Applied Stochastic Analysis

Bangyan Liao
liaobangyan@westlake.edu.cn

Oct 2024

Outline

Outline

Definition (Probability Space)

A probability space is defined as a 3-element tuple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- Ω is the sample space, i.e. the set of possible outcomes. For example, for a coin toss $\Omega = \{\text{Head}, \text{Tails}\}$
- The σ-algebra $\mathcal{F}$ represents the set of events we may want to consider. Continuing the coin toss example, we may have $\Omega = \{\emptyset, \text{Head}, \text{Tails}, \{\text{Head}, \text{Tails}\}\}$
- A probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a function which assigns a number in $[0, 1]$ to any set in the σ-algebra $\mathcal{F}$. The function $\mathbb{P}$ must be σ-additive and $\mathbb{P}(\Omega) = 1$

Definition (σ-algebra)

A σ-algebra $\mathcal{F}$ is a collection of sets satisfying the property

- $\mathcal{F}$ contains $\Omega : \Omega \in \mathcal{F}$.
- $\mathcal{F}$ is closed under complements: if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.
- $\mathcal{F}$ is closed under countable union: if $\forall i \; A_i \in \mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$.

We use the notation $\mathcal{B}(\mathbb{R}^d)$ for the Borel σ-algebra of $\mathbb{R}^d$, which we can think of as the canonical σ-algebra for $\mathbb{R}^d$ - it is the most compact representation of all measurable sets in $\mathbb{R}^d$.

Definition (Probability Measure)

A probability measure $\mathbb{P} : \mathcal{F} \to [0,1]$ is a function which assigns a number in $[0,1]$ to any set in the σ-algebra $\mathcal{F}$.

- For every $A \in \mathcal{F}$, $\mathbb{P}(A)$ is non-negative.
- $\mathbb{P}(\Omega) = 1$.
- For all incompatible set $A_n \in \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n) \tag{1}$$

Random Variable

Probability Space Formalism

Definition (Random Variable)

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a real-valued random variable $\mathrm{x}(\omega)$ is a function $\mathrm{x} : \Omega \to \mathbb{R}^d$, requiring that $\mathrm{x}(\omega)$ is a measurable function, meaning that the pre-image of $\mathrm{x}(\omega)$ lies within the σ-algebra $\mathcal{F}$:

$$\mathrm{x}^{-1}(B) = \{\omega : \mathrm{x}(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}^d) \tag{2}$$

Definition (Probability Distribution)

This allows us to assign a numerical representation to outcomes in Ω. Then, we can ask questions such as what is the probability $P : \mathbb{R}^d \to [0, 1]$ that x is contained within a set $B \subseteq \mathbb{R}^d$

$$P(\mathrm{x}(\omega) \in B) = \mathbb{P}(\{\omega : \mathrm{x}(\omega) \in B\}) \tag{3}$$

Definition (Lebesgue–Stieltjes Integral)

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a measurable function $f : \Omega \to \mathbb{R}$ and a subset $A \in \mathcal{F}$, the Lebesgue–Stieltjes integral

$$\int_A f(x) d\mathbb{P}(x) \tag{4}$$

is a Lebesgue integral with respect to the probability measure $\mathbb{P}$.

If $A = \Omega$, then $\mathbb{E}_{\mathbb{P}}[f(x)] = \int_\Omega f(x) d\mathbb{P}(x)$.
Let $f(x) = \mathbf{1}(x \in A)$, then $\mathbb{E}_{\mathbb{P}}[\mathbf{1}(x \in A)] = \int_A d\mathbb{P}(x) = \mathbb{P}(A)$.

Outline

Definition (Stochastic Process)

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a stochastic process is a collection of random variables X_t or $x(\omega, t) : \Omega \times T \to \mathbb{R}$ indexed by T, which can be written as

$$\{x(\omega, t) : t \in T\} \tag{5}$$

Definition (Filtration)

A filtration $\mathfrak{F} = (\mathcal{F}_t)_{t \in T}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence of indexed sub-σ-algebra of $\mathcal{F}$:

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \quad \forall s \leq t \tag{6}$$

We then call the space $(\Omega, \mathcal{F}, \mathfrak{F}, \mathbb{P})$ an $\mathfrak{F}$-filtered probability space. This allows us to define processes that only depend on the past and present.

Definition (Adapted Process)

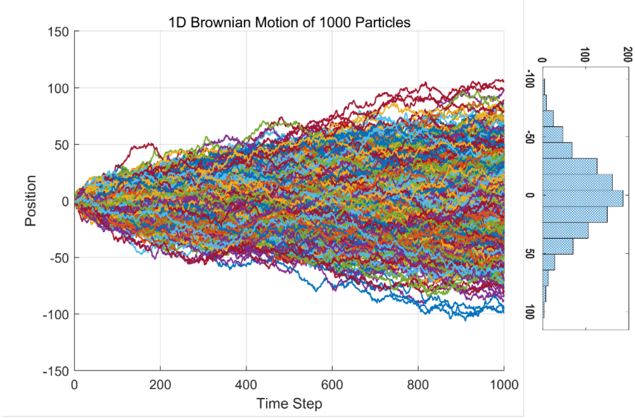
A stochastic process x is $\mathcal{F}_t$-adapted if $x(\omega, t)$ is $\mathcal{F}_t$-measurable:

$$\{\omega : x(\omega, t) \in B\} \in \mathcal{F}_t, \quad \forall t \in T, \forall B \in \mathcal{B}(\mathbb{R}^d) \tag{7}$$

Definition (Wiener Process)

An $\mathcal{F}_t$-adapted Wiener process (Brownian motion) is a stochastic process W_t with the following properties:

- $W_{t_0} = 0$.
- If $[t_1, t_2] \cap [s_1, s_2] = \emptyset$, then $W_{t_2} - W_{t_1}$ and $W_{s_2} - W_{s_1}$ are independent
- $W_{t_2} - W_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$ for $t_2 \geq t_1$



Definition (Stochastic Differential Equation)

For $\mathcal{F}_t$-adapted stochastic processes $\mu(t, X_t)$ and $\sigma(t, X_t)$, an Itô process X_t is defined as

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s, \tag{8}$$

which is often notationally simplified to

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t. \tag{9}$$

Outline

Naively defining the integral with respect to Brownian motion as before is problematic, since the limit is no longer well-defined (unique) for this case:

$$\int_a^b X_t \, dW_t = \lim_{n \to \infty} \sum_{i=0}^{n-1} X_{t_i^*} \left(W_{t_{i+1}} - W_{t_i} \right), \tag{10}$$

where, $t_1 = a < t_2 < ... < t_n = b$, $t_i^* \in [t_i, t_{i+1}]$. For the above limit to exist, we require that the function W_{t_i} has a bounded total variation in t, which does not happen, since Brownian-motion paths do not have bounded total variation.

Definition (Itô Integral)

If we fix the choice $t_i^* = t_i$, it can be shown that this limit will converge in the mean-square sense.

$$\int_a^b X_t \, dW_t = \lim_{n \to \infty} \sum_{i=0}^{n-1} X_{t_i} \left(W_{t_{i+1}} - W_{t_i} \right) . \tag{11}$$

Remark.

The Itô integral is special because it is a martingale.

$$\mathbb{E} \left[\int_0^t Y_s \, dW_s | \mathfrak{F}_r \right] = \int_0^r Y_s \, dW_s, \quad r \leq t \tag{12}$$

when $\mathfrak{F}_r$ is the filtration generated by $\{W_s, Y_s\}_{s \leq r}$.

Lemma (Quadratic Variation)

For a partition $\Pi = \{t_0, t_1, ..., t_j\}$ *of an interval* $[0, T]$, *let* $|\Pi| = max_i(t_{i+1} - t_i)$. *A Brownian motion* W_t *satisfies the following equation with probability 1:*

$$\lim_{|\Pi| \to 0} \sum_i (W_{t_{i+1}} - W_{t_i})^2 = T \tag{13}$$

Remark.

To view it informally, we can say

$$(dW)^2 = dt \tag{14}$$

which is a core transformation in the following proof of Itô Lemma.

Theorem (Itô's lemma)

Let $f(x)$ be a smooth function of two variables, and let X_t be a stochastic process satisfying $dX_t = \mu_t\, dt + \sigma_t\, dW_t$ for a Brownian motion W_t. Then

$$df(t, X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \frac{\partial f}{\partial x} \sigma_t dW_t. \tag{15}$$

Proof.

Following the Taylor expansion, we have

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2 \\ &= \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma_t \frac{\partial f}{\partial x} dW_t \end{aligned} \tag{16}$$

Remark.

For some more complicated SDE

$$dY_t = \mu(t, Y_t)\, dt + \sigma(t, Y_t)\, dB_t, \tag{17}$$

we can define a function such that $Y_t = f(t, X_t)$ and use Itô Lemma to identify the dY_t.

Outline

Kolmogorov Equations

In probability theory, Kolmogorov equations, including Kolmogorov forward equations and Kolmogorov backward equations, characterize continuous-time Markov processes. In particular, they describe how the probability of a continuous-time Markov process in a certain state changes over time. – WikiPedia

For the case of a countable state space and denote the probability from state x at time s to state y at some later time t to be $p(s, x; t, y)$. The Kolmogorov forward equations read

$$\frac{\partial p(s, x; t, y)}{\partial t} = \sum_z p(s, x; t, z) A_{zy}(t), \tag{18}$$

while the Kolmogorov backward equations are

$$\frac{\partial p(s, x; t, y)}{\partial s} = -\sum_z p(s, z; t, y) A_{xz}(t), \tag{19}$$

where $A(t)$ is the generator and $A_{xy}(t) = \left[\frac{\partial p(s,x;t,y)}{\partial t}\right]_{t=s}, \quad \sum_z A_{yz}(t) = 0.$

Theorem (Kolmogorov Backward Equation)

For a stochastic process following the form of $dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$. *The Kolmogorov Backward Equation has the form*

$$\begin{cases} -\frac{\partial u(x,s)}{\partial s} = \mu(s,x)\frac{\partial u(x,s)}{\partial x} + \frac{1}{2}\sigma^2(s,x)\frac{\partial^2 u(x,s)}{\partial x^2}, \quad s < t \\ u(x,t) = f(x) \end{cases} \tag{20}$$

Then, if $f(x) = \delta_y(x)$, *we can derive the transition probability density* $p(s, x; t, y)$ *through the propagation of Kolmogorov Backward Equation.*

$$\begin{cases} -\frac{\partial p(s,x;t,y)}{\partial s} = \mu(s,x)\frac{\partial p(s,x;t,y)}{\partial x} + \frac{1}{2}\sigma^2(s,x)\frac{\partial^2 p(s,x;t,y)}{\partial x^2}, \quad s < t \\ p(t,x;t,y) = \delta_y(x) \end{cases} \tag{21}$$

Proof.

Let us recall the Itô Lemma

$$\begin{aligned} df(X_t) &= \left(\mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2}\right) dt + \frac{\partial f}{\partial x} dW_t \\ &= \mathcal{L} f(X_t) + \frac{\partial f}{\partial x} dW_t \end{aligned} \tag{22}$$

Then, suppose $u(t, x)$ solves the partial differential equation (PDE)

$$\partial_t u + \mathcal{L} u = 0, \quad \text{for } t \leq T \text{ with } u(T, x) = f(x) \tag{23}$$

Proof.

By Ito with $X_t = x$

$$
\begin{aligned}
f(X_T) &= u(T, X_T) \\
&= u(t,x) + \int_t^T \left(\partial_t u(s, X_s) + \partial_{X_s} u(s, X_s)\right) ds \\
&= u(t,x) + \int_t^T \left(\partial_t u(s, X_s) + \mathcal{L} u(s, X_s)\right) ds + \int_t^T \partial_x u(s, X_s)\, \sigma_s(X_s)\, dW_s \\
\mathbb{E}\left[f(X_T) | X_t = x\right] &= u(t,x)
\end{aligned}
\tag{22}
$$

□

Remark.

The Kolmogorov Backward Equation can seen as the optimality condition of the "mean field dynamic programming" problem.
To demonstrate that, recall the expectation explaining $u(x,s) = \mathbb{E}\left[f(X_t)|X_s = x\right]$. The optimality condition states that

$$\mathbb{E}\left[f(X_t)|X_s = x\right] = \mathbb{E}\left[\mathbb{E}\left[f(X_t)|X_{s+\Delta}\right]|X_s = x\right] = \mathbb{E}\left[u(X_{s+\Delta}, s+\Delta)|X_s = x\right] \quad (23)$$

Then, if we denote $du(X_s, s) = \lim_{\Delta \to 0} u(X_{s+\Delta}, s+\Delta) - u(X_s, s)$, the optimality condition $\mathbb{E}\left[du(X_s, s)|X_s = x\right] = 0$ can be stated as

$$-\frac{\partial u(x,s)}{\partial s} = -\mathbb{E}\left[\frac{\partial u(X_s, s)}{\partial s}|X_s = x\right] = \mathbb{E}\left[\frac{\partial u(X_s, s)}{\partial X_s}|X_s = x\right] \quad (24)$$

Theorem (Fokker-Planck (FPK) Equation)

For a stochastic process following the form of $dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t$. *The Fokker-Planck (FPK) equation has the form*

$$\begin{cases} \frac{\partial u(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left(\mu(y,t)u(y,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)u(y,t)\right), \quad s < t \\ u(y,s) = p(y) \end{cases} \tag{25}$$

Then, if $p(y) = \delta_x(y)$, *we can derive the transition probability density* $p(s,x;t,y)$ *through the propagation of Fokker-Planck Equation.*

$$\begin{cases} \frac{\partial p(s,x;t,y)}{\partial t} = -\frac{\partial}{\partial y}\left(\mu(y,t)p(s,x;t,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)p(s,x;t,y)\right), \quad s < t \\ p(t,x;t,y) = \delta_x(y) \end{cases} \tag{26}$$

Proof.

According to the definition

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[u(X_t)|X_s\right] &= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[u(X_{t+\Delta}) - u(X_t)|X_s\right] \\
&= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\mathbb{E}\left[u(X_{t+\Delta}) - u(X_t)|X_t\right]|X_s\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\partial u(X_t, t)}{\partial X_t}|X_t = x\right]|X_s\right] \\
&= \mathbb{E}\left[\mu(s, x)\frac{\partial}{\partial x}u(X_t, t) + \frac{1}{2}\sigma^2(X_t, t)\frac{\partial^2}{\partial x^2}u(X_t, t)|X_s\right]
\end{aligned} \tag{27}$$

Proof.

$$\frac{d}{dt}\mathbb{E}\left[u(X_t)|X_s = x\right] = \mathbb{E}\left[\mu(s,x)\frac{\partial}{\partial x}u(X_t,t) + \frac{1}{2}\sigma^2(X_t,t)\frac{\partial^2}{\partial x^2}u(X_t,t)|X_s = x\right]$$

$$\begin{aligned}\int u(y)\frac{\partial p(s,x;t,y)}{\partial t}dy &= \int\left[\mu(y,t)\frac{\partial}{\partial y}u(y,t) + \frac{1}{2}\sigma^2(y,t)\frac{\partial^2}{\partial y^2}u(y,t)\right]p(s,x;t,y)dy \\ &= \int u(y)\left[-\frac{\partial}{\partial y}(\mu(y,t)p(s,x;t,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)p(s,x;t,y)\right)\right]dy\end{aligned} \tag{27}$$

which shows that

$$\frac{\partial p(s,x;t,y)}{\partial t} = -\frac{\partial}{\partial y}(\mu(y,t)p(s,x;t,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)p(s,x;t,y)\right) \tag{28}$$

□

Corollary (Master Equation.)

If X_0 has density function $p_0(x)$, then the density function $p(t,y)$ of X_t can be get by propagating the Fokker-Planck equation.

$$\begin{cases} \frac{\partial p(t,y)}{\partial t} = -\frac{\partial}{\partial y}\left(\mu(y,t)p(t,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)p(t,y)\right), \quad s < t \\ p(0,y) = p_0(y) \end{cases} \tag{29}$$

Proof.

$$\begin{aligned} \mathbb{E}(f(X_t)) &= \mathbb{E}(\mathbb{E}[f(X_t)]|X_0) \\ &= \int \left[\int f(y)p(0,x;t,y)dy\right] p_0(x)dx \\ \int f(y)p(t,y)dy &= \int f(y)\left[\int p_0(x)p(0,x;t,y)dx\right] dy \end{aligned} \tag{30}$$

Reverse-time SDE

Kolmogorov Equations - Some Corollaries

Definition. Given the stochastic process $X(\cdot) : dX = F(X,t)dt + G(X,t)dW$ and the marginal probability density $p_t(X(t))$ at time t, the reverse-time stochastic process is defined as

$$dX = -\left\{F(X,\tilde{t}) - \nabla \cdot \left[G(X,\tilde{t})G(X,\tilde{t})^T\right] - G(X,\tilde{t})G(X,\tilde{t})^T \nabla_x \log p_{\tilde{t}}(x)\right\} d\tilde{t} + G(X,\tilde{t})d\tilde{W}$$

when $n = 1$ and $G(X,t) = G(t)$

$$dX = -\left[F(X,\tilde{t}) - G^2(\tilde{t})\nabla_x \log p_{\tilde{t}}(x)\right] d\tilde{t} + G(\tilde{t})d\tilde{W}$$

where $\tilde{W}(\cdot)$ represents the standard Wiener process when time flows backwards, and $d\tilde{t}$ is an infinitesimal negative timestep from T to 0.

Proof. For some stochastic process $X(\cdot) : dX = F(X,t)dt + G(t)dW$, the corresponding Fokker-Planck equation is defined as

$$\frac{\partial p_t(X)}{\partial t} = -\frac{\partial}{\partial x}\left[F(X,t)p_t(X)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[G^2(t)p_t(X)\right]$$

We also define the reverse-time stochastic process $Y(\cdot) : dY = F(Y,\tilde{t})dt + G(\tilde{t})d\tilde{W}$, and the corresponding $q_t(Y)$ is defined as

$$\frac{\partial q_t(Y)}{\partial t} = -\frac{\partial p_{T-t}(X)}{\partial t} = \frac{\partial}{\partial x}\left[F(X,T-t)p_{T-t}(X)\right] - \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[G^2(T-t)p_{T-t}(X)\right]$$

$$= \frac{\partial}{\partial x}\left[\left(F(X,T-t) - G^2(T-t)\nabla_x \log p_{T-t}(x)\right)p_{T-t}(X)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[G^2(T-t)p_{T-t}(X)\right]$$

$$= \frac{\partial}{\partial y}\left[\left(F(X,\tilde{t}) - G^2(\tilde{t})\nabla_x \log p_{\tilde{t}}(x)\right)q_t(Y)\right] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[G^2(\tilde{t})q_t(Y)\right]$$

then, according the FK equation, we have

$$F(Y,\tilde{t}) = -F(X,\tilde{t}) + G^2(\tilde{t})\nabla_x \log p_{\tilde{t}}(x), \quad G(t) = G(\tilde{t})$$

Definition. For each reverse-time stochastic process, the probabilistic flow ODE can be defined as followed whose trajectories share the marginal probability densities $p_t(X(t))$.

$$dX = -\left\{ F(X,\tilde{t}) - \frac{1}{2}\nabla \cdot \left[G(X,\tilde{t})G(X,\tilde{t})^T \right] - \frac{1}{2}G(X,\tilde{t})G(X,\tilde{t})^T \nabla_x \log p_{\tilde{t}}(x) \right\} d\tilde{t}$$

when $n = 1$ and $G(X,t) = G(t)$

$$dX = -\left[F(X,\tilde{t}) - \frac{1}{2}G^2(\tilde{t})\nabla_x \log p_{\tilde{t}}(x) \right] d\tilde{t}$$

where $d\tilde{t}$ is an infinitesimal negative timestep from T to 0.

Proof. For some stochastic process $X(\cdot): dX = F(X,t)dt + G(t)dW$, the corresponding Fokker-Planck equation is defined as

$$\frac{\partial p_t(X)}{\partial t} = -\frac{\partial}{\partial x}\left[F(X,t)p_t(X)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[G^2(t)p_t(X)\right]$$

We also define the reverse-time ode process $Y(\cdot): dY = F(Y,\tilde{t})d\tilde{t}$, and the corresponding $q_t(Y)$ is defined as

$$\begin{aligned}\frac{\partial q_t(Y)}{\partial t} = -\frac{\partial p_{T-t}(X)}{\partial t} &= \frac{\partial}{\partial x}\left[F(X,T-t)p_{T-t}(X)\right] - \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[G^2(T-t)p_{T-t}(X)\right] \\ &= \frac{\partial}{\partial x}\left[\left(F(X,T-t) - \frac{1}{2}G^2(T-t)\nabla_x \log p_{T-t}(x)\right)p_{T-t}(X)\right] \\ &= \frac{\partial}{\partial y}\left[\left(F(X,\tilde{t}) - \frac{1}{2}G^2(\tilde{t})\nabla_x \log p_{\tilde{t}}(x)\right)q_t(Y)\right]\end{aligned}$$

then, according the continuity equation, we have

$$F(Y,\tilde{t}) = -F(X,\tilde{t}) + \frac{1}{2}G^2(\tilde{t})\nabla_x \log p_t(x)$$

Outline

Theorem (Generator Definition)

$$\left.\frac{d}{dh}\right|_{h=0} \left\langle k_{t+h|t}, f \right\rangle (x) = \lim_{h \to 0} \frac{\left\langle k_{t+h|t}, f \right\rangle (x) - f(x)}{h} \stackrel{def}{=} [\mathcal{L}_t f](x) \tag{31}$$

where function f is the integrable test function and $k_{t+h|t}$ represents the transition kernel from time t to time $t+h$. We can define the linear action $\langle \cdot, \cdot \rangle$ to be

$$\begin{aligned} \langle p_t, f \rangle &\stackrel{def}{=} \int f(x) p_t(dx) = \mathbb{E}_{x \sim p_t}[f(x)] \\ \left\langle k_{t+h|t}, f \right\rangle (x) &\stackrel{def}{=} \left\langle k_{t+h|t}(\cdot|x), f \right\rangle = \mathbb{E}[f(X_{t+h}) \mid X_t = x] \end{aligned} \tag{32}$$

The tower property implies that $\left\langle p_t, \left\langle k_{t+h|t}, f \right\rangle \right\rangle = \langle p_{t+h}, f \rangle$.

Corollary (Flow)

Given the ODE $dX_t = u(X_t, t)dt$, *the generator is*

$$\begin{aligned} [\mathcal{L}_t f](x) &= \lim_{h \to 0} \frac{\mathbb{E}[f(X_t + hu_t(X_t) + o(h))|X_t = x] - f(x)}{h} \\ &= \lim_{h \to 0} \frac{h\nabla f(x)^T u_t(x) + o(h)}{h} \\ &= \nabla f(x)^T u_t(x) \end{aligned} \tag{33}$$

Corollary (Diffusion)

Given the SDE $dX_t = \sigma(X_t, t)dB$, *the generator is*

$$
\begin{aligned}
\left[\mathcal{L}_t f\right](x) &= \lim_{h \to 0} \frac{\mathbb{E}\left[f(X_t + h\sigma_t(X_t)\epsilon_t + o(h)) | X_t = x\right] - f(x)}{h} \\
&= \frac{1}{2}\sigma_t^2(x) \cdot \nabla^2 f(x)
\end{aligned}
\tag{34}
$$

Theorem (Kolmogorov Forward Equation)

$$\partial_t \langle p_t, f \rangle = \frac{d}{dh}\bigg|_{h=0} \langle p_{t+h}, f \rangle = \left\langle p_t, \frac{d}{dh}\bigg|_{h=0} \left\langle k_{t+h|t}, f \right\rangle \right\rangle = \langle p_t, \mathcal{L}_t f \rangle \tag{35}$$

Theorem (Adjoint KFE)

Let us define the adjoint generator $\mathcal{L}_t^*$ as

$$\langle p_t, \mathcal{L}_t f \rangle = \langle \mathcal{L}_t^* p_t, f \rangle \tag{36}$$

Then, we have this adjoint Kolmogorov Forward Equation

$$\partial_t p_t(x) = [\mathcal{L}_t^* p_t](x) \tag{37}$$

Proof of Adjoint KFE.

$$\begin{aligned}\partial_t \langle p_t, f \rangle &= \partial_t \int f(x) p_t(dx) \\ &= \int f(x) \partial_t p_t(dx) \\ &= \langle p_t, \mathcal{L}_t f \rangle \\ &= \langle \mathcal{L}_t^* p_t, f \rangle \\ &= \int f(x) \mathcal{L}_t^* p_t(dx)\end{aligned} \tag{38}$$

□

Corollary (Flow)

The adjoint generator is $\mathcal{L}_t^* p_t = -\nabla \cdot [u_t(x) p_t(x)]$, *which leads to the well-known continuity equation:*

$$\partial_t p_t(x) = -\nabla \cdot [u_t(x) p_t(x)] \tag{39}$$

Proof.

$$\begin{aligned}
\langle p_t, \mathcal{L}_t f \rangle = \mathbb{E}_{x \sim p_t}[\mathcal{L}_t f(x)] = \int \mathcal{L}_t f(x) p_t(x)\, dx = \int \nabla f(x)^T u_t(x) p_t(x)\, dx \\
= \int f(x)[-\nabla \cdot [u_t(x) p_t(x)]]\, dx \\
= \int f(x)[\mathcal{L}_t^* p_t](x)\, dx
\end{aligned} \tag{40}$$

□

Corollary (Diffusion)

The adjoint generator is $\mathcal{L}_t^* p_t = \frac{1}{2}\nabla^2 \cdot [\sigma_t^2(x) p_t(x)]$*, which leads to the well-known Fokker-Planck equation:*

$$\partial_t p_t(x) = \frac{1}{2}\nabla^2 \cdot [\sigma_t^2(x) p_t(x)] \tag{41}$$

Proof.

$$\begin{aligned}
\langle p_t, \mathcal{L}_t f\rangle &= \mathbb{E}_{x\sim p_t}[\mathcal{L}_t f(x)] = \int \mathcal{L}_t f(x) p_t(x)\, dx = \frac{1}{2}\int \sigma_t^2(x) \cdot \nabla^2 f(x) p_t(x)\, dx \\
&= \frac{1}{2}\int f(x)\nabla^2 \cdot [\sigma_t^2(x) p_t(x)]\, dx \\
&= \int f(x)[\mathcal{L}_t^* p_t](x)\, dx
\end{aligned} \tag{42}$$

Theorem (Kolmogorov Backward Equation)

$$\frac{\partial}{\partial s} \langle k_{t|s}, f \rangle (x) = -\mathcal{L}_s \langle k_{t|s}, f \rangle (x) \tag{43}$$

Proof of Kolmogorov Backward Equation.

Let us first expand the transition kernel from $s \to t$ to $s \to s+h \to t$:

$$\begin{aligned} \left\langle k_{t|s}, f \right\rangle (x) &= \left\langle k_{s+h|s}, \left\langle k_{t|s+h}, f \right\rangle \right\rangle (x) \\ \mathbb{E}\left[f(X_t) \mid X_s = x\right] &= \mathbb{E}\left[\left\langle k_{t|s+h}, f \right\rangle (X_{s+h}) \mid X_s = x\right] \\ &= \mathbb{E}\left[f(X_t) \mid X_s = x\right] \end{aligned} \tag{44}$$

Then, take derivative on both side

$$\begin{aligned} \frac{d}{dh}|_{h=0} \left\langle k_{t|s}, f \right\rangle (x) = \frac{d}{dh}|_{h=0} \left\langle k_{s+h|s}, \left\langle k_{t|s+h}, f \right\rangle \right\rangle (x) \\ 0 = [\mathcal{L}_s \left\langle k_{t|s}, f \right\rangle](x) + \frac{d}{dh}|_{h=0} \left\langle k_{t|s+h}, f \right\rangle (x) \\ = \mathcal{L}_s \left\langle k_{t|s}, f \right\rangle (x) + \frac{\partial}{\partial s} \left\langle k_{t|s}, f \right\rangle (x) \end{aligned} \tag{45}$$

Proof of Kolmogorov Backward Equation.

This concludes that

$$\frac{\partial}{\partial s} \left\langle k_{t|s}, f \right\rangle (x) = -\mathcal{L}_s \left\langle k_{t|s}, f \right\rangle (x) \tag{44}$$

Let $u(x, s) = \mathbb{E}\left[f(X_t) | X_s = x\right]$, this is equivalent to the KBE in section 4.1. □

Remark.

Let $p(s, x; t, y)$ represents the probability transition from time s at x to time t at y. Then the Kolmogorov Forward and Backward Equation can be written as:

$$\begin{aligned} \frac{\partial}{\partial t} p(s, x; t, \cdot) &= +\mathcal{L}_t^* \, p(s, x; t, \cdot) \\ \frac{\partial}{\partial s} p(s, \cdot; t, y) &= -\mathcal{L}_s \, p(s, \cdot; t, y) \end{aligned} \tag{45}$$

Outline

Radon-Nikodym Derivative

Theorem (Radon-Nikodym Theorem)

Given probability measures $\mathbb{P}$ and $\mathbb{Q}$, defined on the measurable space $(\Omega, \mathcal{F})$, there exists a measurable function $\frac{d\mathbb{P}}{d\mathbb{Q}} : \Omega \to [0, \infty)$, and for any set $A \subseteq \mathcal{F}$:

$$\mathbb{P}(A) = \int_A \frac{d\mathbb{P}}{d\mathbb{Q}}(x)\, d\mathbb{Q}(x), \tag{46}$$

where the function $\frac{d\mathbb{P}}{d\mathbb{Q}}(x)$ is known as the RN-derivative.

A direct consequence of this result is

$$\int_A f(x)\, d\mathbb{P}(x) = \int_A f(x) \frac{d\mathbb{P}}{d\mathbb{Q}}(x)\, d\mathbb{Q}(x). \tag{47}$$

Theorem (Disintegration Theorem)

Disintegration Theorem for continuous probability measures: For a probability space

$Z, \mathcal{B}(Z), \mathbb{P}$ where Z is a product space: $Z = Z_x \times Z_y$, and

- $Z_x \subseteq \mathbb{R}^d$, $Z_y \subseteq \mathbb{R}^d$,
- $\pi_i : Z \to Z_i$ is a measurable function known as the canonical projection operator (i.e., $\pi_x(z_x, z_y) = z_x$ and $\pi_x^{-1}(z_x) = \{y|\pi_x(z_x) = z\}$),

there exists a measure $\mathbb{P}_{y|x}(\cdot|x)$, such that

$$\int_{Z_x \times Z_y} f(x, y)\, d\mathbb{P}(y) = \int_{Z_x} \int_{Z_y} f(x, y)\, d\mathbb{P}_{y|x}(y|x)\, d\mathbb{P}(\pi_x^{-1}(x)) \tag{48}$$

where $\mathbb{P}_x(\cdot) = \mathbb{P}(\pi^{-1}(\cdot))$ is a probability measure, typically referred to as a pullback measure, and corresponds to the marginal distribution.

Corollary

The disintegration theorem implies a very interesting corollary as:

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(x, y) = \frac{d\mathbb{P}_{y|x}}{d\mathbb{Q}_{y|x}}(y)\frac{d\mathbb{P}_x}{d\mathbb{Q}_x}(x) \tag{49}$$

Remarks.

The disintegration theorem can be seen as the conditional probability on measure space.

Definition (Path Measure)

For an Itô process of the form $dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t$ defined in $[0, T]$, we call $\mathbb{P}$ the path measure of the above process, with outcome space $\Omega = C([0, T], \mathbb{R}^d)$, if the distribution $\mathbb{P}$ describes a weak solution to the above SDE.

Theorem (Girsanov Theorem)

Given two Itô processes with the same constant volatility: $dX_t = \mu_1(t, X_t)\,dt + \sigma\,dW_t$ *and* $dY_t = \mu_2(t, X_t)\,dt + \sigma\,dW_t$, *the RN derivative of their respective path measures* $\mathbb{P}, \mathbb{Q}$ *is given by*

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(\cdot) = \exp\left(-\frac{1}{2\sigma^2}\int_0^t \|\mu_1(s,\cdot) - \mu_2(s,\cdot)\|^2\,ds + \frac{1}{\sigma^2}\int_0^t (\mu_1(s,\cdot) - \mu_2(s,\cdot))^\top\,dW_s\right) \tag{50}$$

where the type signature of this RN derivative is $\frac{d\mathbb{P}}{d\mathbb{Q}} : C(T, \mathbb{R}^d) \to \mathbb{R}$.

Outline

Theorem (Feynman-Kac Formulation [Discounting])

For a stochastic process following the form of $dX_t = \mu_1(t, X_t)\, dt + \sigma\, dW_t$. *If* $u(x,t)$ *satisfies the form*

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} + \mu(x,t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 u(x,t)}{\partial x^2} - q(x,t)u(x,t) = -g(x,t) \\ u(x,T) = f(x) \end{cases} \tag{51}$$

Then, Feynman-Kac Formulation tells us that under the Wiener process $dX_t = dW_t$

$$u(x,t) = \mathbb{E}\left[f(\xi_T)e^{-\int_t^T q(\theta,\xi_\theta)d\theta} + \int_t^T g(s,\xi_s)e^{-\int_t^T q(\theta,\xi_\theta)d\theta} ds | \xi_t = x\right] \tag{52}$$

Proof.

Recall the Itô formula

$$\begin{aligned}\mathrm{d}u(\xi_s, s) &= \left(\frac{\partial u(\xi_s, s)}{\partial s} + \mu(\xi_s, s)\frac{\partial u(\xi_s, s)}{\partial x} + \frac{1}{2}\sigma^2(\xi_s, s)\frac{\partial^2 u(\xi_s, s)}{\partial x^2}\right)\mathrm{d}s \\ &\quad + \frac{\partial u(\xi_s, s)}{\partial x}\sigma(\xi_s, s)\mathrm{d}W_t \\ &= q(\xi_s, s)u(\xi_s, s)\mathrm{d}s - g(\xi_s, s)\mathrm{d}s + \frac{\partial u(\xi_s, s)}{\partial x}\sigma(\xi_s, s)\mathrm{d}W_s\end{aligned} \tag{53}$$

Proof.

multiplying both sides of the above equation by the integrating factor $e^{-\int_t^s q(\xi_\theta,\theta)\mathrm{d}\theta}$, and using the Itô formula, we have

$$\begin{aligned}\mathrm{d}\left(u(\xi_s,s)e^{-\int_t^s q(\xi_\theta,\theta)\mathrm{d}\theta}\right) &= -q(\xi_s,s)e^{-\int_t^s q(x_\theta,\theta)\mathrm{d}\theta}u(\xi_s,s)\mathrm{d}s + e^{-\int_t^s q(\xi_\theta,\theta)\mathrm{d}\theta}\mathrm{d}u(\xi_s,s)\\ &= e^{-\int_t^s q(\xi_\theta,\theta)\mathrm{d}\theta}\left(-g(\xi_s,s)\mathrm{d}s + \frac{\partial u(\xi_s,s)}{\partial x}\sigma(\xi_s,s)\mathrm{d}W_s\right)\end{aligned} \tag{53}$$

Substituting the initial time t and terminal time T, we obtain

$$u(t,\xi_t) = f(\xi_T)e^{-\int_t^T q(\xi_\theta,\theta)\mathrm{d}\theta} + \int_t^T e^{-\int_t^s q(\xi_\theta,\theta)\mathrm{d}\theta}\left(g(\xi_s,s)\mathrm{d}s - \frac{\partial u(\xi_s,s)}{\partial x}\sigma(\xi_s,s)\mathrm{d}W_s\right).$$

Taking the expectation $\mathbb{E}(\cdot \mid \xi_t = x)$ over Wiener process yields the desired result. □

Theorem (Feynman-Kac Formulation [Non-Discounting])

For a stochastic process following the form of $dX_t = \mu(t, X_t)\, dt + \sigma\, dW_t$. *If* $u(x, t)$ *satisfies the form*

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} + \mu(x, t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2(x, t)\frac{\partial^2 u(x,t)}{\partial x^2} = -g(x, t) \\ u(x, T) = f(x) \end{cases} \tag{54}$$

Then, Feynman-Kac Formulation tells us that

$$u(x, t) = \mathbb{E}\left[f(\xi_T) + \int_t^T g(s, \xi_s) ds | \xi_t = x\right] \tag{55}$$

Proof.

Recall the Itô formula

$$\begin{aligned} \mathrm{d}u(\xi_s, s) &= \left(\frac{\partial u(\xi_s, s)}{\partial s} + \mu(\xi_s, s)\frac{\partial u(\xi_s, s)}{\partial x} + \frac{1}{2}\sigma^2(\xi_s, s)\frac{\partial^2 u(\xi_s, s)}{\partial x^2} \right) \mathrm{d}s \\ &\quad + \frac{\partial u(\xi_s, s)}{\partial x}\sigma(\xi_s, s)\mathrm{d}W_t \\ &= -g(\xi_s, s)\mathrm{d}s + \frac{\partial u(\xi_s, s)}{\partial x}\sigma(\xi_s, s)\mathrm{d}W_s \end{aligned} \tag{56}$$

Proof.

Then, integrate the $u(x, t)$ from time t at x to the terminal time T, we get

$$u(x,t) = \mathbb{E}\left[f(\xi_T) + \int_t^T g(s,\xi_s)ds|\xi_t = x\right] \tag{56}$$

□

Theorem (Nonlinear Feynman-Kac Lemma)

Given the non-linear extension of Feynman-Kac PDE

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} + \mu(x,t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 u(x,t)}{\partial x^2} = -g(x,t,u,\nabla u) \\ u(x,T) = f(x) \end{cases} \tag{57}$$

and the Forward-Backward Differential Equation

$$\begin{cases} dX_t = \mu(t, X_t)\, dt + \sigma\, dW_t,\ X_0 = x_0 \\ dY_t = -g(x,t,u,\nabla u)\, dt + Z_t \cdot dW_t,\ Y_T = f(X_T) \end{cases} \tag{58}$$

If the PDE has unique solution, then we have

$$u(X_t, t) = Y_t,\ \sigma\nabla u = Z_t \tag{59}$$

Given a process X_t that solves $dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t$ and assuming that we want to condition its solution to hit X_T at time $t = T$, then the h-transform provides us with the following SDE for the conditioned process:

$$dX = [\mu(t, X_t) + \sigma(t, X_t)Q\sigma(t, X_t)\nabla \log p(X_T \mid X_t)]\,dt + \sigma(t, X_t)dW_t,$$

Let us define a forward process X_t that solves $dX_t = \mu_+(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$ and a backward process $X_{\tilde{t}}$ that solves $dX_{\tilde{t}} = \mu_-(\tilde{t}, X_{\tilde{t}})\, d\tilde{t} + \sigma(\tilde{t}, X_{\tilde{t}})\, dW_{\tilde{t}}$. We can also define the corresponding probability measure as $p_t(x)$ and $p_{\tilde{t}}(x)$ respectively. Then, if $p_{T-t}(x) = p_{\tilde{t}}(x)$. The Nelson's Duality tells us that

$$\mu_+(t, x) + \mu_-(\tilde{t}, x) = \sigma^2 \nabla_x \log p_{\tilde{t}}(x) = \sigma^2 \nabla_x \log p_t(x) \tag{60}$$

Expected Grad-Log-Prob Lemma

Suppose that P_θ is a parameterized probability distribution over a random variable x.

$$\mathbb{E}_{x \sim P_\theta}[\nabla_\theta \log P_\theta(x)] = 0 \tag{61}$$

Proof.

$$\begin{aligned} \nabla_\theta \int_x P_\theta(x) = \nabla_\theta 1 = 0 \\ \int_x \nabla_\theta P_\theta(x) = 0 \\ \int_x P_\theta(x) \nabla_\theta \log P_\theta(x) = 0 \\ \mathbb{E}_{x \sim P_\theta}[\nabla_\theta \log P_\theta(x)] = 0 \end{aligned} \tag{62}$$

□

For any point $x \in \mathbb{R}^d$ such that $p(x) \neq 0$, it holds that

$$\begin{aligned}
\frac{1}{p(x)}\Delta p(x) &= \frac{1}{p(x)}\nabla \cdot \nabla p(x) = \frac{1}{p(x)}\nabla \cdot (p(x)\nabla \log p(x)) \\
&= \frac{1}{p(x)}(\nabla p(x) \cdot \nabla \log p(x) + p(x)\Delta \log p(x)) \\
&= \|\nabla \log p(x)\|^2 + \Delta \log p(x)
\end{aligned} \tag{63}$$

Under mild assumptions such that all distributions approach zero at a sufficient speed as $\|x\| \to \infty$, and that all integrands are bounded, we have

$$\mathbb{E}_{x\sim p(x)}\left[\Delta \log q(x)\right] = \mathbb{E}_{x\sim p(x)}\left[\nabla \cdot \nabla \log q(x)\right] = \mathbb{E}_{x\sim p(x)}\left[-\nabla \log p(x) \cdot \nabla \log q(x)\right] \tag{64}$$

where the second equality follows by integration by parts and reparameterization trick

$$\begin{aligned} \int p(x) \left(\nabla \cdot \nabla \log q(x)\right) dx &= \int -\left(\nabla p(x) \cdot \nabla \log q(x)\right) dx \\ &= \int -p(x) \left(\nabla \log p(x) \cdot \nabla \log q(x)\right) dx. \end{aligned} \tag{65}$$

More generally, under the same regularity, it holds for a vector field $Z : \mathbb{R}^d \to \mathbb{R}^d$ that

$$\mathbb{E}_{x\sim p(x)}\left[\nabla \cdot Z(x)\right] = \mathbb{E}_{x\sim p(x)}\left[-\nabla \log p(x) \cdot Z(x)\right]. \tag{66}$$

Reference

- Applied Stochastic Calculus
- Topics in Mathematics with Applications in Finance
- Machine-learning approaches for the empirical Schrodinger bridge problem
- AN INTRODUCTION TO STOCHASTIC DIFFERENTIAL EQUATIONS