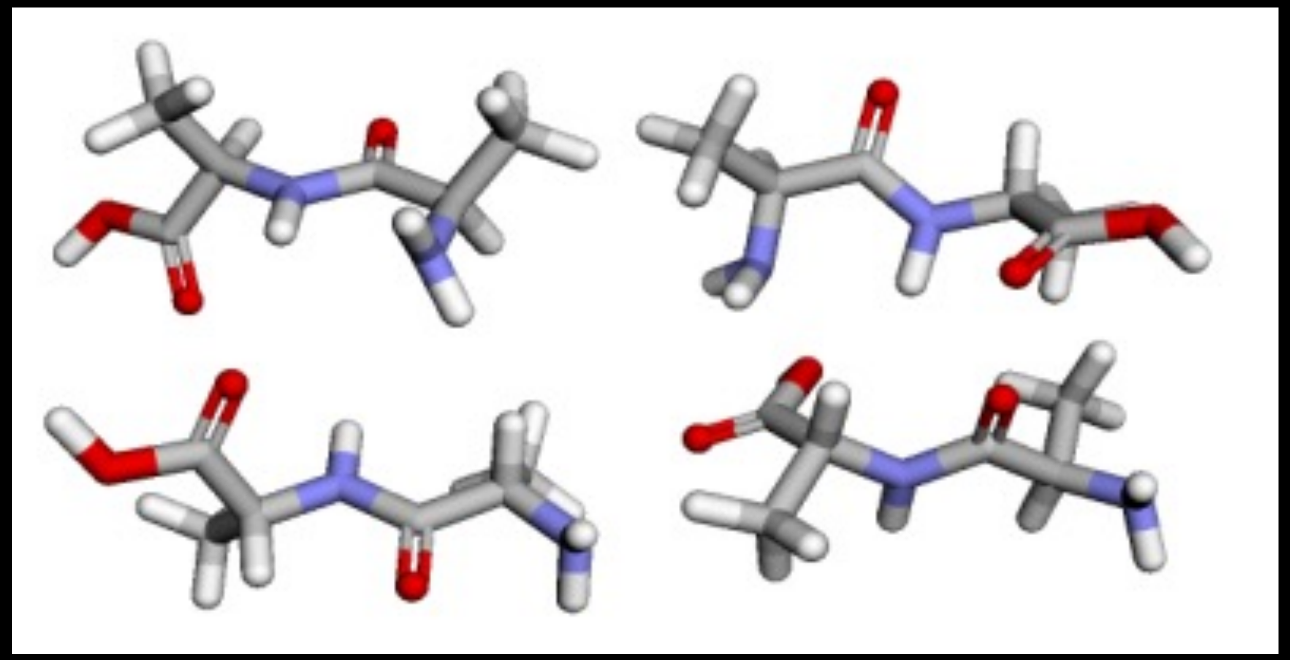
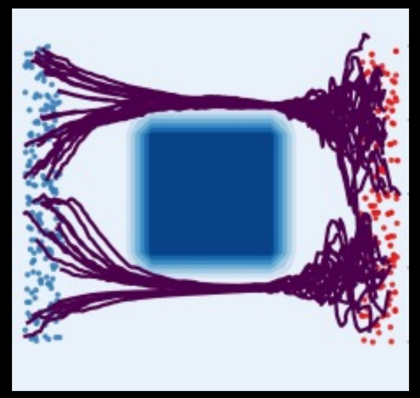
Generative Modelling


Bayesian Inference / Protein Folding



Filtering / Data Assimilation



# Diffusion Models and SDEs

**Lecture 1:**

**A very fast paced introduction to the foundations.**

# Quick Probability Recap

## Probability Space

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

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

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

- Sample Space

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

- 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

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

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

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

- Sample Space

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

- 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:

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

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

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

# Quick Probability Recap

## Stochastic Process

$$A \in \mathcal{F}_s \implies X^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$$

- Collection of Random Variables (Measurable Maps) !

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

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

# 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 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$

# Quick Probability Recap

## Lebesgue Integral

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

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

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

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

# Quick Probability Recap

## Lebesgue Integral Matches Traditional Riemann Integral

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

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

Also note we can replace lambda with a probability distribution:

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

# 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 (Exercise)

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

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 (Exercise:  Probability Density)

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

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

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 (e.g. 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 Convergence

$$
\begin{tikzcd}
\lim_{n\to\infty} \mathbb{P}\left(|X - X_n| > \epsilon\right) = 0 \arrow[dr, Rightarrow] & \lim_{n\to\infty} \mathbb{P}\left(|X - X_n| = 0\right) = 1 \arrow[l, Rightarrow] \arrow[d, Rightarrow] \\
\lim_{n\to\infty} \mathbb{E}\left[|X - X_n|^p\right] = 0 \arrow[u, Rightarrow] \arrow[r, Rightarrow] & \mathrm{Law}\, X = \lim_{n\to\infty} \mathrm{Law}\, X_n
\end{tikzcd}
$$

# SDEs

## Heuristic 1 - Discrete Time Markov Chain

$$X_0 \sim \pi,$$

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

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

# 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 - Definition

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

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

- Then the Ito Integral is defined as:

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

# SDEs

## Stochastic Integrals - Counter Example

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

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

# 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 - Martingales

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

$$\Downarrow$$

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

# SDEs

## Stochastic Integrals - Martingales

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

# SDEs

## 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 $\mathscr{L}^p(\mathbb{P})$ ):

$$X_0 \sim \pi$$

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

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