Basics of (Stochastic) Differential Equation

Il-Chul Moon
Department of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Differential Equations

- Equation to specify the relation between functions and their derivatives
  - Function to describe heat transfer, mechanical movements, social dynamics…
  - Derivative to describe the gradual changes of such functions
    - For example) More diseases susceptible people and a few infected people → More infected people
      - Two Functions : Actual population count at susceptible and infected status
      - Two Derivatives : Increment and decrement of the population count at such status

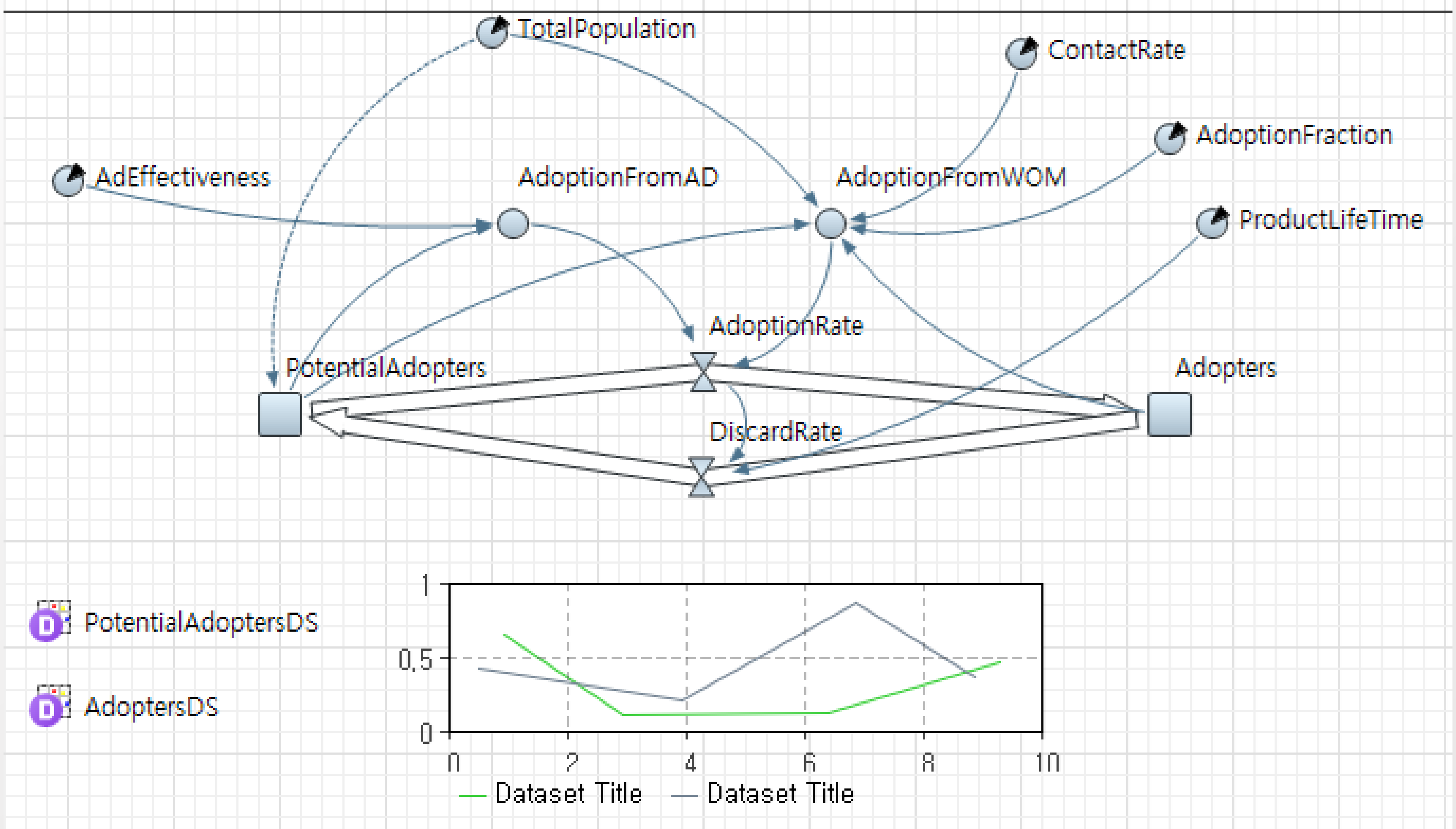
- Our main query on differential equations
  - What would be the value of functions at any given time t?
    - Given the differential equation
    - With an arbitrarily provided function values at time 0
  - The answer is approached in two directions
    - Simulation : by creating an actual instantiation of functions and their derivatives; and by running over-time
    - Calculation : by deriving a closed-form solution to calculate the function value at any time t

©Wikipedia

Types of Differential Equation

- Ordinary Differential Equation(ODE)
  - $F\left(t, y, y', \dots, y^{(n)}\right) = 0$
    - y: dependent variable, t: independent variable, $y = f(t)$: unknown function of t
  - For example) SIR model in disease spread
  - Linear differential equation
    - $F\left(t, y, y', \dots, y^{(n)}\right) = 0$
      - is specified as a linear combination of $y, y', \dots, y^{(n)}$
  - Initial Value Problem(IVP)
    - On some interval I containing t , the problem $\frac{dy}{dt} = f(t, y)$ and $y(t_0) = y_0$ is called and Initial Value Problem where y_0 is an arbitrary specified real constant.
    - $y(t_0) = y_0$ is called initial conditions
- Partial Differential Equation(PDE)
  - $F\left(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n, \mathrm{y}, \frac{\partial \mathrm{y}}{\partial \mathrm{x}_1}, \dots, \frac{\partial \mathrm{y}}{\partial \mathrm{x}_n}\right) = 0$, $\mathrm{y}(\mathrm{x})$: unknown function of x
    - First-order partial differential equation because of the limited first-order partial derivative relations
  - Parabolic (time-dependent) Partial Differential Equation: $y = f(t, x), F\left(x, t, y, \frac{\partial y}{\partial x}, \frac{\partial y}{\partial t}, \dots\right) = 0$
  - For example) Heat equation model in heat transfer
    - $\frac{\partial u}{\partial t} = \Delta u \ (= \frac{\partial^2 u}{\partial x^2})$
- Stochastic Differential Equation(SDE)
  - $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$
    - X_t, B_t: stochastic process
    - B_t: Brownian motion
  - For example) Black-Scholes model in finance, Ornstein-Uhlenbeck process
- And many more…………………..

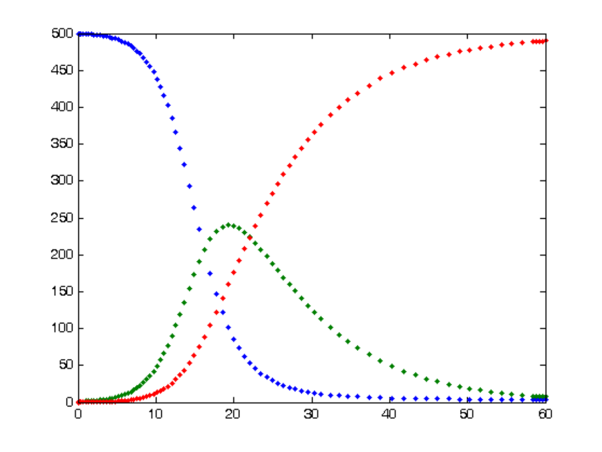
Detour) Simulation with System Dynamics



Ordinary Differential Equation

- Ordinary Differential Equation
  - Differential equation with a set of functions on an independent variable and the derivatives of the functions
  - $F\left(t, y, y', \dots, y^{(n)}\right) = 0$
    - y: dependent variable
    - t: independent variable
    - $y = f(t)$: unknown function of t
- System of Differential Equations
  - Coupled differential equations
    - $\boldsymbol{y}$: $\boldsymbol{y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]$
    - $$\begin{pmatrix} f_1\left(t, y, y', \dots, y^{(n)}\right) \\ f_2\left(t, y, y', \dots, y^{(n)}\right) \\ \dots \\ f_m\left(t, y, y', \dots, y^{(n)}\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}$$
  - Example : SIR Model
    - $\frac{dS}{dt} = -\beta IS, \frac{dI}{dt} = \beta IS - \nu \mathrm{I}, \frac{dR}{dt} = \nu \mathrm{I}$
      - $\boldsymbol{y} = [S, I, R]$
- How to calculate $\boldsymbol{y}$ for an arbitrary t?
  - Exact closed-form solution → Not known, Asymptotic approximation → published in 2020

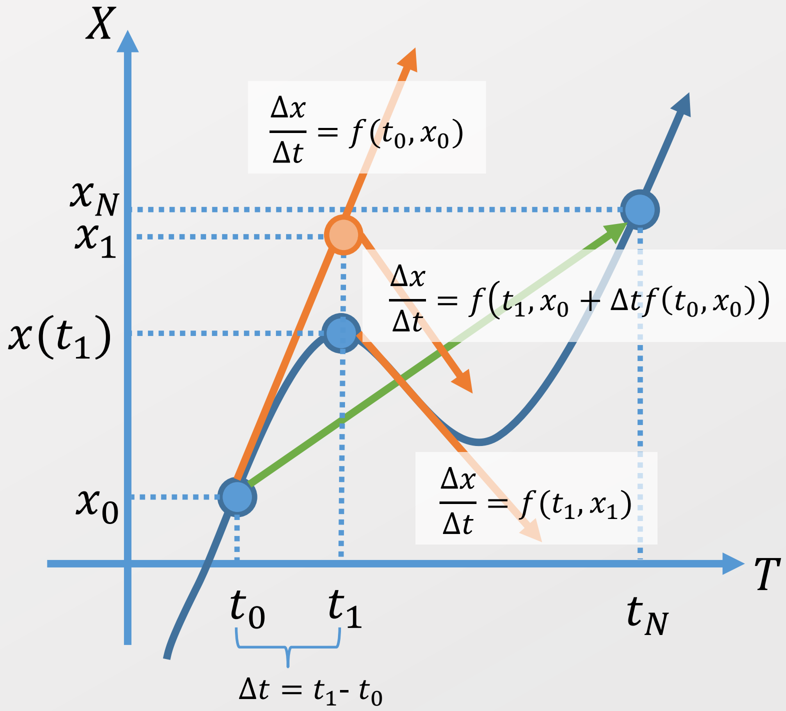
SIR Model from (Ross, 1916), (Kermack and McKendrick, 1927)…

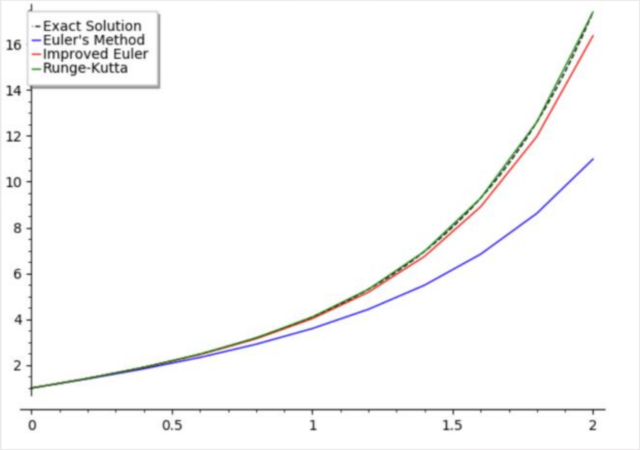


https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

Euler Method

- Let's start from the simplest version: Ordinary Differential Equation
- How to (Approximately) Calculate Differential Equation
  - "Difference" is specified by the equation
  - $\frac{\mathrm{d}x}{\mathrm{d}t} = f(t,x),\ x(t_0) = x_0$
- Euler method
  - Taking a small step of $\Delta t = t_{n+1} - t_n$
  - Propagate until the target time of t_N
  - $x(t_{n+1}) = x_n + \Delta t f(t_n, x_n) + O((\Delta t)^2)$
  - $x_{n+1} = x_n + \Delta t f(t_n, x_n)$
- Modified Euler method
  - Taking a single point of differentiation could be too biased.
  - Take two neighboring differentiations and average them
  - $x_{n+1} = x_n + \frac{\Delta t}{2}\left[f(t_n, x_n) + f\left(t_n + \Delta t, x_{n+1}^p\right)\right]$
    - $x_{n+1}^p = x_n + \Delta t f(t_n, x_n)$





Runge-Kutta Method

Differential equation:

$\frac{\mathrm{d}x}{\mathrm{d}t} = f(t,x),\ x(t_0) = x_0$

Total derivation:

$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n$

- Modified Euler method
  - $x_{n+1} = x_n + \frac{\Delta t}{2}\left[f(t_n, x_n) + f\left(t_n + \Delta t, x_{n+1}^p\right)\right],\ x_{n+1}^p = x_n + \Delta t f(t_n, x_n)$
  - $x_{n+1} = x_n + \frac{\Delta t}{2}[f(t_n, x_n) + f(t_n + \Delta t, x_n + \Delta t f(t_n, x_n))]$
    - Let's say $k_1 = \Delta t f(t_n, x_n),\ k_2 = \Delta t f(t_n + \alpha\Delta t, x_n + \beta k_1)$
  - Then, $x(t_{n+1}) \approx x_n + \frac{1}{2}k_1 + \frac{1}{2}k_2$ assuming $\alpha = \beta = 1$
  - Further generalizing with coefficient,
    - $x(t_{n+1}) \approx x_n + ak_1 + bk_2$ when say $k_1 = \Delta t f(t_n, x_n),\ k_2 = \Delta t f(t_n + \alpha\Delta t, x_n + \beta k_1)$
    - This is called Runge-Kutta 2 (RK2) method
- RK2 : $x(t_{n+1}) \approx x_n + a\Delta t f(t_n, x_n) + b\Delta t f(t_n + \alpha\Delta t, x_n + \beta k_1)$
  - Taylor series on $f(t_n + \alpha\Delta t, x_n + \beta k_1)$: Taylor series on multivariate function
    - $f(t_n + \alpha\Delta t, x_n + \beta\Delta t f(t_n, x_n)) = f(t_n, x_n) + \alpha\Delta t \frac{\partial}{\partial t} f(t_n, x_n) + \beta\Delta t f(t_n, x_n)\frac{\partial}{\partial x} f(t_n, x_n) + O((\Delta t)^2)$
- $x(t_{n+1}) \approx x_n + a\Delta t f(t_n, x_n) + b\Delta t f(t_n + \alpha\Delta t, x_n + \beta k_1)$

$$= x_n + a\Delta t f(t_n, x_n) + b\Delta t\left[f(t_n, x_n) + \alpha\Delta t\frac{\partial}{\partial t} f(t_n, x_n) + \beta\Delta t f(t_n, x_n)\frac{\partial}{\partial x} f(t_n, x_n) + O((\Delta t)^2)\right]$$

$$= x_n + (a+b)\Delta t f(t_n, x_n) + (\Delta t)^2[\alpha b f_t(t_n, x_n) + \beta \mathrm{b} f(t_n, x_n) f_x(t_n, x_n)] + O((\Delta t)^3)$$

Constraints of Runge-Kutta Method

Differential equation:
$\frac{\mathrm{d}x}{\mathrm{d}t} = f(t,x),\ x(t_0) = x_0$

- Taylor Expansion from RK2
  - $x(t_{n+1}) \approx x_n + a\Delta t f(t_n, x_n) + b\Delta t f(t_n + \alpha\Delta t, x_n + \beta k_1)$
    $= x_n + (a+b)\Delta t f(t_n, x_n) + (\Delta t)^2[\alpha b f_t(t_n, x_n) + \beta b f(t_n, x_n) f_x(t_n, x_n)] + O((\Delta t)^3)$
- Taylor Expansion from $x(t_{n+1})$, directly
  - $x(t_{n+1}) = x(t_n + \Delta t) = x(t_n) + \frac{\Delta t}{1!}\frac{d}{dt}x(t)|_{t=t_n} + \frac{(\Delta t)^2}{2!}\frac{d^2}{(dt)^2}x(t)|_{t=t_n} + O((\Delta t)^3)$
    $= x(t_n) + \Delta t f(t_n, x_n) + \frac{1}{2}(\Delta t)^2 \frac{d}{dt} f(t_n, x_n) + O((\Delta t)^3)$
    $= x(t_n) + \Delta t f(t_n, x_n) + \frac{1}{2}(\Delta t)^2 \left[\frac{\partial}{\partial t} f(t_n, x_n) + f(t_n, x_n)\frac{\partial}{\partial x} f(t_n, x_n)\right] + O((\Delta t)^3)$
    $\approx x_n + \Delta t f(t_n, x_n) + \frac{1}{2}(\Delta t)^2[f_t(t_n, x_n) + f(t_n, x_n) f_x(t_n, x_n)]$

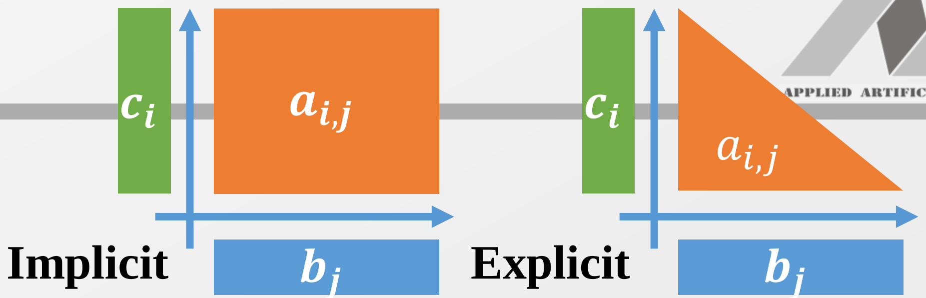
Total derivation:
$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n$

- Now, we can see the constraints on RK2
  - Remember that direct taylor expansion has no approximation as RK2 besides of high-order terms
- $a + b = 1,\ \alpha b = \frac{1}{2},\ \beta b = \frac{1}{2}$

Modified Euler Method ($a = b = \frac{1}{2}, \alpha = \beta = 1$):
$x_{n+1} = x_n + \frac{\Delta t}{2}[f(t_n, x_n) + f(t_n + \Delta t, x_n + \Delta t f(t_n, x_n))]$

Higher-Order Runge-Kutta Method

- RK2 : $x_{n+1} \approx x_n + ak_1 + bk_2$
  - $k_1 = \Delta t f(t_n, x_n),\ k_2 = \Delta t f(t_n + \alpha \Delta t, x_n + \beta k_1)$
  - $a + b = 1,\ \alpha b = \frac{1}{2},\ \beta b = \frac{1}{2}$
- General form RK
  - h : small step between partitions
  - Implicit form : $x_{n+1} \approx x_n + h\sum_{i=1}^{s} b_i k_i$
    - $k_i = f(t_n + c_i h, x_n + h\sum_{j=1}^{s} a_{i,j} k_j)$
  - Explicit form : $x_{n+1} \approx x_n + h\sum_{i=1}^{s} b_i k_i$
    - $k_i = f(t_n + c_i h, x_n + h\sum_{j=1}^{i-1} a_{i,j} k_j$
  - Then, how to represent the constraints?
    - $\sum_{i=1}^{s} b_i = 1,\ \sum_{j=1}^{i-1} a_{i,j} = c_i,\ i = 2, \dots, s$
    - Butcher tableau!
- RK4 : $x_{n+1} = x_n + \frac{1}{6}hk_1 + \frac{1}{3}hk_2 + \frac{1}{3}hk_3 + \frac{1}{6}hk_4$
  - $k_1 = f(t_n, x_n), k_2 = f\left(t_n + \frac{h}{2}, x_n + h\frac{k_1}{2}\right)$
  - $k_3 = f(t_n + \frac{h}{2}, x_n + h\frac{k_2}{2}),\ k_4 = f(t_n + \frac{2h}{2}, x_n + hk_3)$
- This is only a single example. There are many RKs…………



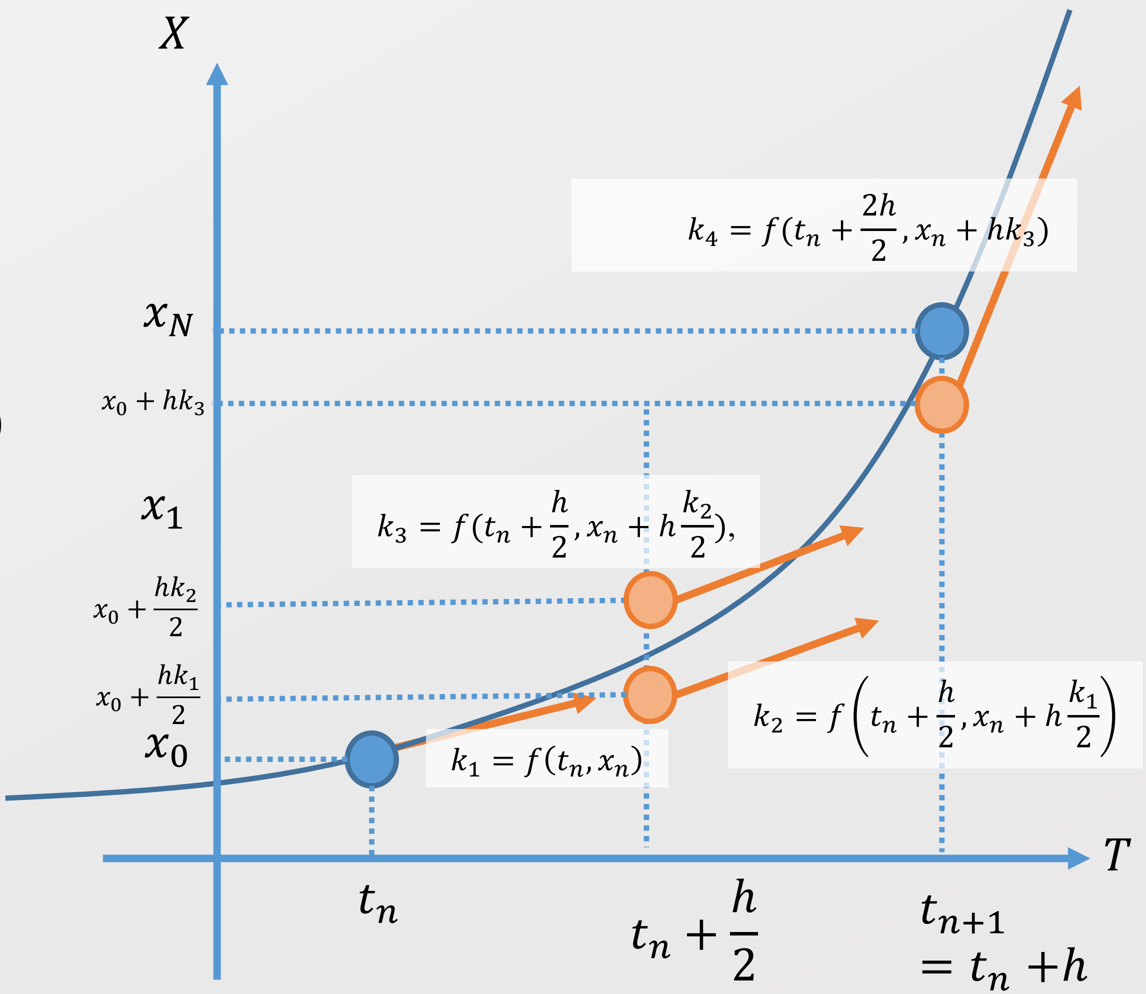
s dim.

0				
1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	1/3	1/3	1/6

s dim.

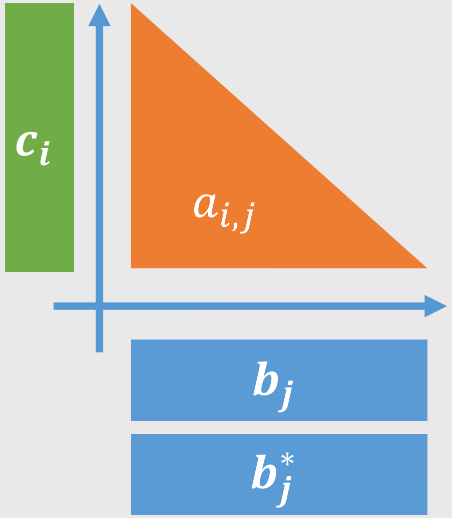
Geometry of RK4 and Reuse of K

- RK2 : $x_{n+1} \approx x_n + \frac{1}{2}k_1 + \frac{1}{2}k_2$
  - $k_1 = f(t_n, x_n)$, $k_2 = f\left(t_n + \frac{h}{2}, x_n + h\frac{k_1}{2}\right)$
- RK4 : $x_{n+1} = x_n + \frac{1}{6}hk_1 + \frac{1}{3}hk_2 + \frac{1}{3}hk_3 + \frac{1}{6}hk_4$
  - $k_1 = f(t_n, x_n), k_2 = f\left(t_n + \frac{h}{2}, x_n + h\frac{k_1}{2}\right)$
  - $k_3 = f(t_n + \frac{h}{2}, x_n + h\frac{k_2}{2}),\ k_4 = f(t_n + \frac{2h}{2}, x_n + hk_3)$
- Eventually, RK is a numerical method of solving an initial value problem. RK searches the optimal weights with respect to the order of accuracy given the number of function evaluations.
  - RK4 can reuse k_1 and k_2 from RK2
  - Common intermediate steps!
- We can increase the order for further accuracy
  - In spite of Complexity



Adaptive Runge-Kutta

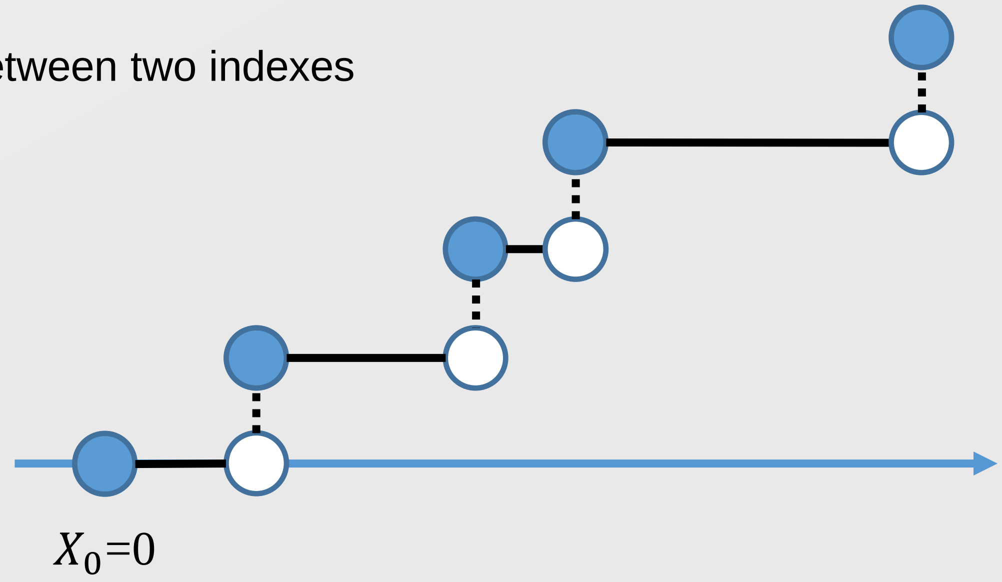
- RK is a numerical method of an IVP with fixed step-size
- Let's design a framework for adaptive step size
  - Two RK methods with p and $p-1$ order, which shares common intermediate steps
  - $(p-1)$ order RK : $x_{n+1}^{*} = x_n + h\sum_{i=1}^{s} b_i^{*} k_i$, $k_i = f(t_n + c_i h, x_n + h\sum_{j=1}^{i-1} a_{i,j} k_j)$
    - k_i is shared with p order RK
  - Error between two RKs
    - $e_{n+1} = x_{n+1} - x_{n+1}^{*} = h\sum_{i=1}^{s}(b_i - b_i^{*})k_i$
    - This is not the actual error, but a **simulated** error from the lower-order RK to the high-order RK
  - Now, we have more parameter of b_i^{*} → expanding the Butcher tableau
- Adaptive step-size
  - If the error goes beyond a user-defined threshold → reduce step-size
  - If the error becomes less than a user-defined threshold → increase step-size
- There are many variations of this framework
  - Dormand-Prince Method
    - Two RKs of RK4 and RK5
    - You can look up the Butcher tableau from Web (too many numbers…)



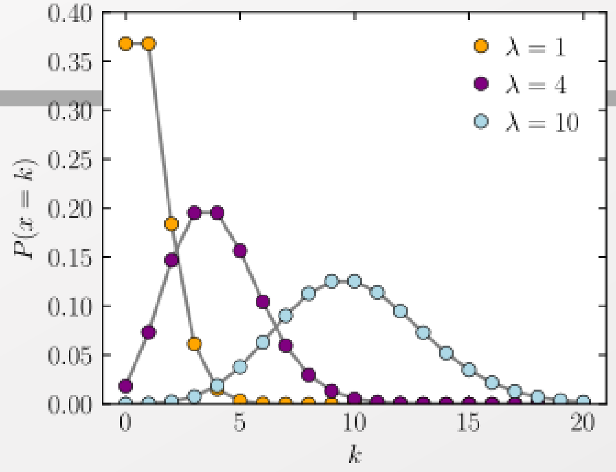
STOCHASTIC DIFFERENTIAL EQUATIONS

Lévy Process

- Stochastic Process with the below conditions
  - $X = \{X_t : t \geq 0\}$
  - $X_0 = 0$
  - Independence of increments
    - $0 \leq t_1 < t_2 < \cdots < t_n < \infty \rightarrow X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent
      - The increments between two random variables with neighbor index values are independent
  - Stationary increments
    - For any $s < t, X_t - X_s$ is equal in distribution to X_{t-s}
      - The distribution of the increment only depends upon the length between two indexes
  - Continuity in Probability
    - For any $\varepsilon > 0$ and $t \geq 0$, $\lim_{h\to 0} P(|X_{t+h} - X_t| > \varepsilon) = 0$
- Instantiation of Levy Process
  - Wiener process: $X_t - X_s \sim N(0, t-s)$
  - Poisson process: $X_t - X_s \sim Poisson(\lambda(t-s))$
  - Gamma process: $X_t - X_s \sim Gamma(v(t) - v(s), u)$

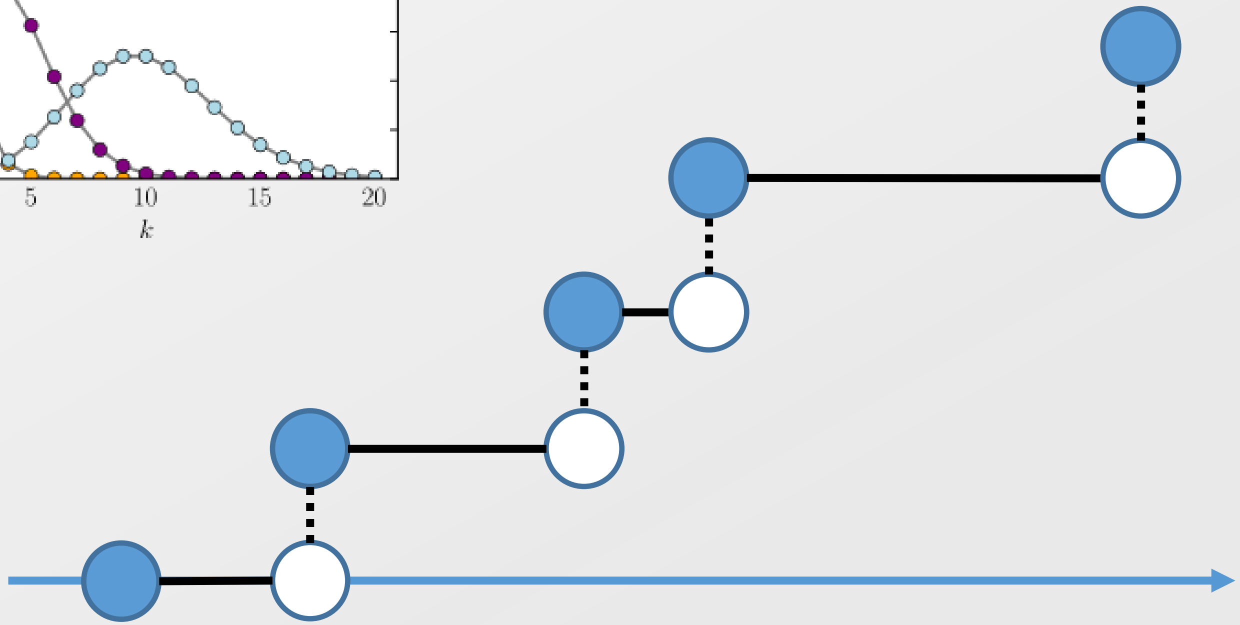


Poisson Process

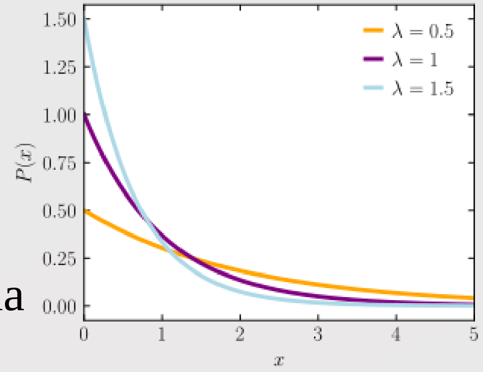


©wikipedia

- Poisson distribution
  - $Poisson(\lambda)$
  - $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$
    - $k \in \mathbb{N}_0$: Natural number with 0
    - $\lambda \in (0, \infty)$
    - Mean= Variance=λ
- Poisson process: $P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$
  - $X_0 = 0$
  - Independence of increments
  - Stationary increments
    - $X_{t-s} \sim Poisson(\lambda(t-s))$
- Some properties
  - Counting process with arrival rate (λ)
  - Memoryless property : $P(X > t + x) = P(X > x)P(X > t)$
  - Inter-arrival of event follows exponential distribution : $Exp(\lambda)$



X_0=0

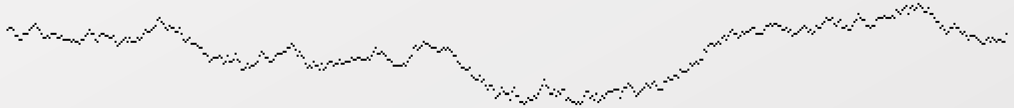


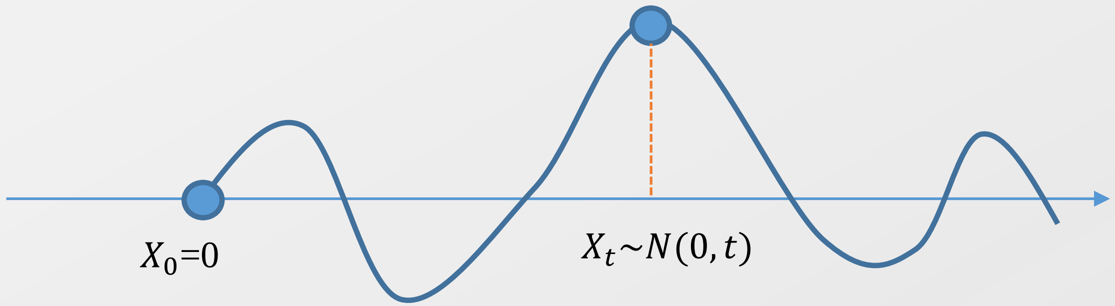
©wikipedia

Wiener Process a.k.a. Brownian Motion

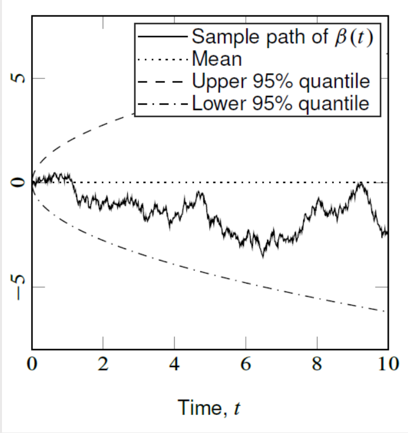
- Wiener process : (One-dimensional case) $P(X_t = x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right)$
  - $X_0 = 0$
  - Independence of increments
    - For every $t > 0$, $(X_{t+u} - X_t) \perp X_s$ if $s \leq t, u \geq 0$
  - Stationary increments
    - $X_{t+u} - X_t \sim \mathbb{N}(0, u)$
- Some properties
  - $\mathbb{E}[X_t] = 0$
  - $Var[X_t] = t$
  - $cov(X_{t_1}, X_{t_2}) = t_1$
    - $= \mathbb{E}[(X_{t_1} - \mathbb{E}[X_{t_1}]) \cdot (X_{t_2} - \mathbb{E}[X_{t_2}])] = \mathbb{E}[X_{t_1} \cdot X_{t_2}] = \mathbb{E}\left[X_{t_1} \cdot \left((X_{t_2} - X_{t_1}) + X_{t_1}\right)\right]$
    - $= \mathbb{E}[X_{t_1} \cdot (X_{t_2} - X_{t_1})] + \mathbb{E}[X_{t_1} \cdot X_{t_1}] = \mathbb{E}[X_{t_1} \cdot X_{t_1}] = t_1$ (using the independence of increment)
    - $X_{t_2} = X_{t_1} + \sqrt{t_2 - t_1} \cdot Z$, $Z \sim N(0,1)$
  - Self-similarity by scaling: $V_t = \frac{1}{\sqrt{c}} W_{ct}, c > 0$

Self-Similarity ©wikipedia





©Applied Stochastic Differential Equations



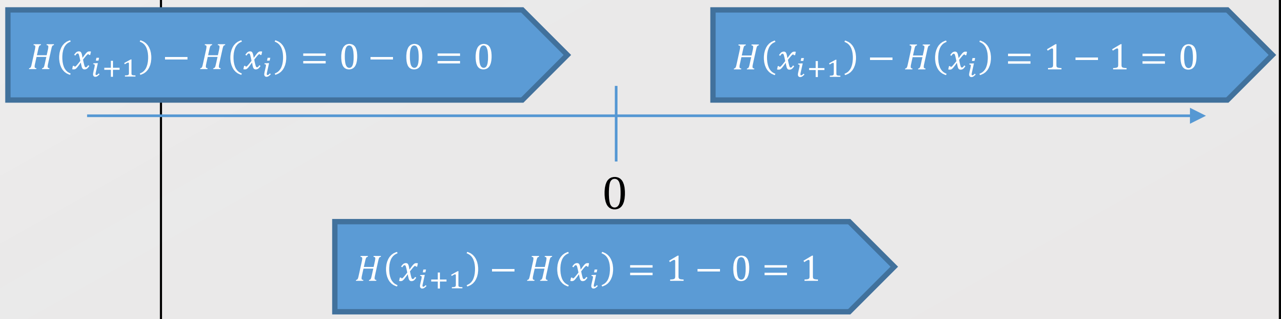
Riemann-Stieltjes Integral

- A generalization of the Riemann integral, $\int_a^b f(x)dx$

$$A = \int_{x=a}^{b} f(x)dg(x)$$

  - f: a real-valued function : integrand
  - g: a real-to-real function : integrator
    - Should be bounded-variation
  - Partition: $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$
    - $S(P, f, g) = \sum_{i=0}^{n-1} f(c_i)|g(x_{i+1}) - g(x_i)|$, $c_i \in [x_i, x_{i+1}]$
    - $\int_{x=a}^{b} f(x)dg(x) = \lim_{\sup|x_{i+1}-x_i|\to 0} \sum_{i=0}^{n-1} f(c_i)|g(x_{i+1}) - g(x_i)|$
  - If g is continuously differentiable over $\mathbb{R}$
    - $\int_a^b f(x)dg(x) = \int_a^b f(x)g'(x)dx$
- Application to probability calculation
  - $\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)g'(x)dx$
    - $g(x)$: CDF of X
    - This does not work if g is not differentiable, i.e. discrete density
  - $\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)dg(x)$

- Example) $\int_{-1}^{1} f(x)dH(x)$
  - f is continuous at 0
  - When $H(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$
- Answer)
  - $\int_{-1}^{1} f(x)dH(x) =$
    $\lim_{\sup|x_{i+1}-x_i|\to 0} \sum_{i=0}^{n-1} f(c_i)|H(x_{i+1}) - H(x_i)|$
  - $P = \{-1 = x_0 < x_1 < \cdots < x_n = 1\}$



  - $\int_{-1}^{1} f(x)dH(x) = f(0) \times 1 = f(0)$

Stochastic Integral: Ito vs Stratonovich

- Diverse integrators
  - Integrator of a deterministic variable: $\int_a^b f(x)dx$
  - Integrator of a deterministic function: $\int_{x=a}^b f(x)dg(x)$
  - Integrator of a stochastic process: $\int_{t=0}^T X_t dX_t$
- Any distinctive character of stochastic process integrator?
  - Total variation: $V_a^b(f) = \sup_{p\in\mathcal{P}} \sum_{i=0}^{n_p-1} |f(x_{i+1}) - f(x_i)|$

    $\mathcal{P} = \{P = \{x_0, \dots, x_{n_p}\} | P \text{ is a partition of } [a,b], x_i \le x_{i+1}, 0 \le i \le n_p - 1\}$
    - Deterministic integrator: Bounded total variation
      - Bounded to be zero: Riemann integral
      - Bounded to be a positive constant: Riemann-Stieltjes integral
    - Stochastic integrator: Unbounded total variation
      - A potential of large deviation between $f(x_{i+1})$ and $f(x_i)$
- $\int_{t=0}^T X_t dX_t = \lim_{n\to\infty} \sum_{k=1}^n X_{t_{k'}}\left(X_{t_k} - X_{t_{k-1}}\right), 0 = t_0 < t_1 \dots < t_n = T, t_{k-1} \le t_{k'} \le t_k$
  - Ito's definition: $\sum_{k=1}^n X_{t_{k-1}}\left(X_{t_k} - X_{t_{k-1}}\right)$
  - Stratonovich's definition: $\sum_{k=1}^n \left(\frac{X_{t_{k-1}} + X_{t_k}}{2}\right)\left(X_{t_k} - X_{t_{k-1}}\right)$

- e.g.) Wiener process
  - $X_0 = 0$
  - Independence of increments
    - For every $t > 0$, $(X_{t+u} - X_t) \perp X_s$ if $s \le t, u \ge 0$
  - Stationary increments
    - $X_{t+u} - X_t \sim N(0, u)$

Detour) Definition, Existence and Rules of Integral

- Definition of integral → Expanding the integrator from a deterministic variable to a stochastic process
  - Depending on the limit of summation over partitions
- Existence of integral
  - The convergence between the lower sum and the upper sum
- Our perception of integral
  - A list of derivation rules memorized since high school
    - $y = f(x) \to dy = f'(x)dx$
  - Chain rule
    - Single variable: when $h(x) = f(g(x))$, $h'(x) = f'\big(g(x)\big)g'(x), \frac{dh}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$
    - Multi-variable: $\frac{d}{dx} f\big(g_1(x), \dots, g_k(x)\big) = \sum_{i=1}^{k} \left(\frac{d}{dx} g_i(x)\right) \frac{\partial f}{\partial g_i}\big(g_1(x), \dots, g_k(x)\big)$
  - Integral by substitution
    - Single variable: $\int_a^b f\big(\psi(x)\big)\psi'(x)dx = \int_{\psi(a)}^{\psi(b)} f(u)du$
      - $\psi: [a, b] \to I$: differentiable function with continuous derivative, $I \subset R$: interval, $f: I \to R$: a continuous function
    - Multi-variable: $\int_U f(\psi(\mathrm{u})) |\det(D\psi)(\mathrm{u})| d\mathrm{u} = \int_{\psi(U)} f(\mathrm{v})d\mathrm{v}$
      - with similar constraints like the single variable case

Ito's Integration

- When $X_t = B_t$, $\int_{t=0}^{T} X_t dX_t = \lim_{n\to\infty} \sum_{k=1}^{n} X_{t_{k'}}\left(X_{t_k} - X_{t_{k-1}}\right) = \begin{cases} -\frac{1}{2}\boldsymbol{t} + \frac{1}{2}\boldsymbol{X_t^2}, \text{if } X_{t_{k'}} = X_{t_{k-1}}, (\text{ito}) \\ \frac{1}{2}\boldsymbol{t} + \frac{1}{2}\boldsymbol{X_t^2}, \text{if } X_{t_{k'}} = X_{t_k} \\ \frac{1}{2}\boldsymbol{X_t^2}, \text{if } X_{t_{k'}} = \frac{X_{t_{k-1}} + X_{t_k}}{2}, (\text{Stratonovich}) \end{cases}$
  - $0 = t_0 < t_1 \ldots < t_n = T, t_{k-1} \le t_{k'} \le t_k$
- Characteristics of Ito's integration
  - Left point → Non-anticipative
  - Independent increments → Martingale, Martingale $\nrightarrow$ Independent increments
  - However, Ito's integration requires different differentiations rules
    - $\int_{t=0}^{T} X_t dX_t = -\frac{1}{2}\boldsymbol{t} + \frac{1}{2}\boldsymbol{X_t^2} \neq \frac{1}{2}\boldsymbol{X_t^2}$
- $I(f) = \int_0^T f(t, X_t) dX_t \approx I(S) = \sum_{i=0}^{n} c_i (X_{t_i} - X_{t_{i-1}})$
  - c_i is measurable at t_{i-1}
  - Ito's integral with a simple function
    - Simple function : $f(x) = \sum_{k=1}^{n} a_k 1_{A_k}(x)$, 1_A is an indicator function, A_k is a disjoint measurable set, $a_k \in R$
  - Using the mean square convergence to check the existence of the Ito integral

KAIST

Ito's Lemma

- Observation of inner product on vectors and stochastic processes
  - $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$
    - $< x, y >= \sum_{i=1}^{n} x_i y_i$
    - $||x - y|| = \sqrt{< x - y, x - y >} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
  - X, Y : two stochastic processes
    - $< X, Y > = E[XY]$
    - $||x - y|| = \sqrt{E[|X - Y|^2]}$

- Our perception of integral : a list of derivation rules
  - Previously, $dy = f'(x)dx$
- Then, our perception of the Ito integral == a list of derivation rules
  - (Ito's Lemma) When $f(t, X_t), df = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial X_t}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}(t, X_t) \cdot (dX_t)^2$
    - X_t: 1-dimensional semi-martingale (which goes over this lecture)
    - $dX_t = udt + vdB_t$
    - $f: R_{\geq 0} \times R \rightarrow R$: twice continuously differentiable function
    - $(dX_t)^2 = (dX_t) \cdot (dX_t)$ is computed according to the rules
      - $dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, dB_t \cdot dB_t = dt$
  - (Integration by part) $X_t Y_t = X_0 Y_0 + \int_0^T X_{t-} dY_t + \int_0^T Y_{t-} dX_t + [X, Y]_t$
    - $[X, Y]_t$: Quadratic covariance process
  - (Ito isometry) $\mathbb{E}\left[\left(\int_0^T H_t dB_t\right)^2\right] = \mathbb{E}\left[\int_0^T H_t^2 dt\right]$
    - H : deterministic sequence or a function, left-continuous, locally bounded
    - B : Wiener process, a.k.a. Brownian motion

Ito Process and Geometric Brownian Motion

- Ito Process
  - Deterministic drift over time + Stochastic diffusion over time
  - $\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t$
    - $X_t = X_0 + \int_0^t \mu_t dt + \int_0^t \sigma_t dB_t$
    - Assume $X_0 = 0$, B : Wiener process (Brownian motion)
  - $\mathbb{E}[X_t] = \int_0^t \mu_t dt$, $Var[X_t] = \int_0^t \sigma_t^2 dt$ by Ito's isometry
- Geometric Brownian Motion
  - Assuming constant volatility σ, constant drift μ
  - $\mathrm{d}X_t = \mu \mathrm{X_t} \mathrm{d}t + \sigma \mathrm{X_t} \mathrm{d}B_t$
    - Assuming $f(X_t) = \log X_t$
    - By (Ito's Lemma) $df(X_t) = \sum_{i=1}^n f_i(X_t) dX_t^i + \frac{1}{2}\sum_{i,j=1}^n f_{i,j}(X_t) d\left[X^i, X^j\right]_t$
    - $df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{(X_t)^2}(X_t)^2\sigma^2 dt = \frac{1}{X_t}(\mu \mathrm{X_t}\mathrm{d}t + \sigma \mathrm{X_t} \mathrm{d}B_t) - \frac{1}{2}\sigma^2 dt$
    $$= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma \mathrm{d}B_t$$
    - $f(X_t) = \log X_t = \log X_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t$
  - $X_t = X_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right)$: closed-form estimation of geometric Brownian motion at any index of t
    - $X_t \sim N\left(X_0 \exp(\mu t), X_0^2 \exp(\mu t)\left(\exp(\sigma^2 t) - 1\right)\right)$