Diffusion Process

赵尉辰

南开大学 统计与数据科学学院

目录

目录

扩散过程

扩散是一种物理现象，指任何物质（如原子、离子、分子、能量）从浓度较高的区域向浓度较低的区域的净移动。例如，颜料滴进水中，在水里扩散开来的运动过程，扩散过程并没有统一的数学定义，但其核心是轨道连续的马氏过程。

历史上，研究扩散过程主要有两种方法：

- 通过建立转移概率满足的方程来刻画其宏观性质的演化，例如，将墨水滴入水中，墨水浓度在不同时刻和位置遵从的规律，这是 Kolmogorov 的分析方法。
- 利用追踪每个花粉粒子或墨水粒子的轨迹，构造概率空间, 建立随机微分方程从微观角度描述其服从的运动规律，这是 Itô 的随机分析方法。

Itô扩散过程

定义 1 (Itô扩散过程)

Itô 扩散过程为一个随机过程 $X_t(\omega) = X(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ 满足如下形式的随机微分方程

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad t \geq s; \quad X_s = x$$

其中B_t为一个m维布朗运动，并且 $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$，$\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{m \times n}$满足随机微分方程的强解存在唯一性条件

(1) 存在常数C, 使得 $|b(t,x)| + |\sigma(t,x)| \leqslant C(1+|x|), \quad \forall x \in \mathbb{R}^n, t \in [0,T]$,

(2) 存在常数K, 使得

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant K|x-y|, \quad \forall x, y \in \mathbb{R}^n, t \in [0,T].$$

如果b, σ只依赖于X_t，并且满足

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leqslant D|x-y|; \quad x, y \in \mathbb{R}^n,$$

我们称X_t为一个时齐的 *Itô* 扩散过程。

Markov性

Markov过程：设$\{X_t : t \in \mathcal{T}\}$为一个$(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathrm{P})$上$\mathscr{F}_t$适应过程，状态空间为$(E, \mathscr{E})$，称$X$ 为一个Markov过程，如果对所有E上有界可测函数f， $s < t \in \mathcal{T}$

$$\mathbb{E}[f(X_t)|\mathscr{F}_s] = \mathbb{E}[f(X_t)|X_s] \quad a.s.$$

定理 1

Itô 扩散过程是一个 *Markov* 过程。

目录

Markov半群

对状态空间为$(E,\mathscr{E})$的 Markov 过程(X_t)，定义:

$$(P_t f)(x) \triangleq \mathbb{E}[f(X_t)|X_0 = x] = \int f(y)p(t,x,y)dy.$$

其中

$$P_t : \mathcal{M}_b(E) \to \mathcal{M}_b(E),\ f \mapsto P_t f$$

为$\mathcal{M}_b(E)$上的算子，$\mathcal{M}_b(E)$ 为E上所有有界可测函数的集合。 那么由KC方程， P_t满足半群(semigroup)性质

$$P_{t+s} f = P_t \circ P_s f.$$

称(P_t) 为一个Markov半群。

设P_t为一个Markov半群，定义

$$\mathcal{L} f := \lim_{t \downarrow 0} \frac{P_t f - f}{t}.$$

我们称$\mathcal{L}$ 为P_t的无穷小生成元，简称生成元。

伴随半群

定义 2 (伴随半群)

设μ为X_0的分布，定义算子：

$$(P_t^*\mu)(A) \triangleq \int P(X_t \in A|X_0 = x)d\mu(x) = \int p(t,x,A)d\mu(x).$$

那么P_t^*满足

$$\int (P_t f)(x)d\mu(x) = \int f(x)d(P_t^*\mu)(x)$$

我们称算子P_t 和 P_t^*是伴随的，P_t^*为Markov半群P_t的伴随半群。记 $\mu_t \triangleq P_t^*\mu$，那么$\mu_t$为 Markov 过程 X_t 的分布。

注. 设 T 为内积空间 H 上的算子，则 T 的伴随算子 T^* 是满足以下关系的唯一算子：

$$\langle T(x), y\rangle = \langle x, T^*(y)\rangle, \quad \forall x, y \in H,$$

把 $d\mu(x)$ 写成密度函数 $g(x)dx$，则 $d(P_t^*\mu)(x) = P_t^* g(x)dx$，伴随性质即为

$$\int P_t f(x)g(x)dx = \int f(x)P_t^* g(x)dx.$$

不变测度

定义 3 (不变测度)

给定一个*Markov* 半群 P_t，我们称测度 π 是P_t不变的，如果对任意正有界可测函数 $f : E \to \mathbb{R}$以及任意$t \geq 0$，有

$$\int_E P_t f d\pi = \int_E f d\pi.$$

或者等价地，

$$P_t^* \pi = \pi$$

目录

扩散算子

定义 4 (扩散算子)

一个*Markov*扩散算子*(diffusion operator)* $\mathcal{L}$为一个如下形式的二阶微分算子

$$\begin{aligned}\mathcal{L}f &= \sum_i b_i(x)\frac{\partial}{\partial x_i}f + \frac{1}{2}\sum_{i,j}\Sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}f \\ &= b(x)\cdot\nabla f + \frac{1}{2}Tr(\Sigma(x)^T\nabla^2 f)\end{aligned} \tag{1}$$

其中 $\Sigma(x) = (\Sigma_{ij}(x))_{1\le i,j\le n}$ 和 $b(x) = (b_i(x))_{1\le i\le n}$分别为$x$的光滑 $n\times n$ 对称矩阵值和 $\mathbb{R}^n$值函数。有时也记 $\Sigma : \nabla^2 f \triangleq Tr(\Sigma(x)^T\nabla^2 f)$为矩阵的*Frobenius*内积。

经典扩散过程

定义 5 (经典一维扩散过程)

一个$\mathbb{R}$上转移函数为$p(s,x;t,A)$ 的 *Markov* 过程 X_t 称为一个扩散过程，如果：

(1) (连续性Continuity). 对于任意x和$\epsilon > 0$,

$$\lim_{t-s\downarrow 0} \frac{1}{t-s} p(s,x;t,\{|X_t - x| > \varepsilon\}) = 0.$$

(2) (漂移系数的定义). 存在函数$b(s,x)$使得，对于任意x和$\epsilon > 0$,

$$\lim_{t-s\downarrow 0} \frac{1}{t-s} \int_{|y-x|\leq\varepsilon} (y-x)p(s,x;t,\mathrm{d}y) = b(s,x).$$

(3) (扩散系数的定义). 存在函数$\Sigma(x,s)$使得，对于任意x和$\epsilon > 0$,

$$\lim_{t-s\downarrow 0} \frac{1}{t-s} \int_{|y-x|\leq\varepsilon} (y-x)^2 p(s,x;t,\mathrm{d}y) = \Sigma(x,s).$$

注．在物理学上，一阶矩主要描述动力学趋势，二阶矩主要给出随机波动。

Itô扩散过程

接下来我们看看Itô用SDE构造的Itô扩散过程和经典的是否一致，也就是说我们下面的定理：它说明了Itô扩散过程的生成元就是扩散算子。

定理 2

设X_t为一个生成元为$\mathcal{A}$的时齐*Itô*扩散过程

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

那么对于所有的二次连续可微函数 f，有

$$\mathcal{A}f = \mathcal{L}f,$$

其中

$$\begin{aligned}\mathcal{L}f &= \sum_i b_i(x)\frac{\partial}{\partial x_i}f + \frac{1}{2}\sum_{i,j}\Sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}f \\ &= b(x)\cdot\nabla f + \frac{1}{2}Tr(\Sigma^T(x)\nabla^2 f),\end{aligned}$$

并且 $\Sigma_{ij}(x) = \sum_{k=1}^n \sigma_{ik}(x)\sigma_{kj}(x) = \sigma(x)\sigma^T(x)$.

Itô扩散过程

注. 之后我们继续使用记号$\mathcal{L}$表示Itô 扩散过程X_t的生成元。记

$$u(t,x) \triangleq \mathbb{E}[f(X_t)|X_0 = x]$$

那么$u(t,x)$满足Kolmogorov向后方程

$$\frac{\partial u(t,x)}{\partial t} = \mathcal{L}u(t,x) = b(x)\cdot\nabla u(t,x) + \frac{1}{2}Tr(\Sigma^T(x)\nabla^2 u(t,x)), \quad t > 0.$$

回忆连续时间Markov链X_t的Kolmogorov向前方程：

$$\frac{\mathrm{d}}{\mathrm{d}t}P(t) = P(t)Q$$

设μ_t为X_t在t时刻的概率分布，$\mu_t = \mu_0 P(t)$，于是向前方程可以写成关于分布的演化：

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t = \mu_t Q.$$

这即是物理学家常说的 Fokker-Planck 方程。

Fokker-Planck 方程

定理 3 (Fokker-Planck 方程)

设X_t为一个时齐 *Itô* 扩散过程

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

并且 $\mu_t(x) = \mu(x,t)$ 为 X_t 在 t 时刻的概率分布，且$\mu_t(x)$关于 x 和 t 二次连续可微，设$\mu(x)$为初始概率分布，那么$\mu(x,t)$是下面方程的解

$$\partial_t \mu(x,t) = \mathcal{L}^* \mu(x,t) \quad (t>0), \quad \mu(x,0) = \mu(x),$$

其中算子

$$\begin{aligned} \mathcal{L}^* g &\triangleq -\sum_j \frac{\partial}{\partial x_j}(b_j(x)g) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}(\Sigma_{ij}(x)g) \\ &= -\nabla \cdot (b(x)g) + \frac{1}{2}\nabla \cdot \nabla \cdot (\Sigma(x)g) \end{aligned}$$

为$\mathcal{L}$的伴随算子。

Fokker-Planck 方程

注. 运用 Fokker-Planck 方程，我们可以快速计算不变测度。由不变测度的定义：

$$\int_E P_t f d\pi = \int_E f d\pi.$$

等价地，

$$P_t^* \pi = \pi$$

那么由 Fokker-Planck 方程

$$\mathcal{L}^* \pi = 0$$

由此可以解得不变测度 π.

目录

Ornstein–Uhlenbeck过程

设$(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathrm{P})$为一个带流的概率空间，$B_t$ 为一个$\mathscr{F}_t$-布朗运动，回顾OU过程

$$dX_t = -bX_t dt + \sigma dB_t, \quad X_0,$$

我们计算过它的解为

$$X_t = X_0 e^{-bt} + \sigma e^{-bt} \int_0^t e^{bs} dB_s$$

本节我们考察一个具体地OU过程，

$$dX_t = -X_t dt + \sqrt{2} dB_t. \tag{2}$$

那么它的解为

$$\begin{aligned} X_t &= e^{-t} X_0 + \sqrt{2} e^{-t} \int_0^t e^s dB_s \\ &\overset{d}{=} e^{-t} X_0 + e^{-t} B_{e^{2t}-1} \qquad \text{等号在分布意义下成立} \\ &\overset{d}{=} e^{-t} X_0 + \sqrt{1 - e^{-2t}} z \qquad z \sim N(0,1) \\ &\sim N(e^{-t} X_0, 1 - e^{-2t}) \end{aligned}$$

OU算子

我们已经证明Itô扩散过程

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

的生成元为

$$\begin{aligned}\mathcal{L}f &= \sum_i b_i(x)\frac{\partial}{\partial x_i}f + \frac{1}{2}\sum_{i,j}\Sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}f \\ &= b(x)\cdot\nabla f + \frac{1}{2}Tr(\Sigma^T(x)\nabla^2 f)\end{aligned}$$

对于OU过程 $dX_t = -X_t dt + \sqrt{2}dB_t$，其生成元OU算子为

$$\mathcal{L}_{OU}f = -x\cdot\nabla f + \Delta f.$$

(按生成元定义证明，留作作业)，其伴随算子为

$$\mathcal{L}_{OU}^* g = \nabla\cdot(xg) + \Delta g.$$

OU过程的不变测度

解方程

$$\mathcal{L}_{OU}^{*}\pi = \nabla \cdot (x\pi) + \Delta\pi = 0,$$

可以得到OU过程的不变测度

$$d\pi(x) \propto e^{-\frac{1}{2}x^2}dx$$

这即是标准高斯测度 $N(0, I)$。

注. 一维情形课后练习：

$$\frac{d}{dx}(x\pi(x)) + \frac{d^2\pi(x)}{dx^2} = 0.$$

转移函数与OU半群

从OU过程的解

$$X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}z \qquad z \sim N(0,1)$$

可以得到OU过程的转移函数

$$p(t,x,y) = \frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{(y-xe^{-t})^2}{2(1-e^{-2t})}\right).$$

那么转移半群为

$$\begin{aligned} P_t f(x) &= \mathbb{E}[f(X_t) \mid X_0 = x] = \int f(y)p(t,x,dy) \\ &= \mathbb{E}[f(e^{-t}x + \sqrt{1-e^{-2t}}z)] \qquad z \sim N(0,1) \\ &= \int f(e^{-t}x + \sqrt{1-e^{-2t}}z)d\pi(z) \end{aligned}$$

称为OU半群。

Kolmogorov向前向后方程

OU过程的Kolmogorov向后方程：

$$\frac{\partial}{\partial t} P_t f(x) = \mathcal{L}_{OU} P_t f(x) = -x \cdot \nabla P_t f(x) + \Delta P_t f(x)$$

Fokker-Planck方程：

$$\partial_t \mu(x,t) = \mathcal{L}_{OU}^* \mu(x,t) = \nabla \cdot (x\mu(x,t)) + \Delta \mu(x,t)$$

更多参考：Bakry D, Gentil I, Ledoux M. Analysis and geometry of Markov diffusion operators[M]. Springer Science & Business Media, 2013.

目录

Langevin Dynamics

定义 6 (Langevin Dynamics)

给定能量函数 $V(x)$，*Langevin Dynamics*是如下形式的*SDE*

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \tag{3}$$

其解一般称为 *Langevin* 扩散。

Langevin 扩散的生成元为

$$\mathcal{L}_{LD} f = -\nabla V \cdot \nabla f + \Delta f$$

生成元的伴随为

$$\mathcal{L}_{LD}^* g = \nabla \cdot (g \nabla V) + \Delta g$$

Kolmogorov向后方程：

$$\frac{\partial}{\partial t} P_t f(x) = \mathcal{L}_{LD} P_t f(x) = -\nabla V(x) \cdot \nabla P_t f(x) + \Delta P_t f(x)$$

Fokker-Planck方程：

$$\partial_t \mu(x,t) = \mathcal{L}_{LD}^* \mu(x,t) = \nabla \cdot (\mu(x,t) \nabla V(x)) + \Delta \mu(x,t)$$

Langevin Dynamics的不变测度

命题 1

Langevin 扩散

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$$

的不变测度为

$$d\pi(x) \propto e^{-V(x)}dx$$

Proof.

注. 事实上，对于任意给定连续分布$p(x)$，我们可以把它写成

$$p(x) = e^{\log p(x)}$$

这表明通过适当的选取能量函数$V(x)$，我们可以构造 Langevin 扩散使得它的不变测度是目标分布$p(x)$，也就是说，通过Langevin Dynamics的迭代，我们可以从服从任意分布的样本X_0产生服从目标分布$p(x)$的样本X_T。

Mixing 速度

我们首先给出刻画分布之间距离的数学量：

定义 7 (Wasserstein距离)

概率测度 μ 和 ν 之间的*2-Wasserstein* 距离定义为：

$$W_2(\mu,\nu) := \inf_{\gamma\in\mathcal{C}(\mu,\nu)} \left(\int \|x-y\|^2 \gamma(\mathrm{d}x,\mathrm{d}y)\right)^{\frac{1}{2}}. \tag{4}$$

其中$\mathcal{C}(\mu,\nu)$是μ和ν的所有耦合*(Couplings)*构成的空间，$\|\cdot\|$是欧式范数。

注1. 我们称 $\gamma(x,y)$ 是测度 $\mu(x)$ 和 $\nu(y)$ 的耦合，如果它关于 x 的边缘分布是 μ，关于 y 的边缘分布是 ν。

注2. 我们也可以把W_2推广到W_p

$$W_p(\mu,\nu) := \inf_{\gamma\in\mathcal{C}(\mu,\nu)} \left(\int \|x-y\|^p \gamma(\mathrm{d}x,\mathrm{d}y)\right)^{\frac{1}{p}}.$$

Mixing 速度

Condition. 设$\alpha > 0$为一个常数，称一个测度$\mu \propto e^{-V}$是α-强 log-concave 的，如果V是α-强凸的，即

$$\nabla^2 V \succeq \alpha I$$

定理 4

设 X_t 为初值为 $X_0 \sim \mu_0$，平稳分布为 $\mu \propto e^{-V}$ 的*Langevin*扩散，假设 μ 是 α-强 *log-concave* 的，那么

$$W_2^2(\mu_t, \mu) \leq \exp(-2\alpha t) W_2^2(\mu_0, \mu).$$

Proof.