应用随机分析

Applied Stochastic Calculus

(部分讲义 2019)

刘　勇

June 10, 2019

主要参考教材：

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

参考书目：

1. Øksendal, B., *Stochastic Differential Equations: An Introduction with Application.* 6th ed. Springer 2005
2. Klebaner, F. C., *Introduction to Stochastic Calculus with Applications.* 2nd ed.
   随机分析及其应用 人民邮电出版社 2008 (英文影印版)
3. Mikosch, T., *Elementary Stochastic Calculus, with Finance in View.* World Scientific 1998
   随机分析基础　世界图书进出口公司
4. Lawler, G. F., *Introduction to Stochastic Processes.* 2nd ed. Chapman & Hall/CRC 2006
   随机过程导论 张景肖 译 机械工业出版社 2010
5. Pavliotis G. A., *Stochastic Processes and Applications, Diffusion Processes, the Fokker-Planck and Langevin Equtions.* Springer 2014

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

Contents

1　第一章　引言和预备知识

在《概率论》和《应用随机过程》中主要是讲一些特殊的过程，例如：随机游动、Poisson 过程、Brown 运动、马氏链等的性质。那么，对于较为一般随时间演化的随机现象（即, 随机过程），我们将如何处理研究？

类比于在中学时我们学习一些初等函数及其性质，例如：二次函数、三角函数；在大学时我们是对较为一般的一些"好"函数（光滑）进行研究的，例如，利用积分、微分，进一步使用牛顿-莱布尼兹公式，从而我们能研究常微分方程 (ODE)、偏微分方程 (PDE) 等。那么这些方法的内在想法又是什么？

我们学过的《应用随机过程》就相当于中学学的若干初等函数和他们的性质，而《应用随机分析》这门课就相当于我们大学学的微积分。具体地说，首先，应该从数学上来定义什么是"随机性"？也就是需要用公理化，对一类现象来定义"随机性"，并且量化。那么，从数学上对一大类随机现象定义关于时间的随机性就是"鞅 (martingale)"。为了定义"鞅"，我们不得不借助严格公理化意义下的条件期望的概念来完成。为了介绍严格公理化定义下的条件期望的概念，我们必须一开始就要了解概率论的公理化是如何建立的（测度论）。

其次，有了"鞅"之后，一个自然的问题就是"鞅"是否可以用我们已知的一些"初等"、"简单"、"特殊"的过程表示出来，这就是随机积分，例如：关于 Brown 运动，Poisson 点过程的随机积分。有了随机积分之后，就要有关于随机积分的计算公式，Ito 公式。它相当于微积分中的牛顿-莱布尼兹公式。有了这些准备之后，那么我们就可以把一大类随机现象表示成一个"确定性部分"+"鞅 (随机部分)"，即，Ito 过程。Ito 过程中有一个特殊类型，Ito 随机微分方程 (Stochastic Differential Equation, SDE). 如同 ODE、PDE 一样，我们可以通过对 SDE 系数的一系列相关问题的研究来知道 SDE 的很多性质。而且，如同 ODE,PDE, 通过 SDE 我们也可以来描述更多的随机现象。

1.1　概率论的公理系统

a) 样本点：一个随机试验 (trial) 可能出现的一个结果，记为 ω;

b) 样本空间：全体样本点的集合，记为 Ω;

c) 事件 (event)：定义为样本点的集合，即，Ω 的一个子集。称某事件发生当且仅当它所包含的某个样本点出现。Ω 也是一个事件，必然事件；$\emptyset$ 也是一个事件，不可能事件。

d) 事件的运算

1) $A \subset B$　若 A 中的每一个样本点都包含在 B 中，称事件 B 包含了事件 A；或者说事件 A 发生蕴含事件 B 发生。
2) A、A^c　对立事件由所有不包含在 A 中的样本点所组成的事件称为 A 的对立事件，事件 A^c 表示事件 A 不发生；
3) $A \cap B$　AB　事件 A 和 B 同时发生
4) $A \cup B$　事件 A 和 B 至少有一个发生
5) $A \cap B = \emptyset$　事件 A 和 B 不同时发生

6) $A+B$ $(A\cap B=\emptyset,\ A\cup B)$

7) $A\setminus B$ 事件 A 发生但 B 不发生

以上可以看出事件的运算和集合论中集合的运算是一致的。事实上, 事件就是样本点的集合。

e) 概率：是度量一个事件发生可能性大小的量。因此，概率实际上就是在某些事件（Ω 的某些子集）上取值 [0,1] 的函数——以集合（事件）为自变量，取值 [0,1] 的函数。

因此，研究一个随机现象（试验），我们需要知道下面三个条件：

1. Ω: 所有样本点组成的集合

2. $\mathcal{F}$: 所有关心的事件全体，但不一定是全体子集

3. P: $\mathcal{F}\to[0,1]$

具体地，我们需要知道 $\mathcal{F}$ 的结构,P 的良定性等问题。P 应该满足

1. $P(A)\geq 0$；

2. $P(\Omega)=1$；

3. 对于 $\forall A_n\in\mathcal{F}$ $(n\geq 1)$，$A_n\cap A_m=\emptyset$ $(\forall n\neq m)$，则

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

$\mathcal{F}$ 被称为事件域，事件体，σ-域，σ-代数，从我们朴素的直观上看

1. 当我们观测到事件 A 发生，那么我们可以推知 A^c 不发生

2. 当我们观测到 A_1,A_2 之一发生，那么我们可以推知 $A_1\cup A_2$ 发生；$A_1,A_2,\cdots,A_n$ 之一发生，我们就可以推知 $A_1\cup A_2\cup\cdots\cup A_n$ 发生；$A_n,n\geq 1$ 之一发生，$\bigcup_{n\geq 1}A_n$ 发生（有限或可数无穷）

3. 至少应该关心一个事件 (哪怕是 $\emptyset$)，即，$\mathcal{F}$ 应该非空。

由此，我们知道 $\mathcal{F}$ 应该满足以下性质

1. $\mathcal{F}$ 非空；

2. 若 $A\in\mathcal{F}$，则 $A^c\in\mathcal{F}$；

3. 若 $A_n,n\geq 1$，$A_n\in\mathcal{F}$，则 $\bigcup_{n\geq 1}A_n\in\mathcal{F}$。

推论 1.1. $\Omega\in\mathcal{F}$.

称 $(\Omega,\mathcal{F},P)$ 为一个概率空间，

注记 1.1. P 的良定性将是《测度论》课程的主要内容之一。$(\Omega,\mathcal{F})$ 有时也被称为概率可测空间，可测空间。

注记 1.2. 为什么不取 Ω 的全体子集 $\mathcal{P}(\Omega)$ 作为事件域？

1. $\mathcal{P}(\Omega)$ 太大，有时包含了太多我们并不能"观测"事件或不可能"达到"的事件。

   例 1.1. $\Omega = \{1,2,3,4\}$

   图 1:　　　　　　　　　　图 2:

   (1) $\mathcal{F}_1 = \mathcal{P}(\Omega) = \{\Omega, \emptyset, \{1\}, \cdots\}$

   (2) $\mathcal{F}_2 = \{\Omega, \emptyset, \{1,2\}, \{3\}, \{4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}\}$。$\mathcal{F}_2$ 中永远不可能通过交并取余运算来达到 $\{1\}$，$\{2\}$。

2. 另一方面，在 $(\Omega, \mathcal{P}(\Omega))$ 并不一定可以给出一个基于基本观测恰当的概率测度。

   例 1.2. 回忆 Lebesgue 测度的建立，知道存在非 Lebesgue 可测的集合。

注记 1.3. 在 $(\Omega, \mathcal{F})$ 上可以建立不同的概率测度成为不同的概率空间。

1.2 随机变量

直接研究 $(\Omega, \mathcal{F}, P)$ 并不方便，试验结果往往用一个数字 ξ 来表示，即 ξ 是定义域为样本空间的一个函数。但 ξ 仅仅是一个函数是不行的。当我们考虑一个随机试验 (现象) 时，除了样本空间 Ω 外，还有我们关心的事件构成的事件域 $\mathcal{F}$。而 ξ 应该满足 $\forall x \in \mathbb{R}$，$\{\omega : \xi(\omega) \le x\} \in \mathcal{F}$，即 $\{\omega : \xi(\omega) \le x\} \in \mathcal{F}$ 应该是一个我们关心的事件。

在定义随机变量和随机向量之前，我们需要给出一些必要、有用的概念：

1. **示性函数**：集合 A 的示性函数定义为：

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

2. 在 $\mathbb{R}$ 中，包含所有开区间的最小事件域 (σ-代数、σ-域、事件体) 的集合 (也称为集族)，称为 Borel 事件域 (σ-代数、σ-域、事件体)，其中的每一个元素称为 Borel 集。在 $\mathbb{R}^d$ 中，包含所有开的 $d-$ 维开矩形的最小事件域 ($\sigma-$ 代数、σ-域、事件体) 的集合（也称为集族），称为 d-维 Borel 事件域 (σ-代数、σ-域、事件体)，记为 $\mathcal{B}(\mathbb{R}^d)$. 其中的每一个元素称为 $d-$ 维 Borel 集。

3. 包含所有连续函数，且对于 (取逐点) 极限运算封闭和线性运算封闭的最小函数类，称为 Borel 函数类，其中的函数称为 Borel 函数。

   注记 1.4. 根据实变函数的知识，f 是从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的一个函数，f 是 Borel 函数定义等价于对于任意的 $\mathbb{R}^m$ 上的 Borel 集合 A, $\{x \in \mathbb{R}^n : f(x) \in A\} \in \mathcal{B}(\mathbb{R}^n)$.

定义 1.1. **随机变量**：$\xi(\omega) : \Omega \to \mathbb{R}$ 的函数，满足 $\forall x \in \mathbb{R}$,

$$\{\omega : \xi(\omega) \le x\} \in \mathcal{F}.$$

因此，随机变量的定义依赖于事件域 $\mathcal{F}$。

例 1.3. （接例1.1）$\xi(i)=i,\ \ i=1,2,3,4.$ 易证 ξ 是 $(\Omega,\mathcal{F}_1)$ 的随机变量，但不是关于 $(\Omega,\mathcal{F}_2)$ 的随机变量，因为 $\{\omega:\xi(\omega)\le 1.5\}=\{1\}\notin\mathcal{F}$。

定义 1.2. 分布函数： $F(x)=P(\omega:\xi(\omega)\le x)$, 称为 ξ 的分布函数。

1. **离散型随机变量：** ξ 的概率分布列为

$$p(x)=P(\omega:\xi(\omega)=x),\quad (x=x_1,x_2,\cdots).$$

2. **连续型随机变量：** ξ 的分布函数 $F(x)$ 满足

$$F(x)=\int_{-\infty}^{x}p(t)dt,\ p(t)\text{被称为密度函数}.$$

3. **奇异型随机变量**:... ...

随机变量的一些数字特征：

1. **数学期望：**(加权平均) 设 g 为 Borel 可测函数

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

$$Eg(\xi(\omega))=\int_{\Omega}g(\xi(\omega))dP(\omega).$$

特别地，

离散型: $Eg(\xi(\omega))=\sum_i g(x_i)p(x_i)$。

连续性: $Eg(\xi(\omega))=\int_{-\infty}^{+\infty}g(x)p(x)dx$。

2. **方差**: $\mathrm{Var}\xi=\mathrm{E}(\xi-\mathrm{E}\xi)^2=\mathrm{E}\xi^2-(\mathrm{E}\xi)^2$；**标准差**:$\sqrt{\mathrm{Var}\xi}$

3. **矩母函数**: $M(z)=Ee^{z\xi}$. 若有限。

4. **特征函数：** $\phi(t)=Ee^{it\xi}$。$\phi(t)$ 与 ξ 的分布函数一一对应。

1.3　随机向量

设 $\xi=(\xi_1,\cdots,\xi_d)^T$ 为 $d-$ 维列随机向量，即 $\xi_i,\ \ i=1,\cdots,d$ 为随机变量。其中，T 表示转置.

定义 1.3. 分布函数： $\forall x=(x_1,\cdots,x_d)\in\mathbb{R}^d$,

$$F(x)=P(\xi\le x)=P(\xi_1\le x_1,\cdots,\xi_d\le x_d).$$

随机向量的一些数字特征：

1. **数学期望：** $E\xi=(E\xi_1,\cdots,E\xi_d)^T$。

2. **协方差矩阵**: $\Sigma_\xi = (\mathrm{cov}(\xi_\mathrm{i}, \xi_\mathrm{j}))_{1\le \mathrm{i},\mathrm{j}\le \mathrm{d}}$, 其中

$$\mathrm{cov}(\xi,\eta) = \mathrm{E}(\xi - \mathrm{E}\xi)(\eta - \mathrm{E}\eta) = \mathrm{E}(\xi\eta) - \mathrm{E}\xi\mathrm{E}\eta.$$

3. **相关系数**:

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

4. ξ **的矩母函数**: $M(z) = Ee^{\langle z,\xi\rangle}\quad z = (z_1,\cdots,z_d)$。

5. ξ 的**特征函数**: $\Phi(t) = Ee^{i\langle t,\xi\rangle}\quad t = (t_1,\cdots,t_d)$。

随机变量 $\xi_1,\cdots,\xi_d$ 相互独立，如果 $\forall x = (x_1,\cdots,x_d)\in\mathbb{R}^d$

$$P(\xi_1\le x_1,\cdots,\xi_d\le x_d) = P(\xi_1\le x_1)\cdots P(\xi_d\le x_d).$$

设 $\xi_1,\cdots,\xi_d$ 分别为 $n_1,\cdots,n_d$ 维随机向量。若对于任意的 n_i 维 Borel 集 $A_i,i=1,\cdots,d$,

$$P(\xi_1\in A_1,\cdots,\xi\in A_d) = P(\xi_1\in A_1)\cdots P(\xi_d\in A_d).$$

则称随机向量 $\xi_1,\cdots,\xi_d$ 相互独立。

1.4 随机过程

1.4.1 随机过程的一种构造方式

结合我们前面对概率论公理系统的了解，从《应用随机过程》的知识我们知道，一个随机过程是定义在一个概率空间 $(\Omega,\mathcal{F},P)$ 上的以 T 为指标集的随机变量族 $\{X_t,\ t\in T\}$(或记为 $\{X(t),\ t\in T\}$)。T 解释为时间。T 可以是有限集、非负整数集、非负实数、整数和实数。对于任意固定的 $t\in T$，X_t 的取值空间称为状态空间，记为 S，其中的元素称为状态。

一个自然的问题就是这样的概率空间是否存在？(参考思考题**??**)

Kolmogorov 相容性定理说这样的概率空间是存在的，而且告诉我们如何构造这个的一个 $(\Omega,\mathcal{F},P)$。事实上，可以把 $\{X_t,\ t\in T\}$ 写成 $\{X(t,\omega),\ t\in T,\omega\in\Omega\}$。可以把整个过程看成 $T\times\Omega$ 到 S 的函数 (映射)。当 $t\in T$ 固定时，$X(t,\cdot)$ 是一个 $(\Omega,\mathcal{F},P)$ 上的随机变量；当 $\omega\in\Omega$ 固定时，$X(\cdot,t)$ 是一个从 T 到 S 的函数 (映射)，称为对于 ω 的样本轨道。

例 1.4. 考虑一个先后掷 4 次硬币这个过程，正面记为 1，反面记为 0，那么整个过程的全体样本点是 $\omega = (\omega_1\omega_2\omega_3\omega_4)$, 其中 $\omega_i = 0$ 或 1, $i = 1,2,3,4$. 即，

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

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

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

或等价于 $\{1,2,3,4\}$ 到 $\{0,1\}$ 的映射全体。

一般地，对于一个状态空间为 S 的随机过程 $\{X_t,\ t\in T\}$，Ω 可以取成 $S^T=\{T\text{到}S\text{的映射}\}$(在此，我们略去 $\mathcal{F},P$ 的构造方式)。现在的一个问题就是仅有 $(\Omega,\mathcal{F},P)$ 对于我们研究随机过程就够了吗？

注记 1.5. 在一般的随机过程理论中,Ω 可以不是 S^T, 可以比 S^T 大, 也可以比 S^T 小。

1.4.2 信息、事件域和带流（滤子）的概率空间

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

考虑掷三次硬币这一随机过程，样本空间是

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

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

在掷硬币之前，我们来设想我们知道的事件是什么？信息是什么？我们仅知道，对于每一个样本点 ω, $\omega\in\Omega$，但 $\omega\notin\emptyset$。即此时 (掷之前) 我们能知道的事件域（信息）为 $\mathcal{F}_0=(\Omega,\emptyset)$。

当第一次掷完之后，虽然我们不能完全知道 ω，但我们可以知道 ω 的部分“信息”。例如，若我们知道了 $\omega_1=1$ ，则我们可以推断出什么样的事件（信息)?

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

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

若 $\omega_1=0$，类似地，

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

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

因此，当第一次掷完之后，对于每一个样本点 ω，我们总是能判断出 $\omega\in A$, $A\in\mathcal{F}_1=\{\Omega,\emptyset,A_1,A_0\}$ 是否成立。此时，我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}$。

当第一次，第二次掷之后，若 $\omega_1=1,\ \omega_2=0$, 则我们可以推断出的事件 (信息) 为

$$\omega\in A_{10}=\{100,101\},\quad \omega\notin A_{11}=\{110,111\},\quad \omega\in\Omega,$$

$$\omega\notin A_{00}=\{000,001\},\quad \omega\notin A_{01}=\{010,011\},\quad \omega\notin\emptyset,$$

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

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

同样，若 $\omega_1=1,\omega_2=1;\omega_1=0,\omega_2=0;\omega_1=0,\omega_2=1$ ，我们可以知道我们能得到的事件域为 $\mathcal{F}_2$。即，此时对于任意的样本点 ω，我们总能判断出对于任意的 $A\in\mathcal{F}_2$，$\omega\in A$ 是否成立。此时，我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}$。

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

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

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

$\mathcal{F}_0 = \{\Omega, \emptyset\}$; $\mathcal{F}_1 = \{\Omega, \emptyset, A_1, A_0\}$，其中，$A_1 = \{\omega|\omega_1 = 1\} = \{\omega|\omega = (1\omega_2\omega_3\cdots),\ \ \omega_i = 0\text{或}1,\ \ i \geq 2\}$, $A_0 = \{\omega|\omega_1 = 0\} = \{\omega|\omega = (0\omega_2\omega_3\cdots),\ \ \omega_i = 0,\text{或}1,\ \ i \geq 2\}$；$\mathcal{F}_2 = ...$；$\mathcal{F}_\infty$ 包含 $\cup_{n=1}^\infty \mathcal{F}_n$ 的最小事件域（但不是 $\mathcal{P}(\Omega)$）。我们可以取 $\mathcal{F} = \mathcal{F}_\infty$。因此，我们发现 $(\Omega, (\mathcal{F}_n)_{n\geq 1}, \mathcal{F}, P)$ 可以比 $(\Omega, \mathcal{F}, P)$ 更好地描述掷多次硬币这个随机试验 (随机过程)。

注记 1.6. 在实际研究中，有时 $\mathcal{F}$ 可以为 $\mathcal{F}_\infty$，也可以取的比 $\mathcal{F}_\infty$ 大。

因此研究随机过程时，我们通常要考虑一个带流的概率空间 $(\Omega, (\mathcal{F}_t)_{t\in T}, \mathcal{F}, P)$ (Filtered Probability Space)，满足 $t_1 < t_2$ 时，$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subset \mathcal{F}$。$(\mathcal{F}_t)_{t\in T}$ 被称为事件域流 ($\sigma-$ 代数流，$\sigma-$ 域流) 或滤子。

注记 1.7. 但为什么我们在学习《应用随机过程》时不提带流的概率空间？主要是因为当时考虑的问题简单，用的是所谓的“自然 σ- 代数流”。在研究复杂的随机现象时，考虑带流的概率空间就非常有用了。

1.5 方差有限的随机变量空间和 Gauss 系

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

1.5.1 $L^2(\Omega, \mathcal{F}, P)$

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

首先，回忆在《高等代数》或《线性代数》中欧氏空间 (Euclid 空间) 的概念。欧氏空间有两个基本的要点：

(1) 线性空间;

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

注记 1.8. 欧氏距离的定义本质上来源于勾股定理，内积和欧氏距离的关系是通过平行四边形法则相联系。

注记 1.9. 一般的实线性空间中抽象的内积定义是：$\langle \mathbf{x}, \mathbf{y}\rangle$ 是对称双线性型，且满足 (1) $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$; (2) $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ 当且仅当 $\mathbf{x} = 0$。

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

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

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

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

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

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

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

其实，这就是概率论中的两个随机变量的均方距离。

注记 1.10. 但与有限维欧氏空间不同，$L^2(\Omega,\mathcal{F},P)$ 一般找不到有限个元，使之构成一组基，(即 $L^2(\Omega,\mathcal{F},P)$ 中元素一般不可能用有限个“坐标”线性表示) 因此，$L^2(\Omega,\mathcal{F},P)$ 是一个无穷维欧氏空间。

注记 1.11. 在概率论中 $\mathrm{cov}(\xi,\eta)$ 是对随机变量 ξ 与 η 之间“线性相关性”的一种“度量”，为什么可以这么说？在有限维欧氏空间中，$\frac{\langle X,Y\rangle}{\|X\|\|Y\|}$ 表示两个向量 X 和 Y 之间夹角的余弦，可以理解为 X 与 Y 中心化之后线性相关性的一种度量。

（图 3）

而相关系数 $r=\dfrac{\mathrm{cov}(\xi,\eta)}{\sqrt{\mathrm{Var}\xi}\sqrt{\mathrm{Var}\eta}}$, $|r|\le 1$, 事实上就是两个向量 X 和 Y 之间夹角的余弦这一概念在无穷维欧氏空间中的推广。因此，可以用协方差 $\mathrm{cov}(\xi,\eta)$ 或相关系数 r 来表示两个随机变量之间“线性相关”的程度。

定义 1.4. 在 $L^2(\Omega,\mathcal{F},P)$ 中随机变量序列 $\{\xi_n\}$ 称为收敛到 $\xi\in L^2(\Omega,\mathcal{F},P)$，如果当 $n\to\infty$ 时，有 $\|\xi_n-\xi\|\to 0$，即 $E(\xi_n-\xi)^2\to 0$, 也就是随机变量序列的均方收敛。

定义 1.5. 在 $L^2(\Omega,\mathcal{F},P)$ 中的随机变量序列 $\{\xi_n\}$ 称为 Cauchy 序列，如果当 $n,m\to\infty$ 时，有 $\|\xi_n-\xi_m\|\to 0$。

性质 1.1. $L^2(\Omega,\mathcal{F},P)$ 中的任意 Cauchy 列必有极限，即如果 $L^2(\Omega)$ 中的随机变量序列满足 $n,m\to\infty$ 时，恒有 $\|\xi_n-\xi_m\|\to 0$，则一定存在 $L^2(\Omega)$ 中的元素 ξ，使得 $\|\xi_n-\xi\|\to 0$。因此，$L^2(\Omega)$ 是完备的，即 $L^2(\Omega)$ 是一个 Hilbert 空间。

注记 1.12. 这种理解对我们以后直观理解条件数学期望有巨大的帮助。

1.5.2 Gauss 分布、Gauss 系和 Gauss 过程

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

为什么说 Gauss 分布或者说服从 Gauss 分布的随机变 (向) 量非常重要呢？我们可以从下面的一些事实来理解：

(1) 回忆中心极限定理；

(2) 另一方面从统计物理和信息论的角度看，在给定方差且密度函数存在，$p(x) > 0,\ \forall x \in \mathbb{R}$ 的所有随机变量中，正态分布的熵是最 (小) 大。

定义 1.6. 设 $\eta_1, \cdots, \eta_m$ 为 m 个相互独立的标准正态分布，$(a_{ij})_{1\le i\le d, 1\le j\le m}$, μ_i, $(1 \le i \le d)$ 为常数，设

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

则称 $d-$ 维随机向量 $\xi = (\xi_1, \cdots, \xi_d)^T$ 服从 $d-$ 维 Gauss 分布，其矩阵形式为

$$\xi = A\eta + \mu \qquad A = (a_{ij})_{1\le i\le d, 1\le j\le m} \quad \mu = (\mu_1, \cdots, \mu_d)^T.$$

1. 其中，期望 $E\xi = \mu$，(μ 为一列向量)；协方差矩阵 $\Sigma = E(\xi-\mu)(\xi-\mu)^T = AA^T$（非负定线性算子）。

   注记 1.13. 设 μ 为 d 维列向量，Σ 为 d 维非负定矩阵，$\forall t = (t_1, \cdots, t_d)^T \in \mathbb{R}^d$，则以
   $$\phi(t) = \exp\{i\mu^T t - \frac{1}{2}t^T \Sigma t\}$$
   为特征函数的分布函数称为 d 维 Gauss 分布。

2. 若 Σ 可逆时，称为（多维、多元）正态分布 (Normal)
   $$p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$
   其中，$|\Sigma|$ 表示 Σ 的行列式，记 $\xi \sim N(\mu, \Sigma)$.

3. 服从 Gauss 分布的随机向量，称为 Gauss 随机向量。

注记 1.14. 对于 Gauss 分布和正态分布，我们有

(1) 常数（常向量）a 并不是正态分布，但是为方差是 0 的 Gauss 分布。

(2) 正态分布就是不退化的 Gauss 分布，只有此时才存在分布密度。一般的 Gauss 分布可能只分布在一个低维的超平面上。若 Σ 的秩为 r，$r < d$，则这时概率分布集中在一个 r 维子空间上，此时仍记 $\xi \sim N(\mu, \Sigma)$。

(3) 设 $\xi=(\xi_1,\cdots,\xi_d)^T$ 服从 $d-$ 维正态分布，则其任意的 r 维（$r<d$）边际分布还是 r 维正态分布，反之不然。反例：

$$\begin{aligned}\phi(x) &= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \qquad x\in\mathbb{R}\\ g(x) &= \begin{cases}\cos x & |x|<\pi\\ 0 & |x|\ge\pi\end{cases}\end{aligned}$$

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

命题 1.1. (1) ξ 服从 d 维 Gauss 分布 $N(\mu,\Sigma)$, C 为 $m\times d$ 维矩阵，$\mathbf{b}$ 为 m 维向量，则 $C\xi+\mathbf{b}\sim N(C\mu+\mathbf{b},C\Sigma C^T)$.

(2) ξ_1,ξ_2 是两个相互独立的 d-维随机变量，若 $\xi_1\sim N(\mu_1,\Sigma_1)$, $\xi_2\sim N(\mu_2,\Sigma_2)$，则

$$\xi_1+\xi_2\sim N(\mu_1+\mu_2,\Sigma_1+\Sigma_2).$$

(3) 一维 Gauss 随机变量 ξ, 或者是正态随机变量，或者为常数

$$Ee^{\xi}=\exp\{E\xi+\frac{1}{2}\mathrm{Var}\xi\}.$$

(4) Gauss 随机向量对依分布收敛有封闭性，即，若 $\xi_n\sim N(\mu_n,\Sigma_n)$, 且 ξ_n 依分布收敛到 ξ, ($\xi_n\xrightarrow{d}\xi$), 即指其相应的特征函数

$$\varphi_{\xi_n}(\lambda)=Ee^{i\lambda^T\xi_n}\to\varphi_\xi(\lambda)=Ee^{i\lambda^T\xi},$$

则存在 μ,Σ 使得 $\mu_n\to\mu,\ \ \Sigma_n\to\Sigma$ 且 $\xi\sim N(\mu,\Sigma)$. 进一步, 若 $\xi^{(n)}\xrightarrow{P}\xi$(指所有分量都依概率收敛), 则 $\xi^{(n)}\xrightarrow{d}\xi$, 且 ξ 服从 Gauss 分布.

命题 1.2. 下面 4 个叙述彼此等价：

(1) 随机向量 $\xi=(\xi_1,\cdots,\xi_d)^T$ 服从 d 维 Gauss 分布。

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

(3) ξ 的特征函数 $\varphi(\lambda)=Ee^{i\lambda^T\xi}=e^{i\lambda^T\mu-\frac{1}{2}\lambda^T\Sigma\lambda}$.

(4) ξ 的矩母函数 $M(z)=Ee^{z^T\xi}=e^{z^T\mu+\frac{1}{2}z^T\Sigma z}$.

定义 1.7. 随机变量族 $\{\xi_t,t\in I\}$ 称为 Gauss 系,如果对于任意 n 以及任意 $t_1,t_2,\cdots,t_n\in I$，随机向量 $(\xi_{t_1},\cdots,\xi_{t_n})$ 服从 Gauss 分布。如果 Gauss 系中的指标集 $I=[0,+\infty)$，则称为 Gauss 过程。

定义 1.8. 随机过程 $\{\xi_t,t\ge0\}$，当固定样本点 ω 时，$\xi_t(\omega)$ 为 t 的函数，称为对应于样本点 ω 的样本轨道。

若 $\{\xi_t,t\ge0\}$ 具有有限方差，即对于任意 $t\ge0$，恒有 $\mathrm{var}\xi_\mathrm{t}<+\infty$, 记

$$\mu(t)=E\xi_t, \qquad R(s,t)=E(\xi_s\xi_t)$$

$$C(s,t)=\mathrm{cov}(\xi_s,\xi_t)=R(s,t)-\mu(s)\mu(t)$$

分别称为 $\{\xi_t\}$ 的均值函数，相关函数和协方差函数。

定理 1.1. 随机过程 $\{\xi_t, t \geq 0\}$ 是 Gauss 过程，当且仅当在任意有限个时刻的任意线性组合的分布都是 Gauss 分布，即对于任意 n，任意实数 $a_1, a_2, \cdots, a_n$ 及任意时刻 $t_1, t_2, \cdots, t_n$，随机变量 $a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$ 或者是常数，或者服从正态分布。

对于 Gauss 过程，其均值函数和相关函数完全的确定了它的有限维分布族，即 Gauss 过程的概率特征完全由其均值函数和协方差函数所决定。

定义 1.9. 随机系 $\{\xi_t, t \in I\}$ 具有有限方差，那么 $L^2(\Omega, \mathcal{F}, P)$ 中的如下子集 $L^2(\xi) \equiv \{\{\xi_t, t \in I\}$中任意有限个元素的实线性组合$\}$，称为 $\{\xi_t, t \in I\}$ 的线性包。$\bar{L}(\xi)$ 为包含 $L^2(\xi)$，且对于均方极限封闭的最小集合，即，若 $\eta^{(n)} \in \bar{L}(\xi)$，$E|\eta^{(n)} - \eta|^2 \to 0$，则 $\eta \in \bar{L}(\xi)$，那么 $\bar{L}(\xi)$ 是 $\{\xi_t, t \in I\}$ 的线性闭包，称为随机系 $\{\xi_t, t \in I\}$ 的线性均方信息空间。

命题 1.3. （封闭性）如果 $\{\xi_t, t \in I\}$ 是 Gauss 系，则其均方线性信息空间 $\bar{L}(\xi)$ 也是 Gauss 系，即 $\bar{L}(\xi)$ 为 $L^2(\Omega, \mathcal{F}, P)$ 的一个闭子空间，$\bar{L}(\xi)$ 本身也是一个 Hilbert 空间.

命题 1.4. （独立性）

1. 若 $\{\xi_\alpha,\ \alpha \in I\}$ 与 $\{\eta_\beta,\ \beta \in J\}$ 独立，那么 $\bar{L}(\xi)$ 与 $\bar{L}(\eta)$ 独立。(指任意各自有限个元素组成的随机向量独立)。

2. 若 $\{\xi_\alpha,\ \eta_\beta,\ \alpha \in I,\ \beta \in J\}\}$ 是 Gauss 系，则 $\{\xi_\alpha,\ \alpha \in I\}$ 与 $\{\eta_\beta,\ \beta \in J\}$ 独立的充要条件为对于任意的 ξ_α 和 η_β 都有 $\mathrm{cov}(\xi_\alpha, \eta_\beta) = 0$。

3. 若 Gauss 系 $\{\xi_\alpha,\ \alpha \in I\}$ 与 Gauss 系 $\{\eta_\beta,\ \beta \in J\}$ 相互独立，那么 $\{\xi_\alpha,\ \eta_\beta,\ \alpha \in I,\ \beta \in J\}\}$ 是也 Gauss 系。

定理 1.2. 设 $\{\xi^{(n)}\}_{t\geq 0}$ 是一列 Gauss 过程，又对于任意的 t 都有

$$\lim_{n\to\infty} E|\xi_t^{(n)} - \xi_t|^2 = 0,$$

则 $\{\xi_t\}_{t\geq 0}$ 是 Gauss 过程。

命题 1.5. 设 $\Xi = \{\xi_t, t \in I\}$ 是 Gauss 系，又设 $\xi_{t_n} \in \Xi$, 且当 $n \to \infty$，ξ_{t_n} 依概率收敛到 ξ_0，则 $\lim_{n\to\infty} E|\xi_{t_n} - \xi_0|^2 = 0$。

2　第二章　条件数学期望

2.1　引言和条件数学期望的直观推导

我们先看两个例子。

例 2.1. 在一个简单的博弈中，某人根据前面各次博弈的结果决定下一次的博弈金额。设在开始时，有本金 ξ_0，又将第 n 次博弈后某人所有的本金记为 ξ_n。将这个简单博弈用一个数学模型表示，可以有以下两个要素：一，设

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

在此我们先假定 $\{\eta_n,\ n=1,2,3,\cdots\}$ 是独立同分布的随机变量序列。二，所谓“根据前面各次博弈的结果决定下一次的博弈金额”，指在第 n 次下注时，是根据前面 $n-1$ 次的输赢，也就是第 n 次的赌注是前面 $n-1$ 次输赢 (或本金) 的函数:

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

从第一章的知识，我们知道该函数应该是一个 Borel 函数，这样才能保证 b_n 或者 $\tilde{b}_n$ 是随机变量。那么，第 n 次博弈后某人所有的本金为

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

现在设想，有张三和李四两人各自给出了博弈“策略”:

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

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

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

表示我们“预计”能赢钱；如果

$$f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1}) - \xi_{n-1}^{\text{张三}} > f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1}) - \xi_{n-1}^{\text{李四}}, \tag{2}$$

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

但问题又出现了，每个人都可以给一个“预测方法 (函数)” $f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$，谁说了算呢? 当然是要找一个“最佳预测方法 (函数)”，在这个“最佳预测方法 (函数)”下，若 (2) 还满足的话，我们就可以大胆放心地用张三此时的博弈“策略”来下注了。

但问题再一次出现，什么是“最佳预测方法 (函数)”? 我们知道预测总是有误差的，“最佳预测方法 (函数)”应该是“误差最小”的。更为关键的是，如何选取一个合理、自然、易于计算的误差评判标准! 写成数学的语言就是，要找到某种“度量”或

“距离” $d(\cdot,\cdot)$，使得“最佳预测方法 (函数)” $f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$ 应该满足，对于任意一个“预测方法 (函数)” $g(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$,

$$d(\xi_n, f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})) \le d(\xi_n, g(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})). \tag{3}$$

因此，如果我们能有一个这样的误差标准 $d(\cdot,\cdot)$, 使得 $f(\xi_0,\eta_1,\eta_2,\cdots,\eta_{n-1})$ 满足 (3), 而且对于这个误差标准 $d(\cdot,\cdot)$ 意义下的“最佳预测方法 (函数)” f，如果我们还能得到不等式 (2)，那么我们就可以选用张三的博弈“策略”来下注了。

我们再来看一个概率论的例子。

例 2.2. 设随机向量 (X,Y) 表示人的身高和体重。我们经常遇到这样的问题，告诉一个人的身高，要求判断他 (或她) 的体重。即需要给出一个关于 X 的函数 $f(X)$，当然 $f(X)$ 与 Y 之间有误差。同样的问题是我们需要找到一个误差的评判标准，也就是说一种“度量”或“距离” $d(\cdot,\cdot)$，使得对于任意的 Borel 函数 g

$$d(f(X),Y) \le d(g(X),Y) \tag{4}$$

成立。

我们先来看一些具有启发性的类比。直线上两点 x 和 y 之间自然的“度量”、“距离” $d(x,y)=|x-y|$; 由勾股定理，我们知道平面 $\mathbb{R}^2$ 上两点 $x=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $y=\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ 之间自然距离是 $d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$；$n$-维欧氏空间的 $\mathbb{R}^n$ 上两点 $x=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ 之间自然距离是

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

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

因此，类比地对于两个随机变量 X,Y，若 $EX^2<\infty$，$EY^2<\infty$，则 X 和 Y 之间的一种自然的距离应该是

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

可以理解为每个“坐标 ω”之间距离平方“加权平均”再开方，实际上，这就是 $L^2(\Omega,\mathcal{F},P)$ 中的两个随机变量的距离。

这样，在 $L^2(\Omega,\mathcal{F},P)$ 范围内看，我们实际是确定了一种“自然”的误差估计方法—“均方误差”。

再回到上面的例 (2.1) 和例 (2.2)，对于例 (2.1) 中的不等式 (3) 和例 (2.2) 中的不等式 (4) 所表述的就是统计学中通常使用的“均方误差最小准则”。因此，在 $L^2(\Omega,\mathcal{F},P)$ 范围内看，对上面的例 (2.1)，我们实质上希望找到“预测函数” $f(\xi_0,\eta_1,\cdots,\eta_{n-1})$，对例 (2.2)，我们实质上希望找到“估计函数” $f(X)$，使得它们在“均方误差最小准则”下达到最小。

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

$$E\big(X-f(Y_1,Y_2,\cdots,Y_n)\big)^2\le E\big(X-g(Y_1,Y_2,\cdots,Y_n)\big)^2, \tag{5}$$

其中 g 为任一关于 $(Y_1,Y_2,\cdots,Y_n)$ 的函数，且使得 $Eg^2(Y_1,Y_2,\cdots,Y_n)<\infty$，或者表述为

$$E\big(X-f(Y_1,Y_2,\cdots,Y_n)\big)^2=\inf_{Eg^2(Y_1,Y_2,\cdots,Y_n)<\infty}E\big(X-g(Y_1,Y_2,\cdots.Y_n)\big)^2, \tag{6}$$

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

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

$$E\big(X-f(Y_t,t\in I)\big)^2=\inf_{Eg^2(Y_t,t\in I)<\infty}E\big(X-g(Y_t,t\in I)\big)^2. \tag{7}$$

但问题是"$g(Y_t,t\in I)$"该如何定义？数学上该如何表达？我们将在第三节中讲述这个问题。

首先，我们来分析概率空间 $(\Omega,\mathcal{F},P)$，随机向量 $(Y_1,Y_2,\cdots,Y_n,X)$ 和 $L^2(\Omega,\mathcal{F},P)$ 空间。考虑 $g(Y_1,Y_2,\cdots,Y_n)$ 且满足 $E(g(Y_1,Y_2,\cdots,Y_n))^2<\infty$ 的随机变量全体，将其记为 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$.

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

2) $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$ 是一个线性空间，即

$$f_i(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P),\quad i=1,2$$

   则 $\forall\alpha,\beta,\ \ \alpha f_1(Y_1,Y_2,\cdots,Y_n)+\beta f_2(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$；

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

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

因此，例 (2.1)、例 (2.2) 中的问题实质从几何的观点来看就是在 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$ 这个 $L^2(\Omega,\mathcal{F},P)$ 的子空间中找一个元素 $f(Y_1,Y_2,\cdots,Y_n)$, 使得 X 到 $f(Y_1,Y_2,\cdots,Y_n)$ 的距离是 X 到 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$ 中距离最短。

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

作为向量来看一个等价说法就是，$\vec{K}$ 应该满足 $\vec{X}-\vec{K}$ 与 (Y,Z) 平面垂直，即 $\vec{X}-\vec{K}\perp(Y,Z)$ 平面。在有限维欧氏空间，$\vec{X}-\vec{K}\perp(Y,Z)$ 平面的数学表达为

$$\langle\vec{X}-\vec{K},\vec{L}\rangle=0,\ \ 其中\vec{L}是(Y,Z)平面上的任意向量.$$

根据泛函分析知识，或者类比于有限维欧氏空间理论，对应于 $L^2(\Omega,\mathcal{F},P)$ 和 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$，应该有 $X-f(Y_1,Y_2,\cdots,Y_n)$ 垂直于 $L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$，也就是说对于任意的 $g(Y_1,Y_2,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$，

$$E\big((X-f(Y_1,Y_2,\cdots,Y_n))g(Y_1,Y_2,\cdots,Y_n)\big)=0$$

即 $f(Y_1,Y_2,\cdots,Y_n)$ 应该满足如下的方程，

$$E(Xg(Y_1,Y_2,\cdots,Y_n))=E(f(Y_1,Y_2,\cdots,Y_n)g(Y_1,Y_2,\cdots,Y_n)).$$

由此，我们可以定义

定义 2.1. $X,Y_1,\cdots,Y_n$ 为概率空间 $(\Omega,\mathcal{F},P)$ 上的随机变量且满足 $EX^2<\infty$，称由 $Y_1,\cdots,Y_n$ 决定的随机变量 $f(Y_1,\cdots,Y_n)$ 为 X 关于 $Y_1,\cdots,Y_n$ 的**条件数学期望**，若 $f(Y_1,\cdots,Y_n)$ 满足

1) $Ef^2(Y_1,\cdots,Y_n)<\infty$;

2) 任意的 $g(Y_1,\cdots,Y_n)\in L^2(\Omega,\sigma(Y_1,Y_2,\cdots,Y_n),P)$(也就是说 g 满足 $Eg^2(Y_1,\cdots,Y_n)<\infty$)，下式成立

$$E\big(Xg(Y_1,\cdots,Y_n)\big)=E\big(f(Y_1,\cdots,Y_n)g(Y_1,\cdots,Y_n)\big).\tag{8}$$

把 $f(Y_1,\cdots,Y_n)$ 记为 $E[X|Y_1,\cdots,Y_n]$。

注记 2.1. (1) 为什么叫条件数学期望？

(2) $g(Y_1,\cdots,Y_n)$ 可以看做是一个复合函数 $g\circ(Y_1,\cdots,Y_n)(\omega)$，$g:\mathbb{R}^n\to\mathbb{R}$ 的 Borel 函数。(图???)

(3) 不是“期望值”，而是一个随机变量。

回头看例 (2.1)，实质上我们是用第 n 次博弈后的本金 ξ_n 关于 $(\xi_0,\eta_1,\cdots,\eta_{n-1})$ 的条件数学期望 $f(\xi_0,\eta_1,\cdots,\eta_{n-1})(=E[\xi_n|\xi_0,\eta_1,\cdots,\eta_{n-1}])$ 来预测 ξ_n 的。对于例 (2.2)，实质是用体重 Y 关于身高 X 的条件数学期望 $f(X)(=E[Y|X])$ 来预测 Y 的。

首先，我们来看一个简单的例子。

例 2.3. 设 (X,Y) 为离散型随机变量，其分布列为 $P(X=x_i,Y=y_j)=p_{ij},\ \ i=1,\cdots,n,\ \ j=1,\cdots,m,$ 且假定 $p_{ij}>0$, 求 $E[X|Y]$，即求 $f(Y)$，使得对于任意的 $g(Y)$，满足

$$\begin{aligned} E(E[X|Y]g(Y)) &= E(Xg(Y)),\\ \text{即}\ \ E(f(Y)g(Y)) &= E(Xg(Y)).\end{aligned}$$

解： 首先，我们知道

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

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

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

特别地，取

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

代人上面的式子得

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

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

因此，解得

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

故而

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

□

例 2.4. 设 (X,Y) 服从密度函数为 $p(x,y)$ 的连续型分布函数，且假定 $p(x,y) > 0$, $(x,y) \in \mathbb{R}^2$, $EX^2 < \infty$，求 $E[X|Y]$.

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

$$
E(f(Y)g(Y)) = E(Xg(Y)).
$$

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

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

因为 $g(y)$ 的任意性，由实变函数知识得

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

即

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

即 $f(y)$ 为已知 $Y=y$ 时 X 的条件期望，即为 $E(X|Y=y)$。

有时 $E[X|Y]$ 记为 $E(X|Y=y)|_{y=Y}$，或者 $E[X|Y] = E(X|Y=y)|_{y=Y}$。 □

例 2.5. (X,Y) 服从二元正态分布，其分布密度如下

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

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

$(X,Y) \sim N(\vec{a},\ \Sigma)$, $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$,

求 $E[X|Y]$ 的密度函数。

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

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

这是正态分布 $N\left(a_1 + r\frac{\sigma_1}{\sigma_2}(y-a_2),\ \ \sigma_1^2(1-r^2)\right)$ 的密度函数，因此 $E(X|Y=y) = a_1 + r\frac{\sigma_1}{\sigma_2}(y-a_2)$, 即 $E[X|Y] = a_1 + r\frac{\sigma_1}{\sigma_2}(Y-a_2)$ 是 Y 的一个线性函数，则 $E[X|Y]$ 还是一个正态分布，其期望和方差分别是 a_1 和 $r^2\sigma_1^2$, 因此其密度函数为 $\frac{1}{\sqrt{2\pi}|r|\sigma_1}e^{-\frac{(x-a_1)^2}{2r^2\sigma_1^2}}$。 □

2.2 条件数学期望的定义和性质 (I)

2.2.1 定义与例子

定义 2.2. (接定义2.1) 设 $X, Y_1, \cdots, Y_n$ 为概率空间 $(\Omega, \mathcal{F}, \mathcal{P})$ 上的随机变量，满足 $E|X| < \infty$, f 为 $\mathbb{R}^n \to \mathbb{R}$ 的 Borel 函数，称 $f(Y_1, \cdots, Y_n)$ 为 X 关于 $(Y_1, \cdots, Y_n)$ 的**条件数学期望**，若

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

2) 对于任意的 n 维有界 Borel 函数 g, 下式成立

$$E\big(Xg(Y_1, \cdots, Y_n)\big) = E\big(f(Y_1, \cdots, Y_n)g(Y_1, \cdots, Y_n)\big). \tag{9}$$

把 $f(Y_1, \cdots, Y_n)$ 记为 $E[X|Y_1, \cdots, Y_n]$。

注记 2.2. 上式 (9) 等价于对于任意的 n 维 Borel 集 A，

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

例 2.6. (X, Y) 为离散型随机变量，其分布列为

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

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

解 1： 先求 $(h(X), Y)$ 的联合分布列，再利用例 (2.3) 的方法即可。

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

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

再求出,

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

因此

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

□

例 2.7. (X, Y) 服从密度函数为 $p(x, y)$ 的连续型随机向量，且 $p(x, y) > 0, \quad (x, y) \in \mathbb{R}^2$, $h(x)$ 为 Borel 函数，满足 $E|h(X)| < \infty$，求 $E[h(X)|Y]$。

我们知道 $(h(X), Y)$ 并不一定是连续型随机向量，类似于例 (2.4) 中先求 $(h(X), Y)$ 的联合密度的方法并不适用。由例 (2.6) 可以猜测

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

现在只需证明一下就行了。

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

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

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

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

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

□

$\frac{\int_{\mathbb{R}}h(z)p(z,y)dz}{\int_{\mathbb{R}}p(z,y)dz}=\frac{\int_{-\infty}^{x}p(z,y)dz}{\int_{\mathbb{R}}p(z,y)dz}$, 当 x 变动时就是条件分布函数。

例 2.8. 在初等概率论中，我们经常考虑已知事件 A 发生时事件 B 的条件概率 $P(B|A)$，那么如何用现在的条件数学期望的语言来表示？我们设

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

(X,Y) 的联合分布列为

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

则

$$\begin{aligned}E[X|Y](\omega) &= E(X|Y=1)\mathbf{1}_A(\omega)+E(X|Y=0)\mathbf{1}_{A^c}(\omega)\\ &= \frac{P(A\bigcap B)}{P(A)}\mathbf{1}_A(\omega)+\frac{P(A^c\bigcap B)}{P(A^c)}\mathbf{1}_{A^c}(\omega)\\ &= P(B|A)\mathbf{1}_A(\omega)+P(B|A^c)\mathbf{1}_{A^c}(\omega).\end{aligned}$$

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

注记 2.3. 在初等概率论中，我们考虑的条件概率、条件期望或条件分布，都是“静态”的，即**在给定** $(Y_1(\omega),\cdots,Y_n(\omega))$ **的某个具体取值** $(y_1,\cdots,y_2)$ **的条件下**，事件 B 的条件概率或随机变量或 X 的条件期望。而现在我们定义的条件数学期望是“动态”的，是一个“函数”，它给出了关于 $(Y_1(\omega),Y_2(\omega),\cdots,Y_n(\omega))$ 的每一个取值发生的条件下，事件 B 的条件期望或 X 的条件概率。也就是说在初等概率论中定义的条件期望是现在定义的条件数学期望的一个“具体的函数值”。

2.2.2　性质

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

性质 2.1. (全期望公式) $E\big(E[X|\mathbf{Y}]\big)=EX$。

证明： 在定义2.2的 (9) 式中取 $g(Y_1,\cdots,Y_n)\equiv 1$ 即得。□

我们来看这个公式的概率含义：

例 2.9. $X(\omega)=\mathbf{1}_A(\omega),\quad Y(\omega)\sim\begin{pmatrix} y_1 & \cdots & y_n & \cdots \\ p_1 & \cdots & p_n & \cdots \end{pmatrix}$, $p_i>0$。容易计算出

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

事实上，

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

即性质2.1就是全概率公式。

例 2.10. 设 (X,Y) 是连续型随机变量，其密度函数满足 $p(x,y)>0,(x,y)\in\mathbb{R}^2$，$p_X(x)=\int_{\mathbb{R}}p(x,y)dy$，$p_Y(y)=\int_{\mathbb{R}}p(x,y)dx$。

设 h 为 Borel 函数，$E(h(X)|Y=y)=\frac{\int_{\mathbb{R}}h(x)p(x,y)dx}{\int_{\mathbb{R}}p(x,y)dx}$，$E[h(X)|Y]=\frac{\int_{\mathbb{R}}h(x)p(x,Y)dx}{\int_{\mathbb{R}}p(x,Y)dx}$。则由性质2.1得到

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

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

例 2.11. 设随机变量 X 服从 $[0,1]$ 上的均匀分布，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为 $N(x,x^2)$。求 EY，$\mathrm{Var}(Y)$。

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

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

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

性质 2.2. 线性性质与单调性:

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

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

性质 2.3. 若 h 为 n 维 Borel 函数，且 $E|h(\mathbf{Y})|<\infty$，则 $E[h(\mathbf{Y})|\mathbf{Y}]=h(\mathbf{Y})$。

这个性质的证明是显然的，理解也是很有意思的。

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

(2) 从注记2.3的观点来看，当“静态地”给定 $\mathbf{Y}=\mathbf{y}$ 的条件下，$h(\mathbf{Y})$ 的期望 $E(h(\mathbf{Y})|\mathbf{Y}=\mathbf{y})$ 只能是 $h(\mathbf{y})$，因此，当让 $\mathbf{Y}$ 动起来时，即 $E[X|\mathbf{Y}]$ 就是 $h(\mathbf{Y})$ 了。

性质 2.4. $(X_1,\cdots,X_m)^T(\equiv\mathbf{X})$ 与 $(Y_1,\cdots,Y_n)^T$ 相互独立，且对于 m 维 Borel 函数 h，满足 $E|h(\mathbf{X})|<\infty$，则

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

注记 2.4. 事实上由性质2.3、性质2.4都有若 $X=c$，c 为常数，则 $E[X|\mathbf{Y}]=c$。

证明： 由定义2.2，只需验证式子 (9)。由独立性，并注意到 $Eh(\mathbf{X})$ 是常数，则对任意的有界 Borel 可测函数 $g(\mathbf{y})$ 得到

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

□

这条性质可以这样理解：因为 $\mathbf{X}$ 与 $\mathbf{Y}$ 独立，所以当 $\mathbf{Y}$ 已知时，$h(\mathbf{X})$ 的条件期望与 $\mathbf{Y}$ 没有任何关系，所以应该是一个常数，而且就是 $h(\mathbf{X})$ 的期望本身。

性质 2.5. h 为 n 维 Borel 函数，满足 $E|h(\mathbf{Y})|<\infty$, $E|X|<\infty$ 与 $E|h(\mathbf{Y})X|<\infty$，则

$$E[h(\mathbf{Y})X|\mathbf{Y}]=h(\mathbf{Y})E[X|\mathbf{Y}]. \tag{10}$$

进一步，由全期望公式得到

$$E\big(h(\mathbf{Y})X\big)=E\big(h(\mathbf{Y})E[X|\mathbf{Y}]\big). \tag{11}$$

注记 2.5. 若 h 和 X 满足性质2.5的条件时，利用测度论的知识可以严格证明该性质。其关键在于因为并不知道 $h(\mathbf{Y})E[X|\mathbf{Y}]$ 的可积性，所以证明时需要有一个取极限的过程，要用到抽象的测度空间上的 Lebesgue 控制收敛定理，这不在本讲义讨论的范围内。

这条性质可以这样地直观理解：当 $\mathbf{Y}=\mathbf{y}$ 已知时，那么 $h(\mathbf{Y})$ 应该是等于常数 $h(\mathbf{y})$，因此 $E(h(\mathbf{Y})X|\mathbf{Y}=\mathbf{y})$ 应该就是 $h(\mathbf{y})E(X|\mathbf{Y}=\mathbf{y})$。再让 $\mathbf{Y}$“动起来”时，就得到了式子 (10)。

例 2.12. (接例2.11) 求 cov(X,Y)。

解： 因为 $E(Y|X=x)=x$，所以 $E[Y|X]=X$。

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

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

□

性质 2.6. 记 $(X_1,\cdots,X_m)$ 为 $\mathbf{X}$，若 h 为 m 维 Borel 函数，且 $Eh^2(\mathbf{X})<\infty$，则

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

注记 2.6. 这从另一个角度说明定义2.1和定义2.2在 $L^2(\Omega)$ 的框架下是一致的。而且说明 $E\big(E[h(\mathbf{X})|\mathbf{Y}]\big)^2 < \infty$。

证明：

$$\begin{aligned}
& E\big(g(\mathbf{Y}) - h(\mathbf{X})\big)^2 \\
= \ & E\big((g(\mathbf{Y}) - E[h(\mathbf{X})|\mathbf{Y}] + E[h(\mathbf{X})|\mathbf{Y}] - h(\mathbf{X})\big)^2 \\
= \ & E\big(g(\mathbf{Y}) - E[h(\mathbf{X})|\mathbf{Y}]\big)^2 + E\big(E[h(\mathbf{X})|\mathbf{Y}] - h(\mathbf{X})\big)^2 \\
& +2E\Big(\big(g(\mathbf{Y}) - E[h(\mathbf{X})|\mathbf{Y}]\big)\big(E[h(\mathbf{X})|\mathbf{Y}] - h(\mathbf{X})\big)\Big) \\
= \ & I + II + 2III.
\end{aligned}$$

因此只需证明 $III = 0$ 即可。

$$\begin{aligned}
III &= E\Big(g(\mathbf{Y})h(\mathbf{X})\Big) - E\Big(E[h(\mathbf{X})|\mathbf{Y}]h(\mathbf{X})\Big) - E\Big(g(\mathbf{Y})E[h(\mathbf{X})|\mathbf{Y}]\Big) + E\Big(E[h(\mathbf{X})|\mathbf{Y}]\Big)^2 \\
&= III_1 - III_2 - III_3 + III_4 \\
III_1 &= E\Big(g(\mathbf{Y})h(\mathbf{X})\Big) \\
III_2 &= E\Big(h(\mathbf{X})E[h(\mathbf{X})|\mathbf{Y}]\Big) \\
III_3 &= E\Big(g(\mathbf{Y})h(\mathbf{X})\Big) \ \text{(利用性质2.5的式子(11))} \\
III_4 &= E\Big(E\Big[h(\mathbf{X})E[h(\mathbf{X})|\mathbf{Y}]\Big|\mathbf{Y}\Big]\Big) \ \text{(利用性质2.5的式子(11))} \\
&= E\Big(h(\mathbf{X})E[h(\mathbf{X})|\mathbf{Y}]\Big) \ \text{(利用全期望公式)}.
\end{aligned}$$

故 $III = 0$。命题成立。 □

性质 2.7. (平滑性、投影性) 若 $E|X| < \infty$，记 $(Y_1, \cdots, Y_n)$ 为 $\mathbf{Y}$, $(Z_1, \cdots, Z_k)$ 为 $\mathbf{Z}$，则

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

这条性质可以这样直观理解：将 $L^2(\Omega, \mathcal{F}, P)$ 想象成 3-维欧氏空间。向量 X 朝向量 Z 决定的方向投影相当于 X 先向 Y, Z 张成的平面投影，再向 Z 方向投影；或者对于 X 在 Z 方向的投影 $E[X|Z]$，这个投影在 Y, Z 张成的平面上的投影就是它自己。

证明： 由性质2.3，易得 $E\big[E[X|\mathbf{Z}]\big|\mathbf{Y}, \mathbf{Z}\big] = E[X|\mathbf{Z}]$。下面证明第一个等号成立，即只需验证 $E[X|\mathbf{Z}]$ 为 $E[X|\mathbf{Y}, \mathbf{Z}]$ 关于 $\mathbf{Z}$ 的条件数学期望。

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

(2) 对于任意的有界 Borel 函数 g，

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

□

性质 2.8. 记 $(X_1, \cdots, X_m)$ 为 $\mathbf{X}$,$(Y_1, \cdots, Y_n)$ 为 $\mathbf{Y}$。设 $E|h(\mathbf{X}, \mathbf{Y})| < \infty$,$E|h(\mathbf{X}, \mathbf{y})| < \infty$, 设 $E[h(\mathbf{X}, \mathbf{y})|\mathbf{Y}] = g(\mathbf{Y}, \mathbf{y})$, 则

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

有时也记为

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

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

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

有时也记为

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

例 2.13. 若随机变量 X 与 Y 相互独立，且分别服从参数为 λ 与 μ 的指数分布，定义随机变量

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

求 EZ.

解： 取 $h(x, y) = \mathbf{1}_{\{x \le y\}}$, $Z = h(X, Y)$. $p_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0,\ y \ge 0 \\ 0, & \text{其它} \end{cases}$, 则

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

因为 X 与 Y 独立, 由性质2.8,

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

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

□

例 2.14. (接例2.11，例2.12) 证明 $\frac{Y}{X}$ 与 X 独立。

证明： 对于任意的 Borel 有界函数 f 和 g，

$$\begin{aligned} E\Big(f\big(\frac{Y}{x}\big)\big|X = z\Big) &= \int_{\mathbb{R}} f\big(\frac{y}{x}\big) \frac{1}{\sqrt{2\pi} z} e^{-\frac{(y-z)^2}{2z^2}} dy, \\ E\Big[f\big(\frac{Y}{x}\big)\big|X\Big] &= \int_{\mathbb{R}} f\big(\frac{y}{x}\big) \frac{1}{\sqrt{2\pi} X} e^{-\frac{(y-X)^2}{2X^2}} dy, \\ E\Big[f\big(\frac{Y}{X}\big)\big|X\Big] &\overset{\text{性质2.8}}{=} E\Big[f\big(\frac{Y}{x}\big)\big|X\Big]\Big|_{x=X} \\ &= \int_{\mathbb{R}} f\big(\frac{y}{X}\big) \frac{1}{\sqrt{2\pi} X} e^{-\frac{(y-X)^2}{2X^2}} dy, \\ &\overset{\text{令}\tilde{y}=\frac{y}{X}}{=} \int_{\mathbb{R}} f(\tilde{y}) \frac{1}{\sqrt{2\pi}} e^{-\frac{(\tilde{y}-1)^2}{2}} d\tilde{y} = Ef(Z), \quad \text{其中} Z \sim N(1, 1). \end{aligned}$$

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

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

□

此命题证明是非平凡的，要用到测度论典型方法和正则条件概率等概念，我们这里略去它的严格证明，只用连续型随机变量验证一下。

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

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

$$\begin{aligned}E(h(X,y)|Y=z) &= \int h(x,y)p_{X|Y}(x|z)dx\\ &= \int h(x,y)\frac{p(x,z)}{\int_{\mathbb{R}}p(x,z)dx}dx.\end{aligned}$$

$$\begin{aligned}E[h(X,y)|Y] &= \int_{\mathbb{R}} h(x,y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)dx}dx\\ E[h(X,y)|Y]\Big|_{y=Y} &= \int_{\mathbb{R}} h(x,Y)\frac{p(x,Y)}{\int_{\mathbb{R}}p(x,Y)dx}dx \qquad (12)\end{aligned}$$

下面验证 (12) 式右端就是 $E[h(X,Y)|Y]$.

(1) 首先验证可积性。

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

(2) 对于任意有界的 $g(Y)$，

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

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

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

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

下面只需验证:

(1)

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

(2) 对于任意有界的 $g(Y)$，

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

例 2.15. (回到例2.1) 设 $P(\eta_n=1)=p,\ \ P(\eta_n=-1)=1-p=q$, 设 b_k 非负有界，

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

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

解:

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

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

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

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

(3) $k=n+1$ 时，由性质2.4、性质2.5

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

因此

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

当 $p=q=\frac{1}{2}$,

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

当 $p > q$,

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

当 $p < q$,

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

例 2.16. 如何用条件数学期望来定义马氏链 (过程)。

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

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

或者等价于对于任意的有界 Borel 函数 f，

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

那么现在我们改用条件数学期望来看上面的表达，其实就可以理解为

$$E[f(X_{n+1})|X_n, X_{n-1}, \cdots, X_1] = E[f(X_{n+1})|X_n] \tag{13}$$

因此可以重新给马氏链一个定义：称随机过程 $\{X_n\}_{n\geq 1}$ 是一个马氏链，如果对于任意的有界 Borel 函数 f，(13) 式成立。

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

2.3 条件数学期望的定义和性质 (II)

2.3.1 定义

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

因此，对于任意 Borel 函数 f，$f(X)$ 构成的随机变量全体就是对 X 描述的随机现象“全部信息的提取”。由实变函数知道 f 是一列简单 Borel 可测函数的极限，而简单 Borel 可测函数是 Borel 示性函数的线性组合 $\sum_{i=1}^n a_i \mathbf{1}_{A_i}$，$A_i \in \mathcal{B}(\mathbb{R})$，而且满足 $A_i \cap A_j = \emptyset$。故而，若承认线性运算和极限运算是我们可以“操作”的话，那么对于任意的 Borel 集 A，$\mathbf{1}_A(X)$ 的全体就是对 X 所描述的随机现象“全部信息的提取”。而 $\mathbf{1}_A(X)$ 是和事件 $\{\omega:\ X(\omega) \in A\}$ 等价的，因此，事件构成的集合 $\big\{\{\omega : X(\omega) \in A\},\ \ A \in \mathcal{B}(\mathbb{R})\big\}$ 就是 X 描述的随机现象的“全部信息的提取”；也就是说，当我们只关心 X 描述的随机现象时，我们只要完全清楚 $\big\{\{\omega : X(\omega) \in A\}, A \in \mathcal{B}(\mathbb{R})\big\}$ 就可以了。事实上，可以很容易证明 $\big\{\{\omega : X(\omega) \in A\}, A \in \mathcal{B}(\mathbb{R})\big\}$ 是一个 $\sigma-$ 代数，即事件域，**这个事件域称为由 X 生成的事件域 (σ-代数)，记为** $\sigma(X)$。易证 $\sigma(X) \subset \mathcal{F}$。

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

命题 2.1. 设 F 为 $(\Omega,\sigma(X),P)$ 上的一个随机变量，则存在一个 Borel 函数 f，使得 $F(\omega)=f(X(\omega))$。反之，若 f 为 $\mathbb{R}$ 上的 Borel 函数，则 $f(X)$ 为 $(\Omega,\sigma(X),P)$ 上的随机变量。

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

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

命题 2.2. 设 F 为 $(\Omega,\sigma(X_1,\cdots,X_n),P)$ 上的一个随机变量，则存在一个 n 维 Borel 函数 f，使得 $F(\omega)=f(X_1(\omega),\cdots X_n(\omega))$。反之，若 f 为 n 维 Borel 函数，则 $f(X_1,\cdots,X_n)$ 为 $(\Omega,\sigma(X_1,\cdots,X_n),P)$ 上的随机变量。

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

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

定义 2.3. 设 $\{X_t,\,t\in I\}$ 为 $(\Omega,\mathcal{F},P)$ 上的一族随机变量，称包含 $\big\{\{\omega:X_t\in(a_t,b_t)\},\forall t\in I,\forall a_t,b_t\in\mathbb{R}\big\}$ 的最小事件域 (σ-代数) 为 $\{X_t,\,t\in I\}$ 生成的事件域，记为 $\sigma(X_t,\,t\in I)$。

因此，类似于命题2.1和命题2.2，一个关于 $\{X_t,\ \ t\in I\}$ 的“可测函数”，实质上就是关于 $(\Omega,\sigma(X_t,t\in I))$ 的一个随机变量。

定义 2.4. 设 $\{Y_t,t\in I\}$ 为 $(\Omega,\mathcal{F},P)$ 上的一族随机变量，X 为 $(\Omega,\mathcal{F},P)$ 上的随机变量，且 $E|X|<\infty$，称 $(\Omega,\sigma(Y_t,t\in I),P))$ 上的随机变量 Z 是 X 关于 $\{Y_t,\,t\in I\}$ 的**条件数学期望**，若 Z 满足

(1) $E|Z|<\infty$;

(2) 对于任意有界的 $(\Omega, \sigma(Y_t, t\in I), P))$ 上的随机变量 F，下式成立

$$E(XF) = E(ZF). \tag{14}$$

记 $Z = E[X|Y_t,\ t\in I]$ 或 $Z = E[X|\sigma(Y_t,\ t\in I)]$ 。

注记 2.7. (2) 等价于对于任意的 $A\in\sigma(Y_t,\ t\in I)$, $E(X\mathbf{1}_A) = E(Z\mathbf{1}_A)$。

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

定义 2.5. 设 X 为 $(\Omega,\mathcal{F},P)$ 上的一个随机变量，满足 $E|X|<\infty$，$\mathcal{G}$ 为 $\mathcal{F}$ 的一个子事件域，称 Z 为 X 关于子事件域 $\mathcal{G}$ 的**条件数学期望**，若 Z 满足

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

(2) 对于任意的 $(\Omega,\mathcal{G},P)$ 上的有界随机变量 F，

$$E(XF) = E(ZF).$$

记 $Z = E[X|\mathcal{G}]$。

注记 2.8. 1) (2) 等价于对于任意 $A\in\mathcal{G}$，$E(X\mathbf{1}_A) = E(Z\mathbf{1}_A)$。

2) $E[X|\mathcal{G}]$ 的存在唯一性由 Radon-Nikodym 定理得到的。

3) X 为 $(\Omega,\mathcal{F})$ 随机变量，但 X 不一定为 $(\Omega,\mathcal{G})$ 的随机变量。

2.3.2 性质

性质 2.9. (全期望公式) $E\big(E[X|\mathcal{G}]\big) = EX.$

证明： 在定义2.4的 (14) 式中取 $F=\mathbf{1}_\Omega$ 即可。 □

例 2.17. X 为 $(\Omega,\mathcal{F},P)$ 上的随机变量，$A_1,A_2,\cdots,A_n,\cdots$ 为 Ω 的一个分割，且 $A_i\in\mathcal{F}$, $i=1,2,\cdots$，即 $A_n\cap A_m=\Phi$, $n\ne m$，且 $\sum\limits_{n=1}^{\infty}A_n=\Omega$。令 $\mathcal{G}$ 为包含 $A_1,A_2,\cdots,A_n,\cdots$ 的最小事件域，由此来验证性质 1.

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

由 3）取 $F(\omega)=\mathbf{1}_{A_k}(\omega)$, 则

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

则

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

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

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

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

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

性质 2.10. 线性性质与单调性：

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

(ii) 若 $X_1\ge X_2$, 则 $E[X_1|\mathcal{G}]\ge E[X_2|\mathcal{G}]$; 若 $X\ge 0$，则 $E[X|\mathcal{G}]\ge 0$；$\big|E[X|\mathcal{G}]\big|\le E[\big|X\big||\mathcal{G}]$。

性质 2.11. 若 X 为 $(\Omega,\mathcal{G})$ 上的随机变量，且 $E|X|<\infty$，则 $E[X|\mathcal{G}]=X$。

性质 2.12. 若 X 与 $\mathcal{G}$ 独立, 即对于任意的 Borel 集 A 和 $B\in\mathcal{G}$, $P\big(\{\omega:X\in A\}\cap B\big)=P\big(\{\omega:X\in A\}\big)P(B)$。若 $E|X|<\infty$，则 $E[X|\mathcal{G}]=EX$。

性质 2.13. Y 为 $(\Omega,\mathcal{G})$ 上的随机变量, 满足 $E|Y|<\infty$，$E|X|<\infty$，$E|YX|<\infty$，则 $E[YX|\mathcal{G}]=YE[X|\mathcal{G}]$。进一步 $E(YX)=E\big(YE[X|\mathcal{G}]\big)$。

性质 2.14. $X\in L^2(\Omega,\mathcal{F},P)$, $\mathcal{G}\subset\mathcal{F}$ 为子事件域，则

$$E\big(X-E[X|\mathcal{G}]\big)^2=\inf_{Y\in L^2(\Omega,\mathcal{G},P)}E(X-Y)^2.$$

性质 2.15. (平滑性、投影性) $E|X|<\infty$, $\mathcal{G}_1\subset\mathcal{G}_2\subset\mathcal{F}$,

$$E[X|\mathcal{G}_1]=E\Big[E[X|\mathcal{G}_2]\Big|\mathcal{G}_1\Big]=E\Big[E[X|\mathcal{G}_1]\Big|\mathcal{G}_2\Big].$$

性质 2.16. 设 $\mathbf{X},\mathbf{Y}$ 为 $(\Omega,\mathcal{F},P)$ 上的随机向量,对于 Borel 函数 $h(\mathbf{x},\mathbf{y})$ 满足 $E|h(\mathbf{X},\mathbf{Y})|<\infty$，$\mathcal{G}\subset\mathcal{F}$，若 $\mathbf{Y}$ 为 $(\Omega,\mathcal{G})$ 上的随机向量，则

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

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

性质 2.17. (Jensen 不等式) φ 为 $\mathbb{R}\to\mathbb{R}$ 上的凸函数 (φ 为有限凸)。若 X 及 $\varphi(X)$ 为随机变量，满足 $E|X|<\infty$，$E|\varphi(X)|<\infty$，则 $\varphi\big(E[X|\mathcal{G}]\big)\le E[\varphi(X)|\mathcal{G}]$。特别地 $\varphi(EX)\le E\varphi(X)$。

例 2.18.

A_2
A_1

A_{21}	A_{22}
A_{11}	A_{12}

$(\Omega,\mathcal{F},P)$, $\Omega=[0,1]\times[0,1]$, $\mathcal{F}=[0,1]^2$ 上的 Borel 域，$P=[0,1]^2$ 上的 Lebesgue 测度。$\mathcal{G}_1=(\Omega,\Phi,A_1,A_2)$，$\mathcal{G}_2=\sigma(\Omega,\Phi,A_{11},A_{12},A_{21},A_{22})$ 是包含 $A_{11},A_{12},A_{21},A_{22}$ 的最小 σ-代数，其中

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

f 为 $(\Omega,\mathcal{F},P)$ 一个随机变量，$E|f|<\infty$，求 $E[f|\mathcal{G}_1]$，$E[f|\mathcal{G}_2]$。

解:

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

□

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

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

在 A_1 上的平均值。

$\mathcal{G}_2 \supset \mathcal{G}_1$, 可以认为 $\mathcal{G}_2$ 包含的信息比 $\mathcal{G}_1$ 多，分辨率高，因此 $E[X|\mathcal{G}_2]$ 比 $E[X|\mathcal{G}_1]$ 复杂，$E[X|\mathcal{G}_1]$ 看起来比 $E[X|\mathcal{G}_2]$ 平滑些，所以性质2.7和性质2.15也叫平滑性质。

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

定义 2.6. (适应性、可知性) 称随机过程 $(X_t)_{t\in T}$ 关于 $(\mathcal{F}_t)_{t\in T}$ 是适应 (adapted) 的 (或可知的)，如果对于任意的 $t \in T$, X_t 是 $(\Omega, \mathcal{F}_t)$ 上的随机变量，即 $\forall a \in \mathbb{R}$,

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

注记 2.9. 在考虑复杂的随机现象时，往往可能有多个噪声源 $\big(\xi_t^{(1)}\big)_{t\in T}, \big(\xi_t^{(2)}\big)_{t\in T}, \cdots$，而观测到的随机过程 X_t 往往是噪声源的函数 $F_t = (\xi_s^{(1)}, \xi_s^{(2)}, s \le t)$(这点在随机微分方程中就很明显)，因此，虽然我们看到的是 X_t，但真正起作用的事件域流可能是 $\mathcal{F}_t = \sigma(\xi_s^{(1)}, \xi_s^{(2)}, s \le t)$。

定义 2.7. (马氏过程) 考虑一个带流的概率空间 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, P)$, $(X_t)_{t\in T}$ 是关于 $(\mathcal{F}_t)_{t\in T}$ 适应的随机过程，称 $(X_t)_{t\in T}$ 是关于 $(\mathcal{F}_t)_{t\in T}$ 的马氏过程，如果对于任意有界的 Borel 函数 f, 任意的 $t > s$,

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

注记 2.10. 称 $\mathcal{F}_t^X = \sigma(X_s, s \le t)$ 为 $(X_t)_{t\in T}$ 生成的自然 σ-代数流。

3　第三章　鞅论浅说

首先来解释一下“鞅 (martingale)”这个词。Martingale 是一个法语词。它有两个释义：(一) 古代的马拉车时安在马脖子上的皮套子 (马的璎珞)；(二) 公平赌博，原意是所谓倍赌策略。关于单词“martingale”的一些起源可以参看如下文献：

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

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

3.1　鞅的定义和例子

在上一章中我们考虑了如下博弈的例子.

$(\eta_n)_{n\geq 1}$ 为 i.i.d. 的随机变量序列，满足 $P(\eta_n=1)=p, P(\eta_n=-1)=1-p=q$，设

$$\xi_n=\xi_0+\sum_{k=1}^{n} b_k(\xi_0,\eta_1,\cdots,\eta_{k-1})\cdot\eta_k, \tag{15}$$

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

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

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

即在 n 时刻，预测 $n+1$ 时刻的“输赢情况”，不论是取什么样的博弈“策略”，都不可能得到任何关于“输赢情况”的信息。什么都没有预测到，也可以理解为，这个博弈的“不确定性”是“纯随机”。而这正是一个“公平的博弈 (fair game)”所要求的。因此有以下定义：

定义 3.1. (鞅) 设 $(X_t)_{t\in T}$ 为 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},P)$ 上的关于 $(\mathcal{F}_t)_{t\in T}$ 适应的随机过程(即 $\forall t\in T$, X_t 是关于 $(\Omega,\mathcal{F}_t)$ 可测的随机变量)，若对于 $\forall s<t$，$E[X_t|\mathcal{F}_s]=X_s$，则称 $(X_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅 (martingale)，或称 $(X_t,\mathcal{F}_t)_{t\in T}$ 为鞅，有时在明确了 $(\mathcal{F}_t)_{t\in T}$ 之后，可以直接称 $(X_t)_{t\in T}$ 为鞅。

若 $T=\{1,2,3,\cdots\}$ 时，$(X_t)_{t\in T}$ 称为鞅列。

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

例如上面的例中，取 $\mathcal{F}_0=\sigma(\xi_0)$, $\mathcal{F}_n=\sigma(\xi_0,\eta_1,\cdots,\eta_n)$。

1. 若 $p=q=\frac{1}{2}$ 时，$E[\xi_{n+1}|\mathcal{F}_n]=\xi_n$，则 $(\xi_n)_{n\geq 1}$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 鞅 (列)。

2. 若 $p>q$ 时，$E[\xi_{n+1}|\mathcal{F}_n]=\xi_n+b_n(\xi_0,\eta_1,\cdots,\eta_n)(p-q)\geq\xi_n$，则 $(\xi_n)_{n\geq 1}$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 下鞅 (列)。

3. 若 $p<q$ 时，$E[\xi_{n+1}|\mathcal{F}_n]\leq\xi_n$，则 $(\xi_n)_{n\geq 1}$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 上鞅 (列)。

下面来看更多鞅、鞅列的例子。

例 3.1. $\eta_n = a + \xi_1 + \cdots + \xi_n$，$\xi_1, \cdots, \xi_n, \cdots$ 相互独立，且 $\forall n \geq 1$, $E\xi_n = 0$，则 $(\eta_n)_{n\geq 1}$ 为 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)_{n\geq 1}$ 鞅列，也为 $\big(\mathcal{F}_n^1 = \sigma(\eta_1, \cdots, \eta_n)\big)_{n\geq 1}$ 鞅列，事实上 $\mathcal{F}_n = \mathcal{F}_n^1$。

例 3.2. 考虑一个带流的 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, P)$。若 ξ 满足 $E|\xi| < \infty$, 令 $\xi_t = E[\xi|\mathcal{F}_t]$, 则 $(\xi_t, \mathcal{F}_t)_{t\in T}$ 为鞅。

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

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

□

例 3.3. $(\xi_n)_{n\geq 1}$ 关于 $(\mathcal{F}_n)_{n\geq 1}$ 适应 (可知)，定义

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

则 $(\zeta_n, \mathcal{F}_n)_{n\geq 1}$ 为鞅。

分析：

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

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

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

证明：

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

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

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

□

例 3.4. 在例3.1中，$E\xi_i^2 = \sigma^2 < +\infty$,

$$\zeta_n = \eta_n^2 - \mathrm{var}(\eta_n) = \eta_n^2 - n\sigma^2,$$

则 $(\zeta_n)_{n\geq 1}$ 为 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)$ 鞅列。

证明：

(1) 显然 $(\eta_n^2 - n\sigma^2)_{n\geq 1}$ 为 $(\mathcal{F}_n)_{n\geq 1}$ 适应。

(2)

$$\begin{aligned} E[\eta_{n+1}^2|\mathcal{F}_n] &= E[(\eta_n + \xi_{n+1})^2|\mathcal{F}_n] \\ &= E[\eta_n^2|\mathcal{F}_n] + 2E[\eta_n\xi_{n+1}|\mathcal{F}_n] + E[\xi_{n+1}^2|\mathcal{F}_n] \\ &= \eta_n^2 + 2\eta_n E[\xi_{n+1}|\mathcal{F}_n] + E\xi_{n+1}^2 \\ &= \eta_n^2 + 2\eta_n E(\xi_{n+1}) + E\xi_{n+1}^2 = \eta_n^2 + \sigma^2 \end{aligned}$$

由此易证 $E[\eta_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n] = \eta_n^2 + \sigma^2 - (n+1)\sigma^2 = \eta_n^2 - n\sigma^2.$

□

例 3.5. (倍赌策略，martingale betting strategy)

在我们的基本例子 (15) 中设 $\xi_0 = 0$，$p = q = \frac{1}{2}$。赌徒的策略是一旦首次猜中获胜，不再下注，立即停止博弈，而在首次获胜之前每一次的赌注相应地在上一次赌注的基础上翻一倍，这个博弈策略具体的数学模型可以写成：

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

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

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

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

则 $(\xi_n)_{n\geq 1}$ 是鞅。

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

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

而且，

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

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

思考题：这种“最终只赢不输”的倍赌策略是不是和鞅所刻画的博弈的“公平性”相矛盾？

例 3.6. (随机利率) 先看确定性、离散时间：设银行的利率为 r ，第 n 天时，张三在银行中的存款为 X_n 时，在第 $n+1$ 天时存款就为 $(1+r)X_n$。

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

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

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

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

$$\frac{dX_t}{dt}=\sigma X_t, \quad X_t=X_0e^{\sigma t}.$$

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

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

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

将 n 时刻的资金 ξ_n 在开始时刻的折现价值记为 ζ_n, 即

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

则 $(\zeta_n,\mathcal{F}_n)$ 为鞅列。

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

现在假设张三、李四两人工资一样，消费也一样。张三把结余的钱 ξ_0 存入银行，李四把结余的钱进行某种投资。n 天之后，张三的资产为

$$\xi_n^{\text{张三}}=e^{(\sigma_0+\sigma_1+\cdots+\sigma_{n-1})}\xi_0,$$

李四的资产为 $\xi_n^{\text{李四}}$ 。从朴素的对于金钱的观念来看，若 $\xi_n^{\text{李四}}>\xi_n^{\text{张三}}$, 则在 n 时刻李四的投资是成功的。

另一种看法是将李四 n 时刻的资产 $\xi_n^{\text{李四}}$ 用银行利率折现成开始时刻的折现价值，即 $\tilde{\xi}_0^{\text{李四}}=e^{-(\sigma_0+\sigma_1+\cdots+\sigma_{n-1})}\xi_n^{\text{李四}}$ ；若 $\tilde{\xi}_0^{\text{李四}}>\xi_0$，那么我们也认为李四的投资增值了。

若李四的资金 $\xi_n^{\text{李四}}$ 满足

$$E[\xi_{n+1}^{\text{李四}}|\mathcal{F}_n]=e^{\sigma_n}\xi_n^{\text{李四}}.$$

而想象张三把钱都存在了银行，张三的存款应满足

$$\xi_{n+1}^{\text{张三}}=e^{\sigma_n}\xi_n^{\text{张三}}.$$

“平均来看”张三的存款与李四投资之间有某种“平等”性，或者说虽然市场上有“随机性”，但这种“随机性”对张三、李四来说是“公平的”，在数学上如何来体现这种公平性呢？**寻找鞅**。

把 $\xi_n^{\text{李四}}$ 按银行的储蓄折现成开始时刻的价值 $\zeta_n^{\text{李四}}$ 。若 $(\zeta_n^{\text{李四}},\mathcal{F}_n)$ 成为一个鞅，则可以认为李四在这个“有风险的市场上”没有赚，也没有赔。这个市场对于张三、李四来说是公平的。

例 3.7. 在例3.1中，可以不假设 $E\xi_n = 0$。令 $\zeta_n = \prod\limits_{k=1}^{n} \frac{e^{s\xi_k}}{Ee^{s\xi_k}}$，则 (ζ_n) 为 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)$ 的鞅列。

证明：

$$
\begin{aligned}
E[\zeta_{n+1}|\mathcal{F}_n] &= E\Big[\prod_{k=1}^{n+1} \frac{e^{s\xi_k}}{Ee^{s\xi_k}}\Big|\mathcal{F}_n\Big] \\
&= \frac{\prod\limits_{k=1}^{n} e^{s\xi_k}}{\prod\limits_{k=1}^{n+1} Ee^{s\xi_k}} E[e^{s\xi_{n+1}}|\mathcal{F}_n] \\
&= \zeta_n \frac{1}{E(e^{s\xi_{n+1}})} E(e^{s\xi_{n+1}}) = \zeta_n
\end{aligned}
$$

□

例 3.8. (分支过程) 第 n 代第 k 个细胞分裂的个数 $\eta_{n,k}$，对于任意的 n,k，$\eta_{n,k}$ 独立同分布。设 $E\eta_{n,k} = \mu < \infty$。

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

则 $\big(\frac{\xi_n}{\mu^n}\big)$ 是 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)$ 鞅。

证明：由马氏性

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

□

例 3.9. (Polya罐子模型，Polya's urn) 假设一个罐子中开始时有一个红球，一个绿球。每次随机的从罐子中摸出一个球，若是红球，则将此红球放回的同时再往罐子中放入一个红球；类似地，若摸出的是绿球，则将此绿球放回的同时再往罐子中放入一个绿球。设 X_n 表示第 n 次摸球后，罐子中红球的个数，$X_0 = 1$。可以证明 (X_n) 是一个马氏过程，其转移概率是

$$
P\big(X_{n+1} = k+1\big|X_n = k\big) = \frac{k}{n+2},\ P\big(X_{n+1} = k\big|X_n = k\big) = \frac{n+2-k}{n+2}. \tag{16}
$$

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

证明： 利用 (16) 容易计算，

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

再由马氏性，

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

□

例 3.10. (似然比序列) 设 (ξ_n) 为 i.i.d. 随机变量序列，其分布密度为 f，g 为另一个分布密度。设

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

则 (η_n) 为 $\big(\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)\big)$ 鞅列。

证明：

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

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

□

例 3.11. (Lévy 鞅) 设时齐 Markov 链 (ξ_n) 的转移矩阵为 $P = (p_{ij})$。对于状态空间上的有界函数 f，定义映射 $\mathbf{P}f$

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

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

$$X_n = f(\xi_n) - f(\xi_0) - \sum_{k=0}^{n-1} \big((\mathbf{P} - \mathbf{I})f\big)(\xi_k), \tag{17}$$

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

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

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

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

因此，

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

□

3.2 Doob 下鞅列分解定理

再次回到上一章中考虑的博弈的例子。此时，我们对模型略作些调整，$(\eta_n)_{n\geq 1}$ 为 i.i.d. 的随机变量序列, $P(\eta_n=1)=p,\quad P(\eta_n=-1)=1-p=q$,

$$
\xi_n = \xi_0 + \sum_{k=1}^{n} b_k(\xi_0,\xi_1,\cdots,\xi_{k-1})\cdot\eta_k. \tag{18}
$$

即，此时我们每一步的“投资策略” b_k 是依赖前面 $k-1$ 次的本金 $(\xi_0,\xi_1,\cdots,\xi_{k-1})$ 来做出的.

注记 3.1. 因为

$$
\xi_1 = \xi_0 + b_1(\xi_0)\eta_1,
$$

所以, ξ_1 为 $\sigma(\xi_0,\eta_1)$ 可测的随机变量，那么，对于第二步“投资策略” $b_2(\xi_0,\xi_1)$ 就应该也是 $\sigma(\xi_0,\eta_1)$ 可测的随机变量. 进一步

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

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

由上面一节基本例子 (15) 的推导，我们可以取 η_k 相互独立，且 $E\eta_k=\mu$ 来考虑更广的模型

$$
\xi_n = \xi_0 + \sum_{k=1}^{n} b_k(\xi_0,\xi_1,\cdots,\xi_{k-1})\cdot\eta_k, \tag{19}
$$

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

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

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

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

其中

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

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

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

定义 3.2. 对于 $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, P)$, $(A_n)_{n\geq 0}$ 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量序列，若 $\forall n \geq 1$，满足 A_n 是关于 $\mathcal{F}_{n-1}$ 可测的随机变量，则称 $(A_n)_{n\geq 1}$ 为可料随机序列 (predictable)。

定理 3.1. (Doob 下鞅分解) $(\xi_n)_{n\geq 0}$ 为 $(\mathcal{F}_n)_{n\geq 0}$ 下鞅列，则存在唯一 $(\mathcal{F}_n)_{n\geq 0}$ 鞅列，满足

(1) $M_0 = \xi_0$ (初值)；

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

证明： 参见例 (3.3) 的证明。 □

注记 3.2. 例 (3.3) 实际上是更广的 Doob 下鞅分解定理。它是说对于一般的期望有限的适应过程都可以分解成鞅与可料过程的和。

注记 3.3. Doob 分解在推广到连续时间时有本质的困难。例如“可料性”该如何推广？在连续时间时称为 Doob-Meyer 分解，它是现代鞅论的起点。

例 3.12. (ξ_n) 相互独立，$E|\xi_n| < \infty$，$E\xi_n = \mu_n > 0$，则 $X_n = \sum_{k=1}^{n} \xi_k$ 为 $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$ 下鞅，(X_n) 的 Doob 分解为

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

若进一步 $E\xi_n = \mu_n = 0$, $E\xi_n^2 < \infty$, $\forall n$, $X_n = \sum_{k=1}^{n} \xi_k$ 为鞅列，X_n^2 为下鞅列，则 X_n^2 的 Doob 分解为

$$X_n^2 = \Big(X_n^2 - \sum_{k=1}^{n} E\xi_k^2\Big) + \sum_{k=1}^{n} E\xi_k^2.$$

对于基本模型 (19)，下面介绍一些它的特例。

例 3.13. 设银行的利率为 R(离散时间)，即存入银行的资金增长满足

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

如果设

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

则这就是基本模型 (19) 非常特殊的一个例子。

例 3.14. Cox-Ross-Rubinstein 模型，二叉 (树) 模型 (binomial-tree model) **(缺图形)**

假定风险证券价格增长满足

$$S_{n+1} = (1+\eta_{n+1})S_n = S_n + \eta_{n+1}S_n, \tag{20}$$

其中 (η_n) 是 i.i.d 的随机变量序列，且 $1+\eta_{n+1} = \frac{S_{n+1}}{S_n} \sim \begin{pmatrix} a & b \\ p & 1-p \end{pmatrix}$，$0<a<b$，$0<p<1$，$(\eta_n)_{n\ge 1}$ 与 S_0 独立。

如果设 $\xi_n = S_n$，$b_n(S_0, S_1, \cdots, S_{n-1}) = S_{n-1}$，则这个模型也是基本模型 (19) 的一个特例。

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

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

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

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

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

Cox-Ross-Rubinstein 模型 (20) 的 Doob 分解为

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

或者写成

$$\frac{S_{n+1}-S_n}{S_n} = \mu + \sigma\frac{\eta_{n+1}-\mu}{\sigma}. \tag{21}$$

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

例 3.15. (Black-Scholes 模型的直观推导)

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

我们用黎曼积分的想法来考察 0 到 t 时间段内 $(S_u)_{0\le u\le t}$ 的变化。首先对 $[0,t]$ 做划分 $0=t_0<t_1<\cdots<t_n=t$. 参考上例中的 (21) 式，在时刻 t_{k-1} 单位时间单位价格的收益应该近似地满足下式

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

类比 (21) 中 (η_n) 满足的统计性质，可以假设 (ξ_t) 应该满足如下的条件：

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

2. $\xi_s-\xi_u$ 的分布只与 $s-u$ 有关。

即，(ξ_t) 是一个独立增量过程，而且增量平稳。我们还可以进一步假定其二阶矩有限，因此 $E(\xi_t-\xi_0)=\mu t, \mathrm{var}(\xi_t-\xi_0)=\sigma^2 t$。近似地可以得到下式

$$S_t-S_0\approx\mu\sum_{k=1}^{n}S_{t_{k-1}}(t_k-t_{k-1})+\sigma\sum_{k=1}^{n}S_{t_{k-1}}\frac{\xi_{t_k}-\xi_{t_{k-1}}-\mu(t_k-t_{k-1})}{\sigma}. \qquad (22)$$

当 $\max_k(t_k-t_{k-1})\to 0$ 时，(22) 式的第一项在黎曼积分的意义下收敛到 $\mu\int_0^t S_s ds$，而第二项可以设想按某种意义收敛到“积分” $\int_0^t S_s$“$d\bar{\xi}_s$”。形式地，我们可以认为 $(S_t)_{t\geq 0}$ 满足下面的方程

$$S_t-S_0=\mu\int_0^t S_s ds+\sigma\int_0^t S_s\text{``}d\bar{\xi}_s\text{''}. \qquad (23)$$

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

从 Doob 下鞅分解定理的角度来看，在 (22) 式直观推导收敛到 (23) 式的过程中，如何保持第二项的“鞅性”不消失，如何在连续时间参数下理解第一项中的“可料性”？这将是第五章以及《随机分析》课程中的重要内容。

例 3.16. 我们将例3.13和例3.14综合起来考虑。风险证券在 n 时刻的价格为 S_n，但并不代表其实际的“价值”。直观地看，1 元钱开始存在银行中，在 n 时刻就增值到 $(1+R)^n$ 元，同样用 1 元钱在开始时刻买了风险证券，在 n 时刻它的价格是 S_n，但其与存在银行中的那一元钱比，也就是说，如果我们用 $\frac{1}{(1+R)^n}$ 作为 n 时刻的计价单位的话，S_n 的“价值”就是 $\tilde{S}_n=\frac{S_n}{(1+R)^n}$。用金融数学的术语说，$\tilde{S}_n$ 为该证券在初始时刻的折现价格。

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

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

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

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

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

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

注记 3.4. 通过上面这个简单的例子 (3.16) 希望给大家些许直观，鞅是如何刻画一些金融数学中的观念和性质的。事实上现代金融数学理论中建立模型的起点是规定不存在套利的机会，而无套利的数学描述依赖于“鞅”或“鞅测度”。

3.3 测度变换与 Girsanov 定理

先看一个简单的博弈例子，掷不均匀的硬币，每次下注 1 元钱，即 $(\eta_n)_{n\geq 1}$ 是 i.i.d 的随机变量序列满足：

$$P(\eta_n=1)=p,\ P(\eta_n=-1)=1-p,\quad 0<p\neq\frac{1}{2}<1.$$

设 $\mathcal{F}_n=\sigma(\eta_1,\cdots,\eta_n)$，$\xi_n=\sum_{k=1}^n\eta_k$。$(\xi_n,\mathcal{F}_n)_{n\geq 1}$ 不是鞅，这个博弈不是公平的。那么如何使得 $(\xi_n,\mathcal{F}_n)_{n\geq 1}$ 变得公平呢？换成均匀的硬币即可。从数学的角度来看就是给"输""赢"重新赋予一个权重 $\hat{P}$，

$$\hat{P}(\eta_n=1)=\frac{1}{2},\ \hat{P}(\eta_n=-1)=\frac{1}{2}.$$

也可以理解为做了一个测度变换

$$\hat{P}(\eta_n=1)=\frac{1}{2p}P(\eta_n=1),\ \hat{P}(\eta_n=-1)=\frac{1}{2(1-p)}P(\eta_n=-1).$$

如果从随机过程的角度来看上面的问题，就是对过程的轨道或称为路径 (path) 重新赋予一个权重。直观上看，$p>\frac{1}{2}$ 时，图形1中向上走比朝下走次数多的轨道的概率更大些。如果 $p=\frac{1}{2}$ 在 n 之前每条轨道应该是等可能的。

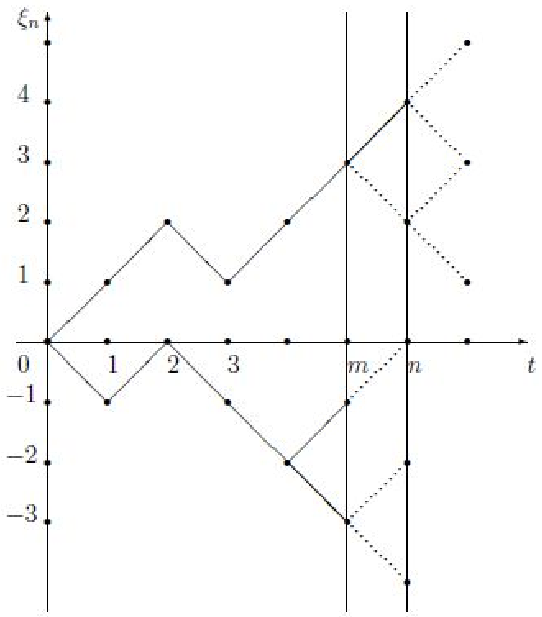

Figure 1: 随机游动轨道

由简单随机游动的知识，我们知道当 $n\geq k$,

$$P(\xi_n=k)=\begin{cases} C_n^{\frac{n+k}{2}}p^{\frac{n+k}{2}}(1-p)^{\frac{n-k}{2}} & n,k\text{同奇偶}\\ 0 & n,k\text{不同奇偶}\end{cases}.$$

对于所有在 $1,2,\cdots,n$ 时刻，$\xi_1(\omega),\cdots,\xi_n(\omega)$ 取值都相同，但 $\xi_{n+1}(\omega),\xi_{n+2}(\omega),\cdots$ 可以不同的轨道组成的集合 (这样的集合有 2^n 个，记为 $A_i,i=1,\cdots,2^n$)，其上的概率为

$$p^{\frac{n+\xi_n}{2}}(1-p)^{\frac{n-\xi_n}{2}}=[4p(1-p)]^{\frac{n}{2}}\Big(\frac{p}{1-p}\Big)^{\frac{\xi_n}{2}}\big(\frac{1}{2}\big)^n.$$

记 $M_n = [4p(1-p)]^{-\frac{n}{2}}\left(\frac{1-p}{p}\right)^{\frac{\xi_n}{2}}$。给每个 A_i 这样的集合赋予一个新的概率 $\hat{P}$:

$$\left(\frac{1}{2}\right)^n = M_n p^{\frac{n+\xi_n}{2}}(1-p)^{\frac{n-\xi_n}{2}},$$

则在这个新的概率 $\hat{P}$ 下，$\eta_1, \cdots, \eta_n$ 是 i.i.d 的随机变量，而且 $\hat{P}(\eta_n = 1) = \hat{P}(\eta_n = -1) = \frac{1}{2}$。容易知道，此时 $(\xi_n, \mathcal{F}_n)$ 在 $\hat{P}$ 下是鞅。

对于 $A \in \mathcal{F}_N$，A 是一些 A_i 的并，因此

$$\hat{P}(A) = E(\mathbf{1}_A M_n).$$

一个自然的问题就是对于 $m < n$，$A \in \mathcal{F}_m \subset \mathcal{F}_n$，那么此时 $\hat{P}(A)$ 就可以有两个定义 $E(\mathbf{1}_A M_m)$ 和 $E(\mathbf{1}_A M_n)$，这两个定义是否一致？也就是说 $\hat{P}$ 的定义是否适定？

由习题知道 $(M_n, \mathcal{F}_n)$ 在概率 P 下是鞅，则

$$\begin{aligned} E(\mathbf{1}_A M_n) &= E\big(E[\mathbf{1}_A M_n | \mathcal{F}_m]\big) \\ &= E\big(\mathbf{1}_A E[M_n | \mathcal{F}_m]\big) \\ &= E(\mathbf{1}_A M_m), \end{aligned}$$

所以，这个新概率的定义 $\hat{P}$ 是适定的。

将上述想法推广得到下面的定理。为了避免一些数学技术上的难度，这里假设所有的 b_k, c_k 都是有界 Borel 可测函数。

定理 3.2. 设 $(\xi_0, \eta_n,\ n \geq 1)$ 相互独立且 $E|\xi_0| < \infty$, $E|\eta_n| < \infty$，对于任意的 k，存在 s_k 使得 $E(\eta_k e^{s_k \eta_k}) = 0$。设

$$z_n = \prod_{k=1}^{n} \frac{e^{s_k \eta_k}}{E e^{s_k \eta_k}}, \quad \mathcal{F}_n = \sigma(\xi_0, \eta_1, \cdots, \eta_n),$$

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

对于任意的 $A \in \mathcal{F}_N$，定义新的概率 $\hat{P}(A) = E(\mathbf{1}_A z_N)$，则 $(\xi_n, \mathcal{F}_n)_{1 \leq n \leq N}$ 在 $\hat{P}$ 下是鞅。

引理 3.1. $(z_n, \mathcal{F}_n)_{n \geq 1}$ 在 P 下是鞅，而且 $E z_n = 1$。

证明： 可以参考例 (3.7). □

定理证明： 首先上面的引理说明 $\hat{P}$ 的定义是适定的, 再由例 (3.1) 知道只需要证明 $(\eta_1, \eta_2, \cdots, \eta_n, \cdots)$ 在 $\hat{P}$ 下是相互独立的，且 $\hat{E}\eta_n = 0, n = 1, 2, \cdots$. 对于任意的

$A_n \in \mathcal{B}(\mathbb{R}), n = 1, 2, \cdots$.

$$\begin{aligned}
\hat{P}(\eta_1 \in A_1, \cdots, \eta_n \in A_n) &= E\big(\mathbf{1}_{A_1}(\eta_1)\cdots\mathbf{1}_{A_n}(\eta_n)\prod_{k=1}^n \frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\big) \\
&= \prod_{k=1}^n E\big(\mathbf{1}_{A_k}(\eta_k)\frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\big) \\
&= E\big(\mathbf{1}_{A_1}(\eta_1)\frac{e^{s_1\eta_1}}{Ee^{s_1\eta_1}}\big)\prod_{k=2}^n E\big(\mathbf{1}_{A_k}(\eta_k)\frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\big) \\
&= \hat{P}(\eta_1 \in A_1)\prod_{k=2}^n \Big(\prod_{j=1}^k E\big(\frac{e^{s_j\eta_j}}{Ee^{s_j\eta_j}}\big)E\big(\mathbf{1}_{A_k}(\eta_k)\frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\Big) \\
&= \hat{P}(\eta_1 \in A_1)\prod_{k=2}^n \Big(E\big(\mathbf{1}_{A_k}(\eta_k)\prod_{j=1}^k \frac{e^{s_j\eta_j}}{Ee^{s_j\eta_j}}\Big) \\
&= \hat{P}(\eta_1 \in A_1)\prod_{k=2}^n \Big(E\big(\mathbf{1}_{A_k}(\eta_k)z_k\big)\Big) \\
&= \prod_{k=1}^n \hat{P}(\eta_k \in A_k).
\end{aligned}$$

进一步,

$$\hat{E}\eta_n = E(\eta_n z_n) = E\big(\eta_n \prod_{k=1}^n \frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\big) = \Big(\prod_{k=1}^{n-1} E\big(\frac{e^{s_k\eta_k}}{Ee^{s_k\eta_k}}\big)\Big)\frac{E\eta_n e^{s_n\eta_n}}{Ee^{s_n\eta_n}} = 0.$$

□

特别地，对于服从正态分布的随机变量，还有更强的结论。

定理 3.3. (Girsanov-Cameron-Martin 定理)

设 $(\xi_0, \eta_n,\ n \geq 1)$ 相互独立，$\mathcal{F}_n = \sigma(\xi_0, \eta_1, \cdots, \eta_n)$，$E|\xi_0| < \infty$，$\eta_n \sim N(0, \sigma_n^2)$，$b_n \neq 0$，$\frac{c_n}{b_n}$ 是有界 Borel 可测函数。

$$\begin{aligned}
\xi_n &= \xi_0 + \sum_{k=1}^n b_k(\xi_0, \ldots, \xi_{k-1})\cdot\eta_k + \sum_{k=1}^n c_k(\xi_0, \xi_1, \cdots, \xi_{k-1}) \\
&= \xi_0 + \sum_{k=1}^n b_k(\xi_0, \ldots, \xi_{k-1})\Big[\eta_k + \frac{c_k(\xi_0, \xi_1, \cdots, \xi_{k-1})}{b_k(\xi_0, \xi_1 \ldots, \xi_{k-1})}\Big].
\end{aligned}$$

$$z_n = \exp\Big(-\sum_{k=1}^n \frac{c_k(\xi_0, \xi_1, \cdots, \xi_{k-1})}{b_k(\xi_0, \xi_1, \cdots, \xi_{k-1})\sigma_k^2}\eta_k - \sum_{k=1}^n \frac{1}{2}\big(\frac{c_k(\xi_0, \xi_1, \cdots, \xi_{k-1})}{b_k(\xi_0, \xi_1, \cdots, \xi_{k-1})\sigma_k}\big)^2\Big).$$

对于任意的 $A \in \mathcal{F}_N$，定义新的概率 $\hat{P}(A) = E(\mathbf{1}_A z_N)$，则在 $\hat{P}$ 下，

$$(\eta_1 + \frac{c_1}{b_1}, \cdots, \eta_N + \frac{c_N}{b_N}.)$$

是相互独立的随机变量序列，而且 $\eta_n + \frac{c_n}{b_n} \sim N(0, \sigma_n^2)$。更进一步，$(\xi_n, \mathcal{F}_n)_{1\leq n\leq N}$ 在 $\hat{P}$ 下是鞅。

引理 3.2. 设 $(\xi_0, \eta_n,\ n \geq 1)$ 相互独立，$E|\xi_0| < \infty$，$\eta_n \sim N(\mu_n, \sigma_n^2)$。设 $f(\xi_0, \eta_1, \cdots, \eta_n)$ 是有界 Borel 函数。$\mathcal{F}_n = \sigma(\xi_0, \eta_1, \cdots, \eta_n)$，则

$$\begin{aligned} & E\big[\exp\big(f(\xi_0, \eta_1, \cdots, \eta_{n-1})\eta_n\big)\big|\mathcal{F}_{n-1}\big] \\ = \ & \exp\Big(f(\xi_0, \eta_1, \cdots, \eta_{n-1})\mu_n + \frac{1}{2}f^2(\xi_0, \eta_1, \cdots, \eta_{n-1})\sigma_n^2\Big). \end{aligned}$$

引理 3.3. $(z_n, \mathcal{F}_n)_{n\geq 1}$ 在 P 下是鞅，而且 $Ez_n = 1$。

定理证明： 只需要证明 $\forall(\lambda_1, \cdots, \lambda_n) \in \mathbb{R}^n$，在 $\hat{P}$ 下

$$\sum_{k=1}^n \lambda_k\big(\eta_k + \frac{c_k}{b_k}\big) \sim N(0, \sum_{k=1}^n \lambda_k^2\sigma_k^2),$$

即在 $\hat{P}$ 下

$$\hat{E}\exp\Big(\sum_{k=1}^n \lambda_k\big(\eta_k + \frac{c_k}{b_k}\big)\Big) = \exp\big(\frac{1}{2}\sum_{k=1}^n \lambda_k^2\sigma_k^2\big).$$

$\cdots\cdots$ □

3.4 选样定理和 Doob 停时定理

3.4.1 停时的概念

在我们日常生活中，除了用秒、分、时、日、月、年等时间刻度记录我们日常生活的轨迹外，有时还用一些标志性事件发生的时刻来记录生活轨迹。例如，出生的时刻、进入幼儿园的时刻、第二次参加某社团活动的时刻、在大学中倒数第二次参加考试的时刻、最后一次遇到某人的时刻等等。这些时刻，虽然都是对同一事件的描述，但对于不同的人，发生的时刻可能是不一样的。在随机过程中也类似，可以用 $0, 1, 2\cdots$ 或 $t \in [0, \infty)$ 等确定性的时间来刻画一个过程，也可以用过程类似于第 i 次击中某集合或最后一次击中某集合等具有随机因素的时刻来描述一个随机过程。这就是随机时间的概念。

定义 3.3. (随机时间) 随机时间 ϑ 是一个从 $(\Omega, \mathcal{F})$ 映射到 $[0, \infty]$ 的随机变量，或 $(\Omega, \mathcal{F})$ 映射到 $\{0, 1, 2, \cdots, \infty\}$ 的随机变量，即 $\forall t \in \{0, 1, 2, \cdots\}$ 或 $t \in [0, \infty)$，$\{\omega : \vartheta(\omega) \leq t\} \in \mathcal{F}$.

注记 3.5. 在定义随机时间 ϑ 时，可以要求 ϑ 取 ∞ 值。例如，设 $(\eta_n)_{n\geq 1}$ 是 i.i.d 的随机变量序列满足 $P(\eta_k = 1) = p > \frac{1}{2}$，$P(\eta_k = -1) = 1 - p < \frac{1}{2}$，考虑非对称简单随机游动 $\xi_n = \sum_{k=1}^n \eta_k$ 首次到达某个负整数的时刻 ϑ，则 ϑ 有一个正概率等于 ∞。

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

定义 3.4. (停时，Stopping time) 设 τ 为一个随机时间，τ 为 $\Omega \to \bar{T} \equiv T \cup \{\infty\}$ 的一个随机变量，若对于 $\forall t \in T, \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$，则称 τ 是 $(\mathcal{F}_t)_{t\in T}$ 停时。

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

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

例如，全班同学上午 8 点从北大出发以不同方式前往王府井，第一次碰到尾号是 7 的汽车的时刻是一个停时，第二次碰到尾号是 7 的汽车时刻也是一个停时。比如每一时刻问所有同学是否碰到了尾号是 7 的汽车时，则每个同学都可以肯定回答“是”或“否”。但最后一次，倒数第二次碰到尾号是 7 的汽车的时刻不是停时，因为在时刻 t，无法判断 t 之后还有没有尾号是 7 的汽车开出来，因此是无法给出“是”或“否”这样确定的答案的；再如，到达王府井所用总时间的一半这个时刻也不是停时，因为总时间依赖于整个过程所有的信息 $\sigma(\cup_{t\in T}\mathcal{F}_t) = \mathcal{F}_\infty$，其一半也依赖于 $\mathcal{F}_\infty$，因此在某个确定性的时刻，比如在上午 8 点半时，是无法给出“此刻是否已经用了总时间的一半？”这个问题“是”或“否”这样确定的答案的。

命题 3.1. 常数 t 是 $(\mathcal{F}_t)_{t\in T}$ 停时，若 τ, σ 都是 $(\mathcal{F}_t)_{t\in T}$ 停时，则 $\tau \wedge \sigma = \min(\tau, \sigma)$，$\tau \vee \sigma = \max(\tau, \sigma)$ 和 $\tau + \sigma$ 也是 $(\mathcal{F}_t)_{t\in T}$ 停时。因此 $\tau \wedge t$，$\tau \vee t$ 也是停时。

证明： 只证 $\tau \wedge \sigma$，其它情况作为习题。

要证 $\tau \wedge \sigma$ 为停时，即要证 $\forall t \in T$,

$$\{\tau \wedge \sigma \le t\} \in \mathcal{F}_t.$$

事实上，

$$\{\tau \wedge \sigma \le t\} = \{\tau \le t\} \cup \{\sigma \le t\}$$

因为 τ, σ 是停时，$\{\tau \le t\} \in \mathcal{F}_t, \{\sigma \le t\} \in \mathcal{F}_t$，所以 $\{\tau \le t\} \cup \{\sigma \le t\} \in \mathcal{F}_t$。

□

定义 3.5. 设 $(\xi_n)_{n\ge1}$ 是 $(\Omega, \mathcal{F}, (\mathcal{F})_{n\ge1}, P)$ 上适应的随机序列，τ 是关于 σ-域流 $(\mathcal{F}_n)_{n\ge1}$ 的停时，设 τ 几乎处处取有限值，即 $P(\tau < \infty) = 1$，定义 $(\xi_\tau)(\omega) = \xi_{\tau(\omega)}(\omega)$，则 ξ_τ 是 $(\Omega, \mathcal{F})$ 上的随机变量，称为随机序列 $(\xi_n)_{n\ge1}$ 在停时 τ 上的取值。

$(\xi_t)_{t\ge0}$ 是 $(\Omega, \mathcal{F}, (\mathcal{F})_{t\ge0}, P)$ 上的适应随机序列，τ 是关于 σ-域流 $(\mathcal{F}_t)_{t\ge0}$ 的停时，设 τ 几乎处处取有限值，即 $P(\tau < \infty) = 1$，定义 $(\xi_\tau)(\omega) = \xi_{\tau(\omega)}(\omega)$，则在一定的条件下 (ξ_τ) 是随机变量，在本讲义中我们总假定 (ξ_τ) 是随机变量，称为随机过程 $(\xi_t)_{t\ge0}$ 在停时 τ 上的取值。

注记 3.6. 严格地说，对于 $(\Omega, \mathcal{F}, (\mathcal{F})_{n\ge1}, P)$，在以上定义中，应该要求 $(\Omega, \mathcal{F}, P)$ 是一个完备的概率空间。

注记 3.7. 在连续时间参数是要证明 ξ_τ 为一个随机变量并不是一件平凡的事情，这是随机分析中比较精细的工作。因此在本讲义中我们假定 ξ_τ 都是随机变量。

3.4.2　选样定理

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

如果博弈是公平的, 即为“鞅”时, 直观上想, 不论博弈者采取怎么样的退出策略, 他都不可能获利, 从统计平均来看应该有

$$E\xi_\tau = E\xi_0, \tag{24}$$

这就是所谓“停时定理”的本质含义.

但事实上这是不对的, 例如倍赌策略的例子 (3.5)；再如考虑下面的简单对称随机游动: $(\eta_n)_{n\geq 1}$ 是 i.i.d 的随机变量满足 $P(\eta_n=1)=P(\eta_n=-1)=\frac{1}{2}$，$\xi_0=1$, 设

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

则 $(\xi_n,\mathcal{F}_n)$ 为鞅. ξ_n 表示最初博弈者有一块钱时, 在进行了简单博弈后, 在 n 时刻的资金. 令 $\tau=\inf\{n\geq 1,\xi_n=2\}$, 即$\tau$表示博弈者首次赚到一块钱. 我们知道 $\xi_\tau=2$, $E\xi_\tau=2$，但 $E\xi_0=1$, $E\xi_\tau\neq E\xi_0$.

但在一定的条件下, (24) 成立.

定理 3.4. (选样定理) 设 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 是鞅，τ 是一个取有限值的 $(\mathcal{F}_n)_{n\geq 0}$ 停时，而且满足

(1) τ 是有界停时，即存在常数 $M<\infty$, $P(\omega:\ \tau(\omega)\leq M)=1$；或更一般地

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

则有 $E\xi_\tau=E\xi_0$.

上面的条件 (1) 用赌博术语说就是：在一场公平赌博中, 不可能利用有界停时策略来增加个人的财富期望. 条件 (2) 此处暂时理解为一个技术性条件, 事实上这个条件有更深刻的含义.

例 3.17. (赌徒输光问题) 设 $(Y_n)_{n\geq 1}$ 是 i.i.d 的随机变量序列, $P(Y_k=1)=p$, $P(Y_k=-1)=1-p$，$\xi_n=a+\sum\limits_{k=1}^{n}Y_k$. 这个模型可以想象为张三和李四用掷硬币的方式进行赌博, “出现正面”则张三赢李四一元钱, “出现反面”则张三输给李四一元钱. 在 0 时刻, 张三有 a 元钱, 李四有 b 元钱. ξ_n 表示张三在 n 时刻的资金. 若存在正整数 k_1 使得 $\xi_{k_1}=0$, 则表示张三在 k_1 时刻输光; 若存在 k_2 使得 $\xi_{k_2}=a+b$, 则表示李四在 k_2 时刻输光; 上面两种情况中一种发生, 则博弈结束.

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

令 $p_a=P(\xi_\tau=0)$,　$q_{a+b}=P(\xi_\tau=a+b)$，显然有

$$p_a+q_{a+b}=1. \tag{25}$$

求 p_a, q_{a+b} 和 $E\tau$.

解： 设 $\mathcal{F}_n=\sigma(Y_1,\cdots,Y_n)$.

(1) 若 $p=\frac{1}{2}$, 则 $(\xi_n,\mathcal{F}_n)_{n\geq 0}$ 为鞅列. 由《应用随机过程》的知识知道：(1) τ 不是一个有界停时；(2) $P(\tau<\infty)=1$. 进一步有

$$E\big(|\xi_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)\leq(a+b)E\mathbf{1}_{\{\tau>n\}}=(a+b)P(\tau>n)\to 0,\ \text{ 当 } n\to\infty.$$

故由选样定理，

$$E\xi_\tau=E\xi_0=a.$$

另一方面

$$E\xi_\tau = 0\cdot P(\xi_\tau=0)+(a+b)\cdot P(\xi_\tau=a+b)=(a+b)q_{a+b},$$

因此，与 (25) 联立求解得到

$$q_{a+b}=\frac{a}{a+b},\quad p_a=\frac{b}{a+b}.$$

进一步, 由《应用随机过程》知识得 $E\tau<\infty$, 令 $\eta_n=\xi_n^2-n\mathrm{var}(Y_1)=\xi_n^2-n$, 则 $(\eta_n,\mathcal{F}_n)_{n\ge 0}$ 是鞅. 当 $n<\tau$ 时, $|\eta_n|\le(a+b)^2+n$, 且

$$E\big(n\mathbf{1}_{\{\tau>n\}}\big)\le E\big(\tau\mathbf{1}_{\{\tau>n\}}\big)\to 0,\quad n\to+\infty.$$

由此得

$$\lim_{n\to\infty}E\Big(|\eta_n|\mathbf{1}_{\{\tau>n\}}\Big)=0.$$

由选样定理 $E\eta_\tau=E\eta_0=a^2$, 注意到 $\eta_\tau=\xi_\tau^2-\tau$, 则

$$a^2=E\xi_\tau^2-E\tau=0\cdot p_a+(a+b)^2q_{a+b}-E\tau=(a+b)a-E\tau.$$

故 $E\tau=ab$. 而且, 这两个结论都很符合我们的直观.

(2) $p>q,\quad E(Y_1)=p-q\equiv\mu>0,\quad 0<\frac{q}{p}<1$, 这不是一个公平博弈. 设

$$\zeta_n=\left(\frac{q}{p}\right)^{\xi_n},$$

则 $(\zeta_n,\mathcal{F}_n)_{n\ge0}$ 为鞅. 事实上

$$\begin{aligned}E\Big[\zeta_{n+1}\Big|\mathcal{F}_n\Big] &= E\Big[\Big(\frac{q}{p}\Big)^{\xi_{n+1}}\Big|\mathcal{F}_n\Big]=E\Big[\Big(\frac{q}{p}\Big)^{\xi_n+Y_{n+1}}\Big|\mathcal{F}_n\Big]\\ &= E\Big[\Big(\frac{q}{p}\Big)^{y+Y_{n+1}}\Big]\Big|_{y=\xi_n}\\ &= \Big[p\Big(\frac{q}{p}\Big)^{y+1}+q\Big(\frac{q}{p}\Big)^{y-1}\Big]\Big|_{y=\xi_n}\\ &= \Big(\frac{q}{p}\Big)^{y}\Big|_{y=\xi_n}=\Big(\frac{q}{p}\Big)^{\xi_n}.\end{aligned}$$

类似地也有 $\lim\limits_{n\to\infty}E\big(|\xi_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)=0$, 故

$$E\zeta_\tau=E\zeta_0=\left(\frac{q}{p}\right)^a.$$

另一方面

$$E\zeta_\tau=\left(\frac{q}{p}\right)^0p_a+\left(\frac{q}{p}\right)^{a+b}(1-p_a),$$

故而，与 (25) 联立求解得到

$$p_a=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}},\qquad q_{a+b}=\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}}.$$

另一方面
$$E\zeta_\tau=\left(\frac{q}{p}\right)^0 p_a+\left(\frac{q}{p}\right)^{a+b}(1-p_a),$$
故而，与 (25) 联立求解得到
$$p_a=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}},\qquad q_{a+b}=\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}}.$$

进一步, 求此赌博的平均持续时间. 设 $\theta_n=\xi_n-n\mu$, 则 $(\theta_n,\mathcal{F}_n)_{n\ge 0}$ 是鞅. 类似地
$$\lim_{n\to\infty}E\big(|\theta_n|\cdot\mathbf{1}_{\{\tau>n\}}\big)=0.$$
由选样定理
$$0\cdot p_a+(a+b)q_{a+b}-\mu E\tau=a,$$
因此, $E\tau=\frac{1}{\mu}[(a+b)q_{a+b}-a]$. □

在上面的例子中, 如果设 $\xi_0=0$, $\tau_1=\inf\{n\ge 0,\ \xi_n=1\}$, 虽然 τ_1 并不满足选样定理所给的条件, 但利用上面的结果, 我们也能计算出 $E\tau_1$.

设 $\tau_{[-N,1]}=\inf\{n\ge 0,\ \xi_n=-N\text{ 或 }\xi_n=1\}$, 则 $\tau_{[-N,1]}$ 是 $(\mathcal{F}_n)_{n\ge 1}$ 停时. 类似地, 可以证明
$$\lim_{n\to\infty}E\big(|\eta_n|\cdot\mathbf{1}_{\{\tau_{[-N,1]}>n\}}\big)=0,\quad \lim_{n\to\infty}E\big(|\theta_n|\cdot\mathbf{1}_{\{\tau_{[-N,1]}>n\}}\big)=0.$$
由选样定理 $E\big(\eta_{\tau_{[-N,1]}}\big)=E\eta_0=0,\ E\big(\theta_{\tau_{[-N,1]}}\big)=E\theta_0=0$, 即
$$\begin{cases} N^2p_N+1\cdot(1-p_N)-E\tau_{[-N,1]}=0, & p=\frac{1}{2},\\ -Np_N+1\cdot(1-p_N)-\mu E\tau_{[-N,1]}=0, & p\ne\frac{1}{2}.\end{cases}$$
其中 $p_N=P(\omega:\ \eta_{\tau_{[-N,1]}}=-N)$ 或者 $p_N=P(\omega:\ \theta_{\tau_{[-N,1]}}=-N)$.
$$\begin{cases} E\tau_{[-N,1]}=N, & p=\frac{1}{2},\\ E\tau_{[-N,1]}=\frac{1}{\mu}[1-(N+1)p_N]=\frac{1}{\mu}\Big[1-(N+1)\frac{\left(\frac{q}{p}\right)^N-\left(\frac{q}{p}\right)^{N+1}}{1-\left(\frac{q}{p}\right)^{N+1}}\Big], & p\ne\frac{1}{2}.\end{cases}$$
注意到 $\tau_{[-N,1]}\uparrow\tau_1$, 由 Levi 单调收敛定理
$$\begin{cases} E\tau_1=\lim\limits_{N\to+\infty}E\tau_{[-N,1]}=\infty, & p=\frac{1}{2},\\ E\tau_1=\lim\limits_{N\to+\infty}E\tau_{[-N,1]}=\frac{1}{\mu}=\frac{1}{p-q}, & p\ne\frac{1}{2}.\end{cases}$$

注记 3.8. $\tau_{[-N,1]}\uparrow\tau_1$ 并非直接可以看出来的. $\tau_{[-N,1]}=\tau_{-N}\wedge\tau_1$, 因此需证明 $\tau_{-N}\uparrow\infty$, 当 $N\to\infty$, 这可以利用《应用随机过程》中的方法来证.

3.4.3 Doob 停时定理

$\big(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},P\big)$, $\forall t$, $\mathcal{F}_t$ 为在 t 时刻之前的信息, 那么对于一个停时 τ, 在 τ 之前可以掌握的“信息”(随机事件全体) 又是什么？令 $\mathcal{F}_\infty=\sigma\big(\bigcup\limits_{t\in T}\mathcal{F}_t\big)$.

定义 3.6. $(\mathcal{F}_t)_{t\in T}$ 停时 τ 的事前事件域

$$\begin{aligned}\mathcal{F}_\tau &\equiv \left\{A\in\mathcal{F}_\infty \middle|\ \forall t\in T,\ \ A\cap\{\tau\le t\}\in\mathcal{F}_t\right\}\\ &= \left\{A\in\mathcal{F}_\infty \middle|\ \forall t\in T,\ \ \mathbf{1}_{A\cap\{\tau\le t\}}\text{为}\mathcal{F}_t\text{可测的随机变量}\right\}\end{aligned}$$

特别的，当 $T=\mathbb{Z}^+$,

$$\mathcal{F}_\tau\equiv\left\{A\in\mathcal{F}_\infty \middle|\ \forall n\in\mathbb{Z}^+,\ \ A\cap\{\tau=n\}\in\mathcal{F}_n\right\}.$$

解释：

(1) $(\mathcal{F}_t)_{t\in T}$, $\mathcal{F}_s\subset\mathcal{F}_t$, $\ s\le t$, $\forall A\in\mathcal{F}_t$ 即“到了”时刻 t, 我总能判断每一个样本是否在 A 中.

(2) 对于停时 τ 在任意的时刻 t, 应该首先判断这个随机时间 τ 是否已经“到达”, 即 $\{\tau(\omega)\le t\}$ 是否已经发生。如果 $\{\tau(\omega)\le t\}$ 发生，再看事件 A 是否属于时刻 t 已知的信息, $A\cap\{\tau(\omega)\le t\}\in\mathcal{F}_t$. 对于停时 τ 到了任意的时刻 t, 我们总能判断每一个样本是否在 $A\cap\{\tau(\omega)\le t\}$ 中, 因此, $\mathcal{F}_\tau$ 表示 τ 之前知道的“信息”.

性质 3.1. 若 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 适应过程，τ 为 $(\mathcal{F}_t)$ 停时，

(1) 若 $T=\mathbb{Z}^+$，则 ξ_τ 为 $\mathcal{F}_\tau$ 可测的随机变量；

(2) 若 $T=\mathbb{R}^+$，且 $(\xi_t)_{t\ge 0}$ 为右连续过程，即 $\forall\omega\in\Omega$, $\xi_t(\omega)$ 为右连续函数，则 ξ_τ 为 $\mathcal{F}_\tau$ 可测的随机变量。

性质 3.2. 设 τ 和 σ 为 $(\mathcal{F}_t)_{t\in T}$ 停时。若 $\tau\le\sigma$，则 $\mathcal{F}_\tau\subset\mathcal{F}_\sigma$。

问题：为什么会有选样定理？

“鞅”是一个“纯随机”的过程, 是一种“客观现象”. 它的这种“客观属性”不应该依赖于“时间坐标系”, 即 $(\mathcal{F}_t)_{t\in T}$ 的选取. 即换成另一个关于 $(\mathcal{F}_t)_{t\in T}$ 适应的停时列 $\{\tau_0\le\tau_1\le\tau_2\le\cdots\tau_n\cdots\}$ 以及 $\{\tau_n\}_{n\in\mathbb{Z}}$ 的事前事件域 $(\mathcal{F}_{\tau_n})_{n\in\mathbb{Z}}$, 它的“鞅性”还应该保持 (在一定条件下). 即 $E[\xi_t|\mathcal{F}_s]=\xi_s$, $\forall s\le t$ 为鞅. $E[\xi_\tau|\mathcal{F}_\sigma]=\xi_\sigma$, 若 $\sigma\le\tau\Rightarrow E(\xi_\tau)=E(\xi_\sigma)$. 这就是 Doob 停时定理.

定理 3.5. (Doob 有界停时定理)

$(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, $\sigma\le\tau<M<\infty$ 为两个有界 $(\mathcal{F}_t)_{t\in T}$ 停时, 则

$$E[\xi_\tau|\mathcal{F}_\sigma]=\xi_\sigma.$$

证明：仅对 $T=\mathbb{Z}^+$ 时证明.

首先, 要证 $E|\xi_\tau|<\infty$, $E|\xi_\sigma|<\infty$,

$$E|\xi_\tau|=\sum_{k=0}^{M}E\big|\xi_k\mathbf{1}_{\{\tau=k\}}\big|\le\sum_{k=0}^{M}E|\xi_k|<\infty.$$

同理, $E|\xi_\sigma|<\infty$.

其次, 要证 $\forall A\in\mathcal{F}_\sigma$, $E\big(\xi_\tau\mathbf{1}_A\big)=E\big(\xi_\sigma\mathbf{1}_A\big)$.

$$E\big(\xi_\tau\mathbf{1}_A\big)=\sum_{k=0}^{M}E\big(\xi_\tau\mathbf{1}_{\{\tau=k\}}\mathbf{1}_A\big)=\sum_{k=0}^{M}E\big(\xi_k\mathbf{1}_{A\cap\{\tau=k\}}\big).$$

因为 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, 则

$$\sum_{k=0}^{M} E\big(\xi_k \mathbf{1}_{A\cap\{\tau=k\}}\big) = \sum_{k=0}^{M} E\big(\xi_M \mathbf{1}_{A\cap\{\tau=k\}}\big) = E\big(\xi_M \mathbf{1}_A\big).$$

同理 $E\big(\xi_\sigma \mathbf{1}_A\big) = E\big(\xi_M \mathbf{1}_A\big)$.

□

定理 3.6. (Doob 停时定理)

若 $(\xi_t)_{t\in T}$ 为 $(\mathcal{F}_t)_{t\in T}$ 鞅, 且满足存在随机变量 ξ, $E|\xi| < \infty$, 使得 $\xi_t = E[\xi|\mathcal{F}_t]$, 则任意两个停时 $\sigma \le \tau$

$$E[\xi_\tau|\mathcal{F}_\sigma] = \xi_\sigma.$$

注记 3.9. 当 $T = \mathbb{R}^+$ 时, 要求 (ξ_t) 右连续.

定理 3.7. $(\xi_n, \mathcal{F}_n)_{n\ge 0}$ 鞅, 则对于任意 $(\mathcal{F}_n)$ 停时 τ, $(\xi_{\tau\wedge n}, n \ge 0)$ 是 $(\mathcal{F}_{\tau\wedge n})$ 鞅. 事实上, 还可以证明 $(\xi_{\tau\wedge n})$ 为 $(\mathcal{F}_n)$ 鞅.

定理 3.8. 若 $(\xi_t, \mathcal{F}_t)_{t\in T}$ 为下鞅, $\sigma \le \tau < M$ 为两个有界 $(\mathcal{F}_t)$ 停时, 则

$$E[\xi_\tau|\mathcal{F}_\sigma] \ge \xi_\sigma.$$

4　第四章　布朗运动

4.1　布朗运动的定义

以下是大家熟知的布朗运动的定义.

定义 4.1. 设 $\{B_t,\ t \geq 0\}$ 是一个随机过程，如果它满足以下三条

1. $B_{t+s} - B_t$ 服从正态分布 $N(0,\ \sigma^2 s)$；
2. 对任意 $0 < t_1 < t_2 < \cdots < t_n$, $B_0, B_{t_1} - B_0, \cdots, B_{t_n} - B_{t_{n-1}}$ 相互独立；
3. 对所有的 ω, $B_t(\omega)$ 关于 t 连续。

则称 $\{B_t,\ t \geq 0\}$ 是一个 (一维)Brown 运动 (Brownian motion)。当 $\sigma^2 = 1, B_0 = 0$ 时，称为标准 Brown 运动。

为什么会有这个定义呢？我们回到 1905 年 Einstein 论文，假定 (花粉) 粒子的运动是由分子的热运动引起的不规则运动，Einstein 做了如下的一些假定：

1. 每一个粒子所进行的运动，同其他一切粒子的运动都无关；
2. 同一个粒子在各个不同的时间间隔中的运动，都必须被看作是相互独立的过程，只要我们设想选择的这些时间间隔不要太小就行了。

   用现在随机过程的话来说，就是粒子运动是独立增量过程。做这样的假定是基于观测时间间隔在宏观尺度上小，但在微观尺度大，因此可以忽略掉一些在非常小尺度上的物理细节。

3. 假设液体中共有 n 个粒子，经过时间间隔 τ，单个粒子的 X 坐标将要增加 Δ，此处 Δ 对于每个粒子都有一个不同的值 (可正可负)，对于 Δ，某种分布定律成立。在时间间隔 τ 内经历了处于 Δ 和 $\Delta + d\Delta$ 之间位移的粒子数 dn，可由如下形式的方程表示
   $$dn = n\varphi(\Delta)d\Delta, \quad 此处 \int_{-\infty}^{+\infty} \varphi(\Delta)d\Delta = 1.$$
   而 φ 只是对非常小的 Δ 值才不是零，并且满足条件 $\varphi(\Delta) = \varphi(-\Delta)$。

   $\varphi(\Delta)d\Delta$ 其实就是在时间间隔 τ 内经历了处于 Δ 和 $\Delta + d\Delta$ 之间位移的粒子数占总粒子数的比例，而这个比例正比于区间间隔 $d\Delta$，比例系数与区间的位置有关。

4. 假设单位体积的粒子数 ν 只同 x 和 t 有关，设
   $$\nu = f(x,t).$$
   也就是说处于 x 位置，在时刻 t 粒子数的密度是 $f(x,t)$。

我们要从粒子在 t 时刻分布计算出 $t + \tau$ 时刻的分布。由 $\varphi(\Delta)$ 的定义可以计算出 $t+\tau$ 时刻位于 X 轴上 $[x,\ x+dx]$ 之间的粒子数，

$$\begin{aligned} f(x,t+\tau)dx &= \int_{-\infty}^{+\infty} f(x-\Delta,t)dx\varphi(\Delta)d\Delta \\ &= \int_{-\infty}^{+\infty} f(x-\Delta,t)\varphi(\Delta)d\Delta dx \\ &= \int_{-\infty}^{+\infty} f(x+\Delta,t)\varphi(\Delta)d\Delta dx. \end{aligned} \tag{26}$$

利用泰勒展开，τ 很小，因此可以设

$$f(x,t+\tau)=f(x,t)+\tau\frac{\partial f}{\partial t}. \tag{27}$$

对 $f(x+\Delta,t)$ 关于 x 变量用泰勒展开

$$f(x+\Delta,t)=f(x,t)+\Delta\frac{\partial f(x,t)}{\partial x}+\frac{\Delta^2}{2}\frac{\partial^2 f(x,t)}{\partial x^2}+\cdots \tag{28}$$

把 (27) (28) 代入 (26)

$$\begin{aligned} & f(x,t)+\frac{\partial f}{\partial t}\tau \qquad (29)\\ = \ & f(x,t)\cdot\int_{-\infty}^{+\infty}\varphi(\Delta)d\Delta+\frac{\partial f}{\partial x}\int_{-\infty}^{+\infty}\Delta\varphi(\Delta)d\Delta+\frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\varphi(\Delta)d\Delta+\cdots \end{aligned}$$

既然 $\varphi(\Delta)=\varphi(-\Delta)$，则上面偶数项为 0。因为只有很小的 Δ 值才对积分有贡献，所以第 5,7,$\cdots$ 项都比第 1 项小很多。设

$$\frac{1}{\tau}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\varphi(\Delta)d\Delta=D.$$

只考虑 (29) 式右端中的第 1 项和第 3 项，有

$$\frac{\partial f}{\partial t}=D\frac{\partial^2 f}{\partial x^2}.$$

假定在 $t=0$ 时刻，n 个粒子都集中在 $x=0$ 点，即

$$f(x,0)=\begin{cases} 0, & x\neq 0,\\ \int_{-\infty}^{+\infty}f(x,0)dx=n. & \end{cases}$$

因此

$$f(x,t)=\frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}.$$

$f(x,t)dx$ 表示 t 时刻 $[x,x+dx]$ 内粒子数，$\dfrac{f(x,t)dx}{n}=\dfrac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}dx$ 表示 t 时刻在 $[x,x+dx]$ 内粒子数占总粒子数的比例，我们现在给出 $f(x,t)dx$ 概率论的解释 (实质上是“偷换概念”) 就是在 t 时刻在 $[x,x+dx]$ 内发现某个粒子的概率，这就对应了布朗运动。

进一步计算，一个粒子在 X 轴方向上平均经历的位移 λ_X，或者比较准确的说就是在 X 轴方向上这些位移平方的算术平均的平方根

$$\lambda_X=\sqrt{\overline{X^2}}=\sqrt{2Dt}.\ \text{(平均自由程)}$$

从概率论的角度理解上面这段话可以理解为，在利用时间 $[0,t]$ 内一个粒子的平均位移替代多个粒子位移的算术平均，利用一个粒子位移平方的平均替代多个粒子位移平方的算术平均，即

$$\overline{X}=\int xf(x,t)dx=0,\quad \text{(均值)}$$

$$\sqrt{\overline{X^2}}=\Big(\int x^2f(x,t)dx\Big)^{\frac{1}{2}}=\sqrt{2Dt}.\quad \text{(标准差)}$$

利用 $\sqrt{\overline{X^2}}$ 来表示平均位移的大小。因此平均位移与时间的平方根成正比，$\sqrt{2Dt}\sim\sqrt{t}$，这点很重要。

注记 4.1. 花粉粒子的位移大致满足

$$X(t+\tau)-X(t)\sim\sqrt{\tau},\quad \frac{X(t+\tau)-X(t)}{\tau}\approx(\sqrt{\tau})'=\frac{1}{2\sqrt{\tau}}.$$

因此，当时间间隔很小时，花粉粒子运动方向变化极快速度无穷大，这就是观测到所谓的“运动的不规则性”。从数学上看就是花粉粒子的运动轨迹每一点都不可微。

由此我们抽象出布朗运动的定义

1. 粒子轨道连续；
2. $\forall 0<t_1<t_2<\cdots<t_n$，在 $[0,t_1],(t_1,t_2],\cdots,(t_{n-1},t_n]$ 之间过程独立；
3. 在 $[0,\tau]$ 和 $[t,t+\tau]$ 间运动规律相同；
4. 粒子运动规律关于 0 点对称。

我们将其严格写成数学语言，就有下面的定理。

定理 4.1. (一维情况) 设随机过程 $\{B_t,\ t\geq 0\}$ 满足

(1) $\forall 0<t_1<t_2<\cdots<t_n, B_0, B_{t_1}-B_0,\cdots,B_{t_n}-B_{t_{n-1}}$ 相互独立；

(2) $\forall s$，B_t-B_0 和 $B_{t+s}-B_s$ 同分布；

(3) 对于所有的 ω，$B_t(\omega)$ 关于 t 连续；

(4) $\forall t$，B_t-B_0 的分布关于 0 点对称。

则 $\{B_t,\ t\geq 0\}$ 一定是布朗运动或者 B_t-B_0 恒等于 0。(**这才是布朗运动本质的定义**.)

推论 4.1. 若进一步只假定 $\{B_t,\ t\geq 0\}$ 只满足上面的 (1)(2)(3)，则 $\{B_t,\ t\geq 0\}$ 一定可以表示成如下形式

$$B_t=B_0+\sigma\bar{B}_t+\mu t, \tag{30}$$

其中 $\bar{B}_t$ 为一个标准布朗运动，σ 和 μ 为常数。当 σ 和 μ 都不为零时，$\{B_t,\ t\geq 0\}$ 称为漂移布朗运动。

定义 4.2. (n 维标准布朗运动) $\{B_t^{(1)},\ t\geq 0\},\cdots,\{B_t^{(n)},\ t\geq 0\}$ 为 n 个相互独立的一维标准布朗运动，则称 $\{(B_t^{(1)},\cdots,B_t^{(n)}),\ t\geq 0\}$ 为 n 维标准布朗运动。

推论 4.2. 对于 n 维随机过程 $\{B_t=(B_t^1,\cdots,B_t^n),t\geq 0\}$，满足

(1) $\forall 0<t_1<t_2<\cdots<t_n, B_0, B_{t_1}-B_0,\cdots,B_{t_n}-B_{t_{n-1}}$ 相互独立；

(2) $\forall s$，B_t-B_0 和 $B_{t+s}-B_s$ 同分布；

(3) 对于所有的 ω，$B_t(\omega)$ 关于 t 连续。

则

$$B_t=B_0+A\cdot\bar{B}_t+b\cdot t, \tag{31}$$

其中 A 为一个 $n\times r$ 常数矩阵，$\bar{B}_t$ 为 r 维标准布朗运动，$b=(b_1,\cdots,b_n)$ 为一个 n 维常数向量。当 A 不是零矩阵，b 不是零向量时，我们称 (31) 为漂移布朗运动。

注记 4.2. 在定理4.1中，我们没有做任何定量的假设，只是尽可少地在几个定性的假设下得到了布朗运动。这从一个侧面说明了布朗运动这个概念的深刻性与广泛性。

(30) 式和 (31) 式也是 Doob 分解定理。

4.2 布朗运动的性质

命题 4.1. 对于运动 $(B_t, t \geq 0)$, 设 $\mathcal{F}_t = \sigma(B_s,\ \ s \leq t)$，则

1. $(B_t, t \geq 0)$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅；
2. $(B_t^2 - t)_{t\geq 0}$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅；
3. $(z_t = e^{CB_t - \frac{C^2 t}{2}}, t \geq 0)$ 是 $(\mathcal{F}_t)_{t\geq 0}$ 鞅。

注记 4.3. 设 $\{\xi_t,\ \ t \geq 0\}$ 为独立增量过程。即 $\forall t_1 < t_2 < \cdots < t_n$, $\{\xi_{t_k} - \xi_{t_{k-1}},\ \ k = 1, 2, \cdots\}$ 相互独立，则 $\xi_{t+s} - \xi_s$ 与 $\mathcal{F}_s = \sigma(\xi_u, u \leq s)$ 独立。

证明：

(1)

$$
\begin{aligned}
E[B_t|\mathcal{F}_s] &= E[B_t - B_s + B_s|\mathcal{F}_s] \\
&= E(B_t - B_s) + B_s = B_s
\end{aligned}
$$

(2)

$$
\begin{aligned}
E[B_t^2 - t|\mathcal{F}_s] &= E\big[(B_t - B_s)^2 + B_s^2 + 2(B_t - B_s)B_s\big|\mathcal{F}_s\big] - t \\
&= E\big[(B_t - B_s)^2\big|\mathcal{F}_s\big] + E\big[B_s^2\big|\mathcal{F}_s\big] + 2E\big[(B_t - B_s)B_s\big|\mathcal{F}_s\big] - t \\
&= E(B_t - B_s)^2 + B_s^2 + 2B_s E(B_t - B_s) - t = B_s^2 - s
\end{aligned}
$$

(3)

$$
\begin{aligned}
E[z_t|\mathcal{F}_s] &= E\big[z_s \frac{z_t}{z_s}\big|\mathcal{F}_s\big] \\
&= z_s E\big[\frac{z_t}{z_s}\big|\mathcal{F}_s\big] \\
&= z_s E\Big[e^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}\Big|\mathcal{F}_s\Big] \\
&= z_s E\Big(e^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}\Big) = z_s
\end{aligned}
$$

□

命题 4.2. $(B_t,\ t \geq 0)$ 为 $(\mathcal{F}_t)_{t\geq 0}$ 马氏过程.

命题 4.3. 设 $\big\{p(t, x, y),\ \ x \in \mathbb{R}^n,\ y \in \mathbb{R}^n\big\}$ 为 n 维标准布朗运动的转移密度，则

$$
\frac{\partial p(t,x,y)}{\partial t} = \frac{1}{2}\Delta_x p(t,x,y) = \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 p(t,x,y)}{\partial x_i^2}, \tag{32}
$$

$$
\frac{\partial p(t,x,y)}{\partial t} = \frac{1}{2}\Delta_y p(t,x,y) = \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 p(t,x,y)}{\partial y_i^2}. \tag{33}
$$

(33) 被称为 Kolmogorov 向前方程 (第二方程) 或者 Fokker-Planck 方程；(32) 被称为 Kolmogorov 向后方程 (第一方程)。

命题 4.4. Donsker 不变原理

$$\{Y_k\} \quad i.i.d. \quad E(Y_k)=0, \quad E(Y_k^2)=1$$

对于固定时刻 t，Δt 时间间隔 Δx 步长，

$$\eta_t^{\Delta} = \Delta x\big(Y_1+Y_2+\cdots+Y_{[\frac{t}{\Delta x}]}\big)$$

固定 $\Delta x, \Delta t, \omega$, 让 t 变化，则 η_t^{Δ} 为一个连续函数。

$$E\eta_t^{\Delta}=0 \quad \mathrm{var}(\eta_t^{\Delta}) = (\Delta x)^2\big[\frac{t}{\Delta t}\big].$$

若取 $h>0$ 充分小，$\Delta t = h, \quad \Delta x = \sqrt{h}, \quad h \to 0$, 则

$$(\eta_t^{\Delta}) \to (B_t)_{t\geq 0}.$$

5 第五章　随机微积分和 Itô 公式

5.1 引言

首先我们来看看微积分。由牛顿运动第一定律 (惯性定律) 告诉我们每一个物体都会继续保持其静止或沿一直线作匀速运动状态，除非有力加于其上，迫使它改变这种状态。那么，什么是匀速运动？其实就是质点的运动速度任何时刻都一样；那么接着问：什么是速度？就是单位时间内质点的位移，即设 $x(t)$ 表示质点的位移，则 $\frac{\Delta x}{\Delta t}$ 就是速度。因此，对于匀速运动我们有如下的运动方程：

$$\frac{\Delta x}{\Delta t}=v\,(\text{常数})\ \ \text{或}\ \ x(t)-x(0)=vt\ \ \text{或}\ \ x(t)-x(s)=v(t-s). \tag{34}$$

当然，这世界上质点的运动并非都是匀速直线运动。而牛顿运动第二定律 $F=ma$ 给出了非惯性运动的物理规律。那么，对于一个非匀速运动的质点，如何得到其运动方程？这就是牛顿-莱布尼茨公式:

$$x(t)-x(s)=\int_s^t \dot{x}(u)du\ \ \text{其中}\ \dot{x}(u)\equiv\lim_{\Delta u\to 0}\frac{x(u+\Delta u)-x(u)}{\Delta u}.$$

如何理解牛顿-莱布尼茨公式？其实就是用简单的运动来逼近复杂的运动，用一些匀速运动的线性叠加来表示一个非匀速运动？即认为 Δu 很小时，在 $[u,u+\Delta u]$ 时间间隔内，质点的运动 $x(t)$ 近似是一个匀速运动，其速度用 $\dot{x}(u)$ 或 $\dot{x}(\xi)$，$\xi\in[u,u+\Delta u]$ 来表示，因此质点的运动方程就近似写成

$$x(t)-x(s)=\sum x(t_{i+1})-x(t_i)\approx\sum\dot{x}(\xi_i)(t_{i+1}-t_i). \tag{35}$$

当 $\max_i(t_{i+1}-t_i)$ 越来越小趋近于 0 时，$\dot{x}(\xi)\Delta t$ 与 $x(u+\Delta t)-x(u)$ 的误差之和也越来越小收敛到 0。因此将 (35) 式右端的极限定义为 $\int_0^t\dot{x}(u)du$。

在微积分中，我们定义一个函数的微分、导数等概念，定义曲线、曲面的切线、切平面等概念。例如：$f(x)$ 在 x_0 处的切线方程是

$$z(x)=f(x_0)+f'(x_0)(x-x_0),$$

实际在说，函数 $f(x)$ 在 x_0 附近的值与 $z(x)$ 在 x_0 附近的值很接近，即 $\Delta f(x)=f'(x)\Delta x+o(\Delta x)$，写成微分形式就是 $df(x)=f'(x)dx$。

函数 $f(x,y)$ 在 (x_0,y_0) 点切平面方程为

$$z(x,y)=f(x_0,y_0)+\frac{\partial f(x,y)}{\partial x}\Big|_{(x_0,y_0)}(x-x_0)+\frac{\partial f(x,y)}{\partial y}\Big|_{(x_0,y_0)}(y-y_0),$$

也就是说，函数在 $f(x,y)$ 在 (x_0,y_0) 附近的值与 $z(x,y)$ 在 (x_0,y_0) 附近的值很接近，即

$$\Delta f(x,y)=\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+o(\sqrt{\Delta x^2+\Delta y^2})$$

写成微分形式

$$df(x,y)=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy=\Big(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\Big)\begin{pmatrix}dx\\ dy\end{pmatrix}.$$

然后，我们回到随机过程。首先问什么随机过程？对比牛顿力学中质点运动的轨迹是 $\mathbb{R}^n$ 中的一条曲线，随机过程 $\{X_t(\omega):\ t\in T\}$ 可以看成随机变量空间上的一条“曲

线”或者概率测度空间上的一条“曲线”。那么自然我们又会问：(1) 什么是这种曲线的切线？(2) 或者本质地问，什么是随机变量空间或概率测度空间中的“直线”？如何理解随机变量空间或概率测度空间中的“直线”? 什么是随机过程中的“匀速直线运动”?

匀速直线运动是“最简单”、“最理想化”的质点运动。(34) 的第三式说明在 $[s,t]$ 时间间隔内，质点的位移只与 $t-s$ 有关，与起点和终点的位置无关。在随机过程中我们将此理解为 $X_t - X_s$ 的统计性质只与 $t-s$ 有关而与起点和终点的位置无关。

而对于“最简单”、“最理想化”这样的描述在随机过程中可以有多元化的理解，在此我们可以看成“最随机”，也就是说对于任意互不相交的区间 $(s_i, t_i]$, $i = 1, \cdots, n$，

$$X_{t_n} - X_{s_n}, X_{t_{n-1}} - X_{s_{n-1}}, \cdots, X_{t_1} - X_{s_1}, X_{s_1}$$

相互独立。因此，我们可以将时齐独立增量过程理解成随机过程中的“直线”。进一步，再对轨道做一些限制，任意的 ω，$X_t(\omega)$ 是关于 t 的连续函数。由上一章定理

$$X_t - X_0 = A \cdot B_t + b \cdot t,$$

其中 $B_t = \left(B_t^{(1)} \cdots B_t^{(r)}\right)$ 是 r 维标准布朗运动，$A = (a_{ij})_{d\times r}$ 常数矩阵，$b = (b_1 \cdots b_d)$ 常数向量。$A = 0$ 时退化为 $\mathbb{R}^n$ 中的直线。特别 $d = 1$ 时，$X_t - X_0 = \sigma B_t + bt$，$(B_t)$ 为一维标准布朗运动，σ 为 0 时，则退化为 $\mathbb{R}$ 中的直线。$d = 1$ 时，设 $\{p(t,x,y), t > 0, x, y \in \mathbb{R}\}$ 为 $\sigma B_t + bt$ 的转移密度函数族，则

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{\sigma^2}{2}\Delta_x p(t,x,y) + b\frac{\partial}{\partial x}p(t,x,y), \tag{36}$$

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{\sigma^2}{2}\Delta_y p(t,x,y) - b\frac{\partial}{\partial y}p(t,x,y). \tag{37}$$

在这节的最后，我们看看随机微积分要做什么？

1. 一个轨道连续的随机过程是否可以用漂移布朗运动表示出来，应该如何表示出来？

2. 随机微积分的牛顿-莱布尼茨公式是什么？

我们看看大家熟知的轨道连续的 $\mathbb{R}$ 上的 Markov 过程是否可以用漂移布朗运动表示出来。这也是伊藤清 1942 年建立随机积分的原始想法。设 $\{p(s,x;t,y)\}$ 为 Markov 过程 $(X_t)_{t\geq 0}$ 的转移密度函数族，即对于任意有界可测函数 f

$$E\big[f(X_t)\big|\mathcal{F}_s\big] = E\big[f(X_t)\big|\sigma(X_s)\big] = \int f(y)p(s,X_s;t,y)dy.$$

我们不知道如何在随机变量空间或者概率测度空间上“求导”，但对于随机过程的一个观测 $(f(X_t))_{t\geq 0}$，对于任意的样本点 ω，它定义了 $\mathbb{R}$ 上的一条连续曲线 (但不一定可微)。进一步，我们看这些曲线的平均

$$E^{s,x}f(X_t) = \int f(y)p(s,x;t,y)dy,$$

上面的式子定义了一条从 (s,x) 出发的曲线，它是否可微？我们来看 $E^{s,x}f(X_t)$ 在无穷小未来时间内的变化，

$$E^{s,x}f(X_{s+\tau}) = \int f(y)p(s,x;s+\tau,y)dy, \qquad \tau > 0.$$

$f(X_s)$ 从 s 时刻 x 位置出发的无穷小未来的平均变化率为

$$\lim_{\tau\to 0^+}\frac{E^{s,x}f(X_{s+\tau})-E^{s,x}f(X_s)}{\tau}=\lim_{\tau\to 0^+}\frac{E^{s,x}f(X_{s+\tau})-f(x)}{\tau},$$

形式地记为 $\frac{dE^{s,x}f(X_{s+\tau})}{ds}\Big|_s$。

与 Einstein 的想法类似，由于轨道的连续性，在 τ 很小时，过程跑出 x 的一个小领域的概率非常得小，因此当假设观测函数 f 有很好的可微性时，我们可以对 $f(x)$ 在 x 附近泰勒展开到二阶，形式上得到

$$\begin{aligned} & E^{s,x}f(X_{s+\tau})-E^{s,x}f(X_s)=E^{s,x}f(X_{s+\tau})-f(x)\\ = & \int(f(y)-f(x))p(s,x;s+\tau,y)dy\\ \approx & \int(y-x)f'(x)p(s,x;s+\tau,y)dy+\frac{1}{2}\int(y-x)^2f''(x)p(s,x;s+\tau,y)dy. \end{aligned}$$

因此

$$\begin{aligned}\frac{dE^{s,x}f(X_{s+\tau})}{ds}\Big|_s = & \lim_{\tau\to 0^+}\frac{1}{\tau}\Big[\int(y-x)p(s,x;s+\tau,y)dy\Big]f'(x)\\ & +\frac{1}{2}\lim_{\tau\to 0^+}\frac{1}{\tau}\Big[\int(y-x)^2p(s,x;s+\tau,y)dy\Big]f''(x).\end{aligned}$$

若假设

$$\lim_{\tau\to 0^+}\frac{1}{\tau}\int(y-x)p(s,x;s+\tau,y)dy=b(s,x), \tag{38}$$

$$\lim_{\tau\to 0^+}\frac{1}{\tau}\int(y-x)^2p(s,x;s+\tau,y)dy=\sigma^2(s,x), \tag{39}$$

则

$$\frac{dE^{s,x}f(X_{s+\tau})}{ds}\Big|_s=\frac{1}{2}\sigma^2(s,x)\frac{d^2f}{dx^2}+b(s,x)\frac{df}{dx}.$$

事实上，直接可以计算出漂移布朗运动 σB_t+bt 的“无穷小未来的变化率”是

$$\lim_{\tau\downarrow 0^+}\frac{1}{\tau}\Big(\int f(y)p^{(\sigma,b)}(s,x;s+\tau,y)dy-f(x)\Big)=\frac{1}{2}\sigma^2\frac{d^2f}{dx^2}+b\frac{df}{dx}.$$

对比 $(\sigma B_t+bt)_{t\geq 0}$ 和 $(X_t)_{t\geq 0}$ 的“微分特征”

$$\frac{1}{2}\sigma^2\frac{d^2f}{dx^2}+b\frac{df}{dx},$$

$$\frac{1}{2}\sigma^2(s,x)\frac{d^2f}{dx^2}+b(s,x)\frac{df}{dx},$$

可以清晰地看到，$(X_t)_{t\geq 0}$ 在 (s,x) 处的无穷小未来的变化可以近似的看成一个漂移布朗运动，也就是说 $(X_t)_{t\geq 0}$ 在 (s,x) 处的“切线”或“切过程”是 $\forall t\in[s,s+\tau]$

$$\sigma(s,x)(B_t-B_s)+b(s,x)(t-s).$$

进一步，$(X_t(\omega))_{t\geq 0}$ 在 $(s,X_s(\omega))$ 处的无穷小未来可以近似看成

$$\sigma\big(s,X_s(\omega)\big)\big(B_t(\omega)-B_s(\omega)\big)+b\big(s,X_s(\omega)\big)(t-s).$$

因此，对于时间划分 $0 = t_0 < t_1 < \cdots < t_n = t$ 时，$\lambda = \max_k |t_k - t_{k-1}|$ 很小时，$X_t(\omega) - X_0(\omega)$ 近似等于

$$\begin{aligned} X_t(\omega) - X_0(\omega) &= \sum_{k=1}^{n} \big(X_{t_k}(\omega) - X_{t_{k-1}}(\omega)\big) \qquad (40)\\ &\approx \sum_{k=0}^{n-1} \Big\{\sigma\big(t_k, X_{t_k}(\omega)\big)\big[B_{t_{k+1}}(\omega) - B_{t_k}(\omega)\big] + b\big(t_k, X_{t_k}(\omega)\big)(t_{k+1} - t_k)\Big\}. \end{aligned}$$

自然的问题是当划分 $\lambda \to 0$ 时，上式是否收敛？在什么意义下收敛？第二项实际上是关于 t 的 Riemann 积分，所以收敛到 $\int_0^t b(s, X_s(\omega))ds$。第一项

$$\sum_{k=0}^{n-1} a\big(t_k, X_{t_k}(\omega)\big)\big[B_{t_{k+1}}(\omega) - B_{t_k}(\omega)\big] = \sum_{k=0}^{n-1} \sigma\big(t_k, X_{t_k}(\omega)\big)\Delta B_{t_k}$$

是否也按 Riemann-Stieltjes 积分的定义收敛？答案是否定的，那么它应该如何收敛？这就是 Itô 随机积分所要回答的问题。

注记 5.1. (38)(39) 式的物理意义是：$b(s,x)$ 表示在 s 时刻 x 位置，过程向未来运动时的平均瞬时**速度**；$\sigma^2(s,x)$ 表示在 s 时刻 x 位置，过程向未来运动时偏离初始位置大小 (方差) 的瞬时变化率 (differential variance)。缺图！

5.2 实值函数的 Stieltjes 积分

5.2.1 对单调函数的 Stieltjes 积分

命题 5.1. 设 $F(t)$ 为区间 $[a,b]$ 上的单调函数，$g(t)$ 为连续函数，对于 $[a,b]$ 的一个划分

$$a = t_0^{(n)} < t_1^{(n)} < \cdots < t_{N_n}^{(n)} = b,$$

$g(t)$ 对 $F(t)$ 的 Riemann-Stieltjes 和定义为

$$\sum_{i=1}^{N_n - 1} g(\xi_i)\big[F(t_{i+1}^{(n)}) - F(t_i^{(n)})\big], \quad t_i \le \xi_i \le t_{i+1}.$$

那么，当划分的最大半径 $\lambda_n = \max_i \{t_{i+1}^{(n)} - t_i^{(n)}\}$ 趋于 0 时，可以证明 Stieltjes 和具有不依赖于划分以及 ξ_i 的选取的极限，将这个极限记为 $\int_a^b g(t)dF(t)$，称为 $g(t)$ 对 $F(t)$ 的 Riemann-Stieltjes 积分。

注记 5.2. $\int_{[a,b]} g(t)dF(t)$ 的记法更标准，在 Stieltjes 积分中，取 $[a,b]$, $[a,b)$ 是有差别的。例如

$$F(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}.$$

设 $g(t) = 1$，则

$$\int_{[-1,0]} g(t)dF(t) = F(0) - F(-1) = 1, \quad \int_{[-1,0)} g(t)dF(t) = 0.$$

例 5.1. 设 $F(t)$ 为随机变量 ξ 的分布函数，如果

$$\int |t|dF(t) \overset{\text{def}}{=} \lim_{\substack{a\to-\infty\\ b\to+\infty}} \int_a^b |t|dF(t)$$

存在，则

$$\int tdF(t) \overset{\text{def}}{=} \lim_{\substack{a\to-\infty\\ b\to+\infty}} \int_a^b tdF(t)$$

也存在，这时有 $E\xi = \int_{-\infty}^{\infty} tdF(t)$. 若进一步，$F(t)$ 有导数 $f(t) = \frac{dF}{dt}$ 且 $|f(t)|$ Lebesgue 可积，那么 Stieltjes 积分就可以转化为普通的定积分

$$\int_a^b g(t)dF(t) = \int_a^b g(t)f(t)dt.$$

若 $g(t)$ 连续可微，那么

$$\int_{[a,b]} g(t)dF(t) = \big[g(b)F(b) - g(a)F(a)\big] - \int_a^b F(t)g'(t)dt.$$

5.2.2 对有界变差函数的 Stieltjes 积分

定义 5.1. 对于区间 $[a,b]$ 如下的一个划分，$P_{t_0^{(n)}\cdots t_{N_n}^{(n)}}:\quad a = t_0^{(n)} < \cdots < t_{N_n}^{(n)} = b$。

$$\bigvee\big(P_{t_0^{(n)}\cdots t_{N_n}^{(n)}}\big) \overset{\text{def}}{=} \big|f\big(t_1^{(n)}\big) - f\big(t_0^{(n)}\big)\big| + \cdots + \big|f\big(t_{N_n}^{(n)}\big) - f\big(t_{N_n-1}^{(n)}\big)\big|$$

称为函数 $f(t)$ 在划分 $P_{t_0^{(n)}\cdots t_{N_n}^{(n)}}$ 上的变差；

$$\bigvee_{[a,b]}(f) \overset{\text{def}}{=} \sup_{[a,b]\text{的所有划分}} \bigvee\big(P_{t_0^{(n)}\cdots t_{N_n}^{(n)}}\big)$$

称为函数 $f(t)$ 在区间 $[a,b]$ 上的全变差。若 $f(t)$ 在 $[a,b]$ 上全变差为有限值时，称为 $f(t)$ 为 $[a,b]$ 上的有界变差函数。

设 $F(t)$ 有界变差函数，则总能表示为两个递增函数的差

$$F(t) = F_1(t) - F_2(t).$$

因此可以定义关于 $F(t)$ 的 Stieltjes 积分

$$\int_{[a,b]} g(t)dF(t) = \int_{[a,b]} g(t)dF_1(t) - \int_{[a,b]} g(t)dF_2(t).$$

直观上，对于不同的 ξ_i 的选取，其误差之和

$$\begin{aligned}
& \Big|\sum g(\xi_i)\big[F(t_{i+1}) - F(t_i)\big] - \sum g(\xi_i')\big[F(t_{i+1}) - F(t_i)\big]\Big| \\
= \; & \Big|\sum \big(g(\xi_i) - g(\xi_i')\big)\big[F(t_{i+1}) - F(t_i)\big]\Big| \le \sum \big|g(\xi_i) - g(\xi_i')\big|\big|F(t_{i+1}) - F(t_i)\big| \\
\le \; & \max_i\{\big|g(\xi_i) - g(\xi_i')\big|\} \sum \big|F(t_{i+1}) - F(t_i)\big| \le \max_i\{\big|g(\xi_i) - g(\xi_i')\big|\} \bigvee_{[a,b]}(F).
\end{aligned}$$

所以，当 g 为 $[a,b]$ 上连续函数，F 为有界变差函数时，上式随着划分越来越细，可以任意得小，由此得 Stieltjes 积分不依赖于 ξ_i 的选取和划分的选取。

注记 5.3. 若 $g(t)$ 不是连续函数，对于有界变差函数 $F(t)$，则当划分的最大半径 $\lambda_n = \max_i\{\big|t_{i+1}^{(n)} - t_i^{(n)}\big|\}$ 趋于 0 时，只要 Riemann-Stieltjes 和具有不依赖于划分以及 ξ_i 选取的极限，则称 $g(t)$ 对 $F(t)$ Stieltjes 可积，并称此极限为 $g(t)$ 对 $F(t)$ 的 Stieltjes 积分。

5.2.3 Brown 运动的轨道性质

我们在《应用随机过程》中知道对于几乎处处的 ω，Brown 运动的轨道 $B_t(\omega)$ 是一个无处可微的连续函数，因此不是有界变差函数，我们不能在 Stieltjes 积分的意义下定义 $\int f(t)dB_t(\omega)$。当然，我们也可以从其它的角度来证明 Brown 运动的轨道不是有界变差函数。

引理 5.1. (Lévy 的振动性质) 设 $(B_t)_{t\geq 0}$ 为 Brown 运动，又

$$s_1 = t_0 < t_1 < \cdots < t_n = s_2,$$

$$\Delta t_k = t_{k+1} - t_k, \quad \Delta B_{t_k} = B_{t_k+1} - B_{t_k}, \quad h = \max_{0\leq k\leq n-1} \Delta t_k,$$

那么，

$$E\Big|\sum_{k=0}^{n-1}(\Delta B_{t_k})^2 - (s_2 - s_1)\Big|^2 \leq 2h(s_2 - s_1). \tag{41}$$

证明：

$$\begin{aligned}
(41)\text{式左端} &= E\Big|\sum_{k=0}^{n-1}[(\Delta B_{t_k})^2 - \Delta t_k]\Big|^2 \\
&= \sum_{k=0}^{n-1} E\big[(\Delta B_{t_k})^2 - \Delta t_k\big]^2 + \sum_{k\neq j} E\Big(\big[(\Delta B_{t_k})^2 - \Delta t_k\big]\big[(\Delta B_{t_j})^2 - \Delta t_j\big]\Big).
\end{aligned}$$

由于当 $k \neq j$ 时，ΔB_{t_k} 与 ΔB_{t_j} 相互独立且 $E(\Delta B_{t_k})^2 = \Delta t_k$，故第二项等于零。而第一项为

$$\begin{aligned}
\sum_{k=0}^{n-1} E\big[(\Delta B_{t_k})^2 - \Delta t_k\big]^2 &= \sum_{k=0}^{n-1} E\big[(\Delta B_{t_k})^4 + (\Delta t_k)^2 - 2(\Delta B_{t_k})^2 \cdot \Delta t_k\big] \\
&= \sum_{k=0}^{n-1} \Big[3\big(E(\Delta B_{t_k})^2\big)^2 - (\Delta t_k)^2\Big] = 2\sum_{k=0}^{n-1}(\Delta t_k)^2 \\
&= 2\sum_{k=0}^{n-1} |\Delta t_k| \cdot |\Delta t_k| \\
&\leq 2h\sum_{k=0}^{n-1} \Delta t_k = 2h(s_2 - s_1).
\end{aligned}$$

□

推论 5.1. (Brown 运动平方变差有限)

$$\lim_{h\to 0} \sum_{k=0}^{n-1}(\Delta B_{t_k})^2 \xrightarrow{L^2} (s_2 - s_1).$$

即

$$\lim_{h\to 0} E\Big|\sum_{k=0}^{n-1}(\Delta B_{t_k})^2 - (s_2 - s_1)\Big|^2 = 0.$$

引理 5.2. 设 $[s_1, s_2]$ 的 2^n 等分点为

$$s_1 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{2^n}^{(n)} = s_2,$$

那么

$$P\big(\omega : \lim_{n\to\infty} \sum_{k=0}^{2^n-1} (\Delta B_{t_k})^2 = s_2 - s_1\big) = 1.$$

引理 5.3. 以概率为 1 的 Brown 运动的轨道在 t 的任何区间内都不是有界变差的。

证明： 令 $\{t_k^{(n)}\}$ 为引理5.2中的划分，记

$$\lambda_n(\omega) = \max_k\{\big|\Delta B_{t_k^{(n)}}(\omega)\big|\},$$

因为 $B_t(\omega)$ 在 $[s_1, s_2]$ 上一致连续，所以当 $\max\limits_k \Delta t_k^{(n)} \to 0$ 时 $\lambda_n(\omega) \to 0$。因此，

$$\sum_{k=0}^{2^n-1} (\Delta B_{t_k^{(n)}})^2 \le \lambda_n(\omega) \sum_k \big|\Delta B_{t_k^{(n)}}(\omega)\big|,$$

由引理5.2，对于几乎处处的轨道有

$$\sum_k \big|\Delta B_{t_k^{(n)}}(\omega)\big| \ge \frac{\sum\limits_{k=0}^{2^n-1} (\Delta B_{t_k^{(n)}})^2}{\lambda_n(\omega)} \xrightarrow{n\to\infty} \infty.$$

□

5.3 关于 Brown 运动的 Itô 积分

由前面的引言以及积分的思想，我们知道，首先我们应该对某种“简单过程”来定义“随机积分”，然后再看“一般过程”是否可以表示成这种“简单过程”的“极限”，最后再看对于逼近的“简单过程”的“积分”是否收敛？若收敛，则定义这个极限就是“一般过程”的“积分”。

定义 5.2. 设 $(B_t)_{t\ge 0}$ 为一个 Brown 运动，称一个随机过程 $(\phi_t)_{t\ge 0}$ 关于 $(B_t)_{t\ge 0}$ 适应 (可知) 是指 $\forall t \ge 0$，ϕ_t 关于 $\sigma(B_s,\ s \le t)$ 可测。记 $\sigma(B_s,\ s\le t) = \mathcal{F}_t^B$.

定义 5.3. 对于区间 $[0, T]$ 的一个划分 $0 = t_0 < t_1 < \cdots < t_n = T$，当随机过程 $(\phi_t)_{t\ge 0}$ 是阶梯过程，即它可以写成如下形式时

$$\phi_t = \sum_{i=0}^{n-1} \phi_{t_i} \mathbf{1}_{(t_i, t_{i+1}]}(t) + \phi_0 \mathbf{1}_{\{0\}}(t),$$

若其中的系数随机变量 ϕ_{t_i} 为 $\sigma(B_u,\ u \le t_i)$ 可测的，而且 ϕ_{t_i}，$i = 0, \cdots, n-1$ 为有界的随机变量 (或 $E|\phi_{t_i}|^2 < \infty$)，则称 $(\phi_t)_{t\ge 0}$ 为 $(B_t)_{t\ge 0}$ 可料阶梯过程。

注记 5.4. (1) 有些书上称可料阶梯过程为简单 (可料) 适应过程或初等 (可料) 适应过程。

(2) 可料性在关于跳过程的随机积分理论中特别重要。

定义 5.4. 设 $(\phi_t)_{t\geq 0}$ 为 $(B_t)_{t\geq 0}$ 可料的阶梯过程。定义 $(\phi_t)_{t\geq 0}$ 关于 Brown 运动 $(B_t)_{t\geq 0}$ 的 Itô 随机积分如下

$$\int_0^T \phi_t dB_t = \sum_{i=0}^{n-1} \phi_{t_i}(B_{t_{i+1}} - B_{t_i}),$$

它是一个随机变量。

引理 5.4. $(B_t)_{t\geq 0}$ 可料阶梯过程 $(\phi_t)_{t\geq 0}$ 的 Itô 随机积分满足

(1) $E\big(\int_0^T \phi_t dB_t\big) = 0$;

(2) Itô 等距

$$E\Big|\int_0^T \phi_t dB_t\Big|^2 = \Big\|\int_0^T \phi_t dB_t\Big\|^2 = \sum_{i=0}^{n-1} \|\phi_{t_i}\|^2(t_{i+1}-t_i) = \int_0^T \|\phi_t\|^2 dt = E\int_0^T |\phi_t|^2 dt,$$

其中 $\|\phi_{t_i}\|^2 = E|\phi_t|^2$。

(3) 线性性质:

$$\int_0^T (\phi+\psi)dB = \int_0^T \phi dB + \int_0^T \psi dB,$$
$$\int_0^T c\phi dB = c\int_0^T \phi dB,\ \ c\text{为常数}.$$

证明:

(1)

$$E\Big(\int_0^T \phi_t dB_t\Big) = E\sum_{i=0}^{n-1} \phi_{t_i}(B_{t_{i+1}} - B_{t_i}) \tag{42}$$
$$= \sum_{i=0}^{n-1} E\phi_{t_i}(B_{t_{i+1}} - B_{t_i}).$$

注意到 ϕ_{t_i} 为 $\sigma(B_u,\ u\leq t_i)$ 可测，故 ϕ_{t_i} 与 $B_{t_{i+1}} - B_{t_i}$ 独立，因此

$$(42)\text{的右端} = \sum_{i=0}^{n-1} E\phi_{t_i}E(B_{t_{i+1}} - B_{t_i}) = 0.$$

(2)

$$
\begin{aligned}
& E\Big|\int_0^T \phi_t dB_t\Big|^2 = E\Big(\sum_{i=0}^{n-1}\phi_{t_i}(B_{t_{i+1}}-B_{t_i})\Big)^2 \\
= \; & E\Big(\sum_{i=0}^{n-1}\phi_{t_i}^2(B_{t_{i+1}}-B_{t_i})^2 + 2\sum_{j<k}\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\Big) \\
= \; & \sum_{i=0}^{n-1} E\Big(E\big[\phi_{t_i}^2(B_{t_{i+1}}-B_{t_i})^2\big|\mathcal{F}_{t_i}^B\big]\Big) \\
& +2\sum_{j<k}E\Big(E\big[\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\big|\mathcal{F}_{t_k}^B\big]\Big) \\
= \; & \sum_{i=0}^{n-1} E\Big(\phi_{t_i}^2 E\big[(B_{t_{i+1}}-B_{t_i})^2\big|\mathcal{F}_{t_i}^B\big]\Big) \\
& +2\sum_{j<k}E\Big(E\big[\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})(B_{t_{k+1}}-B_{t_k})\big|\mathcal{F}_{t_k}^B\big]\Big) \\
= \; & \sum_{i=0}^{n-1} E\big(\phi_{t_i}^2 E(B_{t_{i+1}}-B_{t_i})^2\big) \\
& +2\sum_{j<k}E\Big(\phi_{t_j}\phi_{t_k}(B_{t_{j+1}}-B_{t_j})E\big[B_{t_{k+1}}-B_{t_k}\big|\mathcal{F}_{t_k}^B\big]\Big) \\
= \; & \sum_{i=0}^{n-1} E\big(\phi_{t_i}^2(t_{i+1}-t_i)\big) = \sum_{i=0}^{n-1}\|\phi_{t_i}\|^2(t_{i+1}-t_i) = \int_0^T \|\phi_t\|^2 dt.
\end{aligned}
$$

□

对于一般的过程应该如何定义？

$\mathscr{L}_T^2 = \big\{(\phi_t)_{0\le t\le T}$ 是 $(\mathcal{F}_t^B)_{t\ge 0}$ 适应的随机过程且 $\int_0^T E|\phi_t(\omega)|^2 dt < +\infty\big\}$。

(1) $\mathscr{L}_T^2$ 是一个线性空间；

(2) $\mathscr{L}_T^2$ 中可以定义内积

$$\phi,\ \psi\in\mathscr{L}_T^2,\qquad \text{则}\langle\phi,\ \psi\rangle = \int_0^T E\big(\phi_t(\omega)\psi_t(\omega)\big)dt.$$

(3) 因此 $\mathscr{L}_T^2$ 在内积 $\langle\phi,\ \psi\rangle$ 下是一个 (无穷维) 欧氏空间，$\mathscr{L}_T^2$ 在距离

$$\|\phi-\psi\| = \Big(\int_0^T E\big|\phi_t-\psi_t\big|^2 dt\Big)^{\frac{1}{2}},$$

下成为一个完备的距离空间，即若 $(\phi^{(n)})_{n\ge 1}$ 是 Cauchy 列，当 $n\to\infty$, $m\to\infty$ 时，$\|\phi_n-\psi_m\|\to 0$，则必有 $\phi\in\mathscr{L}_T^2$ 使得 $\|\phi^{(n)}-\psi\|\to 0$, $n\to\infty$，即 $\mathscr{L}_T^2$ 是一个 Hilbert 空间。

考虑 $\mathscr{L}_T^2$ 中的连续随机过程 $(\phi_t)_{t\ge 0}$ 及其在 $[0,T]$ 的一列划分

$$0=t_0^{(n)}<t_1^{(n)}<\cdots<t_{N_n}^{(n)}=T,\quad \lambda_n=\max_{0\le k\le N_n-1}\big\{t_{k+1}^{(n)}-t_k^{(n)}\big\}.$$

对于固定的 n, 设

$$\phi_t^{(n)} = \sum_{k=0}^{N_n-1} \phi_{t_k} \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t) + \phi_0 \mathbf{1}_{\{0\}}(t)$$

称 $(\phi_t^{(n)})_{t\geq 0}, n \geq 1$ 为近似 $(\phi_t)_{t\geq 0}$ 的 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料阶梯随机过程。

引理 5.5. 对于 $\mathscr{L}_T^2$ 中的有界连续随机过程 $(\phi_t)_{t\geq 0}$ (即 $|\phi_t(\omega)| < M, a.s. \omega, \forall t \in [0,T]$) 及其近似的 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料的阶梯随机过程列 $(\phi_t^{(n)})_{t\geq 0}$，有

$$\|\phi^{(n)} - \phi\|^2 = E\int_0^T \left|\phi_t^{(n)} - \phi_t\right|^2 dt \to 0, \qquad \text{当}\lambda_n \to 0\text{时}.$$

引理 5.6. 对于 $\mathscr{L}_T^2$ 中的随机过程 $(\phi_t)_{t\geq 0}$ 必存在 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料的阶梯随机过程列 $(\phi_t^{(n)})_{t\geq 0}$，当 $n \to \infty$ 时

$$\|\phi^{(n)} - \phi\|^2 = E\int_0^T \left|\phi_t^{(n)} - \phi_t\right|^2 dt \to 0,$$

从而 $(\phi^{(n)})_{n\geq 1}$ 也是 $\mathscr{L}_T^2$ 中的 Cauchy 列，称为近似 $(\phi_t)_{t\geq 0}$ 的 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料阶梯随机过程。

注记 5.5. 为什么非要取可料的近似列?

(1) 从这一章的引言中我们知道 Itô 当年就是利用漂移 Brown 运动来近似 Markov 过程在 s 时刻无穷小未来的想法来逼近整个 Markov 过程，

$$\sum a(t_k, X_{t_k})(B_{t_{k+1}} - B_{t_k}) + b(t_k, X_{t_k})(t_{k+1} - t_k),$$

从而定义了 Itô 随机积分。

(2) 从引理5.4的证明知，当 (ϕ_t) 为 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料时，随机积分有很好的性质，例如均值为零和 Itô 等距。

(3) 后面我们还可以看作随机积分当 T 变动时成为一个鞅。

推论 5.2. 对于引理5.5和引理5.6中的随机过程 $(\phi_t)_{t\geq 0}$ 及其近似的可料阶梯随机过程列 $(\phi_t^{(n)})_{t\geq 0}$ 有

$$\|\int_0^T \phi_t^{(n)} dB_t\|^2 = \int_0^T \|\phi_t^{(n)}\|^2 dt = \|\phi^{(n)}\|^2$$

且 $\int_0^T \phi_t^{(n)} dB_t$ 是 $L^2(\Omega)$ 中的 Cauchy 列。

定义 5.5. 对于 $\mathscr{L}_T^2$ 中的随机过程 $(\phi_t)_{t\geq 0}$，$(\phi_t^{(n)})_{t\geq 0}$ 为其近似的 $(\mathcal{F}_t^B)_{t\geq 0}$ 的可料阶梯随机过程列，满足

$$\|\phi^{(n)} - \phi\|^2 = E\int_0^T \left|\phi_t^{(n)} - \phi_t\right|^2 dt \to 0$$

(这时 $\int_0^T \phi_t^{(n)} dB_t$ 是 $L^2(\Omega)$ 中的 Cauchy 列)。定义

$$\int_0^T \phi_t dB_t = L^2(\Omega) - \lim_{n\to\infty} \int_0^T \phi_t^{(n)} dB_t.$$

称为 Itô 随机积分。

从下面的图表我们不能看出随机积分的定义与 Riemann 积分或 Lebesgue 积分的定义想法是类似的

Riemann 积分或 Lebesgue 积分	**随机积分**
定义简单函数的积分	定义可料阶梯 (简单适应) 过程的随机积分
简单函数逼近一般函数 $f_n \to f$	可料阶梯过程逼近适应过程 $\phi^{(n)} \to \phi$
$\int f_n dx$ 在$\mathbb{R}$收敛($\mathbb{R}$完备)	$\int \phi^{(n)} dB$在$L^2(\Omega)$收敛 ($L^2(\Omega)$完备)
$\int f_n dx$的极限定义为$\int f dx$	$\int \phi^{(n)} dB$的极限定义为$\int \phi dB$

注记 5.6. 1. 可以证明使用不同的 $(\mathcal{F}_t^B)_{t\geq 0}$ 可料阶梯随机过程近似列，在不计零概率事件差异下不会影响随机积分的定义。

$$\int_0^T \phi_t dB_t \quad \text{常常简记为} \quad \int_0^T \phi dB.$$

2. $(\phi_t^{(n)})$ 的可料性起了关键性作用，由引理5.4的 Itô 等距，

$$E\Big(\int_0^T (\phi_t^{(n)} - \phi_t^{(m)}) dB_t\Big)^2 = \int_0^T E\big(\phi_t^{(n)} - \phi_t^{(m)}\big)^2 dt = \|\phi^{(n)} - \phi^{(m)}\|^2.$$

3. 类似地可以定义

$$\int_a^b \phi dB = \int_0^b \phi dB - \int_0^a \phi dB.$$

或者直接类似于以上的定义方式直接在区间 $[a, b]$ 上定义随机积分。

例 5.2. 求 $\int_0^t B_s dB_s$。

解： $B_t^{(n)} = \sum\limits_{k=0}^{N_n-1} B_{t_k^{(n)}} \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t)$，则

$$\begin{aligned}
\int_0^t B_s^{(n)} dB_s &= \sum_k B_{t_k^{(n)}}\big(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}\big) \\
&= \frac{1}{2}\sum_k \big(2B_{t_k^{(n)}} B_{t_{k+1}^{(n)}} - 2B^2_{t_k^{(n)}}\big) \\
&= \frac{1}{2}\sum_k \Big(\big(B^2_{t_{k+1}^{(n)}} - B^2_{t_k^{(n)}}\big) - \big(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}\big)^2\Big) \\
&= \frac{1}{2}\sum_k \big(B^2_{t_{k+1}^{(n)}} - B^2_{t_k^{(n)}}\big) - \frac{1}{2}\sum_k \big(\Delta B_{t_k^{(n)}}\big)^2 \\
&= \frac{1}{2}(B_t^2 - B_0^2) - \frac{1}{2}\sum_k \big(\Delta B_{t_k^{(n)}}\big)^2
\end{aligned}$$

由上一节引理5.1的 Lévy 振动性质知 $E\big(\sum\limits_k \big(\Delta B_{t_k^{(n)}}\big)^2\big)^2 \to t$, 因此

$$E\Big(\int_0^t B_s^{(n)} dB_s - \big(\frac{1}{2}(B_t^2 - B_0^2) - \frac{t}{2}\big)\Big)^2 \to 0,$$

故而

$$\int_0^t B dB = \frac{1}{2}(B_t^2 - B_0^2) - \frac{1}{2}t.$$

当 $B_0 = 0$ 时，$\int_0^t B dB = \frac{1}{2}B_t^2 - \frac{1}{2}t$。

例 5.3. 为什么非要取划分左端点的值，用可料逼近呢？为什么不能类似 Riemann-Stieltjes 积分在划分区间中的任意点上取值呢？下面我们设 $(B_t)_{t\geq 0})$ 标准 Brown 运动。

在例5.2中取的是划分左端点的值 $\sum_{k=0}^{N_n-1} B_{t_k^{(n)}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t)$ 来逼近的 $(B_t)_{t\geq 0}$。现在用划分右端点的值来做逼近序列看看会有什么结果呢？

$$\tilde{\phi}_t^{(n)} = \sum_k B_{t_{k+1}^{(n)}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t).$$

考虑

$$H_n^{\tilde{\phi}} \equiv \sum_k B_{t_{k+1}^{(n)}}\Delta B_{t_k^{(n)}} = \sum_{k=0}^{N_n-1} B_{t_{k+1}^{(n)}}\big(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}\big),$$

的 $L^2(\Omega)$ 极限，则

$$H_n^{\tilde{\phi}} \xrightarrow{L^2} \frac{1}{2}B_t^2 + \frac{1}{2}t.$$

若取

$$\Psi_t^{(n)} = \sum_{k=0}^{N_n-1} \frac{B_{t_k^{(n)}} + B_{t_{k+1}^{(n)}}}{2}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t),$$

$$H_n^{\Psi} \equiv \sum_{k=0}^{N_n-1} \frac{B_{t_k^{(n)}} + B_{t_{k+1}^{(n)}}}{2}\Delta B_{t_k^{(n)}} \xrightarrow{L^2} \frac{1}{2}B_t^2.$$

若取

$$\xi_t^{(n)} = \sum_{k=0}^{N_n-1} B_{\frac{t_k^{(n)}+t_{k+1}^{(n)}}{2}}\mathbf{1}_{(t_k^{(n)},t_{k+1}^{(n)}]}(t),$$

$$H_n^{\xi} \equiv \sum_{k=0}^{N_n-1} B_{\frac{t_k^{(n)}+t_{k+1}^{(n)}}{2}}\Delta B_{t_k^{(n)}} \xrightarrow{L^2} \frac{1}{2}B_t^2.$$

□

因此，从以上例子可以看出，对于随机积分并不是非要在 $[t_k^{(n)}, t_{k+1}^{(n)}]$ 取左端点作为“代表点”才能保证收敛性的，也可以在其它位置取点，但其求和的极限是不同的。问题是那么取左端点逼近定义的随机积分又有什么好处呢？我们马上就可以通过随机积分的性质看到。

性质 5.1. 随机积分性质 I

1. 可测性: $\int_0^t \phi dB$ 关于 $\mathcal{F}_t^B$ 可测。

2. 线性性质:

$$\int_0^t (\phi+\psi)dB = \int_0^t \phi dB + \int_0^t \psi dB,$$
$$\int_0^t c\phi dB = c\int_0^t \phi dB,\quad c\text{为常数}.$$

对于 $\sigma(B_u,\ u\leq a)$ 可测的有界随机变量 η,

$$\int_a^b \eta\phi dB = \eta\int_a^b \phi dB.$$

3. 可加性: 对 $a<b<c$ 有

$$\int_a^c \phi dB = \int_a^b \phi dB + \int_b^c \phi dB.$$

4. Itô 随机积分的数学期望，协方差和方差

$$E\Big(\int_0^t \phi_s dB_s\Big) = 0,$$

$$E\Big(\int_0^t \phi_s dB_s \int_0^t \psi_s dB_s\Big) = \int_0^t E(\phi_s\psi_s)ds.$$

特别 $\phi=\psi$ 时，

$$E\Big(\int_0^t \phi_s dB_s\Big)^2 = E\Big(\int_0^t \phi_s^2 ds\Big). \tag{43}$$

(43) 被称为 Itô 等距，逼近序列的“可料性”在证明中起了关键的作用。

进一步，我们可以考虑 $0<t\le T$ ，对于 $\phi\in\mathscr{L}_T^2$，则可以定义一个随机过程 $\big(\int_0^t \phi_s dB_s\big)_{t\in[0,T]}$.

注记 5.7. 对于 $\forall t$, $\int_0^t \phi_s dB_s$ 是几乎处处定义的，但总可以找到这样一个 $(\mathcal{F}_t^B)_{t\ge 0}$ 适应过程 $(z_t)_{t\in[0,T]}$ 满足 $(z_t)_{0\le t\le T}$ 是连续过程，而且 $\forall t$

$$P\Big(\int_0^t \phi_s dB_s = z_t\Big) = 1.$$

因此，以后总认为 $\big(\int_0^t \phi_s dB_s\big)_{t\in[0,T]}$ 这个过程是一个连续的关于 $(\mathcal{F}_t^B)_{t\ge 0}$ 适应的随机过程。

由 Itô 随机积分定义的随机过程 $\big(\int_0^t \phi_s dB_s\big)_{t\in[0,T]}$ 还有以下的性质。

性质 5.2. 随机积分性质 II

1. τ 是 $\mathcal{F}_t^B=\sigma(B_u,\ u\le t)$ 停时，若 $\tau\le T$，则

$$\int_0^\tau \phi_t dB_t = \int_0^T \phi_t \mathbf{1}_{(0,\tau]}(t) dB_t.$$

2. $\big\{\xi_t \stackrel{\text{def}}{=} \int_0^t \phi_s dB_s\big\}$ 是 $(\mathcal{F}_t^B)_{t\in[0,T]}$ 鞅。

3. $\big\{\eta_t \stackrel{\text{def}}{=} (\int_0^t \phi_s dB_s)^2 - \int_0^t \phi_s^2 ds\big\}$ 是 $(\mathcal{F}_t^B)_{t\in[0,T]}$ 鞅，且

$$E\Big[\Big(\int_s^t \phi_u dB_u\Big)^2\Big|\mathcal{F}_s^B\Big] = E\Big[\int_s^t \phi_u^2 du\Big|\mathcal{F}_s^B\Big].$$

4. 若 $(\phi_s)_{s\in[0,T]}$ 是关于 $(\mathcal{F}_t^B)$ 适应的有界随机过程，即 $\forall t$, $|\phi_t|<M<\infty$，则

$$\Big\{\zeta_t \stackrel{\text{def}}{=} e^{\int_0^t \phi_s dB_s - \frac{1}{2}\int_0^t \phi_s^2 ds}\Big\}$$

为 $(\mathcal{F}_t^B)_{t\in[0,T]}$ 鞅。

注记 5.8. 上面 ϕ 的有界性条件可以进一步减弱，例如 Novikov 条件 $Ee^{\frac{1}{2}\int_0^t \phi_s^2 ds} < +\infty$ 就可以保证鞅性.

证明：我们只证 (2)(3)。当 $(\phi_t)_{t\geq 0}$ 是可料阶梯过程时

$$\phi_t = \sum_{i=0}^{n-1} \phi_{t_i} \mathbf{1}_{(t_i, t_{i+1}]}(t) + \phi_0 \mathbf{1}_{\{0\}}(t),$$

对于 $s < t'$，将 s 和 t' 加入到 $0 = t_0 < t_1 < \cdots < t_j < \cdots$ 成为

$$0 = t_0 < \cdots < t_i < s < t' \leq t_{i+1} < \cdots$$

或者

$$0 = t_0 < \cdots < t_i < s \leq t_{i+1} < \cdots < t_j < t' \leq t_{j+1} \cdots .$$

重新记这些划分点为

$$0 = t'_0 < t'_1 < \cdots < t'_p < \cdots < t'_q < \cdots ,$$

其中 $s = t'_p$, $t' = t'_q$。

$$\begin{aligned}
& E\Big[\int_0^{t'} \phi_u dB_u \Big| \mathcal{F}_s^B\Big] \\
= \;& \sum_{k=0}^{q-1} E\Big[\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\Big|\mathcal{F}^B_{t'_p}\Big] \\
= \;& \sum_{k=0}^{p-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k}) + \sum_{k=p}^{q-1} E\Big[\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\Big|\mathcal{F}^B_{t'_p}\Big] \\
= \;& \sum_{k=0}^{p-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k}) + \sum_{k=p}^{q-1} E\Big[E\big[\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\big|\mathcal{F}^B_{t'_k}\big]\Big|\mathcal{F}^B_{t'_p}\Big] \\
= \;& \sum_{k=0}^{p-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k}) + \sum_{k=p}^{q-1} E\Big[E\phi_{t'_k}E(B_{t'_{k+1}} - B_{t'_k})\Big|\mathcal{F}^B_{t'_p}\Big] \\
= \;& \sum_{k=0}^{p-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k}) \\
= \;& \int_0^s \phi_u dB_u.
\end{aligned}$$

对于 $\phi \in \mathscr{L}_T^2$，存在着一列可料阶梯过程 $\phi^{(n)}$ 在 $\mathscr{L}_T^2$ 中逼近 ϕ。对于任意的 $A \in \mathcal{F}_s^B$，

$$\begin{aligned}
& \Big|E\Big(\int_0^{t'} \phi dB \mathbf{1}_A\Big) - E\Big(\int_0^{t'} \phi^{(n)} dB \mathbf{1}_A\Big)\Big| \\
\leq \;& \Big(E\Big|\int_0^{t'} (\phi - \phi^{(n)}) dB\Big|^2\Big)^{\frac{1}{2}} P(A)^{\frac{1}{2}} \\
= \;& \Big(E\int_0^{t'} (\phi - \phi^{(n)})^2 du\Big)^{\frac{1}{2}} P(A)^{\frac{1}{2}} \\
= \;& \Big(\int_0^{t'} E(\phi - \phi^{(n)})^2 du\Big)^{\frac{1}{2}} P(A)^{\frac{1}{2}} \\
\overset{n\to\infty}{\longrightarrow} \;& 0.
\end{aligned}$$

同理

$$\Big|E\Big(\int_0^s \phi dB\mathbf{1}_A\Big) - E\Big(\int_0^s \phi^{(n)} dB\mathbf{1}_A\Big)\Big| \stackrel{n\to\infty}{\longrightarrow} 0.$$

注意到 $E\big(\int_0^{t'} \phi^{(n)} dB\mathbf{1}_A\big) = E\big(\int_0^s \phi^{(n)} dB\mathbf{1}_A\big)$，因此

$$E\Big(\int_0^{t'} \phi dB\mathbf{1}_A\Big) = E\Big(\int_0^s \phi dB\mathbf{1}_A\Big).$$

这意味着 $E\big[\int_0^{t'} \phi_u dB_u\big|\mathcal{F}_s^B\big] = \int_0^s \phi_u dB_u$.

为证明 (3)，我们先证明如下的等式，对于 $s < t'$

$$E\Big[\int_0^s \phi dB \int_s^{t'} \phi dB\Big|\mathcal{F}_s^B\Big] = 0, \tag{44}$$

$$E\Big[\Big(\int_s^{t'} \phi_u dB_u\Big)^2 - \int_s^{t'} \phi_u^2 du\Big|\mathcal{F}_s^B\Big] = 0. \tag{45}$$

事实上

$$\begin{aligned}
(44)\text{式的左端} &= \Big(\int_0^s \phi dB\Big)E\Big[\int_s^{t'} \phi dB\Big|\mathcal{F}_s^B\Big] \\
&= \Big(\int_0^s \phi dB\Big)E\Big[\int_0^{t'} \phi dB - \int_0^s \phi dB\Big|\mathcal{F}_s^B\Big] \\
&= 0.
\end{aligned}$$

对于 (45)，当 ϕ 是可料阶梯过程时

$$\begin{aligned}
& E\Big[\Big(\int_s^{t'} \phi_u dB_u\Big)^2\Big|\mathcal{F}_s^B\Big] \\
= \ & E\Big[\Big(\sum_{k=p}^{q-1} \phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\Big)^2\Big|\mathcal{F}_{t'_p}^B\Big] \\
= \ & \sum_{k=p}^{q-1} E\Big[\big(\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\big)^2\Big|\mathcal{F}_{t'_p}^B\Big] + 2\sum_{p\le k<j\le q-1} E\Big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})(B_{t'_{j+1}} - B_{t'_j})\Big|\mathcal{F}_{t'_p}^B\Big] \\
= \ & \sum_{k=p}^{q-1} E\Big[E\big[\big(\phi_{t'_k}(B_{t'_{k+1}} - B_{t'_k})\big)^2\big|\mathcal{F}_{t'_k}^B\big]\Big|\mathcal{F}_{t'_p}^B\Big] \\
& +2\sum_{p\le k<j\le q-1} E\Big[E\big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})(B_{t'_{j+1}} - B_{t'_j})\big|\mathcal{F}_{t'_j}^B\big]\Big|\mathcal{F}_{t'_p}^B\Big] \\
= \ & \sum_{k=p}^{q-1} E\Big[\phi_{t'_k}^2(t'_{k+1} - t'_k)\Big|\mathcal{F}_{t'_p}^B\Big] + 2\sum_{p\le k<j\le q-1} E\Big[\phi_{t'_k}\phi_{t'_j}(B_{t'_{k+1}} - B_{t'_k})E\big[(B_{t'_{j+1}} - B_{t'_j})\big|\mathcal{F}_{t'_j}^B\big]\Big|\mathcal{F}_{t'_p}^B\Big] \\
= \ & E\Big[\sum_{k=p}^{q-1} \phi_{t'_k}^2(t'_{k+1} - t'_k)\Big|\mathcal{F}_{t'_p}^B\Big] \\
= \ & E\Big[\int_s^{t'} \phi_u^2 du\Big|\mathcal{F}_s^B\Big].
\end{aligned}$$

类似于性质 (2) 的证明中的逼近方式，当 $\phi \in \mathscr{L}_2^T$ 时也可以得到 (45) 成立。

注意到 $\int_0^s \phi_u^2 du$ 关于 $\mathcal{F}_s^B$ 可测，利用 (44)(45)

$$\begin{aligned}
& E\Big[\Big(\int_0^{t'} \phi_u dB_u\Big)^2 - \int_0^{t'} \phi_u^2 du \Big| \mathcal{F}_s^B\Big] \\
= \;& E\Big[\Big(\int_0^{s} \phi_u dB_u + \int_s^{t'} \phi_u dB_u\Big)^2 - \int_0^{s} \phi_u^2 du - \int_s^{t'} \phi_u^2 du \Big| \mathcal{F}_s^B\Big] \\
= \;& \Big(\int_0^{s} \phi_u dB_u\Big)^2 - \int_0^{s} \phi_u^2 du + 2E\Big[\int_0^s \phi dB \int_s^{t'} \phi dB \Big| \mathcal{F}_s^B\Big] \\
& \quad + E\Big[\Big(\int_s^{t'} \phi_u dB_u\Big)^2 - \int_s^{t'} \phi_u^2 du \Big| \mathcal{F}_s^B\Big] \\
= \;& \Big(\int_0^{s} \phi_u dB_u\Big)^2 - \int_0^{s} \phi_u^2 du.
\end{aligned}$$

□

在这一章引言的 (40) 式中

$$X_t - X_0 \approx \sum_k \sigma(t_k, X_{t_k}(\omega))(B_{t_{k+1}}(\omega) - B_{t_k}(\omega)) + \sum_k b(t_k, X_{t_k}(\omega))(t_{k+1} - t_k),$$

除了第一项关于 Brown 运动的 Itô 随机积分，我们还遇到了第二项关于随机过程轨道的 Riemann 积分。

令 $\mathscr{L}_T^1 = \big\{\Phi : (\mathcal{F}_t^B)_{t\in[0,T]}$适应的随机过程, 且 $\int_0^T E|\Phi_t(\omega)|dt < \infty\big\}$. 在 $\mathscr{L}_T^1$ 中可以定义如下的距离

$$d(\Psi, \Theta) \overset{\text{def}}{=} \int_0^T E|\Psi_t - \Theta_t| dt.$$

对于 $\mathscr{L}_T^1$ 中的元 $\Phi = \{\Phi_t,\ \ t \geq 0\}$，可以证明

$$P\Big(\omega : \ \int_0^T \Phi_t(\omega) dt \text{ 存在}\Big) = 1.$$

因此对于几乎处处的样本轨道 ω，可以按照普通的 Riemann 积分定义 $\int_0^T \Phi_t(\omega)dt$。而且还可以证明这样的积分满足

$$E\int_0^T \Phi_t(\omega) dt = \int_0^T (E\Phi_t) dt.$$

并且由 Lebesgue 积分的绝对连续性知

$$P\Big(\omega : \ \int_0^t \Phi_s(\omega) ds \text{ 关于}t\text{是连续函数}\Big) = 1.$$

5.4 Itô 公式

5.4.1 定义特殊类型的 Itô 过程及其 Itô 公式

定义 5.6. (特殊类型的)Itô 过程

设

$$\xi_t = x + \int_0^t \phi_s dB_s + \int_0^t \psi_s ds, \tag{46}$$

其中随机过程 ϕ,ψ 分别为 $\mathscr{L}_T^2$ 和 $\mathscr{L}_T^1$ 中的元。这时称 $(\xi_t)_{t\geq 0}$ 是初值为 x, 系数是 ϕ_t,ψ_t 的特殊类型的 Itô 过程.

可以将 (46) 记为 Itô(形式) 微分

$$d\xi_t = \phi_t dB_t + \psi_t dt, \quad \xi_0 = x.$$

例如，当 $B_0 = x$ 时

$$B_t^2 = x^2 + 2\int_0^t B_s dB_s + t, \text{ 或者表示为 } dB_t^2 = 2B_t dB_t + dt, \quad B_0^2 = x^2.$$

或者设 $f(x) = x^2$，则

$$f(B_t) - f(B_0) = 2\int_0^t B_s dB_s + t, \text{ 或者表示为 } df(B_t) = 2B_t dB_t + dt, \quad f(B_0) = x^2.$$

从上一节的例子我们可以知道

1. 从定义直接计算一个 Itô 积分太繁琐。
2. 与微积分中的链式法则 (牛顿-莱布尼茨公式) 不一样，B_t^2 不再是一个 Itô 积分的形式，而成为“dB_s”+“ds”积分组合，即成为一个 Itô 过程。
   (a) 牛顿-莱布尼茨公式、链式法则、一阶微分的不变性

   $$F(b) - F(a) = \int_a^b F'(x)dx,$$

   $$g\big(f(b)\big) - g\big(f(a)\big) = \int_{f(a)}^{f(b)} g'(z)dz = \int_a^b g'\big(f(x)\big)f'(x)dx = \int_a^b g'\big(f(x)\big)df(x).$$

   Riemann 积分之所以好用，就是它有很好的可计算性，即牛顿-莱布尼茨公式 (或链式法则)，由此联系了积分与微分。

   (b) Itô 随机积分是否有链式法则、一阶微分的不变性？

   $$g(B_t) - g(B_0) \stackrel{?}{=} \int_0^t g'(B_u)dB_u,$$

   答案是否定的，例如：

   $$B_t^2 - B_0^2 = 2\int_0^t B_s dB_s + t, \quad g(x) = x^2,$$

   $$g(B_t) - g(B_0) = 2\int_0^t g'(B_s)dB_s + t.$$

Itô 随机积分的牛顿-莱布尼茨公式应该是什么？首先我们回忆一下 Riemann 积分中牛顿-莱布尼茨公式是如何证明？设 $f(x) \in C[a,b]$，又设 $F \in C[a,b]$，并在 (a,b) 上 $F(x)$ 是 $f(x)$ 的原函数，$F'(x) = f(x)$，则

$$\int_a^b f(x)dx = F(b) - F(a).$$

证明： $a=x_0<x_1<\cdots<x_n=b$ 为 $[a,b]$ 的一个划分，利用微分中值定理

$$F(b)-F(a)=\sum_{k=1}^{n}[F(x_k)-F(x_{k-1})]=\sum_{k=1}^{n}f(\eta_k)\Delta x_k,$$

其中 $\Delta x_k=x_k-x_{k-1}$，$\eta_k\in[x_{k-1},x_k]$。

由于 $f(x)$ 在 $[a,b]$ 中一致连续，所以任给 $\varepsilon>0$，存在 $\delta>0$，只需 $\xi,\eta\in[a,b]$，$|\xi-\eta|<\delta$，就有

$$|f(\xi)-f(\eta)|<\frac{\varepsilon}{b-a}.$$

于是，当 $\lambda=\max\limits_{1\le k\le n}\Delta x_k<\delta$ 时，$\forall\xi_k\in[x_{k-1},x_k]$ 有

$$\Big|\sum_{k=1}^{n}f(\xi_k)\Delta x_k-(F(b)-F(a))\Big|=\Big|\sum_{k=1}^{n}[f(\xi_k)-f(\eta_k)]\Delta x_k\Big|\le\frac{\varepsilon}{b-a}\sum_{k=1}^{n}|\Delta x_k|$$

因此, $f(x)$ 在 $[a,b]$ 上可积，且 $\int_a^b f(x)dx=F(b)-F(a)$。 □

利用这个方法来计算一下 $F(B_t)-F(B_0)=?$

$$\begin{aligned}
F(B_t)-F(B_0) &= \sum_{k=0}^{n-1}\left[F(B_{t_{k+1}})-F(B_{t_k})\right]\\
&= \sum_{k=0}^{n-1}\left[F'(\xi_{t_k})(B_{t_{k+1}}-B_{t_k})\right]\qquad \text{其中}\xi_{t_k}\in[B_{t_{k+1}},B_{t_k}]\text{或}[B_{t_k},B_{t_{k+1}}])\\
&= \sum_{k=0}^{n-1}\left[F'(B_{t_k})\Delta B_{t_k}-\big(F'(B_{t_k})-F'(\xi_{t_k})\big)\Delta B_{t_k}\right]\\
&= \sum_{k=0}^{n-1}\left[F'(B_{t_k})\Delta B_{t_k}\right]-\sum_{k=0}^{n-1}\left[F'(B_{t_k})-F'(\xi_{t_k})\right]\Delta B_{t_k}\\
&= I+II.
\end{aligned}$$

由 Itô 随机积分的定义

$$I\to\int_0^t F'(B_s)dB_s.$$

类似 Riemann 积分中的证明过程，我们看

$$II=\Big|\sum_{k=0}^{n-1}\left[F'(B_{t_k})-F'(\xi_{t_k})\right]\Delta B_{t_k}\Big|\to ?.$$

利用 $F'(B_t)$ 在有界闭区间的一致连续性，即使我们有 $|F'(B_{t_k})-F'(\xi_{t_k})|<\frac{\varepsilon}{t}$ 这样的估计，也仅仅能得到

$$\Big|\sum_{k=0}^{n-1}\left[F'(B_{t_k})-F'(\xi_{t_k})\right]\Delta B_{t_k}\Big|\le\frac{\varepsilon}{t}\sum_{k=0}^{n-1}|\Delta B_{t_k}|.$$

因此单以类似 Riemann 积分的角度来考虑随机积分的牛顿-莱布尼茨公式是行不通的，关键在于 Brown 运动的轨道不是有界变差的，$\sum\limits_{k=0}^{n-1}|\Delta B_{t_k}|$ 发散。

但是我们又知道，$(B_t)_{t\geq 0}$ 是均方变差收敛的。在 Riemann 积分中我们对 F 利用泰勒公式展开到一阶就行了，但在Itô 积分中，我们需对它展开到二阶 (或三阶) 才行。

$$
\begin{aligned}
F(B_t)-F(B_0) &= \sum_{k=0}^{n-1}\left[F(B_{t_{k+1}})-F(B_{t_k})\right] \\
&= \sum_{k=0}^{n-1}\left[F'(B_{t_k})\Delta B_{t_k}+\frac{1}{2}F''(\xi_{t_k})(\Delta B_{t_k})^2\right] \\
&= \sum_{k=0}^{n-1}F'(B_{t_k})\Delta B_{t_k}+\sum_{k=0}^{n-1}\frac{1}{2}F''(\xi_{t_k})(\Delta B_{t_k})^2 \\
&= I+II.
\end{aligned}
$$

此时的关键问题是 II 是否收敛？

引理 5.7. 设 $g(x)$ 连续有界，则

$$
\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2 \xrightarrow{L^2} \int_0^t g(B_s)ds.
$$

证明：

$$
\begin{aligned}
& E\Big(\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2-\int_0^t g(B_s)ds\Big)^2 \\
\leq\ & 2E\Big(\sum_{k=0}^{n-1}g(B_{t_k})(\Delta B_{t_k})^2-\sum_{k=0}^{n-1}g(B_{t_k})\Delta t_k\Big)^2+2E\Big(\sum_{k=0}^{n-1}g(B_{t_k})\Delta t_k-\int_0^t g(B_s)ds\Big)^2 \\
=\ & 2II_1+2II_2.
\end{aligned}
$$

利用 g 的有界性，类似 Lévy 振动性质的证明 (引理5.1) 得到 $II_1\to 0$。

注意到 $\Big(\sum\limits_{k=0}^{n-1}g(B_{t_k})\mathbf{1}_{(t_k,t_{k+1}]}(t)\Big)_{t\geq 0}$ 是 $\big(g(B_t)\big)_{t\geq 0}$ 在 $\mathscr{L}_T^2$ 的可料阶梯过程近似列，由引理5.5，$II_2\to 0$。

□

引理 5.8. 设 $g(x)$ 有界一致连续，$\{t_k^{(n)}\}$ 为 $[0,t]$ 的 2^n 等分点，$\forall \xi_{t_k^{(n)}}\in[B_{t_k^{(n)}},B_{t_{k+1}^{(n)}}]$ 或 $[B_{t_{k+1}^{(n)}},B_{t_k^{(n)}}]$，

$$
P\Big(\lim_{n\to\infty}\sum_{k=0}^{2^n-1}\big(g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big)(\Delta B_{t_k^{(n)}})^2=0\Big)=1.
$$

注记 5.9. $g(\xi_{t_k^{(n)}})$ 不一定是随机变量。可以假设概率空间 $(\Omega,\mathcal{F},P)$ 是一个完备的概率空间，就不会有证明逻辑的问题了。一般在随机分析中我们都要求概率空间是完备的。

证明： 在 $[0,t]$ 内，对于几乎处处的 ω，$g(B_t(\omega))$ 一致连续，因此，可以证明 $\forall\varepsilon>0$，$\exists\delta(\omega)>0$，当 $\frac{t}{2^n}<\delta(\omega)$ 时，对于 $\forall t_k$

$$
\big|g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big|<\varepsilon,
$$

$$
\Big|\sum_{k=0}^{n-1}\big(g(\xi_{t_k^{(n)}})-g(B_{t_k^{(n)}})\big)(\Delta B_{t_k^{(n)}})^2\Big|\leq\varepsilon\sum_{k=0}^{n-1}(\Delta B_{t_k^{(n)}})^2.
$$

由引理5.2，

$$\sum_{k=0}^{n-1}(\Delta B_{t_k^{(n)}})^2 \to t \qquad \text{以概率 1 成立。}$$

因此，引理得证。

□

定理 5.1. Itô 公式 I

若 F 为二次连续可微，则

$$F(B_t)-F(B_0)=\int_0^t F'(B_s)dB_s+\frac{1}{2}\int_0^t F''(B_s)ds,$$

其形式地记为

$$dF(B_t)=F'(B_t)dB_t+\frac{1}{2}F''(B_t)dt.$$

证明： 此处只对 F', F'' 有界且一致连续证明。取 $0=t_0^{(n)}<t_1^{(n)}<\cdots<t_{2^n}^{(n)}=t$ 为 $[0,t]$ 中的 2^n 等分点

$$\begin{aligned}
&F(B_t)-F(B_0)\\
=&\sum_{k=0}^{2^n-1}[F(B_{t_k})-F(B_{t_{k-1}})]\\
=&\sum_{k=0}^{2^n-1}[F'(B_{t_k})\Delta B_{t_k}]+\frac{1}{2}\sum_{k=0}^{2^n-1}F''(B_{t_k})(\Delta B_{t_k})^2+\frac{1}{2}\sum_{k=0}^{2^n-1}\big(F''(\xi_{t_k})-F''(B_{t_k})\big)(\Delta B_{t_k})^2,
\end{aligned}$$

则存在一个子列 n'，使得 $n'\to\infty$ 上式右端以概率 1 收敛到

$$\int_0^t F'(B_s)dB_s+\frac{1}{2}\int_0^t F''(B_s)ds.$$

□

定理 5.2. Itô 公式 II

设 $F(t,x)$ 关于 t 一次连续可微，关于 x 二次连续可微，则

$$\begin{aligned}
&F(t,B_t)-F(0,B_0)\\
=&\int_0^t\frac{\partial F(s,B_s)}{\partial s}ds+\int_0^t\frac{\partial F(s,B_s)}{\partial x}dB_s+\frac{1}{2}\int_0^t\frac{\partial^2F(s,B_s)}{\partial x^2}ds\\
=&\int_0^t F_t'(s,B_s)ds+\int_0^t F_x'(s,B_s)dB_s+\frac{1}{2}\int_0^t F_{xx}''(s,B_s)ds.
\end{aligned}$$

其形式地记为

$$dF(t,B_t)=F_t'(t,B_t)dt+\frac{1}{2}F_{xx}''(t,B_t)dt+F_x'(t,B_t)dB_t.$$

例 5.4. 设 $(B_t)_{t\geq 0}$ 为标准 Brown 运动，证明 $e^{cB_t-\frac{c^2}{2}t}$ 为 $(\mathcal{F}_t^B)_{t\geq 0}$ 鞅。

解：

(1) 首先验证对于 $\forall T>0$，$e^{cB_t-\frac{c^2}{2}t}\in\mathscr{L}_T^2$.

$$E\big(e^{cB_t-\frac{c^2}{2}t}\big)^2=e^{-c^2t}Ee^{2cB_t}=e^{-c^2t}e^{2c^2t}<\infty.$$

(2) 设 $F(t,x)=e^{-\frac{c^2}{2}t+cx}$

$$
\begin{aligned}
F(t,B_t)-F(0,0) &= \int_0^t F_s'(s,B_s)ds+\int_0^t F_x'(s,B_s)dB_s+\frac{1}{2}\int_0^t F_{xx}''(s,B_s)ds \\
&= \int_0^t \big(-\frac{c^2}{2}\big)e^{-\frac{c^2}{2}t}e^{cB_s}ds+c\int_0^t e^{-\frac{c^2}{2}t}e^{cB_s}dB_s+\frac{c^2}{2}\int_0^t e^{-\frac{c^2}{2}}e^{cB_s}ds \\
&= c\int_0^t F(s,B_s)dB_s
\end{aligned}
$$

即

$$e^{cB_t-\frac{c^2}{2}t}=1+c\int_0^t e^{cB_s-\frac{c^2}{2}s}dB_s.$$

因此，由性质5.2的 (2) 知道 $(e^{cB_t-\frac{c^2}{2}t})_{t\geq 0}$ 为 $(\mathcal{F}_t^B)_{t\geq 0}$ 鞅。

□

对于一般的 Itô 过程 $X_t=x_0+\int_0^t\phi_s dB_s+\int_0^t\Psi_s ds$，$F(t,X_t)-F(0,X_0)$ 的 Itô 公式是什么？

根据前面的分析和我们积累起来的经验可以知道包含 t 和 Riemann 积分项 $\int_0^t\Psi_s ds$ 的运算规律应该还是遵从微积分中的牛顿-莱布尼茨公式。因此，问题的关键就在于如何处理包含 $\int_0^t\phi_s dB_s$ 的项。我们将问题简化成若 $X_t=x_0+\int_0^t\phi_s dB_s+\int_0^t\Psi_s ds$，$F(X_t)-F(x_0)$ 应该等于什么？

虽然 $dX_t=\phi_t dB_t+\Psi_t dt$ 只是 Itô 过程 $(X_t)_{t\geq 0}$ 的"形式的微分"，并不具有严格的数学定义，但从这一章引言的一些介绍，我们可以直观的把它理解成当 $\Delta t>0$ 很小的时候，$dX_t=\phi_t dB_t+\Psi_t dt$ 近似地是

$$\Delta X_t=X(t+\Delta t)-X(t)\approx\phi_t(B_{t+\Delta t}-B_t)+\Psi_t\Delta t=\phi_t\Delta B_t+\Psi_t\Delta t.$$

当划分 $0=t_0<t_1<\cdots<t_n=t$ 充分细时，我们有如下近似的看法

$$
\begin{aligned}
\int_0^t g(X_s)\phi_s dB_s+\int_0^t g(X_s)\Psi_s ds &= \int_0^t g(X_s)dX_s \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})\Delta X_{t_i} \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})(\phi_{t_i}\Delta B_{t_i}+\Psi_{t_i}\Delta t_i).
\end{aligned}
$$

当划分充分细时，对于任意的 $\varepsilon>0$，

$$|\Delta B_{t_i}(\omega)|<\frac{\varepsilon}{t},\ \ \sum_{i=0}^{n-1}\big|\Delta B_{t_i}\Delta t_i\big|<\varepsilon\sum_{i=0}^{n-1}\Delta t_i=\varepsilon.$$

对于有界的一致连续函数 $g(x)$，ϕ_t 和 Ψ_t 有界，我们可以近似地得到

$$\sum_{i=0}^{n-1}\big|\Delta B_{t_i}\Delta t_i\big|<\varepsilon\sum_{i=0}^{n-1}\Delta t_i=\varepsilon.$$

$$\begin{aligned}
\int_0^t g(X_s)(dX_s)^2 &\approx \sum_{i=0}^{n-1} g(X_{t_i})(\Delta X_{t_i})^2 \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})(\phi_{t_i}\Delta B_{t_i} + \Psi_{t_i}\Delta t_i)^2 \\
&= \sum_{i=0}^{n-1} g(X_{t_i})\big[\phi_{t_i}^2(\Delta B_{t_i})^2 + 2\phi_{t_i}\Psi_{t_i}\Delta B_{t_i}\Delta t_i + \Psi_{t_i}^2(\Delta t_i)^2\big] \\
&\approx \sum_{i=0}^{n-1} g(X_{t_i})\phi_{t_i}^2(\Delta B_{t_i})^2 + O(\varepsilon) + O(\max_i\{(\Delta t_i)^2\}) \\
&\approx \int_0^t g(X_s)\phi_s^2 ds \quad \text{类似于引理5.7的证明并忽略掉无穷小量}
\end{aligned}$$

$$\begin{aligned}
& F(X_t) - F(x_0) \\
= & \sum_{i=0}^{n-1} \big(F(X_{t_{i+1}}) - F(X_{t_i})\big) \\
\overset{\text{泰勒展开}}{=} & \sum_{i=0}^{n-1} \Big(F'(X_{t_i})\Delta X_{t_i} + \frac{1}{2}F''(X_{t_i})(\Delta X_{t_i})^2 + \frac{1}{2}\big(F''(\xi_{t_i}) - F''(X_{t_i})\big)(\Delta X_{t_i})^2\Big) \\
\approx & \sum_{i=0}^{n-1} \Big(F'(X_{t_i})\Delta X_{t_i} + \frac{1}{2}F''(X_{t_i})(\Delta X_{t_i})^2\Big) \quad \text{类似于引理5.8的证明} \\
\approx & \int_0^t F'(X_s)dX_s + \frac{1}{2}\int_0^t F''(X_s)(dX_s)^2 \\
= & \int_0^t F'(X_s)\phi_s dB_s + \int_0^t F'(X_s)\phi_s ds + \frac{1}{2}\int_0^t F''(X_s)\phi_s^2 ds.
\end{aligned}$$

将上述步骤严格化，并做一些必要的技术上的处理，就有下面的定理。

定理 5.3. Itô 公式 III

设 $X_t = x_0 + \int_0^t \phi_s dB_s + \int_0^t \Psi_s ds$, $dX_t = \phi_t dB_t + \Psi_s ds$, $F(t,x)$ 关于 t 一次连续可微，关于 x 二次连续可微，则

$$F(t,X_t) - F(0,x_0) = \int_0^t \frac{\partial F}{\partial s}(s,X_s)ds + \int_0^t \frac{\partial F}{\partial x}(s,X_s)dX_s + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial x^2}(s,X_s)(dX_s)^2$$

写成“形式微分”

$$dF(t,X_t) = \frac{\partial F}{\partial t}(t,X_t)dt + \frac{\partial F}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X_t)(dX_t)^2,$$

其中 dt, dB_t 的平方和乘积由乘法表

乘积	dt	dB_t
dt	0	0
dB_t	0	dt

给出，即

$$dF(t,X_t) = \Big[\frac{\partial F}{\partial t}(t,X_t) + \Psi_t\frac{\partial F}{\partial x}(t,X_t) + \frac{1}{2}\phi_t^2\frac{\partial^2 F}{\partial x^2}(t,X_t)\Big]dt + \phi_t\frac{\partial F}{\partial x}(t,X_t)dB_t$$

写成真正的积分形式

$$F(t,X_t)-F(0,x_0)=\int_0^t\big[\frac{\partial F}{\partial s}(s,X_s)+\Psi_t\frac{\partial F}{\partial x}(s,X_s)+\frac{1}{2}\phi_s^2\frac{\partial^2 F}{\partial x^2}(s,X_s)\big]ds+\int_0^t\phi_s\frac{\partial F}{\partial x}(s,X_s)dB_s.$$

其中

$$(dX_t)^2=\big(\phi_t dB_t+\Psi_t dt\big)^2=(\phi_t dB_t)^2+2\phi_t\Psi_t dB_t dt+\Psi_t^2(dt)^2=\phi_t^2 dt.$$

例 5.5. 若 $(\phi_t)_{t\geq 0}$ 为关于 $(\mathcal{F}_t^B)_{t\geq 0}$ 适应的有界随机过程，证明 $\xi_t=\exp\{\int_0^t\phi_s dB_s-\frac{1}{2}\int_0^t\phi_s^2 ds\}$ 是关于 $(\mathcal{F}_t^B)_{t\geq 0}$ 的鞅。

证明： 利用解例子5.4的第一步以及可料阶梯过程的逼近程序，可以证明 $\{\xi_t\}\in\mathscr{L}_T^2$。设 $F(x)=e^x$,

$$X_t=\int_0^t\phi_s dB_s-\frac{1}{2}\int_0^t\phi_s^2 ds,\ \text{即 } dX_t=\phi_t dB_t-\frac{1}{2}\phi_t^2 dt.$$

利用 Itô 公式

$$\begin{aligned}F(X_t)-F(0) &= \int_0^t F'(X_s)dX_s+\frac{1}{2}\int_0^t F''(X_s)(dX_s)^2\\ &= \int_0^t F(X_s)\phi_s dB_s-\frac{1}{2}\int_0^t F(X_s)\phi_s^2 dt+\frac{1}{2}\int_0^t F(X_s)\phi_s^2 dt\\ &= \int_0^t F(X_s)\phi_s dB_s,\end{aligned}$$

所以由性质5.2的 (2) 知道 $(\exp\{\int_0^t\phi_s dB_s-\frac{1}{2}\int_0^t\phi_s^2 ds)_{t\geq 0}$ 为 $(\mathcal{F}_t^B)_{t\geq 0}$ 鞅。 □

注记 5.10. $\big|\int_0^t\phi_s dB_s\big|\leq\int_0^t|\phi_s|dB_s$ 是错误的。

5.4.2　一般的 Itô 过程

前面我们对 $\mathscr{L}_T^2$ 中的过程定义了 Itô 随机积分并建立了 Itô 公式，但问题是对被积的过程限制太多。在 Riemann 积分，例如，若在 $[a,b]$，当 $x\to b$ 时，$f(x)\uparrow\infty$，$b_n<b$，$f(x)$ 在 $[a,b_n]$ 内有界且 Riemann 可积，我们如何来定义 $\int_a^b f(x)dx$? 事实上定义 $\int_0^t f(x)dx=\lim\limits_{b_n\uparrow b}\int_a^{b_n}f(x)dx$。

类似的想法

$$\mathscr{L}_T^{2,\mathrm{loc}}\overset{\mathrm{def}}{=}\big\{(\mathcal{F}_t^B)_{t\geq 0}\text{适应的随机过程}\phi_t,\text{ 且}P\big(\omega:\ \int_0^T(\phi_t)^2(\omega)dt<\infty\big)=1\big\}$$

可以证明

(1) 对于任意 $\phi\in\mathscr{L}_T^{2,\mathrm{loc}}$ 存在 $(\mathcal{F}_t^B)_{t\geq 0}$ 停时序列 $\tau_1\leq\cdots\leq\tau_n\uparrow\infty$，使得对于固定的 n，随机过程 $\phi_t^{(n)}\overset{\mathrm{def}}{=}\phi_{t\wedge\tau_n}\in\mathscr{L}_T^2$.

(2) 由此，当 $t\leq\tau_n$ 时可以定义 $\int_0^t\phi_s dB_s$:

$$\Big(\int_0^t\phi_s dB_s\Big)(\omega)\overset{\mathrm{def}}{=}\Big(\int_0^t\phi_s^{(n)}dB_s\Big)(\omega).$$

进一步，

$$\int_0^t\phi_s dB_s\overset{\mathrm{def}}{=}\lim_{n\to\infty}\int_0^t\phi_s^{(n)}dB_s.$$

(3) $(\xi_t \stackrel{\text{def}}{=} \int_0^t \phi_s dB_s)_{t\geq 0}$ 是 $(\mathcal{F}_t^B)$ 局部鞅，即存在 $(\mathcal{F}_t^B)_{t\geq 0}$ 停时序列 $\tau_1 \leq \cdots \leq \tau_n \uparrow \infty$，使得对于固定的 n，$(\zeta_t^{(n)} \stackrel{\text{def}}{=} \xi_{t\wedge\tau_n})_{t\geq 0}$ 是 $(\mathcal{F}_t^B)_{t\geq 0}$ 鞅。但局部鞅不一定是鞅。

(4) 设 $\mathscr{L}_T^{1,\text{loc}} \stackrel{\text{def}}{=} \big\{(\mathcal{F}_t^B)_{t\geq 0}\text{适应的随机过程}\Psi_t \;\; P\big(\omega: \; \int_0^T |\Psi_t|dt < +\infty\big) = 1\big\}$，则对于任意 $\Psi \in \mathscr{L}_T^{1,\text{loc}}$ 积分 $\int_0^t \Psi_s(\omega)ds$ 几乎处处有定义。

定义 5.7. 一般的 Itô 过程

设 $\zeta_t = x + \int_0^t \phi_s dB_s + \int_0^t \Psi_s ds$，其中 ϕ_s, Ψ_s 分别是 $\mathscr{L}_T^{2,\text{loc}}$ $\mathscr{L}_T^{1,\text{loc}}$ 中的元，则 ξ_t 称为初值为 x 系数为 ϕ_t，Ψ_t 的一般的 Itô 过程，简称 Itô 过程。

对 Itô 过程 $\{\xi_t,\ t \geq 0\}$ 的 Itô 积分实质上为

$$\int_0^t \Theta_t d\xi_t = \int_0^t \Theta_t \phi_t dB_t + \int_0^t \Theta_t \Psi_t dt.$$

定理 5.4. Itô 公式 Ⅱ, Ⅲ 对一般的 Itô 过程也成立。

5.4.3 多维 Brown 运动的 Itô 随机积分和 Itô 公式

设 $\mathbf{B}_t = \begin{pmatrix} B_t^{(1)} \\ \vdots \\ B_t^{(m)} \end{pmatrix}$ 为一个 m-维 Brown 运动，$\mathcal{F}_t^{\mathbf{B}} = \sigma(\mathbf{B}_s,\ \ s \leq t)$.

$$\mathscr{L}_T^2 = \big\{(\phi_t)_{t\geq 0}\text{为}(\mathcal{F}_t^{\mathbf{B}})_{t\geq 0}\text{适应的随机过程, 且}\int_0^T E\phi_s^2 ds < \infty\big\}.$$

$$\mathscr{L}_T^1 = \big\{(\phi_t)_{t\geq 0}\text{为}(\mathcal{F}_t^{\mathbf{B}})_{t\geq 0}\text{适应的随机过程, 且}\int_0^T E|\phi_s| ds < \infty\big) = 1\big\}.$$

$$\mathscr{L}_T^{2,\text{loc}} = \big\{(\phi_t)_{t\geq 0}\text{为}(\mathcal{F}_t^{\mathbf{B}})_{t\geq 0}\text{适应的随机过程, 且}P\big(\omega: \int_0^T \phi_s^2 ds < \infty\big) = 1\big\}.$$

$$\mathscr{L}_T^{1,\text{loc}} = \big\{(\phi_t)_{t\geq 0}\text{为}(\mathcal{F}_t^{\mathbf{B}})_{t\geq 0}\text{适应的随机过程, 且}P\big(\omega: \int_0^T |\phi_s| ds < +\infty\big) = 1\big\}.$$

类似地可以定义

$$\int_0^T \phi_t dB_t^{(i)} \qquad i = 1, \cdots, m.$$

定义 5.8. (多维 Brown 运动的 Itô 过程)

设

$$\xi_t = x + \int_0^t \phi_s^T d\mathbf{B}_s + \int_0^t \Psi_s ds,$$

其中 $\phi_t = \begin{pmatrix} \phi_t^{(1)} \\ \vdots \\ \phi_t^{(m)} \end{pmatrix}$, $\xi_t = x + \sum\limits_{k=1}^m \int_0^t \phi_s^{(k)} dB_s^{(k)} + \int_0^t \Psi_s ds$, 而 $(\phi_t^{(i)})_{t\geq 0}$, $(\Psi_t)_{t\geq 0}$ 分别是 $\mathscr{L}_T^{2,\text{loc}}$, $\mathscr{L}_T^{1,\text{loc}}$ 中的元，那么称 $(\xi_t)_{t\geq 0}$ 为初值为 x 的多维 Brown 运动的 Itô 过程。

引理 5.9. 设 $(B_t^{(1)})$, $(B_t^{(2)})$ 是两个独立的 Brown 运动，对于 $[0,t]$ 中的一个划分 $0 = t_0 < \cdots < t_n = t$，

$$T_n = \sum_{k=0}^{n-1} \big(B_{t_{k+1}}^{(1)} - B_{t_k}^{(1)}\big)\big(B_{t_{k+1}}^{(2)} - B_{t_k}^{(2)}\big),$$

则 $E(T_n) = 0$, $E(T_n)^2 \to 0$.

注记 5.11. 上面的引理形式地看成

$$\sum_{k=0}^{n}\Delta B_{t_k}^{(1)}\Delta B_{t_k}^{(2)}\approx 0,\quad dB_t^{(1)}dB_t^{(2)}=0.$$

性质 5.3. 设 $\phi,\psi\in\mathscr{L}_T^2$, $E\big(\int_0^T\phi dB^{(i)}\int_0^t\psi dB^{(j)}\big)=\delta_{ij}E\big(\int_0^t\phi\psi ds\big)$，$\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$ 。

定理 5.5. Itô 公式 Ⅳ

$$d\xi_t=\Phi_t^T d\mathbf{B}_t+\Psi_t dt=\sum_{i=1}^{m}\phi_t^{(i)}dB_t^{(i)}+\Psi_t dt$$

则

$$dF(t,\xi_t)=\frac{\partial F(t,\xi_t)}{\partial t}dt+\frac{\partial F(t,\xi_t)}{\partial x}d\xi_t+\frac{1}{2}\frac{\partial^2 F(t,\xi_t)}{\partial x^2}(d\xi_t)^2$$

即

$$F(t,\xi_t)-F(0,\xi_0)=\int_0^t\frac{\partial F(s,\xi_s)}{\partial s}ds+\int_0^t\frac{\partial F(s,\xi_s)}{\partial x}d\xi_s+\frac{1}{2}\int_0^t\frac{\partial^2 F(s,\xi_s)}{\partial x^2}(d\xi_s)^2$$

其中 $dt, dB_t^{(i)}$ 由下面的乘法表给出

	dt	$dB_t^{(i)}$
dt	0	0
$dB_t^{(i)}$	0	$\delta_{ij}dt$

$$\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases},$$

即

$$\begin{aligned}&F(t,\xi_t)-F(0,\xi_0)\\ =\;&\int_0^t\Big[\frac{\partial F(s,\xi_s)}{\partial s}+\frac{\partial F(s,\xi_s)}{\partial x}\Psi_s+\frac{1}{2}\frac{\partial^2 F(s,\xi_s)}{\partial x^2}\sum_{i=1}^{m}(\phi_s^{(i)})^2\Big]ds+\sum_{i=1}^{m}\int_0^t\frac{\partial F(s,\xi_s)}{\partial x}\phi_s^{(i)}dB_s^{(i)}.\end{aligned}$$

推论 5.3. 乘积公式

ξ_t,η_t 都是多维 Brown 运动 $\{\mathbf{B}_t,\ t\geq 0\}$ 的 Itô 过程，则

$$d(\xi_t\eta_t)=\xi_t d\eta_t+\eta_t d\xi_t+(d\xi_t)(d\eta_t).$$

进一步，若 $\xi_t=\big(\xi_t^{(1)},\xi_t^{(2)},\cdots,\xi_t^{(d)}\big)^T$，$(\xi_t^{(i)})$, $i\leq d$ 都是同一个 m 维 Brown 运动 $\mathbf{B}_t=\begin{pmatrix}B_t^{(1)}\\ \vdots\\ B_t^{(m)}\end{pmatrix}$ 的 Itô 过程，即

$$d\xi_t^{(i)}=(\Phi_t^{(i)})^T d\mathbf{B}_t+\Psi_t^{(i)}dt,\ i\leq d,\ \Psi=\begin{pmatrix}\Psi_t^{(1)}\\ \vdots\\ \Psi_t^{(d)}\end{pmatrix}.$$

写成矩阵的形式

$$\Phi^{(i)}=\begin{pmatrix}\phi^{i1}\\ \vdots\\ \phi^{im}\end{pmatrix},1\le i\le d;\ \Psi=\begin{pmatrix}\Psi_t^{(1)}\\ \vdots\\ \Psi_t^{(d)}\end{pmatrix},$$

$$\Phi=(\Phi^{(1)},\cdots,\Phi^{(d)})=\begin{pmatrix}\phi_t^{11}&\phi_t^{21}&\cdots&\phi_t^{d1}\\ \vdots&\vdots&\vdots&\vdots\\ \phi_t^{1m}&\phi_t^{2m}&\cdots&\phi_t^{dm}\end{pmatrix},$$

$$d\xi=\Phi_t^T\,d\mathbf{B}_t+\Psi\,dt,$$

$$\begin{pmatrix}d\xi_t^{(1)}\\ \vdots\\ d\xi_t^{(d)}\end{pmatrix}=\begin{pmatrix}\phi_t^{11}&\phi_t^{12}&\cdots&\phi_t^{1m}\\ \vdots&\vdots&\vdots&\vdots\\ \phi_t^{d1}&\phi_t^{d2}&\cdots&\phi_t^{dm}\end{pmatrix}\begin{pmatrix}dB_t^{(1)}\\ \vdots\\ dB_t^{(m)}\end{pmatrix}+\begin{pmatrix}\Psi_t^{(1)}\\ \vdots\\ \Psi_t^{(d)}\end{pmatrix}dt.$$

定理 5.6. Itô 公式 V

$F(x_1,\cdots,x_d)$ 为一个二次连续可微的函数，则

$$\begin{aligned}dF(\xi_t^{(1)},\cdots,\xi_t^{(d)})&=\sum_{k=1}^d\frac{\partial}{\partial x_k}F(\xi_t^{(1)},\cdots,\xi_t^{(d)})d\xi_t^{(k)}+\frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\frac{\partial^2}{\partial x_i\partial x_j}F(\xi_t^{(1)},\cdots,\xi_t^{(d)})d\xi_t^{(i)}d\xi_t^{(j)},\\ d\xi_t^{(i)}d\xi_t^{(j)}&=a_{ij}dt,\\ (a_{ij})_{i,j\le d}&=(\Phi^T\Phi)_{i,j\le d}=(\phi^{ij})^T_{\substack{i\le m\\ j\le d}}(\phi^{ij})_{\substack{i\le m\\ j\le d}}.\end{aligned}$$

5.5 Stratonovich-Fisk 对称积分

5.5.1 背景

5.5.2 定义及性质

定义 5.9. (Stratonovich-Fisk 随机积分)

设 $(\xi_t)_{t\ge0}$ 是特殊类型的 Itô 过程, $\xi_t=\xi_0+\int_0^t\phi_sdB_s+\int_0^t\psi_sds$, 对于 $0=t_0^{(n)}<\cdots<t_{N^{(n)}}^{(n)}=t$, $\lambda_n=\max\Delta t_k^{(n)}$, Sratonovich-Fisk 对称积分定义为

$$\int_0^t\xi_s\circ dB_s=L^2(\Omega)-\lim_{\lambda_n\to0}\sum_{k=0}^{N^{(n)}-1}\frac{\xi_{t_k^{(n)}}+\xi_{t_{k+1}^{(n)}}}{2}\Delta B_{t_k^{(n)}};\tag{47}$$

对于一般的 Itô 过程 $(\xi_t)_{t\ge0}$, Stratonovich-Fisk 对称积分定义为

$$\int_0^t\xi_s\circ dB_s=(P)-\lim_{\lambda_n\to0}\sum_{k=0}^{N^{(n)}-1}\frac{\xi_{t_k^{(n)}}+\xi_{t_{k+1}^{(n)}}}{2}\Delta B_{t_k^{(n)}}.\tag{48}$$

性质 5.4. $\int_0^t\xi_s\circ dB_s=\int_0^t\xi_sdB_s+\frac12\int_0^t\phi_sds$, 即

$$\begin{aligned}\xi_t\circ dB_t&=\xi_tdB_t+\frac12d\xi_tdB_t\\ &=\xi_tdB_t+\frac12(\phi_tdB_t)dB_t\\ &=\xi_tdB_t+\frac12\phi_tdt.\end{aligned}$$

例 5.6.

$$\begin{aligned}\int_0^t B_s \circ dB_s &= \int_0^t B_s dB_s + \frac{1}{2}\int_0^t 1 \cdot ds \\ &= \int_0^t B_s dB_s + \frac{t}{2} = B_t^2 - B_0^2.\end{aligned}$$

定义 5.10. 设 $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ 是 Itô 过程

$$X_t = x_0 + \int_0^t \phi_s^X dB_s + \int_0^t \psi_s^X ds,$$

$$Y_t = y_0 + \int_0^t \phi_s^Y dB_s + \int_0^t \psi_s^Y ds.$$

$$\begin{aligned}\int_0^t Y_s \circ dX_s &= (P) - \lim_{\lambda_n \to 0} \sum_{k=0}^{N^{(n)}-1} \frac{Y_{t_k^{(n)}} + Y_{t_{k+1}^{(n)}}}{2} \big(X_{t_{k+1}^{(n)}} - X_{t_k^{(n)}}\big) \\ &= \int_0^t Y dX + \frac{1}{2}\int_0^t dY dX \\ &= \int_0^t Y_s \phi_s^X dB_s + \int_0^t Y_s \psi_s^X ds + \frac{1}{2}\int_0^t \phi_s^X \phi_s^Y ds.\end{aligned}$$

$$\begin{aligned}Y_t \circ dX_t &= Y_t dX_t + \frac{1}{2} dY dX \\ &= Y_t \phi_t^X dB_t + Y \psi_t^X ds + \frac{1}{2}\phi_s^X \phi_s^Y ds.\end{aligned}$$

定理 5.7. (Itô 公式 VI) 设 $\xi_t = \xi_0 + \int_0^t \phi_s dB_s + \int_0^t \psi_s ds$, 即 $d\xi_t = \phi_t dB_t + \psi_t dt$. 若 f 三次连续可微, 则

$$df(\xi_t) = f'(\xi_t) \circ d\xi_t,$$

即

$$f(\xi_t) - f(\xi_0) = \int_0^t f'(\xi_s) \circ d\xi_s = \int_0^t f'(\xi_s) \circ (\phi_s dB_s) + \int_0^t f'(\xi_s)\psi_s ds.$$

若 $f(t,x)$ 对 x 三次连续可微, 对 t 一次连续可微, 则

$$df(t,\xi_t) = f_t' dt + f_x' \circ d\xi_t.$$

6 第六章　随机微分方程和扩散过程

如同常微分方程、偏微分方程一样，随机微分方程 (stochastic differential equation, SDE) 可以描述相当广泛的一类物理现象、经济现象、生物现象。

随机微分方程的一般形式为

$$d\xi_t = b(t,\xi)dt + \sigma(t,\xi)dB_t, \tag{49}$$

即

$$\xi_t = \xi_0 + \int_0^t b(s,\xi(\omega))ds + \int_0^t \sigma(s,\xi(\omega))dB_s.$$

其实这是一个随机积分方程，随机微分方程只是约定俗成的叫法而已。其中 b, σ 满足对于任意的 $s>0$，$b(s,\xi(\omega))$, $\sigma(s,\xi(\omega))$ 是关于 $\mathcal{F}_s^B$ 可测的随机变量，即 $(b(s,\xi(\omega)))_{s\geq 0}$，$(\sigma(s,\xi(\omega)))_{s\geq 0}$ 是 $(\mathcal{F}_s^B)_{s\geq 0}$ 适应的。也就是说 $b(s,\xi(\omega))$，$\sigma(s,\xi(\omega))$ 是由 $\xi(\omega)$ 在 $[0,s]$ 上所有的值决定的，或者说 $b(s,\xi(\omega))$, $\sigma(s,\xi(\omega))$ 是 $\xi(\cdot,\omega), \cdot\in[0,s]$ 这个函数的泛函。

我们这里只研究如下形式的 SDE

$$d\xi_t = b(t,\xi_t)dt + \sigma(t,\xi_t)dB_t$$

即

$$\xi_t = \xi_0 + \int_0^t b(s,\xi_s(\omega))ds + \int_0^t \sigma(s,\xi_s(\omega))dB_s$$

其中 $b(s,x)$, $\sigma(s,x)$, $\mathbb{R}^+\times\mathbb{R}\to\mathbb{R}$ 的 Borel 可测函数，也就是说 $b(s,\xi_s)$，$\sigma(s,\xi_s)$ 的值由仅由 ξ_s 决定。

6.1 随机微分方程的例子和解的存在唯一性

6.1.1 几个重要的例子

例 6.1. Ornstein-Uhlenbeck 方程，Langevin 方程

Langevin 等人用 Einstein 的 Brown 运动数学模型描述 R.Brown 观察到的液体中微小花粉粒子的随机运动，效果并不十分理想。1908 年，Langevin 等人对模型进行了改进。

一个粒子在液体中运动，其速度为 v，则由 Stokes 定律，它受到液体对它的粘滞力为 $-\alpha v$，设 m 为粒子的质量，则由牛顿运动第二定律

$$m\dot{v} = -\alpha v. \tag{50}$$

在粒子质量很大，但还不是极大时 (例如，花粉粒子的质量是液体分子质量的 20 万倍，2×10^5 倍)，除了宏观的粘滞力外，液体分子对粒子杂乱的作用效果可以十分清楚的表现出来。因此，上述方程可以改写成

$$m\dot{v} = -\alpha v + F(t), \tag{51}$$

$F(t)$ 表示分子杂乱碰撞的力。上式两边同除以 m

$$\dot{v} = -\gamma v + \Gamma(t), \quad \gamma = \frac{\alpha}{m}, \quad \Gamma(t) = \frac{F(t)}{m},$$

其中 γ 表示单位质量的阻尼系数，$\Gamma(t)$ 表示单位质量的花粉粒子受到液体分子碰撞的“涨落力”，被称为 Langevin 力。上述方程被称为 Langevin 方程，也被称为 Ornstein-Uhlenbeck 方程。可以形式地记为

$$\frac{dv}{dt} = -\gamma v + \Gamma(t), \text{ 或 } dv = -\gamma v dt + \Gamma(t)dt.$$

可以根据不同的物理系统假定 $\Gamma(t)$ 不同的统计性质，再根据通过实验检验修正这些统计假设。Langevin 当年假设 $\Gamma(t)$ 是所谓的白噪声，即满足如下的统计性质：

(1) 由于 $\Gamma(t)$ 为涨落力，可以认为它的统计平均值为 0, 即 $E\Gamma(t) = 0$ (或用统计物理的记号 $\langle\Gamma(t)\rangle = 0$).

(2) 同时因为宏观观测时间间隔远远大于微观分子对花粉例子的碰撞时间，所以不同时刻的 $\Gamma(t)$ 可以近似认为是相互独立的，$E(\Gamma(t)\Gamma(t')) = 2D\delta(t-t')$ (或 $\langle\Gamma(t)\Gamma(t')\rangle = 2D\delta_0 t - t')$, 其中 $\delta_0(x)$ 为 Dirac 函数。

   从概率论的角度上式可以解释为在任意小的两个不相交区间内，液体分子对花粉粒子碰撞的涨落力是不相关，在任意小的区间内涨落力的方差为 $2D$。

(3) 假定 $\Gamma(t)$ 有 Gauss 性。

在数学上我们可以认为“$\Gamma(t)dt$”就是 dB_t。而 $\Gamma(t)$ 被称为“白噪声”，$\Gamma(t)$ 有时形式上记为“$\frac{dB_t}{dt}$”或“$\dot{B}_t$”。因此，Langevin 方程可以写成一个标准的随机微分方程 (取 $2D = 1$)

$$\begin{aligned} dv_t &= -\gamma v_t dt + dB_t \qquad \text{或} \\ v_t - v_0 &= -\int_0^t \gamma v_s ds + B_t, \quad (\text{设} B_0 = 0). \end{aligned} \tag{52}$$

下面考虑如何解这个方程，形式的

$$\frac{dv_t}{dt} = -\gamma v_t + \dot{B}_t, \qquad dv_t = -\gamma v_t dt + dB_t.$$

物理的解法：借助常微分方程 $\frac{dv_t}{dt} = -\gamma v_t + f(t)$ 的解法。

(1) 先考虑齐次方程

$$\frac{dv_t}{dt} = -\gamma v_t \quad \Rightarrow \quad v_t = c_0 e^{-\gamma t}.$$

(2) 再考虑非齐次方程 $\frac{dv_t}{dt} = -\gamma v_t + f(t)$, 或 $dv_t + \gamma v_t dt = f(t)dt$. 两端乘以 $e^{\gamma t}$

$$e^{\gamma t} dv_t + e^{\gamma t}\gamma v_t dt = e^{\gamma t} f(t)dt,$$

即

$$d(e^{\gamma t} v_t) = e^{\gamma t} f(t)dt.$$

$$\begin{aligned} e^{\gamma t} v_t &= \int_0^t e^{\gamma s} f(s)ds + v_0, \\ v_t &= \int_0^t e^{\gamma(s-t)} f(s)ds + v_0 e^{-\gamma t}. \end{aligned}$$

(3) 用 $\dot{B}_t$ 替代 $f(t)$ 或 $f(t)dt$ 替代 dB_t 得

$$v_t = v_0 e^{-\gamma t} + \int_0^t e^{\gamma(s-t)} dB_s = v_0 e^{-\gamma t} + e^{-\gamma t} \int_0^t e^{\gamma s} dB_s. \tag{53}$$

数学严格的解法：由 Itô 公式

$$\begin{aligned} d(e^{\gamma t} v_t) &= e^{\gamma t} dv_t + v_t de^{\gamma t} \\ &= -\gamma e^{\gamma t} v_t dt + e^{\gamma t} dB_t + \gamma e^{\gamma t} v_t dt \\ &= e^{\gamma t} dB_t. \end{aligned}$$

写出积分形式即得 (53).

注记 6.1. 在上面的例子为什么把 $\Gamma(t)$ 称为白噪声呢?$\Gamma(t)$ 的时间相关函数是 $D\delta_0(t-t')$，它的功率谱密度是常数，用数学的话说就是 δ_0 的 Fourier 变换是常函数 1 。也就是说 $2D\delta_0(t-t')$ 在各个频率上的分量的大小都一样，这与白光的性质类似，所以当 $\Gamma(t)$ 的时间相关函数是 $2D\delta_0(t-t')$ 时，$\Gamma(t)$ 被称为白噪声。

例 6.2. 在上面的例子中，如果系数都是非随机的数值函数，方程成为

$$dX_t = F(t) X_t dt + C(t) dB_t, \tag{54}$$

其中假设系数在任意有限区间内有界，求解此方程。

解：类似于上面例子，先考虑齐次方程

$$dx_t - F(t) x_t dt = 0, \quad \text{解得 } x_t = x_0 e^{\int_0^t F(s)ds}.$$

将 $e^{-\int_0^t F(s)ds}$ 作为恰当因子乘在方程 (54) 两边并利用 Itô 公式，

$$\begin{aligned} d\big(X_t e^{-\int_0^t F(s)ds}\big) &= e^{-\int_0^t F(s)ds} dX_t - X_t F(t) e^{-\int_0^t F(s)ds} dt \\ &= C(t) e^{-\int_0^t F(s)ds} dB_t. \end{aligned}$$

因此，

$$\begin{aligned} X_t &= X_0 e^{\int_0^t F(s)ds} + e^{\int_0^t F(s)ds} \int_0^t C(u) e^{-\int_0^u F(s)ds} dB_u \\ &= X_0 e^{\int_0^t F(s)ds} + \int_0^t C(u) e^{-\int_u^t F(s)ds} dB_u. \end{aligned}$$

□

例 6.3. 随机调和振子

调和振子的运动方程为

$$\begin{cases} \frac{d^2\xi_t}{dt^2} = -\lambda^2 \xi_t \\ \xi_0 = x, \quad \frac{d\xi_t}{dt}\big|_{t=0} = v_0 \end{cases} \tag{55}$$

若振子受到白噪声的随机干扰，形式上

$$\begin{cases} \frac{d^2\xi_t}{dt^2} = -\lambda^2 \xi_t + \dot{B}_t \\ \xi_0 = x, \quad \frac{d\xi_t}{dt}\big|_{t=0} = v_0 \end{cases}. \tag{56}$$

随机微分方程实质上是随机积分方程，因此随机微分方程的一般形式 (49) 中是没有 $\frac{d^2\xi_t}{dt^2}$ 或 $d^2\xi_t$ 这样的项。那么 (56) 这个形式方程对应的随机微分方程应该是什么？

在常微分方程理论中，一个高阶的常微分方程总可以化成一阶的方程组，例如在 (55) 中，设 $\frac{d\xi_t}{dt}=v_t$, 则 (55) 等价于

$$\begin{cases} \frac{dv_t}{dt}=-\lambda^2\xi_t \\ \frac{d\xi_t}{dt}=v_t \\ \xi_0=x \\ v_t|_{t=0}=v_0 \end{cases}.$$

因此，我们可以把 (56) 转化成一个有明确数学含义的随机微分方程组，

$$\begin{cases} dv_t=-\lambda^2\xi_t dt+dB_t \\ d\xi_t=v_t dt \\ \xi_0=x \\ v_t|_{t=0}=v_0 \end{cases}. \tag{57}$$

或者写成随机积分的形式

$$\begin{cases} v_t=v_0-\lambda^2\int_0^t\xi_s ds+B_t & (B_t\text{设为标准布朗运动}) \\ d\xi_t=v_t dt,\ \text{或}\ \ \xi_t=x+\int_0^t v_s ds \end{cases}. \tag{58}$$

接下来的问题是如何解随机微分方程 (57) 或 (58)。我们还是先参照一下在常微分方程理论。若 (55) 中有一个确定性的强迫力

$$\begin{cases} \frac{d^2\xi_t}{dt^2}=-\lambda^2\xi_t+f(t) \\ \xi_0=x,\quad \frac{d\xi_t}{dt}\big|_{t=0}=v_0 \end{cases}.$$

由常微分方程理论知道，

$$\xi_t=x\cos\lambda t+\frac{v_0}{\lambda}\sin\lambda t+\frac{1}{\lambda}\int_0^t\sin\lambda(t-s)f(s)ds.$$

想象存在光滑的 $f_n(s)$ 满足 $f_n(s)ds\to\dot{B}_s ds=dB_s$，则形式地

$$\xi_t=x\cos\lambda t+\frac{v_0}{\lambda}\sin\lambda t+\frac{1}{\lambda}\int_0^t\sin\lambda(t-s)dB_s. \tag{59}$$

问题是 (59) 中定义的 ξ_t 真的是 (58) 的解吗？v_t 又是什么？下面验证 (59) 满足方程 (57) 即可。

首先求 $\frac{d\xi_t}{dt}=v_t$ 时我们遇到的问题是如何对 $\int_0^t\sin\lambda(t-s)dB_s$ 求导？事实上，利用分部积分公式或乘法公式

$$\int_0^t\sin\lambda(t-s)dB_s=\sin\lambda(t-s)B_s\big|_0^t-\int_0^t B_s d\sin\lambda(t-s)=\lambda\int_0^t B_s\cos\lambda(t-s)ds. \tag{60}$$

由此得到

$$\xi_t=x\cos\lambda t+\frac{v_0}{\lambda}\sin\lambda t+\int_0^t B_s\cos\lambda(t-s)ds. \tag{61}$$

$$\frac{d\xi_t}{dt}=-\lambda x\sin\lambda t+v_0\cos\lambda t-\lambda\int_0^t B_s\sin\lambda(t-s)ds+B_t=v_t. \tag{62}$$

另一方面, 对 (61) 式两边关于 t 做积分, 得到

$$\begin{aligned}\int_0^t \xi_s ds &= \frac{x}{\lambda}\sin\lambda t + \frac{v_0}{\lambda^2}\big(1-\cos\lambda t\big) + \int_0^t\int_0^u B_s\cos\lambda(u-s)dsdu \\ &= \frac{x}{\lambda}\sin\lambda t + \frac{v_0}{\lambda^2}\big(1-\cos\lambda t\big) + \int_0^t B_s\int_s^t \cos\lambda(u-s)duds \\ &= \frac{1}{\lambda^2}\big(v_0 + \lambda x\sin\lambda t - v_0\cos\lambda t + \lambda\int_0^t B_s\sin\lambda(t-s)ds\big),\end{aligned}$$

将上式与 (62) 比较, 即得

$$v_t = v_0 - \lambda^2\int_0^t \xi_s ds + B_t,$$

因此 $(\xi_t, v_t)_{t\geq 0}$ 是方程 (57) 或 (58) 的解。

注记 6.2. $\int_0^t \sin\lambda(t-s)dB_s$ 不是鞅 (也不是局部鞅). 分部积分公式 (60) 证明有两种方式。第一种是对于固定的 t 利用随机积分的定义程序，先对阶梯函数做随机积分得到等式左边的项，再利用 Able 求和得到右边的项，两边取 L^2 极限就可以了。第二种方式是对于固定的 t，在 $[0,t]$ 内时间定义 $f^{(t)}(s) = \sin\lambda(t-s)$，由于 f 是确定性的数值函数，所以是 $(\mathcal{F}_u^B)_{0\leq u\leq t}$ 适应的。由乘积公式

$$d\big(f^{(t)}(s)B_s\big) = f^{(t)}(s)dB_s + B_s df^{(t)}(s) + dB_s df^{(t)}(s).$$

右边第三项为零，写成积分形式

$$f^{(t)}(u)B_u - f^{(t)}(0)B_0 = \int_0^u f^{(t)}(s)dB_s + \int_0^t B_s df^{(t)}(s), \quad u\in[0,t].$$

即

$$\sin\lambda(t-u)B_u - \sin\lambda(t-0)B_0 = \int_0^u \sin\lambda(t-s)dB_s + \int_0^u B_s d\sin\lambda(t-s).$$

当 $u=t, B_0=0$ 时，就得到了 (60).

例 6.4. Black-Scholes 方程

在第三章例子 (2.4) 中我们讨论了连续时间的 Black-Scholes 模型的直观推导，得到了风险证卷价格 $(S_t)_{t\geq 0}$ 形式上应该满足

$$S_t - S_0 = \mu\int_0^t S_s ds + \sigma\int_0^t S_s\text{``}d\bar{\xi}_s\text{''}. \tag{63}$$

现在，我们可以将 $\int_0^t S_s\text{``}d\bar{\xi}_s\text{''}$ 解释成关于布朗运动的 Itô 随机积分了，称

$$\begin{cases} dS_t = \mu S_t dt + \sigma S_t dB_t \\ S_0 = s_0 \end{cases} \tag{64}$$

为 Black-Scholes 方程。

下面我们关心如何求解这个方程和解的长时间行为这两个问题。

(1) 先形式地求解，设 $g(x)=\ln x$

$$\begin{aligned} dg(S_t) &= g'(S_t)dS_t+\frac{1}{2}g''(S_t)(dS_t)^2 \\ &= \frac{1}{S_t}[\mu S_t dt+\sigma S_t dB_t]-\frac{1}{2}\frac{1}{S_t^2}(\mu S_t dt+\sigma S_t dB_t)^2 \\ &= [\mu-\frac{1}{2}\sigma^2]dt+\sigma dB_t. \end{aligned}$$

$$\begin{aligned} \ln S_t-\ln s_0 &= [\mu-\frac{1}{2}\sigma^2]t+\sigma B_t, \quad (\text{设}B_0=0) \\ \ln\frac{S_t}{s_0} &= [\mu-\frac{1}{2}\sigma^2]t+\sigma B_t. \\ S_t &= s_0\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma B_t\}. \end{aligned} \tag{65}$$

从数学的严格性上来说，我们上面的计算只是猜出解的形式。我们还要面对两个问题：第一，解的唯一性。这个问题下面的随机微分方程解的定理存在唯一性定理6.1就可以给出肯定的答案；第二，若解有正概率小于等于零，则 $\ln S_t$ 是不能定义的。这个问题的解决方案是，既然方程解是唯一的，不论初值为正还是为负，我们对 (65) 式用 Itô 公式直接验证满足方程即可。

(2) 显然，当 $s_0=0$ 时，$S_t=0$ 是方程的解。下面我们只考虑 $s_0>0$ 的情况。

$$S_t=s_0\exp\big\{\big[(\mu-\frac{1}{2}\sigma^2)+\frac{\sigma B_t}{t}\big]t\big\}.$$

已知布朗运动的强大数律

$$P\big(\lim_{t\to\infty}\frac{B_t}{t}=0\big)=1.$$

当 $\mu-\frac{1}{2}\sigma^2>0$ 时，

$$P\big(\lim_{t\to\infty}\big[(\mu-\frac{1}{2}\sigma^2)+\frac{\sigma B_t}{t}\big]>0\big)=1.$$

因此，

$$P\big(\lim_{t\to\infty}S_t=+\infty\big)=1.$$

当 $\mu-\frac{1}{2}\sigma^2<0$ 时，类似地可以得到

$$P\big(\lim_{t\to\infty}S_t=0\big)=1.$$

当 $\mu-\frac{1}{2}\sigma^2=0$ 时，

$$S_t=s_0\exp\big\{\sigma B_t\big\}=s_0\exp\big\{\sigma\frac{B_t}{\sqrt{2t\ln\ln t}}\sqrt{2t\ln\ln t}\big\}.$$

由布朗运动的重对数律

$$P\big(\limsup_{t\to\infty}\frac{B_t}{\sqrt{2t\ln\ln t}}=1\big)=1,$$

$$P\big(\liminf_{t\to\infty}\frac{B_t}{\sqrt{2t\ln\ln t}}=-1\big)=1,$$

得到

$$P\big(\limsup_{t\to\infty}S_t=+\infty,\ \liminf_{t\to\infty}S_t=0\big)=1.$$

注记 6.3. 我们还可以将 (63) 在数学上解释成 Stratonovich 型的随机微分方程

$$\begin{cases} dS_t = \mu S_t dt + \sigma S_t \circ dB_t \\ S_0 = s_0 \end{cases}$$

这个方程的解是 $S_t = s_0 \exp\left\{\mu t + \sigma B_t\right\}$。

通常把 $\exp(\mu t + \nu B_t)$ 称为几何布朗运动，其名字的来源于几何级数概念。

6.1.2 解的存在唯一性

定理 6.1. 若 $\sigma(t,x)$，$b(t,x)$ 满足 Lipschitz 条件，即对于任意的 $T>0$ 存在 $C_T>0$，使得任意的 $t<T$，

$$|\sigma(t,x)-\sigma(t,y)|+|b(t,x)-b(t,y)| \le C_T|x-y|.$$

对于初始随机变量 ξ_0，ξ_0 与 $(B_t,\ \ t\ge 0)$ 独立，则方程

$$\xi_t = \xi_0 + \int_0^t b(s,\xi_s)ds + \int_0^t \sigma(s,\xi_s)dB_s$$

有唯一解，即若还有另一个解 ξ_t'，只要 $P(\xi_0'=\xi_0)=1$, 则

$$P(\xi_t'=\xi_t,\ \ \forall t>0)=1.$$

证明： Picard 迭代，Gronwall 不等式. □

定理 6.2. 若 $\sigma(t,x)$，$b(t,x)$ 满足:

(1) 局部 Lipschitz 条件，即对于任意的 $T,N>0$ 存在 $C_{T,N}>0$，使得对于任意的 $|x|\le N$，$|y|\le N$，$t\le T$

$$|\sigma(t,x)-\sigma(t,y)|+|b(t,x)-b(t,y)| \le C_{T,N}|x-y|.$$

(2) 线性增长条件，即对于任意的 $T>0$ 存在 $C_T>0$，使得任意的 $t<T$ (T 可以取 ∞)

$$|\sigma(t,x)|+|b(t,x)| \le C_T(1+|x|),$$

对于初始随机变量 ξ_0，ξ_0 与 $(B_t,\ \ t\ge 0)$ 独立，则方程

$$\xi_t = \xi_0 + \int_0^t b(s,\xi_s)ds + \int_0^t \sigma(s,\xi_s)dB_s.$$

有唯一解，即若还有另一个解 ξ_t'，只要 $P(\xi_0'=\xi_0)=1$, 则

$$P(\xi_t'=\xi_t,\ \ \forall t>0)=1.$$

6.2 扩散过程及解的 Markov 性

定义 6.1. 扩散过程

随机微分方程

$$d\xi_t = b(t,\xi_t)dt + \sigma(t,\xi_t)dB_t, \tag{66}$$

的唯一解称为扩散过程，其中 $b(t,x)$ 称为漂移系数，$a(t,x)=\sigma^2(t,x)$ 称为扩散系数。

注记 6.4. (1)“扩散过程”这个名词来源于物理等学科，在不同的书上有不同的严格数学定义，但其核心是指 (右) 连续的强马氏过程。我们这里简单地粗率地把它看成无记忆性随机微分方程 (66) 的解。

(2)“扩散过程”缺图！

定理 6.3. 随机微分方程 (66) 的解若存在唯一，则这个解是一个 Markov 过程。而且当系数不含时间 t 时，即 $\sigma(t,x)=\sigma(x)$，$b(t,x)=b(x)$ 时，此时 Markov 过程是时齐的。

证明大意与直观：将随机微分方程在 (66) 写出积分形式，

$$\xi_t = \xi_0 + \int_0^t b(u,\xi_u)du + \int_0^t \sigma(u,\xi_u)dB_u, \tag{67}$$

$$\xi_t^{s,x} = x + \int_s^t b(u,\xi_u^{s,x})du + \int_s^t \sigma(u,\xi_u^{s,x})dB_u, \quad s<t. \tag{68}$$

其中，$\xi_t^{s,x}$ 表示从 s 时刻 x 位置出发的方程的解，它只与 $\big(s,x,t,(B_u-B_s, s<u\le t)\big)$ 有关，即 $\xi_t^{s,x}$ 是 $\big(s,x,t,(B_u-B_s, s\le u\le t)\big)$ 的函数 (泛函)，记为

$$G\big(s,x,t,(B_u-B_s, s\le u\le t)\big). \tag{69}$$

另一方面

$$\xi_t = \xi_0 + \int_0^s b(u,\xi_u)du + \int_0^s \sigma(u,\xi_u)dB_u + \int_s^t b(u,\xi_u)du + \int_s^t \sigma(u,\xi_u)dB_u. \tag{70}$$

因此，若方程 (66) 的解存在唯一，则 (67) 与 (70) 决定了同一个过程，即

$$\xi_t = \xi_s + \int_s^t b(u,\xi_u)du + \int_s^t \sigma(u,\xi_u)dB_u.$$

由 (69)，ξ_t 可以表示成 $G(s,\xi_s,t,(B_u-B_s,s<u\le t))$.

注意到 $(B_u-B_s, s<u\le t)$ 与 s 之前的事件域 $\sigma(B_v,\ v\le s)$ 独立。因此，ξ_t 在已知 ξ_s 的条件下与 $\sigma(B_v, v\le s)$ 独立。这说明 $(\xi_t)_{t\ge 0}$ 是一个 $(\mathcal{F}_t^B)_{t\ge0}$Markov 过程。

□

注记 6.5. 若 $b(u,\xi)$，$\sigma(u,\xi)$ 是有记忆型时，则 ξ_t 是不能表示成 $G(s,\xi_s,t,(B_u-B_s,s<u\le t))$。原因是 $b(u,\xi),\sigma(u,\xi)$ 当 $u>s$ 是还可能依赖于 $\sigma(B_v,\ v\le s)$.

Markov 过程的直观说法就是已知现在，过去和将来独立。

例 6.5. 接例6.1

我们将 Langevin 方程 (54) 写成稍微一般一点儿的形式

$$d\eta_t = -b\eta_t dt + \sigma dB_t,$$

则解为

$$\eta_t = e^{-bt}\big(\eta_0 + \sigma\int_0^t e^{bs}dB_s\big),$$

称为 OU 过程 (Ornstein-Uhlenbeck 过程，严格地说这个方程的解从其不变概率分布出发时才能称为 OU 过程)。

由定理6.1和定理6.3，$(\eta_t)_{t\geq 0}$ 是一个时齐的 Markov 过程。自然的我们可以问它的转移概率密度是什么，它的是否有不变概率分布？

Markov 过程 $(\eta_t)_{t\geq 0}$ 的转移概率分布 $P(s,x;s+t,A)$ 可以理解为过程从 s 时刻 x 位置出发在 $s+t$ 时刻到达集合 A 的概率，因此我们只需要计算

$$\eta_{t+s}^{s,x} = e^{-bt}\big(x+\sigma\int_s^{t+s} e^{b(u-s)}dB_u\big)$$

的分布函数. 因为 e^{bu} 是确定性函数，所以 $\int_s^{t+s} e^{bu}dB_u$ 是一个 Gauss 分布，利用 Itô 等距，

$$\int_s^{t+s} e^{bu}dB_u \sim N\Big(0,\int_s^t e^{2bu}du\Big) = N\Big(0,\frac{e^{2b(t+s)}-e^{2bs}}{2b}\Big).$$

那么

$$\eta_{t+s}^{s,x} \sim N\Big(xe^{-bt},\ \frac{\sigma^2}{2b}(1-e^{-2bt})\Big).$$

由此得 $(\eta_t)_{t\geq 0}$ 的转移密度函数是

$$p(t,x,y)\equiv p(s,x;s+t,y) = \sqrt{\frac{b}{\pi\sigma^2(1-e^{-2bt})}}\exp\Big(-\frac{(y-xe^{-bt})^2}{\frac{\sigma^2(1-e^{-2bt})}{b}}\Big).$$

当 $b>0$ 时，

$$\lim_{t\to\infty} p(t,x,y) = \frac{1}{\sigma\sqrt{\frac{\pi}{b}}}e^{-\frac{y^2}{\frac{\sigma^2}{b}}} \equiv \pi(y).$$

由此可得 $N(0,\frac{\sigma^2}{2b})$ 是 OU 过程的不变概率分布，也就是说以 π 为初始分布密度，则 OU 过程为一个平稳的马氏过程其平稳分布为 π。(详见习题)

注记 6.6. (1) 由时间齐次性，事实上可以通过计算

$$\eta_t = e^{-bt}x+\sigma\int_0^t e^{b(s-t)}dB_s,$$

的密度来求的转移概率密度函数。

(2) 设 $b,\sigma>0$, B_t 是一个标准布朗运动.，则平稳的 OU 过程另一种表示为

$$\frac{\sigma}{\sqrt{2b}}e^{-bt}B(e^{2bt}).$$

例 6.6. 接例6.4

已知 Black-Scholes 方程 (64) 的解

$$S_t = s_0\exp\Big\{\big[\mu-\frac{1}{2}\sigma^2\big]t+\sigma B_t\Big\}.$$

同样，由定理6.1和定理6.3知道，它是一个 Markov 过程。特别地，当 $s_0=0$ 时，过程恒等于零。因此我们不妨考虑 $s_0>0$，也就是说这个过程的状态空间是 $(0,\infty)$。对于

$f(z)=\mathbf{1}_{(-\infty,y]}(z)$，

$$\begin{aligned}
& E[f(S_{t+s})|S_s] \\
= \ & E\big[f\big(s_0\exp\{[\mu-\frac{1}{2}\sigma^2]s+\sigma B_s\}\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big|S_s\big] \\
= \ & E\big[f\big(S_s\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big|S_s\big] \\
= \ & E\big[f\big(x\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\big)\big]\Big|_{x=S_s} \\
= \ & P\Big(x\exp\{[\mu-\frac{1}{2}\sigma^2]t+\sigma(B_{t+s}-B_s)\}\le y\Big)\Big|_{x=S_s} \\
= \ & P\Big(\frac{B_{t+s}-B_s}{\sqrt{t}}\le\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big)\Big|_{x=S_s} \\
= \ & \Phi\Big(\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big)\Big|_{x=S_s}.
\end{aligned}$$

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$。$(S_t)_{t\ge0}$ 的转移概率分布函数

$$F(t,x,y)=\Phi\Big(\frac{1}{\sigma\sqrt{t}}\big(\ln\frac{y}{x}-[\mu-\frac{1}{2}\sigma^2]t\big)\Big).$$

6.3 Kolmogorov 向前方程、Fokker-Planck 方程、Kolmogorov 向后方程

6.3.1 扩散过程古典的分析描述与随机微分方程

我们描述随机现象可以有两个层次的方式，

1) 概率空间及其上的随机变量 X，对实际问题建立较为精确数学模型；

2) 分布或分布函数 (定义于观测空间，取值在实数空间，联系着统计性质，通过观测得到的)，易得易算，例如 $F(x)=P(X\le x)$.

同样，对于扩散过程也有两个层次的描述方式，

(1) 由随机微分方程描述其演化规律：设有 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\ge0},P)$, 布朗运动 $(B_t)_{t\ge0}$；

$$\xi_t=\xi_0+\int_0^t b(s,\xi_s)ds+\int_0^t\sigma(s,\xi_s)dB_s. \tag{71}$$

(2) 转移密度 (分布) 函数 $p(s,x;t,y)$ 描述其统计规律。

它们分别从不同的角度（“微观”与“宏观”）来描述具有如下运动规律的具有连续运动轨迹的物理“扩散过程”的:

$$\lim_{\Delta\downarrow0}\frac{1}{\Delta}\int(y-x)p(s,x;s+\Delta,y)dy=b(s,x), \tag{72}$$

$$\lim_{\Delta\downarrow0}\frac{1}{\Delta}\int(y-x)^2p(s,x;s+\Delta,y)dy=\sigma^2(s,x). \tag{73}$$

一个自然的问题是，两种描述方式是否是一致？在回答这个问题之前，我们先来回答一个更基本的问题，是否随机微分方程定义的扩散过程存在转移密度函数。由上一节的例子我们知道只有极少数的扩散过程可以得到显式的转移概率密度，但事实上有相当大的一类扩散过程存在转移概率密度的。

定理 6.4. (并非严格的数学定理)

相当一般的 (例如：解存在唯一，系数光滑，扩散系数具有一个正下界) 扩散过程是具有转移密度 $p(s,x;t,y)$ 的 Markov 过程。

下面我们验证当扩散过程 (71) 有转移密度函数时，它满足 (72)(73). 此时假设 $\int_0^t \sigma(s,\xi_s)dB_s$ 和 $\int_0^t \sigma(s,\xi_s)\xi_s dB_s$ 是鞅，

$$\begin{aligned}
&\int (y-\xi_s)p(s,\xi_s;s+\Delta,y)dy \\
=\;& E\big[\xi_{s+\Delta}-\xi_s\big|\xi_s\big] \\
=\;& E\big[\xi_{s+\Delta}-\xi_s\big|\mathcal{F}_s^B\big] \\
=\;& E\big[\int_s^{s+\Delta} b(u,\xi_u)ds\big|\mathcal{F}_s^B\big] \\
=\;& E\big[\int_s^{s+\Delta} b(u,\xi_u)ds\big|\xi_s\big]
\end{aligned} \tag{74}$$

$$\begin{aligned}
&\frac{1}{\Delta}\int (y-x)p(s,x;s+\Delta,y)dy \\
=\;& \frac{1}{\Delta}E[\xi_{s+\Delta}-\xi_s|\xi_s=x] = \frac{1}{\Delta}E\Big[\int_s^{s+\Delta} b(u,\xi_u)du\Big|\xi_s=x\Big] \\
=\;& E\Big[\frac{1}{\Delta}\int_s^{s+\Delta} b(u,\xi_u)du\Big|\xi_s=x\Big] \\
\overset{\Delta\downarrow 0}{=}\;& E[b(s,\xi_s)|\xi_s=x] = b(s,x)
\end{aligned} \tag{75}$$

类似地，

$$\begin{aligned}
&\frac{1}{\Delta}\int (y-x)^2 p(s,x;s+\Delta,y)dy \\
=\;& \frac{1}{\Delta}E[(\xi_{s+\Delta}-\xi_s)^2|\xi_s=x] = \frac{1}{\Delta}E[(\xi_{s+\Delta}-x)^2|\xi_s=x] \\
=\;& \frac{1}{\Delta}E[\xi_{s+\Delta}^2 - 2x\xi_{s+\Delta}+x^2|\xi_s=x] \\
=\;& \frac{1}{\Delta}E\Big[\int_s^{s+\Delta} 2\xi_u d\xi_u + \int_s^{s+\Delta}\sigma^2(u,\xi_u)du - 2x\xi_{s+\Delta}+2x^2\Big|\xi_s=x\Big] \\
=\;& \frac{1}{\Delta}E\Big[\int_s^{s+\Delta} 2\xi_u b(u,\xi_u)du + 2\int_s^{s+\Delta}\xi_u\sigma(u,\xi_u)dB_u + \int_s^{s+\Delta}\sigma^2(u,\xi_u)du - 2x(\xi_{s+\Delta}-x)\Big|\xi_s=x\Big] \\
\overset{\Delta\downarrow 0}{=}\;& 2xb(u,x)+\sigma^2(u,x)-2xb(u,x) = \sigma^2(u,x).
\end{aligned}$$

至少，对于具有转移密度函数的扩散过程，两种描述是一致的。事实上，利用随机微分方程可以构造出更多更广的“扩散过程”。

既然大多数情况下不能给出 $p(s,x;t,y)$ 的显式表达，一个自然的问题是它应该遵循怎样的时间演化规律。这个问题我们并非第一次遇到，例如，我们在时间齐次的连续时间 Markov 链理论中定义了转移概率矩阵 $(p_{ij}(t))$ 和速率矩阵 Q，我们知道它们满足

$$P'(t)=P(t)Q \ \ (\text{向前方程}), \qquad\qquad P'(t)=QP(t) \ \ (\text{向后方程}).$$

6.3.2 Kolmogorov 向前方程、Fokker-Planck 方程

定理 6.5. (Kolmogorov 向前方程、Fokker-Planck 方程、Kolmogorov 第二方程)

在定理6.4同样的假定下，扩散过程

$$d\xi_t = b(t,\xi_t)dt + \sigma(t,\xi_t)dB_t$$

的转移密度函数是如下的偏微分方程的基本解

$$\frac{\partial}{\partial t}p(s,x;t,y) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma(t,y)^2 p(s,x;t,y)\big) - \frac{\partial}{\partial y}\big(b(t,y)p(s,x;t,y)\big). \tag{76}$$

或者记为

$$\frac{\partial p(s,x;t,y)}{\partial t} = \mathcal{L}_y^* p(s,x;t,y).$$

它还满足

$$\int p(s,x;s,y)f(y)dy = f(x), \quad \forall f(x)\in C_0^\infty(\mathbb{R}), \tag{77}$$

即

$$p(s,x;s,y) = \delta_x.$$

其中 C_0^∞ 为 $\mathbb{R}$ 上无穷次可微，且 $\{x:\ f(x)\neq 0\}$ 是一个紧集 (有界闭集)。进一步，若

$$d\xi_t = b(\xi_t)dt + \sigma(\xi_t)dB_t$$

则

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma(y)^2 p(t,x,y)\big) - \frac{\partial}{\partial y}\big(b(y)p(t,x,y)\big).$$

证明： 对于 $\forall \varphi\in C_0^\infty$, 考虑过程 $\varphi(\xi(t))$，由 Itô 公式

$$d\varphi(\xi(t)) = \varphi'(\xi(t))d\xi(t) + \frac{1}{2}\varphi''(\xi(t))\sigma^2(t,\xi(t))dt,$$

$$\varphi(\xi(t'))-\varphi(\xi(t)) = \int_t^{t'}\varphi'(\xi(u))b(u,\xi(u))du + \frac{1}{2}\int_t^{t'}\varphi''(\xi(u))\sigma^2(u,\xi(u))du + \int_t^{t'}\varphi'(\xi(u))\sigma(u,\xi(u))dB_u.$$

对于 $t'>t\geq s$，类似于 (74) 和 (75) 推导中可以把鞅的项通过取期望去掉，

$$\begin{aligned}
&E[\varphi(\xi_{t'})-\varphi(\xi_t)|\xi_s=x]\\
=\ & E\Big[\int_t^{t'}\big[b(u,\xi_u)\varphi'(\xi_u)+\frac{1}{2}\sigma^2(u,\xi_u)\varphi''(\xi_u)\big]du\Big|\xi_s=x\Big]
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\Delta}E[\varphi(\xi_{t+\Delta})-\varphi(\xi_t)|\xi_s=x]\\
=\ & \frac{1}{\Delta}E\Big[\int_t^{t+\Delta}\big[b(u,\xi_u)\varphi'(\xi_u)+\frac{1}{2}\sigma^2(u,\xi_u)\varphi''(\xi_u)\big]du\Big|\xi_s=x\Big]\\
=\ & E\Big[\frac{1}{\Delta}\int_t^{t+\Delta}\big[b(u,\xi_u)\varphi'(\xi_u)+\frac{1}{2}\sigma^2(u,\xi_u)\varphi''(\xi_u)\big]du\Big|\xi_s=x\Big]
\end{aligned}$$

$\Delta \downarrow 0$，得到

$$
\begin{aligned}
\text{左} &= \frac{\partial \int \varphi(y) p(s,x;t,y) dy}{\partial t} = \int \varphi(y) \frac{\partial p(s,x;t,y)}{\partial t} dy \\
\text{右} &= E\big[b(t,\xi_t)\varphi'(\xi_t) + \frac{1}{2}\sigma^2(t,\xi_t)\varphi''(\xi_t)\big|\xi_s = x\big] \\
&= \int \big(b(t,y)\varphi'(y) + \frac{1}{2}\sigma^2(t,y)\varphi''(y)\big) p(s,x;t,y) dy \\
&\overset{\text{分部积分}}{=} -\int \varphi(y) \frac{\partial}{\partial y}[b(t,y)p(s,x;t,y)]dy + \frac{1}{2}\int \varphi(y) \frac{\partial^2}{\partial y^2}\big(\sigma(t,y)^2 p(s,x;t,y)\big) dy.
\end{aligned}
$$

即

$$
\frac{\partial p(s,x;t,y)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma(t,y)^2 p(s,x;t,y)\big) - \frac{\partial}{\partial y}\big(b(t,y)p(s,x;t,y)\big).
$$

再由扩散过程轨道的连续性和控制收敛定理，易证

$$
\lim_{t \downarrow s} \int p(s,x;t,y) f(y) dy = \lim_{t \downarrow s} E[f(\xi_t)|\xi_s = x] = E[f(\xi_s)|\xi_s = x] = f(x). \tag{78}
$$

即

$$
p(s,x;s,y) = \delta_x(y).
$$

□

推论 6.1. 对于时齐的扩散过程

$$
\xi_t = \xi_0 + \int_0^t b(\xi_s) ds + \int_0^t \sigma(\xi_s) dB_s
$$

ξ_0 与 $(B_t)_{t \geq 0}$ 独立，其分布密度为 $p(0,x)$，则 ξ_t 的分布密度为

$$
p(t,y) = \int p(0,x) p(t,x,y) dx
$$

则 $p(t,y)$ 满足以下的 Master 方程 (主方程)

$$
\frac{\partial p}{\partial t} = \frac{\partial^2}{\partial y^2}\big(\frac{\sigma^2(y)}{2} p\big) - \frac{\partial}{\partial y}\big(b(y)p\big).
$$

证明：利用 Fokker-Planck 方程

$$
\begin{aligned}
\frac{\partial p}{\partial t}(t,y) &= \int p(0,x) \frac{\partial p(t,x,y)}{\partial t} dx \\
&= \int p(0,x) \mathcal{L}_y^* p(t,x,y) dx \\
&= \int \mathcal{L}_y^* p(0,x) p(t,x,y) dx \\
&= \mathcal{L}_y^* \int p(0,x) p(t,x,y) dx \\
&= \mathcal{L}_y^* p(t,y) = \frac{\partial^2}{\partial y^2}\big(\frac{\sigma^2(y)}{2} p(t,y)\big) - \frac{\partial}{\partial y}\big(b(y)p(t,y)\big).
\end{aligned}
$$

□

注记 6.7. 主方程表达了 ξ_t 的密度 $p(t,y)$ 应该满足的演化规律。

6.3.3 Kolmogorov 向后方程、Kolmogorov 第一方程

定理 6.6. Kolmogorov 向后方程、Kolmogorov 第一方程

$$-\frac{\partial p(s,x;t,y)}{\partial s}=\frac{1}{2}\sigma^2(s,x)\frac{\partial^2 p(s,x;t,y)}{\partial x^2}+b(s,x)\frac{\partial p(s,x;t,y)}{\partial x}, \tag{79}$$

$$\lim_{s\uparrow t}\int p(s,x;t,y)f(y)dy=f(x). \tag{80}$$

特别地，当 $(\xi_t)_{t\geq 0}$ 为时齐马氏过程时

$$\frac{\partial p(t,x,y)}{\partial t}=\frac{1}{2}\sigma^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}+b(x)\frac{\partial p(t,x,y)}{\partial x}. \tag{81}$$

$$\lim_{t\downarrow 0}\int f(y)p(t,x,y)=f(x),\ \forall f(x)\in C_0^\infty(\mathbb{R}),\ \text{即}\ p(0,x,y)=\delta_x(y). \tag{82}$$

证明： $\varphi(x)\in C_0^\infty$，令 $u(s,x)=E[\varphi(\xi_t)|\xi_s=x]=\int\varphi(y)p(s,x;t,y)dy$，则

$$u(s+\Delta,\xi_{s+\Delta})=E[\varphi(\xi_t)|\xi_{s+\Delta}]=\int\varphi(y)p(s+\Delta,\xi_{s+\Delta};t,y)dy.$$

利用 Itô 公式

$$\begin{aligned}&\frac{1}{\Delta}[u(s+\Delta,\xi_{s+\Delta})-u(s,\xi_s)] \qquad (83)\\ =\ &\frac{1}{\Delta}\Big[\int_s^{s+\Delta}\Big(\frac{\partial u}{\partial l}(l,\xi_l)+\frac{\partial u}{\partial x}(l,\xi_l)b(l,\xi_l)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(l,\xi_l)\sigma^2(l,\xi_l)\Big)dl+\int_s^{s+\Delta}\frac{\partial u}{\partial x}(l,\xi_l)\sigma dB_l\Big].\end{aligned}$$

注意到

$$E[u(s+\Delta,\xi_{s+\Delta})|\xi_s=x]=E\big[E[\varphi(\xi_t)|\xi_{s+\Delta}]\big|\xi_s=x\big]\overset{\text{Chapman-Kolmogorov 方程 (马氏性)}}{=}E[\varphi(\xi_t)|\xi_s=x].$$

$$E[u(s,\xi_s)|\xi_s=x]=u(s,x)=E[\varphi(\xi_t)|\xi_s=x].$$

对 (83) 两边取条件期望 $E[\cdot|\xi_s=x]$，

$$E\Big[\frac{1}{\Delta}\int_s^{s+\Delta}[\quad]dl\Big|\xi_s=x\Big]=0.$$

$\Delta\downarrow 0$，得到

$$E\Big[\frac{\partial u}{\partial s}(s,\xi_s)+\frac{\partial u}{\partial x}(s,\xi_s)b(s,\xi_s)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(s,\xi_s)\sigma^2(s,\xi_s)\Big|\xi_s=x\Big]=0.$$

即

$$-\frac{\partial u}{\partial s}(s,x)=\frac{\partial u}{\partial x}(s,x)b(s,x)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(s,x)\sigma^2(s,x).$$

求导与积分交换顺序得

$$-\int\varphi(y)\frac{\partial p(s,x;t,y)}{\partial s}dy=\int\varphi(y)\Big(b(s,x)\frac{\partial p(s,x;t,y)}{\partial x}+\frac{1}{2}\sigma^2(s,x)\frac{\partial^2 p(s,x;t,y)}{\partial x^2}\Big)dy.$$

有 φ 的任意性就得到 (79).

进一步, 既然 $\int\varphi(y)p(s,x;t.y)dy=E[\varphi(\xi_t)|\xi_s=x]$，

$$
\begin{aligned}
& \lim_{s\uparrow t}\big(\int\varphi(y)p(s,x;t.y)dy-\varphi(x)\big)\\
= & \lim_{s\uparrow t}E\big[\varphi(\xi_t)-\varphi(x)|\xi_s=x\big]\\
= & \lim_{s\uparrow t}E\Big[\int_s^t\big(\varphi'(\xi_u)b(u,\xi_u)du+\frac{1}{2}\varphi''(\xi_u)\sigma^2(s,\xi_u)\big)du+\int_s^t\varphi'(\xi_u)\sigma(u,\xi_u)dB_u\big|\xi_s=x\Big]\\
= & \lim_{s\uparrow t}E\Big[\int_s^t\big(\varphi'(\xi_u)b(u,\xi_u)du+\frac{1}{2}\varphi''(\xi_u)\sigma^2(s,\xi_u)\big)du\big|\xi_s=x\Big]\\
& \text{(由积分的性质和有界控制收敛定理)}\\
= & 0.
\end{aligned}
$$

此即 (80).

时齐情形的证明

$$f(\xi_t)-f(\xi_0)=\int_0^t\big[b(\xi_u)f'(\xi_u)+\frac{1}{2}\sigma^2(\xi_u)f''(\xi_u)\big]du+\int_0^t f'(\xi_u)\sigma(\xi_u)dB_u,$$

$$\lim_{\Delta\to0}\frac{1}{\Delta}E\big[f(\xi_\Delta)-f(\xi_0)\big|\xi_0=x\big]=b(x)f'(x)+\frac{1}{2}\sigma^2(x)f''(x). \tag{84}$$

令

$$
\begin{aligned}
u(t,x) &= E[\varphi(\xi_t)|\xi_0=x]=\int\varphi(y)p(t,x,y)dy\\
&= \int\varphi(y)p(s,x,s+t,y)dy
\end{aligned}
$$

则利用 (84),

$$
\begin{aligned}
& \lim_{\Delta\to0}\frac{1}{\Delta}E\big[u(t,\xi_\Delta)-u(t,\xi_0)\big|\xi_0=x\big]\\
= & b(x)u'(t,x)+\frac{1}{2}\sigma^2(x)u''(t,x).
\end{aligned}
$$

另一方面，

$$
\begin{aligned}
& E[u(t,\xi_\Delta)|\xi_0=x]\\
= & \int u(t,z)p(\Delta,x,z)dz\\
= & \int\int\varphi(y)p(t,z,y)dyp(\Delta,x,z)dz\\
= & \int\varphi(y)p(t+\Delta,x,y)dy=u(t+\Delta,x)\\
& E[u(t,\xi_0)|\xi_0=x]=u(t,x)
\end{aligned}
$$

故

$$
\begin{aligned}
& \lim_{\Delta\to0}\frac{1}{\Delta}E\big[u(t,\xi_\Delta)-u(t,\xi_0)\big|\xi_0=x\big]\\
= & \lim_{\Delta\to0}\frac{1}{\Delta}\big(u(t+\Delta,x)-u(t,x)\big)=\frac{\partial u(t,x)}{\partial t}
\end{aligned}
$$

由 φ 的任意性得到 (81). 与 (78) 的证明类似可以得到 (82). □

在上面的推导中事实上我们证明了如下的**终值问题** (final value problem).

推论 6.2. 对于有界连续函数 $f(x)$，设

$$u(t,x)=E[f(\xi_T)|\xi_t=x],\quad (t\le T),$$

则它是方程

$$\begin{cases}\frac{\partial u}{\partial t}+\frac{1}{2}\sigma^2(t,x)\frac{\partial^2 u}{\partial x^2}+b(t,x)\frac{\partial u}{\partial x}=0\\ \lim_{t\uparrow T}u(t,x)=f(x)\end{cases}.$$

的解。

若设 $v(t,x)=u(T-t,x)=E[f(\xi_T)|\xi_{T-t}=x]$, 则对于 $0\le t\le T$, $v(t,x)$ 满足

$$\begin{cases}\frac{\partial v}{\partial t}=\frac{1}{2}\sigma^2(t,x)\frac{\partial^2 v}{\partial x^2}+b(t,x)\frac{\partial v}{\partial x}\\ \lim_{t\downarrow 0}v(t,x)=f(x)\end{cases}. \tag{85}$$

若 ξ_t 是时齐的扩散过程时，$u(t,x)=E[\xi_t|\xi_0=x]$ 是如下线性抛物方程 Cauchy 问题的解

$$\begin{cases}\frac{\partial u}{\partial t}=\frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}+b(x)\frac{\partial u}{\partial x}\\ u(0,x)=f(x)\end{cases}.$$

6.4 多维扩散过程 (Multidimensional Diffusion Processes)

6.4.1 随机微分方程描述

$$\begin{pmatrix}d\xi_t^{(1)}\\ \vdots\\ d\xi_t^{(n)}\end{pmatrix}=\begin{pmatrix}b_1(t,\xi_t)\\ \vdots\\ b_n(t,\xi_t)\end{pmatrix}dt+\begin{pmatrix}\sigma_{11}(t,\xi_t)&\cdots&\sigma_{1r}(t,\xi_t)\\ \vdots&\vdots&\vdots\\ \sigma_{n1}(t,\xi_t)&\cdots&\sigma_{nr}(t,\xi_t)\end{pmatrix}\begin{pmatrix}dB_t^{(1)}\\ \vdots\\ dB_t^{(r)}\end{pmatrix}. \tag{86}$$

设

$$\mathbf{b}(t,x)=\begin{pmatrix}b_1(t,x)\\ \vdots\\ b_n(t,x)\end{pmatrix},\qquad \Sigma(t,x)=\begin{pmatrix}\sigma_{11}(t,x)&\cdots&\sigma_{1r}(t,x)\\ \vdots&\vdots&\vdots\\ \sigma_{n1}(t,x)&\cdots&\sigma_{nr}(t,x)\end{pmatrix}_{n\times r}.$$

可以将方程 (86) 记为

$$d\xi_t=\mathbf{b}(t,\xi_t)dt+\Sigma(t,\xi_t)d\mathbf{B}_t.$$

6.4.2 古典的分析描述方式

设 $x\in\mathbb{R}^n$, $y\in\mathbb{R}^n$.

$$\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_{|y-x|<1}(y-x)p(t,x;t+\Delta,y)dy=\mathbf{b}(t,x)=\begin{pmatrix}b_1(t,x)\\ \vdots\\ b_n(t,x)\end{pmatrix}.$$

$$\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_{|y-x|<1}(y-x)(y-x)^Tp(t,x;t+\Delta,y)dy=A(t,x)=(a_{ij}(t,x))_{1\le i,j\le n}.$$

6.4.3 物理的描述方式

为简单起见，考虑时齐情况。定义第 i 个方向的概率流 (probability current, flux).

$$J_i(t,x) = b_i(x)p(t,x) - \frac{1}{2}\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\big(a_{ij}(x)p(t,x)\big).$$

设概率流 $\mathbf{J} = (J_1, \cdots, J_n)^T$. 设 $V \subset \mathbb{R}^n$ 是具有光滑边界的单连通区域，则对于 $t_2 > t_1$，$\mathbf{n}$ 为 V 的单位内法向量，由概率守恒 (总概率为 1) 得到，

$$\begin{aligned}
\int_V \big(p(t_2,x) - p(t_1,x)\big)dx &= \int_{t_1}^{t_2} dt \int_V \frac{\partial}{\partial t}p(t,x)dx \\
&= \int_{t_1}^{t_2} dt \int_{\partial V} \mathbf{J}\cdot\mathbf{n}\, dS(x) \\
&\overset{\text{Gauss-Green 公式}}{=} \int_{t_1}^{t_2} \int_V -\mathrm{div}(\mathbf{J})\, dx.
\end{aligned}$$

由此即得

$$\frac{\partial p}{\partial t} + \mathrm{div}(\mathbf{J}) = 0,$$

或者

$$\frac{\partial p}{\partial t} = -\mathrm{div}(\mathbf{J}) = -\nabla\cdot\mathbf{J} \quad .$$

此即是 Fokker-Planck 方程。

事实上很容易利用上式和 Gauss-Green 公式验证

$$\frac{d}{dt}\int_{\mathbb{R}^n} p(x,t)dx = -\int_{\mathbb{R}^n} \mathrm{div}(\mathbf{J})dx = 0.$$

由此得，当 $\int_{\mathbb{R}^n} p(0,x)dx = 1$ 时，对于任意的 $t \geq 0$，$\int_{\mathbb{R}^n} p(0,x)dx = 1$。

事实上，Fokker-Planck 方程是一个守恒律方程 (equation of conservation law)，它保持概率守恒，也被称为连续性方程 (continuity equation)。

6.4.4 向前、向后方程

当系数满足一定光滑性要求时，多维扩散过程的 Kolmogorov 向前方程 (第二方程，Fokker-Planck 方程) 为

$$\frac{\partial p(s,x;t,y)}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i \partial y_j}\Big(a_{ij}(t,y)p(s,x;t,y)\Big) - \sum_{i=1}^{d}\frac{\partial}{\partial y_i}\Big(b_i(t,y)p(s,x;t,y)\Big), \tag{87}$$

其中，$(a_{ij})_{1\leq i,j\leq d} = \Sigma\Sigma^T$.

其非散度形式为

$$\frac{\partial p(s,x;t,y)}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,y)\frac{\partial^2 p(s,x;t,y)}{\partial y_i \partial y_j} + \sum_{i=1}^{d}\tilde{b_i}(t,y)\frac{\partial p(s,x;t,y)}{\partial y_i} + \tilde{c}(t,y)p(s,x;t,y), \tag{88}$$

其中，

$$\tilde{b_i}(t,y) = -b_i(t,y) + \sum_{j=1}^{d}\frac{\partial a_{ij}(t,y)}{\partial y_j},$$

$$\tilde{c}(t,y)=\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 a_{ij}(t,y)}{\partial y_i\partial y_j}-\sum_{i=1}^{d}\frac{\partial b_i(t,y)}{\partial y_i}.$$

Kolmogorov 向后方程 (第一方程) 为

$$-\frac{\partial p(s,x;t,y)}{\partial s}=\frac{1}{2}\sum_{i,j=1}^{d}a_{ij}(t,x)\frac{\partial^2 p(s,x;t,y)}{\partial x_i\partial x_j}+\sum_{i=1}^{d}b_i(t,x)\frac{\partial p(s,x;t,y)}{\partial x_i}. \tag{89}$$

6.5 Feynman-Kac 公式

6.5.1 Feynman-Kac 公式 I (时间齐次情形))

定理 6.7. 设

$$d\xi_t=\mathbf{b}(\xi_t)dt+\Sigma(\xi_t)d\mathbf{B}_t,$$

记 $\mathcal{L}=\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(x)\frac{\partial}{\partial x_i}$, $(a_{ij})_{i,j=1,\cdots,n}=\Sigma\Sigma^T$. 设 $f\in C_0^2(\mathbb{R}^n)$, $q(x)\in C(\mathbb{R}^n)$, $q(x)$ 有下界。

1. 设 $v(t,x)=E\Big(\exp\big(-\int_0^t q(\xi_s)ds\big)f(\xi_t)\Big|\xi_0=x\Big)$, 则

$$\begin{cases}\frac{\partial v}{\partial t}=\mathcal{L}v-qv & t>0,\ x\in\mathbb{R}^n\\ v(0,x)=f(x) & x\in\mathbb{R}^n\end{cases}. \tag{90}$$

2. 若 $w(t,x)\in C^{1,2}(\mathbb{R}\times\mathbb{R}^n)$ 在 $K\times\mathbb{R}^n$ 上有界, $K\subset\subset\mathbb{R}$, w 为 (90) 的解, 则 $w(t,x)=v(t,x)$.

注记 6.8. 当 $q(x)=0, x\in\mathbb{R}^n$ 时, 上述定理实质是时间齐次扩散过程的 Kolmogorov 向后方程.

引理 6.1. 设 $v(t,x)=E\Big(f(\xi_t)\exp\big(-\int_0^t q(\xi_s)ds\big)\Big|\xi_0=x\Big)$, 则

$$E\Big(f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta}g(\xi_s)ds\big)\Big|\mathcal{F}_\Delta^{\mathbf{B}}\Big)=v(t,\xi_\Delta).$$

引理证明：

$$\begin{aligned}&E\Big[f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta}g(\xi_s)ds\big)\Big|\mathcal{F}_\Delta^{\mathbf{B}}\Big]\\ \overset{\text{马氏性}}{=}\ &E\Big[f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta}g(\xi_s)ds\big)\Big|\xi_\Delta\Big]\\ =\ &E\Big(f(\xi_{t+\Delta})\exp\big(-\int_\Delta^{t+\Delta}g(\xi_s)ds\big)\Big|\xi_\Delta=y\Big)\Big|_{y=\xi_\Delta}\\ \overset{\text{时齐性}}{=}\ &E\Big(f(\xi_t)\exp\big(-\int_0^t g(\xi_s)ds\big)\Big|\xi_0=y\Big)\Big|_{y=\xi_\Delta}\\ =\ &v(t,y)\Big|_{y=\xi_\Delta}\\ =\ &v(t,\xi_\Delta).\end{aligned}$$

□

定理 6.8. $D \subset \mathbb{R}^n$ 为一个有界开区域, $q(x)$ 连续下有界, 考虑如下偏微分方程的边值问题

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i} - q(x)u = -g(x) & x \in D \\ u(x) = \varphi(x) & x \in \partial D \end{cases}, \tag{91}$$

则

$$u(x) = E\Big(\varphi(\xi_{\tau_x})e^{-\int_0^{\tau_x} q(\xi_s)ds}\Big|\xi_0 = x\Big) + E\Big(\int_0^{\tau_x} g(\xi_t)\exp\Big(-\int_0^t q(\xi_s)ds\Big)dt\Big|\xi_0 = x\Big), \tag{92}$$

其中

$$d\xi_t = \mathbf{b}(\xi_t)dt + \Sigma(\xi_t)d\mathbf{B}_t, \quad (a_{ij}) = \Sigma\Sigma^T.$$

$$\tau_x = \inf\{t \geq 0, \xi_t^x \notin \overline{D}\}.$$

在一定条件下:

1. 当 $q(x) = 0$ 时，$g(x) = 0$, 方程 (91) 和 $u(x)$ 的表达式 (92) 为

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i} = 0 & x \in D \\ u(x) = \varphi(x) & x \in \partial D \end{cases}, \tag{93}$$

$$u(x) = E\big(\varphi(\xi_{\tau_x})\big|\xi_0 = x\big).$$

特别地, 设 $\Gamma \in \partial D$ 是一个 Borel 可测集合, 取 $\varphi(x) = \mathbf{1}_\Gamma(x)$, 则

$$P(\xi_{\tau_x} \in \Gamma\big|\xi_0 = x) = u(x).$$

这也就是说, 可以利用偏微分方程 (93) 的解给出扩散过程 ξ_t 首达 ∂D 地点的分布.

2. 当 $q(x) = 0$, $g(x) = 1$, $\varphi(x) = 0$, 方程 (91) 和 $u(x)$ 的表达式 (92) 为

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i} = -1 & x \in D \\ u(x) = 0 & x \in \partial D \end{cases}, \tag{94}$$

$$u(x) = E\big(\tau_x\big|\xi_0 = x\big). \tag{95}$$

一方面, 可以认为偏微分方程 (94) 的 Dirichlet 边值问题可以有概率表达式 (95); 另一方面，也可以认为偏微分方程 (94) 的 Dirichlet 边值问题的解给出了扩散过程 ξ_t 首次到达 ∂D 停时 τ_x 的期望的表达式.

3. 进一步, 取 $\varphi(x) = 0$, $q(x) \geq 0$, 考虑如下偏微分方程的 Dirichlet 边值问题

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial u}{\partial x_i} - q(x)u = -g(x) & x \in D \\ u(x) = 0 & x \in \partial D \end{cases},$$

则

$$u(x) = E\Big(\int_0^{\tau_x} g(\xi_t)\exp\Big(-\int_0^t q(\xi_s)ds\Big)dt\Big|\xi_0 = x\Big). \tag{96}$$

那么, 利用 (96) 可以给出 Feynman-Kac 公式的概率含义.

6.5.2 Feynman-Kac 公式 II ((时间非齐次情形))

设 $\mathbf{b}, \Sigma$ 满足全局 Lipschitz 条件和线性增长条件。

$$d\xi_t = \mathbf{b}(t, \xi_t)dt + \Sigma(t, \xi_t)d\mathbf{B}_t, \tag{97}$$

$$\mathcal{L}_t = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(t,x)\frac{\partial}{\partial x_i} \tag{98}$$

定理 6.9. 设 $u(t,x)$ 关于 x 二次连续可微，关于 t 一次连续可微且满足 $\forall t \in (0,T]$,

$$\begin{cases} \mathcal{L}_t u(x,t) + \frac{\partial u}{\partial t}(x,t) - q(t,x)u(x,t) = -g(x,t) \\ u(x,T) = f(x) \end{cases}. \tag{99}$$

则

$$u(x,t) = E\Big(f(\xi(T))e^{-\int_t^T q(\xi_s,s)ds} + \int_t^T g(\xi_s,s)e^{-\int_t^s q(\xi_\theta,\theta)d\theta}ds\Big|\xi_t = x\Big) \tag{100}$$