随机游动与鞅

应坚刚
复旦大学数学系

2020 年 12 月 16 日

序　言

最近二十年来, 随机过程由于其在数理金融等领域的应用而倍受关注, 特别是Merton 和Scholes 在期权定价方面的工作获得诺贝尔奖之后更是如此. 因此我们觉得有必要对学生介绍随机过程的一些基本知识.

这个简短的讲义是为数学系本科高年级学生开设的随机过程选修课而写的, 目的在于让他们对随机过程的经典问题和方法有一个初步的了解. 讲义主要介绍随机游动, 鞅以及Markov 链的基本理论, 作为最简单的随机过程, 它们的研究有悠久的历史和直观自然的背景. 实际上, 许多连续时间的随机过程的结论都可以在这些随机序列的经典结果中发现它们的影子.

第一章讲述随机游动, 也就是独立同分布随机变量的和, 我们将介绍一些经典的方法和问题. 在第二章中我们将介绍鞅理论, 某种意义上说, 鞅是随机游动的一个自然推广, 比起随机游动, 鞅是较为现代的理论, 它的主要发展是在二十世纪下半叶, 鞅在许多领域都有重要应用, 这里我们简单介绍鞅在金融理论中的应用, 也就是简二项期权定价, 虽然简单, 但足以表达Black-Schloes 及Merton 理论的基本思想. 第三章比较抽象, 它讲述随机过程的基本理论和存在性定理, 使得本书建立在一个坚实的基石之上. 而且从逻辑角度考虑, 它是后面讨论Markov 链所必需的, 但是实际上给定转移函数的Markov 链的存在性在直观上是自然的, 因此即使读者不能很好地理解这一章, 也不会影响他们对其它内容的理解. 第四章介绍Markov 链的基本理论和经典方法, 它的研究一直是概率论中最为活跃的领域. 我们还讲简要介绍虽然早已有所论述但最近才被人关注的Markov 链与电路网络的本质联系.

本讲义所涉及的大部分内容可以说是简单直观, 解决其中的问题也不需要太多的数学工具, 对稍有基础的读者来说不难理解. 阅读本讲义所需的概率论知识, 可参考[9]. 对随机过程理论感兴趣并且想进一步阅读的读者可以参考作者的研究生随机过程教材[10], Billingsley[1], 或者更全面的Kallenberg[6].

目　录

序　言 **i**

第一章　预备知识 **1**

1.1 基本框架与单调类定理 1
1.2 一致可积与控制收敛定理 4
1.3 条件期望 6

第二章　随机游动 **10**

2.1 简单对称随机游动 10
2.2 随机游动与Markov 性 17

第三章　经典鞅论 **28**

3.1 鞅序列及其应用 28
3.2 二项期权定价理论 40

第四章　随机过程的构造 **49**

4.1 测度扩张定理 49
4.2 Kolmogorov 相容定理 55

第五章　Markov 链 **66**

5.1 Markov 链 66
5.2 Galton-Watson 分支过程 86
5.3 可逆Markov 链与电路网络 94

第六章　Levy 过程 **104**

6.1 平稳独立增量过程 104
6.2 Lévy-Khinchin 表示 106
6.3 Poisson 过程 111

第七章　Brown 运动 **114**

7.1 Brown 运动的构造 114

7.2 反射原理 116
7.3 二次变差 119
7.4 鞅性质 122
7.5 停时 124

参考文献 **131**

第一章　预备知识

§1.1　基本框架与单调类定理

概率论的基本框架是建立在测度的基础上的, 这一点虽然不是所有人都同意, 但在数学角度看基本上没有异议. 尽管有这样那样的其它更一般的框架存在, 比如非线性的概率论, 但主流的概率论显然需要测度论的基础.

设$(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间, 也就是说, Ω 是非空集合, 概率论中被称为样本空间, $\mathcal{F}$ 是Ω 上的σ-代数, $\mathbb{P}$ 是可测空间$(\Omega, \mathcal{F})$ 上的概率测度. 设ξ 是一个随机变量, 对$B \in \mathcal{B}(\mathbf{R})$, 定义

$$\mu(B) = \mathbb{P}(\xi \in B). \tag{1.1.1}$$

那么μ 是$(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上的概率测度, 也就是说$(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mu)$ 是概率空间, μ 被称为ξ 的分布, 对应的函数

$$F(x) = \mu((-\infty, x]) = \mathbb{P}(\xi \leqslant x),\ x \in \mathbf{R} \tag{1.1.2}$$

被称为ξ 的分布函数. 分布函数与分布是由(1.1.2) 相互唯一决定的.

下面首先要介绍的Dynkin 引理和单调类方法是概率论中常用的一个系统方法, 它省去了证明中所需的许多相似的推理. 单调类方法由Dynkin 首先写在他的书[**?**] 上, 所以也称为Dynkin 引理.

设$\mathcal{F}_0$ 是Ω 的某些子集组成的集合, 也称为集类, 那么用$\sigma(\mathcal{F}_0)$ 表示$\mathcal{F}_0$ 生成的σ-代数, 它等同于Ω 上包含$\mathcal{F}_0$ 的最小σ-代数. 称一个子集类是π-类, 如果它对有限交封闭. 例如Euclid 空间上开集全体或闭集全体都是π-类. 而称一个子集类是Dynkin 类或λ-类, 如果它包含有$\varnothing, \Omega$ 且对于补集运算与不交可列并运算封闭. 显然, 环是π-类, σ-代数是Dynkin 类, 反之不对. 容易看出任意多个Dynkin 类的交仍是Dynkin 类, 因此对Ω 的任何子集类$\mathcal{A}$, 唯一存在一个包含$\mathcal{A}$ 的最小Dynkin 类, 记为$\delta(\mathcal{A})$, 也类似地称为由$\mathcal{A}$ 生成的Dynkin 类.

Dynkin 类与σ-代数的差别就在于后者要求对任意可列并封闭, 前者只要求对不交可列并封闭, 在后面我们会看到, 在许多涉及概率测度的场合我们都无法直接证明一个集类是σ-代数, 而很容易证明它是Dynkin 类, 而Dynkin 引理就是告诉我们一个保证Dynkin 类是σ-代数的一个简单的充分条件.

定理1.1.1 (Dynkin) 设$\mathcal{F}_0$ 是一个π-类, 则$\delta(\mathcal{F}_0)$ 是一个σ-代数, 因此$\sigma(\mathcal{F}_0)=\delta(\mathcal{F}_0)$.

证明. 由定义, 仅须验证$\delta(\mathcal{F}_0)$ 对有限交运算封闭. 任取$A\in\delta(\mathcal{F}_0)$, 定义

$$\kappa[A]:=\{B\in\delta(\mathcal{F}_0):\ A\cap B\in\delta(\mathcal{F}_0)\}.$$

先验证$\kappa[A]$ 是一个Dynkin 类. 事实上, 只需证明$\kappa[A]$ (1) 对补集运算封闭; (2) 对不相交集列的可列并运算封闭.

对(1), 取$B\in\kappa[A]$, 则$A, A^c, A\cap B\in\delta(\mathcal{F}_0)$, 因此$A\cap B^c=[A^c\cup(A\cap B)]^c\in\delta(\mathcal{F}_0)$, 因此$B^c\in\kappa[A]$.

为证明(2), 取$\{B_n\}\subset\kappa[A]$ 是不交集列, 则显然$\{A\cap B_n\}$ 是$\delta(\mathcal{F}_0)$ 中不交集列, 因此$A\bigcap(\bigcup_n B_n)\in\delta(\mathcal{F}_0)$, 推出$\bigcup_n B_n\in\kappa[A]$.

因$\mathcal{F}_0$ 是π-类, 故$A\in\mathcal{F}_0$ 蕴含着$\kappa[A]\supset\mathcal{F}_0$ 即$\kappa[A]\supset\delta(\mathcal{F}_0)$. 这意味着当$A\in\delta(\mathcal{F}_0)$ 时, $\kappa[A]\supset\mathcal{F}_0$. 因此$\kappa[A]\supset\delta(\mathcal{F}_0)$, 即$\delta(\mathcal{F}_0)$ 中元素对有限交运算封闭. □

单调类定理是Dynkin 引理的一种等价叙述, 但看起来要更强大, 因为它似乎跨越更大. 它更多的是用于怎么证明对集合的概率测度成立的结论对于随机变量的数学期望也成立. 有了单调类定理, 这只不过是一句话的事情.

定理1.1.2 设$\mathcal{L}$ 是$(\Omega,\mathcal{F})$ 的上一些可测函数组成的线性空间, $\mathcal{F}_0$ 是一个π-类, 满足

(1) 对任何$A\in\mathcal{F}_0$ 有$1_A\in\mathcal{L}$;

(2) $1\in\mathcal{L}$;

(3) 如果$\{f_n\}$ 是$\mathcal{L}$ 中一致有界的非负函数列且关于n 单调, 则有$\sup_n f_n\in\mathcal{L}$.

那么$\mathcal{L}$ 包含$\sigma(\mathcal{F}_0)$-可测的有界函数全体.

Proof. 首先证明由Dynkin 引理可得$\mathcal{L}$ 包含$\sigma(\mathcal{F}_0)$ 中任意集合的示性函数. 事实上, 令

$$\mathcal{G}=\{A\subset\Omega:1_A\in\mathcal{L}\},$$

条件(1) 蕴含着$\mathcal{G}\supset\mathcal{F}_0$. 条件(2) 说明$\Omega\in\mathcal{G}$, $\mathcal{L}$ 是个线性空间的假设推出如果$A\in\mathcal{G}$, 那么$1_{A^c}=1-1_A\in\mathcal{L}$ 或者$A^c\in\mathcal{G}$, 即$\mathcal{G}$ 对补集运算封闭. 条件(3) 推出$\mathcal{G}$ 对不交可列并运算封闭. 因此$\mathcal{G}$ 是一个Dynkin 类, 那么

$$\mathcal{G}\supset\delta(\mathcal{F}_0).$$

再由Dynkin 定理, $\delta(\mathcal{F}_0)=\sigma(\mathcal{F}_0)$, 推出$\mathcal{G}\supset\sigma(\mathcal{F}_0)$.

现在, 因为任何一个$\sigma(\mathcal{F}_0)$ 的非负有界可测函数都可以表示为$\sigma(\mathcal{F}_0)$ 上递增简单函数列的极限, 所以由$\mathcal{L}$ 的性质(3) 推出$\mathcal{L}$ 包含非负有界可测函数, 再由$\mathcal{L}$ 是线性空间推出定理结论成立. □

作为例子, 我们来看变量替换公式. 设$\mathbb{P}$ 是$(\Omega,\mathcal{F})$ 上的概率测度, ξ 是随机变量, ξ 的分布是$(\mathbf{R},\mathcal{B}(\mathbf{R}))$ 上满足对任何$a,b\in\mathbf{R}$, 有$\mathbb{P}(\xi\in(a,b))=\mu(a,b)$ 的唯一概率测度. 注意到Borel σ-代数是由区间全体生成的, 而区间全体是一个π-类, 那么由单调类定理立刻可得对$\mathbf{R}$ 上任何有界Borel 可测函数f 有

$$\mathbb{E}[f(\xi)]=\int f(x)\mu(dx).$$

另外一个需要用到Dynkin 引理的是独立的概念. 独立性是概率论中一个非常直观且有用的概念, 它是概率论有别于测度论的第一个标签. 独立的概念最初是从两个事件开始的, 事件$A,B\in\mathcal{F}$ 被称为相互独立的, 如果

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B).$$

独立的概念显然是独立的随机试验这个概念的推广, 掷一个硬币与掷一个骰子这两件事情是独立的, 它们的结果不会相互影响. n 个事件$A_1,\cdots,A_n\in\mathcal{F}$ 相互独立是指其中任意个事件同时发生的概率等于各自发生的概率的乘积, 即对任何$k\geqslant 1$ 及$1\leqslant n_1<\cdots<n_k\leqslant n$ 有

$$\mathbb{P}(A_{n_1}\cap\cdots\cap A_{n_k})=\mathbb{P}(A_{n_1})\cdots\mathbb{P}(A_{n_k}).$$

另外, 随机变量$\xi_1,\cdots,\xi_n$ 独立是指对任何$x_1,\cdots,x_n\in\mathbf{R}$ 有

$$\mathbb{P}\left(\bigcap_{k=1}^{n}\{\xi_k\leqslant x_k\}\right)=\prod_{k=1}^{n}\mathbb{P}(\xi_k\leqslant x_k).$$

为了把这些概念统一起来, 我们引入事件类的独立性概念.

设$\mathcal{A}_1,\cdots,\mathcal{A}_n\subset\mathcal{F}$ 是事件类, 那么$\mathcal{A}_1,\cdots,\mathcal{A}_n$ 相互独立是指从每个事件类中取出一个事件$A_k\in\mathcal{A}_k$, $1\leqslant k\leqslant n$ 有$A_1,\cdots,A_n$ 相互独立. 自然地, 类似于线性代数中的线性无关性, 无限多个事件类的相互独立性是指其中任意有限个是相互独立的. 那么随机变量的独立性可以用事件类的独立性表达.

引理1.1.1 随机变量$\xi_1,\cdots,\xi_n$ 独立当且仅当事件类$\sigma(\xi_1),\cdots,\sigma(\xi_n)$ 相互独立.

Proof. 设$n=2$, 我们只需证明对任何Borel 集$B_1,B_2\in\mathcal{B}(\mathbf{R})$ 有

$$\mathbb{P}(\xi_1\in B_1,\xi_2\in B_2)=\mathbb{P}(\xi_1\in B_1)\mathbb{P}(\xi_2\in B_2).\tag{1.1.3}$$

应用Dynkin 引理即可. 由定义ξ_1, ξ_2 独立是指对任何x, y 有

$$\mathbb{P}(\xi_1 \leqslant x, \xi_2 \leqslant y) = \mathbb{P}(\xi_1 \leqslant x)\mathbb{P}(\xi_2 \leqslant y). \tag{1.1.4}$$

用$\mathcal{G}$ 表示$\mathbf{R}$ 中满足

$$\mathbb{P}(\xi_1 \in B_1, \xi_2 \leqslant y) = \mathbb{P}(\xi_1 \in B_1)\mathbb{P}(\xi_2 \leqslant y). \tag{1.1.5}$$

的子集B_1 的全体, 那么$\mathcal{G}$ 包含区间全体且是一个Dynkin 类, 因此$\mathcal{G}$ 包含由区间全体生成的Dynkin 类$\mathcal{B}$. 由Dynkin 定理知区间全体作为一个π-类生成的Dynkin 类是一个σ-代数, 就是Borel 子集的全体. 因此(1.1.5) 对于任何Borel 集B_1 成立, 固定任何的B_1, 同理可以证明(1.1.3)对任何Borel 集B_2 成立. □

上面对于随机向量也是成立的. 也就是说, 谈论随机变量的集合之间的独立性时实际上就是谈论它们生成的σ-代数之间的独立性. 这样独立性就有了统一的语言.

§1.2 一致可积与控制收敛定理

如果ξ 可积, 那么$\lim_n \mathbb{E}(|\xi|; |\xi| \geqslant n) = 0$. 如果考虑一族随机变量, 那么这个极限一致收敛就是一致可积的概念.

定义1.2.1 一个可积随机变量族$\{\xi_i : i \in I\}$ 称为是一致可积的, 若

$$\lim_{N\to\infty} \sup_{I} \mathbb{E}(|\xi_i|; |\xi_i| \geqslant N) = 0.$$

不难验证, 若$\{\xi_i : i \in I\}$ 被一个可积随机变量所控制, 则$\{\xi_i\}$ 是一致可积的. 下面定理给出一致可积的一个等价条件, 先证明一个引理, 说明可积随机变量是绝对连续的.

引理1.2.1 如果随机变量ξ 可积, 则对任何$\epsilon > 0$, 存在$\delta > 0$, 使得对任何$A \in \mathcal{F}$, $\mathbb{P}(A) < \delta$, 有$\mathbb{E}(|\xi|; A) < \epsilon$.

Proof. 因为

$$\begin{aligned}\mathbb{E}(|\xi|; A) &= \mathbb{E}(|\xi|; A, |\xi| \geqslant N) + \mathbb{E}(|\xi|; A, |\xi| < N)\\ &\leqslant \mathbb{E}(|\xi|; |\xi| \geqslant N) + N \cdot \mathbb{P}(A),\end{aligned}$$

故对任何$\epsilon > 0$, 存在N 使得$\mathbb{E}(|\xi|; |\xi| \geqslant N) < \frac{1}{2}\epsilon$, 再取$0 < \delta < \frac{1}{2N}\epsilon$, 因此只要$\mathbb{P}(A) < \delta$, 就有$\mathbb{E}(|\xi|; A) < \epsilon$. □

定理1.2.1 设$\{\xi_i : i \in I\}$ 是随机变量族. 则它是一致可积的充要条件是

(1) L^1-有界: $\sup_{i\in I} \mathbb{E}|\xi_i| < \infty$.

(2) 一致绝对连续: 对任何$\epsilon > 0$, 存在$\delta > 0$ 使得当$A \in \mathcal{F}$, $\mathbb{P}(A) < \delta$ 时, $\sup_{i\in I} \mathbb{E}(|\xi_i|; A) < \epsilon$.

Proof. 必要性. 对任意$A \in \mathcal{F}$, $N > 0$, 有

$$\begin{aligned}\mathbb{E}(|\xi_i|; A) &= \mathbb{E}(|\xi_i|; A \cap \{|\xi_i| \geqslant N\}) + \mathbb{E}(|\xi_i|; A \cap \{|\xi_i| < N\}) \\ &\leqslant \mathbb{E}(|\xi_i|; \{|\xi_i| \geqslant N\}) + N\mathbb{P}(A).\end{aligned}$$

运用一致可积性, 推出$\{\xi_i\}$ 是一致绝对连续的. 再在上式令$A = \Omega$, 得

$$\mathbb{E}|\xi_i| \leqslant \mathbb{E}(|\xi_i|; \{|\xi_i| \geqslant N\}) + N,$$

得到$\{\xi_i\}$ 的L^1-有界性.

充分性. 设$\{\xi_i\}$ 是一致绝对连续且L^1-有界的. 由Chebyshev 不等式,

$$\sup_i \mathbb{P}(|\xi_i| \geqslant N) \leqslant \frac{1}{N} \sup_i \mathbb{E}|\xi_i|.$$

因此对任何$\epsilon > 0$, 存在N 使得对任何$i \in I$, $\mathbb{P}(|\xi_i| \geqslant N) < \delta$. 然后由$\{\xi_i\}$ 的一致绝对连续性得到, 对任何$\epsilon > 0$, 存在$N > 0$, 使得$\mathbb{E}(|\xi_i|; \{|\xi_i| \geqslant N\}) < \epsilon$ 对所有$i \in I$ 成立, 即一致可积性. □

定理1.2.2 可积随机变量序列$\{\xi_n\}$ L^1-收敛于ξ 的充要条件是$\{\xi_n\}$ 是一致可积的且$\xi_n \xrightarrow{p} \xi$.

Proof. 必要性. 首先$\xi_n \xrightarrow{p} \xi$ 是显然的, $\{\xi_n\}$ 的L^1-有界性也是显然的. 只需验证$\{\xi_n\}$ 的一致绝对连续性就够了. 对任意$A \in \mathcal{F}$,

$$\mathbb{E}(|\xi_n|; A) \leqslant \mathbb{E}(|\xi|; A) + \mathbb{E}(|\xi_n - \xi|).$$

对任何$\epsilon > 0$, 存在N, 当$n > N$ 时, $\mathbb{E}(|\xi_n - \xi|) < \frac{1}{2}\epsilon$. 另外因$\xi, \xi_1, \cdots, \xi_N$ 可积, 故存在$\delta > 0$, 只要$\mathbb{P}(A) < \delta$ 便有$\mathbb{E}(|\xi|; A)$, 及所有$\mathbb{E}(|\xi_i|; A)$, $1 \leqslant i \leqslant N$, 小于$\frac{1}{2}\epsilon$. 因此只要$\mathbb{P}(A) < \delta$, 就有$\sup_n \mathbb{E}(|\xi_n|; A) < \epsilon$.

充分性. 首先要证明ξ 是可积的, 因为依概率收敛, 故存在子列$\{\xi_{k_n}\}$ 几乎处处收敛于ξ. 由Fatou 引理及L^1-有界性, $\mathbb{E}|\xi| \leqslant \liminf_n \mathbb{E}|\xi_{k_n}| < \infty$. 因此$\xi$ 可积. 对

任意$\epsilon > 0$, 由一致绝对连续性, 存在$\delta > 0$ 只要$\mathbb{P}(A) < \delta$ 就有$\mathbb{E}(|\xi_n|; A) < \epsilon$ 对任何n 成立且$\mathbb{E}(|\xi|; A) < \epsilon$ 成立. 再由于$\xi_n \xrightarrow{P} \xi$, 存在$N$ 当$n > N$ 时就有$\mathbb{P}(|\xi_n - \xi| > \epsilon) < \delta$. 因此

$$\begin{aligned}\mathbb{E}|\xi_n - \xi| &\leqslant \mathbb{E}(|\xi_n - \xi|; \{|\xi_n - \xi| \leqslant \epsilon\}) + \mathbb{E}(|\xi_n - \xi|; \{|\xi_n - \xi| > \epsilon\}) \\ &\leqslant \epsilon + \mathbb{E}(|\xi_n|; \{|\xi_n - \xi| > \epsilon\}) + \mathbb{E}(|\xi|; \{|\xi_n - \xi| > \epsilon\}) < 3\epsilon.\end{aligned}$$

证明了$\xi_n \xrightarrow{L^1} \xi$. □

如果ξ_n 以L^1 收敛于ξ, 那么$\lim_n \mathbb{E}\xi_n = \mathbb{E}\xi$, 即积分与极限可以交换. 关于积分极限交换, 我们通常使用Lebesgue 控制收敛定理. 由于随机序列$\{\xi_n\}$ 被一个可积随机变量η 控制时, $\{\xi_n\}$ 是一致可积的. 实际上, 对任何$N > 0$ 有

$$\mathbb{E}[|\xi_n|; |\xi_n| > N] \leqslant \mathbb{E}(\eta; \eta > N).$$

因此Lebesgue 控制收敛定理是上述定理的一个特殊情况.

§1.3 条件期望

现代概率论(特别是鞅与Markov 链)显然离不开条件期望这个重要概念, 条件期望是从我们熟知的条件概率, 全概率公式抽象和发展起来的. 设$(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, $A, B \in \mathcal{F}$, 定义

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

称为是B 发生的条件下B 发生的概率或者简单地, A 关于B 的条件概率. 此定义中当然要假设$\mathbb{P}(B) > 0$. 概率论中最重要的公式之一是全概率公式: 设$\{\Omega_n : n \geqslant 1\}$ 是Ω 的可测分类, 则对任何$A \in \mathcal{F}$, 有

$$\mathbb{P}(A) = \sum_{n \geqslant 1} \mathbb{P}(A|\Omega_n)\mathbb{P}(\Omega_n).$$

因为事件A 可以看成为随机变量1_A, 而$\mathbb{P}(A) = \mathbb{E}(1_A)$. 因此我们可以把全概率公式推广到随机变量情况

$$\mathbb{E}\xi = \sum_{n \geqslant 1} \mathbb{E}(\xi|\Omega_n)\mathbb{P}(\Omega_n), \tag{1.3.1}$$

其中ξ 是一个可积随机变量, 而对于任何事件B, 类似地定义B 上的平均

$$\mathbb{E}(\xi|B) := \frac{\mathbb{E}(\xi; B)}{\mathbb{P}(B)}.$$

注意$\mathbb{E}(\xi;B):=\mathbb{E}(\xi\cdot 1_B)$.

仔细观察(1.3.1), 如果定义一个新的随机变量

$$\xi':=\sum_n \mathbb{E}(\xi|\Omega_n)1_{\Omega_n},$$

那么全概率公式可以写成为$\mathbb{E}\xi=\mathbb{E}\xi'$. 而且$\xi'$ 是在某种意义上最接近ξ 的随机变量, 实际上, 任意给定一个序列$\{x_n\}$, 定义$\eta:=\sum_n x_n 1_{\Omega_n}$, 那么

$$\mathbb{E}(\xi-\eta)^2=\mathbb{E}(\xi-\xi')^2+\mathbb{E}(\xi'-\eta)^2,$$

也就是说, ξ' 在L^2 距离的意义下在所有类似η 这样的随机变量中离ξ 最近, 而且ξ' 在这个最近的意义下是(几乎处处相等的意义下)唯一决定的. 证明上面等式的关键是

$$\mathbb{E}[(\xi-\xi')(\xi'-\eta)]=0.$$

而此式成立的原因是两点(1) ξ' 在Ω_n 上为常数; (2) 对任何n,

$$\mathbb{E}(\xi';\Omega_n)=\mathbb{E}(\xi;\Omega_n).$$

条件期望的概念正是从这里抽象出来的, 考虑一个可积随机变量ξ 与$\mathcal{F}$ 的一个子σ-域$\mathcal{A}$, 我们来定义在某种意义上最接近ξ 的$\mathcal{A}$ 可测的随机变量, 称为ξ 关于$\mathcal{A}$ 的条件期望.

定义1.3.1 设ξ 是可积的随机变量, $\mathcal{A}$ 是$\mathcal{F}$ 的子σ-域, 那么一个关于$\mathcal{A}$ 可测且满足对任何$A\in\mathcal{A}$, $\mathbb{E}(\eta;A)=\mathbb{E}(\xi;A)$ 的随机变量η 称为是ξ 关于$\mathcal{A}$ 的条件期望, 记为$\mathbb{E}(\xi|\mathcal{A})$. 如果$\eta$ 是一个随机变量, 那么$\mathbb{E}(\xi|\eta):=\mathbb{E}(\xi|\sigma(\eta))$.

定义中关于$\mathcal{A}$ 可测且满足对任何$A\in\mathcal{A}$, $\mathbb{E}(\eta;A)=\mathbb{E}(\xi;A)$ 的随机变量η 是几乎处处唯一决定的, 也就是说任何两个满足条件的随机变量是几乎处处相等的, 上面所说的唯一性来自下面随机变量唯一性, 如果η_1, η_2 都是ξ 关于$\mathcal{A}$ 的条件期望, 那么他们都是关于$\mathcal{A}$ 可测的随机变量且对任何$A\in\mathcal{A}$ 有$\mathbb{E}(\eta_1;A)=\mathbb{E}(\eta_2;A)$, 因此$\eta_1=\eta_2$ a.s.

练习1.3.1 证明: 如果η_1,η_2 都是关于$\mathcal{A}$ 可测的随机变量且满足对任何$A\in\mathcal{A}$ 有$\mathbb{E}(\eta_1;A)=\mathbb{E}(\eta_2;A)$, 那么$\eta_1=\eta_2$ a.s.

由唯一性, 要证明$\eta=\mathbb{E}(\xi|\mathcal{A})$. 只需证明(1) η 是$\mathcal{A}$ 可测的; (2) 对任何$A\in\mathcal{A}$ 有$\mathbb{E}(\eta;A)=\mathbb{E}(\xi;A)$.

由此可见条件期望$\mathbb{E}(\xi|\mathcal{A})$ 是几乎处处相等的意义下唯一决定的, 因此下面关于条件期望的等式或不等式都是在几乎处处的意义下叙述并证明的. 例如由定义容易验证, 如果ξ 是$\mathcal{A}$ 可测可积随机变量, 则$\mathbb{E}(\xi|\mathcal{A}) = \xi$ a.s. 我们简单地写成$\mathbb{E}(\xi|\mathcal{A}) = \xi$.

条件期望的存在性的证明要更难些, 它是测度论中的Radon-Nikodym 定理的推论. 随机变量关于一般σ-域的条件期望有下面的性质, 这些性质完全是由定义推导的.

定理1.3.1 (1) $\xi \mapsto \mathbb{E}(\xi|\mathcal{A})$ 作为关于σ-域$\mathcal{F}$ 可测的可积随机变量空间到关于σ-域$\mathcal{A}$ 可测的可积随机变量空间的映射是线性的, 保序的;

(2) 全概率公式的推广: $\mathbb{E}(\mathbb{E}(\xi|\mathcal{A})) = \mathbb{E}\xi$;

(3) 如果ξ, η 是随机变量, ξ 与$\xi\eta$ 可积且η 是关于σ-域$\mathcal{A}$ 的随机变量, 那么$\mathbb{E}(\xi\eta|\mathcal{A}) = \eta\mathbb{E}(\xi|\mathcal{A})$;

(4) 如果ξ 与$\mathcal{A}$ 独立, 那么$\mathbb{E}(\xi|\mathcal{A}) = \mathbb{E}\xi$. ξ 与$\mathcal{A}$ 独立是指对任何$A \in \mathcal{A}$, ξ 与1_A 独立;

(5) 设$\mathcal{A}_1, \mathcal{A}_2$ 是子σ-域, 且$\mathcal{A}_1 \subset \mathcal{A}_2$. 那么

$$\mathbb{E}(\mathbb{E}(\xi|\mathcal{A}_1)|\mathcal{A}_2) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{A}_2)|\mathcal{A}_1) = \mathbb{E}(\xi|\mathcal{A}_1);$$

(6) $\mathbb{E}(\xi|\{\emptyset, \Omega\}) = \mathbb{E}\xi$, $\mathbb{E}(\xi|\mathcal{F}) = \xi$.

证明. (1) 留作习题.

(2) 按照条件期望的定义

$$\mathbb{E}(\mathbb{E}(\xi|\mathcal{A})) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{A}); \Omega) = \mathbb{E}(\xi; \Omega) = \mathbb{E}(\xi).$$

(3) 只需对$\eta = 1_B$, $B \in \mathcal{A}$, 证明就可以了. 而这由定义是显然的.

(4) 对任何$A \in \mathcal{A}$, $\mathbb{E}(\xi; A) = \mathbb{E}(\xi 1_A) = \mathbb{E}\xi \cdot \mathbb{P}(A) = \mathbb{E}(\mathbb{E}\xi; A)$.

(5) 因为$\mathbb{E}(\xi|\mathcal{A}_1)$ 是$\mathcal{A}_2$ 可测的, 故$\mathbb{E}(\mathbb{E}(\xi|\mathcal{A}_1)|\mathcal{A}_2) = \mathbb{E}(\xi|\mathcal{A}_1)$ 是显然的. 要证明另一个等式, 两者都是$\mathcal{A}_1$ 可测的, 只需证明对任何$A \in \mathcal{A}_1$, $\mathbb{E}(\xi; A) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{A}_2); A)$, 而这由于$A \in \mathcal{A}_2$ 是显然的.

(6) 由定义显然. □

下面定理说明条件期望$\mathbb{E}(\xi|\mathcal{A})$ 是$\mathcal{A}$ 可测的随机变量中离ξ 最近的那个.

定理1.3.2 设ξ 是平方可积随机变量, $\mathcal{A}$ 是$\mathcal{F}$ 的子σ-域, 那么

$$\mathbb{E}(\xi-\mathbb{E}(\xi|\mathcal{A}))^2=\inf\{\mathbb{E}(\xi-\zeta)^2:\ \zeta \text{ 是关于}\mathcal{A}\text{ 可测的平方可积随机变量}\},$$

即条件期望是所有关于$\mathcal{A}$ 可测的平方可积随机变量中与ξ 的均方距离最近的.

证明. 因为ζ 与$\mathbb{E}(\xi|\mathcal{A})$ 都是关于$\mathcal{A}$ 可测的平方可积随机变量, 故由上面的讨论得

$$\begin{aligned}\mathbb{E}(\xi-\zeta)^2&=\mathbb{E}(\xi-\mathbb{E}(\xi|\mathcal{A}))^2+\mathbb{E}(\mathbb{E}(\xi|\mathcal{A})-\zeta)^2+2\mathbb{E}[(\xi-\mathbb{E}(\xi|\mathcal{A}))(\mathbb{E}(\xi|\mathcal{A})-\zeta)]\\&=\mathbb{E}(\xi-\mathbb{E}(\xi|\mathcal{A}))^2+\mathbb{E}(\mathbb{E}(\xi|\mathcal{A})-\zeta)^2,\end{aligned}$$

由此看出结论是显然的. □

从几何的角度来看, Euclid 空间上的一个点到其一个子空间的最短距离是这个点到这个空间的投影的距离. 平方可积随机变量全体是一个很好的线性空间, 关于$\mathcal{A}$ 可测的平方可积随机变量全体是它的一个线性子空间, 因此$\mathbb{E}(\xi|\mathcal{A})$ 就是ξ 到这个空间的投影. 这也给条件期望另一个直观解释. 把随机变量ξ 看成一个需要预测的随机变量, $\mathcal{A}$ 看成已知信息, 那么$\mathcal{A}$ 可测的随机变量全体是在已知$\mathcal{A}$ 的前提下对ξ 的预测全体, 而$\mathbb{E}(\xi|\mathcal{A})$ 恰好是其中离ξ 距离最近的一个预测, 或者说是最佳预测. 因此说条件期望$\mathbb{E}(\xi|\mathcal{A})$ 是已知信息$\mathcal{A}$ 的条件下对ξ 的最佳预测. 这样定理1.3.1的性质就有很好的解释: (6) 说明当没有信息, 即$\mathcal{A}=\{\emptyset,\Omega\}$ 时, 对ξ 的最佳预测就是它的期望, 而当有足够信息, 即$\mathcal{A}=\mathcal{F}$ 时, ξ 的最佳预测就是ξ; (4) 说明当$\mathcal{A}$ 独立于ξ 时, 它透露的信息对预测ξ 没有帮助, 此信息和没有信息一样. 粗略地说, σ-域$\mathcal{A}$ 作为信息是给出Ω 的分类, 即对于$\omega\in\Omega$, $A\in\mathcal{A}$, 我们能判断是否$\omega\in A$.

设$\{\xi_i:i\in I\}$ 是一个随机变量族, 它们生成的最小σ-代数是Ω 上使得它们可测的最小σ-代数, 记为$\sigma(\xi_i:i\in I)$. 可积随机变量ξ 关于族$\{\xi:i\in I\}$ 的条件期望就是它关于这个族生成的σ-代数的条件期望, 即

$$\mathbb{E}[\xi|\xi_i:i\in I]=\mathbb{E}[\xi|\sigma(\xi_i:i\in I)].$$

第二章　随机游动

随机过程直观上是随机现象中按时间顺序记录的数据, 它的研究对象通常是依序排列的无限多个随机变量. 最简单的随机过程是随机游动, 它是指独立同分布随机变量序列组成的部分和序列. 在实际中有许多例子, 如重复一个同样输赢可能性的赌博. 但它的理论研究起源于1708 年法国学者Remond de Montmort. 所以, 随机游动的研究已有很长的历史.

在这一章中, 我们将用不同的方法研究随机游动. 首先是用经典的组合方法来讨论对称简单随机游动, 整数上对称简单随机游动是随机游动中最简单的一类, 它在每个点处向左和向右的可能性是一样的, 这种随机游动可以简单地由掷一枚公正的硬币的方式来实现. 在第二部分我们介绍强Markov 性, 然后用这个方法来研究简单随机游动的一些问题.

§2.1　简单对称随机游动

随机游动是最早被数学家关注和研究的随机现象, 因为它非常的常见, 比如两个人可以用一个硬币开始, 正面赢一块, 反面输一块, 这样一直下去就是一个简单的随机游动, 当然硬币只是起一个随机发生器的作用, 我们可以用掷骰子, 下棋, 打球, 猜数字等各种方法代替. 把其中某个人持有的钱数按时间顺序记录下来画在黑板上, 赢一块钱向上一格, 输一块钱向下一格, 这样在黑板上就出现一条忽上忽下的轨迹, 称为一条样本轨道. 下次再赌, 又会有一条样本轨道, 所有这些可能的样本轨道是等可能出现的, 这实际上已经给我们关于简单对称随机游动的基本图景.

考虑X-Y 平面上的整数格点组成的空间, 设$m < n$, a, b 是整数, 一条(m, a)到(n, b) 的格点轨道是指整数列$(s_m, s_{m+1}, \ldots, s_n)$ 满足:

(i) $s_m = a$, $s_n = b$;

(ii) 对$m < k \leqslant n$, $s_k - s_{k-1}$ 取值为1 或-1, 即$\{s_k\}$ 以单位1 向上或向下.

用直线将其相邻的点$(k-1, s_{k-1})$ 与(k, s_k) 连接, 形成为一条折线. $n-m$ 称为是轨道的长度.

显然, 经过平移, 从(m,a) 到(n,b) 的格点轨道总数与从$(0,0)$ 到$(n-m,b-a)$ 的格点轨道总数是相同的. 下面我们用$N_{n,x}$ 表示$(0,0)$ 到(n,x) 的格点轨道总数.

显然, 存在连接$(0,0)$ 与(n,x) 的格点轨道当且仅当存在$p,q\in\mathbf{Z}_+$ 使得$n=p+q$, $x=p-q$. 实际上, 只要用p 和q 表示格点轨道中向上和向下的折线数就可以了. 这样一条从$(0,0)$ 出发的格点轨道会经过(n,x) 当且仅当它有p 条向上的折线, 因此由排列的思想得

$$N_{n,x}=\binom{p+q}{p}=\binom{n}{\frac{n+x}{2}}$$

反射原理

反射原理是一个简单朴素的结果, 在许多领域都有应用, 一般认为是Maxwell 和Kelvin 首先引入并运用的.

定理2.1.1 设$a,b>0$, $m<n$, 则从(m,a) 到(n,b) 的且与x-轴相遇的格点轨道总数与从(m,a) 到$(n,-b)$ 的格点轨道总数相同.

证明. 分别用A, B 表示从(m,a) 到(n,b) 的且与x-轴相遇的格点轨道全体与从(m,a) 到$(n,-b)$ 的格点轨道全体. 取A 的任何轨道$(s_m,\ldots,s_n)$, 设最迟遇到x-轴是在k 时刻, 那么$m<k<n$. 将轨道k 到n 部分按x-轴反射得格点轨道

$$(s_m,\ldots,s_k,-s_{k+1},\ldots,-s_n).$$

它是B 中的格点轨道. 容易验证映射

$$(s_m,\ldots,s_n)\mapsto(s_m,\ldots,s_k,-s_{k+1},\ldots,-s_n)$$

建立了A 到B 上的一一对应. 因此结论成立. □

练习**2.1.1** 设$b>a>0$, 证明: $(0,0)$ 到(n,a) 不遇到$y=b$ 的格点轨道数为$N_{n,a}-N_{n,2b-a}$.

练习**2.1.2** (重复反射) 设$a,b>0,-b<c<a$, 证明: $(0,0)$ 到(n,c) 的不遇到$y=-b$ 和$y=a$ 的格点轨道总数为

$$\sum_k(N_{n,2k(a+b)+c}-N_{n,2k(a+b)+2a-c}).$$

格点轨道给出一个概率空间, 精确地讲, 对任何$n \geqslant 1$, 用W_n 表示$(0,0)$ 出发长度为n 的格点轨道全体, $\mathbb{P}_n$ 是其上的古典概率, 即若$B \subset W_n$, 那么

$$\mathbb{P}_n(B) = \frac{|B|}{|W_n|} = \frac{|B|}{2^n}.$$

下面的计票问题的解法是反射原理的应用, 它由Whitworth (1878) 和Bertrand (1887) 提出.

例2.1.1 在一次投票中, 候选人P,Q 的得票分别为m, n 且$m > n$, 那么在整个投票过程中, P 的票数一直领先于(多于) Q 的票数的概率为$(m-n)/(m+n)$.

记在第n 个人投票后, P,Q 的票数差额为s_n. 那么投票过程是一条$(0,0)$ 到$(m+n, m-n)$ 的格点轨道$(s_0, s_1, \ldots, s_{m+n})$, 这样的轨道总数为$N_{m+n,m-n}$. 所有投票过程是等概率的, 是一个古典概率问题: $(0,0)$ 到$(m+n, m-n)$ 的格点轨道除起点外不遇到x-轴的概率. 这个事件等同于$(1,1)$ 到$(m+n, m-n)$ 的轨道不遇到x-轴. 其中的轨道总数等于$N_{m+n-1,m-n-1}$ 减去$(1,1)$ 到$(m+n, m-n)$ 遇到x-轴的轨道总数, 应用反射原理, 后者是$N_{m+n-1,m-n+1}$. 因此概率为

$$\frac{N_{m+n-1,m-n-1} - N_{m+n-1,m-n+1}}{N_{m+n,m-n}} = \frac{m-n}{m+n}.$$

那么怎么计算P 的票数一直不少于Q 的票数的概率呢? 那要计算$(0,0)$ 出发到$(m+n, m-n)$ 不跑到x-轴下面去的轨道总数, 往上平移1, 这等于$(0,1)$ 到$(m+n, m-n+1)$ 不遇到x-轴的轨道总数, 由反射原理, 这等于$N_{m+n,m-n} - N_{m+n,m-n+2}$, 因此概率为

$$\frac{N_{m+n,m-n} - N_{m+n,m-n+2}}{N_{m+n,m-n}} = \frac{m+1-n}{m+1}.$$

∎

现在介绍最简单的随机游动, 或者说一维对称简单随机游动. 固定概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上独立同分布随机变量序列$\{X_n : n \geqslant 1\}$, 其中

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

令

$$S_0 = 0, S_n := \sum_{k=1}^{n} X_k, \ n \geqslant 1,$$

称为1 维对称的简单随机游动. 对$\omega \in \Omega$,

$$(S_0(\omega), S_1(\omega), \ldots, S_n(\omega), \ldots)$$

称为是ω 的样本轨道, 它对应一条$(0,0)$ 出发的格点轨道.

不仅如此, 如果$(s_0,\ldots,s_n)\in W_n$, 那么$\mathbb{P}(S_i=s_i;0\leqslant i\leqslant n)=\frac{1}{2^n}$, 因此如果$B\subset W_n$, 那么$\mathbb{P}((S_0,\ldots,S_n)\in B)=\mathbb{P}_n(B)$. 换句话说, 考虑$\{S_n\}$ 行为(有限长度)的事件的概率和考虑前面所建立的格点轨道的概率是一致的. 因此, 对称的简单随机游动的概率空间上的问题等价于格点轨道构成的概率空间上的问题, 可以用数格点轨道的方法计算.

下面我们来计算随机游动的一些有趣的量的分布. 首先讨论首次返回时间的分布, 让我们先算格点轨道位于x-轴上方的概率. 引入符号$u_{2n}:=\mathbb{P}(S_{2n}=0)$, $n\geqslant 0$. 那么$u_0=1$,

$$u_{2n}=N_{2n,0}/2^{2n}=\frac{1}{2^{2n}}\binom{2n}{n}.$$

引理2.1.1 对$n\geqslant 1$,

$$\mathbb{P}(S_1>0,S_2>0,\ldots,S_{2n}>0)=\frac{1}{2}u_{2n},$$
$$\mathbb{P}(S_1\geqslant 0,S_2\geqslant 0,\ldots,S_{2n}\geqslant 0)=u_{2n}.$$

证明. 显然

$$\begin{aligned}&\mathbb{P}(S_1>0,S_2>0,\ldots,S_{2n}>0)\\&=\sum_{k=1}^{n}\mathbb{P}(S_1>0,S_2>0,\ldots,S_{2n-1}>0,S_{2n}=2k).\end{aligned}$$

由反射原理, 满足$S_1>0,S_2>0,\ldots,S_{2n-1}>0,S_{2n}=2k$ 的格点轨道总数是

$$N_{2n-1,2k-1}-N_{2n-1,2k+1},$$

因此

$$\begin{aligned}&\mathbb{P}(S_1>0,S_2>0,\ldots,S_{2n}>0)\\&=\frac{1}{2^{2n}}\sum_{k=1}^{n}(N_{2n-1,2k-1}-N_{2n-1,2k+1})\\&=\frac{N_{2n-1,1}}{2^{2n}}=\frac{1}{2}u_{2n}.\end{aligned}$$

对任何$k\geqslant 0$, 满足$S_1\geqslant 0,S_2\geqslant 0,\cdots,S_{2n}=2k$ 的格点轨道就是$(0,0)$ 到$(2n,2k)$ 与直线$y=-1$ 不交的格点轨道, 再由反射原理, 总数等于$N_{2n,2k}-N_{2n,2k+2}$. 因

此

$$\mathbb{P}(S_1 \geqslant 0, S_2 \geqslant 0, \cdots, S_{2n} \geqslant 0) = \frac{1}{2^{2n}} \sum_{k=0}^{n} (N_{2n,2k} - N_{2n,2k+2}) = N_{2n,0}/2^{2n}.$$

完成证明. □

用τ 表示首次返回零点的时间, 即

$$\tau := \inf\{n > 0 : S_n = 0\}.$$

注意τ 取值是'时间', 是轨道上的函数, 也就是随机变量, 且我们无法排除它取无穷为值, 因为我们不知道轨道是否一定会再回到0 点, 另外如果τ 有限, 它一定是偶数. 后面我们会经常遇到这样的随机变量, 它简称为0 点的首中时(首次命中的时间).

定理2.1.2 对任何$n \geqslant 1$,

(1) $\mathbb{P}(\tau = 2n) = u_{2n-2} - u_{2n} = \frac{1}{2n-1} u_{2n}$, $n \geqslant 1$.

(2) $u_{2n} = \sum_{r=1}^{n} \mathbb{P}(\tau = 2r) \cdot u_{2n-2r}$.

证明. 由引理立刻得知,

$$\mathbb{P}(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0).$$

故随机游动在$2n$ 时刻首次返回零点的概率为

$$\begin{aligned}
\mathbb{P}(\tau = 2n) &= \mathbb{P}(S_1 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} = 0) \\
&= \mathbb{P}(S_1 \neq 0, \ldots, S_{2n-2} \neq 0) - \mathbb{P}(S_1 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} \neq 0) \\
&= \mathbb{P}(S_{2n-2} = 0) - \mathbb{P}(S_{2n} = 0).
\end{aligned}$$

公式(2) 由全概率公式推出. □

通常用符号$f_{2n} := \mathbb{P}(\tau = 2n)$. 由定理看出$\mathbb{P}(\tau < \infty) = 1$, 这说明随机游动以概率1 在有限时间内返回零点. 由Stirling 公式, $u_{2n} \sim 1/\sqrt{\pi n}$, 因此$\mathbb{E}\tau = \infty$, 也就是说平均返回时间是无限的.

接着讨论极大游程和首次通过时间的分布. 令

$$A_n := \max\{S_0, S_1, \ldots, S_n\},$$

$$T_x := \inf\{n \geqslant 0: \ S_n = x\},$$

分别称为$\{S_n\}$ 的极大游程与首次通过x 的时间(x 的首中时). 显然如果$x \geqslant 0$, 则$A_n \geqslant x$ 当且仅当$T_x \leqslant n$. 设$x > 0$, $k \leqslant x$, 由反射原理, 从$(0,0)$ 到(n,k) 的遇到直线$X = x$ 的格点轨道总数等于从$(0,0)$ 到$(n, 2x-k)$ 的格点轨道总数. 因此

$$\mathbb{P}(S_n = k, A_n \geqslant x) = \mathbb{P}(S_n = 2x - k),$$

那么

$$\begin{aligned}\mathbb{P}(S_n = k, A_n = x) &= \mathbb{P}(S_n = k, A_n \geqslant x) - \mathbb{P}(S_n = k, A_n \geqslant x+1) \\ &= \mathbb{P}(S_n = 2x - k) - \mathbb{P}(S_n = 2x + 2 - k).\end{aligned}$$

极大游程的分布律

$$\begin{aligned}\mathbb{P}(A_n = x) &= \sum_{k \leqslant x} \mathbb{P}(S_n = k, A_n = x) \\ &= \mathbb{P}(S_n = x) + \mathbb{P}(S_n = x+1).\end{aligned}$$

显然$\mathbb{P}(S_n = x)$ 与$\mathbb{P}(S_n = x+1)$ 只有一个非零. 然后计算首次通过时的分布律

$$\{T_x = n\} = \{S_1 < x, \dots, S_{n-1} < x, S_n = x\}.$$

只需要计算$(0,0)$ 到$(n-1, x-1)$ 的不遇到直线$X = x$ 的格点轨道总数, 由反射原理, 它等于

$$N_{n-1,x-1} - N_{n-1,x+1}.$$

因此$\mathbb{P}(T_x = n) = (N_{n-1,x-1} - N_{n-1,x+1})/2^n$.

最后讨论末离时和集合的滞留时的分布. 用L_{2n} 表示长度为$2n$ 的格点轨道最后遇到0 的时间, 即

$$L_{2n} := \sup\{k \leqslant 2n: \ S_k = 0\}$$

被称为0 点的(在时刻$2n$ 前的)末离时, 它必是偶数.

定理2.1.3 $\mathbb{P}(L_{2n} = 2k) = u_{2k} \cdot u_{2n-2k}$, $0 \leqslant k \leqslant n$.

证明. 令$S'_j = \sum_{i=1}^{j} X_{2k+i}$. 显然

$$\{L_{2n} = 2k\} = \{S_{2k} = 0, S_{2k+1} \neq 0, S_{2k+2} \neq 0, \cdots, S_{2n} \neq 0\}$$

$$= \{S_{2k} = 0, S_1' \neq 0, S_2' \neq 0, \cdots, S_{2n-2k}' \neq 0\},$$

因此由独立性和引理2.1.1,

$$\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_1' \neq 0, S_2' \neq 0, \cdots, S_{2n-2k}' \neq 0) = u_{2k} \cdot u_{2n-2k}.$$

完成证明. □

显然对任何$n \geqslant 0$, 或者S_{n-1} 与S_n 都是非负的, 我们说第n 个时段是正的; 或者S_{n-1} 与S_n 都是非正的, 说第n 个时段是负的. 令σ_{2n} 是0 到$2n$ 时段正的时段数, 即

$$\sigma_{2n} := \sum_{k=1}^{2n} 1_{\{S_{k-1} \geqslant 0, S_k \geqslant 0\}},$$

称为随机游动在正集上的逗留时, 自然σ_{2n} 也必是偶数.

定理2.1.4 $\mathbb{P}(\sigma_{2n} = 2k) = u_{2k} \cdot u_{2n-2k}$, $0 \leqslant k \leqslant n$.

证明. 对n 用归纳法. 当$n = 1$ 时, 显然$\mathbb{P}(\sigma_2 = 0) = \mathbb{P}(\sigma_2 = 2) = 1/2$, 结论成立. 另外由引理2.1.1知$\mathbb{P}(\sigma_{2n} = 0) = \mathbb{P}(\sigma_{2n} = 2n) = u_{2n}$, 结论成立, 故只需对$0 < k < n$ 证明即可. 现在设$\mathbb{P}(\sigma_{2m} = 2k) = u_{2k} \cdot u_{2m-2k}$ 对任何$m < n$ 和$0 < k < m$ 成立. 用τ 表示首次返回0 的时间, 那么当$0 < k < n$ 时, 由全概率公式和归纳假设,

$$\begin{aligned}
\mathbb{P}(\sigma_{2n} = 2k) &= \sum_{r=1}^{n} \mathbb{P}(\sigma_{2n} = 2k | \tau = 2r)\mathbb{P}(\tau = 2r) \\
&= \sum_{r=1}^{n} \frac{1}{2}(\mathbb{P}(\sigma_{2n-2r} = 2k - 2r) + \mathbb{P}(\sigma_{2n-2r} = 2k))\mathbb{P}(\tau = 2r) \\
&= \frac{1}{2}\sum_{r=1}^{n}(u_{2k-2r}u_{2n-2k} + u_{2k}u_{2n-2r-2k})\mathbb{P}(\tau = 2r) \\
&= \frac{1}{2}(u_{2k}u_{2n-2k} + u_{2k}u_{2n-2k}),
\end{aligned}$$

由此推出结论, 而最后一个等号由定理2.1.2(2)得到. □

令

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \; x \in (0, 1),$$

那么由Stirling 公式$u_{2k} \cdot u_{2n-2k} \sim \frac{1}{n} f(k/n)$. 即有

$$\sum_{k<xn} u_{2k} \cdot u_{2n-2k} \sim \sum_{k/n<x} \frac{1}{n} f(k/n) \longrightarrow \int_0^x f(y)dy = \frac{2}{\pi} \arcsin\sqrt{x},$$

由此推出著名的反正弦律, 也就是说$L_{2n}/2n$ 和$\sigma_{2n}/2n$ 的分布函数渐近地是上面的反正弦函数. 注意到密度函数f 的形状, 我们发现L_{2n} 与σ_{2n} 分布都较集中在两端. 令人惊奇的是, 反正弦律在随机过程理论有不可思议的普遍性.

§2.2 随机游动与Markov 性

设有概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上的Bernoulli 序列

$$X_1, X_2, \ldots, X_n, \ldots,$$

其中$\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = -1) = q$ 且$p + q = 1$. 这样的独立同分布随机序列的存在性由推论2.3.1[9] 推出, 它也是后面介绍的更一般的Kolmogorov 相容定理的特例, 这里我们给一个简单初等的证明. 这样的概率空间与随机序列的存在性是很直观的, 以致于人们常常忘记它也是需要证明的, 比如考虑一个连续不断的抛硬币的随机试验, 它将产生这样一个序列. 这样的随机序列称为(成功概率p)的Bernoulli 序列.

定理2.2.1 设$p, q \geqslant 0$, $p + q = 1$, 则存在概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上独立同分布随机变量序列$X_1, X_2, \ldots, X_n, \ldots$, 其中

$$\mathbb{P}(X_n = 1) = p, \ \mathbb{P}(X_n = -1) = q.$$

证明. 取$[0, 1]$ 上的Lebesgue 测度空间作为概率空间$(\Omega, \mathcal{F}, \mathbb{P})$. 首先按长度比$q : p$ 分区间$[0, 1]$ 为两段$[0, q)$, $[q, 1]$; 定义X_1 在第1 个区间上为-1, 第2 个区间上为1; 然后再按长度比$q : p$ 将它们分为四段$[0, qq)$, $[qq, q)$, $[q, q + pq)$, $[q + pq, 1]$; 定义X_2 在奇数段为-1, 偶数段为1; 一直继续这个过程, 即每次都按比$q : p$ 将每个区间划分, 第n 次时将区间划分为2^n 个区间, 定义X_n 在奇数段上为-1, 在偶数段上为1, 现在$\{X_1 = \epsilon_1, \ldots, X_n = \epsilon_n\}$ 表示第n 次划分后的一个小区间, 其长度是$q^k p^{n-k}$, 其中$k = |\{0 \leqslant i \leqslant n : \epsilon_i = -1\}|$, 因此容易验证$X_1, X_2, \ldots$ 是满足要求的独立同分布随机变量. □

对$x \in \mathbf{Z}$, 令

$$S_0^x := x, \ S_n^x := x + \sum_{k=1}^{n} X_k, \ n \geqslant 1.$$

当$x = 0$ 时, 简单地写为S_n. 因此$S_n^x = x + S_n$. $\{S_n^x\}$ 称为是由x 出发的简单随机游动, 它向右移动一个单位的概率是p, 而向左移动一个单位的概率是q. 当$p > q$ 时, 说它偏向右的, 反之偏向左的. $p = q$ 即是前面讨论过的对称随机游动.

现在我们看到, 在概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 下, 我们有一族随机序列$\{(S_n^x : n\geqslant 0) : x\in\mathbf{Z}\}$, 也就是说, 固定的概率测度下有不同的随机序列, 下面我们将用另外的观点来看同样的问题, 也就是说同一个随机序列, 而在不同的概率测度下. 两者在分布的意义下没有区别, 而后一种看法要更方便和直观.

为了不使读者觉得这种观点太过抽象且突兀, 我们来看看大家熟悉的分布函数. 我们可以用两种视点来看分布函数的实现, 首先存在概率空间$(\Omega,\mathcal{F},\mathbb{P})$, 满足对任何分布$F$, 存在随机变量$\xi$ 使得它的分布恰是F(见定理2.1.3[9]), 这是说同一个概率, 但不同的随机变量实现所有的分布. 另外, 我们也可以找到一个固定的可测空间$(\Omega,\mathcal{F})$ 和一个固定的随机变量ξ 满足对任何分布F, 存在概率μ_F, 使得ξ 在$(\Omega,\mathcal{F},\mu_F)$ 下的分布函数是F, 即同一个随机变量, 不同的概率实现所有的分布. 应该说这种做法在数学甚至在生活中也是经常的. 比如为了表示速度, 可以固定距离说一公里走二十分钟, 也可以说一小时走三公里, 也就是说可以约定距离用时间表示速度, 也可以约定时间用距离表示速度, 速度有多种表示方式, 甚至可以说两小时走六公里. 再举个例子, 在美国, 当问到汽车的油耗时, 人们经常简单地回答三十英里, 这是说30 miles/gallon, 而在中国对同样的问题, 人们常回答说八升, 这是说8 升/百公里.

现在我们就想找随机过程的另外一种表示. 用W 表示的整数值轨道全体, 精确地说是如下映射的全体:

$$w : \mathbf{Z}_+ \longrightarrow \mathbf{Z}.$$

对$w\in W$, 定义$\xi_n(w) := w(n)$, 是W 上的整数值函数, 表示轨道w 的n-坐标, $\mathcal{B}$ 表示W 上使得所有的ξ_n : $n\geqslant 0$ 都成为随机变量的最小事件域. 那么$\{\xi_n\}$ 是W 上关于事件域$\mathcal{B}$ 的随机变量序列(简称随机序列), W 与$\{\xi_n\}$ 通常称为是典则样本空间与典则随机序列. 下面我们从$(\Omega,\mathcal{F},\mathbb{P})$ 给空间$(W,\mathcal{B})$ 定义概率.

对任何给定的$\omega\in\Omega$, $x\in\mathbf{Z}$, 映射$n\mapsto S_n^x(\omega)$, $n\geqslant 0$ 是一个整数值轨道, 即W 中的元素, 记为$\phi_x(\omega)$. 故对任何$x\in\mathbf{Z}$, $\omega\mapsto\phi_x(\omega)$ 是样本空间Ω 到典则样本空间W 的一个映射. 那么实际上有$\xi_n\circ\phi_x = S_n^x$.

引理2.2.1 对任何$B\in\mathcal{B}$, $\{\phi_x\in B\} = \{\omega\in\Omega : \phi_x(\omega)\in B\}\in\mathcal{F}$.

证明. 先取特殊的$B=\{\xi_n = y\}$, 那么$\{\phi_x\in B\} = \{\xi_n\circ\phi_x = y\} = \{S_n^x = y\}\in\mathcal{F}$. 用$\mathcal{A}$ 表示$\mathcal{B}$ 中使得引理中关系成立的元素全体, 则$\mathcal{A}\subset\mathcal{B}$. 因为$\{\xi_n = y\}\in\mathcal{A}$, 所以每个$\xi_n$ 是关于$\mathcal{A}$ 的随机变量, 另外可以验证$\mathcal{A}$ 本身也是一个事件域, 因此推

出$\mathcal{B} \subset \mathcal{A}$, 因为$\mathcal{B}$ 是最小的. 完成证明. $\square$

这样的映射常称为是随机映射. 这时对任何$x \in \mathbf{Z}$, 定义概率$\mathbb{P}^x$ 为

$$\mathbb{P}^x(B) := \mathbb{P}(\{\omega : \phi_x\omega \in B\}) = \mathbb{E}(1_B \circ \phi_x),\ B \in \mathcal{B}.$$

$(W, \mathcal{B})$ 上的概率$\mathbb{P}^x$ 称为是$\mathbb{P}$ 在映射ϕ_x 下的象. 显然$\{\xi_n\}$ 在$\mathbb{P}^x$ 下的有限维联合分布与$\{S_n^x\}$ 在$\mathbb{P}$ 下的有限维联合分布一致, 即对任何$n \geqslant 0$, $x_0, \ldots, x_n \in \mathbf{Z}$, 有

$$\mathbb{P}^x(\xi_i = x_i : 0 \leqslant i \leqslant n) = \mathbb{P}(S_i^x = x_i : 0 \leqslant i \leqslant n).$$

对$x, y \in \mathbf{Z}$, 令

$$p(x, y) := \begin{cases} q, & y = x - 1; \\ p, & y = x + 1; \\ 0, & \text{其他}. \end{cases}$$

那么我们可以确切地写出有限维分布.

引理2.2.2 对任何$n \geqslant 0$, $x_0, \ldots, x_n \in \mathbf{Z}$, 有

$$\mathbb{P}^x(\xi_i = x_i : 0 \leqslant i \leqslant n) = 1_{\{x = x_0\}} p(x, x_1) p(x_1, x_2) \cdots p(x_{n-1}, x_n).$$

证明. 直接地计算

$$\begin{aligned} &\mathbb{P}^x(\xi_i = x_i : 0 \leqslant i \leqslant n) = \mathbb{P}(S_i^x = x_i : 0 \leqslant i \leqslant n) \\ &\quad = 1_{\{x = x_0\}} \mathbb{P}(S_i = x_i - x : 1 \leqslant i \leqslant n) \\ &\quad = 1_{\{x = x_0\}} \mathbb{P}(X_1 = x_1 - x, X_2 = x_2 - x_1, \ldots, X_n = x_n - x_{n-1}) \\ &\quad = 1_{\{x = x_0\}} \mathbb{P}(X_1 = x_1 - x) \mathbb{P}(X_2 = x_2 - x_1) \cdots \mathbb{P}(X_n = x_n - x_{n-1}) \\ &\quad = 1_{\{x = x_0\}} p(x, x_1) p(x_1, x_2) \cdots p(x_{n-1}, x_n). \end{aligned}$$

$\square$

轨道空间$(W, \mathcal{B})$ 上的概率集$\mathbf{D} = \{\mathbb{P}^x : x \in \mathbf{Z}\}$ 称为标准的(或典则)简单随机游动. 在$\mathbb{P}^x$ 下, $\{\xi_n\}$ 是x 出发的随机游动, 如同$\{S_n^x\}$ 在概率$\mathbb{P}$ 之下的行为. 实际上$\mathbb{P}^x$ 是支撑在格点轨道上的. (见习题.) 这种标准化的方法适用于一般随机过程, 特别是Markov 过程, 这种方法是由E.B.Dynkin 引入的. 相对于$\{S_n^x\}$ 与$\mathbb{P}$ 来说, 使

用$\{\xi_n\}$ 与$\mathbb{P}^x$ 的好处是前者概率是一个而随机序列是随x 变化的, 后者我们只是处理一个固定的随机序列, 前者是一个抽象样本空间, 而后者是具体的样本空间.

Markov 性

由上面的引理可用条件概率表示

$$\mathbb{P}^x(\xi_{n+1}=x_{n+1}|\xi_n=x_n,\dots,\xi_0=x_0)=p(x_n,x_{n+1})=\mathbb{P}^{x_n}(\xi_1=x_{n+1}). \quad \text{(M1)}$$

这就是**D** 的Markov 性: 已知现在$\xi_n=x_n$, 将来ξ_{n+1} 的可能位置与过去$\xi_0,\dots,\xi_{n-1}$ 的位置独立.

让我们用条件期望的方法来表示Markov 性, 用$\mathcal{B}_n$ 表示n+1-维随机变量$(\xi_0,\dots,\xi_n)$ 生成的事件域, 那么$\mathcal{B}_n\subset\mathcal{B}_{n+1}$. 注意$\mathcal{B}_n$ 是离散事件域, 而$\mathcal{B}$ 一般不是离散的.

引理2.2.3 对任何$n\geqslant 0$, $x,y\in\mathbf{Z}$, 有

$$\mathbb{P}^x(\xi_{n+1}=y|\mathcal{B}_n)=p(\xi_n,y)=\mathbb{P}^{\xi_n}(\xi_1=y), \quad \text{(M2)}$$

注意右边$\mathbb{P}^{\xi_n}(\xi_1=y)$ 是映射$w\mapsto\xi_n(w)$ 与$x\mapsto\mathbb{P}^x(\xi_1=y)$ 两者的复合, 后面类似的记号类似理解.

证明. 对任何$x_0,\cdots,x_n\in\mathbf{Z}$, 如果事件$\{s_n=x_n,\dots,s_0=x_0\}$ 的概率不等于0, 那么它是$\mathcal{B}_n$ 的原子. 由(M1) 与定理2.3.1[9] 推出(M2) 成立. □

让我们再引入W 上的推移算子, 对$n\geqslant 0$, $w\in W$, 定义

$$\theta_n w(k):=w(k+n),\ n\geqslant 0,$$

用通常的方法可以验证θ_n 是$(W,\mathcal{B})$ 到自身的可测映射, 并且

$$\xi_k\circ\theta_n=\xi_{k+n},\ k,n\geqslant 0.$$

θ_n 截去轨道w 的开头n 个点, 新的轨道$\theta_n w$ 从原轨道的n 时刻开始, 因此称为(左)推移算子. 推移后的轨道的k 坐标是原轨道的$n+k$ 坐标.

当然θ_n 是W 到自身的映射, 同引理2.2.1 的证明一样可验证对任何$B\in\mathcal{B}$, $\{w\in W:\theta_n(w)\in B\}\in\mathcal{B}$. 实际上, 令$\mathcal{G}_n$ 表示$\xi_n,\xi_{n+1},\dots$ 生成的事件域, 那么如果$B\in\mathcal{B}$, $\{w\in W:\theta_n(w)\in B\}\in\mathcal{G}_n$. 如果$Y$ 是W 上随机变量, 那么$Y\circ\theta_n$ 是$\mathcal{G}_n$ 随机变量. 通常$\mathcal{B}_n$ 代表过去, ξ_n 代表现在, 而$\mathcal{G}_n$ 代表将来. 用这种语言, (M2) 可以写成为

$$\mathbb{E}^x(1_{\{\xi_1=y\}}\circ\theta_n|\mathcal{B}_n)=\mathbb{P}^{\xi_n}(\xi_1=y).$$

更一般地有

定理2.2.2 对任意$n \geqslant 0$, $x \in \mathbf{Z}$ 与有界(或者非负)随机变量Y, 有

$$\mathbb{E}^x(Y \circ \theta_n | \mathcal{B}_n) = \mathbb{E}^{\xi_n}(Y). \tag{M3}$$

证明. 因为$w \mapsto \mathbb{E}^{\xi_n}(Y)$ 是$\mathcal{B}_n$ 随机变量, 故由定义只需验证对任何原子$A \in \mathcal{B}_n$,

$$\mathbb{E}^x(Y \circ \theta_n; A) = \mathbb{E}^x(\mathbb{E}^{\xi_n}(Y); A) \tag{2.2.1}$$

如果(2.2.1) 对示性函数$Y = 1_B$, $B \in \mathcal{B}$ 成立, 那么由条件期望的性质, 它对非负简单随机变量也成立, 取递增趋于Y 的简单随机变量列Y_n, 因此由单调收敛定理推出(2.2.1) 成立.

首先令$Y = 1_B$, 其中$B = \{\xi_0 = x_0, \ldots, \xi_m = x_m\}$ 与$A = \{\xi_0 = y_0, \ldots, \xi_n = y_n\}$, 那么由引理 2.2.2,

$$\begin{aligned}
&\mathbb{E}^x(1_B \circ \theta_n; A) \\
&\qquad = \mathbb{P}^x(\xi_0 = y_0, \ldots, \xi_n = y_n, \xi_n = x_0, \xi_{n+1} = x_1, \ldots, \xi_{n+m} = x_m) \\
&\qquad = 1_{\{x=y_0\}} p(y_0, y_1) \cdots p(y_{n-1}, y_n) 1_{\{x_0 = y_n\}} p(x_0, x_1) \cdots p(x_{m-1}, x_m) \\
&\qquad = 1_{\{x=y_0\}} p(y_0, y_1) \cdots p(y_{n-1}, y_n) \mathbb{P}^{y_n}(B) \\
&\qquad = \mathbb{E}^x(\mathbb{P}^{\xi_n}(B); A).
\end{aligned}$$

这样因为$\mathcal{B}_n$ 是离散的且由上面形式的A 的全体生成, 故(2.2.1) 对$Y = 1_B$, 其中B 是上面的形式, 与任何$A \in \mathcal{B}_n$ 成立.

从上面形式的1_B 到非负$\mathcal{B}$ 可测的随机变量Y 看上去有遥远的距离, 但实际上就是只要应用单调类定理1.1.2就足够了. 因为使得(2.2.1)成立的有界随机变量Y 全体是一个线性空间$\mathcal{L}$, 而且单调收敛定理成立, 而上面形式的B 全体对于有限交是封闭的且生成$\mathcal{B}$, 因此$\mathcal{L}$ 包含全体$\mathcal{B}$ 可测的非负或者有界的随机变量全体. □

定理证明中所遇到的问题是概率论中经常遇到, 所用的从示性函数到非负随机变量的方法也是概率论的标准方法, 在后面还将多次用到, 那时我们不再详细说明. 公式(M1), (M2), (M3) 都是Markov 性的表示形式.

强Markov 性

比固定时间$n \in \mathbf{Z}_+$ 更有用的是所谓停时的随机时间, W 上取值在$\mathbf{Z}_+ \cup \{+\infty\}$ 的(广义)随机变量称为是停时, 如果对任何$n \geqslant 0$, 有$\{T \leqslant n\} \in \mathcal{B}_n$. 然后定义

$$\mathcal{B}_T := \{B \in \mathcal{B}: \ B \cap \{T \leqslant n\} \in \mathcal{B}_n, \ n \geqslant 0\}.$$

我们来验证$\mathcal{B}_T$ 也是一个事件域. 事实上, 显然$\emptyset, W \in \mathcal{B}_T$, 另外若$B \in \mathcal{B}_T$, 则$B^c \cap \{T \leqslant n\} = \{T \leqslant n\} - B \cap \{T \leqslant n\} \in \mathcal{B}_n$, 故$\mathcal{B}_T$ 对补运算封闭, 最后$(\bigcup B_k) \cap \{T \leqslant n\} = \bigcup_k (B_k \cap \{T \leqslant n\})$, 因此$\mathcal{B}_T$ 对可列并运算也封闭.

下面事实的验证是简单的, 留作习题. (停时的详细讨论可见第四章.)

(1) 当T 是一个固定时间n 时, T 是停时且$\mathcal{B}_T = \mathcal{B}_n$.

(2) 设T, S 是停时, 那么$T \wedge S$ 也是.

(3) 设T, S 是停时且$S \leqslant T$, 那么$\mathcal{B}_S \subset \mathcal{B}_T$.

上面的性质(1) 说明使用符号$\mathcal{B}_T$ 是合理的.

例2.2.1 对$y \in \mathbf{Z}$, 定义$T_y := \inf\{n \geqslant 0 : \xi_n = y\}$, 称为是$y$ 的首次通过时. 显然

$$\{T_y \leqslant n\} = \cup_{0 \leqslant k \leqslant n} \{\xi_k = y\} \in \mathcal{B}_n,$$

因此T_y 是停时. 首次通过时是否在时间n 前发生可以从观察0 到n 的格点轨道得知, 这是停时的本质. 另外$\tau_y := \inf\{n \geqslant 1 : \xi_n = y\}$ 也是停时, 称为是y 的首中时. 显然$\tau_y \geqslant 1$ 且τ_y 与T_y 只是在那些从y 出发的轨道上有区别, 在这些轨道上$T_y = 0$. 我们可以验证下面的性质.

(1) 在事件$\{T_y \geqslant n\}$ 上有$T_y = T_y \circ \theta_n + n$ 成立. 但$T_y \circ \theta_n + n \geqslant T_y$ 恒成立.

(2) 在事件$\{T_y \geqslant T_x\}$ 上有$T_y = T_y \circ \theta_{T_x} + T_x$ 成立.

(3) 在事件$\{\tau_y > n\}$ 上有$\tau_y = \tau_y \circ \theta_n + n$ 成立, 而在$\{\tau_y = n\}$ 上不成立.

例如, 验证(1), 因为

$$\begin{aligned} T_y \circ \theta_n &= \inf\{k \geqslant 0 : \xi_k \circ \theta_n = y\} \\ &= \inf\{k \geqslant 0 : \xi_{k+n} = y\} = \inf\{k \geqslant n : \xi_k = y\} - n, \end{aligned}$$

而当$T_y \geqslant n$ 时, $\inf\{k \geqslant n : \xi_k = y\} = T_y$, 所以(1) 成立. ∎

在事件$\{T=n\}$ 上定义$\xi_T(w):=\xi_n(w)$, 即$\xi_T:=\sum_{n\geqslant 0}\xi_n 1_{\{T=n\}}$. 注意$\xi_T$ 只在$\{T<\infty\}$ 上, 它是随机游动在停时T 时刻的位置. 我们说样本空间上的函数X 在事件A 上是$\mathcal{F}$ 随机变量, 是指对任何$x\in\mathbf{R}$, $\{X\leqslant x\}\cap A\in\mathcal{F}$.

引理2.2.4 如果T 是停时, 那么ξ_T 在$\{T<\infty\}$ 上是关于事件域$\mathcal{B}_T$ 的随机变量.

证明. 需要验证对$y\in\mathbf{Z}$, $\{\xi_T=y\}\cap\{T<\infty\}\in\mathcal{B}_T$. 对任何$n\geqslant 0$,

$$\{\xi_T=y\}\cap\{T<\infty\}\cap\{T\leqslant n\}=\cup_{0\leqslant k\leqslant n}\{\xi_k=y,T=k\}\in\mathcal{B}_n.$$

□

再定义停时的推移$\theta_T w(n)=w(n+Tw)$, 即$\theta_T=\theta_n$ 如果$T=n$. 同样θ_T 也是当$T<\infty$ 时才有意义, 它是$(W,\mathcal{B})$ 到$(W,\mathcal{B})$ 的随机映射, 因为对$B=\{\xi_n=x\}$, $\{w\in W:\theta_T(w)\in B\}\cap\{T<\infty\}=\bigcup_k\{\xi_n\circ\theta_k=x\}\cap\{T=k\}=\bigcup_k\{\xi_{n+k}=x\}\cap\{T=k\}\in\mathcal{B}$. 下面定理是关于停时的Markov 性, 称为**D** 的强Markov 性.

定理2.2.3 设T 是停时, Y 是非负随机变量, 那么在$\{T<\infty\}$ 上有

$$\mathbb{E}^x(Y\circ\theta_T|\mathcal{B}_T)=\mathbb{E}^{\xi_T}(Y),$$

即对任何$A\in\mathcal{B}_T$ 有

$$\mathbb{E}^x(Y\circ\theta_T;A,T<\infty)=\mathbb{E}^x(\mathbb{E}^{\xi_T}(Y);A,T<\infty). \tag{M4}$$

证明. 类似上面定理的证明, 我们只需对$Y=1_B$, $B\in\mathcal{B}$ 验证(M4) 成立就足够了. 利用(M3), 因为$A\cap\{T=n\}\in\mathcal{B}_n$, 故

$$\begin{aligned}
&\mathbb{E}^x(1_B\circ\theta_T;A,T<\infty)\\
&\quad=\sum_{n\geqslant 0}\mathbb{E}^x(1_B\circ\theta_T;A\cap\{T=n\})\\
&\quad=\sum_{n\geqslant 0}\mathbb{E}^x(1_B\circ\theta_n;A\cap\{T=n\})\\
&\quad=\sum_{n\geqslant 0}\mathbb{E}^x(\mathbb{P}^{\xi_n}(B);A\cap\{T=n\})\\
&\quad=\sum_{n\geqslant 0}\mathbb{E}^x(\mathbb{P}^{\xi_T}(B);A\cap\{T=n\})
\end{aligned}$$

$$= \mathbb{E}^x(\mathbb{P}^{\xi_T}(B); A, T < \infty).$$

完成了证明. $\square$

空间平移不变性

随机游动有一个重要的性质, 空间的平移不变性, 或空间齐性: 从x 出发轨道落在一个集合B 内的概率等于从$x+y$ 出发轨道落在$B+y$ 内的概率. 随机游动平移一个单位这是因为$p(x,y)$ 有平移不变性: $p(x+z, y+z) = p(x,y)$, $x, y, z \in \mathbf{Z}$. 要确切地描述, 需要引入平移算子, 对任何$y \in \mathbf{Z}$, $w \in W$, 令$\gamma_y w(n) := w(n) + y$, $n \geqslant 0$. γ_y 将整个轨道平移y 单位, 它是W 上的随机映射, 且满足$\xi_n \circ \gamma_y = \xi_n + y$, $\gamma_y^{-1} = \gamma_{-y}$.

练习**2.2.1** 证明: $T_x \circ \gamma_y = T_{x-y}$, $x, y \in \mathbf{Z}$.

定理2.2.4 空间平移不变性: 对W 上非负随机变量Y 有: $\mathbb{E}^x Y \circ \gamma_y = \mathbb{E}^{x+y} Y$, $x, y \in \mathbf{Z}$.

证明. 同样只需对$Y = 1_B$, $B = \{\xi_0 = x_0, \ldots, \xi_n = x_n\}$, 验证定理就足够了. 因为$p(x,y) = p(0, y-x)$, 故

$$\begin{aligned}
\mathbb{E}^x(1_B \circ \gamma_y) &= \mathbb{P}^x(\xi_0 \circ \gamma_y = x_0, \ldots, \xi_n \circ \gamma_y = x_n) \\
&= \mathbb{P}^x(\xi_0 = x_0 - y, \ldots, \xi_n = x_n - y) \\
&= 1_{\{x = x_0 - y\}} p(x_0 - y, x_1 - y) \ldots p(x_{n-1} - y, x_n - y) \\
&= 1_{\{x + y = x_0\}} p(x_0, x_1) \ldots p(x_{n-1}, x_n) \\
&= \mathbb{P}^{x+y}(B).
\end{aligned}$$

完成了证明. $\square$

因为$1_{\{\gamma_y \in B\}} = 1_B \circ \gamma_y$, 上式用概率表示就是$\mathbb{P}^x(\{\gamma_y \in B\}) = \mathbb{P}^{x+y}(B)$, 特别地$\mathbb{P}^x(B) = \mathbb{P}^0(\{\gamma_x \in B\})$, 即所有的分布律$\mathbb{P}^x$ 由$\mathbb{P}^0$ 平移得到.

首次通过时的母函数

这节中, 我们将应用随机游动的强Markov 性来计算首次通过时的母函数. 对任何$x, y \in \mathbf{Z}$, 令$\phi_{x,y}$ 是T_y 在概率空间$(W, \mathcal{B}, \mathbb{P}^x)$ 上的母函数

$$\phi_{x,y}(t) := \mathbb{E}^x t^{T_y}.$$

首先由空间平移不变性

$$\phi_{x,y}(t) = \mathbb{E}^x t^{T_y} = \mathbb{E}^0 t^{T_y \circ \gamma_x} = \mathbb{E}^0 t^{T_{y-x}} = \phi_{0,y-x}(t)$$
$$\phi_{x,x}(t) = \phi_{0,0}(t) = \mathbb{E}^0 t^{T_0} = 1.$$

设$x > 0$, 因为随机游动每次移动一个单位, 故从0 出发时通过x 必须先通过1, 即$T_x \geqslant T_1$, 这时$T_x = T_x \circ \theta_{T_1} + T_1$, 因此由强Markov 性,

$$\begin{aligned}\phi_{0,x}(t) &= \mathbb{E}^0 t^{T_x} = \mathbb{E}^0 t^{T_1 + T_x \circ \theta_{T_1}} \\ &= \mathbb{E}^0 t^{T_1} \cdot t^{T_x} \circ \theta_{T_1} \\ &= \mathbb{E}^0 (t^{T_1} \mathbb{E}^{\xi_{T_1}} t^{T_x}) \\ &= \mathbb{E}^0 t^{T_1} \cdot \mathbb{E}^1 t^{T_x} = \phi_{0,1}(t) \cdot \phi_{1,x}(t).\end{aligned}$$

记$\phi := \phi_{0,1}$, 那么$\phi_{0,x} = \phi^x$.

另一方面, 从0 出发也必有$T_x \geqslant 1$, 这时$T_x = T_x \circ \theta_1 + 1$, 由Markov 性

$$\begin{aligned}\phi_{0,x}(t) &= \mathbb{E}^0 t^{T_x} = \mathbb{E}^0 t^{1 + T_x \circ \theta_1} \\ &= t\mathbb{E}^0 (\mathbb{E}^{\xi_1} t^{T_x}) \\ &= t((\mathbb{E}^1 t^{T_x})p + (\mathbb{E}^{-1} t^{T_x})q) \\ &= tp\phi_{0,x-1}(t) + tq\phi_{0,x+1}(t).\end{aligned}$$

令$x = 1$, 我们得

$$\phi(t) = tp + tq\phi(t)^2,$$

因$\phi(t) \leqslant 1$, 故解得

$$\phi(t) = \frac{1 - \sqrt{1 - 4t^2pq}}{2tq}.$$

由对偶性得

$$\phi_{0,-1}(t) = \mathbb{E}^0 t^{T_{-1}} = \frac{1 - \sqrt{1 - 4t^2pq}}{2tp}.$$

由此推出, 0 点出发在有限步内到达1 的概率为

$$\mathbb{P}^0(T_1 < +\infty) = \phi(1) = \frac{1 - |p - q|}{2q} = \begin{cases} 1, & p \geqslant q; \\ \frac{p}{q}, & p < q. \end{cases}$$

即当随机移动偏向右时, 它必定在有限时间内到达1. 这时计算其期望值h

$$\mathbb{E}^0 T_1 = \phi'(1) = \begin{cases} +\infty, & p = q; \\ \frac{1}{p-q}, & p > q. \end{cases}$$

为了计算0 出发0 的首中时τ_0 的母函数, 因为从0 出发$\tau_0 \geqslant 2 > 1$, 故$\tau_0 = 1 + \tau_0 \circ \theta_1$, 再由Markov 性,

$$\begin{aligned} \mathbb{E}^0 t^{\tau_0} &= \mathbb{E}^0 t^{1+\tau_0\circ\theta_1} = t\mathbb{E}^0 t^{\tau_0\circ\theta_1} \\ &= t\mathbb{E}^0(\mathbb{E}^{\xi_1} t^{\tau_0}) = tp\mathbb{E}^1 t^{\tau_0} + tq\mathbb{E}^{-1} t^{\tau_0} \\ &= tp\phi_{0,-1}(t) + tq\phi_{0,1}(t) = 1 - \sqrt{1 - 4t^2pq}. \end{aligned}$$

0 出发在有限步内返回0 的概率

$$\mathbb{P}^0(\tau_0 < \infty) = 1 - |p - q| = \begin{cases} 2q, & p > q \\ 1, & p = q \\ 2p, & p < q, \end{cases}$$

注意只有在对称简单随机游动时, 此概率为1.

习　题

1. 设$\{\xi_n\}$ 是简单随机游动, 如果T 是停时, $T < \infty$, 定义$\xi'_n = \xi_{n+T} - \xi_T$, $n \geqslant 0$, 证明: ξ' 是一个从0 出发的简单随机游动. 形象地说, 随机游动从停时T 重新开始.

2. 一个粒子在$\{-N, -N+1, \cdots, N-1, N\}$ 上作简单随机游动($n \geqslant 1$), 向右一步的概率为p, 向左一步的概率为$q = 1 - p$, $-N$ 与N 是吸收状态. 设一个粒子从0 点出发, 计算它在返回0 前被吸收的概率.

3. A, B 两人玩一个骰子赌博, 两人共有L 元钱, 如果骰子是偶数, 则互不给钱, 如果骰子是1, 则A 给B 一元钱, 如果骰子是3 或5, 则B 给A 一元钱. 游戏一直到一人输光结束. 用p_k 表示在开始时A 有k 元钱而最后赢的概率, 写出关于$\{p_k\}$ 的差分方程.

4. 一个随机游动向前两格的概率是p 向后一格的概率是$q = 1 - p$, 对$a > 0$, 用$\pi(a)$ 表示这个随机游动有限步到达0 的概率, 证明:

$$\pi(a) = p\pi(a + 2) + q\pi(a - 1),$$

当$p \leqslant \frac{1}{3}$ 时, $\pi(a) = 1$.

第三章　经典鞅论

鞅是在二十世纪初引入的, 但五十年代Doob 的工作给了鞅真正的生命力, 现在鞅是随机分析中的中心内容和不可缺少的工具. 在这一章中, 我们将简单介绍鞅及其在现代金融理论中的应用.

§3.1　鞅序列及其应用

为了讨论鞅, 我们需要对条件期望等概念有更仔细的理解. 回忆可积随机变量ξ 对子σ-域$\mathcal{A}$ 的条件期望$\mathbb{E}(\xi|\mathcal{A})$ 就是满足对任何$A \in \mathcal{A}$, $\mathbb{E}(\xi; A) = \mathbb{E}(\eta; A)$ 的$\mathcal{A}$ 可测的随机变量η. 对随机变量集合$\{\eta_i : i \in I\}$, 用$\sigma(\eta_i : i \in I)$ 表示$\{\eta_i^{-1}(B) : i \in I, B \in \mathcal{B}\}$ 生成的σ-域, 注意它也是使得$\{\eta_i : i \in I\}$ 都可测的最小σ-域. 记$\mathbb{E}(\xi|\eta_i : i \in I) := \mathbb{E}(\xi|\sigma(\eta_i : i \in I))$.

在随机过程理论中, 鞅是一个陌生的词, 直观地讲, 鞅是公平游戏的代名词. 拿随机游动为例, 设$\{\xi_n\}$ 是独立同分布随机序列, 分布为$\mathbb{P}(\xi_n = 1) = p$, $\mathbb{P}(\xi_n = -1) = 1 - p$. 可以理解为简单的博弈游戏. 令$X_n := \sum_{i=1}^n \xi_i$, 为$n$ 局游戏中总的输赢数. 显然当$p = 1/2$ 时, 游戏是公平的. 这时

$$\mathbb{E}(X_n|X_{n-1}, \cdots, X_1) = \mathbb{E}(\xi_n|X_{n-1}, \cdots, X_1) + X_{n-1} = X_{n-1},$$

也就是说, 如果我们以已知的前$n-1$ 局游戏的结果来预测下一局输赢, 期望是0.

设$(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间, 时间集I 是整数集合的一个子集, 不妨设它是整数的一个区间, 其实这不是很重要, 如果不是区间, 我们不妨把$n+1$ 理解为集I 中整数n 的下一个整数. 设$\{X_n : n \in I\}$ 是随机序列. 为了方便, 对任何$n \in I$, 定义

$$\mathcal{F}_n := \sigma(\{X_i : i \in I, i \leqslant n\}),$$

是时间n 前的随机变量所生成的σ-域, 当然$\mathcal{F}_n$ 关于n 是递增的, 集合$\{\mathcal{F}_n : n \in I\}$ 被称为是随机序列决定的自然流.

定义3.1.1 可积的实值随机序列$\{X_n : n \in I\}$ 称为鞅序列, 如果对任何$i, j \in I$ 且$i < j$, 有

$$\mathbb{E}(X_j|\mathcal{F}_i) = X_i.$$

另外$\geqslant$ 号成立时, 称为下鞅, $\leqslant$ 号成立时, 称为上鞅.

鞅的定义等价于对任何$n \in I$, $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$. 一个鞅的期望与$n$ 无关, 因为$\mathbb{E}X_n = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \mathbb{E}X_{n+1}$. 而下鞅的期望是$n$ 的递增序列.

例3.1.1 独立随机序列诱导的鞅: 设$I = \mathbf{N}$, $\{\xi_n\}$ 是可积的独立随机序列, 且$\mathbb{E}\xi_n = 0$, 那么$\{\xi_1 + \cdots + \xi_n\}_{n\geqslant 1}$ 是鞅序列. 如果$\xi_n \geqslant 0$, 有界且$\mathbb{E}\xi_n = 1$, 那么$\{\xi_1 \cdot \xi_2 \cdots \xi_n\}_{n\geqslant 1}$ 是鞅序列. ▮

例3.1.2 (Wald 鞅)设$\{\xi_n\}$是上一节中定义的简单随机游动, 那么其中的σ-域列$\{\mathcal{B}_n\}$是递增的, 对$\lambda > 0$, 令

$$Y_n := \lambda^{\xi_n}, \ n \geqslant 0.$$

因ξ_n 是$\mathcal{B}_n$ 可测的, 故Y_n 也是.

$$\begin{aligned}\mathbb{E}^x(Y_{n+1}|\mathcal{B}_n) &= \mathbb{E}^x(\lambda^{\xi_n + X_{n+1}}|\mathcal{B}_n) = \lambda^{\xi_n} \cdot \mathbb{E}^x(\lambda^{X_{n+1}}|\mathcal{B}_n) \\ &= Y_n \cdot (\lambda p + \frac{q}{\lambda}).\end{aligned}$$

因此$\{\lambda^{\xi_n}(\lambda p + \frac{q}{\lambda})^{-n}\}$ 是鞅序列, 称为Wald 鞅. 当$p \neq q$ 时, 取$\lambda = \frac{q}{p}$, $\lambda p + \frac{q}{\lambda} = 1$, 因此$\{\left(\frac{q}{p}\right)^{\xi_n}\}$ 是鞅序列. ▮

例3.1.3 (Doob 鞅) 设ξ 是$(\Omega, \mathcal{F}, \mathbb{P})$ 上可积随机变量, $\{\mathcal{F}_n : n \in I\}$ 是$\mathcal{F}$ 的一个关于n 递增的子σ-域的集合, 令

$$\xi_n := \mathbb{E}(\xi|\mathcal{F}_n),$$

那么$\{\xi_n\}$ 是一个关于流$(\mathcal{F}_n)$ 的鞅. ▮

简单地设I 是离散时间集$\{0, 1, 2, \cdots\}$, 但只要重新编号, 结果显然对连续整数的集合也是成立的. 设$(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, $(\mathcal{F}_n : n \in I)$ 是流. 一个随机序列$\{H_n : n \geqslant 0\}$ 称为是可预料的, 如果H_0 是$\mathcal{F}_0$ 可测的且对任何$n \geqslant 1$, H_n 是$\mathcal{F}_{n-1}$ 可测的. 设X 是适应过程, H_n 是可预料过程, 定义

$$\begin{aligned}&(H \bullet X)_0 := H_0 X_0, \\ &(H \bullet X)_n := (H \bullet X)_{n-1} + H_n(X_n - X_{n-1}), \ n \geqslant 1.\end{aligned}$$

称为是过程H 关于X 的随机积分, 它是一般随机积分的离散形式. (为了在符号上区别乘积与随机积分, 除非必需, 我们写乘积时一般不用点.)

定理3.1.1 设X 是一个适应过程, H 是可预料有界过程. 如果X 是鞅, 那么过程$H \bullet X$ 是鞅. 如果X 是下鞅且H 非负, 那么$H \bullet X$ 是下鞅.

证明. 显然$(H\bullet X)_n$ 是可积的并且$\mathcal{F}_n$ 可测的, 且对$n \geqslant 1$,

$$\begin{aligned}\mathbb{E}[(H\bullet X)_n - (H\bullet X)_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}(H_n(X_n - X_{n-1})|\mathcal{F}_{n-1})\\ &= H_n\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}),\end{aligned}$$

如果X 是鞅, 则右边是零, 即$H\bullet X$ 是鞅; 如果X 是下鞅且H 非负, 则右边也非负, 即$H\bullet X$ 是下鞅. □

上面的定理是随机分析中非常本质的结果, 千百年来, 赌徒们总是想在赌桌上发现对自己有利的策略, 经验说明这是徒劳的. 以上定理从理论上满意地解释了这个经验, 也再次诠释了[9]第30页中所引用的Feller 的话. 设X_n 是第n 次赌博后某赌徒A 的所有赌资, 则$X_n - X_{n-1}$ 是A 第n 次赌博中输赢的数目, 另一个赌徒B 赌A 的运气, H_n 是乘子, 也就是B 的策略. 但B 也不可能预知下一局A 的输赢, H_n 只能根据$X_0, X_1, \cdots, X_{n-1}$ 的结果决定, 即H_n 是$\mathcal{F}_{n-1}$ 可测的, (从这个意义解释, 可预料也许应称为不可预料.) 定理指出B 的运气不可能比A 更好, 也不可能更坏.

在某个时间停止赌博是一种简单策略. 让我们引入停时的概念, 它是概率论中最重要的概念之一.

定义3.1.2 一个值域为I 的随机变量T 称为是(相对于流$(\mathcal{F}_n : n \in I)$ 的)停时, 如果对任何$n \in I$, $\{T = n\} \in \mathcal{F}_n$.

典型的停时是首中时, 如果A 是Borel 集, 定义T 是序列$\{X_n : n \in I\}$ 首次遇到A 的时间, 即$T := \inf\{n \in I : X_n \in A\}$, 那么$T$ 是停时, 理由是

$$\{T = n\} = \{X_n \in A\} \cap \{X_{n-1} \notin A\} \cap \cdots \cap \{X_0 \notin A\} \in \mathcal{F}_n.$$

因为$\mathcal{F}_n$ 关于n 递增, 故T 是停时等价于对任何$n \in I$, $\{T \leqslant n\} \in \mathcal{F}_n$. 如果$T$ 与S 是停时, 那么$\{T \wedge S \leqslant n\} = \{T \leqslant n\} \cup \{S \leqslant n\}$, 故有下面引理.

引理3.1.1 如果T 与S 是停时, 那么$T \wedge S$ 与$T \vee S$ 也是停时.

对于随机序列$\{X_n : n \in I\}$, 自然地在集合$\{T = n\}$ 上定义$X_T := X_n$, $n \in I$, 被称为X 在停时T 处的位置. 定义T-停止序列

$$X_n^T(\omega) := X_{n\wedge T}(\omega),\ n \geqslant 0,$$

实际上, 它可以写为

$$X_n^T = \sum_{k=0}^{n-1} X_k 1_{\{T=k\}} + X_n 1_{\{T\geqslant n\}}$$

$$
\begin{aligned}
&= \sum_{k=0}^{n-1} X_k(1_{\{T \geqslant k\}} - 1_{\{T \geqslant k+1\}}) + X_n 1_{\{T \geqslant n\}} \\
&= X_0 + \sum_{k=1}^{n} 1_{\{T \geqslant k\}}(X_k - X_{k-1}).
\end{aligned}
$$

而$\{T \geqslant n\} = \{T < n\}^c \in \mathcal{F}_{n-1}$, 应用定理3.1.1 得到有界停止定理.

定理3.1.2 (Doob) (1) 如果$\{X_n : n \in I\}$ 是鞅, T 是停时, 那么鞅的停止序列$\{X_n^T : n \in I\}$ 也是鞅. 进一步, 如果T 是有界的, 那么$\mathbb{E}X_T = \mathbb{E}X_0$. (2) 设$X$ 是下鞅, S, T 是停时且$S \leqslant T$, 则$\{X_n^T - X_n^S : n \in I\}$ 是下鞅. 因此$\mathbb{E}X_n^S \leqslant \mathbb{E}X_n^T$.

证明. 只需证明(2). 由上面X_n^T 的表达式, 结合条件$S \leqslant T$ 得

$$
X_n^T - X_n^S = \sum_{k=1}^{n} 1_{\{T \geqslant k\} \setminus \{S \geqslant k\}}(X_k - X_{k-1}).
$$

由定理3.1.1推出结论. □

定理中(1)的第二个结论通常称为Doob 有界停止定理, 是非常有用的, 但问题是停时一般都不会是有界的, 所以研究什么情况下$\mathbb{E}X_T = \mathbb{E}X_0$ 成立是非常有意义的问题. 首先举个例子说明, 结论一般是不对的.

例3.1.4 设$\{X_n : n \geqslant 0\}$ 是直线上0 出发的的简单对称随机游动, 它是鞅. 定义T 是点1 的首中时, 那么$X_T = 1$, 所以$\mathbb{E}X_T = 1 \neq 0 = \mathbb{E}X_0$. ▮

但下面的定理说明在随机游动的情况下, T 的可积性能保证等式成立.

定理3.1.3 (Wald) 设$\{\xi_n : n \geqslant 1\}$ 是可积独立同分布随机序列且$\mathbb{E}\xi_1 = 0$, T 是可积停时, 则$\mathbb{E}\sum_{n=1}^{T} \xi_n = 0$.

证明. 定义$X_n := \sum_{i=1}^{n} \xi_i$, 那么$\{X_n : n \geqslant 1\}$ 是鞅. 由Doob 停时定理, 对任何n, $\mathbb{E}X_{T\wedge n} = 0$. 因此如果$T$ 有界, 定理结论成立. 下面我们证明当T 可积时, X_T 可积. 事实上,

$$
\mathbb{E}\left(\sum_{i=1}^{T\wedge n} |\xi_i|\right) = \mathbb{E}(T \wedge n) \cdot \mathbb{E}|\xi_1|,
$$

因此由单调收敛定理得

$$
\mathbb{E}|X_T| \leqslant \mathbb{E}\sum_{i=1}^{T} |\xi_i| = \lim_{n\to\infty} \mathbb{E}(T \wedge n) \cdot \mathbb{E}|\xi_1| = \mathbb{E}T \cdot \mathbb{E}|\xi_1| < \infty.
$$

而$\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_T; T \leqslant n) + \mathbb{E}(X_n; T > n)$. 首先由控制收敛定理算右边第一项的极限, $\lim_n \mathbb{E}(X_T; T \leqslant n) = \mathbb{E}X_T$. 另一方面, $\mathbb{E}(|X_n|; T > n) \leqslant \mathbb{E}(\sum_{i=1}^{T} |\xi_i|; T >$

n), 因为$\sum_{i=1}^{T}|\xi_i|$ 可积, 故再用控制收敛定理推出$\mathbb{E}(|X_n|; T > n) \longrightarrow 0$. 因此推出$\mathbb{E}X_T = 0$. □

下面我们将证明鞅有很好的收敛性质, 出发点是Doob 的上窜不等式. 给定N, 不妨取$I = \{n \in \mathbf{Z} : 0 \leqslant n \leqslant N\}$, $X = \{X_n : n \in I\}$ 是实值适应随机序列, 对$-\infty < a < b < \infty$, 定义

$$\begin{aligned}
\tau_0 &:= 0\\
\tau_1 &:= \inf\{n \geqslant 0 : X_n \leqslant a\};\\
\tau_2 &:= \inf\{n \geqslant \tau_1 : X_n \geqslant b\};\\
&\cdots\ \cdots\\
\tau_{2k+1} &:= \inf\{n \geqslant \tau_{2k} : X_n \leqslant a\};\\
\tau_{2k+2} &:= \inf\{n \geqslant \tau_{2k+1} : X_n \geqslant b\};\\
&\cdots\ \cdots
\end{aligned}$$

习惯地, 总是假设空集的下确界是$+\infty$, 则$\{\tau_n : n \geqslant 0\}$ 是一个递增的停时序列, 可以看出$\tau_{n+1} \geqslant n$, 当$n > N$ 时, $\tau_{n+1} \equiv +\infty$. 定义

$$U_I^X[a,b] := \max\{k : \tau_{2k} < +\infty\},$$

随机变量$U_I^X[a,b]$ 记录了随机序列X 在时间I 内从a 下跳至b 上的上窜次数. 下面是著名的Doob 上窜不等式.

定理3.1.4 (Doob) 如果X 是一个下鞅, 则

$$\mathbb{E}U_I^X[a,b] \leqslant \frac{1}{b-a}[\mathbb{E}(X_N - a)^+ - \mathbb{E}(X_0 - a)^+].$$

证明. 令$Y_n := (X_n - a)^+$, 由Jensen 不等式, $Y = (Y_n)$ 也是一个下鞅. 让$\tau_1, \tau_2, \cdots$ 是将$0, b-a, Y$ 分别取代a, b, X 后如上定义的停时列, 自然$U_I^X[a,b] = U_I^Y[0, b-a]$. 取$k$ 使$2k-1 \geqslant N$, 那么$\tau_{2k} \geqslant 2k-1 \geqslant N$, 因此

$$\begin{aligned}
Y_N - Y_0 &= \sum_{n=1}^{\infty}(Y_{\tau_n \wedge N} - Y_{\tau_{n-1}\wedge N})\\
&= \sum_{n\geqslant 1}(Y_{\tau_{2n}\wedge N} - Y_{\tau_{2n-1}\wedge N}) + \sum_{n\geqslant 0}(Y_{\tau_{2n+1}\wedge N} - Y_{\tau_{2n}\wedge N})\\
&\geqslant (b-a)U_I^Y[0, b-a] + \sum_{n\geqslant 0}(Y_{\tau_{2n+1}\wedge N} - Y_{\tau_{2n}\wedge N}),
\end{aligned}$$

其中的和是有限和, 最多只有N 个, 而由定理3.1.2, $\mathbb{E}Y_{\tau_{2n+1}\wedge N} \geqslant \mathbb{E}Y_{\tau_{2n}\wedge N}$, 因此

$$\begin{aligned}\mathbb{E}Y_N - \mathbb{E}Y_0 &\geqslant (b-a)\mathbb{E}U_I^X[a,b] + \sum_{n=0}^{k-1}(\mathbb{E}Y_{\tau_{2n+1}\wedge N} - \mathbb{E}Y_{\tau_{2n}\wedge N})\\ &\geqslant (b-a)\mathbb{E}U_I^X[a,b].\end{aligned}$$

□

Doob 上窜不等式是证明所有的鞅或下鞅收敛定理的基本工具.

定理3.1.5 设$\{X_n : n \geqslant 0\}$ 是下鞅且$K = \sup_n \mathbb{E}|X_n| < \infty$. 则$X_n \to X$ a.s., 其中X 是一个可积随机变量. 另外若$\{X_n\}$ 是一个一致可积鞅, 则$X_n \xrightarrow{L^1} X$ 且$X_n = \mathbb{E}(X|\mathcal{F}_n)$.

证明. 设X^*, X_* 分别是$n \to +\infty$ 时$\{X_n\}$ 的上极限与下极限. 显然

$$\{X^* > X_*\} = \bigcup_{a,b\in\mathbb{Q}} \{X_* < a < b < X^*\}.$$

右边是可列并, 由上窜不等式

$$\mathbb{E}U_{0,N}^X[a,b] \leqslant \frac{1}{b-a}(\mathbb{E}|X_N| + |a|) \leqslant \frac{K+|a|}{b-a}.$$

由单调收敛定理$\mathbb{E}\lim_N U_{0,N}^X[a,b] < +\infty$. 因此$\lim_N U_{0,N}^X[a,b] < +\infty$ a.s. 而

$$\{X_* < a < b < X^*\} \subset \{\lim_N U_{0,N}^X[a,b] = +\infty\},$$

故有$\mathbb{P}(\{X_* < a < b < X^*\}) = 0$, 再由可列可加性推出$X^* = X_*$ a.s. 极限的可积性由Fatou 引理得到. 如果$\{X_n\}$ 是一致可积, 则由定理1.2.2, $X_n \xrightarrow{L^1} X$ 且$X_n = \lim_m \mathbb{E}(X_m|\mathcal{F}_n) = \mathbb{E}(X|\mathcal{F}_n)$. □

上面定理讨论的是鞅或下鞅向右收敛的问题, 即$n \longrightarrow +\infty$ 时的收敛, 下面定理说明下鞅的左边无限时的收敛要容易得多, 即$n \longrightarrow -\infty$ 时的收敛.

定理3.1.6 设$X = (X_n : n \leqslant 0)$ 是关于流$(\mathcal{F}_n : n \leqslant 0)$ 的下鞅且$\inf_n \mathbb{E}X_n > -\infty$. 则$X$ 是一致可积的且当$n \to -\infty$ 时, X_n 几乎处处且L^1 收敛于一个可积随机变量$X_{-\infty}$.

证明. 设$n \leqslant 0$, 因$\mathbb{E}X_n \geqslant \mathbb{E}X_{n-1}$, 故$\inf_n \mathbb{E}X_n > -\infty$ 蕴含着$x = \lim_{n\to-\infty}\mathbb{E}X_n$ 存在且有限. 对给定$\epsilon > 0$, 取k 使得$\mathbb{E}X_k - x < \epsilon$, 那么当$n \leqslant k$ 时,

$$\mathbb{E}(|X_n|; |X_n| > \lambda) = \mathbb{E}(X_n; X_n > \lambda) - \mathbb{E}(X_n; X_n < -\lambda)$$

$$\begin{aligned}
&= \mathbb{E}(X_n; X_n > \lambda) + \mathbb{E}(X_n; X_n \geqslant -\lambda) - \mathbb{E}X_n \\
&\leqslant \mathbb{E}(X_k; X_n > \lambda) + \mathbb{E}(X_k; X_n \geqslant -\lambda) - \mathbb{E}X_k + \epsilon \\
&\leqslant \mathbb{E}(X_k; X_n > \lambda) + \mathbb{E}(-X_k; X_n < -\lambda) + \epsilon \\
&\leqslant \mathbb{E}(|X_k|; |X_n| > \lambda) + \epsilon,
\end{aligned}$$

另外

$$\begin{aligned}
\mathbb{P}(|X_n| > \lambda) &\leqslant \frac{1}{\lambda}\mathbb{E}|X_n| = \frac{1}{\lambda}\mathbb{E}(2X_n^+ - X_n) \\
&= \frac{1}{\lambda}(2\mathbb{E}X_n^+ - \mathbb{E}X_n) \leqslant \frac{1}{\lambda}(2\mathbb{E}X_0^+ - x),
\end{aligned}$$

由此推出X 是一致可积的.

(2) 类似于定理 3.1.5 的证明. 设X^*, X_* 分别是当$n \to -\infty$ 时X_n 的上极限与下极限. 显然

$$\{X^* > X_*\} = \bigcup_{a,b\in\mathbb{Q}} \{X_* < a < b < X^*\}.$$

对负整数N, 将上窜不等式用于下鞅$\{X_N, X_{N+1}, \cdots, X_0\}$,

$$\mathbb{E}U_{N,0}^X[a,b] \leqslant \frac{1}{b-a}(\mathbb{E}|X_0| + |a|).$$

因此同样有$\lim_{N\to-\infty} U_{N,0}^X[a,b] < +\infty$ a.s. 但是

$$\{X_* < a < b < X^*\} \subset \{\lim_{N\to-\infty} U_{N,0}^X[a,b] = +\infty\},$$

故有$\mathbb{P}(\{X_* < a < b < X^*\}) = 0$, 推出$X^* = X_*$ a.s. 即$\lim_{n\to-\infty} X_n(\omega)$ 几乎处处存在, 记极限为$X_{-\infty}$. 再由$\{X_n : n \leqslant 0\}$ 的一致可积性, X_n 也是L^1 收敛于$X_{-\infty}$, 因此$X_{-\infty}$ 是可积的. □

作为负指标鞅的一个应用，我们来证明Kolmogorov 强大数定律. 它推广了Borel 强大数律(参考定理2.4.4[9]).

定理3.1.7 (Kolmogorov) 设$\{\xi_n\}$ 是独立同分布可积随机序列. 则

$$\frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)$$

几乎处处且L^1 收敛于$\mathbb{E}\xi_1$.

证明. 对$n \geqslant 1$ 定义$X_{-n} := \frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)$, $\mathcal{F}_{-n} := \sigma(X_{-k} : k \geqslant n)$. 那么$(X_{-n}, \mathcal{F}_{-n} : n \geqslant 1)$ 是鞅. 实际上, 对任何$n \geqslant 1$, 条件期望$\mathbb{E}(\xi_k | \sum_{i=1}^n \xi_i)$ 关于$1 \leqslant k \leqslant n$ 是不变的. 因此,

$$\mathbb{E}(\xi_1 | \sum_{i=1}^n \xi_i) = \frac{1}{n} \sum_{i=1}^n \xi_i = X_{-n}.$$

由于随机变量族$\{\sum_{i=1}^k \xi_i : k \geqslant n\}$ 与$\{\sum_{i=1}^n \xi_i, \xi_{n+1}, \xi_{n+2}, \cdots\}$ 可以互相线性表示, 故而$\mathcal{F}_{-n} = \sigma(\sum_{i=1}^n \xi_i, \xi_{n+1}, \xi_{n+2}, \cdots)$, 由独立性不难推出

$$\mathbb{E}(X_{-1} | \mathcal{F}_{-n}) = \mathbb{E}(\xi_1 | \mathcal{F}_{-n}) = \mathbb{E}(\xi_1 | \sum_{i=1}^n \xi_i) = X_{-n},$$

即$\{X_{-n}, \mathcal{F}_{-n} : n \geqslant 1\}$ 是鞅.

由可积性和上面的定理3.1.6推出X_{-n} a.s. 且L^1 收敛于一个可积随机变量X. 下面要证明$X = \mathbb{E}X$ a.s. 不妨设$\mathbb{E}\xi_1 = 0$, 则有$\mathbb{E}X = 0$. 让ϕ 表示ξ_1 的特征函数. 那么ϕ 在0 点可微且$\phi'(0) = 0$. 因此

$$\mathbb{E}e^{itX} = \lim_n \mathbb{E}e^{itX_{-n}} = \lim_n \left(\phi(\frac{t}{n})\right)^n = 1.$$

即$X = 0$ a.s. □

下面介绍的鞅不等式也是经典的, 且非常有用.

定理3.1.8 (Doob) 设$\{X_n : n \geqslant 0\}$ 是鞅.

(1) 对任何$\lambda > 0$ 及正整数N,

$$\lambda \mathbb{P}(\max_{0 \leqslant n \leqslant N} |X_n| \geqslant \lambda) \leqslant \mathbb{E}|X_N|.$$

(2) 对任何$p > 1$ 及正整数N,

$$\mathbb{E} \max_{0 \leqslant n \leqslant N} |X_n|^p \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_N|^p.$$

特别的, $p = 2$ 是最常用的场合,

$$\mathbb{E} \max_{0 \leqslant n \leqslant N} X_n^2 \leqslant 4\mathbb{E}X_N^2.$$

证明. 因为$|X|$ 实际上是非负下鞅, 所以下面不妨设X 是非负下鞅.

(1) 令$\tau := \min\{0 \leqslant n \leqslant N : X_n \geqslant \lambda\}$, 则$\tau$ 是一个停时且$\tau \leqslant N$, 故

$$\begin{aligned}\mathbb{E}X_N &\geqslant \mathbb{E}X_\tau \\ &= \mathbb{E}(X_\tau; \max_{0\leqslant n\leqslant N} X_n \geqslant \lambda) + \mathbb{E}(X_\tau; \max_{0\leqslant n\leqslant N} X_n < \lambda) \\ &\geqslant \lambda\mathbb{P}(\max_{0\leqslant n\leqslant N} X_n \geqslant \lambda) + \mathbb{E}(X_N; \max_{0\leqslant n\leqslant N} X_n < \lambda),\end{aligned}$$

因此

$$\lambda\mathbb{P}(\max_{0\leqslant n\leqslant N} X_n \geqslant \lambda) \leqslant \mathbb{E}(X_N; \max_{0\leqslant n\leqslant N} X_n \geqslant \lambda) \leqslant \mathbb{E}X_N.$$

(2) 令$\xi := X_N$, $\eta := \max_{n\leqslant N} X_n$, $q := \frac{p}{p-1}$. 则由(1) 的证明中知道$t\mathbb{P}(\eta \geqslant t) \leqslant \mathbb{E}(\xi; \{\eta \geqslant t\})$, 再结合Fubini 定理和Hölder 不等式

$$\begin{aligned}\mathbb{E}\eta^p &= \mathbb{E}\int_0^\eta pt^{p-1}dt = \int_0^\infty pt^{p-1}\mathbb{P}(\eta \geqslant t)dt \\ &\leqslant \int_0^\infty pt^{p-2}\mathbb{E}(\xi; \{\eta \geqslant t\})dt \\ &= p\mathbb{E}\xi\int_0^\eta t^{p-2}dt = \frac{p}{p-1}\mathbb{E}\xi\eta^{p-1} \\ &\leqslant q(\mathbb{E}\xi^p)^{\frac{1}{p}}(\mathbb{E}\eta^{(p-1)q})^{\frac{1}{q}} \\ &= q(\mathbb{E}\xi^p)^{\frac{1}{p}}(\mathbb{E}\eta^p)^{\frac{1}{q}}.\end{aligned}$$

两边同除$(\mathbb{E}\eta^p)^{\frac{1}{q}}$ 即得. □

下面两个例子显示鞅在解决经典问题时的应用.

例3.1.5 输光问题: 下面我们回到前面讨论的简单随机游动$\mathbf{D}$. 任取$a \in \mathbf{Z}$, $a > 0$, 令$T := T_0 \wedge T_a$, 即首次通过0 或a 其一的时间, 或$\{0, a\}$ 的进入时. 从上一节的结论知, 对任何$x \in \mathbf{Z}$, $0 \leqslant x \leqslant a$, $\mathbb{P}^x(T_0 < \infty)$ 与$\mathbb{P}^x(T_a < \infty)$ 至少有一个是1, 故$\mathbb{P}^x(T < \infty) = 1$. 显然$\{T < \infty\} = \{T_0 < T_a\} \cup \{T_0 > T_a\}$, 我们令

$$q_x := \mathbb{P}^x(T_0 < T_a),$$

即从x 出发的随机游动, 到达点0 在到达点a 之前的概率. 形象地, 这相当于A, B 两人各带$x, a-x$ 枚硬币参加一个简单随机游动形式的赌博游戏, 游戏至其中某人输光所有的硬币时结束, 因此q_x 通常也称为输光概率, T 是游戏持续时间. 理论上, 这称为是一个具吸收壁$\{0, a\}$ 的简单随机游动.

如果$p=q=\frac{1}{2}$, 那么$\{\xi_n\}$ 是一个鞅, 那么$\{\xi_n^T\}$ 也是鞅, 故

$$\mathbb{E}^x\xi_{n\wedge T}=\mathbb{E}^x\xi_0=x,$$

而

$$\begin{aligned}\mathbb{E}^x\xi_{n\wedge T}&=\mathbb{E}^x(\xi_{n\wedge T};T\geqslant n)+\mathbb{E}^x(\xi_{n\wedge T};T<n)\\&=\mathbb{E}^x(\xi_n;T\geqslant n)+\mathbb{E}^x(\xi_T;T<n).\end{aligned}$$

当$T\geqslant n$ 时, $\xi_n\leqslant a$, 故

$$\mathbb{E}^x(\xi_n;T\geqslant n)\leqslant a\mathbb{P}^x(T\geqslant n)\longrightarrow 0,$$

由单调收敛定理,

$$\mathbb{E}^x(\xi_T;T<n)\uparrow\mathbb{E}^x(\xi_T;T<\infty)=\mathbb{E}^x\xi_T.$$

而

$$\mathbb{E}^x\xi_T=\mathbb{E}^x(\xi_T;T_0<T_a)+\mathbb{E}^x(\xi_T;T_0>T_a)=a\mathbb{P}^x(T_0>T_a),$$

因此

$$q_x=1-\mathbb{P}(T_0>T_a)=1-\frac{x}{a}=\frac{a-x}{a}.$$

如果$p\neq q$, 不妨设$p>q$, 那么$\{(\frac{q}{p})^{\xi_n}\}$ 是一个鞅, 且$\{(\frac{q}{p})^{\xi_{n\wedge T}}\}$ 是有界的. 类似地由单调收敛定理, 我们有

$$\begin{aligned}(\frac{q}{p})^x&=\mathbb{E}^x(\frac{q}{p})^{\xi_{n\wedge T}}\\&=\mathbb{E}^x[(\frac{q}{p})^{\xi_n};n\leqslant T]+\mathbb{E}^x[(\frac{q}{p})^{\xi_T};n>T].\end{aligned}$$

而且

$$\begin{aligned}&\mathbb{E}^x[(\frac{q}{p})^{\xi_n};n\leqslant T]\leqslant\mathbb{P}^x(n\leqslant T)\longrightarrow 0,\\&\mathbb{E}^x[(\frac{q}{p})^{\xi_T};n>T]\uparrow\mathbb{E}^x(\frac{q}{p})^{\xi_T}\\&=\mathbb{E}^x[(\frac{q}{p})^{\xi_T};T_0<T_a]+\mathbb{E}^x[(\frac{q}{p})^{\xi_T};T_0>T_a]\\&=q_x+(\frac{q}{p})^a\cdot\mathbb{P}^x(T_0>T_a)\end{aligned}$$

$$= q_x[1-(\frac{q}{p})^a] + (\frac{q}{p})^a,$$

因此

$$q_x = \frac{(\frac{q}{p})^x - (\frac{q}{p})^a}{1-(\frac{q}{p})^a}.$$

∎

例3.1.6 持续时间: 要计算持续时间T 的母函数, 我们需要一个含有T 的鞅. 从例3.1.2 知道, 对任何$\lambda > 0$,

$$\{\lambda^{\xi_n}(\lambda p + \frac{q}{\lambda})^{-n}\}$$

是一个鞅. 不妨设$p \geqslant q$, 取$\lambda \notin (q/p, 1)$, 必有$\lambda p + \frac{q}{\lambda} \geqslant 1$. 利用Doob 停止定理

$$\begin{aligned}\lambda^x &= \mathbb{E}^x \lambda^{\xi_{n\wedge T}}(\lambda p + \frac{q}{\lambda})^{-(n\wedge T)} \\ &= \mathbb{E}^x[\lambda^{\xi_{n\wedge T}}(\lambda p + \frac{q}{\lambda})^{-(n\wedge T)}; n \leqslant T] + \mathbb{E}^x[\lambda^{\xi_{n\wedge T}}(\lambda p + \frac{q}{\lambda})^{-(n\wedge T)}; n > T]\end{aligned}$$

而

$$\mathbb{E}^x[\lambda^{\xi_{n\wedge T}}(\lambda p + \frac{q}{\lambda})^{-n\wedge T}; n \leqslant T] \leqslant (\lambda^a \wedge 1)\mathbb{P}^x(n \leqslant T) \longrightarrow 0,$$

因此, 我们有

$$\begin{aligned}\lambda^x &= \mathbb{E}^x \lambda^{\xi_T}(\lambda p + \frac{q}{\lambda})^{-T} \\ &= \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 < T_a] + \lambda^a \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 > T_a].\end{aligned}$$

令$\mu = \frac{q}{\lambda p}$, 那么$\lambda p + \frac{q}{\lambda} = \mu p + \frac{q}{\mu}$. 代入

$$(\frac{q}{\mu p})^x = \mathbb{E}^x[(\mu p + \frac{q}{\mu})^{-T}; T_0 < T_a] + (\frac{q}{\mu p})^a \mathbb{E}^x[(\mu p + \frac{q}{\mu})^{-T}; T_0 > T_a],$$

将μ 换写为λ 得两个方程

$$\begin{cases}\lambda^x = \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 < T_a] + \lambda^a \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 > T_a], \\ (\frac{q}{\lambda p})^x = \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 < T_a] + (\frac{q}{\lambda p})^a \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 > T_a].\end{cases}$$

解出

$$\begin{aligned}\mathbb{E}^x(\lambda p + \frac{q}{\lambda})^{-T} &= \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 < T_a] + \mathbb{E}^x[(\lambda p + \frac{q}{\lambda})^{-T}; T_0 > T_a] \\ &= \frac{\lambda^{x-a}(\frac{q}{p})^a - \lambda^x - \lambda^{a-x}(\frac{q}{p})^x + \lambda^{-x}(\frac{q}{p})^x}{(\frac{q}{p})^a \lambda^{-a} - \lambda^a}.\end{aligned}$$

取$|t| \leqslant 1$, 令$\lambda p + \frac{q}{\lambda} = \frac{1}{t}$, 那么$\lambda$ 有两个根

$$\lambda_1(t) = \frac{1+\sqrt{1-4t^2pq}}{2pt},\ \lambda_2(t) = \frac{1-\sqrt{1-4t^2pq}}{2pt}.$$

显然$\lambda_1 = \frac{q}{\lambda_2 p}$, 因此

$$\mathbb{E}^x t^T = \frac{\lambda_2(t)^x(\lambda_1(t)^a - 1) - \lambda_1(t)^x(\lambda_2(t)^a - 1)}{\lambda_1(t)^a - \lambda_2(t)^a}.$$

我们用它来计算T 的数学期望

$$D_x := \mathbb{E}^x T = \lim_{t\uparrow 1} \frac{1-\mathbb{E}^x t^T}{1-t}.$$

设$p > q$, 那么$t \uparrow 1$ 时, $\lambda_1 \longrightarrow 1$, $\lambda_2 \longrightarrow \frac{q}{p}$, 且$t = \frac{\lambda_1}{\lambda_1^2 p + q}$, 因此

$$\begin{aligned}
D_x &= \lim_{t\uparrow 1} \frac{1 - \frac{\lambda_2^x(\lambda_1^a - 1) - \lambda_1^x(\lambda_2^a - 1)}{\lambda_1^a - \lambda_2^a}}{1 - \frac{\lambda_1}{\lambda_1^2 p + q}} \\
&= \lim_{t\uparrow 1} \frac{\lambda_1^x(\lambda_1^{a-x} - 1) + \lambda_2^a(\lambda_1^x - 1) + \lambda_2^x(1 - \lambda_1^a)}{(\lambda_1 p - q)(\lambda_1 - 1)(\lambda_1^a - \lambda_2^a)} (\lambda_1^2 p + q) \\
&= \frac{-(a-x) + x(\frac{q}{p})^a + a(\frac{q}{p})^x}{(p-q)(1-(\frac{q}{p})^a)} \\
&= \frac{a}{p-q} \cdot \frac{1-(\frac{q}{p})^x}{1-(\frac{q}{p})^a} - \frac{x}{p-q}.
\end{aligned}$$

如果$p = q = \frac{1}{2}$, 那么$\lambda_2 = \lambda_1^{-1}$, 因此如用$\lambda$ 表示λ_2 (或λ_1), 我们得

$$\begin{aligned}
\mathbb{E}^x t^T &= \frac{\lambda^x(\lambda^{-a} - 1) - \lambda^{-x}(\lambda^a - 1)}{\lambda^{-a} - \lambda^a} \\
&= \frac{\lambda^x(1-\lambda^a) - \lambda^{a-x}(\lambda^a - 1)}{1 - \lambda^{2a}} \\
&= \frac{\lambda^x + \lambda^{a-x}}{1+\lambda^a}.
\end{aligned}$$

而类似地, 持续时间的期望

$$\begin{aligned}
D_x &= \lim_{t\uparrow 1} \frac{1 - \frac{\lambda^x + \lambda^{a-x}}{1+\lambda^a}}{1 - \frac{2\lambda}{1+\lambda^2}} \\
&= \lim_{t\uparrow 1} \frac{(1-\lambda^x)(1-\lambda^{a-x})}{(1-\lambda)^2} \cdot \frac{1+\lambda^2}{1+\lambda^a} = x(a-x).
\end{aligned}$$

∎

§3.2 二项期权定价理论

在本节的最后, 我们将介绍鞅在金融理论中的应用, 鞅的理论在金融市场特别是期权定价理论上的应用是如此的自然和漂亮, 使得金融数学的研究成为近年来的一个热点, 理论的创始人Scholes 与Merton 获得Nobel 经济学奖(另一主要创始人Black 在此时已过世).

我们在此考虑的是简单金融市场与二项期权定价, 但是其思想可平行地应用于连续时间的金融理论. 首先来看股票市场, 我们使用离散时间$\{0,1,2,\cdots,n\}$, 用S_k 表示股票在k 时刻的价格, 是一个随机变量, 在$k+1$ 时刻价格分别以概率p, q 上涨至uS_k 和下跌至dS_k, 当然u, d 是常数满足$0 < d < 1 < u$, $p, q > 0$ 且$p+q=1$. 确切地说, 让$\{\xi_1,\cdots,\xi_n\}$ 是概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 上的独立同分布随机变量且分布为

$$\mathbb{P}(\xi_k = u) = p,\ \mathbb{P}(\xi_k = d) = q.$$

S_0 是一个正常数, 令$S_k := S_0\cdot\xi_1\cdots\xi_k$, $k=1,2,\cdots,n$. 过程$\{S_0,S_1,\cdots,S_n\}$ 描述了股票价格的变化, 称为是一个股票市场, 它实际上由常数u,d,p,q 决定. 其中上涨与下跌的不确定性说明股票市场是有风险的市场. 任何人都可以买卖股票, 在k 时刻买一股股票, 在下一时刻卖出, 获利为$S_{k+1}-S_k$, 反之, 在k 时刻卖一股股票, 在下一时刻买入, 获利为$S_k - S_{k+1}$. 注意两个数都可能是负的.

用$\mathcal{F}_k$ 表示$\{\xi_1,\cdots,\xi_k\}$ 生成的事件域, $1 \leqslant k \leqslant n$, $\mathcal{F}_0 := \{\varnothing,\Omega\}$. 因为$\{S_k\}$ 与$\{\xi_k\}$ 可以互相表示, 故$\mathcal{F}_k = \sigma(\{S_0,S_1,\cdots,S_k\})$, 这时任何$\mathcal{F}_k$ 可测的随机变量X 都可以表示为$\xi_1,\cdots,\xi_k$ 的函数, 常简单地写成为

$$X = X(\xi_1,\cdots,\xi_k).$$

设$\overline{\mathbb{P}}$ 是$(\Omega,\mathcal{F})$ 上另外一个概率, 如果对任何n 及$A\in\mathcal{F}_n$, $\mathbb{P}(A)>0$ 当且仅当$\overline{\mathbb{P}}(A)>0$, 我们说$\overline{\mathbb{P}}$ 等价于$\mathbb{P}$.

再假设存在另一个市场, 让我们称为货币市场, 就是说任何人都可以以固定的利息率r (向银行)存款或(从银行) 贷款, 即存入银行(对应地, 贷出)的1 单位的货币在下一时刻可以取回(对应地, 要归还) $1+r$ 单位, 货币市场是不存在风险的, 它由常数r 决定.

这样, 上面描述的有风险的股票市场和无风险的货币市场组成了我们所谓的简单金融市场. 存款贷款买卖股票是市场的通常投资行为. 当然这是一个极其简化的模型, 不足以演绎真正的股票市场, 但隐含在其中的思想是深刻的, 对真正的

市场有启发作用. 现在任何人都可以在这个市场上通过买卖股票以及存贷款来达到赚钱的目的, 人的趋利性是指他总希望以最小的代价赚取最大利润.

一个投资策略是指平衡在两个市场的投资, 也就是决定在每个时刻k 拥有的股票数量Δ_k, $0 \leqslant k \leqslant n-1$, 并把剩余资金存入银行或不够时向银行贷款. 在这个时刻投资人所有的信息是0 到k 时刻股票的价格走势, 故Δ_k 是$\mathcal{F}_k$ 可测的, 或者说是$\xi_1, \cdots, \xi_k$ 的函数.

定义3.2.1 随机过程$\Delta = \{\Delta_0, \Delta_1, \cdots, \Delta_{n-1}\}$ 称为是一个投资策略(portfolio), 如果Δ_k 是$\mathcal{F}_k$ 可测的, $0 \leqslant k \leqslant n-1$. 一个投资策略是可许的, 如果任何$\Delta_k$ 有界.

我们说一个投资是自融资方式的, 如果除初始投资X_0 外, 在后面的交易过程中不再从市场外取得资金也不从中抽出资金. 用X_k 表示投资人在k 时刻拥有的财富(货币加股票价值), 则

$$\begin{aligned}
X_1 &= \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0), \\
X_2 &= \Delta_1 S_2 + (1+r)(X_1 - \Delta_1 S_1), \\
&\cdots\cdots\cdots\cdots \\
X_{k+1} &= \Delta_k S_{k+1} + (1+r)(X_k - \Delta_k S_k), \\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

过程$X = \{X_k : 0 \leqslant k \leqslant n\}$ 称为是初始投资为X_0, 投资策略为Δ 的自融资方式下的财富过程, 简称为财富过程. 上面表达式可以写成为

$$(1+r)^{-k-1}X_{k+1} - (1+r)^{-k}X_k = \Delta_k((1+r)^{-k-1}S_{k+1} - (1+r)^{-k}S_k).$$

因此, 折现了的财富过程是投资策略关于折现了的股票价格的随机积分.

我们说一个市场有套利机会, 是指存在一个投资策略使人可以无风险地(无论股票市场任何变化)从一无所有而赚到钱, 否则我们说市场没有套利机会. 比如说如果一个银行的存款利率高于另一个银行的贷款利率, 那么这里就存在套利机会(假设不考虑交易成本), 因为你可以从后一个银行贷款存入前一个银行而获得无风险的利率差额. 合理的假设是, 一个成熟的市场没有套利机会. 实际上, 市场有时的确有套利机会, 但人们的趋利性会使市场上的任何套利机会迅速消失. 现在我们给套利这个概念一个严格的数学定义.

定义3.2.2 一个市场存在套利是指存在一个可许的投资策略Δ 使得其初始投资$X_0 = 0$, 自融资方式下的财富过程X 满足$X_n \geqslant 0$ 且$\mathbb{P}(X_n > 0) > 0$.

首先我们证明, 假设市场无套利, 则必有

$$d < 1 + r < u.$$

事实上, 如果$1+r \geqslant u$, 也就是说利率过高, 那么人们可以先卖出(买空) 1 股股票, 获得资金S_0, 存入银行, 在下一时刻取出存款买回股票, 获利为$(1+r)S_0 - S_1$, 因为$1+r \geqslant u > d$, 所以此随机变量是非负的, 而且以正概率大于0, 因此市场存在套利. 同样我们可以证明如果利率过低$1+r \leqslant d$, 那么人们可以贷款去买股票而获得无风险收益, 故市场也有套利机会.

金融市场中还存在大量的衍生证券, 也称为金融产品, 它们的存在极大地活跃了金融市场. 衍生证券是依附于股票价格的合约. 让我们以期权为例来说明. 欧式期权是指在商定的时刻可以以商定的价格(不管此时刻股票的真正价格)来购买股票的一份合约. 期权是个权利, 但不是义务(不象期货的合约必须要被执行). 显然, 一个在商定时刻m 以商定价格K 购买股票的欧式期权在时刻m 的价值为

$$V_m := (S_m - K)^+,$$

也就是说如果到商定的时刻m, 股票价格S_m 高于商定价格K, 那么执行合约, 以价K 购买市场价为S_m 的股票而获利$S_m - K$, 如果此时股票价格S_m 低于商定价格K 而使得合约无利可图, 那么选择放弃执行合约, 获利为0. V_m 是S_m 的函数, 是$\mathcal{F}_m$ 可测的随机变量. 市场无套利推出期权不能是无偿的, 期权定价问题就是这个权利在时刻0 价值几何, 应该卖多少钱?

为了说明问题的本质, 让我们设$m = 1$. 注意到期权的购买者的动机是基于自己对股票价格走势的判断, 也就是说用购买期权获得的收益可以通过市场通常投资行为获得, 这时我们说期权可以被复制(replicate), 而出卖期权的人也认为出卖期权的风险可以通过市场通常投资行为来规避, 这时我们说期权可以对冲(hedge). 我们就基于这个思想来确定期权在0 时刻的价值V_0.

假设有资金V_0, 我们有两种方法进入市场(1) 购买期权; (2) 通常投资. 购买期权后在时刻1 的价值为V_1, 如果通常投资我们需要决定购买(或卖)多少股票, 然后把剩余的资金存入银行(或不够的话从银行贷款), 假设购买Δ_0 股(此数为负意味着卖出股票), 然后将资金$V_0 - \Delta_0 S_0$ 存入银行(此数为负意味着向银行贷款). 那么V_0, Δ_0 都是待定的. 通常投资方式在时刻1 的价值为

$$\Delta_0 S_1 + (1+r)(V_0 - \Delta_0 S_0),$$

第一部分是股票收益, 第二部分为利息.

大家都相信, 适当选取Δ_0 可以使得期权获利与通常投资获利一致,

$$V_1 = \Delta_0 S_1 + (1+r)(V_0 - \Delta_0 S_0).$$

注意V_1, S_1 是随机变量, 所以用此式直接解出Δ_0 是没有用的, 因为在时刻0 购买股票数的决定要依赖时刻1 的股票价格是可笑的. 但是我们可以把上式分两种情况考虑: 股票涨或者跌. 写$V_1^u := V_1|_{\xi_1=u} = (uS_0 - K)^+$, $V_1^d := V_1|_{\xi_1=d} = (dS_0 - K)^+$, 那么得到

$$V_1^u = \Delta_0 u S_0 + (1+r)(V_0 - \Delta_0 S_0), \tag{3.2.1}$$

$$V_1^d = \Delta_0 d S_0 + (1+r)(V_0 - \Delta_0 S_0), \tag{3.2.2}$$

相减得,

$$V_1^u - V_1^d = \Delta_0 S_0 (u-d),\ \Delta_0 = \frac{V_1^u - V_1^d}{uS_0 - dS_0}.$$

代入方程(3.2.1), 解得V_0 为

$$V_0 = \frac{1}{1+r}\left(\frac{1+r-d}{u-d}V_1^u + \frac{u-(1+r)}{u-d}V_1^d\right).$$

同时我们也找到了对冲这个期权的策略: 用资金V_0 开始, 购买Δ_0 股股票, 将剩余的资金存入银行. 另外可以证明如果期权的价格不是V_0 那么一定有套利机会存在. 事实上, 设市场上可以以价格为W_0 购买这样一份期权. 若售价过低$W_0 < V_0$, 则我可以从一无所有卖出Δ_0 股股票用所得的资金$\Delta_0 S_0$ 购买一份期权, 其余投入货币市场, 在下一时刻买回Δ_0 股, 那么我的获利为

$$V_1 + (1+r)(\Delta_0 S_0 - W_0) - \Delta_0 S_1 = (1+r)(V_0 - W_0) > 0,$$

即市场有套利机会, 同样可以证明若$W_0 > V_0$, 市场也存在套利机会. 也就是说, 一个无套利的市场中, 期权V_1 在时刻0 的价格一定是上面的V_0.

令人感到惊奇的是V_0 仅依赖于u, d, r, K, S_0, 而与市场中最无法捉摸的, 体现市场风险的概率p, q 无关. 因此我们说V_0 是风险中性的. 现在令$\overline{p} := \frac{1+r-d}{u-d}$, $\overline{q} := \frac{u-(1+r)}{u-d}$, 显然$\overline{p} + \overline{q} = 1$, 在市场无套利的假设下, $\overline{p}, \overline{q} > 0$. 因此在此假设下, 在空间$(\Omega, \mathcal{F}_n)$ 上存在概率$\overline{\mathbb{P}}$ 使得$\xi_1, \cdots, \xi_n$ 是独立同分布的, 但是分布为

$$\overline{\mathbb{P}}(\xi_1 = u) = \overline{p},\ \overline{\mathbb{P}}(\xi_1 = d) = \overline{q},$$

那么我们有

$$V_0 = \frac{1}{1+r}\overline{\mathbb{E}}V_1,$$

系数$\frac{1}{1+r}$ 是财富的折现率, 即由于货币市场的无风险利率, 下一个时刻单位1 的财富在现时刻的价值为$\frac{1}{1+r}$. 这样, 上式可以直观地解释为在新的概率下, 期权在现时刻的价值等于下一时刻其期望价值的折现.

概率$\overline{\mathbb{P}}$ 也称为市场的风险中性概率, 是一个非常有用的概率, 实际上是它决定了期权的定价, 为什么呢?

定理3.2.1 在概率$\overline{\mathbb{P}}$ 下, 经过折现的股票价格$\{(1+r)^{-k}S_k : k = 0, 1, 2, \cdots, n\}$ 是一个鞅.

证明. 证明是简单的, 因为ξ_{k+1} 与$\mathcal{F}_k$ 独立, 故有

$$\begin{aligned}
\overline{\mathbb{E}}((1+r)^{-k-1}S_{k+1}|\mathcal{F}_k) &= (1+r)^{-k}S_k\overline{\mathbb{E}}(\frac{1}{1+r}\xi_{k+1}) \\
&= (1+r)^{-k}S_k \cdot \frac{1}{1+r}\left(\frac{1+r-d}{u-d}u + \frac{u-(1+r)}{u-d}d\right) \\
&= (1+r)^{-k}S_k.
\end{aligned}$$

完成了证明. □

根据定理3.1.1, 我们立刻有下面的结论.

定理3.2.2 一个可许投资策略的自融资方式下的折现后的财富过程$\{(1+r)^{-k}X_k : 0 \leqslant k \leqslant n\}$ 在风险中性概率之下是一个鞅.

下面我们证明定理3.2.1 的逆也对.

定理3.2.3 市场无套利当且仅当存在一个等价于$\mathbb{P}$ 的概率$\overline{\mathbb{P}}$ 使得折现后的股票价格$\{(1+r)^{-k}S_k\}$ 在$\overline{\mathbb{P}}$ 之下是一个鞅.

证明. 只需证充分性, 如果折现后的股票价格在$\overline{\mathbb{P}}$ 之下是鞅, 则其任何自融资方式下的财富过程X 都是鞅, 因此$\overline{\mathbb{E}}X_n = (1+r)^n\overline{\mathbb{E}}X_0$. 若初始投资$X_0 = 0$, 则$\overline{\mathbb{E}}X_n = 0$, 因此$X_n \geqslant 0$ 蕴含有$\overline{\mathbb{P}}(X_n > 0) = 0$. 因为$\mathbb{P}$ 与$\overline{\mathbb{P}}$ 等价, 故推出$\mathbb{P}(X_n > 0) = 0$. 因此市场无套利. □

定义3.2.3 一个在时刻m 到期的简单欧式衍生证券是指一个$\mathcal{F}_m$ 可测的随机变量. 一个在时刻m 到期的简单欧式衍生证券V_m 称为可对冲(或者说复制)的, 是指存在一个常数V_0 与一个投资策略Δ 使其自融资方式下的财富过程X 满足$X_0 = V_0$, $X_m = V_m$. 一个市场是完备的是指任何简单欧式衍生证券都可对冲.

上面所讨论的欧式期权是典型的简单欧式衍生证券. 我们的目的是证明任何简单欧式衍生证券都可对冲, 而且我们可以具体构造出对冲的方法.

我们首先假设V_m 是一个可对冲的简单欧式衍生证券, 由定义存在一个常数V_0 与一个投资策略Δ 使其自融资方式下的财富过程X 满足$X_0 = V_0$, $X_m = V_m$. 设市场无套利, 那么在风险中性的概率$\overline{\mathbb{P}}$ 之下, $\{(1+r)^{-k}X_k\}$ 是鞅, 即

$$\begin{aligned}(1+r)^{-k}X_k &= \overline{\mathbb{E}}((1+r)^{-m}X_m|\mathcal{F}_k)\\ &= \overline{\mathbb{E}}((1+r)^{-m}V_m|\mathcal{F}_k),\\ X_k &= (1+r)^k\overline{\mathbb{E}}((1+r)^{-m}V_m|\mathcal{F}_k).\end{aligned}$$

而自融资性说明

$$X_{k+1} - (1+r)X_k = \Delta_k(S_{k+1} - (1+r)S_k).$$

更确切地

$$\begin{aligned}&X_{k+1}(\xi_1,\cdots,\xi_k,\xi_{k+1}) - (1+r)X_k(\xi_1,\cdots,\xi_k)\\ &\quad= \Delta_k(\xi_1,\cdots,\xi_k)(S_k\xi_{k+1} - (1+r)S_k),\end{aligned}$$

让ξ_{k+1} 分别取值u, d, 为了方便, 引入记号

$$\begin{aligned}X^u_{k+1} &:= X_{k+1}|_{\xi_{k+1}=u} = X_{k+1}(\xi_1,\xi_2,\cdots,\xi_k,u),\\ X^d_{k+1} &:= X_{k+1}|_{\xi_{k+1}=d} = X_{k+1}(\xi_1,\xi_2,\cdots,\xi_k,d),\end{aligned}$$

它们都是$\mathcal{F}_k$ 可测的, 那么我们得

$$\begin{aligned}X^u_{k+1} - (1+r)X_k &= \Delta_k(uS_k - (1+r)S_k),\\ X^d_{k+1} - (1+r)X_k &= \Delta_k(dS_k - (1+r)S_k),\end{aligned}$$

解出

$$\Delta_k = \frac{X^u_{k+1} - X^d_{k+1}}{(u-d)S_k}.$$

这证明了若V_m 可对冲, 则初始投资$V_0 = (1+r)^{-m}\overline{\mathbb{E}}V_m$, 投资策略$\Delta$ 如上.

上面的讨论为欧式简单衍生证券论证提供了一个构造对冲的思路. 设V_m 是一个欧式简单衍生证券, 先定义

$$V_k := (1+r)^k \overline{\mathbb{E}}((1+r)^{-m} V_m | \mathcal{F}_k),\ k = 0, 1, \cdots, m-1.$$

当然V_k 是$\xi_1, \cdots, \xi_k$ 的函数, 类似地记

$$\begin{aligned} V_{k+1}^u &:= V_{k+1}|_{\xi_{k+1}=u} = V_{k+1}(\xi_1, \xi_2, \cdots, \xi_k, u), \\ V_{k+1}^d &:= V_{k+1}|_{\xi_{k+1}=d} = V_{k+1}(\xi_1, \xi_2, \cdots, \xi_k, d), \end{aligned}$$

定义

$$\Delta_k := \frac{V_{k+1}^u - V_{k+1}^d}{(u-d)S_k},\ k = 0, 1, \cdots, m-1. \tag{3.2.3}$$

显然$\Delta = \{\Delta_0, \Delta_1, \cdots, \Delta_{m-1}\}$ 是一个投资策略. 下面我们证明初始投资为V_0 投资策略为Δ 的自融资方式下的财富过程$\{X_0, X_1, \cdots, X_m\}$ 满足

$$X_k = V_k,\ k = 0, 1, 2, \cdots, m,$$

恰好对冲V_m.

事实上, 用归纳法, 显然$V_0 = X_0$. 设$X_k = V_k$, 我们证明$X_{k+1} = V_{k+1}$. 首先因为$\{(1+r)^{-k} V_k\}$ 是$\overline{\mathbb{P}}$-鞅, 故

$$V_k = \frac{1}{1+r} \overline{\mathbb{E}}(V_{k+1} | \mathcal{F}_k) = \frac{1}{1+r}(\overline{p} V_{k+1}^u + \overline{q} V_{k+1}^d),$$

由财富过程的定义

$$\begin{aligned} X_{k+1} &= (1+r)X_k + \Delta_k(S_{k+1} - (1+r)S_k) \\ &= (1+r)V_k + \frac{V_{k+1}^u - V_{k+1}^d}{(u-d)S_k}(S_k \xi_{k+1} - (1+r)S_k) \\ &= \overline{p} V_{k+1}^u + \overline{q} V_{k+1}^d + \frac{V_{k+1}^u - V_{k+1}^d}{(u-d)}(\xi_{k+1} - (1+r)) \\ &= (\overline{p} + \frac{\xi_{k+1} - (1+r)}{u-d}) V_{k+1}^u + (\overline{q} - \frac{\xi_{k+1} - (1+r)}{u-d}) V_{k+1}^d \\ &= \frac{\xi_{k+1} - d}{u-d} V_{k+1}^u + \frac{u - \xi_{k+1}}{u-d} V_{k+1}^d \\ &= 1_{\{\xi_{k+1}=u\}} V_{k+1}^u + 1_{\{\xi_{k+1}=d\}} V_{k+1}^d = V_{k+1}. \end{aligned}$$

这样我们证明了下面的定理.

定理3.2.4 一个无套利市场是完备的, 且一个m 时刻到期的简单欧式衍生证券V_m在0 时刻的价值是

$$V_0 = \frac{1}{(1+r)^m}\mathbb{E}V_m,$$

它可以以初始投资V_0 投资策略为Δ 的自融资方式来对冲, 其中Δ_k 由(3.2.3) 定义.

证明. 我们只需证明如果V_m 在0 时刻的售价不是V_0, 那么市场就有套利机会. 假设售价$W_0 > V_0$, 则卖出一份衍生证券获得资金W_0, 然后用其中的V_0 开始用以上给出的投资策略, 那么无论市场如何变动, 最终总可以对冲卖出的衍生证券, 因此获得一份无风险的收益$W_0 - V_0 > 0$, 故市场有套利. 类似可以证明当售价$W_0 < V_0$ 时市场也有套利机会. 完成证明. □

上面我们简单地介绍了离散金融市场的理论, 有兴趣的读者可参考相关文献. 应该注意的是这只是金融市场的一种假设性模型, 它只可以帮助解释金融市场的某些现象, 距离真正的金融市场的应用还差得很远.

习 题

1. 一个袋子中在时刻0 有一个红球与一个白球. 随机地从袋子中取一个球, 然后将它放回并放入一个相同颜色的球, 无限地重复此过程. 记X_n 为n 次后袋中白球数与总球数之比. 证明: $\{X_n\}$ 是鞅.

2. 证明均值为零的平方可积鞅$\{X_n\}$ 是正交增量过程: 即对任何$n_1 < n_2 \leqslant n_3 < n_4$ 有
$$\mathbb{E}(X_{n_4} - X_{n_3})(X_{n_2} - X_{n_1}) = 0.$$

3. 设$\{X_n : n \geqslant 0\}$ 是$(\mathcal{F}_n)$ 适应的可积随机序列, 满足
$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \alpha X_n + \beta X_{n-1},\ n \geqslant 1,$$
其中$\alpha > 0, \beta > 0, \alpha + \beta = 1$. 问$a$ 为何值时, 序列$Y_0 := X_0$, $Y_n := aX_n + X_{n-1}$ 是$(\mathcal{F}_n)$ 鞅.

4. 设$\{Y_n\}$ 是独立同分布正随机变量序列使得$\mathbb{E}Y_n = 1$. 记$X_n := Y_1 Y_2 \cdots Y_n$.

   (a) 证明: (X_n) 是鞅且几乎处处收敛于一个随机变量X;

(b) 设Y_n 以概率$\frac{1}{2}$ 分别取值$\frac{1}{2}$ 与$\frac{3}{2}$. 验证$X = 0$ a.s. 因此$\mathbb{E}\prod_{n\geqslant 1} Y_n \neq \prod_{n\geqslant 1}\mathbb{E}Y_n$.

5. 设$\{X_n : n \geqslant 1\}$ 是一个均值0 且平方可积的鞅. 验证: $\mathbb{E}[(X_{n+k} - X_n)^2] = \sum_{i=1}^{k}\mathbb{E}[(X_{n+i} - X_{n+i-1})^2]$. 证明: 如果$\sum_n \mathbb{E}[(X_{n+1} - X_n)^2] < \infty$, 那么$X_n$ 以概率1 收敛.

6. 设$\{X_n\}$ 是平方可积鞅且$\sup_n \mathbb{E}X_n^2 < \infty$, 证明: X_n 平方收敛.

7. 证明Wald 定理: 设ξ_n 是独立同分布随机序列, $\mathbb{E}\xi_n = 0$, $\mathbb{E}\xi_n^2 = 1$, 令$X_n := \sum_{i=1}^{n}\xi_i$, T 是$\{X_n\}$ 的可积停时, 则$\mathbb{E}X_T^2 = \mathbb{E}T$.

8. 设$\{X_n\}$ 是鞅且$|X_0(\omega)|$, $|X_n(\omega) - X_{n-1}(\omega)|$ 被一个与ω 及n 无关的常数控制. 如果τ 是一个具有限均值的停时, 证明: X_τ 可积且$\mathbb{E}X_\tau = \mathbb{E}X_0$.

9. 设$\{X_n : n \geqslant 1\}$ 是鞅, 定义$\xi_n = X_n - X_{n-1}$, $\xi_1 = X_1$, 证明: $DX_n = \sum_{i=1}^{n} D\xi_i$.

10. 证明: 持续时间的母函数$\mathbb{E}^x t^T$ 是t 的有理函数.

11. 直接地证明: 输光概率q_x 满足差分方程

$$\begin{cases} q_x = pq_{x+1} + qq_{x-1}, & 0 < x < a \\ q_0 = 1, q_a = 0. \end{cases}$$

12. 直接地证明: 持续时间T 的期望D_x 满足差分方程

$$\begin{cases} D_x = 1 + pD_{x+1} + qD_{x-1}, & 0 < x < a, \\ D_0 = D_a = 0. \end{cases}$$

13. 证明: 股票价格过程$\{S_n\}$ 是一个Markov 过程.

14. 设$u = 2, d = \frac{1}{2}, S_0 = 4, r = \frac{1}{4}$. 一份欧式期权$(S_2 - 5)^+$ 在0 时刻价值多少? 再问一份简单欧式衍生证券$\max_{0\leqslant k\leqslant 2}(S_k - 5)^+$ 又价值多少?

第四章　随机过程的构造

§4.1　测度扩张定理

首先介绍测度构造理论, 为了方便, 我们重复概率空间的定义.

定义4.1.1 设Ω 是非空集合, 一个子集类$\mathcal{F}$ 称为是Ω 上的一个σ-域(或σ-代数, 事件域等), 如果

(1) $\varnothing, \Omega \in \mathcal{F}$;

(2) $A \in \mathcal{F}$ 蕴含$A^c \in \mathcal{F}$;

(3) $A_n \in \mathcal{F}, n \geqslant 1$ 蕴含$\bigcup_n A_n \in \mathcal{F}$.

称$(\Omega, \mathcal{F})$ 是一个可测空间. $\mathcal{F}$ 上的非负集函数$\mathbb{P}$ 称为是测度, 如果

(1) $\mathbb{P}(\emptyset) = 0$;

(2) 对如何互不相容的集列$\{A_n : n \geqslant 1\} \subset \mathcal{F}$, 有

$$\mathbb{P}(\bigcup_{n\geqslant 1} A_n) = \sum_{n\geqslant 1} \mathbb{P}(A_n).$$

这时, $(\Omega, \mathcal{F}, \mathbb{P})$ 称为是一个测度空间. 进一步, 如果$\mathbb{P}(\Omega) = 1$, 那么我们说$\mathbb{P}$ 是概率测度, 简称概率.

概率测度无非是一种特殊的测度而已, 前面很多对概率测度叙述并证明的结果并没有用到$\mathbb{P}(\Omega) = 1$, 因此它们对于一般的测度也是对的, 需要的时候我们将特别地指出.

古典概率所涉及的概率空间通常是简单的, 我们甚至可以把样本空间列举出来, 但一般的概率空间是相当复杂的, 要列举其σ-域和概率测度几乎是不可能的, 这里导致复杂的主要原因是可列运算和测度的可列可加性的要求. 所以我们先引入一个简单的子集类, 一个子集类$\mathcal{A}$ 称为是Ω 上的域, 如果它包含空集和Ω, 对补集运算封闭且对有限并运算封闭. 域与σ-域的定义看上去差别并不大, 一个对有限并封闭而另一个对可列并封闭, 但实际上域的结构要比σ-域的结构简单许多. 因此我们的思路是在域上构造一个我们所期望的概率测度然后扩张到域生成

的σ-域上. 这本质上就是实变函数课程中介绍的所谓Caratheodory 的测度构造理论.

把测度从域扩张到σ-域的一个桥梁是外测度, 外测度虽然不是测度, 但是它有许多优点.

定义4.1.2 定义在Ω 的幂集(或者所有子集)上的非负集函数μ 称为是Ω 上的一个外测度, 如果

(1) 单调性: 对任何$A, B \subset \Omega$, 如果$A \subset B$, 那么$\mu(A) \leqslant \mu(B)$;

(2) 次可列可加: 对任何$A_n \subset \Omega$, $n \geqslant 1$, 有$\mu(\bigcup_{n\geqslant 1} A_n) \leqslant \sum_{n\geqslant 1} \mu(A_n)$;

(3) $\mu(\varnothing) = 0$.

外测度的第一个优点是它很容易产生, 下面引理说明任何一个子集类上的非负集函数都可以诱导一个外测度.

引理4.1.1 设$\mathcal{A}$ 是Ω 的一个包含空集的子集类, μ 是$\mathcal{A}$ 上的非负集函数满足$\mu(\varnothing) = 0$. 对任何$A \subset \Omega$, 定义

$$\mu^*(A) := \inf\Big\{\sum_{n\geqslant 1} \mu(A_n) : \{A_n\} \subset \mathcal{A}, \bigcup_{n\geqslant 1} A_n \supset A\Big\}, \ \inf \emptyset := \infty,$$

则μ^* 是一个外测度. 称为是由$(\mu, \mathcal{A})$ 诱导的外测度.

证明. 非负性是显然的, 单调性来自于下确界的定义. 关键是证明次可列可加性. 任意给定子集列$\{B_n\}$, 要证明$\mu^*(\bigcup_{n\geqslant 1} B_n) \leqslant \sum_{n\geqslant 1} \mu^*(B_n)$. 如果右边有一个是无穷, 那么不等式显然成立, 所以我们不妨设右边每一个都是有限的. 对任何$\epsilon > 0$, 存在$A_{n,k} \in \mathcal{A}$ 使得$\bigcup_{k\geqslant 1} A_{n,k} \supset B_n$ 且

$$\sum_{k\geqslant 1} \mu(A_{n,k}) < \mu^*(B_n) + \epsilon/2^n.$$

那么$\bigcup_{n,k} A_{n,k} \supset \bigcup_{n\geqslant 1} B_n$, 因此$\sum_{n,k} \mu(A_{n,k}) \geqslant \mu^*(\bigcup_n B_n)$. 由此推出

$$\sum_n \mu^*(B_n) + \epsilon \geqslant \mu^*(\bigcup_n B_n),$$

最后ϵ 的任意性蕴涵着次可列可加. □

外测度的第二个优点是它总是可以得到一个测度. 设μ^* 是外测度, 给定$A \subset \Omega$. 如果对任何$E \subset \Omega$ 有

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c),$$

那么我们说A 是μ^*-可测集. 注意由于外测度的次可加性, 上面的等号等价于$\geqslant$号. 用$\mathcal{M}^*$ 表示μ^*-可测集全体.

引理4.1.2 如果μ^* 是外测度, 那么$(\Omega, \mathcal{M}^*, \mu^*)$ 是测度空间.

证明. 首先证明$\mathcal{M}^*$ 是σ-域. 很显然它包含空集和Ω 并且对补运算封闭, 关键是验证对可列并运算封闭. 先验证$\mathcal{M}^*$ 对有限并运算封闭. 若$A, B \in \mathcal{M}^*$, 则对$E \subset \Omega$,

$$\begin{aligned}
\mu^*(E) &= \mu^*(A \cap E) + \mu^*(A^c \cap E) \\
&= \mu^*(A \cap E) + [\mu^*(B \cap A^c \cap E) + \mu^*(B^c \cap A^c \cap E)] \\
&= [\mu^*(A \cap (A \cup B) \cap E) + \mu^*((A \cup B) \cap A^c \cap E)] \\
&\qquad + \mu^*((B \cup A)^c \cap E) \\
&= \mu^*((A \cup B) \cap E) + \mu^*((B \cup A)^c \cap E),
\end{aligned}$$

推出$A \cup B \in \mathcal{M}^*$. 这蕴涵着$\mathcal{M}^*$ 对有限交运算也封闭. 故下面我们仅须验证$\mathcal{M}^*$ 对不相交集列的可列并运算封闭. 设$\{A_n\}$ 是$\mathcal{M}^*$ 中不相交集列, 令

$$B_n := \bigcup_{i=1}^{n} A_i, \ \ A := \bigcup_{n=1}^{\infty} A_n,$$

由外测度的单调性得

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \\
&\geqslant \mu^*(E \cap B_n) + \mu^*(E \cap A^c) \\
&= \mu^*(E \cap B_n \cap B_{n-1}) + \mu^*(E \cap B_n \cap B_{n-1}^c) + \mu^*(E \cap A^c) \\
&= \mu^*(E \cap B_{n-1}) + \mu^*(E \cap A_n) + \mu^*(E \cap A^c) \\
&= \cdots\cdots = \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \cap A^c).
\end{aligned}$$

由n 的任意性与μ^* 的次可列可加性推出

$$\begin{aligned}
\mu^*(E) &\geqslant \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \\
&\geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c),
\end{aligned}$$

因此$A \in \mathcal{M}^*$. 不仅如此, 从上面的证明过程中可以看出对任何$E \subset \Omega$,

$$\mu^*\left(E \cap (\bigcup_{i=1}^{n} A_i)\right) = \sum_{i=1}^{n} \mu^*(E \cap A_i),$$

故推出μ^* 有有限可加性

$$\mu^*\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu^*(A_i).$$

结合μ^* 的次可列可加性, 推出μ^* 在$\mathcal{M}^*$ 上有可列可加性. 因此μ^* 是$(\Omega, \mathcal{M}^*)$ 上的测度. □

子集类$\mathcal{A}$ 上的非负集函数μ 诱导外测度μ^*, 但一般地, μ^* 和μ 在$\mathcal{A}$ 是不相同的, 那么问题就是什么时候它们是相同的? 也就是说什么时候μ^* 是μ 的扩张? 答案是可列可加性.

定义4.1.3 我们称定义在域$\mathcal{A}$ 上的非负集函数μ 是可加的, 如果它满足(1) $\mu(\emptyset) = 0$; (2) 对任何$\mathcal{A}$ 中互斥的A, B 有$\mu(A \cup B) = \mu(A) + \mu(B)$. 如果它还满足: 对任何$\mathcal{A}$ 中互不相容且并在$\mathcal{A}$ 中的集列有可列可加性, 即若$\{A_n : n \geqslant 1\} \subset \mathcal{A}$ 不相容且$\bigcup_n A_n \in \mathcal{A}$, 有

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n),$$

那么称它是预测度.

显然σ-域上的预测度就是测度. 另外一个非负可加集函数成为预测度当且仅当它在$\emptyset$ 处上连续.

引理4.1.3 设μ 是域$\mathcal{A}$ 上的预测度, 那么(1) $\mu^*|_{\mathcal{A}} = \mu$; (2) $\mathcal{A} \subset \mathcal{M}^*$.

证明. 先证明对任何$A \in \mathcal{A}$, 有$\mu^*(A) = \mu(A)$. 由定义$\mu^*(A) \leqslant \mu(A)$ 是显然的. 要证不等式另一边, 我们设$\mu^*(A) < \infty$, 那么对任何$\epsilon > 0$, 存在覆盖A 的集列$\{A_n\} \subset \mathcal{A}$ 使得

$$\mu^*(A) + \epsilon \geqslant \sum_{n \geqslant 1} \mu(A_n) \geqslant \mu(A).$$

推出结论. 注意这只用到μ 在$\mathcal{A}$ 上的单调性和次可列可加性, 还没有用到预测度的全部.

下面证明$\mathcal{A} \subset \mathcal{M}^*$. 设$A \in \mathcal{A}$, 对任何$E \subset \Omega$, 不妨设$\mu^*(E) < \infty$, 则对任何$\epsilon > 0$, 存在子集列$\{A_n\} \subset \mathcal{A}$ 满足$\bigcup_n A_n \supset E$ 且$\sum_n \mu(A_n) < \mu^*(E) + \epsilon$, 因而

$$\begin{aligned}
&\mu^*(E \cap A) + \mu^*(E \cap A^c) \\
&\qquad \leqslant \mu^*((\bigcup_n A_n) \cap A) + \mu^*((\bigcup_n A_n) \cap A^c)
\end{aligned}$$

$$
\begin{aligned}
&= \mu^*(\bigcup_n (A_n \cap A)) + \mu^*(\bigcup_n (A_n \cap A^c)) \\
&\leqslant \sum_n [\mu^*(A_n \cap A) + \mu^*(A_n \cap A^c)] \\
&\leqslant \sum_n [\mu(A_n \cap A) + \mu(A_n \cap A^c)] \\
&= \sum_n \mu(A_n) < \mu^*(E) + \epsilon,
\end{aligned}
$$

其中最后的等式来自μ 在$\mathcal{A}$ 上的有限可加性. 由此推出$\mu^*(E \cap A) + \mu^*(E \cap A^c) \leqslant \mu^*(E)$, 即$A \in \mathcal{M}^*$. □

现在我们先定义扩张的概念. 设$\mathcal{A}$ 是一个子集类, μ 是$\mathcal{A}$ 上一个集函数, 如果μ^* 是$\sigma(\mathcal{A})$ 上的测度且$\mu^*|_{\mathcal{A}} = \mu$, 那么我们说$\mu^*$ 是μ 的一个扩张. 现在可以来叙述Caratheodaory 的测度扩张定理. 由上面的引理容易推出, $\mathcal{M}^*$ 是一个包含$\mathcal{A}$ 的σ-域, 因此外测度μ^* 是μ 的一个扩张, 称为Caratheodory 扩张.

定理4.1.1 (Caratheodory) 设μ 是域$\mathcal{A}$ 上的预测度, 那么μ 有一个扩张.

自然地, 我们也要问什么时候扩张是唯一的. 下面定理说预概率测度的扩张一定是唯一的.

定理4.1.2 如果μ 是域$\mathcal{A}$ 上的预测度且$\mu(\Omega) = 1$, 那么μ 的扩张是唯一的.

证明. 如果μ 有两个扩张μ_1, μ_2, 定义

$$\mathcal{F} := \{A \subset \sigma(\mathcal{A}) : \ \mu_1(A) = \mu_2(A)\}.$$

那么$\mathcal{F} \supset \mathcal{A}$. 另外任意验证$\mathcal{F}$ 是Dynkin 类而$\mathcal{A}$ 对交封闭, 因此由Dynkin 定理得$\mathcal{F} \supset \sigma(\mathcal{A})$. 也就是说$\mu_1 = \mu_2$. □

这两个定理是说σ-域上的测度构造和域上测度的构造是等价的, 但是我们后面会体会到域的结构要简单许多, 因此域上测度的构造也要简单得多. 下面我们看看$[0,1]$ 上的Lebesgue 测度的存在性的严格证明.

设$\Omega = [0,1]$. 说$I \subset \Omega$ 是一个左开右闭子区间, 是指存在一个$\mathbf{R}$ 的左开右闭区间I' 使得$I = I' \cap \Omega$. 这里等价于I 是$\mathbf{R}$ 的左开右闭区间或者是包含0 的闭区间. 用$\mathcal{A}_0$ 表示Ω 的左开右闭子区间全体, $\mathcal{A}$, $\mathcal{F}$ 分别表示$\mathcal{A}_0$ 生成的域和σ-域. 那么$\mathcal{A}$ 是有限个不交左开右闭子区间并的全体(这只需验证这样的子集类是域就足够了), 而$\mathcal{F}$ 就是$[0,1]$ 上的Borel 域. 现在$\mathcal{A}$ 上有一个自然的集函数, 就是定义为

区间长度的和

$$\left|\bigcup_{i=1}^{n} I_i\right| := \sum_{i=1}^{n} |I_i|,$$

其中$\{I_i : 1 \leqslant i \leqslant n\} \subset \mathcal{A}_0$ 且互不相交. 下面的关键是证明这样定义的集函数是个预测度.

取不相容集列$\{A_n\} \subset \mathcal{A}$ 且$\bigcup_n A_n = A \in \mathcal{A}$, 我们要证明$\sum_n |A_n| = |A|$. 可以认为$A_n, A$ 都是区间, 比如$A_n = (a_n, b_n]$, $A = (a, b]$. 因为对任何n, $\bigcup_{i=1}^{n} A_i \subset A$, 故$\sum_{i \geqslant 1} |A_i| \leqslant |A|$ 是显然的. 反过来, 任取$\epsilon > 0$, 因为$\bigcup_n (a_n, b_n + \epsilon/2^n) \supset [a + \epsilon, b]$, 由有限覆盖定理, 存在$n$ 使得$\bigcup_{i=1}^{n} (a_i, b_i + \epsilon/2^i) \supset [a + \epsilon, b]$, 因此

$$\sum_{i=1}^{\infty} (b_i + \epsilon/2^i - a_i) \geqslant \sum_{i=1}^{n} (b_i + \epsilon/2^i - a_i) \geqslant b - a - \epsilon.$$

推出$\sum_{i=1}^{\infty} |A_i| \geqslant |A| - 2\epsilon$, 由$\epsilon$ 的任意性得知$\sum_{i=1}^{\infty} |A_i| \geqslant |A|$.

我们也能够证明多维分布函数是可以实现的. 让我们考虑二维情形作为例子, 更高维或者一维的情形类似. 首先什么函数可以称为是一个二维分布函数呢? 先假设有一个概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 和二维随机变量(X, Y). 令

$$F(x, y) := \mathbb{P}(X \leqslant x, Y \leqslant y), \;\; x, y \in \mathbf{R}.$$

那么它满足

(1) 它是右上连续的;

(2) 对任何$(x, y) \in \mathbf{R}^2$, $a, b \geqslant 0$, 定义

$$\square_{(x,y)}^{(x+a,y+b)} F := F(x + a, y + b) + F(x, y) - F(x + a, y) - F(x, y + b),$$

那么$\square_{(x,y)}^{(x+a,y+b)} F \geqslant 0$.

(3)

$$\lim_{x \text{ or } y \to -\infty} F(x, y) = 0, \;\; \lim_{x, y \to +\infty} F(x, y) = 1;$$

其中(2) 是本质的, 它比F 依分量单调递增要强, 它的成立是因为

$$\mathbb{P}\left((X, Y) \in (x, x + a] \times (y, y + b]\right) = \square_{(x,y)}^{(x+a,y+b)} F.$$

我们把满足上面三点的函数F 称为二维分布函数.

给定一个二维分布函数F. 用$\mathcal{A}_0$ 表示$\mathbf{R}^2$ 上矩形$(x,x+a]\times(y,y+b]$ 的全体, 定义

$$\mu((x,x+a]\times(y,y+b]):=\square_{(x,y)}^{(x+a,y+b)}F.$$

用$\mathcal{A}$, $\mathcal{B}^2$ 分别表示$\mathcal{A}_0$ 生成的域和σ-域. 同样, $\mathcal{A}$ 中的元素可以表示为有限个不交矩形的并, 故μ 的定义很容易延拓到$\mathcal{A}$ 上. 类似地, 要扩张到$\mathcal{B}^2$ 上的关键是证明这样的μ 是预概率测度.

证明和$[0,1]$ 上的类似, 取不相容集列$\{A_n\}\subset\mathcal{A}$ 且$\bigcup_n A_n=A\in\mathcal{A}$, 我们要证明$\sum_n\mu(A_n)=\mu(A)$. 可以认为$A_n, A$ 都是矩形, 比如

$$A_n=(x_n,x_n+a_n]\times(y_n,y_n+b_n],\ A=(x,x+a]\times(y,y+b].$$

因为对任何n, $\bigcup_{i=1}^n A_i\subset A$, 故$\sum_{i\geqslant1}\mu(A_i)\leqslant\mu(A)$ 是显然的. 反过来, 任取$\epsilon>0$, 由右上连续性, 对任何n, 存在$\delta_n>0$ 使得

$$F(x_n+a_n+\delta_n,y_n+b_n+\delta_n)-F(x_n+a_n,y_n+b_n)<\epsilon/2^n,$$

且存在$\delta>0$ 使得$F(x+\delta,y+\delta)-F(x,y)<\epsilon$. 这蕴涵着

$$\begin{aligned}&0\leqslant\square_{(x_n,y_n)}^{(x_n+a_n+\delta_n,y_n+b_n+\delta_n)}F-\mu(A_n)<\epsilon/2^n;\\&0\leqslant\mu(A)-\square_{(x+\delta,y+\delta)}^{(x+a,y+b)}F<\epsilon.\end{aligned}$$

因为$\{(x_n,x_n+a_n+\delta_n)\times(y_n,y_n+b_n+\delta_n):n\geqslant1\}$ 覆盖$[x+\delta,x+a]\times[y+\delta,y+b]$, 由有限覆盖定理, 存在$n$ 使得$\{(x_i,x_i+a_i+\delta_i)\times(y_i,y_i+b_i+\delta_i):1\leqslant i\leqslant n\}$ 足以覆盖, 因此$\{(x_i,x_i+a_i+\delta_i]\times(y_i,y_i+b_i+\delta_i]:1\leqslant i\leqslant n\}$ 覆盖了$(x+\delta,x+a]\times(y+\delta,y+b]$, 由次可加性推出

$$\sum_{i=1}^{\infty}\mu((x_i,x_i+a_i+\delta_i]\times(y_i,y_i+b_i+\delta_i])\geqslant\mu((x+\delta,x+a]\times(y+\delta,y+b]).$$

由上面的右上连续性得$\sum_{i=1}^\infty\mu(A_i)\geqslant\mu(A)-2\epsilon$, 由$\epsilon$ 的任意性得知$\sum_{i=1}^\infty\mu(A_i)\geqslant\mu(A)$.

§4.2 Kolmogorov 相容定理

上面的讨论说明, 一个有限维Euclid 空间上的分布函数总是可以实现的. 现在我们将讨论随机过程的一些基本问题, 以把理论建立在一个坚实的基础之上.

首先观察随机变量或者随机向量, 它们实际上是满足某些条件的Ω 到Euclid 空间的映射而已. 在一些实际的问题中我们需要考虑更一般的空间, 这样的推广显然没有本质上的困难, 下面的定义也是自然的.

定义4.2.1 设$(\Omega,\mathcal{F})$ 和$(E,\mathcal{E})$ 是可测空间, 一个映射$\xi:\Omega\longrightarrow E$ 称为可测映射, 如果$\xi^{-1}(\mathcal{E})\subset\mathcal{F}$. 如果需要, 那么我们指明是$\mathcal{F}/\mathcal{E}$-可测映射.

随机过程的定义很简单, 它是一个概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 到一个可测空间$(E,\mathcal{E})$ 的一族有序排列的可测映射, $\{\xi_t:t\in T\}$, 其中T 是一个全序集. 这时候, E 称为随机过程的状态空间(或相空间), T 称为是随机过程的时间或指标集. 当T 是$\mathbf{Z}$ 的子集时, 通常称为随机序列. 前面讨论的随机游动和鞅都是随机序列, 都是在一个给定概率空间上满足一定条件的随机序列. 从一个已知的随机序列出发, 可以得到新的随机序列.

例4.2.1 设F 是$\mathbf{R}$ 上分布函数, 那么存在概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 和独立随机序列$\{\xi_n\}$ 且它们的分布函数都是F. 进一步地, 定义$S_1=\xi_1,\cdots,S_n:=\sum_{i=1}^n\xi_i$, 那么$\{S_n\}$ 也是随机序列, 它也称为$\mathbf{R}$ 上的一个随机游动. 如果F 的期望等于0, 则$\{S_n\}$ 是一个鞅列. ∎

在一些简单情况下, 类似于§2.2 中那样可以证明独立随机序列的存在性, 但在许多情况下的随机过程不能这样简单地利用独立随机序列来构成, 让我们先考察下面的模型.

例4.2.2 蚂蚁迷宫: 设$G=(V,E)$ 是一个无向图, V, E 分别是顶点和边的集合, 假设它是局部有限的, 即每个顶点只有有限多个邻居. 现在有一个蚂蚁从某个顶点出发在G 上游走, 规则是这样的: (1) *它总是沿着边走*; (2) *在到达某个顶点后, 它总是等可能地选取一个邻居走过去.* 是不是有一个概率空间可以把这个问题描述清楚? 比如说, 让ξ_n 表示n 时刻(我们总是把蚂蚁达到某个顶点的时间称为时刻) 蚂蚁的位置, 我们需要这样一个概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 来安放随机序列$\{\xi_n\}$, 使得它要满足

$$\mathbb{P}(\xi_{n+1}=y|\xi_n=x)=\begin{cases}\frac{1}{c(x)}; & \text{如果}y\text{ 是}x\text{ 的邻居;}\\ 0, & \text{否则,}\end{cases}$$

这里$c(x)$ 表示顶点x 的邻居个数. ∎

例4.2.3 随机带位: 设一个餐厅有编号为$1,2,3,\cdots$ 的无穷个桌子, 某个桌子可以坐无穷多个客人, 餐厅经理总是坐在没有客人的桌子中号码最小的桌子上, 当下

个客人来的时候, 餐厅按这样的规则带位: (1)客人总是被带到有人的桌子上; (2)客人被带到某个桌子的概率和桌子上已有的人数成比例. 我们是否可以用随机过程来描述这个模型? 例如让ξ_n 表示第n 个客人所带到的桌子号码, 同样地, 是否有一个概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 使得$\xi_1 = 1$,

$$\mathbb{P}(\xi_{n+1} = i|\xi_1, \cdots, \xi_n) = \begin{cases} \frac{|\{k:\xi_k=i\}|}{n+1}; & 若 i \leqslant \max(\xi_1, \cdots, \xi_n); \\ \frac{1}{n+1}; & 若 i = \max(\xi_1, \cdots, \xi_n) + 1; \\ 0; & 其它. \end{cases}$$

∎

上面的例子基本清楚地说明了下面我们要做的事情, 问题类似于§2.1 中所问的一个分布是否可以实现. 虽然例子中的模型是如此直观, 使我们无法怀疑这样的概率空间总是存在的. 但作为数学的一个部分, 我们需要严格地证明它. 为此, 先介绍解释一个随机过程所需要的基本语言.

首先引入象测度的概念, 它很简单, 实际上类似前面的分布函数. 如果$(\Omega, \mathcal{F})$ 上给定了一个概率测度$\mathbb{P}$, 那么利用可测映射ξ, 很容易在$(E, \mathcal{E})$ 上定义一个集函数μ_ξ 如下:

$$\mu_\xi(A) := \mathbb{P}(\xi^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : \xi(\omega) \in A\}), \ A \in \mathcal{E}.$$

利用逆象的性质, 容易验证μ_ξ 是一个概率测度, 它被称为$\mathbb{P}$ 在ξ 下的象测度, 或者ξ 的分布, 也记为$\xi\mathbb{P}$ 或者$\mathbb{P} \circ \xi^{-1}$, 形象地说, ξ 把$\mathbb{P}$ 诱导到$(E, \mathcal{E})$ 上. 不难看出, 如果η 是E 到另一个可测空间F 的可测映射, 那么

$$\mathbb{P} \circ (\eta \circ \xi)^{-1} = (\mathbb{P} \circ \xi^{-1}) \circ \eta^{-1}.$$

特别地, 设ξ 是随机变量, 那么μ_ξ 是$(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上的概率测度, 是ξ 的分布.

再介绍乘积可测空间的概念, 设$(E_1, \mathcal{E}_1)$, $(E_1, \mathcal{E}_2)$ 是可测空间, 对$A_1 \subset E_1$, $A_2 \subset E_2$, 记

$$A_1 \times A_2 := \{(x_1, x_2) : x_1 \in A_1, x_2 \in A_2\}.$$

它是乘积空间$E_1 \times E_2$ 上的矩形, 如果$A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$, 这个矩形被称为可测矩形. 显然可测矩形全体对交封闭, 是π 类, 但不是σ-域. 那么我们自然地取以可测矩形生成的σ-域, 记为$\mathcal{E}_1 \times \mathcal{E}_2$, 作为$E_1 \times E_2$ 上的自然的σ-域. 空间$(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2)$ 被

称为乘积(可测)空间. 定义映射

$$\pi_i(x_1, x_2) = x_i,\ i = 1, 2,$$

它们分别称为是乘积空间到E_1, E_2 上的投影. 乘积σ-域可以如下刻画.

引理4.2.1 σ-域$\mathcal{E}_1 \times \mathcal{E}_2$ 是$E_1 \times E_2$ 上使得两个投影可测的最小σ-域.

特别地, 我们考虑空间的自乘积, 首先容易定义可测空间$(E, \mathcal{E})$ 的n 次自乘积可测空间, 让我们记它为$(E^n, \mathcal{E}^n)$. 空间的一个点通常写为$(x_1, \cdots, x_n)$, 也就是坐标表示, 我们也可以把它看成是集合$[n] = \{1, 2, \cdots, n\}$ 到E 的一个映射$x : i \mapsto x_i$. 反过来, 一个这样的映射也可以写成乘积空间的一个点, 换句话说E^n 和$[n]$ 到E 的映射全体是一一对应的. 所以我们不区分乘积空间的一个点和它所对应的一个映射, 而乘积σ-域恰是使得所有投影可测的最小σ-域.

两种不同角度看同一个问题的好处是后者可以很容易被推广到无穷的情况. 上面的集合$[n]$ 并非本质的, 它可以被任何一个元素个数为n 的集合替代, 因为它只不过是用来标识空间中点的坐标顺序的. 所以任意取一个全序集合(实际上, 序并不是必须的) T, 用符号E^T 表示T 到E 的所有映射的集合, 也被称为E 的T 乘积空间, 特别地$E^n = E^{\{1,\cdots,n\}}$. 当T 是无限集, E^T 被称为无穷乘积空间, 最简单的无穷乘积空间是$E^{\mathbf{N}}$, 它是取值E 上的点列全体. 所以无穷乘积空间的概念并不陌生. 怎么来定义它上面的σ-域呢? 还是用可测矩形的方法或者说投影的方法. 在本章后面, 我们实际上只是考虑T 是$\mathbf{N}$ 的情况, 但在叙述下面的Kolmogorov 相容定理时, 两者没有本质的难度差别, 所以我们是对一般全序集T 论述的. 需要的话, 读者就把它当作$\mathbf{N}$ 来理解.

对映射$x : T \to E, t \in T$, 符号$x(t)$ 或x_t 是x 在时刻t 处的值或坐标. 任意给定T 的有限子集I, (后面我们写$I \subset T$ 时, 都指I 是有限子集.) 定义投影$\pi_I : E^T \longrightarrow E^I$,

$$\pi_I(x) := (x(t) : t \in I),\ x \in E^T.$$

简单地, $\pi_t := \pi_{\{t\}}$. 定义$\mathcal{E}^T$ 为E^T 上使得所有$\{\pi_t : t \in T\}$ 可测的最小σ-域. 这个定义类似于前面两个空间乘积的情形. 我们需要$\mathcal{E}^T$ 的一个等价刻画. 对任何$A \subset E^I$, 它在投影π_I 下的逆像$\pi_I^{-1}(A)$ 称为是A 决定的柱集. 这个名称的理由是平面上的一个图形在三维空间对其投影的逆像是一个柱体. (可测)柱集全体是指

$$\mathcal{E}_0^T := \bigcup_I \pi_I^{-1}(\mathcal{E}^I),$$

其中I 取遍T 的有限子集. 下面引理说明$\mathcal{E}^T$ 就是由柱集生成的σ-域.

引理4.2.2 $\mathcal{E}^T = \sigma(\mathcal{E}_0^T)$.

这样我们得到了一个无穷乘积可测空间$(E^T, \mathcal{E}^T)$ 且$\{\pi_t : t \in T\}$ 是以E 为状态空间的可测映射. 现在假设给定随机过程, 即概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 上以$(E, \mathcal{E})$ 为状态空间的随机过程$\{\xi_t : t \in T\}$. 定义

$$\xi(\omega) := (\xi_t(\omega) : t \in T),\ \omega \in \Omega.$$

显然$\xi(\omega) \in E^T$, 它被称为是ω 的(样本)轨道. 这个ξ 是Ω 到空间E^T 的映射. 下面的引理说明ξ 是可测映射.

引理4.2.3 上面定义的$\xi : \Omega \longrightarrow E^T$ 是可测映射.

用μ^ξ 表示$\mathbb{P}$ 在ξ 下的象概率测度, 它使得$(E^T, \mathcal{E}^T, \mu_\xi)$ 成为概率空间, 在这个概率空间看来$\{\pi_t : t \in T\}$ 是状态空间为E 的随机过程, 称为典则过程, 实际上就是随机过程的分布, 因为它与原来过程的关系就象通常的随机变量与分布的关系一样, 或者更象第一节中介绍的随机游动和格点轨道的关系.

有限维Euclid 空间上的分布函数是可以实现的, 具体如何实现在哪里实现并不重要, 比如0, 1 上的均匀分布, 可以用抛硬币实现, 可以摸球实现, 可以抽签实现, 也可以掷骰子实现. 上面的构造说明分布函数实际上可以在Euclid 空间自身上实现的, 分布函数的作用是表达所需要的性质的, 但真正能表达信息的不是任意的Borel 集而是简单的区间, 一般Borel 集上的性质是抽象的, 由测度的定义来演绎的, 也就是说区间是表达信息的媒介. 自然地对于随机过程, 我们应该在典则空间$(E^T, \mathcal{E}^T)$ 来构造概率实现. 那么具体用什么媒介来表达我们所需要的性质呢? 那就是柱集的集合和随机过程在其上的投影. 下面先证明柱集的集合是域.

引理4.2.4 如果I, J 都是T 的有限子集且$I \subset J$, 那么$\pi_I^{-1}(\mathcal{E}^I) \subset \pi_J^{-1}(\mathcal{E}^J)$, 故$\mathcal{E}_0^T$ 是一个域.

给定概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机过程$\{\xi_t : t \in T\}$ 和T 的有限子集I (我们总是假设其中元素按顺序排列), 定义μ^I 为$(\xi_t : t \in I)$ 的分布,

$$\mu^I(A) = \mathbb{P}((\xi_i : i \in I) \in A),\ A \in \mathcal{E}^I,$$

也就是$\mathbb{P}$ 在映射$\pi_I \circ \xi$ 下的象测度, 或μ^ξ 在π_I 下的象测度:

$$\mu^I = \mathbb{P} \circ (\pi_I \circ \xi)^{-1} = (\mathbb{P} \circ \xi^{-1}) \circ \pi_I^{-1} = \mu^\xi \circ \pi_I^{-1}.$$

它是$(E^I,\mathcal{E}^I)$ 上的概率测度. 称为$\mathbb{P}$ 的一个有限维分布或者$\mathbb{P}$ 在E^I 上的投影, 它体现了过程的部分信息. 概率空间的集合$\{(E^I,\mathcal{E}^I,\mu^I):\ I\subset T\}$ 称为是随机过程$\{\xi_t:t\in T\}$ 的有限维分布族. 有限维分布族就是随机过程的分布μ^ξ 在柱集$\mathcal{E}_0^T$ 上的表现, 相当于Lebesgue 测度在区间上的体现, 所以也就是随机变量的分布函数的角色. 形象地说, 随机过程的分布就是一幅太大的画, 无法在一张画布上表达出来, 有限维分布族就象一张张用来表达同一幅画的画布一样.

如同分布函数有递增右连续性一样, 有限维分布也有一个重要的性质: 相容性. 直观地说, 不同画布上的画须有某种相容性才能拼成一幅完整的画. 设I 和J 都是T 的有限子集且$I\subset J$, 定义映射$\pi_I^J:E^J\longrightarrow E^I$ 为

$$\pi_I^J(x_t:t\in J):=(x_t:t\in I).$$

它是空间E^J 到E^I 的投影. 显然$\pi_I^J\circ\pi_J=\pi_I$, 因此

$$\mu^J\circ(\pi_I^J)^{-1}=\mu^\xi\circ\pi_J^{-1}\circ(\pi_I^J)^{-1}=\mu^\xi\circ\pi_I^{-1}=\mu^I.$$

即μ^I 等于μ^J 在E^I 上的投影, 这说明如果I 是J 的子集, μ^I 是完全由μ^J 决定的. 这就是下面的引理.

引理4.2.5 随机序列的有限维分布族满足相容性: 设I 和J 都是T 的有限子集且$I\subset J$, 那么$\mu^J\circ(\pi_I^J)^{-1}=\mu^I$.

相容性是一个重要概念, 我们给它一个定义.

定义4.2.2 给定时间集T 和可测空间$(E,\mathcal{E})$, 如果对T 的任何有限子集I, μ^I 是$(E^I,\mathcal{E}^I)$ 上的概率, 那么称概率空间的集合$\{(E^I,\mathcal{E}^I,\mu^I):\ I\subset T\}$ 是空间E 和时间T 上的一个有限维分布族. 另外如果它们满足引理4.2.5中的相容性, 那么被称为是相容的有限维分布族.

综上所述, 概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 上以E 为状态空间的随机过程$\{\xi_t:t\in T\}$ 决定了一个相容的有限维分布族$\{(E^I,\mathcal{E}^I,\mu^I):\ I$ 是T 的有限子集$\}$. 现在可以说清楚关键的问题了, 假设给定一个相容的有限维分布族, 是否存在概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 上以E 为状态空间的随机序列$\{\xi_t:t\in T\}$ 使得它的有限维分布族就是给定的那个呢? 如果有, 就说这个相容的有限维分布族是可以实现的.

答案是肯定的, 如果E 是一个完备可分度量空间. 注意当我们谈论一个拓扑空间时, 如果没有其它说明, 自然地取$\mathcal{E}$ 是其上的Borel σ-域.

定理4.2.1 (Kolmogorov) 如果E 是完备可分度量空间, 那么它上面的相容概率空间列是可以实现的.

这是现代概率论中的最主要的定理之一. 有了上面的准备工作, 我们差不多可以证明这个定理了, 差的就是分析的一个重要定理, 我们在附录里给出其证明.

定理4.2.2 设E 是完备可分度量空间, 则E 上任何概率测度μ 是正则的: 对任何$A \in \mathcal{E}$, 有

$$\mu(A) = \sup\{\mu(K): \ K \subset A, \ K \text{ 紧}\}.$$

定理4.2.1的证明. 我们打算在可测空间$(E^T, \mathcal{E}^T)$ 上构造出所需要的概率测度μ, 它要满足对$I \subset T$ (有限子集, 下面类似), $\mu \circ \pi_I^{-1} = \mu^I$. 也就是说, 对任何$H \in \mathcal{E}^I$,

$$\mu(\pi_I^{-1}(H)) = \mu^I(H).$$

这立刻告诉我们μ 在柱集的域$\mathcal{E}_0^T$ 上应该怎么定义. 取$B \in \mathcal{E}_0^T$, 那么存在$I \subset T$ 和$H \in \mathcal{E}^I$ 使得$B = \pi_I^{-1}(H)$. 然后定义

$$\mu(B) := \mu^I(H).$$

但这定义至少看上去与柱集B 的表示有关, 因此为了定义的合理性, 必须证明$\mu(B)$ 不依赖于B 的表示. 换句话说, 如果有另外的$I' \subset T$ 和$H' \in \mathcal{E}^{I'}$ 使得$B = \pi_{I'}^{-1}(H')$, 那么$\mu^{I'}(H') = \mu^I(H)$. 事实上, 因为$\pi_I^{-1}(\mathcal{E}^I)$ 关于I 是递增的, 故不妨设$I \subset I'$, 因为$\pi_I = \pi_I^{I'} \circ \pi_{I'}$, 故

$$\pi_{I'}^{-1}((\pi_I^{I'})^{-1}(H)) = \pi_I^{-1}(H) = \pi_{I'}^{-1}(H'),$$

且因投影一定是满射, 推出$(\pi_I^{I'})^{-1}(H) = H'$. 然后相容性保证

$$\mu^I(H) = \mu^{I'}((\pi_I^{I'})^{-1}(H)) = \mu^{I'}(H').$$

这说明μ 是域$\mathcal{E}_0^T$ 上良定义的非负集函数, 它满足

(1) $\mu(E^T) = 1$; 事实上, 对任何t, $E^T = \pi_t^{-1}(E)$, 故$\mu(E^T) = \mu_{\{t\}}(E) = 1$.

(2) 如果$A, B \in \mathcal{E}_0^T$ 且$A \cap B = \emptyset$, 那么$\mu(A \cup B) = \mu(A) + \mu(B)$. 事实上, 存在$I \subset T$ 使得$A, B \in \pi_I^{-1}(\mathcal{E}^I)$, 即有$H, K \in \mathcal{E}^I$, 互斥, 使得$A = \pi_I^{-1}(H)$, $B = \pi_I^{-1}(K)$. 因此$\mu(A \cup B) = \mu(\pi_I^{-1}(H \cup K)) = \mu^I(H \cup K) = \mu^I(H) + \mu^I(K) = \mu(A) + \mu(B)$.

现在由定理4.1.1, 我们只需要证明μ 在空集处有上连续性就够了, 即我们要证明, 如果$\{A_n\}\subset\mathcal{E}_0^T$ 且$A_n\downarrow\emptyset$, 则$\mu(A_n)\downarrow 0$.

当然$\mu(A_n)$ 是递减的, 反设存在$\epsilon>0$ 使得$\mu(A_n)>\epsilon$. 首先存在$I_n\subset T$, $H_n\in\mathcal{E}^{I_n}$ 使得$A_n=\pi_{I_n}^{-1}(H_n)$. 还是因为$\pi_I^{-1}(\mathcal{E}^I)$ 关于I 递增, 所以不妨认为集列I_n 关于n 递增. 现在$A_n=\pi_{I_n}^{-1}(H_n)$ 其中$H_n\in\mathcal{E}^{I_n}$. 那么$\mu(A_n)=\mu^{I_n}(H_n)$. 因为完备可分度量空间的乘积仍然是完备可分度量空间, 故存在紧集$K_n\subset H_n$ 使得$\mu^{I_n}(H_n\setminus K_n)<\epsilon/2^n$. 令$B_n=\pi_{I_n}^{-1}(K_n)$, 它是包含在$A_n$ 中的柱集但未必关于n 递减. 因此再记$C_n:=\bigcap_{i=1}^n B_i$, 它是含在A_n 中的柱集且关于n 递减. 现在

$$\mu(A_n\setminus C_n)=\mu\left(\bigcup_{i=1}^n(A_n\setminus B_i)\right)\leqslant\sum_{i=1}^n\mu(A_i\setminus B_i)=\sum_{i=1}^n\mu^{I_i}(H_i\setminus K_i)<\epsilon.$$

这蕴涵着对任何n, $\mu(C_n)>0$, 推出C_n 非空. 取点$x^{(n)}\in C_n$ 组成E^T 的点列. 因为C_n 递减, 故对任何$m\leqslant n$, $C_n\subset C_m\subset B_m=\pi_{I_m}^{-1}(K_m)$, 这说明$\{(x^{(n)}(t):t\in I_m):n\geqslant m\}\subset K_m$. 因$K_m$ 紧, 故它有在K_m 中收敛的子列. 因此我们可以得到E 的一个点列$\{x(t):t\in\bigcup_n I_n\}$ 使得对任何m, $(x(t):t\in I_m)$ 是$\{(x^{(n)}(t):t\in I_m):n\geqslant m\}$ 的一个极限点, 推出$(x(t):t\in I_m)\in K_m$, 随意取点$e\in E$, 对$t\notin T\setminus\bigcup_n I_n$, 定义$x(t)=e$, 那么$x=(x(t):t\in T)\in E^T$, 而上面的结论说明$\pi_{I_m}(x)\in K_m$ 或者说$x\in\pi_{I_m}^{-1}(K_m)\subset A_m$, 这导致矛盾, 因为$x\in\bigcap_{n\geqslant 1}A_n$. □

显然, Euclid 空间是完备可分度量空间. 另外赋以离散拓扑的可列集也是完备可分度量空间, 但是离散拓扑的不可列集就不是可分的. 怎么具体来给定一个相容的有限维分布族呢? 首先在时间离散$T=\mathbf{N}$ 的情况下, 事情实际上简单得多, 令$\mathcal{E}_1^{\mathbf{N}}:=\bigcup_n\pi_{[n]}^{-1}(\mathcal{E}^{[n]})$. 显然$\mathcal{E}_1^{\mathbf{N}}\subset\mathcal{E}_0^{\mathbf{N}}$ 并且$\mathcal{E}^{\mathbf{N}}=\sigma(\mathcal{E}_1^{\mathbf{N}})$, 即柱集$\mathcal{E}_1^{\mathbf{N}}$ 就足够生成$\mathcal{E}^{\mathbf{N}}$ 了, 另外当有限维分布族$\{(E^I,\mathcal{E}^I,\mu^I):I\subset\mathbf{N}\}$ 相容时, 它由它的子集$\{(E^{[n]},\mathcal{E}^{[n]},\mu^{[n]}):n\in\mathbf{N}\}$ (简单地称它为有限维分布族的主子列)完全决定了, 因为任何有限子集I 总是某个$[n]$ 的子集.

概率空间列$\{(E^{[n]},\mathcal{E}^{[n]},\mu^{[n]}):n\in\mathbf{N}\}$ 称为相容的, 如果对任何$n\in\mathbf{N}$, $\mu^{[n+1]}\circ(\pi_{[n]}^{[n+1]})^{-1}=\mu^{[n]}$, 更具体地说, 对任何$A_1,\cdots,A_n\in\mathcal{E}$ 有

$$\mu^{[n+1]}(A_1\times A_2\times\cdots\times A_n\times E)=\mu^{[n]}(A_1\times A_2\times\cdots\times A_n).$$

很显然, 一个相容的有限维分布族中的主子列一定是相容的, 反过来, 从一个相容的概率空间列可以唯一构造一个相容的有限维分布族. 事实上, 对任何有限子集, 取n 充分大使得$I\subset[n]$, 定义$\mu^I:=\mu^{[n]}\circ(\pi_I^{[n]})^{-1}$, 由相容性推出定义与$n$ 的取法

无关, 容易验证这样定义的μ^I 构成一个相容的有限维分布族. 也就是说当时间集离散时, 相容的有限维分布族可以用更容易构造的相容概率空间列等价表示.

再假设E 是离散可列集. 给定一个随机序列$\{\xi_n : n \geqslant 0\}$, 由乘积公式推知

$$\begin{aligned}&\mathbb{P}(\xi_0 = x_0, \cdots, \xi_{n-1} = x_{n-1}, \xi_n = x_n)\\&= \mathbb{P}(\xi_0 = x_0)\prod_{i=1}^{n}\mathbb{P}(\xi_i = x_i | \xi_0 = x_0, \cdots, \xi_{i-1} = x_{i-1}),\end{aligned}$$

这说明联合分布是由ξ_0 的分布和条件概率决定的. 直观地, 我们会问过程在起始时刻在哪里? 当知道前n 个时刻过程的行为后, 过程下一刻又会怎么走? 习惯地, 一个随机序列的初始时刻通常记为0, 所以我们下面假设$\mathbf{N} = \{n : n \geqslant 0\}$. 反过来, 如果我们知道过程在0 时刻以概率$\mu_0(x)$ 位于x 点($x \in E$); 如果已知过程的n 时刻前的行为$\xi_i = x_i$, $0 \leqslant i < n$, 那么它在下一步以概率$p_n((x_0, \cdots, x_{n-1}), x)$ 移动到x 点($x \in E$), 那么这样应该足可以描述一个随机序列了. 让我们写下并证明这个结论. 下面说的分布律总是指一个和等于1 的非负数列.

推论4.2.1 设E 是离散可列集, $\{\mu_0(x) : x \in E\}$ 是E 上分布律, 对任何的$n \geqslant 1$ 及$\{x_i : 0 \leqslant i < n\} \subset E$, $\{p_n((x_0, \cdots, x_{n-1}), x) : x \in E\}$ 是E 上分布律. 那么存在概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上随机序列$\{\xi_n : n \in \mathbf{N}\}$ 使得$\mathbb{P}(\xi_0 = x) = \mu_0(x)$ 且(当下面条件概率有意义时)

$$\mathbb{P}(\xi_n = x | \xi_0 = x_0, \cdots, \xi_{n-1} = x_{n-1}) = p_n((x_0, \cdots, x_{n-1}), x),$$

它们分布称为过程的初始分布和转移概率.

证明. 定义$\mu^{[0]} = \mu_0$, 对任何$n \geqslant 1$, $x_i \in E$, $0 \leqslant i \leqslant n$, 归纳定义

$$\mu^{[n]}(x_0, \cdots, x_{n-1}, x_n) := \mu^{[n-1]}(x_0, \cdots, x_{n-1})p_n((x_0, \cdots, x_{n-1}), x_n).$$

容易验证$\mu^{[n]}$ 是概率测度且

$$\sum_{x \in E}\mu^{[n]}(x_0, \cdots, x_{n-1}, x) = \mu^{[n-1]}(x_0, \cdots, x_{n-1}).$$

因此$\{\mu^{[n]} : n \in \mathbf{N}\}$ 是显然概率测度列. 由定理4.2.1推出存在概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上随机序列$\{\xi_n : n \in \mathbf{N}\}$ 满足

$$\mathbb{P}(\xi_0 = x_0, \cdots, \xi_{n-1} = x_{n-1}, \xi_n = x_n) = \mu^{[n]}(x_0, \cdots, x_{n-1}, x_n).$$

由此容易推出上面的条件概率. □

再看例4.2.2 和例4.2.3, 随机过程的存在性是显然的. 对例4.2.2, 任意固定$a \in E$, 定义$\mu_0(x) = 1_{\{x=a\}}$,

$$p_n((x_0, \cdots, x_{n-1}), x) = \frac{1_{\{x \sim x_{n-1}\}}}{c(x_{n-1})},$$

其中$\sim$ 表示两者相邻. 对例4.2.3, 时间从1 开始, 定义$\mu_1(x) = 1_{\{x=1\}}$,

$$p_{n+1}((x_1, \cdots, x_n), x) = \begin{cases} \frac{1}{n+1}\sum_{i=1}^n 1_{\{x_i=x\}}, & x \leqslant \max(x_1, \cdots, x_n); \\ \frac{1}{n+1}, & x = \max(x_1, \cdots, x_n) + 1; \\ 0, & x > \max(x_1, \cdots, x_n) + 1. \end{cases}$$

容易验证两者都满足推论的条件.

附录4.2.1. 定理4.2.2由下面两个定理推出. 读者需要了解完备度量空间的一个性质: 完备度量空间的子集是相对紧的当且仅当它是完全有界的. 这个结论在泛函分析教科书开头可以找到.

定理4.2.3 设μ 是度量空间E 上的概率测度. 则对任何$A \in \mathcal{B}(E)$ 和$\epsilon > 0$, 存在开集G 和闭集F 使得$F \subset A \subset G$ 且$\mu(G \setminus F) < \epsilon$.

证明. 设d 是E 上度量, 如果A 闭, 那么取$F = A$, $G_n = \{x : d(x, A) < n^{-1}\}$. 则$G_n$ 开, 且$G_n \downarrow A$. 因此$\mu(G_n \setminus F) \downarrow 0$, 故结论对闭集成立. 现在我们需要证明满足条件的A 是一个σ-域就够了. 这不难验证, 作为习题. □

定理4.2.4 设μ 是完备可分度量空间E 上的概率, 则对任何$\epsilon > 0$, 存在紧集$K \subset E$, 使得$\mu(K) > 1 - \epsilon$.

证明. 由可分性, 对任何n, 存在半径为$1/n$ 的可列个球$\{A_{n,k} : k \geqslant 1\}$ 覆盖E. 那么

$$\lim_{i \to \infty} \mu\left(\bigcup_{i \geqslant k \geqslant 1} A_{n,k}\right) = 1,$$

故存在i_n 使得$\mu\left(\bigcup_{k=1}^{i_n} A_{n,k}\right) > 1 - \epsilon/2^n$. 令$A = \bigcap_{n \geqslant 1} \bigcup_{k=1}^{i_n} A_{n,k}$, 那么$A$ 是完全有界集且

$$\mu(A^c) \leqslant \sum_{n \geqslant 1} \mu\left\{\left(\bigcup_{k=1}^{i_n} A_{n,k}\right)^c\right\} < \epsilon,$$

用K 表示A 的闭包, 那么K 是紧集且$\mu(K) \geqslant \mu(A) > 1 - \epsilon$. □

习　　题

1. 设ϕ 是集合E' 到E 的满射, 证明: 对任何$A, B \subset E$, $\phi^{-1}(A) = \phi^{-1}(B)$ 蕴涵着$A = B$.

2. 证明: 一个非负可加集函数成为预测度当且仅当它在$\emptyset$ 处上连续.

3. 设E_1, E_2 是可分度量空间, 证明: $\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \times \mathcal{B}(E_2)$, 即Borel σ-域的乘积σ-域就是乘积空间上的Borel σ-域. 如果空间没有可分性, 结论是否成立?

4. 证明: 如果I, J 是有限子集且$I \subset J$, 那么$\pi_I^{-1}(\mathcal{E}^I) \subset \pi_J^{-1}(\mathcal{E}^J)$.

第五章　Markov 链

§5.1　Markov 链

设E 是一个有限集或者可列集, 自然地, E 上的拓扑是离散拓扑, σ-域就取E 的全体子集. 设$\{X_n : n \geqslant 0\}$ 是概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机序列, 那么在上一节中我们说过, 只要知道它在至时刻n 之前每个时刻的位置, 下一步的位置由一个分布律决定, 就是条件分布律$\{\mathbb{P}(X_{n+1} = y|X_i = x_i, 0 \leqslant i \leqslant n) : y \in E\}$. 按照时间的序, 如果我们把一个时刻$n$ 称为现在, 那么小于n 的称为过去, 大于n 的称为将来. 换句话说, 将来的行为规则由过去和现在共同决定.

但在许多随机模型中, 只要知道现在的位置, 下一步的运动规则就完全确定了. 例如在例4.2.2中, 蚂蚁下一步怎么走只与它现在所处的位置有关, 与它前面怎么走到现在的位置无关. (而例4.2.3没有这样的性质.) §3.1中讨论的随机游动显然也有这样的性质. 这个性质首先由Markov 抽象出来, 故称为Markov 性.

定义5.1.1 随机序列$\{X_n\}$ 满足Markov 性, 是指对任何时刻n 和$y \in E$,

$$\mathbb{P}(X_{n+1} = y|X_0, \cdots, X_n) = \mathbb{P}(X_{n+1} = x|X_n).$$

由条件期望的性质, 这等价于对任何$x_0, \cdots, x_n \in E$ 有

$$\mathbb{P}(X_{n+1} = y|X_i = x_i, 0 \leqslant i \leqslant n) = \mathbb{P}(X_{n+1} = y|X_n = x_n).$$

也就是说条件分布律只依赖于现在的位置x_n. 条件概率$\mathbb{P}(X_{n+1} = y|X_n = x)$ 称为是(由时刻n 的位置x 到下时刻的位置y)的转移概率. 这个转移不仅与位置有关也与时间有关. 怎么理解与时间有关这个问题呢? 假设甲参加一个简单的掷硬币赌博, 第n 局的输赢是n 元钱, 那么甲在第n 局后手中的赌资X_n 满足Markov 性, 但是甲在下一局输或者赢的钱数是和局数有关的. 如果规则是每一局的输赢是固定的钱数, 那么甲在下一局输或赢的钱数和局数无关. 为了简化问题, 我们假设条件概率$\mathbb{P}(X_{n+1} = y|X_n = x)$ (有意义时)与时间n 无关, 这个假设的性质称为时间齐性.

定义5.1.2 一个状态空间为E 的满足Markov 性和时间齐性的随机序列称为(状态空间E 的) Markov 链.

一个Markov 链的时间集视方便取为整数, 正整数或者非负整数, 下面通常使用非负整数作为时间. 现在设$\{X_n : n \geqslant 0\}$ 是Markov 链, X_n 是时刻n 的位置, X_0 是初始位置. 令

$$p_{x,y} := \mathbb{P}(X_{n+1} = y | X_n = x),\ x, y \in E,$$

为了方便, 有时也写为$p(x, y)$, 它称为是Markov 链的转移概率. 显然由乘法公式, Markov 链的有限维分布可以写成为

$$\mathbb{P}(X_i = x_i, 0 \leqslant i \leqslant n) = \mathbb{P}(X_0 = x_0) p(x_0, x_1) \cdots p(x_{n-1}, x_n),$$

也就是说, 有限维分布由转移概率和X_0 的分布(初始分布)所决定. 容易验证, 转移概率$p : E \times E \longrightarrow \mathbf{R}$ 满足(1) $p_{x,y} \geqslant 0$; (2) $\sum_{y \in E} p_{x,y} = 1$.

反过来, $E \times E$ 上的函数$(x, y) \mapsto p_{x,y}$ 称为是E 上的转移函数, 如果它满足(1) $p_{x,y} \geqslant 0$; (2) $\sum_{y \in E} p_{x,y} = 1$. 需要时$p_{x,y}$ 可写为$p(x, y)$. 另外我们也将$p_{x,y}$ 看作一个矩阵$\mathbf{P}$ 在位置(x, y) 上的元素, 即

$$\mathbf{P} := (p_{x,y} : x, y \in E),$$

那么$\mathbf{P}$ 称为是E 上的转移矩阵. 这样, 转移函数与转移矩阵是一一对应的. 注意它是一个可能有可列多行与列的矩阵, 但是它的元素是非负的. 且每行的和是1. 很容易验证转移矩阵的n 次幂$\mathbf{P}^n$ 仍然是转移矩阵, 记位置(x, y) 的元素为$p^{(n)}_{x,y}$, 称为是n 步转移概率. 现在我们叙述并证明Markov 链的存在性定理.

定理5.1.1 给定E 上一个转移函数$(p_{x,y} : x, y \in E)$, 对任何E 上的分布律$\mu = (\mu_x : x \in E)$, 存在一个可测空间$(\Omega, \mathcal{F})$, 其上的一个可测序列$\{X_n : n \geqslant 0\}$ 与概率$\mathbb{P}^\mu$, 满足:

(1) 对任何$x \in E$, $\mathbb{P}^\mu(X_0 = x) = \mu_x$;

(2) 对任何$n \geqslant 0$, $x_0, x_1, \cdots, x, y \in E$,

$$\mathbb{P}^\mu(X_{n+1} = y | X_n = x, \cdots, X_1 = x_1, X_0 = x_0) = p_{x,y}.$$

证明. 不妨取$\Omega = E^{\mathbf{N}}$, $\mathcal{F} = \mathcal{E}^{\mathbf{N}}$, $X_i = \pi_i$. 任意取$x_0, x_1, \cdots, x_n \in E$, 定义概率

$$\mu^{[n]}(x_0, \cdots, x_n) := \mu_{x_0} p(x_0, x_1) \cdots p(x_{n-1}, x_n),\ x_i \in E, 0 \leqslant i \leqslant n.$$

那么由转移函数的性质推出$(E^{[n]}, \mathcal{E}^{[n]}, \mu_x^{[n]})$ 是相容的概率空间列. 因此由定理4.2.1知, 在$(\Omega, \mathcal{F})$ 上存在概率测度$\mathbb{P}^{\mu}$ 使得它在$E^{[n]}$ 上的投影是$\mu^{[n]}$, 即

$$\mathbb{P}^{\mu}(X_0 = x_0, \cdots, X_n = x_n) = \mu_{x_0} p(x_0, x_1) \cdots p(x_{n-1}, x_n).$$

然后很容易验证$\mathbb{P}^{\mu}$ 满足定理中的两个性质. □

定理中的(1) 说明初始分布是μ, (2) 是移动规则, 不依赖于μ. 我们得到了一族概率$\{\mathbb{P}^{\mu} : \mu$ 是E 上分布$\}$. 特别地, 用$\mathbb{P}^x$ 表示初始分布是x 点的单点分布对应的概率, 即$\mathbb{P}^x(X_0 = x) = 1$. 那么在概率$\mathbb{P}^x$ 看来, Markov 链$\{X_n\}$ 是从x 点出发的, 并且

$$\mathbb{P}^{\mu}(A) = \sum_{x \in E} \mu(x) \mathbb{P}^x(A), \ A \in \mathcal{E}^{\mathbf{N}},$$

即$\mathbb{P}^{\mu}$ 由$\{\mathbb{P}^x : x \in E\}$ 决定. 这样我们把

$$(\Omega, \mathcal{F}, \{X_n : n \geqslant 0\}, \{\mathbb{P}^x : x \in E\})$$

称为是联系着转移函数$(p_{x,y})$ 的Markov 链, 简单记为X.

任意给定$n \geqslant 1$, $x = x_0, x_1, \cdots, x_n \in E$, 由性质(2) 得联合分布

$$\mathbb{P}^x(X_n = x_n, \cdots, X_1 = x_1, X_0 = x_0) = p(x, x_1) \cdot \cdots \cdot p(x_{n-1}, x_n), \tag{5.1.1}$$

左边解释为X 沿着路线$x, x_1, \cdots, x_n$ 走的概率. 这联合分布是一个非常有用的公式, 首先让$x_n = y$, 对$x_1, \cdots, x_{n-1}$ 走遍E 上求和得

$$\mathbb{P}^x(X_n = y) = \sum_{x_1, \cdots, x_{n-1} \in E} p(x, x_1) \cdot \cdots \cdot p(x_{n-1}, y) = p^{(n)}(x, y),$$

即解释了$p^{(n)}(x, y)$ 为何称为是n 步转移概率. 另外对$n \geqslant 1$,

$$\mathbb{P}^z(X_{n+1} = y | X_n = x) = \frac{\mathbb{P}^z(X_{n+1} = y, X_n = x)}{\mathbb{P}^z(X_n = x)} = p(x, y),$$

右边与n 无关, 也就是说从x 转移到y 的概率与转移的时间无关, 这个性质称为Markov 链的时间齐性.

用$\mathcal{F}_n$ 表示$X_0, X_1, \cdots, X_n$ 生成的事件域, 显然它是离散的, 且不妨认为$\mathcal{F} = \sigma(X_n : n \geqslant 1)$. 那么由性质(2)和定理2.3.1[9],

$$\mathbb{P}^x(X_{n+1} = y | \mathcal{F}_n) = p(X_n, y) = \mathbb{P}^{X_n}(X_1 = y).$$

再引入Ω 上的推移算子θ_n, $n \geqslant 0$, 满足对任何$k \geqslant 0$, $X_k \circ \theta_n = X_{k+n}$.

定理5.1.2 对任何有界随机变量Y, $n \geqslant 0$, $x \in E$,

$$\mathbb{E}^x(Y \circ \theta_n | \mathcal{F}_n) = \mathbb{E}^{X_n}(Y).$$

证明. 因为$\omega \mapsto \mathbb{E}^{X_n(\omega)}(Y)$ 是$\mathcal{F}_n$ 随机变量, 故由定义只需验证对任何原子$A \in \mathcal{F}_n$,

$$\mathbb{E}^x(Y \circ \theta_n; A) = \mathbb{E}^x(\mathbb{E}^{X_n}(Y); A) \tag{5.1.2}$$

如果(5.1.2) 对示性函数$Y = 1_B$, $B \in \mathcal{F}$ 成立, 那么由条件期望的性质, 它对非负简单随机变量也成立, 取递增趋于Y 的简单随机变量列Y_n, 因此由单调收敛定理推出(5.1.2) 成立.

首先令$Y = 1_B$, 其中$B = \{X_0 = x_0, \ldots, X_m = x_m\}$ 与$A = \{X_0 = y_0, \ldots, X_n = y_n\}$, 那么由(5.1.1),

$$\begin{aligned}
\mathbb{E}^x(1_B \circ \theta_n; A) &= \mathbb{P}^x(X_0 = y_0, \ldots, X_n = y_n, X_n = x_0, X_{n+1} = x_1, \ldots, X_{n+m} = x_m) \\
&= 1_{\{x=y_0\}} p(y_0, y_1) \cdots p(y_{n-1}, y_n) 1_{\{x_0 = y_n\}} p(x_0, x_1) \cdots p(x_{m+1}, x_m) \\
&= 1_{\{x=y_0\}} p(y_0, y_1) \cdots p(y_{n-1}, y_n) \mathbb{P}^{y_n}(B) \\
&= \mathbb{E}^x(\mathbb{P}^{X_n}(B); A).
\end{aligned}$$

这样因为$\mathcal{F}_n$ 是离散的且由上面形式的A 的全体生成, 故(5.1.2) 对$Y = 1_B$, 其中B 是上面的形式, 与任何$A \in \mathcal{F}_n$ 成立.

我们要证明对任何$B \in \mathcal{F}$, (5.1.2) 对$Y = 1_B$ 成立. 用$\mathcal{H}$ 表示上面形式的B 的全体, $\mathcal{A}$ 表示使(5.1.2) 成立的$Y = 1_B$ 的事件$B \in \mathcal{F}$ 的全体. 当然我们证明了$\mathcal{H} \subset \mathcal{A}$. 如果我们能够证明$\mathcal{A}$ 也是σ-域, 那么此包含关系可说明任何X_n 是$\mathcal{A}$ 随机变量, 而$\mathcal{F}$ 是最小的, 故有$\mathcal{F} = \mathcal{A}$, 就完成证明了.

直接验证$\mathcal{A}$ 是事件域很困难. 但可以容易地看出(1) $\mathcal{H}$ 是π 类: 关于有限交运算是封闭的. 由概率的性质推出(2) $\mathcal{A}$ 是一个Dynkin 系: 即(a) $\varnothing, W \in \mathcal{A}$; (b) $B \in \mathcal{A}$ 蕴含有$B^c \in \mathcal{A}$; (c) 如果$B_n \in \mathcal{A}$ 且互斥, 则$\bigcup_n B_n \in \mathcal{A}$. 再应用附录2.3.1 中Dynkin 定理. 它断言$\mathcal{A}$ 包含由$\mathcal{H}$ 生成的σ-域, 即有$\mathcal{A} \supset \mathcal{F}$. □

上面的时间n 可以用一类随机时间代替, 也就是强Markov 性成立. 一个映射$\tau : \Omega \longrightarrow [0, \infty]$ 称为是一个停时, 如果对任何$n \geqslant 0$, 有$\{\tau \leqslant n\} \in \mathcal{F}_n$. 这等价于对任何$n \geqslant 0$, $\{\tau = n\} \in \mathcal{F}_n$. 注意, τ 可以取$+\infty$ 为值, 且一定是$\mathcal{F}_\infty$ (广义)随机变量. 容易验证固定时间n 是一个停时. 定义$\mathcal{F}_\tau$ 是$\mathcal{F}_\infty$ 中满足对任何$n \geqslant 0$, $A \cap \{\tau \leqslant n\} \in \mathcal{F}_n$ 的元素A 的全体, 那么(1) $\mathcal{F}_\tau$ 是一个σ-代数; (2) 当$\tau \equiv n$ 时,

$\mathcal{F}_\tau = \mathcal{F}_n$. (请读者自行验证.) 再定义停止位置$X_\tau := X_n$ 当$\tau = n$ 时, $n \geqslant 0$. 这样X_τ 在集合$\{\tau < \infty\}$ 上被定义, 且在此集合上是$\mathcal{F}_\tau$ 可测的. 类似定义$\theta_\tau := \theta_n$ 当$\tau = n$ 时. 同样θ_τ 也在集合$\{\tau < \infty\}$ 上被定义. X 有下面的强Markov 性: 设τ 是一个停时, 则对任何$B \in \mathcal{F}_\infty$, $x \in E$ 有

$$\mathbb{E}^x(1_B \circ \theta_\tau; \tau < +\infty | \mathcal{F}_\tau) = \mathbb{E}^{X_\tau}(1_B) 1_{\{\tau < +\infty\}}.$$

证明是简单的. 任取$A \in \mathcal{F}_\tau$, 那么对任何$n \geqslant 0$, $A \cap \{\tau = n\} \in \mathcal{F}_n$, 由Markov 性,

$$\begin{aligned}
\mathbb{E}^x(1_B \circ \theta_\tau; A \cap \{\tau < \infty\}) &= \sum_{n=0}^{\infty} \mathbb{E}^x(1_B \circ \theta_\tau; A \cap \{\tau = n\}) \\
&= \sum_{n=0}^{\infty} \mathbb{E}^x(1_B \circ \theta_n; A \cap \{\tau = n\}) \\
&= \sum_{n=0}^{\infty} \mathbb{E}^x(\mathbb{E}^{X_n}(1_B); A \cap \{\tau = n\}) \\
&= \mathbb{E}^x(\mathbb{E}^{X_\tau}(1_B); A \cap \{\tau < \infty\}).
\end{aligned}$$

这就证明了强Markov 性. 因此在Markov 链的情形下, Markov 性与强Markov 性是等价的, 但强Markov 性是一个很有用的工具.

下面我们将主要讨论转移函数的性质和状态的分类.

例5.1.1 (Bernoulli-Laplace 扩散模型) 设A, B 两箱中各有r 个球, 其中r 个白, r 个黑. 记X_0 是开始时A 箱中白球个数, 然后各任取一球对换, X_n 是经n 次对换后A 箱中白球个数. $X = (X_n : n \geqslant 1)$ 是随机序列, 状态空间是$E = \{0, 1, 2, \cdots, r\}$. 容易验证$X$ 是Markov 链且

$$p_{x,x-1} = \left(\frac{x}{r}\right)^2, \; p_{x,x} = 2\frac{x(r-x)}{r^2}, \; p_{x,x+1} = \left(\frac{r-x}{r}\right)^2.$$

这是一个关于两种液体混合的概率模型. ∎

例5.1.2 (无限制随机游动) 取独立同分布随机序列$\{\xi_n\}$, 满足$\mathbb{P}(\xi_n = 1) = p$, $\mathbb{P}(\xi_n = -1) = q = 1 - p$, 令$X_0 = x$, $X_n = X_0 + \sum_{i=1}^n \xi_i$, 则$X = (X_n : n \geqslant 0)$

是一个从x 出发的Markov 链. 转移矩阵为

$$\mathbb{P} := \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & q & 0 & p & 0 & 0 & \cdots \\ \cdots & 0 & q & 0 & p & 0 & \cdots \\ \cdots & 0 & 0 & q & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

更一般地, 设E 是可列离散加群(例如$\mathbf{Z}^d$), $\{\xi_n\}$ 是状态空间为E 的独立同分布随机序列, 定义$X_0 = x$, $X_n = X_0 + \sum_{i=1}^n \xi_i$, 则$X = \{X_n : n \geqslant 0\}$ 是从x 出发的Markov 链. 设ξ_i 的分布为$\mathbb{P}(\xi_i = x) = p_x$, $x \in E$. 因为$\mathbb{P}(X_1 = y|X_0 = x) = \mathbb{P}(\xi_1 = y - x) = p_{y-x}$, 故Markov 链的转移概率为$p_{x,y} = p_{y-x}$, $x, y \in E$. ∎

例5.1.3 (具吸收壁的随机游动) 设Markov 链X 有状态空间$E = \{0, 1, \cdots, r\}$, 转移矩阵

$$P := \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix},$$

$p + q = 1$. 直观地, 当X_n 位于状态1 与$r - 1$ 之间时, 则下一步向左的概率为q, 向右的概率为p, 而当X_n 达到边界0 或r 时, 它将不再离开. ∎

例5.1.4 A, B 两人比赛, A 赢的概率是p, 输的概率是$q = 1 - p$. (1) 如果规则是比赛到某人胜出两场为赢, 求A 赢的概率. (2) 如果规则是某人连续赢两场为赢, 求A 赢的概率.

对规则(1). 用X_n 表示到第n 局结束时, A 的净赢局数. 显然X_n 是个具吸收壁Markov 链, 状态依次为-2,-1,0,1,2, 其转移矩阵为

$$P := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

用$p^A(x)$ 表示从状态x 开始最终A 赢的概率. 那么$p^A(-2)=0$, $p^A(2)=1$. 由全概率公式

$$\begin{aligned}p^A(-1)&=p\cdot p^A(0)\\ p^A(0)&=q\cdot p^A(-1)+p\cdot p^A(1)\\ p^A(1)&=q\cdot p^A(0)+p\end{aligned}$$

可以容易地推出A 最终赢的概率是$p^A(0)=\frac{p^2}{1-2pq}$.

对规则(2). 这个Markov 链不能解决问题. 让我们用$X_n=A$ (或$X_n=B$)表示到第n 局A 赢(B 赢). $Y_n=(X_{n-1},X_n), n\geqslant 2$. 则$Y_n$ 有四个状态依次为AA,AB,BA,BB. 转移矩阵为

$$P:=\begin{bmatrix}1&0&0&0\\0&0&p&q\\p&q&0&0\\0&0&0&1\end{bmatrix}.$$

显然$p^A(AA)=1$, $p^A(BB)=0$, $p^A(AB)=0+p\cdot p^A(BA)$, $p^A(BA)=p+q\cdot p^A(AB)$. 先观察头两局比赛, 由全概率公式,

$$p^A=p^2+pq\cdot p^A(AB)+qp\cdot p^A(BA)=\frac{p^2(1+q)}{1-pq}.$$

∎

例5.1.5 (对称随机游动) 设E 是$\mathbf{R}^d$ 上的整数格点, 任意$x\in E$ 都有$2d$ 个邻居(与x 的距离等于1), 如果y 是x 的邻居, 定义$p_{x,y}:=\frac{1}{2d}$, 否则定义$p_{x,y}:=0$. 则$\mathbb{P}=(p_{x,y}:x,y\in E)$ 是一个转移矩阵. 一个具有此转移矩阵的Markov 链称为是对称随机游动. ∎

例5.1.6 (与图对应的简单随机游动) 设E 是一个简单图的顶点集合, 对任何$x,y\in E$, 若x,y 有线直接连接, 则令$S(x,y)=1$ 否则$S(x,y)=0$. 则S 是对称的. 设

$$p_{x,y}=\frac{S(x,y)}{\sum_{z\in E}S(x,z)},$$

那么$(p_{x,y}:x,y\in E)$ 定义了E 上的一个转移函数, 对应的Markov 链称为是图上的简单随机游动. 上例中的对称随机游动就是整数格点图上的简单随机游动. ∎

现在设$(p_{x,y}:x,y\in E)$ 是转移函数, X 是对应的Markov 链. 设$x,y\in E$, 称x 可达y, 如果存在$n\geqslant 1$, 使$p^{(n)}_{x,y}>0$; 称x,y 互达, 如果x 可达y 且y 可达x.

练习5.1.1 证明: x 可达y 当且仅当存在$n \geqslant 1, x_i \in E, 0 \leqslant i \leqslant n$ 使得$p_{x_{i-1},x_i} > 0$, $1 \leqslant i \leqslant n$ 且$x_0 = x$, $x_n = y$.

由Chapman-Kolmogorov 方程, 状态间互达关系是等价关系. 称X 既约, 如果E 的任何两个状态可互达. 对$y \in E$, 令$\tau_y := \inf\{n \geqslant 1 : X_n = y\}$ (约定$\inf \varnothing = \infty$), 容易验证$\tau_y$ 是一个停时, 它被称为是状态y 的首达时. 这是一个非常基本的停时, 有许多性质和应用. 对$n \geqslant 0$, 令$f_{x,y}^{(n)} := \mathbb{P}^x(\tau_y = n)$, 显然$f_{x,y}^{(n)} \leqslant p_{x,y}^{(n)}$. 因为$\{\tau_y = k\} \in \mathcal{F}_k$, $X_{\tau_y} = y$, 故对$n \geqslant 1$,

$$\begin{aligned}
p_{x,y}^{(n)} &= \mathbb{P}^x(\{X_n = y\}) = \mathbb{P}^x(X_n = y, \tau_y \leqslant n) \\
&= \sum_{k=1}^{n} \mathbb{P}^x(X_{n-k} \circ \theta_k = y, \tau_y = k) \\
&= \sum_{k=1}^{n} \mathbb{E}^x(\mathbb{P}^{X_k}(X_{n-k} = y); \tau_y = k) \\
&= \sum_{k=1}^{n} \mathbb{P}^x(\tau_y = k)\mathbb{P}^y(X_{n-k} = y) \\
&= \sum_{k=1}^{n} f_{x,y}^{(k)} p_{y,y}^{(n-k)}.
\end{aligned}$$

再定义

$$f_{x,y} := \mathbb{P}^x(\tau_y < \infty) = \sum_{n \geqslant 1} f_{x,y}^{(n)},$$

这是过程从状态x 经有限步到达y 的概率, 可以看出当$x \neq y$ 时, x 可达y 当且仅当$f_{x,y} > 0$.

练习5.1.2 证明两个公式(1) 对任何$x, y \in E$,

$$f_{x,y} \sum_{n=0}^{\infty} p_{y,y}^{(n)} = \sum_{n=1}^{\infty} p_{x,y}^{(n)}.$$

(2) $f_{x,y}^{(n+1)} = \sum_{z \neq y} p_{x,z} f_{z,y}^{(n)}$.

一个状态$x \in E$ 称为是常返的, 如果从x 出发以概率1 回到x, 即$f_{x,x} = 1$, 否则x 称为暂留的. Markov 链X 称为是常返的, 如果其所有状态常返; 称为暂留, 如果其所有状态暂留.

练习5.1.3 证明: $y \in E$ 是常返的当且仅当对任何$k \geqslant 1$ 有$\mathbb{P}^y(\tau_y \circ \theta_k < \infty) = 1$.

定理5.1.3 我们有下列准则.

(1) 状态x 是常返的等价于$\mathbb{P}^x(\limsup\{X_n=x\})=1$ 也等价于$\sum_n p_{x,x}^{(n)}=\infty$;

(2) 状态x 是暂留的等价于$\mathbb{P}^x(\limsup\{X_n=x\})=0$ 也等价于$\sum_n p_{x,x}^{(n)}<\infty$.

证明. 对$y\in E$, 令

$$\begin{aligned}\tau^{(0)}&:=0,\ \tau:=\tau_y,\\ \tau^{(k)}&:=\tau^{(k-1)}+\tau\circ\theta_{\tau^{(k-1)}},\ k\geqslant 1.\end{aligned}$$

那么$\tau^{(k)}$ 实际上是X 第k 次遇到y 的时刻, 它也是一个停时. 再定义$A_k(y):=\{\tau^{(k)}<\infty\}$, 即表示事件: X 至少有k 次到达y. 则$\{A_k(y):k\geqslant 1\}$ 是单降集列, 且

$$\limsup_n\{X_n=y\}=\bigcap_k A_k(y),$$

由强Markov 性且因为$X_{\tau^{(k-1)}}=y$,

$$\begin{aligned}\mathbb{P}^x(A_k(y))&=\mathbb{P}^x(\tau^{(k)}<\infty)=\mathbb{P}^x(\tau^{(k-1)}<\infty,\tau\circ\theta_{\tau^{(k-1)}}<\infty)\\&=\mathbb{E}^x(\mathbb{P}^{X(\tau^{(k-1)})}(\tau<\infty);\tau^{(k-1)}<\infty)\\&=f_{y,y}\mathbb{P}^x(\tau^{(k-1)}<\infty)\\&=\cdots=f_{x,y}(f_{y,y})^{k-1}.\end{aligned}$$

因此, 当$y=x$ 时

$$\mathbb{P}^x(\limsup_n\{X_n=x\})=\lim_k(f_{x,x})^k=\begin{cases}0, & f_{x,x}<1,\\ 1, & f_{x,x}=1.\end{cases}$$

又由Borel-Cantelli 引理, $\sum_n p_{x,x}^{(n)}<\infty$ 蕴含着

$$\mathbb{P}^x(\limsup\{X_n=x\})=0,$$

故为完成定理的证明, 我们仅需证明$f_{x,x}<1$ 蕴含着$\sum_n p_{x,x}^{(n)}<\infty$. 用$P(t)$, $F(t)$ 分别表示$(p_{x,x}^{(n)}:n\geqslant 0)$, $(f_{x,x}^{(n)}:n\geqslant 0)$ 的母函数. 由上面推导的公式

$$P(t)=1+\sum_{n=1}^{\infty}p_{x,x}^{(n)}t^n$$

$$= 1 + \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_{x,x}^{(k)} p_{x,x}^{(n-k)} t^n$$
$$= 1 + \sum_{k=1}^{\infty} f_{x,x}^{(k)} t^k \sum_{n=k}^{\infty} p_{x,x}^{(n-k)} t^{n-k}$$
$$= 1 + F(t)P(t),$$

因此

$$P(t) = \frac{1}{1-F(t)},$$

推出$f_{x,x} < 1$ 当且仅当$\sum_n p_{x,x}^{(n)} < \infty$. □

下列定理表明在一个互达等价类中, 或者所有状态是常返的, 或者所有状态是暂留的.

定理5.1.4 如果$x, y \in E$ 是互达的, 则下列两者之一成立.

(1) x, y 都是常返的, $\mathbb{P}^x(\limsup\{X_n = y\}) = 1$ 且$\sum_n p_{x,y}^{(n)} = \infty$.

(2) x, y 都是暂留的, $\mathbb{P}^x(\limsup\{X_n = y\}) = 0$ 且$\sum_n p_{x,y}^{(n)} < \infty$.

证明. 因x, y 互达, 存在$i, j \geqslant 1$ 使得$p_{x,y}^{(i)} \cdot p_{y,x}^{(j)} > 0$, 而对任何$n \geqslant 1$,

$$p_{x,x}^{(i+n+j)} \geqslant p_{x,y}^{(i)} p_{y,y}^{(n)} p_{y,x}^{(j)},\ \ p_{y,y}^{(i+n+j)} \geqslant p_{y,x}^{(i)} p_{x,x}^{(n)} p_{x,y}^{(j)},$$

因此定理 5.1.3 推出x, y 或都是常返或都是暂留, 由强Markov 性

$$\begin{aligned}\mathbb{P}^x(\limsup\{X_n = y\}) &= \mathbb{P}^x(\limsup\{X_n = y\}, \tau_y < \infty)\\ &= \mathbb{E}^x(1_{\limsup\{X_n=y\}} \circ \theta_{\tau_y}; \tau_y < \infty)\\ &= \mathbb{E}^x(\mathbb{P}^y(\limsup\{X_n = y\}); \tau_y < \infty)\\ &= f_{x,y}\mathbb{P}^y(\limsup\{X_n = y\}),\end{aligned}$$

因此若x, y 是暂留的, 则$\mathbb{P}^x(\limsup\{X_n = y\}) = 0$ 且

$$(\sum_n p_{x,y}^{(n)}) p_{y,x}^{(j)} \leqslant \sum_n p_{x,x}^{(n+j)} \leqslant \sum_n p_{x,x}^{(n)} < \infty,$$

推出$\sum_n p_{x,y}^{(n)} < \infty$.

若x, y 是常返的, 则用相似方法可推出$\sum_n p_{x,y}^{(n)} = \infty$ 及

$$\mathbb{P}^x(\limsup\{X_n = y\}) = f_{x,y},$$

故我们仅需验证$f_{x,y}=1$ 即可.

事实上, 设y 常返, 对任何的$k\geqslant 1$, $\mathbb{P}^y(\limsup\{X_n=y\})=1$ 蕴含着$\mathbb{P}^y(\bigcup_{n>k}\{X_n=y\})=1$ 这又蕴含着$\mathbb{P}^y(\tau_y\circ\theta_k<\infty)=1$, 故由Markov 性得

$$\begin{aligned}
p^{(k)}_{y,x}&=\mathbb{P}^y(X_k=x)\\
&=\mathbb{P}^y(X_k=x,\tau_y\circ\theta_k<\infty)\\
&=\mathbb{E}^y(\mathbb{P}^y(X_k=x,\tau_y\circ\theta_k<\infty|\mathcal{F}_k))\\
&=\mathbb{E}^y(\mathbb{P}^{X_k}(\tau_y<\infty);X_k=x)\\
&=f_{x,y}p^{(k)}_{y,x}.
\end{aligned}$$

由y 可达x 推出$f_{x,y}=1$. □

练习5.1.4 证明: (1) 如果$y\in E$ 暂留, 则$\sum_{n=1}^{\infty}p^{(n)}_{x,y}<\infty$, $x\in E$. (2) 有限状态Markov 链至少有一个状态是常返的. 因此既约有限状态Markov 链必是常返的.

子集$C\subset E$ 称为是闭的, 如果对任何$x\in C$, $\sum_{y\in C}p_{x,y}=1$, 或等价地, 对任何$x\in C,y\notin C$, 有$p_{x,y}=0$, 即C 内的状态不可到达C 外的状态. 那么X 既约当且仅当E 没有非平凡闭子集. 上面证明的最后实际上证明了, 如果常返状态x 可达状态y, 那么y 也可达x 因此推出y 也常返. 因此所有X 的常返状态全体是个闭子集.

练习5.1.5 证明: (1) 如果C 是闭的, 那么对任何$n\geqslant 1$, $x\in C$, $y\notin C$ 有$p^{(n)}_{x,y}=0$; (2) 一个Markov 链既约当且仅当它没有真闭子集. (3) 所有常返状态全体是闭集. (4) 两个状态x,y 称为连通的, 如果x 可达y 或y 可达x. 连通分支是指一个连通等价类. 证明: 一个连通分支是闭的.

例5.1.7 让我们考虑例5.1.2 中的无限制随机游动. 显然对所有x, 从x 出发经$2n$ 步回到x 的概率$p^{(2n)}_{x,x}=\binom{2n}{n}p^nq^n$, 当然$p^{(2n+1)}_{x,x}=0$. 由Stirling 公式: $n!\sim\frac{n^n}{e^n}\sqrt{2\pi n}$, 得$\binom{2n}{n}\sim\frac{4^n}{\sqrt{\pi n}}$, 故

$$p^{(2n)}_{x,x}\sim\frac{4^n}{\sqrt{\pi n}}(pq)^n.$$

因此当$p=q=\frac{1}{2}$ 时, $\sum_n p^{(n)}_{x,x}=\infty$, 所有状态常返, 即$X$ 常返; 否则所有状态暂留, 即X 暂留. 实际上, 我们可以计算$f_{0,0}$. 由展开式

$$\frac{1}{\sqrt{1-4t}}=\sum_{n=0}^{\infty}\binom{2n}{n}t^n$$

推出$\{p_{0,0}^{(n)}\}$ 的母函数为

$$P(t)=\frac{1}{\sqrt{1-4pqt^2}}.$$

因此$\{f_{0,0}^{(n)}\}$ 的母函数为

$$F(t)=1-\sqrt{1-4pqt^2}.$$

推出$\mathbb{P}^0(\tau_0<\infty)=f_{0,0}=F(1)=1-|p-q|$.

对例5.1.3中的有吸收壁的随机游动, 容易验证从任何状态$0<i<r$ 出发, X 总在有限步内达到0 或r, 然后不再离开. 因此, 状态$0,r$ 常返, 而其他状态暂留. ∎

定理5.1.5 (Polya) $\mathbf{R}^d$ 上的对称随机游动当$d=1,2$ 时是常返的, 而当$d\geqslant 3$ 时是暂留的.

证明. 显然n 返回的概率与维数d 有关而与状态无关, 记为$q_n^{(d)}$. $d=1$ 的情况已在例5.1.7 中证明. 当$d=2$ 时, 同样奇数步返回的概率是零, 偶数$2n$ 返回必定是k 步向上, k 步向下, $n-k$ 步向左, $n-k$ 步向右, 因此

$$\begin{aligned}q_{2n}^{(2)}&=\frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{(2n)!}{k!k!(n-k)!(n-k)!}\\&=\frac{1}{4^{2n}}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k}\\&=\frac{1}{4^{2n}}\binom{2n}{n}^2\sim\frac{1}{\pi n}.\end{aligned}$$

因此$\sum_n q_n^{(2)}=\infty$.

当$d=3$ 时, 同样的分析得

$$\begin{aligned}q_{2n}^{(3)}&=\frac{1}{6^{2n}}\sum_{k_1+k_2+k_3=n}\frac{(2n)!}{k_1!k_1!k_2!k_2!k_3!k_3!}\\&=\frac{1}{6^{2n}}\binom{2n}{n}\sum_{k_1+k_2+k_3=n}\left[\frac{n!}{k_1!k_2!k_3!}\right]^2\\&=\frac{1}{2^{2n}}\binom{2n}{n}\sum_{k_1+k_2+k_3=n}\left[\frac{1}{3^n}\frac{n!}{k_1!k_2!k_3!}\right]^2.\end{aligned}$$

中括号中是一个三项分布, 其和为1. 因此

$$\sum_{k_1+k_2+k_3=n}\left[\frac{1}{3^n}\frac{n!}{k_1!k_2!k_3!}\right]^2$$

$$\leqslant \frac{1}{3^n} \cdot \max\{\frac{n!}{k_1!k_2!k_3!}: \ k_1 + k_2 + k_3 = n\}.$$

右边最大在$k_1 = k_2 = k_3$ 或最接近处达到, 由Stirling 公式, 最大值与$3^n \cdot n^{-1}$ 同级. 因此$q_{2n}^{(3)}$ 与$n^{-\frac{3}{2}}$ 同级, 故$\sum_n q_n^{(3)} < \infty$, 即3 维对称随机游动是暂留的. 当$d \geqslant 4$ 时, 证明类似. □

对$x \in E$, 定义x 的周期, 记$d(x)$, 为集合$\{n \geqslant 1 : p_{x,x}^{(n)} > 0\}$ 的最大公因数, (如集合是空, $d(x) := 0$.) 若$d(x) = 1$, 称x 是非周期的. 称Markov 链是非周期的, 如果其所有状态都是非周期的.

引理5.1.1 (1) 如果x, y 互达, 则$d(x) = d(y)$. (2) 如果X 是既约非周期的Markov 链, 则对$x, y \in E$, 存在n_0, 使当$n > n_0$ 时, $p_{x,y}^{(n)} > 0$.

证明. (1) 存在$s, t \in T$, 使得$p_{x,y}^{(s)} p_{y,x}^{(t)} > 0$, 对$n \geqslant 1$, 如果$p_{x,x}^{(n)} > 0$, 则$p_{y,y}^{(t+n+s)} \geqslant p_{y,x}^{(t)} p_{x,x}^{(n)} p_{x,y}^{(s)} > 0$, 故$d(y)$ 整除$s + n + t$, 而这时$p_{x,x}^{(n)} > 0$, 故$d(y)$ 也整除$s + 2n + t$, 因此$d(y)$ 整除n, 即$d(y) \leqslant d(x)$, 同理有$d(x) \leqslant d(y)$.

(2) 设$G_y := \{n \geqslant 1 : p_{y,y}^{(n)} > 0\}$, 则$G_y$ 对于加法封闭, 且因G_y 的最大公因数是1, 故存在m_0, 使G_y 包含比m_0 大的所有自然数. (这是一个数论问题, G_y 的最大公因子是1 蕴含着其中存在有限个元素$\{n_i : 1 \leqslant i \leqslant k\}$, 它们互素. 因此存在整数$\{\alpha_i : 1 \leqslant i \leqslant k\}$ 使得$\sum_i \alpha_i n_i = 1$, 故存在n 使得$n, n + 1 \in G_y$, 然后利用对加法封闭的性质证明结论.) 另外存在$s \in T$, 使$p_{x,y}^{(s)} > 0$, 故对任何$n > n_0 := m_0 + s$, $p_{x,y}^{(n)} > 0$. □

E 上的一个概率分布$(\pi_y : y \in E)$ 称为是转移矩阵$\mathbf{P}$ 的平稳分布, 如果

$$\sum_{x \in E} \pi_x p_{x,y} = \pi_y, \ y \in E.$$

定理5.1.6 如果一个既约非周期的Markov 链X 存在有平稳分布$(\pi_x : x \in E)$, 那么

(1) X 是常返的;

(2) 对任何$x, y \in E$, $\lim_n p_{x,y}^{(n)} = \pi_y$, 平稳分布是唯一的;

(3) 对任何$x \in E$, $\pi_x > 0$.

证明. (1) 由平稳分布的定义, 对任何$n \geqslant 1$,

$$\sum_{x \in E} \pi_x p_{x,y}^{(n)} = \pi_y, \ y \in E.$$

假设X 是暂留的, 则$\lim_n p_{x,y}^{(n)}=0$, 由控制收敛定理, 对任何$y\in E$, $\pi_y=0$, 与(π_y) 是概率分布矛盾. 注意这里不需要X 是非周期的条件.

(2) 我们首先构造$E\times E$ 上的转移矩阵, 对$(x,y),(u,v)\in E\times E$, 定义

$$p((x,y),(u,v)):=p_{x,u}p_{y,v},$$

容易验证$(p((x,y),(u,v)):(x,y),(u,v)\in E\times E)$ 是$E\times E$ 上的转移矩阵, 称为是$\mathbb{P}$ 的耦合矩阵, 存在以$E\times E$ 为状态空间的Markov 链$((Y_n,Z_n):n\in T)$ (对某个给定的初始分布), 以耦合矩阵为转移矩阵, 称其为耦合链, 显然(Y_n) 与(Z_n) 是两个独立的Markov 链, 因此

$$p^{(n)}((x,y),(u,v))=p_{x,u}^{(n)}p_{y,v}^{(n)},$$

由引理 5.1.1, 耦合链也是既约非周期的. 且$\pi_{(x,y)}:=\pi_x\pi_y$, $x,y\in E\times E$ 是耦合链的平稳分布, 故耦合链也是常返的. 固定$z\in E$, 令

$$\tau:=\inf\{n\geqslant 1:\ (Y_n,Z_n)=(z,z)\}$$

首次命中(z,z) 的时刻, 则对$m\leqslant n$,

$$\begin{aligned}
&\mathbb{P}^{(x,y)}(\{(Y_n,Z_n)=(u,v),\tau=m\})\\
&\quad=\mathbb{P}^{(x,y)}(\{(Y_n,Z_n)=(u,v)|\tau=m\})\mathbb{P}^{(x,y)}(\tau=m)\\
&\quad=\mathbb{P}^{(z,z)}(\{(Y_{n-m},Z_{n-m})=(u,v)\})\mathbb{P}^{(x,y)}(\tau=m)\\
&\quad=p_{z,u}^{(n-m)}p_{z,v}^{(n-m)}\mathbb{P}^{(x,y)}(\tau=m),
\end{aligned}$$

由此推出

$$\mathbb{P}^{(x,y)}(Y_n=u,\tau=m)=\mathbb{P}^{(x,y)}(Z_n=u,\tau=m).$$

因此

$$\begin{aligned}
p_{x,u}^{(n)}=\mathbb{P}^{(x,y)}(Y_n=u)&\leqslant\mathbb{P}^{(x,y)}(Y_n=u,\tau\leqslant n)+\mathbb{P}^{(x,y)}(\tau>n)\\
&=\mathbb{P}^{(x,y)}(Z_n=u,\tau\leqslant n)+\mathbb{P}^{(x,y)}(\tau>n)\\
&\leqslant\mathbb{P}^{(x,y)}(Z_n=u)+\mathbb{P}^{(x,y)}(\tau>n)\\
&=p_{y,u}^{(n)}+\mathbb{P}^{(x,y)}(\tau>n),
\end{aligned}$$

同理可证

$$p_{y,u}^{(n)}\leqslant p_{x,u}^{(n)}+\mathbb{P}^{(x,y)}(\tau>n),$$

故$|p_{x,u}^{(n)}-p_{y,u}^{(n)}|\leqslant \mathbb{P}^{(x,y)}(\tau>n)$, 而耦合链是常返的蕴含着$\mathbb{P}^{(x,y)}(\tau<\infty)=1$, 因此推出

$$\lim_n |p_{x,z}^{(n)}-p_{y,z}^{(n)}|=0,\ x,y,z\in E.$$

另一方面

$$\begin{aligned}\pi_z-p_{x,z}^{(n)}&=\sum_{y\in E}\pi_y p_{y,z}^{(n)}-p_{x,z}^{(n)}\\&=\sum_{y\in E}\pi_y(p_{y,z}^{(n)}-p_{x,z}^{(n)}),\end{aligned}$$

由控制收敛定理, $\lim_n p_{x,z}^{(n)}=\pi_z,\ x,z\in E$.

(3) 因x,y 互达, 存在$i,j\geqslant 1$ 使得$p_{x,y}^{(i)}\cdot p_{y,x}^{(j)}>0$, 而对任何$n\geqslant 0$,

$$p_{x,x}^{(i+n+j)}\geqslant p_{x,y}^{(i)}p_{y,y}^{(n)}p_{y,x}^{(j)},$$

令$n\to\infty$, $\pi_x\geqslant p_{x,y}^{(i)}p_{y,x}^{(j)}\pi_y$, 因此如果某个$\pi_x$ 是零, 则所有其他都是零, 故$\pi_x: x\in E$ 一定全是严格正的. □

非周期的条件是本质的, 例如我们取$E=\{0,1\}$, $p_{0,1}=1, p_{1,0}=1$. 则容易验证它是既约常返链, 平稳分布$\pi_0=\pi_1=\frac{1}{2}$, 但$p_{x,y}^{(n)}$ 极限不存在.

定理5.1.7 如果一个既约非周期的Markov 链X 没有平稳分布, 则对任何$x,y\in E$, $\lim_n p_{x,y}^{(n)}=0$.

证明. 我们仅需考虑X 是常返的情况, 这时如果在上面定理证明中构造的耦合链是暂留的, 则由定理 5.1.4

$$\sum_n (p_{x,y}^{(n)})^2=\sum_n p^{(n)}((x,x),(y,y))<\infty,$$

故$\lim_n p_{x,y}^{(n)}=0$.

现在设耦合链是常返的, 假设结论不对, 因E 是至多可列的, 故存在一个子列(n_u), 使$\lim_u p_{x,y}^{(n_u)}$ 对所有$x,y\in E$ 存在, 且不全为零, 由上面定理的证明中可知极限值与x 无关, 令$\alpha_y:=\lim_u p_{x,y}^{(n_u)}$. 对$E$ 的任意有限子集M,

$$\sum_{y\in M}\alpha_y=\lim_u\sum_{y\in M}p_{x,y}^{(n_u)}\leqslant 1,$$

故$\alpha:=\sum_{y\in E}\alpha_y\leqslant 1$. 另外

$$\sum_{z\in M}p_{x,z}^{(n_u)}p_{z,y}\leqslant p_{x,y}^{(n_u+1)}=\sum_{z\in E}p_{x,z}p_{z,y}^{(n_u)},$$

由控制收敛定理,
$$\sum_{z\in M}\alpha_z p_{z,y}\leqslant\sum_{z\in E}p_{x,z}\alpha_y=\alpha_y,$$
这推出$\sum_{z\in E}\alpha_z p_{z,y}\leqslant\alpha_y$, 如果其中有一个严格小于, 则
$$\alpha=\sum_{y\in E}\alpha_y>\sum_{y\in E}\sum_{z\in E}\alpha_z p_{z,y}=\sum_{z\in E}\alpha_z=\alpha,$$
导致矛盾, 故对任何$y\in E$, $\sum_{z\in E}\alpha_z p_{z,y}=\alpha_y$, 因$\alpha>0$, 记$\pi_y:=\frac{\alpha_y}{\alpha}$, 那么$(\pi_y: y\in E)$ 是$(p_{x,y}:x,y\in E)$ 的一个平稳分布, 与条件矛盾. □

为了进一步区别常返的类型, 我们引入正常返与零常返的概念. 设状态y 是常返的, 称它是正(positive)常返的, 如果平均回转时间有限
$$\mu_y:=\mathbb{E}^y(\tau_y)=\sum_{n\geqslant1}nf_{y,y}^{(n)}<\infty,$$
否则称y 是零(null)常返的.

练习5.1.6 证明: (1) 零常返态可达的状态一定也是零常返态. (2) 一个有限状态Markov 链没有零常返状态. (3) 既约有限状态Markov 链是正常返的.

定理5.1.8 设y 是常返的且$\lim_n p_{y,y}^{(n)}$ 存在记为u_y, 则$u_y>0$ 当且仅当$\mu_y<\infty$, 这时$u_y=\frac{1}{\mu_y}$.

证明. 数列$(p_{y,y}^{(n)})$ 与$(f_{y,y}^{(n)})$ 的母函数$P(t)$ 与$F(t)$ 满足$P(t)(1-F(t))=1$. 因为$\lim_n p_{y,y}^{(n)}$ 存在, 由Abel 定理(见§1.5习题)或其推论推出
$$\lim_n p_{y,y}^{(n)}=\lim_{t\uparrow1}\frac{1-t}{1-F(t)}=\frac{1}{F'(1)}.$$
因此定理的结论是显然的. □

从以上结果可以看出, 一个既约非周期Markov 链可分为三种情况:

(1) 链是暂留的, 这时$\sum_n p_{y,y}^{(n)}<\infty$, $y\in E$, 不存在平稳分布.

(2) 链是常返, 但是零常返, 这时$\sum_n p_{y,y}^{(n)}=\infty$, 但$\lim_n p_{y,y}^{(n)}=0$, 不存在平稳分布.

(3) 链是常返且是正常返的, 这时$\sum_n p_{y,y}^{(n)}=\infty$, 且$\lim_n p_{y,y}^{(n)}=\pi_y>0$, 存在平稳分布.

例5.1.8 设E 是非负整数集, 令

$$\mathbb{P} := \begin{bmatrix} q_0 & p_0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & \cdots \\ q_2 & 0 & 0 & p_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

其中p_x, q_x 是正的, 且$p_x + q_x = 1$, 因此$\mathbb{P}$ 是一个转移矩阵. 设X 是一个以$\mathbb{P}$ 为转移矩阵的Markov 链, 容易验证X 是既约非周期的. 对$n \geqslant 1$, $\sum_{k=1}^n f_{0,0}^{(k)} = 1 - \mathbb{P}^0(\tau_0 > n) = 1 - p_0p_1\cdots p_{n-1}$, 当$n \to \infty$ 时, 乘积$p_0p_1\cdots p_{n-1}$ 极限存在, 记为a, 显然$a \in [0,1]$, 且X 是常返的当且仅当$a = 0$.

解方程$\sum_{x\in E} \pi_x p_{x,y} = \pi_y$, $y \in E$ 得$\pi_x = \pi_0 p_0 p_1 \cdots p_{x-1}$, $x \geqslant 1$. 因此, 上述Markov 链有平稳分布当且仅当级数

$$\sum_{x\geqslant 1} p_0p_1\cdots p_{x-1}$$

收敛. ∎

最后我们简单介绍Markov 链的不变测度的存在唯一性. 因为E 可列, 所以E 上的测度μ 由其在单点上的测度$\mu_x = \mu(\{x\})$, $x \in E$ 决定, σ-有限性意味着所有μ_x 都是有限的. E 上的一个σ-有限测度$(\mu_x : x \in E)$ 称为转移函数为$(p_{x,y})$ 的Markov 链的不变测度, 如果对任何$y \in E$, $\mu_y = \sum_{x\in E} \mu_x p_{x,y}$. 显然, 这时对任何$n \geqslant 1$, 有

$$\mu_y = \sum_{x\in E} \mu_x p_{x,y}^{(n)},$$

且一个不变的概率测度就是平稳测度. 假设Markov 链X 既约, 则其不变测度如果有一个正, 那么全体都是正的. 所以下面说不变测度存在是指正的不变测度存在. 而唯一性是指任何两个不同的不变测度相差一个常数倍. 下面定理证明既约常返的Markov 链的不变测度存在而且唯一, 注意不需要非周期的条件, 实际上非周期的条件为了保证平稳性, 也就是说转移概率的收敛性.

定理5.1.9 (1) 既约常返的Markov 链的不变测度存在而且唯一. (2) 既约Markov 链是正常返的当且仅当它有平稳分布, 且其平稳分布$\pi_x = 1/\mathbb{E}^x\tau_x$, 其中$\tau_x$ 是x 的首中时.

证明. (1) 设X 是既约常返的Markov 链. 先证明存在性, 任取点$o \in E$, 记$\tau :=$

$\inf\{n \geqslant 1 : X_n = o\}$, 对任何$x \in E$,

$$\mu_x := \mathbb{E}^0 \sum_{n=1}^{\tau} 1_{\{X_n = x\}}.$$

因为X 常返, 故$X_\tau = 0$. 因此$\mu_o = 1$ 且若X 从o 出发, 则

$$\sum_{n=1}^{\tau} 1_{\{X_n = x\}} = \sum_{n=1}^{\tau} 1_{\{X_{n-1} = x\}}.$$

另外$\{\tau \geqslant n\} = \{\tau < n\}^c \in \mathcal{F}_{n-1}$, 现在由Markov 性, 当$n \geqslant 1$ 时,

$$\begin{aligned}
\mathbb{E}^o 1_{\{X_n = y, n \leqslant \tau\}} &= \mathbb{P}^o(X_1 \circ \theta_{n-1} = y; n \leqslant \tau) \\
&= \mathbb{E}^o(\mathbb{P}^{X_{n-1}}(X_1 = y); n \leqslant \tau) \\
&= \sum_{x \in E} (\mathbb{E}^o 1_{\{X_{n-1} = x, n \leqslant \tau\}}) \cdot p_{x,y},
\end{aligned}$$

两边对$n \geqslant 1$ 求和, 得$\mu_y = \sum_{x \in E} \mu_x p_{x,y}$, 由此方程结合$\mu_o = 1$ 以及X 的既约性推出对任何$x \in E$, $\mu_x < \infty$. 因此(μ_x) 是X 的不变测度.

下面证明唯一性, 设$(\nu_x : x \in E)$ 也是X 的不变测度, 定义(类似于Doob 的h-变换)

$$q_{x,y} := \frac{\nu_y}{\nu_x} p_{y,x}, \ x, y \in E.$$

由不变测度的性质推出$Q = (q_{x,y} : x, y \in E)$ 也是转移函数, 而且满足对任何$n \geqslant 1$,

$$q_{x,y}^{(n)} = \frac{\nu_y}{\nu_x} p_{y,x}^{(n)},$$

且$q_{x,x}^{(n)} = p_{x,x}^{(n)}$. 因此$Q$ 对应的Markov 链Y 也是既约常返的. 用$g_{x,y}^{(n)}$ 表示Y 从x 出发在时刻n 首次到达y 的概率. 那么$g_{x,y}^{(n+1)} = \sum_{z \neq y} q_{x,z} g_{z,y}^{(n)}$ 或者

$$g_{x,y}^{(n+1)} \nu_x = \sum_{z \neq y} \nu_z g_{z,y}^{(n)} \cdot p_{z,x}.$$

令$r_x^{(n)} = \mathbb{P}^o(X_n = x, n \leqslant \tau)$, 那么上面在证明存在性的时候证明了当$n \geqslant 1$ 时

$$r_x^{(n+1)} = \sum_{z \in E} r_z^{(n)} \cdot p_{z,x} = \sum_{z \in E, z \neq o} r_z^{(n)} \cdot p_{z,x}.$$

当$n = 1$ 时, $g_{x,y}^{(1)} = q_{x,y}$, $r_x^{(1)} = p_{o,x}$, 因此$g_{x,o}^{(1)} \nu_x = \nu_o r_x^{(1)}$, 由此看出$\{g_{x,o}^{(n)} \nu_x : n \geqslant 1\}$ 和$\{\nu_o r_x^{(n)} : n \geqslant 1\}$ 满足同样的回归方程且有相同的初值, 故对任何$n \geqslant 1$,

$g_{x,o}^{(n)}\nu_x = \nu_o r_x^{(n)}$. 现在两边对$n$ 求和, 由常返性推出左边等于ν_x, 右边由定义等于$\nu_o \cdot \mu_x$, 从而$\nu_x = \nu_o \cdot \mu_x$ 对任何$x \in E$ 成立.

(2) 设X 是既约正常返的Markov 链, 那么它有唯一的不变测度

$$\mu_x = \mathbb{E}^o \sum_{n=1}^{\tau} 1_{\{X_n = x\}},\ x \in E,$$

因此$\sum_{x\in E} \mu_x = \mathbb{E}^o \tau < \infty$, 故$\pi_x := \mu_x / \sum_x \mu_x$ 是平稳分布, 且$\pi_o = 1/\mathbb{E}^o \tau$. 因为$o$ 是任意取的, 且平稳分布是唯一的, 故$\pi_x = 1/\mathbb{E}^x \tau_x$ 对任何$x \in E$ 成立.

反过来, 如果既约Markov 链有平稳分布π_x, 那么定理5.1.6(1)告诉我们X 是常返的. 由唯一性, 存在常数c 使得

$$\mathbb{E}^o \sum_{n=1}^{\tau} 1_{\{X_n = x\}} = c\pi_x,\ x \in E,$$

因此$\mathbb{E}^o \tau < \infty$, 即$X$ 是正常返的. □

练习**5.1.7** 设X 是转移函数为$p_{x,y}$ 的Markov 链. 对任何$x \in E$, 定义

$$M_n := \sum_{i=1}^{n} 1_{\{X_i = y\}} - \sum_{x\in E} \left(\sum_{i=1}^{n} 1_{\{X_{i-1} = x\}} \right) \cdot p_{x,y},$$

证明: (1) 对任何$0 \in E$, $(M_n : n \geqslant 1)$ 是$\mathbb{P}^0$ 鞅. (2) 将Doob 有界停止定理应用于(M_n) 证明既约常返的X 有不变测度.

习　题

1. 证明: 例5.1.3 中具吸收壁的随机游动必定在有限步达到状态0 或r.

2. 设$E = \{0, 1\}$.

   (a) 验证: $p_{0,0}^{(n)} = p_{1,0} + (p_{0,0} - p_{1,0})p_{0,0}^{(n-1)}$;

   (b) 写出$p_{0,0}^{(n)}$ 的表达式并求出极限;

   (c) 求出$p_{x,y}^{(n)}$ 及其极限.

3. 甲袋中有3 个黑球, 乙袋中有3 个白球, 每次从两袋中各任取一球, 交换放入另一袋中, 用p_n 表示交换n 次后甲袋中的球颜色一致的概率, 求$\lim_n p_n$.

4. 设$X=(X_n)$ 是概率空间$(\Omega,\mathcal{F},P)$ 上以E 为状态空间的Markov 链. 令$Y_n:=(X_n,X_{n+1})$. 记$F:=\{(x,y)\in E\times E:\ p_{x,y}>0\}$.

   (a) 证明: (Y_n) 是一个以F 为状态空间的Markov 链.

   (b) 写出转移概率并证明: 如果X 是既约与非周期, 则Y 也是.

   (c) 证明: 如果$\{\pi_x\}$ 是X 的平稳分布, 则$\{\pi_x p_{x,y}:(x,y)\in F\}$ 是Y 的平稳分布.

5. 某人有r 把伞放在家或者办公室用于来往于家与办公室之间. 当且仅当天下雨且手边有伞时, 带一把伞走, 到达后放下. 下雨的概率等于p.

   (a) 用X_n 表示他出(家或者办公室)门时手边的伞的数目, 说明它是一个Markov 链, 写出其转移概率. 它是否有平稳分布? 如有, 求其平稳分布.

   (b) 计算他被淋湿的概率的极限.

   (c) 证明: 对任何p, 5 把伞可以保证他以95% 以上的概率不被淋湿.

6. 依次等可能地往M 个盒子里放球, ξ_n 表示放入第n 个球后空盒的个数. 证明$\{\xi_n\}$ 是Markov 链并写出它的转移矩阵.

7. 考虑例5.1.2中可列离散加群上的随机游动X, 它的转移函数是$p_{x,y}=p_{y-x}$, 其中$(p_x:x\in E)$ 是概率分布. (1) 问(p_x) 满足什么条件时, X 是既约的? (1) 证明: 如果E 无限, 那么X 不可能有平稳分布. 因此它不可能是正常返的.

8. 设测度(μ_x) 满足对任何$y\in E$,
$$\sum_{x\in E}\mu_x p_{x,y}\leqslant \mu_y.$$
   (1) 如果X 既约, 证明: (μ_x) 或者恒等于0, 或者恒等于无穷, 或者全部是有限正数. (2) 满足上述条件的σ-有限测度称为Markov 链的过分测度. 证明: 如果E 有限, 那么过分测度是不变测度. (3) 设(μ_x) 是过分测度, 定义
$$\mu_y^{(n)}:=\sum_{x\in E}\mu_x p_{x,y}^{(n)},\ y\in E.$$
   证明: $\{\mu_y^{(n)}:n\geqslant 1\}$ 递减. 令$\mu_y^i:=\lim_n \mu_y^{(n)}$, $y\in E$. 证明: $(\mu_y^i:y\in E)$ 是不变测度.

9. 设X 是E 上既约常返Markov 链, 任取有限集$A \subset E$, 定义σ 是A 的首中时, σ_n 是第n 次命中A 的时间, 也就是说

$$\sigma_n = \sigma_{n-1} + \sigma \circ \theta_{\sigma_{n-1}}.$$

令$Y_n := X(\sigma_n)$. 证明: (1) Y 是以A 为状态空间的既约常返Markov 链(事实上是X 的一个时间变换, 称为是X 在A 上的迹). (2) 如果$(\mu_x : x \in E)$ 是X 的不变测度, 证明它在A 上的限制$(\mu_x : x \in A)$ 是Y 的不变测度.

10. 举简单随机游动的例子说明既约暂留的Markov 链也可能有不变测度, 但未必唯一.

§5.2 Galton-Watson 分支过程

如果不加限制, 生物种群的增长是指数式的, 一个指数增长的最简单的随机模型是所谓的Galton-Watson 分支过程, 虽然简单, 但它有重要的理论价值.

Galton-Watson 分支过程, 简称为分支过程, 的描述是直观的, 设随机变量X 的取值是非负整数, 代表某种生物种群个体繁殖后代数目. 设X 的分布律为

$$p_n = \mathbb{P}(X = n),\ n \geqslant 0,\ \sum_n p_n = 1.$$

用f 表示X 的母函数, 即

$$f(s) = \mathbb{E}[s^X] = \sum_{n \geqslant 0} p_n s^n,\ s \in [0, 1].$$

母函数是研究分支过程最重要的工具. 当X 是常数时, 称为平凡情形, 因为这时的增长模式是确定的. 下面总是假设非平凡的情形.

假设种群的每个个体都以同样的模式且独立于其它个体来繁殖后代. 用数学语言, 设$(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间, $\{X_n^{(k)} : n, k \geqslant 1\}$ 是其上独立同分布随机变量, 都与X 有相同分布. 实际上, $X_n^{(k)}$ 简单视为第k 代中第n 个个体所繁殖的后代数目. 记

$$Z_0^{(j)} = j,\ Z_{n+1}^{(j)} = \sum_{i=1}^{Z_n^{(j)}} X_i^{(n+1)},\ n \geqslant 0, \tag{5.2.1}$$

即$Z_n^{(j)}$ 表示一个起始个体数为j 的种群的第n 代后代的数目, 记$Z_n = Z_n^{(1)}$, 实际上, $\{Z_n^{(j)}\}$ 是$\{Z_n\}$ 的j 个独立复制的和. 另外, 直观地看, 在给定$Z_n = j$ 的条件下, $\{Z_{n+k} : k \geqslant 0\}$ 是独立于$\mathcal{G}_n := \sigma(\{X_i^{(k)} : i \geqslant 1, 1 \leqslant k \leqslant n\})$ 且与$\{Z_n^{(j)} : n \geqslant 0\}$ 同分布的. 因此$\{Z_n\}$ 是Markov 链.

定理5.2.1 $Z = \{Z_n^{(j)} : n \geqslant 0\}$ 是一个Markov 链.

Proof. 记$\mathcal{F}_n = \sigma(Z_1, Z_2, \cdots, Z_n)$, 则$\mathcal{F}_n \subset \mathcal{G}_n$. 首先, 对任何$n \geqslant 1$, 非负整数$k, j, j_{n-1}, \cdots, j_1$, 因为在$\{Z_n = j\}$ 上,

$$\mathbb{P}(Z_{n+1} = k|Z_n) = \mathbb{P}(Z_{n+1} = k|Z_n = j) = \mathbb{P}(\sum_{i=1}^{j} X_i^{(n+1)} = k),$$

故有

$$\begin{aligned}
&\mathbb{P}(Z_{n+1} = k, Z_n = j, Z_{n-1} = j_{n-1}, \cdots, Z_1 = j_1) \\
&= \mathbb{P}(\sum_{i=1}^{j} X_i^{(n+1)} = k)\mathbb{P}(Z_n = j, Z_{n-1} = j_{n-1}, \cdots, Z_1 = j_1) \\
&= \mathbb{E}(\mathbb{P}(\sum_{i=1}^{j} X_i^{(n+1)} = k); Z_n = j, Z_{n-1} = j_{n-1}, \cdots, Z_1 = j_1) \\
&= \mathbb{E}(\mathbb{P}(Z_{n+1} = k|Z_n); Z_n = j, Z_{n-1} = j_{n-1}, \cdots, Z_1 = j_1),
\end{aligned}$$

即有Markov 性,

$$\mathbb{P}(Z_{n+1} = k|\mathcal{F}_n) = \mathbb{P}(Z_{n+1} = k|Z_n).$$

时间齐次性是显然的, 因为

$$\mathbb{P}(Z_{n+1} = k|Z_n = j) = \mathbb{P}(\sum_{i=1}^{j} X_i^{(n+1)} = k) = p_k^{*j},$$

其中$p^{*j} = \{p_k^{*j} : k \geqslant 0\}$ 是$\{p_k : k \geqslant 0\}$ 的j-重卷积, 它是X 的j 个独立复制的和的分布律, p^{*0} 就是0 点的Dirac 测度. 因此$\{Z_n\}$ 是从1 出发的(对应地, $\{Z_n^{(j)} : n \geqslant 0\}$ 是从j 出发的)以$p_{j,k} = p_k^{*j}$, $j, k \geqslant 0$ 为转移函数的(时齐) Markov 链. □

练习5.2.1 Z 的非零状态互达当且仅当$p_0 > 0, p_1 > 0, p_0 + p_1 < 1$.

现在我们来计算Z_n 的母函数, 记它的母函数为$f_{(n)}$, 再记f 的n-重复合函数为f_n, f_0 是恒等映射. 这样

$$f_{(n+1)}(s) = \mathbb{E}\left[s^{Z_{n+1}}\right] = \mathbb{E}\left[s^{\sum_{i=1}^{Z_n} X_i^{(n+1)}}\right]$$

$$
\begin{aligned}
&= \sum_{j\geqslant 0} \mathbb{E}\left[s^{\sum_{i=1}^{j} X_i^{(n+1)}}\right] \cdot \mathbb{P}(Z_n = j) \\
&= \sum_{j\geqslant 0} (f(s))^j \mathbb{P}(Z_n = j) = \mathbb{E}\left[f(s)^{Z_n}\right] = f_{(n)}(f(s)).
\end{aligned}
$$

这推出$f_{(n)} = f_n$, 即Z_n 的母函数就是X 的母函数的n-重复合. 用m, σ^2 分别表示X 的期望与方差. 那么

$$
m = f'(1),\ \sigma^2 = f''(1) + f'(1) - f'(1)^2. \tag{5.2.2}
$$

练习5.2.2 证明:

$$
\mathbb{E}[Z_n] = m^n,\ \mathrm{var}(Z_n) = \sigma^2 m^{n-1} \sum_{i=0}^{n-1} m^i.
$$

注意X 是非平凡的, 任意验证下面的性质.

(1) f 在$[0,1]$ 上递增且凸;

(2) $f(0) = p_0$, $f(1) = 1$;

(3) 若$m \leqslant 1$, 则对任何$s \in [0,1)$ 有$f(s) > s$;

(4) 若$m > 1$, 则$f(s) = s$ 在$[0,1)$ 上有唯一的根.

练习5.2.3 证明上述性质.

也就是说$f(s) = s$ 在$[0,1]$ 上包括1 在内最多有两个根, 记小的那个为q.

引理5.2.1 (1) 若$m \leqslant 1$, 则$q = 1$. 若$m > 1$, 则$q < 1$;

(2) 若$s \in [0,q)$, 则$f_n(s) \uparrow q$;

(3) 若$s \in (q,1]$, 则$f_n(s) \downarrow q$.

Proof. (1) 是显然的. (2) 由f 的递增性与凸性, 当$s \in [0,q)$ 时, $q = f(q) > f(s) > s$. 由归纳法推出$s < f(s) < f_2(s) < \cdots < f_n(s) < q$, 即$f_n(s)$ 递增, 记极限为x, 则由f 的连续性得$x = \lim_n f_n(s) = f(x)$. 因$x \leqslant q$, 故由最小性, $x = q$. (3) 得证明类似. □

例5.2.1 设$0 < p < 1$, $0 < b < 1 - p$, 记

$$
\begin{aligned}
&p_k = bp^{k-1}, k \geqslant 1; \\
&p_0 = 1 - \sum_{k\geqslant 1} p_k = (1 - p - b)(1 - p)^{-1}.
\end{aligned}
$$

设X 的分布律为$\{p_k : k \geqslant 0\}$. 则

$$f(s) = 1 - \frac{b}{1-p} + \frac{bs}{1-ps},\ m = \frac{b}{(1-p)^2}.$$

具体地算Z_n 的母函数f_n. 这里用到分式函数的一个性质, 容易验证对任何实数u, v,

$$\frac{f(s)-f(u)}{f(s)-f(v)} = \frac{\frac{bs}{1-ps} - \frac{bu}{1-pu}}{\frac{bs}{1-ps} - \frac{bv}{1-pv}} = \frac{s-u}{s-v} \cdot \frac{1-pv}{1-pu}.$$

当$m \neq 1$, $f(s) = s$ 有两个不同的根s_0 与1. 取$u = s_0$, $v = 1$, 得

$$\frac{f(s)-s_0}{f(s)-1} = \frac{s-s_0}{s-1} \cdot \frac{1-p}{1-ps_0}.$$

两边同乘$f(s)-1$, 然后让$s \uparrow 1$, 可知$\frac{1-p}{1-ps_0} = \frac{1}{m}$. 因此有

$$\frac{f(s)-s_0}{f(s)-1} = \frac{1}{m} \cdot \frac{s-s_0}{s-1},$$

迭代

$$\frac{f_n(s)-s_0}{f_n(s)-1} = \frac{1}{m} \frac{f_{n-1}(s)-s_0}{f_{n-1}(s)-1} = \cdots = m^{-n} \frac{s-s_0}{s-1},$$

由此解出f_n 得

$$f_n(s) = \frac{s_0 - m^{-n} \cdot \frac{s-s_0}{s-1}}{1 - m^{-n} \cdot \frac{s-s_0}{s-1}}.$$

当$m = 1$ 时, $f(s) = \frac{p-(2p-1)s}{1-ps}$, 用归纳法不难得到$f_n$ 的表达式. ∎

令$\tau = \inf\{n : Z_n = 0\}$. 则$\tau$ 称为是Z 的灭绝时间, $\mathbb{P}(\tau < \infty)$ 称为是分支过程Z 的灭绝概率. 因为$\{\tau < \infty\} = \bigcup_n \{Z_n = 0\}$, 由于0 是个陷阱, 故$\{Z_n = 0\}$ 是递增的, 因而

$$\mathbb{P}(\tau < \infty) = \lim_n \mathbb{P}(Z_n = 0) = \mathbb{P}(\lim_n Z_n = 0),$$

而Z_n 的母函数是f_n, 那么$\mathbb{P}(Z_n = 0) = f_n(0)$, 因此

$$\mathbb{P}(\tau < \infty) = \lim_n f_n(0) = q.$$

即有如下定理.

定理5.2.2 分支过程Z 的灭绝概率为母函数f 在$[0, 1]$ 上小的根q. 因此当$m \leqslant 1$ 时, $q = 1$, 即灭绝是几乎必然的. 当$m > 1$ 时, $q < 1$.

练习5.2.4 证明: (1) 任何非零状态是暂留的, 因为

$$\mathbb{P}(Z_{n+i} \neq k, i \geqslant 1 | Z_n = k) \geqslant \begin{cases} p_{k,0}, & p_0 > 0, \\ 1 - p_{k,k}, & p_0 = 0; \end{cases}$$

(2) 依概率1 $Z_n \to 0$ 或$Z_n \to \infty$.

下面我们来讨论Z_n 的增长速度. 如果$m \leqslant 1$, 则已经知道$\mathbb{P}(Z_n \to 0) = 1$.

引理5.2.2 记$W_n = Z_n \cdot m^{-n}$. 则$\{W_n\}$ 是一个鞅.

Proof. 由Markov 性,

$$\begin{aligned} &\mathbb{E}(Z_{n+1}|Z_n = j, Z_{n-1} = j_{n-1}, \cdots, Z_1 = j_1) \\ &\qquad = \mathbb{E}(Z_{n+1}|Z_n = j) = \mathbb{E}\left[\sum_{i=1}^{j} X_i^{n+1}\right] = jm, \end{aligned}$$

由此可知

$$\mathbb{E}(Z_{n+1} \cdot m^{-(n+1)}|Z_n, Z_{n-1}, \cdots, Z_1) = Z_n \cdot m^{-n},$$

即$\{W_n\}$ 是鞅. □

因为$\{W_n\}$ 是非负鞅, 由鞅收敛定理, 存在可积非负随机变量W 使得$W_n \xrightarrow{\text{a.s.}} W$.

定理5.2.3 若$m > 1$, $\sigma^2 < \infty$, 则

(1) $\lim_n \mathbb{E}(W_n - W)^2 = 0$;

(2)

$$\mathbb{E}W = 1, \ \mathrm{var}W = \frac{\sigma^2}{m^2 - m};$$

(3) $\mathbb{P}(W = 0) = q$.

Proof. 对于(1), (2), 由鞅的Doob 不等式, 只需证明$\mathbb{E}W_n^2$ 有界就足够了.

$$\mathbb{E}W_n^2 = \frac{\mathbb{E}Z_n^2}{m^{2n}} = \frac{\sigma^2(1 - m^{-n})}{m^2 - m} + 1,$$

由此有

$$\sup \mathbb{E}W_n^2 = \frac{\sigma^2}{m^2 - m} + 1 < \infty.$$

(3) 令$r = \mathbb{P}(W = 0)$, 因为$\mathbb{E}W = 1$, 故$r < 1$. 另外注意到前面说过的当$Z_1 = k$ 时, $\{Z_n + 1 : n \geqslant 0\}$ 与$\{Z_n\}$ 的k 个独立复制的和是同分布的,

$$r = \mathbb{P}(W = 0) = \sum_{k \geqslant 0} \mathbb{P}(W = 0|Z_1 = k)\mathbb{P}(Z_1 = k)$$

$$= \sum_{k \geqslant 0} \mathbb{P}(W=0)^k \mathbb{P}(Z_1 = k) = f(r).$$

由此推出$r = q$. □

最后我们讨论分支过程的非零状态弱意义下的不变测度问题. 对于任意的Markov 链$X = (X_n)$, 状态空间E, 转移函数$p_{x,y}$. 取E 的子集F, 定义$q_{x,y} = p_{x,y}$, $x, y \in F$, 再取任意点$\Delta \notin F$, 定义$q(x, \Delta) = 1 - q(x, F)$, $x \in F$; $q(\Delta, \Delta) = 1$. 那么$\{q_{x,y} : x, y \in F \cup \{\Delta\}\}$ 是$F \cup \{\Delta\}$ 上的转移函数, 记对应的Markov 链为X^Δ. 显然Δ 是X^Δ 的陷阱. 满足

$$\pi_y = \sum x \in F \pi_x p_{x,y}, \ y \in F$$

如果的测度$\{\pi_x : x \in F\}$ 被称为X^Δ 的不变测度. 现在对于分支过程, 我们考虑它的非零状态上的Markov 链的不变测度存在唯一性问题.

引理5.2.3 设$p_1 > 0$, 则对任意$j \geqslant 1$, 有$p_{1,j}^{(n)} / p_{1,1}^{(n)}$ 递增.

Proof. $p_1 > 0$ 蕴含着$p_{1,1}^{(n)}$ 总是正的. 由母函数的性质

$$\frac{p_{1,j}^{(n+1)}}{p_{1,1}^{(n+1)}} = \frac{1}{j!} \frac{f_{n+1}^{(j)}(0)}{f'_{n+1}(0)}.$$

而由求导运算

$$\begin{aligned}
&f_{n+1}^{(1)}(s) = f'(f_n(s)) \cdot f'_n(s);\\
&f_{n+1}^{(2)}(s) = f''(f_n(s)) \cdot (f'_n(s))^2 + f'(f_n(s)) \cdot f_n^{(2)}(s)\\
&\qquad \cdots\cdots\\
&f_{n+1}^{(j)}(s) = a_{n,j}(s) + f'(f_n(s)) \cdot f_n^{(j)}(s),
\end{aligned}$$

其中$a_{n,j}$ 是一个系数非负的幂级数. 因此

$$\begin{aligned}
\frac{p_{i,j}^{(n+1)}}{p_{1,1}^{(n+1)}} &= \frac{1}{j!} \cdot \frac{a_{n,j}(0) + f'(f_n(0)) f_n^{(j)}(0)}{f'(f_n(0)) f'_n(0)}\\
&\geqslant \frac{1}{j!} \cdot \frac{f_n^{(j)}(0)}{f'_n(0)} = \frac{p_{1,j}^{(n)}}{p_{1,1}^{(n)}}.
\end{aligned}$$

完成证明. □

记$\pi_j = \lim_n \frac{p_{1,j}^{(n)}}{p_{1,1}^{(n)}}$, $j \geqslant 1$; $\mathcal{P}(s)$ 是$\{\pi_j : j \geqslant 1\}$ 的母函数; $\gamma = f'(q)$, 当$m = 1$ 时, $\gamma = 1$, 否则$\gamma < 1$.

定理5.2.4 $\{\pi_j : j \geqslant 1\}$ 满足以下不变性

$$\gamma\pi_j = \sum_{k\geqslant 1} \pi_k p_{k,j},\ j \geqslant 1. \tag{5.2.3}$$

因此, 若$p_0, p_1 > 0$ 时, 则对任意$j \geqslant 1$, $\pi_j < \infty$.

Proof. 因为

$$p_{1,1}^{(n+1)} = f'_{n+1}(0) = f'(f_n(0)) \cdot p_{1,1}^{(n)},$$

且$f_n(0) \to q$, 故

$$\frac{p_{1,1}^{(n+1)}}{p_{1,1}^{(n)}} = f'(f_n(0)) \uparrow \gamma.$$

再看C-K 方程

$$p_{1,j}^{(n+1)} = \sum_{k\geqslant 0} p_{1,k}^{(n)} p_{k,j},$$

两边同除$p_{1,1}^{(n)}$ 得

$$\frac{p_{1,j}^{(n+1)}}{p_{1,1}^{(n+1)}} \cdot \frac{p_{1,1}^{(n+1)}}{p_{1,1}^{(n)}} = \sum_{k\geqslant 0} \frac{p_{1,k}^{(n)}}{p_{1,1}^{(n)}} \cdot p_{k,j},$$

让$n \to \infty$, 由单调收敛定理推出定理的结论.

因为$\pi_1 = 1$, 故$1 \geqslant \gamma = \gamma\pi_1 = \sum_{k\geqslant 1} \pi_k p_{k,1}$, 而$p_{k,1} = p_0^{k-1} p_1 > 0$, 推出$\pi_k < \infty$. □

练习**5.2.5** 若$k \geqslant 1$, 且存在n 使得$p_{1,k}^{(n)} > 0$, 证明: $\pi_k > 0$.

现在设$p_0, p_1 > 0$. 在(5.2.3) 的两边乘s^j, 然后对$j \geqslant 1$ 求和得

$$\begin{aligned}
\gamma\mathcal{P}(s) &= \sum_{k\geqslant 1} \pi_k \sum_{j\geqslant 1} p_{k,j} s^j = \sum_{k\geqslant 1} \pi_k \left(\sum_{j\geqslant 0} p_{k,j} s^j - p_{k,0} \right) \\
&= \sum_{k\geqslant 1} \pi_k \mathbb{E}\left(s^{\sum_{i\geqslant 1}^{k} X_i^{(1)}} \right) - \sum_{k\geqslant 1} \pi_k p_0^k \\
&= \sum_{k\geqslant 1} \pi_k (f(s))^k - \mathcal{P}(p_0),
\end{aligned}$$

应用(5.2.3)

$$\gamma = \gamma\pi_1 = \sum_{k\geqslant 1} \pi_k p_{k,1} = \sum_{k\geqslant 0} \pi_k k p_0^{k-1} p_1 \geqslant \frac{p_1}{p_0} \mathcal{P}(p_0),$$

因此$\mathcal{P}(p_0)<\infty$, 即有下面的等式

$$\mathcal{P}(f(s))=\gamma\mathcal{P}(s)+\mathcal{P}(p_0). \tag{5.2.4}$$

再由(5.2.4), 可知对任意$n\geqslant 1$, 有$\mathcal{P}(f_n(p_0))<\infty$. 从而对任意$s\in[0,q)$, 有$\mathcal{P}(s)<\infty$, 也就是说等式(5.2.4) 对所有$0\leqslant s<q$ 成立.

对(5.2.4) 进行迭代得

$$\mathcal{P}(f_n(s))=\gamma^n\mathcal{P}(s)+(\gamma^{n-1}+\cdots\gamma+1)\mathcal{P}(p_0). \tag{5.2.5}$$

当$m\leqslant 1$ 时, $q=1$ 且$f_n(s)\uparrow 1$. 因此当$m=1$ 时, $\gamma=f'(1)=m=1$, 从而$\sum_{k\geqslant 1}\pi_k=\mathcal{P}(1)=\infty$, 即$\{\pi_k\}$ 是无穷测度; 当$m<1$ 时, $\gamma=f'(1)=m<1$, 那么$\sum_{k\geqslant 1}\pi_k=\mathcal{P}(1)<\infty$, 即$\{\pi_k\}$ 是有限测度.

定理5.2.5 $\{\pi_j:j\geqslant 1\}$ 作为方程(5.2.3) 的解在常数倍的意义下唯一.

Proof. 只需证明作为(5.2.4) 的解, 非负系数的幂级数$\mathcal{P}$ 在常数倍意义下是唯一的. 设$\mathcal{R}$ 是另外一个解. 首先由(5.2.4), $\mathcal{P}(0)=\mathcal{R}(0)=0$, 所以它们有相同的初始值. 另外对(5.2.5) 求导, 可知它们都满足下面方程

$$\mathcal{P}'(f_n(s))f_n'(s)=\gamma^n\mathcal{P}'(s). \tag{5.2.6}$$

对任何$s\in[0,q)$, 存在$k\geqslant 0$ 使得$f_k(0)\leqslant s\leqslant f_{k+1}(0)$, 因此

$$\begin{aligned}\frac{\mathcal{R}'(s)}{\mathcal{P}'(s)}&=\frac{\mathcal{R}'(f_n(s))}{\mathcal{P}'(f_n(s))}\leqslant\frac{\mathcal{R}'(f_{n+k+1}(0))}{\mathcal{P}'(f_{n+k}(0))}\\&=\frac{\mathcal{R}'(f_{n+k+1}(0))}{\mathcal{P}'(f_{n+k+1}(0))}\frac{\mathcal{P}'(f_{n+k+1}(0))}{\mathcal{P}'(f_{n+k}(0))}.\end{aligned}$$

先从第一个等号推出

$$\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}=\frac{\mathcal{R}'(f_n(0))}{\mathcal{P}'(f_n(0))},$$

右边与n 无关. 因此, 再应用(5.2.6) 得

$$\begin{aligned}\frac{\mathcal{R}'(s)}{\mathcal{P}'(s)}&\leqslant\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}\cdot\frac{\mathcal{P}'(f_{n+k+1}(0))}{\mathcal{P}'(f_{n+k}(0))}\\&=\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}\cdot\frac{f'_{n+k}(0)}{f'_{n+k+1}(0)}\cdot\gamma=\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}\cdot\frac{\gamma}{f(f'_{n+k}(0))}.\end{aligned}$$

让$n\to\infty$, 得

$$\frac{\mathcal{R}'(s)}{\mathcal{P}'(s)}\leqslant\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}.$$

类似地, 应用
$$\frac{\mathcal{R}'(s)}{\mathcal{P}'(s)}=\frac{\mathcal{R}'(f_n(s))}{\mathcal{P}'(f_n(s))}\geqslant\frac{\mathcal{R}'(f_{n+k}(0))}{\mathcal{P}'(f_{n+k+1}(0))},$$
可推得
$$\frac{\mathcal{R}'(s)}{\mathcal{P}'(s)}\geqslant\frac{\mathcal{R}'(0)}{\mathcal{P}'(0)}.$$
即导数$\mathcal{R}'$ 是导数$\mathcal{P}'$ 的常数倍, 而$\mathcal{R}$ 与$\mathcal{P}$ 的初值都是零, 故而$\mathcal{R}$ 是$\mathcal{P}$ 的常数倍, 定理得证. □

§5.3 可逆Markov 链与电路网络

设$X=\{X_n:n\geqslant 0\}$ 是状态空间S 上的一个既约Markov 链, 对任何两点$x,y\in S$, 存在道路$x_0=x,x_1,\cdots,x_{n-1},x_n=y$, 使得对任何$0\leqslant i\leqslant n-1$, $p(x_i,x_{i+1})>0$. 由Markov 性,
$$\mathbb{P}^x(X_i=x_i,1\leqslant i\leqslant n)=\Pi_{i=1}^n p(x_{i-1},x_i),$$
这个概率称为是X 沿上述道路从x 转移到y 的概率.

如果存在S 上的一个非负的非零函数π, 使得对任何$x,y\in S$, $\pi(x)p_{x,y}=\pi(y)p_{y,x}$, 那么称$X$ 是S 上可逆Markov 链, $(\pi(x):x\in S)$ 称为是对称测度. 注意我们总可以假设对任何$x\in S$, $\pi(x)>0$. 理由是如果$\pi(x)=0$, 那么对任何$y\in S$, $p_{y,x}=0$, 即把x 从S 中去除后X 依然是可逆的Markov 链. 设$x=x_0,x_1,\cdots x_{n-1},x_n=x$ 是x 回到x 的路径, 且对任何$1\leqslant i\leqslant n$, $p(x_{i-1},x_i)>0$. 那么
$$\frac{\pi(x_{i-1})}{\pi(x_i)}=\frac{p(x_i,x_{i-1})}{p(x_{i-1},x_i)},$$
因此推出
$$\Pi_{i=1}^n p(x_{i-1},x_i)=\Pi_{i=n}^1 p(x_i,x_{i-1}),$$
也就是说X 沿回路转移的概率与方向无关. 反过来, 如果X 沿回路转移的概率与方向无关(也说X 满足Kolmogorov 的回路测试), 那么X 是可逆的. 事实上, 任取$x\in S$, 取$\pi(x)>0$, 因为X 的既约性, 对任意$y\in S$, 都有一条如上的道路从x 转移到y, 因此我们定义
$$\pi(y):=\frac{\Pi_{i=1}^n p(x_{i-1},x_i)}{\Pi_{i=n}^1 p(x_i,x_{i-1})}\pi(x).$$
首先X 沿回路转移的概率与方向无关这个性质保证上述定义与路径无关, 其次可逆性是显然的. 这证明了下面定理.

定理5.3.1 既约Markov 链X 可逆当且仅当X 满足Kolmogorov 回路测试.

现在让我们看电路网络. 设$G=(S,E)$ 是一个连通的有限图, S 是节点的集合, $E\subset\{(x,y):x,y\in S\}$ 是边的集合, 说点$x,y\in S$ 相邻, 如果$(x,y)\in E$. 假设图是无向的, 即$(x,y)\in E$ 当且仅当$(y,x)\in E$. 注意这样同一条边在E 中两次出现. 再设r 是定义在E 上的一个正对称函数, 即对任何边$(x,y)\in E$, $r(x,y)=r(y,x)>0$. 称$r(x,y)$ 称为是边(x,y) 上的电阻率或电阻, r 称为是G 上的电阻函数. 定义边上的电导率或电导为电阻的倒数$c(x,y):=r(x,y)^{-1}$, $(x,y)\in E$. 自动地, 如果$(x,y)\notin E$, $x\neq y$, 那么不妨认为$r(x,y)=+\infty$, $c(x,y)=0$. 当然电阻和电导是互相唯一决定的. 图G 和函数r (或c) 一起称为是一个电路网络. 实际上电路网络必须在连接电源后才是真正的电路网络, 而与选择连接电源的节点有关.

给定这样一个电路网络. 定义

$$c(x):=\sum_{y\in S}c(x,y),\ p(x,y)=\frac{c(x,y)}{c(x)},\ x,y\in S.$$

那么$(p(x,y):\ x,y\in S)$ 是S 上转移函数, 且$c(x)p(x,y)=c(x,y)=c(y,x)=c(y)p(y,x)$, 因此所以诱导了一个$S$ 上的可逆Markov 链. 反过来, 设X 是有限集合S 上一个既约可逆Markov 链, 转移函数为$p(x,y)$, 对称测度是$\pi(x),x\in S$. 那么定义$E:=\{(x,y):p(x,y)>0\}$, 电导率

$$c(x,y):=\pi(x)p(x,y),\ x,y\in S.$$

那么由可逆性, c 是E 上对称正函数. 图G 及其电阻函数或电导函数一起(G,r) 或(G,c) 称为一个电路网络. 因此电路网络与可逆Markov 链是一一对应的. 直观地解释, Markov 链就如同在电路网络上跑动的电荷, 它自然地更容易向电导性更好的邻居转移.

下面我们就设(G,r) 是一个连通的电路网络, X 是对应的不可分可逆Markov 链. 下面我们主要讨论两者的关系. 用符号$x\sim y$ 表示x,y 相邻, 即$(x,y)\in E$. 设$H\subset S$, 说y 与H 相邻, 如果y 与H 中的某点相邻, 记为$y\sim H$. 定义$\overline{H}:=H\cup\{y\in S:y\sim H\}$, 称为是$H$ 的闭包, 它是H 及与H 相邻的点的全体. 因为图是连通的且若H 是S 的非空真子集, 那么H 一定也是其闭包$\overline{H}$ 的真子集.

定义5.3.1 说状态空间S 上的函数f 在点$x\in S$ 处调和, 如果

$$f(x)=\sum_{y:y\sim x}p(x,y)f(y).$$

说f 在集合$H \subset S$ 上调和, 如果它在H 的任何点上都调和.

显然f 在x 点处调和等价于$f(x) = \mathbb{E}^x f(X_1)$. 调和函数的定义借自古典分析中关于调和函数的定义, 因此这里调和函数也有类似的性质.

定理5.3.2 1. (最大值原理) 设H 是S 的一个连通子集, $f : G \to \mathbf{R}$ 在H 上调和. 如果f 在H 上达到最大值$\max f := \max_{x \in G} f(x)$, 那么$f$ 在$\overline{H}$ 上是常数.

2. (唯一性)设H 是连通图G 的有限真子图. 如果G 上的函数f, g 在H 上调和且在$G \backslash H$ 上相等, 那么$f = g$.

3. 设H 是S 的有限真子集, f_0 是$G \backslash H$ 上的函数, 那么存在G 上的函数f, 它在H 上调和且$f|_{G \backslash H} = f_0$.

证明. (1) 如果f 在点x 调和且$f(x) = \max f$, 那么由调和的定义$f(x) = \sum_{y:y \sim x} p(x, y) f(y)$ 推出当$y \sim x$ 时, $f(y) = \max f$. 因为H 是连通的, 故f 在$\overline{H}$ 上都等于$\max f$.

(2) 令$h = f - g$. 那么h 在H 上调和且在$G \backslash H$ 上为零. 假设h 在H 的某点处正, 则h 在H 的某点x_0 处达到最大值, 用K 表示x_0 在H 中的连通分支, 由(1) 推出h 在$\overline{K}$ 上是正常数. 现在因为G 连通且$K \neq G$, 故$\overline{K} \backslash K$ 非空. 又因为K 是H 的连通分支, 故$\overline{K} \backslash K \subset G \backslash H$, 因此导出矛盾. 推出$h \leqslant 0$, 但同理可以推出$h \geqslant 0$, 故$h = 0$.

(3) 设$\tau := \tau_{G \backslash H}$ 是集合$G \backslash H$ 的进入时. 那么我们来验证$f(x) := \mathbb{E}^x f_0(X_\tau)$ 符合要求. 首先$X_\tau \in G \backslash H$, 故定义有意义. 另外若$x \notin H$, 则$\mathbb{P}^x(\tau = 0) = 1$, 故$f(x) = f_0(x)$; 若$x \in H$, 则由Markov 性$\mathbb{P}^x(\tau \geqslant 1) = 1$, 故$f(x) = \mathbb{E}^x(f_0(X_\tau)\theta_1) = \mathbb{E}^x f_0(X_1)$, 即$f$ 在x 处调和. □

现在取非空子集$A, B \subset S$, 满足$A \cap B = \varnothing$ 且$(A \cup B)^c$ 有限连通. 加一个电源连接A, B 使得A 上每个点的电压是1, B 上每个点的电压是0. 由物理的观察, 存在电压函数, 也就是S 上的一个函数v 使得$v|_A = 1$, $v|_B = 0$, 及电流函数i, 它是边E 上的函数, 和电阻率一起满足下面的Kirchhoff 定律和Ohm 定律:

(1) Kirchhoff 节点定律: 对于不直接联在电池上的点x, 即$x \notin A \cup B$, 有

$$\sum_{y \sim x} i(x, y) = 0,$$

即流出的电流和等于流入的电流和.

(2) Kirchhoff 回路定律: 如果$x = x_0 \sim x_1 \sim \cdots \sim x_n = x$ 是一个回路, 那么

$$\sum_{i=1}^{n} i(x_{i-1}, x_i) r(x_{i-1}, x_i) = 0.$$

(3) Ohm 定律: 如果$x \sim y$, 那么$v(x) - v(y) = i(x, y) r(x, y)$.

Ohm 定律蕴含着i 是E 上的反对称函数: $i(x, y) = -i(y, x)$. 另外Ohm 定律等价于Kirchhoff 回路定律. 事实上, Ohm 定律显然蕴含着Kirchhoff 回路定律, 反过来, 如果Kirchhoff 回路定律成立, 取$b \in B$, 对任何$x \notin A \cup B$, 存在x 到b 的路径$b = x_0 \sim x_1 \sim \cdots \sim x_n = x$, 定义

$$v(x) := \sum_{i=1}^{n} i(x_i, x_{i-1}) r(x_i, x_{i-1}).$$

由Kirchhoff 回路定律, $v(x)$ 与选择的路径无关, 因为$v|_B = 0$, 它也与b 的选择无关. 容易推出v, i, r 满足Ohm 定律. 下面定理说明电压可以用Markov 链来表示.

定理5.3.3 (1) 电压函数v 在$(A \cup B)^c$ 上调和. (2) 对任何$x \in S$, $v(x) = \mathbb{P}^x(\tau_A < \tau_B)$.

证明. (1) 取$x \notin A \cup B$, 由节点定律, $\sum_{y \sim x} i(x, y) = 0$, 再由Ohm 定律$i(x, y) = (v(x) - v(y)) c(x, y)$, 故$\sum_{y \sim x} (v(x) - v(y)) c(x, y) = 0$. 因此$v(x) c(x) = \sum_{y \sim x} c(x, y) v(y)$, 推出$v$ 在x 处调和. (2) 容易验证$x \mapsto \mathbb{P}^x(\tau_A < \tau_B)$ 也在$(A \cup B)^c$ 上且在$A \cup B$ 上与v 一样, 由唯一性知两者恒等. □

因为i 反对称, 故

$$\sum_{x, y \in S : x \sim y} i(x, y) = \sum_{(x, y) \in E} i(x, y) = 0.$$

由Kirchhoff 节点定律,

$$\sum_{x \in A} \sum_{y \sim x} i(x, y) = \sum_{x \in B} \sum_{y \sim x} i(y, x),$$

左边是由A 流出的电流, 右边是流入B 的电流, 他们是相等的, 记它为$i(A, B)$. 整个电路的电压是1, 所以A, B 间的电阻是$i(A, B)^{-1}$, 或者说电导是$i(A, B)$, 分别称为是A, B 间的有效电阻和有效电导, 记为$R_{\text{eff}}(A, B)$, $C_{\text{eff}}(A, B)$. A, B 间有效电导等于在A, B 两端加单位电压后的电流量, 有效电阻等于从A 流入单位电流时A, B 两端的电压差.

有效电导(或电阻)可以通过以下三种电路等价变换的方式来计算, 请读者自行用Ohm 定律和Kirchhoff 定律验证.

1. 串联电路: 两个串联的电阻r_1, r_2 的电路的电阻等于一个$r_1 + r_2$;
2. 并联电路: 两个并联的电阻r_1, r_2 的电路的电阻等于$\frac{r_1 r_2}{r_1+r_2}$;
3. 星形电路: 设a_1, a_2, a_3 与b 是如图连接的星形, 那么它可以变换为a_1, a_2, a_3 三角形连接, 且

$$c(a_1, a_2) = \frac{\gamma}{c(b, a_3)}, c(a_2, a_3) = \frac{\gamma}{c(b, a_1)}, c(a_3, a_1) = \frac{\gamma}{c(b, a_2)},$$

其中

$$\gamma = \frac{\Pi_{i=1}^3 c(b, a_i)}{\sum_{i=1}^3 c(b, a_i)}.$$

现在设$A = \{a\}$, $a \notin B$. 在记号中直接用a 代替A. 下面的定理建立了有效电导与概率的关系.

定理5.3.4 设$\tau_a^+ := \inf\{n \geqslant 1 : X_n = a\}$. a 出发回到a 前到达B 的概率

$$\mathbb{P}^a(\tau_B < \tau_a^+) = \frac{C_{\text{eff}}(a, B)}{c(a)}.$$

证明. 加电池, 使得$v(a) = 1$, $v|_B = 0$. 首先由上面的定理, $v(x) = \mathbb{P}^x(\tau_a < \tau_B)$. 显然$\tau_a^+ = \tau_a \circ \theta_1 + 1 \geqslant 1$, 而从$a$ 出发时, $\tau_B \geqslant 1$. 因此

$$\begin{aligned}
\mathbb{P}^a(\tau_B < \tau_a^+) &= \mathbb{E}^a(1_{\{\tau_B < \tau_a\}} \circ \theta_1) \\
&= \sum_{x \sim a} \mathbb{P}^x(\tau_B < \tau_a) p(a, x) \\
&= \sum_{x \sim a} (1 - v(x)) p(a, x) \\
&= \frac{1}{c(a)} \sum_{x \sim a} (v(a) - v(x)) c(a, x) \\
&= \frac{1}{c(a)} \sum_{x \sim a} i(a, x) = \frac{C_{\text{eff}}(a, B)}{c(a)}.
\end{aligned}$$

□

下面的例子说明怎样用这个定理来计算概率.

例5.3.1 设a_1, a_2, a_3, a_4 是正方形的四个顶点, 每边的电导都是1. 那么相邻两点a_1, a_2 间的有效电导为$C_{\text{eff}}(a_1, a_2) = 1 + \frac{1}{3} = \frac{4}{3}$. 因此$a_1$ 出发回到a_1 前到达a_2 的概率$\mathbb{P}^{a_1}(\tau_{a_2} < \tau_{a_1}^+) = \frac{C_{\text{eff}}(a_1, a_2)}{c(a_1)} = \frac{2}{3}$. 而相对的两点$a_1, a_3$ 间的有效电导为$C_{\text{eff}}(a_1, a_3) = \frac{1}{2} + \frac{1}{2} = 1$, 而对应的概率为$\frac{1}{2}$. 这个例子比较简单, 上面的概率可直接计算. ▮

例5.3.2 设有一个正方体网络, 每边的电导依然是1. 让a, b 是相邻两点. 通过一系列的等价变换, $C_{\text{eff}}(a,b) = \frac{12}{7}$, a 出发回到a 前到达b 的概率为

$$\mathbb{P}^a(\tau_b < \tau_a^+) = \frac{C_{\text{eff}}(a,b)}{c(a)} = \frac{4}{7}.$$

对于这个例子, 这个概率的直接计算几乎是不可能的. ∎

现在设S 是可列集, 而图(S,E) 是连通且局部有限的, 即每个顶点的邻居有限多, c 是E 上的对称正函数. 那么仍然如上对应有一个S 上的可逆Markov 链X. 取S_n 是递增到S 的有限子集列, $B_n := S_n^c$, $a \in S_1$. 上面的定理仍然成立:

$$\mathbb{P}^a(\tau_{B_n} < \tau_a^+) = \frac{C_{\text{eff}}(a,B_n)}{c(a)}.$$

当$n \to +\infty$ 时, $\tau_{B_n} \uparrow +\infty$. 因此$\mathbb{P}^a(\tau_a^+ = \infty) = \frac{1}{c(a)} \lim_n C_{\text{eff}}(a,B_n)$. 称极限$\lim_n C_{\text{eff}}(a,B_n)$ 是a 到∞ 的电导, 记为$C_{\text{eff}}(a,\infty)$. 因为X 常返当且仅当$\mathbb{P}^a(\tau_a^+ < \infty) = 1$, 故推出$X$ 常返当且仅当$C_{\text{eff}}(a,\infty) = 0$, 即$a$ 与无穷远的电导为零.

例5.3.3 设$X^{(1)}$ 是直线整数格点上的简单随机游动. 对应的网络是整数格点网络, 每边的电导是1. 取$S_n := \{-n, -n+1, \cdots, n-1, n\}$, $B_n := S_n^c$. 那么$C_{\text{eff}}(0,B_n) = \frac{2}{n}$, 因此$C_{\text{eff}}(0,\infty) = 0$. 故$X^{(1)}$ 常返. ∎

定义5.3.2 给定一个图$G = (S,E)$. 设A, B 是S 的非空子集且$A \cap B = \varnothing$. E 上的一个反对称函数j 称为是A 到B 的流, 如果对任何$x \notin A \cup B$,

$$\sum_{(x,y)\in E} j(x,y) = 0.$$

类似地验证

$$\sum_{x\in A}\sum_{y\sim x} j(x,y) = \sum_{x\in B}\sum_{y\sim x} j(y,x),$$

因此令$j(A,B) := \sum_{x\in A}\sum_{y\sim x} j(x,y)$ 称为是A 到B 的流量. 说j 是A 到B 的一个单位流, 是指$j(A,B) = 1$. 给定图G 上的电阻r, 对任何A 到B 的流j 定义j 产生的能量

$$\mathcal{E}(j) = \mathcal{E}_{A,B}(j) := \frac{1}{2}\sum_{(x,y)\in E} j(x,y)^2 r(x,y).$$

下面我们证明Thompson 定律.

定理5.3.5 (Thompson) 设i 是在A 上加电压$v > 0$, B 接地时产生的单位电流, 那么在所有A 到B 的单位流中, i 产生的能量最小.

证明. 设j 是A 到B 的单位流, 令$i' := j - i$. 因为i, j 是流量相同的流, 故i' 是A 到B 的一个零流量的流. 计算能量

$$\begin{aligned}\mathcal{E}(j) &= \frac{1}{2}\sum_{(x,y)\in E} j(x,y)^2 r(x,y)\\ &= \mathcal{E}(i) + \mathcal{E}(i') + \sum_{(x,y)\in E} i(x,y)i'(x,y)r(x,y).\end{aligned}$$

而最后一项是零, 因为i 是电流, 由Ohm 定律, 流的定义, v 在A 上是常数v, 在B 上是零等事实得

$$\begin{aligned}\sum_{(x,y)\in E} i(x,y)i'(x,y)r(x,y) &= \sum_{(x,y)\in E} (v(x) - v(y))i'(x,y)\\ &= \sum_{(x,y)\in E} v(x)i'(x,y) - \sum_{(x,y)\in E} v(y))i'(x,y)\\ &= 2\sum_{(x,y)\in E} v(x)i'(x,y)\\ &= 2\sum_{x\in S} v(x)\sum_{y\sim x} i'(x,y)\\ &= 2\sum_{x\in A\cup B} v(x)\sum_{y\sim x} i'(x,y)\\ &= 2v\cdot i'(A,B) = 0,\end{aligned}$$

因此$\mathcal{E}(j) = \mathcal{E}(i) + \mathcal{E}(i') \geqslant \mathcal{E}(i)$, 即电流产生的能量最小. □

由上面的证明过程可以看出: 如果i 是这样的一个单位电流, 那么

$$\mathcal{E}(i) = v\cdot i(A,B) = R_{\text{eff}}(A,B),$$

即单位电流产生的能量等于A, B 间的有效电阻. 确切地说, 下面结论成立.

推论5.3.1 如果i 是A 到B 的单位电流, 那么

$$R_{\text{eff}}(A,B) = \frac{1}{2}\sum_{(x,y)\in E} i(x,y)^2 r(x,y).$$

由此容易地推出下面的Rayleigh 单调原理.

定理5.3.6 (Rayleigh) 有效电阻关于电阻函数是递增的.

证明. 设r, r' 是G 上两个电阻函数, $r' \geqslant r$ 是指对任何$(x, y) \in E$, $r'(x, y) \geqslant r(x, y)$. 用$R_{\text{eff}}(A, B)$ 与$R'_{\text{eff}}(A, B)$ 分别表示A, B 间对应于电阻函数r 与r' 的有效电阻, i 与i' 表示对应的单位电流, $\mathcal{E}$ 与$\mathcal{E}'$ 对应的能量. 那么由Thompson 定律,

$$R'_{\text{eff}}(A, B) = \mathcal{E}'(i') \geqslant \mathcal{E}(i') \geqslant \mathcal{E}(i) = R_{\text{eff}}(A, B),$$

其中第一个$\geqslant$ 号是因为$r' \geqslant r$, 第二个等号是因为对于电路网络(G, r), i 是电流而i' 只是流. □

把在G 上任何两点x, y 的连线去掉称为是一个断路, 相反, 把在G 上两点x, y 间接一个电阻为零的线称为是一个短路. 断路不妨看成电阻$r(x, y) \uparrow +\infty$, 因此电阻函数是增加的, 而短路不妨看成$r(x, y) \downarrow 0$, 因此电阻函数减少. 这样由Rayleigh 单调定理推出断路后, 有效电阻只会增加; 短路后, 有效电阻只会减少. 因为Markov 链X 暂留当且仅当有效电导$C_{\text{eff}}(a, \infty) > 0$, 等价于有效电阻$R_{\text{eff}}(a, \infty) < \infty$, 显然此有效电阻也关于电阻函数递增, 故有下面的比较定理.

定理5.3.7 暂留的Markov 链短路后仍然是暂留的, 常返的Markov 链断路后仍然是常返的.

换句话说, 如果某个短路后的链常返, 那么原来的链也常返, 如果某个断路后的链暂留, 那么原来的链也暂留. 如果G 是无限网络, 那么物理意义上的电路是不存在了, 谈论电流是没有意义的. 但是流依然可以定义, 边E 上的反对称函数j 称为是一个a 到无穷远的单位流, 如果对任何$x \in S$, 有$\sum_{y \sim x} |j(x, y)| < \infty$ 且$\sum_{y \sim x} j(x, y) = 1_{\{a\}}(x)$. 定义$j$ 产生的能量

$$\mathcal{E}(j) := \frac{1}{2} \sum_{(x,y) \in E} j(x, y)^2 r(x, y).$$

定理5.3.8 (T. Lyons, 1983) 电路网络上的Markov 链暂留当且仅当存在一个从任意点a 到无穷远处的能量有限的单位流.

证明. 同样, 取S_n 是递增到S 的有限子集列, $B_n := S_n^c$, $a \in S_1$. 用i_n 表示在a 和B_n 间加电压产生的从a 流入的单位电流, 那么$R_{\text{eff}}(a, B_n) = \mathcal{E}(i_n)$, 因此Markov 链暂留当且仅当$\sup_n \mathcal{E}(i_n) < \infty$. 如果$j$ 是一个从任意点a 到无穷远处的能量有限的单位流, 那么j 是从a 到B_n 的单位流, 由Thompson 定律, $\mathcal{E}(i_n) \leqslant \mathcal{E}(j) < \infty$. 反之, 若$\sup \mathcal{E}(i_n) < \infty$, 则看出对任何边$(x, y) \in E$, $\{i_n(x, y) : n \geqslant 1\}$ 有界, 而E 是可列集, 因此存在子列$\{k_n\}$ 使得$\{i_{k_n}(x, y) : n \geqslant 1\}$ 在每条边上都收敛,

记$j(x,y) := \lim_n i_{k_n}(x,y)$, 那么$j$ 是a 到无穷远的单位流且由Fatou 引理, $\mathcal{E}(j) \leqslant \sup_n \mathcal{E}(i_n)$. □

让我们用T. Lyons 对于Polya 定理的一个漂亮证明结束这一节.

例5.3.4 Polya 定理: $\mathbf{Z}^d$ 上简单随机游动当$d \leqslant 2$ 常返, 当$d \geqslant 3$ 时暂留. 由比较定理, 只需对$d = 2, 3$ 的情况证明. 如果$d = 2$, 只需证明某个短路后的链是常返就够了. 对任何$n \geqslant 0$, 把格点$\{(x,y) : |x| \vee |y| = n\}$ 短路看成一个点a_n, 因为每条边的电导是1, 而a_n 有$4(2n+1)$ 条边连接a_{n+1}, 故$c(a_n, a_{n+1}) = 4(2n+1)$, 从$a_0$ 到无穷远的有效电阻为

$$R_{\text{eff}}(a_0, \infty) = \sum_{n \geqslant 0} \frac{1}{4(2n+1)} = +\infty,$$

因此短路后的链是常返的.

设$d = 3$, 对任何边(x,y), 把它等同于$\mathbf{R}^3$ 的一个x 指向y 的单位向量, v_{xy} 表示原点指向(x,y) 中点的向量, S_{xy} 表示中心在v_{xy}, 垂直于(x,y) 且边与坐标轴平行的单位正方形. 定义$j(x,y)$ 等于S_{xy} 在原点为中心的单位球面上径向投影的面积, 并赋予和向量(x,y) 与v_{xy} 内积相同的符号. 简单的几何观察说明当$x \neq 0$ 时, $\sum_{y \sim x} j(x,y) = 0$. 因此$j$ 是原点到无穷远的一个流. 显然存在一个常数A, 使得$|j(x,y)| \leqslant A \cdot |v_{xy}|^{-2}$, 其中$|\cdot|$ 表示向量长度. 另外存在一个常数B 使得到原点的距离在n 与$n+1$ 间的边的数量不超过Bn^2. 因此j 产生的能量

$$\mathcal{E}(j) = \frac{1}{2} \sum_{(x,y)} j(x,y)^2 \leqslant \sum_n A^2 n^{-4} B n^2 < \infty,$$

即j 是能量有限的流, 故而对应的链是暂留的. ▮

习 题

1. 设$a \notin B$. 记ξ 从a 出发在到达B 前回到a 的次数, 求ξ 的分布.

2. 设$x, y \notin B$, $g(x,y)$ 是从x 出发在到达B 前到y 的平均次数. 证明:

$$c(x)g(x,y) = c(y)g(y,x).$$

3. 考虑具有顶点$\{0,1\}^n$ 的超正方体上的简单电网, 证明: $R_{\text{eff}}(a,b)$ 是距离$d(a,b)$ 的递增函数.

4. 给定一个电路网络G, r, X 是对应的Markov 链, $a \in S$, $Z \subset S$, $a \notin Z$. 对任何$x, y \in S$, $x \sim y$, 用S_{xy} 表示X 在碰到Z 之前从x 转移到y 的次数, 即$S_{xy} := \sum_{n \geqslant 0} 1_{\{X_n = x, Z_{n+1} = y, n < \tau\}}$, 其中$\tau$ 是Z 的首中时. 证明: 如果在a 和Z 间加电压形成一个a 到Z 的单位电流i, 那么$\mathbb{E}^a(S_{xy} - S_{yx}) = i(x, y)$.

5. 证明: (1) 对$a \in S$, 有效电导的极限$\lim_{n\to\infty} C_{\text{eff}}(a, S_n^c)$ 与递增到S 的子集列$\{S_n\}$ 的选取无关. (2) 有效电阻$R_{\text{eff}}(a, +\infty$ 的有限性与a 的选取无关.

6. 用Kirchhoff 定律与Ohm 定律证明串联电路, 并联电路与星形电路的电阻计算公式.

7. 取点$a \in S$, 边E 的子集Π 称为a 与∞ 的断开集, 如果去掉Π 后, a 所在的分支是有限图. 证明: (Nash-Williams 准则) 设$\{\Pi_n\}$ 是互不相交的a 与∞ 的断开集, 那么
$$R_{\text{eff}}(a, \infty) \geqslant \sum_n \left(\sum_{(x,y) \in \Pi_n} c(x, y) \right)^{-1}.$$

8. 用T_k 表示每个顶点都有k 个后代的正则树. 证明: 当$k \geqslant 2$ 时, T_k 上的简单随机游动是暂留的.

9. 证明: 平面上的六角形(蜂窝)格点网络和三角形网络上的简单随机游动是常返的.

10. 对任何$r > 0$, 构造一个局部有限的无限网络包含顶点a 使得每个边的电阻都是1 且$R_{\text{eff}}(a, \infty) = r$.

第六章　Levy 过程

前面介绍的随机过程都是离散时间的随机过程, 也就是随机序列, 它的结构比较简单, 实际上它的可测性比较简单, 因为所涉及基本上是可列运算. 但是连续时间的随机过程是不一样的, 它的方法与随机序列所用的方法有很大的不同. 我们将在后面介绍两类重要的连续时间的随机过程: Poisson 过程与Brown 运动.

设$(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, $X = (X_t : t \geqslant 0)$ 称为是一个d-维随机过程, 如果对任何$t \geqslant 0$, X_t 是一个d-维随机变量. 当$d = 1$ 时, 简称为实值随机过程.

§6.1　平稳独立增量过程

最简单的随机过程是平稳独立增量过程, 它是独立随机变量和的推广. 设$\{\xi_n : n \geqslant 1\}$ 是独立同分布随机序列, $X_0 = 0$, $X_n = \xi_1 + \cdots + \xi_n$. 那么$X = (X_n : n \geqslant 0)$ 满足

(1) 对任意$n > m \geqslant 0$, 增量$X_n - X_m$ 独立于$\{X_1, \cdots, X_m\}$;

(2) 对任意$n > m \geqslant 0$, 增量$X_n - X_m$ 与$X_{n-m} - X_0$ 同分布.

其中的第一个性质称为独立增量性, 第二个性质称为平稳增量性, 合起来X 被称为是平稳独立增量序列. 因此对连续时间随机过程, 我们有下面自然的定义.

定义6.1.1 d-维随机过程$X = (X_t : t \geqslant 0)$ 被称为是平稳独立增量过程, 如果它满足

(1) 对任何$t > s \geqslant 0$, $X_t - X_s$ 独立于$\mathcal{F}_s := \sigma(X_u : u \leqslant s)$;

(2) 对任何$t > s \geqslant 0$, $X_t - X_s$ 与$X_{t-s} - X_0$ 同分布.

不失一般性, 对于平稳独立增量过程我们总是设$X_0 = 0$, 就是过程总是从0 点出发. 平稳独立增量过程也被称为空间齐次过程, 它是最重要的一类随机过程, 其中包含有众所周知的Poisson 过程与Brown 运动.

引理6.1.1 设$X = (X_t : t \geqslant 0)$ 为d-维平稳独立增量过程, 对任意$t \geqslant 0$, 定义μ_t 为X_t 的分布. 那么$\mathbf{R}^d$ 上的概率测度族$\{\mu_t : t \geqslant 0\}$ 满足对任何$t, s \geqslant 0$,

$$\mu_{t+s} = \mu_t * \mu_s. \tag{6.1.1}$$

Proof. 因为$X_{t+s}=(X_{t+s}-X_s)+X_s$, 由平稳独立增量性, 右边两个随机变量独立且$X_{t+s}-X_s$ 的分布就是μ_t, 故$\mu_{t+s}=\mu_t*\mu_s$. □

引理6.1.2 设$\{\mu_t : t\geqslant 0\}$ 是$\mathbf{R}^d$ 上的概率测度族, 满足对任何$t,s\geqslant 0$, $\mu_{t+s}=\mu_t*\mu_s$, 那么存在概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 以及其上的d-维平稳独立增量过程$X=(X_t : t\geqslant 0)$ 使得对任何$t\geqslant 0$, X_t 的分布是μ_t.

Proof. 证明显然是需要使用Kolmogorov 的相容性定理. 首先对于任何的概率分布μ, 对于$x\in\mathbf{R}^d$, $B\in\mathcal{B}(\mathbf{R}^d)$, 定义$p(x,B)=\mu(B-x)$, 那么$p$ 是一个核, 即固定x, p 关于B 是一个概率测度, 固定B, p 关于x 是可测函数. 因此对于$0<t_1<t_2<\cdots<t_n$, 定义乘积空间$(\mathbf{R}^d)^n$ 上的概率测度$\mu_{t_1,\cdots,t_n}$ 为

$$\begin{aligned}&\mu_{t_1,\cdots,t_n}(A) &(6.1.2)\\ :=&\int_{(\mathbf{R}^d)^n}1_A(x_1,\cdots,x_n)\mu_{t_1}(dx_1)\mu_{t_2-t_1}(dx_2-x_1)\cdots\mu_{t_n-t_{n-1}}(dx_n-x_{n-1})\\ =&\int_{(\mathbf{R}^d)^n}1_A(x_1,x_1+x_2,\cdots,x_1+\cdots+x_n)\mu_{t_1}(dx_1)\mu_{t_2-t_1}(dx_2)\cdots\mu_{t_n-t_{n-1}}(dx_n),\end{aligned}$$

其中A 是$(\mathbf{R}^d)^n$ 的任意Borel 集. 由$\{\mu_t\}$ 的半群性不难验证概率测度族

$$\{\mu_{t_1,\cdots,t_n} : n\geqslant 1, 0<t_1<\cdots<t_n\}$$

是$\mathbf{R}^d$ 上的相容的有限维分布族, 因此由Kolmogorov 相容定理, 存在概率空间$(\Omega,\mathcal{F},\mathbb{P})$ 及其上随机过程$X=(X_t : t\geqslant 0)$, 使得对任何$n\geqslant 1$, $0<t_1<\cdots<t_n$, $(\mathbf{R}^d)^n$ 的Borel 集A 有

$$\begin{aligned}&\mathbb{P}((X_{t_1},\cdots,X_{t_n})\in A) &(6.1.3)\\ =&\int_{(\mathbf{R}^d)^n}1_A(x_1,\cdots,x_n)\mu_{t_1}(dx_1)\mu_{t_2-t_1}(dx_2-x_1)\cdots\mu_{t_n-t_{n-1}}(dx_n-x_{n-1})\\ & &(6.1.4)\end{aligned}$$

那么$(X_{t_1},X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}})$ 的联合分布为

$$\begin{aligned}&\mathbb{P}((X_{t_1},X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}})\in A)\\ =&\int_{(\mathbf{R}^d)^n}1_A(x_1,\cdots,x_n)\mu_{t_1}(dx_1)\mu_{t_2-t_1}(dx_2)\cdots\mu_{t_n-t_{n-1}}(dx_n),\end{aligned}$$

从此式可以容易地推出X 是平稳独立增量过程且X_t 的分布是μ_t. □

从上面两个引理可以看出平稳独立增量过程等同于一个满足(6.1.1) 的概率测度族.

§6.2 Lévy-Khinchin 表示

给定$\mathbf{R}^d$ 上这样一个概率测度族$\{\mu_t\}$, 用$\hat{\mu}_t$ 表示μ_t 的特征函数, 那么由半群性以及特征函数的性质得

$$\hat{\mu}_{t+s}(x) = \hat{\mu}_t(x) \cdot \hat{\mu}_s(x), \ x \in \mathbf{R}^d. \tag{6.2.1}$$

这告诉我们$t \mapsto \hat{\mu}_t(x)$ 有指数函数的乘性. 当它未必是指数函数, 简单的数学分析知识可以得知当$\hat{\mu}_t$ 关于t 有连续性条件时, 它必是一个指数型函数.

定义6.2.1 一个平稳独立增量过程$X = (X_t)$ 称为Lévy 过程, 如果当$t \downarrow 0$ 时, X_t 依概率收敛于零.

因为X_t 依概率收敛于零等价于依分布收敛于零, 因此平稳独立增量过程X 是Lévy 过程当且仅当其对应的测度族$\{\mu_t\}$ 满足当$t \downarrow 0$ 时μ_t 弱收敛于零点的单点测度δ_0.

定义6.2.2 $\mathbf{R}^d$ 上一个概率测度族$\{\mu_t : t \geqslant 0\}$ 称为是卷积半群如果它满足(6.1.1) 且当$t \downarrow 0$ 时μ_t 弱收敛于δ_0, 即对任何$\mathbf{R}^d$ 上有界连续函数f 有$\mu_t(f) = f(0)$.

现在设$\{\mu_t\}$ 是一个卷积半群, 那么当$t \downarrow 0$ 时有$\hat{\mu}_t \to 1$. 故而存在函数$\phi(x)$ 使得

$$\hat{\mu}_t(x) = \exp(-t\phi(x)), \ x \in \mathbf{R}^d, \tag{6.2.2}$$

因为特征函数是连续的, 故ϕ 是$\mathbf{R}^d$ 上的复数值连续函数, 还因为$|\hat{\mu}_t| \leqslant 1$, ϕ 的实部是非负的. 函数ϕ 被称为卷积半群$\{\mu_t\}$ 或对应的过程的Lévy 指数. 由特征函数唯一性, Lévy 指数唯一决定卷积半群$\{\mu_t\}$. 那么一个自然的问题就是Lévy 指数有什么样的表示形式. 先看两个例子.

例6.2.1 第一个例子是经典的热半群, $\mathbf{R}^d$ 上热传导方程

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u$$

的基本解为

$$p_t(x) = \left(\frac{1}{\sqrt{2\pi t}}\right)^d \exp\left(-\frac{|x|^2}{2t}\right), \ t > 0, x \in \mathbf{R}^d.$$

它是期望为零方差为t 的正态分布的密度函数, 定义概率测度

$$\mu_t(dx) = p_t(x)dx.$$

那么$\{\mu_t : t \geqslant 0\}$ 是卷积半群. 事实上, 只需验证对任何$t, s > 0$, $p_{t+s} = p_t * p_s$. 这就是说独立的方差分别为t, s 的正态分布随机变量的和是方差为$t + s$ 的随机变量, 这由正态分布的再生性推出. 然后众所周知

$$\hat{\mu}_t(x) = \hat{p}_t(x) = e^{-t\frac{|x|^2}{2}},$$

因此热半群的Lévy 指数是$\phi(x) = \frac{|x|^2}{2}$.

第二个例子的名字也是熟悉的Poisson 半群. 设$\lambda > 0$, 对于$t > 0$, μ_t 是期望为λt 的Poisson 分布. 由Poisson 分布的再生性, 得对任何$t, s > 0$ 有$\mu_t * \mu_s = \mu_{t+s}$, 即$\{\mu_t\}$ 是卷积, 被称为Poisson 半群. 其特征函数$\hat{\mu}_t(x) = \exp(-t\lambda(1 - e^{ix}))$, 因此其Lévy 指数为$\phi(x) = \lambda(1 - e^{ix})$. ∎

例6.2.2 (复合Poisson 半群) 设J 是$\mathbf{R}^d$ 上的一个概率测度, $\lambda > 0$ 是一个常数, 对$t \geqslant 0$, 令

$$\mu_t := e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} J^{*n},$$

其中J^{*n} 是n-重卷积, 而$J^{*0} := \epsilon_0$, 容易验证μ_t 是概率, 且对$t, s \geqslant 0$, $\mu_{t+s} = \mu_t * \mu_s$, 而对$f \in C_b(\mathbf{R}^d)$, 由控制收敛定理,

$$\lim_{t \downarrow 0} \mu_t(f) = \lim_{t \downarrow 0} e^{-\lambda t}(f(0) + \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} J^{*n}(f)) = f(0).$$

因此$\{\mu_t\}$ 是一个卷积半群, 称为是复合Poisson 半群, λ 称为过程的强度, J 称为是跃度分布. $\{\mu_t\}$ 的特征函数为

$$\hat{\mu}_t = e^{-\lambda t} e^{\lambda t \hat{J}} = e^{-\lambda(1-\hat{J})t},$$

其中$\hat{J}$ 是J 的Fourier 变换. 因此复合Poisson 半群的Lévy 指数是$\lambda(1 - \hat{J})$. ∎

实际上任何一个卷积半群的Lévy 指数就是由Brown 运动与复合Poisson 半群两个部分组成的. 下面我们就$d = 1$ 时来给出它的表示. 容易验证两个卷积半群的卷积仍然是卷积半群, 因此Lévy 指数的和也仍然是Lévy 指数. 我们引入无穷可分分布, 它实际上是卷积半群另外一个名词. $\mathbf{R}^d$ 上的分布μ 称为是无穷可分的, 如果对任何$n \geqslant 1$, 存在分布μ_n 使得$\mu = \mu_n^{*n}$. 对应的分布函数称为无穷可分分布函数, 而对应的特征函数称为是无穷可分特征函数. 显然一个特征函数ϕ 无穷可分当且仅当对任何n 存在特征函数ϕ_n 使得$\phi = \phi_n^n$. 显然, 如果$\{\pi_t\}$ 是对卷积具有半群性, 则任何π_t 都是无穷可分分布. 先给出一个引理, 它可以由初等的不等式$1 - \cos 2\alpha \leqslant 4(1 - \cos \alpha)$ 推出. 留给读者作为练习.

引理6.2.1 设f 是一个特征函数, 且实部为u, 则对$\xi \in \mathbf{R}$,

$$1 - u(2\xi) \leqslant 4(1 - u(\xi)).$$

下面的定理是重要的.

定理6.2.1 设$\{f_n\}$ 是特征函数列, 则函数列f_n^n 有一个连续极限(记为f) 当且仅当$n(1 - f_n)$ 有连续极限(记为ϕ). 这时候$f = e^{-\phi}$.

证明. 先设$n(1 - f_n)$ 有一个连续极限ϕ, 则$f_n \longrightarrow 1$ 且在有限区间上一致. 那么对任何$a > 0$, 对充分大的n, $|1 - f_n| < r < 1$ 在$[-a, a]$ 上成立. 因此由Taylor 展开, 对$|\xi| \leqslant a$,

$$\begin{aligned} n \log f_n(\xi) &= n \log(1 - (1 - f_n(\xi))) \\ &= -n(1 - f_n(\xi)) - \frac{n}{2}(1 - f_n(\xi))^2 - \cdots , \end{aligned}$$

右边第一项的极限为$-\phi(\xi)$, 其余的为零. 由一致性及a 的任意性推出$n \log f_n$ 点点收敛于$-\phi$, 即$f_n^n \to e^{-\phi}$.

反之设f_n^n 有一个连续极限f, 我们先证f 不能有零点, 因特征函数列$|f_n|^{2n}$ 的极限是$|f|^2$, 故不妨设f_n, f 是实的非负的. 因f 连续且f_n^n 在任何有限区间上一致收敛, 故存在$a > 0$, f 在$[-a, a]$ 上严格正, 那么当n 充分大时, 在闭区间$[-a, a]$ 上, f_n^n 正且与零有一个正距离, 因此$-n \log f_n$, 同时$n(1 - f_n)$, 在$[-a, a]$ 上一致有界. 由引理6.2.1 推出$n(1 - f_n)$ 在$[-2a, 2a]$ 上一致有界, 从而由上述Taylor 展开推出$-n \log f_n$ 在$[-2a, 2a]$ 上一致有界. 因此f 在$[-2a, 2a]$ 上严格正, 重复这个过程证明了f 在$\mathbf{R}$ 上严格正.

现在在$\mathbf{R}$ 的任何有限区间I 上f_n^n 一致收敛于连续的f. 因f 严格正, 故$n \log f_n$ 在I 上一致收敛于$\log f$. 这蕴含着f_n 点点收敛于1 且在任何有限区间上一致. 再由上面的Taylor 展开推出

$$-n \log f_n(\xi) = n(1 - f_n(\xi))(1 + \Delta_n(\xi)),$$

其中$\Delta_n(\xi) = \frac{1}{2}(1 - f_n(\xi)) + \frac{1}{3}(1 - f_n(\xi))^2 + \cdots \longrightarrow 0$, 因此$n(1 - f_n(\xi)) \to -\log f$. □

这个定理有许多有用的推论. 首先如果f 是无穷可分特征函数, 则$f = f_n^n$, 因此$f = \lim_n e^{n(f_n - 1)}$, 即$f$ 是复合Poisson 特征函数的极限. 另外如果f 是形如f_n^n 的连续极限, 则对任何$t \geqslant 0$, 复合Poisson 特征函数$\exp(tn(f_n - 1))$ 有连续极限f^t, 即f^t 是特征函数, 推出f 是无穷可分的且如果f^t 对应的分布为π_t, 那么$\{\pi_t\}$ 是卷积半群.

推论6.2.1 下面的结论成立.

(1) 特征函数f 无穷可分当且仅当存在特征函数列$\{f_n\}$ 使得f_n^n 收敛于f.

(2) 特征函数是无穷可分的当且仅当它是复合Poisson 特征函数列的极限.

(3) 无穷可分特征函数列的连续极限是无穷可分的.

(4) 对任何无穷可分分布μ, 存在唯一的卷积半群$\{\pi_t\}$ 使得$\pi_1 = \mu$.

于是研究Lévy 指数的表示与研究无穷可分特征函数的表示是等价的. 下面是著名的Lévy-Khinchin 公式.

定理6.2.2 $\mathbf{R}^d$ 上复值函数ϕ 是一个卷积半群的Lévy 指数当且仅当ϕ 可表示为

$$\phi(x) = i(a, x) + \frac{1}{2}(Sx, x) + \int_{\mathbf{R}^d\setminus\{0\}} \left(1 - e^{i(x,y)} + \frac{i(x, y)}{1 + |y|^2}\right) J(dy), \tag{6.2.3}$$

其中$a \in \mathbf{R}^d$, S 是$\mathbf{R}^d$ 上对称非负定线性算子, J 是$\mathbf{R}^d\setminus\{0\}$ 上测度且满足

$$\int \frac{|x|^2}{1 + |x|^2} J(dx) < \infty,$$

称为卷积半群的Lévy 测度. 另外S, J 由ϕ 唯一决定.

证明. 为叙述简单, 我们不妨设$d = 1$. 先设ϕ 有上述表示, 则ϕ 是连续的且对任何$n > 1$, J 在$\{x : |x| > \frac{1}{n}\}$ 上的限制是有限测度, 这样ϕ 可写为

$$\phi(x) = \lim_n \left[\int_{|y|>\frac{1}{n}} (1 - e^{ixy}) J(dy) + ix \cdot \int_{|y|>\frac{1}{n}} \frac{y}{1 + y^2} J(dy)\right] + iax + \frac{1}{2}Sx^2.$$

而右边方括号内是一个复合Poisson 型Lévy 指数和一个一致平移Lévy 指数的和, 故仍然是Lévy 指数. 由上面的推论(3), ϕ 也是Lévy 指数.

下面证明ϕ 唯一决定S, J. 定义

$$\nu(dy) := \frac{y^2}{1 + y^2} J(dy)$$

且$\nu(\{0\}) := S$, 那么ν 是一个有限测度且

$$\phi(x) = iax + \int_{\mathbf{R}} \left(1 - e^{ixy} + \frac{ixy}{1 + y^2}\right) \frac{1 + y^2}{y^2} \nu(dy),$$

其中上面的积分函数在$y=0$ 处是可去连续的, 定义此处为其极限值$\frac{x^2}{2}$ 使函数连续. 直观地如果ϕ 二阶可微, 则二阶导数是$(1+x^2)\nu(dx)$ 的Fourier 变换, 唯一地决定ν. 一般地我们定义ϕ 的二阶差分

$$\begin{aligned}\theta(x) &:= \int_0^1 [\phi(x) - \frac{1}{2}(\phi(x+h) + \phi(x-h))]dh \\ &= -\int_0^1 \int_{\mathbf{R}} e^{ixy}(1-\cos hy)\frac{1+y^2}{y^2}\nu(dy)dh \\ &= -\int_{\mathbf{R}} e^{ixy}\frac{1+y^2}{y^2}\nu(dy)\int_0^1 (1-\cos hy)dh \\ &= -\int_{\mathbf{R}} e^{ixy}\frac{1+y^2}{y^2}(1-\frac{\sin y}{y})\nu(dy),\end{aligned}$$

令$k(y) := \frac{1+y^2}{y^2}(1-\frac{\sin y}{y})$, 则存在$c>0$ 使得$c \leqslant k(y) \leqslant \frac{1}{c}$. 因此$k \cdot \nu$ 仍然是有限测度且其Fourier 变换是θ, 推出它由θ 唯一决定, 故ν 由θ 从而由ϕ 唯一决定.

现在设ϕ 是Lévy 指数, 则ϕ 是连续的且由定理6.2.1, 存在特征函数列$\{f_n\}$, 对应分布列$\{\mu_n\}$, 使得$n(1-f_n) \to \phi$. 则

$$\begin{aligned}\phi(x) &= \lim_n n \int_{\mathbf{R}} (1-e^{ixy})\mu_n(dy) \\ &= \lim_n n \int_{\mathbf{R}} \big(1-e^{ixy} + \frac{ixy}{1+y^2}\big)\mu_n(dy) - ix \cdot n \int_{\mathbf{R}} \frac{y}{1+y^2}\mu_n(dy).\end{aligned}$$

令

$$a_n := -n\int_{\mathbf{R}} \frac{y}{1+y^2}\mu_n(dy),\ \ \nu_n(dy) := \frac{ny^2}{1+y^2}\mu_n(dy),\ \ \phi_n := n(1-f_n).$$

我们再用二阶差分,

$$\theta_n(x) := \int_0^1 [\phi_n(x) - \frac{1}{2}(\phi_n(x+h)+\phi_n(x-h))]dh = -\int_{\mathbf{R}} e^{ixy}k(y)\nu_n(dy).$$

因ϕ_n 有连续极限, 故θ_n 也有连续极限, 记为θ. 由Fourier 变换的连续性, $k \cdot \nu_n$ 弱收敛, 推出ν_n 弱收敛于一个有限测度记为ν, 那么

$$\int_{\mathbf{R}} \big(1-e^{ixy}+\frac{ixy}{1+y^2}\big)\frac{1+y^2}{y^2}\nu_n(dy) \longrightarrow \int_{\mathbf{R}} \big(1-e^{ixy}+\frac{ixy}{1+y^2}\big)\frac{1+y^2}{y^2}\nu(dy),$$

同时推出a_n 收敛于某实数a. □

定理中的Lévy 测度所满足的条件等价于对某个(因此对所有) $\epsilon > 0$, 有

$$\int_{|x|<1} |x|^2 J(dx) < \infty,\ \ J(\{x : |x| > \epsilon\}) < \infty.$$

因此Lévy 指数的表达式的形式通常也写成为

$$\phi(x) = i(a,x) + \frac{1}{2}(Sx,x) + \int_{\mathbf{R}^d\setminus\{0\}} \left(1 - e^{i(x,y)} + i(x,y)1_{\{|y|<1\}}\right) J(dy), \tag{6.2.4}$$

因为不难验证

$$\int \left| i(x,y) \left[\frac{1}{1+|y|^2} - 1_{\{|y|<1\}} \right] \right| J(dy) < \infty.$$

在这两个表达式中a 的取值可能是不同的. 当然还可以有许多其他的等价表达式. 定理也说明Lévy 过程的所有信息都嵌入在其Lévy 指数上.

§6.3 Poisson 过程

简单地说, Poisson 半群对应的平稳独立增量过程是Poisson 过程.

定义6.3.1 随机过程$N = (N_t)$ 称为是参数为λ 的Poisson 过程, 如果

(1) $N_0 = 0$ a.s.;

(2) N 是平稳独立增量过程且N_t 是参数为λt 的Poisson 分布;

(3) N 是右连续的.

但是一般的平稳独立增量过程没有右连续性, 但是我们可以证明定义中这样的Poisson 过程是存在的. 设$\{T_n : n \geqslant 1\}$ 是非负独立随机序列, 其分布函数F 有密度函数f. 记S_n 是其部分和过程, 即

$$S_0 = 0,\ S_n = \sum_{k=1}^{n} T_k.$$

定义连续时间整数值随机过程$N = (N_t : t \geqslant 0)$ 如下: 对$\omega \in \Omega$,

$$N_t(\omega) := \sum_{n=0}^{\infty} n 1_{\{t \in [S_n(\omega), S_{n+1}(\omega))\}},\ t \geqslant 0. \tag{6.3.1}$$

那么对任何$t \geqslant 0$, N_t 是整数值随机变量, 对任何$\omega \in \Omega$, $t \mapsto N_t(\omega)$ 是$[0,\infty)$ 上整数值的递增函数. 随机过程$N = (N_t)$ 被称为更新过程. 把S_n 看成为第n 个更新, N 可以看成为对于更新的记录, 它在$[S_n, S_{n+1})$ 上的值是n. 如果把T_n 理解为某个服务柜台完成一个服务所需要的时间, S_n 就是完成第n 个服务的时间, N_t 就是t 时刻该柜台已经完成服务的数目. 由于有这样的背景, 更新过程通常用于排队模型. 什么条件下更新过程是独立增量过程呢?

引理6.3.1 如果更新过程$N = (N_t)$ 是平稳独立增量过程, 则T_n 是指数分布的. 反过来, 如果T_n 是参数为$\lambda > 0$ 的指数分布, 则N 是平稳独立增量过程, 且N_t 是参数为λ 的Poisson 分布.

Proof. 设更新过程N 是平稳独立增量的, 那么任取$t, s > 0$,

$$\begin{aligned}\mathbb{P}(N_{t+s} = 0) &= \mathbb{P}(N_{t+s} - N_s = 0, N_s = 0) \\ &= \mathbb{P}(N_{t+s} - N_s = 0)\mathbb{P}(N_s = 0) \\ &= \mathbb{P}(N_t = 0)\mathbb{P}(N_s = 0).\end{aligned}$$

因为$N_t = 0$ 当且仅当$T_1 = S_1 > t$, 因此有

$$\mathbb{P}(T_1 > t + s) = \mathbb{P}(T_1 > t)\mathbb{P}(T_1 > s),$$

即得T_1 是指数分布的.

反过来, T_1 是指数分布, 则S_k 是Gamma 分布的. 对任何$t > 0$, 非负整数k, 有

$$\begin{aligned}\mathbb{P}(N_t = k) = \mathbb{P}(S_k \leqslant t < S_{k+1}) &= \mathbb{P}(S_k \leqslant t < S_k + T_{k+1}) \\ &= \int_0^t \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} dx \int_{t-x}^{\infty} \lambda e^{-\lambda y} dy \\ &= \frac{\lambda^k}{(k-1)!} \int_0^t x^{k-1} e^{-\lambda x} e^{-\lambda(t-x)} dx \\ &= \frac{(\lambda t)^k}{k!} e^{-\lambda t}.\end{aligned}$$

下面我们证明$\mathbb{P}(N_{t+s} - N_s = k, N_s = n) = \mathbb{P}(N_t = k)\mathbb{P}(N_s = n)$, 一般的场合留给读者作为习题. 首先我们容易算出$(S_1, S_2, \cdots, S_n)$ 的联合分布密度为

$$p(y_1, \cdots, y_k) = \lambda^k e^{-\lambda y_k} 1_{\{0 \leqslant y_1 < \cdots < y_k\}}. \tag{6.3.2}$$

然后,

$$\begin{aligned}&\mathbb{P}(N_{t+s} - N_s = k, N_s = n) = \mathbb{P}(S_{k+n} \leqslant t + s < S_{k+n+1}, S_n < \leqslant s < S_{n+1}) \\ &= \int_{0 \leqslant y_1 < \cdots < y_{n+k+1}} \lambda^{n+k+1} e^{-\lambda y_{n+k+1}} 1_{\{y_n \leqslant s < y_{n+1}, y_{n+k} \leqslant t+s < y_{n+k+1}\}} dy_1 \cdots dy_{n+k+1} \\ &= \lambda^{n+k} e^{-\lambda(t+s)} \int_{0 < y_1 < \cdots < y_n < s, s < y_{n+1} < \cdots < y_{n+k} < t+s} dy_1 \cdots dy_{n+k}\end{aligned}$$

$$= \lambda^{n+k} e^{-\lambda(t+s)} \frac{s^n}{n!} \frac{t^k}{k!}$$
$$= \mathbb{P}(N_s = n)\mathbb{P}(N_t = k).$$

完成证明. □

因为更新过程是右连续的, 由上面的引理, Poisson 过程实际上是间隔时间是指数分布的更新过程. 比Poisson 更一般的是复合Poisson 过程, 自然地, 复合Poisson 半群对应的平稳独立增量过程是复合Poisson 过程. 复合Poisson 过程是由Poisson 过程来构造的, 它的直观思想是这样的, 一个质点在某个点开始在等待一个参数为λ 的指数分布的时间后, 按照某个固定的分布在空间中选择一个点跳过去, 然后再等待一段时间, 如此反复, 这样的移动称为复合Poisson 过程.

由Kolmogorov 定理, 存在概率空间$(\Omega, \mathcal{F}, \mathbb{P})$, 在这个概率空间上有(1) 一个参数为$\lambda$ 的指数分布的独立同分布随机序列$\{T_n : n \geqslant 1\}$; (2) 取值于$\mathbf{R}^d$ 的分布为J 的独立同分布随机序列$\{\xi_n : n \geqslant 1\}$; (3) 序列$\{T_n\}$ 与序列$\{\xi_n\}$ 独立. 设$N = (N_t : t \geqslant 0)$ 是$\{T_n\}$ 诱导的参数为λ 的Poisson 过程, 它独立于$\{\xi_n\}$. 然后定义

$$X_t = \sum_{n=1}^{N_t} \xi_n,\ t \geqslant 0. \tag{6.3.3}$$

Proposition 6.3.1 $X = (X_t : t \geqslant 0)$ 是d-维平稳独立增量过程, 且其卷积半群是复合Poisson 半群

$$\mu_t = \sum_{n=0}^{\infty} e^{\lambda t} \frac{t^n J^{*n}}{n!},$$

其中J^{*n} 是卷积的n-次幂, $J^{*0} = \delta_0$.

第七章　Brown 运动

如果数学概念有血统的话, 那么Brown 运动绝对有辉煌的血统. 它是生物学家R.Brown 首先提出来研究的一种现象: 花粉在液体表面的运动. 然后由历史上最伟大的物理学家之一A. Einstein 在研究热传导现象的时候给出其转移密度函数, 最终由天才的数学家, 控制论创始人N.Wiener 证明了其轨道的连续性, 证明了它是花粉运动的一个恰当的数学模型. 当然Brown 运动也无愧于其血统, 它绝对是概率论中最重要, 被用得最多的一个随机过程.

§7.1　Brown 运动的构造

定义7.1.1 d-维随机过程$B = (B_t : t \geqslant 0)$ 被称为d-维标准Brown 运动, 如果

(1) $B_0 = 0$ a.s.;

(2) B 是平稳独立增量过程且B_t 的卷积半群是热半群;

(3) B 是连续的, 即存在零概率集N, 使得对任何$\omega \notin N$, $t \mapsto B_t(\omega)$ 是连续的.

由引理6.1.2 存在一个满足(1) 与(2) 的d-维随机过程是容易的, 关键是要证明同时满足(3) 的随机过程的存在性, 这是Wiener 的贡献.

定理7.1.1 (N. Wiener) $\mathbb{R}^d$ 上存在有标准Brown 运动.

Proof. 我们设$d = 1$, 对高维的证明是类似的. 首先我们应用Kolmogorov 相容性定理构造一个概率空间$(\Omega, \mathcal{F}, \mathbb{P})$ 及其上的随机过程过程$X = (X_t)$ 使得它的有限维分布是由(**??**) 的右边给出. 容易验证$(X_t)_{t\geqslant 0}$ 满足Brown 运动定义中除连续性外的其它条件. 最重要的是, 有下面的矩等式

$$\mathbb{E}|B_t - B_s|^{2n} = (2n-1)!!|t-s|^n. \tag{7.1.1}$$

因此我们需要修改X_t 使得它连续. 设$D = \{\frac{j}{2^n} : j \in \mathbb{Z}^+, n \in \mathbb{N}\}$ 非负二分点全体. 显然D 是$\mathbb{R}^+$ 的可数稠子集. 定义

$$H = \bigcup_{N=1}^{\infty} \bigcap_{l=1}^{\infty} \bigcup_{n=l}^{\infty} \bigcup_{j=1}^{N2^n} \left(\left| X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}} \right| \geqslant \frac{1}{2^{n/8}} \right).$$

对固定正整数N, 令

$$A_l = \bigcup_{n=l}^{\infty} \bigcup_{j=1}^{N2^n} \left(\left| X_{\frac{j}{2^n}} - X_{\frac{j-1}{2^n}} \right| \geqslant \frac{1}{2^{n/8}} \right).$$

我们将证明$\bigcap_{l=1}^{\infty} A_l$ 的概率为零, 因此作为其可列并$\mathbb{P}(H)=0$.

应用矩等式(7.1.1) 中$n=2$ 的场合,

$$\begin{aligned}
&\mathbb{P}\left\{\bigcup_{j=1}^{N2^n}\left(\left|X_{\frac{j}{2^n}}-X_{\frac{j-1}{2^n}}\right|\geqslant\frac{1}{2^{n/8}}\right)\right\}\\
&\leqslant\sum_{j=1}^{N2^n}\mathbb{P}\left(\left|X_{\frac{j}{2^n}}-X_{\frac{j-1}{2^n}}\right|\geqslant\frac{1}{2^{n/8}}\right)\\
&=N2^n\mathbb{P}\left(\left|X_{\frac{1}{2^n}}\right|\geqslant\frac{1}{2^{n/8}}\right)\\
&\leqslant N2^n\left(2^{n/8}\right)^4\mathbb{E}\left|X_{\frac{1}{2^n}}\right|^4\\
&=\left(2^{n/8}\right)^4N2^n3\left(\frac{1}{2^n}\right)^2\\
&=\frac{3N}{2^{n/2}},
\end{aligned}$$

故

$$\begin{aligned}
\mathbb{P}(A_l)&\leqslant\sum_{n=l}^{\infty}\mathbb{P}\left\{\bigcup_{j=1}^{N2^n}\left(\left|X_{\frac{j}{2^n}}_X_{\frac{j-1}{2^n}}\right|\geqslant\frac{1}{2^{n/8}}\right)\right\}\\
&\leqslant 3N\sum_{n=l}^{\infty}\frac{1}{2^{n/2}}\\
&=\frac{3N\sqrt{2}}{\sqrt{2}-1}\frac{1}{\left(\sqrt{2}\right)^l}.
\end{aligned}$$

因此

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{l=1}^{\infty}A_l\right)&=\lim_{l\to\infty}\mathbb{P}\{A_l\}\\
&\leqslant\frac{3N\sqrt{2}}{\sqrt{2}-1}\lim_{l\to\infty}\frac{1}{\left(\sqrt{2}\right)^l}=0\ .
\end{aligned}$$

由此推出$\mathbb{P}(H)=0$, 即$\mathbb{P}(H^c)=1$. 另一方面, 由De Morgan 律

$$H^c=\bigcap_{N=1}^{\infty}\bigcup_{l=1}^{\infty}\bigcap_{n=l}^{\infty}\bigcap_{j=1}^{N2^n}\left\{\omega:\left|X_{\frac{j}{2^n}}(\omega)\_X_{\frac{j-1}{2^n}}(\omega)\right|<\frac{1}{2^{n/8}}\right\}$$

因此, 若$\omega\in H^c$, 则对任何N, 存在l 使得对任何$n>l$ 与$j=1,\cdots,N2^n$ 有

$$\left|X_{\frac{j}{2^n}}(\omega)-X_{\frac{j-1}{2^n}}(\omega)\right|<\frac{1}{2^{n/8}}\ .$$

我们因此可以作为一个分析习题证明对任何$\omega \in H^c$, $X_t(\omega)$ 在任何有界区间上一致连续, 故对任何$t \geqslant 0$, 当s 沿着D 趋于t 时, $X_s(\omega)$ 的极限存在. 而D 在$[0, \infty)$ 中稠, 因此对任何$t \in [0, \infty)$ 我们可以定义

$$B_t(\omega) = \lim_{s \in D \to t} X_s(\omega) \quad \text{if } \omega \in H^c$$

若$\omega \in H$, 定义$B_t(\omega) = 0$. 由定义, $(B_t)_{t \geqslant 0}$ 连续, 且对任何$t \geqslant 0$, X_t 几乎处处收敛于B_t. 剩下的事情就是证明$(B_t)_{t \geqslant 0}$ 是X 的修正, 因此它是Brown 运动, 留作习题. □

我们需要一个分析结果以证明$(X_t(\omega) : t \in D \cap [0, n])$ 对任何n 一致连续.

练习7.1.1 设$\alpha > 0$ 且f 是D 上的函数满足对任何N, 存在l 使得对任何$n > l$ 与$j = 1, \cdots, N2^n$ 有

$$\left| f(\frac{j}{2^n}) - f(\frac{j-1}{2^n}) \right| \leqslant \left(\frac{1}{2^n} \right)^\alpha .$$

那么对任何$n > 0$, 存在常数C_n 使得对任何$s, t \in D \cap [0, n]$,

$$|f(s) - f(t)| \leqslant C_n |s - t|^\alpha ,$$

即f 是α-阶局部Hölder 连续的.

§7.2 反射原理

Brown 运动从一个停时开始还是Brown 运动, 即其Markov 性在停时处仍然成立. 因此Brown 运动具有强Markov 性, 这是一个非常重要的性质, 首先是Paul Lévy 以反射原理的形式加以应用, 反射原理是远比Markov 性质更为古老的一个概念. 我们将应用这个原理来计算Brown 运动的极大游程的分布.

在许多应用实例中, 特别在统计中, 我们需要估计随机过程的极大游程的分布. 对于Brown 运动$B = (B_t)_{t \geqslant 0}$ 来说, 极大游程$\sup_{s \in [0,t]} B_s$ 的分布可以由反射原理的方法来导出.

设$B = (B_t)_{t \geqslant 0}$ 在概率空间$(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ 上一维标准Brown 运动. 设$b \geqslant a > 0$, 并设

$$T_b = \inf\{t > 0 : B_t = b\} .$$

那么T_b 是停时, 且

$$\{ \sup_{s \in [0,t]} B_s > b \} = \{T_b < t\}.$$

Brown 运动从到达b 开始后实际上就是标准Brown 运动加上b, 因此有

$$\begin{aligned}\mathbb{P}\{T_b < t, B_t < a\} &= \mathbb{P}\{T_b < t, B_t < 2b - a\} \\ &= \mathbb{P}\{B_t < 2b - a\}\ ,\end{aligned} \tag{7.2.1}$$

其中第一个等式就是所谓的反射原理, 和随机游动的反射原理一样, 在Brown 运动上次到达b 的时间点T_b 看, (从位置b 开始: $B_{T_b} = b$) 跑得就似Brown 运动, 它向下跑$a-b$ 到a 之下和向上跑$b-a$ 到$2b-a$ 之上的概率应该是一样的, 或者说它在T_b 后的轨道关于$y = b$ 是对称的, 就如同标准Brown 运动关于原点对称. 第二个等式是因为$2b - a > b$, 故$\{B_t > 2b - a\} \subset \{T_b < t\}$. 上面这个方程称为反射原理.

当$a = b$ 时, 反射原理推出

$$\begin{aligned}\mathbb{P}(T_a < t) &= \mathbb{P}(T_a < t, B_t < a) + \mathbb{P}(T_a < t, B_t > a) \\ &= 2\mathbb{P}(B_t > a) = 2\mathbb{P}(B_1 > a/\sqrt{t}),\end{aligned}$$

因此

$$\mathbb{P}(T_a < \infty) = 1,$$

即Brown 运动以概率1 会到达a 点.

上面的方法对$a = 0$ 偏偏无效. 自然想到取极限, 当$a \downarrow 0$ 时, 是否有$T_a \downarrow T_0$? 如果对的话, 那么对任何$t > 0$,

$$\mathbb{P}(T_0 < t) = 2\mathbb{P}(B_1 > 0) = 1.$$

因此$T_0 = 0$. 但上面的极限无法证明其成立, 怎么办呢? 取$t > 0$, 用T^t 表示t 时间Brown 运动首次回到0 的时间, 那么由连续函数的性质

$$\lim_{t \to 0} T^t = T_0.$$

利用Brown 运动平移不变的性质, 当$B_t = x$ 时, 之后它首达0 相当于从0 出发首达$-x$, 因此我们可以先算条件期望

$$\begin{aligned}\mathbb{P}(T^t < u | B_t = x) &= \mathbb{P}(T_{-x} < u - t) \\ &= 2\mathbb{P}(B_1 > |x|/\sqrt{u-t}),\end{aligned}$$

再取期望得

$$\begin{aligned}\mathbb{P}(T^t < u) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{\frac{-|x|^2}{2t}} 2\mathbb{P}(B_1 > \frac{|x|}{\sqrt{u-t}})dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-|x|^2}{2}} 2\mathbb{P}(B_1 > \frac{\sqrt{t}|x|}{\sqrt{u-t}})dx.\end{aligned}$$

令$t \downarrow 0$, 推出

$$\mathbb{P}(T_0 < u) = 2\mathbb{P}(B_1 > 0) = 1$$

对任何$u > 0$ 成立, 因此$T_0 = 0$ a.s. 这说明B 的轨道出发时无穷多次与x-轴相交.

考虑轨道的零点集

$$Z(\omega) := \{t : B_t(\omega) = 0\},$$

它是一个随机集合, 对几乎所有的轨道, Z 是闭集, 且

$$\mathbb{E}|Z| = \mathbb{E}\int 1_{\{B_t=0\}}dt = \int \mathbb{P}(B_t = 0)dt = 0,$$

因此几乎每条轨道的零点集Z 的Lebesgue 测度是零. 实际上我们还可以证明几乎所有轨道的零点集都没有孤立点, 或者说Z 是测度零的完美集.

由反射原理(7.2.1) 可得

$$\mathbb{P}\left\{\sup_{s\in[0,t]} B_s \geqslant b, B_t \leqslant a\right\} = \frac{1}{\sqrt{2\pi t}}\int_{2b-a}^{+\infty} e^{-\frac{x^2}{2t}}dt\ ,$$

它给出了Brown 运动t 时刻的位移B_t 与其极大游程$\sup_{s\in[0,t]} B_s$ 的联合分布. 要知道密度, 我们对a 与b 求导.

定理7.2.1 设$B = (B_t)_{t\geqslant 0}$ 是一维标准Brown 运动, $t > 0$. 那么随机变量$(M_t = \sup_{s\in[0,t]} B_s, B_t)$ 的联合密度函数为

$$\mathbb{P}\{M_t \in db, B_t \in da\} = \frac{2(2b-a)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(2b-a)^2}{2t}\right\} dadb,$$

其中(b, a) 位于$\mathbb{R}^2$ 的区域$\{(b, a) : a \leqslant b, b > 0\}$ 中.

练习**7.2.1** 证明

$$\mathbb{E}\left(e^{-sT_b}\right) = e^{-\sqrt{2s}b}.$$

§7.3 二次变差

如同我们看到的那样, 一维Brown 运动B_t 与$M_t \equiv B_t^2 - t$ 都是鞅, 那么

$$B_t^2 = M_t + A_t$$

其中$A_t = t$. 因此, 连续下鞅B_t^2 是一个鞅与一个增适应过程的和. 我们将会看到这个分解是Itô 随机积分理论得以建立的关键.

引理7.3.1 设

$$D = \{0 = t_0 < t_1 < \cdots < t_n = t\}$$

是区间$[0, t]$ 的有限划分, 且设

$$V_D = \sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2$$

称为B 在分划D 上的二次变差, 它是非负随机变量. 那么

$$\mathbb{E}V_D = t$$

且V_D 的方差为

$$\mathbb{E}\left\{(V_D - \mathbb{E}V_D)^2\right\} = 2\sum_{l=1}^{n}(t_l - t_{l-1})^2 \ .$$

Proof. 事实上

$$\begin{aligned}\mathbb{E}V_D &= \sum_{l=1}^{n} \mathbb{E}|B_{t_l} - B_{t_{l-1}}|^2 \\ &= \sum_{l=1}^{n}(t_l - t_{l-1}) \\ &= t \ .\end{aligned}$$

为证明第二个公式, 我们如下计算

$$\begin{aligned}&\mathbb{E}\left\{(V_D - \mathbb{E}V_D)^2\right\} \\ &\quad = \mathbb{E}\left\{\left(\sum_{l=1}^{n} |B_{t_l} - B_{t_{l-1}}|^2 - t\right)^2\right\}\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\left(\sum_{l=1}^{n}\left(|B_{t_l}-B_{t_{l-1}}|^2-(t_l-t_{l-1})\right)\right)^2\right\} \\
&= \sum_{k,l=1}^{n}\mathbb{E}\left\{\left(|B_{t_k}-B_{t_{k-1}}|^2-(t_k-t_{k-1})\right)\left(|B_{t_l}-B_{t_{l-1}}|^2-(t_l-t_{l-1})\right)\right\} \\
&= \sum_{l=1}^{n}\mathbb{E}\left\{\left(|B_{t_l}-B_{t_{l-1}}|^2-(t_l-t_{l-1})\right)^2\right\} \\
&+ \sum_{k\neq l}\mathbb{E}\left\{\left(|B_{t_k}-B_{t_{k-1}}|^2-(t_k-t_{k-1})\right)\left(|B_{t_l}-B_{t_{l-1}}|^2-(t_l-t_{l-1})\right)\right\} \ .
\end{aligned}
$$

因为不同区间的增量是独立的, 以上和式中的每个乘积的期望会等于期望的乘积, 故等于零. 因此我们有

$$
\begin{aligned}
&\mathbb{E}\left\{(V_D-\mathbb{E}V_D)^2\right\} \\
&\quad= \sum_{l=1}^{n}\mathbb{E}\left\{\left(|B_{t_l}-B_{t_{l-1}}|^2-(t_l-t_{l-1})\right)^2\right\} \\
&\quad= \sum_{l=1}^{n}\mathbb{E}\left\{|B_{t_l}-B_{t_{l-1}}|^4-2(t_l-t_{l-1})|B_{t_l}-B_{t_{l-1}}|^2+(t_l-t_{l-1})^2\right\} \\
&\quad= \sum_{l=1}^{n}\left\{\mathbb{E}|B_{t_l}-B_{t_{l-1}}|^4-2(t_l-t_{l-1})\mathbb{E}|B_{t_l}-B_{t_{l-1}}|^2+(t_l-t_{l-1})^2\right\} \\
&\quad= 2\sum_{l=1}^{n}(t_l-t_{l-1})^2
\end{aligned}
$$

这里我们应用了积分

$$
\mathbb{E}|B_{t_l}-B_{t_{l-1}}|^4 = 3(t_l-t_{l-1})^2 \ .
$$

□ □

我们现在可以叙述下面的定理.

定理7.3.1 设$B=(B_t)_{t\geqslant 0}$ 是一维标准Brown 运动. 那么对任何$t>0$,

$$
\lim_{m(D)\to 0}\sum_{l}|B_{t_l}-B_{t_{l-1}}|^2 = t \qquad \text{in } L^2(\Omega,\mathbb{P})
$$

其中D 是区间$[0,t]$ 上的有限分划, 且

$$
m(D) = \max_{l}|t_l-t_{l-1}| \ .
$$

因此

$$\lim_{m(D)\to 0}\sum_l |B_{t_l}-B_{t_{l-1}}|^2 = t \qquad \text{in probability.}$$

Proof. 基于前面的引理, 我们有

$$\begin{aligned}
\mathbb{E}\left|\sum_l |B_{t_l}-B_{t_{l-1}}|^2 - t\right|^2 &= \mathbb{E}\left|V_D-\mathbb{E}\left(V_D\right)\right|^2 \\
&= 2\sum_{l=1}^n (t_l-t_{l-1})^2 \\
&\leqslant 2m(D)\sum_{l=1}^n (t_l-t_{l-1}) \\
&= 2tm(D),
\end{aligned}$$

因此

$$\lim_{m(D)\to 0}\mathbb{E}\left|\sum_l |B_{t_l}-B_{t_{l-1}}|^2 - t\right|^2 = 0\ .$$

□ □

上面的收敛是依概率收敛. 但若定理中的分划取得更好的话, 上面的收敛可以变成是几乎处处的.

Proposition 7.3.2 设$(B_t)_{t\geqslant 0}$ 是一维标准Brown 运动. 那么对任何$t>0$, 当n 趋于无穷时, 有

$$\sum_{j=1}^{2^n}\left|B_{\frac{j}{2^n}t}-B_{\frac{j-1}{2^n}t}\right|^2 \to t \quad \text{a.s.} \tag{7.3.1}$$

Proof. 设D_n 是$[0,t]$ 上的二分点分划

$$D_n = \{0=\frac{0}{2^n}t<\frac{1}{2^n}t<\cdots<\frac{2^n}{2^n}t=t\}\ .$$

且用V_n 表示V_{D_n}. 那么, 按引理Lemma 7.3.1, $\mathbb{E}V_n = t$ and

$$\begin{aligned}
\mathbb{E}\left|V_n-\mathbb{E}V_n\right|^2 &= 2\sum_{l=1}^{2^n}\left(\frac{l}{2^n}t-\frac{l-1}{2^n}t\right)^2 \\
&= 2^{n+1}\left(\frac{1}{2^n}t\right)^2
\end{aligned}$$

$$= \frac{1}{2^{n-1}} t^2 \ .$$

因此由Markov 不等式,

$$\mathbb{P}\left\{ |V_n - \mathbb{E}V_n| \geqslant \frac{1}{n} \right\} \leqslant n^2 \mathbb{E} |V_n - \mathbb{E}V_n|^2$$
$$= \frac{n^2}{2^{n-1}} t^2$$

故而

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{ |V_n - \mathbb{E}V_n| \geqslant \frac{1}{n} \right\} = t^2 \sum_{n=1}^{\infty} \frac{n^2}{2^{n-1}} < +\infty \ .$$

这样由Borel-Cantelli 引理得$V_n \to t$ 几乎处处收敛. □ □

事实上, 几乎处处收敛得结论对于单调分划列都成立. 更确切地说, 对任何n 设

$$D_n = \{0 = t_{0,n} < t_{1,n} < \cdots < t_{k_n,n} = t\}$$

是$[0,t]$ 的有限分划. 如果$D_{n+1} \supset D_n$ 且

$$\lim_{n\to\infty} m(D_n) = \lim_{n\to\infty} \max |t_{n_i,n} - t_{n_{i-1},n}| = 0,$$

那么当n 趋于无穷时

$$\sum_{i=1}^{n} \left| B_{t_{k_i,n}} - B_{t_{k_{i-1},n}} \right|^2 \to t \quad \text{a.s.} \tag{7.3.2}$$

此结论由下面的命题与非负鞅列收敛定理直接推出, 证明参见习题.

Proposition 7.3.3 用M_n 表示(7.3.2) 的左边, 则时间反转的序列

$$\{\cdots\cdots, M_n, \cdots, M_2, M_1\}$$

是一个非负鞅.

§7.4 鞅性质

设$B = (B_t^i)_{t\geqslant 0}$ $(i = 1, \cdots, d)$ 是$\mathbb{R}^d$ 上标准Brown 运动, 其自然流为$(\mathcal{F}_t)_{t\geqslant 0}$.

Proposition 7.4.1 1) 对任何$p>0$, 每个B_t 是p-阶可积的, 且对$t>s$

$$\mathbb{E}(|B_t-B_s|^p)=c_{p,d}|t-s|^{p/2}\ . \tag{7.4.1}$$

2) $(B_t)_{t\geqslant 0}$ 连续平方可积鞅.

3) 对任何i,j, $M_t=B_t^iB_t^j-\delta_{ij}t$ 也是连续鞅.

Proof. 第一部分以前证明过. 当$t>s$ 时, 因为B_t-B_s 独立于$\mathcal{F}_s$. 我们因此有

$$\mathbb{E}(B_t-B_s|\mathcal{F}_s)=\mathbb{E}(B_t-B_s)=0.$$

那么

$$\mathbb{E}(B_t|\mathcal{F}_s)=\mathbb{E}(B_s|\mathcal{F}_s)=B_s$$

也就是说, $(B_t)_{t\geqslant 0}$ 是连续鞅.

显然我们只需对一维Brown 运动证明3). 这种情形下,

$$\begin{aligned}\mathbb{E}(B_t^2-B_s^2|\mathcal{F}_s)&=\mathbb{E}((B_t-B_s)^2\,|\mathcal{F}_s)\\&\quad+\mathbb{E}(2B_s\,(B_t-B_s)\,|\mathcal{F}_s)\\&=\mathbb{E}((B_t-B_s)^2)+2B_s\mathbb{E}((B_t-B_s)\,|\mathcal{F}_s)\\&=\mathbb{E}\,(B_t-B_s)^2\\&=t-s\end{aligned}$$

故而

$$\begin{aligned}\mathbb{E}(B_t^2-t|\mathcal{F}_s^0)&=\mathbb{E}(B_s^2-s|\mathcal{F}_s)\\&=B_s^2-s\end{aligned}$$

这已经证明了B_t^2-t 是一个鞅. □ □

定理7.4.2 设$B=(B_t)_{t\geqslant 0}$ 是$\mathbb{R}$ 上连续随机过程, 使得$B_0=0$. 那么$(B_t)_{t\geqslant 0}$ 是一维标准Brown 运动当且仅当对任何$\xi\in\mathbb{R}$ 与$t>s$

$$\mathbb{E}\left\{\exp\left(\sqrt{-1}\langle\xi,B_t-B_s\rangle\right)|\mathcal{F}_s\right\}=\exp\left(-\frac{(t-s)|\xi|^2}{2}\right)\ . \tag{7.4.2}$$

Proof. 我们观察到(7.4.2) 蕴含着$B_t - B_s$ 独立于$\mathcal{F}_s$ 且是方差为$t-s$ 的正态分布. 反过来, 如果$(B_t)_{t\geqslant 0}$ 是一维标准Brown 运动, 那么$B_t - B_s$ 独立于$\mathcal{F}_s$, 且$B_t - B_s$ 是期望零方差$t-s$ 的正态分布, 故

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left(\sqrt{-1}\langle \xi, B_t - B_s\rangle\right)|\mathcal{F}_s\right\} \\
&\quad = \mathbb{E}\left\{\exp\left(\sqrt{-1}\langle \xi, B_t - B_s\rangle\right)\right\} \\
&\quad = \frac{1}{\sqrt{2\pi(t-s)}}\int_{\mathbb{R}} e^{\sqrt{-1}\langle \xi, x\rangle - \frac{|x|^2}{2(t-s)}}dx \\
&\quad = \exp\left(-\frac{(t-s)|\xi|^2}{2}\right) .
\end{aligned}$$

□ □

推论7.4.1 设(B_t) 是$\mathbb{R}^d$ 上标准Brown 运动. 如果$\xi \in \mathbb{R}^d$, 那么

$$M_t \equiv \exp\left(\sqrt{-1}\langle \xi, B_t\rangle + \frac{|\xi|^2}{2}t\right)$$

是鞅.

注释**7.4.1** 注意到(7.4.2) 的两边都是ξ 的解析函数, 故等式对任何复值ξ 也成立. 特别地, 用$-\sqrt{-1}\xi$ 代替ξ, 我们得

$$\mathbb{E}\{\exp\left(\langle \xi, B_t - B_s\rangle\right)|\mathcal{F}_s\} = \exp\left(\frac{(t-s)|\xi|^2}{2}\right).$$

因此对任何向量ξ

$$\exp\left(\langle \xi, B_t\rangle - \frac{|\xi|^2}{2}t\right)$$

是一个连续鞅. 这个结论也可以推广到$\mathbf{R}^d$ 上的向量场$\xi = (\xi(t))$, 由此得到的等式被称为Cameron-Martin 公式.

§7.5 停时

连续时间也一样有停时的概念, 停时可以相对于一个流定义, 设有概率空间$(\Omega, \mathcal{F}, \mathbb{P})$, 流的定义非常简单, 它就是$\mathcal{F}$ 的一个递增的子σ-代数族$(\mathcal{F}_t : t > 0)$. 定义右极限流

$$\mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_u,\ t \geqslant 0,$$

那么当对任何$t \geqslant 0$ 有$\mathcal{F}_{t+} = \mathcal{F}_t$ 时, 称流$(\mathcal{F}_t)$ 是右连续的. 另外如果$\mathcal{F}_0$ 中含有所有$\mathcal{F}$ 的零概率集及其子集, 那么我们说流$(\mathcal{F}_t)$ 是被强化了的. 一个强化了的右连续流被称为是满足通常条件的. 从定义我们可以看出, 流和概率其实是没有关系的, 被强化的流才和概率发生关系. 设$B = (B_t)$ 是标准Brown 运动, 那么自然的流是它生成的流

$$\mathcal{F}_t^0 := \sigma(B_s : s \leqslant t),$$

尽管它是自然流, 但它与Brown 运动没有多少关系, 因为B 只有在概率$\mathbb{P}$ 之下才成为Brown 运动. 自然流结果强化后记为$(\mathcal{F}_t)$, 称为Brown 流. 可以证明Brown 流一定是右连续的.

练习**7.5.1** 对任何$t \geqslant 0$, $\{\tau < t\} \in \mathcal{F}_t$ 当且仅当对任何$t \geqslant 0$, $\{\tau \leqslant t\} \in \mathcal{F}_{t+}$.

一个随机时间τ 称为是$(\mathcal{F}_t)$-停时, 如果对任何$t \geqslant 0$, 有

$$\{\tau \leqslant t\} \in \mathcal{F}_t.$$

如果说τ 表示某个事件发生的时间, 那么停时是说可以用t 时刻前的消息来判断这个事件是否在t 时刻前发生. 直观地, 首次到达某个集合的时间就应该是停时了, 因为t 时刻前的轨道足以判断它是否碰到过这个集合, 而最后一次到达某集合的时间应该不是停时, 因为t 时刻前的轨道无法判断是否是最后. 但是要严格的证明不是那么简单.

定理7.5.1 如果A 是开集或者闭集, 那么A 的首中时

$$\tau_A(\omega) := \inf\{t > 0 : \ B_t(\omega) \in A\} \tag{7.5.1}$$

是Brown 流的停时.

Proof. 如果A 是开集, 事情比较简单. 因为对于连续的轨道$B_\cdot(\omega)$, $\tau_A(\omega) < t$ 当且仅当存在一个有理数$r < t$ 使得$B_r(\omega) \in A$. 因此$\{\tau_A < t\}$ 与

$$\bigcup_{r \in \mathbb{Q}, r < t} \{B_r \in A\}$$

只相差一个零概率集, 而后者是一个$\mathcal{F}_t$ 中集合的可列并, 故它在$\mathcal{F}_t$ 中, 而$\mathcal{F}_t$ 中含有所有零概率集及其子集, 所以$\{\tau_A < t\} \in \mathcal{F}_t$, 由流的右连续性推出$\tau_A$ 是停时. 其实如果没有右连续性, 读者可以想象, 命中一个开集的时间的下确界等于t 并不意味着你能在t 时刻或之前能够观察到轨道命中开集本身.

闭集的情况要复杂一些, 对于一条连续轨道来说, 因为A 是闭集, 所以B_{τ_A} 一定在A 中. 那么$\tau_A \leqslant t$ 等价于说$\{B_s : 0 < s \leqslant t\}$ 与A 相交. 我们需要把时间放到可数集上. 它是不是等价于$\{B_s : 0 < s \leqslant t, s \in \mathbb{Q}\}$ 与A 的距离为零. 这个一般不对, 后者推不出前者. 问题出在时间0 这个地方, 换种说法是可以的, 对于任何$\delta > 0$, $\{B_s : \delta \leqslant s \leqslant t\}$ 与A 相交等价于$\{B_s : \delta \leqslant s \leqslant t, s \in \mathbb{Q}\}$ 与A 的距离等于零. 由于$\mathcal{F}_t$ 包含零概率集, 故对任何$t > \delta > 0$ 有

$$\{B_s : \delta \leqslant s \leqslant t\} \in \mathcal{F}_t.$$

因而在忽略零概率集的前提下

$$\{\tau_A \leqslant t\} = \bigcup_n \{\{B_s : 1/n \leqslant s \leqslant t\} \cap A \neq \emptyset\} \in \mathcal{F}_t,$$

即τ_A 是停时. □

从这个定理的证明可以看出, 即使是开集或者闭集的首中时, 它们也必须是在流满足通常条件时才能是停时, 所以流满足通常条件是非常重要的. 下面的定理是离散时间相应定理的推广, 它的证明也是要应用连续时间鞅的离散化方法.

定理7.5.2 (Doob) 如果$M = (M_t)$ 是右连续鞅, τ 是有界停时, 那么M_τ 可积且

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

这个定理在处理停时相关问题时非常有用.

例7.5.1 设$a > 0$, τ_a 是a 的首中时

$$\tau_a = \inf\{t > 0 : \ B_t = a\}.$$

那么我们已经看到

$$\mathbb{P}(\tau_a < \infty) = 1.$$

而且显然有$B_{\tau_a} = a$. 当然τ_a 不能是有界停时, 但是对任何n, $\tau_a \wedge n$ 是有界停时. 因为$(\exp(xB_t - \frac{1}{2}x^2 t) : t > 0)$ 是鞅, 应用Doob 定理得

$$\mathbb{E}[\exp(xB_{n\wedge\tau_a} - \frac{1}{2}x^2(\tau_a \wedge n))] = 1,$$

当$x > 0$ 时, $xB_{n\wedge\tau_n} \leqslant xa$, 让$n$ 趋于无穷, 应用控制收敛定理得

$$1 = \mathbb{E}[\exp(xB_{\tau_a} - \frac{1}{2}x^2\tau_a)] = e^{xa}\mathbb{E}[e^{-\frac{1}{2}x^2\tau_a}],$$

令$\frac{1}{2}x^2 = s$, 推出τ_a 的Laplace 变换为

$$\mathbb{E}[e^{-s\tau_a}] = e^{-\sqrt{2s}a}.$$

两边对s 求导且让s 趋于零得$\mathbb{E}[\tau_a] = +\infty$.

现在取$a < 0 < b$,

$$\tau := \tau_{\{a,b\}} = \tau_a \wedge \tau_b.$$

当然τ 也是有限值的但不是有界的. 我们先算Brown 运动先到a 的概率$\mathbb{P}(\tau_a < \tau_b)$. 因为$B = (B_t)$ 是鞅, 所以

$$\mathbb{E}[B_{\tau\wedge n}] = 0,$$

而$|B_{\tau\wedge n}| \leqslant \max -a, b$, 让$n$ 趋于无穷, 得

$$0 = \mathbb{E}[B_\tau] = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_a > \tau_b),$$

因此

$$\mathbb{P}(\tau_a < \tau_b) = \frac{b}{b-a}.$$

再算τ 的期望, 那需要利用$(B_t^2 - t)$ 是鞅的事实, 得

$$\mathbb{E}[B_{\tau\wedge n}^2] = \mathbb{E}[\tau \wedge n],$$

还是应用控制收敛定理

$$\begin{aligned}
\mathbb{E}[\tau] &= \mathbb{E}[B_\tau^2] \\
&= \mathbb{E}[B_\tau^2; \tau_a < \tau_b] + \mathbb{E}[B_\tau^2; \tau_a > \tau_b] \\
&= a^2\frac{b}{b-a} + b^2\frac{-a}{b-a} \\
&= -ab.
\end{aligned}$$

最后我们来算τ 的Laplace 变换. 那要用指数鞅,

$$\mathbb{E}\left(\exp(xB_{\tau\wedge n} - \frac{1}{2}x^2(\tau \wedge n))\right) = 1.$$

再用控制收敛定理得

$$\mathbb{E}\left(\exp(xB_\tau - \frac{1}{2}x^2\tau)\right) = 1.$$

还是分成为两个事件$\{\tau_a < \tau_b\}$ 及其余集, 得

$$e^{xa}\mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a < \tau_b] + e^{xb}\mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a > \tau_b] = 1.$$

利用Brown 对称的性质, $(-B_t)$ 是和(B_t) 分布相同的, 得到类似的

$$e^{-xa}\mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a < \tau_b] + e^{-xb}\mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a > \tau_b] = 1.$$

这样

$$\begin{aligned}\mathbb{E}[e^{-\frac{1}{2}x^2\tau}] &= \mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a < \tau_b] + \mathbb{E}[e^{-\frac{1}{2}x^2\tau}; \tau_a > \tau_b] \\ &= \frac{e^{-xb} - e^{xb} + e^{xa} - e^{-xa}}{e^{-x(b-a)} - e^{x(b-a)}},\end{aligned}$$

令$s = \frac{1}{2}x^2$,

$$\mathbb{E}\left[e^{-s\tau}\right] = \frac{\sinh(b\sqrt{2s}) - \sinh(a\sqrt{2s})}{\sinh((b-a)\sqrt{2s})},$$

其中

$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

设$k \in \mathbf{R}$, $a > 0$, 定义

$$T_{k,a} := \inf\{t > 0 :\ B_t = kt + a\},$$

即时空平面上的Brown 运动首次碰到直线$y = kt + a$ 的时间, 或者漂移Brown 运动$(B_t - kt)$ 首次碰到点a 的时间. 当$k = 0$ 时, 它就是a 的首中时, 它的情况已经清楚, 所以我们只讨论k 非零的情况. 当$k < 0$ 时, 显然$T_{k,a} \leqslant \tau_a$, 所以

$$\mathbb{P}(T_{k,a} < \infty) = 1.$$

我们可以来求它的Laplace 变换. 类似地对指数鞅应用Doob 定理

$$\mathbb{E}\left[\exp(xB_{T_{k,a}\wedge n} - \frac{1}{2}x^2(T_{k,a} \wedge n))\right] = 1, \tag{7.5.2}$$

当$x > 0$ 时$xB_{T_{k,a}\wedge n}$ 上有界, 当n 趋于无穷时应用控制收敛定理得

$$\mathbb{E}\left[\exp(xB_{T_{k,a}} - \frac{1}{2}x^2 T_{k,a})\right] = 1,$$

而$B_{T_{k,a}} = kT_{k,a} + a$, 因此

$$\mathbb{E}\left(e^{-(\frac{1}{2}x^2 - kx)T_{k,a}}\right) = e^{-xa},$$

令$\frac{1}{2}x^2-kx=s>0$, x 有正解

$$x=k+\sqrt{k^2+2s},$$

故而有

$$\mathbb{E}\left[e^{-sT_{k,a}}\right]=\exp(-(k+\sqrt{k^2+2s})a).$$

当$k>0$ 时, 我们需要知道Brown 运动是否一定会碰到直线$y=kt+a$, 也就说需要算概率

$$\mathbb{P}(T_{k,a}<\infty).$$

方程(7.5.2) 依然成立, 但是现在当n 趋于无穷时不能保证期望内的量的有界性, 因此无法用控制收敛定理, 但是

$$xB_{T_{k,a}\wedge n}\leqslant x(k(T_{k,a}\wedge n)+a),$$

故而

$$\exp(xB_{T_{k,a}}-\frac{1}{2}x^2T_{k,a})\leqslant\exp(xa+(xk-\frac{1}{2}x^2)(T_{k,a}\wedge n)),$$

那么$xk-\frac{1}{2}x^2<0$, 也就是说$x>2k$ 时, 能够保证上面这个量有界了. 让n 趋于无穷, 在$\{T_{k,a}<\infty\}$ 上, $B_{T_{k,a}}=kT_{k,a}+a$, 因此有

$$\mathbb{E}[\exp((xk-\frac{1}{2}x^2)T_{k,a});T_{k,a}<\infty]=e^{-xa},$$

然后让$x\uparrow 2k$, 得

$$\mathbb{P}(T_{k,a}<\infty)=e^{-2ka}.$$

关于反射原理的严格证明要用到Brown 运动的强Markov 性.

定理7.5.3 对任何$(\mathcal{F}^0_{t+})$-停时T, $(B_{t+T}-B_T:t\geqslant 0)$ 是一个独立于$\mathcal{F}^0_{T+}$ 的标准Brown 运动. 或者说对任何Borel 集A, 有

$$\mathbb{P}(B_{t+T}-B_T\in A|\mathcal{F}_T)=\mathbb{P}(B_t\in A)\quad \text{a.s. on } \{T<\infty\}. \tag{7.5.3}$$

Proof. 设$T^{(n)}$ 是T 的离散化

$$T^{(n)}=\sum_{k\geqslant 1}\frac{k}{2^n}1_{\{\frac{k-1}{2^n}\leqslant T<\frac{k}{2^n}\}}.$$

那么对任何$H\in\mathcal{F}^0_{T+}$, 有$H\cap\{T<t\}\in\mathcal{F}_t$, 因此对任何$n\geqslant 1$,

$$\mathbb{P}(B_{t+T^{(n)}}-B_{T^{(n)}}\in A;H,T<\infty)$$

$$
\begin{aligned}
&= \sum_{k\geqslant 1}\mathbb{P}(B_{t+k/2^n}-B_{k/2^n}\in A; H\cap\{T^{(n)}=k/2^n\})\\
&= \sum_{k\geqslant 1}\mathbb{P}(B_{t+k/2^n}-B_{k/2^n}\in A)\mathbb{P}(H\cap\{T^{(n)}=k/2^n\})\\
&= \sum_{k\geqslant 1}\mathbb{P}(B_t\in A)\mathbb{P}(H\cap\{T^{(n)}=k/2^n\})\\
&= \mathbb{P}(B_t\in A)\mathbb{P}(T<\infty)\ .
\end{aligned}
$$

第二个等式成立是因为$(B_{t+s}-B_s: t\geqslant 0)$ 是独立于$\mathcal{F}_s^0$ 的标准Brown 运动. 让n 趋于无穷, 由Brown 运动连续性推出(7.5.3). □

让我们证明反射原理(7.2.1)

$$
\begin{aligned}
\mathbb{P}(T_b\leqslant t, B_t\leqslant a) &= \mathbb{P}(T_b\leqslant t, B_{(t-T_b)+T_b}-B_{T_b}\leqslant a-b)\\
&= \mathbb{P}(T_b\leqslant t)\mathbb{P}(B_{(t-T_b)+T_b}-B_{T_b}\leqslant a-b)\\
&= \mathbb{P}(T_b\leqslant t)\mathbb{P}(B_{(t-T_b)+T_b}-B_{T_b}\geqslant -(a-b))\\
&= \mathbb{P}(T_b\leqslant t, B_t\geqslant 2b-a).
\end{aligned}
$$

第三个等式来自标准Brown 运动的对称性.

参考文献

[1] Billingsley, P., PROBABILITY AND MEASURE, John Wiley & Sons, 1986

[2] Billingsley, P., CONVERGENCE OF PROBABILITY MEASURES, John Wiley & Sons, 1968

[3] Bremaud, MARKOV CHAINS

[4] DOYLE, P.G., SNELL, J.L., RANDOM WALKS AND ELECTRIC NETWORKS, Mathematical Association of America, Washington, DC, 1984

[5] Feller, W., PROBABILITY THEORY AND ITS APPLICATION, Vol. I(1959: Third edition), Vol. II(1970), Wiley & Son

[6] Kallengberg, O., FOUNDATIONS OF MODERN PROBABILITY, 科学出版社, Springer, 2001

[7] Lyons, R., PROBABILITY ON TREES AND NETWORKS, Probability web. 2003

[8] Shreve, S., STOCHASTIC CALCULUS AND FINANCE, Probability web, 2003

[9] 应坚刚, 何萍, 概率论, 复旦大学出版社, 上海, 2005

[10] 应坚刚, 金蒙伟, 随机过程基础, 复旦大学出版社, 上海, 2005