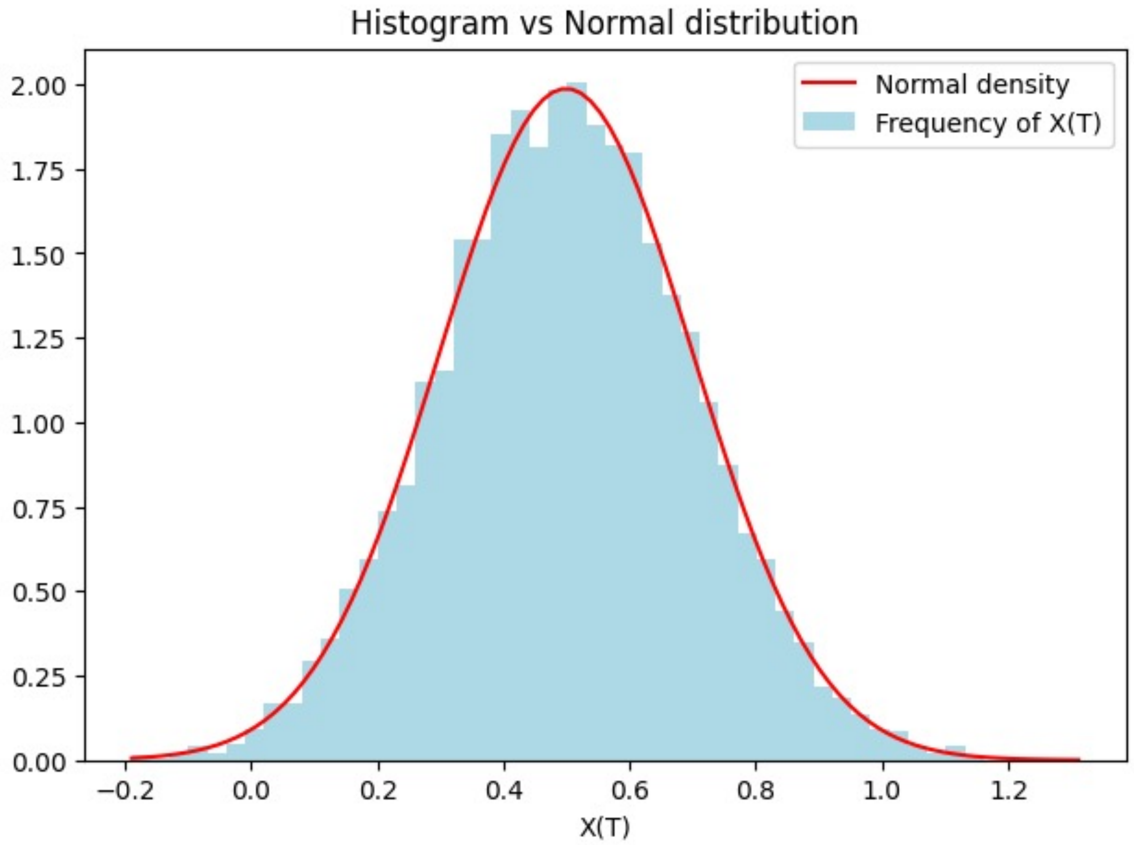
$$X_t = X_0 e^{-\alpha t} + (1 - e^{-2\alpha t})^{1/2} W_1$$

$$X_t = X_0 e^{-\alpha t} + W_{1-e^{-2\alpha t}}$$

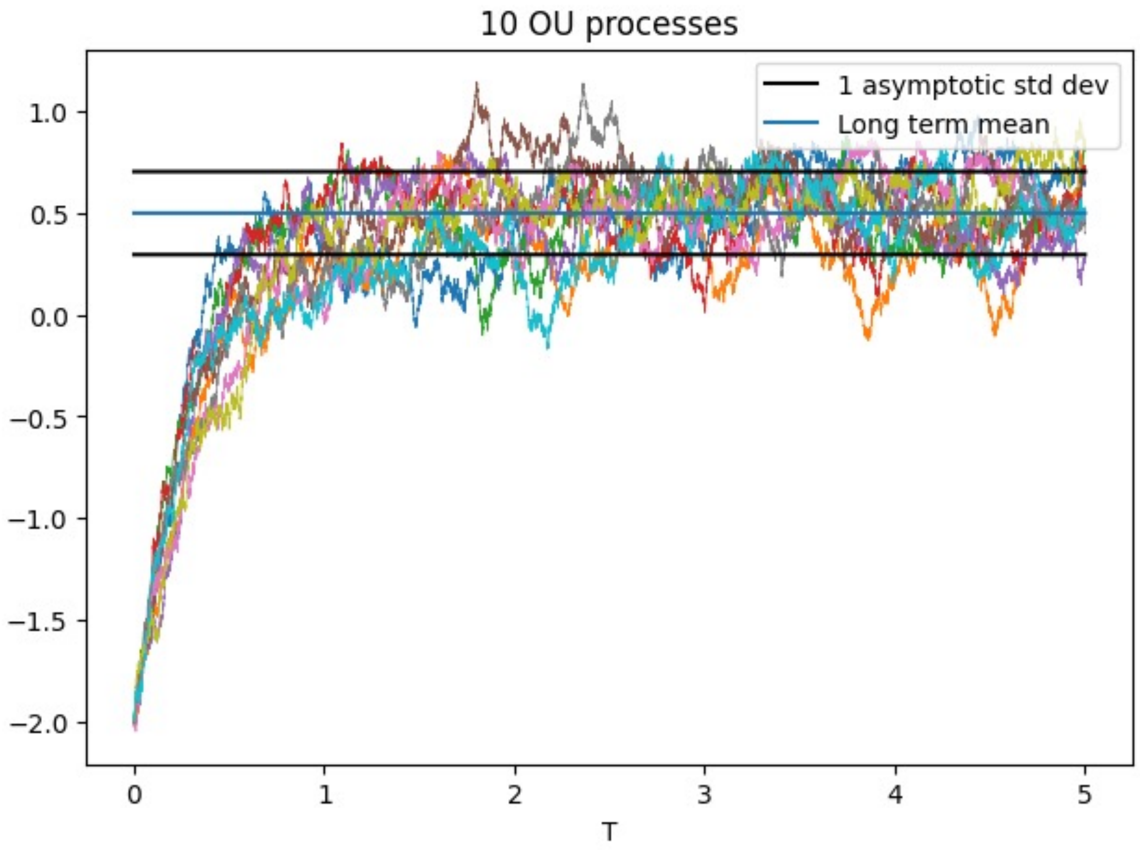
```
std_dt = np.sqrt(sigma**2 / (2 * kappa) * (1 - np.exp(-2 * kappa * dt)))
for t in range(0, N - 1):
    X[:, t + 1] = theta + np.exp(-kappa * dt) * (X[:, t] - theta) + std_dt * W[:, t]
```

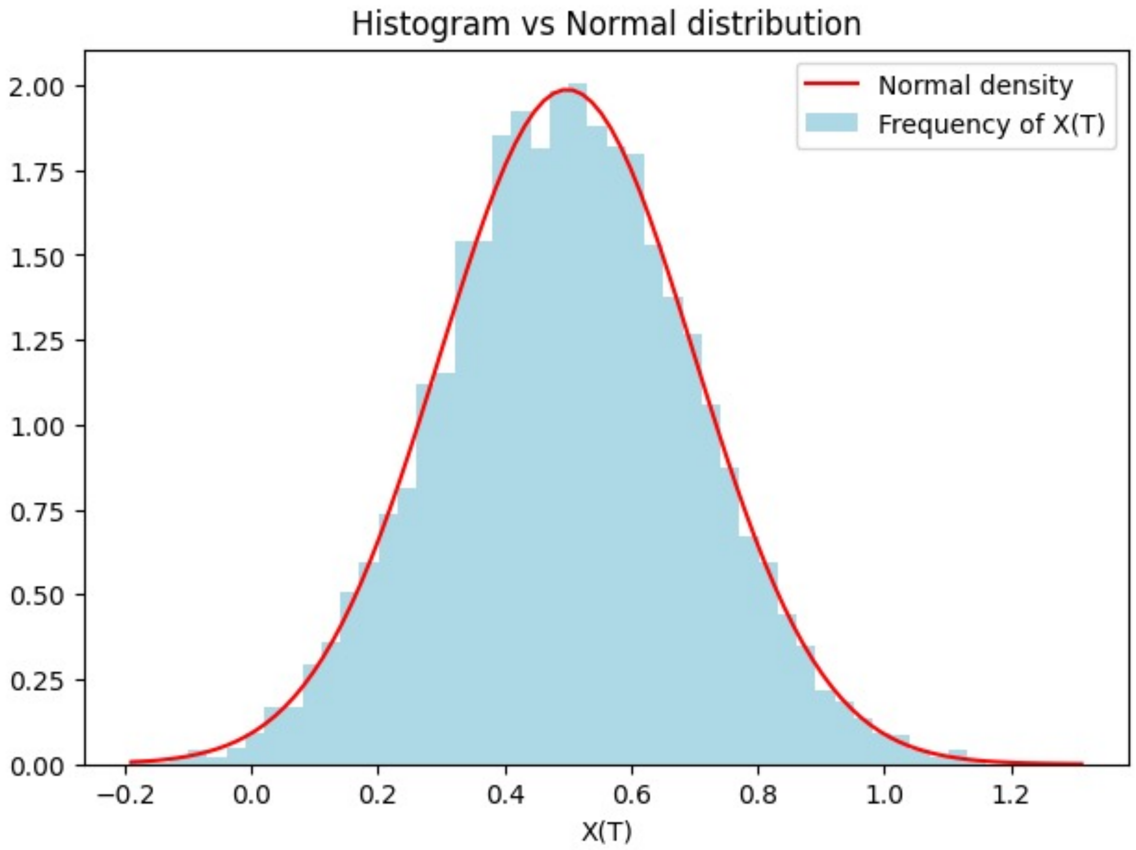
OU Process - Simulation





OU Process - Simulation





Linear SDEs

OU - Process

Intuitively you can see how the limit behaves:

$$\lim_{t \to \infty} X_t \overset{??}{=} W_1 \sim \mathcal{N}(0, I)$$

This is a completely informal/heuristic treatment. Calling it a heuristic is kind, but you can see where it is going.

Linear SDEs

OU - Process

More formal arguments can be made:

$$||\mathrm{Law} X_t - \mathcal{N}(0, I)||_{\mathrm{TV}} \leq C e^{-\alpha^{1/2} t}$$

Can be a bit tricky to show from scratch, typically involves working with the Fokker Plank Equation + Using an Eigen decomposition of its semi group. Alternatively, Martingale methods have also been used.

Convergence in KL, W_p can also be attained see Bakry, Gentil, Ledoux **Analysis and Geometry of Markov Diffusion Operators**.

Non Linear SDEs - Simply Discretise

Euler Maruyama (EM) Discretisation

To solve SDEs of the form

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t$$

We simply discretize them via EM

$$X_0 \sim \pi,$$
$$\epsilon_{t_k} \sim \mathcal{N}(0, \gamma I)$$
$$X_{t_{k+1}} = X_{t_k} + \mu(X_{t_k}, t_k)\delta t + \sqrt{\delta t}\sigma(X_{t_k}, t_k)\epsilon_{t_k},$$

Can prove convergence in $\mathscr{L}^p(\mathbb{P})$. Can we design better integrators ?

Fokker Planck Equation

Definition: The Fokker-Planck Equation (FPE) describes the evolution of the probability density of an SDE. For a general SDE of the form $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$, it is given by

$$\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x}[\mu(x, t)p(x, t)] + \frac{\partial^2}{\partial x^2}[D(x, t)p(x, t)]$$

where $p(x, t)$ is the probability density of the SDE at time t and x and $D(x, t) = \frac{\sigma^2(X_t, t)}{2}$ is defined as the diffusion coefficient.

Fokker Plank Equation

How does the marginal density evolve (SDEs ⇔ Parabolic PDEs)

What is the probability density of the SDE solution at a given time ?

$$\mathrm{Law} X_t = p_t(x) = ???$$

There's a special PDE (think heat equation) whose solution yield the marginal density:

$$\partial_t p_t(x) = -\sum_{i=1}^{d} \partial_{x_i}[\mu_i(t, x_i) p_t(x)] + \sum_{i,j=1}^{d} \partial_{x_i, x_j}[\sigma\sigma_{ij}^\top(t, x) p_t(x)]$$

Fokker Plank Equation

How does the marginal density evolve (SDEs ⇔ Parabolic PDEs)

What is the probability density of the SDE solution at a given time ?

$$\mathrm{Law} X_t = p_t(x) = ???$$

There's a special PDE (think heat equation) whose solution yield the marginal density:

$$\partial_t p_t(x) = \mathcal{P}(p_t)$$

FPE for Brownian Motion

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[\mu(x,t)p(x,t)] + \frac{\partial^2}{\partial x^2}[D(x,t)p(x,t)]$$

$$\frac{\partial p(x,t)}{\partial t} - \frac{1}{2}\frac{\partial^2 p(x,t)}{\partial x^2} = 0$$

Infinitesimal Generator

Uniquely Characterises PDE and Adjoint to FPK Operator

Consider the following operator for a given SDE

$$\mathcal{A}_t[f(x)] = \lim_{t \to 0} \frac{\mathbb{E}[f(X_t)] - x}{t}$$

Can be shown to reduce to:

$$\mathcal{A}_t[f] = \partial_t f + \mu \cdot \nabla f + \frac{1}{2} \sum_{ij} [\sigma \sigma^\top]_{ij}(x,t) \partial_{x_i, x_j} f$$

$$= \partial_t f + \mathcal{P}^\dagger(f)$$