Diffusion: Kolmogorov Equations, Feynman-Kac the Martingale Problem

Math653: SDEs and applications, Chapter 8

Fei Lu

2026 Spring

Overview

Review: Itô Diffusions & Generator

Itô Diffusions & Generator

Itô diffusion

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \qquad X_0 = x,$$

where B_t is m-dimensional, Lipshitz and linear growth.

Generator: The Infinitesimal Expectation Operator

$$\mathbb{E}^x[f(X_h)] = f(x) + hAf(x) + o(h) \Leftrightarrow Af(x) = \lim_{h\to 0} \frac{\mathbb{E}^x[f(X_h)] - f(x)}{h}.$$

$$\boxed{Af = b \cdot \nabla f + \frac{1}{2} a : D^2 f.}$$

$$Af(x) = \sum_{i=1}^{d} b_i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\partial_{ij} f(x).$$

Derivation from Itô's Formula For $f \in C_b^2(\mathbb{R}^d)$, Itô's formula gives

$$df(X_t) = \nabla f(X_t)^T b(X_t)\, dt + \frac{1}{2} a(X_t) : D^2 f(X_t)\, dt + \nabla f(X_t)^T \sigma(X_t)\, dB_t.$$

Integrating from 0 to h and taking expectation,

$$\mathbb{E}^x[f(X_h)] - f(x) = \mathbb{E}^x\left[\int_0^h Af(X_s)\, ds\right].$$

Taking $h \to 0$ gives the generator formula.

Key point

The generator captures the local behavior of the diffusion by averaging over an infinitesimal time interval.

Dynkin's Formula: The Accumulated Version

Question: How do infinitesimal changes accumulate over $[0, \tau]$?

Let τ be a bounded stopping time. Integrating Itô's formula gives

$$f(X_\tau) - f(x) = \int_0^\tau Af(X_s)\, ds + \int_0^\tau \nabla f(X_s)^T \sigma(X_s)\, dB_s.$$

Dynkin's Formula

$$\boxed{\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau Af(X_s)\, ds\right].}$$

- Generator: local (infinitesimal) behavior. Dynkin: global behavior.
- Class of f: from $C_b^2(\mathbb{R}^d)$ to general C^2.
- Dynkin's formula bridges the SDE to PDEs and potential theory.

Running Example: Brownian Motion with Drift

Model

Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be constants:

$$dX_t = \mu\, dt + \sigma\, dB_t, \qquad X_0 = x,$$

Solution: $X_t = x + \mu t + \sigma B_t$.

Its generator is

$$Af(x) = \mu f'(x) + \frac{1}{2}\sigma^2 f''(x).$$

Kolmogorov's Equations

Backward Equation: What PDE Evolves Observables?

Key question

How does $u(t,x) = \mathbb{E}^x[f(X_t)]$ evolve in time?

Define the semigroup ($P_{t+s} = P_t P_s$; $P_0 = I$)

$$P_t f(x) := \mathbb{E}^x[f(X_t)].$$

Then $u(t,x) = P_t f(x)$ and $u(0,x) = f(x)$.
Formally,

$$\frac{u(t+h,x) - u(t,x)}{h} = \frac{P_t(P_h f - f)(x)}{h}.$$

By Dynkin's formula on the short interval $[0,h]$,

$$P_h f(x) - f(x) = \mathbb{E}^x\left[\int_0^h Af(X_s)\,ds\right] = hAf(x) + o(h).$$

$$\Rightarrow \qquad \boxed{\partial_t u(t,x) = Au(t,x), \qquad u(0,x) = f(x).}$$

- The backward equation evolves test functions or observables
- $\partial_t P_t f = AP_t f$; $P_t = e^{tA}$

Example: Backward Equation in 1D $dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$, then

$$Af(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).$$

Hence

$$\boxed{\partial_t u = b(x)\partial_x u + \frac{1}{2}\sigma^2(x)\partial_{xx} u, \qquad u(0,x) = f(x).}$$

For the running example $dX_t = \mu\,dt + \sigma\,dB_t$,

$$\boxed{\partial_t u = \mu\partial_x u + \frac{1}{2}\sigma^2\partial_{xx} u, \qquad u(0,x) = f(x).}$$

Note

The drift contributes the first-order transport and the diffusion contributes the second-order term.

Forward Equation: What PDE Evolves the Law?

Key question

How does the probability density of X_t evolve?

Assume X_t^x has transition density $p_t(x, y)$:

$$P_t f(x) = \int_{\mathbb{R}^d} f(y) p_t(x, y)\, dy.$$

We want

- A PDE for $p_t(x, y)$ in the variable y.
- A PDE for $\rho(t, y)$, the density of X_t, obtained by integrating over x.

Differentiate $P_t f$ in time:

$$\int f(y)\partial_t p_t(x,y)\,dy = \partial_t P_t f(x) = P_t A f(x).$$

Rewrite the right-hand side by integration by parts in y:

$$P_t A f(x) = \int A f(y) p_t(x,y)\,dy = \int f(y) A_y^* p_t(x,y)\,dy.$$

[A^* is the adjoint operator; see next slide.]
The equation holds for all $f \in C_c^2(\mathbb{R}^d)$, so by calculus of variation,

$$\boxed{\partial_t p_t(x,y) = A_y^* p_t(x,y).}$$

- The forward equation is the dual version of the backward equation.
- It holds for each x. Integrating over x, we get the PDE for $\rho(t,y)$.

Adjoint: The Dual Pairing Behind the Formula

In one dimension, let

$$Af = b(x)f'(x) + \frac{1}{2}a(x)f''(x), \qquad a(x) = \sigma^2(x).$$

The dual action is defined by

$$\langle f, \rho \rangle := \int_{\mathbb{R}} f(x)\rho(x)\,dx.$$

Integration by parts gives

$$\langle Af, \rho \rangle = \left\langle f, -\partial_x(b\rho) + \frac{1}{2}\partial_{xx}(a\rho) \right\rangle.$$

Therefore,

$$\boxed{A^*\rho = -\partial_x(b\rho) + \frac{1}{2}\partial_{xx}(a\rho).}$$

Backward vs. Forward: The Two PDE Viewpoints

Backward PDE

Average observable f:

$$u(t,x) = \mathbb{E}^x[f(X_t)]$$

Equation:

$$\partial_t u = Au, \qquad u(0,x) = f(x).$$

Forward PDE

Density:

$$\rho(t,\cdot) = \text{law density of } X_t$$

Equation:

$$\partial_t \rho = A^* \rho, \qquad \rho(0,\cdot) = \rho_0.$$

Key point

The same diffusion gives two PDEs: one evolves observables, the other evolves probability mass.

Running Example: $dX_t = \mu\, dt + \sigma\, dB_t$. Its density evolves by

$$\boxed{\partial_t \rho = A^* \rho = -\mu \partial_x \rho + \frac{1}{2}\sigma^2 \partial_{xx} \rho.}$$

General 1D equation $dX_t = b(X_t)\, dt + \sigma(X_t)\, dB_t$, the density solves

$$\boxed{\partial_t \rho = -\partial_x (b\rho) + \frac{1}{2}\partial_{xx}(a\rho).}$$

- $-\partial_x(b\rho)$ transports mass by the drift, $\frac{1}{2}\partial_{xx}(a\rho)$ diffuses mass.
- A heat equation with advection: the density is translated by the drift and spread by the Brownian noise.

Further question

What can solutions of $A^*\rho = 0$ tell us about invariant measures and long-time behavior?

The Resolvent Operator

Resolvent Operator: Compression in time

Key question

The semigroup P_t tells us what happens at each time t. Can we compress all times into one operator?

The answer is yes: discount the future and integrate in time.
For $\alpha > 0$, define the resolvent

$$R_\alpha g(x) := \mathbb{E}^x \left[\int_0^\infty e^{-\alpha t} g(X_t)\, dt \right] = \int_0^\infty e^{-\alpha t} P_t g(x)\, dt.$$

Meaning

$R_\alpha g(x)$ is the expected total discounted payoff accumulated along the path. It is the Laplace transform of the semigroup.

Resolvent Identity. If $f \in C_b^2$ and $g = (\alpha - A)f$, then Dynkin's formula applied to $e^{-\alpha t} f(X_t)$ yields

$$\mathbb{E}^x \left[\int_0^\infty e^{-\alpha t} (\alpha - A) f(X_t)\, dt \right] = f(x).$$

Thus

$$\boxed{R_\alpha (\alpha - A) f = f.}$$

Equivalently, if $u = R_\alpha g$, then formally

$$(\alpha - A) R_\alpha g = (\alpha - A) u = g.$$

- The resolvent R_α is the Laplace transform of $P_t = e^{tA}$ (formally):

$$R_\alpha = \int_0^\infty e^{-\alpha t} P_t\, dt = (\alpha - A)^{-1}.$$

- It solves an elliptic problem by damped-averaging along trajectories.

Further question

How much spectral information about A is encoded in R_α?

The Feynman-Kac Formula

The Feynman-Kac Formula: solve PDE by SDE

Key question

Can we solve a parabolic PDE by averaging over random paths?

Let $q \geq 0$ and define

$$\boxed{v(t,x) := \mathbb{E}^x \left[\exp\left(-\int_0^t q(X_s)\, ds \right) f(X_t) \right].}$$

Then $v(0,x) = f(x)$ and, under standard regularity assumptions,

$$\boxed{\partial_t v = Av - qv, \qquad v(0,x) = f(x).}$$

- When $q = 0$, we recover the backward equation.
- The potential term q discounts the future along the path.
- Example: If $q \equiv \lambda > 0$, then $v(t,x) = \mathbb{E}^x[e^{-\lambda t} f(X_t)]$.

Theorem (Theorem 8.2.1)

Let $f \in C_0^2(\mathbb{R}^n)$ and $q \in C^1(\mathbb{R}^n)$ be bounded from below. Define

$$v(t,x) := \mathbb{E}^x \left[\exp\left(-\int_0^t q(X_s)\, ds \right) f(X_t) \right].$$

Then:

1. *v solves the IVP*

$$\partial_t v = Av - qv; \quad v(0,x) = f(x); \quad t > 0, x \in \mathbb{R}^n.$$

2. *If $w(t,x) \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$ is bounded on $K \times \mathbb{R}_+$ for each compact $K \subset \mathbb{R}^n$, and if w solves the IVP, then $w = v$.*

Sketch of the proof.
Part (1). Let

$$Y_t := f(X_t), \qquad Z_t := \exp\left(-\int_0^t q(X_s)\, d\right).$$

Then $dY_t = Af(X_t)\,dt + \nabla f(X_t)\sigma(X_t)\,dB_t$ and $dZ_t = -Z_t q(X_t)\,dt$. By Itô's product rule,

$$d(Y_t Z_t) = Z_t\big[Af(X_t) - q(X_t)f(X_t)\big]\,dt + Z_t \nabla f(X_t)\sigma(X_t)\,dB_t.$$

Then, $v(t,x) = \mathbb{E}^x[Y_t Z_t]$ is differentiable in t and

$$\begin{aligned}\partial_t v(t,x) &= \lim_{h\to 0} \frac{\mathbb{E}^x[Y_{t+h}Z_{t+h}] - \mathbb{E}^x[Y_t Z_t]}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{E}^x\big[\int_t^{t+h} Z_s\big[Af(X_s) - q(X_s)f(X_s)\big]\,ds\big]}{h} \text{"} = \text{"}\, Av - qv.\end{aligned}$$

Instead, use definition of the generator and the Markov property to get

$$
\begin{aligned}
Av(t,x) &= \lim_{r\to 0} \frac{1}{r}\big(\mathbb{E}^x[v(t,X_r)] - v(t,x)\big) \\
&= \lim_{r\to 0} \frac{1}{r}\big(\mathbb{E}^x\big[\mathbb{E}^{X_r}[Z_t f(X_t)]\big] - \mathbb{E}^x[Z_t f(X_t)]\big)
\end{aligned}
$$

Since $e^{-\int_0^t q(X_{s+r})\,ds} = e^{-\int_r^{t+r} q(X_s)\,ds} = Z_{t+r}/Z_r$ and the Markov property,

$$
\begin{aligned}
&\mathbb{E}^x\big[\mathbb{E}^{X_r}[Z_t f(X_t)]\big] - \mathbb{E}^x[Z_t f(X_t)] \\
=&\mathbb{E}^x\big[\mathbb{E}^x[f(X_{t+r})e^{-\int_0^t q(X_{s+r})\,ds} \mid \mathcal{F}_r]\big] - \mathbb{E}^x[Z_t f(X_t)] \\
=&\mathbb{E}^x\big[\mathbb{E}^x[f(X_{t+r})Z_{t+r}/Z_r \mid \mathcal{F}_r]\big] - \mathbb{E}^x[Z_t f(X_t)] \\
=&\mathbb{E}^x\big[\mathbb{E}^x[f(X_{t+r})Z_{t+r} - f(X_t)Z_t \mid \mathcal{F}_r]\big] + \mathbb{E}^x[(Z_r^{-1} - 1)Z_{t+r} f(X_{t+r})] \\
=&\mathbb{E}^x\big[f(X_{t+r})Z_{t+r} - f(X_t)Z_t\big] + \mathbb{E}^x[(e^{\int_0^r q(X_s)\,ds} - 1)Z_{t+r} f(X_{t+r})] \\
\frac{1}{r}* \to&\partial_t v(t,x) + q(x)v(t,x), \quad \text{as } r\to 0,
\end{aligned}
$$

where the second term converges to $q(x)v(t,x)$ by the bounded convergence theorem.

Proof of Part (2). Consider extended process

$$H_t = (s - t, X_t^{0,x}, Z_t), \quad Z_t = z + \int_0^t q(X_r)\, dr,$$

and apply Dynkin's formula to $\phi(H_t))$ with $\phi(s, x, z) = e^{-z} w(s, x)$. See the textbook for details.

Generalization.

- **Feynman-Kac with time-dependent potential**: $q = q(t,x)$, and it can be even stochastic (fractional noise). Hu+Nualart+. 2008-2012.
- **Feynman-Kac with terminal-value**. For $T > 0$ and $g \in C_b^2(\mathbb{R}^d)$,

$$u(t,x) = \mathbb{E}^x \left[e^{-\int_0^{T-t} q(X_s)\, ds} g(X_{T-t}) \right]$$

solves

$$\partial_t u + Au - qu = 0, \qquad u(T,x) = g(x).$$

The Martingale Problem

Martingale Problem: weak solu. and mG characterization

Key question

Characterize the law of a diffusion without fixing a Brownian motion?
Weak solution to SDE v.s. generator with a martingale term?

For Ito diffusion $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ with $X_0 = x$ (generator A)

$$M_t = f(X_t) - \int_0^t Af(X_s)ds = f(x) + \int_0^t \nabla f(X_s)^T \sigma(X_s)dB_s$$

is a martingale for any $f \in C_0^2(\mathbb{R}^d)$.
This only requires the existence of a weak solution, i.e., a probability measure $\mathbb{P}^x$ on the path space such that the coordinate process ω_t satisfies the SDE in distribution.

Martingale Problem: Definition

Definition

Let

$$Lf(x) = \sum_{i=1}^{d} b_i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\partial_{ij} f(x),$$

where a and b are locally bounded Borel measurable; $a(x) \geq 0$ (PSD) for all x. A probability measure $\mathbb{P}^x$ on path space (all start from x) solves the martingale problem for L if for every test function $f \in C_c^\infty(\mathbb{R}^d)$,

$$M_t^f := f(\omega_t) - f(\omega_0) - \int_0^t Lf(\omega_s)\, ds$$

is a martingale in the canonical filtration.
The martingale problem is well-posed if there is a unique solution $\mathbb{P}^x$.

Strook and Varadhan (1979):

- weak soluiton $\Leftrightarrow$ mG problem solution
- weak solution + Markov $\Leftrightarrow$ well-posed mG problem

Itô Processes vs. Itô Diffusions

Itô Process and Itô Diffusion?

Diffusion:

$$\begin{aligned} \text{Itô diffusion:} \qquad & dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t. \\ \text{Itô process:} \qquad & dY_t = u(t,\omega)\,dt + v(t,\omega)\,dB_t, \end{aligned} \tag{1}$$

where u and v are adapted processes satisfying integrability conditions, but may depend on the whole past (not just the current state).

Question

If X_t is an Itô process, then $\phi(X_t)$ is also an Itô process for any $\phi \in C^2$.
If X_t is an Itô diffusion, is $\phi(X_t)$ also an Itô diffusion?

If its local characteristics depend only on the present state, then it may share the law of a diffusion.

Theorem (Theorem 8.4.3: Itô Process vs. Diffusion)

Let X_t and Y_t be as in (1) with $X_0 = Y_0 = x$. Then, X_t has the same law as Y_t for each $t \geq 0$ if and only if

$$\mathbb{E}[u(t,\omega) \mid \mathcal{N}_t^Y] = b(Y_t), \qquad \mathbb{E}[v(t,\omega)v(t,\omega)^T \mid \mathcal{N}_t^Y] = \sigma\sigma^\top(Y_t)$$

a.s. for each $t \geq 0$, where $\mathcal{N}_t^Y$ is the σ-algebra generated by the path of Y up to time t.

The proof is based on the martingale problem and the following Brownian motion characterization.

Theorem (Theorem 8.4.2)

An Ito process $dY_t = vdB_t$; $Y_0 = 0$ coincides in law with an n-dimensional Brownian motion if and only if $vv^\top(t,\omega) = I_n$ a.s. for each $t \geq 0$.

Chapter 8 Takeaways

Main messages

Path dynamics $\rightarrow$ expectation $\rightarrow$ generator $\rightarrow$ backward/forward PDEs.
Feynman-Kac: weighted paths average gives the solution to a PDE.

- The martingale problem characterizes the law.
- Ito process vs. diffusion.

Big picture

Four equivalent lenses to study a diffusion: paths, generators, PDEs, and laws.