概率 · 动力学 · 偏微分方程

随机微积分
课程笔记

从布朗运动到分枝扩散

Brownian Motion, SDEs, Itô Calculus,
Fokker–Planck Equations and Branching Diffusions

这份讲义从“怎样描述一条随机轨道”出发，逐步建立随机积分与随机微分的计算规则，再把轨道层面的模型转化为概率密度与种群分布的演化方程。

- 用二次变差理解随机链式法则中的修正项；
- 用生成元把 SDE、期望与 PDE 联系起来；
- 用概率流理解边界条件、平稳分布与质量守恒；
- 用一阶矩与归一化区分分枝、选择和随机种群比例。

中文详解版 · 含证明、例题、数值实验及习题解答

依据所提供的英文草稿重组、扩写与校正

目录

1 学习目标与全篇路线 **1**

1.1 同一个系统的三种描述 1

1.2 阅读层次与建议课时 1

1.3 符号与约定 2

1.4 怎样理解讲义中的证明 2

2 概率预备：随机性、时间与信息 **2**

2.1 概率空间与随机过程 3

2.2 滤过与非预见性 3

2.3 条件期望：给定已有信息后的最佳平均预测 3

2.4 鞅与局部鞅 4

3 布朗运动：连续但粗糙的随机驱动 **4**

3.1 相对于滤过的定义 4

3.2 为什么随机游走要按平方根缩放 5

3.3 协方差与鞅性质 5

3.4 二次变差：二阶项为什么不能扔掉 6

3.5 不可微与白噪声的正确理解 6

4 Itô 随机积分：如何对布朗运动积分 **7**

4.1 从左端点和开始 7

4.2 Itô 等距及其证明 7

4.3 从简单过程延拓到一般过程 8

4.4 随机积分的性质与局部化 9

4.5 核心例子：为什么积分 $W\,\mathrm{d}W$ 多出一个时间项 9

5 随机微分方程：从微分记号到随机轨道 **10**

5.1 真正的定义是积分方程 10

5.2 强解、弱解与两种唯一性 10

5.3 存在唯一性：条件与证明框架 10

5.4 例一：带漂移的布朗运动 11

5.5 例二：Ornstein–Uhlenbeck 过程的完整计算 11

5.6 OU 的长期分布不等于每条路径收敛到均值 12

6 Itô 公式：随机微积分的链式法则 **13**

6.1 为什么普通链式法则会少一项 . . . 13
6.2 公式、条件与积分形式 . . . 13
6.3 从泰勒展开到公式：证明框架 . . . 14
6.4 二次协变差与乘积法则 . . . 14
6.5 多维 Itô 公式与扩散矩阵 . . . 15
6.6 典型例子：多项式、指数与几何布朗运动 . . . 16

7 Stratonovich 积分与两种 SDE 的转换 17
7.1 对称和及其与 Itô 积分的关系 . . . 17
7.2 一维转换：逐项计算修正漂移 . . . 18
7.3 为什么 Stratonovich 形式恢复普通链式法则 . . . 19
7.4 多维转换：按噪声方向求导 . . . 19

8 无穷小生成元：从随机轨道到期望变化 19
8.1 短时条件期望与生成元 . . . 20
8.2 Dynkin 公式与鞅分解 . . . 20
8.3 矩方程：何时可以闭合 . . . 20

9 Kolmogorov 后向方程：预测未来的观测量 21
9.1 时间齐次情形：以经过时间为变量 . . . 21
9.2 终端值问题：以日历时间为变量 . . . 21

10 Fokker–Planck 方程：概率分布如何演化 22
10.1 先写对分布总是更自然的弱形式 . . . 22
10.2 逐步推导一维前向方程 . . . 23
10.3 前向与后向：同一个转移核的两侧 . . . 23
10.4 概率流、质量与边界条件 . . . 24
10.5 多维前向方程 . . . 24
10.6 例一：布朗运动、热核与扩散系数 . . . 25
10.7 例二：带漂移的高斯传播 . . . 25
10.8 零流平稳分布：一般一维公式 . . . 25
10.9 例三：梯度漂移与 Gibbs 型密度 . . . 26
10.10 直接从 Stratonovich 方程写前向方程 . . . 26
10.11 解题时的统一工作顺序 . . . 27

11 分枝扩散：从粒子数到归一化分布 27

11.1 先区分三个不同的对象 . . . 27
11.2 模型假设与净增长率 . . . 28
11.3 用测试函数推导一阶矩方程 . . . 28
11.4 归一化为何产生非线性 . . . 29
11.5 一个必须澄清的抽样区别 . . . 30
11.6 Feynman–Kac：用一条路径计算整个种群的一阶矩 . . . 31
11.7 无运动时：复制子方程与显式解 . . . 32
11.8 均值如何改变：运动项与协方差项 . . . 32
11.9 平稳分布与特征值问题 . . . 33
11.10 非守恒边界会改变归一化项 . . . 33
11.11 计算思路：运动、加权、再归一化 . . . 34

12 把公式变成数值实验 34
12.1 Euler–Maruyama：噪声的尺度不能写错 . . . 34
12.2 强误差、弱误差和 Monte Carlo 误差 . . . 35
12.3 Milstein 格式为什么多一项 . . . 35
12.4 实验一：轨道很不规则，二次变差却稳定 . . . 35
12.5 实验二：用粒子直方图检查 Fokker–Planck 密度 . . . 36
12.6 实验三：同一条布朗路径上的数值方法比较 . . . 36
12.7 OU 的离散平稳分布：一个额外的检查 . . . 37

13 选读：与扩散生成模型的联系 37
13.1 从 Fokker–Planck 直接得到概率流速度 . . . 38
13.2 反向 SDE：先定义一个递增的反向时钟 . . . 38
13.3 用一个高斯例子同时核对三个模型 . . . 38

14 综合习题与解答 39

A 核心公式速查 42

B 常见错误与自查顺序 43

C 进一步阅读 44

1 学习目标与全篇路线

1.1 同一个系统的三种描述

设一个粒子受到确定性作用力和随机扰动。轨道描述写为

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t.$$

这里 X_t 是随机状态，b 是漂移，σ 是噪声振幅，W_t 是标准布朗运动。把同一个初始条件重复实验许多次，可以观察状态的分布。若它有密度 $p(x,t)$，则密度满足

$$\partial_t p = -\partial_x(bp) + \tfrac{1}{2}\partial_{xx}(\sigma^2 p).$$

若粒子还会复制、死亡，期望数密度 $f(x,t)$ 则满足

$$\partial_t f = L_t^* f + rf.$$

这三行看似分属概率论、偏微分方程和种群动力学，其实由同一条逻辑链连接：

$$\boxed{\text{布朗运动}} \longrightarrow \boxed{\text{随机积分与 SDE}} \longrightarrow \boxed{\text{Itô 公式}}$$
$$\longrightarrow \boxed{\text{生成元}L_t} \longrightarrow \boxed{\text{伴随算子}L_t^* \text{ 与密度方程}}.$$

其中最关键的桥梁是

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\varphi(X_t)] = \mathbb{E}[(L_t\varphi)(X_t)] = \int (L_t\varphi)(x)p(x,t)\,\mathrm{d}x. \tag{1.1}$$

把右端的微分从测试函数 φ 转移到密度 p，就得到前向方程。若反过来固定终端观测、改变出发位置，则得到后向方程。

1.2 阅读层次与建议课时

本讲义面向学过多元微积分、线性代数、基础概率与常微分方程的读者。测度论不是开始阅读的硬性门槛，但严谨理解随机积分需要可测性、条件期望和 L^2 收敛。前面的概率预备会说明这些概念在本课程中具体承担什么作用。

课次	主题	应当能够独立完成的任务
1–2	概率预备、布朗运动与二次变差	证明协方差公式与二次变差的 L^2 收敛
3–4	随机积分与 SDE	使用等距公式；解出 OU 过程并计算均值、方差
5–6	随机微积分与积分约定	正确使用链式、乘积法则和漂移转换
7–8	生成元、前向与后向方程	由 SDE 推导 PDE；写出概率流与边界条件
9–10	分枝扩散与归一化	推导选择项，解释两种抽样分布的区别
11–12	数值实验、综合练习与选读	检验模拟，理解强弱误差及扩散模型的联系

每次课可安排约 90 分钟；这是组织材料的建议，不表示每一节都需要相同时间。第一次阅读可先把握主公式和三个核心例子：布朗运动、OU 过程、几何布朗运动。第二次阅读再补可测性、局部化、弱形式与边界条件。

1.3 符号与约定

符号	含义与需要注意的区别
$(\Omega, \mathcal{F}, \mathbb{P})$	概率空间；$\omega \in \Omega$ 表示一次随机实现
$(\mathcal{F}_t)_{t\geq 0}$	随时间增长的信息；默认满足完备性与右连续性
W_t	一维标准布朗运动；多维时各分量为独立标准布朗运动
b, σ, Σ	Itô 漂移、一维噪声系数、多维噪声矩阵
$a = \Sigma\Sigma^{\mathsf{T}}$	瞬时协方差矩阵；一维时 $a = \sigma^2$
L_t, L_t^*	作用于观测函数的生成元、作用于测度或密度的伴随算子
p, f, M	概率密度、期望粒子数密度、期望总粒子数
$r, \bar{r}$	净增长率、相对于当前归一化密度的平均净增长率
$C^{2,1}$	本讲义按 (x, t) 排列：空间二阶、时间一阶导数连续
$\langle \nu, \varphi \rangle$	测度与函数的配对 $\int \varphi \, \mathrm{d}\nu$

除非明确写出 $\circ\,\mathrm{d}W_t$，随机微分方程一律按 Itô 意义解释。对密度进行分部积分时，会说明无穷远衰减、无流边界或测试函数紧支撑等条件。“在每个有限时间区间上成立”与“存在跨越所有时间的统一常数”是不同要求。

1.4 怎样理解讲义中的证明

二次变差、等距公式、显式解、伴随算子以及归一化方程都会给出可逐行检查的推导。涉及布朗运动的严格构造、几乎处处路径性质、随机积分的完备化技术、一般 SDE 的存在理论时，会标明“证明框架”或“结论，证明略”。符号 $(\mathrm{d}W_t)^2 = \mathrm{d}t$ 是二次变差的计算缩写，不能把它当作普通无穷小数的代数恒等式。

对原始材料主要作了四类整理：统一漂移与扩散的符号；合并重复章节并修复 LaTeX 文档结构；补充积分与鞅结论所需的假设；校正分枝系统归一化后的抽样解释。数值实验和扩散生成模型的联系为新增内容，后者可在掌握主线后选读。

2 概率预备：随机性、时间与信息

随机微分方程研究的不是一条预先确定的曲线，而是一族可能发生的轨道。在写下 $\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t$ 之前，必须先说明三个问题：随机结果属于什么集合，时刻 t 已经知道哪些信息，以及哪些量允许进入当前的计算。

2.1 概率空间与随机过程

概率空间记为 $(\Omega, \mathcal{F}, \mathbb{P})$。其中 $\omega \in \Omega$ 表示一个完整的随机结果，$\mathcal{F}$ 是可以赋予概率的事件所组成的 σ-代数，$\mathbb{P}$ 是概率测度。随机变量 $X\colon \Omega \to \mathbb{R}$ 是可测映射；随机过程则是一族随机变量 $X = (X_t)_{t\geq 0}$。

同一个过程有两个互补的观察角度：固定 t，X_t 是一个随机变量，可以讨论它的分布；固定 ω，映射 $t \mapsto X_t(\omega)$ 是一条样本路径。例如，$X_t \sim \mathcal{N}(0,t)$ 只是每个时刻的边缘分布信息，不能单独确定整条随机路径的联合分布。后面出现的 SDE 主要描述路径，Fokker–Planck 方程主要描述边缘分布。

2.2 滤过与非预见性

定义 2.1 (滤过与适应过程)**.** 滤过 $(\mathcal{F}_t)_{t\geq 0}$ 是一族递增的 σ-代数：若 $s \leq t$，则 $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$。它表示信息随时间积累。若对每个 t，X_t 都是 $\mathcal{F}_t$-可测的，就称 X 适应于该滤过。

自然滤过 $\sigma(X_s : 0 \leq s \leq t)$ 只包含截至 t 的观测信息。通常还对它作完备化与右连续化；本文所说的通常条件就是滤过完备且 $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$。适应性的直觉是：在时刻 t 作出的决定可以使用过去与当前的信息，但不能偷看未来。例如 $H_t = W_t$ 是适应的，而 $H_t = W_T$ 在 $t < T$ 时通常不是适应的。

适应性只逐时刻描述可测性，构造时间积分时还要控制 (t,ω) 的联合可测性。

定义 2.2 (渐进可测与可预测)**.** 过程 H 称为渐进可测，如果对任意 $T > 0$，限制在 $[0,T]\times\Omega$ 上的映射 $(t,\omega) \mapsto H_t(\omega)$ 对 $\mathcal{B}([0,T]) \otimes \mathcal{F}_T$ 可测。可预测 σ-代数由所有左连续适应过程生成；对这个 σ-代数可测的过程称为可预测过程。

可预测过程是渐进可测的，渐进可测过程是适应的。连续的适应过程同时是渐进可测和可预测的，因而本讲义的大多数 SDE 例子不会遇到额外的可测性困难。但在定理中，不能把“适应”直接当作完整的随机可积条件。

2.3 条件期望：给定已有信息后的最佳平均预测

若 Y 可积，$\mathcal{G} \subseteq \mathcal{F}$，条件期望 $\mathbb{E}[Y \mid \mathcal{G}]$ 是一个 $\mathcal{G}$-可测随机变量，并满足

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}[Y \mid \mathcal{G}]] = \mathbb{E}[\mathbf{1}_A Y], \qquad A \in \mathcal{G}.$$

它不是把随机性全部去掉，而是只平均掉 $\mathcal{G}$ 尚未揭示的随机性。后面最常用的三个规则是

$$\begin{aligned} \mathbb{E}[ZY \mid \mathcal{G}] &= Z\mathbb{E}[Y \mid \mathcal{G}], && Z \text{ 为 } \mathcal{G}\text{-可测且满足所需可积性,} \\ \mathbb{E}[Y \mid \mathcal{G}] &= \mathbb{E}[Y], && Y \text{ 与 } \mathcal{G} \text{ 独立,} \\ \mathbb{E}[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}] &= \mathbb{E}[Y \mid \mathcal{G}], && \mathcal{G} \subseteq \mathcal{H}. \end{aligned}$$

最后一条称为塔式性质。证明随机积分中交叉项的期望为零时，关键往往就是先对适当的过去信息取条件期望。

2.4 鞅与局部鞅

定义 2.3 (鞅)**.** 适应过程 M 称为鞅，如果 $\mathbb{E}|M_t| < \infty$，且对任意 $0 \le s \le t$，

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s.$$

这表示在已知当前信息后，对未来值的平均预测就是当前值。它不意味着路径保持不变，也不意味着未来值与过去独立。由全期望公式，鞅满足 $\mathbb{E}[M_t] = \mathbb{E}[M_0]$，但反过来“均值不变”并不能保证鞅性质。

定义 2.4 (停止时刻与局部鞅)**.** 随机时间 τ 称为停止时刻，如果 $\{\tau \le t\} \in \mathcal{F}_t$。连续适应过程 M 称为局部鞅，如果存在停止时刻 $\tau_n \uparrow \infty$ 几乎必然，使得每个停止过程 $M_{t\wedge\tau_n}$ 都是鞅。

停止时刻可以理解为一个仅根据已获得的信息判断是否停止的规则。例如连续适应过程第一次离开区间 $[-n, n]$ 的时间就是停止时刻。局部鞅允许先把过程限制在可控范围内，再使用鞅性质。从停止后的结论去掉停止，往往需要一致可积性、平方可积估计或其他控制。

说明 2.5 (后面反复使用的逻辑)**.** 对 $M_t = M_0 + \int_0^t H_s \,\mathrm{d}W_s$，假设 M_0 可积，一般首先得到的是局部鞅。若进一步有 $M_0 \in L^2$（例如 $M_0 = 0$），且 $\mathbb{E}\int_0^T H_t^2 \,\mathrm{d}t < \infty$，则在 $[0, T]$ 上它是平方可积鞅。因此，“没有 $\mathrm{d}t$ 项”与“可以直接取期望并得到常数”之间还差一个可积性检查。

3 布朗运动：连续但粗糙的随机驱动

3.1 相对于滤过的定义

定义 3.1 (标准布朗运动)**.** 实值过程 $W = (W_t)_{t\ge 0}$ 称为相对于 $(\mathcal{F}_t)$ 的标准布朗运动，如果

(i) W 适应于 $(\mathcal{F}_t)$，且 $W_0 = 0$ 几乎必然；

(ii) 样本路径几乎必然连续；

(iii) 对任意 $0 \le s < t$，$W_t - W_s$ 与 $\mathcal{F}_s$ 独立，且 $W_t - W_s \sim \mathcal{N}(0, t-s)$。

第三条同时包含高斯增量、时间平稳增量和对过去信息的独立性。只要求不相交时间区间上的增量彼此独立，可以定义过程本身的布朗性质；若还要在一个指定的滤过下做随机积分，就必须明确增量与该滤过的过去信息独立。过度扩大滤过，例如在时刻 0 就加入未来的 W_T，会破坏这一性质。

由定义立即得到

$$W_t \sim \mathcal{N}(0, t), \qquad \mathbb{E}[W_t] = 0, \qquad \mathrm{Var}(W_t) = t, \qquad \mathbb{E}[(W_t - W_s)^2] = |t - s|.$$

增量的典型尺度是 $\sqrt{t-s}$，而不是 $t - s$。这里的“尺度”是分布或均方意义的描述，并不是逐条路径的确定上界。

3.2 为什么随机游走要按平方根缩放

令 $\xi_1, \xi_2, \dots$ 独立同分布，且 $\mathbb{P}(\xi_k = 1) = \mathbb{P}(\xi_k = -1) = 1/2$。定义

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k.$$

在长度 t 的时间内大约走 nt 步，每一步的空间大小是 $n^{-1/2}$，所以

$$\mathrm{Var}(S_n(t)) = \frac{\lfloor nt \rfloor}{n} \longrightarrow t.$$

若把空间大小改成 n^{-1}，方差会趋于零；若不缩放，方差又会发散。平方根缩放正好保留非平凡的随机波动。中心极限定理解释了固定时刻的高斯极限；更强的函数型中心极限定理说明，作线性插值后，整个过程在有限时间区间的连续函数空间中依分布收敛到布朗运动。这一函数型收敛还需要紧性论证，不能只由单个时刻的中心极限定理推出。

同样的尺度关系给出布朗运动的缩放性质：对任意 $c > 0$，

$$\widetilde{W}_t = \frac{1}{\sqrt{c}} W_{ct}$$

仍是标准布朗运动。这可以直接检查它的连续性和独立高斯增量。

3.3 协方差与鞅性质

命题 3.2. 对任意 $s, t \geq 0$，$\mathrm{Cov}(W_s, W_t) = \min(s, t)$。

证明. 不妨设 $s \leq t$。将 W_t 分解为 $W_s + (W_t - W_s)$，得到

$$\mathbb{E}[W_s W_t] = \mathbb{E}[W_s^2] + \mathbb{E}[W_s(W_t - W_s)].$$

第二项等于 $\mathbb{E}[W_s \mathbb{E}[W_t - W_s \mid \mathcal{F}_s]] = 0$，而第一项为 s。由于两个时刻的均值都为零，协方差就是 s。 □

说明 3.3 (独立的是增量，不是取值)**.** 当 $0 < s < t$ 时，W_s 与 W_t 的协方差为 $s > 0$，因此它们不独立。未来的位置包含过去已经积累的位移；与过去独立的是新增位移 $W_t - W_s$。

命题 3.4. W_t 与 $W_t^2 - t$ 都是鞅。

证明. 首先，

$$\mathbb{E}[W_t \mid \mathcal{F}_s] = W_s + \mathbb{E}[W_t - W_s \mid \mathcal{F}_s] = W_s.$$

其次，将 W_t^2 展开为 $W_s^2 + 2W_s(W_t - W_s) + (W_t - W_s)^2$，条件期望为

$$\mathbb{E}[W_t^2 \mid \mathcal{F}_s] = W_s^2 + (t - s).$$

减去 t 即得 $\mathbb{E}[W_t^2 - t \mid \mathcal{F}_s] = W_s^2 - s$。所需可积性由高斯二阶矩保证。 □

第二个鞅揭示了一个重要事实：W_t^2 本身平均增长，增长速度恰好是 1。后面的 Itô 修正项正是这种二阶效应的微分表达。

3.4 二次变差：二阶项为什么不能扔掉

固定 $T>0$。对确定性分割 $\Pi=\{0=t_0<t_1<\cdots<t_N=T\}$，记

$$|\Pi|=\max_k(t_{k+1}-t_k),\qquad Q_\Pi(W)=\sum_{k=0}^{N-1}(W_{t_{k+1}}-W_{t_k})^2.$$

定理 3.5 (布朗运动的二次变差)**.** 若确定性分割 Π_n 的网格大小趋于零，则

$$Q_{\Pi_n}(W)\longrightarrow T\qquad \text{在}L^2\text{ 中成立.}$$

特别地，该收敛也依概率成立。

证明. 记 $\Delta_kW=W_{t_{k+1}}-W_{t_k}$、$\Delta_kt=t_{k+1}-t_k$。由独立高斯增量及高斯矩公式，

$$\mathbb{E}[(\Delta_kW)^2]=\Delta_kt,\qquad \mathbb{E}[(\Delta_kW)^4]=3(\Delta_kt)^2.$$

因此

$$\mathbb{E}[Q_\Pi(W)]=\sum_k\Delta_kt=T,\qquad \mathrm{Var}((\Delta_kW)^2)=2(\Delta_kt)^2.$$

不同区间增量的平方仍然独立，所以

$$\begin{aligned}\mathbb{E}\big[(Q_\Pi(W)-T)^2\big]&=\mathrm{Var}(Q_\Pi(W))=2\sum_k(\Delta_kt)^2\\&\le 2|\Pi|\sum_k\Delta_kt=2T|\Pi|\longrightarrow 0.\end{aligned}$$

□

通常将这个结果记作 $[W]_T=T$。对于二进等分 $t_k=kT/2^n$，还有几乎必然收敛：因为

$$\sum_{n=1}^{\infty}\mathbb{P}\big(|Q_{\Pi_n}(W)-T|>\varepsilon\big)\le\frac{2T^2}{\varepsilon^2}\sum_{n=1}^{\infty}2^{-n}<\infty,$$

由 Borel–Cantelli 引理，再对可数个 $\varepsilon>0$ 使用这一结论即可。这里区分了两种强度：任意给定的确定性细化分割可得 L^2 收敛；上述估计还给出了二进分割上的几乎必然收敛。

对比一条连续且有限变差的确定性路径 x。若它的总变差为 $V_T(x)<\infty$，则

$$\sum_k(x_{t_{k+1}}-x_{t_k})^2\le\max_k|x_{t_{k+1}}-x_{t_k}|\,V_T(x)\longrightarrow 0.$$

布朗路径沿二进分割的二次变差为 $T>0$，所以它在任何非零长度的固定时间区间上几乎必然不是有限变差路径。这解释了为什么不能直接把布朗运动当成普通有限变差积分器。

3.5 不可微与白噪声的正确理解

布朗路径几乎必然连续且处处不可微。这是一个需要专门证明的路径性质。下面只给出理解它的尺度直觉，而不是证明：对固定 t 与 $h>0$，

$$\frac{W_{t+h}-W_t}{h}\sim\mathcal{N}(0,h^{-1}).$$

当 $h \downarrow 0$，差商的分布尺度越来越大，因而不呈现普通光滑曲线的局部斜率行为。需要注意，方差发散本身并不构成“几乎必然处处不可微”的完整证明。

形式上常把 $\xi_t = \dot{W}_t$ 称为白噪声，但它不是逐点定义的普通随机函数。更合适的定义是：对确定性测试函数 $\varphi \in L^2([0,T])$，令

$$\xi(\varphi) := \int_0^T \varphi(t)\,\mathrm{d}W_t.$$

这是一族中心高斯随机变量，满足

$$\mathbb{E}[\xi(\varphi)\xi(\psi)] = \int_0^T \varphi(t)\psi(t)\,\mathrm{d}t.$$

因此物理中的 $\mathbb{E}[\xi_t\xi_s] = \delta(t-s)$ 应理解为对测试函数积分后的恒等式。写 $\dot{X} = b + \sigma\xi$ 只是便于记忆的表示，数学定义仍来自后面的随机积分方程。

4 Itô 随机积分：如何对布朗运动积分

4.1 从左端点和开始

在确定性积分中，只要函数足够光滑，左端点、右端点与中点取样的区别会随网格消失。布朗运动的二次变差不为零，所以随机积分必须明确取样约定。Itô 积分使用每个小区间开始时已知的系数。

定义 4.1 (简单可预测过程)**.** 在 $[0,T]$ 上，令

$$H_s = \sum_{k=0}^{N-1} \xi_k \mathbf{1}_{(t_k,t_{k+1}]}(s), \qquad 0 = t_0 < \cdots < t_N = T,$$

其中 ξ_k 为 $\mathcal{F}_{t_k}$-可测且 $\mathbb{E}[\xi_k^2] < \infty$。定义

$$I_t(H) := \int_0^t H_s\,\mathrm{d}W_s = \sum_{k=0}^{N-1} \xi_k\big(W_{t_{k+1}\wedge t} - W_{t_k\wedge t}\big), \qquad 0 \le t \le T.$$

每一项表示：用区间起点已有信息决定系数，再乘以该区间随后发生的随机增量。随机积分的积分器是 $\mathrm{d}W_s$，不是 $W_s\,\mathrm{d}s$；两者的量纲和性质完全不同。

4.2 Itô 等距及其证明

定理 4.2 (Itô 等距)**.** 对上述简单过程及任意 $t \in [0,T]$，

$$\boxed{\mathbb{E}\left[\left(\int_0^t H_s\,\mathrm{d}W_s\right)^2\right] = \mathbb{E}\int_0^t H_s^2\,\mathrm{d}s.}$$

并且 $\mathbb{E}[I_t(H)] = 0$。

证明. 先证 $t=T$，记 $\Delta_k W = W_{t_{k+1}} - W_{t_k}$。对角项由条件期望计算：

$$\mathbb{E}[\xi_k^2(\Delta_k W)^2] = \mathbb{E}\big[\xi_k^2 \mathbb{E}[(\Delta_k W)^2 \mid \mathcal{F}_{t_k}]\big] = \mathbb{E}[\xi_k^2](t_{k+1}-t_k).$$

当 $i<j$ 时，$\xi_i \Delta_i W\, \xi_j$ 是 $\mathcal{F}_{t_j}$-可测的，故

$$\mathbb{E}[\xi_i \Delta_i W\, \xi_j \Delta_j W] = \mathbb{E}\big[\xi_i \Delta_i W\, \xi_j \mathbb{E}[\Delta_j W \mid \mathcal{F}_{t_j}]\big] = 0.$$

因此，展开平方后只剩对角项：

$$\mathbb{E}[I_T(H)^2] = \sum_k \mathbb{E}[\xi_k^2](t_{k+1}-t_k) = \mathbb{E}\int_0^T H_s^2 \,\mathrm{d}s.$$

同样地，$\mathbb{E}[\xi_k \Delta_k W] = 0$，故积分均值为零。若某个系数不有界，可先截断系数，再用 L^2 收敛证明以上计算；这也避免了对条件期望中间乘积的可积性作隐含假设。一般 t 的情况对截断过程 $H_s \mathbf{1}_{(0,t]}(s)$ 应用同一结论即可。 □

等距式把难处理的随机积分平方，变成普通时间积分。例如，若 $H_s = s$，则

$$\mathbb{E}\left[\left(\int_0^t s\,\mathrm{d}W_s\right)^2\right] = \int_0^t s^2\,\mathrm{d}s = \frac{t^3}{3}.$$

由于这里的被积函数是确定性的，积分还是高斯变量：$\int_0^t s\,\mathrm{d}W_s \sim \mathcal{N}(0, t^3/3)$。一般随机被积函数的 Itô 积分则不必服从高斯分布。

4.3 从简单过程延拓到一般过程

定义平方可积可预测过程空间

$$\mathcal{H}^2([0,T]) = \left\{ H:\ H \text{ 可预测},\quad \|H\|_{\mathcal{H}^2}^2 := \mathbb{E}\int_0^T H_s^2\,\mathrm{d}s < \infty \right\},$$

其中相差一个 $\mathrm{d}s \otimes \mathbb{P}$-零测集的过程视为同一个元素。简单可预测过程在这个空间中稠密。取 $H^{(n)}$ 满足 $\mathbb{E}\int_0^T |H_s^{(n)} - H_s|^2\,\mathrm{d}s \to 0$，则等距式给出

$$\mathbb{E}\left| I_T(H^{(n)}) - I_T(H^{(m)}) \right|^2 = \mathbb{E}\int_0^T |H_s^{(n)} - H_s^{(m)}|^2\,\mathrm{d}s \longrightarrow 0.$$

因此 $I_T(H^{(n)})$ 是 $L^2(\Omega)$ 中的 Cauchy 序列。利用 L^2 的完备性，定义

$$\int_0^T H_s\,\mathrm{d}W_s := L^2\text{-}\lim_{n\to\infty} I_T(H^{(n)}).$$

等距式还保证这个定义与逼近序列无关。这就是随机积分构造的核心：先定义容易计算的对象，再通过保持范数的映射完成延拓。

具有有限 $\mathcal{H}^2$ 范数的渐进可测过程也可以用于布朗 Itô 积分；它们可在 $\mathrm{d}s \otimes \mathbb{P}$ 意义下由可预测过程代表或由相应简单过程逼近。教材使用“渐进可测”或“可预测”两种表述并不矛盾，但都比单独的适应性更充分。

4.4 随机积分的性质与局部化

对 $H \in \mathcal{H}^2([0,T])$，$M_t = \int_0^t H_s \,\mathrm{d}W_s$ 有连续版本，且为平方可积鞅。特别地，

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s, \qquad \mathbb{E}[M_t] = 0, \qquad [M]_t = \int_0^t H_s^2 \,\mathrm{d}s.$$

连续版本可通过简单过程的连续积分与 Doob 不等式构造：

$$\mathbb{E}\sup_{t\le T} |I_t(H^{(n)}) - I_t(H^{(m)})|^2 \le 4\mathbb{E}\int_0^T |H_s^{(n)} - H_s^{(m)}|^2 \,\mathrm{d}s.$$

这一估计比单个时刻的等距式更强，因为它同时控制整个时间区间。二次变差公式则由简单过程情形和相应的收敛论证得到。

若只知道

$$\int_0^T H_s^2 \,\mathrm{d}s < \infty \qquad \text{几乎必然},$$

而该量的期望可能无穷，也可以局部化。例如定义 $\tau_n = \inf\{t : \int_0^t H_s^2 \,\mathrm{d}s \ge n\} \wedge n$，对 $H\mathbf{1}_{[0,\tau_n]}$ 构造积分，再把彼此一致的停止积分拼接起来。这样得到连续局部鞅，但不自动得到全局均值为零的结论。

说明 4.3 (随机系数并不意味着独立)**.** 当 H 依赖 W 时，不能先“固定整条 H 路径”，就断言积分条件于 H 是 $\mathcal{N}(0, \int H^2)$。这种条件高斯结论需要额外的独立性结构。等距式始终是正确的二阶矩工具，但它并不决定完整分布。

4.5 核心例子：为什么积分 $W\,\mathrm{d}W$ 多出一个时间项

沿分割使用代数恒等式

$$W_{t_{k+1}}^2 - W_{t_k}^2 = 2W_{t_k}\Delta_k W + (\Delta_k W)^2.$$

求和得到

$$\sum_k W_{t_k}\Delta_k W = \frac{1}{2}\left(W_T^2 - \sum_k (\Delta_k W)^2\right).$$

左侧在 L^2 中趋于 Itô 积分，右侧由定理 3.5 趋于 $\frac{1}{2}(W_T^2 - T)$，故

$$\boxed{\int_0^T W_s \,\mathrm{d}W_s = \frac{1}{2}(W_T^2 - T).} \tag{4.1}$$

左端点逼近确实可用等距式验证，因为

$$\sum_k \int_{t_k}^{t_{k+1}} \mathbb{E}|W_s - W_{t_k}|^2 \,\mathrm{d}s = \frac{1}{2}\sum_k (t_{k+1} - t_k)^2 \longrightarrow 0.$$

取期望后，式 (4.1) 的右侧均值为零，与 Itô 积分性质一致。若直接套普通链式法则得到 $W_T^2/2$，均值就会错误地变成 $T/2$。

右端点和则满足

$$\sum_k W_{t_{k+1}}\Delta_k W = \sum_k W_{t_k}\Delta_k W + \sum_k (\Delta_k W)^2 \longrightarrow \frac{1}{2}(W_T^2 + T).$$

两个取样约定的极限相差 T，并没有随网格消失。这个例子既解释了随机积分必须选择约定，也预示了 Itô 与 Stratonovich 积分之间的修正项。

5 随机微分方程：从微分记号到随机轨道

5.1 真正的定义是积分方程

定义 5.1 (一维 Itô SDE)**.** 方程

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t, \qquad X_0 \text{ 给定}, \tag{5.1}$$

表示一个连续适应过程 X 满足

$$X_t = X_0 + \int_0^t b(X_s,s)\,\mathrm{d}s + \int_0^t \sigma(X_s,s)\,\mathrm{d}W_s.$$

在每个有限时间区间内，系数沿路径必须满足普通积分及随机积分所需的可测性与可积性。等式按连续版本理解为几乎必然对所有 t 同时成立。

漂移 b 描述有方向的平均运动，噪声系数 σ 描述随机波动的幅度。在足够正则的情形下，给定 $X_t = x$，短时间内有局部近似

$$X_{t+h} - X_t \approx b(x,t)h + \sigma(x,t)\sqrt{h}\,Z, \qquad Z \sim \mathcal{N}(0,1).$$

因此局部条件均值约为 $b(x,t)h$，局部条件方差约为 $\sigma^2(x,t)h$。这只是冻结系数后的短时近似，不是一般非线性 SDE 的精确转移分布。

说明 5.2 (三个经常混用的系数)**.** σ 是噪声振幅；σ^2 是单位时间的局部方差率；Fokker–Planck 方程中的二阶项为 $\frac{1}{2}\partial_{xx}(\sigma^2 p)$。物理中若把扩散常数记为 D 并写 $\partial_t p = D\partial_{xx}p$，对应的 SDE 是 $\mathrm{d}X_t = \sqrt{2D}\,\mathrm{d}W_t$。不同文献把“扩散系数”用于 σ、σ^2 或 D，必须先看定义。

5.2 强解、弱解与两种唯一性

强解要求在预先给定的概率空间、滤过、布朗运动及初值上构造 X，并且 X 可以由初值与该布朗运动生成的信息确定。弱解允许把概率空间和驱动布朗运动也作为待构造对象的一部分。这里的“强、弱”描述的是对随机输入的要求，不是轨道光滑程度，也不是 PDE 的弱解概念。

路径唯一性是指：在同一个概率空间上，由同一个 W 与同一个初值驱动的两个解，几乎必然对所有时刻都相同。依分布唯一性则只要求解过程的分布相同。二者不是定义上的同义词；在下面的全局 Lipschitz 条件下，强解存在并具有路径唯一性，也有相应的依分布唯一性。

5.3 存在唯一性：条件与证明框架

定理 5.3 (全局 Lipschitz 条件下的强解)**.** 设 $b,\sigma\colon \mathbb{R}\times[0,\infty)\to\mathbb{R}$ 是 Borel 可测函数。对任意 $T<\infty$，存在常数 K_T，使所有 $t\in[0,T]$、$x,y\in\mathbb{R}$ 满足

$$|b(x,t)-b(y,t)|^2 + |\sigma(x,t)-\sigma(y,t)|^2 \le K_T^2|x-y|^2,$$
$$|b(x,t)|^2 + |\sigma(x,t)|^2 \le K_T^2(1+|x|^2).$$

设 X_0 为 $\mathcal{F}_0$-可测，$\mathbb{E}|X_0|^2<\infty$，且 W 是相对于该滤过的布朗运动。则式 (5.1) 有唯一强解，且

$$\mathbb{E}\sup_{0\le t\le T}|X_t|^2 < \infty \qquad (T<\infty).$$

证明框架. 先从 $X_t^{(0)} = X_0$ 开始作 Picard 迭代：

$$X_t^{(n+1)} = X_0 + \int_0^t b(X_s^{(n)}, s)\,\mathrm{d}s + \int_0^t \sigma(X_s^{(n)}, s)\,\mathrm{d}W_s.$$

线性增长条件保证每一步在有限时间内平方可积。由 Cauchy–Schwarz 不等式、Doob 的 L^2 不等式和 Itô 等距，存在仅依赖 T, K_T 的常数 C_T，使

$$D_n(t) := \mathbb{E}\sup_{u\le t}|X_u^{(n+1)} - X_u^{(n)}|^2 \le C_T\int_0^t D_{n-1}(s)\,\mathrm{d}s.$$

递推积分给出阶乘衰减的界 $D_n(T) \le C(C_T T)^n/n!$。因此 $\sum_n \sqrt{D_n(T)} < \infty$，迭代在连续适应过程的均方上确界范数中收敛。利用 Lipschitz 条件和等距式，将极限传入两个积分，得到解。

若 X, Y 是同输入、同初值的两个解，同样的估计得到

$$D(t) := \mathbb{E}\sup_{u\le t}|X_u - Y_u|^2 \le C_T\int_0^t D(s)\,\mathrm{d}s.$$

Gronwall 不等式给出 $D(t) = 0$，故路径唯一。对解本身使用线性增长估计，再应用 Gronwall 不等式，便得到所述矩界并排除有限时爆炸。本段保留了证明的主要估计；完备空间上的极限构造与版本选择细节从略。 □

全局条件是容易检查的充分条件，并不是存在唯一解的必要条件。若系数仅在空间上局部 Lipschitz，通常先在 $\tau_n = \inf\{t : |X_t| \ge n\}$ 前构造唯一解，再考察爆炸时间 $\tau_\infty = \lim_n \tau_n$。没有增长控制时，$\tau_\infty$ 可能有限。例如无噪声方程 $\mathrm{d}X_t = X_t^2\,\mathrm{d}t$、$X_0 = x_0 > 0$ 的解 $X_t = x_0/(1 - x_0 t)$ 在 $t = 1/x_0$ 爆炸。因此“系数很光滑”不能替代“解在所有时间存在”的论证。

5.4 例一：带漂移的布朗运动

考虑常系数方程

$$\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = x_0.$$

直接积分得到

$$X_t = x_0 + \mu t + \sigma W_t, \quad \mathbb{E}[X_t] = x_0 + \mu t, \quad \mathrm{Var}(X_t) = \sigma^2 t.$$

当 $\sigma \neq 0$ 时，转移密度为

$$q_h(x, y) = \frac{1}{\sqrt{2\pi\sigma^2 h}}\exp\left[-\frac{(y - x - \mu h)^2}{2\sigma^2 h}\right], \qquad h > 0.$$

当 $\sigma = 0$ 时分布退化为点质量，不能继续把它写成普通 Lebesgue 密度。若初值本身随机，且与未来布朗增量独立，则上述条件分布仍成立；无条件分布要对初值分布作混合，不一定仍然高斯。

5.5 例二：Ornstein–Uhlenbeck 过程的完整计算

例 5.4 (均值回复模型)**.** 考虑

$$\mathrm{d}X_t = -\theta(X_t - m)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad \theta > 0, \quad X_0 = x_0. \tag{5.2}$$

漂移把状态拉向 m，噪声持续把它推开；θ^{-1} 是回复的时间尺度。

令 $Y_t = e^{\theta t}(X_t - m)$。由于 $e^{\theta t}$ 是确定性的有限变差函数，它与连续半鞅的二次协变差为零，故普通积分因子法适用：

$$\mathrm{d}Y_t = \theta e^{\theta t}(X_t - m)\,\mathrm{d}t + e^{\theta t}\,\mathrm{d}X_t = \sigma e^{\theta t}\,\mathrm{d}W_t.$$

积分并乘回 $e^{-\theta t}$，得到

$$\boxed{X_t = m + (x_0 - m)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s.} \tag{5.3}$$

积分因子法在这里成立有具体原因，并不表示可以对任意随机乘积忽略二次协变差。

最后一项是确定性函数的布朗积分，因此为中心高斯变量。利用等距式，

$$\mathbb{E}[X_t] = m + (x_0 - m)e^{-\theta t},$$

$$v_t := \mathrm{Var}(X_t) = \sigma^2 \int_0^t e^{-2\theta(t-s)}\,\mathrm{d}s = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}).$$

若 $\sigma \neq 0$，则对 $t > 0$，

$$X_t \sim \mathcal{N}\left(m + (x_0 - m)e^{-\theta t}, \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})\right).$$

给定时刻 s 的 $X_s = x$，将式 (5.3) 的起点移到 s，可得精确转移关系

$$X_{s+h} = m + (x - m)e^{-\theta h} + \sigma \int_s^{s+h} e^{-\theta(s+h-u)}\,\mathrm{d}W_u.$$

积分项与 $\mathcal{F}_s$ 独立。当 $\sigma \neq 0$ 时，转移密度为

$$q_h(x, y) = \frac{1}{\sqrt{2\pi v_h}} \exp\left[-\frac{\{y - m - (x - m)e^{-\theta h}\}^2}{2v_h}\right].$$

5.6 OU 的长期分布不等于每条路径收敛到均值

当 $t \to \infty$，均值趋于 m，方差趋于 $\sigma^2/(2\theta)$，所以

$$X_t \implies \mathcal{N}\left(m, \frac{\sigma^2}{2\theta}\right).$$

这是分布收敛。只要 $\sigma \neq 0$，长期随机波动并不消失，因此不能据此声称 $X_t \to m$ 依概率或几乎必然成立。

若从一开始就取 $X_0 \sim \mathcal{N}(m, \sigma^2/(2\theta))$，且 X_0 与未来布朗运动独立，则 X_t 的分布对每个 t 都保持不变。由于转移机制时间齐次，这样得到的过程是平稳过程。当 $s \le t$ 时，由条件期望

$$\mathbb{E}[X_t - m \mid \mathcal{F}_s] = e^{-\theta(t-s)}(X_s - m)$$

可得平稳协方差

$$\mathrm{Cov}(X_s, X_t) = \frac{\sigma^2}{2\theta} e^{-\theta|t-s|}. \tag{5.4}$$

如果初值固定为 x_0，过程尚未平稳，则应改为

$$\mathrm{Cov}(X_s, X_t) = \frac{\sigma^2}{2\theta}\left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right).$$

这两个公式的差别来自初始随机性：平稳版本在 $t = 0$ 已经具有平衡方差，确定初值版本则从零方差开始积累噪声。

说明 5.5 (本阶段应掌握的计算路线)**.** 看到一个 SDE，先写清积分含义与随机输入，再检查解的条件。对线性方程优先尝试积分因子；对确定性核的布朗积分，用零均值、高斯性和 Itô 等距计算分布。若要对非线性函数 $f(X_t)$ 作变量替换，下一步必须使用 Itô 公式。

6 Itô 公式：随机微积分的链式法则

本节解决一个基本问题：已经知道随机状态 X_t 怎样变化，如何求观测量 $f(X_t,t)$ 的变化？答案中的二阶导数项，是连接随机轨道与扩散型偏微分方程的关键。

6.1 为什么普通链式法则会少一项

先看 $f(x)=x^2$。对于任意一段增量，都有精确恒等式

$$(X_{t_{k+1}})^2-(X_{t_k})^2=2X_{t_k}\Delta_k X+(\Delta_k X)^2.$$

对光滑曲线求和并细化分割，最后一项的总和趋于零；对布朗运动，它却趋于 t。因此

$$W_t^2-W_0^2=2\int_0^t W_s\,\mathrm{d}W_s+t.$$

若套用普通链式法则，就会漏掉 t。问题并非泰勒公式失效，而是不能丢弃二阶增量的累积。

对扩散过程 $\mathrm{d}X_t=b_t\,\mathrm{d}t+\sigma_t\,\mathrm{d}W_t$，其连续有限变差部分不贡献二次变差，而随机积分部分满足

$$[X]_t=\int_0^t\sigma_s^2\,\mathrm{d}s. \tag{6.1}$$

这通常被压缩为计算规则

$$(\mathrm{d}W_t)^2=\mathrm{d}t,\qquad \mathrm{d}t\,\mathrm{d}W_t=0,\qquad (\mathrm{d}t)^2=0,\qquad (\mathrm{d}X_t)^2=\sigma_t^2\,\mathrm{d}t.$$

这些是二次变差的符号记法，不能理解为普通实数微分的代数等式。

6.2 公式、条件与积分形式

定理 6.1 (一维 Itô 公式)**.** 设

$$X_t=X_0+\int_0^t b_s\,\mathrm{d}s+\int_0^t\sigma_s\,\mathrm{d}W_s,$$

其中系数具有所需的渐进可测性，且对每个有限 T，几乎必然有

$$\int_0^T|b_s|\,\mathrm{d}s<\infty,\qquad \int_0^T\sigma_s^2\,\mathrm{d}s<\infty.$$

若 $f\in C^{2,1}(\mathbb{R}\times[0,T])$，这里的上标依次表示空间二次、时间一次连续可微，则

$$\mathrm{d}f(X_t,t)=\left(f_t+b_tf_x+\frac{1}{2}\sigma_t^2f_{xx}\right)(X_t,t)\,\mathrm{d}t+\sigma_tf_x(X_t,t)\,\mathrm{d}W_t. \tag{6.2}$$

公式的实际含义是

$$f(X_t,t)-f(X_0,0)=\int_0^t\left[f_s(X_s,s)+b_sf_x(X_s,s)+\frac{1}{2}\sigma_s^2f_{xx}(X_s,s)\right]\mathrm{d}s + \int_0^t\sigma_sf_x(X_s,s)\,\mathrm{d}W_s. \tag{6.3}$$

必要时先停在有界区域内，再通过局部化解释各项。

对于马尔可夫型 SDE，只需在公式中代入 $b_t=b(X_t,t)$、$\sigma_t=\sigma(X_t,t)$。应用公式本身只需局部条件；若要对随机积分取期望并把它删去，则要再检查真鞅条件，例如

$$\mathbb{E}\int_0^T\sigma_s^2f_x(X_s,s)^2\,\mathrm{d}s<\infty. \tag{6.4}$$

能够写出 *Itô* 公式与能够不加说明地交换期望、积分及极限是不同的论证步骤。

6.3 从泰勒展开到公式：证明框架

下面说明各项的来源。完整证明还需建立随机积分和加权二次变差的收敛；这里把这些分析步骤单独指出，避免用形式微分替代证明。

取 $[0,t]$ 的确定性分割。利用时间与空间增量的分拆以及泰勒展开，可将总增量写为

$$f(X_t,t)-f(X_0,0)=\sum_k f_t(X_{t_k},t_k)\Delta_kt+\sum_k f_x(X_{t_k},t_k)\Delta_kX + \frac{1}{2}\sum_k f_{xx}(X_{t_k},t_k)(\Delta_kX)^2+R_\Pi.$$

先把过程停在一个紧区域，并控制 $\int|b_s|\,\mathrm{d}s$ 与 $\int\sigma_s^2\,\mathrm{d}s$。分割细化时：

(1) 第一项是通常的时间积分。

(2) 第二项中，漂移部分趋于 $\int f_xb\,\mathrm{d}s$，随机部分依随机积分的收敛理论趋于 $\int f_x\sigma\,\mathrm{d}W$。

(3) 第三项的极限是

$$\frac{1}{2}\int_0^t f_{xx}(X_s,s)\,\mathrm{d}[X]_s=\frac{1}{2}\int_0^t f_{xx}(X_s,s)\sigma_s^2\,\mathrm{d}s.$$

这一步使用的是加权二次变差收敛。

(4) 在紧区域上利用导数的一致连续性，泰勒余项可以用趋于零的模函数乘以 $\sum(\Delta X)^2$ 及时间增量和控制，从而 $R_\Pi\to0$（依概率）。最后解除停时。

直觉上，ΔW 的典型尺度是 $\sqrt{\Delta t}$，所以 $(\Delta W)^2$ 累积到时间尺度；但这个尺度直觉本身并不证明上述极限。

6.4 二次协变差与乘积法则

对连续半鞅 X,Y，二次协变差定义为增量乘积和的概率极限：

$$[X,Y]_t=\lim_{|\Pi|\to0}\sum_k\Delta_kX\,\Delta_kY.$$

它也可由极化恒等式 $[X,Y]=\frac{1}{4}([X+Y]-[X-Y])$ 得到。如果

$$\mathrm{d}X_t=b_t\,\mathrm{d}t+\sigma_t\,\mathrm{d}W_t,\qquad \mathrm{d}Y_t=c_t\,\mathrm{d}t+\eta_t\,\mathrm{d}W_t,$$

则

$$\mathrm{d}[X,Y]_t=\sigma_t\eta_t\,\mathrm{d}t.$$

同一个噪声会同时驱动两个过程，因此它们的增量乘积能够累积为非零极限。

命题 6.2 (随机乘积法则)**.** 对连续半鞅，

$$\mathrm{d}(X_tY_t)=X_t\,\mathrm{d}Y_t+Y_t\,\mathrm{d}X_t+\,\mathrm{d}[X,Y]_t. \tag{6.5}$$

特别地，若 A_t 连续且具有有限变差，则 $[A,X]=0$，从而

$$\mathrm{d}(A_tX_t)=A_t\,\mathrm{d}X_t+X_t\,\mathrm{d}A_t.$$

这解释了为什么求解 OU 方程时可以使用确定性积分因子 $e^{\theta t}$，而不会再出现修正项。另一方面，若 X、Y 都含随机噪声，就不能随意删去协变差。

例 6.3 (相关噪声)**.** 设两个布朗运动满足 $[W^1,W^2]_t=\rho t$，其中 $-1\le\rho\le1$，则

$$\mathrm{d}(W_t^1W_t^2)=W_t^1\,\mathrm{d}W_t^2+W_t^2\,\mathrm{d}W_t^1+\rho\,\mathrm{d}t.$$

当它们独立时 $\rho=0$；当二者就是同一个布朗运动时 $\rho=1$，便恢复平方公式。

6.5 多维 Itô 公式与扩散矩阵

令状态维度为 d，独立布朗噪声的维度为 m：

$$\mathrm{d}X_t=b(X_t,t)\,\mathrm{d}t+\Sigma(X_t,t)\,\mathrm{d}W_t,\qquad b\in\mathbb{R}^d,\quad \Sigma\in\mathbb{R}^{d\times m}.$$

定义

$$a(x,t)=\Sigma(x,t)\Sigma(x,t)^{\mathsf{T}},\qquad a_{ij}=\sum_{k=1}^{m}\Sigma_{ik}\Sigma_{jk}. \tag{6.6}$$

于是 a 为半正定矩阵，且

$$\mathrm{d}[X^i,X^j]_t=a_{ij}(X_t,t)\,\mathrm{d}t.$$

对空间二次、时间一次连续可微的标量函数 f，

$$\begin{aligned}\mathrm{d}f(X_t,t)=&\left[f_t+b\cdot\nabla f+\frac{1}{2}\sum_{i,j=1}^{d}a_{ij}\partial_{ij}f\right](X_t,t)\,\mathrm{d}t\\&+\nabla f(X_t,t)^{\mathsf{T}}\Sigma(X_t,t)\,\mathrm{d}W_t.\end{aligned} \tag{6.7}$$

二阶项也可写为 $\frac{1}{2}\operatorname{tr}(a\nabla^2 f)$。对称求和中的非对角项包含两个相同贡献，前面的 $1/2$ 会抵消这一次重复计数。

例 6.4 (二维相关扩散的交叉导数)**.** 如果 $a=\left(\begin{smallmatrix}1&\rho\\ \rho&1\end{smallmatrix}\right)$，那么二阶项为

$$\frac{1}{2}f_{x_1x_1}+\rho f_{x_1x_2}+\frac{1}{2}f_{x_2x_2}.$$

把扩散矩阵误写成仅含对角项，会丢失两个分量的瞬时相关性。

6.6 典型例子：多项式、指数与几何布朗运动

例 6.5 (四阶矩)**.** 取 $f(x) = x^4$，则

$$\mathrm{d}(W_t^4) = 4W_t^3\,\mathrm{d}W_t + 6W_t^2\,\mathrm{d}t.$$

在有限时间区间上相应高斯矩有限，随机积分是真鞅，故

$$\mathbb{E}[W_t^4] = 6\int_0^t \mathbb{E}[W_s^2]\,\mathrm{d}s = 6\int_0^t s\,\mathrm{d}s = 3t^2.$$

Itô 公式因此也可用来递推高阶矩。

命题 6.6 (常系数指数鞅)**.** 对任意实常数 λ，

$$M_t = \exp\!\left(\lambda W_t - \frac{1}{2}\lambda^2 t\right)$$

是真鞅，且 $\mathbb{E}[M_t] = 1$。

证明. Itô 公式给出 $\mathrm{d}M_t = \lambda M_t\,\mathrm{d}W_t$，这首先说明它是局部鞅。要证明真鞅，令 $0 \le s \le t$，利用独立高斯增量：

$$\begin{aligned}\mathbb{E}[M_t \mid \mathcal{F}_s] &= M_s\mathbb{E}\Big[e^{\lambda(W_t - W_s) - \lambda^2(t-s)/2} \mid \mathcal{F}_s\Big] \\ &= M_s e^{\lambda^2(t-s)/2}e^{-\lambda^2(t-s)/2} = M_s.\end{aligned}$$

高斯矩母函数同时保证这里的可积性。也可以验证

$$\mathbb{E}\int_0^T \lambda^2 M_s^2\,\mathrm{d}s = \lambda^2\int_0^T e^{\lambda^2 s}\,\mathrm{d}s < \infty$$

并使用平方可积随机积分的鞅性质。 □

说明 6.7. “微分中没有 $\mathrm{d}t$ 项”一般只推出局部鞅，不能直接推出期望守恒。随机被积函数的指数需要额外的可积性条件；常数 λ 的上述计算不能不加检查地推广到所有随机系数。

考虑几何布朗运动（GBM）

$$\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \qquad X_0 = x_0 > 0. \tag{6.8}$$

稳妥的推导顺序是先构造候选解

$$X_t = x_0\exp\!\left[(\mu - \sigma^2/2)t + \sigma W_t\right], \tag{6.9}$$

再对指数函数应用 Itô 公式验证其满足 (6.8)。系数全局 Lipschitz，因此唯一性说明这就是解。构造式也直接证明 $X_t > 0$，此后才可安全地对解使用 log 变换：

$$\mathrm{d}\log X_t = \frac{\mathrm{d}X_t}{X_t} - \frac{\mathrm{d}[X]_t}{2X_t^2} = \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

如果从一开始就写 $\log X_t$，则应说明先停在 $X_t > 0$ 的区域内，并补充不会碰到零的论证。

由显式解，对任意实数 q 和有限 t，

$$\mathbb{E}[X_t^q] = x_0^q \exp\left[q\mu t + \frac{1}{2}q(q-1)\sigma^2 t\right]. \tag{6.10}$$

特别地，

$$\mathbb{E}[X_t] = x_0 e^{\mu t}, \qquad \operatorname{Var}(X_t) = x_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1).$$

若 $\sigma \neq 0$，则 $\log X_t$ 是均值 $\log x_0 + (\mu - \sigma^2/2)t$、方差 $\sigma^2 t$ 的高斯变量，X_t 具有对数正态分布。

说明 6.8. 平均值的增长率为 μ，而对数的漂移为 $\mu - \sigma^2/2$。两者是不同统计量，不能把 $\log \mathbb{E}[X_t]$ 与 $\mathbb{E}[\log X_t]$ 混为一谈。

7 Stratonovich 积分与两种 SDE 的转换

Itô 积分从左端点读取信息，特别适合条件期望与鞅计算。另一个常用定义采用对称取值，它把一部分二次变差吸收进积分定义，并恢复经典形式的链式法则。两种定义都一致自洽；描述同一个过程时，必须同时转换漂移。

7.1 对称和及其与 Itô 积分的关系

对连续半鞅 Y, W，在适当局部化下，Stratonovich 积分可定义为

$$\int_0^t Y_s \circ \mathrm{d}W_s := \int_0^t Y_s \,\mathrm{d}W_s + \frac{1}{2}[Y, W]_t. \tag{7.1}$$

它也等于对称和

$$\sum_k \frac{Y_{t_k} + Y_{t_{k+1}}}{2}(W_{t_{k+1}} - W_{t_k})$$

在确定性分割细化时的概率极限。其来源是精确分拆

$$\sum_k \frac{Y_{t_k} + Y_{t_{k+1}}}{2}\Delta_k W = \sum_k Y_{t_k}\Delta_k W + \frac{1}{2}\sum_k \Delta_k Y\, \Delta_k W.$$

右侧分别趋于 Itô 积分和半个二次协变差。

这里“对称取值”指两个端点函数值的平均。把它称为“中点规则”有助于记忆，但不同类型的中点近似需要各自的收敛条件，不能在没有说明时把所有中点方案视为定义上完全相同。

例 7.1 (同一个被积函数，两种结果)**.** 取 $Y = W$，因为 $[W]_t = t$，

$$\int_0^t W_s \circ \mathrm{d}W_s = \frac{1}{2}(W_t^2 - t) + \frac{1}{2}t = \frac{1}{2}W_t^2.$$

因此 Stratonovich 积分的期望为 $t/2$，而 Itô 积分的期望为零。*Stratonovich* 随机积分一般不是鞅。

7.2 一维转换：逐项计算修正漂移

先从已知的 Itô 方程开始：

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t.$$

为了直接使用 Itô 公式，暂假定 $\sigma \in C^{2,1}$；事实上转换关系可在较弱条件下成立。令 $Y_t = \sigma(X_t,t)$，则

$$\mathrm{d}Y_t = \left(\partial_t\sigma + b\sigma_x + \tfrac{1}{2}\sigma^2\sigma_{xx}\right)(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\sigma_x(X_t,t)\,\mathrm{d}W_t.$$

只有噪声部分贡献与 W 的二次协变差，所以

$$[\sigma(X_\cdot,\cdot),W]_t = \int_0^t \sigma(X_s,s)\sigma_x(X_s,s)\,\mathrm{d}s. \tag{7.2}$$

代入 (7.1)：

$$\int_0^t \sigma(X_s,s)\circ \mathrm{d}W_s = \int_0^t \sigma(X_s,s)\,\mathrm{d}W_s + \frac{1}{2}\int_0^t \sigma\sigma_x(X_s,s)\,\mathrm{d}s.$$

于是得到下面的对应关系。

定理 7.2 (一维 Itô–Stratonovich 转换)**.** 在所需正则性和局部可积性条件下，

$$\begin{aligned}\mathrm{d}X_t &= b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t,\\ \mathrm{d}X_t &= b_{\mathrm{S}}(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\circ \mathrm{d}W_t\end{aligned} \tag{7.3}$$

在其漂移按下式对应时描述同一个过程：

$$\boxed{\,b = b_{\mathrm{S}} + \frac{1}{2}\sigma\partial_x\sigma, \qquad b_{\mathrm{S}} = b - \frac{1}{2}\sigma\partial_x\sigma.\,} \tag{7.4}$$

若只知局部存在性，则等价关系在爆炸时间之前理解。

特别注意：修正项中没有 $\partial_t\sigma$。时间方向具有有限变差，不会与布朗运动产生二次协变差。修正项来自*噪声系数随随机状态变化*，而不是来自其确定性时间变化。

例 7.3 (加性噪声与乘性噪声)**.** 若 $\sigma = \sigma(t)$ 与状态无关，则 $\sigma_x = 0$，两种写法的漂移相同。若 $\sigma(x) = \gamma x$，则修正项为 $\gamma^2 x/2$。因此

$$\mathrm{d}X_t = \lambda X_t\,\mathrm{d}t + \gamma X_t \circ \mathrm{d}W_t$$

等价于

$$\mathrm{d}X_t = (\lambda + \gamma^2/2)X_t\,\mathrm{d}t + \gamma X_t\,\mathrm{d}W_t.$$

前式的解为 $X_t = x_0 e^{\lambda t + \gamma W_t}$。把同一个 λ 原封不动地放进 Itô 方程，会得到另一个过程。

例 7.4 (原稿中的非线性噪声)**.** 对

$$\mathrm{d}X_t = (1 - X_t)\,\mathrm{d}t + X_t^2 \circ \mathrm{d}W_t,$$

有 $\sigma(x) = x^2$、$\sigma_x(x) = 2x$，故

$$\frac{1}{2}\sigma\sigma_x = x^3, \qquad \mathrm{d}X_t = (1 - X_t + X_t^3)\,\mathrm{d}t + X_t^2\,\mathrm{d}W_t.$$

这是一项*局部代数转换*。多项式系数在局部 Lipschitz，但不满足前面使用的全局线性增长条件；仅凭转换公式不能宣称解对所有时间都不爆炸。

7.3 为什么 Stratonovich 形式恢复普通链式法则

为使下面每个复合被积过程都可直接按半鞅处理，这段推导先取 f 与 σ 足够光滑；较弱正则性下的链式法则可通过逼近建立。

从 $b = b_{\mathrm{S}} + \sigma\sigma_x/2$ 出发，Itô 公式给出

$$\mathrm{d}f(X_t,t) = \big(f_t + b_{\mathrm{S}} f_x + \tfrac{1}{2}\sigma\sigma_x f_x + \tfrac{1}{2}\sigma^2 f_{xx}\big)\,\mathrm{d}t + \sigma f_x\,\mathrm{d}W_t.$$

另一方面，将被积函数 σf_x 的积分转成 Stratonovich 形式，其修正项为

$$\frac{1}{2}\sigma\partial_x(\sigma f_x) = \frac{1}{2}\sigma\sigma_x f_x + \frac{1}{2}\sigma^2 f_{xx}.$$

两项恰好被吸收，因此

$$\boxed{\mathrm{d}f(X_t,t) = (f_t + b_{\mathrm{S}} f_x)(X_t,t)\,\mathrm{d}t + \sigma(X_t,t) f_x(X_t,t) \circ \mathrm{d}W_t.} \tag{7.5}$$

也就是 $\mathrm{d}f = f_t\,\mathrm{d}t + f_x \circ \mathrm{d}X_t$。恢复普通形式并不意味着二次变差消失了，而是它已包含在圆圈所表示的积分定义中。

7.4 多维转换：按噪声方向求导

设 W 有 m 个独立分量，Stratonovich 方程为

$$\mathrm{d}X_t^i = b_{\mathrm{S},i}(X_t,t)\,\mathrm{d}t + \sum_{k=1}^{m} \Sigma_{ik}(X_t,t) \circ \mathrm{d}W_t^k.$$

对固定的 i,k，计算 $\Sigma_{ik}(X_t,t)$ 与 W^k 的协变差，得到等价 Itô 漂移

$$b_i = b_{\mathrm{S},i} + \frac{1}{2}\sum_{k=1}^{m}\sum_{j=1}^{d} \Sigma_{jk}\partial_j \Sigma_{ik}. \tag{7.6}$$

如果把 Σ 的第 k 列记为向量场 V_k，则修正向量为 $\frac{1}{2}\sum_k (DV_k)V_k$。方向导数沿每个噪声向量场进行，因此一般不能把一维公式机械地写成 $\frac{1}{2}\nabla a$。

说明 7.5. Itô 方程的生成元只依赖 b 与 $a = \Sigma\Sigma^{\mathsf{T}}$；Stratonovich 漂移的转换还涉及具体的噪声向量场 Σ。因此在多维状态相关噪声下，更换 a 的平方根分解时，不能默认 Stratonovich 漂移保持不变。

8 无穷小生成元：从随机轨道到期望变化

研究一个复杂随机过程时，往往不需要先求出每条轨道。可以先问：任意光滑观测量 $\varphi(X_t)$ 的平均变化率是什么？回答这个问题的算子就是生成元。

8.1 短时条件期望与生成元

对于良定的马尔可夫扩散，记 $\mathbb{E}_{t,x}$ 表示在时刻 t 从 x 出发的期望。对合适的测试函数，定义

$$L_t\varphi(x)=\lim_{h\downarrow 0}\frac{\mathbb{E}_{t,x}[\varphi(X_{t+h})]-\varphi(x)}{h}. \tag{8.1}$$

在系数及测试函数具有足够正则性时，Itô 公式与局部化给出

$$L_t\varphi(x)=b(x,t)\varphi'(x)+\frac{1}{2}\sigma^2(x,t)\varphi''(x) \qquad (\text{一维}), \tag{8.2}$$

$$L_t\varphi(x)=b(x,t)\cdot\nabla\varphi(x)+\frac{1}{2}\sum_{i,j}a_{ij}(x,t)\partial_{ij}\varphi(x) \qquad (\text{多维}). \tag{8.3}$$

例如，在局部条件适当时，短时条件矩满足

$$\mathbb{E}_{t,x}[X_{t+h}-x]=b(x,t)h+o(h),$$

$$\mathbb{E}_{t,x}[(X_{t+h}-x)(X_{t+h}-x)^{\mathsf{T}}]=a(x,t)h+o(h).$$

因此漂移描述一阶平均位移，a 描述单位时间内的二阶随机扩散。

严格的算子理论还须指定生成元所在的函数空间及定义域。这里首先把 L_t 作为作用于光滑紧支撑函数的微分表达式使用；不能仅凭表达式便略去边界条件和算子定义域。

8.2 Dynkin 公式与鞅分解

将 Itô 公式中的漂移记成 $L_t\varphi$，可得

$$\varphi(X_t)-\varphi(X_0)-\int_0^t L_s\varphi(X_s)\,\mathrm{d}s=\int_0^t\sigma(X_s,s)\varphi'(X_s)\,\mathrm{d}W_s. \tag{8.4}$$

右边首先是局部鞅。若 $\mathbb{E}|\varphi(X_0)|<\infty$，$\mathbb{E}\int_0^t\sigma(X_s,s)^2\varphi'(X_s)^2\,\mathrm{d}s<\infty$，且 $\mathbb{E}\int_0^t|L_s\varphi(X_s)|\,\mathrm{d}s<\infty$，则右边为平方可积鞅，可以取期望得到 Dynkin 公式

$$\boxed{\mathbb{E}[\varphi(X_t)]-\mathbb{E}[\varphi(X_0)]=\int_0^t\mathbb{E}[L_s\varphi(X_s)]\,\mathrm{d}s.} \tag{8.5}$$

若被积函数关于时间连续，则可以进一步写成

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\varphi(X_t)]=\mathbb{E}[L_t\varphi(X_t)]. \tag{8.6}$$

若只有可积性，导数等式通常理解为几乎处处成立，而积分式仍是首要表述。对有界停时也可写出相应停止版本，但仍须保证停止后的随机积分具有所需可积性。

8.3 矩方程：何时可以闭合

令 $m(t)=\mathbb{E}[X_t]$。在相关矩有限并能合法取期望时，选择 $\varphi(x)=x$ 得

$$m'(t)=\mathbb{E}[b(X_t,t)]. \tag{8.7}$$

选择 $\varphi(x)=x^2$ 得

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t^2]=\mathbb{E}[2X_t b(X_t,t)+\sigma^2(X_t,t)]. \tag{8.8}$$

于是方差满足

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Var}(X_t)=2\operatorname{Cov}(X_t,b(X_t,t))+\mathbb{E}[\sigma^2(X_t,t)]. \tag{8.9}$$

注意 $\mathbb{E}[b(X_t,t)]$ 一般不等于 $b(\mathbb{E}[X_t],t)$。如果 $b(x)=-x^3$，均值方程涉及三阶矩，三阶矩又可能涉及更高阶矩，形成矩层级；这时低阶矩通常不能单独构成闭合方程。

例 8.1 (OU 过程的矩闭合)**.** 对 $b(x)=-\theta x$、常数噪声强度 σ，

$$m'=-\theta m,\qquad v'=-2\theta v+\sigma^2.$$

若 $X_0=x_0$ 为确定值，则

$$m(t)=x_0e^{-\theta t},\qquad v(t)=\frac{\sigma^2}{2\theta}(1-e^{-2\theta t}).$$

这与直接求解 SDE 的结果一致，但完全不需要先写出随机积分解。

9 Kolmogorov 后向方程：预测未来的观测量

“后向”指算子作用在起始状态上，或者终端条件向较早时间传播。它并不表示布朗运动的时间参数简单地变成负数。

9.1 时间齐次情形：以经过时间为变量

若 b,σ 与时间无关，定义转移半群

$$P_\tau\varphi(x)=\mathbb{E}_x[\varphi(X_\tau)],\qquad u(x,\tau)=P_\tau\varphi(x).$$

马尔可夫性质给出 $P_{\tau+h}=P_hP_\tau$。若 $u(\cdot,\tau)$ 位于生成元的适当定义域，则

$$\frac{u(x,\tau+h)-u(x,\tau)}{h}=\frac{P_h[u(\cdot,\tau)](x)-u(x,\tau)}{h}\longrightarrow Lu(x,\tau).$$

因而

$$\partial_\tau u=Lu,\qquad u(x,0)=\varphi(x). \tag{9.1}$$

这里 τ 是从起点算起的经过时间，所以出现正号。

9.2 终端值问题：以日历时间为变量

现在允许系数随时间变化，并固定终点 T。定义

$$v(x,t)=\mathbb{E}_{t,x}[\varphi(X_T)],\qquad 0\le t\le T.$$

动态规划关系是

$$v(x,t)=\mathbb{E}_{t,x}[v(X_{t+h},t+h)].$$

如果 v 足够光滑，且允许使用短时生成元展开，右边等于

$$v(x,t)+h\big(\partial_t v(x,t)+L_t v(x,t)\big)+o(h).$$

因此

$$\boxed{\partial_t v+L_t v=0,\qquad v(x,T)=\varphi(x).} \tag{9.2}$$

在时间齐次情形，$v(x,t)=u(x,T-t)$，两个符号约定便完全一致。

命题 9.1 (后向方程的验证原理)**.** 若 v 是 (9.2) 的经典解，且沿 X 的随机积分是真鞅，则

$$v(X_s,s),\qquad t\le s\le T,$$

是真鞅，并有 $v(x,t)=\mathbb{E}_{t,x}[\varphi(X_T)]$。

证明. 对 $v(X_s,s)$ 应用 Itô 公式，漂移为 $(\partial_s v+L_s v)(X_s,s)=0$。在可积性条件下从 t 到 T 取条件期望，利用终端条件即得。 □

该原理从已知 PDE 解推出随机表示；反过来，由期望定义的函数是否为经典解还需要正则性理论，不能仅靠一个形式展开便断言光滑性。

例 9.2 (布朗运动的终端平方观测)**.** 若 $\mathrm{d}X_s=\mathrm{d}W_s$，$\varphi(x)=x^2$，则

$$v(x,t)=x^2+(T-t).$$

直接计算 $v_t=-1$、$v_{xx}=2$，故 $v_t+\frac12 v_{xx}=0$，且 $v(x,T)=x^2$。预测时间越长，平方观测的期望就因方差增长而增加。

10 Fokker–Planck 方程：概率分布如何演化

后向方程传播观测量；前向方程传播状态分布。二者由期望的配对关系连接。需要先强调：一个 SDE 有解，并不自动意味着其分布在每个时刻都有光滑密度。例如无噪声的确定性初值系统，其分布可以始终是一个点质量。

10.1 先写对分布总是更自然的弱形式

记 $\nu_t=\mathrm{Law}(X_t)$。对合适的 $\varphi\in C_c^\infty(\mathbb{R})$，Dynkin 公式写成

$$\int\varphi\,\mathrm{d}\nu_t-\int\varphi\,\mathrm{d}\nu_0=\int_0^t\int L_s\varphi(x)\,\nu_s(\mathrm{d}x)\,\mathrm{d}s. \tag{10.1}$$

这个式子不需要假定 ν_t 有密度，它表达了分布意义下的前向演化。若 $\nu_t(\mathrm{d}x)=p(x,t)\,\mathrm{d}x$ 且足够正则，则可把它改写成经典 PDE。

10.2 逐步推导一维前向方程

设

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t.$$

从期望变化率出发，

$$\begin{aligned}\int_{\mathbb{R}} \varphi(x)\partial_t p(x,t)\,\mathrm{d}x &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\varphi(X_t)] \\ &= \int_{\mathbb{R}} \left(b\varphi' + \frac{1}{2}\sigma^2\varphi'' \right) p\,\mathrm{d}x.\end{aligned}$$

对第一项分部积分一次，对第二项分部积分两次：

$$\begin{aligned}\int_{\mathbb{R}} bp\,\varphi'\,\mathrm{d}x &= -\int_{\mathbb{R}} \varphi\,\partial_x(bp)\,\mathrm{d}x, \\ \int_{\mathbb{R}} \sigma^2 p\,\varphi''\,\mathrm{d}x &= -\int_{\mathbb{R}} \partial_x(\sigma^2 p)\varphi'\,\mathrm{d}x \\ &= \int_{\mathbb{R}} \varphi\,\partial_{xx}(\sigma^2 p)\,\mathrm{d}x.\end{aligned}$$

由于测试函数紧支撑，此处没有无穷远边界项。因此对所有这样的 φ，

$$\int_{\mathbb{R}} \varphi \left[\partial_t p + \partial_x(bp) - \frac{1}{2}\partial_{xx}(\sigma^2 p) \right] \mathrm{d}x = 0.$$

由分布的定义得到

$$\boxed{\partial_t p = L_t^* p = -\partial_x(bp) + \frac{1}{2}\partial_{xx}(\sigma^2 p).} \tag{10.2}$$

若各项连续，等式进一步逐点成立。算子 L_t^* 是 L_t 关于积分配对的形式伴随：

$$\int (L_t\varphi)p\,\mathrm{d}x = \int \varphi(L_t^* p)\,\mathrm{d}x. \tag{10.3}$$

说明 10.1 (两个最常见的符号错误)**.** 漂移项前的负号来自一次分部积分，扩散项前的正号来自两次分部积分。若 σ 依赖于 x，必须对整个 $\sigma^2 p$ 求两次导数；一般不能写成 $\frac{1}{2}\sigma^2 p_{xx}$。即使把算子符号写对，忽略边界条件也可能得到错误的质量演化。

10.3 前向与后向：同一个转移核的两侧

设 $q(s,x;t,y)$ 是从 (s,x) 到 (t,y) 的转移密度。在足够正则条件下，

$$\partial_t q(s,x;t,y) = L_{t,y}^* q(s,x;t,y), \tag{10.4}$$

$$\partial_s q(s,x;t,y) + L_{s,x} q(s,x;t,y) = 0. \tag{10.5}$$

下标 y 或 x 指明微分所作用的空间变量。前向方程改变到达时间与到达状态的分布；后向方程改变出发时间与出发位置。

对象	方程	给定的数据
密度 $p(x,t)$	$\partial_t p = L_t^* p$	初始分布 p_0
终端观测期望 $v(x,t)$	$\partial_t v + L_t v = 0$	$v(x,T) = \varphi(x)$
齐次半群 $u(x,\tau)$	$\partial_\tau u = Lu$	$u(x,0) = \varphi(x)$

这个表中的符号差异由变量含义决定，而非相互矛盾。

10.4 概率流、质量与边界条件

定义概率流

$$J(x,t) = b(x,t)p(x,t) - \frac{1}{2}\partial_x(\sigma^2(x,t)p(x,t)). \tag{10.6}$$

则前向方程是连续性方程

$$\partial_t p + \partial_x J = 0. \tag{10.7}$$

对区间 (ℓ, r) 积分得到

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\ell^r p(x,t)\,\mathrm{d}x = J(\ell,t) - J(r,t). \tag{10.8}$$

左边界向右的流入增加区间内质量，右边界向右的流出减少区间内质量。

常见情形如下。

- **全空间**。若无穷远处的流满足足够衰减，边界项为零，总概率守恒。真正需要检查的是流及相关积分，不只是口头上说“密度趋于零”。
- **反射边界**。要求 $J(\ell,t) = J(r,t) = 0$，概率留在区域内。对应的轨道模型通常还要加入反射所产生的边界局部时间；仅写内点 SDE 不足以完整指定反射过程。
- **吸收或杀死边界**。对非退化扩散在规则端点，杀死过程常对应 $p(\ell,t) = p(r,t) = 0$。区域内密度的质量是存活概率，会因外流而减少。若把死亡后的状态放在“墓地状态”，则包括墓地在内的总概率仍为一。
- **周期边界**。首尾相接，并匹配密度、系数和流。此时 $J(\ell,t) = J(r,t)$ 保证质量守恒，但流本身未必为零。

说明 10.2. “吸收到端点并停在端点”和“离开区域后被杀死”必须区分。前者在边界可能积累点质量，单独的内点密度无法表示完整分布。后者的内点密度是总质量小于等于一的次概率密度。

10.5 多维前向方程

对 $\mathrm{d}X_t = b\,\mathrm{d}t + \Sigma\,\mathrm{d}W_t$，由 (8.3) 的伴随得到

$$\partial_t p = -\sum_{i=1}^{d}\partial_i(b_i p) + \frac{1}{2}\sum_{i,j=1}^{d}\partial_i\partial_j(a_{ij}p), \qquad a = \Sigma\Sigma^{\mathsf{T}}. \tag{10.9}$$

定义流的各分量

$$J_i = b_i p - \frac{1}{2}\sum_{j=1}^{d}\partial_j(a_{ij}p),$$

则 $\partial_t p + \nabla\cdot J = 0$。对于区域 D 及其外法向量 n，

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_D p\,\mathrm{d}x = -\int_{\partial D} J\cdot n\,\mathrm{d}S.$$

本讲义采用的无通量反射边界为 $J\cdot n = 0$，不必要求所有切向流分量都为零。在各向异性扩散中，对应轨道的反射方向应与扩散矩阵相容，例如采用共法向反射；一般斜反射需要另行确定其伴随边界条件。

10.6 例一：布朗运动、热核与扩散系数

对 $\mathrm{d}X_t = \sqrt{2D}\,\mathrm{d}W_t$，$D > 0$，方程为

$$\partial_t p = D\partial_{xx} p. \tag{10.10}$$

若初始状态为 x_0，则对 $t > 0$，

$$p(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\left[-\frac{(x-x_0)^2}{4Dt}\right]. \tag{10.11}$$

直接求导：

$$\partial_t p = p\left[-\frac{1}{2t} + \frac{(x-x_0)^2}{4Dt^2}\right],$$
$$\partial_{xx} p = p\left[-\frac{1}{2Dt} + \frac{(x-x_0)^2}{4D^2t^2}\right],$$

可见 $\partial_t p = D\partial_{xx} p$。并且 $\int p\,\mathrm{d}x = 1$，其方差为 $2Dt$。

初值应写为 $p(\cdot,0) = \delta_{x_0}$，这里的 δ 是点质量，不是普通函数。准确含义是对有界连续 φ，

$$\lim_{t\downarrow 0} \int \varphi(x) p(x,t)\,\mathrm{d}x = \varphi(x_0).$$

若初始分布本来有密度 p_0，则解是热核与 p_0 的卷积。

说明 10.3. 标准布朗运动的噪声幅度是 1，对应热方程系数 $D = 1/2$。物理中写的“扩散系数”常指 D，SDE 中的噪声幅度却是 $\sqrt{2D}$；这正是许多 2 或 $1/2$ 错误的来源。

10.7 例二：带漂移的高斯传播

对常系数方程 $\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$，$\sigma \neq 0$，

$$\partial_t p = -\mu\partial_x p + \frac{1}{2}\sigma^2 \partial_{xx} p,$$

且确定性初值的密度为

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{(x-x_0-\mu t)^2}{2\sigma^2 t}\right].$$

漂移将密度整体平移，噪声使其宽度随 $\sqrt{t}$ 增长。令 $y = x - \mu t$ 即可将 PDE 化为热方程。

10.8 零流平稳分布：一般一维公式

若系数与时间无关，平稳密度满足 $L^* p_\infty = 0$，等价于

$$\partial_x J_\infty = 0.$$

因此稳态概率流是常数。平稳不自动意味着零流；在全空间的适当无流边界条件或反射区间上才有 $J_\infty = 0$。在 $\sigma^2(x) > 0$ 的区间内，零流方程为

$$b(x)p_\infty(x) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}(\sigma^2(x)p_\infty(x)) = 0.$$

令 $q(x)=\sigma^2(x)p_\infty(x)$，则

$$\frac{q'(x)}{q(x)}=\frac{2b(x)}{\sigma^2(x)}.$$

积分得到

$$\boxed{p_\infty(x)=\frac{1}{Z}\frac{1}{\sigma^2(x)}\exp\!\left(\int_{x_*}^{x}\frac{2b(y)}{\sigma^2(y)}\,\mathrm{d}y\right).} \tag{10.12}$$

其中 Z 是归一化常数，且必须有 $0<Z<\infty$。参考点 x_* 的改变仅改变被 Z 吸收的常数。

这个推导有三个必须检查的条件：噪声在所考虑区间不退化，候选密度可积，以及候选密度满足实际过程的边界条件。找到形式上的稳态并不自动证明过程从任意初态都收敛到它，后者需要另外的遍历性论证。

例 10.4 (OU 稳态)**.** 对 $b(x)=-\theta x$，$\theta>0$，常数 $\sigma>0$，

$$p_\infty(x)\propto e^{-\theta x^2/\sigma^2},\qquad p_\infty=\mathcal{N}\!\left(0,\frac{\sigma^2}{2\theta}\right).$$

代回 $J_\infty=-\theta x p_\infty-(\sigma^2/2)p_\infty'$，两项抵消。这与 OU 转移分布的长期极限一致。

例 10.5 (圆周上的非零稳态流)**.** 在周长为 ℓ 的圆周上，常漂移 $c\neq 0$ 加常扩散的稳态密度为 $p_\infty=1/\ell$，但概率流是 $J_\infty=c/\ell\neq 0$。密度虽然不随时间改变，粒子仍在持续绕圈运动。

10.9 例三：梯度漂移与 Gibbs 型密度

考虑过阻尼 Langevin 方程

$$\mathrm{d}X_t=-V'(X_t)\,\mathrm{d}t+\sqrt{2D}\,\mathrm{d}W_t,\qquad D>0.$$

其前向方程为

$$\partial_t p=\partial_x(V'p)+D\partial_{xx}p.$$

根据 (10.12)，零流候选为

$$p_\infty(x)=Z^{-1}e^{-V(x)/D},\qquad Z=\int_{\mathbb{R}}e^{-V(x)/D}\,\mathrm{d}x. \tag{10.13}$$

当 $Z<\infty$ 且过程良定、边界行为相容时，这是不变密度。V 的低值区域权重较大，D 控制分布在不同区域之间的扩散尺度。取 $V(x)=\theta x^2/2$，便回到方差 D/θ 的 OU 稳态。

在 d 维，$\mathrm{d}X_t=-\nabla V(X_t)\,\mathrm{d}t+\sqrt{2D}\,\mathrm{d}W_t$ 仍有同样的 Gibbs 候选，因为

$$J=-p\nabla V-D\nabla p=0$$

在 $p\propto e^{-V/D}$ 时逐点成立。对一般非梯度漂移，则不能直接套用这个势函数表达式。

10.10 直接从 Stratonovich 方程写前向方程

若给定的是

$$\mathrm{d}X_t=b_{\mathrm{S}}(X_t,t)\,\mathrm{d}t+\sigma(X_t,t)\circ\mathrm{d}W_t,$$

最稳妥的方法是先转换漂移，再取生成元的伴随：

$$\partial_t p = -\partial_x\left[\left(b_{\mathrm{S}} + \frac{1}{2}\sigma\sigma_x\right)p\right] + \frac{1}{2}\partial_{xx}(\sigma^2 p). \tag{10.14}$$

利用乘积法则，也可写成

$$\boxed{\partial_t p = -\partial_x(b_{\mathrm{S}} p) + \frac{1}{2}\partial_x[\sigma\partial_x(\sigma p)]\,.} \tag{10.15}$$

确实，

$$\partial_x(\sigma^2 p) - \sigma\sigma_x p = \sigma\sigma_x p + \sigma^2 p_x = \sigma\partial_x(\sigma p),$$

再对两边求一次导数便验证了二式等价。

这提供了一个实用的交叉检查：若同一个过程的 Itô 与 Stratonovich 写法导出了不同的概率密度方程，通常是漏掉了漂移转换，或把状态相关噪声的导数移出了乘积。

10.11 解题时的统一工作顺序

面对一个新的扩散模型，可以按以下次序组织计算：

(1) 先说明随机积分的约定、状态空间、初始条件和边界行为。

(2) 若需要，转成 Itô 形式，明确漂移 b 与协方差系数 a。

(3) 研究单个观测量时使用 Itô 公式或生成元；研究整个分布时使用前向方程。

(4) 对取期望、分部积分、归一化及稳态候选，分别检查可积性与边界项。

(5) 用显式可解例子、矩方程或概率流检查符号与常数因子。

这样做将轨道、观测量、概率分布三个层次连在一起，也为后面的分枝与选择机制准备了统一的线性算子语言。

11 分枝扩散：从粒子数到归一化分布

前面的 Fokker–Planck 方程描述粒子如何在状态空间中移动，通常保持总概率质量不变。现在增加两种事件：粒子可以繁殖，也可以死亡。这样得到的第一条方程描述期望粒子数，总质量不再固定；将其归一化以后，才得到一个积分为一的分布。这个归一化步骤产生了一个非线性项，也是本节的核心。

11.1 先区分三个不同的对象

用 $\mathcal{I}_t$ 表示时刻 t 仍然存活的粒子集合，$X_i(t)$ 表示粒子 i 的状态。定义随机计数测度和随机总粒子数

$$Z_t := \sum_{i\in\mathcal{I}_t}\delta_{X_i(t)}, \qquad N_t := Z_t(\mathbb{R}) = |\mathcal{I}_t|. \tag{11.1}$$

这里 δ_x 是位于 x 的 Dirac 测度。对集合 A，$Z_t(A)$ 就是在 A 内的粒子数；对测试函数 φ，记

$$\langle Z_t, \varphi\rangle := \int \varphi(x) Z_t(\,\mathrm{d}x) = \sum_{i\in\mathcal{I}_t}\varphi(X_i(t)).$$

对随机计数测度取期望，得到确定性的有限测度

$$\nu_t(A) := \mathbb{E}[Z_t(A)], \qquad M(t) := \nu_t(\mathbb{R}) = \mathbb{E}[N_t]. \tag{11.2}$$

当 ν_t 对 Lebesgue 测度绝对连续时，写作 $\nu_t(\mathrm{d}x) = f(x,t)\,\mathrm{d}x$。因此

$$f(x,t)\,\mathrm{d}x \quad \text{表示位置在}(x, x+\mathrm{d}x)\text{ 内的期望粒子数}, \qquad \int f(x,t)\,\mathrm{d}x = M(t).$$

注意 f 本身不是概率密度。只要 $0 < M(t) < \infty$，就可以定义

$$p(x,t) := \frac{f(x,t)}{M(t)}, \qquad \int p(x,t)\,\mathrm{d}x = 1. \tag{11.3}$$

如果没有密度，上述定义应写成概率测度 $\pi_t := \nu_t/M(t)$；后面的弱形式仍然适用。

还可以在某一次实际种群中定义经验概率测度

$$\widehat{\pi}_t := \frac{Z_t}{N_t}, \qquad N_t > 0.$$

它仍然是随机的。Z_t、ν_t、$\widehat{\pi}_t$ 分别对应随机粒子数、平均粒子数和一次种群内的相对比例，不能相互替代。

11.2 模型假设与净增长率

为集中讨论基本机制，设每个粒子的运动由同一个非爆炸扩散过程描述：

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad L\varphi = b\varphi' + \frac{1}{2}\sigma^2\varphi''.$$

粒子位于 x 时，以速率 $\lambda(x) \geq 0$ 分裂，以速率 $\mu(x) \geq 0$ 死亡。一次分裂指原粒子被两个同位置的子粒子替代；分裂后子粒子独立演化。因此一次分裂使总数增加 $2-1=1$，一次死亡使总数减少一。净增长率是

$$r(x) := \lambda(x) - \mu(x). \tag{11.4}$$

如果分裂时的平均子代数为 $m(x)$，对应项应改成 $\lambda(x)(m(x)-1) - \mu(x)$；只知道“产生两个子粒子”而不明确母粒子是否保留，会导致系数错误。

本节先假设 λ, μ 有界、$\mathbb{E}N_0 < \infty$，运动保持质量，且子粒子在分裂位置出生。这是一组方便的充分条件：有限时间内不会出现无穷多粒子的爆炸，并且一阶矩有限。对无界速率可以建立更一般理论，但必须另行验证非爆炸性与所需的可积性。若繁殖率依赖整个种群密度或粒子相互作用，则下面的一阶矩方程通常不再自动闭合。

11.3 用测试函数推导一阶矩方程

考虑统计量 $\langle Z_t, \varphi\rangle$。条件于当前粒子配置，在短时间 h 内，对位置为 x 的一个粒子，三种机制的平均一阶变化如下：

(1) 运动贡献为 $hL\varphi(x) + o(h)$。

(2) 分裂以概率 $\lambda(x)h + o(h)$ 发生，统计量由 $\varphi(x)$ 变为 $2\varphi(x)$，净增量为 $\varphi(x)$。

(3) 死亡以概率 $\mu(x)h+o(h)$ 发生，统计量由 $\varphi(x)$ 变为零，净增量为 $-\varphi(x)$。

将粒子贡献相加，形式上得到

$$\mathbb{E}[\langle Z_{t+h},\varphi\rangle-\langle Z_t,\varphi\rangle\mid\mathcal{F}_t]=h\langle Z_t,L\varphi+r\varphi\rangle+o(h).$$

严格推导可以先对有限粒子数和有界区域停止，再用可积性取极限。其积分形式是

$$\langle\nu_t,\varphi\rangle=\langle\nu_0,\varphi\rangle+\int_0^t\langle\nu_s,L\varphi+r\varphi\rangle\,\mathrm{d}s. \tag{11.5}$$

这已经是一阶矩演化方程，不要求存在光滑密度。若 f 足够光滑，利用生成元与前向算子的对偶关系，可写成

$$\boxed{\partial_t f=L^*f+r(x)f,\qquad L^*f=-\partial_x(bf)+\frac{1}{2}\partial_{xx}(\sigma^2 f).} \tag{11.6}$$

第一项负责重新分配位置，第二项负责增加或减少质量。方程对 f 是线性的，因为本模型中的 b,σ,r 都不依赖 f。

说明 11.1 (一阶矩方程没有描述全部随机性)**.** 不同的出生率和死亡率可能有相同的净增长率 r。例如 $(\lambda,\mu)=(1,0)$ 和 $(101,100)$ 都给出 $r=1$，因而具有相同的一阶矩方程，但粒子数涨落和灭绝概率不同。只求解 (11.6) 无法恢复整个随机种群的分布。

11.4 归一化为何产生非线性

首先计算总质量。在全空间零远场通量、周期边界或反射边界等守恒条件下，

$$\int L^*f\,\mathrm{d}x=0.$$

因此由 (11.6) 得到

$$\begin{aligned}\dot M(t)&=\int\partial_t f(x,t)\,\mathrm{d}x=\int r(x)f(x,t)\,\mathrm{d}x\\&=M(t)\int r(x)p(x,t)\,\mathrm{d}x.\end{aligned} \tag{11.7}$$

定义相对于归一化强度的平均增长率

$$\bar r(t):=\int r(x)p(x,t)\,\mathrm{d}x. \tag{11.8}$$

于是 $\dot M/M=\bar r$。再对 $p=f/M$ 使用普通商法则：

$$\begin{aligned}\partial_t p&=\frac{\partial_t f}{M}-\frac{f\dot M}{M^2}\\&=\frac{L^*f+rf}{M}-p\frac{\dot M}{M}\\&=L^*p+rp-p\bar r.\end{aligned} \tag{11.9}$$

第三行用了 M 仅依赖 t、不依赖 x，所以 $L^*(Mp)=ML^*p$。最终得到

$$\boxed{\partial_t p=L^*p+\big(r(x)-\bar r(t)\big)p,\qquad \bar r(t)=\int r(y)p(y,t)\,\mathrm{d}y.} \tag{11.10}$$

这常被理解为“运动加选择”方程：局部增长率高于平均值的位置得到相对增益，低于平均值的位置失去相对占比。即使所有位置都在增长，增长较慢的位置仍可能失去比例。

命题 11.2 (质量与形状可以相互恢复)**.** 在上述守恒条件下，若 p 满足 (11.10) 且初始积分为一，则其积分保持为一。总质量由

$$M(t) = M(0)\exp\left(\int_0^t \bar{r}(s)\,\mathrm{d}s\right) \tag{11.11}$$

恢复，而 $f = Mp$ 满足线性方程 (11.6)。

证明. 令 $S(t) = \int p(x,t)\,\mathrm{d}x$。积分 (11.10) 后有

$$S'(t) = \bar{r}(t) - \bar{r}(t)S(t) = \bar{r}(t)(1 - S(t)).$$

初值 $S(0) = 1$ 给出 $S(t) = 1$。式 (11.11) 直接解出 $\dot{M} = \bar{r}M$。最后，

$$\partial_t(Mp) = \bar{r}Mp + M\big(L^*p + (r - \bar{r})p\big) = L^*(Mp) + rMp.$$

□

非线性来自 $p \mapsto \bar{r}[p]p$：平均增长率本身随未知分布改变。它不是额外假设的粒子相互作用，而是将线性强度方程除以其总质量的结果。若 $r \equiv c$ 为常数，选择项消失，p 按普通 Fokker–Planck 方程演化，$M(t) = M(0)e^{ct}$。

11.5 一个必须澄清的抽样区别

原始种群的随机性没有在取比例时消失。对集合 A，一般有

$$\underbrace{\frac{\mathbb{E}[Z_t(A)]}{\mathbb{E}[N_t]}}_{\pi_t(A)} \neq \underbrace{\mathbb{E}\left[\frac{Z_t(A)}{N_t}\,\middle|\, N_t > 0\right]}_{\text{先抽一次存活种群，再在其中均匀抽一个粒子}}. \tag{11.12}$$

左边是“期望之比”，右边是“随机比例的条件期望”，两种操作没有交换律。

例 11.3 (两种抽样机制给出不同答案)**.** 某时刻的种群有两个等概率的可能结果：结果一包含一个 A 型粒子，结果二包含九个 B 型粒子。这里没有灭绝，因而

$$\mathbb{E}[N_t] = \tfrac{1}{2}\cdot 1 + \tfrac{1}{2}\cdot 9 = 5, \qquad \mathbb{E}[Z_t(A)] = \tfrac{1}{2}.$$

归一化强度给出的 A 型概率是 $\pi_t(A) = 1/10$。但是先以原始概率选择一次种群，再在其中均匀抽样，选到 A 型的概率是 $1/2$。这是两个明确定义且不同的实验。

那么 π_t 对应什么抽样？在固定时刻 t，按总粒子数对种群结果重新加权：

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}_t}{\mathrm{d}\mathbb{P}} = \frac{N_t}{\mathbb{E}N_t}. \tag{11.13}$$

这称为按种群大小加权。它使大种群更容易被选中，并赋予灭绝结果零权重。再在选中的种群内均匀抽一个粒子，则

$$\widetilde{\mathbb{E}}_t\left[\frac{Z_t(A)}{N_t}\right] = \mathbb{E}\left[\frac{N_t}{M(t)}\frac{Z_t(A)}{N_t}\mathbf{1}_{\{N_t>0\}}\right] = \frac{\mathbb{E}Z_t(A)}{M(t)} = \pi_t(A).$$

等价地，在许多独立重复实验中，将所有终点粒子合并后均匀抽样，其大样本分布趋于 π_t，前提是一阶矩有限且 $M(t) > 0$。

说明 11.4 (灭绝与有限种群效应)**.** $M(t)>0$ 不意味着 $N_t>0$ 几乎必然；平均种群数增长也不排除一部分样本已经灭绝。一次有限种群的 $\widehat{\pi}_t$ 会有随机涨落。它不能直接当作确定性方程 (11.10) 的精确解，条件于存活后的均值也通常不满足同一闭合方程。若希望证明大量独立祖先或相互作用粒子系统收敛到该方程，需要相应的大数定律或均场极限定理。

11.6　Feynman–Kac：用一条路径计算整个种群的一阶矩

分枝系统含有很多条随机轨迹，但其一阶矩可以用一条不分枝的扩散轨迹表达。为简洁起见，仍设 r 有界。令 X 只按生成元 L 运动，定义路径权重

$$w_t := \exp\!\left(\int_0^t r(X_s)\,\mathrm{d}s\right). \tag{11.14}$$

轨迹在高增长率区域停留得越久，权重越大。若初始期望计数测度为 ν_0，则对有界测试函数 φ，

$$\boxed{\langle \nu_t,\varphi\rangle = \int \nu_0(\,\mathrm{d}x)\,\mathbb{E}_x[w_t\varphi(X_t)]\,.} \tag{11.15}$$

左边需要对全部后代求和并取期望，右边只需跟踪一个不分枝粒子；这是一阶矩的 Feynman–Kac 表示，也称 many-to-one 公式。

下面解释它为什么与前向方程一致。固定终点时间 T，定义

$$u(x,s) := \mathbb{E}_{x,s}\!\left[\exp\!\left(\int_s^T r(X_v)\,\mathrm{d}v\right)\varphi(X_T)\right], \qquad 0\le s\le T. \tag{11.16}$$

在足够光滑的条件下，u 满足带势项的后向终值问题

$$\partial_s u + Lu + ru = 0, \qquad u(x,T)=\varphi(x). \tag{11.17}$$

这里终值问题中 ru 的符号为正；若使用杀灭率 $V\ge 0$，则 $r=-V$，相应路径权重为 $e^{-\int V}$。

Itô 验证。　设 $A_s=\exp(\int_0^s r(X_v)\,\mathrm{d}v)$。它是有限变差过程，所以与 $u(X_s,s)$ 的二次协变差为零。乘积法则给出

$$\begin{aligned}\mathrm{d}\big(A_s u(X_s,s)\big) &= A_s\big(\partial_s u + Lu + ru\big)(X_s,s)\,\mathrm{d}s \\ &\quad + A_s\sigma(X_s)\partial_x u(X_s,s)\,\mathrm{d}W_s.\end{aligned}$$

当 u 满足 (11.17) 时漂移消失。先局部化，再在有界 r,φ 等条件下验证真鞅性质，就可从终值推出 (11.16)。这一步同时说明为什么指数中的增长率必须沿整条路径积分，而不是简单写成 $e^{Tr(X_T)}$。

前后向配对。　若 f 满足 (11.6)、u 满足 (11.17)，并且边界项消失，则

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\int u(x,s)f(x,s)\,\mathrm{d}x &= \int\big[(-Lu-ru)f + u(L^*f + rf)\big]\,\mathrm{d}x \\ &= 0.\end{aligned}$$

因此

$$\int \varphi(x) f(x,T)\,\mathrm{d}x = \int u(x,0) f(x,0)\,\mathrm{d}x,$$

代入 (11.16) 就得到 (11.15)。对非光滑初值可以在测度和半群框架下表述同一恒等式。

将 $\nu_0 = M(0)\pi_0$ 代入并分别取 φ 和 1，得到归一化形式

$$\boxed{\mathbb{E}_{\pi_t}[\varphi] = \frac{\mathbb{E}_{\pi_0}[w_t \varphi(X_t)]}{\mathbb{E}_{\pi_0}[w_t]}, \qquad M(t) = M(0)\mathbb{E}_{\pi_0}[w_t].} \tag{11.18}$$

这里右边的期望是对普通扩散路径取的，左边是对其加权终点分布取的。它将分枝、路径加权和非线性归一化方程连接在一起。

11.7 无运动时：复制子方程与显式解

设粒子不移动，$L^* = 0$，且 $r = r(x)$ 不随时间变化。未归一化方程可逐点求解：

$$f(x,t) = f(x,0)e^{tr(x)}.$$

除以总质量可得

$$p(x,t) = \frac{p(x,0)e^{tr(x)}}{\int p(y,0)e^{tr(y)}\,\mathrm{d}y}. \tag{11.19}$$

这要求分母有限，且对给定时间为正。高增长率位置的相对权重会指数增加；初始分布完全没有质量的位置，在没有运动或突变时仍然不会出现质量。

例 11.5 (两种类型的比例演化)**.** 只有 A, B 两种类型，净增长率为 r_A, r_B，初始 A 型比例为 $q_0 \in (0,1)$。则

$$q(t) = \frac{q_0 e^{r_A t}}{q_0 e^{r_A t} + (1-q_0)e^{r_B t}} = \frac{1}{1 + \frac{1-q_0}{q_0} e^{-(r_A - r_B)t}}. \tag{11.20}$$

直接求导得到

$$q'(t) = (r_A - r_B) q(t)(1 - q(t)).$$

这在形式上是 logistic 方程，但它描述的是相对比例，不是总粒子数的承载容量。只要 $r_A > r_B$，A 型比例就趋向一；即使两种类型都在衰退，例如 $r_A = -1, r_B = -2$，这一结论仍成立，此时总期望粒子数却趋向零。

给所有位置的增长率同时加上常数 c，即用 $r + c$ 替代 r，不会改变归一化方程，因为 $(r + c) - (\bar{r} + c) = r - \bar{r}$；它只把总质量乘以 e^{ct}。因此归一化形状识别的是相对增长优势。

11.8 均值如何改变：运动项与协方差项

设 φ 与时间无关且满足所需的可积性和边界条件。对 (11.10) 乘以 φ 并积分：

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{p_t}[\varphi] &= \int \varphi L^* p_t\,\mathrm{d}x + \int \varphi(r - \bar{r}) p_t\,\mathrm{d}x \\ &= \mathbb{E}_{p_t}[L\varphi] + \mathbb{E}_{p_t}[\varphi r] - \mathbb{E}_{p_t}[\varphi]\mathbb{E}_{p_t}[r]. \end{aligned} \tag{11.21}$$

于是

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{p_t}[\varphi] = \mathbb{E}_{p_t}[L\varphi] + \mathrm{Cov}_{p_t}(\varphi, r).} \tag{11.22}$$

第一项是运动对统计量的影响，第二项是选择对统计量的影响。若 $\varphi(x) = x$ 且矩有限，

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{p_t}[X] = \mathbb{E}_{p_t}[b(X)] + \mathrm{Cov}_{p_t}(X, r(X)).$$

即使漂移 b 为零，只要大 x 与高增长率相关，分布均值仍会向较大的 x 移动。

若 r 不随时间变化、足够光滑且各项有限，令 $\varphi = r$ 可得

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{r}(t) = \mathbb{E}_{p_t}[Lr] + \mathrm{Var}_{p_t}(r). \tag{11.23}$$

在没有运动时，$Lr = 0$，平均增长率因此单调不减。存在运动时，$\mathbb{E}[Lr]$ 可能为负，不能只保留方差项便断言平均增长率总在上升。若 $r = r(x,t)$，右边还应增加 $\mathbb{E}_{p_t}[\partial_t r]$。

11.9 平稳分布与特征值问题

若存在归一化平稳解 p_*，则

$$0 = L^* p_* + (r - \rho)p_*, \qquad \rho := \int r(x)p_*(x)\,\mathrm{d}x.$$

等价地，

$$(L^* + r)p_* = \rho p_*, \qquad p_* \geq 0, \qquad \int p_* = 1. \tag{11.24}$$

归一化后的稳态形状，因而对应于线性增长算子的非负特征函数；在这个形状下，总质量以速率 ρ 指数增长或衰减。

但“存在平稳解”“它对应主特征值”和“任意初值都收敛到它”是三件需要分别证明的事。在有界连通区域、适当的反射或周期边界、一致椭圆性及正性、紧性等条件下，可以利用正算子的主特征值理论得到相应结论。无界空间中可能没有可归一化的正特征函数，也可能没有收敛；不能单靠写出 (11.24) 就宣布已经找到长期极限。

11.10 非守恒边界会改变归一化项

考虑有界区域 D，令 J_f 是强度的向外概率流，定义净流出量

$$B(t) := \int_{\partial D} J_f \cdot n \,\mathrm{d}S.$$

若 $\partial_t f = -\nabla \cdot J_f + rf$，则

$$\dot{M} = M\bar{r} - B.$$

例如吸收边界使粒子离开区域后被删除，通常产生非零质量损失。此时商法则给出

$$\partial_t p = L^* p + \left(r - \bar{r} + \frac{B(t)}{M(t)}\right) p. \tag{11.25}$$

因为 $\int_D L^* p\,\mathrm{d}x = -B/M$，新增项正好抵消边界损失并维持归一化。反射边界下 $B = 0$，才恢复 (11.10)。若只研究被杀灭前的路径，Feynman–Kac 表示也必须加入存活事件的指示函数。

11.11 计算思路：运动、加权、再归一化

式 (11.18) 提供了一种直接的数值理解。把时间分成小步长 h，每一步依次进行：

(1) 按扩散运动更新粒子，例如用 Euler–Maruyama 步；

(2) 以 $e^{hr(x)}$ 更新权重，近似这一步累积的增长；

(3) 将权重除以总和，使其成为概率权重；如需固定粒子数，可按这些权重重采样。

在只有选择而没有运动的一小步中，更新是精确的：

$$p^{\mathrm{new}}(x) = \frac{e^{hr(x)}p^{\mathrm{old}}(x)}{\int e^{hr(y)}p^{\mathrm{old}}(y)\,\mathrm{d}y}.$$

展开 $e^{hr} = 1 + hr + O(h^2)$ 以及分母的倒数，得到

$$p^{\mathrm{new}} = p^{\mathrm{old}} + h(r - \bar{r})p^{\mathrm{old}} + O(h^2),$$

恰好对应归一化方程中的选择项。运动与选择依次执行属于算子分裂，步长误差与有限粒子数的 Monte Carlo 误差需要分别考虑。

真实分枝模拟的粒子数是随机的，可能灭绝或迅速膨胀；固定数量的加权、重采样算法则是另一种近似计算方法。两者能够在适当条件下逼近同一个归一化一阶矩分布，但它们并不是同一个有限粒子随机模型。

本节自检。 遇到“分枝后的分布”时，先明确三个问题：研究的是计数测度还是概率测度；取期望在归一化之前还是之后；运动是否真的守恒。确定这三点后，方程中的增长项、归一化项和边界项才有唯一含义。

12 把公式变成数值实验

12.1 Euler–Maruyama：噪声的尺度不能写错

把时间区间分为 $t_k = kh$，取相互独立的 $\xi_k \sim \mathcal{N}(0,1)$，令

$$X_{k+1} = X_k + b(X_k, t_k)h + \sigma(X_k, t_k)\sqrt{h}\,\xi_k. \tag{12.1}$$

这称为 Euler–Maruyama 格式。它把两个积分中的系数冻结在左端点。因为 $\Delta W_k \sim \mathcal{N}(0,h)$，噪声必须乘 $\sqrt{h}$。若错写成 $h\xi_k$，在单位时间内累计的噪声方差只有 $(1/h)h^2 = h$，会随步长消失。对 m 维独立布朗运动，每一步应抽取 m 个独立标准正态数，再乘矩阵 Σ。

说明 12.1. 对 Stratonovich 方程，不能不加转换就直接套用式 (12.1)。最直接的做法是先把漂移改成等价的 Itô 漂移，再用这个格式。数值上得到一条曲线，也不意味着方法保持正性、边界约束或原方程的平稳分布。

12.2 强误差、弱误差和 Monte Carlo 误差

强误差比较同一噪声驱动下的两条轨道，例如固定 $T = Nh$，

$$e_{\text{strong}}(h) = \big(\mathbb{E}|X_N - X_T|^2\big)^{1/2}.$$

弱误差比较观测的期望，例如对适当光滑的 φ，

$$e_{\text{weak}}(h;\varphi) = |\mathbb{E}\varphi(X_N) - \mathbb{E}\varphi(X_T)|.$$

在标准的全局 Lipschitz、矩有界和时间正则性等条件下，Euler–Maruyama 的典型强阶为 $1/2$；再加上足够的系数及测试函数光滑性，典型弱阶为 1。这些是有假设的定理，不对任意非线性或超线性系数自动成立。加性噪声等特殊结构下，强阶可能提高。

使用 K 个独立样本估计期望还会产生统计误差。若样本方差有限，标准误差为

$$\text{SE} = \sqrt{\text{Var}(\varphi(X_N))/K}.$$

它与步长偏差是两种不同来源。如果步长已经很小，但 K 不够大，观察到的误差可能由抽样噪声主导。验证强收敛时，粗网格增量必须由同一组细网格增量求和构造；独立模拟两条轨道不能检验强误差。

12.3 Milstein 格式为什么多一项

一维情形把噪声系数再展开一阶，会出现嵌套积分

$$\sigma(X_k,t_k)\partial_x\sigma(X_k,t_k)\int_{t_k}^{t_{k+1}}\int_{t_k}^{s} \mathrm{d}W_u\,\mathrm{d}W_s.$$

由 $\int W\,\mathrm{d}W$ 的公式，内外积分合起来等于 $\frac{1}{2}((\Delta W_k)^2 - h)$。因此 Milstein 格式是

$$\begin{aligned} X_{k+1} = X_k &+ b(X_k,t_k)h + \sigma(X_k,t_k)\Delta W_k \\ &+ \tfrac{1}{2}\sigma(X_k,t_k)\partial_x\sigma(X_k,t_k)\big((\Delta W_k)^2 - h\big). \end{aligned} \tag{12.2}$$

在相应光滑性与增长控制假设下，它的一维典型强阶为 1。这是对随机积分展开的更精细近似，不能把这里的随机校正项与 Stratonovich 漂移修正混为一谈。一般多维噪声的高阶格式还涉及交叉迭代积分，不能机械地逐坐标套用一维公式。

12.4 实验一：轨道很不规则，二次变差却稳定

图 1 左边显示六条布朗轨道的网格值，右边检验 $T = 1$ 时

$$Q_n = \sum_{k=0}^{n-1}(W_{(k+1)/n} - W_{k/n})^2, \qquad \mathbb{E}[(Q_n - 1)^2] = \frac{2}{n}.$$

右图对 4000 条独立轨道取平均；它检验的是平方误差期望，而不是要求每条轨道的 Q_n 单调趋于 1。

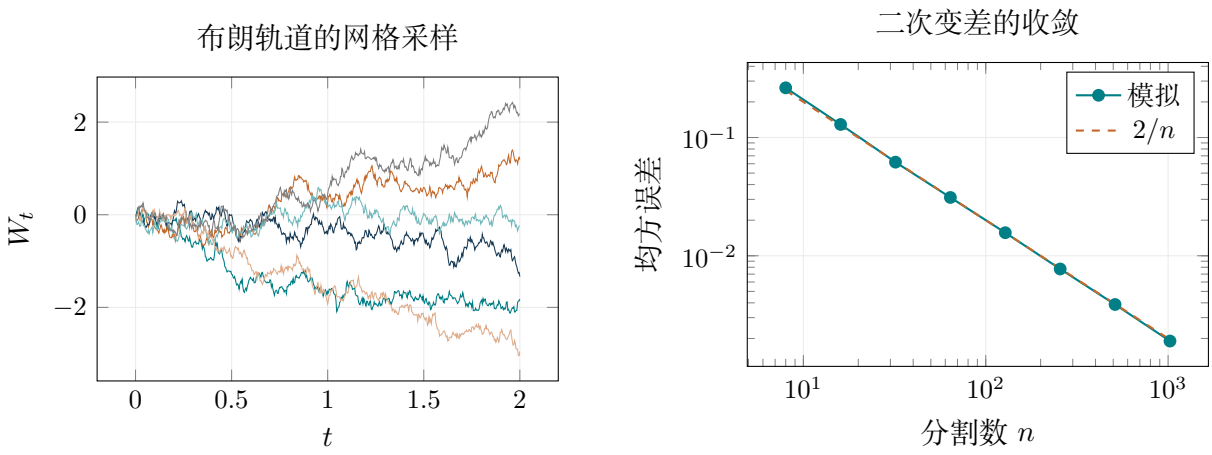


图 1: 固定随机种子生成的可复现实验。左图的线段只是离散点的可视化，真实布朗轨道不是分段线性曲线。

12.5　实验二：用粒子直方图检查 Fokker–Planck 密度

考虑 $\mathrm{d}X_t = -1.2X_t\,\mathrm{d}t + 0.8\,\mathrm{d}W_t$，$X_0 = 2$，终点 $T = 1.5$。用步长 $h = 0.002$ 模拟 20000 条 Euler–Maruyama 轨道，终点直方图应接近

$$\mathcal{N}\left(2e^{-1.2T}, \frac{0.8^2}{2(1.2)}(1 - e^{-2(1.2)T})\right).$$

这里比较的是原 SDE 的精确密度与离散模拟，误差同时包含时间离散偏差、抽样波动和直方图分箱误差。

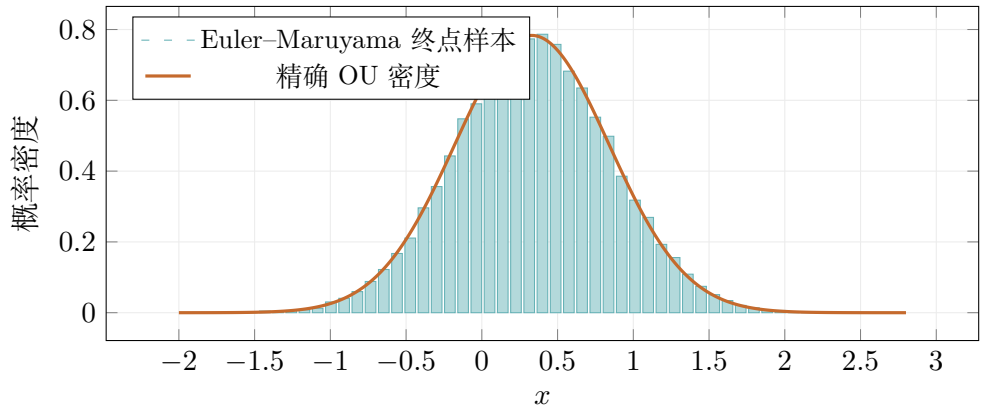


图 2: 随机轨道的集合与确定性密度方程描述同一系统的两个层面。

12.6　实验三：同一条布朗路径上的数值方法比较

取几何布朗运动 $\mathrm{d}X_t = 0.3X_t\,\mathrm{d}t + 0.6X_t\,\mathrm{d}W_t$，$X_0 = 1$，$T = 1$。对每个样本先生成 1024 个细网格增量，再相加生成所有粗网格增量。精确终值为 $X_T = \exp((0.3 - 0.6^2/2)T + 0.6W_T)$。图 3 用 10000 个样本估计终点均方根误差。

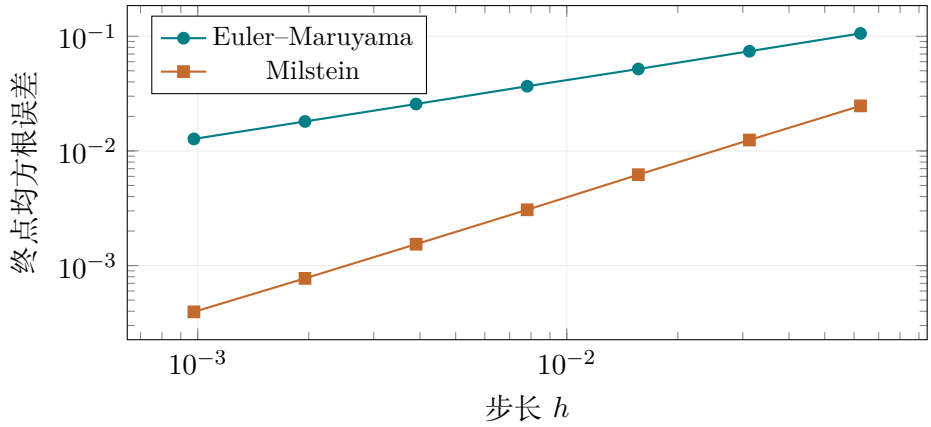


图 3: 同一批耦合噪声下的强误差。对数坐标中的斜率给出该实验的观测收敛阶。

检验量	数值结果	理论值或典型阶
OU 终端均值	0.3371	0.3306
OU 终端方差	0.2623	0.2594
Euler–Maruyama 观测强阶	0.509	1/2
Milstein 观测强阶	0.998	1

表中的观测阶由全部七个步长的对数误差作直线拟合得到，属于数值证据，不代替收敛定理。全部数据由同目录的 `generate_experiments.py` 生成，依赖 Python 与 NumPy，随机种子固定为 20261122。`data/summary.json` 保存样本量、理论值与实测结果。

12.7 OU 的离散平稳分布：一个额外的检查

对常参数 OU，Euler–Maruyama 给出

$$X_{k+1} = (1-\theta h)X_k + \sigma\sqrt{h}\,\xi_k.$$

只有当 $|1-\theta h|<1$，即 $0<\theta h<2$，离散系统才有稳定的方差递推。其平稳方差为

$$v_h = (1-\theta h)^2 v_h + \sigma^2 h \quad\Longrightarrow\quad v_h = \frac{\sigma^2}{2\theta-\theta^2 h}.$$

连续 SDE 的平稳方差是 $\sigma^2/(2\theta)$。因此即便离散模拟很稳定，有限步长也会带来稳态偏差。OU 还有精确网格更新

$$X_{k+1} = e^{-\theta h}X_k + \sigma\sqrt{\frac{1-e^{-2\theta h}}{2\theta}}\,\xi_k,$$

它精确保持 OU 的转移分布。一般 SDE 没有这样的闭式更新，所以理解离散误差十分必要。

13 选读：与扩散生成模型的联系

本节把已有的概率流公式用于扩散生成模型，作为后续阅读的入口。主要考虑 $\mathbb{R}^d$ 上各向同性、空间无关的噪声：

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t, \qquad a(t) = g(t)^2 I_d. \tag{13.1}$$

假设密度 p_t 严格为正且足够光滑，过程不爆炸，并且以下 PDE、流和时间反演均适定。在数据分布不光滑的端点，通常要先考虑 $t > 0$ 或从一个小的正时间开始。这类 SDE、反向过程与 probability flow ODE 的联系可参阅原始论文 [5]。

13.1 从 Fokker–Planck 直接得到概率流速度

定义 score 为 $s_t(x) = \nabla \log p_t(x)$。因为 $\nabla p_t = p_t s_t$，

$$\begin{aligned} \partial_t p_t &= -\nabla \cdot (b p_t) + \tfrac{1}{2} g(t)^2 \Delta p_t \\ &= -\nabla \cdot \left[\left(b - \tfrac{1}{2} g(t)^2 s_t \right) p_t \right]. \end{aligned} \tag{13.2}$$

因此概率流速度是

$$v(x,t) = b(x,t) - \tfrac{1}{2} g(t)^2 \nabla \log p_t(x). \tag{13.3}$$

若 ODE $\dot{Y}_t = v(Y_t, t)$ 的流存在、其连续性方程有适当唯一性，且 Y_0 与 X_0 同分布，则 Y_t 与 X_t 在每个时刻具有相同边缘分布。这一结论说的是**单时刻分布相同**，不是轨道相同、联合分布相同或路径测度相同。当 $g \neq 0$ 时，SDE 的轨道有非零二次变差；光滑 ODE 轨道的二次变差为零。

13.2 反向 SDE：先定义一个递增的反向时钟

设总时间为 T，定义 $s = T - t$，$Y_s = X_{T-s}$，所以新的时间 s 从 0 增加到 T。在通常的时间反演正则条件下，反向过程满足

$$\mathrm{d}Y_s = \left[-b(Y_s, T-s) + g(T-s)^2 s_{T-s}(Y_s) \right] \mathrm{d}s + g(T-s)\, \mathrm{d}B_s, \tag{13.4}$$

其中 B_s 是相对于反向滤过的布朗运动，不能直接视作原布朗增量的机械倒序。这是时间反演定理的结论；本讲义不证明该定理。可以用前向方程检查它的符号：对 $q_s = p_{T-s}$，反向漂移 $\beta = -b + g^2 \nabla \log p$ 给出

$$-\nabla \cdot (\beta q_s) + \tfrac{1}{2} g^2 \Delta q_s = \nabla \cdot (b q_s) - \tfrac{1}{2} g^2 \Delta q_s = -\partial_t p_t \big|_{t=T-s} = \partial_s q_s.$$

这个检查验证边缘密度方程的方向，单独并不构成完整的路径时间反演证明。

常见文献也沿原时间从 T 向 0 积分，此时漂移写作 $b - g^2 s_t$，且 $\mathrm{d}t < 0$。使用公式时必须连同时间方向一起阅读。反向 SDE 中 score 的系数是 1，概率流 ODE 中是 1/2；这与 Itô/Stratonovich 的积分约定转换不是一回事。这里 g 不依赖空间，Itô/Stratonovich 漂移修正恰好为零，但时间反演仍会改变漂移。

13.3 用一个高斯例子同时核对三个模型

取一维 $b = 0$、$g = 1$、$X_0 \sim \mathcal{N}(0, v_0)$，其中 $v_0 > 0$。则 $p_t = \mathcal{N}(0, v_0 + t)$，score 为 $s_t(x) = -x/(v_0 + t)$。对应的概率流 ODE 为

$$\dot{Y}_t = \frac{Y_t}{2(v_0 + t)}, \qquad Y_t = Y_0 \sqrt{\frac{v_0 + t}{v_0}}.$$

因此 $\mathrm{Var}(Y_t) = v_0 + t$，与前向布朗扩散完全一致；但这个 ODE 的整条轨道只由 Y_0 决定。对反向时钟，式 (13.4) 变成

$$\mathrm{d}Z_s = -\frac{Z_s}{v_0 + T - s}\,\mathrm{d}s + \mathrm{d}B_s, \qquad Z_0 \sim \mathcal{N}(0, v_0 + T).$$

若记方差为 $V(s)$，矩方程给出

$$V'(s) = -\frac{2V(s)}{v_0 + T - s} + 1.$$

代入 $V(s) = v_0 + T - s$，两边都等于 -1，所以反向过程恰好收缩回初始高斯。这个例子直接展示：扩散的前向加噪、反向去噪、同边缘的确定性输运可以描述同一族密度，却通过不同的随机轨道机制实现。

说明 13.1 (空间相关噪声的附加项)**.** 若 $a = a(x,t)$，概率流速度的一般表达式为

$$v_i = b_i - \frac{1}{2p}\sum_j \partial_j(a_{ij}p) = b_i - \tfrac{1}{2}\sum_j \partial_j a_{ij} - \tfrac{1}{2}\sum_j a_{ij}\partial_j \log p.$$

反向时钟下的漂移相应为 $-b + \nabla\cdot a + a\nabla\log p$，其中 $(\nabla\cdot a)_i = \sum_j \partial_j a_{ij}$。所以空间相关扩散不能直接套用只含 $g(t)^2\nabla\log p$ 的简式。

14 综合习题与解答

建议先遮住解答独立完成。每题不仅要给出最后一个公式，还应说明用了哪个独立性、积分或边界条件。

习题 1 (条件期望与相关性)**.** 令 $0 < s < t$。求 $\mathbb{E}[W_t^2 \mid \mathcal{F}_s]$、$\mathrm{Cov}(W_s, W_t)$ 与两者的相关系数。

解答. 写 $W_t = W_s + Z$，其中 $Z = W_t - W_s$ 与 $\mathcal{F}_s$ 独立，$\mathbb{E}Z = 0$，$\mathbb{E}Z^2 = t - s$。于是

$$\mathbb{E}[W_t^2 \mid \mathcal{F}_s] = W_s^2 + t - s, \quad \mathrm{Cov}(W_s, W_t) = s, \quad \mathrm{Corr}(W_s, W_t) = \sqrt{s/t}.$$

特别地，$W_t^2 - t$ 是鞅，W_s 与 W_t 通常并不独立。 □

习题 2 (确定性与随机被积函数)**.** 分别求 $A = \int_0^T s\,\mathrm{d}W_s$ 和 $B = \int_0^T W_s\,\mathrm{d}W_s$ 的均值与方差。它们是否都是高斯变量?

解答. 等距给出 $\mathbb{E}A = 0$、$\mathrm{Var}(A) = \int_0^T s^2\,\mathrm{d}s = T^3/3$。确定性简单函数的积分是高斯变量，其 L^2 极限仍为相应高斯变量，故 $A \sim \mathcal{N}(0, T^3/3)$。另一方面

$$B = \tfrac{1}{2}(W_T^2 - T), \qquad \mathbb{E}B = 0, \qquad \mathrm{Var}(B) = \int_0^T \mathbb{E}W_s^2\,\mathrm{d}s = T^2/2.$$

$T > 0$ 时 $B \geq -T/2$ 且非退化，因此不可能是高斯变量。随机积分不因为驱动是布朗运动就自动服从高斯分布。 □

习题 3 (四阶矩)**.** 不用直接计算高斯积分，用 Itô 公式求 $\mathbb{E}W_t^4$。

解答. 取 $f(x)=x^4$ 得 $\mathrm{d}(W_t^4)=4W_t^3\,\mathrm{d}W_t+6W_t^2\,\mathrm{d}t$。在有限时间区间，高斯六阶矩有限，随机积分的期望为零，因此

$$\mathbb{E}W_t^4=6\int_0^t\mathbb{E}W_s^2\,\mathrm{d}s=6\int_0^t s\,\mathrm{d}s=3t^2.$$

必要时也可先停时再去掉局部化。 □

习题 4 (共享噪声的乘积法则)**.** 设 $\mathrm{d}X_t=a_t\,\mathrm{d}t+u_t\,\mathrm{d}W_t$，$\mathrm{d}Y_t=c_t\,\mathrm{d}t+v_t\,\mathrm{d}W_t$。求 $\mathrm{d}(X_tY_t)$。若两者由独立布朗运动驱动，会改变哪一项?

解答. 因为 $\mathrm{d}[X,Y]_t=u_tv_t\,\mathrm{d}t$，

$$\mathrm{d}(X_tY_t)=(X_tc_t+Y_ta_t+u_tv_t)\,\mathrm{d}t+(X_tv_t+Y_tu_t)\,\mathrm{d}W_t.$$

独立驱动时二次协变差项为零，且随机项要保留为相应的两个随机积分。只看普通乘积法则会漏掉 u_tv_t。 □

习题 5 (OU 的矩方程)**.** 对 $\mathrm{d}X_t=-\theta X_t\,\mathrm{d}t+\sigma\,\mathrm{d}W_t$、$X_0=x_0$、$\theta>0$，不用显式解推导均值与方差。

解答. 取观测 x 与 x^2：$m'=-\theta m$，$q'=-2\theta q+\sigma^2$，初值分别为 x_0 和 x_0^2。解得

$$m=x_0e^{-\theta t},\quad q=x_0^2e^{-2\theta t}+\frac{\sigma^2}{2\theta}(1-e^{-2\theta t}).$$

减去 m^2 即得到方差。线性漂移与常数噪声让这两条低阶矩方程闭合。 □

习题 6 (几何布朗运动的方差与典型增长)**.** 设 $X_t=x_0\exp((\mu-\sigma^2/2)t+\sigma W_t)$，$x_0>0$。求 $\mathrm{Var}(X_t)$，并解释为什么均值增长率与 $\mathbb{E}\log X_t$ 的增长率不同。

解答. 高斯矩母函数给出 $\mathbb{E}X_t^2=x_0^2e^{(2\mu+\sigma^2)t}$，故

$$\mathrm{Var}(X_t)=x_0^2e^{2\mu t}(e^{\sigma^2t}-1).$$

均值按 $e^{\mu t}$ 增长，而 $\mathbb{E}\log X_t=\log x_0+(\mu-\sigma^2/2)t$。指数变换会放大少数很大的样本；因此先取均值与先取对数不交换。当 $0<\mu<\sigma^2/2$ 时，均值增加而平均对数减少并不矛盾。 □

习题 7 (乘性噪声转换与局部性)**.** 在 $X_t>0$ 的区域，转换 $\mathrm{d}X_t=\alpha\,\mathrm{d}t+\beta\sqrt{X_t}\circ\mathrm{d}W_t$ 为 Itô 形式。为什么不能由此直接推出 X_t 永远为正?

解答. 在 $x>0$ 上，$\sigma(x)=\beta\sqrt{x}$，$\sigma'(x)=\beta/(2\sqrt{x})$，所以 $\frac12\sigma\sigma'=\beta^2/4$。转换后为

$$\mathrm{d}X_t=(\alpha+\beta^2/4)\,\mathrm{d}t+\beta\sqrt{X_t}\,\mathrm{d}W_t.$$

导数在 0 处奇异；转换论证在离开正半轴之前适用。边界能否到达、到达以后怎样定义，需要额外的边界分类或存在唯一性论证。 □

习题 8 (不要把变量系数移出导数)**.** 对 $x>0$ 上的 SDE $\mathrm{d}X_t=\mu X_t\,\mathrm{d}t+\sigma X_t\,\mathrm{d}W_t$ 写出 FP 方程，并展开所有空间导数。

解答.

$$\partial_t p = -\mu\partial_x(xp) + \tfrac{1}{2}\sigma^2\partial_{xx}(x^2 p) = (-\mu+\sigma^2)p + (-\mu+2\sigma^2)x\partial_x p + \tfrac{1}{2}\sigma^2 x^2\partial_{xx}p.$$

最后一式的三项必须一起保留。单独写 $-\mu x\partial_x p + \frac{1}{2}\sigma^2 x^2\partial_{xx}p$ 是把生成元作用于密度，通常得不到正确的前向方程。 □

习题 9 (反射区间的平稳分布)**.** 在 $[0,1]$ 上考虑漂移 c、噪声振幅 $\sqrt{2D}$ 的反射扩散，其中 $D>0$。求其平稳密度。

解答. 反射边界要求两端 $J = cp - Dp' = 0$。平稳时 $J' = 0$，所以整个区间上 $J = 0$，即 $p' = (c/D)p$。归一化得

$$p_*(x) = \begin{cases} \dfrac{c}{D(e^{c/D}-1)}e^{cx/D}, & c \neq 0, \\ 1, & c = 0. \end{cases}$$

这里反射过程在轨道层面需要边界反射项；只写无约束的 $\mathrm{d}X = c\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W$ 不会把粒子限制在区间内。 □

习题 10 (终端后向方程)**.** 设 $\mathrm{d}X_s = \mu\,\mathrm{d}s + \sigma\,\mathrm{d}W_s$，终端观测为 X_T^2。求 $v(x,t) = \mathbb{E}[X_T^2 \mid X_t = x]$ 并验证后向方程。

解答. 令 $\tau = T - t$，则

$$v(x,t) = (x+\mu\tau)^2 + \sigma^2\tau.$$

因此 $v_t = -2\mu(x+\mu\tau) - \sigma^2$、$v_x = 2(x+\mu\tau)$、$v_{xx} = 2$，代入得 $v_t + \mu v_x + \frac{1}{2}\sigma^2 v_{xx} = 0$，且 $v(x,T) = x^2$。 □

习题 11 (无运动的两类型选择)**.** 两类型增长率分别为 r_A, r_B，初始归一化比例为 $q_0, 1-q_0$，其中 $0 < q_0 < 1$。求 $q(t)$，并判断常数增长率给所有类型同时增加 c 会产生什么影响。

解答. 未归一化质量分别按 $e^{r_A t}$ 和 $e^{r_B t}$ 增长，故

$$q(t) = \frac{q_0 e^{r_A t}}{q_0 e^{r_A t} + (1-q_0)e^{r_B t}} = \frac{1}{1 + \frac{1-q_0}{q_0}e^{-(r_A - r_B)t}}.$$

它满足 $q' = (r_A - r_B)q(1-q)$。共同增加 c 只给总质量乘 e^{ct}，不改变归一化比例。 □

习题 12 (一阶矩与随机比例不是同一个量)**.** 一个随机种群以概率 $1/2$ 包含 1 个 A，以概率 $1/2$ 包含 9 个 B。分别计算归一化期望强度中 A 的比例与一次随机种群中均匀抽样得到 A 的概率。

解答. 前者是 $\mathbb{E}N_A / \mathbb{E}N = (1/2)/5 = 1/10$。后者是 $\mathbb{E}(N_A/N) = \frac{1}{2}(1) + \frac{1}{2}(0) = 1/2$。若先按种群大小偏置选择实现，再在其中均匀选择粒子，才得到前一种分布。差异来自分母本身是随机的。 □

习题 13 (选择对均值的作用)**.** 设 $\partial_t p = L^* p + (r - \bar{r})p$。推导 $\mathrm{d}\mathbb{E}_p[X]/\,\mathrm{d}t$。若 $r(x) = \kappa x$ 且 $b(x) = -\theta x$，结果是什么？

解答. 取观测 $\varphi(x)=x$，在矩可积和边界项消失条件下，

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_p[X]=\mathbb{E}_p[b(X)]+\mathrm{Cov}_p(X,r(X)).$$

所以本题为 $m'=-\theta m+\kappa\,\mathrm{Var}_p(X)$。这说明漂移对均值的贡献与选择对均值的贡献是两个可区分的项。由于 r 无界，此公式要求相应分枝一阶矩与加权矩在所考察时间内有限。 □

习题 14 (非线性漂移与矩闭合)**.** 对 $\mathrm{d}X_t=-X_t^3\,\mathrm{d}t+\sigma\,\mathrm{d}W_t$，写出一阶、二阶矩的演化。能否只求解两个常微分方程就得到这两个矩?

解答. 在所需矩有限时，

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}X_t=-\mathbb{E}X_t^3,\qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}X_t^2=-2\mathbb{E}X_t^4+\sigma^2.$$

右侧涉及更高阶矩，系统一般不闭合。以 $(\mathbb{E}X)^3$ 替代 $\mathbb{E}X^3$ 是额外近似，并非生成元给出的恒等式。 □

习题 15 (数值稳定并不等于稳态精确)**.** 对 OU 的 Euler–Maruyama 模拟，若 $\theta h=1$，离散平稳方差是连续平稳方差的多少倍? 若 $\theta h=2$ 会怎样?

解答. 方差比为

$$\frac{\sigma^2/(2\theta-\theta^2h)}{\sigma^2/(2\theta)}=\frac{1}{1-\theta h/2}.$$

$\theta h=1$ 时是 2 倍；$\theta h=2$ 时递推系数的平方等于 1，每一步都增加 σ^2h 的方差（假设 $\sigma\neq0$），不再有有限的平稳方差。 □

习题 16 (同边缘分布不确定路径规律，选做)**.** 在选读高斯例中，取 $0<s<t$，比较布朗扩散与概率流 ODE 的 $\mathrm{Cov}(X_s,X_t)$。

解答. 布朗扩散 $X_t=X_0+W_t$ 给出 $\mathrm{Cov}(X_s,X_t)=v_0+s$。ODE 给出 $Y_t=Y_0\sqrt{(v_0+t)/v_0}$，因此

$$\mathrm{Cov}(Y_s,Y_t)=\sqrt{(v_0+s)(v_0+t)}.$$

两个模型的单时刻方差都等于 v_0+t，但跨时刻协方差不同。所以只知道每个时刻的概率密度，不足以确定整条随机路径的联合规律。 □

A 核心公式速查

以下公式都沿用正文的正则性、可积性和边界条件。

$$\Delta W\sim\mathcal{N}(0,\Delta t),\qquad [W]_t=t,$$

$$\mathbb{E}\left(\int_0^T H_s\,\mathrm{d}W_s\right)^2=\mathbb{E}\int_0^T H_s^2\,\mathrm{d}s,\qquad \int_0^T W_s\,\mathrm{d}W_s=\tfrac12(W_T^2-T).$$

对 $\mathrm{d}X_t=b\,\mathrm{d}t+\sigma\,\mathrm{d}W_t$，

$$\boxed{\mathrm{d}f(X_t,t)=\left(f_t+bf_x+\tfrac12\sigma^2f_{xx}\right)\,\mathrm{d}t+\sigma f_x\,\mathrm{d}W_t.}$$

$$\mathrm{d}(XY) = X\,\mathrm{d}Y + Y\,\mathrm{d}X + \mathrm{d}[X,Y], \qquad \int Y \circ \mathrm{d}W = \int Y\,\mathrm{d}W + \tfrac{1}{2}[Y,W].$$

$$\boxed{b_{\mathrm{Ito}} = b_{\mathrm{S}} + \tfrac{1}{2}\sigma\partial_x\sigma.}$$

$$\begin{aligned} L_t\varphi &= b\varphi_x + \tfrac{1}{2}\sigma^2\varphi_{xx}, \\ L_t^* p &= -\partial_x(bp) + \tfrac{1}{2}\partial_{xx}(\sigma^2 p), \\ J &= bp - \tfrac{1}{2}\partial_x(\sigma^2 p), \qquad \partial_t p = -\partial_x J, \\ \partial_t v + L_t v = 0, &\quad v(\cdot,T) = \varphi, \\ \partial_t p = L_t^* p, &\quad p(\cdot,0) = p_0. \end{aligned}$$

多维时 $a = \Sigma\Sigma^{\mathsf{T}}$，

$$L_t\varphi = \sum_i b_i\partial_i\varphi + \tfrac{1}{2}\sum_{i,j} a_{ij}\partial_{ij}\varphi, \qquad L_t^* p = -\sum_i \partial_i(b_i p) + \tfrac{1}{2}\sum_{i,j}\partial_{ij}(a_{ij}p).$$

对守恒运动上的分枝系统，

$$\boxed{\partial_t f = L_t^* f + rf, \quad p = \frac{f}{M}, \quad M = \int f, \quad \partial_t p = L_t^* p + (r - \bar{r})p.}$$

$$\bar{r} = \int rp, \qquad M' = M\bar{r}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_p\varphi = \mathbb{E}_p L_t\varphi + \mathrm{Cov}_p(\varphi, r).$$

$\dfrac{\mathbb{E}Z_t}{\mathbb{E}N_t}$ 通常不等于 $\mathbb{E}\left[\dfrac{Z_t}{N_t}\middle| N_t > 0\right]$.

B 常见错误与自查顺序

1. **先检查时间与积分约定**。 t 是正向时间、剩余时长，还是递增的反向时钟？噪声项是 $\mathrm{d}W$ 还是 $\circ\mathrm{d}W$？
2. **再检查系数**。 SDE 中的噪声振幅是 σ；瞬时方差率是 σ^2；热算子的系数是 $\sigma^2/2$。
3. **做变量变换时检查二阶项**。对数和平方是最容易暴露遗漏修正项的两个测试函数。
4. **取期望前检查鞅条件**。零漂移首先意味着局部鞅；平方可积或其他充分条件才保证所需的真鞅性质。
5. **从 SDE 写 PDE 时检查导数位置**。 L 作用于观测函数，L^* 作用于密度。变量系数必须留在 FP 的乘积导数中。
6. **积分 PDE 时检查边界**。质量守恒由总通量决定。杀死边界可以让域内质量减少；保留吸收边界粒子的模型还可能出现边界原子。
7. **解释分枝方程时检查对象**。 f 是期望强度，M 是期望总数；二者之比不自动等于随机经验概率测度的期望。
8. **解释数值结果时检查误差来源**。步长误差、样本误差、分箱误差与模型误差不是同一件事。

C 进一步阅读

正文主要围绕用户提供的草稿作独立重组与推导。以下公开的一手课程资料与原始论文供进一步查阅：

参考文献

[1] Benjamin Fehrman. *Stochastic Differential Equations*. University of Oxford lecture notes, 2019/20. 公开讲义。可继续阅读连续时间鞅、随机积分、Itô 公式的完整证明和 SDE 适定性。

[2] Maria Bruna and Tomislav Plesa. *Stochastic Processes in Biology*, Part III, Lent Term 2024. 公开讲义。包含扩散方程、数值积分、平稳过程和边界条件。

[3] Andreas Eberle. *Stochastic Analysis: An Introduction*, Winter Term 2010/11. 公开讲义。可对照阅读 Feynman–Kac 表示、显式 SDE 与 OU 过程。

[4] Pierre Del Moral. *Advanced Monte Carlo Methods*, FGV lecture notes, 2018. 公开讲义。用于进一步了解 Feynman–Kac 模型、归一化与粒子近似。

[5] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon and Ben Poole. *Score-Based Generative Modeling through Stochastic Differential Equations*. ICLR 2021. arXiv:2011.13456。选读部分关于反向 SDE 和 probability flow ODE 的原始研究来源。