Orals Notes, Applied Stochastic Analysis

Sonia Reilly

October 15, 2023

These notes summarize the content of Miranda Holmes-Cerfon's lecture notes for her Spring 2022 Applied Stochastic Analysis class.

Lecture 1 - Probability Review

- The covariance is $\text{cov}(X, X) = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T$, a matrix with components $\text{cov}(X_1, X_j) = \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)^T$. If X, Y independent, $\text{cov}(X, Y) = 0$, but the converse is not true.
- The marginal pdf of a component X_i of X is the integral of the joint pdf over all variables except x_i.
- Conditional probability: $P(A|B) = P(A \cap B)/P(B)$.
- Law of Total Probability: $P(A) = \sum P(A \cap B_i) = \sum P(A|B_i)P(B_i)$.
- Bernoulli: $f(0) = 1 - p, \quad f(1) = p, \quad \mathbb{E}X = p, \quad \text{var}(X) = p(1 - p)$.
- Binomial: $f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \mathbb{E}X = np, \quad \text{var}(X) = np(1 - p)$.
- Poisson: $f(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \mathbb{E}X = \lambda, \text{var}(X) = \lambda$.
- Exponential: $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \mathbb{E}X = \lambda, \quad \text{var}(X) = \lambda$.
- Gaussian:
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad \mathbb{E}X = \mu, \text{var}(X) = \sigma^2.$$
- Multivariate Gaussian:
$$f(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

Lecture 2 – Discrete-time Markov Chains

- A *discrete-time Markov chain* satisfies the *Markov property*:
$$P(X_{n+1} = s | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = s | X_n = x_n).$$
- The *transition matrix*, $P_{ij}(n) = P(X_{n+1} = j | X_n = i)$, is a stochastic matrix, i.e., it has nonzero entries and its rows sum to 1.
- A Markov chain is *time-homogeneous* if P does not depend on the time n.

- *Forward Kolmogorov equation* for a time-homogeneous, discrete-time Markov chain:

$$\alpha^{(n+1)} = \alpha^{(n)} P,$$

where $\alpha^{(n)}$ is the probability distribution of X_n as a row vector. [Proof: law of total probability.]

- *Backward Kolmogorov equation* for a time-homogeneous, discrete-time Markov chain:

$$u^{(n+1)} = Pu^{(n)}, \quad u_i^{(0)} = f(i) \ \forall i \in S,$$

where $u_i^{(n)} = \mathbb{E}[f(X_n)|X_0 = i]$, a column vector of the expected value of (some function of) the state of the Markov chain. [Proof: definition of expectation, LoTP, Markov property, manipulating sums]

- *Chapman-Kolmogorov equation* (all Markov processes):

$$P(X_n = j|X_0 = i) = \sum_k P(X_n = j|X_m = k)P(X_m = k|X_0 = i)$$

- *Forward Kolmogorov equation* (general Markov chain):

$$\alpha^{(t+1)} = \alpha^{(t)} P(t)$$

[Proof: Chapman-Kolmogorov expanding $P(j, t+1|i, s) = P(X_{t+1} = j|X_s = i)$ in terms of t, evolving the transition matrix forward in time.]

- *Backward Kolmogorov equation* (general Markov chain):

$$u^{(s)} = P(s)u^{(s+1)}$$

[Proof: Chapman-Kolmogorov expanding $P(j, t|i, s)$ in terms of $s + 1$.]

- A *limiting distribution* of a time-homogeneous MC is a distribution λ that satisfies $\lim_{n\to\infty}(P^n)_{ij} = \lambda_j$. If it exists, it is unique (otherwise, either the limit doesn't exist or it doesn't converge to a distribution, e.g. in a random walk on an infinite line, it may converge to 0 everywhere.)

- A *stationary distribution* π satisfies

$$\pi = \pi P \quad \iff \quad \pi_j = \sum_i \pi_i P_{ij}$$

- The limiting distribution is always a stationary distribution.

- A stochastic matrix P is *irreducible* if for every (i, j) there is an $s > 0$ such that $(P^s)_{ij} > 0$ (i.e., every state can eventually be reached from every other state).

- An irreducible Markov chain over a finite state space has a unique stationary distribution. Over an infinite space, there is a unique stationary distribution if the chain is irreducible and has finite mean first passage times for every state. The stationary distribution is limiting if the chain is also aperiodic.

- The *mean first passage time* τ_i is the expected time for an event A to occur when the chain starts at i. The vector τ is the minimal nonnegative solution to

$$\begin{cases} \tau_i = 0 & i \in A \\ \tau_i = 1 + \sum_j P_{ij}\tau_j & i \notin A. \end{cases}$$

- The *hitting probability* h_i of A is the probability of X_t reaching the event A in finite time if it starts at i. It is the minimal nonnegative solution to

$$\begin{cases} h_i = 1 & i \in A \\ h_i = \sum_j P_{ij}h_j & i \notin A. \end{cases}$$

Lecture 3 – Detailed Balance and MCMC

- A Markov chain X_t satisfies *detailed balance* (aka is reversible with respect to) a stationary distribution π if it satisfies the detailed balance equations:

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \forall i, j$$

  Detailed balance requires that the flow of probability along each edge in each time step be equal in both directions.

- A distribution that satisfies the detailed balance equations must be a stationary distribution.

- Metropolis-Hastings: proposal probability $h(y|x)$, acceptance probability

$$a(y|x) = \min\left(1, \frac{\pi(y)h(x|y)}{\pi(x)h(y|x)}\right).$$

  The resulting Markov chain has the stationary distribution π. [Proof: plug $P_{ij} = H_{ij}a_{ij}$ for $i \neq j$ into the detailed balance equations, split into two cases to handle the minimum.]

Lecture 4 – Continuous-time Markov Chains

- *Chapman-Kolmogorov equation* (continuous-time, time-homogeneous):

$$P(t+s) = P(t)P(s) \quad \iff \quad P_{ij}(t+s) = \sum_{k\in S} P_{ik}(t)P_{kj}(s),$$

  where $P_{ij}(t) = P(X_{s+t} = j|X_s = i)$. [Proof: split up $P_{ij}(s+t)$ using LoTP, then Markov property and time-homogeneity.]

- The generator is the matrix

$$Q = \lim_{h\to 0^+} \frac{P(h) - I}{h}.$$

  (Rows sum to 0, diagonals are negative or zero, off-diagonals are positive or zero. Interpret q_{ij} as a rate of jumping from i to j, and $-q_{ii}$ as a rate of leaving i.) Equivalently,

$$P_{ij}(h) = \delta_{ij} + q_{ij}h + o(h).$$

- E.g., a Poisson process has a generator with $-\lambda$ on every diagonal and λ on every entry directly above the diagonal.

- Alternatively, we can view the continuous chain as a discrete *jump chain* Y^n with *jump times* J_m and *holding times* $S_m = J_m - J_{m-1}$.

- A *stopping time* for a continuous-time process X is any nonnegative random variable T such that for any $t \geq 0$, the event $T \leq t$ depends only on $(X_s : s \leq t)$. Jump times are stopping times.

- *Strong Markov property*: If we stop X at a stopping time T, the rest of the chain after T is also a continuous Markov chain with the same transition probabilities.

- The holding time S_m is an exponentially distributed random variable with parameter $-q_{ii}$, where $X_{J_{m-1}} = i$. [Proof: show that $P(S_m > r + t | S_m > r, X_{J_{m-1}} = i) = P(S_m > t | X_{J_{m-1}} = i)$ using strong Markov property, because the exponential distribution is the only one with this lack-of-memory property.]

- The jump/embedded chain Y_m has a transition matrix $\tilde{P}$ with $\tilde{P}_{ii} = 0$ if $q_{ii} \neq 0$, $\tilde{P}_{ii} = 1$ if $q_{ii} = 0$, and $\tilde{P}_{ij} = -q_{ij}/q_{ii}$ for $i \neq j$.

- *Gillespie algorithm* (aka stochastic simulation, Kinetic Monte Carlo): choose S_m from exponential distribution, choose Y_m from $\tilde{P}$, repeat.

- *Forward Kolmogorov equation*: The transition probability evolves as
$$\frac{dP}{dt} = PQ, \quad P(0) = I.$$
[Proof: definition of derivative, Chapman-Kolmogorov, definition of Q.] Equivalently, since $\mu(t) = \mu(0)P(t)$,
$$\frac{d\mu}{dt} = \mu Q.$$

- *Backward Kolmogorov equation*: The transition probability evolves as
$$\frac{dP}{dt} = QP, \quad P(0) = I.$$
[Proof: same as for forward, but factor out $P(t)$ on the right instead.] Equivalently, since $u(t) = P(t)u(0)$,
$$\frac{du}{dt} = Qu.$$

- Solving the Kolmogorov equations gives
$$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{1}{n!} Q^n t^n.$$

- A distribution π is a stationary distribution if and only if $\pi Q = 0$. [Proof: $\pi = \pi P(t) \implies \pi P'(t) = 0 \implies \pi P(t) Q = 0 \implies \pi Q = 0$.]

- The detailed balance equations for ctMCs are $\pi_i q_{ij} = \pi_j q_{ji}$. Stationary distributions are limiting when the chain is irreducible (equivalent to the embedded chain Y_m being irreducible). [Example of non-irreducible chain: Poisson process.]

- The mean first passage time (mfpt) satisfies

$$\begin{cases} \tau_i = 0 & i \in A \\ 1 + \sum_j Q_{ij}\tau_j = 0 & i \notin A. \end{cases}$$

Lecture 5 – Gaussian Processes and Stationary Processes

- A *finite dimensional distribution* (fdd) of a stochastic process X_t is the joint probability distribution of the values of X_t at a finite number of times t_n.
- The mean $m(t)$ and covariance function $B(s,t) = \mathbb{E}X_sX_t - m(s)m(t)$ are the one-point and two-point fdds.
- The covariance function is positive semidefinite (i.e., the corresponding matrix of $B(t_i, t_j)$'s for any finite number of times t_i is psd). For a complex-valued stochastic process, it is Hermitian ($B(s,t) = \overline{B(t,s)}$). [Example: for a Poisson process, $B(s,t) = \lambda \min(s,t)$.]
- A *Gaussian process* is a process whose fdds are all Gaussian.
- *Strongly stationary*: all fdds are invariant with shifts in time. Example: Markov chains in a stationary distribution are strongly stationary.
- *Weakly stationary*: mean and covariance are invariant with shifts in time, i.e., $m(t) = \text{const.}$, $B(s,t) = C(s-t)$.
- Properties of the covariance function $C(t)$: the variance is $C(0)$, and $|C(t)| \le C(0)$ for all t. (Proof: Cauchy-Schwartz: $|C(t)| = |\mathbb{E}X_{t+s}X_s| \le (\mathbb{E}X_{t+s}^2 \mathbb{E}X_s^2)^{1/2} = C(0)$.) And $C(t)$ is a *positive semidefinite function*.
- The Fourier transform of a positive semidefinite function (if it exists) is always nonnegative. Conversely, we can construct a positive semidefinite function (hence a possible covariance function) by taking the inverse Fourier transform of a nonnegative function.
- Skipped: Ergodic Theorem (go back if necessary?)

Lecture 6 – Brownian Motion

- A *Brownian motion* or *Wiener process* W_t is a stochastic process with
  - $W_0 = 0$
  - Independent increments: $W_v - W_u$ is independent from $W_t - W_s$ whenever (u,v) and (s,t) are disjoint intervals
  - Normal increments: $W_{s+t} - W_s \sim N(0,t)$
  - Continuous sample paths: the function $t \to W_t$ is continuous with probability 1.
- Alternately, Brownian motion can be defined as a Gaussian process that is continuous with probability 1 and has $W_0 = 0$, $m(t) = 0$, $B(s,t) = \min(s,t)$.

- A discrete approximation to W_t is the random walk $S_{nt} = \frac{1}{\sqrt{n}} \sum_{j=1}^{nt} X_j$, for $X_j = \pm 1$ at each step. This is a random walk on the integers rescaled with $\Delta t = 1/n$ and $\Delta x = 1/\sqrt{n}$ to get the correct variance. We interpolate linearly between integer values of nt to create a continuous approximation.

- Properties of Brownian motion:
  - Symmetry: $-W_t$ is a Brownian motion
  - Translation: $W_{t+s} - W_s$ is a Brownian motion
  - Scaling: $\frac{1}{\sqrt{c}} W_{ct}$ is a Brownian motion
  - Time-inversion: $tW_{1/t}$ is a Brownian motion

- Brownian motion W_t is bounded in its growth between $\sqrt{t}$ and t – specifically the *Law of Iterated Logarithms* says that

$$\limsup_{t \to 0^+} \frac{W_t}{\sqrt{2t \log \log 1/t}} = 1 \quad a.s.$$

- With probability 1, sample paths of a Brownian motion are not Lipschitz continuous, and hence not differentiable. Basic reason:

$$\xi_t = \frac{dW_t}{dt} = \lim_{h \to 0} \frac{W_{t+h} - W_t}{h},$$

  but $W_{t+h} - W_t \sim N(0, h)$, so $\frac{W_{t+h} - W_t}{h} \sim N(0, \frac{1}{h})$. Hence the limit does not converge, since it is a Gaussian with a variance that blows up to infinity.

- We can however calculate the mean and covariance of the process that would be the derivative if it existed, which we call white noise. White noise has mean 0 and covariance function $\delta(t)$. The spectrum is constant, since the Fourier transform of the delta function is constant.

- Brownian motion has infinite total variation since it is nowhere differentiable. The quadratic variation of Brownian motion converges to t in mean-square as $\max |t_{i+1} - t_i| \to 0$:

$$Q_t^{\sigma}(W) = \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|^2 \to t$$

  in the sense that $X_n \to X$ in mean-square iff $\mathbb{E}|X_n - X|^2 \to 0$ as $n \to \infty$. (A stricter sense than convergence in probability or distribution, not as strict as convergence almost surely.)

- Brownian motion is continuous-time Markov process on a continuous state space, so we define a transition density $p(y, t|x, s)$ s.t.

$$P(X_t \in A | X_s = x) = \int_{x \in A} p(y, t|x, s) dx,$$

  which for Brownian motion is a Gaussian density. In general integration against p replaces multiplication by P.

- The infinitesimal generator of a Markov process is the operator $(\mathcal{L}f)(x) = \lim_{t \to 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t}$. For Brownian motion the operator is the Laplacian, $\mathcal{L}f = \frac{1}{2} \frac{d^2 f}{dx^2}$. Then the backward Kolmogorov equation becomes the heat equation in $u(x, t) = \mathbb{E}_x f(X_t)$, and the forward Kolmogorov equation becomes the heat equation in $p(x, t|y, 0)$, since the operator is self-adjoint.

Lecture 7 – Stochastic Integration

- A general SDE has the form

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\eta(t),$$

where $\eta(t)$ is a white noise, so that we can integrate to get

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

- The *Itô integral* is the mean-square limit of the left-hand Riemann sum. For a partition $0 = t_1, \ldots, t_n = t$, the Itô integral is

$$\int_0^t f(s, \boldsymbol{\omega})dW_s = \lim_{\max_j |\Delta t_j| \to 0} \sum_{j=0}^{n-1} f(t_j, \boldsymbol{\omega})\Delta W_j.$$

- The *Stratonovich integral* is the mean-square limit of the trapezoidal Riemann sum:

$$\int_0^t f(s, \boldsymbol{\omega}) \circ dW_s = \lim_{\max_j |\Delta t_j| \to 0} \sum_{j=0}^{n-1} \frac{f(t_j, \boldsymbol{\omega}) + f(t_{j+1}, \boldsymbol{\omega})}{2}\Delta W_j.$$

- Example:

$$\begin{aligned}
\int_0^t W_s dW_s &= \lim_{\max_j |\Delta t_j| \to 0} \sum_{j=0}^{n-1} W_j \Delta W_j \\
&= \lim_{\max_j |\Delta t_j| \to 0} \left[\frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}^2 - W_j^2) - \frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1} - W_j)^2\right] \\
&= \lim_{\max_j |\Delta t_j| \to 0} \left[\frac{1}{2}(W_t^2 - W_0^2) - \frac{1}{2}Q_t^\sigma(W)\right] \\
&= \frac{1}{2}W_t^2 - \frac{1}{2}t.
\end{aligned}$$

By a similar calculation,

$$\int_0^t W_s \circ dW_s = \frac{1}{2}W_t^2.$$

- Skipped: Formal proof that the Itô integral exists.
- Properties of the Itô integral:
  - Linearity in f
  - Nonanticipating property: $\mathbb{E}\int_0^\infty f dW_t = 0$.
  - Itô isometry:

$$\mathbb{E}\left(\int_0^\infty f(t, \boldsymbol{\omega})dW_t\right)^2 = \mathbb{E}\int_0^\infty f^2(t, \boldsymbol{\omega})dt.$$

– General Itô isometry:

$$\mathbb{E}\left(\int_0^t g(s,\boldsymbol{\omega} dW_s \int_0^t h(s,\boldsymbol{\omega})dW_s\right) = \int_0^t \mathbb{E}[g(s,\boldsymbol{\omega})h(s,\boldsymbol{\omega})]ds.$$

- The *Itô formula* is the chain rule for Itô processes (solutions to SDEs). Let X_t be the solution to

$$dX_t = b(t,\boldsymbol{\omega})dt + \sigma(t,\boldsymbol{\omega})dW_t,$$

where b, σ are adapted functions (depend only on $\boldsymbol{\omega}$ before t). The process $Y_t = g(t, X_t)$ solves the equation

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dX_t)^2,$$

where $(dX_t)^2$ is computed according to the rules

$$dt \cdot dt = dt \cdot dW_t = dW_t \cdot t = 0, \quad dW_t \cdot dW_t = dt,$$

so that

$$dY_t = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}b + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\sigma^2\right)dt + \frac{\partial g}{\partial x}\sigma dW_t.$$

The second-order term is the drift term, introduced by Taylor expanding ΔY to $O(\Delta t)$. Since $\Delta W \sim O(\Delta t^{1/2})$, we go to second order in x to keep terms that are first order in t.

- In higher dimensions, if W_t is a vector of n independent Brownian motions and $Y_t = f(X_t)$,

$$dY_t = \nabla f(X_t) \cdot dX_t + \frac{1}{2}(dX_t)^T \nabla^2 f(X_t) dX_t,$$

which reduces to

$$dY_t = \left(b \cdot f + \frac{1}{2}\sigma\sigma^T : \nabla^2 f\right)dt + (\nabla f)^T \sigma dW_t$$

with the additional fact that $dW_t^{(i)} \cdot dW_t^{(j)} = 0$ for $i \neq j$. Here $A : B = Tr(A^T B) = \sum_{ij} A_{ij}B_{ij}$.

Lecture 8 – Stochastic Differential Equations

- The drift term $b(X_t)$ describes the deterministic part of the equation and the diffusion term $\sigma(X_t)$ describes random motion proportional to Brownian motion. The noise is additive if $\frac{\partial \sigma}{\partial x} = 0$, otherwise multiplicative.

- *Ornstein-Uhlenbeck process* (OU):

$$dX_t = -aX_t dt + \sigma dW_t, \quad X_0 = \xi,$$

with $a, \sigma \in \mathbb{R}$ constants and ξ independent of W_t.

- To solve, multiply both sides by e^{at} and integrate from 0 to t:

$$\underbrace{e^{at}dX_t + ae^{at}X_t dt}_{=d(e^{at}X_t)} = \sigma e^{at}dW_t \implies e^{at}X_t - X_0 = \int_0^t \sigma e^{as}dW_s,$$

which gives solution

$$X_t = e^{-at}X_0 + \sigma\int_0^t e^{-a(t-s)}dW_s.$$

- The initial condition is "forgotten" exponentially quickly. The mean is $\mathbb{E}X_t = e^{-at}\mathbb{E}X_0$. The stationary distribution is $N(0, \frac{\sigma^2}{2a})$, and in this stationary distribution the process has mean and covariance
$$\mathbb{E}X_t = 0, \qquad B(s,t) = \frac{\sigma^2}{2a}e^{-a|s-t|}.$$
This is the only process that is stationary, Markov, Gaussian, and has continuous paths. It is a linearization of many SDEs.

- *Geometric Brownian Motion* (GBM):
$$dN_t = rN_t dt + \alpha N_t dW_t, \qquad N_0 = \xi.$$
Models a process with average growth rate r and fluctuations of magnitude α in the growth rate. Multiplicative noise as opposed to additive noise of OU.

- To solve, notice by Itô's formula
$$d(\log N_t) = \frac{1}{N_t}dN_t - \frac{(dN_t)^2}{2N_t^2} = \frac{1}{N_t}dN_t - \frac{\alpha^2}{2}dt.$$
Plugging in dN_t from the GBM equation gives
$$d(\log N_t) = (r - \frac{\alpha^2}{2})dt + \alpha dW_t,$$
so
$$N_t = N_0 e^{(r-\frac{\alpha^2}{2})t + \alpha W_t}.$$

- Using the GBM equation in integral form and the nonanticipating property, we get $\mathbb{E}N_t = (\mathbb{E}N_0)e^{rt}$. If $r > \alpha^2/2$, then $N_t \to \infty$ a.s. as $t \to \infty$. If $r < \alpha^2/2$, then $N \to 0$ a.s. but $\mathbb{E}N_t \to \infty$, because increasingly large and increasingly rare oscillations dominate the expectation. If $r = \alpha^2/2$, N_t oscillates between arbitrarily high and arbitrarily close to 0.

- Itô-Stratonovich conversion:
$$dX_t = b\,dt + \sigma \circ dW_t \implies dX_t = \left(b + \frac{1}{2}\sigma\frac{\partial \sigma}{\partial x}\right)dt + \sigma dW_t$$
and
$$dX_t = b\,dt + \sigma dW_t \implies dX_t = \left(b - \frac{1}{2}\sigma\frac{\partial \sigma}{\partial x}\right)dt + \sigma \circ dW_t.$$

- The Stratonovich integral satisfies the regular chain rule, $df(X_t) = f'(X_t) \circ dX_t$, but does not satisfy the Itô isometry or the non-anticipating property.

Lecture 9 – Numerically Solving SDEs

- *Euler-Maruyama*: Approximate X_t with $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ by Y_n with
$$Y_{n+1} = Y_n + b(Y_n)\Delta t + \sigma(Y_n)\delta W_n, \qquad \delta W_n \sim N(0, \Delta t) \quad \text{i.i.d.}$$

- To derive higher-order methods, we use the *stochastic Itô-Taylor expansion*, based on an integral form of a Taylor series. Given X_t satisfying the standard SDE, we can apply Itô's formula to find $df(X_t)$ and integrate to get

$$f(X_t) = f(X_0) + \int_0^t \mathcal{L}_0 f(X_s)ds + \int_0^t \mathcal{L}_1 f(X_s)dW_s,$$

where

$$\mathcal{L}_0 = b(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}, \qquad \mathcal{L}_1 = \sigma(x)\frac{\partial}{\partial x}.$$

Applying this to b, σ, and $\mathcal{L}_1\sigma$ in the integral form of the SDE gives

$$\begin{aligned} X_t &= X_0 + b(X_0)\int_0^t ds + \sigma(X_0)\int_0^t dW_s + O(t) \\ &= X_0 + b(X_0)\int_0^t ds + \sigma(X_0)\int_0^t dW_s + \mathcal{L}_1\sigma(X_0)\int_0^t\int_0^s dW_z dW_s + O(t^{3/2}) \\ &= X_0 + b(X_0)\int_0^t ds + \sigma(X_0)\int_0^t dW_s + \mathcal{L}_1\sigma(X_0)\left[\frac{1}{2}(W_t^2 - t)\right] + O(t^{3/2}). \end{aligned}$$

- The result is the *Milstein* scheme:

$$Y_{n+1} = Y_n + b(Y_n)\Delta t + \sigma(Y_n)\delta W_n + \frac{1}{2}\sigma(Y_n)\sigma'(Y_n)((\delta W_n)^2 - \Delta t), \qquad \delta W_n \sim N(0, \Delta t) \quad \text{i.i.d.}$$

- $Y^{\Delta t}$ *converges strongly* to X at time T with order α if for some $C, \delta_0 > 0$ independent of Δt

$$\mathbb{E}|Y_N^{\Delta t} - X_T| \leq C(\Delta t)^\alpha \quad \forall \Delta t < \delta_0,$$

i.e., if individual realizations converge as $\Delta t \to 0$.

- $Y^{\Delta t}$ *converges weakly* to X at T with order β with respect to a class of functions $\mathcal{C}$ if for each $f \in \mathcal{C}$ there exist constants $C_f, \delta_0 > 0$ independent of Δt such that

$$|\mathbb{E}f(Y_N^{\Delta t}) - \mathbb{E}f(X_T)| \leq C_f(\Delta t)^\beta \quad \forall \Delta t < \delta_0,$$

i.e., if the distribution of the approximation converges to the distribution of the exact process.

- For Lipschitz continuous functions f, the weak order $\geq$ the strong order.

- Euler-Maruyama: Strong order 1/2, weak order 1

- Milstein: Strong order 1, weak order 1

- For additive noise, EM will converge with strong order 1 since $\sigma\sigma' = 0 \implies$ EM and Milstein are equivalent.

- For ODEs, we would typically consider a method stable in the region of $\lambda\Delta t$ where a numerical solution to $\frac{dX}{dt} = \lambda X$ near $X = 0$ limits to 0. For SDEs the corresponding linearization is Geometric Brownian Motion

$$dX_t = \lambda X_t dt + \mu X_t dW_t,$$

and we consider two notions of limiting to 0.

- The solution $X_t = 0$ is *mean-square stable* if $\lim_{t\to\infty} \mathbb{E}X_t^2 = 0$ for any X_0. (For GBM, if $\lambda + \frac{1}{2}\mu^2 < 0$.)
- The solution $X_t = 0$ is *asymptotically stable* if $P(\lim_{t\to\infty} X_t = 0) = 1$ for any X_0. (For GBM, if $\lambda - \frac{1}{2}\mu^2 < 0$, so for GBM mean-square $\implies$ asymptotic.)
- Numerical methods like Euler-Maruyama are stable in subsets of the regions of SDE stability. (Plot on $\lambda\Delta t$ and $\mu^2\Delta t$ axes.)

Lecture 10 – Forward and Backward Equations for SDEs

- Solutions to a multidimensional SDE $dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$ are Markov processes (the value $X_t - X_s$ depends only X in $[s, t]$) that satisfy the strong Markov property. The transition density of X (probability density of being at x at time t after being at y at time s) is $p(x, t|y, s)$.
- The generator of the process is

$$(\mathcal{L}f)(x) = b(x) \cdot \nabla f(x) + \left(\frac{1}{2}\sigma(x)\sigma^T(x)\right) : \nabla^2 f(x)$$

- Backward Kolmogorov equation:

$$\partial_t u = \mathcal{L}u \quad \text{for} \quad u(x, t) = \mathbb{E}[f(X_t)|X_0 = x].$$

  For a time-inhomogeneous equation, solve backwards:

$$\partial_s u(y, s) + \mathcal{L}u(y, s) = 0, \quad s < t \quad \text{for} \quad u(y, s) = \mathbb{E}[f(X_t)|X_s = y].$$

  The transition density $p(x, t|y, s)$ is the Green's function of the inhomogeneous backward equation:

$$\partial_s p + \mathcal{L}_y p = 0, \quad s < t, \quad p(x, t|y, t) = \delta(x - y).$$

- The adjoint of the generator, satisfying $\langle \mathcal{L}f, g\rangle = \langle f, \mathcal{L}^* g\rangle$, is

$$\mathcal{L}^* g(x, t) = -\nabla \cdot (b(x, t)g(x)) + \nabla \cdot \nabla \cdot (a(x, t)g(x)).$$

- Forward Kolmogorov (Fokker-Planck) equation for the probability density $\rho(x, t)$ of X_t:

$$\partial_t \rho = \mathcal{L}^* \rho, \quad t > 0$$

  Forward Kolmogorov for the transition density:

$$\partial_t p = \mathcal{L}_x^* p, \quad p(x, s|y, s) = \delta(x - y).$$

- For Brownian motion (SDE $dX_t = dW_t$), $\mathcal{L} = \mathcal{L}^* = \frac{1}{2}\partial_{xx}$, so the backward and forward equations are both the heat equation and the transition density is Gaussian.
- Physical interpretation of the Fokker-Planck equation:

$$\partial_t \rho + \nabla \cdot j = 0, \quad \text{where} \quad j = b(x, t)\rho - \nabla \cdot (a(x, t)\rho),$$

  for probability flux j. The probability is advected at speed b and diffuses with rate a.

- Forward equation boundary conditions: reflecting ($j \cdot n = 0$ on ∂D), absorbing ($\rho = 0$ on ∂D), periodic on $[a, b]$ ($j|_{b^-} = j|_{a^+}, \rho|_{b^-} = \rho|_{a^+}$), etc.
- Backward equation BCs are derived from forward BCs by ensuring that the second term disappears in the integration-by-parts expansion

$$\langle \mathcal{L}f, g \rangle = \langle f, \mathcal{L}^* g \rangle + \int_{\partial D} f j \cdot n + g(a \cdot \nabla f) \cdot n.$$

- Stationary distribution satisfies $\mathcal{L}^* \rho = 0 \iff \nabla \cdot j = 0$.