应用随机分析

Applied Stochastic Calculus

部分讲义 2023 版

2023 版讲义停止更新修订

刘勇　编著

北京大学数学科学学院

2023 年 6 月 18 日

注意事项

2023 版讲义中蓝色标注部分是对 2022 版有较大调整的地方

参考教材

主要参考教材:

龚光鲁 随机微分方程及其应用概要 清华大学出版社 2014 年 1 月第 3 次印刷

参考书目:

1. Øksendal, B., *Stochastic Differential Equations: An Introduction with Application.* 6th ed. Springer 2005
2. Klebaner, F. C., *Introduction to Stochastic Calculus with Applications.* 3rd ed.
随机分析及其应用 人民邮电出版社 2008 (英文影印版)
3. Lawler, G. F., *Introduction to Stochastic Processes.* 2nd ed. Chapman & Hall/CRC 2006
随机过程导论 张景肖 译 机械工业出版社 2010
4. Pavliotis G. A., *Stochastic Processes and Applications, Diffusion Processes, the Fokker-Planck and Langevin Equtions.* Springer 2014

http://www.math.pku.edu.cn/teachers/liuyong/teachingindex.html

课程设置和要求

所有内容解释权归任课老师

成绩: 平时成绩 20%; 期中测验 30%; 期末考试 50%. 平时成绩包括作业与出勤情况.

作业: 请交纸质版作业

收发时间: 每单周周二上课时交 (第一周除外), 双周周四发作业.

补交规定: 补交作业只记录不批改, 补交作业的最后期限 2023 年 6 月 18 日期末考试发试卷前, 过此时间点补交的作业一律不计入平时成绩.

其它事项 如果有同学不愿意做作业, 请写书面声明, 2023 年 3 月 7 日上课时交给任课老师. 总评成绩另行计算.

出勤: 遵守学校相关规定. 可能会有临时点名.

期中测验: 期中测验根据进度定在第 6-9 周，一个学时 (50 分钟). 提前两周通知.

答疑: 答疑时间每周二 15:00-16:00, 地点理科一号楼 1588. 也可以用邮件与老师预约答疑时间.

预约邮件要求:

邮件标题: 应用随机分析答疑

邮件内容: 姓名, 学号, 下周有空时间段 (缺一不可). 请尽量详述问题和你的困惑

截止时间: 每周五上午 10 点

其它事项: 不留姓名和学号的邮件将被视为垃圾邮件

更改邮件标题将可能导致你的邮件被遗漏

目录

第一章　概率论和随机过程的基本概念

1.1　引言

我们在概率论和应用随机过程课程中主要讲一些特殊的过程及其性质, 例如: 随机游动、Poisson 过程、Brown 运动和马氏链等. 那么, 对于较为一般的随时间演化的随机现象, 也就是随机过程, 我们将如何研究呢?

中学时我们学习的一些初等函数及其性质, 例如: 二次函数、三角函数; 在大学微积分中, 我们是对较为一般的 “好” 函数进行研究的. 引入积分和微分, 进一步使用牛顿-莱布尼兹公式, 从而我们能研究常微分方程 (ordinary differential equation, ODE)、偏微分方程 (partial differential equation, PDE) 等. 那么这些方法的内在想法又是什么?

我们学过的应用随机过程就相当于中学学的若干初等函数和它们的性质, 而应用随机分析这门课就相当于我们大学学的微积分. 具体地说, 首先, 应该从数学上来定义什么是 “随机性”? 也就是需要用公理化的方法, 对一类现象来定义 “随机性”, 并且量化. 在数学上定义一类随机现象关于时间的随机性, 这就是 “鞅 (martingale)”. 为了定义 “鞅”, 我们需要借助严格公理化意义下条件期望的概念来实现. 因此, 我们必须一开始就要了解概率论的公理化是如何建立的.

接着, 有了鞅之后, 一个自然的问题就是鞅是否可以用已知的一些 “初等” 或是 “特殊” 的过程 “简单” 的表示出来. 这就是随机积分, 例如: 关于 Brown 运动或 Poisson 点过程的随机积分. 有了随机积分之后, 就要有关于随机积分的计算公式, 这就是 Itô 公式. 它相当于微积分中的牛顿-莱布尼兹公式. 有了这些准备之后, 我们就可以把一大类随机现象表示成一个 “确定性部分”+“鞅 (随机积分) 部分”, 也就是 Itô 过程, 而 Itô 随机微分方程 (stochastic differential equation, SDE) 是其中一种特殊的过程. 如同常微分方程和偏微分方程, 我们可以通过随机微分方程来描述更多的随机现象, 利用相关的数学方法得到随机微分方程的很多性质.

1.2　概率论的公理化结构

在自然界和社会生活中, 有很多不确定的现象. 在一组固定的条件下, 既可能发生, 又可能不发生的现象称为随机现象. 对随机现象的研究必然要进行 “调查”、“观测” 和 “实验”, 称为随机试验 (trial). 在概率论中, 我们假定随机试验可以在相同条件下重复地进行, 每次试验的结果可能不止一个, 并且能事先确定试验的所有可能结果, 但每次试验的结果事先又不可预测. 我们通过对随机试验建立数学模型来研究随机现象的规律性. 这个数学模型应该具有哪些要素?

首先, 需要知道随机试验的所有可能的结果是什么? 因此, 我们引入以下的概念.

定义 1.2.1. *把随机试验每一个可能的结果称为一个***样本点** *(sample point), 通常用 ω 表示; 所有可能的结果组成的集合称为***样本空间** *(sample space), 通常用 Ω 表示.*

例如, 先后掷两次硬币这个随机试验可能出现的结果是 (正, 正)(正, 反)(反, 正)(反, 反), 把这四个结果作为样本点构成样本空间.

接着, 我们感兴趣试验中出现的一些事儿, 比如，在上面的例子中可能感兴趣 “两次出现的结果相同” 这件事儿. 它是指 (正, 正)(反, 反) 这两个样本点之一出现. 再比如, “第二次不出现反面” 这件事儿. 它是指 (正, 正)(反, 正) 这两个样本点之一出现. 实际上, 它们都是一些样本点的集合.

这里的给定一个点的集合 A, 是指对于任何一个点 ω, 都可以判断它是不是属于 A. 如果是, 则记为 $\omega \in A$; 如果不是, 则记为 $\omega \notin A$[①]. 我们约定不包含任何点的集合也是点集, 称为空集, 记为 $\emptyset$. 我们引入以下的概念.

定义 1.2.2. 我们把**事件** (event) 定义为样本点的某个集合. 称某事件发生当且仅当它所包含的某个样本点出现.

虽然试验的样本点在试验前就很明确, 但是只有试验之后, 才能确定某个给定的事件是否发生.

我们把样本空间 Ω 本身也作为一个事件. 每次试验必然有 Ω 中的某个样本点出现, 即 Ω 必然发生. 因此, 我们称 Ω 为**必然事件**. 我们把空集 $\emptyset$ 也作为一个事件. 每次试验中, 它都不发生, 因此, 称为 $\emptyset$ 为**不可能事件**.

通常我们希望通过一些简单事件来表示复杂事件, 利用简单事件的信息来了解复杂事件的信息. 因此, 我们还需要知道事件的关系和事件的运算. 事件作为样本空间的子集, 本质上它们之间的关系和运算就是集合之间的关系与运算. 我们在此需要强调的是用概率论的语言来解释这些关系及运算.

(1) 若事件 A 中的每一个样本点都包含在事件 B 中, 称事件B **包含了事件** A; 或者说**事件** A **发生蕴含事件** B **发生**, 记为 $A \subset B$. 若 $A \subset B$ 和 $B \subset A$ 同时成立, 称**事件** A **与事件** B **等价**或A **等于** B, 记为 $A = B$.

(2) 由所有不包含在事件 A 中的样本点所组成的事件称为事件 A 的**逆事件**或**对立事件**, 记为 A^c. A^c 表示事件 A 不发生.

(3) 用 $A \cap B$ 或者 AB 表示事件 A 和事件 B 同时发生, 即 $A \cap B$ 或者 AB 表示所有同时属于 A 和 B 的样本点的集合, 称为A **与** B **的交**. 用 $\cap_{n=1}^{\infty} A_n$ 表示事件 $A_n, n = 1, 2, \ldots$ 同时发生, 即所有同时属于 $A_n, n = 1, 2, \ldots$ 的样本点的集合, 称为$A_n, n = 1, 2, \ldots$ **的交**

(4) 用 $A \cup B$ 表示事件 A 和和事件 B 至少有一个发生, 即 $A \cup B$ 表示至少属于 A 或 B 中的一个的所有样本点的集合, 称为A **与** B **的并**. 用 $\cup_{n=1}^{\infty} A_n$ 表示事件 $A_n, n = 1, 2, \ldots$ 中至少一个发生, 即表示至少属于 $A_n, n = 1, 2, \ldots$ 中一个的所有样本点的集合, 称为$A_n, n = 1, 2, \ldots$ **的并** .

(5) 若 $AB = \emptyset$, 表示事件 A 和事件 B 不同时发生, 称A **与** B **为互不相容**, 即属于 A 的样本点必不属于 B, 属于 B 的样本点必不属于 A. 若事件 $A_n, n = 1, 2, \ldots$ 两两互不相容, 即任意的 $n \neq m$, $A_n \cap A_m = \emptyset$, 称 $A_n, n = 1, 2, \ldots$ **互不相容**.

(6) 用 $A \setminus B$ 表示事件 A 发生但事件 B 不发生, 即 $A \setminus B$ 表示所有属于 A 但不属于 B 的样本点的集合, 称为A **与** B **的差**.

在本讲义中, 对于事件的运算顺序做如下的约定: 先进行逆运算, 再进行交运算, 最后是并运算或差运算. 进一步, 事件之间的关系还成立如下的运算律:

交换律: $AB = BA$; $A \cup B = B \cup A$.

结合律: $(AB)C = A(BC)$; $(A \cup B) \cup C = A \cup (B \cup C)$.

分配律: $(A \cup B) \cap C = (AC \cup BC)$; $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

[①]这段话的理解可以借鉴如下英文表达: A particular set A is well defined if it is possible to tell whether any given point *belongs to* it or not. These two cases are denoted respectively by $\omega \in A$, $\omega \notin A$. Thus a set is determined by specified rule of membership.
英文摘自 K. L. Chung(钟开莱), F. AitSahlia, *Elementary Probability Theory.* 4th ed. Springer 2003. page 2-3

De Morgan 定理 (对偶原理): $(A \cup B)^c = A^c \cap B^c$; $(A \cap B)^c = A^c \cup B^c$.

我们还关心一些事件 (发生) 的**概率**. 所谓概率就是度量事件发生的可能性大小的量, 实际上就是以某些事件为自变量的非负函数, 必然事件的概率是 1, 不可能事件的概率是 0, 互不相容事件并的概率等于每个事件概率之和.

概括起来, 建立随机试验的数学模型时, 我们必须知道: (1) 试验的样本空间 Ω. 它应该是一个非空的集合; (2) 可以 "观测" 到的或感兴趣的事件以及这些事件通过事件的运算得到的事件的全体, 记为 $\mathcal{F}$; (3) 这些事件的概率. 更进一步, 还需要知道 $\mathcal{F}$ 满足怎样的数学结构以及各事件的概率 P 之间的关系.

当研究某个随机试验时, 首先应该有我们感兴趣的事件, 而且知道

- 如果事件 A 发生, 则可以推知 A^c 不发生. 也就是说, 如果 A 是我们感兴趣的事件, 则 A^c 也应该是我们感兴趣的事件.
- 如果 $A_n, n = 1, 2, \ldots$ 之一发生, 则可以推知事件 $\cup_{n=1}^{\infty} A_n$ 发生. 也就是说, 如果 $A_n, n = 1, 2, \ldots$ 是我们感兴趣的事件, 则 $\cup_{n=1}^{\infty} A_n$ 也应该是我们感兴趣的事件.

由此, 我们引入以下的概念.

定义 1.2.3. *$\mathcal{F}$ 是由样本空间 Ω 的一些子集组成的集合, 如果满足*

(1) *$\mathcal{F}$ 非空;*

(2) *$A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$, 即 $A \in \mathcal{F}$ 蕴含 $A^c \in \mathcal{F}$;*

(3) *$A_n \in \mathcal{F}$, $n = 1, 2, \cdots \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}$, 即 $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ 蕴含 $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$,*

则称 $\mathcal{F}$ 为事件域 (event field).

由事件域的定义, 有如下简单的性质, 证明留作习题.

性质 1.2.1. *$\mathcal{F}$ 是事件域, 则 $\Omega \in \mathcal{F}$; $\emptyset \in \mathcal{F}$.*

结合事件域的概念, 我们引入概率的公理化定义.

定义 1.2.4. *定义在事件域 $\mathcal{F}$ 上的一个集合函数 P 称为***概率** (或**概率测度**), *如果它满足*

(1) *非负性: 对于任意的 $A \in \mathcal{F}$, $\mathrm{P}(A) \geq 0$;*

(2) *规范性: $\mathrm{P}(\Omega) = 1$;*

(3) *可列可加性或完全可加性: 若 $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ 互不相容, 则*

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n). \tag{1.2.1}$$

利用概率的性质, 有如下简单的结论, 证明留作习题.

性质 1.2.2. $\mathrm{P}(\emptyset) = 0$.

定义 1.2.5. *称三元体 $(\Omega, \mathcal{F}, \mathrm{P})$ 为***概率空间**, *其中 Ω 是样本空间, $\mathcal{F}$ 是关于这个样本空间的一个事件域, P 是定义在 $\mathcal{F}$ 上的概率.*

这是 Kolmogorov 在 1933 年建立的概率论公理化结构. 一个概率空间就是描述一个随机试验的数学模型. 对于具体的随机试验如何构建概率空间, 需要视具体问题而定. 我们特别强调: 此后, 当概率空间确定好, 只有**事件域 $\mathcal{F}$ 中的元素才称为事件**, 也就是说当表述中涉及到事件时, 一定要明确它是哪个事件域的元素. 例如:

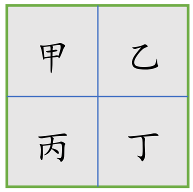


图 1

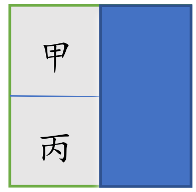


图 2

例 1.2.1. 有一个正方形盒子, 盒盖是透明的, 上面画"十"字线, 从盒子正上方看将盒子分成四块区域 (如图 1). 盒中有一个小球可以滚动. 随意晃动盒子, 小球停止运动后, 随机地停留在这四块区域中的某块中 (设想是理想状态, 忽略小球处于"十"字线上的情况).

当考虑盒子经晃动后小球停留在哪些区域这个试验时, 样本空间是 $\Omega = \{$甲, 乙, 丙, 丁$\}$, 事件域 $\mathcal{F}_1$ 为 Ω 的全体子集组成的集合, 其中共有 16 个事件. 对于试验的每一个结果 ω, 对于 Ω 的每个子集 A, 总能判断出 ω 是否属于 A, 也就是说每次实验后, 总能知道事件 A 是否发生.

如果再设想用一块不透明的材料将"乙"和"丁"区域遮盖住 (如图 2). 同样考虑将盒子晃动后小球停留在哪几块区域的试验, 我们也可以将试验的样本空间取为 $\Omega = \{$甲, 乙, 丙, 丁$\}$. 但是, 此时事件域

$$\mathcal{F}_2 = \big\{\Omega, \emptyset, \{\text{甲}\}, \{\text{丙}\}, \{\text{甲, 丙}\}, \{\text{乙, 丁}\}, \{\text{甲, 乙, 丁}\}, \{\text{丙, 乙, 丁}\}\big\},$$

其中只包含了 8 个事件.

对于试验的某些结果, 虽然可以看到小球停留在被遮蔽的区域, 但是我们不能判断此时小球是否在"乙"区域, 也就是说, 不知道 $\{$乙$\}$ 这事儿是否发生了. 因此, 对于事件域 $\mathcal{F}_2$, Ω 的子集 $\{$乙$\}$ 就不是事件. 同样的, $\{$丁$\}$ 和 $\{$甲, 乙$\}$ 等也不是事件.

在测度论中, 设 X 是一个非空集合, $\mathcal{C}$ 是 X 的一些子集组成的集合, 称为集合系 (或集类、集族). 若 $\mathcal{C}$ 满足定义1.2.3中的 (1)(2)(3), 则称 $\mathcal{C}$ 为**σ-域** (或**σ-代数**). 若 $\mathcal{C}$ 上的集合函数 μ 满足定义1.2.4中的非负性和可列可加性, 则称 μ 为 $\mathcal{C}$ 上的**测度**.

设 $\mathcal{F}$ 为 X 上的 σ-域, 称 $(\mathrm{X}, \mathcal{F})$ 为**可测空间**. X 的子集 A 称为 $\mathcal{F}$ 可测的, 若 $A \in \mathcal{F}$. 设 μ 是 σ-域 $\mathcal{F}$ 上的测度, 则称 $(\mathrm{X}, \mathcal{F}, \mu)$ 为**测度空间**. 概率空间是一类特殊的测度空间.

在测度论或概率论的研究中, 往往可以较直接的知道某个集合系 $\mathcal{C}$ 上的测度 μ, 但 $\mathcal{C}$ 未必是 σ-域. 为了构造可测空间需要增添适量的子集将 $\mathcal{C}$ 扩充成 σ-域, 还需要给新增加的子集合理地赋予数值, 成为一个测度. 这是测度论中测度扩张定理来具体实现的, 需要较多的概念和知识, 已超出了这本书的范围. 但下面的概念对本书的后续讲解是有帮助的.

定义 1.2.6. $\mathcal{C}$ 是 X 的非空集合系, $\mathcal{S}$ 是 X 的 σ-域. 若 $\mathcal{S}$ 满足下面的条件:

(1) $\mathcal{S} \supset \mathcal{C}$;

(2) 对于 X 上的任一 σ-域 $\mathcal{S}'$, 都有

$$\mathcal{S}' \supset \mathcal{C} \Longrightarrow \mathcal{S}' \supset \mathcal{S}.$$

则称 $\mathcal{S}$ 是由集合系 $\mathcal{C}$ **生成的 σ-域**, 或**包含 $\mathcal{C}$ 的最小 σ-域**. 记为 $\sigma(C)$.

性质 1.2.3. 由任何集合系 $\mathcal{C}$ 生成的 σ-域 $\sigma(C)$ 存在且唯一.

这个性质的证明留作习题.

我们介绍在概率论中常用的 Borel σ-域. 设 $\mathbb{R}$ 是实数全体, 记 $\mathbb{R}$ 上左开右闭区间组成的集合系为

$$\mathcal{C} = \{(a, b] : a, b \in \mathbb{R}, a < b\}.$$

称 $\sigma(\mathcal{C})$ 为 $\mathbb{R}$ 上的 Borel σ-域, 记为 $\mathcal{B}(\mathbb{R})$, 其中的元素称为 Borel 集. 不难验证对于左闭右开区间组成的集合系、开区间组成的集合系以及闭区间组成的集合等都生成 $\mathcal{B}(\mathbb{R})$.

例 1.2.2. 考虑 $\mathcal{C}$ 上的测度 μ:

$$\mu\big((a,b]\big) = b-a.$$

利用测度扩张定理可以唯一的将 $\mathcal{C}$ 上的测度 μ 扩张到 $\mathcal{B}(\mathbb{R})$ 上. 即, 在 $\mathcal{B}(\mathbb{R})$ 存在唯一的测度 $\widetilde{\mu}$ 满足, 对于任意的 $(a,b]\in\mathcal{C}$,

$$\widetilde{\mu}\big((a,b]\big) = b-a.$$

$\widetilde{\mu}$ 称为 $\mathbb{R}$ 上 Lebesgue 测度.

在 Lebesgue 测度建立过程中, 有两点需要强调一下: (1) 由实变函数知道, 并不是 $\mathbb{R}$ 的任意子集都是 Borel 集. 也就是说, 虽然 $\mathbb{R}$ 上的子集全体 $\mathscr{P}(\mathbb{R})$ 也是一个 σ-域, 但是并不能在其上构造一个测度 $\widetilde{\mu}$ 满足 $\widetilde{\mu}\big((a,b]\big) = b-a$ [①]

这也说明无论是测度论还是概率论, 恰当地确定 σ-域或事件域是非常关键的; (2) 在利用测度扩张定理时, 对于集合系 $\mathcal{C}$ 及其上的测度 μ 都是有要求的. 有兴趣的读者可以参阅相关教材.

在 n-维欧氏空间 $\mathbb{R}^n$ 中, 如果 $\mathbf{a}=(a_1,\dots,a_n),\mathbf{b}=(b_1,\dots,b_n)$ 满足 $a_i<b_i, i=1,\dots,n$, 记 $(\mathbf{a},\mathbf{b}]=\{\mathbf{x}\in\mathbb{R}^n : a_i<x_i\le b_i, i=1,\dots,n\}$. 设 $\mathcal{C}=\{(\mathbf{a},\mathbf{b}] : \mathbf{a},\mathbf{b}\in\mathbb{R}^n\}$, 称 $\sigma(\mathcal{C})$ 为 $\mathbb{R}^n$ 上的 Borel σ-域, 记为 $\mathcal{B}(\mathbb{R}^n)$, 其中的元素称为 Borel 集.

f 是从 $\mathbb{R}^n$ 映射到 $\mathbb{R}^d$ 的函数, 如果对于任意的 $\mathbf{a}\in\mathbb{R}^d$, 都有

$$\{\mathbf{x}\in\mathbb{R}^n : f_i(\mathbf{x})\le a_i,\ i=1,\dots,d\}\in\mathcal{B}(\mathbb{R}^n),$$

则称 f 是 Borel 函数. 根据测度论, Borel 函数的定义等价于: 对于任意的 $A\in\mathcal{B}(\mathbb{R}^d)$, $\{\mathbf{x}\in\mathbb{R}^n : f(\mathbf{x})\in A\}\in\mathcal{B}(\mathbb{R}^n)$.

1.3 随机变量

当建立好一个随机试验的概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 之后, 如果我们能直接得到每个事件的概率, 那么相当于完全地知道了这个试验的全部信息. 但在现实中这样的试验并不多见, 例如一些简单的古典概型. 然而, 随机现象纷繁复杂丰富多彩, 使得相应的样本空间千差万别. 有些试验结果可以用数字表示, 例如: 测量误差; 有些不是, 例如: 掷硬币的结果用正面和反面来表示. 对于复杂的试验, 我们总是希望利用简单事件的概率来推算出复杂事件的概率. 这样, 需要使用更多统一的数学方法来研究, 比如将试验结果 "数字化", 用一个数字 ξ 来表示, 也就是说 ξ 是以样本空间为定义域取值于 $\mathbb{R}$ 的函数. 但 ξ 仅仅是一个函数是不够的, 对于随机试验除了样本空间, 还需要关注事件域 $\mathcal{F}$, ξ 应该与 $\mathcal{F}$ 发生联系, 即 $\{\omega : \xi(\omega)\le x\}$ 应该是事件. 由此, 我们引入以下的概念:

定义 1.3.1. 对于给定概率空间 $(\Omega,\mathcal{F},\mathrm{P})$, ξ 是从 Ω 到 $\mathbb{R}$ 的函数, 如果任意的 $x\in\mathbb{R}$, 都有 $\{\omega:\xi(\omega)\le x\}\in\mathcal{F}$, 则称 ξ 是**随机变量** (random variable).

根据测度论, 随机变量的定义等价于任意的 $A\in\mathcal{B}(\mathbb{R})$, 都有 $\{\omega:\xi(\omega)\in A\}\in\mathcal{F}$, 即它是一个事件. 需要强调的是, 随机变量的定义依赖于事件域 $\mathcal{F}$. 在本书以后的行文中, 如果需要, 我们会指明随机变量是关于哪个事件域的随机变量. 如果 ξ 是关于 $\mathcal{F}$ 的随机变量, f 是 Borel 函数, 则 $f(\xi)$ 还是关于 $\mathcal{F}$ 的随机变量.

设 $A\subset\Omega$, 集合 A 的**示性函数**定义为:

$$\mathbf{1}_A(\omega)=\begin{cases}1 & \omega\in A,\\ 0 & \omega\notin A.\end{cases}$$

当 $A\in\mathcal{F}$ 时, $\mathbf{1}_A$ 是关于 $\mathcal{F}$ 的随机变量, 否则不是.

[①] 严格地说, 利用测度扩张定理可以将 μ 扩张到一个比 $\mathcal{B}(\mathbb{R})$ 更大的 σ- 域 $\mathcal{L}$ 上. $\mathcal{L}$ 中的元素称为 Lebesgue 可测集. 但 $\mathcal{L}$ 也不是 $\mathscr{P}(\mathbb{R})$. 设 $\mathcal{N}=\{N\subset\mathbb{R} : \exists A\in\mathcal{B}(\mathbb{R}), \widetilde{\mu}(A)=0 \text{ 使得 } A\supset N\}$, 则

$$\mathcal{L}=\{A\cup N : A\in\mathcal{B}(\mathbb{R}), N\in\mathcal{N}\}.$$

例 1.3.1. (接例1.2.1) 设

$$\xi(\omega)=\begin{cases}1 & \omega=\text{甲},\\ 2 & \omega=\text{乙},\\ 3 & \omega=\text{丙},\\ 4 & \omega=\text{丁}.\end{cases}$$

显然, ξ 是关于 $\mathcal{F}_1$ 的随机变量. 但它不是关于 $\mathcal{F}_2$ 的随机变量, 这是因为 $\{\omega:\xi(\omega)\le 2.5\}=\{\text{甲, 乙}\}\notin\mathcal{F}_2$.

本节中我们仅简单地介绍与随机变量相关的基本概念和本书中常用的数字特征及必要的一些计算公式. 这部分内容在概率论或概率统计的相关教材都有详细的讲解.

定义 1.3.2. 设 ξ 是概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量. 任意的 $A\in\mathcal{B}(\mathbb{R})$, $\mathrm{P}(\omega:\xi(\omega)\in A)$ 构成 $\mathcal{B}(\mathbb{R})$ 上的一个概率, 称为 ξ 的**概率分布** (probability distribution), 简称**分布** (distribution). 两个随机向量 ξ 和 η 如果有相同的概率分布, 则称 ξ 和 η **同分布**.

定义 1.3.3. 设 ξ 是概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量, 称 x 的函数

$$F(x)=\mathrm{P}(\omega:\xi(\omega)\le x),\ x\in\mathbb{R}$$

为 ξ 的**分布函数** (distribution function).

♣♣♣20230221

ξ 和 η 可以是两个不同概率空间上的随机变量, 但它们可以有相同的分布函数. 由测度论可以知道, ξ 和 η 的分布函数相同等价于它们同分布.

在本书中我们主要涉及到以下两类随机变量[①]:

(i) 如果随机变量 ξ 只取有限个值 $x_1,\dots,x_n$ 或可列个值 $x_1,x_2,\dots,x_n,\dots$, 则称 ξ 是**离散型随机变量**. 记 $p(x_i)$ 为 $\mathrm{P}(\omega:\xi(\omega)=x_i)$, $i=1,2,\dots$, 则 ξ 的概率分布可以通过如下的分布列来表示:

$\xi(\omega)$	x_1	x_2	$\dots$	x_i	$\dots$
P	$p(x_1)$	$p(x_2)$	$\dots$	$p(x_i)$	$\dots$

(1.3.1)

(ii) 对于随机变量 ξ, 如果存在非负可积函数 $p(x)$ 使其分布函数 $F(x)$ 可以表示为

$$F(x)=\int_{-\infty}^{x}p(y)\mathrm{d}y,\ x\in\mathbb{R},\tag{1.3.2}$$

称 ξ 为**连续型随机变量**. $p(x)$ 称为 ξ 的**密度函数** (density function). $p(x)$ 满足

$$\int_{-\infty}^{\infty}p(y)\mathrm{d}y=1,$$

而且 $F(x)$ 连续且 $p(x)=F'(x)$.

定义 1.3.4. 设 ξ 是 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量, 如果 $\int_\Omega|\xi(\omega)|\mathrm{dP}(\omega)<\infty$, 则称 ξ 的**数学期望存在**, 定义

$$\mathrm{E}\xi=\int_\Omega\xi(\omega)\mathrm{dP}(\omega)$$

为 ξ 的**数学期望** (简称**期望**或**均值**). 如果 $\int_\Omega|\xi(\omega)|\mathrm{dP}(\omega)=\infty$, 则称 ξ 的**数学期望不存在**.

我们这里直接引用 $(\Omega,\mathcal{F},\mathrm{P})$ 上抽象 Lebesgue 积分 $\int_\Omega\xi(\omega)\mathrm{dP}(\omega)$ 给出数学期望的定义. Lebesgue 积分实质上可以理解为加权平均. 直观上, ξ 的值落入 x 的 ε-邻域 $[x-\varepsilon,x+\varepsilon)$ 的概率是 $\mathrm{P}\big(\omega:\xi\in[x-\varepsilon,x+\varepsilon)\big)$, 即 ξ 在 x 附近取值的 "权重" 近似地取成 $\mathrm{P}\big(\omega:\xi(\omega)\in[x-\varepsilon,x+\varepsilon)\big)$. 将 ξ 的值域划分成至多可列个这样互不相交的 ε-邻域 $[x_i-\varepsilon,x_i+\varepsilon)$, 那么, ξ 的加权平均就近似地等于

$$\sum_i x_i\mathrm{P}\big(\omega:\xi(\omega)\in[x_i-\varepsilon,x_i+\varepsilon)\big).$$

[①]事实上, 利用测度论中 Lebesgue 分解定理对随机变量的分布作分解, 除离散型和连续型随机变量外, 还可以得到一类奇异型随机变量. 它的分布函数是连续的, 但不存在密度函数使其可以表示成 (1.3.2) 中不定积分的形式. 同时 Lebesgue 分解定理也告诉我们任何随机变量的分布都是这三种类型随机变量分布的混合.

当划分越来越细, ε 越来越小时, 如果上面的求和式收敛, 则其极限就是数学期望 (Lebesgue 积分)[①].

由此可以知道, 设 f 是 Borel 函数, 对于随机变量的函数 $f(\xi)$, 如果 $\int_\Omega |f(\xi(\omega))|\mathrm{dP}(\omega) < \infty$, 则 $\mathrm{E}f(\xi) = \int_\Omega f(\xi(\omega))\mathrm{dP}(\omega)$. 本书中更多地是讨论离散型和连续型随机变量, 我们给出这两类随机变量数学期望如下的公式.

性质 1.3.1. 设 ξ 是离散型随机变量, 其分布列为 (1.3.1), 如果 $\sum_i |x_i|p(x_i)$ 收敛, 则

$$\mathrm{E}\xi = \sum_i x_i p(x_i).$$

设 f 是 Borel 函数, 如果 $\sum_i |f(x_i)|p(x_i)$ 收敛, 则

$$\mathrm{E}f(\xi) = \sum_i f(x_i)p(x_i).$$

设 ξ 连续型随机变量, 其密度函数为 $p(x)$, 如果 $\int_{-\infty}^{\infty} |x|p(x)\mathrm{d}x < \infty$, 则

$$\mathrm{E}\xi = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x.$$

如果 $\int_{-\infty}^{\infty} |f(x)|p(x)\mathrm{d}x < \infty$, 则

$$\mathrm{E}f(\xi) = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x.$$

在本节和下一节中, 当讨论某随机变量的数学期望时, 如无特殊说明均表示其数学期望存在.

性质 1.3.2. 数学期望的基本性质:

(1) 对于事件 A, $\mathrm{E}\mathbf{1}_A = \mathrm{P}(A)$.

(2) 若 $\xi \geq \eta$, 则 $\mathrm{E}\xi \geq \mathrm{E}\eta$. 特别地, $\xi \geq 0$, 则 $\mathrm{E}\xi \geq 0$.

(3) $\mathrm{E}(a\xi + b\eta) = a\mathrm{E}\xi + b\mathrm{E}\eta$, 其中 a, b 是常数.

(4) $|\mathrm{E}\xi| \leq \mathrm{E}|\xi|$.

(5) Cauchy-Schwarz 不等式: 如果 $\mathrm{E}\xi^2 < \infty, \mathrm{E}\eta^2 < \infty$, 则

$$\big(\mathrm{E}(\xi\eta)\big)^2 \leq \mathrm{E}\xi^2\mathrm{E}\eta^2.$$

等号成立的充分必要条件是存在不全为 0 的常数 a, b, 使得

$$\mathrm{P}\big(\omega : a\xi(\omega) + b\eta(\omega) = 0\big) = 1.$$

可以利用抽象 Lebesgue 积分的理论证明以上性质, 也可以通过离散型或连续型随机变量数学期望的公式验证.

定义 1.3.5. 称 $\mathrm{var}\xi = \mathrm{E}(\xi - \mathrm{E}\xi)^2$ 为随机变量 ξ 的**方差** (variance); $\sqrt{\mathrm{var}\xi}$ 称为**标准差** (standard deviation).

定义 1.3.6. 设 ξ, η 是 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的两个随机变量.

(1) 称 $\mathrm{cov}(\xi, \eta) = \mathrm{E}\big((\xi - \mathrm{E}\xi)(\eta - \mathrm{E}\eta)\big)$ 为 ξ 和 η 的**协方差** (covariance). 当 $\mathrm{cov}(\xi, \eta) = 0$ 时, 称 ξ 和 η **不相关**.

(2) 当 $0 < \mathrm{var}\xi < \infty, 0 < \mathrm{var}\eta < \infty$ 时, 称

$$\rho_{\xi\eta} = \frac{\mathrm{cov}(\xi, \eta)}{\sqrt{\mathrm{var}\xi}\sqrt{\mathrm{var}\eta}},$$

称为 ξ 和 η 的**相关系数** (correlation coefficient).

由 Cauchy-Schwarz 不等式 $|\rho_{\xi\eta}| \leq 1$. 当 $\xi = \eta$ 时, 方差是协方差的特例.

定义 1.3.7. 称 $\phi(\theta) = \mathrm{E}e^{\mathrm{i}\theta\xi}, \theta \in \mathbb{R}$ 为 ξ 的**特征函数** (characteristic function).

定义 1.3.8. 如果 $\mathrm{E}e^{s\xi} < \infty$, 则称 $\mathrm{M}(s) = \mathrm{E}e^{s\xi}$ 为 ξ 的**矩母函数** (moment generating fuction).

[①] 严格地说, 应该是 $\sum_i |x_i|\mathrm{P}(\omega : \xi \in [x - \varepsilon, x + \varepsilon))$ 收敛, $\sum_i x_i\mathrm{P}(\omega : \xi \in [x - \varepsilon, x + \varepsilon))$ 的极限值定义为数学期望.

1.4 随机向量

除了随机变量, 我们还经常需要研究随机向量.

定义 1.4.1. 设 $\xi_1,\xi_2,\cdots,\xi_n$ 是定义在同一概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量, 则称

$$\boldsymbol{\xi}(\omega)=\big(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega)\big)$$

称为 n **维随机向量**. 一维随机向量就是随机变量.

定义 1.4.2. 设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的 n 维随机变量. 任意的 $A\in\mathcal{B}(\mathbb{R}^n)$, $\mathrm{P}(\omega:\boldsymbol{\xi}(\omega)\in A)$ 构成 $\mathcal{B}(\mathbb{R}^n)$ 上的一个概率, 称为 $\boldsymbol{\xi}$ 的**概率分布** (probability distribution), 简称**分布** (distribution). 两个随机向量 $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 如果有相同的概率分布, 则称 $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ **同分布**.

定义 1.4.3. 设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的 n 维随机变量, 称 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1,\cdots,x_n)=\mathrm{P}(\omega:\xi_1(\omega)\le x_1,\cdots,\xi_n\le x_n),\ (x_1,\cdots,x_n)\in\mathbb{R}^n$$

为 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的**联合分布函数** (joint distribution function).

同样的, $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 可以是两个不同概率空间上的随机向量, 但它们可以有相同的联合分布函数. 由测度论可以知道, $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 的分布函数相同等价于它们同分布.

定义 1.4.4. 设 $\boldsymbol{\xi}=\{\xi_1,\cdots,\xi_n\}$ 是 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机向量, $d<n$, $1\le i_1<\cdots<i_d\le n$, 则任意的 $A\in\mathcal{B}(\mathbb{R}^d)$,

$$\mathrm{P}\big(\omega:(\xi_{i_1},\cdots,\xi_{i_d})\in A\big)$$

构成 $\mathcal{B}(\mathbb{R}^d)$ 上的一个概率, 称为 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1},\cdots,\xi_{i_d})$ 的**边际分布** (marginal distribution).

称 $(x_{i_1},\cdots,x_{i_d})\in\mathbb{R}^d$ 的函数

$$F_{i_1,\cdots,i_d}(x_{i_1},\cdots,x_{i_d})=\mathrm{P}\big(\omega:\xi_{i_1}(\omega)\le x_{i_1},\cdots,\xi_{i_d}(\omega)\le x_{i_d}\big)$$

为 $\boldsymbol{\xi}$ 关于 $(x_{i_1},\cdots,x_{i_d})$ 的**边际分布函数**.

在本书中我们更多关心的是两类随机向量:

(i) 设 $\xi_1,\cdots,\xi_n$ 是离散型随机变量, 则称 $\boldsymbol{\xi}=\{\xi_1,\cdots,\xi_n\}$ 为离散型随机向量. 设 ξ_i 的取值为 $x_i(k)$, $k\ge1$, $i=1,\cdots,n$, 则 $\boldsymbol{\xi}$ 的取值为 $\big(x_1(k_1),\cdots,x_n(k_n)\big)$, $k_1,\cdots,k_n\ge1$. $\boldsymbol{\xi}$ 的分布列记为

$$p\big(x_1(k_1),\cdots,x_n(k_n)\big)=\mathrm{P}\big(\omega:\xi_1(\omega)=x_1(k_1),\cdots,\xi_n(\omega)=x_n(k_n)\big).$$

设 $d<n$, $j_1<\cdots<j_{n-d}$, 且 $\{j_1,\cdots,j_{n-d}\}=\{1,\cdots,n\}\setminus\{i_1,\cdots,i_d\}$, 则 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1},\cdots,\xi_{i_d})$ 的边际分布列为

$$p\big(x_{i_1}(k_{i_1}),\cdots,x_{i_d}(k_{i_d})\big)=\sum_{k_{j_1}}\cdots\sum_{k_{j_{n-d}}}p\big(x_1(k_1),\cdots,x_n(k_n)\big).$$

(ii) 对于 n 维随机变量 $\boldsymbol{\xi}$, 如果存在非负可积函数 $p(x_1,\cdots,x_n)$ 使其联合分布函数 $F(x_1,\cdots,x_n)$ 可以表示为

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n,$$

称 $\boldsymbol{\xi}$ 为**连续型随机向量**. $p(x_1,\cdots,x_n)$ 称为 $\boldsymbol{\xi}$ 的**联合密度函数** (joint density function). 且满足 $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1$.

设 $d<n$, $j_1<\cdots<j_{n-d}$, 且 $\{j_1,\cdots,j_{n-d}\}=\{1,\cdots,n\}\setminus\{i_1,\cdots,i_d\}$, 则 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1},\cdots,\xi_{i_d})$ 的边际密度函数为

$$p_{i_1,\cdots,i_d}(x_{i_1},\cdots,x_{i_d})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n)\mathrm{d}x_{j_1}\cdots\mathrm{d}x_{j_{n-d}}.$$

性质 1.4.1. 设 $f(x_1,\cdots,x_n)$ 是 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的 Borel 函数, $\boldsymbol{\xi}$ 是 n 维连续型随机向量, 如果 $f(\boldsymbol{\xi})$ 的数学期望存在, 则

$$\mathrm{E}f(\boldsymbol{\xi})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

定义 1.4.5. 如果 $\mathrm{E}\xi_i$, $i=1,\cdots,n$ 都存在, 则称

$$\mathrm{E}\boldsymbol{\xi}=(\mathrm{E}\xi_1,\cdots,\mathrm{E}\xi_n)$$

为随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的**数学期望**.

定义 1.4.6. 如果 $\mathrm{E}\xi_i^2<\infty$, $i=1,\cdots,n$, 称

$$\boldsymbol{\Sigma}=\begin{pmatrix}\mathrm{cov}(\xi_1,\xi_1) & \mathrm{cov}(\xi_1,\xi_2) & \cdots & \mathrm{cov}(\xi_1,\xi_n)\\ \mathrm{cov}(\xi_2,\xi_1) & \mathrm{cov}(\xi_2,\xi_2) & \cdots & \mathrm{cov}(\xi_2,\xi_n)\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{cov}(\xi_n,\xi_1) & \mathrm{cov}(\xi_n,\xi_2) & \cdots & \mathrm{cov}(\xi_n,\xi_n)\end{pmatrix}$$

为随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的**协方差矩阵**.

协方差阵是非负定矩阵.

定义 1.4.7. 称 $\phi(\boldsymbol{\theta})=\mathrm{e}^{\mathrm{i}\sum_{i=1}^n\theta_i\xi_i},\boldsymbol{\theta}=(\theta_1,\cdots,\theta_n)\in\mathbb{R}^n$ 为 n 维随机向量 $\boldsymbol{\xi}$ 的**特征函数**.

定义 1.4.8. 如果 $\mathrm{E}\mathrm{e}^{\sum_{i=1}^n s_i\xi_i}<\infty$, $\mathbf{s}=(s_1,\cdots,s_n)\in\mathbb{R}^n$, 称 $\mathrm{M}(\mathbf{s})=\mathrm{E}\mathrm{e}^{\sum_{i=1}^n s_i\xi_i}$ 为 n 维随机向量 $\boldsymbol{\xi}$ 的**矩母函数**.

定义 1.4.9. 设 $\xi_1,\cdots,\xi_n$ 是随机变量. 如果对任意实数 $x_1,\cdots,x_n$,

$$\mathrm{P}\big(\omega:\xi_1(\omega)\le x_1,\cdots,\xi_n(\omega)\le x_n\big)=\mathrm{P}\big(\omega:\xi_1(\omega)\le x_1\big)\cdots\mathrm{P}\big(\omega:\xi_n(\omega)\le x_n\big),$$

则称随机变量 $\xi_1,\cdots,\xi_n$ **相互独立**.

$\xi_1,\cdots,\xi_n$ 相互独立等价于对于任意的 Borel 集 $A_i\in\mathcal{B}(\mathbb{R})$, $i=1,\cdots,n$,

$$\mathrm{P}\big(\omega:\xi_1(\omega)\in A_1,\cdots,\xi_n(\omega)\in A_n\big)=\mathrm{P}\big(\omega:\xi_1(\omega)\in A_1\big)\cdots\mathrm{P}\big(\omega:\xi_n(\omega)\in A_n\big).$$

定义 1.4.10. 设 $\boldsymbol{\xi}_i$ 是 d_i 维随机向量,$i=1,\cdots,n$. 如果对于任意的 Borel 集 $A_i\in\mathcal{B}(\mathbb{R}^{d_i})$,

$$\mathrm{P}\big(\omega:\boldsymbol{\xi}_1(\omega)\in A_1,\cdots,\boldsymbol{\xi}_n(\omega)\in A_n\big)=\mathrm{P}\big(\omega:\boldsymbol{\xi}_1(\omega)\in A_1\big)\cdots\mathrm{P}\big(\omega:\boldsymbol{\xi}_n(\omega)\in A_n\big).$$

则称随机向量 $\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n$ **相互独立**.

定义 1.4.11. 设 $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots$ 是一列随机向量, 如果对于任意的 n, $\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n$ 相互独立, 则称 $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots$ **相互独立**.

1.5 随机过程

1.5.1 随机过程的一种构造方式

结合我们前面对概率论公理系统的了解, 从应用随机过程的知识知道, 一个随机过程是定义在同一个概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的以 T 为指标集的随机变量族 $\{X_t,\ t\in T\}$(或记为 $\{X(t),\ t\in T\}$). T 解释为时间, 可以是 $\{0,1,2\cdots,n\}$、非负整数集、非负实数、整数或实数. 对于任意固定的 $t\in T$, X_t 的取值空间称为状态空间, 记为 S, 其中的元素称为状态.

我们想知道一个随机过程的样本空间和样本点可以是什么样子, 具体些想了解 Ω 与 T 和 S 之间关系. 我们可以把 $\{X_t,\ t\in T\}$ 写成 $\{X(t,\omega),\ t\in T,\omega\in\Omega\}$, 因此将过程看成 $T\times\Omega$ 到 S 的函数 (映射). 当 $t\in T$ 固定时, $X(t,\cdot)$ 是一个 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量; 当 $\omega\in\Omega$ 固定时, $X(\cdot,\omega)$ 是一个从 T 到 S 的函数 (映射), 称为对于 ω 的样本轨道.

例 1.5.1. 考虑一个先后掷 4 次硬币这个过程, 正面记为 1, 反面记为 0, 则整个过程的全体样本点是 $\omega = (\omega_1\omega_2\omega_3\omega_4)$, 其中 $\omega_i = 0$ 或 1, $i = 1, 2, 3, 4$. 即,

$$\Omega = \begin{pmatrix} (0000) & (0001) & (0010) & (0011) \\ (0100) & (0101) & (0110) & (0111) \\ (1000) & (1001) & (1010) & (1011) \\ (1100) & (1101) & (1110) & (1111) \end{pmatrix}.$$

那么过程实际就是 $X(t,\omega) = \omega_t$. 例如: $X\big(3, (\omega_1\omega_2\omega_3\omega_4)\big) = \omega_3$, 更具体地, $X\big(3, (0010)\big) = 1$. 因此, Ω 看成

$$\{0,1\}^4 = \{0,1\} \times \{0,1\} \times \{0,1\} \times \{0,1\} = \{\omega = (\omega_1\omega_2\omega_3\omega_4),\ \omega_i = 0 \text{ 或 } 1,\quad i = 1,2,3,4\}$$

或等价于 $\{1,2,3,4\}$ 到 $\{0,1\}$ 的映射全体.

一般地, 对于一个状态空间为 S 的随机过程 $\{X_t,\ t \in T\}$, Ω 可以取成 $S^T = \{T\text{到}S\text{的映射}\}$, 进一步利用 Kolmogorov 相容性定理, 可以构造出事件域 $\mathcal{F}$ 以及事件域上的概率测度 P, 使得 P 应该满足 $\{X_t,\ t \in T\}$ 的统计性质, 比如, 独立同分布性质等.

1.5.2 信息、事件域和带流 (滤子) 的概率空间

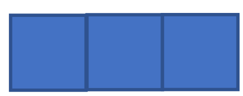
在 "随机世界" 中, 与 "确定性世界" 不同, 由于不确定性, 人们不能精确预测 "运动" 未来的 "轨迹", 但人们总是希望通过已知过去和现在的信息来预测 "未来". 如何在概率空间的框架下来精确定义这种 "信息"? 这就是事件域流. 在这里, 我们仅用下面简单的例子来说明这一思想.

考虑掷三次硬币这一随机过程, 样本空间是

$$\Omega = \{000, 001, 010, 011, 111, 110, 101, 100\} = \{0,1\}^3;$$

样本点表示成 $\omega = (\omega_1\omega_2\omega_3)$, 事件域可以取为 $\mathcal{F} = \mathscr{P}(\Omega)$.

在掷硬币之前, 也就是在试验之前, 我们只能确定所有的样本点是什么, 但并不知道哪个样本点将出现. 我们所能了解的信息仅仅是必然事件 Ω 发生, 不可能事件 $\emptyset$ 不发生, 也就是说, 这时我们能知道的事件域是 $\mathcal{F}_0 = (\Omega, \emptyset)$.



图一

当第一次掷完之后, 虽然, 试验还未完成, 我们不能预测具体某个样本点 ω 是否最终出现, 但这时已经知道 ω 的部分 "信息". 例如, 若观测到正面出现, 也就是说 $\omega_1 = 1$, 那么我们可以知道什么样的 "信息"? 推断出哪些事件发生或不发生?

图二

$$\omega \in A_1 = \{\text{第一次是正面}\} = \{111, 110, 101, 100\},\ \omega \in \Omega,$$

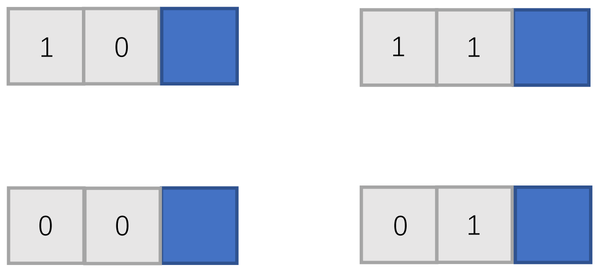
$$\omega \notin A_0 = \{\text{第一次是反面}\} = \{000, 001, 010, 011\},\ \omega \notin \emptyset;$$

若 $\omega_1=0$, 类似地,

$$\omega\notin A_1=\{\text{第一次是正面}\}=\{111,110,101,100\},\ \omega\in\Omega,$$
$$\omega\in A_0=\{\text{第一次是反面}\}=\{000,001,010,011\},\ \omega\notin\emptyset.$$

当第一次掷完之后, 对于每一个样本点 ω, 对于任意的 $A\in\mathcal{F}_1=\{\Omega,\emptyset,A_1,A_0\}$ 我们总是能判断出 $\omega\in A$ 是否成立, 也就是说能判断事件 A 是否发生. 此时, 我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}$.

当第一次, 第二次掷完之后, 若 $\omega_1=1$, $\omega_2=0$, 则我们可以推断出的信息是



图三

$$\omega\in A_{10}=\{100,101\},\ \omega\notin A_{11}=\{110,111\},\ \omega\in\Omega,$$
$$\omega\notin A_{00}=\{000,001\},\ \omega\notin A_{01}=\{010,011\},\ \omega\notin\emptyset,$$

并且可以推断出 ω 是否属于这几个事件的交、并、对立事件以及其交、并、对立事件再交、并和对立事件, 如此往复下去, 也就是说总能推断出此时 ω 是否属于 A, A 是事件域 $\mathcal{F}_2$ 中的任一事件.

$$\begin{aligned}\mathcal{F}_2 \quad=\quad &\{\Omega,\emptyset,A_1,A_0,A_{11},A_{10},A_{01},A_{00},A_{11}^c,A_{10}^c,A_{01}^c,A_{00}^c,\\ &A_{11}\cup A_{01},A_{11}\cup A_{00},A_{10}\cup A_{01},A_{01}\cup A_{00}\}.\end{aligned}$$

若 $\omega_1=1,\omega_2=1$; $\omega_1=0,\omega_2=0$ 或 $\omega_1=0,\omega_2=1$ 我们同样可以推断出它们是否属于 $\mathcal{F}_2$ 中的事件. 即, 此时对于每一个的样本点 ω, 对于任意的 $A\in\mathcal{F}_2$, 我们总能判断出 $\omega\in A$ 是否成立. 此时, 我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}$.

当三次硬币都掷完之后, 对于任意的样本点 ω, 我们总能判断出对于任意的 $A\in\mathcal{F}_3$, $\omega\in A$ 是否成立, 即事件 A 是否发生, 其中 $\mathcal{F}_3=\mathscr{P}(\Omega)$. 此时, 我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}_3\subset\mathcal{F}$.

若进一步考虑 $T=\mathbb{Z}^+,S=\{0,1\}$, 即先后掷无穷多次硬币, 则 Ω 可以取成

$$\Omega=\{0,1\}^{\mathbb{Z}^+}=\{\omega|\omega=(\omega_1\omega_2\omega_3\cdots),\omega_i=0\,\text{或}\,1,i=1,2,3,\cdots\}.$$

$\mathcal{F}_0=\{\Omega,\emptyset\}$; $\mathcal{F}_1=\{\Omega,\emptyset,A_1,A_0\}$, 其中, $A_1=\{\omega|\omega_1=1\}=\{\omega|\omega=(1\omega_2\omega_3\cdots),\ \omega_i=0\text{或}1,\ i\geq 2\}$, $A_0=\{\omega|\omega_1=0\}=\{\omega|\omega=(0\omega_2\omega_3\cdots),\ \omega_i=0,\text{或}1,\ i\geq 2\}$. 类似地, 可以定义 $\mathcal{F}_n,n\geq 2$.

设 $\mathcal{F}_\infty$ 是包含 $\cup_{n=1}^\infty\mathcal{F}_n$ 的最小事件域 (但不是 $\mathscr{P}(\Omega)$). 我们可以取 $\mathcal{F}=\mathcal{F}_\infty$. 因此, 我们发现 $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\geq 1},\mathrm{P})$ 可以比 $(\Omega,\mathcal{F},\mathrm{P})$ 更好地描述掷多次硬币这个随机试验 (随机过程).

因此研究随机过程时, 我们通常要考虑一个带流的概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$ (filtered probability space), 满足 $t_1<t_2$ 时, $\mathcal{F}_{t_1}\subset\mathcal{F}_{t_2}\subset\mathcal{F}$. $(\mathcal{F}_t)_{t\in T}$ 被称为事件域流 (σ-代数流, σ-域流) 或滤子[①,②].

[①] 在实际研究中, 有时 $\mathcal{F}$ 可以为 $\mathcal{F}_\infty$, 也可以取的比 $\mathcal{F}_\infty$ 大.

[②] 因为在学习应用随机过程时遇到的过程简单些, 用的是所谓的“自然 σ-域流”, 所以当时并不强调带流的概率空间的概念. 在研究复杂的随机现象时, 考虑带流的概率空间就非常有用了.

1.6 方差有限的随机变量空间和 Gauss 系

1.6.1 $L^2(\Omega, \mathcal{F}, \mathrm{P})$

数学期望是随机变量的一个重要数字特征, 它表示随机变量取值的平均水平, 但是仅用数学期望描述随机变量通常是不够的, 通常还要利用方差. 方差描述了随机变量关于其数学期望的离散程度. 方差越大, 说明这个随机变量越 “随机”, 可以认为是随机现象 “随机性” 的一种刻画. 因此, 在一定程度上可以用期望和方差两个数字特征来描述随机现象. 有时, 在金融和经济领域, 方差有时也被理解成对 “风险” 的一种度量. 我们知道, 当一个随机变量的二阶矩有限时, 它的数学期望和方差一定存在. 因此, 我们下面来看这样一类随机变量有什么样的数学结构.

概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上二阶矩有限, 也就是说方差有限的全体随机变量组成的空间记为 $L^2(\Omega, \mathcal{F}, \mathrm{P})$(或 $L^2(\Omega)$).

♣♣♣20230223

首先, 回忆在高等代数或线性代数中 n 维欧氏空间 (Euclid 空间) 的概念. 在本节中 n 维向量指列向量, t 表示矩阵转置. 欧氏空间有两个基本的要点:

(1) 线性空间;

(2) 内积: $\mathbf{x} = (x_1, \cdots, x_d)^{\mathrm{t}}, \mathbf{y} = (y_1, \cdots, y_d)^{\mathrm{t}}$, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i$. 由内积可以定义欧氏距离: $\|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2} = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$[①],[②].

事实上, $L^2(\Omega, \mathcal{F}, \mathrm{P})$ 也是一个欧氏空间, 也就是说:

(1) 线性空间: 若 $\xi, \eta \in L^2(\Omega, \mathcal{F}, \mathrm{P})$, 即 $\mathrm{E}\xi^2 < \infty$, $\mathrm{E}\eta^2 < \infty$, 则由 Cauchy-Schwarz 不等式, 对于任意的 $a, b \in \mathbb{R}$,

$$\begin{aligned} \mathrm{E}(a\xi + b\eta)^2 &\le a^2 \mathrm{E}\xi^2 + b^2 \mathrm{E}\eta^2 + 2|ab| \mathrm{E}(\xi\eta) \\ &\le a^2 \mathrm{E}\xi^2 + b^2 \mathrm{E}\eta^2 + 2|ab| \sqrt{\mathrm{E}\xi^2 \mathrm{E}\eta^2} \\ &< \infty. \end{aligned}$$

(2) 内积: 可以在 $L^2(\Omega, \mathcal{F}, \mathrm{P})$ 上定义内积, 对于任意的 $\xi, \eta \in L^2(\Omega, \mathcal{F}, \mathrm{P})$,

$$\langle \xi, \eta \rangle = \mathrm{E}(\xi\eta) \le \sqrt{\mathrm{E}\xi^2 \mathrm{E}\eta^2} < \infty.$$

很容易验证 $\langle \xi, \eta \rangle$ 满足内积的定义. 此时欧氏距离定义为

$$\|\xi - \eta\| = \sqrt{\langle \xi - \eta, \xi - \eta \rangle} = \sqrt{\mathrm{E}(\xi - \eta)^2}.$$

其实, 这就是概率论中的两个随机变量的均方距离.

在 $L^2(\Omega, \mathcal{F}, \mathrm{P})$, 若 $\mathrm{P}\big(\omega : \xi(\omega) = \eta(\omega)\big) = 1$, 则 ξ 和 η 视为同一元素.

在概率论中 $\mathrm{cov}(\xi, \eta)$ 是对随机变量 ξ 与 η 之间 “线性相关程度” 的一种 “度量”, 为什么可以这么说?

在有限维欧氏空间中,$\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$ 表示两个向量 $\mathbf{x}$ 和 $\mathbf{y}$ 之间夹角的余弦, 可以理解为它们之间线性相关性的一种度量. 而相关系数 $\rho_{\xi\eta} = \dfrac{\mathrm{cov}(\xi, \eta)}{\sqrt{\mathrm{var}\xi}\sqrt{\mathrm{var}\eta}}$, $|\rho_{\xi\eta}| \le 1$, 事实上就是两个向量之间夹角余弦这一概念在 $L^2(\Omega, \mathcal{F}, \mathrm{P})$ 中的推广. 因此, 可以用协方差 $\mathrm{cov}(\xi, \eta)$ 或相关系数 $\rho_{\xi\eta}$ 来表示两个随机变量之间 “线性相关” 的程度.

$L^2(\Omega, \mathcal{F}, \mathrm{P})$ 可能是一个无穷维的欧氏空间. 也就是说可能在 $L^2(\Omega, \mathcal{F}, \mathrm{P})$ 中找不到有限个元, 使之构成一组基将其中的元素线性表示出来.

[①]欧氏距离的定义本质上来源于勾股定理, 内积和欧氏距离的关系是通过平行四边形法则相联系.

[②]一般的实线性空间中抽象的内积定义是:$\langle \mathbf{x}, \mathbf{y} \rangle$ 是对称双线性型, 且满足 (1) $\langle \mathbf{x}, \mathbf{x} \rangle \ge 0$; (2) $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ 当且仅当 $\mathbf{x} = 0$.

定义 1.6.1. 在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中随机变量序列 $\{\xi_n\}$ 称为收敛到 $\xi\in L^2(\Omega,\mathcal{F},\mathrm{P})$, 如果当 $n\to\infty$ 时, 有 $\|\xi_n-\xi\|\to 0$, 即 $\mathrm{E}(\xi_n-\xi)^2\to 0$, 也就是随机变量序列的均方收敛.

定义 1.6.2. 在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的随机变量序列 $\{\xi_n\}$ 称为 Cauchy 序列, 如果当 $n,m\to\infty$ 时, 有 $\|\xi_n-\xi_m\|\to 0$.

性质 1.6.1. $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的任意 Cauchy 序列必有极限, 即如果 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的随机变量序列满足 $n,m\to\infty$ 时, 恒有 $\|\xi_n-\xi_m\|\to 0$, 则一定存在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的元素 ξ, 使得 $\|\xi_n-\xi\|\to 0$. 因此, $L^2(\Omega,\mathcal{F},\mathrm{P})$ 是完备的, 即 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 是一个 Hilbert 空间.

1.6.2 Gauss 分布、Gauss 系和 Gauss 过程

在方差有限的随机变量 (随机向量) 类中, 有一类特殊的随机变量 (随机向量), 这就是服从 Gauss 分布的随机变量 (向量).

定义 1.6.3. 设 $\eta_1,\cdots,\eta_m$ 为 m 个相互独立服从标准正态分布的随机变量, $(a_{ij})_{1\le i\le d,1\le j\le m}$, μ_i, $1\le i\le d$ 为常数, 设

$$\begin{cases}\xi_1 &= a_{11}\eta_1+\cdots+a_{1m}\eta_m+\mu_1\\ \cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\\ \xi_d &= a_{d1}\eta_1+\cdots+a_{dm}\eta_m+\mu_d\end{cases},$$

则称 d 维随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_d)^{\mathrm{t}}$ 服从 d 维 Gauss 分布, 其矩阵形式为

$$\begin{pmatrix}\xi_1\\ \vdots\\ \xi_d\end{pmatrix}=\begin{pmatrix}a_{11}&\cdots&a_{1m}\\ \vdots&\ddots&\vdots\\ a_{d1}&\cdots&a_{dm}\end{pmatrix}\begin{pmatrix}\eta_1\\ \vdots\\ \eta_m\end{pmatrix}+\begin{pmatrix}\mu_1\\ \vdots\\ \mu_d\end{pmatrix}.$$

即

$$\boldsymbol{\xi}=A\boldsymbol{\eta}+\mu,\ A=(a_{ij})_{1\le i\le d,1\le j\le m},\ \boldsymbol{\eta}=(\eta_1,\cdots,\eta_m)^{\mathrm{t}},\ \boldsymbol{\mu}=(\mu_1,\cdots,\mu_d)^{\mathrm{t}}.$$

$\boldsymbol{\xi}$ 有如下的一些简单性质:

(1) $\mathrm{E}\boldsymbol{\xi}=\boldsymbol{\mu}$, 协方差矩阵 $\boldsymbol{\Sigma}=AA^{\mathrm{t}}$.

(2) $\boldsymbol{\xi}$ 的特征函数 $\phi(\boldsymbol{\theta})=\exp\big(\mathrm{i}\langle\boldsymbol{\theta},\boldsymbol{\mu}\rangle-\frac{1}{2}\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{\Sigma}\boldsymbol{\theta}\big)$.

(3) $\boldsymbol{\xi}$ 的矩母函数 $\mathrm{M}(\mathbf{s})=\exp\big(\langle\mathbf{s},\boldsymbol{\mu}\rangle+\frac{1}{2}\mathbf{s}^{\mathrm{t}}\boldsymbol{\Sigma}\mathbf{s}\big)$.

(4) 若 $\boldsymbol{\Sigma}$ 可逆, 称 $\boldsymbol{\xi}$ 的分布称为 d 维正态分布 (normal distribution), 其密度函数

$$p(\mathbf{x})=\frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{t}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big),$$

其中, $|\boldsymbol{\Sigma}|$ 表示 $\boldsymbol{\Sigma}$ 的行列式, 记 $\boldsymbol{\xi}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

(5) 若 $\mathrm{P}\big(\omega:\boldsymbol{\xi}(\omega)=\boldsymbol{\mu}\big)=1$, $\boldsymbol{\mu}$ 是一个常向量, 则 $\boldsymbol{\xi}$ 是一个协方差阵为 0 的 Gauss 随机向量.

性质 1.6.2. (1) $\boldsymbol{\xi}$ 服从 d 维 Gauss 分布, 则对于任意的 $r<d$, 其任意的 r 维边际分布还是 Gauss 分布. 反之不然.

(2) $\boldsymbol{\xi}$ 服从 d 维正态分布, 则对于任意的 $r<d$, 其任意的 r 维边际分布还是正态分布. 反之不然.

例 1.6.1. 设 $\phi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$, $x\in\mathbb{R}$ 是标准正态分布的密度函数, $g(x)=\begin{cases}\cos x & |x|<\pi\\ 0 & |x|\ge\pi\end{cases}$. 设二元随机向量 (ξ,η) 的密度函数为

$$p(x,y)=\phi(x)\phi(y)+\frac{1}{2\pi}\mathrm{e}^{-\pi^2}g(x)g(y),\quad (x,y)\in\mathbb{R}^2.$$

(ξ,η) 不是 Gauss 随机向量, 但 ξ 服从标准正态分布, η 亦然.

性质 1.6.3. 下列命题等价.

(1) 随机向量 $\xi=(\xi_1,\cdots,\xi_d)^{\mathrm{t}}$ 服从 d 维 Gauss 分布.

(2) 对于任意 $a_1,\cdots,a_d$, 线性组合 $\sum_{k=1}^{d}a_k\xi_k$ 服从一维 Gauss 分布.

这个性质说明 Gauss 分布的一个重要性质: **关于线性运算封闭**, 也就是说如果 $(\xi_1,\cdots,\xi_d)^{\mathrm{t}}$ 服从多元 Gauss 分布, 则 $\xi_1,\cdots,\xi_d$ 的任意线性组合 $\sum_{k=1}^{d}a_k\xi_k$ 服从一维 Gauss 分布. 另外, 服从 Gauss 分布的随机变量还有一个重要性质: **关于均方收敛意义下封闭**, 也就是说:

性质 1.6.4. 设 $\xi_n, n=1,2\cdots$ 是一列服从 Gauss 分布的随机变量, 若存在 ξ 使得 $\lim_{n\to\infty}\mathrm{E}(\xi_n-\xi)^2=0$, 则 ξ 也服从 Gauss 分布.

定义 1.6.4. 随机变量族 $\{\xi_t, t\in I\}$ 称为 Gauss 系, 如果对于任意 n 以及任意 $t_1,t_2,\cdots,t_n\in I$, 随机向量 $(\xi_{t_1},\cdots,\xi_{t_n})^{\mathrm{t}}$ 服从 Gauss 分布. 如果 Gauss 系中的指标集 I 取成时间参数, 则之称为 Gauss 过程.

第二章　条件数学期望

2.1　条件数学期望的直观推导

我们先看两个例子.

例 2.1.1. 在一个简单博弈中, 某人根据前面各次博弈的结果决定下一次博弈的金额. 设在开始时, 有本金ξ_0, 把第 n 次博弈后某人所有的本金记为 ξ_n. 将这个简单博弈用一个数学模型表示, 需要有以下两个要素: (1) 设

$$\eta_n = \begin{cases} 1 & 某人赢得第n次博弈; \\ -1 & 某人输掉第n次博弈. \end{cases}$$

在此假定 $\{\eta_n,\ n = 1, 2, 3, \cdots\}$ 是独立同分布的随机变量序列; (2) 所谓"根据前面各次博弈的结果决定下一次博弈的金额"是指在第 n 次下注时, 根据前面 $n-1$ 次的输赢决定赌注的多少, 也就是第 n 次的赌注是前面 $n-1$ 次输赢 (或本金) 的函数:

$$b_n(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})\ \ (或者 = \tilde{b}_n(\xi_0, \xi_1, \xi_2, \cdots, \xi_{n-1}))$$

设 b_n(或 $\widetilde{b}_n$) 是 Borel 函数, 那么, 第 n 次博弈后某人所有的本金为

$$\begin{aligned} \xi_n &= \xi_{n-1} + b_n(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1}) \cdot \eta_n \\ &= \xi_0 + \sum_{k=1}^{n} b_k(\xi_0, \eta_1, \eta_2, \cdots, \eta_{k-1}) \cdot \eta_k\ . \end{aligned} \tag{2.1.1}$$

在本书中, 我们将上述模型称为基本模型.

现在设想, 有张三和李四两人各自给出了博弈"策略":

$$\{b_n^{张三}, n = 1, 2, \cdots\},\ \ \{b_n^{李四}, n = 1, 2, \cdots\}.$$

问题是若让我们选, 选谁的博弈"策略"? 当然要选能够让我们赢更多的那个"策略". 但由于"随机性", 我们并**不知道在 n 时刻**采用博弈"策略"$b_n(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})$之后, 能赢多少或输多少. 而我们更实际的做法是针对基本模型 (2.1.1), 根据前面 $n-1$ 次的输赢构造一个"预测方法 (函数)"$f(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})$ 来表示在 n 时刻我们对本金 ξ_n 的预测. 如果

$$f(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1}) - \xi_{n-1} > 0,$$

表示我们"预计"能赢钱; 针对张三和李四的策略分别构造一个对 ξ_n 的"预测方法 (函数)", 如果

$$f^{张三}(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1}) - \xi_{n-1} > f^{李四}(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1}) - \xi_{n-1}, \tag{2.1.2}$$

表示张三此时的博弈"策略"比李四的好, 我们可以选张三此时的博弈"策略"来下注.

但问题又出现了, 每个人都可以给一个"预测方法 (函数)"$f(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})$, 谁说了算呢? 当然是要找一个"最佳预测方法 (函数)", 在这个"最佳预测方法 (函数)"下, 若 (2.1.2) 还满足的话, 我们就可以大胆放心地用张三此时的博弈"策略"来下注了.

但问题再一次出现, 什么是"最佳预测方法 (函数)"? 我们知道预测总是有误差的,"最佳预测方法 (函数)"应该是"误差最小"的.

更为关键的是, 应该选取一个合理、自然的误差评判标准! 写成数学的语言就是, 要找到某种"度量"或"距离" $d(\cdot,\cdot)$, 使得"最佳预测方法 (函数)"$f(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})$ 应该满足, 对于任意一个"预测方法

(函数)” $g(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$,

$$d\big(\xi_n, f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})\big) \le d\big(\xi_n, g(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})\big). \tag{2.1.3}$$

因此, 如果我们能有一个这样的误差标准 $d(\cdot,\cdot)$, 使得 $f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$ 满足 (2.1.3), 而且对于这个误差标准 $d(\cdot,\cdot)$ 下的“最佳预测方法 (函数)”f, 还能得到不等式 (2.1.2), 那么我们就可以选用张三的博弈“策略”来下注了.

我们再来看一个概率论的例子.

例 2.1.2. 设随机向量 (X,Y) 表示人的身高和体重. 我们经常遇到这样的问题, 告诉一个人的身高, 要判断他 (或她) 的体重. 即需要给出一个关于 X 的函数 $f(X)$, 当然 $f(X)$ 与 Y 之间有误差. 同样的问题是我们需要找到一个误差的评判标准, 也就是说一种“度量”或“距离”$d(\cdot,\cdot)$, 使得对于任意的 Borel 函数 g

$$d(f(X),Y) \le d(g(X),Y) \tag{2.1.4}$$

成立.

我们先来看一些具有启发性的类比. 直线上两点 x 和 y 之间自然的距离 $d(x,y)=|x-y|$; 由勾股定理, 平面 $\mathbb{R}^2$ 上两点 $(x_1,x_2)^{\mathrm{t}}$, $(y_1,y_2)^{\mathrm{t}}$ 之间自然距离是 $d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$; n-维欧氏空间的 $\mathbb{R}^n$ 上两点 $(x_1,\cdots,x_n)^{\mathrm{t}}$, $(y_1,\cdots,y_n)^{\mathrm{t}}$ 之间自然距离是

$$d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}.$$

相当于每个坐标之间的距离平方“加权”相加再开方. 此时, 它们的每个坐标的“权重”相等.

因此, 类比地对于两个随机变量 X,Y, 若 $\mathrm{E}X^2<\infty$, $\mathrm{E}Y^2<\infty$, 则 X 和 Y 之间可以定义一种距离

$$d(X,Y)=\sqrt{\mathrm{E}\big(X(\omega)-Y(\omega)\big)^2}.$$

可以理解为每个“坐标 ω”之间距离平方“加权平均”再开方, 实际上, 这就是 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的两个随机变量的距离. 这样, 在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中, 我们实际是确定了一种误差估计方法, 即**均方误差**.

再回到上面的例 (2.1.1) 和例 (2.1.2). 对于例 (2.1.1) 中的不等式 (2.1.3) 和例 (2.1.2) 中的不等式 (2.1.4) 所表述的就是统计学中通常使用的“均方误差最小准则”. 因此, 在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中, 对上面的例 (2.1.1), 我们希望找到“预测函数”$f(\xi_0,\eta_1,\cdots,\eta_{n-1})$, 对例 (2.1.2), 希望找到“估计函数”$f(X)$, 使得它们在“均方误差最小准则”下达到最小.

综合上面的分析, 实质上, 我们希望做的是对于满足 $\mathrm{E}X^2<\infty$ 的随机变量 X 和随机向量 $(Y_1,Y_2,\cdots,Y_n)$, 找到一个关于 $(Y_1,Y_2,\cdots,Y_n)$ 的函数, 使得 $\mathrm{E}f^2(Y_1,Y_2,\cdots,Y_n)<\infty$, 且

$$\mathrm{E}\big(X-f(Y_1,Y_2,\cdots,Y_n)\big)^2 \le \mathrm{E}\big(X-g(Y_1,Y_2,\cdots,Y_n)\big)^2, \tag{2.1.5}$$

其中 g 为任意关于 $(Y_1,Y_2,\cdots,Y_n)$ 的函数, 且使得 $\mathrm{E}g^2(Y_1,Y_2,\cdots,Y_n)<\infty$, 或者表述为

$$\mathrm{E}\big(X-f(Y_1,Y_2,\cdots,Y_n)\big)^2=\inf_{\mathrm{E}g^2(Y_1,Y_2,\cdots,Y_n)<\infty}\mathrm{E}\big(X-g(Y_1,Y_2,\cdots.Y_n)\big)^2, \tag{2.1.6}$$

即 $f(Y_1,Y_2,\cdots,Y_n)$ 是 X 关于 $(Y_1,Y_2,\cdots,Y_n)$ 在**均方误差最小准则下的最佳逼近**.

这里需要明确一下 “g 为关于 $(Y_1,Y_2,\cdots,Y_n)$ 的函数” 的含义是 $g(x_1,x_2,\cdots,x_n)$ 为 $\mathbb{R}^n$ 上的 Borel 函数, $g(Y_1,\cdots,Y_n)$ 是 g 与随机向量 $(Y_1,Y_2,\cdots,Y_n)$ 的复合函数 $g\circ(Y_1,\cdots,Y_n)(\omega)$, 它是一个随机变量, 在本书中我们也称其为由 $(Y_1,\cdots,Y_n)$ 决定的随机变量.

更一般地, 对于随机变量 X 满足 $\mathrm{E}X^2<\infty$ 和一族随机变量 $(Y_t,t\in I)$(I 是一个指标集) 希望找到一个关于 $(Y_t,t\in I)$ 的“函数”$f(Y_t,t\in I)$, 满足 $\mathrm{E}f^2(Y_t,t\in I)<\infty$ 使得对于任意的关于 $(Y_t,t\in I)$ 的“函数”$g(Y_t,t\in I)$,

$$\mathrm{E}\big(X-f(Y_t,t\in I)\big)^2=\inf_{\mathrm{E}g^2(Y_t,t\in I)<\infty}\mathrm{E}\big(X-g(Y_t,t\in I)\big)^2. \tag{2.1.7}$$

但问题是“$g(Y_t,t\in I)$”该如何定义? 数学上该如何表达? 我们将在本章第三节中讲述这个问题.

考虑 $g(Y_1,\cdots,Y_n)$ 且满足 $\mathrm{E}g^2(Y_1,\cdots,Y_n)<\infty$ 的随机变量全体, 记为 $L^2(\Omega,\sigma(Y_1,\cdots,Y_n),\mathrm{P})$. 那么,

1) $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})\subset L^2(\Omega,\mathcal{F},\mathrm{P})$;

2) $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 是一个线性空间, 即
$$f_i(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P}),\quad i=1,2,$$
则 $\forall\alpha,\beta\in\mathbb{R}$
$$\alpha f_1(Y_1,Y_2,\cdots,Y_n)+\beta f_2(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P});$$

3) $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 可以从 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 继承下来内积结构;

4) 可以证明 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 是一个闭集, 即 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 是 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 的闭子空间, 即为一个 Hilbert 空间.

因此, 例 (2.1.1) 和例 (2.1.2) 中的问题实质从几何的观点来看, 就是在 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 这个 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 的子空间中找一个元素 $f(Y_1,Y_2,\cdots,Y_n)$, 使得 X 到 $f(Y_1,Y_2,\cdots,Y_n)$ 的距离是 X 到 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ 中距离最短.

回到有限维欧氏空间 (以 3 维空间为例), 在 (Y,Z) 平面找一个向量 $\vec{K}$ 使得 $\vec{X}$ 到 $\vec{K}$ 的距离最短, 那么 $\vec{K}$ 应该是什么? 我们知道 $\vec{K}$ 应该是 $\vec{X}$ 在平面 (Y,Z) 上的投影.

作为向量来看一个等价说法就是: $\vec{K}$ 应该满足 $\vec{X}-\vec{K}$ 与 (Y,Z) 平面垂直, 即 $\vec{X}-\vec{K}\perp(Y,Z)$ 平面. 在有限维欧氏空间, $\vec{X}-\vec{K}\perp(Y,Z)$ 平面的数学表达为
$$\langle\vec{X}-\vec{K},\vec{L}\rangle=0,$$
其中 $\vec{L}$ 是 (Y,Z) 平面上的任意向量.

类比于有限维欧氏空间理论, 对于 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 和 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$, 应该有
$$X-f(Y_1,Y_2,\cdots,Y_n)$$
垂直于 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$, 也就是说对于任意的 $g(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$,
$$\mathrm{E}\big((X-f(Y_1,Y_2,\cdots,Y_n))g(Y_1,Y_2,\cdots,Y_n)\big)=0,$$
也就是说 $f(Y_1,Y_2,\cdots,Y_n)$ 应该满足如下的方程,
$$\mathrm{E}\big(Xg(Y_1,Y_2,\cdots,Y_n)\big)=\mathrm{E}\big(f(Y_1,Y_2,\cdots,Y_n)g(Y_1,Y_2,\cdots,Y_n)\big).$$

♣♣♣20230302

由此, 我们可以定义

定义 2.1.1. $X,Y_1,\cdots,Y_n$ 为概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量且满足 $\mathrm{E}X^2<\infty$, 称由 $Y_1,\cdots,Y_n$ 决定的随机变量 $f(Y_1,\cdots,Y_n)$ 为 X 关于 $Y_1,\cdots,Y_n$ 的**条件数学期望**, 若 $f(Y_1,\cdots,Y_n)$ 满足

1) $\mathrm{E}f^2(Y_1,\cdots,Y_n)<\infty$;

2) 任意的 $g(Y_1,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),\mathrm{P})$ (即 g 满足 $\mathrm{E}g^2(Y_1,\cdots,Y_n)<\infty$), 下式成立
$$\mathrm{E}\big(Xg(Y_1,\cdots,Y_n)\big)=\mathrm{E}\big(f(Y_1,\cdots,Y_n)g(Y_1,\cdots,Y_n)\big).\tag{2.1.8}$$
把 $f(Y_1,\cdots,Y_n)$ 记为 $\mathrm{E}[X|Y_1,\cdots,Y_n]$.

再看例 (2.1.1), 实质上我们是用第 n 次博弈后的本金 ξ_n 关于 $(\xi_0,\eta_1,\cdots,\eta_{n-1})$ 的条件数学期望 $f(\xi_0,\eta_1,\cdots,\eta_{n-1})(=\mathrm{E}[\xi_n|\xi_0,\eta_1,\cdots,\eta_{n-1}])$ 来预测 ξ_n 的. 对于例 (2.1.2), 实质是用体重 Y 关于身高 X 的条件数学期望 $f(X)(=\mathrm{E}[Y|X])$ 来估计 Y 的.

我们再来看一个简单的例子.

例 2.1.3. 设 (X,Y) 为离散型随机向量, 其分布列为 $\mathrm{P}(X=x_i,Y=y_j)=p_{ij}$, $i=1,\cdots,n;j=1,\cdots,m$, 且假定 $p_{ij}>0$, 求 $\mathrm{E}[X|Y]$, 即求 $f(Y)$, 使得对于任意的 $g(Y)$, 满足
$$\begin{aligned}\mathrm{E}\big(\mathrm{E}[X|Y]g(Y)\big)&=\mathrm{E}\big(Xg(Y)\big),\\ \text{即}\ \mathrm{E}\big(f(Y)g(Y)\big)&=\mathrm{E}\big(Xg(Y)\big).\end{aligned}$$

解: 首先

$$
\begin{aligned}
f(Y(\omega)) &= \begin{cases} f(y_1), & Y(\omega)=y_1 \\ \vdots & \vdots \\ f(y_m), & Y(\omega)=y_m \end{cases} \\
&= \sum_{i=1}^{m} f(y_i)\mathbf{1}_{\{y_i\}}(Y(\omega)).
\end{aligned}
$$

Y 的边际分布为 $\mathrm{P}(Y=y_j)=\sum\limits_{i=1}^{n} p_{ij}$, 则

$$
\begin{aligned}
&\mathrm{E}\big(f(Y)g(Y)\big)\ \big(\text{或}\,\mathrm{E}\big(\mathrm{E}[X|Y]g(Y)\big)\big) \\
=\ & \sum_{j=1}^{m} f(y_j)g(y_j)\sum_{i=1}^{n} p_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} f(y_j)g(y_j)p_{ij} = \sum_{i,j} f(y_j)g(y_j)p_{ij},
\end{aligned}
$$

而

$$
\mathrm{E}(Xg(Y)) = \sum_{i,j} x_i g(y_j) p_{ij}.
$$

特别地, 取

$$
\begin{aligned}
g_k(Y) &= \begin{cases} 1, & Y(\omega)=y_k \\ 0, & Y(\omega)\neq y_k \end{cases} \\
&= \mathbf{1}_{\{y_k\}}(Y(\omega)).
\end{aligned}
$$

代入上面的式子得

$$
\mathrm{E}(f(Y)g_k(Y)) = \sum_i f(y_k)p_{ik} = f(y_k)\sum_i p_{ik},
$$

$$
\mathrm{E}(Xg_k(Y)) = \sum_i x_i p_{ik}.
$$

因此, 解得

$$
f(y_k) = \frac{\sum\limits_i x_i p_{ik}}{\sum\limits_i p_{ik}} = \mathrm{E}(X|Y=y_k).
$$

也就是说, $f(y)$ 在 y_k 处的值是已知 $Y=y_k$ 时, X 的条件期望. 进一步, 得到

$$
\begin{aligned}
f(Y(\omega)) = \mathrm{E}[X|Y](\omega) &= \begin{cases} \frac{\sum_i x_i p_{i1}}{\sum_i p_{i1}}, & Y(\omega)=y_1 \\ \vdots & \vdots \\ \frac{\sum_i x_i p_{im}}{\sum_i p_{im}}, & Y(\omega)=y_m \end{cases} \\
&= \begin{cases} \mathrm{E}(X|Y=y_1), & Y(\omega)=y_1 \\ \vdots & \vdots \\ \mathrm{E}(X|Y=y_m), & Y(\omega)=y_m \end{cases} \\
&= \sum_{j=1}^{m} \mathrm{E}(X|Y=y_j)\mathbf{1}_{\{y_i\}}(Y(\omega)).
\end{aligned}
$$

□

例 2.1.4. 设 (X,Y) 服从密度函数为 $p(x,y)$ 的连续型随机向量, 且假定 $p(x,y)>0$, $(x,y)\in\mathbb{R}^2$, $\mathrm{E}X^2<\infty$, 求 $\mathrm{E}[X|Y]$.

解: 事实上, 我们需要求一个 Borel 可测函数 f, 使得 $\mathrm{E}[X|Y]=f(Y)$, 满足对于任意的 $g\in L^2(\Omega,\sigma(Y),\mathrm{P})$,

$$
\mathrm{E}\big(f(Y)g(Y)\big) = \mathrm{E}\big(Xg(Y)\big).
$$

Y 的边际密度为 $p_Y(y)=\int_{\mathbb{R}}p(x,y)\mathrm{d}x$, 且假定对任意的 y, $p_Y(y)>0$,

$$\begin{aligned}\mathrm{E}\big(f(Y)g(Y)\big) &= \int_{\mathbb{R}} g(y)f(y)p_Y(y)\mathrm{d}y,\\ \mathrm{E}\big(Xg(Y)\big) &= \int_{\mathbb{R}}\int_{\mathbb{R}} xg(y)p(x,y)\mathrm{d}x\mathrm{d}y\\ &= \int_{\mathbb{R}} g(y)\left(\int_{\mathbb{R}} xp(x,y)\mathrm{d}x\right)\mathrm{d}y.\end{aligned}$$

因为 $g(y)$ 的任意性, 由实变函数知识得

$$f(y)p_Y(y)=\int_{\mathbb{R}} xp(x,y)\mathrm{d}x,$$

即

$$f(y)=\frac{\int_{\mathbb{R}} xp(x,y)\mathrm{d}x}{p_Y(y)},$$

也就是说 $f(y)$ 为已知 $Y=y$ 时, X 的条件期望, 即为 $\mathrm{E}(X|Y=y)$.

有时将 $\mathrm{E}[X|Y]$ 记为 $\mathrm{E}(X|Y=y)|_{y=Y}$, $\mathrm{E}[X|Y]=\mathrm{E}(X|Y=y)|_{y=Y}$. □

例 2.1.5. (X,Y) 服从二元正态分布, 其分布密度如下

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\cdot\left[\frac{(x-a_1)^2}{\sigma_1^2}-\frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\right]\right\}$$

a_1, a_2, σ_1, σ_2, r 为常数, $\sigma_1>0$, $\sigma_2>0$, $|r|<1$,

$(X,Y)\sim N(\vec{a},\ \Sigma)$, $\vec{a}=\begin{pmatrix}a_1\\a_2\end{pmatrix}$, $\Sigma=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$. 求 $\mathrm{E}[X|Y]$ 的密度函数.

解: 先求出 $\mathrm{E}(X|Y=y)$, 即函数 $f(y)$, 再求 $f(Y)$ 的密度函数. 第一步求出已知 $Y=y$ 的条件下, X 的条件密度函数 $p_{X|Y}$.

$$p_{X|Y}(x|y)=\frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-r^2)}\cdot\left[x-\left(a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2)\right)\right]^2\right\}.$$

这是正态分布 $N\left(a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2),\ \sigma_1^2(1-r^2)\right)$ 的密度函数. 因此

$$\mathrm{E}(X|Y=y)=a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2),$$

即

$$\mathrm{E}[X|Y]=a_1+r\frac{\sigma_1}{\sigma_2}(Y-a_2). \tag{2.1.9}$$

这是 Y 的一个线性函数, 则 $\mathrm{E}[X|Y]$ 还是一个正态分布, 其期望和方差分别是 a_1 和 $r^2\sigma_1^2$, 其密度函数为 $\frac{1}{\sqrt{2\pi}|r|\sigma_1}\mathrm{e}^{-\frac{(x-a_1)^2}{2r^2\sigma_1^2}}$.

□

根据定义条件数学期望的思想, 我们还可以思考这样的问题.

一. 在条件数学期望定义中, 我们是将取 X 在 $L^2(\Omega,\sigma(Y_1,\cdots,Y_2),\mathrm{P})$ 的投影作为数学期望的定义的. 如果 $\mathrm{E}Y_i^2<\infty, i=1,\cdots,n$, 设

$$\overline{L}(Y_1,\cdots,Y_n)=\{b_0+b_1Y_1+\cdots+b_nY_n,\ \ (b_0,b_1,\cdots,b_n)^{\mathrm{t}}\in\mathbb{R}^{n+1}\},$$

则 $\overline{L}(Y_1,\cdots,Y_n)$ 也是 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 的一个闭的线性子空间. 我们考虑在这个空间中找一个线性函数 $a_0+a_1Y_1+\cdots+a_nY_n$, 使得

$$\mathrm{E}[X-(a_0+a_1Y_1+\cdots+a_nY_n)]^2=\inf_{(b_0,b_1,\cdots,b_n)^{\mathrm{t}}\in\mathbb{R}^{n+1}}\mathrm{E}[X-(b_0+b_1Y_1+\cdots+b_nY_n)]^2.$$

$a_0+a_1Y_1+\cdots+a_nY_n$ 被称为 X 的最佳线性估计. 一般的, 最佳线性估计 (预测), 即 X 关于 $(Y_1,\cdots,Y_n)$ 的线性回归与 $\mathrm{E}[X|Y_1,\cdots,Y_n]$ 不同, 从上面的例子 (2.1.5) 可以看出, 在 Gauss 分布场合, 二者重合. 这是 Gauss 分布一个重要且有趣的性质.

性质 2.1.1. 设 $(X, Y_1, \cdots, Y_n)$ 服从联合 Gauss 分布. 则 $\mathrm{E}[X|Y_1, \cdots, Y_n]$ 与 X 关于 $(Y_1, \cdots, Y_n)$ 的最佳线性估计一致. (表述待完善, 证明待补充.)

利用上面的性质, 不但可以得到例子 (2.1.5) 一个简单的解法, 而且可以给出 $r^2\sigma_1^2$ 概率意义和几何意义. 已知 $\mathrm{E}[X|Y] = a + bY$, 它还是 Gauss 随机变量. 故而只需求出它的期望和方差. 将其中的随机变量标准化, 即

$$X = \mathrm{E}X + \sigma_1 \frac{X - \mathrm{E}X}{\sigma_1},\ Y = \mathrm{E}Y + \sigma_2 \frac{Y - \mathrm{E}Y}{\sigma_2}.$$

注意到, 在 $L^2(\Omega, \mathcal{F}, \mathrm{P})$ 中, 常值随机变量构成的线性空间与零均值随机变量构成的线性空间是正交的. 因此, (1) X 在常值随机变量构成的线性空间上的投影就是 $\mathrm{E}X$; (2) X 在线性空间 $\{c(Y - \mathrm{E}Y), c \in \mathbb{R}\}$ 的投影就是 $X - \mathrm{E}X$ 在 $\{c(Y - \mathrm{E}Y), c \in \mathbb{R}\}$ 上的投影. 相关系数 r 看成 $X - \mathrm{E}X$ 和 $Y - \mathrm{E}Y$"夹角"的余弦, 故 $X - \mathrm{E}X$ 在 $Y - \mathrm{E}Y$ 上投影后得到的随机变量的标准差为 $|r|\sigma_1$, 其表示成 $r\sigma_1 \frac{Y - \mathrm{E}Y}{\sigma_2}$. 因此, $\mathrm{E}[X|Y] = \mathrm{E}X + r\frac{\sigma_1}{\sigma_2}(Y - \mathrm{E}Y)$, 这就是 (2.1.9) 中得到的结论, 它服从 $N(a_1, r^2\sigma_1^2)$.

二. 我们是对 $\mathrm{E}X^2 < \infty$ 的定义了条件数学期望, 我们能否将这一定义推广到 $\mathrm{E}|X| < \infty$ 的随机变量类

$$L^2(\Omega, \mathcal{F}, \mathrm{P}) \subset L^1(\Omega, \mathcal{F}, \mathrm{P}) = \{X\text{为}(\Omega, \mathcal{F}, \mathrm{P})\text{上的随机变量, 且}\mathrm{E}|X| < \infty\}.$$

虽然 $L^1(\Omega, \mathcal{F}, \mathrm{P})$ 也是一个线性空间, 但在 $L^1(\Omega, \mathcal{F}, \mathrm{P})$ 中不能定义内积结构, 而我们知道投影本质上依赖于内积的. 在 $L^2(\Omega, \sigma(Y_1, \cdots, Y_n), \mathrm{P})$ 中, 注意到 (2.1.8) 式中要求 $\mathrm{E}X^2 < \infty, \mathrm{E}f^2(Y_1, \cdots, Y_n) < \infty, \mathrm{E}g^2(Y_1, \cdots, Y_n) < \infty$. 如果我们只考虑 $\mathrm{E}|X| < \infty, \mathrm{E}|f(Y_1, \cdots, Y_n)| < \infty$ 时, 那么当 $g(Y_1, \cdots, Y_n)$ 为有界随机变量时, (2.1.8) 式两端仍有意义. 进一步, 在此条件下, 由测度论中的 Radom-Nikodym 定理知道存在唯一的 $f(Y_1, \cdots, Y_n)$ 满足 (2.1.8). 此时我们认为 $f(Y_1, \cdots, Y_n)$ 就是 X 关于 $Y_1, \cdots, Y_n$ 的"最佳的预测函数", 即 $f(Y_1, \cdots, Y_n)$ 为 X 的条件数学期望①.

2.2　条件数学期望的定义和性质 (I)

2.2.1　定义与例子

定义 2.2.1. (接定义2.1.1) 设 $X, Y_1, \cdots, Y_n$ 为概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量, 满足 $\mathrm{E}|X| < \infty$, f 为 $\mathbb{R}^n \to \mathbb{R}$ 的 Borel 函数, 称 $f(Y_1, \cdots, Y_n)$ 为 X 关于 $(Y_1, \cdots, Y_n)$ 的**条件数学期望**, 若

1) $\mathrm{E}|f(Y_1, \cdots, Y_n)| < \infty$;

2) 对于任意的 n 维有界 Borel 函数 g, 下式成立

$$\mathrm{E}\big(Xg(Y_1, \cdots, Y_n)\big) = \mathrm{E}\big(f(Y_1, \cdots, Y_n)g(Y_1, \cdots, Y_n)\big). \tag{2.2.1}$$

把 $f(Y_1, \cdots, Y_n)$ 记为 $\mathrm{E}[X|Y_1, \cdots, Y_n]$.

上式 (2.2.1) 等价于对于任意的 n 维 Borel 集 A,

$$\mathrm{E}\big(X\mathbf{1}_A(Y_1, \cdots, Y_n)\big) = \mathrm{E}\big(f(Y_1, \cdots, Y_n)\mathbf{1}_A(Y_1, \cdots, Y_n)\big). \tag{2.2.2}$$

对于以上的定义, 很自然我们面临两个问题:

一. 若 $\mathrm{E}X^2 < \infty$, 则 $\mathrm{E}|X| < \infty$, 那么定义2.1.1中的 $f(Y_1, \cdots, Y_n)$ 与定义2.2.1中的 $f(Y_1, \cdots, Y_n)$ 是否一致? 答案是它们是一致的. 我们将在性质2.2.7中给出解答.

二. 类似于 (2.1.6), 也可以在 $L^1(\Omega, \mathcal{F}, \mathrm{P})$ 距离下找 X 的关于 $(Y_1, \cdots, Y_n)$ 的最佳逼近 $\widetilde{f}(Y_1, \cdots, Y_n)$, 也就是说

$$\mathrm{E}\big|X - \widetilde{f}(Y_1, Y_2, \cdots, Y_n)\big| = \inf_{\mathrm{E}|g(Y_1, Y_2, \cdots, Y_n)| < \infty} \mathrm{E}\big|X - g(Y_1, Y_2, \cdots .Y_n)\big|, \tag{2.2.3}$$

① $f(Y_1, \cdots, Y_n)$ 唯一性指, 在相差一个零概率集的意义下相等, 即若 $\mathrm{P}\big(\omega : f_1(Y_1, \cdots, Y_n) \neq f_2(Y_1, \cdots, Y_n)\big) = 0$, 则此时把 f_1 和 f_2 看作同一个随机变量.

那么 (2.2.1) 中定义的 f 与 $\widetilde{f}$ 是否相等? $\widetilde{f}$ 会有什么性质? 这个问题留给读者思考.

例 2.2.1. (X,Y) 为离散型随机向量, 其分布列为

$$\begin{pmatrix} (x_i,y_j) \\ p_{ij} \end{pmatrix}_{i=1,2,\cdots,\ j=1,2,\cdots}$$

且假定 $p_{ij}>0$, 若 $\mathrm{E}|X|<\infty$, 求 $\mathrm{E}[X|Y]$.

解: 与例 (2.1.3) 的解法一样, 只需要保证等式两边的量有意义即可.

$$\mathrm{E}[X|Y](\omega)=\sum_{j=1}^{\infty}\frac{\sum\limits_{i=1}^{\infty}x_ip_{ij}}{\sum\limits_{i=1}^{\infty}p_{ij}}\mathbf{1}_{\{y_j\}}(Y(\omega))=\sum_{j=1}^{\infty}\mathrm{E}(X|Y=y_j)\mathbf{1}_{\{y_j\}}(Y(\omega)).$$

□

例 2.2.2. (X,Y) 是服从密度函数为 $p(x,y)>0$ 的连续型随机向量, $(x,y)\in\mathbb{R}^2$, $\mathrm{E}|X|<\infty$, 求 $\mathrm{E}[X|Y]$.

解: 与例 (2.1.4) 的解法一样, 只需要保证等式两边的量有意义即可

$$\mathrm{E}[X|Y](\omega)=\frac{\int_{\mathbb{R}}xp(x,Y(\omega))\mathrm{d}x}{\int_{\mathbb{R}}p(x,Y(\omega))\mathrm{d}x}.$$

□

例 2.2.3. (X,Y) 为离散型随机变量, 其分布列为

$$\begin{pmatrix} (x_i,y_j) \\ p_{ij} \end{pmatrix}_{i=1,2,\cdots,\ j=1,2,\cdots}$$

且假定 $p_{ij}>0$. 设 h 为 Borel 函数, 若 $\mathrm{E}|h(X)|<\infty$, 求 $\mathrm{E}[h(X)|Y]$.

解 1: 先求 $(h(X),Y)$ 的联合分布列, 再利用例 (2.2.1) 的方法即可.

解 2: 注意到 $\mathrm{E}[h(X)|Y](\omega)=\sum\limits_{j=1}^{\infty}\mathrm{E}(h(X)|Y=y_j)\mathbf{1}_{\{y_j\}}(Y(\omega))$. 其实问题就转化成了求 $\mathrm{E}(h(X)|Y=y_j)$, 即已知 $Y=y_j$ 时 $h(X)$ 的条件期望. 而由初等概率论知, 先求出已知 $Y=y_j$ 时, X 的条件分布列是

$$\mathrm{P}(X=x_i|Y=y_j)=\frac{p_{ij}}{\sum\limits_{i=1}^{\infty}p_{ij}}.$$

再求出,

$$\mathrm{E}(h(X)|Y=y_j)=\frac{\sum\limits_{i=1}^{\infty}h(x_i)p_{ij}}{\sum\limits_{i=1}^{\infty}p_{ij}}.$$

因此

$$\mathrm{E}[h(X)|Y](\omega)=\sum_{j=1}^{\infty}\frac{\sum\limits_{i=1}^{\infty}h(x_i)p_{ij}}{\sum\limits_{i=1}^{\infty}p_{ij}}\mathbf{1}_{\{y_j\}}(Y(\omega)).$$

□

例 2.2.4. (X,Y) 服从密度函数为 $p(x,y)$ 的连续型随机向量, 且 $p(x,y)>0$, $(x,y)\in\mathbb{R}^2$, $h(x)$ 为 Borel 函数, 满足 $\mathrm{E}|h(X)|<\infty$, 求 $\mathrm{E}[h(X)|Y]$.

我们知道 $(h(X),Y)$ 并不一定是连续型随机向量, 类似于例 (2.2.2) 中先求 $(h(X),Y)$ 的联合密度的方法并不适用. 由例 (2.2.3) 可以猜测

$$\mathrm{E}[h(X)|Y](\omega)=\frac{\int_{\mathbb{R}}h(x)p(x,Y(\omega))\mathrm{d}x}{\int_{\mathbb{R}}p(x,Y(\omega))\mathrm{d}x}=\frac{\int_{\mathbb{R}}h(x)p(x,y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}\Big|_{y=Y(\omega)}.$$

现在只需验证就行了.

解: 根据条件数学期望的定义, 首先证明 $\mathrm{E}\big(\big|\mathrm{E}[h(X)|Y]\big|\big)<\infty$.

$$\begin{aligned}
&\mathrm{E}\left|\frac{\int_{\mathbb{R}}h(x)p(x,Y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}\right| \\
=&\int_{\mathbb{R}}\left|\frac{\int_{\mathbb{R}}h(x)p(x,y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}\right|\cdot\big[\int_{\mathbb{R}}p(x,y)\mathrm{d}x\big]\mathrm{d}y \\
\le&\int_{\mathbb{R}}\frac{\int_{\mathbb{R}}|h(x)|p(x,y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}\cdot\big[\int_{\mathbb{R}}p(x,y)\mathrm{d}x\big]\mathrm{d}y \\
=&\int_{\mathbb{R}}\int_{\mathbb{R}}|h(x)|p(x,y)\mathrm{d}x\mathrm{d}y \\
=&\int_{\mathbb{R}}|h(x)|\big[\int_{\mathbb{R}}p(x,y)\mathrm{d}y\big]\mathrm{d}x=\mathrm{E}|h(X)|<\infty.
\end{aligned}$$

其次验证对于任意有界的 Borel 函数 g, $\mathrm{E}\big(h(X)g(Y)\big)=\mathrm{E}\big(\mathrm{E}[h(X)|Y]g(Y)\big)$ 成立.

$$\mathrm{E}\big(h(X)g(Y)\big)=\int_{\mathbb{R}}\int_{\mathbb{R}}h(x)g(y)p(x,y)\mathrm{d}x\mathrm{d}y.$$

$$\begin{aligned}
\mathrm{E}\Big(\frac{\int_{\mathbb{R}}h(x)p(x,Y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}g(Y)\Big)&=\int_{\mathbb{R}}\frac{\int_{\mathbb{R}}h(x)p(x,y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}g(y)\cdot\big[\int_{\mathbb{R}}p(x,y)\mathrm{d}x\big]\mathrm{d}y \\
&=\int_{\mathbb{R}}\int_{\mathbb{R}}h(x)g(y)p(x,y)\mathrm{d}x\mathrm{d}y \\
&=\mathrm{E}(h(X)g(Y)).
\end{aligned}$$

□

特别当 $h(x)=\mathbf{1}_{(-\infty,x]}$ 时, $\frac{\int_{\mathbb{R}}h(z)p(z,y)\mathrm{d}z}{\int_{\mathbb{R}}p(z,y)\mathrm{d}z}=\frac{\int_{-\infty}^{x}p(z,y)\mathrm{d}z}{\int_{\mathbb{R}}p(z,y)\mathrm{d}z}$, 当 x 变动时就是已知 $Y=y$ 时 X 的条件分布函数.

例 2.2.5. *在初等概率论中, 我们经常考虑已知事件 A 发生时事件 B 的条件概率 $\mathrm{P}(B|A)$, 那么如何用现在的条件数学期望的语言来表示? 设*

$$X=\mathbf{1}_B(\omega),\ Y=\mathbf{1}_A(\omega),$$

(X,Y) 的联合分布列为

$$\begin{pmatrix} (1,1) & (0,1) & (1,0) & (0,0) \\ \mathrm{P}(A\bigcap B) & \mathrm{P}(A\bigcap B^c) & \mathrm{P}(A^c\bigcap B) & \mathrm{P}(A^c\bigcap B^c)\end{pmatrix},$$

则

$$\begin{aligned}
\mathrm{E}[X|Y](\omega)&=\frac{\mathrm{P}(A\bigcap B)}{\mathrm{P}(A)}\mathbf{1}_A(\omega)+\frac{\mathrm{P}(A^c\bigcap B)}{\mathrm{P}(A^c)}\mathbf{1}_{A^c}(\omega) \\
&=\mathrm{P}(B|A)\mathbf{1}_A(\omega)+\mathrm{P}(B|A^c)\mathbf{1}_{A^c}(\omega).
\end{aligned}$$

这个式子不但给出了当 A 发生时, 即 $\omega\in A$ 时, B 发生的条件概率, 同时也给出了当 A 不发生时, 即 $\omega\in A^c$ 时,B 发生的条件概率.

在初等概率论中, 我们考虑的条件概率、条件期望或条件分布, 都是“静态”的, 即**在给定** $(Y_1(\omega),\cdots,Y_n(\omega))$ **的某个具体取值** $(y_1,\cdots,y_2)$ **的条件下**, 事件 B 的条件概率或随机变量或 X 的条件期望. 而现在我们定义的条件数学期望是“动态”的, 是一个“函数”, 它给出了关于 $(Y_1(\omega),Y_2(\omega),\cdots,Y_n(\omega))$ 的每一个取值发生的条件下, 事件 B 的条件期望或 X 的条件概率. 也就是说在初等概率论中定义的条件期望是现在定义的条件数学期望的一个“具体的函数值”.

♣♣♣20230307

2.2.2 性质

为方便, 下面用 $\mathbf{Y}$ 表示随机向量 $(Y_1,\cdots,Y_n)$, $\mathbf{y}$ 表示向量 $(y_1,\cdots,y_n)$. 例如 $\mathrm{E}[X|Y_1,\cdots,Y_n]$ 记为 $\mathrm{E}[X|\mathbf{Y}]$, $\mathrm{E}(X|Y_1=y_1,\cdots,Y_n=y_n)$ 记为 $\mathrm{E}(X|\mathbf{Y}=\mathbf{y})$.

性质 2.2.1. (全期望公式) $\mathrm{E}\big(\mathrm{E}[X|\mathbf{Y}]\big)=\mathrm{E}X$.

证明: 在定义2.2.1的 (2.2.1) 式中取 $g(Y_1,\cdots,Y_n)\equiv 1$ 即得. □

我们来看这个公式的概率含义:

例 2.2.6. $X(\omega)=\mathbf{1}_A(\omega)$, $Y(\omega)\sim\begin{pmatrix} y_1 & \cdots & y_n & \cdots \\ p_1 & \cdots & p_n & \cdots\end{pmatrix}$, $p_i>0$. 容易计算出

$$\begin{aligned}\mathrm{E}[X|Y](\omega) &= \mathrm{E}(X|Y=y_1)\mathbf{1}_{\{y_1\}}(Y(\omega))+\cdots+\mathrm{E}(X|Y=y_n)\mathbf{1}_{\{y_n\}}(Y(\omega))+\cdots \\ &= \mathrm{P}(A|Y=y_1)\mathbf{1}_{\{y_1\}}(Y(\omega))+\cdots+\mathrm{P}(A|Y=y_n)\mathbf{1}_{\{y_n\}}(Y(\omega))+\cdots.\end{aligned}$$

事实上,

$$\begin{aligned}\mathrm{E}\big(\mathrm{E}[X|Y]\big) &= \mathrm{P}(A|Y=y_1)\mathrm{E}\big(\mathbf{1}_{\{y_1\}}(Y(\omega))\big)+\cdots+\mathrm{P}(A|Y=y_n)\mathrm{E}\big(\mathbf{1}_{\{y_n\}}(Y(\omega))\big)+\cdots \\ &= \mathrm{P}(A|Y=y_1)\mathrm{P}(Y(\omega)=y_1)+\cdots+\mathrm{P}(A|Y=y_n)\mathrm{P}(Y(\omega)=y_n)+\cdots \\ &= \mathrm{P}(A)=\mathrm{E}X,\end{aligned}$$

即性质2.2.1就是全概率公式.

例 2.2.7. 设 (X,Y) 是连续型随机向量, 密度函数满足 $p(x,y)>0$, $(x,y)\in\mathbb{R}^2$. X 和 Y 的边际分布分别是 $p_X(x)=\int_{\mathbb{R}}p(x,y)\mathrm{d}y$ 和 $p_Y(y)=\int_{\mathbb{R}}p(x,y)\mathrm{d}x$.

设 h 为 Borel 函数,

$$\mathrm{E}(h(X)|Y=y)=\frac{\int_{\mathbb{R}}h(x)p(x,y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x},\ \mathrm{E}[h(X)|Y]=\frac{\int_{\mathbb{R}}h(x)p(x,Y)\mathrm{d}x}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x},$$

则由性质2.2.1得到

$$\mathrm{E}h(X)=\mathrm{E}\big(\mathrm{E}[h(X)|Y]\big)=\int_{\mathbb{R}}\mathrm{E}(h(X)|Y=y)p_Y(y)\mathrm{d}y.$$

这就是连续型随机变量的全期望公式, 同时它也给出了利用条件数学期望计算随机变量期望值的一个办法.

例 2.2.8. 设随机变量 X 服从 $[0,1]$ 上的均匀分布, 而在随机变量 $X=x$ 条件下, 随机变量 Y 的条件分布为 $N(x,x^2)$. 求 $\mathrm{E}Y$, $\mathrm{var}(Y)$.

解: $p_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi}x}\mathrm{e}^{-\frac{(y-x)^2}{2x^2}}$, 容易得 $\mathrm{E}(Y|X=x)=x$.

$$\mathrm{E}Y=\mathrm{E}\big(\mathrm{E}[Y|X]\big)=\int_{\mathbb{R}}\mathrm{E}(Y|X=x)p_X(x)\mathrm{d}x=\int_0^1 x\mathrm{d}x=\frac{1}{2}.$$

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{E}\big(\mathrm{E}[(Y-\frac{1}{2})^2|X]\big) \\ &= \int_0^1\int_{\mathbb{R}}(y-\frac{1}{2})^2\frac{1}{\sqrt{2\pi}x}\mathrm{e}^{-\frac{(y-x)^2}{2x^2}}\mathrm{d}y\mathrm{d}x \\ &= \int_0^1(2x^2-x+\frac{1}{4})\mathrm{d}x \\ &= \frac{5}{12}.\end{aligned}$$

性质 2.2.2. 线性性质与单调性:

(i) 设 $a,\ b\in\mathbb{R}$, 则 $\mathrm{E}[aX_1+bX_2|\mathbf{Y}]=a\mathrm{E}[X_1|\mathbf{Y}]+b\mathrm{E}[X_2|\mathbf{Y}]$;

(ii) 若 $X_1 \geq X_2$, 则 $\mathrm{E}[X_1|\mathbf{Y}] \geq \mathrm{E}[X_2|\mathbf{Y}]$; 若 $X \geq 0$, 则 $\mathrm{E}[X|\mathbf{Y}] \geq 0$; $\big|\mathrm{E}[X|\mathbf{Y}]\big| \leq \mathrm{E}[|X||\mathbf{Y}]$.

性质 2.2.3. 若 h 为 n 维 Borel 函数, 且 $\mathrm{E}|h(\mathbf{Y})| < \infty$, 则 $\mathrm{E}[h(\mathbf{Y})|\mathbf{Y}] = h(\mathbf{Y})$.

这个性质的证明是显然的, 理解也是很有趣的.

(1) 从投影角度来看 $h(Y_1, \cdots, Y_n)$ 在 $L^2(\Omega, \sigma(Y_1, \cdots, Y_n), \mathrm{P})$ 上的投影只能是它本身.

(2) 从 "静态-动态" 的观点看, 当 "静态地" 给定 $\mathbf{Y} = \mathbf{y}$ 的条件下, $h(\mathbf{Y})$ 的期望 $\mathrm{E}(h(\mathbf{Y})|\mathbf{Y} = \mathbf{y})$ 只能是 $h(\mathbf{y})$. 因此, 再让 $\mathbf{Y}$ "动起来", 就是 $\mathrm{E}[X|\mathbf{Y}]$ 就是 $h(\mathbf{Y})$ 了.

性质 2.2.4. $(X_1, \cdots, X_m)^{\mathrm{t}} (\equiv \mathbf{X})$ 与 $(Y_1, \cdots, Y_n)^{\mathrm{t}}$ 相互独立, 且对于 m 维 Borel 函数 h, 满足 $\mathrm{E}|h(\mathbf{X})| < \infty$, 则

$$\mathrm{E}[h(\mathbf{X})|\mathbf{Y}] = \mathrm{E}h(\mathbf{X}).$$

这条性质可以这样理解: 因为 $\mathbf{X}$ 与 $\mathbf{Y}$ 独立, 当 $\mathbf{Y}$ 已知时, $h(\mathbf{X})$ 的条件期望与 $\mathbf{Y}$ 没有任何关系, 所以应该是一个常数, 而且就是 $h(\mathbf{X})$ 的期望本身. 由性质2.2.3、性质2.2.4都有若 $X = c$, c 为常数, 则 $\mathrm{E}[X|\mathbf{Y}] = c$.

证明: 由定义2.2.1, 只需验证式子 (2.2.1). 由独立性, 并注意到 $\mathrm{E}h(\mathbf{X})$ 是常数, 则对任意的有界 Borel 可测函数 $g(\mathbf{y})$ 得到

$$\mathrm{E}\big(h(\mathbf{X})g(\mathbf{Y})\big) = \mathrm{E}h(\mathbf{X})\,\mathrm{E}g(\mathbf{Y}) = \mathrm{E}\big[\big(\mathrm{E}h(\mathbf{X})\big)g(\mathbf{Y})\big].$$

□

性质 2.2.5. [1] h 为 n 维 Borel 函数, g 为 m 维 Borel 函数, 满足 $\mathrm{E}|h(\mathbf{Y})| < \infty$, $\mathrm{E}|g(\mathbf{X})| < \infty$ 与 $\mathrm{E}|h(\mathbf{Y})g(\mathbf{X})| < \infty$, 则

$$\mathrm{E}[h(\mathbf{Y})g(\mathbf{X})|\mathbf{Y}] = \mathbf{h}(\mathbf{Y})\mathrm{E}[\mathbf{g}(\mathbf{X})|\mathbf{Y}]. \tag{2.2.4}$$

进一步, 由全期望公式得到

$$\mathrm{E}\big(h(\mathbf{Y})g(\mathbf{X})\big) = \mathrm{E}\big(\mathbf{h}(\mathbf{Y})\mathrm{E}[\mathbf{g}(\mathbf{X})|\mathbf{Y}]\big). \tag{2.2.5}$$

这条性质可以这样地直观理解: 当 $\mathbf{Y} = \mathbf{y}$ 已知时, 那么 $h(\mathbf{Y})$ 应该是等于常数 $h(\mathbf{y})$, 因此 $\mathrm{E}(h(\mathbf{Y})g(\mathbf{X})|\mathbf{Y} = \mathbf{y})$ 应该就是 $h(\mathbf{y})\mathrm{E}(g(\mathbf{X})|\mathbf{Y} = \mathbf{y})$. 再让 $\mathbf{Y}$ "动起来" 时, 就得到了式子 (2.2.4).

例 2.2.9. (接例2.2.8) 求 $\mathrm{cov}(X, Y)$.

解: 因为 $\mathrm{E}(Y|X = x) = x$, 所以 $\mathrm{E}[Y|X] = X$.

$$\mathrm{E}(XY) = \mathrm{E}\big(\mathrm{E}[XY|X]\big) = \mathrm{E}\big(X\,\mathrm{E}[Y|X]\big) = \mathrm{E}X^2 = \frac{1}{3}.$$

从而 $\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y = \frac{1}{12}$.

□

我们来将性质2.2.5推广到更广的函数上. 首先, 设 $f(\mathbf{Y}, \mathbf{y}) = \mathrm{E}[g(\mathbf{X})\mathbf{h}(\mathbf{y})|\mathbf{Y}]$, 那么 $f(\mathbf{Y}, \mathbf{y})$ 也等于 $h(\mathbf{y})\mathrm{E}[g(\mathbf{X})|\mathbf{Y}]$. 因此,

$$\mathrm{E}[g(\mathbf{X})h(\mathbf{Y})|\mathbf{Y}] = f(\mathbf{Y}, \mathbf{Y}).$$

对于两列 Borel 函数 $(g_k(\mathbf{x}))_{k\geq 1}$ 和 $(h_l(\mathbf{y}))_{l\geq 1}$, 使用上面的结论, 则 $f_{k,l}(\mathbf{Y}, \mathbf{y}) = \mathrm{E}[g_k(\mathbf{X})h_l(\mathbf{y})|\mathbf{Y}]$ 满足

$$\mathrm{E}[g_k(\mathbf{X})h_l(\mathbf{Y})|\mathbf{Y}] = f_{k,l}(\mathbf{Y}, \mathbf{Y}).$$

---

[1] 若 $h(\mathbf{Y})$ 和 $g(\mathbf{X})$ 满足性质2.2.5的条件时, 利用测度论的知识可以严格证明该性质. 其关键在于此时并未假设 $h(\mathbf{Y})\mathrm{E}[g(\mathbf{X})|\mathbf{Y}]$ 的可积性, 所以证明时需要有一个取极限的过程, 要用到抽象的测度空间上的 Lebesgue 控制收敛定理, 这不在本讲义讨论的范围内.

依据条件数学期望的线性性质, 我们形式的有

$$\mathrm{E}\Big[\sum_{k=1,l=1}^{\infty} g_k(\mathbf{X})h_l(\mathbf{Y})|\mathbf{Y}\Big] = \sum_{k=1,l=1}^{\infty} f_{k,l}(\mathbf{Y},\mathbf{Y}).$$

我们想象一般关于 $(\mathbf{x},\mathbf{y})$ 的 Borel 函数 $h(\mathbf{x},\mathbf{y})$ 可以表示成 $\sum_{k=1,l=1}^{\infty} g_k(\mathbf{x})h_l(\mathbf{y})$, 设 $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathbf{Y}] = f(\mathbf{Y},\mathbf{y})$, 那么猜想可以有

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}] = f(\mathbf{Y},\mathbf{Y}).$$

如果, $\mathbf{X}$ 与 $\mathbf{Y}$ 独立, 那么 $g(\mathbf{X})$ 也与 $\mathbf{Y}$ 独立. 进一步利用性质2.2.4, 设 $f(\mathbf{y}) = \mathrm{E}(g(\mathbf{X})h(\mathbf{y}))$, 则

$$\mathrm{E}[g(\mathbf{X})h(\mathbf{Y})|\mathbf{Y}] = f(\mathbf{Y}).$$

同样, 对于一般的 $h(\mathbf{x},\mathbf{y})$, 设 $f(\mathbf{y}) = \mathrm{E}(h(\mathbf{X},\mathbf{y}))$, 我们猜测

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}] = f(\mathbf{Y}).$$

事实上, 这个猜测是正确的. 我们有如下的性质.

性质 2.2.6. 记 $(X_1,\cdots,X_m)$ 为 $\mathbf{X}$, $(Y_1,\cdots,Y_n)$ 为 $\mathbf{Y}$. 设 $\mathrm{E}|h(\mathbf{X},\mathbf{Y})| < \infty$, 任意的 $\mathbf{y}\in\mathbb{R}^n$, $\mathrm{E}|h(\mathbf{X},\mathbf{y})| < \infty$, 设 $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathbf{Y}](\omega) = f(\mathbf{Y}(\omega),\mathbf{y})$, 则

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}](\omega) = f(\mathbf{Y}(\omega),\mathbf{Y}(\omega)).$$

有时也记为

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}] = \mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathbf{Y}]\Big|_{\mathbf{y}=\mathbf{Y}}.$$

特别当 $\mathbf{X}$ 与 $\mathbf{Y}$ 相互独立时, $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathbf{Y}](\omega) = \mathrm{E}\Big(h(\mathbf{X},\mathbf{y})\Big) = f(\mathbf{y})$, 则

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}](\omega) = f(\mathbf{Y}(\omega)).$$

有时也记为

$$\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}] = \mathrm{E}\Big(h(\mathbf{X},\mathbf{y})\Big)\Big|_{\mathbf{y}=\mathbf{Y}}.$$

例 2.2.10. 若随机变量 X 与 Y 相互独立, 且分别服从参数为 λ 与 μ 的指数分布, 定义随机变量

$$Z = \begin{cases} 1, & X \le Y, \\ 0, & X > Y. \end{cases}$$

求 $\mathrm{E}Z$.

解: 取 $h(x,y) = \mathbf{1}_{\{x\le y\}}$, $Z = h(X,Y)$. X 的密度函数为

$$p_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x\ge 0,\ y\ge 0 \\ 0, & \text{其它} \end{cases},$$

则

$$\mathrm{E}h(X,y) = \int_{\{x\le y\}} p_X(x)\mathrm{d}x = \begin{cases} \int_0^y \lambda \mathrm{e}^{-\lambda x}\mathrm{d}x = 1-\mathrm{e}^{-\lambda y}, & y\ge 0, \\ 0, & y<0. \end{cases}$$

因为 X 与 Y 独立, 由性质2.2.6,

$$\mathrm{E}[h(X,Y)|Y] = (\mathrm{E}h(X,y))\big|_{y=Y} = \begin{cases} 1-\mathrm{e}^{-\lambda Y}, & Y\ge 0, \\ 0, & Y<0. \end{cases}$$

$$\begin{aligned} \mathrm{E}h(X,Y) &= \mathrm{E}\big(\mathrm{E}[h(X,Y)|Y]\big) \\ &= \mathrm{E}(1-\mathrm{e}^{-\lambda Y}) = \int_0^{\infty}(1-\mathrm{e}^{-\lambda y})\mu e^{-\mu y}\mathrm{d}y \\ &= \frac{\lambda}{\lambda+\mu}. \end{aligned}$$

□

例 2.2.11. (接例2.2.8, 例2.2.9) 证明 $\frac{Y}{X}$ 与 X 独立.

证明: 对于任意的 Borel 有界函数 f 和 g,

$$\begin{aligned}
\mathrm{E}\Big(f\big(\frac{Y}{x}\big)\big|X=z\Big) &= \int_{\mathbb{R}} f\big(\frac{y}{x}\big)\frac{1}{\sqrt{2\pi z}}\mathrm{e}^{-\frac{(y-z)^2}{2z^2}}\mathrm{d}y,\\
\mathrm{E}\Big[f\big(\frac{Y}{x}\big)\big|X\Big] &= \int_{\mathbb{R}} f\big(\frac{y}{x}\big)\frac{1}{\sqrt{2\pi X}}\mathrm{e}^{-\frac{(y-X)^2}{2X^2}}\mathrm{d}y,\\
\mathrm{E}\Big[f\big(\frac{Y}{X}\big)\big|X\Big] &\overset{\text{性质2.2.6}}{=} \mathrm{E}\Big[f\big(\frac{Y}{x}\big)\big|X\Big]\Big|_{x=X}\\
&= \int_{\mathbb{R}} f\big(\frac{y}{X}\big)\frac{1}{\sqrt{2\pi X}}\mathrm{e}^{-\frac{(y-X)^2}{2X^2}}\mathrm{d}y,\\
&\overset{\text{令}\tilde{y}=\frac{y}{X}}{=} \int_{\mathbb{R}} f(\tilde{y})\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(\tilde{y}-1)^2}{2}}\mathrm{d}\tilde{y} = \mathrm{E}f(Z),\quad \text{其中} Z\sim N(1,1).
\end{aligned}$$

即 $\mathrm{E}\big[f\big(\frac{Y}{X}\big)\big|X\big]=\mathrm{E}f(Z)$, 而且 $\mathrm{E}f\big(\frac{Y}{X}\big)=\mathrm{E}f(Z)$. 由此得到

$$\mathrm{E}\Big(f\big(\frac{Y}{X}\big)g(X)\Big)=\mathrm{E}\Big(\mathrm{E}\big[f\big(\frac{Y}{X}\big)g(X)\big|X\big]\Big)=\mathrm{E}\Big(g(X)\mathrm{E}\big[f\big(\frac{Y}{X}\big)\big|X\big]\Big)=\mathrm{E}f\big(\frac{Y}{X}\big)\mathrm{E}g(X).$$

□

虽然我们之前形式的导出了性质2.2.6, 但这个性质的证明是非平凡的, 要用到测度论典型方法和正则条件概率等概念, 我们这里略去它的严格证明, 只用连续型随机变量验证一下.

验证: 设 (X,Y) 密度函数满足 $p(x,y)>0$, $(x,y)\in\mathbb{R}^2$, 则已知 $Y=z$ 条件下, X 的条件密度为

$$p_{X|Y}(x|z)=\frac{p(x,z)}{\int_{\mathbb{R}}p(x,z)\mathrm{d}x}.$$

$$\begin{aligned}
\mathrm{E}(h(X,y)|Y=z) &= \int h(x,y)p_{X|Y}(x|z)\mathrm{d}x\\
&= \int h(x,y)\frac{p(x,z)}{\int_{\mathbb{R}}p(x,z)\mathrm{d}x}\mathrm{d}x.\\
\mathrm{E}[h(X,y)|Y] &= \int_{\mathbb{R}} h(x,y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}\mathrm{d}x\\
\mathrm{E}[h(X,y)|Y]\Big|_{y=Y} &= \int_{\mathbb{R}} h(x,Y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}\mathrm{d}x \qquad (2.2.6)
\end{aligned}$$

下面验证 (2.2.6) 式右端就是 $\mathrm{E}[h(X,Y)|Y]$.

(1) 首先验证可积性.

$$\begin{aligned}
\mathrm{E}\Big|\int_{\mathbb{R}} h(x,Y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}\mathrm{d}x\Big| &\le \int_{\mathbb{R}}\int_{\mathbb{R}}|h(x,y)|\frac{p(x,y)}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}\cdot\Big[\int_{\mathbb{R}}p(x,y)\mathrm{d}x\Big]\mathrm{d}x\mathrm{d}y\\
&= \mathrm{E}|h(X,Y)|<\infty
\end{aligned}$$

(2) 对于任意有界的 $g(Y)$,

$$\begin{aligned}
&\mathrm{E}\Big(\int_{\mathbb{R}} h(x,Y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)\mathrm{d}x}\mathrm{d}x\cdot g(Y)\Big)\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y)\frac{p(x,y)}{\int_{\mathbb{R}}p(x,y)\mathrm{d}x}\cdot g(y)\Big[\int_{\mathbb{R}}p(x,y)\mathrm{d}x\Big]\mathrm{d}x\mathrm{d}y\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y)g(y)p(x,y)\mathrm{d}x\mathrm{d}y=\mathrm{E}\Big(h(X,Y)g(Y)\Big).
\end{aligned}$$

进一步, 若 X 与 Y 相互独立, 则 $p(x,y)=p_X(x)p_Y(y)$.

$$\mathrm{E}(h(X,y)|Y=z)=\mathrm{E}\big(h(X,y)\big)=\int_{\mathbb{R}}h(x,y)p_X(x)\mathrm{d}x,$$

$$\mathrm{E}\big(h(X,y)\big)\Big|_{y=Y}=\int_{\mathbb{R}}h(x,Y)p_X(x)\mathrm{d}x.$$

下面只需验证:

(1)

$$\begin{aligned}\mathrm{E}\Big|\int_{\mathbb{R}} h(x,Y)p_X(x)\mathrm{d}x\Big| &\le \int_{\mathbb{R}}\int_{\mathbb{R}} |h(x,y)|p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y \\ &\le \mathrm{E}|h(X,Y)| < \infty\end{aligned}$$

(2) 对于任意有界的 $g(Y)$,

$$\begin{aligned}&\mathrm{E}\Big(\int_{\mathbb{R}} h(x,Y)p_X(x)\mathrm{d}x\cdot g(Y)\Big)\\ =& \int\Big[\int h(x,y)p_X(x)\mathrm{d}x\Big]g(y)p_Y(y)\mathrm{d}y\\ =& \int\int h(x,y)g(y)p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y = \mathrm{E}\Big(h(X,Y)g(Y)\Big).\end{aligned}$$

□

性质 2.2.7. 记 $(X_1,\cdots,X_m)$ 为 $\mathbf{X}$, 若 h 为 m 维 Borel 函数, 且 $\mathrm{E}h^2(\mathbf{X})<\infty$, $\mathrm{E}[h(\mathbf{X})|\mathbf{Y}]$ 是定义2.2.1中定义的 $h(\mathbf{X})$ 关于 $\mathbf{Y}$ 的条件数学期望, 则

$$\mathrm{E}\big(h(\mathbf{X})-\mathrm{E}[h(\mathbf{X})|\mathbf{Y}]\big)^2 = \inf_{g(\mathbf{Y})\text{满足}\mathrm{E}g^2(\mathbf{Y})<\infty} \mathrm{E}\big(g(\mathbf{Y})-h(\mathbf{X})\big)^2$$

这个性质不但说明当 $\mathrm{E}h^2(\mathbf{X})<\infty$, $h(\mathbf{X})$ 按照定义2.2.1在 $L^1(\Omega,\mathcal{F},\mathrm{P})$ 上定义的条件数学期望与按照定义2.1.1在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 上定义的条件数学期望一致, 而且还说明 $\mathrm{E}\big(\mathrm{E}[h(\mathbf{X})|\mathbf{Y}]\big)^2<\infty$.

事实上, 如果 $f(\mathbf{Y})$ 是 $h(\mathbf{X})$ 在 $L^2(\Omega,\sigma(\mathbf{Y}),\mathrm{P})$ 上的投影, 那么

$$\mathrm{E}\big(h(\mathbf{X})-f(\mathbf{Y})\big)^2 = \inf_{g(\mathbf{Y})\text{满足}\mathrm{E}g^2(\mathbf{Y})<\infty} \mathrm{E}\big(g(\mathbf{Y})-h(\mathbf{X})\big)^2.$$

由上一节的推导知道, 对于任意的满足 $\mathrm{E}g^2(\mathbf{Y})<\infty$ 的 $g(\mathbf{Y})$, 我们得到

$$\mathrm{E}\big(h(\mathbf{X})g(\mathbf{Y})\big) = \mathrm{E}\big(f(\mathbf{Y})g(\mathbf{Y})\big).$$

而对于有界的 $g(\mathbf{Y})$,

$$\mathrm{E}\big(\mathrm{E}[h(\mathbf{X})|\mathbf{Y}]g(\mathbf{Y})\big) = \mathrm{E}\big(f(\mathbf{Y})g(\mathbf{Y})\big).$$

我们知道, 如果 $g(\mathbf{Y})$ 有界, 一定 $\mathrm{E}g^2(\mathbf{Y})<\infty$, 那么,

$$\mathrm{E}\big((f(\mathbf{Y})-\mathrm{E}[h(\mathbf{X})|\mathbf{Y}])g(\mathbf{Y})\big) = 0.$$

利用 $g(\mathbf{Y})$ 的任意性, 由测度论的知识我们得到 $f(\mathbf{Y})$ 就是 $\mathrm{E}[h(\mathbf{X})|\mathbf{Y}]$.

性质 2.2.8. (平滑性、投影性) 若 $\mathrm{E}|X|<\infty$, 记 $(Y_1,\cdots,Y_n)$ 为 $\mathbf{Y}$, $(Z_1,\cdots,Z_k)$ 为 $\mathbf{Z}$, 则

$$\mathrm{E}[X|\mathbf{Z}] = \mathrm{E}\Big[\mathrm{E}[X|\mathbf{Y},\mathbf{Z}]\Big|\mathbf{Z}\Big] = \mathrm{E}\Big[\mathrm{E}[X|\mathbf{Z}]\Big|\mathbf{Y},\mathbf{Z}\Big].$$

这条性质可以这样直观理解: 将 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 想象成三维欧氏空间. 向量 X 朝向量 Z 决定的方向投影, 相当于 X 先向 Y,Z 张成的平面投影, 再向 Z 方向投影; 或者对于 X 在 Z 方向的投影 $\mathrm{E}[X|Z]$, 这个投影在 Y,Z 张成的平面上的投影就是它自己.

证明: 由性质2.2.3, 易得 $\mathrm{E}\big[\mathrm{E}[X|\mathbf{Z}]\big|\mathbf{Y},\mathbf{Z}\big] = \mathrm{E}[X|\mathbf{Z}]$. 下面证明第一个等号成立, 即只需验证 $\mathrm{E}[X|\mathbf{Z}]$ 为 $\mathrm{E}[X|\mathbf{Y},\mathbf{Z}]$ 关于 $\mathbf{Z}$ 的条件数学期望.

(1) 显然 $\mathrm{E}\big|\mathrm{E}[X|\mathbf{Z}]\big|<\infty$;

(2) 对于任意的有界 Borel 函数 g,

$$\begin{aligned}
\mathrm{E}\Big(\mathrm{E}[X|\mathbf{Z}]g(\mathbf{Z})\Big) &= \mathrm{E}\Big(\mathrm{E}[Xg(\mathbf{Z})|\mathbf{Z}]\Big)\\
&= \mathrm{E}(Xg(\mathbf{Z}))\\
\mathrm{E}\Big(\mathrm{E}[X|\mathbf{Y},\mathbf{Z}]g(\mathbf{Z})\Big) &= \mathrm{E}\Big(\mathrm{E}[Xg(\mathbf{Z})|\mathbf{Y},\mathbf{Z}]\Big)\\
&= \mathrm{E}(Xg(\mathbf{Z})).
\end{aligned}$$

□

例 2.2.12. (回到例2.1.1) 设 $\mathrm{P}(\eta_n=1)=p$, $\mathrm{P}(\eta_n=-1)=1-p=q$, 设 b_k 非负有界,

$$\xi_{n+1}=\xi_0+\sum_{k=1}^{n+1}b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k,$$

求 $\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]$, 即求 ξ_{n+1} 关于 $(\xi_0,\eta_1,\cdots,\eta_n)$ 的预测函数.

解:

$$\begin{aligned}
&\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]\\
=\ &\mathrm{E}\Big[\xi_0+\sum_{k=1}^{n+1}b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k\Big|\xi_0,\eta_1,\cdots,\eta_n\Big]\\
=\ &\mathrm{E}[\xi_0|\xi_0,\eta_1,\cdots,\eta_n]+\sum_{k=1}^{n+1}\mathrm{E}\Big[b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k\Big|\xi_0,\eta_1,\cdots,\eta_n\Big]
\end{aligned}$$

(1) ξ_0 是 $(\xi_0,\eta_1,\cdots,\eta_n)$ 的函数, 故 $\mathrm{E}[\xi_0|\xi_0,\eta_1,\cdots,\eta_n]=\xi_0$;

(2) $k\le n$ 时, $b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k$ 是 $(\xi_0,\eta_1,\cdots,\eta_n)$ 的函数, 则

$$\mathrm{E}[b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k|\xi_0,\eta_1,\cdots,\eta_n]=b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k.$$

(3) $k=n+1$ 时, 由性质2.2.4、性质2.2.5

$$\begin{aligned}
&\mathrm{E}[b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)\eta_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]\\
=\ &b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)\,\mathrm{E}[\eta_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]\\
=\ &b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)\,\mathrm{E}(\eta_{n+1}).
\end{aligned}$$

因此

$$\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_0+\sum_{k=1}^{n}b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k+b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)\mathrm{E}(\eta_{n+1}).$$

当 $p=q=\frac{1}{2}$,

$$\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_0+\sum_{k=1}^{n}b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\eta_k=\xi_n;$$

当 $p>q$,

$$\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_n+b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)(p-q)\ge\xi_n;$$

当 $p<q$,

$$\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_n+b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)(p-q)\le\xi_n.$$

这说明：当 $p=q=\frac{1}{2}$ 时, 即博弈是公平时, ξ_{n+1} 的最佳预测函数是 ξ_n,

$$\mathrm{E}[\eta_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]-\xi_n=0,$$

那么我们预测到了什么? 任何有趋势性的结果都没有预测到. 这正是“公平博弈”的体现, 也是“纯粹的随机性”的表现. 当 $p>q$ 时, 最佳预测函数是 $\xi_n+b_n(\xi_0,\eta_1,\cdots,\eta_n)(p-q)$, 我们预测到了比 n 时刻多赢了 $b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)(p-q)$; 当 $p<q$ 时, $\cdots$. **未完待续, 重写**

□

♣♣♣20230309

例 2.2.13. 如何用条件数学期望来定义马氏链 (过程).

$\{X_n\}_{n\geq 1}$ 为一个随机过程, 状态空间为 $S=\{1,2,3\cdots\}$. 在应用随机过程的课程中是如下定义马氏链的: 对于任意的时刻 n, 任意的 $i,j_k\in S$, $1\leq k\leq n$,

$$\mathrm{P}(X_{n+1}=i|X_n=j_n,X_{n-1}=j_{n-1},\cdots,X_1=j_1)=\mathrm{P}(X_{n+1}=i|X_n=j_n).$$

或者等价于对于任意的有界 Borel 函数 f,

$$\mathrm{E}(f(X_{n+1})|X_n=j_n,X_{n-1}=j_{n-1},\cdots,X_1=j_1)=\mathrm{E}(f(X_{n+1})|X_n=j_n).$$

那么现在我们改用条件数学期望来看上面的表达, 其实就可以理解为

$$\mathrm{E}[f(X_{n+1})|X_n,X_{n-1},\cdots,X_1]=\mathrm{E}[f(X_{n+1})|X_n] \tag{2.2.7}$$

因此可以重新给马氏链一个定义: *称随机过程 $\{X_n\}_{n\geq 1}$ 是一个马氏链, 如果对于任意的有界 Borel 函数 f, (2.2.7) 式成立.*

如何将上面两个例子推广到连续时间. 这里便涉及到如何理解关于 $\{X_t,\, t\in I\}$ 一族随机变量的函数,"信息" 等问题.

2.3 条件数学期望的定义和性质 (II)

2.3.1 定义

设 X 是 $(\Omega,\mathcal{F},\mathrm{P})$ 上的一个随机变量, 若只考虑由 X 来描述的随机现象时, 我们由此可以得到什么样的 "信息"? 事实上, 对于任意的 Borel 函数 f, $f(X)$ 可以理解成对 X 所描述的随机现象的一个"信息提取".

因此, 对于任意 Borel 函数 f, $f(X)$ 构成的随机变量全体就是对 X 描述的随机现象"全部信息的提取". 由实变函数知道 f 是一列简单 Borel 可测函数的极限, 而简单 Borel 可测函数是 Borel 示性函数的线性组合 $\sum_{i=1}^n a_i\mathbf{1}_{A_i}$, $A_i\in\mathcal{B}(\mathbb{R})$, 而且满足 $A_i\cap A_j=\emptyset$. 故而, 若承认线性运算和极限运算是我们的 "认知" 手段, 可以操作来提取信息的话, 那么对于任意的 Borel 集 A, $\mathbf{1}_A(X)$ 的全体就是对 X 所描述的随机现象"全部信息的提取". 而 $\mathbf{1}_A(X)$ 是和事件 $\{\omega:\ X(\omega)\in A\}$ 等价的, 因此, 事件构成的集合 $\big\{\{\omega:X(\omega)\in A\},\ A\in\mathcal{B}(\mathbb{R})\big\}$ 就是 X 描述的随机现象的"全部信息的提取"; 也就是说, 当我们只关心 X 描述的随机现象时, 我们只要完全清楚 $\big\{\{\omega:X(\omega)\in A\},\ A\in\mathcal{B}(\mathbb{R})\big\}$ 就可以了. 可以很容易证明 $\big\{\{\omega:X(\omega)\in A\},\ A\in\mathcal{B}(\mathbb{R})\big\}$ 是事件域, **这个事件域称为由 X 生成的事件域 (σ-域), 记为 $\sigma(X)$.** 易证 $\sigma(X)\subset\mathcal{F}$.

当我们只考虑 X 描述的随机现象时, 可以限制在概率空间 $(\Omega,\sigma(X),\mathrm{P})$ 上就行了.

性质 2.3.1. *设 F 为 $(\Omega,\sigma(X),\mathrm{P})$ 上的一个随机变量, 则存在一个 Borel 函数 f, 使得 $F(\omega)=f(X(\omega))$. 反之, 若 f 为 $\mathbb{R}$ 上的 Borel 函数, 则 $f(X)$ 为 $(\Omega,\sigma(X),\mathrm{P})$ 上的随机变量.*

$$\begin{tikzcd}
(\Omega,\sigma(X)) \arrow[r, "X"] \arrow[dr, "F", "F=f(X)"'] & (\mathbb{R},\mathcal{B}(\mathbb{R})) \arrow[d, "f"] \\
 & (\mathbb{R},\mathcal{B}(\mathbb{R}))
\end{tikzcd}$$

这个命题把随机变量 X 的"信息"提取 $f(X)$ 与 X 生成的事件域联系在了一起. 类似的, 当我们考虑 $(\Omega,\mathcal{F},\mathrm{P})$ 上的有限个随机变量 $X_1,\cdots,X_n$ 描述的随机现象时, 对于全体 n 维 Borel 函数 $f(x_1,\cdots,x_n)$, 考虑 $f(X_1,\cdots,X_n)$ 全体, 则它们把 $(X_1,\cdots,X_n)$ 的"全部信息都提取了出来". 同样由实变函数知道, 事件域 $\big\{\{\omega:(X_1(\omega),\cdots,X_n(\omega))\in A\},\ A\in\mathcal{B}(\mathbb{R}^n)\big\}$ 提取了 $(X_1,\cdots,X_n)$ 的"全部信息", **这个事件域称为由 $(X_1,\cdots,X_n)$ 生成的事件域 (σ-域), 记为 $\sigma(X_1,\cdots,X_n)$.** 易证 $\sigma(X_1,\cdots,X_n)\subset\mathcal{F}$.

性质 2.3.2. 设 F 为 $(\Omega, \sigma(X_1, \cdots, X_n), \mathrm{P})$ 上的一个随机变量, 则存在一个 n 维 Borel 函数 f, 使得 $F(\omega) = f(X_1(\omega), \cdots X_n(\omega))$. 反之, 若 f 为 n 维 Borel 函数, 则 $f(X_1, \cdots, X_n)$ 为 $(\Omega, \sigma(X_1, \cdots, X_n), \mathrm{P})$ 上的随机变量.

$$
\begin{tikzcd}
(\Omega,\sigma(X_1,\cdots,X_n)) \arrow[r, "{(X_1,\cdots,X_n)}"] \arrow[dr, "F", "{F=f(X_1,\cdots,X_n)}"'] & (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n)) \arrow[d, "f"] \\
 & (\mathbb{R},\mathcal{B}(\mathbb{R}))
\end{tikzcd}
$$

由上面的命题我们知道, $f(X_1, \cdots, X_n)$ 可以通过 $(\Omega, \sigma(X_1, \cdots, X_n))$ 上的一个随机变量 F 来表达. 下面我们再分析 $\sigma(X_1, \cdots, X_n)$ 的结构. 由实变函数知 n-维 Borel 集族 $\mathcal{B}(\mathbb{R}^n)$ 是包含 n 维开矩形的最小 σ-域. 所谓开矩阵是指 $(a_1, b_1) \times \cdots \times (a_n, b_n)$. 因此 $\sigma(X_1, \cdots, X_n)$ 就是包含 $\{\omega : X_1 \in (a_1, b_1), \cdots, X_n \in (a_n, b_n)\}$ 的最小事件域, 或者等价于包含 $\{\omega : X_1(\omega) < b_1, \cdots, X_n(\omega) < b_n\}$, 或者等价包含 $\{\omega : X_1(\omega) \le b_1, \cdots, X_n(\omega) \le b_n\}$ 的最小事件域. 由此, 将这一结论推广到一族随机变量 $\{X_t, t \in I\}$ 就有如下的定义:

定义 2.3.1. 设 $\{X_t,\, t \in I\}$ 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的一族随机变量, 称包含 $\big\{\{\omega : X_t \in (a_t, b_t)\}, \forall t \in I, \forall a_t, b_t \in \mathbb{R}\big\}$ 的最小事件域 (σ-域) 为 $\{X_t,\, t \in I\}$ 生成的事件域, 记为 $\sigma(X_t,\, t \in I)$.

因此, 类似于性质2.3.1和性质2.3.2, 一个关于 $\{X_t,\, t \in I\}$ 的"可测函数", 实质上就是关于 $(\Omega, \sigma(X_t, t \in I))$ 的一个随机变量.

定义 2.3.2. 设 $\{Y_t, t \in I\}$ 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的一族随机变量, X 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量, 且 $\mathrm{E}|X| < \infty$, 称 $(\Omega, \sigma(Y_t, t \in I), \mathrm{P}))$ 上的随机变量 Z 是 X 关于 $\{Y_t,\, t \in I\}$ 的**条件数学期望**, 若 Z 满足

(1) $\mathrm{E}|Z| < \infty$;

(2) 对于任意有界的 $(\Omega, \sigma(Y_t, t \in I), \mathrm{P}))$ 上的随机变量 F, 下式成立
$$\mathrm{E}(XF) = \mathrm{E}(ZF). \tag{2.3.1}$$
记 $Z = \mathrm{E}[X|Y_t,\ t \in I]$ 或 $Z = \mathrm{E}[X|\sigma(Y_t,\, t \in I)]$.

(2.3.1) 等价于对于任意的 $A \in \sigma(Y_t,\ t \in I)$, $\mathrm{E}(X\mathbf{1}_A) = \mathrm{E}(Z\mathbf{1}_A)$.

若一个随机现象仅用一个事件域 $\mathcal{G}(\subset \mathcal{F})$ 描述时, 我们有条件数学期望的一般定义.

定义 2.3.3. 设 X 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的一个随机变量, 满足 $\mathrm{E}|X| < \infty$, $\mathcal{G}$ 为 $\mathcal{F}$ 的一个子事件域, 称 Z 为 X 关于子事件域 $\mathcal{G}$ 的**条件数学期望**, 若 Z 满足

(1) Z 是 $(\Omega, \mathcal{G}, \mathrm{P})$ 上的随机变量, 即 $\forall a \in \mathbb{R}$, $\{\omega : Z(\omega) \le a\} \in \mathcal{G}$, 而且 $\mathrm{E}|Z| < \infty$.

(2) 对于任意的 $(\Omega, \mathcal{G}, \mathrm{P})$ 上的有界随机变量 F,
$$\mathrm{E}(XF) = \mathrm{E}(ZF). \tag{2.3.2}$$
记 $Z = \mathrm{E}[X|\mathcal{G}]$.

(2.3.2) 等价于对于任意 $A \in \mathcal{G}$, $\mathrm{E}(X\mathbf{1}_A) = \mathrm{E}(Z\mathbf{1}_A)$. $\mathrm{E}[X|\mathcal{G}]$ 的存在唯一性由 Radon-Nikodym 定理得到的. X 为 $(\Omega, \mathcal{F})$ 随机变量, 但 X 不一定为 $(\Omega, \mathcal{G})$ 的随机变量.

2.3.2 性质

性质 2.3.3. (全期望公式) $\mathrm{E}\big(\mathrm{E}[X|\mathcal{G}]\big) = \mathrm{E}X$.

证明: 在定义2.3.3的 (2.3.2) 式中取 $F = \mathbf{1}_\Omega$ 即可. $\square$

例 2.3.1. X 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量, $A_1, A_2, \cdots, A_n, \cdots$ 为 Ω 的一个分割, 且 $A_i \in \mathcal{F}$, $i = 1, 2, \cdots$, 即 $A_n \cap A_m = \Phi$, $n \neq m$, 且 $\sum\limits_{n=1}^{\infty} A_n = \Omega$. 令 $\mathcal{G}$ 为包含 $A_1, A_2, \cdots, A_n, \cdots$ 的最小事件域, 由此来验证上述性质.

解: $(\Omega, \mathcal{G})$ 的任何一个随机变量 Z 一定能写成 $Z(\omega) = \sum\limits_{n=1}^{\infty} a_n \mathbf{1}_{A_n}(\omega)$ 的形式, 则 $\mathrm{E}[X|\mathcal{G}] = \sum\limits_{n=1}^{\infty} a_n \mathbf{1}_{A_n}(\omega)$. 下面求 a_n 即可.

由 (2.3.2) 取 $F(\omega) = \mathbf{1}_{A_k}(\omega)$, 则

$$\begin{aligned}\mathrm{E}\Big(X\mathbf{1}_{A_k}(\omega)\Big) &= \mathrm{E}\Big(\sum_{n=1}^{\infty} a_n \mathbf{1}_{A_n}(\omega)\cdot \mathbf{1}_{A_k}(\omega)\Big) \\ &= \mathrm{E}\Big(a_k \mathbf{1}_{A_k}(\omega)\Big) = a_k \mathrm{E}\Big(\mathbf{1}_{A_k}(\omega)\Big)\end{aligned}$$

则

$$a_k = \begin{cases} \frac{\mathrm{E}(X\mathbf{1}_{A_k}(\omega))}{\mathrm{P}(A_k)} = \mathrm{E}[X|A_k], & \mathrm{P}(A_k) > 0 \\ c_k(\text{任意常数}), & \mathrm{P}(A_k) = 0 \end{cases}.$$

因此 $\mathrm{E}[X|\mathbf{1}_G] = \sum\limits_{n=1}^{\infty} \mathrm{E}[X|A_n]\mathbf{1}_{A_n}(\omega)$, 其中

$$\mathrm{E}[X|A_n] = \begin{cases} \mathrm{E}[X|A_n], & \text{若}\mathrm{P}(A_k) > 0 \\ c_k(\text{任意常数}), & \mathrm{P}(A_k) = 0 \end{cases}.$$

$$\begin{aligned}\mathrm{E}\Big(\mathrm{E}[X|\mathbf{1}_G]\Big) &= \mathrm{E}\Big(\sum_{\substack{n=1\\ \mathrm{P}(A_n)>0}}^{\infty} \frac{\mathrm{P}\Big(X\mathbf{1}_{A_n}(\omega)\Big)}{\mathrm{P}(A_n)}\cdot \mathbf{1}_{A_n} + \sum_{\substack{n=1\\ \mathrm{P}(A_n)=0}}^{\infty} c_n \mathbf{1}_{A_n}\Big) \\ &= \sum_{\substack{n=1\\ \mathrm{P}(A_n)>0}}^{\infty} \frac{\mathrm{E}\Big(X\mathbf{1}_{A_n}(\omega)\Big)}{\mathrm{P}(A_n)}\cdot \mathrm{P}(A_n) + \sum_{n=1}^{\infty} c_n \mathrm{P}(A_n) \\ &= \sum_{\substack{n=1\\ \mathrm{P}(A_n)>0}}^{\infty} \mathrm{E}\Big(X\mathbf{1}_{A_n}(\omega)\Big)\end{aligned}$$

$$\begin{aligned}\mathrm{E}X &= \mathrm{E}\Big(X\sum_{n=1}^{\infty}\mathbf{1}_{A_n}(\omega)\Big) = \sum_{n=1}^{\infty}\mathrm{E}\Big(X\mathbf{1}_{A_n}(\omega)\Big) \\ &= \sum_{\substack{n=1\\ \mathrm{P}(A_n)>0}}^{\infty} \mathrm{E}\Big(X\mathbf{1}_{A_n}(\omega)\Big) + \sum_{\substack{n=1\\ \mathrm{P}(A_n)=0}}^{\infty} \mathrm{E}\Big(X\mathbf{1}_{A_n}(\omega)\Big) = \sum_{\substack{n=1\\ \mathrm{P}(A_n)>0}}^{\infty} \mathrm{P}\Big(X\mathbf{1}_{A_n}(\omega)\Big)\end{aligned}$$

性质 2.3.4. 线性性质与单调性:

(i) $\mathrm{E}[aX + bY|\mathcal{G}] = a\mathrm{E}[X|\mathcal{G}] + b\mathrm{E}[Y|\mathcal{G}]$;

(ii) 若 $X_1 \geq X_2$, 则 $\mathrm{E}[X_1|\mathcal{G}] \geq \mathrm{E}[X_2|\mathcal{G}]$; 若 $X \geq 0$, 则 $\mathrm{E}[X|\mathcal{G}] \geq 0$;$\big|\mathrm{E}[X|\mathcal{G}]\big| \leq \mathrm{E}[|X||\mathcal{G}]$.

定义 2.3.4. 称 n 维随机向量 $\mathbf{X}$ 与事件域 $\mathcal{G}$ 相互独立, 如果对于任意的 Borel 集 $A \in \mathcal{B}(\mathbb{R}^n)$ 和 $B \in \mathcal{G}$,

$$\mathrm{P}(\{\omega : \mathbf{X}(\omega) \in A\} \cap B) = \mathrm{P}(\{\omega : \mathbf{X}(\omega) \in A\})\mathrm{P}(B).$$

性质 2.3.5. 设 $\mathbf{X}, \mathbf{Y}$ 为 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机向量, 对于 Borel 函数 $h(\mathbf{x}, \mathbf{y})$ 满足 $\mathrm{E}|h(\mathbf{X}, \mathbf{Y})| < \infty$, $\mathcal{G} \subset \mathcal{F}$, 若 $\mathbf{Y}$ 为 $(\Omega, \mathcal{G})$ 上的随机向量, 则

i) $\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathcal{G}]=\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathcal{G}]\big|_{\mathbf{y}=\mathbf{Y}}$. 即先求出 $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathcal{G}]$ 是一个以 $\mathbf{y}$ 为参数的 $(\Omega,\mathcal{G})$ 上的随机变量 $Z(\mathbf{y},\omega)$, 则 $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathcal{G}]\big|_{\mathbf{y}=\mathbf{Y}}=Z(\mathbf{Y},\omega)$, 即 $\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathcal{G}](\omega)=Z(\mathbf{Y}(\omega),\omega)$.

ii) 进一步, 若 $\mathbf{X}$ 与 $\mathcal{G}$ 相互独立, 则 $\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathcal{G}]=\mathrm{E}\big(h(\mathbf{X},\mathbf{y})\big)\big|_{\mathbf{y}=\mathbf{Y}}$. 即先求出 $\mathrm{E}[h(\mathbf{X},\mathbf{y})|\mathcal{G}]$ 是一个以 $\mathbf{y}$ 为自变量的可测函数 $Z(\mathbf{y})$, $\mathrm{E}\big(h(\mathbf{X},\mathbf{y})\big)\big|_{\mathbf{y}=\mathbf{Y}}=Z(\mathbf{Y})$, 即 $\mathrm{E}[h(\mathbf{X},\mathbf{Y})|\mathcal{G}](\omega)=Z(\mathbf{Y}(\omega))$.

我们利用性质2.3.5可以得到以下的推论.

推论 2.3.1. 当 $h(\mathbf{x},\mathbf{y})$ 不依赖于 $\mathbf{x}$ 时, 记 $h(\mathbf{x},\mathbf{y})=h(\mathbf{y})$, 则
$$\mathrm{E}[h(\mathbf{Y})|\mathcal{G}]=\mathrm{E}(h(\mathbf{y})|\mathcal{G})\big|_{\mathbf{y}=\mathbf{Y}}=h(\mathbf{y})\Big|_{\mathbf{y}=\mathbf{Y}}=h(\mathbf{Y}).$$

推论 2.3.2. 若 $\mathbf{X}$ 与 $\mathcal{G}$ 相互独立. $\mathrm{E}|h(\mathbf{X})|<\infty$ 时, 记 $h(\mathbf{x},\mathbf{y})=h(\mathbf{x})$, 则
$$\mathrm{E}[h(\mathbf{X})|\mathcal{G}]=\mathrm{E}[h(\mathbf{x})]\Big|_{\mathbf{y}=\mathbf{Y}}=\mathrm{E}h(\mathbf{X}).$$

推论 2.3.3. 若 $\mathrm{E}|h(\mathbf{Y})|<\infty$, $\mathrm{E}|g(\mathbf{X})|<\infty$ 与 $\mathrm{E}|h(\mathbf{Y})g(\mathbf{X})|<\infty$, 则
$$\mathrm{E}[h(\mathbf{Y})g(\mathbf{X})|\mathcal{G}]=\mathrm{E}[h(\mathbf{y})g(\mathbf{X})|\mathcal{G}]\big|_{\mathbf{y}=\mathbf{Y}}=\big(h(\mathbf{y})\mathrm{E}[g(\mathbf{X})|\mathcal{G}]\big|_{\mathbf{y}=\mathbf{Y}}\big)=h(\mathbf{Y})\mathrm{E}[g(\mathbf{X})|\mathcal{G}].$$
进一步, 利用全期望公式
$$\mathrm{E}\big(h(\mathbf{Y})g(\mathbf{X})\big)=\mathrm{E}\big(h(\mathbf{Y})\mathrm{E}[g(\mathbf{X})|\mathcal{G}]\big).$$

性质 2.3.6. $X\in L^2(\Omega,\mathcal{F},\mathrm{P})$, $\mathcal{G}\subset\mathcal{F}$ 为子事件域, 则
$$\mathrm{E}\big(X-\mathrm{E}[X|\mathcal{G}]\big)^2=\inf_{Y\in L^2(\Omega,\mathcal{G},\mathrm{P})}\mathrm{E}(X-Y)^2.$$

性质 2.3.7. (平滑性、投影性) $\mathrm{E}|X|<\infty$, $\mathcal{G}_1\subset\mathcal{G}_2\subset\mathcal{F}$,
$$\mathrm{E}[X|\mathcal{G}_1]=\mathrm{E}\Big[\mathrm{E}[X|\mathcal{G}_2]\Big|\mathcal{G}_1\Big]=\mathrm{E}\Big[\mathrm{E}[X|\mathcal{G}_1]\Big|\mathcal{G}_2\Big].$$

性质 2.3.8. (Jensen 不等式) φ 为 $\mathbb{R}\to\mathbb{R}$ 上的凸函数 (φ 为有限凸). 若 X 及 $\varphi(X)$ 为随机变量, 满足 $\mathrm{E}|X|<\infty$, $\mathrm{E}|\varphi(X)|<\infty$, 则 $\varphi\big(\mathrm{E}[X|\mathcal{G}]\big)\le\mathrm{E}[\varphi(X)|\mathcal{G}]$. 特别地 $\varphi(\mathrm{E}X)\le\mathrm{E}\varphi(X)$.

例 2.3.2. 考虑概率空间 $(\Omega,\mathcal{F},\mathrm{P})$, 其中 $\Omega=[0,1]\times[0,1]$, $\mathcal{F}=[0,1]^2$ 上的 Borel 域, $\mathrm{P}=[0,1]^2$ 上的 Lebesgue 测度. f 为 $(\Omega,\mathcal{F},\mathrm{P})$ 一个随机变量, $\mathrm{E}|f|<\infty$. 设 $\mathcal{G}_1=(\Omega,\Phi,A_1,A_2)$, $\mathcal{G}_2=\sigma(\Omega,\Phi,A_{11},A_{12},A_{21},A_{22})$ 是包含 $A_{11},A_{12},A_{21},A_{22}$ 的最小 σ-代数, 其中

图 1

$$A_1=[0,1]\times[0,\frac{1}{2}),\ \ A_2=[0,1]\times[\frac{1}{2},1].$$
$$A_{11}=[0,\frac{1}{2})\times[0,\frac{1}{2}),\ \ A_{12}=[\frac{1}{2},1]\times[0,\frac{1}{2}),$$
$$A_{21}=[0,\frac{1}{2})\times[\frac{1}{2},1],\ \ A_{22}=[\frac{1}{2},1]\times[\frac{1}{2},1].$$

求 $\mathrm{E}[f|\mathcal{G}_1]$, $\mathrm{E}[f|\mathcal{G}_2]$.

解:

$$\begin{aligned}
\mathrm{E}[f|\mathcal{G}_1] &= a_1\mathbf{1}_{A_1}+a_2\mathbf{1}_{A_2}.\\
\mathrm{E}[f|\mathcal{G}_1](x,y) &= \frac{\iint_{A_1} f(u,v)\mathrm{d}uv}{\mathrm{P}(A_1)}\mathbf{1}_{A_1}(x,y)+\frac{\iint_{A_2} f(u,v)\mathrm{d}u\mathrm{d}v}{\mathrm{P}(A_2)}\mathbf{1}_{A_2}(x,y)\\
&= 2\Big(\iint_{A_1} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_1}(x,y)+\iint_{A_2} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_2}(x,y)\Big).\\
\mathrm{E}[f|\mathcal{G}_2](x,y) &= 4\Big(\iint_{A_{11}} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_{11}}(x,y)+\iint_{A_{12}} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_{12}}(x,y)\\
&\quad+\iint_{A_{21}} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_{21}}(x,y)+\iint_{A_{22}} f(u,v)\mathrm{d}u\mathrm{d}v\mathbf{1}_{A_{22}}(x,y)\Big).
\end{aligned}$$

□

注意到 $\frac{\iint_{A_{11}} f(u,v)\mathrm{d}u\mathrm{d}v}{\mathrm{P}(A_{11})}=4\iint_{A_{11}} f(u,v)\mathrm{d}u\mathrm{d}v$, 实质上就是 $f(x,y)$ 在 A_{11} 上的平均值. $\frac{\iint_{A_1} f(u,v)\mathrm{d}u\mathrm{d}v}{\mathrm{P}(A_1)}$ 是 $f(x,y)$ 在 A_1 上的平均值, 也是

$$\frac{\iint_{A_{11}} f(u,v)\mathrm{d}u\mathrm{d}v}{\mathrm{P}(A_{11})}\mathbf{1}_{A_{11}}(x,y)+\frac{\iint_{A_{12}} f(u,v)\mathrm{d}u\mathrm{d}v}{\mathrm{P}(A_{12})}\mathbf{1}_{A_{12}}(x,y)$$

在 A_1 上的平均值.

$\mathcal{G}_2\supset\mathcal{G}_1$, 可以认为 $\mathcal{G}_2$ 包含的信息比 $\mathcal{G}_1$ 多, 分辨率高, 因此 $\mathrm{E}[X|\mathcal{G}_2]$ 比 $\mathrm{E}[X|\mathcal{G}_1]$ 复杂,$\mathrm{E}[X|\mathcal{G}_1]$ 看起来比 $\mathrm{E}[X|\mathcal{G}_2]$ 平滑些, 所以性质2.2.8和性质2.3.7也叫平滑性质.

例 2.3.3. G 为 $[0,1)\times[0,1)$ 上的一个黑白图像. $G(x,y)$ 表示 (x,y) 点灰度大小, $G(x,y)$ 有界 Borel 函数. 构造概率空间 $\Omega=[0,1)\times[0,1)$, $\mathcal{F}=\mathcal{B}\big([0,1)\times[0,1)\big)$, $\mathrm{P}=m$, 即 $[0,1)\times[0,1)$ 上的 Lebesgue 测度, G 可以看成 $(\Omega,\mathcal{F},\mathrm{P})$ 上的一个随机变量.

$$\begin{aligned}
\mathcal{G}_0 &= (\Omega,\emptyset)\\
\mathcal{G}_1 &= \sigma\big(\Omega,\emptyset,[0,\tfrac{1}{2})\times[0,\tfrac{1}{2}),[0,\tfrac{1}{2})\times[\tfrac{1}{2},1),[\tfrac{1}{2},1)\times[0,\tfrac{1}{2}),[\tfrac{1}{2},1)\times[\tfrac{1}{2},1)\big)\\
&\vdots\\
\mathcal{G}_n &= \sigma\Big([\frac{k}{2^n},\frac{k+1}{2^n})\times[\frac{j}{2^n},\frac{j+1}{2^n})\ k=0,\cdots,2^n-1,\ j=0,\cdots,2^n-1\Big)
\end{aligned}$$

事实上, $[\frac{k}{2^n},\frac{k+1}{2^n})$ 就是一个像素, $\mathcal{G}_{n-1}\subset\mathcal{G}_n$. 在图像处理中, 实质上要找的是一个以"像素"为基本视觉单元的图像 $G_n(x,y)$, 使得该图像最真实的逼近 $G(x,y)$.

$$G_n(x,y)=\sum_{\substack{k=0,\cdots,2^n-1\\ j=0,\cdots,2^n-1}} a_{k,j}\mathbf{1}_{[\frac{k}{2^n},\frac{k+1}{2^n})\times[\frac{j}{2^n},\frac{j+1}{2^n})}(x,y)$$

$a_{k,j}$ 为像素在 $[\frac{k}{2^n},\frac{k+1}{2^n})\times[\frac{j}{2^n},\frac{j+1}{2^n})$ 的灰度. 即 $G_n(x,y)$ 为 $(\Omega,\mathcal{G}_n)$ 上的随机变量, 如果我们设想这种逼近使用函数间的 L^2 距离, 则

$$a_{k,j}=\frac{1}{\left(\frac{1}{2^n}\right)^2}\iint_{[\frac{k}{2^n},\frac{k+1}{2^n})\times[\frac{j}{2^n},\frac{j+1}{2^n})} G(x,y)\mathrm{d}x\mathrm{d}y$$

即为 $G(x,y)$ 在 $[\frac{k}{2^n},\frac{k+1}{2^n})\times[\frac{j}{2^n},\frac{j+1}{2^n})$ 上灰度的平均. 事实上, 由上例的计算可以知道

$$G_n(x,y)=\mathrm{E}[G(x,y)|\mathcal{G}_n]$$

也就是说图像 $G(x,y)$ 在像素集 $\mathcal{G}_n$ 上的最佳效果就是 $G_n(x,y)$, 即 $G(x,y)$ 在 $\mathcal{G}_n$ 上的条件数学期望. 随着 n 的增加, 分辨率不断加细, $G_n(x,y)$ 中恢复 $G(x,y)$ 的效果越好.

2.3.3 马氏过程理论中的一些应用

由第一章知识我们知道, 在研究随机过程时, 需要考虑一个带流的概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$ 以及一个 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机过程 $(X_t)_{t\in T}$.

定义 2.3.5. (适应性) 称随机过程 $(X_t)_{t\in T}$ 关于 $(\mathcal{F}_t)_{t\in T}$ 是适应的 (adapted), 如果对于任意的 $t\in T$, X_t 是 $(\Omega,\mathcal{F}_t)$ 上的随机变量, 即 $\forall a\in\mathbb{R}$,

$$\{\omega:\ X_t(\omega)\le a\}\in\mathcal{F}_t.$$

在考虑复杂的随机现象时, 往往可能有多个噪声源 $(\xi_t^{(1)})_{t\in T},(\xi_t^{(2)})_{t\in T},\cdots$, 而观测到的随机过程 X_t 往往是噪声源的函数 $F_t=(\xi_s^{(1)},\xi_s^{(2)},\cdots,s\le t)$(这点在随机微分方程中就很明显). 因此, 虽然我们看到的是 X_t, 但真正起作用的事件域流可能是 $\mathcal{F}_t=\sigma(\xi_s^{(1)},\xi_s^{(2)},s\le t)$.

♣♣♣20230316

定义 2.3.6. (马氏过程) 考虑一个带流的概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$, $(X_t)_{t\in T}$ 是关于 $(\mathcal{F}_t)_{t\in T}$ 适应的随机过程, 称 $(X_t)_{t\in T}$ 是关于 $(\mathcal{F}_t)_{t\in T}$ 的马氏过程, 如果对于任意有界 Borel 函数 f, 任意的 $t>s$,

$$\mathrm{E}[f(X_t)|\mathcal{F}_s]=\mathrm{E}[f(X_t)|\sigma(X_s)]\Big(=\mathrm{E}[f(X_t)|X_s]\Big).$$

定义 2.3.7. 称 $\mathcal{F}_t^X=\sigma(X_s,s\le t)$ 为 $(X_t)_{t\in T}$ 生成的自然事件域流.

$(X_t,Y_t)_{t\in T}$ 是关于 $(\mathcal{F}_t^{(X,Y)})_{t\in T}$ 的马氏过程, 但 $(X_t)_{t\in T}$ 不一定是关于 $(\mathcal{F}_t^X)_{t\in T}$ 的马氏过程.

定义 2.3.8. 设 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$ 是带流的概率空间, $(X_t)_{t\in T}$ 是状态空间为 $\mathbb{R}^n$ 关于 $(\mathcal{F}_t)_{t\in T}$ 的马氏过程. 可以证明对于任意的 $s<t$, 存在函数族 $\{P(s,x;t,A):x\in\mathbb{R}^n,A\in\mathcal{B}(\mathbb{R}^n)\}$ 满足

(1) 对于任意的 $A\in\mathcal{B}(\mathbb{R}^n)$, $P(s,x;t,A)$ 关于 x 是 Borel 函数;

(2) 对于任意的 $x\in\mathbb{R}^n$, $P(s,x;t,A)$ 是 $\mathcal{B}(\mathbb{R}^n)$ 上的概率测度;

(3) 对于 $\mathbb{R}^n$ 上任意有界 Borel 函数 f,

$$\mathrm{E}[f(X_t)|\mathcal{F}_s]=\mathrm{E}[f(X_t)|X_s]=\int f(y)P(s,X_s;t,\mathrm{d}y). \tag{2.3.3}$$

称 $\big(P(s,x;t,A):s<t,x\in\mathbb{R}^n,A\in\mathcal{B}(\mathbb{R}^n)\big)$ 为马氏过程 $(X_t)_{t\in T}$ 的转移概率族.

当对于任意给定的 $x\in\mathbb{R}^n$ 和 $A\in\mathcal{B}(\mathbb{R}^n)$, $P(s,x;t,A)$ 只与 $t-s$ 有关, 则称 $(X_t)_{t\in T}$ 为时间齐次马氏过程 (简称时齐马氏过程), 记 $P(s,x;t,A)$ 为 $P(t-s,x,A)$.

当对于任意的 $s,t\in T$ 和 $x\in\mathbb{R}^n$, 若概率测度 $P(s,x;t,A)$ 有密度函数 $p(s,x;t,y)$, 则称 $\big(p(s,x;t,y):s<t,x,y\in\mathbb{R}^n\big)$ 为马氏过程 $(X_t)_{t\in T}$ 的转移概率密度族, 当 $(X_t)_{t\in T}$ 为时齐马氏过程时记为 $P(t-s,x,y)$.

性质 2.3.9. (Chapman-Kolmogorov 方程) 对于任意有界 Borel 函数 f, $s<r<t$,

$$\int_{\mathbb{R}^n}f(y)P(s,x;t,\mathrm{d}y)=\int_{\mathbb{R}^n}\Big(\int_{\mathbb{R}^n}f(y)P(r,z;t,\mathrm{d}y)\Big)P(s,x;r,\mathrm{d}z).$$

特别地, 若 $A\in\mathcal{B}(\mathbb{R}^n)$ 时,

$$P(s,x;t,A)=\int_{\mathbb{R}^n}P(r,z;t,A)P(s,x;r,\mathrm{d}z).\text{[①]}$$

[①] 有些教科书上, 也将以上两式写成:

$$\int_{\mathbb{R}^n}f(y)P(s,x;t,\mathrm{d}y)=\int_{\mathbb{R}^n}P(s,x;r,\mathrm{d}z)\int_{\mathbb{R}^n}f(y)P(r,z;t,\mathrm{d}y).$$

$$P(s,x;t,A)=\int_{\mathbb{R}^n}P(s,x;r,\mathrm{d}z)P(r,z;t,A).$$

证明: 利用马氏性、条件期望的投影性质2.3.7和 (2.3.3) 得

$$\begin{aligned}\int_{\mathbb{R}^n} f(y)P(s,X_s;t,\mathrm{d}y) &= \mathrm{E}[f(X_t)|X_s] \\ &= \mathrm{E}[f(X_t)|\mathcal{F}_s] \\ &= \mathrm{E}[\mathrm{E}\big[f(X_t)|\mathcal{F}_r\big]\big|\mathcal{F}_s\big] \\ &= \mathrm{E}[\mathrm{E}\big[f(X_t)|X_r\big]\big|X_s\big] \\ &= \int_{\mathbb{R}^n}\Big(\int_{\mathbb{R}^n} f(y)P(r,z;t,\mathrm{d}y)\Big)P(s,X_s;r,\mathrm{d}z).\end{aligned}$$

□

例 2.3.4. 设 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathrm{P})$ 是带流的概率空间, $(X_t)_{t\geq 0}$ 是状态空间为 $\mathbb{R}^n$ 关于 $(\mathcal{F}_t)_{t\geq 0}$ 的马氏过程, 而且其转移概率族满足空间平移不变性, 即任意的 $s<t$, $x,y\in\mathbb{R}^n$ 和 $A\in\mathcal{B}(\mathbb{R}^n)$,

$$P(s,x;t,A)=P(s,x+y;t,A+y),$$

其中 $A+y=\{z|z=x+y,x\in A\}$. 证明: $(X_t)_{t\geq 0}$ 是独立增量过程, 即任意的 $s<t$, X_t-X_s 与 $\mathcal{F}_s$ 独立.

证明: 首先证明对于任意的 $A\in\mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned}\mathrm{E}[\mathbf{1}_A(X_t-X_s)|\mathcal{F}_s] &= \mathrm{E}[\mathbf{1}_A(X_t-z)|\mathcal{F}_s]\Big|_{z=X_s} \\ &= \mathrm{E}[\mathbf{1}_{A+z}(X_t)|\mathcal{F}_s]\Big|_{z=X_s} \\ &= \int \mathbf{1}_{A+z}(y)P(s,X_s,t,\mathrm{d}y)\Big|_{z=X_s} \\ &= P(s,X_s;t,A+z)\Big|_{z=X_s} \\ &= P(s,X_s;t,A+X_s) \\ &= P(s,0;t,A).\end{aligned}$$

其次, 证明 $\mathrm{P}(X_t-X_s\in A)=P(s,0;t,A)$.

$$\begin{aligned}\mathrm{P}(X_t-X_s\in A) &= \mathrm{E}\big(\mathrm{E}[\mathbf{1}_A(X_t-X_s)|\mathcal{F}_s]\big) \\ &= \mathrm{E}\big(P(s,0;t,A)\big) \\ &= P(s,0;t,A).\end{aligned}$$

最后, 证明对于任意的 $A\in\mathcal{B}(\mathbb{R}^n)$ 和 $B\in\mathcal{F}_s$, $\mathrm{E}\big(\mathbf{1}_A(X_t-X_s)\mathbf{1}_B(\omega)\big)=\mathrm{P}(X_t-X_s\in A)\mathrm{P}(B)$ 即可.

$$\begin{aligned}\mathrm{E}\big(\mathbf{1}_A(X_t-X_s)\mathbf{1}_B(\omega)\big) &= \mathrm{E}\big(\mathrm{E}[\mathbf{1}_A(X_t-X_s)\mathbf{1}_B(\omega)|\mathcal{F}_s]\big) \\ &= \mathrm{E}\big(\mathbf{1}_B(\omega)\mathrm{E}[\mathbf{1}_A(X_t-X_s)|\mathcal{F}_s]\big) \\ &= \mathrm{E}\big(\mathbf{1}_B(\omega)P(s,0;t,A)\big) \\ &= \mathrm{P}(X_t-X_s\in A)\mathrm{P}(B).\end{aligned}$$

□

第三章　鞅论浅说

3.1　鞅的定义和例子

首先来解释一下“鞅 (martingale)”这个词. Martingale 是一个法语词. 它有两个释义: (一) 古代的马拉车时安在马脖子上的皮套子 (马的璎珞); (二) 公平赌博, 原意是所谓倍赌策略. 关于“martingale”这一概念的起源, 可以参看文献[①,②].

在上一章中我们考虑了如下博弈的例子. $(\eta_n)_{n\geq 1}$ 为独立同分布的随机变量序列, 满足 $\mathrm{P}(\eta_n=1)=p$, $\mathrm{P}(\eta_n=-1)=1-p=q$. 设

$$\xi_n=\xi_0+\sum_{k=1}^{n} b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\cdot\eta_k, \tag{3.1.1}$$

则 $\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_n+b_{n+1}(\xi_0,\eta_1,\cdots,\eta_n)\mathrm{E}(\eta_{n+1})$.

当 $p=q=\frac{1}{2}$ 时, $\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]=\xi_n$. 这说明“当博弈是公平时”($p=q=\frac{1}{2}$) 无论取什么样的博弈“策略”, ξ_{n+1} 的最佳预测函数是 ξ_n, 即

$$\mathrm{E}[\xi_{n+1}|\xi_0,\eta_1,\cdots,\eta_n]-\xi_n=0,$$

也就是说, 在任意的时刻 n 时, 预测时刻 $n+1$ 时的“输赢情况”, 不论是取什么样的博弈“策略”, 都不可能得到任何关于“输赢情况”的信息. 任何具有趋势性的结论都没有预测到, 也可以理解为, 这个博弈的“不确定性”是“纯随机”. 而这正是“公平的博弈 (fair game)”所要求的. 因此有以下定义:

定义 3.1.1. (鞅) 设 $(X_t)_{t\in T}$ 为 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$ 上的关于 $(\mathcal{F}_t)_{t\in T}$ 适应的随机过程, 若对于 $\forall s<t$, $\mathrm{E}[X_t|\mathcal{F}_s]=X_s$, 则称 $(X_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅 (martingale), 或称 $(X_t,\mathcal{F}_t)_{t\in T}$ 为鞅, 有时在明确了 $(\mathcal{F}_t)_{t\in T}$ 之后, 可以直接称 $(X_t)_{t\in T}$ 为鞅.

若 $\mathrm{E}[X_t|\mathcal{F}_s]\geq X_s$, 则称 $(X_t,\mathcal{F}_t)_{t\in T}$ 为下鞅 (submartingale); 若 $\mathrm{E}[X_t|\mathcal{F}_s]\leq X_s$, 则称 $(X_t,\mathcal{F}_t)_{t\in T}$ 为上鞅 (subermartingale).

若 $T=\{0,1,2,3,\cdots\}$ 时, $(X_n)_{n\geq 0}$ 称为 (下、上) 鞅列, $(X_{n+1}-X_n)_{n\geq 0}$ 称为 (下、上) 鞅差序列[③].

性质 3.1.1. 设 $(X_n)_{n\geq 0}$ 是 (下、上) 鞅列, 也就是说对任意的 n 和 m,

$$\mathrm{E}[\xi_{n+m}|\mathcal{F}_n]=(\geq,\leq)\,\xi_n, \tag{3.1.2}$$

则 (3.1.2) 等价于对任意的 n,

$$\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]=(\geq,\leq)\,\xi_n.$$

这个性质的证明留作习题.

例如在 (3.1.1) 中, 取 $\mathcal{F}_0=\sigma(\xi_0)$, $\mathcal{F}_n=\sigma(\xi_0,\eta_1,\cdots,\eta_n)$.

[①] R. Mansuy. *The origins of the word “Martinage”*. Electronic Journal of History of Probability and Statistics. vol.5. no.1 Juin/June 2009 www.jehps.net

[②] L. Bienvenu, G. Shafer, A. Shen. *On the history of martingales in the study of randomness.* Electronic Journal of History of Probability and Statistics. vol.5. no.1 Juin/June 2009 www.jehps.net

[③] 在本书中, 对于离散时间的随机过程, 有时会根据实际例子的设定, T 可以取为 $\{1,2,\cdots\}$.

1. 若 $p=q=\frac{1}{2}$ 时, $\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]=\xi_n$, 则 $(\xi_n)_{n\ge 1}$ 为 $(\mathcal{F}_n)_{n\ge 1}$ 鞅 (列).
2. 若 $p>q$ 时, $\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]=\xi_n+b_n(\xi_0,\eta_1,\cdots,\eta_n)(p-q)\ge\xi_n$, 则 $(\xi_n)_{n\ge 1}$ 为 $(\mathcal{F}_n)_{n\ge 1}$ 下鞅 (列).
3. 若 $p<q$ 时, $\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]\le\xi_n$, 则 $(\xi_n)_{n\ge 1}$ 为 $(\mathcal{F}_n)_{n\ge 1}$ 上鞅 (列).

下面来看更多鞅或鞅列的例子.

例 3.1.1. (与独立随机变量序列相关的鞅)

设 $(\eta_n)_{n\ge 1}$ 相互独立, 且 $\forall n\ge 1$, $\mathrm{E}|\eta_n|<\infty$, $\mathcal{F}_n=\sigma(\eta_1,\cdots,\eta_n)$.

(1) 设 $X_n=\eta_1+\cdots+\eta_n$, $\xi_n=X_n-\sum_{k=1}^n\mathrm{E}\eta_k$, 则 $(\xi_n)_{n\ge 1}$ 是关于 $(\mathcal{F}_n)_{n\ge 1}$ 的鞅列. (验证留作习题)

(2) 若对于任意的 $n\ge 1$, $\mathrm{E}\eta_n^2<\infty$, 设 $\zeta_n=\xi_n^2-\sum_{k=1}^n\mathrm{var}\eta_k$, 则 $(\zeta_n)_{n\ge 1}$ 是关于 $(\mathcal{F}_n)_{n\ge 1}$ 的鞅列.

证明: 显然 $(\zeta_n)_{n\ge 1}$ 为 $(\mathcal{F}_n)_{n\ge 1}$ 适应. 设 $\widetilde{\eta}_n=\eta_n-\mathrm{E}\eta_n$, 则 $\mathrm{E}\widetilde{\eta}_n=0, \mathrm{var}(\eta_n)=\mathrm{E}\widetilde{\eta}_n^2$, $\xi_n=\sum_{k=1}^n\widetilde{\eta}_k$.

$$\begin{aligned}\mathrm{E}[\zeta_{n+1}|\mathcal{F}_n] &= \mathrm{E}\big[\big(\xi_n+\widetilde{\eta}_{n+1}\big)^2-\sum_{k=1}^{n+1}\mathrm{E}\widetilde{\eta}_k^2\big|\mathcal{F}_n\big]\\ &= \mathrm{E}[\xi_n^2|\mathcal{F}_n]+2\mathrm{E}[\xi_n\widetilde{\eta}_{n+1}|\mathcal{F}_n]+\mathrm{E}[\widetilde{\eta}_{n+1}^2|\mathcal{F}_n]-\sum_{k=1}^{n+1}\mathrm{E}\widetilde{\eta}_k^2\\ &= \xi_n^2+2\xi_n\mathrm{E}(\widetilde{\eta}_{n+1})+\mathrm{E}\widetilde{\eta}_{n+1}^2-\sum_{k=1}^{n+1}\mathrm{E}\widetilde{\eta}_k^2\\ &= \xi_n^2-\sum_{k=1}^{n}\mathrm{E}\widetilde{\eta}_k^2\\ &= \zeta_n.\end{aligned}$$

□

(3) 若对于某个实数 s, $\mathrm{E}e^{s\eta_n}<\infty$, $n\ge 1$, 设 $\phi_n=\prod_{k=1}^n\frac{e^{s\eta_k}}{\mathrm{E}e^{s\eta_k}}$, 则 $(\phi_n)_{n\ge 1}$ 是关于 $(\mathcal{F}_n)_{n\ge 1}$ 的鞅列. (验证留作习题)

定义 3.1.2. 对于 $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\ge 0},\mathrm{P})$, $(A_n)_{n\ge 0}$ 为 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量序列, 若 $\forall n\ge 1$, 满足 A_n 是关于 $\mathcal{F}_{n-1}$ 可测的随机变量, 则称 $(A_n)_{n\ge 1}$ 为可料随机序列 (predictable).

例 3.1.2. 设 $(\xi_n)_{n\ge 0}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 鞅列, $(V_n)_{n\ge 1}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 可料随机序列 $V_0\in\mathcal{F}_0$. 定义

$$\phi_n=V_0\xi_0+\sum_{k=1}^n V_k(\xi_k-\xi_{k-1}),\ n\ge 1,\quad \phi_0=V_0\xi_0. \tag{3.1.3}$$

若对任意的 $n\ge 0$, $\mathrm{E}|\phi_n|<\infty$, 则 $(\phi_n)_{n\ge 0}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 鞅列. 在一些教科书上将 (3.1.3) 式称为鞅变换.[①]

若 $(\xi_n)_{n\ge 0}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 下鞅列, $(V_n)_{n\ge 1}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 非负可料随机序列 $V_0\in\mathcal{F}_0$, 对任意的 $n\ge 0$, $\mathrm{E}|\phi_n|<\infty$, 则 (3.1.3) 中的 $(\phi_n)_{n\ge 0}$ 是 $(\mathcal{F}_n)_{n\ge 0}$ 下鞅列. (验证留作习题)

基本模型 (3.1.1) 是 (3.1.3) 的特例.

例 3.1.3. 考虑一个带流的 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$. 若 ξ 满足 $\mathrm{E}|\xi|<\infty$, 令 $\xi_t=\mathrm{E}[\xi|\mathcal{F}_t]$, 则 $(\xi_t,\mathcal{F}_t)_{t\in T}$ 为鞅.

证明: 对于任意的 $t_2\ge t_1$, $\mathcal{F}_{t_2}\supset\mathcal{F}_{t_1}$, 则

$$\mathrm{E}[\xi_{t_2}|\mathcal{F}_{t_1}]=\mathrm{E}\big[\mathrm{E}[\xi|\mathcal{F}_{t_2}]\big|\mathcal{F}_{t_1}\big]=\mathrm{E}[\xi|\mathcal{F}_{t_1}]=\xi_{t_1}.$$

□

[①] 本质上, 这是离散时间情形下, 随机可料序列关于鞅的随机积分.

例 3.1.4. $(\xi_n)_{n\geq 0}$ 关于 $(\mathcal{F}_n)_{n\geq 0}$ 适应. 设 $\zeta_0=\xi_0$, 对于 $n\geq 1$, 定义

$$\zeta_n=\xi_n-\sum_{k=0}^{n-1}\Big(\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k\Big),$$

则 $(\zeta_n,\mathcal{F}_n)_{n\geq 0}$ 为鞅.

分析:

(1) $\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k$ 是第 $k+1$ 步 "比鞅多出的部分" 而且是关于 $\mathcal{F}_k$ 可测的随机变量;

(2) $\sum\limits_{k=0}^{n-1}\Big(\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k\Big)$ 前 n 步 "比鞅多出的部分", 而且它是关于 $\mathcal{F}_{n-1}$ 可测的随机变量;

(3) ξ_n 除去前面所有不是鞅的部分, 即 $\xi_n-\sum\limits_{k=0}^{n-1}\Big(\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k\Big)$, 则剩余部分为鞅.

证明:

(1) 由上面的分析知, ζ_n 为关于 $\mathcal{F}_n$ 可测的随机变量.

(2) 注意到, 当 $k\leq n$ 时, $\mathcal{F}_k\subset\mathcal{F}_n$, 则

$$\begin{aligned}
&\mathrm{E}[\zeta_{n+1}|\mathcal{F}_n]\\
=\;&\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]-\sum_{k=0}^{n}\Big(\mathrm{E}\Big[\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]\Big|\mathcal{F}_n\Big]-\mathrm{E}[\xi_k|\mathcal{F}_n]\Big)\\
=\;&\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]-\Big(\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]-\xi_n+\sum_{k=0}^{n-1}\big(\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k\big)\Big)\\
=\;&\xi_n-\sum_{k=0}^{n-1}\big(\mathrm{E}[\xi_{k+1}|\mathcal{F}_k]-\xi_k\big)\\
=\;&\zeta_n.
\end{aligned}$$

□

例 3.1.5. (倍赌策略, martingale betting strategy)

在我们的基本模型 (3.1.1) 中设 $\xi_0=0$, $p=q=\frac{1}{2}$. 赌徒的策略是首次获胜, 不再下注, 立即停止博弈, 而在首次获胜之前每一次的赌注相应地在上一次赌注的基础上翻一倍, 这个博弈策略具体的数学模型可以写成:

$$b_1(\xi_0)=1,$$

$$b_2(\xi_0,-1)=2,\ b_2(\xi_0,1)=0,$$

$$b_3(\xi_0,-1,-1)=4,\ b_3(\xi_0,-1,1)=0,\ b_3(\xi_0,1,-1)=0,\ b_3(\xi_0,1,1)=0,\cdots$$

$$b_n(\xi_0,\underbrace{-1,\cdots,-1}_{n-1\text{次}})=2^{(n-1)},\ b_n(\xi_0,\text{其它})=0\ .$$

则 $(\xi_n)_{n\geq 1}$ 是鞅.

另一方面, 记 $A_n=\{\overbrace{\eta_1=-1,\cdots,\eta_{n-1}=-1}^{\text{前}n-1\text{次均为}-1},\eta_n=1\}$, 也就是说 A_n 表示在 n 时刻, 赌徒首次获胜这一事件, 则

$$\begin{aligned}
&\mathrm{P}\Big(\xi_0=0,\xi_1=-1,\cdots,\xi_{n-1}=-\sum_{k=1}^{n-1}2^{k-1},\xi_n=\xi_{n-1}+2^{n-1}\Big|\xi_0=0,A_n\Big)\\
=\;&\mathrm{P}\big(\xi_n=1\big|\xi_0=0,A_n\big)\\
=\;&1.
\end{aligned}$$

而且, 如果在 n 时刻获胜, 则之后不再下注博弈, 可以认为 $\xi_{n+k}=1, k\geq 0$, 也就是说

$$\mathrm{P}\big(\xi_n=1,\xi_{n+1}=1,\cdots,\xi_{n+k}=1\big|\xi_0=0,A_n\big)=1.$$

由对称简单随机游动的常返性, 我们知道 $\mathrm{P}(\cup_{n=1}^{\infty}A_n)=1$. 这就说明, 即使在公平博弈的情况下, 也存在着所谓的"倍赌策略"保证赌徒最终只赢不输.

一个值得思考的问题是, 这种"最终只赢不输"的倍赌策略是不是和鞅所刻画的博弈的"公平性"相矛盾?

♣♣♣20230321

例 3.1.6. (随机利率)

先看确定性离散时间的模型. 设银行的利率为 r, 第 n 天时, 张三在银行中的存款为 X_n 时, 在第 $n+1$ 天时存款就为 $(1+r)X_n$.

$$X_{n+1}=(1+r)X_n,\ X_{n+1}-X_n=rX_n.$$

所谓利率就是在单位时间内, 单位存款 (货币) 的利息. 若知道在第 n 天时, 张三的银行账户有存款 X_n, 则开始时张三存入的钱 X_0 为多少?

$$X_n=(1+r)^nX_0,\ X_0=X_n(1+r)^{-n}.$$

而对于连续时间模型, 张三时刻 t 时在银行中的存款是 X_t, 则在单位时间内, 单位存款的利息就是 $\lim\limits_{\Delta t\to 0}\frac{X_{t+\Delta t}-X_t}{\Delta tX_t}=\sigma$, 即

$$\frac{\mathrm{d}X_t}{\mathrm{d}t}=\sigma X_t,\quad X_t=X_0\mathrm{e}^{\sigma t}.$$

若 $\sigma=\ln(1+r)$, $X_{n+1}=\mathrm{e}^{\sigma}X_n$ 称为连续利率增长关系.

设 $\mathcal{F}_n=\sigma(\eta_0,\eta_1,\cdots,\eta_n)$, ξ_n 为关于 $(\Omega,\mathcal{F}_n)$ 的非负随机变量 (也就是说$(\xi_n)_{n\geq 0}$ 为 $(\mathcal{F}_n)_{n\geq 0}$ 适应), 随机利率序列 $(\sigma_n)_{n\geq 0}$ 也是关于 $(\mathcal{F}_n)_{n\geq 0}$ 适应. 若资金 ξ_n 满足条件连续利率增长关系

$$\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]=\mathrm{e}^{\sigma_n}\xi_n,$$

将 n 时刻的资金 ξ_n 在开始时刻的折现价值记为 ζ_n, 即

$$\zeta_n=\mathrm{e}^{-(\sigma_0+\sigma_1+\cdots+\sigma_{n-1})}\xi_n,$$

则 $(\zeta_n,\mathcal{F}_n)$ 为鞅列.

$$\begin{aligned}
\mathrm{E}[\zeta_{n+1}|\mathcal{F}_n] &= \mathrm{E}[\mathrm{e}^{-(\sigma_0+\sigma_1+\cdots+\sigma_n)}\xi_{n+1}|\mathcal{F}_n]\\
&= \mathrm{e}^{-(\sigma_0+\sigma_1+\cdots+\sigma_n)}\mathrm{E}[\xi_{n+1}|\mathcal{F}_n]\\
&= \mathrm{e}^{-(\sigma_0+\sigma_1+\cdots+\sigma_n)}\mathrm{e}^{\sigma_n}\xi_n\\
&= \mathrm{e}^{-(\sigma_0+\sigma_1+\cdots+\sigma_{n-1})}\xi_n=\zeta_n.
\end{aligned}$$

例 3.1.7. (分支过程) 第 n 代第 k 个细胞分裂的个数 $\eta_{n,k}$, 对于任意的 n 和 k, $\eta_{n,k}$ 独立同分布. 设 $\mathrm{E}\eta_{n,k}=\mu<\infty$.

$$\begin{aligned}
\xi_0 &= 1\\
\xi_n &= \eta_{n-1,1}+\cdots+\eta_{n-1,\xi_{n-1}}
\end{aligned}$$

则 $\left(\frac{\xi_n}{\mu^n}\right)_{n\geq 1}$ 是 $\big(\mathcal{F}_n=\sigma(\xi_1,\cdots,\xi_n)\big)_{n\geq 1}$ 鞅.

证明: $(\eta_{n,k})_{n\geq 1,k\geq 1}$ 相互独立, $\xi_1,\cdots,\xi_n$ 是由 $(\eta_{l,k})_{l\leq n-1,k\geq 1}$ 决定的, 则 $(\eta_{n,k})_{k\geq 1}$ 与 $(\xi_1,\cdots,\xi_n)$ 独立,

也就是说第 n 代中每个细胞的分裂数与第 1 代到第 n 代的细胞数独立. 由第二章性质2.2.6,

$$\begin{aligned}
\mathrm{E}\Big[\frac{\xi_{n+1}}{\mu^{n+1}}\Big|\mathcal{F}_n\Big] &= \frac{1}{\mu^{n+1}}\mathrm{E}[\eta_{n,1}+\cdots+\eta_{n,\xi_n}|\mathcal{F}_n] \\
&= \frac{1}{\mu^{n+1}}\mathrm{E}\big(\eta_{n,1}+\cdots+\eta_{n,k}\big)\big|_{k=\xi_n} \\
&= \frac{1}{\mu^{n+1}}\big(k\mu\big)\big|_{k=\xi_n} \\
&= \frac{\xi_n\mu}{\mu^{n+1}} \\
&= \frac{\xi_n}{\mu^n}.
\end{aligned}$$

□

例 3.1.8. (Polya罐子模型, Polya's urn) 假设罐子中开始时有一个红球, 一个绿球. 每次随机地从罐子中摸出一个球, 若是红球, 则将此红球放回的同时再往罐子中放入一个红球; 类似地, 若摸出的是绿球, 则将此绿球放回的同时再往罐子中放入一个绿球. 设 X_n 表示第 n 次摸球后, 罐子中红球的个数, $X_0=1$. 可以证明 (X_n) 是一个马氏过程, 其转移概率是

$$\mathrm{P}\big(X_{n+1}=k+1\big|X_n=k\big)=\frac{k}{n+2},\ \mathrm{P}\big(X_{n+1}=k\big|X_n=k\big)=\frac{n+2-k}{n+2}. \tag{3.1.4}$$

设 $M_n=\frac{X_n}{n+2}$ 表示第 n 次摸球后罐子中红球的比例, 则 $\big(M_n,\mathcal{F}_n=\sigma(X_1,\cdots,X_n\}\big)$ 是鞅.

证明: 利用 (3.1.4) 容易计算,

$$\mathrm{E}[X_{n+1}|X_n]=\mathrm{E}(X_{n+1}|X_n=k)\Big|_{k=X_n}=\big(k+\frac{k}{n+2}\big)\Big|_{k=X_n}=X_n+\frac{X_n}{n+2}.$$

再由马氏性,

$$\begin{aligned}
\mathrm{E}[M_{n+1}|\mathcal{F}_n] &= \mathrm{E}[M_{n+1}|X_n] \\
&= \mathrm{E}\Big[\frac{X_{n+1}}{n+3}\Big|X_n\Big] \\
&= \frac{1}{n+3}\Big(X_n+\frac{X_n}{n+2}\Big) \\
&= \frac{X_n}{n+2}.
\end{aligned}$$

□

例 3.1.9. (似然比序列) 设 (ξ_n) 为独立同分布随机变量序列, 其分布密度为 f. g 为另一个分布密度. 设

$$\eta_n=\prod_{k=1}^n\frac{g(\xi_k)}{f(\xi_k)},$$

则 $(\eta_n)_{n\ge1}$ 为 $\big(\mathcal{F}_n=\sigma(\xi_1,\cdots,\xi_n)\big)_{n\ge1}$ 鞅列.

证明:

$$\mathrm{E}\Big(\frac{g(\xi_k)}{f(\xi_k)}\Big)=\int\frac{g(y)}{f(y)}f(y)dy=1,$$

$$\mathrm{E}[\eta_n|\mathcal{F}_{n-1}]=\eta_{n-1}\mathrm{E}\Big[\frac{g(\xi_n)}{f(\xi_n)}\Big|\mathcal{F}_{n-1}\Big]=\eta_{n-1}\mathrm{E}\Big(\frac{g(\xi_n)}{f(\xi_n)}\Big)=\eta_{n-1}.$$

□

例 3.1.10. 设时齐 Markov 链 (ξ_n) 的转移矩阵为 $P=(p_{ij})$. 对于状态空间上的有界函数 f, 定义映射 $\mathbf{P}f$,

$$(\mathbf{P}f)(i)=\sum_j p_{ij}f(j).$$

$\mathbf{P}f$ 仍是状态空间上的有界函数. 对于任意的有界函数 f, 设

$$X_n = f(\xi_n) - f(\xi_0) - \sum_{k=0}^{n-1} \big((\mathbf{P}-\mathbf{I})f\big)(\xi_k), \tag{3.1.5}$$

其中 $\mathbf{I}$ 为恒同映射, 即对于任意 f, $\mathbf{I}f = f$, 那么 $(X_n)_{n\geq 1}$ 是 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)_{n\geq 1}$ 鞅.

特别地, 当 f 满足方程 $\mathbf{P}f = f$ 时, 即 f 是关于 $\mathbf{P}-\mathbf{I}$ 的调和函数时, $f(\xi_n) - f(\xi_0)$ 是鞅.

证明: 首先我们求 $\mathrm{E}[f(\xi_{n+1})|\mathcal{F}_n]$, 由马氏性

$$\begin{aligned}
\mathrm{E}[f(\xi_{n+1})|\mathcal{F}_n] &= \mathrm{E}[f(\xi_{n+1})|\xi_n] \\
&= \mathrm{E}(f(\xi_{n+1})|\xi_n = i)\big|_{i=\xi_n} \\
&= \sum_j p_{ij} f(j)\big|_{i=\xi_n} \\
&= (\mathbf{P}f)(\xi_n).
\end{aligned}$$

因此,

$$\begin{aligned}
\mathrm{E}[X_{n+1}|\mathcal{F}_n] &= \mathrm{E}[f(\xi_{n+1})|\mathcal{F}_n] - f(\xi_0) - \sum_{k=0}^{n} \big((\mathbf{P}-\mathbf{I})f\big)(\xi_k) \\
&= (\mathbf{P}f)(\xi_n) - f(\xi_0) - \sum_{k=0}^{n} \big((\mathbf{P}-\mathbf{I})f\big)(\xi_k) \\
&= f(\xi_n) - f(\xi_0) - \sum_{k=0}^{n-1} \big((\mathbf{P}-\mathbf{I})f\big)(\xi_k).
\end{aligned}$$

□

3.2 Doob 下鞅列分解定理

再次回到基本模型 (3.1.1). 此时, 我们对模型略作些调整, $(\eta_n)_{n\geq 1}$ 为独立同分布的随机变量序列, $\mathrm{P}(\eta_n = 1) = p$, $\mathrm{P}(\eta_n = -1) = 1 - p = q$.

$$\xi_n = \xi_0 + \sum_{k=1}^{n} b_k(\xi_0, \xi_1, \cdots, \xi_{k-1}) \cdot \eta_k. \tag{3.2.1}$$

也就是说, 这个模型中每一步的 "投资策略" b_k 是依赖前面 $k-1$ 次的本金 $(\xi_0, \xi_1, \cdots, \xi_{k-1})$ 的[①].

由上面一节基本模型 (3.1.1) 的推导, 我们可以取 η_k 相互独立, 且 $\mathrm{E}\eta_k = \mu$ 来考虑更广的模型

$$\xi_n = \xi_0 + \sum_{k=1}^{n} b_k(\xi_0, \xi_1, \cdots, \xi_{k-1}) \cdot \eta_k, \tag{3.2.2}$$

$$\mathrm{E}[\xi_{n+1}|\xi_0, \eta_1, \cdots, \eta_n] = \xi_n + b_{n+1}(\xi_0, \xi_1, \cdots, \xi_n)\mathrm{E}(\eta_{n+1}).$$

当 $\mathrm{E}(\eta_{n+1}) = \mu > 0$, 从直观上, 我们知道这个博弈对我们有利, 也就是说"排除不确定因素", 或者说"排除随机性", 或者"排除风险"后, 每次都要赢. 在第 $n+1$ 次博弈之前, 根据前面 n 次已知的输赢结果进行下注, 我们就知道"排除随机性"后, 第 $n+1$ 次获利 $b_{n+1}(\xi_0, \cdots, \xi_n)\mathrm{E}(\eta_{n+1})$.

这就是 Doob 下鞅分解定理所表达的直观含义. 更具体些 $(\xi_n)_{n\geq 1}$ 可以做如下分解:

$$\begin{aligned}
\xi_{n+1} &= \xi_0 + \sum_{k=1}^{n+1} b_k(\xi_0, \cdots, \xi_{k-1}) \cdot \big(\eta_k - \mathrm{E}(\eta_k)\big) + \sum_{k=1}^{n+1} b_k(\xi_0, \cdots, \xi_{k-1}) \cdot \mathrm{E}(\eta_k) \\
&= M_{n+1} + A_{n+1}.
\end{aligned}$$

[①] 因为 $\xi_1 = \xi_0 + b_1(\xi_0)\eta_1$, 所以, ξ_1 为 $\sigma(\xi_0, \eta_1)$ 可测的随机变量, 那么, 对于第二步 "投资策略" $b_2(\xi_0, \xi_1)$ 就应该也是 $\sigma(\xi_0, \eta_1)$ 可测的随机变量. 进一步

$$\xi_2 = \xi_0 + b_1(\xi_0)\eta_1 + b_2(\xi_0, \xi_1)\eta_2 = \xi_1 + b_2(\xi_0, \xi_1)\eta_2,$$

则 ξ_1 是 $\sigma(\xi_0, \eta_1, \eta_2)$ 可测的随机变量. 以此类推, $b_n(\xi_0, \xi_1, \cdots, \xi_{n-1})$ 是 $\sigma(\xi_0, \eta_1, \cdots, \eta_{n-1})$ 可测的随机变量, ξ_n 是 $\sigma(\xi_0, \eta_1, \cdots, \eta_n)$ 可测的随机变量.

其中

$$M_{n+1} = \xi_0 + \sum_{k=1}^{n+1} b_k(\xi_0, \cdots, \xi_{k-1}) \cdot \big(\eta_k - \mathrm{E}(\eta_k)\big),$$

$$A_{n+1} = \sum_{k=1}^{n+1} b_k(\xi_0, \cdots, \xi_{k-1}) \cdot \mathrm{E}(\eta_k).$$

容易知道 $(M_n, \mathcal{F}_n)_{n\geq 1}$ 是一个鞅, 而 A_n 是 $\mathcal{F}_{n-1}$ 可测的随机变量, 也就是说 $(A_n)_{n\geq 1}$ 是可料的随机变量序列. 事实上, 这个结论是具有一般性的. 这就是著名的 Doob 下鞅分解定理.

定理 3.2.1. (Doob 下鞅分解) $(\xi_n)_{n\geq 0}$ 为 $(\mathcal{F}_n)_{n\geq 0}$ 下鞅列, 则存在唯一 $(\mathcal{F}_n)_{n\geq 0}$ 鞅列, 满足

(1) $M_0 = \xi_0$ (初值);

(2) 存在唯一的可料递增随机序列 $(A_n)_{n\geq 0}$, $A_0 = 0$, 使得

$$\xi_n = M_n + A_n. \tag{3.2.3}$$

分解式 (3.2.3) 还可以写成下鞅差序列的形式:

$$\xi_{n+1} - \xi_n = M_{n+1} - M_n + A_{n+1} - A_n, \ n \geq 0, \tag{3.2.4}$$

其中, $M_{n+1} - M_n = \xi_{n+1} - \mathrm{E}[\xi_{n+1}|\mathcal{F}_n]$, $A_{n+1} - A_n = \mathrm{E}[\xi_{n+1}|\mathcal{F}_n] - \xi_n$. 有时也记 $\Delta\xi_n = \xi_{n+1} - \xi_n$, $\Delta M_n = M_{n+1} - M_n$ 和 $\Delta A_n = A_{n+1} - A_n$, (3.2.4) 也可以表示为

$$\Delta\xi_n = \Delta M_n + \Delta A_n, \ n \geq 0, \tag{3.2.5}$$

证明: 参见例 (3.1.4) 的证明. 唯一性证明留作习题. □

例 (3.1.4) 实际上是更广的 Doob 下鞅分解定理. 它是说对于一般的期望有限的离散时间适应过程都可以分解成鞅与可料过程的和. Doob 分解在推广到连续时间时有本质的困难. 例如“可料性”该如何推广? 在连续时间时称为 Doob-Meyer 分解, 它是现代鞅论的起点.

例 3.2.1. 接例 (3.1.1) (ξ_n) 相互独立, $\mathrm{E}|\xi_n| < \infty$, 设 $\mathrm{E}\xi_n > 0$, 则 $X_n = \sum\limits_{k=1}^{n} \xi_k$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 下鞅. (X_n) 的 Doob 分解为

$$X_n = \sum_{k=1}^{n}(\xi_k - \mathrm{E}\xi_k) + \sum_{k=1}^{n}\mathrm{E}\xi_k = \eta_n + \sum_{k=1}^{n}\mathrm{E}\xi_k.$$

若进一步对于任意的 n, $\mathrm{E}\xi_n^2 < \infty$, $(\eta_n^2)_{n\geq 1}$ 为下鞅列, 则 $(\eta_n^2)_{n\geq 1}$ 的 Doob 分解为

$$\eta_n^2 = \Big(\eta_n^2 - \sum_{k=1}^{n}\mathrm{var}\xi_k\Big) + \sum_{k=1}^{n}\mathrm{var}\xi_k = \zeta_n + \sum_{k=1}^{n}\mathrm{var}\xi_k.$$

例 3.2.2. 设 T 概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上取正整数值的随机变量. 对任意的 $k \geq 1$, $\mathrm{P}(T = k) > 0$. 设 $X_0 = 0$, $X_n(\omega) = \begin{cases} 1 & n \geq T \\ 0 & n < T \end{cases}$. 这个过程可以用来描述突发事件何时发生这一随机现象. 设 $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$, 则 $(X_n)_{n\geq 0}$ 关于 $(\mathcal{F}_n)_{n\geq 0}$ 适应, 关于时间单调非减. 因此, $(X_n, \mathcal{F}_n)_{n\geq 0}$ 是下鞅. 求其 Doob 分解. 未完待续...

例 3.2.3. (接例3.1.7, 分支过程的 Doob 分解)

对于例3.1.7中的分支过程 $\xi_0 = 1$, $\xi_n = \sum_{k=1}^{\xi_{n-1}} \eta_{n,k}$, 设 $\mathrm{E}\eta_{n,k} = \mu, \mathrm{var}\eta_{n,k} = \gamma^2$, $\Delta\xi_n = \xi_{n+1} - \xi_n$. 考虑 $(\xi_n)_{n\geq 0}$ 的 Doob 分解, 则由第二章性质2.2.6 得,

$$\mathrm{E}[\Delta\xi_n|\xi_n] = \mathrm{E}\Big[\sum_{k=1}^{\xi_n}\eta_{n,k} - \xi_n|\xi_n\Big] = \mathrm{E}\Big[\sum_{k=1}^{m}\eta_{n,k} - m\Big]\Big|_{m=\xi_n} = (\mu - 1)\xi_n.$$

$$\begin{aligned}\operatorname{var}(\Delta\xi_n|\xi_n) &= \mathrm{E}\big[(\Delta\xi_n-\mathrm{E}[\Delta\xi_n|\xi_n])^2|\xi_n\big]\\ &= \mathrm{E}\big[\big(\sum_{k=1}^{m}\eta_{n,k}-m-(\mu-1)m\big)^2\big]\big|_{m=\xi_n}\\ &= \gamma^2 m\big|_{m=\xi_n}\\ &= \gamma^2\xi_n.\end{aligned}$$

即 $(\xi_n)_{n\geq 0}$ 在 n 时刻 "人口数" 平均的一步变化正比于 n 时刻的 "人口数"$(\mu\neq 1)$; "人口数" 变化的方差正比于 n 时刻的 "人口数", 也就是说 Doob 分解的鞅差形式为

$$\Delta\xi_n=(\mu-1)\xi_n+\gamma\sqrt{\xi_n}\frac{\Delta\xi_n-(\mu-1)\xi_n}{\gamma\sqrt{\xi_n}}. \tag{3.2.6}$$

或者

$$\xi_{n+1}-\xi_0=\sum_{k=1}^{n}\Delta\xi_k=(\mu-1)\sum_{k=1}^{k}\xi_k+\gamma\sum_{k=1}^{n}\sqrt{\xi_k}\frac{\Delta\xi_k-(\mu-1)\xi_k}{\gamma\sqrt{\xi_k}}. \tag{3.2.7}$$

保持上述 "分支机制", 对离散时间和离散状态分支过程 (6.3.22) 取适当的时间-空间尺度极限就得到 Feller 分支扩散过程

$$\mathrm{d}X_t=\alpha X_t\mathrm{d}t+\sigma\sqrt{X_t}\text{“}\mathrm{d}\tilde{\xi}_t\text{”}.$$

或者

$$X_t-X_0=\alpha\int_0^t X_s\mathrm{d}s+\sigma\int_0^t\sqrt{X_s}\text{“}\mathrm{d}\tilde{\xi}_s\text{”}.$$

同样, 这里也需要解释积分 $\int_0^t\sqrt{X_s}$“$\mathrm{d}\tilde{\xi}_s$” 的意义.

对于基本模型 (3.2.2), 下面介绍一些它的特例.

例 3.2.4. 设银行的利率为 R(离散时间), 即存入银行的资金增长满足

$$S_{n+1}=S_n+RS_n=(1+R)S_n.$$

如果设

$$\xi_n=S_n,\ \eta_n=R(\text{常数}),\ b_n(S_0,S_1,\cdots,S_{n-1})=S_{n-1},$$

则这就是基本模型 (3.2.2) 非常特殊的一个例子.

例 3.2.5. Cox-Ross-Rubinstein 模型, 二叉 (树) 模型 (binomial-tree model) **(缺图形)**

假定价格增长满足

$$S_{n+1}=(1+\eta_{n+1})S_n=S_n+\eta_{n+1}S_n, \tag{3.2.8}$$

其中 (η_n) 是独立同分布的随机变量序列, 且 $1+\eta_{n+1}=\frac{S_{n+1}}{S_n}\sim\begin{pmatrix}a & b\\ p & 1-p\end{pmatrix}$, $0<a<b$, $0<p<1$, $(\eta_n)_{n\geq 1}$ 与 S_0 独立.

如果设 $\xi_n=S_n$, $b_n(S_0,S_1,\cdots,S_{n-1})=S_{n-1}$, 则这个模型也是基本模型 (3.2.2) 的一个特例.

定义 $\frac{1}{S_n}\mathrm{E}\big[S_{n+1}-S_n\big|\eta_n,\eta_{n-1},\cdots,S_0\big]$ 为这支的**收益率** (yield), 它表示单位时间单位价格的平均收益.

$$\begin{aligned}&\frac{1}{S_n}\mathrm{E}\big[S_{n+1}-S_n\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\\ =\ &\mathrm{E}\big[\eta_{n+1}\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\\ =\ &\mathrm{E}\eta_{n+1}=ap+b(1-p)-1\equiv\mu.\end{aligned}$$

定义 $\frac{1}{S_n}\left(\mathrm{var}\big[(S_{n+1}-S_n)\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\right)^{\frac{1}{2}}$ 为这支的的**波动率** (volatility), 它表示单位时间单位价格的平均波动幅度, 是一种对“随机性”或者“风险”的度量指标.

$$
\begin{aligned}
&\frac{1}{S_n}\Big(\mathrm{var}\big[(S_{n+1}-S_n)\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\Big)^{\frac{1}{2}}\\
=\ &\frac{1}{S_n}\Big(\mathrm{E}\Big[\big(\eta_{n+1}S_n-\mathrm{E}\big[\eta_{n+1}S_n\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\big)^2\Big|\eta_n,\eta_{n-1},\cdots,S_0\Big]\Big)^{\frac{1}{2}}\\
=\ &\frac{1}{S_n}\Big(\mathrm{E}\big[S_n^2(\eta_{n+1}-\mathrm{E}\eta_{n+1})^2\big|\eta_n,\eta_{n-1},\cdots,S_0\big]\Big)^{\frac{1}{2}}\\
=\ &\sqrt{\mathrm{E}(\eta_{n+1}-\mathrm{E}\eta_{n+1})^2}\equiv\sigma.
\end{aligned}
$$

在一般情况下, μ 和 σ 可以是随机的随时间变化的量, 因此也可以是随机过程.

Cox-Ross-Rubinstein 模型 (3.2.8) 的 Doob 分解为

$$
S_{n+1}-S_n=(\mathrm{E}\eta_{n+1})S_n+(\eta_{n+1}-\mathrm{E}\eta_{n+1})S_n=\mu S_n+\sigma\frac{(\eta_{n+1}-\mu)}{\sigma}S_n,
$$

或者写成

$$
\frac{S_{n+1}-S_n}{S_n}=\mu+\sigma\frac{\eta_{n+1}-\mu}{\sigma}. \tag{3.2.9}
$$

其中 $\mathrm{E}\Big(\frac{\eta_{n+1}-\mathrm{E}\eta_{n+1}}{\sigma}\Big)^2=1$, $\mathrm{E}\Big(\frac{\eta_{n+1}-\mathrm{E}\eta_{n+1}}{\sigma}\Big)=0$. 这也 Black-Scholes 模型相对应得离散时间版本.

例 3.2.6. (Black-Scholes 模型的直观推导)

若考虑连续时间参数的价格 $(S_t)_{t\geq 0}$, 它应该满足怎样的“方程”呢? 设 $\mathcal{F}_t=\sigma(S_0,\xi_s,s\leq t)$, S_0 与 $(\xi_t)_{t\geq 0}$ 独立, $(S_t)_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 适应过程.

我们用黎曼积分的想法来考察 0 到 t 时间段内 $(S_u)_{0\leq u\leq t}$ 的变化. 首先对 $[0,t]$ 做划分 $0=t_0<t_1<\cdots<t_n=t$, $S_t-S_0=\sum_{k=1}^n(S_{t_k}-S_{t_{k-1}})$. 利用例 (3.2.5) 中 Doob 分解的想法, 参考 (3.2.9) 式, 在时刻 t_{k-1} 单位时间单位价格的收益应该近似地满足下式

$$
\frac{S_{t_k}-S_{t_{k-1}}}{(t_k-t_{k-1})S_{t_{k-1}}}\approx\frac{\xi_{t_k}-\xi_{t_{k-1}}}{t_k-t_{k-1}},
$$

类比 (3.2.9) 中 (η_n) 满足的统计性质[①], 可以假设 (ξ_t) 应该满足如下的条件:

1. 对于任意的划分 $0=t_0<t_1<\cdots<t_n=t$, $\xi_0,\xi_{t_1}-\xi_0,\cdots,\xi_{t_n}-\xi_{t_{n-1}}$ 相互独立;

2. $\xi_s-\xi_u$ 的分布只与 $s-u$ 有关.

也就是说 (ξ_t) 是一个独立增量过程, 而且增量平稳. 我们还可以进一步假定其二阶矩有限, 期望和方差关于时间 t 连续. 因此 $\mathrm{E}(\xi_t-\xi_0)=\mu t$, $\mathrm{var}(\xi_t-\xi_0)=\sigma^2 t$. 故而近似地可以得到下式

$$
S_t-S_0\approx\mu\sum_{k=1}^n S_{t_{k-1}}(t_k-t_{k-1})+\sigma\sum_{k=1}^n S_{t_{k-1}}\frac{\xi_{t_k}-\xi_{t_{k-1}}-\mu(t_k-t_{k-1})}{\sigma}. \tag{3.2.10}
$$

当 $\max_k(t_k-t_{k-1})\to 0$ 时, (3.2.10) 式的第一项在黎曼积分的意义下收敛到 $\mu\int_0^t S_s\mathrm{d}s$, 而第二项可以设想按某种意义收敛到“积分”$\int_0^t S_s$“$\mathrm{d}\bar{\xi}_s$”. 形式地, 我们可以认为 $(S_t)_{t\geq 0}$ 满足下面的方程

$$
S_t-S_0=\mu\int_0^t S_s\mathrm{d}s+\sigma\int_0^t S_s\text{“}\mathrm{d}\bar{\xi}_s\text{”}. \tag{3.2.11}
$$

这就是连续时间的 Black-Scholes 方程 (模型), 我们将在第五章具体定义第二项中的积分. 另一方面, 由对以上直观推导的理解, 也可以按照实际情况假定 (ξ_t) 满足其它合理的统计假设, 这样就可以得到更广的 Black-Scholes 方程.

[①] 在例3.2.5中, 如果记 $\widetilde{\xi}_n=\sum_{k=1}^n\eta_k$, 则 $(\widetilde{\xi}_n)_{n\geq 1}$ 是一个离散时间独立增量过程. 在例3.2.6中, $(\xi_t)_{t\geq 0}$ 的作用相当于 $(\widetilde{\xi}_n)_{n\geq 1}$, $\frac{\xi_{t_{k+1}}-\xi_{t_k}}{t_{k+1}-t_k}$ 的作用可以理解为 $\widetilde{\xi}_{k+1}-\widetilde{\xi}_k=\eta_{k+1}$ 的作用, 表示在时刻 t_k 附近单位时间内单位价格的收益.

在连续时间过程中引入过程 $(\xi_t)_{t\geq 0}$, 而不是引入相互独立的随机变量族 $(\eta_t,t\geq 0)$ 作为过程的原因在于, $(\eta_t,t\geq 0)$ 作为随机过程往往具有很糟糕的 (轨道) 性质, 甚至这样的过程不存在.

从 Doob 下鞅分解定理的角度来看, 在 (3.2.10) 式直观推导收敛到 (3.2.11) 式的过程中, 如何保持第二项的"鞅性"不消失, 如何在连续时间参数下理解第一项中的"可料性"? 这将是第五章以及随机分析课程中的重要内容.

♣♣♣20230323

例 3.2.7. 我们将例3.2.4和例3.2.5综合起来考虑. 在 n 时刻的价格为 S_n, 但并不代表其实际的"价值". 直观地看, 1 元钱开始存在银行中, 在 n 时刻就增值到 $(1+R)^n$ 元. 同样用 1 元钱在开始时刻买了, 在 n 时刻它的价格是 S_n, 但其与存在银行中的那一元钱比, 也就是说, 如果我们用 $\frac{1}{(1+R)^n}$ 作为 n 时刻的计价单位的话, S_n 的"价值"就是 $\tilde{S}_n = \frac{S_n}{(1+R)^n}$. 用金融数学的术语说, $\tilde{S}_n$ 为该证券在初始时刻的折现价格.

当人们希望投资这个, 自然关心某些时刻之后这个的价值是否增加, 是否减少. 也就是说, 对于投资这场"赌博", 人们关心这场"赌博"在什么条件下是"公平"的, 即其折现价格什么时候是"鞅".

记 $\mathcal{F}_n = \sigma(S_0, \eta_1, \cdots, \eta_n)$. 首先设 $(\tilde{S}_n, \mathcal{F}_n)$ 是鞅来看看 η_n 的分布应该满足的条件.

$$\tilde{S}_n = \mathrm{E}[\tilde{S}_{n+1}|\mathcal{F}_n] = \tilde{S}_n \mathrm{E}\Big[\frac{\tilde{S}_{n+1}}{\tilde{S}_n}\Big|\mathcal{F}_n\Big] = \tilde{S}_n \mathrm{E}\Big[\frac{1+\eta_{n+1}}{1+R}\Big|\mathcal{F}_n\Big] = \tilde{S}_n \mathrm{E}\Big(\frac{1+\eta_{n+1}}{1+R}\Big).$$

解出 $p = \frac{b-(1+R)}{b-a}$, 为使 $0 < p < 1$, 还需要满足 $a < 1+R < b$.

当然 $p = \frac{b-(1+R)}{b-a}$ 时, 很容易也能证明 $(\tilde{S}_n, \mathcal{F}_n)$ 是鞅. 此时还有 $\mu = R$, 这个结论也很符合直观, 当银行利率和收益率相等的时候, 投资这个证券才是不赔也不赚的"赌博".

$\frac{b-(1+R)}{b-a}$ 称为 Cox-Ross-Rubinstein 模型的公平概率, 或者此时 $\frac{S_{n+1}}{S_n}$ 的分布称为 Cox-Ross-Rubinstein 模型的风险中性概率.

通过上面这个简单的例子 (3.2.7) 希望给大家些许直观, 鞅是如何刻画一些金融数学中的观念和性质的. 事实上现代金融数学理论中建立模型的起点是规定不存在套利的机会, 而无套利的数学描述依赖于"鞅"或"鞅测度".

3.3 选样定理和 Doob 停时定理

3.3.1 停时的概念

在我们日常生活中, 除了用秒、分、时、日、月、年等时间刻度记录我们日常生活的轨迹外, 有时还用一些标志性事件发生的时刻来记录生活轨迹. 例如, 出生的时刻、进入幼儿园的时刻、第二次参加某社团活动的时刻、在大学中倒数第二次参加考试的时刻、最后一次遇到某人的时刻等等. 这些时刻, 虽然都是对同一件事儿的描述, 但对于不同的人, 发生的时刻可能是不一样的. 在随机过程中也类似, 可以用 $0, 1, 2\cdots$ 或 $t \in [0, \infty)$ 等确定性的时间来刻画一个过程, 也可以用过程第 i 次击中某集合或最后一次击中某集合等具有随机因素的时刻来描述一个随机过程. 这就是随机时间的概念.

定义 3.3.1. (随机时间) 随机时间 ϑ 是一个从 $(\Omega, \mathcal{F})$ 映射到 $[0, \infty]$ 的随机变量或 $(\Omega, \mathcal{F})$ 映射到 $\{0, 1, 2, \cdots, \infty\}$ 的随机变量[①], 即 $\forall t \in \{0, 1, 2, \cdots\}$ 或 $t \in [0, \infty)$, $\{\omega : \vartheta(\omega) \le t\} \in \mathcal{F}$.

当我们在一个带流的概率空间 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, \mathrm{P})$ 上研究一个随机过程 $(\xi_t)_{t\in T}$ 时, 有一类特殊的随机时间非常重要, 这就是停时的概念.

定义 3.3.2. (停时, stopping time) 设 τ 为一个随机时间, 若对于 $\forall t \in T$, $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$, 则称 τ 是 $(\mathcal{F}_t)_{t\in T}$ 停时.

特别当 $T = \{0, 1, 2, \cdots\}$ 时, 上述定义可表述为 $\forall n \ge 0$, $\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n$ 或者 $\mathbf{1}_{\{n\}}(\tau)$ 是关于 $\mathcal{F}_n$ 可测的随机变量.

[①] 在定义随机时间 ϑ 时, 可以要求 ϑ 取 ∞ 值. 例如, 设 $(\eta_n)_{n\ge1}$ 是独立同分布的随机变量序列满足 $\mathrm{P}(\eta_k = 1) = p > \frac{1}{2}$, $\mathrm{P}(\eta_k = -1) = 1 - p < \frac{1}{2}$, 考虑非对称简单随机游动 $\xi_n = \sum_{k=1}^{n} \eta_k$ 首次到达某个负整数的时刻 ϑ, 则 ϑ 有一个正概率等于 ∞.

停时就是一个随机时间, 它在确定性的 t 时刻之前 (包含 t) 是否已经达到, 即事件 $\{\omega : \tau(\omega) \leq t\}$ 是否发生是被t之前的 “信息” $\mathcal{F}_t$ 所决定的.

例如, 全班同学上午八点从北大出发以不同方式前往王府井, 第一次碰到尾号是 7 的汽车的时刻是一个停时, 第二次碰到尾号是 7 的汽车时刻也是一个停时. 比如每一时刻问所有同学是否碰到了尾号是 7 的汽车时, 每个同学都可以回答 “是” 或 “否”. 但最后一次, 倒数第二次碰到尾号是 7 的汽车的时刻不是停时. 在时刻 t, 无法判断 t 之后还有没有尾号是 7 的汽车出现, 因此是无法给出“是”或“否”这样的答案的; 再如, 到达王府井所用总时间的一半这个时刻也不是停时. 总时间依赖于整个过程所有的信息 $\sigma(\cup_{t\in T}\mathcal{F}_t) = \mathcal{F}_\infty$, 其一半也依赖于 $\mathcal{F}_\infty$, 因此在某个确定性的时刻, 比如在上午八点半时, 是无法给出 “此刻是否已经用了总时间的一半? ” 这个问题 “是” 或 “否” 这样确定的答案的.

性质 3.3.1. 常数 t 是 $(\mathcal{F}_t)_{t\in T}$ 停时. 若 τ 和 σ 都是 $(\mathcal{F}_t)_{t\in T}$ 停时, 则 $\tau\wedge\sigma = \min(\tau,\sigma), \tau\vee\sigma = \max(\tau,\sigma)$ 和 $\tau+\sigma$ 也是 $(\mathcal{F}_t)_{t\in T}$ 停时. 因此 $\tau\wedge t, \tau\vee t$ 也是停时.

证明: 只证 $\tau\wedge\sigma$, 其它情况作为习题. 要证 $\tau\wedge\sigma$ 为停时, 即要证 $\forall t\in T$,
$$\{\tau\wedge\sigma\leq t\}\in\mathcal{F}_t.$$
事实上,
$$\{\tau\wedge\sigma\leq t\} = \{\tau\leq t\}\cup\{\sigma\leq t\}.$$
因为 τ,σ 是停时, $\{\tau\leq t\}\in\mathcal{F}_t$, $\{\sigma\leq t\}\in\mathcal{F}_t$, 所以 $\{\tau\leq t\}\cup\{\sigma\leq t\}\in\mathcal{F}_t$.

□

定义 3.3.3. 设 $(\xi_n)_{n\geq 1}$ 是 $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\geq1},\mathrm{P})$ 上适应的随机序列, τ 是关于事件域流 $(\mathcal{F}_n)_{n\geq1}$ 的停时, 设 τ 几乎处处取有限值, 即 $\mathrm{P}(\tau<\infty)=1$, 定义 $(\xi_\tau)(\omega)=\xi_{\tau(\omega)}(\omega)$, 则 ξ_τ 是 $(\Omega,\mathcal{F})$ 上的随机变量, 称为随机序列 $(\xi_n)_{n\geq1}$ 在停时 τ 上的取值[①].

$(\xi_t)_{t\geq0}$ 是 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq0},\mathrm{P})$ 上适应的随机过程, τ 是关于事件域流 $(\mathcal{F}_t)_{t\geq0}$ 的停时. 设 τ 几乎处处取有限值, 即 $\mathrm{P}(\tau<\infty)=1$, 定义 $(\xi_\tau)(\omega)=\xi_{\tau(\omega)}(\omega)$, 则在一定的条件下 (ξ_τ) 是随机变量, 在本讲义中我们总假定 ξ_τ 是随机变量, 称为随机过程 $(\xi_t)_{t\geq0}$ 在停时 τ 上的取值[②].

3.3.2 选样定理

参加博弈的人根据前面历史给自己定制一个退出策略. 例如, 当赢了 500 元钱之后退出以图获利, 或者是当输掉初始本金的 20% 时退出以求 “保本”. 这个退出的时刻τ即为一个停时.

如果博弈是公平的, 即为 “鞅” 时, 直观上想, 不论博弈者采取怎么样的退出策略, 他都不可能获利, 从统计平均来看应该有
$$\mathrm{E}\xi_\tau = \mathrm{E}\xi_0, \tag{3.3.1}$$
这就是所谓 “停时定理” 的本质含义.

但事实上这是不对的. 我们回到倍赌策略的例子具体分析.

例 3.3.1. (接例 (3.1.5)) 设 $\tau=\inf\{n\geq1,\eta_1=-1,\cdots,\eta_{n-1}=-1,\eta=1\}$, 则
$$\mathrm{P}(\tau>n)=\frac{1}{2^n},\ \mathrm{P}(\tau<\infty)=1,\ \mathrm{E}\tau=\sum_{k=1}^{\infty}k\mathrm{P}(\tau=k)=\sum_{k=1}^{\infty}k\frac{1}{2^k}=2<\infty,$$
而且,
$$\xi_\tau=1,\ \ \mathrm{E}\xi_\tau\neq\mathrm{E}\xi_0=0.$$
这说明上述的想法并不成立.

[①] 严格地说, 对于 $(\Omega,\mathcal{F},(\mathcal{F})_{n\geq1},\mathrm{P})$, 在以上定义中应该要求 $(\Omega,\mathcal{F},\mathrm{P})$ 是一个完备的概率空间.

[②] 在连续时间参数是要证明 ξ_τ 为一个随机变量并不是一件平凡的事情, 这是随机分析中比较精细的工作. 因此在本讲义中我们假定 ξ_τ 都是随机变量.

我们来分析, $\mathrm{E}\big(|\xi_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)=\mathrm{E}\big((2^n-1)\mathbf{1}_{\{\tau>n\}}\big)=1-\frac{1}{2^n}$. 这说明, 虽然这个博弈持续很长时间的概率很小, 但是在持续长时间之后博弈者负债额 $|\xi_n|=2^n-1$ 却非常大. 进一步分析, $\xi_{\tau-1}$ 表示博弈者在首次获利之前的负债额 (注意 $\tau-1$ 不是停时),

$$\xi_{\tau-1}=\begin{cases}\xi_0=0, & \tau=1\\ \xi_0-b_1(\xi_0)=-1, & \tau=2\\ \xi_0-b_1(\xi_0)-b_2(\xi_0,-1)=-3, & \tau=3\\ \cdots & \cdots\\ \xi_0-b_1(\xi_0)-\cdots-b_{n-1}(\xi_0,\underbrace{-1,\cdots,-1}_{n-2\text{次}})=-(2^{n-1}-1), & \tau=n\end{cases}.$$

由此得到

$$\begin{aligned}\mathrm{E}\xi_{\tau-1} &= \sum_{k=1}^{\infty}\mathrm{E}\xi_{k-1}\mathbf{1}_{\{\tau=k\}}\\ &= \sum_{k=1}^{\infty}-(2^{k-1}-1)\mathrm{P}(\tau=k)\\ &= \sum_{k=1}^{\infty}-(2^{k-1}-1)\frac{1}{2^k}\\ &= -\infty.\end{aligned}$$

这说明平均来看在首次获利之前, 博弈者负债是无穷大, 或者也可以直观理解为博弈者在初始时刻就要有无穷大的财富 (即 $\mathrm{E}\xi_0=\infty$) 才可以保证执行这种“见好就收”的退出博弈策略时, 博弈的“公平性”. 这不符合鞅对数学期望有限性的要求.

再如考虑下面的简单对称随机游动:

例 3.3.2. $(\eta_n)_{n\geq1}$ 是独立同分布随机变量序列满足 $\mathrm{P}(\eta_n=1)=\mathrm{P}(\eta_n=-1)=\frac{1}{2}$, $\xi_0=0$, 设

$$\xi_n=\xi_0+\sum_{k=1}^{n}\eta_k,\quad \mathcal{F}_n=\sigma(\xi_0,\eta_1,\cdots,\eta_n),$$

则 $(\xi_n,\mathcal{F}_n)$ 为鞅. ξ_n 表示博弈者开始时身无分文, 在进行了简单博弈后, 其在 n 时刻的资金. 令 $\tau=\inf\{n\geq1,\ \xi_n=1\}$, 即 τ 表示博弈者首次赚到一块钱. 我们知道简单对称随机游动是零常返的, 也就是说 $\mathrm{P}(\tau<\infty)=1$, 但是 $\mathrm{E}\tau=\infty$. 很显然 $\xi_\tau=1$, $\mathrm{E}\xi_\tau=1$, 但 $\mathrm{E}\xi_0=0$, $\mathrm{E}\xi_\tau\neq\mathrm{E}\xi_0$.

这说明, 平均来看需要等无穷长的时间才能获利.

上面的两个例子说明, 在博弈时间持续无穷长和博弈的初始资金是无穷大的假设下, 直观的想法 (3.3.1) 并不成立. 因此, 可以提出下面的条件来保证 (3.3.1) 成立.

定理 3.3.1. (选样定理) 设 $(\xi_n,\mathcal{F}_n)_{n\geq0}$ 是鞅, τ 是一个取有限值的 $(\mathcal{F}_n)_{n\geq0}$ 停时, 而且满足

(1) τ 是有界停时, 即存在常数 $M<\infty$, $\mathrm{P}(\omega:\tau(\omega)\leq M)=1$;

或更一般地

(2) ξ_τ 的期望有限, 且 $\lim\limits_{n\to\infty}\mathrm{E}\big(|\xi_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)=0$,

则有 $\mathrm{E}\xi_\tau=\mathrm{E}\xi_0$.

上面的条件中, (1) 用“博弈术语”说就是: 在一场公平博弈中, 不可能利用有界停时策略来增加个人的财富期望. 事实上, (2) 是一个相当广的条件[①]. 这个条件可以理解为对于充分大的时刻 n, 如果此刻停时 τ 还未到达 (即 $\tau>n$), $|\xi_n|$ 不能过分的大; 而且, 在停时处, 平均来看 $|\xi_\tau|$ 不能是无穷.

[①] 条件 (2) 还可以减弱为: ξ_τ 的期望有限, 且 $\liminf\limits_{n\to\infty}\mathrm{E}\big(|\xi_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)=0$.

选样定理第一部分也称为有界选样定理, 它是 Doob 有界停时定理 (定理3.3.2) 的直接推论; 第二部分的证明留作习题. 若 τ 是有界停时, 条件 (2) 自动满足.

例 3.3.3. (推广的赌徒输光问题)

设 $(\eta_n)_{n\geq 1}$ 是独立同分布的随机变量序列, $\mathrm{P}(\eta_k=1)=p$, $\mathrm{P}(\eta_k=-1)=q$, $\mathrm{P}(\eta=0)=r$. 这里假设 $p>0, p+q+r=1$. 设 $\xi_n=a+\sum_{k=1}^n\eta_k$. 这个模型可以想象为张三和李四利用某种方式进行博弈, 每次出现 "张三胜"、"张三负" 与 "和局" 三种可能性. 若 "张三胜", 则张三赢得李四一元钱; 若 "张三负", 则张三输给李四一元钱; 若是 "和局", 则双方互不 "付钱". 在开始时刻, 张三有 a 元钱, 李四有 b 元钱. ξ_n 表示张三在 n 时刻的资金. 设

$$\tau=\inf\{n\geq 0,\ \xi_n=0 \text{ 或 } \xi_n=a+b\},\ \inf\emptyset\stackrel{\text{def}}{=}+\infty.$$

τ 表示张三或李四输光资金的时刻. 令 $p_0=\mathrm{P}(\xi_\tau=0)$, 表示张三输光 (破产) 的概率; $q_{a+b}=\mathrm{P}(\xi_\tau=a+b)$, 表示李四输光 (破产) 的概率.

求赌徒输光的概率 p_0 和 q_{a+b}, 以及博弈平均持续时间 $\mathrm{E}\tau$.

当 $r=0$ 时, 这就是经典的简单随机游动的 "赌徒输光问题".

解: 设 $\mathcal{F}_n=\sigma(\eta_1,\cdots,\eta_n)$.

(1) 首先需要证明 $\mathrm{P}(\tau<\infty)=1$, 使得 ξ_τ 有意义而且

$$p_0+q_{a+b}=1 \tag{3.3.2}$$

成立.

对于任意正整数 n, 考虑 n 个相继的时间段 $\{1,\cdots,a+b\}$, $\{a+b+1,\cdots,2(a+b)\}$,$\cdots$, $\{(n-1)(a+b)+1,\cdots,n(a+b)\}$, 设 $A_k=\{\eta_{(k-1)(a+b)+1}=1,\cdots,\eta_{k(a+b)}=1\}$, 则 $\mathrm{P}(A_k)=p^{a+b-1}>0$, 而且

$$\bigcup_{k=1}^n A_k\subset\{\tau<\infty\}.$$

由 $A_k, k=1,\cdots,n$ 的相互独立性, 我们得到

$$\begin{aligned}\mathrm{P}(\tau<\infty) &\geq \mathrm{P}(\bigcup_{k=1}^n A_k)\\ &= 1-\mathrm{P}(\bigcap_{k=1}^n A_k^c)\\ &= 1-\prod_{k=1}^n \mathrm{P}(A_k^c)\\ &= 1-(1-p^{a+b-1})^n\\ &\stackrel{n\to\infty}{\longrightarrow} =1.\end{aligned}$$

(2) 当 $\mathrm{E}\eta_1=0$, 即 $p=q$ 时, 由例3.1.1的 (1) 和 (2) 知道 $(\xi_n,\mathcal{F}_n)_{n\geq 1}$ 和 $(\xi_n^2-2pn,\mathcal{F}_n)_{n\geq 1}$ 是鞅列. 虽然 τ 不是有界停时, 但注意到 $\tau\geq n$ 时, $0\leq\xi_n\leq a+b$, 使得我们可以比较方便地利用有界选样定理和控制收敛定理得到所需结论. 对于任意的 $n\geq 1$, $\tau\wedge n$ 是有界停时, 一方面由有界选样定理得到

$$\mathrm{E}\xi_{\tau\wedge n}=\mathrm{E}\xi_1=a; \tag{3.3.3}$$

$$\mathrm{E}(\xi_{\tau\wedge n}^2)-2p\mathrm{E}(\tau\wedge n)=\mathrm{E}\xi_1^2-2p=a^2. \tag{3.3.4}$$

另一方面

$$0\leq\xi_{\tau\wedge n}\leq a+b,\ \text{当}n\to\infty\text{ 时}, \tau\wedge n\uparrow\tau,\ \xi_{\tau\wedge n}\to\xi_\tau.$$

利用有界控制收敛定理和 Levi 单调收敛定理对 (3.3.3)(3.3.4) 取极限,

$$\mathrm{E}\xi_\tau = (a+b)q_{a+b} = a; \tag{3.3.5}$$

$$\mathrm{E}\xi_\tau^2 - 2p\mathrm{E}\tau = (a+b)^2 q_{a+b} - 2p\mathrm{E}\tau = a^2. \tag{3.3.6}$$

将 (3.3.5) 和 (3.3.2) 联立求解得到

$$q_{a+b} = \frac{a}{a+b}, \quad p_0 = \frac{b}{a+b}. \tag{3.3.7}$$

由 (3.3.6) 得到

$$\mathrm{E}\tau = \frac{ab}{2p}. \tag{3.3.8}$$

(3) 当 $\mathrm{E}\eta_1 \neq 0$, 即 $p \neq q$ 时, 我们先利用例3.1.1的 (3) 构造一个依赖于参数 s 的鞅. 记 $\alpha(s) = \mathrm{E}e^{s\eta_1} = e^s p + e^{-s} q + r$,

$$\zeta_n(s) = \alpha(s)^{-n} \prod_{k=1}^{n} e^{s\eta_k} = e^{-sa}\alpha(s)^{-n} e^{s\xi_n}.$$

利用有界选样定理得到

$$\mathrm{E}\zeta_{\tau\wedge n}(s) = e^{-sa}\mathrm{E}\big(\alpha(s)^{-(\tau\wedge n)} e^{s\xi_{\tau\wedge n}}\big) = \alpha(s)^{-1}\mathrm{E}e^{s\eta_1} = 1. \tag{3.3.9}$$

这里有需要计算两个不独立随机变量的数学期望. 我们可以取 $s = \ln\frac{q}{p}$ 使得 $\alpha(s) = 1$, 则

$$\Big(\zeta_n\big(\ln\frac{q}{p}\big) = \Big(\frac{q}{p}\Big)^{\xi_n}\Big(\frac{q}{p}\Big)^{-a}, \mathcal{F}_n\Big)_{n\geq 1}$$

为鞅, 而且由 (3.3.9)

$$\mathrm{E}\Big(\frac{q}{p}\Big)^{\xi_{\tau\wedge n}} = \Big(\frac{q}{p}\Big)^{a}.$$

再利用有界收敛定理得到,

$$\mathrm{E}\Big(\frac{q}{p}\Big)^{\xi_\tau} = \Big(\frac{q}{p}\Big)^{a},$$

即

$$\Big(\frac{q}{p}\Big)^{0} p_0 + \Big(\frac{q}{p}\Big)^{a+b} q_{a+b} = \Big(\frac{q}{p}\Big)^{a}. \tag{3.3.10}$$

与 (3.1.1) 联立求解得到

$$p_0 = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}}, \quad q_{a+b} = \frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}}. \tag{3.3.11}$$

由例3.1.1的 (1) 知道 $(\xi_n - (p-q)n, \mathcal{F}_n)_{n\geq 1}$, 类似的有

$$\mathrm{E}\xi_\tau - (p-q)\mathrm{E}\tau = a, \tag{3.3.12}$$

因此, 得到

$$\mathrm{E}\tau = \frac{1}{p-q}(\mathrm{E}\xi_\tau - a) = \frac{1}{p-q}\Big[(a+b)\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}} - a\Big]. \tag{3.3.13}$$

□

在以上的推广的赌徒输光问题中我们利用鞅的方法得到了所要的结果. 我们还可以借助这些结果, 进一步得到一些有趣的结论, 包括简单随机游动的经典结论.

♣♣♣20230330

例 3.3.4. 我们继续例3.3.2的讨论. 考虑 ξ_n 从 $a>0$ 出发可以到达 0 点的概率和首中 0 点的平均时间. 将上例中的 τ 记为 τ_a^b, 并设

$$\tau_0=\inf\{n\ge 1,\xi_n=0\},\ \tau_{a+b}=\inf\{n\ge 1,\xi_n=a+b\},$$

则 $\tau_a^b=\tau_0\wedge\tau_{a+b}$, 而且当 $b\uparrow\infty$ 时, $\tau_a^b\uparrow\tau_0$[①].

设 A 表示 ξ_n 可以到达 0 点这个事件, 则 $A=\bigcup_{b=1}^{\infty}\{\xi_{\tau_a^b}=0\}$, 而且 $\{\xi_{\tau_a^b}=0\}\subset\{\xi_{\tau_a^{b+1}}=0\}$.

当 $p=q$ 时, $\mathrm{P}(A)=\lim_{b\uparrow\infty}\mathrm{P}(\xi_{\tau_a^b}=0)=\lim_{b\uparrow\infty}\frac{b}{a+b}=1$, $\mathrm{E}\tau_0=\lim_{b\uparrow\infty}\mathrm{E}\tau_a^b=\lim_{b\to\infty}\frac{ab}{2p}=\infty$[②].

当 $p>q$ 时, 由 (3.3.11) 得 $\mathrm{P}(A)=\left(\frac{q}{p}\right)^a$, 进一步得 $\mathrm{E}\tau_0=\infty$.

当 $p<q$ 时, 由 (3.3.11) 得 $\mathrm{P}(A)=1$[③], 由 (3.3.13) 得 $\mathrm{E}\tau_0=\frac{a}{q-p}$.

这也说明, 当 $p=q$ 时, $(\xi_n)_{n\ge1}$ 是一个常返马氏链, 平均返回时间是无穷大; 当 $p\ne q$ 时, $(\xi_n)_{n\ge1}$ 是一个非常返马氏链.

例 3.3.5. (利用选样定理证明 Wald 等式)

设 $(\eta_n)_{n\ge1}$ 是独立同分布的随机变量序列, $\mathrm{E}|\eta_1|<\infty$. 设 $\mathcal{F}_n=\sigma(\eta_1,\cdots,\eta_n)$, τ 是 $(\mathcal{F}_n)_{n\ge1}$ 停时, 满足 $\mathrm{E}\tau<\infty$, $\xi_n=\sum_{k=1}^n\eta_k$. 证明:

$$\mathrm{E}\xi_\tau=\mathrm{E}\tau\mathrm{E}\eta_1. \tag{3.3.14}$$

若进一步 $\mathrm{E}\eta_1=0$, $\mathrm{E}\eta_1^2=\sigma^2<\infty$. 证明:

$$\mathrm{E}\xi_\tau^2=\sigma^2\mathrm{E}\tau. \tag{3.3.15}$$

证明: 只证明 (3.3.14), (3.3.15) 的证明留作习题.

$(\xi_n-n\mathrm{E}\eta_1,\mathcal{F}_n)_{n\ge1}$ 是鞅. 我们利用选样定理3.3.1的条件 (2) 只需要验证

1. 先验证 $\mathrm{E}|\xi_\tau-\tau\mathrm{E}\eta_1|<\infty$. 注意到 $\mathrm{E}\tau<\infty$, 因此这需要验证 $\mathrm{E}|\xi_\tau|<\infty$.

$$\begin{aligned}\mathrm{E}|\xi_\tau| = \mathrm{E}\Big|\sum_{k=1}^{\tau}\eta_k\Big| &\le \mathrm{E}\sum_{k=1}^{\tau}|\eta_k| \\ &= \sum_{n=1}^{\infty}\mathrm{E}\sum_{k=1}^{n}|\eta_k|\mathbf{1}_{\{\tau=n\}} \\ &= \sum_{n=1}^{\infty}\sum_{k=1}^{n}\mathrm{E}|\eta_k|\mathbf{1}_{\{\tau=n\}} \\ &= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty}\mathrm{E}|\eta_k|\mathbf{1}_{\{\tau=n\}} \\ &= \sum_{k=1}^{\infty}\mathrm{E}|\eta_k|\mathbf{1}_{\{\tau\ge k\}}.\end{aligned} \tag{3.3.16}$$

因为 $\{\tau\ge k\}=\{\tau<k\}^c\in\mathcal{F}_{k-1}$, 所以 η_k 与 $\{\tau\ge k\}$ 独立, $\mathrm{E}|\eta_k|\mathbf{1}_{\{\tau\ge k\}}=\mathrm{E}|\eta_k|\mathrm{P}\big(\tau\ge k\big)$. 由此得到,

$$\mathrm{E}|\xi_\tau|\le\mathrm{E}|\eta_k|\sum_{k=1}^{\infty}\mathrm{P}\big(\tau\ge k\big)=\mathrm{E}|\eta_k|\mathrm{E}\tau<\infty.$$ [④]

[①] τ_a^b 关于 b 单调递增, 由 $\tau_a^b=\tau_0\wedge\tau_{a+b}$, $\tau_a^b\le\tau_0$, 则 $\lim_{b\uparrow\infty}\tau_a^b\le\tau_0$. 另一方面 $\tau_{a+b}\ge b$, $\tau_a^b\ge b\wedge\tau_0$, 则 $\lim_{b\uparrow\infty}\tau_q^b\ge\tau_0$.

[②] 在例3.3.2 中我们引用简单对称随机游动平均返回时间无穷大的事实得到 $\mathrm{E}\tau_1=\infty$ 来说明问题的. 此处, 我们可以直接利用 (有界) 选样定理得到 $\mathrm{E}\tau_1=\infty$.

[③] 当然也可以通过强大数律得到这个结论.

[④] 事实上, 当 $(\xi_n)_{n\ge1}$ 是非负随机变量时, 上面的计算过程可以直接得到这个 Wald 等式.

2. 再验证 $\lim_{n\to\infty}\mathrm{E}\big(\big|\xi_n - n\mathrm{E}\eta_1\big|\mathbf{1}_{\{\tau>n\}}\big)=0$. 首先,

$$\begin{aligned}\mathrm{E}\big(\big|\xi_n\big|\mathbf{1}_{\{\tau>n\}}\big) &= \mathrm{E}\big(\big|\sum_{k=1}^{n}\eta_k\big|\mathbf{1}_{\{\tau>n\}}\big)\\ &\le \mathrm{E}\big(\sum_{k=1}^{n}|\eta_k|\mathbf{1}_{\{\tau>n\}}\big)\\ &\le \mathrm{E}\big(\sum_{k=1}^{\tau}|\eta_k|\mathbf{1}_{\{\tau>n\}}\big)\end{aligned}\tag{3.3.17}$$

在 (3.3.16) 中已经证明 $\mathrm{E}\sum_{k=1}^{\tau}\big|\eta_k\big|<\infty$, 而且 $\mathrm{E}\tau<\infty$ 意味着 $\mathbf{1}_{\{\tau>n\}}$ 几乎处处收敛到 0. 因此 $\sum_{k=1}^{\tau}\big|\eta_k\big|\mathbf{1}_{\{\tau>n\}}$ 几乎处处收敛到 0, 由 Lebesgue 控制收敛定理

$$\lim_{n\to\infty}\mathrm{E}\sum_{k=1}^{\tau}\big|\eta_k\big|\mathbf{1}_{\{\tau>n\}}=0.$$

由 (3.3.17) 式说明 $\lim_{n\to\infty}\mathrm{E}\big(\big|\xi_n\big|\mathbf{1}_{\{\tau>n\}}\big)=0$.

其次, 因为 $\mathrm{E}\tau=\sum_{m=1}^{\infty}m\mathrm{P}(\tau=m)<\infty$, 所以

$$\begin{aligned}n\mathrm{P}(\tau>n) &\le n\mathrm{P}(\tau\ge n)\\ &= n\sum_{m=n}^{\infty}\mathrm{P}(\tau=m)\\ &\le \sum_{m=n}^{\infty}m\mathrm{P}(\tau=m)\\ &\overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

由此得到 $\lim_{n\to\infty}\mathrm{E}\big(\big|\xi_n - n\mathrm{E}\eta_1\big|\mathbf{1}_{\{\tau>n\}}\big)=0$.

□

例 3.3.6. (更新过程)

设 $(X_n)_{n\ge1}$ 是独立同分布的随机变量序列, $\mathrm{P}(X_1>0)=1$ 和 $\mathrm{E}X_1<\infty$. 记 $\mathrm{E}X_1=\mu$, $\mathcal{F}_n=\sigma(X_1,\cdots,X_n)$. 设

$$T_n = X_1+\cdots+X_n.$$

X_n 表示从第 $n-1$ 个粒子出现到第 n 个粒子出现的持续时间, T_n 表示第 n 粒子出现的时刻.

$$N_t=\sum_{n=1}^{\infty}\mathbf{1}_{\{T_n\le t\}}=\max\{n,T_n\le t\},$$

称为更新过程. N_t 表示出现在区间 $[0,t]$ 上的粒子数, 易知 $\{N_t\ge n\}=\{T_n\le t\}$, 而且第 N_t 个粒子到达时间一定不大于 t, 第 N_t+1 个粒子在时刻 t 还没有到达, 即

$$T_{N_t}\le t<T_{N_t+1},\ t>0.\tag{3.3.18}$$

称 $\mathrm{E}N_t$ 为更新函数, 记为 $m(t)$. 可以证明对于任意的 $t\ge0$, $\mathrm{E}N_t<\infty$[①]. 虽然由强大数律可以得到 $\mathrm{P}(\lim_{t\to\infty}\frac{N_t}{t}=\frac{1}{\mu})=1$, 但并不能由此直接证明更新理论的基本定理, 即

$$\lim_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu}.\tag{3.3.19}$$

下面借助 Wald 等式证明 (3.3.19).

(1) 首先证明 $\liminf_{t\to\infty}\frac{m(t)}{t}\ge\frac{1}{\mu}$.

[①]这个结论的证明非平凡, 严格证明留作习题. 可以这样来直观地看为什么这个结论是有可能能对的, $\mathrm{P}(X_1>0)=1$, 意味着存在 $\delta>0$ 和 $\epsilon>0$ 使得 $\mathrm{P}(X_1>\delta)>\epsilon$, 即 $\mathrm{P}(X_1\le\delta)\le1-\epsilon$. 也就是说在 $[0,t]$ 内出现 m 个粒子的概率 (当 m 很大时) 应该比 $(1-\epsilon)^m$ 小. 因此, $\mathrm{E}N_t=\sum_{m=1}^{\infty}m\mathrm{P}(N_t=m)\approx\sum_{m=1}^{\infty}m(1-\epsilon)^m<\infty$.

事件 $\{N_t=n\}$ 表示 $[0,t]$ 时间段内恰好只有 n 个粒子, 但第 $n+1$ 个粒子还没到, 因此它是被事件域 $\mathcal{F}_{n+1}$ 决定的, 所以直观上 N_t 不是停时, 但是 $\{N_t=n-1\}$ 表示 $[0,t]$ 内有 $n-1$ 个粒子出现, 而第 n 个粒子还没有到达这一事件, 因此, 它应该被 $\mathcal{F}_n$ 决定, 也就是说, $\{N_t+1=n\}\in\mathcal{F}_n$, 这意味着 N_t+1 是 $(\mathcal{F}_n)_{n\geq 1}$ 停时. 我们将上述想法按如下方法严格地写出来:

$$\{N_t+1\leq n\}=\{N_t\leq n-1\}=\{N_t<n\}=\{T_n>t\}=\Big\{\sum_{k=1}^n X_k>t\Big\}=\Big\{\sum_{k=1}^n X_k\leq t\Big\}^c\in\mathcal{F}_n,$$

由此证明了 N_t+1 是停时. 由 Wald 等式

$$\mathrm{E}T_{N_t+1}=\mu(m(t)+1).$$

利用 (3.3.18), 得到 $t<\mu(m(t)+1)$, 因此

$$\frac{m(t)}{t}>\frac{1}{\mu}-\frac{1}{t}\Longrightarrow\liminf_{t\to\infty}\frac{m(t)}{t}\geq\frac{1}{\mu}.$$

(2) 其次证明 $\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\mu}$. 先设 X_n 是有界随机变量, 即存在常数 M 使得 $\mathrm{P}(X_n<M)=1$. 注意到 $T_{N_t+1}=T_{N_t}+X_{N_t+1}$, 由 (3.3.18), 得到

$$t\geq\mathrm{E}T_{N_t}=\mathrm{E}\big(T_{N_t+1}-X_{N_t+1}\big)=\mu(m(t)+1)-\mathrm{E}X_{N_t+1}.$$

因此,

$$t\geq\mu(m(t)+1)-M\Longrightarrow\frac{m(t)}{t}\leq\frac{1}{\mu}-\frac{1}{t}\frac{M-\mu}{\mu}\Longrightarrow\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\mu}.$$

对于一般的 X_n, 定义 $X_n^M=X_n\mathbf{1}_{\{X_n\leq M\}}+M\mathbf{1}_{\{X_n>M\}}$, $T_n^M=\sum_{k=1}^n X_k^M$, $N_t^M=\sum_{n=1}^\infty\mathbf{1}_{\{T_n^M\leq t\}}$. 因为 $X_n^M\leq X_n$, 所以

$$N_t^M\geq N_t,\ m^M(t)=\mathrm{E}N_t^M\geq\mathrm{E}N_t=m(t).$$

设 $\mu^M=\mathrm{E}X_1^M$,

$$\frac{m(t)}{t}\leq\frac{m^M(t)}{t}\leq\frac{1}{\mu^M}+\frac{1}{t}\frac{M-\mu^M}{\mu^M}\Longrightarrow\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\mu^M}.$$

注意到 $\lim_{M\to\infty}\mathrm{E}X_1^M=\mathrm{E}X_1=\mu$, 由此得 $\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\mu}$.

□

3.3.3 Doob 停时定理

对于事件域流 $(\mathcal{F}_t)_{t\in T}$, 任意时刻 t, $\mathcal{F}_t$ 为在 t 时刻之前的信息, 那么对于一个停时 τ, 在 τ 之前可以掌握的"信息"(随机事件全体) 又是什么? 令 $\mathcal{F}_\infty=\sigma\big(\bigcup_{t\in T}\mathcal{F}_t\big)$.

定义 3.3.4. $(\mathcal{F}_t)_{t\in T}$ 停时 τ 的事前事件域

$$\begin{aligned}\mathcal{F}_\tau &\equiv \Big\{A\in\mathcal{F}_\infty\Big|\ \forall t\in T,\ A\cap\{\tau\leq t\}\in\mathcal{F}_t\Big\}\\ &= \Big\{A\in\mathcal{F}_\infty\Big|\ \forall t\in T,\ \mathbf{1}_{A\cap\{\tau\leq t\}}\text{为}\mathcal{F}_t\text{可测的随机变量}\Big\}\end{aligned}$$

特别地, 当 $T=\mathbb{Z}^+$,

$$\mathcal{F}_\tau\equiv\Big\{A\in\mathcal{F}_\infty\Big|\ \forall n\in\mathbb{Z}^+,\ A\cap\{\tau=n\}\in\mathcal{F}_n\Big\}.$$

解释:

(1) 先考察 $\mathcal{F}_t$. 对于确定性的时刻 t, 对于任意的 $\forall A\in\mathcal{F}_t$, "到了"时刻 t, 总能判断每一个样本点是否在 A 中.

(2) 对于停时 τ, 在任意时刻 t, 应该首先判断这个停时 τ 是否已经"到达", 即 $\{\tau(\omega)\leq t\}$ 是否已经发生. 如果 $\{\tau(\omega)\leq t\}$ 发生, 再看事件 A 是否属于时刻 t 已知的信息, $A\cap\{\tau(\omega)\leq t\}\in\mathcal{F}_t$. 也就是说对于停时 τ 到了时刻 t, 总能判断每一个样本是否在 $A\cap\{\tau(\omega)\leq t\}$ 中, 因此, $\mathcal{F}_\tau$ 表示 τ 之前知道的"信息".

性质 3.3.2. 若 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 适应过程, τ 为 $(\mathcal{F}_t)$ 停时,

(1) 若 $T=\mathbb{Z}^+$, 则 ξ_τ 为 $\mathcal{F}_\tau$ 可测的随机变量;

(2) 若 $T=\mathbb{R}^+$, 且 $(\xi_t)_{t\geq 0}$ 为右连续过程, 即 $\forall\omega\in\Omega$, $\xi_t(\omega)$ 为右连续函数, 则 ξ_τ 为 $\mathcal{F}_\tau$ 可测的随机变量.

性质 3.3.3. 设 τ 和 σ 为 $(\mathcal{F}_t)_{t\in T}$ 停时. 若 $\sigma\leq\tau$, 则 $\mathcal{F}_\sigma\subset\mathcal{F}_\tau$.

为什么会有选样定理? "鞅" 是一个 "纯随机" 的过程, 是一种 "客观现象". 它的这种 "客观属性" 不应该依赖于 "时间坐标系", 即 $(\mathcal{F}_t)_{t\in T}$ 的选取, 也就是说换成另一个关于 $(\mathcal{F}_t)_{t\in T}$ 适应的停时列 $\{\tau_0\leq\tau_1\leq\tau_2\leq\cdots\leq\tau_n\cdots\}$ 以及 $\{\tau_n\}_{n\in\mathbb{Z}}$ 的事前事件域流 $(\mathcal{F}_{\tau_n})_{n\in\mathbb{Z}}$, 它的 "鞅性" 在一定条件下还应该保持. 具体地说, 若 $(\xi_t,\mathcal{F}_t)_{t\in T}$ 为鞅, 即对于任意的 $t\geq s$, $\mathrm{E}[\xi_t|\mathcal{F}_s]=\xi_s$, 那么当停时满足一定的条件下, $\mathrm{E}[\xi_{\tau_{n+m}}|\mathcal{F}_{\tau_n}]=\xi_{\tau_n}$. 这就是 Doob 停时定理的内在含义.

定理 3.3.2. (Doob 有界停时定理)

$(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, $\sigma\leq\tau<M<\infty$ 为两个有界 $(\mathcal{F}_t)_{t\in T}$ 停时, 则

$$\mathrm{E}[\xi_\tau|\mathcal{F}_\sigma]=\xi_\sigma.$$

证明: 仅对 $T=\mathbb{Z}^+$ 时证明. 首先, 要证 $\mathrm{E}|\xi_\tau|<\infty$, $\mathrm{E}|\xi_\sigma|<\infty$,

$$\mathrm{E}|\xi_\tau|=\mathrm{E}\Big|\sum_{k=0}^{M}\xi_k\mathbf{1}_{\{\tau=k\}}\Big|\leq\sum_{k=0}^{M}\mathrm{E}|\xi_k|<\infty.$$

同理, $\mathrm{E}|\xi_\sigma|<\infty$.

其次, 我们只需证明 $\forall A\in\mathcal{F}_\sigma$, $\mathrm{E}\big(\xi_\tau\mathbf{1}_A\big)=\mathrm{E}\big(\xi_\sigma\mathbf{1}_A\big)$ 就可以了. 取 M 为正整数, 则

$$\mathrm{E}\big(\xi_\tau\mathbf{1}_A\big)=\sum_{k=0}^{M}\mathrm{E}\big(\xi_\tau\mathbf{1}_{\{\tau=k\}}\mathbf{1}_A\big)=\sum_{k=0}^{M}\mathrm{E}\big(\xi_k\mathbf{1}_{A\cap\{\tau=k\}}\big).$$

因为 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, 所以对于任意的 $B\in\mathcal{F}_k$, $\mathrm{E}(\xi_M\mathbf{1}_B)=\mathrm{E}(\xi_k\mathbf{1}_B)$; 同时, 由于 $\sigma\leq\tau$, 那么 $\mathcal{F}_\sigma\subset\mathcal{F}_\tau$, 即 $A\in\mathcal{F}_\tau$, 故而由停时 τ 的事前事件域 $\mathcal{F}_\tau$ 的定义知道, 对于任意的 k, $A\cap\{\tau=k\}\in\mathcal{F}_k$, 因此, 我们得到

$$\sum_{k=0}^{M}\mathrm{E}\big(\xi_k\mathbf{1}_{A\cap\{\tau=k\}}\big)=\sum_{k=0}^{M}\mathrm{E}\big(\xi_M\mathbf{1}_{A\cap\{\tau=k\}}\big)=\mathrm{E}\big(\xi_M\mathbf{1}_A\big).$$

同理 $\mathrm{E}\big(\xi_\sigma\mathbf{1}_A\big)=\mathrm{E}\big(\xi_M\mathbf{1}_A\big)$.

□

♣♣♣20230404

很显然, 选样定理3.3.1的 (1) 是 Doob 有界停时定理的推论. 在选样定理3.3.1表述之后, 我们解释了它的直观意思是对于一个鞅 (公平博弈), 不可能通过有界停时策略增加个人财富的期望; 反过来, 如果一个博弈, 对于任何有界停时策略都不可能增加个人财富的期望, 那么, 直观上这个博弈应该是 "公平" 的, 也就是鞅. 这个直观是正确的, 可以表述为如下的定理.

定理 3.3.3. 若 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 适应过程且对于任意的 $t\in T$, $\mathrm{E}|\xi_t|<\infty$, $\sigma\leq\tau$ 是任意两个有界停时, 则 $(\xi_t,\mathcal{F}_t)_{t\in T}$ 是鞅当且仅当

$$\mathrm{E}\xi_\sigma=\mathrm{E}\xi_\tau.$$

上述定理是有界选样定理一种深刻的推广, 它给出鞅的一种等价表述.

定理 3.3.4. (Doob 停时定理)

若 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, 且满足存在随机变量 ξ, $\mathrm{E}|\xi|<\infty$, 使得 $\xi_t=\mathrm{E}[\xi|\mathcal{F}_t]$, 则任意两个停时 $\sigma\leq\tau$

$$\mathrm{E}[\xi_\tau|\mathcal{F}_\sigma]=\xi_\sigma.$$

在此我们强调, 当 $T=\mathbb{R}^+$ 时, 要求 $(\xi_t)_{t\in\mathbb{R}^+}$ 右连续.

定理 3.3.5. $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 鞅, 则对于任意 $(\mathcal{F}_n)$ 停时 τ, $(\xi_{\tau\wedge n}, n\geq 0)$ 是 $(\mathcal{F}_{\tau\wedge n})$ 鞅. 事实上, 还可以证明 $(\xi_{\tau\wedge n})_{n\geq 1}$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 鞅.

定理 3.3.6. 若 $(\xi_t,\mathcal{F}_t)_{t\in T}$ 为下鞅, $\sigma\leq\tau<M$ 为两个有界 $(\mathcal{F}_t)$ 停时, 则

$$\mathrm{E}[\xi_\tau|\mathcal{F}_\sigma]\geq\xi_\sigma.$$

3.4 鞅不等式

定理 3.4.1. (Doob 不等式 I) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是下鞅, 对任意的正整数 N, 对任意的 $a>0$, 则

$$a\mathrm{P}(\max_{0\leq n\leq N}\xi_n\geq a)\leq \mathrm{E}\big(\xi_N\mathbf{1}_{\{\max_{0\leq n\leq N}\xi_n\geq a\}}\big)\leq \mathrm{E}\xi_N^+.$$

$$a\mathrm{P}(\min_{0\leq n\leq N}\xi_n\leq -a)\leq -\mathrm{E}\xi_0+\mathrm{E}\big(\xi_N\mathbf{1}_{\{\min_{0\leq n\leq N}\xi_n> -a\}}\big)\leq -\mathrm{E}\xi_0+\mathrm{E}\xi_N^+.$$

其中, $\xi_N^+=\max(0,\xi_N)$.

证明: 只证明第一个不等式, 第二个留作习题.

设 $\tau=\inf\{n\geq 0,\,\xi_n\geq a\}\wedge N$, $A=\{\max_{0\leq n\leq N}\xi_n\geq a\}$. 由 Doob 停时定理,

$$\mathrm{E}\xi_N\geq\mathrm{E}\xi_\tau=\mathrm{E}\xi_\tau\mathbf{1}_A+\mathrm{E}\xi_\tau\mathbf{1}_{A^c}\geq a\mathrm{P}(A)+\mathrm{E}\xi_N\mathbf{1}_{A^c},$$

因此,

$$a\mathrm{P}(A)\leq\mathrm{E}\xi_N-\mathrm{E}\xi_N\mathbf{1}_{A^c}=\mathrm{E}\xi_N\mathbf{1}_A\leq\mathrm{E}\xi_N^+.$$

□

定理 3.4.2. (Doob 不等式 II) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是鞅, 对某个 $p\geq 1$, 任意的 $n\geq 0$, 有 $\mathrm{E}|\xi_n|^p<\infty$, 则对任意的 $a>0$ 和正整数 N,

$$\mathrm{P}(\max_{0\leq n\leq N}|\xi_n|\geq a)\leq\frac{\mathrm{E}|\xi_N|^p}{a^p}.$$

对于任意的 $a,b>0$,

$$\mathrm{P}(\max_{0\leq n\leq N}\xi_n\geq a)\leq\frac{\mathrm{E}\mathrm{e}^{b\xi_N}}{\mathrm{e}^{ba}}.$$

例 3.4.1. 设 $(\eta_n)_{n\geq 1}$ 是独立同分布的随机变量序列, $\mathrm{P}(\eta_1=1)=\mathrm{P}(\eta_1=-1)=\frac{1}{2}$. 设 $\xi_n=\sum_{k=1}^n\eta_k$, $(\xi_n)_{n\geq 1}$ 是鞅. 取 $b=\frac{1}{\sqrt{N}}$, 则

$$\mathrm{P}(\max_{0\leq n\leq N}\xi_n\geq a\sqrt{N})\leq\mathrm{e}^{-a}\mathrm{E}(\mathrm{e}^{\frac{\xi_N}{\sqrt{N}}}).$$

而

$$\mathrm{E}(\mathrm{e}^{\frac{\xi_N}{\sqrt{N}}})=\mathrm{E}(\mathrm{e}^{\frac{\eta_1+\cdots+\xi_N}{\sqrt{N}}})=\big(\mathrm{E}(\mathrm{e}^{\frac{\eta_1}{\sqrt{N}}})\big)^N=\Big(\frac{\mathrm{e}^{\frac{1}{\sqrt{N}}}+\mathrm{e}^{-\frac{1}{\sqrt{N}}}}{2}\Big)^N.$$

由泰勒级数展开

$$\Big(\frac{\mathrm{e}^{\frac{1}{\sqrt{N}}}+\mathrm{e}^{-\frac{1}{\sqrt{N}}}}{2}\Big)^N=1+\frac{1}{2N}+O(\frac{1}{N^2}),$$

由此得到

$$\lim_{N\to\infty}\mathrm{E}(\mathrm{e}^{\frac{\xi_N}{\sqrt{N}}})=\lim_{N\to\infty}\Big(1+\frac{1}{2N}+O(\frac{1}{N^2})\Big)^N=\mathrm{e}^{\frac{1}{2}}.$$

因此, 存在常数 $C>0$ 使得对于充分大的 N 和任意的 $a>0$

$$\mathrm{P}(\max_{0\leq n\leq N}\xi_n\geq a\sqrt{N})\leq C\mathrm{e}^{-a}.$$

定理 3.4.3. (Doob 不等式 III, 极大 L^p 不等式) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是鞅或非负下鞅, 任意的 $n\geq 0$, $\mathrm{E}|\xi_n|^p<\infty$, $1<p<\infty$, 则

$$\big(\mathrm{E}(\max_{0\leq n\leq N}|\xi_n|^p)\big)^{\frac{1}{p}}\leq\frac{p}{p-1}\big(\mathrm{E}|\xi_N|^p\big)^{\frac{1}{p}}.$$

3.5 鞅收敛定理

$(\xi_n)_{n\geq 0}$ 是一个关于 $(\mathcal{F}_n)_{n\geq 0}$ 的适应过程, 设 $a<b$, 定义一列停时

$$\begin{aligned}
\tau_0 &= 0,\\
\tau_1 &= \inf\{n\geq 0,\ \xi\leq a\},\\
\tau_2 &= \inf\{n\geq \tau_1,\ \xi\geq b\},\\
\tau_3 &= \inf\{n\geq \tau_2,\ \xi\leq a\},\\
\tau_4 &= \inf\{n\geq \tau_3,\ \xi\geq b\},\\
&\cdots\\
\tau_{2m-1} &= \inf\{n\geq \tau_{2m-2},\ \xi_n\leq a\},\\
\tau_{2m} &= \inf\{n\geq \tau_{2m-1},\ \xi_n\geq b\}.
\end{aligned}$$

定义 ξ_n 在时刻 N 之前的上穿次数, 参考图3.1,

$$U_N[a,b]=\begin{cases}0, & \tau_2>N,\\ \max\{m,\ \tau_{2m}\leq N\}, & \tau_2\leq N.\end{cases}$$

定理 3.5.1. (Doob 上穿不等式) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是下鞅, 则对于任意的 $N\geq 1$,

$$\mathrm{E}U_N[a,b]\leq\frac{\mathrm{E}(\xi_N-a)^+}{b-a}. \tag{3.5.1}$$

证明: 下鞅 $(\xi_n,\mathcal{F}_n)_{n\geq 1}$ 在区间 $[a,b]$ 的上穿次数等于下鞅 $((\xi_n-a)^+,\mathcal{F}_n)$ 在 $[0,b-a]$ 的上穿次数, 参考图3.1. 因此, 我们只需证明下述命题即可.

如果 $(\xi_n)_{n\geq 0}$ 为非负下鞅, 则 $\mathrm{E}U_N(0,b)\leq\frac{\mathrm{E}\xi_N}{b}$. 设

$$\phi_i=\begin{cases}1 & \tau_m<i\leq\tau_{m+1},\ \text{且 } m \text{ 为奇数}\\ 0 & \tau_m<i\leq\tau_{m+1},\ \text{且 } m \text{ 为偶数}\end{cases}.$$

若 m 是满足 $2m\leq N$ 的最大正整数, 可以参考图3.1, 我们得到,

$$\begin{aligned}
U_N(0,b) &\leq \frac{\xi_{\tau_2}-\xi_{\tau_1}}{b}+\frac{\xi_{\tau_4}-\xi_{\tau_3}}{b}+\cdots\frac{\xi_{\tau_{2m}}-\xi_{\tau_{2m-1}}}{b}\\
&= \frac{1}{b}\sum_{i=1}^{N}\phi_i(\xi_i-\xi_{i-1})
\end{aligned}$$

同时,

$$\{\phi_i=1\}=\bigcup_{m\text{ 为奇数}}\big\{\{\tau_m<i\}\setminus\{\tau_{m+1}<i\}\big\}\in\mathcal{F}_{i-1},$$

也就是说 (ϕ_i) 是可料序列. 设 $\eta_n=\sum_{i=1}^n\phi_i(\xi_i-\xi_{i-1})$, 注意到 $\xi_n=\sum_{i=1}^n(\xi_i-\xi_{i-1})$, $1-\phi_i\geq 0$, 那么 $\xi_n-\eta_n=\sum_{i=1}^n(1-\phi_i)(\xi_i-\xi_{i-1})$, 利用例3.1.2中的结论, $(\xi_n-\eta_n,\mathcal{F}_n)$ 是下鞅, 因此 $\mathrm{E}(\xi_n-\eta_n)\geq\mathrm{E}(\xi_0-\eta_0)=\mathrm{E}\xi_0\geq 0$, 这就意味着

$$\mathrm{E}U_N(0,b)\leq\frac{1}{b}\mathrm{E}\eta_N\leq\frac{1}{b}\mathrm{E}\xi_N.$$

□

定理 3.5.2. (Doob 下鞅收敛定理) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是下鞅, 且 $\sup_n\mathrm{E}|\xi_n|<\infty$, 则 $\lim_{n\to\infty}\xi_n=\xi_\infty$ 几乎处处收敛, 且 $\mathrm{E}|\xi_\infty|<\infty$.

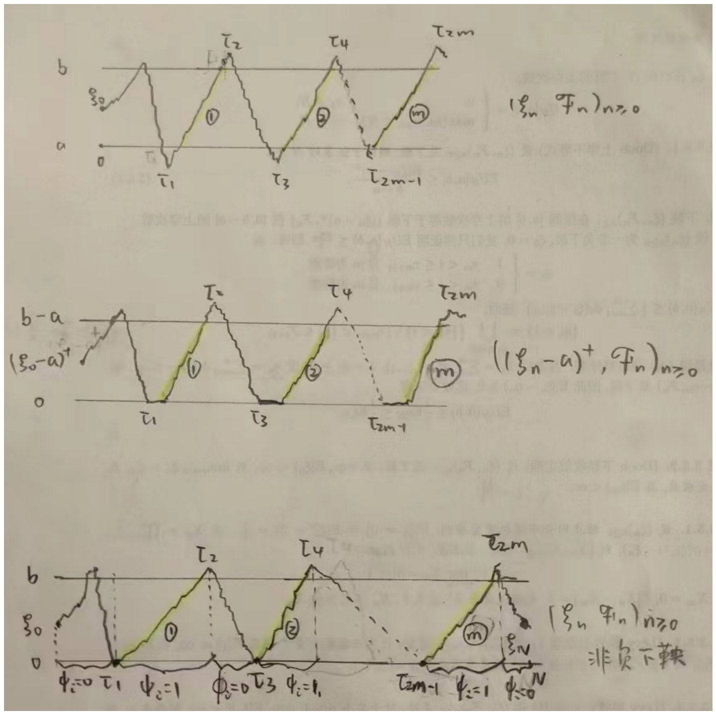


图 3.1: 上传不等式示意图

证明: 未完待续 □

例 3.5.1. 设 $(\xi_n)_{n\geq 1}$ 独立同分布随机变量序列, $\mathrm{P}(\xi_1 = 0) = \mathrm{P}(\xi_1 = 2) = \frac{1}{2}$. 设 $X_n = \prod_{k=1}^n \xi_k$, $\mathcal{F}_N = \sigma(\xi_1, \cdots, \xi_n)$, 则 $(X_n, \mathcal{F}_n)_{n\geq 1}$ 是鞅, 且 $\mathrm{E}X_n = 1$, 因此

$$\mathrm{P}(\lim_{n\to\infty} X_n = 0) = 1.$$

若记 $X_\infty = 0$, $\mathrm{E}|X_n - X_\infty| = 1$, 也就是说在 L^1 意义下, X_n 不收敛到 X_∞.

推论 3.5.1. 设 $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ 是非负鞅, 则存在 ξ_∞ 使得 $\lim_{n\to\infty} \xi_n = \xi_\infty$ 几乎处处收敛, 且 $\mathrm{E}\xi_\infty < \infty$.

推论 3.5.2. 设 $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ 是非负上鞅, 则存在 ξ_∞ 使得 $\lim_{n\to\infty} \xi_n = \xi_\infty$ 几乎处处收敛, 且 $\mathrm{E}\xi_\infty < \infty$.

定理 3.5.3. (Lévy 鞅收敛定理 I) 设 $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ 是鞅, 且存在随机变量 ξ 满足 $\mathrm{E}|\xi| < \infty$, 使得 $\xi_n = \mathrm{E}[\xi|\mathcal{F}_n]$, 则 ξ_n 在几乎处处和 L^1 的意义下收敛到 $\mathrm{E}[\xi|\mathcal{F}_\infty]$.

定理 3.5.4. (Lévy 鞅收敛定理 II) 设 $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ 是鞅, 对于某个 $p > 1$, $\sup_n \mathrm{E}|\xi_n|^p < \infty$, 则存在 ξ, 使得 $\mathrm{E}[\xi|\mathcal{F}_n] = \xi_n$, [①] 则 ξ_n 在几乎处处和 L^1 的意义下收敛到 $\mathrm{E}[\xi|\mathcal{F}_\infty]$.

例 3.5.2. 设 $(\xi_n)_{n\geq 1}$ 独立同分布随机变量序列, $\mathrm{P}(\xi_1 = 1) = \mathrm{P}(\xi_1 = -1) = \frac{1}{2}$. 记 $X_n = \sum_{k=1}^n \frac{\xi_k}{k}$, $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$ 则 $(X_n, \mathcal{F}_n)_{n\geq 1}$ 是鞅.

$$\mathrm{E}X_n^2 = \mathrm{E}\xi_1^2 \sum_{k=1}^n \frac{1}{k^2} < \infty.$$

由 Lévy 鞅收敛定理 II 知道存在 X_∞ 使得

$$X_n = \mathrm{E}[X_\infty|\mathcal{F}_n], \text{ 而且 } \mathrm{P}(\lim_{n\to\infty} X_n = X_\infty) = 1, \ \mathrm{E}|X_n - X_\infty| = 0.$$

[①] 一般的是 $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ 是一致可积鞅时, 则存在 ξ, 使得 $\mathrm{E}[\xi|\mathcal{F}_n] = \xi_n$.

♣♣♣20230406

✠✠✠✠✠✠

第四章　布朗运动和扩散过程的分析方法初步

4.1　布朗运动的定义

以下是大家熟知的布朗运动的定义.

定义 4.1.1. 设 $(B_t)_{t\geq 0}$ 是一个随机过程, 如果它满足以下三条:

1. $B_{t+s}-B_t$ 服从正态分布 $N(0,\sigma^2 s)$;
2. 对于任意 $0<t_1<t_2<\cdots<t_n$, $B_0, B_{t_1}-B_0, \cdots, B_{t_n}-B_{t_{n-1}}$ 相互独立;
3. 对于所有的 ω, $B_t(\omega)$ 关于 t 连续.

则称 $(B_t)_{t\geq 0}$ 是一个 (一维) 布朗运动 (Brownian motion). 在本讲义中, 我们总假定布朗运动满足 $\sigma^2=1$; 当 $B_0=0$ 时, 称为标准布朗运动.

为什么会有这个定义呢? 我们回到 1905 年 Einstein 论文[①], 假定 (花粉) 粒子的运动是由分子的热运动引起的不规则运动, Einstein 做了如下的一些假定:

1. 每一个粒子所进行的运动, 同其他一切粒子的运动都无关;
2. 同一个粒子在各个不同时间间隔中的运动, 都必须被看作是相互独立的过程, 只要我们设想选择的这些时间间隔不要太小就行了.

   用现在随机过程的话来说, 就是粒子运动是独立增量过程. 做这样的假定是基于观测时间间隔在宏观尺度上小, 但在微观尺度大, 因此可以忽略掉一些在非常小尺度上的物理细节.

3. 假设液体中共有 n 个粒子, 经过时间间隔 τ, 单个粒子的 X 坐标将要增加 Δ, 此处 Δ 对于每个粒子都有一个不同的值 (可正可负), 对于 Δ, 某种分布定律成立. 在时间间隔 τ 内经历了处于 Δ 和 $\Delta+\mathrm{d}\Delta$ 之间位移的粒子数 $\mathrm{d}n$, 可由如下形式的方程表示
$$\mathrm{d}n = n\varphi(\Delta)\mathrm{d}\Delta, \quad 此处 \int_{-\infty}^{+\infty}\varphi(\Delta)\mathrm{d}\Delta = 1.$$
   而 φ 只是对非常小的 Δ 值才不是零, 并且满足条件 $\varphi(\Delta)=\varphi(-\Delta)$.

   $\varphi(\Delta)\mathrm{d}\Delta$ 其实就是在时间间隔 τ 内经历了处于 Δ 和 $\Delta+\mathrm{d}\Delta$ 之间位移的粒子数占总粒子数的比例, 而这个比例正比于区间间隔 $\mathrm{d}\Delta$, 比例系数与坐标增量区间的位置 Δ 有关.

[①] 参见:

[1] Einstein, A,: *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeitensus-pendienrten Teilchen.* Ann. Phys. 17 (1905) 549-560

中译文参见:

热的分子运动论所要求的静液体中悬浮粒子的运动. 爱因斯坦文集 第二卷　范岱年 赵中立 许良英 编译 商务印书馆 1977, 北京 p72-82

[2] Einstein, A,: *Investigations on the theory of the Brownian movement.* Edited by R. Fürth, Dover, Now York 1956

原文发表在　Ann. Phys. 19 (1906) 371-381

中译文参见:

关于布朗运动的理论. 爱因斯坦文集 第二卷　范岱年 赵中立 许良英 编译 商务印书馆 1977, 北京 p119-129

4. 假设单位体积的粒子数 ν 只同 x 和 t 有关, 设

$$\nu = f(x,t).$$

也就是说处于 x 位置, 在时刻 t 粒子数的密度是 $f(x,t)$.

我们要从粒子在 t 时刻分布计算出 $t+\tau$ 时刻的分布. 由 $\varphi(\Delta)$ 的定义可以计算出 $t+\tau$ 时刻位于 X 轴上 $[x,\ x+\mathrm{d}x]$ 之间的粒子数,

$$\begin{aligned} f(x,t+\tau)\mathrm{d}x &= \int_{-\infty}^{+\infty} f(x-\Delta,t)\mathrm{d}x\varphi(\Delta)\mathrm{d}\Delta \\ &= \int_{-\infty}^{+\infty} f(x-\Delta,t)\varphi(\Delta)\mathrm{d}\Delta\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f(x+\Delta,t)\varphi(\Delta)\mathrm{d}\Delta\mathrm{d}x. \end{aligned} \tag{4.1.1}$$

利用泰勒展开, τ 很小, 因此可以设

$$f(x,t+\tau) = f(x,t) + \tau\frac{\partial f}{\partial t}. \tag{4.1.2}$$

对 $f(x+\Delta,t)$ 关于 x 变量用泰勒展开

$$f(x+\Delta,t) = f(x,t) + \Delta\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 f(x,t)}{\partial x^2} + \cdots \tag{4.1.3}$$

$$\begin{aligned} & f(x,t) + \frac{\partial f}{\partial t}\tau \\ = \ & f(x,t)\cdot\int_{-\infty}^{+\infty}\varphi(\Delta)\mathrm{d}\Delta + \frac{\partial f}{\partial x}\int_{-\infty}^{+\infty}\Delta\varphi(\Delta)\mathrm{d}\Delta + \frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\varphi(\Delta)\mathrm{d}\Delta + \cdots \end{aligned} \tag{4.1.4}$$

既然 $\varphi(\Delta) = \varphi(-\Delta)$, 则上面偶数项为 0. 因为只有很小的 Δ 值才对积分有贡献, 所以第 5,7,$\cdots$ 项都比第 1 项和第 3 项小很多. 设

$$\frac{1}{\tau}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\varphi(\Delta)\mathrm{d}\Delta = D.$$

只考虑 (4.1.4) 式右端中的第 1 项和第 3 项, 有

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}.$$

假定在 $t=0$ 时刻, n 个粒子都集中在 $x=0$ 点, 即

$$f(x,0) = \begin{cases} 0, & x \neq 0, \\ \int_{-\infty}^{+\infty} f(x,0)\mathrm{d}x = n. & \end{cases}$$

因此

$$f(x,t) = \frac{n}{\sqrt{4\pi Dt}}\mathrm{e}^{-\frac{x^2}{4Dt}}.$$

$f(x,t)\mathrm{d}x$ 表示 t 时刻 $[x,x+\mathrm{d}x]$ 内粒子数, $\dfrac{f(x,t)\mathrm{d}x}{n} = \dfrac{1}{\sqrt{4\pi Dt}}\mathrm{e}^{-\frac{x^2}{4Dt}}\mathrm{d}x$ 表示 t 时刻在 $[x,x+\mathrm{d}x]$ 内粒子数占总粒子数的比例. 我们现在给出 $\frac{f(x,t)\mathrm{d}x}{n}$ 概率论的解释 (实质上是“偷换概念”) 就是在 t 时刻在 $[x,x+\mathrm{d}x]$ 内发现某个粒子的概率, 这就对应了布朗运动.

进一步计算, 一个粒子在 X 轴方向上平均经历的位移 λ_X, 或者比较准确的说就是在 X 轴方向上位移平方的算术平均的平方根

$$\lambda_X = \sqrt{\overline{X^2}} = \sqrt{2Dt}. \ (\text{平均自由程})$$

从概率论的角度上面这段话可以理解为, 利用时间 $[0,t]$ 内多个粒子位移的算术平均替代一个粒子在任意时间段 $[s,s+t]$ 的平均位移, 利用多个粒子位移平方的算术平均替代一个粒子在任意时间段 $[s,s+t]$ 的位移平方的平均, 即

$$\overline{X} = \int xf(x,t)\mathrm{d}x = 0, \quad (\text{均值})$$

$$\sqrt{\overline{X^2}}=\left(\int x^2 f(x,t)\mathrm{d}x\right)^{\frac{1}{2}}=\sqrt{2Dt}.\quad(\text{标准差})$$

利用平均自由程 $\sqrt{\overline{X^2}}$ 来表示平均位移的大小. 因此平均位移与时间的平方根成正比, $\sqrt{2Dt}\sim\sqrt{t}$. 这点很重要 [1].

由此我们抽象出布朗运动应该具有以下的性质:

1. 粒子轨道连续;
2. $\forall 0<t_1<t_2<\cdots<t_n$, 在 $[0,t_1],(t_1,t_2],\cdots,(t_{n-1},t_n]$ 之间过程独立;
3. 在 $[0,\tau]$ 和 $[t,t+\tau]$ 间运动规律相同;
4. 粒子运动规律关于 0 点对称.

我们将其严格写成数学语言, 就有下面的定理.

定理 4.1.1. (一维情况) 设随机过程 $(B_t)_{t\geq 0}$ 满足:

(1) $\forall 0<t_1<t_2<\cdots<t_n$, $B_0,B_{t_1}-B_0,\cdots,B_{t_n}-B_{t_{n-1}}$ 相互独立;

(2) $\forall s$, B_t-B_0 和 $B_{t+s}-B_s$ 同分布;

(3) 对于所有的 ω, $B_t(\omega)$ 关于 t 连续;

(4) $\forall t$, B_t-B_0 的分布关于 0 点对称.

则 $(B_t)_{t\geq 0}$ 一定是布朗运动或者 B_t-B_0 恒等于 0.[2]

推论 4.1.1. 若进一步只假定 $(B_t)_{t\geq 0}$ 只满足上面的 (1)(2)(3), 则 $(B_t)_{t\geq 0}$ 一定可以表示成如下形式,

$$B_t=B_0+\sigma\bar{B}_t+\mu t,\tag{4.1.5}$$

其中 $(\bar{B}_t)_{t\geq 0}$ 为一个标准布朗运动, σ 和 μ 为常数. 当 σ 和 μ 都不为零时, $(B_t)_{t\geq 0}$ 称为漂移布朗运动.

定义 4.1.2. (n 维标准布朗运动) $(B_t^{(1)})_{t\geq 0},\cdots,(B_t^{(n)})_{t\geq 0}$ 为 n 个相互独立的一维标准布朗运动, 则称 $\left(B_t^{(1)},\cdots,B_t^{(n)}\right)_{t\geq 0}$ 为 n 维标准布朗运动.

推论 4.1.2. 对于 n 维随机过程 $\left(B_t^{(1)},\cdots,B_t^{(n)}\right)_{t\geq 0}$, 满足

(1) $\forall 0<t_1<t_2<\cdots<t_n$, $B_0,B_{t_1}-B_0,\cdots,B_{t_n}-B_{t_{n-1}}$ 相互独立;

(2) $\forall s$, B_t-B_0 和 $B_{t+s}-B_s$ 同分布;

(3) 对于所有的 ω, $B_t(\omega)$ 关于 t 连续,

---

[1] 花粉粒子的位移大致满足

$$X(s+t)-X(s)\sim\sqrt{t},\quad \frac{X(s+t)-X(s)}{t}\approx\frac{1}{\sqrt{t}}.$$

因此, 当时间间隔很小时, 花粉粒子运动方向变化极快速度无穷大, 这就是观测到所谓的 “运动的不规则性”. 从数学上看就是花粉粒子的运动轨迹每一点都不可微.

[2] 这才是布朗运动本质的定义. 在定理4.1.1中, 我们没有做任何定量的假设, 只是尽可少地在几个定性的假设下得到了布朗运动. 这从一个侧面说明了布朗运动这个概念的深刻性与广泛性.

这个定理的证明可以参见

概率论基础和随机过程, 王寿仁 编著 科学出版社 1986

第四章第 7 节的内容, 例如, 定理 4.7.1.

则

$$B_t = B_0 + A \cdot \bar{B}_t + b \cdot t, \tag{4.1.6}$$

其中 A 为一个 $n \times r$ 常数矩阵, $\bar{B}_t$ 为 r 维标准布朗运动, $b = (b_1, \cdots, b_n)$ 为一个 n 维常数向量. 当 A 不是零矩阵, b 不是零向量时, 称 (4.1.6) 为漂移布朗运动.

(4.1.5) 式和 (4.1.6) 式也可以认为是 Doob 分解定理.

4.2 布朗运动的一些性质

可以将布朗运动的定义推广至带流的概率空间 $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathrm{P})$.

定义 4.2.1. 设 $(B_t)_{t\geq 0}$ 关于事件域流 $(\mathcal{F}_t)_{t\geq 0}$ 适应, $B_0 = 0$,

1. $B_{t+s} - B_t$ 服从正态分布 $N(0, s)$;
2. 对于任意的 $0 \leq s < t$, $B_t - B_s$ 与 $\mathcal{F}_s$ 独立; [①]
3. 对于所有的 ω, $B_t(\omega)$ 关于 t 连续,

则称 $(B_t)_{t\geq 0}$ 是关于事件域流 $(\mathcal{F}_t)_{t\geq 0}$ 的 (一维) 标准布朗运动. $(B_t^{(1)})_{t\geq 0}, \cdots, (B_t^{(n)})_{t\geq 0}$ 为 n 个相互独立的 (一维)$(\mathcal{F}_t)_{t\geq 0}$ 标准布朗运动, 则称 $\left(B_t^{(1)}, \cdots, B_t^{(n)}\right)_{t\geq 0}$ 为 n 维 $(\mathcal{F}_t)_{t\geq 0}$ 标准布朗运动.

性质 4.2.1. 对于 $(\mathcal{F}_t)_{t\geq 0}$ 标准布朗运动 $(B_t)_{t\geq 0}$,

1. $(B_t)_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅;
2. $(B_t^2 - t)_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅;
3. $(z_t = \mathrm{e}^{cB_t - \frac{c^2 t}{2}})_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅.

证明:

(1)

$$\begin{aligned} \mathrm{E}[B_t|\mathcal{F}_s] &= \mathrm{E}[B_t - B_s + B_s|\mathcal{F}_s] \\ &= \mathrm{E}(B_t - B_s) + B_s = B_s \end{aligned}$$

(2)

$$\begin{aligned} \mathrm{E}[B_t^2 - t|\mathcal{F}_s] &= \mathrm{E}\big[(B_t - B_s)^2 + B_s^2 + 2(B_t - B_s)B_s\big|\mathcal{F}_s\big] - t \\ &= \mathrm{E}\big[(B_t - B_s)^2\big|\mathcal{F}_s\big] + \mathrm{E}\big[B_s^2\big|\mathcal{F}_s\big] + 2\mathrm{E}\big[(B_t - B_s)B_s\big|\mathcal{F}_s\big] - t \\ &= \mathrm{E}(B_t - B_s)^2 + B_s^2 + 2B_s\mathrm{E}(B_t - B_s) - t = B_s^2 - s \end{aligned}$$

(3)

$$\begin{aligned} \mathrm{E}[z_t|\mathcal{F}_s] &= \mathrm{E}\big[z_s \frac{z_t}{z_s}\big|\mathcal{F}_s\big] \\ &= z_s\mathrm{E}\big[\frac{z_t}{z_s}\big|\mathcal{F}_s\big] \\ &= z_s\mathrm{E}\Big[\mathrm{e}^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}\Big|\mathcal{F}_s\Big] \\ &= z_s\mathrm{E}\Big(\mathrm{e}^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}\Big) = z_s \end{aligned}$$

[①] 这里需要强调一下, 由任意的 $0 \leq s < t$, $B_t - B_s$ 与 $\mathcal{F}_s$ 独立可以推出 $\forall 0 = t_0 < t_1 < t_2 < \cdots < t_n$, $\{B_{t_k} - B_{t_{k-1}}, k = 1, 2, \cdots, n\}$ 相互独立.

□

性质 4.2.2. 设 $\left(B_t^{(1)},\cdots,B_t^{(n)}\right)_{t\geq 0}$ 为 n 维 $(\mathcal{F}_t)_{t\geq 0}$ 标准布朗运动, 则它为 $(\mathcal{F}_t)_{t\geq 0}$ 时间齐次的马氏过程, 其转移密度函数为

$$p(t,x,y)=\frac{1}{(2\pi t)^{\frac{n}{2}}}\mathrm{e}^{-\frac{\|x-y\|^2}{2t}},\ x\in\mathbb{R}^n,\ y\in\mathbb{R}^n.$$

性质 4.2.3. 设 $\{p(t,x,y),\ x\in\mathbb{R}^n,\ y\in\mathbb{R}^n\}$ 为 n 维标准布朗运动的转移密度, 则

$$\frac{\partial p(t,x,y)}{\partial t}=\frac{1}{2}\Delta_x p(t,x,y)=\frac{1}{2}\sum_{i=1}^n\frac{\partial^2 p(t,x,y)}{\partial x_i^2}, \tag{4.2.1}$$

$$\frac{\partial p(t,x,y)}{\partial t}=\frac{1}{2}\Delta_y p(t,x,y)=\frac{1}{2}\sum_{i=1}^n\frac{\partial^2 p(t,x,y)}{\partial y_i^2}. \tag{4.2.2}$$

(4.2.1) 被称为 Kolmogorov 向后方程 (第一方程), (4.2.2) 被称为 Kolmogorov 向前方程 (第二方程) 或者 Fokker-Planck 方程.

性质 4.2.4. (Donsker 不变原理) 设 $(Y_k)_{n\geq 1}$ 独立同分布随机变量序列, $\mathrm{E}Y_k=0$, $\mathrm{E}Y_k^2=1$. 对于固定时刻 t, Δt 表示时间间隔, Δx 表示步长,

$$\eta_t^{\Delta}=\Delta x\big(Y_1+Y_2+\cdots+Y_{[\frac{t}{\Delta t}]}\big).$$

固定 Δx, Δt 和 ω, 让 t 变化, 则 η_t^{Δ} 为一个右连续左极限存在的函数.

$$\mathrm{E}\eta_t^{\Delta}=0,\mathrm{var}(\eta_t^{\Delta})=(\Delta x)^2\big[\frac{t}{\Delta t}\big].$$

若取 $h>0$ 充分小, $\Delta t=h$, $\Delta x=\sqrt{h}$, $h\to 0$, 则

$$(\eta_t^{\Delta})_{0\leq t\leq 1}\to(B_t)_{0\leq t\leq 1}.$$

♣♣♣20230413

4.3 扩散过程的分析方法简介

扩散 (diffusion) 是一种物理现象, 描述混合在一起的几种物质运动趋向于平衡的过程. 例如, 布朗运动描述悬浮在液体中的花粉粒子在液体中逐渐 “扩散” 至 “均匀” 分布的运动过程. 一个自然的问题是, 如果液体在不同时刻、不同位置的物理性态对花粉粒子的影响不同时, 例如, 在流动的液体中, 花粉粒子将遵从怎样的运动规律?

扩散过程并没有统一的数学定义, 但其核心是轨道连续的马氏过程. 与布朗运动相同, 可以通过建立转移概率满足的方程来刻画其宏观性质的演化, 例如, 将墨水滴入水中, 墨水浓度在不同时刻和位置遵从的规律; 也可以利用追踪每个花粉粒子或墨水粒子的轨迹, 构造概率空间, 建立随机微分方程从微观角度描述其服从的运动规律, 这是第五章和第六章的内容.

对比布朗运动, 给出关于扩散过程的三个条件: 对于任意的 $\varepsilon>0$

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\sup_{x\in\mathbb{R}^d}\mathrm{P}(|X_t-x|>\varepsilon|X_s=x)=0; \tag{4.3.1}$$

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\int_{|y-x|\leq\varepsilon}(y-x)P(s,x;t,\mathrm{d}y)=b(s,x); \tag{4.3.2}$$

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\int_{|y-x|\leq\varepsilon}(y-x)^2P(s,x;t,\mathrm{d}y)=\sigma^2(s,x). \tag{4.3.3}$$

当 $(X_t)_{t\geq 0}$ 是时间齐次的马氏过程时, $b(s,x)$ 和 $\sigma(s,x)$ 不依赖于 s, 分别记为 $b(x)$ 和 $\sigma(x)$. 称 $b(x,s)$ 和 $\sigma^2(s,x)$ 分别称为扩散过程 $(X_t)_{t\geq 0}$ 的漂移系数 (drift coefficient) 和扩散系数 (diffusion coefficient). 其中 (4.3.1) 是为了保证 $(X_t)_{t\geq 0}$ 是一个轨道连续的过程.

我们利用 C-K 方程来推导转移概率应该满足的演化规律.

4.3.1 Kolmogorov 向后方程

这里我们以状态空间为 $\mathbb{R}$ 为例. 对于固定的时刻 t, 对于任意的 $s<t$, $f\in C_b(\mathbb{R})$, f 看作对 $(X_t)_{t\geq 0}$ 的一个观测函数. 设 $u(s,x)=\mathrm{E}(f(X_t)|X_s=x)$, $u(s,X_s)=\mathrm{E}[f(X_t)|X_s]$. 我们将 $u(s,X_s)$ 考虑成 s 的函数, 形式上它求关于 s 的微分:

$$\mathrm{d}u(s,X_s)=\frac{\partial u(s,X_s)}{\partial s}\mathrm{d}s+\frac{\partial u(s,X_s)}{\partial x}\text{“}\mathrm{d}X_s\text{”}.$$

此时, 上式对于每一条样本轨道没有明确的数学含义, 但可以研究其在 “宏观平均” $\mathrm{E}(\cdot|X_s=x)$ 下的数学意义.

一方面, $\mathrm{E}(f(X_t)|X_s=x)$ 表示从 s 时刻, x 位置出发, 观测量 $f(X_t)$ 的均值, 还可以表示成从 s 时刻, x 位置出发, 先经过时间 $\Delta(>0)$, 到达 $X_{s+\Delta}$, 再从时刻 $s+\Delta$ 出发, 到达时刻 t, 观测量 $f(X_t)$ 的均值. 因此由 Chapman-Kolmogorov 方程,

$$\mathrm{E}(f(X_t)|X_s=x)=\mathrm{E}\big(\mathrm{E}[f(X_t)|X_{s+\Delta}]\big|X_s=x\big)=\mathrm{E}\big(u(s+\Delta,X_{s+\Delta})\big|X_s=x\big).$$

另一方面, 从 s 时刻, x 出发, 量 $u(s,X_s)$ 的均值就是 $u(s,x)$, 即

$$\mathrm{E}(u(s,X_s)|\xi_s=x)=u(s,x). \tag{4.3.4}$$

因此, $\mathrm{E}\big(u(s+\Delta,X_{s+\Delta})-u(s,X_s)|X_s=x\big)=0$. 而形式上,

$$\mathrm{d}u(s,X_s)\approx u(s+\Delta,X_{s+\Delta})-u(s,X_s),$$

所以, 可以认为 $\mathrm{E}(\mathrm{d}u(s,X_s)|X_s=x)=0$, 并得到

$$-\frac{\partial u(s,x)}{\partial s}=-\mathrm{E}\Big(\frac{\partial u(s,X_s)}{\partial s}\Big|X_s=x\Big)=\mathrm{E}\Big(\frac{\partial u(s,X_s)}{\partial x}\text{“}\frac{\mathrm{d}X_s}{\mathrm{d}s}\text{”}\Big|X_s=x\Big).$$

由此, 我们可以通过计算 $-\frac{\partial u(s,X_s)}{\partial s}$ 得到上式右端的明确的数学意义.

定理 4.3.1. Kolmogorov 向后方程

设 $b(s,x)$, $\sigma(s,x)$ 和 $u(s,x)$ 是关于 s 和 x 的光滑函数, 则 $u(s,x)$ 是如下方程终值问题的解,

$$\begin{cases}-\dfrac{\partial u(s,x)}{\partial s}=b(s,x)\dfrac{\partial u(s,x)}{\partial x}+\dfrac{1}{2}\sigma^2(s,x)\dfrac{\partial^2u(s,x)}{\partial x^2}, & s<t\\ u(t,x)=f(x)\end{cases}. \tag{4.3.5}$$

若 $P(s,x;t,\mathrm{d}y)$ 还有密度函数 $p(s,x;t,y)$, 则

$$\begin{cases}-\dfrac{\partial p(s,x;t,y)}{\partial s}=b(s,x)\dfrac{\partial p(s,x;t,y)}{\partial x}+\dfrac{1}{2}\sigma^2(s,x)\dfrac{\partial^2p(s,x;t,x)}{\partial x^2}, & s<t\\ p(t,x;t,y)=\delta_x(y),\ \text{即}\displaystyle\int f(y)p(t,x;t,y)\mathrm{d}y=f(x),\ f\in C_b(\mathbb{R})\end{cases}. \tag{4.3.6}$$

证明: 对于 $s<s+\Delta<t$, 利用泰勒展开得到

$$\begin{aligned}&\frac{u(s,x)-u(s+\Delta,x)}{\Delta}\\
=&\frac{1}{\Delta}\Big\{\int u(s+\Delta,y)P(s,x;s+\Delta,\mathrm{d}y)-u(s+\Delta,x)\Big\}\\
=&\frac{1}{\Delta}\Big\{\int\big[u(s+\Delta,y)-u(s+\Delta,x)\big]P(s,x;s+\Delta,\mathrm{d}y)\Big\}\\
=&\frac{1}{\Delta}\Big\{\int_{|x-y|<\varepsilon}\big[u(s+\Delta,y)-u(s+\Delta,x)\big]P(s,x;s+\Delta,\mathrm{d}y)\\
&\quad+\int_{|x-y|\geq\varepsilon}\big[u(s+\Delta,y)-u(s+\Delta,x)\big]P(s,x;s+\Delta,\mathrm{d}y)\Big\}\\
=&\frac{1}{\Delta}\Big\{\frac{\partial u(s+\Delta,x)}{\partial x}\int_{|x-y|<\varepsilon}(y-x)P(s,x;s+\Delta,\mathrm{d}y)\\
&\quad+\frac{1}{2}\frac{\partial^2u(s+\Delta,x)}{\partial x^2}\int_{|x-y|<\varepsilon}(y-x)^2P(s,x;s+\Delta,\mathrm{d}y)\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\int_{|x-y|<\varepsilon}\left[\frac{\partial^2 u(s+\Delta,z)}{\partial x^2}-\frac{\partial^2 u(s+\Delta,x)}{\partial x^2}\right](y-x)^2P(s,x;s+\Delta,\mathrm{d}y)\\
&+\int_{|x-y|\ge\varepsilon}\big[u(s+\Delta,y)-u(s+\Delta,x)\big]P(s,x;s+\Delta,\mathrm{d}y)\bigg\}\\
=&\frac{1}{\Delta}\{I+II+III+IV\}.
\end{aligned}
$$

这里 $z\in\{y:|x-y|\le\varepsilon\}$.

首先, $u(s,x)$ 是光滑函数, 因此, 当 $\Delta\to 0,\varepsilon\to 0$ 时

$$
\sup_{|x-y|<\varepsilon}\left|\frac{\partial^2 u(s+\Delta,y)}{\partial x^2}-\frac{\partial^2 u(s+\Delta,x)}{\partial x^2}\right|\le\sup_{s\in[0,t],|x-y|<\varepsilon}\left|\frac{\partial^2 u(s,z)}{\partial x^2}-\frac{\partial^2 u(s,x)}{\partial x^2}\right|\to 0.
$$

其次, $\sup_{s\in[0,t];x\in\mathbb{R}}|u(s,x)|\le\sup_{s\in[0,t];x\in\mathbb{R}}\int_{\mathbb{R}}|f(y)|P(s,x;t,\mathrm{d}y)\le\sup_{y\in\mathbb{R}}|f(y)|$. 因此,

$$
|IV|\le 2\sup_{y\in\mathbb{R}}|f(y)|\int_{|x-y|\ge\varepsilon}P(s,x;s+\Delta,\mathrm{d}y).
$$

利用 (4.3.1), (4.3.2) 和 (4.3.3), 当 $\Delta\to 0,\varepsilon\to 0$ 时

$$
\frac{1}{\Delta}I\to b(x,s)\frac{\partial u(s,x)}{\partial x},\quad \frac{1}{\Delta}II\to\frac{1}{2}\sigma^2(s,x)\frac{\partial^2 u(s,x)}{\partial x^2},\quad \frac{1}{\Delta}III\to 0,\quad \frac{1}{\Delta}IV\to 0.
$$

由此就得到 (4.3.20). 进一步, 当 $P(s,x;t,\mathrm{d}y)$ 有密度函数 $p(s,x;t,y)$ 时, (4.3.20) 成为

$$
-\int f(y)\frac{\partial p(s,x;t,y)}{\partial s}\mathrm{d}y=\int f(y)\Big[b(s,x)\frac{\partial p(s,x;t,y)}{\partial x}+\frac{1}{2}\sigma^2(s,x)\frac{\partial^2 p(s,x;t,y)}{\partial x^2}\Big]\mathrm{d}y.
$$

由 $f(y)$ 的任意性得到 (4.3.22)

□

如果 $(X_t)_{t\ge t}$ 是时间齐次扩散过程, 设 $t'=t-s$, $\widetilde{u}(t',x)=u(s,x)=\mathrm{E}(f(X_t)|X_s=x)$. 因为, $\frac{\partial\widetilde{u}(t',x)}{\partial t'}=-\frac{\partial u(s,x)}{\partial s}$, 所以 (4.3.20) 成为,

$$
\begin{cases}
\dfrac{\partial\widetilde{u}(t',x)}{\partial t'}=b(x)\dfrac{\partial\widetilde{u}(t',x)}{\partial x}+\dfrac{1}{2}\sigma^2(x)\dfrac{\partial^2\widetilde{u}(t',x)}{\partial x^2},\ t'>0\\
\widetilde{u}(0,x)=f(x)
\end{cases}.
$$

进一步得到时间齐次扩散过程的 Kolmogorov 向后方程.

定理 4.3.2. (时间齐次)

如果定理4.3.1中的 $(X_t)_{t\ge 0}$ 是时间齐次扩散过程, 设 $u(t,x)=\mathrm{E}(f(X_t)|X_0=x)$, 则 $u(t,x)$ 是如下方程初值问题的解,

$$
\begin{cases}
\dfrac{\partial u(t,x)}{\partial t}=b(x)\dfrac{\partial u(t,x)}{\partial x}+\dfrac{1}{2}\sigma^2(x)\dfrac{\partial^2 u(t,x)}{\partial x^2},\ t>0\\
u(0,x)=f(x)
\end{cases}.
$$

若 $P(t,x,\mathrm{d}y)$ 还有密度函数 $p(t,x,y)$, 则

$$
\begin{cases}
\dfrac{\partial p(t,x,y)}{\partial t}=b(x)\dfrac{\partial p(t,x,y)}{\partial x}+\dfrac{1}{2}\sigma^2(x)\dfrac{\partial^2 p(t,x,y)}{\partial x^2},\ t>0\\
p(0,x,y)=\delta_x(y),\ \text{即}\displaystyle\int f(y)p(t,x,y)\mathrm{d}y=f(x),\ f\in C_b(\mathbb{R})
\end{cases}. \tag{4.3.7}
$$

对于特殊形式的向后方程, 可以直接求解得到转移密度函数或分布函数的显式表达.

例 4.3.1. 设扩散过程 $(X_t)_{t\ge 0}$ 的漂移系数为 0, 扩散系数与空间变量 x 无关, 记为 $\sigma^2(t)$, 并且假设对于任意的 $t\ge 0$, $\sigma^2(t)>0$, $\int_0^\infty\sigma^2(t)\mathrm{d}t=\infty$. 求 $(X_t)_{t\ge 0}$ 的转移密度函数.

解: 我们需要求解如下的偏微分方程:

$$
\begin{cases}
-\dfrac{\partial u(s,x)}{\partial s}=\dfrac{1}{2}\sigma^2(s)\dfrac{\partial^2 u(s,x)}{\partial x^2},\ s<t\\
u(t,x)=f(x),\quad f\in C_b(\mathbb{R})
\end{cases}. \tag{4.3.8}
$$

设 $\alpha(s)=\tilde{s}$, $\tilde{u}(\tilde{s},x)=u(s,x)$, 其中 $\alpha(s)$ 的具体形式待定. 因为

$$\frac{\partial u(s,x)}{\partial s}=\frac{\partial \tilde{u}(\tilde{s},x)}{\partial \tilde{s}}\frac{\mathrm{d}\alpha(s)}{\mathrm{d}s},\quad \frac{\partial^2 u(s,x)}{\partial x^2}=\frac{\partial^2 \tilde{u}(\tilde{s},x)}{\partial x^2},$$

所以

$$-\frac{\mathrm{d}\alpha(s)}{\mathrm{d}s}\frac{\partial \tilde{u}(\tilde{s},x)}{\partial \tilde{s}}=\frac{1}{2}\sigma^2(s)\frac{\partial^2 \tilde{u}(\tilde{s},x)}{\partial x^2}.$$

取 $\alpha(s)=\int_s^t \sigma^2(l)\mathrm{d}l$, 也就是 $-\frac{\mathrm{d}\alpha(s)}{\mathrm{d}s}=\sigma^2(s)$, 则 (4.3.8) 转化成热方程

$$\begin{cases}\dfrac{\partial \tilde{u}(\tilde{s},x)}{\partial \tilde{s}}=\dfrac{1}{2}\dfrac{\partial^2 \tilde{u}(\tilde{s},x)}{\partial x^2}, & \tilde{s}>0\\ \tilde{u}(0,x)=f(x), & f\in C_b(\mathbb{R})\end{cases}.$$

因此,

$$\tilde{u}(\tilde{s},x)=\int_{\mathbb{R}} f(y)\frac{1}{\sqrt{2\pi\tilde{s}}}\exp\Big(-\frac{(x-y)^2}{2\tilde{s}}\Big)\mathrm{d}y.$$

由此解得,

$$u(s,x)=\int_{\mathbb{R}} f(y)\frac{1}{\sqrt{2\pi\int_s^t \sigma^2(l)\mathrm{d}l}}\exp\Big(-\frac{(x-y)^2}{2\int_s^t \sigma^2(l)\mathrm{d}l}\Big)\mathrm{d}y.$$

这说明 $(X_t)_{t\geq 0}$ 的转移密度函数为

$$p(s,x;t,y)=\frac{1}{\sqrt{2\pi\int_s^t \sigma^2(l)\mathrm{d}l}}\exp\Big(-\frac{(x-y)^2}{2\int_s^t \sigma^2(l)\mathrm{d}l}\Big).$$

□

设 $(B_t)_{t\geq 0}$ 是标准布朗运动. 事实上, 我们还可以看出 X_t-X_0 与 $B_{\int_0^t \sigma^2(l)\mathrm{d}l}$ 和 $(\int_0^t \sigma^2(l)\mathrm{d}l)^{\frac{1}{2}}B_t$ 同分布.

这里有一个有趣的问题, 如果在某些时刻 $\sigma^2(s)=0$ 或者 $\int_0^\infty \sigma^2(t)\mathrm{d}t<\infty$, $(X_t)_{t\geq 0}$ 会有什么样的性态? 这个问题留给读者思考.

例 4.3.2. 设时间齐次扩散过程 $(X_t)_{t\geq 0}$ 的漂移系数 $b(x)=\begin{cases}\mu x & x>0\\ 0 & x\leq 0\end{cases}$, 扩散系数 $\sigma^2(x)=\begin{cases}\sigma^2x^2 & x>0\\ 0 & x\leq 0\end{cases}$. 其中 μ 和 σ 为常数, $\sigma>0$. 求 $(X_t)_{t\geq 0}$ 的转移密度函数和转移分布函数.

4.3.2 Kolmogorov 向前方程, Fokker-Planck 方程

Kolmogorov 向后方程是刻画 $P(s,x;t,\mathrm{d}y)$ 对于 “过去的” 时空变量 (s,x) 的变化规律. 那么, 相对于 “未来的” 时空变量 (t,y) 的变化规律是什么呢?

设 $f\in C_b^2(\mathbb{R})$, $t>s$, $\Delta>0$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[f(X_t)|\mathcal{F}_s]&=\lim_{\Delta\to 0}\frac{1}{\Delta}\mathrm{E}[f(X_{t+\Delta})-f(X_t)|\mathcal{F}_s]\\&=\lim_{\Delta\to 0}\frac{1}{\Delta}\mathrm{E}\big[\mathrm{E}[f(X_{t+\Delta})-f(X_t)|\mathcal{F}_t]\big|\mathcal{F}_s\big].\end{aligned}\tag{4.3.9}$$

类似于向后方程的方法, 我们计算

$$\begin{aligned}&\lim_{\Delta\to 0}\frac{1}{\Delta}\mathrm{E}[f(X_{t+\Delta})-f(X_t)|\mathcal{F}_t]\\=&\lim_{\Delta\to 0}\frac{1}{\Delta}\mathrm{E}[f(X_{t+\Delta})-f(X_t)|X_t]\\=&\lim_{\Delta\to 0}\frac{1}{\Delta}\int\big(f(y)-f(x)\big)p(t,x;t+\Delta,y)\mathrm{d}y\Big|_{x=X_t}\\=&\frac{1}{2}\sigma^2(t,x)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x)+b(t,x)\frac{\mathrm{d}f}{\mathrm{d}x}(x)\Big|_{x=X_t}.\end{aligned}\tag{4.3.10}$$

将 (4.3.10) 代入 (4.3.9), 得到

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[f(X_t)|\mathcal{F}_s] = \mathrm{E}\Big[\frac{1}{2}\sigma^2(t,X_t)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_t) + b(t,X_t)\frac{\mathrm{d}f}{\mathrm{d}x}(X_t)\Big|\mathcal{F}_s\Big]. \tag{4.3.11}$$

也就是说, 对于任意的 $t_2 > t_1 > s$,

$$\mathrm{E}\Big[f(X_{t_2}) - f(X_{t_1}) - \int_{t_1}^{t_2} \frac{1}{2}\sigma^2(t,X_t)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_t) + b(t,X_t)\frac{\mathrm{d}f}{\mathrm{d}x}(X_t)\,\mathrm{d}t\Big|\mathcal{F}_s\Big] = 0. \tag{4.3.12}$$

由此得到下面非常重要的一个结论:

定理 4.3.3. 对于任意的 $f \in C_b^2(\mathbb{R})$,

$$\big(f(X_t) - f(X_0) - \int_0^t \frac{1}{2}\sigma^2(s,X_s)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_s) + b(s,X_s)\frac{\mathrm{d}f}{\mathrm{d}x}(X_s)\,\mathrm{d}s\big)_{t\geq 0}$$

是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅.

定理 4.3.4. Kolmogorov 向前方程, Fokker-Planck 方程

设 $(X_t)_{t\geq 0}$ 的转移概率 $P(s,x;t,\mathrm{d}y)$ 还有密度函数 $p(s,x;t,y)$, $b(t,y)$, $\sigma(t,y)$ 和 $p(s,x;t,y)$ 是关于 t 和 y 的光滑函数, 则 $p(s,x;t,y)$ 是如下初值问题的解,

$$\begin{cases} \dfrac{\partial p(s,x;t,y)}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}\big(\sigma^2(t,y)p(s,x;t,y)\big) - \dfrac{\partial}{\partial y}\big(b(t,y)p(s,x;t,y)\big) \\ p(s,x;s,y) = \delta_x(y), \text{ 即} \displaystyle\int p(s,x;s,y)f(y)\mathrm{d}y = f(x), \quad f(x) \in C_0^\infty(\mathbb{R}) \end{cases}. \tag{4.3.13}$$

当 $(X_t)_{t\geq 0}$ 还是时间齐次扩散过程, 则

$$\begin{cases} \dfrac{\partial p(t,x,y)}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}\big(\sigma(y)^2 p(t,x,y)\big) - \dfrac{\partial}{\partial y}\big(b(y)p(t,x,y)\big) \\ p(0,x,y) = \delta_x(y) \end{cases}. \tag{4.3.14}$$

证明: 对于任意的 $f \in C_0^2(\mathbb{R})$, 利用马氏性, (4.3.11) 成为

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}(f(X_t)|X_s = x) = \mathrm{E}\Big(\frac{1}{2}\sigma^2(t,X_t)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_t) + b(t,X_t)\frac{\mathrm{d}f}{\mathrm{d}x}(X_t)\Big|X_s = x\Big).$$

也就是,

$$\begin{aligned} &\int f(y)\frac{\partial p(s,x;t,y)}{\partial t}\mathrm{d}y \\ =& \int \Big[\frac{1}{2}\sigma^2(t,y)\frac{\mathrm{d}^2 f}{\mathrm{d}y^2}(y) + b(t,y)\frac{\mathrm{d}f}{\mathrm{d}y}(y)\Big] p(s,x;t,y)\mathrm{d}y \\ =& \int f(y)\Big[\frac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(t,y)p(s,x;t,y)\big) + -\frac{\partial}{\partial y}\big(b(t,y)p(s,x;t,y)\big)\Big]\mathrm{d}y. \end{aligned} \tag{4.3.15}$$

其中第二个等号中第一项使用两次分部积分公式, 第二项使用了一次分部积分公式. 由 $f(y)$ 的任意性, 就得到了 (4.3.13).

当 $(X_t)_{t\geq 0}$ 是时间齐次时, 对于任意的 $t > 0$, $p(t,x,y) = p(s,x;s+t,y)$,

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{\partial p(s,x;s+t,y)}{\partial(s+t)}\frac{\mathrm{d}(s+t)}{\mathrm{d}t},$$

就得到了 (4.3.24).

□

♣♣♣20230418

时间齐次扩散过程的 Kolmogorov 向后方程和向前方程右端项关于 t 的导数可以理解为是对出发时刻与到达时刻的时间间隔的求导.

同样, 对于某些特殊形式的向前方程, 也可以直接求解得到转移密度函数或分布函数的显式表达.

例 4.3.3. 设时间齐次扩散过程 $(X_t)_{t\geq 0}$ 的漂移系数是 $-bx$, 扩散系数是常数 σ^2. 求 $(X_t)_{t\geq 0}$ 的转移密度函数.

推论 4.3.1. 主方程, Master equation

$(X_t)_{t\geq 0}$ 如定理4.3.4. X_0 有密度函数 $p_0(x)$, 则 X_t 的密度函数 $p(t,y)$ 是 $\int p_0(x)p(0,x;t,y)\mathrm{d}x$, 而且 $p(t,y)$ 满足:

$$\begin{cases} \dfrac{\partial p(t,y)}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}\big(\sigma(t,y)^2 p(t,y)\big) - \dfrac{\partial}{\partial y}\big(b(t,y)p(t,y)\big) \\ p(0,y) = p_0(y) \end{cases}. \tag{4.3.16}$$

证明: 对于任意的 $f(x)\in C_b(\mathbb{R})$,

$$\begin{aligned} \mathrm{E}(f(X_t)) &= \mathrm{E}\big(\mathrm{E}[f(X_t)|X_0]\big) \\ &= \int\Big(\int f(y)p(0,x;t,y)\mathrm{d}y\Big)p_0(x)\mathrm{d}x \\ &= \int f(y)\Big(\int p_0(x)p(0,x;t,y)\mathrm{d}x\Big)\mathrm{d}y. \end{aligned}$$

因此, $p(t,y) = \int p_0(x)p(0,x;t,y)\mathrm{d}x$. 进一步, 向前方程 (4.3.13) 两端乘以 $p_0(x)$, 再关于 x 积分就得到了 (4.3.25).

□

方程 (4.3.25) 表达了 X_t 的密度 $p(t,y)$ 应该满足的演化规律. 它在化学等学科中也被称为主方程 (Master equation).

例 4.3.4. 设 $(X_t)_{t\geq 0}$ 是时间齐次一维扩散过程, $p(t,x)$ 为 X_t 的密度函数. 由推论4.3.1, $p(t,x)$ 满足:

$$\begin{cases} \dfrac{\partial p(t,y)}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}\big(\sigma(y)^2 p(t,y)\big) - \dfrac{\partial}{\partial y}\big(b(y)p(t,y)\big) \\ p(0,y) = p_0(y) \end{cases}. \tag{4.3.17}$$

如果对任意的 $t\geq 0$, $\frac{\partial p(t,y)}{\partial t} = 0$, 这意味着 $p(t,y) = p_0(y)$, 也就是说 $(X_t)_{t\geq 0}$ 不但是平稳过程, 而且 $p_0(y)$ 是其不变概率密度函数, 即 $\int p_0(x)p(t,x,y)\mathrm{d}x = p_0(y)$. 此时记 $p_0(y) = \pi(y)$.

设 $\inf_{\mathbb{R}}\sigma(y)^2 > 0$, $\int_{\mathbb{R}} \frac{1}{\sigma(y)^2}\mathrm{e}^{\int_c^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}\mathrm{d}y < \infty$. 我们利用 (4.3.17) 可以求出 $\pi(y)$ 的显式表达. 注意到 $\pi(y)$ 满足如下方程:

$$\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2}\big(\sigma(y)^2\pi(y)\big) - \frac{\mathrm{d}}{\mathrm{d}y}\big(b(y)\pi(y)\big) = 0.$$

进一步,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}y}\big(\sigma(y)^2\pi(y)\big) - b(y)\pi(y) = \frac{\alpha}{2},$$

α 是待定常数. 设 $\gamma(y) = \sigma(y)^2\pi(y)$, 上式化为

$$\frac{\mathrm{d}}{\mathrm{d}y}\gamma(y) = \frac{2b(y)}{\sigma(y)^2}\gamma(y) + \alpha.$$

这个一阶非齐次常微分方程的解为

$$\gamma(y) = \alpha\int_c^y \mathrm{e}^{\int_s^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}\mathrm{d}s + \beta\mathrm{e}^{\int_c^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}.$$

其中 $c\in\mathbb{R}$ 是一个取定的常数, β 为依赖于 c 的待定常数.

$$\pi(y) = \frac{\alpha}{\sigma(y)^2}\int_c^y \mathrm{e}^{\int_s^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}\mathrm{d}s + \frac{\beta}{\sigma(y)^2}\mathrm{e}^{\int_c^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}.$$

因为 $\int_{\mathbb{R}}\pi(y)\mathrm{d}y = 1$, $\lim_{y\to\pm\infty}\pi(y) = 0$, 由此得到 $\alpha = 0$, 即

$$\pi(y) = \frac{\beta}{\sigma(y)^2}\mathrm{e}^{\int_c^y \frac{2b(u)}{\sigma(u)^2}\mathrm{d}u}.$$

β 使得 $\int_{\mathbb{R}}\pi(y)\mathrm{d}y = 1$.

特别, 如果 $\sigma(y)^2$ 是常数, 记为 $\sigma^2>0$, $b(y)$ 是线性函数, 记为 $-by$, 其中 $b>0$ 是常数, 则

$$\pi_(y)=\frac{\sqrt{b}}{\sigma\sqrt{\pi}}\mathrm{e}^{-\frac{by^2}{\sigma^2}}.$$

此时, $(X_t)_{t\geq 0}$ 的不变概率分布服从 $N(0,\frac{\sigma^2}{2b})$.

4.3.3 多维扩散

多维形式, 扩散算子??

如果 $X_t=(X_t^{(1)},\cdots,X_t^{(n)})$ 是取值 $\mathbb{R}^n$ 的马氏过程, 其转移概率记为 $P(s,x;t,\mathrm{d}y)$, $x\in\mathbb{R}^n$, $y\in\mathbb{R}^n$. 设 $P(s,x;t,\mathrm{d}y)$ 满足:

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\sup_{x\in\mathbb{R}^n}\mathrm{P}(\|X_t-x\|>\varepsilon|X_s=x)=0;$$

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\int_{\|y-x\|<\varepsilon}(y-x)P(s,x;t,\mathrm{d}y)=\mathbf{b}(t,x)=\begin{pmatrix}b_1(t,x)\\ \vdots\\ b_n(t,x)\end{pmatrix}.$$

$$\lim_{t-s\downarrow 0}\frac{1}{t-s}\int_{\|y-x\|<\varepsilon}(y-x)(y-x)^{\mathrm{t}}P(s,x;t,\mathrm{d}y)=\mathbf{A}(t,x)=(a_{ij}(t,x))_{1\leq i,j\leq n}.$$

定义扩散算子:

$$\mathcal{L}_{t,x}=\frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,x)\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^n b_i(t,x)\frac{\partial}{\partial x_i}.$$

即 $f\in C^2(\mathbb{R}^n)$

$$\mathcal{L}_{t,x}f(x)=\frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,x)\frac{\partial^2 f}{\partial x_i\partial x_j}(x)+\sum_{i=1}^n b_i(t,x)\frac{\partial f}{\partial x_i}(x).$$[①]

定义 $\mathcal{L}_t$ 的形式共轭算子

$$\mathcal{L}^*_{t,y}f(y)=\frac{1}{2}\sum_{i,j=1}^n\frac{\partial^2}{\partial y_i\partial y_j}\Big(a_{ij}(t,y)f(y)\Big)-\sum_{i=1}^d\frac{\partial}{\partial y_i}\Big(b_i(t,y)f(y)\Big) \tag{4.3.18}$$

$$=\frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,y)\frac{\partial^2 f}{\partial y_i\partial y_j}(y)+\sum_{i=1}^d\tilde{b}_i(t,y)\frac{\partial f}{\partial y_i}(y)+\tilde{c}(t,y)f(y). \tag{4.3.19}$$

其中,

$$\tilde{b_i}(t,y)=-b_i(t,y)+\sum_{j=1}^n\frac{\partial a_{ij}(t,y)}{\partial y_j},$$

$$\tilde{c}(t,y)=\frac{1}{2}\sum_{i,j=1}^n\frac{\partial^2 a_{ij}(t,y)}{\partial y_i\partial y_j}-\sum_{i=1}^n\frac{\partial b_i(t,y)}{\partial y_i}.$$

(4.3.18) 是记成散度型算子[②], (4.3.19) 是非散度型算子.

[①]设 ∇ 为梯度算子, 即, 如果 $f\in C^2(\mathbb{R}^n)$, $\nabla f(x)=\big(\frac{\partial f}{\partial x_1}(x),\cdots,\frac{\partial f}{\partial x_n}(x)\big)^{\mathrm{t}}$; 如果 $v(x)=(v_1(x),\cdots,v_m(x)):\mathbb{R}^n\to\mathbb{R}^m$, $v_i(x)\in C^2(\mathbb{R}^n)$, $\nabla v(x)=\big(\frac{\partial v_i}{\partial x_j}\big)_{i=1,\cdots,m,\ j=1,\cdots,n}$. 对于 $f\in C^2(\mathbb{R}^n)$, 其 Hessian 矩阵为 $\nabla^2 f(x):=\nabla(\nabla f(x))$, 则 $\nabla^2 f(x)=\Big(\frac{\partial^2 f}{\partial x_i\partial x_j}\Big)_{i=1,\cdots,n,\ j=1,\cdots,n}$. 由此, 扩散算子 $\mathcal{L}_t$ 可以写成如下更为简洁的形式:

$$\mathcal{L}_{t,x}f(x)=\frac{1}{2}\mathrm{Tr}\big(\mathbf{A}(\nabla^2 f(x))^{\mathrm{t}}\big)+\big\langle\mathbf{b}(t,x)^{\mathrm{t}},\nabla f(x)\big\rangle.$$

其中 Tr 表示矩阵的迹.

[②]设 div 是散度算子, 即对于 $\mathbf{v}(y)=(v_1(y),\cdots,v_m(y))$: $\mathbb{R}^n\to\mathbb{R}^n$, $v_i(x)\in C^1(\mathbb{R}^n)$, $\operatorname{div}\mathbf{v}(y)=\sum_{i=1}^n\frac{\partial v_i}{\partial y_i}(y)$. $\mathbf{V}(y)=(\mathbf{v}_1(y),\cdots,\mathbf{v}_n(y))$, $\mathbf{v}_i:\mathbb{R}^n\to\mathbb{R}^n$, $\mathbf{v}_i\in C^1(\mathbb{R}^n)$, 则 $\operatorname{div}\mathbf{V}(y)=(\operatorname{div}\mathbf{v}_1,\cdots,\operatorname{div}\mathbf{v}_n)$. (4.3.18) 可以写成如下更为简洁的形式:

$$\mathcal{L}^*_{t,y}f(y)=\operatorname{div}_y\Big(\frac{1}{2}\operatorname{div}_y\big(\mathbf{A}(t,y)f(y)\big)-\mathbf{b}(t,y)^{\mathrm{t}}f(x)\Big)$$

在应用中, 当明确对具体的空间变量求导时, $\mathcal{L}_{t,x}$ 和 $\mathcal{L}^*_{t,y}$ 可以分别记为 $\mathcal{L}_t$ 和 $\mathcal{L}^*_t$; 当 $(X_t)_{t\geq 0}$ 是时间齐次扩散过程时, $\mathcal{L}_{t,x}$ 和 $\mathcal{L}^*_{t,y}$ 分别记为 $\mathcal{L}_x$ 和 $\mathcal{L}^*_y$, 甚至记为 $\mathcal{L}$ 和 $\mathcal{L}^*$.

设 $f(x_1,\cdots,x_n)\in C_b(\mathbb{R}^n)$, 记 $u(t,x)=\mathrm{E}(f(X_t)|X_s=x)$, $x=(x_1,\cdots,x_n)$. 类似于定理4.3.1, 我们得到多维扩散过程的 Kolmogorov 向后方程.

定理 4.3.5. Kolmogorov 向后方程

设 $\mathbf{b}(s,x)$, $\mathbf{A}(s,x)$ 和 $u(s,x)$ 是关于 s 和 x 的光滑函数, 则 $u(s,x)$ 是如下方程终值问题的解,

$$\begin{cases} -\dfrac{\partial u(s,x)}{\partial s}=\mathcal{L}_s u(s,x),\ \ s<t \\ u(t,x)=f(x) \end{cases}. \tag{4.3.20}$$

若 $P(s,x;t,\mathrm{d}y)$ 还有密度函数 $p(s,x;t,y)$, 则

$$\begin{cases} -\dfrac{\partial p(s,x;t,y)}{\partial s}=\mathcal{L}_{s,x}p(s,x;t,y),\ \ s<t \\ p(t,x;t,y)=\delta_x(y),\ \text{即} \displaystyle\int f(y)p(t,x;t,y)\mathrm{d}y=f(x),\ \ f\in C_b(\mathbb{R}^n) \end{cases}. \tag{4.3.21}$$

定理 4.3.6. (时间齐次)

如果定理4.3.5中的 $(X_t)_{t\geq 0}$ 是时间齐次扩散过程, 设 $u(t,x)=\mathrm{E}(f(X_t)|X_0=x)$, 则 $u(t,x)$ 是如下方程初值问题的解,

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t}=\mathcal{L}u(t,x),\ \ t>0 \\ u(0,x)=f(x) \end{cases}.$$

若 $P(t,x,\mathrm{d}y)$ 还有密度函数 $p(t,x,y)$, 则

$$\begin{cases} \dfrac{\partial p(t,x,y)}{\partial t}=\mathcal{L}_x p(t,x,y),\ \ t>0 \\ p(0,x,y)=\delta_x(y),\ \text{即} \displaystyle\int f(y)p(t,x,y)\mathrm{d}y=f(x),\ \ f\in C_b(\mathbb{R}) \end{cases}. \tag{4.3.22}$$

类似于定理4.3.3的推导, 我们得到多维扩散鞅的刻画:

定理 4.3.7. 对于任意的 $f\in C_b^2(\mathbb{R}^n)$,

$$\Big(f(X_t)-f(X_0)-\int_0^t \mathcal{L}_{s,x}f(X_s)\,\mathrm{d}s\Big)_{t\geq 0}$$

是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅.

定理 4.3.8. Kolmogorov 向前方程, Fokker-Planck 方程

设 $(X_t)_{t\geq 0}$ 是 $\mathbb{R}^n$ 的扩散过程, 其转移概率 $P(s,x;t,\mathrm{d}y)$ 还有密度函数 $p(s,x;t,y)$, $\mathbf{b}(t,y)$, $\mathbf{A}(t,y)$ 和 $p(s,x;t,y)$ 是关于 t 和 y 的光滑函数, 则 $p(s,x;t,y)$ 是如下初值问题的解,

$$\begin{cases} \dfrac{\partial p(s,x;t,y)}{\partial t}=\mathcal{L}^*_{t,y}p(s,x;t,y) \\ p(s,x;s,y)=\delta_x(y),\ \text{即} \displaystyle\int p(s,x;s,y)f(y)\mathrm{d}y=f(x),\ \ \ f(x)\in C_0^\infty(\mathbb{R}) \end{cases}. \tag{4.3.23}$$

当 $(X_t)_{t\geq 0}$ 还是时间齐次扩散过程, 则

$$\begin{cases} \dfrac{\partial p(t,x,y)}{\partial t}=\mathcal{L}^*_y p(t,x,y) \\ p(0,x,y)=\delta_x(y) \end{cases}. \tag{4.3.24}$$

推论 4.3.2. 主方程, Master equation

$(X_t)_{t\geq 0}$ 如定理4.3.8. X_0 有密度函数 $p_0(x)$, 则 X_t 的密度函数 $p(t,y)$ 是 $\int p_0(x)p(0,x;t,y)\mathrm{d}x$, 而且 $p(t,y)$ 满足:

$$\begin{cases} \dfrac{\partial p(t,y)}{\partial t} = \mathcal{L}^* p(t,y) \\ p(0,y) = p_0(y) \end{cases}. \tag{4.3.25}$$

我们这里给出多维扩散过程主方程一种物理想法的推导.

设

$$J_i(t,x) = b_i(t,x)p(t,x) - \frac{1}{2}\sum_{j=1}^n \frac{\partial}{\partial x_j}\big(a_{ij}(t,x)p(t,x)\big).$$

称 $\mathbf{J} = (J_1,\cdots,J_n)^{\mathrm{t}}$ 为概率流 (probability current, flux), J_i 称为第 i 个方向的概率流. 设 $V\subset\mathbb{R}^n$ 是具有光滑边界的有界单连通区域, $\mathbf{n}$ 为 V 边界 $S_V(x)$ 的单位内法向量. 对于 $t_2 > t_1$, 由概率守恒 (总概率为 1), 在时刻 t_1 与 t_2, 区域 V 上的概率的变化, 应该是由概率流从 $S_V(x)$ 流入和流出产生的, 由此得到

$$\begin{aligned} \int_V p(t_2,x)\mathrm{d}x - \int_V p(t_1,x)\mathrm{d}x &= \int_V \int_{t_1}^{t_2} \frac{\partial}{\partial t}p(t,x)\mathrm{d}t\,\mathrm{d}x \\ &= \int_{t_1}^{t_2} \mathrm{d}t \int_V \frac{\partial}{\partial t}p(t,x)\mathrm{d}x \\ &= \int_{t_1}^{t_2} \mathrm{d}t \int_{\partial V} \mathbf{J}\cdot\mathbf{n}\,\mathrm{d}S(x) \\ &= \int_{t_1}^{t_2} \int_V -\mathrm{div}(\mathbf{J})\,\mathrm{d}x. \end{aligned}$$

上式的最后一个等号利用了 Gauss-Green 公式. 由此即得

$$\frac{\partial p}{\partial t} + \mathrm{div}(\mathbf{J}) = 0,$$

或者

$$\frac{\partial p}{\partial t} = -\mathrm{div}(\mathbf{J}) = -\nabla\cdot\mathbf{J} \quad .$$

此即是 Fokker-Planck 方程.

事实上, Fokker-Planck 方程是一个守恒律方程 (equation of conservation law), 它保持概率守恒, 也被称为连续性方程 (continuity equation).

♣♣♣20230420、期中考试

✠✠✠✠✠✠

第五章　随机微积分和 Itô 公式

5.1　引言

首先我们来回顾一下微积分. 牛顿运动第一定律 (惯性定律) 告诉我们*每一个物体都会继续保持其静止或沿一直线作匀速运动状态, 除非有力加于其上, 迫使它改变这种状态*. 什么是匀速运动? 其实就是质点的运动速度任何时刻都一样; 接着问: 什么是速度? 就是单位时间内质点的位移, 即设 $x(t)$ 表示质点的位移, 则 $\frac{\Delta x}{\Delta t}$ 就是速度. 因此, 对于匀速运动我们有如下的运动方程:

$$\frac{\Delta x}{\Delta t} = v\,(\text{常数}) \quad \text{或} \quad x(t) - x(0) = vt \quad \text{或} \quad x(t) - x(s) = v(t-s). \tag{5.1.1}$$

当然, 这世界上质点的运动并非都是匀速直线运动. 而牛顿运动第二定律 $F = ma$ 给出了非惯性运动的物理规律. 对于一个非匀速运动的质点, 如何得到其运动方程? 这就是牛顿-莱布尼茨公式:

$$x(t) - x(s) = \int_s^t \dot{x}(u)\mathrm{d}u \quad \text{其中 } \dot{x}(u) \equiv \lim_{\Delta u \to 0} \frac{x(u+\Delta u) - x(u)}{\Delta u}.$$

如何理解牛顿-莱布尼茨公式? 其实就是用简单的运动来逼近复杂的运动, 用一些匀速运动的线性叠加来表示一个非匀速运动? 也就是说, 认为当 Δu 很小时, 在 $[u, u+\Delta u]$ 时间间隔内, 质点的运动 $x(t)$ 近似是一个匀速运动, 其速度用 $\dot{x}(u)$ 或 $\dot{x}(\xi)$, $\xi \in [u, u+\Delta u]$ 来表示, 此质点的运动方程就近似写成

$$x(t) - x(s) = \sum x(t_{i+1}) - x(t_i) \approx \sum \dot{x}(\xi_i)(t_{i+1} - t_i). \tag{5.1.2}$$

当 $\max_i(t_{i+1} - t_i)$ 越来越小趋近于 0 时, $\dot{x}(\xi)\Delta u$ 与 $x(u+\Delta u) - x(u)$ 的误差之和也越来越小收敛到 0. 因此将 (5.1.2) 式右端的极限定义为 $\int_0^t \dot{x}(u)\mathrm{d}u$.

在微积分中, 我们定义一个函数的微分、导数等概念, 定义曲线、曲面的切线、切平面等概念. 例如: $f(x)$ 在 x_0 处的切线方程是

$$z(x) = f(x_0) + f'(x_0)(x - x_0),$$

实际在说, 函数 $f(x)$ 在 x_0 附近的值与 $z(x)$ 在 x_0 附近的值很接近, 即 $\Delta f(x) = f'(x)\Delta x + o(\Delta x)$, 写成微分形式就是 $\mathrm{d}f(x) = f'(x)\mathrm{d}x$.

函数 $f(x,y)$ 在 (x_0, y_0) 点切平面方程为

$$z(x,y) = f(x_0,y_0) + \frac{\partial f(x,y)}{\partial x}\Big|_{(x_0,y_0)}(x - x_0) + \frac{\partial f(x,y)}{\partial y}\Big|_{(x_0,y_0)}(y - y_0),$$

也就是说, 函数在 $f(x,y)$ 在 (x_0,y_0) 附近的值与 $z(x,y)$ 在 (x_0,y_0) 附近的值很接近, 即

$$\Delta f(x,y) = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$$

写成微分形式

$$\mathrm{d}f(x,y) = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y = \Big(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\Big)\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \end{pmatrix}.$$

现在我们回到随机过程. 什么随机过程? 对比牛顿力学中质点运动的轨迹是 $\mathbb{R}^n$ 中的一条曲线, 随机过程 $\{X_t(\omega):\ t \in T\}$ 可以看成*随机变量空间*上的一条 "曲线", X_t 的分布 P_t 关于时间 t 可以看成是概率测度空间上的一条曲线. 自然我们又会问: (1) 什么是随机变量空间上曲线的切线? (2) 或者本质地问, 什么是随机变量空间或中的 "直线"? 如何理解随机变量空间中的 "直线"? 什么是随机过程中的 "匀速直线运动"?

匀速直线运动是“最简单”、“最理想化”的质点运动. (5.1.1) 的第三式说明在 $[s,t]$ 时间间隔内, 质点的位移只与 $t-s$ 有关, 与起点和终点的位置无关. 在随机过程中我们将此理解为 $X_t - X_s$ 的统计性质只与 $t-s$ 有关而与起点和终点的位置无关.

而对于“最简单”、“最理想化”这样的描述在随机过程中可以有多元化的理解, 在此我们可以看成“最随机”, 也就是说对于任意互不相交的区间 $(s_i, t_i]$, $i = 1, \cdots, n$,

$$X_{t_n} - X_{s_n}, X_{t_{n-1}} - X_{s_{n-1}}, \cdots, X_{t_1} - X_{s_1}, X_{s_1}$$

相互独立. 因此, 我们可以将时齐独立增量过程理解成随机过程中的“直线”. 进一步, 再对轨道做一些限制, 任意的 ω, $X_t(\omega)$ 是关于 t 的连续函数. 由上一章定理

$$X_t - X_0 = A \cdot B_t + b \cdot t,$$

其中 $B_t = \left(B_t^{(1)} \cdots B_t^{(r)}\right)^{\mathrm{t}}$ 是 r 维标准布朗运动, $A = (a_{ij})_{\mathrm{d}\times r}$ 常数矩阵, $b = (b_1 \cdots b_d)^{\mathrm{t}}$ 常数向量. $A = 0$ 时退化为 $\mathbb{R}^n$ 中的直线. 特别 $d = 1$ 时, $X_t - X_0 = \sigma B_t + bt$, (B_t) 为一维标准布朗运动, σ 为 0 时, 则退化为 $\mathbb{R}$ 中的直线. $d = 1$ 时, 设 $\{p(t,x,y), t > 0, x, y \in \mathbb{R}\}$ 为 $\sigma B_t + bt$ 的转移密度函数族, 则

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{\sigma^2}{2}\Delta_x p(t,x,y) + b\frac{\partial}{\partial x}p(t,x,y), \tag{5.1.3}$$

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{\sigma^2}{2}\Delta_y p(t,x,y) - b\frac{\partial}{\partial y}p(t,x,y). \tag{5.1.4}$$

在这节的最后, 我们看看随机微积分要做什么?

1. 一个轨道连续的随机过程是否可以用漂移布朗运动表示出来, 应该如何表示出来?
2. 随机微积分的牛顿-莱布尼茨公式是什么?

我们看看大家熟知的轨道连续的 $\mathbb{R}$ 上的 Markov 过程是否可以用漂移布朗运动“表示”出来. 这也是伊藤清 1942 年建立随机积分的原始想法. 设 $\{p(s,x;t,y), s < t, x, y \in \mathbb{R}\}$ 为 Markov 过程 $(X_t)_{t\geq 0}$ 的转移密度函数族, 即对于任意有界可测函数 f

$$\mathrm{E}\big[f(X_t)\big|\mathcal{F}_s\big] = \mathrm{E}\big[f(X_t)\big|\sigma(X_s)\big] = \int f(y)p(s,X_s;t,y)\mathrm{d}y.$$

我们不知道如何在随机变量空间上“求导”, 而且对于随机过程的一个观测 $(f(X_t))_{t\geq 0}$, 当 f 连续可微时, 关于任意的样本点 ω, 它定义了 $\mathbb{R}$ 上的一条连续曲线, 这条曲线也不一定可微. 进一步, 我们看这条曲线的平均

$$\mathrm{E}(f(X_t)|X_s = x) = \int f(y)p(s,x;t,y)\mathrm{d}y.$$

为方便我们记: $\mathrm{E}^{s,x}f(X_t) = \mathrm{E}(f(X_t)|X_s = x)$. $\mathrm{E}^{s,x}f(X_t)$ 定义了一条从 (s,x) 出发的曲线, 它是否可微? 我们来看 $\mathrm{E}^{s,x}f(X_t)$ 在无穷小未来时间内的变化,

$$\mathrm{E}^{s,x}f(X_{s+\tau}) = \int f(y)p(s,x;s+\tau,y)\mathrm{d}y, \ \tau > 0.$$

$f(X_s)$ 从 s 时刻 x 位置出发在无穷小未来的平均变化率为

$$\lim_{\tau\to 0^+}\frac{\mathrm{E}^{s,x}f(X_{s+\tau}) - \mathrm{E}^{s,x}f(X_s)}{\tau} = \lim_{\tau\to 0^+}\frac{\mathrm{E}^{s,x}f(X_{s+\tau}) - f(x)}{\tau},$$

形式地记为 $\frac{\mathrm{d}\mathrm{E}^{s,x}f(X_{s+\tau})}{\mathrm{d}s}\Big|_s$.

与 Einstein 的想法类似, 由于轨道的连续性, 在 τ 很小时, 过程跑出 x 的一个小邻域的概率非常得小, 因此当假设观测函数 f 有很好的可微性时, 我们可以对 $f(x)$ 在 x 附近泰勒展开到二阶, 形式上得到

$$\begin{aligned}
&\mathrm{E}^{s,x}f(X_{s+\tau}) - \mathrm{E}^{s,x}f(X_s) = \mathrm{E}^{s,x}f(X_{s+\tau}) - f(x)\\
= \ &\int (f(y) - f(x))p(s,x;s+\tau,y)\mathrm{d}y\\
\approx \ &\int (y-x)f'(x)p(s,x;s+\tau,y)\mathrm{d}y + \frac{1}{2}\int (y-x)^2 f''(x)p(s,x;s+\tau,y)\mathrm{d}y.
\end{aligned}$$

因此

$$\frac{\mathrm{dE}^{s,x}f(X_{s+\tau})}{\mathrm{d}s}\Big|_s = \lim_{\tau\to 0^+}\frac{1}{\tau}\Big[\int(y-x)p(s,x;s+\tau,y)\mathrm{d}y\Big]f'(x) + \frac{1}{2}\lim_{\tau\to 0^+}\frac{1}{\tau}\Big[\int(y-x)^2p(s,x;s+\tau,y)\mathrm{d}y\Big]f''(x).$$

若假设

$$\lim_{\tau\to 0^+}\frac{1}{\tau}\int(y-x)p(s,x;s+\tau,y)\mathrm{d}y = b(s,x), \tag{5.1.5}$$

$$\lim_{\tau\to 0^+}\frac{1}{\tau}\int(y-x)^2p(s,x;s+\tau,y)\mathrm{d}y = \sigma^2(s,x), \tag{5.1.6}$$

则

$$\frac{\mathrm{dE}^{s,x}f(X_{s+\tau})}{\mathrm{d}s}\Big|_s = \frac{1}{2}\sigma^2(s,x)\frac{\mathrm{d}^2f}{\mathrm{d}x^2}+b(s,x)\frac{\mathrm{d}f}{\mathrm{d}x}.$$

(5.1.5) 式中 $b(s,x)$ 的物理意义是表示在 s 时刻 x 位置, 过程向未来运动的平均瞬时**速度**; (5.1.6) 式中 $\sigma^2(s,x)$ 的物理意义表示在 s 时刻 x 位置, 过程向未来运动时偏离初始位置大小 (方差) 的瞬时变化率 (differential variance). (缺图!)

事实上, 直接可以计算出漂移布朗运动 $\sigma B_t + bt$ 的 “无穷小未来的变化率” 是

$$\lim_{\tau\downarrow 0^+}\frac{1}{\tau}\Big(\int f(y)p^{(\sigma,b)}(s,x;s+\tau,y)\mathrm{d}y - f(x)\Big) = \frac{1}{2}\sigma^2\frac{\mathrm{d}^2f}{\mathrm{d}x^2}+b\frac{\mathrm{d}f}{\mathrm{d}x}.$$

对比 $(\sigma B_t+bt)_{t\geq 0}$ 和 $(X_t)_{t\geq 0}$ 的 “微分特征”

$$\frac{1}{2}\sigma^2\frac{\mathrm{d}^2f}{\mathrm{d}x^2}+b\frac{\mathrm{d}f}{\mathrm{d}x},$$
$$\frac{1}{2}\sigma^2(s,x)\frac{\mathrm{d}^2f}{\mathrm{d}x^2}+b(s,x)\frac{\mathrm{d}f}{\mathrm{d}x},$$

可以清晰地看到, $(X_t)_{t\geq 0}$ 在 (s,x) 处的无穷小未来的变化可以近似的看成一个漂移布朗运动, 也就是说 $(X_t)_{t\geq 0}$ 在 (s,x) 处的 “切线” 或 “切过程” 是 $\forall t\in[s,s+\tau]$

$$\sigma(s,x)(B_t-B_s)+b(s,x)(t-s).$$

进一步, $(X_t(\omega))_{t\geq 0}$ 在 $(s,X_s(\omega))$ 处的无穷小未来可以近似看成

$$\sigma\big(s,X_s(\omega)\big)\big(B_t(\omega)-B_s(\omega)\big)+b\big(s,X_s(\omega)\big)(t-s).$$

因此, 对于时间划分 $0=t_0<t_1<\cdots<t_n=t$ 时, $\lambda=\max_k|t_k-t_{k-1}|$ 很小时, $X_t(\omega)-X_0(\omega)$ 近似等于

$$\begin{aligned} X_t(\omega)-X_0(\omega) &= \sum_{k=1}^{n}\big(X_{t_k}(\omega)-X_{t_{k-1}}(\omega)\big) \\ &\approx \sum_{k=0}^{n-1}\Big\{\sigma\big(t_k,X_{t_k}(\omega)\big)\big[B_{t_{k+1}}(\omega)-B_{t_k}(\omega)\big]+b\big(t_k,X_{t_k}(\omega)\big)(t_{k+1}-t_k)\Big\}. \end{aligned} \tag{5.1.7}$$

自然的问题是当划分 $\lambda\to 0$ 时, 上式是否收敛? 在什么意义下收敛? 第二项实际上是关于 t 的 Riemann 积分, 所以收敛到 $\int_0^t b(s,X_s(\omega))\mathrm{d}s$. 第一项

$$\sum_{k=0}^{n-1}a\big(t_k,X_{t_k}(\omega)\big)\big[B_{t_{k+1}}(\omega)-B_{t_k}(\omega)\big]=\sum_{k=0}^{n-1}\sigma\big(t_k,X_{t_k}(\omega)\big)\Delta B_{t_k}$$

是否也按 Riemann-Stieltjes 积分的定义收敛? 答案是否定的, 那么它应该如何收敛? 这就是 Itô 随机积分所要回答的问题.

5.2 Riemann-Stieltjes 积分

$[a,b]$ 是 $\mathbb{R}$ 上的一个有界闭区间. $\Pi:=\{a=t_0<t_1<\cdots<t_n=b\}$ 是 $[a,b]$ 的一个划分, 记 $|\Pi|=\max_{j=1\cdots,n}\{t_j-t_{j-1}\}$; ξ_j^Π 是 $[t_j,t_{j+1}]$, $j=1\cdots,n$ 中的任意一个点.

定义 5.2.1. 设 $\alpha(t)$ 和 $f(t)$ 都是 $[a,b]$ 到 $\mathbb{R}$ 的函数, 对于 $[a,b]$ 的划分 Π, 称

$$S^{\Pi}(f;\alpha) := \sum_{t_j,t_{j-1}\in\Pi} f(\xi_{j-1}^{\Pi})\big(\alpha(t_j)-\alpha(t_{j-1})\big)$$

为 f 的 Riemann-Stieltjes 和.

定义 5.2.2. 设 $\alpha(t)$ 和 $f(t)$ 都是 $[a,b]$ 到 $\mathbb{R}$ 的函数, 称 f 关于 α 是 Riemann-Stieltjes 可积的, 若当 $|\Pi|\to 0$ 时, $S^{\Pi}(f;\alpha)$ 收敛到不依赖于 ξ_j^{Π} 选取的有限值. 这个值称为 f 关于 α 的 Riemann-Stieltjes 积分, 记为

$$\int_a^b f(t)\,\mathrm{d}\alpha(t) := \lim_{|\Pi|\to 0}\sum_{t_j,t_{j-1}\in\Pi} f(\xi_{j-1}^{\Pi})\big(\alpha(t_j)-\alpha(t_{j-1})\big).$$

定义 5.2.3. $\Pi=\{a=t_0<t_1<\cdots<t_n=b\}$ 为 $[a,b]$ 的一个划分. 设 $\mathrm{V}_\Pi f := \sum_{i=1}^n |f(t_i)-f(t_{i-1})|$. 称 $\mathrm{V}f = \sup\limits_{[a,b]\text{的所有划分}\Pi} \mathrm{V}_\Pi f$ 为 f 的全变差. 若 $\mathrm{V}f<\infty$, 则称 f 为有界变差函数.

例 5.2.1. 设 f 是 $[a,b]$ 上的连续函数, α 是单调函数, 则 f 关于 α 是 Riemann-Stieltjes 可积的.

定理 5.2.1. 设 $\alpha(t)$ 是 $[a,b]$ 到 $\mathbb{R}$ 的函数. 任意的 $[a,b]$ 上的连续函数 f 都是关于 α 是 Riemann-Stieltjes 可积的, 则 α 是有界变差函数.

例 5.2.2. 设 f 是 $[a,b]$ 上的连续函数, α 是有界变差函数, 则 f 关于 α 是 Riemann-Stieltjes 可积的.

例 5.2.3. 设 $\alpha(t)=\begin{cases} 1 & t\in[\frac{1}{2},1] \\ 0 & t\in[0,\frac{1}{2}) \end{cases}$, 则 α 关于 α 自己不是 Riemann-Stieltjes 可积的[①].

5.3 布朗运动的平方变差性质

在应用随机过程中知道对于几乎处处的 ω, 布朗运动的轨道 $B_t(\omega)$ 是一个无处可微的连续函数, 因此不是有界变差函数, 我们不能在 Stieltjes 积分的意义下定义 $\int f(t)\mathrm{d}B_t(\omega)$. 当然, 我们也可以从其它的角度来证明布朗运动的轨道不是有界变差函数.

引理 5.3.1. (Lévy 的振动性质) 设 $(B_t)_{t\geq 0}$ 为布朗运动, 又

$$s_1=t_0<t_1<\cdots<t_n=s_2,$$
$$\Delta t_k = t_{k+1}-t_k,\ \Delta B_{t_k}=B_{t_k+1}-B_{t_k},\ h=\max_{0\leq k\leq n-1}\Delta t_k,$$

那么,

$$\mathrm{E}\Big|\sum_{k=0}^{n-1}(\Delta B_{t_k})^2-(s_2-s_1)\Big|^2 \leq 2h(s_2-s_1). \tag{5.3.1}$$

证明:

$$\begin{aligned}
\text{(5.3.1)式左端} \quad &= \quad \mathrm{E}\Big|\sum_{k=0}^{n-1}[(\Delta B_{t_k})^2-\Delta t_k]\Big|^2 \\
&= \quad \sum_{k=0}^{n-1}\mathrm{E}\big[(\Delta B_{t_k})^2-\Delta t_k\big]^2+\sum_{k\neq j}\mathrm{E}\Big(\big[(\Delta B_{t_k})^2-\Delta t_k\big]\big[(\Delta B_{t_j})^2-\Delta t_j\big]\Big).
\end{aligned}$$

[①]但 α 关于 α 生成的 Lebesgue-Stietljes 测度 μ_α 是 Lebesgue 可积的, $\int_{[0,1]}\alpha(s)\mathrm{d}\mu_\alpha(s)=1$.

由于当 $k \neq j$ 时, ΔB_{t_k} 与 ΔB_{t_j} 相互独立且 $\mathrm{E}(\Delta B_{t_k})^2 = \Delta t_k$, 故第二项等于零. 而第一项为

$$\begin{aligned}\sum_{k=0}^{n-1} \mathrm{E}\big[(\Delta B_{t_k})^2 - \Delta t_k\big]^2 &= \sum_{k=0}^{n-1} \mathrm{E}\big[(\Delta B_{t_k})^4 + (\Delta t_k)^2 - 2(\Delta B_{t_k})^2 \cdot \Delta t_k\big] \\ &= \sum_{k=0}^{n-1} \Big[3\big(\mathrm{E}(\Delta B_{t_k})^2\big)^2 - (\Delta t_k)^2\Big] = 2\sum_{k=0}^{n-1} (\Delta t_k)^2 \\ &= 2\sum_{k=0}^{n-1} |\Delta t_k| \cdot |\Delta t_k| \\ &\leq 2h \sum_{k=0}^{n-1} \Delta t_k = 2h(s_2 - s_1).\end{aligned}$$

□

推论 5.3.1. (布朗运动平方变差有限)

$$\lim_{h\to 0} \sum_{k=0}^{n-1} (\Delta B_{t_k})^2 \xrightarrow{L^2} (s_2 - s_1).$$

即

$$\lim_{h\to 0} \mathrm{E}\Big|\sum_{k=0}^{n-1} (\Delta B_{t_k})^2 - (s_2 - s_1)\Big|^2 = 0.$$

引理 5.3.2. *设 $[s_1, s_2]$ 的 2^n 等分点为*

$$s_1 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{2^n}^{(n)} = s_2,$$

那么

$$\mathrm{P}\Big(\omega : \lim_{n\to\infty} \sum_{k=0}^{2^n-1} (\Delta B_{t_k})^2 = s_2 - s_1\Big) = 1.$$

引理 5.3.3. *以概率为 1 的布朗运动的轨道在 t 的任何区间内都不是有界变差的.*

证明: 令 $\{t_k^{(n)}\}$ 为引理5.3.2中的划分, 记

$$\lambda_n(\omega) = \max_k \{|\Delta B_{t_k^{(n)}}(\omega)|\},$$

因为 $B_t(\omega)$ 在 $[s_1, s_2]$ 上一致连续, 所以当 $\max_k \Delta t_k^{(n)} \to 0$ 时 $\lambda_n(\omega) \to 0$. 因此,

$$\sum_{k=0}^{2^n-1} (\Delta B_{t_k^{(n)}})^2 \leq \lambda_n(\omega) \sum_k |\Delta B_{t_k^{(n)}}(\omega)|,$$

由引理5.3.2, 对于几乎处处的轨道有

$$\sum_k |\Delta B_{t_k^{(n)}}(\omega)| \geq \frac{\sum_{k=0}^{2^n-1} (\Delta B_{t_k^{(n)}})^2}{\lambda_n(\omega)} \xrightarrow{n\to\infty} \infty.$$

□

♣♣♣20230427

5.4 关于布朗运动的 Itô 积分

由前面的引言以及积分的思想, 我们知道, 首先应该对某种 “简单过程” 来定义 “随机积分”, 然后再看 “一般过程” 是否可以表示成这种 “简单过程” 的 “极限”, 最后再看对于逼近的 “简单过程” 的 “积分” 是否收敛? 若收敛, 则定义这个极限就是 “一般过程” 的 “积分”.

设 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathrm{P})$ 是一个带流的概率空间, $(B_t)_{t\geq 0}$ 为关于事件域流 $(\mathcal{F}_t)_{t\geq 0}$ 的布朗运动, 随机过程 $(\phi_t)_{t\geq 0}$ 关于 $(\mathcal{F}_t)_{t\geq 0}$ 适应.

定义 5.4.1. 对于区间 $[0,T]$ 的一个划分 $0=t_0<t_1<\cdots<t_n=T$, 当随机过程 $(\phi_t)_{t\geq 0}$ 是阶梯过程, 即它可以写成如下形式时

$$\phi_t = \sum_{i=0}^{n-1} \phi_{t_i} \mathbf{1}_{(t_i, t_{i+1}]}(t) + \phi_0 \mathbf{1}_{\{0\}}(t),$$

若其中的系数随机变量 ϕ_{t_i} 为 $\mathcal{F}_{t_i}$ 可测的, 而且 ϕ_{t_i}, $i=0,\cdots,n-1$ 为有界的随机变量 (或 $\mathrm{E}|\phi_{t_i}|^2<\infty$), 则称 $(\phi_t)_{t\geq 0}$ 为 $(\mathcal{F}_t)_{t\geq 0}$ 可料阶梯过程[①].

定义 5.4.2. 设 $(\phi_t)_{t\geq 0}$ 为 $(\mathcal{F}_t)_{t\geq 0}$ 可料的阶梯过程. 定义 $(\phi_t)_{t\geq 0}$ 关于布朗运动 $(B_t)_{t\geq 0}$ 的 Itô 随机积分如下

$$\int_0^T \phi_t \mathrm{d}B_t = \sum_{i=0}^{n-1} \phi_{t_i}(B_{t_{i+1}} - B_{t_i}),$$

对于任意固定的时刻 T, 它是一个随机变量.

引理 5.4.1. 可料阶梯过程 $(\phi_t)_{t\geq 0}$ 关于 $(B_t)_{t\geq 0}$ 的 Itô 随机积分满足

(1) $\mathrm{E}\big(\int_0^T \phi_t \mathrm{d}B_t\big) = 0$;

(2) Itô 等距

$$\mathrm{E}\Big|\int_0^T \phi_t \mathrm{d}B_t\Big|^2 = \Big\|\int_0^T \phi_t \mathrm{d}B_t\Big\|^2 = \mathrm{E}\int_0^T |\phi_t|^2 \mathrm{d}t;$$

(3) 线性性质

$$\begin{aligned}\int_0^T (\phi+\psi)\,\mathrm{d}B &= \int_0^T \phi\,\mathrm{d}B + \int_0^T \psi\,\mathrm{d}B,\\ \int_0^T c\,\phi\,\mathrm{d}B &= c\int_0^T \phi\,\mathrm{d}B,\quad c\text{为常数}.\end{aligned}$$

证明:

(1)

$$\begin{aligned}\mathrm{E}\Big(\int_0^T \phi_t \mathrm{d}B_t\Big) &= \mathrm{E}\sum_{i=0}^{n-1} \phi_{t_i}(B_{t_{i+1}} - B_{t_i}) \qquad (5.4.1)\\ &= \sum_{i=0}^{n-1} \mathrm{E}\,\phi_{t_i}\,(B_{t_{i+1}} - B_{t_i}).\end{aligned}$$

注意到 ϕ_{t_i} 为 $\mathcal{F}_{t_i}$ 可测, 故 ϕ_{t_i} 与 $B_{t_{i+1}} - B_{t_i}$ 独立, 因此

$$(5.4.1)\text{的右端} = \sum_{i=0}^{n-1} \mathrm{E}\,\phi_{t_i}\,\mathrm{E}(B_{t_{i+1}} - B_{t_i}) = 0.$$

[①] 有些书上称可料阶梯过程为简单 (可料) 适应过程或初等 (可料) 适应过程. 可料性在关于跳过程的随机积分理论中特别重要.

(2)

$$
\begin{aligned}
&\mathrm{E}\Big|\int_0^T \phi_t \mathrm{d}B_t\Big|^2 = \mathrm{E}\Big(\sum_{i=0}^{n-1}\phi_{t_i}(B_{t_{i+1}}-B_{t_i})\Big)^2 \\
= \quad & \mathrm{E}\Big(\sum_{i=0}^{n-1}\phi_{t_i}^2(B_{t_{i+1}}-B_{t_i})^2 + 2\sum_{j<k}\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\Big) \\
= \quad & \sum_{i=0}^{n-1}\mathrm{E}\Big(\mathrm{E}\big[\phi_{t_i}^2(B_{t_{i+1}}-B_{t_i})^2\big|\mathcal{F}_{t_i}\big]\Big) \\
& +2\sum_{j<k}\mathrm{E}\Big(\mathrm{E}\big[\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\big|\mathcal{F}_{t_k}\big]\Big) \\
= \quad & \sum_{i=0}^{n-1}\mathrm{E}\Big(\phi_{t_i}^2\mathrm{E}\big[(B_{t_{i+1}}-B_{t_i})^2\big|\mathcal{F}_{t_i}\big]\Big) \\
& +2\sum_{j<k}\mathrm{E}\Big(\mathrm{E}\big[\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\big|\mathcal{F}_{t_k}\big]\Big) \\
= \quad & \sum_{i=0}^{n-1}\mathrm{E}\big(\phi_{t_i}^2\mathrm{E}(B_{t_{i+1}}-B_{t_i})^2\big) \\
& +2\sum_{j<k}\mathrm{E}\Big(\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})\mathrm{E}\big[B_{t_{k+1}}-B_{t_k}\big|\mathcal{F}_{t_k}\big]\Big) \\
\overset{\text{第二项为 0}}{=} \quad & \sum_{i=0}^{n-1}\mathrm{E}\big(\phi_{t_i}^2(t_{i+1}-t_i)\big) \\
= \quad & \mathrm{E}\sum_{i=0}^{n-1}\big(\phi_{t_i}^2(t_{i+1}-t_i)\big) \\
= \quad & \mathrm{E}\int_0^T|\phi_t|^2\mathrm{d}t.
\end{aligned}
$$

□

对于一般的过程应该如何定义?

$\mathscr{L}_T^2=\big\{(\phi_t)_{0\le t\le T}$ 是 $(\mathcal{F}_t)_{t\ge0}$ 适应的随机过程, 且 $\mathrm{E}\int_0^T|\phi_t(\omega)|^2\mathrm{d}t<+\infty\big\}$[①]

(1) $\mathscr{L}_T^2$ 是一个线性空间;

(2) $\mathscr{L}_T^2$ 中可以定义内积

$$\phi,\psi\in\mathscr{L}_T^2,\ \text{则}\langle\phi,\psi\rangle=\mathrm{E}\int_0^T\big(\phi_t(\omega)\psi_t(\omega)\big)\mathrm{d}t.$$

(3) 因此 $\mathscr{L}_T^2$ 在内积 $\langle\phi,\psi\rangle$ 下是一个 (无穷维) 欧氏空间, $\mathscr{L}_T^2$ 在距离

$$\|\phi-\psi\|=\Big(\mathrm{E}\int_0^T\big|\phi_t-\psi_t\big|^2\mathrm{d}t\Big)^{\frac{1}{2}},$$

下成为一个完备的距离空间, 也就是说 $(\phi^{(n)})_{n\ge1}$ 是 Cauchy 列, 当 $n\to\infty$, $m\to\infty$ 时, $\|\phi_n-\psi_m\|\to0$, 则必有 $\phi\in\mathscr{L}_T^2$ 使得 $\|\phi^{(n)}-\psi\|\to0$, $n\to\infty$, 即 $\mathscr{L}_T^2$ 是一个 Hilbert 空间.

引理 5.4.2. 对于 $\mathscr{L}_T^2$ 中的随机过程 $(\phi_t)_{t\ge0}$ 必存在 $(\mathcal{F}_t)_{t\ge0}$ 可料的阶梯随机过程列 $(\phi_t^{(n)})_{t\ge0}$, 当 $n\to\infty$ 时

$$\|\phi^{(n)}-\phi\|^2=\mathrm{E}\int_0^T\big|\phi_t^{(n)}-\phi_t\big|^2\mathrm{d}t\to0,$$

[①] 严格说, $(\phi_t)_{0\le t\le T}$ 应该是循序可测 (progressively measurable) 过程.

从而 $(\phi^{(n)})_{n\geq 1}$ 也是 $\mathscr{L}_T^2$ 中的 Cauchy 列, 称为近似 $(\phi_t)_{t\geq 0}$ 的 $(\mathcal{F}_t)_{t\geq 0}$ 可料阶梯随机过程[①] .

例 5.4.1. 考虑 $\mathscr{L}_T^2$ 中的有界连续随机过程 $(\phi_t)_{t\geq 0}$ 及其在 $[0,T]$ 的一列划分

$$0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{N_n}^{(n)} = T,\ \lambda_n = \max_{0\leq k\leq N_n-1}\{t_{k+1}^{(n)} - t_k^{(n)}\}.$$

对于固定的 n, 设

$$\phi_t^{(n)} = \sum_{k=0}^{N_n-1} \phi_{t_k}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t) + \phi_0\mathbf{1}_{\{0\}}(t)$$

则

$$\|\phi^{(n)} - \phi\|^2 = \mathrm{E}\int_0^T \left|\phi_t^{(n)} - \phi_t\right|^2 \mathrm{d}t \to 0,\ \text{当}\lambda_n \to 0\text{时}.$$

因此, $(\phi_t^{(n)})_{t\geq 0}, n\geq 1$ 为一个近似 $(\phi_t)_{t\geq 0}(\mathcal{F}_t)_{t\geq 0}$ 可料阶梯随机过程序列.

引理 5.4.3. 设 $(\phi_t)_{t\geq 0}\in\mathscr{L}_T^2$, $(\phi_t^{(n)})_{t\geq 0}$ 是 $(\phi_t)_{t\geq 0}$ 近似的可料阶梯过程序列, 则 $(\int_0^T \phi_t^{(n)}\mathrm{d}B_t)_{n\geq 0}$ 是 $L^2(\Omega)$ 中的 Cauchy 列, 即当 $n,m\to\infty$ 时

$$\mathrm{E}\Big(\int_0^T \phi_t^{(n)}\mathrm{d}B_t - \int_0^T \phi_t^{(m)}\mathrm{d}B_t\Big)^2 \to 0.$$

证明: 由可料阶梯过程随机积分的线性性质和 Itô 等距性质我们得到:

$$\mathrm{E}\Big(\int_0^T \phi_t^{(n)}\mathrm{d}B_t - \int_0^T \phi_t^{(m)}\mathrm{d}B_t\Big)^2 = \mathrm{E}\Big(\int_0^T (\phi_t^{(n)} - \phi_t^{(m)})\mathrm{d}B_t\Big)^2 = \mathrm{E}\int_0^T (\phi_t^{(n)} - \phi_t^{(m)})^2\mathrm{d}t = \|\phi^{(n)} - \phi^{(m)}\|^2.$$

由于 $(\phi_t^{(n)})_{t\geq 0}$ 是 $(\phi_t)_{t\geq 0}$ 的近似可料阶梯序列, 从而 $(\phi_t^{(n)})_{t\geq 0}$ 是 $\mathscr{L}_T^2$ 中的 Cauchy 序列. 因此, 当 $n,m\to\infty$ 时, $\|\phi^{(n)} - \phi^{(m)}\|^2 \to 0$.

□

定义 5.4.3. 对于 $\mathscr{L}_T^2$ 中的随机过程 $(\phi_t)_{t\geq 0}$, $(\phi_t^{(n)})_{t\geq 0}$ 为其近似的 $(\mathcal{F}_t)_{t\geq 0}$ 可料阶梯随机过程列, 满足

$$\|\phi^{(n)} - \phi\|^2 = \mathrm{E}\int_0^T \left|\phi_t^{(n)} - \phi_t\right|^2\mathrm{d}t \to 0,\ \text{当}n\to\infty.$$

(这时 $\int_0^T \phi_t^{(n)}\mathrm{d}B_t$ 是 $L^2(\Omega)$ 中的 Cauchy 列), 记 $\int_0^T \phi_t^{(n)}\mathrm{d}B_t$ 在 $L^2(\Omega)$ 中的极限为 $\int_0^T \phi_t\,\mathrm{d}B_t$, 即

$$\int_0^T \phi_t\mathrm{d}B_t := L^2(\Omega) - \lim_{n\to\infty}\int_0^T \phi_t^{(n)}\mathrm{d}B_t,$$

称之为 $(\phi_t)_{0\leq t\leq T}$ 关于 $(B_t)_{t\geq 0}$ 的 Itô 随机积分[②].

在 Itô 随机积分定义中, 引理5.4.1的 Itô 等距起到了关键性作用. 而引理5.4.1的 Itô 等距成立的本质原因是 $(\phi_t^{(n)})_{t\geq 0}$ 的可料性.

事实上还可以证明使用不同的 $(\mathcal{F}_t)_{t\geq 0}$ 可料阶梯随机过程近似列, 在不计零概率事件差异下不会影响随机积分的定义. $\int_0^T \phi_t\,\mathrm{d}B_t$ 常常简记为 $\int_0^T \phi\,\mathrm{d}B$.

[①] 为什么非要取可料的近似列?

(1) 从这一章引言中我们知道 Itô 当年就是利用漂移布朗运动来近似 Markov 过程在 s 时刻无穷小未来的想法来逼近整个 Markov 过程,

$$\sum a(t_k, X_{t_k})(B_{t_{k+1}} - B_{t_k}) + b(t_k, X_{t_k})(t_{k+1} - t_k),$$

从而定义了 Itô 随机积分.

(2) 从引理5.4.1的证明知, 当 (ϕ_t) 为 $(\mathcal{F}_t)_{t\geq 0}$ 可料时, 随机积分有很好的性质, 例如均值为零和 Itô 等距.

(3) 后面我们还可以看作随机积分当 T 变动时成为一个鞅.

[②] 类似地可以定义

$$\int_a^b \phi\,\mathrm{d}B = \int_0^b \phi\,\mathrm{d}B - \int_0^a \phi\,\mathrm{d}B.$$

或者直接类似于以上的定义方式直接在区间 $[a,b]$ 上定义随机积分.

性质 5.4.1. 对于 $\mathscr{L}_T^2$ 中的随机过程列 $\big((\phi_t^{(n)})_{t\geq 0}\big)_n$ 和随机过程 $(\phi_t)_{t\geq 0}$, 若

$$\|\phi^{(n)}-\phi\|^2=\mathrm{E}\int_0^T\left|\phi_t^{(n)}-\phi_t\right|^2\mathrm{d}t\to 0,\ \text{当}n\to\infty,$$

则 $\int_0^T\phi_t\mathrm{d}B_t=L^2(\Omega)-\lim_{n\to\infty}\int_0^T\phi_t^{(n)}\mathrm{d}B_t$.

从下面的图表我们不能看出随机积分的定义与 Riemann 积分或 Lebesgue 积分的定义想法是类似的

Riemann 积分或 Lebesgue 积分	**随机积分**
定义简单函数的积分	定义可料阶梯 (简单适应) 过程的随机积分
简单函数逼近一般函数 $f_n\to f$	可料阶梯过程逼近适应过程 $\phi^{(n)}\to\phi$
$\int f_n\mathrm{d}x$ 在$\mathbb{R}$收敛($\mathbb{R}$完备)	$\int\phi^{(n)}\mathrm{d}B$在$L^2(\Omega)$收敛 ($L^2(\Omega)$完备)
$\int f_n\mathrm{d}x$的极限定义为$\int f\mathrm{d}x$	$\int\phi^{(n)}\mathrm{d}B$的极限定义为$\int\phi\mathrm{d}B$

例 5.4.2. 求 $\int_0^t B_s\mathrm{d}B_s$.

解: $B_t^{(n)}=\sum\limits_{k=0}^{N_n-1}B_{t_k^{(n)}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t)$, $B_t^{(n),M}=\sum\limits_{k=0}^{N_n-1}(-M\vee B_{t_k^{(n)}}\wedge M)\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t)$, 其中 M 是正整数. 可以证明当 $M\to\infty$, $\mathrm{E}\int_0^T|\ B_t^{(n),M}-B_t^{(n)}|^2\mathrm{d}t\to 0$. 由可料阶梯过程随机积分的定义,

$$\int_0^t B_s^{(n),M}\mathrm{d}B_s=\sum_k(-M\vee B_{t_k^{(n)}}\wedge M)\big(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}}\big)$$

当 $M\to\infty$ 时, 利用上述性质5.4.1我们得到上式左端为 $\int_0^t B_s^{(n)}\mathrm{d}B_s$; 进一步可以证明当 $M\to\infty$ 时, 上式右端为 $\sum\limits_k B_{t_k^{(n)}}\big(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}}\big)$. 对上式两端关于 $M\to\infty$ 取 L^2 的极限, 我们得到

$$\begin{aligned}\int_0^t B_s^{(n)}\mathrm{d}B_s &= \sum_k B_{t_k^{(n)}}\big(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}}\big)\\ &= \frac{1}{2}\sum_k\big(2B_{t_k^{(n)}}B_{t_{k+1}^{(n)}}-2B_{t_k^{(n)}}^2\big)\\ &= \frac{1}{2}\sum_k\Big(\big(B_{t_{k+1}^{(n)}}^2-B_{t_k^{(n)}}^2\big)-\big(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}}\big)^2\Big)\\ &= \frac{1}{2}\sum_k\big(B_{t_{k+1}^{(n)}}^2-B_{t_k^{(n)}}^2\big)-\frac{1}{2}\sum_k\big(\Delta B_{t_k^{(n)}}\big)^2\\ &= \frac{1}{2}(B_t^2-B_0^2)-\frac{1}{2}\sum_k\big(\Delta B_{t_k^{(n)}}\big)^2\end{aligned}$$

由上一节引理5.3.1的 Lévy 振动性质知 $\mathrm{E}\big(\sum\limits_k\big(\Delta B_{t_k^{(n)}}\big)^2\big)^2\to t$, 因此

$$\mathrm{E}\Big(\int_0^t B_s^{(n)}\mathrm{d}B_s-\big(\frac{1}{2}(B_t^2-B_0^2)-\frac{t}{2}\big)\Big)^2\to 0,$$

故

$$\int_0^t B\mathrm{d}B=\frac{1}{2}(B_t^2-B_0^2)-\frac{1}{2}t.$$

当 $B_0=0$ 时, $\int_0^t B\mathrm{d}B=\frac{1}{2}B_t^2-\frac{1}{2}t$.

例 5.4.3. 为什么非要取划分左端点的值, 用可料逼近呢? 为什么不能类似 Riemann-Stieltjes 积分在划分区间中的任意点上取值呢? 下面我们设 $(B_t)_{t\geq 0}$ 标准布朗运动.

在例5.4.2中取的是划分左端点的值 $\sum\limits_{k=0}^{N_n-1}B_{t_k^{(n)}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t)$ 来逼近的 $(B_t)_{t\geq 0}$. 现在用划分右端点的值来做逼近序列看看会有什么结果呢?

$$\tilde{\phi}_t^{(n)}=\sum_k B_{t_{k+1}^{(n)}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t).$$

考虑

$$H_n^{\tilde{\phi}} \equiv \sum_k B_{t_{k+1}^{(n)}} \Delta B_{t_k^{(n)}} = \sum_{k=0}^{N_n-1} B_{t_{k+1}^{(n)}} \big(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}\big),$$

的 $L^2(\Omega)$ 极限, 则

$$H_n^{\tilde{\phi}} \xrightarrow{L^2} \frac{1}{2}B_t^2 + \frac{1}{2}t.$$

若取

$$\Psi_t^{(n)} = \sum_{k=0}^{N_n-1} \frac{B_{t_k^{(n)}} + B_{t_{k+1}^{(n)}}}{2} \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t),$$

$$H_n^{\Psi} \equiv \sum_{k=0}^{N_n-1} \frac{B_{t_k^{(n)}} + B_{t_{k+1}^{(n)}}}{2} \Delta B_{t_k^{(n)}} \xrightarrow{L^2} \frac{1}{2}B_t^2.$$

若取

$$\xi_t^{(n)} = \sum_{k=0}^{N_n-1} B_{\frac{t_k^{(n)} + t_{k+1}^{(n)}}{2}} \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t),$$

$$H_n^{\xi} \equiv \sum_{k=0}^{N_n-1} B_{\frac{t_k^{(n)} + t_{k+1}^{(n)}}{2}} \Delta B_{t_k^{(n)}} \xrightarrow{L^2} \frac{1}{2}B_t^2.$$

□

因此, 从以上例子可以看出, 对于随机积分并不是非要在 $[t_k^{(n)}, t_{k+1}^{(n)}]$ 取左端点作为“代表点”才能保证收敛性的, 也可以在其它位置取点, 但其求和的极限是不同的. 问题是那么取左端点逼近定义的随机积分又有什么好处呢? 我们马上就可以通过随机积分的性质看到.

性质 5.4.2. 随机积分性质 I

1. 可测性: $\int_0^t \phi\, \mathrm{d}B$ 关于 $\mathcal{F}_t$ 可测.

2. 线性性质:

$$\int_0^t (\phi + \psi) \mathrm{d}B = \int_0^t \phi\, \mathrm{d}B + \int_0^t \psi\, \mathrm{d}B,$$

$$\int_0^t c\,\phi\, \mathrm{d}B = c\int_0^t \phi\, \mathrm{d}B,\ c\text{为常数}.$$

   对于 $\mathcal{F}_a$ 可测的有界随机变量 η,

$$\int_a^b \eta\phi\, \mathrm{d}B = \eta\int_a^b \phi\, \mathrm{d}B.$$

3. 可加性: 对 $a < b < c$ 有

$$\int_a^c \phi\, \mathrm{d}B = \int_a^b \phi\, \mathrm{d}B + \int_b^c \phi\, \mathrm{d}B.$$

4. Itô 随机积分的数学期望, 协方差和方差

$$\mathrm{E}\big(\int_0^t \phi_s \mathrm{d}B_s\big) = 0,$$

$$\mathrm{E}\big(\int_0^t \phi_s \mathrm{d}B_s \int_0^t \psi_s \mathrm{d}B_s\big) = \mathrm{E}\int_0^t (\phi_s\psi_s) \mathrm{d}s.$$

   特别 $\phi = \psi$ 时,

$$\mathrm{E}\big(\int_0^t \phi_s \mathrm{d}B_s\big)^2 = \mathrm{E}\big(\int_0^t \phi_s^2 \mathrm{d}s\big). \tag{5.4.2}$$

   (5.4.2) 被称为 Itô 等距, 逼近序列的“可料性”在证明中起了关键的作用.

另外, $\left|\int_0^t \phi_s \mathrm{d}B_s\right| \le \int_0^t |\phi_s| \mathrm{d}B_s$ 是错误的.

进一步, 我们可以考虑 $0 < t \le T$, 对于 $\phi \in \mathscr{L}_T^2$, 则可以定义一个随机过程 $\left(\int_0^t \phi_s \mathrm{d}B_s\right)_{t\in[0,T]}$. 对于 $\forall t$, $\int_0^t \phi_s \mathrm{d}B_s$ 是几乎处处定义的, 但总可以找到这样一个 $(\mathcal{F}_t)_{t\ge 0}$ 适应过程 $(z_t)_{t\in[0,T]}$ 满足 $(z_t)_{0\le t\le T}$ 是连续过程, 而且 $\forall t$

$$\mathrm{P}\Big(\int_0^t \phi_s \mathrm{d}B_s = z_t\Big) = 1.$$

因此, 以后总认为 $\left(\int_0^t \phi_s \mathrm{d}B_s\right)_{t\in[0,T]}$ 是一个连续的关于 $(\mathcal{F}_t)_{t\ge 0}$ 适应的随机过程.

由 Itô 随机积分定义的随机过程 $\left(\int_0^t \phi_s \mathrm{d}B_s\right)_{t\in[0,T]}$ 还有以下的性质.

性质 5.4.3. 随机积分性质 II

1. τ 是 $(\mathcal{F}_t)_{t\ge 0}$ 停时, 若 $\tau \le T$, 则
$$\int_0^\tau \phi_t \mathrm{d}B_t = \int_0^T \phi_t \mathbf{1}_{(0,\tau]}(t) \mathrm{d}B_t.$$
2. $\left\{\xi_t \overset{\text{def}}{=} \int_0^t \phi_s \mathrm{d}B_s\right\}$ 是 $(\mathcal{F}_t)_{t\in[0,T]}$ 鞅.
3. $\left\{\eta_t \overset{\text{def}}{=} (\int_0^t \phi_s \mathrm{d}B_s)^2 - \int_0^t \phi_s^2 \mathrm{d}s\right\}$ 是 $(\mathcal{F}_t)_{t\in[0,T]}$ 鞅, 且
$$\mathrm{E}\Big[\Big(\int_s^t \phi_u \mathrm{d}B_u\Big)^2\Big|\mathcal{F}_s\Big] = \mathrm{E}\Big[\int_s^t \phi_u^2 \mathrm{d}u\Big|\mathcal{F}_s\Big].$$
4. 若 $(\phi_t)_{t\in[0,T]}$ 是关于 $(\mathcal{F}_t)_{t\in[0,T]}$ 适应的有界随机过程, 即 $\forall t$, $|\phi_t| < M < \infty$, 则
$$\Big\{\zeta_t \overset{\text{def}}{=} \mathrm{e}^{\int_0^t \phi_s \mathrm{d}B_s - \frac{1}{2}\int_0^t \phi_s^2 \mathrm{d}s}\Big\}$$
为 $(\mathcal{F}_t)_{t\in[0,T]}$ 鞅[①].

证明: 只证 (2)(3). 当 $(\phi_t)_{t\ge 0}$ 是可料阶梯过程时

$$\phi_t = \sum_{i=0}^{n-1} \phi_{t_i} \mathbf{1}_{(t_i,t_{i+1}]}(t) + \phi_0 \mathbf{1}_{\{0\}}(t),$$

对于 $s < t'$, 将 s 和 t' 加入到 $0 = t_0 < t_1 < \cdots < t_j < \cdots$ 成为

$$0 = t_0 < \cdots < t_i < s < t' \le t_{i+1} < \cdots$$

或者

$$0 = t_0 < \cdots < t_i < s \le t_{i+1} < \cdots < t_j < t' \le t_{j+1} \cdots.$$

重新记这些划分点为

$$0 = t_0' < t_1' < \cdots < t_p' < \cdots < t_q' < \cdots,$$

[①] 上面 ϕ 的有界性条件可以进一步减弱, 例如 Novikov 条件 $\mathrm{E}\mathrm{e}^{\frac{1}{2}\int_0^t \phi_s^2 \mathrm{d}s} < +\infty$ 就可以保证鞅性.

其中 $s=t'_p,\ t'=t'_q$.

$$\begin{aligned}
&\mathrm{E}\Big[\int_0^{t'}\phi_u\mathrm{d}B_u\Big|\mathcal{F}_s\Big]\\
=\ &\sum_{k=0}^{q-1}\mathrm{E}\Big[\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})\Big|\mathcal{F}_{t'_p}\Big]\\
=\ &\sum_{k=0}^{p-1}\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})+\sum_{k=p}^{q-1}\mathrm{E}\Big[\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})\Big|\mathcal{F}_{t'_p}\Big]\\
=\ &\sum_{k=0}^{p-1}\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})+\sum_{k=p}^{q-1}\mathrm{E}\Big[\mathrm{E}\big[\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})\big|\mathcal{F}_{t'_k}\big]\Big|\mathcal{F}_{t'_p}\Big]\\
=\ &\sum_{k=0}^{p-1}\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})+\sum_{k=p}^{q-1}\mathrm{E}\Big[\mathrm{E}\phi_{t'_k}\mathrm{E}(B_{t'_{k+1}}-B_{t'_k})\Big|\mathcal{F}_{t'_p}\Big]\\
=\ &\sum_{k=0}^{p-1}\phi_{t'_k}(B_{t'_{k+1}}-B_{t'_k})\\
=\ &\int_0^s\phi_u\mathrm{d}B_u.
\end{aligned}$$

对于 $\phi\in\mathscr{L}_T^2$, 存在着一列可料阶梯过程 $\phi^{(n)}$ 在 $\mathscr{L}_T^2$ 中逼近 ϕ. 对于任意的 $A\in\mathcal{F}_s$,

$$\begin{aligned}
&\Big|\mathrm{E}\Big(\int_0^{t'}\phi\,\mathrm{d}B\mathbf{1}_A\Big)-\mathrm{E}\Big(\int_0^{t'}\phi^{(n)}\mathrm{d}B\mathbf{1}_A\Big)\Big|\\
\le\ &\Big(\mathrm{E}\Big|\int_0^{t'}(\phi-\phi^{(n)})\mathrm{d}B\Big|^2\Big)^{\frac12}\mathrm{P}(A)^{\frac12}\\
=\ &\Big(\mathrm{E}\int_0^{t'}(\phi-\phi^{(n)})^2\mathrm{d}u\Big)^{\frac12}\mathrm{P}(A)^{\frac12}\\
\stackrel{n\to\infty}{\longrightarrow}\ &0.
\end{aligned}$$

同理

$$\Big|\mathrm{E}\Big(\int_0^{s}\phi\,\mathrm{d}B\mathbf{1}_A\Big)-\mathrm{E}\Big(\int_0^{s}\phi^{(n)}\mathrm{d}B\mathbf{1}_A\Big)\Big|\stackrel{n\to\infty}{\longrightarrow}0.$$

注意到 $\mathrm{E}\big(\int_0^{t'}\phi^{(n)}\mathrm{d}B\mathbf{1}_A\big)=\mathrm{E}\big(\int_0^{s}\phi^{(n)}\mathrm{d}B\mathbf{1}_A\big)$, 因此

$$\mathrm{E}\Big(\int_0^{t'}\phi\,\mathrm{d}B\mathbf{1}_A\Big)=\mathrm{E}\Big(\int_0^{s}\phi\,\mathrm{d}B\mathbf{1}_A\Big).$$

这意味着 $\mathrm{E}\big[\int_0^{t'}\phi_u\mathrm{d}B_u\big|\mathcal{F}_s\big]=\int_0^s\phi_u\mathrm{d}B_u$.

为证明 (3), 先证明如下的等式, 对于 $s<t'$

$$\mathrm{E}\Big[\int_0^s\phi\,\mathrm{d}B\int_s^{t'}\phi\,\mathrm{d}B\Big|\mathcal{F}_s\Big]=0, \tag{5.4.3}$$

$$\mathrm{E}\Big[\Big(\int_s^{t'}\phi_u\,\mathrm{d}B_u\Big)^2-\int_s^{t'}\phi_u^2\,\mathrm{d}u\Big|\mathcal{F}_s\Big]=0. \tag{5.4.4}$$

事实上

$$\begin{aligned}
\text{(5.4.3)式的左端}\ &=\ \Big(\int_0^s\phi\,\mathrm{d}B\Big)\mathrm{E}\Big[\int_s^{t'}\phi\,\mathrm{d}B\Big|\mathcal{F}_s\Big]\\
&=\ \Big(\int_0^s\phi\,\mathrm{d}B\Big)\mathrm{E}\Big[\int_0^{t'}\phi\,\mathrm{d}B-\int_0^{s}\phi\,\mathrm{d}B\Big|\mathcal{F}_s\Big]\\
&=\ 0.
\end{aligned}$$

对于 (5.4.4), 当 ϕ 是可料阶梯过程时

$$
\begin{aligned}
& \mathrm{E}\Big[\Big(\int_s^{t'} \phi_u \,\mathrm{d}B_u\Big)^2\Big|\mathcal{F}_s\Big] \\
= \;& \mathrm{E}\Big[\Big(\sum_{k=p}^{q-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\Big)^2\Big|\mathcal{F}_{t'_p}\Big] \\
= \;& \sum_{k=p}^{q-1} \mathrm{E}\Big[\big(\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\big)^2\Big|\mathcal{F}_{t'_p}\Big] + 2\sum_{p\le k<j\le q-1} \mathrm{E}\Big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})(B_{t'_{j+1}} - B_{t'_j})\Big|\mathcal{F}_{t'_p}\Big] \\
= \;& \sum_{k=p}^{q-1} \mathrm{E}\Big[\mathrm{E}\big[\big(\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\big)^2\big|\mathcal{F}_{t'_k}\big]\Big|\mathcal{F}_{t'_p}\Big] \\
& +2\sum_{p\le k<j\le q-1} \mathrm{E}\Big[\mathrm{E}\big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})(B_{t'_{j+1}} - B_{t'_j})\big|\mathcal{F}_{t'_j}\big]\Big|\mathcal{F}_{t'_p}\Big] \\
= \;& \sum_{k=p}^{q-1} \mathrm{E}\Big[\phi_{t'_k}^2(t'_{k+1} - t'_k)\Big|\mathcal{F}_{t'_p}\Big] + 2\sum_{p\le k<j\le q-1} \mathrm{E}\Big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})\mathrm{E}\big[(B_{t'_{j+1}} - B_{t'_j})\big|\mathcal{F}_{t'_j}\big]\Big|\mathcal{F}_{t'_p}\Big] \\
= \;& \mathrm{E}\Big[\sum_{k=p}^{q-1} \phi_{t'_k}^2(t'_{k+1} - t'_k)\Big|\mathcal{F}_{t'_p}\Big] \\
= \;& \mathrm{E}\Big[\int_s^{t'} \phi_u^2 \,\mathrm{d}u\Big|\mathcal{F}_s\Big].
\end{aligned}
$$

类似于性质 (2) 的证明中的逼近方式, 当 $\phi \in \mathscr{L}_2^T$ 时也可以得到 (5.4.4) 成立.

注意到 $\int_0^s \phi_u^2 \,\mathrm{d}u$ 关于 $\mathcal{F}_s$ 可测, 利用 (5.4.3)(5.4.4)

$$
\begin{aligned}
& \mathrm{E}\Big[\Big(\int_0^{t'} \phi_u \,\mathrm{d}B_u\Big)^2 - \int_0^{t'} \phi_u^2 \,\mathrm{d}u\Big|\mathcal{F}_s\Big] \\
= \;& \mathrm{E}\Big[\Big(\int_0^{s} \phi_u \,\mathrm{d}B_u + \int_s^{t'} \phi_u \,\mathrm{d}B_u\Big)^2 - \int_0^{s} \phi_u^2 \,\mathrm{d}u - \int_s^{t'} \phi_u^2 \,\mathrm{d}u\Big|\mathcal{F}_s\Big] \\
= \;& \Big(\int_0^{s} \phi_u \,\mathrm{d}B_u\Big)^2 - \int_0^{s} \phi_u^2 \,\mathrm{d}u + 2\mathrm{E}\Big[\int_0^{s} \phi \,\mathrm{d}B \int_s^{t'} \phi \,\mathrm{d}B\Big|\mathcal{F}_s\Big] \\
& +\mathrm{E}\Big[\Big(\int_s^{t'} \phi_u \,\mathrm{d}B_u\Big)^2 - \int_s^{t'} \phi_u^2 \,\mathrm{d}u\Big|\mathcal{F}_s\Big] \\
= \;& \Big(\int_0^{s} \phi_u \,\mathrm{d}B_u\Big)^2 - \int_0^{s} \phi_u^2 \,\mathrm{d}u.
\end{aligned}
$$

□

在这一章引言的 (5.1.7) 式中

$$X_t - X_0 \approx \sum_k \sigma(t_k, X_{t_k}(\omega))(B_{t_{k+1}}(\omega) - B_{t_k}(\omega)) + \sum_k b(t_k, X_{t_k}(\omega))(t_{k+1} - t_k),$$

除了第一项关于布朗运动的 Itô 随机积分, 我们还遇到了第二项关于随机过程轨道的 Riemann 积分.

令 $\mathscr{L}_T^1 = \big\{\Phi : (\mathcal{F}_t)_{t\in[0,T]}\text{适应的随机过程, 且} \,\mathrm{E}\int_0^T |\Phi_t(\omega)|\mathrm{d}t < \infty\big\}$ ①. 在 $\mathscr{L}_T^1$ 中可以定义如下的距离

$$d(\Psi, \Theta) \overset{\text{def}}{=} \mathrm{E}\int_0^T |\Psi_t - \Theta_t|\mathrm{d}t.$$

对于 $\mathscr{L}_T^1$ 中的元 $\Phi = \{\Phi_t,\, t \ge 0\}$, 可以证明

$$\mathrm{P}\Big(\omega : \int_0^T \Phi_t(\omega)\mathrm{d}t \text{ 存在}\Big) = 1.$$

①严格说, $(\phi_t)_{0\le t\le T}$ 应该是循序可测 (progressively measurable) 过程.

因此对于几乎处处的样本轨道 ω, 可以按照普通的 Riemann 积分定义 $\int_0^T \Phi_t(\omega)\mathrm{d}t$. 并且由 Lebesgue 积分的绝对连续性知

$$\mathrm{P}\Big(\omega: \int_0^t \Phi_s(\omega)\mathrm{d}s \text{ 关于}t\text{是绝对连续函数}\Big) = 1.$$

♣♣♣20230513

5.5 Itô 公式

5.5.1 定义特殊类型的 Itô 过程及其 Itô 公式

定义 5.5.1. (特殊类型的) Itô 过程

设

$$\xi_t = x + \int_0^t \phi_s \,\mathrm{d}B_s + \int_0^t \psi_s \,\mathrm{d}s, \tag{5.5.1}$$

其中随机过程 ϕ, ψ 分别为 $\mathscr{L}_T^2$ 和 $\mathscr{L}_T^1$ 中的元. 这时称 $(\xi_t)_{t\geq 0}$ 是初值为 x, 系数是 ϕ_t, ψ_t 的特殊类型的 Itô 过程.

可以将 (5.5.1) 记为 Itô (形式) 微分

$$\mathrm{d}\xi_t = \phi_t \mathrm{d}B_t + \psi_t \mathrm{d}t, \quad \xi_0 = x.$$

例如, 当 $B_0 = x$ 时

$$B_t^2 = x^2 + 2\int_0^t B_s \mathrm{d}B_s + t, \text{ 或者表示为 } \mathrm{d}B_t^2 = 2B_t\mathrm{d}B_t + \mathrm{d}t, \quad B_0^2 = x^2.$$

或者设 $f(x) = x^2$, 则

$$f(B_t) - f(B_0) = 2\int_0^t B_s \mathrm{d}B_s + t, \text{ 或者表示为 } \mathrm{d}f(B_t) = 2B_t\mathrm{d}B_t + \mathrm{d}t, \quad f(B_0) = x^2.$$

从上一节的例子我们可以知道

1. 从定义直接计算一个 Itô 积分太繁琐.
2. 与微积分中的链式法则 (牛顿-莱布尼茨公式) 不一样, B_t^2 不再是一个 Itô 积分的形式, 而成为 “$\mathrm{d}B_s$”+“$\mathrm{d}s$” 积分组合, 即成为一个 Itô 过程.
   (a) 牛顿-莱布尼茨公式、链式法则、一阶微分的不变性
   $$F(b) - F(a) = \int_a^b F'(x)\mathrm{d}x,$$
   $$g\big(f(b)\big) - g\big(f(a)\big) = \int_{f(a)}^{f(b)} g'(z)\mathrm{d}z = \int_a^b g'\big(f(x)\big)f'(x)\mathrm{d}x = \int_a^b g'\big(f(x)\big)\mathrm{d}f(x).$$
   Riemann 积分之所以有广泛的应用, 原因之一是它有很好的可计算性, 即牛顿-莱布尼茨公式 (或链式法则), 由此联系了积分与微分.
   (b) Itô 随机积分是否有链式法则、一阶微分的不变性?
   $$g(B_t) - g(B_0) = \int_0^t g'(B_u)\mathrm{d}B_u,$$
   答案是否定的, 例如:
   $$B_t^2 - B_0^2 = 2\int_0^t B_s dB_s + t, \quad g(x) = x^2,$$
   $$g(B_t) - g(B_0) = 2\int_0^t g'(B_s)\mathrm{d}B_s + t.$$

Itô 随机积分的牛顿-莱布尼茨公式应该是什么? 首先我们回忆一下 Riemann 积分中牛顿-莱布尼茨公式是如何证明? 设 $f(x) \in C[a,b]$, 又设 $F \in C[a,b]$, 并在 (a,b) 上 $F(x)$ 是 $f(x)$ 的原函数, $F'(x) = f(x)$, 则

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

证明: $a = x_0 < x_1 < \cdots < x_n = b$ 为 $[a,b]$ 的一个划分, 利用微分中值定理

$$F(b) - F(a) = \sum_{k=1}^{n}[F(x_k) - F(x_{k-1})] = \sum_{k=1}^{n} f(\eta_k)\Delta x_k,$$

其中 $\Delta x_k = x_k - x_{k-1}$, $\eta_k \in [x_{k-1}, x_k]$.

由于 $f(x)$ 在 $[a,b]$ 中一致连续, 所以任给 $\varepsilon > 0$, 存在 $\delta > 0$, 只需 $\xi, \eta \in [a,b]$, $|\xi - \eta| < \delta$, 就有

$$|f(\xi) - f(\eta)| < \frac{\varepsilon}{b-a}.$$

于是, 当 $\lambda = \max\limits_{1\le k\le n} \Delta x_k < \delta$ 时, $\forall \xi_k \in [x_{k-1}, x_k]$ 有

$$\Big|\sum_{k=1}^{n} f(\xi_k)\Delta x_k - (F(b) - F(a))\Big| = \Big|\sum_{k=1}^{n}[f(\xi_k) - f(\eta_k)]\Delta x_k\Big| \le \frac{\varepsilon}{b-a}\sum_{k=1}^{n}|\Delta x_k|$$

因此, $f(x)$ 在 $[a,b]$ 上可积, 且 $\int_a^b f(x)\mathrm{d}x = F(b) - F(a)$. □

利用这个方法来计算一下 $F(B_t) - F(B_0) =$

$$\begin{aligned}
F(B_t) - F(B_0) &= \sum_{k=0}^{n-1}\left[F(B_{t_{k+1}}) - F(B_{t_k})\right] \\
&= \sum_{k=0}^{n-1}\left[F'(\xi_{t_k})(B_{t_{k+1}} - B_{t_k})\right] \ (\text{其中}\xi_{t_k} \in [B_{t_{k+1}}, B_{t_k}]\text{或}[B_{t_k}, B_{t_{k+1}}]) \\
&= \sum_{k=0}^{n-1}\left[F'(B_{t_k})\Delta B_{t_k} - \big(F'(B_{t_k}) - F'(\xi_{t_k})\big)\Delta B_{t_k}\right] \\
&= \sum_{k=0}^{n-1}\left[F'(B_{t_k})\Delta B_{t_k}\right] - \sum_{k=0}^{n-1}\left[F'(B_{t_k}) - F'(\xi_{t_k})\right]\Delta B_{t_k} \\
&= I + II.
\end{aligned}$$

由 Itô 随机积分的定义

$$I \to \int_0^t F'(B_s)\mathrm{d}B_s.$$

类似 Riemann 积分中的证明过程, 我们看

$$II = \Big|\sum_{k=0}^{n-1}\left[F'(B_{t_k}) - F'(\xi_{t_k})\right]\Delta B_{t_k}\Big| \to .$$

利用 $F'(B_t)$ 在有界闭区间的一致连续性, 即使我们有 $|F'(B_{t_k}) - F'(\xi_{t_k})| < \frac{\varepsilon}{t}$ 这样的估计, 也仅仅能得到

$$\Big|\sum_{k=0}^{n-1}\left[F'(B_{t_k}) - F'(\xi_{t_k})\right]\Delta B_{t_k}\Big| \le \frac{\varepsilon}{t}\sum_{k=0}^{n-1}\big|\Delta B_{t_k}\big|.$$

因此仅仅以类似 Riemann 积分的角度来考虑随机积分的牛顿-莱布尼茨公式是行不通的, 关键在于布朗运动的轨道不是有界变差的, $\sum\limits_{k=0}^{n-1} |\Delta B_{t_k}|$ 发散.

在 Riemann 积分中我们对 F 利用泰勒公式展开到一阶就行了. 虽然 $(B_t)_{t\ge 0}$ 的轨道不是有界变差

的, 但它是均方变差收敛的. 因此对于 Itô 积分, 我们需对它展开更高阶才行.

$$\begin{aligned}F(B_t)-F(B_0) &= \sum_{k=0}^{n-1}\big[F(B_{t_{k+1}})-F(B_{t_k})\big]\\ &= \sum_{k=0}^{n-1}\big[F'(B_{t_k})\Delta B_{t_k}+\frac{1}{2}F''(\xi_{t_k})(\Delta B_{t_k})^2\big]\\ &= \sum_{k=0}^{n-1}F'(B_{t_k})\Delta B_{t_k}+\sum_{k=0}^{n-1}\frac{1}{2}F''(\xi_{t_k})(\Delta B_{t_k})^2\\ &= I+II.\end{aligned}$$

此时的关键问题是 II 是否收敛?

引理 5.5.1. 设 $g(x)$ 连续有界, 则

$$\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2 \xrightarrow{L^2} \int_0^t g(B_s)\mathrm{d}s.$$

证明:

$$\begin{aligned}&\mathrm{E}\Big(\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2-\int_0^t g(B_s)\mathrm{d}s\Big)^2\\ \le\ & 2\mathrm{E}\Big(\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2-\sum_{k=0}^{n-1}g(B_{t_k})\Delta t_k\Big)^2+2\mathrm{E}\Big(\sum_{k=0}^{n-1}g(B_{t_k})\Delta t_k-\int_0^t g(B_s)\mathrm{d}s\Big)^2\\ =\ & 2II_1+2II_2.\end{aligned}$$

利用 g 的有界性, 类似 Lévy 振动性质的证明 (引理5.3.1) 得到 $II_1\to 0$.

注意到 $\Big(\sum_{k=0}^{n-1}g(B_{t_k})\mathbf{1}_{(t_k,t_{k+1}]}(t)\Big)_{t\ge 0}$ 是 $\big(g(B_t)\big)_{t\ge 0}$ 在 $\mathscr{L}_T^2$ 的可料阶梯过程近似列, 由例5.4.1, $II_2\to 0$.

□

引理 5.5.2. 设 $g(x)$ 有界一致连续, $\{t_k^{(n)}\}$ 为 $[0,t]$ 的 2^n 等分点, $\forall\xi_{t_k^{(n)}}\in[B_{t_k^{(n)}},B_{t_{k+1}^{(n)}}]$ 或 $[B_{t_{k+1}^{(n)}},B_{t_k^{(n)}}]$ [①],

$$\mathrm{P}\Big(\lim_{n\to\infty}\sum_{k=0}^{2^n-1}\big(g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big)(\Delta B_{t_k^{(n)}})^2=0\Big)=1.$$

证明: 在 $[0,t]$ 内, 对于几乎处处的 ω, $g(B_t(\omega))$ 一致连续, 因此, 可以证明 $\forall\varepsilon>0$, $\exists\delta(\omega)>0$, 当 $\frac{t}{2^n}<\delta(\omega)$ 时, 对于 $\forall t_k^{(n)}$

$$\big|g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big|<\varepsilon,$$

$$\Big|\sum_{k=0}^{n-1}\big(g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big)(\Delta B_{t_k^{(n)}})^2\Big|\le\varepsilon\sum_{k=0}^{n-1}(\Delta B_{t_k^{(n)}})^2.$$

由引理5.3.2,

$$\sum_{k=0}^{n-1}(\Delta B_{t_k^{(n)}})^2\to t,\ \text{以概率 1 成立.}$$

□

[①] $g(\xi_{t_k^{(n)}})$ 不一定是随机变量. 可以假设概率空间 $(\Omega,\mathcal{F},P)$ 是一个完备的概率空间, 就不会有证明逻辑的问题了. 一般在随机分析中我们都要求概率空间是完备的. 此处证明还待完善

定理 5.5.1. (Itô 公式 I) 若 F 为二次连续可微, 则

$$F(B_t)-F(B_0)=\int_0^t F'(B_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t F''(B_s)\mathrm{d}s,$$

其形式地记为

$$\mathrm{d}F(B_t)=F'(B_t)\mathrm{d}B_t+\frac{1}{2}F''(B_t)\mathrm{d}t.$$

证明: 此处只对 F', F'' 有界且一致连续证明. 取 $0=t_0^{(n)}<t_1^{(n)}<\cdots<t_{2^n}^{(n)}=t$ 为 $[0,t]$ 中的 2^n 等分点

$$\begin{aligned}
& F(B_t)-F(B_0)\\
= & \sum_{k=0}^{2^n-1}\big[F(B_{t_k})-F(B_{t_{k-1}})\big]\\
= & \sum_{k=0}^{2^n-1}\big[F'(B_{t_k})\Delta B_{t_k}\big]+\frac{1}{2}\sum_{k=0}^{2^n-1}F''(B_{t_k})(\Delta B_{t_k})^2+\frac{1}{2}\sum_{k=0}^{2^n-1}\big(F''(\xi_{t_k})-F''(B_{t_k})\big)(\Delta B_{t_k})^2,
\end{aligned}$$

则存在一个子列 n', 使得 $n'\to\infty$ 上式右端以概率 1 收敛到

$$\int_0^t F'(B_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t F''(B_s)\mathrm{d}s.$$

□

定理 5.5.2. (Itô 公式 II) 设 $F(t,x)$ 关于 t 一次连续可微, 关于 x 二次连续可微, 则

$$\begin{aligned}
& F(t,B_t)-F(0,B_0)\\
= & \int_0^t\frac{\partial F(s,B_s)}{\partial s}\mathrm{d}s+\int_0^t\frac{\partial F(s,B_s)}{\partial x}\mathrm{d}B_s+\frac{1}{2}\int_0^t\frac{\partial^2F(s,B_s)}{\partial x^2}\mathrm{d}s\\
= & \int_0^t F_t'(s,B_s)\mathrm{d}s+\int_0^t F_x'(s,B_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t F_{xx}''(s,B_s)\mathrm{d}s.
\end{aligned}$$

其形式地记为

$$\mathrm{d}F(t,B_t)=F_t'(t,B_t)\mathrm{d}t+\frac{1}{2}F_{xx}''(t,B_t)\mathrm{d}t+F_x'(t,B_t)\mathrm{d}B_t.$$

例 5.5.1. 设 $(B_t)_{t\geq0}$ 为标准布朗运动, 证明 $\mathrm{e}^{cB_t-\frac{c^2}{2}t}$ 为 $(\mathcal{F}_t)_{t\geq0}$ 鞅.

解:

(1) 首先验证对于 $\forall T>0$, $\mathrm{e}^{cB_t-\frac{c^2}{2}t}\in\mathscr{L}_T^2$.

$$\mathrm{E}\big(\mathrm{e}^{cB_t-\frac{c^2}{2}t}\big)^2=\mathrm{e}^{-c^2t}\mathrm{E}\,\mathrm{e}^{2cB_t}=\mathrm{e}^{-c^2t}\mathrm{e}^{2c^2t}<\infty.$$

(2) 设 $F(t,x)=\mathrm{e}^{-\frac{c^2}{2}t+cx}$

$$\begin{aligned}
F(t,B_t)-F(0,0) &= \int_0^t F_s'(s,B_s)\mathrm{d}s+\int_0^t F_x'(s,B_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t F_{xx}''(s,B_s)\mathrm{d}s\\
&= \int_0^t\big(-\frac{c^2}{2}\big)\mathrm{e}^{-\frac{c^2}{2}t}\mathrm{e}^{cB_s}\mathrm{d}s+c\int_0^t\mathrm{e}^{-\frac{c^2}{2}t}\mathrm{e}^{cB_s}\mathrm{d}B_s+\frac{c^2}{2}\int_0^t\mathrm{e}^{-\frac{c^2}{2}}\mathrm{e}^{cB_s}\mathrm{d}s\\
&= c\int_0^t F(s,B_s)\mathrm{d}B_s
\end{aligned}$$

即

$$\mathrm{e}^{cB_t-\frac{c^2}{2}t}=1+c\int_0^t\mathrm{e}^{cB_s-\frac{c^2}{2}s}\mathrm{d}B_s.$$

因此, 由性质5.4.3的 (2) 知道 $(\mathrm{e}^{cB_t-\frac{c^2}{2}t})_{t\geq0}$ 为 $(\mathcal{F}_t)_{t\geq0}$ 鞅. □

对于一般的 Itô 过程 $X_t=x_0+\int_0^t\phi_s\mathrm{d}B_s+\int_0^t\Psi_s\mathrm{d}s$, $F(t,X_t)-F(0,X_0)$ 的 Itô 公式是什么?

根据前面的分析和我们积累起来的经验可以知道包含 t 和 Riemann 积分项 $\int_0^t \Psi_s \mathrm{d}s$ 的运算规律应该还是遵从微积分中的牛顿-莱布尼茨公式. 因此, 问题的关键就在于如何处理包含 $\int_0^t \phi_s \mathrm{d}B_s$ 的项. 我们将问题简化成若 $X_t = x_0 + \int_0^t \phi_s \mathrm{d}B_s + \int_0^t \Psi_s \mathrm{d}s,\ F(X_t) - F(x_0)$ 应该等于什么?

虽然 $\mathrm{d}X_t = \phi_t \mathrm{d}B_t + \Psi_t \mathrm{d}t$ 只是 Itô 过程 $(X_t)_{t\geq 0}$ 的 "形式的微分", 并不具有严格的数学定义, 但从这一章引言的一些介绍, 我们可以直观的把它理解成当 $\Delta t > 0$ 很小的时候, $\mathrm{d}X_t = \phi_t \mathrm{d}B_t + \Psi_t \mathrm{d}t$ 近似地是

$$\Delta X_t = X(t+\Delta t) - X(t) \approx \phi_t (B_{t+\Delta t} - B_t) + \Psi_t \Delta t = \phi_t \Delta B_t + \Psi_t \Delta t.$$

当划分 $0 = t_0 < t_1 < \cdots < t_n = t$ 充分细时, 我们有如下近似的看法

$$\begin{aligned}
\int_0^t g(X_s)\phi_s \mathrm{d}B_s + \int_0^t g(X_s)\Psi_s \mathrm{d}s &= \int_0^t g(X_s)\mathrm{d}X_s \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})\Delta X_{t_i} \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})(\phi_{t_i}\Delta B_{t_i} + \Psi_{t_i}\Delta t_i).
\end{aligned}$$

对于任意的 $\varepsilon > 0$, 当划分充分细时, 使得 $\max_i\{|\Delta B_i|, \Delta t_i\} < \varepsilon$

$$|\Delta B_{t_i}(\omega)| < \varepsilon,\ \ \sum_{i=0}^{n-1}\left|\Delta B_{t_i}\Delta t_i\right| < \varepsilon\sum_{i=0}^{n-1}\Delta t_i = \varepsilon t,\ \ \sum_{i=0}^{n-1}(\Delta t_i)^2 \leq \max_i\{\Delta t_i\}\sum_{i=1}^{n-1}\Delta t_i < \varepsilon t.$$

对于有界一致连续函数 $g(x)$, ϕ_t 和 Ψ_t 有界, 我们可以近似地得到

$$\begin{aligned}
\int_0^t g(X_s)(\mathrm{d}X_s)^2 &\approx \sum_{i=0}^{n-1} g(X_{t_i})(\Delta X_{t_i})^2 \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})(\phi_{t_i}\Delta B_{t_i} + \Psi_{t_i}\Delta t_i)^2 \\
&= \sum_{i=0}^{n-1} g(X_{t_i})\left[\phi_{t_i}^2(\Delta B_{t_i})^2 + 2\phi_{t_i}\Psi_{t_i}\Delta B_{t_i}\Delta t_i + \Psi_{t_i}^2(\Delta t_i)^2\right] \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})\phi_{t_i}^2(\Delta B_{t_i})^2 + O(\varepsilon) \\
&\approx \int_0^t g(X_s)\phi_s^2 \mathrm{d}s \quad \text{类似于引理5.5.1的证明并忽略掉无穷小量}
\end{aligned}$$

$$\begin{aligned}
& F(X_t) - F(x_0) \\
=\ & \sum_{i=0}^{n-1}\left(F(X_{t_{i+1}}) - F(X_{t_i})\right) \\
\overset{\text{泰勒展开}}{=}\ & \sum_{i=0}^{n-1}\Big(F'(X_{t_i})\Delta X_{t_i} + \frac{1}{2}F''(X_{t_i})(\Delta X_{t_i})^2 + \frac{1}{2}\big(F''(\xi_{t_i}) - F''(X_{t_i})\big)(\Delta X_{t_i})^2\Big) \\
\approx\ & \sum_{i=0}^{n-1}\Big(F'(X_{t_i})\Delta X_{t_i} + \frac{1}{2}F''(X_{t_i})(\Delta X_{t_i})^2\Big) \quad \text{类似于引理5.5.2的证明} \\
\approx\ & \int_0^t F'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t F''(X_s)(\mathrm{d}X_s)^2 \\
=\ & \int_0^t F'(X_s)\phi_s \mathrm{d}B_s + \int_0^t F'(X_s)\phi_s \mathrm{d}s + \frac{1}{2}\int_0^t F''(X_s)\phi_s^2 \mathrm{d}s.
\end{aligned}$$

将上述步骤严格化, 并做一些必要的技术上的处理, 就有下面的定理.

定理 5.5.3. (Itô 公式 Ⅲ) 设 $X_t = x_0 + \int_0^t \phi_s \mathrm{d}B_s + \int_0^t \Psi_s \mathrm{d}s$, $\mathrm{d}X_t = \phi_t \mathrm{d}B_t + \Psi_s \mathrm{d}s$, $F(t,x)$ 关于 t 一次

连续可微, 关于 x 二次连续可微, 则

$$F(t,X_t)-F(0,x_0)=\int_0^t\frac{\partial F}{\partial s}(s,X_s)\mathrm{d}s+\int_0^t\frac{\partial F}{\partial x}(s,X_s)\mathrm{d}X_s+\frac{1}{2}\int_0^t\frac{\partial^2 F}{\partial x^2}(s,X_s)(\mathrm{d}X_s)^2.$$

写成"形式微分"

$$\mathrm{d}F(t,X_t)=\frac{\partial F}{\partial t}(t,X_t)\mathrm{d}t+\frac{\partial F}{\partial x}(t,X_t)\mathrm{d}X_t+\frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X_t)(\mathrm{d}X_t)^2,$$

其中 $\mathrm{d}t$ 和 $\mathrm{d}B_t$ 的平方与乘积由乘法表

乘积	$\mathrm{d}t$	$\mathrm{d}B_t$
$\mathrm{d}t$	0	0
$\mathrm{d}B_t$	0	$\mathrm{d}t$

给出, 即

$$\mathrm{d}F(t,X_t)=\Big[\frac{\partial F}{\partial t}(t,X_t)+\Psi_t\frac{\partial F}{\partial x}(t,X_t)+\frac{1}{2}\phi_t^2\frac{\partial^2 F}{\partial x^2}(t,X_t)\Big]\mathrm{d}t+\phi_t\frac{\partial F}{\partial x}(t,X_t)\mathrm{d}B_t$$

写成真正的积分形式

$$F(t,X_t)-F(0,x_0)=\int_0^t\Big[\frac{\partial F}{\partial s}(s,X_s)+\Psi_t\frac{\partial F}{\partial x}(s,X_s)+\frac{1}{2}\phi_s^2\frac{\partial^2 F}{\partial x^2}(s,X_s)\Big]\mathrm{d}s+\int_0^t\phi_s\frac{\partial F}{\partial x}(s,X_s)\mathrm{d}B_s.$$

其中

$$(\mathrm{d}X_t)^2=\big(\phi_t\mathrm{d}B_t+\Psi_t\mathrm{d}t\big)^2=(\phi_t\mathrm{d}B_t)^2+2\phi_t\Psi_t\mathrm{d}B_t\mathrm{d}t+\Psi_t^2(\mathrm{d}t)^2=\phi_t^2\mathrm{d}t.$$

例 5.5.2. 若 $(\phi_t)_{t\geq 0}$ 为关于 $(\mathcal{F}_t)_{t\geq 0}$ 适应的有界随机过程, 证明 $\xi_t=\exp\{\int_0^t\phi_s\mathrm{d}B_s-\frac{1}{2}\int_0^t\phi_s^2\mathrm{d}s\}$ 是关于 $(\mathcal{F}_t)_{t\geq 0}$ 的鞅.

证明: 利用求解例5.5.1的第一步以及可料阶梯过程的逼近程序, 可以证明 $\{\xi_t\}\in\mathscr{L}_T^2$. 设 $F(x)=\mathrm{e}^x$,

$$X_t=\int_0^t\phi_s\mathrm{d}B_s-\frac{1}{2}\int_0^t\phi_s^2\mathrm{d}s,\quad 即\ \mathrm{d}X_t=\phi_t\mathrm{d}B_t-\frac{1}{2}\phi_t^2\mathrm{d}t.$$

利用 Itô 公式

$$\begin{aligned}F(X_t)-F(0)&=\int_0^t F'(X_s)\mathrm{d}X_s+\frac{1}{2}\int_0^t F''(X_s)(\mathrm{d}X_s)^2\\&=\int_0^t F(X_s)\phi_s\mathrm{d}B_s-\frac{1}{2}\int_0^t F(X_s)\phi_s^2\mathrm{d}t+\frac{1}{2}\int_0^t F(X_s)\phi_s^2\mathrm{d}t\\&=\int_0^t F(X_s)\phi_s\mathrm{d}B_s,\end{aligned}$$

所以由性质5.4.3的 (2) 知道 $(\exp\{\int_0^t\phi_s\mathrm{d}B_s-\frac{1}{2}\int_0^t\phi_s^2\mathrm{d}s)_{t\geq 0}$ 为 $(\mathcal{F}_t)_{t\geq 0}$ 鞅. □

5.5.2 一般的 Itô 过程

前面我们对 $\mathscr{L}_T^2$ 中的过程定义了 Itô 随机积分并建立了 Itô 公式, 但问题是对被积的过程限制太多. 在 Riemann 积分, 例如: 若在 $[a,b]$, 当 $x\to b$ 时, $f(x)\uparrow\infty$, 存在 $b_n\uparrow b$, 使得 $f(x)$ 在 $[a,b_n]$ 内有界且 Riemann 可积, 我们如何来定义 $\int_a^b f(x)\mathrm{d}x$? 事实上定义 $\int_a^b f(x)\mathrm{d}x=\lim\limits_{b_n\uparrow b}\int_a^{b_n}f(x)\mathrm{d}x$.

类似的想法

$$\mathscr{L}^{2,\mathrm{loc}}\overset{\mathrm{def}}{=}\Big\{(\mathcal{F}_t)_{t\geq 0}适应的随机过程(\phi_t)_{t\geq 0}, 且对于任意的t>0,\mathrm{P}\Big(\omega:\int_0^T(\phi_t)^2(\omega)\mathrm{d}t<\infty\Big)=1\Big\}.$$

可以证明

(1) 对于任意 $\phi\in\mathscr{L}^{2,\mathrm{loc}}$ 存在 $(\mathcal{F}_t)_{t\geq 0}$ 停时序列 $\tau_1\leq\cdots\leq\tau_n\uparrow\infty$, 使得对于固定的 n, 随机过程 $\phi_t^{(n)}\overset{\mathrm{def}}{=}\phi_{t\wedge\tau_n}\in\mathscr{L}^2=\big\{(\mathcal{F}_t)_{t\geq 0}适应的随机过程(\phi_t)_{t\geq 0}, 且对于任意的t>0,\mathrm{E}\big(\int_0^T(\phi_t)^2(\omega)\mathrm{d}t\big)<\infty\big\}$.

(2) 由此, 当 $t \le \tau_n$ 时可以定义 $\int_0^t \phi_s \mathrm{d}B_s$:

$$\Big(\int_0^t \phi_s \mathrm{d}B_s\Big)(\omega) \overset{\text{def}}{=} \Big(\int_0^t \phi_s^{(n)} \mathrm{d}B_s\Big)(\omega).$$

进一步,

$$\int_0^t \phi_s \mathrm{d}B_s \overset{\text{def}}{=} \lim_{n\to\infty} \int_0^t \phi_s^{(n)} \mathrm{d}B_s.$$

(3) $(\xi_t \overset{\text{def}}{=} \int_0^t \phi_s \mathrm{d}B_s)_{t\ge 0}$ 是 $(\mathcal{F}_t)$ 局部鞅, 即存在 $(\mathcal{F}_t)_{t\ge 0}$ 停时序列 $\tau_1 \le \cdots \le \tau_n \uparrow \infty$, 使得对于固定的 n, $(\zeta_t^{(n)} \overset{\text{def}}{=} \xi_{t\wedge\tau_n})_{t\ge 0}$ 是 $(\mathcal{F}_t)_{t\ge 0}$ 鞅. 但局部鞅不一定是鞅.

(4) 设 $\mathscr{L}^{1,\text{loc}} \overset{\text{def}}{=} \big\{(\mathcal{F}_t)_{t\ge 0}$适应的随机过程, 且对于任意的 $t>0, (\Psi_t)_{t\ge 0}, \mathrm{P}\big(\omega : \int_0^t |\Psi_t| \mathrm{d}t < +\infty\big) = 1\big\}$, 则对于任意 $\Psi \in \mathscr{L}^{1,\text{loc}}$ 积分 $\int_0^t \Psi_s(\omega)\mathrm{d}s$ 几乎处处有定义.

定义 5.5.2. 一般的 Itô 过程

设 $\zeta_t = x + \int_0^t \phi_s \mathrm{d}B_s + \int_0^t \Psi_s \mathrm{d}s$, 其中 ϕ_s 和 Ψ_s 分别是 $\mathscr{L}^{2,\text{loc}}$ 和 $\mathscr{L}^{1,\text{loc}}$ 中的元, 则 ξ_t 称为初值为 x 系数为 ϕ_t, Ψ_t 的一般的 Itô 过程, 简称 Itô 过程.

对 Itô 过程 $\{\zeta_t,\ t\ge 0\}$ 的 Itô 积分实质上为

$$\int_0^t \Theta_t \mathrm{d}\zeta_t = \int_0^t \Theta_t\phi_t \mathrm{d}B_t + \int_0^t \Theta_t \Psi_t \mathrm{d}t.$$

定理 5.5.4. Itô 公式 Ⅱ, Ⅲ 对一般的 Itô 过程也成立.

5.5.3 多维布朗运动的 Itô 随机积分和 Itô 公式

设 $\mathbf{B}_t = \begin{pmatrix} B_t^{(1)} \\ \vdots \\ B_t^{(m)} \end{pmatrix}$ 为 $(\mathcal{F}_t)_{t\ge 0}$ 关于事件域流 $(\mathcal{F}_t)_{t\ge 0}$ 的 m-维布朗运动.

$$\mathscr{L}_T^2 = \Big\{(\phi_t)_{t\ge 0}\text{为}(\mathcal{F}_t)_{t\ge 0}\text{适应的随机过程, 且}\,\mathrm{E}\int_0^T \phi_s^2\,\mathrm{d}s < \infty\Big\}.$$

$$\mathscr{L}_T^1 = \Big\{(\phi_t)_{t\ge 0}\text{为}(\mathcal{F}_t)_{t\ge 0}\text{适应的随机过程, 且}\,\mathrm{E}\int_0^T\ |\phi_s|\mathrm{d}s < \infty\big) = 1\Big\}.$$

$$\mathscr{L}^{2,\text{loc}} = \Big\{(\phi_t)_{t\ge 0}\text{为}(\mathcal{F}_t)_{t\ge 0}\text{适应的随机过程, 且对于任意的}\,t>0,\ \mathrm{P}\Big(\omega : \int_0^t \phi_s^2 \mathrm{d}s < \infty\Big) = 1\Big\}.$$

$$\mathscr{L}^{1,\text{loc}} = \Big\{(\phi_t)_{t\ge 0}\text{为}(\mathcal{F}_t)_{t\ge 0}\text{适应的随机过程, 且对于任意的}\,t>0,\ \mathrm{P}\Big(\omega : \int_0^t |\phi_s| \mathrm{d}s < +\infty\Big) = 1\Big\}.$$

类似地可以定义

$$\int_0^t \phi_s \mathrm{d}B_s^{(i)}\ i = 1, \cdots, m.$$

定义 5.5.3. (多维布朗运动的 Itô 过程)

设

$$\xi_t = x + \int_0^t \phi_s^{\mathrm{t}} \mathrm{d}\mathbf{B}_s + \int_0^t \Psi_s \mathrm{d}s,$$

其中 $\phi_t = \begin{pmatrix} \phi_t^{(1)} \\ \vdots \\ \phi_t^{(m)} \end{pmatrix}$, $\xi_t = x + \sum_{k=1}^m \int_0^t \phi_s^{(k)} \mathrm{d}B_s^{(k)} + \int_0^t \Psi_s \mathrm{d}s$, 而 $(\phi_t^{(i)})_{t\ge 0}$, $(\Psi_t)_{t\ge 0}$ 分别是 $\mathscr{L}^{2,\text{loc}}$, $\mathscr{L}^{1,\text{loc}}$ 中的元, 那么称 $(\xi_t)_{t\ge 0}$ 为初值为 x 的多维布朗运动的 Itô 过程.

引理 5.5.3. 设 $(B_t^{(1)})$, $(B_t^{(2)})$ 是两个独立的布朗运动, 对于 $[0,t]$ 中的一个划分 $0=t_0<\cdots<t_n=t$,
$$T_n=\sum_{k=0}^{n-1}\big(B^{(1)}_{t_{k+1}}-B^{(1)}_{t_k}\big)\big(B^{(2)}_{t_{k+1}}-B^{(2)}_{t_k}\big),$$
则 $\mathrm{E}(T_n)=0$, $\mathrm{E}(T_n)^2\to 0$.

上面的引理形式地看成 $\sum\limits_{k=0}^{n}\Delta B^{(1)}_{t_k}\Delta B^{(2)}_{t_k}\approx 0,\quad \mathrm{d}B_t^{(1)}\mathrm{d}B_t^{(2)}=0.$

性质 5.5.1. 设 $\phi,\psi\in\mathscr{L}_T^2$, $\mathrm{E}\big(\int_0^T\phi\mathrm{d}B^{(i)}\int_0^T\psi\mathrm{d}B^{(j)}\big)=\delta_{ij}\mathrm{E}\big(\int_0^T\phi\psi\mathrm{d}s\big)$, 其中 $\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$.

定理 5.5.5. Itô 公式 Ⅳ

$$\mathrm{d}\xi_t=\Phi_t^{\mathrm{t}}d\mathbf{B}_t+\Psi_t\mathrm{d}t=\sum_{i=1}^{m}\phi_t^{(i)}\mathrm{d}B_t^{(i)}+\Psi_t\mathrm{d}t$$

则
$$\mathrm{d}F(t,\xi_t)=\frac{\partial F(t,\xi_t)}{\partial t}\mathrm{d}t+\frac{\partial F(t,\xi_t)}{\partial x}\mathrm{d}\xi_t+\frac{1}{2}\frac{\partial^2F(t,\xi_t)}{\partial x^2}(\mathrm{d}\xi_t)^2$$
即
$$F(t,\xi_t)-F(0,\xi_0)=\int_0^t\frac{\partial F(s,\xi_s)}{\partial s}\mathrm{d}s+\int_0^t\frac{\partial F(s,\xi_s)}{\partial x}\mathrm{d}\xi_s+\frac{1}{2}\int_0^t\frac{\partial^2F(s,\xi_s)}{\partial x^2}(\mathrm{d}\xi_s)^2$$
其中 $\mathrm{d}t$, $\mathrm{d}B_t^{(i)}$ 由下面的乘法表给出

	$\mathrm{d}t$	$\mathrm{d}B_t^{(i)}$
$\mathrm{d}t$	0	0
$\mathrm{d}B_t^{(i)}$	0	$\delta_{ij}\mathrm{d}t$

$$\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases},$$

即
$$\begin{aligned}&F(t,\xi_t)-F(0,\xi_0)\\ =&\int_0^t\Big[\frac{\partial F(s,\xi_s)}{\partial s}+\frac{\partial F(s,\xi_s)}{\partial x}\Psi_s+\frac{1}{2}\frac{\partial^2F(s,\xi_s)}{\partial x^2}\sum_{i=1}^{m}(\phi_s^{(i)})^2\Big]\mathrm{d}s+\sum_{i=1}^{m}\int_0^t\frac{\partial F(s,\xi_s)}{\partial x}\phi_s^{(i)}\mathrm{d}B_s^{(i)}.\end{aligned}$$

♣♣♣20230516

推论 5.5.1. 分部积分公式, 乘积公式

ξ_t, η_t 都是多维布朗运动 $\{\mathbf{B}_t,\,t\geq 0\}$ 的 Itô 过程, 则
$$\mathrm{d}(\xi_t\eta_t)=\xi_t\mathrm{d}\eta_t+\eta_t\mathrm{d}\xi_t+(\mathrm{d}\xi_t)(\mathrm{d}\eta_t).$$
写成积分形式
$$\xi_t\eta_t-\xi_0\eta_0=\int_0^t\xi_s\mathrm{d}\eta_s+\int_0^t\eta_s\mathrm{d}\xi_s+\int_0^t(\mathrm{d}\xi_s)(\mathrm{d}\eta_s).$$

进一步, 若 $\xi_t=\big(\xi_t^{(1)},\xi_t^{(2)},\cdots,\xi_t^{(d)}\big)^{\mathrm{t}}$, $(\xi_t^{(i)}),i\leq d$ 都是同一个 m 维布朗运动 $\mathbf{B}_t=\begin{pmatrix}B_t^{(1)}\\ \vdots\\ B_t^{(m)}\end{pmatrix}$ 的 Itô 过程, 即
$$\mathrm{d}\xi_t^{(i)}=(\Phi_t^{(i)})^{\mathrm{t}}\mathrm{d}\mathbf{B}_t+\Psi_t^{(i)}\mathrm{d}t,\ i\leq d,\ \Psi=\begin{pmatrix}\Psi_t^{(1)}\\ \vdots\\ \Psi_t^{(d)}\end{pmatrix}.$$

写成矩阵的形式

$$\Phi^{(i)} = \begin{pmatrix} \phi^{i1} \\ \vdots \\ \phi^{im} \end{pmatrix}, 1 \le i \le d;\ \Psi = \begin{pmatrix} \Psi_t^{(1)} \\ \vdots \\ \Psi_t^{(d)} \end{pmatrix},$$

$$\Phi = (\Phi^{(1)}, \cdots, \Phi^{(d)}) = \begin{pmatrix} \phi_t^{11} & \phi_t^{21} & \cdots & \phi_t^{d1} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_t^{1m} & \phi_t^{2m} & \cdots & \phi_t^{dm} \end{pmatrix},$$

$$\mathrm{d}\xi = \Phi_t^{\mathrm{t}}\,\mathrm{d}\mathbf{B}_t + \Psi\,\mathrm{d}t,$$

$$\begin{pmatrix} \mathrm{d}\xi_t^{(1)} \\ \vdots \\ \mathrm{d}\xi_t^{(d)} \end{pmatrix} = \begin{pmatrix} \phi_t^{11} & \phi_t^{12} & \cdots & \phi_t^{1m} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_t^{d1} & \phi_t^{d2} & \cdots & \phi_t^{dm} \end{pmatrix} \begin{pmatrix} \mathrm{d}B_t^{(1)} \\ \vdots \\ \mathrm{d}B_t^{(m)} \end{pmatrix} + \begin{pmatrix} \Psi_t^{(1)} \\ \vdots \\ \Psi_t^{(d)} \end{pmatrix} \mathrm{d}t.$$

定理 5.5.6. Itô 公式 V

$F(x_1, \cdots, x_d)$ 为一个二次连续可微的函数, 则

$$\begin{aligned} \mathrm{d}F(\xi_t^{(1)}, \cdots, \xi_t^{(d)}) &= \sum_{k=1}^{d} \frac{\partial}{\partial x_k} F(\xi_t^{(1)}, \cdots, \xi_t^{(d)}) \mathrm{d}\xi_t^{(k)} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} F(\xi_t^{(1)}, \cdots, \xi_t^{(d)}) \mathrm{d}\xi_t^{(i)} \mathrm{d}\xi_t^{(j)}, \\ \mathrm{d}\xi_t^{(i)} \mathrm{d}\xi_t^{(j)} &= a_{ij} \mathrm{d}t, \\ (a_{ij})_{i,j \le d} &= (\Phi^{\mathrm{t}}\Phi)_{i,j\le d} = (\phi^{ij})^{\mathrm{t}}_{\substack{i \le m \\ j \le d}} (\phi^{ij})_{\substack{i \le m \\ j \le d}}. \end{aligned}$$

5.6 Stratonovich-Fisk 对称积分

5.6.1 背景

5.6.2 定义及性质

定义 5.6.1. (Stratonovich-Fisk 随机积分) 设 $(\xi_t)_{t\ge 0}$ 是特殊类型的 Itô 过程, $\xi_t = \xi_0 + \int_0^t \phi_s \mathrm{d}B_s + \int_0^t \psi_s \mathrm{d}s$, 对于 $0 = t_0 < \cdots < t_n = t$, $\lambda_n = \max \Delta t_k$, Sratonovich-Fisk 对称积分定义为

$$\int_0^t \xi_s \circ \mathrm{d}B_s = L^2(\Omega) - \lim_{\lambda_n \to 0} \sum_{k=0}^{n-1} \frac{\xi_{t_k} + \xi_{t_{k+1}}}{2} \Delta B_{t_k}; \tag{5.6.1}$$

对于一般的 Itô 过程 $(\xi_t)_{t\ge 0}$, Stratonovich-Fisk 对称积分定义为如下的依概率收敛的极限:

$$\int_0^t \xi_s \circ \mathrm{d}B_s = \mathrm{P} - \lim_{\lambda_n \to 0} \sum_{k=0}^{n-1} \frac{\xi_{t_k} + \xi_{t_{k+1}}}{2} \Delta B_{t_k}. \tag{5.6.2}$$

注意到:

$$\sum_{k=0}^{n-1} \frac{\xi_{t_k} + \xi_{t_{k+1}}}{2} \Delta B_{t_k} = \sum_{k=0}^{n-1} \xi_{t_k} \Delta B_{t_k} + \frac{1}{2} \sum_{k=0}^{n-1} (\xi_{t_{k+1}} - \xi_{t_k}) \Delta B_{t_k}.$$

形式上有

$$\begin{aligned} \xi_t \circ \mathrm{d}B_t &= \xi_t \mathrm{d}B_t + \frac{1}{2} \mathrm{d}\xi_t \mathrm{d}B_t \\ &= \xi_t \mathrm{d}B_t + \frac{1}{2} (\phi_t \mathrm{d}B_t) \mathrm{d}B_t \\ &= \xi_t \mathrm{d}B_t + \frac{1}{2} \phi_t \mathrm{d}t. \end{aligned}$$

将上面的过程严格化, 得到如下的性质:

性质 5.6.1. $\int_0^t \xi_s \circ \mathrm{d}B_s = \int_0^t \xi_s \mathrm{d}B_s + \frac{1}{2}\int_0^t \phi_s \mathrm{d}s.$

例 5.6.1.

$$\begin{aligned}\int_0^t B_s \circ \mathrm{d}B_s &= \int_0^t B_s \mathrm{d}B_s + \frac{1}{2}\int_0^t 1\cdot \mathrm{d}s \\ &= \int_0^t B_s \mathrm{d}B_s + \frac{t}{2} = \frac{1}{2}(B_t^2 - B_0^2).\end{aligned}$$

定义 5.6.2. 设 $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ 是 Itô 过程

$$X_t = x_0 + \int_0^t \phi_s^X \mathrm{d}B_s + \int_0^t \psi_s^X \mathrm{d}s,$$
$$Y_t = y_0 + \int_0^t \phi_s^Y \mathrm{d}B_s + \int_0^t \psi_s^Y \mathrm{d}s.$$

$$\begin{aligned}\int_0^t Y_s \circ \mathrm{d}X_s &= \mathrm{P} - \lim_{\lambda_n\to 0}\sum_{k=0}^{n-1}\frac{Y_{t_k}+Y_{t_{k+1}}}{2}\left(X_{t_{k+1}} - X_{t_k}\right) \\ &= \int_0^t Y\mathrm{d}X + \frac{1}{2}\int_0^t \mathrm{d}Y\mathrm{d}X \\ &= \int_0^t Y_s\phi_s^X \mathrm{d}B_s + \int_0^t Y_s\psi_s^X\mathrm{d}s + \frac{1}{2}\int_0^t \phi_s^X\phi_s^Y\mathrm{d}s.\end{aligned}$$

$$\begin{aligned}Y_t\circ \mathrm{d}X_t &= Y_t\mathrm{d}X_t + \frac{1}{2}\mathrm{d}Y\mathrm{d}X \\ &= Y_t\phi_t^X\mathrm{d}B_t + Y\psi_t^X\mathrm{d}s + \frac{1}{2}\phi_s^X\phi_s^Y\mathrm{d}s.\end{aligned}$$

定理 5.6.1. (Itô 公式 VI) 设 $\xi_t = \xi_0 + \int_0^t \phi_s\mathrm{d}B_s + \int_0^t \psi_s\mathrm{d}s$, 即 $\mathrm{d}\xi_t = \phi_t\mathrm{d}B_t + \psi_t\mathrm{d}t$. 若 f 三次连续可微, 则

$$\mathrm{d}f(\xi_t) = f'(\xi_t)\circ \mathrm{d}\xi_t,$$

即

$$f(\xi_t) - f(\xi_0) = \int_0^t f'(\xi_s)\circ\mathrm{d}\xi_s = \int_0^t f'(\xi_s)\circ(\phi_s\mathrm{d}B_s) + \int_0^t f'(\xi_s)\psi_s\mathrm{d}s.$$

若 $f(t,x)$ 对 x 三次连续可微, 对 t 一次连续可微, 则

$$\mathrm{d}f(t,\xi_t) = f_t'\mathrm{d}t + f_x'\circ\mathrm{d}\xi_t.$$

第六章　随机微分方程和扩散过程

6.1　随机微分方程的概念和例子

如同常微分方程、偏微分方程一样, 随机微分方程 (stochastic differential equation, SDE) 可以描述相当广泛的一类物理现象、经济现象、生物现象.

随机微分方程的一般形式为

$$\mathrm{d}\xi_t = b(t,\xi)\mathrm{d}t + \sigma(t,\xi)\mathrm{d}B_t, \tag{6.1.1}$$

即

$$\xi_t = \xi_0 + \int_0^t b(s,\xi(\omega))\mathrm{d}s + \int_0^t \sigma(s,\xi(\omega))\mathrm{d}B_s.$$

其实这是一个随机积分方程, "随机微分方程" 只是约定俗成的叫法而已. 其中 b, σ 满足对于任意的 $s>0$, $b(s,\xi(\omega))$, $\sigma(s,\xi(\omega))$ 是关于 $\mathcal{F}_s$ 可测的随机变量, 即 $(b(s,\xi(\omega)))_{s\geq 0}$, $(\sigma(s,\xi(\omega)))_{s\geq 0}$ 是 $(\mathcal{F}_s)_{s\geq 0}$ 适应的. 也就是说 $b(s,\xi(\omega))$, $\sigma(s,\xi(\omega))$ 依赖于 $\xi(\omega)$ 在 $[0,s]$ 的历史信息, 由 $\xi(\omega)$ 在 $[0,s]$ 上所有的值决定的, 或者说 $b(s,\xi(\omega))$, $\sigma(s,\xi(\omega))$ 是 $\xi(\cdot,\omega), \cdot\in[0,s]$ 这个函数的泛函.

本章着重讨论如下形式的 SDE

$$\mathrm{d}\xi_t = b(t,\xi_t)\mathrm{d}t + \sigma(t,\xi_t)\mathrm{d}B_t$$

即

$$\xi_t = \xi_0 + \int_0^t b(s,\xi_s(\omega))\mathrm{d}s + \int_0^t \sigma(s,\xi_s(\omega))\mathrm{d}B_s$$

其中 $b(s,x)$, $\sigma(s,x)$, $\mathbb{R}^+\times\mathbb{R}\to\mathbb{R}$ 的 Borel 可测函数, 也就是说 $b(s,\xi_s)$, $\sigma(s,\xi_s)$ 的值由仅由 ξ_s 决定.

例 6.1.1. Ornstein-Uhlenbeck 方程, Langevin 方程

Langevin 等人用 Einstein 的布朗运动数学模型描述 R.Brown 观察到的液体中微小花粉粒子的不规则运动, 效果并不十分理想. 1908 年, Langevin 等人对模型进行了改进.

一个粒子在液体中运动, 其速度为 v. 由 Stokes 定律, 它受到液体对它的粘滞力为 $-\alpha v$, 设 m 为粒子的质量, 则由牛顿运动第二定律

$$m\dot{v} = -\alpha v. \tag{6.1.2}$$

在粒子质量很大, 但还不是极大时 (例如, 花粉粒子的质量是液体分子质量的 20 万倍, 2×10^5 倍), 除了宏观的粘滞力外, 液体分子对粒子杂乱的作用效果可以十分清楚的表现出来. 因此, 上述方程可以改写成

$$m\dot{v} = -\alpha v + F(t), \tag{6.1.3}$$

$F(t)$ 表示分子杂乱碰撞的力. 上式两边同除以 m

$$\dot{v} = -\gamma v + \Gamma(t), \quad \gamma = \frac{\alpha}{m}, \quad \Gamma(t) = \frac{F(t)}{m},$$

其中 γ 表示单位质量的阻尼系数, $\Gamma(t)$ 表示单位质量的花粉粒子受到液体分子碰撞 "涨落力", 被称为 Langevin 力. 上述方程被称为 Langevin 方程, 也被称为 Ornstein-Uhlenbeck 方程. 可以形式地记为

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\gamma v + \Gamma(t), \text{ 或 } \mathrm{d}v = -\gamma v\mathrm{d}t + \Gamma(t)\mathrm{d}t.$$

可以根据不同的物理系统假定 $\Gamma(t)$ 不同的统计性质, 再根据实验检验修正这些统计假设. Langevin 当年假设 $\Gamma(t)$ 是所谓的白噪声, 即满足如下的统计性质:

(1) 由于 $\Gamma(t)$ 为涨落力, 可以认为它的统计平均值为 0, 即 $\mathrm{E}\,\Gamma(t)=0$ (或用统计物理的记号 $\langle\Gamma(t)\rangle=0$).

(2) 同时因为宏观观测时间间隔远远大于微观分子对花粉粒子的碰撞时间, 所以不同时刻的 $\Gamma(t)$ 可以近似认为是相互独立的, $\mathrm{E}(\Gamma(t)\Gamma(t'))=2D\delta_0(t-t')$ (或 $\langle\Gamma(t)\Gamma(t')\rangle=2D\delta_0(t-t')$), 其中 $\delta_0(x)$ 为 Dirac 函数.

从概率论的角度上式可以解释为在任意小的两个不相交区间内, 液体分子对花粉粒子碰撞的涨落力是不相关, 在任意小的区间内涨落力的方差为 $2D$.

(3) 假定 $\Gamma(t)$ 有 Gauss 性.

在数学上我们可以认为 "$\Gamma(t)\mathrm{d}t$" 就是 $\mathrm{d}B_t$. 而 $\Gamma(t)$ 被称为 "白噪声"[①], $\Gamma(t)$ 有时形式上记为 "$\frac{\mathrm{d}B_t}{\mathrm{d}t}$" 或 "$\dot{B}_t$". 因此, Langevin 方程可以写成一个标准的随机微分方程 (取 $2D=1$):

$$\begin{aligned}\mathrm{d}v_t &= -\gamma v_t\mathrm{d}t+\mathrm{d}B_t \text{ 或} \\ v_t-v_0 &= -\int_0^t \gamma v_s\mathrm{d}s+B_t,\ (\text{设} B_0=0).\end{aligned} \tag{6.1.4}$$

下面考虑如何解这个方程, 形式地

$$\frac{\mathrm{d}v_t}{\mathrm{d}t}=-\gamma v_t+\dot{B}_t,\ \text{或}\ \mathrm{d}v_t=-\gamma v_t\mathrm{d}t+\mathrm{d}B_t.$$

物理的解法: 借助常微分方程 $\frac{\mathrm{d}v_t}{\mathrm{d}t}=-\gamma v_t+f(t)$ 的解法.

(1) 先考虑齐次方程

$$\frac{\mathrm{d}v_t}{\mathrm{d}t}=-\gamma v_t\ \Rightarrow\ v_t=c_0\mathrm{e}^{-\gamma t}.$$

(2) 再考虑非齐次方程 $\frac{\mathrm{d}v_t}{\mathrm{d}t}=-\gamma v_t+f(t)$, 或 $\mathrm{d}v_t+\gamma v_t\mathrm{d}t=f(t)\mathrm{d}t$. 两端乘以 $\mathrm{e}^{\gamma t}$

$$\mathrm{e}^{\gamma t}\mathrm{d}v_t+\mathrm{e}^{\gamma t}\gamma v_t\mathrm{d}t=\mathrm{e}^{\gamma t}f(t)\mathrm{d}t,$$

即

$$\mathrm{d}(\mathrm{e}^{\gamma t}v_t)=\mathrm{e}^{\gamma t}f(t)\mathrm{d}t.$$

$$\begin{aligned}\mathrm{e}^{\gamma t}v_t &= \int_0^t \mathrm{e}^{\gamma s}f(s)\mathrm{d}s+v_0, \\ v_t &= \int_0^t \mathrm{e}^{\gamma(s-t)}f(s)\mathrm{d}s+v_0\mathrm{e}^{-\gamma t}.\end{aligned}$$

(3) 用 $\dot{B}_t$ 替代 $f(t)$ 或 $f(t)\mathrm{d}t$ 替代 $\mathrm{d}B_t$ 得

$$v_t=v_0\mathrm{e}^{-\gamma t}+\int_0^t \mathrm{e}^{\gamma(s-t)}\mathrm{d}B_s=v_0\mathrm{e}^{-\gamma t}+\mathrm{e}^{-\gamma t}\int_0^t \mathrm{e}^{\gamma s}\mathrm{d}B_s. \tag{6.1.5}$$

数学严格的解法: 由 Itô 公式

$$\begin{aligned}\mathrm{d}(\mathrm{e}^{\gamma t}v_t) &= \mathrm{e}^{\gamma t}\mathrm{d}v_t+v_t\mathrm{d}\mathrm{e}^{\gamma t} \\ &= -\gamma\mathrm{e}^{\gamma t}v_t\mathrm{d}t+\mathrm{e}^{\gamma t}\mathrm{d}B_t+\gamma\mathrm{e}^{\gamma t}v_t\mathrm{d}t \\ &= \mathrm{e}^{\gamma t}\mathrm{d}B_t.\end{aligned}$$

写出积分形式即得 (6.1.5).

[①] 在上面的例子为什么把 $\Gamma(t)$ 称为白噪声呢? $\Gamma(t)$ 的时间相关函数是 $D\delta_0(t-t')$, 它的功率谱密度是常数, 用数学的话说就是 δ_0 的 Fourier 变换是常函数 1. 也就是说 $2D\delta_0(t-t')$ 在各个频率上的分量的大小都一样, 这与白光的性质类似, 所以当 $\Gamma(t)$ 的时间相关函数是 $2D\delta_0(t-t')$ 时, $\Gamma(t)$ 被称为白噪声.

例 6.1.2. 在上面的例子中, 如果系数都是非随机的数值函数, 方程成为

$$\mathrm{d}X_t = F(t)X_t\mathrm{d}t + C(t)\mathrm{d}B_t, \tag{6.1.6}$$

其中假设系数在任意有限区间内有界, 求解此方程.

解: 类似于上面例子, 先考虑齐次方程

$$\mathrm{d}x_t - F(t)x_t\mathrm{d}t = 0, \quad \text{解得 } x_t = x_0\mathrm{e}^{\int_0^t F(s)\mathrm{d}s}.$$

将 $\mathrm{e}^{-\int_0^t F(s)\mathrm{d}s}$ 作为恰当因子乘在方程 (6.1.6) 两边并利用 Itô 公式,

$$\begin{aligned}\mathrm{d}\big(X_t\mathrm{e}^{-\int_0^t F(s)\mathrm{d}s}\big) &= \mathrm{e}^{-\int_0^t F(s)\mathrm{d}s}\mathrm{d}X_t - X_tF(t)\mathrm{e}^{-\int_0^t F(s)\mathrm{d}s}\mathrm{d}t \\ &= C(t)\mathrm{e}^{-\int_0^t F(s)\mathrm{d}s}\mathrm{d}B_t.\end{aligned}$$

因此,

$$\begin{aligned}X_t &= X_0\mathrm{e}^{\int_0^t F(s)\mathrm{d}s} + \mathrm{e}^{\int_0^t F(s)\mathrm{d}s}\int_0^t C(u)\mathrm{e}^{-\int_0^u F(s)\mathrm{d}s}\mathrm{d}B_u \\ &= X_0\mathrm{e}^{\int_0^t F(s)\mathrm{d}s} + \int_0^t C(u)\mathrm{e}^{-\int_u^t F(s)\mathrm{d}s}\mathrm{d}B_u.\end{aligned}$$

□

例 6.1.3. 随机调和振子

调和振子的运动方程为

$$\begin{cases}\frac{\mathrm{d}^2\xi_t}{\mathrm{d}t^2} = -\lambda^2\xi_t \\ \xi_0 = x,\ \frac{\mathrm{d}\xi_t}{\mathrm{d}t}\big|_{t=0} = v_0\end{cases} \tag{6.1.7}$$

若振子受到白噪声的随机干扰, 形式上

$$\begin{cases}\frac{\mathrm{d}^2\xi_t}{\mathrm{d}t^2} = -\lambda^2\xi_t + \dot{B}_t \\ \xi_0 = x,\ \frac{\mathrm{d}\xi_t}{\mathrm{d}t}\big|_{t=0} = v_0\end{cases}. \tag{6.1.8}$$

随机微分方程实质上是随机积分方程, 因此随机微分方程的一般形式 (6.1.1) 中是没有 $\frac{\mathrm{d}^2\xi_t}{\mathrm{d}t^2}$ 或 $\mathrm{d}^2\xi_t$ 这样的项. 那么 (6.1.8) 这个形式方程对应的随机微分方程应该是什么?

在常微分方程理论中, 一个高阶的常微分方程总可以化成一阶的方程组, 例如在 (6.1.7) 中, 设 $\frac{\mathrm{d}\xi_t}{\mathrm{d}t} = v_t$, 则 (6.1.7) 等价于

$$\begin{cases}\frac{\mathrm{d}v_t}{\mathrm{d}t} = -\lambda^2\xi_t \\ \frac{\mathrm{d}\xi_t}{\mathrm{d}t} = v_t \\ \xi_0 = x \\ v_t|_{t=0} = v_0\end{cases}.$$

因此, 我们可以把 (6.1.8) 转化成一个有明确数学含义的随机微分方程组,

$$\begin{cases}\mathrm{d}v_t = -\lambda^2\xi_t\mathrm{d}t + \mathrm{d}B_t \\ \mathrm{d}\xi_t = v_t\mathrm{d}t \\ \xi_0 = x \\ v_t|_{t=0} = v_0\end{cases}. \tag{6.1.9}$$

或者写成随机积分的形式

$$\begin{cases}v_t = v_0 - \lambda^2\int_0^t \xi_s\mathrm{d}s + B_t & (B_t\text{设为标准布朗运动}) \\ \mathrm{d}\xi_t = v_t\mathrm{d}t,\ \text{或}\ \xi_t = x + \int_0^t v_s\mathrm{d}s\end{cases}. \tag{6.1.10}$$

接下来的问题是如何解随机微分方程 (6.1.9) 或 (6.1.10). 我们还是先参照一下在常微分方程理论. 若 (6.1.7) 中有一个确定性的强迫力

$$\begin{cases}\frac{\mathrm{d}^2\xi_t}{\mathrm{d}t^2} = -\lambda^2\xi_t + f(t) \\ \xi_0 = x,\ \frac{\mathrm{d}\xi_t}{\mathrm{d}t}\big|_{t=0} = v_0\end{cases}.$$

由常微分方程理论知道,

$$\xi_t = x\cos\lambda t + \frac{v_0}{\lambda}\sin\lambda t + \frac{1}{\lambda}\int_0^t \sin\lambda(t-s)f(s)\mathrm{d}s.$$

想象存在光滑的 $f_n(s)$ 满足 $f_n(s)\mathrm{d}s \to \dot{B}_s\mathrm{d}s = \mathrm{d}B_s$, 则形式地

$$\xi_t = x\cos\lambda t + \frac{v_0}{\lambda}\sin\lambda t + \frac{1}{\lambda}\int_0^t \sin\lambda(t-s)\mathrm{d}B_s. \tag{6.1.11}$$

问题是 (6.1.11) 中定义的 ξ_t 真的是 (6.1.10) 的解吗? v_t 又是什么? 下面验证 (6.1.11) 满足方程 (6.1.9) 即可.

首先求 $\frac{\mathrm{d}\xi_t}{\mathrm{d}t} = v_t$ 时我们遇到的问题是如何对 $\int_0^t \sin\lambda(t-s)\mathrm{d}B_s$ 求导? 事实上, 利用分部积分公式或乘积公式 [①]

$$\int_0^t \sin\lambda(t-s)\mathrm{d}B_s = \sin\lambda(t-s)B_s\big|_0^t - \int_0^t B_s\mathrm{d}\sin\lambda(t-s) = \lambda\int_0^t B_s\cos\lambda(t-s)\mathrm{d}s. \tag{6.1.12}$$

由此得到

$$\xi_t = x\cos\lambda t + \frac{v_0}{\lambda}\sin\lambda t + \int_0^t B_s\cos\lambda(t-s)\mathrm{d}s. \tag{6.1.13}$$

$$\frac{\mathrm{d}\xi_t}{\mathrm{d}t} = -\lambda x\sin\lambda t + v_0\cos\lambda t - \lambda\int_0^t B_s\sin\lambda(t-s)\mathrm{d}s + B_t = v_t. \tag{6.1.14}$$

另一方面, 对 (6.1.13) 式两边关于 t 做积分, 得到

$$\begin{aligned}\int_0^t \xi_s\mathrm{d}s &= \frac{x}{\lambda}\sin\lambda t + \frac{v_0}{\lambda^2}\big(1-\cos\lambda t\big) + \int_0^t\int_0^u B_s\cos\lambda(u-s)\mathrm{d}s\mathrm{d}u \\ &= \frac{x}{\lambda}\sin\lambda t + \frac{v_0}{\lambda^2}\big(1-\cos\lambda t\big) + \int_0^t B_s\int_s^t \cos\lambda(u-s)\mathrm{d}u\mathrm{d}s \\ &= \frac{1}{\lambda^2}\Big(v_0 + \lambda x\sin\lambda t - v_0\cos\lambda t + \lambda\int_0^t B_s\sin\lambda(t-s)\mathrm{d}s\Big),\end{aligned}$$

将上式与 (6.1.14) 比较, 即得

$$v_t = v_0 - \lambda^2\int_0^t \xi_s\mathrm{d}s + B_t,$$

因此 $(\xi_t, v_t)_{t\geq 0}$ 是方程 (6.1.9) 或 (6.1.10) 的解.

例 6.1.4. Black-Scholes 方程

在第三章例子 (2.4) 中我们讨论了连续时间的 Black-Scholes 模型的直观推导, 得到了风险证券券价格 $(S_t)_{t\geq 0}$ 形式上应该满足

$$S_t - S_0 = \mu\int_0^t S_s\mathrm{d}s + \sigma\int_0^t S_s\text{``}\mathrm{d}\bar{\xi}_s\text{''}. \tag{6.1.15}$$

[①] $\int_0^t \sin\lambda(t-s)\mathrm{d}B_s$ 不是 (局部鞅) 鞅, 也不是 Itô 过程, 并不能直接利用分部积分公式 (5.5.1).

分部积分公式 (6.1.12) 证明有两种方式. 第一种是对于固定的 t 利用随机积分的定义程序, 先对阶梯函数做随机积分得到等式左边的项, 再利用 Able 求和得到右边的项, 两边取 L^2 极限就可以了. 第二种方式是对于固定的 t, 在 $[0,t]$ 内时间定义 $f^{(t)}(s) = \sin\lambda(t-s)$, 由于 f 是确定性的数值函数, 所以是 $(\mathcal{F}_u)_{0\leq u\leq t}$ 适应的. 由乘积公式

$$\mathrm{d}\big(f^{(t)}(s)B_s\big) = f^{(t)}(s)\mathrm{d}B_s + B_s\mathrm{d}f^{(t)}(s) + \mathrm{d}B_s\mathrm{d}f^{(t)}(s).$$

右边第三项为零, 写成积分形式

$$f^{(t)}(u)B_u - f^{(t)}(0)B_0 = \int_0^u f^{(t)}(s)\mathrm{d}B_s + \int_0^t B_s\mathrm{d}f^{(t)}(s), \quad u\in[0,t].$$

即

$$\sin\lambda(t-u)B_u - \sin\lambda(t-0)B_0 = \int_0^u \sin\lambda(t-s)\mathrm{d}B_s + \int_0^u B_s\mathrm{d}\sin\lambda(t-s).$$

当 $u=t$, $B_0=0$ 时, 就得到了 (6.1.12).

现在, 我们可以将 $\int_0^t S_s$“$\mathrm{d}\bar{\xi}_s$” 解释成关于布朗运动的 Itô 随机积分了, 称

$$\begin{cases} \mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}B_t \\ S_0 = s_0 \end{cases} \tag{6.1.16}$$

为 Black-Scholes 方程.

下面我们关心如何求解这个方程和解的长时间行为这两个问题.

(1) 先形式地求解, 设 $g(x) = \ln x$

$$\begin{aligned} \mathrm{d}g(S_t) &= g'(S_t)\mathrm{d}S_t + \frac{1}{2}g''(S_t)(\mathrm{d}S_t)^2 \\ &= \frac{1}{S_t}[\mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}B_t] - \frac{1}{2}\frac{1}{S_t^2}(\mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}B_t)^2 \\ &= [\mu - \frac{1}{2}\sigma^2]\mathrm{d}t + \sigma \mathrm{d}B_t. \end{aligned}$$

$$\begin{aligned} \ln S_t - \ln s_0 &= [\mu - \frac{1}{2}\sigma^2]t + \sigma B_t, \quad (\text{设} B_0 = 0) \\ \ln \frac{S_t}{s_0} &= [\mu - \frac{1}{2}\sigma^2]t + \sigma B_t. \\ S_t &= s_0 \exp\{[\mu - \frac{1}{2}\sigma^2]t + \sigma B_t\}. \end{aligned} \tag{6.1.17}$$

从数学的严格性上来说, 我们上面的计算只是猜出解的形式. 我们还要面对两个问题: 第一, 解的唯一性. 随机微分方程解的存在唯一性定理6.2.1就可以给出肯定的答案; 第二, 若解有正概率小于等于零, 则 $\ln S_t$ 是不能定义的. 这个问题的解决方案是, 既然方程解是唯一的, 不论初值为正还是为负, 我们对 (6.1.17) 式用 Itô 公式直接验证满足方程即可.

♣♣♣20230518

(2) 显然, 当 $s_0 = 0$ 时, $S_t = 0$ 是方程的解. 下面我们只考虑 $s_0 > 0$ 的情况.

$$S_t = s_0 \exp\big\{\big[(\mu - \frac{1}{2}\sigma^2) + \frac{\sigma B_t}{t}\big]t\big\}.$$

已知布朗运动的强大数律

$$\mathrm{P}\big(\lim_{t\to\infty} \frac{B_t}{t} = 0\big) = 1.$$

当 $\mu - \frac{1}{2}\sigma^2 > 0$ 时,

$$\mathrm{P}\big(\lim_{t\to\infty} \big[(\mu - \frac{1}{2}\sigma^2) + \frac{\sigma B_t}{t}\big] > 0\big) = 1.$$

因此,

$$\mathrm{P}\big(\lim_{t\to\infty} S_t = +\infty\big) = 1.$$

当 $\mu - \frac{1}{2}\sigma^2 < 0$ 时, 类似地可以得到

$$\mathrm{P}\big(\lim_{t\to\infty} S_t = 0\big) = 1.$$

当 $\mu - \frac{1}{2}\sigma^2 = 0$ 时,

$$S_t = s_0 \exp\big\{\sigma B_t\big\} = s_0 \exp\big\{\sigma \frac{B_t}{\sqrt{2t\ln\ln t}}\sqrt{2t\ln\ln t}\big\}.$$

由布朗运动的重对数律

$$\mathrm{P}\big(\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} = 1\big) = 1,$$

$$\mathrm{P}\big(\liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} = -1\big) = 1,$$

得到

$$\mathrm{P}\big(\limsup_{t\to\infty} S_t = +\infty,\ \liminf_{t\to\infty} S_t = 0\big) = 1.$$

我们还可以将 (6.1.15) 在数学上解释成 Stratonovich 型的随机微分方程

$$\begin{cases} \mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \circ \mathrm{d}B_t \\ S_0 = s_0 \end{cases}$$

这个方程的解是 $S_t = s_0 \exp\{\mu t + \sigma B_t\}$.

通常把 $\exp(\mu t + \nu B_t)$ 称为几何布朗运动.

我们可以将上面的例子进行推广. 如果收益率 μ 和波动率 σ 依赖于历史信息, 也就是说收益率和波动率是关于事件域流 $(\mathcal{F}_t)_{t\geq 0}$ 适应的随机过程, 分别记为 $\mu_t(\omega)$ 和 $\sigma_t(\omega)$, 那么可以将 Black-Scholes 方程推广成

$$\mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sigma_t S_t \mathrm{d}B_t, \quad S_0 = s_0. \tag{6.1.18}$$

在一定的条件下, (例如对于任意的 $t > 0$, $\sup_{s\leq t}\{|\mu_s|, |\sigma_s|\} < M_t$, M_t 是一个非随机的函数), 这个方程是有显式解的. 求解留作习题.

6.2 扩散过程的 Itô 理论和鞅问题简介

6.2.1 解的存在唯一性

考虑随机微分方程

$$\mathrm{d}\xi_t = b(t, \xi_t)\mathrm{d}t + \sigma(t, \xi_t)\mathrm{d}B_t, \tag{6.2.1}$$

定理 6.2.1. 若 $\sigma(t,x), b(t,x)$ 满足 Lipschitz 条件, 即对于任意的 $T > 0$ 存在 $C_T > 0$, 使得任意的 $t < T$,

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq C_T|x - y|.$$

对于初始随机变量 ξ_0, ξ_0 与 $(B_t, t \geq 0)$ 独立, 则方程

$$\xi_t = \xi_0 + \int_0^t b(s, \xi_s)\mathrm{d}s + \int_0^t \sigma(s, \xi_s)\mathrm{d}B_s$$

有唯一解, 即若还有另一个解 ξ_t', 只要 $\mathrm{P}(\xi_0' = \xi_0) = 1$, 则

$$\mathrm{P}(\xi_t' = \xi_t, \ \forall t > 0) = 1.$$

证明: Picard 迭代, Gronwall 不等式. □

定理 6.2.2. 若 $\sigma(t,x)$, $b(t,x)$ 满足:

(1) 局部 Lipschitz 条件, 即对于任意的 $T, N > 0$ 存在 $C_{T,N} > 0$, 使得对于任意的 $|x| \leq N$, $|y| \leq N$, $t \leq T$

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq C_{T,N}|x - y|.$$

(2) 线性增长条件, 即对于任意的 $T > 0$ 存在 $C_T > 0$, 使得任意的 $t < T$ (T 可以取 ∞)

$$|\sigma(t,x)| + |b(t,x)| \leq C_T(1 + |x|),$$

对于初始随机变量 ξ_0, ξ_0 与 $(B_t, t \geq 0)$ 独立, 则方程

$$\xi_t = \xi_0 + \int_0^t b(s, \xi_s)\mathrm{d}s + \int_0^t \sigma(s, \xi_s)\mathrm{d}B_s.$$

有唯一解, 即若还有另一个解 ξ_t', 只要 $\mathrm{P}(\xi_0' = \xi_0) = 1$, 则

$$\mathrm{P}(\xi_t' = \xi_t, \forall t > 0) = 1.$$

6.2.2 解的 Markov 性

定理 6.2.3. 随机微分方程 (6.2.1) 的解若存在唯一, 则这个解是一个 Markov 过程. 而且当系数不含时间 t 时, 即 $\sigma(t,x)=\sigma(x)$, $b(t,x)=b(x)$ 时, 此时 Markov 过程是时齐的.

证明的直观解释: 将随机微分方程 (6.2.1) 写成积分形式,

$$\xi_t = \xi_0+\int_0^t b(u,\xi_u)\mathrm{d}u+\int_0^t \sigma(u,\xi_u)\mathrm{d}B_u, \tag{6.2.2}$$

$$\xi_t^{s,x} = x+\int_s^t b(u,\xi_u^{s,x})\mathrm{d}u+\int_s^t \sigma(u,\xi_u^{s,x})\mathrm{d}B_u,\quad s<t. \tag{6.2.3}$$

其中 $\xi_t^{s,x}$ 表示从 s 时刻 x 位置出发的方程的解. 它只与 $\big(s,x,t,(B_u-B_s,s<u\le t)\big)$ 有关, 即 $\xi_t^{s,x}$ 是 $\big(s,x,t,(B_u-B_s,s\le u\le t)\big)$ 的函数 (泛函), 记为

$$G\big(s,x,t,(B_u-B_s,s\le u\le t)\big). \tag{6.2.4}$$

另一方面

$$\xi_t=\xi_0+\int_0^s b(u,\xi_u)\mathrm{d}u+\int_0^s \sigma(u,\xi_u)\mathrm{d}B_u+\int_s^t b(u,\xi_u)\mathrm{d}u+\int_s^t \sigma(u,\xi_u)\mathrm{d}B_u. \tag{6.2.5}$$

因此, 若方程 (6.2.1) 的解存在唯一, 则 (6.2.2) 与 (6.2.5) 决定了同一个过程, 即

$$\xi_t=\xi_s+\int_s^t b(u,\xi_u)\mathrm{d}u+\int_s^t \sigma(u,\xi_u)\mathrm{d}B_u.$$

由 (6.2.4), ξ_t 可以表示成 $G(s,\xi_s,t,(B_u-B_s,s<u\le t))$.

注意到 $(B_u-B_s,s<u\le t)$ 与 s 之前的事件域 $\sigma(B_v,\,v\le s)$ 独立. 因此, ξ_t 在已知 ξ_s 的条件下与 $\sigma(B_v,v\le s)$ 独立, 也就是说, 对于任意的有界可测函数 f,

$$\mathrm{E}[f(\xi_t)|\mathcal{F}_s^B]=\mathrm{E}[f(G(s,\xi_s,t,(B_u-B_s,s<u\le t)))|\mathcal{F}_s^B]=\mathrm{E}\big(f(G(s,x,t,(B_u-B_s,s<u\le t)))\big)\big|_{x=\xi_s},$$

利用条件数学期望的投影性质

$$\begin{aligned}\mathrm{E}[f(\xi_t)|\xi_s]&=\mathrm{E}\big[\mathrm{E}[f(\xi_t)|\mathcal{F}_s^B]\big|\xi_s\big]\\&=\mathrm{E}\big[\mathrm{E}\big(f(G(s,x,t,(B_u-B_s,s<u\le t)))\big)\big|_{x=\xi_s}\big|\xi_s\big]\\&=\mathrm{E}[f(\xi_t)|\mathcal{F}_s^B]=\mathrm{E}[f(G(s,\xi_s,t,(B_u-B_s,s<u\le t)))|\mathcal{F}_s^B].\end{aligned}$$

这说明 $(\xi_t)_{t\ge0}$ 是一个 $(\mathcal{F}_t^B)_{t\ge0}$ Markov 过程. □

若 $b(u,\xi)$, $\sigma(u,\xi)$ 是有记忆型时, 则 ξ_t 是不能表示成 $G(s,\xi_s,t,(B_u-B_s,s<u\le t))$. 原因是 $b(u,\xi),\sigma(u,\xi)$ 当 $u>s$ 是还可能依赖于 $\sigma(B_v,\,v\le s)$.

若随机微分方程

$$\mathrm{d}\xi_t=b(t,\xi_t)\mathrm{d}t+\sigma(t,\xi_t)\mathrm{d}B_t, \tag{6.2.6}$$

有唯一解, 有时也称这个解为扩散过程[①], 其中 $b(t,x)$ 称为漂移系数, $a(t,x)=\sigma^2(t,x)$ 称为扩散系数. $a(t,x)$ 有时被称为弥漫系数 (dispersion coefficient).

下面我们需要验证随机微分方程 (6.2.1) 定义的扩散过程也就是我们在第五章引言中希望直观构造的扩散过程, 与第四章章用分析方法给出的扩散过程在一定条件下是一致的. 待补充

例 6.2.1. 接例6.1.1

我们将 Langevin 方程 (6.1.4) 写成稍微一般一点儿的形式

$$\mathrm{d}\eta_t=-b\eta_t\mathrm{d}t+\sigma\mathrm{d}B_t,$$

则解为

$$\eta_t=\mathrm{e}^{-bt}\big(\eta_0+\sigma\int_0^t \mathrm{e}^{bs}\mathrm{d}B_s\big),$$

[①]扩散过程这个名词来源于物理等学科, 在不同的书上有不同的严格数学定义, 但其核心是指 (右) 连续的强马氏过程. 我们这里简单地粗率地把它看成无记忆性随机微分方程 (6.2.1) 的解.

称为 OU 过程 (Ornstein-Uhlenbeck 过程, 严格地说这个方程的解从其不变概率分布出发时才能称为 OU 过程[①]).

由定理6.2.1和定理6.2.3, $(\eta_t)_{t\geq 0}$ 是一个时齐的 Markov 过程. 自然的我们可以问它的转移概率密度是什么, 它的是否有不变概率分布?

Markov 过程 $(\eta_t)_{t\geq 0}$ 的转移概率分布 $\mathrm{P}(s,x;s+t,A)$ 可以理解为过程从 s 时刻 x 位置出发在 $s+t$ 时刻到达集合 A 的概率, 因此我们只需要计算

$$\eta_{t+s}^{s,x}=\mathrm{e}^{-bt}\big(x+\sigma\int_s^{t+s}\mathrm{e}^{b(u-s)}\mathrm{d}B_u\big)$$

的分布函数. 因为 e^{bu} 是确定性函数, 所以 $\int_s^{t+s}\mathrm{e}^{bu}\mathrm{d}B_u$ 是服从 Gauss 分布的随机变量, 利用 Itô 等距,

$$\int_s^{t+s}\mathrm{e}^{bu}\mathrm{d}B_u\sim N\Big(0,\int_s^{t+s}\mathrm{e}^{2bu}\mathrm{d}u\Big)=N\Big(0,\frac{\mathrm{e}^{2b(t+s)}-\mathrm{e}^{2bs}}{2b}\Big).$$

那么

$$\eta_{t+s}^{s,x}\sim N\Big(x\mathrm{e}^{-bt},\ \frac{\sigma^2}{2b}(1-\mathrm{e}^{-2bt})\Big).$$

由此得 $(\eta_t)_{t\geq 0}$ 的转移密度函数是

$$p(t,x,y)\equiv p(s,x;s+t,y)=\sqrt{\frac{b}{\pi\sigma^2(1-\mathrm{e}^{-2bt})}}\exp\Big(-\frac{(y-x\mathrm{e}^{-bt})^2}{\frac{\sigma^2(1-\mathrm{e}^{-2bt})}{b}}\Big).$$

当 $b>0$ 时,

$$\lim_{t\to\infty}p(t,x,y)=\frac{1}{\sigma\sqrt{\frac{\pi}{b}}}\mathrm{e}^{-\frac{y^2}{\frac{\sigma^2}{b}}}\equiv\pi(y).$$

由此可得 $N(0,\frac{\sigma^2}{2b})$ 是 OU 过程的不变概率分布, 也就是说以 π 为初始分布密度, 则 OU 过程为一个平稳的马氏过程其平稳分布为 π.

由时间齐次性, 事实上可以通过计算

$$\eta_t=\mathrm{e}^{-bt}x+\sigma\int_0^t\mathrm{e}^{b(s-t)}\mathrm{d}B_s,$$

的密度来求转移概率密度函数.

例 6.2.2. 接例6.1.4

已知 Black-Scholes 方程 (6.1.16) 的解

$$S_t=s_0\exp\Big\{\big[\mu-\frac{1}{2}\sigma^2\big]t+\sigma B_t\Big\}.$$

同样, 由定理6.2.1和定理6.2.3知道, 它是一个 Markov 过程. 特别地, 当 $s_0=0$ 时, 过程恒等于零. 因此我们不妨考虑 $s_0>0$, 也就是说这个过程的状态空间是 $(0,\infty)$. 对于 $f(z)=\mathbf{1}_{(-\infty,y]}(z)$,

$$\begin{aligned}
&\mathrm{E}[f(S_{t+s})|S_s]\\
=\;&\mathrm{E}\big[f\big(s_0\exp\{[\mu-\frac{1}{2}\sigma^2]s+\sigma B_s\}\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big|S_s\big]\\
=\;&\mathrm{E}\big[f\big(S_s\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big|S_s\big]\\
=\;&\mathrm{E}\big[f\big(x\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big]\Big|_{x=S_s}\\
=\;&\mathrm{P}\Big(x\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\leq y\Big)\Big|_{x=S_s}\\
=\;&\mathrm{P}\Big(\frac{B_{t+s}-B_s}{\sqrt{t}}\leq\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big)\Big|_{x=S_s}\\
=\;&\Phi\Big(\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big)\Big|_{x=S_s}.
\end{aligned}$$

[①]设 $b,\sigma>0$, B_t 是一个标准布朗运动, 则平稳的 OU 过程另一种表示为 $\frac{\sigma}{\sqrt{2b}}\mathrm{e}^{-bt}B(\mathrm{e}^{2bt})$.

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y$. $(S_t)_{t\geq 0}$ 的转移概率分布函数

$$F(t,x,y)=\Phi\Big(\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big). \tag{6.2.7}$$

6.2.3 鞅问题

多维扩散过程的随机微分方程形式如下:

$$\begin{pmatrix}\mathrm{d}\xi_t^{(1)}\\ \vdots\\ \mathrm{d}\xi_t^{(n)}\end{pmatrix}=\begin{pmatrix}b_1(t,\xi_t)\\ \vdots\\ b_n(t,\xi_t)\end{pmatrix}\mathrm{d}t+\begin{pmatrix}\sigma_{11}(t,\xi_t) & \cdots & \sigma_{1r}(t,\xi_t)\\ \vdots & \vdots & \vdots\\ \sigma_{n1}(t,\xi_t) & \cdots & \sigma_{nr}(t,\xi_t)\end{pmatrix}\begin{pmatrix}\mathrm{d}B_t^{(1)}\\ \vdots\\ \mathrm{d}B_t^{(r)}\end{pmatrix}. \tag{6.2.8}$$

设

$$\mathbf{b}(t,x)=\begin{pmatrix}b_1(t,x)\\ \vdots\\ b_n(t,x)\end{pmatrix},\quad \Sigma(t,x)=\begin{pmatrix}\sigma_{11}(t,x) & \cdots & \sigma_{1r}(t,x)\\ \vdots & \vdots & \vdots\\ \sigma_{n1}(t,x) & \cdots & \sigma_{nr}(t,x)\end{pmatrix}_{n\times r}.$$

可以将方程 (6.2.8) 记为

$$\mathrm{d}\boldsymbol{\xi}_t=\mathbf{b}(t,\boldsymbol{\xi}_t)\mathrm{d}t+\Sigma(t,\boldsymbol{\xi}_t)\mathrm{d}\mathbf{B}_t. \tag{6.2.9}$$

考虑随机微分方程 (6.2.8)(6.2.9). 先定义扩散算子

$$\mathcal{L}_t=\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}(t,x)\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(t,x)\frac{\partial}{\partial x_i},\quad (a_{ij}(t,x))_{1\leq i,j\leq n}=\Sigma\Sigma^{\mathrm{t}}. \tag{6.2.10}$$

对于任意的 $f(t,x)\in C_{\mathbf{b}}^{1,2}([0,\infty)\times\mathbb{R}^n)$, 由 Itô 公式,

$$\begin{aligned}&f(t,\boldsymbol{\xi}_t)-f(0,\boldsymbol{\xi}_0)-\int_0^t\big[\frac{\partial f(s,\boldsymbol{\xi})}{\partial s}+\sum_{i=1}^{n}b_i(s,\boldsymbol{\xi}_s)\frac{\partial f(s,\boldsymbol{\xi}_s)}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}(s,\boldsymbol{\xi}_s)\frac{\partial^2 f(s,\boldsymbol{\xi}_s)}{\partial x_i\partial x_j}\big]\mathrm{d}s\\ =\;&\int_0^t\sum_{i=1}^{n}\frac{\partial f(s,\boldsymbol{\xi}_s)}{\partial x_i}\sum_{k=1}^{r}\sigma_{ik}(s,\boldsymbol{\xi}_s)\mathrm{d}B_s^{(k)}.\end{aligned}$$

即

$$f(t,\boldsymbol{\xi}_t)-f(0,\boldsymbol{\xi}_0)-\int_0^t\big(\frac{\partial}{\partial s}+\mathcal{L}_s\big)f(s,\boldsymbol{\xi}_s)\mathrm{d}s \tag{6.2.11}$$

是连续局部鞅.

反过来, 若仅仅知道形如 (6.2.10) 的扩散算子, 是否可以构造一个带流的概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathrm{P})$ 以及其上马氏过程 $(\boldsymbol{\xi}_t)_{t\geq 0}$ 满足对于任意的 $f(t,x)\in C_{\mathbf{b}}^{1,2}([0,\infty)\times\mathbb{R}^n)$, (6.2.11) 是连续的 $(\mathcal{F}_t)_{t\geq 0}$ 局部鞅且其转移概率密度 (函数) 满足向前、向后方程? 进一步, 是否存在 $(\mathcal{F}_t)_{t\geq 0}$ 布朗运动 $(\mathbf{B}_t)_{t\geq 0}$ 使得 $(\boldsymbol{\xi}_t,\mathbf{B}_t)_{t\geq 0}$ 满足随机微分方程 (6.2.8).

这是 Stroock 和 Varadhan 建立的多维扩散过程鞅问题的基本想法①.

6.3 随机微分方程与扩散过程的一些应用

6.3.1 Dirichlet 边值问题的概率表示

设 $(\xi_t)_{t\geq 0}$ 为一个时齐的 n-维扩散过程,

$$\mathrm{d}\boldsymbol{\xi}_t=\mathbf{b}(\boldsymbol{\xi}_t)\mathrm{d}t+\Sigma(\boldsymbol{\xi}_t)\mathrm{d}\mathbf{B}_t. \tag{6.3.1}$$

①这一理论被称为没有随机积分的随机分析. 其优点是在更一般的状态空间上, 针对马氏过程的无穷小特征和利用鞅来构造马氏过程和研究过程的相关性质.

其中,

$$\mathbf{b}(x)=\begin{pmatrix} b_1(x)\\ \vdots \\ b_n(x)\end{pmatrix},\quad \Sigma(x)=\begin{pmatrix}\sigma_{11}(x) & \cdots & \sigma_{1r}(x)\\ \vdots & \vdots & \vdots \\ \sigma_{n1}(x) & \cdots & \sigma_{nr}(x)\end{pmatrix}_{n\times r}.$$

定义扩散算子:

$$\mathcal{L}=\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}(x)\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(x)\frac{\partial}{\partial x_i},\quad (a_{ij}(x))_{1\le i,j\le n}=\Sigma\Sigma^{\mathrm{t}}.$$

定理 6.3.1. 设 $D\subset\mathbf{R}^n$ 为一个有界开集. $\mathcal{L}$ 的系数连续且在区域 D 上一致椭圆, 即存在 $\delta>0$, 使得 $\sum_{1\le i,j\le n}a_{ij}\eta_i\eta_j\ge\delta\|\eta\|^2$ 成立. g 是 D 上的连续函数.

$$\begin{cases}\mathcal{L}v(x)=-g(x) & x\in D\\ v(x)=\phi(x) & x\in\partial D\end{cases},\tag{6.3.2}$$

具有直到边界 ∂D 的二阶连续可导的解 $v(x)$, 那么 $v(x)$ 有如下的概率表示:

$$v(x)=\mathrm{E}\big[\phi(\boldsymbol{\xi}_\tau)+\int_0^\tau g(\boldsymbol{\xi}_u)\mathrm{d}u|\boldsymbol{\xi}_0=x\big]\tag{6.3.3}$$

其中 $\tau=\inf\{t>0,\ \boldsymbol{\xi}_t\notin D\}$.

证明: 记 $\mathrm{E}[\cdot|\boldsymbol{\xi}_0=x]=\mathrm{E}_x(\cdot)$. 首先, 将 v 开拓到全空间成为一个 C_0^2 中的函数, 考虑

$$\boldsymbol{\eta}_t=v(\boldsymbol{\xi}_t)-\int_0^t\mathcal{L}v(\boldsymbol{\xi}_u)\mathrm{d}u,$$

则 $(\boldsymbol{\eta}_t,\mathcal{F}_t^{\mathbf{B}})_{t\ge0}$ 是鞅. 由停时定理,

$$\mathrm{E}_x(\boldsymbol{\eta}_{t\wedge\tau})=\mathrm{E}_x\boldsymbol{\eta}_0=v(x),$$

即

$$v(x)=\mathrm{E}_x\big(v(\boldsymbol{\xi}_{t\wedge\tau})-\int_0^{t\wedge\tau}\mathcal{L}v(\boldsymbol{\xi}_u)\mathrm{d}u\big).\tag{6.3.4}$$

由于 $t\le\tau$ 时, $\boldsymbol{\xi}_t\in\overline{D}$, 并且 v 和 $\mathcal{L}v$ 在紧集 $\overline{D}$ 上连续, 因此有界. 从而 $v(\boldsymbol{\xi}_{t\wedge\tau})$ 和 $\mathcal{L}v(\boldsymbol{\xi}_{t\wedge\tau})$ 有界, 也就是说存在常数 c, 使得

$$\int_0^{t\wedge\tau}\mathcal{L}v(\boldsymbol{\xi}_u)\mathrm{d}u\le c\tau.$$

若 $\mathrm{E}_x\tau<\infty$, 在 (6.3.4) 中令 $t\to\infty$, 由控制收敛定理和 $\boldsymbol{\xi}_t$ 的连续性得

$$v(x)=\mathrm{E}_x\big(v(\boldsymbol{\xi}_\tau)-\int_0^\tau\mathcal{L}v(\boldsymbol{\xi}_u)\mathrm{d}u\big)=\mathrm{E}_x\big(\phi(\xi_\tau)+\int_0^\tau g(\boldsymbol{\xi}_u)\mathrm{d}u\big).$$

♣♣♣20230525

下面验证 $\mathrm{E}_x\tau<\infty$. 由一致椭圆性条件可以得, 存在正整数 l 满足 $1\le l\le n$ 使得

$$\min_{x\in\overline{D}}a_{ll}>0.$$

设 $b:=\max_{x\in\overline{D}}\{|b_1(x)|,\cdots,|b_n(x)|\}$, $a:=\min_{x\in\overline{D}}a_{ll}(x)$ 和 $q:=\min_{x\in\overline{D}}x_l$. 取 $\nu>\frac{2b}{a}$, 设

$$h(x)=-\mu\mathrm{e}^{\nu x_l},\ x=(x_1,\cdots,x_n)\in\overline{D},$$

其中 μ 是一个待定常数.

$$\begin{aligned}-(\mathcal{L}h)(x) &= \mu\mathrm{e}^{\nu x_l}\big[\frac{1}{2}\nu^2a_{ll}(x)+\nu b_l(x)\big]\\ &\ge \frac{1}{2}\mu\nu a\mathrm{e}^{\nu q}(\nu-\frac{2b}{a})\end{aligned}$$

取 μ 充分大, 使得 $(\mathcal{L}h)(x)\le-1$, $x\in\overline{D}$.

因为 h 及它的各阶导数在 $\overline{D}$ 中均有界, 所以 $h(\boldsymbol{\xi}_{t\wedge\tau}) - h(x) - \int_0^{t\wedge\tau}(\mathcal{L}h)(\boldsymbol{\xi}_s)\mathrm{d}s$ 为鞅. 因而

$$\mathrm{E}_x h(\boldsymbol{\xi}_{t\wedge\tau}) - h(x) - \mathrm{E}_x \int_0^{t\wedge\tau}(\mathcal{L}h)(\boldsymbol{\xi}_s)\mathrm{d}s = 0,$$

从而

$$\mathrm{E}_x(t\wedge\tau) \le h(x) - \mathrm{E}_x h(\boldsymbol{\xi}_{t\wedge\tau}) \le 2\max_{x\in\overline{D}}|h(x)| < \infty.$$

令 $t\to\infty$ 得 $\mathrm{E}_x\tau < \infty$. □

(6.3.3) 式给出了二阶椭圆偏微分方程 Dirichlet 边值问题 (6.3.2) 解的概率表示. 它的一个应用是利用随机模拟提供 Dirichlet 问题解数值计算的一种方法; 另一方面, 利用 (6.3.3) 也可以将某些扩散过程相关问题通过偏微分方程的解来表示.

问题 1 设 $(\xi_t)_{t\ge0}$ 从 $x\in D$ 出发, 对于 ∂D 上的一个可测集 Γ, 求 $\mathrm{P}(\xi_\tau\in\Gamma|\boldsymbol{\xi}_0 = x)$, 即首达边界地点的分布.

问题 2 设 $(\xi_t)_{t\ge0}$ 从 $x\in D$ 出发, 求 $\mathrm{E}[\tau|\boldsymbol{\xi}_0 = x]$, 即首达时的期望.

问题 3 给定 $z\in D$, 若 $g(z)$ 表示系统在状态 z 时的 “支出费用率” 或 “成本率”(cost rate), 那么 $C(\omega) = \int_0^\tau g(\xi_s(\omega))\mathrm{d}s$ 表示当 $(\xi_t)_{t\ge0}$ 跑出边界 ∂D 时总的支出费用或成本, 求平均支出费用或成本.

特别地, 当 $g(z)=1, z\in D$ 时, $C(\omega)=\tau(\omega)$, 问题 2 是问题 3 的一个特殊形式.

由此, 定理 6.3.1可以得到下面的推论.

推论 6.3.1. (首达地点的分布) *对于边界 ∂D 上的一个可测集 Γ. 若方程*

$$\begin{cases} \mathcal{L}u = 0 & x\in D \\ u(x) = 1 & x\in\Gamma \\ u(x) = 0 & x\in\partial D\setminus\Gamma \end{cases}, \tag{6.3.5}$$

在 D 上有二次连续可微的解 u_Γ. 则

$$\mathrm{P}(\xi_\tau\in\Gamma|\boldsymbol{\xi}_0 = x) = u_\Gamma(x).$$

证明: 在定理 6.3.1中取 $g(x)=0$, $\phi(x)=\mathbf{1}_\Gamma(x), x\in\partial D$ 即得. □

推论 6.3.2. (首达时的期望)

$$\begin{cases} \mathcal{L}u = -1 & x\in D \\ u(x) = 0 & x\in\partial D \end{cases}, \tag{6.3.6}$$

在 D 上有二次连续可微的解 u_Γ. 则

$$\mathrm{E}[\tau|\xi_0 = x] = u(x).$$

证明: 在定理 6.3.1中取 $g(x)=1$, $\phi(x)=0, x\in\partial D$, 则

$$u(x) = \mathrm{E}[\int_0^\tau g(\xi_u)\mathrm{d}u|\boldsymbol{\xi}_0 = x] = \mathrm{E}[\tau|\xi_0 = x].$$

□

推论 6.3.3. 设 $g(z), z\in D$ 是有界连续函数. 若方程

$$\begin{cases} \mathcal{L}u = -g & x\in D \\ u(x) = 0 & x\in\partial D \end{cases}, \tag{6.3.7}$$

具有直到边界 ∂D 的二阶连续可导的解 $u(x)$, 则

$$u(x) = \mathrm{E}[\int_0^\tau g(\xi_u)\mathrm{d}u|\boldsymbol{\xi}_0 = x].$$

例 6.3.1. 设 $B(0,r)=\{x\in\mathbf{R}^n:\|x\|<r\}$, u 是方程

$$\begin{cases}\frac{1}{2}\Delta u=-1 & x\in B(0,r)\\ u(x)=0 & x\in\partial B(0,r)\end{cases},$$

的解, 即 $u(x)=\frac{r^2-\|x\|^2}{n},\ \ x\in B(0,r)$. 则对于 n-维 Brown 运动 $(\xi_t)_{t\geq 0}$, $\tau=\inf\{t\geq 0,\,\xi_t\notin B(0,r)\}$,

$$\mathrm{E}(\tau|\xi_0=x)=\frac{r^2-\|x\|^2}{n},\ \ x\in B(0,r).$$

6.3.2 一维扩散过程的 Feller 自然尺度函数和标准测度及其应用

对于一维扩散过程

$$\mathrm{d}\xi_t=b(\xi_t)\mathrm{d}t+\sigma(\xi_t)\mathrm{d}B_t, \tag{6.3.8}$$

此时 $\mathcal{L}=\frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2}+b(x)\frac{\mathrm{d}}{\mathrm{d}x}$. 考虑区间 $[l,r]\subset\mathbb{R}$, 当 $\xi_0=x\in(l,r)$ 时, 设

$$\tau_l=\inf\{t>0,\xi_t=l\},\ \tau_r=\inf\{t>0,\xi_t=r\},$$

$$\tau_{(l,r)}=\inf\{t>0,\xi_t\notin(l,r)\}=\min\{\tau_l,\tau_r\}=\tau_l\wedge\tau_r.$$

设 $\sigma(x),b(x)$ 是 $[l,r]$ 上的连续函数, 且 $\sigma^2(x)>0$, $x\in[l,r]$, 则推论6.3.1、推论6.3.2和推论6.3.3中分别对应的偏微分方程 (6.3.5)、(6.3.6) 和 (6.3.7) 是可解的.

首先, 在一维情况下方程 (6.3.5) 可以写成

$$\begin{cases}\frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2u(x)}{\mathrm{d}x^2}+b(x)\frac{\mathrm{d}u(x)}{\mathrm{d}x}=0 & x\in(l,r)\\ u(x)=1 & x=l\\ u(x)=0 & x=r\end{cases}. \tag{6.3.9}$$

事实上, 可以很容易给出方程

$$\frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2u(x)}{\mathrm{d}x^2}+b(x)\frac{\mathrm{d}u(x)}{\mathrm{d}x}=0 \tag{6.3.10}$$

的通解. 设 $s(x)=\frac{\mathrm{d}u(x)}{\mathrm{d}x}$, 则 $\frac{\mathrm{d}s(x)}{\mathrm{d}x}+\frac{2b(x)}{\sigma^2(x)}s(x)=0$, 即

$$s(x)=\exp\big(\int_\alpha^x-\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\big). \tag{6.3.11}$$

进一步方程 (6.3.10) 的通解是

$$u(x)=\int_\beta^x s(y)\mathrm{d}y=\int_\beta^x\exp\Big(\int_\alpha^y-\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\Big)\mathrm{d}y,\ \ \text{其中}\alpha,\beta\text{是不定常数}. \tag{6.3.12}$$

将 (6.3.9) 中的边值条件 $u(l)=1,u(r)=0$ 代入 (6.3.12) 来确定 α 和 β, 解得 $\beta=r$ 以及 α 是使得 $\int_r^l\exp\Big(\int_\alpha^y-\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\Big)\mathrm{d}y=1$ 成立的唯一常数, 由此即得方程 (6.3.9) 的解.

我们可以将 (6.3.11) 和 (6.3.12) 右端的不定积分形式地记为

$$s(x)=\exp\big(\int^x-\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\big),\ S(x)=\int^x\exp\Big(\int^y-\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\Big)\mathrm{d}y. \tag{6.3.13}$$

通过以上的求解过程, 我们不难发现, 可以将 $\mathcal{L}=\frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2}+b(x)\frac{\mathrm{d}}{\mathrm{d}x}$ 写成如下的形式

$$\mathcal{L}f(x)=\frac{1}{2}\Big([\sigma^2(x)s(x)]\frac{\mathrm{d}}{\mathrm{d}x}\big[\frac{1}{s(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x}\big]\Big).$$

注意到 $\frac{\mathrm{d}S(x)}{\mathrm{d}x}=s(x)$, 即 $\frac{\mathrm{d}x}{\mathrm{d}S(x)}=\frac{1}{s(x)}$. 设 $m(x)=\frac{1}{\sigma^2(x)s(x)}$, M 是以 $m(x)$ 为密度函数的测度, 即 $\frac{\mathrm{d}M}{\mathrm{d}x}(x)=m(x)$ 或者 $\mathrm{d}M=m(x)\mathrm{d}x$. 可以定义如下的微分算子:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}S}=\frac{1}{s(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x},\ \frac{\mathrm{d}f(x)}{\mathrm{d}M}=\frac{1}{m(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x}.$$

由此我们可以将 $\mathcal{L}$ 写成紧凑和更有意义的关于 $\frac{\mathrm{d}}{\mathrm{d}S}$ 与 $\frac{\mathrm{d}}{\mathrm{d}M}$ 相继微分运算的表示形式:

$$\mathcal{L}f(x)=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}M}\Big[\frac{\mathrm{d}f(x)}{\mathrm{d}S}\Big]. \tag{6.3.14}$$

(6.3.14) 称为扩散算子的典则表示 (canonical representation). 通过引入 Feller 自然尺度 $S(x)$、速度密度 $m(x)$ 和速度测度 M 可以给出问题 1-3 的解简洁和更具概率意义的表示.

定义 6.3.1. 称

$$S(x)=\int^x \exp\Big(\int^y -\frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\Big)\mathrm{d}y$$

是扩散算子 $\mathcal{L}=\frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2}+b(x)\frac{\mathrm{d}}{\mathrm{d}x}$ 在 (l,r) 上的 Feller 自然尺度 (natural scale).

称

$$m(x)=\frac{1}{\sigma^2(x)s(x)}$$

是扩散算子 $\mathcal{L}$ 在 (l,r) 上的 Feller 速度密度 (speed density).

称

$$M(\cdot)=\int_{\cdot} m(x)\mathrm{d}x$$

是扩散算子 $\mathcal{L}$ 在 (l,r) 上的 Feller 速度测度 (speed measure). 对于不同的 $s(x)$, 对应着不同的标准测度 M, 但它们只差一个常数倍.

重新考虑方程 (6.3.9),

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}M}\Big[\frac{\mathrm{d}u(x)}{\mathrm{d}S}\Big]=0,\ u(l)=1,\ u(r)=0.$$

对上式做一次积分得到 $\frac{\mathrm{d}u(x)}{\mathrm{d}S}=\beta$, 再做一次积分得到 $u(x)=\alpha+\beta S(x)$, 其中 α 和 β 是积分常数, $l\le x\le r$. 带入边界条件 $u(r)=0$ 得到 $\alpha=-\beta S(r)$, 带入 $u(l)=1$ 得到 $\beta=\frac{1}{S(l)-S(r)}$. 因此, 得到

推论 6.3.4. 方程 (6.3.9) 的解可以表示为

$$u(x)=\frac{S(r)-S(x)}{S(r)-S(l)}. \tag{6.3.15}$$

记 $\mathrm{P}_x(\cdot)=\mathrm{P}(\cdot|\xi_0=x)$, 则有

$$\mathrm{P}_x(\xi_{\tau_{(l,r)}}=l)=\frac{S(r)-S(x)}{S(r)-S(l)},\ \mathrm{P}_x(\xi_{\tau_{(l.r)}}=r)=\frac{S(x)-S(l)}{S(r)-S(l)}. \tag{6.3.16}$$

或者

$$\mathrm{P}_x(\tau_l<\tau_r)=\frac{S(r)-S(x)}{S(r)-S(l)},\ \mathrm{P}_x(\tau_r<\tau_l)=\frac{S(x)-S(l)}{S(r)-S(l)}. \tag{6.3.17}$$

对于 (6.3.13) 中的不定积分, 取不同的一对不定常数 (α_1,β_1) 和 (α_2,β_2) 可以定义不同的自然尺度 $S_1(x)$ 和 $S_2(x)$. 计算可得存在 A 和 B 使得 $S_1(x)=AS_2(x)+B$. 因此, 不同自然尺度的取法不影响方程 (6.3.9) 的解和 (6.3.16) 和 (6.3.17) 的概率值.

特别地, 若 $(\xi_t)_{t\ge0}$ 是一维布朗运动, 则

$$\mathrm{P}_x(\xi_{\tau_{(l,r)}}=l)=\frac{r-x}{r-l},\ \mathrm{P}_x(\xi_{\tau_{(l,r)}}=r)=\frac{x-l}{r-l}.$$

与布朗运动做对比, 对于一般的一维扩散过程 $(\xi_t)_{t\ge0}$ 首中 l 或 r 的概率是按照 $S(x)$ 作比例分配的. 这就不难理解为什么称 $S(x)$ 为扩散算子 (过程) 的自然尺度了.

更进一步, 设 $S(x)$ 为扩散过程 (6.3.8) 对应的自然尺度. $S(x)$ 在 $[S(l),S(r)]$ 内严格单调, 二次可微而且满足 $\frac{1}{2}\sigma^2(x)S''(x)+b(x)S'(x)=0$. 设 $Y_t=S(\xi_t)$, 由 Itô 公式

$$\begin{aligned}\mathrm{d}S(\xi_t)&=S'(\xi_t)b(\xi_t)\mathrm{d}t+S'(\xi_t)\sigma(\xi_t)\mathrm{d}B_t+\frac{1}{2}S''(\xi_t)\sigma^2(\xi_t)\mathrm{d}t\\&=s(\xi_t)\sigma(\xi_t)\mathrm{d}B_t.\end{aligned}$$

记 $S^{-1}(y)$ 是 $S(x)$ 的反函数, 则由上式 $(Y_t)_{t\ge0}$ 满足

$$\mathrm{d}Y_t=s(S^{-1}(Y_t))\sigma(S^{-1}(Y_t))\mathrm{d}B_t.$$

容易计算出扩散过程 $(Y_t)_{t\ge0}$ 在 $(S(l).S(r))$ 上对应的自然尺度是线性函数 $S_Y(y)=A+By, B\ne0$ (或者可以简单地取 $S_Y(y)=y$). 对于 $S(l)\le a<b\le S(r)$, $y\in(a,b)$,

$$\mathrm{P}(Y_{\tau_{(a,b)}}=a|Y_0=y)=\frac{b-y}{b-a},\quad \mathrm{P}(Y_{\tau_{(a,b)}}=b|Y_0=y)=\frac{y-a}{b-a},$$

也就是说 $(\xi_t)_{t\geq 0}$ 在区间 (l,r) 上用 $S(x)$ 重新 "标度" 成 $Y_t = S(\xi_t)$ 之后, 其首中时的概率分布按照出发点距离边界的比例分配, 这点与布朗运动相同[①].

接下来我们考虑问题 3 的一维情形, 其中 $g(x)$ 是有界连续函数.

$$\begin{cases} \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}M}\left[\frac{\mathrm{d}v(x)}{\mathrm{d}S}\right] = -g(x) & x \in (l,r) \\ v(x) = 0 & x = l \\ v(x) = 0 & x = r \end{cases}, \tag{6.3.18}$$

则 $v(x) = \mathrm{E}(\int_0^{\tau_{(l,r)}} g(\xi_u)\mathrm{d}u|\xi_0 = x)$. 问题 2 是它的特殊情况, 即 $g(x) = 1$, $v(x) = \mathrm{E}(\tau_{(l,r)}|\xi_0 = x)$.

对于上式做第一次积分得到:

$$\frac{\mathrm{d}v(y)}{\mathrm{d}S} = -2\int_l^y g(z)\mathrm{d}M(z) + \beta = -2\int_l^y g(z)m(z)\mathrm{d}z + \beta.$$

再做第二次积分得到:

$$v(x) = -2\int_l^x \big[\int_l^y g(z)m(z)\mathrm{d}z\big]\mathrm{d}S(y) + \beta[S(x) - S(l)] + \alpha.$$

由 $v(l) = 0$ 得到 $\alpha = 0$; $v(r) = 0$ 得到 $\beta = \frac{2}{S(r)-S(l)}\int_l^r \big[\int_l^y g(z)m(z)\mathrm{d}z\big]\mathrm{d}S(y)$. 积分交换顺序, 可以得到 $v(x)$ 如下对称的表达形式:

$$\begin{aligned} v(x) &= -2\int_l^x \big[\int_l^y g(z)m(z)\mathrm{d}z\big]\mathrm{d}S(y) + 2\frac{S(x)-S(l)}{S(r)-S(l)}\int_l^r \big[\int_l^y g(z)m(z)\mathrm{d}z\big]\mathrm{d}S(y) \\ &= -2\int_l^x \big[\int_z^x \mathrm{d}S(y)\big]g(z)m(z)\mathrm{d}z + 2\frac{S(x)-S(l)}{S(r)-S(l)}\int_l^r \big[\int_z^r \mathrm{d}S(y)\big]g(z)m(z)\mathrm{d}z \\ &= -2\int_l^x [S(x)-S(z)]g(z)m(z)\mathrm{d}z + 2\frac{S(x)-S(l)}{S(r)-S(l)}\int_l^r [S(r)-S(z)]g(z)m(z)\mathrm{d}z \\ &= -2\int_l^x [S(x)-S(z)]g(z)m(z)\mathrm{d}z + 2\frac{S(x)-S(l)}{S(r)-S(l)}\int_l^x [S(r)-S(z)]g(z)m(z)\mathrm{d}z \\ &\quad +2\frac{S(x)-S(l)}{S(r)-S(l)}\int_x^r [S(r)-S(z)]g(z)m(z)\mathrm{d}z \\ &= 2\Big\{\frac{S(r)-S(x)}{S(r)-S(l)}\int_l^x [S(z)-S(l)]g(z)m(z)\mathrm{d}z + \frac{S(x)-S(l)}{S(r)-S(l)}\int_x^r [S(r)-S(z)]g(z)m(z)\mathrm{d}z\Big\} \\ &= 2\Big\{u(x)\int_l^x [S(z)-S(l)]g(z)m(z)\mathrm{d}z + [1-u(x)]\int_x^r [S(r)-S(z)]g(z)m(z)\mathrm{d}z\Big\}. \end{aligned}$$

此处 $S(x)$ 与 $m(x)$ 是相互匹配的, 其不同的取法不影响上面式子的值. 特别地, 当 $g(x) = 1$ 时,

$$\begin{aligned} v(x) &= 2\Big\{\frac{S(r)-S(x)}{S(r)-S(l)}\int_l^x [S(z)-S(l)]m(z)\mathrm{d}z + \frac{S(x)-S(l)}{S(r)-S(l)}\int_x^r [S(r)-S(z)]m(z)\mathrm{d}z\Big\} \\ &= 2\Big\{u(x)\int_l^x [S(z)-S(l)]m(z)\mathrm{d}z + [1-u(x)]\int_x^r [S(r)-S(z)]m(z)\mathrm{d}z\Big\}. \end{aligned} \tag{6.3.19}$$

例 6.3.2. 若 $(\xi_t)_{t\geq t}$ 是布朗运动, $\xi_0 = x$, 则 $s(x) = 1$, $m(x) = 1$, 取 $S(x) = x$. 由 (6.3.19),

$$\mathrm{E}(\tau_{(l,r)}|\xi_0 = x) = 2\Big\{\frac{r-x}{r-l}\int_l^x (z-l)\mathrm{d}z + \frac{x-l}{r-l}\int_x^r (r-z)\mathrm{d}z\Big\} = (r-x)(x-l).$$

考虑布朗运动从 x 出发逃出 $(x-\varepsilon, x+\varepsilon)$ 的平均时间

$$\mathrm{E}(\tau_{(x-\varepsilon,x+\varepsilon)}|\xi_0 = x) = \varepsilon^2,\ \text{即}\ \frac{1}{\varepsilon^2}\mathrm{E}(\tau_{(x-\varepsilon,x+\varepsilon)}|\xi_0 = x) = 1.$$

若进一步 $(\xi_t)_{t\geq 0}$ 的自然尺度 S 是线性函数时, 特别地取为 $S(x) = x$, $s(x) = 1$, 考虑 ξ_t 从 x 出发逃出 $(x-\varepsilon, x+\varepsilon)$ 的平均时间, 由 (6.3.19),

$$\mathrm{E}(\tau_{(x-\varepsilon,x+\varepsilon)}|\xi_0 = x) = \int_x^{x+\varepsilon}(x+\varepsilon-z)m(z)\mathrm{d}z + \int_{x-\varepsilon}^x (z-x+\varepsilon)m(z)\mathrm{d}z,$$

[①] 在一些书中, 对于 $(\xi)_{t\geq 0}$, 将 $S(x)$ 仅仅称为尺度函数 (scale function), 当 $S(x)$ 为线性函数时, 称 $S(x)$ 为自然尺度 (natural scale) 或者典则尺度 (canonical scale).

进一步计算得到

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon^2}\mathrm{E}(\tau_{(x-\varepsilon,x+\varepsilon)}|\xi_0=x)=m(x). \tag{6.3.20}$$

速度密度 $m(x)$ 是表达 ξ_t 从位置 x 的一个小邻域逃逸出来“快慢”的量, 或者理解为过程 ξ_t 的“时钟”在 x 位置运行的“速度”. 布朗运动速度密度恒为 1, 可以认为是“标准时钟”, 对于具有线性自然尺度的扩散过程 ξ_t, 从 $(x-\varepsilon,x+\varepsilon)$ 逃逸出来所花费的平均时间的主阶项是 $\varepsilon^2 m(x)$[①]. 此处的 $m(x)$ 对应于 Poisson 过程强度参数 λ 和连续时间马氏链中第 i 个位置处 q_i. $\frac{1}{\lambda}$ 或 $\frac{1}{q_i}$ 表示在 i 位置停留的平均时间, λ 或 q_i 表示从 i 位置起跳的强度, 也就是表示从 i 位置逃逸的快慢的量.

记

$$G(x,y)=\begin{cases}2\dfrac{\big[S(r)-S(x)\big]\big[S(y)-S(l)\big]}{S(r)-S(l)}\dfrac{1}{\sigma^2(y)s(y)}, & l\le y\le x\le r\\ 2\dfrac{\big[S(x)-S(l)\big]\big[S(r)-S(y)\big]}{S(r)-S(l)}\dfrac{1}{\sigma^2(y)s(y)}, & l\le x\le y\le r\end{cases}.$$

称 $G(x,y)$ 为扩散过程 ξ_t 在 (l,r) 上的 Green 函数[②]. 二阶微分方程 (6.3.18) 的解可以表示为:

$$v(x)=\mathrm{E}(\int_0^{\tau(l,r)}g(\xi_u)\mathrm{d}u|\xi_0=x)=\int_l^r G(x,y)g(y)\mathrm{d}y. \tag{6.3.21}$$

设 $g(x)=\mathbf{1}_{[z,z+\Delta)}(x)$, 利用连续函数逼近 $g(x)$, 由 (6.3.21) 得到

$$\mathrm{E}\Big(\int_0^{\tau(l,r)}\mathbf{1}_{[z,z+\Delta)}(\xi_u)\mathrm{d}u\Big|\xi_0=x\Big)=\int_z^{z+\Delta}G(x,y)\mathrm{d}y.$$

$G(x,y)\mathrm{d}y$ 的概率意义是 ξ_t 从 x 出发在逃逸出 (l,r) 之前在 y 的无穷小领域 $[y,y+\mathrm{d}y)$ 内停留的平均时间.

♣♣♣20230530

例 6.3.3. (Feller 分支扩散过程)

先考虑第三章例3.1.7和例3.2.3中的分支过程 $\xi_0=1$, $\xi_n=\sum_{k=1}^{\xi_{n-1}}\eta_{n,k}$. 设 $\mathrm{E}\eta_{n,1}=\mu, \mathrm{var}\eta_{n,1}=\gamma^2$. 考虑 $(\xi_n)_{n\ge 0}$ 的 Doob 分解.

$$\Delta\xi_n=(\mu-1)\xi_n+\gamma\sqrt{\xi_n}\frac{\Delta\xi_n-(\mu-1)\xi_n}{\gamma\sqrt{\xi_n}}. \tag{6.3.22}$$

保持上述“分支机制”, 对离散时间和离散状态分支过程 (6.3.22) 取适当的时间-空间尺度极限就得到 Feller 分枝扩散过程

$$\mathrm{d}X_t=\alpha X_t\mathrm{d}t+\sigma\sqrt{X_t}\mathrm{d}B_t. \tag{6.3.23}$$

它的扩散系数不是局部 Lipschitz 连续的, 但利用 Yamada-Watanabe 定理[③]可以证明方程 (6.3.23) 的 (强) 解存在唯一, 并且可以利用强马氏性可以证明:

定理 6.3.2. 设 X_t 是方程 (6.3.23) 的解且 $X_0>0$, 则

$$\mathrm{P}(X_t\ge 0,\ \forall t\ge 0)=1.$$

[①]对于取不同的 $S(x)$, (6.3.20) 得到的结果都一样, 均是 $s(x)=1$ 时对应的速度密度 $m(x)$.

[②]$G(x,y)$ 也是扩散算子 $\mathcal{L}$ 在区间 (l,r) 上关于 Dirichlet 零边值问题对应的 Green 函数. 即将 $G(x,y)$ 考虑为 y 的函数, 它是看成如下方程的解:

$$\begin{cases}\frac{1}{2}\sigma^2(y)\frac{\mathrm{d}^2G(x,y)}{\mathrm{d}y^2}+b(y)\frac{\mathrm{d}G(x,y)}{\mathrm{d}y}=-\delta_x, & y\in(l,r)\\ G(x,y)=0, & y=l,\ y=r\end{cases}.$$

其中 δ_x 是 x 点的 Dirac 测度. $G(x,y)$ 的物理意义可以解释为在杆 (l,r) 的内部 x 处放一个单位点热源, l 端和 r 端与外界接触保持零度. 那么杆关于 $\mathcal{L}$ 的稳定温度场就是 Green 函数; 或者是导体 (l,r) 两端接地, 在内部 x 位置放一个单位正电荷, 那么导体内部关于 $\mathcal{L}$ 的电势分布就是 Green 函数.

[③]Yamada-Watanabe 定理: 对于一维随机微分方程 $\mathrm{d}X_t=b(X_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}B_t$, 若 $b(x)$ 是 Lipschitz 连续的, $\sigma(x)$ 是 α 阶 Hölder 连续的, $\alpha\ge\frac{1}{2}$, 即 $|\sigma(x)-\sigma(y)|<K|x-y|^{\alpha}$, 则上述随机微分方程存在唯一的 (强) 解.

进一步, 设 $\tau=\inf\{t>0,X_t=0\}$, 则

$$\mathrm{P}(X_t=0,\forall t\geq\tau)=1.$$

定理 6.3.3. 设 X_t 是方程 (6.3.23) 的解且 $X_0>0$, 则

$$\mathrm{E}X_t=X_0\mathrm{e}^{\alpha t},$$

$X_t\mathrm{e}^{-\alpha t}$ 为非负鞅, 且当 $\alpha>0$, $t\to\infty$ 时, 几乎处处收敛到非退化的极限.

定理 6.3.4. 设 X_t 是方程 (6.3.23) 的解且 $X_0=x>0$, 则 X_t 的灭绝概率 $\mathrm{P}_x(\tau<\infty)=\begin{cases}\mathrm{e}^{-\frac{2\alpha}{\sigma^2}x} & \alpha>0\\ 1 & \alpha\leq 0\end{cases}$.

证明: 取 $0<a_n<b_n$, $\tau=\inf\{t>0,X(t)=0\}$, $\tau_{a_n}=\inf\{t>0,X(t)=a_n\}$, $\tau_{b_n}=\inf\{t>0,X(t)=b_n\}$ 和 $\tau_\infty=\inf\{t>0,X(t)=\infty\}$, 则

$$\mathrm{P}_x(\tau_{a_n}<\tau_{b_n})=\frac{S(b_n)-S(x)}{S(b_n)-S(a_n)}=\frac{\mathrm{e}^{-\frac{2\alpha}{\sigma^2}x}-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_n}}{\mathrm{e}^{-\frac{2\alpha}{\sigma^2}a_n}-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_n}}.$$

令 $a_n\downarrow 0$, 则 $\tau=\lim_{n\to\infty}\tau_{a_n}$, $\{\tau<\tau_{b_m}\}=\bigcap_{n=1}^\infty\{\tau_{a_n}<\tau_{b_m}\}$. 由概率的从上连续性得到

$$\mathrm{P}_x(\tau<\tau_{b_m})=\lim_{n\to\infty}\frac{\mathrm{e}^{-\frac{2\alpha}{\sigma^2}x}-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_n}}{\mathrm{e}^{-\frac{2\alpha}{\sigma^2}a_n}-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_n}}=\frac{\mathrm{e}^{-\frac{2\alpha}{\sigma^2}x}-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_m}}{1-\mathrm{e}^{-\frac{2\alpha}{\sigma^2}b_m}}.$$

令 $b_m\uparrow\infty$, 则 $\tau_\infty=\lim_{m\to\infty}\tau_{b_m}$, $\{\tau<\tau_\infty\}=\bigcup_{m=1}^\infty\{\tau<\tau_{b_m}\}$. 因此, 由概率的从下连续性得到

$$\mathrm{P}_x(\tau<\tau_\infty)=\lim_{m\to\infty}\mathrm{P}_x(\tau<\tau_{b_m}).$$

注意到方程 (6.3.23) 在 $[0,\infty)$ 存在唯一, 即 $\mathrm{P}_x(\tau_\infty=\infty)=1$, 由此得到 $X(t)$ 的灭绝概率

$$\mathrm{P}_x(\tau<\infty)=\begin{cases}\mathrm{e}^{-\frac{2\alpha}{\sigma^2}x} & \alpha>0\\ 1 & \alpha\leq 0\end{cases}.$$

□

6.3.3 首出时的分布和向后方程

需要重写

6.4 Feynman-Kac 公式

6.4.1 背景

6.4.2 Feynman-Kac 公式 I (时间齐次情形)

定理 6.4.1. 设

$$\mathrm{d}\xi_t=\mathbf{b}(\xi_t)\mathrm{d}t+\Sigma(\xi_t)d\mathbf{B}_t, \tag{6.4.1}$$

记 $\mathcal{L}=\frac{1}{2}\sum_{i,j=1}^n a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^n b_i(x)\frac{\partial}{\partial x_i}$, $(a_{ij})_{i,j=1,\cdots,n}=\Sigma\Sigma^{\mathrm{t}}$. 设 $f\in C_0^2(\mathbb{R}^n)$, $q(x)\in C(\mathbb{R}^n)$, $q(x)$ 有下界.

1. 设 $v(t,x)=\mathrm{E}\Big(f(\xi_t)\exp\big(-\int_0^t q(\xi_s)\mathrm{d}s\big)\Big|\xi_0=x\Big)$, 则

$$\begin{cases}\frac{\partial v}{\partial t}=\mathcal{L}v-qv & t>0,\ x\in\mathbb{R}^n\\ v(0,x)=f(x) & x\in\mathbb{R}^n\end{cases}. \tag{6.4.2}$$

2. 若 $w(t,x)\in C^{1,2}(\mathbb{R}\times\mathbb{R}^n)$ 在 $K\times\mathbb{R}^n$ 上有界, $K\subset\subset\mathbb{R}$, w 为 (6.4.2) 的解, 则 $w(t,x)=v(t,x)$.

当 $q(x)=0, x\in\mathbb{R}^n$ 时, 上述定理实质是时间齐次扩散过程的 Kolmogorov 向后方程. 类似于时齐的向后方程, 我们可以形式地验证 (6.4.2).

(1) $v(t,x)$ 表示 $(\xi_t)_{t\geq 0}$ 从 0 时刻, x 位置出发, 在时间区间 $[0,t]$ 内, 量 $f(\xi_t)\exp(-\int_0^t q(\xi_s)\mathrm{d}s)$ 的均值. 由时间齐次性, $v(t,x)$ 实质表示 $(\xi_t)_{t\geq 0}$ 从 s 时刻, x 位置出发, 在时间区间 $(s,s+t]$ 内, 量 $f(\xi_{s+t})\exp(-\int_s^{s+t} q(\xi_s)\mathrm{d}s)$ 的均值.

(2) 因此, $v(t,\xi_\Delta)$ 表示过程从 Δ 时刻, 位置 ξ_Δ 出发, 在 $(\Delta,\Delta+t]$ 内 $f(\xi_{\Delta+t})\exp(-\int_\Delta^{\Delta+t} q(s)\mathrm{d}s)$ 的均值. 具体地有如下性质.

性质 6.4.1. $\mathrm{E}\Big[f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta} q(\xi_s)\mathrm{d}s\big)\Big|\mathcal{F}_\Delta^{\mathbf{B}}\Big]=v(t,\xi_\Delta)$.

验证留作习题.

(3) 将 $v(\Delta+t,x)$ 分成两步计算, 先从 0 时刻, x 位置出发到达 ξ_Δ, 再从时刻 ξ_Δ 出发到达 $\xi_{t+\Delta}$. 由马氏性和时间齐次性有如下性质.

性质 6.4.2. $v(t+\Delta,x)=\mathrm{E}[\exp\big(-\int_0^\Delta q(\xi_s)\mathrm{d}s\big)v(t,\xi_\Delta)\big|\xi_0=x]$.

验证留作习题.

(4) 利用以上的性质有

$$\begin{aligned}
& v(t+\Delta,x)-v(t,x)\\
=\ & \mathrm{E}\big[\exp\big(-\int_0^\Delta q(\xi_s)\mathrm{d}s\big)v(t,\xi_\Delta)-\exp\big(-\int_0^t q(\xi_s)\mathrm{d}s\big)f(\xi_t)|\xi_0=x\big]\\
=\ & \mathrm{E}\big[\exp\big(-\int_0^\Delta q(\xi_s)\mathrm{d}s\big)v(t,\xi_\Delta)-v(t,\xi_\Delta)+v(t,\xi_\Delta)-\exp\big(-\int_0^t q(\xi_s)\mathrm{d}s\big)f(\xi_t)|\xi_0=x\big]\\
=\ & \mathrm{E}\big[\exp\big(-\int_0^\Delta q(\xi_s)\mathrm{d}s\big)\big(v(t,\xi_\Delta)-1\big)|\xi_0=x\big]+\mathrm{E}\big[v(t,\xi_\Delta)-v(t,x)|\xi_0=x\big]\\
=\ & I+II.
\end{aligned}$$

对上式两端同乘以 $\frac{1}{\Delta}$, 再令 $\Delta\downarrow 0$, 得到

$$\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\big(v(t+\Delta,x)-v(t,x)\big)=\frac{\partial v(t,x)}{\partial t},$$

$$\lim_{\Delta\downarrow 0}\frac{1}{\Delta}I=-q(x)v(t,x),$$

先对 $v(t,\xi_\Delta)-v(t,x)$ 用 Itô 公式, 再求如下极限

$$\lim_{\Delta\downarrow 0}\frac{1}{\Delta}II=\mathcal{L}v(t,x).$$

引理 6.4.1. 设 $v(t,x)=\mathrm{E}\Big(f(\xi_t)\exp\big(-\int_0^t q(\xi_s)\mathrm{d}s\big)\Big|\xi_0=x\Big)$, 则

$$\mathrm{E}\Big(f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta} q(\xi_s)\mathrm{d}s\big)\Big|\mathcal{F}_\Delta^{\mathbf{B}}\Big)=v(t,\xi_\Delta).$$

引理证明: 证明留作习题 □

定理 6.4.2. $D\subset\mathbb{R}^n$ 为一个有界开区域, $\mathcal{L}$ 的系数连续且在区域 D 上一致椭圆, 即存在 $\delta>0$, 使得 $\sum_{1\leq i,j\leq n}a_{ij}\eta_i\eta_j\geq\delta\|\eta\|^2$ 成立. $q(x)$ 在区域 D 上连续, 而且下有界, g 是 D 上的连续函数. 若下面的偏微分方程边值问题

$$\begin{cases}\frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i\partial x_j}+\sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i}-q(x)u=-g(x) & x\in D\\ u(x)=\varphi(x) & x\in\partial D\end{cases}\tag{6.4.3}$$

具有直到边界 ∂D 的二阶连续可导的解 $u(x)$, 那么 $u(x)$ 有如下的概率表示:

$$u(x)=\mathrm{E}\Big(\varphi(\xi_\tau)\mathrm{e}^{-\int_0^\tau q(\xi_s)\mathrm{d}s}\Big|\xi_0=x\Big)+\mathrm{E}\Big(\int_0^\tau g(\xi_t)\exp\Big(-\int_0^t q(\xi_s)\mathrm{d}s\Big)\mathrm{d}t\Big|\xi_0=x\Big),\tag{6.4.4}$$

其中

$$d\xi_t = \mathbf{b}(\xi_t)dt + \Sigma(\xi_t)d\mathbf{B}_t,\quad (a_{ij}) = \Sigma\Sigma^{\mathrm{t}}.$$

$$\tau = \inf\{t \geq 0,\ \xi_t \notin \overline{D}\}.$$

取 $\varphi(x) = 0$, $q(x) \geq 0$, 考虑如下偏微分方程的 Dirichlet 边值问题

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i} - q(x)u = -g(x) & x \in D \\ u(x) = 0 & x \in \partial D \end{cases},$$

则

$$u(x) = \mathrm{E}\Big(\int_0^\tau g(\xi_t)\exp\Big(-\int_0^t q(\xi_s)\mathrm{d}s\Big)\mathrm{d}t\Big|\xi_0 = x\Big). \tag{6.4.5}$$

利用 (6.4.5) 可以给出 Feynman-Kac 公式的概率含义. 想象一个粒子按照扩散过程 (6.4.1) 的轨道作运动, 在一个小的时间段 $[t, t+h)$ 以概率 $q(\xi_t)h + o(h)$ 被“斩杀”. 将时间段 $[0, t]$ 等分成 n 份,

$$0 = t_0 < t_1 < \cdots < t_n,\ \text{取}\ h = t_{i+1} - t_i = \frac{t}{n},$$

那么, 到达时刻 t 时, 粒子还存活的概率近似地为

$$\begin{aligned} \prod_{k=1}^{n-1}\big(1 - q(t_k)h\big) &= \exp\Big(\sum_{k=1}^{n-1}\log(1 - q(\xi_{t_k})h)\Big) \\ &\approx \exp\Big(-\sum_{k=1}^{n-1} q(\xi_{t_k})h\Big) \\ &\xrightarrow{n\to 0} \exp\Big(-\int_0^t q(\xi_s)\mathrm{d}s\Big). \end{aligned}$$

因此, (6.4.5) 中的 $u(x)$ 可以理解成粒子们沿随机轨道到达边界 ∂D 前还未被斩杀的粒子的轨道函数 $g(\xi_t)$ 值的时间累积的均值.

折损率的解释???

例 6.4.1. (Lévy ArcSine law, 反正弦律)

$(B_t)_{t\geq 0}$ 是布朗运动, 设 $\Gamma(t) = \int_0^t \mathbf{1}_{(0,\infty)}(B_s)\mathrm{d}s$, 求 $\mathrm{P}(\Gamma(t) \leq \theta | B_0 = 0)$, 即 $\Gamma(t)$ 在 $B_0 = 0$ 条件下的分布函数.

解: 记 $\mathrm{P}_x = \mathrm{P}(\cdot | B_0 = x)$, $\mathrm{E}_x = \mathrm{E}(\cdot | B_0 = x)$.

$$\mathrm{P}_0(\Gamma(t) \leq \theta) = \frac{2}{\pi}\arcsin\Big(\sqrt{\frac{\theta}{t}}\Big),\ 0 \leq \theta \leq t.$$

6.4.3 Feynman-Kac 公式 II (时间非齐次情形)

设 $\mathbf{b}, \Sigma$ 满足全局 Lipschitz 条件和线性增长条件.

$$d\xi_t = \mathbf{b}(t, \xi_t)dt + \Sigma(t, \xi_t)d\mathbf{B}_t, \tag{6.4.6}$$

$$\mathcal{L}_t = \frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(t,x)\frac{\partial}{\partial x_i} \tag{6.4.7}$$

定理 6.4.3. 设 $u(t,x)$ 关于 x 二次连续可微, 关于 t 一次连续可微且满足 $\forall t \in (0, T]$,

$$\begin{cases} \mathcal{L}_t u(t,x) + \frac{\partial u}{\partial t}(t,x) - q(t,x)u(t,x) = -g(t,x) \\ u(T,x) = f(x) \end{cases}. \tag{6.4.8}$$

则

$$u(t,x) = \mathrm{E}\Big(f(\xi_T)\mathrm{e}^{-\int_t^T q(\theta,\xi_\theta)\mathrm{d}\theta} + \int_t^T g(s,\xi_s)\mathrm{e}^{-\int_t^s q(\theta,\xi_\theta)\mathrm{d}\theta}\mathrm{d}s\Big|\xi_t = x\Big). \tag{6.4.9}$$

形式验证: 以一维扩散为例.

$$\begin{aligned} du(s,\xi_s) &= \Big(\frac{\partial u(s,\xi_s)}{\partial s} + \mathcal{L}_s u(s,\xi_s)\Big)ds + \frac{\partial u(s,\xi_s)}{\partial x}\sigma(s,\xi_s)dB_s \\ &= q(s,\xi_s)u(s,\xi_s)ds - g(s,\xi_s)ds + \frac{\partial u(s,\xi_s)}{\partial x}\sigma(s,\xi_s)dB_s. \end{aligned}$$

与例6.1.2类似, 上式两端同乘以积分因子 $e^{-\int_t^s q(\theta,\xi_\theta)d\theta}$, 利用 Itô 公式,

$$\begin{aligned} d\big(u(s,\xi_s)e^{-\int_t^s q(\theta,\xi_\theta)d\theta}\big) &= -q(s,\xi_s)e^{-\int_t^s q(\theta,\xi_\theta)d\theta}u(s,\xi_s)ds + e^{-\int_t^s q(\theta,\xi_\theta)d\theta}du(s,\xi_s) \\ &= e^{-\int_t^s q(\theta,\xi_\theta)d\theta}\Big(-g(s,\xi_s)ds + \frac{\partial u(s,\xi_s)}{\partial x}\sigma(s,\xi_s)dB_s\Big) \end{aligned}$$

将起点时刻 t 和终点时刻 T 代入上式,

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

上式两边取 $E\big(\cdot|\xi_t = x\big)$ 即得.

例 6.4.2. (Black-Scholes 模型的欧式未定权益的定价公式)

设某支风险证券的价格满足如下的 Black-Scholse 方程

$$dS_t = rS_t dt + \sigma S_t dB_t. \tag{6.4.10}$$

其中 r 为银行的利率, 也就是说 (6.4.10) 是风险中性模型; $(B_t)_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 布朗运动.

设 $f(S_T)$ 表示 T 时刻的收益, 即未定权益; $v(t,S_t)$ 为价格函数, 即未定权益 $f(S_T)$ 在 $t<T$ 时刻的价格, 则 $v(t,S_t)$ 应该满足:

1. $v(t,S_t)$ 在 $t=T$ 时刻就是 $f(S_T)$;
2. $f(S_T)e^{-r(T-t)}$ 是 $f(S_T)$ 在经历 $[t,T]$ 时间段后 $f(S_T)$ 乘以"损耗 (折现)"的价值, 为确保公平性 (无套利), 对它在 t 时刻关于 $\mathcal{F}_t$ 做预测就应该是它此刻的定价 $v(t,S_t)$, 也就是说 $v(t,S_t)$ 的折现价格 $Y_t = e^{-rt}v(t,S_t)$ 应该是鞅, 即 $E[Y_T|\mathcal{F}_t] = Y_t$. 因此得到,
$$E[f(S_T)e^{-r(T-t)}|\mathcal{F}_t] = v(t,S_t). \tag{6.4.11}$$

♣♣♣20230530

由 $(S_t)_{t\geq 0}$ 的马氏性得到,

$$E[f(S_T)e^{-r(T-t)}|\mathcal{F}_t] = E[f(S_T)e^{-r(T-t)}|S_t] = E[f(S_T)e^{-r(T-t)}|S_t = x]\big|_{x=S_t}.$$

这说明

$$v(t,x) = E\big(f(S_T)e^{-r(T-t)}|S_t = x\big). \tag{6.4.12}$$

由 Feynman-Kac 公式, $v(t,x)$ 满足如下方程

$$\begin{cases} \frac{\partial v(t,x)}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v(t,x)}{\partial x^2} + rx\frac{\partial v(t,x)}{\partial x} - rv(t,x) = 0 \\ v(T,x) = f(x) \end{cases}. \tag{6.4.13}$$

我们可以利用 (6.4.12) 和例6.2.2中解得的 S_t 的分布函数 (6.2.7), 给出偏微分方程 (6.4.13) 的显式解

$$v(t,x) = \frac{e^{-r(T-t)}}{\sqrt{2\pi}}\int f\big(xe^{(r-\frac{\sigma^2}{2})(T-t)+\sigma\sqrt{T-t}z}\big)e^{-\frac{-z^2}{2}}dz. \tag{6.4.14}$$

上面的推导中 $e^{-rt}v(t,S_t)$ 的鞅性实质依赖于 (6.4.2) 是风险中性的, 即 $(e^{-rt}S_t,\mathcal{F}_t)_{t\geq 0}$ 是鞅.

但当 $(S_t)_{t\geq 0}$ 不是风险中性时, $e^{-rt}S_t$ 就不是鞅, 这是就需要构造一个新的概率测度 $\widetilde{P}$, 使得 $e^{-rt}S_t$ 在 $\widetilde{P}$ 成为鞅, 在 $\widetilde{P}$ 下预测 $f(S_T)e^{-r(T-t)}$,

$$\widetilde{E}[f(S_T e^{-r(T-t)})|\mathcal{F}_t] = \widetilde{E}[f(S_T e^{-r(T-t)})|S_t] = v(t,x)|_{x=S_t}, \tag{6.4.15}$$

此时的 $v(t,x)$ 才是所求的未定权益.

(6.4.13) 就是欧式未定权益的 Black-Scholes 偏微分方程, 它与风险证券的平均收益率无关, 只与银行利率和风险证券的波动率有关.

6.5 扩散过程的长时间行为简介和例子

6.5.1 基本理论

定理 6.5.1. （Birkhoff 个别遍历定理）

设 n 维时齐扩散过程

$$\mathrm{d}\boldsymbol{\xi}_t = \mathbf{b}(\boldsymbol{\xi}_t)\mathrm{d}t + \Sigma(\boldsymbol{\xi}_t)d\mathbf{B}_t,$$

记 $\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial}{\partial x_i}$, $(a_{ij})_{i,j=1,\cdots,n} = \Sigma\Sigma^{\mathrm{t}}$. 若存在 $0 < r_1 < r_2$, 使得

$$r_1 I_n \le \Sigma\Sigma^{\mathrm{t}} \le r_2(1+|x|^2)I_n,\ I_n\text{为单位矩阵},$$

而且

$$\mathcal{L}^*\phi(x) = \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\Big(a_{ij}(x)\phi(x)\Big) - \sum_{i=1}^n \frac{\partial}{\partial x_i}\Big(b_i(x)\phi(x)\Big) = 0,$$

有一个非负非零的可积解, 即 $\phi(x) > 0, x \in \mathbb{R}^n$, $\int_{\mathbb{R}^n}\phi(x)\mathrm{d}x < \infty$, 不妨记 $\int_{\mathbb{R}^n}\phi(x)\mathrm{d}x = 1$, 则概率测度 $\pi(\cdot) = \int_\cdot \phi(x)\mathrm{d}x$ 是 $\boldsymbol{\xi}_t$ 的不变概率测度, 并且对于任意的有界 Borel 函数 f 或非负 Borel 函数满足 $\int_{\mathbb{R}^n} f(x)\mathrm{d}x < \infty$, 对于任意的初值随机变量 $\boldsymbol{\xi}_0$,

$$\mathrm{P}\Big(\lim_{T\to\infty}\frac{1}{T}\int_0^T f(\boldsymbol{\xi}_t)\mathrm{d}t = \int_{\mathbb{R}^n} f(x)\phi(x)\mathrm{d}x\Big) = 1.$$

定理 6.5.2. (平均遍历定理)

条件同如上的 Birkhoff 个别遍历定理6.5.1, 对于任意的有界 Borel 函数 f 或非负 Borel 函数满足 $\int_{\mathbb{R}^n} f(x)\mathrm{d}x < \infty$, 对于任意的初值随机变量 $\boldsymbol{\xi}_0$,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{E}\big(f(\boldsymbol{\xi}_t)\big)\mathrm{d}t = \int_{\mathbb{R}^n} f(x)\phi(x)\mathrm{d}x.$$

定理 6.5.3. （渐近稳定性）

条件同如上的 Birkhoff 个别遍历定理6.5.1, 对于任意的有界连续函数 f, 对于任意的 $x \in \mathbb{R}^n$,

$$\lim_{t\to\infty}\mathrm{E}\big(f(\boldsymbol{\xi}_t)|\boldsymbol{\xi}_0 = x\big) = \int_{\mathbb{R}^n} f(x)\phi(x)\mathrm{d}x.$$

6.5.2 应用的例子

例 6.5.1. 随机梯度下降 (stochastic gradient descent), 模拟退火算法 (simulated annealing algorithm) 的扩散过程版本

设 U 是 $\mathbb{R}^n$ 上的一个二阶光滑的势函数, 满足 $U(x) \ge 0$, $\lim_{|x|\to\infty}U(x) = \infty$, $\nabla U(x)$ 是整体 Lipschitz 连续的, 对于任意 $\varepsilon > 0$, $Z_\varepsilon := \int_{\mathbb{R}^n}\mathrm{e}^{-\frac{2U(x)}{\varepsilon}}\mathrm{d}x < \infty$, 并假设 U 的局部极小值点只有有限个. 考虑如下的扩散过程:

$$\mathrm{d}X_t^\varepsilon = -\nabla U(X_t^\varepsilon)\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}B_t, \tag{6.5.1}$$

其中, $(B_t)_{t\ge 0}$ 是 n 维布朗运动.

第一步 先考虑无随机扰动的梯度下降系统

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = -\nabla U(X_t). \tag{6.5.2}$$

如果 $\widetilde{x}$ 是 U 的某个局部极小点, $A(\widetilde{x})$ 是 $\widetilde{x}$ 的吸引域, 即任意的 $x \in A(\widetilde{x})$, $X(t)$ 从 x 出发, $\lim_{t\to\infty}X_t = \widetilde{x}$. 对于不同的 $\widetilde{x}, \widetilde{x}$ 的吸引域不同.

第二步 对于任意的 $0 < T < \infty$, 若 $X_0 = X_0^\varepsilon = x$, 则任给 $\delta > 0$,

$$\lim_{\varepsilon\to 0}\mathrm{P}\Big(\sup_{t\in[0,T]}|X_t^\varepsilon - X_t| < \delta\Big) = 1.$$

第三步 设 $\pi^\varepsilon(\cdot)=\frac{1}{Z_\varepsilon}\int_\cdot \mathrm{e}^{-\frac{2U(x)}{\varepsilon}}\mathrm{d}x$, 则它是 $(X_t^\varepsilon)_{t\geq 0}$ 的不变概率测度.

第四步 由 Birkhoff 个别遍历定理6.5.1, 对于 $\mathbb{R}^n$ 上的任意 Borel 集 K,

$$\mathrm{P}\Big(\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{1}_K(X_t^\varepsilon)\mathrm{d}t=\pi^\varepsilon(K)\Big)=1.$$

特别地, 对于任意的开集 O, $\pi^\varepsilon(O)>0$, 这说明 $(X_t)_{t\geq 0}$ 是正常返的. 这点与无扰动系统 (6.5.2) 不同. 进一步, 由稳定性定理6.5.3, 对于任意的有界连续函数 f,

$$\lim_{t\to\infty}\mathrm{E}\big(f(X_t^\varepsilon)|X_0=x\big)=\int_{\mathbb{R}^n}f(x)\pi^\varepsilon(x)\mathrm{d}x.$$

特别地, 设 ∂K 为 K 边界集, 若 ∂K 的 Lebesgue 测度为零[①], 则对于任意的 $x\in\mathbb{R}^n$,

$$\lim_{t\to\infty}\mathrm{E}\big(X_t^\varepsilon\in K|X_0=x\big)=\pi^\varepsilon(K). \tag{6.5.3}$$

第五步 π^ε 的渐近行为. 设 $A=\{x:U(x)=\inf_{y\in\mathbb{R}^n}U(y)\}$, 即 A 是 U 的全局最小值点集, O_δ 为 A 的 δ 领域.

性质 6.5.1. $\lim_{\varepsilon\to 0}\pi^\varepsilon(O_\delta)=1$.

第六步 当 ε 很小时, X_t^ε 的长时间行为:

$$\frac{1}{t}\int_0^t \mathbf{1}_{O_\delta}(X_s^\varepsilon)\mathrm{d}s\xrightarrow{t\to\infty}\pi^\varepsilon(O_\delta)\approx 1,\ a.s.\mathrm{P}.$$
$$\mathrm{E}\big(X_t\in O_\delta|X_0=x\big)\xrightarrow{t\to\infty}\pi^\varepsilon(O_\delta)\approx 1.$$

但需要强调 $t\to\infty$ 和 $\varepsilon\to 0$ 是不能交换顺序的.

第七步 因此, 考虑如下的模拟退火扩散过程

$$\mathrm{d}X_t=-\nabla U(X_t)\mathrm{d}t+\sqrt{\varepsilon(t)}\mathrm{d}B_t.$$

如何精确地设计退火时间表 (freezing schedule) $\varepsilon(t)\to 0$, 使得

$$\lim_{t\to\infty}\mathrm{P}\big(X_t\in O_\varepsilon\big)=1\,?$$

第八步 其它问题, 例如: 最可几路径, 逃逸时间等.

第九步 不同的噪声扰动方式, 可能产生不同的结果.

如 (6.5.1) 中噪声扰动的方法称为加法噪声或加性噪声 (additive noise); 而

$$\mathrm{d}X_t^\varepsilon=-\nabla U(X_t^\varepsilon)\mathrm{d}t+\sqrt{\varepsilon}\sigma(X_t^\varepsilon)\mathrm{d}B_t,$$

噪声扰动的方式被称为乘法噪声或乘性噪声 (multiplicative noise). 我们可以通过下面的简单例子展示一下.

例 6.5.2. 随机 Landau-Stuart 方程

考虑 Landau-Stuart 方程

$$\frac{\mathrm{d}X_t}{\mathrm{d}t}=X_t(\lambda-X_t^2), \tag{6.5.4}$$

其中 λ 是参数. 设 $U(x)=-\frac{\lambda}{2}x^2+\frac{x^4}{4}$, $-U'(x)=x(\lambda-x^2)$, 则 Landau-Stuart 方程是一个梯度下降系统.

当 $\lambda\leq 0$ 时, $U(x)$ 只有一个稳定平衡点 $x^*=0$, $\lim_{t\to\infty}X_t=0$.

当 $\lambda>0$ 时, $U(x)$ 有三个平衡点. $x^*=0$ 是不稳定的, $x^-=-\sqrt{\lambda}$ 和 $x^+=\sqrt{\lambda}$ 是稳定的, 而且 $U(x)$ 在 x^- 和 x^+ 处达到全局最小.

[①] 此时由于 π^ε 关于 Lebesgue 测度绝对连续, $\pi^\varepsilon(\partial K)=0$. 利用概率测度弱收敛理论的 Portmanteau 定理, 就得到 (6.5.3).

若 $x_0 = 0$**，**可以证明对于任意的 $t \geq 0$, $X_t = 0$;

若 $x_0 > 0$**，**可以证明对于任意的 $t \geq 0$, $\lim_{t\to\infty} X_t = \sqrt{\lambda}$;

若 $x_0 < 0$**，**可以证明对于任意的 $t \geq 0$, $\lim_{t\to\infty} X_t = -\sqrt{\lambda}$.

这个系统可以研究两种受随机扰动的方式:

一. 整个系统处于某种环境白噪声中, 可以通过如下有加法噪声的随机微分方程来建模

$$\mathrm{d}X_t = X_t(\lambda - X_t^2)\mathrm{d}t + \sigma \mathrm{d}B_t. \tag{6.5.5}$$

这个方程的长时间行为可以通过上面例子 (6.5.1) 的方式来研究, 不再赘述.

二. 参数 λ 随着时间不断地受到随机涨落的影响或者观测者不能确定参数的实际值, 即 λ 可以看成 $\lambda + \sigma \dot{B}_t$, 这种情况下可以通过如下有乘法噪声的随机微分方程来建模

$$\mathrm{d}X_t = X_t(\lambda - X_t^2)\mathrm{d}t + \sigma X_t \mathrm{d}B_t. \tag{6.5.6}$$

当 $\lambda \leq 0$ **时，**可以证明 $\lim_{t\to\infty} X_t = 0$.

当 $\lambda > 0$ **时，**我们与无随机扰动的系统 (6.5.4) 的长时间行为进行比较.

若 $X_0 = 0$**，**对于任意的 $t \geq 0$, $X_t = 0$;

若 $X_0 < 0$**，**可以证明对于任意的 $t \geq 0$, $X_t \leq 0$;

若 $X_0 > 0$**，**可以证明对于任意的 $t \geq 0$, $X_t \geq 0$.

因此我们仅需要讨论 (6.5.6) 在 $[0,\infty)$ 上的长时间行为. 设 $z = \int_0^\infty x^r \mathrm{e}^{-\frac{x^2}{\sigma^2}}\mathrm{d}x$, 其中 $r = 2(\frac{\lambda}{\sigma^2} - 1)$. 当 $r > -1$ 时 (即 $2\lambda > \sigma^2$), $z < \infty$. 此时设

$$p_s(x) = \frac{1}{z} x^r \mathrm{e}^{-\frac{x^2}{\sigma^2}},$$

则它是 X_t 的不变概率密度. 由遍历定理和渐近稳定定理可以得到它的长时间行为的刻画[①]. 当 $\sigma^2 > 2\lambda$ 时, 可以证明 $\lim_{t\to\infty} X_t = 0, a.s.$P.. 当 $\sigma^2 = 2\lambda$, $X_0 > 0$ 时, 可以证明 $\limsup_{t\to\infty} X_t = +\infty, a.s.$P., $\liminf_{t\to\infty} X_t = 0, a.s.$P..[②] 这些结论与无随机扰动 (6.5.4) 明显不同.

例 6.5.3. 分子马达的概率模型[③]

设 $V(x)$ 是周期为 2π 的周期函数. 为简单记, 取势场 $V(x) = \cos x$. 考虑 $\frac{\mathrm{d}X_t}{\mathrm{d}t} = -\sin X_t$. 进一步考虑加外力情形:

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = I - \sin X_t, \quad U(x) = Ix + \cos x, \ I \geq 0. \tag{6.5.7}$$

同时考虑 $\phi_t = X_t \mod 2\pi$, 则 $(\phi_t)_{t\geq 0}$ 是一个取值于 $[0.2\pi)$ 的函数 (或者取值于 S^1 的函数).

当 $I > 1$ **时，**$I - \sin x > 0$, 即 $\frac{\mathrm{d}X_t}{\mathrm{d}t} > 0$, 则 $\lim_{t\to\infty} X_t = \infty$, 并且可以证明

$$\lim_{t\to\infty} \frac{X_t}{2\pi t} = \frac{\sqrt{I^2-1}}{2\pi} > 0.$$

事实上, 可以证明 ϕ_t 是以 $\frac{2\pi}{\sqrt{I^2-1}}$ 为周期的函数, 或者说 ϕ_t 的转动频率是 $\frac{\sqrt{I^2-1}}{2\pi}$.

当 $I = 1$ **时，**$1 - \sin x \geq 0$, $1 - \sin(\frac{\pi}{2} + 2n\pi) = 0$, $n \in \mathbb{Z}$. 因此, $\forall x \in (2n\pi + \frac{\pi}{2}, 2(n+1)\pi + \frac{\pi}{2}]$, $X_0 = x$, $\lim_{t\to\infty} X_t = 2(n+1)\pi + \frac{\pi}{2}$, 也就是说 $\lim_{t\to\infty} \phi_t = \frac{\pi}{2}$, $\lim_{t\to\infty} \frac{X_t}{2\pi t} = 0$.

[①] 虽然此时方程的系数并不满足定理6.5.1、定理6.5.2和定理6.5.3的条件, 但可以证明它满足这些定理的结论.

[②] 蓝色字体部分的结论可以依据如下参考书第 355 页定理 5.9 来证明.

随机微分方程引论 (第二版) 龚光鲁编著北京大学出版社 1995

[③] 参看 Dai Wang, Shu Zhu and Minping Qian. Rotation number of a system of a single oscillator in definite and white noise perturbed cases. *Commun. Nonlinear Sci. Numer. Simul.* 2 (1997), 91–95.

当 $0 \le I < 1$ 时, 方程 (6.5.7) 有两组平衡点:

- 其中 $2n\pi + \arcsin I$, $n \in \mathbb{Z}$ 是稳定的. 即, 当 $x \in ((2n-1)\pi - \arcsin I, 2n\pi + \arcsin I)$ 时, $I - \sin x > 0$. 因此 $X_0 = x$, $\lim_{t\to\infty} X_t = 2n\pi + \arcsin I$;
  当 $x \in (2n\pi + \arcsin I, 2n\pi - \arcsin I)$ 时, $I - \sin x < 0$. 因此 $X_0 = x$, $\lim_{t\to\infty} X_t = 2n\pi + \arcsin I$;
  特别地, 当 $X_0 = 2n\pi + \arcsin I$ 时, $X_t = 2n\pi + \arcsin I$, $\forall t \ge 0$.
  也就是说, $\lim_{t\to\infty} \phi_t = \arcsin I$.
- 其中 $(2n+1)\pi - \arcsin I$, $n \in \mathbb{Z}$ 是不稳定的, 当 $X_0 = (2n+1)\pi - \arcsin I$ 时, $X_t = (2n+1)\pi - \arcsin I$, $\forall t \ge 0$.
  也就是说, $\lim_{t\to\infty} \phi_t = \pi - \arcsin I$.

总之, $\lim_{t\to\infty} \frac{X_t}{2\pi t} = 0$.

考虑受到热噪声影响的系统

$$\mathrm{d}X_t = (I - \sin X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.$$

考虑 $\phi_t = X_t \mod 2\pi$, 则此时 $(\phi_t)_{t\ge 0}$ 是一个以 $[0,2\pi)$(或者 S^1) 为状态空间的马氏过程.

虽然, X_t 的不变概率测度不存在, 但 ϕ_t 的不变概率测度存在, 其不变概率密度函数

$$\pi(x) = \frac{1}{z} w(x)\Big(\int_0^x \frac{2}{\sigma^2 w(y)}\mathrm{d}y + \int_x^{2\pi} \frac{2k}{\sigma^2 w(y)}\mathrm{d}y\Big).$$

其中, $w(x) = \exp\big(\frac{1}{\sigma^2}(Ix + \cos x - 1)\big)$, $k = \exp\big(\frac{4\pi I}{\sigma^2}\big)$, z 是归一化常数[①].

进一步, 由遍历定理对于任意的 $[0,2\pi)$ 上的有界可测函数 $f(x)$,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\phi_s)\mathrm{d}s = \int_0^{2\pi} f(x)\pi(x)\mathrm{d}x.$$

特别地, 对于任意的 $I > 0$, $\sigma > 0$ 均有

$$\begin{aligned}
\frac{1}{2\pi t}X_t &= \frac{1}{2\pi t}X_0 + \frac{1}{2\pi t}\int_0^t \big[I - \sin\big(X(s)\big)\big]\mathrm{d}s + \sigma\frac{B_t}{2\pi t} \\
&= \frac{1}{2\pi t}\int_0^t \big[I - \sin\big(\phi(s)\big)\big]\mathrm{d}s + \sigma\frac{B_t}{2\pi t} \\
& \text{关于}\phi_t\text{用遍历定理; 关于}B_t\text{用强大数律} \\
&\xrightarrow{t\to\infty} \frac{1}{2\pi}\int_0^{2\pi} \big(I - \sin x\big)\pi(x)\mathrm{d}x \\
&= \frac{1}{z}\big(\mathrm{e}^{\frac{4\pi I}{\sigma^2}} - 1\big) > 0.
\end{aligned}$$

这点与无随机扰动时的结果不一样. 从物理上解释是将热噪声转化成了机械能, 将无序运动转化成了有序运动.

[①] 对于一般的扩散过程,

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

若系数 $b(x)$ 和 $\sigma(x)$ 以 2π 为周期, 则 $\phi_t = X_t \mod 2\pi$ 可以看成是 $[0,2\pi)$ 或 S^1 上的扩散过程, 其扩散算子为

$$\mathcal{L} = \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x},\ x \in [0,2\pi)\text{或}S^1.$$

如果 $\sigma(x) > 0$, 则 ϕ_t 的不变概率密度为

$$\pi(x) = \frac{1}{z}w(x)\Big(\int_0^x \frac{2}{\sigma(y)^2 w(y)}\mathrm{d}y + \int_x^{2\pi} \frac{2k}{\sigma(y)^2 w(y)}\mathrm{d}y\Big).$$

其中 $w(x) = \exp\big(\int_0^x \frac{2(I-\sin x)}{\sigma(x)^2}\mathrm{d}y\big)$, $k = w(2\pi)$, z 是归一化常数.