

积分、概率与随机过程

独立中文讲义

Jean-François Le Gall

依据 2006 年 9 月课程讲义翻译、重新排版

中文译校与排版: Yuguang XIAO (肖煜光)

原始法文讲义:

<https://www.imo.universite-paris-saclay.fr/~jean-francois.le-gall/IPPA2.pdf>

阅读说明

本书是原讲义的独立中文版本，不与法文逐段对照。正文保留原讲义的章节结构、定理编号、证明顺序和主要记号；对原稿中可以严格确认的排印错误，已在正文中改正，并以“校订注”说明。

适合的读者与预备知识。本书面向已经学过一元与多元微积分、线性代数，并接触过度量空间基本概念的本科生。第一次学习时，建议按“定义—例子—命题—证明”的顺序阅读；带星号或明显偏深入的证明，可以在掌握主线后再读。

记号与阅读约定

- $\mathbb{N}$ 含 0； $\mathcal{B}(E)$ 表示拓扑空间 E 的 Borel σ -代数。
- “几乎处处”总是相对于当前测度而言；“几乎必然”专用于概率测度。
- 原稿中的区间记号 $]a, b[$ 、 $]a, b]$ 在公式中予以保留，分别等同于 (a, b) 、 $(a, b]$ 。这样既忠于原稿，也便于核对端点。
- 1_A 表示集合 A 的示性函数； δ_x 表示点 x 处的 Dirac 测度。
- 测度论中，只取有限多个值的可测函数称为“简单函数”；黎曼积分语境中的分段常值函数称为“阶梯函数”。

建议的学习主线。第一遍可重点掌握： σ -代数与测度、Lebesgue 积分及收敛定理、乘积测度、随机变量与独立性、条件期望、鞅、马尔可夫链和布朗运动。每读完一个定理，先确认假设中哪些条件用于“存在”、哪些用于“有限性”或“极限交换”，再进入证明细节。

目录

阅读说明	ii
第一部分 积分	1
第一章 测度空间	2
1.1 可测集合	2
1.2 正测度	3
1.3 可测函数	5
1.4 单调类	6
第二章 关于测度的积分	9
2.1 非负函数的积分	9
2.2 可积函数	14
2.3 含参数的积分	16
第三章 测度的构造	19
3.1 外测度	19
3.2 Lebesgue 测度	21
3.3 与黎曼积分的联系	25
3.4 不可测集合的一个例子	26
3.5 斯蒂尔切斯积分	27
3.6 里斯表示定理	28
第四章 L^p 空间	30
4.1 定义与 Hölder 不等式	30
4.2 Banach 空间 L^p	32
4.3 L^p 空间中的稠密性定理	35
4.4 Radon-Nikodym 定理	37
第五章 乘积测度	40
5.1 乘积空间概述	40
5.2 乘积测度的构造	41
5.3 Fubini 定理	42
5.4 应用	45

5.4.1	分部积分	45
5.4.2	卷积	45
5.4.3	单位球体积的计算	48
第六章	有符号测度	50
6.1	定义与全变差	50
6.2	Jordan 分解	52
6.3	L^p - L^q 对偶	54
6.4	里斯表示定理	57
第七章	变量替换公式与补充内容	58
7.1	变量替换公式	58
7.2	单位球面上的 Lebesgue 测度	61
第二部分	概率	64
第八章	概率论基础	65
8.1	一般定义	65
8.1.1	概率空间	65
8.1.2	随机变量	66
8.1.3	数学期望	67
8.1.4	例: Bertrand 悖论	68
8.1.5	常见分布	69
8.1.6	由随机变量生成的 σ -代数	71
8.2	随机变量的矩	71
8.2.1	p 阶矩与方差	71
8.2.2	线性回归	73
8.2.3	特征函数	74
8.2.4	母函数	77
第九章	独立性	78
9.1	独立事件	78
9.2	独立随机变量与独立 σ -代数	79
9.3	Borel–Cantelli 引理	84
9.4	独立随机变量之和	86
第十章	随机变量的收敛	91
10.1	不同的收敛概念	91
10.2	强大数律	93

10.3 依分布收敛	96
10.4 两个应用	101
10.4.1 经验测度的收敛	101
10.4.2 中心极限定理	102
10.4.3 向量情形的推广	103
第十一章 条件期望	106
11.1 离散条件化	106
11.2 条件期望的定义	108
11.2.1 可积随机变量的情形	108
11.2.2 非负随机变量的情形	109
11.2.3 平方可积随机变量的特殊情形	112
11.3 条件期望的特殊性质	112
11.4 条件期望的计算	114
11.4.1 离散条件化	114
11.4.2 具有密度的随机变量	115
11.4.3 高斯条件化	116
11.5 转移概率与条件分布	118
第三部分 随机过程	121
第十二章 离散时间鞅论	122
12.1 定义与例	122
12.2 停时	125
12.3 鞅的几乎必然收敛	127
12.4 $p > 1$ 时的 L^p 收敛	132
12.5 一致可积性与鞅	136
12.6 反向鞅	141
第十三章 马尔可夫链	146
13.1 定义与基本性质	146
13.2 若干例子	148
13.2.1 独立随机变量	148
13.2.2 $\mathbb{Z}^d$ 上的随机游走	148
13.2.3 图上的简单随机游走	149
13.2.4 分枝过程	149
13.3 典范马尔可夫链	149
13.4 状态分类	153

13.5 不变测度	159
13.6 渐近行为	164
13.7 鞅与马尔可夫链	167
第十四章 布朗运动导论	170
14.1 作为随机游走极限的布朗运动	170
14.2 布朗运动的构造	172
14.3 Wiener 测度	175
14.4 布朗运动的基本性质	176
14.5 强马尔可夫性质	179
14.6 调和函数与 Dirichlet 问题	182
14.7 调和函数与布朗运动	189

第一部分

积分

第一章 测度空间

测度论的出发点，是给定一个集合，并试图为它的每个子集赋予一个非负实数（称为该子集的测度），同时满足若干自然的可加性要求；例如，两两不交的子集之并的测度应等于各子集测度之和。由于深刻的原因，一般不可能为所有子集同时定义这样的测度，因此必须只考虑某一类子集——可测子集；这类子集构成一个 σ -代数。配备了 σ -代数的集合称为可测空间。本章介绍 σ -代数（即可测集合族）、可测空间上的测度以及可测函数等基本概念；稍后将对可测函数定义积分。最后一节给出单调类引理的一种形式，它在测度论和概率论中都十分重要。

1.1 可测集合

定义 1.1.1 设 E 为任意集合。 E 上的一个 σ -代数，是指 E 的一个子集族 $\mathcal{A}$ ，满足：

$$(i) E \in \mathcal{A};$$

$$(ii) A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A};$$

$$(iii) \text{ 若 } A_n \in \mathcal{A} \text{ 对每个 } n \in \mathbb{N} \text{ 都成立, 则 } \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}.$$

$\mathcal{A}$ 的元素称为可测集；需要强调所用 σ -代数时，也称为 $\mathcal{A}$ -可测集。二元组 $(E, \mathcal{A})$ 称为可测空间。

由定义立即得到：

$$\emptyset \in \mathcal{A}.$$

$$(2) \text{ 若 } A_n \in \mathcal{A} \ (n \in \mathbb{N}), \text{ 则 } \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}.$$

(3) 在可数并中令充分大的 n 所对应的 $A_n = \emptyset$ ，便知 $\mathcal{A}$ 对有限并封闭；由 De Morgan 公式，它也对有限交封闭。

例。

$$\bullet \mathcal{A} = \mathcal{P}(E);$$

$\bullet \mathcal{A} = \{\emptyset, E\}$ 是平凡 σ -代数；

$\bullet E$ 的所有至多可数子集，以及补集至多可数的子集，共同组成 E 上的一个 σ -代数。

任意一族 σ -代数的交仍是 σ -代数。由此得到下一个定义。

定义 1.1.2 设 $\mathcal{C} \subset \mathcal{P}(E)$ 。存在唯一的、包含 $\mathcal{C}$ 的最小 σ -代数，记为 $\sigma(\mathcal{C})$ ，并可写成

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{A} \text{ 为 } \sigma\text{-代数} \\ \mathcal{C} \subset \mathcal{A}}} \mathcal{A}.$$

$\sigma(\mathcal{C})$ 称为由 $\mathcal{C}$ 生成的 σ -代数。

Borel σ -代数。 设 E 是拓扑空间，这是生成 σ -代数的第一个重要例子。

定义 1.1.3 设 $\mathcal{O}$ 为拓扑空间 E 的全体开集。由 $\mathcal{O}$ 生成的 σ -代数 $\sigma(\mathcal{O})$ 称为 E 的 Borel σ -代数，记为 $\mathcal{B}(E)$ 。

所以， $\mathcal{B}(E)$ 是包含 E 的所有开集的最小 σ -代数；其中的元素称为 E 的 Borel 集。以后考察 $\mathbb{R}$ 、 $\mathbb{R}^d$ 等拓扑空间时，除非另有说明，都默认配备其 Borel σ -代数。

练习。 证明 $\mathcal{B}(\mathbb{R})$ 可由下列任一集合族生成：

$$\{(a, b) : a, b \in \mathbb{R}, a < b\}, \quad \{(-\infty, a) : a \in \mathbb{R}\}, \quad \{(-\infty, a) : a \in \mathbb{Q}\}.$$

也可把这些开区间换成相应的闭区间。

积 σ -代数。 这是生成 σ -代数的第二个重要例子。

定义 1.1.4 设 $(E_1, \mathcal{A}_1)$ 、 $(E_2, \mathcal{A}_2)$ 是两个可测空间。定义 $E_1 \times E_2$ 上的积 σ -代数为

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}).$$

练习。 证明

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}).$$

1.2 正测度

本节固定一个可测空间 $(E, \mathcal{A})$ 。

定义 1.2.1 $(E, \mathcal{A})$ 上的正测度，是满足下列条件的映射 $\mu : \mathcal{A} \rightarrow [0, \infty]$ ：

$$(i) \mu(\emptyset) = 0;$$

(ii) 对任意两两不交的可测集列 $(A_n)_{n \in \mathbb{N}}$,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

允许测度取 $+\infty$ 十分重要。条件 (ii) 称为可列可加性或 σ -可加性。若从某一项起 $A_n = \emptyset$ ，它就退化为有限可加性。

基本性质。 (1) 若 $A \subset B$ ，则 $\mu(A) \leq \mu(B)$ ；若还满足 $\mu(A) < \infty$ ，则

$$\mu(B \setminus A) = \mu(B) - \mu(A);$$

(2) 若 $A, B \in \mathcal{A}$ ，则

$$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B);$$

(3) 若 $A_n \in \mathcal{A}$ 且 $A_n \subset A_{n+1}$ ，则

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mu(A_n);$$

(4) 若 $B_n \in \mathcal{A}$, $B_{n+1} \subset B_n$, 且 $\mu(B_0) < \infty$, 则

$$\mu\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \downarrow \mu(B_n);$$

(5) 若 $A_n \in \mathcal{A}$, 则

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

只证明 (3) — (5)。证明 (3) 时, 令 $C_0 = A_0$, 且对 $n \geq 1$ 令

$$C_n = A_n \setminus A_{n-1}$$

于是 $\bigcup_n A_n = \bigcup_n C_n$, 而各 C_n 两两不交, 故

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \mu\left(\bigcup_{n \in \mathbb{N}} C_n\right) = \sum_{n \in \mathbb{N}} \mu(C_n) = \lim_{N \rightarrow \infty} \uparrow \sum_{n=0}^N \mu(C_n) = \lim_{N \rightarrow \infty} \uparrow \mu(A_N).$$

证明 (4) 时, 对每个 n 令 $A_n = B_0 \setminus B_n$, 则 (A_n) 递增。因此

$$\mu(B_0) - \mu\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \mu\left(B_0 \setminus \bigcap_{n \in \mathbb{N}} B_n\right) = \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mu(A_n) = \lim_{n \rightarrow \infty} \uparrow (\mu(B_0) - \mu(B_n)).$$

这里 $\mu(B_0) < \infty$ 正是为了合法地写出 $\mu(A_n) = \mu(B_0) - \mu(B_n)$ 。最后证明 (5): 令 $C_0 = A_0$, 并对 $n \geq 1$ 令

$$C_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k.$$

各 C_n 两两不交, 而且 $C_n \subset A_n$, 所以

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \mu\left(\bigcup_{n \in \mathbb{N}} C_n\right) = \sum_{n \in \mathbb{N}} \mu(C_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

例。 (1) 若 $E = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$, 则计数测度定义为

$$\mu(A) = \text{Card}(A).$$

对任意集合 E , 也可在 $(E, \mathcal{P}(E))$ 上定义计数测度。这个例子还说明性质 (4) 中的条件 $\mu(B_0) < \infty$ 不能删去: 取

$$B_n = \{n, n+1, n+2, \dots\}$$

则每个 $\mu(B_n) = \infty$, 但 $\bigcap_n B_n = \emptyset$, 所以 $\mu(\bigcap_n B_n) = 0$ 。(2) 设 $(E, \mathcal{A})$ 是任意可测空间, $x \in E$ 。定义 x 处的 Dirac 测度为

更一般地, 若 $x_n \in E$, $\alpha_n \in [0, \infty]$ ($n \in \mathbb{N}$), 则可定义正测度 $\sum_n \alpha_n \delta_{x_n}$:

$$(\sum \alpha_n \delta_{x_n})(A) = \sum \alpha_n \delta_{x_n}(A) = \sum \alpha_n 1_A(x_n).$$

(3)**Lebesgue 测度**。在 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上存在唯一正测度 λ , 使每个开区间 (a, b) 都满足 $\lambda((a, b)) = b - a$ 。它的存在性和唯一性将在后文证明。

若干术语。 $\mu(E) < \infty$ 时称 μ 为有限测度, $\mu(E)$ 称为总质量; $\mu(E) = 1$ 时称 μ 为概率测度。若存在递增可测集列 (E_n) , 使 $E = \bigcup_n E_n$ 且 $\mu(E_n) < \infty$, 则称 μ 为 σ -有限测度。

若 $\{x\} \in \mathcal{A}$ 且 $\mu(\{x\}) > 0$, 则称 x 是 μ 的原子。没有原子的测度称为无原子测度 (或弥散测度)。

1.3 可测函数

定义 1.3.1 设 $(E, \mathcal{A})$ 、 $(F, \mathcal{B})$ 是可测空间。映射 $f: E \rightarrow F$ 称为可测映射, 如果

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

若 E, F 是配备 Borel σ -代数的拓扑空间, 也称 f 为 Borel 可测映射。

命题 1.3.1 两个可测映射的复合仍可测。

证明。若 $f: E \rightarrow F$ 、 $g: F \rightarrow G$ 可测, 则对每个可测集 $C \subset G$,

$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$$

是 E 中的可测集。

命题 1.3.2 要证明 $f: E \rightarrow F$ 可测, 只需找到 $\mathcal{C} \subset \mathcal{B}$, 使 $\sigma(\mathcal{C}) = \mathcal{B}$, 并验证对每个 $B \in \mathcal{C}$ 都有 $f^{-1}(B) \in \mathcal{A}$ 。

证明。令

$$\mathcal{G} = \{B \in \mathcal{B} : f^{-1}(B) \in \mathcal{A}\}.$$

容易验证 $\mathcal{G}$ 是 σ -代数。假设给出 $\mathcal{C} \subset \mathcal{G}$, 故 $\mathcal{B} = \sigma(\mathcal{C}) \subset \mathcal{G}$, 结论成立。

例。(1) 当 $(F, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 时, 要证明 f 可测, 只需验证 $f^{-1}((a, b))$ 对所有 $a < b$ 可测; 甚至只验证 $f^{-1}((-\infty, a))$ 对所有 $a \in \mathbb{R}$ 可测即可。(2) 若 E, F 都配备 Borel σ -代数, 则连续映射 $f: E \rightarrow F$ 一定可测, 因为开集的原像仍是开集。

可测函数的运算。

引理 1.3.3 设 $f_1: (E, \mathcal{A}) \rightarrow (F_1, \mathcal{B}_1)$ 、 $f_2: (E, \mathcal{A}) \rightarrow (F_2, \mathcal{B}_2)$ 可测。则乘积映射

$$f: (E, \mathcal{A}) \longrightarrow (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2), \quad f(x) = (f_1(x), f_2(x))$$

也可测。

证明。应用命题 1.3.2, 并取生成族

$$\mathcal{C} = \{B_1 \times B_2; B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}.$$

因为

$$f^{-1}(B_1 \times B_2) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \in \mathcal{A},$$

所以 f 可测。反过来, 若乘积映射 $f = (f_1, f_2)$ 可测, 则坐标投影连续, 从而 f_1, f_2 也可测。

推论 1.3.4 若 $f, g : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 可测, 则 $f + g$ 、 fg 、 $\inf(f, g)$ 、 $f^+ = \sup(f, 0)$ 以及 $f^- = \sup(-f, 0)$ 都可测。

例如, $f + g$ 是可测映射 $x \mapsto (f(x), g(x))$ 与连续映射 $(a, b) \mapsto a + b$ 的复合。再回忆: 若 $(a_n) \subset \mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}$, 则定义

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} a_k \right), \quad \liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} a_k \right).$$

其中的单调极限都在 $\overline{\mathbb{R}}$ 中理解。 $\limsup a_n$ 与 $\liminf a_n$ 分别是数列 (a_n) 的最大聚点与最小聚点。

命题 1.3.5 若 (f_n) 是从 E 到 $\overline{\mathbb{R}}$ 的可测函数列, 则

$$\sup_n f_n, \quad \inf_n f_n, \quad \limsup_{n \rightarrow \infty} f_n, \quad \liminf_{n \rightarrow \infty} f_n$$

也都可测。特别地, 若 f_n 逐点收敛, 则其极限可测; 一般地, 集合 $\{x \in E : \lim_n f_n(x) \text{ 存在}\}$ 也是可测集。

证明。 令 $f(x) = \inf_n f_n(x)$ 。对任意 $a \in \mathbb{R}$,

$$f^{-1}([-\infty, a)) = \{x : \inf_n f_n(x) < a\} = \bigcup_n \{x : f_n(x) < a\}.$$

右端可测, 故 f 可测; $\sup_n f_n$ 同理。因此

$$\liminf_{n \rightarrow \infty} f_n = \sup_{n \geq 0} \left(\inf_{k \geq n} f_k \right).$$

也可测。最后令 $G(x) = (\liminf_n f_n(x), \limsup_n f_n(x))$ 。则

像测度。

定义 1.3.2 设 $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ 可测, μ 是 $(E, \mathcal{A})$ 上的正测度。称 μ 在 f 下的像测度为 $f(\mu)$, 它是 $(F, \mathcal{B})$ 上由下式定义的正测度:

$$f(\mu)(B) = \mu(f^{-1}(B)).$$

容易验证上式确实定义了测度。并且 $f(\mu)(F) = \mu(E)$, 所以像测度与原测度总质量相同。但 μ 为 σ -有限测度时, $f(\mu)$ 未必 σ -有限。

1.4 单调类

定义 1.4.1 集合族 $\mathcal{M} \subset \mathcal{P}(E)$ 称为单调类, 如果

$$(i) \ E \in \mathcal{M};$$

- (ii) 若 $A, B \in \mathcal{M}$ 且 $A \subset B$, 则 $B \setminus A \in \mathcal{M}$;
 (iii) 若 $A_n \in \mathcal{M}$ 且 $A_n \subset A_{n+1}$, 则 $\bigcup_n A_n \in \mathcal{M}$ 。

每个 σ -代数都是单调类; 任意一族单调类的交仍是单调类。因此, 对 $\mathcal{C} \subset \mathcal{P}(E)$, 可定义由 $\mathcal{C}$ 生成的最小单调类

$$\mathcal{M}(\mathcal{C}) = \bigcap_{\substack{\mathcal{M} \text{ 为单调类} \\ \mathcal{C} \subset \mathcal{M}}} \mathcal{M}.$$

定理 1.4.1 (单调类引理) 若 $\mathcal{C} \subset \mathcal{P}(E)$ 对有限交封闭, 则 $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$ 。

证明。 每个 σ -代数都是单调类, 所以 $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$ 。反过来, 只需证明 $\mathcal{M}(\mathcal{C})$ 是 σ -代数。一个单调类若对有限交封闭, 就对补集、有限并以及可数并封闭, 因而是 σ -代数。因此只需证明 $\mathcal{M}(\mathcal{C})$ 对有限交封闭。先固定 $A \in \mathcal{C}$, 并令

$$\mathcal{M}_1 = \{B \in \mathcal{M}(\mathcal{C}) : A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

由于 $\mathcal{C}$ 对有限交封闭, $\mathcal{C} \subset \mathcal{M}_1$ 。下面验证 $\mathcal{M}_1$ 是单调类:

- $E \in \mathcal{M}_1$;
- 若 $B, B' \in \mathcal{M}_1$ 且 $B \subset B'$, 则

$$A \cap (B' \setminus B) = (A \cap B') \setminus (A \cap B) \in \mathcal{M}(\mathcal{C}),$$

从而

$$B' \setminus B \in \mathcal{M}_1.$$

- 若 $B_n \in \mathcal{M}_1$ 且 $B_n \uparrow$, 则

$$A \cap \bigcup_n B_n = \bigcup_n (A \cap B_n) \in \mathcal{M}(\mathcal{C}),$$

故 $\bigcup_n B_n \in \mathcal{M}_1$ 。

因此 $\mathcal{M}_1$ 是包含 $\mathcal{C}$ 的单调类, 故 $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_1$ 。也就是说,

$$\forall A \in \mathcal{C}, \forall B \in \mathcal{M}(\mathcal{C}), A \cap B \in \mathcal{M}(\mathcal{C}).$$

还需把两个变量的角色交换。固定 $B \in \mathcal{M}(\mathcal{C})$, 并令

$$\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{C}) : A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

第一步说明 $\mathcal{C} \subset \mathcal{M}_2$ 。用完全相同的论证可知 $\mathcal{M}_2$ 是单调类, 因此 $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_2$ 。这说明任意 $A, B \in \mathcal{M}(\mathcal{C})$ 都满足 $A \cap B \in \mathcal{M}(\mathcal{C})$, 定理得证。

推论 1.4.2 设 μ, ν 是 $(E, \mathcal{A})$ 上的两个测度。假设 $\mathcal{C} \subset \mathcal{A}$ 对有限交封闭, $\sigma(\mathcal{C}) = \mathcal{A}$, 且 $\mu(A) = \nu(A)$ 对每个 $A \in \mathcal{C}$ 成立。

- (1) 若 $\mu(E) = \nu(E) < \infty$, 则 $\mu = \nu$ 。
- (2) 若存在递增集合列 $E_n \in \mathcal{C}$, 使 $E = \bigcup_n E_n$ 且 $\mu(E_n) = \nu(E_n) < \infty$, 则 $\mu = \nu$ 。

证明。(1) 令 $\mathcal{G} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$ 。由假设 $\mathcal{C} \subset \mathcal{G}$ 。容易验证 $\mathcal{G}$ 是单调类。例如, 若 $A, B \in \mathcal{G}$ 且 $A \subset B$, 则

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A),$$

从而 $B \setminus A \in \mathcal{G}$ 。¹这里用到了 μ 与 ν 都是有限测度。因此 $\mathcal{G}$ 包含 $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A}$, 故 $\mathcal{G} = \mathcal{A}$, 即 $\mu = \nu$ 。(2) 对每个 n , 记 μ_n, ν_n 分别为 μ, ν 在 E_n 上的限制:

$$\forall A \in \mathcal{A}, \mu_n(A) = \mu(A \cap E_n), \nu_n(A) = \nu(A \cap E_n).$$

由 (1) 得 $\mu_n = \nu_n$ 。于是对任意 $A \in \mathcal{A}$,

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A \cap E_n) = \lim_{n \rightarrow \infty} \nu(A \cap E_n) = \nu(A).$$

推论: Lebesgue 测度的唯一性。至多存在一个 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的测度 λ , 使每个非空开区间 (a, b) 都满足 $\lambda((a, b)) = b - a$ 。事实上, 若 λ' 也有此性质, 在推论 (2) 中取

$$\mathcal{C} = \{\emptyset\} \cup \{(a, b) : a, b \in \mathbb{R}, a < b\}, \quad E_n = (-n, n),$$

即可得到 $\lambda = \lambda'$ 。同理, $\mathbb{R}$ 上的有限测度 μ 由分布函数值 $\mu((-\infty, a])$ ($a \in \mathbb{R}$) 唯一确定。

¹校订注: 原文此处把结论误排为 $B \setminus A \in E$; 按本段所定义的集合族, 应为 $B \setminus A \in \mathcal{G}$ 。

第二章 关于测度的积分

本章首先构造可测函数的积分。对只取有限多个值的简单函数，积分可以直接定义；再把非负可测函数的积分定义为所有不超过它的非负简单函数积分的上确界。对可取正负值的函数，则先要求其绝对值积分有限，再利用正部与负部定义积分。随后将证明三个基本收敛结果：单调收敛定理、Fatou 引理和支配收敛定理。它们给出把极限与积分交换的常用充分条件。本章最后讨论含参数积分的连续性与可微性。

2.1 非负函数的积分

固定测度空间 $(E, \mathcal{A}, \mu)$ ，即 $(E, \mathcal{A})$ 是可测空间， μ 是其上的正测度。

简单函数。若可测函数 $f: E \rightarrow \mathbb{R}$ 只取有限多个值，则称 f 为简单函数。设它的不同取值为 $\alpha_1 < \alpha_2 < \cdots < \alpha_n$ ，则

$$f(x) = \sum_{i=1}^n \alpha_i 1_{A_i}(x)$$

其中 $A_i = f^{-1}(\{\alpha_i\}) \in \mathcal{A}$ ($1 \leq i \leq n$)。这称为 f 的规范表示。

定义 2.1.1 设简单函数 $f: E \rightarrow \mathbb{R}_+$ 的规范表示如上。定义 f 关于 μ 的积分为

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

若 $\alpha_i = 0$ 、 $\mu(A_i) = \infty$ ，采用约定 $0 \cdot \infty = 0$ 。

由定义， $\int f d\mu \in [0, \infty]$ 。若 f 还有另一种表示

$$f = \sum_{j=1}^m \beta_j 1_{B_j}$$

其中 $B_1, \dots, B_m$ 仍构成 E 的可测分割，但 β_j 不要求彼此不同，则仍有

$$\int f d\mu = \sum_{j=1}^m \beta_j \mu(B_j).$$

事实上，对每个 i ，集合 A_i 是所有满足 $\beta_j = \alpha_i$ 的 B_j 的不交并。因此由测度的可加性，

$$\mu(A_i) = \sum_{\{j: \beta_j = \alpha_i\}} \mu(B_j)$$

代回求和式即可得到结论。

性质。设 f, g 是非负简单函数。(1) 对任意 $a, b \geq 0$,

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

(2) 若 $f \leq g$, 则

$$\int f d\mu \leq \int g d\mu.$$

证明。 (1) 写出 f, g 的规范表示

$$f = \sum_{i=1}^n \alpha_i 1_{A_i}, g = \sum_{k=1}^m \alpha'_k 1_{A'_k}$$

把每个 A_i 与每个 A'_k 再按交集细分, 可找到同一组两两不交可测集 $B_1, \dots, B_p$, 使得

$$f = \sum_{j=1}^p \beta_j 1_{B_j}, g = \sum_{j=1}^p \gamma_j 1_{B_j}$$

其中 β_j, γ_j 不要求互异。根据定义后的说明,

$$\int f d\mu = \sum_{j=1}^p \beta_j \mu(B_j), \int g d\mu = \sum_{j=1}^p \gamma_j \mu(B_j).$$

并且

$$\int (af + bg) d\mu = \sum_{j=1}^p (a\beta_j + b\gamma_j) \mu(B_j),$$

线性性随即成立。

(2) 由 (1) 和 $g = f + (g - f)$ 得

$$\int g d\mu = \int f d\mu + \int (g - f) d\mu \geq \int f d\mu.$$

记 $\mathcal{E}_+$ 为所有非负简单函数组成的集合。

定义 2.1.2 设 $f: E \rightarrow [0, \infty]$ 可测。定义

$$\int f d\mu = \sup_{h \in \mathcal{E}_+, h \leq f} \int h d\mu.$$

上面的单调性说明, 当 f 本身是简单函数时, 这一定义与定义 2.1.1 一致。以下记号等价:

$$\int f d\mu = \int f(x) d\mu(x) = \int f(x) \mu(dx)$$

有时也写成 $\langle \mu, f \rangle$ 或 $\mu(f)$ 。

性质。 (1) 若 $0 \leq f \leq g$, 则 $\int f d\mu \leq \int g d\mu$ 。(2) 若 $\mu(\{f > 0\}) = 0$, 则 $\int f d\mu = 0$ 。这两点都可直接由定义得到。

定理 2.1.1 (单调收敛定理) 设 (f_n) 是非负可测函数列, $f_n \uparrow f$, 即对每个 x , $f_n(x)$ 单调递增并收敛于 $f(x)$ 。则

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

证明。由单调性立即有

$$\int f d\mu \geq \lim_{n \rightarrow \infty} \uparrow \int f_n d\mu.$$

只需证明反向不等式。任取满足 $0 \leq h \leq f$ 的非负简单函数

$$h = \sum_{i=1}^m \alpha_i 1_{A_i}$$

再固定 $a \in [0, 1)$ ，并令

$$E_n = \{x \in E : ah(x) \leq f_n(x)\}.$$

集合 E_n 可测，而且 $E_n \uparrow E$ 。确实，若 $h(x) = 0$ 则 $x \in E_n$ 对所有 n 成立；若 $h(x) > 0$ ，由 $ah(x) < h(x) \leq f(x)$ 与 $f_n(x) \uparrow f(x)$ ，可知 x 最终属于 E_n 。

又有 $f_n \geq a 1_{E_n} h$ ，所以

$$\int f_n d\mu \geq \int a 1_{E_n} h d\mu = a \sum_{i=1}^m \alpha_i \mu(A_i \cap E_n).$$

因为 $E_n \uparrow E$ ，所以对每个 i ， $A_i \cap E_n \uparrow A_i$ ，从而 $\mu(A_i \cap E_n) \uparrow \mu(A_i)$ 。令 $n \rightarrow \infty$ 得

$$\lim_{n \rightarrow \infty} \uparrow \int f_n d\mu \geq a \sum_{i=1}^m \alpha_i \mu(A_i) = a \int h d\mu.$$

再令 $a \uparrow 1$ ，得到

$$\lim_{n \rightarrow \infty} \uparrow \int f_n d\mu \geq \int h d\mu.$$

最后对所有满足 $0 \leq h \leq f$ 的非负简单函数取上确界，即得所需反向不等式。以后“非负可测函数”均指取值于 $[0, \infty]$ 的可测函数。

命题 2.1.2 (1) 每个非负可测函数 f 都存在非负简单函数列 (f_n) ，使 $f_n \uparrow f$ 。(2) 若 f, g 非负可测，且 $a, b \in \mathbb{R}_+$ ，则

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

(3) 若 (f_n) 是任意非负可测函数列，则

$$\int \sum_n f_n d\mu = \sum_n \int f_n d\mu.$$

证明。(1) 对每个 $n \geq 1$ 以及 $i \in \{0, 1, \dots, n2^n - 1\}$ ，令

$$A_n = \{x \in E : f(x) \geq n\}$$

$$B_{n,i} = \{x \in E : i2^{-n} \leq f(x) < (i+1)2^{-n}\}.$$

定义简单函数

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1_{B_{n,i}} + n 1_{A_n}.$$

直接验证可得 $f_n(x) \uparrow f(x)$ 对每个 $x \in E$ 成立。

(2) 由 (1), 可取非负简单函数列 (f_n) 、 (g_n) , 使 $f_n \uparrow f$ 、 $g_n \uparrow g$ 。于是 $af_n + bg_n \uparrow af + bg$ 。分别使用单调收敛定理和非负简单函数积分的线性性,

$$\int (af + bg) d\mu = \lim_{n \rightarrow \infty} \int (af_n + bg_n) d\mu = \lim_{n \rightarrow \infty} \left(a \int f_n d\mu + b \int g_n d\mu \right) = a \int f d\mu + b \int g d\mu.$$

(3) 该命题由 (2) (有限和的情况) 以及单调收敛定理推得。

注。 考虑特殊情形 $E = \mathbb{N}$, 并设 μ 为计数测度。此时容易看出

$$\int f d\mu = \sum_{k \in \mathbb{N}} f(k)$$

而性质 (3) 正好给出下列熟知结论: 对任意非负实数二重数列 $(a_{n,k})_{n,k \in \mathbb{N}}$,

$$\sum_{k \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{n,k} \right) = \sum_{n \in \mathbb{N}} \left(\sum_{k \in \mathbb{N}} a_{n,k} \right).$$

推论 2.1.3 设 f 为非负可测函数。对每个 $A \in \mathcal{A}$, 定义

$$\nu(A) = \int 1_A f d\mu := \int_A f d\mu.$$

则 ν 是 $(E, \mathcal{A})$ 上的正测度, 称为以 f 为关于 μ 的密度所定义的测度, 记为 $\nu = f \cdot \mu$ 。

证明。 显然 $\nu(\emptyset) = 0$ 。此外, 若 $(A_n)_{n \in \mathbb{N}}$ 是一列两两不交的可测集, 则

$$\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \int \sum_{n \in \mathbb{N}} 1_{A_n} f d\mu = \sum_{n \in \mathbb{N}} \int 1_{A_n} f d\mu = \sum_{n \in \mathbb{N}} \nu(A_n)$$

这里使用了上面的性质 (3)。

注。 若 $\mu(A) = 0$, 则

$$\nu(A) = \int 1_A f d\mu = 0.$$

若一个性质在某个 μ -零测集之外成立, 就称该性质 μ -几乎处处成立, 简记为 μ -a.e.; 在不会混淆时也简称“几乎处处”。例如, 对两个可测函数 f, g , $f = g$ μ -几乎处处的含义是

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0.$$

命题 2.1.4 设 f 为非负可测函数。(1) 对任意 $a > 0$,

$$\mu(\{x \in E : f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu.$$

(2) 我们有

$$\int f d\mu < \infty \implies f < \infty \quad \mu\text{-几乎处处.}$$

(3) 我们有

$$\int f d\mu = 0 \iff f = 0 \quad \mu\text{-几乎处处.}$$

(4) 如果 g 是另一个非负可测函数,

$$f = g \quad \mu\text{-几乎处处} \implies \int f d\mu = \int g d\mu.$$

证明。 (1) 令 $A_a = \{x \in E : f(x) \geq a\}$ 。于是 $f \geq a1_{A_a}$, 故

$$\int f d\mu \geq \int a1_{A_a} d\mu = a\mu(A_a).$$

(2) 对每个 $n \geq 1$, 令 $A_n = \{x \in E : f(x) \geq n\}$, 并令 $A_\infty = \{x \in E : f(x) = \infty\}$ 。由 (1),

$$\mu(A_\infty) = \mu\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mu(A_n) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \int f d\mu = 0.$$

(3) “ $\Leftarrow$ ”已经在前面证明。为证明“ $\Rightarrow$ ”, 对每个 $n \geq 1$ 令 $B_n = \{x \in E : f(x) \geq n^{-1}\}$ 。由 (1),

$$\mu(B_n) \leq n \int f d\mu = 0$$

于是 $\mu(B_n) = 0$, 从而 $\mu(\{x : f(x) > 0\}) = \mu\left(\bigcup_{n \geq 1} B_n\right) = 0$ 。

(4) 记 $f \vee g = \sup(f, g)$, $f \wedge g = \inf(f, g)$ 。若 $f = g$ 几乎处处, 则 $f \vee g = f \wedge g$ 几乎处处, 因而

$$\int (f \vee g) d\mu = \int (f \wedge g) d\mu + \int (f \vee g - f \wedge g) d\mu = \int (f \wedge g) d\mu,$$

因为 $f \vee g - f \wedge g = 0$ 几乎处处, 且 $f \wedge g \leq f \leq f \vee g$, 对 g 亦然, 故

$$\int f d\mu = \int (f \vee g) d\mu = \int g d\mu.$$

定理 2.1.5 (Fatou 引理) 设 $(f_n)_{n \geq 1}$ 是非负可测函数列。则

$$\int (\liminf_{n \rightarrow \infty} f_n) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

证明。 逐点有

$$\liminf_{n \rightarrow \infty} f_n = \lim_{k \rightarrow \infty} \uparrow \left(\inf_{n \geq k} f_n \right).$$

由单调收敛定理,

$$\int \liminf_{n \rightarrow \infty} f_n d\mu = \lim_{k \rightarrow \infty} \uparrow \int \left(\inf_{n \geq k} f_n \right) d\mu.$$

另一方面, 对每个 $p \geq k$, $\inf_{n \geq k} f_n \leq f_p$, 故

$$\int \left(\inf_{n \geq k} f_n \right) d\mu \leq \inf_{p \geq k} \int f_p d\mu.$$

令 $k \rightarrow \infty$, 即得 Fatou 不等式。

2.2 可积函数

定义 2.2.1 设 $f: E \rightarrow \mathbb{R}$ 可测。若

$$\int |f| d\mu < \infty,$$

则称 f 关于 μ 可积, 或称 f 为 μ -可积。此时定义

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu,$$

其中 $f^+ = f \vee 0$ 、 $f^- = (-f) \vee 0$ 分别是 f 的正部与负部。它们可测, 并满足

$$f = f^+ - f^-, \quad |f| = f^+ + f^-.$$

注。由 $\int f^\pm d\mu \leq \int |f| d\mu < \infty$, 上述差有定义。当 $f \geq 0$ 时, 这一定义与前一节的定义一致。记 $\mathcal{L}^1(E, \mathcal{A}, \mu)$ 为全体 μ -可积函数; 其非负部分有时记为 $\mathcal{L}_+^1(E, \mathcal{A}, \mu)$ 。

基本性质。

(a) 若 $f \in \mathcal{L}^1$, 则

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

(b) $\mathcal{L}^1(E, \mathcal{A}, \mu)$ 是向量空间, $f \mapsto \int f d\mu$ 是其上的线性泛函。

(c) 若 $f, g \in \mathcal{L}^1$ 且 $f \leq g$, 则

$$\int f d\mu \leq \int g d\mu.$$

(d) 若 $f, g \in \mathcal{L}^1$ 且 $f = g$, μ -几乎处处, 则

$$\int f d\mu = \int g d\mu.$$

证明。(a) 由三角不等式,

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \int f^+ d\mu + \int f^- d\mu = \int |f| d\mu.$$

(b) 若 $a \in \mathbb{R}$, 则 $\int |af| d\mu = |a| \int |f| d\mu < \infty$, 而直接区分 $a \geq 0$ 与 $a < 0$ 可得 $\int af d\mu = a \int f d\mu$ 。若 $f, g \in \mathcal{L}^1$, 则 $|f+g| \leq |f| + |g|$, 故 $f+g \in \mathcal{L}^1$ 。又由

$$(f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^-$$

以及非负函数积分的可加性, 得到

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu.$$

(c) 因为 $g-f \geq 0$,

$$\int g d\mu = \int f d\mu + \int (g-f) d\mu \geq \int f d\mu.$$

(d) 由 $f = g$ 几乎处处可知 $f^+ = g^+$ 、 $f^- = g^-$ 几乎处处, 再使用非负函数情形即可。

结合 (c)、(d) 还可知: 若 $f \leq g$ 几乎处处, 则 $\int f d\mu \leq \int g d\mu$ 。

复值函数。若可测函数 $f: E \rightarrow \mathbb{C}$ 满足

$$\int |f| d\mu < \infty,$$

则称 f 可积, 并记为 $f \in \mathcal{L}_{\mathbb{C}}^1(E, \mathcal{A}, \mu)$ 。

此时定义

$$\int f d\mu = \int \operatorname{Re}(f) d\mu + i \int \operatorname{Im}(f) d\mu.$$

上述性质 (a)、(b) 和 (d) 在把 $\mathcal{L}^1$ 替换为 $\mathcal{L}_{\mathbb{C}}^1$ 后仍成立。为证明 (a), 注意

$$\left| \int f d\mu \right| = \sup_{\substack{a \in \mathbb{C} \\ |a|=1}} a \cdot \int f d\mu = \sup_{\substack{a \in \mathbb{C} \\ |a|=1}} \int (a \cdot f) d\mu \leq \int |f| d\mu,$$

其中 $a \cdot z$ 表示把 $\mathbb{C}$ 识别为 $\mathbb{R}^2$ 后的实内积。

定理 2.2.1 (支配收敛定理) 设 (f_n) 是 $\mathcal{L}^1(E, \mathcal{A}, \mu)$ (或 $\mathcal{L}_{\mathbb{C}}^1(E, \mathcal{A}, \mu)$) 中的函数列。假设:

(1) 存在取值于 $\mathbb{R}$ (或 $\mathbb{C}$) 的可测函数 f , 使

$$f_n(x) \rightarrow f(x) \quad \mu\text{-几乎处处};$$

(2) 存在可测函数 $g: E \rightarrow \mathbb{R}_+$, 使 $\int g d\mu < \infty$, 且对所有 n ,

$$|f_n| \leq g \quad \mu\text{-几乎处处}.$$

则 $f \in \mathcal{L}^1$ (或 $f \in \mathcal{L}_{\mathbb{C}}^1$), 并且

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu, \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

证明。先假设下列更强的条件成立: 对每个 $x \in E$, $f_n(x) \rightarrow f(x)$; 并且存在可测函数 $g: E \rightarrow \mathbb{R}_+$, 满足 $\int g d\mu < \infty$, 使对所有 n 及 $x \in E$ 都有 $|f_n(x)| \leq g(x)$ 。因为 $|f| \leq g$, 所以 $f \in \mathcal{L}^1$ 。又因为 $|f - f_n| \leq 2g$ 且 $|f - f_n| \rightarrow 0$, 对非负函数 $2g - |f - f_n|$ 应用 Fatou 引理, 得

$$\liminf_{n \rightarrow \infty} \int (2g - |f - f_n|) d\mu \geq \int \liminf_{n \rightarrow \infty} (2g - |f - f_n|) d\mu = 2 \int g d\mu.$$

由积分的线性性,

$$2 \int g d\mu - \limsup_{n \rightarrow \infty} \int |f - f_n| d\mu \geq 2 \int g d\mu.$$

因此

$$\limsup_{n \rightarrow \infty} \int |f - f_n| d\mu = 0,$$

也就是 $\int |f - f_n| d\mu \rightarrow 0$ 。最后, 由

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu,$$

得到积分的收敛。

现在只假设定理中的 (1) 和 (2)。令

$$A = \{x \in E : f_n(x) \rightarrow f(x) \text{ 且对所有 } n, |f_n(x)| \leq g(x)\}.$$

则 $\mu(A^c) = 0$ 。定义

$$\tilde{f}_n(x) = 1_A(x)f_n(x), \quad \tilde{f}(x) = 1_A(x)f(x).$$

对这些函数可应用证明的第一部分。又因为 $f = \tilde{f}$ 几乎处处，且对每个 n ， $f_n = \tilde{f}_n$ 几乎处处，所以

$$\int f_n d\mu = \int \tilde{f}_n d\mu, \quad \int f d\mu = \int \tilde{f} d\mu, \quad \int |f_n - f| d\mu = \int |\tilde{f}_n - \tilde{f}| d\mu.$$

所求结论便由已经证明的情形推出。

2.3 含参数的积分

给定一个度量空间 (U, d) ，把它视为参数空间。再给定映射 $f : U \times E \rightarrow \mathbb{R}$ (或 $\mathbb{C}$)。

定理 2.3.1 设 $u_0 \in U$ 。^a假设：

- (i) 对每个 $u \in U$ ，函数 $x \mapsto f(u, x)$ 可测；
- (ii) 对 μ -几乎所有的 x ，函数 $u \mapsto f(u, x)$ 在 u_0 处连续；
- (iii) 存在 $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$ ，使对所有 $u \in U$ ，

$$|f(u, x)| \leq g(x) \quad \mu\text{-几乎处处}.$$

则

$$F(u) = \int f(u, x) \mu(dx)$$

对每个 $u \in U$ 都有定义，并且在 u_0 处连续。

^a校订注：原讲义此处把 u_0 误写为属于 E ；但 u_0 是参数点，正确条件应为 $u_0 \in U$ 。

证明。条件 (iii) 表明，对每个 $u \in U$ ，函数 $x \mapsto f(u, x)$ 可积，因而 $F(u)$ 有定义。设 $u_n \rightarrow u_0$ 。由 (ii)，

$$f(u_n, x) \rightarrow f(u_0, x) \quad \mu\text{-几乎处处}.$$

由支配条件 (iii) 及支配收敛定理，

$$\lim_{n \rightarrow \infty} \int f(u_n, x) \mu(dx) = \int f(u_0, x) \mu(dx).$$

因此 F 在 u_0 处连续。

例。(a) 设 μ 是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的无原子测度，且 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ 。则函数

$$F(u) = \int_{(-\infty, u]} \varphi(x) \mu(dx) = \int 1_{(-\infty, u]}(x) \varphi(x) \mu(dx)$$

连续。事实上，在定理 2.3.1 中取 $f(u, x) = 1_{(-\infty, u]}(x) \varphi(x)$ 与 $g = |\varphi|$ 。对固定的 $u_0 \in \mathbb{R}$ ，当 $x \neq u_0$ 时， $u \mapsto f(u, x)$ 在 u_0 处连续；而 $\mu(\{u_0\}) = 0$ 。

(b) **傅里叶变换。**若 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ ，则

$$\widehat{\varphi}(u) = \int e^{iux} \varphi(x) \lambda(dx)$$

在 $\mathbb{R}$ 上连续。

(c) **卷积。**设 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ ，且 $h : \mathbb{R} \rightarrow \mathbb{R}$ 有界连续。定义

$$(h * \varphi)(u) = \int h(u - x) \varphi(x) \lambda(dx).$$

则 $h * \varphi$ 连续且有界。

下面讨论积分号下求导。设 $U = I$ 是 $\mathbb{R}$ 中的开区间，并给定映射 $f: I \times E \rightarrow \mathbb{R}$ (或 $\mathbb{C}$)。

定理 2.3.2 (积分号下求导) 设 $u_0 \in I$ 。假设：

- (i) 对每个 $u \in I$, $x \mapsto f(u, x)$ 属于 $\mathcal{L}^1(E, \mathcal{A}, \mu)$;
- (ii) 对 μ -几乎所有的 x , 函数 $u \mapsto f(u, x)$ 在 u_0 处可导, 导数记为 $\frac{\partial f}{\partial u}(u_0, x)$;
- (iii) 存在 $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$, 使对每个 $u \in I$,

$$|f(u, x) - f(u_0, x)| \leq g(x)|u - u_0| \quad \mu\text{-几乎处处}.$$

则函数 $F(u) = \int f(u, x) \mu(dx)$ 在 u_0 处可导, 且

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x) \mu(dx).$$

注。根据 (ii), $\frac{\partial f}{\partial u}(u_0, x)$ 起初只在某个零测集的补集上定义。可将它在该零测集上任意延拓 (例如取值 0), 从而使给出 $F'(u_0)$ 的积分有定义。

证明。取 $I \setminus \{u_0\}$ 中任意收敛到 u_0 的序列 $(u_n)_{n \geq 1}$, 并令

$$\varphi_n(x) = \frac{f(u_n, x) - f(u_0, x)}{u_n - u_0}.$$

由 (ii), $\varphi_n(x) \rightarrow \frac{\partial f}{\partial u}(u_0, x)$ 对 μ -几乎所有的 x 成立。条件 (iii) 给出 $|\varphi_n(x)| \leq g(x)$ 几乎处处, 故支配收敛定理表明

$$\lim_{n \rightarrow \infty} \frac{F(u_n) - F(u_0)}{u_n - u_0} = \lim_{n \rightarrow \infty} \int \varphi_n(x) \mu(dx) = \int \frac{\partial f}{\partial u}(u_0, x) \mu(dx).$$

因为序列 (u_n) 任意, 结论成立。

注。在许多应用中, 条件 (ii) 和 (iii) 由下列更强条件代替:

- (ii)' 对 μ -几乎所有的 x , 函数 $u \mapsto f(u, x)$ 在 I 上可导;
- (iii)' 存在 $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$, 使对 μ -几乎所有的 x ,

$$\forall u \in I, \quad \left| \frac{\partial f}{\partial u}(u, x) \right| \leq g(x).$$

由中值定理, (iii)' 蕴含 (iii)。在这些条件下, F 在整个 I 上可导。不过, 下面的练习说明, 定理 2.3.2 所给出的较精细形式有时确实必要。

例。(a) 设 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, 且

$$\int |x\varphi(x)| \lambda(dx) < \infty.$$

则傅里叶变换 $\hat{\varphi}$ 在 $\mathbb{R}$ 上可导, 并且

$$\hat{\varphi}'(u) = i \int x e^{iux} \varphi(x) \lambda(dx).$$

(b) 设 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, 并设 $h: \mathbb{R} \rightarrow \mathbb{R}$ 属于 C^1 , 且 h, h' 都有界。则卷积 $h * \varphi$ 在 $\mathbb{R}$ 上可导, 并满足

$$(h * \varphi)' = h' * \varphi.$$

这个结论可以迭代使用。例如，若 $h \in C_c^\infty(\mathbb{R})$ ，则 $h * \varphi \in C^\infty(\mathbb{R})$ 。

练习。 设 μ 是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的无原子测度，并设 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ 满足

$$\int |x\varphi(x)| \mu(dx) < \infty.$$

对每个 $u \in \mathbb{R}$ ，定义

$$F(u) = \int_{\mathbb{R}} (u - x)^+ \varphi(x) \mu(dx).$$

证明 F 在 $\mathbb{R}$ 上可导，且

$$F'(u) = \int_{(-\infty, u]} \varphi(x) \mu(dx).$$

第三章 测度的构造

上一章假定测度已经给定。本章讨论如何真正构造测度，重点是 Lebesgue 测度。先引入外测度：它满足的条件比测度弱，却定义在所有子集上。Carathéodory 方法将从外测度中选出适当的 σ -代数，使外测度在其上成为真正的测度。随后讨论 Lebesgue 测度的性质、它与黎曼积分的关系，以及实轴上有限测度所对应的 Stieltjes 积分。

3.1 外测度

定义 3.1.1 设 E 为任意集合。映射 $\mu^* : \mathcal{P}(E) \rightarrow [0, \infty]$ 称为外测度，如果：

- (i) $\mu^*(\emptyset) = 0$;
- (ii) μ^* 单调：若 $A \subset B$ ，则 $\mu^*(A) \leq \mu^*(B)$;
- (iii) μ^* 可列次可加：对任意子集列 $(A_k)_{k \in \mathbb{N}} \subset \mathcal{P}(E)$,

$$\mu^*\left(\bigcup_{k \in \mathbb{N}} A_k\right) \leq \sum_{k \in \mathbb{N}} \mu^*(A_k).$$

外测度的条件比测度更弱，但外测度定义在 E 的所有子集上，而非只定义在某个 σ -代数上。本节固定一个外测度 μ^* ，并构造依赖于它的 σ -代数 $\mathcal{M}(\mu^*)$ ，使 μ^* 限制在该集合族上成为测度。

定义 3.1.2 若 $B \subset E$ 满足：对每个 $A \subset E$,

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

则称 B 为 μ^* -可测集。所有 μ^* -可测集组成的集合族记为 $\mathcal{M}(\mu^*)$ 。

注。由可列次可加性，总有

$$\mu^*(A) \leq \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

因此，要验证 B 的 μ^* -可测性，只需证明反向不等式。

定理 3.1.1 (1) $\mathcal{M}(\mu^*)$ 是 σ -代数，并且包含每个满足下式的集合 $B \subset E$ ：

$$\mu^*(B) = 0.$$

(2) μ^* 在 $\mathcal{M}(\mu^*)$ 上的限制是测度。

证明。(1) 简记 $\mathcal{M} = \mathcal{M}(\mu^*)$ 。若 $\mu^*(B) = 0$ ，则由单调性，

$$\mu^*(A) \geq \mu^*(A \cap B^c) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$$

故 $B \in \mathcal{M}$ 。特别地, $\emptyset \in \mathcal{M}$ 。由定义的对称性, $\mathcal{M}$ 对取补封闭。下面先证明它对有限并封闭。取 $B_1, B_2 \in \mathcal{M}$ 和任意 $A \subset E$ 。把 B_1 的可测性应用于 $A \cap (B_1 \cup B_2)$, 得到

$$\mu^*(A \cap (B_1 \cup B_2)) = \mu^*(A \cap (B_1 \cup B_2) \cap B_1) + \mu^*(A \cap (B_1 \cup B_2) \cap B_1^c) = \mu^*(A \cap B_1) + \mu^*(A \cap B_2 \cap B_1^c).$$

再依次使用 B_2 与 B_1 的可测性,

$$\begin{aligned} & \mu^*(A \cap (B_1 \cup B_2)) + \mu^*(A \cap (B_1 \cup B_2)^c) \\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2) + \mu^*(A \cap B_1^c \cap B_2^c) \\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) = \mu^*(A). \end{aligned}$$

因此 $B_1 \cup B_2 \in \mathcal{M}$ 。结合取补封闭性, $\mathcal{M}$ 也对有限交封闭; 特别地, 若 $B, B' \in \mathcal{M}$, 则

$$B' \setminus B = B' \cap B^c \in \mathcal{M}.$$

于是只需证明: 若 $B_k \in \mathcal{M}$ 两两不交, 则 $\bigcup_k B_k \in \mathcal{M}$ 。为此, 对 m 归纳证明: 对任意 $A \subset E$,

$$\mu^*(A) = \sum_{k=0}^m \mu^*(A \cap B_k) + \mu^*(A \cap \left(\bigcap_{k=0}^m B_k^c \right)).$$

$m=0$ 时这就是 B_0 的可测性。若结论对 m 成立, 则利用 B_{m+1} 的可测性写成

$$\begin{aligned} \mu^*(A \cap \left(\bigcap_{k=0}^m B_k^c \right)) &= \mu^*(A \cap \left(\bigcap_{k=0}^m B_k^c \right) \cap B_{m+1}) + \mu^*(A \cap \left(\bigcap_{k=0}^{m+1} B_k^c \right)) \\ &= \mu^*(A \cap B_{m+1}) + \mu^*(A \cap \left(\bigcap_{k=0}^{m+1} B_k^c \right)) \end{aligned}$$

这里使用了 $B_0, B_1, \dots$ 两两不交。由上面的归纳式可得

$$\mu^*(A) \geq \sum_{k=0}^m \mu^*(A \cap B_k) + \mu^*(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right))$$

再令 $m \rightarrow \infty$, 得到

$$\begin{aligned} \mu^*(A) &\geq \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^*(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right)) \\ &\geq \mu^*(A \cap \left(\bigcup_{k=0}^{\infty} B_k \right)) + \mu^*(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right)), \end{aligned}$$

第二个不等式来自外测度的可列次可加性。这正是 $\bigcup_{k=0}^{\infty} B_k$ 的 μ^* -可测性。

(2) 记 $\mu = \mu^*|_{\mathcal{M}}$ 。已有 $\mu(\emptyset) = 0$ 。设 $B_k \in \mathcal{M}$ 两两不交。(1) 的证明说明, 对任意 $A \subset E$,

$$\mu^*(A) \geq \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^*(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right))$$

取 $A = \bigcup_{k=0}^{\infty} B_k$, 得到

$$\mu^*\left(\bigcup_{k=0}^{\infty} B_k\right) \geq \sum_{k=0}^{\infty} \mu^*(B_k).$$

反向不等式由 μ^* 的可列次可加性给出。因此 μ 可列可加, 确为测度。

3.2 Lebesgue 测度

对任意 $A \subset \mathbb{R}$, 定义

下确界取遍 A 的所有可数开区间覆盖 $((a_i, b_i))_{i \in \mathbb{N}}$ 。这样的覆盖总是存在。

定理 3.2.1 (i) λ^* 是 $\mathbb{R}$ 上的外测度; (ii) $\mathcal{M}(\lambda^*) \supset \mathcal{B}(\mathbb{R})$; (iii) 对任意 $a \leq b$,

$$\lambda^*([a, b]) = \lambda^*((a, b)) = b - a.$$

λ^* 在 $\mathcal{B}(\mathbb{R})$ (或在 $\mathcal{M}(\lambda^*)$) 上的限制称为 $\mathbb{R}$ 上的 Lebesgue 测度, 以后简记为 λ 。由第 1 章末的唯一性结果, λ 是 $\mathcal{B}(\mathbb{R})$ 上唯一满足 $\lambda((a, b)) = b - a$ ($a < b$) 的测度。

证明。(i) 显然 $\lambda^*(\emptyset) = 0$, 且 λ^* 具有单调性。只需再证可列次可加性。为此, 任取 $\mathbb{R}$ 的子集列 $(A_n)_{n \in \mathbb{N}}$ ¹。可设对每个 n 都有 $\lambda^*(A_n) < \infty$, 否则无须证明。给定 $\varepsilon > 0$ 。对每个 $n \in \mathbb{N}$, 可取一列开区间 $]a_i^{(n)}, b_i^{(n)}[$ ($i \in \mathbb{N}$), 使得²

$$A_n \subset \bigcup_{i \in \mathbb{N}}]a_i^{(n)}, b_i^{(n)}[.$$

以及

$$\sum_{i \in \mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \lambda^*(A_n) + \frac{\varepsilon}{2^n}.$$

这些区间 $(a_i^{(n)}, b_i^{(n)})$ ($n, i \in \mathbb{N}$) 构成 $\bigcup_{n \in \mathbb{N}} A_n$ 的可数开覆盖, 因此

$$\lambda^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \sum_{i \in \mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \sum_{n \in \mathbb{N}} \lambda^*(A_n) + 2\varepsilon.$$

由于 $\varepsilon > 0$ 任意, 便得到可列次可加性。

(ii) 因为 $\mathcal{M}(\lambda^*)$ 是 σ -代数, 只需证明它包含一个生成 $\mathcal{B}(\mathbb{R})$ 的集合族。取 $\{(-\infty, \alpha] : \alpha \in \mathbb{R}\}$ 。固定 $\alpha \in \mathbb{R}$, 令 $B = (-\infty, \alpha]$ 。需要验证: 对 $\mathbb{R}$ 的任意子集 A ,

$$\lambda^*(A) \geq \lambda^*(A \cap B) + \lambda^*(A \cap B^c).$$

设 $((a_i, b_i))_{i \in \mathbb{N}}$ 是 A 的一个开覆盖, 并取 $\varepsilon > 0$ 。区间

$$(a_i \wedge \alpha, (b_i \wedge \alpha) + \varepsilon 2^{-i})$$

覆盖 $A \cap B$, 而区间

$$(a_i \vee \alpha, b_i \vee \alpha)$$

覆盖 $A \cap B^c$ 。因此

$$\begin{aligned} \lambda^*(A \cap B) &\leq \sum_{i \in \mathbb{N}} ((b_i \wedge \alpha) - (a_i \wedge \alpha)) + 2\varepsilon, \\ \lambda^*(A \cap B^c) &\leq \sum_{i \in \mathbb{N}} ((b_i \vee \alpha) - (a_i \vee \alpha)). \end{aligned}$$

¹校订注: 原文此处误作“ $\mathbb{N}$ 的子集列”; 按定理所在空间及后续论证改为“ $\mathbb{R}$ 的子集列”。

²校订注: 原文随后一个估计式把误差写为 $\varepsilon/2^i$, 其中 i 已是区间的求和指标, 因而留下自由指标。为了在随后对 n 求和, 正确写法应为 $\varepsilon/2^n$, 此处已改正。

对各项求和可得

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \sum_{i \in \mathbb{N}} (b_i - a_i) + 2\varepsilon.$$

由于 ε 是任意的, 我们有

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \sum_{i \in \mathbb{N}} (b_i - a_i).$$

而 $\lambda^*(A)$ 按定义是对 A 的所有可数开区间覆盖所得右端和式的下确界, 故所需不等式成立。

(iii) 由定义立即有

$$\lambda^*([a, b]) \leq b - a.$$

反过来, 设

$$[a, b] \subset \bigcup_{i \in \mathbb{N}} (a_i, b_i).$$

由紧性, 可从中选出有限子覆盖; 重新编号后, 存在 N , 使

$$[a, b] \subset \bigcup_{i=0}^N (a_i, b_i).$$

初等的区间长度估计给出

$$b - a \leq \sum_{i=0}^N (b_i - a_i) \leq \sum_{i=0}^{\infty} (b_i - a_i).$$

对所有可数开覆盖取下确界, 得到 $b - a \leq \lambda^*([a, b])$ 。最后, 因为 $\lambda^*({a}) = \lambda^*({b}) = 0$, 所以 $\lambda^*((a, b)) = \lambda^*([a, b])$ 。

d 维推广。 称形如

$$P = \prod_{j=1}^d (a_j, b_j) \quad \left(\text{相应地, } P = \prod_{j=1}^d [a_j, b_j] \right)$$

的 $\mathbb{R}^d$ 子集为开长方体 (相应地, 闭长方体)。定义其体积为

$$\text{vol}(P) = \prod_{j=1}^d (b_j - a_j).$$

对 $\mathbb{R}^d$ 的任意子集 A , 再定义

$$\lambda^*(A) = \inf \left\{ \sum_{i \in \mathbb{N}} \text{vol}(P_i) : A \subset \bigcup_{i \in \mathbb{N}} P_i \right\},$$

其中下确界取遍 A 的所有可数开长方体覆盖。于是有前一定理的 d 维版本。

定理 3.2.2 (i) λ^* 是 $\mathbb{R}^d$ 上的外测度; (ii) $\mathcal{M}(\lambda^*)$ 包含 $\mathcal{B}(\mathbb{R}^d)$; (iii) 对任意开长方体或闭长方体 P , $\lambda^*(P) = \text{vol}(P)$ 。

λ^* 在 $\mathcal{B}(\mathbb{R}^d)$ (或在 $\mathcal{M}(\lambda^*)$) 上的限制称为 $\mathbb{R}^d$ 上的 Lebesgue 测度, 仍简记为 λ 。

证明。 (i) 的证明与 $d = 1$ 时完全相同。为证 (ii), 只需证明形如

$$A = \mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, a] \times \mathbb{R} \times \cdots \times \mathbb{R}$$

的集合属于 $\mathcal{M}(\lambda^*)$, 因为这些集合生成 $\mathcal{B}(\mathbb{R}^d)$ 。证明与一维情形完全类似。为证 (iii), 可归结为如下事实: 若闭长方体 P 被有限个开长方体覆盖,

$$P \subset \bigcup_{i=1}^n P_i,$$

则

$$\text{vol}(P) \leq \sum_{i=1}^n \text{vol}(P_i).$$

这个事实留作练习。

注。第 5 章还会给出另一种从一维情形构造 d 维 Lebesgue 测度的方法。

自然要问, $\mathcal{M}(\lambda^*)$ 是否比 $\mathcal{B}(\mathbb{R}^d)$ 大得多。下面会看到, 从完备化的意义说, 二者相差并不大。先给出一个一般性的预备命题。

命题 3.2.3 设 $(E, \mathcal{A}, \mu)$ 为测度空间。定义可忽略集族

$$\mathcal{N} = \{A \subset E : \exists B \in \mathcal{A}, A \subset B, \mu(B) = 0\}.$$

称 $\overline{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N})$ 为 $\mathcal{A}$ 关于 μ 的完备化。存在唯一的测度 $\overline{\mu}$ 定义在 $(E, \overline{\mathcal{A}})$ 上, 并且延拓 μ 。

证明。先给出 $\overline{\mathcal{A}}$ 的具体描述。令

$$\mathcal{D} = \{A \subset E : \exists B, B' \in \mathcal{A}, B \subset A \subset B', \mu(B' \setminus B) = 0\}.$$

容易验证 $\mathcal{D}$ 是一个 σ -代数。显然 $\mathcal{A} \subset \mathcal{D}$ 且 $\mathcal{N} \subset \mathcal{D}$, 所以 $\overline{\mathcal{A}} \subset \mathcal{D}$ 。反过来, 若 $A \in \mathcal{D}$, 取上式中的 B, B' , 则 $A = B \cup (A \setminus B)$, 其中 $B \in \mathcal{A}$, $A \setminus B \in \mathcal{N}$, 故 $A \in \overline{\mathcal{A}}$ 。因此 $\overline{\mathcal{A}} = \mathcal{D}$ 。现在对 $A \in \overline{\mathcal{A}} = \mathcal{D}$, 选取满足上述条件的 B, B' , 定义 $\overline{\mu}(A) = \mu(B) = \mu(B')$ 。该定义与 B, B' 的选择无关: 若 $\tilde{B}, \tilde{B}'$ 是另一组选择, 则由夹逼关系可得 $\mu(\tilde{B}) \leq \mu(B') = \mu(B)$ 以及 $\mu(B) \leq \mu(\tilde{B}') = \mu(\tilde{B})$, 故两者相等。最后验证可列可加性。若 (A_n) 是 $\overline{\mathcal{A}}$ 中两两不交的集合, 选取 $B_n \in \mathcal{A}$, $B_n \subset A_n$, 使 $A_n \setminus B_n$ 可忽略, 则

$$\sum_n \overline{\mu}(A_n) = \sum_n \mu(B_n) = \mu\left(\bigcup_n B_n\right) = \overline{\mu}\left(\bigcup_n A_n\right),$$

其中最后一个等号来自 $\bigcup_n (A_n \setminus B_n)$ 可忽略。故 $\overline{\mu}$ 是测度。唯一性由同一夹逼表示立即得到。

命题 3.2.4 Carathéodory 可测集的 σ -代数 $\mathcal{M}(\lambda^*)$ 等于 Borel σ -代数 $\mathcal{B}(\mathbb{R}^d)$ 关于 Lebesgue 测度 λ 的完备化 $\overline{\mathcal{B}}(\mathbb{R}^d)$ 。

证明。包含关系 $\overline{\mathcal{B}}(\mathbb{R}^d) \subset \mathcal{M}(\lambda^*)$ 是直接的。事实上, 若 $A \subset B$, 其中 $B \in \mathcal{B}(\mathbb{R}^d)$ 且 $\lambda(B) = 0$, 则 $\lambda^*(A) \leq \lambda^*(B) = 0$; 由本节开头的定理, $A \in \mathcal{M}(\lambda^*)$ 。反过来, 设 $A \in \mathcal{M}(\lambda^*)$ 。先设 $A \subset (-K, K)^d$; 一般情形可把 A 写成 $A \cap (-n, n)^d$ 的递增并。此时 $\lambda^*(A) < \infty$ 。对每个 $n \geq 1$, 可取一列包含于 $(-K, K)^d$ 的开长方体 $(P_i^n)_{i \in \mathbb{N}}$, 使

$$A \subset \bigcup_i P_i^n, \quad \sum_i \text{vol}(P_i^n) \leq \lambda^*(A) + \frac{1}{n}.$$

令 $B_n = \bigcup_i P_i^n$, $B = \bigcap_n B_n$ 。则 $B \in \mathcal{B}(\mathbb{R}^d)$, $A \subset B$, 并且对每个 n ,

$$\lambda^*(B) \leq \sum_i \text{vol}(P_i^n) \leq \lambda^*(A) + \frac{1}{n},$$

从而 $\lambda^*(B) = \lambda^*(A)$ 。对 $(-K, K)^d \setminus A$ 作同样构造, 得到 Borel 集 $\tilde{B} \subset (-K, K)^d$, 满足 $(-K, K)^d \setminus A \subset \tilde{B}$ 且两者外测度相等。令 $B' = (-K, K)^d \setminus \tilde{B}$, 则 $B' \subset A$ 且 $\lambda^*(B') = \lambda^*(A)$ 。故存在 Borel 集 B', B 使 $B' \subset A \subset B$ 且 $\lambda(B \setminus B') = 0$, 即 $A \in \overline{\mathcal{B}}(\mathbb{R}^d)$ 。

定理 3.2.5 $\mathbb{R}^d$ 上的 Lebesgue 测度具有平移不变性: 对任意 $A \in \mathcal{B}(\mathbb{R}^d)$ 和 $x \in \mathbb{R}^d$,

$$\lambda(x + A) = \lambda(A).$$

反过来, 若 μ 是 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上在有界集上有限且具有平移不变性的测度, 则存在常数 $c \geq 0$, 使 $\mu = c\lambda$ 。

证明。 记 $\sigma_x(y) = y - x$ 。像测度 $\sigma_x(\lambda)$ 定义为

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \sigma_x(\lambda)(A) = \lambda(\sigma_x^{-1}(A)) = \lambda(x + A).$$

对任意长方体 A , 都有 $\sigma_x(\lambda)(A) = \lambda(A)$, 因为 A 与 $x + A$ 体积相同。由第 1 章的单调类引理, $\sigma_x(\lambda) = \lambda$, 这就证明了第一部分。

反过来, 设 μ 是 $\mathcal{B}(\mathbb{R}^d)$ 上的平移不变测度, 并令

$$c = \mu([0, 1]^d).$$

集合 $[0, 1]^d$ 是 n^d 个彼此不交、且均为 $[0, 1/n]^d$ 的平移的长方体之并, 故对每个整数 $n \geq 1$,

$$\mu([0, 1/n]^d) = \frac{c}{n^d}.$$

再取 $a_1, \dots, a_d \geq 0$ 。记 $[x]$ 为 x 的整数部分。由

$$\prod_{j=1}^d \left[0, \frac{[na_j]}{n}\right) \subset \prod_{j=1}^d [0, a_j) \subset \prod_{j=1}^d \left[0, \frac{[na_j] + 1}{n}\right),$$

得到

$$\left(\prod_{j=1}^d [na_j]\right) \frac{c}{n^d} \leq \mu\left(\prod_{j=1}^d [0, a_j)\right) \leq \left(\prod_{j=1}^d ([na_j] + 1)\right) \frac{c}{n^d}.$$

令 $n \rightarrow \infty$, 便有

$$\mu\left(\prod_{j=1}^d [0, a_j)\right) = c \prod_{j=1}^d a_j = c\lambda\left(\prod_{j=1}^d [0, a_j)\right).$$

³利用平移不变性, μ 与 $c\lambda$ 在所有半开长方体 $\prod_{j=1}^d [a_j, b_j)$ 上相等。再用单调类引理, 得到 $\mu = c\lambda$ 。

命题 3.2.6 Lebesgue 测度在 $\mathbb{R}^d$ 上是正则的。更准确地, 对任意 $A \in \overline{\mathcal{B}}(\mathbb{R}^d)$,

$$\lambda(A) = \inf\{\lambda(U) : U \text{ 为开集, } A \subset U\} = \sup\{\lambda(F) : F \text{ 为紧集, } F \subset A\}.$$

证明。 显然, $\inf\{\lambda(U) : U \text{ 为开集且 } A \subset U\} \geq \lambda(A)$ 。为证反向不等式, 可设 $\lambda(A) < \infty$ 。由 $\lambda(A) = \lambda^*(A)$ 的定义, 对每个 $\varepsilon > 0$, 可用开方体 $(P_i)_i$ 覆盖 A , 使 $\sum_i \lambda(P_i) \leq \lambda(A) + \varepsilon$ 。令 $U = \bigcup_i P_i$, 则 U 是包含 A 的开集, 并且 $\lambda(U) \leq \sum_i \lambda(P_i) \leq \lambda(A) + \varepsilon$ 。令 $\varepsilon \downarrow 0$ 即得所需不等式。

³校订注: 原讲义这一等式中间的乘积上限误排为 n ; 乘积遍历的是 d 个坐标, 故应为 d 。

为证明命题中的第二个等式, 可先假设 A 包含在某个紧集 C 中; 否则使用 $\lambda(A) = \lim_{n \rightarrow \infty} \uparrow \lambda(A \cap [-n, n]^d)$ 。给定 $\varepsilon > 0$, 由第一部分可取包含 $C \setminus A$ 的开集 U , 使 $\lambda(U) < \lambda(C \setminus A) + \varepsilon$ 。于是 $F = C \setminus U$ 是包含在 A 中的紧集, 并且

$$\lambda(F) \geq \lambda(C) - \lambda(U) \geq \lambda(C) - (\lambda(C \setminus A) + \varepsilon) = \lambda(A) - \varepsilon,$$

这就得到了第二个等式。前一个命题可以推广到一个更加一般的框架。我们仅限于有限测度的情况。

命题 3.2.7 设 (E, d) 为度量空间, μ 为 $(E, \mathcal{B}(E))$ 上的有限测度。则对任意 $A \in \mathcal{B}(E)$,

$$\mu(A) = \inf\{\mu(U) : U \text{ 为开集}, A \subset U\} = \sup\{\mu(F) : F \text{ 为闭集}, F \subset A\}.$$

证明。 记 $\mathcal{O}$ 为 E 中所有开集构成的类, 记 $\mathcal{C}$ 为满足命题中两个等式的 Borel 集构成的类。因为 $\mathcal{B}(E) = \sigma(\mathcal{O})$, 只需证明 $\mathcal{O} \subset \mathcal{C}$ 且 $\mathcal{C}$ 是 σ -代数。

若 $A \in \mathcal{O}$, 外逼近等式显然。对每个 $n \geq 1$, 令

$$F_n = \{x \in E : d(x, A^c) \geq 1/n\}.$$

则 F_n 闭, 且 $F_n \uparrow A$, 所以

$$\mu(A) = \lim_{n \rightarrow \infty} \uparrow \mu(F_n).$$

这证明了内逼近等式, 故 $\mathcal{O} \subset \mathcal{C}$ 。

还需证明 $\mathcal{C}$ 是 σ -代数。显然 $\emptyset \in \mathcal{C}$; 利用开集与闭集在取补下的对称性, $\mathcal{C}$ 对取补封闭。设 $(A_n)_{n \in \mathbb{N}} \subset \mathcal{C}$, 并取 $\varepsilon > 0$ 。对每个 n , 可取开集 $U_n \supset A_n$, 使 $\mu(U_n) \leq \mu(A_n) + \varepsilon 2^{-n}$ 。于是

$$\mu\left(\bigcup_{n \in \mathbb{N}} U_n \setminus \bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(U_n \setminus A_n) \leq 2\varepsilon.$$

因为 $\bigcup_n U_n$ 是开集, 这证明了 $\bigcup_n A_n$ 的外逼近等式。再取充分大的 N , 使

$$\mu\left(\bigcup_{n=0}^N A_n\right) \geq \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) - \varepsilon.$$

对每个 $n \leq N$, 可取闭集 $F_n \subset A_n$, 使 $\mu(A_n \setminus F_n) \leq \varepsilon 2^{-n}$ 。令 $F = \bigcup_{n=0}^N F_n$, 则 F 闭且

$$\mu\left(\bigcup_{n=0}^N A_n \setminus F\right) < 2\varepsilon,$$

从而

$$\mu\left(\bigcup_{n=0}^{\infty} A_n \setminus F\right) < 3\varepsilon.$$

这证明了内逼近等式, 故 $\bigcup_n A_n \in \mathcal{C}$ 。

3.3 与黎曼积分的联系

固定实数轴上的非退化闭区间 $[a, b]$ 。若函数 $h : [a, b] \rightarrow \mathbb{R}$ 满足: 存在分点 $a = x_0 < x_1 < \dots < x_N = b$ 以及实数 $y_1, \dots, y_N$, 使

$$\forall i \in \{1, \dots, N\}, \quad \forall x \in (x_{i-1}, x_i), \quad h(x) = y_i,$$

则称 h 为阶梯函数, 记作 $h \in \text{Esc}$ 。⁴定义

$$I(h) = \sum_{i=1}^N y_i (x_i - x_{i-1}).$$

显然 $I(h) = \int_{[a,b]} h d\lambda$ 。称有界函数 $f : [a, b] \rightarrow \mathbb{R}$ 为黎曼可积, 如果

$$\sup_{\substack{h \in \text{Esc} \\ h \leq f}} I(h) = \inf_{\substack{h \in \text{Esc} \\ h \geq f}} I(h).$$

这个公共值记为 $I(f)$ 。

命题 3.3.1 设 f 在 $[a, b]$ 上黎曼可积。则 f 对完备化 $\overline{\mathcal{B}}([a, b])$ 可测, 并且

$$I(f) = \int_{[a,b]} f d\lambda.$$

证明。可取定义在 $[a, b]$ 上的阶梯函数列 (h_n) , 使得 $h_n \geq f$ 且 $I(h_n) \downarrow I(f)$ 。把 h_n 替换为 $h_1 \wedge \cdots \wedge h_n$ 后, 不妨设 (h_n) 单调递减, 因而可以定义

$$h_\infty = \lim \downarrow h_n \geq f.$$

同样, 可取递增的阶梯函数列 $(\tilde{h}_n)$, 使 $\tilde{h}_n \leq f$ 且 $I(\tilde{h}_n) \uparrow I(f)$, 并令

$$\tilde{h}_\infty = \lim \uparrow \tilde{h}_n \leq f.$$

函数 h_∞ 与 $\tilde{h}_\infty$ 都是有界 Borel 函数。由支配收敛定理,

$$\int_{[a,b]} h_\infty d\lambda = \lim_{n \rightarrow \infty} \downarrow \int_{[a,b]} h_n d\lambda = I(f),$$

$$\int_{[a,b]} \tilde{h}_\infty d\lambda = \lim_{n \rightarrow \infty} \uparrow \int_{[a,b]} \tilde{h}_n d\lambda = I(f).$$

所以

$$\int_{[a,b]} (h_\infty - \tilde{h}_\infty) d\lambda = 0.$$

由于 $h_\infty \geq \tilde{h}_\infty$, 可知二者几乎处处相等。又因 $h_\infty \geq f \geq \tilde{h}_\infty$, 所以 f 几乎处处等于一个 Borel 函数, 进而对完备化 $\overline{\mathcal{B}}([a, b])$ 可测。最后,

$$\int_{[a,b]} f d\lambda = \int_{[a,b]} h_\infty d\lambda = I(f).$$

3.4 不可测集合的一个例子

考虑商空间 $\mathbb{R}/\mathbb{Q}$, 即实数模有理数的等价类。对每个 $a \in \mathbb{R}/\mathbb{Q}$, 从该等价类中选取一个位于 $[0, 1]$ 的代表元 x_a , 并令

$$F = \{x_a : a \in \mathbb{R}/\mathbb{Q}\} \subset [0, 1].$$

⁴校订注: 原讲义在这个定义的公式中误写为 $f(x) = y_i$; 本段正在定义的是 h , 故改为 $h(x) = y_i$ 。

则 F 不是 Borel 集, 甚至不属于 Lebesgue σ -代数 $\overline{\mathcal{B}}(\mathbb{R})$ 。反设 F 可测。由构造,

$$\mathbb{R} \subset \bigcup_{q \in \mathbb{Q}} (q + F),$$

故 $\lambda(F) > 0$; 否则右端是可数个零测集之并。另一方面, 集合 $q + F$ ($q \in \mathbb{Q}$) 两两不交: 若 $q + x_a = q' + x_{a'}$, 则 $x_a - x_{a'} = q' - q \in \mathbb{Q}$, 因而 $a = a'$, 继而 $q = q'$ 。又有

$$\bigcup_{q \in \mathbb{Q} \cap [0, 1]} (q + F) \subset [0, 2],$$

所以

$$\sum_{q \in \mathbb{Q} \cap [0, 1]} \lambda(q + F) \leq 2.$$

由平移不变性, 和式中的每一项都等于 $\lambda(F)$; 指标集无限, 故只能有 $\lambda(F) = 0$, 矛盾。

3.5 斯蒂尔切斯积分

下面的定理刻画 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的所有有限测度。这一结果也容易推广到 Radon 测度。

定理 3.5.1

(i) 设 μ 是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的有限测度。对每个 $x \in \mathbb{R}$, 令

$$F_\mu(x) = \mu((-\infty, x]).$$

则 F_μ 单调递增、有界、右连续, 并满足 $F_\mu(-\infty) := \lim_{x \rightarrow -\infty} F_\mu(x) = 0$ 。

(ii) 反过来, 若 $F: \mathbb{R} \rightarrow \mathbb{R}_+$ 单调递增、有界、右连续, 且 $F(-\infty) = 0$, 则存在唯一的 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的有限测度 μ , 使 $F = F_\mu$ 。

注。 当 $F = F_\mu$ 时, 通常记

$$\int f(x) \mu(dx) = \int f(x) dF(x).$$

右端称为 f 关于 F 的 Stieltjes 积分。特别地,

$$\int_{(a, b]} dF(x) = F(b) - F(a),$$

以及

$$\int_{[a, b]} dF(x) = \lim_{n \rightarrow \infty} \int_{(a-n^{-1}, b]} dF(x) = F(b) - F(a-),$$

其中 $F(a-)$ 表示 F 在 a 处的左极限。

证明。 (i) 这些性质都很容易验证。例如, 若 $x_n \downarrow x$, 则 $(-\infty, x_n] \downarrow (-\infty, x]$, 因而

$$F_\mu(x_n) = \mu((-\infty, x_n]) \downarrow \mu((-\infty, x]) = F_\mu(x).$$

同理, 若 $x_n \downarrow -\infty$, 则 $(-\infty, x_n] \downarrow \emptyset$, 所以 $F_\mu(x_n) \downarrow 0$ 。

(ii) 唯一性由第 1 章的单调类引理推出, 因为集合族 $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ 对有限交封闭并生成 $\mathcal{B}(\mathbb{R})$ 。为证明存在性, 对每个 $A \subset \mathbb{R}$ 定义

$$\mu^*(A) = \inf \left\{ \sum_{i \in \mathbb{N}} (F(b_i) - F(a_i)) : A \subset \bigcup_{i \in \mathbb{N}} (a_i, b_i] \right\}.$$

这里用左开右闭区间覆盖 A 。与构造 Lebesgue 测度时相同的论证表明 μ^* 是外测度, 并且每个 $(-\infty, \alpha]$ 都属于 $\mathcal{M}(\mu^*)$ 。⁵ 因此 $\mathcal{M}(\mu^*)$ 包含 $\mathcal{B}(\mathbb{R})$, 而 μ^* 在 Borel σ -代数上的限制给出一个测度 μ 。最后只需证明: 对任意 $a < b$,

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

上界由 μ^* 的定义立即得到。

为证下界, 设 $((x_i, y_i])_{i \in \mathbb{N}}$ 是 $(a, b]$ 的可数覆盖, 并取 $\varepsilon \in (0, b - a)$ 。对每个 i , 由 F 的右连续性可取 $y'_i > y_i$, 使 $F(y'_i) \leq F(y_i) + \varepsilon 2^{-i}$ 。开区间族 $((x_i, y'_i))_{i \in \mathbb{N}}$ 覆盖紧区间 $[a + \varepsilon, b]$, 故存在有限子覆盖。重新编号后,

$$\begin{aligned} F(b) - F(a + \varepsilon) &\leq \sum_{i=0}^{N_\varepsilon} (F(y'_i) - F(x_i)) \\ &\leq \sum_{i=0}^{\infty} (F(y'_i) - F(x_i)) \\ &\leq \sum_{i=0}^{\infty} (F(y_i) - F(x_i)) + 2\varepsilon. \end{aligned}$$

令 $\varepsilon \downarrow 0$, 并利用 F 的右连续性, 得到

$$F(b) - F(a) \leq \sum_{i=0}^{\infty} (F(y_i) - F(x_i)).$$

对所有覆盖取下确界, 便有 $\mu((a, b]) \geq F(b) - F(a)$ 。

Radon 测度的情形。 公式

$$F(x) = \begin{cases} \mu((0, x]), & x \geq 0, \\ -\mu((x, 0]), & x < 0 \end{cases}$$

给出如下双射: $\mathbb{R}$ 上的 Radon 测度 μ , 对应于满足 $F(0) = 0$ 的右连续单调递增函数 $F: \mathbb{R} \rightarrow \mathbb{R}$ 。这可由有限测度情形推出, 且仍有 $\mu((a, b]) = F(b) - F(a)$ 。特别地, 当 $F(x) = x$ 时, 对应测度就是 Lebesgue 测度。

3.6 里斯表示定理

设 X 为度量空间, $C_c(X)$ 表示 X 上具有紧支集的连续函数空间。定义在 $C_c(X)$ 上的线性泛函 J 称为正线性泛函, 如果 $f \geq 0$ 时总有 $J(f) \geq 0$ 。若 μ 是 X 上的 Radon 测度, 则可在 $C_c(X)$ 上定义线性泛函

$$J(f) = \int f d\mu.$$

这个积分有定义, 因为存在紧集 $K \subset X$ 与常数 C , 使 $|f| \leq C 1_K$, 而 Radon 测度在紧集上有限。此外, J 是正的。Riesz 表示定理说明, 在适当条件下, $C_c(X)$ 上每个正线性泛函都来自这种积分。

⁵校订注: 原讲义此处误写成 $\mathcal{M}(\lambda^*)$; 本段构造的外测度是 μ^* , 故应为 $\mathcal{M}(\mu^*)$ 。

定理 3.6.1 (Riesz 表示定理) 设 X 是局部紧致可分度量空间, J 是 $C_c(X)$ 上的正线性泛函。则存在唯一的 $(X, \mathcal{B}(X))$ 上的 Radon 测度 μ , 使

$$\forall f \in C_c(X), J(f) = \int f d\mu.$$

测度 μ 是正则的, 即对任意 $A \in \mathcal{B}(X)$,

$$\mu(A) = \inf\{\mu(U) : U \text{ 为开集且 } A \subset U\} = \sup\{\mu(K) : K \text{ 为紧集且 } K \subset A\}.$$

^a

此外, 对任意开集 $U \subset X$,

$$\mu(U) = \sup\{J(f) : f \in C_c(X), 0 \leq f \leq 1_U\}.$$

^a校订注: 原讲义此处三次把测度符号误排为 λ ; 本定理所构造的是一般 Radon 测度 μ , 故统一改为 μ 。

例。若 $X = \mathbb{R}$, 可取 $J(f) = I(f)$, 其中 $I(f)$ 是前面定义的黎曼积分。容易验证 J 是 $C_c(\mathbb{R})$ 上的正线性泛函, 对应测度正是 Lebesgue 测度。因此, 在已经构造连续函数黎曼积分的前提下, 这又给出 Lebesgue 测度的一种构造。

本讲义不证明定理 3.6.1; 可参见 Briane 与 Pagès [2] 的定理 10.1, 或 Rudin [7] 第 2 章中更精确的表述。

第四章 L^p 空间

本章主要研究 L^p 空间，即绝对值的 p 次方可积的函数所成的空间。Hölder、Minkowski 与 Jensen 不等式是基本工具。我们将研究 L^p 的 Banach 空间结构，以及 $p = 2$ 时 L^2 的 Hilbert 空间结构。稠密性定理说明， L^p 中任意函数都可由更光滑的函数逼近，这在分析中有许多应用。最后利用 L^2 的 Hilbert 空间结构证明 Radon–Nikodym 定理：给定参考测度，可把另一测度分解为关于参考测度有密度的部分与奇异部分之和。

4.1 定义与 Hölder 不等式

本章始终固定一个测度空间 $(E, \mathcal{A}, \mu)$ 。对每个实数 $p \geq 1$ ，定义

$$\mathcal{L}^p(E, \mathcal{A}, \mu) = \left\{ f : E \rightarrow \mathbb{R} : f \text{ 可测且 } \int |f|^p d\mu < \infty \right\}.$$

并且我们还定义

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) = \{ f : E \rightarrow \mathbb{R} : f \text{ 可测, 且存在 } C < \infty \text{ 使 } |f| \leq C \text{ 几乎处处} \}.$$

也可以类似定义复值函数空间 $\mathcal{L}_\mathbb{C}^p$ 与 $\mathcal{L}_\mathbb{C}^\infty$ ，但本章主要讨论实值情形。对每个 $p \in [1, \infty]$ ，在 $\mathcal{L}^p$ 上定义等价关系

$$f \sim g \iff f = g \text{ } \mu\text{-几乎处处}.$$

由此定义商空间

$$L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu) / \sim.$$

因此， $L^p(E, \mathcal{A}, \mu)$ 的元素是由几乎处处相等的函数组成的等价类。以后通常把 L^p 的元素与其某个代表元视为同一对象；凡涉及点值的陈述，都应理解为选取代表元后、在几乎处处意义下成立。

对任意可测函数 $f : E \rightarrow \mathbb{R}$ ，当 $p \in [1, \infty)$ 时记

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{1/p},$$

其中约定 $\infty^{1/p} = \infty$ ；并定义本性上确界范数

$$\|f\|_\infty = \inf\{C \in [0, \infty] : |f| \leq C \text{ } \mu\text{-几乎处处}\}.$$

于是 $|f| \leq \|f\|_\infty$ 几乎处处，且 $\|f\|_\infty$ 是具有这一性质的最小扩充实数。若 $p, q \in [1, \infty]$ 满足

$$\frac{1}{p} + \frac{1}{q} = 1,$$

则称 p, q 为共轭指数。特别地，1 与 ∞ 共轭。

定理 4.1.1 (Hölder 不等式) 设 p, q 为共轭指数。对任意可测函数 $f, g: E \rightarrow \mathbb{R}$,

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

特别地, 若 $f \in L^p(E, \mathcal{A}, \mu)$ 、 $g \in L^q(E, \mathcal{A}, \mu)$, 则 $fg \in L^1(E, \mathcal{A}, \mu)$ 。

证明。 若 $\|f\|_p = 0$, 则 $f = 0$ 几乎处处, 从而 $\int |fg| d\mu = 0$, 不等式显然成立。因此可设 $\|f\|_p > 0, \|g\|_q > 0$ 。不妨还设 $f \in L^p(E, \mathcal{A}, \mu), g \in L^q(E, \mathcal{A}, \mu)$ 。当 $p = 1, q = \infty$ 时, $|fg| \leq \|g\|_\infty |f|$ 几乎处处, 故

$$\int |fg| d\mu \leq \|g\|_\infty \int |f| d\mu = \|g\|_\infty \|f\|_1.$$

现设 $1 < p < \infty$ (因而 $1 < q < \infty$)。任取 $\alpha \in (0, 1)$ 。对每个 $x \in \mathbb{R}_+$, 有

$$x^\alpha - \alpha x \leq 1 - \alpha.$$

事实上, 函数 $\varphi_\alpha(x) = x^\alpha - \alpha x$ 在 $(0, \infty)$ 上的导数为 $\varphi'_\alpha(x) = \alpha(x^{\alpha-1} - 1)$ 。该导数在 $(0, 1)$ 上为正, 在 $(1, \infty)$ 上为负, 因此 φ_α 在 $x = 1$ 处取得最大值。把这一不等式应用于 $x = u/v$, 其中 $u \geq 0, v > 0$, 得到

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1 - \alpha)v,$$

这个不等式在 $v = 0$ 时仍成立。现在取 $\alpha = 1/p$ (于是 $1 - \alpha = 1/q$), 再令

$$u = \frac{|f(x)|^p}{\|f\|_p^p}, \quad v = \frac{|g(x)|^q}{\|g\|_q^q}$$

由此得到

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}.$$

对上述不等式关于 μ 积分, 得到

$$\frac{1}{\|f\|_p \|g\|_q} \int |fg| d\mu \leq \frac{1}{p} + \frac{1}{q} = 1.$$

练习。 设 $1 < p < \infty$ 。证明 Hölder 不等式取等号, 当且仅当存在不全为零的非负实数 α, β , 使 $\alpha|f|^p = \beta|g|^q$ 几乎处处成立。

Hölder 不等式在 $p = q = 2$ 时就是 Cauchy-Schwarz 不等式:

$$\int |fg| d\mu \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int |g|^2 d\mu \right)^{1/2}.$$

再考虑 $\mu(E) < \infty$ 的情形。在 Hölder 不等式中取 $g = 1$, 得到

$$\int |f| d\mu \leq \mu(E)^{1/q} \|f\|_p$$

这表明, 对每个 $p \in (1, \infty]$, 都有 $L^p \subset L^1$ 。将 $|f|$ 替换为 $|f|^r$ ($r \geq 1$), 并令 $r' = pr$, 便得到: 对任意 $1 \leq r \leq r' \leq \infty$,

$$\|f\|_r \leq \mu(E)^{\frac{1}{r} - \frac{1}{r'}} \|f\|_{r'},$$

所以 $L^{r'} \subset L^r$ (这里仍假设 μ 有限)。特别地, 当 μ 是概率测度时, 对所有 $1 \leq r \leq r' \leq \infty$, 都有 $\|f\|_r \leq \|f\|_{r'}$ 。最后这个不等式也可视为 Jensen 不等式的一个特例。

定理 4.1.2 (Jensen 不等式) 设 μ 是概率测度, 且 $\varphi: \mathbb{R} \rightarrow \mathbb{R}_+$ 为凸函数。则对任意 $f \in L^1(E, \mathcal{A}, \mu)$,

$$\int \varphi \circ f d\mu \geq \varphi\left(\int f d\mu\right).$$

注。 因为 $\varphi \circ f$ 是非负可测函数, 所以左端积分总有定义, 允许取值 $+\infty$ 。

证明。 令

$$\mathcal{E}_\varphi = \{(a, b) \in \mathbb{R}^2 : \forall x \in \mathbb{R}, \varphi(x) \geq ax + b\}.$$

凸函数的众所周知的性质保证了

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \sup_{(a,b) \in \mathcal{E}_\varphi} (ax + b).$$

因此,

$$\begin{aligned} \int \varphi \circ f d\mu &\geq \sup_{(a,b) \in \mathcal{E}_\varphi} \int (af + b) d\mu \\ &= \sup_{(a,b) \in \mathcal{E}_\varphi} \left(a \int f d\mu + b \right) \\ &= \varphi\left(\int f d\mu\right) \end{aligned}$$

练习。 证明: 若 $\mu(E) < \infty$, 则

$$\|f\|_\infty = \lim_{p \rightarrow \infty} \|f\|_p.$$

4.2 Banach 空间 L^p

定理 4.2.1 (Minkowski 不等式) 设 $p \in [1, \infty]$, $f, g \in L^p(E, \mathcal{A}, \mu)$ 。则 $f + g \in L^p(E, \mathcal{A}, \mu)$, 并且

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

证明。 $p = 1$ 与 $p = \infty$ 的情形只需使用 $|f + g| \leq |f| + |g|$, 结论很容易。以下设 $1 < p < \infty$ 。由

$$|f + g|^p \leq 2^p(|f|^p + |g|^p)$$

可知 $\int |f + g|^p d\mu < \infty$, 故 $f + g \in L^p$. 再把下列不等式关于 μ 积分:

$$|f + g|^p \leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}$$

我们得到

$$\int |f + g|^p d\mu \leq \int |f| |f + g|^{p-1} d\mu + \int |g| |f + g|^{p-1} d\mu.$$

将 Hölder 不等式应用于共轭指数 p 与 $q = p/(p-1)$, 得到

$$\int |f + g|^p d\mu \leq \|f\|_p \left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}} + \|g\|_p \left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}}.$$

若 $\int |f + g|^p d\mu = 0$, 则定理中的不等式显然成立。否则, 将前式两边同除以

$$\left(\int |f + g|^p d\mu \right)^{(p-1)/p},$$

即得所需结论。

定理 4.2.2 (Riesz) 对每个 $p \in [1, \infty]$, 赋予范数 $f \mapsto \|f\|_p$ 的空间 $L^p(E, \mathcal{A}, \mu)$ 是 Banach 空间, 即完备赋范向量空间。

证明。 只证 $1 \leq p < \infty$; $p = \infty$ 的情形更容易。先验证 $\|\cdot\|_p$ 是 L^p 上的范数。由

$$\|f\|_p = 0 \implies \int |f|^p d\mu = 0 \implies f = 0 \quad \mu\text{-几乎处处},$$

可知正定性成立。齐次性显然, 三角不等式就是 Minkowski 不等式。

下面证明完备性。设 (f_n) 是 L^p 中的 Cauchy 列。可选取严格递增整数列 (k_n) , 使

$$\|f_{k_{n+1}} - f_{k_n}\|_p \leq 2^{-n} \quad (n \geq 1).$$

令 $g_n = f_{k_n}$ 。由单调收敛定理与 Minkowski 不等式,

$$\begin{aligned} \int \left(\sum_{n=1}^{\infty} |g_{n+1} - g_n| \right)^p d\mu &= \lim_{N \rightarrow \infty} \int \left(\sum_{n=1}^N |g_{n+1} - g_n| \right)^p d\mu \\ &\leq \lim_{N \rightarrow \infty} \left(\sum_{n=1}^N \|g_{n+1} - g_n\|_p \right)^p < \infty. \end{aligned}$$

因此

$$\sum_{n=1}^{\infty} |g_{n+1} - g_n| < \infty \quad \mu\text{-几乎处处}.$$

在这个满测集上定义

$$h = g_1 + \sum_{n=1}^{\infty} (g_{n+1} - g_n),$$

并在余下零测集上令 $h = 0$ 。于是 h 可测且 $g_N \rightarrow h$ 几乎处处。Fatou 引理给出

$$\int |h|^p d\mu \leq \liminf_{N \rightarrow \infty} \int |g_N|^p d\mu \leq \sup_N \|g_N\|_p^p < \infty,$$

故 $h \in L^p$ 。再次用 Fatou 引理，并注意对 $N > n$,

$$\|g_N - g_n\|_p \leq \sum_{j=n}^{N-1} \|g_{j+1} - g_j\|_p \leq 2^{-n+1},$$

得到

$$\|h - g_n\|_p^p \leq \liminf_{N \rightarrow \infty} \|g_N - g_n\|_p^p \leq (2^{-n+1})^p.$$

所以子列 $g_n \rightarrow h$ 于 L^p 。原列 (f_n) 是 Cauchy 列，且有一个子列收敛到 h ，故整个序列也在 L^p 中收敛到 h 。

例。若 $E = \mathbb{N}$ ， μ 为计数测度，则对每个 $p \in [1, \infty)$ ，空间 L^p 就是满足

$$\sum_{n=0}^{\infty} |a_n|^p < \infty$$

的实数序列 $a = (a_n)_{n \in \mathbb{N}}$ 所成的空间，其范数为

$$\|a\|_p = \left(\sum_{n=0}^{\infty} |a_n|^p \right)^{1/p}.$$

而 L^∞ 是全体有界序列所成的空间，范数为 $\|a\|_\infty = \sup_n |a_n|$ 。在这个例子中，除空集外没有零测集，所以函数等价类可直接视为序列本身。该空间通常记为 $\ell^p = \ell^p(\mathbb{N})$ ，在 Banach 空间理论中十分重要。上一证明还给出一个值得单独陈述的中间结论。

命题 4.2.3 设 $p \in [1, \infty)$ ，并设函数列 (f_n) 在 $L^p(E, \mathcal{A}, \mu)$ 中收敛于 f 。则存在子列 (f_{k_n}) ，使 $f_{k_n} \rightarrow f$ 在 μ -几乎处处意义下成立。

注。该结论对 $p = \infty$ 也成立；此时甚至无须抽取子列，因为 L^∞ 收敛等价于在去掉一个零测集后的一致收敛。反过来，几乎处处收敛一般不能推出 L^p 收敛。不过，支配收敛定理说明：若

- (i) $f_n \rightarrow f$ ， μ -几乎处处；
 - (ii) 存在可测函数 $g \geq 0$ ，使 $\int g^p d\mu < \infty$ ，且对每个 n ，都有 $|f_n| \leq g$ ， μ -几乎处处，
- 则 $f_n \rightarrow f$ 在 L^p 中成立。

练习。设 $\mu(E) < \infty$ 、 $p \in [1, \infty)$ 。证明：若

- (i) $f_n \rightarrow f$ ， μ -几乎处处；
 - (ii) 存在 $r > p$ ，使 $\sup_n \int |f_n|^r d\mu < \infty$ ，
- 则 $f_n \rightarrow f$ 在 L^p 中成立。

Riesz 定理中 $p = 2$ 的情形尤其重要，因为 L^2 具有 Hilbert 空间结构。

定理 4.2.4 空间 $L^2(E, \mathcal{A}, \mu)$ 配备内积

$$\langle f, g \rangle = \int fg d\mu$$

后，是一个实 Hilbert 空间。

证明。由 Cauchy-Schwarz 不等式，若 $f, g \in L^2$ ，则 fg 可积，所以 $\langle f, g \rangle$ 良定义。显然，映射 $(f, g) \mapsto \langle f, g \rangle$ 是正定对称双线性型，其诱导范数正是 $\|f\|_2$ 。完备性由 Riesz 定理得到。因此，

Hilbert 空间的经典结论都可用于 $L^2(E, \mathcal{A}, \mu)$ 。特别地, 若

$$\Phi : L^2(E, \mathcal{A}, \mu) \rightarrow \mathbb{R}$$

是连续线性泛函, 则存在唯一的 $g \in L^2(E, \mathcal{A}, \mu)$, 使得

$$\forall f \in L^2(E, \mathcal{A}, \mu), \quad \Phi(f) = \langle f, g \rangle.$$

这一结果将在本章后面用到。**注:** 与前述结果一样, 上述定理也可推广到复数情形。复空间 $L^2_{\mathbb{C}}(E, \mathcal{A}, \mu)$ 在内积

$$\langle f, g \rangle = \int f \bar{g} d\mu$$

下是复 Hilbert 空间。

4.3 L^p 空间中的稠密性定理

若 (E, d) 是度量空间, 则称 $(E, \mathcal{B}(E))$ 上的测度 μ 是外正则的, 如果

$$\forall A \in \mathcal{B}(E), \quad \mu(A) = \inf\{\mu(U) : U \text{ 为开集且 } A \subset U\}.$$

称函数 $f : E \rightarrow \mathbb{R}$ 为 Lipschitz 函数, 如果存在常数 K , 使得

$$\forall x, y \in E, \quad |f(x) - f(y)| \leq Kd(x, y).$$

定理 4.3.1 设 $p \in [1, \infty)$ 。(1) 可积简单函数在 $L^p(E, \mathcal{A}, \mu)$ 中稠密。(2) 若 (E, d) 是度量空间, 且 μ 是 $(E, \mathcal{B}(E))$ 上的外正则测度, 则属于 L^p 的有界 Lipschitz 函数在 $L^p(E, \mathcal{B}(E), \mu)$ 中稠密。(3) 若 (E, d) 是局部紧致可分度量空间, 且 μ 是 E 上的 Radon 测度, 则紧支集 Lipschitz 函数在 $L^p(E, \mathcal{B}(E), \mu)$ 中稠密。

证明。(1) 写成 $f = f^+ - f^-$ 后, 只需考虑 $f \geq 0$ 。由简单函数逼近定理, 存在非负简单函数列 (φ_n) , 使得

$$f = \lim_{n \rightarrow \infty} \uparrow \varphi_n.$$

这里对每个 n , 都有 $0 \leq \varphi_n \leq f$, 且 φ_n 是简单函数。于是

$$\int |\varphi_n|^p d\mu \leq \int |f|^p d\mu < \infty,$$

所以 $\varphi_n \in L^p$ (对简单函数而言, 这等价于 $\varphi_n \in L^1$)。又因 $|f - \varphi_n|^p \leq f^p$, 支配收敛定理给出

$$\lim_{n \rightarrow \infty} \int |f - \varphi_n|^p d\mu = 0.$$

(2) 只需证明每个可积简单函数都能在 L^p 中由有界 Lipschitz 函数逼近。容易归结到 $f = 1_A$, $A \in \mathcal{B}(E)$ 、 $\mu(A) < \infty$ 的情形。给定 $\varepsilon > 0$, 由外正则性可取包含 A 的开集 O , 使

$$\mu(O \setminus A) < (\varepsilon/2)^p, \quad \|1_O - 1_A\|_p < \varepsilon/2.$$

对每个 $k \geq 1$, 令

$$\varphi_k(x) = (kd(x, O^c)) \wedge 1.$$

则 φ_k 是 Lipschitz 函数, 并且当 $k \rightarrow \infty$ 时, $\varphi_k \uparrow 1_O$ 。由支配收敛定理,

$$\int |1_O - \varphi_k|^p d\mu \rightarrow 0.$$

故可取充分大的 k , 使 $\|1_O - \varphi_k\|_p < \varepsilon/2$ 。最后,

$$\|1_A - \varphi_k\|_p \leq \|1_A - 1_O\|_p + \|1_O - \varphi_k\|_p < \varepsilon.$$

(3) 使用下面的引理, 其证明留到本证明的最后。回忆: 若 $A \subset E$, 则 A° 表示 A 的内部。

引理 4.3.2 设 E 是局部紧致可分度量空间。则存在递增的紧集列 $(L_n)_{n \geq 1}$, 使对每个 n ,

$$L_n \subset L_{n+1}^\circ, \quad E = \bigcup_{n \geq 1} L_n = \bigcup_{n \geq 1} L_n^\circ.$$

由该引理容易推出: E 上的每个 Radon 测度 μ 都是外正则的 (Riesz 表示定理的陈述中已经使用过这一事实, 但当时未给出证明)。事实上, 若 A 是 E 的 Borel 集, 则把 μ 限制在 L_n° 上可得有限测度。由前章的外正则性结论, 对每个 n 可取开集 $O_n \subset L_n^\circ$, 使 $A \cap L_n^\circ \subset O_n$, 并且

$$\mu(O_n \setminus (A \cap L_n^\circ)) \leq \varepsilon 2^{-n}.$$

令 $O = \bigcup_n O_n$ 。则 O 是 E 的开集, 并且

$$\mu(O \setminus A) \leq \sum_{n \geq 1} \mu(O_n \setminus (A \cap L_n^\circ)) \leq \varepsilon.$$

现在, 由于 μ 外正则, 可以应用定理的第 (2) 部分。因此只需证明: 任意有界 Lipschitz 函数 f , 若 $\int |f|^p d\mu < \infty$, 都能在 L^p 中由紧支集 Lipschitz 函数逼近。

由支配收敛定理,

$$\lim_{n \rightarrow \infty} \int_{(L_n^\circ)^c} |f|^p d\mu = 0,$$

故 $\|f - f1_{L_n^\circ}\|_p \rightarrow 0$ 。另一方面, 固定 n , 对每个 $k \geq 1$ 定义

$$\varphi_{n,k}(x) = (kd(x, (L_n^\circ)^c)) \wedge 1.$$

由于 $\varphi_{n,k} \leq 1_{L_n^\circ}$, 故 $\varphi_{n,k} \in L^p$ 。再次使用支配收敛定理可知: 对每个固定的 n , 当 $k \rightarrow \infty$ 时, $\varphi_{n,k} \rightarrow 1_{L_n^\circ}$ 在 L^p 中成立。最后,

$$\begin{aligned} \|f - f\varphi_{n,k}\|_p &\leq \|f - f1_{L_n^\circ}\|_p + \|f1_{L_n^\circ} - f\varphi_{n,k}\|_p \\ &\leq \|f - f1_{L_n^\circ}\|_p + \|f\|_\infty \|1_{L_n^\circ} - \varphi_{n,k}\|_p. \end{aligned}$$

依次取充分大的 n 与 k , 便可用紧支集 Lipschitz 函数 $f\varphi_{n,k}$ 在 L^p 中逼近 f 。**引理的证明。**先证明 E 是一列递增紧集 $(K_n)_{n \geq 1}$ 的并。取 E 中稠密序列 $(x_p)_{p \geq 0}$, 并定义整数对集合

$$I = \{(p, k) \in \mathbb{N}^2 : \overline{B}(x_p, 2^{-k}) \text{ 是紧集}\}.$$

这里 $\overline{B}(x, r)$ 表示以 x 为中心、半径为 r 的闭球。由 E 的局部紧致性以及序列 (x_p) 的稠密性, 容易看出

$$E = \bigcup_{(p,k) \in I} \overline{B}(x_p, 2^{-k}).$$

集合 I 可数, 故存在 I 的递增有限子集列 $(I_n)_{n \geq 1}$, 其并为 I 。令

$$K_n = \bigcup_{(p,k) \in I_n} \overline{B}(x_p, 2^{-k}),$$

便得到所需的递增紧集列。

然后递归构造 (L_n) 。先取 $L_1 = K_1$ 。若已构造 L_n , 则用有限多个具有紧闭包的开邻域 $V_1, \dots, V_q$ 覆盖紧集 $K_{n+1} \cup L_n$, 并令

$$L_{n+1} = \overline{V}_1 \cup \dots \cup \overline{V}_q.$$

这就得到引理要求的紧集列。

推论。对 $p \in [1, \infty)$, 有: (i) $\mathbb{R}^d$ 上的紧支集连续函数空间 $C_c(\mathbb{R}^d)$ 在 $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ 中稠密。这里的 Lebesgue 测度 λ 也可替换为 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上的任意 Radon 测度。

(ii) 紧支集阶梯函数在 $L^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ 中稠密。事实上, 只需证明每个 $f \in C_c(\mathbb{R})$ 都可由阶梯函数在 L^p 中逼近; 一致连续性给出

$$f = \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{Z}} f\left(\frac{k}{n}\right) 1_{[k/n, (k+1)/n)} \quad \text{于 } L^p.$$

应用 (Riemann–Lebesgue 引理)。 若 $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, 并定义

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{ix\xi} dx,$$

则

$$\widehat{f}(\xi) \longrightarrow 0 \quad (|\xi| \rightarrow \infty).$$

证明。 由稠密性, 可取紧支集阶梯函数 φ_n , 使 $\|f - \varphi_n\|_1 \rightarrow 0$ 。于是

$$\sup_{\xi \in \mathbb{R}} |\widehat{f}(\xi) - \widehat{\varphi}_n(\xi)| \leq \|f - \varphi_n\|_1 \longrightarrow 0.$$

因此只需处理阶梯函数。若

$$f = \sum_{j=1}^m \lambda_j 1_{(x_j, x_{j+1})},$$

则

$$\widehat{f}(\xi) = \sum_{j=1}^m \lambda_j \frac{e^{i\xi x_{j+1}} - e^{i\xi x_j}}{i\xi} \longrightarrow 0 \quad (|\xi| \rightarrow \infty),$$

结论得证。

4.4 Radon–Nikodym 定理

定义 4.4.1 设 μ, ν 是可测空间 $(E, \mathcal{A})$ 上的测度。

1. 若对每个 $A \in \mathcal{A}$,

$$\mu(A) = 0 \implies \nu(A) = 0,$$

则称 ν 关于 μ 绝对连续, 记作 $\nu \ll \mu$ 。

2. 若存在 $N \in \mathcal{A}$, 使

$$\mu(N) = 0, \quad \nu(N^c) = 0,$$

则称 ν 与 μ 互相奇异, 记作 $\nu \perp \mu$ 。

例。若 $f: E \rightarrow [0, \infty]$ 可测, 并定义

$$\nu(A) = \int_A f d\mu, \quad A \in \mathcal{A},$$

则 $\nu = f \cdot \mu \ll \mu$ 。

定理 4.4.1 (Radon–Nikodym 定理与 Lebesgue 分解) 设 μ, ν 是 $(E, \mathcal{A})$ 上的 σ -有限测度。则存在唯一的一对 σ -有限测度 (ν_a, ν_s) , 使

$$\nu = \nu_a + \nu_s, \quad \nu_a \ll \mu, \quad \nu_s \perp \mu.$$

此外, 存在可测函数 $g: E \rightarrow [0, \infty]$, 使

$$\nu_a(A) = \int_A g d\mu, \quad A \in \mathcal{A}.$$

函数 g 在 μ -几乎处处意义下唯一, 记作 $g = d\nu_a/d\mu$ 。

证明。 先设 μ, ν 都是有限测度; σ -有限情形将在最后处理。先考虑 $\nu \leq \mu$ 。定义 $L^2(E, \mathcal{A}, \mu)$ 上的线性泛函

$$\Phi(f) = \int f d\nu.$$

这个积分是良定义的: 有限测度情形下 $L^2(\mu) \subset L^1(\mu)$, 且 $\int |f| d\nu \leq \int |f| d\mu$ 。若 $f = \tilde{f}$ 在 μ 意义下几乎处处成立, 则因 $\nu \leq \mu$, 该等式也在 ν 意义下几乎处处成立, 所以 Φ 与代表元的选择无关。由 Cauchy–Schwarz 不等式,

$$|\Phi(f)| \leq \left(\int f^2 d\nu \right)^{1/2} \nu(E)^{1/2} \leq \nu(E)^{1/2} \|f\|_{L^2(\mu)}.$$

故 Φ 连续。由 Hilbert 空间的 Riesz 表示定理, 存在 $h \in L^2(\mu)$, 使对每个 $f \in L^2(\mu)$,

$$\Phi(f) = \langle f, h \rangle = \int fh d\mu.$$

特别地, 取 $f = 1_A$ 得

$$\nu(A) = \int_A h d\mu, \quad A \in \mathcal{A}.$$

还可取一个代表使 $0 \leq h \leq 1$ 处处成立。事实上, 对每个 $\varepsilon > 0$,

$$\mu\{h \geq 1 + \varepsilon\} \geq \nu\{h \geq 1 + \varepsilon\} = \int_{\{h \geq 1 + \varepsilon\}} h d\mu \geq (1 + \varepsilon)\mu\{h \geq 1 + \varepsilon\},$$

故该集合为零测集; 同理 $h \geq 0$ 几乎处处。现在处理一般的有限测度情形。把上述结论应用于 $\nu \leq \mu + \nu$, 得到可测函数 h , $0 \leq h \leq 1$, 使对任意 $f \in L^2(\mu + \nu)$,

$$\int f d\nu = \int fh d(\mu + \nu).$$

因而对任意有界可测函数 f ,

$$\int f(1 - h) d\nu = \int fh d\mu.$$

由单调收敛定理, 该等式对所有非负可测函数仍成立。令 $N = \{h = 1\}$ 。取 $f = 1_N$ 可得 $\mu(N) = 0$ 。定义

$$\nu_s = 1_N \nu, \quad g = 1_{N^c} \frac{h}{1-h}, \quad \nu_a = 1_{N^c} \nu = g\mu.$$

于是 $\nu = \nu_a + \nu_s$ 、 $\nu_a \ll \mu$, 且 $\nu_s \perp \mu$ 。再证唯一性。若 $(\tilde{\nu}_a, \tilde{\nu}_s)$ 是另一组分解, 则

$$\nu_a(A) - \tilde{\nu}_a(A) = \nu_s(A) - \tilde{\nu}_s(A), \quad A \in \mathcal{A}.$$

取 μ -零测集 $N, \tilde{N}$, 分别支撑 $\nu_s, \tilde{\nu}_s$ 。右端只依赖于 $A \cap (N \cup \tilde{N})$, 而 ν_a 与 $\tilde{\nu}_a$ 都对 μ 绝对连续, 故右端为零。因此两部分各自唯一。若 $g, \tilde{g}$ 都是 ν_a 关于 μ 的密度, 则

$$\int_{\{\tilde{g} > g\}} \tilde{g} d\mu = \nu_a\{\tilde{g} > g\} = \int_{\{\tilde{g} > g\}} g d\mu,$$

所以 $\tilde{g} \leq g$ 几乎处处; 交换二者即得 $g = \tilde{g}$ 几乎处处。最后设 μ, ν 仅为 σ -有限。可取 E 的可数可测分割 (E_n) , 使 $\mu(E_n), \nu(E_n) < \infty$ 。将有限测度情形用于限制测度 $\mu_n = \mu|_{E_n}$ 、 $\nu_n = \nu|_{E_n}$, 写成 $\nu_n = \nu_{a,n} + \nu_{s,n}$, 其中 $\nu_{a,n} = g_n \mu_n$ 、 $\nu_{s,n} \perp \mu_n$, 并令 g_n 在 E_n^c 上为零。于是

$$\nu_a = \sum_n \nu_{a,n}, \quad \nu_s = \sum_n \nu_{s,n}, \quad g = \sum_n g_n$$

给出所需分解; 由于 (E_n) 是分割, 每一点至多有一个 g_n 非零。唯一性仍由上面的论证得到。

第五章 乘积测度

给定两个各自带测度的可测空间，我们可以在它们的笛卡尔积上构造一个称为乘积测度的测度。此外，定义在乘积空间上的函数的积分可以通过先对第一个空间的测度积分，然后对第二个空间的测度积分来计算，或者按相反的顺序进行：这就是著名的 Fubini 定理。除了在分析（分部积分、卷积等）或概率论中的重要应用外，Fubini 定理还是计算积分的一个基本工具。

5.1 乘积空间概述

设 $(E, \mathcal{A})$ 与 $(F, \mathcal{B})$ 为两个可测空间。在 $E \times F$ 上定义乘积 σ -代数

$$\mathcal{A} \otimes \mathcal{B} = \sigma(A \times B; A \in \mathcal{A}, B \in \mathcal{B}).$$

形如 $A \times B$ 的集合称为可测矩形。容易验证， $\mathcal{A} \otimes \mathcal{B}$ 是使两个典范投影 $\pi_1 : E \times F \rightarrow E$ 、 $\pi_2 : E \times F \rightarrow F$ 都可测的最小 σ -代数。若 $(G, \mathcal{C})$ 是第三个可测空间， $f : G \rightarrow E \times F$ ，并写成 $f(x) = (f_1(x), f_2(x))$ ，则 f 可测当且仅当 f_1, f_2 都可测。乘积 σ -代数可直接推广到有限多个可测空间 $(E_i, \mathcal{A}_i)$ ：

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_n = \sigma(A_1 \times \cdots \times A_n; A_i \in \mathcal{A}_i)$$

并具有预期的结合律。例如当 $n = 3$ 时，

$$(\mathcal{A}_1 \otimes \mathcal{A}_2) \otimes \mathcal{A}_3 = \mathcal{A}_1 \otimes (\mathcal{A}_2 \otimes \mathcal{A}_3) = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3.$$

命题 5.1.1 若 E, F 是可分度量空间，则

$$\mathcal{B}(E \times F) = \mathcal{B}(E) \otimes \mathcal{B}(F).$$

证明。 包含关系 $\mathcal{B}(E \times F) \supset \mathcal{B}(E) \otimes \mathcal{B}(F)$ 不需要可分性：坐标投影 π_1, π_2 连续，因而关于 $\mathcal{B}(E \times F)$ 可测。反过来，由可分性， E 有可数拓扑基 $\mathcal{U} = \{U_n : n \geq 1\}$ ， F 也有可数拓扑基 $\mathcal{V} = \{V_m : m \geq 1\}$ 。例如，可取以稠密可数集中的点为中心、以正有理数为半径的开球。若 $O \subset E \times F$ 为开集，则 O 是所有满足 $U_n \times V_m \subset O$ 的基元素 $U_n \times V_m$ 的并；这是可数并。因此 $O \in \mathcal{B}(E) \otimes \mathcal{B}(F)$ ，从而得到反向包含关系。

下面回到任意可测空间 $(E, \mathcal{A})$ 、 $(F, \mathcal{B})$ 。对 $C \subset E \times F$ ，定义竖直截面与水平截面

$$C_x = \{y \in F : (x, y) \in C\}, \quad C^y = \{x \in E : (x, y) \in C\}.$$

若 f 定义在 $E \times F$ 上，则记 $f_x(y) = f(x, y)$ 、 $f^y(x) = f(x, y)$ 。

定理 5.1.2 (i) 若 $C \in \mathcal{A} \otimes \mathcal{B}$ ，则对每个 $x \in E$ ， $C_x \in \mathcal{B}$ ；对每个 $y \in F$ ， $C^y \in \mathcal{A}$ 。

(ii) 若 $f : E \times F \rightarrow G$ 关于积 σ -代数 $\mathcal{A} \otimes \mathcal{B}$ 可测，则对每个 $x \in E$ ， f_x 为 $\mathcal{B}$ -可测；对每个 $y \in F$ ， f^y 为 $\mathcal{A}$ -可测。

证明。(i) 固定 $x \in E$, 令

$$\mathcal{C}_x = \{C \in \mathcal{A} \otimes \mathcal{B} : C_x \in \mathcal{B}\}.$$

$\mathcal{C}_x$ 是 σ -代数, 并包含所有可测矩形, 因为

$$(A \times B)_x = \begin{cases} B, & x \in A, \\ \emptyset, & x \notin A. \end{cases}$$

故 $\mathcal{C}_x = \mathcal{A} \otimes \mathcal{B}$ 。水平截面同理。(ii) 若 $D \subset G$ 可测, 则

$$f_x^{-1}(D) = \{y \in F : (x, y) \in f^{-1}(D)\} = (f^{-1}(D))_x \in \mathcal{B},$$

其中最后一步使用 (i); f^y 的证明相同。

5.2 乘积测度的构造

定理 5.2.1 设 μ, ν 分别是 $(E, \mathcal{A})$ 、 $(F, \mathcal{B})$ 上的 σ -有限测度。

(i) 存在唯一的 $(E \times F, \mathcal{A} \otimes \mathcal{B})$ 上的测度 m , 使

$$m(A \times B) = \mu(A)\nu(B) \quad (A \in \mathcal{A}, B \in \mathcal{B}),$$

其中约定 $0 \cdot \infty = 0$ 。测度 m 也是 σ -有限的, 记为 $\mu \otimes \nu$ 。

(ii) 对每个 $C \in \mathcal{A} \otimes \mathcal{B}$,

$$(\mu \otimes \nu)(C) = \int_E \nu(C_x) \mu(dx) = \int_F \mu(C^y) \nu(dy).$$

证明。唯一性。由 σ -有限性, 可取递增可测集列 $A_n \uparrow E, B_n \uparrow F$, 使 $\mu(A_n) < \infty, \nu(B_n) < \infty$ 。令 $C_n = A_n \times B_n$, 则 $C_n \uparrow E \times F$ 。若 m, m' 都满足 (i), 则二者在所有可测矩形上相等, 而且

$$m(C_n) = \mu(A_n)\nu(B_n) = m'(C_n) < \infty.$$

可测矩形族对有限交封闭, 并生成 $\mathcal{A} \otimes \mathcal{B}$ 。由推论 1.4.2, $m = m'$ 。¹

存在性。对 $C \in \mathcal{A} \otimes \mathcal{B}$, 拟定义

$$m(C) = \int_E \nu(C_x) \mu(dx). \quad (5.1)$$

定理 5.1.2 保证 $C_x \in \mathcal{B}$ 。还需证明 $x \mapsto \nu(C_x)$ 是 $\mathcal{A}$ -可测的。先设 ν 有限, 并令

$$\mathcal{G} = \{C \in \mathcal{A} \otimes \mathcal{B} : x \mapsto \nu(C_x) \text{ 是 } \mathcal{A}\text{-可测的}\}.$$

若 $C = A \times B$, 则 $\nu(C_x) = 1_A(x)\nu(B)$, 所以 $\mathcal{G}$ 包含所有可测矩形。它还是单调类: 若 $C \subset C'$ 且 $C, C' \in \mathcal{G}$, 则

$$\nu((C' \setminus C)_x) = \nu(C'_x) - \nu(C_x),$$

这里使用了 ν 的有限性; 若 $C_n \uparrow C$, 则

$$\nu(C_x) = \lim_{n \rightarrow \infty} \uparrow \nu((C_n)_x).$$

¹校订注: 原讲义在选择 B_n 时把有限性条件误写为 $\mu(B_n) < \infty$; $B_n \subset F$, 正确条件是 $\nu(B_n) < \infty$ 。

由单调类定理, $\mathcal{G} = \mathcal{A} \otimes \mathcal{B}$ 。²一般地, 若 ν 仅为 σ -有限, 令 $\nu_n(D) = \nu(D \cap B_n)$ 。有限测度情形说明 $x \mapsto \nu_n(C_x)$ 可测, 而

$$\nu(C_x) = \lim_{n \rightarrow \infty} \uparrow \nu_n(C_x),$$

所以所需可测性仍成立。

下面验证 (5.1) 定义了测度。若 (C_n) 两两不交, 则对每个 x , $((C_n)_x)$ 也两两不交。因此由非负函数积分的可列可加性,

$$\begin{aligned} m\left(\bigcup_n C_n\right) &= \int_E \nu\left(\bigcup_n (C_n)_x\right) \mu(dx) \\ &= \int_E \sum_n \nu((C_n)_x) \mu(dx) \\ &= \sum_n \int_E \nu((C_n)_x) \mu(dx) = \sum_n m(C_n). \end{aligned}$$

并且

$$m(A \times B) = \int_E 1_A(x) \nu(B) \mu(dx) = \mu(A) \nu(B).$$

集合 $C_n = A_n \times B_n$ 以有限 m -测度递增覆盖 $E \times F$, 所以 m 为 σ -有限测度。交换两个空间的角色, 同样得到测度

$$m'(C) = \int_F \mu(C^y) \nu(dy).$$

m' 在可测矩形上的取值与 m 相同, 故由唯一性 $m = m'$, 这也证明了 (ii)。

注。(i) 至少对结论 (ii) 而言, σ -有限性不可缺少。取 $(E, \mathcal{A}) = (F, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 令 $\mu = \lambda$, ν 为计数测度, 并取对角线 $C = \{(x, x) : x \in \mathbb{R}\}$ 。则

$$\int_{\mathbb{R}} \nu(C_x) \lambda(dx) = \infty, \quad \int_{\mathbb{R}} \lambda(C^y) \nu(dy) = 0.$$

(ii) 对 σ -有限测度 $\mu_1, \dots, \mu_n$, 可递归定义

$$\mu_1 \otimes \cdots \otimes \mu_n = \mu_1 \otimes (\mu_2 \otimes \cdots \otimes \mu_n).$$

括号的放置不影响结果, 因为该测度由矩形上的取值唯一确定:

$$(\mu_1 \otimes \cdots \otimes \mu_n)(A_1 \times \cdots \times A_n) = \prod_{j=1}^n \mu_j(A_j).$$

例。若两个因子都是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, 则 $\lambda \otimes \lambda$ 正是 $\mathbb{R}^2$ 上的 Lebesgue 测度, 因为二者在每个矩形 $[a, b] \times [c, d]$ 上取值相同。同理, 一维 Lebesgue 测度的乘积给出各维 Lebesgue 测度。

5.3 Fubini 定理

先给出非负函数版本。沿用上一节的记号: $(E, \mathcal{A})$ 与 $(F, \mathcal{B})$ 是可测空间, $E \times F$ 配备 $\mathcal{A} \otimes \mathcal{B}$ 。

²校订注: 原讲义在集合差公式中把 $C' \setminus C$ 误排为 $C \setminus C'$, 并相应打乱了等式右端; 此处已改正。

定理 5.3.1 (Fubini–Tonelli) 设 μ, ν 分别是 $(E, \mathcal{A})$ 、 $(F, \mathcal{B})$ 上的 σ -有限测度, $f: E \times F \rightarrow [0, \infty]$ 可测。

(i) 函数

$$x \mapsto \int_F f(x, y) \nu(dy), \quad y \mapsto \int_E f(x, y) \mu(dx)$$

分别为 $\mathcal{A}$ -可测与 $\mathcal{B}$ -可测。

(ii) 有

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) = \int_F \left(\int_E f(x, y) \mu(dx) \right) \nu(dy).$$

证明。 (i) 先设 $f = 1_C$, 其中 $C \in \mathcal{A} \otimes \mathcal{B}$ 。上一节已经证明

$$x \mapsto \int_F f(x, y) \nu(dy) = \nu(C_x)$$

是 $\mathcal{A}$ -可测的; 另一个函数同理。由线性性, 结论对非负简单函数成立。一般地, 取非负简单函数 $f_n \uparrow f$, 并使用

$$\int_F f(x, y) \nu(dy) = \lim_{n \rightarrow \infty} \int_F f_n(x, y) \nu(dy).$$

(ii) 当 $f = 1_C$ 时, 所需等式就是

$$(\mu \otimes \nu)(C) = \int_E \nu(C_x) \mu(dx) = \int_F \mu(C^y) \nu(dy),$$

已由上一节证明。³由线性性, 等式对非负简单函数成立; 再取 $f_n \uparrow f$, 两次应用单调收敛定理, 即得一般情形。

下面转向可取任意符号的函数, 仍保留上一条定理的假设。

定理 5.3.2 (Fubini–Lebesgue) 设 $f \in \mathcal{L}^1(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ (复值情形同理)。则:

(a) 对 μ -几乎所有的 x , 函数 $y \mapsto f(x, y)$ 属于 $\mathcal{L}^1(F, \mathcal{B}, \nu)$; 对 ν -几乎所有的 y , 函数 $x \mapsto f(x, y)$ 属于 $\mathcal{L}^1(E, \mathcal{A}, \mu)$ 。

(b) 在零测集上任意补定义后, 函数

$$x \mapsto \int_F f(x, y) \nu(dy), \quad y \mapsto \int_E f(x, y) \mu(dx)$$

分别属于 $\mathcal{L}^1(E, \mathcal{A}, \mu)$ 与 $\mathcal{L}^1(F, \mathcal{B}, \nu)$ 。

(c) 有

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) = \int_F \left(\int_E f(x, y) \mu(dx) \right) \nu(dy).$$

证明。 (a) 把 Fubini–Tonelli 定理应用于 $|f|$, 得到

$$\int_E \left(\int_F |f(x, y)| \nu(dy) \right) \mu(dx) = \int |f| d\mu \otimes \nu < \infty.$$

因为右端有限, 所以对 μ -几乎所有 x ,

$$\int_F |f(x, y)| \nu(dy) < \infty$$

³校订注: 原讲义该式最后一个截面误排为 C^x ; 积分变量是 y , 且按定义应为水平截面 C^y 。

因此, 对 μ -几乎所有 x , 可测函数 $y \mapsto f(x, y)$ 属于 $\mathcal{L}^1(F, \mathcal{B}, \nu)$ 。交换 x, y 的角色可得另一半结论。(b) 写成 $f = f^+ - f^-$, 并分别使用 Fubini–Tonelli 定理, 可知

$$x \longrightarrow \int f(x, y) \nu(dy) = \int f^+(x, y) \nu(dy) - \int f^-(x, y) \nu(dy)$$

是可测的。严格地说, 对满足 $\int_F |f(x, y)| \nu(dy) = \infty$ 的 x , 应给 $\int_F f(x, y) \nu(dy)$ 任取一个值 (例如 0); 这些 x 构成一个零测集。此外,

$$\int_E \left| \int_F f(x, y) \nu(dy) \right| \mu(dx) \leq \int_E \left(\int_F |f(x, y)| \nu(dy) \right) \mu(dx) = \int |f| d(\mu \otimes \nu) < \infty.$$

(c) 只需在等式中逐项相减即可

$$\begin{aligned} \int_E \left(\int_F f^+(x, y) \nu(dy) \right) \mu(dx) &= \int_{E \times F} f^+ d\mu \otimes \nu \\ \int_E \left(\int_F f^-(x, y) \nu(dy) \right) \mu(dx) &= \int_{E \times F} f^- d\mu \otimes \nu. \end{aligned}$$

注。假设 $f \in \mathcal{L}^1(\mu \otimes \nu)$ 至关重要。即使 (a)、(b) 中的两个累次积分各自都有意义, 它们也可能不相等:

$$\int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) \quad \text{与} \quad \int_F \left(\int_E f(x, y) \mu(dx) \right) \nu(dy).$$

这两个累次积分都可能有意义, 却不一定相等。下面给出一个典型反例。在 $(0, \infty) \times (0, 1]$ 上定义

$$f(x, y) = 2e^{-2xy} - e^{-xy}.$$

对每个 $y \in (0, 1]$, 有

$$\int_0^\infty f(x, y) dx = 2 \int_0^\infty e^{-2xy} dx - \int_0^\infty e^{-xy} dx = 0.$$

而对每个 $x > 0$, 有

$$\int_0^1 f(x, y) dy = 2 \int_0^1 e^{-2xy} dy - \int_0^1 e^{-xy} dy = \frac{e^{-x} - e^{-2x}}{x}.$$

因此

$$\int_0^1 \left(\int_0^\infty f(x, y) dx \right) dy = 0,$$

但

$$\int_0^\infty \left(\int_0^1 f(x, y) dy \right) dx = \int_0^\infty \frac{e^{-x} - e^{-2x}}{x} dx = \log 2 > 0.$$

这里确有

$$\int_{(0, \infty) \times (0, 1]} |f(x, y)| dx dy = \infty,$$

所以 f 不满足 Fubini–Lebesgue 定理的绝对可积条件。实际使用时应记住: Fubini–Tonelli 定理可直接用于非负可测函数; 对可取正负值的函数, 通常要先验证

$$\int |f| d(\mu \otimes \nu) < \infty.$$

只有在相应换序确有依据时, 才省略括号写成

$$\int f d(\mu \otimes \nu) = \int_E \int_F f(x, y) \nu(dy) \mu(dx).$$

5.4 应用

5.4.1 分部积分

设 $f, g: \mathbb{R} \rightarrow \mathbb{R}$ 是局部可积函数。对 $x \in \mathbb{R}$, 定义

$$F(x) = \int_0^x f(t) dt = \begin{cases} \int_{[0,x]} f(t) dt, & x \geq 0, \\ -\int_{[x,0]} f(t) dt, & x < 0, \end{cases}$$

$$G(x) = \int_0^x g(t) dt.$$

于是, 对任意 $a < b$, 分部积分公式为

$$F(b)G(b) = F(a)G(a) + \int_a^b f(t)G(t)dt + \int_a^b F(t)g(t)dt.$$

我们可以很容易地看到这个等式等价于

$$\int_a^b f(t)(G(t) - G(a)) dt = \int_a^b (F(b) - F(t))g(t) dt.$$

为了证明最后这个等式, 我们写成

$$\begin{aligned} \int_a^b f(t)(G(t) - G(a)) dt &= \int_a^b f(t) \left(\int_a^t g(s) ds \right) dt \\ &= \int_a^b \left(\int_a^b 1_{\{s \leq t\}} f(t) g(s) ds \right) dt \\ &= \int_a^b \left(\int_a^b 1_{\{s \leq t\}} f(t) g(s) dt \right) ds \\ &= \int_a^b g(s) \left(\int_s^b f(t) dt \right) ds \\ &= \int_a^b g(s)(F(b) - F(s)) ds. \end{aligned}$$

在第三个等式中, 我们将 Fubini-Lebesgue 定理应用于函数

$$\varphi(s, t) = 1_{\{s \leq t\}} f(t) g(s)$$

注意到, 由于 Fubini-Tonelli 定理,

$$\int_{[a,b]^2} |\varphi(s, t)| ds dt \leq \int_{[a,b]^2} |f(t)| |g(s)| ds dt = \left(\int_{[a,b]} |f(t)| dt \right) \left(\int_{[a,b]} |g(s)| ds \right)$$

5.4.2 卷积

若 $f, g: \mathbb{R}^d \rightarrow \mathbb{R}$ 可测, 则在下列绝对收敛条件成立时定义卷积

$$f * g(x) = \int_{\mathbb{R}^d} f(x - y) g(y) dy$$

在满足以下条件下是良确定的

$$\int_{\mathbb{R}^d} |f(x - y) g(y)| dy < \infty.$$

在这种情形下, 利用 Lebesgue 测度在平移及反射 $y \mapsto -y$ 下的不变性作变量替换, 可知 $(g * f)(x)$ 也有定义, 且 $(g * f)(x) = (f * g)(x)$ 。

命题 5.4.1 若 $f, g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, 则对 λ -几乎所有 $x \in \mathbb{R}^d$, 卷积 $(f * g)(x)$ 有定义。此外, $f * g \in L^1(\lambda)$, 且

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

注。 这里把只在几乎处处有定义的函数视为 L^1 元素是没有问题的: 只需在未定义的零测集上任意补定义, 例如令其为 0。

证明。 由 Fubini–Tonelli 定理,

$$\begin{aligned} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt \right) dx &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dx \right) dt \\ &= \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x-t)| dx \right) dt \\ &= \left(\int_{\mathbb{R}^d} |g(t)| dt \right) \left(\int_{\mathbb{R}^d} |f(x)| dx \right) \\ &< \infty \end{aligned}$$

所以

$$\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt < \infty \quad \text{对 } dx\text{-几乎所有 } x \text{ 成立。}$$

这证明了第一条断言。对第二条, 继续使用同一计算可得

$$\int_{\mathbb{R}^d} |f * g(x)| dx \leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt \right) dx = \|f\|_1 \|g\|_1 < \infty.$$

下一命题给出卷积处处有定义的另一组充分条件。

命题 5.4.2 设 $p \in [1, \infty)$ 、 $q \in (1, \infty]$, 且 $p^{-1} + q^{-1} = 1$ 。若 $f \in L^p(\mathbb{R}^d)$ 、 $g \in L^q(\mathbb{R}^d)$, 则对每个 $x \in \mathbb{R}^d$, 卷积 $(f * g)(x)$ 都有定义; 并且 $f * g$ 在 $\mathbb{R}^d$ 上有界且一致连续。

证明。 Hölder 不等式给出

$$\int_{\mathbb{R}^d} |f(x-y)g(y)| dy \leq \left(\int_{\mathbb{R}^d} |f(x-y)|^p dy \right)^{1/p} \|g\|_q = \|f\|_p \|g\|_q.$$

这既证明了卷积处处有定义, 也给出 $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$ 。为证明一致连续性, 使用下列引理。

引理 5.4.3 记 $\sigma_x(y) = y - x$ 。若 $p \in [1, \infty)$ 且 $f \in L^p(\mathbb{R}^d)$, 则映射

$$x \mapsto f \circ \sigma_x$$

从 $\mathbb{R}^d$ 到 $L^p(\mathbb{R}^d)$ 一致连续。

由该引理, 对任意 $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} |f * g(x) - f * g(x')| &\leq \int |f(x-y) - f(x'-y)| |g(y)| dy \\ &\leq \|g\|_q \left(\int |f(x-y) - f(x'-y)|^p dy \right)^{1/p} \\ &= \|g\|_q \|f \circ \sigma_{-x} - f \circ \sigma_{-x'}\|_p \end{aligned}$$

当 $x - x' \rightarrow 0$ 时, 最后一项趋于 0, 故 $f * g$ 一致连续。下面证明引理。先设 $f \in C_c(\mathbb{R}^d)$, 则

$$\int |f \circ \sigma_x - f \circ \sigma_y|^p d\lambda = \int |f(z-x) - f(z-y)|^p dz = \int |f(z) - f(z-(y-x))|^p dz.$$

当 $y-x \rightarrow 0$ 时, 上式由支配收敛定理趋于 0。一般情形下, 可取 $f_n \in C_c(\mathbb{R}^d)$, 使 f_n 在 $L^p(\lambda)$ 中收敛到 f (见第 4 章)。于是

$$\begin{aligned} \|f \circ \sigma_x - f \circ \sigma_y\|_p &\leq \|f \circ \sigma_x - f_n \circ \sigma_x\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p + \|f_n \circ \sigma_y - f \circ \sigma_y\|_p \\ &= 2\|f - f_n\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p. \end{aligned}$$

给定 $\varepsilon > 0$, 先取 n 使 $\|f - f_n\|_p < \varepsilon/4$, 再取 $\delta > 0$, 使得当 $|x-y| < \delta$ 时,

$$\|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p \leq \varepsilon/2.$$

由前面的估计可知, 只要 $|x-y| < \delta$, 就有 $\|f \circ \sigma_x - f \circ \sigma_y\|_p \leq \varepsilon$ 。

Dirac 测度的逼近。 若函数列 $(\varphi_n) \subset C_c(\mathbb{R}^d)$ 满足:

1. 存在紧集 $K \subset \mathbb{R}^d$, 使 $\text{supp}(\varphi_n) \subset K$ 对所有 n 成立;
2. $\varphi_n \geq 0$, 且 $\int_{\mathbb{R}^d} \varphi_n(x) dx = 1$;
3. 对每个 $\delta > 0$,

$$\lim_{n \rightarrow \infty} \int_{\{|x| > \delta\}} \varphi_n(x) dx = 0.$$

就称 (φ_n) 为 δ_0 的一个逼近恒等族。这种函数列容易构造: 取非负函数 $\varphi \in C_c(\mathbb{R}^d)$, 使 $\int_{\mathbb{R}^d} \varphi(x) dx = 1$, 并令

$$\varphi_n(x) = n^d \varphi(nx).$$

还可令每个 $\varphi_n \in C_c^\infty(\mathbb{R}^d)$ 。例如取

$$\varphi(x) = c \exp\left(-\frac{1}{1-|x|^2}\right) 1_{\{|x| < 1\}},$$

其中 $c > 0$ 由归一化条件 $\int \varphi(x) dx = 1$ 确定。

命题 5.4.4 设 (φ_n) 是 δ_0 的一个近似。(i) 若 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 连续, 则 $\varphi_n * f \rightarrow f$ 在每个紧集上一致收敛。(ii) 若 $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, 其中 $p \in [1, \infty)$, 则 $\varphi_n * f \rightarrow f$ 于 L^p 。

证明。 第 (i) 部分很容易, 通过写出

$$\varphi_n * f(x) = \int_{|y| \leq \delta} f(x-y) \varphi_n(y) dy + \int_{|y| > \delta} f(x-y) \varphi_n(y) dy$$

并使用 f 的连续性。对于第 (ii) 部分, 我们注意到如果 $f, g \in L^p(\mathbb{R}^d, \lambda)$,

$$\begin{aligned} \int |\varphi_n * f(x) - \varphi_n * g(x)|^p dx &\leq \int \left(\int \varphi_n(x-y) |f(y) - g(y)| dy \right)^p dx \\ &\leq \int \left(\int \varphi_n(x-y) |f(y) - g(y)|^p dy \right) dx \\ &= \int |f(y) - g(y)|^p \left(\int \varphi_n(x-y) dx \right) dy \\ &= \int |f(y) - g(y)|^p dy. \end{aligned}$$

第二个不等式来自 Jensen 不等式：对固定的 x ， $\varphi_n(x-y)dy$ 是关于变量 y 的概率测度。这个估计使问题归结为 $f \in C_c(\mathbb{R}^d)$ 的情形；此时结论由 (i) 与支配收敛定理得到。

应用。在维数 $d = 1$ 时，我们可以取

$$\varphi_n(x) = c_n(1-x^2)^n 1_{\{|x| \leq 1\}}$$

其中常数 c_n 取为使 $\int \varphi_n(x) dx = 1$ 。现取闭区间 $[a, b] \subset (0, 1)$ ，并设 f 在 $[a, b]$ 上连续。可以把 f 延拓为 $\mathbb{R}$ 上的连续函数，使其紧支集包含在 $[0, 1]$ 中（例如令 f 在 $[0, a]$ 与 $[b, 1]$ 上为仿射函数）。于是

$$\varphi_n * f(x) = c_n \int (1-(x-y)^2)^n 1_{\{|x-y| \leq 1\}} f(y) dy \rightarrow f(x)$$

该收敛在 $[a, b]$ 上一致。对 $x \in [a, b]$ ，显然可以去掉指示函数 $1_{\{|x-y| \leq 1\}}$ 。因此 f 是 $[a, b]$ 上一列多项式的一致极限。这就给出了 Stone-Weierstrass 定理在本情形下的结论。

5.4.3 单位球体积的计算

记 B_d 为 $\mathbb{R}^d$ 的闭单位球， λ_d 为 $\mathbb{R}^d$ 上的 Lebesgue 测度，并令 $\gamma_d = \lambda_d(B_d)$ 。先注意，对任意 $a > 0$ ，映射 $x \mapsto ax$ 下 λ_d 的像测度为 $a^{-d}\lambda_d$ ，即对任意 Borel 集 A ，

$$\lambda_d(a^{-1}A) = a^{-d}\lambda_d(A)$$

这个等式只需对长方体检验。特别地，

$$\lambda_d(aB_d) = a^d\lambda_d(B_d).$$

当 $d \geq 2$ 时，由 Fubini 定理，

$$\begin{aligned} \gamma_d &= \int_{\mathbb{R}^d} 1_{\{x_1^2 + \dots + x_d^2 \leq 1\}} dx_1 \cdots dx_d \\ &= \int_{-1}^1 \lambda_{d-1}\left(\sqrt{1-x_d^2} B_{d-1}\right) dx_d \\ &= \gamma_{d-1} \int_{-1}^1 (1-x_d^2)^{(d-1)/2} dx_d = \gamma_{d-1} I_{d-1}, \end{aligned}$$

其中对整数 $n \geq 0$ ，定义

$$I_n = \int_{-1}^1 (1-x^2)^{n/2} dx.$$

分部积分给出：对 $n \geq 2$ ，

$$I_n = \frac{n}{n+1} I_{n-2}.$$

再由 $I_0 = 2$ 、 $I_1 = \pi/2$ ，归纳可得

$$I_{d-1} I_{d-2} = \frac{2\pi}{d}.$$

因此对 $d \geq 3$ ，

$$\gamma_d = I_{d-1} I_{d-2} \gamma_{d-2} = \frac{2\pi}{d} \gamma_{d-2}.$$

从 $\gamma_1 = 2$ 、 $\gamma_2 = \pi$ 出发，得到

$$\gamma_{2k} = \frac{\pi^k}{k!}, \quad \gamma_{2k+1} = \frac{\pi^k}{(k + \frac{1}{2})(k - \frac{1}{2}) \cdots \frac{3}{2} \cdot \frac{1}{2}}.$$

利用 Gamma 函数关系 $\Gamma(s+1) = s\Gamma(s)$ 及 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, 两式可统一写成

$$\gamma_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)}.$$

第六章 有符号测度

与前几章不同，本章研究可同时取正、负值的带符号测度。主要结果是 Jordan 分解：它把带符号测度以最小方式写成两个由互不相交可测集支撑的正测度之差。作为应用，将证明如下对偶定理：若 p, q 为共轭指数，即 $1/p + 1/q = 1$ ，且 $p < \infty$ ，则 L^p 的连续对偶可与 L^q 等距识别。

6.1 定义与全变差

定义 6.1.1 设 $(E, \mathcal{A})$ 为可测空间。映射 $\mu : \mathcal{A} \rightarrow \mathbb{R}$ 称为带符号测度，如果 $\mu(\emptyset) = 0$ ，并且对任意两两不交的可测集列 $(A_n)_{n \in \mathbb{N}}$ ，级数 $\sum_{n \in \mathbb{N}} \mu(A_n)$ 绝对收敛，且

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

定理 6.1.1 设 μ 是 $(E, \mathcal{A})$ 上的带符号测度。对每个 $A \in \mathcal{A}$ ，定义

$$|\mu|(A) = \sup \left\{ \sum_{n \in \mathbb{N}} |\mu(A_n)| : A = \bigcup_{n \in \mathbb{N}} A_n, (A_n) \text{ 两两不交且可测} \right\}.$$

则 $|\mu|$ 是 $(E, \mathcal{A})$ 上的有限正测度，并且

$$|\mu(A)| \leq |\mu|(A) \quad (A \in \mathcal{A}).$$

证明。先证 $|\mu|$ 是正测度。设 $(B_i)_{i \in \mathbb{N}}$ 为两两不交的可测集族，并令 $B = \bigcup_{i \in \mathbb{N}} B_i$ 。按定义，若 $t_i \in [0, |\mu|(B_i)]$ （当 $|\mu|(B_i) = 0$ 时取 $t_i = 0$ ），则可取一个可测划分

若 $|\mu|(B_i) = 0$ ，取 $t_i = 0$ ；否则可选 $t_i \in [0, |\mu|(B_i))$ 。由上确界的定义，可取 B_i 的可测划分 $B_i = \bigcup_{n \in \mathbb{N}} A_{n,i}$ ，使

$$\sum_{n \in \mathbb{N}} |\mu(A_{n,i})| \geq t_i.$$

于是 $(A_{n,i})_{n,i \in \mathbb{N}}$ 是 B 的可数分割，因此

$$|\mu|(B) \geq \sum_{i \in \mathbb{N}} \sum_{n \in \mathbb{N}} |\mu(A_{n,i})| \geq \sum_{i \in \mathbb{N}} t_i.$$

由于 t_i 可以任意接近 $|\mu|(B_i)$ ，因此得出

$$|\mu|(B) \geq \sum_{i \in \mathbb{N}} |\mu|(B_i).$$

为了得到相反的不等式，令 $(A_n)_{n \in \mathbb{N}}$ 是 B 的一个分割。那么

$$\begin{aligned}
\sum_{n \in \mathbb{N}} |\mu(A_n)| &= \sum_{n \in \mathbb{N}} |\sum_{i \in \mathbb{N}} \mu(A_n \cap B_i)| \\
&\leq \sum_{n \in \mathbb{N}} \sum_{i \in \mathbb{N}} |\mu(A_n \cap B_i)| \\
&= \sum_{i \in \mathbb{N}} \sum_{n \in \mathbb{N}} |\mu(A_n \cap B_i)| \\
&\leq \sum_{i \in \mathbb{N}} |\mu|(B_i),
\end{aligned}$$

最后一个不等式来自如下事实：对固定的 i ，集合族 $(A_n \cap B_i)_{n \in \mathbb{N}}$ 构成 B_i 的一个划分；再用 $|\mu|(B_i)$ 的定义即可。对 B 的所有划分 $(A_n)_{n \in \mathbb{N}}$ 取上确界，得到

$$|\mu|(B) \leq \sum_{i \in \mathbb{N}} |\mu|(B_i)$$

这完成了对 $|\mu|$ 是一个正测度的证明。由于不等式 $|\mu(A)| \leq |\mu|(A)$ 是显然的，剩下的就是证明 $|\mu|$ 是有限测度。

引理 6.1.2 若 $A \in \mathcal{A}$ 且 $|\mu|(A) = \infty$ ，则可把 A 分成两个不交的可测集 B, C ，使

$$A = B \cup C, \quad |\mu(B)| > 1, \quad |\mu|(C) = \infty.$$

证明。 由 $|\mu|(A) = \infty$ ，可取 A 的可测划分 $A = \bigcup_{n \in \mathbb{N}} A_n$ ，使

$$\sum_{n \in \mathbb{N}} |\mu(A_n)| > 2(1 + |\mu(A)|).$$

正项和与负项和中至少有一个大于 $1 + |\mu(A)|$ 。不妨设

$$\sum_{n \in \mathbb{N}} \mu(A_n)^+ > 1 + |\mu(A)|$$

(对称情况 $\sum_{n \in \mathbb{N}} \mu(A_n)^- > 1 + |\mu(A)|$ 的处理方式相同)。然后设定

$$B = \bigcup_{\{n: \mu(A_n) > 0\}} A_n$$

以便

$$\mu(B) = \sum_{n \in \mathbb{N}} \mu(A_n)^+ > 1 + |\mu(A)|.$$

此外，如果 $C = A \setminus B$,

$$|\mu(C)| = |\mu(A) - \mu(B)| \geq |\mu(B)| - |\mu(A)| > 1.$$

另一方面，由于 $A = B \cup C$ ，且 $|\mu|$ 是测度，必有 $|\mu|(B) = \infty$ 或 $|\mu|(C) = \infty$ 。必要时交换 B 与 C ，便得到引理的结论。**现在完成定理的证明。** 假设 $|\mu|(E) = \infty$ 。可取两两不交的可测集 B_0, C_0 ，使 $|\mu(B_0)| > 1$ 且 $|\mu|(C_0) = \infty$ 。再对 C_0 应用引理，可取两两不交的 B_1, C_1 ，满足 $C_0 = B_1 \cup C_1$ 、 $|\mu(B_1)| > 1$ 及 $|\mu|(C_1) = \infty$ 。递推下去，得到两两不交的可测集列 $(B_n)_{n \in \mathbb{N}}$ ，且对每个 n 都有 $|\mu(B_n)| > 1$ 。这与下列级数按带符号测度的定义必须绝对收敛相矛盾：

$$\sum_{n \in \mathbb{N}} \mu(B_n)$$

我们得出结论

$$|\mu|(E) < \infty.$$

例。设 ν 是 $(E, \mathcal{A})$ 上的正测度，且 $g \in L^1(E, \mathcal{A}, \nu)$ 。定义

$$\mu(A) = \int_A g d\nu, \quad A \in \mathcal{A}.$$

¹则 μ 是带符号测度。事实上，若 $A = \bigcup_{n \in \mathbb{N}} A_n$ 是不交并，则 $g1_A$ 是 $g1_{\bigcup_{n \leq k} A_n}$ 在 $L^1(\nu)$ 中的极限，故支配收敛定理给出

$$\mu(A) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

稍后将证明，在此情形

$$|\mu| = |g| \cdot \nu.$$

6.2 Jordan 分解

设 μ 是 $(E, \mathcal{A})$ 上的有限符号测度。容易验证

$$\mu^+ = \frac{1}{2}(\mu + |\mu|), \quad \mu^- = \frac{1}{2}(|\mu| - \mu)$$

定义了 $(E, \mathcal{A})$ 上的两个有限正测度，并且

$$\mu = \mu^+ - \mu^-, \quad |\mu| = \mu^+ + \mu^-.$$

定理 6.2.1 (Jordan 分解) 存在可测集 $B \subset E$ ，它在相差一个 $|\mu|$ -零测集的意义下唯一，并满足

$$\mu^+ = 1_B \cdot |\mu|, \quad \mu^- = 1_{B^c} \cdot |\mu|.$$

等价地， μ^+ 是 $|\mu|$ 在 B 上的限制， μ^- 是 $|\mu|$ 在 B^c 上的限制。并且对每个 $A \in \mathcal{A}$ ，

$$\mu^+(A) = \mu^+(A \cap B) = \mu(A \cap B), \quad \mu^-(A) = \mu^-(A \cap B^c) = -\mu(A \cap B^c),$$

从而

$$\mu(A) = \mu^+(A \cap B) - \mu^-(A \cap B^c), \quad |\mu|(A) = \mu^+(A \cap B) + \mu^-(A \cap B^c).$$

证明。显然 $\mu^+ \leq |\mu|$ 、 $\mu^- \leq |\mu|$ ，故 $\mu^+, \mu^- \ll |\mu|$ 。由 Radon-Nikodym 定理，存在有限值非负可测函数 h_1, h_2 ，使

$$\mu^+ = h_1 \cdot |\mu|, \quad \mu^- = h_2 \cdot |\mu|.$$

上述测度不等式还给出 $0 \leq h_1, h_2 \leq 1$ ， $|\mu|$ -几乎处处成立。令 $h = h_1 - h_2$ ，则对每个 $A \in \mathcal{A}$ ，

$$\mu(A) = \int_A (h_1 - h_2) d|\mu|.$$

下面证明 $|h_1 - h_2| = 1$ ， $|\mu|$ -几乎处处。固定 $r < 1$ ，令

$$E_r = \{x \in E : |h_1(x) - h_2(x)| \leq r\}.$$

¹校订注：原讲义该公式左端误排为 $\nu(A)$ ；这无法定义新的带符号测度，依上下文应为 $\mu(A)$ 。

若 $(A_n)_{n \in \mathbb{N}}$ 是 E_r 的任意可测分割, 则

$$\sum_n |\mu(A_n)| = \sum_n \left| \int_{A_n} (h_1 - h_2) d|\mu| \right| \leq \sum_n \int_{A_n} |h_1 - h_2| d|\mu| \leq r |\mu|(E_r).$$

对所有这样的分割取上确界, 由全变差的定义得到 $|\mu|(E_r) \leq r |\mu|(E_r)$, 故 $|\mu|(E_r) = 0$. 令 $r \uparrow 1$ 可知 $|h_1 - h_2| \geq 1$, 而反向不等式由 $0 \leq h_1, h_2 \leq 1$ 立即得到。

因此, $|\mu|$ -几乎处处, (h_1, h_2) 只能等于 $(1, 0)$ 或 $(0, 1)$ 。² 令 $B = \{x \in E : h_1(x) = 1\}$. 则 $h_1 = 1_B$ 、 $h_2 = 1_{B^c}$, $|\mu|$ -几乎处处, 因而

$$\mu^+ = 1_B \cdot |\mu|, \quad \mu^- = 1_{B^c} \cdot |\mu|.$$

B 的唯一性来自 Radon–Nikodym 密度的几乎处处唯一性; 定理中的其余等式随即得到。

注。 若 $\mu = \mu_1 - \mu_2$ 还是两个有限正测度之差, 则必有 $\mu_1 \geq \mu^+$ 、 $\mu_2 \geq \mu^-$ 。例如

$$\mu_1(A) \geq \mu_1(A \cap B) \geq \mu(A \cap B) = \mu^+(A \cap B) = \mu^+(A).$$

关于符号测度的积分。 若 $f \in L^1(E, \mathcal{A}, |\mu|)$, 定义

$$\int f d\mu := \int f d\mu^+ - \int f d\mu^- = \int f(1_B - 1_{B^c}) d|\mu|.$$

于是

$$\left| \int f d\mu \right| \leq \int |f| d|\mu|.$$

3

命题 6.2.2 设 ν 是 $(E, \mathcal{A})$ 上的正测度, $g \in L^1(E, \mathcal{A}, \nu)$, 并定义符号测度

$$\mu(A) = \int_A g d\nu.$$

则 $|\mu| = |g| \cdot \nu$ 。此外, 对每个 $f \in L^1(E, \mathcal{A}, |\mu|)$, 都有 $fg \in L^1(E, \mathcal{A}, \nu)$, 且

$$\int f d\mu = \int fg d\nu.$$

证明。 沿用定理 6.2.1 的记号, 令 $h = 1_B - 1_{B^c}$ 。对每个 $A \in \mathcal{A}$,

$$|\mu|(A) = \mu(A \cap B) - \mu(A \cap B^c) = \int_A gh d\nu.$$

在这个等式中取 $A = \{gh < 0\}$, 可知 $gh \geq 0$, ν -几乎处处。由于 $|h| = 1$, 故 $gh = |gh| = |g|$, ν -几乎处处, 从而

$$|\mu|(A) = \int_A |g| d\nu.$$

特别地,

$$\int |f| d|\mu| = \int |f| |g| d\nu,$$

所以 $f \in L^1(|\mu|)$ 蕴含 $fg \in L^1(\nu)$ 。

²校订注: 原讲义这里第二种情形误印为 $h_1(x) = 0$, $h_1(x) = 0$; 应为 $h_1(x) = 0$, $h_2(x) = 1$ 。

³校订注: 原讲义这条估计的右端排成了 $\int |f| d\mu$; 对符号测度而言正确写法是 $\int |f| d|\mu|$ 。

若 f 是简单函数, 则由定义直接有

$$\int f d\mu = \int fg d\nu.$$

一般地, 可取简单函数列 (f_n) , 使 $f_n \rightarrow f$ 且 $|f_n| \leq |f|$ 。分别对 μ^+, μ^- 和 ν 使用支配收敛定理, 便得到同一等式。

符号测度的 Radon–Nikodym 定理。 设 ν 是正测度, μ 是符号测度。若

$$\forall A \in \mathcal{A}, \quad \nu(A) = 0 \implies \mu(A) = 0,$$

则称 μ 关于 ν 绝对连续, 记为 $\mu \ll \nu$ 。

定理 6.2.3 设 μ 是符号测度, ν 是 σ -有限正测度。下列三条等价:

(i) $\mu \ll \nu$;

(ii) 对每个 $\varepsilon > 0$, 存在 $\delta > 0$, 使

$$\forall A \in \mathcal{A}, \quad \nu(A) \leq \delta \implies |\mu|(A) \leq \varepsilon;$$

(iii) 存在 $g \in L^1(E, \mathcal{A}, \nu)$, 使

$$\forall A \in \mathcal{A}, \quad \mu(A) = \int_A g d\nu.$$

证明。 (ii) $\Rightarrow$ (i) 显然。若 (i) 成立, 则 $\mu^+, \mu^- \ll \nu$ 。对两个正测度使用 Radon–Nikodym 定理, 可写成 $\mu^+ = g_1 \cdot \nu$ 、 $\mu^- = g_2 \cdot \nu$, 其中 $g_1, g_2 \geq 0$, 并且

$$\int g_1 d\nu = \mu^+(E) < \infty, \quad \int g_2 d\nu = \mu^-(E) < \infty.$$

取 $g = g_1 - g_2$ 即得 (iii)。最后设 (iii) 成立。由命题 6.2.2, $|\mu| = |g| \cdot \nu$ 。支配收敛定理给出

$$\lim_{N \rightarrow \infty} \int_{\{|g| \geq N\}} |g| d\nu = 0.$$

给定 $\varepsilon > 0$, 取充分大的 N , 使上式中的积分小于 $\varepsilon/2$, 再令 $\delta = \varepsilon/(2N)$ 。若 $\nu(A) \leq \delta$, 则

$$|\mu|(A) = \int_A |g| d\nu \leq \int_{\{|g| \geq N\}} |g| d\nu + \int_{A \cap \{|g| < N\}} |g| d\nu \leq \frac{\varepsilon}{2} + N\delta = \varepsilon.$$

所以 (ii) 成立。

6.3 L^p – L^q 对偶

设 ν 是 $(E, \mathcal{A})$ 上的正测度, $p \in [1, \infty]$, q 是 p 的共轭指数。固定 $g \in L^q(E, \mathcal{A}, \nu)$ 后,

$$\Phi_g(f) = \int fg d\nu$$

定义了 $L^p(E, \mathcal{A}, \nu)$ 上的连续线性泛函。Hölder 不等式给出

$$|\Phi_g(f)| \leq \|g\|_q \|f\|_p, \quad \|\Phi_g\| := \sup_{\|f\|_p \leq 1} |\Phi_g(f)| \leq \|g\|_q.$$

自然的问题是: L^p 上的每个连续线性泛函是否都这样得到? 当 $p = q = 2$ 时, Hilbert 空间理论已经给出肯定答案; 下述定理处理 $p < \infty$ 的情形。

定理 6.3.1 (L^p - L^q 对偶) 设 ν 是 σ -有限测度, $p \in [1, \infty)$, q 是其共轭指数。若 Φ 是 $L^p(E, \mathcal{A}, \nu)$ 上的连续线性泛函, 则存在唯一的 $g \in L^q(E, \mathcal{A}, \nu)$, 使

$$\Phi(f) = \int f g d\nu \quad (f \in L^p).$$

而且 $\|\Phi\| = \|g\|_q$ 。因此映射 $g \mapsto \Phi_g$ 把 $L^q(\nu)$ 等距同构到 $L^p(\nu)$ 的连续对偶空间。 $p = \infty$ 时这一结论一般不成立。

证明。 先设 $\nu(E) < \infty$ 。对 $A \in \mathcal{A}$, 令

$$\mu(A) = \Phi(1_A).$$

因为 $1_A \in L^p(\nu)$, 定义有意义。下面验证 μ 是符号测度。若 (A_n) 两两不交, $A = \bigcup_n A_n$, 则

$$1_A = \lim_{k \rightarrow \infty} \sum_{n \leq k} 1_{A_n} \quad \text{于 } L^p(\nu).$$

这里可使用支配收敛定理, 因为 $\nu(E) < \infty$, 故 $1 \in L^p(\nu)$ 。由 Φ 的连续性,

$$\mu(A) = \lim_{k \rightarrow \infty} \sum_{n \leq k} \mu(A_n).$$

还需说明级数 $\sum_n \mu(A_n)$ 绝对收敛。令 $A'_n = A_n$ (若 $\mu(A_n) > 0$), 否则令 $A'_n = \emptyset$, 并设 $A' = \bigcup_n A'_n$ 。则

$$\sum_n \mu(A_n)^+ = \sum_n \mu(A'_n) = \lim_{k \rightarrow \infty} \sum_{n \leq k} \mu(A'_n) = \mu(A') < \infty.$$

对负项同理。因此 μ 是有限符号测度。若 $\nu(A) = 0$, 则 $1_A = 0$ 于 $L^p(\nu)$, 从而 $\mu(A) = 0$ 。所以 $\mu \ll \nu$ 。由定理 6.2.3, 存在 $g \in L^1(\nu)$, 使

$$\Phi(1_A) = \mu(A) = \int_A g d\nu \quad (A \in \mathcal{A}).$$

利用线性性, 这一等式先对简单函数 f 成立; 又因有限测度空间上的有界可测函数可由简单函数一致逼近, 所以它对所有有界可测函数也成立。

下面证明 $g \in L^q(\nu)$ 。若 $p = 1$, 对任意 $A \in \mathcal{A}$,

$$\left| \int_A g d\nu \right| = |\Phi(1_A)| \leq \|\Phi\| \|1_A\|_1 = \|\Phi\| \nu(A).$$

分别取 $A = \{g > \|\Phi\| + \varepsilon\}$ 与 $A = \{g < -\|\Phi\| - \varepsilon\}$, 可知 $|g| \leq \|\Phi\|$, ν -几乎处处。因此 $\|g\|_\infty \leq \|\Phi\|$ 。

若 $1 < p < \infty$, 令

$$E_n = \{x : |g(x)| \leq n\}, \quad f_n = 1_{E_n} |g|^{q-1} \operatorname{sgn}(g).$$

f_n 有界, 所以

$$\int_{E_n} |g|^q d\nu = \int f_n g d\nu = \Phi(f_n) \leq \|\Phi\| \|f_n\|_p = \|\Phi\| \left(\int_{E_n} |g|^q d\nu \right)^{1/p}.$$

因此

$$\int_{E_n} |g|^q d\nu \leq \|\Phi\|^q.$$

令 $n \rightarrow \infty$ ，由单调收敛定理得到 $g \in L^q(\nu)$ 且 $\|g\|_q \leq \|\Phi\|$ 。

在两种情形下都已得到 $g \in L^q(\nu)$ 。映射 $f \mapsto \Phi(f)$ 与 $f \mapsto \int fg d\nu$ 都是 $L^p(\nu)$ 上的连续线性泛函，并且在有界可测函数这个稠密子集上一致，故在整个 $L^p(\nu)$ 上一致。另一方面，Hölder 不等式给出 $\|\Phi\| \leq \|g\|_q$ ，结合前述反向不等式可得 $\|\Phi\| = \|g\|_q$ 。表示函数 g 的唯一性也随之成立。

现在只假设 ν 为 σ -有限。把 E 写成两两不交可测集的并

$$E = \bigcup_{n \in \mathbb{N}} E_n, \quad \nu(E_n) < \infty.$$

记 $\nu_n = \nu|_{E_n}$ 。映射 $f \mapsto f1_{E_n}$ 把 $L^p(\nu_n)$ 等距嵌入 $L^p(\nu)$ 。对

$$\Phi_n(f) = \Phi(f1_{E_n})$$

应用有限测度情形，得到 $g_n \in L^q(\nu_n)$ ，使

$$\Phi(f1_{E_n}) = \int f g_n d\nu.$$

把 g_n 在 E_n^c 上延拓为 0。若 $f \in L^p(\nu)$ ，则

$$f = \lim_{k \rightarrow \infty} \sum_{n \leq k} f1_{E_n} \quad \text{于 } L^p(\nu),$$

所以

$$\Phi(f) = \lim_{k \rightarrow \infty} \int f \left(\sum_{n \leq k} g_n \right) d\nu.$$

又由

$$\left| \int f \left(\sum_{n \leq k} g_n \right) d\nu \right| = \left| \Phi \left(f \sum_{n \leq k} 1_{E_n} \right) \right| \leq \|\Phi\| \|f\|_p,$$

使用有限测度情形中相同的检验函数，可得

$$\left\| \sum_{n \leq k} g_n \right\|_q \leq \|\Phi\| \quad (k \geq 0).$$

令

$$g(x) = \sum_{n \in \mathbb{N}} g_n(x).$$

由于 E_n 两两不交，每个 x 至多对应一个非零项。若 $q = \infty$ ，前一估计直接给出 $\|g\|_\infty \leq \|\Phi\|$ ；若 $q < \infty$ ，则

$$\int |g|^q d\nu = \sum_n \int |g_n|^q d\nu = \lim_{k \rightarrow \infty} \int \left| \sum_{n \leq k} g_n \right|^q d\nu \leq \|\Phi\|^q.$$

所以总有 $g \in L^q(\nu)$ 。而 $\sum_{n \leq k} g_n \rightarrow g$ 逐点，且 $\left| \sum_{n \leq k} g_n \right| \leq |g|$ 。Hölder 不等式说明 $|fg| \in L^1$ ，故支配收敛定理给出

$$\Phi(f) = \lim_{k \rightarrow \infty} \int f \left(\sum_{n \leq k} g_n \right) d\nu = \int fg d\nu.$$

范数等式与唯一性同有限测度情形，定理得证。

注。 $p = \infty$ 时, 定理一般不成立。以 ℓ^∞ 为例, 它是所有有界实数列 $a = (a_k)_{k \in \mathbb{N}}$ 组成的 Banach 空间, 范数为 $\|a\|_\infty = \sup_k |a_k|$ 。令

$$H = \{a \in \ell^\infty : \lim_{k \rightarrow \infty} a_k \text{ 存在}\}, \quad \Phi(a) = \lim_{k \rightarrow \infty} a_k \quad (a \in H).$$

显然 $|\Phi(a)| \leq \|a\|_\infty$ 。Hahn-Banach 定理把 Φ 延拓为 ℓ^∞ 上的连续线性泛函, 并保持该估计。它不可能写成

$$\Phi(a) = \sum_{k \in \mathbb{N}} a_k b_k \quad (b \in \ell^1).$$

否则, 对第 n 个单位向量 $e^{(n)}$, 有 $b_n = \Phi(e^{(n)}) = 0$ 对所有 n 成立, 于是 $b = 0$; 但常数列 $\mathbf{1}$ 满足 $\Phi(\mathbf{1}) = 1$, 矛盾。

6.4 里斯表示定理

本节始终假设 E 是局部紧致可分度量空间。记 $C_0(E)$ 为在无穷远处趋于 0 的连续函数空间: $f \in C_0(E)$ 当且仅当 f 连续, 并且对每个 $\varepsilon > 0$, 存在紧集 $K \subset E$, 使

$$|f(x)| < \varepsilon \quad (x \in E \setminus K).$$

配备上确界范数

$$\|f\|_\infty = \sup_{x \in E} |f(x)|,$$

$C_0(E)$ 是 Banach 空间。若 μ 是 $(E, \mathcal{B}(E))$ 上的有限符号测度, 则

$$\Phi_\mu(f) = \int_E f d\mu \quad (f \in C_0(E))$$

定义了连续线性泛函, 而且

$$|\Phi_\mu(f)| \leq \int_E |f| d|\mu| \leq |\mu|(E) \|f\|_\infty.$$

因此 $\|\Phi_\mu\| \leq |\mu|(E)$ 。

定理 6.4.1 (Riesz 表示定理) 若 Φ 是 $C_0(E)$ 上的连续线性泛函, 则存在唯一的有限符号测度 μ 使

$$\Phi(f) = \int_E f d\mu \quad (f \in C_0(E)).$$

本讲义不证明这一结果; 可参见 Rudin [7] 第 6 章, 其中还处理了更一般的复数情形。

注。记 $\mathcal{M}(E)$ 为 E 上所有有限符号测度组成的向量空间。映射 $\mu \mapsto |\mu|(E)$ 是其上的范数, 且 $\mathcal{M}(E)$ 在这个范数下完备。Riesz 表示定理可以改述为: $\mathcal{M}(E)$ 与 $C_0(E)$ 的连续对偶空间等距同构。若 E 紧致, 则 $C_0(E) = C_b(E)$, 所以 $\mathcal{M}(E) = C_b(E)^*$ 。当 E 非紧时, 这个结论对 $C_b(E)$ 一般不成立; 例如 $E = \mathbb{R}$ 时, 存在不能由有限符号测度表示的 $C_b(E)$ 上连续线性泛函。可把上一节末的 ℓ^∞ 例子作适当修改来构造。

第七章 变量替换公式与补充内容

变量替换公式描述 Lebesgue 测度在 $\mathbb{R}^d$ 的开集之间经微分同胚变换后的像测度。继 Fubini 定理之后，它是计算多重积分的第二个基本工具。本章还将推出 $\mathbb{R}^d$ 中的极坐标积分公式，并由此引入单位球面上的自然表面测度。

7.1 变量替换公式

我们首先处理一个重要的特殊情形——仿射映射。

命题 7.1.1 设 $b \in \mathbb{R}^d$, M 为可逆的 $d \times d$ 实矩阵, 并令 $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 为 $f(x) = Mx + b$ 。则对任意 Borel 集 $A \subset \mathbb{R}^d$,

$$\lambda(f(A)) = |\det M| \lambda(A).$$

注意。若 M 不可逆, 则 $f(A) \subset f(\mathbb{R}^d)$, 而后者包含在某个低维仿射子空间中, 其 d 维 Lebesgue 测度为零 (练习)。

证明。若 A 为 Borel 集, 则 $f(A) = (f^{-1})^{-1}(A)$ 仍为 Borel 集。由 Lebesgue 测度的平移不变性, 只需考虑 $b = 0$ 。此时, 对任意 $a \in \mathbb{R}^d$ 与 Borel 集 A ,

$$\lambda(f(a + A)) = \lambda(f(a) + f(A)) = \lambda(f(A)).$$

因此, $A \mapsto \lambda(f(A))$ 是平移不变测度; 由 Lebesgue 测度的唯一性, 存在常数 $c \geq 0$, 使

$$\lambda(f(A)) = c\lambda(A).$$

下面证明 $c = |\det M|$ 。若 M 正交, 设 B_d 为闭单位球, 则 $f(B_d) = B_d$, 故 $c = 1 = |\det M|$ 。若 M 为正定对称矩阵, 则存在正交矩阵 P , 使

$$P^T M P = \text{diag}(\alpha_1, \dots, \alpha_d), \quad \alpha_i > 0.$$

于是

$$f(P([0, 1]^d)) = \left\{ Py : y \in \prod_{i=1}^d [0, \alpha_i] \right\}.$$

利用正交情形和 $\lambda(P([0, 1]^d)) = 1$, 得到

$$c = \lambda(f(P([0, 1]^d))) = \lambda\left(\prod_{i=1}^d [0, \alpha_i]\right) = \prod_{i=1}^d \alpha_i = \det M.$$

最后, 对一般可逆矩阵使用极分解

$$M = PS, \quad S = (M^T M)^{1/2}, \quad P = MS^{-1}.$$

这里 P 正交、 S 正定对称。两个特殊情形合并后给出 $c = |\det P| |\det S| = |\det M|$ 。

设 U, D 是 $\mathbb{R}^d$ 中的开集。若映射 $\varphi: U \rightarrow D$ 双射, 且 φ 与 φ^{-1} 都属于 C^1 , 则称 φ 为 C^1 微分同胚。此时对每个 $u \in U$, 导数矩阵 $\varphi'(u)$ 都可逆。

定理 7.1.2(变量替换公式) 设 $\varphi: U \rightarrow D$ 是 C^1 微分同胚。则对任意 Borel 函数 $f: D \rightarrow \mathbb{R}_+$,

$$\int_D f(x) dx = \int_U f(\varphi(u)) |J_\varphi(u)| du,$$

其中 $J_\varphi(u) = \det(\varphi'(u))$ 是 φ 在 u 处的雅可比行列式。

证明。 通过非负简单函数逼近和单调收敛定理, 只需证明 $f = 1_A$ 的情形, 其中 $A \in \mathcal{B}(D)$ 。此时所需等式为

$$\lambda(A) = \int_{\varphi^{-1}(A)} |J_\varphi(u)| du.$$

把 $\varphi^{-1}(A)$ 重新记为 A , 等价地只需证明: 对每个 $A \in \mathcal{B}(U)$,

$$\lambda(\varphi(A)) = \int_A |J_\varphi(u)| du. \quad (7.1)$$

注意 $\varphi(A) = (\varphi^{-1})^{-1}(A)$ 确为 Borel 集。

引理 7.1.3 设 $K \subset U$ 为紧集, $\varepsilon > 0$ 。则可取充分小的 $\delta > 0$, 使对每个以 $u_0 \in K$ 为中心、各面平行于坐标超平面且边长小于 δ 的方体 C , 都有

$$(1 - \varepsilon) |J_\varphi(u_0)| \lambda(C) \leq \lambda(\varphi(C)) \leq (1 + \varepsilon) |J_\varphi(u_0)| \lambda(C).$$

引理证明。 由 φ' 的连续性, 可取充分小的 $\delta > 0$, 使

$$\delta < \frac{1}{d} \text{dist}(K, U^c),$$

并且当 $u_0 \in K$ 、 $|u - u_0| < d\delta$ 时,

$$|\varphi(u) - \varphi(u_0) - \varphi'(u_0)(u - u_0)| \leq \eta |u - u_0|,$$

其中辅助参数 $\eta > 0$ 稍后按给定的 ε 选取。令

$$F(v) = \varphi(u_0) + \varphi'(u_0)v.$$

则

$$\varphi(u) = F(u - u_0 + h(u, u_0)),$$

其中

$$h(u, u_0) = \varphi'(u_0)^{-1}(\varphi(u) - \varphi(u_0) - \varphi'(u_0)(u - u_0)).$$

紧致性给出 $a := \sup_{v \in K} \|\varphi'(v)^{-1}\| < \infty$, 故 $|h(u, u_0)| \leq a\eta |u - u_0|$ 。设 C 是以 u_0 为中心、边长 $r \leq \delta$ 的方体, $\tilde{C} = C - u_0$ 。于是

$$\varphi(C) \subset F((1 + da\eta)\tilde{C}).$$

由命题 7.1.1,

$$\lambda(\varphi(C)) \leq (1 + da\eta)^d |J_\varphi(u_0)| \lambda(C).$$

对逆映射作同样估计可得相应下界; 等价地, 对适当常数 $c' > 0$,

$$F((1 - c'\eta)\tilde{C}) \subset \varphi(C).$$

最后把 η 取得足够小, 使两个体积因子分别不超过 $1 + \varepsilon$ 且不小于 $1 - \varepsilon$, 即得引理。¹

回到定理的证明。对整数 $n \geq 1$, 称形如

$$C = \prod_{j=1}^d (k_j 2^{-n}, (k_j + 1) 2^{-n}], \quad k_j \in \mathbb{Z},$$

的方体为 n 阶基本方体, 并以 $\mathcal{C}_n$ 表示其全体。固定一个 n_0 阶基本方体 C_0 , 使 $\overline{C_0} \subset U$, 并取 $\varepsilon > 0$ 。选充分大的 $n \geq n_0$, 使引理对 $K = \overline{C_0}$ 、 $\delta = 2^{-n}$ 成立, 且当 $u, v \in K$ 、 $|u - v| \leq d\delta$ 时,

$$(1 - \varepsilon)|J_\varphi(u)| \leq |J_\varphi(v)| \leq (1 + \varepsilon)|J_\varphi(u)|. \quad (7.2)$$

那么, 记 x_C 为立方体 C 的中心

$$\begin{aligned} \lambda(\varphi(C_0)) &= \sum_{\substack{C \in \mathcal{C}_n \\ C \subset C_0}} \lambda(\varphi(C)) \\ &\leq (1 + \varepsilon) \sum_{\substack{C \in \mathcal{C}_n \\ C \subset C_0}} |J_\varphi(x_C)| \lambda(C) \\ &\leq (1 + \varepsilon)^2 \sum_{\substack{C \in \mathcal{C}_n \\ C \subset C_0}} \int_C |J_\varphi(u)| du \\ &= (1 + \varepsilon)^2 \int_{C_0} |J_\varphi(u)| du. \end{aligned}$$

在第一个不等式中我们使用了引理, 在第二个不等式中使用了 (7.2)。同样可以得到下界

$$\lambda(\varphi(C_0)) \geq (1 - \varepsilon)^2 \int_{C_0} |J_\varphi(u)| du.$$

由于 ε 是任意的, 我们可以得出结论

$$\lambda(\varphi(C_0)) = \int_{C_0} |J_\varphi(u)| du.$$

因此, 当 A 是闭包包含在 U 中的基本方体时, 式 (7.1) 已经成立。²一般情形由单调类论证得到。记 μ 为 D 上 Lebesgue 测度在 φ^{-1} 下的像测度:

$$\mu(A) = \lambda(\varphi(A))$$

对于 U 中的任意 Borel 集 A 。也设

$$\tilde{\mu}(A) = \int_A |J_\varphi(u)| du.$$

由此已经对闭包包含在 U 中的每个基本立方体 C 证明了 $\mu(C) = \tilde{\mu}(C)$ 。另一方面, 令 U_n 为闭包包含在 $U \cap \{u : |u| \leq n\}$ 中的 n 阶基本立方体之不相交并。则 $U_n \uparrow U$, 且对每个 n , $\mu(U_n) = \tilde{\mu}(U_n) < \infty$ 。闭包包含在 U 中的基本立方体类对有限交封闭并生成 $\mathcal{B}(U)$ 。应用第 1 章最后的推论, 得到 $\mu = \tilde{\mu}$, 即为所需结论。

¹学习提示: 证明中的 η 是辅助误差, 而定理给定的是 ε 。原讲义沿用同一符号; 这里分开记号, 以明确最后需要先缩小辅助误差。

²校订注: 原讲义此处把环境开集 U 误排成了待处理的 Borel 集 A ; 基本方体的闭包应包含在微分同胚的定义域 U 中。

极坐标积分的应用。取 $d = 2$ 、 $U = (0, \infty) \times (-\pi, \pi)$ ，以及 $D = \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\}$ 。于是映射

$$\varphi(r, \theta) = (r \cos \theta, r \sin \theta), \quad (r, \theta) \in U$$

它是从 U 到 D 的 C^1 微分同胚，并且

$$\varphi'(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}, \quad J_\varphi(r, \theta) = r.$$

因此，对任意 Borel 函数 $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ ，变量替换公式给出

$$\begin{aligned} \int_D f(x, y) dx dy &= \int_U f(r \cos \theta, r \sin \theta) r dr d\theta \\ &= \int_0^\infty \int_{-\pi}^\pi f(r \cos \theta, r \sin \theta) r d\theta dr. \end{aligned}$$

负半轴在 $\mathbb{R}^2$ 中的 Lebesgue 测度为零，故也有

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_0^\infty \int_{-\pi}^\pi f(r \cos \theta, r \sin \theta) r d\theta dr.$$

例。取 $f(x, y) = \exp(-x^2 - y^2)$ 。一方面，Fubini–Tonelli 定理给出

$$\int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \left(\int_{-\infty}^\infty e^{-x^2} dx \right)^2.$$

另一方面，极坐标公式给出

$$\int_0^\infty \int_{-\pi}^\pi e^{-r^2} r d\theta dr = 2\pi \int_0^\infty e^{-r^2} r dr = \pi.$$

由于积分为正，得到经典的 Gauss 积分

$$\boxed{\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}}.$$

7.2 单位球面上的 Lebesgue 测度

本节记 λ_d 为 $\mathbb{R}^d$ 上的 Lebesgue 测度，并记

$$S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$$

为单位球面。若 $A \in \mathcal{B}(S^{d-1})$ ，定义其锥体

$$\Gamma(A) = \{rx : r \in [0, 1], x \in A\}.$$

定理 7.2.1 对每个 $A \in \mathcal{B}(S^{d-1})$ ，令

$$\omega_d(A) = d \lambda_d(\Gamma(A)).$$

则 ω_d 是 S^{d-1} 上的有限正测度, 并且在线性等距变换下不变。此外, 对任意 Borel 函数 $f: \mathbb{R}^d \rightarrow \mathbb{R}_+$,

$$\int_{\mathbb{R}^d} f(x) dx = \int_0^\infty \int_{S^{d-1}} f(rz) r^{d-1} \omega_d(dz) dr. \quad (7.3)$$

最后, ω_d 的总质量 (即单位球面的表面积) 为

$$\omega_d(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

注意。 还可以证明: S^{d-1} 上任何在线性等距变换下不变的有限测度, 都与 ω_d 成比例。

证明。 ω_d 显然是 S^{d-1} 上的有限正测度; 忽略零点这个 Lebesgue 零测集后, 它就是测度 $d\lambda_d|_{B_d}$ 在映射 $x \mapsto x/|x|$ 下的像测度。由 λ_d 在线性等距变换下不变, 可知 ω_d 也具有同样的不变性。具体地, 若 φ 是线性等距变换, 则

$$\lambda_d(\Gamma(\varphi^{-1}(A))) = \lambda_d(\varphi^{-1}(\Gamma(A))) = \lambda_d(\Gamma(A)).$$

其总质量为

$$\omega_d(S^{d-1}) = d\lambda_d(B_d) = d \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

下面证明式 (7.3)。只需处理 $f = 1_B$ 的情形, 其中 $B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ 。公式

$$\mu(B) = \int_0^\infty \int_{S^{d-1}} 1_B(rz) r^{d-1} \omega_d(dz) dr$$

在 $\mathbb{R}^d \setminus \{0\}$ 上定义了一个测度 μ ; 要证 $\mu = \lambda_d$ 。先考虑

$$B = \left\{ x \in \mathbb{R}^d \setminus \{0\} : a < |x| \leq b, \frac{x}{|x|} \in A \right\},$$

其中 $A \in \mathcal{B}(S^{d-1})$, 且 $0 < a \leq b$ 。于是

$$\mu(B) = \omega_d(A) \int_a^b r^{d-1} dr = \frac{b^d - a^d}{d} \omega_d(A).$$

为计算 $\lambda_d(B)$, 令 $\alpha = a/b \in (0, 1)$, 并对 $n \geq 0$ 定义

$$\Gamma_n(A) = \{y = rx : \alpha^{n+1} < r \leq \alpha^n, x \in A\}.$$

则 $\lambda_d(\Gamma_n(A)) = \alpha^{nd} \lambda_d(\Gamma_0(A))$, 且

$$\lambda_d(\Gamma(A)) = \sum_{n=0}^{\infty} \lambda_d(\Gamma_n(A)).$$

因此

$$\lambda_d(\Gamma_0(A)) = (1 - \alpha^d) \lambda_d(\Gamma(A)) = \frac{1 - \alpha^d}{d} \omega_d(A).$$

又因 $B = b\Gamma_0(A)$,

$$\lambda_d(B) = b^d \lambda_d(\Gamma_0(A)) = \frac{b^d - a^d}{d} \omega_d(A) = \mu(B).$$

最后, 上述形状的集合类对有限交封闭, 并生成 $\mathbb{R}^d \setminus \{0\}$ 上的 Borel σ -代数。由通常的单调类论证, $\mu = \lambda_d$, 式 (7.3) 得证。

若 $f: \mathbb{R}^d \rightarrow \mathbb{R}_+$ 是径向函数, 即存在 $\tilde{f}: [0, \infty) \rightarrow \mathbb{R}_+$ 使 $f(x) = \tilde{f}(|x|)$, 则定理给出

$$\int_{\mathbb{R}^d} f(x) dx = c_d \int_0^\infty \tilde{f}(r) r^{d-1} dr, \quad c_d = \omega_d(S^{d-1}).$$

3

概率

³学习提示: 原讲义为简洁起见把径向函数及其一维轮廓都记成 f 。这里改记为 $\tilde{f}$, 以避免定义域混淆。

第二部分

概率

第八章 概率论基础

本章介绍概率论的基本概念：随机变量、数学期望、分布、矩和特征函数等。概率空间本质上是总质量为 1 的测度空间，因此许多概念都可由前面的积分论直接得到：随机变量是可测函数，数学期望是关于概率测度的积分。不过，概率论关注的是随机试验及其可观测量。分布作为像测度，因而成为核心对象：它给出随机变量落入任一可测集合的概率。

8.1 一般定义

8.1.1 概率空间

设 $(\Omega, \mathcal{A})$ 为可测空间， P 为其上的概率测度。称 $(\Omega, \mathcal{A}, P)$ 为概率空间。概率空间是总质量为 1 的特殊测度空间；但概率论的观点不同于一般积分论：我们要用它为“随机试验”建立数学模型。

- Ω 是所有可能结果组成的集合；其中一个元素 $\omega \in \Omega$ 表示随机试验的一种确定结果。
- $\mathcal{A}$ 是事件组成的 σ -代数。事件 $A \in \mathcal{A}$ 是 Ω 的一个子集，表示满足某项性质的所有试验结果。

对 $A \in \mathcal{A}$ ， $P(A)$ 表示事件 A 发生的概率。早期概率论著作在测度论建立之前，常用频率来说明概率：设同一随机试验重复 N 次， N_A 表示事件 A 出现的次数，则当 $N \rightarrow \infty$ 时，比例 N_A/N 收敛到 $P(A)$ 。后文将说明这种历史定义与现代公理化方法之间的联系。

例子。(1) 掷两次骰子：

$$\Omega = \{1, 2, \dots, 6\}^2, \quad \mathcal{A} = \mathcal{P}(\Omega), \quad P(A) = \frac{\text{Card}(A)}{36}.$$

这里的概率测度体现了一个建模假设：两次掷骰的 36 种有序结果等可能。(2) 不断掷一枚骰子，直到首次出现 6。由于所需次数没有先验上界，自然的做法是把一次试验结果看成无限序列，并取

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}^*}$$

于是 $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ 记录各次掷骰结果。在 Ω 上取乘积 σ -代数 $\mathcal{A}$ ，即使所有下列柱集可测的最小 σ -代数：

$$\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}$$

其中 $n \geq 1$ ，且 $i_1, \dots, i_n \in \{1, \dots, 6\}$ 。这个 σ -代数也等于乘积拓扑对应的 Borel σ -代数。最后， P 是 Ω 上唯一满足下式的概率测度：

$$P(\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}) = \left(\frac{1}{6}\right)^n.$$

P 的唯一性是单调类引理的直接推论，存在性则是无限乘积测度构造的一个特例。也可从 $[0, 1]$ 上的 Lebesgue 测度直接构造 P ：对几乎每个 $x \in [0, 1]$ ，存在唯一序列 $(\varepsilon_k)_{k \geq 1} \in \Omega$ ，使

$$x = \sum_{k=1}^{\infty} (\varepsilon_k - 1)6^{-k}.$$

于是 P 就是 Lebesgue 测度在映射 $x \mapsto (\varepsilon_k)_{k \geq 1}$ 下的像测度。

(3) 考虑受随机扰动的质点在空间中的运动。若时间限制在 $[0, 1]$ ，自然的样本空间是 $\Omega = C([0, 1], \mathbb{R}^3)$ ：每个 $\omega \in \Omega$ 都是一条可能的连续轨道。取使所有坐标映射 $\omega \mapsto \omega(t)$ ($t \in [0, 1]$) 可测的最小 σ -代数；它恰好是由一致收敛拓扑生成的 Borel σ -代数。还需在其上选取概率测度，而最重要的例子是

Wiener 测度；它就是布朗运动的分布。

重要提示。在接下来的内容中，我们经常不会明确指定概率空间的选择。重要的信息将是定义在该空间上的函数的性质，即随机变量。

8.1.2 随机变量

定义 8.1.1 设 $(E, \mathcal{E})$ 为可测空间。可测映射 $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ 称为取值于 E 的随机变量。

例。回到上面的三个例子：(1) $X((i, j)) = i + j$ 定义了一个取值于 $\{1, 2, \dots, 12\}$ 的随机变量。(2) 令 $X(\omega) = \inf\{j : \omega_j = 6\}$ ，约定 $\inf \emptyset = \infty$ 。它是取值于 $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ 的随机变量。为验证可测性，注意对每个 $k \geq 1$ ，

$$X^{-1}(\{k\}) = \{\omega \in \Omega : \omega_1 \neq 6, \omega_2 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\}.$$

(3) 对固定的 $t \in [0, 1]$ ，映射 $X(\omega) = \omega(t)$ 是取值于 $\mathbb{R}^3$ 的随机变量。(这里尚未构造 P ，但随机变量的可测性只依赖 $(\Omega, \mathcal{A})$ ，与 P 无关。)

定义 8.1.2 随机变量 $X : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{E})$ 的分布，是 P 在 X 下的像测度，记为 P_X ：

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{E}.$$

在实际操作中更常写作：

$$P_X(B) = P(X \in B) (= P(\{\omega \in \Omega : X(\omega) \in B\})).$$

分布 P_X 用来计算由随机变量 X 决定的事件的概率。对每个 $\omega \in \Omega$ ， $X(\omega)$ 是 E 中的一个“随机点”； $P_X(B) = P(X \in B)$ 就是该随机点落入 B 的概率。

注。给定任意可测空间 $(E, \mathcal{E})$ 上的概率测度 μ ，都可典范地构造一个分布为 μ 的随机变量：取 $\Omega = E$ 、 $\mathcal{A} = \mathcal{E}$ 、 $P = \mu$ ，并令 $X(\omega) = \omega$ 。显然 $P_X = \mu$ 。

特殊情况。

• **离散随机变量。**若 E 可数且 $\mathcal{E} = \mathcal{P}(E)$ ，则

$$P_X = \sum_{x \in E} p_x \delta_x, \quad p_x = P(X = x).$$

其中 $p_x = P(X = x)$ ， δ_x 表示点 x 处的 Dirac 测度。事实上，

$$P_X(B) = P(X \in B) = P\left(\bigcup_{x \in B} \{X = x\}\right) = \sum_{x \in B} P(X = x) = \sum_{x \in E} p_x \delta_x(B).$$

在实际计算中，求离散随机变量的分布，就是求出每个 $x \in E$ 的概率 $P(X = x)$ 。**例。**回到前面的例 (2)，令 $X(\omega) = \inf\{j : \omega_j = 6\}$ 。则对每个 $k \geq 1$ ，

$$P(X = k) = P\left(\bigcup_{i_1, \dots, i_{k-1} \neq 6} \{\omega_1 = i_1, \dots, \omega_{k-1} = i_{k-1}, \omega_k = 6\}\right) = 5^{k-1} \left(\frac{1}{6}\right)^k = \frac{1}{6} \left(\frac{5}{6}\right)^{k-1}.$$

注意, $\sum_{k=1}^{\infty} P(X = k) = 1$, 所以 $P(X = \infty) = 1 - P(X \in \mathbb{N}) = 0$ 。但集合 $\{X = \infty\}$ 绝非空集; 例如, 任何从不取值 6 的序列 $(i_1, i_2, \dots)$ 都属于该集合。

有密度的随机变量。取值于 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 的随机变量 X , 若其分布 P_X 关于 Lebesgue 测度 λ 绝对连续, 就称 X 有密度。由 Radon–Nikodym 定理, 存在 Borel 函数 $p: \mathbb{R}^d \rightarrow \mathbb{R}_+$, 使

$$P_X(B) = \int_B p(x) dx.$$

特别地, $\int_{\mathbb{R}^d} p(x) dx = P(X \in \mathbb{R}^d) = 1$ 。函数 p 在相差一个 Lebesgue 零测集的意义下唯一, 称为 X (或其分布) 的密度。当 $d = 1$ 时, 对任意 $\alpha \leq \beta$,

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x) dx.$$

8.1.3 数学期望

定义 8.1.3 设 X 为实随机变量, 即取值于 $\mathbb{R}$ 。定义

$$E[X] = \int_{\Omega} X(\omega) P(d\omega),$$

它在以下两种情况下有定义:

- $X \geq 0$, 此时 $E[X] \in [0, \infty]$;
- X 可正可负, 但 $E[|X|] = \int |X| dP < \infty$ 。

若 $X = (X_1, \dots, X_d)$ 取值于 $\mathbb{R}^d$, 则定义 $E[X] = (E[X_1], \dots, E[X_d])$, 前提是每个分量的期望均有定义。

注。若 $X = 1_B$, 则 $E[X] = P(B)$ 。一般地, $E[X]$ 可理解为随机变量 X 的平均值。特别地, 当 Ω 有限且 P 对每个单点赋予相同概率时, 这正是 X 所有取值的通常算术平均。

命题 8.1.1 设 X 是取值于可测空间 $(E, \mathcal{E})$ 的随机变量。对任意可测函数 $f: E \rightarrow [0, \infty]$,

$$\mathbb{E}[f(X)] = \int_E f(x) P_X(dx).$$

证明。对 $f = 1_B$ ($B \in \mathcal{E}$), 等式就是像测度 $P_X = P \circ X^{-1}$ 的定义; 由线性性, 等式对非负简单函数成立。再以非负简单函数单调逼近一般的非负可测函数, 并使用单调收敛定理, 即可得到一般情形。若 f 可取正负值, 而 $\mathbb{E}[|f(X)|] < \infty$, 同一公式仍成立。

因而, 已知 P_X 就能计算 $f(X)$ 的期望。反过来, 若对足够丰富的一类检验函数 f 都有

$$\mathbb{E}[f(X)] = \int_E f d\nu,$$

则可据此认出 $\nu = P_X$ 。

命题 8.1.2 设 $X = (X_1, \dots, X_d)$ 为 $\mathbb{R}^d$ -值随机变量, 并设 X 的分布关于 d 维 Lebesgue 测度具有密度 p 。则对每个 $j \in \{1, \dots, d\}$, 边缘随机变量 X_j 的密度为

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_d) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_d.$$

特别地, 当 $d = 2$ 时,

$$p_1(x) = \int_{\mathbb{R}} p(x, y) dy, \quad p_2(y) = \int_{\mathbb{R}} p(x, y) dx.$$

证明。 对任意非负 Borel 函数 $f: \mathbb{R} \rightarrow [0, \infty]$, 由 Fubini-Tonelli 定理,

$$\begin{aligned} \mathbb{E}[f(X_j)] &= \int_{\mathbb{R}^d} f(x_j) p(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \int_{\mathbb{R}} f(x_j) p_j(x_j) dx_j. \end{aligned}$$

由命题 8.1.1, p_j 正是 X_j 的密度。

注。 若 $X = (X_1, \dots, X_d)$ 取值于 $\mathbb{R}^d$, 则 X_j 的分布 P_{X_j} (常称为 X 的第 j 个边缘分布) 由 X 的分布确定, 因为 $P_{X_j} = \pi_j(P_X)$ 。但务必注意:

反之则不成立!

例如, 取 $\mathbb{R}$ 上的概率密度 q , 则 $p(x_1, x_2) = q(x_1)q(x_2)$ 是 $\mathbb{R}^2$ 上的概率密度。可构造 $X = (X_1, X_2)$, 使 P_X 关于 Lebesgue 测度的密度为 p 。随机向量 X 与 $X' = (X_1, X_1)$ 有相同的边缘分布, 因为 $P_{X_1}(dx) = P_{X_2}(dx) = q(x) dx$; 但 P_X 与 $P_{X'}$ 完全不同: $P_{X'}$ 集中在 $\mathbb{R}^2$ 的对角线上, 而该对角线的二维 Lebesgue 测度为零。

8.1.4 例: Bertrand 悖论

为了说明前面段落中引入的概念, 我们考虑以下问题。我们关注的是, 在一个圆上随机选择一条弦, 其长度大于内接等边三角形边长的概率。没有损失一般性, 我们可以假设该圆是单位圆。贝特朗提出了两种计算方法:

(a) 在圆周上独立、均匀地选取弦的两个端点。固定第一个端点后, 弦长大于内接等边三角形边长, 当且仅当第二个端点落在张角为 $2\pi/3$ 的圆弧内, 故概率为 $(2\pi/3)/(2\pi) = 1/3$ 。

(b) 在单位圆盘内均匀选取弦的中点。所求事件等价于中点落在以原点为圆心、半径 $1/2$ 的圆盘内; 面积比为 $1/4$, 故概率为 $1/4$ 。

因此, 在两种情况下得到的结果不同。解释在于, 两种方法对应不同的随机实验, 由不同的概率空间选择表示。因此, 没有理由认为所考虑的随机变量 (弦长) 的分布在两种情况下是相同的。为了证明这一点, 我们明确概率空间的选择。

(a) 严格地说, 取

$$\Omega = [0, 2\pi)^2, \quad \mathcal{A} = \mathcal{B}([0, 2\pi)^2), \quad P(d\theta d\theta') = \frac{1}{4\pi^2} d\theta d\theta'.$$

若 $\omega = (\theta, \theta')$, 弦长为

$$X(\omega) = 2 \left| \sin \frac{\theta - \theta'}{2} \right|.$$

对任意非负 Borel 函数 f ,

$$\begin{aligned} E[f(X)] &= \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} f\left(2 \left| \sin \frac{\theta - \theta'}{2} \right| \right) d\theta d\theta' \\ &= \frac{1}{\pi} \int_0^\pi f\left(2 \sin \frac{u}{2}\right) du \\ &= \frac{1}{\pi} \int_0^2 f(x) \frac{dx}{\sqrt{1-x^2/4}}. \end{aligned}$$

因此 X 的密度为

$$p(x) = \frac{1}{\pi \sqrt{1-x^2/4}} 1_{[0,2]}(x),$$

从而 $P(X \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x) dx = 1/3$ 。

(b) 现在取

$$\Omega = \{(y, z) \in \mathbb{R}^2 : y^2 + z^2 < 1\}, \quad \mathcal{A} = \mathcal{B}(\Omega), \quad P(dy dz) = \frac{1}{\pi} 1_\Omega(y, z) dy dz.$$

弦长为 $X(y, z) = 2\sqrt{1-y^2-z^2}$ 。极坐标变换给出

$$\begin{aligned} E[f(X)] &= \frac{1}{\pi} \int_{\mathbb{R}^2} f\left(2\sqrt{1-y^2-z^2}\right) 1_{\{y^2+z^2<1\}} dy dz \\ &= 2 \int_0^1 f\left(2\sqrt{1-r^2}\right) r dr = \frac{1}{2} \int_0^2 f(x) x dx. \end{aligned}$$

故 X 的密度为

$$p(x) = \frac{x}{2} 1_{[0,2]}(x),$$

并且 $P(X \geq \sqrt{3}) = 1/4$ 。¹

练习。 处理 Bertrand 的第三种方法：随机选择与弦垂直的半径方向，再在该半径上均匀选取弦的中点。

8.1.5 常见分布

本节列出若干重要分布。

离散分布。

(a) **均匀分布。** 若有限集合 E 有 n 个元素，则 X 在 E 上均匀分布是指

$$P(X = x) = \frac{1}{n}, \quad x \in E.$$

(b) **Bernoulli 分布。** 参数 $p \in [0, 1]$ ，取值于 $\{0, 1\}$ ，并满足

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

可把 X 理解为一次偏硬币投掷的正面指示变量。

¹校订注：原讲义在方法 (b) 的密度公式右端多排了一个微分 dx ；密度 $p(x)$ 是函数， dx 只应出现在积分或测度 $p(x)dx$ 中。

(c) **二项分布** $\mathcal{B}(n, p)$ 。其中 $n \in \mathbb{N}^*$ 、 $p \in [0, 1]$ ，且

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

它表示独立投掷上述硬币 n 次所得的正面次数。²

(d) **几何分布**。按原讲义的参数约定，若 $p \in (0, 1)$ ，

$$P(X = k) = (1-p)p^k, \quad k \in \mathbb{N}.$$

这里 X 表示连续出现正面（每次概率为 p ）直到第一次反面之前的正面次数。³

(e) **Poisson 分布**。参数 $\lambda > 0$ ，

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

容易算得 $E[X] = \lambda$ 。若 $X_n \sim \mathcal{B}(n, p_n)$ 且 $np_n \rightarrow \lambda$ ，则对每个固定的 $k \in \mathbb{N}$ ，

$$P(X_n = k) \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

这就是 Poisson 分布对稀有事件计数的二项近似。

连续分布。以下随机变量均取值于 $\mathbb{R}$ ，密度记为 $p(x)$ 。

(a) 区间 $[a, b]$ ($a < b$) 上的均匀分布：

$$p(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

(b) **指数分布**。参数 $\lambda > 0$ ，密度为

$$p(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x).$$

它满足无记忆性质：对 $a, b > 0$ ，

$$P(X > a+b) = P(X > a)P(X > b).$$

等价地， $P(X - a > b \mid X > a) = P(X > b)$ 。

(c) **高斯（正态）分布** $\mathcal{N}(m, \sigma^2)$ 。其中 $m \in \mathbb{R}$ 、 $\sigma > 0$ ，密度为

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

参数满足 $m = E[X]$ 、 $\sigma^2 = E[(X-m)^2]$ 。并且 $X - m \sim \mathcal{N}(0, \sigma^2)$ 。约定常值随机变量 m 服从 $\mathcal{N}(m, 0)$ 。若 $X \sim \mathcal{N}(m, \sigma^2)$ ，则对 $a, b \in \mathbb{R}$ ，

$$aX + b \sim \mathcal{N}(am + b, a^2\sigma^2).$$

实随机变量的分布函数。实随机变量 X 的分布函数定义为

$$F_X(t) = P(X \leq t) = P_X((-\infty, t]), \quad t \in \mathbb{R}.$$

函数 F_X 单调递增、右连续，并满足 $F_X(t) \rightarrow 0$ ($t \rightarrow -\infty$)、 $F_X(t) \rightarrow 1$ ($t \rightarrow +\infty$)。反之，任何满足这些性质的函数都是某个唯一概率测度的分布函数，因此 F_X 唯一确定 P_X 。特别地，若 $a \leq b$ ，

$$P(a \leq X \leq b) = F_X(b) - F_X(a-),$$

$$P(a < X < b) = F_X(b-) - F_X(a).$$

并且 $P(X = t) = F_X(t) - F_X(t-)$ ，所以分布函数的跳跃恰对应于分布的原子。

²校订注：原讲义把二项分布的值域写成 $\{1, \dots, n\}$ ，漏掉了可能值 0；公式本身也要求 $k = 0$ 的情形。

³学习提示：不少中国教材把几何分布写成 $p(1-p)^k$ ，此时参数 p 表示“首次成功”的概率。两种约定只交换了 p 与 $1-p$ ，使用时须先看清定义。

8.1.6 由随机变量生成的 σ -代数

设 X 取值于可测空间 $(E, \mathcal{E})$ 。由 X 生成的 σ -代数是使 X 可测的最小 σ -代数：

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{E}\}.$$

注。 若 $(X_i)_{i \in I}$ 是一族随机变量，其中 X_i 取值于 $(E_i, \mathcal{E}_i)$ ，则定义

$$\sigma(X_i : i \in I) = \sigma(X_i^{-1}(B) : B \in \mathcal{E}_i, i \in I).$$

^a

^a校订注：原讲义在这个多变量公式左端仍写 $\sigma(X)$ ，但此处并没有单个随机变量 X ；正确记号应明确表示整族 $(X_i)_{i \in I}$ 。

命题 8.1.3。 设 X 取值于 $(E, \mathcal{E})$ ， Y 为实随机变量。下列条件等价：

- (i) Y 关于 $\sigma(X)$ 可测；
- (ii) 存在可测函数 $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ，使 $Y = f(X)$ 。

证明。 (ii) $\Rightarrow$ (i) 由可测函数复合的可测性立即得到。反过来，先设 $Y = \sum_{i=1}^n \lambda_i 1_{A_i}$ 是 $\sigma(X)$ -可测简单函数。对每个 i ，可取 $B_i \in \mathcal{E}$ 使 $A_i = X^{-1}(B_i)$ 。于是

$$Y = \sum_{i=1}^n \lambda_i 1_{B_i} \circ X = f \circ X, \quad f = \sum_{i=1}^n \lambda_i 1_{B_i}.$$

一般地，取 $\sigma(X)$ -可测简单函数 $Y_n \rightarrow Y$ 逐点收敛，并写成 $Y_n = f_n(X)$ 。定义

$$f(x) = \begin{cases} \lim_{n \rightarrow \infty} f_n(x), & \text{若极限存在,} \\ 0, & \text{否则.} \end{cases}$$

则 f 可测；而对每个 ω ， $f_n(X(\omega)) = Y_n(\omega) \rightarrow Y(\omega)$ ，故 $f(X(\omega)) = Y(\omega)$ 。

8.2 随机变量的矩

8.2.1 p 阶矩与方差

设 X 为实随机变量， $p \geq 1$ 为整数。 X 的 p 阶矩定义为 $E[X^p]$ ；它只在 $E[|X|^p] < \infty$ 或 $X \geq 0$ 时有定义。量 $E[|X|^p]$ 称为 p 阶绝对矩。特别地，一阶矩就是期望。若实随机变量 X 可积且 $E[X] = 0$ ，则称 X 为中心化随机变量。数学期望是关于正测度的积分，因而可直接使用积分论中的一般定理；特别是以下收敛定理经常使用。

单调收敛： $X_n \geq 0, X_n \uparrow X \Rightarrow E[X_n] \uparrow E[X]$ 。

Fatou 引理：若 $X_n \geq 0$ ，则

$$E[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} E[X_n].$$

支配收敛定理：若 $|X_n| \leq Z$ 、 $E[Z] < \infty$ ，且 $X_n \rightarrow X$ 几乎必然成立，则 $E[X_n] \rightarrow E[X]$ 。

在概率论中，通常把测度论中的“几乎处处”称为“几乎必然”。空间 $L^p(\Omega, \mathcal{A}, P)$ ($p \in [1, \infty]$) 仍按积分论中的方式定义。Hölder 不等式写为

$$E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q},$$

其中 $p^{-1} + q^{-1} = 1$ 。取 $Y = 1$ ，得到 $\|X\|_1 \leq \|X\|_p$ ；更一般地，若 $r \leq p$ ，则 $\|X\|_r \leq \|X\|_p$ 。特别地， $L^p(\Omega, \mathcal{A}, P) \subset L^r(\Omega, \mathcal{A}, P)$ 。Cauchy-Schwarz 不等式写为

$$E[|XY|] \leq E[X^2]^{1/2} E[Y^2]^{1/2}$$

以及特殊情况 $Y=1$

$$E[|X|]^2 \leq E[X^2]$$

是非常常用的。

定义 8.2.1 设 $X \in L^2(\Omega, \mathcal{A}, P)$ 。X 的方差定义为

$$\text{Var}(X) = E[(X - E[X])^2]$$

而X的标准差为

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

直观地， $\text{Var}(X)$ 衡量 X 围绕均值 $E[X]$ 的离散程度。并且， $\text{Var}(X) = 0$ 当且仅当 X 几乎必然为常数。

命题 8.2.1 还有 $\text{Var}(X) = E[X^2] - (E[X])^2$ ，且对任意 $a \in \mathbb{R}$ ，

$$E[(X - a)^2] = \text{Var}(X) + (E[X] - a)^2.$$

因此，

$$\text{Var}(X) = \inf_{a \in \mathbb{R}} E[(X - a)^2].$$

证明。 展开平方并整理，得到

$$E[(X - a)^2] = E[X^2] - 2a E[X] + a^2 = E[X^2] - (E[X])^2 + (E[X] - a)^2.$$

这些结论由上式立即得到；其中令 $a = E[X]$ 即得 $\text{Var}(X) = E[X^2] - (E[X])^2$ 。

Markov 不等式。 若 $X \geq 0$ 、 $a > 0$ ，则

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Bienaymé-Chebyshev 不等式。 若 $X \in L^2(\Omega, \mathcal{A}, P)$ 、 $a > 0$ ，则

$$P(|X - E[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

这是把 Markov 不等式应用于非负随机变量 $(X - E[X])^2$ 得到的。

定义 8.2.2 设 $X, Y \in L^2(\Omega, \mathcal{A}, P)$ 。二者的协方差定义为

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[X(Y - E[Y])] = E[XY] - E[X]E[Y].$$

若 $X = (X_1, \dots, X_d)$ 取值于 $\mathbb{R}^d$ ，且每个分量都属于 $L^2(\Omega, \mathcal{A}, P)$ （等价于 $E[|X|^2] < \infty$ ），则 X 的协方差矩阵定义为

$$K_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}.$$

直观地， $\text{Cov}(X, Y)$ 衡量 X 与 Y 之间的线性相关程度。注意 $\text{Cov}(X, X) = \text{Var}(X)$ ，且由 Cauchy-Schwarz 不等式，

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}.$$

映射 $(X, Y) \mapsto \text{Cov}(X, Y)$ 是 $L^2(\Omega, \mathcal{A}, P)$ 上的双线性形式。在向量情形 $X = (X_1, \dots, X_d)$ 中，矩阵 K_X 对称且半正定：对任意 $\lambda_1, \dots, \lambda_d \in \mathbb{R}$ ，

$$\sum_{i,j=1}^d \lambda_i \lambda_j K_X(i, j) = \text{Var}\left(\sum_{i=1}^d \lambda_i X_i\right) \geq 0.$$

练习。 设 A 是 $n \times d$ 的确定性矩阵， $Y = AX$ 。证明

$$K_Y = AK_X A^\top.$$

8.2.2 线性回归

设 $X, Y_1, \dots, Y_n \in L^2(\Omega, \mathcal{A}, P)$ 。我们要用 $Y_1, \dots, Y_n$ 的仿射函数来最好地逼近 X ；也就是说，要在所有实数 $(\beta_0, \dots, \beta_n)$ 中最小化

$$E[(X - (\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n))^2]$$

其中最小值取遍所有 $(\beta_0, \dots, \beta_n) \in \mathbb{R}^{n+1}$ 。

命题 8.2.2 上述最小二乘问题的最小值由下面的随机变量 Z 取到：

$$\inf_{\beta_0, \dots, \beta_n \in \mathbb{R}} E[(X - (\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n))^2] = E[(X - Z)^2],$$

其中

$$Z = E[X] + \sum_{j=1}^n \alpha_j (Y_j - E[Y_j]). \quad (1)$$

系数 $\alpha_1, \dots, \alpha_n$ 可取为下列线性方程组的任意一组解：

$$\sum_{j=1}^n \alpha_j \text{Cov}(Y_j, Y_k) = \text{Cov}(X, Y_k), \quad 1 \leq k \leq n.$$

特别地，若 K_Y 非退化，则矩阵形式为 $\alpha = \text{Cov}(X, Y) K_Y^{-1}$ 。

证明。令 H 为 $L^2(\Omega, \mathcal{A}, P)$ 中由 $1, Y_1, \dots, Y_n$ 张成的线性子空间。使 $\|X - U\|_2$ ($U \in H$) 最小的随机变量 Z ，正是 X 在 H 上的正交投影。可把 Z 写成

$$Z = \alpha_0 + \sum_{j=1}^n \alpha_j (Y_j - E[Y_j]).$$

由正交投影的定义， $X - Z$ 与 H 正交。因此

$$E[(X - Z) \cdot 1] = 0,$$

故 $\alpha_0 = E[X]$ 。同理，对每个 $k \in \{1, \dots, n\}$,

$$E[(X - Z) \cdot (Y_k - E[Y_k])] = 0,$$

这等价于 $\text{Cov}(Z, Y_k) = \text{Cov}(X, Y_k)$ ，也就是

$$\sum_{j=1}^n \alpha_j \text{Cov}(Y_j, Y_k) = \text{Cov}(X, Y_k).$$

反过来，若 α_j 满足上述方程组，则式 (1) 定义的 $Z \in H$ ，且 $X - Z \perp H$ ，故 Z 正是 X 在 H 上的正交投影。**注。**若 $n = 1$ 且 Y 不几乎必然为常数，则用 Y 的仿射函数对 X 作最佳 L^2 逼近所得

$$Z = E[X] + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} (Y - E[Y]).$$

这有时称为 X 关于 Y 的回归直线。

8.2.3 特征函数

定义 8.2.3 设 X 为取值于 $\mathbb{R}^d$ 的随机变量。 X 的特征函数是映射 $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ ，定义为

$$\Phi_X(\xi) = E[\exp(i\xi \cdot X)], \quad \xi \in \mathbb{R}^d.$$

也可以写作

$$\Phi_X(\xi) = \int e^{i\xi \cdot x} P_X(dx)$$

因此， Φ_X 可看作 X 的分布 P_X 的傅里叶变换，有时也写成 $\Phi_X(\xi) = \hat{P}_X(\xi)$ 。支配收敛定理表明 Φ_X 在 $\mathbb{R}^d$ 上连续且有界。我们的目标是证明特征函数唯一确定分布。先计算一个重要的特殊情形。

引理 8.2.3 设 $X \sim \mathcal{N}(0, \sigma^2)$ 。则对每个 $\xi \in \mathbb{R}$,

$$\Phi_X(\xi) = \exp\left(-\frac{\sigma^2 \xi^2}{2}\right), \quad \xi \in \mathbb{R}.$$

证明。按定义，

$$\Phi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} e^{i\xi x} dx.$$

作变量替换 $x = \sigma y$, 可化归到 $\sigma = 1$. 由奇偶性, $\Phi_X(\xi)$ 的虚部为零, 因此只需计算

$$f(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(\xi x) dx.$$

在积分号下求导, 得到

$$f'(\xi) = - \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \sin(\xi x) dx$$

这里可在积分号下求导, 因为 $|x \sin(\xi x) e^{-x^2/2}| \leq |x| e^{-x^2/2}$, 而右端可积. 再作分部积分, 得到

$$f'(\xi) = - \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \xi \cos(\xi x) dx = -\xi f(\xi).$$

因此 f 满足微分方程 $f'(\xi) = -\xi f(\xi)$ 及初值 $f(0) = 1$, 所以 $f(\xi) = e^{-\xi^2/2}$.

定理 8.2.4 (特征函数唯一性定理) $\mathbb{R}^d$ -值随机变量的特征函数唯一确定其分布。等价地, $\mathbb{R}^d$ 上概率测度的傅里叶变换是单射。

证明。 先处理 $d = 1$. 对 $\sigma > 0$, 记 g_σ 为高斯分布 $\mathcal{N}(0, \sigma^2)$ 的密度:

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

若 μ 是 $\mathbb{R}$ 上的概率测度, 定义平滑后的密度 f_σ 及概率测度 μ_σ :

$$\begin{aligned} f_\sigma(x) &= \int_{\mathbb{R}} g_\sigma(x-y) \mu(dy) := (g_\sigma * \mu)(x), \\ \mu_\sigma(dx) &= f_\sigma(x) dx. \end{aligned}$$

为证明定理, 只需验证下列两点:

- (1) μ_σ 由 $\hat{\mu}$ 唯一确定;
- (2) 对每个 $\varphi \in C_b(\mathbb{R})$, 当 $\sigma \rightarrow 0$ 时, $\int \varphi d\mu_\sigma \rightarrow \int \varphi d\mu$.

为证明 (1), 由上一引理, 对每个 $x \in \mathbb{R}$,

$$\sigma\sqrt{2\pi} g_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) = \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) d\xi.$$

于是

$$\begin{aligned} f_\sigma(x) &= \int_{\mathbb{R}} g_\sigma(x-y) \mu(dy) \\ &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{i\xi(x-y)} g_{1/\sigma}(\xi) d\xi \right) \mu(dy) \\ &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \left(\int_{\mathbb{R}} e^{-i\xi y} \mu(dy) \right) d\xi \\ &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \hat{\mu}(-\xi) d\xi. \end{aligned}$$

倒数第二个等式使用了 Fubini 定理; 这是合法的, 因为 μ 是概率测度, 而 $g_{1/\sigma}$ 关于 Lebesgue 测度可积. 因此 f_σ , 从而 μ_σ , 都由 $\hat{\mu}$ 唯一确定. 为证明 (2), 对任意 $\varphi \in C_b(\mathbb{R})$ 写成

$$\begin{aligned}\int_{\mathbb{R}} \varphi(x) \mu_{\sigma}(dx) &= \int_{\mathbb{R}} \varphi(x) \left(\int_{\mathbb{R}} g_{\sigma}(y-x) \mu(dy) \right) dx \\ &= \int_{\mathbb{R}} (g_{\sigma} * \varphi)(y) \mu(dy).\end{aligned}$$

这里同样由 Fubini 定理交换积分次序。再使用高斯核的性质

$$\begin{aligned}\int g_{\sigma}(x) dx &= 1, \\ \lim_{\sigma \rightarrow 0} \int_{\{|x|>\varepsilon\}} g_{\sigma}(x) dx &= 0, \quad \forall \varepsilon > 0,\end{aligned}$$

由此得到⁴

$$\lim_{\sigma \rightarrow 0} (g_{\sigma} * \varphi)(y) = \varphi(y).$$

这正是第 5 章中 δ_0 的逼近恒等族。又有 $|g_{\sigma} * \varphi| \leq \|\varphi\|_{\infty}$ ，故对 μ 使用支配收敛定理可得

$$\lim_{\sigma \rightarrow 0} \int \varphi(x) \mu_{\sigma}(dx) = \int \varphi(x) \mu(dx),$$

这就完成了 $d = 1$ 的证明。一般 d 的论证完全类似，只需使用乘积高斯核

$$g_{\sigma}^{(d)}(x_1, \dots, x_d) = \prod_{j=1}^d g_{\sigma}(x_j)$$

并注意，对每个 $\xi \in \mathbb{R}^d$ ，有⁵

$$\int_{\mathbb{R}^d} g_{\sigma}^{(d)}(x) e^{i\xi \cdot x} dx = \prod_{j=1}^d \int_{\mathbb{R}} g_{\sigma}(x_j) e^{i\xi_j x_j} dx_j = \left(\frac{2\pi}{\sigma^2} \right)^{d/2} g_{1/\sigma}^{(d)}(\xi).$$

命题 8.2.5 设 $X = (X_1, \dots, X_d)$ 是平方可积的 $\mathbb{R}^d$ -值随机变量。则 $\Phi_X \in C^2$ ，并且

$$\Phi_X(\xi) = 1 + i \sum_{j=1}^d \xi_j E[X_j] - \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \xi_j \xi_k E[X_j X_k] + o(|\xi|^2)$$

当 $\xi = (\xi_1, \dots, \xi_d) \rightarrow 0$ 时成立。

证明。对积分符号下进行求导，我们得到

$$\frac{\partial \Phi_X}{\partial \xi_j}(\xi) = i E[X_j e^{i\xi \cdot X}],$$

这是因为 $|i X_j e^{i\xi \cdot X}| = |X_j|$ ，且 $X_j \in L^2 \subset L^1$ 。同理，由 Cauchy-Schwarz 不等式，

$$E[|X_j X_k|] \leq E[X_j^2]^{1/2} E[X_k^2]^{1/2} < \infty.$$

故可再求一次导数，得到

⁴校订注：原讲义这里把趋于零的参数误写成 ε ；实际变化的是高斯核尺度 σ 。

⁵校订注：原讲义此处的比例因子误排为 $(2\pi\sigma)^{d/2}$ 。由一维恒等式逐坐标相乘，正确因子应为 $(2\pi/\sigma^2)^{d/2}$ 。

$$\frac{\partial^2 \Phi_X}{\partial \xi_j \partial \xi_k}(\xi) = -E[X_j X_k e^{i\xi \cdot X}].$$

此外，积分号下连续性定理保证 $\frac{\partial^2 \Phi_X}{\partial \xi_j \partial \xi_k}(\xi)$ 是 ξ 的连续函数。最后一个断言正是 Φ_X 在原点的二阶 Taylor 展开。**注意：**若 X 的 p 次矩可积，其中整数 $p \geq 1$ ，同样的论证说明 $\Phi_X \in C^p$ 。后文最常用的是 $p = 2$ 。

8.2.4 母函数

对取值于 $\mathbb{N}$ 的随机变量，通常使用概率母函数，它比特征函数更直接地编码各点概率。

定义 8.2.4 设 X 是 $\mathbb{N}$ -值随机变量。 X 的概率母函数是定义在 $[0, 1]$ 上的函数 g_X ：

$$g_X(r) = E[r^X] = \sum_{n=0}^{\infty} P(X = n) r^n.$$

由支配收敛定理， g_X 在 $[0, 1]$ 上连续，并且 $g_X(0) = P(X = 0)$ 、 $g_X(1) = 1$ 。定义式中的幂级数收敛半径至少为 1；其在 0 处的 Taylor 系数为 $P(X = n)$ ，所以 g_X 唯一确定 X 的分布。函数 g_X 在 1 处总有左导数（允许为 $+\infty$ ），并且

因此，概率母函数在 1 处的各阶左导数给出 X 的各阶下降阶乘矩。

第九章 独立性

独立性是概率论中第一个明显有别于积分论的重要概念。两个事件或两个随机变量的独立性最容易直观理解，但更基本的对象是两个（或多个）子 σ -代数的独立性。本章的一个关键结果是：两个随机变量独立，当且仅当它们组成的随机向量的分布等于各自分布的乘积测度。借助 Fubini 定理，这一结果常可给出独立性的有用等价表述。作为应用，本章将证明著名的 Borel–Cantelli 引理（其中一个有趣的应用揭示随机选取的实数之二进制展开所具有的若干性质），并给出大数定律的一个初步形式。这足以把概率论的公理化方法同概率的“历史定义”联系起来：大量重复进行同一随机试验时，事件的概率由该事件出现的频率来刻画。

9.1 独立事件

本章始终固定概率空间 $(\Omega, \mathcal{A}, P)$ 。若 $A, B \in \mathcal{A}$ 是两个事件，则称 A 与 B 相互独立，如果

$$P(A \cap B) = P(A)P(B).$$

当 $P(B) > 0$ 时，可用条件概率

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

来解释独立性：等式 $P(A | B) = P(A)$ 表明，已知事件 B 发生并不改变事件 A 的概率。独立性的定义关于 A, B 对称。

例。(i) 掷两枚骰子： $\Omega = \{1, 2, \dots, 6\}^2$ ，且对任意 $\omega \in \Omega$ ， $P(\{\omega\}) = 1/36$ 。事件

$$A = \{6\} \times \{1, 2, \dots, 6\}, \quad B = \{1, 2, \dots, 6\} \times \{6\}$$

相互独立。事实上， P 的构造正是为了使只涉及第一次掷骰的事件与只涉及第二次掷骰的事件相互独立。

(ii) 掷一次骰子： $\Omega = \{1, 2, \dots, 6\}$ ，且对每个 $\omega \in \Omega$ ， $\mathbb{P}(\{\omega\}) = 1/6$ 。事件 $A = \{1, 2\}$ 与 $B = \{1, 3, 5\}$ 相互独立。

定义 9.1.1 称 n 个事件 $A_1, \dots, A_n$ 相互独立，如果对 $\{1, \dots, n\}$ 的每个非空子集 $\{j_1, \dots, j_p\}$ ，都有

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \cdots \mathbb{P}(A_{j_p}).$$

注。仅有下面一个等式并不足以推出 $A_1, \dots, A_n$ 相互独立：

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

即使每一对 $\{i, j\} \subset \{1, \dots, n\}$ 中的事件 A_i 与 A_j 都独立，也仍然不够。例如，考虑连续抛掷两次均匀硬币的样本空间，令

$$\begin{aligned} A &= \{\text{第一次出现正面}\}, \\ B &= \{\text{第二次出现正面}\}, \\ C &= \{\text{两次结果相同}\}. \end{aligned}$$

事件 A, B, C 两两独立，却不是相互独立的。

命题 9.1.1 事件 $A_1, \dots, A_n$ 相互独立，当且仅当

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \cdots \mathbb{P}(B_n)$$

对每个 $i \in \{1, \dots, n\}$ 以及任意 $B_i \in \sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$ 都成立。

证明。 上述条件显然蕴含定义中的条件：只要对 $i \in \{j_1, \dots, j_p\}$ 取 $B_i = A_i$ ，其余指标取 $B_i = \Omega$ 即可。反之，设 $A_1, \dots, A_n$ 相互独立。验证命题中的等式时，不妨假设每个 $B_i \neq \emptyset$ 。令 $\{j_1, \dots, j_p\} = \{i : B_i \neq \Omega\}$ ，问题便化为证明

$$\mathbb{P}(B_{j_1} \cap \dots \cap B_{j_p}) = \mathbb{P}(B_{j_1}) \cdots \mathbb{P}(B_{j_p}),$$

其中每个 B_{j_k} 等于 A_{j_k} 或 $A_{j_k}^c$ 。归根结底，只需证明：若 $C_1, C_2, \dots, C_p$ 相互独立，则 $C_1^c, C_2, \dots, C_p$ 也相互独立。事实上，对 $\{2, \dots, p\}$ 的任意子集 $\{i_1, \dots, i_q\}$ ，

$$\begin{aligned} \mathbb{P}(C_1^c \cap C_{i_1} \cap \dots \cap C_{i_q}) &= \mathbb{P}(C_{i_1} \cap \dots \cap C_{i_q}) - \mathbb{P}(C_1 \cap C_{i_1} \cap \dots \cap C_{i_q}) \\ &= \mathbb{P}(C_{i_1}) \cdots \mathbb{P}(C_{i_q}) - \mathbb{P}(C_1) \mathbb{P}(C_{i_1}) \cdots \mathbb{P}(C_{i_q}) \\ &= \mathbb{P}(C_1^c) \mathbb{P}(C_{i_1}) \cdots \mathbb{P}(C_{i_q}). \end{aligned}$$

9.2 独立随机变量与独立 σ -代数

更一般地，可以定义若干个 σ -代数的独立性。

定义 9.2.1 设 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 是 $\mathcal{A}$ 的子 σ -代数。称它们相互独立，如果

$$\begin{aligned} \forall A_1 \in \mathcal{B}_1, \dots, \forall A_n \in \mathcal{B}_n, \\ \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n). \end{aligned}$$

设 $X_1, \dots, X_n$ 分别是取值于下列可测空间的随机变量：

$$(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n).$$

若 σ -代数 $\sigma(X_1), \dots, \sigma(X_n)$ 相互独立，就称 $X_1, \dots, X_n$ 相互独立。等价地，

$$\forall F_1 \in \mathcal{E}_1, \dots, \forall F_n \in \mathcal{E}_n, \quad \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in F_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in F_i). \quad (9.1)$$

这里用到了 $\sigma(X_i) = \{X_i^{-1}(F) : F \in \mathcal{E}_i\}$ 。

直观地说， $X_1, \dots, X_n$ 相互独立，意味着知道其中某些变量的取值不会提供关于其余变量的

信息。**注：**(i) 若 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 是相互独立的子 σ -代数，且对每个 i ，随机变量 X_i 是 $\mathcal{B}_i$ -可测的，则 $X_1, \dots, X_n$ 相互独立。(ii) 事件 $A_1, \dots, A_n$ 相互独立当且仅当 σ -代数 $\sigma(A_1), \dots, \sigma(A_n)$ 相互独立 (参见上一命题)。

若 $X_1, \dots, X_n$ 分别取值于 $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ ，则随机向量 $(X_1, \dots, X_n)$ 取值于配备乘积 σ -代数 $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ 的乘积空间 $E_1 \times \dots \times E_n$ 。

定理 9.2.1 随机变量 $X_1, \dots, X_n$ 相互独立，当且仅当随机向量 $(X_1, \dots, X_n)$ 的分布等于各分量分布的乘积：

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

并且此时

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$$

只要对每个 $i \in \{1, \dots, n\}$ ，函数 $f_i : (E_i, \mathcal{E}_i) \rightarrow [0, \infty]$ 可测。

证明。任取 $F_i \in \mathcal{E}_i$ ($1 \leq i \leq n$)。一方面，

$$\mathbb{P}_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = \mathbb{P} \left(\bigcap_{i=1}^n \{X_i \in F_i\} \right)$$

另一方面，

$$\begin{aligned} (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(F_1 \times \dots \times F_n) &= \prod_{i=1}^n \mathbb{P}_{X_i}(F_i) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in F_i). \end{aligned}$$

与式 (9.1) 比较可知， $X_1, \dots, X_n$ 相互独立，当且仅当两个概率测度 $\mathbb{P}_{(X_1, \dots, X_n)}$ 与 $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ 在所有矩形集 $F_1 \times \dots \times F_n$ 上取值相同。由单调类引理，乘积空间上的概率测度由其在这些矩形集上的取值唯一确定，故这又等价于 $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ 。定理的第二个断言则由 Fubini–Tonelli 定理得到：

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] &= \int_{E_1 \times \dots \times E_n} \prod_{i=1}^n f_i(x_i) \mathbb{P}_{X_1}(dx_1) \dots \mathbb{P}_{X_n}(dx_n) \\ &= \prod_{i=1}^n \int_{E_i} f_i(x_i) \mathbb{P}_{X_i}(dx_i) \\ &= \prod_{i=1}^n \mathbb{E}[f_i(X_i)]. \end{aligned}$$

上述定理还给出了构造独立随机变量的方法。以实随机变量为例，设 $\mu_1, \dots, \mu_n$ 是 $\mathbb{R}$ 上的概率测度。由第 8 章，可以构造一个取值于 $\mathbb{R}^n$ 的随机向量 $Y = (Y_1, \dots, Y_n)$ ，使其分布为 $\mu_1 \otimes \dots \otimes \mu_n$ 。根据前一定理，分量 $Y_1, \dots, Y_n$ 相互独立，且各自的分布依次为 $\mu_1, \dots, \mu_n$ 。**注。**若函数 f_i 不必非负，则等式

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$$

在每个 $\mathbb{E}[|f_i(X_i)|] < \infty$ 时仍成立；此时还有

$$\mathbb{E} \left[\prod_{i=1}^n |f_i(X_i)| \right] = \prod_{i=1}^n \mathbb{E}[|f_i(X_i)|] < \infty$$

这也说明了前一公式左端确实有定义。特别地，若 $X_1, \dots, X_n$ 是相互独立且属于 L^1 的实随机变量，则 $X_1 \cdots X_n \in L^1$ ，并且

$$\mathbb{E}[X_1 \cdots X_n] = \prod_{i=1}^n \mathbb{E}[X_i].$$

一般而言，若干个 L^1 随机变量的乘积未必仍属于 L^1 ；这里的结论依赖于独立性。

推论 9.2.2 若 X_1, X_2 是两个相互独立且属于 L^2 的实随机变量，则 $\text{Cov}(X_1, X_2) = 0$ 。

这是因为 $\text{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0$ 。

该推论的逆命题不成立：对两个 L^2 随机变量而言，协方差为零远弱于相互独立。取实随机变量 X_1 ，其分布具有对称密度 p ，即 $p(x) = p(-x)$ ，并满足 $\int_{\mathbb{R}} x^2 p(x) dx < \infty$ ，从而 $X_1 \in L^2$ 。例如可取 $X_1 \sim \mathcal{N}(0, \sigma^2)$ 。再取与 X_1 独立、在 $\{-1, 1\}$ 上等概率取值的随机变量 ε ，即 $P(\varepsilon = 1) = P(\varepsilon = -1) = 1/2$ ，并令 $X_2 = \varepsilon X_1$ 。由对称性，

$$\text{Cov}(X_1, X_2) = E[\varepsilon X_1^2] = E[\varepsilon]E[X_1^2] = 0.$$

但 X_1 与 X_2 并不独立。否则 $|X_1|$ 将与 $|X_2| = |X_1|$ 独立；一个随机变量若与自身独立，就必为几乎处处常数（请读者证明），这与 $|X_1|$ 具有密度 $2p(x)\mathbf{1}_{\mathbb{R}_+}(x)$ 矛盾。

推论 9.2.3 设 $X_1, \dots, X_n$ 为 n 个实随机变量。(i) 若对每个 $i \in \{1, \dots, n\}$ ， X_i 的分布有密度 p_i ，且 $X_1, \dots, X_n$ 相互独立，则随机向量 $(X_1, \dots, X_n)$ 的分布具有密度

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i).$$

(ii) 反过来，若 $(X_1, \dots, X_n)$ 的分布具有乘积形式的密度

$$p(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i),$$

其中 $q_1, \dots, q_n: \mathbb{R} \rightarrow [0, \infty)$ 是 Borel 函数，则 $X_1, \dots, X_n$ 相互独立；并且对每个 i ， X_i 的密度可写成 $p_i = C_i q_i$ ，其中 $C_i > 0$ 为常数。

证明。(i) 若 $P_{X_i}(dx_i) = p_i(x_i) dx_i$ ，则由 Fubini–Tonelli 定理，

$$P_{X_1} \otimes \cdots \otimes P_{X_n}(dx_1 \dots dx_n) = \left(\prod_{i=1}^n p_i(x_i) \right) dx_1 \dots dx_n.$$

(ii) 仍由 Fubini–Tonelli 定理，

$$\prod_{i=1}^n \left(\int_{\mathbb{R}} q_i(x) dx \right) = \int_{\mathbb{R}^n} p(x_1, \dots, x_n) dx_1 \dots dx_n = 1,$$

特别地, 对每个 $i \in \{1, \dots, n\}$, $K_i := \int_{\mathbb{R}} q_i(x) dx \in (0, \infty)$ 。由第 8 章的边缘密度公式, X_i 的密度为

$$p_i(x_i) = \int_{\mathbb{R}^{n-1}} p(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n = \left(\prod_{j \neq i} K_j \right) q_i(x_i) = \frac{1}{K_i} q_i(x_i).$$

因此 $(X_1, \dots, X_n)$ 的联合密度可改写为

$$p(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i) = \prod_{i=1}^n p_i(x_i)$$

于是 $P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}$, 故 $X_1, \dots, X_n$ 相互独立。

例。 设 U 服从参数为 1 的指数分布, V 在 $[0, 1]$ 上均匀分布, 并且 U, V 相互独立。定义

$$X = \sqrt{U} \cos(2\pi V), \quad Y = \sqrt{U} \sin(2\pi V),$$

则 X 与 Y 相互独立。为此计算 (X, Y) 的分布。对任意非负 Borel 函数 $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}_+$,

$$\begin{aligned} E[\varphi(X, Y)] &= \int_0^\infty \int_0^1 \varphi(\sqrt{u} \cos(2\pi v), \sqrt{u} \sin(2\pi v)) e^{-u} dv du \\ &= \frac{1}{\pi} \int_0^\infty \int_0^{2\pi} \varphi(r \cos \theta, r \sin \theta) r e^{-r^2} d\theta dr \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} \varphi(x, y) e^{-x^2-y^2} dx dy. \end{aligned}$$

因此 (X, Y) 的密度为 $\pi^{-1} e^{-x^2-y^2}$, 它具有乘积形式, 故由推论 9.2.3, X 与 Y 独立。还可看出二者具有相同的密度

$$p(x) = \frac{1}{\sqrt{\pi}} \exp(-x^2)$$

所以 X 与 Y 都服从高斯分布 $\mathcal{N}(0, 1/2)$ 。

注。 对实随机变量 $X_1, \dots, X_n$, 下列条件等价:

- (i) $X_1, \dots, X_n$ 相互独立;
- (ii) 对任意 $a_1, \dots, a_n \in \mathbb{R}$,

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) = \prod_{i=1}^n P(X_i \leq a_i);$$

- (iii) 对任意连续紧支集函数 $f_1, \dots, f_n: \mathbb{R} \rightarrow \mathbb{R}_+$,

$$E\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n E[f_i(X_i)].$$

(iv) 随机向量 $X = (X_1, \dots, X_n)$ 的特征函数满足

$$\Phi_X(\xi_1, \dots, \xi_n) = \prod_{i=1}^n \Phi_{X_i}(\xi_i)$$

(为证明 (iv) 蕴含 (i), 可使用傅里叶变换的单射性, 参见第 8 章。)

现在我们转向一个非常有用的技术性结果。

命题 9.2.4 设 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 是 $\mathcal{A}$ 的子 σ -代数。对每个 $i \in \{1, \dots, n\}$, 设 $\mathcal{C}_i \subset \mathcal{B}_i$ 对有限交封闭、包含 Ω , 且 $\sigma(\mathcal{C}_i) = \mathcal{B}_i$ 。假设

$$\forall C_1 \in \mathcal{C}_1, \dots, \forall C_n \in \mathcal{C}_n, \quad P(C_1 \cap C_2 \cap \dots \cap C_n) = P(C_1)P(C_2) \dots P(C_n)$$

则 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 相互独立。

证明。 先固定 $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$, 并令

$$\mathcal{M}_1 = \{B_1 \in \mathcal{B}_1 : P(B_1 \cap C_2 \cap \dots \cap C_n) = P(B_1)P(C_2) \dots P(C_n)\}.$$

由假设 $\mathcal{C}_1 \subset \mathcal{M}_1$, 且容易验证 $\mathcal{M}_1$ 是单调类。单调类定理给出 $\mathcal{B}_1 = \sigma(\mathcal{C}_1) \subset \mathcal{M}_1$, 故

$$\begin{aligned} \forall B_1 \in \mathcal{B}_1, \forall C_2 \in \mathcal{C}_2, \dots, \forall C_n \in \mathcal{C}_n, \\ P(B_1 \cap C_2 \cap \dots \cap C_n) = P(B_1)P(C_2) \dots P(C_n). \end{aligned}$$

接着固定 $B_1 \in \mathcal{B}_1, C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$, 并令

$$\mathcal{M}_2 = \{B_2 \in \mathcal{B}_2 : P(B_1 \cap B_2 \cap C_3 \cap \dots \cap C_n) = P(B_1)P(B_2)P(C_3) \dots P(C_n)\}.$$

¹ 同理, $\mathcal{M}_2$ 是包含 $\mathcal{C}_2$ 的单调类, 故包含 $\mathcal{B}_2$ 。依次对其余坐标重复, 即得结论。

推论 (分组原理)。设 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 是相互独立的 σ -代数, 并设 $n_0 = 0 < n_1 < \dots < n_p = n$ 。则下列 σ -代数

$$\begin{aligned} \mathcal{D}_1 &= \mathcal{B}_1 \vee \dots \vee \mathcal{B}_{n_1} = \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1}), \\ \mathcal{D}_2 &= \mathcal{B}_{n_1+1} \vee \dots \vee \mathcal{B}_{n_2}, \\ &\vdots \\ \mathcal{D}_p &= \mathcal{B}_{n_{p-1}+1} \vee \dots \vee \mathcal{B}_{n_p}. \end{aligned}$$

为证明这一点, 在命题 9.2.4 中, 对每个 j 取 $\mathcal{C}_j$ 为所有形如

$$B_{n_{j-1}+1} \cap \dots \cap B_{n_j}$$

的集合, 其中每个 $B_i \in \mathcal{B}_i, n_{j-1}+1 \leq i \leq n_j$ 。特别地, 若 $X_1, \dots, X_n$ 相互独立, 则分块随机向量

$$Y_1 = (X_1, \dots, X_{n_1}), \dots, Y_p = (X_{n_{p-1}+1}, \dots, X_{n_p})$$

也相互独立。**例。**若 $X_1, \dots, X_4$ 相互独立, 则

¹校订注: 原文在此处把新定义的单调类误记为 $\mathcal{M}_1$; 下句及递推结构均表明应为 $\mathcal{M}_2$ 。

$$Z_1 = X_1 X_3, \quad Z_2 = X_2^3 + X_4$$

是独立的。

无限族的独立性

设 $(\mathcal{B}_i)_{i \in I}$ 是 $\mathcal{A}$ 的任意一族子 σ -代数。若它的每个有限子族都相互独立，则称整个族相互独立。类似地，随机变量族 $(X_i)_{i \in I}$ 称为相互独立，如果 $(\sigma(X_i))_{i \in I}$ 相互独立。

命题 9.2.5 设 $(X_n)_{n \in \mathbb{N}}$ 是相互独立的随机变量列。则对每个 $p \in \mathbb{N}$ ，下列两个 σ -代数相互独立：

$$\mathcal{B}_1 = \sigma(X_0, \dots, X_p), \quad \mathcal{B}_2 = \sigma(X_{p+1}, X_{p+2}, \dots)$$

。

证明。应用命题 9.2.4，取

$$\begin{aligned} \mathcal{C}_1 &= \sigma(X_0, \dots, X_p) = \mathcal{B}_1 \\ \mathcal{C}_2 &= \bigcup_{k=p+1}^{\infty} \sigma(X_{p+1}, X_{p+2}, \dots, X_k) \subset \mathcal{B}_2 \end{aligned}$$

由有限个独立随机变量的分组原理，命题 9.2.4 的假设成立，故 $\mathcal{B}_1$ 与 $\mathcal{B}_2$ 相互独立。

9.3 Borel–Cantelli 引理

若 $(A_n)_{n \in \mathbb{N}}$ 是事件序列，记

$$\begin{aligned} \limsup A_n &= \bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right) \\ \liminf A_n &= \bigcup_{n=0}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k \right) \end{aligned}$$

引理 9.3.1 设 $(A_n)_{n \in \mathbb{N}}$ 为事件序列。

(i) 若 $\sum_{n \in \mathbb{N}} P(A_n) < \infty$ ，则

$$P(\limsup A_n) = 0$$

等价地，几乎必然只有有限多个 A_n 发生，即 $\{n \in \mathbb{N} : \omega \in A_n\}$ 几乎必然为有限集。

(ii) 若 $\sum_{n \in \mathbb{N}} P(A_n) = \infty$ ，且事件 $(A_n)_{n \in \mathbb{N}}$ 相互独立，则

$$P(\limsup A_n) = 1$$

等价地，几乎必然有无穷多个 A_n 发生，即 $\{n \in \mathbb{N} : \omega \in A_n\}$ 几乎必然为无限集。

注。第 (ii) 部分必须加入独立性或其他适当条件。例如若对所有 n 都取 $A_n = A$ ，其中 $0 < P(A) < 1$ ，则 $\sum_n P(A_n) = \infty$ ，但 $P(\limsup_n A_n) = P(A) < 1$ 。

证明。(i) 若 $\sum_n P(A_n) < \infty$, 则

$$E\left[\sum_{n \in \mathbb{N}} 1_{A_n}\right] = \sum_{n \in \mathbb{N}} P(A_n) < \infty$$

故 $\sum_n 1_{A_n} < \infty$ 几乎必然, 也就是只有有限多个 A_n 发生。

(ii) 固定 $n_0 \in \mathbb{N}$ 。对任意 $n \geq n_0$, 由独立性,

$$P\left(\bigcap_{k=n_0}^n A_k^c\right) = \prod_{k=n_0}^n P(A_k^c) = \prod_{k=n_0}^n (1 - P(A_k)).$$

因为 $\log(1-x) \leq -x$ ($0 \leq x < 1$), 而 $\sum_k P(A_k) = \infty$, 所以 $\prod_{k=n_0}^n (1 - P(A_k)) \rightarrow 0$ 。由上连续性得到

$$P\left(\bigcap_{k=n_0}^{\infty} A_k^c\right) = 0.$$

这对每个 $n_0 \in \mathbb{N}$ 都成立, 因此

$$P\left(\bigcup_{n_0=0}^{\infty} \left(\bigcap_{k=n_0}^{\infty} A_k^c\right)\right) = 0$$

取补集便得到

$$P\left(\bigcap_{n_0=0}^{\infty} \left(\bigcup_{k=n_0}^{\infty} A_k\right)\right) = 1,$$

右端事件正是 $\limsup_n A_n$, 故第 (ii) 部分得证。

两个应用。(1) 不存在 $\mathbb{N}$ 上的概率测度 P , 使每个整数 $n \geq 1$ 的倍数集 $n\mathbb{N}$ 都满足 $P(n\mathbb{N}) = 1/n$ 。反设存在。记 $\mathcal{P}$ 为素数集, 对 $p \in \mathcal{P}$ 令 $A_p = p\mathbb{N}$ 。若 $p_1, \dots, p_k$ 是互不相同的素数, 则

$$\begin{aligned} P(A_{p_1} \cap \dots \cap A_{p_k}) &= P(p_1\mathbb{N} \cap \dots \cap p_k\mathbb{N}) \\ &= P((p_1 \cdots p_k)\mathbb{N}) = \frac{1}{p_1 \cdots p_k} = \prod_{j=1}^k P(A_{p_j}). \end{aligned}$$

此外, Euler 的经典结果给出

$$\sum_{p \in \mathcal{P}} P(A_p) = \sum_{p \in \mathcal{P}} \frac{1}{p} = \infty.$$

于是由 Borel–Cantelli 引理 (ii), P -几乎每个整数都属于无穷多个 A_p , 即能被无穷多个不同素数整除, 矛盾。(2) 考虑概率空间

$$(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda).$$

对每个 $n \geq 1$, 定义

$$\forall \omega \in [0, 1], \quad X_n(\omega) = \lfloor 2^n \omega \rfloor - 2 \lfloor 2^{n-1} \omega \rfloor.$$

其中 $\lfloor x \rfloor$ 表示 x 的整数部分。于是 $X_n(\omega) \in \{0, 1\}$ 。对 n 作简单归纳可得, 对每个 $\omega \in [0, 1]$,

$$0 \leq \omega - \sum_{k=1}^n X_k(\omega) 2^{-k} < 2^{-n},$$

因此

$$\omega = \sum_{k=1}^{\infty} X_k(\omega) 2^{-k}.$$

因此, 数 $X_k(\omega)$ 是 ω 的规范二进展开的系数。把事件 $\{X_n = 1\}$ 写出来, 容易证明对每个 $n \geq 1$,

$$P(X_n = 0) = P(X_n = 1) = \frac{1}{2}.$$

最后, $(X_n)_{n \geq 1}$ 相互独立。只需验证对任意 $i_1, \dots, i_p \in \{0, 1\}$,

$$P(X_1 = i_1, \dots, X_p = i_p) = \frac{1}{2^p} = \prod_{j=1}^p P(X_j = i_j).$$

然而, 我们立即看到

$$\{X_1 = i_1, \dots, X_p = i_p\} = \left[\sum_{j=1}^p i_j 2^{-j}, \sum_{j=1}^p i_j 2^{-j} + 2^{-p} \right).$$

因此独立性得证。现在固定整数 $p \geq 1$ 以及 $(i_1, \dots, i_p) \in \{0, 1\}^p$ 。Borel–Cantelli 引理将给出

$$\text{Card}\{k \geq 0 : X_{k+1} = i_1, \dots, X_{k+p} = i_p\} = \infty \quad \text{几乎必然.} \quad (9.2)$$

这说明: 对 Lebesgue 测度意义下几乎每个 $\omega \in [0, 1)$, 其二进制展开中任意给定的有限 0-1 字块都会出现无穷多次。为证明式 (9.2), 对每个 $n \in \mathbb{N}$ 定义

$$Y_n = (X_{np+1}, X_{np+2}, \dots, X_{np+p}).$$

由分组原理, 序列 $(Y_n)_{n \in \mathbb{N}}$ 相互独立。把 Borel–Cantelli 引理应用于事件

$$A_n = \{Y_n = (i_1, \dots, i_p)\}$$

即可得到结论, 因为这些事件相互独立, 且都有概率 2^{-p} 。再利用可数个零概率集合之并仍为零概率, 可把式 (9.2) 加强为

$$\mathbb{P}(\forall p \geq 1, \forall (i_1, \dots, i_p) \in \{0, 1\}^p, \text{Card}\{k \geq 0 : X_{k+1} = i_1, \dots, X_{k+p} = i_p\} = \infty) = 1.$$

换言之, 对几乎每个 $x \in [0, 1)$, 任意有限 0-1 字块都会在 x 的二进制展开中出现无穷多次。

9.4 独立随机变量之和

独立随机变量之和在概率论中极为重要, 并将在下一章继续研究。先记下卷积的定义。若 μ, ν 是 $\mathbb{R}^d$ 上的概率测度, 则 $\mu * \nu$ 定义为乘积测度 $\mu \otimes \nu$ 在映射 $(x, y) \mapsto x + y$ 下的像测度。于是对任意非负可测函数 $\varphi: \mathbb{R}^d \rightarrow [0, \infty]$,

$$\int_{\mathbb{R}^d} \varphi(z) \mu * \nu(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x+y) \mu(dx) \nu(dy).$$

命题 9.4.1 设 X, Y 是两个相互独立的 $\mathbb{R}^d$ -值随机变量。

- (i) $X+Y$ 的分布为 $P_X * P_Y$ 。特别地, 若 X 与 Y 的密度分别为 p_X 与 p_Y , 则 $X+Y$ 的密度为 $p_X * p_Y$ 。
- (ii) $X+Y$ 的特征函数满足 $\Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi)$ 。等价地, 若 μ, ν 是 $\mathbb{R}^d$ 上的概率测度, 则 $\widehat{\mu * \nu} = \widehat{\mu} \widehat{\nu}$ 。
- (iii) 若 X, Y 平方可积, 则 $K_{X+Y} = K_X + K_Y$ 。特别地, 当 $d=1$ 时,

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

证明。(i) 由独立性, $P_{(X,Y)} = P_X \otimes P_Y$ 。因此对任意非负可测函数 $\varphi: \mathbb{R}^d \rightarrow [0, \infty]$,

$$\int \varphi(x+y) P_{(X,Y)}(dxdy) = \int \int \varphi(x+y) P_X(dx) P_Y(dy) = \int \varphi(z) P_{X*P_Y}(dz)$$

右端按定义就是 $\int \varphi d(P_X * P_Y)$, 故 $P_{X+Y} = P_X * P_Y$ 。若 X, Y 还分别具有密度 p_X, p_Y , 则

$$E[\varphi(X+Y)] = \iint \varphi(x+y) p_X(x) p_Y(y) dx dy = \int \varphi(z) \left(\int p_X(x) p_Y(z-x) dx \right) dz.$$

这表明 $X+Y$ 的密度为 $p_X * p_Y$ 。这里 $p_X * p_Y$ 作为两个 $L^1(\mathbb{R}^d, \lambda)$ 函数的卷积, 几乎处处有定义。(ii) 只需写出

$$\Phi_{X+Y}(\xi) = E[e^{i\xi \cdot X} e^{i\xi \cdot Y}] = E[e^{i\xi \cdot X}] E[e^{i\xi \cdot Y}] = \Phi_X(\xi) \Phi_Y(\xi).$$

(iii) 若 $X = (X_1, \dots, X_d)$ 、 $Y = (Y_1, \dots, Y_d)$, 则 X 与 Y 的独立性蕴含 $\text{Cov}(X_i, Y_j) = 0$, 其中 $i, j \in \{1, \dots, d\}$ 。因此, 由协方差的双线性性,

$$\text{Cov}(X_i + Y_i, X_j + Y_j) = \text{Cov}(X_i, X_j) + \text{Cov}(Y_i, Y_j)$$

这正是矩阵恒等式 $K_{X+Y} = K_X + K_Y$ 。

定理 9.4.2 (弱大数定律) 设 $(X_n)_{n \geq 1}$ 是独立同分布的实随机变量, 且 $E[X_1^2] < \infty$ 。则

$$\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow[n \rightarrow \infty]{L^2} E[X_1].$$

证明。由数学期望的线性性,

$$E\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = E[X_1].$$

因此,

$$\begin{aligned} E\left[\left(\frac{1}{n}(X_1 + \dots + X_n) - E[X_1]\right)^2\right] &= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) \\ &= \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{1}{n} \text{Var}(X_1). \end{aligned}$$

当 $n \rightarrow \infty$ 时, 右端趋于 0。

注。证明表明, 结论在更弱的假设下仍成立。无需假设 X_n 同分布, 只要对每个 n 有 $E[X_n] = E[X_1]$, 且序列 $(E[X_n^2])$ 有界即可。独立性也可弱化为: 当 $n \neq m$ 时, $\text{Cov}(X_n, X_m) = 0$ 。“弱”大数定律中的“弱”, 是指这里得到的是 L^2 收敛; 从概率论观点看, 更强也更自然的是几乎必然收敛, 即除去一个零概率集合后的逐点收敛, 称为强大数定律。下面先给出一个初步结果, 下一章将大幅推广。

命题 9.4.3 在定理 9.4.2 的假设下, 再假设 $E[X_1^4] < \infty$ 。则几乎必然有

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{} E[X_1].$$

证明。把每个 X_n 替换为 $X_n - E[X_n]$, 不改变要证的结论, 故不妨设 $E[X_n] = 0$ 。于是

$$E\left[\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)^4\right] = \frac{1}{n^4} \sum_{i_1, \dots, i_4 \in \{1, \dots, n\}} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}].$$

利用独立性和性质 $E[X_k] = 0$, 我们看到和中唯一非零的项是那部分四元组 (i_1, i_2, i_3, i_4) 的每个分量值在该四元组中至少出现两次的项。利用 X_k 服从相同分布的事实, 我们得到

$$E\left[\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)^4\right] = \frac{1}{n^4} \left(nE[X_1^4] + 3n(n-1)E[X_1^2 X_2^2] \right) \leq \frac{C}{n^2}$$

对于某个常数 $C < \infty$ 。由此可得

$$\sum_{n=1}^{\infty} E\left[\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)^4\right] < \infty.$$

交换求和与期望, 我们获得

$$E\left[\sum_{n=1}^{\infty} \left(\frac{1}{n}(X_1 + \cdots + X_n)\right)^4\right] < \infty,$$

因此

$$\sum_{n=1}^{\infty} \left(\frac{1}{n}(X_1 + \cdots + X_n)\right)^4 < \infty, \quad \text{几乎必然}$$

可见样本均值的四次方几乎必然趋于 0, 从而样本均值本身几乎必然趋于 0; 还原中心化之前的均值即得命题。

推论 9.4.4 若 $(A_n)_{n \geq 1}$ 是相互独立且概率相同的事件列, 则

$$\frac{1}{n} \sum_{i=1}^n 1_{A_i} \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} P(A_1).$$

这个推论将我们现代公理化方法与概率的历史定义联系起来, 即当我们重复大量随机实验时, 事件出现的频率。回到上面给出的 Borel–Cantelli 引理的第二个应用, 它涉及一个实数的二进制展开

$$\omega = \sum_{k=1}^{\infty} X_k(\omega) 2^{-k}$$

若固定 $p \geq 1$, 我们已经看到随机变量

$$Y_1 = (X_1, \dots, X_p), \quad Y_2 = (X_{p+1}, \dots, X_{2p}), \quad \dots$$

相互独立且同分布。因此, 由上述推论, 对任意 $i_1, \dots, i_p \in \{0, 1\}$, 有

$$\frac{1}{n} \text{Card}\{j \leq n : Y_j(\omega) = (i_1, \dots, i_p)\} \xrightarrow{n \rightarrow \infty} \frac{1}{2^p}, \quad d\omega\text{-几乎处处.}$$

对每个 $\ell \in \{1, \dots, p\}$, 把同一论证应用于随机变量

$$(X_\ell, X_{\ell+1}, \dots, X_{p+\ell-1}), (X_{p+\ell}, X_{p+\ell+1}, \dots, X_{2p+\ell-1}), \dots$$

便得到

$$\frac{1}{n} \text{Card}\{j \leq n : X_{jp+\ell}(\omega) = i_1, \dots, X_{(j+1)p+\ell-1}(\omega) = i_p\} \xrightarrow{n \rightarrow \infty} \frac{1}{2^p}, \quad d\omega\text{-几乎处处.}$$

合并这些结果可得

$$\frac{1}{n} \text{Card}\{k \leq n : X_{k+1}(\omega) = i_1, \dots, X_{k+p}(\omega) = i_p\} \xrightarrow{n \rightarrow \infty} \frac{1}{2^p}, \quad d\omega\text{-几乎处处.}$$

由于可数个零测集之并仍是零测集, 故除去一个零测集后, 对每个 $\omega \in [0, 1)$ 都同时有

$$\forall p \geq 1, \forall i_1, \dots, i_p \in \{0, 1\}, \quad \frac{1}{n} \text{Card}\{k \leq n : X_{k+1}(\omega) = i_1, \dots, X_{k+p}(\omega) = i_p\} \xrightarrow{n \rightarrow \infty} \frac{1}{2^p}. \quad (9.3)$$

换言之, 对 $[0, 1)$ 中几乎每个实数 ω , 在它的二进制展开中, 任意一个由 0 与 1 组成的有限长字块都有渐近频率; 若字块长度为 p , 该频率就是 2^{-p} . 要直接写出满足性质 (9.3) 的实数并不容易。事实上, 证明这类实数存在的最快方法, 很可能正是上述概率论证。这是概率方法处理存在性问题的典型例子: 为证明具有某些性质的对象存在, 只需证明按照适当概率分布随机抽取的对象以正概率 (这里甚至以概率一) 具有这些性质。

卷积半群

设 $I = \mathbb{N}$ 或 $I = \mathbb{R}_+$ 。

定义 9.4.1 设 $(\mu_t)_{t \in I}$ 是 $\mathbb{R}$ (或 $\mathbb{R}^d$) 上的一族概率测度。若 $\mu_0 = \delta_0$, 并且

$$\mu_t * \mu_{t'} = \mu_{t+t'}, \quad \forall t, t' \in I,$$

则称 $(\mu_t)_{t \in I}$ 为卷积半群。

其概率解释如下: 若 X 的分布为 μ_t , Y 的分布为 $\mu_{t'}$, 且 X, Y 相互独立, 则 $X + Y$ 的分布为 $\mu_{t+t'}$ (参见本部分的第一个命题)。

引理 9.4.5 要使 $(\mu_t)_{t \in I}$ 成为卷积半群, 只需存在函数 $\varphi: \mathbb{R} \rightarrow \mathbb{C}$, 满足:

- 若 $I = \mathbb{N}$, 则 $\hat{\mu}_t(\xi) = \varphi(\xi)^t$, 对所有 $t \in I$ 成立;

- 若 $I = \mathbb{R}_+$, 则 $\hat{\mu}_t(\xi) = \exp(-t\varphi(\xi))$, 对所有 $t \in I$ 成立。

a

^a校订注: 原讲义第二项把指标集写成 $I = \mathbb{R}$, 但本节一开始只允许 $I = \mathbb{N}$ 或 $I = \mathbb{R}_+$, 而且卷积半群参数 t 在这里取非负值; 故应为 $I = \mathbb{R}_+$ 。

证明。 上述两种情形都立即给出

$$\widehat{\mu_{t+t'}} = \widehat{\mu_t} \widehat{\mu_{t'}} = \widehat{\mu_t * \mu_{t'}}.$$

由傅里叶变换的单射性, 得到 $\mu_{t+t'} = \mu_t * \mu_{t'}$ 。

例。 (1) 取 $I = \mathbb{N}$ 。固定 $p \in [0, 1]$, 对每个 $n \geq 1$, 令 μ_n 为二项分布 $\mathcal{B}(n, p)$, 并令 $\mu_0 = \delta_0$ 。由二项分布的概率解释, $\mu_{n+m} = \mu_n * \mu_m$ 。也可使用前一引理, 因为

$$\hat{\mu}_n(\xi) = (pe^{i\xi} + 1 - p)^n.$$

(2) 取 $I = \mathbb{R}_+$ 。对每个 $t \geq 0$, 令 μ_t 为参数为 t 的 Poisson 分布。此时

$$\hat{\mu}_t(\xi) = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{ik\xi} e^{-t} = \exp(-t(1 - e^{i\xi})).$$

(3) 取 $I = \mathbb{R}_+$ 。对 $t > 0$, 令 μ_t 为高斯分布 $\mathcal{N}(0, t)$, 并约定 $\mu_0 = \delta_0$ 。第 8 章已经算得²

$$\hat{\mu}_t(\xi) = \exp(-\frac{t\xi^2}{2}).$$

一个重要推论如下。若 X 与 X' 是相互独立的实随机变量, 则:

- 若 $X \sim \text{Poisson}(\lambda)$ 、 $X' \sim \text{Poisson}(\lambda')$, 则 $X + X' \sim \text{Poisson}(\lambda + \lambda')$;
- 若 $X \sim \text{Poisson}(\lambda)$ 、 $X' \sim \text{Poisson}(\lambda')$, 则 $X + X' \sim \text{Poisson}(\lambda + \lambda')$;
- 若 $X \sim \mathcal{N}(m, \sigma^2)$ 、 $X' \sim \mathcal{N}(m', \sigma'^2)$, 则

$$X + X' \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2).$$

(对 $X - m$ 与 $X' - m'$ 应用中心情形即可。)

更一般地, 相互独立的高斯随机变量的任意线性组合仍为高斯随机变量。

²校订注: 原讲义只写了 $t > 0$, 但卷积半群还必须包含 $t = 0$; 这里补出 $\mu_0 = \delta_0$ 。

第十章 随机变量的收敛

本章的第一部分介绍了随机变量收敛的不同概念，以及这些概念之间的联系。随后介绍了大数定律，这是概率论的两个基本极限定理之一。第三段介绍了随机变量的分布收敛：这种收敛类型可能是最难理解的，部分原因在于它是一种测度收敛（收敛的是随机变量的分布，而不是随机变量本身）。分布收敛的概念，以及将这种收敛与特征函数收敛联系起来的重要定理，使得我们能够得到第二个基本极限定理，即中心极限定理。

10.1 不同的收敛概念

设 $(X_n)_{n \geq 1}$ 与 X 是定义在概率空间 $(\Omega, \mathcal{A}, P)$ 上、取值于 $\mathbb{R}^d$ 的随机变量。前面已经遇到过若干种 $X_n \rightarrow X$ 的含义。特别地，

称 (X_n) 几乎必然收敛到 X ，记作 $X_n \xrightarrow{\text{a.s.}} X$ ，如果

$$P\left(\left\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\right\}\right) = 1.$$

而对 $p \in [1, \infty)$ ，记 $X_n \xrightarrow{L^p} X$ ，如果 $\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0$ 。

定义 10.1.1 若对每个 $\varepsilon > 0$ 都有下式成立，则称 (X_n) 依概率收敛到 X ，记作

$$X_n \xrightarrow[n \rightarrow \infty]{(P)} X$$

即

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

命题 10.1.1 令 $\mathcal{L}_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, P)$ 为所有 $\mathbb{R}^d$ -值随机变量的空间，令 $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, P)$ 为它关于等价关系 $X \sim Y \iff X = Y$ 几乎必然的商空间。则公式

$$d(X, Y) = E[|X - Y| \wedge 1]$$

这就在 $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, P)$ 上定义了一个距离。它与概率收敛相容： (X_n) 依概率收敛到 X ，当且仅当 $d(X_n, X) \rightarrow 0$ 。此外，空间 $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, P)$ 关于距离 d 是完备的。

证明。容易验证 d 在几乎必然相等的等价类上定义良好，并满足距离的三条公理。若 X_n 依概率收敛到 X ，则对每个 $\varepsilon > 0$ ，

$$\begin{aligned} d(X_n, X) &= E[|X_n - X| \wedge 1] \\ &\leq E[|X_n - X| 1_{\{|X_n - X| \leq \varepsilon\}}] \\ &\quad + E[(|X_n - X| \wedge 1) 1_{\{|X_n - X| > \varepsilon\}}] \\ &\leq \varepsilon + P(|X_n - X| > \varepsilon). \end{aligned}$$

因此 $\limsup_{n \rightarrow \infty} d(X_n, X) \leq \varepsilon$ 。令 $\varepsilon \downarrow 0$ ，得到 $d(X_n, X) \rightarrow 0$ 。反过来，若 $d(X_n, X) \rightarrow 0$ ，则对

每个 $\varepsilon \in (0, 1]$,

$$P(|X_n - X| > \varepsilon) \leq \varepsilon^{-1} E[|X_n - X| \wedge 1] = \varepsilon^{-1} d(X_n, X) \xrightarrow{n \rightarrow \infty} 0.$$

下面证明完备性。设 (X_n) 是距离 d 下的 Cauchy 列。可抽取子列 $Y_k = X_{n_k}$, 使对每个 $k \geq 1$,

$$d(Y_k, Y_{k+1}) \leq 2^{-k}.$$

于是

$$E\left[\sum_{k=1}^{\infty} (|Y_{k+1} - Y_k| \wedge 1)\right] = \sum_{k=1}^{\infty} d(Y_k, Y_{k+1}) < \infty,$$

故 $\sum_{k \geq 1} (|Y_{k+1} - Y_k| \wedge 1) < \infty$ 几乎必然。特别地, 几乎必然只有有限多个 k 满足 $|Y_{k+1} - Y_k| \geq 1$; 去掉这些有限项后, 截断值等于原值, 所以 $\sum_{k \geq 1} |Y_{k+1} - Y_k| < \infty$ 几乎必然。因而可定义 L^0 中的随机变量

$$X = Y_1 + \sum_{k=1}^{\infty} (Y_{k+1} - Y_k).$$

按构造, $Y_k \rightarrow X$ 几乎必然。由于 $|Y_k - X| \wedge 1 \leq 1$, 支配收敛定理给出

$$d(Y_k, X) = E[|Y_k - X| \wedge 1] \xrightarrow{k \rightarrow \infty} 0,$$

故 Y_k 依概率收敛到 X 。又因为原列 (X_n) 是 Cauchy 列, 对任意 $\varepsilon > 0$, 先固定充分大的 k , 再令 $n \rightarrow \infty$, 由三角不等式可得 $d(X_n, X) \rightarrow 0$ 。这证明了完备性。以上抽取过程还表明: 任何依概率收敛的序列, 都可抽出一个几乎必然收敛到同一极限的子列。

命题 10.1.2 若 $X_n \rightarrow X$ 几乎必然, 或在 L^p 中收敛 ($p \in [1, \infty]$), 则 $X_n \rightarrow X$ 依概率成立。反过来, 若 $X_n \rightarrow X$ 依概率成立, 则存在子列 (X_{n_k}) 几乎必然收敛到 X 。

证明。 反向断言已由上一命题的证明得到。若 $X_n \rightarrow X$ 几乎必然, 则由支配收敛定理,

$$d(X_n, X) = E[|X_n - X| \wedge 1] \xrightarrow{n \rightarrow \infty} 0,$$

若 $X_n \rightarrow X$ 于 L^p , 则

$$d(X_n, X) \leq \|X_n - X\|_1 \leq \|X_n - X\|_p \xrightarrow{n \rightarrow \infty} 0.$$

综上, 依概率收敛弱于几乎必然收敛, 也弱于任意 L^p 收敛; 但依概率收敛总能抽出几乎必然收敛到同一极限的子列。下面给出由依概率收敛推出 L^p 收敛的常用充分条件。

命题 10.1.3 设 $X_n \rightarrow X$ 依概率成立。若存在 $r \in (1, \infty)$, 使 $\sup_n \|X_n\|_r < \infty$, 则对每个 $p \in [1, r)$, 都有 $X_n \rightarrow X$ 于 L^p 。

证明。 由假设, 存在常数 $C < \infty$, 使 $\sup_n E[|X_n|^r] \leq C$ 。从任意子列都可再抽取一个几乎必然收敛到 X 的子列; 对该子列使用 Fatou 引理可得 $E[|X|^r] \leq C$, 故 $X \in L^r$ 。固定 $p \in [1, r)$ 与 $\varepsilon > 0$ 。由 Hölder 不等式,

$$\begin{aligned}
E[|X_n - X|^p] &= E[|X_n - X|^p 1_{\{|X_n - X| \leq \varepsilon\}}] + E[|X_n - X|^p 1_{\{|X_n - X| > \varepsilon\}}] \\
&\leq \varepsilon^p + E[|X_n - X|^r]^{p/r} P(|X_n - X| > \varepsilon)^{1-p/r} \\
&\leq \varepsilon^p + 2^p C^{p/r} P(|X_n - X| > \varepsilon)^{1-p/r}.
\end{aligned}$$

由于 $X_n \rightarrow X$ 依概率成立, 令 $n \rightarrow \infty$ 可得

$$\limsup_{n \rightarrow \infty} E[|X_n - X|^p] \leq \varepsilon^p$$

最后令 $\varepsilon \downarrow 0$, 便得到 $E[|X_n - X|^p] \rightarrow 0$ 。

10.2 强大数律

我们的目标是证明, 如果 (X_n) 是一个独立同分布随机变量序列, 在 L^1 中, 那么平均值 $\frac{1}{n}(X_1 + \cdots + X_n)$ 会几乎必然收敛到 $E[X_1]$ 。我们已经在附加假设 $E[|X_1|^4] < \infty$ 的情况下得到了这个结果, 但现在我们希望在最优假设下建立它。我们从一个重要的初步结果开始。

定理 10.2.1 (Kolmogorov 0-1 律) 设 $(X_n)_{n \geq 1}$ 是取值于任意可测空间的一列相互独立的随机变量。对每个 $n \geq 1$, 令

$$\mathcal{B}_n = \sigma(X_k; k \geq n).$$

并定义尾 σ -代数

$$\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n$$

则 $\mathcal{B}_\infty$ 是平凡的, 即对每个 $B \in \mathcal{B}_\infty$, 都有 $P(B) \in \{0, 1\}$ 。

证明。 令

$$\mathcal{D}_n = \sigma(X_k; k \leq n).$$

第 9 章已经说明, $\mathcal{D}_n$ 与 $\mathcal{B}_{n+1}$ 独立, 因而与 $\mathcal{B}_\infty \subset \mathcal{B}_{n+1}$ 独立。所以

$$\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n, \forall B \in \mathcal{B}_\infty, \quad P(A \cap B) = P(A)P(B).$$

集合族 $\bigcup_{n \geq 1} \mathcal{D}_n$ 对有限交封闭。由第 9 章的独立性判别, $\mathcal{B}_\infty$ 与

$$\sigma\left(\bigcup_{n=1}^{\infty} \mathcal{D}_n\right) = \sigma(X_n; n \geq 1).$$

独立。由于 $\mathcal{B}_\infty$ 本身包含于右端的 σ -代数, 它与自身独立。因此对任意 $B \in \mathcal{B}_\infty$,

$$P(B) = P(B \cap B) = P(B)^2,$$

故 $P(B) = 0$ 或 1 。另外, 关于平凡 σ -代数可测的实随机变量必几乎必然为常数; 例如可从其分布函数只能取 $0, 1$ 看出。因此, 对任意一列相互独立的实随机变量 (X_n) , 随机变量

$$\limsup_{n \rightarrow \infty} \frac{1}{n}(X_1 + \cdots + X_n)$$

关于 $\mathcal{B}_\infty$ 可测, 故它作为取值于 $[-\infty, \infty]$ 的扩充实随机变量, 几乎必然为常数。特别地, 若 $n^{-1}(X_1 + \cdots + X_n)$ 几乎必然收敛, 则其极限几乎必然是常数。下面先把 0-1 律用于简单对称随机游走。

命题 10.2.2 设 $(X_n)_{n \geq 1}$ 相互独立, 且对每个 n ,

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

令

$$S_n = X_1 + X_2 + \cdots + X_n.$$

则几乎必然有

$$\sup_{n \geq 1} S_n = +\infty, \quad \inf_{n \geq 1} S_n = -\infty.$$

特别地, 几乎必然存在任意大的整数 n , 使得 $S_n = 0$ 。

换言之, 若玩家每轮以各 $1/2$ 的概率赢得或输掉一欧元, 则 S_n 是前 n 轮的累计盈亏。命题说明, 当时间趋于无穷时, 累计盈亏会到达任意大的正值和任意小的负值。

证明。 先证明对每个整数 $p \geq 1$,

$$P(-p \leq \inf_n S_n \leq \sup_n S_n \leq p) = 0.$$

固定整数 $k > 2p$, 注意到

$$\bigcup_{j=0}^{\infty} \{X_{jk+1} = X_{jk+2} = \cdots = X_{jk+k} = 1\} \subset (\{-p \leq \inf_n S_n \leq \sup_n S_n \leq p\})^c.$$

左端各事件相互独立, 且概率均为 $2^{-k} > 0$ 。由第二 Borel–Cantelli 引理, 其并集的概率为 1, 因而得到上式。令 $p \rightarrow \infty$, 得到

$$P(\{\inf_n S_n > -\infty\} \cap \{\sup_n S_n < \infty\}) = 0,$$

因此

$$P(\{\inf_n S_n = -\infty\} \cup \{\sup_n S_n = \infty\}) = 1,$$

特别地

$$P(\{\inf_n S_n = -\infty\}) + P(\{\sup_n S_n = \infty\}) \geq 1.$$

一个对称性论证表明

$$P(\{\inf_n S_n = -\infty\}) = P(\{\sup_n S_n = \infty\})$$

根据上述内容, 这两个概率都是严格大于零的。总之, 我们注意到

$$\left\{ \sup_n S_n = \infty \right\} \in \mathcal{B}_\infty.$$

另一方面, 对任意 $k \geq 1$,

$$\{\sup_n S_n = \infty\} = \{\sup_{n \geq k} (X_k + X_{k+1} + \cdots + X_n) = \infty\} \in \mathcal{B}_k$$

所以事件 $\{\sup_n S_n = \infty\}$ 属于 $\bigcap_{k \geq 1} \mathcal{B}_k = \mathcal{B}_\infty$ 。前面已经说明它具有严格正概率，故由 0-1 律，其概率必为 1。对称性给出 $P(\inf_n S_n = -\infty) = 1$ 。由于随机游走每步只变动 1，它从任意正值走到负值时必须经过 0，于是最后一个断言也成立。

现在我们进入本段的主要结果。

定理 10.2.3 (强大数律) 设 $(X_n)_{n \geq 1}$ 是一列属于 L^1 的独立同分布实随机变量。则

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} E[X_1].$$

注。(i) 可积性假设是最优的，因为它保证 $E[X_1]$ 有定义且有限。若 $X_n \geq 0$ 且 $E[X_1] = \infty$ ，把定理应用于 $X_n \wedge K$ ，再令 $K \rightarrow \infty$ ，可得

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} +\infty$$

(令 $K \rightarrow \infty$ 即可。)

(ii) 事实上，定理中的收敛也在 L^1 中成立。这里暂不证明；第 12 章末将由鞅理论推出。就概率解释而言，几乎必然收敛最为重要。

证明。记 $S_n = X_1 + \cdots + X_n$ 、 $S_0 = 0$ 。固定 $a > E[X_1]$ ，并令

$$M = \sup_{n \in \mathbb{N}} (S_n - na)$$

这是取值于 $[0, \infty]$ 的随机变量。我们先证明

$$M < \infty \quad \text{几乎必然.} \tag{10.1}$$

由于对每个 n 都有 $S_n \leq na + M$ ，立刻得到

$$\limsup_{n \rightarrow \infty} \frac{1}{n} S_n \leq a, \quad \text{几乎必然}$$

让 a 沿一列数递减到 $E[X_1]$ ，得到

$$\limsup_{n \rightarrow \infty} \frac{1}{n} S_n \leq E[X_1], \quad \text{几乎必然}$$

再把同一论证用于 $(-X_n)$ ，得到反向不等式

$$\liminf_{n \rightarrow \infty} \frac{1}{n} S_n \geq E[X_1], \quad \text{几乎必然}$$

结合最后两个不等式即得强大数律。还需证明 (10.1)。按照 0-1 律中的记号，事件 $\{M < \infty\}$ 属于 $\mathcal{B}_\infty$ 。事实上，对每个 $k \geq 0$ ，

$$\{M < \infty\} = \{\sup_{n \in \mathbb{N}} (S_n - na) < \infty\} = \{\sup_{n \geq k} (S_n - S_k - (n-k)a) < \infty\}$$

而最后一个事件关于 $\sigma(X_{k+1}, X_{k+2}, \dots)$ 可测。因此，只需证明 $P(M < \infty) > 0$ ，等价地，证明 $P(M = \infty) < 1$ 。下面用反证法完成。

让我们从一些符号记法开始。对于所有整数 $k \in \mathbb{N}$, 设

$$M_k = \sup_{0 \leq n \leq k} (S_n - na),$$

$$M'_k = \sup_{0 \leq n \leq k} (S_{n+1} - S_1 - na).$$

随机变量 M_k 与 M'_k 同分布。事实上, 向量 $(X_1, \dots, X_k)$ 与 $(X_2, \dots, X_{k+1})$ 同分布; 另一方面, 存在同一个确定性函数 $F_k: \mathbb{R}^k \rightarrow \mathbb{R}$, 使

$$M_k = F_k(X_1, \dots, X_k), \quad M'_k = F_k(X_2, \dots, X_{k+1}).$$

因此

$$M = \lim_{k \rightarrow \infty} \uparrow M_k$$

和

$$M' = \lim_{k \rightarrow \infty} \uparrow M'_k$$

也同分布。事实上,

$$P(M' \leq x) = \lim_{k \rightarrow \infty} \downarrow P(M'_k \leq x) = \lim_{k \rightarrow \infty} \downarrow P(M_k \leq x) = P(M \leq x).$$

另一方面, 由定义可知, 对每个 $k \geq 1$,

$$M_{k+1} = \sup \left(0, \sup_{1 \leq n \leq k+1} (S_n - na) \right) = \sup(0, M'_k + X_1 - a),$$

这可以进一步重写为

$$M_{k+1} = M'_k - \inf(a - X_1, M'_k).$$

由于 M'_k 与 M_k 具有相同的分布 (并且这两个随机变量显然属于 L^1), 我们得到

$$E[\inf(a - X_1, M'_k)] = E[M'_k] - E[M_{k+1}] = E[M_k] - E[M_{k+1}] \leq 0$$

因为显然有 $M_k \leq M_{k+1}$ 。现在对随机变量列 $\inf(a - X_1, M'_k)$ 应用支配收敛定理; 它们的绝对值由 $|a - X_1|$ 控制 (回忆 $M'_k \geq 0$)。于是

$$E[\inf(a - X_1, M')] = \lim_{k \rightarrow \infty} E[\inf(a - X_1, M'_k)] \leq 0.$$

若 $P(M = \infty) = 1$, 则因 M 与 M' 同分布, 也有 $P(M' = \infty) = 1$ 。于是 $\inf(a - X_1, M') = a - X_1$ 几乎必然成立。前一不等式便给出 $E[a - X_1] \leq 0$, 这与所取 $a > E[X_1]$ 矛盾。证明完毕。

10.3 依分布收敛

记 $C_b(\mathbb{R}^d)$ 为 $\mathbb{R}^d$ 上全体实值有界连续函数组成的空间, 并赋以上确界范数

$$\|\varphi\| = \sup_{x \in \mathbb{R}^d} |\varphi(x)|.$$

定义 10.3.1 设 (μ_n) 与 μ 是 $\mathbb{R}^d$ 上的概率测度。若对每个 $\varphi \in C_b(\mathbb{R}^d)$, 都有 $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$, 则称 μ_n 弱收敛于 μ , 记作 $\mu_n \xrightarrow{w} \mu$ 。

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \int \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi d\mu.$$

若 $\mathbb{R}^d$ -值随机变量列 (X_n) 的分布 P_{X_n} 弱收敛于 P_X , 则称 X_n 依分布收敛于 X , 记作 $X_n \xrightarrow{\text{law}} X$ 。等价地,

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad E[\varphi(X_n)] \xrightarrow{n \rightarrow \infty} E[\varphi(X)].$$

注。(i) “ X_n 依分布收敛于 X ”只确定极限分布 P_X , 并不确定极限随机变量本身。因此也可说 X_n 依分布收敛于概率测度 μ , 意即 P_{X_n} 弱收敛于 μ 。依分布收敛允许各随机变量定义在不同概率空间上, 这是它与前述收敛方式的重要区别。

(ii) 概率测度可视为 $C_b(\mathbb{R}^d)$ 上的线性泛函; 弱收敛就是这些泛函在每个测试函数上的逐点收敛。

例。(a) 若随机变量 X_n, X 取值于 $\mathbb{Z}^d$, 则 X_n 依分布收敛于 X , 当且仅当

$$\forall x \in \mathbb{Z}^d, \quad P(X_n = x) \xrightarrow{n \rightarrow \infty} P(X = x)$$

反向蕴含可由稍后证明的命题 10.3.3 直接得到。

(b) 设 $P_{X_n}(dx) = p_n(x) dx$ 。若

$$p_n(x) \rightarrow p(x) \quad dx\text{-几乎处处},$$

且存在可积函数 $q \geq 0$, 使对每个 n ,

$$p_n(x) \leq q(x) \quad dx\text{-几乎处处},$$

则由支配收敛定理, p 是概率密度, 且 X_n 依分布收敛于密度为 p 的分布。

(c) 若 X_n 在有限集合 $\{2^{-n}, 2 \cdot 2^{-n}, \dots, 2^n \cdot 2^{-n}\}$ 上均匀分布, 则 X_n 依分布收敛于 $[0, 1]$ 上的均匀分布; 这就是黎曼和逼近积分。

(d) 若 $X_n \sim \mathcal{N}(0, \sigma_n^2)$ 且 $\sigma_n \rightarrow 0$, 则 X_n 依分布收敛于常数 0。

命题 10.3.1 若 X_n 依概率收敛于 X , 则 X_n 依分布收敛于 X 。

证明。若 $X_n \rightarrow X$ 几乎必然, 则对任意 $\varphi \in C_b(\mathbb{R}^d)$, 有 $\varphi(X_n) \rightarrow \varphi(X)$ 几乎必然; 由支配收敛定理, $E[\varphi(X_n)] \rightarrow E[\varphi(X)]$ 。一般地, 反设依分布收敛不成立, 则存在 $\varphi \in C_b(\mathbb{R}^d)$, $\varepsilon > 0$ 及子列 (n_k) , 使

$$|E[\varphi(X_{n_k})] - E[\varphi(X)]| \geq \varepsilon, \quad k \geq 1.$$

但依概率收敛可从 (X_{n_k}) 中再抽出几乎必然收敛于 X 的子列; 对该子列应用前一论证即得矛盾。

注。若极限 $X = a \in \mathbb{R}^d$ 几乎必然为常数, 则逆命题成立: 若 X_n 依分布收敛于 a , 由下面命题的开集不等式, 对每个 $\varepsilon > 0$,

$$\liminf_{n \rightarrow \infty} P_{X_n}(B(a, \varepsilon)) \geq 1$$

其中 $B(a, \varepsilon)$ 是以 a 为中心、半径为 ε 的开球；这正说明 X_n 依概率收敛于 a 。但一般的依分布收敛并不保证

$$P(X_n \in B) \xrightarrow{n \rightarrow \infty} P(X \in B).$$

对所有 Borel 集 $B \subset \mathbb{R}^d$ 都成立；例如在上例 (d) 中取 $B = \{0\}$ 即可看到。精确的判别如下。

命题 10.3.2 (Portmanteau 定理) 设 (μ_n) 与 μ 是 $\mathbb{R}^d$ 上的概率测度。下列四个条件等价。

(i) μ_n 弱收敛于 μ 。

(ii) 对每个开集 $G \subset \mathbb{R}^d$,

$$\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G).$$

(iii) 对每个闭集 $F \subset \mathbb{R}^d$,

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F).$$

(iv) 对每个满足 $\mu(\partial B) = 0$ 的 Borel 集 $B \subset \mathbb{R}^d$,

$$\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B).$$

证明。先证 (i) $\Rightarrow$ (ii)。若 $G \subset \mathbb{R}^d$ 为开集，则存在一列有界连续函数 (φ_p) ，使 $0 \leq \varphi_p \leq 1_G$ 且 $\varphi_p \uparrow 1_G$ 。例如可取 $\varphi_p(x) = p \operatorname{dist}(x, G^c) \wedge 1$ 。于是

$$\liminf_{n \rightarrow \infty} \mu_n(G) \geq \sup_p \left(\liminf_{n \rightarrow \infty} \int \varphi_p d\mu_n \right) = \sup_p \left(\int \varphi_p d\mu \right) = \mu(G).$$

等价性 (ii) $\iff$ (iii) 由取补集立即得到。下面由 (ii) 和 (iii) 推出 (iv)。任取 $B \in \mathcal{B}(\mathbb{R}^d)$ ，有

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu_n(B) &\leq \limsup_{n \rightarrow \infty} \mu_n(\overline{B}) \leq \mu(\overline{B}), \\ \liminf_{n \rightarrow \infty} \mu_n(B) &\geq \liminf_{n \rightarrow \infty} \mu_n(B^\circ) \geq \mu(B^\circ). \end{aligned}$$

若 $\mu(\partial B) = 0$ ，则 $\mu(\overline{B}) = \mu(B^\circ) = \mu(B)$ ，从而得到 (iv)。下面证明 (iv) $\implies$ (i)。任取 $\varphi \in C_b(\mathbb{R}^d)$ 。把 φ 分解成 $\varphi = \varphi^+ - \varphi^-$ ，可只考虑 $\varphi \geq 0$ 的情形。取 $K > 0$ ，使 $0 \leq \varphi \leq K$ 。由 Fubini–Tonelli 定理，

$$\int_{\mathbb{R}^d} \varphi(x) \mu(dx) = \int_{\mathbb{R}^d} \left(\int_0^K \mathbf{1}_{\{t \leq \varphi(x)\}} dt \right) \mu(dx) = \int_0^K \mu(E_t^\varphi) dt.$$

其中 $E_t^\varphi = \{x \in \mathbb{R}^d : \varphi(x) \geq t\}$ 。同样，对每个 n ，

$$\int_{\mathbb{R}^d} \varphi(x) \mu_n(dx) = \int_0^K \mu_n(E_t^\varphi) dt.$$

注意 $\partial E_t^\varphi \subset \{x \in \mathbb{R}^d : \varphi(x) = t\}$ ，并且满足

$$\mu(\{x \in \mathbb{R}^d : \varphi(x) = t\}) > 0$$

的 t 至多可数。事实上，对每个整数 $k \geq 1$ ，满足 $\mu(\{x : \varphi(x) = t\}) \geq 1/k$ 的不同 t 至多有 k 个。因此，由 (iv)，对 Lebesgue 几乎处处的 $t \in [0, K]$ ， $\mu_n(E_t^\varphi) \rightarrow \mu(E_t^\varphi)$ 。再由支配收敛定理，

$$\int_{\mathbb{R}^d} \varphi(x) \mu_n(dx) = \int_0^K \mu_n(E_t^\varphi) dt \xrightarrow{n \rightarrow \infty} \int_0^K \mu(E_t^\varphi) dt = \int_{\mathbb{R}^d} \varphi(x) \mu(dx).$$

¹ 推论 实随机变量序列 (X_n) 依分布收敛于实随机变量 X ，当且仅当其分布函数 $F_{X_n}(x)$ 在 F_X 的每个连续点 x 收敛于 $F_X(x)$ 。正向结论由命题 10.3.2 (iv) 立即得到。反过来，若上述分布函数收敛条件成立，则对任意 $x \in \mathbb{R}$ ，有

$$\liminf_{n \rightarrow \infty} F_{X_n}(x-) \geq F_X(x-), \quad \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

由此先对开区间验证命题 10.3.2 (ii)，其中 $\mu_n = P_{X_n}$ 、 $\mu = P_X$ 。再将任意实轴上的开集写成至多可数个两两不交开区间的并，即得所需结论。最后回顾： $C_c(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$ 上全体具有紧支集的连续函数。

命题 10.3.3 设 (μ_n) 与 μ 是 $\mathbb{R}^d$ 上的概率测度。设 $H \subset C_b(\mathbb{R}^d)$ ，并且 H 在上确界范数下的闭包包含 $C_c(\mathbb{R}^d)$ 。则以下三条性质等价：

- (i) (μ_n) 弱收敛于 μ ；
- (ii) 对每个 $\varphi \in C_c(\mathbb{R}^d)$ ，有

$$\forall \varphi \in C_c(\mathbb{R}^d), \quad \int \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi d\mu.$$

- (iii) 对每个 $\varphi \in H$ ，有

$$\forall \varphi \in H, \quad \int \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi d\mu.$$

证明。显然 (i) 蕴含 (ii) 和 (iii)。现设 (ii) 成立。取 $\varphi \in C_b(\mathbb{R}^d)$ ，再取 $(f_k) \subset C_c(\mathbb{R}^d)$ ，使 $0 \leq f_k \leq 1$ ，且当 $k \rightarrow \infty$ 时 $f_k \uparrow 1$ 。对每个 k ，都有 $\varphi f_k \in C_c(\mathbb{R}^d)$ ，因而

$$\int \varphi f_k d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi f_k d\mu.$$

此外，

$$\begin{aligned} \left| \int \varphi d\mu_n - \int \varphi f_k d\mu_n \right| &\leq \|\varphi\|_\infty \left(1 - \int f_k d\mu_n \right), \\ \left| \int \varphi d\mu - \int \varphi f_k d\mu \right| &\leq \|\varphi\|_\infty \left(1 - \int f_k d\mu \right). \end{aligned}$$

² 所以，对每个 $k \geq 1$ ，

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| &\leq \|\varphi\|_\infty \left[\limsup_{n \rightarrow \infty} \left(1 - \int f_k d\mu_n \right) + \left(1 - \int f_k d\mu \right) \right] \\ &= 2\|\varphi\|_\infty \left(1 - \int f_k d\mu \right). \end{aligned}$$

令 $k \rightarrow \infty$ ，由单调收敛定理，右端趋于 0；于是 $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$ ，故 (i) 成立。最后证明 (iii) 蕴含 (ii)。设 (iii) 成立。任取 $\varphi \in C_c(\mathbb{R}^d)$ 。由闭包假设，对每个整数 $k \geq 1$ ，可取 $\varphi_k \in H$ ，使 $\|\varphi - \varphi_k\|_\infty \leq 1/k$ 。于是

¹校订注：原讲义此式在箭头右侧仍误排为 $\mu_n(E_t^\varphi)$ ；支配收敛定理给出的极限被积函数应为 $\mu(E_t^\varphi)$ 。

²校订注：原讲义在这两处上确界的下标误排为 $x \in \mathbb{R}$ ；本命题讨论 $\mathbb{R}^d$ 上的函数，应为 $x \in \mathbb{R}^d$ 。

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| &\leq \limsup_{n \rightarrow \infty} \left| \int (\varphi - \varphi_k) d\mu_n \right| \\
&\quad + \limsup_{n \rightarrow \infty} \left| \int \varphi_k d\mu_n - \int \varphi_k d\mu \right| + \left| \int (\varphi_k - \varphi) d\mu \right| \\
&\leq \frac{1}{k} + 0 + \frac{1}{k} = \frac{2}{k}.
\end{aligned}$$

由于 k 任意, 令 $k \rightarrow \infty$ 即得 $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$, 故 (ii) 成立。

定理 10.3.4 (Lévy 连续性定理) $\mathbb{R}^d$ 上的概率测度列 (μ_n) 弱收敛于概率测度 μ , 当且仅当

$$\forall \xi \in \mathbb{R}^d, \quad \hat{\mu}_n(\xi) \xrightarrow{n \rightarrow \infty} \hat{\mu}(\xi).$$

等价地, $\mathbb{R}^d$ -值随机变量列 (X_n) 依分布收敛于 X , 当且仅当

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_{X_n}(\xi) \xrightarrow{n \rightarrow \infty} \Phi_X(\xi).$$

证明。 只需证明测度版本。若 μ_n 弱收敛于 μ , 则对每个 $\xi \in \mathbb{R}^d$, 函数 $x \mapsto e^{i\xi \cdot x}$ 连续有界, 所以由弱收敛的定义,

$$\hat{\mu}_n(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot x} \mu_n(dx) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}^d} e^{i\xi \cdot x} \mu(dx) = \hat{\mu}(\xi).$$

反过来, 假设对每个 $\xi \in \mathbb{R}^d$, 都有 $\hat{\mu}_n(\xi) \rightarrow \hat{\mu}(\xi)$ 。要证明 (μ_n) 弱收敛于 μ 。为简化记号, 只处理 $d = 1$; 一般情形相同。取 $f \in C_c(\mathbb{R})$, 并对每个 $\sigma > 0$ 令

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

第 8 章末已说明, 当 $\sigma \downarrow 0$ 时, $g_\sigma * f \rightarrow f$ 逐点成立; 对紧支连续函数, 事实上收敛在 $\mathbb{R}$ 上一致。另一方面, 若 ν 是 $\mathbb{R}$ 上的概率测度, 第 8 章傅里叶变换单射性定理的证明给出

$$\begin{aligned}
\int_{\mathbb{R}} (g_\sigma * f) d\nu &= \int_{\mathbb{R}} f(x) (g_\sigma * \nu)(x) dx \\
&= \int_{\mathbb{R}} f(x) \left[\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \hat{\nu}(-\xi) d\xi \right] dx.
\end{aligned}$$

由于 $\hat{\mu}_n(\xi) \rightarrow \hat{\mu}(\xi)$, 且概率测度的特征函数模不超过 1, 对固定的 x 和 $\sigma > 0$, 支配收敛定理给出

$$\int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \hat{\mu}_n(-\xi) d\xi \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \hat{\mu}(-\xi) d\xi.$$

上式两边的内层积分绝对值一致有界。再利用前一表示式, 并对 x 应用一次支配收敛定理, 得到

$$\int_{\mathbb{R}} (g_\sigma * f) d\mu_n \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} (g_\sigma * f) d\mu.$$

最后, 令

$$H = \{ \varphi = g_\sigma * f : \sigma > 0, f \in C_c(\mathbb{R}) \} \subset C_b(\mathbb{R}).$$

由于对每个 $f \in C_c(\mathbb{R})$, 都有 $\|g_\sigma * f - f\|_\infty \rightarrow 0$, 故 H 的闭包包含 $C_c(\mathbb{R})$ 。而上面已经证明, 对所有 $\varphi \in H$, $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$ 。由命题 10.3.3, (μ_n) 弱收敛于 μ 。

10.4 两个应用

10.4.1 经验测度的收敛

设 $(X_n)_{n \geq 1}$ 为一列取值于 $\mathbb{R}^d$ 的独立同分布随机变量。它们可表示同一随机试验的独立重复结果。统计学中的一个基本问题是：给定同一个样本点 ω 所产生的观测 $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ ，如何估计 X_1 的分布。

例（抽样调查）。 设总体由编号为 $1, 2, \dots, N$ 的 N 个个体组成，其中 N 很大。给每个个体 i 赋予参数 $a(i) \in \mathbb{R}^d$ ，例如年龄、投票意向和月收入等。对任意 $A \in \mathcal{B}(\mathbb{R}^d)$ ，我们关注总体比例

$$\mu(A) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_A(a(i)).$$

它等于总体中满足 $a(i) \in A$ 的个体比例（原讲义举例为：年龄超过五十岁、准备投票给 Chirac 且月收入超过 2000 欧元者的比例）。当 N 很大时，直接计算 $\mu(A)$ 不现实。抽样调查随机选取 n 个个体，其中 n 较大但远小于 N ，并用样本比例估计总体比例。为精确表述，取相互独立、均在 $\{1, \dots, N\}$ 上服从均匀分布的随机变量 $Y_1, \dots, Y_n$ ，它们代表入样个体；令 $X_j = a(Y_j)$ 。于是 $X_1, \dots, X_n$ 独立同分布，而且

$$P_{X_1}(A) = P(a(Y_1) \in A) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_A(a(i)) = \mu(A).$$

另一方面，样本中满足参数落入 A 的比例为

$$\frac{1}{n} \sum_{j=1}^n \mathbf{1}_A(X_j(\omega)) = \frac{1}{n} \sum_{j=1}^n \delta_{X_j(\omega)}(A).$$

因此，要判断样本比例是否接近总体真实比例 $\mu(A)$ ，等价于研究经验测度

$$\mu_{n,\omega} := \frac{1}{n} \sum_{j=1}^n \delta_{X_j(\omega)}$$

当 $n \rightarrow \infty$ 时是否接近真实分布 P_{X_1} 。下一定理给出肯定答案。

定理 10.4.1 设 $(X_n)_{n \geq 1}$ 是一列取值于 $\mathbb{R}^d$ 的独立同分布随机变量。对每个 $\omega \in \Omega$ 及 $n \geq 1$ ，定义 $\mathbb{R}^d$ 上的经验概率测度

$$\mu_{n,\omega} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}.$$

则对几乎所有 ω ，

$$\mu_{n,\omega} \xrightarrow[n \rightarrow \infty]{w} P_{X_1}.$$

注。 实际应用还需要收敛速度的估计：在抽样调查中，希望当 n 远小于总体规模 N 时，经验测度已经足够接近 P_{X_1} 。原讲义给出的数量级示例是 $N \approx 10^7$ 、 $n \approx 10^3$ 。

证明。 取 $C_c(\mathbb{R}^d)$ 在上确界范数下的一个可数稠密子集 H 。对每个 $\varphi \in H$ ，把强大数律应用于独立同分布随机变量 $\varphi(X_i)$ ，得到

$$\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}[\varphi(X_1)].$$

按经验测度的定义，上式等价于

$$\int \varphi d\mu_{n,\omega} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \int \varphi dP_{X_1}.$$

由于 H 可数，可以舍去一个零概率集合，使得余下每个 ω 同时对所有 $\varphi \in H$ 满足

$$\forall \varphi \in H, \quad \int \varphi d\mu_{n,\omega} \xrightarrow[n \rightarrow \infty]{} \int \varphi dP_{X_1}.$$

由命题 10.3.3，这说明对几乎所有 ω ，经验测度 $\mu_{n,\omega}$ 弱收敛于 P_{X_1} 。

10.4.2 中心极限定理

设 $(X_n)_{n \geq 1}$ 为一列属于 L^1 的独立同分布实随机变量。强大数定律给出

$$\frac{1}{n} (X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} \mathbb{E}[X_1].$$

然后我们要研究这种收敛发生的速度，也就是说差值的数量级

$$\frac{1}{n} (X_1 + \cdots + X_n) - \mathbb{E}[X_1]$$

当 n 很大时会发生什么？若再假设随机变量 $X_i \in L^2$ ，就可以像证明大数弱定律时那样，通过下面的计算猜出正确的归一化尺度：

$$\mathbb{E}[(X_1 + \cdots + X_n - n\mathbb{E}[X_1])^2] = \text{Var}(X_1 + \cdots + X_n) = n \text{Var}(X_1).$$

这个计算表明， $(X_1 + \cdots + X_n - n\mathbb{E}[X_1])^2$ 的期望随 n 线性增长。因此， $X_1 + \cdots + X_n - n\mathbb{E}[X_1]$ 的典型数量级应为 $\sqrt{n}$ ；等价地，样本均值 $n^{-1}(X_1 + \cdots + X_n) - \mathbb{E}[X_1]$ 的典型数量级应为 $n^{-1/2}$ 。中心极限定理把这一判断精确化。

定理 10.4.2 (中心极限定理) 设 $(X_n)_{n \geq 1}$ 是一列相互独立、同分布且属于 L^2 的实随机变量，并令 $\sigma^2 = \text{Var}(X_1)$ 。则

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{law}} \mathcal{N}(0, \sigma^2).$$

其中 $\mathcal{N}(0, \sigma^2)$ 表示均值为 0、方差为 σ^2 的高斯分布。若 $\sigma > 0$ ，则等价地，对任意 $a, b \in \mathbb{R}$ 且 $a < b$ ，^a

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_1 + \cdots + X_n \in [n\mathbb{E}[X_1] + a\sqrt{n}, n\mathbb{E}[X_1] + b\sqrt{n}]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

^a校订注：原讲义没有在下方的密度积分公式前注明 $\sigma > 0$ 。当 $\sigma = 0$ 时，极限分布是退化分布 $\delta_0 = \mathcal{N}(0, 0)$ ，而含 $1/\sigma$ 的密度公式没有意义；中心极限定理本身仍然成立。

证明。第二种表述由第一种表述以及用分布函数刻画依分布收敛的结论立即得到；这里用到了极限高斯分布的分布函数连续这一事实。下面证明第一种表述。把 X_n 替换成 $X_n - \mathbb{E}[X_n]$ ，不妨设 $\mathbb{E}[X_1] = 0$ 。令

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Z_n 的特征函数为

$$\Phi_{Z_n}(\xi) = \mathbb{E}\left[\exp\left(i\xi \frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right] = \left(\mathbb{E}\left[e^{i(\xi/\sqrt{n})X_1}\right]\right)^n = \Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right)^n.$$

第二个等号使用了 X_i 相互独立且同分布这一事实。由第 8 章的结果，当 $\xi \rightarrow 0$ 时，

$$\Phi_{X_1}(\xi) = 1 + i\xi\mathbb{E}[X_1] - \frac{1}{2}\xi^2\mathbb{E}[X_1^2] + o(\xi^2) = 1 - \frac{\sigma^2\xi^2}{2} + o(\xi^2).$$

因此，对每个固定的 $\xi \in \mathbb{R}$ ，当 $n \rightarrow \infty$ 时也有

$$\Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{\sigma^2\xi^2}{2n} + o\left(\frac{1}{n}\right).$$

结合上式可知，对任意 $\xi \in \mathbb{R}$ ，

$$\lim_{n \rightarrow \infty} \Phi_{Z_n}(\xi) = \lim_{n \rightarrow \infty} \left(1 - \frac{\sigma^2\xi^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = \exp\left(-\frac{\sigma^2\xi^2}{2}\right) = \Phi_U(\xi).$$

其中 $U \sim \mathcal{N}(0, \sigma^2)$ 。由 Lévy 连续性定理， Z_n 依分布收敛到 U ，定理得证。

特殊情形：de Moivre 定理。设 (X_n) 是一列相互独立、参数为 $1/2$ 的 Bernoulli 随机变量，即 $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 0) = 1/2$ 。那么 $S_n = X_1 + \cdots + X_n$ 服从二项分布 $\mathcal{B}(n, 1/2)$ ：

$$\mathbb{P}(S_n = k) = \binom{n}{k} 2^{-n}.$$

在此情形下 $\sigma^2 = 1/4$ ，因而对任意 $a < b$ ，中心极限定理给出

$$2^{-n} \sum_{\frac{n}{2} + a\sqrt{n} \leq k \leq \frac{n}{2} + b\sqrt{n}} \binom{n}{k} \xrightarrow{n \rightarrow \infty} \sqrt{\frac{2}{\pi}} \int_a^b e^{-2x^2} dx.$$

利用 Stirling 公式可以直接验证上述收敛，尽管计算稍显繁琐。事实上还可证明如下更精确的估计：

$$\sqrt{n} 2^{-n} \binom{n}{k} = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{2}{n} \left(k - \frac{n}{2}\right)^2\right) + o(1).$$

其中余项 $o(1)$ 对所有 $k \in \{0, 1, \dots, n\}$ 一致。

10.4.3 向量情形的推广

现在设 $(X_n)_{n \geq 1}$ 是一列取值于 $\mathbb{R}^d$ 、相互独立、同分布且可积的随机变量。逐坐标应用大数强定律可得

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}[X_1].$$

这里, 若写作 $X_1 = (X_1^1, \dots, X_1^d)$, 则极限 $\mathbb{E}[X_1]$ 就是向量 $(\mathbb{E}[X_1^1], \dots, \mathbb{E}[X_1^d])$ 。进一步假设 X_n 平方可积。多维中心极限定理不能仅靠逐坐标应用一维结论得到: 与几乎必然收敛不同, 一个 $\mathbb{R}^d$ 值随机变量列的每个坐标分别依分布收敛, 并不足以推出向量本身依分布收敛; 而且极限的联合分布也不由其各个边际分布唯一确定。为了把中心极限定理推广到 $\mathbb{R}^d$ 值随机变量, 首先需要推广高斯分布的概念。

定义 10.4.1 设 C 是实对称半正定 $d \times d$ 矩阵。平方可积的 $\mathbb{R}^d$ -值随机变量 X 称为协方差矩阵为 C 的中心高斯向量, 如果

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_X(\xi) = \mathbb{E}[e^{i\xi \cdot X}] = \exp\left(-\frac{1}{2}\xi^\top C \xi\right).$$

也称 X 服从 $\mathcal{N}(0, C)$ 分布。

注。 若 $a \in \mathbb{R}^d$, 且 $X - a \sim \mathcal{N}(0, C)$, 则称 $X \sim \mathcal{N}(a, C)$ 。

第 8 章已经证明: 若平方可积随机向量 $X = (X^1, \dots, X^d)$ 取值于 $\mathbb{R}^d$, 则当 $\xi \rightarrow 0$ 时, 其特征函数有二阶展开

$$\Phi_X(\xi) = 1 + i \sum_{j=1}^d \xi_j \mathbb{E}[X^j] - \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \xi_j \xi_k \mathbb{E}[X^j X^k] + o(|\xi|^2).$$

因此, 若 $X \sim \mathcal{N}(0, C)$, 比较一阶项与二阶项可得 $\mathbb{E}[X] = 0$ 以及

$$K_X = C.$$

命题 10.4.3 任给实对称半正定矩阵 C , 存在协方差矩阵为 C 的中心高斯向量。

证明。 首先回忆一下 (见第 9 章末尾), 独立高斯随机变量的线性组合仍然是高斯随机变量。

令 $A = C^{1/2}$, 则 A 是对称半正定矩阵且 $A^2 = C$ 。再取相互独立、均服从 $\mathcal{N}(0, 1)$ 的实随机变量 $Y^1, \dots, Y^d$, 令 $Y = (Y^1, \dots, Y^d)^\top$, 并置 $X = AY$ 。下面证明 $X \sim \mathcal{N}(0, C)$ 。对任意 $\xi \in \mathbb{R}^d$, $\xi \cdot X$ 是 $Y^1, \dots, Y^d$ 的线性组合, 因而是中心高斯随机变量。更精确地, $\xi \cdot X \sim \mathcal{N}(0, \sigma^2)$, 其中

$$\begin{aligned} \sigma^2 &= \mathbb{E}[(\xi \cdot X)^2] = \mathbb{E}[(\xi^\top AY)(Y^\top A\xi)] \\ &= \xi^\top A \mathbb{E}[YY^\top] A\xi = \xi^\top A^2 \xi = \xi^\top C \xi. \end{aligned}$$

这里使用矩阵运算, 以及 $\mathbb{E}[YY^\top] = I_d$; 后者来自 Y 的各坐标相互独立且均服从 $\mathcal{N}(0, 1)$ 。最后, 由一维高斯分布 $\mathcal{N}(0, \sigma^2)$ 的特征函数公式, 对每个 $u \in \mathbb{R}$, 有

$$\mathbb{E}[e^{iu\xi \cdot X}] = \exp\left(-\frac{\sigma^2 u^2}{2}\right) = \exp\left(-\frac{u^2}{2}\xi^\top C \xi\right).$$

取 $u = 1$ 即得所需结论。**注。** (i) 沿用上述证明的记号, Y 服从 $\mathcal{N}(0, I_d)$ 。(ii) 取值于 $\mathbb{R}^d$ 的随机变量 X 是中心高斯向量, 当且仅当它的各分量的每个线性组合都是中心高斯随机变量。事实上, 此时

$$\Phi_X(\xi) = \mathbb{E}[e^{i\xi \cdot X}] = \exp\left(-\frac{1}{2}\mathbb{E}[(\xi \cdot X)^2]\right) = \exp\left(-\frac{1}{2}\xi^\top K_X \xi\right).$$

练习。 设 X 为中心高斯向量。证明： X 具有关于 Lebesgue 测度的密度，当且仅当协方差矩阵 K_X 非退化；并在非退化时求出该密度。

定理 10.4.4 (多维中心极限定理) 设 $(X_n)_{n \geq 1}$ 是一列取值于 $\mathbb{R}^d$ 的独立同分布平方可积随机向量。则

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - nE[X_1]) \xrightarrow[n \rightarrow \infty]{\text{law}} \mathcal{N}(0, K_{X_1}).$$

证明。 论证与一维情形相同。中心化后可设 $E[X_1] = 0$ 。对任意 $\xi \in \mathbb{R}^d$ ，由独立性，

$$\mathbb{E} \left[\exp \left(i\xi \cdot \frac{X_1 + \cdots + X_n}{\sqrt{n}} \right) \right] = \left(\mathbb{E} \left[e^{i(\xi/\sqrt{n}) \cdot X_1} \right] \right)^n = \Phi_{X_1} \left(\frac{\xi}{\sqrt{n}} \right)^n.$$

另一方面，特征函数在原点的二阶展开给出

$$\Phi_{X_1} \left(\frac{\xi}{\sqrt{n}} \right) = 1 - \frac{1}{2n} \xi^\top K_{X_1} \xi + o\left(\frac{1}{n}\right).$$

于是

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\exp \left(i\xi \cdot \frac{X_1 + \cdots + X_n}{\sqrt{n}} \right) \right] = \exp \left(-\frac{1}{2} \xi^\top K_{X_1} \xi \right).$$

右端是 $\mathcal{N}(0, K_{X_1})$ 的特征函数；由 Lévy 连续性定理即得结论。

第十一章 条件期望

本章构造条件期望并研究其性质。直观地说，给定子 σ -代数 $\mathcal{B}$ 后， $\mathbb{E}[X | \mathcal{B}]$ 是 $\mathcal{B}$ -可测、并在适当意义下最接近 X 的随机变量。在预测和不完全观测等问题中，我们只掌握部分信息，却要估计随机变量；条件期望正是描述这种估计的基本工具。本章先从离散情形提炼其特征性质，再给出一般定义，并讨论若干可显式计算的例子，尤其是高斯情形；最后介绍条件分布。

11.1 离散条件化

与前几章一样，固定概率空间 $(\Omega, \mathcal{A}, P)$ 。设 $B \in \mathcal{A}$ 且 $P(B) > 0$ 。定义给定 B 时的条件概率测度 $P(\cdot | B)$ ：对每个 $A \in \mathcal{A}$,

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

同样，若 $X \geq 0$ 或 $X \in L^1(\Omega, \mathcal{A}, P)$ ，定义给定事件 B 时 X 的条件期望为

$$E[X | B] = \frac{E[X \mathbf{1}_B]}{P(B)}.$$

这个量也是在概率 $P(\cdot | B)$ 下 X 的期望，它可以解释为当 B 发生时 X 的平均值。

下面把“已知事件”推广为“已知随机变量”。先设随机变量 Y 取值于可数集合 E 。

令 $E' = \{y \in E : P(Y = y) > 0\}$ 。对每个 $y \in E'$ 以及任意 $X \in L^1(\Omega, \mathcal{A}, P)$ ，可以把前面的定义特化为

$$\mathbb{E}[X | Y = y] = \frac{\mathbb{E}[X \mathbf{1}_{\{Y=y\}}]}{P(Y = y)}.$$

定义 11.1.1 设 $X \in L^1(\Omega, \mathcal{A}, P)$ 。给定离散随机变量 Y 时， X 的条件期望定义为

$$\mathbb{E}[X | Y] = \varphi(Y),$$

其中 $\varphi : E \rightarrow \mathbb{R}$ 定义为

$$\varphi(y) = \begin{cases} \mathbb{E}[X | Y = y], & y \in E', \\ 0, & y \in E \setminus E'. \end{cases}$$

注。 φ 在 $E \setminus E'$ 上的取值可以任意指定；因为

$$P(Y \in E \setminus E') = \sum_{y \in E \setminus E'} P(Y = y) = 0.$$

所以改变 φ 在 $E \setminus E'$ 上的定义，只会在零概率集合上改变 $\mathbb{E}[X | Y]$ 。一般的条件期望也总只在几乎必然相等的意义下唯一。与给定单个事件时得到一个数不同， $\mathbb{E}[X | Y]$ 是随机变量：几乎必然地，

$$\mathbb{E}[X | Y](\omega) = \mathbb{E}[X | Y = y] \quad \text{若 } Y(\omega) = y.$$

此外, $\mathbb{E}[X | Y] = \varphi(Y)$ 是 Y 的函数, 因而关于 $\sigma(Y)$ 可测。稍后会说明, 它是用 Y 的函数逼近 X 的最佳方式。

例 (掷骰子)。取 $\Omega = \{1, 2, \dots, 6\}$, 并令 $P(\{\omega\}) = 1/6$ 。定义

$$Y(\omega) = \begin{cases} 1, & \omega \text{ 为奇数,} \\ 0, & \omega \text{ 为偶数,} \end{cases} \quad X(\omega) = \omega.$$

于是

$$\mathbb{E}[X | Y](\omega) = \begin{cases} 3, & \omega \in \{1, 3, 5\}, \\ 4, & \omega \in \{2, 4, 6\}. \end{cases}$$

也就是说, 已知骰子点数的奇偶性后, 条件期望就是相应三个可能点数的平均值。

命题 11.1.1 有 $E[|E[X | Y]|] \leq E[|X|]$ 。特别地, $E[X | Y] \in L^1(\Omega, \mathcal{A}, P)$ 。此外, 对任意有界的 $\sigma(Y)$ -可测随机变量 Z ,

$$E[ZX] = E[Z E[X | Y]].$$

证明。由定义,

$$\begin{aligned} E[|E[X | Y]|] &= \sum_{y \in E'} P(Y = y) \frac{|\mathbb{E}[X \mathbf{1}_{\{Y=y\}}]|}{P(Y = y)} \\ &\leq \sum_{y \in E} \mathbb{E}[|X| \mathbf{1}_{\{Y=y\}}] = E[|X|]. \end{aligned}$$

为证明最后一个断言, 注意任意有界的 $\sigma(Y)$ -可测随机变量都可写成 $Z = \psi(Y)$, 其中 ψ 有界。于是

$$\begin{aligned} \mathbb{E}[\psi(Y) \mathbb{E}[X | Y]] &= \sum_{y \in E} \psi(y) \mathbb{E}[X \mathbf{1}_{\{Y=y\}}] \\ &= \sum_{y \in E} \mathbb{E}[\psi(Y) X \mathbf{1}_{\{Y=y\}}] = \mathbb{E}[\psi(Y) X]. \end{aligned}$$

推论。若 Y' 是另一个离散随机变量, 且 $\sigma(Y) = \sigma(Y')$, 则

$$\mathbb{E}[X | Y] = \mathbb{E}[X | Y'] \quad \text{几乎必然.}$$

事实上, 在前一命题中取 $Z = \mathbf{1}_{\{E[X|Y] > E[X|Y']\}}$ 。由于 $\sigma(Y) = \sigma(Y')$, 且 $E[X | Y]$ 、 $E[X | Y']$ 都对这个 σ -代数可测, 所以 Z 确实可测。于是得到

$$\mathbb{E}[\mathbf{1}_{\{E[X|Y] > E[X|Y']\}} (\mathbb{E}[X | Y] - \mathbb{E}[X | Y'])] = 0.$$

从而 $E[X | Y] \leq E[X | Y']$ 几乎必然成立; 交换 Y 与 Y' 可得反向不等式。这个论证还说明: 命题最后给出的积分恒等式, 在所有 $\sigma(Y)$ -可测且可积的随机变量中唯一刻画 $E[X | Y]$ 。这提示我们, 合适的条件化对象是一个 σ -代数。下面将以上述刻画性质为基础发展这一概念。

11.2 条件期望的定义

11.2.1 可积随机变量的情形

定理与定义 11.2.1 设 $\mathcal{B} \subset \mathcal{A}$ 是子 σ -代数, 且 $X \in L^1(\Omega, \mathcal{A}, P)$ 。存在唯一的随机变量 $E[X | \mathcal{B}] \in L^1(\Omega, \mathcal{B}, P)$, 使

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X \mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X | \mathcal{B}] \mathbf{1}_B]. \quad (11.1)$$

更一般地, 对任意有界的 $\mathcal{B}$ -可测随机变量 Z , 有

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X | \mathcal{B}]Z]. \quad (11.2)$$

若 $X \geq 0$, 则 $\mathbb{E}[X | \mathcal{B}] \geq 0$ 几乎必然。

关键要求是 $\mathbb{E}[X | \mathcal{B}]$ 关于 $\mathcal{B}$ 可测。在 $L^1(\Omega, \mathcal{B}, P)$ 中, 式 (11.1) 或式 (11.2) 中的任一条都唯一刻画条件期望。

这两个等式都可作为条件期望在 $L^1(\Omega, \mathcal{B}, P)$ 中的特征性质。若 $\mathcal{B} = \sigma(Y)$, 则不加区别地写

$$E[X | \mathcal{B}] = E[X | \sigma(Y)] = E[X | Y].$$

这种表示与前一部分处理中离散情况是一致的: 比较 (11.2) 和上述命题。

证明。先证唯一性。设 $X', X'' \in L^1(\Omega, \mathcal{B}, P)$ 都满足条件期望的特征性质。取事件 $\{X' > X''\} \in \mathcal{B}$, 得到

$$\forall B \in \mathcal{B}, \quad E[X' \mathbf{1}_B] = E[X \mathbf{1}_B] = E[X'' \mathbf{1}_B].$$

取 $B = \{X' > X''\}$ (它确实是 $\mathcal{B}$ 可测的, 因为 X' 和 X'' 可测), 得到

$$E[(X' - X'') \mathbf{1}_{\{X' > X''\}}] = 0$$

所以 $X' \leq X''$ 几乎必然; 交换二者即得相等。再证存在性。先设 $X \geq 0$, 并在 $(\Omega, \mathcal{B})$ 上定义有限测度

$$\forall B \in \mathcal{B}, \quad Q(B) = E[X \mathbf{1}_B].$$

把 P 也看成 $(\Omega, \mathcal{B})$ 上的概率测度, 则显然 $Q \ll P$ 。在该可测空间上应用 Radon-Nikodym 定理, 得到非负的 $\mathcal{B}$ -可测随机变量 X' , 使

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X \mathbf{1}_B] = Q(B) = \mathbb{E}[X' \mathbf{1}_B].$$

取 $B = \Omega$, 得 $E[X'] = E[X] < \infty$, 所以 $X' \in L^1(\Omega, \mathcal{B}, P)$, 并且 X' 满足定理中的性质。一般的实值 X 只需写成 $X = X^+ - X^-$, 并分别应用上述构造。

$$E[X | \mathcal{B}] = E[X^+ | \mathcal{B}] - E[X^- | \mathcal{B}].$$

最后, 从 (11.1) 推出 (11.2), 只需用简单函数逐点逼近可测函数。**例。**取 $\Omega = (0, 1]$ 、 $\mathcal{A} = \mathcal{B}((0, 1])$ 、 $P(d\omega) = d\omega$ 。固定 $n \geq 1$, 令 $\mathcal{B}$ 为由区间 $((i-1)/n, i/n]$ ($i = 1, \dots, n$) 生成的 σ -代数。若 $f \in L^1(\Omega, \mathcal{A}, P)$, 则

$$E[f | \mathcal{B}] = \sum_{i=1}^n f_i 1_{\left[\frac{i-1}{n}, \frac{i}{n}\right]},$$

其中 $f_i = n \int_{(i-1)/n}^{i/n} f(\omega) d\omega$, 即 f 在区间 $((i-1)/n, i/n]$ 上的平均值。条件期望具有下列基本性质。

条件期望的性质。

(a) 若 X 是 $\mathcal{B}$ -可测的, 则 $E[X | \mathcal{B}] = X$ 。

(b) 映射 $X \mapsto E[X | \mathcal{B}]$ 是线性的。

$$(c) E[E[X | \mathcal{B}]] = E[X].$$

(d) $|E[X | \mathcal{B}]| \leq E[|X| | \mathcal{B}]$ 几乎必然; 因此 $E[|E[X | \mathcal{B}]|] \leq E[|X|]$ 。

(e) 若 $X \geq X'$, 则 $E[X | \mathcal{B}] \geq E[X' | \mathcal{B}]$ 几乎必然。

证明。(a) 由上一定理中的唯一性立即得到。(b) 同样如此: 若 $X, X' \in L^1$ 且 $\alpha, \alpha' \in \mathbb{R}$, 则随机变量

$$\alpha E[X | \mathcal{B}] + \alpha' E[X' | \mathcal{B}]$$

满足 $\alpha X + \alpha' X'$ 的特征性质 (11.1)。(c) 是 (11.1) 中取 $B = \Omega$ 的特殊情形。对 (d), 回忆若 $X \geq 0$, 则 $E[X | \mathcal{B}] \geq 0$ 。因此

$$\begin{aligned} |E[X | \mathcal{B}]| &= |E[X^+ | \mathcal{B}] - E[X^- | \mathcal{B}]| \\ &\leq E[X^+ | \mathcal{B}] + E[X^- | \mathcal{B}] = E[|X| | \mathcal{B}]. \end{aligned}$$

最后, (e) 由线性性立即得到。

11.2.2 非负随机变量的情形

定理 11.2.2 设 X 是取值于 $[0, \infty]$ 的随机变量。公式

$$E[X | \mathcal{B}] = \lim_{n \rightarrow \infty} \uparrow E[X \wedge n | \mathcal{B}] \quad \text{几乎必然}$$

定义了一个取值于 $[0, \infty]$ 的随机变量。它由下列性质在相差零概率集的意义下唯一确定: 对每个非负的 $\mathcal{B}$ -可测随机变量 Z ,

$$E[XZ] = E[E[X | \mathcal{B}]Z]. \quad (11.3)$$

若 X 还可积, 则比较上式与 (11.1) 可知, 这一定义与上一小节中的 $E[X | \mathcal{B}]$ 完全一致。与可积情形相同, 式 (11.3) 称为条件期望的特征性质。

证明。定义中的极限单调递增, 这是上面的性质 (e) 的直接结果。若 Z 为非负的 $\mathcal{B}$ -可测随机变量, 则由单调收敛定理,

$$\begin{aligned} E[E[X | \mathcal{B}]Z] &= \lim_{n \rightarrow \infty} E[E[X \wedge n | \mathcal{B}](Z \wedge n)] \\ &= \lim_{n \rightarrow \infty} E[(X \wedge n)(Z \wedge n)] = E[XZ]. \end{aligned}$$

还需证明唯一性。设 X' 与 X'' 是两个取值于 $[0, \infty]$ 的 $\mathcal{B}$ -可测随机变量，并且

$$\mathbb{E}[X'Z] = \mathbb{E}[X''Z]$$

对每个非负的 $\mathcal{B}$ -可测随机变量 Z 都成立。固定有理数 $a, b \in \mathbb{Q}_+$ 且 $a < b$ ，取

$$Z = 1_{\{X' \leq a < b \leq X''\}}.$$

于是

$$a\mathbb{P}(X' \leq a < b \leq X'') \geq b\mathbb{P}(X' \leq a < b \leq X'').$$

这只有在 $\mathbb{P}(X' \leq a < b \leq X'') = 0$ 时才可能。因此

$$\mathbb{P}\left(\bigcup_{\substack{a, b \in \mathbb{Q}_+ \\ a < b}} \{X' \leq a < b \leq X''\}\right) = 0.$$

从而 $X' \geq X''$ 几乎必然。交换 X' 与 X'' ，同理得到 $X'' \geq X'$ 几乎必然，唯一性得证。

注。可能同时有 $X < \infty$ 几乎必然，以及 $\mathbb{P}(\mathbb{E}[X | \mathcal{B}] = \infty) > 0$ 。例如，当 $\mathcal{B} = \{\emptyset, \Omega\}$ 时， $\mathbb{E}[X | \mathcal{B}] = \mathbb{E}[X]$ ，而一个几乎必然有限的非负随机变量完全可能有无穷期望。再给出一个较不平凡的例子。仍取 $\Omega = (0, 1]$ ， $P(d\omega) = d\omega$ ，并令 $\mathcal{B}$ 由区间 $((i-1)/n, i/n]$ ($1 \leq i \leq n$) 生成。若 $X(\omega) = 1/\omega$ ，则

$$\mathbb{E}[X | \mathcal{B}] = \infty 1_{(0, 1/n]} + \sum_{i=2}^n n \log\left(\frac{i}{i-1}\right) 1_{((i-1)/n, i/n]}.$$

非负随机变量的条件期望还具有下列性质。

(a) 若 X, X' 非负且 $a, b \geq 0$ ，则

$$\mathbb{E}[aX + bX' | \mathcal{B}] = a\mathbb{E}[X | \mathcal{B}] + b\mathbb{E}[X' | \mathcal{B}].$$

(b) 若 X 是 $\mathcal{B}$ -可测的，则 $\mathbb{E}[X | \mathcal{B}] = X$ 。

(c) 若 (X_n) 是递增的非负随机变量列，且 $X = \lim_{n \rightarrow \infty} \uparrow X_n$ ，则

$$\mathbb{E}[X | \mathcal{B}] = \lim_{n \rightarrow \infty} \uparrow \mathbb{E}[X_n | \mathcal{B}] \quad \text{几乎必然.}$$

(d) 若 (X_n) 是一列非负随机变量，则

$$\mathbb{E}[\liminf_{n \rightarrow \infty} X_n | \mathcal{B}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}] \quad \text{几乎必然.}$$

(e) 设可积随机变量列 (X_n) 几乎必然收敛到 X 。若存在非负随机变量 Z ，使得对每个 n 都有 $|X_n| \leq Z$ 几乎必然，且 $\mathbb{E}[Z] < \infty$ ，则

$$\mathbb{E}[X | \mathcal{B}] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}] \quad \text{几乎必然且在 } L^1 \text{ 中.}$$

(f) 若 f 是非负凸函数，且 $X \in L^1$ ，则

$$\mathbb{E}[f(X) | \mathcal{B}] \geq f(\mathbb{E}[X | \mathcal{B}]).$$

注。严格说来，每个涉及条件期望的等式或不等式都应注明“几乎必然”，因为条件期望只在相差零概率集的意义下定义。通常省略这一说明，如上面的 (a)、(b) 和 (f)。

证明。(a)、(b) 由条件期望的特征性质直接得到。(c) 由 (a)，若 $X_1 \geq X_2 \geq 0$ ，则 $E[X_1 | \mathcal{B}] \geq E[X_2 | \mathcal{B}]$ 。在 (c) 的条件下，可定义 $X' = \lim_{n \rightarrow \infty} \uparrow E[X_n | \mathcal{B}]$ 。它是取值于 $[0, \infty]$ 的 $\mathcal{B}$ -可测随机变量。对每个非负的 $\mathcal{B}$ -可测随机变量 Z ，有

$$E[ZX'] = \lim_{n \rightarrow \infty} \uparrow E[ZE[X_n | \mathcal{B}]] = \lim_{n \rightarrow \infty} \uparrow E[ZX_n] = E[ZX].$$

由定理的特征性质， $X' = E[X | \mathcal{B}]$ 。(d) 利用 (c)，可写成

$$\begin{aligned} E[\liminf_{n \rightarrow \infty} X_n | \mathcal{B}] &= E\left[\lim_{k \rightarrow \infty} \uparrow \inf_{n \geq k} X_n \mid \mathcal{B}\right] \\ &= \lim_{k \rightarrow \infty} \uparrow E[\inf_{n \geq k} X_n | \mathcal{B}] \\ &\leq \lim_{k \rightarrow \infty} \uparrow \inf_{n \geq k} E[X_n | \mathcal{B}] \\ &= \liminf_{n \rightarrow \infty} E[X_n | \mathcal{B}]. \end{aligned}$$

(e) 把 (d) 分别应用于 $Z - X_n$ 和 $Z + X_n$ ，得到

$$\begin{aligned} E[\liminf_n (Z - X_n) | \mathcal{B}] &\leq E[Z | \mathcal{B}] - \limsup_n E[X_n | \mathcal{B}], \\ E[\liminf_n (Z + X_n) | \mathcal{B}] &\leq E[Z | \mathcal{B}] + \liminf_n E[X_n | \mathcal{B}]. \end{aligned}$$

于是

$$E[X | \mathcal{B}] \leq \liminf_n E[X_n | \mathcal{B}] \leq \limsup_n E[X_n | \mathcal{B}] \leq E[X | \mathcal{B}].$$

所以 $E[X_n | \mathcal{B}] \rightarrow E[X | \mathcal{B}]$ 几乎必然。再由 $|E[X_n | \mathcal{B}]| \leq E[|X_n| | \mathcal{B}] \leq E[Z | \mathcal{B}]$ 以及 $E[E[Z | \mathcal{B}]] = E[Z] < \infty$ ，支配收敛定理还给出 L^1 收敛。

(f) 令

$$\mathcal{E}_f = \{(a, b) \in \mathbb{R}^2 : \forall x \in \mathbb{R}, f(x) \geq ax + b\}.$$

容易验证，对每个 $x \in \mathbb{R}$ ，

$$f(x) = \sup_{(a,b) \in \mathcal{E}_f} (ax + b) = \sup_{(a,b) \in \mathcal{E}_f \cap \mathbb{Q}^2} (ax + b).$$

利用 $\mathbb{Q}^2$ 的可数性，几乎必然有

$$\begin{aligned} E[f(X) | \mathcal{B}] &= E\left[\sup_{(a,b) \in \mathcal{E}_f \cap \mathbb{Q}^2} (aX + b) \mid \mathcal{B}\right] \\ &\geq \sup_{(a,b) \in \mathcal{E}_f \cap \mathbb{Q}^2} E[aX + b | \mathcal{B}] = f(E[X | \mathcal{B}]). \end{aligned}$$

注。 类比 $P(A) = E[1_A]$, 对任意 $A \in \mathcal{A}$, 常写

$$P(A | \mathcal{B}) := E[1_A | \mathcal{B}].$$

但应注意, 如此定义的 $P(A | \mathcal{B})$ 是随机变量, 而不是一个数。

11.2.3 平方可积随机变量的特殊情形

当 X 平方可积时, $E[X | \mathcal{B}]$ 还有一个重要解释。注意 $L^2(\Omega, \mathcal{B}, P)$ 可自然视为 $L^2(\Omega, \mathcal{A}, P)$ 的闭子空间: 它由那些至少有一个 $\mathcal{B}$ -可测代表元的等价类组成。

定理 11.2.3 若 $X \in L^2(\Omega, \mathcal{A}, P)$, 则 $E[X | \mathcal{B}]$ 是 X 在 $L^2(\Omega, \mathcal{B}, P)$ 上的正交投影。

证明。 上面的性质 (f) 表明 $E[X | \mathcal{B}]^2 \leq E[X^2 | \mathcal{B}]$ 几乎必然成立。因此 $E[E[X | \mathcal{B}]^2] \leq E[X^2] < \infty$, 从而随机变量 $E[X | \mathcal{B}]$ 属于 $L^2(\Omega, \mathcal{B}, P)$ 。此外, 对任意有界的 $\mathcal{B}$ -可测随机变量 Z ,

$$E[Z(X - E[X | \mathcal{B}])] = E[ZX] - E[ZE[X | \mathcal{B}]] = 0,$$

这里仍使用了条件期望的特征性质。因此 $X - E[X | \mathcal{B}]$ 与所有有界的 $\mathcal{B}$ -可测随机变量正交; 再由稠密性, 它与整个 $L^2(\Omega, \mathcal{B}, P)$ 正交, 结论得证。这个定理也给出了不借助 Radon–Nikodym 定理构造条件期望的方法: 先处理平方可积随机变量, 再作逼近。它还说明, 当 $X \in L^2$ 时, $E[X | \mathcal{B}]$ 是所有 $\mathcal{B}$ -可测随机变量中对 X 的最佳 L^2 逼近。

11.3 条件期望的特殊性质

上述建立的性质类似于期望 (或可测函数的积分) 的性质。在本段中, 我们将建立关于条件期望的更具体的性质。

命题 11.3.1 设 X 是实随机变量, Y 是 $\mathcal{B}$ -可测的实随机变量。则

$$E[YX | \mathcal{B}] = Y E[X | \mathcal{B}]$$

只要式中条件期望有定义, 上式就成立; 具体地, 只需假设 $X, Y \geq 0$, 或者假设 X 与 YX 都可积。

证明。 先设 $X \geq 0, Y \geq 0$ 。对任意非负的 $\mathcal{B}$ -可测随机变量 Z ,

$$E[Z(YE[X | \mathcal{B}])] = E[(ZY)E[X | \mathcal{B}]] = E[ZYX].$$

由于 $YE[X | \mathcal{B}]$ 是非负的 $\mathcal{B}$ -可测随机变量, 这个等式与条件期望的特征性质给出

$$YE[X | \mathcal{B}] = E[YX | \mathcal{B}].$$

若 X 与 YX 可积, 把 $X = X^+ - X^-$ 、 $Y = Y^+ - Y^-$ 代入非负情形并利用线性性, 即得一般结论。

命题 11.3.2 设 $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{A}$ 为两个子 σ -代数。则对任意非负或可积随机变量 X ,

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

注。在相同假设下，还有 $\mathbb{E}[E[X | \mathcal{B}_1] | \mathcal{B}_2] = E[X | \mathcal{B}_1]$ ；这是因为 $E[X | \mathcal{B}_1]$ 已经是 $\mathcal{B}_2$ -可测的。

证明。只证 $X \geq 0$ 的情形。若 Z 是非负的 $\mathcal{B}_1$ -可测随机变量，则它也 $\mathcal{B}_2$ -可测，因而

$$E[Z E[X | \mathcal{B}_2] | \mathcal{B}_1] = E[Z E[X | \mathcal{B}_2]] = E[ZX].$$

这足以证明所宣布的等式。

定理 11.3.3 两个子 σ -代数 $\mathcal{B}_1, \mathcal{B}_2$ 相互独立，当且仅当对每个非负的 $\mathcal{B}_2$ -可测随机变量 X ，有下式成立。等价地，只要求它对每个 $X \in L^1(\Omega, \mathcal{B}_2, P)$ 成立，或者只对 $X = 1_A, A \in \mathcal{B}_2$ 成立即可。

$$E[X | \mathcal{B}_1] = E[X].$$

证明。先设 $\mathcal{B}_1$ 与 $\mathcal{B}_2$ 独立。若 X 是非负的 $\mathcal{B}_2$ -可测随机变量，则对任意非负的 $\mathcal{B}_1$ -可测随机变量 Z ，

$$E[ZX] = E[Z]E[X] = E[Z E[X]],$$

因此常数随机变量 $E[X]$ 满足 $E[X | \mathcal{B}_1]$ 的特征性质。若 X 可积，再用线性性。反过来，假设

$$\forall A \in \mathcal{B}_2, \quad E[1_A | \mathcal{B}_1] = E[1_A] = P(A).$$

则对任意 $B \in \mathcal{B}_1$ ，

$$P(A \cap B) = E[1_A 1_B] = E[E[1_A | \mathcal{B}_1] 1_B] = E[P(A) 1_B] = P(A)P(B)$$

这正说明 $\mathcal{B}_1$ 与 $\mathcal{B}_2$ 独立。

注。设 X, Y 是实随机变量。由于 $\sigma(X)$ -可测随机变量恰为 X 的可测函数，前一定理说明： X 与 Y 独立，当且仅当

$$E[h(X) | Y] = E[h(X)]$$

对每个满足 $\mathbb{E}|h(X)| < \infty$ 的 Borel 函数 h 成立。这里 $E[h(X) | Y]$ 是 $E[h(X) | \sigma(Y)]$ 的简写。特别地，若 X 可积，则

$$E[X | Y] = E[X].$$

但仅凭最后这个等式不能推出 X 与 Y 独立。例如，设 $X \sim \mathcal{N}(0, 1)$ ，并令 $Y = |X|$ 。每个有界的 $\sigma(Y)$ -可测随机变量都可写成 $Z = g(Y)$ ，其中 g 有界；于是

$$E[ZX] = E[g(|X|)X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(|y|) y e^{-y^2/2} dy = 0.$$

所以 $E[X | Y] = 0 = E[X]$ ，但 X 与 Y 显然并不独立。下面给出另一个连接条件期望与独立性的常用定理。

定理 11.3.4 设随机变量 X, Y 分别取值于可测空间 E, F 。假设 X 与 $\mathcal{B}$ 独立, 且 Y 是 $\mathcal{B}$ -可测的。则对任意可测函数 $g: E \times F \rightarrow \mathbb{R}_+$,

$$E[g(X, Y) | \mathcal{B}] = \int g(x, Y) P_X(dx),$$

其中 P_X 是 X 的分布。右端是把随机变量 Y 代入可测函数

$$\Phi(y) = \int_E g(x, y) P_X(dx)$$

所得的随机变量; Φ 的可测性来自 Fubini 定理。

注。非正式地说, 在给定子 σ -代数 $\mathcal{B}$ 时, $\mathcal{B}$ -可测随机变量 Y 相当于常量; 而 X 与 $\mathcal{B}$ 独立, 意味着已知 $\mathcal{B}$ 不会提供关于 X 的额外信息。因此, $g(X, Y)$ 的条件期望只需保持 Y 不变, 并对第一个变量按 X 的分布积分。

证明。只需对任意非负的 $\mathcal{B}$ -可测随机变量 Z 验证

$$E[g(X, Y)Z] = E[\Phi(Y)Z].$$

记 $P_{(X, Y, Z)}$ 为三元组 (X, Y, Z) 在 $E \times F \times \mathbb{R}_+$ 上的分布。由于 X 与 (Y, Z) 独立,

$$P_{(X, Y, Z)} = P_X \otimes P_{(Y, Z)}$$

因此, 利用 Fubini 定理,

$$\begin{aligned} E[g(X, Y)Z] &= \int g(x, y)z P_{(X, Y, Z)}(dxdydz) \\ &= \int g(x, y)z P_X(dx)P_{(Y, Z)}(dydz) \\ &= \int_{F \times \mathbb{R}_+} z \left(\int_E g(x, y)P_X(dx) \right) P_{(Y, Z)}(dydz) \\ &= \int_{F \times \mathbb{R}_+} z \Phi(y) P_{(Y, Z)}(dydz) \\ &= E[\Phi(Y)Z] \end{aligned}$$

这就是所要求的结果。

11.4 条件期望的计算

11.4.1 离散条件化

设 Y 取值于可数空间 E , 且 $X \in L^1(\Omega, \mathcal{A}, P)$ 。前面已经证明

$$E[X | Y] = \varphi(Y)$$

其中

$$\varphi(y) = \frac{E[X 1_{\{Y=y\}}]}{P(Y=y)}$$

对每个满足 $P(Y=y) > 0$ 的 $y \in E$ 成立; 当 $P(Y=y) = 0$ 时, $\varphi(y)$ 可任意指定。

11.4.2 具有密度的随机变量

设随机变量 X, Y 分别取值于 $\mathbb{R}^m, \mathbb{R}^n$ 。假设向量 (X, Y) 的密度为 $p(x, y)$ 。于是对任意非负 Borel 函数 $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}_+$,

$$E[f(X, Y)] = \int_{\mathbb{R}^m \times \mathbb{R}^n} f(x, y) p(x, y) dx dy.$$

则 Y 的密度为

$$q(y) = \int_{\mathbb{R}^m} p(x, y) dx$$

严格地说, 在 $\int_{\mathbb{R}^m} p(x, y) dx = \infty$ 的那些 y 上应把 $q(y)$ 重新定义为 0; 这样的 y 构成 Lebesgue 零测集。后面的计算中略去这一技术细节。现在设 $h: \mathbb{R}^m \rightarrow \mathbb{R}_+$ 可测。为计算 $E[h(X) | Y]$, 取任意 Borel 函数 $g: \mathbb{R}^n \rightarrow \mathbb{R}_+$, 则

$$\begin{aligned} E[h(X)g(Y)] &= \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x) g(y) p(x, y) dx dy \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} h(x) p(x, y) dx \right) g(y) dy \\ &= \int_{\mathbb{R}^n} \left(\frac{\int_{\mathbb{R}^m} h(x) p(x, y) dx}{q(y)} \right) g(y) q(y) 1_{\{q(y) > 0\}} dy \\ &= \int_{\mathbb{R}^n} \varphi(y) g(y) q(y) 1_{\{q(y) > 0\}} dy \\ &= E[\varphi(Y) g(Y)], \end{aligned}$$

其中我们设定

$$\varphi(y) = \begin{cases} \frac{1}{q(y)} \int_{\mathbb{R}^m} h(x) p(x, y) dx, & q(y) > 0, \\ h(0), & q(y) = 0. \end{cases}$$

当 $q(y) = 0$ 时, $\varphi(y)$ 的值任意; 为了后面的陈述方便取为 $h(0)$ 。上面的计算还隐含使用了: 若 $q(y) = 0$, 则 $p(x, y) = 0$ 对 dx -几乎处处成立, 因而 $\int h(x) p(x, y) dx = 0$ 。由条件期望的特征性质可得

$$E[h(X) | Y] = \varphi(Y).$$

我们以略微不同的形式重新陈述这一结果。

命题 11.4.1 对每个 $y \in \mathbb{R}^n$, 在 $\mathbb{R}^m$ 上定义概率测度

$$\nu(y, dx) = \begin{cases} \frac{1}{q(y)} p(x, y) dx, & q(y) > 0, \\ \delta_0(dx), & q(y) = 0. \end{cases}$$

则对任意 Borel 函数 $h: \mathbb{R}^m \rightarrow \mathbb{R}_+$,

$$E[h(X) | Y] = \int \nu(Y, dx) h(x).$$

通常略作记号滥用: 对任意满足 $q(y) > 0$ 的 $y \in \mathbb{R}^n$, 写成

$$E[h(X) | Y = y] = \int \nu(y, dx) h(x) = \frac{1}{q(y)} \int h(x) p(x, y) dx$$

并称 $\nu(y, dx)$ 为给定 $Y = y$ 时 X 的条件分布。函数

$$x \longrightarrow \frac{p(x, y)}{q(y)}$$

称为给定 $Y = y$ 时 X 的条件密度。

练习。在上述假设下证明：更一般地，对任意 Borel 函数 $h: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ ，都有

$$E[h(X, Y) | Y] = \int h(x, Y) \nu(Y, dx).$$

11.4.3 高斯条件化

设 $X, Y_1, \dots, Y_p \in L^2(\Omega, \mathcal{A}, P)$ 。如第 11.2.3 节所述，条件期望

$$E[X | Y_1, \dots, Y_p]$$

是 X 在 $L^2(\Omega, \sigma(Y_1, \dots, Y_p), P)$ 上的正交投影；除平凡情形外，该子空间是无限维的。因此，这也是在 L^2 范数意义下，用形如 $\varphi(Y_1, \dots, Y_p)$ 的随机变量对 X 所作的最佳逼近。

在高斯情形，上一节把条件期望看作 L^2 中正交投影的抽象结论，还可以进一步化为有限维线性代数问题。这里要用到中心高斯向量的特征：取值于 $\mathbb{R}^k$ 的随机向量 $Z = (Z_1, \dots, Z_k)$ 是中心高斯向量，当且仅当它的任意线性组合都是中心高斯随机变量；等价地，其特征函数满足

$$\forall \xi \in \mathbb{R}^k, \quad E[e^{i\xi \cdot Z}] = \exp\left(-\frac{1}{2}\xi^\top K_Z \xi\right).$$

例如，当 $Z_1, \dots, Z_k$ 是相互独立的中心高斯随机变量时， $(Z_1, \dots, Z_k)$ 就是中心高斯向量。

命题 11.4.2 设 $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ 为中心高斯向量。则两个子向量 $(X_1, \dots, X_m)$ 与 $(Y_1, \dots, Y_n)$ 相互独立，当且仅当对所有 $1 \leq i \leq m$ 、 $1 \leq j \leq n$ ，都有

$$\text{Cov}(X_i, Y_j) = 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

证明。反向结论对任意具有二阶矩的独立随机向量都成立。现在假设式 (11.4) 成立。对任意 $\xi = (\eta_1, \dots, \eta_m, \zeta_1, \dots, \zeta_n) \in \mathbb{R}^{m+n}$ ，中心高斯向量的特征函数为

$$\Phi_W(\xi) = \exp\left(-\frac{1}{2}\xi^\top K_W \xi\right).$$

并且在条件 (11.4) 下，

$$\xi^\top K_W \xi = \sum_{j,k=1}^m \eta_j \eta_k \text{Cov}(X_j, X_k) + \sum_{j,k=1}^n \zeta_j \zeta_k \text{Cov}(Y_j, Y_k).$$

因此

$$E[\exp(i\xi \cdot (X_1, \dots, X_m, Y_1, \dots, Y_n))] = E[\exp(i \sum_{j=1}^m \eta_j X_j)] E[\exp(i \sum_{j=1}^n \zeta_j Y_j)],$$

或者说

$$\begin{aligned}\widehat{P}_{(X_1, \dots, X_m, Y_1, \dots, Y_n)}(\eta_1, \dots, \eta_m, \zeta_1, \dots, \zeta_n) &= \widehat{P}_{(X_1, \dots, X_m)}(\eta_1, \dots, \eta_m) \\ &\times \widehat{P}_{(Y_1, \dots, Y_n)}(\zeta_1, \dots, \zeta_n).\end{aligned}$$

利用傅里叶变换的单射性，我们得到

$$P_{(X_1, \dots, X_m, Y_1, \dots, Y_n)} = P_{(X_1, \dots, X_m)} \otimes P_{(Y_1, \dots, Y_n)}$$

这就是所需的独立性。

推论 设 $(X_1, \dots, X_n)$ 为中心高斯向量，且当 $j \neq k$ 时 $\text{Cov}(X_j, X_k) = 0$ 。则随机变量 $X_1, \dots, X_n$ 相互独立。事实上，前一命题立即给出这一结论。

由前一命题，可依次看出 X_n 与 $(X_1, \dots, X_{n-1})$ 独立、 X_{n-1} 与 $(X_1, \dots, X_{n-2})$ 独立，等等，从而得到推论。更一般地，若中心高斯向量 $(X_1, \dots, X_n)$ 的协方差矩阵为块对角矩阵，各对角块的大小依次为 $i_1, \dots, i_\ell$ ，其中 $i_1 + \dots + i_\ell = n$ ，则与这些块对应的子向量相互独立。

定理 11.4.3 设 $(Y_1, \dots, Y_n, X)$ 为中心高斯向量。则 $E[X | Y_1, \dots, Y_n]$ 等于 X 在由 $Y_1, \dots, Y_n$ 张成的向量子空间上的正交投影。因此存在实数 $\lambda_1, \dots, \lambda_n$ ，使得

$$E[X | Y_1, \dots, Y_n] = \sum_{j=1}^n \lambda_j Y_j.$$

此外，对任意 Borel 函数 $h: \mathbb{R} \rightarrow \mathbb{R}_+$ ，

$$E[h(X) | Y_1, \dots, Y_n] = \int_{\mathbb{R}} h(x) q_{\sum_{j=1}^n \lambda_j Y_j, \sigma^2}(x) dx,$$

其中

$$\sigma^2 = E[(X - \sum_{j=1}^n \lambda_j Y_j)^2]$$

并且对任意 $m \in \mathbb{R}$ ，

$$q_{m, \sigma^2}(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

是正态分布 $\mathcal{N}(m, \sigma^2)$ 的概率密度。

注。当且仅当 $X = \sum_{j=1}^n \lambda_j Y_j$ 几乎必然成立时，才有 $\sigma = 0$ 。此时 X 关于 $\sigma(Y_1, \dots, Y_n)$ 可测，定理中的第二个公式应理解为 $E[h(X) | Y_1, \dots, Y_n] = h(X)$ 。下面的证明中排除这一平凡情形，即假设 $\sigma > 0$ 。

证明。设 $\widehat{X} = \sum_{j=1}^n \lambda_j Y_j$ 为 X 在由 $Y_1, \dots, Y_n$ 张成的向量空间上的正交投影。于是对每个 $j \in \{1, \dots, n\}$ ，

$$\text{Cov}(X - \widehat{X}, Y_j) = E[(X - \widehat{X})Y_j] = 0.$$

由于向量 $(Y_1, \dots, Y_n, X - \widehat{X})$ 是中心高斯向量（其任意线性组合都是 $Y_1, \dots, Y_n, X$ 的线性组合），前一命题表明 $X - \widehat{X}$ 与 $(Y_1, \dots, Y_n)$ 独立。因此

$$E[X | Y_1, \dots, Y_n] = E[X - \widehat{X} | Y_1, \dots, Y_n] + \widehat{X} = E[X - \widehat{X}] + \widehat{X} = \widehat{X}.$$

这里先用了 $\hat{X}$ 关于 $\sigma(Y_1, \dots, Y_n)$ 可测, 再用了 $X - \hat{X}$ 与 $(Y_1, \dots, Y_n)$ 的独立性; 因此 $E[X - \hat{X} | Y_1, \dots, Y_n] = E[X - \hat{X}] = 0$ 。

为证明最后一个断言, 令 $Z = X - \hat{X}$ 。于是 Z 与 $(Y_1, \dots, Y_n)$ 独立, 且服从 $\mathcal{N}(0, \sigma^2)$: 它是中心高斯变量, 并且按定义 $\sigma^2 = E[Z^2]$ 。由定理 11.3.4,

$$\begin{aligned} E[h(X) | Y_1, \dots, Y_n] &= E \left[h \left(\sum_{j=1}^n \lambda_j Y_j + Z \right) \middle| Y_1, \dots, Y_n \right] \\ &= \int_{\mathbb{R}} h \left(\sum_{j=1}^n \lambda_j Y_j + z \right) P_Z(dz). \end{aligned}$$

写成 $P_Z(dz) = q_{0, \sigma^2}(z) dz$, 再作变量代换 $x = \sum_{j=1}^n \lambda_j Y_j + z$, 便得到定理中的公式。

11.5 转移概率与条件分布

前面的条件期望计算可以通过使用转移概率的概念以更加简洁的方式重新表述。

定义 11.5.1 设 $(E, \mathcal{E})$ 与 $(F, \mathcal{F})$ 为两个可测空间。从 E 到 F 的转移概率 (或转移核) 是映射

$$\nu : E \times \mathcal{F} \longrightarrow [0, 1]$$

它满足以下两个性质:

- (i) 对每个 $x \in E$, $\nu(x, \cdot)$ 是 $(F, \mathcal{F})$ 上的概率测度;
- (ii) 对每个 $A \in \mathcal{F}$, 映射 $x \mapsto \nu(x, A)$ 是 $\mathcal{E}$ -可测的。

直观地说, 固定 $x \in E$ 后, 概率测度 $\nu(x, \cdot)$ 给出在状态空间 F 中随机选取下一状态的方法。第 13 章研究马尔可夫链时, 若时刻 n 的状态为 x , 则时刻 $n+1$ 的条件分布正由 $\nu(x, dy)$ 给出。

例。 设 λ 是 $(F, \mathcal{F})$ 上的 σ -有限正测度, 且 $f : E \times F \rightarrow \mathbb{R}_+$ 可测并满足

$$\int_F f(x, y) \lambda(dy) = 1, \quad \forall x \in E.$$

那么

$$\nu(x, A) = \int_A f(x, y) \lambda(dy)$$

便定义了从 E 到 F 的转移概率。性质 (ii) 可由 Tonelli 定理 (或参数积分的可测性) 得到。

命题 11.5.1 (i) 若 h 是 $(F, \mathcal{F})$ 上的非负可测函数 (或有界可测函数), 则

$$\varphi(x) := \int \nu(x, dy) h(y), \quad x \in E$$

是 $(E, \mathcal{E})$ 上的非负可测函数 (相应地, 为有界可测函数)。(ii) 若 λ 是 $(E, \mathcal{E})$ 上的概率测度, 则

$$\mu(A) := \int \lambda(dx) \nu(x, A), \quad A \in \mathcal{F}$$

是 $(F, \mathcal{F})$ 上的概率测度。

这些性质很容易验证。对 (i)，先对非负简单函数 h 验证，再用单调收敛定理；有界函数的情况可由正部、负部分解得到。下面说明转移概率与条件期望之间的联系。

定义 11.5.2 设随机变量 X 与 Y 分别取值于 $(E, \mathcal{E})$ 与 $(F, \mathcal{F})$ 。若从 E 到 F 的转移概率 ν 满足：对每个 $(F, \mathcal{F})$ 上的非负可测函数 h ，都有下式，则称 ν 为给定 X 时 Y 的一个条件分布：

$$E[h(Y) | X] = \int \nu(X, dy) h(y).$$

注。 随机变量 $\int \nu(X, dy) h(y)$ 是 X 与可测映射

$$x \mapsto \int \nu(x, dy) h(y)$$

的复合。因此它确实是 X 的函数，正如条件期望 $E[h(Y) | X]$ 所必须满足的那样。

按照定义，对任意 $A \in \mathcal{F}$ ，

$$P(Y \in A | X) = \nu(X, A), \quad \text{几乎必然.}$$

为了表达直觉，人们常把它形式化地写成

$$P(Y \in A | X = x) = \nu(x, A),$$

对每个 $x \in E$ ，把上式直观地写成 $P(Y \in A | X = x) = \nu(x, A)$ 很自然，但这个等式通常不能用初等条件概率来定义：除非 X 是离散随机变量，否则常有 $P(X = x) = 0$ 。因此严格表述应是随机变量等式 $P(Y \in A | X) = \nu(X, A)$ 。现在讨论唯一性。若 ν 与 ν' 都是给定 X 时 Y 的条件分布，则对每个 $A \in \mathcal{F}$ ，

$$\nu(X, A) = P(Y \in A | X) = \nu'(X, A), \quad \text{几乎必然,}$$

等价地，对每个 $A \in \mathcal{F}$ ，

$$\nu(x, A) = \nu'(x, A), \quad P_X(dx)\text{-几乎处处.}$$

再假设 $(F, \mathcal{F})$ 上的概率测度可由某个可数可测集族上的取值唯一确定；例如 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 可取端点为有理数的半开矩形族。把上述零测集对这个可数集族取并，便得到

$$\nu(x, \cdot) = \nu'(x, \cdot), \quad P_X(dx)\text{-几乎处处.}$$

这就是条件分布的唯一性：只能要求在 P_X -几乎处处意义下唯一，不可能要求对每个 x 都唯一。尽管如此，通常仍简称为“给定 X 时 Y 的条件分布”。下面转向其存在性。

定理 11.5.2 设 E 与 F 都是完备可分度量空间，并分别赋予其 Borel σ -代数 $\mathcal{E}$ 与 $\mathcal{F}$ 。则给定 X 时 Y 的条件分布总是存在。

这个定理是测度论中较深的结果，本讲义不作证明。后文也不必直接调用它，因为所需条件分布都能显式构造。下面重述上一节的三个例子；注意其中 X 与 Y 的角色已经互换。

(1) 若 X 是离散随机变量, 即 E 可数, 令 $E' = \{a \in E : P(X = a) > 0\}$. 可定义

$$\nu(x, A) = \begin{cases} P(Y \in A | X = x), & x \in E', \\ \delta_{y_0}(A), & x \notin E', \end{cases} \quad A \in \mathcal{F},$$

其中 $y_0 \in F$ 是任意固定点。

(2) 设 X 与 Y 分别取值于 $\mathbb{R}^m$ 与 $\mathbb{R}^n$, 并且 (X, Y) 具有联合密度 $p(x, y)$. 则 X 的密度为

$$q(x) = \int_{\mathbb{R}^n} p(x, y) dy.$$

由命题 11.4.1, 给定 X 时 Y 的条件分布可定义为

$$\nu(x, A) = \begin{cases} \frac{1}{q(x)} \int_A p(x, y) dy, & q(x) > 0, \\ \delta_0(A), & q(x) = 0. \end{cases}$$

(3) 最后, 设 $(X_1, \dots, X_n, Y)$ 为中心高斯向量, 并把

$$\sum_{j=1}^n \lambda_j X_j$$

记作 Y 在由 $X_1, \dots, X_n$ 张成的向量空间上的正交投影。再令

$$\sigma^2 = E[(Y - \sum_{j=1}^n \lambda_j X_j)^2].$$

定理 11.4.3 表明, 给定 $X = (X_1, \dots, X_n)$ 后, Y 的条件分布为

$$\nu(x_1, \dots, x_n; A) = \int_A q_{\sum_{j=1}^n \lambda_j x_j, \sigma^2}(y) dy,$$

其中 q_{m, σ^2} 是高斯分布 $\mathcal{N}(m, \sigma^2)$ 的密度。略作术语上的简化, 我们也说: 在给定 $(X_1, \dots, X_n)$ 的条件下, Y 服从分布

$$\mathcal{N}\left(\sum_{j=1}^n \lambda_j X_j, \sigma^2\right).$$

随机过程

第三部分

随机过程

第十二章 离散时间鞅论

12.1 定义与例

固定概率空间 $(\Omega, \mathcal{F}, P)$ 。随机过程是定义在这个概率空间上的随机变量序列 $(X_n)_{n \in \mathbb{N}}$ 。本章中的随机过程均取实值。

定义 12.1.1 $(\Omega, \mathcal{F})$ 上的滤过，是一列递增的子 σ -代数 $(\mathcal{F}_n)_{n \in \mathbb{N}}$ ，即

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$$

此时也称 $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, P)$ 为滤过概率空间。

通常把参数 n 理解为时间，而 $\mathcal{F}_n$ 表示截至时刻 n 已获得的全部信息。**例。**(a) 设 $(X_n)_{n \in \mathbb{N}}$ 是定义在 $(\Omega, \mathcal{F}, P)$ 上的任意随机变量序列。令 $\mathcal{F}_n^X$ 为使 $X_0, X_1, \dots, X_n$ 均可测的最小 σ -代数，即

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \dots, X_n).$$

于是 $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$ 是一个滤过，称为随机过程 $(X_n)_{n \in \mathbb{N}}$ 的典范滤过。(b) 设 $\Omega = [0, 1)$ ， $\mathcal{F}$ 是其 Borel σ -代数，而 P 是 Lebesgue 测度。令

$$\mathcal{F}_n = \sigma\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right) : i = 1, 2, \dots, 2^n\right).$$

则 $(\mathcal{F}_n)_{n \in \mathbb{N}}$ 是 $[0, 1)$ 上的二进滤过。

定义 12.1.2 若对每个 $n \in \mathbb{N}$ ，随机变量 X_n 都关于 $\mathcal{F}_n$ 可测，则称过程 $(X_n)_{n \in \mathbb{N}}$ 适应于滤过 $(\mathcal{F}_n)_{n \in \mathbb{N}}$ 。

典范滤过按构造是使该过程适应的最小滤过。本章以下部分（第 12.6 节除外）固定一个滤过概率空间 $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, P)$ ；在例子中有时会明确指定它。下列概念均相对于这个空间而言。

定义 12.1.3 设 $(X_n)_{n \in \mathbb{N}}$ 是适应过程，并且对每个 $n \in \mathbb{N}$ 都有 $E[|X_n|] < \infty$ 。称该过程为：

- 鞅，如果对每个 $n \in \mathbb{N}$ ，

$$E[X_{n+1} \mid \mathcal{F}_n] = X_n;$$

- 上鞅，如果对每个 $n \in \mathbb{N}$ ，

$$E[X_{n+1} \mid \mathcal{F}_n] \leq X_n;$$

- 下鞅，如果对每个 $n \in \mathbb{N}$ ，

$$E[X_{n+1} \mid \mathcal{F}_n] \geq X_n.$$

鞅的定义立即推出一个表面上更强的性质：对任意 $0 \leq n \leq m$ ，

$$E[X_m | \mathcal{F}_n] = X_n$$

可对 $m - n$ 作归纳证明。 $m = n$ 时显然； $m = n + 1$ 时就是定义；若 $m - n \geq 2$ ，由条件期望的塔式性质，

$$E[X_m | \mathcal{F}_n] = E[E[X_m | \mathcal{F}_{m-1}] | \mathcal{F}_n] = E[X_{m-1} | \mathcal{F}_n].$$

特别地，式 (12.1) 蕴含 $E[X_m] = E[X_n] = E[X_0]$ 。同理，若 (X_n) 是上鞅（相应地，下鞅），则对所有 $0 \leq n \leq m$ ，

$$E[X_m | \mathcal{F}_n] \leq X_n \quad (\text{相应地, } E[X_m | \mathcal{F}_n] \geq X_n).$$

从而 $E[X_m] \leq E[X_n]$ （相应地， $E[X_m] \geq E[X_n]$ ）。把鞅理解为公平游戏很有帮助： X_n 表示玩家在时刻 n 的资产， $\mathcal{F}_n$ 表示截至该时刻掌握的信息。等式 $E[X_{n+1} | \mathcal{F}_n] = X_n$ 表示在已知过去的条件下，下一时刻资产的平均值等于当前资产；平均而言既不赢也不输。上鞅相应于不利游戏。显然， (X_n) 是上鞅当且仅当 $(-X_n)$ 是下鞅。因此，后文只为其中一类陈述的结果，都有符号相反的对版本。

例。(i) 若 $X \in L^1(\Omega, \mathcal{F}, P)$ ，令

$$X_n = E[X | \mathcal{F}_n].$$

则 $(X_n)_{n \in \mathbb{N}}$ 是鞅，因为

$$E[X_{n+1} | \mathcal{F}_n] = E[E[X | \mathcal{F}_{n+1}] | \mathcal{F}_n] = E[X | \mathcal{F}_n] = X_n.$$

这种鞅称为闭鞅。(ii) 若 (X_n) 是由可积随机变量组成的递减适应过程，则 (X_n) 是上鞅：

$$E[X_{n+1} | \mathcal{F}_n] \leq E[X_n | \mathcal{F}_n] = X_n.$$

(iii) $\mathbb{R}$ 上的随机游走。设 $x \in \mathbb{R}$ ，并设 $(Y_n)_{n \geq 1}$ 是一列独立同分布的实随机变量，其共同分布为 μ ，且 $E[|Y_1|] < \infty$ 。令

$$X_0 = x, \quad X_n = x + Y_1 + \cdots + Y_n \quad (n \geq 1).$$

再定义滤过

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \text{ 并且当 } n \geq 1 \text{ 时, } \mathcal{F}_n = \sigma(Y_1, \dots, Y_n).$$

这正是 (X_n) 的典范滤过。于是 (X_n) 在下列情形分别为：

- 鞅，如果 $E[Y_1] = 0$ ；
- 上鞅，如果 $E[Y_1] \leq 0$ ；
- 下鞅，如果 $E[Y_1] \geq 0$ 。

例如，在 $E[Y_1] = 0$ 时，

$$E[X_{n+1} | \mathcal{F}_n] = E[X_n + Y_{n+1} | \mathcal{F}_n] = X_n + E[Y_{n+1}] = X_n,$$

因为按构造 Y_{n+1} 与 $\mathcal{F}_n$ 独立。过程 (X_n) 称为从 x 出发、跳跃分布为 μ 的 $\mathbb{R}$ 上随机游走。(iv) 回到例 (b) 的滤过概率空间。设 μ 是 $[0, 1)$ 上的有限测度，且 $P = \lambda$ 是 Lebesgue 测度。对每个 $n \in \mathbb{N}$ ，令

$$f_n = \frac{d(\mu|_{\mathcal{F}_n})}{d(\lambda|_{\mathcal{F}_n})}.$$

这里 f_n 是把 μ 与 λ 限制在 $\mathcal{F}_n$ 上以后, μ 关于 λ 的 Radon-Nikodym 导数。在有限 σ -代数 $\mathcal{F}_n$ 上, 每个测度都关于 λ 绝对连续。容易验证

$$f_n(\omega) = \sum_{i=1}^{2^n} \frac{\mu([(i-1)2^{-n}, i2^{-n}))}{2^{-n}} 1_{[(i-1)2^{-n}, i2^{-n})}(\omega).$$

于是 $(f_n)_{n \in \mathbb{N}}$ 是鞅。事实上, 对任意 $A \in \mathcal{F}_n$,

$$E[1_A f_{n+1}] = \int 1_A(\omega) f_{n+1}(\omega) d\omega = \mu(A) = \int 1_A(\omega) f_n(\omega) d\omega = E[1_A f_n],$$

这正是 $f_n = E[f_{n+1} | \mathcal{F}_n]$ 的刻画。特别地, 若 $\mu \ll \lambda$ 在 $\mathcal{F}$ 上成立, 记 $f = d\mu/d\lambda$, 则

$$f_n = E[f | \mathcal{F}_n],$$

因此这里的鞅属于例 (i) 中的闭鞅。

鞅的两种变换

两个鞅的变换。

命题 12.1.1 设 $\varphi: \mathbb{R} \rightarrow \mathbb{R}_+$ 为凸函数, $(X_n)_{n \in \mathbb{N}}$ 为适应过程, 并且对每个 n 都有 $E[\varphi(X_n)] < \infty$ 。

(i) 若 (X_n) 是鞅, 则 $(\varphi(X_n))$ 是下鞅。

(ii) 若 (X_n) 是下鞅且 φ 单调递增, 则 $(\varphi(X_n))$ 是下鞅。

特别地, 若 (X_n) 是鞅, 则 $(|X_n|)$ 是下鞅; 若另外 $E[X_n^2] < \infty$, 则 (X_n^2) 也是下鞅。若 (X_n) 是下鞅, 则 (X_n^+) 仍是下鞅。

证明。 (i) 由条件 Jensen 不等式,

$$E[\varphi(X_{n+1}) | \mathcal{F}_n] \geq \varphi(E[X_{n+1} | \mathcal{F}_n]) = \varphi(X_n).$$

(ii) 同理, 因为 $X_n \leq E[X_{n+1} | \mathcal{F}_n]$ 且 φ 单调递增,

$$E[\varphi(X_{n+1}) | \mathcal{F}_n] \geq \varphi(E[X_{n+1} | \mathcal{F}_n]) \geq \varphi(X_n).$$

定义 12.1.4 若对每个 $n \geq 1$, 实随机变量 H_n 都有界且关于 $\mathcal{F}_{n-1}$ 可测, 则称随机变量族 $(H_n)_{n \geq 1}$ 是可预测的。

命题 12.1.2 设 $(X_n)_{n \in \mathbb{N}}$ 为适应过程, $(H_n)_{n \geq 1}$ 为可预测族。令 $(H \cdot X)_0 = 0$, 且对 $n \geq 1$,

$$(H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1}), \quad n \geq 1.$$

则:

(i) 若 (X_n) 是鞅, 则 $((H \cdot X)_n)$ 也是鞅。

(ii) 若 (X_n) 是上鞅（相应地，是下鞅），且对每个 $n \geq 1$ 都有 $H_n \geq 0$ ，则 $((H \cdot X)_n)$ 仍是上鞅（相应地，仍是下鞅）。

证明。(i) 由于 H_n 有界， $(H \cdot X)_n$ 可积；按构造它还是适应过程。因此只需验证对每个 $n \in \mathbb{N}$,

$$E[(H \cdot X)_{n+1} - (H \cdot X)_n \mid \mathcal{F}_n] = 0.$$

而 $(H \cdot X)_{n+1} - (H \cdot X)_n = H_{n+1}(X_{n+1} - X_n)$ 。由于 H_{n+1} 关于 $\mathcal{F}_n$ 可测，

$$E[H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] = H_{n+1} E[X_{n+1} - X_n \mid \mathcal{F}_n] = 0.$$

(ii) 的证明完全类似。若把鞅 X_n 解释为玩家在时刻 n 的资产，则 $X_{n+1} - X_n$ 是第 $n+1$ 轮的收益。玩家在时刻 n 依据已有信息选择赌注倍数 H_{n+1} ，因此该倍数必须 $\mathcal{F}_n$ -可测。变换后的收益是 $H_{n+1}(X_{n+1} - X_n)$ ，公平性仍保持；这解释了随机积分 $(H \cdot X)_n$ 的定义。

12.2 停时

定义 12.2.1 取值于 $\bar{\mathbb{N}} := \mathbb{N} \cup \{+\infty\}$ 的随机变量 T 称为滤过 $(\mathcal{F}_n)$ 的停时，如果对每个 $n \in \mathbb{N}$ ，都有

$$\{T = n\} \in \mathcal{F}_n.$$

这等价于对每个 $n \in \mathbb{N}$ 都有 $\{T \leq n\} \in \mathcal{F}_n$ 。下文可任用这两种刻画。注意停时允许取值 $+\infty$ 。写成

$$\{T = +\infty\} = \Omega \setminus \bigcup_{n \in \mathbb{N}} \{T = n\}$$

可知 $\{T = +\infty\} \in \mathcal{F}_\infty$ ，其中

$$\mathcal{F}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{F}_n = \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right).$$

从博弈观点看，停时是能依据当前信息决定停止的随机时刻：在时刻 n 只能使用 $\mathcal{F}_n$ 中的信息。例如，不能把“股票达到全年最高价的时刻”作为停时，因为这需要预知未来。**例。**(i) 任意常值时间 $T = k$ ($k \in \mathbb{N}$) 都是停时。(ii) 若 (Y_n) 是适应过程， $A \in \mathcal{B}(\mathbb{R})$ ，则

$$T_A := \inf\{n \in \mathbb{N} : Y_n \in A\}$$

是一种停时，称为进入 A 的时间。事实上，对于任意整数 $n \geq 0$,

$$\{T_A = n\} = \{Y_0 \notin A, Y_1 \notin A, \dots, Y_{n-1} \notin A, Y_n \in A\} \in \mathcal{F}_n.$$

注意，在 T_A 的定义中约定 $\inf \emptyset = +\infty$ ；下文始终采用这一约定。(iii) 反过来，固定整数 $N > 0$ ，并定义最后访问时刻

$$L_A := \sup\{n \leq N : Y_n \in A\}, \quad \sup \emptyset := 0.$$

一般情况下 L_A 不是停时。事实上，对于 $n \in \{1, \dots, N-1\}$,

$$\{L_A = n\} = \{Y_n \in A, Y_{n+1} \notin A, \dots, Y_N \notin A\}$$

原则上不属于 $\mathcal{F}_n$.

命题 12.2.1 (i) 若 S, T 是停时, 则 $S \vee T$ 与 $S \wedge T$ 也是停时. (ii) 若 $(T_k)_{k \in \mathbb{N}}$ 是一列停时, 则 $\inf_k T_k$ 、 $\sup_k T_k$ 、 $\limsup_k T_k$ 与 $\liminf_k T_k$ 也都是停时.

证明. (i) 对每个 $n \in \mathbb{N}$,

$$\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}, \quad \{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}.$$

(ii) 同理, $\{\inf_k T_k \leq n\} = \bigcup_k \{T_k \leq n\}$. 例如,

$$\{\liminf_k T_k \leq n\} = \bigcap_{m=0}^{\infty} \bigcup_{k=m}^{\infty} \{T_k \leq n\}.$$

其余情形完全类似.

定义 12.2.2 设 T 为停时. 直到时刻 T 的过去 σ -代数定义为

$$\mathcal{F}_T = \{A \in \mathcal{F} : \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

容易验证 $\mathcal{F}_T$ 是一个 σ -代数; 若 $T = n$ 为常值停时, 则 $\mathcal{F}_T = \mathcal{F}_n$.

命题 12.2.2 设 S, T 是两个停时, 并且 $S \leq T$. 则 $\mathcal{F}_S \subset \mathcal{F}_T$.

证明. 设 $A \in \mathcal{F}_S$. 对任意 $n \in \mathbb{N}$,

$$A \cap \{T = n\} = \bigcup_{k=0}^n (A \cap \{S = k\}) \cap \{T = n\} \in \mathcal{F}_n.$$

命题 12.2.3 设 $(Y_n)_{n \in \mathbb{N}}$ 是适应过程, T 是停时. 定义

$$1_{\{T < \infty\}} Y_T(\omega) = \begin{cases} Y_n(\omega), & T(\omega) = n \in \mathbb{N}, \\ 0, & T(\omega) = +\infty. \end{cases}$$

则随机变量 $1_{\{T < \infty\}} Y_T$ 是 $\mathcal{F}_T$ -可测的.

证明. 设 $B \subset \mathbb{R}$ 是 Borel 集. 对每个 $n \in \mathbb{N}$,

$$\{1_{\{T < \infty\}} Y_T \in B\} \cap \{T = n\} = \{Y_n \in B\} \cap \{T = n\} \in \mathcal{F}_n,$$

这说明 $\{1_{\{T < \infty\}} Y_T \in B\} \in \mathcal{F}_T$ (先处理 $0 \notin B$ 的情形). 若 $0 \in B$, 只需写成

$$\{1_{\{T < \infty\}} Y_T \in B\} = \{1_{\{T < \infty\}} Y_T \in B^c\}^c.$$

若停时 $T < \infty$ 几乎必然, 当然简写 Y_T , 不再写 $1_{\{T < \infty\}} Y_T$. 特别地, 对任意停时 T , $n \wedge T$ 也是停时 (引理 12.2.1). 由前一命题, $Y_{n \wedge T}$ 是 $\mathcal{F}_{n \wedge T}$ -可测的; 再由命题 12.2.2, 它也是 $\mathcal{F}_n$ -可测的.

定理 12.2.4 (停时定理) 设 $(X_n)_{n \in \mathbb{N}}$ 是鞅 (相应地, 是上鞅), T 是停时。则 $(X_{n \wedge T})_{n \in \mathbb{N}}$ 仍是鞅 (相应地, 仍是上鞅)。特别地, 若 T 有界, 则 $X_T \in L^1$, 且

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] \quad (\text{上鞅情形为 } \mathbb{E}[X_T] \leq \mathbb{E}[X_0]) .$$

证明。 对每个 $n \geq 1$, 令

$$H_n = \mathbf{1}_{\{T \geq n\}} = 1 - \mathbf{1}_{\{T \leq n-1\}} .$$

于是 $(H_n)_{n \geq 1}$ 是可预测过程。又因为

$$X_{n \wedge T} = X_0 + (H \cdot X)_n$$

定理第一部分由命题 12.1.2 得到。若 $T \leq N$, 则 $\mathbb{E}[X_T] = \mathbb{E}[X_{N \wedge T}] = \mathbb{E}[X_0]$; 上鞅情形则相应得到不等式。 T 有界的假设不能删去。设 $X_n = Y_1 + \cdots + Y_n$ 是从 0 出发的简单对称随机游走, 其中 $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = \frac{1}{2}$ 。则 (X_n) 是鞅。若令

$$T = \inf\{n \geq 0 : X_n = 1\}$$

这里 $T < \infty$ 几乎必然 (参见命题 10.2.2; 下一节还会给出另一种证明), 但是

$$1 = \mathbb{E}[X_T] \neq \mathbb{E}[X_0] = 0.$$

当然, 这里的停时 T 没有有界性, 所以与定理并不矛盾。

12.3 鞅的几乎必然收敛

下面研究鞅或下鞅在 $n \rightarrow \infty$ 时的几乎必然收敛性。先从一个确定性数列 $\alpha = (\alpha_n)_{n \in \mathbb{N}}$ 开始。

给定实数 $a < b$, 对数列 $\alpha = (\alpha_n)_{n \in \mathbb{N}}$ 定义两列取值于 $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ 的时刻 $S_k(\alpha)$ 与 $T_k(\alpha)$ 。先令

$$\begin{aligned} S_1(\alpha) &= \inf\{n \geq 0 : \alpha_n \leq a\} \\ T_1(\alpha) &= \inf\{n \geq S_1(\alpha) : \alpha_n \geq b\} \end{aligned}$$

再递归地令

$$\begin{aligned} S_{k+1}(\alpha) &= \inf\{n \geq T_k(\alpha) : \alpha_n \leq a\} \\ T_{k+1}(\alpha) &= \inf\{n \geq S_{k+1}(\alpha) : \alpha_n \geq b\}. \end{aligned}$$

这些定义都采用 $\inf \emptyset = +\infty$ 的约定。对每个整数 $n \geq 0$, 令

$$\begin{aligned} N_n([a, b], \alpha) &= \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k(\alpha) \leq n\}}, \\ N_{\infty}([a, b], \alpha) &= \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k(\alpha) < \infty\}}. \end{aligned}$$

量 $N_{\infty}([a, b], \alpha)$ 称为数列 (α_n) 对区间 $[a, b]$ 的上穿次数: 每一次先到达不高于 a 的位置、随后到达不低于 b 的位置, 就完成一次上穿。下面使用一个初等分析引理。

引理 12.3.1 数列 $(\alpha_n)_{n \in \mathbb{N}}$ 在扩充实数集 $\bar{\mathbb{R}}$ 中收敛, 当且仅当对每一对满足 $a < b$ 的有理数 a, b , 都有 $N_\infty([a, b], \alpha) < \infty$ 。

现在令 $(X_n)_{n \in \mathbb{N}}$ 为适应过程。此时 $S_k(X)$ 与 $T_k(X)$ 是取值于 $\bar{\mathbb{N}}$ 的随机变量, 而且都是停时。例如,

$$\{T_k(X) \leq n\} = \bigcup_{0 \leq m_1 < n_1 < \dots < m_k < n_k \leq n} \{X_{m_1} \leq a, X_{n_1} \geq b, \dots, X_{m_k} \leq a, X_{n_k} \geq b\}$$

所以 $\{T_k(X) \leq n\} \in \mathcal{F}_n$ 。特别地, $N_n([a, b], X)$ 是 $\mathcal{F}_n$ -可测随机变量。

引理 12.3.2 (Doob 上穿不等式) 设 $(X_n)_{n \in \mathbb{N}}$ 是下鞅。则对任意实数 $a < b$ 及任意 $n \in \mathbb{N}$,

$$(b - a) E[N_n([a, b], X)] \leq E[(X_n - a)^+ - (X_0 - a)^+].$$

证明。 令 $Y_n = (X_n - a)^+$ 。由命题 12.1.1, (Y_n) 仍是下鞅。为简化记号, 写 $N_n = N_n([a, b], X)$, 并将 $S_k(X), T_k(X)$ 简写为 S_k, T_k 。定义可预测过程 $(H_n)_{n \geq 1}$:

$$H_n = \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k < n \leq T_k\}} \leq 1$$

注意, $\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \setminus \{T_k \leq n-1\} \in \mathcal{F}_{n-1}$, 因为 S_k, T_k 都是停时。因此 (H_n) 可预测, 并且直接计算得

$$\begin{aligned} (H \cdot Y)_n &= \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) + \mathbf{1}_{\{S_{N_n+1} < n\}} (Y_n - Y_{S_{N_n+1}}) \\ &\geq \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) \geq N_n(b - a). \end{aligned}$$

第一个不等式成立, 是因为在 $\{S_{N_n+1} < n\}$ 上有 $Y_{S_{N_n+1}} = 0$, 且 $Y_n \geq 0$ 。所以

$$E[(H \cdot Y)_n] \geq (b - a) E[N_n].$$

另一方面, 令 $K_n = 1 - H_n$ 。则 (K_n) 是非负可预测过程, 命题 12.1.2 说明 $(K \cdot Y)$ 是下鞅, 故 $\mathbb{E}[(K \cdot Y)_n] \geq 0$ 。又有

$$(K \cdot Y)_n + (H \cdot Y)_n = ((K + H) \cdot Y)_n = Y_n - Y_0.$$

因此

$$(b - a) \mathbb{E}[N_n] \leq \mathbb{E}[(H \cdot Y)_n] \leq \mathbb{E}[(K \cdot Y)_n + (H \cdot Y)_n] = \mathbb{E}[Y_n - Y_0].$$

这就是引理中的不等式。

定理 12.3.3 设 $(X_n)_{n \in \mathbb{N}}$ 为下鞅, 并满足

$$\sup_{n \in \mathbb{N}} E[(X_n)^+] < \infty.$$

则当 $n \rightarrow \infty$ 时, X_n 几乎必然收敛; 其极限 X_∞ 满足 $\mathbb{E}[|X_\infty|] < \infty$ 。

注。 由 $\mathbb{E}[X_n] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n^-]$ 以及下鞅性质 $\mathbb{E}[X_n] \geq \mathbb{E}[X_0]$, 对每个 $k \geq 0$,

$$\mathbb{E}[(X_k)^-] \leq \left(\sup_{n \in \mathbb{N}} \mathbb{E}[(X_n)^+] \right) - \mathbb{E}[X_0].$$

所以条件 (12.2) 等价于

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < \infty$$

即 (X_n) 在 L^1 中有界。**证明。** 固定有理数 $a < b$ 。由引理 12.3.2, 对每个 $n \geq 1$,

$$(b-a)\mathbb{E}[N_n([a, b], X)] \leq \mathbb{E}[(X_n - a)^+] \leq |a| + \mathbb{E}[X_n^+] \leq |a| + \sup_{k \in \mathbb{N}} \mathbb{E}[X_k^+].$$

令 $n \rightarrow \infty$, 并使用条件 (12.2), 得到

$$(b-a) \mathbb{E}[N_\infty([a, b], X)] < \infty$$

所以 $N_\infty([a, b], X) < \infty$ 几乎必然。由于有理数对 (a, b) 只有可数多个, 删去一个零概率集后, 可以同时保证: 对所有有理数 $a < b$, 都有 $N_\infty([a, b], X) < \infty$ 。由引理 12.3.1, (X_n) 几乎必然在扩充实数集 $\mathbb{R}$ 中存在极限, 记为 X_∞ 。再由 Fatou 引理以及定理后的注,

$$\mathbb{E}[|X_\infty|] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[|X_n|] \leq \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < \infty$$

因此 X_∞ 实际上是有限值随机变量, 且 $|X_\infty| < \infty$ 几乎必然。¹

推论 12.3.4 设 $(X_n)_{n \in \mathbb{N}}$ 是非负上鞅。则 X_n 几乎必然收敛, 其极限 $X_\infty \in L^1$, 并且对每个 $n \in \mathbb{N}$,

$$X_n \geq \mathbb{E}[X_\infty | \mathcal{F}_n] \quad \text{几乎必然.}$$

证明。 把定理 12.3.3 应用于下鞅 $X'_n = -X_n$ 。由于 $(X'_n)^+ = X_n^- = 0$, 条件 (12.2) 自动成立。最后, 对固定 n 和 $m \geq n$, 上鞅性质给出 $X_n \geq \mathbb{E}[X_m | \mathcal{F}_n]$; 再使用条件 Fatou 引理,

$$X_n \geq \liminf_{m \rightarrow \infty} \mathbb{E}[X_m | \mathcal{F}_n] \geq \mathbb{E}[\liminf_{m \rightarrow \infty} X_m | \mathcal{F}_n] = \mathbb{E}[X_\infty | \mathcal{F}_n].$$

例。 (1) 令 $Y_n = 1 + Z_1 + \cdots + Z_n$ 为从 1 出发的简单对称随机游走。前面已知 (Y_n) 关于其规范滤过是鞅。令

$$T = \inf\{n \geq 0 : Y_n = 0\}.$$

则 T 是停时。由定理 12.2.4, $X_n = Y_{n \wedge T}$ 是正鞅, 故可应用上述推论: X_n 几乎必然收敛到有限随机变量 X_∞ 。然而在 $\{T = \infty\}$ 上, 对每个 n 都有 $|X_{n+1} - X_n| = |Y_{n+1} - Y_n| = 1$, 这与收敛矛盾。因此 $T < \infty$ 几乎必然。结合对称性, 这也证明了上一节最后一个例子中用到的性质。此例中 $X_\infty = 0$ 几乎必然, 所以尽管 (X_n) 是鞅, $X_n \geq \mathbb{E}[X_\infty | \mathcal{F}_n] = 0$ 通常并非等式。它还说明上述几

¹校订注: 原讲义在使用 Fatou 引理之前便说极限在 $\mathbb{R}$ 中; 上穿次数论证此时只排除了振荡, 严格地说先得到扩充实数值极限, 再由上式排除 $\pm\infty$ 。

乎必然收敛未必是 L^1 收敛：对每个 n , $\mathbb{E}[X_n] = 1$, 而 $\mathbb{E}[X_\infty] = 0$ 。² (2) 分枝过程。设 μ 是 $\mathbb{N}$ 上的概率测度, 且

$$m = \sum_{k=1}^{\infty} k \mu(k) < \infty.$$

排除 $\mu = \delta_0$ 与 $\mu = \delta_1$ 两个退化情形。取一族相互独立、同服从 μ 的随机变量 $(\xi_{n,j})_{n,j \in \mathbb{N}}$, 再固定整数 $\ell \geq 1$, 递归定义取值于 $\mathbb{N}$ 的随机变量列 (X_n) :

$$\begin{aligned} X_0 &= \ell \\ X_{n+1} &= \sum_{j=1}^{X_n} \xi_{n,j}, \quad \forall n \in \mathbb{N}. \end{aligned}$$

X_n 表示种群第 n 代的个体数; 每个个体的子代数服从 μ , 且不同个体的子代数相互独立。则 $(m^{-n} X_n)$ 关于下列滤过是鞅:

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_n = \sigma(\xi_{k,j} : k < n, j \in \mathbb{N}) \quad (n \geq 1).$$

容易看出 (X_n) 适应于该滤过, 因为 X_n 的定义只用到 $k < n$ 时的 $\xi_{k,j}$ 。又对每个 $n \geq 0$,

$$E[X_{n+1} | \mathcal{F}_n] = E\left[\sum_{j=1}^{\infty} \mathbf{1}_{\{j \leq X_n\}} \xi_{n,j} \mid \mathcal{F}_n\right] = \sum_{j=1}^{\infty} \mathbf{1}_{\{j \leq X_n\}} E[\xi_{n,j} | \mathcal{F}_n] = m X_n$$

这是因为 $\xi_{n,j}$ 与 $\mathcal{F}_n$ 独立, 所以 $E[\xi_{n,j} | \mathcal{F}_n] = E[\xi_{n,j}] = m$ 。于是

$$E[m^{-(n+1)} X_{n+1} | \mathcal{F}_n] = m^{-n} X_n.$$

这首先说明 $X_n \in L^1$; 事实上, 简单递推给出 $E[X_n] = \ell m^n$ 。同时, $(m^{-n} X_n)$ 是正鞅。以下分三种情形讨论。

(1) $m < 1$ 。由于 X_n 取整数值, 而 $m^{-n} X_n$ 收敛到有限值, 故充分大的 n 必有 $X_n = 0$, 即种群几乎必然灭绝。

(2) $m = 1$ 。此时 (X_n) 本身是正鞅。只要验证

$$P(\exists N \geq 1, p \geq 1 : \forall n \geq N, X_n = p) = 0.$$

由 Borel–Cantelli 引理容易得到这一点; 这里使用了 $\mu(1) < 1$ 。(3) $m > 1$ 。此时

$$m^{-n} X_n \xrightarrow{n \rightarrow \infty} Z$$

在 $\{Z > 0\}$ 上, X_n 在 n 很大时具有 m^n 的数量级。还希望证明 $\mathbb{P}(Z > 0) > 0$, 并证明在非灭绝事件 $\{\liminf_n X_n > 0\}$ 上几乎必然有 $Z > 0$ 。若 (12.3) 的收敛在 L^1 中成立, 则 $\mathbb{P}(Z > 0) > 0$, 因为

$$\mathbb{E}[Z] = \lim_{n \rightarrow \infty} m^{-n} \mathbb{E}[X_n] = \ell.$$

Kesten–Stigum 定理说明, (12.3) 在 L^1 中收敛当且仅当

$$\sum_{k=1}^{\infty} k \log(k) \mu(k) < \infty$$

²校订注: 原讲义在路径增量等式中有一处左定界符误排; 此处统一写为绝对值。

此时在非灭绝事件上几乎必然有 $Z > 0$ 。第 12.4 节将证明一个稍弱的结果。若 (X_n) 是 L^1 中有界的鞅，定理 12.3.3 保证 X_n 几乎必然收敛到 X_∞ ，但前面的例子说明未必在 L^1 中收敛。下一个定理刻画恰好具有 L^1 收敛性的鞅。

定理 12.3.5 设 $(X_n)_{n \in \mathbb{N}}$ 为鞅。下列条件等价：

(i) X_n 几乎必然且在 L^1 中收敛到 X_∞ ；

(ii) 存在 $Z \in L^1(\Omega, \mathcal{F}, P)$ ，使得对每个 $n \in \mathbb{N}$ ， $X_n = E[Z | \mathcal{F}_n]$ 。

若这些条件成立，则 (ii) 中可以取 $Z = X_\infty$ 。此时称鞅 (X_n) 为闭鞅。

证明。 首先假设 (i)。写出

$$X_n = E[X_m | \mathcal{F}_n], \quad \forall m > n$$

再利用映射 $Y \mapsto E[Y | \mathcal{F}_n]$ 是 L^1 上的压缩，即 $E[|E[Y | \mathcal{F}_n]|] \leq E[|Y|]$ ，令 $m \rightarrow \infty$ ，得到 $X_n = E[X_\infty | \mathcal{F}_n]$ 。反过来，设 (ii) 成立。则 (X_n) 在 L^1 中有界，因而由定理 12.3.3 几乎必然收敛。为证明 L^1 收敛，先设 $|Z| \leq K < \infty$ 。此时所有 $|X_n| \leq K$ ，由支配收敛定理即得结论。一般情形下，固定 $\varepsilon > 0$ ，取充分大的 $M > 0$ ，使

$$E[|Z - Z \mathbf{1}_{\{|Z| \leq M\}}|] < \varepsilon.$$

于是对每个 n ，

$$E[|X_n - E[Z \mathbf{1}_{\{|Z| \leq M\}} | \mathcal{F}_n]|] = E[|E[Z - Z \mathbf{1}_{\{|Z| \leq M\}} | \mathcal{F}_n]|] < \varepsilon.$$

由有界情形，鞅 $(E[Z \mathbf{1}_{\{|Z| \leq M\}} | \mathcal{F}_n])$ 在 L^1 中收敛。因此可取 n_0 充分大，使得对所有 $m, n \geq n_0$ ，

$$E[|E[Z \mathbf{1}_{\{|Z| \leq M\}} | \mathcal{F}_m] - E[Z \mathbf{1}_{\{|Z| \leq M\}} | \mathcal{F}_n]|] < \varepsilon.$$

结合前两个估计，对所有 $m, n \geq n_0$ ，

$$E[|X_m - X_n|] < 3\varepsilon.$$

由于 $\varepsilon > 0$ 任意， (X_n) 是 L^1 中的 Cauchy 列，故在 L^1 中收敛。

推论 12.3.6 设 $Z \in L^1(\Omega, \mathcal{F}, P)$ 。则鞅 $(E[Z | \mathcal{F}_n])_{n \in \mathbb{N}}$ 几乎必然且在 L^1 中收敛到

$$X_\infty = E[Z | \mathcal{F}_\infty], \quad \mathcal{F}_\infty = \bigvee_{n=1}^{\infty} \mathcal{F}_n.$$

证明。 由前一定理，只需证明 $X_\infty = E[Z | \mathcal{F}_\infty]$ 。首先， X_∞ 是 $\mathcal{F}_\infty$ -可测的，因为每个 X_n 都是。其次，对任意 $n \in \mathbb{N}$ 和 $A \in \mathcal{F}_n$ ，

$$E[Z \mathbf{1}_A] = E[X_n \mathbf{1}_A] = E[X_\infty \mathbf{1}_A].$$

一个简单的单调类论证（见定理 1.4.1）表明：等式 $E[Z \mathbf{1}_A] = E[X_\infty \mathbf{1}_A]$ 已知对 $A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n$ 成立，因而也对 $A \in \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n) = \mathcal{F}_\infty$ 成立。条件期望的刻画性质随即给出结论。

例（由二进滤过得到 Lebesgue 分解）。 取 $\Omega = [0, 1)$ ， $\mathcal{F} = \mathcal{B}([0, 1))$ ， $P = \lambda$ ，其中 λ 是 Lebesgue 测度。考虑二进滤过

$$\mathcal{F}_n = \sigma\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right) : i = 1, \dots, 2^n\right).$$

设 μ 是 $[0, 1)$ 上的有限正测度，并令

$$f_n(\omega) = \frac{d\mu}{d\lambda} \Big|_{\mathcal{F}_n}(\omega) = \sum_{i=1}^{2^n} \frac{\mu([(i-1)2^{-n}, i2^{-n}))}{2^{-n}} 1_{[(i-1)2^{-n}, i2^{-n})}(\omega).$$

前面已经说明 (f_n) 是非负鞅，因此由推论 12.3.4,

$$f_n \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} f_\infty, \quad \int f_\infty d\lambda < \infty.$$

并且 $f_n \geq E[f_\infty | \mathcal{F}_n]$ 。所以对每个 $A \in \mathcal{F}_n$,

$$\mu(A) = \int f_n 1_A d\lambda \geq \int E[f_\infty | \mathcal{F}_n] 1_A d\lambda = \int f_\infty 1_A d\lambda.$$

二进区间指示函数的非负线性组合，对 $[0, 1)$ 上任意有限测度 γ ，在 $L_+^1([0, 1), \gamma)$ 中稠密。因此前一不等式推广为：对每个有界非负可测函数 g ,

$$\int g d\mu \geq \int g f_\infty d\lambda.$$

于是 $\nu = \mu - f_\infty \cdot \lambda$ 是 $[0, 1)$ 上的正测度。下面证明 $\nu \perp \lambda$ 。对每个 $n \geq 0$ ，令

$$h_n = \frac{d\nu}{d\lambda} \Big|_{\mathcal{F}_n} = f_n - E[f_\infty | \mathcal{F}_n].$$

这里 $\mathcal{F}_\infty = \mathcal{F}$ ，故推论 12.3.6 给出 $E[f_\infty | \mathcal{F}_n] \rightarrow f_\infty$ 几乎必然。所以 $h_n \rightarrow 0$ ， λ -几乎处处，即

$$\lambda\left(\left\{x \in [0, 1) : \limsup_{n \rightarrow \infty} h_n(x) > 0\right\}\right) = 0. \quad (12.4)$$

另一方面，对任意 $\varepsilon > 0$,

$$\nu(\{h_n \leq \varepsilon\}) = \int 1_{\{h_n \leq \varepsilon\}} h_n d\lambda \leq \varepsilon.$$

因此

$$\nu\left(\left\{x \in [0, 1) : \limsup_{n \rightarrow \infty} h_n(x) < \varepsilon\right\}\right) \leq \nu\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{h_n \leq \varepsilon\}\right) \leq \varepsilon.$$

令 $\varepsilon \downarrow 0$ ，得到

$$\nu\left(\left\{x \in [0, 1) : \limsup_{n \rightarrow \infty} h_n(x) = 0\right\}\right) = 0.$$

与 (12.4) 比较可知， λ 与 ν 集中在两个互不相交的 Borel 集上，即 $\nu \perp \lambda$ 。所以

$$\mu = f_\infty \cdot \lambda + \nu$$

正是 μ 关于 Lebesgue 测度的 Lebesgue 分解：第一项绝对连续，第二项与 λ 奇异。此外， $\mu \ll \lambda \iff \nu = 0$ ，这又等价于鞅 (f_n) 是闭鞅。

12.4 $p > 1$ 时的 L^p 收敛

现在研究鞅 (X_n) 在什么条件下于 $p > 1$ 时在 L^p 中收敛。这将导出关于上确界 $\sup_{n \in \mathbb{N}} X_n$ 取大值概率的重要估计。

引理 12.4.1 设 $(X_n)_{n \in \mathbb{N}}$ 是下鞅, S, T 是有界停时, 且 $S \leq T$. 则

$$\mathbb{E}[X_S] \leq \mathbb{E}[X_T].$$

注. $S = 0$ 的情形已经包含在定理 12.2.4 中。

证明. 已知 $X_S, X_T \in L^1$. 对 $n \geq 1$, 令

$$H_n = 1_{\{S < n \leq T\}} = 1_{\{S \leq n-1\}} - 1_{\{T \leq n-1\}}.$$

则 (H_n) 非负且可预测。取整数 N 使 $S \leq T \leq N$, 便有

$$(H \cdot X)_N = X_T - X_S.$$

由命题 12.1.2, $(H \cdot X)$ 是从 0 出发的下鞅, 故

$$\mathbb{E}[X_T - X_S] = \mathbb{E}[(H \cdot X)_N] \geq 0.$$

定理 12.4.2 (Doob 最大不等式) 设 $(X_n)_{n \in \mathbb{N}}$ 是下鞅。则对任意 $a > 0$ 和 $n \in \mathbb{N}$,

$$aP\left(\sup_{0 \leq k \leq n} X_k \geq a\right) \leq E\left[X_n 1_{\{\sup_{0 \leq k \leq n} X_k \geq a\}}\right] \leq E[X_n^+].$$

证明. 引入停时

$$T = \inf\{n \geq 0 : X_n \geq a\}.$$

令

$$A = \left\{ \sup_{0 \leq k \leq n} X_k \geq a \right\}$$

则 $A = \{T \leq n\}$ 。把引理 12.4.1 应用于有界停时 $T \wedge n$ 与常值停时 n , 得到

$$E[X_{T \wedge n}] \leq E[X_n]$$

另一方面, 在 A 上 $X_{T \wedge n} \geq a$, 在 A^c 上 $T > n$, 故

$$X_{T \wedge n} \geq a 1_A + X_n 1_{A^c}.$$

合并以上两个不等式, 得到

$$E[X_n] \geq aP(A) + E[X_n 1_{A^c}]$$

移去等式两边共同的 $E[X_n 1_{A^c}]$, 得到 $aP(A) \leq E[X_n 1_A]$ 。又因 $X_n 1_A \leq X_n^+$, 第二个不等式随即成立。

命题 12.4.3 设 $p > 1$, $(X_n)_{n \in \mathbb{N}}$ 是非负下鞅, 并令

$$\tilde{X}_n = \sup_{0 \leq k \leq n} X_k.$$

则对每个 $n \geq 0$,

$$E[(\tilde{X}_n)^p] \leq \left(\frac{p}{p-1}\right)^p E[(X_n)^p].$$

因此, 若 $(Y_n)_{n \in \mathbb{N}}$ 是鞅, 并令

$$Y_n^* = \sup_{0 \leq k \leq n} |Y_k|$$

则对每个 $n \geq 0$,

$$E[(Y_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p E[|Y_n|^p].$$

证明。第二条是把第一条应用于非负下鞅 $X_n = |Y_n|$ 得到的。证明第一条时, 可假设 $E[X_n^p] < \infty$, 否则结论平凡。对每个 $0 \leq k \leq n$, 条件 Jensen 不等式给出

$$\mathbb{E}[X_k^p] \leq \mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_k]^p] \leq \mathbb{E}[\mathbb{E}[X_n^p | \mathcal{F}_k]] = \mathbb{E}[X_n^p]. \quad (12.5)$$

³ 因此还有 $E[(\tilde{X}_n)^p] < \infty$ 。由定理 12.4.2, 对每个 $a > 0$ 都有

$$a P(\tilde{X}_n \geq a) \leq E[X_n \mathbf{1}_{\{\tilde{X}_n \geq a\}}].$$

把不等式两端乘以 a^{p-2} , 再对 $a \in (0, \infty)$ 积分。左端为

$$\int_0^\infty a^{p-1} P(\tilde{X}_n \geq a) da = \mathbb{E} \left[\int_0^{\tilde{X}_n} a^{p-1} da \right] = \frac{1}{p} \mathbb{E}[(\tilde{X}_n)^p].$$

这里使用了 Fubini-Tonelli 定理。右端同理为

$$\begin{aligned} \int_0^\infty a^{p-2} \mathbb{E}[X_n \mathbf{1}_{\{\tilde{X}_n \geq a\}}] da &= \mathbb{E} \left[X_n \int_0^{\tilde{X}_n} a^{p-2} da \right] \\ &= \frac{1}{p-1} \mathbb{E}[X_n (\tilde{X}_n)^{p-1}] \\ &\leq \frac{1}{p-1} \mathbb{E}[X_n^p]^{1/p} \mathbb{E}[(\tilde{X}_n)^p]^{(p-1)/p}. \end{aligned}$$

其中最后一步使用 Hölder 不等式。因此

$$\frac{1}{p} \mathbb{E}[(\tilde{X}_n)^p] \leq \frac{1}{p-1} \mathbb{E}[X_n^p]^{1/p} \mathbb{E}[(\tilde{X}_n)^p]^{(p-1)/p}.$$

约去有限的因子 $\mathbb{E}[(\tilde{X}_n)^p]^{(p-1)/p}$, 便得到命题的第一条不等式。

对任意随机过程 $(X_n)_{n \in \mathbb{N}}$, 以后记

$$X_\infty^* = \sup_{n \in \mathbb{N}} |X_n|.$$

³校订注: 原讲义在条件 Jensen 不等式的第二个条件期望中误排为 $\mathcal{F}_n$; 此处应始终以 $\mathcal{F}_k$ 为条件。

定理 12.4.4 设 $(X_n)_{n \in \mathbb{N}}$ 是鞅。假设存在 $p > 1$, 使

$$\sup_{n \in \mathbb{N}} E[|X_n|^p] < \infty.$$

则 X_n 几乎必然且在 L^p 中收敛到某个随机变量 X_∞ , 并且

$$E[|X_\infty|^p] = \sup_{n \in \mathbb{N}} E[|X_n|^p]$$

此外,

$$E[(X_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p E[|X_\infty|^p].$$

证明。 该鞅在 L^1 中有界, 所以定理 12.3.3 已给出 $X_n \rightarrow X_\infty$ 几乎必然成立。又由命题 12.4.3, 对每个 $n \in \mathbb{N}$,

$$E[(X_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{k \in \mathbb{N}} E[|X_k|^p].$$

令 $n \uparrow \infty$, 由单调收敛定理,

$$E[(X_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{k \in \mathbb{N}} E[|X_k|^p] < \infty$$

故 $X_\infty^* \in L^p$ 。因为 $|X_n| \leq X_\infty^*$, 支配收敛定理给出 $X_n \rightarrow X_\infty$ 于 L^p 。最后, 式 (12.5) 说明 $(E[|X_n|^p])$ 单调递增, 所以

$$E[|X_\infty|^p] = \lim_{n \rightarrow \infty} E[|X_n|^p] = \sup_{n \in \mathbb{N}} E[|X_n|^p].$$

例。 回到上一节的分支过程 $(X_n)_{n \in \mathbb{N}}$ 。设初始种群大小为 $X_0 = \ell$, 并假设繁殖分布 μ 满足

$$m = \sum_{k=0}^{\infty} k \mu(k) \in (1, \infty)$$

以及

$$\sum_{k=0}^{\infty} k^2 \mu(k) < \infty.$$

再令 $\sigma^2 = \text{Var}(\mu) = \sum_{k=0}^{\infty} k^2 \mu(k) - m^2$ 。前面已经证明 $(m^{-n} X_n)$ 是鞅。下面验证它在 L^2 中有界。直接计算得

$$\begin{aligned} \mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] &= \mathbb{E} \left[\sum_{j,k=1}^{\infty} \mathbf{1}_{\{j \leq X_n, k \leq X_n\}} \xi_{n,j} \xi_{n,k} \mid \mathcal{F}_n \right] \\ &= \sum_{j,k=1}^{\infty} \mathbf{1}_{\{j \leq X_n, k \leq X_n\}} \mathbb{E}[\xi_{n,j} \xi_{n,k}] \\ &= \sum_{j,k=1}^{\infty} \mathbf{1}_{\{j \leq X_n, k \leq X_n\}} (m^2 + \sigma^2 \mathbf{1}_{\{j=k\}}) \\ &= m^2 X_n^2 + \sigma^2 X_n. \end{aligned}$$

取全期望，得到

$$\mathbb{E}[X_{n+1}^2] = m^2 \mathbb{E}[X_n^2] + \ell \sigma^2 m^n.$$

令 $a_n = m^{-2n} \mathbb{E}[X_n^2]$ ，则

$$a_{n+1} = a_n + \ell \sigma^2 m^{-n-2}.$$

由于 $m > 1$ ，级数 $\sum_n m^{-n-2}$ 收敛，故 (a_n) 收敛。于是鞅 $(m^{-n} X_n)$ 在 L^2 中有界。由定理 12.4.4，它在 L^2 中收敛于某个 Z 。特别地， $\mathbb{E}[Z] = \mathbb{E}[X_0] = \ell$ ，从而 $\mathbb{P}(Z > 0) > 0$ 。事实上还可证明：在种群不灭事件上， $Z > 0$ 几乎必然成立。

12.5 一致可积性与鞅

定义 12.5.1 若随机变量族 $(X_i)_{i \in I} \subset L^1(\Omega, \mathcal{F}, P)$ 满足下式，则称它一致可积：

$$\lim_{a \rightarrow +\infty} \left(\sup_{i \in I} \mathbb{E}[|X_i| \mathbf{1}_{\{|X_i| > a\}}] \right) = 0.$$

显然，一族一致可积的随机变量在 L^1 中是有界的：只需选择足够大的 a 使得

$$\left(\sup_{i \in I} \mathbb{E}[|X_i| \mathbf{1}_{\{|X_i| > a\}}] \right) \leq 1$$

于是 $\sup_i \mathbb{E}[|X_i|] \leq a + 1$ 。反之， L^1 中有界的族未必一致可积。下面给出几个基本例子。

- (1) 单点族一致可积；更一般地， L^1 的每个有限子集都一致可积。
- (2) 若 $Z \in L^1$ 且 $Z \geq 0$ ，则所有满足 $|X| \leq Z$ 的随机变量组成一致可积族。
- (3) 设 $\Phi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ 满足 $\Phi(x)/x \rightarrow \infty$ ($x \rightarrow \infty$)。则对任意 $C > 0$ ，

$$\{X \in L^1 : \mathbb{E}[\Phi(|X|)] \leq C\}$$

一致可积。确实，

$$\mathbb{E}[|X| \mathbf{1}_{\{|X| > a\}}] \leq \left(\sup_{x > a} \frac{x}{\Phi(x)} \right) \mathbb{E}[\Phi(|X|)].$$

- (4) 若 $p > 1$ ，则 L^p 的每个有界子集都一致可积；这是 (3) 取 $\Phi(x) = x^p$ 的特例。

下一命题解释了“一致可积”这一名称。

命题 12.5.1 设 $(X_i)_{i \in I}$ 是 $L^1(\Omega, \mathcal{F}, P)$ 中的有界族。下列两条等价：

- (i) $(X_i)_{i \in I}$ 一致可积；
- (ii) 对每个 $\varepsilon > 0$ ，存在 $\delta > 0$ ，使得只要 $A \in \mathcal{F}$ 且 $P(A) < \delta$ ，就有

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbf{1}_A] < \varepsilon.$$

证明。 (i) $\Rightarrow$ (ii)。固定 $\varepsilon > 0$ 。由一致可积性，可取 $a > 0$ ，使

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbf{1}_{\{|X_i| > a\}}] < \frac{\varepsilon}{2}.$$

令 $\delta = \varepsilon/(2a)$ 。若 $P(A) < \delta$ ，则对每个 $i \in I$ ，

$$\begin{aligned} E[|X_i|1_A] &\leq E[|X_i|1_{A \cap \{|X_i| \leq a\}}] + E[|X_i|1_{\{|X_i| > a\}}] \\ &\leq aP(A) + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

(ii) $\Rightarrow$ (i)。令 $C = \sup_{i \in I} E[|X_i|] < \infty$ 。由 Markov 不等式，

$$P(|X_i| > a) \leq \frac{C}{a} \quad (i \in I).$$

给定 $\varepsilon > 0$ ，取条件 (ii) 所对应的 $\delta > 0$ 。若 a 充分大，使 $C/a < \delta$ ，把 (ii) 应用于事件 $\{|X_i| > a\}$ ，便有

$$\sup_{i \in I} E[|X_i|1_{\{|X_i| > a\}}] < \varepsilon.$$

这正是一致可积性。

推论 12.5.2 设 $X \in L^1(\Omega, \mathcal{F}, P)$ 。当 $\mathcal{G}$ 遍历 $\mathcal{F}$ 的所有子 σ -代数时，随机变量族 $\{E[X | \mathcal{G}]\}$ 一致可积。

证明。任取 $\varepsilon > 0$ 。因为单点族 $\{X\}$ 一致可积，由命题 12.5.1 可选取 $\delta > 0$ ，使得对每个 $A \in \mathcal{F}$ ，若 $P(A) < \delta$ ，则

$$E[|X|1_A] \leq \varepsilon.$$

另一方面，对每个 $a > 0$ ，由 Markov 不等式，

$$P(|E[X | \mathcal{G}]| > a) \leq \frac{1}{a} E[|E[X | \mathcal{G}]|] \leq \frac{E[|X|]}{a}.$$

因此，只要 a 足够大，使 $E[|X|]/a < \delta$ ，就可利用条件期望的特征性质得到

$$\begin{aligned} E[|E[X | \mathcal{G}]|1_{\{|E[X | \mathcal{G}]| > a\}}] &\leq E[E[|X| | \mathcal{G}]1_{\{|E[X | \mathcal{G}]| > a\}}] \\ &= E[|X|1_{\{|E[X | \mathcal{G}]| > a\}}] < \varepsilon. \end{aligned}$$

这正说明所述随机变量族一致可积。

定理 12.5.3 设 $(X_n)_{n \in \mathbb{N}} \subset L^1$ 依概率收敛到 X_∞ 。则下列两条等价：

- (i) $X_n \rightarrow X_\infty$ 在 L^1 中成立；
- (ii) 序列 $(X_n)_{n \in \mathbb{N}}$ 一致可积。

注。支配收敛定理说明：若 $X_n \rightarrow X_\infty$ 几乎必然，并存在非负可积随机变量 Z ，使所有 $|X_n| \leq Z$ ，则 $X_n \rightarrow X_\infty$ 在 L^1 中成立。这个支配条件比一致可积更强；而本定理说明，在已知依概率收敛时，一致可积恰好是 L^1 收敛的充要条件。

证明。(i) $\Rightarrow$ (ii)。首先， (X_n) 在 L^1 中有界。给定 $\varepsilon > 0$ ，可取 N 足够大，使得对每个 $n \geq N$ ，

$$E[|X_n - X_N|] < \frac{\varepsilon}{2}.$$

有限族 $\{X_0, X_1, \dots, X_N\}$ 一致可积，因此可取 $\delta > 0$ ，使得每当事件 A 满足 $P(A) < \delta$ 时，

$$\forall n \in \{0, 1, \dots, N\}, \quad E[|X_n|1_A] < \frac{\varepsilon}{2}.$$

若 $n > N$, 则还有

$$E[|X_n|1_A] \leq E[|X_N|1_A] + E[|X_n - X_N|] < \varepsilon.$$

所以命题 12.5.1 的条件 (ii) 成立, (X_n) 一致可积。

(ii) $\Rightarrow$ (i) 由命题 12.5.1 (ii) 对一致可积性的刻画, 随机变量族 $(X_n - X_m)_{n,m \in \mathbb{N}}$ 也一致可积。因此, 固定 $\varepsilon > 0$ 后, 可取 a 足够大, 使得对所有 $m, n \in \mathbb{N}$,

$$E[|X_n - X_m|1_{\{|X_n - X_m| > a\}}] < \varepsilon.$$

于是, 对任意 $m, n \in \mathbb{N}$,

$$\begin{aligned} E[|X_n - X_m|] &\leq E[|X_n - X_m|1_{\{|X_n - X_m| \leq \varepsilon\}}] \\ &\quad + E[|X_n - X_m|1_{\{\varepsilon < |X_n - X_m| \leq a\}}] \\ &\quad + E[|X_n - X_m|1_{\{|X_n - X_m| > a\}}] \\ &\leq 2\varepsilon + aP(|X_n - X_m| > \varepsilon). \end{aligned}$$

又因为 (X_n) 依概率收敛到 X_∞ , 所以

$$P(|X_n - X_m| > \varepsilon) \leq P\left(|X_n - X_\infty| > \frac{\varepsilon}{2}\right) + P\left(|X_m - X_\infty| > \frac{\varepsilon}{2}\right) \xrightarrow{m, n \rightarrow \infty} 0.$$

因此

$$\limsup_{m, n \rightarrow \infty} E[|X_n - X_m|] \leq \varepsilon.$$

由于 ε 任意, (X_n) 在 L^1 范数下是 Cauchy 列。而 L^1 完备, 且 L^1 收敛蕴含依概率收敛; 极限的唯一性表明其 L^1 极限就是 X_∞ 。

注。作为推论, 若 (X_n) 依概率收敛, 并且对某个 $p > 1$ 在 L^p 中有界, 则 $X_n \rightarrow X_\infty$ 在 L^1 中成立; 事实上, 对每个 $q < p$, 还在 L^q 中收敛。后一个结论可把本定理应用于 $|X_n - X_\infty|^q$ 得到。

鞅的应用。把前一定理与定理 12.3.5 结合起来, 可知对鞅 $(X_n)_{n \in \mathbb{N}}$, 下列三条等价:

- (i) X_n 几乎必然且在 L^1 中收敛到 X_∞ ;
- (ii) 序列 $(X_n)_{n \in \mathbb{N}}$ 一致可积;
- (iii) 该鞅是闭鞅: 存在 $Z \in L^1(\Omega, \mathcal{F}, P)$, 使得对所有 $n \in \mathbb{N}$, 都有 $X_n = E[Z | \mathcal{F}_n]$ 。

也可直接由推论 12.5.2 从 (iii) 推出 (ii)。因此, 一致可积鞅恰好就是闭鞅。回忆在这种情况下, 对每个 n 都有 $X_n = E[X_\infty | \mathcal{F}_n]$ 。

停时定理。若适应过程 $(X_n)_{n \in \mathbb{N}}$ 几乎必然收敛到 X_∞ , 则对每个可能取无穷值的停时 T , 定义

$$X_T = \sum_{n=0}^{\infty} 1_{\{T=n\}} X_n + 1_{\{T=\infty\}} X_\infty.$$

命题 12.2.3 的直接推广表明, X_T 是 $\mathcal{F}_T$ -可测的。

定理 12.5.4 设 $(X_n)_{n \in \mathbb{N}}$ 是一致可积鞅。则对任意停时 T (允许 $T = \infty$)，

$$X_T = E[X_\infty | \mathcal{F}_T],$$

特别地，对每个 $n \in \mathbb{N}$ ， $\mathbb{E}[X_T] = \mathbb{E}[X_\infty] = \mathbb{E}[X_n]$ 。若 S, T 是停时且 $S \leq T$ ，则

$$X_S = E[X_T | \mathcal{F}_S].$$

注。(i) 由本定理和推论 12.5.2，随机变量族 $\{X_T : T \text{ 为停时}\}$ 一致可积。(ii) 对任意鞅 $(X_n)_{n \in \mathbb{N}}$ 和固定整数 $N \geq 0$ ，可把本定理应用于一致可积的停止鞅 $(X_{n \wedge N})_{n \in \mathbb{N}}$ ，从而重新得到前面的一些结果。

证明。先验证 $X_T \in L^1$ ：

$$\begin{aligned} E[|X_T|] &= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|X_n|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\ &= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|E[X_\infty | \mathcal{F}_n]|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\ &\leq \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}E[|X_\infty| | \mathcal{F}_n]] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\ &= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|X_\infty|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\ &= E[|X_\infty|] < \infty. \end{aligned}$$

再取任意 $A \in \mathcal{F}_T$ 。由于 $X_T \in L^1$ ，

$$\begin{aligned} E[\mathbf{1}_A X_T] &= \sum_{n \in \mathbb{N} \cup \{\infty\}} E[\mathbf{1}_{A \cap \{T=n\}} X_T] \\ &= \sum_{n \in \mathbb{N} \cup \{\infty\}} E[\mathbf{1}_{A \cap \{T=n\}} X_n] \\ &= \sum_{n \in \mathbb{N} \cup \{\infty\}} E[\mathbf{1}_{A \cap \{T=n\}} X_\infty] \\ &= E[\mathbf{1}_A X_\infty]. \end{aligned}$$

第一个等式利用 $X_T \in L^1$ ，从而可由 Fubini-Tonelli 定理交换求和与期望；第三个等式利用 $X_n = E[X_\infty | \mathcal{F}_n]$ 以及停时 σ -代数的定义性质 $A \cap \{T=n\} \in \mathcal{F}_n$ 。由于 X_T 是 $\mathcal{F}_T$ -可测的，上式即证明 $X_T = E[X_\infty | \mathcal{F}_T]$ 。其余断言随即得到；特别地， $\mathcal{F}_S \subset \mathcal{F}_T$ 蕴含

$$X_S = E[X_\infty | \mathcal{F}_S] = E[E[X_\infty | \mathcal{F}_T] | \mathcal{F}_S] = E[X_T | \mathcal{F}_S].$$

定理 12.5.5 设 $(X_n)_{n \in \mathbb{N}}$ 为上鞅，并假设下列条件之一成立：

(i) 对每个 $n \in \mathbb{N}$ ， $X_n \geq 0$ ；

(ii) 序列 $(X_n)_{n \in \mathbb{N}}$ 一致可积。

则对任意停时 T (允许取值 ∞)，都有 $X_T \in L^1$ 。此外，若停时 $S \leq T$ ，则：

(i) 情形下， $\mathbf{1}_{\{S < \infty\}} X_S \geq E[\mathbf{1}_{\{T < \infty\}} X_T | \mathcal{F}_S]$ ；

(ii) 情形下， $X_S \geq E[X_T | \mathcal{F}_S]$ 。

证明。先处理 (i)。若停时 T 有界，定理 12.2.4 给出 $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ 。对任意停时 T ，由 Fatou 引理，

$$\mathbb{E}[X_T] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[X_{T \wedge k}] \leq \mathbb{E}[X_0].$$

故 $X_T \in L^1$ 。再设 $S \leq T$ 。先假设 $S, T \leq N$ 。引理 12.4.1 给出 $\mathbb{E}[X_S] \geq \mathbb{E}[X_T]$ 。更一般地，任取

$A \in \mathcal{F}_S$, 定义

$$S^A(\omega) = \begin{cases} S(\omega), & \omega \in A, \\ N, & \omega \notin A, \end{cases} \quad T^A(\omega) = \begin{cases} T(\omega), & \omega \in A, \\ N, & \omega \notin A. \end{cases}$$

由于 $A \in \mathcal{F}_S \subset \mathcal{F}_T$, 二者都是停时。把 $\mathbb{E}[X_{S^A}] \geq \mathbb{E}[X_{T^A}]$ 展开, 得到

$$\forall A \in \mathcal{F}_S, \quad \mathbb{E}[X_S \mathbf{1}_A] \geq \mathbb{E}[X_T \mathbf{1}_A].$$

现在回到任意停时 $S \leq T$ 。取 $B \in \mathcal{F}_S$, 并把上述结论应用于 $S \wedge k$ 、 $T \wedge k$ 以及 $A = B \cap \{S \leq k\}$, 便有

$$\begin{aligned} \mathbb{E}[X_{S \wedge k} \mathbf{1}_{B \cap \{S \leq k\}}] &\geq \mathbb{E}[X_{T \wedge k} \mathbf{1}_{B \cap \{S \leq k\}}] \\ &\geq \mathbb{E}[X_{T \wedge k} \mathbf{1}_{B \cap \{T \leq k\}}]. \end{aligned}$$

因为 $\{T \leq k\} \subset \{S \leq k\}$, 且 $X_{T \wedge k} \geq 0$ 。所以

$$\mathbb{E}[X_S \mathbf{1}_{B \cap \{S \leq k\}}] \geq \mathbb{E}[X_T \mathbf{1}_{B \cap \{T \leq k\}}].$$

令 $k \rightarrow \infty$, 由支配收敛定理得到

$$\mathbb{E}[X_S \mathbf{1}_{B \cap \{S < \infty\}}] \geq \mathbb{E}[X_T \mathbf{1}_{B \cap \{T < \infty\}}].$$

记 $\tilde{X}_S = \mathbf{1}_{\{S < \infty\}} X_S$ 、 $\tilde{X}_T = \mathbf{1}_{\{T < \infty\}} X_T$ 。于是对每个 $B \in \mathcal{F}_S$,

$$\mathbb{E}[\tilde{X}_S \mathbf{1}_B] \geq \mathbb{E}[\tilde{X}_T \mathbf{1}_B] = \mathbb{E}[\mathbb{E}[\tilde{X}_T | \mathcal{F}_S] \mathbf{1}_B].$$

由于 $\tilde{X}_S$ 关于 $\mathcal{F}_S$ 可测, 这正说明 $\tilde{X}_S \geq \mathbb{E}[\tilde{X}_T | \mathcal{F}_S]$ 。

下面处理 (ii)。一致可积性蕴含 (X_n) 在 L^1 中有界, 故它几乎必然收敛于某个 X_∞ ; 由定理 12.5.3, 收敛也在 L^1 中成立。在上鞅不等式 $X_n \geq \mathbb{E}[X_{n+m} | \mathcal{F}_n]$ 中令 $m \rightarrow \infty$, 得到

$$X_n \geq \mathbb{E}[X_\infty | \mathcal{F}_n], \quad n \in \mathbb{N}.$$

另一方面, 闭鞅 $Z_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ 几乎必然收敛到 X_∞ (注意 X_∞ 关于 $\mathcal{F}_\infty$ 可测)。令 $Y_n = X_n - Z_n$, 则 (Y_n) 是非负上鞅, 且 $Y_\infty = 0$ 。由 (i) 以及定理 12.5.4, 先得 $X_T = Y_T + Z_T \in L^1$, 再得

$$Y_S \geq \mathbb{E}[Y_T | \mathcal{F}_S].$$

这里无须写有限性指示函数, 因为 $Y_\infty = 0$ 。再把定理 12.5.4 应用于一致可积鞅 (Z_n) , 得到

$$Z_S = \mathbb{E}[Z_T | \mathcal{F}_S].$$

把以上两式相加, 便得到

$$X_S \geq \mathbb{E}[X_T | \mathcal{F}_S].$$

例 (赌徒破产)。考虑从 $X_0 = k \geq 0$ 出发的一维简单对称随机游走。固定整数 $m \geq 1$, 且 $0 \leq k \leq m$, 令

$$T = \inf\{n \geq 0 : X_n = 0 \text{ 或 } X_n = m\}.$$

由前面的随机游走例可知 $T < \infty$ 几乎必然。停止鞅 $Y_n = X_{n \wedge T}$ 有界，因而一致可积。所以 $\mathbb{E}[Y_\infty] = \mathbb{E}[Y_0] = k$ ，即

$$m \mathbb{P}(X_T = m) = k.$$

从而

$$\mathbb{P}(X_T = m) = \frac{k}{m}, \quad \mathbb{P}(X_T = 0) = 1 - \frac{k}{m}.$$

还可处理有偏随机游走。设 $X_n = k + Y_1 + \cdots + Y_n$ ，其中 Y_i 独立同分布，且

$$\mathbb{P}(Y_1 = 1) = p, \quad \mathbb{P}(Y_1 = -1) = 1 - p =: q.$$

其中 $p \in (0, 1) \setminus \{1/2\}$ 。容易验证

$$Z_n = \left(\frac{q}{p}\right)^{X_n}$$

是鞅。仍以 T 表示首次到达 0 或 m 的时刻。由于 $Z_{n \wedge T}$ 收敛而其在 $\{T = \infty\}$ 上不可能最终保持常值，可知 $T < \infty$ 几乎必然。对有界鞅 $Z_{n \wedge T}$ 应用停时定理，得到

$$\left(\frac{q}{p}\right)^k = \mathbb{E}\left[\left(\frac{q}{p}\right)^{X_T}\right] = \left(\frac{q}{p}\right)^m \mathbb{P}(X_T = m) + \mathbb{P}(X_T = 0).$$

解这个二元方程，得到

$$\mathbb{P}(X_T = m) = \frac{(q/p)^k - 1}{(q/p)^m - 1}, \quad \mathbb{P}(X_T = 0) = \frac{(q/p)^m - (q/p)^k}{(q/p)^m - 1}.$$

12.6 反向鞅

反向滤过是由负整数指标的一族子 σ -代数 $(\mathcal{F}_n)_{n \in -\mathbb{N}}$ ，满足对任意 $m, n \in -\mathbb{N}$ ，

$$n \leq m \Rightarrow \mathcal{F}_n \subset \mathcal{F}_m.$$

记作

$$\mathcal{F}_{-\infty} = \bigcap_{n \in -\mathbb{N}} \mathcal{F}_n$$

$\mathcal{F}_{-\infty}$ 仍是 $\mathcal{F}$ 的子 σ -代数。与正向滤过不同，当 $n \rightarrow -\infty$ 时， $\mathcal{F}_n$ 越来越小。若对每个 $n \in -\mathbb{N}$ ， X_n 关于 $\mathcal{F}_n$ 可测且 $E[|X_n|] < \infty$ ，并且对任意 $m, n \in -\mathbb{N}$ 满足

$$n \leq m \implies X_n = E[X_m | \mathcal{F}_n],$$

则称 $(X_n)_{n \in -\mathbb{N}}$ 为反向鞅。把等号分别改为 $\geq$ 或 $\leq$ ，就得到反向上鞅或反向下鞅。

定理 12.6.1 设 $(X_n)_{n \in -\mathbb{N}}$ 为反向上鞅，并假设

$$\sup_{n \in -\mathbb{N}} E[|X_n|] < \infty.$$

则 $(X_n)_{n \in -\mathbb{N}}$ 一致可积；当 $n \rightarrow -\infty$ 时，它几乎必然且在 L^1 中收敛到某个 X_∞ 。此外，对每个 $n \in -\mathbb{N}$,

$$E[X_n | \mathcal{F}_{-\infty}] \leq X_\infty.$$

注。(a) 对反向鞅，条件 (12.6) 自动成立，因为 $X_n = E[X_0 | \mathcal{F}_n]$ ，从而 $E[|X_n|] \leq E[|X_0|]$ 。其一致可积性也可由推论 12.5.2 得到。(b) 正向上鞅即使在 L^1 中有界，也未必一致可积；这是反向情形与正向情形的关键差别。

证明。先由 Doob 上升次数不等式证明几乎必然收敛。固定 $K \geq 1$ ，对 $n \in \{0, 1, \dots, K\}$ 令

$$\begin{aligned} Y_n^K &= X_{-K+n}, \\ \mathcal{G}_n^K &= \mathcal{F}_{-K+n}. \end{aligned}$$

当 $n > K$ 时再令 $Y_n^K = X_0$ 、 $\mathcal{G}_n^K = \mathcal{F}_0$ 。于是 $(Y_n^K)_{n \in \mathbb{N}}$ 关于滤过 $(\mathcal{G}_n^K)$ 是上鞅。把引理 12.3.2 应用于下鞅 $(-Y_n^K)$ ，对任意 $a < b$ 得到

$$(b-a) \mathbb{E}[N_K([a, b], -Y^K)] \leq \mathbb{E}[(-Y_K^K - a)^+] = \mathbb{E}[(-X_0 - a)^+] \leq |a| + \mathbb{E}[|X_0|].$$

⁴ 容易验证，当 $K \uparrow \infty$ 时， $N_K([a, b], -Y^K)$ 单调增加到

$$\begin{aligned} N([a, b], -X) &:= \sup\{k \in \mathbb{N} : \exists m_1 < n_1 < \dots < m_k < n_k \leq 0, \\ &\quad -X_{m_1} \leq a, -X_{n_1} \geq b, \dots, -X_{m_k} \leq a, -X_{n_k} \geq b\}. \end{aligned}$$

它就是过程 $(-X_n)_{n \in -\mathbb{N}}$ 穿越区间 $[a, b]$ 的总上升次数。由单调收敛定理，

$$(b-a) E[N([a, b], -X)] \leq |a| + E[|X_0|] < \infty.$$

所以对每对有理数 $a < b$ ，几乎必然有 $N([a, b], -X) < \infty$ 。对引理 12.3.1 作同样修改，得到 (X_n) 在 $n \rightarrow -\infty$ 时几乎必然收敛；Fatou 引理还说明极限 X_∞ 满足 $E[|X_\infty|] < \infty$ 。下面证明一致可积。固定 $\varepsilon > 0$ 。由 (12.6)，递增数列 $(E[X_{-n}])_{n \in \mathbb{N}}$ 有上界，故可取充分小的整数 $K \leq 0$ ，使对每个 $n \leq K$ ，

$$E[X_n] \leq E[X_K] + \frac{\varepsilon}{2}.$$

有限族 $X_K, X_{K+1}, \dots, X_0$ 一致可积，因此可取 $a > 0$ 充分大，使对每个 $n \in \{K, K+1, \dots, 0\}$ ，

$$E[|X_n| \mathbf{1}_{\{|X_n| > a\}}] < \varepsilon.$$

还可取 $\delta > 0$ 充分小，使得只要 $A \in \mathcal{F}$ 且 $P(A) < \delta$ ，就有

$$E[|X_K| \mathbf{1}_A] < \frac{\varepsilon}{2}.$$

现在设 $n < K$ 。则

⁴校订注：原讲义在上升次数中误排为 $-Y^K$ ；本段固定的截断过程是 Y^K ，故应为 $-Y^K$ 。

$$\begin{aligned}
E[|X_n| \mathbf{1}_{\{|X_n|>a\}}] &= E[-X_n \mathbf{1}_{\{X_n < -a\}}] + E[X_n \mathbf{1}_{\{X_n > a\}}] \\
&= -E[X_n \mathbf{1}_{\{X_n < -a\}}] + E[X_n] - E[X_n \mathbf{1}_{\{X_n \leq a\}}] \\
&\leq -E[E[X_K | \mathcal{F}_n] \mathbf{1}_{\{X_n < -a\}}] + E[X_K] + \frac{\varepsilon}{2} - E[E[X_K | \mathcal{F}_n] \mathbf{1}_{\{X_n \leq a\}}] \\
&= -E[X_K \mathbf{1}_{\{X_n < -a\}}] + E[X_K] + \frac{\varepsilon}{2} - E[X_K \mathbf{1}_{\{X_n \leq a\}}] \\
&= -E[X_K \mathbf{1}_{\{X_n < -a\}}] + E[X_K \mathbf{1}_{\{X_n > a\}}] + \frac{\varepsilon}{2} \\
&\leq E[|X_K| \mathbf{1}_{\{|X_n|>a\}}] + \frac{\varepsilon}{2}.
\end{aligned}$$

第一个不等式使用了 $E[X_n] \leq E[X_K] + \varepsilon/2$ ，以及反向上鞅性质 $X_n \geq E[X_K | \mathcal{F}_n]$ 。另一方面，

$$P(|X_n| > a) \leq \frac{1}{a} E[|X_n|] \leq \frac{C}{a}$$

其中 $C = \sup_k E[|X_k|] < \infty$ 。必要时再增大 a ，可设 $C/a < \delta$ ，于是

$$E[|X_K| \mathbf{1}_{\{|X_n|>a\}}] < \frac{\varepsilon}{2}$$

并结合前述内容，

$$E[|X_n| \mathbf{1}_{\{|X_n|>a\}}] < \varepsilon,$$

这对每个 $n < K$ 成立；而对 $n \in \{K, K+1, \dots, 0\}$ ，同一不等式已由 a 的选择保证。因此 $(X_n)_{n \in -\mathbb{N}}$ 一致可积。结合几乎必然收敛，便得到 L^1 收敛。最后，对 $m \leq n$ 及 $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_m$ ，有

$$E[X_n \mathbf{1}_A] \leq E[X_m \mathbf{1}_A]$$

令 $m \rightarrow -\infty$ ，得到

$$E[X_n \mathbf{1}_A] \leq E[X_{\infty} \mathbf{1}_A], \quad \forall A \in \mathcal{F}_{-\infty}.$$

因此我们也有

$$E[E[X_n | \mathcal{F}_{-\infty}] \mathbf{1}_A] \leq E[X_{\infty} \mathbf{1}_A], \quad \forall A \in \mathcal{F}_{-\infty}.$$

因为 X_{∞} 关于 $\mathcal{F}_{-\infty}$ 可测，这恰好说明 $E[X_n | \mathcal{F}_{-\infty}] \leq X_{\infty}$ 。

推论 12.6.2 设 $Z \in L^1$ ， $(\mathcal{G}_n)_{n \in \mathbb{N}}$ 是一列递减的子 σ -代数。则

$$E[Z | \mathcal{G}_n] \xrightarrow[n \rightarrow \infty]{\text{a.s.}, L^1} E[Z | \mathcal{G}_{\infty}].$$

其中

$$\mathcal{G}_{\infty} = \bigcap_{n \in \mathbb{N}} \mathcal{G}_n.$$

证明。对每个 $n \in \mathbb{N}$ ，令 $X_{-n} = \mathbb{E}[Z | \mathcal{G}_n]$ ， $\mathcal{F}_{-n} = \mathcal{G}_n$ 。则 $(X_n)_{n \in -\mathbb{N}}$ 关于逆向滤过 $(\mathcal{F}_n)_{n \in -\mathbb{N}}$ 是鞅。定理保证当 $n \rightarrow -\infty$ 时， X_n 几乎必然且在 L^1 中收敛。又由定理最后一条，极限满足

$$X_{\infty} = \mathbb{E}[X_0 | \mathcal{F}_{-\infty}] = \mathbb{E}[\mathbb{E}[Z | \mathcal{F}_0] | \mathcal{F}_{-\infty}] = \mathbb{E}[Z | \mathcal{F}_{-\infty}].$$

应用。(A) 强大数律。设 $\xi_1, \xi_2, \dots$ 是一列属于 L^1 的独立同分布实随机变量。令 $S_0 = 0$, 并对 $n \geq 1$ 令

$$S_n = \xi_1 + \dots + \xi_n.$$

关键恒等式是

$$E[\xi_1 | S_n] = \frac{1}{n} S_n.$$

事实上, 存在可测函数 g , 使 $E[\xi_1 | S_n] = g(S_n)$ 。任取 $k \in \{1, \dots, n\}$, 随机向量 (ξ_k, S_n) 与 (ξ_1, S_n) 同分布。因此, 对任意有界可测函数 h ,

$$E[\xi_k h(S_n)] = E[\xi_1 h(S_n)] = E[g(S_n) h(S_n)].$$

这说明 $E[\xi_k | S_n] = g(S_n)$ 也成立。于是

$$ng(S_n) = E[\xi_1 + \dots + \xi_n | S_n] = S_n. \quad (12.7)$$

这就证明了式 (12.7)。另外, 对每个 $n \geq 1$, 还有

$$E[\xi_1 | S_n, \xi_{n+1}, \xi_{n+2}, \dots] = \frac{1}{n} S_n. \quad (12.8)$$

这由式 (12.7) 和下面的引理立即得到: 应用引理时取 $Z = \xi_1$, $\mathcal{H}_1 = \sigma(S_n)$ 及 $\mathcal{H}_2 = \sigma(\xi_{n+1}, \xi_{n+2}, \dots)$ 。

引理 12.6.3 设 $Z \in L^1$, 且 $\mathcal{H}_1, \mathcal{H}_2$ 是 $\mathcal{F}$ 的两个子 σ -代数。若 $\mathcal{H}_2$ 与 $\sigma(Z) \vee \mathcal{H}_1$ 独立, 则

$$E[Z | \mathcal{H}_1 \vee \mathcal{H}_2] = E[Z | \mathcal{H}_1].$$

证明。由独立性, 对形如 $A = B \cap C$ 的集合, 其中 $B \in \mathcal{H}_1, C \in \mathcal{H}_2$, 有 $E[\mathbf{1}_A Z] = E[\mathbf{1}_A E[Z | \mathcal{H}_1]]$ 。对这样的矩形事件应用单调类引理 (定理 1.4.1), 即可把该恒等式推广到所有 $A \in \mathcal{H}_1 \vee \mathcal{H}_2$ 。现在在推论 12.6.2 中取 $Z = \xi_1$, 并对 $n \geq 1$ 令

$$\mathcal{G}_n = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \dots).$$

由式 (12.8), $n^{-1} S_n = E[Z | \mathcal{G}_n]$ 。于是 $(n^{-1} S_n)$ 几乎必然且在 L^1 中收敛。该极限不受任何有限个 ξ_i 的改变影响, 因而对每个尾 σ -代数都可测; Kolmogorov 全或无律说明它几乎必然为常数。再取期望便知该常数为 $\lim_n n^{-1} E[S_n] = E[\xi_1]$ 。

(B) Hewitt–Savage 全或无律。设 $\xi_1, \xi_2, \dots$ 是一列取值于可测空间 $(E, \mathcal{E})$ 的独立同分布随机变量。映射 $\omega \mapsto (\xi_1(\omega), \xi_2(\omega), \dots)$ 是取值于 $E^{\mathbb{N}^*}$ 的随机变量; 这里的乘积 σ -代数, 是使每个坐标映射 $(x_1, x_2, \dots) \mapsto x_i$ 可测的最小 σ -代数。若可测函数 $F: E^{\mathbb{N}^*} \rightarrow \mathbb{R}$ 满足

$$F(x_1, x_2, x_3, \dots) = F(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}, \dots).$$

对 $\mathbb{N}^*$ 的每个有限支撑置换 π 都成立, 则称 F 为对称函数。所谓“有限支撑”, 是指只有有限多个 i 满足 $\pi(i) \neq i$ 。

定理 12.6.4 (Hewitt–Savage 全或无律) 若 $F: E^{\mathbb{N}^*} \rightarrow \mathbb{R}$ 是对称可测函数, 则 $F(\xi_1, \xi_2, \dots)$ 几乎必然为常数。

例。若 $\xi_1, \xi_2, \dots$ 取值于 $\mathbb{R}^d$, 考虑 d 维随机游走

$$X_n = \xi_1 + \dots + \xi_n.$$

任取 Borel 集 $B \in \mathcal{B}(\mathbb{R}^d)$, 指标随机变量

$$\mathbf{1}_{\{\text{card}\{n \geq 1: X_n \in B\} = \infty\}}$$

它是 $\xi_1, \xi_2, \dots$ 的对称函数。

因此

$$P(\text{Card}\{n \geq 1: X_n \in B\} = \infty) \in \{0, 1\}.$$

证明。不失一般性, 可设 F 有界。令

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n), \quad \mathcal{G}_n = \sigma(\xi_{n+1}, \xi_{n+2}, \dots).$$

记 $Y = F(\xi_1, \xi_2, \dots)$, 并对每个 $n \in \mathbb{N}$ 令

$$X_n = \mathbb{E}[Y | \mathcal{F}_n], \quad Z_n = \mathbb{E}[Y | \mathcal{G}_n].$$

推论 12.3.6 给出 $X_n \rightarrow \mathbb{E}[Y | \mathcal{F}_\infty] = Y$ 几乎必然且在 L^1 中成立; 推论 12.6.2 给出 $Z_n \rightarrow \mathbb{E}[Y | \mathcal{G}_\infty] = \mathbb{E}[Y]$ 几乎必然且在 L^1 中成立, 其中 $\mathcal{G}_\infty$ 由 Kolmogorov 全或无律是平凡的。因此, 对任意 $\varepsilon > 0$, 可取充分大的 n , 使

$$\mathbb{E}[|X_n - Y|] < \varepsilon, \quad \mathbb{E}[|Z_n - \mathbb{E}[Y]|] < \varepsilon. \quad (12.9)$$

另一方面, 存在可测函数 $g: E^n \rightarrow \mathbb{R}$, 使 $X_n = g(\xi_1, \dots, \xi_n)$ 。式 (12.9) 的第一个估计写成

$$\mathbb{E}[|F(\xi_1, \xi_2, \dots) - g(\xi_1, \dots, \xi_n)|] < \varepsilon.$$

由于序列 $(\xi_{n+1}, \dots, \xi_{2n}, \xi_1, \dots, \xi_n, \xi_{2n+1}, \dots)$ 与 $(\xi_1, \xi_2, \dots)$ 同分布, 上式还蕴含

$$\mathbb{E}[|F(\xi_{n+1}, \dots, \xi_{2n}, \xi_1, \dots, \xi_n, \xi_{2n+1}, \dots) - g(\xi_{n+1}, \dots, \xi_{2n})|] < \varepsilon.$$

由 F 的对称性, 前一式中出现的 F 的值就是 $F(\xi_1, \xi_2, \dots) = Y$ 。因此得到

$$\mathbb{E}[|Y - g(\xi_{n+1}, \dots, \xi_{2n})|] < \varepsilon. \quad (12.10)$$

对 $\mathcal{G}_n$ 取条件期望, 并利用条件 Jensen 不等式, 得到

$$\mathbb{E}[\mathbb{E}[Y | \mathcal{G}_n] - \mathbb{E}[g(\xi_{n+1}, \dots, \xi_{2n}) | \mathcal{G}_n]] < \varepsilon.$$

即

$$\mathbb{E}[|Z_n - g(\xi_{n+1}, \dots, \xi_{2n})|] < \varepsilon. \quad (12.11)$$

把式 (12.10)、(12.11) 与式 (12.9) 的第二个估计合并, 得到

$$\mathbb{E}[|Y - \mathbb{E}[Y]|] < 3\varepsilon.$$

由于 $\varepsilon > 0$ 任意, 必有 $\mathbb{E}[|Y - \mathbb{E}[Y]|] = 0$, 即 $Y = \mathbb{E}[Y]$ 几乎必然。

第十三章 马尔可夫链

13.1 定义与基本性质

本章始终设 E 是有限或可数集合，并赋予离散 σ -代数 $\mathcal{P}(E)$ 。 E 上的随机矩阵，是一族实数 $(Q(x, y))_{x, y \in E}$ ，满足：(i) 对每个 $x, y \in E$ ， $0 \leq Q(x, y) \leq 1$ ；

(ii) 对每个 $x \in E$ ， $\sum_{y \in E} Q(x, y) = 1$ 。

这个概念等价于从 E 到 E 的转移概率。事实上，令

$$\nu(x, A) = \sum_{y \in A} Q(x, y), \quad x \in E, A \subset E.$$

则 ν 是从 E 到 E 的转移概率（见第 11 章）。反过来，从任一这样的转移概率 ν 出发，公式 $Q(x, y) = \nu(x, \{y\})$ 定义了 E 上的随机矩阵。对每个整数 $n \geq 1$ ，可定义矩阵幂 Q^n ：先令 $Q^1 = Q$ ，再递归地令

$$Q^{n+1}(x, y) = \sum_{z \in E} Q^n(x, z)Q(z, y).$$

容易验证， Q^n 仍是 E 上的随机矩阵。另令 $Q^0(x, y) = 1_{\{x=y\}}$ 。对任意函数 $f: E \rightarrow \mathbb{R}_+$ ，记 Qf 为函数

$$Qf(x) = \sum_{y \in E} Q(x, y)f(y).$$

定义 13.1.1 设 Q 是 E 上的随机矩阵， $(X_n)_{n \in \mathbb{N}}$ 是取值于 E 的随机过程。若对每个 $n \geq 0$ ，给定 $(X_0, X_1, \dots, X_n)$ 后， X_{n+1} 的条件分布为 $Q(X_n, \cdot)$ ，就称 (X_n) 是转移矩阵为 Q 的马尔可夫链。等价地，

$$P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = Q(x_n, y)$$

对所有 $x_0, x_1, \dots, x_n, y \in E$ ，只要 $P(X_0 = x_0, \dots, X_n = x_n) > 0$ ，上式就成立。

注。(i) 一般地，给定 $X_0, \dots, X_n$ 后， X_{n+1} 的条件分布可能依赖整个过去。马尔可夫性质断言：这里它只依赖当前状态 X_n 。也就是说，在预测下一时刻时，已知现在以后，过去不再提供额外信息。(ii) $Q(x, \cdot)$ 不依赖时刻 n ，这称为齐次性。若从时刻 n 到 $n+1$ 的转移机制还依赖 n ，则得到非齐次马尔可夫链。

命题 13.1.1 取值于 E 的过程 $(X_n)_{n \in \mathbb{N}}$ 是转移矩阵为 Q 的马尔可夫链，当且仅当对每个 $n \geq 0$ 及任意 $x_0, \dots, x_n \in E$ ，

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0)Q(x_0, x_1)Q(x_1, x_2) \cdots Q(x_{n-1}, x_n). \quad (13.1)$$

特别地, 若 $\mathbb{P}(X_0 = x_0) > 0$, 则

$$P(X_n = x_n \mid X_0 = x_0) = Q_n(x_0, x_n).$$

证明。 若 (X_n) 是转移矩阵为 Q 的马尔可夫链, 对 n 作归纳, 并在归纳步骤中写成

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_{n+1} = x_{n+1}) &= \\ &= P(X_0 = x_0, \dots, X_n = x_n) \times P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n). \end{aligned}$$

反过来, 若式 (13.1) 成立, 则只要条件事件具有正概率, 就有

$$\begin{aligned} \mathbb{P}(X_{n+1} = y \mid X_0 = x_0, \dots, X_n = x_n) &= \\ &= \frac{\mathbb{P}(X_0 = x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n)Q(x_n, y)}{\mathbb{P}(X_0 = x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n)} = Q(x_n, y). \end{aligned}$$

最后一个结论来自恒等式

$$Q_n(x_0, x_n) = \sum_{x_1, x_2, \dots, x_{n-1} \in E} Q(x_0, x_1)Q(x_1, x_2) \cdots Q(x_{n-1}, x_n).$$

注。 公式 (13.1) 表明, 马尔可夫链 (X_n) 的有限维分布完全由初始分布与转移矩阵 Q 决定。下面的命题汇集若干基本性质。在 (ii) 中用 $\mathbb{P}(A \mid Z)$ 表示条件期望 $\mathbb{E}[\mathbf{1}_A \mid Z]$ 。

命题 13.1.2 设 $(X_n)_{n \in \mathbb{N}}$ 是转移矩阵为 Q 的马尔可夫链。

(i) 对任意整数 $n \geq 0$ 与任意函数 $f: E \rightarrow \mathbb{R}_+$,

$$E[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = E[f(X_{n+1}) \mid X_n] = Qf(X_n).$$

更一般地, 若 $\{i_1, \dots, i_k\} \subset \{0, \dots, n-1\}$, 则

$$E[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_k}, X_n] = Qf(X_n).$$

(ii) 对任意整数 $n \geq 0, p \geq 1$ 以及任意 $y_1, \dots, y_p \in E$,

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0, \dots, X_n) = Q(X_n, y_1)Q(y_1, y_2) \cdots Q(y_{p-1}, y_p).$$

因此

$$P(X_{n+p} = y_p \mid X_n) = Q_p(X_n, y_p).$$

固定 n , 若对每个 $p \in \mathbb{N}$ 令 $Y_p = X_{n+p}$, 则 $(Y_p)_{p \in \mathbb{N}}$ 仍是转移矩阵为 Q 的马尔可夫链。

证明。 (i) 由定义,

$$E[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = \sum_{y \in E} Q(X_n, y)f(y) = Qf(X_n).$$

再设 $\{i_1, \dots, i_k\} \subset \{0, \dots, n-1\}$ 。利用条件期望的塔式性质,

$$\begin{aligned} \mathbb{E}[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_k}, X_n] &= \mathbb{E}[\mathbb{E}[f(X_{n+1}) \mid X_0, \dots, X_n] \mid X_{i_1}, \dots, X_{i_k}, X_n] \\ &= \mathbb{E}[Qf(X_n) \mid X_{i_1}, \dots, X_{i_k}, X_n] = Qf(X_n). \end{aligned}$$

(ii) 式 (13.1) 立即给出

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0 = x_0, \dots, X_n = x_n) = Q(x_n, y_1)Q(y_1, y_2) \cdots Q(y_{p-1}, y_p).$$

对 $y_1, \dots, y_{p-1}$ 的所有可能取值求和, 便得到 $\mathbb{P}(X_{n+p} = y_p \mid X_n) = Q^p(X_n, y_p)$ 。至于最后一个断言, 前面的结论给出

$$P(Y_0 = y_0, Y_1 = y_1, \dots, Y_p = y_p) = P(X_n = y_0)Q(y_0, y_1)Q(y_1, y_2) \cdots Q(y_{p-1}, y_p)$$

再应用命题 13.1.1 的刻画即可。

13.2 若干例子

13.2.1 独立随机变量

若 $(X_n)_{n \in \mathbb{N}}$ 是一列取值于 E 、相互独立且同服从概率测度 μ 的随机变量, 则它是转移矩阵如下的马尔可夫链:

$$Q(x, y) = \mu(y), \quad \forall x, y \in E.$$

验证直接来自独立性; 这是一个退化但有用的例子。

13.2.2 $\mathbb{Z}^d$ 上的随机游走

设 $\eta, \xi_1, \xi_2, \dots$ 是取值于 $\mathbb{Z}^d$ 的相互独立随机变量, 并设 $\xi_1, \xi_2, \dots$ 同服从概率测度 μ 。对 $n \geq 0$, 令

$$X_n = \eta + \xi_1 + \xi_2 + \cdots + \xi_n.$$

则 $(X_n)_{n \in \mathbb{N}}$ 是马尔可夫链, 转移矩阵为

$$Q(x, y) = \mu(y - x), \quad x, y \in \mathbb{Z}^d.$$

事实上, ξ_{n+1} 与 $(X_0, \dots, X_n)$ 独立, 因此

$$\begin{aligned} P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= P(\xi_{n+1} = y - x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= P(\xi_{n+1} = y - x_n) \\ &= \mu(y - x_n). \end{aligned}$$

设 $(e_1, \dots, e_d)$ 为 $\mathbb{R}^d$ 的标准基。若对每个 $i \in \{1, \dots, d\}$ 都有

$$\mu(e_i) = \mu(-e_i) = \frac{1}{2d},$$

则所得马尔可夫链称为 $\mathbb{Z}^d$ 上的简单随机游走。

13.2.3 图上的简单随机游走

记 $\mathcal{P}_2(E)$ 为 E 的所有二元子集所成的集合, 并取 $A \subset \mathcal{P}_2(E)$ 。对每个 $x \in E$, 令

$$A_x = \{y \in E : \{x, y\} \in A\}.$$

假设每个 A_x 都有限且非空。定义 E 上的转移矩阵

$$Q(x, y) = \begin{cases} \frac{1}{\text{Card } A_x}, & \{x, y\} \in A, \\ 0, & \text{否则}. \end{cases}$$

转移矩阵为 Q 的马尔可夫链称为图 (E, A) 上的简单随机游走。

13.2.4 分枝过程

回顾上一章的定义。设 μ 是 $\mathbb{N}$ 上的概率测度, $\ell \in \mathbb{N}$ 。递归定义取值于 $\mathbb{N}$ 的过程 (X_n) :

$$\begin{aligned} X_0 &= \ell \\ X_{n+1} &= \sum_{j=1}^{X_n} \xi_{n,j}, \quad \forall n \in \mathbb{N}, \end{aligned}$$

其中 $(\xi_{n,j})_{n,j \in \mathbb{N}}$ 相互独立且同服从 μ 。则 (X_n) 是状态空间 $E = \mathbb{N}$ 上的马尔可夫链, 转移矩阵为

$$Q(x, y) = \mu^{*x}(y), \quad \forall x, y \in \mathbb{N},$$

这里 μ^{*x} 是 μ 的 x 次卷积, 也就是 x 个独立、同服从 μ 的随机变量之和的分布; 特别地, $\mu^{*0} = \delta_0$ 。由于 $(\xi_{n,j})_{j \in \mathbb{N}}$ 与 $(X_0, \dots, X_n)$ 独立,

$$\begin{aligned} P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= P\left(\sum_{j=1}^{x_n} \xi_{n,j} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\right) \\ &= P\left(\sum_{j=1}^{x_n} \xi_{n,j} = y\right) \\ &= \mu^{*x_n}(y). \end{aligned}$$

13.3 典范马尔可夫链

我们从一个关于给定转移矩阵的马尔可夫链存在性结果开始。

命题 13.3.1 设 Q 是可数集 E 上的随机转移矩阵。存在一个概率空间 $(\Omega, \mathcal{F}, P)$, 使得对每个 $x \in E$, 其上都有一个过程 $(X_n^x)_{n \in \mathbb{N}}$, 它是转移矩阵为 Q 、从 $X_0^x = x$ 出发的马尔可夫链。

证明。可取 $\Omega' = [0, 1)$, 赋予 Borel σ -代数与 Lebesgue 测度。对 $\omega \in [0, 1)$ 取其规范二进展开 (即不以无穷多个 1 结尾的展开)

$$\omega = \sum_{n=0}^{\infty} \varepsilon_n(\omega) 2^{-n-1}, \quad \varepsilon_n(\omega) \in \{0, 1\}$$

由此构造出一列相互独立、同服从 Bernoulli($\frac{1}{2}$) 分布的随机变量 $(\varepsilon_n)_{n \in \mathbb{N}}$, 即

$$\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = 0) = \frac{1}{2}.$$

若 $\varphi: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ 是单射, 则 $\eta_{i,j} = \varepsilon_{\varphi(i,j)}$ ($i, j \in \mathbb{N}$) 仍是一族相互独立、同分布的随机变量。令

$$U_i = \sum_{j=0}^{\infty} \eta_{i,j} 2^{-j-1}$$

由此得到一系列相互独立、在 $[0, 1]$ 上均匀分布的随机变量 $U_0, U_1, U_2, \dots$ 。为验证 U_i 的均匀分布, 只需注意, 对每个整数 $p \geq 0$, 有限和 $\sum_{j=0}^p \eta_{i,j} 2^{-j-1}$ 与 $\sum_{n=0}^p \varepsilon_n 2^{-n-1}$ 同分布, 再令 $p \rightarrow \infty$ 。把 E 的元素枚举为 $y_1, y_2, \dots$, 并固定 $x \in E$ 。令 $X_0^x = x$; 若

$$\sum_{1 \leq j < k} Q(x, y_j) < U_1 \leq \sum_{1 \leq j \leq k} Q(x, y_j),$$

则定义 $X_1^x = y_k$ 。

因此, 对任意 $y \in E$, 都有 $\mathbb{P}(X_1^x = y) = Q(x, y)$ 。继续递归定义: 若 $\sum_{1 \leq j < k} Q(X_n^x, y_j) < U_{n+1} \leq \sum_{1 \leq j \leq k} Q(X_n^x, y_j)$, 则 $X_{n+1}^x = y_k$ 。

利用随机变量 U_i 的独立性, 我们可以非常容易地验证, 对于所有的 $k \geq 1$,

$$\begin{aligned} & \mathbb{P}(X_{n+1}^x = y_k \mid X_0^x = x_0, \dots, X_n^x = x_n) \\ &= \mathbb{P}\left(\sum_{1 \leq j < k} Q(x_n, y_j) < U_{n+1} \leq \sum_{1 \leq j \leq k} Q(x_n, y_j) \mid X_0^x = x_0, \dots, X_n^x = x_n\right) \\ &= \mathbb{P}\left(\sum_{1 \leq j < k} Q(x_n, y_j) < U_{n+1} \leq \sum_{1 \leq j \leq k} Q(x_n, y_j)\right) = Q(x_n, y_k). \end{aligned}$$

因此, $(X_n^x)_{n \in \mathbb{N}}$ 是转移矩阵为 Q 的马尔可夫链。

后面固定一个规范概率空间来同时研究所有初始状态。取

$$\Omega = E^{\mathbb{N}}.$$

元素 $\omega \in \Omega$ 是序列 $\omega = (\omega_0, \omega_1, \dots)$, 其中 $\omega_n \in E$ 。定义坐标映射

$$X_n(\omega) = \omega_n.$$

在 Ω 上取使所有坐标映射可测的最小 σ -代数 $\mathcal{F}$ 。等价地, $\mathcal{F}$ 由下列圆柱集生成:

$$C = \{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \dots, \omega_n = x_n\}$$

其中 $n \in \mathbb{N}$, 且 $x_0, \dots, x_n \in E$ 。

引理 13.3.2 设 $(G, \mathcal{G})$ 是可测空间, $\psi: G \rightarrow \Omega$ 。则 ψ 可测, 当且仅当对每个 $n \in \mathbb{N}$, $X_n \circ \psi$ 都可测。

证明。 只需证明充分性。若每个 $X_n \circ \psi$ 都可测，则集合族

$$\{A \in \mathcal{F} : \psi^{-1}(A) \in \mathcal{G}\}$$

是 Ω 上的 σ -代数，并包含每个 $X_n^{-1}(\{y\})$ ($n \in \mathbb{N}, y \in E$)。因此它使所有坐标映射可测；由 $\mathcal{F}$ 的最小性，它就是 $\mathcal{F}$ 。

定理 13.3.3 设 Q 是 E 上的转移矩阵。对每个 $x \in E$ ，在 $\Omega = E^{\mathbb{N}}$ 上存在唯一的概率测度 $\mathbb{P}_x$ ，使坐标过程 (X_n) 在 $\mathbb{P}_x$ 下是转移矩阵为 Q 的马尔可夫链，并且 $\mathbb{P}_x(X_0 = x) = 1$ 。

证明。 取 $x \in E$ 。命题 13.3.1 允许我们在概率空间 $(\Omega', \mathcal{F}', \mathbb{P}')$ 上构造过程 $(X_n^x)_{n \in \mathbb{N}}$ ，它是转移矩阵为 Q 的马尔可夫链，且 $X_0^x = x$ 。定义 $\mathbb{P}_x$ 为 $\mathbb{P}'$ 在下列映射下的像测度：

$$\begin{aligned} \Omega' &\longrightarrow \Omega \\ \omega' &\longrightarrow (X_n^x(\omega'))_{n \in \mathbb{N}}. \end{aligned}$$

由前一引理，该映射可测。并且 $\mathbb{P}_x(X_0 = x) = \mathbb{P}'(X_0^x = x) = 1$ 。此外，对任意 $x_0, x_1, \dots, x_n \in E$ ，

$$\begin{aligned} \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}'(X_0^x = x_0, \dots, X_n^x = x_n) \\ &= \mathbb{P}'(X_0^x = x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n). \end{aligned}$$

这表明在 $\mathbb{P}_x$ 下，坐标过程是转移矩阵为 Q 的马尔可夫链（见命题 13.1.1）。关于唯一性，若 $\mathbb{P}'_x$ 是另一满足定理条件的概率测度，则 $\mathbb{P}_x$ 与 $\mathbb{P}'_x$ 在所有柱集上一致。柱集在有限交下稳定，并生成 $\mathcal{F}$ 。由单调类引理（见推论 1.4.2）， $\mathbb{P}_x = \mathbb{P}'_x$ 。**注：**（a）由命题 13.1.1 的最后一条，对任意 $n \geq 0$ 以及 $x, y \in E$ ，

$$\mathbb{P}_x(X_n = y) = Q_n(x, y).$$

（b）若 μ 是 E 上的概率测度，记

$$\mathbb{P}_\mu = \sum_{x \in E} \mu(x) \mathbb{P}_x$$

这在 Ω 上定义了一个概率测度。写出 $\mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n)$ 的显式公式，立即可见在 $\mathbb{P}_\mu$ 下， $(X_n)_{n \in \mathbb{N}}$ 是转移矩阵为 Q 的马尔可夫链，且 X_0 的分布为 μ 。（c）若 $(X'_n)_{n \in \mathbb{N}}$ 是转移矩阵为 Q 、初始分布为 μ 的马尔可夫链，则对 $\Omega = E^{\mathbb{N}}$ 中任意可测集 B ，有

$$P((X'_n)_{n \in \mathbb{N}} \in B) = \mathbb{P}_\mu(B).$$

事实上，当 B 是一个圆柱集合时，这个等式成立，然后我们可以使用与上述证明结尾相同的论证。这个等式表明，我们随后为典范马尔可夫链（由定理 13.3.3 提供的那个）建立的所有结果，都可以推广到任意具有相同转移矩阵的马尔可夫链。

典范马尔可夫链的一个重要优点，是可以使用平移算子。对每个 $k \in \mathbb{N}$ ，定义映射 $\theta_k : \Omega \rightarrow \Omega$ 为

$$\theta_k((\omega_n)_{n \in \mathbb{N}}) = (\omega_{k+n})_{n \in \mathbb{N}}.$$

引理 13.3.2 表明这些映射都是可测的。记

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$$

为 Ω 上的规范滤过；并用 $\mathbb{E}_x$ 表示概率 $\mathbb{P}_x$ 下的期望。

定理 13.3.4 (简单马尔可夫性质) 设 $F, G : \Omega \rightarrow [0, \infty]$ 可测, $n \geq 0$, 并且 F 关于 $\mathcal{F}_n$ 可测。则对每个 $x \in E$,

$$\mathbb{E}_x[F \cdot G \circ \theta_n] = \mathbb{E}_x[F \mathbb{E}_{X_n}[G]].$$

等价地,

$$\mathbb{E}_x[G \circ \theta_n | \mathcal{F}_n] = \mathbb{E}_{X_n}[G],$$

换言之, 在给定过去 $(X_0, \dots, X_n)$ 后, 移位轨道 $\theta_n(\omega)$ 的条件分布为 $\mathbb{P}_{X_n}$ 。

注。对任意初始分布 μ , 把 $\mathbb{E}_x$ 换成 $\mathbb{E}_\mu$, 结论仍成立; 下一命题也一样。

证明。只需证明第一个等式。先设

$$F = \mathbf{1}_{\{X_0=x_0, X_1=x_1, \dots, X_n=x_n\}}$$

其中 $x_0, \dots, x_n \in E$ 。再先考虑 G 也是柱事件指示函数:

$$G = \mathbf{1}_{\{X_0=y_0, X_1=y_1, \dots, X_p=y_p\}}$$

其中 $p \geq 0$, $y_0, \dots, y_p \in E$ 。此时, 对任意 $y \in E$,

$$\mathbb{E}_y[G] = \mathbf{1}_{\{y_0=y\}} Q(y_0, y_1) \cdots Q(y_{p-1}, y_p)$$

另一方面

$$\begin{aligned} \mathbb{E}_x[F G \circ \theta_n] &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n, X_n = y_0, X_{n+1} = y_1, \dots, X_{n+p} = y_p) \\ &= \mathbf{1}_{\{x_0=x\}} Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) \mathbf{1}_{\{y_0=x_n\}} Q(y_0, y_1) \cdots Q(y_{p-1}, y_p). \end{aligned}$$

于是容易得到结论。再用单调类论证, 可将结果推广到任意 $G = \mathbf{1}_A$ ($A \in \mathcal{F}$), 从而完成证明。

前一定理给出了简单马尔可夫性质的一般形式: 在已知过去 $(X_0, X_1, \dots, X_n)$ 时, 未来 $\theta_n(\omega)$ 的条件分布只依赖于当前状态 X_n 。把确定时刻 n 换成随机时刻 T 仍能成立, 这一点非常重要。为说明其作用, 考虑从某点 x 出发的链是否会无穷多次返回 x 。换言之, 记

$$N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}$$

问题是: 是否有 $\mathbb{P}_x(N_x = \infty) = 1$? 事实上, 只需验证链至少返回 x 一次。令

$$H_x = \inf\{n \geq 1 : X_n = x\}$$

仍约定 $\inf \emptyset = +\infty$ 。则

$$\mathbb{P}_x(N_x = \infty) = 1 \Leftrightarrow \mathbb{P}_x(H_x < \infty) = 1.$$

从左到右显然。反过来, 设 $\mathbb{P}_x(H_x < \infty) = 1$ 。若强马尔可夫性质已经成立, 则 $\theta_{H_x}(\omega)$ 在 $\mathbb{P}_x$ 下仍服从 $\mathbb{P}_x$ 。又有

$$N_x(\omega) = 1 + N_x(\theta_{H_x}(\omega))$$

所以 N_x 与 $1 + N_x$ 同分布；这迫使 $N_x = \infty$ 几乎必然。下面的强马尔可夫性质使这一论证严格成立。

定理 13.3.5 (强马尔可夫性质) 设 T 是滤过 $(\mathcal{F}_n)$ 的停时, $F, G : \Omega \rightarrow \mathbb{R}_+$ 可测, 并且 F 关于 $\mathcal{F}_T$ 可测。则对每个 $x \in E$,

$$\mathbb{E}_x[\mathbf{1}_{\{T < \infty\}} F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbf{1}_{\{T < \infty\}} F \mathbb{E}_{X_T}[G]].$$

等价地,

$$\mathbb{E}_x[\mathbf{1}_{\{T < \infty\}} G \circ \theta_T \mid \mathcal{F}_T] = \mathbf{1}_{\{T < \infty\}} \mathbb{E}_{X_T}[G].$$

注。 随机变量 X_T 定义在 $\{T < \infty\}$ 上, 且关于 $\mathcal{F}_T$ 可测。虽然命题 12.2.3 写的是实值过程, 其证明对可数状态空间完全相同。随机变量 $\mathbb{E}_{X_T}[G]$ 是映射 $\omega \mapsto X_T(\omega)$ 与 $y \mapsto \mathbb{E}_y[G]$ 的复合。

证明。 对每个整数 $n \geq 0$,

$$\mathbb{E}_x[\mathbf{1}_{\{T=n\}} F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbf{1}_{\{T=n\}} F \cdot G \circ \theta_n] = \mathbb{E}_x[\mathbf{1}_{\{T=n\}} F \mathbb{E}_{X_n}[G]]$$

这里使用了简单马尔可夫性质；因为 F 关于 $\mathcal{F}_T$ 可测, 所以 $\mathbf{1}_{\{T=n\}} F$ 关于 $\mathcal{F}_n$ 可测。最后对 $n \in \mathbb{N}$ 求和即可。

推论 13.3.6 设 T 是停时, $\mathbb{P}_x(T < \infty) = 1$, 并且存在 $y \in E$ 使 $\mathbb{P}_x(X_T = y) = 1$ 。则在 $\mathbb{P}_x$ 下, $\theta_T(\omega)$ 与 $\mathcal{F}_T$ 独立, 且其分布为 $\mathbb{P}_y$ 。

证明。 用定理的记号,

$$\mathbb{E}_x[F \cdot G(\theta_T(\omega))] = \mathbb{E}_x[F \mathbb{E}_{X_T}[G]] = \mathbb{E}_x[F \mathbb{E}_y[G]] = \mathbb{E}_x[F] \mathbb{E}_y[G]$$

这正是独立性及分布断言。

13.4 状态分类

从现在起, 除特别说明外均使用上一节构造的规范马尔可夫链。对 $x \in E$, 记

$$H_x = \inf\{n \geq 1 : X_n = x\}$$

$$N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}.$$

命题 13.4.1 (状态分类) 固定 $x \in E$ 。恰有下列两种情形之一:

- 或者 $\mathbb{P}_x(H_x < \infty) = 1$, 此时

$$N_x = \infty, \quad \mathbb{P}_x\text{-几乎必然,}$$

并称 x 为常返态;

- 或者 $\mathbb{P}_x(H_x < \infty) < 1$, 此时

$$N_x < \infty, \quad \mathbb{P}_x\text{-几乎必然},$$

更准确地,

$$\mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(H_x = \infty)} < \infty;$$

此时称 x 为暂态。

证明。 对每个整数 $k \geq 1$, 强马尔可夫性质给出

$$\begin{aligned} \mathbb{P}_x(N_x \geq k+1) &= \mathbb{E}_x[\mathbf{1}_{\{H_x < \infty\}} \mathbf{1}_{\{N_x \geq k\}} \circ \theta_{H_x}] \\ &= \mathbb{P}_x(H_x < \infty) \mathbb{P}_x(N_x \geq k). \end{aligned}$$

由于 $\mathbb{P}_x(N_x \geq 1) = 1$, 递推得到

$$\mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(H_x < \infty)^{k-1}.$$

若 $\mathbb{P}_x(H_x < \infty) = 1$, 则 $\mathbb{P}_x(N_x = \infty) = 1$ 。若 $\mathbb{P}_x(H_x < \infty) < 1$, 则

$$\mathbb{E}_x[N_x] = \sum_{k=1}^{\infty} \mathbb{P}_x(N_x \geq k) = \frac{1}{\mathbb{P}_x(H_x = \infty)} < \infty.$$

定义 13.4.1 链的势核是函数 $U: E \times E \rightarrow [0, \infty]$, 定义为

$$U(x, y) = \mathbb{E}_x[N_y].$$

命题 13.4.2 (i) 对任意 $x, y \in E$,

$$U(x, y) = \sum_{n=0}^{\infty} Q_n(x, y).$$

(ii) $U(x, x) = \infty$ 当且仅当 x 为常返态。(iii) 对任意互异的 $x, y \in E$,

$$U(x, y) = \mathbb{P}_x(H_y < \infty) U(y, y).$$

证明。 (i) 直接计算

$$U(x, y) = \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = y\}} \right] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=0}^{\infty} Q_n(x, y).$$

(ii) 由命题 13.4.1 与势核的定义立即得到。(iii) 由强马尔可夫性质,

$$\mathbb{E}_x[N_y] = \mathbb{E}_x[\mathbf{1}_{\{H_y < \infty\}} N_y \circ \theta_{H_y}] = \mathbb{E}_x[\mathbf{1}_{\{H_y < \infty\}} \mathbb{E}_y[N_y]] = \mathbb{P}_x(H_y < \infty) U(y, y)$$

例。 考虑 $\mathbb{Z}^d$ 上转移矩阵如下的马尔可夫链:

$$Q((x_1, \dots, x_d), (y_1, \dots, y_d)) = \frac{1}{2^d} \prod_{i=1}^d \mathbf{1}_{\{|y_i - x_i| = 1\}}$$

这是 $\mathbb{Z}^d$ 上随机游走的一个特例。从 0 出发时，它与 $(Y_n^1, \dots, Y_n^d)_{n \in \mathbb{N}}$ 同分布，其中 $Y^1, \dots, Y^d$ 是相互独立的一维简单对称随机游走。因此¹

$$Q_n(0, 0) = P(Y_n^1 = 0, \dots, Y_n^d = 0) = P(Y_n^1 = 0)^d.$$

一维简单随机游走满足：当 n 为奇数时 $\mathbb{P}(Y_n^1 = 0) = 0$ ；当 $n = 2k$ 为偶数时，简单计数给出

$$P(Y_{2k}^1 = 0) = 2^{-2k} C_{2k}^k.$$

$$U(0, 0) = \sum_{k=0}^{\infty} Q_{2k}(0, 0) = \sum_{k=0}^{\infty} (2^{-2k} C_{2k}^k)^d.$$

斯特林公式表明

$$2^{-2k} C_{2k}^k \underset{k \rightarrow \infty}{\sim} \frac{\left(\frac{2k}{e}\right)^{2k} \sqrt{4\pi k}}{2^{2k} \left(\left(\frac{k}{e}\right)^k \sqrt{2\pi k}\right)^2} \underset{k \rightarrow \infty}{\sim} \sqrt{\frac{1}{\pi k}}.$$

因此，当 $d = 1$ 或 2 时，状态 0 常返；当 $d \geq 3$ 时，状态 0 暂态。

记 R 为常返状态的集合。

引理 13.4.3 设 $x \in R$ ，且 $U(x, y) > 0$ 。则 $y \in R$ ，并且 $\mathbb{P}_y(H_x < \infty) = 1$ ；特别地， $U(y, x) > 0$ 。

证明。 先证 $\mathbb{P}_y(H_x < \infty) = 1$ 。写成

$$\begin{aligned} 0 = \mathbb{P}_x(N_x < \infty) &\geq \mathbb{P}_x(H_y < \infty, H_x \circ \theta_{H_y} = \infty) \\ &= \mathbb{E}_x[\mathbf{1}_{\{H_y < \infty\}} \mathbf{1}_{\{H_x = \infty\}} \circ \theta_{H_y}] \\ &= \mathbb{E}_x[\mathbf{1}_{\{H_y < \infty\}} \mathbb{P}_y(H_x = \infty)] \\ &= \mathbb{P}_x(H_y < \infty) \mathbb{P}_y(H_x = \infty). \end{aligned}$$

假设 $U(x, y) > 0$ 蕴含 $\mathbb{P}_x(H_y < \infty) > 0$ ，于是由前面的等式可知 $\mathbb{P}_y(H_x = \infty) = 0$ ，即 $\mathbb{P}_y(H_x < \infty) = 1$ 。接着可取整数 $n_1, n_2 \geq 1$ ，使 $Q^{n_1}(x, y) > 0$ 且 $Q^{n_2}(y, x) > 0$ 。对每个整数 $p \geq 0$ ，

$$Q^{n_2+p+n_1}(y, y) \geq Q^{n_2}(y, x) Q^p(x, x) Q^{n_1}(x, y).$$

从而

$$\begin{aligned} U(y, y) &\geq \sum_{p=0}^{\infty} Q^{n_2+p+n_1}(y, y) \\ &\geq Q^{n_2}(y, x) \left(\sum_{p=0}^{\infty} Q^p(x, x) \right) Q^{n_1}(x, y) = \infty. \end{aligned}$$

这里用到了 $x \in R$ 蕴含 $\sum_{p=0}^{\infty} Q^p(x, x) = U(x, x) = \infty$ 。所以 $y \in R$ 。作为该引理的推论，若 $x \in R$ 而 $y \in E \setminus R$ ，则 $U(x, y) = 0$ ：从常返状态不可能到达暂态状态。这个事实在下面的状态分类定理中至关重要。

¹学习提示：这里每一步的 d 个坐标都同时独立改变 ± 1 ，与每一步只选择一个坐标方向的通常 $\mathbb{Z}^d$ 简单随机游走不同。

定理 13.4.4 (状态分类) 常返状态集 R 存在一个划分 $R = \bigcup_{i \in I} R_i$, 并满足:

- 若 $x \in R_i$, 则在 P_x 下几乎必然, 对每个 $y \in R_i$ 有 $N_y = +\infty$, 而对每个 $y \in E \setminus R_i$ 有 $N_y = 0$;
- 若 $x \in E \setminus R$, 并令 $T = \inf\{n \geq 0 : X_n \in R\}$, 则在 P_x 下几乎必然, 或者 $T = \infty$ 且每个 $y \in E$ 的访问次数 $N_y < \infty$, 或者 $T < \infty$, 并存在一个随机指标 $j \in I$, 使得对所有 $n \geq T$ 都有 $X_n \in R_j$ 。

证明。 对 $x, y \in R$, 定义 $x \sim y$ 当且仅当 $U(x, y) > 0$ 。由上一引理, 这在 R 上定义了一个等价关系。传递性来自: 若 $Q^n(x, y) > 0$ 且 $Q^m(y, z) > 0$, 则 $Q^{n+m}(x, z) > 0$ 。定理中的划分就是该等价关系的等价类, 它们称为马尔可夫链的常返类。固定 $i \in I$ 与 $x \in R_i$ 。对所有 $y \in E \setminus R_i$ 都有 $U(x, y) = 0$ (当 $y \in E \setminus R$ 时用上一引理), 因而在 P_x 下 $N_y = 0$ 几乎必然。若 $y \in R_i$, 上一引理给出 $P_x(H_y < \infty) = 1$, 强马尔可夫性质又给出

$$P_x(N_y = \infty) = E_x[1_{\{H_y < \infty\}} 1_{\{N_y = \infty\}} \circ \theta_{H_y}] = P_x(H_y < \infty) P_y(N_y = \infty) = 1.$$

若 $x \in E \setminus R$ 且 $T = \infty$, 由强马尔可夫性质容易推出每个 $y \in E \setminus R$ 都只被访问有限次。若 $T < \infty$, 令 j 为满足 $X_T \in R_j$ 的随机指标; 在 T 处应用强马尔可夫性质和定理的第一部分, 便得到对所有 $n \geq T$, 都有 $X_n \in R_j$ 。

定义 13.4.2 若对任意 $x, y \in E$ 都有 $U(x, y) > 0$, 则称该链不可约。

推论 13.4.5 若马尔可夫链不可约, 则必有且仅有下列两种情形之一:

(i) 所有状态都常返, 并且只有一个常返类; 此时对每个 $x \in E$,

$$P_x(N_y = \infty, \forall y \in E) = 1.$$

(ii) 所有状态都暂态; 此时对每个 $x \in E$,

$$P_x(N_y < \infty, \forall y \in E) = 1.$$

当 E 有限时, 只可能出现第一种情形。

证明。 若存在一个常返状态, 引理 13.4.3 立即说明所有状态都常返; 又因任意 $x, y \in E$ 都有 $U(x, y) > 0$, 所以只有一个常返类。其余结论由定理 13.4.4 得到, 只需证明最后一句。若 E 有限而所有状态都暂态, 则在 P_x 下几乎必然有 $\sum_{y \in E} N_y < \infty$, 这与下式矛盾:

$$\sum_{y \in E} N_y = \sum_{y \in E} \sum_{n=0}^{\infty} 1_{\{X_n=y\}} = \sum_{n=0}^{\infty} \sum_{y \in E} 1_{\{X_n=y\}} = \infty.$$

所有状态都常返的不可约马尔可夫链称为不可约常返链。

例。 下面回到前面介绍的例子, 讨论其中的状态分类。先强调一点: 关于典范马尔可夫链得到的结果, 可立即转化为任意转移矩阵为 Q 的马尔可夫链 $(Y_n)_{n \in \mathbb{N}}$ 的结论, 反之亦然。例如, 若 $Y_0 = y$, 并记 $N_x^Y = \sum_{n=0}^{\infty} 1_{\{Y_n=x\}}$, 则对每个 $k \in \mathbb{N}$,

$$P(N_x^Y = k) = P_y(N_x = k)$$

因为左端也可以写成

$$P((Y_n)_{n \in \mathbb{N}} \in B)$$

其中 $B = \{\omega \in E^{\mathbb{N}} : N_x(\omega) = k\}$, 再使用定理 13.3.3 后的注 (b) 即可。(1) 独立同分布随机变量的情形。若共同分布为 μ , 则 $Q(x, y) = \mu(y)$ 。容易看出, y 为常返状态当且仅当 $\mu(y) > 0$, 并且只有一个常返类。该链不可约当且仅当对每个 $y \in E$, $\mu(y) > 0$ 。

(2) 在 $\mathbb{Z}$ 上的随机游走。我们有

$$Y_n = Y_0 + \sum_{i=1}^n \xi_i$$

其中 (ξ_i) 是一列取值于 $\mathbb{Z}$ 、相互独立且同服从 μ 的随机变量, 并且与 Y_0 独立。由于

$$Q(x, y) = \mu(y - x),$$

容易看出 $U(x, y)$ 只依赖于 $y - x$ 。因此所有状态类型相同: 它们要么全都常返, 要么全都暂态。

定理 13.4.6 设 $\mathbb{E}[|\xi_1|] < \infty$, 并令 $m = \mathbb{E}[\xi_1]$ 。

(i) 若 $m \neq 0$, 则所有状态都是暂态。

(ii) 若 $m = 0$, 则所有状态都常返。此外, 链不可约当且仅当由 $\{y \in \mathbb{Z} : \mu(y) > 0\}$ 生成的子群等于 $\mathbb{Z}$ 。

证明。(i) 若 $m \neq 0$, 强大数律给出 $|Y_n| \rightarrow \infty$ 几乎必然, 故所有状态都暂态。(ii) 反设 $m = 0$ 且 0 暂态, 即 $U(0, 0) < \infty$ 。以下取 $Y_0 = 0$ 。对每个 $x \in \mathbb{Z}$,

$$U(0, x) \leq U(x, x) = U(0, 0)$$

第一个不等式来自命题 13.4.2 (iii)。因此, 对每个 $n \geq 1$,

$$\sum_{|x| \leq n} U(0, x) \leq (2n+1)U(0, 0) \leq Cn, \quad n \geq 1. \quad (13.2)$$

其中 $C = 3U(0, 0) < \infty$ 。另一方面, 强大数律给出 $n^{-1}Y_n \rightarrow 0$ 几乎必然, 因而也依概率收敛。令 $\varepsilon = (4C)^{-1}$, 可取 N 足够大, 使对每个 $n \geq N$,

$$P(|Y_n| \leq \varepsilon n) > \frac{1}{2},$$

等价地,

$$\sum_{|x| \leq \varepsilon n} Q_n(0, x) > \frac{1}{2}.$$

若 $n \geq p \geq N$, 同理有

$$\sum_{|x| \leq \varepsilon n} Q_p(0, x) \geq \sum_{|x| \leq \varepsilon p} Q_p(0, x) > \frac{1}{2}$$

再对 $p = N, \dots, n$ 求和,

$$\sum_{|x| \leq \varepsilon n} U(0, x) \geq \sum_{p=N}^n \sum_{|x| \leq \varepsilon p} Q_p(0, x) > \frac{n-N}{2}.$$

另一方面, 由式 (13.2), 当 $\varepsilon n \geq 1$ 时,

$$\sum_{|x| \leq \varepsilon n} U(0, x) \leq C\varepsilon n = \frac{n}{4}.$$

当 n 足够大时, 这与前一个下界矛盾。最后证明不可约性的刻画。记 G 为由 $\{x \in \mathbb{Z} : \mu(x) > 0\}$ 生成的子群。显然

$$P(Y_n \in G, \forall n \in \mathbb{N}) = 1$$

这里仍取 $Y_0 = 0$ 。若 $G \neq \mathbb{Z}$, 链不可能不可约。反过来, 设 $G = \mathbb{Z}$, 并令

$$H = \{x \in \mathbb{Z} : U(0, x) > 0\}$$

下面验证 H 是 $\mathbb{Z}$ 的子群:

- 若 $x, y \in H$, 则不等式

$$Q_{n+p}(0, x+y) \geq Q_n(0, x) Q_p(x, x+y) = Q_n(0, x) Q_p(0, y)$$

所以 $x+y \in H$;

• 若 $x \in H$, 由于 0 常返, $U(0, x) > 0$ 由引理 13.4.3 蕴含 $U(x, 0) > 0$ 。又因 $U(x, 0) = U(0, -x)$, 故 $-x \in H$ 。

最后, 由于 H 包含 $\{x \in \mathbb{Z} : \mu(x) > 0\}$, 必有 $H = \mathbb{Z}$ 。例如, 若

$$\mu = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2,$$

则所有状态都常返, 但有两个常返类: 偶数类与奇数类。

(3) 图上的随机游走。这里只考虑有限图: E 是有限集, 且 $A \subset \mathcal{P}_2(E)$, 并满足对每个 $x \in E$, 集合

$$A_x := \{y \in E : \{x, y\} \in A\}$$

非空。若对任意 $x, y \in E$, 都存在整数 $p \geq 0$ 以及 E 中的元素 $x_0 = x, x_1, \dots, x_p = y$, 使得 $\{x_{i-1}, x_i\} \in A$ 对每个 $i \in \{1, \dots, p\}$ 成立, 则称该图连通。

命题 13.4.7 有限连通图上的简单随机游走是常返态、不可约的。

证明。 图连通立即给出链不可约, 再由推论 13.4.5 得到常返性。(4) 分枝过程。此时 $E = \mathbb{N}$, 且 $Q(x, y) = \mu^{*x}(y)$ 。状态 0 是吸收态, 因为

$$\mathbb{P}_0(\forall n \in \mathbb{N}, X_n = 0) = 1.$$

因此, 0 也是常返态。下面排除平凡情形 $\mu = \delta_1$; 在该情形下, 所有状态都是吸收态。

命题 13.4.8 状态 0 是唯一的常返状态。因此, 几乎必然有且只有下列两种情形之一:

- 或者存在 N , 使对所有 $n \geq N$ 都有 $X_n = 0$;
- 或者当 $n \rightarrow \infty$ 时 $X_n \rightarrow +\infty$ 。

注。令 $m = \sum_k k\mu(k)$ 。上一章已说明：若 $m \leq 1$ ，第一种情形以概率 1 发生；若 $m > 1$ ，第二种情形以严格正概率发生（还需假设 μ 有二阶矩）。

证明。先设 $\mu(0) > 0$ 。若 $x \geq 1$ ，则

$$U(x, 0) \geq \mathbb{P}_x(X_1 = 0) = \mu(0)^x > 0, \quad U(0, x) = 0.$$

因此 x 必为暂态。再设 $\mu(0) = 0$ 。由于排除了 $\mu = \delta_1$ ，存在 $k \geq 2$ 使 $\mu(k) > 0$ 。于是对每个 $x \geq 1$ ， $\mathbb{P}_x(X_1 > x) > 0$ ，从而存在 $y > x$ 使 $U(x, y) > 0$ 。但 $U(y, x) = 0$ ，所以 x 仍为暂态。其余结论由定理 13.4.4 得到。

13.5 不变测度

定义 13.5.1 设 μ 是 E 上非零正测度，且对每个 $x \in E$ 都有 $\mu(x) < \infty$ 。若

$$\forall y \in E, \quad \mu(y) = \sum_{x \in E} \mu(x)Q(x, y).$$

则称 μ 是 Q 的不变测度。矩阵形式为 $\mu Q = \mu$ ，迭代可得 $\mu Q^n = \mu$ ($n \in \mathbb{N}$)。^a

^a校订注：原讲义在平稳分布的解释段落中误写成 $\mu Q^n = Q$ ；正确关系是 $\mu Q^n = \mu$ 。

概率解释。再假设 $\mu(E) < \infty$ ，并归一化为 $\mu(E) = 1$ 。对任意 $f: E \rightarrow \mathbb{R}_+$ ，

$$\begin{aligned} f: E &\longrightarrow \mathbb{R}_+. \\ \mathbb{E}_\mu[f(X_1)] &= \sum_{x \in E} \mu(x) \sum_{y \in E} Q(x, y) f(y) \\ &= \sum_{y \in E} f(y) \sum_{x \in E} \mu(x) Q(x, y) = \sum_{y \in E} \mu(y) f(y). \end{aligned}$$

这说明在 $\mathbb{P}_\mu$ 下， X_1 与 X_0 同服从 μ 。由 $\mu Q^n = \mu$ ，每个 X_n 都服从 μ 。更精确地，对任意可测函数 $F: \Omega \rightarrow \mathbb{R}_+$ ，

$$\mathbb{E}_\mu[F \circ \theta_1] = \mathbb{E}_\mu[\mathbb{E}_{X_1}[F]] = \sum_{x \in E} \mu(x) \mathbb{E}_x[F] = \mathbb{E}_\mu[F]$$

所以在 $\mathbb{P}_\mu$ 下， $(X_{1+n})_{n \in \mathbb{N}}$ 与 $(X_n)_{n \in \mathbb{N}}$ 同分布；任意整数平移也同样成立。

例。对 $\mathbb{Z}^d$ 上任意随机游走 $Q(x, y) = \gamma(y - x)$ ，计数测度是不变测度。

定义 13.5.2 设 μ 是 E 上的非零正测度，且 $\mu(x) < \infty$ 对每个 $x \in E$ 成立。若

$$\forall x, y \in E, \quad \mu(x)Q(x, y) = \mu(y)Q(y, x).$$

命题 13.5.1 每个可逆测度都是不变的。

证明。若 μ 可逆，

$$\sum_{x \in E} \mu(x)Q(x, y) = \sum_{x \in E} \mu(y)Q(y, x) = \mu(y).$$

反之, 不变测度未必可逆。前面已知, 计数测度对 $\mathbb{Z}^d$ 上任意随机游走都不变; 但它仅在跳跃分布 γ 对称, 即 $\gamma(x) = \gamma(-x)$ 时可逆。

例。(a) 有偏简单随机游走。这是 $\mathbb{Z}$ 上转移矩阵如下的随机游走:

$$Q(i, i+1) = p, \quad Q(i, i-1) = q = 1 - p.$$

其中 $p \in (0, 1)$ 。容易验证, 测度

$$\mu(i) = \left(\frac{p}{q}\right)^i, \quad i \in \mathbb{Z}$$

是可逆测度, 因而是不变测度。注意, 除 $p = 1/2$ 外, 它不同于计数测度 (后者也是不变测度)。

(b) 图上的随机游走。测度

$$\mu(x) = \text{card}(A_x)$$

是可逆的。事实上, 如果 $\{x, y\} \in A$,

$$\mu(x)Q(x, y) = \text{card}(A_x) \frac{1}{\text{card}(A_x)} = 1 = \mu(y)Q(y, x).$$

(c) Ehrenfest 球盒模型。这是状态空间 $\{0, 1, \dots, k\}$ 上的马尔可夫链, 其转移矩阵满足

$$Q(j, j+1) = \frac{k-j}{k} \quad (0 \leq j \leq k-1), \quad Q(j, j-1) = \frac{j}{k} \quad (1 \leq j \leq k).$$

测度 μ 可逆, 当且仅当

$$\mu(j) \frac{k-j}{k} = \mu(j+1) \frac{j+1}{k}.$$

对每个 $0 \leq j \leq k-1$ 成立。容易验证

$$\mu(j) = C_k^j$$

适用。

定理 13.5.2 设 x 是常返状态。公式

$$\mu(y) = \mathbb{E}_x \left[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}} \right]$$

定义了一个不变测度。此外, $\mu(y) > 0$ 当且仅当 y 属于 x 所在的常返类。

证明。若 y 不属于 x 的常返类, 则 $\mathbb{E}_x[N_y] = U(x, y) = 0$, 故当然有 $\mu(y) = 0$ 。接着, 对任意 $y \in E$, 写成

$$\begin{aligned}
\mu(y) &= \mathbb{E}_x \left[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_k=y\}} \right] \\
&= \sum_{z \in E} \mathbb{E}_x \left[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_{k-1}=z, X_k=y\}} \right] \\
&= \sum_{z \in E} \sum_{k=1}^{\infty} \mathbb{E}_x [\mathbf{1}_{\{k \leq H_x, X_{k-1}=z\}} \mathbf{1}_{\{X_k=y\}}] \\
&= \sum_{z \in E} \sum_{k=1}^{\infty} \mathbb{E}_x [\mathbf{1}_{\{k \leq H_x, X_{k-1}=z\}}] Q(z, y) \\
&= \sum_{z \in E} \mathbb{E}_x \left[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_{k-1}=z\}} \right] Q(z, y) = \sum_{z \in E} \mu(z) Q(z, y).
\end{aligned}$$

第四个等式中使用了事件 $\{k \leq H_x, X_{k-1} = z\}$ 是 $\mathcal{F}_{k-1}$ -可测的，从而能在时刻 $k-1$ 应用马尔可夫性质。由此得到 $\mu Q = \mu$ ；迭代可得对每个整数 $n \geq 0$ ， $\mu Q^n = \mu$ 。特别地，

$$\mu(x) = 1 = \sum_{z \in E} \mu(z) Q_n(z, x).$$

设 y 属于 x 的常返类。则存在 $n \geq 0$ 使 $Q^n(y, x) > 0$ ，于是前式给出 $\mu(y) < \infty$ 。又存在 m 使 $Q^m(x, y) > 0$ ，并且

$$\mu(y) = \sum_{z \in E} \mu(z) Q_m(z, y) \geq Q_m(x, y) > 0.$$

注。若存在多个常返类 R_i ($i \in I$)，则可在每个 R_i 中选取一点 x_i ，并令

$$\mu_i(y) = \mathbb{E}_{x_i} \left[\sum_{k=0}^{H_{x_i}-1} \mathbf{1}_{\{X_k=y\}} \right]$$

这样便构造出支撑两两不交的不变测度。

定理 13.5.3 设链常返且不可约，则不变测度在相差一个正常数因子的意义下唯一。

证明。设 μ 为不变测度。对 $p \geq 0$ 作归纳，证明对任意 $x, y \in E$ ，

$$\mu(y) \geq \mu(x) \mathbb{E}_x \left[\sum_{k=0}^{p \wedge (H_x-1)} \mathbf{1}_{\{X_k=y\}} \right].$$

若 $y = x$ ，不等式显然成立；以下设 $y \neq x$ 。当 $p = 0$ 时式 (13.3) 显然。假设它对 p 成立，则

$$\begin{aligned}
\mu(y) &= \sum_{z \in E} \mu(z) Q(z, y) \\
&\geq \mu(x) \sum_{z \in E} \mathbb{E}_x \left[\sum_{k=0}^{p \wedge (H_x - 1)} \mathbf{1}_{\{X_k = z\}} \right] Q(z, y) \\
&= \mu(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x [\mathbf{1}_{\{X_k = z, k \leq H_x - 1\}}] Q(z, y) \\
&= \mu(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x [\mathbf{1}_{\{X_k = z, k \leq H_x - 1\}} \mathbf{1}_{\{X_{k+1} = y\}}] \\
&= \mu(x) \mathbb{E}_x \left[\sum_{k=0}^{p \wedge (H_x - 1)} \mathbf{1}_{\{X_{k+1} = y\}} \right] \\
&= \mu(x) \mathbb{E}_x \left[\sum_{k=1}^{(p+1) \wedge H_x} \mathbf{1}_{\{X_k = y\}} \right].
\end{aligned}$$

这就是 $p+1$ 阶的结论。这里利用了 $\{X_k = z, k \leq H_x - 1\} \in \mathcal{F}_k$ ，从而可在时刻 k 应用马尔可夫性质。令 $p \rightarrow \infty$ ，由单调收敛得到

$$\mu(y) \geq \mu(x) \mathbb{E}_x \left[\sum_{k=0}^{H_x - 1} \mathbf{1}_{\{X_k = y\}} \right].$$

固定 $x \in E$ ，并记

$$\nu_x(y) = \mathbb{E}_x \left[\sum_{k=0}^{H_x - 1} \mathbf{1}_{\{X_k = y\}} \right]$$

它由定理 13.5.2 知是不变测度，且 $\mu(y) \geq \mu(x) \nu_x(y)$ 对每个 $y \in E$ 成立。于是对每个 $n \geq 1$ ，

$$\mu(x) = \sum_{z \in E} \mu(z) Q_n(z, x) \geq \sum_{z \in E} \mu(x) \nu_x(z) Q_n(z, x) = \mu(x) \nu_x(x) = \mu(x),$$

首尾相等，故对每个满足 $Q^n(z, x) > 0$ 的 z ，必有 $\mu(z) = \mu(x) \nu_x(z)$ 。不可约性保证每个 $z \in E$ 都满足这一条件（对某个 n ），所以 $\mu = \mu(x) \nu_x$ 。

推论 13.5.4 设链常返且不可约，则恰有下列两种情形之一：

(i) 存在不变概率测度 μ ，并且对每个 $x \in E$ ，

$$\mathbb{E}_x[H_x] = \frac{1}{\mu(x)}.$$

(ii) 每个不变测度的总质量都为无穷，并且对每个 $x \in E$ ，

$$\mathbb{E}_x[H_x] = \infty.$$

情形 (i) 称为正常返，情形 (ii) 称为零常返。

注。若 E 有限，只会出现情形 (i)。

证明。由定理 13.5.3, 所有不变测度彼此成比例, 所以它们要么全为无限测度, 要么全为有限测度; 后一种情形可归一化为概率测度。

$$\nu_x(y) = \mathbb{E}_x \left[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}} \right],$$

在情形 (i) 中, 令 μ 为唯一的不变概率测度。由定理 13.5.3, $\mu = C\nu_x$ 。从 $1 = \mu(E) = C\nu_x(E)$ 得 $C = \nu_x(E)^{-1}$, 故

$$\mu(x) = \frac{\nu_x(x)}{\nu_x(E)} = \frac{1}{\nu_x(E)}.$$

另一方面,

$$\nu_x(E) = \sum_{y \in E} \mathbb{E}_x \left[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}} \right] = \mathbb{E}_x \left[\sum_{k=0}^{H_x-1} \left(\sum_{y \in E} \mathbf{1}_{\{X_k=y\}} \right) \right] = \mathbb{E}_x[H_x].$$

在情形 (ii) 中, $\nu_x(E) = \infty$, 由同一计算得到

$$\mathbb{E}_x[H_x] = \nu_x(E) = \infty.$$

命题 13.5.5 设链不可约。若存在有限不变测度, 则链常返, 因而正常返。

证明。设 γ 为有限不变测度, 并取 $y \in E$ 使 $\gamma(y) > 0$ 。对每个 $x \in E$, 命题 13.4.2(iii) 给出

$$\sum_{n=0}^{\infty} Q_n(x, y) = U(x, y) \leq U(y, y).$$

将不等式两端乘以 $\gamma(x)$, 再对 $x \in E$ 求和, 得到

$$\sum_{n=0}^{\infty} \gamma Q_n(y) \leq \gamma(E) U(y, y).$$

由于 γ 不变, 对每个 $n \geq 0$ 都有 $\gamma Q^n(y) = \gamma(y) > 0$ 。因此

$$\gamma(E) U(y, y) = \infty.$$

因为 $\gamma(E) < \infty$, 必有 $U(y, y) = \infty$, 所以 y 常返; 链不可约, 于是整条链常返 (推论 13.4.5)。

注。存在无限不变测度时不能据此判断常返性。例如, 有偏简单随机游走只有在 $p = 1/2$ 时常返。

例。取 $p \in (0, 1)$, 考虑状态空间 $E = \mathbb{N}$ 上转移矩阵为

$$Q(k, k+1) = p, \quad Q(k, k-1) = 1-p \quad (k \geq 1), \quad Q(0, 1) = 1.$$

该链不可约。直接验证可知, 由下式定义的测度 μ

$$\mu(k) = \left(\frac{p}{1-p} \right)^{k-1} \quad (k \geq 1), \quad \mu(0) = 1-p.$$

是可逆的, 因而不变。若 $p < 1/2$, 则 μ 有限, 命题 13.5.5 说明该链正常返。

练习：证明 $p = 1/2$ 时链零常返，而 $p > 1/2$ 时链暂态。

13.6 渐近行为

我们继续考虑与转移矩阵 Q 相关的规范马尔可夫链。

定理 13.6.1 设链常返且不可约， μ 为不变测度。设 f, g 是 E 上的非负函数，并满足 $\int f d\mu < \infty$ 与 $0 < \int g d\mu < \infty$ 。则对每个 $x \in E$ ，在 $\mathbb{P}_x$ -几乎必然意义下，

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \xrightarrow{n \rightarrow \infty} \frac{\int f d\mu}{\int g d\mu}.$$

注。即使 $\int f d\mu = \infty$ ，比值极限定理仍成立。可取 $0 \leq f_k \uparrow f$ ，且 $\int f_k d\mu < \infty$ ，再作比较。

推论 13.6.2 若链不可约且正常返， μ 是唯一的不变概率测度，则对每个 $x \in E$ ，在 $\mathbb{P}_x$ -几乎必然意义下，

$$\frac{1}{n+1} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \int f d\mu.$$

在定理中取 $g \equiv 1$ 即得推论。下面证明定理 13.6.1。先定义停时

$$T_0 = 0, \quad T_1 = H_x,$$

并递归地令

$$T_{n+1} = \inf\{k > T_n : X_k = x\}.$$

T_n 是第 n 次返回 x 的时刻。由于 x 常返，这些停时均几乎必然有限。对 $k \geq 0$ ，定义第 k 个回返周期的报酬

$$Z_k(f) = \sum_{n=T_k}^{T_{k+1}-1} f(X_n).$$

引理 13.6.3 随机变量 $Z_0(f), Z_1(f), \dots$ 相互独立且同分布。

证明。取 $\mathbb{R}_+$ 上有界可测函数 $g_0, g_1, \dots$ 。只需证明对每个 $k \geq 0$,

$$\mathbb{E}_x \left[\prod_{i=0}^k g_i(Z_i(f)) \right] = \prod_{i=0}^k \mathbb{E}_x [g_i(Z_0(f))].$$

对 k 作归纳。 $k=0$ 显然。若结论对 $k-1$ 成立，则注意：

- $Z_0(f), \dots, Z_{k-1}(f)$ 关于 $\mathcal{F}_{T_k}$ 可测；
- 由强马尔可夫性质与 $X_{T_k} = x$ ， θ_{T_k} 与 $\mathcal{F}_{T_k}$ 独立，且分布为 $\mathbb{P}_x$ ；
- 按定义， $Z_k(f) = Z_0(f) \circ \theta_{T_k}$ 。

对 $\mathcal{F}_{T_k}$ 取条件期望, 得到

$$\begin{aligned}\mathbb{E}_x \left[\prod_{i=0}^k g_i(Z_i(f)) \right] &= \mathbb{E}_x \left[\left(\prod_{i=0}^{k-1} g_i(Z_i(f)) \right) g_k(Z_0(f) \circ \theta_{T_k}) \right] \\ &= \mathbb{E}_x \left[\prod_{i=0}^{k-1} g_i(Z_i(f)) \right] \mathbb{E}_x[g_k(Z_0(f))].\end{aligned}$$

再使用归纳假设即可。

回到定理的证明。 令 ν_x 是定理 13.5.2 构造的不变测度。因为 $\nu_x(x) = 1$, 且所有不变测度彼此成比例 (定理 13.5.3), 所以 $\mu = \mu(x)\nu_x$ 。于是

$$\begin{aligned}\mathbb{E}_x[Z_0(f)] &= \mathbb{E}_x \left[\sum_{k=0}^{H_x-1} \sum_{y \in E} f(y) \mathbf{1}_{\{X_k=y\}} \right] \\ &= \sum_{y \in E} f(y) \nu_x(y) = \frac{\int f d\mu}{\mu(x)}.\end{aligned}$$

由引理 13.6.3 和强大数律, 在 P_x -几乎必然意义下,

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(f) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{\int f d\mu}{\mu(x)}. \quad (13.4)$$

对每个整数 n , 记 $N_x(n)$ 为时刻 n 以前完成的回返次数, 即 $T_{N_x(n)} \leq n < T_{N_x(n)+1}$ 。由于链常返, $N_x(n) \rightarrow \infty$ 几乎必然。又因 $f \geq 0$,

$$\frac{\sum_{k=0}^{T_{N_x(n)}-1} f(X_k)}{N_x(n)} \leq \frac{\sum_{k=0}^n f(X_k)}{N_x(n)} \leq \frac{\sum_{k=0}^{T_{N_x(n)+1}-1} f(X_k)}{N_x(n)}.$$

等价地,

$$\frac{\sum_{j=0}^{N_x(n)-1} Z_j(f)}{N_x(n)} \leq \frac{\sum_{k=0}^n f(X_k)}{N_x(n)} \leq \frac{\sum_{j=0}^{N_x(n)} Z_j(f)}{N_x(n)}.$$

结合式 (13.4) 与夹逼定理, 得到

$$\frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{\int f d\mu}{\mu(x)}.$$

对 g 重复同一论证, 再把两个极限相除, 即完成证明。

推论 13.6.4 设链常返且不可约。对每个 $x \in E$:

(i) 若链正常返, 则

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mu(x).$$

其中 μ 是唯一的不变概率测度;

(ii) 若链零常返, 则

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n \rightarrow \infty]{\text{几乎必然}} 0.$$

上述收敛对任意初始分布均成立。

定义 13.6.1 设 x 是常返状态, 并令

$$L_x = \{n \geq 0 : Q_n(x, x) > 0\}.$$

x 的周期 $d(x)$ 定义为集合 L_x 中所有整数的最大公因数。

注。 L_x 对加法封闭, 因为 $Q^{n+m}(x, x) \geq Q^n(x, x)Q^m(x, x)$ 。故 $L_x - L_x = d(x)\mathbb{Z}$ 。

命题 13.6.5 设链常返且不可约。

(i) 所有状态具有同一周期, 称为链的周期, 记为 d 。

(ii) 若 $d = 1$, 则称链无周期; 此时对任意 $x, y \in E$, 存在 n_0 , 使对所有 $n \geq n_0$ 都有 $Q^n(x, y) > 0$ 。

证明。 (i) 取 $x, y \in E$ 。由不可约性, 可取 n_1, n_2 使 $Q^{n_1}(x, y) > 0$ 、 $Q^{n_2}(y, x) > 0$ 。若 $n \in L_x$, 则 $n_1 + n + n_2 \in L_y$, 所以 $L_x - L_x \subset L_y - L_y$, 即 $d(y) \mid d(x)$ 。交换 x, y 得 $d(x) = d(y)$ 。(ii) 只需处理 $y = x$ 。由 $d(x) = 1$ 与 $L_x - L_x = \mathbb{Z}$, 可取 $n_1, m_1 \in L_x$ 使 $n_1 - m_1 = 1$, 即

$$Q^{n_1}(x, x) > 0, \quad Q^{m_1}(x, x) > 0.$$

若 $m_1 = 0$, 则 $n_1 = 1$, 结论显然。若 $m_1 \geq 1$, 则对 $j = 0, \dots, m_1 - 1$,

$$Q^{m_1+j}(x, x) \geq (Q^{n_1}(x, x))^j (Q^{m_1}(x, x))^{m_1-j} > 0.$$

故从 $n_0 = m_1^2$ 起的每个整数都属于 L_x , 即

$$Q^{n_0+j}(x, x) > 0, \quad j \geq 0.$$

定理 13.6.6 设链不可约、正常返且无周期, μ 为唯一的不变概率测度。则对每个 $x \in E$,

$$\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \mu(y)| \xrightarrow[n \rightarrow \infty]{} 0.$$

证明。 在 $E \times E$ 上定义

$$\overline{Q}((x_1, x_2), (y_1, y_2)) = Q(x_1, y_1)Q(x_2, y_2)$$

它是转移矩阵。记相应的规范链为 $((X_n^1, X_n^2), \overline{\mathbb{P}}_{(x_1, x_2)})$ 。由命题 13.6.5 (ii), $\overline{Q}$ 不可约: 两个坐标可以在同一个充分大的时刻分别到达指定状态。

此外, 乘积测度 $\mu \otimes \mu$ 对 $\overline{Q}$ 不变:

$$\begin{aligned} \sum_{(x_1, x_2) \in E \times E} \mu(x_1) \mu(x_2) Q(x_1, y_1) Q(x_2, y_2) &= \left(\sum_{x_1 \in E} \mu(x_1) Q(x_1, y_1) \right) \left(\sum_{x_2 \in E} \mu(x_2) Q(x_2, y_2) \right) \\ &= \mu(y_1) \mu(y_2). \end{aligned}$$

由命题 13.5.5, 乘积链 (X_n^1, X_n^2) 正常返。另一方面,

$$\begin{aligned} \mathbb{P}_x(X_n = y) - \mu(y) &= \bar{\mathbb{P}}_{\mu \otimes \delta_x}(X_n^2 = y) - \bar{\mathbb{P}}_{\mu \otimes \delta_x}(X_n^1 = y) \\ &= \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}}]. \end{aligned}$$

令相遇时间 $T = \inf\{n \geq 0 : X_n^1 = X_n^2\}$ 。则

$$\begin{aligned} \mathbb{P}_x(X_n = y) - \mu(y) &= \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})] \\ &\quad + \sum_{k=0}^n \sum_{z \in E} \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T=k, X_k^1=X_k^2=z\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]. \end{aligned} \quad (13.5)$$

对每个 $k \in \{0, \dots, n\}$ 与 $z \in E$, 马尔可夫性质给出

$$\begin{aligned} \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T=k, X_k^1=X_k^2=z\}} \mathbf{1}_{\{X_n^2=y\}}] &= \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T=k, X_k^1=X_k^2=z\}}] Q_{n-k}(z, y) \\ &= \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T=k, X_k^1=X_k^2=z\}} \mathbf{1}_{\{X_n^1=y\}}], \end{aligned}$$

因此式 (13.5) 的第二项为零, 进而

$$\begin{aligned} \sum_{y \in E} |\mathbb{P}_x(X_n = y) - \mu(y)| &= \sum_{y \in E} |\bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]| \\ &\leq \sum_{y \in E} \bar{\mathbb{E}}_{\mu \otimes \delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} + \mathbf{1}_{\{X_n^1=y\}})] \\ &= 2 \bar{\mathbb{P}}_{\mu \otimes \delta_x}(T > n). \end{aligned}$$

乘积链常返, 所以 $T < \infty$ 几乎必然; 于是右端随 $n \rightarrow \infty$ 趋于零。

13.7 鞅与马尔可夫链

继续考虑转移矩阵为 Q 的规范马尔可夫链。

定义 13.7.1 函数 $f: E \rightarrow \mathbb{R}_+$ 称为调和函数 (相应地, 超调和函数), 若对每个 $x \in E$,

$$f(x) = Qf(x) \quad (\text{相应地, } f(x) \geq Qf(x)).$$

更一般地, 若上述等式 (相应地, 不等式) 对每个 $x \in F \subset E$ 成立, 就称 f 在 F 上调和 (相应地, 超调和)。

备注。 我们可以更一般地考虑任意符号的调和函数或超调和函数。

命题 13.7.1 (i) 函数 f 调和 (分别超调和), 当且仅当对每个 $x \in E$, 过程 $(f(X_n))_{n \in \mathbb{N}}$ 在 $\mathbb{P}_x$ 下、关于滤过 $(\mathcal{F}_n)$ 是鞅 (分别上鞅)。 (ii) 设 $F \subset E$, $G = E \setminus F$, 并记停时

$$T_G = \inf\{n \geq 0 : X_n \in G\}.$$

若 f 在 F 上调和 (分别超调和), 则对每个 $x \in F$, 停止过程 $(f(X_{n \wedge T_G}))_{n \in \mathbb{N}}$ 在 $\mathbb{P}_x$ 下是鞅 (分别上鞅)。

证明。 (i) 首先假设 f 是调和的。那么根据命题 13.1.2(i),

$$\mathbb{E}_x[f(X_{n+1}) | \mathcal{F}_n] = Qf(X_n) = f(X_n)$$

因此 $\mathbb{E}_x[f(X_n)] = \mathbb{E}_x[f(X_0)] = f(x)$, 故 $f(X_n) \in L^1$ 。反过来, 若 $(f(X_n))$ 在 $\mathbb{P}_x$ 下是鞅, 则立即有

$$f(x) = \mathbb{E}_x[f(X_0)] = \mathbb{E}_x[f(X_1)] = Qf(x).$$

超调和函数的情况以同样的方式处理。 (ii) 我们来处理调和函数的情况。对于 $x \in F$ 写出

$$\begin{aligned} \mathbb{E}_x[f(X_{(n+1) \wedge T_G}) | \mathcal{F}_n] &= \mathbb{E}_x[f(X_{n+1}) \mathbf{1}_{\{T_G > n\}} | \mathcal{F}_n] + \mathbb{E}_x[f(X_{T_G}) \mathbf{1}_{\{T_G \leq n\}} | \mathcal{F}_n] \\ &= \mathbf{1}_{\{T_G > n\}} \mathbb{E}_x[f(X_{n+1}) | \mathcal{F}_n] + f(X_{T_G}) \mathbf{1}_{\{T_G \leq n\}} \\ &= \mathbf{1}_{\{T_G > n\}} Qf(X_n) + f(X_{T_G}) \mathbf{1}_{\{T_G \leq n\}} \\ &= \mathbf{1}_{\{T_G > n\}} f(X_n) + f(X_{T_G}) \mathbf{1}_{\{T_G \leq n\}} \\ &= f(X_{n \wedge T_G}). \end{aligned}$$

我们利用了 $f(X_{T_G}) \mathbf{1}_{\{T_G \leq n\}} = f(X_{T_G \wedge n}) \mathbf{1}_{\{T_G \leq n\}}$ 是 $\mathcal{F}_n$ 可测的事实。

定理 13.7.2 设 F 是 E 的非空子集, $G = E \setminus F$ 。设 $g : G \rightarrow \mathbb{R}_+$ 为有界函数。

(i) 函数

$$h(x) = \mathbb{E}_x[g(X_{T_G}) \mathbf{1}_{\{T_G < \infty\}}], \quad x \in E$$

在 F 上调和。

(ii) 设对每个 $x \in F$, 都有 $T_G < \infty$ 在 $\mathbb{P}_x$ -几乎必然意义下成立。则 h 是 E 上唯一的有界函数, 满足:

- 在 F 上调和;
- 在 G 上等于 g 。

证明。 (i) 若 $x \in F$, 则在 $\mathbb{P}_x$ -几乎必然意义下,

$$g(X_{T_G}) \mathbf{1}_{\{T_G < \infty\}} = g(X_{T_G} \circ \theta_1) \mathbf{1}_{\{T_G \circ \theta_1 < \infty\}}.$$

换言之, 若 $U(\omega) = g(X_{T_G}(\omega)) \mathbf{1}_{\{T_G(\omega) < \infty\}}$, 则 $U = U \circ \theta_1$, $\mathbb{P}_x$ -几乎必然。因此由定理 13.3.4,

$$h(x) = \mathbb{E}_x[U] = \mathbb{E}_x[U \circ \theta_1] = \mathbb{E}_x[\mathbb{E}_{X_1}[U]] = \mathbb{E}_x[h(X_1)] = Qh(x),$$

所以 h 在 F 上调和。 (ii) 若 $x \in G$, 显然 $h(x) = g(x)$ 。设 h' 是另一个有界函数, 在 F 上调和且在 G 上等于 g 。对 $x \in F$, 命题 13.7.1 表明

$$Y_n = h'(X_{n \wedge T_G})$$

是 $\mathbb{P}_x$ 下的有界鞅，因而一致可积。又因 $T_G < \infty$ 几乎必然， $Y_n \rightarrow g(X_{T_G})$ 几乎必然。由第 12 章的一致可积收敛定理，

$$h'(x) = \mathbb{E}_x[Y_0] = \mathbb{E}_x[Y_\infty] = \mathbb{E}_x[g(X_{T_G})] = h(x).$$

例（离散 Dirichlet 问题）。 设 $F \subset \mathbb{Z}^d$ 为有限集，定义边界

$$\partial F = \{y \in \mathbb{Z}^d \setminus F : \exists x \in F, |y - x| = 1\}.$$

记 $\bar{F} = F \cup \partial F$ 。若函数 $h : \bar{F} \rightarrow \mathbb{R}$ 在每个 $x \in F$ 处的值，等于它在 $2d$ 个最近邻点处函数值的平均，就称 h 在 F 上离散调和。这正对应于 $\mathbb{Z}^d$ 上的简单随机游走

$$Q(x, x \pm e_j) = \frac{1}{2d}, \quad j = 1, \dots, d.$$

因此，对任意 $g : \partial F \rightarrow \mathbb{R}_+$ ，满足以下边界值问题的函数 $h : \bar{F} \rightarrow \mathbb{R}_+$ 唯一：

• h 在 F 上是调和的

$$\bullet h(y) = g(y), \forall y \in \partial F,$$

由下式给出

$$h(x) = \mathbb{E}_x[g(X_{T_{\partial F}})], \quad x \in F,$$

其中

$$T_{\partial F} = \inf\{n \geq 0 : X_n \in \partial F\}.$$

为应用定理 13.7.2，形式上需要把 g 定义在整个 $\mathbb{Z}^d \setminus F$ 上，而不仅是 ∂F 上。但 g 在 $\mathbb{Z}^d \setminus \bar{F}$ 上的取值不会影响 h 在 F 上的取值。

第十四章 布朗运动导论

14.1 作为随机游走极限的布朗运动

布朗运动的物理图景是：粒子不断受到周围介质的碰撞，运动方向频繁改变，因而轨迹显得极其无序、难以预测。从数学上看，这启发我们先考察离散时间模型：一个点粒子在格点 $\mathbb{Z}^d$ 上运动，并在每一步独立于过去选择新的位移。具体地，考虑从 0 出发的随机游走 $(S_n)_{n \in \mathbb{N}}$ ：

$$S_n = Y_1 + \cdots + Y_n$$

其中 $Y_1, Y_2, \dots$ 是取值于 $\mathbb{Z}^d$ 的独立同分布随机变量，共同分布为 μ 。假设 μ 满足：

- $\sum_{k \in \mathbb{Z}^d} |k|^2 \mu(k) < \infty$;
- $\sum_{k \in \mathbb{Z}^d} k \mu(k) = 0$ ，即 μ 的均值为零。

再加上如下各向同性条件：

- 存在常数 $\sigma > 0$ ，使得对任意 $i, j \in \{1, \dots, d\}$,

$$\sum_{k \in \mathbb{Z}^d} k_i k_j \mu(k) = \sigma^2 \delta_{ij}.$$

$\mathbb{Z}^d$ 上的简单随机游走（见上一章）满足这些假设，且 $\sigma^2 = 1/d$ ；满足这些条件的分布还有很多。

我们研究映射 $k \mapsto S_k$ 在长时间区间上的整体行为。为此作如下尺度变换：对任意整数 $n \geq 1$ 和实数 $t \geq 0$ ，令

$$S_t^{(n)} = \frac{1}{\sqrt{n}} S_{[nt]}$$

其中 $[x]$ 表示实数 x 的整数部分。

命题 14.1.1 任取整数 $p \geq 1$ 与实数 $0 = t_0 < t_1 < \cdots < t_p$ ，有

$$(S_{t_1}^{(n)}, \dots, S_{t_p}^{(n)}) \xrightarrow[n \rightarrow \infty]{\text{law}} (U_1, \dots, U_p).$$

其极限分布如下所示：

- 随机变量 $U_1, U_2 - U_1, \dots, U_p - U_{p-1}$ 相互独立；
- 对每个 $j \in \{1, \dots, p\}$ ， $U_j - U_{j-1}$ 是中心高斯向量，协方差矩阵为 $\sigma^2(t_j - t_{j-1})I_d$ （约定 $U_0 = 0$ ）。

注。 极限分布的密度可以显式写出。 $U_j - U_{j-1}$ 的密度是 $p_{\sigma^2(t_j - t_{j-1})}(x)$ 。对任意 $a > 0$ ，令

$$p_a(x) = \frac{1}{(2\pi a)^{d/2}} \exp\left(-\frac{|x|^2}{2a}\right), \quad x \in \mathbb{R}^d.$$

¹ 是协方差矩阵为 aI_d 的高斯向量的密度 (这种向量的各坐标是相互独立、服从 $\mathcal{N}(0, a)$ 的实随机变量, 见命题 11.4.2 及其后的注)。由随机变量 $U_1, U_2 - U_1, \dots, U_p - U_{p-1}$ 的独立性, 向量 $(U_1, U_2 - U_1, \dots, U_p - U_{p-1})$ 的密度为

$$g(x_1, \dots, x_p) = p_{\sigma^2 t_1}(x_1) p_{\sigma^2(t_2-t_1)}(x_2) \cdots p_{\sigma^2(t_p-t_{p-1})}(x_p),$$

并且通过一个简单的变量变换, $(U_1, U_2, \dots, U_p)$ 的密度为

$$g(y_1, y_2 - y_1, \dots, y_p - y_{p-1}) = p_{\sigma^2 t_1}(y_1) p_{\sigma^2(t_2-t_1)}(y_2 - y_1) \cdots p_{\sigma^2(t_p-t_{p-1})}(y_p - y_{p-1}).$$

证明。 只需证明, 对于所有 $\xi_1, \dots, \xi_p \in \mathbb{R}^d$,

$$\mathbb{E} \left[\exp \left(i \sum_{j=1}^p \xi_j \cdot S_{t_j}^{(n)} \right) \right] \longrightarrow \mathbb{E} \left[\exp \left(i \sum_{j=1}^p \xi_j \cdot U_j \right) \right].$$

这等价于说, 对于所有 $\eta_1, \dots, \eta_p \in \mathbb{R}^d$,

$$\mathbb{E} \left[\exp \left(i \sum_{j=1}^p \eta_j \cdot (S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)}) \right) \right] \longrightarrow \mathbb{E} \left[\exp \left(i \sum_{j=1}^p \eta_j \cdot (U_j - U_{j-1}) \right) \right]. \quad (14.1)$$

由 $U_1, U_2 - U_1, \dots, U_p - U_{p-1}$ 的独立性, 已知

$$\begin{aligned} \mathbb{E} \left[\exp \left(i \sum_{j=1}^p \eta_j \cdot (U_j - U_{j-1}) \right) \right] &= \prod_{j=1}^p \mathbb{E}[e^{i\eta_j \cdot (U_j - U_{j-1})}] \\ &= \exp \left(- \sum_{j=1}^p \frac{\sigma^2 |\eta_j|^2 (t_j - t_{j-1})}{2} \right). \end{aligned}$$

这里使用了高斯分布的傅里叶变换公式。另一方面,

$$S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=[nt_{j-1}]+1}^{[nt_j]} Y_k$$

这首先说明 $S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)}$ ($1 \leq j \leq p$) 相互独立; 其次, 对每个固定的 j ,

$$\begin{aligned} S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)} &\stackrel{\text{law}}{=} \frac{1}{\sqrt{n}} S_{[nt_j] - [nt_{j-1}]} \\ &= \frac{\sqrt{[nt_j] - [nt_{j-1}]}}{\sqrt{n}} \frac{S_{[nt_j] - [nt_{j-1}]}}{\sqrt{[nt_j] - [nt_{j-1}]}}. \end{aligned}$$

由多维中心极限定理, 上式最后一个随机变量在 $n \rightarrow \infty$ 时依分布收敛到 $\sqrt{t_j - t_{j-1}} N$, 其中 N 是协方差矩阵为 $\sigma^2 I_d$ 的中心高斯向量。这里也用了如下简单事实: 若 X_n 依分布收敛到 X , 且实数列 $a_n \rightarrow a$, 则 $a_n X_n$ 依分布收敛到 aX 。因此, 对每个固定的 j ,

$$\begin{aligned} \mathbb{E} \left[e^{i\eta_j \cdot (S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)})} \right] &\longrightarrow \mathbb{E}[e^{i\sqrt{t_j - t_{j-1}} \eta_j \cdot N}] \\ &= \exp \left(- \frac{\sigma^2 |\eta_j|^2 (t_j - t_{j-1})}{2} \right). \end{aligned}$$

¹校订注: 原讲义此式指数的分母排成了 $2at$; 协方差为 aI_d 的高斯密度应为 $2a$ 。

再利用这些增量的独立性，便得到所需的 (14.1)。

定义 14.1.1 设 $(B_t)_{t \in \mathbb{R}_+}$ 是某概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上一族取值于 $\mathbb{R}^d$ 的随机变量。若满足下列两条性质，就称它为从 0 出发的 d 维布朗运动。

(P1) $B_0 = 0$ 几乎必然。并且对任意整数 $p \geq 1$ 与 $0 = t_0 < t_1 < \cdots < t_p$ ，随机变量 $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_p} - B_{t_{p-1}}$ 相互独立；对每个 $j \in \{1, \dots, p\}$ ，增量 $B_{t_j} - B_{t_{j-1}}$ 是均值为零、协方差矩阵为 $(t_j - t_{j-1})I_d$ 的高斯向量。

(P2) 对每个 $\omega \in \Omega$ ，函数 $t \mapsto B_t(\omega)$ 连续。

注。(i) 暂且承认布朗运动存在（下一节将给出构造）。命题 14.1.1 可改写为：对任意 $0 \leq t_1 < \cdots < t_p$,

$$(S_{t_1}^{(n)}, S_{t_2}^{(n)}, \dots, S_{t_p}^{(n)}) \xrightarrow[n \rightarrow \infty]{\text{依分布}} (\sigma B_{t_1}, \sigma B_{t_2}, \dots, \sigma B_{t_p}).$$

除去常数因子 σ ，布朗运动正是离散随机游走经过适当时空尺度变换后的连续极限。

从物理上看，这对应于由微观碰撞模型过渡到宏观连续运动。(ii) 由独立高斯增量可知，向量 $(B_{t_1}, \dots, B_{t_p})$ 的分布满足

$$\mathbb{P}((B_{t_1}, \dots, B_{t_p}) \in A) = \int_A p_{t_1}(y_1) p_{t_2-t_1}(y_2 - y_1) \cdots p_{t_p-t_{p-1}}(y_p - y_{p-1}) dy_1 \cdots dy_p. \quad (14.2)$$

对每个 Borel 集 $A \in \mathcal{B}((\mathbb{R}^d)^p)$ 都成立。

14.2 布朗运动的构造

定理 14.2.1 布朗运动存在。即存在适当的概率空间以及满足 (P1)、(P2) 的过程 $(B_t)_{t \in \mathbb{R}_+}$ 。

证明。先处理 $d = 1$ ，并先在 $[0, 1]$ 上构造 $(B_t)_{t \in [0, 1]}$ 。只需选取一个能够承载一系列独立标准高斯随机变量的概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 。例如，上一章已说明可在 $[0, 1]$ 上构造一系列独立均匀随机变量，再通过分位数变换得到所需的独立标准高斯变量。下面引入 Haar 函数，先令

$$h_0(t) = 1, \quad \forall t \in [0, 1]$$

再对任意整数 $n \geq 0$ 和 $k \in \{0, 1, \dots, 2^n - 1\}$ ，定义

$$h_n^k(t) = 2^{n/2} \mathbf{1}_{[(2k)2^{-n-1}, (2k+1)2^{-n-1})}(t) - 2^{n/2} \mathbf{1}_{[(2k+1)2^{-n-1}, (2k+2)2^{-n-1})}(t), \quad t \in [0, 1].$$

令 $\mathcal{H} = L^2([0, 1], \mathcal{B}([0, 1]), \lambda)$ ，其中 λ 是 Lebesgue 测度。容易验证， h_0 与 h_n^k 构成 $\mathcal{H}$ 中的规范正交系。它还是完备的：固定 n 后，在各区间 $[i2^{-n}, (i+1)2^{-n})$ 上为常数的阶梯函数，都可写成 h_0 与 h_p^k ($p < n$) 的线性组合。因此

$$h_0, (h_n^k)_{n \geq 0, 0 \leq k \leq 2^n - 1}$$

构成 $\mathcal{H}$ 的规范正交基。记 $\langle \cdot, \cdot \rangle$ 为 $\mathcal{H}$ 的内积。于是对任意 $f \in \mathcal{H}$,

$$f = \langle f, h_0 \rangle h_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle h_n^k.$$

另一方面, 在 $(\Omega, \mathcal{F}, \mathbb{P})$ 上取一系列独立标准高斯随机变量; 重新编号后写成

$$N_0, (N_n^k)_{n \geq 0, 0 \leq k \leq 2^n-1}.$$

这个随机变量族在 $L^2(\Omega, \mathcal{F}, \mathbb{P})$ 中也是规范正交系。因此存在唯一的线性等距映射

$$\mathcal{B} : L^2([0, 1], \mathcal{B}([0, 1]), \lambda) \longrightarrow L^2(\Omega, \mathcal{F}, \mathbb{P})$$

使 $\mathcal{B}(h_0) = N_0$, 且 $\mathcal{B}(h_n^k) = N_n^k$ ($n \geq 0, 0 \leq k < 2^n$)。具体地,

$$\mathcal{B}(f) = \langle f, h_0 \rangle N_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k,$$

对每个 $f \in L^2([0, 1])$ 都成立, 且右端级数在 $L^2(\Omega, \mathcal{F}, \mathbb{P})$ 中收敛。由等距性,

$$E[\mathcal{B}(f)^2] = \|f\|_2^2$$

又因所有 N_0, N_n^k 都中心化, $E[\mathcal{B}(f)] = 0$ 。下一个引理说明 $\mathcal{B}(f)$ 仍是高斯随机变量。

引理 14.2.2 若高斯随机变量序列 (U_n) 在 L^2 中收敛于 U , 则 U 也是高斯随机变量。

证明。 令 $m_n = E[U_n]$, $\sigma_n^2 = \text{Var}(U_n)$ 。 L^2 收敛给出

$$m_n \rightarrow m := E[U], \quad \sigma_n^2 \rightarrow \sigma^2 := \text{Var}(U).$$

同时, L^2 收敛蕴含依分布收敛, 所以对每个 $\xi \in \mathbb{R}$,

$$e^{im_n \xi - \sigma_n^2 \xi^2 / 2} = E[e^{i\xi U_n}] \longrightarrow E[e^{i\xi U}]$$

令 $n \rightarrow \infty$, 便得到 U 的特征函数

$$E[e^{i\xi U}] = e^{im\xi - \sigma^2 \xi^2 / 2}$$

因此 $U \sim \mathcal{N}(m, \sigma^2)$ 。再把 $\mathcal{B}(f)$ 写成

$$\mathcal{B}(f) = \lim_{m \rightarrow \infty} \left(\langle f, h_0 \rangle N_0 + \sum_{n=0}^m \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k \right),$$

有限个独立高斯变量的线性组合仍为高斯变量, 故由引理可知 $\mathcal{B}(f) \sim \mathcal{N}(0, \|f\|_2^2)$ 。此外, 对任意 $f, f' \in L^2([0, 1])$,

$$\text{Cov}(\mathcal{B}(f), \mathcal{B}(f')) = E[\mathcal{B}(f)\mathcal{B}(f')] = \langle f, f' \rangle.$$

² 这是等距性与极化恒等式的直接结果。现在对每个 $t \in [0, 1]$, 令

$$B_t = \mathcal{B}(\mathbf{1}_{[0,t]}).$$

²校订注: 原讲义此式左端把第二个随机变量误排为 $\mathcal{B}'(f)$; 应为 $\mathcal{B}(f')$ 。

特别地, $B_0 = \mathcal{B}(\mathbf{1}_{\{0\}}) = \mathcal{B}(0) = 0$ 几乎必然。现在验证 $(B_t)_{t \in [0,1]}$ 在区间 $[0,1]$ 上满足 (P1)。取 $0 = t_0 < t_1 < \cdots < t_p \leq 1$ 。由线性性,

$$B_{t_i} - B_{t_{i-1}} = \mathcal{B}(\mathbf{1}_{[t_{i-1}, t_i]})$$

它服从 $\mathcal{N}(0, t_i - t_{i-1})$ 。此外, 若 $i \neq j$,

$$\text{Cov}(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}) = \langle \mathbf{1}_{[t_{i-1}, t_i]}, \mathbf{1}_{[t_{j-1}, t_j]} \rangle = 0.$$

向量 $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_p} - B_{t_{p-1}})$ 还是高斯向量: 任取 $\lambda_1, \dots, \lambda_p \in \mathbb{R}$, 有

$$\sum_{j=1}^p \lambda_j (B_{t_j} - B_{t_{j-1}}) = \mathcal{B}\left(\sum_{j=1}^p \lambda_j \mathbf{1}_{[t_{j-1}, t_j]}\right)$$

服从高斯分布。由命题 11.4.2, 协方差矩阵为对角矩阵蕴含随机变量 $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_p} - B_{t_{p-1}}$ 相互独立, 故 (P1) 得证。下面证明路径连续性 (P2)。目前, $B_t = \mathcal{B}(\mathbf{1}_{[0,t]})$ 只是 $L^2(\Omega, \mathcal{F}, P)$ 中的一个元素, 即一类几乎必然相等的随机变量。为了讨论 (P2), 必须对每个 $t \in [0,1]$ 选定一个代表元; 这个选择不影响 (P1), 但会影响 (P2)。为此, 要进一步分析定义 B_t 的级数。先引入 Schauder 函数。

$$\begin{aligned} g_0(t) &= \langle \mathbf{1}_{[0,t]}, h_0 \rangle = t \\ g_n^k(t) &= \langle \mathbf{1}_{[0,t]}, h_n^k \rangle = \int_0^t h_n^k(s) ds. \end{aligned}$$

根据构造, 对于所有 $t \in [0,1]$, 有

$$B_t = \mathcal{B}(\mathbf{1}_{[0,t]}) = tN_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k$$

该级数先验地对每个固定的 $t \in [0,1]$ 在 $L^2(\Omega, \mathcal{F}, P)$ 中收敛。下面将证明更强的结论: 除去某个概率为零的集合 $A \in \mathcal{F}$ 外, 对每个 $\omega \in \Omega$, 该级数在 $[0,1]$ 上一致收敛。若 $\omega \in A^c$, 定义 $B_t(\omega)$ 为上述级数之和; 若 $\omega \in A$, 则对所有 $t \in [0,1]$ 令 $B_t(\omega) = 0$ 。这只是从与 $\mathcal{B}(\mathbf{1}_{[0,t]})$ 几乎必然相等的随机变量等价类中选定代表, 不影响 (P1)。由于连续函数的一致极限仍连续, 这将给出 $t \mapsto B_t(\omega)$ 的连续性。先注意到 $0 \leq g_n^k \leq 2^{-n/2}$; 对固定 n , 函数 g_n^k ($0 \leq k \leq 2^n - 1$) 的支集互不相交, 而且仅当 $k2^{-n} < t < (k+1)2^{-n}$ 时 $g_n^k(t) > 0$ 。因此

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \right| \leq 2^{-n/2} \sup_{0 \leq k \leq 2^n-1} |N_n^k|.$$

引理 14.2.3 若 $N \sim \mathcal{N}(0,1)$, 则对每个 $a \geq 1$,

$$P(|N| \geq a) \leq e^{-a^2/2}.$$

证明。直接估计高斯密度的尾积分:

$$P(|N| \geq a) = \frac{2}{\sqrt{2\pi}} \int_a^\infty dx e^{-x^2/2} \leq \frac{2}{\sqrt{2\pi}} \int_a^\infty dx \frac{x}{a} e^{-x^2/2} = \frac{2}{a\sqrt{2\pi}} e^{-a^2/2}.$$

因为所有 $N_n^k \sim \mathcal{N}(0,1)$, 由并集界与上述引理,

$$P\left(\sup_{0 \leq k \leq 2^n-1} |N_n^k| > 2^{n/4}\right) \leq \sum_{k=0}^{2^n-1} P(|N_n^k| > 2^{n/4}) \leq 2^n \exp(-2^{\frac{n}{2}-1}).$$

记

$$A_n = \left\{ \sup_{0 \leq k \leq 2^n-1} |N_n^k| > 2^{n/4} \right\}$$

由前一估计可知 $\sum_n P(A_n) < \infty$ 。Borel–Cantelli 引理因而给出

$$P(\limsup A_n) = 0.$$

令 $A = \limsup_{n \rightarrow \infty} A_n$ ，则 $P(A) = 0$ 。若 $\omega \notin A$ ，则对所有充分大的 n ，

$$\sup_{0 \leq k \leq 2^n-1} |N_n^k| \leq 2^{n/4}$$

于是

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \right| \leq 2^{-n/4}$$

右端关于 n 可求和，所以定义 B_t 的函数级数在 $[0,1]$ 上一致收敛，从而得到 (P2)。该构造甚至保证对每个 $\omega \in \Omega$ 都有 $B_0(\omega) = 0$ 。最后把构造从 $[0,1]$ 推广到 $\mathbb{R}_+$ ，并推广到任意维数。取彼此独立的一维过程 $(B_t^{(1)})_{t \in [0,1]}, (B_t^{(2)})_{t \in [0,1]}, \dots$ ，每个过程都按上法由一组新的、且与其他各组独立的高斯变量构造，并令

$$B_t = B_1^{(1)} + \dots + B_1^{(k)} + B_{t-k}^{(k+1)}, \quad t \in [k, k+1).$$

由独立平稳增量和逐段连续性，容易验证 $(B_t)_{t \geq 0}$ 是一维布朗运动。对任意 $d \geq 1$ ，再取 d 个相互独立的一维布朗运动 $(B_t^j)_{t \geq 0}$ ($1 \leq j \leq d$)，并令

$$B_t = (B_t^1, B_t^2, \dots, B_t^d)$$

这就构造出了 d 维布朗运动。

若 $x \in \mathbb{R}^d$ ，称 $(B_t)_{t \geq 0}$ 是从 x 出发的布朗运动，如果 $(B_t - x)_{t \geq 0}$ 是从 0 出发的布朗运动。

14.3 Wiener 测度

设 $C(\mathbb{R}_+, \mathbb{R}^d)$ 为从 $\mathbb{R}_+$ 到 $\mathbb{R}^d$ 的连续函数空间。赋予它最小的 σ -代数 $\mathcal{C}$ ，使每个坐标映射 $w \mapsto w(t)$ ($t \in \mathbb{R}_+$) 均可测。

引理 14.3.1 当 $C(\mathbb{R}_+, \mathbb{R}^d)$ 赋予在每个紧集上一致收敛的拓扑时， $\mathcal{C}$ 恰为其 Borel σ -代数。

证明。 记 $\mathcal{B}$ 为 Borel σ -代数。坐标映射连续，故 $\mathcal{C} \subset \mathcal{B}$ 。反过来，空间上的一个度量可定义为

$$d(w, w') = \sum_{n=1}^{\infty} 2^{-n} \sup_{0 \leq t \leq n} (|w(t) - w'(t)| \wedge 1).$$

空间 $C(\mathbb{R}_+, \mathbb{R}^d)$ 可分, 故每个开集都是可数个开球之并。因此只需证明所有开球属于 $\mathcal{C}$; 等价地, 对固定 $w_0 \in C(\mathbb{R}_+, \mathbb{R}^d)$, 映射 $w \mapsto d(w_0, w)$ 是 $\mathcal{C}$ -可测的。对每个 $n \geq 1$,

$$\sup_{t \in [0, n]} (|w(t) - w_0(t)| \wedge 1) = \sup_{t \in [0, n] \cap \mathbb{Q}} (|w(t) - w_0(t)| \wedge 1),$$

我们立刻得到这个可测性性质。

定义 14.3.1 设 $(B_t)_{t \in \mathbb{R}_+}$ 是定义在概率空间 $(\Omega, \mathcal{F}, P)$ 上、从 0 出发的 d 维布朗运动。 d 维 Wiener 测度是 $C(\mathbb{R}_+, \mathbb{R}^d)$ 上的概率测度 $\mathbb{P}_0$, 即 $P(dw)$ 在下列映射下的像测度:

$$\begin{aligned} \Phi : \Omega &\longrightarrow C(\mathbb{R}_+, \mathbb{R}^d), \\ \omega &\longmapsto (B_t(\omega))_{t \geq 0}. \end{aligned}$$

映射 Φ 是可测的: 只需考察它与各坐标映射 $w \mapsto w(t)$ 的复合; 该复合正是随机变量 B_t 。上述定义必须与所选布朗运动 B 无关。为此, 设 $0 = t_0 < t_1 < \cdots < t_p$, 并任取 $\mathbb{R}^d$ 中的 Borel 集 $A_0, A_1, \dots, A_p$ 。则

$$\begin{aligned} \mathbb{P}_0\{w : w(t_0) \in A_0, \dots, w(t_p) \in A_p\} &= P(B_{t_0} \in A_0, \dots, B_{t_p} \in A_p) \\ &= 1_{A_0}(0) \int_{A_1 \times \cdots \times A_p} p_{t_1}(y_1) \\ &\quad \times p_{t_2-t_1}(y_2 - y_1) \cdots p_{t_p-t_{p-1}}(y_p - y_{p-1}) dy_1 \cdots dy_p. \end{aligned}$$

根据式 (14.2), 上式对任意布朗运动 B 都成立 (这正是 (P1) 的重述)。单调类引理表明, $C(\mathbb{R}_+, \mathbb{R}^d)$ 上的概率测度由其在柱集上的取值唯一确定; 柱集即形如

$$\{w : w(t_0) \in A_0, \dots, w(t_p) \in A_p\}$$

的集合。因此, 上式唯一确定 $\mathbb{P}_0$, 从而 Wiener 测度与用来定义它的布朗运动无关。以后只把 $\mathbb{P}_0$ 称为 Wiener 测度。对每个 $x \in \mathbb{R}^d$, 记 $\mathbb{P}_x(dw)$ 为 $\mathbb{P}_0(dw)$ 在平移映射 $w \mapsto x + w$ 下的像测度。现在取典范空间 $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$, 赋予 σ -代数 $\mathcal{C}$ 和概率测度 $\mathbb{P}_0$, 并定义坐标过程

$$B_t(w) = w(t), \quad t \geq 0.$$

由 Wiener 测度的定义, $(B_t)_{t \geq 0}$ 在 $\mathbb{P}_0$ 下是从 0 出发的 d 维布朗运动。事实上, 对任意 $0 = t_0 < t_1 < \cdots < t_p$ 及 Borel 集 $A_0, \dots, A_p$,

$$\mathbb{P}_0(B_{t_0} \in A_0, \dots, B_{t_p} \in A_p) = 1_{A_0}(0) \int_{A_1 \times \cdots \times A_p} p_{t_1}(y_1) \cdots p_{t_p-t_{p-1}}(y_p - y_{p-1}) dy_1 \cdots dy_p.$$

同理, 在 $\mathbb{P}_x$ 下, 坐标过程 $(B_t)_{t \geq 0}$ 是从 x 出发的 d 维布朗运动。

14.4 布朗运动的基本性质

以后记 $\mathcal{F}_s = \sigma(B_r : 0 \leq r \leq s)$, 以及 $\mathcal{F}_\infty = \sigma(B_r : r \geq 0)$ 。

命题 14.4.1 设 $(B_t)_{t \geq 0}$ 是从 0 出发的 d 维布朗运动。(i) 若 $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 是线性等距映射, 则 $(\varphi(B_t))_{t \geq 0}$ 仍是从 0 出发的 d 维布朗运动。
(ii) 若 $\gamma > 0$, 则由 $B_t^\gamma = \gamma^{-1} B_{\gamma^2 t}$ ($t \geq 0$) 定义的过程 $(B_t^\gamma)_{t \geq 0}$ 仍是从 0 出发的 d 维布朗运动。
(iii) 若 $s \geq 0$, 则由 $B_t^{(s)} = B_{s+t} - B_s$ ($t \geq 0$) 定义的过程 $(B_t^{(s)})_{t \geq 0}$ 是从 0 出发的 d 维布朗运动, 并且独立于 $\mathcal{F}_s$ 。

证明。(i) 和 (ii) 非常容易。对于 (iii) 中的独立性, 我们注意到对于任意选择的 $t_1 < t_2 < \dots < t_p$ 和 $r_1 < r_2 < \dots < r_q \leq s$, 性质 (P1) 推出向量

$$(B_{t_1}^{(s)}, \dots, B_{t_p}^{(s)})$$

是独立的

$$(B_{r_1}, \dots, B_{r_q}).$$

由命题 9.2.4, 整个未来增量过程 $(B_t^{(s)})_{t \in \mathbb{R}_+}$ 独立于过去 $(B_r)_{0 \leq r \leq s}$, 也即独立于 $\mathcal{F}_s$ 。

定理 14.4.2 (布鲁门塔尔的全或无律) 设

$$\mathcal{F}_{0+} = \bigcap_{s>0} \mathcal{F}_s.$$

则 $\mathcal{F}_{0+}$ 是平凡 σ -代数: 对任意 $A \in \mathcal{F}_{0+}$, 都有 $P(A) \in \{0, 1\}$ 。

证明。设 $A \in \mathcal{F}_{0+}$, 并取 $t_1, \dots, t_p > 0$ 。当 $\varepsilon > 0$ 充分小时, 简单马尔可夫性质 (命题 14.4.1 (iii)) 表明 $(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)$ 与 $\mathcal{F}_\varepsilon$ 独立, 因而也与 $\mathcal{F}_{0+}$ 独立。因此, 对任意定义在 $(\mathbb{R}^d)^p$ 上的有界连续函数 f ,

$$E[\mathbf{1}_A f(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)] = P(A) E[f(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)].$$

令 $\varepsilon \downarrow 0$, 由路径连续性与支配收敛定理得到

$$E[\mathbf{1}_A f(B_{t_1}, \dots, B_{t_p})] = P(A) E[f(B_{t_1}, \dots, B_{t_p})],$$

因此 $(B_{t_1}, \dots, B_{t_p})$ 独立于 $\mathcal{F}_{0+}$ 。再由命题 9.2.4, $\mathcal{F}_\infty$ 独立于 $\mathcal{F}_{0+}$ 。但 $\mathcal{F}_{0+} \subset \mathcal{F}_\infty$, 故 $\mathcal{F}_{0+}$ 独立于自身, 只能是平凡的。

推论 14.4.3 设 $d = 1$ 。则几乎必然地, 对每个 $\varepsilon > 0$,

$$\sup_{0 \leq s \leq \varepsilon} B_s > 0, \quad \inf_{0 \leq s \leq \varepsilon} B_s < 0.$$

对每个 $a \in \mathbb{R}$, 令 $T_a = \inf\{t \geq 0 : B_t = a\}$ (约定 $\inf \emptyset = \infty$)。则几乎必然地, 对每个 $a \in \mathbb{R}$, 都有 $T_a < \infty$ 。
因此, 几乎处处,

$$\limsup_{t \rightarrow \infty} B_t = +\infty, \quad \liminf_{t \rightarrow \infty} B_t = -\infty.$$

注。随机变量 $\sup_{0 \leq s \leq \varepsilon} B_s$ 的可测性并非先验显然，因为它是不可数族可测函数的上确界。但布朗路径连续，所以可把上确界限制在 $[0, \varepsilon] \cap \mathbb{Q}$ 上，从而化为可数族随机变量的上确界（也可直接使用引理 14.3.1）。

证明。取严格递减至 0 的正实数列 (ε_p) ，并令

$$A = \bigcap_p \left\{ \sup_{0 \leq s \leq \varepsilon_p} B_s > 0 \right\}.$$

显然 $A \in \mathcal{F}_{0+}$ 。又由事件的单调性，

$$\mathbb{P}(A) = \lim_{p \rightarrow \infty} \mathbb{P} \left(\sup_{0 \leq s \leq \varepsilon_p} B_s > 0 \right).$$

$$P \left(\sup_{0 \leq s \leq \varepsilon_p} B_s > 0 \right) \geq P(B_{\varepsilon_p} > 0) = \frac{1}{2},$$

因为 $B_{\varepsilon_p} \sim \mathcal{N}(0, \varepsilon_p)$ ，其分布关于 0 对称，所以 $P(A) \geq 1/2$ 。由定理 14.4.2 的全或无律， $P(A) = 1$ ，从而

$$\mathbb{P}\text{-几乎必然地, 对每个 } \varepsilon > 0, \quad \sup_{0 \leq s \leq \varepsilon} B_s > 0.$$

关于下确界的结论只需把 B 换成 $-B$ 。接下来，

$$1 = \mathbb{P} \left(\sup_{0 \leq s \leq 1} B_s > 0 \right) = \lim_{\delta \downarrow 0} \mathbb{P} \left(\sup_{0 \leq s \leq 1} B_s > \delta \right).$$

在命题 14.4.1 (ii) 中取 $\gamma = \delta$ ，由尺度不变性，

$$P \left(\sup_{0 \leq s \leq 1} B_s > \delta \right) = P \left(\sup_{0 \leq s \leq 1/\delta^2} B_s^\delta > 1 \right) = P \left(\sup_{0 \leq s \leq 1/\delta^2} B_s > 1 \right)$$

最后一个等号用到了布朗运动分布的唯一性。令 $\delta \downarrow 0$ ，得到

$$P(\sup_{s \geq 0} B_s > 1) = 1.$$

再次使用尺度不变性，对任意 $A > 0$ ，

$$P(\sup_{s \geq 0} B_s > A) = 1$$

再把 B 换成 $-B$ ，还有

$$P(\inf_{s \geq 0} B_s < -A) = 1.$$

这些结论说明布朗运动几乎必然会到达每个实数，并且 $\limsup_{t \rightarrow \infty} B_t = +\infty$ 、 $\liminf_{t \rightarrow \infty} B_t = -\infty$ 。

再结合简单马尔可夫性质可得：几乎必然地，样本路径 $t \mapsto B_t$ 在任何非退化区间上都不单调。

14.5 强马尔可夫性质

我们的目标是把简单马尔可夫性质（命题 14.4.1 (iii)）从确定时刻 s 推广到随机时刻 T 。为此，先明确允许使用的随机时刻。沿用前面的 $\mathcal{F}_t$ 与 $\mathcal{F}_\infty$ 。

定义 14.5.1 取值于 $[0, \infty]$ 的随机变量 T 称为停时，如果

$$\forall t \geq 0, \{T \leq t\} \in \mathcal{F}_t.$$

注。若 T 是停时，则对任意 $t \geq 0$,

$$\{T < t\} = \bigcup_{q \in \mathbb{Q} \cap [0, t)} \{T \leq q\} \in \mathcal{F}_t.$$

例。在一维情形， $T_a = \inf\{t \geq 0 : B_t = a\}$ 是停时。事实上，

$$\{T_a \leq t\} = \left\{ \inf_{r \in \mathbb{Q} \cap [0, t]} |B_r - a| = 0 \right\} \in \mathcal{F}_t.$$

定义 14.5.2 设 T 是停时。定义“停时 T 之前的信息”所组成的 σ -代数为

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty; \forall t \geq 0, A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

容易验证， T 与 $\mathbf{1}_{\{T < \infty\}} B_T$ 都是 $\mathcal{F}_T$ -可测的。对后一个结论，可使用近似式

$$\mathbf{1}_{\{T < \infty\}} B_T = \lim_{n \rightarrow \infty} \sum_{i=0}^{\infty} \mathbf{1}_{\{i2^{-n} \leq T < (i+1)2^{-n}\}} B_{i2^{-n}},$$

并注意对每个 $s \geq 0$ ，随机变量 $B_s \mathbf{1}_{\{s \leq T\}}$ 都是 $\mathcal{F}_T$ -可测的。

定理 14.5.1 (强马尔可夫性质) 设 T 是停时且 $P(T < \infty) > 0$ 。在条件概率 $P(\cdot | T < \infty)$ 下，由

$$B_t^{(T)} = B_{T+t} - B_T$$

定义的过程 $(B_t^{(T)})_{t \geq 0}$ 是从 0 出发的布朗运动，并且独立于 $\mathcal{F}_T$ 。

注。严格说，还需在 $\{T = \infty\}$ 上定义 $B^{(T)}$ 。例如，可在 $T(\omega) = \infty$ 时令 $B_t^{(T)}(\omega) = 0$ ($t \geq 0$)；此选择不影响上述条件结论。

证明。先设 $T < \infty$ 几乎必然。对 $A \in \mathcal{F}_T$ 、 $0 \leq t_1 < \dots < t_p$ ，以及 $F : (\mathbb{R}^d)^p \rightarrow \mathbb{R}_+$ 有界连续，我们证明

$$E[\mathbf{1}_A F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})] = P(A) E[F(B_{t_1}, \dots, B_{t_p})].$$

这足以证明定理：取 $A = \Omega$ 得 $B^{(T)}$ 是布朗运动（其轨道连续）；而 (14.3) 表明，对任意 $0 \leq t_1 < \dots < t_p$ ，向量 $(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})$ 与 $\mathcal{F}_T$ 独立，因此 $B^{(T)}$ 与 $\mathcal{F}_T$ 独立。为证 (14.3)，先注意到几乎必然有

$$\begin{aligned}
& F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)}) \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \mathbf{1}_{\{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \dots, B_{k2^{-n}+t_p} - B_{k2^{-n}}),
\end{aligned}$$

因此, 根据支配收敛定理,

$$\begin{aligned}
& E[\mathbf{1}_A F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})] \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} E[\mathbf{1}_A \mathbf{1}_{\{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \dots, B_{k2^{-n}+t_p} - B_{k2^{-n}})]
\end{aligned}$$

对 $A \in \mathcal{F}_T$, 事件

$$A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}$$

是 $\mathcal{F}_{k2^{-n}}$ -可测的。由简单马尔可夫性质,

$$\begin{aligned}
& E[\mathbf{1}_{A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \dots, B_{k2^{-n}+t_p} - B_{k2^{-n}})] \\
&= P(A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}) E[F(B_{t_1}, \dots, B_{t_p})],
\end{aligned}$$

对 k 求和即得 (14.3)。若 $P(T = \infty) > 0$, 同样的逼近给出

$$E[\mathbf{1}_{A \cap \{T < \infty\}} F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})] = P(A \cap \{T < \infty\}) E[F(B_{t_1}, \dots, B_{t_p})]$$

并且由此再次得出了所需的结果。马尔可夫强性质的一个重要应用是在下面定理证明中所示的反射原理。

定理 14.5.2 设 $d = 1$ 。对每个 $t \geq 0$, 记

$$S_t = \sup_{0 \leq s \leq t} B_s.$$

若 $a \geq 0$ 且 $b \leq a$, 则

$$P(S_t \geq a, B_t \leq b) = P(B_t \geq 2a - b).$$

特别地, S_t 与 $|B_t|$ 同分布。

证明。 我们将马尔可夫强性质应用于停时

$$T_a = \inf\{t \geq 0, B_t = a\}.$$

由推论 14.4.3, $T_a < \infty$ 几乎必然。于是

$$P(S_t \geq a, B_t \leq b) = P(T_a \leq t, B_t \leq b) = P(T_a \leq t, B_{t-T_a}^{(T_a)} \leq b - a),$$

因为 $B_{t-T_a}^{(T_a)} = B_t - B_{T_a} = B_t - a$ 。记 $B' = B^{(T_a)}$ 。由定理 14.5.1, B' 是与 $\mathcal{F}_{T_a}$ 独立的布朗运动, 特别地与 T_a 独立。又因 B' 与 $-B'$ 同分布, (T_a, B') 与 $(T_a, -B')$ 也同分布。令

$$H = \{(s, w) \in \mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R}) : s \leq t, w(t-s) \leq b-a\}.$$

前面的概率等于

$$\begin{aligned}
\mathbb{P}((T_a, B') \in H) &= \mathbb{P}((T_a, -B') \in H) \\
&= \mathbb{P}(T_a \leq t, -B_{t-T_a}^{(T_a)} \leq b-a) \\
&= \mathbb{P}(T_a \leq t, B_t \geq 2a-b) \\
&= \mathbb{P}(B_t \geq 2a-b),
\end{aligned}$$

最后一个等式使用了事件 $\{B_t \geq 2a-b\} \subset \{T_a \leq t\}$ 。为证明第二条，注意

$$\begin{aligned}
\mathbb{P}(S_t \geq a) &= \mathbb{P}(S_t \geq a, B_t \geq a) + \mathbb{P}(S_t \geq a, B_t \leq a) \\
&= 2\mathbb{P}(B_t \geq a) = \mathbb{P}(|B_t| \geq a).
\end{aligned}$$

因此二元随机变量 (S_t, B_t) 的联合密度为

$$g(a, b) = \frac{2(2a-b)}{\sqrt{2\pi t^3}} \exp\left(-\frac{(2a-b)^2}{2t}\right) \mathbf{1}_{\{a>0, b<a\}}.$$

推论 14.5.3 ($d=1$) 对每个 $a>0$ ，停时 T_a 与 a^2/B_1^2 同分布，因此其密度为

$$f(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right) \mathbf{1}_{\{t>0\}}.$$

证明。我们写出

$$\begin{aligned}
\mathbb{P}(T_a \leq t) &= \mathbb{P}(S_t \geq a) = \mathbb{P}(|B_t| \geq a) \\
&= \mathbb{P}(B_t^2 \geq a^2) = \mathbb{P}(tB_1^2 \geq a^2) = \mathbb{P}\left(\frac{a^2}{B_1^2} \leq t\right).
\end{aligned}$$

又因 $B_1 \sim \mathcal{N}(0, 1)$ ，对变量 a^2/B_1^2 作直接换元，即可得到上述密度。

规范空间上的改写。为便于后续应用，把马尔可夫性质改写 to 规范空间 $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ 上。从现在起，在该空间上取坐标过程 $B_t(w) = w(t)$ 与规范滤过 $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t)$ 。回忆：对每个 $x \in \mathbb{R}^d$ ，在 $\mathbb{P}_x$ 下， $(B_t)_{t \geq 0}$ 是从 x 出发的布朗运动。

再引入平移算子。对每个 $s \geq 0$ ，定义 $\theta_s : \Omega \rightarrow \Omega$ 为 $(\theta_s \omega)(t) = \omega(s+t)$ ($t \geq 0$)。

等价地， $B_t \circ \theta_s = B_{s+t}$ 。

定理 14.5.4 设 T 是停时， F, G 是 Ω 上的两个非负可测函数，并假设 F 是 $\mathcal{F}_T$ -可测的。则对任意 $x \in \mathbb{R}^d$ ，

$$E_x[1_{\{T<\infty\}} F \cdot G \circ \theta_T] = E_x[1_{\{T<\infty\}} F \cdot E_{B_T}[G]].$$

注。可将此式与离散时间的定理 13.3.5 对照。

证明。由空间平移只需处理 $x=0$ 。为简化记号，先设 $P_0(T < \infty) = 1$ 。关键观察是：当 $T(\omega) < \infty$ 时，

$$(\theta_T \omega)(t) = \omega(T+t) = \omega(T) + (\omega(T+t) - \omega(T)) = B_T(\omega) + B_t^{(T)}(\omega).$$

于是

$$\begin{aligned} E_0[1_{\{T<\infty\}} F \cdot G \circ \theta_T] &= E_0[1_{\{T<\infty\}} F \cdot G(B_T + B_{\bullet}^{(T)})] \\ &= E_0\left[1_{\{T<\infty\}} F \cdot E_0[G(B_T + B_{\bullet}^{(T)}) \mid \mathcal{F}_T]\right]. \end{aligned}$$

这里 $B_{\bullet}^{(T)}$ 表示连续函数 $(B_t^{(T)})_{t \geq 0}$, 把它看成取值于 $C(\mathbb{R}_+, \mathbb{R}^d)$ 的随机变量. B_T 是 $\mathcal{F}_T$ -可测的; 另一方面, 由定理 14.5.1, $B_{\bullet}^{(T)}$ 独立于 $\mathcal{F}_T$, 且其分布为 P_0 . 应用定理 11.3.4,

$$E_0[G(B_T + B_{\bullet}^{(T)}) \mid \mathcal{F}_T] = \int P_0(d\omega) G(B_T + \omega) = E_{B_T}[G].$$

代回上一式即得结论。

14.6 调和函数与 Dirichlet 问题

第 7 章已经引入球面 S^{d-1} 上的 Lebesgue 测度 ω_d . 把 ω_d 归一化, 得到 S^{d-1} 上的均匀概率测度 σ_d . 根据第 7 章, 它与 $\mathbb{R}^d$ 上的 Lebesgue 测度 λ_d 满足

$$\sigma_d(A) = \frac{\Gamma(\frac{d}{2} + 1)}{\pi^{d/2}} \lambda_d(\{rx : 0 \leq r \leq 1, x \in A\}).$$

对每个 Borel 集 $A \subset S^{d-1}$ 都成立. 和 ω_d 一样, σ_d 在线性等距变换作用下不变. 此外, 定理 7.2.1 给出极坐标积分公式: 对每个非负 Borel 函数 $f: \mathbb{R}^d \rightarrow \mathbb{R}_+$,

$$\int_{\mathbb{R}^d} f(x) dx = c_d \int_0^\infty \int_{S^{d-1}} f(rz) r^{d-1} dr \sigma_d(dz). \quad (14.4)$$

其中 $c_d = 2\pi^{d/2}/\Gamma(d/2)$.

引理 14.6.1 σ_d 是球面 S^{d-1} 上唯一在线性等距变换作用下不变的概率测度。

证明。 设 μ 是 S^{d-1} 上另一个在线性等距变换作用下不变的概率测度. 对任意 $\xi \in \mathbb{R}^d$ 和任意线性等距映射 Φ ,

$$\begin{aligned} \hat{\mu}(\xi) &= \int_{S^{d-1}} e^{i\xi \cdot x} \mu(dx) = \int_{S^{d-1}} e^{i\xi \cdot \Phi^{-1}(x)} \mu(dx) \\ &= \int_{S^{d-1}} e^{i\Phi(\xi) \cdot x} \mu(dx) = \hat{\mu}(\Phi(\xi)). \end{aligned}$$

由此可知, $\hat{\mu}(\xi)$ 只依赖于 $|\xi|$. 因此存在函数 $f: \mathbb{R}_+ \rightarrow \mathbb{C}$, 使得对每个 $\xi \in \mathbb{R}^d$,

$$\hat{\mu}(\xi) = f(|\xi|).$$

同理, 存在函数 $g: \mathbb{R}_+ \rightarrow \mathbb{C}$, 使得

$$\hat{\sigma}_d(\xi) = g(|\xi|).$$

于是, 对每个 $r \geq 0$,

$$\int_{S^{d-1}} \left(\int_{S^{d-1}} e^{ir\xi \cdot x} \mu(dx) \right) \sigma_d(d\xi) = \int_{S^{d-1}} f(r) \sigma_d(d\xi) = f(r)$$

根据 Fubini 定理, 这也等于

$$\int_{S^{d-1}} \left(\int_{S^{d-1}} e^{irx \cdot \xi} \sigma_d(d\xi) \right) \mu(dx) = \int_{S^{d-1}} g(r) \mu(dx) = g(r).$$

因此 $f = g$, 从而 $\hat{\mu} = \hat{\sigma}_d$. 由定理 8.2.4, $\mu = \sigma_d$. 若 $x \in \mathbb{R}^d$, $r > 0$, 记 $B(x, r)$ 与 $\bar{B}(x, r)$ 分别为以 x 为中心、半径为 r 的开球与闭球。以 x 为中心、半径为 r 的球面上的均匀概率测度记为 $\sigma_{x,r}$; 按定义, 它是 $\sigma_d(dy)$ 在映射 $y \mapsto x + ry$ 下的像测度。再次说明: 本章以下部分始终使用第 14.3 节末构造的典范空间上的布朗运动。

命题 14.6.2 设 $x \in \mathbb{R}^d$, $r > 0$, 并令停时

$$S = \inf\{t \geq 0 : |B_t - x| \geq r\}.$$

则在 $\mathbb{P}_x$ 下, B_S 的分布为均匀概率测度 $\sigma_{x,r}$.

证明. 通过空间平移与尺度变换, 只需处理 $x = 0, r = 1$. 由一维布朗运动的无界性可知 $S < \infty$ 几乎必然; 路径连续性给出 $|B_S| = 1$. 布朗运动的旋转不变性又说明 B_S 的分布在线性等距变换下不变。由引理 14.6.1, 该分布只能是 σ_d . 回忆: 区域是 $\mathbb{R}^d$ 中的连通开集。函数 $h : D \rightarrow \mathbb{R}$ 若 D 的每个紧子集上有界, 就称为局部有界。

定义 14.6.1 设 D 是 $\mathbb{R}^d$ 中的区域。局部有界可测函数 $h : D \rightarrow \mathbb{R}$ 称为调和函数, 如果对任意 $x \in D$ 和 $r > 0$, 只要 $\bar{B}(x, r) \subset D$, 就有

$$h(x) = \int h(y) \sigma_{x,r}(dy).$$

换言之, 只要闭球 $\bar{B}(x, r)$ 包含于 D , $h(x)$ 就等于函数 h 在以 x 为中心、半径为 r 的球面上的平均值。

经典 Dirichlet 问题. 给定有界区域 D 和连续边界函数 $g : \partial D \rightarrow \mathbb{R}$, 希望找到 $h : D \rightarrow \mathbb{R}$, 使得:

- h 以连续极限方式取得边界值 g : 对每个 $y \in \partial D$,

$$g(y) = \lim_{x \rightarrow y, x \in D} h(x);$$

- h 在 D 上调和。

下列定理先给出一个自然的候选函数。

定理 14.6.3 设 D 是有界区域, g 是 ∂D 上的有界可测函数。记退出时刻

$$T = \inf\{t \geq 0 : B_t \notin D\}.$$

则函数

$$h(x) = \mathbb{E}_x[g(B_T)], \quad x \in D$$

在 D 上调和。

这与上一章末离散调和函数的结果完全对应。

证明。由路径连续性, 对任意 $t \geq 0$,

$$\{T \leq t\} = \left\{ \inf_{\substack{0 \leq s \leq t \\ s \in \mathbb{Q}}} \text{dist}(B_s, D^c) = 0 \right\},$$

故 T 是停时。因为 D 有界, 利用任一坐标的一维布朗运动无界性可知 $T < \infty$, P_x -几乎必然。再由路径连续性可知 $B_T \in \partial D$, 所以 $E_x[g(B_T)]$ 有定义, 且绝对值不超过 $\sup_{y \in \partial D} |g(y)|$ 。下面验证 h 可测。记 $\mathcal{C}$ 为 $C(\mathbb{R}_+, \mathbb{R}^d)$ 上的典范 σ -代数。对任意 $A \in \mathcal{C}$, 映射 $x \mapsto P_x(A)$ 可测: 先对柱集用有限维分布的显式公式验证, 再用单调类定理推广。因此, 对路径空间上的任意有界可测函数 F , 映射 $x \mapsto E_x[F]$ 可测。把这一结论应用于

$$F(w) = \mathbf{1}_{\{T(w) < \infty\}} g(B_{T(w)}(w)) = \mathbf{1}_{\{T(w) < \infty\}} g(w(T(w))),$$

即可得到 h 的可测性。

现在固定 $x \in D$ 与 $r > 0$, 使闭球 $\overline{B}(x, r) \subset D$ 。令

$$S = \inf\{t \geq 0 : B_t \notin B(x, r)\} = \inf\{t \geq 0 : |B_t - x| \geq r\}.$$

显然, 在 P_x 下几乎必然有 $S \leq T$; 事实上, 对典范空间 $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ 中每个 w , 都有 $S(w) \leq T(w)$ 。此外, 在 P_x 下几乎必然有 $B_T = B_T \circ \theta_S$ 。

这只是说: 若 $t \mapsto w(t)$ 是一条从 x 出发的路径, 那么它离开 D 的位置, 和删去从出发时刻到离开球 $B(x, r)$ 这一段后所得路径离开 D 的位置相同。这是因为 $\overline{B}(x, r) \subset D$ 。于是可把强马尔可夫性质按定理 14.5.4 的形式应用于停时 S , 得到

$$\begin{aligned} h(x) &= \mathbb{E}_x[g(B_T)] = \mathbb{E}_x[g(B_T) \circ \theta_S] = \mathbb{E}_x[\mathbb{E}_{B_S}[g(B_T)]] \\ &= \mathbb{E}_x[h(B_S)] = \int h(y) \sigma_{x,r}(dy). \end{aligned}$$

最后一个等号使用了命题 14.6.2, 证明完成。要使前一定理给出的函数成为经典 Dirichlet 问题的解 (再假设边界函数 g 连续), 还需证明对每个 $y \in \partial D$,

$$g(y) = \lim_{x \rightarrow y, x \in D} \mathbb{E}_x[g(B_T)].$$

直观上, 若 $x \in D$ 接近 $y \in \partial D$, 从 x 出发的布朗运动会很快离开 D , 因而退出点 B_T 既接近 x , 也接近 y ; 由 g 的连续性, $g(B_T)$ 将接近 $g(y)$, 从而得到所需结论。在附加适当假设并严格化这一论证之前, 我们先处理解的唯一性。下一个命题说明调和函数自动具有很强的正则性。

命题 14.6.4 若 h 在 D 上调和, 则 $h \in C^\infty(D)$ 。此外, 若 $x \in D$ 、 $r > 0$ 且 $\overline{B}(x, r) \subset D$, 则

$$h(x) = \frac{1}{\lambda_d(B(x, r))} \int_{B(x, r)} h(y) dy.$$

证明。取 $r_0 > 0$, 并令

$$D_0 = \{x \in D : \text{dist}(x, D^c) > r_0\}.$$

只需证明 $h \in C^\infty(D_0)$ 。取非零函数 $\varphi \in C_c^\infty((0, r_0))$, 且 $\varphi \geq 0$ 。对任意 $x \in D_0$ 和 $r \in (0, r_0)$,

$$h(x) = \int \sigma_{x,r}(dz) h(z) = \int \sigma_d(dy) h(x + ry).$$

把这个等式的首、末两端乘以 $r^{d-1}\varphi(r)$, 再对 $r \in (0, r_0)$ 积分。由公式 (14.4), 存在只依赖于 φ 的常数 $c > 0$, 使对每个 $x \in D_0$,

$$\begin{aligned} ch(x) &= c_d \int_0^{r_0} r^{d-1} \varphi(r) dr \int_{S^{d-1}} h(x + ry) \sigma_d(dy) \\ &= \int_{B(0, r_0)} \varphi(|z|) h(x + z) dz \\ &= \int_{B(x, r_0)} \varphi(|z - x|) h(z) dz \\ &= \int_{\mathbb{R}^d} \varphi(|z - x|) \tilde{h}(z) dz. \end{aligned}$$

³ 在最后一个等式中, $\tilde{h}$ 表示把 h 在 D^c 上延拓为 0 所得的函数。这个取值不影响积分, 因为若 $x \in D_0$ 、 $z \in D^c$, 则 $\varphi(|z - x|) = 0$ 。因此, 在 D_0 上, h 等于函数 $z \mapsto \varphi(|z|)$ 与可测有界函数 $\tilde{h}$ 的卷积; 前者属于 C^∞ 且具有紧支集。第 2 章末关于积分号下求导的讨论说明, 这一卷积属于 C^∞ 。还需证明第二个断言。令 $\varphi = 1_{[0, r_0]}$, 重复上述计算, 则对 $x \in D_0$ 有

$$h(x) = c' \int_{B(x, r_0)} dy h(y)$$

其中 c' 只依赖于 r_0 。取常值调和函数 $h \equiv 1$, 可知 $c' = \lambda_d(B(x, r_0))^{-1}$, 从而得到 (14.6)。

推论 14.6.5 如果狄利克雷问题存在解, 则该解是唯一的。

证明。 设 h_1, h_2 是两个解, 并令 $f = h_1 - h_2$ 。假设 f 不恒为零。必要时交换 h_1, h_2 , 可设 f 在某点取严格正值。把 f 在 ∂D 上延拓为 0, 所得函数在 $\bar{D}$ 上连续, 因而在 D 内达到其正最大值 M (D 有界, 故 $\bar{D}$ 紧)。取 $x_0 \in D$ 使 $f(x_0) = M$ 。由前一命题, 对每个 $0 < r < \text{dist}(x_0, D^c)$, 有

$$f(x_0) = \frac{1}{\lambda_d(B(x_0, r))} \int_{B(x_0, r)} dy f(y),$$

即

$$\int_{B(x_0, r)} dy (f(x_0) - f(y)) = 0.$$

因为对每个 $y \in D$ 都有 $f(x_0) \geq f(y)$, 所以上式只能在 $f(x_0) = f(y)$ 对 $B(x_0, r)$ 中 λ_d -几乎处处成立时成立。又因 f 连续, 故对每个 $y \in B(x_0, r)$ 都有 $f(x_0) = f(y)$ 。因此集合 $\{x \in D : f(x) = M\}$ 是开集。

另一方面, 该集合也是 D 的闭集。由于 D 连通, 必有

$$\{x \in D : f(x) = M\} = D.$$

这与 $M > 0$ 且 f 在 D 的边界处趋于零矛盾。

³校订注: 原讲义此处换元后的两行分别误排为 $h(x)$ 与 $\tilde{h}(x)$; 积分变量为 z , 正确被积函数应为 $h(z)$ 与 $\tilde{h}(z)$ 。

定义 14.6.2 若对每个 $y \in \partial D$, 都存在 $r > 0$ 和一个以 y 为顶点的开圆锥 C , 使得

$$C \cap B(y, r) \subset D^c,$$

则称区域 D 满足外锥条件。

定理 14.6.6 设 D 是满足外锥条件的有界区域, $g \in C(\partial D)$ 。则

$$h(x) = \mathbb{E}_x[g(B_T)], \quad x \in D$$

是经典 Dirichlet 问题的唯一解。

证明。 由定理 14.6.3 和推论 14.6.5, 只需证明对每个固定的 $y \in \partial D$,

$$\lim_{x \rightarrow y, x \in D} h(x) = g(y).$$

取 $\varepsilon > 0$ 。由 g 在紧集 ∂D 上一致连续, 可取 $\delta > 0$, 使得只要 $z \in \partial D$ 且 $|z - y| < \delta$, 就有

$$|g(z) - g(y)| < \frac{\varepsilon}{3}.$$

再取 $M > \sup_{z \in \partial D} |g(z)|$ 。对任意 $\eta > 0$,

$$\begin{aligned} |\mathbb{E}_x[g(B_T)] - g(y)| &\leq \mathbb{E}_x[|g(B_T) - g(y)|\mathbf{1}_{\{T \leq \eta\}}] + \mathbb{E}_x[|g(B_T) - g(y)|\mathbf{1}_{\{T > \eta\}}] \\ &\leq \mathbb{E}_x\left[|g(B_T) - g(y)|\mathbf{1}_{\{T \leq \eta\}}\mathbf{1}_{\{\sup_{t \leq \eta} |B_t - x| \leq \delta/2\}}\right] \\ &\quad + 2M\mathbb{P}_x\left(\sup_{t \leq \eta} |B_t - x| > \frac{\delta}{2}\right) + 2MP_x(T > \eta) \\ &=: I + II + III. \end{aligned}$$

分别估计 I, II, III 。若 $|x - y| < \delta/2$, 则在事件 $\{T \leq \eta\} \cap \{\sup_{t \leq \eta} |B_t - x| \leq \delta/2\}$ 上,

$$|B_T - y| \leq |B_T - x| + |x - y| < \delta$$

所以 δ 的选择给出 $I \leq \varepsilon/3$ 。由平移不变性,

$$II = 2M\mathbb{P}_0\left(\sup_{t \leq \eta} |B_t| > \frac{\delta}{2}\right)$$

最后只需再取足够小的 $\alpha \in (0, \delta/2]$, 使得当 $|x - y| < \alpha$ 时, $III = 2MP_x(T > \eta) \leq \varepsilon/3$ 。下面的引理正好保证这一点, 从而定理得证。

引理 14.6.7 在外锥条件下, 对任意 $y \in \partial D$ 与 $\eta > 0$, 都有

$$\lim_{\substack{x \rightarrow y \\ x \in D}} P_x(T > \eta) = 0.$$

注。 验证边界条件 (14.7) 的关键, 是证明从 D 的边界附近出发的布朗运动以高概率迅速离开 D 。本引理精确表达了这一事实。外锥条件不是最优条件, 但足以支撑本讲义的应用。

证明。先重写点 $y \in \partial D$ 处的外锥条件。对 $u \in S^{d-1}$ 、 $\gamma > 0$ ，令

$$C(u, \gamma) = \{z \in \mathbb{R}^d : z \cdot u > (1 - \gamma)|z|\}.$$

它是以 0 为顶点、方向为 u 、开口参数为 γ 的开圆锥。可取 $r > 0$ 、 $u \in S^{d-1}$ 和 $\gamma > 0$ ，使得

$$y + (C(u, \gamma) \cap B(0, r)) \subset D^c.$$

为简化记号，令 $C = C(u, \gamma) \cap B(0, r)$ ，并再令

$$\tilde{C} = \{z \in \mathbb{R}^d : z \cdot u > (1 - \gamma/2)|z|\} \cap B(0, r/2).$$

它是把 $C(u, \gamma)$ 缩窄并截在 $B(0, r/2)$ 中所得的圆锥。若 $T_{\tilde{C}} = \inf\{t \geq 0 : B_t \in \tilde{C}\}$ ，则由布鲁门塔尔全或无律可得 $P_0(T_{\tilde{C}} = 0) = 1$ 。⁴

事实上，若 (ε_n) 严格递减到 0，则事件 $\limsup_n \{B_{\varepsilon_n} \in \tilde{C}\}$ 属于 $\mathcal{F}_{0+}$ ；与推论 14.4.3 的证明相同的论证表明这个事件的概率严格为正，故由全或无律其概率为 1。对 $0 < a < r/2$ ，记

$$\tilde{C}_a = \tilde{C} \cap B(0, a)^c.$$

当 $a \downarrow 0$ 时， $\tilde{C}_a \uparrow \tilde{C}$ ，故 $T_{\tilde{C}_a} \downarrow T_{\tilde{C}} = 0$ 几乎必然。因此，对任意 $\beta > 0$ ，可把 a 取得足够小，使

$$P_0(T_{\tilde{C}_a} \leq \eta) > 1 - \beta.$$

利用 $y + C \subset D^c$ ，并采用显然的 hitting time 记号，可得

$$P_x(T \leq \eta) \geq P_x(T_{y+C} \leq \eta) = P_0(T_{y-x+C} \leq \eta).$$

一个简单的几何观察表明，当 $|y - x|$ 足够小时，平移圆锥 $y - x + C$ 包含 $\tilde{C}_a$ 。于是

$$P_x(T \leq \eta) \geq P_0(T_{\tilde{C}_a} \leq \eta) > 1 - \beta.$$

其中最后一个不等式来自 a 的选择。由于 $\beta > 0$ 任意，引理得证。下面讨论调和函数的另一个解析刻画，它通常也被直接用作调和函数的定义。

命题 14.6.8 设 h 是区域 D 上的局部有界函数。则 h 调和，当且仅当 $h \in C^2(D)$ 且 $\Delta h = 0$ 。

证明。先设 h 调和。由命题 14.6.4， $h \in C^\infty(D)$ 。固定 $x \in D$ ，并取 $r_0 > 0$ 使 $\bar{B}(x, r_0) \subset D$ 。仍由命题 14.6.4，对每个 $r \in (0, r_0)$ ，

$$h(x) = \frac{1}{\lambda_d(B(x, r))} \int_{B(x, r)} h(y) dy.$$

另一方面，对 $y \in B(x, r)$ 使用二阶 Taylor 公式，得到

$$h(y) = h(x) + \sum_{i=1}^d \frac{\partial h}{\partial y_i}(x) (y_i - x_i) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 h}{\partial y_i \partial y_j}(x) (y_i - x_i)(y_j - x_j) + o(r^2)$$

其中余项 $o(r^2)$ 对 $y \in B(x, r)$ 一致。对上式在 $B(x, r)$ 上积分，并利用球的对称性，得到

⁴校订注：原讲义此处把较大的圆锥误写成了 $C(u, r)$ ；圆锥的开口参数应为 γ 。

$$\int_{B(x,r)} h(y) dy = \lambda_d(B(x,r)) h(x) + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 h}{\partial y_i^2}(x) \int_{B(x,r)} (y_i - x_i)^2 dy + o(r^{d+2}).$$

令 $C_1 = \int_{B(0,1)} y_1^2 dy > 0$ 。前式与 (14.8) 给出

$$\frac{C_1}{2} \Delta h(x) r^{d+2} + o(r^{d+2}) = 0$$

这只有在 $\Delta h(x) = 0$ 时才可能。反过来, 设 $h \in C^2(D)$ 且 $\Delta h = 0$ 。只需证明: 若 U 是满足 $\bar{U} \subset D$ 的开球, 则 h 在 U 上调和。由定理 14.6.6, 存在唯一函数 $\tilde{h}$, 它在 $\bar{U}$ 上连续、在 U 内调和, 且在 ∂U 上与 h 相等。证明的第一部分又说明 $\Delta \tilde{h} = 0$ 于 U 。把下面的最大值原理分别应用于 $h - \tilde{h}$ 与 $\tilde{h} - h$, 便得 $h = \tilde{h}$ 于 $\bar{U}$, 命题得证。

引理 14.6.9 (最大值原理) 设 $V \subset \mathbb{R}^d$ 是有界开集, $u \in C(\bar{V}) \cap C^2(V)$, 并且 $\Delta u \geq 0$ 于 V 。则

$$\sup_{x \in \bar{V}} u(x) = \sup_{x \in \partial V} u(x).$$

证明。先假设更强的条件 $\Delta u > 0$ 于 V 。反设

$$\sup_{x \in \bar{V}} u(x) > \sup_{x \in \partial V} u(x).$$

则存在 $x_0 \in V$ 使

$$u(x_0) = \sup_{x \in V} u(x).$$

此时

$$\frac{\partial u}{\partial y_j}(x_0) = 0, \quad \forall j \in \{1, \dots, d\}$$

而且二阶 Taylor 公式说明 Hessian 矩阵

$$M_{x_0} = \left(\frac{\partial^2 u}{\partial y_i \partial y_j}(x_0) \right)_{i,j \in \{1, \dots, d\}}$$

是半负定的, 即相应二次型只取非正值。因此 M_{x_0} 的所有特征值均非正, 其迹也非正。但 $\text{tr}(M_{x_0}) = \Delta u(x_0) > 0$, 矛盾。若只假设 $\Delta u \geq 0$, 对每个 $\varepsilon > 0$ 令

$$u_\varepsilon(x) = u(x) + \varepsilon x_1^2,$$

则 $\Delta u_\varepsilon = \Delta u + 2\varepsilon > 0$ 。由已证情形,

$$\sup_{x \in \bar{V}} u_\varepsilon(x) = \sup_{x \in \partial V} u_\varepsilon(x),$$

最后令 $\varepsilon \downarrow 0$ 即得结论。

14.7 调和函数与布朗运动

先说明连续时间鞅。仍在第 14.3 节末的布朗运动规范空间上，令 $\mathcal{F}_t = \sigma(B_s : s \leq t)$ 。一族可积随机变量 $(M_t)_{t \geq 0}$ 称为鞅，如果对每个 $t \geq 0$ ， M_t 是 $\mathcal{F}_t$ -可测的，并且对任意 $0 \leq s \leq t$,

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

若 $U \subset \mathbb{R}^d$ 是开集，记 $H_U = \inf\{t \geq 0 : B_t \notin U\}$ 。

定理 14.7.1 设 $D \subset \mathbb{R}^d$ 是区域。连续函数 $h : D \rightarrow \mathbb{R}$ 调和，当且仅当对每个满足 $\bar{U} \subset D$ 且具有外锥条件的有界开集 U ，过程

$$(h(B_{t \wedge H_U}))_{t \geq 0}$$

对每个 $x \in U$ ，在 $\mathbb{P}_x$ 下都是鞅。

直观地说，调和函数与布朗运动复合后产生鞅。上述外锥条件对结论本身不是必需的，只是本讲义的证明需要它。**证明。**先设 h 调和，并取满足定理条件的开集 U 。简记 $H = H_U$ ，固定 $x \in U$ 。

随机变量 $h(B_{t \wedge H})$ 在 $\mathbb{P}_x$ 下几乎必然由 $\sup\{|h(y)| : y \in \bar{U}\} < \infty$ 控制。设 $s \leq t$ 。注意随机变量 $B_{s \wedge H}$ 是 $\mathcal{F}_{s \wedge H}$ -可测的，因而也是 $\mathcal{F}_s$ -可测的。要证明所需等式 $\mathbb{E}_x[h(B_{t \wedge H}) | \mathcal{F}_s] = h(B_{s \wedge H})$ ，只需证明对任意有界的 $\mathcal{F}_s$ -可测随机变量 F ，都有

$$\mathbb{E}_x[F h(B_{s \wedge H})] = \mathbb{E}_x[F h(B_{t \wedge H})].$$

可把 h 看成区域 U 上、边界值为 $h|_{\partial U}$ 的 Dirichlet 问题的唯一解。由定理 14.6.6，对每个 $y \in U$ ，

$$h(y) = \mathbb{E}_y[h(B_H)].$$

因此得出

$$\mathbb{E}_x[F \mathbf{1}_{\{s < H\}} h(B_{s \wedge H})] = \mathbb{E}_x[F \mathbf{1}_{\{s < H\}} h(B_s)] = \mathbb{E}_x[F \mathbf{1}_{\{s < H\}} \mathbb{E}_{B_s}[h(B_H)]].$$

由于 $F \mathbf{1}_{\{s < H\}}$ 是 $\mathcal{F}_s$ -可测的（练习），把定理 14.5.4 应用于常停时 s ，得到

$$\mathbb{E}_x[F \mathbf{1}_{\{s < H\}} \mathbb{E}_{B_s}[h(B_H)]] = \mathbb{E}_x[F \mathbf{1}_{\{s < H\}} h(B_H)].$$

由此我们得到

$$\mathbb{E}_x[F h(B_{s \wedge H})] = \mathbb{E}_x[F \mathbf{1}_{\{s < H\}} h(B_s)] + \mathbb{E}_x[F \mathbf{1}_{\{s \geq H\}} h(B_H)] = \mathbb{E}_x[F h(B_H)].$$

显然，同样的论证表明

$$\mathbb{E}_x[F h(B_{t \wedge H})] = \mathbb{E}_x[F h(B_H)] = \mathbb{E}_x[F h(B_{s \wedge H})]$$

反过来更简单。若 h 满足定理中的鞅性质，取闭包包含于 D 的开球 U 。对 $x \in U$ ，鞅性质给出

$$h(x) = \mathbb{E}_x[h(B_{t \wedge H}) | \mathcal{F}_0] = \mathbb{E}_x[h(B_{t \wedge H})].$$

令 $t \rightarrow \infty$ ，由有界收敛定理得 $h(x) = \mathbb{E}_x[h(B_H)]$ 。定理 14.6.3 表明 h 在 U 上调和；这足以推出 h 在 D 上调和。以下假设 $d \geq 2$ ；在一维，调和函数就是仿射函数。

命题 14.7.2 设 $0 \leq a < b$, 并令 $D_{a,b}$ 为区域

$$D_{a,b} = B(0, b) \setminus \bar{B}(0, a).$$

若径向函数 $f: D_{a,b} \rightarrow \mathbb{R}$ 只依赖于 $|x|$, 则 f 调和, 当且仅当存在 $C, C' \in \mathbb{R}$, 使

$$f(x) = \begin{cases} C + C' \log |x|, & d = 2, \\ C + C'|x|^{2-d}, & d \geq 3. \end{cases}$$

证明。调和性先给出 $f \in C^\infty$ 。令 $g: (a, b) \rightarrow \mathbb{R}$ 满足 $f(x) = g(|x|)$ 。径向函数的拉普拉斯算子为

$$\Delta f(x) = g''(|x|) + \frac{d-1}{|x|} g'(|x|).$$

由命题 14.6.8, f 调和当且仅当 g 满足微分方程

$$g''(r) + \frac{d-1}{r} g'(r) = 0.$$

解这个方程就得到命题中的两种表达式。以下两个陈述中, 对任意闭集 $A \subset \mathbb{R}^d$, 记

$$T_A = \inf\{t \geq 0 : B_t \in A\}.$$

命题 14.7.3 设 $x \in \mathbb{R}^d \setminus \{0\}$, 且 $\varepsilon, R > 0$ 满足 $\varepsilon < |x| < R$ 。则

$$P_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c}) = \begin{cases} \frac{\log R - \log |x|}{\log R - \log \varepsilon}, & d = 2, \\ \frac{|x|^{2-d} - R^{2-d}}{\varepsilon^{2-d} - R^{2-d}}, & d \geq 3. \end{cases} \quad (14.9)$$

注。一维的对应结论是: 若 $a < x < b$, 则

$$P_x(T_a < T_b) = \frac{b-x}{b-a}$$

证明。考虑满足外锥条件的区域 $D_{\varepsilon,R} = B(0, R) \setminus \bar{B}(0, \varepsilon)$, 并在 $\partial D_{\varepsilon,R}$ 上定义连续函数

$$g(y) = \begin{cases} 1, & |y| = \varepsilon, \\ 0, & |y| = R. \end{cases}$$

定理 14.6.6 表明, 函数

$$h(x) = P_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c}), \quad \varepsilon < |x| < R,$$

是边界条件为 g 的 Dirichlet 问题的唯一解。而由命题 14.7.2, 式 (14.9) 右端也是同一个 Dirichlet 问题的解, 故两者相等。由此可以推出布朗运动路径在长时间下的重要性质。

推论 14.7.4 (i) 若 $d \geq 3$, 则对任意 $\varepsilon > 0$ 及 $x \in \mathbb{R}^d$ 满足 $\varepsilon < |x|$,

$$P_x(T_{\bar{B}(0,\varepsilon)} < \infty) = \left(\frac{\varepsilon}{|x|}\right)^{d-2}.$$

此外, 对任意 $x \in \mathbb{R}^d$, 在 P_x 下几乎必然有 $\lim_{t \rightarrow \infty} |B_t| = \infty$ 。

(ii) 若 $d = 2$, 则对任意 $\varepsilon > 0$ 及 $x \in \mathbb{R}^2$ 满足 $\varepsilon < |x|$,

$$P_x(T_{\bar{B}(0,\varepsilon)} < \infty) = 1$$

但是

$$P_x(T_{\{0\}} < \infty) = 0.$$

此外, 在 P_x 下几乎必然, 对 $\mathbb{R}^2$ 的每个开集 U , 集合 $\{t \geq 0 : B_t \in U\}$ 都无界。

与马尔可夫链类似, $d \geq 3$ 时称布朗运动暂态, $d = 2$ 时称其常返。平面布朗运动的常返性并不表示它会访问每个点; 相反, 除起点外, 任一预先指定的点都几乎必然不会被访问。

证明。(i) 首先,

$$P_x(T_{\bar{B}(0,\varepsilon)} < \infty) = \lim_{n \rightarrow \infty} P_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,n)^c}).$$

对每个整数 $n \geq 1$, 再令

$$T^{(n)} = T_{B(0,2^n)^c}.$$

在停时 $T^{(n)}$ 处应用强马尔可夫性质, 并再次利用式 (14.9)。若 $|x| \leq 2^n$, 则

$$P_x\left(\inf_{t \geq T^{(n)}} |B_t| \leq n\right) = E_x\left[P_{B_{T^{(n)}}}(T_{\bar{B}(0,n)} < \infty)\right] = \left(\frac{n}{2^n}\right)^{d-2}.$$

Borel–Cantelli 引理说明, 在 P_x 下几乎必然, 当整数 n 足够大时,

$$\inf_{t \geq T^{(n)}} |B_t| > n.$$

因此 $|B_t| \rightarrow \infty$ ($t \rightarrow \infty$)。 (ii) 由式 (14.9), 当 $\varepsilon < |x| < R$ 时,

$$P_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c}) = \frac{\log R - \log |x|}{\log R - \log \varepsilon}.$$

令 $R \rightarrow \infty$, 得到

$$P_x(T_{\bar{B}(0,\varepsilon)} < \infty) = 1.$$

在同一公式中令 $\varepsilon \downarrow 0$, 得到

$$P_x(T_{\{0\}} < T_{B(0,R)^c}) = 0.$$

又因为 $T_{B(0,R)^c} \uparrow \infty$ ($R \uparrow \infty$), 所以

$$P_x(T_{\{0\}} < \infty) = 0.$$

这两条性质先对正有理数 ε 同时成立, 再由单调性推广到所有 $\varepsilon > 0$ 。若到各个缩小球的首次到达时刻没有趋于无穷, 则由路径连续性会实际到达 0, 与第二条性质矛盾。因此, 几乎必然地, 0 是

$t \rightarrow \infty$ 时路径的聚点。故对每个含 0 的开集 U , 集合 $\{t \geq 0 : B_t \in U\}$ 无界。再作空间平移, 并只取一个可数开集基, 即得推论最后一条。

Poisson 核。以下仍设 $d \geq 2$ 。单位球的 Poisson 核是定义在 $B(0, 1) \times S^{d-1}$ 上的函数

$$K(x, y) = \frac{1 - |x|^2}{|x - y|^d}, \quad x \in B(0, 1), y \in S^{d-1}.$$

引理 14.7.5 对每个固定的 $y \in S^{d-1}$, 函数 $x \mapsto K(x, y)$ 在 $B(0, 1)$ 上调和。

证明。 对 $x \in B(0, 1)$, 令 $K_y(x) = K(x, y)$ 。直接计算得到 $\Delta K_y = 0$, 再应用命题 14.6.8 即可。

引理 14.7.6 对每个 $x \in B(0, 1)$, 都有

$$\int_{S^{d-1}} K(x, y) \sigma_d(dy) = 1.$$

证明。 对 $x \in B(0, 1)$, 令

$$F(x) = \int_{S^{d-1}} K(x, y) \sigma_d(dy).$$

由前一引理容易推出 F 在 $B(0, 1)$ 上调和: 可用 Fubini 定理验证平均值性质, 也可在积分号下求导证明 $\Delta F = 0$ 。又因为 σ_d 与 K 在线性等距变换下不变, F 是径向函数。在去心球 $B(0, 1) \setminus \{0\}$ 上, F 必须具有命题 14.7.2 中的形式; 但 F 在 0 点连续, 所以其中的常数 C' 必须为 0。故对所有 $x \in B(0, 1)$, $F(x) = F(0) = 1$ 。

定理 14.7.7 设 g 是 S^{d-1} 上的连续函数。单位球 $B(0, 1)$ 中、边界条件为 g 的 Dirichlet 问题的解为

$$h(x) = \int_{S^{d-1}} K(x, y) g(y) \sigma_d(dy), \quad x \in B(0, 1).$$

此外, 对每个固定的 $x \in B(0, 1)$, 函数 $y \mapsto K(x, y)$ 是布朗运动在 P_x 下离开 $B(0, 1)$ 时出口点分布关于 $\sigma_d(dy)$ 的密度。

证明。 与引理 14.7.6 相同的论证表明, h 在 $B(0, 1)$ 上调和。为验证边界条件, 固定 $y_0 \in S^{d-1}$ 。任取 $\delta > 0$ 。Poisson 核的显式公式表明: 若 $x \in B(0, 1)$ 、 $y \in S^{d-1}$, 且 $|x - y_0| < \delta/2$ 、 $|y - y_0| > \delta$, 则

$$K(x, y) \leq \left(\frac{2}{\delta}\right)^d (1 - |x|^2).$$

因此, 对每个 $\delta > 0$,

$$\lim_{\substack{x \rightarrow y_0 \\ x \in B(0, 1)}} \int_{\{|y - y_0| > \delta\}} K(x, y) \sigma_d(dy) = 0. \quad (14.10)$$

⁵ 给定 $\varepsilon > 0$, 取充分小的 $\delta > 0$, 使当 $y \in S^{d-1}$ 、 $|y - y_0| \leq \delta$ 时, $|g(y) - g(y_0)| \leq \varepsilon$ 。令 $M = \sup\{|g(y)| : y \in S^{d-1}\}$, 则

⁵校订注: 原讲义在式 (14.10) 中把 σ_d 的下标 (d) 漏排; 此处与本节其余记号统一。

$$\begin{aligned}
|h(x) - g(y_0)| &= \left| \int_{S^{d-1}} K(x, y)(g(y) - g(y_0)) \sigma_d(dy) \right| \\
&\leq 2M \int_{\{|y-y_0|>\delta\}} K(x, y) \sigma_d(dy) + \varepsilon.
\end{aligned}$$

第一个等式使用引理 14.7.6, 后一个估计使用了 δ 的选择。结合 (14.10), 得到

$$\limsup_{x \rightarrow y_0, x \in B(0,1)} |h(x) - g(y_0)| \leq \varepsilon.$$

由于 ε 任意, 所需边界条件成立。最后, 定理 14.6.6 还表明, 同一 Dirichlet 问题的解可写成

$$h(x) = \mathbb{E}_x[g(B_T)],$$

其中 $T = \inf\{t \geq 0 : B_t \notin B(0,1)\}$ 。比较 h 的两种表示, 就得到 B_T 的分布为 $K(x, y)\sigma_d(dy)$ 。

参考文献

第一部分：积分

Rudin 的经典书籍 [7] 仍然是一个很好的参考。Briane 和 Pagès 的书 [2] 非常详细且相当完整。

[1] M.R. Adams, V. Guillemin. 《测度论与概率》。Birkhäuser, 1996。

[2] M. Briane, G. Pagès. 《积分理论》。Vuibert, 2000。

[3] D.L. Cohn. 《测度论》。Birkhäuser, 1980。

[4] J.L. Doob. 《测度论》。Springer, 1994。

[5] R.M. Dudley. 《实分析与概率》。Chapman and Hall, 1989。

[6] D. Revuz. 《测度与积分》。Hermann, 1994。

[7] W. Rudin. 《实分析与复分析》。McGraw Hill, 1974。

[8] D.W. Stroock. 《积分理论简明入门》。Birkhäuser, 1994。

第二部分：概率

[9] 和 [18] 是法语著作, 其水平大致相当于课程水平。[10] 和 [11] 是经典书籍, 阅读始终有趣。[13] 和 [17] 是由顶尖概率学家撰写的新书籍。

[9] P. Barbe, M. Ledoux. 《概率》, Belin, 1998。

[10] P. Billingsley. 《概率与测度》, 第 3 版, Wiley, 1995。

[11] L. Breiman. 《概率》, Addison-Wesley, 1968。

[12] K.L. Chung. 《概率理论课程》, Harcourt Brace and World, 1968。

[13] R. Durrett. 《概率与实例》, 第 2 版, Duxbury Press, 1996。

[14] W. Feller. 《概率论及其应用导论》, 第一卷。Wiley。(一本经典著作, 介绍了在不使用测度理论的情况下可以在概率中做的所有事情)

[15] G. Grimmett, D. Stirzaker. 《概率与随机过程》。牛津科学出版, 1992 年。

[16] J. Neveu. 《概率计算的数学基础》。Masson, 1964 年。(关于概率论中测度论工具的参考书)⁶

[17] J. Pitman. 《概率》。Springer, 1993 年。

[18] D. Revuz. 《概率》。Hermann, 1997 年。

⁶校订注：原讲义把出版年份误排为 1064；此处据该书书目信息改为 1964。

[19] D.W. Stroock.《概率论：解析视角》。剑桥大学出版社, 1993 年。(一本更高级的书, 围绕分析与概率的联系)

第三部分：随机过程

[20] J. Neveu,《离散时间鞅》, Masson, 1972。

[21] D. Williams,《含鞅的概率论》, Cambridge University Press, 1991。

[22] C. Dellacherie, P.A. Meyer,《概率与势, 卷 V–VIII: 鞅理论》, Hermann, 1980。(尤其讨论连续时间鞅。)

[23] P. Baldi, L. Mazliak, P. Priouret,《鞅与马尔可夫链》, Hermann, 1998。

[24] K.L. Chung,《具有平稳转移概率的马尔可夫链》, Springer, 1967。

[25] R. Durrett,《随机过程要义》, Springer, 1999。(包含许多马尔可夫链的具体例子。)

[26] D.W. Stroock,《马尔可夫过程导论》, Springer, 2005。(适合进一步学习马尔可夫链与马尔可夫过程。)