随机分析九讲

应坚刚
复旦大学数学系

2012 年 6 月 21 日

序 言

这是为复旦大学数学学院新招收的金融工程专业硕士随机分析课写的讲义, 这是第二次上这个课, 感觉不好讲, 因为学生的数学基础参差不齐, 半数学生只学过微积分和工科概率论, 所以不可能按照数学系的要求来给学生上课, 也不可能象写数学教材那样定义定理证明那种形式来写讲义. 第一次讲下来, 我对上什么内容有了些想法, 这本讲义完全是按照上课的要点来写的, 尽量讲得直观浅显, 实际上是先讲课后写讲义, 然后再进行完善. 我自己总结了此讲义的特色是: 讲概念, 但不太正式, 讲证明, 但不求严格. 说实在的, 我十多年来一直为复旦数学专业的研究生讲授从测度论出发的随机分析, 写这样一本不涉及测度论的随机分析讲义对我的观念和能力来说都是一个挑战. 但是我相信这样一本书是有市场的, 对需要理解随机分析而又没有多少数学基础的金融从业人员和对金融数学有兴趣的人来说看得明白且有帮助的, 至少可以让他们部分地明白随机分析包含些什么内容, 为什么随机分析在金融中是有用的.

为这个选取内容是另外一个问题, 我的想法是化两节课复习概率论的一些基本内容, 比如随机变量与分布, 期望与方差, 正态分布, 大数定律, 中心极限定理以及条件数学期望等. 然后引入离散鞅, 用最浅显的语言叙述 Doob 的离散随机积分定理, 接着介绍二叉树定价模型, 这些内容简单但富有思想. 后五讲主要介绍 Brown 运动, 讨论其轨道性质以及关于 Brown 的随机积分, 著名的 Ito 公式在第九讲自然地引入, 最后还会介绍鞅表示定理. 这样基本上涵盖了随机分析的基本内容.

1. 分析: 通常指极限微积分等关于函数形态的分析理论, 始于 300 年前牛顿和 Leibniz 的工作.

2. 随机分析: 可以说是概率空间上的分析学, 或者说是随机函数的分析学, 始于 1940 年代 Ito 的工作.

3. 金融随机分析: 与金融定价理论相关的随机分析, 始于 1973 年 Black 和 Scholes 发表的期权定价公式.

4. 本课程内容: 与金融定价理论相关的简单随机分析, 包括离散鞅论与基于布朗运动的微积分.

5. 本课程目的: 使学生基本理解金融定价中为什么要用到随机分析以及怎么用随机分析, 也就是金融随机分析的基本思想与方法.

6. 要求的数学基础: 高等数学和概率论.

本讲义里的许多结果的叙述都不是很严格, 读者会有两方面的反应, 对部分读者来说, 这些内容还是太抽象, 理论证明还是太多太难理解, 对这部分读者只能说抱歉了, 我只能做到这个程度, 一点不写证明的随机分析书与我的理念相违. 对另外一部分读者来说, 没有严格证明的结论是很难接受的, 若果是如此, 那自然是值得高兴的, 请这部分读者参考我写的研究生教材 [4], 那里有严密的证明. 感谢2011 级的汪敏同学为大多数习题做了解答.

应坚刚
2011 年秋于复旦大学

目　录

第 一 讲　预备知识

摘要 本章开始主要是讨论随机现象的直观意思, 这对于理解概率论或者随机分析能做什么不能做什么以及怎么做都非常要紧, 然后我们复习概率论的一些基本概念, 如随机变量的分布函数与数学期望, 数学期望的性质, 一些常用的分布, 如二项分布, 均匀分布, 指数分布等, 最重要的是正态分布.

关键词
- 随机现象
- 概率空间
- 客观与主观概率
- 随机变量与分布函数
- 期望
- 正态分布

1. 金融为什么与随机分析有关系? 因为金融中有许多随机现象, 最常见的就是股票价格的走势.

2. 随机现象或随机试验: 指无法预测结果的现象或试验. 如掷骰子, 掷硬币, 彩票, 人的寿命, 股票和天气等等. 几乎每天人们都会碰到随机现象, 可以说, 随机现象无处不在.

3. 是否真的有随机现象? 这是一个很难回答的问题, 科学界对这个问题也有争议, A. Einstein 说: 上帝不掷骰子. 也就是他认为没有随机现象. 关于这个问题的答案更多的是反映一种哲学观点, 因为仔细思考一下的话, 你会发现很多现象之所以无法预测是因为我们的计算能力不够或者对自然概率认识的不够或者观察的时间不够, 比如天气, 我相信明天的天气现在应该已经确定了, 硬币落地时是正面还是反面在它离开手的瞬间也应该已经确定了, 只是系统太复杂, 我们无法把它精确地算出来. 我们这里不关心是不是真的有随机现象的问题, 只关心在需要确定的时候我们能不能确定什么会发生, 如果不能, 就是随机现象.

4. 气候变暖: 究竟有没有气候变暖的问题? 其实气候是一个大尺度的事情, 时间的尺度可以追溯到地球诞生的时候甚至更早. 但我们关于气候的记录充其量不过几百年, 以这样的尺度看问题, 我们很难确定这些年的气候变暖是瞬间的随机变化还是一种趋势. 比如说我们不能因为掷出两次六点就断言说这个骰子每面都是六.

5. 概率论: 概率论是研究随机现象的一种方法. 但是要注意的是, 概率论只是一种无奈的办法, 它不能本质上解决随机现象不能预测的问题, 比如说不要期望一个概率学家能够算出彩票的中奖号码. 概率论真正有点用处通常是在大样本的场合, 它能够告诉我们统计意义上的某种趋势, 但关于个体几乎是毫无意义的.

6. 随机分析学什么: 我认为随机分析最重要的是对于鞅的感觉, 学生应该通过定义定理习题等手段培养起对于鞅的直观认识, 只有真正理解了鞅, 才能真正理解随机分析以及它在金融领域的应用.

7. 什么是 e? e 是一个非常重要的量, 它在随机分析中也是无处不在.

$$\mathrm{e} = \lim_{n\to\infty}(1+1/n)^n.$$

设银行的年利率是 $r>0$, 如果每年结一次利, 那么一年后 1 元钱变成 $1+r$ 元; 如果每月结利, 那么一年后变成 $(1+r/12)^{12}$; 如果每天结利, 那么一年后是 $(1+r/365)^{365}$; 因此如果连续结利, 那么 1 元钱一年后是 e^r 元. 如果 r 元钱存入银行一个时间段后变成 1, 那么我们说 r 是这个时间段的折现率, 折现率就是利息率的倒数.

8. 概率空间: 谈论某个随机现象时, 所有可能结果的全体称为样本空间, 记为 Ω. 例如, 掷硬币的样本空间是 $\{$ 正面, 反面 $\}$; 掷骰子的样本空间是 $\{1,2,3,4,5,6\}$. 概率是一个数量, 取值在 0,1 之间, 通常我们说是事件的概率, 所以我们用一个符号 $\mathscr{F}$ 表示事件的集合, 称为事件域, 一个事件是指样本空间的一个子集, 样本空间中的元素称为基本事件, 例如从一副牌里面摸一张牌, '摸到一个草花' 就是一个事件, 它作为集合是所有草花的全体. 在学概率的时候, 事件域不是那么重要, 但在随机分析中事件域是非常重要的, 它通常理解为信息. 事件用 A 表示, 那么 $\mathbb{P}(A)$ 表示 '事件 A 发生的概率'. 概率最重要的性质是可加性: 如果事件 A,B 不同时发生, 那么其中之一发生的概率是两个概率的和, 即
$$\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B).$$
特别地, 事件 A 发生的概率与不发生的概率之和为 1, 即
$$\mathbb{P}(A^c)=1-\mathbb{P}(A).$$
不仅如此, 数学上更重要的是可列可加性: 如果 $A_1,\cdots,A_n,\cdots$ 两两不同时发生, 那么
$$\mathbb{P}(A_1\cup\cdots\cup A_i\cup\cdots)=\mathbb{P}(A_1)+\cdots+\mathbb{P}(A_i)+\cdots, \tag{1.1}$$
这三个符号放在一起 $(\Omega,\mathscr{F},\mathbb{P})$ 就称为是一个概率空间, 然后我们就可以量化地谈论概率问题并且计算一些有趣事件的概率了.

9. 概率和主观概率: 为了数学上有意义, 也就是后面说的大数定律之下有意义, 我们要求随机试验是可以独立地重复的, 比如掷骰子, 掷硬币, 摸牌等. 显然很多随机现象是不能重复的, 比如明天上海的天气, 每个特定人的寿命等等. 不能重复的随机现象中所谈论的概率很难说有什么具体的意义, 它无非是体现说话者的一种主观推测, 所以称为主观概率 (相对应的, 数学意义上的概率

称为客观概率). 比如有人说: ‘奥巴马的寿命活到 90 岁的概率是 85%’, ‘明天上海下雨的概率是 20%’, 我们对这些话都无法说对也无法说错. 但是如果概率论只能讨论可以重复的现象, 那么应用的范围未免太窄. 很多情况下, 我们可以把一个不能重复的现象模糊地进行重复来应用概率的思想方法. 比如在讨论坏境对人的影响时, 一个特定的人不能重复, 但我们可以把其他在类似坏境下的人当做某个人的重复; 在讨论某只股票走势的时候把它在以往类似情况下的走势当做它的重复等等, 但是这样得到的结果是统计意义下的结果, 这些结论不能用于谈论任何一个特定的人. 人们常常容易无意识地 (或者为了娱乐) 反其道为之, 比如说某个人是什么什么血型, 所以有什么什么性格等, 当然也不排除一些无良的官员和科学家会故意地这么做以误导大众而达到不可告人的目的.

10. 确切地说: 任意选择一个中国人, 他可以活过 80 岁的概率是客观概率, 而具体某个人可以活过 80 岁的概率是主观概率, 人们在讨论类似问题的时候应该注意两者的根本区别. 本书下面在举例的时候也许不特别地去区分可以重复和不可重复的随机现象, 但是读者应该心里有数.

11. 随机变量: 随机变量是样本空间上的函数. 例如人的寿命, 明天的雨量, 掷骰子的点数, 早上上班路上的时间等等. 随机变量是我们经常遇到的事情. 随机变量可以看成是量化的随机现象, 它的取值是随机的. 有的随机变量的量化值的大小是有意义的, 比如寿命, 有的量化仅仅是符号化, 比如掷硬币的正反面分别用 0, 1 表示成为随机变量, 其大小是没有意义的. 随机变量通常用字母 X, Y 等表示. 同一个随机现象中因为关注的问题不同而有不同的随机变量, 比如从一副牌中取 10 张牌, 有人关心对子的个数 X, 有人关心黑桃的个数 Y, 这样就有不同的随机变量. 这些随机变量之间有关系, 比如说黑桃个数多了, 感觉对子就少了, 当然这种关系不是确定的, 只是一种相关性.

12. 随机变量的分布: 当然随机变量的重要特性是它的分布, 分布这个字几乎是不需解释的, 它是指随机变量取值可能性的分布情况, 大多数情况下, 直方图用来表示随机变量的分布情况, 比如考试成绩的分布图, 人口的分布图等等, 但数学上分布是怎么回事呢? 设 X 是一个随机变量, 有两种常用的量来体现它的分布: 分布函数与密度函数.

13. 分布函数: 随机变量 X 的分布函数定义为

$$F(x) = \mathbb{P}(X \leq x),\ x \in \mathbf{R}, \tag{1.2}$$

右边是 X 不超过 x 的概率. 函数 F 是实数集上的函数, 可以证明它是递增且规范的, 我们这里说它是规范的是指

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

显然

$$\mathbb{P}(X \in (a, b]) = F(b) - F(a),\ a, b \in \mathbf{R}, a < b,$$

这是说 X 取值在 a, b 之间的概率是 $F(b) - F(a)$, 体现了分布的本来含义. 当我们看到一个分布函数图象时, 怎么来判断分布情况呢? 因为分布依赖于分布函数的差, 所以图象陡峭的地方分布密集, 反之图象平缓的地方分布稀少. 也可以分布的多少和分布函数的导数成正比.

14. 密度函数: 密度函数可以更直观地体现随机变量的分布. 密度函数定义为分布函数的导数, 即如果 $F(x)$ 是随机变量 X 的分布函数, 那么它的密度函数为 $f(x) = F'(x)$. 显然密度函数总是非负的且

$$\int_{-\infty}^{+\infty} f(x) \mathrm{d}x = 1.$$

通常密度函数本身的大小已经体现 X 的分布情况, 所以它更为直观, 恰如大家熟悉的直方图一样直观.

15. 注意: 不是所有的分布函数都有密度函数, 有密度函数的分布函数称为是连续型的.

16. 几个常用分布:

    (1) 二值分布, 只取两个值的随机变量的分布, 例如掷硬币的结果, 掷骰子时六点是否出现;

    (2) 二项分布, 重复 n 次成功概率为 p 的随机试验其中成功的次数 X, 那么

    $$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

    这里 $\binom{n}{k}$ 是 n 个不同物体中取 k 个的组合数.

(3) 均匀分布: 从区间 $[a,b]$ 上 (均匀) 随机地取个数记为 X, 那么 X 的分布函数是

$$\mathbb{P}(X \le x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a} & a \le x \le b; \\ 1, & x > b, \end{cases}$$

密度函数是

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b); \\ 0, & x \notin (a,b). \end{cases}$$

均匀就体现在其密度函数在 (a,b) 上是常数. 要注意的是, 从 (a,b) 中随机取一个点这件事情只有在数学的世界里能做到, 在现实生活中是做不到的, 但是它可以模拟现实世界, 而且处理起来更简单, 这实际上也是微积分应用的本质.

(4) 指数分布: 随机变量 X 服从指数分布是指其密度函数是 $[0,\infty)$ 上的指数函数

$$f(x) = \begin{cases} \alpha \mathrm{e}^{-\alpha x}, & x \ge 0; \\ 0, & x < 0. \end{cases}$$

其中 $\alpha > 0$ 是参数. 指数分布有遗忘性

$$\mathbb{P}(X > x+y | X > x) = \mathbb{P}(X > y),\ x, y > 0.$$

17. 独立: 独立是非常重要的概念, 随机变量 X, Y 独立, 如果对任何 $x, y \in \mathbf{R}$, 有

$$\mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y), \tag{1.3}$$

也就是说联合分布函数是边缘分布函数的乘积. 如果它们有密度函数, 那么联合密度函数是边缘密度函数的乘积. 直观地说, 独立是指 X 的取值与 Y 的取值无关. 用函数的形式写出来, (1.3) 为

$$\mathbb{E}[1_{(-\infty,x]}(X)1_{(-\infty,y]}(Y)] = \mathbb{E}[1_{(-\infty,x]}(X)] \cdot \mathbb{E}[1_{(-\infty,y]}(Y)].$$

可以证明当上式在 x, y 跑遍所有实数都成立时蕴含着对任何有界 (可测) 函数 f, g 有

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \tag{1.4}$$

记住这个公式, 在处理独立随机变量的时候很有用.

18. 例: 设 $X_1, \cdots, X_n$ 是独立且服从参数为 1 的指数分布, 对每个样本, 按其顺序从小到大记为

$$X_{(1)}, \cdots, X_{(n)},$$

称为顺序统计量. 设 $X_{(0)} = X_0 = 0$, 对任何正数 $x_1, \cdots, x_n$, 由对称性得

$$\begin{aligned}
&\mathbb{P}(X_{(k)} - X_{(k-1)} > x_k, 1 \le k \le n) \\
&= n!\mathbb{P}(X_k - X_{k-1} > x_k, 1 \le k \le n) \\
&= \mathrm{e}^{-x_n - 2x_{n-1} - \cdots - nx_1},
\end{aligned}$$

因此 $X_{(1)}, X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)}$ 独立且分别服从参数 $n, n-1, \cdots, 1$ 的指数分布.

19. 数学期望与方差: 随机变量 X 的期望定义为

$$\mathbb{E}X = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x,$$

其中 f 是 X 的密度函数, 在离散的场合,

$$\mathbb{E}X = \sum_n x_n \mathbb{P}(X = x_n),$$

所以期望也可以说是分布函数或者密度函数或者分布律的期望. 期望也称为平均, 它体现随机变量在某种意义下的平均值. 期望这个词直观上是说, 如果你必须要预测随机变量 X 的取值, 那么说它是 $\mathbb{E}X$ 是最靠谱的, 理由我们后面会说明. 方差定义为

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2],$$

它直观上描述了随机变量对于它的期望的平均偏差, 或者说是随机变量的集中程度. 期望和方差都是数值, 它们是随机变量最重要的两个数字特征或者说统计量, 如同一个人身上的两个特征, 比如身高和体重. 但是它们也仅仅是两个特征, 不能真的精确地重现分布, 不仅如此有时候可能会误导, 比如一个小区里有 100 家人家, 有 99 家收入在 1000 元以下, 1 家收入在一百万, 那么平均收入看起来很高, 但实际情况完全相反, 所以看期望和方差是简单且要紧的, 可是不要过分看重.

20. 在上面收入的例子中, 中数更好地反映实际情况, 中数是指位于中间位置的那个收入.

21. 期望和方差的性质: 期望的运算是线性的, 设 a, b 是常数,

$$\mathbb{E}(aX+bY)=a\mathbb{E}X+b\mathbb{E}Y, \tag{1.5}$$

但提醒读者注意的是这条众所周知的性质并不象看到的那么容易证明, 当然我们不会在这里试图去证明它. 方差不是线性的

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X).$$

但若 X, Y 独立, 则

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y). \tag{1.6}$$

也就是说平移和发射是不改变方差的. 常数的期望不变而方差为零.

22. 最好的预测: 对任何的实数 a, 我们有不等式

$$\mathbb{E}(X-\mathbb{E}X)^2\le\mathbb{E}(X-a)^2, \tag{1.7}$$

也就是说期望是离随机变量最近的一个常数, 用它来预测 (在某种意义上) 是最优的.

23. 例: 设 n 对夫妇参加舞会, 随后他们随机挑选舞伴, X_n 表示其中夫妇配对的数目. 随意给夫妇编个号, 用 A_k 表示第 k 对夫妇恰好配对的事件, 1_{A_k} 表示其示性, 那么

$$X_n=1_{A_1}+\cdots+1_{A_n}.$$

推出 $\mathbb{E}[X_n]=n\mathbb{P}(A_1)=1$, 因为

$$\mathbb{P}(A_1)=\cdots=\mathbb{P}(A_n)=\frac{1}{n}.$$

类似地

$$\mathbb{E}[X_n^2]=\mathbb{E}[X_n]+\sum_{i\neq j}\mathbb{P}(A_i\cap A_j)=2,$$

因此 $\mathrm{Var}(X_n)=1$.

24. 期望计算公式: 很多场合下, 我们需要计算随机变量的函数的期望, 比如 $\mathbb{E}X^2$, 设 X 是随机变量, 密度函数为 $f(x)$, $y=g(x)$ 是实数域上的连续函数, 那么 $g(X)$ 也是随机变量, 如果按照上面的公式算它的期望, 我们需要算出 $g(X)$ 的密度函数, 但其实有一个简单的公式, 可以直接用 X 的密度来计算, 非常重要.

    定理1.1
    $$\mathbb{E}[g(X)] = \int_{\mathbf{R}} g(x)f(x)\mathrm{d}x, \tag{1.8}$$
    这里积分 $\int_{\mathbf{R}}$ 表示从 $-\infty$ 到 $+\infty$ 的积分.

25. 正态分布: 正态分布是概率论以及随机分析中最重要的分布, 不是之一. 几个科学界的天才从不同的道路得到这个分布, 法国的 De Moivre 从中心极限定理的角度, 德国的 Gauss 从考虑测量误差的角度, A. Einstein 从分析分子运动规则的角度, 这也说明了正态分布的适用广泛和重要性. 正态分布的密度函数如下定义
    $$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\ x\in\mathbf{R},$$
    其中 $\mu\in\mathbf{R}$, $\sigma>0$ 是两个参数. 如果随机变量 X 的密度函数如上, 我们就说 X 服从正态分布, 写为 $X\sim N(\mu,\sigma^2)$.

26. 钟型曲线: 正态分布的图像如同一口倒扣的钟, 所以称为 Bell curve, 即钟型曲线. 图像关于 $x=\mu$ 对称, 钟型曲线最高点的高度是 $\frac{1}{\sqrt{2\pi}\sigma}$, 因为与 x 轴所夹的区域面积总是 1, 所以 σ 越小, 曲线就越高而瘦, 随机变量就越集中, 反之 σ 越大, 曲线就低而胖, 随机变量就越分散, 即 σ 体现 X 的集中度.

27. 标准正态分布: 当 $\mu=0$, $\sigma^2=1$ 时的正态分布是标准正态分布, 标准正态分布的图像是偶函数, 体型标准. 标准正态分布的分布函数用 Φ 表示, 即
    $$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t,\ x\in\mathbf{R}.$$
    这个函数没有解析表达式, 也就是说被积函数的不定积分无法用初等函数表达, 通常概率教材最后都会附上标准正态分布的数值表.

28. 习题: 设 $X_1,\cdots,X_n$ 独立且服从标准正态分布, 求 $Y=X_1^2+\cdots+X_n^2$ 的密度函数.

29. 期望方差: 设 $X \sim N(\mu, \sigma^2)$, 那么 $\mathbb{E}X = \mu$, $\mathrm{Var}(X) = \sigma^2$, 所以这两个参数的数学意义与直观是符合的. 实际上, 如果 X 是标准正态分布的, 那么对于正整数 n,

$$\mathbb{E}(X^{2n}) = (2n-1)!!.$$

这可以通过分部积分和数学归纳法证明. 反过来, 上面矩的值也唯一地刻画了标准正态分布, 即如果

$$\mathbb{E}(X^{2n}) = (2n-1)!!, \quad \mathbb{E}(X^{2n-1}) = 0,$$

那么 $X \sim N(0,1)$.

30. 模拟: 在实际计算时, 模拟一个给定分布函数的随机变量是非常有用的. 电脑的编程语言中一般都有一个服从 $[0,1]$ 均匀分布的随机变量, 称为随机数, 虽然它实际上是伪随机的, 但也只能如此了. 怎么用它来生成其它分布的随机变量呢? 用 U 表示随机数, 想要一个分布函数为 F 的随机变量 X, 经典的公式是

$$X = F^{-1}(U),$$

其中 F^{-1} 是 F 的逆函数,

$$F^{-1}(x) := \inf\{y : F(y) \geq x\}.$$

这公式虽然简单漂亮, 但是因为逆函数不总是那么容易表达的, 所以用起来难. 比如正态分布函数的逆就不容易表达的. 有没有更有效的方法? 这里介绍一个. 离散情况就不说了, 比较简单. 我们假设 F 的密度函数是 f. 假设我们已经可以模拟密度为 g 的随机变量了且存在常数 c 使得对所有 y, 当 $g(y) \neq 0$ 时,

$$\frac{f(y)}{g(y)} \leq c,$$

然后运行下面的程序

(a) 产生一个随机数 U, 再产生一个密度是 g 的随机量 Y;

(b) 如果 $U \leq f(Y)/cg(Y)$, 那么 $X = Y$, 否则重复第一步.

证明 X 服从密度 f 的工作留作习题. 这里我们解释一下怎么模拟服从标准正态分布的随机量 Z.

(a) 参数 1 的指数分布随机量 V 容易模拟, 它的分布函数是 $1-\mathrm{e}^{-x}$, 逆函数是 $-\log(1-y)$, 因此产生一个随机数 U, $V=-\log(1-U)$ 服从参数 1 的指数分布.

(b) 不要指望正态密度 f 被指数密度控制. 所以在得到 Y 后再掷个硬币, 正面的话 $Y=V$, 反面的话 $Y=-V$, 那么 Y 的密度函数是

$$g(x)=\frac{1}{2}\mathrm{e}^{-|x|},\ x\in\mathbf{R}.$$

这样

$$\frac{f(x)}{g(x)}=\frac{2}{\sqrt{2\pi}}\mathrm{e}^{x-\frac{1}{2}x^2}=\frac{2\sqrt{\mathrm{e}}}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-1)^2}\le 2.$$

(c) 再产生一个随机数 U, 如果 $U\le f(Y)/2g(Y)$, 那么 $Z=Y$, 否则拒绝并重复上述步骤.

31. 习题: 证明模拟中产生的随机量 X 服从密度函数 f. 证明: 实际上 X 是一个条件分布.

$$\begin{aligned}\mathbb{P}(X\le x)&=\mathbb{P}(Y\le x|U\le f(Y)/cg(Y))\\&=\frac{\mathbb{P}(Y\le x,U\le f(Y)/cg(Y))}{\mathbb{P}(U\le f(Y)/cg(Y))}\\&=\frac{\int_{-\infty}^{x}g(y)f(y)/cg(y)\mathrm{d}y}{\int_{\mathbf{R}}g(y)f(y)/cg(y)\mathrm{d}y}\\&=\int_{-\infty}^{x}f(y)\mathrm{d}y.\end{aligned}$$

32. 习题: 如果随机变量 X 的分布函数 F 连续, 证明随机变量的函数 $F(X)$ 服从 $[0,1]$ 上均匀分布.

第 二 讲　预备知识 II

摘要 本章还是从正态分布的特征函数与矩性质开始, 顺便介绍了特征函数这个重要工具, 然后介绍概率论中最重要的几个极限定理, 包括 Bernoulli 大数定律和中心极限定理, 前者是随机产品一种重要定价方式的依据, 后者说明正态分布的中心地位, 最后引入条件期望的概念.

关键词
- 特征函数
- 正态分布再生性
- 大数定律
- Chebyshev 不等式
- 中心极限定理
- 条件期望

1. 特征函数: 密度函数为 $f(x)$ 的随机变量 X 的特征函数定义为

$$\phi(t)=\mathbb{E}[\mathrm{e}^{itX}]=\int_{-\infty}^{+\infty}\mathrm{e}^{itx}f(x)\mathrm{d}x,\ t\in\mathbf{R},\tag{2.1}$$

其中 $i=\sqrt{-1}$. 注意定义右边说明特征函数有分布函数唯一确定. 特征函数不容易算, 经常要用到复函数的积分性质, 但是特征函数非常重要, 也称为 Fourier 变换, 它有两个重要性质:

(1) 如果 X 与 Y 独立, 那么 $X+Y$ 的特征函数是 X 的特征函数与 Y 的特征函数的乘积.

(2) 如果 X 与 Y 的特征函数相等, 那么它们的分布相等, 也就是说特征函数唯一决定分布函数.

特征函数就像软件里的某种无损压缩方法, 它可以精确地恢复分布函数.

2. 特征函数的唯一性: 这是概率论中最难理解的思想也是最最要的工具. 随机变量的分布是由一个足够丰富的函数类上的期望确定的, 按其定义

$$F_X(x)=\mathbb{P}(X\le x)=\mathbb{E}(1_{(-\infty,x]}(X)),$$

即由示性函数类

$$\{1_{(-\infty,x]}(X):x\in\mathbf{R}\}$$

确定, 这在讲独立的时候已经提到过. 另外就是三角函数类

$$\{\mathrm{e}^{ixX}:x\in\mathbf{R}\},$$

它是全体 (有界可测) 函数的类 $\{g(X):g\}$ 的一个子集, 但是前面两个比较简单的函数类基本上已经足够丰富到可以反映出随机变量 X 的分布特性了. 或者说一个期望的等式对于前面的函数族成立就可以推出对全体函数成立. 在很多情况下, 这三个函数类可以互相替换以得到所需结果. 例如, 如果 X, Y 的特征函数一样, 那么 $\mathbb{E}(\mathrm{e}^{ixX})=\mathbb{E}(\mathrm{e}^{ixY})$ 对所有 x 成立, 这时可以替换为示性函数类得

$$F_X(x)=\mathbb{E}(1_{(-\infty]}(X))=\mathbb{E}(1_{(-\infty]}(Y))=F_Y(x)$$

对所有 x 成立, 因此它们同分布. 再例如, 设 X, Y 是随机变量, 且 $\mathbb{E}[Y\mathrm{e}^{ixX}] = 0$ 对所有 $x \in \mathbf{R}$ 成立, 那么

$$\mathbb{E}[Yg(X)] = 0$$

对所有有界 (可测) 函数成立了. 实际上最后一讲证明鞅表示定理就必须用这个思想, 当然这实际上是 Fourier 变换的思想.

3. 正态分布的特征函数: 设 $X \sim N(\mu, \sigma^2)$, 我们来试着计算 X 的特征函数. 我们不舫先假设 X 是标准正态分布的. 回忆 e^x 的 Taylor 展开式

$$\mathrm{e}^x = 1 + x + \cdots + \frac{x^n}{n!} + \cdots .$$

这个展开对复数也成立, 因此

$$\mathrm{e}^{itX} = 1 + itX + \cdots + \frac{(itX)^n}{n!} + \cdots .$$

然后两边取期望, 利用 29 中的公式, 右边 n 是奇数时期望为零, 故有

$$\begin{aligned}
\phi(t) &= \mathbb{E}[\mathrm{e}^{itX}] \\
&= \sum_{n=0}^{\infty} \frac{(it)^{2n}\mathbb{E}X^{2n}}{(2n)!} \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{t^{2n}}{(2n)!!} \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{t^{2n}}{2^n n!} \\
&= \exp(-\frac{1}{2}t^2)
\end{aligned}$$

有了标准正态分布的特征函数, 一般正态分布的特征函数就简单了, 因为 $(X-\mu)\sigma^{-1}$ 服从标准正态分布. 因此

$$\mathbb{E}[\exp(it\sigma^{-1}X)]\mathrm{e}^{-it\sigma^{-1}\mu} = \mathbb{E}[\exp(it(X-\mu)\sigma^{-1})] = \mathrm{e}^{-\frac{1}{2}t^2},$$

作变换 $z = t\sigma^{-1}$ 推出一般正态分布的特征函数是

$$\phi(z) = \exp(iz\mu - \frac{1}{2}\sigma^2 z^2), \ z \in \mathbf{R}. \tag{2.2}$$

4. 习题: 设 $X \sim N(\mu, \sigma^2)$, 用类似上面的方法计算 $\mathbb{E}\left(\mathrm{e}^{tX}\right)$, $t \in \mathbf{R}$.

5. 对数正态分布: 如果 X 是正随机变量, 且 $\log X$ 服从正态分布, 那么我们说 X 服从对数正态分布. 实际上, 如果 Y 服从正态分布, 那么 e^Y 服从对数正态分布.

6. 正态分布的再生性: 如果 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 且 X, Y 独立, 那么

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2),$$

这个性质称为再生性. 要证明这件事情, 利用特征函数的性质, 只需要证明 $X + Y$ 的特征函数等于

$$\exp\left(it(\mu_1 + \mu_2) - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right)$$

就可以了, 而仔细观察看出这恰好是 X 与 Y 的特征函数的乘积, 再由特征函数的性质, 因为 X 与 Y 独立, $X + Y$ 特征函数就等于它们两个的特征函数的乘积, 这样就证明了再生性.

7. 有限维正态分布: 前面出现过一维和二维正态分布, 但是要弄清楚 Brown 运动的有限维分布, 我们必须复习一下有限维正态分布. 这需要一点点矩阵的知识, 一个 n 阶对称方阵 A 称为正定的, 如果对任何非零 (分量不全为零) 的 n 维向量 (列向量) $x \in \mathbf{R}^n$ 有

$$x^t A x > 0,$$

其中右上角的 t 表示矩阵转置. 如果 $>$ 改为 $\geq$, 那么 A 称为非负定的. 显然一个非负定的矩阵是正定的当且仅当它是非退化的, 即其行列式非零. 关于对称矩阵最重要的结果是说对称矩阵 A 的特征值是实的而且存在正交矩阵 Q 使得 $Q^t A Q$ 是对角矩阵, 对角线的元素恰好是特征值全体. 当 A 是非负定时, 特征值是非负的; 当 A 是正定时, 特征值都是正的. 一个 n 维随机向量 $X = (X_1, \cdots, X_n)$ 是 (非退化) 正态分布的, 如果它的密度函数可以写为如下形式:

$$p(x) := C \cdot \exp\left(-\frac{1}{2}(x - a)^t A^{-1}(x - a)\right), \ x \in \mathbf{R}^n, \tag{2.3}$$

其中 $a \in \mathbf{R}^n$, A 是 n- 阶正定矩阵, A^{-1} 是 A 的逆. 显然向量 a 恰好是 X 的期望向量, 即 $\mathbb{E}[X] = a$, 可见 $X - a$ 的密度函数应该是

$$p(x + a) = C \cdot \exp\left(-\frac{1}{2}x^t A^{-1} x\right).$$

C 是规范常数使得密度的积分等于 1. 这时我们说 X 服从正态分布 $N(a,A)$. 当 A 是单位矩阵 I, $a=0$ 时,

$$p(x)=\left(\frac{1}{\sqrt{2\pi}}\right)^n \mathrm{e}^{-\frac{|x|^2}{2}},$$

其中 $|x|=x^tx=x_1^2+\cdots+x_n^2$, 这时 $(X_1,\cdots,X_n)$ 是独立且都服从标准正态分布的, 我们就说 X 服从 n- 维标准正态分布. 对于正定矩阵 A, 存在唯一的正定矩阵 S 使得 $A=S^2$, 这时也写 $S=\sqrt{A}$. 设 $X\sim N(a,S^2)$, 类似一维情况下的标准化做线性变换

$$Y=S^{-1}(X-a),$$

那么由多元积分的换元公式知 Y 的密度函数为

$$\begin{aligned}p_Y(y)&=p(Sy+a)|S|\\&=C|S|\cdot\exp\left(-\frac{1}{2}y^tSA^{-1}Sy\right)\\&=C|S|\cdot\exp\left(-\frac{1}{2}|y|^2\right),\end{aligned}$$

其中 $|S|$ 是 S 的行列式, 也是换元的 Jacobi 行列式. 由此推出

$$C=\frac{1}{\sqrt{(2\pi)^n|S|}}.$$

向量 a 的意义已经知晓, 矩阵 A 的意义是什么呢? 实际上它是 X 的协方差矩阵. 由定义 X 的协方差矩阵是随机变量矩阵 $(X-a)(X-a)^t$ 的各项求期望得到的矩阵, 记为

$$\mathrm{cov}(X)=\mathbb{E}[(X-a)(X-a)^t].$$

用变换 $(X-a)=SY$ 得

$$\mathrm{cov}(X)=\mathbb{E}[SYY^tS]=S\mathbb{E}[YY^t]S=S^2=A,$$

其中 $\mathbb{E}[YY^t]$ 是标准正态分布的协方差矩阵, 它显然是单位矩阵.

引理2.1　一个 n 维随机向量 X 是正态分布的当且仅当存在一个非退化线性变换 T 和向量 a 使得 $T(X-a)$ 是 n 维标准正态分布的随机向量. 因此正态分布随机向量的非退化线性变换依然服从正态分布.

8. 习题: 如果 X, Y 的联合密度函数是 $f(x, y)$, A 是非退化二阶方阵, 作变换
$$(U, V) = (X, Y)A.$$
求 (U, V) 的联合密度函数.

9. 习题: 设 X_1, X_2 独立且服从标准正态分布, (1) 求 $Y = X_2/X_1$ 的密度函数. (2) 记 $X = (X_1 + X_2)/2$, 证明: X 与 $(X_1 - X)^2 + (X_2 - X)^2$ 独立.

10. 特征函数: 随机向量 $(X_1, \cdots, X_n)$ 的特征函数定义为
$$\phi(x_1, \cdots, x_n) := \mathbb{E}\left(\mathrm{e}^{i(x_1 X_1 + \cdots + x_n X_n)}\right) = \mathbb{E}\left(\mathrm{e}^{ix^t X}\right).$$
我们来算正态分布的特征函数. 首先求标准正态分布的随机向量 X 的特征函数, 这时 $X_1, \cdots, X_n$ 独立标准正态分布的, 因此
$$\phi(x_1, \cdots, x_n) = \mathbb{E}\left(\mathrm{e}^{ix_1 X_1}\right) \cdots \mathbb{E}\left(\mathrm{e}^{ix_n X_n}\right) = \mathrm{e}^{-\frac{1}{2}|x|^2}.$$
一般地, 如果 X 的密度函数是 (2.3), 那么 $\sqrt{A^{-1}}(X - a)$ 是标准正态的, 因此
$$\mathbb{E}\left(\mathrm{e}^{ix^t \sqrt{A^{-1}}(X-a)}\right) = \mathrm{e}^{-\frac{1}{2}x^t x},$$
做变换
$$y = \sqrt{A^{-1}}x,\ x = \sqrt{A}y$$
$$\mathbb{E}\left(\mathrm{e}^{iy^t (X-a)}\right) = \mathrm{e}^{-\frac{1}{2}y^t A y},$$
把 $\mathrm{e}^{-iy^t a}$ 乘到右边再把 y 换成 x, 得到
$$\phi(x) = \mathbb{E}\left(\mathrm{e}^{ix^t X}\right) = \mathrm{e}^{-\frac{1}{2}x^t A x + ix^t a}.$$

11. 大数定律: 大数定律是概率论最重要的结果之一, 是著名数学家 (雅各) Bernoulli 在他死后的 1713 年出版的书中给出的, 它是说如果你重复一个成功概率为 p 的随机试验很多次, 那么成功的频率将逼近概率. 用数学语言来说, 重复一个成功 (事件 A) 概率为 p 的随机试验 n 次, 用 $n(A)$ 表示其中成功的次数, 那么
$$\lim_n \frac{n(A)}{n} = p$$

大数定律给出了检验概率的一个方法, 实际上, 概率是一个随机事件的一个属性, 除了在一些简单的情况和合理的假设下, 这个属性不为人知. 那么频率是我们触摸概率的一种方式. 但是这里的 n 要充分的大, 这也就是我们前面说只有当一个随机试验可以重复时, 我们才有可能触摸到真正的概率, 或者说概率才有意义.

12. 大数定律的一般化: 更一般地, 我们把大数定律这样叙述.

定理2.1 设

$$X_1, X_2, \cdots, X_n, \cdots$$

是独立且分布相同的随机变量序列, 那么

$$\lim_n \frac{X_1 + X_2 + \cdots + X_n}{n} = \mathbb{E}X_1. \tag{2.4}$$

独立且分布相同 (或者说, 独立同分布) 实际上是 '独立地重复' 的意思.

13. 应用: 大数定律应用广泛, 几乎所有的随机产品定价都要用到大数定律. 所谓随机产品, 就是以某个固定价格买卖的产品, 而产品的最终价值是随机的, 这在我们这个商品社会里是极其普遍的, 如彩票保险等, 最典型的是赌场, 例如某个赌场设计一个投币的赌博机 (slot machine), 它预先规定好出现某个图案给多少钱奖励, 当然图案是随机出现的, 所以收益是一个随机变量, 定价是指顾客应该放多少钱玩一次游戏. 那么最好的办法是算玩一次游戏得到的平均收益, 用大数定律, 赌场可以找一个人连续玩一个月, 记下总的所得, 然后除以次数, 大概差不多就是一次游戏的平均收益了, 这个方法简单有效, 称为 Monte Carlo 方法.

14. 证明: 证明不难, 只需要用 Chebyshev 不等式就足够了, 对任何随机变量和任何正数 ε 有

$$\mathbb{P}(|X| > \varepsilon) \leq \frac{\mathbb{E}[X^2]}{\varepsilon^2}. \tag{2.5}$$

让我们从很简单的不等式开始.

(1) 非负随机变量的期望也是非负的. 如果 $X \geq 0$. 那么 $\mathbb{E}X \geq 0$.

(2) 如果 $X \geq Y$, 那么 $\mathbb{E}X \geq \mathbb{E}Y$.

(3) 对于一个事件 A, 定义一个随机变量 1_A, 它在集合 A 上等于 1, 其余地方等于 0. 那么 $\mathbb{E}[1_A] = \mathbb{P}(A)$.

(4) 首先很容易看出

$$X^2 \geq \varepsilon^2 \cdot 1_{\{|X|>\varepsilon\}},$$

所以

$$\mathbb{E}[X^2] \geq \varepsilon^2 \mathbb{E}[1_{\{|X|>\varepsilon\}}] = \varepsilon^2 \mathbb{P}(|X|>\varepsilon).$$

(5) 记

$$Y_n = \frac{X_1+X_2+\cdots+X_n}{n} - \mathbb{E}X_1 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}X_k).$$

用 Y_n 代入 Chebyshev 不等式, 我们只需计算 $\mathbb{E}Y_n^2$. 因为 $\{X_n\}$ 独立同分布性,

$$\mathbb{E}Y_n^2 = n^{-2}\mathrm{Var}(X_1+\cdots+X_n) = n^{-1}\mathrm{Var}(X_1),$$

因此对任何 $\varepsilon > 0$,

$$\lim_n \mathbb{P}(|Y_n| > \varepsilon) \leq \lim_n n^{-1}\varepsilon^{-2}\mathrm{Var}(X_1) = 0.$$

15. 收敛的意义: 本讲义所涉及的随机变量列 $\{X_n\}$ 的收敛有三种不同的含义, 一是几乎处处收敛, 就是对几乎所有样本 ω, $X_n(\omega) \longrightarrow X(\omega)$; 二是均方收敛, 即

$$\mathbb{E}[(X_n - X)^2] \longrightarrow 0;$$

三是依概率收敛, 就是刚刚说的, 对任何 $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0.$$

上面实际上已经证明了均方收敛推出依概率收敛. 不难证明几乎处处收敛也蕴含着依概率收敛. 但是几乎处处收敛和均方收敛没有什么互相蕴含的关系.

16. Borel-Cantelli 引理: 当依概率收敛的速度很快时, 也可以推出几乎处处收敛. 这个结果后面会用到.

引理2.2 如果对任何 $\varepsilon > 0$ 有

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < \infty,$$

则 X_n 几乎处处收敛于 X.

17. 习题: 如果存在趋于零的正数列 $\{\varepsilon_n\}$ 使得

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty,$$

那么 X_n 几乎处处收敛于 X.

18. 中心极限定理: 我们还是假设 $\{X_n\}$ 是独立同分布的, 期望是 μ, 方差是 σ^2. 如果把 $X_i - \mu$ 看成一次测量的误差, 那么 $\sum_{i=1}^n (X_i - \mu)$ 是 n 次测量的误差总和, 再除以 n 是平均误差, 大数定律说平均误差逼近 0. 现在我们来估计总误差的分布, 显然除以 n 太多了, 因为总误差的方差是 $n\sigma^2$, 那么

$$Y_n := \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma}$$

的期望等于 0 而方差等于 1, 是一个标准化的随机变量.

定理2.2 (中心极限定理) 对任何 $x \in \mathbf{R}$,

$$\lim_n \mathbb{P}(Y_n \le x) = \Phi(x), \tag{2.6}$$

其中 Φ 是标准正态分布的分布函数.

De Moivre 最先给出二项分布的中心极限定理, Laplace 改进了证明, 后来中心极限定理不断改进, 上面的形式是 Lindeberg-Lévy 中心极限定理. 中心极限定理的中心两字说明了正态分布在概率论中的中心地位. 证明较难, 需要更多的工具, 读者可以参考其他概率教材, 比如 [3].

19. 模拟: 中心极限定理给了我们一个简单的方法来近似模拟正态分布的随机变量. 取 $\{U_i : 1 \le i \le n\}$ 是 n 个随机数, 定义

$$X_i = 1_{\{U_i < 1/2\}} - 1_{\{U_i > 1/2\}},$$

那么 $\{X_i\}$ 独立同分布, 期望 0, 方差 1, 当 n 很大时

$$X = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$$

近似地标准正态分布. 在计算机越来越强大的今天, 多取一些随机数应该不是太大的问题.

20. 条件概率与条件期望: 在事件 A 发生的条件下, 事件 B 发生的概率为

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

随机变量 X 在事件 A 发生的条件下的期望为

$$\mathbb{E}(X|A) = \frac{\mathbb{E}(X; A)}{\mathbb{P}(A)},$$

其中 $\mathbb{E}(X; A) = \mathbb{E}(X1_A)$. 两者都很直观, 比如已知某人有两个孩子, 在已知至少有一个女儿的条件下, 两个都是女儿的概率是 $1/3$. 假如全国的家庭平均年收入是 50000, 上海的家庭平均年收入是 80000, 任取一个中国人其家庭收入记为 X, 那么 $\mathbb{E}X = 50000$, 如果知道此人是上海人, 那么猜其家庭收入是条件期望 $\mathbb{E}(X|A) = 80000$.

21. 全概率公式的推广: 众所周知, 如果在一个随机试验中, 事件 $\{\Omega_n\}$ 中有且仅有一个发生, 那么

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|\Omega_n)\mathbb{P}(\Omega_n).$$

同样地,

$$\mathbb{E}[X] = \sum_n \mathbb{E}(X|\Omega_n)\mathbb{P}(\Omega_n). \tag{2.7}$$

22. 例: 硬币的正反面分别记为 $1, 0$, 连续掷一个硬币得到一个 01 序列, 我们知道, 等待 1 出现所需时间 T_1 是几何分布, 期望是 $\mathbb{E}[T_1] = 2$, 那么等待 11 出现的时间 T_{11} 呢? 它不是几何分布. 稍一般些, 我们假设硬币正面概率是 p, 那么由上面的公式

$$\begin{aligned}
\mathbb{E}[T_{11}] &= \mathbb{E}[T_{11}|0]q + \mathbb{E}[T_{11}|1]p, \\
\mathbb{E}[T_{11}|0] &= 1 + \mathbb{E}[T_{11}], \\
\mathbb{E}[T_{11}|1] &= \mathbb{E}[T_{11}|10]q + \mathbb{E}[T_{11}|11]p = (\mathbb{E}[T_{11}] + 2)q + 2p
\end{aligned}$$

得到方程

$$\mathbb{E}[T_{11}] = (1 + \mathbb{E}[T_{11}])q + (\mathbb{E}[T_{11}] + 2)pq + 2p^2,$$

因此

$$\mathbb{E}[T_{11}] = \frac{q + 2pq + 2p^2}{1 - q - pq} = \frac{1 + p}{p^2}.$$

23. 习题: 硬币的正反面分别记为 $1, 0$, 连续掷一个硬币得到一个 01 序列, 求在这个序列中, (1) 11 首次出现的平均时间, (2) 01 首次出现的平均时间, (3) 101 首次出现的平均时间.

第 三 讲　条件数学期望

摘要 条件数学期望是随机分析理论中最基本的一个概念, 本章定义并给出了条件期望的各种刻画, 讲述了在具体情况下怎么计算条件期望, 最重要的是列出了条件期望的多条重要性质, 它们在后面的理论中经常用到, 最后从随机游动的性质引出了鞅的概念.

关键词
- 条件期望
- 最优预测
- 条件期望的刻画
- 条件期望的计算
- 条件期望的性质
- 随机游动与鞅

1. 一点历史: 条件期望的定义和基本性质都是由前苏联天才数学家 A. Kolmogorov 建立的.

2. 信息和条件期望: 在预测的时候, 信息是很重要的, 多知道一点信息, 预测就会更准确一些, 所以期望是随信息变化的. 那么什么是信息呢? 我们说信息是样本空间的一个分类, 比如说人类是一个样本空间的话, 男女就是一个分类, 国籍是一个分类, 文化程度可以分类等等. 你知道样本在某个分类中就是知道更多的信息了. 例如你是国家统计局的, 知道各种统计数字, 现在给你拉来一个中国人, 站在外面, 要你猜他/她的年收入, 你对他/她一无所知, 那你只好以全国的平均收入来猜测; 现在他/她说话了, 你知道是个女的, 你自然就以全国女性的平均收入来猜测; 她说她是上海人, 那么你就以上海女性的平均收入来猜测. 也就是说期望是随着信息而改变的, 条件期望是已知信息下所能作出的最佳抉择. 再例如打牌的时候, 根据出牌的情况, 每个人都在调整, 以作出最佳的出牌选择. 从这个角度说, 条件期望是无处不在的.

3. 关于随机变量的条件期望: 我们后面用到的大多数情况都是以随机变量或者随机向量作为信息源的条件期望, 一个随机随机变量实际上给出样本空间的分类, 比如人的收入作为随机变量它给出了人根据收入的分类, 人的籍贯作为随机变量给出了籍贯分类等等. 关于随机变量 X 的一个预测实际上是随机变量 X 的一个函数 $\phi(X)$, 另一个随机变量 Y 在已知 X 的条件下的期望是指在所有已知 X 的预测中离 Y 最近的一个预测 $\phi(X)$, 距离是指均方距离, 即
$$\mathbb{E}[(Y-\phi(X))^2] = \min_g \mathbb{E}[(Y-g(X))^2]. \tag{3.1}$$

4. 随机变量作为信息: 概率论或者随机分析中最重要的概念是可测性. 在数学上严格定义可测性需要很多铺垫, 但我们现在简单地把可测性理解为函数. 设 X 是一个随机变量或者多个随机变量组成的随机向量, Y 是随机变量, 把 Y 由 X 决定或者 Y 是 X 的函数说成为 Y 关于 X 可测, 即

   定义3.1 如果存在一个函数 g 使得 $Y=g(X)$, 那么说 Y 关于 X 可测, 记为 $Y\in\sigma(X)$.

   比如掷两个骰子点数分别为 X_1, X_2, X 是点数的和, 那么 X 关于 X_1, X_2 可测, 或 $X\in\sigma(X_1,X_2)$. 掷一个骰子点数为 X, $X_1 = 1_{\{X=6\}}$, $X_2 = 1_{\{X=2,4,6\}}$, 那么它们两者谁关于谁都不可测.

5. 可测表示与概率: 设随机变量 X 取值 1,2,3, 那么我们说 Y 关于 X 可测当且仅当存在 $a_1, a_2, a_3 \in \mathbf{R}$,

$$Y = a_1 1_{\{X=1\}} + a_2 1_{\{X=2\}} + a_3 1_{\{X=3\}}.$$

这是 Y 的可测表示, 它与概率测度毫无关系. 但若 $\mathbb{P}(X=3)=0$, 那么我们可以说 Y 关于 X 可测当且仅当 $a_1, a_2 \in \mathbf{R}$, 在几乎处处的意义下

$$Y = a_1 1_{\{X=1\}} + a_2 1_{\{X=2\}}.$$

这是 Y 的几乎处处意义下的可测表示, 它比通常的可测表示简单, 与测度大有关系. 所以上面定义中的 $Y = g(X)$ 只要求几乎处处成立就可以了, 随机分析中的鞅表示问题就是这种表示.

6. 优化问题的解: Y 关于 X 的条件期望 $\phi(X)$, 记为 $\mathbb{E}[Y|X]$, 是优化问题

$$\mathbb{E}[(Y-\phi(X))^2] = \min_g \mathbb{E}[(Y-g(X))^2]$$

的解, 其中 ϕ, g 都是实函数. 上面这个最优化问题有没有唯一的解? 怎么求解? 下面的定理可以帮助我们.

定理3.1 $\phi(X)$ 是解当且仅当满足对任何 X 的函数 $h(X)$ 有

$$\mathbb{E}[(Y-\phi(X))h(X)] = 0. \tag{3.2}$$

7. 证明: 先设上面条件满足, 那么有

$$\begin{aligned}\mathbb{E}[(Y-g(X))^2] &= \mathbb{E}[(Y-\phi(X)+\phi(X)-g(X))^2] \\ &= \mathbb{E}[(Y-\phi(X))^2] + \mathbb{E}[(\phi(X)-g(X))^2],\end{aligned}$$

因此 $\phi(X)$ 是优化问题的解, 而且唯一. 反过来, 如果 $\phi(X)$ 是优化问题的解, 那么对任何实数 λ,

$$\mathbb{E}[(Y-\phi(X))^2] \le \mathbb{E}[(Y-\phi(X)+\lambda h(X))^2],$$

等价于

$$2\lambda\mathbb{E}[(Y-\phi(X))h(X)] + \lambda^2\mathbb{E}[h(X)^2] \ge 0,$$

那么 λ 的系数必须是 0, 故而有 (3.2) 成立.

8. 定理或者定义: $\phi(X)$ 是 Y 关于 X 的条件期望当且仅当对任何 h 满足 (3.2). 这个定义在实际使用时非常方便. 在讨论 Y 关于其它随机变量的条件期望时, 一个技术性条件是可积性, 即

$$\mathbb{E}|Y| < \infty.$$

我们不常提到的原因是我们总是假设可积性条件满足.

9. 习题: 设 X, Y 是两个随机变量, 如果对任何 $x \in \mathbf{R}$, 有

$$\mathbb{E}\left(\mathrm{e}^{ixX}|Y\right) = \mathbb{E}\left(\mathrm{e}^{ixX}\right),$$

应用定理 3.1 证明 X 与 Y 独立.

10. 解函数 $y = \phi(x)$ 不需要对任何函数 h 满足 (3.2), 实际上只需要对一类特殊函数满足就可以了, 即对任何实数 x,

$$\mathbb{E}[\phi(X); X \leq x] = \mathbb{E}(Y; X \leq x). \tag{3.3}$$

这个简化的证明比较抽象, 这里省略.

11. 条件期望的计算: 如果我们知道 (X, Y) 的联合密度函数 $f(x, y)$ 以及 X 的密度函数 $f_X(x)$, 那么 (3.3) 可以写为

$$\int_{-\infty}^{x} \phi(t) f_X(t)\mathrm{d}t = \int_{\infty}^{+\infty} \mathrm{d}y \int_{-\infty}^{x} y f(t, y)\mathrm{d}t.$$

两边对 x 求导得

$$\phi(x) f_X(x) = \int_{-\infty}^{+\infty} y f(x, y)\mathrm{d}y,$$

从而得

$$\phi(x) = \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)}\mathrm{d}y. \tag{3.4}$$

因为固定 x 有 $\int f(x, y)\mathrm{d}y = f_X(x)$, 所以 $\frac{f(x,y)}{f_X(x)}$ 关于 y 是密度函数, 称为给定 $X = x$ 时 Y, 有时写成 $Y|_{\{X=x\}}$, 的条件密度函数. 而 $\phi(X)$ 是给定 X 的条件下 Y 的期望, 记为 $\mathbb{E}(Y|X)$, 或者说 $\phi(x)$ 是给定 $X = x$ 的条件下 Y 的期望, 记为 $\mathbb{E}(Y|X = x)$, 由上式可以看出, 它等于给定 $X = x$ 时 Y 的条件密度的期望.

12. 现在我们知道条件期望的直观意义以及怎么计算条件期望了:

$$\phi(X) = \mathbb{E}(Y|X),\ \phi(x) = \mathbb{E}(Y|X=x).$$

直观地看

$$\phi(x) = \frac{\mathbb{E}(Y; X=x)}{\mathbb{P}(X=x)},$$

和前面条件平均的概念相符. 显然如果 X, Y 独立, 那么 X 的信息对预测 Y 没有帮助, 和没有信息是一样的, 因此 $\mathbb{E}(Y|X) = \mathbb{E}Y$. 如果 Y 本身就是 X 的函数, 也就是说 X 的信息完全确定了 Y, 那么 Y 本身是上面最优问题的解, 即 $\mathbb{E}(Y|X) = Y$.

13. 当 X 离散的时候, $\mathbb{E}(Y|X)$ 是离散随机变量, 它在 $X = x$ 时的取值是 $\mathbb{E}(Y|X = x)$, 那么

$$\mathbb{P}(\mathbb{E}(Y|X) = \mathbb{E}(Y|X=x)) = \mathbb{P}(X=x).$$

因此

$$\mathbb{E}[\mathbb{E}(Y|X)] = \sum_x \mathbb{E}(Y|X=x)\mathbb{P}(X=x) = \mathbb{E}Y.$$

这说明, 期望是相容的. 举例说, 假设你有数据, 知道各省居民的平均收入, 但是你不知道某个人是哪个省的, 那么那么你还是只能用全国的平均收入来预测. 这个公式告诉我们全国的平均收入可以根据各省的平均收入来计算, 是各省平均收入的加权平均. 这类似于我们熟悉的全概率公式: 一个事件的概率是不同分类情况下概率的加权平均. 更一般地, 如果某甲知道的信息比某乙多, 某甲预测过的随机变量再让某乙来预测与一开始就让某乙来预测是一样的.

14. 条件完全的性质: 在后面真正计算条件期望的场合不多, 更重要的是使用条件期望的性质来解决问题. 在下面所述的性质中, 我们不妨把 X 想象成随机向量, 或者是有限多个随机变量的集合, 这时上面所述的那些计算公式仍然有效. 另外下面这些性质的证明也许有点抽象, 但是直观上理解是不难的.

定理3.2 条件期望性质:

(a) $\mathbb{E}(1|X) = 1$;

(b) $\mathbb{E}(a_1Y_1 + a_2Y_2|X) = a_1\mathbb{E}(Y_1|X) + a_2\mathbb{E}(Y_2|X)$;

(c) 如果 $Y_1 \le Y_2$ 那么 $\mathbb{E}(Y_1|X) \le \mathbb{E}(Y_2|X)$;

(d) $|\mathbb{E}(Y|X)| \le \mathbb{E}(|Y||X)$;

(e) $\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}[Y]$;

(f) 如果 X_1 由 X_2 确定, 那么 $\mathbb{E}[\mathbb{E}(Y|X_2)|X_1] = \mathbb{E}(Y|X_1)$;

(g) 如果 Y 由 X 决定, 那么 $\mathbb{E}(Y|X) = Y$;

(h) 如果 H 由 X 确定, 那么

$$\mathbb{E}(H \cdot Y|X) = H \cdot \mathbb{E}(Y|X).$$

可能读者很不情愿去证明这些看上去显然的东西,但数学就是这样,显然的东西也是需要证明的, 要真正弄懂, 你就必须能够证明它. 定理 3.1 是我们证明这些性质的主要工具. 让我们来证明最后一个性质. 设 $\phi(X) = \mathbb{E}(Y|X)$, $h(X) = H$, $\psi(X) = h(X)\phi(X)$, 我们要证明 $\mathbb{E}(H \cdot Y|X) = \psi(X)$, 按照定义, 就是证明对任何 g 有

$$\mathbb{E}[(HY - \psi(X))g(X)] = 0.$$

而左边等于 $\mathbb{E}[(Y - \phi(X))h(X)g(X)]$, 应用定义 3.2 即推出它等于 0.

15. 说明: 概率论中关于随机变量的等式或不等式, 一般都是在几乎必然的意义下成立, 也就是说不成立的概率等于 0. 要注意不可能事件和概率 0 事件的微小差别.

16. 例: 设 (X, Y) 的联合密度是

$$f(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right),$$

这是说 (X, Y) 服从相关系数是 r ($|r| < 1$) 的二维正态分布, 那么 X, Y 分别都服从标准正态分布, 且 $Y|_{X=x}$ 的条件密度函数是

$$f(x,y)/f_X(x) = \frac{1}{\sqrt{2\pi(1-r^2)}} \exp\left(-\frac{(y-rx)^2}{2(1-r^2)}\right),$$

即 $Y|_{X=x}$ 服从正态分布 $N(rx, (1-r^2))$, 它的期望是 rx, 因此 $\mathbb{E}(Y|X) = rX$.

17. 掷两个骰子, 按点数大小记为 (X,Y), 那么 $\mathbb{E}(Y|X=1)=41/11$, $\mathbb{E}(Y|X=2)=38/9$, $\mathbb{E}(Y|X=3)=33/7$, ... 实际上

$$\mathbb{E}(Y|X=x)=\frac{42-x^2}{13-2x},\ x=1,2,3,4,5,6,$$

因此

$$\mathbb{E}(Y|X)=\frac{42-X^2}{13-2X}.$$

18. 习题:

(a) 设 (X,Y) 的期望方分别是 μ_1,σ_1^2 和 μ_2,σ_2^2, 且

$$((X-\mu_1)\sigma_1^{-1},(Y-\mu_2)\sigma_2^{-1})$$

服从相关系数为 r 的标准正态分布, 求 $\mathbb{E}(Y|X)$;

(b) 设 $X_1,\cdots,X_n$ 是独立且服从标准正态分布, 求 $\mathbb{E}(X_1|X_1+\cdots+X_n)$.

(c) 掷两个骰子, X 表示两个点数差的绝对值, Y 表示两个点数的和. 求 $\mathbb{E}(Y|X)$ 与 $\mathbb{E}(X|Y)$.

19. 随机序列和信息流: 一个随机序列 $X_1,X_2,\cdots,X_n,\cdots$ 可以看成是一个按离散时间记录的随机过程. 生活中有许多的这样的随机过程, 比如股票, 赌博等. 随机变量的集合 $\{X_1,\cdots,X_n\}$ 可以看成是时刻 n 前的信息, 所以称为一个信息流, 随着时间流逝, 信息越来越多, 我们会不断地调整我们的判断和策略. 有时候, 我们用 $\sigma\{X_1,\cdots,X_n\}$ 表示包含的信息, 用 $X\in\sigma\{X_1,\cdots,X_n\}$ 表示随机变量 X 由 $X_1,\cdots,X_n$ 这些信息决定.

20. 随机游动: 从最简单的赌博开始, 甲乙两人赌博, 每次输赢一元, 甲赢的概率为 p, 输的概率是 $q=1-p$, 用数学来表达的话, 设 $\{\xi_n\}$ 表示独立同分布随机序列,

$$\mathbb{P}(\xi_n=1)=p,\ \mathbb{P}(\xi_n=-1)=q,\ p+q=1,$$

令 $S_n(a)=a+\sum_{i=1}^n\xi_i$, 那么 $\{S_n(a)\}$ 称为从 a 出发的随机游动 (当输赢概率一样时), 或者非对称随机游动 (当输赢概率不同时). 如果 $a=0$, 我们直接写 S_n. 理解随机游动是理解随机分析的第一步.

21. 首次回归时间: 先用非常初等的方法. 甲乙两个赌徒以掷公平硬币每次一元的输赢赌博, $S_0 = a$ 是甲所带的赌资, S_n 是 n 局后甲所剩下的钱, 令

$$T = \min\{n \geq 1 : S_n = a\}$$

称为是首次回归时间, 这是所谓的随机时间, 当 $\{n \geq 1 : S_n = a\}$ 是空集时, 定义 T 为无穷. 我们来计算 T 的分布并证明 $\mathbb{P}(T < \infty) = 1$, 但是 $\mathbb{E}[T] = \infty$, 也就是说它会在有限时间内回归但平均回归时间是无穷. 我们用非常初等的方法来做, 不妨设 $a = 0$, 首先在 XY 平面看

$$\{S_0, S_1, \cdots, S_n, \cdots\}$$

它实际上是一条格点轨道 $\{(n, S_n) : n \geq 0\}$, 如同放大了的股票价格走势, 它的特点是上下一格. 显然 T 只能是偶数, 因为 $S_{2n-1} \neq 0$, 故

$$\begin{aligned}
\mathbb{P}(T = 2) &= 1/2; \\
\mathbb{P}(T = 2n) &= \mathbb{P}(S_2 \neq 0, \cdots, S_{2n-1} \neq 0, S_{2n} = 0) \\
&= 2\mathbb{P}(S_2 > 0, \cdots, S_{2n-2} > 0, S_{2n} = 0)
\end{aligned}$$

怎么算这个概率? 让我们引入反射原理, 它是由伟大的 Maxwell 和 Kelvin 首先应用的.

定理3.3 设 $a, b > 0$, 那么从 (m, a) 到 (n, b) 碰到 x- 轴的格点轨道的总数等于从 (m, a) 到 $(n, -b)$ 的格点轨道的总数.

反射原理的证明很简单, 只要轨道中从碰到 x- 轴的最后一个时间点开始的一段依 x- 轴反射得到一条从 (m, a) 到 $(n, -b)$ 的格点轨道, 这是两个集合之间的一一对应, 因此轨道数是一样的. 证明完毕. 现在我们需要算 $\{S_2 > 0, \cdots, S_{2n-1} > 0, S_{2n} = 0\}$ 中的格点轨道数, 也就是从 $(0,0)$ 到 $(2n, 0)$ 位于 x- 轴上方但中间不碰到 x- 轴的格点轨道总数, 它等于 $(1,1)$ 到 $(2n-1, 1)$ 不碰到 x- 轴的格点轨道总数, 也等于 $(1,1)$ 到 $(2n-1, 1)$ 的格点轨道数减去从 $(1,1)$ 到 $(2n-1, 1)$ 中间碰到 x- 轴的格点轨道数, 后者由反射原理等于 $(1,1)$ 到 $(2n-1, -1)$ 的格点轨道数, 因此

$$\mathbb{P}(S_2 > 0, \cdots, S_{2n-2} > 0, S_{2n} = 0) = \frac{1}{2^{2n}}(N_{2n-2,0} - N_{2n-2,-2}),$$

其中 $N_{n,a}$ 是从 $(0,0)$ 到 (n,a) 的格点轨道总数, 不难计算

$$N_{n,a} = \binom{n}{(n+a)/2}.$$

代入得

$$\begin{aligned}
&\mathbb{P}(S_2 > 0, \cdots, S_{2n-2} > 0, S_{2n} = 0)] \\
&= \frac{1}{2^{2n}} \left[\binom{2n-2}{n-1} - \binom{2n-2}{n-2} \right] \\
&= \frac{1}{2n-1} \frac{1}{2^{2n+1}} \binom{2n}{n},
\end{aligned}$$

因此有 T 的分布律

$$\begin{aligned}
\mathbb{P}(T = 2n) &= \frac{1}{2n-1} \frac{1}{2^{2n}} \binom{2n}{n} \\
&= \left(\frac{2n}{2n-1} - 1 \right) \frac{1}{2^{2n}} \binom{2n}{n} \\
&= \frac{1}{2^{2n-2}} \binom{2n-2}{n-1} - \frac{1}{2^{2n}} \binom{2n}{n},
\end{aligned}$$

从而推出 $\mathbb{P}(T < \infty) = 1$. 应用 Stirling 公式

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} \mathrm{e}^{-n}$$

得

$$\mathbb{E}[T] \sim \sum_{n \geq 1} \frac{1}{\sqrt{n}} = \infty.$$

22. 习题: 求 $\mathbb{P}(S_2 > 0, \cdots, S_{2n} > 0)$.

23. 习题: 设 T_k 是首次抵达 k 的时间, 用反射原理算 T_1 的分布律.

24. 数列及其母函数: 母函数是研究分布律的一种非常有效的方法. 数列 $\{a_n : n \geq 0\}$ 的母函数 (generating function) 定义为

$$G(z) := a_0 + a_1 z + \cdots + a_n z^n + \cdots, \ z \in \mathbf{R}.$$

右边是幂级数, 有收敛半径, 如果收敛半径正, 那么幂级数可以反过来唯一决定数列, 这时母函数才有意义, z 的定义域是右边的收敛区域内. 也可以认为右边是左边的幂级数展开, 几个熟悉的 Taylor 展开

$$\frac{1}{1-z} = 1 + z + z^2 + \cdots + z^n + \cdots;$$

$$
\begin{aligned}
-\log(1-z) &= z + \frac{z^2}{2} + \cdots + \frac{z^n}{n} + \cdots ; \\
\mathrm{e}^z &= 1 + z + \frac{z^2}{2} + \cdots + \frac{z^n}{n!} + \cdots ; \\
&\cdots\cdots\cdots .
\end{aligned}
$$

如果 $\{b_n\}$ 的母函数是 $H(z)$, 那么

$$
\begin{aligned}
G(z)H(z) = a_0b_0 &+ (a_0b_1 + a_1b_0)z \\
&+ \cdots + (a_0b_n + a_1b_{n-1} + \cdots + a_nb_0)z^n + \cdots ,
\end{aligned}
$$

因此两个数列卷积的母函数等于母函数的乘积, 这是母函数最重要的性质. 把母函数的方法引进概率论, 可以用来讨论取非负指数值的随机变量 ξ 的分布律

$$
G(z) = \mathbb{P}(\xi = 0) + z\mathbb{P}(\xi = 1) + \cdots + z^n\mathbb{P}(\xi = n) + \cdots .
$$

那么母函数决定分布律.

25. 随机游动的马氏性: 还是考虑零点出发的随机游动 $\{S_n\}$, T_k 表示首次抵达 k 的时间, S_n 抵达 k 可以分为两步, (1) 抵达 $k-1$; (2) 从 $k-1$ 抵达 k. 当它抵达 $k-1$ 后重新算时间, 它是一个从 $k-1$ 出发的随机游动, 与之前的行为独立, 因此从 $k-1$ 抵达 k, 就相当于从 0 抵达 1, 这说明 T_k 等于 T_{k-1} 加上一个与之独立且与 T_1 同分布的量. 这是随机游动的马氏性, 看起来很直观, 证明不太难, 有兴趣的读者可以自己试试. 用母函数表达就是

$$
\mathbb{E}[z^{T_k}] = \mathbb{E}[z^{T_{k-1}}]\mathbb{E}[z^{T_1}] = \left(\mathbb{E}[z^{T_1}]\right)^k .
$$

好了, 现在让随机游动走一步, 看 S_1, 两种可能, 如果它等于 1, 那么 $T_1 = 1$, 如果它等于 -1, 那么它要从 -1 出发抵达 1, 相当于从 0 出发抵达 2, 由此推出

$$
\begin{aligned}
\mathbb{E}[z^{T_1}] &= \frac{1}{2}\mathbb{E}[z^{T_1}|S_1 = 1] + \frac{1}{2}\mathbb{E}[z^{T_1}|S_1 = -1] \\
&= \frac{1}{2}(z + \mathbb{E}[z^{1+T_2}] \\
&= \frac{1}{2}z\left[1 + \left(\mathbb{E}[z^{T_1}]\right)^2\right],
\end{aligned}
$$

解这个方程

$$
\mathbb{E}[z^{T_1}] = \frac{1-\sqrt{1-z^2}}{z},\ z \in [0,1]. \tag{3.5}
$$

$$\mathbb{P}(T_1 < \infty) = \lim_{z \uparrow 1} \mathbb{E}[z^{T_1}] = 1,$$

类似地, 再用全概率公式

$$\begin{aligned}\mathbb{E}[z^{T_0}] &= \frac{1}{2}(\mathbb{E}[z^{T_0}|S_1 = 1] + \mathbb{E}[z^{T_0}|S_1 = -1]) \\ &= z\mathbb{E}[z^{T_1}] = 1 - \sqrt{1 - z^2}.\end{aligned}$$

右边的展开

$$1 - \sqrt{1 - z^2} = \frac{1}{2}z^2 + \frac{1}{2!2^2}z^4 + \cdots + \frac{(2n-3)!!}{n!2^n}z^{2n} + \cdots,$$

即得到和前面一样的答案

$$\mathbb{P}(T_0 = 2n) = \frac{1}{2n-1}\binom{2n}{n}\frac{1}{2^{2n}}.$$

26. 习题: 设 ξ 的母函数是 $G(z)$, 证明:

$$\mathbb{P}(\xi < \infty) = \lim_{z \uparrow 1} G(z).$$

27. 习题: 对于一个不对称的随机游动 $\{S_n\}$, 用母函数方法算首中时 T_k 的分布律.

28. 习题: 在上面首次回归时间的问题中, 令 $u_{2n} := \mathbb{P}(S_{2n} = 0)$ 已知, 如果 $S_{2n} = 0$, 那么 $T \leq 2n$, 因此

$$\begin{aligned}u_{2n} &= \sum_{k=1}^{n} \mathbb{P}(S_{2n} = 0|T = 2k)\mathbb{P}(T = 2k) \\ &= \sum_{k=1}^{n} \mathbb{P}(S_{2n-2k} = 0)\mathbb{P}(T = 2k) \\ &= \sum_{k=1}^{n} u_{2n-2k}\mathbb{P}(T = 2k),\end{aligned}$$

利用这个关系可否算出 $\mathbb{P}(T = 2k)$?

第 四 讲　鞅与停时

摘要 鞅论是随机分析与金融数学的最重要的理论, 本章主要讲述了 Doob 的鞅基本定理, 就是一个可料序列关于鞅的随机积分仍然是鞅, 这个定理是整个随机分析的基石, 然后我们讲述停时的概念, 由此引出许多有趣的问题, 其中很多的计算和方法都是概率论中的经典.

关键词
- 鞅
- 随机积分
- 鞅基本定理
- Doob 分解
- 停时
- 倍增策略
- 输光问题

1. 一点历史: 鞅的研究是由 P.Lévy 在 1937 年开始的, 主要的工作是 J.L.Doob 在 1940 年完成的, 系统地总结在它 1953 年的专著中.

2. 从简单随机游动开始: 设 $\{S_n\}$ 表示一个公平赌博中 n 局后甲的所有赌资, 它是一个随机游动. 因为第 n 局的输赢结果 $S_n - S_{n-1}$ 与前 n 局独立, 因此有

$$\mathbb{E}(S_n - S_{n-1}|S_1, \cdots, S_{n-1}) = \mathbb{E}(S_n - S_{n-1}) = 0.$$

换句话说, 已知前面的结果对预测下一局结果没有任何帮助, 条件期望是 0. 也可以写成

$$\mathbb{E}(S_n|S_1, \cdots, S_{n-1}) = S_{n-1}.$$

把这个概念抽象出来就得到鞅的定义.

3. 鞅的定义: 设 $\{X_n\}$ 是随机序列, 如果对任何 n 有

$$\mathbb{E}(X_n|X_1, \cdots, X_{n-1}) = X_{n-1}, \tag{4.1}$$

那么我们说 $\{X_n\}$ 是鞅. 从上面的分析可以看出, 鞅就是公平的意思, 也就是你用以前的信息不能得到对你有利的判断. 期望等于 0 的独立同分布随机变量的部分和是鞅. 如果上面的等号换成 $\geq$, 那么是越来越好, 称为下鞅, 如果换成 $\leq$, 那么越来越差, 称为上鞅. 上鞅下鞅都是不公平的. 一般来说, 你去赌场赌博, 赌场是下鞅, 你是上鞅; 人们总是相信股票市场中的股民都是下鞅, 那谁是下鞅呢? 也许大自然是上鞅.

4. 鞅: $\{X_n\}$ 是随机序列, 如果对任何 n 有

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

那么我们说 $\{X_n\}$ 是鞅. 鞅是公平游戏的代名词.

5. 期望不变性: 显然

$$\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}] = \mathbb{E}[X_0],$$

鞅的期望与 n 无关, 称为期望不变性, 它是鞅的一个必要条件.

6. 独立同分布随机变量的积: 设 $Y_1, Y_2, \cdots, Y_n, \cdots$ 是独立同分布随机系列, $\mathbb{E}[Y_n] = 1$. 令

$$X_n = Y_1 Y_2 \cdots Y_n, \ n \geq 1,$$

那么 $\{X_n\}$ 是鞅.

7. 习题: 设 $\{\xi_n\}$ 独立同分布且 $\mathbb{P}(\xi_n = 1) = p$ and $\mathbb{P}(\xi = -1) = q$, $p + q = 1$. 求实数 $a > 0$ 使得 $X_n := a^{\xi_1+\cdots+\xi_n}$ 是鞅.

   对于这个问题, 只要 $\mathbb{E}[a^{\xi_1}] = ap + a^{-1}q = 1$ 就可以了, 也就是二次方程 $a^2p - a + q = 0$, 这个方程有两个解, 一个是 1，另外一个正解可以满足我们的要求了.

8. 一个常识: 随机序列 X_n 表示某甲在一个赌局的 n 时刻的所有, $X_n - X_{n-1}$ 表示第 n 局的所得, 乘一个随机变量 H_{n-1}, $H_{n-1}(X_n - X_{n-1})$ 是第 n 局所得的一个倍数, 可以解释为另外一个人某乙押在甲身上的赌博. 那么

$$Y_n = Y_0 + H_0(X_1 - X_0) + \cdots + H_{n-1}(X_n - X_{n-1}) \tag{4.2}$$

   是乙在这场赌博中直到第 n 局的所有输赢的钱数. 按照常识, 如果甲参加的是公平赌博, 那么不管乙怎么选取他的策略 $\{H_n\}$, 他也是进行一场公平赌博, 他的策略不可能给他带来任何优势.

9. 鞅基本定理: Doob 敏锐地观察到这个常识并且在数学上把它表示出来, 这里的本质是什么呢? 本质是乙和任何人一样, 不可能在第 n 局开始前预知第 n 局的结果, 数学上说, 就是

$$H_{n-1} \in \sigma\{X_1, \cdots, X_{n-1}\},$$

   也就是说, 他只能按照前 $n-1$ 局的情况来确定策略 H_{n-1}. (在很多教材上 H_{n-1} 用 H_n 代替, 那么 H_n 要求关于 $\{X_1, \cdots, X_{n-1}\}$ 可测, 通常称为可料性.)

   定理4.1 (Doob) 在这个假设下, Y_n 也是鞅.

   事实上, 根据条件期望的性质,

$$\begin{aligned}
&\mathbb{E}(Y_n - Y_{n-1}|X_1, \cdots, X_{n-1}) \\
&\qquad = \mathbb{E}(H_{n-1}(X_n - X_{n-1})|X_1, \cdots, X_{n-1}) \\
&\qquad = H_{n-1}\mathbb{E}(X_n - X_{n-1}|X_1, \cdots, X_{n-1}) \\
&\qquad = 0.
\end{aligned}$$

而 $\{Y_1, \cdots, Y_{n-1}\}$ 是由 $\{X_1, \cdots, X_{n-1}\}$ 确定的, 前者的信息少于后者, 因此再用条件期望的性质推出

$$\mathbb{E}(Y_n - Y_{n-1} | Y_1, \cdots, Y_{n-1}) = 0.$$

10. 随机积分: 上面的随机序列 $\{Y_n\}$ 也称为是 $\{H_n\}$ 关于 $\{X_n\}$ 的随机积分. 鞅基本定理是随机分析中最基本的定理, 可以说随机分析理论就是从这个定理开始的. 严格地说, 鞅基本定理是需要条件的, 比如要求每个 H_n (关于样本点) 是有界的.

11. 关于鞅流的鞅: 设 $\{X_n\}$ 是鞅, 那么

$$\mathscr{F}_n = \sigma\{X_1, \cdots, X_n\}$$

称为一个鞅流. 随机序列 $\{Y_n\}$ 被称为是一个关于鞅流 $\{\mathscr{F}_n\}$ 的鞅, 如果对任何 n, Y_n 是关于 $\mathscr{F}_n$ 可测的 (这时我们说 $\{Y_n\}$ 关于流 $\{\mathscr{F}_n\}$ 适应) 且 $\{Y_n\}$ 关于鞅流 $\{\mathscr{F}_n\}$ 是鞅, 即对任何 n

$$\mathbb{E}(Y_n | \mathscr{F}_{n-1}) = Y_{n-1}.$$

由鞅基本定理 $\{H_n\}$ 关于鞅 $\{X_n\}$ 的随机积分是关于鞅流的鞅.

12. 习题: 设 $\{X_n\}$ 是鞅, $\{\mathscr{F}_n\}$ 是鞅流.

    (a) 如果 $\{Y_n : 0 \leq n \leq N\}$ 是关于鞅流的鞅, 那么 $Y_n = \mathbb{E}(Y_N | \mathscr{F}_n)$, $n \leq N$.

    (b) 如果 $Y \in \mathscr{F}_N$, 定义 $Y_n = \mathbb{E}(Y | \mathscr{F}_n)$, 那么 $\{Y_n : 1 \leq n \leq N\}$ 是关于鞅流的鞅.

    (c) 如果 $\{Y_n : 0 \leq n \leq N\}$ 与 $\{Z_n : 0 \leq n \leq N\}$ 都是关于鞅流的鞅且 $Y_N = Z_N$, 那么对所有 $0 \leq n < N$ 有 $Y_n = Z_n$.

13. Doob 分解: 设 $\{X_n\}$ 是鞅, 那么 $\{X_n^2\}$ 是下鞅, 因为

$$\mathbb{E}(X_n^2 - X_{n-1}^2 | \mathscr{F}_{n-1}) = \mathbb{E}[(X_n - X_{n-1})^2 | \mathscr{F}_{n-1}] \geq 0.$$

下面的 Doob 分解定理在离散时间没有什么大用, 但它的连续时间版本在定义随机积分时有不可替代的重要性.

定理4.2 (Doob 分解定理) 存在唯一的增过程 $\{Z_n\}$ 使得 $Z_0=0$, $Z_n \in \mathscr{F}_{n-1}$ 且 $X_n^2 - Z_n$ 是鞅, 即 X_n^2 分解为鞅和可料 ($Z_n \in \mathscr{F}_{n-1}$) 增过程的和

$$X_n^2 = (X_n^2 - Z_n) + Z_n.$$

其实定理对任意下鞅成立. 证明很简单, 归纳地定义 $Z_0 := 0$ (为了保证唯一性),

$$Z_n := Z_{n-1} + \mathbb{E}[X_n^2 - X_{n-1}^2 | \mathscr{F}_{n-1}],$$

然后验证 $\{Z_n\}$ 满足要求就可以了, 唯一性留作习题.

14. 习题: 证明 Doob 分解定理中的唯一性.

15. 鞅表示问题: 设 $\{X_n\}$ 是鞅, 那么关于鞅流的鞅是很多的, 鞅基本定理的一个反问题就是, 关于鞅流的鞅是不是一定是某个随机积分? 也就是说对于任何关于鞅流的鞅 $\{Y_n\}$, 是否存在适应的 $\{H_n\}$ 使得 $\{Y_n\}$ 就是 $\{H_n\}$ 关于 $\{X_n\}$ 的随机积分? 即

$$Y_n - Y_{n-1} = H_{n-1}(X_n - X_{n-1}).$$

鞅表示问题是金融数学中最重要的问题, 它最早是由 Ito 提出并在 Brown 运动的场合证明的. 在离散的场合下, 一般是没有鞅表示的, 但我们后面会看到, 在一些简单的特殊场合时, 鞅表示是成立的.

16. 随机时间: 某个赌徒说:'今天我玩 10 局就走' 是一种策略, 另外一个赌徒说: '今天我赢了 100 元钱就走' 也是一种策略, 前者是固定时间, 后者是随机时间, 他自己也不知道什么时候能停止. 因此随机时间在实际应用中是普遍的. 如果把上面的随机时间称为停时的话, 那么关键的地方是在每一局结束时赌徒可以判断出自己是不是应该停止了. 不是所有的随机时间都有这样的性质, 比如某人说:'如果后面会连输三局我就走', 那么赌徒在一局结束的时候无法判断是不是应该停止, 因为赌徒无从知道后面三局是输是赢, 所以它不是停时. 设 T 是停时, 由停时定义, $\{T=n\}$ 由 $\{X_1, \cdots, X_n\}$ 可以断定. 令 $Y_n = X_{n\wedge T}$, 直观地说, 这是玩到时间 T 停止的策略. 那么

$$\begin{aligned} Y_n - Y_{n-1} &= X_{n\wedge T} - X_{(n-1)\wedge T} \\ &= 1_{\{T\geq n\}} \cdot (X_n - X_{n-1}) \end{aligned}$$

由停时的性质,

$$\begin{aligned}1_{\{T\geq n\}} &= 1 - 1_{\{T\leq n-1\}}\\ &= 1 - \sum_{k=0}^{n-1} 1_{\{T=k\}},\end{aligned}$$

由 $\{X_1, \cdots, X_{n-1}\}$ 决定. 由鞅基本定理, $\{Y_{n\wedge T}\}$ 是鞅.

17. 赌博中的倍增策略: 假设你在赌场参与一个赢的概率为 p 的赌博, 通常 p 会略小于 1/2. 采取一种倍增的策略, 第一次押 K 元, 如果赢了走人, 如果输了下次多押一倍, 依次类推, 那么当你走人的时候, 你总是可以赢 K 元钱, 成为一个所谓的套利机会, 这里的停止时间是随机时间, 服从几何分布, 也就是在独立重复一个成功概率为 p 的随机试验中, 等待成功首次出现的时间, 它几乎肯定是有限的 (但不有界). 这似乎与鞅基本定理所说的公平性矛盾. 实际上不矛盾, 因为鞅基本定理是说赌博的每一局 (有界且固定时间) 是公平的, 倍增策略涉及的是无界的时间. 实际上倍增策略要想成功必须具备两个条件 (1) 无限的赌资; (2) 无限的时间. 如果我们规定赌资是有界的或者赌博的次数有上界, 那么这样的套利机会也是不存在的 (证明需要深刻的数学工具). 所以所有的赌场每局都有最大赌金数的限制, 我在著名的赌城 Las Vegas 看到上限通常是下限的 1000 倍.

18. 倍增策略的数学描述: 设 $\{X_n\}$ 是掷硬币赌博的独立随机变量和, 倍增策略实际上是一个随机积分加上停时, 令

$$\begin{aligned}Y_n - Y_{n-1} &= K2^{n-1}(X_n - X_{n-1});\\ T &= \inf\{n \geq 1 : X_n - X_{n-1} = 1\}.\end{aligned}$$

那么到 T 停止的随机序列 $\{Y_{n\wedge T}\}$ 就是上面说的倍增策略, 它也是鞅, $\mathbb{P}(T < \infty) = 1$ 且 $Y_T = K$, 所以期望的不变性对于随机时间 T 不成立. 由于停时的重要性, 我们讲几个关于停时的例子.

19. 首次超越时间: 某甲在网上卖车, 有人先给了个价是 X_0, 然后甲决定把车卖给下一个出价比 X_0 高的卖主. 设 $X_0, X_1, \cdots, X_n, \cdots$ 是独立同分布随机序列, X_n 是第 n 个人的出价. 令

$$T = \inf\{n \geq 1 : X_n > X_0\},$$

显然 $T > n$ 等价于说

$$X_0 \geq \max\{X_1, \cdots, X_n\}$$

或者说 X_0 是前 $n+1$ 出价中的最高价, 由对称性, 每个人出价最高的概率都是一样的, 因此如果 X_0 的分布是连续型的, 那么

$$\mathbb{P}(T > n) = \frac{1}{n+1}.$$

由此推出 $\mathbb{E}[T] = \sum_n \mathbb{P}(T > n) = \infty$.

20. 期望不变性: 但是鞅在在停时上不一定有期望不变性, 在讨论倍增问题的时候已经提到过. 实际上鞅在一个停时上有没有期望不变性是非常重要和困难的问题, 要证明在某些停时上有期望不变性需要用到真正的数学.

21. 习题: 考虑上面的零点出发的简单随机游动 X_n 和首次回归时间 T, 证明: $Y_n = X_n^2 - n,\ n \geq 0$, 也是鞅, 然后证明 $\{Y_n\}$ 在停时 T 上没有期望不变性.

22. 赌徒输光问题的鞅解法: 这是一个简单而历史悠久的赌博问题, 甲乙两个赌徒以掷硬币每次一元的输赢赌博, 甲乙各带了 a 元和 b 元钱来参加赌博, 中间不容许借钱 (就是金融里的所谓自融资), 赌博到某人输光即止, 问甲输光的概率是多少? 用 X_n 表示甲在 n 局赌博后剩下的钱, 那么 $X_0 = a$, 用 T 表示其中某人输光的时间,

$$T = \inf\{n \geq 1 : X_n \in \{0, a+b\}\},$$

我们要算 $\mathbb{P}(X_T = 0)$. 显然, 因为 X_T 只有取 0 或者 $a+b$ 两种可能, 故

$$\mathbb{E}[X_T] = 0 \cdot \mathbb{P}(X_T = 0) + (a+b)\mathbb{P}(X_T = a+b),$$

因此

$$\mathbb{P}(X_T = 0) = 1 - \frac{\mathbb{E}[X_T]}{a+b},$$

故我们只需算 $\mathbb{E}[X_T]$ 就行了. 如果期望不变性在 T 上成立, 那么 $\mathbb{E}[X_T] = \mathbb{E}[X_0] = a$, 这样推出

$$\mathbb{P}(X_T = 0) = \frac{b}{a+b}.$$

我们可以证明期望不变性在 T 上成立.

23. 习题: 有一个圆形城墙有 $m+1$ 个城门, 以 $0,1,\cdots,m$ 依次编号. 某甲从 0 号门出发, 掷一个硬币, 视硬币正反面选择走向左右的城门, 然后继续掷硬币, 继续选择, 一直到他走遍所有城门后立刻走出, 问他从 i 号城门出来的概率是多少? $1\le i\le m$.

24. 极限和期望的交换 (补充): 要证明上面所说的期望不变性在 T 上成立, 要问何时极限与期望可以交换? 因为由 Doob 的基本定理, $\{X_{n\wedge T}\}$ 是鞅, 所以有

$$\mathbb{E}[X_{n\wedge T}]=\mathbb{E}[X_0]$$

对任何 n 成立. 另外

$$\lim_{n\to\infty}X_{n\wedge T}=X_T,$$

故而只要两者可以交换, 那么

$$\mathbb{E}[X_0]=\lim_n\mathbb{E}[X_{n\wedge T}]=\mathbb{E}[\lim_n X_{n\wedge T}]=\mathbb{E}[X_T].$$

定理4.3 (有界收敛定理) 如果随机序列 $\{X_n\}$ 收敛于 X 且存在常数 C 使得对任何 n 有 $|X_n|\le C$, 那么

$$\mathbb{E}[\lim_n X_n]=\lim_n\mathbb{E}[X_n].$$

利用这个定理, 上面的交换就可以了, 因为 $\{X_{n\wedge T}\}$ 是有界的.

25. 习题: 设在上面的赌徒输光问题中, 甲赢乙赢的概率不同, 各为 p 和 $q=1-p$, 再用鞅方法计算甲输光的概率.

26. 习题: 设 $\{X_n\}$ 是简单随机游动, $H_n\in\sigma(X_1,\cdots,X_n)$ 且存在与 n 和 ω 无关的常数 C 使得 $|H_n(\omega)|\le C$, 再设 T 是停时且 $\mathbb{E}[T]<\infty$, Y_n 是 H 关于 X 的随机积分, 证明: Y_n 在 T 上也有期望不变性: $\mathbb{E}[Y_T]=\mathbb{E}[Y_0]$. 此题是对于倍增策略这个例子的一个补充说明. 提示: 应用控制收敛定理.

27. 习题: 用鞅方法来求零点出发的非对称 ($p\ne 1/2$) 随机游动 $\{S_n\}$ 首次抵达 k 的时间 T_k 的分布以及首次返回零点的时间 T 的分布.

28. 母函数与鞅性: 把鞅方法和母函数方法结合起来用是很强大的. 设 $k>0$, $z\in(0,1)$, 因为

$$\mathbb{E}[z^{\xi_n}]=pz+qz^{-1},\ \mathbb{E}[z^{-\xi_n}]=pz^{-1}+qz$$

所以 $\left\{(pz+qz^{-1})^{-n}z^{S_n}\right\}$ 与 $\left\{(pz^{-1}+qz)^{-n}z^{-S_n}\right\}$ 是鞅. 由强大数定律, 当 $p>q$ 时, S_n 趋于 $+\infty$, 所以 $\mathbb{P}(T_k<\infty)=1$; 而当 $q>p$ 时, S_n 趋于 $-\infty$. 现在有期望不变性

$$\mathbb{E}\left[(pz+qz^{-1})^{-T_k\wedge n}z^{S_{T_k\wedge n}}\right]=1.$$

当 $n\to\infty$ 时, 若 $T_k<\infty$, $S_{T_k\wedge n}$ 的极限是 S_{T_k}, 反之极限是 $-\infty$. 另外必须保证左边方括号里的量有界才能交换极限和期望. 这只要 $pz+qz^{-1}\geq 1$ 且 $\{S_{T_k\wedge n}\}$ 下有界, 但是 $\{S_{T_k\wedge n}\}$ 是上有界的而不下有界, 因此我们上面取鞅的方式不对, 应该取另一个鞅

$$\left\{(pz^{-1}+qz)^{-n}z^{-S_n}\right\},$$

这时只要 $pz^{-1}+qz>1$ 就可以了, 而当 $p>q$ 时, $pz^{-1}+qz$ 在 $(0,1)$ 上大于 1, 由期望不变性

$$\mathbb{E}\left[(pz^{-1}+qz)^{-T_k\wedge n}z^{-S_{T_k\wedge n}}\right]=1$$

再让 n 趋于无穷得

$$1=z^{-k}\mathbb{E}\left[(pz^{-1}+qz)^{-T_k};T_k<\infty\right]=z^{-k}\mathbb{E}\left[(pz^{-1}+qz)^{-T_k}\right],$$

取 $0<y<1$, 解方程 $pz^{-1}+qz=y^{-1}$ 得

$$\mathbb{E}\left[y^{T_k}\right]=\left(\frac{1-\sqrt{1-4pqy^2}}{2qy}\right)^k.$$

把 p,q 位置互换就得到 T_{-k} 的母函数. 为了算 T 的母函数, 让随机游动走一步, 到 1 或者 -1, 这时 T 就分别相当于零出发的 T_{-1} 和 T_1,

$$\mathbb{E}[z^T]=pz\mathbb{E}[z^{T_{-1}}]+qz\mathbb{E}[z^{T_1}]=1-\sqrt{1-4pqz^2}.$$

第 五 讲　马氏链

摘要 马氏链是最简单最直观的随机过程, 在数学和过程计算等领域有广泛的应用, 其中的 Markov 性是 150 年前由俄罗斯数学家 Markov 提出来的, 是随机游动的独立增量性的抽象化. 在这一讲中, 我们将主要介绍时间其次的马氏链的概念, 这样的马氏链结构简单, 由一个转移矩阵确定, 从这个转移矩阵出发, 引出可达性, 不可约性, 常返暂留性, 以及重要的平稳性.

关键词
- 马氏链
- 转移函数和转移矩阵
- 不可约性
- 常返与暂留性
- 正常返与平稳性
- Galton-Watson 分支过程
- 电路网络与马氏链

1. 马氏链是非常重要的随机模型, 它已经有 150 年的历史, 前面所说的随机游动是马氏链的特殊情况. 首先马氏链是一个随机序列

$$\{X_0, X_1, X_2, \cdots, X_n, \cdots\}$$

它的状态空间也就是随机序列的值域 S 可以是一个有限集合也可以是可数集合, 欧氏空间的整数格点 $\mathbb{Z}^d$ 是常见的情况, 直观地说, 如果在时间 n, X_n 位于位置 $x \in S$ 时, 事件 $\{X_{n+1} = y\}$ 与过去独立, 且概率与时间 n 无关, 等于 $p(x, y)$. 这时 $S \times S$ 上的函数 $p(x, y)$ 称为马氏链的转移函数或者转移概率, 它显然满足

$$p(x, y) \geq 0, \ \sum_{y \in S} p(x, y) = 1.$$

更严格的数学定义如下: 对任何

$$n \geq 0, \ x, y, x_{n-1}, \cdots, x_1, x_0 \in S,$$

有

$$\mathbb{P}(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \cdots, X_0 = x) = p(x, y). \tag{5.1}$$

这就是马氏性, 给定现在的状态, 将来与过去是独立的.

2. 转移函数和初始分布: 决定一个随机变量的分布的当然是分布函数, 而决定一个随机序列的分布的是其有限维分布族, 也就是 $(X_0, X_1, \cdots, X_n)$ 的分布, 由乘法公式

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

右边是由转移函数和 X_0 的分布 (初始分布) 决定的. 给状态空间一个顺序, 转移函数 $p(x, y)$ 可以写成矩阵的样子来直观表达

$$\mathbf{P} = (p(x, y))_{x, y \in S},$$

称为转移矩阵. 由概率的可加性推出, n 时间 X_n 的分布

$$\mathbb{P}(X_n = x) = \sum_{x_0, x_1, \cdots, x_{n-1} \in S} \mathbb{P}(X_0 = x_0) p(x_0, x_1) \cdots p(x_{n-1}, x).$$

利用此公式可以计算一步条件概率

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

推出

$$\mathbb{P}(X_{n+1}=y|X_n=x)=\mathbb{P}(X_1=y|X_0=x)=p(x,y),$$

这是为什么把 $p(x,y)$ 称为 (一步) 转移概率的理由. 再计算

$$\begin{aligned}\mathbb{P}(X_2=y|X_0=x)&=\frac{\mathbb{P}(X_2=y,X_0=x)}{\mathbb{P}(X_0=x)}\\&=\frac{\sum_{x_1\in S}\mathbb{P}(X_2=y,X_1=x_1,X_0=x)}{\mathbb{P}(X_0=x)}\\&=\sum_{x_1\in S}p(x,x_1)p(x_1,y),\end{aligned}$$

同理,

$$\mathbb{P}(X_{n+2}=y|X_n=x)=\mathbb{P}(X_2=y|X_0=x),$$

它直观地理解为从 x 出发两步走到 y 的概率, 称为二步转移概率, 记为 $p^{(2)}(x,y)$. 它也形成一个转移矩阵, 记为 $\mathbf{P}^{(2)}$. 根据矩阵的乘法定义立刻可以看出

$$\mathbf{P}^{(2)}=\mathbf{P}^2,$$

右端是矩阵的平方. 类似地可以计算从 x 出发 k 步走到 y 的概率

$$\begin{aligned}p^{(k)}(x,y)&=\mathbb{P}(X_k=y|X_0=x)\\&=\sum_{x_1\in S}\cdots\sum_{x_{k-1}\in S}p(x,x_1)\cdots p(x_{k-1},y),\end{aligned}$$

称为 k 步转移概率, 形成的转移矩阵满足

$$\mathbf{P}^{(k)}=\mathbf{P}^k.$$

约定 $\mathbf{P}^{(0)}$ 是单位矩阵. 自然地对任何 $n,m\geq 0$, $x,y\in E$,

$$p^{(n+m)}(x,y)=\sum_{z\in E}p^{(n)}(x,z)p^{(m)}(z,y), \tag{5.2}$$

此方程就是 Chapman-Kolmogorov 方程. 转移函数对于一个马氏链起主导作用, 因此我们通常不区分马氏链与它的转移函数.

3. 例 1: 最自然的马氏链是随机游动. 设 $\mathbb{Z}^d$ 是 d 维整数格点, 每个点有 $2d$ 个相邻点. 马氏链总是等可能地选择一个相邻点走去, 也就是说其转移函数为

$$p(x,y)=\begin{cases}\frac{1}{2d}, & \text{如果 } y \text{ 与 } x \text{ 相邻};\\ 0, & \text{否则}.\end{cases}$$

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

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

这是一个关于两种液体混合的概率模型.

5. 例 3: 非对称随机游动. 一维整数点上的转移矩阵

$$\mathbf{P}:=\begin{bmatrix}\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ \cdots & q_{-1} & 0 & p_1 & 0 & 0 & \cdots\\ \cdots & 0 & q_0 & 0 & p_2 & 0 & \cdots\\ \cdots & 0 & 0 & q_1 & 0 & p_3 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\end{bmatrix}$$

导出一个马氏链, 其中 $q_x+p_x=1$. 当 $p_x=p$ 时, 就是通常的非对称随机游动. 当 $p_x\equiv 01$ 时, 称为向右平移.

6. 例 4: 具吸收壁的随机游动. 设 Markov 链 X 有状态空间 $E=\{0,1,\cdots,r\}$, 转移矩阵

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

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

7. 例 5: A, B 两人比赛, A 赢的概率是 p, 输的概率是 $q=1-p$. (1) 如果规则是比赛到某人胜出两场为赢, 求 A 赢的概率. (2) 如果规则是某人连续赢两场为

赢. 我们以此分别构造两个马氏链. 对规则 (1). 用 X_n 表示到第 n 局结束时, A 的净赢局数. 显然 X_n 是个具吸收壁 Markov 链, 状态依次为 $-2,-1,0,1,2$, 其转移矩阵为

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

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

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

8. 例 6: 图上简单随机游动. 设 S 是一个局部有限图的顶点集合 (可以是可数的), 对任何 $x, y \in E$, 若 x, y 有线直接连接, 则令 $S(x,y) = 1$ 否则 $S(x,y) = 0$. 则 S 是对称的. 设
$$p_{x,y} = \frac{S(x,y)}{\sum_{z \in E} S(x,z)},$$
则 $(p_{x,y} : x, y \in E)$ 定义了 E 上一个转移函数, 对应的 Markov 链称为是图上简单随机游动. 它总是等可能地选择一个邻居走过去.

9. 可达与不可约: 说状态 x 可达 y, 是指存在 $n \geq 0$ 使得
$$p^{(n)}(x,y) > 0,$$
可达这个概念描述状态 x 是不是可能走到状态 y. 这等价于存在一条从 x 到 y 的通畅道路
$$x = x_0 \to x_1 \to \cdots \to x_{n-1} \to x_n = y$$
使得对任何 $1 \leq i \leq n$, 有 $p(x_{i-1}, x_i) > 0$. 我们总是约定 x 是可达自己的. 如果两个状态互相可达, 称为互达. 一个马氏链称为是不可约的, 如果它的任何

两个状态都是互达的. 在上面的例子中, 例 1,2 中的马氏链都是不可约的. 如果对所有 x, $0 < p_x < 1$, 那么例 3 中的马氏链是不可约的. 图上简单随机游动是不可约当且仅当图是连通的. 有吸收态的马氏链肯定不是不可约的. 显然当 X 的某一个 k 步转移矩阵的所有元素都是正的时候, 那么马氏链一定是不可约的. 这是不可约的一个充分条件, 是不是必要的呢? 同学可以自己思考一下. 互达是一个等价关系, 可以依照互达对状态空间分类为互达等价类. 另外一个分类的概念是闭性. S 的一个子集 E 称为是闭集, 如果马氏链从 E 中的状态不可达 E^c 中的状态. 如果 E 是闭集, 那么把转移概率限制在 E 上仍然是一个转移函数, 因为对任何 $x \in E$, 有

$$\sum_{y \in E} p(x,y) = 1,$$

也就是说 X 作为 E 上的过程仍然是马氏链. 给定状态 x, 用 $C(x)$ 表示 x 由可达的状态全体. 那么 $C(x)$ 是闭的. 不可约等价类不一定是闭的, 闭集也不一定是不可约的.

10. 从 x 出发的马氏链: 对于任何的状态 $x \in S$, 用 $\mathbb{P}^x$ 表示从 x 出发的马氏链的概率, 或者说
$$\mathbb{P}^x(A) = \mathbb{P}(A|X_0 = x).$$

11. 常返与暂留: 在一个马氏链中, 有些状态, 马氏链会无限多次地访问, 称为常返态, 而另外一些状态, 马氏链最多访问有限多次, 称为暂留态. 严格地说, 状态 x 称为常返, 如果
$$\mathbb{P}^x(\{\text{存在无限多个 } n \text{ 有 } X_n = x\}) = 1;$$
称为暂留, 如果上面的概率是 0. 首先我们来证明任何的状态, 或者常返或者暂留. 用 $T = T_x$ 表示状态 x 的首中时,
$$A_n := \{\omega \in \Omega : \sum_{k \geq 1} 1_{\{X_k(\omega) = x\}} \geq n\},$$
即 A_n 是访问 x 至少 n 次的轨道全体.那么
$$\begin{aligned}\mathbb{P}^x(A_n) &= \mathbb{P}^x(A_n, T < \infty) \\ &= \sum_{k \geq 1} \mathbb{P}^x(A_n, T = k)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k \geq 1} \mathbb{P}^x(A_n, X_k = x, X_{k-1} \neq x, \cdots, X_1 \neq x) \\
&= \sum_{k \geq 1} \mathbb{P}^x(\sum_{j \geq 1} 1_{\{X_{k+j}=x\}} \geq n-1, X_k = x, X_{k-1} \neq x, \cdots, X_1 \neq x) \\
&= \sum_{k \geq 1} \mathbb{P}^x(A_{n-1}) \mathbb{P}(T = k) \\
&= \mathbb{P}(A_{n-1}) \mathbb{P}(T < \infty) \\
&= [\mathbb{P}(T < \infty)]^n.
\end{aligned}
$$

用 A_∞ 表示无穷多次访问 x 的轨道全体, 那么

$$
\mathbb{P}^x(A_\infty) = \lim_n [\mathbb{P}(T_x < \infty)]^n,
$$

推出此概率非 1 即 0. 记

$$
f_x = \mathbb{P}^x(T_x < \infty).
$$

那么 x 常返或者暂留分别等价于 f_x 等于 1 或者小于 1.

12. 状态的位势: 状态 x 的位势是马氏链访问 x 的平均次数, 即

$$
u(x) := \mathbb{E}^x \left[\sum_{n \geq 1} 1_{\{X_n = x\}} \right] = \sum_{n \geq 1} p(x, x).
$$

显然 $u(x) < \infty$ 蕴含着

$$
\mathbb{P}(\sum_{k \geq 1} 1_{\{X_k = x\}} < \infty) = 1.
$$

反过来对吗? 也是对的, 先证明首次通过公式: 对 $n \geq 1$,

$$
\begin{aligned}
p^{(n)}(x, y) &= \mathbb{P}^x(X_n = y) \\
&= \sum_{k=1}^n \mathbb{P}^x(X_n = y, T_y = k) \\
&= \sum_{k=1}^n \mathbb{P}^y(X_{n-k} = y) \mathbb{P}^x(T_y = k) \\
&= \sum_{k=1}^n p^{(n-k)}(y, y) \mathbb{P}^x(T_y = k).
\end{aligned}
$$

令 $x=y$, 两边对 n 求和

$$\begin{aligned}u(x) &= \sum_{n\geq 1}\sum_{k=1}^{n} p^{(n-k)}(x,x)\mathbb{P}^x(T=k)\\ &= \sum_{k\geq 1}\mathbb{P}^x(T=k)\sum_{n\geq k} p^{(n-k)}(x,x)\\ &= \mathbb{P}^x(T<\infty)(1+u(x)),\end{aligned}$$

推出 $u(x)=\infty$ 当且仅当 $f_x=1$.

13. 一个状态不是常返就是暂留, 这是对状态空间的一种分类. 一个马氏链的状态一般有常返的也有暂留的. 比如吸收壁的随机游动, 两个吸收壁是常返的, 而其它状态是暂留的. 但是如果两个状态是互达的, 那么它们或者都是常返的或者都是暂留的. 这可以从位势的性质推出. 事实上, 设 x,y 互达, 那么存在 n_1,n_2 使得
$$p^{(n_1)}(x,y)>0,\ p^{(n_2)}(y,x)>0,$$
再有 Chapman-Kolmogorov 方程推出对任何 n 有
$$p^{(n_1+n+n_2)}(x,x)\geq p^{(n_1)}(x,y)p^{(n)}(y,y)p^{(n_2)}(y,x).$$
这说明 $u(x)<\infty$ 蕴含着 $u(y)<\infty$, 反之亦然. 一种由互达状态传递的性质称为类性质, 那么常返暂留是类性质. 对于一个不可约马氏链, 或者所有状态常返, 称为常返马氏链; 或者所有状态暂留, 称为暂留马氏链. 对于例 3 中一维的随机游动, 由 Stirling 逼近 $n!\sim(n/e)^n\sqrt{2\pi n}$ 得
$$p^{(2n)}(0,0)=\binom{2n}{n}p^nq^n\sim\frac{(4pq)^n}{\sqrt{\pi n}},$$
即 $p=q=1/2$ 时, 位势是发散的, 常返; 否则位势收敛, 暂留.

14. 我们还可以证明, 一个常返状态 x 可达的状态 y 肯定是常返的, 而且 y 也可达 x. 事实上, 对任何 $n\geq 1$,
$$\begin{aligned}p^{(n)}(x,y) &= \mathbb{P}^x(X_n=y)\\ &= \sum_{k\geq 1}\mathbb{P}^x(X_n=y, X_{n+i}\neq x, 0<i<k, X_{n+k}=x)\end{aligned}$$

$$= \sum_{k \geq 1} \mathbb{P}^x(X_n = y)\mathbb{P}^y(T_x = k)$$
$$= p^{(n)}(x,y)\mathbb{P}^y(T_x < \infty),$$

因为 x 可达 y, 故

$$\mathbb{P}^y(T_x < \infty) = 1,$$

此事蕴含着 y 可达 x, 所以 y 也是常返的. 由此可以推出, 一个常返状态可达的所有状态都是常返的, 而且是不可约的马氏链.

15. 有限状态马氏链: 有限状态的马氏链比较简单. 首先我们可以证明有限状态马氏链一定有一个常返状态. 这在直观上是显然的. 理论上的证明稍复杂些, 固定 $x \in S$, 首先

$$\sum_{y \in S} p^{(n)}(x,y) = 1,$$

因为 S 有限, 故存在 $y \in S$ 使得

$$\sum_{n \geq 1} p^{(n)}(x,y) = \infty.$$

应用首次通过公式得右边为

$$(1 + u(y))\mathbb{P}^x(T_y < \infty),$$

因此 y 是常返的. 而 y 可达的状态全体组成一个常返马氏链, 所以有限状态马氏链包含一个常返的子链. 进一步看出, 一个不可约的有限状态马氏链一定是常返的.

16. 正常返: 一个常返状态 x 是指从 x 出发的回归时间 T_x 几乎肯定是有限的. 记 π_x 是平均回归周期, 即

$$\pi_x := \mathbb{E}^x[T_x].$$

称常返状态 x 为正常返, 如果 $\pi_x < \infty$, 否则称为零常返. 正常返也是一个类性质. 下面的量

$$\frac{1}{n}\sum_{k=1}^{n} p^{(k)}(x,x) = \mathbb{E}^x\left[\frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k = x\}}\right]$$

可以理解为马氏链访问 x 的平均频率, 直观地, 周期是频率的倒数, 因此

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} p^{(k)}(x,x) = \frac{1}{\pi_x}. \tag{5.3}$$

正整数集 $\{n : p^{(n)}(x,x) > 0\}$ 的公因子称为是 x 的周期, 周期是 1 的状态称为非周期. 周期是类性质. 当 x 是非周期时, 更强大的极限定理成立

$$\lim_n p^{(n)}(x,x) = \frac{1}{\pi_x}. \tag{5.4}$$

这时我们说这个状态有平稳性.

17. 平稳与遍历: 下面的定理是非常经典的.

定理5.1 一个不可约非周期有限状态马氏链必定是正常返的或者说平稳的, 且对任何 $x, y \in S$ 有

$$\lim_n p^{(n)}(x,y) = \pi_y. \tag{5.5}$$

当然 $\sum_y \pi_y = 1$, 所以 $\{\pi_y : y \in S\}$ 被称为是马氏链的平稳分布.

对于一个有限状态马氏链, 如果存在 n 使得 $\mathbf{P}^n$ 的所有元素严格正, 那么它是不可约非周期的马氏链. 因为

$$p^{(n+1)}(x,y) = \sum_{z\in S} p^{(n)}(x,z)p(z,y),$$

让 n 趋于无穷, 对任何 $y \in S$,

$$\pi_y = \sum_{z\in S} \pi_z p(z,y).$$

把 $\pi = (\pi_Z)$ 写成行向量

$$\pi = \pi\mathbf{P}.$$

也就是说平稳分布 π 是上面线性方程的唯一解 (和等于 1). 换句话说 1 是转移矩阵 $\mathbf{P}$ 的特征值, π 是它的左特征向量.

18. 例: 某人有 r 把伞放在家或者办公室用于来往于家与办公室之间. 当且仅当天下雨且手边有伞时, 带一把伞走, 到达后放下. 下雨的概率等于 p. 用 X_n 表示他第 n 次出门时手边伞的数目. 那么这是一个马氏链, 状态空间是

$$S = \{0, 1, 2, \cdots, r\},$$

如果它现在要出门, 手边有 $x > 0$ 把伞, 那么对面有 $r - x$ 把伞, 如果天下雨, 他要带把伞, 那么下次手边有 $r - x + 1$ 把伞; 天不下雨的话, 下次手边会有 $r - x$ 把伞. 当 $x = 0$ 时, 他无伞可拿, 不管下不下雨, 下次手边都是 r 把伞. 转移矩阵为

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

这是一个不可约非周期的马氏链. 解方程 $\pi = \pi\mathbf{P}$ 得

$$\pi_0 = \frac{q}{q+r}, \pi_1 = \cdots = \pi_r = \frac{1}{q+r}.$$

我们来算一下他在时刻 n 会挨雨淋的概率, 也就是天下雨而他手边没有伞的概率, 等于 $p \cdot \mathbb{P}^x(X_n = 0)$, 它的极限是

$$p \cdot \pi_0 = \frac{pq}{q+r}.$$

当 $r \geq 5$ 这个概率不超过 5%.

19. Galton-Watson 分支过程: 考虑一个物种繁殖的模型, 称为 Galton-Watson 分支过程. 设 X 是非负整数值随机变量, 为了排除平凡的情况, 设 X 不恒等于常数, 表示物种个体可能繁殖的数量, 用 f 表示它的母函数, $m = \mathbb{E}X$ 表示其平均繁殖数, 当然 $m = f'(1)$. 现在设有概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 和独立随机变量集 $\{X_k^n : k \geq 1, n \geq 1\}$ 而且它们都与 X 有相同分布. 直观地说个体是独立繁殖且能力没有差别. 定义 n 代物种数量,

$$\begin{aligned} \xi_0 &= 1, \\ \xi_n &= \sum_{i=1}^{\xi_{n-1}} X_i^n, \ n \geq 1. \end{aligned}$$

可以看出 X_k^n 和 ξ_{n-1} 是独立的, 且 $\{\xi_n\}$ 是一个马氏链, 它的状态是非负整数集, 状态 0 是吸收态, 也称为灭绝状态. 显然 $\{\xi_{n-1}=0\}\subset\{\xi_n=0\}$, 概率 $q:=\mathbb{P}(\bigcup_n\{\xi_n=0\})=\lim_n\mathbb{P}(\xi_n=0)$ 称为是物种的灭绝概率. 我们要计算灭绝概率, 先来算 ξ_n 的母函数.

$$\begin{aligned}\mathbb{E}(t^{\xi_n}|\xi_{n-1}=k)&=\mathbb{E}(t^{\sum_{i=1}^k X_i^{(n)}}|\xi_{n-1}=k)\\&=\mathbb{E}(t^{\sum_{i=1}^k X_i^{(n)}})=\prod_{i=1}^k\mathbb{E}t^{X_i^{(n)}}=(f(t))^k,\end{aligned}$$

因此 $\mathbb{E}(t^{\xi_n}|\xi_{n-1})=f(t)^{\xi_{n-1}}$, 且

$$f_n(t):=\mathbb{E}t^{\xi_n}=\mathbb{E}[\mathbb{E}(t^{\xi_n}|\xi_{n-1})]=\mathbb{E}[f(t)]^{\xi_{n-1}}=f_{n-1}(f(t)).$$

令 $f^{\circ 0}(t):=t$, 用 $f^{\circ n}$ 表示 f 的 n 次复合, 那么 $f_n=f^{\circ n}$. 显然 $\mathbb{P}(\xi_n=0)=f_n(0)=f(f_{n-1}(0))$, 由 f 的连续性 $q=f(q)$. 因此灭绝概率 q 是方程 $t=f(t)$ 在 $[0,1]$ 上的根, 而函数 f 是递增凸函数, $f(0)=\mathbb{P}(X=0)\geq 0$, $f(1)=1$, 故此方程最多只有两个根, 1 总是一个根. 而对任何 $n\geq 1$, $f(f_n(0))\geq f_n(0)$, 因此 q 是方程小的根. 故分两种情况,

(1) 当 $m=f'(1)\leq 1$ 时, 此方程只有唯一的根 1, 这时物种依概率 1 灭绝.

(2) 当 $m=f'(1)>1$ 时, $0\leq q<1$. $q=0$ 当且仅当 $\mathbb{P}(X=0)=0$.

这个模型虽然简单但很受关注.

20. 电路和马氏链: 一个电路网络自然地诱导出一个马氏链的转移函数. 如果把图 S 看成一个电路, 函数 $c(x,y)$ 是边的电导, 它是电阻 $r(x,y)$ 的倒数. (当 $c(x,y)=0$ 时, 理解为两点之间没有边相连, 或者电阻是无穷大.) 定义转移函数

$$p(x,y):=\frac{c(x,y)}{\sum_{y\in S}c(x,y)}.$$

它所对应的马氏链是电路上的自然的马氏链. 下面我们看看马氏链和电阻电流电压的联系. 取非空子集 $A,B\subset S$, 满足 $A\cap B=\varnothing$ 且 $(A\cup B)^c$ 有限连通. 加一个电源连接 A,B 使得 A 上每个点的电压是 1, B 上每个点的电压是 0. 由物理的观察, 存在电压函数, 也就是 S 上的一个函数 v 使得 $v|_A=1$, $v|_B=0$, 及电流函数 i, 它是边 E 上的函数, 和电阻率一起满足下面的 Kirchhoff 定律和 Ohm 定律:

(1) Kirchhoff 节点定律: 对于不直接联在电池上的点 x, 即 $x \notin A \cup B$, 有

$$\sum_{y \sim x} i(x,y) = 0,$$

即流出的电流和等于流入的电流和.

(2) Kirchhoff 回路定律: 如果 $x = x_0 \sim x_1 \sim \cdots \sim x_n = x$ 是一个回路, 那么

$$\sum_{i=1}^{n} i(x_{i-1}, x_i) r(x_{i-1}, x_i) = 0.$$

(3) Ohm 定律: 如果 $x \sim y$, 那么 $v(x) - v(y) = i(x,y) r(x,y)$.

Ohm 定律蕴含着 i 是 E 上的反对称函数: $i(x,y) = -i(y,x)$. 另外 Ohm 定律等价于 Kirchhoff 回路定律. 事实上, Ohm 定律显然蕴含着 Kirchhoff 回路定律, 反过来, 如果 Kirchhoff 回路定律成立, 取 $b \in B$, 对任何 $x \notin A \cup B$, 存在 x 到 b 的路径 $b = x_0 \sim x_1 \sim \cdots \sim x_n = x$, 定义

$$v(x) := \sum_{i=1}^{n} i(x_i, x_{i-1}) r(x_i, x_{i-1}).$$

由 Kirchhoff 回路定律, $v(x)$ 与选择的路径无关, 因为 $v|_B = 0$, 它也与 b 的选择无关. 容易推出 v, i, r 满足 Ohm 定律. 下面的结论说明电压可以用 Markov 链来表示.

(a) 电压函数 v 在 $(A \cup B)^c$ 上调和, 即对任何 $x \notin A \cup B$, 有

$$v(x) = \sum_{y \in S} p(x,y) v(y).$$

事实上, 由节点定律, $\sum_{y \sim x} i(x,y) = 0$, 再由 Ohm 定律 $i(x,y) = (v(x) - v(y)) c(x,y)$, 故 $\sum_{y \sim x} (v(x) - v(y)) c(x,y) = 0$. 因此 $v(x)c(x) = \sum_{y \sim x} c(x,y) v(y)$, 推出 v 在 x 处调和.

(b) 对任何 $x \in S$, $v(x) = \mathbb{P}^x(\tau_A < \tau_B)$, 其中 τ 表示集合的进入时. 此命题的证明留作习题.

21. 有效电阻与有效电导: 上面我们看到电压可以表示 $\{\tau_A < \tau_B\}$ 的概率, 下面我们来看看怎么用电阻或者电导来表示概率. 因为电流 i 反对称, 故

$$\sum_{x,y \in S: x \sim y} i(x,y) = \sum_{(x,y) \in E} i(x,y) = 0.$$

由 Kirchhoff 节点定律,

$$\sum_{x\in A}\sum_{y\sim x} i(x,y) = \sum_{x\in B}\sum_{y\sim x} i(y,x),$$

左边是由 A 流出的电流, 右边是流入 B 的电流, 他们是相等的, 记它为 $i(A,B)$. 整个电路的电压是 1, 所以 A, B 间的电阻是 $i(A,B)^{-1}$, 或者说电导是 $i(A,B)$, 分别称为是 A,B 间的有效电阻和有效电导, 记为 $R_{\text{eff}}(A,B)$, $C_{\text{eff}}(A,B)$. A,B 间有效电导等于在 A,B 两端加单位电压后的电流量, 有效电阻等于从 A 流入单位电流时 A,B 两端的电压差. 有效电导 (或电阻) 可以通过以下三种电路等价变换的方式来计算, 请读者自行用 Ohm 定律和 Kirchhoff 定律推导.

(a) 串联电路:

(b) 并联电路:

(c) 星形电路:

用 T 表示 $a\in S$ 的首中时, $T := \inf\{n\geq 1 : X_n = a\}$. 那么 a 出发回到 a 前到达 B 的概率

$$\mathbb{P}^a(T>\tau_B) = \frac{C_{\text{eff}}(a,B)}{c(a)}. \tag{5.6}$$

事实上, 因为 τ_B 和 T 都不会是零, 所以我们可以从时刻 1 开始观察,

$$\begin{aligned}
\mathbb{P}^a(T>\tau_B) &= \sum_{x\in S}\mathbb{P}^a(T>\tau_B|X_1=x)p(a,x)\\
&= \sum_{x}\mathbb{P}^x(\tau_a>\tau_B)p(a,x)\\
&= \sum_{x}(1-v(x))p(a,x)\\
&= \frac{1}{c(a)}\sum_{x}(v(a)-v(x))c(a,x)\\
&= \frac{1}{c(a)}\sum_{x} i(a,x) = \frac{C_{\text{eff}}(a,B)}{c(a)}.
\end{aligned}$$

下面的例子说明怎样用有效电导来计算概率.

22. 例: 设 a_1,a_2,a_3,a_4 是正方形的四个顶点, 每边的电导都是 1. 那么相邻两点 a_1,a_2 间的有效电导为 $C_{\text{eff}}(a_1,a_2) = 1+\frac{1}{3} = \frac{4}{3}$. 因此 a_1 出发回到 a_1 前到达

a_2 的概率 $\mathbb{P}^{a_1}(\tau_{a_2} < \tau_{a_1}^+) = \frac{C_{\text{eff}}(a_1,a_2)}{c(a_1)} = \frac{2}{3}$. 而相对的两点 a_1, a_3 间的有效电导为 $C_{\text{eff}}(a_1,a_3) = \frac{1}{2} + \frac{1}{2} = 1$, 而对应的概率为 $\frac{1}{2}$. 这个例子比较简单, 上面的概率可直接计算.

23. 例: 设有一个正方体网络, 每边的电导依然是 1. 让 a, b 是相邻两点. 通过一系列的等价变换, $C_{\text{eff}}(a,b) = \frac{12}{7}$, a 出发回到 a 前到达 b 的概率为

$$\mathbb{P}^a(\tau_b < \tau_a^+) = \frac{C_{\text{eff}}(a,b)}{c(a)} = \frac{4}{7}.$$

对于这个例子, 这个概率的直接计算几乎是不可能的.

24. 常返性和电导: 利用 (5.6) 取 B_n 递减保持其余集有限连通趋于空集, 那么 $\tau_{B_n} \uparrow \infty$, 所以

$$\mathbb{P}^a(T = \infty) = \frac{\lim_n C_{\text{eff}}(a, B_n)}{c(a)}, \tag{5.7}$$

记 $C(a,\infty) = \lim_n C_{\text{eff}}(a,B_n)$, 它可以理解为 a 到无穷的电导, 这个电导等于零直观地解释为无穷远处是过不去的, 而从上面的公式推出电导为零当且仅当马氏链是常返的.

25. 依次等可能地往 M 个盒子里放球, ξ_n 表示放入第 n 个球后空盒的个数. 证明 $\{\xi_n\}$ 是 Markov 链并写出它的转移矩阵.

26. 习题: 甲袋中有 3 个黑球, 一袋中有三个白球, 每次从两袋中各取一球交换, 用 X_n 表示 n 次交换后甲袋中白球数, 说明 X_n 是马氏链并写出它的转移矩阵和平稳分布. 再计算 n 次交换后甲袋中球颜色一致的概率的极限值.

27. 设 S 是非负整数集, 令

$$\mathbb{P} := \begin{bmatrix} q_0 & p_0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & \cdots \\ q_2 & 0 & 0 & p_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

其中 p_x, q_x 是正的, 且 $p_x + q_x = 1$, 因此 $\mathbb{P}$ 是一个转移矩阵. 设 X 是一个以 $\mathbb{P}$ 为转移矩阵的 Markov 链, 证明: X 是不可约非周期的. 它什么时候是常返的? 什么时候有平稳分布.

28. 证明: 有限状态马氏链没有零常返状态.

第 六 讲　布朗运动

摘要 从离散时间随机过程到连续时间随机过程在思想与数学方法方面更加抽象, 但连续时间随机过程在模拟物体运动时显然更自然. Brown 运动作为模拟由生物学家 R. Brown 在 1927 年观察到并描述的花粉等微粒在液体表面的随机运动是最重要的随机过程, 由 A. Einstein 在 1905 年在讨论热力学现象时给出数学描述并由 N. Wiener 在 1918 年最终给出存在性证明以来, 一直是随机性研究的中心. 本章介绍 Brown 运动的定义及其性质, 还通过一些经典例子介绍其思想方法, 最后说明 Brown 运动是鞅, 而且通过 Brown 可以构造其他的鞅从而解决其他有趣问题.

关键词

- Brown 运动
- 独立平稳增量
- 有限维分布
- 首达时的分布
- 自相似性
- Donsker 不变原理
- 鞅性, 指数鞅

1. 随机过程: 从离散时间随机序列到连续时间随机过程, 直观上差不多, 但方法很不同, 因为物体的运动一般是连续的, 所以连续时间随机过程通常能更好地模拟和研究某些随机现象.

2. 连续时间随机过程: 在某个时间区间上连续记录一个随机试验的结果称为连续时间随机过程. 具体地, 样本空间 Ω 上的随机变量族

$$X = (X_t : t \in [0, T]),$$

称为随机过程, 这里 T 可以是 $+\infty$. 注意 $X_t = X(t, \omega)$, $t \leq T, \omega \in \Omega$ 是两个变量的函数, t 是时间变量, ω 是样本变量, 固定 t 是随机变量, 固定 ω 是 $[0, T]$ 上的函数, 称为样本 ω 的轨道, 或样本轨道. 样本轨道这个词很形象, 记录一个股票的走势是一条样本轨道, 记录自来水的用量是一条样本轨道, 记录到达一个银行的人数是样本轨道, 等等. 随机变量对应的是一个数, 而随机过程对应的是一个函数, 所以随机过程也可以理解为随机函数.

3. 符号: 随机过程标准的写法是 $X = (X_t : t \in [0, T])$, X 是这个随机过程的名字, X_t 是随机过程在时间 t 的取值, 有时我们直接说随机过程 X 或者 (X_t) 或者简单地 X_t 等, 不一一解说.

4. 有限维分布: 任取有限个时间点 $0 \leq t_1 < t_2 < \cdots < t_n$, 随机向量

$$(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$$

的分布称为随机过程在时间点 $(t_1, \cdots, t_n)$ 上的有限维分布, 所有的有限维分布全体称为有限维分布族. 如果两个随机过程在所有相同时间点上的有限维分布相同, 那么我们说它们等价.

5. 最常见最重要的两个随机过程是 Poisson (泊松) 过程与 Brown (布朗) 运动, 前者描述时间段 $[0, t]$ 到达某服务店的人数, 也称为排队过程, 后者描述花粉在液体表面由于分子碰撞引起的运动轨迹, 它由首次描述这种现象的生物学家 Brown 的名字命名, 但实际上数学模型的建立主要归功于物理学家 A. Einstein 和数学家 N. Wiener. Poisson 过程的样本轨道是跳跃的, Brown 运动的样本轨道是连续的. 与这些伟大的名字联系着, 可想而知, 这两个随机过程是多么的重要.

6. 信息流: 时间 t 之前的信息 $\sigma(X_s : 0 \le s \le t)$ 记为 $\mathscr{F}_t$, 虽然其中包含无限多个随机变量, 但是它的直观意义与离散时间类似. $(\mathscr{F}_t)$ 称为信息流. 注意 $\mathscr{F}_t$ 中包含不可数无穷多个随机变量, 更加抽象更难理解, 可是它对于理解随机分析是至关重要的. 让我们从几个不同角度做一个解释, 希望能够帮助读者理解它. 首先, 我们可以证明一个随机变量 ξ 关于它可测当且仅当可以从中取出一个随机序列使得 $\{X_{s_n}\}$ 使得 ξ 可以写成这个序列的函数. 另外随机变量 ξ 与 $\mathscr{F}_t$ 独立是指 ξ 与其中任何有限个随机变量组成的随机向量独立.

7. 关于信息流的条件期望: 怎么理解随机变量 ξ 关于它的条件期望 $\mathbb{E}(\xi|\mathscr{F}_t)$? 直观上, 和前面第三讲所说的关于随机变量的条件期望一样, 它是已知信息条件下的最佳预测. 具体计算时, 它是一个关于 $\mathscr{F}_t$ 可测的随机变量 η, 满足对任何有限个随机变量 $X_{s_1}, \cdots, X_{s_n}$ 和函数 g, 有

$$\mathbb{E}[\xi g(X_{s_1}, \cdots, X_{s_n})] = \mathbb{E}[\eta g(X_{s_1}, \cdots, X_{s_n})].$$

通常我们不需要真的去计算随机变量关于信息流的条件期望, 更多的是应用它的性质. 第三讲里面关于条件期望的那些性质在信息流的情况下还是成立的, 这些性质读者应该理解并记住, 后面会经常用到.

(a) $\mathbb{E}(1|\mathscr{F}_t) = 1$;

(b) $\mathbb{E}(a_1 Y_1 + a_2 Y_2|\mathscr{F}_t) = a_1\mathbb{E}(Y_1|\mathscr{F}_t) + a_2\mathbb{E}(Y_2|\mathscr{F}_t)$;

(c) 如果 $Y_1 \le Y_2$ 那么 $\mathbb{E}(Y_1|\mathscr{F}_t) \le \mathbb{E}(Y_2|\mathscr{F}_t)$;

(d) $|\mathbb{E}(Y|\mathscr{F}_t)| \le \mathbb{E}(|Y||\mathscr{F}_t)$;

(e) $\mathbb{E}[\mathbb{E}(Y|\mathscr{F}_t)] = \mathbb{E}[Y]$;

(f) 如果 $s \le t$, 那么 $\mathbb{E}[\mathbb{E}(Y|\mathscr{F}_t)|\mathscr{F}_s] = \mathbb{E}(Y|\mathscr{F}_s)$;

(g) 如果 Y 是 $\mathscr{F}_t$ 可测的, 那么 $\mathbb{E}(Y|\mathscr{F}_t) = Y$;

(h) 如果 H 是 $\mathscr{F}_t$ 可测的, 那么

$$\mathbb{E}(H \cdot Y|\mathscr{F}_t) = H \cdot \mathbb{E}(Y|\mathscr{F}_t).$$

8. 连续时间鞅: 类似于离散时间, 称随机过程 $X = (X_t)$ 是鞅, 如果对任意 $t > s \ge 0$ 有

$$\mathbb{E}(X_t|\mathscr{F}_s) = X_s. \tag{6.1}$$

等价于增量的条件期望等于零

$$\mathbb{E}(X_t - X_s|\mathscr{F}_s) = 0.$$

同样也有上鞅下鞅的概念.

9. 独立增量: 设 $X = (X_t)$ 是随机过程, X_t 可以取值在高维空间, 比如 Brown 运动描述花粉在液体表面的运动, 因此应该是二维的, 但是为了简单, 我们通常假设它们是一维的. 对于 $t > s \geq 0$, $X_t - X_s$ 称为增量. 一个随机过程称为是独立增量的是指它的增量是无记忆的, 或者说增量 $X_t - X_s$ 与时间 s 前的信息 $\mathscr{F}_s$ 独立, 也就是说, 任意取 s 前的时间 $0 < s_1 < \cdots < s_n \leq s$, 随机变量 $X_t - X_s$ 与随机向量 $(X_{s_1}, \cdots, X_{s_n})$ 独立. Poisson 过程与 Brown 运动都是独立增量的.

10. 马氏性: 马氏性是指给定现在的位置, 将来与过去是独立的. 马氏性是随机过程中最重要的性质之一, 独立增量的随机过程一定是有马氏性的. 马氏性是以俄国数学家 Markov 的名字命名的, 他在 150 年前首次抽象出现在所谓的马氏性. 有马氏性的随机过程称为马氏过程. 独立增量蕴含马氏性, 因为 $X_t - X_s$ 与 s 之前独立, 因此在知道 X_s 后 X_t 与 s 前一定是独立的, 用数学语言来表达: 对任何 $t > s > 0$ 与 $x \in \mathbf{R}$, 有

$$\mathbb{P}(X_t \leq x|\mathscr{F}_s) = \mathbb{P}(X_t \leq x|X_s). \tag{6.2}$$

11. 平稳增量: 如果增量 $X_t - X_s$ 与 $X_{t-s} - X_0$ 同分布, 那么我们说随机过程是平稳增量的. 平稳增量是说增量的分布依赖于时间差 $t - s$, 与时间起终点 s, t 无关. Poisson 过程的增量的分布是参数为 $\lambda(t-s)$ 的 Poisson 分布, 而 Brown 运动的增量的分布是正态分布 $N(0, t-s)$, 都是平稳增量过程.

12. 因此 Poisson 过程与 Brown 运动都是平稳且独立增量过程. 为了方便, 我们假设它们都从 0 出发, 即 $X_0 = 0$. 满足独立增量性与平稳增量性的的随机过程通称为 Lévy 过程, Lévy 包括了一大类随机过程, 它可以用来模拟许多随机现象. 比如在金融市场上, 人们已经不满足于用 Brown 运动来模拟股票价格, 因为通过一些数据分析, 一些学者认为股票走势中包含有跳的现象, 或者股票走势的分布似乎不应该是轻尾的, 这两种现象都不符合 Brown 运动, 所以学者希望能以符合这两中现象的最简单的随机过程来模拟股票走势, 就是 Lévy 过程中的稳定过程.

13. 布朗运动: 满足下面条件的随机过程 $X=(X_t)$ 称为原点出发的 Brown 运动,

   (a) $X_0=0$;

   (b) 独立增量

   (c) 平稳增量且 $X_t-X_s\sim N(0,t-s)$;

   (d) 样本轨道连续.

   当我们掷两个银币的时候, 两个银币的正反面或许不同, 但是它们的分布一样; 同样, 当我们观察两个花粉的运动时, 它们的轨迹不同, 但是它们都是 Brown 运动, 因它们的分布相同. 另外说 $X=(X_t)$ 是从 x 出发的 Brown 运动, 如果 (X_t-x) 是从零点出发的 Brown 运动. 一般地, n 维 Brown 运动是指 n 个独立的 Brown 运动作为分量组成的随机向量过程. 另外, Brown 运动的指数 (e^{X_t}) 称为是几何 Brown 运动.

14. 存在性: 通常随机过程都有一个存在性问题, 对于 Brown 运动尤其如此, 存在性问题曾经困扰数学家十多年, 最后是 N. Wiener 在 1923 年给予证明, 所以 Brown 运动也称为 Wiener 过程. 也许一些读者对 Brown 运动的存在性不以为然, 认为 Brown 运动作为刻画花粉在液体表面的运动轨迹自然地存在着, 但实际上我们这里说的 Brown 运动只是生物中所说的 Brown 运动的一个数学模型, 它是满足给定条件的一个数学概念, 它的存在性必须得到逻辑保证, 而不能只是以生物学观察来保证. 把一个数学概念称为 Brown 运动是因为学者认为它真的很好地模拟了生物中所谓的 Brown 运动. 关于 Brown 运动的存在性证明, 参考 [4].

15. 布朗运动的有限维分布: 设 $X=(X_t)$ 是零点出发的 Brown 运动, 由独立增量性推出对任何 $0<t_1<t_2<\cdots<t_n$, 增量随机向量

   $$(X_{t_1}, X_{t_2}-X_{t_1}, \cdots, X_{t_n}-X_{t_{n-1}})$$

   独立且都是正态分布, 所以它的联合密度函数是 n 维正态分布

   $$p_{t_1}(x_1)p_{t_2-t_1}(x_2)\cdots p_{t_n-t_{n-1}}(x_n),$$

   其中 $p_t(x)$ 表示期望 0 方差 t 的一维正态分布的密度函数. 由于随机向量 $(X_{t_1},\cdots,X_{t_n})$ 是上述随机向量的一个线性变换, 所以它也是正态分布, 且它

的密度函数是

$$\begin{aligned}&p_{t_1}(x_1)p_{t_2-t_1}(x_2-x_1)\cdots p_{t_n-t_{n-1}}(x_n-x_{n-1})\\&=\frac{1}{\sqrt{(2\pi)^n t_1(t_2-t_1)\cdots(t_n-t_{n-1})}}\\&\quad\cdot\exp\left(-\frac{1}{2}\left(\frac{x_1^2}{t_1}+\frac{(x_2-x_1)^2}{t_2-t_1}+\cdots\frac{(x_n-x_{n-1})^2}{t_n-t_{n-1}}\right)\right).\end{aligned}$$

因此 Brown 运动的有限维分布是期望为零的正态分布, 是 Gauss 过程 (所有有限维分布都是正态分布的随机过程称为是一个 Gauss 过程). 对于期望零的正态分布, 知道它的协方差矩阵就足够了, 其 i 行 j 列的元素为

$$\mathbb{E}[X_{t_i}X_{t_j}]=t_i\wedge t_j.$$

16. 协方差: 怎么算协方差 $\mathbb{E}[X_tX_s]$? 利用独立增量性, 如果 $t>s$, 那么

$$\mathbb{E}[X_tX_s]=\mathbb{E}[(X_t-X_s)X_s]+\mathbb{E}[X_s^2]=s.$$

可以这么说, 只要善于用独立增量性, Brown 的问题基本上都可以解决.

17. 习题: 设 $X=(X_t)$ 是原点出发的 Brown 运动, (1) 对任何正数 t_1,t_2,t_3,t_4 求 $\mathbb{E}[X_{t_1}X_{t_2}X_{t_3}]$ 与

$$\mathbb{E}[X_{t_1}X_{t_2}X_{t_3}X_{t_4}];$$

(2) 设 $s<t$, 求 $\mathbb{E}[X_s|X_t]$; (3) 设 $s<t<u$, 求 $\mathbb{E}[X_t|X_s,X_u]$.

18. 习题: Einstein 发现 Brown 运动 X_t 的密度函数 $u(t,x)=p_t(x)$ 应该满足热传导方程

$$\frac{\partial u}{\partial t}=\frac{1}{2}\frac{\partial^2 u}{\partial x^2},$$

证明之.

19. 性质: 因此要验证一个随机过程是 Brown 运动, 一个办法是验证它是 Gauss 过程, 期望为零且方差为 $\mathbb{E}(X_tX_s)=t\wedge s$ 就可以了, 所以下面几个性质很容易验证.

   (a) 对称性: $(-X_t)$ 也是 Brown 运动;

   (b) 相似性: 对任何实数 $c\neq 0$, $Y_t=c^{-1}X_{c^2t}$ 也是 Brown 运动;

(c) 可逆性: 把时间尺度转过来, $Y_t = tX_{1/t}$ 也是 Brown 运动.

(d) 可复制性: 如果 T 是停时, 那么 $Y_t := X_{t+T} - X_T$ 作为随机过程也是零点出发的 Brown 运动.

20. 不变原理: 设 $\{\xi_n\}$ 独立同分布随机变量, $\mathbb{E}\xi_n = 0$, $\mathrm{Var}(\xi_n) = 1$. 令 $S_n = \xi_1 + \cdots + \xi_n$, 用直线把点 (n, S_n) 依次相连, 即

$$S_t := (t-n)(S_{n+1} - S_n) + S_n,\ n \le t < n+1,\ n \ge 0.$$

那么 S_t 是随机过程. 然后把时间和空间的做个尺度上的变换, 对任何 n, 令

$$X_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}, \tag{6.3}$$

当 n 越来越大的时候, 过程 $(X_t^{(n)})$ 看上去会越像一个真正的 Brown 运动. 对于固定的 t, 由中心极限定理就可以看出来 $X_t^{(n)}$ 依分布收敛于正态分布 $N(0,t)$, 而更强大的结果如下.

定理6.1 (Donsker 不变原理) 当 n 趋于无穷时, $X_t^{(n)}$ 收敛于一个 Brown 运动.

21. 习题: Brown 运动很直观, 但是在电脑上无法直接实现, 我们可以用 Donsker 不变原理来模拟 Brown 运动, 这正是它的强大之处. 写一个程序来模拟 Brown 运动.

22. 首中时: 设 T_a 是从 0 出发的 Brown 运动 X 首达 $a \in \mathbf{R}$ 的时间, 精确地说

$$T_a(\omega) := \inf\{t > 0 : X_t(\omega) = a\},$$

虽然 Brown 运动从 0 出发, 但是定义的方式表明 T_0 未必等于 0. 在简单随机游动的场合下, 我们算出了它的分布并且证明了 $\mathbb{P}(T_a < \infty) = 1$, 那么这个方法可否由于 Brown 运动呢? Brown 也有对称性, 即 $(-X_t)$ 是和 (X_t) 分布相同的 Brown 运动, 所以也可以应用反射原理. 设 $a > 0$, 在 $T_a(\omega) < t$ 时, 样本轨道 $X_s(\omega)$ 作为 s 的函数在 t 之前碰到 a, 那么在碰到 a 之后的轨道往上走往下走是等可能的, 推出 $X_t > a + (b-a)$ 和 $X_t < a - (b-a)$ 的概率相同.

定理6.2 (反射原理) 设 $a > 0$, $b \in \mathbf{R}$, $t > 0$, 那么

$$\mathbb{P}(T_a < t, X_t > b) = \mathbb{P}(T_a < t, X_t < 2a - b).$$

这个反射原理可以用来算 T_a 的分布了, 注意 T_{-a} 与 T_a 同分布. 取 $b=a$, 得

$$\mathbb{P}(T_a<t, X_t>a)=\mathbb{P}(T_a<t, X_t<a),$$

那么 T_a 的分布函数为

$$\begin{aligned}\mathbb{P}(T_a<t) &= 2\mathbb{P}(T_a<t, X_t>a)\\ &= 2\mathbb{P}(X_t>a)\\ &= 2\int_a^\infty \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-\frac{|x|^2}{2t}}\mathrm{d}x\\ &= 2\int_{a/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{|x|^2}{2}}\mathrm{d}x\\ &= 2\mathbb{P}(X_1>a/\sqrt{t}),\end{aligned}$$

当 $t\to\infty$ 时,

$$\mathbb{P}(T_a<\infty)=2\mathbb{P}(X_1>0)=1.$$

对 t 求导, 得 T_a 的密度函数为

$$\phi_a(t)=\frac{a}{\sqrt{2\pi t^3}}\exp\left(-\frac{a^2}{2t}\right).$$

23. 回归时间: 上面的方法对 $a=0$ 偏偏无效. 自然想到取极限, 当 $a\downarrow 0$ 时, 是否有 $T_a\downarrow T_0$? 如果对的话, 那么对任何 $t>0$,

$$\mathbb{P}(T_0<t)=2\mathbb{P}(X_1>0)=1.$$

因此 $T_0=0$. 但上面的极限不一定对, 怎么办呢? 取 $t>0$, 用 T^t 表示 t 时间后首次达到 0 的时间, 那么 $\lim_{t\to 0}T^t=T_0$. 由 Brown 运动平移不变的性质, 当 $X_t=x$ 时, 之后它首达 0 相当于从 0 出发首达 $-x$, 先算条件期望

$$\begin{aligned}\mathbb{P}(T^t<u|X_t=x) &= \mathbb{P}(T_{-x}<u-t)\\ &= 2\mathbb{P}(X_1>|x|/\sqrt{u-t}),\end{aligned}$$

再取期望得

$$\mathbb{P}(T^t<u)=\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{\frac{-|x|^2}{2t}}2\mathbb{P}(X_1>\frac{|x|}{\sqrt{u-t}})\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{-|x|^2}{2}} 2\mathbb{P}(X_1 > \frac{\sqrt{t}|x|}{\sqrt{u-t}}) \mathrm{d}x.$$

令 $t \downarrow 0$, 推出

$$\mathbb{P}(T_0 < u) = 2\mathbb{P}(X_1 > 0) = 1$$

对任何 $u > 0$ 成立, 因此 $T_0 = 0$ a.s. 这说明 X 的轨道出发时无穷多次与 x- 轴相交. 实际上 X 的轨道与 x- 轴相交点的集合没有孤立点, 但是它的总长度为零.

24. 习题: 设 $X = (X_t)$ 是原地出发的 Brown 运动, 定义

$$Y_t(\omega) := \int_0^t X_s(\omega)\mathrm{d}s, \ t > 0, \omega \in \Omega.$$

这个积分是通常的 Riemann 积分, 证明: $Y = (Y_t)$ 是 Gauss 过程, 求它的协方差函数 $\mathrm{cov}(Y_s, Y_t)$.

25. Wiener 的构造: Brown 运动的构造方法有许多, 其中 Wiener 不仅首先证明了 Wiener 测度 (也就是 Brown 运动) 的存在性, 还利用独立且服从标准正态分布的随机序列 $\{g_n : n \geq 0\}$ 以另外一种方法构造了 Brown 运动. 利用 Fourier 级数, 定义

$$X_t = \frac{t}{\sqrt{\pi}} g_0 + \sum_{n \geq 1} \sum_{k=2^{n-1}}^{2^n - 1} \sqrt{\frac{2}{\pi}} \frac{\sin kt}{k} g_k,$$

右边实际上在 $[0, \pi]$ 上是以概率 1 一致收敛的, 因此是 Brown 运动.

26. 随机性的问题: 前面我们说过随机变量的信息, 信息看起来和概率没有关系, 但是离开概率谈信息意义肯定不大, 所以我们把概率元素加进去来讨论随机变量的随机性问题, 实际上是分布的随机性问题, 可能更有意义些. 分布所包含的随机性是不同的, 比如掷硬币的随机性没有掷骰子的随机性强. 一般地如若随机变量 Y 是 X 的函数, 我们说 X 更多信息, 也说 X 的分布函数更随机. 如果他们可以互相表示, 那么随机变量的信息一样, 他们的分布函数的随机性一样. 但是不同信息的随机变量可以有相同的随机性, 因为随机性忽略一些概率零的信息. Brown 运动包含有多少随机性? 看起来很多, 因为它包含那么多的正态随机变量, 对每个 $t > 0$, X_t 都是正态的, 但 Wiener 构造告诉我们它实际上建立在可数多个独立的标准正态分布随机变量上, 而可数多

个标准正态分布的随机变量是可以通过 $[0,1]$ 上均匀分布随机变量构造出来的, 所以 Brown 运动所包含的随机性和 $[0,1]$ 上均匀分布的随机变量的随机性是一样的.

27. 与 Brown 运动相关的鞅: 设 $X=(X_t)$ 是 Brown 运动.

(a) (X_t) 是鞅;

(b) 随机过程 (X_t^2-t) 是鞅;

(c) 随机过程 $\big(\exp(aX_t-a^2t/2)\big)$ 是鞅, 这里说明几何 Brown 运动 $\exp(X_t)$ 一般不是鞅, 而往下漂移后的 Brown 运动 $X_t-t/2$ 的指数才是鞅.

要证明 X_t^2-t 是鞅, 取 $t>s>0$,

$$\begin{aligned}\mathbb{E}[(X_t-X_s)^2|\mathscr{F}_s]&=\mathbb{E}[X_t^2-2X_tX_s+X_s^2|\mathscr{F}_s]\\&=\mathbb{E}[X_t^2|\mathscr{F}_s]+X_s^2-2X_s\mathbb{E}[X_t|\mathscr{F}_s]\\&=\mathbb{E}[X_t^2-X_s^2|\mathscr{F}_s],\end{aligned}$$

而 X_t-X_s 与 $\mathscr{F}_s$ 独立, 所以

$$\mathbb{E}[X_t^2-X_s^2|\mathscr{F}_s]=\mathbb{E}[(X_t-X_s)^2]=t-s,$$

看出 X_t^2-t 是鞅. 关于指数鞅, 因为 X_t-X_s 与 $\mathscr{F}_s$ 独立且服从 $N(0,t-s)$, 因此

$$\mathbb{E}\left(\mathrm{e}^{a(X_t-X_s)}|\mathscr{F}_s\right)=\mathbb{E}\left(\mathrm{e}^{a(X_t-X_s)}\right)=\mathrm{e}^{a^2(t-s)/2},$$

推出 (3) 成立.

28. 例: 在连续时间情形, Doob 的随机停止定理仍然成立. 与离散时间情形类似, 说随机时间 $T:\Omega\to[0,\infty]$ 是停时是指对任何 $t\geq 0$ 有 $\{T\leq t\}\in\mathscr{F}_t$.

定理6.3 如果 $M=(M_t)$ 是鞅且 T 是一个停时, 那么 $(M_{t\wedge T})$ 也是鞅.

有期望不变性

$$\mathbb{E}[M_0]=\mathbb{E}[M_{t\wedge T}]$$

对任何 $t>0$ 成立. 只要能交换极限和期望, 鞅可以解决很多问题, 比如前面 $T=T_a$ $(a>0)$ 的分布, 我们用鞅方法来做. 用哪个鞅来做? 其它两个没有用, 必须用指数鞅. 换个参数,

$$\mathrm{e}^{bX_t-\frac{1}{2}b^2t}$$

是鞅, 那么由期望不变性

$$\mathbb{E}\left(\mathrm{e}^{bX_{t\wedge T}-\frac{1}{2}b^2(t\wedge T)}\right)=1.$$

当 $t\to\infty$ 时, $X_{t\wedge T}\to X_T=a$, 问题就可以解决了. 但极限与期望可以交换吗? 由控制收敛定理, 只要期望里面的量被一个与 t 与 ω 无关的界控制就可以用控制收敛定理了. 因为 $t\wedge T\le T$, 所以 $X_{t\wedge T}\le a$, 上有界, 故而只要 $b>0$,

$$\mathrm{e}^{bX_{t\wedge T}}\le \mathrm{e}^{ba},$$

另一部分, $\mathrm{e}^{-b^2(t\wedge T)/2}\le 1$, 因此在 $b>0$ 时可以交换, 得

$$\mathbb{E}\left(\mathrm{e}^{-\frac{1}{2}b^2T}\right)=\mathrm{e}^{-ab},$$

做变换 $\theta=b^2/2$, 推出 T 的 Laplace 变换为

$$\mathbb{E}\left(\mathrm{e}^{-\theta T_a}\right)=\mathrm{e}^{-a\sqrt{2\theta}}.$$

从数学上看, 一个非负随机变量的 Laplace 变换唯一地确定了它的分布.

29. 习题: 设 $a<0<b$, T_a, T_b 分别是原点出发的 Brown 运动首次碰到 a,b 的时间. 用鞅方法来求 $\mathbb{P}(T_a<T_b)$, $\mathbb{E}[T]$ 以及 T 的 Laplace 变换, 其中 $T=T_a\wedge T_b$.
解答: 由鞅的期望不变性, 对任何 $t>0$,

$$\mathbb{E}B_{T\wedge t}=0.$$

当 $t\to\infty$ 时, $a\le B_{T\wedge t}\le b$, 因此由控制收敛定理得 $\mathbb{E}B_T=0$. 而

$$B_T=a1_{\{T_a<T_b\}}+b1_{\{T_b<T_a\}},$$

因此 $a\mathbb{P}(T_a<T_b)+b\mathbb{P}(T_b<T_a)=0$, 即

$$\mathbb{P}(T_a<T_b)=\frac{b}{b-a}.$$

要算 $\mathbb{E}T$, 需要用鞅 (B_t^2-t), 还是由期望不变性,

$$\mathbb{E}[B_{T\wedge t}^2]=\mathbb{E}[T\wedge t].$$

然后让 $t\to\infty$, 左边应用控制收敛定理右边应用单调收敛定理得

$$\mathbb{E}[T]=\mathbb{E}[B_T^2]=a^2\mathbb{P}(T_a<T_b)+b^2\mathbb{P}(T_a>T_b)=-ab.$$

类似地, 要算 T 的 Laplace 变换, 应该用指数鞅 $(\exp(\lambda B_t - \lambda^2 t/2))$. 期望不变性和控制收敛定理得

$$\mathbb{E}\left(\exp(\lambda B_T - \lambda^2 T/2\right) = 1, \tag{6.4}$$

因此

$$\mathrm{e}^{\lambda a}\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}; T_a < T_b\right] + \mathrm{e}^{\lambda b}\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}; T_a > T_b\right] = 1. \tag{6.5}$$

这还不足以算出 $\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}\right]$. 再考虑指数鞅 $(\exp(-\lambda B_t - \lambda^2 t/2))$, 类似地得

$$\mathbb{E}\left(\exp(-\lambda B_T - \lambda^2 T/2\right) = 1 \tag{6.6}$$

和

$$\mathrm{e}^{-\lambda a}\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}; T_a < T_b\right] + \mathrm{e}^{-\lambda b}\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}; T_a > T_b\right] = 1. \tag{6.7}$$

那么

$$\mathbb{E}\left[\mathrm{e}^{-\lambda^2 T/2}\right] = \frac{\sinh(\lambda a) - \sinh(\lambda b)}{\sinh(\lambda(a-b))}$$

其中 $\sinh x = (\mathrm{e}^x - \mathrm{e}^{-x})/2$. 由此立刻得到 T 的 Laplace 变换.

30. 习题: 前面看到 Brown 运动一定会碰到任何一条水平直线. 那么它会肯定碰到一条斜的直线 $x = at + b$ 吗? 显然当 a, b 异号时肯定, 同号时肯定吗? 似乎未必. 取 $a, b > 0$, 那么 $x = at + b$ 是一条在原点之上, 斜率正的直线, 令 T 是 Brown 运动 B 首次碰到这条直线的时间, 即

$$T = \inf\{t > 0 : B_t = at + b\}.$$

求 $\mathbb{P}(T < \infty)$. 我看过在一些股票走势的分析方法中, 有人用直线相交的方法来决定买入或者卖出, 似乎可能用到类似的结果. 我们在这里介绍一个技巧, 一些细节留给读者. 因为 $\exp(cB_t - c^2 t/2)$ 是鞅, 因此由 Doob 鞅基本定理, 对任何 $t > 0$ 有

$$\mathbb{E}\left(\exp(cB_{t\wedge T} - c^2(t\wedge T)/2)\right) = 1,$$

因为 $B_{t\wedge T} \le a(t\wedge T) + b$, 故只要 $ac - c^2/2 \le 0$ 且 $c > 0$ 时, t 可以趋于无穷, 这时

$$\mathbb{E}\left[\exp\left((ac - \frac{c^2}{2})T\right)\right] = \mathrm{e}^{-bc},$$

方程 $ac - c^2/2 = -\lambda$ 当 $\lambda > 0$ 时有正根 $c = a + \sqrt{a^2 + 2\lambda}$, 那么

$$\mathbb{E}\left(\mathrm{e}^{-\lambda T}\right) = \exp(-b(a + \sqrt{a^2 + 2\lambda})).$$

第 七 讲　二次变差过程

摘要 Brown 运动的样本轨道是连续函数, 因为极端的随机性, 它的轨道变得极其不光滑, 本章将引入讨论函数光滑性的量: 一次变差与二次变差. 证明了样本轨道没有有限的一次变差, 但是有很好的二次变差, 这至少说明 Brown 运动的轨道不能定义通常的积分。在 19 世纪时候, 数学界为 Weierstrass 构造的一个连续

关键词

- 样本轨道
- 一次变差
- 二次变差
- 平方可积鞅
- 二次变差过程
- 二次协变差

1. 样本轨道: 观察一个花粉粒子的运动轨迹就是一条样本轨道, 样本空间是所有可能的样本轨道全体, 概率告诉我们哪些样子的轨道发生的可能性的大小. 从统计角度看, 一条样本轨道是不说明多少问题的, 观察许多许多样本才能看到样本轨道的分布情况. 股票的走势是一条样本轨道, 许多所谓专家执着地认为自己可以从这一条轨道看出其中的分布情况和将来的走势从理论上看多少有点滑稽. 不过我们这个讲义是在假设了走势分布的情况下来应用随机分析的, 至于假设是否合理不是随机分析能解决的问题.

2. 光滑与不光滑: 让我们先从数学上对 Brown 运动的样本轨道性质进行分析. 从学数学开始, 我们接触的函数看上去通常是光滑的, 因为通常的初等函数是无穷次可微的, 我们对光滑的函数如此熟悉, 以至于要构造一个处处不可导的函数并不容易, 但是 Brown 运动展现另一个极端, 为了看清楚这个问题, 我们首先需要想一想怎么来衡量函数的光滑程度? 光滑的反义词应该是不光滑? 从 Donsker 的不变原理看, Brown 运动虽然连续, 但肯定是极其不光滑的一类, 它的不光滑的程度使得我们无法用笔把它描绘出来, 因为人的手不可能自然地抖到那种程度. 那我们怎么来描述这种极端的不光滑性呢?

3. 几乎处处: 在概率论中, 大家要学会用几乎处处的思想来思考问题. 在古典概率中, 概率零的事件就是不可能的, 但是在一般的概率空间中, 概率零只是几乎不可能的, 比如在 $[0,1]$ 上取个数, 它是有理数的概率几乎是不可能的; 在 n 维空间中按独立的正态分布取 n 个向量, 那么它们线性相关的概率是几乎不可能的. 但是这些可能结果在样本空间中都存在着. 对于 Brown 运动 $X = (X_t : t \in [0,1])$, 固定样本 ω, 那么 $X_t(\omega)$ 关于 t 是连续函数, 称为样本轨道, 我们要说明几乎所有的样本轨道都是很不光滑的.

4. 函数图像的长度: 考虑 $[0,1]$ 上的连续函数: $x = x(t)$. (其实我们可以考虑任何区间上的函数, $[0,1]$ 只是为了方便.) 对于函数的图像来说, 一个自然的概念是长度, 一个连续函数的图像作为连续曲线应该是可以测量长度的, 但是怎么测量呢? 当然是用无穷小的方法, 考虑任意划分 Δ:

$$0 = t_0 < t_1 < \cdots < t_n = 1.$$

按照直线最短的原则, 函数图像的长度 $L(x)$ 肯定不小于如下的折线长度和,

即

$$L(x) \geq \sum_i \sqrt{(t_i - t_{i-1})^2 + (x(t_i) - x(t_{i-1}))^2}.$$

再利用三角不等式, 右边一定是随着划分的变细而递增的, 所以我们把长度就定义为

$$L(x) = \sup_{\Delta} \sum_i \sqrt{(t_i - t_{i-1})^2 + (x(t_i) - x(t_{i-1}))^2}. \tag{7.1}$$

如果图像的长度无限, 可以想象函数本身一定很不光滑, 因为正常的人大概不可能画出一条长度无限的曲线来, 所以我们就从长度有限还是无限来讨论函数的光滑程度. 下面的结果是微积分的一个著名公式.

引理7.1 如果函数 $x = x(t)$ 在 $[0,1]$ 上连续可导, 那么 $L(x) < \infty$ 且

$$L(x) = \int_0^1 \sqrt{1 + x'(t)}\mathrm{d}t.$$

5. 怎么取划分: 因为涉及积分, 我们后面经常有取划分的需要. 从理论上讲积分存在性应该与划分无关, 这里以及后面所涉及的划分也是如此. 但为了容易理解和避免混乱, 我们可以规定在后面说的划分变细趋于零, 是指对区间 $[0,T]$ 取一列特别的划分: 二分点划分, 对 $n \geq 1$, 定义

$$\Delta_n := \{0, \frac{T}{2^n}, \frac{2T}{2^n}, \cdot, \frac{(2^n-1)T}{2^n}, T\},$$

也就是长度为 $T2^{-n}$ 的等分点, 这样取划分的好处是前一次划分包含在后一次划分中, 分点越来越细, 当 n 趋于无穷时, 划分的长度趋于零. 但是为了书写的方便, 我们总是简单地说 “划分 $\Delta = \{t_i\}$ 变细趋于零”.

6. 函数的变差: 由于不等式

$$\begin{aligned}\sum_i |x(t_i) - x(t_{i-1})| &\leq \sum_i \sqrt{(t_i - t_{i-1})^2 + (x(t_i) - x(t_{i-1}))^2} \\ &\leq 1 + \sum_i |x(t_i) - x(t_{i-1})|,\end{aligned}$$

所以 $L(x)$ 有限无限只和 $\sum_i |x(t_i) - x(t_{i-1})|$ 的性质有关, 它称为是函数 x 在划分 Δ 上的变差, 定义

$$V(x) = \sup_{\Delta} \sum_i |x(t_i) - x(t_{i-1})|, \tag{7.2}$$

它称为是 $x = x(t)$ 的一次变差. 显然 $L(x) < \infty$ 当且仅当 $V(x) < \infty$. 可以证明, 一个函数其一次变差有限当且仅当它可以写成为两个递增函数的差.

7. 轨道积分: 如果 $y = y(t)$ 是 $[0, 1]$ 上的连续函数, 那么我们作划分 Δ 上关于函数 $x = x(t)$ 的 Riemann 和

$$\sum_i y(s_i)(x(t_i) - x(t_{i-1})), \tag{7.3}$$

其中 $s_i \in [t_i, t_{i-1}]$. 假设函数 x 具有有限的一次变差, 那么当划分变细趋于零时, Riemann 和的极限存在而且与 s_i 的取法无关, 通常我们取左端 t_{i-1}, 记这个极限为

$$\int_0^t y(t)\mathrm{d}x(t),$$

称为是 Riemann-Stieltjes 积分. 上面积分的存在性非常依赖于 x 的有限一次变差性, 否则积分是不可能存在的. 当 x 连续可导时, 上面的积分变成通常的 Riemann 积分

$$\int_0^1 y(t)\mathrm{d}x(t) = \int_0^1 y(t)x'(t)\mathrm{d}t.$$

下面定理说明问题所在, 但证明要用到泛函分析中的共鸣定理.

定理7.1 当划分 Δ 变细趋于零时, Riemann 和 (7.3) 对任意连续函数有有限极限, 那么 x 有有限一次变差.

8. 微积分基本定理: 设 f 是 $\mathbf{R}$ 上连续可导的函数, 且 x 有有限一次变差, 那么微积分基本定理成立

$$f(x(t)) - f(x(0)) = \int_0^t f'(x(s))\mathrm{d}x(s),\ \ t \in [0, 1]. \tag{7.4}$$

9. 二次变差: 如果 $\sum_i (x(t_i) - x(t_{i-1}))^2$ 称为是 x 在划分 Δ 上的二次变差, 如果划分变细趋于零时,

$$\lim_{\Delta \to 0} \sum_i (x(t_i) - x(t_{i-1}))^2$$

存在, 那么我们说 x 的二次变差存在, 记为 $\langle x \rangle$, 如果需要, 我们在下标上指定区间以示是这个区间上的二次变差.

引理7.2 如果 x 连续且有有限一次变差, 那么 $\langle x \rangle = 0$.

证明: 令

$$x(\Delta) := \max\{|x(t_i) - x(t_{i-1})| : 1 \le i \le n\},$$

那么当划分变细趋于零时, $x(\Delta)$ 趋于零, 因此

$$\langle x \rangle \le x(\Delta) V(x) \longrightarrow 0.$$

10. 均方收敛: 随机序列 $\{\xi_n\}$ 收敛于 ξ 有多种不同的意义, 其中最重要的有几乎处处收敛与均方收敛, 对每个固定的样本 ω, $\xi_n(\omega)$ 是数列, 如果它收敛于 $\xi(\omega)$, 那么说 ξ_n 在样本 ω 上收敛于 ξ. 几乎处处收敛和点点收敛差不多, 是指不收敛的那些样本全体的概率等于零. 如果

$$\mathbb{E}[\xi_n - \xi]^2 \longrightarrow 0,$$

那么我们说 ξ_n 均方收敛于 ξ. 均方收敛和几乎处处收敛没有蕴含关系, 它们只是不同意义的收敛. 但是均方收敛可以推出依概率收敛, 而且可以推出有一个子列是几乎处处收敛的.

11. Brown 运动的二次变差: 为了方便, 此后我们特别地用 $B = (B_t)$ 表示零点出发的 Brown 运动, Δ 是区间 $[a, b]$ 的划分,

定理7.2 二次变差 $B(\Delta)$ 均方收敛于 $b - a$.

证明: 回忆关于正态分布的矩计算公式

$$\mathbb{E}[(B_t - B_s)^{2n}] = (2n-1)!!(t-s)^n.$$

再利用增量的独立性

$$\begin{aligned}
&\mathbb{E}[(B(\Delta) - (b-a))^2] \\
&\quad = \mathbb{E}[(\sum_i (B_{t_i} - B_{t_{i-1}})^2 - (b-a))^2] \\
&\quad = \mathbb{E}[\sum_i (B_{t_i} - B_{t_{i-1}})^2]^2 \\
&\quad - 2(b-a)\mathbb{E}\left[\sum_i (B_{t_i} - B_{t_{i-1}})^2\right] + (b-a)^2 \\
&\quad = \mathbb{E}\left[\sum_i (B_{t_i} - B_{t_{i-1}})^4\right]
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left[\sum_{i\neq j}(B_{t_i}-B_{t_{i-1}})^2(B_{t_j}-B_{t_{j-1}})^2\right]-(b-a)^2\\
&=3\sum_i(t_i-t_{i-1})^2+\sum_{i\neq j}(t_i-t_{i-1})(t_j-t_{j-1})-(b-a)^2\\
&=2\sum_i(t_i-t_{i-1})^2,
\end{aligned}
$$

因此有

$$\lim_{\Delta\to 0}\mathbb{E}(B(\Delta)-(b-a))^2=\lim_{\Delta\to 0}2\sum_i(t_i-t_{i-1})^2=0.$$

这个定理很清晰地告诉我们, Brown 运动在任何区间上的二次变差都是正的, 因此它的一次变差肯定是无限的, 也就是说它的轨道几乎都很不光滑.

12. 二次变差过程: 定义 $\langle X\rangle_t$ 是随机过程 X 在区间 $[0,t]$ 上的二次变差, 对应的随机过程 $\langle X\rangle=(\langle X\rangle_t,\ t\geq 0)$, 称为 X 的二次变差过程, 因此上面的定理说 Brown 运动的二次变差过程是 t, 即 $\langle X\rangle_t=t$, 这是非常漂亮的一个结果.

13. 几乎处处的意义: 上面所说的二次变差是均方收敛的意义下, 但是不难证明在取某些特殊划分之下, 也可以证明几乎处处收敛. 比如取区间 $[0,1]$ 的二分式划分

$$\Delta_n=\{k\cdot 2^{-n}:0\leq k\leq 2^n\}.$$

那么

$$\mathbb{E}[B(\Delta_n)-1)^2]=2\cdot 2^{-n},$$

由 Chebyshev 不定式

$$\mathbb{P}(|B(\Delta_n)-1|>1/n)\leq\frac{2n^2}{2^n},$$

推出 $\sum_n\mathbb{P}(|B(\Delta_n)-1|>1/n)<\infty$, 最后由 Borel-Cantelli 引理得二次变差 $B(\Delta_n)$ 几乎处处收敛于 1.

14. 例: 设 f 是 $\mathbf{R}$ 上连续可导函数, $B=(B_t)$ 是 Brown 运动, 求随机过程 $X_t=f(B_t)$ 的二次变差过程. 取 $[0,t]$ 上的划分 Δ, 由中值定理

$$\sum_i(f(B_{t_i})-f(B_{t_{i-1}}))^2$$

$$
\begin{aligned}
&= \sum_i [(f'(B_{s_i}))^2 (B_{t_i} - B_{t_{i-1}})^2 - (f'(B_{s_i}))^2 (t_i - t_{i-1})] \\
&+ \sum_i (f'(B_{s_i}))^2 (t_i - t_{i-1}),
\end{aligned}
$$

当 Δ 变细趋于 0 时, 因为 f' 有界, 所以前一个和将均方趋于 0, 第二个和是 Riemann 和, 趋于通常的积分, 因此

$$
\langle f(B) \rangle_t = \int_0^t [f'(B_s)]^2 \mathrm{d}s.
$$

15. 习题: 设 $X_t = tB_t$, 求 X 的二次变差 $\langle X \rangle$. 证明: $X_t := tB_t - \int_0^t B_s \mathrm{d}s$ 是鞅.

16. 在本章中, 设 $B = (B_t)$ 是 Brown 运动, 它生成的信息流记为 $(\mathscr{F}_t)$, $\mathscr{F}_t$ 看起来似乎与概率无关, 但实际上不体现概率的流是没有意义的, 所以我们必须要把所有的 $\mathbb{P}$- 概率等于零的集合及其子集都加入到每个 $\mathscr{F}_t$ 里去. 经过这样处理的流 $(\mathscr{F}_t)$ 称为 Brown 流.

17. 思想: 接着我们想定义一个随机过程 $F = (F_t)$ 关于 Brown 运动的 Riemann-Stieltjes 积分, 但是 Brown 运动是很不光滑的, 想按它的样本轨道来定义积分

$$
\int_0^t X_s(\omega) \mathrm{d}B_s(\omega)
$$

肯定是徒劳的. 但是我们可以在均方收敛的意义下 (或者在更弱的依概率收敛的意义下) 来定义 Riemann 和的收敛, 这真是 Itô 随机积分的思想.

18. 平方可积鞅: 不妨考虑有限区间 $[0, T]$, 一个随机过程 $M = (M_t)$ 称为连续平方可积鞅, 如果

    (a) 连续: 几乎所有轨道连续,

    (b) 平方可积: 对任何 $t \in [0, T]$, $\mathbb{E}M_t^2 < \infty$,

    (c) 适应: 对任何 $t \in [0, T]$, 有 $M_t \in \mathscr{F}_t$,

    (d) 鞅性: 对任何 $T \geq t > s \geq 0$ 有

$$
\mathbb{E}[M_t | \mathscr{F}_s] = M_s.
$$

这里的流是 Brown 流, 当然关于 Brown 流的鞅也是通常的鞅. 牢记一个鞅是由它的终端值 M_T 决定的. 平方可积鞅全体是线性空间, 里面的鞅如同一个向量, 有自然的长度概念,

$$\|M\| := \{\mathbb{E}[M_T^2]\}^{1/2},$$

我们可以证明它有完备性, 也就是说 Cauchy 列有极限. 这对于定义随机积分至关重要.

定理7.3 若 $M^{(n)}$ 是平方可积鞅序列使得当 n, m 都趋于无穷时, $\left\|M^{(n)} - M^{(m)}\right\|$ 趋于零, 那么存在一个平方可积鞅 $M = (M_t)$ 使得 $\left\|M^{(n)} - M\right\|$ 趋于零.

19. 注释: 要证明完备性, 关键是证明极限鞅 $M = (M_t)$ 是连续过程, 这时重要的 Doob 不等式登场了.

$$\mathbb{E}[\max\{M_s^2 : 0 < s < t\}] \leq \frac{1}{4}\mathbb{E}[M_t^2]. \tag{7.5}$$

证明这个不等式就是要用到最初的 Doob 鞅基本定理, 但过程颇为复杂, 感兴趣的读者参考 [4].

20. 一次变差: 首先, 与 Brown 运动一样, 除了恒等于常数 (关于 t), 平方可积鞅的几乎所有轨道的一次变差总是无限的. 粗略地证明如下, 利用恒等式

$$\begin{aligned}\mathbb{E}[M_T^2 - M_0^2] &= \mathbb{E}\left[\sum_i (M_{t_i}^2 - M_{t_{i-1}}^2)\right] \\ &= \mathbb{E}\left[\sum_i (M_{t_i} - M_{t_{i-1}})^2\right].\end{aligned}$$

右边不超过

$$\mathbb{E}[\max_i |M_{t_i} - M_{t_{i-1}}| \cdot V(M)],$$

其中 $V(M)(\omega)$ 是 $M_t(\omega), t \in [0, T]$ 的一次变差. 如果一次变差有限, 当划分变细趋于零时, 此量由连续性会趋于零, 因此

$$\mathbb{E}[(M_T - M_0)^2] = \mathbb{E}[M_T^2 - M_0^2] = 0,$$

推出 $M_T = M_0$, 进而对任何 t 有 $M_t = M_0$.

21. 二次变差: 平方可积鞅的二次变差过程存在唯一, 这是随机分析中最重要的定理. 唯一性不难证明, 但存在性几乎是随机分析最难的一个定理, 是更一般的 Doob-Meyer 定理的特殊情况, Doob-Meyer 定理是离散时间的 Doob 分解定理的连续时间版本.

定理7.4 当 $[0,T]$ 的划分变细趋于零时, 二次变差

$$\sum_i (M_{t_i\wedge t} - M_{t_{i-1}\wedge t})^2$$

收敛于 $\langle M\rangle_t$, 它是唯一一个使得 $\langle M\rangle_0 = 0$ 且使得 $M_t^2 - \langle M\rangle_t$ 是鞅的连续增过程.

证明的思想: 对 $[0,T]$ 上的划分 Δ 定义

$$X_t^\Delta := M_t^2 - \sum_i (M_{t_i\wedge t} - M_{t_{i-1}\wedge t})^2,\ t\in[0,T],$$

证明 X^Δ 是平方可积鞅且当 Δ 变细趋于零时是平方可积鞅空间的 Cauchy 列, 那么它的极限就是 $M^2 - \langle M\rangle$. 参考 [4].

22. 二次协变差: 设 $M=(M_t)$ 和 $N=(N_t)$ 都是平方可积鞅, 二次协变差

$$\langle M,N\rangle_t := \frac{1}{2}(\langle M+N\rangle_t - \langle M\rangle_t - \langle N\rangle_t)$$

是

$$\sum_i (M_{t_i\wedge t} - M_{t_{i-1}\wedge t})(N_{t_i\wedge t} - N_{t_{i-1}\wedge t})$$

当划分趋于零时的极限. 容易验证 $\langle M,N\rangle$ 是

(a) 两个连续增过程的差;

(b) 过程 $M_tN_t - \langle M,N\rangle_t$ 是鞅,

另外它是满足这两个条件的唯一过程. $\langle M,N\rangle$ 有类似内积的性质:

(a) 对称性: $\langle M,N\rangle = \langle N,M\rangle$;

(b) 线性: $\langle cM + c'M', N\rangle = c\langle M,N\rangle + c'\langle M',N\rangle$;

(c) 如果 M,N 独立, 那么 $\langle M,N\rangle = 0$.

23. 习题: 设 $B^{(1)}$ 与 $B^{(2)}$ 是两个独立的 Brown 运动, 证明: 它们的乘积是鞅, 从而证明 $\langle B^{(1)}, B^{(2)}\rangle = 0$.

第 八 讲　随机积分与伊藤公式

摘要 随机积分是 Itô 引入的, 所以也称为 Itô 积分, 它是随机分析最重要的概念, 它看起来象 Riemann-Stieltjes 积分, 但显然不是通常意义下的, 它的存在性是利用平方可积鞅空间的完备性证明的, 当然与 Brown 运动的二次变差过程密切相关. 有了随机积分, 就有怎么计算随机积分的问题, 通过刻画定理, 随机积分的许多性质类似于通常积分, 但是其分部积分公式与变量替换公式不同于通常积分, 后者称为伊藤公式, 它可媲美于微积分中伟大的 Newton-Leibniz 公式, 本章介绍随机积分的定义性质以及分部积分公式与伊藤公式.

关键词
- 随机积分
- 连续半鞅
- 半鞅分解
- 分部积分公式
- 伊藤公式
- 指数鞅

1. 左端点的 Riemann 和: 先简单地假设 $F=(F_t)$ 是连续有界且关于 Brown 流适应的随机过程. Δ 是区间 $[0,T]$ 的划分, 定义取左端点的 Riemann 和随机过程

$$(F^{\Delta}.B)_t := \sum_{i=1}^{n} F_{t_{i-1}}(B_{t_i\wedge t}-B_{t_{i-1}\wedge t}),\ t\le T,$$

其中 F^{Δ} 是分段定义的随机过程,

$$F_t^{\Delta} := F_{t_{i-1}},\ t\in(t_{i-1},t_i].$$

这样写的好处是我们可以把它看成为 n 个随机过程的和.

引理8.1 过程 $(F^{\Delta}.B)$ 是平方可积鞅且有 Itô 等距

$$\mathbb{E}(F^{\Delta}.B)_T^2=\mathbb{E}\left(\int_0^T (F_s^{\Delta})^2\mathrm{d}s\right). \tag{8.1}$$

2. 证明: (1) 平方可积鞅: 我们只要证明对任何 i,

$$\begin{aligned} M_t^{(i)} &:= F_{t_{i-1}\wedge t}(B_{t_i\wedge t}-B_{t_{i-1}\wedge t}) \\ &= \begin{cases} 0, & t\le t_{i-1}; \\ F_{t_{i-1}}(B_t-B_{t_{i-1}}), & t\in(t_{i-1},t_i); \\ F_{t_{i-1}}(B_{t_i}-B_{t_{i-1}}), & t\ge t_i \end{cases} \end{aligned} \tag{8.2}$$

是平方可积鞅就可以了. 连续性是显然的, 平方可积是没有问题的, 因为 F 是有界的. (2) Itô 等距: $(F^{\Delta}.B)_T^2$ 分成平方项和交叉项, 对于平方项, 利用独立增量的性质推出 $F_{t_{i-1}}$ 与 $B_{t_i}-B_{t_{i-1}}$ 是独立的, 故有

$$\begin{aligned} &\mathbb{E}\left(\sum_i F_{t_{i-1}}^2(B_{t_i}-B_{t_{i-1}})^2\right) \\ &=\sum_i \mathbb{E}\left[F_{t_{i-1}}^2\right]\mathbb{E}\left[(B_{t_i}-B_{t_{i-1}})^2\right] \\ &=\sum_i \mathbb{E}\left[F_{t_{i-1}}^2\right](t_i-t_{i-1}) \\ &=\mathbb{E}\left(\int_0^T (F_t^{\Delta})^2\mathrm{d}t\right). \end{aligned}$$

对于交叉项, 也利用条件期望性质可以证明它们都等于 0. 留作习题.

3. 习题: 设 $0 \le t_1 < t_2 \le t_3 < t_4$, 随机变量 ξ, η 分别关于 $\mathscr{F}_{t_1}$ 与 $\mathscr{F}_{t_3}$ 可测, 证明:
$$\mathbb{E}[\xi(B_{t_2} - B_{t_1})\eta(B_{t_4} - B_{t_3})] = 0.$$

4. 收敛: 因为
$$\lim_{\Delta \to 0} \int_0^T (F_s^\Delta - F_s)^2 \mathrm{d}s = 0,$$
所以当 Δ 变细趋于零时 $(F^\Delta.B)_t$ 均方收敛于某个平方可积随机变量, 记为 $(F.B)_t$ 或
$$\int_0^t F_s \mathrm{d}B_s,$$
它被称为是 F 关于 B 在区间 $[0,t]$ 上的随机积分. 由平方可积鞅空间的完备性, 它作为一个随机过程是平方可积鞅且有 Ito 等距
$$\mathbb{E}\left(\int_0^t F_s \mathrm{d}B_s\right)^2 = \mathbb{E}\left(\int_0^t F_s^2 \mathrm{d}s\right). \tag{8.3}$$
如果去掉有界性, 设 F 是连续适应随机过程满足可积性条件
$$\mathbb{E}\left(\int_0^T F_t^2 \mathrm{d}t\right) < \infty, \tag{8.4}$$
这时对任何 n, $F \wedge n$ 关于 B 的随机积分已经被定义了, 且
$$\mathbb{E}\left(\int_0^T (F_t - F_t \wedge n)^2 \mathrm{d}t\right) \longrightarrow 0,$$
因此由 Itô 等距 $(F \wedge n).B$ 有极限, 它就是 F 关于 B 的随机积分 $F.B$. 可以随机积分的随机过程类还可以再扩大, 但我们不想这么做了. 另外上面的积分条件是保证随机积分是平方可积鞅的本质条件.

5. 随机积分: 我们把上面所说的稍加整理, 其实要理解这些定义, 还是需要知道比微积分多一点点的数学知识. 鞅的定义有两个要素: 概率测度和流. 平方可积鞅是一个有内积的线性空间, 用 $\mathscr{H}^2$ 表示, 它由概率测度以及流决定, 它与通常的欧式空间有类似的地方, 向量有长度, 向量之间有角度, 所以有垂直的概念, 另外它是完备的, 所以是 Hilbert 空间, 它里面的连续平方可积鞅全体是它的一个闭子空间, 用 $\mathscr{H}_c^2$. 闭的意思就是 $\mathscr{H}_c^2$ 中的序列取极限后还在它

里面. 把满足上面可积性条件的适应过程 $F=(F_t)$ 全部放在一起, 记为 $\mathscr{L}^2$, 它也是一个 Hilbert 空间, 也是由概率测度和流决定. 现在随机积分就是把空间 $\mathscr{L}^2$ 中的过程 F 变成 $\mathscr{H}_c^2$ 中的过程 $F.B$,

$$F\mapsto F.B,$$

不妨称为 Itô 映射, 它是一个线性映射或变换. 具体怎么做到的呢? 先对 $\mathscr{L}^2$ 中简单的过程全体 $\mathscr{L}_0$ 定义 Itô 映射, 然后建立 Itô 等距

$$\mathbb{E}[(F.B)_T^2]=\mathbb{E}\left(\int_0^T F_t^2\mathrm{d}t\right),$$

最后用极限方法把变换的定义域延拓到 $\mathscr{L}_0$ 的闭包 $\mathscr{L}^2$ 上就是所谓的随机积分了. 以上这个等距变换肯定是单射, 但未必是满射, 在最后一讲我们会证明, 如果流是 Brown 流的话, 那么它就是满射, 这是所谓的鞅表示定理.

6. 运算规则: 以上积分的存在性以及二次变差的存在性的证明都不是容易理解的事情, 尤其对数学基础不好的读者, 不过他们也不必在此过多纠缠, 只需要熟知它们的运算法则就可以了, 如同对于通常的积分一样, 很多非数学专业的学生也只是知道怎么积分, 从来没去想过为什么. 对于任何满足可积性

$$\mathbb{E}\left(\int_0^T F_t^2\mathrm{d}t\right)<\infty$$

的连续适应随机过程 $F=(F_t)$, 它关于 Brown 运动的随机积分是一个关于时间 t 的随机过程

$$\int_0^t F_s\mathrm{d}B_s,$$

它是取左端点的 Riemann 和

$$\sum_i F_{t_{i-1}}(B_{t_i\wedge t}-B_{t_{i-1}\wedge t})$$

当划分变细趋于零时的均方收敛 (或者依概率收敛) 的极限, 这个随机积分的过程也简单记为 $F.B=\{(F.B)_t:t\in[0,T]\}$. 另外它的二次变差过程是

$$\langle F.B\rangle_t=\int_0^t F_s^2\mathrm{d}s. \tag{8.5}$$

7. (8.5) 蕴含着 Itô 等距公式 (8.3), 它们都非常重要, 要证明这个, 只需证明下面两点, 留作习题.

   (a) (8.2) 中定义的 $M^{(i)}$ 有

$$\begin{aligned}\langle M^{(i)}\rangle_t &= F_{t_{i-1}}^2(t\wedge t_i - t\wedge t_{i-1}),\\ &= F_{t_{i-1}}^2\int_{t_{i-1}\wedge t}^{t_i\wedge t}\mathrm{d}s\\ &= \int_0^t F_{t_{i-1}}^2 1_{\{s\in(t_{i-1},t_i]\}}\mathrm{d}s.\end{aligned}$$

   (b) 对不同的 i, j 有

$$\langle M^{(i)}, M^{(j)}\rangle = 0.$$

   由协变差的唯一性, 这等价于证明乘积过程 $M^{(i)}M^{(j)}$ 是鞅.

8. 关于平方可积鞅的随机积分: 类似地可以定义 F 关于平方可积鞅 M 的随机积分

$$\int_0^t F_s \mathrm{d}M_s$$

或者 $F.M$, 它仍然是一个平方可积鞅且有

$$\langle F.M\rangle_t = \int_0^t F_s^2 \mathrm{d}\langle M\rangle_s,$$

右边是一个通常积分, 因为 $\langle M\rangle$ 是递增连续函数. 这时被积过程 $F=(F_t)$ 应该连续适应且满足

$$\mathbb{E}\left(\int_0^T F_t^2\mathrm{d}\langle M\rangle_t\right) < \infty,$$

而且后面我们总是如此假设. 下面的随机积分刻画定理非常强大非常有用.

定理8.1 随机积分 $F.M$ 是满足下面条件的唯一的平方可积鞅: 对任何平方可积鞅 N 有

$$\langle F.M, N\rangle_t = \int_0^t F_s \mathrm{d}\langle M, N\rangle_s.$$

注意右边是通常积分.

9. 性质: 随机积分与通常积分不太一样, 它实际上不是函数关于函数的积分, 写成通常积分的样子可能会导致误解, 但之所以这么写之所以称为积分, 是因为它的性质很像通常的积分, 比如它是线性的

$$\int_0^t (c_1 F_s^{(1)} + c_2 F_s^{(2)}) \mathrm{d}B_s = c_1 \int_0^t F_s^{(1)} \mathrm{d}B_s + c_2 \int_0^t F_s^{(2)} \mathrm{d}B_s, \tag{8.6}$$

其中 c_1, c_2 是常数. 另外, 如果有两个连续的积分, 那么

$$\int_0^t G_s \mathrm{d} \int_0^s F_u \mathrm{d}B_u = \int_0^t G_s F_s \mathrm{d}B_s. \tag{8.7}$$

这两个性质可以简单地用随机积分刻画定理证明. 比如证明 (8.7), 任取平方可积鞅 N,

$$\begin{aligned}
\langle G.(F.B), N \rangle_t &= \int_0^t G_s \mathrm{d} \langle F.B, N \rangle_s \\
&= \int_0^t G_s \mathrm{d} \int_0^s F_u \mathrm{d} \langle B, N \rangle_u \\
&= \int_0^t G_s F_s \mathrm{d} \langle B, N \rangle_s \\
&= \langle GF.B, N \rangle_t,
\end{aligned}$$

由唯一性, $G.(F.B) = GF.B$, 其中 $GF = (G_t F_t : t \in [0, T])$ 表示两个随机过程的乘积. 下面的例子说明它和通常积分很不一样的一面.

10. 习题: 设 F, G 是两个连续适应过程, 证明

$$\langle F.B, G.B \rangle_t = \int_0^t F_s G_s \mathrm{d}s.$$

11. 注释: 其实这里和后面关于随机积分的理论中许多问题都要用局部鞅的概念, 否则很多结论是无法推广的, 局部鞅是需要用停时来定义的, 且关于它的许多结论也是用停时这个工具来证明的, 已经超出了本讲义的范围.

12. 例: 求随机积分

$$\int_0^t B_s \mathrm{d}B_s.$$

不妨设 Δ 是 $[0, t]$ 的划分, 那么

$$\sum_{i=1}^n B_{t_i} (B_{t_i} - B_{t_{i-1}})$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} B_{t_i}^2 - \sum_{i=1}^{n} B_{t_i} B_{t_{i-1}} \\
&= \sum_{i=1}^{n} [B_{t_{i-1}}^2 - B_{t_i} B_{t_{i-1}}] + B_t^2 - B_0^2 \\
&= \sum_{i=1}^{n} B_{t_{i-1}} (B_{t_{i-1}} - B_{t_i}) + B_t^2 - B_0^2,
\end{aligned}
$$

因此左端点的 Riemann 和

$$
\begin{aligned}
&2\sum_{i=1}^{n} B_{t_{i-1}} (B_{t_i} - B_{t_{i-1}}) \\
&\qquad = -\sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2 + B_t^2 - B_0^2,
\end{aligned}
$$

取极限, 左边的极限是随机积分, 右边的和的极限是二次变差, 得

$$
2\int_0^t B_s \mathrm{d}B_s = B_t^2 - B_0^2 - t.
$$

由此看出随机积分与二次变差密切相关, 另外还看出微积分基本定理在这里无效, 因为按照微积分基本定理答案应该是

$$
2\int_0^t B_s \mathrm{d}B_s = B_t^2 - B_0^2.
$$

13. 习题: 设 M 是平方可积鞅, 如果 $M^2 = (M_t^2)$ 也是鞅, 证明: 对任何 t, $M_t = M_0$. 证明: 由鞅的性质 $\mathbb{E}(M_t - M_0)^2 = \mathbb{E}(M_t^2 - M_0^2) = 0$, 因此 $M_t = M_0$.

14. 习题: 设 M 是平方可积鞅, 如果对任何平方可积鞅 N 有 $\langle M, N\rangle = 0$, 证明: 对任何 t, $M_t = M_0$.

15. 设 $B = (B_t)$ 是 Brown 运动, $(\mathscr{F}_t)$ 是 Brown 流. 回忆一个适应的随机过程 X 是指对任何 t, X_t 关于 $\mathscr{F}_t$ 可测, 也就是说 X_t 由 Brown 运动 t 时间前的信息确定. 注意我们下面提到的随机过程都是连续的.

16. 有限一次变差过程: 是指一个适应的随机过程 $V = (V_t)$, 它的几乎所有样本轨道是连续且有有限一次变差的. 这样对于任意连续适应的随机过程 $F = (F_t)$ 我们可以定义通常的积分

$$
\int_0^t F_s \mathrm{d}V_s
$$

它是依样本轨道取左端点的 Riemann 和的极限来定义的, 得到的仍然是一个有限一次变差过程, 也记为 $F.V$.

17. 连续半鞅: 连续半鞅的引入是为了叙述的方便. 如果一个随机过程 X 可以写成一个平方可积鞅 M 和一个有有限一次变差的连续适应随机过程 V 的和,

$$X_t = M_t + V_t,$$

那么我们说 X 是连续半鞅. 我们总是可以要求 $V_0 = 0$ 也总是如此要求, 因为这样可以保证上面这样的分解是唯一的, 否则可能会相差一个常数. 分解 $X = M + V$ 称为半鞅分解.

18. 随机积分: 设 F 是连续适应过程, 自然地定义 F 关于半鞅 X 的随机积分

$$(F.X)_t = \int_0^t F_s \mathrm{d}X_s = \int_0^t F_s \mathrm{d}M_s + \int_0^t F_s \mathrm{d}V_s, \tag{8.8}$$

也就是说

$$F.X = F.M + F.V.$$

它是 F 关于 X 取左端点的 Riemann 和 (至少是依概率收敛) 的极限. 因为 $F.M$ 是鞅, $F.V$ 有有限一次变差, 所以 $F.X$ 仍然是半鞅. 再来看二次变差.

引理8.2 设 $K = (K_t)$ 是连续适应过程, 那么有限一次变差 V 与它的二次协变差 $\langle K, V\rangle = 0$.

证明: 因为在划分 Δ 上的协变差

$$\begin{aligned}\sum_i &|(K_{t_i} - K_{t_{i-1}})(V_{t_i} - V_{t_{i-1}})| \\ &\leq \max_i |K_{t_i} - K_{t_{i-1}}| \cdot V(\Delta),\end{aligned}$$

而 $V(\Delta)$ 有界, K 连续, 所以当划分变细趋于零时, 右边趋于零. 引理说明 V 与任何连续适应过程的二次协变差都等于零, 所以有下面的性质.

(a) $\langle X\rangle = \langle M\rangle$.

(b) 设另一个半鞅 $Y = N + U$, 那么 $\langle X, Y\rangle = \langle M, N\rangle$.

有限一次变差的部分在做协变差时可以忽略.

19. 这样我们可以说关于半鞅的随机积分拓展了通常的积分, 当鞅部分等于零时, 关于半鞅的积分就是通常积分, 只有当鞅部分出现时, 才是真正的随机积分.

20. 随机积分的分部积分公式: 如果 U, V 是有限一次变差过程, 那么分部积分公式可以写成

$$U_tV_t - U_0V_0 = \int_0^t U_s\mathrm{d}V_s + \int_0^t V_s\mathrm{d}U_s. \tag{8.9}$$

应用平行四边形法则

$$U_tV_t = \frac{1}{2}[(U_t+V_t)^2 - V_t^2 - U_t^2],$$

上面的公式可以由 $U = V$ 时的形式来表达

$$U_t^2 - U_0^2 = 2\int_0^t U_s\mathrm{d}U_s.$$

对于 Brown 运动,

$$B_t^2 - B_0^2 = 2\int_0^t B_s\mathrm{d}B_s + t,$$

后面的 t 是 B 的二次变差. 完全类似地可以证明, 如果 $M = (M_t)$ 是平方可积鞅时, 那么

$$M_t^2 - M_0^2 = 2\int_0^t M_s\mathrm{d}M_s + \langle M\rangle_t. \tag{8.10}$$

还是仿照这个证明, 我们得到下面关于半鞅的分部积分公式.

定理8.2 设 X 是连续半鞅, 那么

$$X_t^2 - X_0^2 = 2\int_0^t X_s\mathrm{d}X_s + \langle X\rangle_t. \tag{8.11}$$

利用平行四边形法则可以推出, 对任何连续半鞅 X, Y

$$X_tY_t - X_0Y_0 = \int_0^t X_s\mathrm{d}Y_s + \int_0^t Y_s\mathrm{d}X_s + \langle X, Y\rangle_t. \tag{8.12}$$

如果 X, Y 之一没有鞅部分, 那么 $\langle X, Y\rangle = 0$, 上式就是通常的分部积分公式.

21. 仿照通常的积分, 我们也可以用微元的写法, 比如公式 (8.12) 可写成

$$\mathrm{d}X_tY_t = X_t\mathrm{d}Y_t + Y_t\mathrm{d}X_t + \mathrm{d}\langle X, Y\rangle_t,$$

这样用起来有时候会方便些. 有时候下面的写法更方便

$$XY - X_0Y_0 = X.Y + Y.X + \langle X, Y\rangle,$$

其中 X_0Y_0 看成恒等于随机变量 X_0Y_0 的随机过程.

22. 例: 随机过程 $X_t = tB_t$, 它是两个半鞅的乘积, 用分部积分公式, 得到它的半鞅分解

$$tB_t = \int_0^t s\mathrm{d}B_s + \int_0^t B_s\mathrm{d}s,$$

那么它的二次变差是

$$\langle X\rangle = \int_0^t s^2\mathrm{d}s = \frac{1}{3}t^3.$$

类似地, 随机过程 $X_t = tB_t^2$ 的半鞅分解

$$\begin{aligned} tB_t^2 &= \int_0^t s\mathrm{d}B_s^2 + \int_0^t B_s^2\mathrm{d}s \\ &= 2\int_0^t sB_s\mathrm{d}B_s + \int_0^t (s + B_s^2)\mathrm{d}s, \end{aligned}$$

它的二次变差

$$\langle X\rangle_t = 4\int_0^t s^2B_s^2\mathrm{d}s.$$

求期望

$$\mathbb{E}\langle X\rangle_t = 4\int_0^t s^2\mathbb{E}[B_s^2]\mathrm{d}s = t^4.$$

23. 例: B^2 是半鞅, 它的分解是 $B^2 = 2B.B + t$, 其中 t 表示 t 时刻等于 t 的随机过程. 根据 (8.12)

$$\begin{aligned} B^3 &= B^2.B + B.B^2 + \langle B^2, B\rangle \\ &= B^2.B + 2B^2.B + B.t + \langle 2B.B + t, B\rangle \\ &= 3B^2.B + 3B.t \end{aligned}$$

也就是说 B^3 也是半鞅, $3B^2.B$ 是鞅部分, $3B.t$ 是有限一次变差部分. 用归纳法以及分部积分公式可以证明

$$B_t^n - B_0^n = n\int_0^t B_s^{n-1}\mathrm{d}B_s + \frac{n(n-1)}{2}\int_0^t B_s^{n-2}\mathrm{d}s.$$

这时, 通过观察并且根据随机积分的线性性质可以严格地证明, 如果 f 是一个多项式, 那么

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\mathrm{d}s.$$

这就是著名的伊藤公式之雏形.

24. 伊藤 (Itô) 公式:

定理8.3 设 f 是二次连续可导, 那么

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\mathrm{d}s. \tag{8.13}$$

类似微元的形式

$$\mathrm{d}f(B_t) = f'(B_t)\mathrm{d}B_t + \frac{1}{2}f''(B_t)\mathrm{d}t. \tag{8.14}$$

类似地, 平方可积鞅的伊藤公式是

$$\mathrm{d}f(M_t) = f'(M_t)\mathrm{d}M_t + \frac{1}{2}f''(M_t)\langle M\rangle_t, \tag{8.15}$$

与半鞅的伊藤公式

$$\mathrm{d}f(X_t) = f'(X_t)\mathrm{d}X_t + \frac{1}{2}f''(X_t)\langle X\rangle_t. \tag{8.16}$$

它们的形式完全一致, 当半鞅 X 不含鞅部分时, $\langle X\rangle = 0$, 这就是通常的 Newton-Leibniz 公式.

25. 伊藤公式的证明: Itô 公式的证明方法通常是两种, 一种是通常多项式逼近, 不那么严格地说, 存在一个多项式序列 f_n, 它连同它的一阶导数二阶导数分别一致收敛于 f 及其一阶导数与二阶导数. 这样就推出 Itô 公式. 但是多项式的收敛必须是在有限区间上的, 所以要严格完成上述证明还需要其他工具. 第二种方法是利用二阶中值定理, 在端点 a, b 的区间上存在点 ξ 使得

$$f(b) - f(a) = f'(a)(b-a) + \frac{1}{2}f''(\xi)(b-a)^2,$$

那么

$$f(B_t) - f(B_0) = \sum_{i=1}^n (f(B_{t_i}) - f(B_{t_{i-1}}))$$

$$= \sum_{i=1}^{n} \left[f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2} f''(B_{s_i})(B_{t_i} - B_{t_{i-1}}) \right],$$

其中 $s_i \in [t_{i-1}, t_i]$, 它是由介值定理确定的. 当 n 趋于无穷时, 前一项均方趋于随机积分, 后一项均方趋于二次变差积分. 要证明后一项的极限, 只要证

$$\mathbb{E} \left(\sum_i f''(B_{s_i})(B_{t_i} - B_{t_{i-1}})^2 - \sum_i f''(B_{s_i})(t_i - t_{i-1}) \right)^2$$

收敛于零就够了, 而上式不会超过

$$m^2 \cdot \mathbb{E} \left(\sum_i (B_{t_i} - B_{t_{i-1}})^2 - t) \right)^2$$

其中 m 是二阶导数 f'' 在区间 $[0, t]$ 上的绝对上界, 然后利用 Brown 运动的二次变差的均方收敛就可以了.

26. 半鞅的封闭性: 半鞅的优点是它有很好的封闭性, 首先半鞅是线性空间, 然后有限个半鞅的乘积是半鞅, 由 Itô 公式看出, 半鞅复合一个光滑函数后还是半鞅.

27. 指数鞅公式: 前面在 Brown 运动那一讲中已经说过 Brown 运动的指数鞅性质, 即

$$\mathrm{e}^{aB_t - a^2 t/2}$$

是鞅, 这由 Itô 公式很容易证明, 把 $X_t = aB_t - a^2 t/2$ 当做半鞅用 Itô 公式

$$\begin{aligned} \mathrm{e}^{aB_t - a^2 t/2} - 1 &= \int_0^t \mathrm{e}^{aB_s - a^2 s/2} \mathrm{d}(aB_s - a^2 s/2) \\ &\quad + \frac{1}{2} \int_0^t \mathrm{e}^{aB_s - a^2 s/2} \mathrm{d}\langle aB_s - a^2 s/2 \rangle \\ &= a \int_0^t \mathrm{e}^{aB_s - a^2 s/2} \mathrm{d}B_s, \end{aligned}$$

这个鞅称为是 Brown 运动的指数鞅. 一般地有下面的定理, 证明类似.

定理8.4 如果 M 是平方可积鞅, 那么

$$\varepsilon(M)_t := \exp\left(M_t - \frac{\langle M \rangle}{2} \right) \tag{8.17}$$

是鞅, 称为 M 的指数鞅, 而且它就是自身关于 M 的随机积分

$$\varepsilon(M)_t = \mathrm{e}^{M_0} + \int_0^t \varepsilon(M)_s \mathrm{d}M_s.$$

直观地说, 指数鞅是给定 Brown 运动下的随机微分式

$$\mathrm{d}X_t = aX_t\mathrm{d}B_t \tag{8.18}$$

的解. 类似这样的随机微分式称为随机微分方程. 随机微分方程类似于常微分方程, 有一套关于强解弱解的存在唯一性理论, 但是真正能这样显式解出来的方程是极少的, 而且很难说有什么系统方法.

28. 例: 求满足初值 $X_0 = x$ 和 (Langevin) 随机微分方程

$$\mathrm{d}X_t = -\alpha X_t\mathrm{d}t + \mathrm{d}B_t,$$

的解 X. 令 $Y_t := \mathrm{e}^{\alpha t}X_t$, 那么

$$\mathrm{d}Y_t = \mathrm{e}^{\alpha t}(\alpha X_t\mathrm{d}t + \mathrm{d}X_t) = \mathrm{e}^{\alpha t}\mathrm{d}B_t,$$

因此

$$Y_t = Y_0 + \int_0^t \mathrm{e}^{\alpha u}\mathrm{d}B_u,$$

换成 X_t 得

$$X_t = \mathrm{e}^{-\alpha t}\left(x + \int_0^t \mathrm{e}^{\alpha u}\mathrm{d}B_u\right).$$

这是著名的 Orstein-Uhlenbeck 过程.

29. 习题: $X_t := B_t^3 - 3tB_t$, 证明: (X_t) 是鞅. 提示: 用 Itô 公式.

30. 习题: (Itô 的一个结果) 证明:

$$n!\int_{0\le u_1\le\cdots\le u_n\le t} \mathrm{d}B_{u_1}\cdots\mathrm{d}B_{u_n} = t^{\frac{n}{2}}h_n\left(\frac{B_t}{\sqrt{t}}\right),$$

其中 h_n 是所谓的 Hermite 多项式

$$h_n(x) = (-1)^n\mathrm{e}^{\frac{x^2}{2}}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(\mathrm{e}^{-\frac{x^2}{2}}\right).$$

31. 习题: 计算 Orstein-Uhlenbeck 过程的协方差

$$\mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)].$$

计算: 设 $t > s$, 那么

$$\begin{aligned}
&\mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)] = \mathrm{e}^{-\alpha t}\mathrm{e}^{-\alpha s}\mathbb{E}\left(\int_0^t \mathrm{e}^{\alpha u}\mathrm{d}B_u \int_0^s \mathrm{e}^{\alpha u}\mathrm{d}B_u\right) \\
&= \mathrm{e}^{-\alpha(t+s)}\mathbb{E}\left(\int_0^s \mathrm{e}^{\alpha u}\mathrm{d}B_u\right)^2 \\
&= \mathrm{e}^{-\alpha(t+s)}\int_0^s \mathrm{e}^{2\alpha u}\mathrm{d}u \\
&= \frac{1}{2\alpha}(\mathrm{e}^{-\alpha(t-s)} - \mathrm{e}^{-\alpha(t+s)}).
\end{aligned}$$

第 九 讲　鞅表示与鞅测度

摘要 在这最后一讲中, 我们将介绍随机分析中两个最重要的定理, 鞅表示定理与测度变换定理. 鞅表示定理是说一个关于 Brown 流的平方可积鞅一定可以写成关于 Brown 运动的随机积分, 它保证了期权定价中无风险投资策略的存在性, 测度变换定理最早称为 Cameron-Martin 定理, 是无限维分析的一个重要定理, 后来推广到一般的概率空间上, 也称为 Girsanov 定理, 它与期权定价的第一基本定理相关. 最后我们用这两个定理推导出 Black-Scholes 的期权定价公式.

关键词

- 鞅表示定理
- 测度变换
- Levy 的 Brown 运动刻画定理
- Girsanov 公式
- Black-Scholes 模型
- 定价公式

1. 在这一讲中我们将着重讨论鞅测度的存在问题以及鞅表示问题, 才离散时间的讨论可以看出, 这两者和金融的期权定价有密切联系. 当然, 在离散时间理论中, 这两个结果几乎都是显然的, 证明非常简单. 你将看到, 这两个结果在连续时间下不是那么容易证明, 有的证明部分只好省略了. 其中用到的主要工具就是上一讲建立起来的随机积分的 Itô 公式和分部积分公式.

2. 鞅表示: 关于 Brown 流的鞅可以用 Brown 运动的随机积分积分表示, 这首先是由 Itô 指出的. 设给定 Brown 运动 $B = (B_t)$ 与加入零概率集以及子集后的 Brown 流 $(\mathscr{F}_t)$, $M = (M_t)$ 是平方可积鞅, 不妨设 $M_0 = 0$, 否则考虑平方可积鞅 $M_t - M_0$ 就可以了. 如果我们能证明存在适应过程 $F = (F_t)$, 使得

$$M_T = \int_0^T F_s \mathrm{d}B_s,$$

那么因为两边都是鞅, 由唯一性得鞅表示

$$M_t = \int_0^t F_s \mathrm{d}B_s.$$

而且 $F = (F_t)$ 是被 M 唯一确定的. 我们知道平方可积鞅的终端值 M_T 组成的空间是与关于 $\mathscr{F}_T$ 可测的平方可积随机变量组成的空间 $L^2(\mathscr{F}_T)$ 等距同构的. 因此实际上我们要证明下面的定理.

定理9.1 (Itô) 设随机变量 ξ 平方可积, 且关于 $\mathscr{F}_T$ 可测, 那么存在唯一的适应随机过程 $F = (F_s)$ 使得

$$\xi - \mathbb{E}[\xi] = \int_0^T F_s \mathrm{d}B_s.$$

3. 证明概述: 不妨设 $\mathbb{E}[\xi] = 0$. 证明的思想对于一个有数学基础的人不难理解. 我们把可以这样写的随机变量称为可以表示的. 首先要验证可以表示的随机变量全体是一个闭的线性子空间. 考虑指数鞅, 因为

$$\mathrm{e}^{aB_t - \frac{a^2 t}{2}} - 1 = a\int_0^t \mathrm{e}^{aB_s - \frac{a^2 s}{2}} \mathrm{d}B_s,$$

所以左边这样的随机变量是可以表示的. 再考虑更一般些的指数鞅, 取 f 是 $[0,T]$ 上的分段连续函数, 鞅

$$(if.B)_t := \int_0^t if(s)\mathrm{d}B_s$$

的指数鞅是

$$\varepsilon_t(f) := \exp\left(i\int_0^t f(s)\mathrm{d}B_s + \frac{1}{2}\int_0^t f^2(s)\mathrm{d}s\right),$$

它满足下面的方程

$$\varepsilon_t(f) - 1 = \int_0^t \varepsilon_s(f)\mathrm{d}(if.B)_s = i\int_0^t \varepsilon_s(f)f\mathrm{d}B_s.$$

也就是说, 左边这样的随机变量仍然是可以表示的. 最后只需要证明这样的随机变量全体 $\{\varepsilon_T(f)-1 : f\}$ 在 $L^2(\mathscr{F}_T)$ 中稠密就足够了. 也就是说如果有某个期望为零的平方可积随机变量 η 与这个集合正交, 即对任何这样的 f 有

$$\mathbb{E}[\eta\cdot\varepsilon_T(f)] = \mathbb{E}[\eta\cdot(\varepsilon_T(f)-1)] = 0, \tag{9.1}$$

那么 $\eta = 0$. 而最后这一步在数学上是最难的, 需要使用 Fourier 变换唯一性的思想.

4. 证明的细节: 看得头晕的读者可以忽略了. (9.1) 等价于

$$\mathbb{E}\left[\eta\exp\left(i\int_0^t f(s)\mathrm{d}B_s\right)\right] = 0.$$

任取点 $0 < t_1 < \cdots < t_n = T$, 和 $x_1, \cdots, x_n \in \mathbf{R}$, 代入

$$f(t) = \sum_i x_i 1_{t\in(t_{i-1},t_i]}$$

得

$$\mathbb{E}\left[\eta\exp\left(ix_1B_{t_1} + ix_2(B_{t_2}-B_{t_1}) + \cdots + ix_n(B_{t_n}-B_{t_{n-1}})\right)\right] = 0,$$

由于 $x_1, \cdots, x_n$ 是任意的, 推出对任何 $x_1, \cdots, x_n \in \mathbf{R}$ 有

$$\mathbb{E}\left[\eta\exp\left(ix_1B_{t_1} + ix_2B_{t_2} + \cdots + ix_nB_{t_n}\right)\right] = 0,$$

由在第二讲中解释的关于特征函数唯一性 (参考第 14 页) 得, 后面的指数函数可以用任何有界函数替代推出

$$\mathbb{E}\left[\eta\cdot g(B_{t_1}, \cdots, B_{t_n})\right] = 0,$$

因为 $g(B_{t_1}, \cdots, B_{t_n})$ 这样的随机变量全体在 $L^2(\mathscr{F}_T)$ 中稠密 (弄明白这句话也不是很容易的事情), 故 $\eta = 0$ a.s.

5. 注释: 和离散时间的证明不同, 那里我们是可以把表示中的 H_n 从 $n = N-1$ 开始倒着一步一步地写出来的, 这也是倒向随机微分方程的最初思想. 但上面 Itô 表示定理的证明是一个存在性证明, 不是构造性证明, 因为我们实际上最终并没有把 F 构造出来. 在实际应用中, 特别是金融市场的应用中, 可能更重要的是把 F 计算出来. 这个定理说明关于鞅流的平方可积鞅一定是连续的.

6. 鞅与指数鞅之间的表示: 前面二项期权中实际上已经讨论过鞅表示的问题, 实际上, 如果 $\{X_n : 0 \le n \le N\}$ 是简单随机游动, 那么它是鞅, 设 $\{Y_n\}$ 是关于鞅流 $\{\mathscr{F}_n\}$ 的鞅, 也就是说 Y_n 关于 $X_1, \cdots, X_n$ 可测的. 鞅表示问题是问是否存在 $H_{n-1} \in \mathscr{F}_{n-1}$ 使得

$$Y_n - Y_{n-1} = H_{n-1}(X_n - X_{n-1})?$$

记 $\xi_n = X_n - X_{n-1}$, 它取值 $1, -1$ 的概率各 $1/2$. 假设 Y_n 已经知道了, 那么我们有方程

$$\begin{aligned} Y_n|_{\xi_n=1} - Y_{n-1} &= H_{n-1} \\ Y_n|_{\xi_n=-1} - Y_{n-1} &= -H_{n-1}, \end{aligned}$$

因此

$$H_{n-1} = (Y_n|_{\xi_n=1} - Y_n|_{\xi_n=-1})/2,$$

其中的量只依赖于 $X_1, \cdots, X_{n-1}$, 因为 ξ_n 给定了. 鞅表示成立的主要原因是 ξ_n 恰好取两个值. 鞅也可以通过 $\{X_n\}$ 的指数鞅来表示, 这也很有意义, 因为通常股票价格是用指数鞅来模拟的, 因为它是非负的, 比较符合实际. 例如, 因为

$$\mathbb{E}[2^{\xi_n}] = 5/4,$$

所以 $Z_n = (4/5)^n 2^{X_n}$ 是 $\{X_n\}$ 的一个指数鞅, 它可以由 $\{X_n\}$ 鞅表示如下

$$\begin{aligned} Z_n - Z_{n-1} &= (Z_n|_{\xi_n=1} - Z_n|_{\xi_n=-1})/2 \cdot (X_n - X_{n-1}) \\ &= 3/5 \cdot Z_{n-1}(X_n - X_{n-1}), \end{aligned}$$

因此关于鞅流的鞅 Y_n 可以用指数鞅 $\{Z_n\}$ 来表示

$$Y_n - Y_{n-1} = \frac{5H_{n-1}}{3Z_{n-1}}(Z_n - Z_{n-1}).$$

对于平方可积鞅 $M=(M_t)$, 其指数鞅

$$\varepsilon_a(M)_t=\exp\left(aM_t-\frac{a^2}{2}\langle M\rangle_t\right)$$

满足微分方程

$$\mathrm{d}\varepsilon_a(M)_t=a\varepsilon_a(M)_t\mathrm{d}M_t,$$

这已经说明鞅与指数鞅之间可以互相表示.

引理9.1 鞅与其指数鞅是可以互相表示的.

7. 为什么可以表示: 在离散时间情形, 我们说鞅可以表示是因为模型是二项模型, 表示方程恰好可解. 连续时间情形的解释要复杂得多, 从证明看, 关键的地方是 Brown 运动的二次变差过程是确定性的没有随机性但这可能只是技术性条件, 更本质的地方可能是因为 Brown 运动的留给关于 Brown 流的自由度不多不少. 另外 Brown 运动是平方可积鞅且二次变差过程是确定性的过程 t, 这个性质恰好刻画了 Brown 运动, 这就是下面的 Lévy 刻画定理.

定理9.2 (Lévy) 如果一个随机过程 X 是平方可积鞅而且其二次变差 $\langle X\rangle_t=t$, 那么 X 一定是 Brown 运动.

证明: 根据指数鞅公式, 随机过程

$$\mathrm{e}^{izX_t+\frac{1}{2}z^2t}$$

是鞅, 因此对任何 $t>s$ 与实数 z, 有

$$\mathbb{E}\left(\mathrm{e}^{izX_t+\frac{1}{2}z^2t}|\mathscr{F}_s\right)=\mathrm{e}^{izX_s+\frac{1}{2}z^2s},$$

这样等价于

$$\mathbb{E}\left(\mathrm{e}^{iz(X_t-X_s)}|\mathscr{F}_s\right)=\mathrm{e}^{-\frac{1}{2}z^2(t-s)}.$$

而这正是 Brown 运动的特性, 说明了 X_t-X_s 独立于过去 $\mathscr{F}_s$ 且服从方差为 $(t-s)$ 的正态分布, 也即 $X=(X_t)$ 是 Brown 运动. 证明也用到 Fourier 变换的性质.

8. 鞅测度: 鞅测度的问题就是讨论一个不是鞅的随机过程是不是可以通过变换一个等价测度的方式变成鞅. 我们考虑其中最简单的一种情况, 设 $B=(B_t)$ 是在概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 下的 Brown 运动, $X_t=B_t+bt$, 其中 b 是一个非零

实数. 那么 $X=(X_t)$ 不是鞅, 现在我们是不是可以找到一个与 $\mathbb{P}$ 等价的概率 $\widehat{\mathbb{P}}$ 使得 X 在新测度下是鞅? 如果是鞅, 那么因为 $\langle X\rangle_t=t$, 所以根据 Lévy 的刻画定理, X 必然就是 Brown 运动.

9. 测度的变换: 在第五讲中我们已经对测度变换的问题进行过阐述: 概率测度是对随机事件可能性的一种度量法则, 等价概率是一种新的度量法则, 但它保证不改变随机性, 也就是说不会把随机的事件变成确定性, 也不会把确定性的事件变成随机的.

10. 数学上怎么做: 在离散的情况下, 测度的改变是容易的, 在连续的情况下, 我们通常通过密度函数的方式来改变测度, 取一个 $\mathscr{F}$ 可测的随机变量 ζ, 令

$$\widehat{\mathbb{P}}(A):=\mathbb{E}(\zeta 1_A)=\mathbb{E}(\zeta;A).$$

当且仅当 ζ 严格正且 $\mathbb{E}[\zeta]=1$ 时, $\widehat{\mathbb{P}}$ 是一个等价于 $\mathbb{P}$ 的新概率测度, 这里的等价在数学上是指 $\mathbb{P}(A)=0$ 当且仅当 $\widehat{\mathbb{P}}(A)=0$, 随机的还是随机的, 确定的还是确定的. 用 $\widehat{\mathbb{E}}$ 表示概率 $\widehat{\mathbb{P}}$ 的数学期望, 那么对任何随机变量 X, 有

$$\widehat{\mathbb{E}}[X]=\mathbb{E}[\zeta X].$$

11. 习题: 证明上面定义的 $\widehat{\mathbb{P}}$ 是一个等价于 $\mathbb{P}$ 的新概率测度当且仅当 ζ 严格正且 $\mathbb{E}[\zeta]=1$.

12. 现在回到 Brown 运动的那个概率空间, $\mathscr{F}$ 就是 $\mathscr{F}_T$, ζ 可以作为密度的条件是: 严格正, 关于 $\mathscr{F}_T$ 可测且 $\mathbb{E}[\zeta^2]<\infty$, $\mathbb{E}[\zeta]=1$. 令

$$Z_t:=\mathbb{E}[\zeta|\mathscr{F}_t],$$

那么 $Z=(Z_t)$ 是严格正的平方可积鞅. 设 X_T 是 $\mathscr{F}_T$ 可测的随机变量, 那么 $X_t:=\widehat{\mathbb{E}}[X_T|\mathscr{F}_t]$ 是关于概率 $\widehat{\mathbb{P}}$ 的鞅. 对任何 $\mathscr{F}_t$ 可测的随机变量 A, 由条件期望的性质得

$$\begin{aligned}\mathbb{E}[X_tZ_t;A]&=\widehat{\mathbb{E}}[X_t;A]\\&=\widehat{\mathbb{E}}[\mathbb{E}[X_T|\mathscr{F}_t];A]\\&=\widehat{\mathbb{E}}[X_T;A]=\mathbb{E}[X_TZ_T;A]\\&=\mathbb{E}[\mathbb{E}[X_TZ_T|\mathscr{F}_t];A],\end{aligned}$$

推出

$$X_tZ_t = \mathbb{E}[X_TZ_T|\mathscr{F}_t],$$

也就是说乘积过程 XZ 是关于 $\mathbb{P}$ 的鞅.

引理9.2 $X=(X_t)$ 关于概率 $\widehat{\mathbb{P}}$ 是鞅当且仅当 $XZ=(X_tZ_t)$ 关于概率 $\mathbb{P}$ 是鞅.

13. 指数鞅作为密度: 更具体地, 设

$$\zeta = \exp\left(aB_T - \frac{1}{2}a^2T\right),$$

那么 ζ 符合作为密度的条件, $Z=(Z_t)$ 就是指数鞅, 它满足

$$\mathrm{d}Z_t = aZ_t\mathrm{d}B_t,$$

由 ζ 得到新概率测度 $\widehat{\mathbb{P}}$. 现在 $X=(X_t)$ 是 $\widehat{\mathbb{P}}$ 鞅当且仅当 XZ 是 $\mathbb{P}$ 鞅, 但是由分部积分公式

$$X_tZ_t - X_0Z_0 = \int_0^t Z_s\mathrm{d}X_s + \int_0^t X_s\mathrm{d}Z_s + \langle X, Z\rangle_t.$$

特别地, 如果 X 是漂移 Brown 运动 $X_t = B_t + bt$, 那么

$$\langle X, Z\rangle_t = \int_0^t aZ_s\mathrm{d}\langle B\rangle_s = \int_0^t aZ_s\mathrm{d}s,$$

因此

$$X_tZ_t - X_0Z_0 = \left(\int_0^t Z_s\mathrm{d}B_s + \int_0^t X_s\mathrm{d}Z_s\right) + \int_0^t (a+b)Z_s\mathrm{d}s.$$

因为 B 和 Z 都是关于概率 $\mathbb{P}$ 的平方可积鞅, 故前面部分是 $\mathbb{P}$ 鞅, 于是 XZ 是 $\mathbb{P}$ 鞅当且仅当 $a+b=0$. 也就是说, 我们证明了下面的定理, 称为 Cameron-Martin 定理或者 Girsanov 定理. 直观地说, Brown 运动是对称的, 如果它有了漂移, 那么可以改变概率测度使其重新成为 Brown 运动.

定理9.3 $X_t = B_t - at$ 在新概率测度 $\widehat{\mathbb{P}}$ 下是鞅, 于是它在新概率测度下是 Brown 运动.

14. 期权定价理论概述: 最后我们给出随机分析最漂亮的应用: Black-Scholes 公式. 随机积分等的直观意义在这里可以充分展现. 不管连续时间还是离散时

间, 期权定价的基本理论是类似的, 不同的是模型和定价公式. 设 $S=(S_t)$ 是股票在 t 时刻的价格, $r>0$ 是利率, 连续计利. 一个 T 时刻到期且商定价格是 K 的欧式看涨期权在 T 时刻的价值显然是

$$V_T=(S_T-K)^+=\max(0,S_T-K).$$

一般的衍生证券的在 T 时刻的价值是一个 $\mathscr{F}_T$ 可测的非负随机变量 V_T.

15. 自融资: 一个投资策略为 $H=(H_t)$ 的投资人在时刻 t 的财富为

$$X_t=H_tS_t+b_t,$$

其中 H_tS_t 是投资在股票上的钱, b_t 是放在银行的钱, 银行的利率为 r, 连续复利 $\mathrm{d}b_t=rb_t\mathrm{d}t$. 由离散模型 (**??**) 看出, 自融资的意思是财富的增加仅来自股票的增值和银行利息

$$\mathrm{d}X_t=H_t\mathrm{d}S_t+\mathrm{d}b_t, \tag{9.2}$$

如同离散场合一样, 考虑折现后, 由分部积分公式推出

$$\begin{aligned}\mathrm{d}(\mathrm{e}^{-rt}X_t)&=\mathrm{e}^{-rt}(-rX_t\mathrm{d}t+\mathrm{d}X_t)\\&=\mathrm{e}^{-rt}[-r(H_tS_t+b_t)\mathrm{d}t+H_t\mathrm{d}S_t+\mathrm{d}b_t]\\&=H_t\mathrm{d}(\mathrm{e}^{-rt}S_t),\end{aligned}$$

即折现后财富的增加来自于股票折现后的增值, 因此初始投资为 $X_0=x_0$ 的自融资创造的财富折现后为

$$\mathrm{e}^{-rt}X_t=X_0+\int_0^t H_u\mathrm{d}(\mathrm{e}^{-ru}S_u),$$

折现后的财富是 H 关于折现后股票价格的随机积分. 这里随机积分的直观性得到了淋漓尽致的体现.

16. 定价的思想: 如果我们能够找到一个概率测度 $\widehat{\mathbb{P}}$, 使得折现后股票价格在此概率下是鞅, 那么衍生证券在 0 时刻的价格应该是

$$V_0=\mathrm{e}^{-rT}\widehat{\mathbb{E}}[V_T].$$

如果鞅可以表示, 那么由 V_T 确定的鞅

$$\mathrm{e}^{-rt}X_t:=\mathrm{e}^{-rT}\widehat{\mathbb{E}}[V_T|\mathscr{F}_t]$$

可以由 H 关于鞅 $(\mathrm{e}^{-rt}S_t)$ 来表示, H 就是可以拿来对冲因为卖出一份期权所带来的风险的投资策略, 这时上面的价格称为可复制价格. 不可复制的价格实际上是没有意义的.

17. 例: 作为一个特例, 我们来讨论 Black-Scholes 的期权定价模型, Black 与 Scholes 的贡献就是在假设股票价格满足下面随机微分式的条件下给出期权价格的显示表达式,

$$\frac{\mathrm{d}S_t}{S_t} = \sigma \mathrm{d}B_t + \mu \mathrm{d}t, \tag{9.3}$$

其中 $\sigma > 0$ 称为波动率, $\mu \in \mathbf{R}$ 称为收益率. 由分部积分公式

$$\mathrm{d}(\mathrm{e}^{-rt}S_t) = \mathrm{e}^{-rt}\mathrm{d}S_t - r\mathrm{e}^{-rt}S_t\mathrm{d}t$$

因此折现后的股票价格过程 $(\mathrm{e}^{-rt}S_t)$ 满足随机微分式

$$\begin{aligned}\frac{\mathrm{d}(\mathrm{e}^{-rt}S_t)}{\mathrm{e}^{-rt}S_t} &= \frac{\mathrm{d}S_t}{S_t} - r\mathrm{d}t \\ &= \sigma\mathrm{d}B_t + (\mu - r)\mathrm{d}t \\ &= \sigma\left(\mathrm{d}B_t + \frac{\mu - r}{\sigma}\mathrm{d}t\right).\end{aligned}$$

令

$$\widehat{B}_t = B_t - (r - \mu)\sigma^{-1}t.$$

如果我们能够找一个概率 $\widehat{\mathbb{P}}$ 使得 $\widehat{B} = (\widehat{B}_t)$ 在此概率下是鞅, 那么折现后的股票价格在新概率下是鞅. 由测度变换定理, 这只要定义新概率测度

$$\widehat{\mathbb{P}}(A) := \mathbb{E}\left[\exp(aB_T - \frac{1}{2}a^2T); A\right],$$

就可以做到了, 其中 $a = (r - \mu)\sigma^{-1}$. 即在概率 $\widehat{\mathbb{P}}$ 下, $\widehat{B}$ 是 Brown 运动.

18. 期权定价公式: 随机微分式 (9.4) 是可以解的, 由 Itô 公式,

$$\begin{aligned}\mathrm{d}\log S_t &= \frac{\mathrm{d}S_t}{S_t} - \frac{1}{2S_t^2}\mathrm{d}\langle S\rangle_t \\ &= \sigma\mathrm{d}B_t + \mu\mathrm{d}t - \frac{1}{2S_t^2}\sigma^2 S_t^2\mathrm{d}t \\ &= \sigma\mathrm{d}B_t + (\mu - \sigma^2/2)\mathrm{d}t,\end{aligned}$$

因此有

$$S_t = S_0 \exp\left(\sigma B_t + (\mu - \sigma^2/2)t\right).$$

现在

$$V_0 = \mathrm{e}^{-rT}\widehat{\mathbb{E}}[V_T],$$

如果考虑欧式看涨 $V_T = (S_T - K)^+$, 那么

$$\begin{aligned} V_0 &= \mathrm{e}^{-rT}\widehat{\mathbb{E}}\left((S_T - K)^+\right) \\ &= \mathrm{e}^{-rT}\widehat{\mathbb{E}}\left(S_0 \mathrm{e}^{\sigma B_T + (\mu - \sigma^2/2)T} - K\right)^+ \\ &= \mathrm{e}^{-rT}\widehat{\mathbb{E}}\left(S_0 \mathrm{e}^{\sigma \widehat{B}_T + (r - \sigma^2/2)T} - K\right)^+, \end{aligned}$$

其中仅有的随机变量 $\widehat{B}_T$ 在概率 $\widehat{\mathbb{P}}$ 下服从正态分布 $N(0,T)$, 因此期权价格的显式表达式为

$$V_0 = \mathrm{e}^{-rT}\int_{\mathbf{R}} \left(S_0 \mathrm{e}^{\sigma x\sqrt{T} + (r - \sigma^2/2)T} - K\right)^+ \phi(x)\mathrm{d}x,$$

其中 $\phi(x)$ 是标准正态分布的密度函数. 写得更明确点, 用标准正态分布函数 $\Phi(x)$ 表示, 就是 Black-Scholes 公式

$$V_0 = S_0\Phi(d + \sigma\sqrt{T}) - K\mathrm{e}^{-rT}\Phi(d), \tag{9.4}$$

其中

$$d = \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

19. 习题: 证明 $X_t = |B_t|$ 是半鞅.

20. 习题: 最后, 用自己的语言阐述期权定价的思想并且在不看任何书的情况下把期权定价公式 (9.4) 推导出来.

习题解答

1. 习题: 设 $X_1, \cdots, X_n$ 独立且服从标准正态分布, 求 $Y = X_1^2 + \cdots + X_n^2$ 的密度函数.

2. 习题: 证明模拟中产生的随机量 X 服从密度函数 f.

3. 习题: 如果随机变量 X 的分布函数 F 连续, 证明随机变量的函数 $F(X)$ 服从 $[0,1]$ 上均匀分布.

4. 习题: 设 $X \sim N(\mu, \sigma^2)$, 用类似上面的方法计算 $\mathbb{E}\left(\mathrm{e}^{tX}\right)$, $t \in \mathbf{R}$.

5. 习题: 如果 X, Y 的联合密度函数是 $f(x,y)$, A 是非退化二阶方阵, 作变换

$$(U,V) = (X,Y)A.$$

求 (U,V) 的联合密度函数.

6. 习题: 设 X_1, X_2 独立且服从标准正态分布, (1) 求 $Y = X_2/X_1$ 的密度函数. (2) 记 $X = (X_1 + X_2)/2$, 证明: X 与 $(X_1 - X)^2 + (X_2 - X)^2$ 独立.

7. 习题: 如果存在趋于零的正数列 $\{\varepsilon_n\}$ 使得

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty,$$

那么 X_n 几乎处处收敛于 X.

8. 习题: 硬币的正反面分别记为 $1, 0$, 连续掷一个硬币得到一个 01 序列, 求在这个序列中, (1) 11 首次出现的平均时间, (2) 01 首次出现的平均时间, (3) 101 首次出现的平均时间.

9. 习题: 设 X, Y 是两个随机变量, 如果对任何 $x \in \mathbf{R}$, 有

$$\mathbb{E}\left(\mathrm{e}^{ixX} | Y\right) = \mathbb{E}\left(\mathrm{e}^{ixX}\right),$$

应用定理 3.1 证明 X 与 Y 独立.

10. 习题:

(a) 设 (X,Y) 的期望方分别是 μ_1,σ_1^2 和 μ_2,σ_2^2, 且

$$((X-\mu_1)\sigma_1^{-1},(Y-\mu_2)\sigma_2^{-1})$$

服从相关系数为 r 的标准正态分布, 求 $\mathbb{E}(Y|X)$;

(b) 设 $X_1,\cdots,X_n$ 是独立且服从标准正态分布, 求 $\mathbb{E}(X_1|X_1+\cdots+X_n)$.

(c) 掷两个骰子, X 表示两个点数差的绝对值, Y 表示两个点数的和. 求 $\mathbb{E}(Y|X)$ 与 $\mathbb{E}(X|Y)$.

11. 习题: $\{S_n\}$ 是从零出发的对称随机游动, 求 $\mathbb{P}(S_2>0,\cdots,S_{2n}>0)$.

12. 习题: 在对称随机游动的情形, 设 T_k 是首次抵达 k 的时间, 用反射原理算 T_1 的分布律.

13. 习题: 设 ξ 的母函数是 $G(z)$, 证明:

$$\mathbb{P}(\xi<\infty)=\lim_{z\uparrow 1}G(z).$$

14. 习题: 对于一个不对称的随机游动 $\{S_n\}$, 用母函数方法算首中时 T_k 的分布律.

15. 习题: 在首次回归时间的问题中, 令 $u_{2n}:=\mathbb{P}(S_{2n}=0)$ 已知, 如果 $S_{2n}=0$, 那么 $T\le 2n$, 因此

$$\begin{aligned} u_{2n} &= \sum_{k=1}^{n}\mathbb{P}(S_{2n}=0|T=2k)\mathbb{P}(T=2k) \\ &= \sum_{k=1}^{n}\mathbb{P}(S_{2n-2k}=0)\mathbb{P}(T=2k) \\ &= \sum_{k=1}^{n}u_{2n-2k}\mathbb{P}(T=2k), \end{aligned}$$

利用这个关系可否算出 $\mathbb{P}(T=2k)$?

16. 习题: 设 $\{\xi_n\}$ 独立同分布且 $\mathbb{P}(\xi_n=1)=p$ and $\mathbb{P}(\xi=-1)=q$, $p+q=1$. 求实数 $a>0$ 使得 $X_n:=a^{\xi_1+\cdots+\xi_n}$ 是鞅.

17. 习题: 设 $\{X_n\}$ 是鞅, $\{\mathscr{F}_n\}$ 是鞅流.

(a) 如果 $\{Y_n : 0 \le n \le N\}$ 是关于鞅流的鞅, 那么 $Y_n = \mathbb{E}(Y_N|\mathscr{F}_n)$, $n \le N$.

(b) 如果 $Y \in \mathscr{F}_N$, 定义 $Y_n = \mathbb{E}(Y|\mathscr{F}_n)$, 那么 $\{Y_n : 1 \le n \le N\}$ 是关于鞅流的鞅.

(c) 如果 $\{Y_n : 0 \le n \le N\}$ 与 $\{Z_n : 0 \le n \le N\}$ 都是关于鞅流的鞅且 $Y_N = Z_N$, 那么对所有 $0 \le n < N$ 有 $Y_n = Z_n$.

18. 习题: 证明 Doob 分解定理中的唯一性.

19. 习题: 考虑上面的零点出发的简单随机游动 X_n 和首次回归时间 T, 证明: $Y_n = X_n^2 - n,\ n \ge 0$, 也是鞅, 然后证明 $\{Y_n\}$ 在停时 T 上没有期望不变性.

20. 习题: 有一个圆形城墙有 $m+1$ 个城门, 以 $0, 1, \cdots, m$ 依次编号. 某甲从 0 号门出发, 掷一个硬币, 视硬币正反面选择走向左右的城门, 然后继续掷硬币, 继续选择, 一直到他走遍所有城门后立刻走出, 问他从 i 号城门出来的概率是多少? $1 \le i \le m$.

21. 习题: 设在上面的赌徒输光问题中, 甲赢乙赢的概率不同, 各为 p 和 $q = 1-p$, 再用鞅方法计算甲输光的概率.

22. 习题: 设 $\{X_n\}$ 是简单随机游动, $H_n \in \sigma(X_1, \cdots, X_n)$ 且存在与 n 和 ω 无关的常数 C 使得 $|H_n(\omega)| \le C$, 再设 T 是停时且 $\mathbb{E}[T] < \infty$, Y_n 是 H 关于 X 的随机积分, 证明: Y_n 在 T 上也有期望不变性: $\mathbb{E}[Y_T] = \mathbb{E}[Y_0]$. 此题是对于倍增策略这个例子的一个补充说明. 提示: 应用控制收敛定理.

23. 习题: 用鞅方法来求零点出发的非对称 $(p \ne 1/2)$ 随机游动 $\{S_n\}$ 首次抵达 k 的时间 T_k 的分布以及首次返回零点的时间 T 的分布.

24. 习题: 设 ξ 是非负随机变量, 证明: $\mathbb{P}(\xi > 0) > 0$ 当且仅当 $\mathbb{E}\xi > 0$.

25. 习题: 设有一股票现在时刻 0 的价格是 50 元, 在每个时间段都以 3/4 的概率上涨 20%, 以 1/4 的概率下跌 20%, 假设利率可以忽略, 某个顾客欲在证券代理公司购买在时刻 2 到期的商定价格为 52 元的欧式看跌期权, 也就是说他有在时刻 2 以 52 元的价格卖给代理公司一股股票的权利, 问这份期权应该卖多少钱? 代理公司应该怎么投资来对冲风险, 如果时刻 1 股票下跌, 这份期权还值多少钱? 代理公司又应该怎么操作?

26. 习题: 假设上题中的存贷款的利率是 $r = 5\%$, 相应的问题应该怎么做?

27. 习题: 如果 (??) 中的 ξ_n 取值 $0 < u_1 < u_2 < u_3$ 的概率分别为 p_1, p_2, p_3, 其中 p_1, p_2, p_3 是正的且和为 1, 问存在概率测度使得折现后的价格过程 $\{(1+r)^{-n}S_n\}$ 成为鞅的条件是什么? 唯一吗?

28. 习题: 设 $X = (X_t)$ 是原点出发的 Brown 运动, (1) 对任何正数 t_1, t_2, t_3, t_4 求 $\mathbb{E}[X_{t_1}X_{t_2}X_{t_3}]$ 与
$$\mathbb{E}[X_{t_1}X_{t_2}X_{t_3}X_{t_4}];$$
(2) 设 $s < t$, 求 $\mathbb{E}[X_s|X_t]$; (3) 设 $s < t < u$, 求 $\mathbb{E}[X_t|X_s, X_u]$.

29. 习题: Einstein 发现 Brown 运动 X_t 的密度函数 $u(t,x) = p_t(x)$ 应该满足热传导方程
$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2},$$
证明之.

30. 习题: Brown 运动很直观, 但是在电脑上无法直接实现, 我们可以用 Donsker 不变原理来模拟 Brown 运动, 这正是它的强大之处. 写一个程序来模拟 Brown 运动.

31. 习题: 设 $X = (X_t)$ 是原地出发的 Brown 运动, 定义
$$Y_t(\omega) := \int_0^t X_s(\omega)\mathrm{d}s,\ t > 0, \omega \in \Omega.$$
这个积分是通常的 Riemann 积分, 证明: $Y = (Y_t)$ 是 Gauss 过程, 求它的协方差函数 $\mathrm{cov}(Y_s, Y_t)$.

32. 习题: 设 $a < 0 < b$, T_a, T_b 分别是原点出发的 Brown 运动首次碰到 a, b 的时间. 用鞅方法来求 $\mathbb{P}(T_a < T_b), \mathbb{E}[T]$ 以及 T 的 Laplace 变换, 其中 $T = T_a \wedge T_b$.

33. 习题: 前面看到 Brown 运动一定会碰到任何一条水平直线. 那么它会肯定碰到一条斜的直线 $x = at + b$ 吗? 显然当 a, b 异号时肯定, 同号时肯定吗? 似乎未必. 取 $a, b > 0$, 那么 $x = at + b$ 是一条在原点之上, 斜率正的直线, 令 T 是 Brown 运动 X 首次碰到这条直线的时间, 即
$$T = \inf\{t > 0 : X_t = at + b\}.$$

求 $\mathbb{P}(T < \infty)$. 我看过在一些股票走势的分析方法中, 有人用直线相交的方法来决定买入或者卖出, 似乎可能用到类似的结果. 我们在这里介绍一个技巧, 一些细节留给读者. 因为 $\exp(cX_t - c^2 t/2)$ 是鞅, 因此由 Doob 鞅基本定理, 对任何 $t > 0$ 有

$$\mathbb{E}\left(\exp(cX_{t\wedge T} - c^2(t\wedge T)/2)\right) = 1,$$

因为 $X_{t\wedge T} \le a(t\wedge T) + b$, 故只要 $ac - c^2/2 \le 0$ 且 $c > 0$ 时, t 可以趋于无穷, 这时

$$\mathbb{E}\left[\exp\left((ac - \frac{c^2}{2})T\right)\right] = \mathrm{e}^{-bc},$$

方程 $ac - c^2/2 = -\lambda$ 当 $\lambda > 0$ 时有正根 $c = a + \sqrt{a^2 + 2\lambda}$, 那么

$$\mathbb{E}\left(\mathrm{e}^{-\lambda T}\right) = \exp(-b(a + \sqrt{a^2 + 2\lambda})).$$

34. 习题: 设 $X_t = tB_t$, 求 X 的二次变差 $\langle X\rangle$. 证明: $X_t := tB_t - \int_0^t B_s \mathrm{d}s$ 是鞅.

35. 习题: 设 $B^{(1)}$ 与 $B^{(2)}$ 是两个独立的 Brown 运动, 证明: 它们的乘积是鞅, 从而证明 $\langle B^{(1)}, B^{(2)}\rangle = 0$.

36. 习题: 设 $0 \le t_1 < t_2 \le t_3 < t_4$, 随机变量 ξ, η 分别关于 $\mathscr{F}_{t_1}$ 与 $\mathscr{F}_{t_3}$ 可测, 证明:

$$\mathbb{E}[\xi(B_{t_2} - B_{t_1})\eta(B_{t_4} - B_{t_3})] = 0.$$

37. 习题: 证明.

   (a) (8.2) 中定义的 $M^{(i)}$ 有

$$\langle M^{(i)}\rangle_t = F^2_{t_{i-1}}(t\wedge t_i - t\wedge t_{i-1}),$$

   其中右边就是通常积分

$$\int_0^t F^2_{t_{i-1}} 1_{\{s\in(t_{i-1}, t_i]\}} \mathrm{d}s.$$

   (b) 对不同的 i, j 有

$$\langle M^{(i)}, M^{(j)}\rangle = 0.$$

38. 习题: 设 F, G 是两个连续适应过程, 证明

$$\langle F.B, G.B\rangle_t = \int_0^t F_s G_s \mathrm{d}s.$$

39. 习题: 设 M 是平方可积鞅, 如果 $M^2 = (M_t^2)$ 也是鞅, 证明: 对任何 t, $M_t = M_0$.

40. 习题: 设 M 是平方可积鞅, 如果对任何平方可积鞅 N 有 $\langle M, N\rangle = 0$, 证明: 对任何 t, $M_t = M_0$.

41. 习题: $X_t := B_t^3 - 3tB_t$, 证明: (X_t) 是鞅.

42. 习题: (Itô 的一个结果) 证明:

$$n!\int_{0\le u_1\le\cdots\le u_n\le t} \mathrm{d}B_{u_1}\cdots\mathrm{d}B_{u_n} = t^{\frac{n}{2}} h_n\left(\frac{B_t}{\sqrt{t}}\right),$$

其中 h_n 是所谓的 Hermite 多项式

$$h_n(x) = (-1)^n \mathrm{e}^{\frac{x^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(\mathrm{e}^{-\frac{x^2}{2}}\right).$$

43. 习题: 计算 Orstein-Uhlenbeck 过程的协方差

$$\mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)].$$

44. 习题: 证明 $X_t = |B_t|$ 是半鞅.

45. 习题: 最后, 请读者在不看任何书的情况下把期权定价公式推导出来.

参考文献

[1] A. Etheridge, A Course in Financial Calculus, Cambridge University Press, 2002 (有中文版)

[2] . M. Ross, Introduction to probability models, Academic Press, 1997, (有中文版)

[3] 应坚刚, 何萍, 概率论, 复旦大学出版社, 上海, 2005

[4] 应坚刚, 金蒙伟, 随机过程基础, 复旦大学出版社, 上海, 2005