概率论：理论和例子

Probability: Theory and Examples

原书第五版

原著作者 Rick Durrett（美国）

中文翻译 临江仙

更新发布 https://hangjuan.cn/market

更新日期 2026 年 2 月 18 日

惊鸿舟数海，一笔临江仙。

目录

第一章　测度论 **1**

第 1 节　概率空间 1

第 2 节　分布 9

第 3 节　随机变量 14

第 4 节　积分 17

第 5 节　积分的性质 22

第 6 节　期望值 26

1.6.1　不等式 26

1.6.2　极限积分 27

1.6.3　计算期望值 29

第 7 节　乘积测度、Fubini 定理 34

第二章　大数定律 **39**

第 1 节　独立性 39

2.1.1　独立的充分条件 41

2.1.2　独立性、分布与期望 43

2.1.3　独立随机变量之和 45

2.1.4　构造独立随机变量 48

第 2 节　弱大数定律 52

2.2.1　L^2 弱大数定律 52

2.2.2　三角阵列 54

2.2.3　截断 57

第 3 节　Borel-Cantelli 引理 63

第 4 节　强大数定律 71

第 5 节 随机级数的收敛 * 76
2.5.1 收敛速度 81
2.5.2 无限均值 83
第 6 节 更新理论 * 86
第 7 节 大偏差 * 99

第三章 中心极限定理 **107**
第 1 节 De Moivre-Laplace 定理 107
第 2 节 弱收敛 111
3.2.1 例子 111
3.2.2 理论 113
第 3 节 特征函数 120
3.3.1 定义、反演公式 120
3.3.2 弱收敛 127
3.3.3 矩与导数 128
3.3.4 Polya 准则 * 131
3.3.5 矩问题 * 134
第 4 节 中心极限定理 138
3.4.1 独立同分布序列 138
3.4.2 三角阵列 142
3.4.3 素因子（Erdős-Kac）* 146
3.4.4 收敛速度（Berry-Esseen）* 150
第 5 节 局部极限定理 * 155
第 6 节 Poisson 收敛 161
3.6.1 基本极限定理 161
3.6.2 两个相依的例子 165
第 7 节 Poisson 过程 169
3.7.1 复合 Poisson 过程 171
3.7.2 稀疏 172
3.7.3 条件化 174
第 8 节 稳定分布 * 177
第 9 节 无穷可分分布 * 188
第 10 节 $\mathbb{R}^d$ 中的极限定理 190

第四章　鞅 **197**
第 1 节 条件期望 197
4.1.1 例子 199
4.1.2 性质 202
4.1.3 正则条件概率 * 205
第 2 节 鞅，几乎必然收敛 209
第 3 节 例子 216
4.3.1 有界增量 216
4.3.2 Polya 罐子模型 217
4.3.3 Radon-Nikodym 导数 218
4.3.4 分支过程 221
第 4 节 Doob 不等式和 $p > 1$ 时的 L^p 收敛 226
第 5 节 平方可积鞅 * 231
第 6 节 一致可积性，L^1 收敛 235
第 7 节 反向鞅 241
第 8 节 可选停时定理 247
4.8.1 在随机游走中的应用 248
第 9 节 简单随机游走的组合学 * 253

第五章　Markov 链 **258**
第 1 节 例子 258
第 2 节 构造与 Markov 性质 262
第 3 节 常返与暂态 271
第 4 节 随机游走的常返性 * 277
第 5 节 平稳测度 289
第 6 节 渐近行为 300
第 7 节 周期性，尾 σ 域 * 307
第 8 节 一般状态空间 * 311
5.8.1 常返与暂态 313
5.8.2 平稳测度 314
5.8.3 收敛定理 315
5.8.4 GI/G/1 排队系统 315

第六章　遍历定理 **318**
第 1 节 定义与例子 318
第 2 节 Birkhoff 遍历定理 322

第 3 节 常返性 327
第 4 节 次可加遍历定理 331
第 5 节 应用 336

第七章 布朗运动 **340**
第 1 节 定义与构造 340
第 2 节 Markov 性质，Blumenthal 0-1 律 347
第 3 节 停时，强 Markov 性质 353
第 4 节 路径性质 357
7.4.1 布朗运动的零点 357
7.4.2 击中时 358
第 5 节 鞅 363
第 6 节 Itô 公式 * 368

第八章 随机游走的应用 **376**
第 1 节 Donsker 定理 376
第 2 节 鞅的中心极限定理 384
第 3 节 平稳序列的中心极限定理 391
8.3.1 混合性质 395
第 4 节 经验分布与布朗桥 399
第 5 节 重对数律 406

第九章 多维布朗运动 **410**
第 1 节 鞅 410
第 2 节 热方程 414
第 3 节 非齐次热方程 416
第 4 节 Feynman-Kac 公式 419
第 5 节 Dirichlet 问题 423
9.5.1 退出分布 427
第 6 节 Green 函数与位势核 430
第 7 节 Poisson 方程 434
9.7.1 占据时间 436
第 8 节 薛定谔方程 440

附录 A 测度论细节 **447**
第 1 节 Carathéodory 扩张定理 447

第 2 节 哪些集合是可测的? 453
第 3 节 符号测度 456
第 4 节 积分号下求导 460

附录 B 正态分布表 **463**

参考文献 **465**

术语表 **475**

第一章
测度论

本章回顾测度论中的一些基本定义和结果，旨在为此前未接触过这些概念的读者提供入门介绍，同时也为已有基础的读者复习相关内容. 那些较难且对直观理解帮助不大的证明均置于附录. 具有扎实测度论背景的读者可跳过 1.4、1.5 和 1.7 节——这些节此前均为附录内容.

第 1 节　概率空间

本书中，被定义的术语均以**粗体**表示. 首先引入最基本的概念.

定义. 概率空间是一个三元组 $(\Omega, \mathcal{F}, P)$，其中 Ω 为**结果**集，$\mathcal{F}$ 为**事件**集，而 $P : \mathcal{F} \to [0,1]$ 是给事件赋予概率的函数. 假设 $\mathcal{F}$ 是一个 **σ-域**（或 **σ-代数**），即 Ω 的一个（非空）**子集**族，满足

(i) 若 $A \in \mathcal{F}$，则 $A^c \in \mathcal{F}$;

(ii) 若 $A_i \in \mathcal{F}$ 为一**可数序列**，则 $\bigcup_i A_i \in \mathcal{F}$.

在本书中，**可数的**意指**有限的**或**可数无限的**. 由于 $\bigcap_i A_i = (\bigcup_i A_i^c)^c$，可知 σ-域对**可数交**封闭. 为使验证更简便，最后这一性质从略.

没有 P 时，$(\Omega, \mathcal{F})$ 称为**可测空间**，即可以在其上定义**测度**的空间. **测度**是一个非负的**可数可加集函数**；也就是说，一个函数 $\mu : \mathcal{F} \to \mathbb{R}$ 满足

(i) 对所有 $A \in \mathcal{F}$，$\mu(A) \geq \mu(\emptyset) = 0$;

(ii) 若 $A_i \in \mathcal{F}$ 是可数的不交集合序列，则 $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

若 $\mu(\Omega) = 1$，则称 μ 为**概率测度**. 本书中，概率测度通常记为 P.

下面给出测度定义的一些**推论**，后文将频繁使用. 以下总假设所提及的集合均属于 $\mathcal{F}$.

定理 1.1.1. 设 μ 是 $(\Omega, \mathcal{F})$ 上的一个测度.

(i) **单调性**. 若 $A \subset B$，则 $\mu(A) \leq \mu(B)$.

(ii) **次可加性**. 若 $A \subset \bigcup_{m=1}^{\infty} A_m$，则 $\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$.
(iii) **下连续性**. 若 $A_i \uparrow A$（即 $A_1 \subset A_2 \subset \cdots$ 且 $\bigcup_i A_i = A$），则 $\mu(A_i) \uparrow \mu(A)$.
(iv) **上连续性**. 若 $A_i \downarrow A$（即 $A_1 \supset A_2 \supset \cdots$ 且 $\bigcap_i A_i = A$），且 $\mu(A_1) < \infty$，则 $\mu(A_i) \downarrow \mu(A)$.

证明. (i) 记 $B - A = B \cap A^c$ 为两集合的**差集**. 用 $+$ 表示不交并，则 $B = A + (B - A)$，故 $\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A)$.

(ii) 记 $A'_n = A \cap A_n$，$B_1 = A'_1$，且对 $n > 1$，$B_n = A'_n - \bigcup_{m=1}^{n-1} A'_m$. 由于 B_n 不交且其并为 A，利用测度定义中的 (ii)、$B_m \subset A_m$ 以及本定理的 (i) 可得

$$\mu(A) = \sum_{m=1}^{\infty} \mu(B_m) \leq \sum_{m=1}^{\infty} \mu(A_m).$$

(iii) 记 $B_n = A_n - A_{n-1}$. 则 B_n 不交，且 $\bigcup_{m=1}^{\infty} B_m = A$，$\bigcup_{m=1}^{n} B_m = A_n$，故

$$\mu(A) = \sum_{m=1}^{\infty} \mu(B_m) = \lim_{n\to\infty} \sum_{m=1}^{n} \mu(B_m) = \lim_{n\to\infty} \mu(A_n).$$

(iv) $A_1 - A_n \uparrow A_1 - A$，故由 (iii) 知 $\mu(A_1 - A_n) \uparrow \mu(A_1 - A)$. 由于 $A_1 \supset A$，有 $\mu(A_1 - A) = \mu(A_1) - \mu(A)$，由此即得 $\mu(A_n) \downarrow \mu(A)$. □

最简单的情形在本科概率论中应该很熟悉：

例题 1.1.2. 离散概率空间. 设 Ω 为一**可数集**，即有限或可数无限. 设 $\mathcal{F} = \Omega$ 的所有子集构成的集合. 令

$$P(A) = \sum_{\omega\in A} p(\omega),$$

其中 $p(\omega) \geq 0$ 且 $\sum_{\omega\in\Omega} p(\omega) = 1$. 稍加思考即可发现这是该空间上最一般的概率测度. 在许多 Ω 为有限集的情形中，取 $p(\omega) = 1/|\Omega|$，其中 $|\Omega|$ 表示 Ω 中的元素个数.

下面给出一个需要这种一般性的简单具体例子：考虑 astragali，即古埃及使用的羊踝骨骰子. 这种骰子停在骨头顶面得 4 点，停在底面得 3 点. 骨头侧面略微弯曲，骰子可停在一个又平又窄的面上得 6 点，或停在侧面其他地方得 1 点. 没有理由认为四种结果等可能，因此需要概率 p_1, p_3, p_4, p_6 来描述 P.

为准备下一个定义，注意由定义易知：若 $\mathcal{F}_i$，$i \in I$ 均为 σ-域，则 $\bigcap_{i\in I} \mathcal{F}_i$ 也是. 这里 $I \neq \emptyset$ 为**任意**指标集（即可能是不可数的）. 由此可见，给定集合 Ω 及其子集族 $\mathcal{A}$，必存在包含 $\mathcal{A}$ 的**最小 σ-域**. 称之为**由 $\mathcal{A}$ 生成的 σ-域**，记为 $\sigma(\mathcal{A})$.

设 $\mathbb{R}^d$ 为实数**向量** $(x_1, \ldots, x_d)$ 的集合，$\mathcal{R}^d$ 为 **Borel 集**，即包含所有**开集**的最小 σ-域. 当 $d = 1$ 时，省略上标.

例题 1.1.3. 实直线上的测度. $(\mathbb{R}, \mathcal{R})$ 上的测度通过给出一个 **Stieltjes 测度函数**来定义，该函数具有以下性质：

(i) F 是**非减的**.

(ii) F 是**右连续的**，即 $\lim_{y \downarrow x} F(y) = F(x)$.

定理 1.1.4. 对每个 Stieltjes 测度函数 F，存在唯一的测度 μ 在 $(\mathbb{R}, \mathcal{R})$ 上满足

$$\mu((a, b]) = F(b) - F(a) \tag{1.1.1}$$

当 $F(x) = x$ 时，所得的测度称为 **Lebesgue 测度**.

定理 1.1.1 的证明颇为曲折，因此本节仅描述主要思想，而将剩余细节置于第 A.1 节. 在 $(a, b]$ 中选择 "右闭" 是由以下事实决定的：若 $b_n \downarrow b$，则

$$\bigcap_n (a, b_n] = (a, b].$$

下一个定义将解释 "左开" 的选择.

定义. 集合族 $\mathcal{S}$ 称为**半代数**，如果

(i) 它对**交**封闭，即 $S, T \in \mathcal{S}$ 蕴涵 $S \cap T \in \mathcal{S}$；

(ii) 若 $S \in \mathcal{S}$，则 S^c 是 $\mathcal{S}$ 中集合的**有限不交并**.

半代数的一个重要例子是：

例题 1.1.5. $\mathcal{S}_d =$ 空集加上所有形如

$$(a_1, b_1] \times \cdots \times (a_d, b_d] \subset \mathbb{R}^d$$

的集合，其中 $-\infty \le a_i < b_i \le \infty$.

定义 (1.1.1) 给出了 μ 在半代数 $\mathcal{S}_1$ 上的值. 要从半代数到 σ-代数，需要一个中间步骤.

定义. Ω 的子集族 $\mathcal{A}$ 称为**代数**（或**域**），如果 $A, B \in \mathcal{A}$ 蕴涵 A^c 和 $A \cup B$ 均属于 $\mathcal{A}$.

由于 $A \cap B = (A^c \cup B^c)^c$，可知 $A \cap B \in \mathcal{A}$. 显然 σ-代数是代数. 其逆不成立的一个例子是：

例题 1.1.6. 设 $\Omega = \mathbb{Z} =$ **整数**. $\mathcal{A} = \{A \subset \mathbb{Z} : A$ 或 A^c 是有限的 $\}$ 是一个代数.

引理 1.1.7. 若 $\mathcal{S}$ 是一个半代数，则 $\bar{\mathcal{S}} = \{\mathcal{S}$ 中集合的有限不交并 $\}$ 是一个代数，称为**由 $\mathcal{S}$ 生成的代数**.

证明. 设 $A=+S_i$ 和 $B=+T_j$，其中 $+$ 表示不交并，且假设指标集是有限的. 则 $A\cap B=+_{i,j}S_i\cap T_j\in\bar{\mathcal{S}}$. 至于补集，若 $A=+S_i$，则 $A^c=\bigcap_i S_i^c$. 由 $\mathcal{S}$ 的定义知 $S_i^c\in\bar{\mathcal{S}}$. 已证 $\bar{\mathcal{S}}$ 对交封闭，故由归纳法可知 $A^c\in\bar{\mathcal{S}}$. □

例题 1.1.8. 设 $\Omega=\mathbb{R}$ 且 $\mathcal{S}=\mathcal{S}_1$，则 $\bar{\mathcal{S}}=$ 空集加上所有形如

$$\bigcup_{i=1}^{k}(a_i,b_i]$$

的集合，其中 $-\infty\le a_i<b_i\le\infty$.

给定 $\mathcal{S}$ 上的集函数 μ，可通过

$$\mu(+_{i=1}^n A_i)=\sum_{i=1}^n\mu(A_i)$$

将其延拓到 $\bar{\mathcal{S}}$ 上. 根据代数 $\mathcal{A}$ 上的**测度**，指的是一个集函数 μ 满足

(i) 对所有 $A\in\mathcal{A}$，$\mu(A)\ge\mu(\emptyset)=0$;

(ii) 若 $A_i\in\mathcal{A}$ 不交且其并也在 $\mathcal{A}$ 中，则 $\mu(\bigcup_{i=1}^\infty A_i)=\sum_{i=1}^\infty\mu(A_i)$.

若存在集合序列 $A_n\in\mathcal{A}$ 使得 $\mu(A_n)<\infty$ 且 $\bigcup_n A_n=\Omega$，则称 μ 是 **σ-有限的**. 令 $A_1'=A_1$ 且对 $n\ge2$，$A_n'=\bigcup_{m=1}^n A_m$ 或 $A_n'=A_n\cap\bigcap_{m=1}^{n-1}A_m^c\in\mathcal{A}$，则可**不失一般性**假设 $A_n\uparrow\Omega$ 或 A_n 不交.

下面的结果帮助我们将定义在半代数 $\mathcal{S}$ 上的测度延拓到它生成的 σ-代数 $\sigma(\mathcal{S})$ 上.

定理 1.1.9. 设 $\mathcal{S}$ 是一个半代数，且 μ 定义在 $\mathcal{S}$ 上满足 $\mu(\emptyset)=0$. 假设

(i) 若 $S\in\mathcal{S}$ 是集合 $S_i\in\mathcal{S}$ 的有限不交并，则 $\mu(S)=\sum_i\mu(S_i)$;

(ii) 若 $S,S_i\in\mathcal{S}$ 满足 $S=+_{i\ge1}S_i$，则 $\mu(S)\le\sum_{i\ge1}\mu(S_i)$.

则 μ 有唯一的延拓 $\bar{\mu}$，它是 $\bar{\mathcal{S}}$(由 $\mathcal{S}$ 生成的代数) 上的测度. 若 $\bar{\mu}$ 是 σ-有限的，则存在唯一的延拓 ν，它是 $\sigma(\mathcal{S})$ 上的测度.

在上面的 (ii) 中以及下文中，$i\ge1$ 表示可数并，而普通的下标 i 或 j 表示有限并. 定理 1.1.9 的证明相当复杂，故置于第 A.1 节. 为检验定理中的条件 (ii)，下面的结果很有用.

引理 1.1.10. 假设仅 (i) 成立.

(a) 若 $A,B_i\in\bar{\mathcal{S}}$ 满足 $A=+_{i=1}^n B_i$，则 $\bar{\mu}(A)=\sum_{i=1}^n\bar{\mu}(B_i)$.

(b) 若 $A,B_i\in\bar{\mathcal{S}}$ 满足 $A\subset\bigcup_{i=1}^n B_i$，则 $\bar{\mu}(A)\le\sum_{i=1}^n\bar{\mu}(B_i)$.

证明. 由定义可知，若 $A=+B_i$ 是 $\bar{\mathcal{S}}$ 中集合的有限不交并，且 $B_i=+_j S_{i,j}$，则

$$\bar{\mu}(A)=\sum_{i,j}\mu(S_{i,j})=\sum_i\bar{\mu}(B_i).$$

为证 (b)，先从 $n=1$ 的情形开始，$B=B_1$. $B=A+(B\cap A^c)$ 且 $B\cap A^c\in\bar{\mathcal{S}}$，故

$$\bar{\mu}(A)\le\bar{\mu}(A)+\bar{\mu}(B\cap A^c)=\bar{\mu}(B).$$

为处理 $n>1$ 的情形，令 $F_k=B_1^c\cap\cdots\cap B_{k-1}^c\cap B_k$，注意

$$\bigcup_i B_i=F_1+\cdots+F_n,\quad A=A\cap\left(\bigcup_i B_i\right)=(A\cap F_1)+\cdots+(A\cap F_n).$$

利用 (a)，$n=1$ 时的 (b)，以及再次使用 (a) 可得

$$\bar{\mu}(A)=\sum_{k=1}^n\bar{\mu}(A\cap F_k)\le\sum_{k=1}^n\bar{\mu}(F_k)=\bar{\mu}\left(\bigcup_i B_i\right).$$

□

定理 1.1.4 的证明. 设 $\mathcal{S}$ 是半开区间 $(a,b](-\infty\le a<b\le\infty)$ 构成的半代数. 为在 $\mathcal{S}$ 上定义 μ，首先注意 $F(\infty)=\lim_{x\uparrow\infty}F(x)$ 和 $F(-\infty)=\lim_{x\downarrow-\infty}F(x)$ 存在，且 $\mu((a,b])=F(b)-F(a)$ 对所有 $-\infty\le a<b\le\infty$ 有意义，因为 $F(\infty)>-\infty$ 且 $F(-\infty)<\infty$.

若 $(a,b]=+_{i=1}^n(a_i,b_i]$，则重排区间后必有 $a_1=a$，$b_n=b$，且 $a_i=b_{i-1}$ 对 $2\le i\le n$ 成立，故定理 1.1.9 中的条件 (i) 成立. 为检验 (ii)，先设 $-\infty<a<b<\infty$，且 $(a,b]\subset\bigcup_{i\ge1}(a_i,b_i]$（不失一般性，$-\infty<a_i<b_i<\infty$）. 选取 $\delta>0$ 使得 $F(a+\delta)<F(a)+\varepsilon$，并选取 η_i 使得 $F(b_i+\eta_i)<F(b_i)+\varepsilon2^{-i}$.

开区间 $(a_i,b_i+\eta_i)$ 覆盖 $[a+\delta,b]$，故存在有限子覆盖 (α_j,β_j)，$1\le j\le J$. 由于 $(a+\delta,b]\subset\bigcup_{j=1}^J(\alpha_j,\beta_j]$，引理 1.1.10 中的 (b) 蕴涵

$$F(b)-F(a+\delta)\le\sum_{j=1}^J(F(\beta_j)-F(\alpha_j))\le\sum_{i=1}^\infty(F(b_i+\eta_i)-F(a_i)).$$

由 δ 和 η_i 的选择，

$$F(b)-F(a)\le2\varepsilon+\sum_{i=1}^\infty(F(b_i)-F(a_i)),$$

且由于 ε 是任意的，已证 $-\infty<a<b<\infty$ 时的结果. 为去掉最后一个限制，注意若 $(a,b]\subset\bigcup_i(a_i,b_i]$ 且 $(A,B]\subset(a,b]$ 满足 $-\infty<A<B<\infty$，则

$$F(B)-F(A)\le\sum_{i=1}^\infty(F(b_i)-F(a_i)).$$

由于最后一个结果对任何有限区间 $(A,B]\subset(a,b]$ 成立，故得所需结果. □

0	2/3	1
0	0	2/3
0	0	0

图 1.1: 反例

$\mathbb{R}^d$ 上的测度

下一个目标是证明定理 1.1.4 的一个 $\mathbb{R}^d$ 版本. 第一步是引入关于定义函数 F 的假设. 通过与 $d=1$ 情形的类比，自然假设:

(i) 它是非减的，即若 $x \le y$（意指对所有 i 有 $x_i \le y_i$），则 $F(x) \le F(y)$.

(ii) F 是右连续的，即 $\lim_{y \downarrow x} F(y) = F(x)$（这里 $y \downarrow x$ 意指每个 $y_i \downarrow x_i$）.

(iii) 若 $x_n \downarrow -\infty$（即每个坐标都趋于 $-\infty$），则 $F(x_n) \downarrow 0$. 若 $x_n \uparrow \infty$（即每个坐标都趋于 ∞），则 $F(x_n) \uparrow 1$.

然而这还不够. 考虑下面的 F:

$$F(x_1,x_2) = \begin{cases} 1 & \text{若} x_1, x_2 \ge 1, \\ 2/3 & \text{若} x_1 \ge 1 \text{ 且} 0 \le x_2 < 1, \\ 2/3 & \text{若} x_2 \ge 1 \text{ 且} 0 \le x_1 < 1, \\ 0 & \text{其他情形.} \end{cases}$$

参见图 1.1 中的示意图. 稍加思考即可看出

$$\begin{aligned} \mu((a_1,b_1]\times(a_2,b_2]) &= \mu((-\infty,b_1]\times(-\infty,b_2]) - \mu((-\infty,a_1]\times(-\infty,b_2]) \\ &\quad - \mu((-\infty,b_1]\times(-\infty,a_2]) + \mu((-\infty,a_1]\times(-\infty,a_2]) \\ &= F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2). \end{aligned}$$

用此式取 $a_1 = a_2 = 1-\varepsilon$ 和 $b_1 = b_2 = 1$，令 $\varepsilon \to 0$，可见

$$\mu(\{1,1\}) = 1 - 2/3 - 2/3 + 0 = -1/3.$$

类似的推理表明 $\mu(\{1,0\}) = \mu(\{0,1\}) = 2/3$.

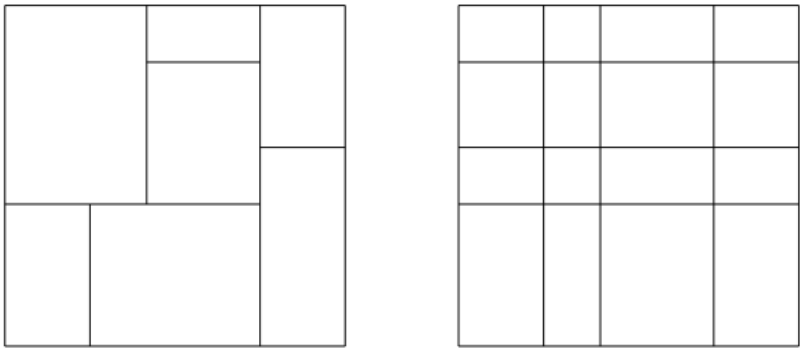

图 1.2: 将一个细分转化为正则细分

为表述 F 定义测度的第三个也是最后一个条件，令

$$A = (a_1, b_1] \times \cdots \times (a_d, b_d], \quad V = \{a_1, b_1\} \times \cdots \times \{a_d, b_d\},$$

其中 $-\infty < a_i < b_i < \infty$. 为强调不允许 ∞，称 A 为**有限矩形**. 则 $V =$ 矩形 A 的**顶点**. 若 $v \in V$，令

$$\operatorname{sgn}(v) = (-1)^{v \text{ 中 } a \text{ 的个数}}, \quad \Delta_A F = \sum_{v \in V} \operatorname{sgn}(v) F(v).$$

令 $\mu(A) = \Delta_A F$，故必须假设

(iv) 对所有矩形 A，$\Delta_A F \geq 0$.

定理 1.1.11. 设 $F : \mathbb{R}^d \to [0,1]$ 满足上述条件 (i)–(iv). 则存在唯一的概率测度 μ 在 $(\mathbb{R}^d, \mathcal{R}^d)$ 上使得 $\mu(A) = \Delta_A F$ 对所有有限矩形 A 成立.

例题 1.1.12. 设 $F(x) = \prod_{i=1}^d F_i(x_i)$，其中 F_i 满足定理 1.1.4 的条件 (i) 和 (ii). 在这种情形下，

$$\Delta_A F = \prod_{i=1}^d (F_i(b_i) - F_i(a_i)).$$

当对所有 i 有 $F_i(x) = x$ 时，所得的测度是 $\mathbb{R}^d$ 上的 **Lebesgue 测度**.

证明. 对所有有限矩形令 $\mu(A) = \Delta_A F$，然后利用单调性将定义延拓到 $\mathcal{S}_d$. 为检验定理 1.1.9 的条件 (i)，称 $A = +B_k$ 为 A 的一个**正则细分**，如果存在序列 $a_i = \alpha_{i,0} < \alpha_{i,1} < \cdots < \alpha_{i,n_i} = b_i$ 使得每个矩形 B_k 具有形式 $(\alpha_{1,j_1-1}, \alpha_{1,j_1}] \times \cdots \times (\alpha_{d,j_d-1}, \alpha_{d,j_d}]$，其中 $1 \leq j_i \leq n_i$.

容易看到对正则细分有 $\mu(A) = \sum_k \mu(B_k)$[1]. （先考虑所有端点都有限的情形，然后取极限得到一般情形. ）为将此结果推广到一般的有限细分 $A = +A_j$，进一步细分得到一个正则细分.

[1]**译注.** 原稿在推导正则细分时误用符号 λ，现统一为定理主体的测度符号 μ，以确保逻辑自洽.

(ii) 的证明与定理 1.1.4 中的几乎完全相同. 为使书写更容易并突出与定理 1.1.4 的类比，对 $x, y \in \mathbb{R}^d$ 令

$$(x, y) = (x_1, y_1) \times \cdots \times (x_d, y_d), \quad (x, y] = (x_1, y_1] \times \cdots \times (x_d, y_d], \quad [x, y] = [x_1, y_1] \times \cdots \times [x_d, y_d].$$

首先设 $-\infty < a < b < \infty$，其中不等式意指每个分量都有限，且设 $(a, b] \subset \bigcup_{i \geq 1} (a_i, b_i]$（**不失一般性**，$-\infty < a_i < b_i < \infty$）. 令 $\bar{1} = (1, \ldots, 1)$，选取 $\delta > 0$ 使得 $\mu((a + \delta\bar{1}, b]) < \mu((a, b]) + \varepsilon$，并选取 η_i 使得 $\mu((a_i, b_i + \eta_i\bar{1}]) < \mu((a_i, b_i]) + \varepsilon 2^{-i}$.

开矩形 $(a_i, b_i + \eta_i\bar{1})$ 覆盖 $[a + \delta\bar{1}, b]$，故存在有限子覆盖 (α_j, β_j)，$1 \leq j \leq J$. 由于 $(a + \delta\bar{1}, b] \subset \bigcup_{j=1}^{J} (\alpha_j, \beta_j]$，引理 1.1.10 中的 (b) 蕴涵

$$\mu([a + \delta\bar{1}, b]) \leq \sum_{j=1}^{J} \mu((\alpha_j, \beta_j]) \leq \sum_{i=1}^{\infty} \mu((a_i, b_i + \eta_i\bar{1}]).$$

由 δ 和 η_i 的选择，

$$\mu((a, b]) \leq 2\varepsilon + \sum_{i=1}^{\infty} \mu((a_i, b_i]),$$

且由于 ε 是任意的，已证 $-\infty < a < b < \infty$ 时的结果. 证明现在可以像以前一样完成. □

习题

习题 1.1.1. 设 $\Omega = \mathbb{R}$，$\mathcal{F}$ = 所有满足 A 或 A^c 可数的子集，$P(A) = 0$ 对第一种情形，$= 1$ 对第二种. 证明 $(\Omega, \mathcal{F}, P)$ 是一个概率空间.

习题 1.1.2. 回顾例 1.1.5 中 $\mathcal{S}_d$ 的定义. 证明 $\sigma(\mathcal{S}_d) = \mathcal{R}^d$，即 $\mathbb{R}^d$ 的 Borel 子集.

习题 1.1.3. σ-域 $\mathcal{F}$ 称为是**可数生成的**，如果存在可数族 $\mathcal{C} \subset \mathcal{F}$ 使得 $\sigma(\mathcal{C}) = \mathcal{F}$. 证明 $\mathcal{R}^d$ 是可数生成的.

习题 1.1.4. (i) 证明若 $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ 是 σ-代数，则 $\bigcup_i \mathcal{F}_i$ 是一个代数.
(ii) 给出一个例子说明 $\bigcup_i \mathcal{F}_i$ 不必是 σ-代数.

习题 1.1.5. 集合 $A \subset \{1, 2, \ldots\}$ 称为有**渐近密度** θ，如果 $\lim_{n \to \infty} |A \cap \{1, 2, \ldots, n\}|/n = \theta$. 设 $\mathcal{A}$ 是渐近密度存在的集合族. $\mathcal{A}$ 是一个 σ-代数吗？是一个代数吗？

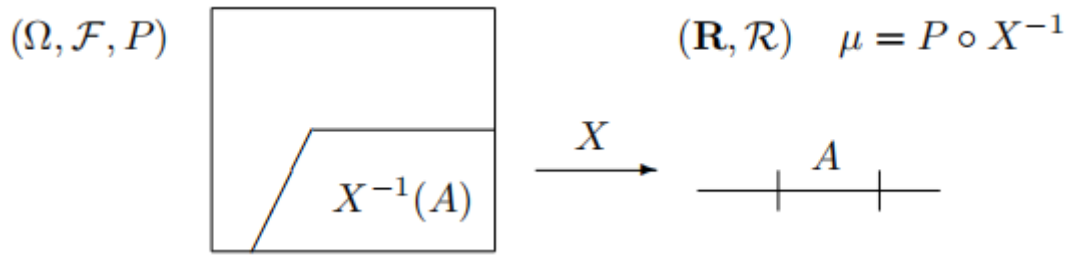

图 1.3: X 的分布的定义

第 2 节 分布

在概率空间上定义**随机变量**后，概率论变得饶有趣味. 定义在 Ω 上的实值函数 X 称为随机变量，如果对每个 Borel 集 $B \subset \mathbb{R}$，都有 $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$. 当需要强调 σ-域时，称 X 是 **$\mathcal{F}$-可测的**，或记作 $X \in \mathcal{F}$. 若 Ω 是**离散概率空间**（见例 1.1.2），则任意函数 $X : \Omega \to \mathbb{R}$ 都是随机变量. 另一类平凡但有用的随机变量是集合 $A \in \mathcal{F}$ 的**指示函数**：

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

该记号意在强调此函数在 A 上取值为 1. 分析学家称此对象为 A 的**特征函数**. 在概率论中，该术语用于完全不同的含义（见 3.3 节）.

若 X 是随机变量，则 X 在 $\mathbb{R}$ 上诱导出一个概率测度，称为它的**分布**，记为 $\mu(A) = P(X \in A)$，其中 A 为 Borel 集. 利用上面引入的记号，右端可写为 $P(X^{-1}(A))$. 换言之，将 $A \in \mathcal{R}$ 拉回到 $X^{-1}(A) \in \mathcal{F}$，再取该集合的 P 值.

为验证 μ 是概率测度，设 A_i 不交，则

$$\mu\left(\bigcup_i A_i\right) = P\left(X \in \bigcup_i A_i\right) = P\left(\bigcup_i \{X \in A_i\}\right) = \sum_i P(X \in A_i) = \sum_i \mu(A_i).$$

这里依次用到 μ 的定义、X 落入并集当且仅当它落入某个 A_i、若集合 $A_i \in \mathcal{R}$ 不交则事件 $\{X \in A_i\}$ 不交，以及 μ 的定义.

随机变量 X 的分布通常通过其**分布函数** $F(x) = P(X \le x)$ 来描述.

定理 1.2.1. 任意分布函数 F 具有以下性质：

(i) F 是**非减的**.

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

(iii) F 是**右连续的**，即 $\lim_{y\downarrow x} F(y) = F(x)$.

(iv) 若 $F(x-) = \lim_{y\uparrow x} F(y)$，则 $F(x-) = P(X < x)$.

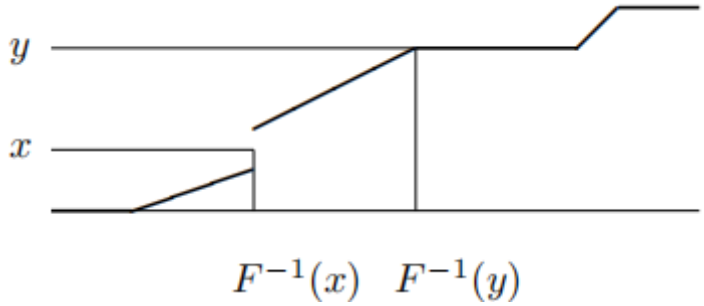

图 1.4: 定理 1.2.2 的证明中定义的逆

(v) $P(X = x) = F(x) - F(x-)$.

证明. (i) 若 $x \le y$，则 $\{X \le x\} \subset \{X \le y\}$，由定理 1.1.1 中的 (i) 得 $P(X \le x) \le P(X \le y)$.
(ii) 若 $x \uparrow \infty$，则 $\{X \le x\} \uparrow \Omega$；若 $x \downarrow -\infty$，则 $\{X \le x\} \downarrow \emptyset$. 再利用定理 1.1.1 的 (iii) 和 (iv).
(iii) 若 $y \downarrow x$，则 $\{X \le y\} \downarrow \{X \le x\}$.
(iv) 若 $y \uparrow x$，则 $\{X \le y\} \uparrow \{X < x\}$.
(v) 由 $P(X = x) = P(X \le x) - P(X < x)$ 及 (iii)、(iv) 即得. □

上述性质足以刻画分布函数.

定理 1.2.2. 若 F 满足定理 1.2.1 中的 (i)、(ii) 和 (iii)，则它是某个随机变量的分布函数.

证明. 取 $\Omega = (0,1), \mathcal{F}$ =Borel 集, P =Lebesgue 测度. 对 $\omega \in (0,1)$, 定义 $X(\omega) = \sup\{y : F(y) < \omega\}$. 下证

$$\{\omega : X(\omega) \le x\} = \{\omega : \omega \le F(x)\} \tag{?}$$

由此即得结论，因为 $P(\omega : \omega \le F(x)) = F(x)$（回忆 P 是 Lebesgue 测度）. 验证如下：若 $\omega \le F(x)$，则 $x \notin \{y : F(y) < \omega\}$，故 $X(\omega) \le x$. 反之，若 $\omega > F(x)$，由 F 的右连续性，存在 $\varepsilon > 0$ 使得 $F(x+\varepsilon) < \omega$，于是 $X(\omega) \ge x + \varepsilon > x$. □

尽管 F 未必是一一对应且满射的，仍称 X 为 F 的逆，记为 F^{-1}. 定理 1.2.2 证明中的方法在计算机上生成随机变量时很有用. 标准算法生成服从**均匀分布**的随机变量 U，然后应用定理 1.2.2 中定义的分布函数的逆，得到具有分布函数 F 的随机变量 $F^{-1}(U)$.

若 X 和 Y 在 $(\mathbb{R}, \mathcal{R})$ 上诱导出相同的分布 μ，则称 X 和 Y 是**同分布的**. 由定理 1.1.4，这等价于 X 和 Y 具有相同的分布函数，即对所有 x 有 $P(X \le x) = P(Y \le x)$. 当 X 和 Y 同分布时，记作 $X \overset{d}{=} Y$，但此记号在文本中太高，故出于排版原因也使用 $X =_d Y$.

若分布函数 $F(x) = P(X \le x)$ 具有形式

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y \tag{1.2.1}$$

则称 X 具有**密度函数** f. 记忆公式时，常将 $f(x)$ 设想为 $P(X=x)$，尽管

$$P(X=x)=\lim_{\varepsilon\to 0}\int_{x-\varepsilon}^{x+\varepsilon} f(y)\,\mathrm{d}y=0.$$

为避免混淆，本书采用 $f_X(x)$ 这类记号表示密度函数.

也可从 f 出发，利用 (1.2.1) 定义分布函数 F. 为使 F 成为分布函数，必要且充分条件是 $f(x)\geq 0$ 且 $\int f(x)\,\mathrm{d}x=1$. 下面三个例子在后续内容中很重要：

例题 1.2.3. $(0,1)$ **上的均匀分布**. $f(x)=1$ 对 $x\in(0,1)$，否则为 0. 分布函数：

$$F(x)=\begin{cases}0 & x\leq 0,\\ x & 0\leq x\leq 1,\\ 1 & x>1.\end{cases}$$

例题 1.2.4. 速率为 λ 的指数分布. $f(x)=\lambda e^{-\lambda x}$ 对 $x\geq 0$，否则为 0. 分布函数：

$$F(x)=\begin{cases}0 & x\leq 0,\\ 1-e^{-\lambda x} & x\geq 0.\end{cases}$$

例题 1.2.5. 标准正态分布.

$$f(x)=(2\pi)^{-1/2}\exp\left(-\frac{x^2}{2}\right).$$

此时，$F(x)$ 没有闭式表达，但对大 x 有以下有用的界：

定理 1.2.6. 对 $x>0$,

$$(x^{-1}-x^{-3})\exp\left(-\frac{x^2}{2}\right)\leq\int_x^\infty \exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y\leq x^{-1}\exp\left(-\frac{x^2}{2}\right).$$

证明. 作变量替换 $y=x+z$ 并利用 $\exp(-z^2/2)\leq 1$ 得

$$\int_x^\infty \exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y\leq\exp\left(-\frac{x^2}{2}\right)\int_0^\infty\exp(-xz)\,\mathrm{d}z=x^{-1}\exp\left(-\frac{x^2}{2}\right).$$

对另一方向，注意

$$\int_x^\infty(1-3y^{-4})\exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y=(x^{-1}-x^{-3})\exp\left(-\frac{x^2}{2}\right).$$

□

图 1.5: Cantor 分布函数

$\mathbb{R}$ 上的分布函数若具有密度则称为**绝对连续的**，若相应的测度关于 Lebesgue 测度**奇异**则称为奇异的. 关于这些概念的更多内容见 A.4 节. 奇异分布的一个例子是：

例题 1.2.7. Cantor 集上的均匀分布. Cantor 集 C 的定义是从 $[0,1]$ 中去掉 $(1/3,2/3)$，然后去掉剩余各区间的中三分之一. 定义一个相关的分布函数如下：$x \le 0$ 时 $F(x)=0$，$x \ge 1$ 时 $F(x)=1$，$x \in [1/3,2/3]$ 时 $F(x)=1/2$，$x \in [1/9,2/9]$ 时 $F(x)=1/4$，$x \in [7/9,8/9]$ 时 $F(x)=3/4$，... 然后利用单调性将 F 延拓到整个 $[0,1]$. 不存在使得 (1.2.1) 成立的 f，因为这样的 f 将在测度为 1 的集合上等于 0. 由定义立即可知相应的测度满足 $\mu(C^c)=0$.

概率测度 P（或其相关的分布函数）称为**离散的**，若存在**可数**集 S 使得 $P(S^c)=0$. 离散分布的最简单例子是

例题 1.2.8. 0 **处的点质量**. $x \ge 0$ 时 $F(x)=1$，$x<0$ 时 $F(x)=0$.

1.6 节将介绍 Bernoulli 分布、Poisson 分布和几何分布. 下面的例子表明，与离散概率测度相关的分布函数可能相当复杂.

例题 1.2.9. 稠密不连续点. 设 $q_1,q_2,\ldots$ 是有理数的一个枚举. 设 $\alpha_i>0$ 满足 $\sum_{i=1}^{\infty}\alpha_i=1$ 且令

$$F(x)=\sum_{i=1}^{\infty}\alpha_i \mathbf{1}_{[q_i,\infty)}(x)$$

其中 $\mathbf{1}_{[\theta,\infty)}(x)=1$ 若 $x\in[\theta,\infty)$，$=0$ 否则.

习题

习题 1.2.1. 设 X 和 Y 是 $(\Omega,\mathcal{F},P)$ 上的随机变量，且 $A\in\mathcal{F}$. 证明若定义 $Z(\omega)=X(\omega)$ 对 $\omega\in A$，$Z(\omega)=Y(\omega)$ 对 $\omega\in A^c$，则 Z 是随机变量.

习题 1.2.2. 设 χ 服从标准正态分布. 利用定理 1.2.6 给出 $P(\chi \geq 4)$ 的上界和下界.

习题 1.2.3. 证明分布函数至多有可数多个不连续点.

习题 1.2.4. 证明若 $F(x) = P(X \leq x)$ 连续，则 $Y = F(X)$ 服从 $(0,1)$ 上的均匀分布，即若 $y \in [0,1]$，$P(Y \leq y) = y$.

习题 1.2.5. 设 X 具有连续密度 f，$P(\alpha \leq X \leq \beta) = 1$ 且 g 是在 (α, β) 上严格递增且可微的函数. 则 $g(X)$ 具有密度 $f(g^{-1}(y))/g'(g^{-1}(y))$ 对 $y \in (g(\alpha), g(\beta))$，否则为 0. 当 $g(x) = ax + b$ 且 $a > 0$ 时，$g^{-1}(y) = (y - b)/a$，故答案为 $(1/a)f((y - b)/a)$.

习题 1.2.6. 设 X 服从正态分布. 利用上题计算 $\exp(X)$ 的密度. (该答案称为**对数正态分布**.)

习题 1.2.7. (i) 设 X 具有密度函数 f. 计算 X^2 的分布函数，然后求导得到其密度函数.
(ii) 当 X 服从标准正态分布时算出答案，从而得到**卡方分布**的密度.

第 3 节 随机变量

本节将建立若干结果，以便后续证明所定义的对象确为随机变量（即可测）时使用. 本节的大部分结论对任意可测空间 $(S,\mathcal{S})$ 上的**随机元**均成立，且证明方法相同（有时更简单），故在此一般性框架下建立这些结果. 首先给出一个定义. 设函数 $X:\Omega\to S$ 满足对所有 $B\in\mathcal{S}$ 都有 $X^{-1}(B)\equiv\{\omega:X(\omega)\in B\}\in\mathcal{F}$，则称 X 为从 $(\Omega,\mathcal{F})$ 到 $(S,\mathcal{S})$ 的**可测映射**. 若 $(S,\mathcal{S})=(\mathbb{R}^d,\mathcal{R}^d)$ 且 $d>1$，则称 X 为**随机向量**. 当然，若 $d=1$，则称 X 为**随机变量**，或简记为 r.v.

下面的结果对证明映射可测很有用.

定理 1.3.1. 若 $\{\omega:X(\omega)\in A\}\in\mathcal{F}$ 对所有 $A\in\mathcal{A}$ 成立，且 $\mathcal{A}$ **生成** $\mathcal{S}$（即 $\mathcal{S}$ 是包含 $\mathcal{A}$ 的最小 σ-域），则 X 可测.

证明. 用 $\{X\in B\}$ 作为 $\{\omega:X(\omega)\in B\}$ 的简写，我们有 $\{X\in\bigcup_i B_i\}=\bigcup_i\{X\in B_i\}$，$\{X\in B^c\}=\{X\in B\}^c$. 因此集合类 $\mathcal{B}=\{B:\{X\in B\}\in\mathcal{F}\}$ 是一个 σ-域. 由于 $\mathcal{B}\supset\mathcal{A}$ 且 $\mathcal{A}$ 生成 $\mathcal{S}$，故 $\mathcal{B}\supset\mathcal{S}$. □

由前一证明中展示的两个等式可知，若 $\mathcal{S}$ 是 σ-域，则 $\{\{X\in B\}:B\in\mathcal{S}\}$ 是 σ-域. 它是使得 X 成为可测映射的 Ω 上最小 σ-域. 称之为由 X **生成的** σ**-域**，记为 $\sigma(X)$. 为后文参考，注意

$$\sigma(X)=\{\{X\in B\}:B\in\mathcal{S}\}\tag{1.3.1}$$

例题 1.3.2. 若 $(S,\mathcal{S})=(\mathbb{R},\mathcal{R})$，则定理 1.3.1 中 $\mathcal{A}$ 可取 $\{(-\infty,x]:x\in\mathbb{R}\}$ 或 $\{(-\infty,x):x\in\mathbb{Q}\}$，其中 $\mathbb{Q}$ 为有理数集.

例题 1.3.3. 若 $(S,\mathcal{S})=(\mathbb{R}^d,\mathcal{R}^d)$，$\mathcal{A}$ 的一个有用取法是 $\{(a_1,b_1)\times\cdots\times(a_d,b_d):-\infty<a_i<b_i<\infty\}$，或者偶尔使用更大的**开集**族.

定理 1.3.4. 若 $X:(\Omega,\mathcal{F})\to(S,\mathcal{S})$ 和 $f:(S,\mathcal{S})\to(T,\mathcal{T})$ 均为可测映射，则 $f(X)$ 是从 $(\Omega,\mathcal{F})$ 到 $(T,\mathcal{T})$ 的可测映射.

证明. 设 $B\in\mathcal{T}$. 则 $\{\omega:f(X(\omega))\in B\}=\{\omega:X(\omega)\in f^{-1}(B)\}\in\mathcal{F}$，这是因为由假设 $f^{-1}(B)\in\mathcal{S}$. □

由定理 1.3.4 立即可知，若 X 是随机变量，则对所有 $c\in\mathbb{R}$，cX 也是随机变量，X^2、$\sin(X)$ 等亦然. 下面的结果说明了为何要对可测映射证明定理 1.3.4.

定理 1.3.5. 若 $X_1,\ldots,X_n$ 是随机变量且 $f:(\mathbb{R}^n,\mathcal{R}^n)\to(\mathbb{R},\mathcal{R})$ 可测，则 $f(X_1,\ldots,X_n)$ 是随机变量.

证明. 由定理 1.3.4，只需证明 $(X_1,\ldots,X_n)$ 是随机向量. 为此，注意若 $A_1,\ldots,A_n$ 为 Borel 集，则 $\{(X_1,\ldots,X_n)\in A_1\times\cdots\times A_n\}=\bigcap_i\{X_i\in A_i\}\in\mathcal{F}$. 由于形如 $A_1\times\cdots\times A_n$ 的集合生成 $\mathcal{R}^n$，故由定理 1.3.1 即得所需结果. □

定理 1.3.6. 若 $X_1,\dots,X_n$ 是随机变量，则 $X_1+\cdots+X_n$ 是随机变量.

证明. 由定理 1.3.5，只需证明 $f(x_1,\dots,x_n)=x_1+\cdots+x_n$ 可测. 为此，利用例 1.3.2 并注意 $\{x: x_1+\cdots+x_n<a\}$ 是开集，因而属于 $\mathcal{R}^n$. □

定理 1.3.7. 若 $X_1,X_2,\dots$ 是随机变量，则 $\inf_n X_n$、$\sup_n X_n$、$\limsup_{n\to\infty} X_n$、$\liminf_{n\to\infty} X_n$ 均为随机变量.

证明. 由于序列的**下确界** $<a$ 当且仅当某项 $<a$（若所有项 $\geq a$ 则下确界亦然），我们有 $\{\inf_n X_n<a\}=\bigcup_n\{X_n<a\}\in\mathcal{F}$. 类似地，$\{\sup_n X_n>a\}=\bigcup_n\{X_n>a\}\in\mathcal{F}$. 对于后两个，注意

$$\liminf_{n\to\infty} X_n=\sup_n\inf_{m\geq n} X_m,\quad \limsup_{n\to\infty} X_n=\inf_n\sup_{m\geq n} X_m.$$

为完成第一个情形的证明，注意对每个 n，$Y_n=\inf_{m\geq n}X_m$ 是随机变量，故 $\sup_n Y_n$ 亦然. □

由定理 1.3.7 可知，

$$\Omega_o\equiv\{\omega:\lim_{n\to\infty} X_n\text{ 存在}\}=\{\omega:\limsup_{n\to\infty} X_n=\liminf_{n\to\infty} X_n\text{ 且均在}\mathbb{R}\text{ 中}\}^2$$

是可测集.（这里 $\equiv$ 表示第一个等式是定义.）若 $P(\Omega_o)=1$，称 X_n **几乎必然**收敛，或简记为 a.s. 收敛. 这种收敛在测度论中称为**几乎处处**收敛. 为在整个空间上定义极限，方便的做法是令 $X_\infty=\limsup_{n\to\infty}X_n$ 在整个 Ω 上，但这个随机变量可能取值 $+\infty$ 或 $-\infty$.

为处理这一情况以及一些其他问题，我们将推广随机变量的定义. 定义域为集合 $D\in\mathcal{F}$ 且**值域**为 $\mathbb{R}^*\equiv[-\infty,\infty]$ 的函数 X，若对所有 $B\in\mathcal{R}^*$ 都有 $X^{-1}(B)=\{\omega:X(\omega)\in B\}\in\mathcal{F}$，则称为随机变量. 这里 $\mathcal{R}^*$ 是 $\mathbb{R}^*$ 的 Borel 子集，其中 $\mathbb{R}^*$ 配备**通常拓扑**，即由形如 $[-\infty,a)$、(a,b) 和 $(b,\infty]$ 的区间生成的拓扑，其中 $a,b\in\mathbb{R}$. 读者应注意 $(\mathbb{R}^*,\mathcal{R}^*)$ 是一个**可测空间**，因此上述所有结果立即可以推广.

习题

习题 1.3.1. 证明若 $\mathcal{A}$ 生成 $\mathcal{S}$，则 $X^{-1}(\mathcal{A})\equiv\{\{X\in A\}:A\in\mathcal{A}\}$ 生成 $\sigma(X)=\{\{X\in B\}:B\in\mathcal{S}\}$.

习题 1.3.2. 当 $n=2$ 时，通过验证 $\{X_1+X_2<x\}\in\mathcal{F}$ 来证明定理 1.3.6.

习题 1.3.3. 证明若 f **连续**且 $X_n\to X$ **几乎必然**，则 $f(X_n)\to f(X)$ 几乎必然.

习题 1.3.4. (i) 证明从 $\mathbb{R}^d\to\mathbb{R}$ 的**连续**函数是从 $(\mathbb{R}^d,\mathcal{R}^d)$ 到 $(\mathbb{R},\mathcal{R})$ 的可测映射.

(ii) 证明 $\mathcal{R}^d$ 是使得所有连续函数可测的最小 σ-域.

[2]**译注.** 原稿采用差值定义收敛，在广义实数 $\mathbb{R}^*$ 框架下存在 $\infty-\infty$ 未定义的逻辑风险，现修正为判别上下极限相等以确保测度判别式的普适性.

习题 1.3.5. 函数 f 称为**下半连续**（简记为 l.s.c.），如果 $\liminf_{y\to x} f(y) \geq f(x)$；称为**上半连续**（简记为 u.s.c.），如果 $-f$ 是下半连续的. 证明 f 下半连续当且仅当 $\{x : f(x) \leq a\}$ 对每个 $a \in \mathbb{R}$ 是闭集，并由此得出半连续函数可测.

习题 1.3.6. 设 $f : \mathbb{R}^d \to \mathbb{R}$ 为任意函数，令 $f^\delta(x) = \sup\{f(y) : |y-x| < \delta\}$，$f_\delta(x) = \inf\{f(y) : |y-x| < \delta\}$，其中 $|z| = (z_1^2 + \cdots + z_d^2)^{1/2}$. 证明 f^δ 是下半连续的，f_δ 是上半连续的. 令 $f^0 = \lim_{\delta\downarrow 0} f^\delta$，$f_0 = \lim_{\delta\downarrow 0} f_\delta$，并由此得出 f 的不连续点集 $= \{f^0 \neq f_0\}$ 是可测的.

习题 1.3.7. 函数 $\varphi : \Omega \to \mathbb{R}$ 称为**简单函数**，如果 $\varphi(\omega) = \sum_{m=1}^n c_m \mathbf{1}_{A_m}(\omega)$，其中 c_m 为实数且 $A_m \in \mathcal{F}$. 证明 $\mathcal{F}$ 可测函数类是包含简单函数类且对逐点极限封闭的最小类.

习题 1.3.8. 利用上一习题得出：Y 关于 $\sigma(X)$ 可测当且仅当 $Y = f(X)$，其中 $f : \mathbb{R} \to \mathbb{R}$ 可测.

习题 1.3.9. 为得到上一结果的一个构造性证明，注意 $\{\omega : m2^{-n} \leq Y < (m+1)2^{-n}\} = \{X \in B_{m,n}\}$ 对某个 $B_{m,n} \in \mathcal{R}$ 成立，令 $f_n(x) = m2^{-n}$ 对 $x \in B_{m,n}$，并证明当 $n \to \infty$ 时 $f_n(x) \to f(x)$ 且 $Y = f(X)$.

第 4 节 积分

设 μ 为 $(\Omega, \mathcal{F})$ 上的 σ-**有限**测度. 我们主要关注 μ 为概率测度的特殊情形，但有时也需要关于无穷测度进行积分，而在一般情形下建立这些结果并不会增加难度.

本节将为一类可测函数定义 $\int f \,\mathrm{d}\mu$. 这是一个四步构造:

(1) 简单函数;
(2) 有界函数;
(3) 非负函数;
(4) 一般函数.

这四步序列在证明积分公式时也很有用，参见定理 1.6.9 和 1.7.2 的证明.

第 1 步. 若 $\varphi(\omega) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ 且 A_i 为不交集合并满足 $\mu(A_i) < \infty$，则称 φ 为**简单函数**. 若 φ 是简单函数，定义

$$\int \varphi \,\mathrm{d}\mu = \sum_{i=1}^n a_i \mu(A_i).$$

φ 的表示不唯一，因为我们并未假设 a_i 互不相同. 然而，容易验证上述定义不会自相矛盾.

我们将对以下三个结论证明四次，但在第一次陈述之前，先给出一个定义. 称 $\varphi \geq \psi$ μ-**几乎处处**成立（或记作 $\varphi \geq \psi$ μ-a.e.），是指 $\mu(\{\omega : \varphi(\omega) < \psi(\omega)\}) = 0$. 当所指的测度明确无疑时，我们省略 μ.

引理 1.4.1. 设 φ 和 ψ 为简单函数.

(i) 若 $\varphi \geq 0$ a.e.，则 $\int \varphi \,\mathrm{d}\mu \geq 0$.
(ii) 对任意 $a \in \mathbb{R}$，$\int a\varphi \,\mathrm{d}\mu = a \int \varphi \,\mathrm{d}\mu$.
(iii) $\int \varphi + \psi \,\mathrm{d}\mu = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\mu$.

证明. (i) 和 (ii) 由定义立即得到. 为证 (iii)，设 $\varphi = \sum_{i=1}^m a_i \mathbf{1}_{A_i}$，$\psi = \sum_{j=1}^n b_j \mathbf{1}_{B_j}$. 为使两个函数的支集相同，令 $A_0 = \bigcup_i B_i - \bigcup_i A_i$，$B_0 = \bigcup_i A_i - \bigcup_i B_i$，且 $a_0 = b_0 = 0$. 于是

$$\varphi + \psi = \sum_{i=0}^m \sum_{j=0}^n (a_i + b_j) \mathbf{1}_{(A_i \cap B_j)},$$

且 $A_i \cap B_j$ 两两不交，故

$$\int (\varphi + \psi) \,\mathrm{d}\mu = \sum_{i=0}^m \sum_{j=0}^n (a_i + b_j) \mu(A_i \cap B_j) = \cdots = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\mu.$$

在倒数第二步，我们用到了 $A_i = \bigsqcup_j (A_i \cap B_j)$ 和 $B_j = \bigsqcup_i (A_i \cap B_j)$，其中 $\sqcup$ 表示不交并. □

随着积分的推广，我们还将三次证明 (i)–(iii). 由 (i)–(iii) 可得到另外三个有用的性质. 为避免重复其不变的证明，我们证明

引理 1.4.2. 若 (i) 和 (iii) 成立，则有：

(iv) 若 $\varphi \le \psi$ a.e.，则 $\int \varphi \,\mathrm{d}\mu \le \int \psi \,\mathrm{d}\mu$.

(v) 若 $\varphi = \psi$ a.e.，则 $\int \varphi \,\mathrm{d}\mu = \int \psi \,\mathrm{d}\mu$.

(vi) $\left|\int \varphi \,\mathrm{d}\mu\right| \le \int |\varphi| \,\mathrm{d}\mu$.

证明. 由 (iii)，$\int \psi \,\mathrm{d}\mu = \int \varphi \,\mathrm{d}\mu + \int (\psi - \varphi) \,\mathrm{d}\mu$，而由 (i) 知第二个积分 ≥ 0，故 (iv) 成立. $\varphi = \psi$ a.e. 蕴涵 $\varphi \le \psi$ a.e. 且 $\psi \le \varphi$ a.e.，所以 (v) 由两次应用 (iv) 得到. 为证 (vi)，注意 $\varphi \le |\varphi|$，故 (iv) 蕴涵 $\int \varphi \,\mathrm{d}\mu \le \int |\varphi| \,\mathrm{d}\mu$. 又 $-\varphi \le |\varphi|$，故 (iv) 和 (ii) 蕴涵 $-\int \varphi \,\mathrm{d}\mu \le \int |\varphi| \,\mathrm{d}\mu$. 由于 $|y| = \max(y, -y)$，即得结论. □

第 2 步. 设 E 为 $\mu(E) < \infty$ 的集合，f 为在 E^c 上**为零**的有界函数. 为定义 f 的积分，我们观察到若 φ, ψ 为满足 $\varphi \le f \le \psi$ 的简单函数，则我们希望有 $\int \varphi \,\mathrm{d}\mu \le \int f \,\mathrm{d}\mu \le \int \psi \,\mathrm{d}\mu$，于是令

$$\int f \,\mathrm{d}\mu = \sup_{\varphi \le f} \int \varphi \,\mathrm{d}\mu = \inf_{\psi \ge f} \int \psi \,\mathrm{d}\mu. \tag{1.4.1}$$

这里及第 2 步的其余部分，均假设 φ 和 ψ 在 E^c 上为零. 为说明定义的合理性，需证明上下确界相等. 由引理 1.4.2 的 (iv) 知

$$\sup_{\varphi \le f} \int \varphi \,\mathrm{d}\mu \le \inf_{\psi \ge f} \int \psi \,\mathrm{d}\mu.$$

为证反向不等式，设 $|f| \le M$，令

$$E_k = \{x \in E : kM/n \ge f(x) > (k-1)M/n\}, \quad -n \le k \le n,$$

$$\psi_n(x) = \sum_{k=-n}^{n} (kM/n)\mathbf{1}_{E_k}, \quad \varphi_n(x) = \sum_{k=-n}^{n} ((k-1)M/n)\mathbf{1}_{E_k}.$$

由定义，$\psi_n(x) - \varphi_n(x) = (M/n)\mathbf{1}_E$，故 $\int (\psi_n(x) - \varphi_n(x)) \,\mathrm{d}\mu = (M/n)\mu(E)$.

由于 $\varphi_n(x) \le f(x) \le \psi_n(x)$，由引理 1.4.1 的 (iii) 知

$$\sup_{\varphi \le f} \int \varphi \,\mathrm{d}\mu \ge \cdots \ge -(M/n)\mu(E) + \inf_{\psi \ge f} \int \psi \,\mathrm{d}\mu.$$

最后一个不等式对所有 n 成立，故证毕.

引理 1.4.3. 设 E 为 $\mu(E) < \infty$ 的集合. 若 f 和 g 为在 E^c 上为零的有界函数，则性质 (i)–(vi) 成立.

证明. 由于可取 $\varphi \equiv 0$，(i) 由定义显然. 为证 (ii)，注意若 $a > 0$，则 $a\varphi \le af$ 当且仅当 $\varphi \le f$，故

$$\int af \,\mathrm{d}\mu = \sup_{\varphi \le f} \int a\varphi \,\mathrm{d}\mu = \sup_{\varphi \le f} a \int \varphi \,\mathrm{d}\mu = a \sup_{\varphi \le f} \int \varphi \,\mathrm{d}\mu = a \int f \,\mathrm{d}\mu.$$

当 $a<0$ 时，注意 $a\varphi \le af$ 当且仅当 $\varphi \ge f$，故 $\int af\,\mathrm{d}\mu = a\int f\,\mathrm{d}\mu$.

为证 (iii)，注意若 $\psi_1 \ge f$ 且 $\psi_2 \ge g$，则 $\psi_1+\psi_2 \ge f+g$，故

$$\inf_{\psi\ge f+g}\int \psi\,\mathrm{d}\mu \le \inf_{\psi_1\ge f,\psi_2\ge g}\int(\psi_1+\psi_2)\,\mathrm{d}\mu.$$

利用简单函数的线性性，可得 $\int f+g\,\mathrm{d}\mu = \cdots \le \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$.

为证反向不等式，注意将上一结论应用于 $-f$ 和 $-g$ 并结合 (ii) 可得

$$-\int f+g\,\mathrm{d}\mu \le -\int f\,\mathrm{d}\mu - \int g\,\mathrm{d}\mu.$$

(iv)–(vi) 由 (i)–(iii) 及引理 1.4.2 得到. □

记号. 我们定义 f 在集合 E 上的积分：$\int_E f\,\mathrm{d}\mu \equiv \int f\cdot \mathbf{1}_E\,\mathrm{d}\mu$.

第 3 步. 若 $f\ge 0$，令

$$\int f\,\mathrm{d}\mu = \sup\left\{\int h\,\mathrm{d}\mu : 0\le h\le f,\ h \text{ 有界且}\mu(\{x:h(x)>0\})<\infty\right\}.$$

最后一个定义是容易理解的，因为其含义显然. 下面的结果将有助于我们计算积分的值.

引理 1.4.4. 设 $E_n\uparrow\Omega$，其中 $\mu(E_n)<\infty$，令 $a\wedge b=\min(a,b)$. 则 $\int_{E_n} f\wedge n\,\mathrm{d}\mu \uparrow \int f\,\mathrm{d}\mu$（当 $n\uparrow\infty$ 时）.

证明. 由引理 1.4.3 的 (iv) 显然可知左边随 n 增加而增加. 由于 $h=(f\wedge n)\mathbf{1}_{E_n}$ 是右边上确界中的一个可能取法，故每一项都小于右边的积分. 为证极限等于 $\int f\,\mathrm{d}\mu$，注意若 $0\le h\le f$，$h\le M$，且 $\mu(\{x:h(x)>0\})<\infty$，则对 $n\ge M$，利用 $h\le M$、(iv) 和 (iii) 有

$$\int_{E_n} f\wedge n\,\mathrm{d}\mu \ge \int_{E_n} h\,\mathrm{d}\mu = \int h\,\mathrm{d}\mu - \int_{E_n^c} h\,\mathrm{d}\mu.$$

现在 $0\le \int_{E_n^c} h\,\mathrm{d}\mu \le M\mu(E_n^c\cap\{x:h(x)>0\})\to 0$（当 $n\to\infty$ 时），故

$$\liminf_{n\to\infty}\int_{E_n} f\wedge n\,\mathrm{d}\mu \ge \int h\,\mathrm{d}\mu,$$

这就证明了所需结果，因为 h 是定义 f 积分的那个类中的任意元素. □

引理 1.4.5. 设 $f,g\ge 0$.

(i) $\int f\,\mathrm{d}\mu \ge 0$.

(ii) 若 $a>0$，则 $\int af\,\mathrm{d}\mu = a\int f\,\mathrm{d}\mu$.

(iii) $\int f+g\,\mathrm{d}\mu=\int f\,\mathrm{d}\mu+\int g\,\mathrm{d}\mu$.
(iv) 若 $0\le g\le f$ a.e.，则 $\int g\,\mathrm{d}\mu\le\int f\,\mathrm{d}\mu$.
(v) 若 $0\le g=f$ a.e.，则 $\int g\,\mathrm{d}\mu=\int f\,\mathrm{d}\mu$.
这里我们略去了 (vi)，因为对于 $f\ge 0$ 它是平凡的.

证明. (i) 由定义显然. (ii) 是清楚的，因为当 $a>0$ 时，$ah\le af$ 当且仅当 $h\le f$，且对定义类中的 h 有 $\int ah\,\mathrm{d}\mu=a\int h\,\mathrm{d}\mu$. 对于 (iii)，注意若 $f\ge h$ 且 $g\ge k$，则 $f+g\ge h+k$，故对 f 和 g 的定义类中的 h 和 k 取上确界得 $\int f+g\,\mathrm{d}\mu\ge\int f\,\mathrm{d}\mu+\int g\,\mathrm{d}\mu$.

为证另一方向，注意 $(a+b)\wedge n\le(a\wedge n)+(b\wedge n)$，故由引理 1.4.3 的 (iv) 和引理 1.4.4 的 (iii) 有

$$\int_{E_n}(f+g)\wedge n\,\mathrm{d}\mu\le\int_{E_n}f\wedge n\,\mathrm{d}\mu+\int_{E_n}g\wedge n\,\mathrm{d}\mu.$$

令 $n\to\infty$ 并利用引理 1.4.4 即得 (iii). 如前所述，(iv) 和 (v) 由 (i)、(iii) 及引理 1.4.2 得到. □

第 4 步. 若 $\int|f|\,\mathrm{d}\mu<\infty$，则称 f 是**可积的**. 令 $f^+(x)=f(x)\vee 0$，$f^-(x)=(-f(x))\vee 0$，其中 $a\vee b=\max(a,b)$. 显然，$f(x)=f^+(x)-f^-(x)$ 且 $|f(x)|=f^+(x)+f^-(x)$.

我们定义 f 的积分为

$$\int f\,\mathrm{d}\mu=\int f^+\,\mathrm{d}\mu-\int f^-\,\mathrm{d}\mu.$$

右边是良定义的，因为 $f^+,f^-\le|f|$，而由引理 1.4.5 的 (iv) 知这两个积分均有限. 最后，我们再证一次上述六条性质. 为此，下面的事实很有用：

引理 1.4.6. 若 $f=f_1-f_2$，其中 $f_1,f_2\ge 0$ 且 $\int f_i\,\mathrm{d}\mu<\infty$，则 $\int f\,\mathrm{d}\mu=\int f_1\,\mathrm{d}\mu-\int f_2\,\mathrm{d}\mu$.

证明. $f_1+f^-=f_2+f^+$，且四个函数均 ≥ 0，故由引理 1.4.5 的 (iii)，

$$\int f_1\,\mathrm{d}\mu+\int f^-\,\mathrm{d}\mu=\int(f_1+f^-)\,\mathrm{d}\mu=\int(f_2+f^+)\,\mathrm{d}\mu=\int f_2\,\mathrm{d}\mu+\int f^+\,\mathrm{d}\mu.$$

整理即得所需结论. □

定理 1.4.7. 设 f 和 g 可积. 则性质 (i)–(vi) 成立.

证明. (i) 是平凡的. (ii) 是清楚的，因为若 $a>0$，则 $(af)^+=a(f^+)$，等等. 为证 (iii)，注意 $f+g=(f^++g^+)-(f^-+g^-)$，故利用引理 1.4.6 和引理 1.4.5：

$$\int f+g\,\mathrm{d}\mu=\int(f^++g^+)\,\mathrm{d}\mu-\int(f^-+g^-)\,\mathrm{d}\mu=\int f^+\,\mathrm{d}\mu+\int g^+\,\mathrm{d}\mu-\int f^-\,\mathrm{d}\mu-\int g^-\,\mathrm{d}\mu.$$

照例，(iv)–(vi) 由 (i)–(iii) 及引理 1.4.2 得到. □

特殊情形的记号. (a) 当 $(\Omega,\mathcal{F},\mu)=(\mathbb{R}^d,\mathcal{R}^d,\lambda)$ 时，我们将 $\int f\,\mathrm{d}\lambda$ 写作 $\int f(x)\,\mathrm{d}x$.

(b) 当 $(\Omega,\mathcal{F},\mu)=(\mathbb{R},\mathcal{R},\lambda)$ 且 $E=[a,b]$ 时，我们将 $\int_E f\,\mathrm{d}\lambda$ 写作 $\int_a^b f(x)\,\mathrm{d}x$.

(c) 当 $(\Omega,\mathcal{F},\mu)=(\mathbb{R},\mathcal{R},\mu)$ 且 $\mu((a,b])=G(b)-G(a)\,(a<b)$ 时，我们将 $\int f\,\mathrm{d}\mu$ 写作 $\int f(x)\,\mathrm{d}G(x)$.

(d) 当 Ω 为可数集，$\mathcal{F}=\Omega$ 的所有子集，且 μ 为**计数测度**时，我们将 $\int f\,\mathrm{d}\mu$ 写作 $\sum_{i\in\Omega} f(i)$.

提及例 (d) 主要是为了说明：关于和的结果可由关于积分的结果导出. μ 为概率测度时的特殊记号将在 1.6 节中讨论.

习题

习题 1.4.1. 证明若 $f\geq 0$ 且 $\int f\,\mathrm{d}\mu=0$，则 $f=0$ a.e.

习题 1.4.2. 设 $f\geq 0$ 且 $E_{n,m}=\{x: m/2^n\leq f(x)<(m+1)/2^n\}$. 证明当 $n\uparrow\infty$ 时，

$$\sum_{m=1}^{\infty}(m/2^n)\mu(E_{n,m})\uparrow\int f\,\mathrm{d}\mu.$$

习题 1.4.3. 设 g 为 $\mathbb{R}$ 上的可积函数，$\varepsilon>0$.

(i) 利用积分的定义推出存在简单函数 $\varphi=\sum_k b_k\mathbf{1}_{A_k}$ 使得 $\int|g-\varphi|\,\mathrm{d}x<\varepsilon$.

(ii) 利用习题 A.2.1，用区间的有限并逼近 A_k 以得到阶梯函数 $q=\sum_{j=1}^{k}c_j\mathbf{1}_{(a_{j-1},a_j)}$（$a_0<a_1<\cdots<a_k$），使得 $\int|\varphi-q|<\varepsilon$.

(iii) 将 q 的角“磨圆”以得到连续函数 r 使得 $\int|q-r|\,\mathrm{d}x<\varepsilon$.

习题 1.4.4. 证明 **Riemann-Lebesgue 引理**：若 g 可积，则 $\lim_{n\to\infty}\int g(x)\cos(nx)\,\mathrm{d}x=0$.

提示：若 g 是阶梯函数，这是容易的. 现在利用上一习题.

第 5 节　积分的性质

本节将建立上一节所定义积分的若干性质. 首先推广定理 1.4.7 中的 (vi).

定理 1.5.1. Jensen 不等式. 设 φ 是**凸的**函数, 即对所有 $\lambda \in (0,1)$ 和 $x, y \in \mathbb{R}$, 有 $\lambda\varphi(x)+(1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y)$. 若 μ 是**概率测度**，且 f 与 $\varphi(f)$ 均**可积**，则

$$\varphi\left(\int f \,\mathrm{d}\mu\right) \leq \int \varphi(f) \,\mathrm{d}\mu.$$

证明. 记 $c = \int f \,\mathrm{d}\mu$，并设 $l(x) = ax + b$ 是一个线性函数，满足 $l(c) = \varphi(c)$ 且 $\varphi(x) \geq l(x)$. 为证明这样的函数存在，回顾凸性蕴涵

$$\lim_{h\downarrow 0} \frac{\varphi(c) - \varphi(c-h)}{h} \leq \lim_{h\downarrow 0} \frac{\varphi(c+h) - \varphi(c)}{h}.$$

(由于序列单调, 上述极限存在.)若取 a 为介于这两个极限之间的任意数, 并令 $l(x) = a(x-c)+\varphi(c)$, 则 l 具有所需性质. 既然 l 的存在性已确立，其余部分极易验证. 定理 1.4.7 中的 (iv) 蕴涵

$$\int \varphi(f) \,\mathrm{d}\mu \geq \int (af+b) \,\mathrm{d}\mu = a\int f \,\mathrm{d}\mu + b = l\left(\int f \,\mathrm{d}\mu\right) = \varphi\left(\int f \,\mathrm{d}\mu\right),$$

这是因为 $c = \int f \,\mathrm{d}\mu$ 且 $l(c) = \varphi(c)$. □

记 $\|f\|_p = \left(\int |f|^p \,\mathrm{d}\mu\right)^{1/p}$ (对 $1 \leq p < \infty$)，并注意对任意实数 c 有 $\|cf\|_p = |c| \cdot \|f\|_p$.

定理 1.5.2. Hölder 不等式. 若 $p, q \in (1,\infty)$ 满足 $1/p + 1/q = 1$，则

$$\int |fg| \,\mathrm{d}\mu \leq \|f\|_p \|g\|_q.$$

证明. 若 $\|f\|_p$ 或 $\|g\|_q = 0$，则 $|fg| = 0$ a.e.，故只需考虑 $\|f\|_p$ 和 $\|g\|_q > 0$ 的情形；或者，两边同除以 $\|f\|_p\|g\|_q$ 后，只需在 $\|f\|_p = \|g\|_q = 1$ 时证明即可. 固定 $y \geq 0$，令

$$\varphi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy \quad \text{对} x \geq 0,$$

则 $\varphi'(x) = x^{p-1} - y$ 且 $\varphi''(x) = (p-1)x^{p-2}$，故 φ 在 $x_0 = y^{1/(p-1)}$ 处取得极小值. 由 $q = p/(p-1)$ 及 $x_0^p = y^{p/(p-1)} = y^q$ 得

$$\varphi(x_0) = y^q\left(\frac{1}{p} + \frac{1}{q}\right) - y^{1/(p-1)}y = 0.$$

由于 x_0 是极小值点，可得 $xy \le x^p/p + y^q/q$. 令 $x = |f|$，$y = |g|$ 并积分得

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

□

注释. 特殊情形 $p = q = 2$ 称为 **Cauchy-Schwarz 不等式**. 对此情形，可直接证明如下：对任意 θ，

$$0 \le \int (f + \theta g)^2 \,\mathrm{d}\mu = \int f^2 \,\mathrm{d}\mu + \theta \left(2 \int fg \,\mathrm{d}\mu \right) + \theta^2 \int g^2 \,\mathrm{d}\mu,$$

故右边的二次式 $a\theta^2 + b\theta + c$ 至多有一个实根. 回忆二次方程求根公式

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

可见 $b^2 - 4ac \le 0$，即为所求结果.

下面给出保证

$$\lim \int f_n \,\mathrm{d}\mu = \int \lim f_n \,\mathrm{d}\mu$$

成立的条件. 首先定义依测度收敛. 称 $f_n \to f$ **依测度**收敛，是指对任意 $\varepsilon > 0$，当 $n \to \infty$ 时 $\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) \to 0$. 在有限测度空间上，这一条件比 $f_n \to f$ a.e. 更弱，但下一个结果在更一般的框架下更容易证明.

定理 1.5.3. 有界收敛定理. 设 E 为满足 $\mu(E) < \infty$ 的集合. 假设 f_n 在 E^c 上为零，$|f_n(x)| \le M$，且 $f_n \to f$ 依测度. 则

$$\int f \,\mathrm{d}\mu = \lim \int f_n \,\mathrm{d}\mu.$$

例题 1.5.4. 考虑实直线 $\mathbb{R}$ 配备 Borel 集 $\mathcal{R}$ 和 Lebesgue 测度 λ 的情形. 函数 $f_n(x) = 1/n$（在 $[0, n]$ 上）且在其他地方为 0，表明当 $\mu(E) = \infty$ 时定理 1.5.3 的结论不成立.

证明. 令 $\varepsilon > 0$，$G_n = \{x : |f_n(x) - f(x)| < \varepsilon\}$，$B_n = E - G_n$. 利用定理 1.4.7 中的 (iii) 和 (vi)，

$$\begin{aligned}
\left| \int f \,\mathrm{d}\mu - \int f_n \,\mathrm{d}\mu \right| = \left| \int (f - f_n) \,\mathrm{d}\mu \right| &\le \int |f - f_n| \,\mathrm{d}\mu \\
&= \int_{G_n} |f - f_n| \,\mathrm{d}\mu + \int_{B_n} |f - f_n| \,\mathrm{d}\mu \\
&\le \varepsilon \mu(E) + 2M\mu(B_n).
\end{aligned}$$

$f_n \to f$ 依测度蕴涵 $\mu(B_n) \to 0$. 由于 $\varepsilon > 0$ 是任意的且 $\mu(E) < \infty$，证毕. □

定理 1.5.5. Fatou 引理. 若 $f_n \geq 0$，则

$$\int \liminf f_n \,\mathrm{d}\mu \geq \liminf \int f_n \,\mathrm{d}\mu.$$

例题 1.5.6. 例 1.5.4 表明定理 1.5.5 中可能出现严格不等式. 函数 $f_n(x) = n \cdot \mathbf{1}_{(0,1/n]}(x)$ 在 $(0,1)$（配备 Borel 集和 Lebesgue 测度）上表明，即使在有限测度空间上这种情况也可能发生.

证明. 记 $g_n(x) = \inf_{m \geq n} f_m(x)$. 有 $f_n(x) \geq g_n(x)$，且当 $n \uparrow \infty$ 时，$g_n(x) \uparrow g(x) = \liminf_{n\to\infty} f_n(x)$. 由于 $\int f_n \,\mathrm{d}\mu \geq \int g_n \,\mathrm{d}\mu$，故只需证明

$$\liminf \int g_n \,\mathrm{d}\mu \geq \int g \,\mathrm{d}\mu.$$

令 $E_m \uparrow \Omega$ 为有限测度集. 由于 $g_n \geq 0$ 且对固定的 m，

$$(g_n \wedge m) \cdot \mathbf{1}_{E_m} \to (g \wedge m) \cdot \mathbf{1}_{E_m} \quad \text{a.e.},$$

有界收敛定理 1.5.3 蕴涵

$$\liminf \int g_n \,\mathrm{d}\mu \geq \int g \wedge m \,\mathrm{d}\mu.$$

对 m 取上确界并利用定理 1.4.4 即得所需结果. □

定理 1.5.7. 单调收敛定理. 若 $f_n \geq 0$ 且 $f_n \uparrow f$，则

$$\int f_n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu.$$

证明. Fatou 引理（定理 1.5.5）蕴涵 $\liminf \int f_n \,\mathrm{d}\mu \geq \int f \,\mathrm{d}\mu$. 另一方面，$f_n \leq f$ 蕴涵 $\limsup \int f_n \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\mu$. □

定理 1.5.8. 控制收敛定理. 若 $f_n \to f$ a.e.，$|f_n| \leq g$ 对所有 n 成立，且 g 可积，则 $\int f_n \,\mathrm{d}\mu \to \int f \,\mathrm{d}\mu$.

证明. $f_n + g \geq 0$，故 Fatou 引理蕴涵

$$\liminf \int (f_n + g) \,\mathrm{d}\mu \geq \int (f + g) \,\mathrm{d}\mu.$$

两边减去 $\int g \,\mathrm{d}\mu$ 得

$$\liminf \int f_n \,\mathrm{d}\mu \geq \int f \,\mathrm{d}\mu.$$

将上述结果应用于 $-f_n$，得

$$\limsup \int f_n \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\mu,$$

证毕. □

习题

习题 1.5.1. 记 $\|f\|_\infty = \inf\{M : \mu(\{x : |f(x)| > M\}) = 0\}$. 证明 $\int |fg| \,\mathrm{d}\mu \leq \|f\|_1 \|g\|_\infty$.

习题 1.5.2. 证明若 μ 是概率测度，则 $\|f\|_\infty = \lim_{p\to\infty} \|f\|_p$.

习题 1.5.3. Minkowski 不等式.

(i) 设 $p \in (1,\infty)$. 不等式 $|f+g|^p \leq 2^p(|f|^p+|g|^p)$ 表明：若 $\|f\|_p$ 和 $\|g\|_p$ 均 $< \infty$，则 $\|f+g\|_p < \infty$. 对 $|f||f+g|^{p-1}$ 和 $|g||f+g|^{p-1}$ 应用 Hölder 不等式，证明 $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.

(ii) 证明当 $p = 1$ 或 $p = \infty$ 时上述结论仍然成立.

习题 1.5.4. 若 f 可积且 E_m 为不交集合，其并为 E，则 $\sum \int_{E_m} f \,\mathrm{d}\mu = \int_E f \,\mathrm{d}\mu$. 因此，若 $f \geq 0$，则 $\nu(E) = \int_E f \,\mathrm{d}\mu$ 定义了一个测度.

习题 1.5.5. 若 $g_n \uparrow g$ 且 $\int g_1^- \,\mathrm{d}\mu < \infty$，则 $\int g_n \,\mathrm{d}\mu \uparrow \int g \,\mathrm{d}\mu$.

习题 1.5.6. 若 $g_m \geq 0$，则 $\int \sum g_m \,\mathrm{d}\mu = \sum \int g_m \,\mathrm{d}\mu$.

习题 1.5.7. 设 $f \geq 0$.

(i) 证明当 $n \to \infty$ 时，$\int f \wedge n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu$.

(ii) 利用 (i) 推出：若 g 可积且 $\varepsilon > 0$，则可选取 $\delta > 0$ 使得 $\mu(A) < \delta$ 蕴涵 $\int_A |g| \,\mathrm{d}\mu < \varepsilon$.

习题 1.5.8. 证明若 f 在 $[a,b]$ 上可积，则 $g(x) = \int_{[a,x]} f(y) \,\mathrm{d}y$ 在 (a,b) 上连续.

习题 1.5.9. 证明若 $\|f\|_p < \infty$，则存在简单函数 φ_n 使得 $\|\varphi_n - f\|_p \to 0$.

习题 1.5.10. 证明若 $\sum \int |f_n| \,\mathrm{d}\mu < \infty$，则 $\int \sum f_n \,\mathrm{d}\mu = \sum \int f_n \,\mathrm{d}\mu$.

第 6 节 期望值

本节专门讨论关于概率测度 P 的积分. 若 $X \geq 0$ 是 $(\Omega, \mathcal{F}, P)$ 上的随机变量，则定义其**期望值**为 $EX = \int X \, \mathrm{d}P$, 这总是有意义的, 但可能是 ∞. 为将一般情形约化到非负情形, 记 $x^+ = \max\{x, 0\}$ 为 x 的**正部**，记 $x^- = \max\{-x, 0\}$ 为 x 的**负部**. 当减法有意义时，即 $EX^+ < \infty$ 或 $EX^- < \infty$，称 EX 存在并令 $EX = EX^+ - EX^-$.

EX 常被称为 X 的**均值**，记作 μ. EX 是通过积分 X 来定义的，因此它具有积分的所有性质. 由定理 1.4.5 和 1.4.7 以及显然的事实 $E(b) = b$（对任意实数 b），可得如下结果：

定理 1.6.1. 设 $X, Y \geq 0$ 或 $E|X|, E|Y| < \infty$.

(a) $E(X + Y) = EX + EY$.

(b) $E(aX + b) = aE(X) + b$，对任意实数 a, b.

(c) 若 $X \geq Y$，则 $EX \geq EY$.

本节用期望值的术语重述上节推导的积分性质，并证明一些新的性质. 为组织材料，将这些发展分为三个小节.

1.6.1 不等式

对于概率测度，定理 1.5.1 变为：

定理 1.6.2. Jensen 不等式. 设 ϕ 是凸函数，即对所有 $\lambda \in (0, 1)$ 和 $x, y \in \mathbb{R}$ 有

$$\lambda\phi(x) + (1 - \lambda)\phi(y) \geq \phi(\lambda x + (1 - \lambda)y).$$

则

$$E(\phi(X)) \geq \phi(EX),$$

只要两个期望均存在，即 $E|X| < \infty$ 且 $E|\phi(X)| < \infty$.

为记住不等式的方向，注意若 $P(X = x) = \lambda$ 且 $P(X = y) = 1 - \lambda$，则

$$E\phi(X) = \lambda\phi(x) + (1 - \lambda)\phi(y) \geq \phi(\lambda x + (1 - \lambda)y) = \phi(EX).$$

两个有用的特例是 $|EX| \leq E|X|$ 和 $(EX)^2 \leq E(X^2)$.

定理 1.6.3. Hölder 不等式. 若 $p, q \in [1, \infty]$ 满足 $1/p + 1/q = 1$，则

$$E|XY| \leq \|X\|_p \|Y\|_q.$$

这里 $\|X\|_r = (E|X|^r)^{1/r}$（对 $r \in [1, \infty)$）；$\|X\|_\infty = \inf\{M : P(|X| > M) = 0\}$.

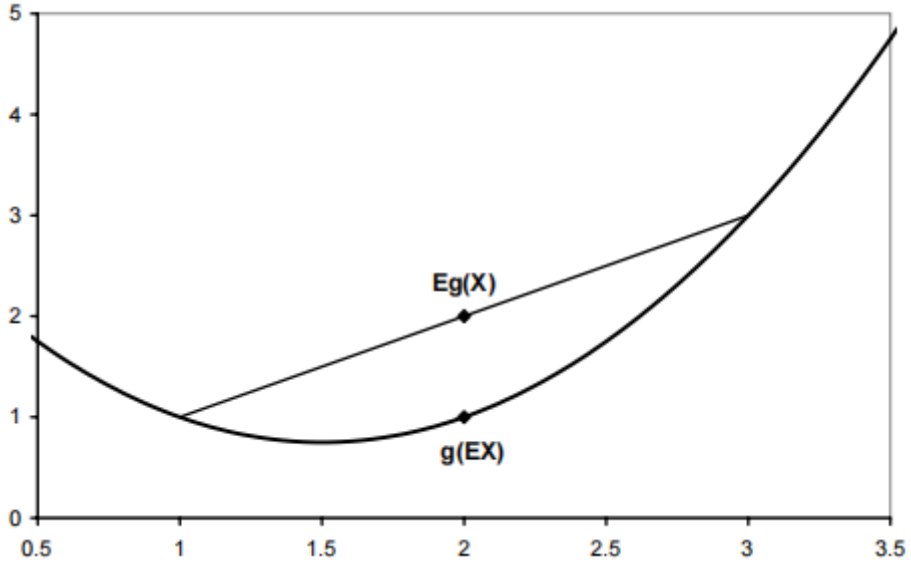

图 1.6: 对 $g(x) = x^2 - 3x + 3, P(X = 1) = P(X = 3) = 1/2$ 的 Jensen 不等式

为陈述下一结果，需要一些记号. 若仅在 $A \subset \Omega$ 上积分，记

$$E(X; A) = \int_A X \, \mathrm{d}P.$$

定理 1.6.4. Chebyshev 不等式. 设 $\phi : \mathbb{R} \to \mathbb{R}$ 满足 $\phi \geq 0$，令 $A \in \mathcal{R}$，并令 $i_A = \inf\{\phi(y) : y \in A\}$. 则

$$i_A P(X \in A) \leq E(\phi(X); X \in A) \leq E\phi(X).$$

证明. i_A 的定义以及 $\phi \geq 0$ 的事实蕴涵

$$i_A \mathbf{1}_{(X \in A)} \leq \phi(X) \mathbf{1}_{(X \in A)} \leq \phi(X).$$

取期望值并应用定理 1.6.1 的 (c) 即得所需结果. □

注释. 有些作者称此结果为 **Markov 不等式**，而将 **Chebyshev 不等式**之名用于 $\phi(x) = x^2$ 且 $A = \{x : |x| \geq a\}$ 的特殊情形：

$$a^2 P(|X| \geq a) \leq EX^2. \tag{1.6.1}$$

1.6.2 极限积分

下面用期望值的术语重述上节关于交换极限与积分次序的三个经典结果.

定理 1.6.5. Fatou 引理. 若 $X_n \geq 0$，则

$$\liminf_{n \to \infty} EX_n \geq E\left(\liminf_{n \to \infty} X_n\right).$$

定理 1.6.6. 单调收敛定理. 若 $0 \le X_n \uparrow X$，则 $EX_n \uparrow EX$.

定理 1.6.7. 控制收敛定理. 若 $X_n \to X$ a.s.，$|X_n| \le Y$ 对所有 n 成立，且 $EY < \infty$，则 $EX_n \to EX$.

定理 1.6.7 中 Y 为常数的特殊情形称为**有界收敛定理**.

在下面的发展中，还需要另一个关于极限积分的结果. 该结果最重要的特殊情形或许是当 $g(x) = |x|^p$ $(p > 1)$ 且 $h(x) = x$ 时.

定理 1.6.8. 设 $X_n \to X$ a.s. 令 g, h 为连续函数，满足

(i) $g \ge 0$ 且当 $|x| \to \infty$ 时 $g(x) \to \infty$,

(ii) 当 $|x| \to \infty$ 时 $|h(x)|/g(x) \to 0$,

且 (iii) $Eg(X_n) \le K < \infty$ 对所有 n 成立. 则 $Eh(X_n) \to Eh(X)$.

证明. 由 h 减去一个常数，不失一般性可设 $h(0) = 0$. 取 M 足够大使得 $P(|X| = M) = 0$ 且当 $|x| \ge M$ 时 $g(x) > 0$. 给定随机变量 Y，令 $\bar{Y} = Y\mathbf{1}_{(|Y| \le M)}$. 由于 $P(|X| = M) = 0$，有 $\bar{X}_n \to \bar{X}$ a.s. 由于 $h(\bar{X}_n)$ 有界且 h 连续，由有界收敛定理得

$$Eh(\bar{X}_n) \to Eh(\bar{X}). \tag{a}$$

为控制截断的影响，使用如下估计：

$$|Eh(\bar{Y}) - Eh(Y)| \le E|h(\bar{Y}) - h(Y)| \le E(|h(Y)|; |Y| > M) \le \epsilon_M Eg(Y), \tag{b}$$

其中 $\epsilon_M = \sup\{|h(x)|/g(x) : |x| \ge M\}$. 为验证第二个不等式，注意当 $|Y| \le M$ 时 $\bar{Y} = Y$，且已设 $h(0) = 0$. 第三个不等式由 ϵ_M 的定义得到.

在 (b) 中取 $Y = X_n$ 并利用 (iii)，可得

$$|Eh(\bar{X}_n) - Eh(X_n)| \le K\epsilon_M. \tag{c}$$

为估计 $|Eh(\bar{X}) - Eh(X)|$，观察到 $g \ge 0$ 且 g 连续，故 Fatou 引理蕴涵

$$Eg(X) \le \liminf_{n\to\infty} Eg(X_n) \le K.$$

在 (b) 中取 $Y = X$ 得

$$|Eh(\bar{X}) - Eh(X)| \le K\epsilon_M. \tag{d}$$

三角不等式给出

$$|Eh(X_n) - Eh(X)| \le |Eh(X_n) - Eh(\bar{X}_n)| + |Eh(\bar{X}_n) - Eh(\bar{X})| + |Eh(\bar{X}) - Eh(X)|.$$

取极限并利用 (a), (c), (d)，得

$$\limsup_{n\to\infty} |Eh(X_n) - Eh(X)| \le 2K\epsilon_M,$$

这就证明了所需结果，因为 $K < \infty$ 且当 $M \to \infty$ 时 $\epsilon_M \to 0$. □

1.6.3 计算期望值

在 $(\Omega, \mathcal{F}, P)$ 上积分在理论上是好的，但要进行计算，必须转移到可以进行微积分的空间. 在大多数情形，将对 $S = \mathbb{R}^d$ 应用下一结果.

定理 1.6.9. 变量替换公式. 设 X 是 $(S, \mathcal{S})$ 上的随机元，具有分布 μ，即 $\mu(A) = P(X \in A)$. 若 f 是从 $(S, \mathcal{S})$ 到 $(\mathbb{R}, \mathcal{R})$ 的可测函数，满足 $f \ge 0$ 或 $E|f(X)| < \infty$，则

$$Ef(X) = \int_S f(y)\,\mu(\mathrm{d}y).$$

注释. 为解释此名称，将 h 写作 X，将 $P \circ h^{-1}$ 写作 μ，得到

$$\int_\Omega f(h(\omega))\,\mathrm{d}P = \int_S f(y)\,\mathrm{d}(P \circ h^{-1}).$$

证明. 下面通过验证四个越来越一般的特殊情形来证明此结果，这四个情形与 1.4 节中定义积分的方式平行. 读者应注意所使用的方法，因为下面还将多次使用它.

情形 1：指示函数. 若 $B \in \mathcal{S}$ 且 $f = \mathbf{1}_B$，则回忆相关定义可得

$$E\mathbf{1}_B(X) = P(X \in B) = \mu(B) = \int_S \mathbf{1}_B(y)\,\mu(\mathrm{d}y).$$

情形 2：简单函数. 设 $f(x) = \sum_{m=1}^n c_m \mathbf{1}_{B_m}$，其中 $c_m \in \mathbb{R}$，$B_m \in \mathcal{S}$. 期望值的线性性、情形 1 的结果以及积分的线性性蕴涵

$$Ef(X) = \sum_{m=1}^n c_m E\mathbf{1}_{B_m}(X) = \sum_{m=1}^n c_m \int_S \mathbf{1}_{B_m}(y)\,\mu(\mathrm{d}y) = \int_S f(y)\,\mu(\mathrm{d}y).$$

情形 3：非负函数. 若 $f \ge 0$，令

$$f_n(x) = ([2^n f(x)]/2^n) \wedge n,$$

其中 $[x]$ = 不超过 x 的最大整数且 $a \wedge b = \min\{a, b\}$，则 f_n 是简单函数且 $f_n \uparrow f$，故利用简单函数

的结果和单调收敛定理：

$$Ef(X) = \lim_n Ef_n(X) = \lim_n \int_S f_n(y)\,\mu(\mathrm{d}y) = \int_S f(y)\,\mu(\mathrm{d}y).$$

情形 4：可积函数. 一般情形现在可以通过写 $f(x) = f(x)^+ - f(x)^-$ 得到. 条件 $E|f(X)| < \infty$ 保证 $Ef(X)^+$ 和 $Ef(X)^-$ 均有限. 故利用非负函数的结果以及期望值和积分的线性性：

$$Ef(X) = Ef(X)^+ - Ef(X)^- = \int_S f(y)^+\,\mu(\mathrm{d}y) - \int_S f(y)^-\,\mu(\mathrm{d}y) = \int_S f(y)\,\mu(\mathrm{d}y),$$

证毕. □

定理 1.6.9 的一个推论是，可以通过实直线上的积分来计算随机变量函数的期望值. 在讨论一些例子之前，需要引入将要计算的术语. 若 k 是正整数，则 EX^k 称为 X 的 k **阶矩**. 一阶矩 EX 通常称为**均值**，记作 μ. 若 $EX^2 < \infty$，则 X 的**方差**定义为 $\mathrm{var}(X) = E(X-\mu)^2$. 为计算方差，以下公式很有用：

$$\mathrm{var}(X) = E(X-\mu)^2 = EX^2 - 2\mu EX + \mu^2 = EX^2 - \mu^2. \tag{1.6.2}$$

由此立即得到

$$\mathrm{var}(X) \le EX^2. \tag{1.6.3}$$

这里 EX^2 是 X^2 的期望值. 当想要 EX 的平方时，会写作 $(EX)^2$. 由于 $E(aX+b) = aEX + b$（由定理 1.6.1 的 (b)），由定义易知

$$\mathrm{var}(aX+b) = E(aX+b-E(aX+b))^2 = a^2E(X-EX)^2 = a^2\,\mathrm{var}(X). \tag{1.6.4}$$

下面转向具体例子，并留给读者前两个例子中的微积分计算.（分部积分.）

例题 1.6.10. 若 X 具有速率 1 的**指数分布**，则

$$EX^k = \int_0^\infty x^k e^{-x}\,\mathrm{d}x = k!$$

故 X 的均值为 1，方差为 $EX^2 - (EX)^2 = 2 - 1^2 = 1$. 若令 $Y = X/\lambda$，则由习题 1.2.5，Y 具有密度 $\lambda e^{-\lambda y}$（$y \ge 0$），即参数 λ 的指数密度. 由定理 1.6.1 的 (b) 和 (1.6.4)，可得 Y 的均值为 $1/\lambda$，方差为 $1/\lambda^2$.

例题 1.6.11. 若 X 具有**标准正态分布**，

$$EX = \int x(2\pi)^{-1/2}\exp(-x^2/2)\,\mathrm{d}x = 0 \quad \text{（由对称性）}$$

$$\text{var}(X) = EX^2 = \int x^2 (2\pi)^{-1/2} \exp(-x^2/2)\, \mathrm{d}x = 1.$$

若令 $\sigma > 0$，$\mu \in \mathbb{R}$，且 $Y = \sigma X + \mu$，则由定理 1.6.1 的 (b) 和 (1.6.4)，可得 $EY = \mu$ 且 $\text{var}(Y) = \sigma^2$. 由习题 1.2.5，Y 具有密度

$$(2\pi\sigma^2)^{-1/2} \exp(-(y-\mu)^2/2\sigma^2),$$

即均值为 μ、方差为 σ^2 的**正态分布**.

下面考虑一些离散分布. 第一个非常简单，但将在下面多次使用，故在此记录它.

例题 1.6.12. 称 X 具有参数 p 的**伯努利分布**，若 $P(X = 1) = p$ 且 $P(X = 0) = 1 - p$. 显然，

$$EX = p \cdot 1 + (1-p) \cdot 0 = p.$$

由于 $X^2 = X$，有 $EX^2 = EX = p$ 且

$$\text{var}(X) = EX^2 - (EX)^2 = p - p^2 = p(1-p).$$

例题 1.6.13. 称 X 具有参数 λ 的**泊松分布**，若

$$P(X = k) = e^{-\lambda}\lambda^k/k! \quad \text{对} k = 0, 1, 2, \ldots$$

为计算泊松随机变量的矩，借助一点灵感观察到对 $k \geq 1$

$$E(X(X-1)\cdots(X-k+1)) = \sum_{j=k}^{\infty} j(j-1)\cdots(j-k+1)e^{-\lambda}\frac{\lambda^j}{j!} = \lambda^k \sum_{j=k}^{\infty} e^{-\lambda}\frac{\lambda^{j-k}}{(j-k)!},$$

其中等式由 (i) 当 $j < k$ 时 $j(j-1)\cdots(j-k+1) = 0$ 的事实，(ii) 约去阶乘的一部分，以及 (iii) 泊松分布的总质量为 1 的事实得到. 利用最后一个公式，可得 $EX = \lambda$ 而

$$\text{var}(X) = EX^2 - (EX)^2 = E(X(X-1)) + EX - \lambda^2 = \lambda.$$

例题 1.6.14. 称 N 具有**成功概率** $p \in (0, 1)$ 的**几何分布**，若

$$P(N = k) = p(1-p)^{k-1} \quad \text{对} k = 1, 2, \ldots$$

N 是观察到概率为 p 的事件所需的独立试验次数. 对恒等式

$$\sum_{k=0}^{\infty} (1-p)^k = 1/p$$

求导并参见例 A.5.6 的论证，得

$$-\sum_{k=1}^{\infty} k(1-p)^{k-1} = -1/p^2, \quad \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = 2/p^3.$$

由此可得

$$EN = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = 1/p,$$

$$EN(N-1) = \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} = 2(1-p)/p^2,$$

$$\text{var}(N) = EN^2 - (EN)^2 = EN(N-1) + EN - (EN)^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

习题

习题 1.6.1. 设 ϕ 是严格凸的, 即对 $\lambda \in (0,1)$ 严格不等式成立. 证明在定理 1.6.2 的假设下, $\phi(EX) = E\phi(X)$ 蕴涵 $X = EX$ a.s.

习题 1.6.2. 设 $\phi : \mathbb{R}^n \to \mathbb{R}$ 是凸的. 模仿定理 1.5.1 的证明，证明

$$E\phi(X_1, \ldots, X_n) \geq \phi(EX_1, \ldots, EX_n),$$

只要 $E|\phi(X_1, \ldots, X_n)| < \infty$ 且对所有 i 有 $E|X_i| < \infty$.

习题 1.6.3. Chebyshev 不等式是且不是精确的.

(i) 证明定理 1.6.4 是精确的，方法是证明若 $0 < b \leq a$ 固定，则存在 X 满足 $EX^2 = b^2$ 使得 $P(|X| \geq a) = b^2/a^2$.

(ii) 证明定理 1.6.4 不是精确的，方法是证明若 X 满足 $0 < EX^2 < \infty$，则

$$\lim_{a\to\infty} a^2 P(|X| \geq a)/EX^2 = 0.$$

习题 1.6.4. 单侧 Chebyshev 界.

(i) 设 $a > b > 0$, $0 < p < 1$, 并设 X 满足 $P(X = a) = p$ 且 $P(X = -b) = 1-p$. 对 $\phi(x) = (x+b)^2$ 应用定理 1.6.4，并得出结论：若 Y 是满足 $EY = EX$ 且 $\text{var}(Y) = \text{var}(X)$ 的任意随机变量，则 $P(Y \geq a) \leq p$，且当 $Y = X$ 时等号成立.

(ii) 设 $EY = 0$，$\text{var}(Y) = \sigma^2$，且 $a > 0$. 证明 $P(Y \geq a) \leq \sigma^2/(a^2 + \sigma^2)$，且存在 Y 使得等号成立.

习题 1.6.5. 两个不存在的下界. 证明：(i) 若 $\epsilon > 0$，则 $\inf\{P(|X| > \epsilon) : EX = 0, \text{var}(X) = 1\} = 0$. (ii) 若 $y \geq 1$，$\sigma^2 \in (0, \infty)$，则 $\inf\{P(|X| > y) : EX = 1, \text{var}(X) = \sigma^2\} = 0$.

习题 1.6.6. 一个有用的下界. 设 $Y \geq 0$ 满足 $EY^2 < \infty$. 对 $Y\mathbf{1}_{(Y>0)}$ 应用 Cauchy-Schwarz 不等式，并得出结论

$$P(Y > 0) \geq (EY)^2/EY^2.$$

习题 1.6.7. 设 $\Omega = (0,1)$ 配备 Borel 集和 Lebesgue 测度. 设 $\alpha \in (1,2)$ 且 $X_n = n^\alpha \mathbf{1}_{(1/(n+1),1/n)} \to 0$ a.s. 证明定理 1.6.8 可应用于 $h(x) = x$ 和 $g(x) = |x|^{2/\alpha}$，但 X_n 不被可积函数控制.

习题 1.6.8. 设概率测度 μ 满足 $\mu(A) = \int_A f(x)\,\mathrm{d}x$ 对所有 $A \in \mathcal{R}$ 成立. 使用定理 1.6.9 的证明技巧，证明对任意满足 $g \geq 0$ 或 $\int |g(x)|\,\mu(\mathrm{d}x) < \infty$ 的 g 有

$$\int g(x)\,\mu(\mathrm{d}x) = \int g(x) f(x)\,\mathrm{d}x.$$

习题 1.6.9. 容斥公式. 设 $A_1, A_2, \ldots, A_n$ 为事件且 $A = \bigcup_{i=1}^n A_i$. 证明 $\mathbf{1}_A = 1 - \prod_{i=1}^n (1 - \mathbf{1}_{A_i})$. 展开右端，然后取期望值以得出

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).$$

习题 1.6.10. Bonferroni 不等式. 设 $A_1, A_2, \ldots, A_n$ 为事件且 $A = \bigcup_{i=1}^n A_i$. 证明 $\mathbf{1}_A \leq \sum_{i=1}^n \mathbf{1}_{A_i}$，等等，然后取期望值以得到

第 7 节 乘积测度、Fubini 定理

设 $(X,\mathcal{A},\mu_1)$ 和 $(Y,\mathcal{B},\mu_2)$ 为两个 **σ-有限测度空间**. 记

$$\Omega = X\times Y = \{(x,y) : x\in X, y\in Y\},$$

$$\mathcal{S} = \{A\times B : A\in\mathcal{A}, B\in\mathcal{B}\}.$$

$\mathcal{S}$ 中的集合称为**矩形**. 容易验证 $\mathcal{S}$ 是一个半代数:

$$(A\times B)\cap(C\times D) = (A\cap C)\times(B\cap D),$$

$$(A\times B)^c = (A^c\times B)\cup(A\times B^c)\cup(A^c\times B^c).$$

记 $\mathcal{F}=\mathcal{A}\times\mathcal{B}$ 为由 $\mathcal{S}$ 生成的 σ-代数.

定理 1.7.1. 存在 $\mathcal{F}$ 上的唯一测度 μ 满足

$$\mu(A\times B) = \mu_1(A)\mu_2(B).$$

记号. μ 常记为 $\mu_1\times\mu_2$.

证明. 由定理 1.1.9，只需验证若 $A\times B = +_i(A_i\times B_i)$ 为有限或可数个不交并，则

$$\mu(A\times B) = \sum_i \mu(A_i\times B_i).$$

对每个 $x\in A$，记 $I(x)=\{i : x\in A_i\}$. 由 $B = +_{i\in I(x)}B_i$ 得

$$\mathbf{1}_A(x)\mu_2(B) = \sum_i \mathbf{1}_{A_i}(x)\mu_2(B_i).$$

对 μ_1 积分并利用习题 1.5.6 可得

$$\mu_1(A)\mu_2(B) = \sum_i \mu_1(A_i)\mu_2(B_i),$$

这就证明了结论. □

利用定理 1.7.1 和归纳法可知，若 $(\Omega_i,\mathcal{F}_i,\mu_i)$，$i=1,\dots,n$ 为 σ-有限测度空间且 $\Omega=\Omega_1\times\cdots\times$

Ω_n，则 $\mathcal{F}$（由形如 $A_1 \times \cdots \times A_n$ 的集合生成，其中 $A_i \in \mathcal{F}_i$）上存在唯一测度 μ 满足

$$\mu(A_1 \times \cdots \times A_n) = \prod_{m=1}^{n} \mu_m(A_m).$$

当 $(\Omega_i, \mathcal{F}_i, \mu_i) = (\mathbb{R}, \mathcal{R}, \lambda)$ 对所有 i 成立时，所得结果即为 $\mathbb{R}^n$ 的 Borel 子集上的 **Lebesgue 测度**.

回到 $(\Omega, \mathcal{F}, \mu)$ 为两个测度空间 $(X, \mathcal{A}, \mu)$ 和 $(Y, \mathcal{B}, \nu)$ 的乘积的情形，下面来证明：

定理 1.7.2. Fubini 定理. 若 $f \geq 0$ 或 $\int |f| \, \mathrm{d}\mu < \infty$，则

$$\int_X \int_Y f(x,y) \, \mu_2(\mathrm{d}y) \, \mu_1(\mathrm{d}x) = \int_{X \times Y} f \, \mathrm{d}\mu = \int_Y \int_X f(x,y) \, \mu_1(\mathrm{d}x) \, \mu_2(\mathrm{d}y). \tag{$*$}$$

证明. 下面证明第一个等式，第二个由对称性可得. 在断言第一个积分有意义之前，先证两个技术性结论：

- 当 x 固定时，$y \mapsto f(x,y)$ 是 $\mathcal{B}$-可测的.
- $x \mapsto \int_Y f(x,y) \, \mu_2(\mathrm{d}y)$ 是 $\mathcal{A}$-可测的.

首先考虑 $f = \mathbf{1}_E$ 的情形. 记 $E_x = \{y : (x,y) \in E\}$ 为 x 处的截面.

引理 1.7.3. 若 $E \in \mathcal{F}$，则 $E_x \in \mathcal{B}$.

证明. 由 $(E^c)_x = (E_x)^c$ 以及 $(\bigcup_i E_i)_x = \bigcup_i (E_i)_x$ 可知，若 $\mathcal{E}$ 是满足 $E_x \in \mathcal{B}$ 的集合 E 构成的集合族，则 $\mathcal{E}$ 是一个 σ-代数. 由于 $\mathcal{E}$ 包含所有矩形，结论得证. □

引理 1.7.4. 若 $E \in \mathcal{F}$，则 $g(x) \equiv \mu_2(E_x)$ 是 $\mathcal{A}$-可测的，且

$$\int_X g \, \mathrm{d}\mu_1 = \mu(E).$$

注意，满足结论的集合构成的集合族不一定是 σ-代数，因为 $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2) - \mu(E_1 \cap E_2)$. **Dynkin 的 π-λ 定理**（A.1.4）正是为这种情况量身打造的.

证明. 若结论对 E_n 成立且 $E_n \uparrow E$，则由定理 1.3.7 和单调收敛定理可知结论对 E 也成立. 由于 μ_1 和 μ_2 是 σ-有限的，故只需对 $E \subset F \times G$ 且 $\mu_1(F) < \infty$、$\mu_2(G) < \infty$ 证明结果，或者取 $\Omega = F \times G$ 并**不失一般性**假设 $\mu(\Omega) < \infty$. 记 $\mathcal{L}$ 为满足结论的集合 E 构成的集合族. 下面验证 $\mathcal{L}$ 是一个 λ-系. λ-系的性质 (i) 是显然的. (iii) 由证明的第一句话可得. 为验证 (ii)，注意

$$\mu_2((A - B)_x) = \mu_2(A_x - B_x) = \mu_2(A_x) - \mu_2(B_x),$$

对 x 积分即得第二个结论. 由于 $\mathcal{L}$ 包含生成 $\mathcal{F}$ 的矩形（一个 π-系），故由 π-λ 定理即得所需结果. □

下面通过四种逐次更一般的情形来证明定理 1.7.2.

情形 1. 若 $E \in \mathcal{F}$ 且 $f = \mathbf{1}_E$，则 $(*)$ 由引理 1.7.4 可得.

情形 2. 由于每个积分关于 f 是线性的，故 $(*)$ 对简单函数成立.

情形 3. 若 $f \geq 0$，记 $f_n(x) = ([2^n f(x)]/2^n) \wedge n$，其中 $[x] =$ 不超过 x 的最大整数，则 f_n 是简单函数且 $f_n \uparrow f$，故由单调收敛定理可知 $(*)$ 对所有 $f \geq 0$ 成立.

情形 4. 一般情形由 $f(x) = f(x)^+ - f(x)^-$ 并对 f^+、f^- 和 $|f|$ 应用情形 3 即得. □

为说明定理 1.7.2 中各假设的必要性，下面给出一些结论失效的例子.

例题 1.7.5. 设 $X = Y = \{1, 2, \ldots\}$，$\mathcal{A} = \mathcal{B} =$ 所有子集，$\mu_1 = \mu_2 =$ 计数测度. 对 $m \geq 1$，令 $f(m, m) = 1$ 且 $f(m+1, m) = -1$，并对其他情形令 $f(m, n) = 0$. 断言

$$\sum_m \sum_n f(m,n) = 1 \quad \text{但} \quad \sum_n \sum_m f(m,n) = 0.$$

一图胜千言:

$$\begin{array}{cccccc} \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 1 & -1 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \cdots \end{array}$$

若先对列求和，第一列给出 1，其余各列给出 0；而若先对行求和，每一行都给出 0.

例题 1.7.6. 设 $X = (0,1)$，$Y = (1, \infty)$，均配备 Borel 集和 Lebesgue 测度. 令 $f(x,y) = e^{-xy} - 2e^{-2xy}$.

$$\int_0^1 \int_1^\infty f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 x^{-1}(e^{-x} - e^{-2x})\,\mathrm{d}x > 0,$$

$$\int_1^\infty \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_1^\infty y^{-1}(e^{-2y} - e^{-y})\,\mathrm{d}y < 0.$$

下一个例子说明 μ_1 和 μ_2 为何必须是 σ-有限的.

例题 1.7.7. 设 $X = (0,1)$，$\mathcal{A} =$Borel 集，$\mu_1 =$Lebesgue 测度. 设 $Y = (0,1)$，$\mathcal{B} =$ 所有子集，$\mu_2 =$ 计数测度. 令 $f(x,y) = 1$ 若 $x = y$，否则为 0.

$$\int_Y f(x,y)\,\mu_2(\mathrm{d}y) = 1 \quad \text{对所有 } x \text{ 成立，故} \quad \int_X \int_Y f(x,y)\,\mu_2(\mathrm{d}y)\,\mu_1(\mathrm{d}x) = 1,$$

$$\int_X f(x,y)\,\mu_1(\mathrm{d}x) = 0 \quad \text{对所有 } y \text{ 成立，故} \quad \int_Y \int_X f(x,y)\,\mu_1(\mathrm{d}x)\,\mu_2(\mathrm{d}y) = 0.$$

最后一个例子表明可测性是重要的，或者集合论的某些公理并不像看起来那样无害.

例题 1.7.8. 由选择公理和连续统假设，可以在 $(0,1)$ 上定义一个序关系 $<_0$，使得对每个 y，$\{x : x <_0 y\}$ 是可数的. 设 $X = Y = (0,1)$，$\mathcal{A} = \mathcal{B}$ =Borel 集，$\mu_1 = \mu_2$ =Lebesgue 测度. 令 $f(x,y) = 1$ 若 $x <_0 y$，否则为 0. 由于 $\{x : x <_0 y\}$ 和 $\{y : x <_0 y\}^c$ 是可数的，

$$\int_X f(x,y)\,\mu_1(\mathrm{d}x) = 0 \quad \text{对所有 } y \text{ 成立},$$

$$\int_Y f(x,y)\,\mu_2(\mathrm{d}y) = 1 \quad \text{对所有 } x \text{ 成立}.$$

习题

习题 1.7.1. 若 $\int_X \int_Y |f(x,y)|\,\mu_2(\mathrm{d}y)\,\mu_1(\mathrm{d}x) < \infty$，则

$$\int_X \int_Y f(x,y)\,\mu_2(\mathrm{d}y)\,\mu_1(\mathrm{d}x) = \int_{X\times Y} f\,\mathrm{d}(\mu_1 \times \mu_2) = \int_Y \int_X f(x,y)\,\mu_1(\mathrm{d}x)\,\mu_2(\mathrm{d}y).$$

推论. 设 $X = \{1,2,\ldots\}$，$\mathcal{A} = X$ 的所有子集，$\mu_1 =$ 计数测度.

若 $\sum_n \int |f_n|\,\mathrm{d}\mu < \infty$，则 $\int \sum_n f_n\,\mathrm{d}\mu = \sum_n \int f_n\,\mathrm{d}\mu$.

习题 1.7.2. 设 $g \geq 0$ 是 $(X,\mathcal{A},\mu)$ 上的可测函数. 利用定理 1.7.2 推出

$$\int_X g\,\mathrm{d}\mu = (\mu\times\lambda)(\{(x,y) : 0 \leq y < g(x)\}) = \int_0^\infty \mu(\{x : g(x) > y\})\,\mathrm{d}y.$$

习题 1.7.3. 设 F、G 为 Stieltjes 测度函数，μ、ν 为 $(\mathbb{R},\mathcal{R})$ 上的相应测度. 证明

(i) $\int_{(a,b]} \{F(y) - F(a)\}\,\mathrm{d}G(y) = (\mu\times\nu)(\{(x,y) : a < x \leq y \leq b\})$

(ii) $\int_{(a,b]} F(y)\,\mathrm{d}G(y) + \int_{(a,b]} G(y)\,\mathrm{d}F(y)$
$= F(b)G(b) - F(a)G(a) + \sum_{x\in(a,b]} \mu(\{x\})\nu(\{x\})$

(iii) 若 $F = G$ 连续，则 $\int_{(a,b]} 2F(y)\,\mathrm{d}F(y) = F^2(b) - F^2(a)$.

为说明 (ii) 中需要第二项，令 $F(x) = G(x) = \mathbf{1}_{[0,\infty)}(x)$ 且 $a < 0 < b$.

习题 1.7.4. 设 μ 为 $\mathbb{R}$ 上的有限测度且 $F(x) = \mu((-\infty,x])$. 证明

$$\int (F(x+c) - F(x))\,\mathrm{d}x = c\mu(\mathbb{R}).$$

习题 1.7.5. 证明 $e^{-xy}\sin x$ 在区域 $0 < x < a$、$0 < y$ 上可积. 按两种次序计算重积分得到：

$$\int_0^a \frac{\sin x}{x}\,\mathrm{d}x = \arctan(a) - (\cos a)\int_0^\infty \frac{e^{-ay}}{1+y^2}\,\mathrm{d}y - (\sin a)\int_0^\infty \frac{ye^{-ay}}{1+y^2}\,\mathrm{d}y,$$

并将 $1+y^2$ 替换为 1 以推出对 $a \geq 1$ 有

$$\left| \int_0^a \frac{\sin x}{x} \, \mathrm{d}x - \arctan(a) \right| \leq \frac{2}{a}.$$

第二章
大数定律

第 1 节　独立性

测度论至此告一段落，概率论则始于**独立性**的定义. 本节先给出一些读者理应熟悉的定义，再逐步引入适合当前框架的表述.

事件 A 和 B **独立**，如果 $P(A\cap B)=P(A)P(B)$.

随机变量 X 和 Y **独立**，如果对所有 $C,D\in\mathcal{R}$，

$$P(X\in C,Y\in D)=P(X\in C)P(Y\in D)$$

即事件 $A=\{X\in C\}$ 和 $B=\{Y\in D\}$ 独立.

σ-域 $\mathcal{F}$ 和 $\mathcal{G}$ **独立**，如果对任意 $A\in\mathcal{F}$ 和 $B\in\mathcal{G}$，事件 A 和 B 独立.

下面的结果表明，第二个定义是第三个定义的特例.

定理 2.1.1. (i) 若 X 和 Y 独立，则 $\sigma(X)$ 和 $\sigma(Y)$ 独立. (ii) 反之，若 $\mathcal{F}$ 和 $\mathcal{G}$ 独立，$X\in\mathcal{F}$ 且 $Y\in\mathcal{G}$，则 X 和 Y 独立.

证明. (i) 若 $A\in\sigma(X)$，则由 $\sigma(X)$ 的定义可知 $A=\{X\in C\}$ 对某个 $C\in\mathcal{R}$. 同理若 $B\in\sigma(Y)$，则 $B=\{Y\in D\}$ 对某个 $D\in\mathcal{R}$，于是利用这些事实及 X 和 Y 的独立性，

$$P(A\cap B)=P(X\in C,Y\in D)=P(X\in C)P(Y\in D)=P(A)P(B)$$

(ii) 反之，若 $X\in\mathcal{F}$，$Y\in\mathcal{G}$ 且 $C,D\in\mathcal{R}$，则由可测性的定义可知 $\{X\in C\}\in\mathcal{F}$，$\{Y\in D\}\in\mathcal{G}$. 由于 $\mathcal{F}$ 和 $\mathcal{G}$ 独立，故 $P(X\in C,Y\in D)=P(X\in C)P(Y\in D)$. □

第一个定义又是第二个定义的特例.

定理 2.1.2. (i) 若 A 和 B 独立，则 A^c 和 B、A 和 B^c、A^c 和 B^c 也都独立. (ii) 反之，事件 A 和 B 独立当且仅当其**示性随机变量** $\mathbf{1}_A$ 和 $\mathbf{1}_B$ 独立.

证明. (i) 由 $P(B) = P(B)$ 减去 $P(A \cap B) = P(A)P(B)$ 得 $P(A^c \cap B) = P(A^c)P(B)$. 第二和第三个结论通过对独立事件对 (B, A) 和 (A, B^c) 应用第一个结论得到.

(ii) 若 $C, D \in \mathcal{R}$，则 $\{\mathbf{1}_A \in C\} \in \{\emptyset, A, A^c, \Omega\}$ 且 $\{\mathbf{1}_B \in D\} \in \{\emptyset, B, B^c, \Omega\}$，因此有 16 种情形需要验证. 当涉及的集合为 $\emptyset$ 或 Ω 时等式成立，故只需考虑四种情形，而这些都由 (i) 涵盖. □

由于第一个定义是第二个定义的特例，第二个定义又是第三个定义的特例，在讨论多个对象的独立性时，我们按相反的顺序展开. 首先归结为有限多个对象的情形. 称无限多个对象（σ-域、随机变量或集合）**独立**，如果其中每个有限子集都独立.

σ-域 $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ **独立**，如果对任意 $A_i \in \mathcal{F}_i$ $(i = 1, \ldots, n)$，有

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

随机变量 $X_1, \ldots, X_n$ **独立**，如果对任意 $B_i \in \mathcal{R}$ $(i = 1, \ldots, n)$，有

$$P\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} P(X_i \in B_i)$$

集合 $A_1, \ldots, A_n$ **独立**，如果对任意 $I \subset \{1, \ldots, n\}$，有

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

从表面看，最后一个定义似乎与其他两个不一致. 然而，稍加思考就会发现，若示性变量 $\mathbf{1}_{A_i}$ $(1 \le i \le n)$ 独立，且对 $i \in I$ 取 $B_i = \{1\}$，对 $i \notin I$ 取 $B_i = \mathbb{R}$，则定义中的条件可得. 反之，

定理 2.1.3. 设 $A_1, A_2, \ldots, A_n$ 独立. (i) $A_1^c, A_2, \ldots, A_n$ 独立；(ii) $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}$ 独立.

证明. (i) 取 $B_1 = A_1^c$ 且对 $i > 1$ 取 $B_i = A_i$. 若 $I \subset \{1, \ldots, n\}$ 不包含 1，则显然 $P(\bigcap_{i \in I} B_i) = \prod_{i \in I} P(B_i)$. 现设 $1 \in I$ 且令 $J = I - \{1\}$. 由 $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$ 减去 $P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$ 得 $P(A_1^c \cap \bigcap_{i \in J} A_i) = P(A_1^c) \prod_{i \in J} P(A_i)$.

(ii) 反复应用 (i) 可知，若 $B_i \in \{A_i, A_i^c\}$，则 $B_1, \ldots, B_n$ 独立. 于是若 $C_i \in \{A_i, A_i^c, \Omega\}$，则 $P(\bigcap_{i=1}^n C_i) = \prod_{i=1}^n P(C_i)$. 当某个 $C_i = \emptyset$ 时最后一个等式显然成立，注意到 $\mathbf{1}_{A_i} \in \{\emptyset, A_i, A_i^c, \Omega\}$ 即得所需结果. □

关于独立事件定义，首先应注意，仅假设对所有 $i \ne j$ 有 $P(A_i \cap A_j) = P(A_i)P(A_j)$ 是不够的. 具有后一性质的序列 $A_1, \ldots, A_n$ 称为**两两独立**. 显然独立事件是两两独立的. 下例表明其逆不成立.

例题 2.1.4. 设 X_1, X_2, X_3 为独立随机变量，满足

$$P(X_i = 0) = P(X_i = 1) = 1/2$$

令 $A_1 = \{X_2 = X_3\}$，$A_2 = \{X_3 = X_1\}$，$A_3 = \{X_1 = X_2\}$. 这些事件两两独立，因为当 $i \neq j$ 时

$$P(A_i \cap A_j) = P(X_1 = X_2 = X_3) = 1/4 = P(A_i)P(A_j)$$

但它们不独立，因为

$$P(A_1 \cap A_2 \cap A_3) = 1/4 \neq 1/8 = P(A_1)P(A_2)P(A_3)$$

为证明随机变量 X 和 Y 独立，需验证对所有 Borel 集 A 和 B 有 $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$. 由于 Borel 集数量众多，下一个主题是

2.1.1 独立的充分条件

主要结果为定理 2.1.7. 为陈述该结果，需要推广前述所有定义的概念.

集合族 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n \subset \mathcal{F}$ 称为**独立**，如果对任意 $A_i \in \mathcal{A}_i$ 及 $I \subset \{1, \ldots, n\}$，有 $P(\bigcap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$. 若每个集合族都是单点集（即 $\mathcal{A}_i = \{A_i\}$），则此定义退化为集合的独立定义.

引理 2.1.5. 不失一般性，可设每个 $\mathcal{A}_i$ 包含 Ω. 在此情形下，条件等价于

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

对任意 $A_i \in \mathcal{A}_i$ 成立，因为对 $i \notin I$ 可令 $A_i = \Omega$.

证明. 若 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ 独立且 $\bar{\mathcal{A}}_i = \mathcal{A}_i \cup \{\Omega\}$，则 $\bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2, \ldots, \bar{\mathcal{A}}_n$ 独立，因为若 $A_i \in \bar{\mathcal{A}}_i$ 且 $I = \{j : A_j = \Omega\}$，则 $\bigcap_i A_i = \bigcap_{i\in I} A_i$. □

定理 2.1.7 的证明基于 Dynkin 的 π-λ 定理. 为陈述该结果，需要两个定义. 称 $\mathcal{A}$ 为 **π-系**，如果它对**交**封闭，即若 $A, B \in \mathcal{A}$ 则 $A \cap B \in \mathcal{A}$. 称 $\mathcal{L}$ 为 **λ-系**，如果：(i) $\Omega \in \mathcal{L}$；(ii) 若 $A, B \in \mathcal{L}$ 且 $A \subset B$ 则 $B - A \in \mathcal{L}$；(iii) 若 $A_n \in \mathcal{L}$ 且 $A_n \uparrow A$ 则 $A \in \mathcal{L}$.

定理 2.1.6. π-λ 定理. 若 $\mathcal{P}$ 是 π-系且 $\mathcal{L}$ 是包含 $\mathcal{P}$ 的 λ-系，则 $\sigma(\mathcal{P}) \subset \mathcal{L}$.

证明见附录 A.1 节.

定理 2.1.7. 设 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ 独立且每个 $\mathcal{A}_i$ 是 π-系，则 $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \ldots, \sigma(\mathcal{A}_n)$ 独立.

证明. 令 $A_2,\ldots,A_n$ 为集合, 其中 $A_i\in\mathcal{A}_i$, 令 $F=A_2\cap\cdots\cap A_n$ 且 $\mathcal{L}=\{A:P(A\cap F)=P(A)P(F)\}$. 由于 $P(\Omega\cap F)=P(\Omega)P(F)$, 有 $\Omega\in\mathcal{L}$.

为验证 λ-系定义中的 (ii), 注意若 $A,B\in\mathcal{L}$ 且 $A\subset B$, 则 $(B-A)\cap F=(B\cap F)-(A\cap F)$. 于是利用定理 1.1.1 中的 (i)、$A,B\in\mathcal{L}$ 的事实以及再次使用定理 1.1.1 中的 (i):

$$\begin{aligned}P((B-A)\cap F)&=P(B\cap F)-P(A\cap F)=P(B)P(F)-P(A)P(F)\\&=\{P(B)-P(A)\}P(F)=P(B-A)P(F)\end{aligned}$$

故 $B-A\in\mathcal{L}$.

为验证 (iii), 令 $B_k\in\mathcal{L}$ 且 $B_k\uparrow B$, 注意 $(B_k\cap F)\uparrow(B\cap F)$, 于是利用定理 1.1.1 中的 (iii)、$B_k\in\mathcal{L}$ 的事实以及再次使用定理 1.1.1 中的 (iii):

$$P(B\cap F)=\lim_k P(B_k\cap F)=\lim_k P(B_k)P(F)=P(B)P(F)$$

现应用 π-λ 定理得 $\mathcal{L}\supset\sigma(\mathcal{A}_1)$. 由此, 若 $A_1\in\sigma(\mathcal{A}_1)$ 且 $A_i\in\mathcal{A}_i$ 对 $2\le i\le n$, 则

$$P\left(\bigcap_{i=1}^n A_i\right)=P(A_1)P\left(\bigcap_{i=2}^n A_i\right)=\prod_{i=1}^n P(A_i)$$

现利用引理 2.1.5, 有:

$(*)$ 若 $\mathcal{A}_1,\mathcal{A}_2,\ldots,\mathcal{A}_n$ 独立, 则 $\sigma(\mathcal{A}_1),\mathcal{A}_2,\ldots,\mathcal{A}_n$ 独立.

将 $(*)$ 应用于 $\mathcal{A}_2,\ldots,\mathcal{A}_n,\sigma(\mathcal{A}_1)$ (它们是独立的, 因为独立的定义不随顺序改变), 这表明 $\sigma(\mathcal{A}_2),\mathcal{A}_3,\ldots,\mathcal{A}_n,\sigma(\mathcal{A}_1)$ 独立, 经 n 次迭代后即得所需结果. □

注释. 读者应注意, 证明 $A,B\in\mathcal{L}$ 时 $A\cap B\in\mathcal{L}$ 或 $A\cup B\in\mathcal{L}$ 并不容易, 但验证 $A,B\in\mathcal{L}$ 且 $A\subset B$ 时 $B-A\in\mathcal{L}$ 却很简单.

建立了定理 2.1.7 后, 下面给出几个推论.

定理 2.1.8. 为使 $X_1,\ldots,X_n$ 独立, 充分条件是对所有 $x_1,\ldots,x_n\in(-\infty,\infty]$,

$$P(X_1\le x_1,\ldots,X_n\le x_n)=\prod_{i=1}^n P(X_i\le x_i)$$

证明. 令 $\mathcal{A}_i=$ 形如 $\{X_i\le x_i\}$ 的集合族. 由于

$$\{X_i\le x\}\cap\{X_i\le y\}=\{X_i\le x\wedge y\},$$

其中 $(x\wedge y)_i=x_i\wedge y_i=\min\{x_i,y_i\}$, 故 $\mathcal{A}_i$ 是 π-系. 由于允许 $x_i=\infty$, 有 $\Omega\in\mathcal{A}_i$. 习题 1.3.1 蕴涵 $\sigma(\mathcal{A}_i)=\sigma(X_i)$, 故结果由定理 2.1.7 得出. □

最后一个结果用分布函数表达了随机变量的独立性. 习题 2.1.1 和 2.1.2 讨论密度函数和离散随机变量.

下一个目标是证明不相交独立随机变量集合的函数仍独立. 确切表述见定理 2.1.10. 首先证明 σ-域的类似结果.

定理 2.1.9. 设 $\mathcal{F}_{i,j}$ $(1 \le i \le n,\ 1 \le j \le m(i))$ 独立，令 $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{i,j})$，则 $\mathcal{G}_1, \ldots, \mathcal{G}_n$ 独立.

证明. 令 $\mathcal{A}_i$ 为形如 $\bigcap_j A_{i,j}$ 的集合族，其中 $A_{i,j} \in \mathcal{F}_{i,j}$. $\mathcal{A}_i$ 是包含 Ω 的 π-系且包含 $\bigcup_j \mathcal{F}_{i,j}$，故定理 2.1.7 蕴涵 $\sigma(\mathcal{A}_i) = \mathcal{G}_i$ 独立. □

定理 2.1.10. 若对 $1 \le i \le n,\ 1 \le j \le m(i),\ X_{i,j}$ 独立且 $f_i : \mathbb{R}^{m(i)} \to \mathbb{R}$ 可测，则 $f_i(X_{i,1}, \ldots, X_{i,m(i)})$ 独立.

证明. 令 $\mathcal{F}_{i,j} = \sigma(X_{i,j})$ 且 $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{i,j})$. 由于 $f_i(X_{i,1}, \ldots, X_{i,m(i)}) \in \mathcal{G}_i$，所需结果由定理 2.1.9 和习题 2.1.1 得出. □

定理 2.1.10 的一个具体特例是：若 $X_1, \ldots, X_n$ 独立，则 $X = X_1$ 和 $Y = X_2 \cdots X_n$ 独立. 稍后研究独立随机变量之和 $S_m = X_1 + \cdots + X_m$ 时，将用定理 2.1.10 得出：若 $m < n$，则 $S_n - S_m$ 与事件 $\{\max_{1 \le k \le m} S_k > x\}$ 的示性函数独立.

2.1.2 独立性、分布与期望

下一个目标是获得独立随机变量分布和期望的公式.

定理 2.1.11. 设 $X_1, \ldots, X_n$ 为独立随机变量且 X_i 具有分布 μ_i，则 $(X_1, \ldots, X_n)$ 具有分布 $\mu_1 \times \cdots \times \mu_n$.

证明. 利用 (i) $A_1 \times \cdots \times A_n$ 的定义、(ii) 独立性、(iii) μ_i 以及 (iv) $\mu_1 \times \cdots \times \mu_n$ 的定义，

$$
\begin{aligned}
&P((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) = P(X_1 \in A_1, \ldots, X_n \in A_n) \\
&= \prod_{i=1}^{n} P(X_i \in A_i) = \prod_{i=1}^{n} \mu_i(A_i) = \mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n)
\end{aligned}
$$

最后一个公式表明 $(X_1, \ldots, X_n)$ 的分布与测度 $\mu_1 \times \cdots \times \mu_n$ 在形如 $A_1 \times \cdots \times A_n$ 的集合上一致，而这些集合构成生成 $\mathbb{R}^n$ 的 π-系. 故定理 2.1.6 蕴涵它们必须一致. □

定理 2.1.12. 设 X 和 Y 独立且具有分布 μ 和 ν. 若 $h : \mathbb{R}^2 \to \mathbb{R}$ 为可测函数且 $h \ge 0$ 或 $E|h(X,Y)| < \infty$，则

$$
Eh(X,Y) = \iint h(x,y)\, \mu(\mathrm{d}x)\, \nu(\mathrm{d}y)
$$

特别地，若 $h(x,y) = f(x)g(y)$，其中 $f, g : \mathbb{R} \to \mathbb{R}$ 为可测函数且 $f, g \ge 0$ 或 $E|f(X)|$ 和 $E|g(Y)| < \infty$，则

$$
Ef(X)g(Y) = Ef(X) \cdot Eg(Y)
$$

证明. 利用定理 1.6.9 然后应用 Fubini 定理（定理 1.7.2）有

$$Eh(X,Y)=\int_{\mathbb{R}^2} h\,\mathrm{d}(\mu\times\nu)=\iint h(x,y)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y)$$

为证明第二个结果，从 $f,g\geq 0$ 的情形开始. 此时，利用第一个结果、$g(y)$ 不依赖于 x 的事实以及两次应用定理 1.6.9 得

$$\begin{aligned}Ef(X)g(Y)&=\iint f(x)g(y)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y)=\int g(y)\int f(x)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y)\\&=\int Ef(X)g(y)\,\nu(\mathrm{d}y)=Ef(X)Eg(Y)\end{aligned}$$

将非负 f 和 g 的结果应用于 $|f|$ 和 $|g|$，表明 $E|f(X)g(Y)|=E|f(X)|E|g(Y)|<\infty$，于是可重复上述论证证明所需结果. □

由定理 2.1.12，只需一小步即可得

定理 2.1.13. 若 $X_1,\ldots,X_n$ 独立且满足 (a) 对所有 i 有 $X_i\geq 0$，或 (b) 对所有 i 有 $E|X_i|<\infty$，则

$$E\left(\prod_{i=1}^n X_i\right)=\prod_{i=1}^n EX_i$$

即左边的期望存在且等于右边给出的值.

证明. 由定理 2.1.10，$X=X_1$ 和 $Y=X_2\cdots X_n$ 独立，故取 $f(x)=|x|$ 和 $g(y)=|y|$ 有 $E|X_1\cdots X_n|=E|X_1|E|X_2\cdots X_n|$，由归纳法可知对 $1\leq m\leq n$ 有

$$E|X_m\cdots X_n|=\prod_{i=m}^n E|X_i|$$

若 $X_i\geq 0$，则 $|X_i|=X_i$，所需结果由 $m=1$ 的特殊情形得出. 为证明一般情形，注意 $m=2$ 的特殊情形蕴涵 $E|Y|=E|X_2\cdots X_n|<\infty$，故利用定理 2.1.12 取 $f(x)=x$ 和 $g(y)=y$ 表明 $E(X_1\cdots X_n)=EX_1\cdot E(X_2\cdots X_n)$，所需结果由归纳法得出. □

例题 2.1.14. 可能出现 $E(XY)=EX\cdot EY$ 但变量不独立的情形. 设 X 和 Y 的联合分布由下表给出

$X\backslash Y$	1	0	−1
1	0	a	0
0	b	c	b
−1	0	a	0

其中 $a,b>0$，$c\geq 0$，且 $2a+2b+c=1$. 表中安排使得 $XY\equiv 0$. 对称性蕴涵 $EX=0$ 和 $EY=0$，故 $E(XY)=0=EXEY$. 随机变量不独立，因为

$$P(X=1,Y=1)=0<ab=P(X=1)P(Y=1)$$

两个随机变量 X 和 Y 满足 $EX^2,EY^2<\infty$ 且 $EXY=EXEY$ 称为**不相关的**. 需要有限的二阶矩以保证由 Cauchy-Schwarz 不等式知 $E|XY|<\infty$.

2.1.3 独立随机变量之和

定理 2.1.15. 若 X 和 Y 独立，$F(x)=P(X\leq x)$，$G(y)=P(Y\leq y)$，则

$$P(X+Y\leq z)=\int F(z-y)\,\mathrm{d}G(y)$$

右边的积分称为 F 和 G 的**卷积**，记为 $F*G(z)$. $\mathrm{d}G(y)$ 的含义将在证明中解释.

证明. 令 $h(x,y)=\mathbf{1}_{(x+y\leq z)}$. 令 μ 和 ν 为分别以 F 和 G 为分布函数的概率测度. 由于对固定的 y，

$$\int h(x,y)\,\mu(\mathrm{d}x)=\int \mathbf{1}_{(-\infty,z-y]}(x)\,\mu(\mathrm{d}x)=F(z-y)$$

利用定理 2.1.12 得

$$P(X+Y\leq z)=\iint \mathbf{1}_{(x+y\leq z)}\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y)=\int F(z-y)\,\nu(\mathrm{d}y)=\int F(z-y)\,\mathrm{d}G(y)$$

最后一个等式只是记号改变：将 $\mathrm{d}G(y)$ 视为 “关于以 G 为分布函数的测度 ν 积分” 的简写. □

为处理具体例子，需要定理 2.1.15 的一个特殊情形.

定理 2.1.16. 设 X 具有密度 f，Y 具有分布函数 G，且二者独立. 则 $X+Y$ 具有密度

$$h(x)=\int f(x-y)\,\mathrm{d}G(y)$$

当 Y 具有密度 g 时，最后一个公式可写为

$$h(x)=\int f(x-y)g(y)\,\mathrm{d}y$$

证明. 由定理 2.1.15、密度函数的定义以及 Fubini 定理（定理 1.7.2），后者可应用因为所有量非负，

得

$$P(X+Y\le z)=\int F(z-y)\,\mathrm{d}G(y)=\int\int_{-\infty}^{z-y}f(x-y)\,\mathrm{d}x\,\mathrm{d}G(y)=\int_{-\infty}^{z}\int f(x-y)\,\mathrm{d}G(y)\,\mathrm{d}x$$

最后一个等式表明 $X+Y$ 具有密度 $h(x)=\int f(x-y)\,\mathrm{d}G(y)$. 回忆定理 2.1.15 证明中 $\mathrm{d}G(y)$ 的含义并利用习题 1.6.8，当 G 有密度 g 时由第一个公式可得第二个公式. □

定理 2.1.16 加上一些繁琐的计算可处理两个标准例子. 这些事实应是本科概率论中熟悉的.

例题 2.1.17. 参数为 α 和 λ 的**伽马密度**为

$$f(x)=\begin{cases}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}/\Gamma(\alpha) & x\ge 0\\ 0 & x<0\end{cases}$$

其中 $\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}\mathrm{d}x$.

定理 2.1.18. 若 $X=\mathrm{gamma}(\alpha,\lambda)$ 和 $Y=\mathrm{gamma}(\beta,\lambda)$ 独立，则 $X+Y$ 服从 $\mathrm{gamma}(\alpha+\beta,\lambda)$ 分布. 因此若 $X_1,\dots,X_n$ 为独立的 $\mathrm{exponential}(\lambda)$ 随机变量，则 $X_1+\cdots+X_n$ 服从 $\mathrm{gamma}(n,\lambda)$ 分布.

证明. 记 $f_{X+Y}(z)$ 为 $X+Y$ 的密度函数，利用定理 2.1.16

$$f_{X+Y}(x)=\int_0^x\frac{\lambda^{\alpha}(x-y)^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda(x-y)}\frac{\lambda^{\beta}y^{\beta-1}}{\Gamma(\beta)}e^{-\lambda y}\,\mathrm{d}y=\frac{\lambda^{\alpha+\beta}e^{-\lambda x}}{\Gamma(\alpha)\Gamma(\beta)}\int_0^x(x-y)^{\alpha-1}y^{\beta-1}\,\mathrm{d}y$$

故只需证明积分等于 $x^{\alpha+\beta-1}\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. 为此，先作变量替换 $y=xu$，$\mathrm{d}y=x\,\mathrm{d}u$ 得

$$x^{\alpha+\beta-1}\int_0^1(1-u)^{\alpha-1}u^{\beta-1}\,\mathrm{d}u=\int_0^x(x-y)^{\alpha-1}y^{\beta-1}\,\mathrm{d}y \tag{2.1.1}$$

此时有两种方式完成证明.

简洁的解法是注意到已证明密度 $f_{X+Y}(x)=c_{\alpha,\beta}e^{-\lambda x}\lambda^{\alpha+\beta}x^{\alpha+\beta-1}$，其中

$$c_{\alpha,\beta}=\frac{1}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1(1-u)^{\alpha-1}u^{\beta-1}\,\mathrm{d}u$$

只有一种归一化常数 $c_{\alpha,\beta}$ 可使之成为概率分布，故回忆**贝塔分布**的定义，必须有 $c_{\alpha,\beta}=1/\Gamma(\alpha+\beta)$.

对记不住贝塔定义的人，较不优雅的办法是用微积分证明最后一个等式. 两边乘以 e^{-x}，从 0

到 ∞ 积分，然后对右边应用 Fubini 定理得

$$\Gamma(\alpha+\beta)\int_0^1 (1-u)^{\alpha-1}u^{\beta-1}\,\mathrm{d}u = \int_0^\infty\int_0^x y^{\beta-1}e^{-y}(x-y)^{\alpha-1}e^{-(x-y)}\,\mathrm{d}y\,\mathrm{d}x$$
$$= \int_0^\infty y^{\beta-1}e^{-y}\int_y^\infty (x-y)^{\alpha-1}e^{-(x-y)}\,\mathrm{d}x\,\mathrm{d}y = \Gamma(\alpha)\Gamma(\beta)$$

即得第一个结果. 第二个结果由 gamma$(1,\lambda)$ 是参数为 λ 的指数分布以及归纳法得出. □

例题 2.1.19. 正态分布. 在例 1.6.11 中，我们引入了均值为 μ、方差为 a 的**正态密度**，

$$(2\pi a)^{-1/2}\exp(-(x-\mu)^2/2a)$$

定理 2.1.20. 若 $X=\text{normal}(\mu,a)$ 和 $Y=\text{normal}(\nu,b)$ 独立，则 $X+Y=\text{normal}(\mu+\nu,a+b)$.

证明. 只需对 $\mu=\nu=0$ 证明结果. 设 $Y_1=\text{normal}(0,a)$ 和 $Y_2=\text{normal}(0,b)$. 由定理 2.1.16，

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}}\int e^{-x^2/2a}e^{-(z-x)^2/2b}\,\mathrm{d}x$$

略去前面的常数，积分可改写为

$$\int \exp\left(\frac{-bx^2-ax^2+2axz-az^2}{2ab}\right)\mathrm{d}x$$
$$= \int \exp\left(\frac{-a+b}{2ab}\left\{x^2-\frac{2a}{a+b}xz+\frac{a}{a+b}z^2\right\}\right)\mathrm{d}x$$
$$= \int \exp\left(\frac{-a+b}{2ab}\left\{\left(x-\frac{a}{a+b}z\right)^2+\frac{ab}{(a+b)^2}z^2\right\}\right)\mathrm{d}x$$

因为 $-\{a/(a+b)\}^2+\{a/(a+b)\}=ab/(a+b)^2$. 提出不依赖于 x 的项，最后一个积分

$$= \exp\left(-\frac{z^2}{2(a+b)}\right)\int \exp\left(\frac{-a+b}{2ab}\left(x-\frac{a}{a+b}z\right)^2\right)\mathrm{d}x = \exp\left(-\frac{z^2}{2(a+b)}\right)\sqrt{2\pi ab/(a+b)}$$

因为最后一个积分是参数为 $\mu=az/(a+b)$ 和 $\sigma^2=ab/(a+b)$ 的正态密度（缺省归一化常数）. 重新引入开始时略去的常数，

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}}\sqrt{2\pi ab/(a+b)}\exp\left(-\frac{z^2}{2(a+b)}\right)$$

□

2.1.4 构造独立随机变量

在能够研究独立随机变量之前，我们还需要回答的最后一个问题是：它们存在吗？（如果不存在，研究它们就毫无意义！）若给定有限个分布函数 F_i（$1 \le i \le n$），很容易构造独立随机变量 $X_1, \ldots, X_n$ 使得 $P(X_i \le x) = F_i(x)$. 令 $\Omega = \mathbb{R}^n$，$\mathcal{F} = \mathcal{R}^n$，$X_i(\omega_1, \ldots, \omega_n) = \omega_i$（$\omega \in \mathbb{R}^n$ 的第 i 个坐标），并令 P 为 $\mathbb{R}^n$ 上的测度，满足

$$P((a_1, b_1] \times \cdots \times (a_n, b_n]) = (F_1(b_1) - F_1(a_1)) \cdots (F_n(b_n) - F_n(a_n))$$

若 μ_i 为以 F_i 为分布函数的测度，则 $P = \mu_1 \times \cdots \times \mu_n$.

为构造具有给定分布函数的无限序列 $X_1, X_2, \ldots$，我们希望在无限乘积空间上执行上述构造

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \ldots) : \omega_i \in \mathbb{R}\} = \{\text{函数}\omega : \mathbb{N} \to \mathbb{R}\}$$

其中 $\mathbb{N} = \{1, 2, \ldots\}$，$\mathbb{N}$ 代表自然数. 定义 $X_i(\omega) = \omega_i$，并赋予 $\mathbb{R}^{\mathbb{N}}$ 乘积 σ-域 $\mathcal{R}^{\mathbb{N}}$，它由有限维集生成，即形如 $\{\omega : \omega_i \in B_i, 1 \le i \le n\}$ 的集合，其中 $B_i \in \mathcal{R}$. 如何对有限维集定义 P 是显然的. 为断言到 $\mathcal{R}^{\mathbb{N}}$ 的唯一延拓存在，我们使用定理 A.3.1：

定理 2.1.21. Kolmogorov 扩张定理. 设在 $(\mathbb{R}^n, \mathcal{R}^n)$ 上给定了概率测度 μ_n，它们相容，即

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

则存在唯一的概率测度 P 在 $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ 上满足

$$P(\omega : \omega_i \in (a_i, b_i], 1 \le i \le n) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

在后续讨论中，我们需要构造取值于其他可测空间 $(S, \mathcal{S})$ 的随机变量序列. 不幸的是，定理 2.1.21 对任意可测空间并不成立. 第一个反例（在不同空间 $\Omega_1 \times \Omega_2 \times \ldots$ 的无限乘积上）由 Andersen 和 Jessen（1948）构造.（见 Halmos（1950）p. 214 或 Neveu（1965）p. 84.）对所有空间 Ω_i 相同的例子见 Wegner（1973）. 幸运的是，存在一类空间足以处理我们所有的结果，且 Kolmogorov 定理在这类空间上的推广是平凡的.

若存在从 S 到 $\mathbb{R}$ 的一一映射 φ 使得 φ 和 φ^{-1} 均可测，则称 $(S, \mathcal{S})$ 为**好空间**.

这样的空间常称为标准 Borel 空间，但我们已经有太多东西以 Borel 命名了. 下述结果表明应用中出现的多数空间都是好空间.

定理 2.1.22. 若 S 是完备可分度量空间 M 的 Borel 子集，$\mathcal{S}$ 是 S 的 Borel 子集族，则 $(S, \mathcal{S})$ 是好空间.

证明. 先从特殊情形 $S=[0,1)^{\mathbb{N}}$ 及度量

$$\rho(x,y)=\sum_{n=1}^{\infty}|x_n-y_n|/2^n$$

开始. 若 $x=(x_1,x_2,x_3,\ldots)$，将每个分量展开为二进制 $x_j=.x_1^jx_2^jx_3^j\ldots$（取含无限多个 0 的展开）. 令

$$\varphi_o(x)=.x_1^1x_1^2x_2^1x_1^3x_2^2x_3^1x_1^4x_2^3x_3^2x_4^1\ldots$$

为处理一般情形，注意通过令

$$d(x,y)=\rho(x,y)/(1+\rho(x,y))$$

（更多细节见习题 2.1.3），可设度量满足对所有 x,y 有 $d(x,y)<1$. 令 $q_1,q_2,\ldots$ 为 S 中的可数稠密集. 令

$$\psi(x)=(d(x,q_1),d(x,q_2),\ldots)$$

$\psi:S\to[0,1)^{\mathbb{N}}$ 连续且一一. $\varphi_o\circ\psi$ 给出所需映射. □

注释. Caveat emptor. 上述证明在细节上有些简略. 更全面的讨论见 Dudley（1989）第 13.1 节. 那里的分析有一个有趣的推论：对于完备可分度量空间的 Borel 子集，连续统假设成立，即所有集合要么是有限的，要么是可数无限的，要么具有实数的势.

习题

习题 2.1.1. 设 $(X_1,\ldots,X_n)$ 具有密度 $f(x_1,x_2,\ldots,x_n)$，即

$$P((X_1,X_2,\ldots,X_n)\in A)=\int_A f(x)\,\mathrm{d}x,\quad A\in\mathbb{R}^n$$

若 $f(x)$ 可写为 $g_1(x_1)\cdots g_n(x_n)$，其中 $g_m\geq 0$ 可测，则 $X_1,X_2,\ldots,X_n$ 独立. 注意 g_m 不必是概率密度.

习题 2.1.2. 设 $X_1,\ldots,X_n$ 为取值于可数集 $S_1,\ldots,S_n$ 的随机变量. 则为使 $X_1,\ldots,X_n$ 独立，充分条件是对任意 $x_i\in S_i$，

$$P(X_1=x_1,\ldots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)$$

习题 2.1.3. 设 $\rho(x,y)$ 为度量. (i) 设 h 可微且 $h(0)=0$，$h'(x)>0$ 对 $x>0$，$h'(x)$ 在 $[0,\infty)$ 上递减. 则 $h(\rho(x,y))$ 是度量. (ii) $h(x)=x/(x+1)$ 满足 (i) 中的假设.

习题 2.1.4. 设 $\Omega=(0,1)$，$\mathcal{F}$ =Borel 集，P =Lebesgue 测度. $X_n(\omega)=\sin(2\pi n\omega)$，$n=1,2,\ldots$ 是不相关的但不是独立的.

习题 2.1.5. (i) 证明若 X 和 Y 独立且具有分布 μ 和 ν，则

$$P(X+Y=0)=\sum_y \mu(\{-y\})\nu(\{y\})$$

(ii) 由此推出若 X 具有连续分布则 $P(X=Y)=0$.

习题 2.1.6. 直接从定义证明若 X 和 Y 独立且 f 和 g 为可测函数，则 $f(X)$ 和 $g(Y)$ 独立.

习题 2.1.7. 设 $K\geq 3$ 为素数且设 X 和 Y 为独立随机变量，均服从 $\{0,1,\ldots,K-1\}$ 上的均匀分布. 对 $0\leq n<K$，令 $Z_n=X+nY \pmod K$. 证明 $Z_0,Z_1,\ldots,Z_{K-1}$ 是**两两独立**的，即每对都独立. 它们不独立，因为若知道其中两个变量的值，就知道所有变量的值.

习题 2.1.8. 找出四个取值于 $\{-1,1\}$ 的随机变量，使得任意三个独立但所有四个不独立. 提示：考虑独立随机变量的乘积.

习题 2.1.9. 设 $\Omega=\{1,2,3,4\}$，$\mathcal{F}=\Omega$ 的所有子集，$P(\{i\})=1/4$. 给出两个集合族 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 的例子，它们是独立的，但其生成的 σ-域不独立.

习题 2.1.10. 证明若 X 和 Y 是独立的整数值随机变量，则

$$P(X+Y=n)=\sum_m P(X=m)P(Y=n-m)$$

习题 2.1.11. 在例 1.6.13 中，我们引入了参数为 λ 的**泊松分布**，由 $P(Z=k)=e^{-\lambda}\lambda^k/k!$ 对 $k=0,1,2,\ldots$ 给出. 利用上一习题证明若 $X=\text{Poisson}(\lambda)$ 和 $Y=\text{Poisson}(\mu)$ 独立，则 $X+Y=\text{Poisson}(\lambda+\mu)$.

习题 2.1.12. 称 X 服从 Binomial(n,p) 分布，若

$$P(X=m)=\binom{n}{m}p^m(1-p)^{n-m}$$

(i) 证明若 $X=\text{Binomial}(n,p)$ 和 $Y=\text{Binomial}(m,p)$ 独立，则 $X+Y=\text{Binomial}(n+m,p)$. (ii) 看例 1.6.12 并用归纳法得出 n 个独立 Bernoulli(p) 随机变量之和服从 Binomial(n,p) 分布.

习题 2.1.13. $X+Y$ 的分布可以是 $F*G$ 但随机变量不独立，这并不奇怪. 设 $X,Y\in\{0,1,2\}$ 且每个值以概率 $1/3$ 取得. (a) 设 X 和 Y 独立，求 $X+Y$ 的分布. (b) 找出所有使 $X+Y$ 分布与 (a) 答案相同的联合分布 (X,Y).

习题 2.1.14. 设 $X, Y \geq 0$ 独立且具有分布函数 F 和 G. 求 XY 的分布函数.

习题 2.1.15. 若我们想要无限次抛硬币的序列，不必使用 Kolmogorov 定理. 设 Ω 为配备 Borel 集 $\mathcal{F}$ 和 Lebesgue 测度 P 的单位区间 $(0,1)$. 令 $Y_n(\omega) = 1$ 若 $[2^n\omega]$ 为奇数，$Y_n(\omega) = 0$ 若 $[2^n\omega]$ 为偶数. 证明 $Y_1, Y_2, \ldots$ 独立且 $P(Y_k = 0) = P(Y_k = 1) = 1/2$.

第 2 节 弱大数定律

本节将证明若干" 弱大数定律". 首先定义定理结论中出现的收敛模式. 称 Y_n **依概率收敛**于 Y，如果对任意 $\varepsilon > 0$，当 $n \to \infty$ 时，$P(|Y_n - Y| > \varepsilon) \to 0$.

2.2.1 L^2 弱大数定律

第一组弱大数定律通过计算方差并运用 Chebyshev 不等式得到. 将例 2.1.14 中对两个随机变量的定义加以推广，称满足 $EX_i^2 < \infty$ 的随机变量族 $X_i, i \in I$ 是**不相关的**，如果

$$E(X_iX_j) = EX_i \cdot EX_j, \quad \text{只要} i \neq j.$$

关于不相关随机变量的弱大数定律（即定理 2.2.3）的关键在于下述结果.

定理 2.2.1. 设 $X_1, \ldots, X_n$ 满足 $E(X_i^2) < \infty$ 且不相关. 则

$$\text{var}(X_1 + \cdots + X_n) = \text{var}(X_1) + \cdots + \text{var}(X_n),$$

其中 $\text{var}(Y)$ 表示 Y 的方差.

证明. 记 $\mu_i = EX_i$，$S_n = \sum_{i=1}^n X_i$. 由于 $ES_n = \sum_{i=1}^n \mu_i$，利用方差的定义，将和的平方写成两个和的乘积再展开，可得

$$\begin{aligned}
\text{var}(S_n) = E(S_n - ES_n)^2 &= E\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2 \\
&= E\sum_{i=1}^n \sum_{j=1}^n (X_i - \mu_i)(X_j - \mu_j) \\
&= \sum_{i=1}^n E(X_i - \mu_i)^2 + 2\sum_{i=1}^n \sum_{j=1}^{i-1} E((X_i - \mu_i)(X_j - \mu_j)),
\end{aligned}$$

其中最后一个等式将 $i = j$ 的对角项分离出来，并利用 $1 \leq i < j \leq n$ 上的和与 $1 \leq j < i \leq n$ 上的和相同这一事实.

第一个和式即为 $\text{var}(X_1) + \cdots + \text{var}(X_n)$，故只需证第二个和式为零. 为此，注意

$$\begin{aligned}
E((X_i - \mu_i)(X_j - \mu_j)) &= EX_iX_j - \mu_iEX_j - \mu_jEX_i + \mu_i\mu_j \\
&= EX_iX_j - \mu_i\mu_j = 0,
\end{aligned}$$

这里用到 X_i 与 X_j 不相关. □

定理 2.2.1 表明，对不相关随机变量而言，和的方差等于方差的和. 证明定理 2.2.3 的第二个要素是下面 (1.6.4) 的推论：

$$\text{var}(cY) = c^2 \,\text{var}(Y).$$

第三个也是最后一个要素是下述引理.

引理 2.2.2. 若 $p > 0$ 且 $E|Z_n|^p \to 0$，则 Z_n 依概率收敛于 0.

证明. 在 Chebyshev 不等式（定理 1.6.4）中取 $\varphi(x) = x^p$，$X = |Z_n|$，可知若 $\varepsilon > 0$ 则

$$P(|Z_n| \geq \varepsilon) \leq \varepsilon^{-p} E|Z_n|^p \to 0.$$

□

由此容易证明下述定理.

定理 2.2.3. L^2 **弱大数定律**. 设 $X_1, X_2, \ldots$ 是不相关随机变量，满足 $EX_i = \mu$ 且 $\text{var}(X_i) \leq C < \infty$. 若 $S_n = X_1 + \cdots + X_n$，则当 $n \to \infty$ 时，$S_n/n \to \mu$（在 L^2 意义下且依概率收敛）.

证明. 为证 L^2 收敛，注意 $E(S_n/n) = \mu$，故

$$E(S_n/n - \mu)^2 = \text{var}(S_n/n) = \frac{1}{n^2}(\text{var}(X_1) + \cdots + \text{var}(X_n)) \leq \frac{Cn}{n^2} \to 0.$$

为得出依概率收敛的结论，将引理 2.2.2 应用于 $Z_n = S_n/n - \mu$ 即可. □

定理 2.2.3 最重要的特殊情形是当 $X_1, X_2, \ldots$ 为独立同分布随机变量时. 按术语习惯，称其为**独立同分布**，简记为 i.i.d.. 定理 2.2.3 告诉我们，若 $EX_i^2 < \infty$，则当 $n \to \infty$ 时，S_n/n 依概率收敛于 $\mu = EX_i$. 定理 2.2.14 将表明 $E|X_i| < \infty$ 足以保证最后的结论，但目前先关注这一较弱结果的推论.

第一个应用是一个表面上与随机性无关的情形.

例题 2.2.4. 多项式逼近. 设 f 是 $[0,1]$ 上的连续函数，令

$$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f(m/n), \quad \text{其中} \binom{n}{m} = \frac{n!}{m!(n-m)!},$$

为与 f 相关联的 n 次 **Bernstein 多项式**. 则当 $n \to \infty$ 时

$$\sup_{x \in [0,1]} |f_n(x) - f(x)| \to 0.$$

证明. 首先注意，若 S_n 是 n 个独立随机变量的和，其中 $P(X_i=1)=p$，$P(X_i=0)=1-p$，则 $EX_i=p$，$\text{var}(X_i)=p(1-p)$，且

$$P(S_n=m)=\binom{n}{m}p^m(1-p)^{n-m},$$

故 $Ef(S_n/n)=f_n(p)$. 定理 2.2.3 告诉我们，当 $n\to\infty$ 时，S_n/n 依概率收敛于 p. 最后这两个观察启发了 $f_n(p)$ 的定义，但要证明所需的结论，还需利用定理 2.2.3 的证明而非定理本身.

将定理 2.2.3 的证明与 X_i 的方差公式以及 $p\in[0,1]$ 时 $p(1-p)\le 1/4$ 的事实相结合，可得

$$P(|S_n/n-p|>\delta)\le\frac{\text{var}(S_n/n)}{\delta^2}=\frac{p(1-p)}{n\delta^2}\le\frac{1}{4n\delta^2}.$$

为得出 $Ef(S_n/n)\to f(p)$ 的结论，记 $M=\sup_{x\in[0,1]}|f(x)|$，取 $\varepsilon>0$，并选取 $\delta>0$ 使得当 $|x-y|<\delta$ 时 $|f(x)-f(y)|<\varepsilon$.（这是可行的，因为连续函数在有界区间上一致连续.）由 Jensen 不等式（定理 1.6.2）可得

$$|Ef(S_n/n)-f(p)|\le E|f(S_n/n)-f(p)|\le\varepsilon+2MP(|S_n/n-p|>\delta).$$

令 $n\to\infty$，有 $\limsup_{n\to\infty}|Ef(S_n/n)-f(p)|\le\varepsilon$，但 ε 是任意的，故得所需结果. □

下面是一个轻松的调剂.

例题 2.2.5. 高维立方体几乎是一个球的边界. 设 $X_1,X_2,\ldots$ 独立且在 $(-1,1)$ 上均匀分布. 令 $Y_i=X_i^2$，它们因是独立随机变量的函数而保持独立. $EY_i=1/3$ 且 $\text{var}(Y_i)\le EY_i^2\le 1$，故定理 2.2.3 蕴涵

$$(X_1^2+\cdots+X_n^2)/n\to 1/3\quad\text{依概率收敛},\quad n\to\infty.$$

令

$$A_{n,\varepsilon}=\left\{x\in\mathbb{R}^n:(1-\varepsilon)\sqrt{n/3}<|x|<(1+\varepsilon)\sqrt{n/3}\right\},$$

其中 $|x|=(x_1^2+\cdots+x_n^2)^{1/2}$. 若以 $|S|$ 表示 S 的 Lebesgue 测度，则上述结论蕴涵对任意 $\varepsilon>0$ 有 $|A_{n,\varepsilon}\cap(-1,1)^n|/2^n\to 1$，换言之，立方体 $(-1,1)^n$ 的大部分体积来自 $A_{n,\varepsilon}$，而后者几乎是半径为 $\sqrt{n/3}$ 的球的边界.

2.2.2 三角阵列

概率论中许多经典极限定理涉及随机变量阵列 $X_{n,k},1\le k\le n$，并研究其**行和** $S_n=X_{n,1}+\cdots+X_{n,n}$ 的极限行为. 在多数情形下，假设每行中的随机变量相互独立，但对下面这个平凡（但有用）的结果，不需要这一假设. 实际上，这里的 S_n 可以是任意随机变量序列.

定理 2.2.6. 设 $\mu_n = ES_n$，$\sigma_n^2 = \mathrm{var}(S_n)$. 若 $\sigma_n^2/b_n^2 \to 0$，则

$$\frac{S_n - \mu_n}{b_n} \to 0 \quad \text{依概率收敛.}$$

证明. 由假设知 $E((S_n - \mu_n)/b_n)^2 = b_n^{-2}\,\mathrm{var}(S_n) \to 0$，故由引理 2.2.2 即得所需结论. □

下面给出定理 2.2.6 的三个应用. 对这三个例子，下列计算很有用：

$$\sum_{m=1}^{n} \frac{1}{m} \geq \int_1^n \frac{\mathrm{d}x}{x} \geq \sum_{m=2}^{n} \frac{1}{m},$$

即

$$\log n \leq \sum_{m=1}^{n} \frac{1}{m} \leq 1 + \log n. \tag{2.2.1}$$

例题 2.2.7. 优惠券收集问题. 设 $X_1, X_2, \ldots$ 独立同分布且在 $\{1, 2, \ldots, n\}$ 上均匀分布. 为解释名称，可设想收集棒球卡（或优惠券）. 假设收集的第 i 件物品从所有可能中随机选取且与先前选择独立. 令 $\tau_k^n = \inf\{m : |\{X_1, \ldots, X_m\}| = k\}$ 为首次收集到 k 种不同物品的时刻. 本问题关注 $T_n = \tau_n^n$ 的渐近行为，即收集完整一套所需的时间. 易见 $\tau_1^n = 1$. 为便于后续公式表达，令 $\tau_0^n = 0$. 对 $1 \leq k \leq n$，令 $X_{n,k} \equiv \tau_k^n - \tau_{k-1}^n$ 表示得到与先前 $k-1$ 种不同的选择所需的时间，故 $X_{n,k}$ 服从参数为 $1 - (k-1)/n$ 的几何分布，且与先前的等待时间 $X_{n,j}, 1 \leq j < k$ 独立. 例 1.6.14 指出，若 X 服从参数为 p 的几何分布，则 $EX = 1/p$ 且 $\mathrm{var}(X) \leq 1/p^2$.

利用期望的线性性、(2.2.1) 中关于 $\sum_{m=1}^n 1/m$ 的估计以及定理 2.2.1，可得

$$\begin{aligned} ET_n &= \sum_{k=1}^{n} \left(1 - \frac{k-1}{n}\right)^{-1} = n \sum_{m=1}^{n} m^{-1} \sim n \log n, \\ \mathrm{var}(T_n) &\leq \sum_{k=1}^{n} \left(1 - \frac{k-1}{n}\right)^{-2} = n^2 \sum_{m=1}^{n} m^{-2} \leq n^2 \sum_{m=1}^{\infty} m^{-2}. \end{aligned}$$

取 $b_n = n \log n$ 并应用定理 2.2.6，可得

$$\frac{T_n - n\sum_{m=1}^n m^{-1}}{n \log n} \to 0 \quad \text{依概率收敛,}$$

从而 $T_n/(n \log n) \to 1$ 依概率收敛.

举一个具体例子，取 $n = 365$，即我们想知道遇到拥有每个生日的人需要见多少人. 此时极限定理表明大约需要 $365 \log 365 = 2153.46$ 次尝试才能收集完整一套. 注意尝试次数是生日数量的 5.89 倍.

例题 2.2.8. 随机排列. 令 Ω_n 为 $n!$ 个排列（即从 $\{1, \ldots, n\}$ 到自身的双射）构成的集合，并假设所

有排列等可能，使之成为概率空间. 下面这个弱大数定律的应用关注随机排列 π 的循环结构，故首先描述排列到循环的分解. 考虑序列 $1, \pi(1), \pi(\pi(1)), \ldots$. 最终 $\pi^k(1) = 1$. 当这种情况发生时，称第一个循环完成，其长度为 k. 为开始第二个循环，选取不在第一个循环中的最小整数 i 并观察 $i, \pi(i), \pi(\pi(i)), \ldots$ 直到回到 i. 重复这一构造直至所有元素都被计入. 例如，若排列为

i	1	2	3	4	5	6	7	8	9
$\pi(i)$	3	9	6	8	2	1	5	4	7

则循环分解为 $(136)(2975)(48)$.

令 $X_{n,k} = 1$ 若在分解中第 k 个数后出现右括号，否则 $X_{n,k} = 0$，并令 $S_n = X_{n,1} + \cdots + X_{n,n}$ 为循环个数.（在上例中，$X_{9,3} = X_{9,7} = X_{9,9} = 1$，其余 $X_{9,m} = 0$.) 下面证明

引理 2.2.9. $X_{n,1}, \ldots, X_{n,n}$ 相互独立且 $P(X_{n,j} = 1) = \frac{1}{n-j+1}$.

直观上这是成立的：无论此前发生了什么，仍有 $n - j + 1$ 个值未出现在值域中，而只有其中之一会完成循环.

证明. 为证明这一点，用一种特殊方式生成排列会很有帮助. 令 $i_1 = 1$. 从 $\{1, \ldots, n\}$ 中随机选取 j_1 并令 $\pi(i_1) = j_1$. 若 $j_1 \neq 1$，令 $i_2 = j_1$；若 $j_1 = 1$，令 $i_2 = 2$. 无论哪种情形，从 $\{1, \ldots, n\} \setminus \{j_1\}$ 中随机选取 j_2. 一般地，若已选定 $i_1, j_1, \ldots, i_{k-1}, j_{k-1}$ 并令 $\pi(i_\ell) = j_\ell$ 对 $1 \leq \ell < k$，则 (a) 若 $j_{k-1} \in \{i_1, \ldots, i_{k-1}\}$（即刚刚完成一个循环），令 $i_k = \inf(\{1, \ldots, n\} \setminus \{i_1, \ldots, i_{k-1}\})$；(b) 若 $j_{k-1} \notin \{i_1, \ldots, i_{k-1}\}$，令 $i_k = j_{k-1}$. 无论哪种情形，从 $\{1, \ldots, n\} \setminus \{j_1, \ldots, j_{k-1}\}$ 中随机选取 j_k 并令 $\pi(i_k) = j_k$.

上述构造写起来或读起来都很繁琐，但现在可以理直气壮地断言 $X_{n,1}, \ldots, X_{n,n}$ 相互独立且 $P(X_{n,k} = 1) = 1/(n-k+1)$：当选取 j_k 时，$\{1, \ldots, n\} \setminus \{j_1, \ldots, j_{k-1}\}$ 中有 $n - k + 1$ 个值，而只有其中之一会完成循环. □

为验证定理 2.2.6 的条件，注意

$$
\begin{aligned}
ES_n &= \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1, \\
\operatorname{var}(S_n) &= \sum_{k=1}^{n} \operatorname{var}(X_{n,k}) \leq \sum_{k=1}^{n} E(X_{n,k}^2) = \sum_{k=1}^{n} E(X_{n,k}) = ES_n,
\end{aligned}
$$

其中第二行的结果来自定理 2.2.1、$\operatorname{var}(Y) \leq EY^2$ 以及 $X_{n,k}^2 = X_{n,k}$. 现 $ES_n \sim \log n$，故若 $b_n = (\log n)^{0.5+\varepsilon}$，$\varepsilon > 0$，则定理 2.2.6 的条件满足，从而

$$
\frac{S_n - \sum_{m=1}^{n} m^{-1}}{(\log n)^{0.5+\varepsilon}} \to 0 \quad \text{依概率收敛.} \tag{$*$}
$$

取 $\varepsilon = 0.5$ 得 $S_n / \log n \to 1$ 依概率收敛，但 (∗) 给出更强的结果. 例 3.4.11 将表明当 $\varepsilon = 0$ 时 (∗) 不成立.

例题 2.2.10. 占位问题. 假设将 r 个球随机放入 n 个盒子中，即所有 n^r 种球的分配方式等可能. 令 A_i 为第 i 个盒子为空的事件，N_n 为空盒个数. 易见

$$P(A_i) = (1-1/n)^r, \quad EN_n = n(1-1/n)^r.$$

简单计算（取对数）表明，若 $r/n \to c$，则 $EN_n/n \to e^{-c}$.（证明见引理 3.1.1.）为计算 N_n 的方差，注意

$$\begin{aligned} EN_n^2 &= E\left(\sum_{m=1}^{n} 1_{A_m}\right)^2 = \sum_{1\le k,m\le n} P(A_k \cap A_m), \\ \mathrm{var}(N_n) &= EN_n^2 - (EN_n)^2 = \sum_{1\le k,m\le n} [P(A_k\cap A_m) - P(A_k)P(A_m)] \\ &= n(n-1)\{(1-2/n)^r - (1-1/n)^{2r}\} + n\{(1-1/n)^r - (1-1/n)^{2r}\}. \end{aligned}$$

第一项来自 $k \neq m$，第二项来自 $k = m$. 由于 $(1-2/n)^r \to e^{-2c}$ 且 $(1-1/n)^r \to e^{-c}$，由最后一式易知 $\mathrm{var}(N_n/n) = \mathrm{var}(N_n)/n^2 \to 0$. 在定理 2.2.6 中取 $b_n = n$，得

$$N_n/n \to e^{-c} \quad \text{依概率收敛.}$$

2.2.3 截断

对随机变量 X 在水平 M 处**截断**是指考虑

$$\bar{X} = X1_{(|X|\le M)} = \begin{cases} X, & |X| \le M, \\ 0, & |X| > M. \end{cases}$$

为将弱大数定律推广到没有有限二阶矩的随机变量，采用截断再运用 Chebyshev 不等式的方法. 先给出一个非常一般但也十分有用的结果，其证明容易是因为已假设了证明所需的条件. 后面在特殊情形下验证假设时需要做一些工作，但一般结果有助于识别证明的本质要素.

定理 2.2.11. 三角阵列的弱大数定律. 对每个 n，设 $X_{n,k}, 1 \le k \le n$ 相互独立. 令 $b_n > 0$ 且 $b_n \to \infty$，并令 $\bar{X}_{n,k} = X_{n,k}1_{(|X_{n,k}|\le b_n)}$. 假设当 $n \to \infty$ 时

(i) $\sum_{k=1}^{n} P(|X_{n,k}| > b_n) \to 0$,

(ii) $b_n^{-2}\sum_{k=1}^{n} E\bar{X}_{n,k}^2 \to 0$.

若令 $S_n = X_{n,1} + \cdots + X_{n,n}$ 且 $a_n = \sum_{k=1}^{n} E\bar{X}_{n,k}$，则

$$(S_n - a_n)/b_n \to 0 \quad \text{依概率收敛.}$$

证明. 记 $\bar{S}_n = \bar{X}_{n,1} + \cdots + \bar{X}_{n,n}$. 显然

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \le P(S_n \neq \bar{S}_n) + P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right).$$

估计第一项，注意

$$P(S_n \neq \bar{S}_n) \le P\left(\bigcup_{k=1}^{n}\{\bar{X}_{n,k} \neq X_{n,k}\}\right) \le \sum_{k=1}^{n} P(|X_{n,k}| > b_n) \to 0$$

由 (i) 成立. 对第二项，注意 $a_n = E\bar{S}_n$，由 Chebyshev 不等式、定理 2.2.1 以及 $\mathrm{var}(X) \le EX^2$ 可得

$$\begin{aligned} P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right) &\le (\varepsilon b_n)^{-2} E\left|\bar{S}_n - a_n\right|^2 = (\varepsilon b_n)^{-2}\,\mathrm{var}(\bar{S}_n) \\ &= (\varepsilon b_n)^{-2} \sum_{k=1}^{n} \mathrm{var}(\bar{X}_{n,k}) \le (\varepsilon b_n)^{-2} \sum_{k=1}^{n} E(\bar{X}_{n,k})^2 \to 0 \end{aligned}$$

由 (ii) 成立，证毕. □

由定理 2.2.11 可得关于单个序列的下述结果.

定理 2.2.12. 弱大数定律. 设 $X_1, X_2, \ldots$ 独立同分布且

$$xP(|X_1| > x) \to 0 \quad \text{当} x \to \infty.$$

令 $S_n = X_1 + \cdots + X_n$，$\mu_n = E(X_1 1_{(|X_1| \le n)})$. 则 $S_n/n - \mu_n \to 0$ 依概率收敛.

注释. 定理中的假设是存在常数 a_n 使得 $S_n/n - a_n \to 0$ 的必要条件. 证明参见 Feller, Vol. II (1971) p. 234–236.

证明. 将定理 2.2.11 应用于 $X_{n,k} = X_k$，$b_n = n$. 验证 (i)，注意

$$\sum_{k=1}^{n} P(|X_{n,k}| > n) = nP(|X_1| > n) \to 0$$

由假设成立. 为验证 (ii)，需证 $n^{-2} \cdot nE\bar{X}_{n,1}^2 \to 0$. 为此，需要下面这个将在后面多次有用的结果.

引理 2.2.13. 若 $Y \ge 0$ 且 $p > 0$，则 $E(Y^p) = \int_0^\infty py^{p-1}P(Y > y)\mathrm{d}y$.

证明. 利用期望值的定义、Fubini 定理（对非负随机变量）然后计算所得积分，可得

$$\begin{aligned}\int_0^\infty py^{p-1}P(Y>y)\mathrm{d}y &= \int_0^\infty py^{p-1}\int_\Omega 1_{(Y>y)}\mathrm{d}P\mathrm{d}y\\ &= \int_\Omega\int_0^\infty py^{p-1}1_{(Y>y)}\mathrm{d}y\mathrm{d}P\\ &= \int_\Omega\int_0^Y py^{p-1}\mathrm{d}y\mathrm{d}P = \int_\Omega Y^p\mathrm{d}P = EY^p,\end{aligned}$$

即为所求. □

回到定理 2.2.12 的证明，由引理 2.2.13 以及 $\bar{X}_{n,1}=X_1 1_{(|X_1|\le n)}$ 这一事实，可得

$$E(\bar{X}_{n,1}^2)=\int_0^\infty 2yP(|\bar{X}_{n,1}|>y)\mathrm{d}y\le\int_0^n 2yP(|X_1|>y)\mathrm{d}y,$$

这是因为 $P(|\bar{X}_{n,1}|>y)=0$ 对 $y\ge n$ 成立，而对 $y\le n$ 有 $P(|\bar{X}_{n,1}|>y)=P(|X_1|>y)-P(|X_1|>n)$. 由 $yP(|X_1|>y)\to 0$ 可知

$$\frac{1}{n}E(\bar{X}_{n,1}^2)=\frac{1}{n}\int_0^n 2yP(|X_1|>y)\mathrm{d}y\to 0,\quad n\to\infty.$$

直观上，这是因为右端是 $g(y)=2yP(|X_1|>y)$ 在 $[0,n]$ 上的平均值，且 $g(y)\to 0$ 当 $y\to\infty$. 详言之，注意 $0\le g(y)\le 2y$ 且 $g(y)\to 0$ 当 $y\to\infty$，故必有 $M=\sup g(y)<\infty$. 令 $g_n(y)=g(ny)$. 由于 g 有界且 $g_n(y)\to 0$ a.s.，

$$\frac{1}{n}\int_0^n g(y)\mathrm{d}y=\int_0^1 g_n(x)\mathrm{d}x\to 0,$$

证毕. □

注释. 将引理 2.2.13 应用于 $p=1-\varepsilon$，$\varepsilon>0$，可见 $xP(|X_1|>x)\to 0$ 蕴涵 $E|X_1|^{1-\varepsilon}<\infty$，故定理 2.2.12 中的假设并不比有限均值弱太多.

最后，给出最熟悉形式的弱大数定律.

定理 2.2.14. 设 $X_1,X_2,\ldots$ 独立同分布且 $E|X_i|<\infty$. 令 $S_n=X_1+\cdots+X_n$，$\mu=EX_1$. 则 $S_n/n\to\mu$ 依概率收敛.

证明. 两次应用控制收敛定理可得

$$\begin{aligned}xP(|X_1|>x)\le E(|X_1|1_{(|X_1|>x)})\to 0,\quad x\to\infty,\\ \mu_n=E(X_1 1_{(|X_1|\le n)})\to E(X_1)=\mu,\quad n\to\infty.\end{aligned}$$

利用定理 2.2.12，可知若 $\varepsilon > 0$ 则 $P(|S_n/n - \mu_n| > \varepsilon/2) \to 0$. 由于 $\mu_n \to \mu$，故 $P(|S_n/n - \mu| > \varepsilon) \to 0$. □

例题 2.2.15. 举一个弱大数定律不成立的例子，假设 $X_1, X_2, \ldots$ 独立且服从**柯西分布**：

$$P(X_i \leq x) = \int_{-\infty}^{x} \frac{\mathrm{d}t}{\pi(1+t^2)}.$$

当 $x \to \infty$ 时，

$$P(|X_1| > x) = 2\int_x^{\infty} \frac{\mathrm{d}t}{\pi(1+t^2)} \sim \frac{2}{\pi}\int_x^{\infty} t^{-2}\mathrm{d}t = \frac{2}{\pi}x^{-1}.$$

由上述条件的必要性，可得出结论：不存在常数列 μ_n 使得 $S_n/n - \mu_n \to 0$. 后面将看到 S_n/n 总与 X_1 同分布.（见习题 3.3.6.）

如下例所示，即使在 $E|X| = \infty$ 的情形，有时仍可有弱大数定律成立.

例题 2.2.16. "St. Petersburg 悖论". 设 $X_1, X_2, \ldots$ 为独立随机变量且

$$P(X_i = 2^j) = 2^{-j}, \quad j \geq 1.$$

换言之，若掷硬币 j 次才出现正面，则赢得 2^j 美元. 这里的悖论在于 $EX_1 = \infty$，但显然你不会愿意支付无限金额来玩这个游戏. 应用定理 2.2.11 可以告诉我们玩 n 次游戏应该支付多少.

本例中，$X_{n,k} = X_k$. 为应用定理 2.2.11，需要选取 b_n. 为此，遵循如下原则：在验证 (ii) 时希望 b_n 尽可能小，同时使 (i) 成立. 依此原则，注意若 m 为整数则

$$P(X_1 \geq 2^m) = \sum_{j=m}^{\infty} 2^{-j} = 2^{-m+1}.$$

令 $m(n) = \log_2 n + K(n)$，其中 $K(n) \to \infty$ 且选取使得 $m(n)$ 为整数（从而使上式有效）. 令 $b_n = 2^{m(n)}$，有

$$nP(X_1 \geq b_n) = n \cdot 2^{-m(n)+1} = 2^{-K(n)+1} \to 0,$$

证得 (i). 为验证 (ii)，注意若 $\bar{X}_{n,k} = X_k 1_{(|X_k| \leq b_n)}$ 则

$$E\bar{X}_{n,k}^2 = \sum_{j=1}^{m(n)} 2^{2j} \cdot 2^{-j} \leq 2^{m(n)} \sum_{k=0}^{\infty} 2^{-k} = 2b_n.$$

故 (ii) 中的表达式小于 $2n/b_n$，而这趋于 0，因为

$$b_n = 2^{m(n)} = n2^{K(n)} \quad \text{且} \quad K(n) \to \infty.$$

习题

习题 2.2.1. 设 $X_1, X_2, \ldots$ 独立同分布且 $P(X_i > x) = e/x$ 对 $x \geq e$ 成立，其中 $e = \exp(1)$. 利用定理 2.2.12 在 a_n 如该定理所示时证明 $(S_n - a_n)/n \Rightarrow 0$.

习题 2.2.2. 设 $X_1, X_2, \ldots$ 独立同分布且 $P(X_i = (-1)^k k) = C/k^2 \log k$ 对 $k \geq 2$ 成立，其中 C 为正则化常数. 证明 $E|X_1| = \infty$，但存在 μ 使得 $S_n/n \to \mu$ 依概率收敛.

习题 2.2.3. (a) 证明若 $X \geq 0$ 为整数值随机变量，则 $EX = \sum_{n\geq 1} P(X \geq n)$. (b) 对非负整数值随机变量 X，求 $EX(X-1)$ 的类似公式. (c) 推广你的结论.

习题 2.2.4. 推广定理 2.2.12 的证明以证明：若对每个 n，$X_{n,1}, \ldots, X_{n,n}$ 相互独立且

$$\lim_{n\to\infty} \sum_{k=1}^{n} P(|X_{n,k}| > \varepsilon) = 0 \quad \text{对所有} \varepsilon > 0,$$

且

$$\lim_{n\to\infty} \sum_{k=1}^{n} E(X_{n,k}^2 1_{(|X_{n,k}|\leq 1)}) = 0,$$

则 $S_n - ES_n \to 0$ 依概率收敛.

习题 2.2.5. 设 $X_1, X_2, \ldots$ 独立同分布且 $P(X_i > x) = e/x \log x$ 对 $x \geq e$ 成立. 证明 $E|X_i| = \infty$，但存在常数列 $\mu_n \to \infty$ 使得 $S_n/n - \mu_n \to 0$ 依概率收敛.

习题 2.2.6. (i) 证明若 $X \geq 0$ 为整值随机变量，则 $EX = \sum_{n\geq 1} P(X \geq n)$. (ii) 求 EX^2 的类似表达式.

习题 2.2.7. 推广引理 2.2.13 以证明：若 $H(x) = \int_{(-\infty,x]} h(y)\mathrm{d}y$ 且 $h(y) \geq 0$，则

$$EH(X) = \int_{-\infty}^{\infty} h(y)P(X \geq y)\mathrm{d}y.$$

一个重要的特例是 $H(x) = \exp(\theta x)$，其中 $\theta > 0$.

习题 2.2.8. 不公平的 "公平游戏". 设 $p_k = 1/2^k k(k+1)$，$k = 1, 2, \ldots$，且 $p_0 = 1 - \sum_{k\geq 1} p_k$.

$$\sum_{k=1}^{\infty} 2^k p_k = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots = 1,$$

故若令 $X_1, X_2, \ldots$ 独立同分布，满足 $P(X_n = -1) = p_0$ 且 $P(X_n = 2^k - 1) = p_k$ 对 $k \geq 1$，则 $EX_n = 0$. 令 $S_n = X_1 + \cdots + X_n$. 在定理 2.2.11 中取 $b_n = 2^{m(n)}$，其中 $m(n) = \min\{m : 2^{-m} m^{-3/2} \leq n^{-1}\}$，证明 $S_n/(n/\log_2 n) \to -1$ 依概率收敛.

习题 2.2.9. 正变量弱大数定律. 设 $X_1, X_2, \ldots$ 独立同分布，$P(0 \leq X_i < \infty) = 1$ 且对所有 x 有 $P(X_i > x) > 0$. 令 $\mu(s) = \int_0^s x\mathrm{d}F(x)$，$\nu(s) = \mu(s)/s(1 - F(s))$. 已知存在常数 a_n 使得 $S_n/a_n \to 1$ 依概率收敛当且仅当 $\nu(s) \to \infty$ 当 $s \to \infty$. 选取 $b_n \geq 1$ 使得 $n\mu(b_n) = b_n$（对大的 n 这可行），并利用定理 2.2.11 证明该条件是充分的.

第 3 节 Borel-Cantelli 引理

若 A_n 是 Ω 的一列子集，定义

$$\limsup A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \text{ 属于无穷多个} A_n\}$$

（极限存在是因为该序列关于 m 递减）以及

$$\liminf A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \text{ 属于除有限个外所有的} A_n\}$$

（极限存在是因为该序列关于 m 递增）. 上极限和下极限的名称可由下式说明:

$$\limsup_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_{(\limsup A_n)} \quad \liminf_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_{(\liminf A_n)}$$

通常记 $\limsup A_n = \{\omega : \omega \in A_n \text{ i.o.}\}$，其中 i.o. 代表无穷多次. 下例说明了这一记号的用法: "$X_n \to 0$ a.s. 当且仅当对所有 $\epsilon > 0$，$P(|X_n| > \epsilon \text{ i.o.}) = 0$." 读者将在下文看到更多例子.

习题 2.3.1. 证明 $P(\limsup A_n) \geq \limsup P(A_n)$ 且 $P(\liminf A_n) \leq \liminf P(A_n)$.

下述结果即便名字可能陌生，其内容在测度论中应该耳熟能详.

定理 2.3.1. Borel-Cantelli 引理. 若 $\sum_{n=1}^{\infty} P(A_n) < \infty$，则 $P(A_n \text{ i.o.}) = 0$.

证明. 记 $N = \sum_k \mathbf{1}_{A_k}$ 为发生的事件个数. Fubini 定理表明 $EN = \sum_k P(A_k) < \infty$，故必有 $N < \infty$ a.s. □

下述结果是 Borel-Cantelli 引理的一个典型应用.

定理 2.3.2. $X_n \to X$ 依概率收敛当且仅当对每个子序列 $X_{n(m)}$，存在进一步的子序列 $X_{n(m_k)}$ 几乎必然收敛于 X.

证明. 设 ϵ_k 是一列递减趋于 0 的正数. 对每个 k，存在 $n(m_k) > n(m_{k-1})$ 使得 $P(|X_{n(m_k)} - X| > \epsilon_k) \leq 2^{-k}$. 由于

$$\sum_{k=1}^{\infty} P(|X_{n(m_k)} - X| > \epsilon_k) < \infty$$

Borel-Cantelli 引理表明 $P(|X_{n(m_k)} - X| > \epsilon_k \text{ i.o.}) = 0$，即 $X_{n(m_k)} \to X$ a.s. 为证明第二个结论，注意若对每个子序列 $X_{n(m)}$ 都存在进一步的子序列 $X_{n(m_k)}$ 几乎必然收敛于 X，则对任意 $\delta > 0$，可将下一定理应用于数列 $y_n = P(|X_n - X| > \delta)$ 以得到所需结果. □

定理 2.3.3. 设 y_n 是拓扑空间中的一列元素. 若每个子序列 $y_{n(m)}$ 都有进一步的子序列 $y_{n(m_k)}$ 收敛于 y，则 $y_n \to y$.

证明. 若 $y_n \not\to y$，则存在包含 y 的开集 G 和子序列 $y_{n(m)}$ 使得对所有 m 都有 $y_{n(m)} \notin G$，但显然 $y_{n(m)}$ 的任何子序列都不收敛于 y. □

注释. 由于存在依概率收敛但不几乎必然收敛的随机变量序列（例子见习题 2.3.14 或 2.3.11），由定理2.3.2可知几乎必然收敛不由度量导出，甚至不由拓扑导出. 习题 2.3.6 将给出一个依概率收敛的度量，习题 2.3.7 将表明随机变量空间在此度量下是完备空间.

定理2.3.2使我们能将依概率收敛提升为几乎必然收敛. 其用途之一是

定理 2.3.4. 若 f 连续且 $X_n \to X$ 依概率收敛，则 $f(X_n) \to f(X)$ 依概率收敛. 此外，若 f 有界，则 $Ef(X_n) \to Ef(X)$.

证明. 若 $X_{n(m)}$ 是子序列，则定理2.3.2表明存在进一步的子序列 $X_{n(m_k)} \to X$ 几乎必然. 由于 f 连续，习题 1.3.3 表明 $f(X_{n(m_k)}) \to f(X)$ 几乎必然，再由定理2.3.2得 $f(X_n) \to f(X)$ 依概率收敛. 若 f 有界，则有界收敛定理表明 $Ef(X_{n(m_k)}) \to Ef(X)$，将定理2.3.3应用于 $y_n = Ef(X_n)$ 即得所需结果. □

习题 2.3.2. 直接从定义证明定理2.3.4的第一个结果.

作为 Borel-Cantelli 引理的第二个应用，我们得到第一个强大数定律：

定理 2.3.5. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = \mu$ 且 $EX_i^4 < \infty$. 若 $S_n = X_1 + \cdots + X_n$，则 $S_n/n \to \mu$ a.s.

证明. 令 $X_i^0 = X_i - \mu$，由对称性可设 $\mu = 0$. 现在

$$ES_n^4 = E\left(\sum_{i=1}^n X_i\right)^4 = E\sum_{1\le i,j,k,\ell\le n} X_iX_jX_kX_\ell$$

形如 $E(X_i^3X_j)$、$E(X_i^2X_jX_k)$ 和 $E(X_iX_jX_kX_\ell)$ 的项（当 i,j,k,ℓ 互不相同时）均为 0，因为乘积的期望等于期望的乘积，且每种情形都有一个项的期望为 0. 唯一不消失的项是形如 EX_i^4 和 $EX_i^2X_j^2 = (EX_i^2)^2$ 的项. 这两类项分别有 n 个和 $3n(n-1)$ 个.（对第二类，两个指标有 $n(n-1)/2$ 种选法，指标固定后，该项总共能以 6 种方式出现.）最后这一观察表明

$$ES_n^4 = nEX_1^4 + 3(n^2-n)(EX_1^2)^2 \le Cn^2$$

其中 $C < \infty$. Chebyshev 不等式给出

$$P(|S_n| > n\epsilon) \le E(S_n^4)/(n\epsilon)^4 \le C/(n^2\epsilon^4)$$

对 n 求和并利用 Borel-Cantelli 引理得 $P(|S_n| > n\epsilon \text{ i.o.}) = 0$. 由于 ϵ 任意，证毕. □

Borel-Cantelli 引理的逆命题显然不成立.

例题 2.3.6. 设 $\Omega = (0,1)$，$\mathcal{F}$ =Borel 集，P =Lebesgue 测度. 若 $A_n = (0, a_n)$，其中 $a_n \to 0$（当 $n \to \infty$），则 $\limsup A_n = \emptyset$，但若 $a_n \geq 1/n$，有 $\sum_n a_n = \infty$.

上面给出的例子表明，对一般集合，我们无法比习题 2.3.1 的结果说得更多.

然而，对独立事件，$P(\limsup A_n) > 0$ 的必要条件足以保证 $P(\limsup A_n) = 1$.

定理 2.3.7. 第二 Borel-Cantelli 引理. 若事件 A_n 相互独立，则 $\sum_n P(A_n) = \infty$ 意味着 $P(A_n \text{ i.o.}) = 1$.

证明. 设 $M < N < \infty$. 由独立性和 $1 - x \leq e^{-x}$ 得

$$P\left(\bigcap_{n=M}^{N} A_n^c\right) = \prod_{n=M}^{N} (1 - P(A_n)) \leq \exp\left(-\sum_{n=M}^{N} P(A_n)\right) \to 0 \quad (\text{当} N \to \infty)$$

故对所有 M 有 $P(\bigcup_{n=M}^{\infty} A_n) = 1$，又因 $\bigcup_{n=M}^{\infty} A_n \downarrow \limsup A_n$，故 $P(\limsup A_n) = 1$. □

第二 Borel-Cantelli 引理的一个典型应用是：

定理 2.3.8. 若 $X_1, X_2, \ldots$ 独立同分布且 $E|X_i| = \infty$，则 $P(|X_n| \geq n \text{ i.o.}) = 1$. 因此若 $S_n = X_1 + \cdots + X_n$，则 $P(\lim S_n/n \text{ 存在} \in (-\infty, \infty)) = 0$.

证明. 由引理 2.2.13 得

$$E|X_1| = \int_0^{\infty} P(|X_1| > x)\, \mathrm{d}x \leq \sum_{n=0}^{\infty} P(|X_1| > n)$$

由于 $E|X_1| = \infty$ 且 $X_1, X_2, \ldots$ 独立同分布，由第二 Borel-Cantelli 引理得 $P(|X_n| \geq n \text{ i.o.}) = 1$. 为证第二个断言，注意

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

且在 $C \equiv \{\omega : \lim_{n\to\infty} S_n/n \text{ 存在} \in (-\infty, \infty)\}$ 上，$S_n/(n(n+1)) \to 0$. 故在 $C \cap \{\omega : |X_n| \geq n \text{ i.o.}\}$ 上，有

$$\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| > 2/3 \text{ i.o.}$$

这与 $\omega \in C$ 矛盾. 由最后这一观察，得 $\{\omega : |X_n| \geq n \text{ i.o.}\} \cap C = \emptyset$，又因 $P(|X_n| \geq n \text{ i.o.}) = 1$，故 $P(C) = 0$. □

定理2.3.8表明 $E|X_i| < \infty$ 是强大数定律的必要条件. 读者需等到定理 2.4.1 才能看到该条件也是充分的. 下述结果推广了第二 Borel-Cantelli 引理并加强了其结论.

定理 2.3.9. 若 $A_1, A_2, \ldots$ 两两独立且 $\sum_{n=1}^{\infty} P(A_n) = \infty$，则当 $n \to \infty$ 时，

$$\frac{\sum_{m=1}^{n} \mathbf{1}_{A_m}}{\sum_{m=1}^{n} P(A_m)} \to 1 \quad \text{a.s.}$$

证明. 令 $X_m = \mathbf{1}_{A_m}$ 且 $S_n = X_1 + \cdots + X_n$. 由于 A_m 两两独立，X_m 不相关，故定理 2.2.1 表明

$$\text{var}(S_n) = \text{var}(X_1) + \cdots + \text{var}(X_n)$$

由 $X_m \in \{0, 1\}$ 得 $\text{var}(X_m) \le E(X_m^2) = E(X_m)$，故 $\text{var}(S_n) \le E(S_n)$. Chebyshev 不等式表明

$$(*) \quad P(|S_n - ES_n| > \delta ES_n) \le \text{var}(S_n)/(\delta ES_n)^2 \le 1/(\delta^2 ES_n) \to 0$$

当 $n \to \infty$.（我们假设了 $ES_n \to \infty$.）

上述计算表明 $S_n/ES_n \to 1$ 依概率收敛. 为得到几乎必然收敛，需要取子序列. 令 $n_k = \inf\{n : ES_n \ge k^2\}$. 令 $T_k = S_{n_k}$，由定义和 $EX_m \le 1$ 得 $k^2 \le ET_k \le k^2 + 1$. 在 $(*)$ 中以 n_k 代 n 并利用 $ET_k \ge k^2$ 得

$$P(|T_k - ET_k| > \delta ET_k) \le 1/(\delta^2 k^2)$$

故 $\sum_{k=1}^{\infty} P(|T_k - ET_k| > \delta ET_k) < \infty$, Borel-Cantelli 引理表明 $P(|T_k - ET_k| > \delta ET_k \text{ i.o.}) = 0$. 由于 δ 任意, 得 $T_k/ET_k \to 1$ a.s. 为证 $S_n/ES_n \to 1$ a.s., 取 ω 使得 $T_k(\omega)/ET_k \to 1$ 并注意若 $n_k \le n < n_{k+1}$ 则

$$\frac{T_k}{ET_{k+1}} \le \frac{S_n}{ES_n} \le \frac{T_{k+1}}{ET_k}$$

为证左右两端趋于 1，将上述不等式改写为

$$\frac{ET_k}{ET_{k+1}} \cdot \frac{T_k(\omega)}{ET_k} \le \frac{S_n(\omega)}{ES_n} \le \frac{T_{k+1}(\omega)}{ET_{k+1}} \cdot \frac{ET_{k+1}}{ET_k}$$

由此可知只需证 $ET_{k+1}/ET_k \to 1$，而这由下式

$$k^2 \le ET_k \le ET_{k+1} \le (k+1)^2 + 1$$

及 $\{(k+1)^2 + 1\}/k^2 = 1 + 2/k + 2/k^2 \to 1$ 可得. □

定理2.3.9证明的启示是，若要证明对递增序列 $c_n, X_n \ge 0$ 有 $X_n/c_n \to 1$ a.s.，只需对满足 $c_{n(k+1)}/c_{n(k)} \to 1$ 的子序列 $n(k)$ 证明该结果.

例题 2.3.10. 记录值. 设 $X_1, X_2, \ldots$ 是一列随机变量，将 X_k 视为某运动员第 k 次跳高或掷铅球的距离，则 $A_k = \{X_k > \sup_{j<k} X_j\}$ 是在时刻 k 创纪录的事件. 忽略运动员可能因经验增加而表现更好

或可能受伤的事实，假设 $X_1, X_2, \ldots$ 独立同分布，具有连续分布函数 $F(x)$. 尽管时刻 k 创纪录似乎会降低时刻 $k+1$ 创纪录的可能性，我们

断言. A_k 相互独立且 $P(A_k) = 1/k$.

为证明此断言，首先注意由于 F 连续，对任意 $j \neq k$ 有 $P(X_j = X_k) = 0$（见习题 2.1.5），故可令 $Y_n^1 > Y_n^2 > \cdots > Y_n^n$ 为 $X_1, \ldots, X_n$ 按递减顺序排列的结果，并定义 $\{1, \ldots, n\}$ 的随机排列 $\pi_n(i) = j$（若 $X_i = Y_n^j$），即第 i 个随机变量的秩为 j. 由于 $(X_1, \ldots, X_n)$ 的分布不因改变随机变量顺序而改变，易见：

(a) 排列 π_n 在 $n!$ 种可能上是均匀分布的.

(a) 的证明. 由对称性这是" 显然" 的，但若想进一步说明，可如下论证. 设 π_n 是由 $(X_1, \ldots, X_n)$ 诱导的排列，σ_n 是与 X 序列独立的随机选取的 $\{1, \ldots, n\}$ 的排列. 则关于 $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ 诱导的排列有两件事可说：(i) 它是 $\pi_n \circ \sigma_n$，(ii) 它与 π_n 同分布. 注意到若 π 是任意排列，$\pi \circ \sigma_n$ 在 $n!$ 种可能上是均匀的，即得所需结果. □

一旦接受 (a)，其余就容易了：

(b) $P(A_n) = P(\pi_n(n) = 1) = 1/n$.

(c) 若 $m < n$ 且 $i_{m+1}, \ldots, i_n$ 是 $\{1, \ldots, n\}$ 的不同元素，则

$$P(A_m \mid \pi_n(j) = i_j \text{ 对} m+1 \leq j \leq n) = 1/m$$

直观上这是成立的，因为若我们条件于 $X_{m+1}, \ldots, X_n$ 的秩，则这确定了 $X_1, \ldots, X_m$ 可用的秩集合，但所有可能的秩排序都是等可能的，故最小秩最终落在 m 处的概率为 $1/m$.

(c) 的证明. 若令 σ_m 为 $\{1, \ldots, m\}$ 的随机选取的排列，则 (i)$\pi_n \circ \sigma_m$ 与 π_n 同分布，(ii) 由于 σ_m 的应用随机重排 $\pi_n(1), \ldots, \pi_n(m)$，即得所需结果. □

若令 $m_1 < m_2 < \ldots < m_k$，则由 (c) 得

$$P(A_{m_1} \mid A_{m_2} \cap \ldots \cap A_{m_k}) = P(A_{m_1})$$

由归纳法即得断言.

利用定理2.3.9和熟知的 $\sum_{m=1}^{n} 1/m \sim \log n$，有

定理 2.3.11. 若 $R_n = \sum_{m=1}^{n} \mathbf{1}_{A_m}$ 为时刻 n 的记录个数，则当 $n \to \infty$ 时，

$$R_n / \log n \to 1 \quad \text{a.s.}$$

读者应注意最后这个结果与分布 F 无关（只要它是连续的）.

注释. 设 $X_1, X_2, \ldots$ 独立同分布且具有连续分布. 令 Y_i 为满足 $j \le i$ 且 $X_j > X_i$ 的 j 的个数. 由 (a) 知 Y_i 是独立随机变量且 $P(Y_i = j) = 1/i$ （$0 \le j < i-1$）.

轻松一刻. 设 $X_0, X_1, \ldots$ 独立同分布, 想象它们是你卖车得到的报价. 令 $N = \inf\{n \ge 1 : X_n > X_0\}$. 对称性表明 $P(N > n) \ge 1/(n+1)$.（分布连续时该概率恰为 $1/(n+1)$，但现在的分布是一般的，平局归于第一个来电者.）利用习题 2.2.7:

$$EN = \sum_{n=0}^{\infty} P(N > n) \ge \sum_{n=0}^{\infty} \frac{1}{n+1} = \infty$$

故你等待优于第一个报价的期望时间为 ∞. 为避免诉讼，我得赶紧补充我并非建议你接受第一个报价!

例题 2.3.12. 连续正面. 设 X_n，$n \in \mathbb{Z}$，独立同分布且 $P(X_n = 1) = P(X_n = -1) = 1/2$. 令 $\ell_n = \max\{m : X_{n-m+1} = \ldots = X_n = 1\}$ 为时刻 n 的连续 $+1$ 游程长度，$L_n = \max_{1 \le m \le n} \ell_m$ 为时刻 n 的最长游程. 我们用双侧序列使得对所有 n 有 $P(\ell_n = k) = (1/2)^{k+1}$ （$k \ge 0$）. 由于 $\ell_1 < \infty$，我们将证明的结果

$$L_n / \log_2 n \to 1 \quad \text{a.s.} \tag{2.3.1}$$

对单侧序列也成立. 为证 (2.3.1)，首先注意对任意 $\epsilon > 0$ 有

$$P(\ell_n \ge (1+\epsilon) \log_2 n) \le n^{-(1+\epsilon)}$$

故由 Borel-Cantelli 引理得 $\ell_n \le (1+\epsilon) \log_2 n$ （$n \ge N_\epsilon$）. 由于 ϵ 任意，得

$$\limsup_{n \to \infty} L_n / \log_2 n \le 1 \quad \text{a.s.}$$

为得到另一方向的结果，将前 n 次试验划分为长度 $[(1-\epsilon) \log_2 n] + 1$ 的不交块，每块上变量全为 1 的概率为

$$2^{-[(1-\epsilon) \log_2 n]-1} \ge n^{-(1-\epsilon)}/2$$

由此得当 n 足够大使得 $[n/\{[(1-\epsilon) \log_2 n] + 1\}] \ge n / \log_2 n$ 时

$$P(L_n \le (1-\epsilon) \log_2 n) \le (1 - n^{-(1-\epsilon)}/2)^{n/(\log_2 n)} \le \exp(-n^{\epsilon}/2 \log_2 n)$$

该式可求和，故 Borel-Cantelli 引理表明

$$\liminf_{n \to \infty} L_n / \log_2 n \ge 1 \quad \text{a.s.}$$

习题 2.3.3. 设 ℓ'_n 为时刻的正面游程长度. 精确定义见例 8.3.1. 证明 $\limsup_{n \to \infty} \ell'_n / \log_2 n = 1$,

$\liminf_{n\to\infty} \ell'_n = 0$ a.s.

习题

习题 2.3.4. Fatou 引理. 设 $X_n \geq 0$ 且 $X_n \to X$ 依概率收敛. 证明 $\liminf_{n\to\infty} EX_n \geq EX$.

习题 2.3.5. 控制收敛. 设 $X_n \to X$ 依概率收敛且 (a)$|X_n| \leq Y$ 且 $EY < \infty$，或 (b) 存在连续函数 g 满足当 $|x|$ 充分大时 $g(x) > 0$ 且 $|x|/g(x) \to 0$（当 $|x| \to \infty$），使得 $Eg(X_n) \leq C < \infty$ 对所有 n 成立. 证明 $EX_n \to EX$.

习题 2.3.6. 依概率收敛的度量. 证明 (a)$d(X,Y) = E(|X-Y|/(1+|X-Y|))$ 在随机变量集合上定义了一个度量，即 (i)$d(X,Y) = 0$ 当且仅当 $X = Y$ a.s.，(ii)$d(X,Y) = d(Y,X)$，(iii)$d(X,Z) \leq d(X,Y) + d(Y,Z)$；且 (b)$d(X_n,X) \to 0$（当 $n \to \infty$）当且仅当 $X_n \to X$ 依概率收敛.

习题 2.3.7. 证明随机变量在上题定义的度量下构成完备空间，即若当 $m,n \to \infty$ 时 $d(X_m,X_n) \to 0$，则存在随机变量 X_∞ 使得 $X_n \to X_\infty$ 依概率收敛.

习题 2.3.8. 设 A_n 是一列独立事件且对所有 n 有 $P(A_n) < 1$. 证明 $P(\bigcup_n A_n) = 1$ 意味着 $\sum_n P(A_n) = \infty$，从而 $P(A_n \text{ i.o.}) = 1$.

习题 2.3.9. (i) 若 $P(A_n) \to 0$ 且 $\sum_{n=1}^{\infty} P(A_n^c \cap A_{n+1}) < \infty$，则 $P(A_n \text{ i.o.}) = 0$. (ii) 找一个序列 A_n 的例子，使得 (i) 的结果可用但 Borel-Cantelli 引理不能用.

习题 2.3.10. Kochen-Stone 引理. 设 $\sum_k P(A_k) = \infty$. 利用习题 1.6.6 和 2.3.1 证明若

$$\limsup_{n\to\infty} \left(\sum_{k=1}^{n} P(A_k) \right)^2 \Big/ \left(\sum_{1\leq j,k\leq n} P(A_j \cap A_k) \right) = \alpha > 0$$

则 $P(A_n \text{ i.o.}) \geq \alpha$. $\alpha = 1$ 的情形包含定理2.3.7.

习题 2.3.11. 设 $X_1, X_2, \ldots$ 独立且 $P(X_n = 1) = p_n$，$P(X_n = 0) = 1 - p_n$. 证明 (i)$X_n \to 0$ 依概率收敛当且仅当 $p_n \to 0$，(ii)$X_n \to 0$ a.s. 当且仅当 $\sum_n p_n < \infty$.

习题 2.3.12. 设 $X_1, X_2, \ldots$ 是 $(\Omega, \mathcal{F}, P)$ 上的随机变量序列，其中 Ω 是可数集且 $\mathcal{F}$ 由 Ω 的所有子集构成. 证明 $X_n \to X$ 依概率收敛意味着 $X_n \to X$ a.s.

习题 2.3.13. 若 X_n 是任意随机变量序列，则存在趋于 ∞ 的常数 c_n 使得 $X_n/c_n \to 0$ a.s.

习题 2.3.14. 设 $X_1, X_2, \ldots$ 独立. 证明 $\sup_n X_n < \infty$ a.s. 当且仅当 $\sum_n P(X_n > A) < \infty$ 对某个 A 成立.

习题 2.3.15. 设 $X_1, X_2, \ldots$ 独立同分布且 $P(X_i > x) = e^{-x}$，$M_n = \max_{1\leq m\leq n} X_m$. 证明：

(i) $\limsup_{n\to\infty} X_n/\log n = 1$ a.s. 且
(ii) $M_n/\log n \to 1$ a.s.

习题 2.3.16. 设 $X_1, X_2, \ldots$ 独立同分布且具有分布函数 F，$\lambda_n \uparrow \infty$，$A_n = \{\max_{1\le m\le n} X_m > \lambda_n\}$. 证明 $P(A_n \text{ i.o.}) = 0$ 或 1 按 $\sum_{n\ge 1}(1 - F(\lambda_n)) < \infty$ 或 $= \infty$ 而定.

习题 2.3.17. 设 $Y_1, Y_2, \ldots$ 独立同分布. 求 (i)$Y_n/n \to 0$ 几乎必然，(ii)$(\max_{m\le n} Y_m)/n \to 0$ 几乎必然，(iii)$(\max_{m\le n} Y_m)/n \to 0$ 依概率收敛，(iv)$Y_n/n \to 0$ 依概率收敛的充要条件.

习题 2.3.18. 设 $0 \le X_1 \le X_2 \ldots$ 是随机变量，$EX_n \sim an^\alpha$ $(a, \alpha > 0)$，$\text{var}(X_n) \le Bn^\beta$ $(\beta < 2\alpha)$. 证明 $X_n/n^\alpha \to a$ a.s.

习题 2.3.19. 设 X_n 是独立的 Poisson 随机变量，$EX_n = \lambda_n$，$S_n = X_1 + \cdots + X_n$. 证明若 $\sum_n \lambda_n = \infty$ 则 $S_n/ES_n \to 1$ a.s.

习题 2.3.20. 证明若 X_n 是 St. Petersburg 游戏（例 2.2.16）第 n 次的结果，则

$$\limsup_{n\to\infty} X_n/(n\log_2 n) = \infty \text{a.s.}$$

从而对 S_n 也有同样结果. 这表明 2.2 节证明的依概率收敛 $S_n/(n\log_2 n) \to 1$ 不几乎必然成立.

第 4 节 强大数定律

下面给出 Etemadi(1981) 对**强大数定律**的证明.

定理 2.4.1. 强大数定律. 设 $X_1, X_2, \ldots$ 为**两两独立的同分布的**随机变量且 $E|X_i| < \infty$. 记 $EX_i = \mu$, $S_n = X_1 + \cdots + X_n$, 则当 $n \to \infty$ 时, $S_n/n \to \mu$ a.s.

证明. 与弱大数定律的证明类似, 首先进行**截断**.

引理 2.4.2. 设 $Y_k = X_k 1_{(|X_k| \le k)}$, $T_n = Y_1 + \cdots + Y_n$. 只需证 $T_n/n \to \mu$ a.s.

证明.

$$\sum_{k=1}^{\infty} P(|X_k| > k) \le \int_0^{\infty} P(|X_1| > t)\,\mathrm{d}t = E|X_1| < \infty$$

故 $P(X_k \ne Y_k \text{ i.o.}) = 0$, 其中 i.o. 表示无穷多次 (infinitely often). 这表明 $|S_n(\omega) - T_n(\omega)| \le R(\omega) < \infty$ a.s. 对所有 n 成立, 由此即得所需结果. □

接下来的步骤并不那么直观, 但它是本证明以及 2.5 节中证明的重要组成部分.

引理 2.4.3. $\sum_{k=1}^{\infty} \mathrm{var}(Y_k)/k^2 \le 4E|X_1| < \infty$.

证明. 为估计这个和, 注意到

$$\mathrm{var}(Y_k) \le E(Y_k^2) = \int_0^{\infty} 2yP(|Y_k| > y)\,\mathrm{d}y \le \int_0^{k} 2yP(|X_1| > y)\,\mathrm{d}y$$

因此利用 Fubini 定理 (因所有项 ≥ 0 且和是关于计数测度在 $\{1, 2, \ldots\}$ 上的积分)

$$\begin{aligned}\sum_{k=1}^{\infty} E(Y_k^2)/k^2 &\le \sum_{k=1}^{\infty} k^{-2} \int_0^{k} 2yP(|X_1| > y)\,\mathrm{d}y \\ &= \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} k^{-2} 1_{(y<k)} \right\} 2yP(|X_1| > y)\,\mathrm{d}y\end{aligned}$$

由于 $E|X_1| = \int_0^{\infty} P(|X_1| > y)\,\mathrm{d}y$, 只需证明

引理 2.4.4. 若 $y \ge 0$ 则 $2y \sum_{k>y} k^{-2} \le 4$.

证明. 首先注意到若 $m \ge 2$ 则

$$\sum_{k \ge m} k^{-2} \le \int_{m-1}^{\infty} x^{-2}\,\mathrm{d}x = (m-1)^{-1}$$

当 $y \geq 1$ 时，和式从 $k = [y] + 1 \geq 2$ 开始，故

$$\sum_{k>y} 2y\, k^{-2} \leq 2y/[y] \leq 4$$

因为 $y/[y] \leq 2$ 对 $y \geq 1$ 成立（最坏情形是 y 接近 2）. 为覆盖 $0 \leq y < 1$ 的情形，注意到此时

$$2y \sum_{k>y} k^{-2} \leq 2\left(1 + \sum_{k=2}^{\infty} k^{-2}\right) \leq 4$$

这就证明了引理 2.4.4，从而完成了引理 2.4.3 和定理的证明. □

□

前两步，即上面的引理 2.4.2 和引理 2.4.3，是标准的. Etemadi 的贡献在于：由于 X_n^+，$n \geq 1$ 和 X_n^-，$n \geq 1$ 都满足定理的条件且 $X_n = X_n^+ - X_n^-$，故**不失一般性**可假设 $X_n \geq 0$.

与定理 2.3.9 的证明类似，先对子序列证明结果，然后利用单调性控制中间值. 但这一次，取 $\alpha > 1$，$k(n) = [\alpha^n]$. Chebyshev 不等式蕴涵若 $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \varepsilon k(n)) \leq \varepsilon^{-2} \sum_{n=1}^{\infty} \text{var}(T_{k(n)})/k(n)^2$$

$$= \varepsilon^{-2} \sum_{n=1}^{\infty} \sum_{m=1}^{k(n)} k(n)^{-2} \,\text{var}(Y_m) = \varepsilon^{-2} \sum_{m=1}^{\infty} \text{var}(Y_m) \sum_{n:k(n)\geq m} k(n)^{-2}$$

其中用 Fubini 定理交换了两个非负项的求和顺序. 现在 $k(n) = [\alpha^n]$ 且 $[\alpha^n] \geq \alpha^n/2$ 对 $n \geq 1$，故对几何级数求和并注意第一项 $\leq m^{-2}$：

$$\sum_{n:\alpha^n \geq m} [\alpha^n]^{-2} \leq 4 \sum_{n:\alpha^n \geq m} \alpha^{-2n} \leq 4(1 - \alpha^{-2})^{-1} m^{-2}$$

综合上述计算得

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \varepsilon k(n)) \leq 4(1 - \alpha^{-2})^{-1} \varepsilon^{-2} \sum_{m=1}^{\infty} E(Y_m^2) m^{-2} < \infty$$

由引理 2.4.3. 由于 ε 任意，$(T_{k(n)} - ET_{k(n)})/k(n) \to 0$ a.s.

控制收敛定理蕴涵当 $k \to \infty$ 时，$EY_k \to EX_1$，故 $ET_{k(n)}/k(n) \to EX_1$ 且已证 $T_{k(n)}/k(n) \to EX_1$ a.s. 为处理中间值，注意到若 $k(n) \leq m < k(n+1)$

$$T_{k(n)}/k(n+1) \leq T_m/m \leq T_{k(n+1)}/k(n)$$

(这里用 $Y_i \geq 0$)，故回忆 $k(n) = [\alpha^n]$，有 $k(n+1)/k(n) \to \alpha$ 且

$$(1/\alpha)EX_1 \leq \liminf_{m\to\infty} T_m/m \leq \limsup_{m\to\infty} T_m/m \leq \alpha EX_1$$

由于 $\alpha > 1$ 任意，证毕. □

下一个结果表明只要 EX_i 存在，强大数定律就成立.

定理 2.4.5. 设 $X_1, X_2, \ldots$ 为**独立同分布**且 $EX_i^+ = \infty$，$EX_i^- < \infty$. 若 $S_n = X_1 + \cdots + X_n$，则 $S_n/n \to \infty$ a.s.

证明. 设 $M > 0$，$X_i^M = X_i \wedge M$. 这些 X_i^M 是**独立同分布**的且 $E|X_i^M| < \infty$，故若 $S_n^M = X_1^M + \cdots + X_n^M$，则定理 2.4.1 蕴涵 $S_n^M/n \to EX_i^M$. 由于 $X_i \geq X_i^M$，可知

$$\liminf_{n\to\infty} S_n/n \geq \lim S_n^M/n = EX_i^M$$

单调收敛定理蕴涵当 $M \uparrow \infty$ 时，$E(X_i^M)^+ \uparrow EX_i^+ = \infty$，故 $EX_i^M = E(X_i^M)^+ - E(X_i^M)^- \uparrow \infty$，从而 $\liminf_{n\to\infty} S_n/n \geq \infty$，这就蕴涵了所需结果. □

本节的其余部分致力于**强大数定律**的应用.

例题 2.4.6. 更新理论. 设 $X_1, X_2, \ldots$ 为**独立同分布**且 $0 < X_i < \infty$. 令 $T_n = X_1 + \cdots + X_n$ 并将 T_n 视为某事件第 n 次发生的时间. 为具体起见，考虑一位勤勉的**看门人**，他在**灯泡一烧坏**就立即更换. 假设第一个灯泡在时间 0 安装，X_i 为第 i 个灯泡的寿命. 在此解释下，T_n 是第 n 个灯泡烧坏的时间，$N_t = \sup\{n : T_n \leq t\}$ 是到时间 t 为止烧坏的灯泡数.

定理 2.4.7. 若 $EX_1 = \mu \leq \infty$，则当 $t \to \infty$ 时，

$$N_t/t \to 1/\mu \text{ a.s.} \quad (1/\infty = 0).$$

证明. 由定理 2.4.1 和 2.4.5，$T_n/n \to \mu$ a.s. 根据 N_t 的定义，$T(N_t) \leq t < T(N_t + 1)$，故除以 N_t 得

$$T(N_t)/N_t \leq t/N_t < T(N_t+1)/(N_t+1) \cdot (N_t+1)/N_t$$

为取极限，注意因 $T_n < \infty$ 对所有 n，有当 $t \to \infty$ 时，$N_t \uparrow \infty$. 强大数定律蕴涵对 $\omega \in \Omega_0$ 且 $P(\Omega_0) = 1$，有 $T_n(\omega)/n \to \mu$，$N_t(\omega) \uparrow \infty$，故

$$T_{N_t(\omega)}(\omega)/N_t(\omega) \to \mu, \quad (N_t(\omega)+1)/N_t(\omega) \to 1$$

由此可知对 $\omega \in \Omega_0$ 有 $t/N_t(\omega) \to \mu$ a.s. □

上述论证表明若 $X_n \to X_\infty$ a.s. 且 $N(n) \to \infty$ a.s. 则 $X_{N(n)} \to X_\infty$ a.s. 仔细写出这一点是因为对**依概率收敛**的类似结果是错误的. 若 $X_n \in \{0,1\}$ 独立且 $P(X_n = 1) = a_n \to 0$，$\sum a_n = \infty$，则 $X_n \to 0$ 依概率，但若令 $N(n) = \inf\{m \geq n : X_m = 1\}$，则 $X_{N(n)} = 1$ a.s.

例题 2.4.8. 经验分布函数. 设 $X_1, X_2, \ldots$ 为**独立同分布**且分布为 F，令

$$F_n(x) = n^{-1} \sum_{m=1}^{n} 1_{(X_m \leq x)}$$

$F_n(x)$ = 观测值 $\leq x$ 的频率，因此得名. 下一个结果表明当 $n \to \infty$ 时，F_n 一**致收敛**到 F.

定理 2.4.9. Glivenko-Cantelli 定理. 当 $n \to \infty$ 时，

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ a.s.}$$

证明. 固定 x，令 $Y_n = 1_{(X_n \leq x)}$. 因 Y_n 是**独立同分布**的且 $EY_n = P(X_n \leq x) = F(x)$，强大数定律蕴涵 $F_n(x) = n^{-1} \sum_{m=1}^{n} Y_m \to F(x)$ a.s. 一般而言，若 F_n 是**非减**函数序列且**逐点收敛**到有界连续极限 F，则 $\sup_x |F_n(x) - F(x)| \to 0$. 然而，分布函数 $F(x)$ 可能有跳跃点，故需做更多工作.

同样固定 x，令 $Z_n = 1_{(X_n < x)}$. 因 Z_n 是**独立同分布**的且 $EZ_n = P(X_n < x) = F(x-) = \lim_{y \uparrow x} F(y)$，强大数定律蕴涵 $F_n(x-) = n^{-1} \sum_{m=1}^{n} Z_m \to F(x-)$ a.s. 对 $1 \leq j \leq k-1$ 令 $x_{j,k} = \inf\{y : F(y) \geq j/k\}$. $F_n(x)$ 和 $F_n(x-)$ 的逐点收敛性表明可以选取 $N_k(\omega)$ 使得若 $n \geq N_k(\omega)$ 则

$$|F_n(x_{j,k}) - F(x_{j,k})| < k^{-1} \text{ 且} |F_n(x_{j,k}-) - F(x_{j,k}-)| < k^{-1}$$

对 $1 \leq j \leq k-1$. 若令 $x_{0,k} = -\infty$，$x_{k,k} = \infty$，则最后两个不等式对 $j = 0$ 或 k 也成立. 若 $x \in (x_{j-1,k}, x_{j,k})$ 且 $1 \leq j \leq k$，$n \geq N_k(\omega)$，则利用 F_n 和 F 的**单调性**，以及 $F(x_{j,k}-) - F(x_{j-1,k}) \leq k^{-1}$，有

$$F_n(x) \leq F_n(x_{j,k}-) \leq F(x_{j,k}-) + k^{-1} \leq F(x_{j-1,k}) + 2k^{-1} \leq F(x) + 2k^{-1}$$

$$F_n(x) \geq F_n(x_{j-1,k}) \geq F(x_{j-1,k}) - k^{-1} \geq F(x_{j,k}-) - 2k^{-1} \geq F(x) - 2k^{-1}$$

故 $\sup_x |F_n(x) - F(x)| \leq 2k^{-1}$，这就证明了结果. □

例题 2.4.10. Shannon 定理. 设 $X_1, X_2, \ldots \in \{1, \ldots, r\}$ 独立且 $P(X_n = k) = p(k) > 0$ 对 $1 \leq k \leq r$. 这里把 $1, \ldots, r$ 看作一个**字母表**的字母，$X_1, X_2, \ldots$ 是一个**信息源**连续产生的字母. 在此独立同分布情形下，这就是传说中的**打字机上的猴子**. 令 $\pi_n(\omega) = p(X_1(\omega)) \cdots p(X_n(\omega))$ 为前 n 次试验中观测到的**实现**的概率. 因 $\log \pi_n(\omega)$ 是独立随机变量之和，由强大数定律可知

$$-n^{-1} \log \pi_n(\omega) \to H \equiv -\sum_{k=1}^{r} p(k) \log p(k) \text{ a.s.}$$

常数 H 称为源的**熵**，是衡量其随机性的量. 最后一个结果是**渐近等分性**：若 $\varepsilon > 0$ 则当 $n \to \infty$ 时

$$P\{\exp(-n(H+\varepsilon)) \le \pi_n(\omega) \le \exp(-n(H-\varepsilon))\} \to 1$$

习题

习题 2.4.1. 懒惰的看门人. 假设第 i 个**灯泡**点亮 X_i 时间然后保持**烧坏**状态 Y_i 时间才被更换. 假设 X_i, Y_i 为正且独立，X 的分布为 F，Y 的分布为 G，两者都有有限均值. 令 R_t 为 $[0,t]$ 区间内有灯泡工作的时间. 证明 $R_t/t \to EX_i/(EX_i + EY_i)$ 几乎必然.

习题 2.4.2. 设 $X_0 = (1,0)$，归纳定义 $X_n \in \mathbb{R}^2$ 如下：X_{n+1} 从以原点为中心、半径 $|X_n|$ 的球中随机选取，即 $X_{n+1}/|X_n|$ 在半径为 1 的球上均匀分布且与 $X_1, \ldots, X_n$ 独立. 证明 $n^{-1}\log|X_n| \to c$ a.s. 并计算 c.

习题 2.4.3. 投资问题. 假设每年初可以购买面值 \$1 年底值 \$$a$ 的债券，或购买年底值随机量 $V_n \ge 0$ 的股票. 若总是将固定比例 p 的财富投资于债券，则第 $n+1$ 年底财富为 $W_{n+1} = (ap + (1-p)V_n)W_n$. 假设 $V_1, V_2, \ldots$ 为**独立同分布**且 $EV_n^2 < \infty$，$E(V_n^{-2}) < \infty$.

(i) 证明 $n^{-1}\log W_n \to c(p)$ a.s.

(ii) 证明 $c(p)$ 是凹函数. [用附录中的定理 A.5.1 证明在期望下求导的合理性.]

(iii) 通过考察 $c'(0)$ 和 $c'(1)$，给出关于 V 的条件以保证最优选择 p 在 $(0,1)$ 内.

(iv) 假设 $P(V=1) = P(V=4) = 1/2$. 求最优 p 作为 a 的函数.

第 5 节 随机级数的收敛 *

本节将采用第二种方法证明**强大数定律**，该方法基于随机级数的收敛. 这一方法的优势在于，在矩条件下，它能给出收敛速度的估计（定理 2.5.11 和 2.5.12），并能得到关于无限均值情形的否定性结果（定理 2.5.13），该结果比定理 2.3.8 中的结论更强. 本节的前两个结果本身就具有相当的独立意义，尽管我们将在引理 3.1.1 和定理 3.4.2 中看到更一般的版本.

为陈述第一个结果，需要引入一些记号. 设 $\mathcal{F}_n^0 = \sigma(X_n, X_{n+1}, \ldots)$ 为时刻 n 后的**未来**，即使得所有 X_m（$m \geq n$）均可测的最小 σ-域. 记 $\mathcal{T} = \bigcap_n \mathcal{F}_n^0$，称之为**尾 σ-域**. 直观上，$A \in \mathcal{T}$ 当且仅当改变有限个变量取值不影响该事件的发生. 照例，我们通过例子来帮助理解这一定义.

例题 2.5.1. 若 $B_n \in \mathcal{R}$，则 $\{X_n \in B_n \text{ i.o.}\} \in \mathcal{T}$. 若令 $X_n = 1_{A_n}$ 且 $B_n = \{1\}$，此例变为 $\{A_n \text{ i.o.}\}$.

例题 2.5.2. 设 $S_n = X_1 + \cdots + X_n$. 容易验证

$$\{\lim_{n\to\infty} S_n \text{ 存在}\} \in \mathcal{T},$$
$$\{\limsup_{n\to\infty} S_n > 0\} \notin \mathcal{T},$$
$$\{\limsup_{n\to\infty} S_n/c_n > x\} \in \mathcal{T}, \text{ 若} c_n \to \infty.$$

下一个结果表明上述例子全都是平凡的.

定理 2.5.3. Kolmogorov 0-1 律. 若 $X_1, X_2, \ldots$ 独立且 $A \in \mathcal{T}$，则 $P(A) = 0$ 或 1.

证明. 下面证明 A 与自身独立，即 $P(A \cap A) = P(A)P(A)$，于是 $P(A) = P(A)^2$，从而 $P(A) = 0$ 或 1. 分两步得出此结论：

(a) $A \in \sigma(X_1, \ldots, X_k)$ 与 $B \in \sigma(X_{k+1}, X_{k+2}, \ldots)$ 独立.

(a) 的证明. 若对某个 j 有 $B \in \sigma(X_{k+1}, \ldots, X_{k+j})$，这由定理 2.1.9 推出. 由于 $\sigma(X_1, \ldots, X_k)$ 与 $\bigcup_j \sigma(X_{k+1}, \ldots, X_{k+j})$ 都是包含 Ω 的 π-系，(a) 由定理 2.1.7 推出.

(b) $A \in \sigma(X_1, X_2, \ldots)$ 与 $B \in \mathcal{T}$ 独立.

(b) 的证明. 由于 $\mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, \ldots)$，若对某个 k 有 $A \in \sigma(X_1, \ldots, X_k)$，这由 (a) 推出. $\bigcup_k \sigma(X_1, \ldots, X_k)$ 与 $\mathcal{T}$ 都是包含 Ω 的 π-系，故 (b) 由定理 2.1.7 推出.

由于 $\mathcal{T} \subset \sigma(X_1, X_2, \ldots)$，(b) 蕴含 $A \in \mathcal{T}$ 与自身独立，从而定理 2.5.3 得证. □

在进入主题之前，先证明一个 0-1 律，它在独立同分布情形下推广了 Kolmogorov 的 0-1 律. 为陈述新的 0-1 律，需要两个定义. 称 $\mathbb{N} = \{1, 2, \ldots\}$ 的一个**有限置换**是指从 $\mathbb{N}$ 到 $\mathbb{N}$ 的映射 π，使得 $\pi(i) \neq i$ 只对有限个 i 成立. 若 π 是 $\mathbb{N}$ 的有限置换且 $\omega \in S^{\mathbb{N}}$，定义 $(\pi\omega)_i = \omega_{\pi(i)}$. 换言之，$\omega$ 的坐标按 π 重排. 由于 $X_i(\omega) = \omega_i$，这与重排随机变量相同. 事件 A 称为**可置换的**，若对任意有限置换 π，$\pi^{-1}A \equiv \{\omega : \pi\omega \in A\}$ 等于 A；换言之，若重排有限个随机变量不影响该事件的发生. 可置换事件的全体构成一个 σ-域，称为**可交换 σ-域**，记作 $\mathcal{E}$.

为说明可置换事件的意义，设 $S=\mathbb{R}$ 并令 $S_n(\omega)=X_1(\omega)+\cdots+X_n(\omega)$. 两个可置换事件的例子是

(i) $\{\omega: S_n(\omega)\in B \text{ i.o.}\}$

(ii) $\{\omega: \limsup_{n\to\infty} S_n(\omega)/c_n \geq 1\}$

每种情形下，该事件都是可置换的，因为对充分大的 n 有 $S_n(\omega)=S_n(\pi\omega)$. 注意到

(iii) 尾 σ-域 $\mathcal{T}$ 中的所有事件都是可置换的.

为验证此点，注意到若 $A\in\sigma(X_{n+1},X_{n+2},\ldots)$，则 A 的发生不受 $X_1,\ldots,X_n$ 置换的影响. (i) 表明 (iii) 的逆命题不成立. 下一个结果表明，对于独立同分布序列，$\mathcal{E}$ 与 $\mathcal{T}$ 没有区别——它们都是平凡的.

定理 2.5.4. Hewitt-Savage 0-1 律. 若 $X_1,X_2,\ldots$ 独立同分布且 $A\in\mathcal{E}$，则 $P(A)\in\{0,1\}$.

证明. 设 $A\in\mathcal{E}$. 如同 Kolmogorov 0-1 律的证明，下面证明 A 与自身独立，即 $P(A)=P(A\cap A)=P(A)P(A)$，于是 $P(A)\in\{0,1\}$. 设 $A_n\in\sigma(X_1,\ldots,X_n)$ 使得

(a) $P(A_n\triangle A)\to 0$

这里 $A\triangle B=(A-B)\cup(B-A)$ 为**对称差**. A_n 的存在性由引理 A.2.1 的第二部分保证. A_n 可写成 $\{\omega:(\omega_1,\ldots,\omega_n)\in B_n\}$，其中 $B_n\in\mathcal{S}^n$. 令

$$\pi(j)=\begin{cases} j+n & \text{若} 1\leq j\leq n,\\ j-n & \text{若} n+1\leq j\leq 2n,\\ j & \text{若} j\geq 2n+1.\end{cases}$$

注意到 π^2 是恒等映射（故无需担心写 π 还是 π^{-1}），且坐标独立同分布（故置换后的坐标亦然），于是

(b) $P(\omega:\omega\in A_n\triangle A)=P(\omega:\pi\omega\in A_n\triangle A)$

现在 $\{\omega:\pi\omega\in A\}=\{\omega:\omega\in A\}$，因为 A 可置换，而

$$\{\omega:\pi\omega\in A_n\}=\{\omega:(\omega_{n+1},\ldots,\omega_{2n})\in B_n\}$$

若用 A_n' 表示最后一个事件，则

(c) $\{\omega:\pi\omega\in A_n\triangle A\}=\{\omega:\omega\in A_n'\triangle A\}$

结合 (b) 和 (c) 得

(d) $P(A_n\triangle A)=P(A_n'\triangle A)$

易见 $|P(B)-P(C)|\leq P(B\triangle C)$，故 (d) 蕴含 $P(A_n),P(A_n')\to P(A)$. 由 $A-C\subset(A-B)\cup(B-C)$ 及关于 $C-A$ 的类似不等式，有 $A\triangle C\subset(A\triangle B)\cup(B\triangle C)$. 由最后的不等式、(d) 和 (a) 得

$$P(A_n\triangle A_n')\leq P(A_n\triangle A)+P(A\triangle A_n')\to 0$$

最后的估计蕴含

$$0 \le P(A_n) - P(A_n \cap A_n') \le P(A_n \cup A_n') - P(A_n \cap A_n') = P(A_n \triangle A_n') \to 0$$

于是 $P(A_n \cap A_n') \to P(A)$. 但 A_n 与 A_n' 独立，故

$$P(A_n \cap A_n') = P(A_n)P(A_n') \to P(A)^2$$

这表明 $P(A) = P(A)^2$，定理 2.5.4 得证. □

若 $A_1, A_2, \ldots$ 独立，则定理 2.5.3 蕴含 $P(A_n \text{ i.o.}) = 0$ 或 1. 对例 2.5.2 应用定理 2.5.3 就得到 $P(\lim_{n\to\infty} S_n \text{ 存在}) = 0$ 或 1. 下一个结果将有助于在某些情形下证明概率为 1.

定理 2.5.5. Kolmogorov 极大不等式. 设 $X_1, \ldots, X_n$ 独立，$EX_i = 0$ 且 $\text{var}(X_i) < \infty$. 若 $S_n = X_1 + \cdots + X_n$，则

$$P\left(\max_{1\le k\le n} |S_k| \ge x\right) \le x^{-2}\,\text{var}(S_n).$$

注释. 在相同假设下，Chebyshev 不等式（定理 1.6.4）仅给出

$$P(|S_n| \ge x) \le x^{-2}\,\text{var}(S_n).$$

证明. 令 $A_k = \{|S_k| \ge x \text{ 但} |S_j| < x \text{ 对} j < k\}$，即按 $|S_k|$ 首次超过 x 的时间分解. 由于 A_k 不交且 $(S_n - S_k)^2 \ge 0$,

$$ES_n^2 \ge \sum_{k=1}^n \int_{A_k} S_n^2\,\mathrm{d}P = \sum_{k=1}^n \int_{A_k} S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\,\mathrm{d}P$$

$$\ge \sum_{k=1}^n \int_{A_k} S_k^2\,\mathrm{d}P + \sum_{k=1}^n \int_{A_k} 2S_k 1_{A_k} \cdot (S_n - S_k)\,\mathrm{d}P$$

$S_k 1_{A_k} \in \sigma(X_1, \ldots, X_k)$ 与 $S_n - S_k \in \sigma(X_{k+1}, \ldots, X_n)$ 由定理 2.1.10 知独立，故利用定理 2.1.13 及 $E(S_n - S_k) = 0$ 得

$$\int_{A_k} 2S_k 1_{A_k} \cdot (S_n - S_k)\,\mathrm{d}P = E(2S_k 1_{A_k}) \cdot E(S_n - S_k) = 0$$

利用在 A_k 上有 $|S_k| \ge x$ 且 A_k 不交,

$$ES_n^2 \ge \sum_{k=1}^n \int_{A_k} S_k^2\,\mathrm{d}P \ge x^2 \sum_{k=1}^n P(A_k) = x^2 P\left(\max_{1\le k\le n} |S_k| \ge x\right)$$

□

现在转向级数收敛的结果. 为陈述它们，需要一个定义. 称 $\sum_{n=1}^{\infty} a_n$ **收敛**，若 $\lim_{N\to\infty}\sum_{n=1}^{N} a_n$ 存在.

定理 2.5.6. 设 $X_1, X_2, \ldots$ 独立且 $EX_n = 0$. 若

$$\sum_{n=1}^{\infty} \mathrm{var}(X_n) < \infty,$$

则 $\sum_{n=1}^{\infty} X_n(\omega)$ 几乎必然收敛.

证明. 设 $S_N = \sum_{n=1}^{N} X_n$. 由定理 2.5.5 得

$$P\left(\max_{M\le m\le N} |S_m - S_M| > \varepsilon\right) \le \varepsilon^{-2}\,\mathrm{var}(S_N - S_M) = \varepsilon^{-2}\sum_{n=M+1}^{N} \mathrm{var}(X_n)$$

令 $N \to \infty$，得

$$P\left(\sup_{m\ge M} |S_m - S_M| > \varepsilon\right) \le \varepsilon^{-2}\sum_{n=M+1}^{\infty} \mathrm{var}(X_n) \to 0 \text{ 当} M \to \infty$$

令 $w_M = \sup_{m,n\ge M} |S_m - S_n|$，则当 $M\uparrow$ 时 $w_M \downarrow$，且

$$P(w_M > 2\varepsilon) \le P\left(\sup_{m\ge M} |S_m - S_M| > \varepsilon\right) \to 0$$

当 $M \to \infty$，于是 $w_M \downarrow 0$ 几乎必然. 但 $w_M(\omega) \downarrow 0$ 蕴含 $S_n(\omega)$ 是 Cauchy 序列，从而 $\lim_{n\to\infty} S_n(\omega)$ 存在. 证毕. □

例题 2.5.7. 设 $X_1, X_2, \ldots$ 独立且

$$P(X_n = n^{-\alpha}) = P(X_n = -n^{-\alpha}) = 1/2.$$

此时 $EX_n = 0$ 且 $\mathrm{var}(X_n) = n^{-2\alpha}$，故若 $\alpha > 1/2$，由定理 2.5.6 知 $\sum_n X_n$ 收敛. 定理 2.5.8 表明 $\alpha > 1/2$ 也是该结论成立的必要条件. 注意到 $\sum_n |X_n| < \infty$ 当且仅当 $\alpha > 1$.

定理 2.5.6 足以应对所有应用，但若不提随机级数收敛的完整结论，本节将显得不够完整.

定理 2.5.8. Kolmogorov 三级数定理. 设 $X_1, X_2, \ldots$ 独立. 令 $A > 0$ 且 $Y_n = X_n 1_{(|X_n|\le A)}$. 为使

$\sum_{n=1}^{\infty} X_n$ 几乎必然收敛，必要且充分的条件是

$$\sum_{n=1}^{\infty} P(|X_n| > A) < \infty, \quad \sum_{n=1}^{\infty} EY_n \text{ 收敛}, \quad \sum_{n=1}^{\infty} \text{var}(Y_n) < \infty.$$

证明. 必要性的证明将在例 3.4.12 中作为中心极限定理的应用给出. 为证充分性，令 $\mu_n = EY_n$. (iii) 与定理 2.5.6 蕴含 $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ 几乎必然收敛. 利用 (ii) 得 $\sum_{n=1}^{\infty} Y_n$ 几乎必然收敛. (i) 与 Borel-Cantelli 引理蕴含 $P(X_n \neq Y_n \text{ i.o.}) = 0$，故 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛. □

级数收敛与强大数定律之间的联系由如下结果给出.

定理 2.5.9. Kronecker 引理. 若 $a_n \uparrow \infty$ 且 $\sum_{n=1}^{\infty} x_n/a_n$ 收敛，则

$$a_n^{-1} \sum_{m=1}^{n} x_m \to 0.$$

证明. 令 $a_0 = 0$，$b_0 = 0$，且对 $m \geq 1$，$b_m = \sum_{k=1}^{m} x_k/a_k$. 则 $x_m = a_m(b_m - b_{m-1})$，于是

$$\begin{aligned}
\sum_{m=1}^{n} a_n^{-1} x_m &= \sum_{m=1}^{n} a_n^{-1} a_m b_m - a_n^{-1} a_m b_{m-1} \\
&= a_n^{-1} a_n b_n + \sum_{m=2}^{n} a_n^{-1} a_{m-1} b_{m-1} - \sum_{m=1}^{n} a_n^{-1} a_m b_{m-1} \\
&= b_n - \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1}
\end{aligned}$$

（回忆 $a_0 = 0$. ）由假设，当 $n \to \infty$ 时 $b_n \to b_\infty$. 由于 $a_m - a_{m-1} \geq 0$，最后一个和是 $b_0, \ldots, b_n$ 的平均. 直观上，若固定 $\varepsilon > 0$ 与 $M < \infty$ 且 n 充分大，该平均对 $m \geq M$ 的 b_m 赋予质量 $\geq 1 - \varepsilon$，故

$$\sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1} \to b_\infty$$

为严格论证，令 $B = \sup_n |b_n|$，选取 M 使得对 $m \geq M$ 有 $|b_m - b_\infty| < \varepsilon/2$，再选取 N 使得对 $n \geq N$ 有 $a_M/a_n < \varepsilon/4B$. 若 $n \geq N$，有

$$\begin{aligned}
\left| \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1} - b_\infty \right| &\leq \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} |b_{m-1} - b_\infty| \\
&\leq \frac{a_M}{a_n} \cdot 2B + \frac{a_n - a_M}{a_n} \cdot \frac{\varepsilon}{2} < \varepsilon
\end{aligned}$$

由于 ε 任意，得证. □

定理 2.5.10. 强大数定律. 设 $X_1, X_2, \ldots$ 独立同分布且 $E|X_i| < \infty$. 令 $EX_i = \mu$，$S_n = X_1 + \cdots + X_n$. 则当 $n \to \infty$ 时 $S_n/n \to \mu$ 几乎必然.

证明. 令 $Y_k = X_k 1_{(|X_k| \le k)}$ 且 $T_n = Y_1 + \cdots + Y_n$. 由定理 2.4.1 证明中的 (a)，只需证 $T_n/n \to \mu$. 令 $Z_k = Y_k - EY_k$，则 $EZ_k = 0$. 注意 $\text{var}(Z_k) = \text{var}(Y_k) \le EY_k^2$，且定理 2.4.1 证明中的 (b) 蕴含

$$\sum_{k=1}^{\infty} \text{var}(Z_k)/k^2 \le \sum_{k=1}^{\infty} EY_k^2/k^2 < \infty$$

现在应用定理 2.5.6，断定 $\sum_{k=1}^{\infty} Z_k/k$ 几乎必然收敛，于是定理 2.5.9 给出

$$n^{-1}\sum_{k=1}^{n}(Y_k - EY_k) \to 0 \text{ 且} \frac{T_n}{n} - n^{-1}\sum_{k=1}^{n} EY_k \to 0 \text{ 几乎必然}$$

控制收敛定理蕴含当 $k \to \infty$ 时 $EY_k \to \mu$. 由此易得

$$n^{-1}\sum_{k=1}^{n} EY_k \to \mu \text{ 从而} T_n/n \to \mu. \qquad \square$$

2.5.1 收敛速度

如前面所提及，随机级数证明法的优势之一在于它给出了 $S_n/n \to \mu$ 的收敛速度估计. 从每个随机变量减去 μ，可以且将**不失一般性**假设 $\mu = 0$.

定理 2.5.11. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$ 且 $EX_i^2 = \sigma^2 < \infty$. 令 $S_n = X_1 + \cdots + X_n$. 若 $\varepsilon > 0$，则

$$S_n/n^{1/2}(\log n)^{1/2+\varepsilon} \to 0 \text{ 几乎必然.}$$

注释. 重对数律（定理 8.5.2）将表明

$$\limsup_{n\to\infty} S_n/n^{1/2}(\log\log n)^{1/2} = \sigma\sqrt{2} \text{ 几乎必然},$$

故上述结果与最优结果相去不远.

证明. 对 $n \ge 2$ 令 $a_n = n^{1/2}(\log n)^{1/2+\varepsilon}$ 且 $a_1 > 0$.

$$\sum_{n=1}^{\infty} \text{var}(X_n/a_n) = \frac{\sigma^2}{a_1^2} + \sum_{n=2}^{\infty} \frac{1}{n(\log n)^{1+2\varepsilon}} < \infty,$$

于是应用定理 2.5.6 得 $\sum_{n=1}^{\infty} X_n/a_n$ 几乎必然收敛，从而由定理 2.5.9 即得所述结果. $\square$

下一个结果属于 Marcinkiewicz 和 Zygmund，处理 $EX_i^2=\infty$ 但 $E|X_i|^p<\infty$（其中 $1<p<2$）的情形.

定理 2.5.12. 设 $X_1,X_2,\ldots$ 独立同分布，$EX_1=0$ 且 $E|X_1|^p<\infty\,(1<p<2)$. 若 $S_n=X_1+\cdots+X_n$，则 $S_n/n^{1/p}\to 0$ 几乎必然.

证明. 令 $Y_k=X_k1_{(|X_k|\le k^{1/p})}$ 且 $T_n=Y_1+\cdots+Y_n$.

$$\sum_{k=1}^{\infty}P(Y_k\neq X_k)=\sum_{k=1}^{\infty}P(|X_k|^p>k)\le E|X_k|^p<\infty,$$

于是 Borel-Cantelli 引理蕴含 $P(Y_k\neq X_k \text{ i.o.})=0$，故只需证 $T_n/n^{1/p}\to 0$. 利用 $\text{var}(Y_m)\le E(Y_m^2)$，引理 2.2.13（取 $p=2$），$P(|Y_m|>y)\le P(|X_1|>y)$，及 Fubini 定理（所有项 ≥ 0），有

$$\begin{aligned}\sum_{m=1}^{\infty}\text{var}(Y_m/m^{1/p})&\le\sum_{m=1}^{\infty}EY_m^2/m^{2/p}\\&\le\sum_{m=1}^{\infty}\sum_{n=1}^{m}\int_{(n-1)^{1/p}}^{n^{1/p}}\frac{2y}{m^{2/p}}P(|X_1|>y)\,\mathrm{d}y\\&=\sum_{n=1}^{\infty}\int_{(n-1)^{1/p}}^{n^{1/p}}\sum_{m=n}^{\infty}\frac{2y}{m^{2/p}}P(|X_1|>y)\,\mathrm{d}y\end{aligned}$$

为估计积分，注意到对 $n\ge 2$，比较和与 $x^{-2/p}$ 的积分得

$$\sum_{m=n}^{\infty}m^{-2/p}\le\frac{p}{2-p}(n-1)^{(p-2)/p}\le Cy^{p-2}$$

当 $y\in[(n-1)^{1/p},n^{1/p}]$. 由于 $E|X_i|^p=\int_0^{\infty}px^{p-1}P(|X_i|>x)\,\mathrm{d}x<\infty$，得

$$\sum_{m=1}^{\infty}\text{var}(Y_m/m^{1/p})<\infty$$

若令 $\mu_m=EY_m$ 并应用定理 2.5.6 与定理 2.5.9，则

$$n^{-1/p}\sum_{m=1}^{n}(Y_m-\mu_m)\to 0 \text{ 几乎必然}$$

为估计 μ_m，注意到由于 $EX_i=0$，有 $\mu_m=-E(X_i;|X_i|>m^{1/p})$，故

$$\begin{aligned}|\mu_m| &\le E(|X_i|;|X_i|>m^{1/p}) = m^{1/p}E(|X_i|/m^{1/p};|X_i|>m^{1/p})\\ &\le m^{1/p}E((|X_i|/m^{1/p})^p;|X_i|>m^{1/p})\\ &\le m^{-1+1/p}p^{-1}E(|X_i|^p;|X_i|>m^{1/p})\end{aligned}$$

现 $\sum_{m=1}^n m^{-1+1/p}\le Cn^{1/p}$ 且当 $m\to\infty$ 时 $E(|X_i|^p;|X_i|>m^{1/p})\to 0$，故

$$n^{-1/p}\sum_{m=1}^n \mu_m \to 0$$

从而得证. □

2.5.2 无限均值

St. Petersburg 博弈(例 2.2.16 与习题 2.3.20 中讨论过)是 $EX_i=\infty$ 的例子,此时 $S_n/(n\log_2 n)\to 1$ 依概率，但

$$\limsup_{n\to\infty} S_n/(n\log_2 n)=\infty \text{ 几乎必然.}$$

下一个结果属于 Feller（1946），表明当 $E|X_1|=\infty$ 时，S_n/a_n 不能几乎必然收敛于非零极限. 在定理 2.3.8 中我们考虑了 $a_n=n$ 的特殊情形.

定理 2.5.13. 设 $X_1,X_2,\dots$ 独立同分布且 $E|X_1|=\infty$，$S_n=X_1+\cdots+X_n$. 令 a_n 为正数列且 a_n/n 递增. 则 $\limsup_{n\to\infty}|S_n|/a_n=0$ 或 ∞，视 $\sum_{n=1}^\infty P(|X_1|\ge a_n)<\infty$ 或 $=\infty$ 而定.

证明. 由于 $a_n/n\uparrow$，对任意整数 k 有 $a_{kn}\ge ka_n$. 利用此点及 $a_n\uparrow$ 得

$$\sum_{n=1}^\infty P(|X_1|\ge ka_n)\ge\sum_{n=1}^\infty \frac{1}{k}P(|X_1|\ge a_{kn})\ge\sum_{m=k}^\infty P(|X_1|\ge a_m),$$

最后一个观察表明若和为无穷，则 $\limsup_{n\to\infty}|X_n|/a_n=\infty$. 由于 $\max\{|S_{n-1}|,|S_n|\}\ge|X_n|/2$，得 $\limsup_{n\to\infty}|S_n|/a_n=\infty$.

为证另一半，由恒等式

$$\sum_{m=1}^\infty mP(a_{m-1}\le|X_i|<a_m)=\sum_{n=1}^\infty P(|X_i|\ge a_{n-1}) \tag{$*$}$$

为此，写 $m=\sum_{n=1}^m 1$，然后用 Fubini 定理. 现令 $Y_n=X_n1_{(|X_n|<a_n)}$，$T_n=Y_1+\cdots+Y_n$. 当和有限

时，$P(Y_n \neq X_n \text{ i.o.}) = 0$，故只需研究 T_n 的行为. 为此，令 $a_0 = 0$ 并计算

$$\begin{aligned}\sum_{n=1}^{\infty} \operatorname{var}(Y_n/a_n) &\leq \sum_{n=1}^{\infty} EY_n^2/a_n^2 \\ &= \sum_{n=1}^{\infty}\sum_{m=1}^{n} a_n^{-2} \int_{[a_{m-1},a_m)} y^2 \,\mathrm{d}F(y) \\ &= \sum_{m=1}^{\infty} \int_{[a_{m-1},a_m)} y^2 \,\mathrm{d}F(y) \sum_{n=m}^{\infty} a_n^{-2}\end{aligned}$$

由于 $a_n \geq na_m/m$，有 $\sum_{n=m}^{\infty} a_n^{-2} \leq (m^2/a_m^2)\sum_{n=m}^{\infty} n^{-2} \leq Cma_m^{-2}$，于是

$$\leq C\sum_{m=1}^{\infty} m \int_{[a_{m-1},a_m)} \mathrm{d}F(y)$$

现用 $(*)$ 得 $\sum_{n=1}^{\infty} \operatorname{var}(Y_n/a_n) < \infty$.

最后一步是证 $ET_n/a_n \to 0$. 首先，注意到若 $E|X_i| = \infty$，$\sum_{n=1}^{\infty} P(|X_i| > a_n) < \infty$，且 $a_n/n \uparrow$，则必有 $a_n/n \uparrow \infty$. 为估计 ET_n/a_n，注意到

$$\begin{aligned}\left| a_n^{-1} \sum_{m=1}^{n} EY_m \right| &\leq a_n^{-1} n \cdot E(|X_i|; |X_i| < a_n) \\ &\leq \frac{na_N}{a_n} + \frac{n}{a_n} E(|X_i|; a_N \leq |X_i| < a_n)\end{aligned}$$

其中最后一个不等式对任意固定的 N 成立. 由于 $a_n/n \to \infty$，第一项趋于 0. 由于 $m/a_m \downarrow$，第二项是

$$\leq \sum_{m=N+1}^{n} \frac{m}{a_m} E(|X_i|; a_{m-1} \leq |X_i| < a_m) \leq \sum_{m=N+1}^{\infty} mP(a_{m-1} \leq |X_i| < a_m)$$

$(*)$ 表明该和有限，故当 N 充分大时它很小，从而得证. □

习题

习题 2.5.1. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$，$\operatorname{var}(X_i) = C < \infty$. 对 $n = m^\alpha$ 用定理 2.5.5，其中 $\alpha(2p-1) > 1$，推出若 $S_n = X_1 + \cdots + X_n$ 且 $p > 1/2$，则 $S_n/n^p \to 0$ 几乎必然.

习题 2.5.2. 定理 2.5.12 的逆命题要简单得多. 设 $p > 0$. 若 $S_n/n^{1/p} \to 0$ 几乎必然，则 $E|X_1|^p < \infty$.

习题 2.5.3. 设 $X_1, X_2, \ldots$ 独立同分布于标准正态. 证明对任意 t

$$\sum_{n=1}^{\infty} X_n \cdot \frac{\sin(n\pi t)}{n} \text{ 几乎必然收敛.}$$

我们将在 8.1 节末尾再次见到这个级数.

习题 2.5.4. 设 $X_1, X_2, \dots$ 独立，$EX_n = 0$，$\mathrm{var}(X_n) = \sigma_n^2$.

(i) 证明若 $\sum_n \sigma_n^2/n^2 < \infty$，则 $\sum_n X_n/n$ 几乎必然收敛，从而 $n^{-1}\sum_{m=1}^n X_m \to 0$ 几乎必然.

(ii) 设 $\sum_n \sigma_n^2/n^2 = \infty$ 且**不失一般性**假设对所有 n 有 $\sigma_n^2 \le n^2$. 证明存在独立随机变量 X_n, $EX_n = 0$ 且 $\mathrm{var}(X_n) \le \sigma_n^2$，使得 X_n/n 从而 $n^{-1}\sum_{m\le n} X_m$ 不几乎必然收敛于 0.

习题 2.5.5. 设 $X_n \ge 0$ 对 $n \ge 1$ 独立. 下列等价：

(i) $\sum_{n=1}^\infty X_n < \infty$ 几乎必然.

(ii) $\sum_{n=1}^\infty [P(X_n > 1) + E(X_n 1_{(X_n \le 1)})] < \infty$.

(iii) $\sum_{n=1}^\infty E(X_n/(1+X_n)) < \infty$.

习题 2.5.6. 设 $\psi(x) = x^2$ 当 $|x| \le 1$，$= |x|$ 当 $|x| \ge 1$. 证明若 $X_1, X_2, \dots$ 独立，$EX_n = 0$ 且 $\sum_{n=1}^\infty E\psi(X_n) < \infty$，则 $\sum_{n=1}^\infty X_n$ 几乎必然收敛.

习题 2.5.7. 设 X_n 独立. 设 $\sum_{n=1}^\infty E|X_n|^{p(n)} < \infty$，其中对所有 n 有 $0 < p(n) \le 2$，且当 $p(n) > 1$ 时 $EX_n = 0$. 证明 $\sum_{n=1}^\infty X_n$ 几乎必然收敛.

习题 2.5.8. 设 $X_1, X_2, \dots$ 独立同分布且不恒为 0. 则幂级数 $\sum_{n\ge1} X_n(\omega) z^n$ 的**收敛半径**（即 $r(\omega) = \sup\{c : \sum_n |X_n(\omega)| c^n < \infty\}$）几乎必然为 1 或 0，视 $E\log^+|X_1| < \infty$ 或 $= \infty$ 而定，其中 $\log^+ x = \max(\log x, 0)$.

习题 2.5.9. 设 $X_1, X_2, \dots$ 独立，$S_{m,n} = X_{m+1} + \cdots + X_n$. 则

$$P\left(\max_{m<j\le n} |S_{m,j}| > 2a\right) \min_{m<k\le n} P(|S_{k,n}| \le a) \le P(|S_{m,n}| > a). \tag{$\star$}$$

习题 2.5.10. 用 ($\star$) 证明 P. Lévy 的一个定理：设 $X_1, X_2, \dots$ 独立，$S_n = X_1 + \cdots + X_n$. 若 $\lim_{n\to\infty} S_n$ 依概率存在，则它也几乎必然存在.

习题 2.5.11. 设 $X_1, X_2, \dots$ 独立同分布，$S_n = X_1 + \cdots + X_n$. 用 ($\star$) 推出若 $S_n/n \to 0$ 依概率，则 $(\max_{1\le m\le n} S_m)/n \to 0$ 依概率.

习题 2.5.12. 设 $X_1, X_2, \dots$ 独立同分布，$S_n = X_1 + \cdots + X_n$. 设 $a_n \uparrow \infty$ 且 a_{2n}/a_{2n-1} 有界.

(i) 用 ($\star$) 证明若 $S_n/a_n \to 0$ 依概率且 $S_{2n}/a_{2n} \to 0$ 几乎必然，则 $S_n/a_n \to 0$ 几乎必然.

(ii) 进一步设 $EX_1 = 0$ 且 $EX_1^2 < \infty$. 用前一习题及 Chebyshev 不等式推出 $S_n/n^{1/2}(\log n)^{1/2+\varepsilon} \to 0$ 几乎必然.

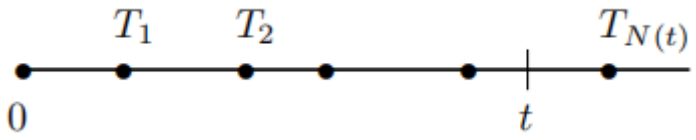

图 2.1: 更新序列

第 6 节 更新理论 *

设 $\xi_1, \xi_2, \ldots$ 为独立同分布的正随机变量（即 $P(\xi_i > 0) = 1$），分布函数为 F，并按如下方式定义时刻序列：$T_0 = 0$，且对 $k \geq 1$ 有 $T_k = T_{k-1} + \xi_k$. 如例 2.4.6 所述，可将 ξ_i 理解为第 i 个灯泡的寿命，T_k 则是第 k 个灯泡烧坏的时刻. 按照第 3.6 节中关于泊松过程的讨论，还有第二种解释：T_k 是第 k 位顾客到达的时刻. 第三种解释来自第 5 章：T_k 是第 k 次访问某状态的时刻.

为使用中性术语，称 T_k 为**更新**. 更新一词指的是该过程在 T_k 处“重新开始”，即 $\{T_{k+j} - T_k, j \geq 1\}$ 与 $\{T_j, j \geq 1\}$ 同分布.

与例 2.4.6 的记号略有不同，记 $N(t) = \inf\{k : T_k > t\}$. $N(t)$ 是区间 $[0, t]$ 中的更新次数，计入时刻 0 处的更新. 这一定义的优点是 $N(t)$ 是一个停时，即 $\{N(t) = k\}$ 关于 $\mathcal{F}_k$ 可测.

在定理 2.4.7 中，已经证明了

定理 2.6.1. 当 $t \to \infty$ 时，$N(t)/t \to 1/\mu$ 几乎必然，其中 $\mu = E\xi_i \in (0, \infty]$，并规定 $1/\infty = 0$.

下面的结果涉及 $U(t) = EN(t)$ 的渐近行为. 为此，需要

定理 2.6.2. Wald 等式. 设 $X_1, X_2, \ldots$ 独立同分布且 $E|X_i| < \infty$. 若 N 为停时且 $EN < \infty$，则 $ES_N = EX_1 \cdot EN$.

证明. 先设 $X_i \geq 0$.

$$ES_N = \int S_N \, \mathrm{d}P = \sum_{n=1}^{\infty} \int S_n 1_{\{N=n\}} \, \mathrm{d}P = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \int X_m 1_{\{N=n\}} \, \mathrm{d}P$$

由于 $X_i \geq 0$，可以交换求和次序（即使用 Fubini 定理）推出上式

$$= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int X_m 1_{\{N=n\}} \, \mathrm{d}P = \sum_{m=1}^{\infty} \int X_m 1_{\{N \geq m\}} \, \mathrm{d}P$$

注意 $\{N \geq m\} = \{N \leq m-1\}^c \in \mathcal{F}_{m-1}$，且与 X_m 独立，故上式

$$= \sum_{m=1}^{\infty} EX_m P(N \geq m) = EX_1 \cdot EN$$

为证明一般情形，逆向运行上述论证. 若 $EN<\infty$，则

$$\infty>\sum_{m=1}^{\infty}E|X_m|P(N\geq m)=\sum_{m=1}^{\infty}\sum_{n=m}^{\infty}\int|X_m|1_{\{N=n\}}\,\mathrm{d}P$$

最后一个公式表明二重和按一种顺序绝对收敛，故 Fubini 定理给出

$$\sum_{m=1}^{\infty}\sum_{n=m}^{\infty}\int X_m1_{\{N=n\}}\,\mathrm{d}P=\sum_{n=1}^{\infty}\sum_{m=1}^{n}\int X_m1_{\{N=n\}}\,\mathrm{d}P$$

利用 $\{N\geq m\}\in\mathcal{F}_{m-1}$ 与 X_m 的独立性，并重写最后一个恒等式，可得

$$\sum_{m=1}^{\infty}EX_mP(N\geq m)=ES_N$$

由于左边等于 $EN\cdot EX_1$，证毕. □

定理 2.6.3. 当 $t\to\infty$ 时，$U(t)/t\to1/\mu$.

证明. 对停时 $N(t)$ 应用 Wald 等式. 第一步是证明 $EN(t)<\infty$. 为此，取 $\delta>0$ 使得 $P(\xi_i>\delta)=\varepsilon>0$，并取 K 使得 $K\delta\geq t$. 由于 K 个连续的 $\xi_i>\delta$ 将使 $T_n>t$，故

$$P(N(t)>mK)\leq(1-\varepsilon^K)^m$$

从而 $EN(t)<\infty$. 若 $\mu<\infty$，应用 Wald 等式即得

$$\mu EN(t)=ET_{N(t)}\geq t$$

所以 $U(t)\geq t/\mu$. 当 $\mu=\infty$ 时最后的不等式显然成立，故该不等式普遍成立.

对于上界，注意若 $P(\xi_i\leq c)=1$，则重复上述论证可得 $\mu EN(t)=ES_{N(t)}\leq t+c$，于是对有界分布结论成立. 若令 $\bar{\xi}_i=\xi_i\wedge c$，并以显然的方式定义 $\bar{T}_n$ 和 $\bar{N}(t)$，则

$$EN(t)\leq E\bar{N}(t)\leq(t+c)/E(\bar{\xi}_i)$$

令 $t\to\infty$ 再令 $c\to\infty$，得 $\limsup_{t\to\infty}EN(t)/t\leq1/\mu$，证毕. □

为更仔细地研究更新发生的时刻，记

$$U(A)=\sum_{n=0}^{\infty}P(T_n\in A)$$

称 U 为**更新测度**. 将旧定义 $U(t)=EN(t)$ 纳入新定义，把 $U(t)$ 理解为 $U([0,t])$ 的简写. 这应不会引起问题，因为 $U(t)$ 是更新测度的分布函数. $U(t)$ 的渐近行为取决于分布 F 是**算术分布**（即集中在 $\{\delta,2\delta,3\delta,\ldots\}$ 上，其中某 $\delta>0$）还是**非算术分布**（即不是算术分布）. 将在第 5 章把第一种情形作为马尔可夫链的应用来处理，故此处仅讨论第二种情形.

定理 2.6.4. Blackwell 更新定理. 若 F 为非算术分布，则当 $t\to\infty$ 时

$$U([t,t+h])\to h/\mu$$

下面采用 “耦合” 方法对 $\mu<\infty$ 的情形加以证明，该方法来自 Lindvall（1977）和 Athreya, McDonald 与 Ney（1978）. 为证明做准备，需要一个定义和一些初步计算. 若 $T_0\geq 0$ 独立于 $\xi_1,\xi_2,\ldots$ 且具有分布 G，则 $T_k=T_{k-1}+\xi_k$（$k\geq 1$）定义了一个**延迟更新过程**，G 称为**延迟分布**. 若如前令 $N(t)=\inf\{k:T_k>t\}$ 并设 $V(t)=EN(t)$，则按 T_0 的值分解可得

$$V(t)=\int_0^t U(t-s)\,\mathrm{d}G(s) \tag{2.6.1}$$

最后一个积分（以及下面所有类似表达式）都包含 G 在 0 处的任何质量. 若令 $U(r)=0$（当 $r<0$），则最后一个方程可写为 $V=U*G$，其中 $*$ 表示卷积.

对 U 应用类似推理得

$$U(t)=1+\int_0^t U(t-s)\,\mathrm{d}F(s) \tag{2.6.2}$$

或引入卷积记号，

$$U=1_{[0,\infty)}(t)+U*F$$

将两边与 G 卷积（并注意 $G*U=U*G$）得

$$V=G*U=G+V*F \tag{2.6.3}$$

已知 $U(t)\sim t/\mu$. 下一步是找一个 G 使得 $V(t)=t/\mu$. 将目标代入 (2.6.3) 得

$$\frac{t}{\mu}=G(t)+\int_0^t\frac{t-y}{\mu}\,\mathrm{d}F(y)$$

于是

$$G(t)=\frac{t}{\mu}-\int_0^t\frac{t-y}{\mu}\,\mathrm{d}F(y)$$

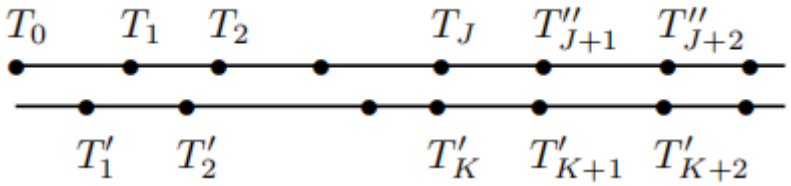

图 2.2: 一对更新过程

分部积分公式为

$$\int_0^t K(y)\,\mathrm{d}H(y) = H(t)K(t) - H(0)K(0) - \int_0^t H(y)\,\mathrm{d}K(y)$$

若令 $H(y) = (y-t)/\mu$，$K(y) = 1 - F(y)$，则

$$\frac{1}{\mu}\int_0^t 1 - F(y)\,\mathrm{d}y = \frac{t}{\mu} - \int_0^t \frac{t-y}{\mu}\,\mathrm{d}F(y)$$

故有

$$G(t) = \frac{1}{\mu}\int_0^t 1 - F(y)\,\mathrm{d}y \tag{2.6.4}$$

值得注意的是 $\mu = \int_0^\infty [1 - F(y)]\,\mathrm{d}y$，故最后一个公式确实定义了一个概率分布. 当延迟分布 G 由 (2.6.4) 给出时，所得过程称为**平稳更新过程**.

当 $F(t) = 1 - \exp(-\lambda t)$（$t \geq 0$，$\lambda > 0$）（即更新过程是速率为 λ 的泊松过程）时，发生非常特殊的事情. 此时 $\mu = 1/\lambda$，故 $G(t) = F(t)$.

定理 2.6.4 的证明（$\mu < \infty$ 的情形）. 设 T_n 为更新过程（$T_0 = 0$），T'_n 为独立的平稳更新过程. 首先找到 J 和 K 使得 $|T_J - T'_K| < \varepsilon$，且增量 $\{T_{J+i} - T_J, i \geq 1\}$ 与 $\{T'_{K+i} - T'_K, i \geq 1\}$ 为独立同分布序列，且独立于之前发生的一切.

令 $\eta_1, \eta_2, \ldots$ 和 $\eta'_1, \eta'_2, \ldots$ 为独立同分布且独立于 T_n 和 T'_n 的随机变量，各以概率 $1/2$ 取值 0 和 1. 令 $\nu_n = \eta_1 + \cdots + \eta_n$，$\nu'_n = 1 + \eta'_1 + \cdots + \eta'_n$，$S_n = T_{\nu_n}$，$S'_n = T'_{\nu'_n}$. $S_n - S'_n$ 的增量以至少 $1/4$ 的概率为 0，且其分布的支撑是对称的并包含 ξ_k 的支撑，故若 ξ_k 的分布为非算术分布，则随机游动 $S_n - S'_n$ 是不可约的. 由于 $S_n - S'_n$ 的增量均值为 0，故 $N = \inf\{n : |S_n - S'_n| < \varepsilon\}$ 满足 $P(N < \infty) = 1$，可以令 $J = \nu_N$，$K = \nu'_N$. 令

$$T''_n = \begin{cases} T_n & \text{若} J \geq n, \\ T_J + T'_{K+(n-J)} - T'_K & \text{若} J < n \end{cases}$$

换言之，对 $i \geq 1$，增量 $T''_{J+i} - T''_J$ 与 $T'_{K+i} - T'_K$ 相同.

由构造易见 T_n 与 T_n'' 同分布. 若记

$$N'[s,t] = |\{n : T_n' \in [s,t]\}| \text{ 且} N''[s,t] = |\{n : T_n'' \in [s,t]\}|$$

表示两个过程在 $[s,t]$ 中的更新次数，则在 $\{T_J \le t\}$ 上有

$$N''[t,t+h] = N'[t+T_K' - T_J, t+h+T_K' - T_J] \begin{cases} \ge N'[t+\varepsilon, t+h-\varepsilon] \\ \le N'[t-\varepsilon, t+h+\varepsilon] \end{cases}$$

为联系两个过程在期望更新次数上的关系，注意即使对 $[0,s]$ 中所有更新的位置取条件，$[s,s+t]$ 中的期望更新次数也不超过 $U(t)$，因为最坏的情况是在时刻 s 有一个更新. 结合最后两个观察，若 $\varepsilon < h/2$（于是 $[t+\varepsilon, t+h-\varepsilon]$ 具有正长度）

$$U([t,t+h]) = EN''[t,t+h] \ge E(N'[t+\varepsilon, t+h-\varepsilon]; T_J \le t) \ge \frac{h-2\varepsilon}{\mu} - P(T_J > t)U(h)$$

由于 $EN'[t+\varepsilon, t+h-\varepsilon] = (h-2\varepsilon)/\mu$ 且 $\{T_J > t\}$ 由 T 在 $[0,t]$ 中的更新和 T' 在 $[0,t+\varepsilon]$ 中的更新决定. 对于另一方向，有

$$U([t,t+h]) \le E(N'[t-\varepsilon, t+h+\varepsilon]; T_J \le t) + E(N''[t,t+h]; T_J > t) \le \frac{h+2\varepsilon}{\mu} + P(T_J > t)U(h)$$

由 $P(T_J > t) \to 0$ 及 $\varepsilon < h/2$ 任意，即得所求结果. □

定理 2.6.4 的证明（$\mu = \infty$ 的情形）. 在这种情形下，不存在平稳更新过程，故不得不采用其他方法. 记

$$\beta = \limsup_{t\to\infty} U(t, t+1] = \lim_{k\to\infty} U(t_k, t_k+1]$$

对某序列 $t_k \to \infty$. 需证明 $\beta = 0$，因为由此通过加法可知前面对 1 的结论对任何整数 n 也成立，并由单调性对任何 $h < n$ 也成立，这就给出了定理 2.6.4 中的结果. 固定 i 并记

$$a_{k,j} = \int_{(j-1,j]} U(t_k - y, t_k + 1 - y] \, \mathrm{d}F^{i*}(y)$$

考虑 T_i 的位置得

$$\text{(a)} \quad \lim_{k\to\infty} \sum_{j=1}^{\infty} a_{k,j} = \lim_{k\to\infty} \int_0^{\infty} U(t_k - y, t_k + 1 - y] \, \mathrm{d}F^{i*}(y) = \beta$$

由于 β 是上极限，必有

$$\text{(b)} \quad \limsup_{k\to\infty} a_{k,j} \le \beta \cdot P(T_i \in (j-1, j])$$

下面由 (a) 和 (b) 推出

$$\text{(c)} \quad \liminf_{k\to\infty} a_{k,j} \ge \beta \cdot P(T_i \in (j-1, j])$$

为此，注意通过考虑 $(j-1, j]$ 中首次更新的位置有

$$\text{(d)} \quad 0 \le a_{k,j} \le U(1) P(T_i \in (j-1, j])$$

当 $\beta = 0$ 时 (c) 显然成立，故可设 $\beta > 0$. 用反证法，假设存在 j_0 和 $\varepsilon > 0$ 使得

$$\liminf_{k\to\infty} a_{k,j_0} \le \beta \cdot \{P(T_i \in (j_0 - 1, j_0]) - \varepsilon\}$$

取 $k_n \to \infty$ 使得

$$a_{k_n,j_0} \to \beta \cdot \{P(T_i \in (j_0 - 1, j_0]) - \varepsilon\}$$

利用 (d)，可取 $J \ge j_0$ 使得

$$\limsup_{n\to\infty} \sum_{j=J+1}^{\infty} a_{k_n,j} \le U(1) \sum_{j=J+1}^{\infty} P(T_i \in (j-1, j]) \le \beta\varepsilon/2$$

一个简单的论证表明

$$\limsup_{n\to\infty} \sum_{j=1}^{J} a_{k_n,j} \le \sum_{j=1}^{J} \limsup_{n\to\infty} a_{k_n,j} \le \beta \left\{ \sum_{j=1}^{J} P(T_i \in (j-1, j]) - \varepsilon \right\}$$

由 (b) 和假设. 将最后两个结果相加得

$$\limsup_{n\to\infty} \sum_{j=1}^{\infty} a_{k_n,j} \le \beta(1 - \varepsilon/2)$$

这与 (a) 矛盾，证明了 (c).

现在，若 $j - 1 < y \le j$，则

$$U(t_k - y, t_k + 1 - y] \le U(t_k - j, t_k + 2 - j]$$

故利用 (c) 推出对满足 $P(T_i \in (j-1, j]) > 0$ 的 j，必有

$$\liminf_{k\to\infty} U(t_k - j, t_k + 2 - j] \geq \beta$$

对 i 求和可见最后结论在 $U(j-1, j] > 0$ 时成立.

U 的支撑对加法封闭.（若 x 在 F^{m*} 的支撑中且 y 在 F^{n*} 的支撑中，则 $x+y$ 在 $F^{(m+n)*}$ 的支撑中.）已假设 F 为非算术分布，故对 $j \geq j_0$ 有 $U(j-1, j] > 0$. 令 $r_k = t_k - j_0$ 并考虑 $[0, r_k]$ 中最后一次更新的位置和 T_i 的指标得

$$1 = \sum_{i=0}^{\infty} \int_0^{r_k} (1-F(r_k-y))\,\mathrm{d}F^{i*}(y) = \int_0^{r_k} (1-F(r_k-y))\,\mathrm{d}U(y) \geq \sum_{n=1}^{\infty} (1-F(2n))U(r_k-2n, r_k+2-2n]$$

由于 $\liminf_{k\to\infty} U(r_k - 2n, r_k + 2 - 2n] \geq \beta$ 且 $\sum_{n=0}^{\infty}(1-F(2n)) \geq \mu/2 = \infty$，必有 $\beta = 0$，证毕. □

注释. 依照 Lindvall（1977），对 $\mu = \infty$ 情形的证明基于 Feller（1961）关于离散更新定理（即对算术分布）证明的一部分. 关于 Feller 的证明，参见 Freedman（1971b）第 22–25 页. 纯粹主义者可以在 Thorisson（1987）中找到完全用耦合完成所有工作的证明.

下一个主题是更新方程：$H = h + H * F$. 在 (2.6.2) 和 (2.6.3) 中见过的两个例子是：

例题 2.6.5. $h \equiv 1$：$U(t) = 1 + \int_0^t U(t-s)\,\mathrm{d}F(s)$

例题 2.6.6. $h(t) = G(t)$：$V(t) = G(t) + \int_0^t V(t-s)\,\mathrm{d}F(s)$

最后一个方程对任意延迟分布都有效. 若令 G 为 (2.6.4) 中的分布，并将上面两个方程相减，得

例题 2.6.7. $H(t) = U(t) - t/\mu$ 满足更新方程，其中 $h(t) = \frac{1}{\mu}\int_t^{\infty}[1-F(s)]\,\mathrm{d}s$.

最后但同样重要的是，下面给出一个更新方程的典型应用.

例题 2.6.8. 设 $x > 0$ 固定，记 $H(t) = P(T_{N(t)} - t > x)$. 考虑 T_1 的值得

$$H(t) = (1 - F(t+x)) + \int_0^t H(t-s)\,\mathrm{d}F(s)$$

上面的例子应能激发起对下述定理的兴趣：

定理 2.6.9. 若 h 有界，则函数

$$H(t) = \int_0^t h(t-s)\,\mathrm{d}U(s)$$

是更新方程的有界区间上有界解的唯一解.

证明. 记 $U_n(A)=\sum_{m=0}^{n}P(T_m\in A)$ 且

$$H_n(t)=\int_0^t h(t-s)\,\mathrm{d}U_n(s)=\sum_{m=0}^{n}(h*F^{m*})(t)$$

这里 F^{m*} 是 T_m 的分布，将 h 延拓为 $h(r)=0$（当 $r<0$）. 由最后一个表达式易见

$$H_{n+1}=h+H_n*F$$

$U(t)<\infty$ 的事实蕴含 $U(t)-U_n(t)\to 0$. 由于 h 有界，

$$|H_n(t)-H(t)|\le\|h\|_\infty|U(t)-U_n(t)|$$

且 $H_n(t)\to H(t)$ 在有界区间上一致收敛. 为估计卷积，注意

$$|H_n*F(t)-H*F(t)|\le\sup_{s\le t}|H_n(s)-H(s)|\le\|h\|_\infty|U(t)-U_n(t)|$$

由于 $U-U_n=\sum_{m=n+1}^{\infty}F^{m*}$ 关于 t 递增. 在 $H_{n+1}=h+H_n*F$ 中令 $n\to\infty$，可见 H 是更新方程的一个在有界区间上有界的解.

为证唯一性，注意若 H_1 和 H_2 是两个解，则 $K=H_1-H_2$ 满足 $K=K*F$. 若 K 在有界区间上有界，迭代得 $K=K*F^{n*}\to 0$（当 $n\to\infty$），故 $H_1=H_2$. □

定理 2.6.9 的证明在 $F(\infty)=P(\xi_i<\infty)<1$ 时仍然有效. 在这种情形下，有一个**终止更新过程**. 经过均值为 $1/(1-F(\infty))$ 的几何多次试验后，$T_n=\infty$. 这个“平凡情形”有一些有趣的应用.

例题 2.6.10. “行人延迟”. 一只鸡想过一条马路（不要问为什么），马路上的车流是速率为 λ 的泊松过程. 她需要 1 个单位时间内没有车辆到达才能安全过马路. 记 $M=\inf\{t\ge 0:$ 在$(t,t+1]$中没有到达$\}$ 为她开始等待过马路的时间. 考虑首次到达的时刻，可见 $H(t)=P(M\le t)$ 满足

$$H(t)=e^{-\lambda}+\int_0^1 H(t-y)\lambda e^{-\lambda y}\,\mathrm{d}y$$

与例 2.6.5 比较并利用定理 2.6.9，得

$$H(t)=e^{-\lambda}\sum_{n=0}^{\infty}F^{n*}(t)$$

不用更新理论，通过注意

$$P(M\le t)=\sum_{n=0}^{\infty}P(T_n\le t,T_{n+1}=\infty)$$

也能得到这一结果. 最后一个表示使我们能计算 M 的均值. 设 μ 为到达间隔时间小于 1 的条件下其条件均值，注意指数分布的无记忆性蕴含

$$\mu = \int_0^1 x\lambda e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda} - (\frac{1}{\lambda}+1)e^{-\lambda}$$

于是，考虑终止更新过程中更新次数的期望，

$$EM = \sum_{n=0}^{\infty} e^{-\lambda}(1-e^{-\lambda})^n n\mu = (e^{\lambda}-1)\mu$$

因为若 X 是成功概率为 $e^{-\lambda}$ 的几何分布，则 $EM = \mu E(X-1)$.

例题 2.6.11. Cramér 破产估计. 考虑一家保险公司以速率 c 收取保费，并在速率为 1 的泊松过程 $N(t)$ 的到达时刻经历独立同分布的理赔 $Y_1, Y_2, \ldots$，理赔分布为 G 且均值为 μ. 若其初始资本为 x，则时刻 t 的财富为

$$W_x(t) = x + ct - \sum_{m=1}^{N(t)} Y_i$$

记

$$R(x) = P(W_x(t) \geq 0 \text{ 对所有} t)$$

表示以资本 x 开始永不破产的概率. 考虑首次理赔的时刻和大小:

$$\text{(a)}\quad R(x) = \int_0^{\infty} e^{-s}\int_0^{x+cs} R(x+cs-y)\,\mathrm{d}G(y)\,\mathrm{d}s$$

这看起来不太像更新方程，但经过一番巧思可将其转化为更新方程. 换元 $t = x+cs$

$$R(x)e^{x/c} = \int_x^{\infty} e^{-t/c}\frac{\mathrm{d}t}{c}\int_0^t R(t-y)\,\mathrm{d}G(y)$$

对 x 求导再乘以 $e^{x/c}$,

$$R'(x) = \frac{1}{c}R(x) - \frac{1}{c}\int_0^x R(x-y)\,\mathrm{d}G(y)$$

将 x 从 0 到 w 积分

$$\text{(b)}\quad R(w) - R(0) = \frac{1}{c}\int_0^w R(x)\,\mathrm{d}x - \frac{1}{c}\int_0^w\int_0^x R(x-y)\,\mathrm{d}G(y)\,\mathrm{d}x$$

交换重积分中的积分次序，记

$$S(w) = \int_0^w R(x)\,\mathrm{d}x$$

利用 $\mathrm{d}G = -\mathrm{d}(1-G)$ 再分部积分

$$-\frac{1}{c}\int_0^w \int_y^w R(x-y)\,\mathrm{d}x\,\mathrm{d}G(y) = -\frac{1}{c}\int_0^w S(w-y)\,\mathrm{d}G(y) = \frac{1}{c}\int_0^w S(w-y)\,\mathrm{d}(1-G)(y)$$

$$= \frac{1}{c}\left\{-S(w) + \int_0^w (1-G(y))R(w-y)\,\mathrm{d}y\right\}$$

代入 (b)，终于得到一个更新方程：

$$\text{(c)} \quad R(w) = R(0) + \int_0^w R(w-y)\frac{1-G(y)}{c}\,\mathrm{d}y$$

得出最后一个方程需要一些巧思，但分析起来却很直接. 首先，排除一个平凡情形. 若 $\mu > c$，则

$$ct - \sum_{m=1}^{N(t)} Y_i \to c - \mu < 0 \quad \text{几乎必然}$$

故 $R(x) \equiv 0$. 当 $\mu < c$ 时，

$$F(x) = \int_0^x \frac{1-G(y)}{c}\,\mathrm{d}y$$

是一个缺欠概率分布，满足 $F(\infty) = \mu/c$. 更新方程可写为

$$\text{(d)} \quad R = R(0) + R * F$$

故与例 2.6.5 比较并利用定理 2.6.9 可知 $R(w) = R(0)U(w)$. 为完成求解，还需计算常数 $R(0)$. 令 $w \to \infty$ 并注意 $R(w) \to 1$，$U(w) \to (1 - F(\infty))^{-1} = (1-\mu/c)^{-1}$，得 $R(0) = 1 - \mu/c$.

更新方程解（在非终止情形）的基本事实是：

定理 2.6.12. 更新定理. 若 F 为非算术分布且 h 为**直接黎曼可积**，则当 $t \to \infty$ 时

$$H(t) \to \frac{1}{\mu}\int_0^\infty h(s)\,\mathrm{d}s$$

直观上，这成立是因为定理 2.6.9 蕴含

$$H(t) = \int_0^t h(t-s)\,\mathrm{d}U(s)$$

而定理 2.6.4 蕴含当 $s \to \infty$ 时 $\mathrm{d}U(s) \to \mathrm{d}s/\mu$. 下面给出直接黎曼可积的定义. 将开始证明，然后确定需要假设什么.

证明. 设

$$h(s)=\sum_{k=0}^{\infty}a_k1_{[k\delta,(k+1)\delta)}(s)$$

其中 $\sum_{k=0}^{\infty}|a_k|<\infty$. 由于 $U([t,t+\delta])\le U([0,\delta])<\infty$，由定理 2.6.4 易见

$$\int_0^t h(t-s)\,\mathrm{d}U(s)=\sum_{k=0}^{\infty}a_kU((t-(k+1)\delta,t-k\delta])\to\frac{1}{\mu}\sum_{k=0}^{\infty}a_k\delta$$

(取 K 使得 $\sum_{k\ge K}|a_k|\le\varepsilon/2U([0,\delta])$，再取 T 使得对 $t\ge T$ 和 $0\le k<K$ 有 $|a_k|\cdot|U((t-(k+1)\delta,t-k\delta])-\delta/\mu|\le\varepsilon/2K$.)

若 h 是 $[0,\infty)$ 上的任意函数，记

$$I^{\delta}=\sum_{k=0}^{\infty}\delta\sup\{h(x):x\in[k\delta,(k+1)\delta)\}$$

$$I_{\delta}=\sum_{k=0}^{\infty}\delta\inf\{h(x):x\in[k\delta,(k+1)\delta)\}$$

为逼近 $[0,\infty)$ 上 h 的积分的上、下 Riemann 和. 将 h 与在 $[k\delta,(k+1)\delta)$ 上为常数的显然上、下界比较并利用特殊情形的结果，

$$\frac{I_{\delta}}{\mu}\le\liminf_{t\to\infty}\int_0^t h(t-s)\,\mathrm{d}U(s)\le\limsup_{t\to\infty}\int_0^t h(t-s)\,\mathrm{d}U(s)\le\frac{I^{\delta}}{\mu}$$

若当 $\delta\to0$ 时 I^{δ} 和 I_{δ} 都趋于同一有限极限 I，则称 h 为**直接黎曼可积**，由此推出

$$\int_0^t h(t-s)\,\mathrm{d}U(y)\to I/\mu$$

□

注释. “直接” 一词的含义是，虽然 $[0,\infty)$ 上的 Riemann 积分通常定义为 $[0,a]$ 上积分的极限，这里直接逼近 $[0,\infty)$ 上的积分.

在检验定理 2.6.12 中的新假设时，下面的结果很有用.

引理 2.6.13. 若 $h(x)\ge0$ 递减，$h(0)<\infty$ 且 $\int_0^{\infty}h(x)\,\mathrm{d}x<\infty$，则 h 为直接黎曼可积.

证明. 由于 h 递减，$I^{\delta}=\sum_{k=0}^{\infty}\delta h(k\delta)$ 且 $I_{\delta}=\sum_{k=0}^{\infty}\delta h((k+1)\delta)$. 故

$$I^{\delta}\ge\int_0^{\infty}h(x)\,\mathrm{d}x\ge I_{\delta}=I^{\delta}-h(0)\delta$$

即得所求结果. □

回到例子，跳过前两个，因为在那些情形中 $h(t) \to 1$（当 $t \to \infty$），故 h 在任何意义下都不可积.

例题 2.6.14. 例 2.6.7 的继续. $h(t) = \frac{1}{\mu}\int_t^\infty [1 - F(s)]\,\mathrm{d}s$. h 递减，$h(0) = 1$，且

$$\mu \int_0^\infty h(t)\,\mathrm{d}t = \int_0^\infty \int_t^\infty [1 - F(s)]\,\mathrm{d}s\,\mathrm{d}t = \int_0^\infty \int_0^s [1 - F(s)]\,\mathrm{d}t\,\mathrm{d}s = \int_0^\infty s(1 - F(s))\,\mathrm{d}s = E(\xi_i^2/2)$$

于是，若 $\nu \equiv E(\xi_i^2) < \infty$，由引理 2.6.13、定理 2.6.12 及例 2.6.7 中的公式得

$$0 \le U(t) - t/\mu \to \nu/2\mu^2 \quad \text{当} t \to \infty$$

当更新过程是速率为 λ 的泊松过程时，即 $P(\xi_i > t) = e^{-\lambda t}$，$N(t) - 1$ 具有均值为 λt 的泊松分布，故 $U(t) = 1 + \lambda t$. 按照 Feller 第二卷（1971）第 385 页，若 ξ_i 服从 $(0, 1)$ 上的均匀分布，则

$$U(t) = \sum_{k=0}^{n} (-1)^k \frac{e^{t-k}(t-k)^k}{k!} \quad \text{对} n \le t \le n+1$$

如他所说，精确表达式“很少揭示 U 的本质. 渐近公式 $0 \le U(t) - 2t \to 2/3$ 要有趣得多. ”

例题 2.6.15. 例 2.6.8 的继续. $h(t) = 1 - F(t+x)$. 同样，h 递减，但这一次 $h(0) \le 1$ 且当 $\mu = E\xi_i < \infty$ 时 h 的积分有限. 现在应用引理 2.6.13 和定理 2.6.12 得

$$P(T_{N(t)} - t > x) \to \frac{1}{\mu}\int_0^\infty h(s)\,\mathrm{d}s = \frac{1}{\mu}\int_x^\infty [1 - F(t)]\,\mathrm{d}t$$

故（当 $\mu < \infty$ 时）剩余等待时间 $T_{N(t)} - t$ 的分布收敛于产生平稳更新过程的延迟分布. 这一事实也可从 2.6.4 的证明中得出.

习题

习题 2.6.1. 证明 $t/E(\xi_i \wedge t) \le U(t) \le 2t/E(\xi_i \wedge t)$.

习题 2.6.2. 由定理 2.6.1 推出定理 2.6.3，通过证明 $\limsup_{t\to\infty} E(N(t)/t)^2 < \infty$.

提示：使用定理 2.6.3 证明中的比较方法.

习题 2.6.3. 顾客在速率为 1 的泊松过程时刻到达. 若服务员正忙，他们就离开. （想象公用电话或风尘女子. ）若不然，他们接受服务，所需服务时间服从均值为 μ 的分布 F. 证明顾客开始接受服务的时刻构成一个均值为 $\mu + 1$ 的更新过程，并用定理 2.6.1 推出渐近顾客服务比例为 $1/(\mu + 1)$.

在下面的问题中，假设 F 为非算术分布，在出现均值的问题中假设均值有限.

习题 2.6.4. 设 $A_t = t - T_{N(t)-1}$ 为时刻 t 的 “年龄”，即自上次更新以来经过的时间. 若固定 $x > 0$，则 $H(t) = P(A_t > x)$ 满足更新方程

$$H(t) = (1 - F(t)) \cdot 1_{(x,\infty)}(t) + \int_0^t H(t-s)\,\mathrm{d}F(s)$$

故 $P(A_t > x) \to \frac{1}{\mu}\int_x^\infty [1 - F(t)]\,\mathrm{d}t$，这是剩余寿命 $B_t = T_{N(t)} - t$ 的极限分布.

习题 2.6.5. 利用上一题的更新方程和定理 2.6.9 推出：若 T 是速率为 λ 的泊松过程，则 A_t 与 $\xi_i \wedge t$ 同分布.

习题 2.6.6. 设 $A_t = t - T_{N(t)-1}$，$B_t = T_{N(t)} - t$. 证明

$$P(A_t > x, B_t > y) \to \frac{1}{\mu}\int_{x+y}^\infty [1 - F(t)]\,\mathrm{d}t$$

习题 2.6.7. 交替更新过程. 设 $\xi_1, \xi_2, \ldots > 0$ 独立同分布于 F_1，$\eta_1, \eta_2, \ldots > 0$ 独立同分布于 F_2. 令 $T_0 = 0$，且对 $k \geq 1$ 令 $S_k = T_{k-1} + \xi_k$，$T_k = S_k + \eta_k$. 换言之，有一台机器工作 ξ_k 时间后故障，然后需要 η_k 单位时间修理. 令 $F = F_1 * F_2$，$H(t)$ 为时刻 t 机器在工作的概率. 证明若 F 为非算术分布，则当 $t \to \infty$ 时

$$H(t) \to \mu_1/(\mu_1 + \mu_2)$$

其中 μ_i 为 F_i 的均值.

习题 2.6.8. 为 $H(t) = P([0,t]$中更新次数为奇数$)$ 写一个更新方程，并利用更新定理证明 $H(t) \to 1/2$. 注：这是上一习题的特殊情形.

习题 2.6.9. 更新密度. 证明若 $F(t)$ 具有直接黎曼可积的密度函数 $f(t)$，则 $V = U - 1_{[0,\infty)}$ 具有密度 v 满足

$$v(t) = f(t) + \int_0^t v(t-s)\,\mathrm{d}F(s)$$

利用更新定理推出若 f 为直接黎曼可积，则 $v(t) \to 1/\mu$（当 $t \to \infty$）.

第 7 节 大偏差 *

设 $X_1, X_2, \ldots$ 独立同分布, $S_n = X_1 + \cdots + X_n$. 本节将研究当 $a > \mu = EX_i$ 时, $P(S_n > na) \to 0$ 的速率. 最终将得到: 若对某个 $\theta > 0$, 矩母函数 $\varphi(\theta) = E\exp(\theta X_i) < \infty$, 则 $P(S_n \geq na)$ 以指数速度趋于 0, 并且将确定

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

首先证明该极限存在. 这一基于观察的论证在下面的讨论中还将多次用到. 令 $\pi_n = P(S_n \geq na)$. 由于 S_m 与 $S_{n+m} - S_m$ 独立,

$$\pi_{m+n} \geq P(S_m \geq ma, S_{n+m} - S_m \geq na) = \pi_m \pi_n$$

令 $\gamma_n = \log \pi_n$, 乘法就转化为加法.

引理 2.7.1. 若 $\gamma_{m+n} \geq \gamma_m + \gamma_n$, 则当 $n \to \infty$ 时, $\gamma_n/n \to \sup_m \gamma_m/m$.

证明. 显然, $\limsup \gamma_n/n \leq \sup \gamma_m/m$. 为完成证明, 只需证对任意 m 有 $\liminf \gamma_n/n \geq \gamma_m/m$. 将 n 写成 $n = km + \ell$, 其中 $0 \leq \ell < m$, 并反复利用假设得 $\gamma_n \geq k\gamma_m + \gamma_\ell$. 除以 $n = km + \ell$ 得

$$\frac{\gamma(n)}{n} \geq \frac{km}{km+\ell} \cdot \frac{\gamma(m)}{m} + \frac{\gamma(\ell)}{n}$$

令 $n \to \infty$, 并注意到 $n = km + \ell$ 且 $0 \leq \ell < m$, 即得所需结果. □

引理 2.7.1 蕴含 $\lim_{n\to\infty} n^{-1} \log P(S_n \geq na) = \gamma(a)$ 存在且 ≤ 0. 由极限公式可得

$$P(S_n \geq na) \leq e^{n\gamma(a)} \tag{2.7.1}$$

最后这个结论对任何分布都成立, 但若 $\gamma(a) = 0$ 则它不是很有用. 对本节其余部分, 假设:

(H1) $\varphi(\theta) = E\exp(\theta X_i) < \infty$ 对某个 $\theta > 0$ 成立

令 $\theta_+ = \sup\{\theta : \varphi(\theta) < \infty\}$, $\theta_- = \inf\{\theta : \varphi(\theta) < \infty\}$, 并注意到当 $\theta \in (\theta_-, \theta_+)$ 时 $\varphi(\theta) < \infty$. (H1) 蕴含 $EX_i^+ < \infty$, 故 $\mu = EX^+ - EX^- \in [-\infty, \infty)$. 若 $\theta > 0$, Chebyshev 不等式给出

$$e^{\theta na} P(S_n \geq na) \leq E\exp(\theta S_n) = \varphi(\theta)^n$$

或令 $\kappa(\theta) = \log \varphi(\theta)$,

$$P(S_n \geq na) \leq \exp(-n\{a\theta - \kappa(\theta)\}) \tag{2.7.2}$$

下面证明:

引理 2.7.2. 若 $a > \mu$ 且 $\theta > 0$ 充分小, 则 $a\theta - \kappa(\theta) > 0$.

证明. $\kappa(0)=\log\varphi(0)=0$，故只需证明 (i) κ 在 0 处连续，(ii) κ 在 $(0,\theta_+)$ 上可微，以及 (iii) 当 $\theta\to 0$ 时 $\kappa'(\theta)\to\mu$. 因为

$$\int_0^\theta (a-\kappa'(x))\,\mathrm{d}x>0$$

对充分小的 θ 成立.

令 $F(x)=P(X_i\leq x)$. 为证 (i)，注意若 $0<\theta<\theta_0<\theta_-$，则

$$e^{\theta x}\leq 1+e^{\theta_0 x} \tag{$*$}$$

故由控制收敛定理，当 $\theta\to 0$ 时

$$\int e^{\theta x}\,\mathrm{d}F\to\int 1\,\mathrm{d}F=1$$

为证 (ii)，注意若 $|h|<h_0$，则

$$|e^{hx}-1|=\left|\int_0^{hx}e^y\,\mathrm{d}y\right|\leq|hx|e^{h_0x}$$

故应用控制收敛定理得

$$\begin{aligned}\varphi'(\theta)&=\lim_{h\to0}\frac{\varphi(\theta+h)-\varphi(\theta)}{h}\\&=\lim_{h\to0}\int\frac{e^{hx}-1}{h}e^{\theta x}\,\mathrm{d}F(x)\\&=\int xe^{\theta x}\,\mathrm{d}F(x),\quad\theta\in(0,\theta_+)\end{aligned}$$

由最后一个等式，$\kappa(\theta)=\log\varphi(\theta)$ 有 $\kappa'(\theta)=\varphi'(\theta)/\varphi(\theta)$. 利用 $(*)$ 和控制收敛定理即得 (iii). □

在得到 $P(S_n\geq na)$ 的上界后，自然要通过求 $\theta a-\kappa(\theta)$ 的最大值来优化它：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\{\theta a-\log\varphi(\theta)\}=a-\varphi'(\theta)/\varphi(\theta)$$

故（假设情况良好）最大值出现在 $a=\varphi'(\theta)/\varphi(\theta)$ 处. 为将括号中的条款转化为数学假设，首先定义

$$F_\theta(x)=\frac{1}{\varphi(\theta)}\int_{-\infty}^x e^{\theta y}\,\mathrm{d}F(y)$$

每当 $\varphi(\theta)<\infty$ 时. 由引理 2.7.2 的证明可知，若 $\theta\in(\theta_-,\theta_+)$，则 F_θ 是一个分布函数，其均值为

$$\int_{-\infty}^\infty x\,\mathrm{d}F_\theta(x)=\frac{1}{\varphi(\theta)}\int_{-\infty}^\infty xe^{\theta x}\,\mathrm{d}F(x)=\frac{\varphi'(\theta)}{\varphi(\theta)}$$

重复引理 2.7.2 中的证明，易见若 $\theta \in (\theta_-, \theta_+)$，则

$$\varphi''(\theta) = \int_{-\infty}^{\infty} x^2 e^{\theta x} \,\mathrm{d}F(x)$$

于是

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\varphi'(\theta)}{\varphi(\theta)} &= \frac{\varphi''(\theta)}{\varphi(\theta)} - \left(\frac{\varphi'(\theta)}{\varphi(\theta)}\right)^2 \\ &= \int x^2 \,\mathrm{d}F_\theta(x) - \left(\int x \,\mathrm{d}F_\theta(x)\right)^2 \geq 0\end{aligned}$$

因其恰为 F_θ 的方差. 假设

(H2) 分布 F 不是集中在 μ 处的单点分布

则 $\varphi'(\theta)/\varphi(\theta)$ 严格递增，且 $a\theta - \log\varphi(\theta)$ 是凹的. 由于 $\varphi'(0)/\varphi(0) = \mu$，这表明对每个 $a > \mu$，至多存在一个 $\theta_a \geq 0$ 满足 $a = \varphi'(\theta_a)/\varphi(\theta_a)$，且这个 θ 值使 $a\theta - \log\varphi(\theta)$ 达到最大. 在讨论 θ_a 的存在性之前，先考虑一些例子.

例题 2.7.3. 正态分布.

$$\int e^{\theta x}(2\pi)^{-1/2}\exp(-x^2/2)\,\mathrm{d}x = \exp(\theta^2/2)\int (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)\,\mathrm{d}x$$

最后一个积分中的被积函数是均值为 θ、方差为 1 的正态分布密度，故 $\varphi(\theta) = \exp(\theta^2/2)$，$\theta \in (-\infty, \infty)$. 此时 $\varphi'(\theta)/\varphi(\theta) = \theta$，且

$$F_\theta(x) = e^{-\theta^2/2}\int_{-\infty}^{x} e^{\theta y}(2\pi)^{-1/2}e^{-y^2/2}\,\mathrm{d}y$$

是均值为 θ、方差为 1 的正态分布.

例题 2.7.4. 参数为 λ 的指数分布. 若 $\theta < \lambda$,

$$\int_0^{\infty} e^{\theta x}\lambda e^{-\lambda x}\,\mathrm{d}x = \lambda/(\lambda-\theta)$$

$\varphi'(\theta)/\varphi(\theta) = 1/(\lambda-\theta)$，且

$$F_\theta(x) = \frac{\lambda}{\lambda-\theta}\int_0^{x} e^{\theta y}\lambda e^{-\lambda y}\,\mathrm{d}y$$

是参数为 $\lambda - \theta$ 的指数分布，因而均值为 $1/(\lambda-\theta)$.

例题 2.7.5. 抛硬币. $P(X_i=1)=P(X_i=-1)=1/2$

$$\varphi(\theta)=(e^{\theta}+e^{-\theta})/2$$

$\varphi'(\theta)/\varphi(\theta)=(e^{\theta}-e^{-\theta})/(e^{\theta}+e^{-\theta})$

$F_\theta(\{x\})/F(\{x\})=e^{\theta x}/\varphi(\theta)$，故

$$F_\theta(\{1\})=e^{\theta}/(e^{\theta}+e^{-\theta}),\quad F_\theta(\{-1\})=e^{-\theta}/(e^{\theta}+e^{-\theta})$$

例题 2.7.6. 奇异指数分布. 令 $g(x)=Cx^{-3}e^{-x}$（当 $x\geq 1$），$g(x)=0$（其他），并选取 C 使 g 成为概率密度. 此时，

$$\int e^{\theta x}g(x)\,\mathrm{d}x<\infty$$

当且仅当 $\theta\leq 1$，且当 $\theta\leq 1$ 时，

$$\frac{\varphi'(\theta)}{\varphi(\theta)}\leq\frac{\varphi'(1)}{\varphi(1)}=\frac{\int_1^{\infty}Cx^{-2}\,\mathrm{d}x}{\int_1^{\infty}Cx^{-3}\,\mathrm{d}x}=2$$

回顾 $\theta_+=\sup\{\theta:\varphi(\theta)<\infty\}$. 在例 2.7.3 和 2.7.4 中，当 $\theta\uparrow\theta_+$ 时 $\varphi'(\theta)/\varphi(\theta)\uparrow\infty$，故对任意 $a>\mu$ 都能解 $a=\varphi'(\theta)/\varphi(\theta)$. 在例 2.7.5 中，当 $\theta\to\infty$ 时 $\varphi'(\theta)/\varphi(\theta)\uparrow 1$，但不能期望更多，因为 F 从而 F_θ 支撑在 $\{-1,1\}$ 上. 例 2.7.6 提出了一个问题：当 $a>2$ 时无法解 $a=\varphi'(\theta)/\varphi(\theta)$. 定理 2.7.10 将处理这种疑难情形，但首先处理可以解方程的情形.

定理 2.7.7. 假设除 (H1) 和 (H2) 外，存在 $\theta_a\in(0,\theta_+)$ 使得 $a=\varphi'(\theta_a)/\varphi(\theta_a)$. 则当 $n\to\infty$ 时，

$$n^{-1}\log P(S_n\geq na)\to -a\theta_a+\log\varphi(\theta_a)$$

证明. 左边 lim sup 不超过右边由 (2.7.2) 可得. 为证另一方向的不等式，取 $\lambda\in(\theta_a,\theta_+)$，令 $X_1^\lambda,X_2^\lambda,\ldots$ 独立同分布，具有分布 F_λ，且 $S_n^\lambda=X_1^\lambda+\cdots+X_n^\lambda$. 记 $\mathrm{d}F/\mathrm{d}F_\lambda$ 为相应测度的 Radon-Nikodym 导数，由定义立即得到 $\mathrm{d}F/\mathrm{d}F_\lambda=e^{-\lambda x}\varphi(\lambda)$. 若令 F_λ^n 和 F^n 分别表示 S_n^λ 和 S_n 的分布，则

引理 2.7.8. $\dfrac{\mathrm{d}F^n}{\mathrm{d}F_\lambda^n}=e^{-\lambda x}\varphi(\lambda)^n$.

证明. 用归纳法证明. 当 $n=1$ 时结果成立. 对 $n>1$，注意

$$\begin{aligned} F^n = F^{n-1} * F(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} \mathrm{d}F^{n-1}(x)\,\mathrm{d}F(y) \\ &= \int\int \mathrm{d}F_\lambda(x)\,\mathrm{d}F_\lambda(y) 1_{(x+y\le z)} e^{-\lambda(x+y)}\varphi(\lambda)^n \\ &= E\left(1_{(S_{n-1}^\lambda + X_n^\lambda \le z)} e^{-\lambda(S_{n-1}^\lambda + X_n^\lambda)}\varphi(\lambda)^n\right) \\ &= \int_{-\infty}^{z} \mathrm{d}F_\lambda^n(u) e^{-\lambda u}\varphi(\lambda)^n \end{aligned}$$

其中最后两个等式对 $(S_{n-1}^\lambda, X_n^\lambda)$ 和 S_n^λ 应用了定理 1.6.9. □

若 $\nu > a$，则由引理和单调性，

$$P(S_n \ge na) \ge \int_{na}^{n\nu} e^{-\lambda x}\varphi(\lambda)^n\,\mathrm{d}F_\lambda^n(x) \ge \varphi(\lambda)^n e^{-\lambda n\nu}(F_\lambda^n(n\nu) - F_\lambda^n(na)) \qquad (*)$$

F_λ 的均值为 $\varphi'(\lambda)/\varphi(\lambda)$，故若 $a < \varphi'(\lambda)/\varphi(\lambda) < \nu$，则由弱大数定律

$$F_\lambda^n(n\nu) - F_\lambda^n(na) \to 1, \quad n\to\infty$$

由最后的结论和 $(*)$ 可得

$$\liminf_{n\to\infty} n^{-1}\log P(S_n > na) \ge -\lambda\nu + \log\varphi(\lambda)$$

由于 $\lambda > \theta_a$ 和 $\nu > a$ 是任意的，证毕. □

为体会答案的形式，考虑例子. 为计算做准备，回顾一些重要信息：

$$\kappa(\theta) = \log\varphi(\theta), \quad \kappa'(\theta) = \varphi'(\theta)/\varphi(\theta), \quad \theta_a \text{ 解} \kappa'(\theta_a) = a$$

$$\gamma(a) = \lim_{n\to\infty}(1/n)\log P(S_n \ge na) = -a\theta_a + \kappa(\theta_a)$$

正态分布. （例 2.7.3）

$$\kappa(\theta) = \theta^2/2, \quad \kappa'(\theta) = \theta, \quad \theta_a = a$$

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a^2/2$$

最后一个结果可通过观察 S_n 具有均值 0 和方差 n 的正态分布，然后应用定理 1.2.6 来验证.

指数分布. （例 2.7.4），取 $\lambda = 1$

$$\kappa(\theta) = -\log(1-\theta), \quad \kappa'(\theta) = 1/(1-\theta), \quad \theta_a = 1 - 1/a$$

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a + 1 + \log a$$

以这两个例子为模型，读者应能处理

抛硬币.（例 2.7.5）. 这里采用不同的方法. 为找到使 F_θ 的均值为 a 的 θ，令 $F_\theta(\{1\}) = e^\theta/(e^\theta+e^{-\theta}) = (1+a)/2$. 令 $x = e^\theta$ 得

$$2x = (1+a)(x + x^{-1})$$

即 $(a-1)x^2 + (1+a) = 0$. 故 $x = \sqrt{(1+a)/(1-a)}$，且 $\theta_a = \log x = \{\log(1+a) - \log(1-a)\}/2$.

$$\varphi(\theta_a) = \frac{e^{\theta_a} + e^{-\theta_a}}{2} = \frac{1}{\sqrt{(1+a)(1-a)}}$$

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -\{(1+a)\log(1+a) + (1-a)\log(1-a)\}/2$$

在习题 3.1.3 中，这一结果将通过直接计算证明. 由于 $\gamma(a)$ 的公式相当复杂，习题 2.7.4 中给出的如下更简单的上界很有用.

现在转向那些无法解 $a = \varphi'(\theta_a)/\varphi(\theta_a)$ 的疑难值.

定理 2.7.9. 设 $x_0 = \sup\{x : F(x) < 1\} < \infty$ 且 F 不是集中在 x_0 处的单点分布. 则对所有 $\theta > 0$ 有 $\varphi(\theta) < \infty$，且当 $\theta \uparrow \infty$ 时 $\varphi'(\theta)/\varphi(\theta) \to x_0$.

证明. 由于 $P(X \le x_0) = 1$，对所有 $\theta > 0$ 有 $Ee^{\theta X} < \infty$. 由于 F_θ 集中在 $(-\infty, x_0]$ 上，显然 $\mu_\theta = \varphi'(\theta)/\varphi(\theta) \le x_0$. 另一方面，若 $\delta > 0$，则 $P(X \ge x_0 - \delta) = c_\delta > 0$，$Ee^{\theta X} \ge c_\delta e^{\theta(x_0-\delta)}$，故

$$F_\theta(x_0 - 2\delta) = \frac{1}{\varphi(\theta)} \int_{-\infty}^{x_0-2\delta} e^{\theta x}\, \mathrm{d}F(x) \le \frac{e^{(x_0-2\delta)\theta}}{c_\delta e^{(x_0-\delta)\theta}} = e^{-\theta\delta}/c_\delta \to 0 \tag{2.7.3}$$

由于 $\delta > 0$ 是任意的，得 $\mu_\theta \to x_0$. □

$a = x_0$ 的结果是平凡的：

$$n^{-1}\log P(S_n \ge nx_0) = \log P(X_i = x_0), \quad \text{对所有} n$$

留给读者证明当 $a \uparrow x_0$ 时 $\gamma(a) \downarrow \log P(X_i = x_0)$.

当 $x_0 = \infty$ 且 $\theta_+ = \infty$ 时，(2.7.3) 中的计算蕴含当 $\theta \uparrow \infty$ 时 $\varphi'(\theta)/\varphi(\theta) \uparrow \infty$，故唯一剩下的情形由以下定理覆盖：

定理 2.7.10. 设 $x_0 = \infty$，$\theta_+ < \infty$，且当 $\theta \uparrow \theta_+$ 时 $\varphi'(\theta)/\varphi(\theta)$ 递增至有限极限 a_0. 若 $a_0 \le a < \infty$，则

$$n^{-1}\log P(S_n \ge na) \to -a\theta_+ + \log\varphi(\theta_+)$$

即 $\gamma(a)$ 对 $a \ge a_0$ 是线性的.

证明. 由于 $(\log\varphi(\theta))' = \varphi'(\theta)/\varphi(\theta)$，从 0 到 θ_+ 积分表明 $\log(\varphi(\theta_+)) < \infty$. 在 (2.7.2) 中令 $\theta = \theta_+$，表明左边的 lim sup 不超过右边. 为得到另一方向，对 $\lambda = \theta_+$ 使用变换后的分布 F_λ. 令 $\theta \uparrow \theta_+$，对 $x \le 0$ 使用控制收敛定理，对 $x \ge 0$ 使用单调收敛定理，可知 F_λ 的均值为 a_0. 由定理 2.7.7 证明中的 $(*)$，若 $a_0 \le a < \nu = a + 3\varepsilon$，

$$P(S_n \ge na) \ge \varphi(\lambda)^n e^{-n\lambda\nu}(F_\lambda^n(n\nu) - F_\lambda^n(na))$$

故

$$\frac{1}{n}\log P(S_n \ge na) \ge \log\varphi(\lambda) - \lambda\nu + \frac{1}{n}\log P(S_n^\lambda \in (na, n\nu])$$

令 $X_1^\lambda, X_2^\lambda, \ldots$ 独立同分布，具有分布 F_λ，且 $S_n^\lambda = X_1^\lambda + \cdots + X_n^\lambda$，有

$$\begin{aligned} P(S_n^\lambda \in (na, n\nu]) &\ge P\{S_{n-1}^\lambda \in ((a_0-\varepsilon)n, (a_0+\varepsilon)n]\} \\ &\quad \cdot P\{X_n^\lambda \in ((a-a_0+\varepsilon)n, (a-a_0+2\varepsilon)n]\} \\ &\ge \frac{1}{2}P\{X_n^\lambda \in ((a-a_0+\varepsilon)n, (a-a_0+\varepsilon)(n+1)]\} \end{aligned}$$

对大的 n 由弱大数定律成立. 为得到最后一式右边的下界，观察到

$$\limsup_{n\to\infty}\frac{1}{n}\log P(X_1^\lambda \in ((a-a_0+\varepsilon)n, (a-a_0+\varepsilon)(n+1)]) = 0$$

因为若 lim sup < 0，则对某个 $\eta > 0$ 有 $E\exp(\eta X_1^\lambda) < \infty$，从而 $E\exp((\lambda+\eta)X_1) < \infty$，与 $\lambda = \theta_+$ 的定义矛盾. 为完成论证，回顾定理 2.7.1 蕴含

$$\lim_{n\to\infty}\frac{1}{n}\log P(S_n \ge na) = \gamma(a)$$

存在，故对 lim sup 的下界已足够好. □

通过改编最后这个结果的证明，可以证明 (H1) 是指数收敛的必要条件：

习题

习题 2.7.1. 考虑 (2.7.1) 中定义的 $\gamma(a)$. 下列等价：(a) $\gamma(a) = -\infty$，(b) $P(X_1 \ge a) = 0$，以及 (c) 对所有 n 有 $P(S_n \ge na) = 0$.

习题 2.7.2. 用定义得出若 $\lambda \in [0,1]$ 是有理数，则 $\gamma(\lambda a + (1-\lambda)b) \ge \lambda\gamma(a) + (1-\lambda)\gamma(b)$. 用单调性得出最后一个关系对所有 $\lambda \in [0,1]$ 成立，故 γ 是凹的，从而在 $\gamma(a) > -\infty$ 的紧子集上是 Lipschitz 连续的.

习题 2.7.3. 设 $X_1, X_2, \ldots$ 独立同分布于均值为 1 的泊松分布，$S_n = X_1 + \cdots + X_n$. 求 $a > 1$ 时的 $\lim_{n\to\infty}(1/n)\log P(S_n \geq na)$. 答案和另一证明可在习题 3.1.4 中找到.

习题 2.7.4. 证明对抛硬币 $\varphi(\theta) \leq \exp(\varphi(\theta) - 1) \leq \exp(\beta\theta^2)$ 对 $\theta \leq 1$ 成立，其中 $\beta = \sum_{n=1}^{\infty} 1/(2n)! \approx 0.586$，并用 (2.7.2) 得出对所有 $a \in [0, 1]$ 有 $P(S_n \geq an) \leq \exp(-na^2/4\beta)$. 通常用 $\beta \leq \sum_{n=1}^{\infty} 2^{-n} = 1$ 进一步简化.

习题 2.7.5. 设 $EX_i = 0$ 且对所有 $\theta > 0$ 有 $E\exp(\theta X_i) = \infty$. 则

$$\frac{1}{n}\log P(S_n \geq na) \to 0, \quad \text{对所有} a > 0$$

习题 2.7.6. 设 $EX_i = 0$. 证明若 $\varepsilon > 0$，则

$$\liminf_{n\to\infty} P(S_n \geq na)/nP(X_1 \geq n(a + \varepsilon)) \geq 1$$

提示：令 $F_n = \{X_i \geq n(a + \varepsilon)$ 对恰好一个$i \leq n\}$.

第三章
中心极限定理

本章前四节旨在建立**中心极限定理**，后五节讨论各种推广与补充. 首先考察可通过初等计算处理的特殊情形.

第 1 节　De Moivre-Laplace 定理

设 $X_1, X_2, \ldots$ 独立同分布，且 $P(X_1 = 1) = P(X_1 = -1) = 1/2$，记 $S_n = X_1 + \cdots + X_n$. 换言之，考虑抛掷一枚均匀硬币时每次下注 \$1 的赌博模型，S_n 表示时刻 n 的累计收益. 若 n 和 k 为整数，则

$$P(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n}$$

这是因为 $S_{2n} = 2k$ 当且仅当前 $2n$ 次抛掷中有 $n+k$ 次为 $+1$、$n-k$ 次为 -1. 第一项给出所有可能结果的总数，第二项给出每种特定结果的概率. **Stirling 公式**（见 Feller，第一卷（1968），p. 52）表明

$$n! \sim n^n e^{-n} \sqrt{2\pi n} \quad \text{当} n \to \infty \tag{3.1.1}$$

其中 $a_n \sim b_n$ 表示当 $n \to \infty$ 时 $a_n / b_n \to 1$，于是

$$\begin{aligned} \binom{2n}{n+k} &= \frac{(2n)!}{(n+k)!(n-k)!} \\ &\sim \frac{(2n)^{2n} (2\pi(2n))^{1/2}}{(n+k)^{n+k}(n-k)^{n-k}(2\pi(n+k))^{1/2}(2\pi(n-k))^{1/2}} \end{aligned}$$

从而有

$$\begin{aligned}\binom{2n}{n+k}2^{-2n} \sim &\left(1+\frac{k}{n}\right)^{-n-k}\left(1-\frac{k}{n}\right)^{-n+k} \\ &\cdot(\pi n)^{-1/2}\cdot\left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2}\end{aligned} \tag{3.1.2}$$

右边前两项等于

$$\left(1-\frac{k^2}{n^2}\right)^{-n}\left(1+\frac{k}{n}\right)^{-k}\left(1-\frac{k}{n}\right)^{k}$$

下述引理可由微积分直接推出：

引理 3.1.1. 若 $c_j \to 0$，$a_j \to \infty$ 且 $a_j c_j \to \lambda$，则 $(1+c_j)^{a_j} \to e^{\lambda}$.

证明. 当 $x \to 0$ 时，$\log(1+x)/x \to 1$，故 $a_j \log(1+c_j) \to \lambda$，由此即得所求结果. □

现应用引理 3.1.1，若 $2k = x\sqrt{2n}$，即 $k = x\sqrt{n/2}$，则

$$\begin{aligned}\left(1-\frac{k^2}{n^2}\right)^{-n} &= \left(1-\frac{x^2}{2n}\right)^{-n} \to e^{x^2/2} \\ \left(1+\frac{k}{n}\right)^{-k} &= \left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}} \to e^{-x^2/2} \\ \left(1-\frac{k}{n}\right)^{k} &= \left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}} \to e^{-x^2/2}\end{aligned}$$

对此选取的 k，有 $k/n \to 0$，故

$$\left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2} \to 1$$

综合以上可得：

定理 3.1.2. 若 $2k/\sqrt{2n} \to x$，则 $P(S_{2n}=2k) \sim (\pi n)^{-1/2} e^{-x^2/2}$.

下一步计算

$$P(a\sqrt{2n} \le S_{2n} \le b\sqrt{2n}) = \sum_{m \in [a\sqrt{2n}, b\sqrt{2n}] \cap 2\mathbb{Z}} P(S_{2n}=m)$$

换元 $m = x\sqrt{2n}$，则上式可写为

$$\approx \sum_{x \in [a,b] \cap (2\mathbb{Z}/\sqrt{2n})} (2\pi)^{-1/2} e^{-x^2/2} \cdot (2/n)^{1/2}$$

其中 $2\mathbb{Z}/\sqrt{2n} = \{2z/\sqrt{2n} : z \in \mathbb{Z}\}$. 此处乘以并除以 $\sqrt{2}$，因求和中相邻点的间距为 $(2/n)^{1/2}$，故当 n 充分大时，上述求和

$$\approx \int_a^b (2\pi)^{-1/2} e^{-x^2/2}\,\mathrm{d}x$$

被积函数是（标准）**正态分布**的**密度**，故改写符号可将最后一式表为 $P(a \le \chi \le b)$，其中 χ 是服从该分布的随机变量.

填补具体细节并不困难，由此得到：

定理 3.1.3. De Moivre-Laplace 定理. 若 $a < b$，则当 $m \to \infty$ 时

$$P(a \le S_m/\sqrt{m} \le b) \to \int_a^b (2\pi)^{-1/2} e^{-x^2/2}\,\mathrm{d}x$$

（为去掉对偶数整数的限制，注意 $S_{2n+1} = S_{2n} \pm 1$. ）最后这个结果是 3.4 节将给出的**中心极限定理**的特例，进一步细节留给读者.

习题

习题 3.1.1. 推广引理 3.1.1 的证明，证明：若 $\max_{1\le j\le n} |c_{j,n}| \to 0$，$\sum_{j=1}^n c_{j,n} \to \lambda$，且 $\sup_n \sum_{j=1}^n |c_{j,n}| < \infty$，则 $\prod_{j=1}^n (1 + c_{j,n}) \to e^\lambda$.

以下三个习题说明 Stirling 公式的应用. 设其中 $X_1, X_2, \ldots$ 独立同分布，$S_n = X_1 + \cdots + X_n$.

习题 3.1.2. 若 X_i 服从均值为 1 的 **Poisson 分布**，则 S_n 服从均值为 n 的 Poisson 分布，即 $P(S_n = k) = e^{-n}n^k/k!$. 用 Stirling 公式证明：若 $(k-n)/\sqrt{n} \to x$，则

$$\sqrt{2\pi n} P(S_n = k) \to \exp(-x^2/2)$$

如同抛硬币情形，由此可得

$$P(a \le (S_n - n)/\sqrt{n} \le b) \to \int_a^b (2\pi)^{-1/2} e^{-x^2/2}\,\mathrm{d}x$$

但证明最后一个结论不是本习题的内容.

以下两例中，应先考察当 $k/n \to a$ 时的 $P(S_n = k)$，然后通过 $P(S_n = j+1)$ 与 $P(S_n = j)$ 的关系，证明 $P(S_n \ge k) \le CP(S_n = k)$.

习题 3.1.3. 设 $P(X_i = 1) = P(X_i = -1) = 1/2$. 证明：若 $a \in (0, 1)$，则

$$\frac{1}{2n} \log P(S_{2n} \ge 2na) \to -\gamma(a)$$

其中 $\gamma(a)=\frac{1}{2}\{(1+a)\log(1+a)+(1-a)\log(1-a)\}$.

习题 3.1.4. 设 $P(X_i=k)=e^{-1}/k!$ $(k=0,1,\ldots)$. 证明: 若 $a>1$, 则

$$\frac{1}{n}\log P(S_n\geq na)\to a-1-a\log a$$

第 2 节 弱收敛

本节将定义中心极限定理中出现的收敛类型，并探讨其部分性质. 称分布函数序列**弱收敛**于极限 F（记作 $F_n \Rightarrow F$），若对所有 F 的**连续点** y 都有 $F_n(y) \to F(y)$. 称随机变量序列 X_n **弱收敛**或**依分布收敛**于极限 X_∞（记作 $X_n \Rightarrow X_\infty$），若其分布函数 $F_n(x) = P(X_n \le x)$ 弱收敛. 为说明在连续点处的收敛足以确定极限，注意 F 右连续，且由习题 1.2.3 知 F 的不连续点至多可数.

3.2.1 例子

前两个弱收敛的例子如下：

例题 3.2.1. 设 $X_1, X_2, \ldots$ 独立同分布，$P(X_i = 1) = P(X_i = -1) = 1/2$，令 $S_n = X_1 + \cdots + X_n$. 则定理 3.1.3 蕴含

$$F_n(y) = P(S_n/\sqrt{n} \le y) \to \int_{-\infty}^{y} (2\pi)^{-1/2} e^{-x^2/2}\, \mathrm{d}x.$$

例题 3.2.2. 设 $X_1, X_2, \ldots$ 独立同分布于 F. **Glivenko-Cantelli 定理**（定理 8.4.1）蕴含对几乎处处的 ω，

$$F_n(y) = n^{-1} \sum_{m=1}^{n} 1_{(X_m(\omega) \le y)} \to F(y) \text{ 对所有} y.$$

在前两个例子中，收敛对所有 y 成立，即使第二个例子中的分布函数可能有不连续点. 下例说明为何要将注意力限制在连续点.

例题 3.2.3. 设 X 的分布为 F. 则 $X + 1/n$ 的分布为

$$F_n(x) = P(X + 1/n \le x) = F(x - 1/n).$$

当 $n \to \infty$ 时，$F_n(x) \to F(x-) = \lim_{y \uparrow x} F(y)$，故仅在连续点处收敛.

例题 3.2.4. 等待稀有事件. 设 X_p 为在成功概率 p 的独立试验序列中首次成功所需的试验次数. 则 $P(X_p \ge n) = (1-p)^{n-1}$（$n = 1, 2, 3, \ldots$），由引理 3.1.1 知当 $p \to 0$ 时，

$$P(pX_p > x) \to e^{-x} \text{ 对所有} x \ge 0.$$

换言之，pX_p 弱收敛于指数分布.

例题 3.2.5. 生日问题. 设 $X_1, X_2, \ldots$ 独立且在 $\{1, \ldots, N\}$ 上均匀分布，令 $T_N = \min\{n : X_n = X_m \text{ 对某个} m < n\}$.

$$P(T_N > n) = \prod_{m=2}^{n} \left(1 - \frac{m-1}{N}\right).$$

当 $N = 365$ 时，这是 n 人中没有两人生日相同的概率（假设所有生日等可能）. 利用习题 3.1.1，不难看出

$$P(T_N/N^{1/2} > x) \to \exp(-x^2/2) \text{ 对所有} x \geq 0.$$

取 $N = 365$ 并注意 $22/\sqrt{365} = 1.1515$ 且 $(1.1515)^2/2 = 0.6630$，这表明

$$P(T_{365} > 22) \approx e^{-0.6630} \approx 0.515.$$

此答案比真实概率 0.524 低 2%.

在给出第六个例子之前，需要一个称为 **Scheffé 定理**的简单结果. 设有概率密度 f_n（$1 \leq n \leq \infty$），且 $f_n \to f_\infty$ 点态收敛. 则对所有 Borel 集 B，

$$\left| \int_B f_n(x)\,\mathrm{d}x - \int_B f_\infty(x)\,\mathrm{d}x \right| \leq \int |f_n(x) - f_\infty(x)|\,\mathrm{d}x = 2\int (f_\infty(x) - f_n(x))^+\,\mathrm{d}x \to 0$$

（由控制收敛定理，等式源于 $f_n \geq 0$ 且积分为 1 的事实）. 记 μ_n 为相应的测度，则已证**全变差范数**

$$\|\mu_n - \mu_\infty\| \equiv \sup_B |\mu_n(B) - \mu_\infty(B)| \to 0$$

这是比弱收敛更强的结论（取 $B = (-\infty, x]$ 即可）. 例子 $\mu_n = 1/n$ 处的点质量（$1/\infty = 0$）表明，可以有 $\mu_n \Rightarrow \mu_\infty$ 但 $\|\mu_n - \mu_\infty\| = 1$ 对所有 n 成立.

例题 3.2.6. 中心顺序统计量. 在 $(0,1)$ 中随机放置 $(2n+1)$ 个点，即位置独立且均匀分布. 令 V_{n+1} 为第 $(n+1)$ 大的点. 不难看出

引理 3.2.7. V_{n+1} 具有密度函数

$$f_{V_{n+1}}(x) = (2n+1)\binom{2n}{n}x^n(1-x)^n.$$

证明. 有 $2n+1$ 种方式选取落在 x 处的观测值，然后需选取 n 个指标使观测值 $< x$，这可以用 $\binom{2n}{n}$ 种方式完成. 一旦确定了哪些指标落在 $< x$ 和 $> x$，相应随机变量按预期行动的概率为 $x^n(1-x)^n$，剩余一个落在 x 处的概率密度为 1. 若不喜欢上一句，可计算 $P(X_1 < x-\varepsilon, \ldots, X_n < x-\varepsilon, x-\varepsilon < X_{n+1} < x+\varepsilon, X_{n+2} > x+\varepsilon, \ldots, X_{2n+1} > x+\varepsilon)$ 再令 $\varepsilon \to 0$. □

为计算 $Y_n = 2(V_{n+1} - 1/2)\sqrt{2n}$ 的密度函数，可用习题 1.2.5，或直接换元 $x = 1/2 + y/2\sqrt{2n}$，$\mathrm{d}x = \mathrm{d}y/2\sqrt{2n}$ 得

$$f_{Y_n}(y) = (2n+1)\binom{2n}{n}\left(\frac{1}{2} + \frac{y}{2\sqrt{2n}}\right)^n\left(\frac{1}{2} - \frac{y}{2\sqrt{2n}}\right)^n\frac{1}{2\sqrt{2n}}$$

$$= \binom{2n}{n} 2^{-2n} \cdot (1 - y^2/2n)^n \cdot \frac{2n+1}{2n} \cdot \sqrt{\frac{n}{2}}.$$

第一个因子是简单随机游走的 $P(S_{2n} = 0)$，故定理 3.1.2 和引理 3.1.1 蕴含

$$f_{Y_n}(y) \to (2\pi)^{-1/2} \exp(-y^2/2) \text{ 当} n \to \infty.$$

这里及以后用 $P(Y_n = y)$ 记 Y_n 的密度函数. 现用 Scheffé 定理，断定 Y_n 弱收敛于标准正态分布.

3.2.2 理论

下面的结果对证明弱收敛有关的事实很有用.

定理 3.2.8. Skorokhod 表示定理. 若 $F_n \Rightarrow F_\infty$，则存在随机变量 Y_n（$1 \le n \le \infty$），分别以 F_n 为分布，使得 $Y_n \to Y_\infty$ 几乎必然.

证明. 设 $\Omega = (0,1)$，$\mathcal{F}$ 为 Borel 集，P 为 Lebesgue 测度，并令 $Y_n(x) = \sup\{y : F_n(y) < x\}$. 由定理 1.2.2，$Y_n$ 的分布为 F_n. 下面证明 $Y_n(x) \to Y_\infty(x)$ 对除可数个点外的所有 x 成立. 为此，方便的做法是将 $Y_n(x)$ 写成 $F_n^{-1}(x)$ 并在 $n = \infty$ 时省略下标. 首先确定例外集. 令 $a_x = \sup\{y : F(y) < x\}$，$b_x = \inf\{y : F(y) > x\}$，以及 $\Omega_0 = \{x : (a_x, b_x) = \emptyset\}$，其中 (a_x, b_x) 为以所示端点开区间. $\Omega - \Omega_0$ 可数，因为 (a_x, b_x) 不交且每个非空区间含有不同的有理数. 若 $x \in \Omega_0$，则 $F(y) < x$ 对 $y < F^{-1}(x)$，且 $F(z) > x$ 对 $z > F^{-1}(x)$. 为证对 $x \in \Omega_0$ 有 $F_n^{-1}(x) \to F^{-1}(x)$，需证两件事：

(a) $\liminf_{n\to\infty} F_n^{-1}(x) \ge F^{-1}(x)$

(a) 的证明. 设 $y < F^{-1}(x)$ 且 F 在 y 连续. 由于 $x \in \Omega_0$，$F(y) < x$ 且当 n 充分大时 $F_n(y) < x$，即 $F_n^{-1}(x) \ge y$. 因这对满足所述限制的所有 y 成立，故得结论.

(b) $\limsup_{n\to\infty} F_n^{-1}(x) \le F^{-1}(x)$

(b) 的证明. 设 $y > F^{-1}(x)$ 且 F 在 y 连续. 由于 $x \in \Omega_0$，$F(y) > x$ 且当 n 充分大时 $F_n(y) > x$，即 $F_n^{-1}(x) \le y$. 因这对满足所述限制的所有 y 成立，故得结论，从而完成证明. □

由定理 3.2.8 可立即推广之前的某些结果.

下面的结果说明定理 3.2.8 的用途，并给出弱收敛的一个等价定义，该定义在任何拓扑空间中都适用.

定理 3.2.9. $X_n \Rightarrow X_\infty$ 当且仅当对每个有界连续函数 g 都有 $Eg(X_n) \to Eg(X_\infty)$.

证明. 设 Y_n 与 X_n 同分布且几乎必然收敛. 由于 g 连续，$g(Y_n) \to g(Y_\infty)$ 几乎必然，由有界收敛定理，

$$Eg(X_n) = Eg(Y_n) \to Eg(Y_\infty) = Eg(X_\infty).$$

为证逆命题，令

$$g_{x,\varepsilon}(y)=\begin{cases}1 & y\le x,\\ 0 & y\ge x+\varepsilon,\\ \text{线性} & x\le y\le x+\varepsilon.\end{cases}$$

由于 $g_{x,\varepsilon}(y)=1$ 对 $y\le x$，$g_{x,\varepsilon}$ 连续，且 $g_{x,\varepsilon}(y)=0$ 对 $y>x+\varepsilon$，

$$\limsup_{n\to\infty}P(X_n\le x)\le\limsup_{n\to\infty}Eg_{x,\varepsilon}(X_n)=Eg_{x,\varepsilon}(X_\infty)\le P(X_\infty\le x+\varepsilon).$$

令 $\varepsilon\to 0$ 得 $\limsup_{n\to\infty}P(X_n\le x)\le P(X_\infty\le x)$. 最后这对任意 x 都成立. 为得到另一方向，注意

$$\liminf_{n\to\infty}P(X_n\le x)\ge\liminf_{n\to\infty}Eg_{x-\varepsilon,\varepsilon}(X_n)=Eg_{x-\varepsilon,\varepsilon}(X_\infty)\ge P(X_\infty\le x-\varepsilon).$$

令 $\varepsilon\to 0$ 得 $\liminf_{n\to\infty}P(X_n\le x)\ge P(X_\infty<x)=P(X_\infty\le x)$ 若 x 是连续点. 关于 lim sup 和 lim inf 的结果结合起来即得所求结论. □

下面的结果是定理 3.2.9 的一个平凡但有用的推广.

定理 3.2.10. 连续映射定理. 设 g 为可测函数，$D_g=\{x:g\text{在}x\text{处不连续}\}$. 若 $X_n\Rightarrow X_\infty$ 且 $P(X_\infty\in D_g)=0$，则 $g(X_n)\Rightarrow g(X)$. 进一步，若 g 有界，则 $Eg(X_n)\to Eg(X_\infty)$.

注释. D_g 总是 Borel 集. 见习题 1.3.6.

证明. 设 $Y_n\overset{d}{=}X_n$ 且 $Y_n\to Y_\infty$ 几乎必然. 若 f 连续，则 $D_{f\circ g}\subset D_g$，故 $P(Y_\infty\in D_{f\circ g})=0$，从而 $f(g(Y_n))\to f(g(Y_\infty))$ 几乎必然. 进一步，若 f 有界，则由有界收敛定理 $Ef(g(Y_n))\to Ef(g(Y_\infty))$. 因这对所有有界连续函数成立，由定理 3.2.9 知 $g(X_n)\Rightarrow g(X_\infty)$.

第二个结论更易证. 由于 $P(Y_\infty\in D_g)=0$，$g(Y_n)\to g(Y_\infty)$ 几乎必然，所求结论由有界收敛定理可得. □

下面的结果给出弱收敛的若干有用等价刻画.

定理 3.2.11. 下列陈述等价：

(i) $X_n\Rightarrow X_\infty$.

(ii) 对所有开集 G，$\liminf_{n\to\infty}P(X_n\in G)\ge P(X_\infty\in G)$.

(iii) 对所有闭集 K，$\limsup_{n\to\infty}P(X_n\in K)\le P(X_\infty\in K)$.

(iv) 对所有满足 $P(X_\infty\in\partial A)=0$ 的 Borel 集 A，$\lim_{n\to\infty}P(X_n\in A)=P(X_\infty\in A)$.

注释. 为帮助记住 (ii) 和 (iii) 中不等式的方向，考虑 $P(X_n=x_n)=1$ 的特殊情形. 在此情形，若 $x_n\in G$ 且 $x_n\to x_\infty\in\partial G$，则 $P(X_n\in G)=1$ 对所有 n 但 $P(X_\infty\in G)=0$. 令 $K=G^c$ 给出 (iii) 的一个例子.

证明. 下面证明四件事，留给读者验证这已证得上述结果.

(i) 推出 (ii)： 设 Y_n 与 X_n 同分布且 $Y_n \to Y_\infty$ 几乎必然. 由于 G 开，

$$\liminf_{n\to\infty} 1_G(Y_n) \geq 1_G(Y_\infty)$$

故 Fatou 引理蕴含

$$\liminf_{n\to\infty} P(Y_n \in G) \geq P(Y_\infty \in G).$$

(ii) 等价于 (iii)： 这易由以下事实推出：A 开当且仅当 A^c 闭，且 $P(A) + P(A^c) = 1$.

(ii) 和 (iii) 推出 (iv)： 令 $K = \bar{A}$ 且 $G = A^\circ$ 分别为 A 的闭包和内部. A 的边界 $\partial A = \bar{A} - A^\circ$ 且 $P(X_\infty \in \partial A) = 0$，故

$$P(X_\infty \in K) = P(X_\infty \in A) = P(X_\infty \in G).$$

现用 (ii) 和 (iii)，

$$\limsup_{n\to\infty} P(X_n \in A) \leq \limsup_{n\to\infty} P(X_n \in K) \leq P(X_\infty \in K) = P(X_\infty \in A)$$

$$\liminf_{n\to\infty} P(X_n \in A) \geq \liminf_{n\to\infty} P(X_n \in G) \geq P(X_\infty \in G) = P(X_\infty \in A).$$

(iv) 推出 (i)： 设 x 满足 $P(X_\infty = x) = 0$，即 x 是 F 的连续点，并令 $A = (-\infty, x]$. □

下面的结果对研究分布序列的极限很有用.

定理 3.2.12. Helly 选择定理. 对每个分布函数序列 F_n，存在子序列 $F_{n(k)}$ 及右连续非减函数 F，使得 $\lim_{k\to\infty} F_{n(k)}(y) = F(y)$ 在 F 的所有连续点 y 处成立.

注释. 极限未必是分布函数. 例如若 $a + b + c = 1$ 且 $F_n(x) = a \cdot 1_{(x\geq n)} + b \cdot 1_{(x\geq -n)} + c \cdot G(x)$，其中 G 是分布函数，则 $F_n(x) \to F(x) = b + cG(x)$，

$$\lim_{x\downarrow -\infty} F(x) = b \text{ 且 } \lim_{x\uparrow\infty} F(x) = b + c = 1 - a.$$

换言之，质量 a 逃逸到 $+\infty$，质量 b 逃逸到 $-\infty$. 定理 3.2.12 中出现的这种收敛有时称为**淡收敛**，这里记作 $\Rightarrow_v$.

证明. 第一步是对角线论证. 设 $q_1, q_2, \ldots$ 为有理数的一个枚举. 因对每个 k，$F_m(q_k) \in [0,1]$ 对所有 m 成立，存在序列 $m_k(i) \to \infty$（它是 $m_{k-1}(j)$ 的子序列，令 $m_0(j) \equiv j$）使得

$$F_{m_k(i)}(q_k) \text{ 收敛到} G(q_k) \text{ 当} i \to \infty.$$

令 $F_{n(k)} = F_{m_k(k)}$. 由构造 $F_{n(k)}(q) \to G(q)$ 对所有有理数 q 成立. 函数 G 可能不右连续，但 $F(x) =$

$\inf\{G(q) : q \in \mathbb{Q}, q > x\}$ 是右连续的，因为

$$\lim_{x_n \downarrow x} F(x_n) = \inf\{G(q) : q \in \mathbb{Q}, q > x_n \text{ 对某个} n\} = \inf\{G(q) : q \in \mathbb{Q}, q > x\} = F(x).$$

为完成证明，设 x 是 F 的连续点. 取有理数 r_1, r_2, s 满足 $r_1 < r_2 < x < s$ 使得

$$F(x) - \varepsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(s) < F(x) + \varepsilon.$$

由于 $F_{n(k)}(r_2) \to G(r_2) \geq F(r_1)$，且 $F_{n(k)}(s) \to G(s) \leq F(s)$，可知若 k 充分大，

$$F(x) - \varepsilon < F_{n(k)}(r_2) \leq F_{n(k)}(x) \leq F_{n(k)}(s) < F(x) + \varepsilon$$

这正是所求结论. □

最后一个结果引出一个问题：何时能在定理 3.2.12 的极限中保证没有质量损失?

定理 3.2.13. 每个子序列极限都是概率测度的分布函数当且仅当序列 F_n **胎紧**，即对所有 $\varepsilon > 0$ 存在 M_ε 使得

$$\limsup_{n\to\infty} [1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)] \leq \varepsilon.$$

证明. 设序列胎紧且 $F_{n(k)} \Rightarrow_v F$. 令 $r < -M_\varepsilon$ 和 $s > M_\varepsilon$ 为 F 的连续点. 由于 $F_n(r) \to F(r)$ 且 $F_n(s) \to F(s)$，我们有

$$1 - F(s) + F(r) = \lim_{k\to\infty} \left[1 - F_{n(k)}(s) + F_{n(k)}(r)\right] \leq \limsup_{n\to\infty} [1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)] \leq \varepsilon.$$

最后这个结果蕴含 $\limsup_{x\to\infty} [1 - F(x) + F(-x)] \leq \varepsilon$. 因 ε 任意，知 F 是概率测度的分布函数.

为证逆命题，现设 F_n 不胎紧. 此时，存在 $\varepsilon > 0$ 和子序列 $n(k) \to \infty$ 使得

$$1 - F_{n(k)}(k) + F_{n(k)}(-k) \geq \varepsilon$$

对所有 k 成立. 通过取进一步的子序列 $F_{n(k_j)}$ 可设 $F_{n(k_j)} \Rightarrow_v F$. 令 $r < 0 < s$ 为 F 的连续点.

$$1 - F(s) + F(r) = \lim_{j\to\infty} \left[1 - F_{n(k_j)}(s) + F_{n(k_j)}(r)\right] \geq \liminf_{j\to\infty} \left[1 - F_{n(k_j)}(k_j) + F_{n(k_j)}(-k_j)\right] \geq \varepsilon.$$

令 $s \to \infty$ 且 $r \to -\infty$，可见 F 不是概率测度的分布函数. □

下面的胎紧充分条件经常有用.

定理 3.2.14. 若存在 $\varphi \geq 0$ 使得 $\varphi(x) \to \infty$ 当 $|x| \to \infty$ 且

$$C = \sup_n \int \varphi(x)\, \mathrm{d}F_n(x) < \infty$$

则 F_n 胎紧.

证明. $1 - F_n(M) + F_n(-M) \leq C/\inf_{|x|\geq M} \varphi(x)$. □

习题 3.2.6 和 3.2.7 定义了依分布收敛的度量. 依分布收敛可由度量导出这一事实立即蕴含

定理 3.2.15. 若 X_n 的每个子序列都有进一步收敛到 X 的子序列，则 $X_n \Rightarrow X$.

我们将在定理 3.3.17 的证明末尾再次证明这一点.

习题

习题 3.2.1. 举例说明存在具有密度 f_n 的随机变量 X_n，使得 X_n 弱收敛于 $(0,1)$ 上的均匀分布，但 $f_n(x)$ 对任何 $x \in [0,1]$ 都不收敛到 1.

习题 3.2.2. 最大值的收敛. 设 $X_1, X_2, \ldots$ 独立同分布于 F，令 $M_n = \max_{m\leq n} X_m$. 则 $P(M_n \leq x) = F(x)^n$. 证明 M_n 的下述极限定律：

(i) 若 $F(x) = 1 - x^{-\alpha}$ 对 $x \geq 1$（其中 $\alpha > 0$），则对 $y > 0$

$$P(M_n/n^{1/\alpha} \leq y) \to \exp(-y^{-\alpha}).$$

(ii) 若 $F(x) = 1 - |x|^{\beta}$ 对 $-1 \leq x \leq 0$（其中 $\beta > 0$），则对 $y < 0$

$$P(n^{1/\beta} M_n \leq y) \to \exp(-|y|^{\beta}).$$

(iii) 若 $F(x) = 1 - e^{-x}$ 对 $x \geq 0$，则对所有 $y \in (-\infty, \infty)$

$$P(M_n - \log n \leq y) \to \exp(-e^{-y}).$$

上述出现的极限称为**极值分布**. 最后一个称为**双指数分布**或 **Gumbel 分布**. Gnedenko（1943）得到了 $(M_n - b_n)/a_n$ 收敛到这些极限的充要条件. 新近的处理见 Resnick（1987）.

习题 3.2.3. 设 $X_1, X_2, \ldots$ 独立同分布于标准正态分布.

(i) 由定理 1.2.6 知

$$P(X_i > x) \sim \frac{1}{\sqrt{2\pi}\, x} e^{-x^2/2} \text{ 当} x \to \infty.$$

利用此结论证明对任意实数 θ

$$P(X_i > x + (\theta/x))/P(X_i > x) \to e^{-\theta}.$$

(ii) 证明若定义 b_n 使得 $P(X_i > b_n) = 1/n$，则

$$P(b_n(M_n - b_n) \le x) \to \exp(-e^{-x}).$$

(iii) 证明 $b_n \sim (2\log n)^{1/2}$ 并推出 $M_n/(2\log n)^{1/2} \to 1$ 依概率.

习题 3.2.4. Fatou 引理. 设 $g \ge 0$ 连续. 若 $X_n \Rightarrow X_\infty$，则

$$\liminf_{n\to\infty} Eg(X_n) \ge Eg(X_\infty).$$

习题 3.2.5. 极限积分. 设 g, h 连续，$g(x) > 0$，且 $|h(x)|/g(x) \to 0$ 当 $|x| \to \infty$. 若 $F_n \Rightarrow F$ 且 $\int g(x)\,\mathrm{d}F_n(x) \le C < \infty$，则

$$\int h(x)\,\mathrm{d}F_n(x) \to \int h(x)\,\mathrm{d}F(x).$$

习题 3.2.6. Lévy 度量. 证明

$$\rho(F, G) = \inf\{\varepsilon : F(x-\varepsilon) - \varepsilon \le G(x) \le F(x+\varepsilon) + \varepsilon \text{ 对所有} x\}$$

在分布空间上定义了一个度量，且 $\rho(F_n, F) \to 0$ 当且仅当 $F_n \Rightarrow F$.

习题 3.2.7. 随机变量上的 **Ky Fan 度量**定义为

$$\alpha(X, Y) = \inf\{\varepsilon \ge 0 : P(|X - Y| > \varepsilon) \le \varepsilon\}.$$

证明若 $\alpha(X, Y) = \alpha$ 则相应分布的 Lévy 距离 $\rho(F, G) \le \alpha$.

习题 3.2.8. 设 $\alpha(X, Y)$ 为前一习题中的度量，$\beta(X, Y) = E(|X - Y|/(1 + |X - Y|))$ 为习题 2.3.6 中的度量. 若 $\alpha(X, Y) = a$ 则

$$a^2/(1 + a) \le \beta(X, Y) \le a + (1 - a)a/(1 + a).$$

习题 3.2.9. 若 $F_n \Rightarrow F$ 且 F 连续，则 $\sup_x |F_n(x) - F(x)| \to 0$.

习题 3.2.10. 若 F 是任意分布函数，则存在形如 $\sum_{m=1}^{n} a_{n,m} 1_{(x_{n,m} \le x)}$ 的分布函数序列使得 $F_n \Rightarrow F$.

习题 3.2.11. 设 X_n（$1 \le n \le \infty$）取整数值. 证明 $X_n \Rightarrow X_\infty$ 当且仅当 $P(X_n = m) \to P(X_\infty = m)$ 对所有 m 成立.

习题 3.2.12. 证明若 $X_n \to X$ 依概率，则 $X_n \Rightarrow X$；反之，若 $X_n \Rightarrow c$（其中 c 为常数），则 $X_n \to c$ 依概率.

习题 3.2.13. 收敛联合引理. 若 $X_n \Rightarrow X$ 且 $Y_n \Rightarrow c$（其中 c 为常数），则 $X_n + Y_n \Rightarrow X + c$. 此结果的一个有用推论是，若 $X_n \Rightarrow X$ 且 $Z_n - X_n \Rightarrow 0$，则 $Z_n \Rightarrow X$.

习题 3.2.14. 设 $X_n \Rightarrow X$，$Y_n \geq 0$，且 $Y_n \Rightarrow c$（其中 $c > 0$ 为常数），则 $X_n Y_n \Rightarrow cX$. 此结果在去掉 $Y_n \geq 0$ 和 $c > 0$ 的假设后仍成立，这里加上这些只是为了使证明不那么繁琐.

习题 3.2.15. 证明若 $X_n = (X_{n1}, \ldots, X_{nn})$ 在 $\mathbb{R}^n$ 中半径为 $\sqrt{n}$ 的球面上均匀分布，则 $X_{n1} \Rightarrow$ 标准正态分布.

提示： 设 $Y_1, Y_2, \ldots$ 为独立同分布于标准正态，令 $X_{ni} = Y_i(n/\sum_{m=1}^{n} Y_m^2)^{1/2}$.

习题 3.2.16. 设 $Y_n \geq 0$，$EY_n^\alpha \to 1$ 且 $EY_n^\beta \to 1$ 对某 $0 < \alpha < \beta$. 证明 $Y_n \to 1$ 依概率.

习题 3.2.17. 对每个 $K < \infty$ 和 $y < 1$，存在 $c_{y,K} > 0$ 使得 $EX^2 = 1$ 且 $EX^4 \leq K$ 蕴含 $P(|X| > y) \geq c_{y,K}$.

第 3 节 特征函数

本节较长，分为五个部分. 前三部分为必读内容，后两部分为选读内容. 第一部分证明特征函数 $\varphi(t)=E\exp(itX)$ 确定了 $F(x)=P(X\leq x)$，并给出从 φ 计算 F 的方法. 第二部分将分布的弱收敛与相应特征函数的性质联系起来. 第三部分将 $\varphi(t)$ 在 0 处的性质与 X 的矩联系起来. 第四部分证明 Polya 准则，并用它构造一些著名而奇特的特征函数例子. 最后，第五部分讨论矩问题，即何时一个分布由其矩唯一确定.

3.3.1 定义、反演公式

若 X 为随机变量，定义其**特征函数**（缩写为 ch.f.）为

$$\varphi(t)=Ee^{itX}=E\cos(tX)+iE\sin(tX).$$

最后一个公式涉及复值随机变量的期望，但正如第二个等式所示，这并不需要引入新的理论. 若 Z 为复值随机变量，定义 $EZ=E(\operatorname{Re}Z)+iE(\operatorname{Im}Z)$，其中 $\operatorname{Re}(a+bi)=a$ 为**实部**，$\operatorname{Im}(a+bi)=b$ 为**虚部**. 还需如下定义：复数 $z=a+bi$ 的**模**为 $|a+bi|=(a^2+b^2)^{1/2}$，而 $z=a+bi$ 的**复共轭**为 $\bar{z}=a-bi$.

定理 3.3.1. 所有特征函数具有以下性质：

(a) $\varphi(0)=1$,

(b) $\varphi(-t)=\overline{\varphi(t)}$,

(c) $|\varphi(t)|=|Ee^{itX}|\leq E|e^{itX}|=1$,

(d) $|\varphi(t+h)-\varphi(t)|\leq E|e^{ihX}-1|$，故 $\varphi(t)$ 在 $(-\infty,\infty)$ 上一致连续.

(e) $Ee^{it(aX+b)}=e^{itb}\varphi(at)$.

证明. (a) 显然. 对于 (b)，注意

$$\varphi(-t)=E(\cos(-tX)+i\sin(-tX))=E(\cos(tX)-i\sin(tX)).$$

(c) 由习题 1.6.2 推出，因为 $\varphi(x,y)=(x^2+y^2)^{1/2}$ 是凸函数.

$$|\varphi(t+h)-\varphi(t)|=|E(e^{i(t+h)X}-e^{itX})|\leq E|e^{i(t+h)X}-e^{itX}|=E|e^{ihX}-1|,$$

故一致连续性由有界收敛定理推出. 对于 (e)，注意 $Ee^{it(aX+b)}=e^{itb}Ee^{i(ta)X}=e^{itb}\varphi(at)$. □

引入特征函数的主要原因是下面这个结果：

定理 3.3.2. 若 X_1 与 X_2 独立且具有特征函数 φ_1 和 φ_2，则 X_1+X_2 的特征函数为 $\varphi_1(t)\varphi_2(t)$.

证明.

$$Ee^{it(X_1+X_2)} = E(e^{itX_1}e^{itX_2}) = Ee^{itX_1}Ee^{itX_2},$$

因为 e^{itX_1} 与 e^{itX_2} 独立. □

下面给出一些例子.

例题 3.3.3. 抛硬币. 若 $P(X=1) = P(X=-1) = 1/2$，则

$$Ee^{itX} = (e^{it} + e^{-it})/2 = \cos t.$$

例题 3.3.4. Poisson 分布. 若 $P(X=k) = e^{-\lambda}\lambda^k/k!$ $(k = 0, 1, 2, \ldots)$，则

$$Ee^{itX} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it}-1)).$$

例题 3.3.5. 正态分布

$$\begin{aligned} \text{密度} \quad & (2\pi)^{-1/2}\exp(-x^2/2) \\ \text{特征函数} \quad & \exp(-t^2/2). \end{aligned}$$

将此结果与定理 3.3.1(e) 结合，可见均值为 μ、方差为 σ^2 的正态分布的特征函数为 $\exp(i\mu t - \sigma^2 t^2/2)$. 类似的缩放可应用于其他例子，因此通常只给出分布族中一个成员的特征函数.

物理证明.

$$\int e^{itx}(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x = e^{-t^2/2}\int (2\pi)^{-1/2}e^{-(x-it)^2/2}\,\mathrm{d}x.$$

该积分为 1，因为被积函数是均值为 it、方差为 1 的正态密度. □

数学证明. 既然已通过 "作弊" 得出了答案，下面用一个形式计算来验证它，而这个计算几乎没有给出为什么成立的任何洞察. 记

$$\varphi(t) = \int e^{itx}(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x = \int \cos(tx)(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x,$$

因为 $i\sin(tx)$ 是奇函数. 对 t 求导（其合理性参见定理 A.5.1），然后分部积分得

$$\begin{aligned} \varphi'(t) &= -\int x\sin(tx)(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x \\ &= -t\int \cos(tx)(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x = -t\varphi(t). \end{aligned}$$

这表明 $\frac{\mathrm{d}}{\mathrm{d}t}\{\varphi(t)\exp(t^2/2)\}=0$，故 $\varphi(t)\exp(t^2/2)=\varphi(0)=1$. □

在接下来的三个例子中，密度在所示范围外为零.

例题 3.3.6. (a,b) **上的均匀分布**

$$\begin{aligned}\text{密度}\quad & 1/(b-a)\quad x\in(a,b)\\ \text{特征函数}\quad & (e^{itb}-e^{ita})/it(b-a).\end{aligned}$$

在特殊情形 $a=-c$，$b=c$ 下，特征函数为 $(e^{itc}-e^{-itc})/2cit=(\sin ct)/ct$.

证明. 回忆 $\int_a^b e^{\lambda x}\,\mathrm{d}x=(e^{\lambda b}-e^{\lambda a})/\lambda$ 对复数 λ 成立，结论立得. □

例题 3.3.7. 三角分布

$$\begin{aligned}\text{密度}\quad & 1-|x|\quad x\in(-1,1)\\ \text{特征函数}\quad & 2(1-\cos t)/t^2.\end{aligned}$$

证明. 注意若 X 与 Y 独立且均在 $(-1/2,1/2)$ 上均匀分布，则 $X+Y$ 服从三角分布. 利用例 3.3.6 和定理 3.3.2，所求特征函数为

$$\{(e^{it/2}-e^{-it/2})/it\}^2=\{2\sin(t/2)/t\}^2.$$

利用三角恒等式 $\cos 2\theta=1-2\sin^2\theta$（取 $\theta=t/2$）可将答案化为上述形式. □

例题 3.3.8. 指数分布

$$\begin{aligned}\text{密度}\quad & e^{-x}\quad x\in(0,\infty)\\ \text{特征函数}\quad & 1/(1-it).\end{aligned}$$

证明. 积分得

$$\int_0^\infty e^{itx}e^{-x}\,\mathrm{d}x=\left.\frac{e^{(it-1)x}}{it-1}\right|_0^\infty=\frac{1}{1-it},$$

因为当 $x\to\infty$ 时 $\exp((it-1)x)\to 0$. □

对于下一个结果，需要以下由 $\int f\,\mathrm{d}(\mu+\nu)=\int f\,\mathrm{d}\mu+\int f\,\mathrm{d}\nu$ 推出的事实.

引理 3.3.9. 若 $F_1,\ldots,F_n$ 具有特征函数 $\varphi_1,\ldots,\varphi_n$ 且 $\lambda_i\geq 0$ 满足 $\lambda_1+\cdots+\lambda_n=1$，则 $\sum_{i=1}^n\lambda_iF_i$ 的特征函数为 $\sum_{i=1}^n\lambda_i\varphi_i$.

例题 3.3.10. 双边指数分布

$$\begin{aligned} \text{密度} \quad & \frac{1}{2}e^{-|x|} \quad x \in (-\infty, \infty) \\ \text{特征函数} \quad & 1/(1+t^2). \end{aligned}$$

证明. 这由引理 3.3.9 推出，其中 F_1 为指数随机变量 X 的分布，F_2 为 $-X$ 的分布，$\lambda_1 = \lambda_2 = 1/2$，然后利用定理 3.3.1(b)，所求特征函数为

$$\frac{1}{2(1-it)} + \frac{1}{2(1+it)} = \frac{(1+it)+(1-it)}{2(1+t^2)} = \frac{1}{1+t^2}. \qquad \square$$

首先要解决的问题是特征函数唯一确定分布. 下面的定理给出了这一点，甚至更多:

定理 3.3.11. 反演公式. 设 $\varphi(t) = \int e^{itx}\mu(\mathrm{d}x)$，其中 μ 为概率测度. 若 $a < b$，则

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,\mathrm{d}t = \mu(a,b) + \frac{1}{2}\mu(\{a,b\}).$$

注释. 极限的存在性是结论的一部分. 若 $\mu = \delta_0$（在 0 处的点质量），$\varphi(t) \equiv 1$. 在此情形下，若 $a = -1$ 且 $b = 1$，则被积函数为 $(2\sin t)/t$，积分不绝对收敛.

证明. 记

$$I_T = \int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,\mathrm{d}t = \int_{-T}^{T}\int\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(\mathrm{d}x)\,\mathrm{d}t.$$

被积函数在 $t = 0$ 附近看起来不太好，但若注意

$$\frac{e^{-ita}-e^{-itb}}{it} = \int_a^b e^{-ity}\,\mathrm{d}y,$$

可见被积函数的模以 $b-a$ 为界. 由于 μ 是概率测度且 $[-T,T]$ 是有限区间，由 Fubini 定理、$\cos(-x) = \cos x$ 和 $\sin(-x) = -\sin x$ 得

$$\begin{aligned} I_T &= \int\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\,\mathrm{d}t\,\mu(\mathrm{d}x) \\ &= \int\left\{\int_{-T}^{T}\frac{\sin(t(x-a))}{t}\,\mathrm{d}t - \int_{-T}^{T}\frac{\sin(t(x-b))}{t}\,\mathrm{d}t\right\}\mu(\mathrm{d}x). \end{aligned}$$

引入 $R(\theta,T) = \int_{-T}^{T}(\sin\theta t)/t\,\mathrm{d}t$，可将最后结果写成

$$I_T = \int\{R(x-a,T) - R(x-b,T)\}\mu(\mathrm{d}x). \qquad (*)$$

若令 $S(T)=\int_0^T(\sin x)/x\,\mathrm{d}x$，则对 $\theta>0$ 作变量替换 $t=x/\theta$ 表明

$$R(\theta,T)=2\int_0^{T\theta}\frac{\sin x}{x}\,\mathrm{d}x=2S(T\theta),$$

而对 $\theta<0$，$R(\theta,T)=-R(|\theta|,T)$. 引入符号函数 $\operatorname{sgn}x$（当 $x>0$ 时为 1，$x<0$ 时为 -1，$x=0$ 时为 0），可将最后两个公式合写为

$$R(\theta,T)=2(\operatorname{sgn}\theta)S(T|\theta|).$$

当 $T\to\infty$ 时，$S(T)\to\pi/2$（见习题 1.7.5），故 $R(\theta,T)\to\pi\operatorname{sgn}\theta$ 且

$$R(x-a,T)-R(x-b,T)\to\begin{cases}2\pi & a<x<b,\\ \pi & x=a\text{ 或}x=b,\\ 0 & x<a\text{ 或}x>b.\end{cases}$$

$|R(\theta,T)|\le 2\sup_y S(y)<\infty$，故将控制收敛定理应用于 $(*)$ 得

$$(2\pi)^{-1}I_T\to\mu(a,b)+\frac{1}{2}\mu(\{a,b\}),$$

证得所需结果. □

反演公式的两个简单推论是：

推论 3.3.12. 若 φ 为实函数，则 X 与 $-X$ 同分布.

推论 3.3.13. 若 X_i（$i=1,2$）独立且服从均值为 0、方差为 σ_i^2 的正态分布，则 X_1+X_2 服从均值为 0、方差为 $\sigma_1^2+\sigma_2^2$ 的正态分布.

当 φ 可积时，反演公式更简单，但如下一结果所示，这仅当底层测度“好”时才会发生.

定理 3.3.14. 若 $\int|\varphi(t)|\,\mathrm{d}t<\infty$，则 μ 具有有界连续密度

$$f(y)=\frac{1}{2\pi}\int e^{-ity}\varphi(t)\,\mathrm{d}t.$$

证明. 如在定理 3.3.11 的证明中所见

$$\left|\frac{e^{-ita}-e^{-itb}}{it}\right|=\left|\int_a^b e^{-ity}\,\mathrm{d}y\right|\le|b-a|,$$

故在此情形下定理 3.3.11 中的积分绝对收敛且

$$\mu(a,b)+\frac{1}{2}\mu(\{a,b\})=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,\mathrm{d}t\le\frac{b-a}{2\pi}\int_{-\infty}^{\infty}|\varphi(t)|\,\mathrm{d}t.$$

最后结果表明 μ 没有点质量且

$$\begin{aligned}\mu(x,x+h)&=\frac{1}{2\pi}\int\frac{e^{-itx}-e^{-it(x+h)}}{it}\varphi(t)\,\mathrm{d}t\\&=\frac{1}{2\pi}\int\left\{\int_x^{x+h}e^{-ity}\,\mathrm{d}y\right\}\varphi(t)\,\mathrm{d}t\\&=\int_x^{x+h}\left\{\frac{1}{2\pi}\int e^{-ity}\varphi(t)\,\mathrm{d}t\right\}\mathrm{d}y,\end{aligned}$$

由 Fubini 定理，故分布 μ 具有密度函数

$$f(y)=\frac{1}{2\pi}\int e^{-ity}\varphi(t)\,\mathrm{d}t.$$

控制收敛定理蕴含 f 连续，证毕. □

定理 3.3.14 及下一结果表明，φ 在无穷远处的性质与底层测度的光滑性有关.

将反演公式定理 3.3.14 应用于例 3.3.7 和 3.3.10 中的特征函数，得到另外两个特征函数的例子. 第一个没有正式名称，给它一个名字以纪念它在证明 Polya 准则中的作用（见定理 3.3.22）.

例题 3.3.15. Polya 分布

$$\begin{aligned}\text{密度}\quad&(1-\cos x)/\pi x^2\\\text{特征函数}\quad&(1-|t|)^+.\end{aligned}$$

证明. 定理 3.3.14 蕴含

$$\frac{1}{2\pi}\int\frac{2(1-\cos s)}{s^2}e^{-isy}\,\mathrm{d}s=(1-|y|)^+.$$

现令 $s=x$，$y=-t$. □

例题 3.3.16. Cauchy 分布

$$\begin{aligned}\text{密度}\quad&1/\pi(1+x^2)\\\text{特征函数}\quad&\exp(-|t|).\end{aligned}$$

证明. 定理 3.3.14 蕴含

$$\frac{1}{2\pi}\int \frac{1}{1+s^2}e^{-isy}\,\mathrm{d}s = \frac{1}{2}e^{-|y|}.$$

现令 $s=x$，$y=-t$ 并将两边乘以 2. □

习题

习题 3.3.1. 证明若 φ 为特征函数，则 $\operatorname{Re}\varphi$ 和 $|\varphi|^2$ 也是特征函数.

习题 3.3.2. (i) 模仿定理 3.3.11 的证明证明

$$\mu(\{a\}) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\varphi(t)\,\mathrm{d}t.$$

(ii) 若 $P(X\in h\mathbb{Z})=1$（其中 $h>0$），则其特征函数满足 $\varphi(2\pi/h+t)=\varphi(t)$，故

$$P(X=x) = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\varphi(t)\,\mathrm{d}t \quad \text{对} x\in h\mathbb{Z}.$$

(iii) 若 $X=Y+b$，则 $E\exp(itX)=e^{itb}E\exp(itY)$. 故若 $P(X\in b+h\mathbb{Z})=1$，(ii) 中的反演公式对 $x\in b+h\mathbb{Z}$ 成立.

习题 3.3.3. 设 X 与 Y 独立且具有特征函数 φ 和分布 μ. 将习题 3.3.2 应用于 $X-Y$ 并利用习题 2.1.5 得

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\varphi(t)|^2\,\mathrm{d}t = P(X-Y=0) = \sum_x \mu(\{x\})^2.$$

注释. 最后一个结果表明，若 $\varphi(t)\to 0$（当 $t\to\infty$），则 μ 没有点质量. 习题 3.3.11 给出一个例子说明逆命题不成立. Riemann-Lebesgue 引理（习题 1.4.4）表明，若 μ 有密度，则 $\varphi(t)\to 0$（当 $t\to\infty$）.

习题 3.3.4. 给出一个测度 μ 具有密度但 $\int|\varphi(t)|\,\mathrm{d}t=\infty$ 的例子. 提示：上面的两个例子具有此性质.

习题 3.3.5. 证明若 $X_1,\dots,X_n$ 独立且在 $(-1,1)$ 上均匀分布，则对 $n\ge 2$，$X_1+\cdots+X_n$ 具有密度

$$f(x) = \frac{1}{\pi}\int_0^\infty (\sin t/t)^n\cos(tx)\,\mathrm{d}t.$$

虽然从公式并不明显，但 f 在每个区间 $(k,k+1)$（$k\in\mathbb{Z}$）上是多项式，且在 $[-n,n]^c$ 上为零.

习题 3.3.6. 利用例 3.3.16 的结果推出，若 $X_1,X_2,\dots$ 独立且服从 Cauchy 分布，则 $(X_1+\cdots+X_n)/n$ 与 X_1 同分布.

3.3.2 弱收敛

走向中心极限定理的下一步是将特征函数的收敛与弱收敛联系起来.

定理 3.3.17. 连续性定理. 设 μ_n（$1 \le n \le \infty$）为具有特征函数 φ_n 的概率测度.

(i) 若 $\mu_n \Rightarrow \mu_\infty$，则对所有 t，$\varphi_n(t) \to \varphi_\infty(t)$.

(ii) 若 $\varphi_n(t)$ 逐点收敛于一个极限 $\varphi(t)$，且 φ 在 0 处连续，则相应的分布序列 μ_n 是胎紧的，且弱收敛于具有特征函数 φ 的测度 μ.

注释. 为说明 (ii) 中为何需要极限在 0 处连续，设 μ_n 服从均值为 0、方差为 n 的正态分布. 在此情形下 $\varphi_n(t) = \exp(-nt^2/2) \to 0$（对 $t \neq 0$），且对所有 n 有 $\varphi_n(0) = 1$，但测度不弱收敛，因为 $\mu_n((-\infty, x]) \to 1/2$ 对所有 x 成立.

证明. (i) 容易. e^{itx} 有界且连续，故若 $\mu_n \Rightarrow \mu_\infty$，则定理 3.2.9 蕴含 $\varphi_n(t) \to \varphi_\infty(t)$. 为证 (ii)，第一个目标是证明胎紧性. 从一些计算开始，它们看起来可能有些神秘，但将证明非常有用.

$$\int_{-u}^{u} 1 - e^{itx}\,\mathrm{d}t = 2u - \int_{-u}^{u} (\cos(tx) + i\sin(tx))\,\mathrm{d}t = 2u - \frac{2\sin(ux)}{x}.$$

两边除以 u，对 $\mu_n(\mathrm{d}x)$ 积分，并对左边用 Fubini 定理得

$$u^{-1}\int_{-u}^{u} (1 - \varphi_n(t))\,\mathrm{d}t = 2\int \left(1 - \frac{\sin(ux)}{ux}\right)\mu_n(\mathrm{d}x).$$

为估计右边，注意

$$|\sin x| = \left|\int_0^x \cos(y)\,\mathrm{d}y\right| \le |x| \quad \text{对所有} x,$$

故 $1 - (\sin(ux)/ux) \ge 0$. 舍弃在 $(-2/u, 2/u)$ 上的积分，并在其余部分用 $|\sin(ux)| \le 1$，右边

$$\ge 2\int_{|x|\ge 2/u} \left(1 - \frac{1}{|ux|}\right)\mu_n(\mathrm{d}x) \ge \mu_n(\{x : |x| > 2/u\}).$$

由于当 $t \to 0$ 时 $\varphi(t) \to 1$，

$$u^{-1}\int_{-u}^{u} (1 - \varphi(t))\,\mathrm{d}t \to 0 \quad \text{当} u \to 0.$$

选取 u 使得积分 $< \varepsilon$. 由于对每个 t 有 $\varphi_n(t) \to \varphi(t)$，由有界收敛定理，对 $n \ge N$

$$2\varepsilon \ge u^{-1}\int_{-u}^{u} (1 - \varphi_n(t))\,\mathrm{d}t \ge \mu_n(\{x : |x| > 2/u\}).$$

由于 ε 任意，序列 μ_n 是胎紧的.

为完成证明，注意若 $\mu_{n(k)} \Rightarrow \mu$，则由证明的第一句，$\mu$ 具有特征函数 φ. 最后这一观察与胎紧性蕴含每个子序列都有进一步子序列收敛于 μ. 下面断言这蕴含整个序列收敛于 μ. 为此，注意

前面已证明，若 f 有界且连续，则 $\int f\,\mathrm{d}\mu_n$ 的每个子序列都有进一步子序列收敛于 $\int f\,\mathrm{d}\mu$，故定理 2.3.3 蕴含整个序列收敛于该极限. 这表明对所有有界连续函数 f 有 $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$，故由定理 3.2.9 得所需结果. □

习题

习题 3.3.7. 设 $X_n \Rightarrow X$ 且 X_n 服从均值为 0、方差为 σ_n^2 的正态分布. 证明 $\sigma_n^2 \to \sigma^2 \in [0,\infty)$.

习题 3.3.8. 证明若对 $1 \le n \le \infty$，X_n 与 Y_n 独立，$X_n \Rightarrow X_\infty$ 且 $Y_n \Rightarrow Y_\infty$，则 $X_n + Y_n \Rightarrow X_\infty + Y_\infty$.

习题 3.3.9. 设 $X_1, X_2, \ldots$ 独立且 $S_n = X_1 + \cdots + X_n$. 令 φ_j 为 X_j 的特征函数，假设 $S_n \to S_\infty$ 几乎必然. 则 S_∞ 的特征函数为 $\prod_{j=1}^{\infty} \varphi_j(t)$.

习题 3.3.10. 反复使用恒等式 $\sin t = 2\sin(t/2)\cos(t/2)$ 可得 $(\sin t)/t = \prod_{m=1}^{\infty} \cos(t/2^m)$. 将每一边解释为特征函数来证明最后一个恒等式.

习题 3.3.11. 设 $X_1, X_2, \ldots$ 独立，各以概率 $1/2$ 取值 0 和 1. $X = \sum_{j\ge1} 2X_j/3^j$ 服从 Cantor 分布. 计算 X 的特征函数 φ，并注意 φ 在 $t = 3k\pi$（$k = 0, 1, 2, \ldots$）处取值相同.

3.3.3 矩与导数

在定理 3.3.17 的证明中，推导了不等式

$$\mu\{x : |x| > 2/u\} \le u^{-1} \int_{-u}^{u} (1 - \varphi(t))\,\mathrm{d}t, \tag{3.3.1}$$

它表明特征函数在 0 处的光滑性与测度在 ∞ 处的衰减有关. 将证明留给读者.

定理 3.3.18. 若 $\int |x|^n \mu(\mathrm{d}x) < \infty$，则其特征函数 φ 具有 n 阶连续导数，由下式给出

$$\varphi^{(n)}(t) = \int (ix)^n e^{itx} \mu(\mathrm{d}x).$$

证明. 这通过在积分号下重复求导并利用定理 A.5.1 来证明其合理性而得到. □

定理 3.3.18 中的结果表明，若 $E|X|^n < \infty$，则其特征函数在 0 处 n 次可微，且 $\varphi^{(n)}(0) = E(iX)^n$. 将 φ 在 0 处展开为 Taylor 级数得

$$\varphi(t) = \sum_{m=0}^{n} \frac{E(itX)^m}{m!} + o(t^n),$$

其中 $o(t^n)$ 表示满足 $g(t)/t^n \to 0$（当 $t \to 0$）的量 $g(t)$. 为下面的目的，对误差项有一个好的估计将很重要，故下面推导这一结果. 出发点是一点微积分.

引理 3.3.19.

$$\left|e^{ix}-\sum_{m=0}^{n}\frac{(ix)^m}{m!}\right|\le\min\left\{\frac{|x|^{n+1}}{(n+1)!},\frac{2|x|^n}{n!}\right\}. \tag{3.3.2}$$

右边第一项是修正项通常期望的量级. 第二项对大的 $|x|$ 更好，并将帮助在不假设有限三阶矩的情况下证明中心极限定理.

证明. 分部积分得

$$\int_0^x(x-s)^ne^{is}\,\mathrm{d}s=\frac{x^{n+1}}{n+1}+\frac{i}{n+1}\int_0^x(x-s)^{n+1}e^{is}\,\mathrm{d}s.$$

当 $n=0$ 时，上式表明

$$\int_0^x e^{is}\,\mathrm{d}s=x+i\int_0^x(x-s)e^{is}\,\mathrm{d}s.$$

左边为 $(e^{ix}-1)/i$，故重排得

$$e^{ix}=1+ix+i^2\int_0^x(x-s)e^{is}\,\mathrm{d}s.$$

现利用 $n=1$ 的结果得

$$e^{ix}=1+ix+\frac{i^2x^2}{2}+\frac{i^3}{2}\int_0^x(x-s)^2e^{is}\,\mathrm{d}s,$$

迭代得

$$e^{ix}-\sum_{m=0}^{n}\frac{(ix)^m}{m!}=\frac{i^{n+1}}{n!}\int_0^x(x-s)^ne^{is}\,\mathrm{d}s. \tag{a}$$

为证明结果，只需估计右边的“误差项”. 由于对所有 s 有 $|e^{is}|\le1$，

$$\left|\frac{i^{n+1}}{n!}\int_0^x(x-s)^ne^{is}\,\mathrm{d}s\right|\le|x|^{n+1}/(n+1)!. \tag{b}$$

最后一个估计对小的 x 是好的. 下一个是为大的 x 设计的. 分部积分

$$\frac{i}{n}\int_0^x(x-s)^ne^{is}\,\mathrm{d}s=-\frac{x^n}{n}+\int_0^x(x-s)^{n-1}e^{is}\,\mathrm{d}s.$$

注意到 $x^n/n=\int_0^x(x-s)^{n-1}\,\mathrm{d}s$，现得

$$\frac{i^{n+1}}{n!}\int_0^x(x-s)^ne^{is}\,\mathrm{d}s=\frac{i^n}{(n-1)!}\int_0^x(x-s)^{n-1}(e^{is}-1)\,\mathrm{d}s.$$

由于 $|e^{ix}-1|\le2$，得

$$\left|\frac{i^{n+1}}{n!}\int_0^x(x-s)^ne^{is}\,\mathrm{d}s\right|\le\frac{2}{(n-1)!}\left|\int_0^x(x-s)^{n-1}\,\mathrm{d}s\right|\le2|x|^n/n!. \tag{c}$$

结合 (a)、(b) 和 (c) 即得所需结果. □

取期望，利用 Jensen 不等式，将引理 3.3.19 应用于 $x = tX$，得

$$\begin{aligned}\left|Ee^{itX} - E\sum_{m=0}^{n}\frac{(itX)^m}{m!}\right| &\le E\left|e^{itX} - \sum_{m=0}^{n}\frac{(itX)^m}{m!}\right| \\ &\le E\min\left\{|tX|^{n+1}, 2|tX|^n\right\},\end{aligned}$$

其中在第二步中舍弃了分母以使界限更简单.

在下一节中，下面的特殊情形将有用.

定理 3.3.20. 若 $E|X|^2 < \infty$，则

$$\varphi(t) = 1 + itEX - t^2E(X^2)/2 + o(t^2).$$

证明. 误差项 $\le t^2E(|t|\cdot|X|^3 \wedge 2|X|^2)$. 括号中的变量小于 $2|X|^2$ 且当 $t \to 0$ 时趋于 0，故由控制收敛定理得所需结论. □

注释. 估计式 (3.3.3)（涉及两项的最小值而非只是 naive 地应用 Taylor 级数得到的第一项）的意义在于，在假设 $E|X|^2 < \infty$ 下（即不必假设 $E|X|^3 < \infty$）就得到了定理 3.3.20 的结论.

下一个结果表明，二阶导数的存在性蕴含二阶矩的存在性.

定理 3.3.21. 若 $\limsup_{h\downarrow 0}\{\varphi(h) - 2\varphi(0) + \varphi(-h)\}/h^2 > -\infty$，则 $E|X|^2 < \infty$.

证明. $(e^{ihx} - 2 + e^{-ihx})/h^2 = -2(1-\cos(hx))/h^2 \le 0$ 且 $2(1-\cos(hx))/h^2 \to x^2$（当 $h \to 0$），故 Fatou 引理和 Fubini 定理蕴含

$$\begin{aligned}\int x^2\,\mathrm{d}F(x) &\le 2\liminf_{h\to 0}\int\frac{1-\cos(hx)}{h^2}\,\mathrm{d}F(x) \\ &= -\limsup_{h\to 0}\frac{\varphi(h) - 2\varphi(0) + \varphi(-h)}{h^2} < \infty,\end{aligned}$$

即证所需结果. □

习题

习题 3.3.12. 利用定理 3.3.18 和 $e^{-t^2/2}$ 的级数展开证明标准正态分布具有

$$EX^{2n} = (2n)!/2^n n! = (2n-1)(2n-3)\cdots 3\cdot 1 \equiv (2n-1)!!.$$

习题 3.3.13. (i) 设测度族 $\{\mu_i, i \in I\}$ 是胎紧的，即 $\sup_i \mu_i([-M, M]^c) \to 0$（当 $M \to \infty$）. 利用定理 3.3.1(d) 和 (3.3.3)（取 $n = 0$）证明其特征函数 φ_i 是等度连续的，即若 $\varepsilon > 0$，可选取 $\delta > 0$ 使得当 $|h| < \delta$ 时，$|\varphi_i(t + h) - \varphi_i(t)| < \varepsilon$.

(ii) 设 $\mu_n \Rightarrow \mu_\infty$. 利用定理 3.3.17 和等度连续性推出特征函数 $\varphi_n \to \varphi_\infty$ 在紧集上一致收敛. [直接论证. 不需要用到 AA.]

(iii) 给出一个例子说明收敛不必在整个实直线上是一致的.

习题 3.3.14. 设 $X_1, X_2, \ldots$ 独立同分布，特征函数为 φ.

(i) 若 $\varphi'(0) = ia$ 且 $S_n = X_1 + \cdots + X_n$，则 $S_n/n \to a$ 依概率.

(ii) 若 $S_n/n \to a$ 依概率，则 $\varphi(t/n)^n \to e^{iat}$（当 $n \to \infty$，n 取整数值）.

(iii) 利用 (ii) 和定理 3.3.1(d) 中建立的一致连续性证明 $(\varphi(h) - 1)/h \to -ia$（当 $h \to 0$，h 取正实数值）. 故弱大数定律成立当且仅当 $\varphi'(0)$ 存在. 这一结果归功于 E.J.G. Pitman（1956），John Walsh 指出应该证明 (iii).

最后一个习题与习题 2.2.4 结合表明，当 $E|X| = \infty$ 时 $\varphi'(0)$ 可能存在.

习题 3.3.15. $\int_0^\infty (1 - \operatorname{Re}\varphi(t))/\pi t^2 \, \mathrm{d}t = \int |y| \, \mathrm{d}F(y)$. 提示：在例 3.3.15 的密度函数中作变量替换 $x = |y|t$，该密度的积分为 1.

习题 3.3.16. 证明若 $\lim_{t \downarrow 0}(\varphi(t) - 1)/t^2 = c > -\infty$，则 $EX = 0$ 且 $E|X|^2 = -2c < \infty$. 特别地，若 $\varphi(t) = 1 + o(t^2)$，则 $\varphi(t) \equiv 1$.

习题 3.3.17. 若 Y_n 为具有特征函数 φ_n 的随机变量，则 $Y_n \Rightarrow 0$ 当且仅当存在 $\delta > 0$ 使得对 $|t| \leq \delta$ 有 $\varphi_n(t) \to 1$.

习题 3.3.18. 设 $X_1, X_2, \ldots$ 独立. 若 $S_n = \sum_{m \leq n} X_m$ 依分布收敛，则它依概率收敛（从而由习题 2.5.10 知几乎必然收敛）. 提示：上一习题蕴含若 $m, n \to \infty$ 则 $S_m - S_n \to 0$ 依概率. 现利用习题 2.5.11.

3.3.4 Polya 准则 *

下一个结果对构造特征函数的例子很有用.

定理 3.3.22. Polya 准则. 设 $\varphi(t)$ 为实非负函数，满足 $\varphi(0) = 1$，$\varphi(t) = \varphi(-t)$，且 φ 在 $(0, \infty)$ 上递减凸，满足

$$\lim_{t \downarrow 0} \varphi(t) = 1, \quad \lim_{t \uparrow \infty} \varphi(t) = 0.$$

则存在 $(0, \infty)$ 上的概率测度 ν，使得

$$\varphi(t) = \int_0^\infty \left(1 - \left|\frac{t}{s}\right|\right)^+ \nu(\mathrm{d}s), \tag{$*$}$$

故 φ 是特征函数.

注释. 在陷入证明的细节之前，读者应注意 $(*)$ 将 φ 表示为例 3.3.15 中形式的特征函数的凸组合，故引理 3.3.9 的推广（下面证明）蕴含这是特征函数.

假设 $\lim_{t\to 0}\varphi(t)=1$ 是必要的，因为函数 $\varphi(t)=1_{\{0\}}(t)$（在 0 处为 1，其余为 0）满足所有其他假设. 可以允许 $\lim_{t\to\infty}\varphi(t)=c>0$，办法是在 0 处有一个大小为 c 的点质量，但把这个推广留给读者.

证明. 记 φ' 为 φ 的右导数，即

$$\varphi'(t)=\lim_{h\downarrow 0}\frac{\varphi(t+h)-\varphi(t)}{h}.$$

由于 φ 是凸函数，该极限存在且为右连续、递增的. 故可令 μ 为 $(0,\infty)$ 上的测度，满足对所有 $0\le a<b<\infty$ 有 $\mu(a,b]=\varphi'(b)-\varphi'(a)$，并令 ν 为 $(0,\infty)$ 上的测度，满足 $\mathrm{d}\nu/\mathrm{d}\mu=s$.

现当 $t\to\infty$ 时 $\varphi'(t)\to 0$（因若 $\varphi'(t)\downarrow -\varepsilon$，则对所有 t 有 $\varphi(t)\le 1-\varepsilon t$），故习题 A.4.7 蕴含

$$-\varphi'(s)=\int_s^\infty r^{-1}\nu(\mathrm{d}r).$$

再次积分并利用 Fubini 定理，对 $t\ge 0$ 有

$$\begin{aligned}\varphi(t)&=\int_t^\infty\int_s^\infty r^{-1}\nu(\mathrm{d}r)\,\mathrm{d}s=\int_t^\infty r^{-1}\int_t^r \mathrm{d}s\,\nu(\mathrm{d}r)\\&=\int_t^\infty\left(1-\frac{t}{r}\right)\nu(\mathrm{d}r)=\int_0^\infty\left(1-\frac{t}{r}\right)^+\nu(\mathrm{d}r).\end{aligned}$$

利用 $\varphi(-t)=\varphi(t)$ 将公式延拓到 $t\le 0$，得到 $(*)$. 在 $(*)$ 中令 $t=0$ 表明 ν 的总质量为 1.

若 φ 是分段线性的，ν 具有有限个原子，结果由例 3.3.15 和引理 3.3.9 推出. 为证明一般结果，设 ν_n 为 $(0,\infty)$ 上具有有限个原子的测度序列，弱收敛于 ν（见习题 3.2.10），并令

$$\varphi_n(t)=\int_0^\infty\left(1-\left|\frac{t}{s}\right|\right)^+\nu_n(\mathrm{d}s).$$

由于 $s\mapsto(1-|t/s|)^+$ 有界连续，$\varphi_n(t)\to\varphi(t)$，故由定理 3.3.17(ii) 得所需结果. □

Polya 准则的一个经典应用是：

习题 3.3.19. 证明 $\exp(-|t|^\alpha)$ 是 $0<\alpha\le 1$ 的特征函数.

（$\alpha=1$ 的情形对应于 Cauchy 分布.）下一个论证（从 Frank Spitzer 处学到）证明这对 $0<\alpha\le 2$ 成立. $\alpha=2$ 的情形对应于正态分布，故该情形在证明中可安全忽略.

例题 3.3.23. $\exp(-|t|^\alpha)$ 是 $0<\alpha<2$ 的特征函数.

证明. 一点微积分表明，对任意 β 和 $|x|<1$

$$(1-x)^{\beta}=\sum_{n=0}^{\infty}\binom{\beta}{n}(-x)^n,$$

其中

$$\binom{\beta}{n}=\frac{\beta(\beta-1)\cdots(\beta-n+1)}{1\cdot 2\cdots n}.$$

令 $\psi(t)=1-(1-\cos t)^{\alpha/2}=\sum_{n=1}^{\infty}c_n(\cos t)^n$，其中

$$c_n=(-1)^{n+1}\binom{\alpha/2}{n},$$

$c_n\geq 0$（这里用到了 $\alpha<2$），且 $\sum_{n=1}^{\infty}c_n=1$（在 ψ 的定义中取 $t=0$）. $\cos t$ 是特征函数（见例 3.3.3），故引理 3.3.9 的简易推广表明 ψ 是特征函数. 有 $1-\cos t\sim t^2/2$（当 $t\to 0$），故

$$1-\cos(t\cdot 2^{1/2}\cdot n^{-1/\alpha})\sim n^{-2/\alpha}t^2.$$

现利用引理 3.1.1 和定理 3.3.17(ii) 得

$$\exp(-|t|^{\alpha})=\lim_{n\to\infty}\{\psi(t\cdot 2^{1/2}\cdot n^{-1/\alpha})\}^n$$

是特征函数. □

习题 3.3.16 表明当 $\alpha>2$ 时 $\exp(-|t|^{\alpha})$ 不是特征函数. 对这些特征函数感兴趣的一个原因由下面习题 3.3.15 的推广解释.

习题 3.3.20. 若 $X_1,X_2,\ldots$ 独立且具有特征函数 $\exp(-|t|^{\alpha})$，则 $(X_1+\cdots+X_n)/n^{1/\alpha}$ 与 X_1 同分布.

将在 3.8 节回到这个话题.

习题 3.3.21. 设 φ_1 和 φ_2 为特征函数. 证明 $A=\{t:\varphi_1(t)=\varphi_2(t)\}$ 是闭集，包含 0，且关于 0 对称. 证明若 A 是具有这些性质的集合且 $\varphi_1(t)=e^{-|t|}$，则存在 φ_2 使得 $\{t:\varphi_1(t)=\varphi_2(t)\}=A$.

例题 3.3.24. 对某些目的，有两个特征函数在 $[-1,1]$ 上一致的显式例子是很好的. 由例 3.3.15，知道 $(1-|t|)^+$ 是密度 $(1-\cos x)/\pi x^2$ 的特征函数. 定义 $\psi(t)$ 在 $[-1,1]$ 上等于 φ 且周期为 2，即 $\psi(t)=\psi(t+2)$. ψ 的 Fourier 级数为

$$\psi(u)=\frac{1}{2}+\sum_{n=-\infty}^{\infty}\frac{2}{\pi^2(2n-1)^2}\exp(i(2n-1)\pi u).$$

右边是具有

$$P(X=0)=1/2,\quad P(X=(2n-1)\pi)=2\pi^{-2}(2n-1)^{-2}\quad (n\in\mathbb{Z})$$

的离散分布的特征函数.

习题 3.3.22. 找独立的随机变量 X,Y 和 Z 使得 Y 与 Z 不同分布但 $X+Y$ 与 $X+Z$ 同分布.

习题 3.3.23. 证明若 X 与 Y 独立且 $X+Y$ 与 X 同分布，则 $Y=0$ 几乎必然.

更多奇闻异事，参见 Feller, Vol. II (1971), Section XV.2a.

3.3.5 矩问题 *

设 $\int x^k\,\mathrm{d}F_n(x)$ 对每个 k 有极限 μ_k. 则由定理 3.2.14，分布序列是胎紧的，且每个子序列极限具有矩 μ_k（由习题 3.2.5），故若具有这些矩的分布唯一，就能得出序列弱收敛的结论. 若 F 集中在有限区间 $[-M,M]$ 上，这容易看出，因为每个连续函数都能在 $[-M,M]$ 上用多项式一致逼近. 一般而言，结论不成立.

反例 1. Heyde (1963) 考虑对数正态密度

$$f_0(x)=(2\pi)^{-1/2}x^{-1}\exp(-(\log x)^2/2)\quad x\geq 0$$

且对 $-1\leq a\leq 1$ 令

$$f_a(x)=f_0(x)\{1+a\sin(2\pi\log x)\}.$$

为看出 f_a 是密度且具有与 f_0 相同的矩，只需证明

$$\int_0^\infty x^r f_0(x)\sin(2\pi\log x)\,\mathrm{d}x=0\quad \text{对}r=0,1,2,\ldots$$

作变量替换 $x=\exp(s+r)$，$s=\log x-r$，$\mathrm{d}s=\mathrm{d}x/x$，积分变为

$$\begin{aligned}&\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp(rs+r^2)\exp(-(s+r)^2/2)\sin(2\pi(s+r))\,\mathrm{d}s\\&=\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp(r^2/2)\exp(-s^2/2)\sin(2\pi s)\,\mathrm{d}s=0.\end{aligned}$$

两个等式成立因为 r 是整数且被积函数是奇函数. 从证明中应可清楚看出，可以令

$$g(x)=f_0(x)\left\{1+\sum_{k=1}^{\infty}a_k\sin(k\pi\log x)\right\}\quad \text{若}|a_k|\leq 1$$

来得到一族具有与对数正态相同矩的密度.

对数正态的矩容易计算. 回忆若 χ 服从标准正态分布，则习题 1.2.6 蕴含 $\exp(\chi)$ 服从对数正态分布.

$$\begin{aligned} EX^n = E\exp(n\chi) &= \int e^{nx}(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x \\ &= e^{n^2/2}\int (2\pi)^{-1/2}e^{-(x-n)^2/2}\,\mathrm{d}x = \exp(n^2/2), \end{aligned}$$

因为最后的被积函数是均值为 n、方差为 1 的正态密度.

相当引人注目的是，有一族离散随机变量具有这些矩. 令 $a>0$ 且

$$P(Y_a = ae^k) = a^{-k}\exp(-k^2/2)/c_a \quad \text{对} k\in\mathbb{Z},$$

其中 c_a 选取使得总质量为 1.

$$\begin{aligned} \exp(-n^2/2)EY_a^n &= \exp(-n^2/2)\sum_k (ae^k)^n a^{-k}\exp(-k^2/2)/c_a \\ &= \sum_k a^{-(k-n)}\exp(-(k-n)^2/2)/c_a = 1, \end{aligned}$$

由 c_a 的定义.

对数正态密度在 $|x|\to\infty$ 时以 $\exp(-(\log x)^2/2)$ 衰减. 下一个反例有更快速的衰减. 由于指数分布 e^{-x}（$x\ge 0$）由其矩确定（见下面习题 3.3.25），不能希望做得更好.

反例 2. 令 $\lambda\in(0,1)$ 且对 $-1\le a\le 1$ 令

$$f_{a,\lambda}(x) = c_\lambda\exp(-|x|^\lambda)\{1+a\sin(\beta|x|^\lambda\operatorname{sgn}(x))\},$$

其中 $\beta=\tan(\lambda\pi/2)$ 且 $1/c_\lambda=\int\exp(-|x|^\lambda)\,\mathrm{d}x$. 为证明这些是密度函数且对固定的 λ 它们具有相同的矩，只需证明

$$\int x^n\exp(-|x|^\lambda)\sin(\beta|x|^\lambda\operatorname{sgn}(x))\,\mathrm{d}x = 0 \quad \text{对} n=0,1,2,\ldots$$

这对偶数 n 显然，因为被积函数是奇函数. 为奇数 n 证明结果，只需在 $[0,\infty)$ 上积分. 利用恒等式

$$\int_0^\infty t^{p-1}e^{-qt}\,\mathrm{d}t = \Gamma(p)/q^p \quad \text{当}\operatorname{Re}q>0$$

取 $p=(n+1)/\lambda$，$q=1+\beta i$，并作变量替换 $t=x^\lambda$，得

$$\begin{aligned}&\Gamma((n+1)/\lambda)/(1+\beta i)^{(n+1)/\lambda}\\&=\int_0^\infty x^{\lambda\{(n+1)/\lambda-1\}}\exp(-(1+\beta i)x^\lambda)\lambda x^{\lambda-1}\,\mathrm{d}x\\&=\lambda\int_0^\infty x^n\exp(-x^\lambda)\cos(\beta x^\lambda)\,\mathrm{d}x-i\lambda\int_0^\infty x^n\exp(-x^\lambda)\sin(\beta x^\lambda)\,\mathrm{d}x.\end{aligned}$$

由于 $\beta=\tan(\lambda\pi/2)$

$$(1+\beta i)^{(n+1)/\lambda}=(\cos(\lambda\pi/2))^{-(n+1)/\lambda}(\exp(i\lambda\pi/2))^{(n+1)/\lambda}.$$

右边是实数，因为 $\lambda<1$ 且 $(n+1)$ 是偶数，故

$$\int_0^\infty x^n\exp(-x^\lambda)\sin(\beta x^\lambda)\,\mathrm{d}x=0.$$

一个分布由其矩确定的实用充分条件是

定理 3.3.25. 若 $\limsup_{k\to\infty}\mu_{2k}^{1/2k}/2k=r<\infty$，则至多存在一个分布函数 F 满足 $\mu_k=\int x^k\,\mathrm{d}F(x)$ 对所有正整数 k 成立.

注释. 这略强于 Carleman 条件

$$\sum_{k=1}^\infty 1/\mu_{2k}^{1/2k}=\infty,$$

该条件对定理 3.3.25 的结论也是充分的.

证明. 令 F 为任意具有矩 μ_k 的分布函数，且 $\nu_k=\int|x|^k\,\mathrm{d}F(x)$. Cauchy-Schwarz 不等式蕴含 $\nu_{2k+1}^2\le\mu_{2k}\mu_{2k+2}$，故

$$\limsup_{k\to\infty}(\nu_k^{1/k})/k=r<\infty.$$

在引理 3.3.19 中取 $x=tX$ 并乘以 $e^{i\theta X}$，有

$$\left|e^{i\theta X}\left(e^{itX}-\sum_{m=0}^{n-1}\frac{(itX)^m}{m!}\right)\right|\le\frac{|tX|^n}{n!}.$$

取期望并利用习题 3.3.18 得

$$\left|\varphi(\theta+t)-\varphi(\theta)-t\varphi'(\theta)-\cdots-\frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(\theta)\right|\le\frac{|t|^n}{n!}\nu_n.$$

利用最后结果、$\nu_k\le(r+\varepsilon)^k k^k$（对大的 k）的事实，以及平凡界限 $e^k\ge k^k/k!$（将左边展开为幂级

数)，可见对任意 θ

$$\varphi(\theta+t)=\varphi(\theta)+\sum_{m=1}^{\infty}\frac{t^m}{m!}\varphi^{(m)}(\theta)\quad \text{对}|t|<1/er. \tag{$*$}$$

令 G 为另一具有给定矩的分布，ψ 为其特征函数. 由于 $\varphi(0)=\psi(0)=1$，由 $(*)$ 和归纳法得 $\varphi(t)=\psi(t)$ 对 $|t|\le k/3r$（对所有 k）成立，故两个特征函数重合，分布相等. □

将定理 3.3.25 与开始考虑矩问题时的讨论结合：

定理 3.3.26. 设 $\int x^k\,\mathrm{d}F_n(x)$ 对每个 k 有极限 μ_k 且

$$\limsup_{k\to\infty}\mu_{2k}^{1/2k}/2k<\infty,$$

则 F_n 弱收敛于具有这些矩的唯一分布.

习题 3.3.24. 令 $G(x)=P(|X|<x)$，$\lambda=\sup\{x:G(x)<1\}$，$\nu_k=E|X|^k$. 证明 $\nu_k^{1/k}\to\lambda$，故若 $\lambda<\infty$，定理 3.3.26 的假设成立.

习题 3.3.25. 设 $|X|$ 在 $(0,\infty)$ 上具有密度 $Cx^\alpha\exp(-x^\lambda)$. 作变量替换 $y=x^\lambda$，$\mathrm{d}x=(1/\lambda)x^{1/\lambda-1}\mathrm{d}x$

$$E|X|^n=C\lambda^{-1}\int_0^\infty y^{(n+\alpha)/\lambda}\exp(-y)y^{1/\lambda-1}\,\mathrm{d}y=C\lambda^{-1}\Gamma((n+\alpha+1)/\lambda).$$

利用恒等式 $\Gamma(x+1)=x\Gamma(x)$（$x\ge0$）推出定理 3.3.26 的假设对 $\lambda\ge1$ 满足但对 $\lambda<1$ 不满足. 这表明正态（$\lambda=2$）和伽马（$\lambda=1$）分布由其矩确定.

结果到目前为止是针对所谓的 **Hamburger 矩问题**. 若事先假设分布集中在 $[0,\infty)$ 上，就有了 **Stieltjes 矩问题**. $X\ge0$ 与 $\mathbb{R}$ 上的对称分布之间存在一一对应，由 $X\mapsto\xi\sqrt{X}$ 给出，其中 $\xi\in\{-1,1\}$ 独立于 X 且以相等概率取两个值. 由此看到

$$\limsup_{k\to\infty}\nu_k^{1/2k}/2k<\infty$$

足以保证存在唯一的 $[0,\infty)$ 上的分布具有给定矩. 下一个例子表明，对非负随机变量，最后一个结果接近最佳可能.

反例 3. 令 $\lambda\in(0,1/2)$，$\beta=\tan(\lambda\pi)$，$-1\le a\le1$ 且

$$f_a(x)=c_\lambda\exp(-x^\lambda)(1+a\sin(\beta x^\lambda))\quad \text{对}x\ge0,$$

其中 $1/c_\lambda=\int_0^\infty\exp(-x^\lambda)\,\mathrm{d}x$.

通过模仿反例 2 中的计算，容易看出 f_a 是具有相同矩的概率密度. 这个例子似乎归功于 Stoyanov (1987) p. 92–93. 特殊情形 $\lambda=1/4$ 广为人知.

第 4 节 中心极限定理

下面讨论本章的主要内容. 首先证明独立同分布序列的**中心极限定理**.

3.4.1 独立同分布序列

定理 3.4.1. 设 $X_1, X_2, \ldots$ 为独立同分布随机变量，$EX_i = \mu$，$\mathrm{var}(X_i) = \sigma^2 \in (0, \infty)$. 若 $S_n = X_1 + \cdots + X_n$，则

$$(S_n - n\mu)/\sigma n^{1/2} \Rightarrow \chi$$

其中 χ 服从**标准正态分布**.

这一记号虽非标准，但颇为方便. 为理解该记号的逻辑，注意正态分布的平方服从卡方分布.

证明. 考虑 $X_i' = X_i - \mu$，只需证明 $\mu = 0$ 时的结果. 由定理 3.3.20

$$\phi(t) = E\exp(itX_1) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$$

于是

$$E\exp(itS_n/\sigma n^{1/2}) = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n$$

由引理 3.1.1，当 $n \to \infty$ 时最后一个量趋于 $\exp(-t^2/2)$，结合定理 3.3.17 即完成证明. 然而，引理 3.1.1 是关于实数的事实，故需将其推广到复数情形才能完成证明. □

定理 3.4.2. 若 $c_n \to c \in \mathbb{C}$，则 $(1 + c_n/n)^n \to e^c$.

证明. 证明基于两个简单的事实：

引理 3.4.3. 设 $z_1, \ldots, z_n$ 和 $w_1, \ldots, w_n$ 为模不超过 θ 的复数. 则

$$\left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m\right| \leq \theta^{n-1} \sum_{m=1}^n |z_m - w_m|$$

证明. $n = 1$ 时结论显然成立. 对 $n > 1$，注意到

$$\begin{aligned}\left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m\right| &\leq \left|z_1 \prod_{m=2}^n z_m - z_1 \prod_{m=2}^n w_m\right| + \left|z_1 \prod_{m=2}^n w_m - w_1 \prod_{m=2}^n w_m\right| \\ &\leq \theta \left|\prod_{m=2}^n z_m - \prod_{m=2}^n w_m\right| + \theta^{n-1}|z_1 - w_1|\end{aligned}$$

再用归纳法即得结论. □

引理 3.4.4. 若 b 为满足 $|b| \le 1$ 的复数，则 $|e^b - (1+b)| \le |b|^2$.

证明. $e^b - (1+b) = b^2/2! + b^3/3! + b^4/4! + \cdots$，故当 $|b| \le 1$ 时

$$|e^b - (1+b)| \le \frac{|b|^2}{2}(1 + 1/2 + 1/2^2 + \cdots) = |b|^2 \qquad \square$$

定理 3.4.2 的证明. 令 $z_m = (1 + c_n/n)$，$w_m = \exp(c_n/n)$，$\gamma > |c|$. 对充分大的 n，$|c_n| < \gamma$. 由于 $1 + \gamma/n \le \exp(\gamma/n)$，由引理 3.4.3 和 3.4.4 得

$$|(1 + c_n/n)^n - e^{c_n}| \le (e^{\gamma/n})^{n-1} n \left|\frac{c_n}{n}\right|^2 \le e^{\gamma}\frac{\gamma^2}{n} \to 0$$

当 $n \to \infty$. 证毕. $\square$

为理解中心极限定理的含义，下面给出几个具体例子.

例题 3.4.5. 轮盘赌. 轮盘有编号 1 至 36 的槽（18 个红色，18 个黑色）以及两个编号 0 和 00 的绿色槽. 玩家可以下注 \$1 赌球落在红色（或黑色）槽中，若赌对则赢得 \$1. 设 X_i 为第 i 次下注的赢利，则 $X_1, X_2, \ldots$ 独立同分布，满足 $P(X_i = 1) = 18/38$，$P(X_i = -1) = 20/38$.

$$EX_i = -1/19, \quad \text{var}(X_i) = EX_i^2 - (EX_i)^2 = 1 - (1/19)^2 = 0.9972$$

所关心的是

$$P(S_n \ge 0) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \ge \frac{-n\mu}{\sigma\sqrt{n}}\right)$$

取 $n = 361 = 19^2$，为简化计算将 σ 取为 1，

$$\frac{-n\mu}{\sigma\sqrt{n}} = \frac{361 \cdot (1/19)}{\sqrt{361}} = 1$$

于是中心极限定理及书后正态分布表给出

$$P(S_n \ge 0) \approx P(\chi \ge 1) = 1 - 0.8413 = 0.1587$$

换言之，转动 361 次轮盘后，赌场平均会赢你 \$19，但约有 0.16 的概率你会领先.

例题 3.4.6. 抛硬币. 设 $X_1, X_2, \ldots$ 独立同分布，$P(X_i = 0) = P(X_i = 1) = 1/2$. 若 $X_i = 1$ 表示第 i 次抛出正面，则 $S_n = X_1 + \cdots + X_n$ 为时刻 n 时正面朝上的总次数.

$$EX_i = 1/2, \quad \text{var}(X_i) = EX_i^2 - (EX_i)^2 = 1/2 - 1/4 = 1/4$$

于是中心极限定理给出 $(S_n - n/2)/\sqrt{n/4} \Rightarrow \chi$. 正态分布表给出

$$P(\chi > 2) = 1 - 0.9773 = 0.0227$$

所以 $P(|\chi| \leq 2) = 1 - 2(0.0227) = 0.9546$，代入中心极限定理得

$$0.95 \approx P\left(\frac{S_n - n/2}{\sqrt{n/4}} \in [-2, 2]\right) = P(S_n - n/2 \in [-\sqrt{n}, \sqrt{n}])$$

取 $n = 10000$，这意味着 95% 的时间里正面次数会在 4900 至 5100 之间.

例题 3.4.7. 二项分布的正态逼近. 设 $X_1, X_2, \ldots$ 和 S_n 如上例. 为利用中心极限定理估计 $P(S_{16} = 8)$，将 8 视为区间 $[7.5, 8.5]$. 由于 $\mu = 1/2$，且 $n = 16$ 时 $\sigma\sqrt{n} = 2$

$$\begin{aligned} P(|S_{16} - 8| \leq 0.5) &= P\left(\frac{|S_{16} - n\mu|}{\sigma\sqrt{n}} \leq 0.25\right) \\ &\approx P(|\chi| \leq 0.25) = 2(0.5987 - 0.5) = 0.1974 \end{aligned}$$

尽管 n 很小，这与精确概率

$$2^{-16}\binom{16}{8} = \frac{13 \cdot 11 \cdot 10 \cdot 9}{65536} = 0.1964$$

吻合得很好.

上述计算启发了**直方图修正**，这对用小 n 进行正态逼近很重要. 例如，若要逼近 $P(S_{16} \leq 11)$，则将其视为 $P(S_{16} \leq 11.5)$. 这样做的一个明显原因是，若视 $P(S_{16} \leq 11) = 1 - P(S_{16} \geq 12)$，应得到相同答案.

例题 3.4.8. Poisson 分布的正态逼近. 设 Z_λ 服从均值为 λ 的 Poisson 分布. 若 $X_1, X_2, \ldots$ 独立且均服从均值为 1 的 Poisson 分布，则 $S_n = X_1 + \cdots + X_n$ 服从均值为 n 的 Poisson 分布. 由于 $\text{var}(X_i) = 1$，中心极限定理蕴含

$$(S_n - n)/n^{1/2} \Rightarrow \chi \text{ 当} n \to \infty$$

对非整数 λ，令 N_1, N_2, N_3 独立且分别服从均值为 $[\lambda]$，$\lambda - [\lambda]$ 和 $[\lambda] + 1 - \lambda$ 的 Poisson 分布. 令 $S_{[\lambda]} = N_1$，$Z_\lambda = N_1 + N_2$，$S_{[\lambda]+1} = N_1 + N_2 + N_3$，则 $S_{[\lambda]} \leq Z_\lambda \leq S_{[\lambda]+1}$，利用 S_n 的极限定理可得

$$(Z_\lambda - \lambda)/\lambda^{1/2} \Rightarrow \chi \text{ 当} \lambda \to \infty$$

例题 3.4.9. 两两独立不足以保证中心极限定理. 两两独立对**强大数定律**已经足够（见定理 2.4.1），但对中心极限定理却不够. 设 $\xi_1, \xi_2, \ldots$ 独立同分布，$P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. 构造如下：对

$n \geq 1$

$$S_{2^n} = \xi_1(1+\xi_2)\cdots(1+\xi_{n+1}) = \begin{cases} \pm 2^n & \text{概率为} 2^{-n-1} \\ 0 & \text{概率为} 1-2^{-n} \end{cases}$$

为此令 $X_1 = \xi_1$，$X_2 = \xi_1\xi_2$，且对 $m = 2^{n-1}+j$，$0 < j \leq 2^{n-1}$，$n \geq 2$，令 $X_m = X_j\xi_{n+1}$. 每个 X_m 是不同 ξ_j 集合的乘积，故它们两两独立.

习题

习题 3.4.1. 设你掷骰子 180 次. 用正态逼近（带直方图修正）估计得到少于 25 个 6 的概率.

习题 3.4.2. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$，$0 < \text{var}(X_i) < \infty$，令 $S_n = X_1 + \cdots + X_n$.

(a) 用中心极限定理和 Kolmogorov 0-1 律证明 $\limsup S_n/\sqrt{n} = \infty$ a.s.

(b) 用反证法证明 $S_n/\sqrt{n}$ 不依概率收敛. 提示：考虑 $n = m!$.

习题 3.4.3. 设 $X_1, X_2, \ldots$ 独立同分布，$S_n = X_1 + \cdots + X_n$. 假设 $S_n/\sqrt{n} \Rightarrow$ 某极限，推出 $EX_i^2 < \infty$. 概要：设 $X_1', X_2', \ldots$ 为原序列的独立复制. 令 $Y_i = X_i - X_i'$，$U_i = Y_i 1_{(|Y_i| \leq A)}$，$V_i = Y_i 1_{(|Y_i| > A)}$，并注意对任意 K

$$P\left(\sum_{m=1}^n Y_m \geq K\sqrt{n}\right) \geq P\left(\sum_{m=1}^n U_m \geq K\sqrt{n}, \sum_{m=1}^n V_m \geq 0\right) \geq \frac{1}{2}P\left(\sum_{m=1}^n U_m \geq K\sqrt{n}\right) \geq \frac{1}{5}$$

对充分大的 n，若 A 充分大. 由于 K 任意，这产生矛盾.

习题 3.4.4. 设 $X_1, X_2, \ldots$ 独立同分布，$X_i \geq 0$，$EX_i = 1$，$\text{var}(X_i) = \sigma^2 \in (0, \infty)$. 证明 $2(\sqrt{S_n} - \sqrt{n}) \Rightarrow \sigma\chi$.

习题 3.4.5. 自归一化和. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$，$EX_i^2 = \sigma^2 \in (0, \infty)$. 则

$$\frac{\sum_{m=1}^n X_m}{(\sum_{m=1}^n X_m^2)^{1/2}} \Rightarrow \chi$$

习题 3.4.6. 随机指标中心极限定理. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$，$EX_i^2 = \sigma^2 \in (0, \infty)$，$S_n = X_1 + \cdots + X_n$. 设 N_n 为一列非负整数值随机变量，a_n 为一列整数满足 $a_n \to \infty$ 且 $N_n/a_n \to 1$ 依概率. 证明

$$S_{N_n}/\sigma\sqrt{a_n} \Rightarrow \chi$$

提示：用 Kolmogorov 不等式（定理 2.5.5）证明若 $Y_n = S_{N_n}/\sigma\sqrt{a_n}$，$Z_n = S_{a_n}/\sigma\sqrt{a_n}$，则 $Y_n - Z_n \to 0$ 依概率.

习题 3.4.7. 更新论中的中心极限定理. 设 $Y_1, Y_2, \ldots$ 为独立同分布正随机变量，$EY_i = \mu$，$\text{var}(Y_i) = \sigma^2 \in (0, \infty)$. 令 $S_n = Y_1 + \cdots + Y_n$，$N_t = \sup\{m : S_m \le t\}$. 将前一习题应用于 $X_i = Y_i - \mu$，证明当 $t \to \infty$ 时

$$(\mu N_t - t)/(\sigma^2 t/\mu)^{1/2} \Rightarrow \chi$$

习题 3.4.8. 更新中心极限定理的另一证明. 设 $Y_1, Y_2, \ldots$，S_n 和 N_t 如上题. 令 $u = [t/\mu]$，$D_t = S_u - t$. 用 Kolmogorov 不等式证明

$$P(|S_{u+m} - (S_u + m\mu)| > t^{2/5} \text{ 对某} m \in [-t^{3/5}, t^{3/5}]) \to 0 \text{ 当} t \to \infty$$

推出 $|N_t - (t - D_t)/\mu|/t^{1/2} \to 0$ 依概率，然后得到前一习题的结果.

3.4.2 三角阵列

下一步将中心极限定理推广到：

定理 3.4.10. Lindeberg-Feller 定理. 对每个 n，设 $X_{n,m}$ $(1 \le m \le n)$ 为独立随机变量，$EX_{n,m} = 0$. 假设

(i) $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$

(ii) 对所有 $\varepsilon > 0$，$\lim_{n\to\infty} \sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) = 0$

则 $S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow \sigma\chi$ 当 $n \to \infty$.

注释. 用语言表述，该定理说明大量小的独立效应之和近似服从正态分布. 为说明定理 3.4.10 蕴含第一个中心极限定理，设 $Y_1, Y_2, \ldots$ 独立同分布，$EY_i = 0$，$EY_i^2 = \sigma^2 \in (0, \infty)$，令 $X_{n,m} = Y_m/n^{1/2}$. 则 $\sum_{m=1}^{n} EX_{n,m}^2 = \sigma^2$，且若 $\varepsilon > 0$

$$\begin{aligned}\sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) &= nE(|Y_1/n^{1/2}|^2; |Y_1/n^{1/2}| > \varepsilon) \\ &= E(|Y_1|^2; |Y_1| > \varepsilon n^{1/2}) \to 0\end{aligned}$$

由控制收敛定理，因 $EY_1^2 < \infty$.

证明. 令 $\phi_{n,m}(t) = E\exp(itX_{n,m})$，$\sigma_{n,m}^2 = EX_{n,m}^2$. 由定理 3.3.17，只需证明

$$\prod_{m=1}^{n} \phi_{n,m}(t) \to \exp(-t^2\sigma^2/2)$$

令 $z_{n,m}=\phi_{n,m}(t)$，$w_{n,m}=(1-t^2\sigma_{n,m}^2/2)$. 由 (3.3.3)

$$\begin{aligned}|z_{n,m}-w_{n,m}| &\le E(|tX_{n,m}|^3\wedge 2|tX_{n,m}|^2)\\ &\le E(|tX_{n,m}|^3;|X_{n,m}|\le\varepsilon)+E(2|tX_{n,m}|^2;|X_{n,m}|>\varepsilon)\\ &\le \varepsilon|t|^3E(|X_{n,m}|^2;|X_{n,m}|\le\varepsilon)+2t^2E(|X_{n,m}|^2;|X_{n,m}|>\varepsilon)\end{aligned}$$

对 $m=1$ 到 n 求和，令 $n\to\infty$，并利用 (i) 和 (ii) 得

$$\limsup_{n\to\infty}\sum_{m=1}^{n}|z_{n,m}-w_{n,m}|\le\varepsilon|t|^3\sigma^2$$

由于 $\varepsilon>0$ 任意，该序列趋于 0. 下一步是用引理 3.4.3（取 $\theta=1$）得

$$\left|\prod_{m=1}^{n}\phi_{n,m}(t)-\prod_{m=1}^{n}(1-t^2\sigma_{n,m}^2/2)\right|\to 0$$

为验证引理 3.4.3 的条件，注意因 $\phi_{n,m}$ 是特征函数，$|\phi_{n,m}(t)|\le 1$ 对所有 n,m 成立. 对第二个乘积中的项，注意

$$\sigma_{n,m}^2\le\varepsilon^2+E(|X_{n,m}|^2;|X_{n,m}|>\varepsilon)$$

ε 任意，故 (ii) 蕴含 $\sup_m\sigma_{n,m}^2\to 0$，于是当 n 充分大时，对所有 m 有 $1\ge 1-t^2\sigma_{n,m}^2/2>-1$.

为完成证明，对 $c_{m,n}=-t^2\sigma_{n,m}^2/2$ 应用习题 3.1.1. 上面刚证明 $\sup_m\sigma_{n,m}^2\to 0$，(i) 蕴含

$$\sum_{m=1}^{n}c_{m,n}\to-\sigma^2t^2/2$$

于是 $\prod_{m=1}^{n}(1-t^2\sigma_{n,m}^2/2)\to\exp(-t^2\sigma^2/2)$，证毕. □

例题 3.4.11. 随机排列中的循环与记录值. 继续例 2.2.8 和 2.3.10 的分析，设 $Y_1,Y_2,\ldots$ 独立，$P(Y_m=1)=1/m$，$P(Y_m=0)=1-1/m$. $EY_m=1/m$，$\mathrm{var}(Y_m)=1/m-1/m^2$. 于是若 $S_n=Y_1+\cdots+Y_n$，则 $ES_n\sim\log n$，$\mathrm{var}(S_n)\sim\log n$. 令

$$X_{n,m}=(Y_m-1/m)/(\log n)^{1/2}$$

$EX_{n,m}=0$，$\sum_{m=1}^{n}EX_{n,m}^2\to 1$，且对任意 $\varepsilon>0$

$$\sum_{m=1}^{n}E(|X_{n,m}|^2;|X_{n,m}|>\varepsilon)\to 0$$

因当 $(\log n)^{-1/2} < \varepsilon$ 时该和为 0. 应用定理 3.4.10 得

$$(\log n)^{-1/2}\left(S_n - \sum_{m=1}^{n}\frac{1}{m}\right) \Rightarrow \chi$$

注意

$$\int_1^n x^{-1}\mathrm{d}x = \log n \geq \sum_{m=2}^{n}\frac{1}{m} \geq \int_2^{n+1} x^{-1}\mathrm{d}x = \log(n+1) - \log 2$$

表明 $|\log n - \sum_{m=1}^{n} 1/m| \leq 1$，结论可写为

$$(S_n - \log n)/(\log n)^{1/2} \Rightarrow \chi$$

例题 3.4.12. 三级数定理的逆. 回忆定理 2.5.8 的设定. 设 $X_1, X_2, \ldots$ 独立，$A > 0$，$Y_m = X_m 1_{(|X_m| \leq A)}$. 为使 $\sum_{n=1}^{\infty} X_n$ 收敛（即 $\lim_{N\to\infty}\sum_{n=1}^{N} X_n$ 存在），必须：

$$\text{(i)}\sum_{n=1}^{\infty} P(|X_n| > A) < \infty, \quad \text{(ii)}\sum_{n=1}^{\infty} EY_n \text{ 收敛}, \quad \text{(iii)}\sum_{n=1}^{\infty}\mathrm{var}(Y_n) < \infty$$

证明. 第一个条件的必要性是显然的，因若该级数发散，$P(|X_n| > A \text{ i.o.}) > 0$，$\lim_{n\to\infty}\sum_{m=1}^{n} X_m$ 不可能存在. 假设 (i) 中级数有限但 (iii) 中级数无限. 令

$$c_n = \sum_{m=1}^{n}\mathrm{var}(Y_m), \quad X_{n,m} = (Y_m - EY_m)/c_n^{1/2}$$

$EX_{n,m} = 0$，$\sum_{m=1}^{n} EX_{n,m}^2 = 1$，且对任意 $\varepsilon > 0$

$$\sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) \to 0$$

因当 $2A/c_n^{1/2} < \varepsilon$ 时该和为 0. 应用定理 3.4.10，若 $S_n = X_{n,1} + \cdots + X_{n,n}$，则 $S_n \Rightarrow \chi$. 现

(i) 若 $\lim_{n\to\infty}\sum_{m=1}^{n} X_m$ 存在，$\lim_{n\to\infty}\sum_{m=1}^{n} Y_m$ 存在.

(ii) 若令 $T_n = (\sum_{m\leq n} Y_m)/c_n^{1/2}$，则 $T_n \Rightarrow 0$.

最后两个结果与习题 3.2.13 蕴含 $(S_n - T_n) \Rightarrow \chi$. 由于

$$S_n - T_n = -\sum_{m\leq n} EY_m/c_n^{1/2}$$

不是随机的，这很荒谬.

最后，假设 (i) 和 (iii) 中级数有限. 定理 2.5.6 蕴含 $\lim_{n\to\infty}\sum_{m=1}^{n}(Y_m - EY_m)$ 存在，故若

$\lim_{n\to\infty}\sum_{m=1}^{n}X_m$ 从而 $\lim_{n\to\infty}\sum_{m=1}^{n}Y_m$ 存在，取差表明 (ii) 成立. □

例题 3.4.13. 无限方差. 设 $X_1, X_2, \ldots$ 独立同分布，满足 $P(X_1 > x) = P(X_1 < -x)$ 且 $P(|X_1| > x) = x^{-2}$ $(x \geq 1)$.

$$E|X_1|^2 = \int_0^\infty 2xP(|X_1| > x)\mathrm{d}x = \infty$$

但 $S_n = X_1 + \cdots + X_n$ 经适当标准化后会收敛到正态分布. 令

$$Y_{n,m} = X_m 1_{(|X_m| \leq n^{1/2}\log\log n)}$$

截断水平 $c_n = n^{1/2}\log\log n$ 选得足够大使得

$$\sum_{m=1}^{n} P(Y_{n,m} \neq X_m) \leq nP(|X_1| > c_n) \to 0$$

但希望 $Y_{n,m}$ 的方差尽可能小，故将截断保持在尽可能低的水平.

下一步证明 $EY_{n,m}^2 \sim \log n$. 为此需要上下界估计. 由于 $P(|Y_{n,m}| > x) \leq P(|X_1| > x)$ 且 $x > c_n$ 时为 0，有

$$\begin{aligned} EY_{n,m}^2 &\leq \int_0^{c_n} 2yP(|X_1| > y)\mathrm{d}y = 1 + \int_1^{c_n} 2/y\mathrm{d}y \\ &= 1 + 2\log c_n = 1 + \log n + 2\log\log\log n \sim \log n \end{aligned}$$

另一方面，注意 $P(|Y_{n,m}| > x) = P(|X_1| > x) - P(|X_1| > c_n)$，右端在 $x \leq \sqrt{n}$ 时 $\geq (1 - (\log\log n)^{-2})P(|X_1| > x)$，故

$$EY_{n,m}^2 \geq (1 - (\log\log n)^{-2}) \int_1^{\sqrt{n}} 2/y\mathrm{d}y \sim \log n$$

若 $S_n' = Y_{n,1} + \cdots + Y_{n,n}$，则 $\mathrm{var}(S_n') \sim n\log n$，于是对 $X_{n,m} = Y_{n,m}/(n\log n)^{1/2}$ 应用定理 3.4.10. 已安排使得 (i) 满足. 由于 $|Y_{n,m}| \leq n^{1/2}\log\log n$，大 n 时 (ii) 中的和为 0，故 $S_n'/(n\log n)^{1/2} \Rightarrow \chi$. 因 c_n 的选择保证 $P(S_n \neq S_n') \to 0$，对 S_n 同样结论成立.

注释. 在 3.6 节，我们将看到若将例 3.4.13 中 $P(|X_1| > x) = x^{-2}$ 替换为 $P(|X_1| > x) = x^{-\alpha}$ $(0 < \alpha < 2)$，则 $S_n/n^{1/\alpha} \Rightarrow$ 某非 χ 极限. 关于正态分布收敛的最终结论是如下 Lévy 的结果.

定理 3.4.14. 设 $X_1, X_2, \ldots$ 独立同分布，$S_n = X_1 + \cdots + X_n$. 为使存在常数 a_n 和 $b_n > 0$ 使得 $(S_n - a_n)/b_n \Rightarrow \chi$，必要且充分的条件是

$$y^2P(|X_1| > y)/E(|X_1|^2; |X_1| \leq y) \to 0$$

证明见 Gnedenko 和 Kolmogorov(1954)，该文献包含许多关于独立随机变量之和的结果的最终结论.

习题

以下五题中 $X_1, X_2, \ldots$ 独立，$S_n = X_1 + \cdots + X_n$.

习题 3.4.9. 设 $P(X_m = m) = P(X_m = -m) = m^{-2}/2$，且 $m \geq 2$ 时 $P(X_m = 1) = P(X_m = -1) = (1 - m^{-2})/2$. 证明 $\text{var}(S_n)/n \to 2$ 但 $S_n/\sqrt{n} \Rightarrow \chi$. 问题在于 $X_{n,m} = X_m/\sqrt{n}$ 不满足定理 3.4.10 的条件 (ii).

习题 3.4.10. 证明若 $|X_i| \leq M$ 且 $\text{var}(S_n) = \infty$，则

$$(S_n - ES_n)/\sqrt{\text{var}(S_n)} \Rightarrow \chi$$

习题 3.4.11. 设 $EX_i = 0$，$EX_i^2 = 1$，$E|X_i|^{2+\delta} \leq C$ 对某 $0 < \delta, C < \infty$. 证明 $S_n/\sqrt{n} \Rightarrow \chi$.

习题 3.4.12. 证明 Lyapunov 定理. 令 $\alpha_n = \{\text{var}(S_n)\}^{1/2}$. 若存在 $\delta > 0$ 使得

$$\lim_{n\to\infty} \alpha_n^{-(2+\delta)} \sum_{m=1}^{n} E(|X_m - EX_m|^{2+\delta}) = 0$$

则 $(S_n - ES_n)/\alpha_n \Rightarrow \chi$. 注意前一习题是本结果的特例.

习题 3.4.13. 设 $P(X_j = j) = P(X_j = -j) = 1/2j^\beta$ 且 $P(X_j = 0) = 1 - j^{-\beta}$ （$\beta > 0$）. 证明：(i) 若 $\beta > 1$，则 $S_n \to S_\infty$ a.s. (ii) 若 $\beta < 1$，则 $S_n/n^{(3-\beta)/2} \Rightarrow c\chi$. (iii) 若 $\beta = 1$，则 $S_n/n \Rightarrow \aleph$，其中

$$E\exp(it\aleph) = \exp\left(-\int_0^1 x^{-1}(1 - \cos xt)\mathrm{d}x\right)$$

3.4.3 素因子（Erdős-Kac）*

这里的目标是证明：从 $\{1, 2, \ldots, n\}$ 中随机选取一个整数，它约有

$$\log\log n + \chi(\log\log n)^{1/2}$$

个素因子. 由于 $\exp(e^4) = 5.15 \times 10^{23}$，这一结果不适用于" 日常生活中" 遇到的大多数数. 推导该结果的第一步是给出

定理 3.4.10 的第二个证明. . 第一步令

$$h_n(\varepsilon) = \sum_{m=1}^{n} E(X_{n,m}^2; |X_{n,m}| > \varepsilon)$$

并注意

引理 3.4.15. 对每个固定的 $\varepsilon>0$，$h_n(\varepsilon)\to 0$，故可选取 $\varepsilon_n\to 0$ 使得 $h_n(\varepsilon_n)\to 0$.

证明. 选取 N_m 使得 $n\geq N_m$ 时 $h_n(1/m)\leq 1/m$ 且 N_m 递增. 令 $\varepsilon_n=1/m$ $(N_m\leq n<N_{m+1})$，$n<N_1$ 时 $\varepsilon_n=1$. 当 $N_m\leq n<N_{m+1}$，$\varepsilon_n=1/m$，故 $|h_n(\varepsilon_n)|=|h_n(1/m)|\leq 1/m$，即得所求. □

令 $X_{n,m}^0=X_{n,m}1_{(|X_{n,m}|>\varepsilon_n)}$，$Y_{n,m}=X_{n,m}1_{(|X_{n,m}|\leq\varepsilon_n)}$，$Z_{n,m}=Y_{n,m}-EY_{n,m}$. 显然 $|Z_{n,m}|\leq 2\varepsilon_n$. 利用 $X_{n,m}=X_{n,m}^0+Y_{n,m}$，$Z_{n,m}=Y_{n,m}-EY_{n,m}$，$EY_{n,m}=-EX_{n,m}^0$，和的方差等于方差之和，以及 $\mathrm{var}(W)\leq EW^2$，有

$$\begin{aligned}E\left(\sum_{m=1}^n X_{n,m}-\sum_{m=1}^n Z_{n,m}\right)^2&=E\left(\sum_{m=1}^n X_{n,m}^0-EX_{n,m}^0\right)^2\\&=\sum_{m=1}^n E(X_{n,m}^0-EX_{n,m}^0)^2\leq\sum_{m=1}^n E(X_{n,m}^0)^2\to 0\end{aligned}$$

当 $n\to\infty$，由 ε_n 的选择.

令 $S_n=\sum_{m=1}^n X_{n,m}$，$T_n=\sum_{m=1}^n Z_{n,m}$. 上述计算表明 $S_n-T_n\to 0$ 在 L^2 中从而依概率（由引理 2.2.2）. 于是由习题 3.2.13，只需证 $T_n\Rightarrow\sigma\chi$. (i) 蕴含 $ES_n^2\to\sigma^2$. 上面刚证明 $E(S_n-T_n)^2\to 0$，故 L^2 范数的三角不等式蕴含 $ET_n^2\to\sigma^2$. 为计算高阶矩，注意

$$T_n^r=\sum_{k=1}^r\sum_{\sum r_i=r}^{r_i}\frac{r!}{r_1!\cdots r_k!k!}\sum_{i_j}Z_{n,i_1}^{r_1}\cdots Z_{n,i_k}^{r_k}$$

其中求和遍历满足 $\sum r_i=r$ 的正整数 k-元组，i_j 求和遍历满足 $1\leq i_j\leq n$ 的不同整数 k-元组. 若令

$$A_n(r_1,\ldots,r_k)=\sum_{i_j}EZ_{n,i_1}^{r_1}\cdots Z_{n,i_k}^{r_k}$$

则

$$ET_n^r=\sum_{k=1}^r\sum_{\sum r_i=r}^{r_i}\frac{r!}{r_1!\cdots r_k!k!}A_n(r_1,\ldots,r_k)$$

为计算 ET_n^r 的极限，注意：

(a) 若某 $r_j=1$，则 $A_n(r_1,\ldots,r_k)=0$，因 $EZ_{n,i_j}=0$.

(b) 若所有 $r_j=2$，则

$$\sum_{i_j}EZ_{n,i_1}^2\cdots EZ_{n,i_k}^2\leq\left(\sum_{m=1}^n EZ_{n,m}^2\right)^k\to\sigma^{2k}$$

为论证另一不等式，注意对任意 $1 \le a < b \le k$，将 EZ_{n,i_a}^2 替换为 $(2\varepsilon_n)^2$ 可估计 $i_a = i_b$ 的和，得（因子 $\binom{k}{2}$ 给出选取 $1 \le a < b \le k$ 的方式数）

$$\left(\sum_{m=1}^{n} EZ_{n,m}^2\right)^k - \sum_{i_j} EZ_{n,i_1}^2 \cdots EZ_{n,i_k}^2 \le \binom{k}{2}(2\varepsilon_n)^2 \sum_{m=1}^{n} EZ_{n,m}^2 \to 0$$

(c) 若所有 $r_i \ge 2$ 但某 $r_j > 2$，则利用

$$E|Z_{n,i_j}|^{r_j} \le (2\varepsilon_n)^{r_j-2} EZ_{n,i_j}^2$$

有

$$|A_n(r_1, \ldots, r_k)| \le (2\varepsilon_n)^{r-2k} A_n(2, \ldots, 2) \to 0$$

r 为奇数时，某 r_j 必为 1 或 ≥ 3，故由 (a) 和 (c) 知 $ET_n^r \to 0$. 若 $r = 2k$ 为偶数，(a)–(c) 蕴含

$$ET_n^r \to \frac{\sigma^{2k}(2k)!}{2^k k!} = E(\sigma\chi)^r$$

由定理 3.3.26 即得结果. □

转向素因子的结果，令 P_n 为 $\{1, \ldots, n\}$ 上的均匀分布. 若 $P_\infty(A) \equiv \lim P_n(A)$ 存在，该极限称为 $A \subset \mathbb{Z}$ 的**密度**. 令 A_p 为被 p 整除的整数集. 显然，若 p 为素数，$P_\infty(A_p) = 1/p$，$q \ne p$ 为另一素数

$$P_\infty(A_p \cap A_q) = 1/pq = P_\infty(A_p)P_\infty(A_q)$$

尽管 P_∞ 不是概率测度（因对所有 i 有 $P_\infty(\{i\}) = 0$），可将此解释为被 p 和 q 整除的事件独立. 令 $\delta_p(n) = 1$（若 n 被 p 整除），否则为 0，且

$$g(n) = \sum_{p \le n} \delta_p(n) \text{ 为} n \text{ 的素因子个数}$$

其中对 p 的求和遍历素数. 直观上，$\delta_p(n)$ 的行为类似于独立 X_p，满足

$$P(X_p = 1) = 1/p, \quad P(X_p = 0) = 1 - 1/p$$

$\sum_{p \le n} X_p$ 的均值和方差分别为 $\sum_{p \le n} 1/p$ 和 $\sum_{p \le n} 1/p(1 - 1/p)$. 已知

$$\sum_{p \le n} 1/p = \log\log n + O(1) \qquad (*)$$

（见 Hardy 和 Wright(1959)，第 XXII 章），而 $\sum_p 1/p^2 < \infty$ 是显然的，故将定理 3.4.10 应用于 X_p

并作小小的信念跳跃得:

定理 3.4.16. Erdős-Kac 中心极限定理. 当 $n \to \infty$ 时

$$P_n\left(m \le n : g(m) - \log\log n \le x(\log\log n)^{1/2}\right) \to P(\chi \le x)$$

证明. 先证可忽略" 靠近"n 的素数. 令 $\alpha_n = n^{1/\log\log n}$

$$\log \alpha_n = \log n / \log\log n$$
$$\log\log \alpha_n = \log\log n - \log\log\log n$$

序列 α_n 有两个良好性质:

(a) $\sum_{\alpha_n < p \le n} 1/p/(\log\log n)^{1/2} \to 0$ (由 $(*)$)

(b) 若 $\varepsilon > 0$, 则大 n 时 $\alpha_n \le n^{\varepsilon}$, 从而对所有 $r < \infty$ 有 $\alpha_n^r/n \to 0$

(a) 的证明: 由 $(*)$

$$\begin{aligned}\sum_{\alpha_n < p \le n} 1/p &= \sum_{p \le n} 1/p - \sum_{p \le \alpha_n} 1/p \\ &= \log\log n - \log\log \alpha_n + O(1) \\ &= \log\log\log n + O(1)\end{aligned}$$

(b) 的证明: $1/\log\log n \to 0$ 当 $n \to \infty$.

令 $g_n(m) = \sum_{p \le \alpha_n} \delta_p(m)$, E_n 表示关于 P_n 的期望.

$$E_n \sum_{\alpha_n < p \le n} \delta_p = \sum_{\alpha_n < p \le n} P_n(m : \delta_p(m) = 1) \le \sum_{\alpha_n < p \le n} 1/p$$

由 (a), 只需对 g_n 证明结果. 令

$$S_n = \sum_{p \le \alpha_n} X_p$$

其中 X_p 为上述独立随机变量. 令 $b_n = ES_n$, $a_n^2 = \text{var}(S_n)$. (a) 告诉我们 b_n 和 a_n^2 均为 $\log\log n + o((\log\log n)^{1/2})$, 故只需证

$$P_n(m : g_n(m) - b_n \le x a_n) \to P(\chi \le x)$$

应用定理 3.4.10 得 $(S_n - b_n)/a_n \Rightarrow \chi$, 且因 $|X_p| \le 1$, 由定理 3.4.10 的第二个证明知

$$E((S_n - b_n)/a_n)^r \to E\chi^r \text{ 对所有} r$$

沿用该证明的记号（将 i_j 换为 p_j）

$$ES_n^r = \sum_{k=1}^{r} \sum_{\substack{r_i \\ \sum r_i = r}} \frac{r!}{r_1! \cdots r_k! k!} \sum_{p_j} E(X_{p_1}^{r_1} \cdots X_{p_k}^{r_k})$$

因 $X_p \in \{0, 1\}$，被加项为

$$E(X_{p_1} \cdots X_{p_k}) = 1/(p_1 \cdots p_k)$$

稍加思考可知

$$E_n(\delta_{p_1} \cdots \delta_{p_k}) \leq \frac{1}{n}[n/(p_1 \cdots p_k)]$$

两矩之差 $\leq 1/n$，故

$$|E(S_n^r) - E_n(g_n^r)| = \sum_{k=1}^{r} \sum_{\substack{r_i \\ \sum r_i = r}} \frac{r!}{r_1! \cdots r_k! k!} \sum_{p_j} \frac{1}{n} \leq 13n \left(\sum_{p \leq \alpha_n} 1 \right)^r \frac{1}{n} \leq \frac{\alpha_n^r}{n} \to 0$$

由 (b). 现

$$E(S_n - b_n)^r = \sum_{m=0}^{r} \binom{r}{m} ES_n^m (-b_n)^{r-m}$$
$$E_n(g_n - b_n)^r = \sum_{m=0}^{r} \binom{r}{m} E_n g_n^m (-b_n)^{r-m}$$

相减并利用 $|E(S_n^m) - E_n(g_n^m)|$（取 $r = m$）的估计得

$$|E(S_n - b_n)^r - E_n(g_n - b_n)^r| \leq \sum_{m=0}^{r} \binom{r}{m} \frac{1}{n} \alpha_n^m b_n^{r-m} = (\alpha_n + b_n)^r / n \to 0$$

因 $b_n \leq \alpha_n$. 这足以推出 $E((g_n - b_n)/a_n)^r \to E\chi^r$，由定理 3.3.26 即得所求结果. □

3.4.4 收敛速度（Berry-Esseen）*

定理 3.4.17. 设 $X_1, X_2, \ldots$ 独立同分布，$EX_i = 0$，$EX_i^2 = \sigma^2$，$E|X_i|^3 = \rho < \infty$. 若 $F_n(x)$ 为 $(X_1 + \cdots + X_n)/\sigma\sqrt{n}$ 的分布，$N(x)$ 为标准正态分布，则

$$|F_n(x) - N(x)| \leq 3\rho/\sigma^3\sqrt{n}$$

注释. 读者应注意该不等式对所有 n 和 x 成立，但由于 $\rho \geq \sigma^3$，它只对 $n \geq 10$ 有非平凡内容. 易见

速度不可能快于 $n^{-1/2}$. 当 $P(X_i=1)=P(X_i=-1)=1/2$ 时，对称性和 (1.4) 蕴含

$$F_{2n}(0)=\frac{1}{2}\{1+P(S_{2n}=0)\}=\frac{1}{2}(1+(\pi n)^{-1/2})+o(n^{-1/2})$$

常数 3 不是已知最好的（van Beek(1972) 得到 0.8），但如 Feller 所言，" 简化的方法即使避免了通常繁琐的数值计算，仍能给出相当好的界." 假设 $E|X|^3$ 是为获得 $n^{-1/2}$ 速度所必需的. Heyde(1967) 证明对 $0<\delta<1$

$$\sum_{n=1}^{\infty}n^{-1+\delta/2}\sup_x|F_n(x)-N(x)|<\infty$$

当且仅当 $E|X|^{2+\delta}<\infty$. 关于收敛速度的更多内容，见 Hall(1982).

证明. 因不等式两边不受尺度变换影响，**不失一般性**可设 $\sigma^2=1$. 论证的第一阶段是推导一个将两分布之差与它们的特征函数之差联系起来的不等式（引理 3.4.19）. Polya 密度（见例 3.3.15 并用定理 3.3.1(e)）

$$h_L(x)=\frac{1-\cos Lx}{\pi Lx^2}$$

的特征函数为 $\omega_L(\theta)=(1-|\theta/L|)^+$（$|\theta|\le L$）. 用 H_L 表示其分布函数. 将用 H_L 与考虑的分布做卷积，以得到具有紧支撑的特征函数. 第一步是证明与 H_L 卷积不会使分布之差减小太多.

引理 3.4.18. 设 F 和 G 为分布函数，$G'(x)\le\lambda<\infty$. 令 $\Delta(x)=F(x)-G(x)$，$\eta=\sup_x|\Delta(x)|$，$\Delta_L=\Delta*H_L$，$\eta_L=\sup_x|\Delta_L(x)|$. 则

$$\eta_L\ge\frac{\eta}{2}-\frac{12\lambda}{\pi L}\quad\text{或}\quad\eta\le 2\eta_L+\frac{24\lambda}{\pi L}$$

证明. Δ 在 $\pm\infty$ 趋于 0，G 连续，F 是分布函数，故存在 x_0 使得 $\Delta(x_0)=\eta$ 或 $\Delta(x_0-)=-\eta$. 对第二种情形考虑 (-1) 乘以相应随机变量的分布函数，**不失一般性**可设 $\Delta(x_0)=\eta$. 由于 $G'(x)\le\lambda$ 且 F 不减，$\Delta(x_0+s)\ge\eta-\lambda s$. 令 $\delta=\eta/2\lambda$，$t=x_0+\delta$，有

$$\Delta(t-x)\ge\begin{cases}(\eta/2)+\lambda x & |x|\le\delta\\ -\eta & \text{否则}\end{cases}$$

为估计卷积 Δ_L，注意

$$2\int_{\delta}^{\infty}h_L(x)\mathrm{d}x\le 2\int_{\delta}^{\infty}\frac{2}{\pi Lx^2}\mathrm{d}x=\frac{4}{\pi L\delta}$$

分开考虑 $(-\delta,\delta)$ 及其补集，并利用对称性 $\int_{-\delta}^{\delta}xh_L(x)\mathrm{d}x=0$，有

$$\eta_L\ge\Delta_L(t)\ge\frac{\eta}{2}\left(1-\frac{4}{\pi L\delta}\right)-\eta\frac{4}{\pi L\delta}=\frac{\eta}{2}-\frac{6\eta}{\pi L\delta}=\frac{\eta}{2}-\frac{12\lambda}{\pi L}$$

即得引理. □

引理 3.4.19. 设 K_1 和 K_2 为均值为 0 且特征函数 κ_i 可积的分布函数

$$K_1(x) - K_2(x) = (2\pi)^{-1} \int e^{-itx} \frac{\kappa_1(t) - \kappa_2(t)}{-it} \mathrm{d}t$$

证明. 因 κ_i 可积，反演公式（定理 3.3.11）蕴含密度 $k_i(x)$ 满足

$$k_i(y) = (2\pi)^{-1} \int e^{-ity} \kappa_i(t) \mathrm{d}t$$

将 $i = 2$ 的表达式从 $i = 1$ 的表达式中减去，然后从 a 到 x 积分，令 $\Delta K = K_1 - K_2$ 得

$$\begin{aligned} \Delta K(x) - \Delta K(a) &= (2\pi)^{-1} \int_a^x \int e^{-ity} \{\kappa_1(t) - \kappa_2(t)\} \mathrm{d}t \mathrm{d}y \\ &= (2\pi)^{-1} \int \{e^{-ita} - e^{-itx}\} \frac{\kappa_1(t) - \kappa_2(t)}{it} \mathrm{d}t \end{aligned}$$

Fubini 定理的应用是合理的，因 κ_i 对 t 可积且考虑 y 的有界区间.

因子 $1/it$ 在零附近可能有问题，但假设 K_i 均值为 0，故 $\{1 - \kappa_i(t)\}/t \to 0$（由习题 3.3.18），从而 $(\kappa_1(t) - \kappa_2(t))/it$ 有界且连续. 因子 $1/it$ 改善了大 t 的可积性，故 $(\kappa_1(t) - \kappa_2(t))/it$ 可积. 令 $a \to -\infty$ 并用 Riemann-Lebesgue 引理（习题 1.4.4）即得结果. □

令 ϕ_F 和 ϕ_G 分别为 F 和 G 的特征函数. 将引理 3.4.19 应用于 $F_L = F * H_L$ 和 $G_L = G * H_L$，得

$$|F_L(x) - G_L(x)| \leq \frac{1}{2\pi} \int \frac{|\phi_F(t)\omega_L(t) - \phi_G(t)\omega_L(t)|}{|t|} \mathrm{d}t \leq \frac{1}{2\pi} \int_{-L}^{L} \frac{|\phi_F(t) - \phi_G(t)|}{|t|} \mathrm{d}t$$

因 $|\omega_L(t)| \leq 1$. 现用引理 3.4.18，有

$$|F(x) - G(x)| \leq \frac{1}{\pi} \int_{-L}^{L} \frac{|\phi_F(\theta) - \phi_G(\theta)|}{|\theta|} \mathrm{d}\theta + \frac{24\lambda}{\pi L}$$

其中 $\lambda = \sup_x G'(x)$. 代入 $F = F_n$ 和 $G = N$ 得

$$|F_n(x) - N(x)| \leq \frac{1}{\pi} \int_{-L}^{L} \frac{|\phi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} \mathrm{d}\theta + \frac{24\lambda}{\pi L} \tag{3.4.1}$$

还需估计右端. 这部分论证相当常规，但有大量代数运算. 为省却读者试图改进不等式以获得更好界的努力，这里指出已用 $C = 3$ 来排除 $n \leq 9$ 的情形，并在 (e) 中用 $n \geq 10$.

为估计 (3.4.1) 中的第二项，注意

$$\sup_x N'(x) = N'(0) = (2\pi)^{-1/2} = 0.39894 < 2/5 \tag{a}$$

对第一项，注意若 $|\alpha|, |\beta| \le \gamma$

$$|\alpha^n - \beta^n| \le \sum_{m=0}^{n-1} |\alpha^{n-m}\beta^m - \alpha^{n-m-1}\beta^{m+1}| \le n|\alpha - \beta|\gamma^{n-1} \tag{b}$$

现用 (3.3.3)（回忆假设 $\sigma^2 = 1$）

$$|\phi(t) - 1 + t^2/2| \le \rho|t|^3/6 \tag{c}$$

故若 $t^2 \le 2$

$$|\phi(t)| \le 1 - t^2/2 + \rho|t|^3/6 \tag{d}$$

令 $L = 4\sqrt{n}/3\rho$. 若 $|\theta| \le L$，由 (d) 及事实 $\rho|\theta|/\sqrt{n} \le 4/3$

$$\begin{aligned} |\phi(\theta/\sqrt{n})| &\le 1 - \theta^2/2n + \rho|\theta|^3/6n^{3/2} \\ &\le 1 - 5\theta^2/18n \le \exp(-5\theta^2/18n) \end{aligned}$$

因 $1 - x \le e^{-x}$. 现对

$$\alpha = \phi(\theta/\sqrt{n}), \quad \beta = \exp(-\theta^2/2n), \quad \gamma = \exp(-5\theta^2/18n)$$

应用 (b). 因假设 $n \ge 10$

$$\gamma^{n-1} \le \exp(-\theta^2/4) \tag{e}$$

对 (b) 的另一部分，写

$$n|\alpha - \beta| \le n|\phi(\theta/\sqrt{n}) - 1 + \theta^2/2n| + n|1 - \theta^2/2n - \exp(-\theta^2/2n)|$$

为估计右端第一项，注意 (c) 蕴含

$$n|\phi(\theta/\sqrt{n}) - 1 + \theta^2/2n| \le \rho|\theta|^3/6n^{1/2}$$

对第二项，注意若 $0 < x < 1$，则有一个递减项的交错级数，故

$$|e^{-x} - (1-x)| = \left|\frac{x^2}{2!} - \frac{x^3}{3!} + \cdots\right| \le \frac{x^2}{2}$$

取 $x=\theta^2/2n$，对 $|\theta|\le L\le 2\sqrt{n}$ 有

$$n|1-\theta^2/2n-\exp(-\theta^2/2n)|\le \theta^4/8n$$

结合第一项的估计得

$$n|\alpha-\beta|\le \rho|\theta|^3/6n^{1/2}+\theta^4/8n \tag{f}$$

在 (b) 中用 (f) 和 (e) 得

$$\frac{|\phi^n(\theta/\sqrt{n})-\exp(-\theta^2/2)|}{|\theta|}\le \exp(-\theta^2/4)\left(\frac{\rho\theta^2}{6n^{1/2}}+\frac{|\theta|^3}{8n}\right)\le \frac{1}{L}\exp(-\theta^2/4)\left(\frac{2\theta^2}{9}+\frac{|\theta|^3}{18}\right)$$

因 $\rho/\sqrt{n}=4/3L$，$1/n=1/\sqrt{n}\cdot 1/\sqrt{n}\le 4/3L\cdot 1/3$（因 $\rho\ge 1$ 且 $n\ge 10$）. 用上述结果和 (a) 于引理 3.4.19 得

$$\pi L|F_n(x)-N(x)|\le \int \exp(-\theta^2/4)\left(\frac{2\theta^2}{9}+\frac{|\theta|^3}{18}\right)\mathrm{d}\theta+9.6$$

回忆 $L=4\sqrt{n}/3\rho$，可见上述结果形如 $|F_n(x)-N(x)|\le C\rho/\sqrt{n}$. 为计算常数，注意

$$\int (2\pi a)^{-1/2}x^2\exp(-x^2/2a)\mathrm{d}x=a$$

写 $x^3=2x^2\cdot x/2$ 并分部积分

$$2\int_0^\infty x^3\exp(-x^2/4)\mathrm{d}x=2\int_0^\infty 4x\exp(-x^2/4)\mathrm{d}x=-16e^{-x^2/4}\Big|_0^\infty=16$$

这给出

$$|F_n(x)-N(x)|\le \frac{1}{\pi}\cdot\frac{3}{4}\cdot\left(\frac{2}{9}\cdot 2\cdot\sqrt{4\pi}+\frac{16}{18}+9.6\right)\frac{\rho}{\sqrt{n}}<3\frac{\rho}{\sqrt{n}}$$

最后一步，你需要拿出计算器或相信 Feller. □

第 5 节 局部极限定理 *

在 3.1 节中，若 $X_1, X_2, \ldots$ 独立同分布且 $P(X_1 = 1) = P(X_1 = -1) = 1/2$，而 k_n 为整数列满足 $2k_n/(2n)^{1/2} \to x$，则有

$$P(S_{2n} = 2k_n) \sim (\pi n)^{-1/2} \exp(-x^2/2).$$

本节证明两个定理以推广该结果. 首先给出两个定义. 若存在常数 b 和 $h > 0$ 使得 $P(X \in b + h\mathbb{Z}) = 1$，其中 $b + h\mathbb{Z} = \{b + hz : z \in \mathbb{Z}\}$，则称随机变量 X 服从**格子点分布**. 使该式成立的最大 h 称为该分布的**跨度**.

例题 3.5.1. 若 $P(X = 1) = P(X = -1) = 1/2$，则 X 服从跨度为 2 的格子点分布. 当 $h = 2$ 时，可取 $b = -1$.

下面的结果将上述定义与特征函数联系起来. 为验证其中的 (ii)，注意在上例中 $E(e^{itX}) = \cos t$ 在 $t = n\pi$ 处满足 $|\cos t| = 1$.

定理 3.5.2. 设 $\varphi(t) = Ee^{itX}$. 只有以下三种可能：

(i) 对所有 $t \neq 0$ 有 $|\varphi(t)| < 1$.

(ii) 存在 $\lambda > 0$ 使得 $|\varphi(\lambda)| = 1$ 且对 $0 < t < \lambda$ 有 $|\varphi(t)| < 1$. 此时 X 服从跨度为 $2\pi/\lambda$ 的格子点分布.

(iii) 对所有 t 有 $|\varphi(t)| = 1$. 此时 $X = b$ 几乎必然对某个 b 成立.

证明. 先证 (ii). 只需证明 $|\varphi(t)| = 1$ 当且仅当 $P(X \in b + (2\pi/t)\mathbb{Z}) = 1$ 对某个 b 成立. 首先，若 $P(X \in b + (2\pi/t)\mathbb{Z}) = 1$，则

$$\varphi(t) = Ee^{itX} = e^{itb} \sum_{n \in \mathbb{Z}} e^{i2\pi n} P(X = b + (2\pi/t)n) = e^{itb}.$$

反之，若 $|\varphi(t)| = 1$，则在不等式 $|Ee^{itX}| \leq E|e^{itX}|$ 中等号成立，故由习题 1.6.1 知 e^{itX} 的分布必集中于某点 e^{itb}，从而 $P(X \in b + (2\pi/t)\mathbb{Z}) = 1$.

再证三分法. 假设 (i) 和 (ii) 均不成立，即存在序列 $t_n \downarrow 0$ 使得 $|\varphi(t_n)| = 1$. 第一段表明存在 b_n 使得 $P(X \in b_n + (2\pi/t_n)\mathbb{Z}) = 1$. 不失一般性，可取 $b_n \in (-\pi/t_n, \pi/t_n]$. 当 $n \to \infty$ 时，$P(X \in (-\pi/t_n, \pi/t_n]) \to 0$，故 $P(X = b_n) \to 1$. 这仅当 $b_n = b$ $(n \geq N)$ 且 $P(X = b) = 1$ 时才可能成立. □

定理 3.5.2 中的三种情形称为：(i) **非格子点**，(ii) **格子点**，(iii) **退化**. 读者应注意，按定义格子点随机变量是非退化的. 在转入本节主题之前，再介绍一种特殊情形. 若 X 服从格子点分布且可取 $b = 0$，即 $P(X \in h\mathbb{Z}) = 1$，则称 X 为**算术的**. 此时若 $\lambda = 2\pi/h$，则 $\varphi(\lambda) = 1$ 且 φ 是周期的：$\varphi(t + \lambda) = \varphi(t)$.

首先考虑格子点情形. 设 $X_1, X_2, \ldots$ 独立同分布, 满足 $EX_i = 0$, $EX_i^2 = \sigma^2 \in (0, \infty)$, 且具有共同的跨度为 h 的格子点分布. 若 $S_n = X_1 + \cdots + X_n$ 且 $P(X_i \in b + h\mathbb{Z}) = 1$, 则 $P(S_n \in nb + h\mathbb{Z}) = 1$. 记

$$p_n(x) = P(S_n/\sqrt{n} = x) \quad \text{对} x \in L_n = \{(nb + hz)/\sqrt{n} : z \in \mathbb{Z}\}$$

且

$$n(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2) \quad \text{对} x \in (-\infty, \infty).$$

定理 3.5.3. 在上述假设下，当 $n \to \infty$ 时

$$\sup_{x \in L_n} \left| \frac{n^{1/2}}{h} p_n(x) - n(x) \right| \to 0.$$

注释. 为解释该结论，注意若沿用例 3.4.7 的方法，则对 $x \in L_n$ 有

$$p_n(x) \approx \int_{x - h/2\sqrt{n}}^{x + h/2\sqrt{n}} n(y)\, \mathrm{d}y \approx \frac{h}{\sqrt{n}} n(x).$$

证明. 设 Y 为满足 $P(Y \in a + \theta\mathbb{Z}) = 1$ 的随机变量且 $\psi(t) = E \exp(itY)$. 由习题 3.3.2 的 (iii) 得

$$P(Y = x) = \frac{1}{2\pi/\theta} \int_{-\pi/\theta}^{\pi/\theta} e^{-itx} \psi(t)\, \mathrm{d}t.$$

将 $\theta = h/\sqrt{n}$，$\psi(t) = E \exp(itS_n/\sqrt{n}) = \varphi^n(t/\sqrt{n})$ 代入该公式，然后两边乘以 $1/\theta$ 得

$$\frac{n^{1/2}}{h} p_n(x) = \frac{1}{2\pi} \int_{-\pi\sqrt{n}/h}^{\pi\sqrt{n}/h} e^{-itx} \varphi^n(t/\sqrt{n})\, \mathrm{d}t.$$

对特征函数为 $\exp(-\sigma^2 t^2/2)$ 的 $n(x)$ 使用反演公式（定理 3.3.14）得

$$n(x) = \frac{1}{2\pi} \int e^{-itx} \exp(-\sigma^2 t^2/2)\, \mathrm{d}t.$$

两式相减得（注意 $\pi > 1$，$|e^{-itx}| \le 1$）

$$\begin{aligned} \left| \frac{n^{1/2}}{h} p_n(x) - n(x) \right| &\le \frac{1}{2\pi} \int_{-\pi\sqrt{n}/h}^{\pi\sqrt{n}/h} |\varphi^n(t/\sqrt{n}) - \exp(-\sigma^2 t^2/2)|\, \mathrm{d}t \\ &\quad + \frac{1}{2\pi} \int_{\pi\sqrt{n}/h}^{\infty} \exp(-\sigma^2 t^2/2)\, \mathrm{d}t. \end{aligned}$$

右边与 x 无关，故为证定理 3.5.3，只需证明其趋于 0. 第二个积分显然 $\to 0$. 为估计第一个积分，注意 $\varphi^n(t/\sqrt{n}) \to \exp(-\sigma^2 t^2/2)$，故被积函数趋于 0，下面只需 “应用控制收敛定理”.

为此，将积分分成三块. 有界收敛定理蕴含对任意 $A<\infty$，$(-A,A)$ 上的积分趋于 0. 为估计 $(-A,A)^c$ 上的积分，注意因 $EX_i=0$ 且 $EX_i^2=\sigma^2$，公式 (3.3.3) 和三角不等式蕴含

$$|\varphi(u)|\le|1-\sigma^2u^2/2|+\frac{u^2}{2}E(\min(|u|\cdot|X|^3,6|X|^2)).$$

最后的期望值当 $u\to0$ 时趋于 0. 这意味着可取 $\delta>0$ 使得当 $|u|<\delta$ 时它 $\le\sigma^2/2$，从而

$$|\varphi(u)|\le1-\sigma^2u^2/2+\sigma^2u^2/4=1-\sigma^2u^2/4\le\exp(-\sigma^2u^2/4)$$

因 $1-x\le e^{-x}$. 将最后结果应用于 $u=t/\sqrt{n}$ 可见对 $t\le\delta\sqrt{n}$ 有

$$|\varphi(t/\sqrt{n})^n|\le\exp(-\sigma^2t^2/4). \tag{$*$}$$

于是 $(-\delta\sqrt{n},\delta\sqrt{n})-(-A,A)$ 上的积分小于

$$2\int_A^{\delta\sqrt{n}}\exp(-\sigma^2t^2/4)\,\mathrm{d}t,$$

当 A 大时这是小的.

为估计积分的剩余部分，注意因 X 的跨度为 h，定理 3.5.2 蕴含 $|\varphi(u)|\ne1$ 对 $u\in[\delta,\pi/h]$. φ 连续，故存在 $\eta<1$ 使得 $|\varphi(u)|\le\eta<1$ 对 $|u|\in[\delta,\pi/h]$ 成立. 再令 $u=t/\sqrt{n}$，可见 $[-\pi\sqrt{n}/h,\pi\sqrt{n}/h]-(-\delta\sqrt{n},\delta\sqrt{n})$ 上的积分小于

$$2\int_{\delta\sqrt{n}}^{\pi\sqrt{n}/h}(\eta^n+\exp(-\sigma^2t^2/2))\,\mathrm{d}t,$$

当 $n\to\infty$ 时趋于 0. 证毕. □

现在转向非格子点情形. 设 $X_1,X_2,\ldots$ 独立同分布，满足 $EX_i=0$，$EX_i^2=\sigma^2\in(0,\infty)$，且具有共同的特征函数 $\varphi(t)$ 满足对所有 $t\ne0$ 有 $|\varphi(t)|<1$. 设 $S_n=X_1+\cdots+X_n$ 且 $n(x)=(2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$.

定理 3.5.4. 在上述假设下，若 $x_n/\sqrt{n}\to x$ 且 $a<b$，则

$$\sqrt{n}P(S_n\in(x_n+a,x_n+b))\to(b-a)n(x).$$

注释. 该结果的证明必须有点迂回，因为上述假设未能给 φ 的行为提供多少控制. 一个不好的例子：设 $q_1,q_2,\ldots$ 为正有理数的一个枚举满足 $q_n\le n$. 假设

$$P(X=q_n)=P(X=-q_n)=1/2^{n+1}.$$

此时 $EX=0$，$EX^2<\infty$，且分布是非格子点的. 然而特征函数满足 $\limsup_{t\to\infty}|\varphi(t)|=1$.

证明. 为驯服病态的特征函数，这里用一个技巧. 设 $\delta>0$

$$h_0(y)=\frac{1}{\pi}\cdot\frac{1-\cos\delta y}{\delta y^2}$$

为 Polya 分布的密度，且令 $h_\theta(x)=e^{i\theta x}h_0(x)$. 若引入 Fourier 变换

$$\hat{g}(u)=\int e^{iuy}g(y)\,\mathrm{d}y,$$

则由例 3.3.15 得

$$\hat{h}_0(u)=\begin{cases}1-|u/\delta| & \text{若}|u|\le\delta,\\ 0 & \text{其他情形,}\end{cases}$$

且易见 $\hat{h}_\theta(u)=\hat{h}_0(u+\theta)$. 下面证明对任意 θ 有

$$\sqrt{n}Eh_\theta(S_n-x_n)\to n(x)\int h_\theta(y)\,\mathrm{d}y. \tag{a}$$

先证 (a) 蕴含定理结论. 令

$$\mu_n(A)=\sqrt{n}P(S_n-x_n\in A),\quad \text{且}\quad \mu(A)=n(x)|A|,$$

其中 $|A|$ 为 A 的 Lebesgue 测度. 令

$$\alpha_n=\sqrt{n}Eh_0(S_n-x_n)\quad \text{且}\quad \alpha=n(x)\int h_0(y)\,\mathrm{d}y=n(x).$$

最后定义概率测度

$$\nu_n(B)=\frac{1}{\alpha_n}\int_B h_0(y)\mu_n(\mathrm{d}y),\quad \text{且}\quad \nu(B)=\frac{1}{\alpha}\int_B h_0(y)\mu(\mathrm{d}y).$$

在 (a) 中取 $\theta=0$ 可见 $\alpha_n\to\alpha$，故 (a) 蕴含

$$\int e^{i\theta y}\nu_n(\mathrm{d}y)\to\int e^{i\theta y}\nu(\mathrm{d}y). \tag{b}$$

因这对所有 θ 成立，由定理 3.3.17 得 $\nu_n\Rightarrow\nu$. 现若 $|a|,|b|<2\pi/\delta$，则函数

$$k(y)=\frac{1}{h_0(y)}\cdot 1_{(a,b)}(y)$$

关于 ν 几乎处处有界连续，故由定理 3.2.10 得

$$\int k(y)\nu_n(\mathrm{d}y) \to \int k(y)\nu(\mathrm{d}y).$$

因 $\alpha_n \to \alpha$，这蕴含

$$\sqrt{n}P(S_n \in (x_n + a, x_n + b)) \to (b - a)n(x),$$

即定理 3.5.4 的结论.

下证 (a)，反演公式（定理 3.3.14）蕴含

$$h_0(x) = \frac{1}{2\pi}\int e^{-iux}\hat{h}_0(u)\,\mathrm{d}u.$$

回忆 h_θ 的定义，利用上式并换元 $u = v + \theta$ 得

$$\begin{aligned} h_\theta(x) = e^{i\theta x}h_0(x) &= \frac{1}{2\pi}\int e^{-i(u-\theta)x}\hat{h}_0(u)\,\mathrm{d}u \\ &= \frac{1}{2\pi}\int e^{-ivx}\hat{h}_\theta(v)\,\mathrm{d}v \end{aligned}$$

因 $\hat{h}_\theta(v) = \hat{h}_0(v + \theta)$. 设 F_n 为 $S_n - x_n$ 的分布并积分得

$$\begin{aligned} Eh_\theta(S_n - x_n) &= \frac{1}{2\pi}\int\int e^{-iux}\hat{h}_\theta(u)\,\mathrm{d}u\,\mathrm{d}F_n(x) \\ &= \frac{1}{2\pi}\int\int e^{-iux}\,\mathrm{d}F_n(x)\hat{h}_\theta(u)\,\mathrm{d}u \end{aligned}$$

由 Fubini 定理.（注意 $\hat{h}_\theta(u)$ 具有紧支集且 F_n 是分布函数.）利用定理 3.3.1(e)，可见上式为

$$\frac{1}{2\pi}\int \varphi(-u)^n e^{iux_n}\hat{h}_\theta(u)\,\mathrm{d}u.$$

为取 $n \to \infty$ 的极限，设 $[-M, M]$ 为 $\hat{h}_\theta(u) = 0$ 对 $u \notin [-M, M]$ 的区间. 由上述 $(*)$，可取 δ 使得对 $|u| < \delta$ 有

$$|\varphi(u)| \le \exp(-\sigma^2u^2/4). \tag{c}$$

令 $I = [-\delta, \delta]$ 且 $J = [-M, M] - I$. 因 $|\varphi(u)| < 1$ 对 $u \ne 0$ 且 φ 连续，存在常数 $\eta < 1$ 使得 $|\varphi(u)| \le \eta < 1$ 对 $u \in J$. 因 $|\hat{h}_\theta(u)| \le 1$，这蕴含

$$\left|\frac{\sqrt{n}}{2\pi}\int_J \varphi(-u)^n e^{iux_n}\hat{h}_\theta(u)\,\mathrm{d}u\right| \le \frac{\sqrt{n}}{2\pi}\cdot 2M\eta^n \to 0$$

当 $n \to \infty$. 对 I 上的积分，换元 $u = t/\sqrt{n}$ 得

$$\frac{1}{2\pi} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \varphi(-t/\sqrt{n})^n e^{itx_n/\sqrt{n}} \hat{h}_\theta(t/\sqrt{n}) \, \mathrm{d}t.$$

中心极限定理蕴含 $\varphi(-t/\sqrt{n})^n \to \exp(-\sigma^2 t^2/2)$. 现利用 (c) 和控制收敛定理得（注意 $x_n/\sqrt{n} \to x$）

$$\begin{aligned} \frac{\sqrt{n}}{2\pi} \int_I \varphi(-u)^n e^{iux_n} \hat{h}_\theta(u) \, \mathrm{d}u &\to \frac{1}{2\pi} \int \exp(-\sigma^2 t^2/2) e^{itx} \hat{h}_\theta(0) \, \mathrm{d}t \\ &= n(x) \hat{h}_\theta(0) = n(x) \int h_\theta(y) \, \mathrm{d}y \end{aligned}$$

由反演公式（定理 3.3.14）和 $\hat{h}_\theta(0)$ 的定义. 这证明了 (a) 从而完成了定理 3.5.4 的证明. □

第 6 节 Poisson 收敛

3.6.1 基本极限定理

首先介绍的结果有时被戏称为“小数弱定律”或“稀有事件定律”. 这些名称源于如下事实：Poisson 分布可作为具有小概率事件之示性变量和的极限出现.

定理 3.6.1. 对每个 n，设 $X_{n,m}$ $(1 \le m \le n)$ 为独立随机变量，满足 $P(X_{n,m}=1)=p_{n,m}$，$P(X_{n,m}=0)=1-p_{n,m}$. 假设

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0,\infty)$,

(ii) $\max_{1\le m\le n} p_{n,m} \to 0$.

若 $S_n = X_{n,1}+\cdots+X_{n,n}$，则 $S_n \Rightarrow Z$，其中 Z 服从 Poisson(λ).

这里 Poisson(λ) 是均值为 λ 的 Poisson 分布的简写，即

$$P(Z=k)=e^{-\lambda}\lambda^k/k!$$

注意，本着 Lindeberg-Feller 定理的精神，没有单独一项对和的贡献很大. 但与该定理不同的是，这里的贡献一旦为正就并非小量.

第一个证明. 令 $\varphi_{n,m}(t)=E(\exp(itX_{n,m}))=(1-p_{n,m})+p_{n,m}e^{it}$，并记 $S_n=X_{n,1}+\cdots+X_{n,n}$. 则

$$E\exp(itS_n)=\prod_{m=1}^{n}(1+p_{n,m}(e^{it}-1))$$

设 $0\le p\le 1$. 由 $|\exp(p(e^{it}-1))|=\exp(p\,\mathrm{Re}(e^{it}-1))\le 1$ 以及 $|1+p(e^{it}-1)|\le 1$（因其位于连接 1 与 e^{it} 的线段上）. 利用引理 3.4.3（取 $\theta=1$）和引理 3.4.4（当 $\max_m p_{n,m}\le 1/2$ 时有效，因 $|e^{it}-1|\le 2$），

$$\begin{aligned}\left|\exp\left(\sum_{m=1}^{n}p_{n,m}(e^{it}-1)\right)-\prod_{m=1}^{n}\{1+p_{n,m}(e^{it}-1)\}\right| &\le \sum_{m=1}^{n}\left|\exp(p_{n,m}(e^{it}-1))-\{1+p_{n,m}(e^{it}-1)\}\right| \\ &\le \sum_{m=1}^{n}p_{n,m}^2|e^{it}-1|^2.\end{aligned}$$

再次利用 $|e^{it}-1|\le 2$，可知最后一个表达式

$$\le 4\left(\max_{1\le m\le n}p_{n,m}\right)\sum_{m=1}^{n}p_{n,m}\to 0$$

由假设 (i) 和 (ii). 最后一个结论与 $\sum_{m=1}^{n} p_{n,m} \to \lambda$ 蕴含

$$E\exp(itS_n) \to \exp(\lambda(e^{it}-1)).$$

为完成证明，查阅例 3.3.4 中 Poisson 分布的特征函数并应用定理 3.3.17. □

下面考虑一些可应用定理 3.6.1 的具体情形. 在每种情形下，取 $p_{n,m} = c/n$，故用均值为 c 的 Poisson 分布来逼近和的分布.

例题 3.6.2. 生日问题. 在一个有 400 名学生的微积分课上，期末考试当天过生日的学生数近似服从均值为 $400/365 = 1.096$ 的 Poisson 分布. 这意味着没有人在该日期出生的概率约为 $e^{-1.096} = 0.334$. 类似推理表明，某天出生的婴儿数，或 1:15 至 1:30 之间到达银行的人数应服从 Poisson 分布.

例题 3.6.3. 假设将两颗骰子掷 36 次. 出现 “双一点”（每颗骰子均为 1 点）的概率为 $1/36$，故该事件发生的次数应近似服从均值为 1 的 Poisson 分布. 比较 Poisson 逼近与精确概率可见，即使试验次数较小，吻合也很好.

k	0	1	2	3
Poisson	0.3678	0.3678	0.1839	0.0613
精确	0.3627	0.3730	0.1865	0.0604

给出定理 3.6.1 的第二个证明后，将讨论收敛速度. 这些结果表明，对大 n，最大偏差出现在 $k = 1$ 处，约为 $1/2en$（在本例中为 0.0051）.

定理 3.6.1 的第二个证明需要更多的工作，但提供了关于收敛速度的信息. 首先定义可数集 S 上两个测度之间的**全变差距离**:

$$\|\mu - \nu\| \equiv \frac{1}{2}\sum_{z} |\mu(z) - \nu(z)| = \sup_{A \subset S} |\mu(A) - \nu(A)|.$$

第一个等式是定义. 为证明第二个等式，注意对任意 A

$$\sum_{z} |\mu(z) - \nu(z)| \geq |\mu(A) - \nu(A)| + |\mu(A^c) - \nu(A^c)| = 2|\mu(A) - \nu(A)|$$

且当 $A = \{z : \mu(z) \geq \nu(z)\}$ 时取等号.

引理 3.6.4. (i) $d(\mu,\nu) = \|\mu - \nu\|$ 定义了 $\mathbb{Z}$ 上概率测度的一个度量，
(ii) $\|\mu_n - \mu\| \to 0$ 当且仅当 $\mu_n(x) \to \mu(x)$ 对每个 $x \in \mathbb{Z}$ 成立，由习题 3.2.11 这等价于 $\mu_n \Rightarrow \mu$.

证明. (i) 显然 $d(\mu,\nu) = d(\nu,\mu)$，且 $d(\mu,\nu) = 0$ 当且仅当 $\mu = \nu$. 为验证三角不等式，注意实数的三角不等式蕴含

$$|\mu(x) - \nu(x)| + |\nu(x) - \pi(x)| \geq |\mu(x) - \pi(x)|$$

然后对 x 求和.

(ii) 一个方向是平凡的. 除非 $\mu_n(x) \to \mu(x)$ 对每个 x 成立，否则不能有 $\|\mu_n - \mu\| \to 0$. 为证反向，注意若 $\mu_n(x) \to \mu(x)$ 则

$$\sum_x |\mu_n(x) - \mu(x)| = 2\sum_x (\mu(x) - \mu_n(x))^+ \to 0$$

由控制收敛定理. □

习题 3.6.1. 证明 $\|\mu-\nu\| \le 2\delta$ 当且仅当存在具有分布 μ 和 ν 的随机变量 X 和 Y 使得 $P(X \ne Y) \le \delta$.

下面三个引理是第二个证明的关键.

引理 3.6.5. 若 $\mu_1 \times \mu_2$ 表示 $\mathbb{Z} \times \mathbb{Z}$ 上的乘积测度，满足 $(\mu_1 \times \mu_2)(x, y) = \mu_1(x)\mu_2(y)$，则

$$\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\| \le \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\|.$$

证明.

$$\begin{aligned}
2\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\| &= \sum_{x,y} |\mu_1(x)\mu_2(y) - \nu_1(x)\nu_2(y)| \\
&\le \sum_{x,y} |\mu_1(x)\mu_2(y) - \nu_1(x)\mu_2(y)| + \sum_{x,y} |\nu_1(x)\mu_2(y) - \nu_1(x)\nu_2(y)| \\
&= \sum_y \mu_2(y) \sum_x |\mu_1(x) - \nu_1(x)| + \sum_x \nu_1(x) \sum_y |\mu_2(y) - \nu_2(y)| \\
&= 2\|\mu_1 - \nu_1\| + 2\|\mu_2 - \nu_2\|
\end{aligned}$$

即得所需结果. □

引理 3.6.6. 若 $\mu_1 * \mu_2$ 表示 μ_1 与 μ_2 的**卷积**，即

$$\mu_1 * \mu_2(x) = \sum_y \mu_1(x - y)\mu_2(y),$$

则 $\|\mu_1 * \mu_2 - \nu_1 * \nu_2\| \le \|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|$.

证明.

$$\begin{aligned}2\|\mu_1*\mu_2-\nu_1*\nu_2\| &= \sum_x\left|\sum_y \mu_1(x-y)\mu_2(y)-\sum_y \nu_1(x-y)\nu_2(y)\right| \\ &\le \sum_x\sum_y |\mu_1(x-y)\mu_2(y)-\nu_1(x-y)\nu_2(y)| \\ &= 2\|\mu_1\times\mu_2-\nu_1\times\nu_2\|\end{aligned}$$

即得所需结果. □

引理 3.6.7. 设 μ 为满足 $\mu(1)=p$ 且 $\mu(0)=1-p$ 的测度. 设 ν 为均值为 p 的 Poisson 分布. 则 $\|\mu-\nu\|\le p^2$.

证明.

$$\begin{aligned}2\|\mu-\nu\| &= |\mu(0)-\nu(0)|+|\mu(1)-\nu(1)|+\sum_{n\ge 2}\nu(n) \\ &= |1-p-e^{-p}|+|p-pe^{-p}|+1-e^{-p}(1+p).\end{aligned}$$

由于 $1-x\le e^{-x}\le 1$ 对 $x\ge 0$ 成立，上式

$$\begin{aligned}&= e^{-p}-1+p+p(1-e^{-p})+1-e^{-p}-pe^{-p} \\ &= 2p(1-e^{-p})\le 2p^2\end{aligned}$$

即得所需结果. □

定理 3.6.1 的第二个证明. 设 $\mu_{n,m}$ 为 $X_{n,m}$ 的分布. 设 μ_n 为 S_n 的分布. 设 $\nu_{n,m}$，ν_n 和 ν 分别为均值 $p_{n,m}$，$\lambda_n=\sum_{m\le n}p_{n,m}$ 和 λ 的 Poisson 分布. 由于 $\mu_n=\mu_{n,1}*\cdots*\mu_{n,n}$ 且 $\nu_n=\nu_{n,1}*\cdots*\nu_{n,n}$，引理 3.6.6、3.6.5 和 3.6.7 蕴含

$$\|\mu_n-\nu_n\|\le\sum_{m=1}^n\|\mu_{n,m}-\nu_{n,m}\|\le 2\sum_{m=1}^n p_{n,m}^2. \tag{3.6.1}$$

利用全变差距离的定义，现在有

$$\sup_A|\mu_n(A)-\nu_n(A)|\le\sum_{m=1}^n p_{n,m}^2.$$

假设 (i) 和 (ii) 蕴含右边 $\to 0$. 由于 $\nu_n\Rightarrow\nu$（当 $n\to\infty$），结论得证. □

注释. 上述证明归功于 Hodges 和 Le Cam (1960). 利用不同方法，C. Stein (1987)（见第 89 页 (43)

式）证明了

$$\sup_A |\mu_n(A) - \nu_n(A)| \le (\lambda \vee 1)^{-1} \sum_{m=1}^{n} p_{n,m}^2.$$

收敛速度. 当 $p_{n,m} = 1/n$ 时，(3.6.1) 变为

$$\sup_A |\mu_n(A) - \nu_n(A)| \le 1/n.$$

为评估此界的质量，比较 k 次成功的 Poisson 概率与二项概率.

k	Poisson	二项
0	e^{-1}	$\left(1-\frac{1}{n}\right)^n$
1	e^{-1}	$n \cdot n^{-1}\left(1-\frac{1}{n}\right)^{n-1} = \left(1-\frac{1}{n}\right)^{n-1}$
2	$e^{-1}/2!$	$\binom{n}{2} n^{-2}\left(1-\frac{1}{n}\right)^{n-2} = \left(1-\frac{1}{n}\right)^{n-1}/2!$
3	$e^{-1}/3!$	$\binom{n}{3} n^{-3}\left(1-\frac{1}{n}\right)^{n-3} = \left(1-\frac{2}{n}\right)\left(1-\frac{1}{n}\right)^{n-2}/3!$

由于 $(1-x) \le e^{-x}$，有 $\mu_n(0) - \nu_n(0) \le 0$. 展开

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$

得

$$(n-1)\log\left(1-\frac{1}{n}\right) = -\frac{n-1}{n} - \frac{n-1}{2n^2} - \cdots = -1 + \frac{1}{2n} + O(n^{-2}).$$

故

$$n\left(1-\frac{1}{n}\right)^{n-1} - e^{-1} = ne^{-1}\exp\{1/2n + O(n^{-2})\} - 1 \to e^{-1}/2$$

且由此

$$n(\mu_n(1) - \nu_n(1)) \to e^{-1}/2, \quad n(\mu_n(2) - \nu_n(2)) \to e^{-1}/4.$$

对 $k \ge 3$，利用 $(1-2/n) \le (1-1/n)^2$ 和 $(1-x) \le e^{-x}$ 可见 $\mu_n(k) - \nu_n(k) \le 0$，故

$$\sup_{A \subset \mathbb{Z}} |\mu_n(A) - \nu_n(A)| \approx 3/4en.$$

关于相依事件的 Poisson 逼近有大量文献. 这里考虑

3.6.2 两个相依的例子

例题 3.6.8. 匹配问题. 设 π 为 $\{1, 2, \ldots, n\}$ 的一个随机排列，设 $X_{n,m} = 1$ 若 m 是一个**不动点**（否则为 0），并令 $S_n = X_{n,1} + \cdots + X_{n,n}$ 为不动点数. 计算 $P(S_n = 0)$. （为更有趣的故事，考虑男士

寄存帽子或妻子交换丈夫.）令 $A_{n,m}=\{X_{n,m}=1\}$. **容斥公式**蕴含

$$\begin{aligned}P\left(\bigcup_{m=1}^{n}A_m\right)&=\sum_m P(A_m)-\sum_{\ell<m}P(A_\ell\cap A_m)\\&\quad+\sum_{k<\ell<m}P(A_k\cap A_\ell\cap A_m)-\cdots\\&=n\cdot\frac{1}{n}-\binom{n}{2}\frac{(n-2)!}{n!}+\binom{n}{3}\frac{(n-3)!}{n!}-\cdots\end{aligned}$$

因为具有 k 个指定不动点的排列数为 $(n-k)!$. 约去一些阶乘得

$$P(S_n>0)=\sum_{m=1}^{n}\frac{(-1)^{m-1}}{m!}\quad\text{故}\quad P(S_n=0)=\sum_{m=0}^{n}\frac{(-1)^m}{m!}.$$

认出第二个和是 e^{-1} 展开式的前 $n+1$ 项，得

$$|P(S_n=0)-e^{-1}|=\left|\sum_{m=n+1}^{\infty}\frac{(-1)^m}{m!}\right|\le\frac{1}{(n+1)!}\left|\sum_{k=0}^{\infty}(n+2)^{-k}\right|=\frac{1}{(n+1)!}\cdot\left(1-\frac{1}{n+2}\right)^{-1}$$

收敛速度远优于 $1/n$. 为计算其他概率，注意通过考虑不动点的位置

$$\begin{aligned}P(S_n=k)&=\binom{n}{k}\frac{1}{n(n-1)\cdots(n-k+1)}P(S_{n-k}=0)\\&=\frac{1}{k!}P(S_{n-k}=0)\to e^{-1}/k!\end{aligned}$$

例题 3.6.9. 占位问题. 设将 r 个球随机放入 n 个盒子. 由二项分布的 Poisson 逼近可知，若 $n\to\infty$ 且 $r/n\to c$，则给定盒子中的球数将趋于均值为 c 的 Poisson 分布. 最后这一观察应能解释为何例 2.2.10 中空盒子的比例趋于 e^{-c}. 这里将证明：

定理 3.6.10. 若 $ne^{-r/n}\to\lambda\in[0,\infty)$，则空盒子数趋于均值为 λ 的 Poisson 分布.

证明. 为看出答案从何而来，注意在 Poisson 逼近中，给定盒子为空的概率是 $e^{-r/n}\approx\lambda/n$，故若各盒子的占用相互独立，结论将由定理 3.6.1 推出. 为证明结果，首先注意

$$P(\text{盒子}i_1,i_2,\ldots,i_k\text{ 为空})=\left(1-\frac{k}{n}\right)^r.$$

若令 $p_m(r,n)$ 为当 r 个球放入 n 个盒子时恰有 m 个盒子为空的概率，则

$$P(\text{无空盒})=1-P(\text{至少一个空盒}).$$

故由容斥公式

$$p_0(r,n)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\left(1-\frac{k}{n}\right)^r. \tag{a}$$

通过考虑空盒的位置

$$p_m(r,n)=\binom{n}{m}\left(1-\frac{m}{n}\right)^r p_0(r,n-m). \tag{b}$$

为求 $p_m(r,n)$ 的极限，首先证明若 $ne^{-r/n}\to\lambda$ 则

$$\binom{n}{m}\left(1-\frac{m}{n}\right)^r\to\lambda^m/m! \tag{c}$$

这一半是容易的. 由于 $(1-x)\le e^{-x}$ 且 $ne^{-r/n}\to\lambda$

$$\binom{n}{m}\left(1-\frac{m}{n}\right)^r\le\frac{n^m}{m!}e^{-mr/n}\to\lambda^m/m! \tag{d}$$

对另一方向，注意

$$\binom{n}{m}\ge(n-m)^m/m!\quad\text{故}$$

$$\binom{n}{m}\left(1-\frac{m}{n}\right)^r\ge\left(1-\frac{m}{n}\right)^{m+r}n^m/m!$$

现 $(1-m/n)^m\to1$（当 $n\to\infty$）且 $1/m!$ 是常数. 为处理其余部分，注意若 $0\le t\le1/2$ 则

$$\log(1-t)=-t-t^2/2-t^3/3\cdots\ge-t-\frac{t^2}{2}(1+2^{-1}+2^{-2}+\cdots)=-t-t^2$$

故有

$$\log\left(n^m\left(1-\frac{m}{n}\right)^r\right)\ge m\log n-rm/n-r(m/n)^2.$$

假设 $ne^{-r/n}\to\lambda$ 意味着

$$r=n\log n-n\log\lambda+o(n)$$

故 $r(m/n)^2\to0$. 将最后显示式乘以 m/n 并重排得 $m\log n-rm/n\to m\log\lambda$. 结合最后两个结果可见

$$\liminf_{n\to\infty}n^m\left(1-\frac{m}{n}\right)^r\ge\lambda^m$$

且 (c) 得证. 由 (a)、(c) 和控制收敛定理（利用 (d) 得到控制），得

$$若ne^{-r/n}\to\lambda\ 则p_0(r,n)\to\sum_{k=0}^{\infty}\frac{(-1)^k\lambda^k}{k!}=e^{-\lambda}. \tag{e}$$

对固定的 m，$(n-m)e^{-r/(n-m)} \to \lambda$，故由 (e) 知 $p_0(r, n-m) \to e^{-\lambda}$. 结合 (b) 和 (c) 完成证明. □

例题 3.6.11. 优惠券收集问题. 设 $X_1, X_2, \ldots$ 为 $\{1, 2, \ldots, n\}$ 上的独立同分布均匀变量，并令 $T_n = \inf\{m : \{X_1, \ldots, X_m\} = \{1, 2, \ldots, n\}\}$. 由于 $T_n \le m$ 当且仅当 m 个球填满所有 n 个盒子，由定理 3.6.10 得

$$P(T_n - n\log n \le nx) \to \exp(-e^{-x}).$$

证明. 若 $r = n\log n + nx$，则 $ne^{-r/n} \to e^{-x}$. □

注意 T_n 是 n 个独立随机变量之和（见例 2.2.7），但 T_n 不收敛于正态分布. 问题在于和中最后几项的量级为 n，故 Lindeberg-Feller 定理的条件不满足.

对前一结果的具体实例，考虑：2190（$= 6 \cdot 365$）人的村庄中，所有生日都被覆盖的概率是多少？认为 1825（$= 5 \cdot 365$）人的答案会有很大不同吗?

解答. 这里 $n = 365$，故 $365\log 365 = 2153$ 且

$$\begin{aligned} P(T_{365} \le 2190) &= P((T_{365} - 2153)/365 \le 37/365) \\ &\approx \exp(-e^{-0.1014}) = \exp(-0.9036) = 0.4051 \\ P(T_{365} \le 1825) &= P((T_{365} - 2153)/365 \le -328/365) \\ &\approx \exp(-e^{0.8986}) = \exp(-2.4562) = 0.085 \end{aligned}$$

第 7 节 Poisson 过程

由定理 3.6.1 可直接推广得到如下结果.

定理 3.7.1. 设 $X_{n,m}, 1 \le m \le n$ 为独立的非负整数值随机变量, 满足 $P(X_{n,m}=1)=p_{n,m}, P(X_{n,m} \ge 2)=\varepsilon_{n,m}$.

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0,\infty)$,

(ii) $\max_{1\le m\le n} p_{n,m} \to 0$,

(iii) $\sum_{m=1}^{n} \varepsilon_{n,m} \to 0$.

若 $S_n = X_{n,1}+\cdots+X_{n,n}$，则 $S_n \Rightarrow Z$，其中 Z 服从 Poisson(λ) 分布.

证明. 记 $X'_{n,m}=1$（若 $X_{n,m}=1$），否则为 0. 又记 $S'_n = X'_{n,1}+\cdots+X'_{n,n}$. 由 (i)-(ii) 和定理 3.6.1 知 $S'_n \Rightarrow Z$，而 (iii) 表明 $P(S_n \ne S'_n) \to 0$，再由习题 3.2.13 的收敛引理即得结果. □

下面的结果利用定理 3.7.1 解释了 Poisson 分布在应用中频繁出现的原因. 设 $N(s,t)$ 为银行或冰淇淋店在时间区间 $(s,t]$ 内的到达人数. 假设

(i) 不相交区间内的到达人数相互独立,

(ii) $N(s,t)$ 的分布仅依赖于 $t-s$,

(iii) $P(N(0,h)=1)=\lambda h+o(h)$,

(iv) $P(N(0,h)\ge 2)=o(h)$.

其中，两个 $o(h)$ 表示满足 $g_i(h)/h \to 0$（当 $h\to 0$ 时）的函数 $g_1(h)$ 和 $g_2(h)$.

定理 3.7.2. 若 (i)-(iv) 成立，则 $N(0,t)$ 服从均值为 λt 的 Poisson 分布.

证明. 令 $X_{n,m}=N((m-1)t/n, mt/n)$（$1\le m\le n$）并应用定理 3.7.1. □

满足以下条件的随机变量族 N_t，$t\ge 0$ 称为**速率为 λ 的 Poisson 过程**:

(i) 若 $0=t_0<t_1<\cdots<t_n$，则 $N(t_k)-N(t_{k-1})$，$1\le k\le n$ 相互独立,

(ii) $N(t)-N(s)$ 服从 Poisson($\lambda(t-s)$) 分布.

为理解 N_t 的行为，另一种构造方法颇有裨益. 设 $\xi_1,\xi_2,\ldots$ 为独立随机变量，满足 $P(\xi_i>t)=e^{-\lambda t}$（$t\ge 0$）. 令 $T_n=\xi_1+\cdots+\xi_n$ 且 $N_t=\sup\{n: T_n\le t\}$，其中 $T_0=0$. 用更新理论的语言（见定理 2.4.7），T_n 是第 n 次到达的时刻，N_t 是到时刻 t 为止的到达次数. 为验证 N_t 是 Poisson 过程，首先回顾（见定理 2.1.18）:

$$f_{T_n}(s)=\frac{\lambda^n s^{n-1}}{(n-1)!}e^{-\lambda s} \quad 对 s\ge 0$$

即 T_n 的分布密度由上式右端给出.

现

$$P(N_t=0)=P(T_1>t)=e^{-\lambda t}$$

且对 $n \geq 1$

$$P(N_t = n) = P(T_n \leq t < T_{n+1}) = \int_0^t P(T_n = s)P(\xi_{n+1} > t - s)\,\mathrm{d}s$$
$$= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{-\lambda(t-s)}\,\mathrm{d}s = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

最后两个公式表明 N_t 服从均值为 λt 的 Poisson 分布. 为验证不相交区间内的到达次数相互独立，注意

$$P(T_{n+1} \geq u \mid N_t = n) = P(T_{n+1} \geq u, T_n \leq t)/P(N_t = n)$$

为计算分子，注意

$$P(T_{n+1} \geq u, T_n \leq t) = \int_0^t f_{T_n}(s)P(\xi_{n+1} \geq u - s)\,\mathrm{d}s$$
$$= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{-\lambda(u-s)}\,\mathrm{d}s = e^{-\lambda u}\frac{(\lambda t)^n}{n!}.$$

分母为 $P(N_t = n) = e^{-\lambda t}(\lambda t)^n/n!$，故

$$P(T_{n+1} \geq u \mid N_t = n) = e^{-\lambda u}/e^{-\lambda t} = e^{-\lambda(u-t)}$$

或改写为 $P(T_{n+1} - t \geq s \mid N_t = n) = e^{-\lambda s}$. 令 $T_1' = T_{N(t)+1} - t$, 且对 $k \geq 2$ 令 $T_k' = T_{N(t)+k} - T_{N(t)+k-1}$. 最后的计算表明

(a) $T_1', T_2', \ldots$ 独立同分布且与 N_t 独立.

上述观察表明，时刻 t 之后的到达与 N_t 独立且具有与原序列相同的分布. 由此易见：

(b) 若 $0 = t_0 < t_1 < \cdots < t_n$，则 $N(t_i) - N(t_{i-1})$，$i = 1, \ldots, n$ 相互独立.

为看出这一点, 注意向量 $(N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1}))$ 是 $\sigma(T_k', k \geq 1)$ 可测的, 因而与 $N(t_1)$ 独立. 然后用归纳法可得

$$P(N(t_i) - N(t_{i-1}) = k_i, i = 1, \ldots, n) = \prod_{i=1}^n \exp(-\lambda(t_i - t_{i-1}))\frac{\lambda(t_i - t_{i-1})^{k_i}}{k_i!}.$$

注释. 证明 (a) 的关键是指数分布的**无记忆性**：

$$P(T > t + s \mid T > t) = P(T > s) \tag{3.7.1}$$

它表明时刻 t 之后第一次到达的位置与时刻 t 之前发生的事情独立且具有指数分布.

习题 3.7.1. 证明若 $P(T > 0) = 1$ 且 (3.7.1) 成立，则存在 $\lambda > 0$ 使得 $P(T > t) = e^{-\lambda t}$ （$t \geq 0$）. 提

示：首先证明这对 $t = m2^{-n}$ 成立.

习题 3.7.2. 若 $S = \text{exponential}(\lambda)$ 且 $T = \text{exponential}(\mu)$ 相互独立，则 $V = \min\{S, T\}$ 服从 exponential$(\lambda + \mu)$ 分布且 $P(U = S) = \lambda/(\lambda + \mu)$.

习题 3.7.3. 设 $T_i = \text{exponential}(\lambda_i)$. 令 $V = \min(T_1, \ldots, T_n)$，$I$ 为 T_i 中最小的那个的（随机）下标.

$$P(V > t) = \exp(-(\lambda_1 + \cdots + \lambda_n)t),$$
$$P(I = i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}.$$

I 与 $V = \min\{T_1, \ldots, T_n\}$ 独立.

3.7.1 复合 Poisson 过程

假设在 12:00 到 1:00 之间，汽车按照速率为 λ 的 Poisson 过程 $N(t)$ 到达一家快餐店. 设 Y_i 为第 i 辆车中的人数，假设 Y_i 独立同分布且与 $N(t)$ 独立. 引入 Y_i 后，自然要考虑时刻 t 为止所见到的 Y_i 之和：

$$S(t) = Y_1 + \cdots + Y_{N(t)}$$

其中当 $N(t) = 0$ 时设 $S(t) = 0$. 在上述例子中，$S(t)$ 给出到时刻 t 为止到达的顾客人数.

定理 3.7.3. 设 $Y_1, Y_2, \ldots$ 独立同分布，N 为与之独立的非负整数值随机变量，令 $S = Y_1 + \cdots + Y_N$（当 $N = 0$ 时 $S = 0$）.

(i) 若 $E|Y_i|, EN < \infty$，则 $ES = EN \cdot EY_i$.

(ii) 若 $EY_i^2, EN^2 < \infty$，则 $\text{var}(S) = EN \cdot \text{var}(Y_i) + \text{var}(N) \cdot (EY_i)^2$.

(iii) 若 N 服从 Poisson(λ) 分布，则 $\text{var}(S) = \lambda EY_i^2$.

其合理性如下. 第一条性质是自然的，因为若 $N = n$ 为非随机的，则 $ES = nEY_i$. (i) 通过令 $n = EN$ 得到. 这一事实称为 **Wald 等式**. (ii) 中的公式更复杂，但它显然具有两个必要的性质：

- 若 $N = n$ 为非随机的，$\text{var}(S) = n \cdot \text{var}(Y_i)$.
- 若 $Y_i = c$ 为非随机的，$\text{var}(S) = c^2 \cdot \text{var}(N)$.

结合这两个观察，可见 $EN \cdot \text{var}(Y_i)$ 是来自 Y_i 变异性的方差贡献，而 $\text{var}(N) \cdot (EY_i)^2$ 是来自 N 变异性的贡献.

证明. 当 $N = n$ 时，$S = X_1 + \cdots + X_n$ 具有 $ES = nEY_i$. 按 N 的取值分解，

$$ES = \sum_{n=0}^{\infty} E(S \mid N = n) \cdot P(N = n) = \sum_{n=0}^{\infty} nEY_i \cdot P(N = n) = EN \cdot EY_i.$$

对第二个公式，注意当 $N=n$ 时，$S=X_1+\cdots+X_n$ 具有 $\text{var}(S)=n\cdot\text{var}(Y_i)$，故

$$E(S^2\mid N=n)=n\cdot\text{var}(Y_i)+(nEY_i)^2.$$

类似计算得

$$\begin{aligned}ES^2&=\sum_{n=0}^{\infty}E(S^2\mid N=n)\cdot P(N=n)\\&=\sum_{n=0}^{\infty}\{n\cdot\text{var}(Y_i)+n^2(EY_i)^2\}\cdot P(N=n)\\&=(EN)\cdot\text{var}(Y_i)+EN^2\cdot(EY_i)^2.\end{aligned}$$

为计算方差，注意

$$\begin{aligned}\text{var}(S)&=ES^2-(ES)^2\\&=(EN)\cdot\text{var}(Y_i)+EN^2\cdot(EY_i)^2-(EN\cdot EY_i)^2\\&=(EN)\cdot\text{var}(Y_i)+\text{var}(N)\cdot(EY_i)^2\end{aligned}$$

其中最后一步利用了 $\text{var}(N)=EN^2-(EN)^2$ 合并第二和第三项.

对 (iii)，注意在 Poisson 的特殊情形，$EN=\lambda$ 且 $\text{var}(N)=\lambda$，故由 $\text{var}(Y_i)+(EY_i)^2=EY_i^2$ 可得结果. □

3.7.2 稀疏

在上一小节中，将 Poisson 过程中到达所关联的 Y_i 相加，以观察时刻 t 为止积累了多少顾客等. 在本小节中，将用 Y_i 将 Poisson 过程分裂成若干个，即令 $N_j(t)$ 为满足 $Y_i=j$ 的 $i\le N(t)$ 的个数. 一个有些令人惊讶的事实是：

定理 3.7.4. $N_j(t)$ 是独立的速率为 $\lambda P(Y_i=j)$ 的 Poisson 过程.

这令人惊讶的原因有二：(i) 所得过程是 Poisson 的；(ii) 它们是相互独立的. 为强调这一点，考虑一个速率为每小时 10 人的 Poisson 过程，然后抛硬币决定到达的顾客是男性还是女性. 有人可能认为一小时内到达 40 名男性意味着生意兴隆，因此女性人数也会多于正常水平，但定理 3.7.4 表明，每小时到达的男性和女性人数是相互独立的.

证明. 首先假设 $P(Y_i=1)=p$ 且 $P(Y_i=2)=1-p$，因此只需考虑两个 Poisson 过程：$N_1(t)$ 和 $N_2(t)$. Poisson 过程的独立增量性质显然意味着增量对

$$(N_1(t_i)-N_1(t_{i-1}),N_2(t_i)-N_2(t_{i-1})),\quad 1\le i\le n$$

相互独立. 由于 $N_1(0)=N_2(0)=0$ 按定义成立，只需验证分量 $X_i=N_i(t+s)-N_i(s)$ 相互独立且具有正确的 Poisson 分布. 为此，注意若 $X_1=j$ 且 $X_2=k$，则在 s 和 $s+t$ 之间必有 $j+k$ 次到达，其中 j 个被赋予 1，k 个被赋予 2，故

$$\begin{aligned}P(X_1=j,X_2=k)&=e^{-\lambda t}\frac{(\lambda t)^{j+k}}{(j+k)!}\cdot\frac{(j+k)!}{j!k!}p^j(1-p)^k\\&=e^{-\lambda pt}\frac{(\lambda pt)^j}{j!}\cdot e^{-\lambda(1-p)t}\frac{(\lambda(1-p)t)^k}{k!}\end{aligned}\tag{3.7.2}$$

因此 $X_1=\text{Poisson}(\lambda pt)$ 且 $X_2=\text{Poisson}(\lambda(1-p)t)$. 对一般情形，利用多项式可得，若 $p_j=P(Y_i=j)$ $(1\leq j\leq m)$，则

$$\begin{aligned}&P(X_1=k_1,\ldots,X_m=k_m)\\&=e^{-\lambda t}\frac{(\lambda t)^{k_1+\cdots+k_m}}{(k_1+\cdots+k_m)!}\cdot\frac{(k_1+\cdots+k_m)!}{k_1!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m}\\&=\prod_{j=1}^m e^{-\lambda p_j t}\frac{(\lambda p_j)^{k_j}}{k_j!}\end{aligned}$$

这就证明了所需结果. □

稀疏结果容易推广到非齐次情形：

定理 3.7.5. 假设在速率为 λ 的 Poisson 过程中，以概率 $p(s)$ 保留落在 s 处的点. 则结果是一个速率为 $\lambda p(s)$ 的**非齐次 Poisson 过程**.

对此的一个应用是考虑

习题 3.7.4. M/G/∞ 排队系统. 走在 Duke 校园里，似乎每个学生都在用智能手机通话. ATM 机到达的论证表明，通话开始遵循 Poisson 过程. 至于通话本身，虽然许多打电话的人表现出缺乏记忆性，但没有理由假设通话时长服从指数分布，因此用一般的分布函数 G（满足 $G(0)=0$ 且均值为 μ）. 证明在长期运行中，系统中的通话数将是 Poisson 分布，均值为

$$\int_{r=0}^{\infty}\lambda(1-G(r))\,\mathrm{d}r=\lambda\mu.\tag{3.7.3}$$

例题 3.7.6. Poisson 化与占位问题. 若将均值为 r 的 Poisson 数个球放入 n 个盒子中，并令 N_i 为第 i 个盒子中的球数，则上一习题意味着 $N_1,\ldots,N_n$ 相互独立且服从均值为 r/n 的 Poisson 分布. 利用这一观察来证明定理 3.6.10.

为获得固定球数 n 的结果，令 $r=n\log n-(\log\lambda)n+o(n)$ 且 $s_i=n\log n-(\log\mu_i)n$，其中 $\mu_2<\lambda<\mu_1$. Poisson 的正态近似表明 $P(\text{Poisson}(s_1)<r<\text{Poisson}(s_2))\to 1$（当 $n\to\infty$）. 若球

数服从均值为 $s = n\log n - n(\log\mu)$ 的 Poisson 分布，则第 i 个盒子中的球数 N_i 相互独立，均值为 $s/n = \log(n/\mu)$，因此它们以概率 $\exp(-s/n) = \mu/n$ 为空. 令 $X_{n,i} = 1$（若第 i 个盒子为空），否则为 0，并利用定理 3.6.1，可知空盒子数收敛于均值为 μ 的 Poisson 分布.

为证明固定球数的结果，注意中心极限定理意味着

$$P(\text{Poisson}(s_1) < r < \text{Poisson}(s_2)) \to 1.$$

由于空盒子数随球数增加而减少，故得所需结果.

习题 3.7.5. 假设均值为 2190 的 Poisson 数个 Duke 学生将出席观看下一场女子篮球比赛. 问对于全部 365 天，每一天观众中至少有一个人过生日的概率是多少.（假设所有生日概率相等且不存在 2 月 29 日.）

有时，例如在 3.8 节中，以下推广颇有裨益.

例题 3.7.7. 测度空间上的 Poisson 过程. 测度空间 $(S, \mathcal{S}, \mu)$ 上的 **Poisson 过程**是一个随机映射 $m : \mathcal{S} \to \{0, 1, \ldots\}$，对每个 ω 是 S 上的测度且具有以下性质：若 $A_1, \ldots, A_n$ 为满足 $\mu(A_i) < \infty$ 的不交集合，则 $m(A_1), \ldots, m(A_n)$ 相互独立且服从均值为 $\mu(A_i)$ 的 Poisson 分布. μ 称为该过程的**均值测度**.

若 $\mu(S) < \infty$，则由定理 3.7.4 知可用如下方法构造 m：令 $X_1, X_2, \ldots$ 为服从分布 $\nu(\cdot) = \mu(\cdot)/\mu(S)$ 的独立同分布 S 中元素，令 N 为与之独立的均值为 $\mu(S)$ 的 Poisson 随机变量，并令 $m(A) = |\{j \le N : X_j \in A\}|$. 为将构造推广到无限测度空间，例如 $S = \mathbb{R}^d$，$\mathcal{S}$ 为 Borel 集，μ 为 Lebesgue 测度，将空间划分为有限测度的不交集并在每个集上放置独立的 Poisson 过程.

下一个例子是正方形上的 Poisson 过程：

例题 3.7.8. 二战期间伦敦南部的飞弹袭击服从 Poisson 分布. 如 Feller, Vol. I (1968), p.160–161 所报告，该区域被划分为 576 个面积为 $1/4$ 平方公里的格子. 总命中数为 537，平均每格 0.9323. 下表比较了命中 k 次的格子数 N_k 与 Poisson 近似的预测值.

k	0	1	2	3	4	≥ 5
N_k	229	211	93	35	7	1
Poisson	226.74	211.39	98.54	30.62	7.14	1.57

3.7.3 条件化

定理 3.7.9. 设 T_n 为速率 λ 的 Poisson 过程中第 n 次到达的时刻. 设 $U_1, U_2, \ldots, U_n$ 为 $(0, t)$ 上独立均匀分布，令 V_k^n 为 $\{U_1, \ldots, U_n\}$ 中第 k 小的数. 若在 $N(t) = n$ 上条件化，则向量 $V = (V_1^n, \ldots, V_n^n)$ 与 $T = (T_1, \ldots, T_n)$ 具有相同分布.

证明. 在 $\{N(t)=n\}$ 上，联合密度

$$f_T(t_1,\dots,t_n)=\prod_{m=1}^{n}\lambda e^{-\lambda(t_m-t_{m-1})}\cdot\lambda e^{-\lambda(t-t_n)}=\lambda^n e^{-\lambda t}.$$

若除以 $P(N(t)=n)=e^{-\lambda t}(\lambda t)^n/n!$，结果为 $n!/t^n$，这正是 V 的联合分布. □

推论 3.7.10. 若 $0<s<t$ 则

$$P(N(s)=m\mid N(t)=n)=\binom{n}{m}\left(\frac{s}{t}\right)^m\left(1-\frac{s}{t}\right)^{n-m}.$$

证明. 一旦注意到集合 $\{U_1,\dots,U_n\}$ 与 $\{V_1,\dots,V_n\}$ 具有相同分布，这就是定理 3.7.9 的直接推论. □

以下是定理 3.7.9 的一个变体.

定理 3.7.11. 设 T_n 为速率 λ 的 Poisson 过程中第 n 次到达的时刻. 设 $U_1,U_2,\dots,U_n$ 为 $(0,1)$ 上独立均匀分布，令 V_k^n 为 $\{U_1,\dots,U_n\}$ 中第 k 小的数. 则向量 $(V_1^n,\dots,V_n^n)$ 与 $(T_1/T_{n+1},\dots,T_n/T_{n+1})$ 具有相同分布.

证明. 作变量替换 $v=r(t)$，其中 $v_i=t_i/t_{n+1}$ $(i\le n)$，$v_{n+1}=t_{n+1}$. 逆函数为

$$s(v)=(v_1v_{n+1},\dots,v_nv_{n+1},v_{n+1})$$

其偏导数矩阵 $\partial s_i/\partial v_j$ 为

$$\begin{pmatrix} v_{n+1} & 0 & \cdots & 0 & v_1\\ 0 & v_{n+1} & \cdots & 0 & v_2\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & v_{n+1} & v_n\\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

该矩阵的行列式为 v_{n+1}^n，故若令 $W=(V_1,\dots,V_{n+1})=r(T_1,\dots,T_{n+1})$，则变量替换公式表明 W 具有联合密度

$$f_W(v_1,\dots,v_n,v_{n+1})=\prod_{m=1}^{n}\lambda e^{-\lambda v_{n+1}(v_m-v_{m-1})}\cdot\lambda e^{-\lambda v_{n+1}(1-v_n)}\cdot v_{n+1}^n.$$

为求 $V=(V_1,\dots,V_n)$ 的联合密度，简化上述公式并对最后一个坐标积分得

$$f_V(v_1,\dots,v_n)=\int_0^\infty\lambda^{n+1}v_{n+1}^n e^{-\lambda v_{n+1}}\,\mathrm{d}v_{n+1}=n!$$

对 $0<v_1<v_2<\cdots<v_n<1$，这正是所需的联合密度. □

间隔. 最后一个结果可用于研究独立同分布均匀分布顺序统计量之间的间隔. 在接下来的四个习题中，使用定理 3.7.11 的记号，取 $\lambda = 1$ 并令 $V_0^n = 0$，$V_{n+1}^n = 1$.

习题 3.7.6. Smirnov (1949) $nV_k^n \Rightarrow T_k$.

习题 3.7.7. Weiss (1955) $n^{-1}\sum_{m=1}^{n} 1_{(n(V_m^n - V_{m-1}^n) > x)} \to e^{-x}$ 依概率.

习题 3.7.8. $(n/\log n)\max_{1\le m\le n+1}(V_m^n - V_{m-1}^n) \to 1$ 依概率.

习题 3.7.9. $P(n^2 \min_{1\le m\le n}(V_m^n - V_{m-1}^n) > x) \to e^{-x}$.

第 8 节 稳定分布 *

设 $X_1, X_2, \ldots$ 为独立同分布随机变量且 $S_n = X_1 + \cdots + X_n$. 定理 3.4.1 表明, 若 $EX_i = \mu$ 且 $\text{var}(X_i) = \sigma^2 \in (0, \infty)$, 则

$$(S_n - n\mu)/\sigma n^{1/2} \Rightarrow \chi$$

本节将研究 $EX_1^2 = \infty$ 的情形，并给出存在常数 a_n 和 b_n 使得

$$(S_n - b_n)/a_n \Rightarrow Y \quad \text{其中 } Y \text{ 为非退化的}$$

的充要条件.

先看一个例子. 设 X_i 的分布满足

$$P(X_1 > x) = P(X_1 < -x) = x^{-\alpha}/2 \quad \text{对 } x \geq 1 \tag{3.8.1}$$

其中 $0 < \alpha < 2$. 若 $\varphi(t) = E\exp(itX_1)$，则

$$\begin{aligned} 1 - \varphi(t) &= \int_1^\infty (1 - e^{itx}) \frac{\alpha}{2|x|^{\alpha+1}} \, \mathrm{d}x + \int_{-\infty}^{-1} (1 - e^{itx}) \frac{\alpha}{2|x|^{\alpha+1}} \, \mathrm{d}x \\ &= \alpha \int_1^\infty \frac{1 - \cos(tx)}{x^{\alpha+1}} \, \mathrm{d}x \end{aligned}$$

作变量替换 $tx = u$，$\mathrm{d}x = \mathrm{d}u/t$，最后一个积分变为

$$= \alpha \int_t^\infty \frac{1 - \cos u}{(u/t)^{\alpha+1}} \frac{\mathrm{d}u}{t} = \frac{\alpha}{t^\alpha} \int_t^\infty \frac{1 - \cos u}{u^{\alpha+1}} \, \mathrm{d}u$$

当 $u \to 0$ 时, $1 - \cos u \sim u^2/2$, 故 $(1 - \cos u)/u^{\alpha+1} \sim u^{-\alpha+1}/2$ 可积, 因为 $\alpha < 2$ 蕴含 $-\alpha + 1 > -1$. 令

$$C = \alpha \int_0^\infty \frac{1 - \cos u}{u^{\alpha+1}} \, \mathrm{d}u < \infty$$

并注意 (3.8.1) 蕴含 $\varphi(t) = \varphi(-t)$, 则上述结果表明

$$1 - \varphi(t) \sim C|t|^\alpha \quad \text{当 } t \to 0 \tag{3.8.2}$$

设 $X_1, X_2, \ldots$ 为具有 (3.8.1) 所给分布的独立同分布随机变量, 并记 $S_n = X_1 + \cdots + X_n$.

$$E\exp(itS_n/n^{1/\alpha}) = \varphi(t/n^{1/\alpha})^n = (1 - \{1 - \varphi(t/n^{1/\alpha})\})^n$$

当 $n \to \infty$ 时, $n(1-\varphi(t/n^{1/\alpha})) \to C|t|^\alpha$, 故由定理 3.4.2 得

$$E\exp(itS_n/n^{1/\alpha}) \to \exp(-C|t|^\alpha)$$

由定理 3.3.17(ii), 右边的表达式是某随机变量 Y 的特征函数, 且

$$S_n/n^{1/\alpha} \Rightarrow Y \tag{3.8.3}$$

为给出一般结果做准备, 下面提供 (3.8.3) 的另一证明. 若 $0 < a < b$ 且 $an^{1/\alpha} > 1$, 则

$$P(an^{1/\alpha} < X_1 < bn^{1/\alpha}) = \frac{1}{2}(a^{-\alpha} - b^{-\alpha})n^{-1}$$

故由定理 3.6.1 得

$$N_n(a,b) \equiv |\{m \le n : X_m/n^{1/\alpha} \in (a,b)\}| \Rightarrow N(a,b)$$

其中 $N(a,b)$ 服从均值为 $(a^{-\alpha} - b^{-\alpha})/2$ 的 Poisson 分布. 该结果的一个简单推广表明, 若 $A \subset \mathbb{R} - (-\delta,\delta)$ 且 $\delta n^{1/\alpha} > 1$, 则

$$P(X_1/n^{1/\alpha} \in A) = n^{-1}\int_A \frac{\alpha}{2|x|^{\alpha+1}}\,\mathrm{d}x$$

故 $N_n(A) \equiv |\{m \le n : X_m/n^{1/\alpha} \in A\}| \Rightarrow N(A)$, 其中 $N(A)$ 服从均值为

$$\mu(A) = \int_A \frac{\alpha}{2|x|^{\alpha+1}}\,\mathrm{d}x < \infty$$

的 Poisson 分布.

称极限随机变量族 $N(A)$ 为 $(-\infty,\infty)$ 上具有均值测度 μ 的 Poisson 过程 (更多关于此过程的内容见例 3.7.7). 注意对任意 $\varepsilon > 0$, $\mu(\varepsilon,\infty) = \varepsilon^{-\alpha}/2 < \infty$, 故 $N(\varepsilon,\infty) < \infty$.

上一段描述了随机集

$$\mathcal{X}_n = \{X_m/n^{1/\alpha} : 1 \le m \le n\}$$

的极限行为. 为描述 $S_n/n^{1/\alpha}$ 的极限, 下面 "对这些点求和". 取 $\varepsilon > 0$ 且

$$I_n(\varepsilon) = \{m \le n : |X_m| > \varepsilon n^{1/\alpha}\}$$
$$\hat{S}_n(\varepsilon) = \sum_{m\in I_n(\varepsilon)} X_m \quad \bar{S}_n(\varepsilon) = S_n - \hat{S}_n(\varepsilon)$$

$I_n(\varepsilon)$ 为 "大项" 的指标集, 即绝对值大于 $\varepsilon n^{1/\alpha}$ 的项的指标集. $\hat{S}_n(\varepsilon)$ 是大项之和, 而 $\bar{S}_n(\varepsilon)$ 是其余部

分. 首先证明当 ε 充分小时 $\bar{S}_n(\varepsilon)$ 的贡献可忽略. 令

$$\bar{X}_m(\varepsilon) = X_m 1(|X_m| \le \varepsilon n^{1/\alpha})$$

对称性蕴含 $E\bar{X}_m(\varepsilon) = 0$, 故 $E(\bar{S}_n(\varepsilon)^2) = nE\bar{X}_1(\varepsilon)^2$.

$$\begin{aligned} E\bar{X}_1(\varepsilon)^2 &= \int_0^\infty 2yP(|\bar{X}_1(\varepsilon)| > y)\,\mathrm{d}y \le \int_0^1 2y\,\mathrm{d}y + \int_1^{\varepsilon n^{1/\alpha}} 2y\cdot y^{-\alpha}\,\mathrm{d}y \\ &= 1 + \frac{2}{2-\alpha}\varepsilon^{2-\alpha}n^{2/\alpha-1} - \frac{1}{2-\alpha} \le \frac{2\varepsilon^{2-\alpha}}{2-\alpha}n^{2/\alpha-1} \end{aligned}$$

其中在计算积分时用到了 $\alpha < 2$, 在最后的不等式中用到了 $\alpha > 0$. 由此得

$$E(\bar{S}_n(\varepsilon)/n^{1/\alpha})^2 \le \frac{2\varepsilon^{2-\alpha}}{2-\alpha} \tag{3.8.4}$$

为计算 $\hat{S}_n(\varepsilon)/n^{1/\alpha}$ 的极限, 注意 $|I_n(\varepsilon)|$ 服从成功概率为 $p = \varepsilon^{-\alpha}/n$ 的二项分布. 给定 $|I_n(\varepsilon)| = m$, $\hat{S}_n(\varepsilon)/n^{1/\alpha}$ 是 m 个独立随机变量之和, 其分布 F_n^ε 关于 0 对称且满足

$$1 - F_n^\varepsilon(x) = P(X_1/n^{1/\alpha} > x \mid |X_1|/n^{1/\alpha} > \varepsilon) = x^{-\alpha}/2\varepsilon^{-\alpha} \quad \text{对 } x \ge \varepsilon$$

最后一个分布与 εX_1 的分布相同, 故若 $\varphi(t) = E\exp(itX_1)$, 则分布 F_n^ε 的特征函数为 $\varphi(\varepsilon t)$. 综合本段的观察得

$$E\exp(it\hat{S}_n(\varepsilon)/n^{1/\alpha}) = \sum_{m=0}^{n}\binom{n}{m}(\varepsilon^{-\alpha}/n)^m(1-\varepsilon^{-\alpha}/n)^{n-m}\varphi(\varepsilon t)^m$$

由

$$\binom{n}{m}\frac{1}{n^m} = \frac{1}{m!}\frac{n(n-1)\cdots(n-m+1)}{n^m} \le \frac{1}{m!}$$

并注意 $(1-\varepsilon^{-\alpha}/n)^n \le \exp(-\varepsilon^{-\alpha})$, 利用控制收敛定理

$$\begin{aligned} E\exp(it\hat{S}_n(\varepsilon)/n^{1/\alpha}) &\to \sum_{m=0}^{\infty}\exp(-\varepsilon^{-\alpha})(\varepsilon^{-\alpha})^m\varphi(\varepsilon t)^m/m! \\ &= \exp(-\varepsilon^{-\alpha}\{1-\varphi(\varepsilon t)\}) \end{aligned} \tag{3.8.5}$$

为证 (3.8.3), 需要用到引理 3.4.15 的下述推广.

引理 3.8.1. 若对每个 $\varepsilon > 0$ 有 $h_n(\varepsilon) \to g(\varepsilon)$ 且当 $\varepsilon \to 0$ 时 $g(\varepsilon) \to g(0)$, 则可选取 $\varepsilon_n \to 0$ 使得 $h_n(\varepsilon_n) \to g(0)$.

证明. 取 N_m 使得当 $n \ge N_m$ 时 $|h_n(1/m) - g(1/m)| \le 1/m$, 且 $m \mapsto N_m$ 递增. 对 $N_m \le n < N_{m+1}$

定义 $\varepsilon_n = 1/m$, 对 $n < N_1$ 定义 $\varepsilon_n = 1$. 当 $N_m \le n < N_{m+1}$ 时，$\varepsilon_n = 1/m$，故由三角不等式和 ε_n 的定义得

$$\begin{aligned} |h_n(\varepsilon_n) - g(0)| &\le |h_n(1/m) - g(1/m)| + |g(1/m) - g(0)| \\ &\le 1/m + |g(1/m) - g(0)| \end{aligned}$$

当 $n \to \infty$ 时，有 $m \to \infty$，结论得证. □

记 $h_n(\varepsilon) = E\exp(it\hat{S}_n(\varepsilon)/n^{1/\alpha})$, $g(\varepsilon) = \exp(-\varepsilon^{-\alpha}\{1-\varphi(\varepsilon t)\})$. (3.8.2) 蕴含当 $t \to 0$ 时 $1-\varphi(t) \sim C|t|^\alpha$，故

$$g(\varepsilon) \to \exp(-C|t|^\alpha) \quad \text{当 } \varepsilon \to 0$$

由引理 3.8.1, 可选取 $\varepsilon_n \to 0$ 使得 $h_n(\varepsilon_n) \to \exp(-C|t|^\alpha)$. 引入满足 $E\exp(itY) = \exp(-C|t|^\alpha)$ 的随机变量 Y, 则 $\hat{S}_n(\varepsilon_n)/n^{1/\alpha} \Rightarrow Y$. 若 $\varepsilon_n \to 0$, 由 (3.8.4) 得

$$\bar{S}_n(\varepsilon_n)/n^{1/\alpha} \Rightarrow 0$$

由收敛在一起引理（习题 3.2.13）得 (3.8.3).

给出最后一个定义后, 下面陈述并证明前述一般结果. 称 L 为**慢变函数**，若

$$\lim_{x\to\infty} L(tx)/L(x) = 1 \quad \text{对所有 } t > 0$$

注意 $L(t) = \log t$ 是慢变函数，但 t^ε 不是（若 $\varepsilon \neq 0$）.

定理 3.8.2. 设 $X_1, X_2, \ldots$ 为独立同分布随机变量，其分布满足

(i) $\lim_{x\to\infty} P(X_1 > x)/P(|X_1| > x) = \theta \in [0,1]$

(ii) $P(|X_1| > x) = x^{-\alpha}L(x)$

其中 $\alpha < 2$ 且 L 为慢变函数. 记 $S_n = X_1 + \cdots + X_n$,

$$a_n = \inf\{x : P(|X_1| > x) \le n^{-1}\} \quad \text{且} \quad b_n = nE(X_1 1(|X_1| \le a_n))$$

则当 $n \to \infty$ 时，$(S_n - b_n)/a_n \Rightarrow Y$，其中 Y 具有非退化分布.

注释. 这并非对例子的实质性推广, 但这些条件是存在常数 a_n 和 b_n 使得 $(S_n - b_n)/a_n \Rightarrow Y$(其中 Y 为非退化的) 的充要条件. 必要性的证明可见 Breiman (1968) 第 9 章或 Gnedenko and Kolmogorov (1954). (3.8.11) 给出 Y 的特征函数. 读者已在 (3.8.3) 的第二证明中见过主要思想, 故可直接跳到该处而无甚损失.

证明. 易见 (ii) 蕴含

$$nP(|X_1| > a_n) \to 1 \tag{3.8.6}$$

为证此, 注意 $nP(|X_1|>a_n)\le 1$ 且取 $\varepsilon>0$. 令 $x=a_n/(1+\varepsilon)$ 和 $t=1+2\varepsilon$, (ii) 蕴含

$$(1+2\varepsilon)^{-\alpha}=\lim_{n\to\infty}\frac{P(|X_1|>(1+2\varepsilon)a_n/(1+\varepsilon))}{P(|X_1|>a_n/(1+\varepsilon))}\le\liminf_{n\to\infty}\frac{P(|X_1|>a_n)}{1/n}$$

因 ε 任意，证得 (3.8.6). 综合 (3.8.6) 与 (i) 和 (ii) 得

$$nP(X_1>xa_n)\to\theta x^{-\alpha}\quad\text{对 } x>0 \tag{3.8.7}$$

故 $|\{m\le n: X_m>xa_n\}|\Rightarrow \text{Poisson}(\theta x^{-\alpha})$. 与前面一样, 最后一个结果蕴含 $\mathcal{X}_n=\{X_m/a_n: 1\le m\le n\}$ 收敛于 $(-\infty,\infty)$ 上具有均值测度

$$\mu(A)=\int_{A\cap(0,\infty)}\theta\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x+\int_{A\cap(-\infty,0)}(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x$$

的 Poisson 过程.

为对这些点求和, 令 $I_n(\varepsilon)=\{m\le n:|X_m|>\varepsilon a_n\}$,

$$\begin{aligned}
&\hat{\mu}(\varepsilon)=EX_m1(\varepsilon a_n<|X_m|\le a_n),\quad \hat{S}_n(\varepsilon)=\sum_{m\in I_n(\varepsilon)}X_m,\\
&\bar{\mu}(\varepsilon)=EX_m1(|X_m|\le\varepsilon a_n),\\
&\bar{S}_n(\varepsilon)=(S_n-b_n)-(\hat{S}_n(\varepsilon)-n\hat{\mu}(\varepsilon))=\sum_{m=1}^{n}\{X_m1(|X_m|\le\varepsilon a_n)-\bar{\mu}(\varepsilon)\}.
\end{aligned}$$

记 $\bar{X}_m(\varepsilon)=X_m1(|X_m|\le\varepsilon a_n)$, 则

$$\begin{aligned}
E(\bar{S}_n(\varepsilon)/a_n)^2&=n\,\mathrm{var}(\bar{X}_1(\varepsilon)/a_n)\le nE(\bar{X}_1(\varepsilon)/a_n)^2\\
E(\bar{X}_1(\varepsilon)/a_n)^2&\le\int_0^{\varepsilon}2yP(|X_1|>ya_n)\,\mathrm{d}y\\
&=P(|X_1|>a_n)\int_0^{\varepsilon}2y\frac{P(|X_1|>ya_n)}{P(|X_1|>a_n)}\,\mathrm{d}y
\end{aligned}$$

希望利用 (3.8.7) 和 (ii) 得到

$$nE(\bar{X}_1(\varepsilon)/a_n)^2\to\int_0^{\varepsilon}2y\cdot y^{-\alpha}\,\mathrm{d}y=\frac{2\varepsilon^{2-\alpha}}{2-\alpha}$$

从而

$$\limsup_{n\to\infty}E(\bar{S}_n(\varepsilon)/a_n)^2\le\frac{2\varepsilon^{2-\alpha}}{2-\alpha} \tag{3.8.8}$$

为证极限与积分的交换合理并完成 (3.8.8) 的证明, 下面证明 (取 $\delta < 2-\alpha$):

引理 3.8.3. 对任意 $\delta > 0$ 存在常数 C 使得对所有 $t \geq t_0$ 和 $y \leq 1$

$$P(|X_1| > yt)/P(|X_1| > t) \leq Cy^{-\alpha-\delta}$$

证明. (ii) 蕴含当 $t \to \infty$ 时

$$P(|X_1| > t/2)/P(|X_1| > t) \to 2^{\alpha}$$

故对 $t \geq t_0$ 有

$$P(|X_1| > t/2)/P(|X_1| > t) \leq 2^{\alpha+\delta}$$

迭代并在第一次满足 $t/2^m < t_0$ 时停止，则对所有 $n \geq 1$

$$P(|X_1| > t/2^n)/P(|X_1| > t) \leq C2^{(\alpha+\delta)n}$$

其中 $C = 1/P(|X_1| > t_0)$. 将最后一个结果应用于满足 $1/2^n < y$ 的第一个 n 并注意 $y \leq 1/2^{n-1}$，得

$$P(|X_1| > yt)/P(|X_1| > t) \leq C2^{\alpha+\delta}y^{-\alpha-\delta}$$

引理得证. □

为计算 $\hat{S}_n(\varepsilon)$ 的极限, 注意 $|I_n(\varepsilon)| \Rightarrow \text{Poisson}(\varepsilon^{-\alpha})$. 给定 $|I_n(\varepsilon)| = m$，$\hat{S}_n(\varepsilon)/a_n$ 是 m 个独立随机变量之和，其分布 F_n^{ε} 满足

$$\begin{aligned} 1 - F_n^{\varepsilon}(x) &= P(X_1/a_n > x \mid |X_1|/a_n > \varepsilon) \to \theta x^{-\alpha}/\varepsilon^{-\alpha} \\ F_n^{\varepsilon}(-x) &= P(X_1/a_n < -x \mid |X_1|/a_n > \varepsilon) \to (1-\theta)|x|^{-\alpha}/\varepsilon^{-\alpha} \end{aligned}$$

对 $x \geq \varepsilon$. 若令 $\psi_n^{\varepsilon}(t)$ 表示 F_n^{ε} 的特征函数，则定理 3.3.17 蕴含当 $n \to \infty$ 时

$$\psi_n^{\varepsilon}(t) \to \psi^{\varepsilon}(t) = \int_{\varepsilon}^{\infty} e^{itx}\theta\varepsilon^{\alpha}\alpha x^{-(\alpha+1)}\,\mathrm{d}x + \int_{-\infty}^{-\varepsilon} e^{itx}(1-\theta)\varepsilon^{\alpha}\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x$$

故重复 (3.8.5) 的证明得

$$\begin{aligned} E\exp(it\hat{S}_n(\varepsilon)/a_n) &\to \exp(-\varepsilon^{-\alpha}\{1-\psi^{\varepsilon}(t)\}) \\ &= \exp\left(\int_{\varepsilon}^{\infty}(e^{itx}-1)\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x + \int_{-\infty}^{-\varepsilon}(e^{itx}-1)(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x\right) \end{aligned}$$

其中用到了 $\varepsilon^{-\alpha}=\int_{\varepsilon}^{\infty}\alpha x^{-(\alpha+1)}\,\mathrm{d}x$. 为引入

$$\hat{\mu}(\varepsilon)=EX_m1(\varepsilon a_n<|X_m|\le a_n)$$

注意 (3.8.7) 蕴含 $nP(xa_n<X_m\le ya_n)\to\theta(x^{-\alpha}-y^{-\alpha})$, 故

$$n\hat{\mu}(\varepsilon)/a_n\to\int_{\varepsilon}^{1}x\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x+\int_{-1}^{-\varepsilon}x(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x$$

由此得 $E\exp(it\{\hat{S}_n(\varepsilon)-n\hat{\mu}(\varepsilon)\}/a_n)\to$

$$\begin{aligned}\exp\Bigg(&\int_1^{\infty}(e^{itx}-1)\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x+\int_{\varepsilon}^{1}(e^{itx}-1-itx)\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x\\&+\int_{-1}^{-\varepsilon}(e^{itx}-1-itx)(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x+\int_{-\infty}^{-1}(e^{itx}-1)(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x\Bigg)\end{aligned}\tag{3.8.9}$$

最后一个表达式较复杂, 但当 $t\to0$ 时 $e^{itx}-1-itx\sim-t^2x^2/2$, 因此需要减去 itx 使得

$$\int_0^1(e^{itx}-1-itx)x^{-(\alpha+1)}\,\mathrm{d}x\quad\text{在 }\alpha\ge1\text{ 时收敛}$$

为将积分个数从四个减为两个，可将 (3.8.9) 右边当 $\varepsilon\to0$ 时的极限写为

$$\begin{aligned}\exp\Bigg(&itc+\int_0^{\infty}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x\\&+\int_{-\infty}^{0}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x\Bigg)\end{aligned}\tag{3.8.10}$$

其中 c 为常数. 综合 (3.8.6) 和 (3.8.9) 并利用引理 3.8.1, 易得 $(S_n-b_n)/a_n\Rightarrow Y$, 其中 Ee^{itY} 由 (3.8.10) 给出.

利用微积分计算 (见 Breiman (1968) 第 204–206 页) 可将 (3.8.10) 改写为

$$\exp(itc-b|t|^{\alpha}\{1+i\kappa\,\mathrm{sgn}(t)w_{\alpha}(t)\})\tag{3.8.11}$$

其中 $-1\le\kappa\le1(\kappa=2\theta-1)$ 且

$$w_{\alpha}(t)=\begin{cases}\tan(\pi\alpha/2) & \text{若 }\alpha\ne1\\(2/\pi)\log|t| & \text{若 }\alpha=1\end{cases}$$

注意, 虽然上述推导中假设 $0<\alpha<2$, 但若设 $\alpha=2$ 则含 κ 的项消失, 且 (3.8.11) 退化为均值为

c、方差为 $2b$ 的正态分布的特征函数.

称特征函数形如 (3.8.11) 或 (3.8.10) 的分布为**稳定律**. α 通常称为**指数**. 当 $\alpha=1$ 且 $\kappa=0$ 时, 得到 Cauchy 分布. 除 Cauchy 分布和正态分布外，只有另一个已知密度的情形：当 $\alpha=1/2$，$\kappa=1$，$c=0$ 且 $b=1$ 时，密度为

$$(2\pi y^3)^{-1/2}\exp(-1/2y) \quad \text{对 } y\geq 0 \tag{3.8.12}$$

可直接计算特征函数来验证该论断. 然而, 稍后 (见第 7.4 节) 将能不费力地验证该论断, 故将繁琐的计算留给读者. □

下面给出若干例子.

例题 3.8.4. 设 $X_1, X_2, \ldots$ 为独立同分布随机变量，其密度关于 0 对称且在 0 处连续为正. 断言

$$\frac{1}{n}\left(\frac{1}{X_1}+\cdots+\frac{1}{X_n}\right)\Rightarrow \text{Cauchy 分布} \quad (\alpha=1,\kappa=0)$$

为验证此, 注意

$$P(1/X_i>x)=P(0<X_i<x^{-1})=\int_0^{x^{-1}} f(y)\,\mathrm{d}y\sim f(0)/x$$

当 $x\to\infty$. 类似计算表明 $P(1/X_i<-x)\sim f(0)/x$, 故定理 3.8.2 中的 (i) 成立 ($\theta=1/2$), (ii) 成立 ($\alpha=1$). 缩放常数 $a_n\sim 2f(0)n$, 而中心化常数因假设 X 的分布关于 0 对称而为零.

注释. 欲挑战的读者可尝试去掉对称性假设, 为简单起见可设 f 在 0 处可微.

例题 3.8.5. 设 $X_1, X_2, \ldots$ 为独立同分布随机变量且 $P(X_i=1)=P(X_i=-1)=1/2$, 记 $S_n=X_1+\cdots+X_n$, 且 $\tau=\inf\{n\geq 1: S_n=1\}$. 在第 4 章 (见 (4.9.2) 后的讨论) 将证明

$$P(\tau>2n)\sim \pi^{-1/2}n^{-1/2} \quad \text{当 } n\to\infty$$

设 $\tau_1, \tau_2, \ldots$ 为与 τ 同分布的独立随机变量, 记 $T_n=\tau_1+\cdots+\tau_n$. 第 4.1 节的结果蕴含 T_n 与 S_m 第 n 次击中 0 的时刻同分布. 断言 T_n/n^2 收敛于 $\alpha=1/2, \kappa=1$ 的稳定律, 注意这是导出 (3.8.12) 的关键. 为证此断言, 注意定理 3.8.2 中的 (i) 成立 ($\theta=1$), (ii) 成立 ($\alpha=1/2$). 缩放常数 $a_n\sim Cn^2$. 因 $\alpha<1$, 习题 3.8.1 蕴含无需中心化常数.

例题 3.8.6. 假设 n 个物体 $X_{n,1}, \ldots, X_{n,n}$ 独立且随机地放置在 $[-n, n]$ 中. 令

$$F_n=\sum_{m=1}^{n}\operatorname{sgn}(X_{n,m})/|X_{n,m}|^p \tag{3.8.13}$$

为施加在 0 上的净力. 下面证明若 $p>1/2$, 则

$$\lim_{n\to\infty} E\exp(itF_n)=\exp(-c|t|^{1/p})$$

为此, 方便起见令 $X_{n,m}=nY_m$, 其中 Y_i 为 $[-1,1]$ 上的独立同分布随机变量. 则

$$F_n=n^{-p}\sum_{m=1}^{n}\operatorname{sgn}(Y_m)/|Y_m|^p$$

令 $Z_m=\operatorname{sgn}(Y_m)/|Y_m|^p$, 则 Z_m 关于 0 对称且 $P(|Z_m|>x)=P(|Y_m|<x^{-1/p})$, 故定理 3.8.2 中的 (i) 成立 ($\theta=1/2$), (ii) 成立 ($\alpha=1/p$). 缩放常数 $a_n\sim Cn^p$, 且由对称性中心化常数为 0.

例题 3.8.7. 在上述例子中，有 $b_n=0$. 为理解中心化常数，考虑独立同分布随机变量 $X_1,X_2,\ldots$ 满足

$$P(X_i>x)=\theta x^{-\alpha}\quad P(X_i<-x)=(1-\theta)x^{-\alpha}$$

其中 $0<\alpha<2$. 此时 $a_n=n^{1/\alpha}$ 且

$$b_n=n\int_1^{n^{1/\alpha}}(2\theta-1)\alpha x^{-\alpha}\,\mathrm{d}x\sim\begin{cases}cn & \alpha>1\\ cn\log n & \alpha=1\\ cn^{1/\alpha} & \alpha<1\end{cases}$$

当 $\alpha<1$ 时中心化与缩放同阶且可忽略. 当 $\alpha>1$ 时, $b_n\sim n\mu$, 其中 $\mu=EX_i$.

下面结果解释稳定律名称的由来. 称随机变量 Y 服从**稳定律**, 若对每个整数 $k>0$ 存在常数 a_k 和 b_k, 使得当 $Y_1,\ldots,Y_k$ 为与 Y 独立同分布的随机变量时, 有 $(Y_1+\cdots+Y_k-b_k)/a_k\overset{d}{=}Y$. 该定义使下述结果的一半是显然的.

定理 3.8.8. Y 是某独立同分布序列 X_i 的 $(X_1+\cdots+X_k-b_k)/a_k$ 的极限当且仅当 Y 具有稳定律.

证明. 若 Y 服从稳定律, 可取 $X_1,X_2,\ldots$ 为与 Y 同分布的独立同分布随机变量. 为证另一方向, 令

$$Z_n=(X_1+\cdots+X_n-b_n)/a_n$$

且 $S_{nj}=X_{(j-1)n+1}+\cdots+X_{jn}$. 简单计算表明

$$\begin{aligned}Z_{nk}&=(S_{n1}+\cdots+S_{nk}-b_{nk})/a_{nk}\\ a_{nk}Z_{nk}&=(S_{n1}-b_n)+\cdots+(S_{nk}-b_n)+(kb_n-b_{nk})\\ a_{nk}Z_{nk}/a_n&=(S_{n1}-b_n)/a_n+\cdots+(S_{nk}-b_n)/a_n+(kb_n-b_{nk})/a_n\end{aligned}$$

右边前 k 项当 $n\to\infty$ 时 $\Rightarrow Y_1+\cdots+Y_k$, 其中 $Y_1,\ldots,Y_k$ 独立且与 Y 同分布, 且 $Z_{nk}\Rightarrow Y$. 取 $W_n=Z_{nk}$ 且

$$W_n'=\frac{a_{kn}}{a_n}Z_{nk}-\frac{kb_n-b_{nk}}{a_n}$$

即得所需结果. □

定理 3.8.9. 收敛类型定理. 若 $W_n \Rightarrow W$ 且存在常数 $\alpha_n > 0$，β_n 使得 $W_n' = \alpha_n W_n + \beta_n \Rightarrow W'$，其中 W 和 W' 为非退化的，则存在常数 α 和 β 使得 $\alpha_n \to \alpha$ 且 $\beta_n \to \beta$.

证明. 记 $\varphi_n(t) = E\exp(itW_n)$.

$$\psi_n(t) = E\exp(it(\alpha_n W_n + \beta_n)) = \exp(it\beta_n)\varphi_n(\alpha_n t)$$

若 φ 和 ψ 分别为 W 和 W' 的特征函数，则

$$\varphi_n(t) \to \varphi(t) \quad \psi_n(t) = \exp(it\beta_n)\varphi_n(\alpha_n t) \to \psi(t) \tag{a}$$

取子列 $\alpha_{n(m)}$ 收敛于极限 $\alpha \in [0,\infty]$. 首先证明 $\alpha = 0$ 不可能. 若此发生，则利用习题 3.3.13 中证明的一致收敛性

$$|\psi_n(t)| = |\varphi_n(\alpha_n t)| \to 1 \tag{b}$$

$|\psi(t)| \equiv 1$, 由定理 3.5.2 知极限是退化的. 令 $t = u/\alpha_n$ 并交换 φ 与 ψ 的位置可知 $\alpha = \infty$ 不可能.

若 α 是子列极限，则如 (b) 中论证得 $|\psi(t)| = |\varphi(\alpha t)|$. 若有两个子列极限 $\alpha_0 < \alpha$，对两个极限用最后一个等式得 $|\varphi(u)| = |\varphi(u\alpha_0/\alpha)|$. 迭代得 $|\varphi(u)| = |\varphi(u(\alpha_0/\alpha)^k)| \to 1$(当 $k \to \infty$), 与 W' 非退化的假设矛盾, 故 $\alpha_n \to \alpha \in [0,\infty)$.

为证 $\beta_n \to \beta$, 注意习题 3.3.13 的 (ii) 蕴含 $\varphi_n \to \varphi$ 在紧集上一致收敛, 故 $\varphi_n(\alpha_n t) \to \varphi(\alpha t)$. 若 δ 足够小使得 $|t| \le \delta$ 时 $|\varphi(\alpha t)| > 0$, 则由 (a) 和习题 3.3.13 的另一应用得

$$\exp(it\beta_n) = \frac{\psi_n(t)}{\varphi_n(\alpha t)} \to \frac{\psi(t)}{\varphi(\alpha t)}$$

在 $[-\delta,\delta]$ 上一致收敛. $\exp(it\beta_n)$ 是 β_n 处点质量的特征函数. 如在定理 3.3.17 的证明中一样利用 (3.3.1), 可知点质量分布列是胎紧的, 即 β_n 有界. 若 $\beta_{n_m} \to \beta$, 则 $\exp(it\beta) = \psi(t)/\varphi(\alpha t)$ 对 $|t| \le \delta$ 成立, 故只可能有一个子列极限. □

定理 3.8.8 证明了将特征函数形如 (3.8.11) 或 (3.8.10) 的分布称为稳定律的合理性. 为完整起见, 应提及这些就是仅有的稳定律. 再次参见 Breiman (1968) 第 9 章或 Gnedenko and Kolmogorov (1954). 下面的例子表明, 知道所有可能的极限有时是有用的.

例题 3.8.10. Holtsmark 分布. $(\alpha = 3/2, \kappa = 0)$. 假设恒星按照密度为 t 的 Poisson 过程分布在空间中, 且它们的质量是独立同分布的. 令 X_t 为密度是 t 时在原点处引力作用的 x 分量. 密度变化 $1 \to t$ 对应长度变化 $1 \to t^{-1/3}$, 而引力服从平方反比律, 故

$$X_t \stackrel{d}{=} t^{3/2} X_1 \tag{3.8.14}$$

若想象用掷一个 n 面骰子来稀疏化 Poisson 过程, 则定理 3.7.4 蕴含

$$X_t \stackrel{d}{=} X^1_{t/n} + \cdots + X^n_{t/n}$$

其中右边的随机变量独立且与 $X_{t/n}$ 同分布. 由定理 3.8.8, X_t 服从稳定律. 缩放性质 (3.8.14) 蕴含 $\alpha = 3/2$. 因 $X_t \stackrel{d}{=} -X_t$, 故 $\kappa = 0$.

习题

习题 3.8.1. 证明当 $\alpha < 1$ 时, 定理 3.8.2 中的中心化不必要, 即可以取 $b_n = 0$.

习题 3.8.2. 考虑 (3.8.13) 中定义的 F_n. 证明 (i) 若 $p < 1/2$ 则 $F_n/n^{1/2-p} \Rightarrow c\chi$，其中 χ 为标准正态. (ii) 若 $p = 1/2$ 则 $F_n/(\log n)^{1/2} \Rightarrow c\chi$.

习题 3.8.3. 设 Y 为 $\kappa = 1$ 的稳定律. 利用极限定理 3.8.2 得出结论：若 $\alpha < 1$ 则 $Y \geq 0$.

习题 3.8.4. 设 X 为指数为 α 的对称稳定随机变量. (i) 利用 (3.3.1) 证明 $E|X|^p < \infty$ 对 $p < \alpha$ 成立. (ii) 利用 (3.8.3) 的第二证明证明 $P(|X| \geq x) \geq Cx^{-\alpha}$, 故 $E|X|^\alpha = \infty$.

习题 3.8.5. 设 $Y, Y_1, Y_2, \ldots$ 为独立同分布随机变量且具有指数为 α 的稳定律. 定理 3.8.8 蕴含存在常数 α_k 和 β_k 使得 $Y_1 + \cdots + Y_k$ 与 $\alpha_k Y + \beta_k$ 同分布. 利用定理 3.8.8 的证明、定理 3.8.2 和习题 3.8.1 得出结论：(i) $\alpha_k = k^{1/\alpha}$，(ii) 若 $\alpha < 1$ 则 $\beta_k = 0$.

习题 3.8.6. 设 Y 为指数 $\alpha < 1$ 且 $\kappa = 1$ 的稳定律. 习题 3.8.3 蕴含 $Y \geq 0$, 故可定义其 Laplace 变换 $\psi(\lambda) = E\exp(-\lambda Y)$. 上一习题蕴含对任意整数 $n \geq 1$ 有 $\psi(\lambda)^n = \psi(n^{1/\alpha}\lambda)$. 利用此结论证明 $E\exp(-\lambda Y) = \exp(-c\lambda^\alpha)$.

习题 3.8.7. (i) 证明若 X 为指数为 α 的对称稳定随机变量且 $Y \geq 0$ 为指数 $\beta < 1$ 的独立稳定随机变量，则 $XY^{1/\alpha}$ 为指数为 $\alpha\beta$ 的对称稳定随机变量. (ii) 设 W_1 和 W_2 为独立标准正态随机变量. 验证 $1/W_2^2$ 具有 (3.8.12) 给出的密度，并利用此得出结论 W_1/W_2 具有 Cauchy 分布.

第 9 节 无穷可分分布 *

上一节刻画了可以表示为独立同分布随机变量归一化和的极限的那些分布. 本节将描述那些可以表示为如下和式极限的分布:

$$S_n = X_{n,1} + \cdots + X_{n,n} \tag{$*$}$$

其中 $X_{n,m}$ 独立同分布. 注意这里使用了动词 "描述". 本节几乎不证明任何结论，只是陈述一些最重要的事实，使读者具备鸡尾酒会上的谈资素养.

Z 是形如 $(*)$ 的和式极限的一个充分条件是 Z 具有**无穷可分分布**，即对每个 n，存在独立同分布序列 $Y_{n,1}, \ldots, Y_{n,n}$ 使得

$$Z \overset{d}{=} Y_{n,1} + \cdots + Y_{n,n}.$$

下一定理表明这一条件也是必要的.

定理 3.9.1. Z 是形如 $(*)$ 的和式极限当且仅当 Z 具有无穷可分分布.

证明. 由上述注记，只需证必要性. 记

$$S_{2n} = (X_{2n,1} + \cdots + X_{2n,n}) + (X_{2n,n+1} + \cdots + X_{2n,2n}) \equiv Y_n + Y_n'.$$

随机变量 Y_n 与 Y_n' 独立同分布. 若 $S_n \Rightarrow Z$，则 Y_n 的分布构成胎紧序列，因为

$$P(Y_n > y)^2 = P(Y_n > y)P(Y_n' > y) \leq P(S_{2n} > 2y),$$

且类似地 $P(Y_n < -y)^2 \leq P(S_{2n} < -2y)$. 若取子列 n_k 使得 $Y_{n_k} \Rightarrow Y$（从而 $Y_{n_k}' \Rightarrow Y'$），则 $Z \overset{d}{=} Y + Y'$. 类似论证表明 Z 可被分成 $n > 2$ 份，证毕. □

确定定理 3.9.1 后，下面转向例子. 前三种情形下的分布均为无穷可分，因其为形如 $(*)$ 的和式极限. 所给数字指明相关极限定理.

例题 3.9.2. 正态分布. 定理 3.4.1

例题 3.9.3. 稳定律. 定理 3.8.2

例题 3.9.4. Poisson 分布. 定理 3.6.1

例题 3.9.5. 复合 Poisson 分布. 设 $\xi_1, \xi_2, \ldots$ 为独立同分布随机变量，$N(\lambda)$ 为与之独立的均值为 λ 的 Poisson 随机变量. 则 $Z = \xi_1 + \cdots + \xi_{N(\lambda)}$ 具有无穷可分分布.（记 $X_{n,j} \overset{d}{=} \xi_1 + \cdots + \xi_{N(\lambda/n)}$.）对于后续发展，注意若 $\phi(t) = E\exp(it\xi_i)$，则

$$E\exp(itZ) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\phi(t)^n = \exp(-\lambda(1-\phi(t))). \tag{3.9.1}$$

例 3.9.5 是例 3.9.4 的特例，又是例 3.9.2 和 3.9.3 的推广. 为解释此说，注意若 ξ 以概率 $1/2$ 分别取值 ϵ 和 $-\epsilon$，则 $\phi(t) = (e^{i\epsilon t} + e^{-i\epsilon t})/2 = \cos(\epsilon t)$. 因此若 $\lambda = \epsilon^{-2}$，则 (3.9.1) 蕴含

$$E\exp(itZ) = \exp(-\epsilon^{-2}(1-\cos(\epsilon t))) \to \exp(-t^2/2)$$

当 $\epsilon \to 0$. 换言之，正态分布是复合 Poisson 分布的极限. 为说明稳定律也是特例（沿用定理 3.8.2 证明中的记号），记

$$\begin{aligned} I_n(\epsilon) &= \{m \le n : |X_m| > \epsilon a_n\}, \\ \hat{S}_n(\epsilon) &= \sum_{m \in I_n(\epsilon)} X_m, \\ \bar{S}_n(\epsilon) &= S_n - \hat{S}_n(\epsilon). \end{aligned}$$

若 $\epsilon_n \to 0$，则 $\bar{S}_n(\epsilon_n)/a_n \Rightarrow 0$. 若 ϵ 固定，则当 $n \to \infty$ 时有 $|I_n(\epsilon)| \Rightarrow \text{Poisson}(\epsilon^{-\alpha})$ 且 $\hat{S}_n(\epsilon)/a_n \Rightarrow$ 复合 Poisson 分布：

$$E\exp(it\hat{S}_n(\epsilon)/a_n) \to \exp(-\epsilon^{-\alpha}\{1-\psi_\epsilon(t)\}).$$

结合上述两个观察并利用定理 3.8.2 的证明，可知稳定律是复合 Poisson 分布的极限. 极限特征函数的公式 (3.8.10)

$$\begin{aligned} &\exp\left(itc + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\theta\alpha x^{-(\alpha+1)}\,\mathrm{d}x \right. \\ &\left. + \int_{-\infty}^0 \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)(1-\theta)\alpha|x|^{-(\alpha+1)}\,\mathrm{d}x\right) \end{aligned} \tag{3.9.2}$$

有助于解释：

定理 3.9.6. Lévy-Khinchin 定理. Z 具有无穷可分分布当且仅当其特征函数具有形式

$$\log\varphi(t) = ict - \frac{\sigma^2 t^2}{2} + \int\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\mu(\mathrm{d}x)$$

其中 μ 是满足 $\mu(\{0\}) = 0$ 且 $\int \frac{x^2}{1+x^2}\,\mu(\mathrm{d}x) < \infty$ 的测度.

证明参见 Breiman (1968) 第 9.5 节，或 Feller II (1971) 第 XVII.2 节. μ 称为该分布的 **Lévy 测度**. 将 (3.9.2) 与此比较并回顾定理 3.8.2 的证明，可得 μ 的如下解释：若 $\sigma^2 = 0$，则 Z 可通过在 $\mathbb{R}$ 上构造均值为测度 μ 的 Poisson 过程，再对所得点求和来构建. 如同稳定律的情形，需对 $[-\epsilon,\epsilon]^c$ 中的点求和，减去适当的常数，再令 $\epsilon \to 0$.

无穷可分分布的理论在方差有限的情形下更为简洁. 对此情形，有以下结论：

定理 3.9.7. Kolmogorov 定理. Z 具有均值为 0 且方差有限的无穷可分分布当且仅当其特征函数具

有形式

$$\log \varphi(t) = \int (e^{itx} - 1 - itx)x^{-2}\,\nu(\mathrm{d}x).$$

这里被积函数在 0 处取值为 $-t^2/2$，ν 称为**典范测度**，且 $\mathrm{var}(Z) = \nu(\mathbb{R})$.

为解释此公式，注意若 Z_λ 服从均值为 λ 的 Poisson 分布，则

$$E\exp(itx(Z_\lambda - \lambda)) = \exp(\lambda(e^{itx} - 1 - itx)),$$

故 $Z = x(Z_\lambda - \lambda)$ 对应的测度满足 $\nu(\{x\}) = \lambda x^2$.

习题

习题 3.9.1. 证明 Gamma 分布是无穷可分的.

习题 3.9.2. 证明有界随机变量 Z 的分布是无穷可分的当且仅当 Z 是常数. 提示：证明 $\mathrm{var}(Z) = 0$.

习题 3.9.3. 证明若 μ 是无穷可分的，则其特征函数 φ 永不为零. 提示：考察 $\psi = |\varphi|^2$，它也是无穷可分的，以避免对复数取 n 次根，然后利用习题 3.3.17.

习题 3.9.4. 习题 3.4.13 第 (iii) 部分中极限 $\aleph$ 的 Lévy 测度是什么?

第 10 节 $\mathbb{R}^d$ 中的极限定理

我们首先考虑一般度量空间 (S, ρ) 中的弱收敛. 在此框架下，$X_n \Rightarrow X_\infty$ 当且仅当对所有有界连续函数 f 有 $Ef(X_n) \to Ef(X_\infty)$. 如同 3.2 节，弱收敛的若干等价定义将十分有用. 称 f 为 **Lipschitz 连续**，若存在常数 C 使得 $|f(x) - f(y)| \le C\rho(x, y)$.

定理 3.10.1. 以下陈述等价：

(i) 对所有有界连续函数 f，$Ef(X_n) \to Ef(X_\infty)$.
(ii) 对所有有界 Lipschitz 连续函数 f，$Ef(X_n) \to Ef(X_\infty)$.
(iii) 对所有闭集 K，$\limsup_{n\to\infty} P(X_n \in K) \le P(X_\infty \in K)$.
(iv) 对所有开集 G，$\liminf_{n\to\infty} P(X_n \in G) \ge P(X_\infty \in G)$.
(v) 对所有满足 $P(X_\infty \in \partial A) = 0$ 的集合 A，$\lim_{n\to\infty} P(X_n \in A) = P(X_\infty \in A)$.
(vi) 设 D_f 为 f 的不连续点集. 对所有满足 $P(X_\infty \in D_f) = 0$ 的有界函数 f，有 $Ef(X_n) \to Ef(X_\infty)$.

证明. (i) 蕴含 (ii)：显然.

(ii) 蕴含 (iii)：令 $\rho(x,K)=\inf\{\rho(x,y):y\in K\}$，$\varphi_j(r)=(1-jr)^+$，以及 $f_j(x)=\varphi_j(\rho(x,K))$. f_j 是 Lipschitz 连续的，取值于 $[0,1]$，且当 $j\uparrow\infty$ 时 $f_j\downarrow \mathbf{1}_K(x)$. 因此

$$\limsup_{n\to\infty} P(X_n\in K)\le \lim_{n\to\infty} Ef_j(X_n)=Ef_j(X_\infty)\downarrow P(X_\infty\in K)\quad (j\uparrow\infty).$$

(iii) 等价于 (iv)：如同定理 3.2.11 的证明，这易从两个事实推出：A 是开集当且仅当 A^c 是闭集；$P(A)+P(A^c)=1$.

(iii) 和 (iv) 蕴含 (v)：令 $K=\bar{A}$，$G=A^\circ$，并如同定理 3.2.11 的证明那样推理.

(v) 蕴含 (vi)：设 $|f(x)|\le K$，并选取 $\alpha_0<\alpha_1<\cdots<\alpha_\ell$ 使得 $P(f(X_\infty)=\alpha_i)=0$ $(0\le i\le \ell)$，$\alpha_0<-K<K<\alpha_\ell$，且 $\alpha_i-\alpha_{i-1}<\epsilon$. 这总是可行的，因为 $\{\alpha:P(f(X_\infty)=\alpha)>0\}$ 是可数集. 令 $A_i=\{x:\alpha_{i-1}<f(x)\le\alpha_i\}$. $\partial A_i\subset\{x:f(x)\in\{\alpha_{i-1},\alpha_i\}\}\cup D_f$，故 $P(X_\infty\in\partial A_i)=0$，由 (v) 可得

$$\sum_{i=1}^{\ell}\alpha_i P(X_n\in A_i)\to\sum_{i=1}^{\ell}\alpha_i P(X_\infty\in A_i).$$

α_i 的定义表明

$$0\le\sum_{i=1}^{\ell}\alpha_i P(X_n\in A_i)-Ef(X_n)\le\epsilon\quad (1\le n\le\infty).$$

由 ϵ 任意，可得 $Ef(X_n)\to Ef(X_\infty)$.

(vi) 蕴含 (i)：显然. □

现在特化到 $\mathbb{R}^d$，令 $X=(X_1,\dots,X_d)$ 为随机向量. 其**分布函数**定义为 $F(x)=P(X\le x)$. 这里 $x\in\mathbb{R}^d$，而 $X\le x$ 表示 $X_i\le x_i$ $(i=1,\dots,d)$. 如同一维情形，F 具有三条显然的性质：

(i) 它是非减的，即若 $x\le y$ 则 $F(x)\le F(y)$.

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

(iii) F 是右连续的，即 $\lim_{y\downarrow x}F(y)=F(x)$.

这里 $x\to\infty$ 表示每个坐标 x_i 趋于 ∞，$x_i\to-\infty$ 表示令 $x_i\to-\infty$ 而保持其他坐标固定，$y\downarrow x$ 表示每个坐标 $y_i\downarrow x_i$.

如 1.1 节所讨论，为保证 F 是某概率测度的分布函数，还需一个额外条件. 令

$$A=(a_1,b_1]\times\cdots\times(a_d,b_d],$$
$$V=\{a_1,b_1\}\times\cdots\times\{a_d,b_d\}$$

V 为矩形 A 的顶点. 若 $v\in V$，令

$$\mathrm{sgn}(v)=(-1)^{\#\text{ of }a\text{'s in }v}.$$

容斥公式蕴含

$$P(X \in A) = \sum_{v \in V} \operatorname{sgn}(v) F(v).$$

故若用 $\Delta_A F$ 表示右端，我们需要

(iv) 对所有矩形 A，$\Delta_A F \geq 0$.

最后一个条件保证赋予每个矩形的测度 ≥ 0. 至此，已在 1.1.5 例中定义的半代数 $\mathcal{S}_d$ 上定义了测度. 定理 1.1.11 现蕴含存在唯一具有分布 F 的概率测度.

若 F_n 和 F 是 $\mathbb{R}^d$ 上的分布函数，我们说 F_n **弱收敛**于 F，记为 $F_n \Rightarrow F$，如果 $F_n(x) \to F(x)$ 在 F 的所有连续点处成立. 我们的首要任务是证明连续点足够多，使这成为一个合理的定义. 为给一个具体例子，考虑

$$F(x,y) = \begin{cases} 1 & \text{若} x \geq 0, y \geq 1, \\ y & \text{若} x \geq 0, 0 \leq y < 1, \\ 0 & \text{其他情形.} \end{cases}$$

F 是 $(0, Y)$ 的分布函数，其中 Y 在 $(0,1)$ 上均匀分布. 注意这个分布没有原子，但当 $y > 0$ 时 F 在 $(0, y)$ 处不连续.

记住最后一个例子，注意若 $x_n < x$（即对所有坐标 i 有 $x_{n,i} < x_i$），且 $x_n \uparrow x$（当 $n \to \infty$），则

$$F(x) - F(x_n) = P(X \leq x) - P(X \leq x_n) \downarrow P(X \leq x) - P(X < x).$$

在 $d = 2$ 时，最后的表达式是 X 落在

$$\{(a, x_2) : a \leq x_1\} \cup \{(x_1, b) : b \leq x_2\}$$

中的概率. 令 $H_i^c = \{x : x_i = c\}$ 为第 i 个坐标等于 c 的超平面. 对每个 i，H_i^c 互不相交，故 $D_i = \{c : P(X \in H_i^c) > 0\}$ 至多可数. 容易看出若 x 满足对所有 i 有 $x_i \notin D_i$，则 F 在 x 处连续. 这给出了足够多的点来重构 F.

定理 3.10.2. 在 $\mathbb{R}^d$ 上，用分布函数收敛 $F_n \Rightarrow F$ 定义的弱收敛等价于一般度量空间中弱收敛的概念.

证明. (v) 蕴含 $F_n \Rightarrow F$: 若 F 在 x 处连续，则 $A = (-\infty, x_1] \times \cdots \times (-\infty, x_d]$ 满足 $\mu(\partial A) = 0$，故 $F_n(x) = P(X_n \in A) \to P(X_\infty \in A) = F(x)$.

$F_n \Rightarrow F$ 蕴含 (iv): 令 $D_i = \{c : P(X_\infty \in H_i^c) > 0\}$，其中 $H_i^c = \{x : x_i = c\}$. 称矩形 $A = (a_1, b_1] \times \cdots \times (a_d, b_d]$ 为**好**的，若对所有 i 有 $a_i, b_i \notin D_i$. ($\Rightarrow$) 表明对所有好矩形有 $P(X_n \in A) \to P(X_\infty \in A)$. 对可表示为好矩形有限不交并的 B 这也成立. 现任何开集 G 都是 B_k 的递增极

限，其中 B_k 为好矩形的有限不交并，故

$$\liminf_{n\to\infty} P(X_n \in G) \geq \liminf_{n\to\infty} P(X_n \in B_k) = P(X_\infty \in B_k) \uparrow P(X_\infty \in G) \quad (k \to \infty).$$

证毕. □

注释. 在 3.2 节中，我们通过构造具有给定分布的随机变量使得 $X_n \to X_\infty$ 几乎必然来证明 (i)–(v) 是弱收敛的推论. 这在 $\mathbb{R}^d$（或任何完备可分度量空间）中也可做到，但构造相当繁琐. 参见 Billingsley (1979) 第 337–340 页关于 $\mathbb{R}^d$ 的证明.

称概率测度序列 μ_n 是**胎紧**的，如果对任意 $\epsilon > 0$，存在 M 使得 $\liminf_{n\to\infty} \mu_n([-M, M]^d) \geq 1-\epsilon$.

定理 3.10.3. 若 μ_n 胎紧，则存在弱收敛子列.

证明. 令 F_n 为相应的分布函数，$q_1, q_2, \ldots$ 为 $\mathbb{Q}^d = \mathbb{R}^d$ 中有理坐标点的枚举. 如同定理 3.2.12 证明中的对角线论证，可选取子列使得 $F_{n(k)}(q) \to G(q)$ 对所有 $q \in \mathbb{Q}^d$ 成立. 令

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}^d, q > x\}$$

其中 $q > x$ 表示对所有 i 有 $q_i > x_i$. 容易看出 F 是右连续的. 为验证它是分布函数，注意若 A 是顶点在 $\mathbb{Q}^d$ 中的矩形，则对所有 n 有 $\Delta_A F_n \geq 0$，故 $\Delta_A G \geq 0$，取极限可知对 F 这一结论对所有矩形 A 成立. 胎紧性蕴含 F 具有分布函数的 (i) 和 (ii) 两条性质. 留给读者验证 $F_n \Rightarrow F$. 若将诸如 $r_1 < r_2 < x < s$ 的不等式读作向量间的相应关系，则定理 3.2.12 的证明在此适用. □

$(X_1, \ldots, X_d)$ 的**特征函数**是 $\varphi(t) = E\exp(it \cdot X)$，其中 $t \cdot X = t_1X_1 + \cdots + t_dX_d$ 是两个向量的通常点积.

定理 3.10.4. 反演公式. 若 $A = [a_1, b_1] \times \cdots \times [a_d, b_d]$ 满足 $\mu(\partial A) = 0$，则

$$\mu(A) = \lim_{T\to\infty} (2\pi)^{-d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j)\varphi(t)\,\mathrm{d}t$$

其中 $\psi_j(s) = (\exp(-isa_j) - \exp(-isb_j))/is$.

证明. Fubini 定理蕴含

$$\int_{[-T,T]^d} \int \prod_{j=1}^{d} \psi_j(t_j) \exp(it_jx_j)\,\mu(\mathrm{d}x)\,\mathrm{d}t = \int \prod_{j=1}^{d} \int_{-T}^{T} \psi_j(t_j) \exp(it_jx_j)\,\mathrm{d}t_j\,\mu(\mathrm{d}x).$$

由定理 3.3.11 的证明可知

$$\int_{-T}^{T} \psi_j(t_j) \exp(it_jx_j)\,\mathrm{d}t_j \to \pi\left(\mathbf{1}_{(a_j,b_j)}(x) + \mathbf{1}_{[a_j,b_j]}(x)\right)$$

故所求结论由有界收敛定理推出. □

定理 3.10.5. 收敛定理. 设 X_n（$1 \le n \le \infty$）为具有特征函数 φ_n 的随机向量. $X_n \Rightarrow X_\infty$ 的充分必要条件为 $\varphi_n(t) \to \varphi_\infty(t)$.

证明. $\exp(it \cdot x)$ 有界连续，故若 $X_n \Rightarrow X_\infty$，则 $\varphi_n(t) \to \varphi_\infty(t)$. 为证另一方向，如同定理 3.3.17 的证明，只需证序列胎紧. 为此，注意若固定 $\theta \in \mathbb{R}^d$，则对所有 $s \in \mathbb{R}$ 有 $\varphi_n(s\theta) \to \varphi_\infty(s\theta)$，故由定理 3.3.17 可知 $\theta \cdot X_n$ 的分布胎紧. 将上述观察应用于 d 个单位向量 $e_1, \ldots, e_d$，即知 X_n 的分布胎紧，证毕. □

注释. 如同前面，若 $\varphi_n(t) \to \varphi_\infty(t)$ 且 φ_∞ 在 0 处连续，则 $\varphi_\infty(t)$ 是某个 X_∞ 的特征函数且 $X_n \Rightarrow X_\infty$.

定理 3.10.5 有一个重要的推论.

定理 3.10.6. Cramér-Wold 方法. $X_n \Rightarrow X_\infty$ 的一个充分条件是对所有 $\theta \in \mathbb{R}^d$ 有 $\theta \cdot X_n \Rightarrow \theta \cdot X_\infty$.

证明. 所指条件蕴含对所有 $\theta \in \mathbb{R}^d$ 有 $E\exp(i\theta \cdot X_n) \to E\exp(i\theta \cdot X_\infty)$. □

定理 3.10.6 立即导出

定理 3.10.7. $\mathbb{R}^d$ 中的中心极限定理. 设 $X_1, X_2, \ldots$ 为独立同分布随机向量，$EX_n = \mu$，且具有有限协方差

$$\Gamma_{ij} = E((X_{n,i} - \mu_i)(X_{n,j} - \mu_j)).$$

若 $S_n = X_1 + \cdots + X_n$，则 $(S_n - n\mu)/n^{1/2} \Rightarrow \chi$，其中 χ 服从均值为 0 且协方差为 Γ 的**多元正态分布**，即

$$E\exp(i\theta \cdot \chi) = \exp\left(-\sum_i \sum_j \theta_i \theta_j \Gamma_{ij}/2\right).$$

证明. 通过考虑 $X_n' = X_n - \mu$，不妨设 $\mu = 0$. 令 $\theta \in \mathbb{R}^d$. $\theta \cdot X_n$ 是均值为 0 且方差为

$$E\left(\sum_i \theta_i X_{n,i}\right)^2 = \sum_i \sum_j E(\theta_i \theta_j X_{n,i} X_{n,j}) = \sum_i \sum_j \theta_i \theta_j \Gamma_{ij}$$

的随机变量，故由一维中心极限定理和定理 3.10.6 可知 $S_n/n^{1/2} \Rightarrow \chi$，其中

$$E\exp(i\theta \cdot \chi) = \exp\left(-\sum_i \sum_j \theta_i \theta_j \Gamma_{ij}/2\right),$$

证得所求结果. □

为说明定理 3.10.7 的用途，下面考虑两个例子. 每个例子中 $e_1, \ldots, e_d$ 为 d 个单位向量.

例题 3.10.8. $\mathbb{Z}^d$ 上的**简单随机游走**. 令 $X_1, X_2, \ldots$ 为独立同分布随机变量，满足

$$P(X_n = +e_i) = P(X_n = -e_i) = 1/2d \quad \text{对} i = 1, \ldots, d.$$

$EX_{n,i} = 0$, 且若 $i \neq j$ 则 $EX_{n,i}X_{n,j} = 0$, 因为两个分量不能同时非零. 故协方差矩阵为 $\Gamma = (1/2d)I$.

例题 3.10.9. 令 $X_1, X_2, \ldots$ 为独立同分布随机变量，满足 $P(X_n = e_i) = 1/6$ $(i = 1, 2, \ldots, 6)$. 换言之，每次掷一颗骰子并记录出现的点数. $EX_{n,i} = 1/6$ 且 $i \neq j$ 时 $EX_{n,i}X_{n,j} = 0$，故 $i = j$ 时 $\Gamma_{ij} = (1/6)(5/6)$，而 $i \neq j$ 时 $\Gamma_{ij} = -(1/6)^2$. 在此情形，极限分布在 $\{x : \sum_i x_i = 0\}$ 上集中.

对中心极限定理的讨论若不对多元正态分布作一些说明则是不完整的. 首先注意 $\Gamma_{ij} = \Gamma_{ji}$，且若 $EX_i = 0$ 且 $EX_iX_j = \Gamma_{i,j}$，则

$$\sum_i \sum_j \theta_i \theta_j \Gamma_{ij} = E\left(\sum_i \theta_i X_i\right)^2 \geq 0,$$

故 Γ 对称且非负定. 一个众所周知的结果蕴含存在正交矩阵 U（即满足 $U^tU = I$）使得 $\Gamma = U^tVU$, 其中 $V \geq 0$ 为对角矩阵. 令 W 为满足 $W^2 = V$ 的非负对角矩阵. 若令 $A = WU$，则 $\Gamma = A^tA$. 令 Y 为 d 维向量，其分量独立且服从均值为 0、方差为 1 的正态分布. 若将向量视为 $1 \times d$ 矩阵并令 $\chi = YA$，则 χ 具有所求的多元正态分布. 为验证此点，注意

$$\theta \cdot YA = \sum_i \theta_i \sum_j Y_j A_{ji}$$

服从均值为 0、方差为

$$\left(\sum_j \sum_i A_{ji}\theta_i\right)^2 = \sum_j \left(\sum_i \theta_i A^t_{ij}\right)\left(\sum_k A_{jk}\theta_k\right) = \theta A^t A \theta^t = \theta \Gamma \theta^t$$

的正态分布，故 $E(\exp(i\theta \cdot \chi)) = \exp(-(\theta\Gamma\theta^t)/2)$.

若协方差矩阵的秩为 d，则称该正态分布是**非退化**的. 在此情形，其密度函数为

$$(2\pi)^{-d/2}(\det \Gamma)^{-1/2} \exp\left(-\sum_{i,j} y_i \Gamma^{-1}_{ij} y_j/2\right).$$

退化情形下的联合分布可通过线性变换约化到非退化情形来计算. 例如，在例 3.10.9 中可考察 $(X_1, \ldots, X_5)$ 的分布.

习题

习题 3.10.1. 若 F 为 $(X_1,\dots,X_d)$ 的分布，则 $F_i(x)=P(X_i\le x)$ 为其**边缘分布**. 如何从 F 得到它们?

习题 3.10.2. 令 $F_1,\dots,F_d$ 为 $\mathbb{R}$ 上的分布. 证明对任意 $\alpha\in[-1,1]$,

$$F(x_1,\dots,x_d)=\left\{1+\alpha\prod_{i=1}^{d}(1-F_i(x_i))\right\}\prod_{j=1}^{d}F_j(x_j)$$

是具有给定边缘分布的分布函数. $\alpha=0$ 的情形对应独立随机变量.

习题 3.10.3. 称分布 F 具有密度 f，若

$$F(x_1,\dots,x_k)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}f(y)\,\mathrm{d}y_k\cdots\mathrm{d}y_1.$$

证明若 f 连续，则 $\partial^k F/\partial x_1\cdots\partial x_k=f$.

习题 3.10.4. 令 X_n 为随机向量. 证明若 $X_n\Rightarrow X$ 则坐标 $X_{n,i}\Rightarrow X_i$.

习题 3.10.5. 令 φ 为 $\mathbb{R}$ 上分布 F 的特征函数. 对应于特征函数 $\psi(t_1,\dots,t_d)=\varphi(t_1+\cdots+t_d)$ 的 $\mathbb{R}^d$ 上的分布是什么?

习题 3.10.6. 证明随机变量 $X_1,\dots,X_k$ 独立当且仅当

$$\varphi_{X_1,\dots,X_k}(t)=\prod_{j=1}^{k}\varphi_{X_j}(t_j).$$

习题 3.10.7. 设 $(X_1,\dots,X_d)$ 服从均值为向量 θ、协方差为 Γ 的多元正态分布. 证明 $X_1,\dots,X_d$ 独立当且仅当 $i\ne j$ 时 $\Gamma_{ij}=0$. 换言之，联合正态的不相关随机变量独立.

习题 3.10.8. 证明 $(X_1,\dots,X_d)$ 服从均值为向量 θ、协方差为 Γ 的多元正态分布当且仅当每个线性组合 $c_1X_1+\cdots+c_dX_d$ 服从均值为 $c\theta^t$、方差为 $c\Gamma c^t$ 的正态分布.

第四章

鞅

鞅 X_n 可理解为参与者在时刻 n 进行**公平博弈**时的财富；下鞅（上鞅）则对应**有利博弈**（**不利博弈**）的情形. 关于鞅有两个基本事实. 其一，无法通过押注鞅来获利（见定理 4.2.8），特别地，若在有界时刻 N 停止游戏，则期望收益 EX_N 等于初始财富 X_0.（此处暂设 X_0 非随机.）其二，定理 4.2.11 涉及下鞅. 借用从 Mike Brennan 处学到的一个启发式说法，“下鞅是单调非减序列的随机类比，因此若它们上有界（准确地说，$\sup_n EX_n^+ < \infty$），则几乎必然收敛.” 如 4.3 节所示，这一结果有广泛应用. 后续各节分别讨论鞅在 L^p 中的收敛（$p > 1$，4.4 节）、在 L^1 中的收敛（4.6 节）、平方可积鞅的特殊情形（4.5 节），以及由 $n \leqslant 0$ 索引的鞅（4.7 节）. 4.8 节给出使 $EX_N = EX_0$ 对有界停时成立的充分条件，这些结果在研究随机游走时颇为有用. 4.9 节则通过组合论证补充 4.8.1 节中由鞅方法得到的随机游走结果.

第 1 节　条件期望

本节首先给出一个对本章及下一章都至关重要的定义，然后通过若干例子加以阐释. 给定概率空间 $(\Omega, \mathcal{F}_0, P)$，$\sigma$-域 $\mathcal{F} \subset \mathcal{F}_0$，以及满足 $E|X| < \infty$ 的随机变量 $X \in \mathcal{F}_0$. 定义 X 关于 $\mathcal{F}$ 的**条件期望** $E(X|\mathcal{F})$ 为满足以下条件的随机变量 Y：

(i) $Y \in \mathcal{F}$，即 Y 是 $\mathcal{F}$ 可测的；

(ii) 对所有 $A \in \mathcal{F}$，$\int_A X \,\mathrm{d}P = \int_A Y \,\mathrm{d}P$.

满足 (i) 和 (ii) 的任一 Y 都称为 $E(X|\mathcal{F})$ 的一个**版本**. 首先需确立条件期望的存在性与唯一性. 先证唯一性，但在此之前需要一个技术性引理.

引理 4.1.1. 若 Y 满足 (i) 和 (ii)，则 Y 可积.

证明. 记 $A=\{Y>0\}\in\mathcal{F}$，两次运用 (ii) 后相加得

$$\int_A Y\,\mathrm{d}P=\int_A X\,\mathrm{d}P\leqslant\int_A|X|\,\mathrm{d}P,$$
$$\int_{A^c}-Y\,\mathrm{d}P=\int_{A^c}-X\,\mathrm{d}P\leqslant\int_{A^c}|X|\,\mathrm{d}P.$$

因此 $E|Y|\leqslant E|X|$. □

唯一性. 若 Y' 亦满足 (i) 和 (ii)，则

$$\int_A Y\,\mathrm{d}P=\int_A Y'\,\mathrm{d}P\quad\text{对所有}A\in\mathcal{F}.$$

取 $A=\{Y-Y'\geqslant\varepsilon>0\}$，可见

$$0=\int_A X-X\,\mathrm{d}P=\int_A Y-Y'\,\mathrm{d}P\geqslant\varepsilon P(A),$$

故 $P(A)=0$. 由于这对所有 ε 成立，有 $Y\leqslant Y'$ 几乎必然，交换 Y 与 Y' 的角色得 $Y=Y'$ 几乎必然. 技术上，所有等式如 $Y=E(X|\mathcal{F})$ 应理解为 $Y=E(X|\mathcal{F})$ 几乎必然，但前几章忽略这一点，今后亦将继续如此.

重复上述论证可得

定理 4.1.2. 若 $X_1=X_2$ 在 $B\in\mathcal{F}$ 上成立，则 $E(X_1|\mathcal{F})=E(X_2|\mathcal{F})$ 几乎必然在 B 上成立.

证明. 记 $Y_1=E(X_1|\mathcal{F})$，$Y_2=E(X_2|\mathcal{F})$. 取 $A=\{Y_1-Y_2\geqslant\varepsilon>0\}$，可见

$$0=\int_{A\cap B}X_1-X_2\,\mathrm{d}P=\int_{A\cap B}Y_1-Y_2\,\mathrm{d}P\geqslant\varepsilon P(A),$$

故 $P(A)=0$，结论如前所述. □

存在性. 首先回顾，若 $\mu(A)=0$ 蕴含 $\nu(A)=0$，则称 ν 关于 μ **绝对连续**（记为 $\nu\ll\mu$）. 下面使用定理 A.4.8：

Radon-Nikodym 定理. 设 μ 和 ν 是 $(\Omega,\mathcal{F})$ 上的 σ-有限测度. 若 $\nu\ll\mu$，则存在 $f\in\mathcal{F}$ 使得对所有 $A\in\mathcal{F}$

$$\int_A f\,\mathrm{d}\mu=\nu(A).$$

f 通常记为 $\mathrm{d}\nu/\mathrm{d}\mu$，称为 **Radon-Nikodym 导数**.

由上述定理易证条件期望的存在性. 先设 $X \geqslant 0$. 令 $\mu = P$，且

$$\nu(A) = \int_A X \,\mathrm{d}P \quad \text{对} A \in \mathcal{F}.$$

控制收敛定理蕴含 ν 是一个测度（见习题 1.5.4），且由积分定义知 $\nu \ll \mu$. Radon-Nikodym 导数 $\mathrm{d}\nu/\mathrm{d}\mu \in \mathcal{F}$ 且对任意 $A \in \mathcal{F}$ 有

$$\int_A X \,\mathrm{d}P = \nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}P.$$

取 $A = \Omega$，可见 $\mathrm{d}\nu/\mathrm{d}\mu \geqslant 0$ 可积，且已证明 $\mathrm{d}\nu/\mathrm{d}\mu$ 是 $E(X|\mathcal{F})$ 的一个版本.

再处理一般情形，写 $X = X^+ - X^-$，记 $Y_1 = E(X^+|\mathcal{F})$，$Y_2 = E(X^-|\mathcal{F})$. 则 $Y_1 - Y_2 \in \mathcal{F}$ 可积，且对所有 $A \in \mathcal{F}$ 有

$$\begin{aligned}\int_A X \,\mathrm{d}P &= \int_A X^+ \,\mathrm{d}P - \int_A X^- \,\mathrm{d}P \\ &= \int_A Y_1 \,\mathrm{d}P - \int_A Y_2 \,\mathrm{d}P = \int_A (Y_1 - Y_2) \,\mathrm{d}P.\end{aligned}$$

这表明 $Y_1 - Y_2$ 是 $E(X|\mathcal{F})$ 的一个版本. 证毕.

4.1.1 例子

直观上，$\mathcal{F}$ 描述了可获得的信息——对每个 $A \in \mathcal{F}$，知道 A 是否发生. $E(X|\mathcal{F})$ 就是给定已有信息时，对 X 值的“最佳猜测”. 若干例子有助于澄清这一点，并将 $E(X|\mathcal{F})$ 与其他条件期望的定义联系起来.

例题 4.1.3. 若 $X \in \mathcal{F}$，则 $E(X|\mathcal{F}) = X$；即，若已知 X，则“最佳猜测”就是 X 本身. 由于 X 恒满足 (ii)，唯一阻止 X 成为 $E(X|\mathcal{F})$ 的是条件 (i). 此例的特殊情形为 $X = c$，其中 c 为常数.

例题 4.1.4. 在完全信息的另一端是无信息情形. 设 X 与 $\mathcal{F}$ 独立，即对所有 $B \in \mathbb{R}$ 和 $A \in \mathcal{F}$

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

断言：在此情形下，$E(X|\mathcal{F}) = EX$；即，若对 X 一无所知，则最佳猜测就是均值 EX. 为验证定义，注意 $EX \in \mathcal{F}$，故 (i) 成立. 为验证 (ii)，观察到若 $A \in \mathcal{F}$，则因 X 与 $1_A \in \mathcal{F}$ 独立，由定理 2.1.13 得

$$\int_A X \,\mathrm{d}P = E(X1_A) = EX \cdot E1_A = \int_A EX \,\mathrm{d}P.$$

读者应注意，此处及以下采用的方法是“猜测并验证”：先给出条件期望的一个表达式，再验证它满足 (i) 和 (ii).

例题 4.1.5. 在此例中，将新的条件期望定义与本科概率课程中教授的第一个定义联系起来. 设 $\Omega_1, \Omega_2, \ldots$ 是 Ω 分成不交子集的有限或无限划分，每个子集都有正概率，令 $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \ldots)$ 为由这些集合生成的 σ-域. 则

$$E(X|\mathcal{F}) = \frac{E(X;\Omega_i)}{P(\Omega_i)} \quad \text{在}\Omega_i \text{ 上.}$$

换言之，Ω_i 中的信息表明结果位于划分的哪个元素中，给定这一信息，对 X 的最佳猜测就是 X 在 Ω_i 上的平均值. 为证明猜测正确，注意到所给公式在每个 Ω_i 上为常数，故关于 $\mathcal{F}$ 可测. 为验证 (ii)，只需对 $A = \Omega_i$ 检验等式，但这显然成立：

$$\int_A \frac{E(X;\Omega_i)}{P(\Omega_i)}\,\mathrm{d}P = E(X;\Omega_i) = \int_{\Omega_i} X\,\mathrm{d}P = \int_A X\,\mathrm{d}P.$$

一个退化但重要的特殊情形是 $\mathcal{F} = \{\emptyset, \Omega\}$，即平凡 σ-域. 在此情形下，$E(X|\mathcal{F}) = EX$.

为继续与本科概念的联系，记

$$P(A|\mathcal{G}) = E(1_A|\mathcal{G}), \quad P(A|B) = P(A \cap B)/P(B),$$

并观察到在上例中 $P(A|\mathcal{F}) = P(A|\Omega_i)$ 在 Ω_i 上成立.

给定 σ-域的条件期望定义包含了给定随机变量的条件期望作为特例. 定义

$$E(X|Y) = E(X|\sigma(Y)),$$

其中 $\sigma(Y)$ 是 Y 生成的 σ-域.

例题 4.1.6. 为继续与本科概率中条件期望定义的联系，设 X 和 Y 有联合密度 $f(x,y)$，即

$$P((X,Y) \in B) = \int_B f(x,y)\,\mathrm{d}x\,\mathrm{d}y \quad \text{对}B \in \mathbb{R}^2,$$

且为简单起见设对所有 y 有 $\int f(x,y)\,\mathrm{d}x > 0$. 断言：在此情形下，若 $E|g(X)| < \infty$ 则 $E(g(X)|Y) = h(Y)$，其中

$$h(y) = \int g(x)f(x,y)\,\mathrm{d}x \Big/ \int f(x,y)\,\mathrm{d}x.$$

为 “猜测” 此公式，注意将概率密度 $P(Y = y)$ 当作真实概率处理时

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x,y)}{\int f(x,y)\,\mathrm{d}x},$$

故关于条件概率密度积分，有

$$E(g(X)|Y=y)=\int g(x)P(X=x|Y=y)\,\mathrm{d}x.$$

为“验证”所给公式，观察到 $h(Y)\in\sigma(Y)$，故 (i) 成立. 为检验 (ii)，注意到若 $A\in\sigma(Y)$ 则对某个 $B\in\mathbb{R}$ 有 $A=\{\omega:Y(\omega)\in B\}$，故

$$\begin{aligned}E(h(Y);A)&=\int_B\int h(y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\int_B\int g(x)f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=E(g(X)1_B(Y))=E(g(X);A).\end{aligned}$$

注释. 为去掉假设 $\int f(x,y)\,\mathrm{d}x>0$，由

$$\int h(y)f(x,y)\,\mathrm{d}x=\int g(x)f(x,y)\,\mathrm{d}x$$

定义 h（即，在 $\int f(x,y)\,\mathrm{d}x=0$ 处 h 可为任意值），并注意到这对证明已足够.

例题 4.1.7. 设 X 和 Y 独立. 令 φ 为满足 $E|\varphi(X,Y)|<\infty$ 的函数，记 $g(x)=E(\varphi(x,Y))$. 下面证明

$$E(\varphi(X,Y)|X)=g(X).$$

证明. 显然 $g(X)\in\sigma(X)$. 为检验 (ii)，注意若 $A\in\sigma(X)$ 则 $A=\{X\in C\}$，故利用变量替换公式（定理 1.6.9）以及 (X,Y) 的分布是乘积测度（定理 2.1.11）这一事实，再由 g 的定义并再次变量替换，

$$\begin{aligned}\int_A\varphi(X,Y)\,\mathrm{d}P&=E\{\varphi(X,Y)1_C(X)\}\\&=\int\int\varphi(x,y)1_C(x)\,\nu(\mathrm{d}y)\,\mu(\mathrm{d}x)\\&=\int 1_C(x)g(x)\,\mu(\mathrm{d}x)=\int_A g(X)\,\mathrm{d}P,\end{aligned}$$

证得所需结果. □

例题 4.1.8. Borel 悖论. 设 X 是地球上随机选取的一点，θ 为其经度，ϕ 为其纬度. 通常取 $\theta\in[0,2\pi)$ 和 $\phi\in(-\pi/2,\pi/2]$，但也可以同样取 $\theta\in[0,\pi)$ 和 $\phi\in(-\pi,\pi]$. 换言之，新的经度指定了该点所在的大圆，然后 ϕ 给出角度.

乍看之下，若 X 在地球上均匀分布，则 θ 和大圆上的角度 ϕ 都应在其可能取值上均匀分布. θ 是均匀的，但 ϕ 不是. 一旦意识到在新坐标或传统坐标中 ϕ 与 θ 独立，悖论就完全消失了，因此条件分布就是无条件分布，它不是均匀的，因为赤道附近的陆地比北极附近多.

4.1.2 性质

条件期望具有许多与普通期望相同的性质.

定理 4.1.9. 在前两部分中假设 $E|X|, E|Y| < \infty$.

(a) 条件期望是线性的:

$$E(aX + Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F}). \tag{4.1.1}$$

(b) 若 $X \leqslant Y$ 则

$$E(X|\mathcal{F}) \leqslant E(Y|\mathcal{F}). \tag{4.1.2}$$

(c) 若 $X_n \geqslant 0$ 且 $X_n \uparrow X$ 满足 $EX < \infty$ 则

$$E(X_n|\mathcal{F}) \uparrow E(X|\mathcal{F}). \tag{4.1.3}$$

注释. 将最后结果应用于 $Y_1 - Y_n$，可见若 $Y_n \downarrow Y$ 且 $E|Y_1|, E|Y| < \infty$，则 $E(Y_n|\mathcal{F}) \downarrow E(Y|\mathcal{F})$.

证明. 为证 (a)，需检验右边是左边的一个版本. 它显然是 $\mathcal{F}$ 可测的. 为检验 (ii)，观察到若 $A \in \mathcal{F}$，则由积分的线性性及 $E(X|\mathcal{F})$ 和 $E(Y|\mathcal{F})$ 的定义性质，

$$\begin{aligned}\int_A \{aE(X|\mathcal{F}) + E(Y|\mathcal{F})\}\,\mathrm{d}P &= a\int_A E(X|\mathcal{F})\,\mathrm{d}P + \int_A E(Y|\mathcal{F})\,\mathrm{d}P \\ &= a\int_A X\,\mathrm{d}P + \int_A Y\,\mathrm{d}P = \int_A (aX + Y)\,\mathrm{d}P,\end{aligned}$$

证得 (4.1.1).

利用定义

$$\int_A E(X|\mathcal{F})\,\mathrm{d}P = \int_A X\,\mathrm{d}P \leqslant \int_A Y\,\mathrm{d}P = \int_A E(Y|\mathcal{F})\,\mathrm{d}P.$$

令 $A = \{E(X|\mathcal{F}) - E(Y|\mathcal{F}) \geqslant \varepsilon > 0\}$，可见所指集合对所有 $\varepsilon > 0$ 概率为 0，证得 (4.1.2).

令 $Y_n = X - X_n$. 只需证明 $E(Y_n|\mathcal{F}) \downarrow 0$. 因 $Y_n \downarrow$，由 (4.1.2) 知 $Z_n \equiv E(Y_n|\mathcal{F}) \downarrow$ 极限 Z_∞. 若 $A \in \mathcal{F}$ 则

$$\int_A Z_n\,\mathrm{d}P = \int_A Y_n\,\mathrm{d}P.$$

令 $n \to \infty$，注意 $Y_n \downarrow 0$，并利用控制收敛定理得 $\int_A Z_\infty\,\mathrm{d}P = 0$ 对所有 $A \in \mathcal{F}$ 成立，故 $Z_\infty \equiv 0$. □

定理 4.1.10. 若 φ 凸且 $E|X|, E|\varphi(X)| < \infty$ 则

$$\varphi(E(X|\mathcal{F})) \leqslant E(\varphi(X)|\mathcal{F}). \tag{4.1.4}$$

证明. 若 φ 线性，结果是平凡的，故设 φ 非线性. 这样做是为了若令 $S = \{(a, b) : a, b \in \mathbb{Q}, ax + b \leqslant \varphi(x)$ 对所有$x\}$，则 $\varphi(x) = \sup\{ax + b : (a, b) \in S\}$. 更多细节见定理 1.6.2 的证明. 若 $\varphi(x) \geqslant ax + b$,

则由 (4.1.2) 和 (4.1.1) 得

$$E(\varphi(X)|\mathcal{F}) \geqslant aE(X|\mathcal{F}) + b \quad \text{几乎必然.}$$

取上确界（关于 $(a,b) \in S$）得

$$E(\varphi(X)|\mathcal{F}) \geqslant \varphi(E(X|\mathcal{F})) \quad \text{几乎必然,}$$

证得所需结果. □

注释. 这里在不平等旁写了 "几乎必然" 以强调对每个 a, b 都有例外集，故必须取一个可数集的上确界.

定理 4.1.11. 条件期望是 L^p 中的压缩映射，$p \geqslant 1$.

证明. (4.1.4) 蕴含 $|E(X|\mathcal{F})|^p \leqslant E(|X|^p|\mathcal{F})$. 取期望值得

$$E(|E(X|\mathcal{F})|^p) \leqslant E(E(|X|^p|\mathcal{F})) = E|X|^p.$$

在最后一个等式中，使用了一个恒等式，它是定义的直接推论（在定义中用性质 (ii) 且 $A = \Omega$）:

$$E(E(Y|\mathcal{F})) = E(Y). \tag{4.1.5}$$

□

条件期望还具有一些 "普通" 期望所没有的性质.

定理 4.1.12. 若 $\mathcal{F} \subset \mathcal{G}$ 且 $E(X|\mathcal{G}) \in \mathcal{F}$ 则 $E(X|\mathcal{F}) = E(X|\mathcal{G})$.

证明. 由假设 $E(X|\mathcal{G}) \in \mathcal{F}$. 为检验定义的另一部分，注意到若 $A \in \mathcal{F} \subset \mathcal{G}$ 则

$$\int_A X \,\mathrm{d}P = \int_A E(X|\mathcal{G}) \,\mathrm{d}P.$$

□

定理 4.1.13. 若 $\mathcal{F}_1 \subset \mathcal{F}_2$ 则

(i) $E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$,

(ii) $E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1)$.

换言之，较小的 σ-域总是取胜. 如证明将显示的，第一个等式是平凡的. 第二个容易证明，但与定理 4.1.14 结合是计算条件期望的有力工具，也见过它多次被误用.

证明. 一旦注意到 $E(X|\mathcal{F}_1) \in \mathcal{F}_2$，(i) 由例 4.1.3 推出. 为证 (ii)，注意到 $E(X|\mathcal{F}_1) \in \mathcal{F}_1$，且若 $A \in \mathcal{F}_1 \subset \mathcal{F}_2$ 则

$$\int_A E(X|\mathcal{F}_1)\,\mathrm{d}P = \int_A X\,\mathrm{d}P = \int_A E(X|\mathcal{F}_2)\,\mathrm{d}P.$$

□

下述结果表明，对于关于 $\mathcal{F}$ 的条件期望，$X \in \mathcal{F}$ 的随机变量如同常数，可以被拿到 “积分” 外面.

定理 4.1.14. 若 $X \in \mathcal{F}$ 且 $E|Y|, E|XY| < \infty$ 则

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F}).$$

证明. 右边 $\in \mathcal{F}$，故需检验 (ii). 为此，使用通常的四步法. 首先，设 $X = 1_B$ 且 $B \in \mathcal{F}$. 在此情形下，若 $A \in \mathcal{F}$

$$\int_A 1_B E(Y|\mathcal{F})\,\mathrm{d}P = \int_{A\cap B} E(Y|\mathcal{F})\,\mathrm{d}P = \int_{A\cap B} Y\,\mathrm{d}P = \int_A 1_B Y\,\mathrm{d}P,$$

故 (ii) 成立. 最后一个结果通过线性性推广到简单 X. 若 $X, Y \geqslant 0$，令 X_n 为简单随机变量且 $X_n \uparrow X$，并用单调收敛定理得

$$\int_A XE(Y|\mathcal{F})\,\mathrm{d}P = \int_A XY\,\mathrm{d}P.$$

为证一般情形的结果，将 X 和 Y 分成它们的正部和负部. □

定理 4.1.15. 设 $EX^2 < \infty$. $E(X|\mathcal{F})$ 是 $Y \in \mathcal{F}$ 中使 “均方误差”$E(X-Y)^2$ 最小的变量.

注释. 这一结果给出了 $E(X|\mathcal{F})$ 的 “几何解释”. $L^2(\mathcal{F}_0) = \{Y \in \mathcal{F}_0 : EY^2 < \infty\}$ 是 Hilbert 空间，$L^2(\mathcal{F})$ 是其闭子空间. 在此情形下，$E(X|\mathcal{F})$ 是 X 到 $L^2(\mathcal{F})$ 的投影，即子空间中最接近 X 的点.

证明. 首先观察到若 $Z \in L^2(\mathcal{F})$，则由定理 4.1.14 得

$$ZE(X|\mathcal{F}) = E(ZX|\mathcal{F}).$$

($E|XZ| < \infty$ 由 Cauchy-Schwarz 不等式.) 取期望值得

$$E(ZE(X|\mathcal{F})) = E(E(ZX|\mathcal{F})) = E(ZX),$$

或重排得

$$E[Z(X - E(X|\mathcal{F}))] = 0 \quad \text{对} Z \in L^2(\mathcal{F}).$$

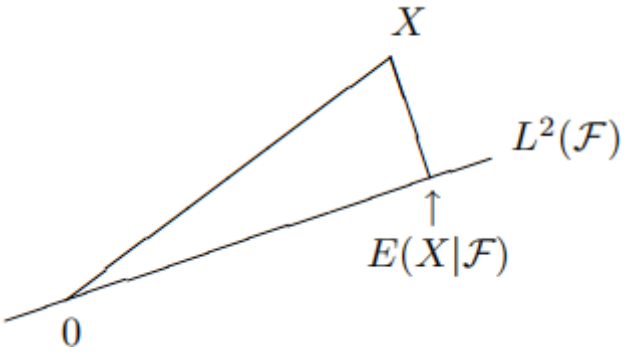

图 4.1: 条件期望作为 L^2 中的投影

若 $Y \in L^2(\mathcal{F})$ 且 $Z = E(X|\mathcal{F}) - Y$ 则

$$E(X-Y)^2 = E\{X - E(X|\mathcal{F}) + Z\}^2 = E\{X - E(X|\mathcal{F})\}^2 + EZ^2,$$

因交叉项为零. 由最后一个公式，易见 $E(X-Y)^2$ 在 $Z=0$ 时最小. □

4.1.3 正则条件概率 *

设 $(\Omega, \mathcal{F}, P)$ 为概率空间，$X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ 为可测映射，$\mathcal{G}$ 为 $\mathcal{F}$ 的子 σ-域. $\mu : \Omega \times \mathcal{S} \to [0,1]$ 称为给定 $\mathcal{G}$ 的 X 的**正则条件分布**，如果

(i) 对每个 A，$\omega \mapsto \mu(\omega, A)$ 是 $P(X \in A|\mathcal{G})$ 的一个版本.

(ii) 对几乎必然 ω，$A \mapsto \mu(\omega, A)$ 是 $(S, \mathcal{S})$ 上的概率测度.

当 $S = \Omega$ 且 X 是恒等映射时，μ 称为**正则条件概率**.

例 4.1.6 的延续. 设 X 和 Y 有联合密度 $f(x,y) > 0$. 若

$$\mu(y, A) = \int_A f(x,y)\,\mathrm{d}x \Big/ \int f(x,y)\,\mathrm{d}x,$$

则 $\mu(Y(\omega), A)$ 是给定 $\sigma(Y)$ 的 X 的正则条件分布（r.c.d.）.

定义中的 (i) 由在例 4.1.6 中取 $h = 1_A$ 推出. 为检验 (ii)，注意控制收敛定理蕴含 $A \mapsto \mu(y, A)$ 是概率测度.

正则条件分布的用处在于，它们允许同时计算 X 的所有函数的条件期望，并以更直接的方式推广普通期望的性质.

定理 4.1.16. 设 $\mu(\omega, A)$ 为给定 $\mathcal{F}$ 的 X 的正则条件分布. 若 $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{R})$ 满足 $E|f(X)| < \infty$ 则

$$E(f(X)|\mathcal{F}) = \int \mu(\omega, \mathrm{d}x) f(x) \quad \text{几乎必然.}$$

证明. 若 $f = 1_A$，这由定义推出. 线性性将结果推广到简单 f，单调收敛到非负 f. 最后通过写 $f = f^+ - f^-$ 得到一般情形的结果. □

不幸的是，正则条件分布并不总是存在. 第一个例子归功于 Dieudonné (1948). 更多近期进展见 Doob (1953), p. 624, 或 Faden (1985). 不深入例子的细节，问题的来源很容易看出. 若 $A_1, A_2, \ldots$ 不交，则由 (4.1.1) 和 (4.1.3) 得

$$P\left(X \in \bigcup_n A_n \middle| \mathcal{G}\right) = \sum_n P(X \in A_n|\mathcal{G}) \quad \text{几乎必然,}$$

但若 S 包含足够多的可数不交集族，例外集可能堆积起来. 幸运的是，

定理 4.1.17. 若 $(S, \mathcal{S})$ 是好空间，则正则条件分布存在.

证明. 由定义，存在 1-1 映射 $\phi : S \to \mathbb{R}$ 使得 ϕ 和 ϕ^{-1} 可测. 利用单调性 (4.1.2) 并去掉可数零测集，发现有集合 Ω_o 满足 $P(\Omega_o) = 1$ 以及随机变量族 $G(q, \omega)$，$q \in \mathbb{Q}$ 使得 $q \mapsto G(q, \omega)$ 非减且 $\omega \mapsto G(q, \omega)$ 是 $P(\phi(X) \leqslant q|\mathcal{G})$ 的一个版本. 令 $F(x, \omega) = \inf\{G(q, \omega) : q > x\}$. 此记号可能使读者想起定理 3.2.12 的证明. 那里给出的论证表明 F 是分布函数. 因 $G(q_n, \omega) \downarrow F(x, \omega)$，定理 4.1.9 后的注记蕴含 $F(x, \omega)$ 是 $P(\phi(X) \leqslant x|\mathcal{G})$ 的一个版本.

现在，对每个 $\omega \in \Omega_o$，存在唯一测度 $\nu(\omega, \cdot)$ 在 $(\mathbb{R}, \mathcal{R})$ 上使得 $\nu(\omega, (-\infty, x]) = F(x, \omega)$. 为检验对每个 $B \in \mathcal{R}$，$\nu(\omega, B)$ 是 $P(\phi(X) \in B|\mathcal{G})$ 的一个版本，观察到使这一陈述成立的 B 类（这包括 $\omega \mapsto \nu(\omega, B)$ 的可测性）是包含所有形如 $(a_1, b_1] \cup \cdots \cup (a_k, b_k]$（其中 $-\infty \leqslant a_i < b_i \leqslant \infty$）的集合的 λ-系，故所需结果由 $\pi - \lambda$ 定理推出.

为提取所需的正则条件分布，注意若 $A \in \mathcal{S}$ 且 $B = \phi(A)$，则 $B = (\phi^{-1})^{-1}(A) \in \mathcal{R}$，并令 $\mu(\omega, A) = \nu(\omega, B)$. □

定理 4.1.17 的以下推广在 6.1 节中需要.

定理 4.1.18. 设 X 和 Y 取值于好空间 $(S, \mathcal{S})$ 且 $\mathcal{G} = \sigma(Y)$. 存在函数 $\mu : S \times \mathcal{S} \to [0, 1]$ 使得

(i) 对每个 A，$\mu(Y(\omega), A)$ 是 $P(X \in A|\mathcal{G})$ 的一个版本，

(ii) 对几乎必然 ω，$A \mapsto \mu(Y(\omega), A)$ 是 $(S, \mathcal{S})$ 上的概率测度.

证明. 如在定理 4.1.17 的证明中，发现有集合 Ω_o 满足 $P(\Omega_o) = 1$ 以及随机变量族 $G(q, \omega)$，$q \in \mathbb{Q}$ 使得 $q \mapsto G(q, \omega)$ 非减且 $\omega \mapsto G(q, \omega)$ 是 $P(\phi(X) \leqslant q|\mathcal{G})$ 的一个版本. 因 $G(q, \omega) \in \sigma(Y)$，可以写 $G(q, \omega) = H(q, Y(\omega))$. 令 $F(x, y) = \inf\{G(q, y) : q > x\}$. 定理 4.1.17 证明中给出的论证表明存在集合 A_0 满足 $P(Y \in A_0) = 1$ 使得当 $y \in A_0$ 时，F 是分布函数且 $F(x, Y(\omega))$ 是 $P(\phi(X) \leqslant x|Y)$ 的一个版本.

对每个 $y \in A_o$，存在唯一测度 $\nu(y, \cdot)$ 在 $(\mathbb{R}, \mathcal{R})$ 上使得 $\nu(y, (-\infty, x]) = F(x, y)$. 为检验对每个 $B \in \mathcal{R}$，$\nu(Y(\omega), B)$ 是 $P(\phi(X) \in B|Y)$ 的一个版本，观察到使这一陈述成立的 B 类（这包括

$\omega \mapsto \nu(Y(\omega), B)$ 的可测性）是包含所有形如 $(a_1, b_1] \cup \cdots \cup (a_k, b_k]$（其中 $-\infty \leqslant a_i < b_i \leqslant \infty$）的集合的 λ-系，故所需结果由 $\pi - \lambda$ 定理推出.

为提取所需的正则条件分布，注意若 $A \in \mathcal{S}$ 且 $B = \phi(A)$，则 $B = (\phi^{-1})^{-1}(A) \in \mathcal{R}$，并令 $\mu(y, A) = \nu(y, B)$. □

习题

习题 4.1.1. Bayes 公式. 设 $G \in \mathcal{G}$ 并证明

$$P(G|A) = \int_G P(A|G)\,\mathrm{d}P \Big/ \int_\Omega P(A|G)\,\mathrm{d}P.$$

当 $\mathcal{G}$ 是由划分生成的 σ-域时，这化为通常的 Bayes 公式

$$P(G_i|A) = P(A|G_i)P(G_i) \Big/ \sum_j P(A|G_j)P(G_j).$$

习题 4.1.2. 证明 Chebyshev 不等式. 若 $a > 0$ 则

$$P(|X| \geqslant a|\mathcal{F}) \leqslant a^{-2}E(X^2|\mathcal{F}).$$

习题 4.1.3. 模仿定理 1.5.2 后注记中的证明来证明条件 Cauchy-Schwarz 不等式.

$$E(XY|\mathcal{G})^2 \leqslant E(X^2|\mathcal{G})E(Y^2|\mathcal{G}).$$

习题 4.1.4. 利用正则条件概率从无条件的得到条件的 Hölder 不等式，即证明若 $p, q \in (1, \infty)$ 且 $1/p + 1/q = 1$ 则

$$E(|XY||\mathcal{G}) \leqslant E(|X|^p|\mathcal{G})^{1/p}E(|Y|^q|\mathcal{G})^{1/q}.$$

习题 4.1.5. 在 $\Omega = \{a, b, c\}$ 上给出一个例子使得

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) \neq E(E(X|\mathcal{F}_2)|\mathcal{F}_1).$$

习题 4.1.6. 证明若 $\mathcal{G} \subset \mathcal{F}$ 且 $EX^2 < \infty$ 则

$$E(\{X - E(X|\mathcal{F})\}^2) + E(\{E(X|\mathcal{F}) - E(X|\mathcal{G})\}^2) = E(\{X - E(X|\mathcal{G})\}^2).$$

去掉左边第二项，得到一个不等式，它几何上表示子空间越大投影越近，或统计上表示更多信息意味着更小的均方误差.

习题 4.1.7. 上题结果的一个重要特殊情形是当 $\mathcal{G} = \{\emptyset, \Omega\}$ 时. 令 $\text{var}(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$. 证明

$$\text{var}(X) = E(\text{var}(X|\mathcal{F})) + \text{var}(E(X|\mathcal{F})).$$

习题 4.1.8. 设 $Y_1, Y_2, \ldots$ 独立同分布，均值为 μ，方差为 σ^2，N 为与 Y_i 独立的正整数值随机变量且 $EN^2 < \infty$，$X = Y_1 + \cdots + Y_N$. 证明 $\text{var}(X) = \sigma^2 EN + \mu^2 \text{var}(N)$. 为理解和帮助记住此公式，考虑 N 或 Y 为常数的两种特殊情形.

习题 4.1.9. 证明若 X 和 Y 是随机变量满足 $E(Y|\mathcal{G}) = X$ 且 $EY^2 = EX^2 < \infty$，则 $X = Y$ 几乎必然.

习题 4.1.10. 上题结果蕴含若 $EY^2 < \infty$ 且 $E(Y|\mathcal{G})$ 与 Y 同分布，则 $E(Y|\mathcal{G}) = Y$ 几乎必然. 在假设 $E|Y| < \infty$ 下证明这一点. 提示：技巧是证明 $\text{sgn}(X) = \text{sgn}(E(X|\mathcal{G}))$ 几乎必然，然后取 $X = Y - c$ 得到所需结果.

第 2 节 鞅，几乎必然收敛

本节定义鞅及其相关概念——上鞅和下鞅，并初步发展其理论. 设 $\mathcal{F}_n$ 为 **σ-域流**，即递增的 σ-域序列. 称序列 X_n 是**适应于** $\mathcal{F}_n$ 的，若对所有 n 有 $X_n \in \mathcal{F}_n$. 若序列 X_n 满足

(i) $E|X_n| < \infty$,

(ii) X_n 适应于 $\mathcal{F}_n$,

(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ 对所有 n 成立,

则称 X 为**鞅**（关于 $\mathcal{F}_n$）. 若在上述定义的最后一个条件中将 $=$ 替换为 $\leqslant$ 或 $\geqslant$，则分别称 X 为**上鞅**或**下鞅**.

首先给出三个与随机游走相关的例子. 设 $\xi_1, \xi_2, \ldots$ 独立同分布. 令 $S_n = S_0 + \xi_1 + \cdots + \xi_n$，其中 S_0 为常数. 令 $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$（$n \geqslant 1$）且取 $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

例题 4.2.1. 线性鞅. 若 $\mu = E\xi_i = 0$，则 S_n（$n \geqslant 0$）是关于 $\mathcal{F}_n$ 的鞅.

证明. 注意 $S_n \in \mathcal{F}_n$，$E|S_n| < \infty$，且 ξ_{n+1} 独立于 $\mathcal{F}_n$. 由条件期望的线性性 (4.1.1) 和例 4.1.4，

$$E(S_{n+1}|\mathcal{F}_n) = E(S_n|\mathcal{F}_n) + E(\xi_{n+1}|\mathcal{F}_n) = S_n + E\xi_{n+1} = S_n. \qquad \square$$

若 $\mu \leqslant 0$，则上述计算表明 $E(X_{n+1}|\mathcal{F}_n) \leqslant X_n$，即 X_n 是上鞅. 此时 X_n 对应于**不利博弈**中的押注，因此上鞅并非" 超级". 这一名称源于如下事实：若 f 是上调和的（即 f 具有不超过二阶的连续导数且 $\partial^2 f/\partial x_1^2 + \cdots + \partial^2 f/\partial x_d^2 \leqslant 0$），则

$$f(x) \geqslant \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\,\mathrm{d}y, \tag{4.2.1}$$

其中 $B(x,r) = \{y : |x-y| \leqslant r\}$ 为半径为 r 的球，$|B(x,r)|$ 是该球的体积.

若 $\mu \geqslant 0$，则 S_n 是下鞅. 将第一个结果应用于 $\xi_i' = \xi_i - \mu$，可见 $S_n - n\mu$ 是鞅.

例题 4.2.2. 二次鞅. 现设 $\mu = E\xi_i = 0$ 且 $\sigma^2 = \mathrm{var}(\xi_i) < \infty$. 此时 $S_n^2 - n\sigma^2$ 是鞅.

证明. 由于 $(S_n + \xi_{n+1})^2 = S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2$ 且 ξ_{n+1} 独立于 $\mathcal{F}_n$，有

$$\begin{aligned} E(S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n) &= S_n^2 + 2S_nE(\xi_{n+1}|\mathcal{F}_n) + E(\xi_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2 \\ &= S_n^2 + 0 + \sigma^2 - (n+1)\sigma^2 = S_n^2 - n\sigma^2. \end{aligned} \qquad \square$$

例题 4.2.3. 指数鞅. 设 $Y_1, Y_2, \ldots$ 为非负独立同分布随机变量且 $EY_m = 1$. 若 $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$，则 $M_n = \prod_{m \leqslant n} Y_m$ 定义了一个鞅. 为证明这一点，注意

$$E(M_{n+1}|\mathcal{F}_n) = M_nE(Y_{n+1}|\mathcal{F}_n) = M_n.$$

现设 $Y_i = e^{\theta\xi_i}$ 且 $\varphi(\theta) = Ee^{\theta\xi_i} < \infty$. $Y_i = \exp(\theta\xi)/\varphi(\theta)$ 的均值为 1，故 $EY_i = 1$ 且

$$M_n = \prod_{i=1}^{n} Y_i = \exp(\theta S_n)/\varphi(\theta)^n$$

是鞅.

下面将看到更多例子，因此现在转向推导鞅的性质. 第一个结果是上鞅定义的直接推论. 可将该结果的结论作为上鞅的定义，但那样定义将更难验证.

定理 4.2.4. 若 X_n 是上鞅，则当 $n > m$ 时，$E(X_n|\mathcal{F}_m) \leqslant X_m$.

证明. 定义直接给出 $n = m + 1$ 的情形. 设 $n = m + k$（$k \geqslant 2$）. 由定理 4.1.2，

$$E(X_{m+k}|\mathcal{F}_m) = E(E(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \leqslant E(X_{m+k-1}|\mathcal{F}_m),$$

其中不等式由定义和 (4.1.2) 得到. 再由归纳法即得所需结果. □

定理 4.2.5. (i) 若 X_n 是下鞅，则当 $n > m$ 时，$E(X_n|\mathcal{F}_m) \geqslant X_m$.
(ii) 若 X_n 是鞅，则当 $n > m$ 时，$E(X_n|\mathcal{F}_m) = X_m$.

证明. 为证 (i)，注意 $-X_n$ 是上鞅并利用 (4.1.1). 对于 (ii)，注意 X_n 既是上鞅又是下鞅. □

注释. 定理 4.2.5 证明中的思想将在下文多次使用. 为避免重复，只对上鞅或下鞅之一陈述结果，而将翻译为另外两者的结果留给读者.

定理 4.2.6. 设 X_n 是关于 $\mathcal{F}_n$ 的鞅，φ 是凸函数且对所有 n 满足 $E|\varphi(X_n)| < \infty$，则 $\varphi(X_n)$ 是关于 $\mathcal{F}_n$ 的下鞅. 因此，若 $p \geqslant 1$ 且对所有 n 有 $E|X_n|^p < \infty$，则 $|X_n|^p$ 是关于 $\mathcal{F}_n$ 的下鞅.

证明. 由 Jensen 不等式和定义，

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \geqslant \varphi(E(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n).$$ □

定理 4.2.7. 设 X_n 是关于 $\mathcal{F}_n$ 的下鞅，φ 是递增凸函数且对所有 n 满足 $E|\varphi(X_n)| < \infty$，则 $\varphi(X_n)$ 是关于 $\mathcal{F}_n$ 的下鞅. 因此
(i) 若 X_n 是下鞅，则 $(X_n - a)^+$ 是下鞅.
(ii) 若 X_n 是上鞅，则 $X_n \wedge a$ 是上鞅.

证明. 由 Jensen 不等式和假设，

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \geqslant \varphi(E(X_{n+1}|\mathcal{F}_n)) \geqslant \varphi(X_n).$$ □

设 $\mathcal{F}_n$（$n \geqslant 0$）为 σ-域流. 称 H_n（$n \geqslant 1$）为**可料序列**，若对所有 $n \geqslant 1$ 有 $H_n \in \mathcal{F}_{n-1}$. 换言之，$H_n$ 的值可由时刻 $n-1$ 已知的信息（确定性地）预见.

本节中，将 H_n 理解为赌徒在时刻 n 押注的金额. 这可以基于时刻 $1, \ldots, n-1$ 的结果，但不能基于时刻 n 的结果!

一旦将 H_n 视为博弈系统，自然会问使用它会赚多少钱. 设 X_n 为每次押注 1 美元时到时刻 n 的净收益. 若按照博弈系统 H 押注，则时刻 n 的收益为

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m (X_m - X_{m-1}),$$

因为若在时刻 m 押注 3 美元，则财富的变化将是押注 1 美元的人的变化的 3 倍. 或者，可以将 X_m 视为股票价值，H_m 为从时刻 $m-1$ 到时刻 m 持有的股数.

现设 $\xi_m = X_m - X_{m-1}$ 满足 $P(\xi_m = 1) = p$ 且 $P(\xi_m = -1) = 1 - p$. 一个著名的博弈系统称为”鞅”（martingale），由 $H_1 = 1$ 且对 $n \geqslant 2$ 定义为：

$$H_n = \begin{cases} 2H_{n-1} & \text{若} \xi_{n-1} = -1, \\ 1 & \text{若} \xi_{n-1} = 1 \end{cases}$$

换言之，输时加倍下注，这样若连输 k 次然后赢，净收益将为 1. 为说明这一点，考虑如下具体情况：

H_n	1	2	4	8	16
ξ_n	-1	-1	-1	-1	1
$(H \cdot X)_n$	-1	-3	-7	-15	1

只要 $P(\xi_m = 1) > 0$，这个系统似乎提供了一种” 稳赚” 的方法. 然而，下述结果表明，不存在能够击败不利博弈的系统.

定理 4.2.8. 设 X_n（$n \geqslant 0$）为上鞅. 若 $H_n \geqslant 0$ 为可料序列且每个 H_n 有界，则 $(H \cdot X)_n$ 是上鞅.

证明. 利用条件期望的线性性，$(H \cdot X)_n \in \mathcal{F}_n$，$H_n \in \mathcal{F}_{n-1}$，以及 (4.1.14)，有

$$\begin{aligned} E((H \cdot X)_{n+1} | \mathcal{F}_n) &= (H \cdot X)_n + E(H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n) \\ &= (H \cdot X)_n + H_{n+1} E((X_{n+1} - X_n) | \mathcal{F}_n) \leqslant (H \cdot X)_n, \end{aligned}$$

因为 $E((X_{n+1} - X_n) | \mathcal{F}_n) \leqslant 0$ 且 $H_{n+1} \geqslant 0$. □

注释. 同样的结论对下鞅和鞅也成立（后一种情形无需限制 $H_n \geqslant 0$）.

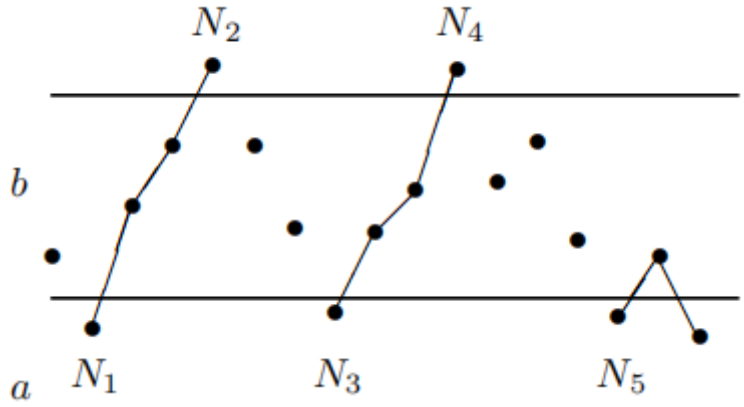

图 4.2: (a,b) 中的上穿，线段代表 $(H\cdot X)_n$ 中的增长，在 Y_n 中小于 a 的点被抬升到 a

下面考虑一个非常特殊的博弈系统：在时刻 $n\leqslant N$ 每次押注 1 美元然后停止. 称随机变量 N 为**停时**，若对所有 $n<\infty$ 有 $\{N=n\}\in\mathcal{F}_n$，即在时刻 n 停止的决定必须关于该时刻已知的信息可测. 若令 $H_n=1_{\{N\geqslant n\}}$，则 $\{N\geqslant n\}=\{N\leqslant n-1\}^c\in\mathcal{F}_{n-1}$，故 H_n 是可料的. 由定理 4.2.8 可知 $(H\cdot X)_n=X_{N\wedge n}-X_0$ 是上鞅. 由于常数序列 $Y_n=X_0$ 是上鞅且两个上鞅之和也是上鞅，故得：

定理 4.2.9. 若 N 是停时且 X_n 是上鞅，则 $X_{N\wedge n}$ 是上鞅.

虽然定理 4.2.8 意味着无法用博弈系统赚钱，但可用其证明定理. 设 X_n（$n\geqslant 0$）为下鞅. 设 $a<b$，令 $N_0=-1$，且对 $k\geqslant 1$ 令

$$N_{2k-1}=\inf\{m>N_{2k-2}:X_m\leqslant a\},$$
$$N_{2k}=\inf\{m>N_{2k-1}:X_m\geqslant b\}.$$

N_j 均为停时且 $\{N_{2k-1}<m\leqslant N_{2k}\}=\{N_{2k-1}\leqslant m-1\}\cap\{N_{2k}\leqslant m-1\}^c\in\mathcal{F}_{m-1}$，故

$$H_m=\begin{cases}1 & \text{若对某个}k\text{有}N_{2k-1}<m\leqslant N_{2k},\\ 0 & \text{其他}\end{cases}$$

定义了一个可料序列. $X(N_{2k-1})\leqslant a$ 且 $X(N_{2k})\geqslant b$，故在时刻 N_{2k-1} 和 N_{2k} 之间，X_m 从 a 下方穿越到 b 上方. H_m 是一个试图利用这些”**上穿**”获利的博弈系统. 用股市术语，在 $X_m\leqslant a$ 时买入，在 $X_m\geqslant b$ 时卖出，因此每次完成一次上穿，就获得至少 $(b-a)$ 的利润. 最后，$U_n=\sup\{k:N_{2k}\leqslant n\}$ 是到时刻 n 完成的上穿次数.

定理 4.2.10. 上穿不等式. 若 X_m（$m\geqslant 0$）是下鞅，则

$$(b-a)EU_n\leqslant E(X_n-a)^+-E(X_0-a)^+.$$

证明. 令 $Y_m = a + (X_m - a)^+$. 由定理 4.2.7，Y_m 是下鞅. 显然，它上穿 $[a,b]$ 的次数与 X_m 相同，且 $(b-a)U_n \leqslant (H\cdot Y)_n$，因为每次上穿产生至少 $(b-a)$ 的利润，而最后的未完成上穿（如果有的话）对右边有非负贡献. 正是出于这一原因，需要将 X_m 替换为 Y_m.

令 $K_m = 1 - H_m$. 显然 $Y_n - Y_0 = (H\cdot Y)_n + (K\cdot Y)_n$，由定理 4.2.8 知 $E(K\cdot Y)_n \geqslant E(K\cdot Y)_0 = 0$，故 $E(H\cdot Y)_n \leqslant E(Y_n - Y_0)$，证得所需不等式. □

上述结果以经典形式给出，尽管这有些误导. 关键事实是 $E(K\cdot Y)_n \geqslant 0$，即无论如何努力，押注下鞅都不会输钱. 由上穿不等式，容易得到

定理 4.2.11. 鞅收敛定理. 若 X_n 是满足 $\sup EX_n^+ < \infty$ 的下鞅，则当 $n \to \infty$ 时，X_n 几乎必然收敛于极限 X 且 $E|X| < \infty$.

证明. 由于 $(X-a)^+ \leqslant X^+ + |a|$，定理 4.2.10 意味着

$$EU_n \leqslant (|a| + EX_n^+)/(b-a).$$

当 $n \uparrow \infty$ 时，$U_n \uparrow U$ 为整个序列上穿 $[a,b]$ 的次数，故若 $\sup EX_n^+ < \infty$ 则 $EU < \infty$，从而 $U < \infty$ 几乎必然. 由最后一个结论对所有有理数 a 和 b 成立，

$$\bigcup_{a,b\in\mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}$$

概率为 0，故 $\limsup X_n = \liminf X_n$ 几乎必然，即 $\lim X_n$ 几乎必然存在. 由 Fatou 引理，$EX^+ \leqslant \liminf EX_n^+ < \infty$，故 $X < \infty$ 几乎必然. 为证 $X > -\infty$，注意

$$EX_n^- = EX_n^+ - EX_n \leqslant EX_n^+ - EX_0$$

（因为 X_n 是下鞅），故再由 Fatou 引理得

$$EX^- \leqslant \liminf_{n\to\infty} EX_n^- \leqslant \sup_n EX_n^+ - EX_0 < \infty,$$

证毕. □

注释. 为准备定理 4.7.1 的证明，读者应注意，上面已证明若 X_n 上穿 (a,b) 的次数对所有 $a,b\in\mathbb{Q}$ 有限，则 X_n 的极限存在.

定理 4.2.11 的一个重要特例是

定理 4.2.12. 若 $X_n \geqslant 0$ 是上鞅，则当 $n \to \infty$ 时，$X_n \to X$ 几乎必然且 $EX \leqslant EX_0$.

证明. $Y_n = -X_n \leqslant 0$ 是满足 $EY_n^+ = 0$ 的下鞅. 由于 $EX_0 \geqslant EX_n$，不等式由 Fatou 引理得出. □

下一节将给出最后两个结果的若干应用. 本节最后用两个" 反例" 作结.

例题 4.2.13. 第一个反例表明，定理 4.2.12（或 4.2.11）的假设不能保证 L^1 收敛. 设 S_n 为对称简单随机游走且 $S_0 = 1$，即 $S_n = S_{n-1} + \xi_n$，其中 $\xi_1, \xi_2, \ldots$ 独立同分布，$P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. 令 $N = \inf\{n : S_n = 0\}$ 且 $X_n = S_{N\wedge n}$. 定理 4.2.9 蕴含 X_n 是非负鞅. 定理 4.2.12 蕴含 X_n 收敛于极限 $X_\infty < \infty$，它必恒为 0，因为收敛于 $k > 0$ 是不可能的.（若 $X_n = k > 0$ 则 $X_{n+1} = k \pm 1$.）由于对所有 n 有 $EX_n = EX_0 = 1$ 而 $X_\infty = 0$，故收敛不可能在 L^1 中发生.

例 4.2.13 是一个重要反例，在阅读本章其余部分时应牢记. 下一个反例没那么重要.

例题 4.2.14. 现给出一个依概率收敛于 0 但不几乎必然收敛的鞅的例子. 令 $X_0 = 0$. 当 $X_{k-1} = 0$ 时，令 $X_k = 1$ 或 -1 的概率各为 $1/2k$，$= 0$ 的概率为 $1 - 1/k$. 当 $X_{k-1} \neq 0$ 时，令 $X_k = kX_{k-1}$ 的概率为 $1/k$，$= 0$ 的概率为 $1 - 1/k$. 由构造，$P(X_k = 0) = 1 - 1/k$，故 $X_k \to 0$ 依概率. 另一方面，第二 Borel-Cantelli 引理蕴含 $P(X_k = 0$ 对所有$k \geqslant K) = 0$，且 $(-1, 1) - \{0\}$ 中的值不可能出现，故 X_k 不几乎必然收敛于 0.

习题

习题 4.2.1. 设 X_n 是关于 $\mathcal{G}_n$ 的鞅且令 $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$，则 $\mathcal{G}_n \supset \mathcal{F}_n$ 且 X_n 是关于 $\mathcal{F}_n$ 的鞅.

习题 4.2.2. 给出一个下鞅 X_n 的例子，使得 X_n^2 是上鞅. 提示：X_n 不必是随机的.

习题 4.2.3. 推广定理 4.2.7 的 (i)，证明若 X_n 和 Y_n 均为关于 $\mathcal{F}_n$ 的下鞅，则 $X_n \vee Y_n$ 也是.

习题 4.2.4. 设 X_n（$n \geqslant 0$）为满足 $\sup X_n < \infty$ 的下鞅. 令 $\xi_n = X_n - X_{n-1}$ 且设 $E(\sup \xi_n^+) < \infty$. 证明 X_n 几乎必然收敛.

习题 4.2.5. 给出一个鞅 X_n 的例子，使得 $X_n \to -\infty$ 几乎必然. 提示：令 $X_n = \xi_1 + \cdots + \xi_n$，其中 ξ_i 独立（但不同分布）且 $E\xi_i = 0$.

习题 4.2.6. 设 $Y_1, Y_2, \ldots$ 为非负独立同分布随机变量，$EY_m = 1$ 且 $P(Y_m = 1) < 1$. 由例 4.2.3 知 $X_n = \prod_{m\leqslant n} Y_m$ 定义了一个鞅.

(i) 利用定理 4.2.12 和反证法证明 $X_n \to 0$ 几乎必然.

(ii) 利用强大数定律得出 $(1/n)\log X_n \to c < 0$.

习题 4.2.7. 设对所有 n 有 $y_n > -1$ 且 $\sum |y_n| < \infty$. 证明 $\prod_{m=1}^{\infty}(1 + y_m)$ 存在.

习题 4.2.8. 设 X_n 和 Y_n 为正可积且适应于 $\mathcal{F}_n$. 设

$$E(X_{n+1}|\mathcal{F}_n) \leqslant (1 + Y_n)X_n$$

且 $\sum Y_n < \infty$ 几乎必然. 通过构造一个与之密切相关的上鞅并应用定理 4.2.12，证明 X_n 几乎必然收敛于有限极限.

习题 4.2.9. 切换原理. 设 X_n^1 和 X_n^2 均为关于 $\mathcal{F}_n$ 的上鞅，N 为停时且满足 $X_N^1 \geqslant X_N^2$. 则

$$Y_n = X_n^1 1_{(N>n)} + X_n^2 1_{(N\leqslant n)} \text{ 是上鞅},$$
$$Z_n = X_n^1 1_{(N\geqslant n)} + X_n^2 1_{(N<n)} \text{ 是上鞅}.$$

习题 4.2.10. Dubins 不等式. 对每个正上鞅 X_n（$n \geqslant 0$），上穿 $[a,b]$ 的次数 U 满足

$$P(U \geqslant k) \leqslant \left(\frac{a}{b}\right)^k E\min(X_0/a, 1).$$

为证明这一点，令 $N_0 = -1$ 且对 $j \geqslant 1$ 令

$$N_{2j-1} = \inf\{m > N_{2j-2} : X_m \leqslant a\},$$
$$N_{2j} = \inf\{m > N_{2j-1} : X_m \geqslant b\}.$$

令 $Y_n = 1$（$0 \leqslant n < N_1$）且对 $j \geqslant 1$

$$Y_n = \begin{cases} (b/a)^{j-1}(X_n/a) & \text{对} N_{2j-1} \leqslant n < N_{2j}, \\ (b/a)^j & \text{对} N_{2j} \leqslant n < N_{2j+1} \end{cases}$$

(i) 利用上一习题的切换原理和归纳法证明 $Z_n^j = Y_{n\wedge N_j}$ 是上鞅.
(ii) 利用 $EY_{n\wedge N_{2k}} \leqslant EY_0$ 并令 $n \to \infty$ 得到 Dubins 不等式.

第 3 节 例子

本节将应用鞅收敛定理来推广第二 Borel-Cantelli 引理, 并研究 Polya 罐子模型、Radon-Nikodym 导数和分支过程. 这四个主题相互独立，按上述顺序展开.

4.3.1 有界增量

首先证明，具有有界增量的鞅要么收敛，要么在 $+\infty$ 与 $-\infty$ 之间振荡.

定理 4.3.1. 设 $X_1, X_2, \ldots$ 为鞅，满足 $|X_{n+1} - X_n| \leqslant M < \infty$. 令

$$C = \{\lim X_n \text{ 存在且有限}\},$$
$$D = \{\limsup X_n = +\infty \text{ 且 } \liminf X_n = -\infty\}.$$

则 $P(C \cup D) = 1$.

证明. 由于 $X_n - X_0$ 是鞅，不失一般性可设 $X_0 = 0$. 令 $0 < K < \infty$，并令 $N = \inf\{n : X_n \leqslant -K\}$. $X_{n\wedge N}$ 是鞅，且 $X_{n\wedge N} \geqslant -K - M$ 几乎必然，因此对 $X_{n\wedge N} + K + M$ 应用定理 4.2.12 可知，在 $\{N = \infty\}$ 上 $\lim X_n$ 存在. 令 $K \to \infty$，可见极限在 $\{\liminf X_n > -\infty\}$ 上存在. 将这一结论应用于 $-X_n$，可知极限在 $\{\limsup X_n < \infty\}$ 上存在. 证毕. □

为应用该结果，先给出

定理 4.3.2. Doob 分解. 任意下鞅 X_n（$n \geqslant 0$）可唯一地表示为 $X_n = M_n + A_n$，其中 M_n 是鞅，A_n 是可料递增序列且 $A_0 = 0$.

证明. 我们希望 $X_n = M_n + A_n$，$E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$，且 $A_n \in \mathcal{F}_{n-1}$. 因此必须有

$$E(X_n|\mathcal{F}_{n-1}) = E(M_n|\mathcal{F}_{n-1}) + E(A_n|\mathcal{F}_{n-1}) = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n,$$

从而

$$A_n - A_{n-1} = E(X_n|\mathcal{F}_{n-1}) - X_{n-1}. \tag{4.3.1}$$

由于 $A_0 = 0$，有

$$A_n = \sum_{m=1}^{n} E(X_n - X_{n-1}|\mathcal{F}_{n-1}). \tag{4.3.2}$$

为验证这一构造有效，注意到 $A_n - A_{n-1} \geqslant 0$（因 X_n 是下鞅）且 $A_n \in \mathcal{F}_{n-1}$. 为证 $M_n = X_n - A_n$ 是鞅，利用 $A_n \in \mathcal{F}_{n-1}$ 和 (4.3.1) 得

$$E(M_n|\mathcal{F}_{n-1}) = E(X_n - A_n|\mathcal{F}_{n-1}) = E(X_n|\mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1},$$

证毕. □

为说明这一结果的用途，我们给出如下重要例子.

例题 4.3.3. 设 $B_n \in \mathcal{F}_n$，利用 (4.3.2) 可得

$$M_n = \sum_{m=1}^{n} \{1_{B_m} - P(B_m|\mathcal{F}_{m-1})\}$$

是鞅.

定理 4.3.4. 第二 Borel-Cantelli 引理, II. 设 $\mathcal{F}_n$ $(n \geqslant 0)$ 为 σ-域流，满足 $\mathcal{F}_0 = \{\emptyset, \Omega\}$，且 B_n $(n \geqslant 1)$ 为事件序列，满足 $B_n \in \mathcal{F}_n$. 则

$$\{B_n \text{ i.o.}\} = \left\{ \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty \right\}.$$

证明. 若令 $X_0 = 0$ 且 $X_n = \sum_{m \leqslant n} 1_{B_m}$，则 X_n 是下鞅. (4.3.2) 蕴含 $A_n = \sum_{m=1}^{n} E(1_{B_m}|\mathcal{F}_{m-1})$，因此若 $M_0 = 0$ 且

$$M_n = \sum_{m=1}^{n} 1_{B_m} - P(B_m|\mathcal{F}_{m-1})$$

对 $n \geqslant 1$ 成立，则 M_n 是鞅，满足 $|M_n - M_{n-1}| \leqslant 1$. 利用定理 4.3.1 的记号，有

$$\text{在 } C \text{ 上}, \quad \sum_{n=1}^{\infty} 1_{B_n} = \infty \text{ 当且仅当} \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty,$$

且在 D 上，$\sum_{n=1}^{\infty} 1_{B_n} = \infty$ 且 $\sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty$. 由于 $P(C \cup D) = 1$，结论得证. □

4.3.2 Polya 罐子模型

罐中最初有 r 个红球和 g 个绿球. 每次从中抽取一球，然后将其放回，并添加 c 个同色球. 设 X_n 为第 n 次抽取后绿球所占比例. 为验证 X_n 是鞅，注意若时刻 n 有 i 个红球和 j 个绿球，则

$$X_{n+1} = \begin{cases} (j+c)/(i+j+c) & \text{以概率} j/(i+j), \\ j/(i+j+c) & \text{以概率} i/(i+j), \end{cases}$$

且有

$$\frac{j+c}{i+j+c} \cdot \frac{j}{i+j} + \frac{j}{i+j+c} \cdot \frac{i}{i+j} = \frac{(j+c+i)j}{(i+j+c)(i+j)} = \frac{j}{i+j}.$$

由于 $X_n \geqslant 0$，定理 4.2.12 蕴含 $X_n \to X_\infty$ 几乎必然. 为计算极限的分布，我们观察到：(a) 前 m 次抽到绿球，接下来 $\ell = n - m$ 次抽到红球的概率为

$$\frac{g}{g+r} \cdot \frac{g+c}{g+r+c} \cdots \frac{g+(m-1)c}{g+r+(m-1)c} \cdot \frac{r}{g+r+mc} \cdots \frac{r+(\ell-1)c}{g+r+(n-1)c},$$

(b) 前 n 次抽取中任意其他包含 m 个绿球和 ℓ 个红球的结果具有相同的概率，因为分母相同而分子只是排列不同. 考虑特例 $c = 1, g = 1, r = 1$. 设 G_n 为第 n 次抽取完成并添加新球后的绿球数. 由 (a) 和 (b) 可知

$$P(G_n = m+1) = \binom{n}{m} \frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1},$$

因此 X_∞ 在 $(0,1)$ 上服从均匀分布.

若设 $c = 1, g = 2, r = 1$，则

$$P(G_n = m+2) = \frac{n!}{m!(n-m)!} \cdot \frac{(m+1)!(n-m)!}{(n+2)!/2} \to 2x$$

当 $n \to \infty$ 且 $m/n \to x$ 时. 一般地，X_∞ 的分布具有密度

$$\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)} x^{(g/c)-1}(1-x)^{(r/c)-1}.$$

这是参数为 g/c 和 r/c 的**贝塔分布**（beta 分布）. 在例 4.5.6 中我们将看到，若除添加 c 个同色球外还总是添加一个异色球，则极限行为会发生剧烈变化.

4.3.3 Radon-Nikodym 导数

设 μ 为有限测度，ν 为概率测度，均定义于 $(\Omega, \mathcal{F})$ 上. 设 $\mathcal{F}_n \uparrow \mathcal{F}$ 为 σ-域（即 $\sigma(\cup \mathcal{F}_n) = \mathcal{F}$）. 设 μ_n 和 ν_n 分别为 μ 和 ν 在 $\mathcal{F}_n$ 上的限制.

定理 4.3.5. 设对所有 n 有 $\mu_n \ll \nu_n$. 令 $X_n = \mathrm{d}\mu_n/\mathrm{d}\nu_n$ 且 $X = \limsup X_n$. 则

$$\mu(A) = \int_A X \,\mathrm{d}\nu + \mu(A \cap \{X = \infty\}).$$

注释. $\mu_r(A) \equiv \int_A X \,\mathrm{d}\nu$ 是 $\ll \nu$ 的测度. 由于定理 4.2.12 蕴含 $\nu(X = \infty) = 0$，$\mu_s(A) \equiv \mu(A \cap \{X = \infty\})$ 关于 ν **奇异**. 因此 $\mu = \mu_r + \mu_s$ 给出 μ 的 Lebesgue 分解（见定理 A.4.7），且 $X_\infty = \mathrm{d}\mu_r/\mathrm{d}\nu$，$\nu$-几乎必然. 此处及证明中需注意关于哪个测度几乎必然.

证明. 先证如下引理：

引理 4.3.6. X_n（定义于 $(\Omega, \mathcal{F}, \nu)$ 上）关于 $\mathcal{F}_n$ 是鞅.

证明. 由定义，$X_n \in \mathcal{F}_n$. 设 $A \in \mathcal{F}_n$. 由于 $X_n \in \mathcal{F}_n$ 且 ν_n 是 ν 在 $\mathcal{F}_n$ 上的限制，

$$\int_A X_n \,\mathrm{d}\nu = \int_A X_n \,\mathrm{d}\nu_n.$$

由 X_n 的定义和习题 A.4.7 可知，

$$\int_A X_n \,\mathrm{d}\nu_n = \mu_n(A) = \mu(A),$$

最后一个等式成立是因为 $A \in \mathcal{F}_n$ 且 μ_n 是 μ 在 $\mathcal{F}_n$ 上的限制. 若 $A \in \mathcal{F}_{m-1} \subset \mathcal{F}_m$，利用上述结果对 $n = m$ 和 $n = m-1$ 得

$$\int_A X_m \,\mathrm{d}\nu = \mu(A) = \int_A X_{m-1} \,\mathrm{d}\nu,$$

故 $E(X_m|\mathcal{F}_{m-1}) = X_{m-1}$. □

由于 X_n 是非负鞅，定理 4.2.12 蕴含 $X_n \to X$ ν-几乎必然. 我们希望验证定理中的等式成立. 以 $\mu(\Omega)$ 除 $\mu(A)$，不失一般性可设 μ 为概率测度. 令 $\rho = (\mu+\nu)/2$，$\rho_n = (\mu_n+\nu_n)/2 = \rho$ 在 $\mathcal{F}_n$ 上的限制. 令 $Y_n = \mathrm{d}\mu_n/\mathrm{d}\rho_n$，$Z_n = \mathrm{d}\nu_n/\mathrm{d}\rho_n$. $Y_n, Z_n \geqslant 0$ 且 $Y_n + Z_n = 2$（由习题 A.4.6），故 Y_n 和 Z_n 是有界鞅，具有极限 Y 和 Z. 由此可得

$$Y = \mathrm{d}\mu/\mathrm{d}\rho, \quad Z = \mathrm{d}\nu/\mathrm{d}\rho. \tag{*}$$

只需证明第一个等式. 由引理 4.3.6 的证明，若 $A \in \mathcal{F}_m \subset \mathcal{F}_n$，

$$\mu(A) = \int_A Y_n \,\mathrm{d}\rho \to \int_A Y \,\mathrm{d}\rho$$

由控制收敛定理. 最后这一计算表明

$$\int_A Y \,\mathrm{d}\rho = \mu(A) \quad \text{对所有} A \in \mathcal{G} = \cup_m \mathcal{F}_m.$$

$\mathcal{G}$ 是 π-系，故 π-λ 定理蕴含等式对所有 $A \in \mathcal{F} = \sigma(\mathcal{G})$ 成立，(*) 得证.

由习题 A.4.8 和 A.4.9，$X_n = Y_n/Z_n$. 此时易见 $X = Y/Z$，但为严谨起见，注意 $Y+Z = 2$（ρ-几乎必然），故 $\rho(Y = 0, Z = 0) = 0$. 排除 $0/0$ 情形后有 $X = Y/Z$ ρ-几乎必然（回忆 $X \equiv \limsup X_n$）.

令 $W = (1/Z) \cdot 1_{(Z>0)}$. 利用 (*)，然后 $1 = ZW + 1_{(Z=0)}$，有

$$\mu(A) = \int_A Y \,\mathrm{d}\rho = \int_A YWZ \,\mathrm{d}\rho + \int_A 1_{(Z=0)} Y \,\mathrm{d}\rho. \tag{a}$$

(*) 蕴含 $\mathrm{d}\nu = Z\,\mathrm{d}\rho$，由定义可知

$$YW = X1_{(Z>0)} = X \quad \nu\text{-几乎必然},$$

第二个等式因 $\nu(\{Z=0\})=0$. 综合得

$$\int_A YWZ\,\mathrm{d}\rho = \int_A X\,\mathrm{d}\nu. \tag{b}$$

为处理另一项，注意 (*) 蕴含 $\mathrm{d}\mu = Y\,\mathrm{d}\rho$，且由定义可知 $\{X=\infty\}=\{Z=0\}$（μ-几乎必然），故

$$\int_A 1_{(Z=0)}Y\,\mathrm{d}\rho = \int_A 1_{(X=\infty)}\,\mathrm{d}\mu. \tag{c}$$

综合 (a)、(b) 和 (c) 即得所需结论. □

例题 4.3.7. 设 $\mathcal{F}_n=\sigma(I_{k,n}:0\leqslant k<K_n)$，其中对每个 n，$I_{k,n}$ 是 Ω 的划分，且第 $(n+1)$ 个划分是第 n 个划分的加细. 此时条件 $\mu_n\ll\nu_n$ 即 $\nu(I_{k,n})=0$ 蕴含 $\mu(I_{k,n})=0$，且鞅 $X_n=\mu(I_{k,n})/\nu(I_{k,n})$ 在 $I_{k,n}$ 上是 Radon-Nikodym 导数的近似. 具体例子：$\Omega=[0,1)$，$I_{k,n}=[k2^{-n},(k+1)2^{-n})$ $(0\leqslant k<2^n)$，ν 为 Lebesgue 测度.

无穷乘积测度的 Kakutani 二分法. 设 μ 和 ν 为序列空间 $(\mathbb{R}^{\mathbb{N}},\mathcal{R}^{\mathbb{N}})$ 上的测度，使得坐标 $\xi_n(\omega)=\omega_n$ 独立. 令 $F_n(x)=\mu(\xi_n\leqslant x)$，$G_n(x)=\nu(\xi_n\leqslant x)$. 设 $F_n\ll G_n$ 并令 $q_n=\mathrm{d}F_n/\mathrm{d}G_n$. 为避免问题，设 $q_n>0$，G_n-几乎必然.

令 $\mathcal{F}_n=\sigma(\xi_m:m\leqslant n)$，$\mu_n$ 和 ν_n 分别为 μ 和 ν 在 $\mathcal{F}_n$ 上的限制，且

$$X_n=\frac{\mathrm{d}\mu_n}{\mathrm{d}\nu_n}=\prod_{m=1}^{n}q_m.$$

定理 4.3.5 蕴含 $X_n\to X$ ν-几乎必然. 由假设 $q_n>0$（G_n-几乎必然）可知 $\sum_{m=1}^{\infty}\log(q_m)>-\infty$ 是尾事件，故由 Kolmogorov 0-1 律可得

$$\nu(X=0)\in\{0,1\}. \tag{4.3.3}$$

由定理 4.3.5 可知要么 $\mu\ll\nu$，要么 $\mu\perp\nu$. 下述结果给出了区分两种情形的具体准则.

定理 4.3.8. $\mu\ll\nu$ 或 $\mu\perp\nu$，依 $\prod_{m=1}^{\infty}\int\sqrt{q_m}\,\mathrm{d}G_m>0$ 或 $=0$ 而定.

证明. Jensen 不等式和习题 A.4.7 蕴含

$$\left(\int\sqrt{q_m}\,\mathrm{d}G_m\right)^2\leqslant\int q_m\,\mathrm{d}G_m=\int\mathrm{d}F_m=1,$$

故积分的无穷积良定义且 $\leqslant 1$. 令

$$X_n = \prod_{m\leqslant n} q_m(\omega_m)$$

如上，回忆 $X_n \to X$ ν-几乎必然. 若无穷积为 0，则

$$\int X_n^{1/2}\,\mathrm{d}\nu = \prod_{m=1}^{n}\int \sqrt{q_m}\,\mathrm{d}G_m \to 0.$$

Fatou 引理蕴含

$$\int X^{1/2}\,\mathrm{d}\nu \leqslant \liminf_{n\to\infty}\int X_n^{1/2}\,\mathrm{d}\nu = 0,$$

故 $X = 0$ ν-几乎必然，定理 4.3.5 蕴含 $\mu \perp \nu$. 为证另一方向，令 $Y_n = X_n^{1/2}$. 由 $\int \sqrt{q_m}\,\mathrm{d}G_m \leqslant 1$，若记 E 为关于 ν 的期望，则 $EY_m^2 = EX_m = 1$，故

$$E(Y_{n+k} - Y_n)^2 = E(X_{n+k} + X_n - 2X_n^{1/2}X_{n+k}^{1/2}) = 2\left(1 - \prod_{m=n+1}^{n+k}\int \sqrt{q_m}\,\mathrm{d}G_m\right).$$

由 $|a-b| = |a^{1/2} - b^{1/2}|\cdot(a^{1/2}+b^{1/2})$，利用 Cauchy-Schwarz 不等式和 $(a+b)^2 \leqslant 2a^2 + 2b^2$ 得

$$\begin{aligned} E|X_{n+k} - X_n| = E(|Y_{n+k} - Y_n|(Y_{n+k} + Y_n)) &\leqslant \left(E(Y_{n+k} - Y_n)^2 E(Y_{n+k} + Y_n)^2\right)^{1/2} \\ &\leqslant \left(4E(Y_{n+k} - Y_n)^2\right)^{1/2}. \end{aligned}$$

由最后两式，若无穷积 > 0，则 X_n 在 $L^1(\nu)$ 中收敛于 X，故 $\nu(X=0) < 1$，(4.3.3) 蕴含概率为 0，所需结果由定理 4.3.5 得出. □

4.3.4 分支过程

设 ξ_i^n（$i, n \geqslant 1$）为独立同分布的非负整数值随机变量. 定义序列 Z_n（$n \geqslant 0$）如下：$Z_0 = 1$ 且

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & \text{若} Z_n > 0, \\ 0 & \text{若} Z_n = 0. \end{cases} \tag{4.3.4}$$

Z_n 称为 **Galton-Watson 过程**. 定义背后的想法是：Z_n 表示第 n 代（或时刻 n）的个体数，该代每个成员独立地产出同分布数量的后代. $p_k = P(\xi_i^n = k)$ 称为**后代分布**.

引理 4.3.9. 设 $\mathcal{F}_n = \sigma(\xi_i^m : i \geqslant 1, 1 \leqslant m \leqslant n)$ 且 $\mu = E\xi_i^m \in (0,\infty)$. 则 Z_n/μ^n 关于 $\mathcal{F}_n$ 是鞅.

证明. 显然 $Z_n \in \mathcal{F}_n$. 由定理 4.1.2，在 $\{Z_n = k\}$ 上有

$$E(Z_{n+1}|\mathcal{F}_n) = E(\xi_1^{n+1} + \cdots + \xi_k^{n+1}|\mathcal{F}_n) = k\mu = \mu Z_n,$$

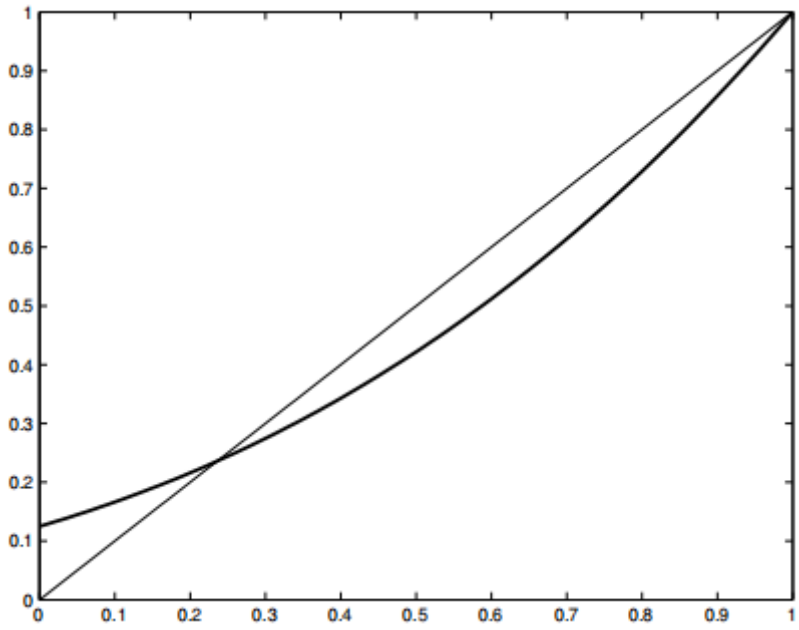

图 4.3: Binomial$(3, 1/2)$ 的生成函数

其中第二个等式利用了 ξ_k^{n+1} 独立于 $\mathcal{F}_n$ 这一事实. □

Z_n/μ^n 是非负鞅，由定理 4.2.12 可知其几乎必然收敛于某极限. 下面先确定该极限为平凡的情形.

定理 4.3.10. 若 $\mu < 1$，则对所有充分大的 n 有 $Z_n = 0$，故 $Z_n/\mu^n \to 0$.

证明. $E(Z_n/\mu^n) = E(Z_0) = 1$，故 $E(Z_n) = \mu^n$. 在 $\{Z_n > 0\}$ 上 $Z_n \geqslant 1$，故

$$P(Z_n > 0) \leqslant E(Z_n; Z_n > 0) = E(Z_n) = \mu^n \to 0$$

当 $\mu < 1$ 时指数衰减. □

最后一个结果是显然的：若每个个体平均产出不到一个孩子，物种将灭绝. 下述结果表明，在排除每个个体恰有一个孩子的平凡情形后，$\mu = 1$ 时同样成立.

定理 4.3.11. 若 $\mu = 1$ 且 $P(\xi_i^m = 1) < 1$，则对所有充分大的 n 有 $Z_n = 0$.

证明. 当 $\mu = 1$ 时，Z_n 本身是非负鞅. 由于 Z_n 取整数值且由定理 4.2.12 收敛于几乎必然有限的极限 Z_∞，对大的 n 必有 $Z_n = Z_\infty$. 若 $P(\xi_i^m = 1) < 1$ 且 $k > 0$，则对任意 N 有 $P(Z_n = k \text{ 对所有} n \geqslant N) = 0$，故必有 $Z_\infty \equiv 0$. □

当 $\mu \leqslant 1$ 时，分支过程灭绝，故 Z_n/μ^n 的极限为 0. 下一步证明若 $\mu > 1$ 则 $P(Z_n > 0 \text{ 对所有} n) > 0$. 对 $s \in [0, 1]$，令 $\varphi(s) = \sum_{k \geqslant 0} p_k s^k$，其中 $p_k = P(\xi_i = k)$. φ 是后代分布 p_k 的**生成函数**.

定理 4.3.12. 设 $\mu > 1$. 若 $Z_0 = 1$，则 $P(\text{对某 } n \text{ 有} Z_n = 0) = \rho$ 为 $\varphi(\rho) = \rho$ 在 $[0,1)$ 中的唯一解.

证明. $\varphi(1) = 1$. 求导并引用定理 A.5.3 得

$$\varphi'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \geqslant 0 \quad (s < 1),$$

故 φ 递增. 当 $s > 1$ 时 $\varphi(s)$ 可能为 ∞，故需谨慎处理.

$$\lim_{s\uparrow 1} \varphi'(s) = \sum_{k=1}^{\infty} k p_k = \mu.$$

积分可知

$$\varphi(1) - \varphi(1-h) = \int_{1-h}^{1} \varphi'(s)\,\mathrm{d}s \sim \mu h$$

当 $h \to 0$ 时成立，故对充分小的 h 有 $\varphi(1-h) < 1-h$. 由 $\varphi(0) \geqslant 0$ 可知 $\varphi(x) = x$ 在 $[0,1)$ 中必有解.

为证唯一性，注意对 $s < 1$

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} > 0,$$

因 $\mu > 1$ 蕴含存在某 $k \geqslant 2$ 使 $p_k > 0$. 令 ρ 为 $[0,1)$ 中 $\varphi(\rho) = \rho$ 的最小解. 由于 $\varphi(1) = 1$ 且 φ 严格凸，对 $x \in (\rho, 1)$ 有 $\varphi(x) < x$，故 $\varphi(\rho) = \rho$ 在 $[0,1)$ 中仅有一解.

综合下述两个结果即完成证明.

(a) 若 $\theta_m = P(Z_m = 0)$，则 $\theta_m = \sum_{k=0}^{\infty} p_k (\theta_{m-1})^k = \varphi(\theta_{m-1})$.

(a) 的证明. 若 $Z_1 = k$（概率为 p_k），则 $Z_m = 0$ 当且仅当所有 k 个家族在剩余 $m-1$ 单位时间内灭绝，这一独立事件的概率为 θ_{m-1}^k. 对各 k 的不交可能性求和即得结果.

(b) 当 $m \uparrow \infty$ 时，$\theta_m \uparrow \rho$.

(b) 的证明. 显然 $\theta_m = P(Z_m = 0)$ 递增. 用归纳法证 $\theta_m \leqslant \rho$：$\theta_0 = 0 \leqslant \rho$，若对 $m-1$ 成立，则

$$\theta_m = \varphi(\theta_{m-1}) \leqslant \varphi(\rho) = \rho.$$

在 $\theta_m = \varphi(\theta_{m-1})$ 中取极限得 $\theta_\infty = \varphi(\theta_\infty)$. 由 $\theta_\infty \leqslant \rho$ 得 $\theta_\infty = \rho$. □

上述结果表明，当 $\mu > 1$ 时，Z_n/μ^n 的极限有非零的机会. 关于此问题的最佳结果属于 Kesten 和 Stigum：

定理 4.3.13. $W = \lim Z_n/\mu^n$ 不 $\equiv 0$ 当且仅当 $\sum_k p_k k \log k < \infty$.

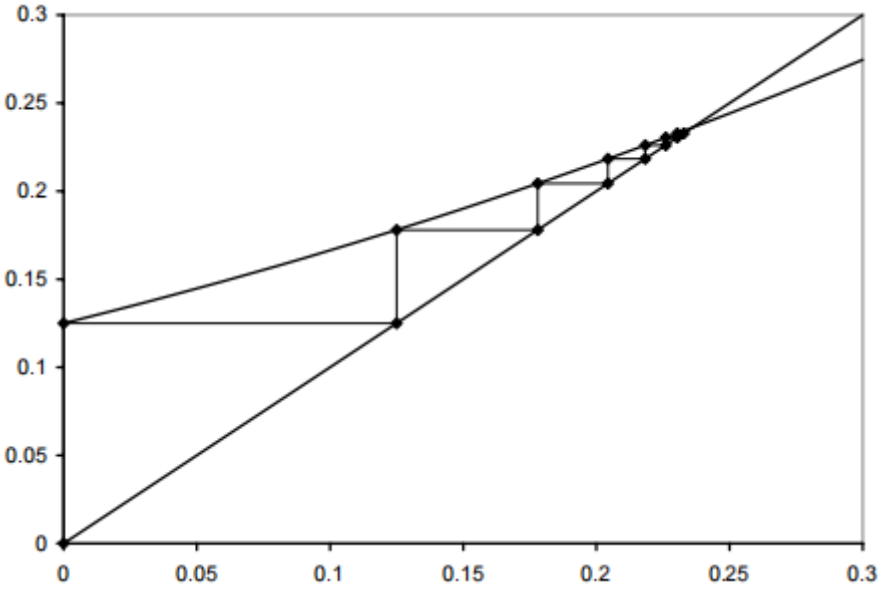

图 4.4: Binomial$(3,1/2)$ 的生成函数对应的 (c) 部分迭代

证明见 Athreya 和 Ney (1972)，第 24–29 页. 下一节将证明 $\sum_k k^2 p_k < \infty$ 足以保证非平凡极限存在.

习题

习题 4.3.1. 给出一个鞅 X_n 的例子，满足 $\sup_n |X_n| < \infty$ 且对 $a = -1, 0, 1$ 都有 $P(X_n = a \text{ i.o.}) = 1$. 此例表明定理 4.3.1 中仅有 $\sup |X_{n+1} - X_n| < \infty$ 是不够的.

习题 4.3.2. （假设熟悉有限状态 Markov 链.）微调上例使得 $P(X_n = 0) \to 1 - 2p$ 且 $P(X_n = -1), P(X_n = 1) \to p$，其中 p 为 $(0, 1/2)$ 中你最喜欢的数，即要求对一个你可选的 p 值完成此构造. 此例表明鞅可以依分布收敛，但未必几乎必然（或依概率）收敛.

习题 4.3.3. 设 X_n 和 Y_n 为正可积且适应于 $\mathcal{F}_n$. 设 $E(X_{n+1}|\mathcal{F}_n) \leqslant X_n + Y_n$，且 $\sum Y_n < \infty$ 几乎必然. 证明 X_n 几乎必然收敛于有限极限. 提示：令 $N = \inf\{k : \sum_{m=1}^{k} Y_m > M\}$，并在时刻 N 停止你的上鞅.

习题 4.3.4. 设 $p_m \in [0, 1)$. 利用 Borel-Cantelli 引理证明

$$\prod_{m=1}^{\infty}(1 - p_m) = 0 \text{ 当且仅当} \sum_{m=1}^{\infty} p_m = \infty.$$

习题 4.3.5. 证明 $\sum_{n=2}^{\infty} P(A_n | \cap_{m=1}^{n-1} A_m^c) = \infty$ 蕴含 $P(\cap_{m=1}^{\infty} A_m^c) = 0$.

习题 4.3.6. 通过直接计算验证例 4.3.7 中的 X_n 是鞅. 证明若去掉条件 $\mu_n \ll \nu_n$ 并当 $\nu(I_{k,n}) = 0$ 时令 $X_n = 0$，则 $E(X_{n+1}|\mathcal{F}_n) \leqslant X_n$.

习题 4.3.7. 对例 4.3.7 应用定理 4.3.5 得到 Radon-Nikodym 定理的 "概率" 证明. 准确地说，设 $\mathcal{F}$ 可数生成（即存在集合列 A_n 使 $\mathcal{F}=\sigma(A_n:n\geqslant 1)$），证明若 μ 和 ν 为 σ-有限测度且 $\mu\ll\nu$，则存在函数 g 使 $\mu(A)=\int_A g\,\mathrm{d}\nu$. 在提出这是循环论证之前（Radon-Nikodym 定理曾被用于定义条件期望!），注意例 4.3.7 所需的条件期望具有初等定义.

Bernoulli 乘积测度. 对下三题，设 F_n,G_n 集中于 $\{0,1\}$，且 $F_n(0)=1-\alpha_n$，$G_n(0)=1-\beta_n$.

习题 4.3.8. (i) 利用定理 4.3.8 找出 $\mu\ll\nu$ 的充要条件. (ii) 设 $0<\varepsilon\leqslant\alpha_n,\beta_n\leqslant 1-\varepsilon<1$. 证明此时条件即为 $\sum(\alpha_n-\beta_n)^2<\infty$.

习题 4.3.9. 证明若 $\sum\alpha_n<\infty$ 且 $\sum\beta_n=\infty$，则 $\mu\perp\nu$. 这表明条件 $\sum(\alpha_n-\beta_n)^2<\infty$ 一般不是 $\mu\ll\nu$ 的充分条件.

习题 4.3.10. 设 $0<\alpha_n,\beta_n<1$. 证明 $\sum|\alpha_n-\beta_n|<\infty$ 是 $\mu\ll\nu$ 的充分条件.

习题 4.3.11. 证明若 $P(\lim Z_n/\mu^n=0)<1$，则 $=\rho$，故 $\{\lim Z_n/\mu^n>0\}=\{Z_n>0\text{ 对所有}n\}$ 几乎必然.

习题 4.3.12. 设 Z_n 为后代分布 p_k 的分支过程（定义见 4.3 节 d 部分），$\varphi(\theta)=\sum p_k\theta^k$. 设 $\rho<1$ 满足 $\varphi(\rho)=\rho$. 证明 ρ^{Z_n} 是鞅，并用此结论证明

$$P(Z_n=0\text{ 对某}n\geqslant 1|Z_0=x)=\rho^x.$$

习题 4.3.13. Galton 和 Watson 提出以他们名字命名的过程时，关心的是家族姓氏的延续. 设每个家庭恰有 3 个孩子，由抛硬币决定性别. 在 19 世纪，只有男性后代继承家族姓氏，故追踪男性后代得到一个分支过程，其中 $p_0=1/8,p_1=3/8,p_2=3/8,p_3=1/8$. 当 $Z_0=1$ 时计算家族姓氏灭绝的概率 ρ.

第 4 节 Doob 不等式和 $p>1$ 时的 L^p 收敛

首先证明定理 4.2.9 的一个推论.

定理 4.4.1. 若 X_n 为下鞅，N 为停时且 $P(N \leqslant k)=1$，则

$$EX_0 \leqslant EX_N \leqslant EX_k.$$

注释. 设 S_n 为简单随机游走，$S_0=1$，$N=\inf\{n: S_n=0\}$（详见例 4.2.13）. $ES_0=1>0=ES_N$，故第一个不等式对无界停时未必成立. 5.7 节将给出保证 $EX_0 \leqslant EX_N$ 对无界 N 成立的条件.

证明. 定理 4.2.9 蕴含 $X_{N\wedge n}$ 是下鞅，因此

$$EX_0 = EX_{N\wedge 0} \leqslant EX_{N\wedge k} = EX_N.$$

为证另一不等式，记 $K_n = 1_{\{N<n\}} = 1_{\{N\leqslant n-1\}}$. K_n 是可料的，故定理 4.2.8 蕴含 $(K\cdot X)_n = X_n - X_{N\wedge n}$ 是下鞅，从而

$$EX_k - EX_N = E(K\cdot X)_k \geqslant E(K\cdot X)_0 = 0. \qquad \square$$

下面将看到定理 4.4.1 非常有用. 首个例证如下：

定理 4.4.2. Doob 不等式. 设 X_m 为下鞅，$\bar{X}_n = \max_{0\leqslant m\leqslant n} X_m^+$，$\lambda>0$，$A=\{\bar{X}_n \geqslant \lambda\}$. 则

$$\lambda P(A) \leqslant EX_n 1_A \leqslant EX_n^+.$$

证明. 令 $N=\inf\{m: X_m \geqslant \lambda \text{ 或} m=n\}$. 由于在 A 上 $X_N \geqslant \lambda$，有

$$\lambda P(A) \leqslant EX_N 1_A \leqslant EX_n 1_A.$$

第二个不等式由定理 4.4.1 及在 A^c 上 $X_N = X_n$ 这一事实推出. 最后一个不等式显然成立，证毕. $\square$

例题 4.4.3. 随机游走. 设 $S_n = \xi_1 + \cdots + \xi_n$，其中 ξ_m 独立且 $E\xi_m=0$，$\sigma_m^2 = E\xi_m^2 < \infty$. S_n 是鞅，故定理 4.2.6 蕴含 $X_n = S_n^2$ 是下鞅. 令 $\lambda = x^2$ 并对 X_n 应用定理 4.4.2，得到 Kolmogorov 极大不等式（定理 2.5.5）：

$$P\left(\max_{1\leqslant m\leqslant n} |S_m| \geqslant x\right) \leqslant x^{-2}\,\mathrm{var}(S_n).$$

对定理 4.4.2 中的不等式积分可得：

定理 4.4.4. L^p **极大不等式**. 若 X_n 为下鞅，则对 $1<p<\infty$，

$$E(\bar{X}_n^p) \leqslant \left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

因此，若 Y_n 为鞅且 $Y_n^* = \max_{0\leqslant m\leqslant n}|Y_m|$，

$$E|Y_n^*|^p \leqslant \left(\frac{p}{p-1}\right)^p E(|Y_n|^p).$$

证明. 第二个不等式通过对 $X_n = |Y_n|$ 应用第一个不等式得到. 为证第一个不等式，下面将说明为何考虑 $\bar{X}_n \wedge M$ 而非 $\bar{X}_n$. 由于 $\{\bar{X}_n \wedge M \geqslant \lambda\}$ 恒为 $\{\bar{X}_n \geqslant \lambda\}$ 或 $\emptyset$，这不改变定理 4.4.2 的应用. 利用引理 2.2.13、定理 4.4.2、Fubini 定理和少许计算，得

$$\begin{aligned} E((\bar{X}_n \wedge M)^p) &= \int_0^\infty p\lambda^{p-1} P(\bar{X}_n \wedge M \geqslant \lambda)\,\mathrm{d}\lambda \\ &\leqslant \int_0^\infty p\lambda^{p-1}\lambda^{-1}\int X_n^+ 1_{(\bar{X}_n\wedge M\geqslant\lambda)}\,\mathrm{d}P\,\mathrm{d}\lambda \\ &= \int X_n^+ \int_0^{\bar{X}_n\wedge M} p\lambda^{p-2}\,\mathrm{d}\lambda\,\mathrm{d}P \\ &= \frac{p}{p-1}\int X_n^+(\bar{X}_n\wedge M)^{p-1}\,\mathrm{d}P. \end{aligned}$$

令 $q = p/(p-1)$ 为 p 的共轭指数，由 Hölder 不等式（定理 1.6.3），上式

$$\leqslant \frac{p}{1-p}(E|X_n^+|^p)^{1/p}(E|\bar{X}_n\wedge M|^p)^{1/q}.$$

将最后不等式两边除以 $(E|\bar{X}_n\wedge M|^p)^{1/q}$（由 $\wedge M$ 知其有限），再两边取 p 次幂，得

$$E(|\bar{X}_n\wedge M|^p) \leqslant \left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

令 $M\to\infty$，由单调收敛定理即得所求结果. □

例题 4.4.5. 不存在 L^1 极大不等式. 反例仍由例 4.2.13 提供. 设 S_n 为从 $S_0 = 1$ 出发的简单随机游走，$N = \inf\{n : S_n = 0\}$，$X_n = S_{N\wedge n}$. 定理 4.4.1 蕴含 $EX_n = ES_{N\wedge n} = ES_0 = 1$ 对所有 n 成立. 利用定理 4.8.7 中简单随机游走的击中概率，有

$$P\left(\max_m X_m \geqslant M\right) = \frac{1}{M}. \tag{4.4.1}$$

故

$$E(\max_m X_m) = \sum_{M=1}^\infty P(\max_m X_m \geqslant M) = \sum_{M=1}^\infty 1/M = \infty.$$

单调收敛定理蕴含 $E\max_{m\leqslant n} X_m \uparrow \infty$（当 $n\uparrow\infty$ 时）.

由定理 4.4.4 可得下述结果：

定理 4.4.6. L^p **收敛定理**. 若 X_n 为鞅且 $\sup E|X_n|^p < \infty$（其中 $p > 1$），则 $X_n \to X$ a.s. 且依 L^p 收敛.

证明. $(EX_n^+)^p \leqslant (E|X_n|)^p \leqslant E|X_n|^p$，故由鞅收敛定理（定理 4.2.11）知 $X_n \to X$ a.s. 由定理 4.4.4 的第二个结论，

$$E\left(\sup_{0\leqslant m\leqslant n}|X_m|\right)^p \leqslant \left(\frac{p}{p-1}\right)^p E|X_n|^p.$$

令 $n \to \infty$ 并利用单调收敛定理得 $\sup|X_n| \in L^p$. 由于 $|X_n - X|^p \leqslant (2\sup|X_n|)^p$，由控制收敛定理知 $E|X_n - X|^p \to 0$. □

本节结果最重要的特例是 $p = 2$ 的情形. 为处理这一情形，下述两个结果很有用.

定理 4.4.7. 鞅增量的正交性. 设 X_n 为鞅且 $EX_n^2 < \infty$ 对所有 n 成立. 若 $m \leqslant n$ 且 $Y \in \mathcal{F}_m$ 满足 $EY^2 < \infty$，则

$$E((X_n - X_m)Y) = 0$$

从而若 $\ell < m < n$，

$$E((X_n - X_m)(X_m - X_\ell)) = 0.$$

证明. 由 Cauchy-Schwarz 不等式知 $E|(X_n - X_m)Y| < \infty$. 由 (4.1.5)、定理 4.1.14 及鞅的定义，

$$E((X_n - X_m)Y) = E[E((X_n - X_m)Y|\mathcal{F}_m)] = E[YE((X_n - X_m)|\mathcal{F}_m)] = 0.$$ □

定理 4.4.8. 条件方差公式. 若 X_n 为鞅且 $EX_n^2 < \infty$ 对所有 n 成立，

$$E((X_n - X_m)^2|\mathcal{F}_m) = E(X_n^2|\mathcal{F}_m) - X_m^2.$$

注释. 这是 $E(X - EX)^2 = EX^2 - (EX)^2$ 的条件版本，证明方式完全相同.

证明. 利用条件期望的线性性及定理 4.1.14，有

$$\begin{aligned} E(X_n^2 - 2X_nX_m + X_m^2|\mathcal{F}_m) &= E(X_n^2|\mathcal{F}_m) - 2X_mE(X_n|\mathcal{F}_m) + X_m^2 \\ &= E(X_n^2|\mathcal{F}_m) - 2X_m^2 + X_m^2 \end{aligned}$$

即得所求结果. □

例题 4.4.9. 分支过程. 继续上节末开始的研究. 采用那里引入的记号，设 $\mu = E(\xi_i^m) > 1$ 且 $\mathrm{var}(\xi_i^m) = \sigma^2 < \infty$. 令 $X_n = Z_n/\mu^n$. 在定理 4.4.8 中取 $m = n - 1$ 并整理，有

$$E(X_n^2|\mathcal{F}_{n-1}) = X_{n-1}^2 + E((X_n - X_{n-1})^2|\mathcal{F}_{n-1}).$$

为计算第二项，注意

$$E((X_n - X_{n-1})^2|\mathcal{F}_{n-1}) = E((Z_n/\mu^n - Z_{n-1}/\mu^{n-1})^2|\mathcal{F}_{n-1}) = \mu^{-2n}E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}).$$

由习题 4.1.2，在 $\{Z_{n-1} = k\}$ 上，

$$E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}) = E\left(\left(\sum_{i=1}^{k}\xi_i^n - \mu k\right)^2 \Bigg| \mathcal{F}_{n-1}\right) = k\sigma^2 = Z_{n-1}\sigma^2.$$

综合最后三个等式得

$$EX_n^2 = EX_{n-1}^2 + E(Z_{n-1}\sigma^2/\mu^{2n}) = EX_{n-1}^2 + \sigma^2/\mu^{n+1}$$

因 $E(Z_{n-1}/\mu^{n-1}) = EZ_0 = 1$. 现 $EX_0^2 = 1$，故 $EX_1^2 = 1 + \sigma^2/\mu^2$，归纳得

$$EX_n^2 = 1 + \sigma^2\sum_{k=2}^{n+1}\mu^{-k}.$$

这表明 $\sup EX_n^2 < \infty$，故 $X_n \to X$ 依 L^2 收敛，从而 $EX_n \to EX$. $EX_n = 1$ 对所有 n 成立，故 $EX = 1$ 且 X 不 $\equiv 0$. 由习题 4.3.11 知 $\{X > 0\} = \{Z_n > 0$ 对所有$n\}$.

习题

习题 4.4.1. 证明若 $j \leqslant k$ 则 $E(X_j; N = j) \leqslant E(X_k; N = j)$，并对 j 求和以得到 $EX_N \leqslant EX_k$ 的第二个证明.

习题 4.4.2. 推广定理 4.4.1 的证明以说明：若 X_n 为下鞅且 $M \leqslant N$ 为满足 $P(N \leqslant k) = 1$ 的停时，则 $EX_M \leqslant EX_N$.

习题 4.4.3. 设 $M \leqslant N$ 为停时. 若 $A \in \mathcal{F}_M$ 则

$$L = \begin{cases} M & \text{在}A\text{ 上,} \\ N & \text{在}A^c\text{ 上} \end{cases}$$

是停时.

习题 4.4.4. 利用前一题中的停时加强习题 4.4.2 的结论为 $X_M \leqslant E(X_N|\mathcal{F}_M)$.

习题 4.4.5. 证明条件方差公式的下述变体. 若 $\mathcal{F} \subset \mathcal{G}$ 则

$$E\left(E[Y|\mathcal{G}] - E[Y|\mathcal{F}]\right)^2 = E(E[Y|\mathcal{G}])^2 - E(E[Y|\mathcal{F}])^2.$$

习题 4.4.6. 设在上述条件基础上补充 $|\xi_m| \leqslant K$，令 $s_n^2 = \sum_{m\leqslant n} \sigma_m^2$. 习题 4.2.2 蕴含 $S_n^2 - s_n^2$ 是鞅. 利用此结论和定理 4.4.1 推出

$$P\left(\max_{1\leqslant m\leqslant n} |S_m| \leqslant x\right) \leqslant (x+K)^2/\operatorname{var}(S_n).$$

习题 4.4.7. 下述结果将定理 4.4.2 推广到 $p=1$ 的情形. 设 X_n 为满足 $X_0 = 0$ 且 $EX_n^2 < \infty$ 的鞅. 证明

$$P\left(\max_{1\leqslant m\leqslant n} X_m \geqslant \lambda\right) \leqslant EX_n^2/(EX_n^2 + \lambda^2).$$

提示：利用 $(X_n + c)^2$ 是下鞅这一事实并对 c 优化.

习题 4.4.8. 设 X_n 为下鞅且 $\log^+ x = \max(\log x, 0)$. 证明

$$E\bar{X}_n \leqslant (1-e^{-1})^{-1}\{1 + E(X_n^+ \log^+(X_n^+))\}.$$

证明通过以下步骤完成：(i) 模仿 4.4.2 的证明，但对 $\lambda \leqslant 1$ 使用平凡界 $P(A) \leqslant 1$ 以说明

$$E(\bar{X}_n \wedge M) \leqslant 1 + \int X_n^+ \log(\bar{X}_n \wedge M)\,\mathrm{d}P.$$

(ii) 利用微积分说明 $a\log b \leqslant a\log a + b/e \leqslant a\log^+ a + b/e$.

习题 4.4.9. 设 X_n 和 Y_n 为满足 $EX_n^2 < \infty$ 和 $EY_n^2 < \infty$ 的鞅.

$$EX_nY_n - EX_0Y_0 = \sum_{m=1}^{n} E(X_m - X_{m-1})(Y_m - Y_{m-1}).$$

习题 4.4.10. 设 X_n $(n \geqslant 0)$ 为鞅，$\xi_n = X_n - X_{n-1}$ $(n \geqslant 1)$. 若 $EX_0^2, \sum_{m=1}^{\infty} E\xi_m^2 < \infty$ 则 $X_n \to X_\infty$ a.s. 且依 L^2 收敛.

习题 4.4.11. 沿用前一题的记号. 若 $b_m \uparrow \infty$ 且 $\sum_{m=1}^{\infty} E\xi_m^2/b_m^2 < \infty$ 则 $X_n/b_n \to 0$ a.s. 特别地，若 $E\xi_n^2 \leqslant K < \infty$ 且 $\sum_{m=1}^{\infty} b_m^{-2} < \infty$ 则 $X_n/b_n \to 0$ a.s.

第 5 节 平方可积鞅 *

本节假设

$$X_n \text{ 是鞅，满足} X_0 = 0 \text{ 且} EX_n^2 < \infty \text{ 对所有} n \text{ 成立.}$$

定理 4.2.6 蕴含 X_n^2 是下鞅. 由 Doob 分解定理 4.3.2，可将 X_n^2 写成 $X_n^2 = M_n + A_n$，其中 M_n 是鞅，且由定理 4.3.2 和 4.4.8 中的公式知

$$A_n = \sum_{m=1}^{n} E(X_m^2|\mathcal{F}_{m-1}) - X_{m-1}^2 = \sum_{m=1}^{n} E((X_m - X_{m-1})^2|\mathcal{F}_{m-1}).$$

A_n 称为与 X_n 相关联的**增过程**. A_n 可视为时刻 n 处方差的逐路径度量，而 $A_\infty = \lim A_n$ 为路径上的总方差. 定理 4.5.2 和 4.5.3 分别描述了鞅在 $\{A_\infty < \infty\}$ 和 $\{A_\infty = \infty\}$ 上的行为. 第一个结果证明的关键是:

定理 4.5.1. $E(\sup_m |X_m|^2) \leqslant 4EA_\infty$.

证明. 对 X_n 应用 L^2 极大不等式（定理 4.4.4）得

$$E \sup_{0\leqslant m\leqslant n} |X_m|^2 \leqslant 4EX_n^2 = 4EA_n$$

因 $EX_n^2 = EM_n + EA_n$ 且 $EM_n = EM_0 = EX_0^2 = 0$. 由单调收敛定理即得所求结果. □

定理 4.5.2. $\lim_{n\to\infty} X_n$ 在 $\{A_\infty < \infty\}$ 上几乎必然存在且有限.

证明. 设 $a > 0$. 由于 $A_{n+1} \in \mathcal{F}_n$，$N = \inf\{n : A_{n+1} > a^2\}$ 是停时. 对 $X_{N\wedge n}$ 应用定理 4.5.1 并注意 $A_{N\wedge n} \leqslant a^2$ 得

$$E \sup_n |X_{N\wedge n}|^2 \leqslant 4a^2,$$

故 L^2 收敛定理 4.4.6 蕴含 $\lim X_{N\wedge n}$ 几乎必然存在且有限. 因 a 任意，所求结果得证. □

下述结果是习题 4.4.11 的一个变体.

定理 4.5.3. 设 $f \geqslant 1$ 为递增函数且 $\int_0^\infty f(t)^{-2}\,\mathrm{d}t < \infty$. 则在 $\{A_\infty = \infty\}$ 上 $X_n/f(A_n) \to 0$ a.s.

证明. $H_m = f(A_m)^{-1}$ 有界且可料，故定理 4.2.8 蕴含

$$Y_n \equiv (H \cdot X)_n = \sum_{m=1}^{n} \frac{X_m - X_{m-1}}{f(A_m)}$$

是鞅. 若 B_n 为与 Y_n 相关联的增过程，则

$$\begin{aligned} B_{n+1}-B_n &= E((Y_{n+1}-Y_n)^2|\mathcal{F}_n) \\ &= E\left(\frac{(X_{n+1}-X_n)^2}{f(A_{n+1})^2}\middle|\mathcal{F}_n\right)=\frac{A_{n+1}-A_n}{f(A_{n+1})^2} \end{aligned}$$

因 $f(A_{n+1})\in\mathcal{F}_n$. 关于 f 的假设蕴含

$$\sum_{n=0}^{\infty}\frac{A_{n+1}-A_n}{f(A_{n+1})^2}\leqslant\sum_{n=0}^{\infty}\int_{[A_n,A_{n+1})}f(t)^{-2}\,\mathrm{d}t<\infty,$$

故由定理 4.5.2 知 $Y_n\to Y_\infty$，而所求结论由 Kronecker 引理（定理 2.5.9）推出. □

例题 4.5.4. 设 $\varepsilon>0$ 且 $f(t)=(t\log^{1+\varepsilon}t)^{1/2}\vee 1$. 则 f 满足定理 4.5.3 的假设. 设 $\xi_1,\xi_2,\ldots$ 独立，$E\xi_m=0$ 且 $E\xi_m^2=\sigma_m^2$. 此时 $X_n=\xi_1+\cdots+\xi_n$ 为平方可积鞅，$A_n=\sigma_1^2+\cdots+\sigma_n^2$，故若 $\sum_{i=1}^{\infty}\sigma_i^2=\infty$，定理 4.5.3 蕴含 $X_n/f(A_n)\to 0$，推广了定理 2.5.11.

由定理 4.5.3 可得 Dubins 和 Freedman (1965) 的一个结果，它推广了前面定理 2.3.7 和 4.3.4 中的两个版本.

定理 4.5.5. 第二 Borel-Cantelli 引理，III. 设 B_n 适应于 $\mathcal{F}_n$ 且令 $p_n=P(B_n|\mathcal{F}_{n-1})$. 则

$$\frac{\sum_{m=1}^n 1_{B(m)}}{\sum_{m=1}^n p_m}\to 1 \text{ a.s. 在}\sum_{m=1}^{\infty}p_m=\infty \text{ 上.}$$

证明. 通过 $X_0=0$ 且对 $n\geqslant 1$ 有 $X_n-X_{n-1}=1_{B_n}-P(B_n|\mathcal{F}_{n-1})$ 定义鞅，于是

$$\frac{\sum_{m=1}^n 1_{B(m)}}{\sum_{m=1}^n p_m}-1=\frac{X_n}{\sum_{m=1}^n p_m}.$$

与 X_n 相关联的增过程满足

$$A_n-A_{n-1}=E((X_n-X_{n-1})^2|\mathcal{F}_{n-1})=E((1_{B_n}-p_n)^2|\mathcal{F}_{n-1})=p_n-p_n^2\leqslant p_n.$$

在 $\{A_\infty<\infty\}$ 上，由定理 4.5.2 知 X_n 趋于有限极限，故在 $\{A_\infty<\infty\}\cap\{\sum_m p_m=\infty\}$ 上

$$\frac{X_n}{\sum_{m=1}^n p_m}\to 0.$$

$\{A_\infty=\infty\}=\{\sum_m p_m(1-p_m)=\infty\}\subset\{\sum_m p_m=\infty\}$，故在 $\{A_\infty=\infty\}$ 上，取 $f(t)=t\vee 1$ 由定理 4.5.3 可得所求结论. □

注释. 平凡例子 $B_n=\Omega$（对所有 n）表明，可能出现 $A_\infty<\infty$ 且 $\sum_m p_m=\infty$ a.s. 的情形。

例题 4.5.6. Bernard Friedman 罐子. 考虑 Polya 罐子（见 5.3 节）的一个变体，其中我们添加 a 个所抽颜色的球和 b 个相反颜色的球，这里 $a\geqslant 0$ 且 $b>0$. 将证明若从 g 个绿球和 r 个红球开始 $(g,r>0)$，则绿球比例 $g_n\to 1/2$. 令 G_n 和 R_n 分别为第 n 次抽取完成后的绿球和红球数. 令 B_n 为第 n 次抽到绿球的事件，D_n 为前 n 次抽取中抽到的绿球数. 由定理 4.5.5，

$$\frac{D_n}{\sum_{m=1}^n g_{m-1}}\to 1 \text{ a.s. 于} \left\{\sum_{m=1}^\infty g_{m-1}=\infty\right\} \text{ 上,} \tag{4.5.1}$$

后者恒成立因 $g_m\geqslant g/(g+r+(a+b)m)$. 至此，论证分为三种情形.

情形 1. $a=b=c$. 此时结果是平凡的，因我们总是添加 c 个每种颜色的球.

情形 2. $a>b$. 首先注意

$$g_{n+1}=\frac{G_{n+1}}{G_{n+1}+R_{n+1}}=\frac{g+aD_n+b(n-D_n)}{g+r+n(a+b)}. \tag{$*$}$$

若 $\limsup_{n\to\infty} g_n\leqslant x$ 则 (4.5.1) 蕴含 $\limsup_{n\to\infty} D_n/n\leqslant x$ 且（因 $a>b$）

$$\limsup_{n\to\infty} g_{n+1}\leqslant \frac{ax+b(1-x)}{a+b}=\frac{b+(a-b)x}{a+b}.$$

右端是斜率 <1 的线性函数且在 $1/2$ 处有不动点，故从平凡上界 $x=1$ 出发迭代得 $\limsup g_n\leqslant 1/2$. 交换红绿角色得 $\liminf_{n\to\infty} g_n\geqslant 1/2$，结果得证.

情形 3. $a<b$. 此情形结果更容易相信，因我们添加更多未抽到颜色的球，但证明略难. 困难在于当 $b>a$ 且 $D_n\leqslant xn$ 时，$(*)$ 右端在 $D_n=0$ 时取最大值，故还需利用红球比例 r_n 满足

$$r_{n+1}=\frac{R_{n+1}}{G_{n+1}+R_{n+1}}=\frac{r+bD_n+a(n-D_n)}{g+r+n(a+b)}$$

这一事实. 将此式与 g_{n+1} 的公式结合，若 $\limsup_{n\to\infty} g_n\leqslant x$ 且 $\limsup_{n\to\infty} r_n\leqslant y$ 则

$$\begin{aligned}\limsup_{n\to\infty} g_n&\leqslant \frac{a(1-y)+by}{a+b}=\frac{a+(b-a)y}{a+b},\\ \limsup_{n\to\infty} r_n&\leqslant \frac{bx+a(1-x)}{a+b}=\frac{a+(b-a)x}{a+b}.\end{aligned}$$

从平凡界 $x=1,y=1$ 出发迭代（注意两个上界恒相同），如情形 2 可得两个 $\limsup$ 均 $\leqslant 1/2$.

注释. B. Friedman (1949) 考虑了若干不同的罐子模型. 上述结果属于 Freedman (1965)，他用不同方法证明了该结果. 上述证明属于 Ornstein，来自 Freedman 论文中的注释.

定理 4.5.1 来自对定理 4.4.4 的应用. 若改用定理 4.4.2，可得稍好的结果.

定理 4.5.7. $E(\sup_n |X_n|) \leqslant 3EA_\infty^{1/2}$.

证明. 如在定理 4.5.2 证明中，设 $a>0$ 且令 $N=\inf\{n : A_{n+1}>a^2\}$. 但这次起点是

$$P\left(\sup_m |X_m|>a\right) \leqslant P(N<\infty)+P\left(\sup_m |X_{N\wedge m}|>a\right).$$

$P(N<\infty)=P(A_\infty>a^2)$. 为界定第二项，对 $X_{N\wedge m}^2$ 应用定理 4.4.2，取 $\lambda=a^2$ 得

$$P\left(\sup_{m\leqslant n} |X_{N\wedge m}|>a\right) \leqslant a^{-2}EX_{N\wedge n}^2 = a^{-2}EA_{N\wedge n} \leqslant a^{-2}E(A_\infty\wedge a^2).$$

在最后一式中令 $n\to\infty$，代入第一式并积分得

$$\int_0^\infty P\left(\sup_m |X_m|>a\right)\,\mathrm{d}a \leqslant \int_0^\infty P(A_\infty>a^2)\,\mathrm{d}a+\int_0^\infty a^{-2}E(A_\infty\wedge a^2)\,\mathrm{d}a.$$

由 $P(A_\infty>a^2)=P(A_\infty^{1/2}>a)$，第一个积分为 $EA_\infty^{1/2}$. 对第二个，利用引理 2.2.13（第一和第四步）、Fubini 定理和计算得

$$\begin{aligned}\int_0^\infty a^{-2}E(A_\infty\wedge a^2)\,\mathrm{d}a &= \int_0^\infty a^{-2}\int_0^{a^2} P(A_\infty>b)\,\mathrm{d}b\,\mathrm{d}a\\ &= \int_0^\infty P(A_\infty>b)\int_{\sqrt{b}}^\infty a^{-2}\,\mathrm{d}a\,\mathrm{d}b = \int_0^\infty b^{-1/2}P(A_\infty>b)\,\mathrm{d}b = 2EA_\infty^{1/2},\end{aligned}$$

证毕. □

例题 4.5.8. 设 $\xi_1,\xi_2,\ldots$ 独立同分布，$P(\xi_i=1)=P(\xi_i=-1)$. 令 $S_n=\xi_1+\cdots+\xi_n$. 定理 4.4.1 蕴含对任意停时 N，$ES_{N\wedge n}=0$. 利用定理 4.5.7 可断定若 $EN^{1/2}<\infty$ 则 $ES_N=0$. 令 $T=\inf\{n : S_n=-1\}$. 因 $S_T=-1$ 均值为 0，故 $ET^{1/2}=\infty$.

第 6 节 一致可积性，L^1 收敛

本节给出鞅在 L^1 中收敛的充分必要条件. 关键在于如下定义.

称随机变量族 X_i（$i \in I$）是**一致可积的**，若

$$\lim_{M\to\infty} \sup_{i\in I} E(|X_i|; |X_i| > M) = 0.$$

若取足够大的 M 使得上确界 < 1，则

$$\sup_{i\in I} E|X_i| \leqslant M + 1 < \infty.$$

这一注记将在下文多次用到.

一致可积族的一个平凡例子是被同一可积随机变量控制的随机变量族，即 $|X_i| \leqslant Y$ 且 $EY < \infty$. 首个结果表明一致可积族可以非常大，下面给出一个有趣的例子.

定理 4.6.1. 给定概率空间 $(\Omega, \mathcal{F}_0, P)$ 及 $X \in L^1$，则 $\{E(X|\mathcal{F}) : \mathcal{F} \text{ 为 } \sigma\text{-域} \subset \mathcal{F}_0\}$ 是一致可积的.

证明. 若 A_n 为一列集合且 $P(A_n) \to 0$，则控制收敛定理蕴含 $E(|X|; A_n) \to 0$. 由这一注记，若 $\varepsilon > 0$，可取 $\delta > 0$ 使得当 $P(A) \leqslant \delta$ 时 $E(|X|; A) \leqslant \varepsilon$.（否则，存在集合 A_n 满足 $P(A_n) \leqslant 1/n$ 且 $E(|X|; A_n) > \varepsilon$，矛盾.）

取足够大的 M 使得 $E|X|/M \leqslant \delta$. Jensen 不等式和条件期望的定义蕴含

$$E(|E(X|\mathcal{F})|; |E(X|\mathcal{F})| > M) \leqslant E(E(|X||\mathcal{F}); E(|X||\mathcal{F}) > M) = E(|X|; E(|X||\mathcal{F}) > M),$$

因为 $\{E(|X||\mathcal{F}) > M\} \in \mathcal{F}$. 利用 Chebyshev 不等式并由 M 的定义，有

$$P\{E(|X||\mathcal{F}) > M\} \leqslant E\{E(|X||\mathcal{F})\}/M = E|X|/M \leqslant \delta.$$

故由 δ 的选取，对所有 $\mathcal{F}$ 有

$$E(|E(X|\mathcal{F})|; |E(X|\mathcal{F})| > M) \leqslant \varepsilon.$$

因 ε 是任意的，该集合族一致可积. □

验证一致可积性的一个常用方法是使用：

定理 4.6.2. 设 $\varphi \geqslant 0$ 为任意满足 $\varphi(x)/x \to \infty$（当 $x \to \infty$）的函数，例如 $\varphi(x) = x^p$（$p > 1$）或 $\varphi(x) = x \log^+ x$. 若对所有 $i \in I$ 有 $E\varphi(|X_i|) \leqslant C$，则 $\{X_i : i \in I\}$ 一致可积.

证明. 令 $\varepsilon_M = \sup\{x/\varphi(x) : x \geqslant M\}$. 对 $i \in I$ 有

$$E(|X_i|; |X_i| > M) \leqslant \varepsilon_M E(\varphi(|X_i|); |X_i| > M) \leqslant C\varepsilon_M,$$

且 $\varepsilon_M \to 0$（当 $M \to \infty$）. □

一致可积性对 L^1 收敛的意义由下述结果阐明:

定理 4.6.3. 设对所有 n 有 $E|X_n| < \infty$. 若 $X_n \to X$ 依概率，则下列命题等价:

(i) $\{X_n : n \geqslant 0\}$ 一致可积.

(ii) $X_n \to X$ 在 L^1 中.

(iii) $E|X_n| \to E|X| < \infty$.

证明.

(i) 蕴含 (ii). 令

$$\varphi_M(x) = \begin{cases} M & \text{若} x \geqslant M, \\ x & \text{若} |x| \leqslant M, \\ -M & \text{若} x \leqslant -M. \end{cases}$$

三角不等式蕴含

$$|X_n - X| \leqslant |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X)| + |\varphi_M(X) - X|.$$

由于 $|\varphi_M(Y) - Y| = (|Y| - M)^+ \leqslant |Y| 1_{(|Y|>M)}$，取期望值得

$$E|X_n - X| \leqslant E|\varphi_M(X_n) - \varphi_M(X)| + E(|X_n|; |X_n| > M) + E(|X|; |X| > M).$$

定理 2.3.4 蕴含 $\varphi_M(X_n) \to \varphi_M(X)$ 依概率, 故第一项 $\to 0$ (由有界收敛定理, 见习题 2.3.5). 若 $\varepsilon > 0$ 且 M 足够大，一致可积性蕴含第二项 $\leqslant \varepsilon$. 为估计第三项，注意一致可积性蕴含 $\sup_n E|X_n| < \infty$, 故 Fatou 引理（习题 2.3.4 中给出的形式）蕴含 $E|X| < \infty$，且取更大的 M 可使第三项 $\leqslant \varepsilon$. 结合上述三个事实可知 $\limsup E|X_n - X| \leqslant 2\varepsilon$. 由于 ε 是任意的，这证明了 (ii).

(ii) 蕴含 (iii). Jensen 不等式蕴含

$$\big|E|X_n| - E|X|\big| \leqslant E\big||X_n| - |X|\big| \leqslant E|X_n - X| \to 0.$$

(iii) 蕴含 (i). 令

$$\psi_M(x)=\begin{cases}x & \text{在}[0,M-1]\text{ 上},\\ 0 & \text{在}[M,\infty)\text{ 上},\\ \text{线性} & \text{在}[M-1,M]\text{ 上}.\end{cases}$$

控制收敛定理蕴含若 M 足够大，$E|X|-E\psi_M(|X|)\leqslant\varepsilon/2$. 如证明第一部分所述，有界收敛定理蕴含 $E\psi_M(|X_n|)\to E\psi_M(|X|)$，故利用 (iii) 可知若 $n\geqslant n_0$ 则

$$E(|X_n|;|X_n|>M)\leqslant E|X_n|-E\psi_M(|X_n|)\leqslant E|X|-E\psi_M(|X|)+\varepsilon/2<\varepsilon.$$

取更大的 M，可使 $E(|X_n|;|X_n|>M)\leqslant\varepsilon$ 对 $0\leqslant n<n_0$ 成立，故 X_n 一致可积.

□

下面给出本节的主要定理. 前述工作已完成，故证明很短.

定理 4.6.4. 对下鞅，下列命题等价：

(i) 它一致可积.

(ii) 它几乎必然收敛且在 L^1 中收敛.

(iii) 它在 L^1 中收敛.

证明.

(i) 蕴含 (ii). 一致可积性蕴含 $\sup_n E|X_n|<\infty$，故鞅收敛定理蕴含 $X_n\to X$ 几乎必然，且定理 4.6.3 蕴含 $X_n\to X$ 在 L^1 中.

(ii) 蕴含 (iii). 显然.

(iii) 蕴含 (i). $X_n\to X$ 在 L^1 中蕴含 $X_n\to X$ 依概率（见引理 2.2.2），故这由定理 4.6.3 推出.

□

在证明鞅的类似结果（即定理 4.6.4）之前，先分离出论证中两个有用的部分.

引理 4.6.5. 若可积随机变量 $X_n\to X$ 在 L^1 中，则 $E(X_n;A)\to E(X;A)$.

证明. $|EX_m1_A-EX1_A|\leqslant E|X_m1_A-X1_A|\leqslant E|X_m-X|\to0$. □

引理 4.6.6. 若鞅 $X_n\to X$ 在 L^1 中，则 $X_n=E(X|\mathcal{F}_n)$.

证明. 鞅性质蕴含若 $m>n$ 则 $E(X_m|\mathcal{F}_n)=X_n$，故若 $A\in\mathcal{F}_n$ 则 $E(X_n;A)=E(X_m;A)$. 引理 4.6.5 蕴含 $E(X_m;A)\to E(X;A)$，故对所有 $A\in\mathcal{F}_n$ 有 $E(X_n;A)=E(X;A)$. 回顾条件期望的定义，由此得 $X_n=E(X|\mathcal{F}_n)$. □

定理 4.6.7. 对鞅，下列命题等价：

(i) 它一致可积.

(ii) 它几乎必然收敛且在 L^1 中收敛.

(iii) 它在 L^1 中收敛.

(iv) 存在可积随机变量 X 使得 $X_n = E(X|\mathcal{F}_n)$.

证明.

(i) 蕴含 (ii). 因鞅也是下鞅，这由定理 4.6.4 推出.

(ii) 蕴含 (iii). 显然.

(iii) 蕴含 (iv). 由引理 4.6.6 推出.

(iv) 蕴含 (i). 这由定理 4.6.1 推出.

□

下述结果与引理 4.6.6 相关，但方向相反.

定理 4.6.8. 设 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$，即 $\mathcal{F}_n$ 是递增的 σ-域序列且 $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$. 当 $n \to \infty$ 时，

$$E(X|\mathcal{F}_n) \to E(X|\mathcal{F}_\infty) \quad \text{几乎必然且在} L^1 \text{中.}$$

证明. 首先注意若 $m > n$ 则定理 4.1.13 蕴含

$$E(E(X|\mathcal{F}_m)|\mathcal{F}_n) = E(X|\mathcal{F}_n),$$

故 $Y_n = E(X|\mathcal{F}_n)$ 是鞅. 定理 4.6.1 蕴含 Y_n 一致可积，故定理 4.6.7 蕴含 Y_n 几乎必然且在 L^1 中收敛到极限 Y_∞. Y_n 的定义和引理 4.6.6 蕴含

$$E(X|\mathcal{F}_n) = Y_n = E(Y_\infty|\mathcal{F}_n),$$

故

$$\int_A X \,\mathrm{d}P = \int_A Y_\infty \,\mathrm{d}P \quad \text{对所有} A \in \mathcal{F}_n.$$

由 X 和 Y_∞ 可积，且 $\bigcup_n \mathcal{F}_n$ 是 π-系，π-λ 定理蕴含最后一个结果对所有 $A \in \mathcal{F}_\infty$ 成立. 又 $Y_\infty \in \mathcal{F}_\infty$，由此得 $Y_\infty = E(X|\mathcal{F}_\infty)$. □

定理 4.6.8 的直接推论如下：

定理 4.6.9. Lévy 0-1 律. 若 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ 且 $A \in \mathcal{F}_\infty$，则

$$E(1_A|\mathcal{F}_n) \to 1_A \quad \text{几乎必然.}$$

借用 Chung 的一句话：“读者宜深思这一结果的意涵，自行判断它是显然的还是不可思议的. ”下面将论证这两种观点.

“它是显然的.” $1_A \in \mathcal{F}_\infty$，且 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$，故给定 $\mathcal{F}_n$ 中信息时对 1_A 的最佳猜测应趋近于给定 $\mathcal{F}_\infty$ 时的最佳猜测 1_A.

“它是不可思议的.” 设 $X_1, X_2, \ldots$ 独立，$A \in \mathcal{T}$，即尾 σ-域. 对每个 n，A 独立于 $\mathcal{F}_n$，故 $E(1_A|\mathcal{F}_n) = P(A)$. 当 $n \to \infty$ 时，左边几乎必然收敛于 1_A，故 $P(A) = 1_A$ 几乎必然，从而 $P(A) \in \{0, 1\}$，即证明了 Kolmogorov 0-1 律.

最后一个论证可能不足以表明定理 4.6.9“过于反常”，但本节下面的应用表明它是一个非常有用的结果.

定理 4.6.8 的另一技术性推论如下：

定理 4.6.10. 条件期望的控制收敛定理. 设 $Y_n \to Y$ 几乎必然且对所有 n 有 $|Y_n| \leqslant Z$, 其中 $EZ < \infty$. 若 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$，则

$$E(Y_n|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty) \quad \text{几乎必然.}$$

证明. 令 $W_N = \sup\{|Y_n - Y_m| : n, m \geqslant N\}$. $W_N \leqslant 2Z$，故 $EW_N < \infty$. 利用单调性 (4.1.2) 并对 W_N 应用定理 4.6.8 得

$$\limsup_{n\to\infty} E(|Y_n - Y||\mathcal{F}_n) \leqslant \lim_{n\to\infty} E(W_N|\mathcal{F}_n) = E(W_N|\mathcal{F}_\infty).$$

最后一个结果对所有 N 成立，且 $N \uparrow \infty$ 时 $W_N \downarrow 0$，故 (4.1.3) 蕴含 $E(W_N|\mathcal{F}_\infty) \downarrow 0$，Jensen 不等式给出

$$|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F}_n)| \leqslant E(|Y_n - Y||\mathcal{F}_n) \to 0 \quad \text{几乎必然（当}n \to \infty\text{）.}$$

定理 4.6.8 蕴含 $E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty)$ 几乎必然. 由最后两个结论和三角不等式即得所需结果. □

例题 4.6.11. 设 $X_1, X_2, \ldots$ 一致可积且几乎必然 $\to X$. 定理 4.6.3 蕴含 $X_n \to X$ 在 L^1 中，结合习题 4.6.7 可知 $E(X_n|\mathcal{F}) \to E(X|\mathcal{F})$ 在 L^1 中. 下证 $E(X_n|\mathcal{F})$ 不必几乎必然收敛. 设 $Y_1, Y_2, \ldots$ 和 $Z_1, Z_2, \ldots$ 为独立随机变量且

$$P(Y_n = 1) = 1/n, \quad P(Y_n = 0) = 1 - 1/n,$$
$$P(Z_n = n) = 1/n, \quad P(Z_n = 0) = 1 - 1/n.$$

令 $X_n = Y_n Z_n$. $P(X_n > 0) = 1/n^2$，故 Borel-Cantelli 引理蕴含 $X_n \to 0$ 几乎必然. $E(X_n; |X_n| \geqslant$

$1) = n/n^2$，故 X_n 一致可积. 令 $\mathcal{F} = \sigma(Y_1, Y_2, \ldots)$.

$$E(X_n|\mathcal{F}) = Y_n E(Z_n|\mathcal{F}) = Y_n EZ_n = Y_n.$$

由 $Y_n \to 0$ 在 L^1 中但不是几乎必然，$E(X_n|\mathcal{F})$ 亦然.

习题

习题 4.6.1. 设 $Z_1, Z_2, \ldots$ 独立同分布且 $E|Z_i| < \infty$，θ 为独立的有限均值随机变量，且 $Y_i = Z_i + \theta$. 若 Z_i 服从 normal$(0,1)$，则在统计术语中，我们有一个方差为 1、均值未知的正态总体样本. θ 的分布称为**先验分布**，$P(\theta \in \cdot|Y_1, \ldots, Y_n)$ 称为 n 次观测后的**后验分布**. 证明 $E(\theta|Y_1, \ldots, Y_n) \to \theta$ 几乎必然.

习题 4.6.2. 在下两个习题中，$\Omega = [0,1)$，$I_{k,n} = [k2^{-n}, (k+1)2^{-n})$，$\mathcal{F}_n = \sigma(I_{k,n} : 0 \leqslant k < 2^n)$.

若 $|f(t) - f(s)| \leqslant K|t-s|$ 对 $0 \leqslant s, t < 1$ 成立，则称 f 是 Lipschitz 连续的. 证明在 $I_{k,n}$ 上由 $X_n = (f((k+1)2^{-n}) - f(k2^{-n}))/2^{-n}$ 定义的 X_n 是鞅，$X_n \to X_\infty$ 几乎必然且在 L^1 中，且

$$f(b) - f(a) = \int_a^b X_\infty(\omega)\, \mathrm{d}\omega.$$

习题 4.6.3. 设 f 在 $[0,1)$ 上可积. $E(f|\mathcal{F}_n)$ 是阶梯函数且在 L^1 中 $\to f$. 由此立即可得，若 $\varepsilon > 0$，存在 $[0,1]$ 上的阶梯函数 g 使得 $\int |f - g|\, \mathrm{d}x < \varepsilon$. 这一逼近比习题 1.4.3 中使用的徒手方法简单得多，但当然使用了很多工具.

习题 4.6.4. 设 X_n 为取值于 $[0, \infty)$ 的随机变量. 令 $D = \{X_n = 0 \text{ 对某个} n \geqslant 1\}$ 且设

$$P(D|X_1, \ldots, X_n) \geqslant \delta(x) > 0 \quad \text{几乎必然在} \{X_n \leqslant x\} \text{ 上.}$$

利用定理 4.6.9 推出 $P(D \cup \{\lim_n X_n = \infty\}) = 1$.

习题 4.6.5. 设 Z_n 为具有后代分布 p_k 的分支过程（见 5.3 节末的定义）. 利用上一个结果证明若 $p_0 > 0$ 则 $P(\lim_n Z_n = 0 \text{ 或} \infty) = 1$.

习题 4.6.6. 设 $X_n \in [0,1]$ 适应于 $\mathcal{F}_n$. 设 $\alpha, \beta > 0$ 且 $\alpha + \beta = 1$，并设

$$P(X_{n+1} = \alpha + \beta X_n|\mathcal{F}_n) = X_n,$$
$$P(X_{n+1} = \beta X_n|\mathcal{F}_n) = 1 - X_n.$$

证明 $P(\lim_n X_n = 0 \text{ 或} 1) = 1$ 且若 $X_0 = \theta$ 则 $P(\lim_n X_n = 1) = \theta$.

习题 4.6.7. 证明若 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ 且 $Y_n \to Y$ 在 L^1 中，则 $E(Y_n|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty)$ 在 L^1 中.

第 7 节 反向鞅

反向鞅（有些作者称之为 reversed）是由负整数索引的鞅，即 X_n（$n \leqslant 0$），适应于递增的 σ-域 $\mathcal{F}_n$ 且满足

$$E(X_{n+1}|\mathcal{F}_n) = X_n \quad 对 n \leqslant -1.$$

因 σ-域随 $n \downarrow -\infty$ 递减，反向鞅的收敛理论特别简单.

定理 4.7.1. $X_{-\infty} = \lim_{n\to-\infty} X_n$ 几乎必然存在且在 L^1 中.

证明. 令 U_n 为 $X_{-n}, \ldots, X_0$ 上穿 $[a,b]$ 的次数. 上穿不等式（定理 4.2.10）蕴含 $(b-a)EU_n \leqslant E(X_0 - a)^+$. 令 $n \to \infty$ 并利用单调收敛定理，有 $EU_\infty < \infty$，故由定理 4.2.11 证明后的注记，极限几乎必然存在. 鞅性质蕴含 $X_n = E(X_0|\mathcal{F}_n)$，故定理 4.6.1 蕴含 X_n 一致可积，且定理 4.6.3 告诉我们收敛在 L^1 中发生. □

下述结果确定了定理 4.7.1 中的极限.

定理 4.7.2. 若 $X_{-\infty} = \lim_{n\to-\infty} X_n$ 且 $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$，则 $X_{-\infty} = E(X_0|\mathcal{F}_{-\infty})$.

证明. 显然 $X_{-\infty} \in \mathcal{F}_{-\infty}$，$X_n = E(X_0|\mathcal{F}_n)$，故若 $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_n$ 则

$$\int_A X_n \,\mathrm{d}P = \int_A X_0 \,\mathrm{d}P.$$

定理 4.7.1 和引理 4.6.5 蕴含 $E(X_n; A) \to E(X_{-\infty}; A)$，故

$$\int_A X_{-\infty} \,\mathrm{d}P = \int_A X_0 \,\mathrm{d}P$$

对所有 $A \in \mathcal{F}_{-\infty}$ 成立，证得所需结论. □

下面是定理 4.6.8 的反向形式.

定理 4.7.3. 若 $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ 当 $n \downarrow -\infty$（即 $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$），则

$$E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_{-\infty}) \quad 几乎必然且在 L^1 中.$$

证明. $X_n = E(Y|\mathcal{F}_n)$ 是反向鞅，故定理 4.7.1 和 4.7.2 蕴含当 $n \downarrow -\infty$ 时，$X_n \to X_{-\infty}$ 几乎必然且在 L^1 中，其中

$$X_{-\infty} = E(X_0|\mathcal{F}_{-\infty}) = E(E(Y|\mathcal{F}_0)|\mathcal{F}_{-\infty}) = E(Y|\mathcal{F}_{-\infty}).$$

□

尽管反向鞅的收敛理论很简单，但有一些很好的应用. 本节余下部分回到 4.1 节使用的特殊空间，以便利用那里给出的定义. 即设

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \ldots) : \omega_i \in S\},\\ \mathcal{F} &= \mathcal{S} \times \mathcal{S} \times \cdots,\\ X_n(\omega) &= \omega_n.\end{aligned}$$

记 $\mathcal{E}_n$ 为由在置换 $n+1, n+2, \ldots$ 下不变的事件生成的 σ-域，令 $\mathcal{E} = \bigcap_n \mathcal{E}_n$ 为**可交换 σ-域**.

例题 4.7.4. 强大数定律. 设 $\xi_1, \xi_2, \ldots$ 独立同分布且 $E|\xi_i| < \infty$. 令 $S_n = \xi_1 + \cdots + \xi_n$，$X_{-n} = S_n/n$，且

$$\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \ldots).$$

为计算 $E(X_{-n}|\mathcal{F}_{-n-1})$，注意若 $j, k \leqslant n+1$，对称性蕴含 $E(\xi_j|\mathcal{F}_{-n-1}) = E(\xi_k|\mathcal{F}_{-n-1})$，故

$$E(\xi_{n+1}|\mathcal{F}_{-n-1}) = \frac{1}{n+1}\sum_{k=1}^{n+1} E(\xi_k|\mathcal{F}_{-n-1}) = \frac{1}{n+1}E(S_{n+1}|\mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n+1}.$$

由 $X_{-n} = (S_{n+1} - \xi_{n+1})/n$，由此得

$$\begin{aligned}E(X_{-n}|\mathcal{F}_{-n-1}) &= E(S_{n+1}/n|\mathcal{F}_{-n-1}) - E(\xi_{n+1}/n|\mathcal{F}_{-n-1})\\ &= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = X_{-n-1}.\end{aligned}$$

最后一个计算表明 X_{-n} 是反向鞅，故由定理 4.7.1 和 4.7.2 得 $\lim_{n\to\infty} S_n/n = E(X_{-1}|\mathcal{F}_{-\infty})$. 由 $\mathcal{F}_{-n} \subset \mathcal{E}_n$，$\mathcal{F}_{-\infty} \subset \mathcal{E}$. Hewitt-Savage 0-1 律（定理 2.5.4）表明 $\mathcal{E}$ 是平凡的，故有

$$\lim_{n\to\infty} S_n/n = E(X_{-1}) \quad \text{几乎必然.}$$

例题 4.7.5. 投票定理. 设 $\{\xi_j, 1 \leqslant j \leqslant n\}$ 为独立同分布的非负整数值随机变量，$S_k = \xi_1 + \cdots + \xi_k$，且 $G = \{S_j < j \text{ 对} 1 \leqslant j \leqslant n\}$. 则

$$P(G|S_n) = (1 - S_n/n)^+. \tag{4.7.1}$$

注释. 为解释名称，考虑一场选举，候选人 B 获得 β 票，A 获得 $\alpha > \beta$ 票. 令 $\xi_1, \xi_2, \ldots, \xi_n$ 独立同分布，各以概率 1/2 取值 0 或 2. 将 0 和 2 解释为对候选人 A 和 B 的投票，可见 $G =$

{计票过程中 A 一直领先 B}，故若 $n=\alpha+\beta$

$$P(G|\text{B 获得}\beta\text{票})=1-\frac{2\beta}{\alpha+\beta}=\frac{\alpha-\beta}{\alpha+\beta},$$

即定理 4.9.2 中的结果.

证明. 当 $S_n\geqslant n$ 时结果是平凡的，故设 $S_n<n$. 例 4.7.4 中的计算表明 $X_{-j}=S_j/j$ 关于 $\mathcal{F}_{-j}=\sigma(S_j,\ldots,S_n)$ 是鞅. 令 $T=\inf\{k\geqslant -n: X_k\geqslant 1\}$ 且若集合为空则令 $T=-1$. 断言 $X_T=1$ 在 G^c 上. 为验证这一点，注意若 $S_{j+1}<j+1$ 则 ξ_i 的非负整数值蕴含 $S_j\leqslant S_{j+1}\leqslant j$. 由 $G\subset\{T=-1\}$ 且 $S_1<1$ 蕴含 $S_1=0$，有 $X_T=0$ 在 G 上. 注意 $\mathcal{F}_{-n}=\sigma(S_n)$ 并利用习题 4.4.4，可见在 $\{S_n<n\}$ 上

$$P(G^c|S_n)=E(X_T|\mathcal{F}_{-n})=X_{-n}=S_n/n.$$

用 1 减去它并回顾这一计算是在假设 $S_n<n$ 下完成的，即得所需结果. □

例题 4.7.6. Hewitt-Savage 0-1 律. 若 $X_1,X_2,\ldots$ 独立同分布且 $A\in\mathcal{E}$，则 $P(A)\in\{0,1\}$.

新证明的关键在于:

引理 4.7.7. 设 $X_1,X_2,\ldots$ 独立同分布且令

$$A_n(\phi)=\frac{1}{(n)_k}\sum_i\phi(X_{i_1},\ldots,X_{i_k}),$$

其中求和取遍所有不同的整数序列 $1\leqslant i_1,\ldots,i_k\leqslant n$，且

$$(n)_k=n(n-1)\cdots(n-k+1)$$

是这样的序列的个数. 若 ϕ 有界，则 $A_n(\phi)\to E\phi(X_1,\ldots,X_k)$ 几乎必然.

证明. $A_n(\phi)\in\mathcal{E}_n$，故

$$A_n(\phi)=E(A_n(\phi)|\mathcal{E}_n)=\frac{1}{(n)_k}\sum_i E(\phi(X_{i_1},\ldots,X_{i_k})|\mathcal{E}_n)=E(\phi(X_1,\ldots,X_k)|\mathcal{E}_n),$$

因为求和中的所有项都相同. 令 $\mathcal{F}_{-m}=\mathcal{E}_m$ $(m\geqslant 1)$，定理 4.7.3 蕴含

$$E(\phi(X_1,\ldots,X_k)|\mathcal{E}_n)\to E(\phi(X_1,\ldots,X_k)|\mathcal{E}).$$

需证极限为 $E(\phi(X_1,\ldots,X_k))$. 首先注意 $A_n(\phi)$ 中有 $k(n-1)_{k-1}$ 项包含 X_1 且 ϕ 有界，故若令 $1\in i$

表示对包含 1 的序列求和，

$$\frac{1}{(n)_k}\sum_{1\in i}\phi(X_{i_1},\ldots,X_{i_k})\leqslant\frac{k(n-1)_{k-1}}{(n)_k}\sup\phi\to 0.$$

这表明

$$E(\phi(X_1,\ldots,X_k)|\mathcal{E})\in\sigma(X_2,X_3,\ldots).$$

对 $2,3,\ldots,k$ 重复上述论证表明

$$E(\phi(X_1,\ldots,X_k)|\mathcal{E})\in\sigma(X_{k+1},X_{k+2},\ldots).$$

直观而言，若随机变量的条件期望独立于该随机变量本身，则

$$\text{(a) } E(\phi(X_1,\ldots,X_k)|\mathcal{E})=E(\phi(X_1,\ldots,X_k)).$$

为证明这一点，先证：

$$\text{(b) 若}EX^2<\infty\text{ 且}E(X|\mathcal{G})\in\mathcal{F}\text{，其中}X\text{ 独立于}\mathcal{F}\text{，则}E(X|\mathcal{G})=EX.$$

证明. 令 $Y=E(X|\mathcal{G})$ 并注意定理 4.1.11 蕴含 $EY^2\leqslant EX^2<\infty$. 由独立性，$EXY=EX\cdot EY=(EY)^2$（因为 $EY=EX$）. 由条件期望的几何解释（定理 4.1.15），$E((X-Y)Y)=0$，故 $EY^2=EXY=(EY)^2$ 且 $\text{var}(Y)=EY^2-(EY)^2=0$. □

(a) 对所有有界 ϕ 成立，故 $\mathcal{E}$ 独立于 $\mathcal{G}_k=\sigma(X_1,\ldots,X_k)$. 因这对所有 k 成立，且 $\bigcup_k\mathcal{G}_k$ 是包含 Ω 的 π-系，定理 2.1.6 蕴含 $\mathcal{E}$ 独立于 $\sigma(\bigcup_k\mathcal{G}_k)\supset\mathcal{E}$，从而得到通常的 0-1 律结论. 若 $A\in\mathcal{E}$，它独立于自身，故 $P(A)=P(A\cap A)=P(A)P(A)$，即 $P(A)\in\{0,1\}$. □

例题 4.7.8. de Finetti 定理. 称序列 $X_1,X_2,\ldots$ 是**可交换的**，若对每个 n 和 $\{1,\ldots,n\}$ 的置换 π，

$$(X_1,\ldots,X_n)\text{ 与}(X_{\pi(1)},\ldots,X_{\pi(n)})\text{ 同分布.}$$

定理 4.7.9. de Finetti 定理. 若 $X_1,X_2,\ldots$ 可交换，则关于 $\mathcal{E}$ 条件化后，$X_1,X_2,\ldots$ 独立同分布.

证明. 重复引理 4.7.7 证明中的第一个计算并利用那里引入的记号，表明对任何可交换序列：

$$A_n(\phi)=E(A_n(\phi)|\mathcal{E}_n)=\frac{1}{(n)_k}\sum_i E(\phi(X_{i_1},\ldots,X_{i_k})|\mathcal{E}_n)=E(\phi(X_1,\ldots,X_k)|\mathcal{E}_n),$$

因为求和中的所有项都相同. 同样，定理 4.7.3 蕴含

$$A_n(\phi) \to E(\phi(X_1, \ldots, X_k)|\mathcal{E}). \tag{4.7.2}$$

然而，这次 $\mathcal{E}$ 可能是非平凡的，故不能希望证明极限是 $E(\phi(X_1, \ldots, X_k))$.

令 f 和 g 分别为 $\mathbb{R}^{k-1}$ 和 $\mathbb{R}$ 上的有界函数. 若令 $I_{n,k}$ 为所有不同整数序列 $1 \leqslant i_1, \ldots, i_k \leqslant n$ 的集合，则

$$\begin{aligned}(n)_k A_{k-1}(f) n A_n(g) &= \sum_{i \in I_{n,k-1}} \sum_m f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_m) \\ &= \sum_{i \in I_{n,k}} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_k}) \\ &\quad + \sum_{i \in I_{n,k-1}} \sum_{j=1}^{k-1} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_j}).\end{aligned}$$

若令 $\phi(x_1, \ldots, x_k) = f(x_1, \ldots, x_{k-1}) g(x_k)$，注意

$$\frac{(n)_k}{(n)_{k-1}} = \frac{n}{n-k+1} \quad \text{且} \quad \frac{(n)_{k-1}}{(n)_k} = \frac{1}{n-k+1},$$

重排后得

$$A_n(\phi) = \frac{n}{n-k+1} A_{k-1}(f) A_n(g) - \frac{1}{n-k+1} \sum_{j=1}^{k-1} A_n(\phi_j),$$

其中 $\phi_j(x_1, \ldots, x_{k-1}) = f(x_1, \ldots, x_{k-1}) g(x_j)$. 对 ϕ, f, g 和所有 ϕ_j 应用 (4.7.2) 得

$$E(f(X_1, \ldots, X_{k-1}) g(X_k)|\mathcal{E}) = E(f(X_1, \ldots, X_{k-1})|\mathcal{E}) E(g(X_k)|\mathcal{E}).$$

由归纳法可知

$$E\left(\prod_{j=1}^k f_j(X_j) \middle| \mathcal{E}\right) = \prod_{j=1}^k E(f_j(X_j)|\mathcal{E}).$$

□

当 X_i 取值于好空间时，给定 $\mathcal{E}$ 存在 $(X_1, X_2, \ldots)$ 的正则条件分布，且序列可表示为独立同分布序列的混合. Hewitt 和 Savage (1956) 称这种情形下的序列是**可表示的**. 对于通常的测度论问题，当 X_i 取值于任意测度空间时，最后一个结果不成立. 反例见 Dubins 和 Freedman (1979) 以及 Freedman (1980).

定理 4.7.9 最简单的特殊情形是 $X_i \in \{0, 1\}$. 在此情形

定理 4.7.10. 若 $X_1, X_2, \ldots$ 可交换且取值于 $\{0,1\}$，则存在 $[0,1]$ 上的概率分布使得

$$P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_0^1 \theta^k (1-\theta)^{n-k} \, \mathrm{d}F(\theta).$$

这一结果对那些关心统计学基础的人很有用（见 Savage (1972) 的 3.7 节），因为从可接受的对称性假设出发，得到了强有力的结论：序列是独立同分布序列的混合. 定理 4.7.10 已被用多种不同方法证明. 见 Feller, Vol. II (1971), p. 228–229，其中给出了与矩问题相关的证明. Diaconis 和 Freedman (1980) 给出了一个很好的证明，从平凡观察出发：有限可交换序列 X_m（$1 \leqslant m \leqslant n$）的分布具有形式 $p_0 H_{0,n} + \cdots + p_n H_{n,n}$，其中 $H_{m,n}$ 是“从有 m 个 1 和 $n-m$ 个 0 的罐子中不放回抽样”. 若 $m \to \infty$ 且 $m/n \to p$，则 $H_{m,n}$ 趋近于密度为 p 的乘积测度. 由此易得定理 4.7.10，且可以得到收敛速度的界.

习题

习题 4.7.1. 证明若反向鞅有 $X_0 \in L^p$，则收敛在 L^p 中发生.

习题 4.7.2. 证明定理 4.6.10 的反向类似结果. 设 $n \to -\infty$ 时 $Y_n \to Y_{-\infty}$ 几乎必然且对所有 n 有 $|Y_n| \leqslant Z$ 几乎必然，其中 $EZ < \infty$. 若 $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$，则 $E(Y_n|\mathcal{F}_n) \to E(Y_{-\infty}|\mathcal{F}_{-\infty})$ 几乎必然.

习题 4.7.3. 直接从定义证明，若 $X_1, X_2, \ldots \in \{0,1\}$ 可交换，则

$$P(X_1 = 1, \ldots, X_k = 1 | S_n = m) = \binom{n-k}{n-m} \Big/ \binom{n}{m}.$$

习题 4.7.4. 若 $X_1, X_2, \ldots \in \mathbb{R}$ 可交换且 $EX_i^2 < \infty$，则 $E(X_1 X_2) \geqslant 0$.

习题 4.7.5. 利用引理 4.7.7 证明的前几行推出，若 $X_1, X_2, \ldots$ 独立同分布，$EX_i = \mu$ 且 $\mathrm{var}(X_i) = \sigma^2 < \infty$，则

$$\binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} (X_i - X_j)^2 \to 2\sigma^2.$$

第 8 节 可选停时定理

本节将证明若干结果，使得当 X_n 为下鞅且 $M \leqslant N$ 为停时，可推出 $EX_M \leqslant EX_N$. 例 4.2.13 表明这一结论并非总是成立，但习题 4.4.2 说明当 N 有界时该结论成立，因此注意力将集中于 N 无界的情形.

定理 4.8.1. 若 X_n 为一致可积下鞅，则对任意停时 N，序列 $X_{N\wedge n}$ 一致可积.

与定理 4.2.5 类似，上述结果蕴含关于上鞅（以 $\geqslant$ 代替 $=$）和鞅（以 $=$ 代替 $\leqslant$）的相应结论. 下述两个定理亦如此.

证明. 由 X_n^+ 是下鞅，定理 4.4.1 蕴含 $EX_{N\wedge n}^+ \leqslant EX_n^+$. 由于 X_n^+ 一致可积，根据定义后的注记知

$$\sup_n EX_{N\wedge n}^+ \leqslant \sup_n EX_n^+ < \infty.$$

鞅收敛定理 (4.2.11) 表明 $X_{N\wedge n} \to X_N$ 几乎必然（此处 $X_\infty = \lim X_n$）且 $E|X_N| < \infty$. 这一结论成立后，其余部分极易验证. 将

$$E(|X_{N\wedge n}|; |X_{N\wedge n}| > K) = E(|X_N|; |X_N| > K, N \leqslant n) + E(|X_n|; |X_n| > K, N > n).$$

由于 $E|X_N| < \infty$ 且 X_n 一致可积，当 K 充分大时每项皆小于 $\varepsilon/2$. □

由定理 4.8.1 证明的最后计算可得：

定理 4.8.2. 若 $E|X_N| < \infty$ 且 $X_n 1_{(N>n)}$ 一致可积，则 $X_{N\wedge n}$ 一致可积，从而 $EX_0 \leqslant EX_N$.

定理 4.8.3. 若 X_n 为一致可积下鞅，则对任意停时 $N \leqslant \infty$，有 $EX_0 \leqslant EX_N \leqslant EX_\infty$，其中 $X_\infty = \lim X_n$.

证明. 定理 4.4.1 蕴含 $EX_0 \leqslant EX_{N\wedge n} \leqslant EX_n$. 令 $n \to \infty$，注意到定理 4.8.1 和 4.6.4 蕴含 $X_{N\wedge n} \to X_N$ 且 $X_n \to X_\infty$ 在 L^1 中成立，即得所求. □

下述结果无需一致可积性假设.

定理 4.8.4. 若 X_n 为非负上鞅且 $N \leqslant \infty$ 为停时，则 $EX_0 \geqslant EX_N$，其中 $X_\infty = \lim X_n$ 存在（由定理 4.2.12 保证）.

证明. 利用定理 4.4.1 及 Fatou 引理，

$$EX_0 \geqslant \liminf_{n\to\infty} EX_{N\wedge n} \geqslant EX_N.$$ □

下述结果在有些情形下颇为有用.

定理 4.8.5. 设 X_n 为下鞅且 $E(|X_{n+1}-X_n|\,|\,\mathcal{F}_n)\leqslant B$ 几乎必然. 若 N 为满足 $EN<\infty$ 的停时，则 $X_{N\wedge n}$ 一致可积，从而 $EX_N\geqslant EX_0$.

证明. 首先注意到

$$|X_{N\wedge n}|\leqslant |X_0|+\sum_{m=0}^{\infty}|X_{m+1}-X_m|1_{(N>m)}.$$

为证一致可积性，只需验证右端具有有限期望，从而 $|X_{N\wedge n}|$ 被一可积随机变量所控制. 由于 $\{N>m\}\in\mathcal{F}_m$，

$$E(|X_{m+1}-X_m|;N>m)=E(E(|X_{m+1}-X_m|\,|\,\mathcal{F}_m);N>m)\leqslant BP(N>m),$$

因此

$$E\sum_{m=0}^{\infty}|X_{m+1}-X_m|1_{(N>m)}\leqslant B\sum_{m=0}^{\infty}P(N>m)=BEN<\infty. \qquad \square$$

4.8.1 在随机游走中的应用

设 $\xi_1,\xi_2,\ldots$ 独立同分布，$S_n=S_0+\xi_1+\cdots+\xi_n$，其中 S_0 为常数，并令 $\mathcal{F}_n=\sigma(\xi_1,\ldots,\xi_n)$. 下面将利用 4.2 节中的三个鞅来推导若干结果.

线性鞅. 若令 $\mu=E\xi_i$，则 $X_n=S_n-n\mu$ 为鞅. （见例 4.2.1.）

将线性鞅与定理 4.8.5 结合，可得

定理 4.8.6. Wald 等式. 若 $\xi_1,\xi_2,\ldots$ 独立同分布，$E\xi_i=\mu$，$S_n=\xi_1+\cdots+\xi_n$，且 N 为满足 $EN<\infty$ 的停时，则 $ES_N=\mu EN$.

证明. 令 $X_n=S_n-n\mu$，注意 $E(|X_{n+1}-X_n|\,|\,\mathcal{F}_n)=E|\xi_i-\mu|$. $\square$

二次鞅. 设 $E\xi_i=0$ 且 $E\xi_i^2=\sigma^2\in(0,\infty)$. 则 $X_n=S_n^2-n\sigma^2$ 为鞅. （见例 4.2.2.）

指数鞅. 设 $\varphi(\theta)=E\exp(\theta\xi_i)<\infty$. 则 $X_n=\exp(\theta S_n)/\varphi(\theta)^n$ 为鞅. （见例 4.2.3.）

定理 4.8.7. 对称简单随机游走指 $P(\xi_i=1)=P(\xi_i=-1)=1/2$ 的特殊情形. 设 $S_0=x$ 并令 $N=\min\{n:S_n\notin(a,b)\}$. 记 P_x 和 E_x 以强调初始位置，

(a) $P_x(S_N=a)=\dfrac{b-x}{b-a}$，$\quad P_x(S_N=b)=\dfrac{x-a}{b-a}$

(b) $E_xN=-ab$，从而 $E_xN=(b-x)(x-a)$.

设 $T_x=\min\{n:S_n=x\}$. 取 $a=0,x=1$ 及 $b=M$，有

$$P_1(T_0<T_M)=\frac{M-1}{M},\quad P_1(T_M<T_0)=\frac{1}{M}.$$

第一个结果证明了 (4.4.1). 在第二个式子中令 $M\to\infty$，得 $P_1(T_0<\infty)=1$.

证明. (a) 为证 $P(N<\infty)=1$，注意若有 $(b-a)$ 个连续的 $+1$ 步，则会离开区间. 由此可得

$$P(N>m(b-a))\leqslant(1-2^{-(b-a)})^m,$$

故 $EN<\infty$.

显见 $E|S_N|<\infty$ 且 $S_n1_{\{N>n\}}$ 一致可积，故由定理 4.8.2，

$$x=ES_N=aP_x(S_N=a)+b[1-P_x(S_N=a)].$$

整理即得 $P_x(S_N=a)=(b-x)/(b-a)$，减去 1 即得 $P_x(S_N=b)=(x-a)/(b-a)$.

(b) 第二个结果由第一个直接推出.

对有界停时 $N\wedge n$ 运用停时定理，有

$$0=E_0S_{N\wedge n}^2-E_0(N\wedge n).$$

单调收敛定理蕴含 $E_0(N\wedge n)\uparrow E_0N$. 由有界收敛定理及 (a) 中 $x=0$ 的结果，

$$E_0S_{N\wedge n}^2\to a^2\frac{b}{b-a}+b^2\frac{-a}{b-a}=-ab+\frac{ab^2-a^2b}{b-a}=-ab.$$

□

注释. 读者应仔细研习 (b) 中的技巧，因其在多种情形下均颇有用. 对有界停时 $T\wedge n$ 应用定理 4.4.1，再令 $n\to\infty$，并使用适当的收敛定理.

定理 4.8.8. 设 S_n 为从 0 出发的对称简单随机游走，令 $T_1=\min\{n:S_n=1\}$. 则

$$Es^{T_1}=\frac{1-\sqrt{1-s^2}}{s}.$$

反演生成函数可得

$$P(T_1=2n-1)=\frac{1}{2n-1}\cdot\frac{(2n)!}{n!n!}2^{-2n}. \tag{4.8.1}$$

证明. 下用指数鞅 $X_n=\exp(\theta S_n)/\varphi(\theta)^n$ 且 $\theta>0$. 定理 4.8.7 后的注记蕴含 $P_0(T_1<\infty)=1$.

$$\varphi(\theta)=E\exp(\theta\xi_i)=(e^\theta+e^{-\theta})/2.$$

$\varphi(0)=1$，$\varphi'(0)=0$ 且 $\varphi''(\theta)>0$，故当 $\theta>0$ 时 $\varphi(\theta)>1$. 这蕴含 $X_{N\wedge n}\in[0,e^\theta]$，由有界收敛定理得 $1=EX(T_1)$，从而

$$e^{-\theta}=E(\varphi(\theta)^{-T_1}).$$

为将其转化为生成函数的公式，令

$$\varphi(\theta)=\frac{e^{\theta}+e^{-\theta}}{2}=\frac{1}{s}.$$

令 $x=e^{\theta}$ 并整理，需 $x+x^{-1}=2/s$ 或

$$sx^2-2x+s=0.$$

二次方程给出

$$x=\frac{2\pm\sqrt{4-4s^2}}{2s}=\frac{1\pm\sqrt{1-s^2}}{s}.$$

由于 $Es^{T_1}=\sum_{k=1}^{\infty}s^kP(T_1=k)$，需要 $s=0$ 时为 0 的解，即 $(1-\sqrt{1-s^2})/s$.

为反演生成函数，使用 Newton 二项公式

$$(1+t)^a=1+\binom{a}{1}t+\binom{a}{2}t^2+\binom{a}{3}t^3+\cdots$$

其中 $\binom{x}{r}=x(x-1)\cdots(x-r+1)/r!$. 取 $t=-s^2$，$a=1/2$，有

$$\sqrt{1-s^2}=1-\binom{1/2}{1}s^2+\binom{1/2}{2}s^4-\binom{1/2}{3}s^6+\cdots$$

$$\frac{1-\sqrt{1-s^2}}{s}=\binom{1/2}{1}s-\binom{1/2}{2}s^3+\binom{1/2}{3}s^5+\cdots$$

s^{2n-1} 的系数为

$$(-1)^{n-1}\frac{(1/2)(-1/2)\cdots(3-2n)/2}{n!}=\frac{1\cdot3\cdots(2n-3)}{n!}\cdot2^{-n}=\frac{1}{2n-1}\frac{(2n)!}{n!n!}2^{-2n},$$

□

定理 4.8.9. 非对称简单随机游走指 $P(\xi_i=1)=p$ 且 $P(\xi_i=-1)=q\equiv1-p$ 而 $p\neq q$ 的特殊情形.

(a) 若 $\varphi(y)=\{(1-p)/p\}^y$，则 $\varphi(S_n)$ 为鞅.

(b) 若令 $T_z=\inf\{n:S_n=z\}$，则对 $a<x<b$

$$P_x(T_a<T_b)=\frac{\varphi(b)-\varphi(x)}{\varphi(b)-\varphi(a)},\quad P_x(T_b<T_a)=\frac{\varphi(x)-\varphi(a)}{\varphi(b)-\varphi(a)}.\tag{4.8.2}$$

对最后两部分设 $1/2<p<1$.

(c) 若 $a<0$ 则 $P\left(\min_n S_n\leqslant a\right)=P(T_a<\infty)=\{(1-p)/p\}^{-a}$.

(d) 若 $b>0$ 则 $P(T_b<\infty)=1$ 且 $ET_b=b/(2p-1)$.

证明. 由于 S_n 与 ξ_{n+1} 独立，由例 4.1.7 知，在 $\{S_n=m\}$ 上

$$\begin{aligned}E(\varphi(S_{n+1})\,|\,\mathcal{F}_n)&=p\cdot\left(\frac{1-p}{p}\right)^{m+1}+(1-p)\cdot\left(\frac{1-p}{p}\right)^{m-1}\\&=\left(\frac{1-p}{p}\right)^{m}\{1-p+p\}=\varphi(S_n),\end{aligned}$$

即得 (a).

令 $N=T_a\wedge T_b$. 由于 $\varphi(S_{N\wedge n})$ 有界，定理 4.8.2 蕴含

$$\varphi(x)=E\varphi(S_N)=P_x(T_a<T_b)\varphi(a)+[1-P_x(T_a<T_b)]\varphi(b).$$

整理即得 $P_x(T_a<T_b)$ 的公式，减去 1 即得 $P_x(T_b<T_a)$ 的公式.

令 $b\to\infty$ 并注意 $\varphi(b)\to 0$ 即得 (c) 中结果，因 $T_a<\infty$ 当且仅当对某个 b 有 $T_a<T_b$. 下证 (d)，注意 $\varphi(a)\to\infty$ 当 $a\to-\infty$，故 $P(T_b<\infty)=1$. 对第二个结论，注意 $X_n=S_n-(p-q)n$ 为鞅. 由于 $T_b\wedge n$ 为有界停时，定理 4.4.1 蕴含

$$0=E(S_{T_b\wedge n}-(p-q)(T_b\wedge n)).$$

由 $b\geqslant S_{T_b\wedge n}\geqslant \min_m S_m$ 及 (c) 蕴含 $E(\inf_m S_m)>-\infty$，控制收敛定理给出 $ES_{T_b\wedge n}\to ES_{T_b}$ 当 $n\to\infty$. 单调收敛定理蕴含 $E(T_b\wedge n)\uparrow ET_b$，故有

$$b=(p-q)ET_b.$$ □

例题 4.8.10. 一赌徒在玩轮盘赌，每次押注 1 美元于黑色. 她赢 1 美元的概率为 18/38，输 1 美元的概率为 20/38. 计算从 20 美元开始，在输光之前达到 40 美元的概率.

由 $(1-p)/p=20/18$，利用 (4.8.2) 有

$$P_{20}(T_{40}<T_0)=\frac{(10/9)^{20}-1}{(10/9)^{40}-1}=\frac{8.225-1}{67.655-1}=0.1083.$$

习题

习题 4.8.1. 推广定理 4.8.2 以证明：若 $L\leqslant M$ 为停时且 $Y_{M\wedge n}$ 为一致可积下鞅，则 $EY_L\leqslant EY_M$，且

$$Y_L\leqslant E(Y_M\,|\,\mathcal{F}_L).$$

习题 4.8.2. 若 $X_n\geqslant 0$ 为上鞅，则 $P(\sup_n X_n>\lambda)\leqslant EX_0/\lambda$.

习题 4.8.3. 设 $S_n = \xi_1 + \cdots + \xi_n$，其中 ξ_i 独立，$E\xi_i = 0$ 且 $\mathrm{var}(\xi_i) = \sigma^2$. $S_n^2 - n\sigma^2$ 为鞅. 令 $T = \min\{n : |S_n| > a\}$. 利用定理 4.8.2 证明 $ET \geqslant a^2/\sigma^2$.

习题 4.8.4. Wald 第二等式. 设 $S_n = \xi_1 + \cdots + \xi_n$，其中 ξ_i 独立，$E\xi_i = 0$ 且 $\mathrm{var}(\xi_i) = \sigma^2$. 利用上题的鞅证明：若 T 为满足 $ET < \infty$ 的停时，则 $ES_T^2 = \sigma^2 ET$.

习题 4.8.5. 赌徒破产时间的方差. 设 $\xi_1, \xi_2, \ldots$ 独立，$P(\xi_i = 1) = p$ 且 $P(\xi_i = -1) = q = 1 - p$，其中 $p < 1/2$. 令 $S_n = S_0 + \xi_1 + \cdots + \xi_n$，并令 $V_0 = \min\{n \geqslant 0 : S_n = 0\}$. 定理 4.8.9 表明 $E_x V_0 = x/(1-2p)$. 本题旨在计算 V_0 的方差. 若令 $Y_i = \xi_i - (p-q)$ 并注意 $EY_i = 0$ 且 $\mathrm{var}(Y_i) = \mathrm{var}(\xi_i) = E\xi_i^2 - (E\xi_i)^2$，则 $(S_n - (p-q)n)^2 - n(1-(p-q)^2)$ 为鞅.

(a) 利用此证明当 $S_0 = x$ 时 V_0 的方差为 $x \cdot \dfrac{1-(p-q)^2}{(q-p)^3}$.

(b) 为什么 (a) 中的答案必为 cx 的形式?

习题 4.8.6. 赌徒破产时间的生成函数. 沿用上一题的设定.

(a) 利用指数鞅及停时定理推出：若 $\theta \leqslant 0$，则 $e^{\theta x} = E_x(\varphi(\theta)^{-V_0})$.

(b) 设 $0 < s < 1$. 解方程 $\varphi(\theta) = 1/s$，然后利用 (a) 推出

$$E_x(s^{V_0}) = \left(\frac{1 - \sqrt{1 - 4pqs^2}}{2ps} \right)^x .$$

(c) 为什么 (b) 中的答案必为 $f(s)^x$ 的形式?

习题 4.8.7. 设 S_n 为从 0 出发的对称简单随机游走，$T = \inf\{n : S_n \notin (-a, a)\}$，其中 a 为整数. 求常数 b 和 c 使得 $Y_n = S_n^4 - 6nS_n^2 + bn^2 + cn$ 为鞅，并利用此计算 ET^2.

习题 4.8.8. 设 $S_n = \xi_1 + \cdots + \xi_n$ 为随机游走. 设 $\varphi(\theta_o) = E\exp(\theta_o \xi_i) = 1$ 对某个 $\theta_o < 0$ 成立且 ξ_i 非常数. 在这一指数鞅的特殊情形 $X_n = \exp(\theta_o S_n)$ 为鞅. 令 $\tau = \inf\{n : S_n \notin (a, b)\}$ 且 $Y_n = X_{n \wedge T}$. 利用定理 4.8.2 推出 $EX_\tau = 1$ 且 $P(S_\tau \leqslant a) \leqslant \exp(-\theta_o a)$.

习题 4.8.9. 沿用上一题的设定，设 ξ_i 取整数值，$P(\xi_i < -1) = 0$，$P(\xi_i = -1) > 0$ 且 $E\xi_i > 0$. 令 $T = \inf\{n : S_n = a\}$ 且 $a < 0$. 利用鞅 $X_n = \exp(\theta_o S_n)$ 推出 $P(T < \infty) = \exp(-\theta_o a)$.

习题 4.8.10. 考虑一个有利博弈，其收益为 $-1, 1$ 或 2，概率各为 $1/3$. 利用上一题的结果计算当从 i 美元开始时最终破产（即盈利 W_n 达到 \$0）的概率.

习题 4.8.11. 设 S_n 为保险公司在第 n 年末的总资产. 在第 n 年，收到保费总计 $c > 0$ 并赔付索赔 ζ_n，其中 ζ_n 服从 Normal(μ, σ^2) 且 $\mu < c$. 准确地说，若 $\xi_n = c - \zeta_n$，则 $S_n = S_{n-1} + \xi_n$. 若资产降至 0 或以下则公司破产. 证明若 $S_0 > 0$ 非随机，则

$$P(\text{破产}) \leqslant \exp(-2(c-\mu)S_0/\sigma^2).$$

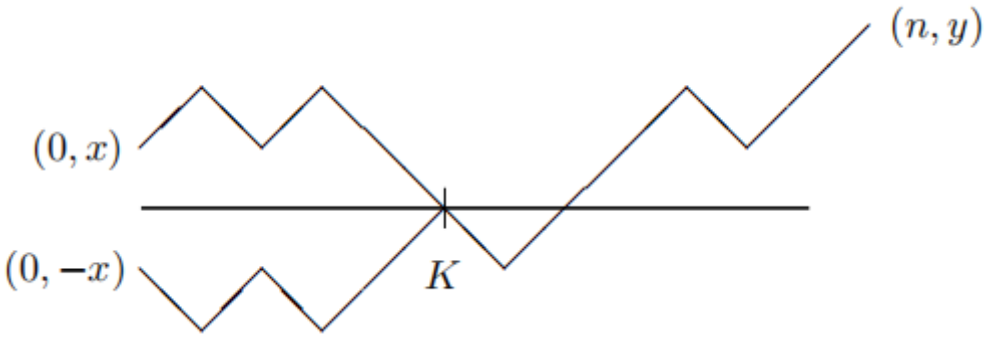

图 4.5: 反射原理

第 9 节 简单随机游走的组合学 *

上节利用鞅证明了一些关于简单随机游走的结果. 本节运用组合论证深入探讨其性质，且不使用鞅，作为下一章 Markov 链研究的过渡.

为方便讨论，将序列 $S_0, S_1, S_2, \ldots, S_n$ 表示为以线段 $(k-1, S_{k-1}) \to (k, S_k)$ 构成的折线. **路径**是指简单随机游走可能的折线结果. 为计算从 $(0,0)$ 到 (n,x) 的路径数，引入 a 和 b: $a = (n+x)/2$ 为路径中正步数，$b = (n-x)/2$ 为负步数. 注意 $n = a+b$ 且 $x = a-b$. 若 $-n \leqslant x \leqslant n$ 且 $n-x$ 为偶数，则上述定义的 a 和 b 为非负整数，从 $(0,0)$ 到 (n,x) 的路径数为

$$N_{n,x} = \binom{n}{a}. \tag{4.9.1}$$

否则，路径数为 0.

定理 4.9.1. 反射原理. 设 $x, y > 0$，则从 $(0,x)$ 到 (n,y) 在某时刻经过 0 的路径数等于从 $(0,-x)$ 到 (n,y) 的路径数.

证明. 设 $(0,s_0), (1,s_1), \ldots, (n,s_n)$ 为从 $(0,x)$ 到 (n,y) 的路径. 令 $K = \inf\{k : s_k = 0\}$. 对 $k \leqslant K$ 令 $s'_k = -s_k$，对 $K \leqslant k \leqslant n$ 令 $s'_k = s_k$. 则 (k, s'_k)，$0 \leqslant k \leqslant n$，为从 $(0,-x)$ 到 (n,y) 的路径.

反之，若 $(0,t_0), (1,t_1), \ldots, (n,t_n)$ 为从 $(0,-x)$ 到 (n,y) 的路径，则它必穿过 0. 令 $K = \inf\{k : t_k = 0\}$. 对 $k \leqslant K$ 令 $t'_k = -t_k$，对 $K \leqslant k \leqslant n$ 令 $t'_k = t_k$. 则 (k, t'_k)，$0 \leqslant k \leqslant n$，为从 $(0,-x)$ 到 (n,y) 且在时刻 K 为 0 的路径. 最后两个观察建立了两类路径之间的一一对应，故其数目相等. □

由定理 4.9.1 可得一个首次证明于 1878 年的结果.

定理 4.9.2. 投票定理. 设一次选举中候选人 A 获得 α 票，候选人 B 获得 β 票，其中 $\beta < \alpha$. A 在计票过程中始终领先 B 的概率为 $(\alpha-\beta)/(\alpha+\beta)$.

证明. 令 $x = \alpha - \beta$，$n = \alpha + \beta$. 显然，此类结果数等于从 $(1,1)$ 到 (n,x) 且永不为 0 的路径数. 反射原理表明从 $(1,1)$ 到 (n,x) 在某时刻为 0 的路径数等于从 $(1,-1)$ 到 (n,x) 的路径数，故由 (4.9.1)，

从 $(1,1)$ 到 (n,x) 且永不为 0 的路径数为

$$\begin{aligned} N_{n-1,x-1} - N_{n-1,x+1} &= \binom{n-1}{\alpha-1} - \binom{n-1}{\alpha} \\ &= \frac{(n-1)!}{(\alpha-1)!(n-\alpha)!} - \frac{(n-1)!}{\alpha!(n-\alpha-1)!} \\ &= \frac{\alpha-(n-\alpha)}{n} \cdot \frac{n!}{\alpha!(n-\alpha)!} = \frac{\alpha-\beta}{\alpha+\beta} N_{n,x}, \end{aligned}$$

由于 $n=\alpha+\beta$，即得所求. □

利用投票定理，可计算简单随机游走击中 0 时刻的分布.

引理 4.9.3. $P(S_1 \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2n}=0)$.

证明. $P(S_1>0,\ldots,S_{2n}>0) = \sum_{r=1}^{\infty} P(S_1>0,\ldots,S_{2n-1}>0, S_{2n}=2r)$. 由定理 4.9.2 的证明可见，从 $(0,0)$ 到 $(2n,2r)$ 且在正时刻永不为 0 的路径数（等于从 $(1,1)$ 到 $(2n,2r)$ 且永不为 0 的路径数）为

$$N_{2n-1,2r-1} - N_{2n-1,2r+1}.$$

令 $p_{n,x}=P(S_n=x)$，则此蕴含

$$P(S_1>0,\ldots,S_{2n-1}>0,S_{2n}=2r) = \frac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1}).$$

对 $r=1$ 到 ∞ 求和得

$$P(S_1>0,\ldots,S_{2n}>0) = \frac{1}{2}p_{2n-1,1} = \frac{1}{2}P(S_{2n}=0).$$

对称性蕴含 $P(S_1<0,\ldots,S_{2n}<0) = (1/2)P(S_{2n}=0)$，证毕. □

令 $R=\inf\{m \geqslant 1 : S_m=0\}$. 将引理 4.9.3 与二项分布的中心极限定理（定理 3.1.2）结合，得

$$P(R>2n) = P(S_{2n}=0) \sim \pi^{-1/2} n^{-1/2}. \tag{4.9.2}$$

由于 $P(R>x)/P(|R|>x)=1$，由定理 3.8.8 知 R 属于指数为 $\alpha=1/2$ 且 $\kappa=1$ 的稳定律的吸引域. 这蕴含若 R_n 为第 n 次返回 0 的时刻，则 $R_n/n^2 \Rightarrow Y$，其中 Y 为所指稳定律. 在例 3.8.5 中，考虑过 $\tau=T_1$，其中 $T_x=\inf\{n : S_n=x\}$. 由于 $S_1 \in \{-1,1\}$ 且 $T_{-1} \stackrel{d}{=} T_1$，$R \stackrel{d}{=} 1+T_1$，从而 $T_n/n^2 \Rightarrow Y^{-1}$，同一稳定律. 在例 8.1.12 中，将利用这一观察证明极限与 Brown 运动击中 1 的时刻同分布，其密度由 (7.4.6) 给出.

由 (4.9.2) 得

$$\begin{aligned} P(T_1 = 2n-1) = P(R = 2n) &= P(R > 2n-2) - P(R > 2n) \\ &= 2^{-2(n-1)} \frac{(2n-2)!}{(n-1)!(n-1)!} - 2^{-2n} \frac{(2n)!}{n!n!} \\ &= 2^{-2n} \frac{(2n)!}{n!n!} \left(4 \frac{n \cdot n}{(2n-1) \cdot 2n} - 1 \right) \\ &= \frac{1}{2n-1} \frac{(2n)!}{n!n!} 2^{-2n} \sim \frac{1}{2} \pi^{-1/2} n^{-3/2}. \end{aligned}$$

以上完成了对访问 0 的讨论. 下面转向反正弦律. 第一个涉及

$$L_{2n} = \sup\{m \leqslant 2n : S_m = 0\}.$$

计算 L_{2n} 的分布出奇地容易.

引理 4.9.4. 令 $u_{2m} = P(S_{2m} = 0)$. 则 $P(L_{2n} = 2k) = u_{2k} u_{2n-2k}$.

证明. $P(L_{2n} = 2k) = P(S_{2k} = 0, S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0)$，故由引理 4.9.3 即得所求. □

定理 4.9.5. 最后访问 0 的反正弦律. 对 $0 < a < b < 1$,

$$P(a \leqslant L_{2n}/2n \leqslant b) \to \int_a^b \pi^{-1} (x(1-x))^{-1/2} \, \mathrm{d}x.$$

为理解名称由来，在积分中作替换 $y = x^{1/2}$，$\mathrm{d}y = (1/2) x^{-1/2} \, \mathrm{d}x$，得

$$\int_{\sqrt{a}}^{\sqrt{b}} \frac{2}{\pi} (1-y^2)^{-1/2} \, \mathrm{d}y = \frac{2}{\pi} \{\arcsin(\sqrt{b}) - \arcsin(\sqrt{a})\}.$$

由于 L_{2n} 是 $2n$ 之前最后一个零点的时刻，答案关于 $1/2$ 对称令人惊讶. 极限分布的对称性蕴含

$$P(L_{2n}/2n \leqslant 1/2) \to 1/2.$$

用赌博术语，若两人每天掷硬币赌 1 美元，则其中一人在概率 $1/2$ 下从 7 月 1 日至年末始终领先，这一事件无疑会令另一人抱怨运气不佳.

定理 4.9.5 的证明. 由 u_{2n} 的渐近公式知，若 $k/n \to x$ 则

$$nP(L_{2n} = 2k) \to \pi^{-1} (x(1-x))^{-1/2}.$$

为由此得到所求，令 $2na_n =$ 大于等于 $2na$ 的最小偶整数，$2nb_n =$ 小于等于 $2nb$ 的最大偶整数，并对 $2k/2n \leqslant x < 2(k+1)/2n$ 令 $f_n(x) = nP(L_{2n} = k)$，则可写

$$P(a \leqslant L_{2n}/2n \leqslant b) = \sum_{k=na_n}^{nb_n} P(L_{2n} = 2k) = \int_{a_n}^{b_n+1/n} f_n(x)\,\mathrm{d}x.$$

第一个结果蕴含在紧集上一致地 $f_n(x) \to f(x) = \pi^{-1}(x(1-x))^{-1/2}$. 一致收敛性给出

$$\sup_{a_n \leqslant x \leqslant b_n+1/n} f_n(x) \to \sup_{a \leqslant x \leqslant b} f(x) < \infty$$

当 $0 < a \leqslant b < 1$，故由有界收敛定理

$$\int_{a_n}^{b_n+1/n} f_n(x)\,\mathrm{d}x \to \int_a^b f(x)\,\mathrm{d}x. \qquad \square$$

下一结果直接涉及一人领先的时间长度.

定理 4.9.6. 位于 0 之上的时间的反正弦律. 令 π_{2n} 为位于横轴之上（即于 $\{(x,y) : y \geqslant 0\}$ 中）的线段 $(k-1, S_{k-1}) \to (k, S_k)$ 的条数，且令 $u_m = P(S_m = 0)$. 则

$$P(\pi_{2n} = 2k) = u_{2k}u_{2n-2k}$$

从而，若 $0 < a < b < 1$

$$P(a \leqslant \pi_{2n}/2n \leqslant b) \to \int_a^b \pi^{-1}(x(1-x))^{-1/2}\,\mathrm{d}x.$$

注释. 由于 $\pi_{2n} \overset{d}{=} L_{2n}$，第二个结论由定理 4.9.5 的证明可得. 读者应注意极限密度 $\pi^{-1}(x(1-x))^{-1/2}$ 在 $x = 1/2$ 处取最小值，且当 $x \to 0$ 或 1 时趋于 ∞. 因此，正负两侧步数均等划分是最不可能的情形，而完全偏向一侧的划分概率最高.

证明. 令 $\beta_{2k,2n}$ 记所讨论的概率. 对 n 用归纳法证明 $\beta_{2k,2n} = u_{2k}u_{2n-2k}$. 当 $n = 1$ 时，显然

$$\beta_{0,2} = \beta_{2,2} = 1/2 = u_0u_2.$$

对一般的 n，首先设 $k = n$. 由引理 4.9.3 的证明，

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

最后一步等式由以下观察得出：若 $S_{2n-1} \geqslant 0$ 则 $S_{2n} \geqslant 1$，从而 $S_{2n} \geqslant 0$.

最后一计算证明了 $k=n$ 时的结果. 由于 $\beta_{0,2n}=\beta_{2n,2n}$，$k=0$ 时结果亦成立. 下设 $1\leqslant k\leqslant n-1$. 在此情形，若 R 为首次返回 0 的时刻，则 $R=2m$ 对某满足 $0<m<n$ 的 m 成立. 令 $f_{2m}=P(R=2m)$ 并按首次游程位于正侧或负侧分解，得

$$\beta_{2k,2n}=\frac{1}{2}\sum_{m=1}^{k}f_{2m}\beta_{2k-2m,2n-2m}+\frac{1}{2}\sum_{m=1}^{n-k}f_{2m}\beta_{2k,2n-2m}.$$

由归纳假设，

$$\beta_{2k,2n}=\frac{1}{2}u_{2n-2k}\sum_{m=1}^{k}f_{2m}u_{2k-2m}+\frac{1}{2}u_{2k}\sum_{m=1}^{n-k}f_{2m}u_{2n-2k-2m}.$$

由首次返回 0 的时刻，可见

$$\sum_{m=1}^{k}f_{2m}u_{2k-2m}=u_{2k},\quad \sum_{m=1}^{n-k}f_{2m}u_{2n-2k-2m}=u_{2n-2k},$$

即得所求. □

习题

习题 4.9.1. 设 $a\in S$，$f_n=P_a(T_a=n)$，$u_n=P_a(X_n=a)$.

(i) 证明 $u_n=\sum_{1\leqslant m\leqslant n}f_m u_{n-m}$.

(ii) 令 $u(s)=\sum_{n\geqslant 0}u_n s^n$，$f(s)=\sum_{n\geqslant 1}f_n s^n$，证明 $u(s)=1/(1-f(s))$.

第五章
Markov 链

本章重点研究（时间齐次）马尔可夫链. 由于该模型的假设在许多实际案例中都能成立，且能由此推导出深入而系统的理论体系，因此具有极高的重要性.

第 1 节　例子

本节从可数状态空间开始讨论. 在此情形下，**Markov 性质**是指：对任意状态 $i_0, \dots, i_{n-1}, i$ 和 j

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i).$$

换言之，给定当前状态，过去的信息对预测未来无关紧要.

本节给出若干例子. 每种情形下，我们直接给出**转移概率**

$$p(i, j) = P(X_{n+1} = j \mid X_n = i)$$

而将验证 Markov 性质的工作留给读者. 下一节将给出更正式的阐述.

例题 5.1.1. 随机游走. 设 $\xi_1, \xi_2, \dots \in \mathbb{Z}^d$ 为独立随机变量，分布为 μ. 令 $X_n = X_0 + \xi_1 + \cdots + \xi_n$，其中 X_0 为常数. 则 X_n 是 Markov 链，其转移概率为

$$p(i, j) = \mu(\{j - i\}).$$

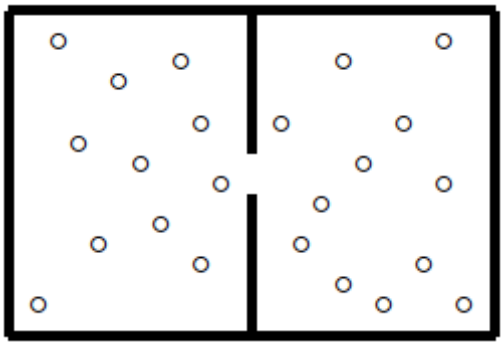

图 5.1: Ehrenfest 链的物理启发

例题 5.1.2. 分支过程. 状态空间 $S = \{0, 1, 2, \ldots\}$

$$p(i, j) = P\left(\sum_{m=1}^{i} \xi_m = j\right),$$

其中 $\xi_1, \xi_2, \ldots$ 为独立同分布的非负整数值随机变量. 换言之，时刻 n（或第 n 代）的 i 个个体各自独立同分布地产生后代.

上述两链不是理想例子，因为正如我们将看到的，它们不收敛到平衡态.

例题 5.1.3. Ehrenfest 链. 状态空间 $S = \{0, 1, \ldots, r\}$

$$\begin{aligned} p(k, k+1) &= (r-k)/r, \\ p(k, k-1) &= k/r, \\ p(i, j) &= 0 \quad \text{其他情形.} \end{aligned}$$

换言之，两个罐子中共有 r 个球；第一个罐子中有 k 个，第二个中有 $r-k$ 个. 随机选取 r 个球中的一个并将其移到另一罐子. Ehrenfest 用此模型描述通过小孔相连的两个相等容器之间的空气分子分布.

例题 5.1.4. Wright-Fisher 模型. 考虑 $N/2$ 个二倍体个体（每个染色体有两个拷贝）或 N 个单倍体个体（每个染色体有一个拷贝）的种群，即 N 个基因的固定种群，可以是两种类型之一：A 或 a. 在此模型的最简单版本中，时刻 $n+1$ 的种群通过从时刻 n 的种群中有放回抽样得到. 此时，若令 X_n 为时刻 n 的 A 等位基因数量，则 X_n 是 Markov 链，其转移概率为

$$p(i, j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j},$$

因右边是 N 次独立试验的二项分布，成功概率为 i/N.

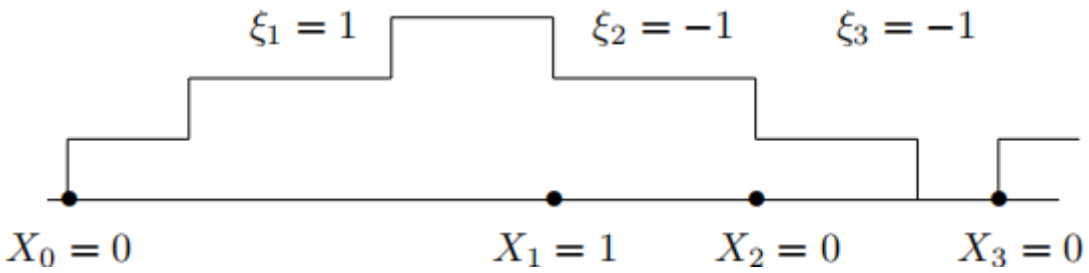

图 5.2: M/G/1 排队系统的实例，黑点代表顾客到达的时间

在此模型中，状态 $x = 0$ 和 N 对应于种群固定在全 a 或全 A 状态，它们是**吸收态**，即 $p(x, x) = 1$. 正如我们将看到的，此链最终会进入状态 0 或状态 N. 为改进这个简单模型，引入**突变**. 即，被抽到的 A 在下一代变成 a 的概率为 u，而被抽到的 a 变成 A 的概率为 v. 此时，一次给定抽取产生 A 的概率为

$$\rho_i = \frac{i}{N}(1-u) + \frac{N-i}{N}v,$$

而转移概率仍具有二项形式

$$p(i,j) = \binom{N}{j}(\rho_i)^j(1-\rho_i)^{N-j}.$$

若 u 和 v 均为正，则 0 和 N 不再是吸收态，故系统可以收敛到平衡分布.

例题 5.1.5. M/G/1 排队系统. 在此模型中，顾客按速率为 λ 的 Poisson 过程到达. （M 表示 Markov，指在 Poisson 过程中不相交时间区间内的到达数独立.）每个顾客需要独立的服务时间，分布为 F. （G 表示一般服务分布；1 表示有一个服务员.）令 X_n 为第 n 个顾客开始服务时队列中等待的顾客数. 确切地，当 $X_0 = x$ 时，链以 x 人排队、顾客 0 刚开始服务开始.

为定义 Markov 链 X_n，令

$$a_k = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^k}{k!}\,\mathrm{d}F(t)$$

为一个服务时间内到达 k 个顾客的概率. 令 $\xi_1, \xi_2, \ldots$ 为独立同分布随机变量，满足 $P(\xi_i = k-1) = a_k$. 我们将 ξ_i 视为第 i 个服务时间内的净到达顾客数，减去完成服务的那个顾客. 定义 X_n 为

$$X_{n+1} = (X_n + \xi_{n+1})^+ \tag{5.1.1}$$

其中正部仅在 $X_n = 0$ 且 $\xi_{n+1} = -1$ 时起作用（例如，图 5.2 中 $X_2 = 0, \xi_3 = -1$），它反映了如下事实：当队列大小为 0 且服务时间内无人到达时，下一个队列大小为 0，因为我们直到下一个顾客到达才开始计数，此时队列长度为 0.

易见 (5.1.1) 中定义的序列是 Markov 链，其转移概率为

$$p(0,0) = a_0 + a_1,$$
$$p(j, j-1+k) = a_k \quad 若j \geqslant 1 或k > 1.$$

a_k 的公式相当复杂，其确切形式并不重要，因此我们简化假设，仅设对所有 $k \geqslant 0$ 有 $a_k > 0$ 且 $\sum_{k\geqslant 0} a_k = 1$.

习题

习题 5.1.1. 设 $\xi_1, \xi_2, \ldots$ 为独立同分布，取值于 $\{1, 2, \ldots, N\}$，每个值概率为 $1/N$. 证明 $X_n = |\{\xi_1, \ldots, \xi_n\}|$ 是 Markov 链并计算其转移概率.

习题 5.1.2. 设 $\xi_1, \xi_2, \ldots$ 为独立同分布，取值于 $\{-1, 1\}$，每个值概率为 $1/2$. 令 $S_0 = 0, S_n = \xi_1 + \cdots + \xi_n$ 且 $X_n = \max\{S_m : 0 \leqslant m \leqslant n\}$. 证明 X_n 不是 Markov 链.

习题 5.1.3. 设 $\xi_0, \xi_1, \ldots$ 为独立同分布，取值于 $\{H, T\}$，每个值概率为 $1/2$. 证明 $X_n = (\xi_n, \xi_{n+1})$ 是 Markov 链并计算其转移概率 p. p^2 是什么?

习题 5.1.4. 兄妹交配. 在此方案中，两只动物交配，在其直系后代中随机选取一对异性个体. 这些动物交配，过程继续. 假设每个个体可以是三种基因型之一：AA, Aa, aa，且后代的类型由从每个亲本选取一个字母决定. 依此规则，第 n 代的基因型对是 Markov 链，具有六个状态：

AA, AA　AA, Aa　AA, aa　Aa, Aa　Aa, aa　aa, aa

计算其转移概率.

习题 5.1.5. Bernoulli-Laplace 扩散模型. 假设左右两个罐子各有 m 个球. b 个（设 $\leqslant m$）球是黑色的，$2m - b$ 个是白色的. 每次从每个罐子中各取一个球并交换. 令时刻 n 的状态为左罐中黑球的数量. 计算转移概率.

习题 5.1.6. 设 $\theta, U_1, U_2, \ldots$ 独立且服从 $(0, 1)$ 上的均匀分布. 令 $X_i = 1$ 若 $U_i \leqslant \theta$，$= -1$ 若 $U_i > \theta$，且令 $S_n = X_1 + \cdots + X_n$. 换言之，首先按均匀分布选取 θ，然后以概率 θ 掷一枚正面朝上的硬币来生成随机游走. 计算 $P(X_{n+1} = 1 \mid X_1, \ldots, X_n)$ 并得出 S_n 是时非齐次 Markov 链的结论. 这是由于“S_n 是估计 θ 的充分统计量”.

第 2 节　构造与 Markov 性质

设 $(S,\mathcal{S})$ 为可测空间. 这将是我们 Markov 链的**状态空间**.

函数 $p : S \times \mathcal{S} \to \mathbb{R}$ 称为**转移概率**，若:

(i) 对每个 $x \in S$，$A \mapsto p(x,A)$ 是 $(S,\mathcal{S})$ 上的概率测度.

(ii) 对每个 $A \in \mathcal{S}$，$x \mapsto p(x,A)$ 是可测函数.

称 X_n 为（关于 $\mathcal{F}_n$ 的）具有转移概率 p 的 **Markov 链**，若

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B).$$

给定转移概率 p 和 $(S,\mathcal{S})$ 上的初始分布 μ，可通过

$$P(X_j \in B_j, 0 \leqslant j \leqslant n) = \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n) \tag{5.2.1}$$

定义相容的有限维分布族.

若假设 $(S,\mathcal{S})$ 是**好空间**，Kolmogorov 扩张定理（定理 2.1.21）允许我们在序列空间

$$(\Omega_o, \mathcal{F}_\infty) = (S^{\{0,1,\ldots\}}, \mathcal{S}^{\{0,1,\ldots\}})$$

上构造概率测度 P_μ，使得坐标映射 $X_n(\omega) = \omega_n$ 具有所需的分布. 注意我们有一组非常简单的随机变量，以及一大族测度，每个初始条件一个. 当然，我们每个状态只需要一个测度，因为

$$P_\mu(A) = \int \mu(\mathrm{d}x) P_x(A).$$

在此**典则概率空间**上构造链的一个优点是我们可以定义**移位算子**

$$\theta_n(\omega_0, \omega_1, \ldots) = (\omega_n, \omega_{n+1}, \ldots).$$

定理 5.2.1. X_n 是（关于 $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ 的）具有转移概率 p 的 Markov 链. 即

$$P_\mu(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B).$$

证明. 下证此结论. 令 $A = \{X_0 \in B_0, X_1 \in B_1, \ldots, X_n \in B_n\}$，$B_{n+1} = B$，并注意利用积分的定义、

A 的定义以及 P_μ 的定义

$$\begin{aligned}\int_A 1_{(X_{n+1}\in B)}\,\mathrm{d}P_\mu &= P_\mu(A, X_{n+1}\in B)\\ &= P_\mu(X_0\in B_0, X_1\in B_1,\dots,X_n\in B_n, X_{n+1}\in B)\\ &= \int_{B_0}\mu(\mathrm{d}x_0)\int_{B_1}p(x_0,\mathrm{d}x_1)\cdots\int_{B_n}p(x_{n-1},\mathrm{d}x_n)p(x_n,B_{n+1}).\end{aligned}$$

下证最后一个表达式等于

$$=\int_A p(X_n,B)\,\mathrm{d}P_\mu.$$

为此，首先注意对任意 $C\in\mathcal{S}$

$$\int_{B_0}\mu(\mathrm{d}x_0)\int_{B_1}p(x_0,\mathrm{d}x_1)\cdots\int_{B_n}p(x_{n-1},\mathrm{d}x_n)1_C(x_n)=\int_A 1_C(X_n)\,\mathrm{d}P_\mu.$$

线性性蕴含对简单函数成立

$$\int_{B_0}\mu(\mathrm{d}x_0)\int_{B_1}p(x_0,\mathrm{d}x_1)\cdots\int_{B_n}p(x_{n-1},\mathrm{d}x_n)f(x_n)=\int_A f(X_n)\,\mathrm{d}P_\mu,$$

而控制收敛定理蕴含对有界可测 f 成立，例如 $f(x)=p(x,B_{n+1})$.

使命等式

$$\int_A 1_{(X_{n+1}\in B)}\,\mathrm{d}P_\mu=\int_A p(X_n,B)\,\mathrm{d}P_\mu$$

成立的集合构成一个 λ-系，而已证明的集合构成一个 π-系，故由 π-λ 定理（定理 2.1.6），等式对所有 $A\in\mathcal{F}_n$ 成立. 这表明

$$P(X_{n+1}\in B\mid\mathcal{F}_n)=p(X_n,B),$$

证毕. □

下证：若 X_n 具有转移概率 p，则对任意有界可测 f

$$E(f(X_{n+1})\mid\mathcal{F}_n)=\int p(X_n,\mathrm{d}y)f(y). \tag{5.2.2}$$

所需结论由下面结果导出，它将节省后续证明的工作量. 令 $\mathcal{H}$ 为使恒等式成立的有界函数集合.

定理 5.2.2. 单调类定理. 设 $\mathcal{A}$ 为包含 Ω 的 π-系，$\mathcal{H}$ 为满足以下条件的实值函数集合：

(i) 若 $A\in\mathcal{A}$，则 $1_A\in\mathcal{H}$.

(ii) 若 $f,g\in\mathcal{H}$，则 $f+g$ 和 $cf\in\mathcal{H}$ 对任意实数 c 成立.

(iii) 若 $f_n \in \mathcal{H}$ 非负且递增到有界函数 f，则 $f \in \mathcal{H}$.
则 $\mathcal{H}$ 包含所有关于 $\sigma(\mathcal{A})$ 可测的有界函数.

证明. 假设 $\Omega \in \mathcal{A}$，(ii) 和 (iii) 蕴含 $\mathcal{G} = \{A : 1_A \in \mathcal{H}\}$ 是 λ-系，故由 (i) 和 π-λ 定理（定理 2.1.6），$\mathcal{G} \supset \sigma(\mathcal{A})$. (ii) 蕴含 $\mathcal{H}$ 包含所有简单函数，(iii) 蕴含 $\mathcal{H}$ 包含所有有界可测函数. □

为推广 (5.2.2)，注意条件期望的熟知性质和 (5.2.2) 蕴含

$$E\left(\prod_{m=0}^{n} f_m(X_m)\right) = E\left(E\left(\prod_{m=0}^{n} f_m(X_m)\,\middle|\,\mathcal{F}_{n-1}\right)\right) = E\left(\prod_{m=0}^{n-1} f_m(X_m) E(f_n(X_n) \mid \mathcal{F}_{n-1})\right).$$

最后一个积分是 X_{n-1} 的有界可测函数，故由归纳法，若 μ 是 X_0 的分布，则

$$E\left(\prod_{m=0}^{n} f_m(X_m)\right) = \int \mu(\mathrm{d}x_0) f_0(x_0) \int p_0(x_0, \mathrm{d}x_1) f_1(x_1) \cdots \int p_{n-1}(x_{n-1}, \mathrm{d}x_n) f_n(x_n). \tag{5.2.3}$$

下证 Markov 性质的两个推广，其中 $\{X_{n+1} \in B\}$ 被替换为未来的有界函数 $h(X_n, X_{n+1}, \ldots)$，而 n 被替换为停时 N. 这些结果，尤其是第二个，将是发展 Markov 链理论的关键.

定理 5.2.3. Markov 性质. 设 $Y : \Omega_o \to \mathbb{R}$ 有界可测.

$$E_\mu(Y \circ \theta_m \mid \mathcal{F}_m) = E_{X_m} Y.$$

注释. 我们用 Y 表示函数，这个通常用于随机变量的字母，是因为 Y 正是如此，定义在我们概率空间 Ω_o 上的可测函数. 这里左边的下标 μ 表示条件期望是关于 P_μ 取的. 右边是函数 $\varphi(x) = E_x Y$ 在 $x = X_m$ 处的值.

证明. 先在特殊情形下证明结果，然后用 π-λ 和单调类定理得到一般结果. 令 $A = \{\omega : \omega_0 \in A_0, \ldots, \omega_m \in A_m\}$，$g_0, \ldots, g_n$ 有界可测. 将 (5.2.3) 应用于 $f_k = 1_{A_k}$ $(k < m)$，$f_m = 1_{A_m} g_0$，以及 $f_k = g_{k-m}$ $(m < k \leqslant m+n)$ 得

$$\begin{aligned} E_\mu\left(\prod_{k=0}^{n} g_k(X_{m+k}); A\right) &= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} p(x_0, \mathrm{d}x_1) \cdots \int_{A_m} p(x_{m-1}, \mathrm{d}x_m) \\ &\quad \cdot g_0(x_m) \int p(x_m, \mathrm{d}x_{m+1}) g_1(x_{m+1}) \cdots \int p(x_{m+n-1}, \mathrm{d}x_{m+n}) g_n(x_{m+n}) \\ &= E_\mu\left(E_{X_m}\left(\prod_{k=0}^{n} g_k(X_k)\right); A\right). \end{aligned}$$

使命等式成立的集合构成 λ-系，而已证明的集合构成 π-系，故由 π-λ 定理（定理 2.1.6），最后一个恒等式对所有 $A \in \mathcal{F}_m$ 成立.

固定 $A \in \mathcal{F}_m$，令 $\mathcal{H}$ 为使

$$(*) \quad E_\mu(Y \circ \theta_m; A) = E_\mu(E_{X_m} Y; A)$$

成立的有界可测 Y 的集合.

上述计算表明 $(*)$ 当

$$Y(\omega) = \prod_{0 \leqslant k \leqslant n} g_k(\omega_k)$$

时成立. 为完成证明，应用单调类定理（定理 5.2.2）. 令 $\mathcal{A}$ 为形如 $\{\omega : \omega_0 \in A_0, \ldots, \omega_k \in A_k\}$ 的集合族. $\mathcal{A}$ 是 π-系，故取 $g_k = 1_{A_k}$ 表明 (i) 成立. $\mathcal{H}$ 显然具有性质 (ii) 和 (iii)，故定理 5.2.2 蕴含 $\mathcal{H}$ 包含关于 $\sigma(\mathcal{A})$ 可测的有界函数，证毕. □

作为定理 5.2.3 的推论，我们得到

定理 5.2.4. Chapman-Kolmogorov 方程.

$$P_x(X_{m+n} = z) = \sum_y P_x(X_m = y) P_y(X_n = z).$$

直观上，为在 $m+n$ 步内从 x 到 z，必须在时刻 m 到达某处 y，而 Markov 性质意味着对给定的 y，旅程的两部分独立.

证明. $P_x(X_{n+m} = z) = E_x(P_x(X_{n+m} = z \mid \mathcal{F}_m)) = E_x(P_{X_m}(X_n = z))$，由 Markov 性质（定理 5.2.3），因为 $1_{(X_n = z)} \circ \theta_m = 1_{(X_{n+m} = z)}$. □

现给出 Markov 性质的第二个推广. 回顾 N 是**停时**，若 $\{N = n\} \in \mathcal{F}_n$. 如第 4 章，令

$$\mathcal{F}_N = \{A : A \cap \{N = n\} \in \mathcal{F}_n \text{ 对所有} n\}$$

为时刻 N 已知的信息，并定义**随机移位算子**

$$\theta_N(\omega) = \begin{cases} \theta_n(\omega) & \text{在} \{N = n\} \text{ 上,} \\ \Delta & \text{在} \{N = \infty\} \text{ 上,} \end{cases}$$

其中 Δ 是额外添加的点，满足对所有 Y 有 $Y(\Delta) = 0$.

定理 5.2.5. 强 Markov 性质. 设 $Y : \Omega_o \to \mathbb{R}$ 有界可测，N 为停时.

$$E_\mu(Y \circ \theta_N \mid \mathcal{F}_N) = E_{X_N} Y \text{ 在} \{N < \infty\} \text{ 上.}$$

证明. 令 $A \in \mathcal{F}_N$. 将 A 分解为 $A = \bigcup_{n=0}^{\infty} A \cap \{N = n\}$，注意 $A \cap \{N = n\} \in \mathcal{F}_n$，有

$$\begin{aligned} E_\mu(Y \circ \theta_N; A \cap \{N < \infty\}) &= \sum_{n=0}^{\infty} E_\mu(Y \circ \theta_n; A \cap \{N = n\}) \\ &= \sum_{n=0}^{\infty} E_\mu(E_{X_n} Y; A \cap \{N = n\}) \\ &= E_\mu(E_{X_N} Y; A \cap \{N < \infty\}). \end{aligned}$$

□

为更好地理解强 Markov 性质，考虑**返回时间**的例子. 对 $y \in S$，令

$$\begin{aligned} T_y^0 &= 0, \\ T_y^k &= \inf\{n > T_y^{k-1} : X_n = y\} \quad 对 k \geqslant 1, \end{aligned}$$

其中 T_y^k 是第 k 次返回 y 的时刻. 读者应注意 $T_y^1 > 0$，故时刻 0 的任何访问不计入. 我们采用此约定，使得若令 $T_y = T_y^1$ 且 $\rho_{xy} = P_x(T_y < \infty)$，则

定理 5.2.6. $P_x(T_y^k < \infty) = \rho_{xy} \rho_{yy}^{k-1}$.

直观上，为访问 y 共 k 次，必须先从 x 到 y，然后返回 $k-1$ 次到 y.

证明. 当 $k = 1$ 时，结论是平凡的，故设 $k \geqslant 2$.

令 $Y(\omega) = 1$ 若 $\omega_n = y$ 对某 $n \geqslant 1$, 否则 $Y(\omega) = 0$. 若 $N = T_y^{k-1}$, 则在 $\{T_y^k < \infty\}$ 上 $Y \circ \theta_N = 1$. 强 Markov 性质（定理 5.2.5）蕴含

$$E_x(Y \circ \theta_N \mid \mathcal{F}_N) = E_{X_N} Y \quad 在 \{N < \infty\} 上.$$

在 $\{N < \infty\}$ 上，$X_N = y$，故右边为 $P_y(T_y < \infty) = \rho_{yy}$，由此

$$\begin{aligned} P_x(T_y^k < \infty) &= E_x(Y \circ \theta_N; N < \infty) \\ &= E_x(E_x(Y \circ \theta_N \mid \mathcal{F}_N); N < \infty) \\ &= E_x(\rho_{yy}; N < \infty) = \rho_{yy} P_x(T_y^{k-1} < \infty). \end{aligned}$$

结果现由归纳法得出. □

下一个例子说明定理 5.2.5 的应用，并解释为什么我们想让应用于移位路径的 Y 依赖于 n.

定理 5.2.7. 反射原理. 设 $\xi_1, \xi_2, \ldots$ 为独立同分布，分布关于 0 对称. 令 $S_n = \xi_1 + \cdots + \xi_n$. 若 $a > 0$ 则

$$P\left(\sup_{m \leqslant n} S_m \geqslant a\right) \leqslant 2P(S_n \geqslant a).$$

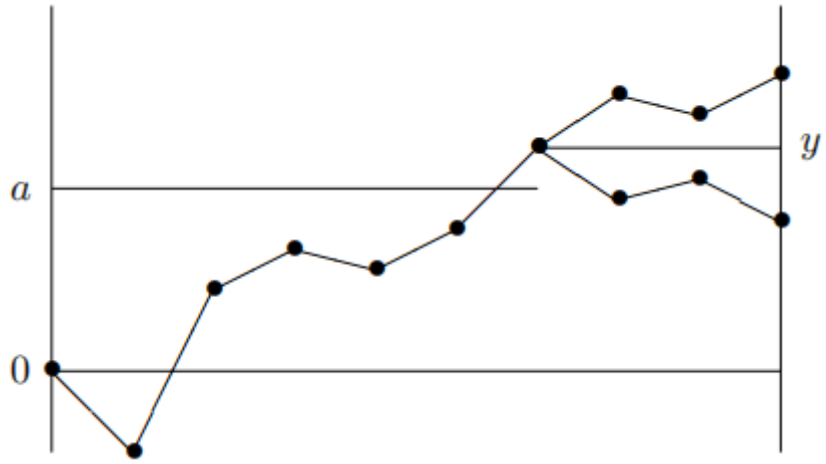

图 5.3: 反射原理的图片证明

我们分两步证明，因这是实践中推导此类公式的方式. 首先，直观计算，然后找出如何从定理 5.2.5 中提取所需公式.

文字证明. 首先注意若 Z 的分布关于 0 对称，则

$$P(Z \geqslant 0) \geqslant P(Z > 0) + \frac{1}{2}P(Z = 0) = \frac{1}{2}.$$

若令 $N = \inf\{m \leqslant n : S_m > a\}$（约定 $\inf\emptyset = \infty$），则在 $\{N < \infty\}$ 上，$S_n - S_N$ 独立于 S_N 且满足 $P(S_n - S_N \geqslant 0) \geqslant 1/2$. 故

$$P(S_n > a) \geqslant \frac{1}{2}P(N \leqslant n).$$

□

证明. 令 $Y_m(\omega) = 1$ 若 $m \leqslant n$ 且 $\omega_{n-m} \geqslant a$，否则 $Y_m(\omega) = 0$. Y_m 的定义选取使得 $(Y_N \circ \theta_N)(\omega) = 1$ 若 $\omega_n \geqslant a$（因此 $N \leqslant n$），否则为 0. 强 Markov 性质蕴含

$$E_0(Y_N \circ \theta_N \mid \mathcal{F}_N) = E_{S_N}Y_N \quad 在\{N < \infty\} = \{N \leqslant n\} 上.$$

为计算右边，注意若 $y > a$，则

$$E_yY_m = P_y(S_{n-m} \geqslant a) \geqslant P_y(S_{n-m} \geqslant y) \geqslant 1/2.$$

故在 $\{N \leqslant n\}$ 上积分并利用条件期望的定义得

$$\frac{1}{2}P(N \leqslant n) \leqslant E_0(E_0(Y_N \circ \theta_N \mid \mathcal{F}_N); N \leqslant n) = E_0(Y_N \circ \theta_N; N \leqslant n),$$

因 $\{N \leqslant n\} \in \mathcal{F}_N$. 回忆 $Y_N \circ \theta_N = 1_{\{S_n \geqslant a\}}$，最后一项

$$= E_0(1_{\{S_n \geqslant a\}}; N \leqslant n) = P_0(S_n \geqslant a),$$

因 $\{S_n > a\} \subset \{N \leqslant n\}$. □

习题

习题 5.2.1. 设 $A \in \sigma(X_0, \ldots, X_n)$ 且 $B \in \sigma(X_n, X_{n+1}, \ldots)$. 用 Markov 性质证明对任意初始分布 μ

$$P_\mu(A \cap B \mid X_n) = P_\mu(A \mid X_n) P_\mu(B \mid X_n).$$

换言之，给定现在，过去与未来条件独立.

提示：将左边写为 $E_\mu(E_\mu(1_A 1_B \mid \mathcal{F}_n) \mid X_n)$.

习题 5.2.2. 设 X_n 为 Markov 链. 用 Lévy 0-1 律证明若

$$P\left(\bigcup_{m=n+1}^{\infty} \{X_m \in B_m\} \middle| X_n \right) \geqslant \delta > 0 \quad \text{在} \{X_n \in A_n\} \text{ 上},$$

则 $P(\{X_n \in A_n \text{ i.o.}\} - \{X_n \in B_n \text{ i.o.}\}) = 0$.

习题 5.2.3. 状态 a 称为**吸收的**，若 $P_a(X_1 = a) = 1$. 令 $D = \{X_n = a \text{ 对某} n \geqslant 1\}$ 且 $h(x) = P_x(D)$. 用上一习题的结果得出 $h(X_n) \to 0$ 几乎必然在 D^c 上成立. 这里几乎必然指 P_μ 几乎必然对任意初始分布 μ.

以下习题假设 S 可数. 注意首次访问的时刻 V_A 和 V_x 是对 $n \geqslant 0$ 取下确界，而击中时 T_A 和 T_x 是对 $n \geqslant 1$ 取下确界.

习题 5.2.4. 首次进入分解. 令 $T_y = \inf\{n \geqslant 1 : X_n = y\}$. 证明

$$p^n(x, y) = \sum_{m=1}^{n} P_x(T_y = m) p^{n-m}(y, y).$$

习题 5.2.5. 证明 $\sum_{m=0}^{n} P_x(X_m = x) \geqslant \sum_{m=k}^{n+k} P_x(X_m = x)$.

习题 5.2.6. 令 $T_C = \inf\{n \geqslant 1 : X_n \in C\}$. 假设 $S - C$ 有限且对每个 $x \in S - C$ 有 $P_x(T_C < \infty) > 0$. 则存在 $N < \infty$ 和 $\epsilon > 0$ 使得对所有 $y \in S - C$，$P_y(T_C > kN) \leqslant (1 - \epsilon)^k$.

习题 5.2.7. 退出分布. 令 $V_C = \inf\{n \geqslant 0 : X_n \in C\}$ 且 $h(x) = P_x(V_A < V_B)$. 假设 $A \cap B = \emptyset$，$S - (A \cup B)$ 有限，且 $P_x(V_{A \cup B} < \infty) > 0$ 对所有 $x \in S - (A \cup B)$.

(i) 证明

$$h(x)=\sum_y p(x,y)h(y)\quad 对x\notin A\cup B. \tag{*}$$

(ii) 证明若 h 满足 $(*)$，则 $h(X(n\wedge V_{A\cup B}))$ 是鞅.

(iii) 用此和习题 5.2.6 得出 $h(x)=P_x(V_A<V_B)$ 是 $(*)$ 的唯一解，在 A 上为 1，在 B 上为 0.

习题 5.2.8. 设 X_n 为 Markov 链，$S=\{0,1,\dots,N\}$，假设 X_n 是鞅. 令 $V_x=\min\{n\geqslant 0:X_n=x\}$ 且假设 $P_x(V_0\wedge V_N<\infty)>0$ 对所有 x. 证明 $P_x(V_N<V_0)=x/N$.

习题 5.2.9. Wright-Fisher 模型. 假设 $S=\{0,1,\dots,N\}$，考虑

$$p(i,j)=\binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j}.$$

证明此链满足习题 5.2.8 的假设.

习题 5.2.10. 在习题 5.1.4 描述的兄妹交配中，AA, AA 和 aa, aa 是吸收态. 证明配对中 A 的数量是鞅，并用此计算从各状态出发被吸收于 AA, AA 的概率.

习题 5.2.11. 退出时间. 令 $V_A=\inf\{n\geqslant 0:X_n\in A\}$ 且 $g(x)=E_xV_A$. 假设 $S-A$ 有限且对每个 $x\in S-A$，$P_x(V_A<\infty)>0$.

(i) 证明

$$g(x)=1+\sum_y p(x,y)g(y)\quad 对x\notin A. \tag{*}$$

(ii) 证明若 g 满足 $(*)$，则 $g(X(n\wedge V_A))+n\wedge V_A$ 是鞅.

(iii) 用此得出 $g(x)=E_x\tau_A$ 是 $(*)$ 的唯一解，在 A 上为 0.

习题 5.2.12. 在此问题中，想象你感到无聊，开始抛硬币，想看看需要抛多少次 $N(HH)$ 才能得到连续两个正面. 令 $\xi_0,\xi_1,\dots$ 为独立同分布，取值于 $\{H,T\}$，每个值概率为 $1/2$，令 $X_n=(\xi_n,\xi_{n+1})$ 为习题 5.1.3 中的 Markov 链. 用上一习题的结果计算 $g(x)=$ 从状态 x 到达状态 (H,H) 的期望时间. 我们想要的答案是 $EN(HH)=2+(1/4)[g(H,H)+g(H,T)+g(T,H)+g(T,T)]$.

习题 5.2.13. 在此问题中，我们考虑得到 HHH 所需的抛掷次数 $N(HHH)$. 令 X_n 为 n 次抛掷后连续正面的数量. X_n 是 Markov 链，转移概率为

	0	1	2	3
0	1/2	1/2	0	0
1	1/2	0	1/2	0
2	1/2	0	0	1/2
3	0	0	0	1

令 T_3 为到达状态 3 的时间. 用问题 5.2.11 计算 E_0T_3.

第 3 节 常返与暂态

设 $T_y^0 = 0$，且对 $k \geqslant 1$，定义

$$T_y^k = \inf\{n > T_y^{k-1} : X_n = y\}$$

回顾 (5.2.6)

$$\rho_{xy} = P_x(T_y < \infty)$$

称状态 y 为**常返的**，若 $\rho_{yy} = 1$；称为**暂态的**，若 $\rho_{yy} < 1$.

若 y 是常返的，由定理 5.2.6 知 $P_y(T_y^k < \infty) = 1$ 对所有 k 成立，故 $P_y(X_n = y \text{ i.o.}) = 1$.

若 y 是暂态的，设 $N(y) = \sum_{n=1}^{\infty} 1_{(X_n=y)}$ 为在正时刻访问 y 的次数，则

$$\begin{aligned} E_x N(y) &= \sum_{k=1}^{\infty} P_x(N(y) \geqslant k) = \sum_{k=1}^{\infty} P_x(T_y^k < \infty) \\ &= \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty \end{aligned} \tag{5.3.1}$$

将上述计算与常返态的结果结合，得到

定理 5.3.1. y 是常返的当且仅当 $E_y N(y) = \infty$.

下面的结果表明常返性具有传染性.

定理 5.3.2. 若 x 是常返的且 $\rho_{xy} > 0$，则 y 也是常返的且 $\rho_{yx} = 1$.

证明. 先证 $\rho_{yx} = 1$. 若 $\rho_{xy} > 0$ 且 $\rho_{yx} < 1$，则 $\rho_{xx} < 1$.

设 $K = \inf\{k : p^k(x, y) > 0\}$. 存在序列 $y_1, \ldots, y_{K-1}$ 使得

$$p(x, y_1)p(y_1, y_2)\cdots p(y_{K-1}, y) > 0$$

由于 K 是最小的，故 $y_i \neq x$ 对 $1 \leqslant i \leqslant K-1$ 成立. 若 $\rho_{yx} < 1$，则有

$$P_x(T_x = \infty) \geqslant p(x, y_1)p(y_1, y_2)\cdots p(y_{K-1}, y)(1-\rho_{yx}) > 0$$

矛盾. 故 $\rho_{yx} = 1$.

为证 y 是常返的，注意 $\rho_{yx} > 0$ 意味着存在 L 使得 $p^L(y, x) > 0$. 于是

$$p^{L+n+K}(y, y) \geqslant p^L(y, x)p^n(x, x)p^K(x, y)$$

对 n 求和，可见

$$\sum_{n=1}^{\infty} p^{L+n+K}(y,y) \geqslant p^L(y,x)p^K(x,y)\sum_{n=1}^{\infty} p^n(x,x) = \infty$$

故由定理 5.3.1 知 y 是常返的. □

下面给出有助于识别常返态的事实. 先引入两个定义.

C 称为**闭集**，若 $x \in C$ 且 $\rho_{xy} > 0$ 蕴含 $y \in C$. 这一名称的由来是：若 C 是闭集且 $x \in C$，则 $P_x(X_n \in C) = 1$ 对所有 n 成立. D 称为**不可约的**，若 $x, y \in D$ 蕴含 $\rho_{xy} > 0$.

定理 5.3.3. 设 C 为有限闭集. 则 C 包含常返态. 若 C 是不可约的，则 C 中所有状态都是常返的.

证明. 由定理 5.3.2，只需证明第一个断言. 假设它不成立，则对所有 $y \in C$，$\rho_{yy} < 1$ 且 $E_xN(y) = \rho_{xy}/(1-\rho_{yy})$，但这导致荒谬的结论：

$$\infty > \sum_{y\in C} E_xN(y) = \sum_{y\in C}\sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty}\sum_{y\in C} p^n(x,y) = \sum_{n=1}^{\infty} 1$$

第一个不等式因 C 有限成立，最后一个等式因 C 是闭集成立. □

下面举例说明上述结果的应用：

例题 5.3.4. 七状态链. 考虑如下转移概率：

	1	2	3	4	5	6	7
1	.3	0	0	0	.7	0	0
2	.1	.2	.3	.4	0	0	0
3	0	0	.5	.5	0	0	0
4	0	0	0	.5	0	.5	0
5	.6	0	0	0	.4	0	0
6	0	0	0	.1	0	.1	.8
7	0	0	0	1	0	0	0

为识别常返态和暂态，先画一个图：若 $p(i,j) > 0$ 且 $i \neq j$，则从 i 到 j 画一条弧. 自环（对应 $p(i,i) > 0$ 的状态）无需画出，因这类转移无法帮助链到达新的地方.

对本例，画弧 $1 \to 5$，$2 \to 1$，$2 \to 3$，$2 \to 4$，$3 \to 4$，$4 \to 6$，$4 \to 7$，$5 \to 1$，$6 \to 4$，$6 \to 7$，$7 \to 4$.

(i) $\rho_{21} > 0$ 而 $\rho_{12} = 0$，故由定理 5.3.2 知 2 必为暂态. 类似地，$\rho_{34} > 0$ 而 $\rho_{43} = 0$，故 3 必为暂态.

(ii) $\{1,5\}$ 和 $\{4,6,7\}$ 是不可约闭集，故由定理 5.3.3 知这些状态是常返的.

上述推理可用于识别有限状态空间 S 中的暂态和常返态. 对 $x \in S$ 只有两种情形：

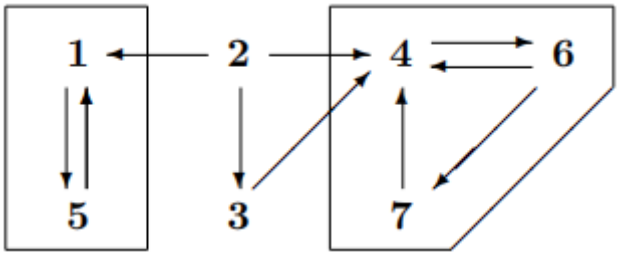

图 5.4: 七状态链

(i) 存在 y 使得 $\rho_{xy} > 0$ 而 $\rho_{yx} = 0$，则 x 必为暂态；或

(ii) $\rho_{xy} > 0$ 蕴含 $\rho_{yx} > 0$.

在情形 (ii) 中，习题 5.3.2 蕴含 $C_x = \{y : \rho_{xy} > 0\}$ 是不可约闭集（若 $y, z \in C_x$ 则 $\rho_{yz} \geqslant \rho_{yx}\rho_{xz} > 0$；若 $\rho_{yw} > 0$ 则 $\rho_{xw} \geqslant \rho_{xy}\rho_{yw} > 0$，故 $w \in C_x$）. 故由定理 5.3.3 知 x 是常返的.

例 5.3.4 启发如下结果：

定理 5.3.5. 分解定理. 设 $R = \{x : \rho_{xx} = 1\}$ 为 Markov 链的常返态集，则 R 可写成 $\cup_i R_i$，其中每个 R_i 都是闭且不可约的.

注释. 这一结果表明，研究常返态时，不失一般性可只考虑单个不可约闭集.

证明. 若 $x \in R$，记 $C_x = \{y : \rho_{xy} > 0\}$. 由定理 5.3.2，$C_x \subset R$，且若 $y \in C_x$ 则 $\rho_{yx} > 0$. 由此易见：要么 $C_x \cap C_y = \emptyset$，要么 $C_x = C_y$. 为证最后一个断言，设 $C_x \cap C_y \neq \emptyset$. 若 $z \in C_x \cap C_y$，则 $\rho_{xy} \geqslant \rho_{xz}\rho_{zy} > 0$，故若 $w \in C_y$ 有 $\rho_{xw} \geqslant \rho_{xy}\rho_{yw} > 0$，可见 $C_x \supset C_y$. 交换 x 与 y 的角色得 $C_y \supset C_x$. 证毕. 若令 R_i 为以某个 C_x 形式出现的集合的列举，即得所需的分解. □

本节余下部分专注于例子，讨论核心问题：如何判断一个状态是常返的还是暂态的？基于定理 5.3.2 的推理偶尔对无限 S 有效.

例题 5.3.6. 分支过程. 若无子女的概率为正，则 $\rho_{k0} > 0$ 而 $\rho_{0k} = 0$ 对 $k \geqslant 1$ 成立，故由定理 5.3.3 知所有状态 $k \geqslant 1$ 都是暂态的. 状态 0 满足 $p(0,0) = 1$，是常返的. 它称为**吸收态**，以反映链一旦进入 0 便永远停留的事实.

若 S 无限且不可约，定理 5.3.2 只表明：要么所有状态都是常返的，要么所有都是暂态的. 因此需判断哪种情形发生.

例题 5.3.7. M/G/1 排队系统. 设 $\mu = \sum_k k a_k$ 为一个服务时间内到达顾客的平均数. 下面证明：若 $\mu > 1$，则链是暂态的（即所有状态都是）；但若 $\mu \leqslant 1$，则是常返的.

对 $\mu > 1$ 的情形，设 $\xi_1, \xi_2, \ldots$ 独立同分布，$P(\xi_m = j) = a_{j+1}$ 对 $j \geqslant -1$，且 $S_n = \xi_1 + \cdots + \xi_n$，则 $X_0 + S_n$ 与 X_n 在时刻 $N = \inf\{n : X_0 + S_n = 0\}$ 之前行为相同. 当 $\mu > 1$ 时，$E\xi_m = \mu - 1 > 0$，故 $S_n \to \infty$ 几乎必然，且 $\inf S_n > -\infty$ 几乎必然. 由最后一个观察，若 x 很大，则 $P_x(N < \infty) < 1$，链是暂态的.

对 $\mu \leqslant 1$ 的情形，由上一段的论证知 $X_{n\wedge N}$ 是上鞅. 设 $T = \inf\{n : X_n \geqslant M\}$. 因 $X_{n\wedge N}$ 是非负上鞅，在时刻 $\tau = T \wedge N$ 使用定理 4.8.4，并注意在 $\{T < N\}$ 上 $X_\tau \geqslant M$，在 $\{N < T\}$ 上 $X_\tau = 0$，得

$$x \geqslant MP_x(T < N)$$

令 $M \to \infty$ 得 $P_x(N < \infty) = 1$，故链是常返的.

注释. 还有另一种方法可看出当 $\mu > 1$ 时 M/G/1 排队系统是暂态的. 若把一个人在服务时间内到达的顾客视为她的子女，则得到一个分支过程. 5.3 节的结果蕴含：当 $\mu \leqslant 1$ 时分支过程以概率 1 灭绝（即排队系统变空），故链是常返的. 当 $\mu > 1$ 时，定理 4.3.12 蕴含 $P_x(T_0 < \infty) = \rho^x$，其中 ρ 是函数 $\varphi(\theta) = \sum_{k=0}^{\infty} a_k\theta^k$ 在 $(0, 1)$ 中的唯一不动点.

下面结果概括了对生灭链和 M/G/1 排队系统使用的技巧.

定理 5.3.8. 设 S 不可约，$\varphi \geqslant 0$ 满足 $E_x\varphi(X_1) \leqslant \varphi(x)$ 对 $x \notin F$（F 为有限集），且 $\varphi(x) \to \infty$ 当 $x \to \infty$（即对任意 $M < \infty$，$\{x : \varphi(x) \leqslant M\}$ 有限），则链是常返的.

证明. 设 $\tau = \inf\{n > 0 : X_n \in F\}$. 由假设知 $Y_n = \varphi(X_{n\wedge\tau})$ 是上鞅. 设 $T_M = \inf\{n > 0 : X_n \in F$ 或$\varphi(X_n) > M\}$. 因 $\{x : \varphi(x) \leqslant M\}$ 有限且链不可约，$T_M < \infty$ 几乎必然. 现在使用定理 4.8.4，可见

$$\varphi(x) \geqslant E_x\varphi(X_{T_M}) \geqslant MP_x(T_M < \tau)$$

因当 $T_M < \tau$ 时 $\varphi(X_{T_M}) \geqslant M$. 令 $M \to \infty$，得 $P_x(\tau < \infty) = 1$ 对所有 $x \notin F$ 成立. 故 $P_y(X_n \in F \text{ i.o.}) = 1$ 对所有 $y \in S$ 成立，而因 F 有限，$P_y(X_n = z \text{ i.o.}) = 1$ 对某个 $z \in F$ 成立. □

例题 5.3.9. $\{0, 1, 2, \ldots\}$ 上的生灭链. 设

$$\begin{aligned} p(i, i+1) &= p_i \\ p(i, i-1) &= q_i \\ p(i, i) &= r_i \end{aligned}$$

其中 $q_0 = 0$. 设 $N = \inf\{n : X_n = 0\}$. 为分析此例，定义函数 φ 使得 $\varphi(X_{N\wedge n})$ 是鞅. 先设 $\varphi(0) = 0$，$\varphi(1) = 1$. 为使鞅性质在 $X_n = k \geqslant 1$ 时成立，必须有

$$\varphi(k) = p_k\varphi(k+1) + r_k\varphi(k) + q_k\varphi(k-1)$$

利用 $r_k = 1 - (p_k + q_k)$，可将最后一个方程改写为

$$q_k(\varphi(k) - \varphi(k-1)) = p_k(\varphi(k+1) - \varphi(k))$$

或

$$\varphi(k+1)-\varphi(k)=\frac{q_k}{p_k}(\varphi(k)-\varphi(k-1))$$

这里及下文设 $p_k,q_k>0$ 对 $k\geqslant 1$ 成立，否则链不是不可约的. 因 $\varphi(1)-\varphi(0)=1$，迭代最后一个结果得

$$\varphi(m+1)-\varphi(m)=\prod_{j=1}^{m}\frac{q_j}{p_j}\quad \text{对}m\geqslant 1$$

$$\varphi(n)=\sum_{m=0}^{n-1}\prod_{j=1}^{m}\frac{q_j}{p_j}\quad \text{对}n\geqslant 1$$

其中约定 $m=0$ 时乘积为 1. 记 $T_c=\inf\{n\geqslant 1:X_n=c\}$.

定理 5.3.10. 若 $a<x<b$ 则

$$P_x(T_a<T_b)=\frac{\varphi(b)-\varphi(x)}{\varphi(b)-\varphi(a)}$$

$$P_x(T_b<T_a)=\frac{\varphi(x)-\varphi(a)}{\varphi(b)-\varphi(a)}$$

证明. 若令 $T=T_a\wedge T_b$，则 $\varphi(X_{n\wedge T})$ 是有界鞅，故由定理 4.8.2 知 $\varphi(x)=E_x\varphi(X_T)$. 因 $X_T\in\{a,b\}$ 几乎必然，

$$\varphi(x)=\varphi(a)P_x(T_a<T_b)+\varphi(b)[1-P_x(T_a<T_b)]$$

解之即得所求公式. □

注释. 答案和证明应使读者想起对称和非对称简单随机游走的结果，它们是本例的特例.

在定理 5.3.10 中令 $a=0$，$b=M$ 得

$$P_x(T_0>T_M)=\varphi(x)/\varphi(M)$$

令 $M\to\infty$ 并注意 $T_M\geqslant M-x$，P_x 几乎必然，得：

定理 5.3.11. 0 是常返的当且仅当 $\varphi(M)\to\infty$ 当 $M\to\infty$，即

$$\varphi(\infty)\equiv\sum_{m=0}^{\infty}\prod_{j=1}^{m}\frac{q_j}{p_j}=\infty$$

若 $\varphi(\infty)<\infty$ 则 $P_x(T_0=\infty)=\varphi(x)/\varphi(\infty)$.

习题

习题 5.3.1. 假设 y 是常返的. 对于 $k \geq 0$，令 $R_k = T_y^k$ 为第 k 次返回 y 的时间；对于 $k \geq 1$，令 $r_k = R_k - R_{k-1}$ 为第 k 次到达间隔时间（interarrival time）. 利用强马尔可夫性质证明，在概率测度 P_y 下，向量序列 $v_k = (r_k, X_{R_{k-1}}, \ldots, X_{R_k-1})(k \geq 1)$ 是独立同分布（i.i.d.）的.

习题 5.3.2. 利用强马尔可夫性质证明 $\rho_{xz} \geq \rho_{xy}\rho_{yz}$.

习题 5.3.3. 证明在埃伦费斯特链（Ehrenfest chain，见例 5.1.3）中，所有状态都是常返的.

习题 5.3.4. 为了探究生灭链（birth and death chains）在常返与非常返之间的边界，假设 $p_j = 1/2 + \epsilon_j$，其中当 $j \to \infty$ 时 $\epsilon_j \sim Cj^{-\alpha}$，且 $q_j = 1 - p_j$. 证明：

(i) 若 $\alpha > 1$，状态 0 是常返的；

(ii) 若 $\alpha < 1$，状态 0 是非常返的（transient）；

(iii) 若 $\alpha = 1$，则当 $C > 1/4$ 时状态 0 是非常返的，当 $C < 1/4$ 时状态 0 是常返的.

习题 5.3.5. 证明：如果我们把定理 5.3.8 中的条件 “$\varphi(x) \to \infty$” 替换为 “$\varphi(x) \to 0$”，并假设对于 $x \in F$ 都有 $\varphi(x) > 0$，那么可以得出该链是非常返的结论.

习题 5.3.6. 设 X_n 是一个生灭链，当 $j \to \infty$ 时满足 $p_j - 1/2 \sim C/j$ 且 $q_j = 1 - p_j$.

(i) 证明：若取 $C < 1/4$，则我们可以选取 $\alpha > 0$ 使得 $\varphi(x) = x^\alpha$ 满足定理 5.3.8 的假设.

(ii) 证明：当 $C > 1/4$ 时，我们可以取 $\alpha < 0$ 并应用练习 5.3.5 的结论.

习题 5.3.7. 若 f 满足 $f(x) \geq \sum_y p(x,y)f(y)$，则称 f 是**超调和的（superharmonic）**，这等价于 $f(X_n)$ 是一个超鞅. 假设 p 是不可约的. 证明：p 是常返的当且仅当每一个非负超调和函数都是常数.

习题 5.3.8. M/M/∞ 排队系统. 考虑一个拥有无限条线路的电话系统. 令 X_n 为在时间 n 正在使用的线路数量，并假设：

$$X_{n+1} = \sum_{m=1}^{X_n} \xi_{n,m} + Y_{n+1}$$

其中 $\xi_{n,m}$ 是独立同分布的，且 $P(\xi_{n,m} = 1) = p$，$P(\xi_{n,m} = 0) = 1 - p$；Y_n 是一个独立的、均值为 λ 的泊松分布随机变量序列. 直观地说：对于每一通通话，我们抛掷一枚正面概率为 p 的硬币，以决定它是否在下一分钟继续.

第 4 节 随机游走的常返性 *

本节中，S_n 总表示随机游走，即 $S_n = X_1 + \cdots + X_n$，其中 $X_1, X_2, \ldots$ 独立同分布. 称 $x \in \mathbb{R}^d$ 为随机游走 S_n 的**常返值**，若对每个 $\epsilon > 0$，$P(\|S_n - x\| < \epsilon \text{ i.o.}) = 1$. 这里 $\|x\| = \sup_i |x_i|$（读者将在引理 5.4.6 的证明中看到选择这一范数的理由）. Hewitt-Savage 0-1 律（定理 2.5.4）蕴含：若最后的概率 < 1，则它为 0. 第一个结果表明，要确定常返值集合，只需检验 $x = 0$. 称 x 为随机游走的**可能值**，若对任意 $\epsilon > 0$，存在 n 使得 $P(\|S_n - x\| < \epsilon) > 0$.

定理 5.4.1. 常返值集合 V 要么是 $\emptyset$，要么是 $\mathbb{R}^d$ 的闭子群. 在第二种情形，$V = U$，即可能值集合.

证明. 设 $V \neq \emptyset$. 显然 V^c 是开集，故 V 是闭集. 为证 V 是群，先证

$$\text{若} x \in U \text{ 且} y \in V \text{ 则} y - x \in V. \tag{*}$$

这一陈述的表述方式使得一旦确立，结果易证. 设

$$p_{\delta,m}(z) = P(\|S_n - z\| \geqslant \delta \text{ 对所有} n \geqslant m)$$

若 $y - x \notin V$，则存在 $\epsilon > 0$ 和 $m \geqslant 1$ 使得 $p_{2\epsilon,m}(y-x) > 0$. 因 $x \in U$，存在 k 使得 $P(\|S_k - x\| < \epsilon) > 0$. 因

$$P(\|S_n - S_k - (y - x)\| \geqslant 2\epsilon \text{ 对所有} n \geqslant k + m) = p_{2\epsilon,m}(y - x)$$

且与 $\{\|S_k - x\| < \epsilon\}$ 独立，故

$$p_{\epsilon,m+k}(y) \geqslant P(\|S_k - x\| < \epsilon) p_{2\epsilon,m}(y - x) > 0$$

与 $y \in V$ 矛盾，故 $y - x \in V$.

为证 $V \neq \emptyset$ 时 V 是群，设 $q, r \in V$，并注意：(i) 在 (*) 中取 $x = y = r$ 得 $0 \in V$；(ii) 取 $x = r, y = 0$ 得 $-r \in V$；(iii) 取 $x = -r, y = q$ 得 $q + r \in V$. 为证此时 $V = U$，注意若 $u \in U$，取 $x = u, y = 0$ 得 $-u \in V$，而因 V 是群，故 $u \in V$. □

若 $V = \emptyset$，称随机游走为**暂态的**；否则称为**常返的**. 在深入处理一般随机游走的技术细节前，先分析 Pólya 于 1921 年考虑的特殊情形. 传说 Pólya 在苏黎世附近的一个公园散步时思考这个问题，注意到他总是遇到同一对年轻夫妇. 历史没有记录这对年轻夫妇的想法.

例题 5.4.2. $\mathbb{Z}^d$ **上的简单随机游走.**

$$P(X_i = e_j) = P(X_i = -e_j) = 1/2d$$

对每个 d 个单位向量 e_j. 为分析此情形，先给出一个对任意随机游走都有效的结果. 设 $\tau_0 = 0$, $\tau_n = \inf\{m > \tau_{n-1} : S_m = 0\}$ 为第 n 次返回 0 的时刻. 由强 Markov 性质，

$$P(\tau_n < \infty) = P(\tau_1 < \infty)^n$$

这一事实易导出：

定理 5.4.3. 对任意随机游走，下列等价：

(i) $P(\tau_1 < \infty) = 1$;

(ii) $P(S_m = 0 \text{ i.o.}) = 1$;

(iii) $\sum_{m=0}^{\infty} P(S_m = 0) = \infty$.

证明. 若 $P(\tau_1 < \infty) = 1$，则 $P(\tau_n < \infty) = 1$ 对所有 n 成立，且 $P(S_m = 0 \text{ i.o.}) = 1$. 设

$$V = \sum_{m=0}^{\infty} 1_{(S_m=0)} = \sum_{n=0}^{\infty} 1_{(\tau_n<\infty)}$$

为访问 0 的次数（计入时刻 0 的访问）. 取期望并利用 Fubini 定理将期望移入求和：

$$EV = \sum_{m=0}^{\infty} P(S_m = 0) = \sum_{n=0}^{\infty} P(\tau_n < \infty) = \sum_{n=0}^{\infty} P(\tau_1 < \infty)^n = \frac{1}{1 - P(\tau_1 < \infty)}$$

第二个等式表明 (ii) 蕴含 (iii)，而结合最后两个等式表明：若 (i) 不成立，则 (iii) 不成立（即 (iii) 蕴含 (i)）. □

定理 5.4.4. 简单随机游走在 $d \leqslant 2$ 时是常返的，在 $d \geqslant 3$ 时是暂态的.

借用 Kakutani（U.C.L.A. 学术报告）的玩笑：“醉汉终将找到回家的路，但醉鸟可能永远迷失.”

证明. 记 $\rho_d(m) = P(S_m = 0)$. $\rho_d(m)$ 在 m 为奇数时为 0. 由定理 3.1.3, $\rho_1(2n) \sim (\pi n)^{-1/2}$ 当 $n \to \infty$. 这与定理 5.4.3 给出一维的结果. 下面证明

二维简单随机游走是常返的. 注意要使 $S_{2n} = 0$，必须有某个 $0 \leqslant m \leqslant n$ 使得 n 步中 m 步向上、m

步向下、$n-m$ 步向左、$n-m$ 步向右，故

$$\begin{aligned}\rho_2(2n) &= 4^{-2n}\sum_{m=0}^{n}\frac{(2n)!}{m!m!(n-m)!(n-m)!}\\ &= 4^{-2n}\binom{2n}{n}\sum_{m=0}^{n}\binom{n}{m}\binom{n}{n-m}\\ &= 4^{-2n}\binom{2n}{n}^2 = \rho_1(2n)^2\end{aligned}$$

倒数第二个等式的解释如下：考虑从 n 个男生和 n 个女生的班级中选 n 个学生，对某个 $0\leqslant m\leqslant n$ 必须选 m 个男生和 $n-m$ 个女生. 利用渐近公式 $\rho_1(2n)\sim(\pi n)^{-1/2}$，得 $\rho_2(2n)\sim(\pi n)^{-1}$. 因 $\sum n^{-1}=\infty$，结果由定理 5.4.3 得出. □

注释. 关于 $\rho_2(2n)=\rho_1(2n)^2$ 的直接证明：若 T_n^1 和 T_n^2 是独立的一维随机游走，则 T_n 以相等概率从 x 跳到 $x+(1,1)$，$x+(1,-1)$，$x+(-1,1)$ 和 $x+(-1,-1)$，故将 T_n 旋转 45 度并除以 $\sqrt{2}$ 即得 S_n.

三维简单随机游走是暂态的. 直观上，这成立是因为 $2n$ 步后返回 0 的概率 $\sim cn^{-3/2}$，而这是可求和的. 下面不精确计算概率，只给出正确数量级的上界. 因在每个方向 $\pm e_i$ 的步数必须相等（$i=1,2,3$），

$$\begin{aligned}\rho_3(2n) &= 6^{-2n}\sum_{j,k}\frac{(2n)!}{(j!k!(n-j-k)!)^2}\\ &= 2^{-2n}\binom{2n}{n}\sum_{j,k}\left(\frac{3^{-n}n!}{j!k!(n-j-k)!}\right)^2\\ &\leqslant 2^{-2n}\binom{2n}{n}\max_{j,k}\frac{3^{-n}n!}{j!k!(n-j-k)!}\end{aligned}$$

其中最后一个不等式利用了：若 $a_{j,k}\geqslant 0$ 且和为 1，则 $\sum_{j,k}a_{j,k}^2\leqslant \max_{j,k}a_{j,k}$. 最后一步是证

$$\max_{j,k}\frac{3^{-n}n!}{j!k!(n-j-k)!}\leqslant Cn^{-1}$$

为此，注意：

(a) 若 j,k 或 $n-j-k$ 中任一个 $<[n/3]$，增大最小的数并减小最大的数会减小分母（因 $x(1-x)$ 在 $1/2$ 处最大），故最大值出现在三个数尽可能接近 $n/3$ 时；

(b) Stirling 公式蕴含

$$\frac{n!}{j!k!(n-j-k)!}\sim\frac{n^n}{j^jk^k(n-j-k)^{n-j-k}}\cdot\sqrt{\frac{n}{jk(n-j-k)}}\cdot\frac{1}{2\pi}$$

取 j 和 k 在 $n/3$ 的 1 范围内，右端第一项 $\leqslant C3^n$，即得所需结果.

高维简单随机游走是暂态的 ($d>3$). 设 $T_n=(S_n^1,S_n^2,S_n^3)$，$N(0)=0$，$N(n)=\inf\{m>N(n-1):T_m\neq T_{N(n-1)}\}$. 易见 $T_{N(n)}$ 是三维简单随机游走. 因 $T_{N(n)}$ 以概率 0 无穷多次返回 0，而前三个坐标在 $N(n)$ 之间保持常数，故 S_n 是暂态的.

本节余下部分证明关于随机游走的下列事实：

- 若 $S_n/n\to 0$ 依概率，则 S_n 在 $d=1$ 时是常返的.
- 若 $S_n/n^{1/2}\Rightarrow$ 非退化正态分布，则 S_n 在 $d=2$ 时是常返的.
- 若随机游走 “真正三维”，则 S_n 在 $d\geqslant 3$ 时是暂态的.

为证最后一个结果，下面给出常返性的一个充要条件.

首先推广定理 5.4.3.

引理 5.4.5. 若 $\sum_{n=1}^{\infty}P(\|S_n\|<\epsilon)<\infty$ 则 $P(\|S_n\|<\epsilon\text{ i.o.})=0$. 若 $\sum_{n=1}^{\infty}P(\|S_n\|<\epsilon)=\infty$ 则 $P(\|S_n\|<2\epsilon\text{ i.o.})=1$.

证明. 第一个结论由 Borel-Cantelli 引理得出. 为证第二个，设 $F=\{\|S_n\|<\epsilon\text{ i.o.}\}^c$. 按 $\|S_n\|<\epsilon$ 的最后时刻分解：

$$P(F)=\sum_{m=0}^{\infty}P(\|S_m\|<\epsilon,\|S_n\|\geqslant\epsilon\text{ 对所有}n\geqslant m+1)$$

$$\geqslant\sum_{m=0}^{\infty}P(\|S_m\|<\epsilon,\|S_n-S_m\|\geqslant 2\epsilon\text{ 对所有}n\geqslant m+1)=\sum_{m=0}^{\infty}P(\|S_m\|<\epsilon)\rho_{2\epsilon,1}$$

其中 $\rho_{\delta,k}=P(\|S_n\|\geqslant\delta\text{ 对所有}n\geqslant k)$. 因 $P(F)\leqslant 1$ 且 $\sum_{m=0}^{\infty}P(\|S_m\|<\epsilon)=\infty$，故 $\rho_{2\epsilon,1}=0$. 为将此结论推广到 $\rho_{2\epsilon,k}$ $(k\geqslant 2)$，设

$$A_m=\{\|S_m\|<\epsilon,\|S_n\|\geqslant\epsilon\text{ 对所有}n\geqslant m+k\}$$

因任一 ω 至多在 k 个 A_m 中，重复上述论证得

$$k\geqslant\sum_{m=0}^{\infty}P(A_m)\geqslant\sum_{m=0}^{\infty}P(\|S_m\|<\epsilon)\rho_{2\epsilon,k}$$

故 $\rho_{2\epsilon,k}=P(\|S_n\|\geqslant 2\epsilon\text{ 对所有}j\geqslant k)=0$，而因 k 任意，即得所需结论. □

下面证明引理 5.4.5 中和的收敛或发散与 ϵ 无关. 前面的证明适用于任意范数，但对下一步，需要 $\|x\|=\sup_i|x_i|$.

引理 5.4.6. 设 m 为整数 $\geqslant 2$.

$$\sum_{n=0}^{\infty}P(\|S_n\|<m\epsilon)\leqslant(2m)^d\sum_{n=0}^{\infty}P(\|S_n\|<\epsilon)$$

证明. 先注意

$$\sum_{n=0}^{\infty} P(\|S_n\| < m\epsilon) \leqslant \sum_{n=0}^{\infty} \sum_{k} P(S_n \in k\epsilon + [0,\epsilon)^d)$$

其中内层求和对 $k \in \{-m, \ldots, m-1\}^d$. 设

$$T_k = \inf\{\ell \geqslant 0 : S_\ell \in k\epsilon + [0,\epsilon)^d\}$$

则按 T_k 的值分解并利用 Fubini 定理得

$$\sum_{n=0}^{\infty} P(S_n \in k\epsilon + [0,\epsilon)^d) = \sum_{n=0}^{\infty} \sum_{\ell=0}^{n} P(S_n \in k\epsilon + [0,\epsilon)^d, T_k = \ell) \leqslant \sum_{\ell=0}^{\infty} \sum_{n=\ell}^{\infty} P(\|S_n - S_\ell\| < \epsilon, T_k = \ell)$$

因 $\{T_k = \ell\}$ 与 $\{\|S_n - S_\ell\| < \epsilon\}$ 独立，最后的和

$$= \sum_{m=0}^{\infty} P(T_k = m) \sum_{j=0}^{\infty} P(\|S_j\| < \epsilon) \leqslant \sum_{j=0}^{\infty} P(\|S_j\| < \epsilon)$$

因在 $\{-m, \ldots, m-1\}^d$ 中有 $(2m)^d$ 个 k 值. 证毕. □

结合引理 5.4.5 和 5.4.6 得：

定理 5.4.7. $\sum_n P(\|S_n\| < \epsilon)$ 对单个 $\epsilon > 0$ 的收敛（或发散）足以判定暂态性（或常返性）.

在 $d = 1$ 时，若 $EX_i = \mu \neq 0$，则强大数定律蕴含 $S_n/n \to \mu$，故 $|S_n| \to \infty$ 而 S_n 是暂态的. 作为逆命题，有

定理 5.4.8. Chung-Fuchs 定理. 设 $d = 1$. 若弱大数定律以 $S_n/n \to 0$ 依概率的形式成立，则 S_n 是常返的.

证明. 对 $x > 0$，记 $u_n(x) = P(|S_n| < x)$. 引理 5.4.6 蕴含

$$\sum_{n=0}^{\infty} u_n(1) \geqslant \frac{1}{2m} \sum_{n=0}^{\infty} u_n(m) \geqslant \frac{1}{2m} \sum_{n=0}^{Am} u_n(n/A)$$

对任意 $A < \infty$，因 $u_n(x) \geqslant 0$ 且关于 x 递增. 由假设 $u_n(n/A) \to 1$，故令 $m \to \infty$ 并注意右端是前 Am 项平均的 $A/2$ 倍，

$$\sum_{n=0}^{\infty} u_n(1) \geqslant A/2$$

因 A 任意，和必为 ∞，所需结论由定理 5.4.7 得出. □

定理 5.4.9. 若 S_n 是 $\mathbb{R}^2$ 中的随机游走且 $S_n/n^{1/2} \Rightarrow$ 非退化正态分布，则 S_n 是常返的.

注释. 若极限是退化的, 结论亦真, 但此时随机游走本质上是一维 (或零维) 的, 结果由 Chung-Fuchs 定理得出.

证明. 记 $u(n,m)=P(\|S_n\|<m)$. 由引理 5.4.6 知

$$\sum_{n=0}^{\infty}u(n,1)\geqslant(4m^2)^{-1}\sum_{n=0}^{\infty}u(n,m)$$

若 $m/\sqrt{n}\to c$，则

$$u(n,m)\to\int_{[-c,c]^2}n(x)\,\mathrm{d}x$$

其中 $n(x)$ 是极限正态分布的密度. 以 $\rho(c)$ 记右端并令 $n=[\theta m^2]$，则 $u([\theta m^2],m)\to\rho(\theta^{-1/2})$. 将

$$m^{-2}\sum_{n=0}^{\infty}u(n,m)=\int_0^{\infty}u([\theta m^2],m)\,\mathrm{d}\theta$$

令 $m\to\infty$ 并用 Fatou 引理，得

$$\liminf_{m\to\infty}(4m^2)^{-1}\sum_{n=0}^{\infty}u(n,m)\geqslant 4^{-1}\int_0^{\infty}\rho(\theta^{-1/2})\,\mathrm{d}\theta$$

因正态密度在 0 处为正且连续,

$$\rho(c)=\int_{[-c,c]^2}n(x)\,\mathrm{d}x\sim n(0)(2c)^2\quad \text{当}c\to 0$$

故 $\rho(\theta^{-1/2})\sim 4n(0)/\theta$ 当 $\theta\to\infty$，积分发散，回溯证明中第一个不等式得 $\sum_{n=0}^{\infty}u(n,1)=\infty$. 证毕. □

下面给出所承诺的常返性充要条件. 这里 $\varphi=E\exp(it\cdot X_j)$ 是随机游走一步的特征函数.

定理 5.4.10. 设 $\delta>0$. S_n 是常返的当且仅当

$$\int_{(-\delta,\delta)^d}\mathrm{Re}\,\frac{1}{1-\varphi(y)}\,\mathrm{d}y=\infty$$

下面证明一个较弱的结果:

定理 5.4.11. 设 $\delta>0$. S_n 是常返的当且仅当

$$\sup_{r<1}\int_{(-\delta,\delta)^d}\mathrm{Re}\,\frac{1}{1-r\varphi(y)}\,\mathrm{d}y=\infty$$

注释. 从第二个结果得到第一个结果所需的一半工作是平凡的.

$$0 \leqslant \operatorname{Re} \frac{1}{1-r\varphi(y)} \to \operatorname{Re} \frac{1}{1-\varphi(y)} \quad \text{当} r \to 1$$

故 Fatou 引理表明：若积分为无穷，则随机游走是常返的. 另一方向相当困难：第二个结果见于 Chung 和 Fuchs (1951)，但第一个结果的证明要等到 Ornstein (1969) 和 Stone (1969) 独立解决此问题. 他们的证明用一个技巧归结到增量有密度的情形，再用第二个技巧处理该情形，故这里不给出细节. 读者可查阅所引文献或 Port 和 Stone (1969)，其中对 Abel 群上的随机游走证明了此结果.

证明. 证明需要以下要素.

引理 5.4.12. Parseval 关系. 设 μ 和 ν 是 $\mathbb{R}^d$ 上特征函数为 φ 和 ψ 的概率测度.

$$\int \psi(t)\,\mu(\mathrm{d}t) = \int \varphi(x)\,\nu(\mathrm{d}x)$$

证明. 因 $e^{it\cdot x}$ 有界，由 Fubini 定理得

$$\int \psi(t)\mu(\mathrm{d}t) = \int\int e^{it\cdot x}\nu(\mathrm{d}x)\mu(\mathrm{d}t) = \int\int e^{it\cdot x}\mu(\mathrm{d}t)\nu(\mathrm{d}x) = \int \varphi(x)\nu(\mathrm{d}x)$$

□

再引入一点微积分.

引理 5.4.13. 若 $|x| \leqslant \pi/3$ 则 $1-\cos x \geqslant x^2/4$.

证明. 只需对 $x>0$ 证之. 若 $z \leqslant \pi/3$，则 $\cos z \geqslant 1/2$，

$$\sin y = \int_0^y \cos z\,\mathrm{d}z \geqslant \frac{y}{2}$$

$$1-\cos x = \int_0^x \sin y\,\mathrm{d}y \geqslant \int_0^x \frac{y}{2}\,\mathrm{d}y = \frac{x^2}{4}$$

即得所需结果. □

由例 3.3.7，密度 $\frac{\delta-|x|}{\delta^2}$（当 $|x| \leqslant \delta$），0（其他）的特征函数为 $2(1-\cos\delta t)/(\delta t)^2$. 以 μ_n 表示 S_n 的分布. 先用引理 5.4.13（注意 $\pi/3 \geqslant 1$），再用引理 5.4.12，有

$$P(\|S_n\| < 1/\delta) \leqslant 4^d \int \prod_{i=1}^d \frac{1-\cos(\delta t_i)}{(\delta t_i)^2}\mu_n(\mathrm{d}t) = 2^d \int_{(-\delta,\delta)^d} \prod_{i=1}^d \frac{\delta-|x_i|}{\delta^2}\varphi^n(x)\,\mathrm{d}x$$

下一步是从 0 到 ∞ 求和. 为能交换求和与积分，先乘以 r^n（$r<1$）.

$$\sum_{n=0}^{\infty} r^n P(\|S_n\|<1/\delta) \leqslant 2^d \int_{(-\delta,\delta)^d} \prod_{i=1}^{d} \frac{\delta-|x_i|}{\delta^2} \frac{1}{1-r\varphi(x)}\,\mathrm{d}x$$

对称性要求右端的积分为实数，故可取实部而不改变其值. 令 $r\uparrow 1$ 并用 $(\delta-|x|)/\delta\leqslant 1$，

$$\sum_{n=0}^{\infty} P(\|S_n\|<1/\delta) \leqslant \left(\frac{2}{\delta}\right)^d \sup_{r<1} \int_{(-\delta,\delta)^d} \mathrm{Re}\,\frac{1}{1-r\varphi(x)}\,\mathrm{d}x$$

再用定理 5.4.7 即得定理 5.4.11 的一半.

为证另一方向，例 3.3.15 表明：密度 $(1-\cos(x/\delta))/\pi x^2/\delta$ 的特征函数为 $1-|\delta t|$（当 $|t|\leqslant 1/\delta$），0（其他）. 由 $1\geqslant\prod_{i=1}^{d}(1-|\delta x_i|)$，再用引理 5.4.12 得

$$P(\|S_n\|<1/\delta) \geqslant \int_{(-1/\delta,1/\delta)^d} \prod_{i=1}^{d}(1-|\delta x_i|)\mu_n(\mathrm{d}x) = \int \prod_{i=1}^{d} \frac{1-\cos(t_i/\delta)}{\pi t_i^2/\delta}\varphi^n(t)\,\mathrm{d}t$$

乘以 r^n 并求和，得

$$\sum_{n=0}^{\infty} r^n P(\|S_n\|<1/\delta) \geqslant \int \prod_{i=1}^{d} \frac{1-\cos(t_i/\delta)}{\pi t_i^2/\delta} \frac{1}{1-r\varphi(t)}\,\mathrm{d}t$$

最后的积分是实的，故若只积分被积函数的实部，其值不变. 这样做并应用引理 5.4.13，得

$$\sum_{n=0}^{\infty} r^n P(\|S_n\|<1/\delta) \geqslant (4\pi\delta)^{-d} \int_{(-\delta,\delta)^d} \mathrm{Re}\,\frac{1}{1-r\varphi(t)}\,\mathrm{d}t$$

令 $r\uparrow 1$ 并用定理 5.4.7，即完成定理 5.4.11 的证明. □

下面考虑若干例子. 对 $d=1$ 和 $d=2$，目的是使读者相信定理 5.4.8 和 5.4.9 中的条件接近最优.

$d=1$**.** 考虑对称稳定律，其特征函数 $\varphi(t)=\exp(-|t|^\alpha)$. 为避免使用尚未证明的事实，下面从定理 5.4.11 得出结论. 在此情形下使用该形式的判别准则并不困难，因为

$$1-r\varphi(t)\downarrow 1-\exp(-|t|^\alpha) \quad \text{当} r\uparrow 1$$
$$1-\exp(-|t|^\alpha)\sim|t|^\alpha \quad \text{当} t\to 0$$

由此可知，相应随机游走在 $\alpha<1$ 时是暂态的，在 $\alpha\geqslant 1$ 时是常返的.

$\alpha>1$ 的情形由定理 5.4.8 覆盖，因这些随机游走均值为 0. $\alpha=1$ 的结果是新的，因为 Cauchy

分布不满足 $S_n/n \to 0$ 依概率. $\alpha < 1$ 的随机游走是有趣的，因为习题 5.4.1 蕴含

$$-\infty = \liminf S_n < \limsup S_n = \infty$$

但 $P(|S_n| < M \text{ i.o.}) = 0$ 对任意 $M < \infty$.

注释. 稳定律的例子在某一方面具有误导性. Shepp (1964) 证明了常返随机游走可能有任意大的尾部. 确切地说，给定函数 $\epsilon(x) \downarrow 0$ 当 $x \uparrow \infty$，存在常返随机游走使得 $P(|X_1| \geqslant x) \geqslant \epsilon(x)$ 对大的 x 成立.

$d = 2$. 设 $\alpha < 2$，令 $\varphi(t) = \exp(-|t|^\alpha)$，其中 $|t| = (t_1^2 + t_2^2)^{1/2}$. φ 是随机向量 (X_1, X_2) 的特征函数，具有两个好性质：

(i) (X_1, X_2) 的分布在旋转下不变；

(ii) X_1 和 X_2 有指标 α 的对称稳定律.

又，$1 - r\varphi(t) \downarrow 1 - \exp(-|t|^\alpha)$ 当 $r \uparrow 1$，且 $1 - \exp(-|t|^\alpha) \sim |t|^\alpha$ 当 $t \to 0$. 换到极坐标并注意

$$2\pi \int_0^\delta \mathrm{d}x\, x\, x^{-\alpha} < \infty \quad \text{当} 1 - \alpha > -1$$

表明特征函数 $\exp(-|t|^\alpha)$（$\alpha < 2$）的随机游走是暂态的. 当 $p < \alpha$ 时，由习题 3.8.4 知 $E|X_1|^p < \infty$，故这些例子表明定理 5.4.9 相当尖锐.

$d \geqslant 3$. 积分 $\int_0^\delta \mathrm{d}x\, x^{d-1}x^{-2} < \infty$，故若随机游走在 $d \geqslant 3$ 时是常返的，其特征函数必须比 t^2 更快地趋于 1. 习题 3.3.16 表明（在一维情形）若 $\varphi(r) = 1 + o(r^2)$ 则 $\varphi(r) \equiv 1$. 考虑 $\varphi(r\theta)$（其中 r 为实数，θ 为固定向量），最后一个结论易推广到 $\mathbb{R}^d$（$d > 1$），并提示：一旦排除停留在过 0 的平面上的随机游走，没有三维随机游走是常返的.

$\mathbb{R}^3$ 中的随机游走称为**真正三维**的，若 X_1 的分布满足 $P(X_1 \cdot \theta \neq 0) > 0$ 对所有 $\theta \neq 0$.

定理 5.4.14. 没有真正三维的随机游走是常返的.

证明. 下面从定理 5.4.11 推出此结果. 先做一些演算. 若 z 是复数，$1 - z$ 的共轭是 $1 - \bar{z}$，故

$$\frac{1}{1-z} = \frac{1-\bar{z}}{|1-z|^2} \quad \text{且} \quad \mathrm{Re}\,\frac{1}{1-z} = \frac{\mathrm{Re}(1-z)}{|1-z|^2}$$

若 $z = a + bi$ 且 $a \leqslant 1$，则用前一个公式并从分母舍去 b^2 得

$$\mathrm{Re}\,\frac{1}{1-z} = \frac{1-a}{(1-a)^2+b^2} \leqslant \frac{1}{1-a}$$

取 $z = r\varphi(t)$ 并设对第二个不等式有 $0 \leqslant \mathrm{Re}\,\varphi(t) \leqslant 1$，有

$$\text{(a)} \quad \mathrm{Re}\,\frac{1}{1-r\varphi(t)} \leqslant \frac{1}{\mathrm{Re}(1-r\varphi(t))} \leqslant \frac{1}{\mathrm{Re}(1-\varphi(t))}$$

最后的计算表明只需估计

$$\mathrm{Re}(1-\varphi(t))=\int\{1-\cos(x\cdot t)\}\mu(\mathrm{d}x)\geqslant\int_{|x\cdot t|<\pi/3}\frac{|x\cdot t|^2}{4}\mu(\mathrm{d}x)$$

其中用了引理 5.4.13. 写 $t=\rho\theta$，其中 $\theta\in S=\{x:|x|=1\}$，得

$$\text{(b)}\quad \mathrm{Re}(1-\varphi(\rho\theta))\geqslant\frac{\rho^2}{4}\int_{|x\cdot\theta|<\pi/3\rho}|x\cdot\theta|^2\mu(\mathrm{d}x)$$

Fatou 引理蕴含：若令 $\rho\to0$ 且 $\theta(\rho)\to\theta$，则

$$\text{(c)}\quad \liminf_{\rho\to0}\int_{|x\cdot\theta(\rho)|<\pi/3\rho}|x\cdot\theta(\rho)|^2\mu(\mathrm{d}x)\geqslant\int|x\cdot\theta|^2\mu(\mathrm{d}x)>0$$

这蕴含：对 $\rho<\rho_0$

$$\text{(d)}\quad \inf_{\theta\in S}\int_{|x\cdot\theta|<\pi/3\rho}|x\cdot\theta|^2\mu(\mathrm{d}x)=C>0$$

为得最后一个结论，注意若它不成立，则对 $\rho=1/n$，存在 θ_n 使得

$$\int_{|x\cdot\theta_n|<n\pi/3}|x\cdot\theta_n|^2\mu(\mathrm{d}x)\leqslant1/n$$

所有 θ_n 位于紧集 S 中，故若取收敛子列则与 (c) 矛盾. 结合 (b) 和 (d) 得

$$\mathrm{Re}(1-\varphi(\rho\theta))\geqslant C\rho^2/4$$

用此结果和 (a)，再换到极坐标，可见若 δ 小（使 $\mathrm{Re}\,\varphi(y)\geqslant0$ 在 $(-\delta,\delta)^d$ 上）

$$\begin{aligned}\int_{(-\delta,\delta)^d}\mathrm{Re}\,\frac{1}{1-r\varphi(y)}\,\mathrm{d}y&\leqslant\int_0^{\delta\sqrt{d}}\mathrm{d}\rho\,\rho^{d-1}\int\mathrm{d}\theta\,\frac{1}{\mathrm{Re}(1-\varphi(\rho\theta))}\\&\leqslant C'\int_0^1\mathrm{d}\rho\,\rho^{d-3}<\infty\end{aligned}$$

当 $d>2$ 时，故所需结果由定理 5.4.11 得出. □

注释. 当考虑 $\mathbb{Z}^d$ 上的随机游走时，分析变得简单得多. 习题 3.3.2 给出的反演公式蕴含

$$P(S_n=0)=(2\pi)^{-d}\int_{(-\pi,\pi)^d}\varphi^n(t)\,\mathrm{d}t$$

乘以 r^n 并求和，得

$$\sum_{n=0}^{\infty} r^n P(S_n = 0) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} \frac{1}{1 - r\varphi(t)}\, \mathrm{d}t$$

对 $d = 3$ 的简单随机游走，$\varphi(t) = \frac{1}{3}\sum_{j=1}^{3} \cos t_j$ 是实的.

$$\frac{1}{1 - r\varphi(t)} \uparrow \frac{1}{1 - \varphi(t)} \quad \text{当} \varphi(t) > 0$$
$$0 \leqslant \frac{1}{1 - r\varphi(t)} \leqslant 1 \quad \text{当} \varphi(t) \leqslant 0$$

故由单调收敛定理和有界收敛定理得

$$\sum_{n=0}^{\infty} P(S_n = 0) = (2\pi)^{-3} \int_{(-\pi,\pi)^3} \left(1 - \frac{1}{3}\sum_{i=1}^{3} \cos x_i\right)^{-1} \mathrm{d}x$$

此积分由 Watson 于 1939 年首次用椭圆积分算出，可在表中找到. Glasser 和 Zucker (1977) 证明它等于

$$(\sqrt{6}/32\pi^3)\Gamma(1/24)\Gamma(5/24)\Gamma(7/24)\Gamma(11/24) = 1.51638606\ldots$$

故由 (5.3.1)，若 $\beta_3 = P_0(T_0 < \infty)$ 则

$$\sum_{n=0}^{\infty} P_0(S_n = 0) = \frac{1}{1 - \beta_3}$$

从而 $\beta_3 = 0.34053733\ldots$ 关于 $4 \leqslant d \leqslant 9$ 的数值结果，见 Kondo 和 Hara (1987).

习题

习题 5.4.1. 对于 $\mathbb{R}$ 上的随机游走，只有以下四种可能性，且其中一种发生的概率为 1：

(i) 对所有的 n，都有 $S_n = 0$.

(ii) $S_n \to \infty$.

(iii) $S_n \to -\infty$.

(iv) $-\infty = \liminf S_n < \limsup S_n = \infty$.

习题 5.4.2. 阶梯变量 (Ladder variables). 令 $\alpha(\omega) = \inf\{n : \omega_1 + \cdots + \omega_n > 0\}$，其中规定 $\inf\emptyset = \infty$，并设 $\alpha(\Delta) = \infty$. 令 $\alpha_0 = 0$ 且对于 $k \geq 1$，令

$$\alpha_k(\omega) = \alpha_{k-1}(\omega) + \alpha(\theta^{\alpha_{k-1}}\omega)$$

在时间 α_k，随机游走处于历史最高值.

(i) 若 $P(\alpha<\infty)<1$，则 $P(\sup S_n<\infty)=1$.

(ii) 若 $P(\alpha<\infty)=1$，则 $P(\sup S_n=\infty)=1$.

习题 5.4.3. 假设 f 在 $\mathbb{R}^d$ 上是超调和的 (superharmonic)，即：

$$f(x)\geq\frac{1}{|B(x,r)|}\int_{B(x,r)}f(y)dy$$

令 $\xi_1,\xi_2,\dots$ 独立同分布，且在单位球 $B(0,1)$ 上服从均匀分布. 定义 S_n 为 $S_n=S_{n-1}+\xi_n$ $(n\geq 1)$，且 $S_0=x$.

(i) 证明 $X_n=f(S_n)$ 是一个超鞅 (supermartingale).

(ii) 由此得出结论：在 $d\leq 2$ 时，非负超调和函数必为常数. 例子 $f(x)=|x|^{2-d}$ 表明当 $d>2$ 时该结论不成立.

习题 5.4.4. 假设 h 在 $\mathbb{R}^d$ 上是调和的 (harmonic)，即：

$$h(x)=\frac{1}{|B(x,r)|}\int_{B(x,r)}h(y)dy$$

令 $\xi_1,\xi_2,\dots$ 独立同分布，且在 $B(0,1)$ 上服从均匀分布. 定义 S_n 为 $S_n=S_{n-1}+\xi_n$ $(n\geq 1)$，且 $S_0=x$.

(i) 证明 $X_n=h(S_n)$ 是一个鞅 (martingale).

(ii) 由此得出结论：在任意维度下，有界调和函数必为常数. Hewitt-Savage 0-1 律在此处会很有用.

第 5 节　平稳测度

称测度 μ 为**平稳测度**，若

$$\sum_x \mu(x)p(x,y) = \mu(y).$$

最后一个等式表明 $P_\mu(X_1 = y) = \mu(y)$. 利用 Markov 性质和归纳法，可得对所有 $n \geqslant 1$ 有 $P_\mu(X_n = y) = \mu(y)$. 若 μ 是概率测度，则称其为**平稳分布**，它代表链的一个可能平衡态. 即，若 X_0 服从分布 μ，则对所有 $n \geqslant 1$，X_n 也服从分布 μ. 稍作想象，这一解释也可应用于 μ 为无限测度的情形.（当总质量有限时，可除以 $\mu(S)$ 得到平稳分布.）在深入理论之前，先看一些例子.

例题 5.5.1. 随机游走. $S = \mathbb{Z}^d$. $p(x,y) = f(y-x)$，其中 $f(z) \geqslant 0$ 且 $\sum f(z) = 1$. 在此情形下，$\mu(x) \equiv 1$ 是平稳测度，因为

$$\sum_x p(x,y) = \sum_x f(y-x) = 1.$$

满足 $\sum_x p(x,y) = 1$ 的转移概率称为**双随机的**. 这显然构成了 $\mu(x) \equiv 1$ 为平稳测度的充要条件.

例题 5.5.2. 非对称简单随机游走. $S = \mathbb{Z}$.

$$\begin{aligned} p(x, x+1) &= p, \\ p(x, x-1) &= q = 1-p. \end{aligned}$$

由上一例，$\mu(x) \equiv 1$ 是平稳测度. 当 $p \neq q$ 时，$\mu(x) = (p/q)^x$ 是另一个平稳测度. 为验证这一点，注意

$$\begin{aligned} \sum_x \mu(x)p(x,y) &= \mu(y+1)p(y+1,y) + \mu(y-1)p(y-1,y) \\ &= (p/q)^{y+1}q + (p/q)^{y-1}p = (p/q)^y[p+q] = (p/q)^y. \end{aligned}$$

为简化接下来两个例子中的计算，引入一个比平稳测度更强的概念. 称 μ 满足**细致平衡条件**，若

$$\mu(x)p(x,y) = \mu(y)p(y,x). \tag{5.5.1}$$

对 x 求和得

$$\sum_x \mu(x)p(x,y) = \mu(y).$$

(5.5.1) 断言从 x 跳到 y 的质量恰好等于从 y 跳到 x 的质量.

满足 (5.5.1) 的测度 μ 称为**可逆测度**. 定理 5.5.5 将解释 “可逆” 一词的含义.

例题 5.5.3. Ehrenfest 链. $S=\{0,1,\ldots,r\}$.

$$\begin{aligned} p(k,k+1)&=(r-k)/r,\\ p(k,k-1)&=k/r. \end{aligned}$$

在此情形下，$\mu(x)=2^{-r}\binom{r}{x}$ 是平稳分布. 无需纸笔即可验证这一点：只需注意到 μ 对应抛 r 枚硬币决定每个球放入哪个罐子，而链的转移对应随机选取一枚硬币翻转. 或者，可验证

$$\begin{aligned} \mu(k+1)p(k+1,k)&=2^{-r}\frac{r!}{(k+1)!(r-k-1)!}\cdot\frac{k+1}{r}\\ &=\frac{(r-1)!}{k!(r-k-1)!}=2^{-r}\frac{r!}{k!(r-k)!}\cdot\frac{r-k}{r}\\ &=\mu(k)p(k,k+1). \end{aligned}$$

例题 5.5.4. 生灭链. $S=\{0,1,2,\ldots\}$

$$\begin{aligned} p(x,x+1)&=p_x,\\ p(x,x)&=r_x,\\ p(x,x-1)&=q_x, \end{aligned}$$

其中 $q_0=0$ 且 $p(i,j)=0$ 对其他情形. 在此情形下，存在测度

$$\mu(x)=\prod_{k=1}^{x}\frac{p_{k-1}}{q_k}$$

满足细致平衡条件：

$$\mu(x)p(x,x+1)=p_x\prod_{k=1}^{x}\frac{p_{k-1}}{q_k}=\mu(x+1)p(x+1,x).$$

下面结果解释了“可逆”一词的由来.

定理 5.5.5. 设 μ 为平稳测度且假设 X_0 服从“分布”μ. 则 $Y_m=X_{n-m}$（$0\leqslant m\leqslant n$）是一个 Markov 链，具有初始测度 μ 和转移概率

$$q(x,y)=\mu(y)p(y,x)/\mu(x).$$

q 称为**对偶转移概率**. 若 μ 是可逆测度，则 $q=p$.

证明.

$$\begin{aligned} P(Y_{m+1}=y \mid Y_m=x) &= P(X_{n-m-1}=y \mid X_{n-m}=x) \\ &= \frac{P(X_{n-m-1}=y, X_{n-m}=x)}{P(X_{n-m}=x)} \\ &= \frac{P(X_{n-m}=x \mid X_{n-m-1}=y)P(X_{n-m-1}=y)}{P(X_{n-m}=x)} \\ &= \frac{\mu(y)p(y,x)}{\mu(x)}, \end{aligned}$$

这就证明了结论. □

虽然许多链具有可逆测度，但大多数链没有. 若存在一对状态满足 $p(x,y)>0$ 且 $p(y,x)=0$，则不可能有 $\mu(x)p(x,y)=\mu(y)p(y,x)$. M/G/1 排队系统也有这个问题. 下面结果给出了链具有可逆测度的充要条件.

定理 5.5.6. Kolmogorov 循环条件. 假设 p 是不可约的. 存在可逆测度的充要条件是 (i) $p(x,y)>0$ 蕴含 $p(y,x)>0$，且 (ii) 对任意环路 $x_0,x_1,\ldots,x_n=x_0$ 满足 $\prod_{1\leqslant i\leqslant n}p(x_i,x_{i-1})>0$，

$$\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}=1.$$

证明. 为证明该条件的必要性，注意不可约性蕴含任何平稳测度都满足 $\mu(x)>0$ 对所有 x 成立，故 (5.5.1) 蕴含 (i) 成立. 为验证 (ii)，注意 (5.5.1) 蕴含对上述序列有

$$\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}=\prod_{i=1}^{n}\frac{\mu(x_i)}{\mu(x_{i-1})}=1.$$

为证明充分性，固定 $a\in S$，令 $\mu(a)=1$，且若 $x_0=a,x_1,\ldots,x_n=x$ 是满足 $\prod_{1\leqslant i\leqslant n}p(x_i,x_{i-1})>0$ 的序列（不可约性保证这样的序列存在），定义

$$\mu(x)=\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}.$$

循环条件保证了该定义与路径无关. 为验证 (5.5.1)，注意若 $p(y,x)>0$，则在到 x 的路径后添加 $x_{n+1}=y$，有

$$\frac{\mu(x)p(x,y)}{p(y,x)}=\mu(y).$$ □

只有特殊的链才有可逆测度，然而如下面结果所示，许多 Markov 链具有平稳测度.

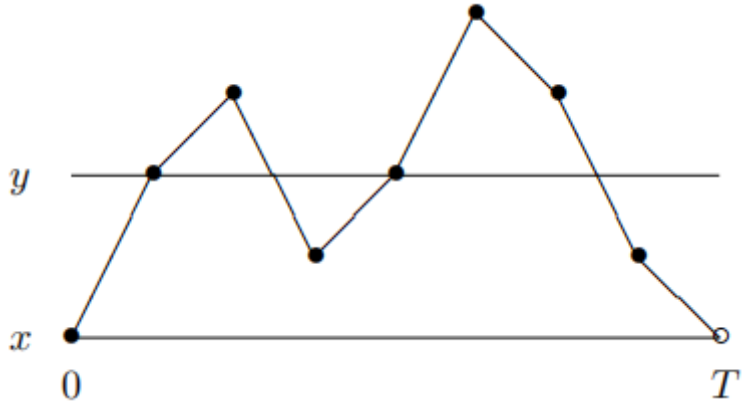

图 5.5: 循环技巧

定理 5.5.7. 设 x 为常返态，且令 $T = \inf\{n \geqslant 1 : X_n = x\}$. 则

$$\mu_x(y) = E_x\left(\sum_{n=0}^{T-1} 1_{\{X_n=y\}}\right) = \sum_{n=0}^{\infty} P_x(X_n = y, T > n)$$

定义了一个平稳测度.

证明. 此即 "**循环技巧**". 用文字描述很简单. $\mu_x(y)$ 是 $\{0,\ldots,T-1\}$ 中访问 y 的期望次数. $\mu_x p(y) \equiv \sum_z \mu_x(z)p(z,y)$ 是 $\{1,\ldots,T\}$ 中访问 y 的期望次数，由于 $X_T = X_0 = x$，它等于 $\mu_x(y)$.

为将这一直观转化为证明，令 $\bar{p}_n(x,y) = P_x(X_n = y, T > n)$ 并利用 Fubini 定理得

$$\sum_y \mu_x(y)p(y,z) = \sum_{n=0}^{\infty}\sum_y \bar{p}_n(x,y)p(y,z).$$

情形 1. $z \neq x$.

$$\sum_y \bar{p}_n(x,y)p(y,z) = \sum_y P_x(X_n = y, T > n, X_{n+1} = z) = P_x(T > n+1, X_{n+1} = z) = \bar{p}_{n+1}(x,z)$$

故 $\sum_{n=0}^{\infty}\sum_y \bar{p}_n(x,y)p(y,z) = \sum_{n=0}^{\infty}\bar{p}_{n+1}(x,z) = \mu_x(z)$，因为 $\bar{p}_0(x,z) = 0$.

情形 2. $z = x$.

$$\sum_y \bar{p}_n(x,y)p(y,x) = \sum_y P_x(X_n = y, T > n, X_{n+1} = x) = P_x(T = n+1)$$

故 $\sum_{n=0}^{\infty}\sum_y \bar{p}_n(x,y)p(y,x) = \sum_{n=0}^{\infty} P_x(T = n+1) = 1 = \mu_x(x)$，因为由定义 $P_x(T = 0) = 0$. □

注释. 若 x 是暂态，则 $\mu_x p(z) \leqslant \mu_x(z)$，且对所有 $z \neq x$ 等号成立.

技术性说明. 为说明没有取巧，应证明 $\mu_x(y) < \infty$ 对所有 y 成立. 首先，观察到 $\mu_x p = \mu_x$ 蕴含 $\mu_x p^n = \mu_x$ 对所有 $n \geqslant 1$ 成立，且 $\mu_x(x) = 1$，故若 $p^n(y,x) > 0$ 则 $\mu_x(y) < \infty$. 由于该结果对所有 n 成立，当 $\rho_{yx} > 0$ 时 $\mu_x(y) < \infty$，但这已足够好. 由定理 5.3.2，当 x 常返时 $\rho_{xy} > 0$ 蕴含 $\rho_{yx} > 0$，由上述论证可得 $\mu_x(y) < \infty$. 若 $\rho_{xy} = 0$ 则 $\mu_x(y) = 0$.

例题 5.5.8. 更新链. $S = \{0, 1, 2, \ldots\}$，$f_k \geqslant 0$，且 $\sum_{k=1}^{\infty} f_k = 1$.

$$\begin{aligned} p(0,j) &= f_{j+1} \quad \text{对} j \geqslant 0, \\ p(i,i-1) &= 1 \quad \text{对} i \geqslant 1, \\ p(i,j) &= 0 \quad \text{其他}. \end{aligned}$$

为解释定义，设 $\xi_1, \xi_2, \ldots$ 独立同分布，$P(\xi_m = j) = f_j$，令 $T_0 = i_0$ 且对 $k \geqslant 1$ 令 $T_k = T_{k-1} + \xi_k$. T_k 是更新过程中第 k 次到达的时刻，该过程的首次到达时刻为 i_0. 令

$$Y_m = \begin{cases} 1 & \text{若} m \in \{T_0, T_1, T_2, \ldots\}, \\ 0 & \text{其他}, \end{cases}$$

且令 $X_n = \inf\{m - n : m \geqslant n, Y_m = 1\}$. $Y_m = 1$ 表示时刻 m 发生更新，X_n 是到首次更新 $\geqslant n$ 的剩余时间.

一个例子有助于澄清定义：

Y_n	0	0	0	1	0	0	1	1	0	0	0	0	1
X_n	3	2	1	0	2	1	0	0	4	3	2	1	0

显然，若 $X_n = i > 0$ 则 $X_{n+1} = i - 1$. 当 $X_n = 0$ 时，下一个 ξ 将以概率 f_k 等于 k. 在此情形下 Y 的下一个正值在时刻 $n + k$，故 $X_{n+1} = k - 1$.

$P_0(T_0 < \infty) = 1$ 故 0 总是常返的. 若 $f_K > 0$ 但 $k > K$ 时 $f_k = 0$，则链在 $\{0, 1, 2, \ldots\}$ 上不是不可约的，但在 $\{0, 1, \ldots, K-1\}$ 上是不可约的.

为计算平稳分布，从 0 出发使用循环技巧. 链击中 0 当且仅当首次跳跃到达某点 $k \geqslant j$，故 $\mu_0(j) = P(\xi > j)$. $E_0 T_0 = \sum_{j=0}^{\infty} P(\xi > j) = E\xi$. 若 $E\xi < \infty$ 则 $\pi(j) = P(\xi > j)/E\xi$ 是平稳分布.

定理 5.5.7 允许为每个常返态闭集构造一个平稳测度. 反之，有：

定理 5.5.9. 若 p 不可约且常返（即所有状态都是常返的），则平稳测度在常数倍意义下唯一.

证明. 设 ν 为平稳测度，$a \in S$.

$$\nu(z) = \sum_y \nu(y) p(y,z) = \nu(a) p(a,z) + \sum_{y \neq a} \nu(y) p(y,z).$$

利用上述恒等式将右端的 $\nu(y)$ 替换，

$$\begin{aligned}\nu(z) &= \nu(a)p(a,z) + \sum_{y\neq a}\nu(a)p(a,y)p(y,z)\\ &\quad + \sum_{x\neq a}\sum_{y\neq a}\nu(x)p(x,y)p(y,z)\\ &= \nu(a)P_a(X_1 = z) + \nu(a)P_a(X_1 \neq a, X_2 = z)\\ &\quad + P_\nu(X_0 \neq a, X_1 \neq a, X_2 = z).\end{aligned}$$

依此类推，得

$$\nu(z) = \nu(a)\sum_{m=1}^{n} P_a(X_k \neq a, 1 \leqslant k < m, X_m = z) + P_\nu(X_j \neq a, 0 \leqslant j < n, X_n = z).$$

最后一项 $\geqslant 0$. 令 $n \to \infty$ 得 $\nu(z) \geqslant \nu(a)\mu_a(z)$，其中 μ_a 是定理 5.5.7 中 $x = a$ 定义的测度.

可能想断言常返性蕴含

$$P_\nu(X_j \neq a, 0 \leqslant j < n) \to 0$$

但 ν 可能是无限测度，故需要另一种方法. 由定理 5.5.7，μ_a 是满足 $\mu_a(a) = 1$ 的平稳测度.（这里从 1 到 T 求和而非从 0 到 $T-1$.）为将最后不等式中的 $\geqslant$ 变为 $=$，注意

$$\nu(a) = \sum_x \nu(x)p^n(x,a) \geqslant \nu(a)\sum_x \mu_a(x)p^n(x,a) = \nu(a)\mu_a(a) = \nu(a).$$

由于 $\nu(x) \geqslant \nu(a)\mu_a(x)$ 且左右两边相等，当 $p^n(x,a) > 0$ 时必须有 $\nu(x) = \nu(a)\mu_a(x)$. 由于 p 不可约，$\nu(x) = \nu(a)\mu_a(x)$ 对所有 $x \in S$ 成立，证毕. □

定理 5.5.7 与 5.5.9 相辅相成. 第一个结果给出了平稳分布的公式（称之为 μ_x），第二个说明它在常数倍意义下唯一. 它们共同使许多公式得以推导.

在考察了平稳测度的存在性和唯一性之后，现转向平稳分布，即满足 $\pi p = \pi$ 的概率测度 π. 平稳测度可能对暂态链存在，例如 $d \geqslant 3$ 的随机游走，但

定理 5.5.10. 若存在平稳分布，则所有满足 $\pi(y) > 0$ 的状态 y 都是常返的.

证明. 由于 $\pi p^n = \pi$，Fubini 定理蕴含

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

当 $\pi(y) > 0$ 时. 利用定理 5.3.1 得

$$\infty = \sum_x \pi(x) \frac{\rho_{xy}}{1-\rho_{yy}} \leqslant \frac{1}{1-\rho_{yy}}$$

因为 $\rho_{xy} \leqslant 1$ 且 π 是概率测度. 故 $\rho_{yy} = 1$. □

定理 5.5.11. 若 p 不可约且有平稳分布 π，则

$$\pi(x) = 1/E_x T_x.$$

注释. 借用 Chung 对定理 4.6.9 的评论，证明将使 $\pi(x) = 1/E_x T_x$ 变得显然，但

$$\sum_x \frac{1}{E_x T_x} p(x,y) = \frac{1}{E_y T_y}$$

看起来难以置信.

证明. 不可约性蕴含 $\pi(x) > 0$，故由定理 5.5.10 所有状态都是常返的. 由定理 5.5.7，

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

定义了满足 $\mu_x(x) = 1$ 的平稳测度，且 Fubini 定理蕴含

$$\sum_y \mu_x(y) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x T_x.$$

由定理 5.5.9，平稳测度在常数倍意义下唯一，故 $\pi(x) = \mu_x(x)/E_x T_x$. 由于 $\mu_x(x) = 1$ 由定义，即得所需结果. □

称状态 x 为**正常返**，若 $E_x T_x < \infty$. 满足 $E_x T_x = \infty$ 的常返态称为**零常返**. 定理 5.6.1 将解释这些名称的由来. 下面结果帮助识别正常返状态.

定理 5.5.12. 若 p 不可约，则下列等价：

(i) 某 x 是正常返的.

(ii) 存在平稳分布.

(iii) 所有状态都是正常返的.

该结果表明正常返是一个**类性质**. 若它对不可约集中某一状态成立，则对所有状态都成立.

证明. (i) 蕴含 (ii). 若 x 正常返则

$$\pi(y)=\sum_{n=0}^{\infty}P_x(X_n=y,T_x>n)/E_xT_x$$

定义了一个平稳分布.

(ii) 蕴含 (iii). 定理 5.5.11 蕴含 $\pi(y)=1/E_yT_y$，不可约性表明 $\pi(y)>0$ 对所有 y 成立，故 $E_yT_y<\infty$.

(iii) 蕴含 (i). 显然. □

例题 5.5.13. 生灭链存在平稳分布当且仅当

$$\sum_x\prod_{k=1}^{x}\frac{p_{k-1}}{q_k}<\infty.$$

由定理 5.3.11，链常返当且仅当

$$\sum_{m=0}^{\infty}\prod_{j=1}^{m}\frac{q_j}{p_j}=\infty.$$

例题 5.5.14. M/G/1 排队系统. 令 $\mu=\sum ka_k$ 为一次服务时间内到达顾客的平均数. 在例 5.3.7 中，证明了链常返当且仅当 $\mu\leqslant 1$. 现证明链正常返当且仅当 $\mu<1$. 首先，设 $\mu<1$. 当 $X_n>0$ 时，链的行为类似于均值为 $\mu-1$ 的随机游走的跳跃，故若 $N=\inf\{n\geqslant 0:X_n=0\}$ 则 $X_{N\wedge n}-(\mu-1)(N\wedge n)$ 是鞅. 若 $X_0=x>0$ 则鞅性质蕴含

$$x=E_xX_{N\wedge n}+(1-\mu)E_x(N\wedge n)\geqslant(1-\mu)E_x(N\wedge n)$$

因为 $X_{N\wedge n}\geqslant 0$，由此 $E_xN\leqslant x/(1-\mu)$.

为证明等号成立，观察到 X_n 每次最多减少一，且对 $x\geqslant 1$ 有 $E_xT_{x-1}=E_1T_0$，故 $E_xN=cx$. 为确定常数，注意到

$$E_1N=1+\sum_{k=0}^{\infty}a_kE_kN$$

故 $c=1+\mu c$ 且 $c=1/(1-\mu)$. 若 $X_0=0$ 则 $p(0,0)=a_0+a_1$ 且 $p(0,k-1)=a_k$ 对 $k\geqslant 2$. 通过考虑首次跳跃发生的情况，看到（第一项可能看起来不对，但注意 $k=1$ 时 $k-1=0$）

$$E_0T_0=1+\sum_{k=1}^{\infty}a_k\frac{k-1}{1-\mu}=1+\frac{\mu-(1-a_0)}{1-\mu}=\frac{a_0}{1-\mu}<\infty.$$

这表明当 $\mu<1$ 时链是正常返的. 为证明逆命题，观察到上述论证表明若 $E_0T_0<\infty$ 则 $E_kN<\infty$ 对所有 k 成立，$E_kN=ck$，且 $c=1/(1-\mu)$，当 $\mu\geqslant 1$ 时这是不可能的.

最后一个结果结合定理 5.5.7 和 5.5.11 可得：平稳分布满足 $\pi(0) = (1-\mu)/a_0$. 这似乎不多，但 $\pi p = \pi$ 的方程为：

$$\begin{aligned}
\pi(0) &= \pi(0)(a_0 + a_1) + \pi(1)a_0, \\
\pi(1) &= \pi(0)a_2 + \pi(1)a_1 + \pi(2)a_0, \\
\pi(2) &= \pi(0)a_3 + \pi(1)a_2 + \pi(2)a_1 + \pi(3)a_0,
\end{aligned}$$

或一般地，对 $j \geqslant 1$

$$\pi(j) = \sum_{i=0}^{j+1} \pi(i)a_{j+1-i}.$$

这些方程具有 “三角” 形式，故知道 $\pi(0)$，可以求解 $\pi(1), \pi(2), \ldots$. 第一个表达式，

$$\pi(1) = \pi(0)(1 - (a_0 + a_1))/a_0$$

很简单，但公式变得越来越复杂，没有好的闭式解.

本节最后给出一个自洽的证明：

定理 5.5.15. 若 p 不可约且有平稳分布 π，则任何其他平稳测度都是 π 的倍数.

注释. 这一结果是定理 5.5.10 和定理 5.5.9 的推论，但证明方法很有趣.

证明. 由 p 的不可约性，$\pi(x) > 0$ 对所有 x 成立. 令 ϕ 为 $(0, \infty)$ 上的凹函数且有界，例如 $\phi(x) = x/(x+1)$. 定义 μ 的**熵**为

$$E(\mu) = \sum_y \phi\left(\frac{\mu(y)}{\pi(y)}\right)\pi(y).$$

名称的由来将在证明过程中阐明.

$$\begin{aligned}
E(\mu p) &= \sum_y \phi\left(\frac{\sum_x \mu(x)p(x,y)}{\pi(y)}\right)\pi(y) \\
&= \sum_y \phi\left(\sum_x \frac{\mu(x)}{\pi(x)} \cdot \frac{\pi(x)p(x,y)}{\pi(y)}\right)\pi(y) \\
&\geqslant \sum_y \sum_x \phi\left(\frac{\mu(x)}{\pi(x)}\right)\frac{\pi(x)p(x,y)}{\pi(y)}\pi(y)
\end{aligned}$$

因为 ϕ 是凹函数，且 $\nu(x) = \pi(x)p(x,y)/\pi(y)$ 是概率分布. 由于 $\pi(y)$ 抵消且 $\sum_y p(x,y) = 1$，最后一式等于 $E(\mu)$，证明了 $E(\mu p) \geqslant E(\mu)$，即任意初始测度 μ 的熵经 p 作用后增加.

若对所有 x, y 有 $p(x,y) > 0$，且 $\mu p = \mu$，则 $\mu(x)/\pi(x)$ 必为常数，否则 Jensen 不等式的应用将产生严格不等式. 为从这一特殊情形得到一般结果，观察到若 p 不可约

$$\bar{p}(x,y) = \sum_{n=1}^{\infty} 2^{-n} p^n(x,y) > 0 \quad \text{对所有} x, y$$

且 $\mu p = \mu$ 蕴含 $\mu \bar{p} = \mu$. □

习题

习题 5.5.1. 求习题 5.1.5 中 Bernoulli-Laplace 扩散模型的平稳分布.

习题 5.5.2. 令 $w_{xy} = P_x(T_y < T_x)$. 证明 $\mu_x(y) = w_{xy}/w_{yx}$.

习题 5.5.3. 证明若 p 不可约且常返则

$$\mu_x(y)\mu_y(z) = \mu_x(z).$$

习题 5.5.4. 设 p 不可约且正常返. 则 $E_x T_y < \infty$ 对所有 x, y 成立.

习题 5.5.5. 设 p 不可约且有平稳测度 μ 满足 $\sum_x \mu(x) = \infty$. 则 p 不是正常返的.

习题 5.5.6. 利用定理 5.5.7 和 5.5.9 证明对简单随机游走，(i) 相继访问 0 之间访问 k 的期望次数对所有 k 均为 1，(ii) 若从 k 出发，击中 0 之前访问 k 的期望次数为 $2k$.

习题 5.5.7. 设 $\xi_1, \xi_2, \ldots$ 独立同分布，$P(\xi_m = k) = a_{k+1}$ 对 $k \geqslant -1$，令 $S_n = x + \xi_1 + \cdots + \xi_n$，其中 $x \geqslant 0$，且令

$$X_n = S_n + \left(\min_{m \leqslant n} S_m \right)^{-}.$$

(5.1.1) 表明 X_n 与从 $X_0 = x$ 出发的 M/G/1 排队系统有相同分布. 利用这一表示得出结论：若 $\mu = \sum k a_k < 1$，则当 $n \to \infty$

$$\frac{1}{n} |\{m \leqslant n : X_{m-1} = 0, \xi_m = -1\}| \to (1 - \mu) \quad \text{a.s.}$$

从而 $\pi(0) = (1-\mu)/a_0$ 如上所证.

习题 5.5.8. M/M/∞ 排队系统. 在此链中，由习题 5.3.8 引入，

$$X_{n+1} = \sum_{m=1}^{X_n} \xi_{n,m} + Y_{n+1}$$

其中 $\xi_{n,m}$ 是均值为 p 的独立同分布 Bernoulli 随机变量，Y_{n+1} 是独立的 Poisson(λ) 随机变量. 当 $X_0 = \text{Poisson}(\mu)$ 时计算时刻 1 的分布，并利用结果求平稳分布.

习题 5.5.9. 设 $X_n \geqslant 0$ 为 Markov 链且假设 $E_x X_1 \leqslant x - \epsilon$ 对 $x > K$，其中 $\epsilon > 0$. 令 $Y_n = X_n + n\epsilon$ 且 $\tau = \inf\{n : X_n \leqslant K\}$. $Y_{n\wedge\tau}$ 是正上鞅且可选停时定理蕴含 $E_x\tau \leqslant x/\epsilon$. 再结合关于链从点 $x \leqslant K$ 出发的行为之假设，可导出正常返的条件.

第 6 节 渐近行为

本节首先研究 $p^n(x,y)$ 的渐近行为. 若 y 为暂态，则 $\sum_n p^n(x,y) < \infty$，故当 $n \to \infty$ 时 $p^n(x,y) \to 0$. 为处理常返态，令

$$N_n(y) = \sum_{m=1}^{n} 1_{\{X_m = y\}}$$

为到时刻 n 为止访问 y 的次数.

定理 5.6.1. 设 y 为常返态. 则对任意 $x \in S$，当 $n \to \infty$ 时

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y} 1_{\{T_y < \infty\}} \quad P_x\text{-a.s.}$$

其中 $1/\infty = 0$.

证明. 首先设从 y 出发. 令 $R(k) = \min\{n \geqslant 1 : N_n(y) = k\}$ 为第 k 次返回 y 的时间. 令 $t_k = R(k) - R(k-1)$，其中 $R(0) = 0$. 因假设 $X_0 = y$，故 $t_1, t_2, \ldots$ 独立同分布，由强大数定律得

$$\frac{R(k)}{k} \to E_y T_y \quad P_y\text{-a.s.}$$

由 $R(N_n(y)) \leqslant n < R(N_n(y)+1)$ 知

$$\frac{R(N_n(y))}{N_n(y)} \leqslant \frac{n}{N_n(y)} < \frac{R(N_n(y)+1)}{N_n(y)+1} \cdot \frac{N_n(y)+1}{N_n(y)}.$$

令 $n \to \infty$，并注意 $N_n(y) \to \infty$ 几乎必然（因 y 常返），得

$$\frac{n}{N_n(y)} \to E_y T_y \quad P_y\text{-a.s.}$$

现推广至 $x \neq y$ 的情形. 注意若 $T_y = \infty$，则对所有 n 有 $N_n(y) = 0$，故在 $\{T_y = \infty\}$ 上

$$\frac{N_n(y)}{n} \to 0.$$

强 Markov 性质蕴含，在 $\{T_y < \infty\}$ 上，$t_2, t_3, \ldots$ 独立同分布且 $P_x(t_k = n) = P_y(T_y = n)$，故

$$\frac{R(k)}{k} = \frac{t_1}{k} + \frac{t_2 + \cdots + t_k}{k} \to 0 + E_y T_y \quad P_x\text{-a.s.}$$

重复 $x = y$ 情形的证明，得

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y} \quad P_x\text{-a.s. 于 } \{T_y < \infty\} \text{ 上,}$$

再与 $\{T_y=\infty\}$ 上的结果结合，即完成证明. □

注释. 定理 5.6.1 有助于解释**正常返**与**零常返**这两个术语. 若从 x 出发，则第一种情形渐近地有正比例的时间花在 x 上，而第二种情形这一比例为 0.

由 $0\leqslant N_n(y)/n\leqslant 1$ 及有界收敛定理得 $E_xN_n(y)/n\to E_x(1_{\{T_y<\infty\}}/E_yT_y)$，故

$$\frac{1}{n}\sum_{m=1}^{n}p^m(x,y)\to\frac{\rho_{xy}}{E_yT_y}. \tag{5.6.1}$$

最后这个结果对常返的 y 已证，但对暂态的 y 亦成立，因此时 $E_yT_y=\infty$ 而极限为 0，由于 $\sum_m p^m(x,y)<\infty$. (5.6.1) 表明序列 $p^n(x,y)$ 在 **Cesaro 意义**下总是收敛的. 下面的例子表明 $p^n(x,y)$ 未必收敛.

例题 5.6.2.

$$p=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad p^2=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad p^3=p,\quad p^4=p^2,\dots$$

Ehrenfest 链亦有类似的问题. 在该链中，若 X_0 为偶数，则 X_1 为奇数，X_2 为偶数，…，故 $p^n(x,x)=0$ 除非 n 为偶数. 易构造例子使得 $p^n(x,x)=0$ 除非 n 是 3 或 17 或 … 的倍数.

下面的定理 5.6.6 将表明，这种 “周期性” 是阻止 $p^n(x,y)$ 收敛的唯一障碍. 首先给出一个定义和两个预备结果. 设 x 为常返态，令 $I_x=\{n\geqslant 1:p^n(x,x)>0\}$，并令 d_x 为 I_x 的**最大公约数**. d_x 称为 x 的**周期**.

例题 5.6.3. 三角形与正方形. 考虑如下转移矩阵：

	-2	-1	0	1	2	3
-2	0	0	1	0	0	0
-1	1	0	0	0	0	0
0	0	0.5	0	0.5	0	0
1	0	0	0	0	1	0
2	0	0	0	0	0	1
3	0	0	1	0	0	0

换言之，从 0 出发，以相等概率到达 1 或 -1. 从 -1 以概率 1 到达 -2 然后返回 0，从 1 到达 2 再到 3 然后返回 0. 名称的由来是 $0\to-1\to-2\to 0$ 构成三角形，而 $0\to 1\to 2\to 3\to 0$ 构成正方形.

显然，$p^3(0,0)>0$ 且 $p^4(0,0)>0$，故 $3,4\in I_0$，从而 $d_0=1$.

下面的结果表明周期是**类性质**. 特别地，由此可得上例中所有状态都有周期 1.

引理 5.6.4. 若 $\rho_{xy}>0$，则 $d_y=d_x$.

证明. 设 K 和 L 满足 $p^K(x,y)>0$ 和 $p^L(y,x)>0$.（x 常返，故 $\rho_{yx}>0$.）由

$$p^{K+L}(y,y)\geqslant p^L(y,x)p^K(x,y)>0,$$

知 d_y 整除 $K+L$，简记为 $d_y\mid(K+L)$. 设 n 满足 $p^n(x,x)>0$，则

$$p^{K+n+L}(y,y)\geqslant p^L(y,x)p^n(x,x)p^K(x,y)>0,$$

故 $d_y\mid(K+n+L)$, 从而 $d_y\mid n$. 因 $n\in I_x$ 任意, $d_y\mid d_x$. 交换 y 与 x 的角色得 $d_x\mid d_y$, 故 $d_x=d_y$. □

若链不可约且 $d_x=1$，则称它为**非周期的**. 检验这一点的最简单方法是找到满足 $p(x,x)>0$ 的状态. M/G/1 排队系统对所有 $k\geqslant 0$ 有 $a_k>0$，故具有此性质. **更新链**当 g.c.d.$\{k:f_k>0\}=1$ 时为非周期的.

引理 5.6.5. 若 $d_x=1$，则对所有 $m\geqslant m_0$ 有 $p^m(x,x)>0$.

例子证明. 考虑更新链（例 5.5.8）满足 $f_5=f_{12}=1/2$. $5,12\in I_0$. 由于 $m,n\in I_0$ 蕴含 $m+n\in I_0$

$$I_0=\{5,10,12,15,17,20,22,24,25,27,29,30,32,$$
$$34,35,36,37,39,40,41,42,43,\ldots\}.$$

为检验这一点，注意 5 产生 $10=5+5$ 和 $17=5+12$，10 产生 15 和 22，12 产生 17 和 24，等等. 一旦 I_0 中有五个连续整数，这里是 39–43，则其余所有整数皆在其中. □

证明. 首先注意，只需证明 I_x 包含两个连续整数 k 和 $k+1$. 因为此时它将包含 $2k,2k+1,2k+2,3k,3k+1,3k+2,3k+3$, 如此继续直到有 k 个连续整数, 然后 $(k-1)k,(k-1)k+1,\ldots,(k-1)k+k-1$, 于是其余所有整数皆在其中.

为证存在两个连续整数，我们在此借用数论中的一个结果：若集合 I_x 的最大公约数为 1，则存在 $i_1,\ldots,i_m\in I_x$ 以及（正或负）整数系数 c_i 使得 $c_1i_1+\cdots+c_mi_m=1$. 令 $a_i=c_i^+$ 和 $b_i=(c_i)^-$. 换言之，a_i 是正系数而 b_i 是 -1 乘以负系数. 重排最后一个方程得

$$a_1i_1+\cdots+a_mi_m=(b_1i_1+\cdots+b_mi_m)+1,$$

于是在 I_x 中找到了两个连续整数. 在数值例子中 $5\cdot 5-2\cdot 12=1$，给出连续整数 24 和 25. □

定理 5.6.6. 收敛定理. 设 p 不可约、非周期（即所有状态满足 $d_x=1$）且有**平稳分布** π. 则当 $n\to\infty$ 时，$p^n(x,y)\to\pi(y)$.

证明. 令 $S^2=S\times S$. 在 $S\times S$ 上定义转移概率 $\bar{p}$ 为

$$\bar{p}((x_1,y_1),(x_2,y_2))=p(x_1,x_2)p(y_1,y_2),$$

即每个坐标独立运动. 第一步是检验 $\bar{p}$ 不可约. 这看起来像是先做傻事，但这是唯一需要非周期性的步骤. 因 p 不可约，存在 K, L 使得 $p^K(x_1, x_2) > 0$ 和 $p^L(y_1, y_2) > 0$. 由引理 5.6.5，若 M 充分大，则 $p^{L+M}(x_2, x_2) > 0$ 和 $p^{K+M}(y_2, y_2) > 0$，故

$$\bar{p}^{K+L+M}((x_1, y_1), (x_2, y_2)) > 0.$$

第二步，注意因两个坐标独立，$\bar{\pi}(a, b) = \pi(a)\pi(b)$ 定义了 $\bar{p}$ 的一个平稳分布，而定理 5.5.10 蕴含对 $\bar{p}$ 所有状态皆常返. 令 (X_n, Y_n) 表示 $S \times S$ 上的链，令 T 为该链首次击中**对角线** $\{(y, y) : y \in S\}$ 的时间. 令 $T_{(x,x)}$ 为击中 (x, x) 的时间. 因 $\bar{p}$ 不可约且常返，$T_{(x,x)} < \infty$ 几乎必然，从而 $T < \infty$ 几乎必然.

最后一步，注意在 $\{T \leqslant n\}$ 上，两个坐标 X_n 和 Y_n 有相同分布. 通过考虑首次相交的时间和位置然后利用 Markov 性质，

$$\begin{aligned} P(X_n = y, T \leqslant n) &= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x, X_n = y) \\ &= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x) P(X_n = y \mid X_m = x) \\ &= \sum_{m=1}^{n} \sum_{x} P(T = m, Y_m = x) P(Y_n = y \mid Y_m = x) \\ &= P(Y_n = y, T \leqslant n). \end{aligned}$$

为完成证明，注意

$$P(X_n = y) = P(Y_n = y, T \leqslant n) + P(X_n = y, T > n) \leqslant P(Y_n = y) + P(X_n = y, T > n),$$

类似地，$P(Y_n = y) \leqslant P(X_n = y) + P(Y_n = y, T > n)$. 故

$$|P(X_n = y) - P(Y_n = y)| \leqslant P(X_n = y, T > n) + P(Y_n = y, T > n),$$

对 y 求和得

$$\sum_{y} |P(X_n = y) - P(Y_n = y)| \leqslant 2P(T > n).$$

若令 $X_0 = x$ 且 Y_0 具有平稳分布 π，则 Y_n 的分布为 π，于是

$$\sum_{y} |p^n(x, y) - \pi(y)| \leqslant 2P(T > n) \to 0,$$

证得所需结果. 若回顾 3.6 节给出的**全变差距离**的定义，最后的结论可写为

$$\|p^n(x,\cdot)-\pi(\cdot)\| \leqslant P(T>n) \to 0.$$

□

乍看之下，通过运行链的独立副本来证明收敛定理似乎奇怪. 一种稍复杂但能更好解释发生何事的途径是定义

$$q((x_1,y_1),(x_2,y_2)) = \begin{cases} p(x_1,x_2)p(y_1,y_2) & 若x_1 \neq y_1, \\ p(x_1,x_2) & 若x_1=y_1, x_2=y_2, \\ 0 & 其他. \end{cases}$$

换言之，两个坐标独立运动直到相遇，然后一起运动. 由定义易见每个坐标都是原过程的复制. 若 T' 为新链 (X'_n,Y'_n) 击中**对角线**的时间，则在 $T' \leqslant n$ 上 $X'_n = Y'_n$，故显然

$$\sum_y |P(X'_n=y)-P(Y'_n=y)| \leqslant 2P(X'_n \neq Y'_n) = 2P(T'>n).$$

另一方面，T 和 T' 同分布，故 $P(T'>n)\to 0$，结论如前所述. 最后证明中使用的技术称为**耦合**. 一般地，这一术语指在同一空间上构造两个序列 X_n 和 Y_n，通过证明 $P(X_n \neq Y_n)\to 0$，或更一般地，对某度量 ρ 有 $\rho(X_n,Y_n)\to 0$ 依概率，来得出 X_n 依分布收敛的结论.

习题

习题 5.6.1. 设 $S=\{0,1\}$ 且

$$p = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

用归纳法证明

$$P_\mu(X_n=0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n \left\{ \mu(0) - \frac{\beta}{\alpha+\beta} \right\}.$$

习题 5.6.2. 证明若 S 有限且 p 不可约且非周期，则存在 m 使得对所有 x,y 有 $p^m(x,y)>0$.

习题 5.6.3. 证明若 S 有限，p 不可约且非周期，且 T 是定理 5.6.6 证明中定义的**耦合时间**，则 $P(T>n) \leqslant Cr^n$ 对某 $r<1$ 和 $C<\infty$ 成立. 故在此情形收敛到平衡以指数速度发生. 提示：首先考虑对所有 x 和 y 有 $p(x,y)>0$ 的情形，然后通过观察 p 的某次幂将一般情形归结为这种情形.

习题 5.6.4. 对任意转移矩阵 p，定义

$$\alpha_n = \sup_{i,j} \frac{1}{2} \sum_k |p^n(i,k) - p^n(j,k)|.$$

此处的 $1/2$ 是因对任意 i 和 j 可定义随机变量 X 和 Y 使得 $P(X = k) = p^n(i,k)$，$P(Y = k) = p^n(j,k)$，且

$$P(X \neq Y) = \frac{1}{2} \sum_k |p^n(i,k) - p^n(j,k)|.$$

证明 $\alpha_{m+n} \leqslant \alpha_n \alpha_m$. 这里**耦合**的解释有助于避免陷入代数运算. 利用引理 2.7.1，可得

$$\frac{1}{n} \log \alpha_n \to \inf_{m \geqslant 1} \frac{1}{m} \log \alpha_m,$$

故若对某 m 有 $\alpha_m < 1$，则它以指数速度趋于 0.

习题 5.6.5. 加性泛函的强大数定律. 设 p 不可约且有**平稳分布** π. 设 f 为满足 $\sum_y |f(y)|\pi(y) < \infty$ 的函数. 令 T_x^k 为第 k 次返回 x 的时间.（i）证明

$$V_k^f = f(X(T_x^k)) + \cdots + f(X(T_x^{k+1} - 1)), \quad k \geqslant 1$$

独立同分布且 $E|V_k^f| < \infty$.（ii）令 $K_n = \inf\{k : T_x^k \geqslant n\}$ 并证明

$$\frac{1}{n} \sum_{m=1}^{K_n} V_m^f \to \frac{EV_1^f}{E_x T_x^1} = \sum_y f(y)\pi(y) \quad P_\mu\text{-a.s.}$$

（iii）证明 $\max_{1 \leqslant m \leqslant n} V_m^{|f|}/n \to 0$ 并推出

$$\frac{1}{n} \sum_{m=1}^{n} f(X_m) \to \sum_y f(y)\pi(y) \quad P_\mu\text{-a.s.}$$

对任意初始分布 μ 成立.

习题 5.6.6. 加性泛函的中心极限定理. 设除习题 5.6.5 的条件外还有 $\sum_y f(y)\pi(y) = 0$ 且 $E_x(V_k^{|f|})^2 < \infty$.（i）利用**随机指标中心极限定理**（习题 3.4.6）得出对任意初始分布 μ

$$\frac{1}{\sqrt{n}} \sum_{m=1}^{K_n} V_m^f \Rightarrow c\chi \quad \text{在} P_\mu \text{ 下.}$$

(ii) 证明 $\max_{1\leqslant m\leqslant n} V_m^{|f|}/\sqrt{n}\to 0$ 依概率并推出

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n} f(X_m)\Rightarrow c\chi \quad \text{在}P_\mu \text{下}.$$

习题 5.6.7. 比值极限定理. 定理 5.6.1 在零常返情形说的不多. 为得到更有信息量的极限定理，设 y 常返且 m 是 $C_y=\{z:\rho_{yz}>0\}$ 上的（在常数倍意义下唯一的）**平稳测度**. 令 $N_n(z)=|\{m\leqslant n: X_m=z\}|$. 在相继返回 y 之间分割路径，并证明

$$\frac{N_n(z)}{N_n(y)}\to\frac{m(z)}{m(y)} \quad P_x\text{-a.s.}$$

对所有 $x,z\in C_y$ 成立. 注意 $n\mapsto N_n(z)$ 递增，故这比前一问题容易得多.

习题 5.6.8. 由定理 5.6.1 取期望得到 (5.6.1). 这对前一练习中的比值行不通，故需另寻他法. 设 $z\neq y$.
(i) 令 $\bar{p}^n(x,z)=P_x(X_n=z,T_y>n)$ 并按 $J=\sup\{j\in[1,m):X_j=y\}$ 的值分解 $p^m(x,z)$ 得

$$\sum_{m=1}^{n} p^m(x,z)=\sum_{m=1}^{n}\bar{p}^m(x,z)+\sum_{j=1}^{n-1}p^j(x,y)\sum_{k=1}^{n-j}\bar{p}^k(y,z).$$

(ii) 证明

$$\frac{\sum_{m=1}^{n} p^m(x,z)}{\sum_{m=1}^{n} p^m(x,y)}\to\frac{m(z)}{m(y)}.$$

第 7 节　周期性，尾 σ 域 *

引理 5.7.1. 假设 p 不可约、常返且所有状态的周期为 d. 固定 $x \in S$，对每个 $y \in S$ 令 $K_y = \{n \geqslant 1 : p^n(x,y) > 0\}$.

(i) 存在 $r_y \in \{0,1,\ldots,d-1\}$ 使得若 $n \in K_y$ 则 $n \equiv r_y \pmod d$，即 $n - r_y$ 是 d 的倍数.

(ii) 令 $S_r = \{y : r_y = r\}$ $(0 \leqslant r < d)$. 若 $y \in S_i$，$z \in S_j$ 且 $p^n(y,z) > 0$，则 $n \equiv j - i \pmod d$.

(iii) $S_0, S_1, \ldots, S_{d-1}$ 是 p^d 的不可约类，且所有状态周期为 1.

证明. (i) 设 $m(y)$ 满足 $p^{m(y)}(y,x) > 0$. 若 $n \in K_y$，则 $p^{n+m(y)}(x,x)$ 为正，故 $d \mid (n+m)$. 令 $r_y = (d - m(y)) \pmod d$.

(ii) 设 m, n 满足 $p^n(y,z), p^m(x,y) > 0$. 由于 $p^{n+m}(x,z) > 0$，由 (i) 知 $n + m = j \pmod d$. 由于 $m = i \pmod d$，结论得证. (iii) 中的不可约性由 (ii) 直接推出. 非周期性源于周期的定义：周期是集合 $\{n \geqslant 1 : p^n(x,x) > 0\}$ 的最大公约数. □

满足引理 5.7.1(ii) 的状态空间划分 $S_0, S_1, \ldots, S_{d-1}$ 称为状态空间的**循环分解**. 除选择哪个集合放在首位外，它是唯一的.（任取 $x \in S$，它属于某个 S_j；一旦知道 j 的值，由不可约性和 (ii) 即可算出所有集合.）

习题 5.7.1. 求具有如下转移概率的 Markov 链的循环分解：

	1	2	3	4	5	6	7
1	0	0	0	.5	.5	0	0
2	.3	0	0	0	0	0	.7
3	0	0	0	0	0	0	1
4	0	0	1	0	0	0	0
5	0	0	1	0	0	0	0
6	0	1	0	0	0	0	0
7	0	0	0	.4	0	.6	0

定理 5.7.2. 周期情形的收敛定理. 假设 p 不可约，具有平稳分布 π，且所有状态周期为 d. 令 $x \in S$，并设 $S_0, S_1, \ldots, S_{d-1}$ 为状态空间的循环分解且 $x \in S_0$. 若 $y \in S_r$，则

$$\lim_{m\to\infty} p^{md+r}(x,y) = \pi(y)d$$

证明. 若 $y \in S_0$，则由引理 5.7.1(iii) 并对 p^d 应用定理 5.6.6 知

$$\lim_{m\to\infty} p^{md}(x,y) \text{ 存在}$$

为确定极限，注意 (5.6.1) 蕴含

$$\frac{1}{n}\sum_{m=1}^{n} p^m(x,y) \to \pi(y)$$

且引理 5.7.1(ii) 蕴含 $p^m(x,y)=0$ 除非 $d \mid m$，故第一个极限必为 $\pi(y)d$. 若 $y \in S_r$ 且 $1 \leqslant r < d$，则

$$p^{md+r}(x,y) = \sum_{z \in S_r} p^r(x,z) p^{md}(z,y)$$

由于 $y, z \in S_r$，由第一种情形的证明知 $p^{md}(z,y) \to \pi(y)d$ $(m \to \infty)$. $\sum_z p^r(x,z)=1$ 且 $p^{md}(z,y) \leqslant 1$，故由控制收敛定理即得结论. □

令 $\mathcal{F}_n^0 = \sigma(X_{n+1}, X_{n+2}, \ldots)$ 且 $\mathcal{T} = \bigcap_n \mathcal{F}_n^0$ 为**尾 σ-域**. 下述结果属于 Orey. 我们给出的证明来自 Blackwell 与 Freedman (1964).

定理 5.7.3. 假设 p 不可约、常返且所有状态周期为 d，则 $\mathcal{T} = \sigma(\{X_0 \in S_r\} : 0 \leqslant r < d)$.

注释. 为精确起见，若 μ 为任一初始分布且 $A \in \mathcal{T}$，则存在 r 使得 $A = \{X_0 \in S_r\}$ P_μ-a.s.

证明. 下面分三步建立一般结果.

情形 1. 设 $P(X_0 = x) = 1$. 令 $T_0 = 0$，且对 $n \geqslant 1$，令 $T_n = \inf\{m > T_{n-1} : X_m = x\}$ 为第 n 次返回 x 的时刻. 令 $V_n = (X(T_{n-1}), \ldots, X(T_n - 1))$. 由习题 5.3.1 知向量 V_n 独立同分布，且尾 σ 域包含于 V_n 的可交换 σ 域中，故由 Hewitt-Savage 0-1 律（定理 2.5.4，那里对取值于一般可测空间的随机变量证明了该结果）知 $\mathcal{T}$ 在此情形下平凡.

情形 2. 设初始分布集中在某个循环类上，比如 S_0. 若 $A \in \mathcal{T}$，则由情形 1 知对每个 x 有 $P_x(A) \in \{0,1\}$. 若对所有 $x \in S_0$ 有 $P_x(A)=0$，则 $P_\mu(A)=0$. 设对某个 $y \in S_0$ 有 $P_y(A)>0$，从而 $=1$. 令 $z \in S_0$. 由于 p^d 在 S_0 上不可约且非周期，存在 n 使得 $p^n(z,y)>0$ 且 $p^n(y,y)>0$. 若写 $1_A = 1_B \circ \theta_n$，则 Markov 性质蕴含

$$1 = P_y(A) = E_y(E_y(1_B \circ \theta_n \mid \mathcal{F}_n)) = E_y(E_{X_n} 1_B)$$

故 $P_y(B)=1$. Markov 性质的另一应用给出

$$P_z(A) = E_z(E_{X_n} 1_B) \geqslant p^n(z,y) > 0$$

故 $P_z(A)=1$，且因 $z \in S_0$ 任意，$P_\mu(A)=1$.

一般情形. 由情形 2 知，$P(A \mid X_0 = y) \equiv 1$ 或 $\equiv 0$ 在每个循环类上. 故 $\{X_0 \in S_r\} \subset A$ 或 $\{X_0 \in S_r\} \cap A = \emptyset$ P_μ-a.s. 反之，显然 $\{X_0 \in S_r\} = \{X_{nd} \in S_r \text{ i.o.}\} \in \mathcal{T}$，证毕. □

下述结果有助于在暂态例子中识别尾 σ 域.

定理 5.7.4. 设 X_0 具有初始分布 μ. 方程

$$h(X_n,n)=E_\mu(Z\mid\mathcal{F}_n)\quad\text{和}\quad Z=\lim_{n\to\infty}h(X_n,n)$$

在有界 $Z\in\mathcal{T}$ 与有界**时空调和函数**之间建立一一对应, 即有界 $h:S\times\{0,1,\ldots\}\to\mathbb{R}$ 使得 $h(X_n,n)$ 是鞅.

证明. 令 $Z\in\mathcal{T}$，写 $Z=Y_n\circ\theta_n$，并令 $h(x,n)=E_xY_n$. 由 Markov 性质

$$E_\mu(Z\mid\mathcal{F}_n)=E_\mu(Y_n\circ\theta_n\mid\mathcal{F}_n)=h(X_n,n)$$

故 $h(X_n,n)$ 是鞅. 反之，若 $h(X_n,n)$ 是有界鞅，利用定理 4.2.11 和 4.6.7 可知 $h(X_n,n)\to Z\in\mathcal{T}$ (当 $n\to\infty$)，且 $h(X_n,n)=E_\mu(Z\mid\mathcal{F}_n)$. □

习题 5.7.2. 满足 $Z=Z\circ\theta$（从而 $Z=Z\circ\theta_n$ 对所有 n）的随机变量 Z 称为**不变**的. 证明有界不变随机变量与有界调和函数之间存在一一对应. 我们将在 7.1 节进一步讨论不变随机变量.

例题 5.7.5. d **维简单随机游走**. 首先为该过程构造一个耦合. 令 $i_1,i_2,\ldots$ 为独立同分布且服从 $\{1,\ldots,d\}$ 上的均匀分布. 令 $\xi_1,\xi_2,\ldots$ 和 $\eta_1,\eta_2,\ldots$ 为独立同分布且服从 $\{-1,1\}$ 上的均匀分布. 令 e_j 为第 j 个单位向量. 通过下式构造一对耦合的 d 维简单随机游走:

$$X_n=X_{n-1}+e_{i_n}\xi_n$$

$$Y_n=\begin{cases}Y_{n-1}+e_{i_n}\xi_n & \text{若}X_{n-1}^{i_n}=Y_{n-1}^{i_n}\\ Y_{n-1}+e_{i_n}\eta_n & \text{若}X_{n-1}^{i_n}\neq Y_{n-1}^{i_n}\end{cases}$$

换言之，两个游走改变的坐标总相同；一旦它们在某坐标上一致，该方向上的未来运动也相同. 易见若对所有 $1\leqslant i\leqslant d$ 有 $X_0^i-Y_0^i$ 为偶数，则两随机游走以概率 1 相遇.

令 $L_0=\{z\in\mathbb{Z}^d:z^1+\cdots+z^d$ 为偶数$\}$ 且 $L_1=\mathbb{Z}^d-L_0$. 虽然我们只在常返情形定义了这一概念，但显然 L_0,L_1 是简单随机游走的状态空间循环分解. 若 $S_n\in L_i$ 则 $S_{n+1}\in L_{1-i}$ 且 p^2 在每个 L_i 上不可约. 为耦合从 $x,y\in L_i$ 出发的两个随机游走，先令它们独立运行，直到所有坐标差首次全为偶数，然后使用上述耦合. 在剩余情形 $x\in L_0,y\in L_1$ 下，耦合是不可能的.

下述结果应能解释我们对耦合两个 d 维简单随机游走的兴趣.

定理 5.7.6. 对 d 维简单随机游走,

$$\mathcal{T}=\sigma(\{X_0\in L_i\},i=0,1)$$

证明. 令 $x,y\in L_i$，并令 X_n,Y_n 为上述耦合在 $X_0=x$ 和 $Y_0=y$ 下的实现. 令 $h(x,n)$ 为有界时空调

和函数. 鞅性质蕴含 $h(x,0)=E_xh(X_n,n)$. 若 $|h|\leqslant C$，则由耦合知

$$|h(x,0)-h(y,0)|=|Eh(X_n,n)-Eh(Y_n,n)|\leqslant 2CP(X_n\neq Y_n)\to 0$$

故 $h(x,0)$ 在 L_0 和 L_1 上为常数. 将最后结果应用于 $h'(x,m)=h(x,n+m)$，可见 $h(x,n)=a_n^i$ 在 L_i 上. 鞅性质蕴含 $a_n^i=a_{n+1}^{1-i}$，由定理 5.7.4 得所需结果. □

例题 5.7.7. Ornstein 耦合. 令 $p(x,y)=f(y-x)$ 为 $\mathbb{Z}$ 上不可约非周期随机游走的转移概率. 为证尾 σ 域平凡,取足够大的 M 使得概率分布 $f_M(x)$ 生成的随机游走不可约且非周期,其中 $f_M(x)=c_Mf(x)$ $(|x|\leqslant M)$，$f_M(x)=0$ $(|x|>M)$. 令 $Z_1,Z_2,\ldots$ 为独立同分布且服从分布 f，$W_1,W_2,\ldots$ 为独立同分布且服从分布 f_M. 令 $X_n=X_{n-1}+Z_n$ $(n\geqslant 1)$. 若 $X_{n-1}=Y_{n-1}$，我们令 $X_n=Y_n$. 否则，令

$$Y_n=\begin{cases}Y_{n-1}+Z_n & \text{若}|Z_n|>m\\ Y_{n-1}+W_n & \text{若}|Z_n|\leqslant m\end{cases}$$

换言之，大跳跃同步进行，小跳跃独立. 一维零均值随机游走的常返性蕴含 $P(X_n\neq Y_n)\to 0$. 重复定理 5.7.6 的证明，可见 $\mathcal{T}$ 是平凡的.

定理 5.7.6 中的尾 σ-域与定理 5.7.3 中的本质上相同. 为得到更有趣的 $\mathcal{T}$，我们考察：

例题 5.7.8. 树上的随机游走. 为方便定义，将该系统视为具有 3 个生成元 a,b,c 的群上的随机游走，满足 $a^2=b^2=c^2=e$（单位元）. 令 $\xi_1,\xi_2,\ldots$ 独立同分布且 $P(\xi_n=x)=1/3$ $(x=a,b,c)$，并令 $X_n=X_{n-1}\xi_n$.（这等价于每个顶点度为 3 的树上的随机游走，但代数表述便于计算.）

令 L_n 为 X_n 尽可能约化后的字长，若 $X_n=e$ 则 $L_n=0$. 约化可在进行过程中完成：若 X_{n-1} 的最后一个字母与 ξ_n 相同则擦除，否则添加新字母. 易见 L_n 是具有 $p(0,1)=1$ 的 Markov 链且

$$p(j,j-1)=1/3,\quad p(j,j+1)=2/3\quad \text{对}j\geqslant 1$$

当 $n\to\infty$ 时 $L_n\to\infty$. 由此易知字 X_n 在如下意义下存在极限：第 i 个字母 X_n^i 对大 n 保持不变. 令 X_∞^i 为极限字，即 $X_\infty^i=\lim X_n^i$. $\mathcal{T}\supset\sigma(X_\infty^i,i\geqslant 1)$，但易见这并非全部. 令 $S_0=$ 偶数长度的字，$S_1=S_0^c$，则 $X_n\in S_i$ 蕴含 $X_{n+1}\in S_{1-i}$，故 $\{X_0\in S_0\}\in\mathcal{T}$. 读者能否证明我们现已找到 $\mathcal{T}$ 的全部? 正如 Fermat 所言：“我有一个证明，但它写不进页边距.”

注释. 此次解答不涉及椭圆曲线，而是使用 “h-路径”. 见 Furstenburg (1970) 或解码以下内容：“给定出口点（无限字）条件，则所得随机游走是一个 h-过程，它以概率 $2/3$ 向边界移动，以概率 $1/3$ 远离（$1/6$ 分别对应两种可能）. 两个这样的随机游走可耦合，只要它们具有相同的奇偶性.” 这段引语来自 Robin Pemantle，他说他曾咨询 Itai Benajamini 和 Yuval Peres.

第 8 节 一般状态空间 *

本节将可数状态空间的结果推广到一类具有不可数状态空间的 Markov 链，称为 **Harris 链**. 本节的发展基于三个想法：其一，若状态空间中存在一点被链以概率 1 击中，则可数状态空间的证明适用（例如，考虑通过循环技巧构造平稳测度）. 其二，常返 Harris 链可被修改以包含这样一点. 其三，Harris 链的集合是一个良好的推广层次：既广泛到包含大量有趣的例子，又足够限制以允许丰富的理论.

若能找到集合 $A, B \in \mathcal{S}$，函数 q 满足 $q(x,y) \geqslant \varepsilon > 0$（$x \in A, y \in B$），以及集中于 B 的概率测度 ρ，使得：

(i) 若 $\tau_A = \inf\{n \geqslant 0 : X_n \in A\}$，则 $P_z(\tau_A < \infty) > 0$ 对所有 $z \in S$ 成立.

(ii) 若 $x \in A$ 且 $C \subset B$ 则 $p(x,C) \geqslant \int_C q(x,y)\,\rho(\mathrm{d}y)$.

则称 Markov 链 X_n 为 **Harris 链**.

为解释该定义，我们转向一些例子：

例题 5.8.1. 可数状态空间. 若 S 可数且存在点 a 使得 $\rho_{xa} > 0$ 对所有 x 成立（比不可约性稍弱的条件），则可取 $A = \{a\}$，$B = \{b\}$（其中 b 为任意满足 $p(a,b) > 0$ 的状态），$\mu = \delta_b$ 为 b 处的点质量，且 $q(a,b) = p(a,b)$.

例题 5.8.2. 具有连续密度的链. 设 $X_n \in \mathbb{R}^d$ 为具有转移概率 $p(x,\mathrm{d}y) = p(x,y)\,\mathrm{d}y$ 的 Markov 链，其中 $(x,y) \mapsto p(x,y)$ 连续. 选取 (x_0,y_0) 使得 $p(x_0,y_0) > 0$. 令 A 和 B 分别为 x_0 和 y_0 的足够小开邻域，使得在 $A \times B$ 上 $p(x,y) \geqslant \varepsilon > 0$. 若令 $\rho(C) = |B \cap C|/|B|$（其中 $|B|$ 为 B 的 Lebesgue 测度），则 (ii) 成立. 若 (i) 成立，则 X_n 是 Harris 链.

具体例子如下：

(a) **扩散过程**是一大类超出本书范围的例子，但重要到不能忽视. 当情况良好时，具体而言，若 X 的生成元具有满足适当增长条件的 Hölder 连续系数，见 Dynkin (1965) 的附录，则 $P_x(X_1 \in \mathrm{d}y) = p(x,y)\,\mathrm{d}y$，且 p 满足上述条件.

(b) **离散 Ornstein-Uhlenbeck 过程**. 令 $\xi_1, \xi_2, \ldots$ 为独立同分布标准正态，令 $V_n = \theta V_{n-1} + \xi_n$. Ornstein-Uhlenbeck 过程是扩散过程 $\{V_t, t \in [0,\infty)\}$，它模拟悬浮在液体中粒子的速度. 见例如 Breiman (1968) 第 16.1 节. 在整数时刻观察 V_t（并除以常数使方差为 1）给出具有所述分布的 Markov 链.

例题 5.8.3. GI/G/1 排队系统，或存储模型. 令 $\xi_1, \xi_2, \ldots$ 独立同分布，并归纳定义 $W_n = (W_{n-1} + \xi_n)^+$. 若 $P(\xi_n < 0) > 0$ 则我们可取 $A = B = \{0\}$ 且 (i) 和 (ii) 成立. 为解释标题中的第一个名字，考虑一个排队系统，其中顾客按更新过程到达，即在时刻 $0 = T_0 < T_1 < T_2 \ldots$ 到达，其中 $\zeta_n = T_n - T_{n-1}$ $(n \geqslant 1)$ 独立同分布. 令 η_n $(n \geqslant 0)$ 为第 n 个顾客所需的服务时间，并令 $\xi_n = \eta_{n-1} - \zeta_n$. 断言：$W_n$ 是第 n 个顾客进入服务前必须等待的时间. 为说明这一点，注意第 $(n-1)$ 个顾客给服务器增加了 η_{n-1} 的工作量，且若服务器在 $[T_{n-1}, T_n)$ 中所有时刻都忙，其工作量减少了 ζ_n. 若 $W_{n-1} + \eta_{n-1} < \zeta_n$ 则服务器有足够时间完成工作，下一位到达的顾客将发现队列为空.

标题中的第二个名字指 W_n 可用于模拟存储设施的内容. 为直观描述，考虑水库. 我们假设暴雨在更新过程 $\{T_n : n \geqslant 1\}$ 的时刻发生，第 n 场暴雨贡献水量 η_n，且水以恒定速率 c 消耗. 若令 $\zeta_n = T_n - T_{n-1}$ 如前，且 $\xi_n = \eta_{n-1} - c\zeta_n$，则 W_n 给出第 n 场暴雨前水库中的水量.

历史回顾. Doeblin 率先证明了关于一般状态空间 Markov 链的结果. 他假设存在 n 使得 $p^n(x, C) \geqslant \varepsilon\rho(C)$ 对所有 $x \in S$ 和 $C \subset S$ 成立. 见 Doob (1953) 第 V.5 节对他结果的叙述. Harris (1956) 推广了 Doeblin 的结果，观察到只要有一个集合 A 使得 (i) 成立，且在 A 上观察的链（$Y_k = X(T_A^k)$，其中 $T_A^k = \inf\{n > T_A^{k-1} : X_n \in A\}$ 且 $T_A^0 = 0$）满足 Doeblin 条件即可. 我们的表述以及本节的大部分证明遵循 Athreya 和 Ney (1978). 对 “传统方法” 的很好描述，见 Revuz (1984).

给定 $(S, \mathcal{S})$ 上的 Harris 链，下面构造一个具有转移概率 $\bar{p}$ 的 Markov 链 $\bar{X}_n$，其状态空间为 $(\bar{S}, \bar{\mathcal{S}})$，其中 $\bar{S} = S \cup \{\alpha\}$，$\bar{\mathcal{S}} = \{B, B \cup \{\alpha\} : B \in \mathcal{S}\}$. 目的如前所述，是构造一个点 α 使得过程在常返情形下以概率 1 击中它.

$$\bar{p}(x, C) = \begin{cases} p(x, C) & \text{对} C \in \mathcal{S} \text{ 若} x \in S - A \\ \varepsilon & \text{若} x \in A \text{ 且} C = \{\alpha\} \\ p(x, C) - \varepsilon\rho(C) & \text{对} C \in \mathcal{S} \text{ 若} x \in A \end{cases}$$

$$\bar{p}(\alpha, D) = \int \rho(\mathrm{d}x)\bar{p}(x, D) \quad \text{对} D \in \bar{\mathcal{S}}$$

直观上，$\bar{X}_n = \alpha$ 对应于 X_n 按 ρ 分布在 B 上. 这里及以下，保留 A 和 B 用于定义中出现的特殊集合，用 C 和 D 表示 $\mathcal{S}$ 的一般元素. 常通过写 $\bar{p}(x, \alpha)$ 代替 $\bar{p}(x, \{\alpha\})$，$\bar{\mu}(\alpha)$ 代替 $\bar{\mu}(\{\alpha\})$ 等来简化记号.

下面证明三个技术引理，为后续理论发展做准备. 定义转移概率 v 为

$$v(x, \{x\}) = 1 \text{ 若} x \in S, \quad v(\alpha, C) = \rho(C)$$

换言之，V 保持 S 中的质量不变，但将 α 处的质量返回 S 并按 ρ 分布.

引理 5.8.4. $v\bar{p} = \bar{p}$ 且 $\bar{p}v = p$.

证明. 在证明之前，提醒读者测度左乘转移概率，即第一种情形要证 $\mu v\bar{p} = \mu\bar{p}$. 若先按 v 转移再按 $\bar{p}$ 转移，这相当于一次按 $\bar{p}$ 转移，因为只有 α 处的质量受 v 影响，且

$$\bar{p}(\alpha, D) = \int \rho(\mathrm{d}x)\bar{p}(x, D)$$

第二个等式也由定义易得. 换言之，若 $\bar{p}$ 先作用然后 v 作用，则 v 将 α 处的质量返回原处. □

由引理 5.8.4 立得:

引理 5.8.5. 令 Y_n 为具有 $p_{2k}=v$ 和 $p_{2k+1}=\bar{p}$ 的非齐次 Markov 链. 则 $\bar{X}_n=Y_{2n}$ 是具有转移概率 $\bar{p}$ 的 Markov 链，且 $X_n=Y_{2n+1}$ 是具有转移概率 p 的 Markov 链.

引理 5.8.5 表明 X_n 与 $\bar{X}_n$ 的渐近行为密切相关. 为量化这一点，需要如下定义. 若 f 是 S 上的有界可测函数，令 $\bar{f}=vf$，即 $\bar{f}(x)=f(x)$（$x\in S$）且 $\bar{f}(\alpha)=\int f\,\mathrm{d}\rho$.

引理 5.8.6. 若 μ 是 $(S,\mathcal{S})$ 上的概率测度，则

$$E_\mu f(X_n)=E_\mu \bar{f}(\bar{X}_n)$$

证明. 注意若 X_n 和 $\bar{X}_n$ 如引理 5.8.5 中构造，且 $P(\bar{X}_0\in S)=1$，则 $X_0=\bar{X}_0$ 且 X_n 由 $\bar{X}_n$ 经按 v 转移得到. □

最后三个引理使得可从 $\bar{X}_n$ 的结果得到 X_n 的结果. 我们现在转向将可数状态空间的结果推广到 $\bar{X}_n$ 的任务. 在发展理论之前，我们将给出一个最后的例子，它解释了为什么某些表述会很复杂.

例题 5.8.7. 病态 O.U. 过程. 取例 5.8.2(b) 中的离散 O.U. 过程并修改整数 $x\geqslant 2$ 处的转移概率使得

$$p(x,\{x+1\})=1-x^{-2},\quad p(x,A)=x^{-2}|A|\ \text{对}A\subset(0,1)$$

p 是 Harris 链的转移概率，但

$$P_2(X_n=n+2\ \text{对所有}n)>0$$

可以理解那些认为这种链不会 “在应用中出现” 的读者，但似乎更容易（也更好）调整理论以包含它们，而非修改假设以排除它们.

5.8.1 常返与暂态

下面从常返与暂态的二分法开始. 令 $R=\inf\{n\geqslant 1:\bar{X}_n=\alpha\}$. 若 $P_\alpha(R<\infty)=1$ 则称链**常返**，否则称**暂态**. 令 $R_1=R$，且对 $k\geqslant 2$，令 $R_k=\inf\{n>R_{k-1}:\bar{X}_n=\alpha\}$ 为第 k 次返回 α 的时刻. 强 Markov 性质蕴含 $P_\alpha(R_k<\infty)=P_\alpha(R<\infty)^k$，故在常返情形 $P_\alpha(\bar{X}_n=\alpha\ \text{i.o.})=1$，在暂态情形 $=0$.

下述结果推广了引理 5.3.2（若 x 常返且 $\rho_{xy}>0$，则 y 常返且 $\rho_{yx}=1$）.

定理 5.8.8. 令 $\lambda(C)=\sum_{n=1}^{\infty}2^{-n}\bar{p}^n(\alpha,C)$. 在常返情形，若 $\lambda(C)>0$ 则 $P_\alpha(\bar{X}_n\in C\ \text{i.o.})=1$. 对 λ-a.e. x，$P_x(R<\infty)=1$.

证明. 第一个结论由引理推出. 对第二个结论，令 $D=\{x:P_x(R<\infty)<1\}$ 并注意若对某个 n 有 $\bar{p}^n(\alpha,D)>0$，则

$$P_\alpha(\bar{X}_m=\alpha\ \text{i.o.})\leqslant\int\bar{p}^n(\alpha,\mathrm{d}x)P_x(R<\infty)<1$$

□

注释. 例 5.8.7 表明我们不能期望对所有 x 有 $P_x(R<\infty)=1$. 为说明即使状态空间可数，我们也未必从 α 出发击中每一点，见习题 5.8.2.

5.8.2 平稳测度

定理 5.8.9. 在常返情形，存在 σ 有限平稳测度 $\bar{\mu}\ll\lambda$.

证明. 令 $R=\inf\{n\geqslant 1:\bar{X}_n=\alpha\}$，并令

$$\bar{\mu}(C)=E_\alpha\sum_{n=0}^{R-1}1_{\{\bar{X}_n\in C\}}=\sum_{n=0}^{\infty}P_\alpha(\bar{X}_n\in C,R>n)$$

重复定理 5.5.7 的证明知 $\bar{\mu}\bar{p}=\bar{\mu}$. 令 $\mu=\bar{\mu}v$，则由引理 5.8.4 知 $\mu\bar{p}=\bar{\mu}\bar{p}v=\bar{\mu}v$，故 $\mu p=\mu$.

为证其 σ 有限，令 $G_{k,\delta}=\{x:\bar{p}^k(x,\alpha)\geqslant\delta\}$. 令 $T_1=\inf\{n:X_n\in G_{k,\delta}\}$，且对 $m\geqslant 2$，令 $T_m=\inf\{n>T_{m-1}:X_n\in G_{k,\delta}\}$，并令 $M=\sup\{m:T_m<R\}$. 由假设 $EM<1/\delta$. 在 T_m 和 T_{m-1} 之间最多有 k 次访问 $G_{k,\delta}$. $\bar{\mu}\ll\lambda$ 由比较定义显然. □

为研究平稳测度的唯一性，从以下引理开始：

引理 5.8.10. 若 ν 是 p 的 σ-有限平稳测度，则 $\nu(A)<\infty$ 且 $\bar{\nu}=\nu\bar{p}$ 是 $\bar{p}$ 的平稳测度满足 $\bar{\nu}(\alpha)<\infty$.

证明. 首先证 $\nu(A)<\infty$. 若 $\nu(A)=\infty$，则定义中 (ii) 蕴含对所有满足 $\rho(C)>0$ 的集合 C 有 $\nu(C)=\infty$. 若 $B=\bigcup_i B_i$ 满足 $\nu(B_i)<\infty$，则由最后观察知 $\rho(B_i)=0$，且由可数次可加性知 $\rho(B)=0$，矛盾. 故 $\nu(A)<\infty$ 且 $\bar{\nu}(\alpha)=\nu\bar{p}(\alpha)=\varepsilon\nu(A)<\infty$. 利用 $\nu p=\nu$，我们求得

$$\nu\bar{p}(C)=\nu(C)-\varepsilon\nu(A)\rho(B\cap C)$$

最后减法良定义，因 $\nu(A)<\infty$；由此知 $\bar{\nu}v=\nu$. 为验证 $\bar{\nu}\bar{p}=\bar{\nu}$，注意引理 5.8.4 和最后结果蕴含 $\bar{\nu}\bar{p}=\bar{\nu}v\bar{p}=\nu\bar{p}=\bar{\nu}$. □

定理 5.8.11. 假设 p 常返. 若 ν 是 σ-有限平稳测度，则 $\nu=\bar{\nu}(\alpha)\mu$，其中 μ 是定理 5.8.9 证明中构造的测度.

证明. 由引理 5.8.10，只需证明若 $\bar{\nu}$ 是 $\bar{p}$ 的平稳测度满足 $\bar{\nu}(\alpha)<\infty$，则 $\bar{\nu}=\bar{\nu}(\alpha)\bar{\mu}$. 重复定理 5.5.9 的证明并以 $a=\alpha$，易证 $\bar{\nu}(C)\geqslant\bar{\nu}(\alpha)\bar{\mu}(C)$. 继续如该证明中计算：

$$\bar{\nu}(\alpha)=\int\bar{\nu}(\mathrm{d}x)\bar{p}^n(x,\alpha)\geqslant\bar{\nu}(\alpha)\int\bar{\mu}(\mathrm{d}x)\bar{p}^n(x,\alpha)=\bar{\nu}(\alpha)\bar{\mu}(\alpha)=\bar{\nu}(\alpha)$$

令 $S_n=\{x:p^n(x,\alpha)>0\}$. 由假设，$\bigcup_n S_n=S$. 若对某个 D 有 $\bar{\nu}(D)>\bar{\nu}(\alpha)\bar{\mu}(D)$，则 $\bar{\nu}(D\cap S_n)>\bar{\nu}(\alpha)\bar{\mu}(D\cap S_n)$，由此知 $\bar{\nu}(\alpha)>\bar{\nu}(\alpha)$，矛盾. □

5.8.3 收敛定理

若 g.c.d.$\{n \geqslant 1 : p^n(\alpha,\alpha) > 0\} = 1$，则称常返 Harris 链 X_n **非周期**. 例如，若定义中可取 $A = B$ 则发生这种情况，因这时 $p(\alpha,\alpha) > 0$.

定理 5.8.12. 设 X_n 为具有平稳分布 π 的非周期常返 Harris 链. 若 $P_x(R < \infty) = 1$，则当 $n \to \infty$ 时，

$$\|p^n(x,\cdot) - \pi(\cdot)\| \to 0$$

注. 这里 $\|\cdot\|$ 表示测度间的全变差距离. 引理 5.8.8 保证对 π-a.e. x 假设成立.

证明. 鉴于引理 5.8.6，只需对 $\bar{p}$ 证明该结果. 首先注意，平稳概率测度的存在性及定理 5.8.11 中的唯一性结果蕴含定理 5.8.9 中构造的测度满足 $E_\alpha R = \bar{\mu}(S) < \infty$. 如定理 5.6.6 的证明，令 X_n 和 Y_n 为具有初始分布 δ_x 和 π 的链的独立复制，并令 $\tau = \inf\{n \geqslant 0 : X_n = Y_n = \alpha\}$. 对 $m \geqslant 0$，令 S_m（相应地 T_m）为 X_n（相应地 Y_n）第 $(m+1)$ 次访问 α 的时刻（$S_0 = T_0 = 0$）. $S_m - T_m$ 是零均值随机游走，故 $M = \inf\{m \geqslant 1 : S_m = T_m\} < \infty$ a.s.，由此知 $\tau < \infty$ a.s.. 定理 5.6.6 证明中的计算表明 $|P(X_n \in C) - P(Y_n \in C)| \leqslant P(\tau > n)$ 这对所有 C 成立，故 $\|p^n(x,\cdot) - \pi(\cdot)\| \leqslant P(\tau > n)$，证毕. □

5.8.4 GI/G/1 排队系统

本节剩余部分集中讨论 GI/G/1 排队系统. 令 $\xi_1, \xi_2, \ldots$ 为独立同分布，令 $W_n = (W_{n-1} + \xi_n)^+$，并令 $S_n = \xi_1 + \cdots + \xi_n$. 回忆 $\xi_n = \eta_{n-1} - \zeta_n$，其中 η 为服务时间，ζ 为到达间隔时间，并设 $E\xi_n < 0$ 使得习题 5.8.5 蕴含存在平稳分布.

一般而言，难以获得 M 分布的显式公式. 然而，若到达或服务分布为指数分布，则可以得到. 原因之一如下：

引理 5.8.13. 设 $X, Y \geqslant 0$ 独立且 $P(X > x) = e^{-\lambda x}$. 证明 $P(X - Y > x) = ae^{-\lambda x}$，其中 $a = P(X - Y > 0)$.

证明. 令 F 为 Y 的分布

$$P(X - Y > x) = \int_0^\infty \mathrm{d}F(y) e^{-\lambda(x+y)} = e^{-\lambda x} \int_0^\infty \mathrm{d}F(y) e^{-\lambda y} = e^{-\lambda x} P(X - Y > 0)$$

其中最后等式由在第一式中令 $x = 0$ 得到. □

例题 5.8.14. 指数服务时间. 设 $P(\eta_n > x) = e^{-\beta x}$ 且 $E\zeta_n > E\eta_n$. 令 $T = \inf\{n : S_n > 0\}$，$L = S_T$；若 $T = \infty$ 则令 $L = -\infty$. 指数分布的无记忆性蕴含 $P(L > x) = re^{-\beta x}$，其中 $r = P(T < \infty)$

为计算最大值 M 的分布，令 $T_1 = T$，且对 $k \geqslant 2$，令 $T_k = \inf\{n > T_{k-1} : S_n > S_{T_{k-1}}\}$. 强 Markov 性质蕴含若 $T_k < \infty$，则 $S(T_{k+1}) - S(T_k) \stackrel{d}{=} L$ 且与 $S(T_k)$ 独立. 利用这一点并按 $K = \inf\{k : L_{k+1} = -\infty\}$ 的值分解，可见对 $x > 0$ 密度函数

$$P(M = x) = \sum_{k=1}^{\infty} r^k(1-r)e^{-\beta x}\beta^k x^{k-1}/(k-1)! = \beta r(1-r)e^{-\beta x(1-r)}$$

为完成计算，需要计算 r. 为此令

$$\varphi(\theta) = E\exp(\theta\xi_n) = E\exp(\theta\eta_{n-1})E\exp(-\theta\zeta_n)$$

它对 $0 < \theta < \beta$ 有限，因 $\zeta_n \geqslant 0$ 且 η_{n-1} 具有指数分布. 易见 $\varphi'(0) = E\xi_n < 0$ 且 $\lim_{\theta\uparrow\beta}\varphi(\theta) = \infty$，故存在 $\theta \in (0,\beta)$ 使 $\varphi(\theta) = 1$. 例 4.2.3 知 $\exp(\theta S_n)$ 是鞅. 定理 4.4.1 知 $1 = E\exp(\theta S_{T\wedge n})$. 令 $n \to \infty$，注意 $(S_n \mid T = n)$ 具有指数分布，且 $S_n \to -\infty$ 在 $\{T = \infty\}$ 上，有

$$1 = r\int_0^{\infty} e^{\theta x}\beta e^{-\beta x}\,\mathrm{d}x = \frac{r\beta}{\beta - \theta}$$

例题 5.8.15. Poisson 到达. 设 $P(\zeta_n > x) = e^{-\alpha x}$ 且 $E\zeta_n > E\eta_n$. 令 $\bar{S}_n = -S_n$. 如习题 5.8.6(ii) 中反转时间，可见（对 $n \geqslant 1$）

$$P\left(\max_{0\leqslant k<n}\bar{S}_k < \bar{S}_n \in A\right) = P\left(\min_{1\leqslant k\leqslant n}\bar{S}_k > 0, \bar{S}_n \in A\right)$$

令 $\psi_n(A)$ 为最后两式的共同值，令 $\psi(A) = \sum_{n\geqslant 0}\psi_n(A)$. $\psi_n(A)$ 是随机游走在时刻 n 于 A 中到达新最大值（或阶梯点，见习题 5.4.2）的概率，故 $\psi(A)$ 是 A 中阶梯点的期望个数且 $\psi(\{0\}) = 1$. 令随机游走再走一步，

$$P\left(\min_{1\leqslant k\leqslant n}\bar{S}_k > 0, \bar{S}_{n+1} \leqslant x\right) = \int F(x-z)\,\mathrm{d}\psi_n(z)$$

最后恒等式对 $n = 0$ 亦成立，若将左边解释为 $F(x)$. 令 $\tau = \inf\{n \geqslant 1 : \bar{S}_n \leqslant 0\}$，$x \leqslant 0$. 对右边分部积分，再对 $n \geqslant 0$ 求和得

$$P(\bar{S}_\tau \leqslant x) = \sum_{n=0}^{\infty} P\left(\min_{1\leqslant k\leqslant n}\bar{S}_k > 0, \bar{S}_{n+1} \leqslant x\right) = \int_{y\leqslant x}\psi[0, x-y]\,\mathrm{d}F(y)$$

极限 $y \leqslant x$ 源于 $\psi((-\infty, 0)) = 0$.

令 $\bar{\xi}_n = \bar{S}_n - \bar{S}_{n-1} = -\xi_n$. 习题 5.8.13 知 $P(\bar{\xi}_n > x) = ae^{-\alpha x}$. 令 $\bar{T} = \inf\{n : \bar{S}_n > 0\}$. $E\bar{\xi}_n > 0$，故 $P(\bar{T} < \infty) = 1$. 令 $J = \bar{S}_{\bar{T}}$. 如上例知 $P(J > x) = e^{-\alpha x}$. 令 $V_n = J_1 + \cdots + J_n$. V_n 是速率为 α 的

Poisson 过程，故对 $x-y\geqslant 0$ 有 $\psi[0,x-y]=1+\alpha(x-y)$. 现利用 (5.8.1) 并分部积分得

$$P(\bar{S}_\tau\leqslant x)=\int_{y\leqslant x}(1+\alpha(x-y))\,\mathrm{d}F(y)=F(x)+\alpha\int_{-\infty}^{x}F(y)\,\mathrm{d}y\quad \text{对}x\leqslant 0$$

由于对 $n\geqslant 1$ 有 $P(\bar{S}_n=0)=0$，故 $-S_\tau$ 与 S_T 同分布，其中 $T=\inf\{n:S_n>0\}$. 结合习题 5.8.6(ii) 即得 $P(M>x)$ 的 “公式”. 直接但略显繁琐的计算表明：若 $B(s)=E\exp(-s\eta_n)$，则

$$E\exp(-sM)=\frac{(1-\alpha\cdot E\eta)s}{s-\alpha+\alpha B(s)}$$

这一结果称为 **Pollaczek-Khintchine 公式**. 我们省略的计算可在 Billingsley (1979) 第 277 页或 Feller, Vol. II (1971) 中找到.

习题

习题 5.8.1. $\bar{X}_n$ 常返当且仅当 $\sum_{n=1}^{\infty}\bar{p}^n(\alpha,\alpha)=\infty$.

习题 5.8.2. 若 X_n 是可数状态空间上的常返 Harris 链，则 S 只能有一个不可约常返类，但可能有非空暂态集. 具体例子：考虑分支过程，其中无后代概率 $p_0>0$，并令 $A=B=\{0\}$.

习题 5.8.3. 利用习题 5.8.1 并模仿定理 5.5.10 的证明，证明具有平稳分布的 Harris 链必为常返.

习题 5.8.4. 设 X_n 为常返 Harris 链. 证明：若 (A',B') 是满足定义条件的另一对集合，则定理 5.8.8 蕴含 $P_\alpha(\bar{X}_n\in A'\text{ i.o.})=1$，故常返或暂态不依赖于 (A,B) 的选择.

习题 5.8.5. 在 GI/G/1 排队系统中，等待时间 W_n 与随机游走 $S_n=X_0+\xi_1+\cdots+\xi_n$ 在 $N=\inf\{n:S_n<0\}$ 前一致，且此时 $W_N=0$. 如例 5.3.7 那样利用这一观察，证明当 $E\xi_n>0$ 时例 5.8.3 暂态，当 $E\xi_n\leqslant 0$ 时常返，且当 $E\xi_n<0$ 时存在平稳分布.

习题 5.8.6. 令 $m_n=\min(S_0,S_1,\ldots,S_n)$，其中 S_n 为如上定义的随机游走.

(i) 证明 $S_n-m_n\overset{d}{=}W_n$.

(ii) 令 $\xi'_m=\xi_{n+1-m}$ （$1\leqslant m\leqslant n$）. 证明 $S_n-m_n=\max(S'_0,S'_1,\ldots,S'_n)$.

(iii) 推断当 $n\to\infty$ 时 $W_n\Rightarrow M\equiv\max(S'_0,S'_1,S'_2,\ldots)$.

习题 5.8.7. 在离散 O.U. 过程中，X_{n+1} 是均值为 θX_n 方差为 1 的正态分布. 若 Y_{n+1} 改为均值为 0 方差为 $\beta^2|Y_n|$ 的正态分布，常返和暂态会发生什么变化?

第六章
遍历定理

称 X_n（$n \geqslant 0$）为**平稳序列**，如果对每个 $k \geqslant 1$，平移序列 X_{n+k}（$n \geqslant 0$）具有相同的分布. 关于这类序列的一个基本事实称为遍历定理：若 $E|f(X_0)| < \infty$，则

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} f(X_m) \text{ 几乎必然存在.}$$

若 X_n 是遍历的（这是 Markov 链不可约性概念的推广），则极限为 $Ef(X_0)$. 6.1 节和 6.2 节发展证明遍历定理所需要的理论. 在 6.3 节中，我们应用遍历定理研究具有平稳序列增量的随机游走的常返性，得到了关于独立同分布情形的惊人推广. 在 6.4 节中，我们证明一个次可加遍历定理. 6.4 节和 6.5 节中的例子将表明，这是遍历定理的一个有用推广.

第 1 节　定义与例子

称 $X_0, X_1, \ldots$ 为**平稳序列**，如果对任意 k，平移序列 $\{X_{k+n}, n \geqslant 0\}$ 具有相同的分布，即对每个 m，

$$(X_0, \ldots, X_m) \text{ 与} (X_k, \ldots, X_{k+m}) \text{ 同分布.}$$

首先给出四个例子，它们将作为本章的主要例子.

例题 6.1.1. $X_0, X_1, \ldots$ 是独立同分布的随机变量.

例题 6.1.2. 设 X_n 是具有转移概率 $p(x, A)$ 和平稳分布 π 的 Markov 链，即 $\pi(A) = \int \pi(\mathrm{d}x)p(x, A)$. 若 X_0 具有分布 π，则 $X_0, X_1, \ldots$ 是平稳序列. 一个需要牢记的特殊反例是状态空间 $S = \{0, 1\}$ 且转移概率 $p(x, \{1 - x\}) = 1$ 的链. 此时平稳分布为 $\pi(0) = \pi(1) = 1/2$，且 $(X_0, X_1, \ldots)$ 以概率 $1/2$ 等于 $(0, 1, 0, 1, \ldots)$ 或 $(1, 0, 1, 0, \ldots)$.

例题 6.1.3. 圆周旋转. 设 $\Omega=[0,1)$，$\mathcal{F}=$ Borel 子集，$P=$ Lebesgue 测度. 令 $\theta\in(0,1)$，且对 $n\geqslant 0$，令 $X_n(\omega)=(\omega+n\theta) \mod 1$，其中 $x \mod 1=x-[x]$，$[x]$ 表示不超过 x 的最大整数. 为理解这一名称的由来，将 $[0,1)$ 通过 $x\mapsto \exp(2\pi i x)$ 映射到 $\mathbb{C}$ 中. 这个例子是上一个例子的特殊情形. 令 $p(x,\{y\})=1$ 当 $y=(x+\theta) \mod 1$.

最后一个例子不仅重要，而且涵盖了所有情形!

例题 6.1.4. 设 $(\Omega,\mathcal{F},P)$ 为概率空间. 称可测映射 $\varphi:\Omega\to\Omega$ 为**保测变换**，如果对任意 $A\in\mathcal{F}$，有 $P(\varphi^{-1}A)=P(A)$. 令 φ^n 表示 φ 的**第 n 次迭代**，递归地定义为 $\varphi^n=\varphi(\varphi^{n-1})$ $(n\geqslant 1)$，其中 $\varphi^0(\omega)=\omega$. 断言：若 $X\in\mathcal{F}$，则 $X_n(\omega)=X(\varphi^n\omega)$ 定义了一个平稳序列. 为验证这一点，令 $B\in\mathcal{R}^{n+1}$ 且 $A=\{\omega:(X_0(\omega),\ldots,X_n(\omega))\in B\}$. 则

$$P((X_k,\ldots,X_{k+n})\in B)=P(\varphi^k\omega\in A)=P(\omega\in A)=P((X_0,\ldots,X_n)\in B).$$

最后一个例子不仅是一个重要的例子. 事实上，它是唯一的一个例子! 若 $Y_0,Y_1,\ldots$ 是取值于一个好空间中的平稳序列，则 Kolmogorov 扩张定理（定理 A.3.1）允许在序列空间 $(S^{\{0,1,\ldots\}},\mathcal{S}^{\{0,1,\ldots\}})$ 上构造一个测度 P，使得序列 $X_n(\omega)=\omega_n$ 与 $\{Y_n,n\geqslant 0\}$ 具有相同的分布. 若令 φ 为**移位算子**，即 $\varphi(\omega_0,\omega_1,\ldots)=(\omega_1,\omega_2,\ldots)$，且令 $X(\omega)=\omega_0$，则 φ 是保测变换且 $X_n(\omega)=X(\varphi^n\omega)$.

鉴于上述观察，只需在例 6.1.4 的框架下给出定义并证明结果. 因此，基本设定为

$$\begin{aligned}&(\Omega,\mathcal{F},P)\text{ 是一个概率空间,}\\&\varphi\text{ 是一个保测映射,}\\&X_n(\omega)=X(\varphi^n\omega)\text{，其中}X\text{是一个随机变量.}\end{aligned}$$

下面给出一些重要的定义. 这里及以下都假设 φ 是保测变换. 称集合 $A\in\mathcal{F}$ 为**不变的**，如果 $\varphi^{-1}A=A$.（这里，如常，若两个集合的对称差概率为 0，则认为它们相等.）令 $\mathcal{I}$ 为不变事件的集合. 在习题 6.1.1 中，读者将证明 $\mathcal{I}$ 是一个 σ-域.

称 $(\Omega,\mathcal{F},P)$ 上的保测变换为**遍历的**，如果 $\mathcal{I}$ 是平凡的，即对任意 $A\in\mathcal{I}$，$P(A)\in\{0,1\}$. 若 φ 不是遍历的，则空间可分解为两个集合 A 和 A^c，每个都具有正测度，使得 $\varphi(A)=A$ 且 $\varphi(A^c)=A^c$. 换言之，φ 不是不可约的.

为深入探讨遍历性的含义，下面回到例子中，重新编号是因为新的重点是验证遍历性.

例题 6.1.5. 独立同分布序列. 首先注意到，若 $\Omega=\mathbb{R}^{\{0,1,\ldots\}}$ 且 φ 是移位算子，则不变集 A 满足 $\{\omega:\omega\in A\}=\{\omega:\varphi\omega\in A\}\in\sigma(X_1,X_2,\ldots)$. 迭代得

$$A\in\bigcap_{n=1}^{\infty}\sigma(X_n,X_{n+1},\ldots)=\mathcal{T}\text{，尾 }\sigma\text{-域,}$$

故 $\mathcal{I} \subset \mathcal{T}$. 对独立同分布序列，Kolmogorov 的 0-1 律蕴含 $\mathcal{T}$ 是平凡的，故 $\mathcal{I}$ 是平凡的，序列是遍历的（即当对应的测度被置于序列空间 $\Omega = \mathbb{R}^{\{0,1,2,\ldots\}}$ 上时，移位是遍历的）.

例题 6.1.6. Markov 链. 假设状态空间 S 是可数的，且平稳分布对所有 $x \in S$ 满足 $\pi(x) > 0$. 由定理 5.5.10 和 5.3.5，所有状态都是常返的，且可将 S 写成 $S = \cup_i R_i$，其中 R_i 是不交不可约闭集. 若 $X_0 \in R_i$，则以概率一，$X_n \in R_i$ 对所有 $n \geqslant 1$ 成立，故 $\{\omega : X_0(\omega) \in R_i\} \in \mathcal{I}$. 由此可知，若 Markov 链不是不可约的，则序列不是遍历的. 为证明逆命题，注意到若 $A \in \mathcal{I}$，则 $1_A \circ \theta_n = 1_A$，其中 $\theta_n(\omega_0, \omega_1, \ldots) = (\omega_n, \omega_{n+1}, \ldots)$. 因此，若令 $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$，则 1_A 的移位不变性和 Markov 性质蕴含

$$E_\pi(1_A|\mathcal{F}_n) = E_\pi(1_A \circ \theta_n|\mathcal{F}_n) = h(X_n),$$

其中 $h(x) = E_x 1_A$. Lévy 的 0-1 律蕴含左边当 $n \to \infty$ 时收敛到 1_A. 若 X_n 是不可约且常返的，则对任意 $y \in S$，右边 $= h(y)$ 无穷多次成立，故要么 $h(x) \equiv 0$，要么 $h(x) \equiv 1$，且 $P_\pi(A) \in \{0, 1\}$. 这个例子也说明 $\mathcal{I}$ 和 $\mathcal{T}$ 可以不同. 当转移概率 p 是不可约的时，$\mathcal{I}$ 是平凡的，但若所有状态具有周期 $d > 1$，则 $\mathcal{T}$ 不是平凡的. 在定理 5.7.3 中，我们证明了若 $S_0, \ldots, S_{d-1}$ 是 S 的循环分解，则 $\mathcal{T} = \sigma(\{X_0 \in S_r\} : 0 \leqslant r < d)$.

例题 6.1.7. 圆周旋转. 若 $\theta = m/n$，其中 $m < n$ 是正整数，则圆周旋转不是遍历的. 若 B 是 $[0, 1/n)$ 的 Borel 子集且

$$A = \bigcup_{k=0}^{n-1} (B + k/n),$$

则 A 是不变的. 反之，若 θ 是无理数，则 φ 是遍历的. 为证明这一点，需要 Fourier 分析的一个事实. 若 f 是 $[0, 1)$ 上的可测函数且 $\int f^2(x)\,\mathrm{d}x < \infty$，则 f 可写成 $f(x) = \sum_k c_k e^{2\pi i k x}$，其中等式含义为当 $K \to \infty$ 时

$$\sum_{k=-K}^{K} c_k e^{2\pi i k x} \to f(x) \text{ 在} L^2[0, 1) \text{ 中},$$

且系数 $c_k = \int f(x) e^{-2\pi i k x}\,\mathrm{d}x$ 是唯一确定的. 现

$$f(\varphi(x)) = \sum_k c_k e^{2\pi i k(x+\theta)} = \sum_k (c_k e^{2\pi i k\theta}) e^{2\pi i k x}.$$

系数的唯一性蕴含 $f(\varphi(x)) = f(x)$ 当且仅当 $c_k(e^{2\pi i k\theta} - 1) = 0$. 若 θ 是无理数，这蕴含 $c_k = 0$ 对 $k \neq 0$ 成立，故 f 是常数. 将上述结果应用于 $f = 1_A$ $(A \in \mathcal{I})$，表明 $A = \emptyset$ 或 $[0, 1)$ 几乎必然.

习题

习题 6.1.1. 证明不变事件类 $\mathcal{I}$ 是一个 σ-域，且 $X \in \mathcal{I}$ 当且仅当 X 是不变的，即 $X \circ \varphi = X$ 几乎必然.

习题 6.1.2. 一些作者称 A 为几乎不变的，如果 $P(A\triangle\varphi^{-1}(A))=0$，并称 C 为严格意义下不变的，如果 $C=\varphi^{-1}(C)$.

(i) 令 A 为任意集合，$B=\cup_{n=0}^{\infty}\varphi^{-n}(A)$. 证明 $\varphi^{-1}(B)\subset B$.

(ii) 令 B 为任意满足 $\varphi^{-1}(B)\subset B$ 的集合，$C=\cap_{n=0}^{\infty}\varphi^{-n}(B)$. 证明 $\varphi^{-1}(C)=C$.

(iii) 证明 A 几乎不变当且仅当存在严格意义下不变的 C 使得 $P(A\triangle C)=0$.

习题 6.1.3. 圆周旋转遍历性的直接证明.

(i) 证明若 θ 是无理数，则 $x_n=n\theta \mod 1$ 在 $[0,1)$ 中稠密. 提示：所有 x_n 互不相同，故对任意 $N<\infty$，存在 $m<n\leqslant N$ 使得 $|x_n-x_m|\leqslant 1/N$.

(ii) 利用习题 A.2.1 证明若 A 是 Borel 集且 $|A|>0$，则对任意 $\delta>0$ 存在区间 $J=[a,b)$ 使得 $|A\cap J|>(1-\delta)|J|$.

(iii) 结合 (i) 和 (ii) 得出 $P(A)=1$.

习题 6.1.4. 任何平稳序列 $\{X_n, n\geqslant 0\}$ 都可以嵌入到一个双边平稳序列 $\{Y_n : n\in\mathbb{Z}\}$ 中.

习题 6.1.5. 若 $X_0, X_1, \ldots$ 是平稳序列且 $g:\mathbb{R}^{\{0,1,\ldots\}}\to\mathbb{R}$ 可测，则 $Y_k=g(X_k, X_{k+1}, \ldots)$ 是平稳序列. 若 X_n 是遍历的，则 Y_n 也是遍历的.

习题 6.1.6. 独立块. 设 $X_1, X_2, \ldots$ 是平稳序列. 令 $n<\infty$ 且 $Y_1, Y_2, \ldots$ 是一个序列，使得这个新的序列 $(Y_{nk+1}, \ldots, Y_{n(k+1)})$ $(k\geqslant 0)$ 是独立同分布的，且 $(Y_1, \ldots, Y_n)=(X_1, \ldots, X_n)$. 最后，令 ν 为 $\{1, 2, \ldots, n\}$ 上的均匀分布，独立于 Y，且 $Z_m=Y_{\nu+m}$ 对 $m\geqslant 1$. 证明 Z_m 是平稳且遍历的.

习题 6.1.7. 连分数. 令 $\varphi(x)=1/x-[1/x]$ 对 $x\in(0,1)$，且 $A(x)=[1/x]$，其中 $[1/x]$ 表示不超过 $1/x$ 的最大整数. $a_n=A(\varphi^n x)$ $(n=0,1,2,\ldots)$ 给出 x 的连分数展开，即

$$x=1/(a_0+1/(a_1+1/(a_2+1/\ldots)))$$

证明 φ 保持测度 $\mu(A)=\frac{1}{\log 2}\int_A\frac{\mathrm{d}x}{1+x}$ 对 $A\subset(0,1)$.

第 2 节 Birkhoff 遍历定理

本节中，φ 是 $(\Omega, \mathcal{F}, P)$ 上的保测变换. 详见例 6.1.4. 首先证明如下结果：

定理 6.2.1. 遍历定理. 对任意 $X \in L^1$，

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m\omega) \to E(X|\mathcal{I}) \text{ 几乎必然且} L^1 \text{ 收敛.}$$

这一结果归功于 Birkhoff（1931），有时称为**点态遍历定理**或**个体遍历定理**，因为结论中包含几乎必然收敛. 当序列是遍历的时，极限为均值 EX. 此时，若取 $X = 1_A$，则 $\varphi^m \in A$ 的渐近时间比例为 $P(A)$.

下述证明基于 Yosida 和 Kakutani（1939）的一个积分不等式. 我们遵循 Garsia（1965）的方法. 这一证明不够直观，但每一步都不困难.

引理 6.2.2. 极大遍历引理. 令 $X_j(\omega) = X(\varphi^j\omega)$，$S_k(\omega) = X_0(\omega) + \cdots + X_{k-1}(\omega)$，且 $M_k(\omega) = \max(0, S_1(\omega), \ldots, S_k(\omega))$. 则 $E(X; M_k > 0) \geqslant 0$.

证明. 若 $j \leqslant k$，则 $M_k(\varphi\omega) \geqslant S_j(\varphi\omega)$，故加 $X(\omega)$ 得

$$X(\omega) + M_k(\varphi\omega) \geqslant X(\omega) + S_j(\varphi\omega) = S_{j+1}(\omega)$$

整理得

$$X(\omega) \geqslant S_{j+1}(\omega) - M_k(\varphi\omega) \text{ 对} j = 1, \ldots, k.$$

显然，$X(\omega) \geqslant S_1(\omega) - M_k(\varphi\omega)$，因为 $S_1(\omega) = X(\omega)$ 且 $M_k(\varphi\omega) \geqslant 0$. 因此

$$\begin{aligned} E(X(\omega); M_k > 0) &\geqslant \int_{\{M_k>0\}} \max(S_1(\omega), \ldots, S_k(\omega)) - M_k(\varphi\omega)\,\mathrm{d}P \\ &= \int_{\{M_k>0\}} M_k(\omega) - M_k(\varphi\omega)\,\mathrm{d}P. \end{aligned}$$

在 $\{M_k > 0\}^c$ 上 $M_k(\omega) = 0$ 且 $M_k(\varphi\omega) \geqslant 0$，故最后一个表达式

$$\geqslant \int M_k(\omega) - M_k(\varphi\omega)\,\mathrm{d}P = 0,$$

因为 φ 是保测变换. □

定理 6.2.1 的证明. $E(X|\mathcal{I})$ 在 φ 下不变（见习题 6.1.1），故令 $\bar{X} = X - E(X|\mathcal{I})$，可以不失一般性地假设 $E(X|\mathcal{I}) = 0$. 令 $\bar{X} = \limsup S_n/n$，令 $\varepsilon > 0$，且 $D = \{\omega : \bar{X}(\omega) > \varepsilon\}$. 目标是证明 $P(D) = 0$.

$\bar{X}(\varphi\omega)=\bar{X}(\omega)$，故 $D\in\mathcal{I}$. 令

$$\begin{aligned}
&X^*(\omega)=(X(\omega)-\varepsilon)1_D(\omega),\quad S_n^*(\omega)=X^*(\omega)+\cdots+X^*(\varphi^{n-1}\omega),\\
&M_n^*(\omega)=\max(0,S_1^*(\omega),\ldots,S_n^*(\omega)),\quad F_n=\{M_n^*>0\},\\
&F=\bigcup_n F_n=\left\{\sup_{k\geqslant 1}S_k^*/k>0\right\}.
\end{aligned}$$

由于 $X^*(\omega)=(X(\omega)-\varepsilon)1_D(\omega)$ 且 $D=\{\limsup S_k/k>\varepsilon\}$，可得

$$F=\left\{\sup_{k\geqslant 1}S_k/k>\varepsilon\right\}\cap D=D.$$

引理 6.2.2 蕴含 $E(X^*;F_n)\geqslant 0$. 由于 $E|X^*|\leqslant E|X|+\varepsilon<\infty$，控制收敛定理蕴含 $E(X^*;F_n)\to E(X^*;F)$，故 $E(X^*;F)\geqslant 0$. 最后一个结论看起来无关紧要，但 $F=D\in\mathcal{I}$，故它蕴含

$$0\leqslant E(X^*;D)=E(X-\varepsilon;D)=E(E(X|\mathcal{I});D)-\varepsilon P(D)=-\varepsilon P(D),$$

因为 $E(X|\mathcal{I})=0$. 最后一个不等式蕴含

$$0=P(D)=P(\limsup S_n/n>\varepsilon),$$

且由于 $\varepsilon>0$ 是任意的，可得 $\limsup S_n/n\leqslant 0$. 将上述结果应用于 $-X$ 表明 $S_n/n\to 0$ 几乎必然.

证明的核心部分已经结束，其余极易验证. 为证明收敛在 L^1 中发生，令

$$X_M^0(\omega)=X(\omega)1_{(|X(\omega)|\leqslant M)}\text{ 且}X_M^{00}(\omega)=X(\omega)-X_M^0(\omega).$$

已证的遍历定理部分蕴含

$$\frac{1}{n}\sum_{m=0}^{n-1}X_M^0(\varphi^m\omega)\to E(X_M^0|\mathcal{I})\text{ 几乎必然.}$$

由于 X_M^0 有界，有界收敛定理蕴含

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1}X_M^0(\varphi^m\omega)-E(X_M^0|\mathcal{I})\right|\to 0.$$

为处理 X_M^{00}，注意

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1}X_M^{00}(\varphi^m\omega)\right|\leqslant\frac{1}{n}\sum_{m=0}^{n-1}E|X_M^{00}(\varphi^m\omega)|=E|X_M^{00}|$$

且 $E|E(X_M^{00}|\mathcal{I})| \leqslant EE(|X_M^{00}||\mathcal{I}) = E|X_M^{00}|$. 故

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1} X_M^{00}(\varphi^m\omega) - E(X_M^{00}|\mathcal{I})\right| \leqslant 2E|X_M^{00}|,$$

且可得

$$\limsup_{n\to\infty} E\left|\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m\omega) - E(X|\mathcal{I})\right| \leqslant 2E|X_M^{00}|.$$

当 $M\to\infty$ 时，$E|X_M^{00}|\to 0$ 由控制收敛定理，证毕. □

下一步是看看定理 6.2.1 对例子意味着什么.

例题 6.2.3. 独立同分布序列. 由于 $\mathcal{I}$ 是平凡的，遍历定理蕴含

$$\frac{1}{n}\sum_{m=0}^{n-1} X_m \to EX_0 \text{ 几乎必然且} L^1 \text{ 收敛.}$$

几乎必然收敛就是强大数定律.

注释. 我们可以不引用遍历定理来证明大数定律的 L^1 收敛. 为此，注意

$$\frac{1}{n}\sum_{m=1}^{n} X_m^+ \to EX_0^+ \text{ 几乎必然}, \quad E\frac{1}{n}\sum_{m=1}^{n} X_m^+ = EX_0^+,$$

并利用定理 4.6.3 得出 $\frac{1}{n}\sum_{m=1}^{n} X_m^+ \to EX_0^+$ 在 L^1 中. 对负部得到类似结果，再用三角不等式即得所需结果.

例题 6.2.4. Markov 链. 设 X_n 是可数状态空间上具有平稳分布 π 的不可约 Markov 链. 设 f 是满足 $\sum_x |f(x)|\pi(x) < \infty$ 的函数. 在例 6.1.6 中，我们证明了 $\mathcal{I}$ 是平凡的，故将遍历定理应用于 $f(X_0(\omega))$ 得

$$\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) \to \sum_x f(x)\pi(x) \text{ 几乎必然且} L^1 \text{ 收敛.}$$

关于几乎必然收敛的另一证明，见习题 5.6.5.

例题 6.2.5. 圆周旋转. $\Omega = [0,1)$，$\varphi(\omega) = (\omega+\theta) \mod 1$. 假设 $\theta \in (0,1)$ 是无理数，故由 7.1 节的结果 $\mathcal{I}$ 是平凡的. 若令 $X(\omega) = 1_A(\omega)$，A 为 $[0,1)$ 的 Borel 子集，则遍历定理蕴含

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_{(\varphi^m\omega\in A)} \to |A| \text{ 几乎必然,}$$

其中 $|A|$ 表示 A 的 Lebesgue 测度. 最后一式对 $\omega = 0$ 的结果通常称为 Weyl 的**等分布定理**，尽管 Bohl 和 Sierpinski 也应获得荣誉. 关于历史和非概率证明，见 Hardy 和 Wright（1959），第 390–393 页.

为恢复数论结果，下面证明：

定理 6.2.6. 若 $A = [a, b)$，则例外集为空.

证明. 令 $A_k = [a + 1/k, b - 1/k)$. 若 $b - a > 2/k$，遍历定理蕴含

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_{A_k}(\varphi^m\omega) \to b - a - \frac{2}{k}$$

对 $\omega \in \Omega_k$ 成立，$P(\Omega_k) = 1$. 令 $G = \cap_k \Omega_k$，其中交取遍满足 $b - a > 2/k$ 的整数 k. $P(G) = 1$，故 G 在 $[0, 1)$ 中稠密. 若 $x \in [0, 1)$ 且 $\omega_k \in G$ 满足 $|\omega_k - x| < 1/k$，则 $\varphi^m\omega_k \in A_k$ 蕴含 $\varphi^m x \in A$，故

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_A(\varphi^m x) \geqslant b - a - \frac{2}{k}$$

对所有足够大的 k 成立. 注意 k 是任意的，并对 A^c 应用类似的推理，表明

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_A(\varphi^m x) \to b - a.$$

□

例题 6.2.7. Benford 定律. 正如 Gelfand 首先观察到的，等分布定理对 2^m 给出了一些有趣的结论. 令 $\theta = \log_{10} 2$，$1 \leqslant k \leqslant 9$，且 $A_k = [\log_{10} k, \log_{10}(k + 1))$，其中 $\log_{10} y$ 表示 y 以 10 为底的对数. 在最后一个结果中取 $x = 0$，我们有

$$\frac{1}{n}\sum_{m=0}^{n-1} 1_{A_k}(\varphi^m 0) \to \log_{10}\frac{k+1}{k}.$$

稍加思考可知，2^m 的首位数字是 k 当且仅当 $m\theta \mod 1 \in A_k$. 极限概率的数值为

1	2	3	4	5	6	7	8	9
.3010	.1761	.1249	.0969	.0792	.0669	.0580	.0512	.0458

$\{1, \ldots, 9\}$ 上的极限分布称为 Benford（1938）定律，尽管它是由 Newcomb（1881）发现的. 如 Raimi（1976）所解释的，在许多表格中，k 作为首位数字出现的观测频率近似为 $\log_{10}((k + 1)/k)$. 据说遵循 Benford 定律的一些例子包括：3259 个县的人口普查数据、Reader's Digest[1]的 308 个数字、335

[1]**译注.** 应该指美国的一家知名刊物，中文翻译为《读者文摘》.

条河流的面积、342 个美国科学界人士的地址. 下表比较了前五个类别中观测值的百分比与 Benford 定律:

	1	2	3	4	5
人口普查	33.9	20.4	14.2	8.1	7.2
Reader's Digest	33.4	18.5	12.4	7.5	7.1
河流	31.0	16.4	10.7	11.3	7.2
Benford 定律	30.1	17.6	12.5	9.7	7.9
地址	28.9	19.2	12.6	8.8	8.5

拟合远非完美，但在每种情况下 Benford 定律都与观测分布的一般形状相匹配. 美国国税局和其他政府机构使用 Benford 定律来检测欺诈. 当记录被编造时，首位数字分布不符合 Benford 定律.

习题

习题 6.2.1. 证明若 $X \in L^p$ $(p > 1)$，则定理 6.2.1 中的收敛在 L^p 中发生.

习题 6.2.2. (i) 证明若 $g_k(\omega) \to g(\omega)$ 几乎必然且 $E(\sup_k |g_k(\omega)|) < \infty$，则

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} g_m(\varphi^m \omega) = E(g|\mathcal{I}) \text{ 几乎必然.}$$

(ii) 证明若只假设 $g_n \to g$ 在 L^1 中，则得到 L^1 收敛.

习题 6.2.3. Wiener 极大不等式. 令 $X_j(\omega) = X(\varphi^j \omega)$，$S_k(\omega) = X_0(\omega) + \cdots + X_{k-1}(\omega)$，$A_k(\omega) = S_k(\omega)/k$，且 $D_k = \max(A_1, \ldots, A_k)$. 利用引理 6.2.2 证明若 $\alpha > 0$ 则

$$P(D_k > \alpha) \leqslant \alpha^{-1} E|X|.$$

第 3 节 常返性

本节研究平稳序列的常返性质. 首先给出遍历定理的一个应用.

设 $X_1, X_2, \ldots$ 为取值于 $\mathbb{R}^d$ 的平稳序列，$S_k = X_1 + \cdots + X_k$，$A = \{S_k \neq 0 \text{ 对所有} k \geqslant 1\}$，$R_n = |\{S_1, \ldots, S_n\}|$ 为时刻 n 访问的点数. Kesten, Spitzer 和 Whitman（见 Spitzer (1964), p. 40）在 X_i 独立同分布时证明了下述结果. 此时 $\mathcal{I}$ 是平凡的，故极限为 $P(A)$.

定理 6.3.1. 当 $n \to \infty$ 时，$R_n/n \to E(1_A|\mathcal{I})$ 几乎必然.

证明. 设 $X_1, X_2, \ldots$ 构造于 $(\mathbb{R}^d)^{\{0,1,\ldots\}}$ 上，$X_n(\omega) = \omega_n$，φ 为移位算子. 显然

$$R_n \geqslant \sum_{m=1}^{n} 1_A(\varphi^m \omega)$$

因为右边等于 $|\{m : 1 \leqslant m \leqslant n, S_\ell \neq S_m \text{ 对所有} \ell > m\}|$.

应用遍历定理得

$$\liminf_{n\to\infty} R_n/n \geqslant E(1_A|\mathcal{I}) \quad \text{几乎必然.}$$

为证反向不等式，记 $A_k = \{S_1 \neq 0, S_2 \neq 0, \ldots, S_k \neq 0\}$. 显然

$$R_n \leqslant k + \sum_{m=1}^{n-k} 1_{A_k}(\varphi^m \omega)$$

因为右边的和等于 $|\{m : 1 \leqslant m \leqslant n-k, S_\ell \neq S_m \text{ 对} m < \ell \leqslant m+k\}|$. 应用遍历定理得

$$\limsup_{n\to\infty} R_n/n \leqslant E(1_{A_k}|\mathcal{I}).$$

当 $k \uparrow \infty$ 时，$A_k \downarrow A$，故由条件期望的单调收敛定理（定理 4.1.9 的 (c)）知

$$E(1_{A_k}|\mathcal{I}) \downarrow E(1_A|\mathcal{I}) \quad \text{当} k \uparrow \infty,$$

证毕. □

由定理 6.3.1 可得关于具有平稳增量的随机游走的常返性结果，（对整数值随机游走而言）它是 Chung-Fuchs 定理 5.4.8 的推广.

定理 6.3.2. 设 $X_1, X_2, \ldots$ 为取值于 $\mathbb{Z}$ 的平稳序列，$E|X_i| < \infty$. 记 $S_n = X_1 + \cdots + X_n$，$A = \{S_1 \neq 0, S_2 \neq 0, \ldots\}$.

(i) 若 $E(X_1|\mathcal{I}) = 0$ 则 $P(A) = 0$.

(ii) 若 $P(A) = 0$ 则 $P(S_n = 0 \text{ i.o.}) = 1$.

注释. 换言之，均值为零蕴含常返性. 条件 $E(X_1|\mathcal{I})=0$ 用于排除一些平凡的例子，这些例子均值为 0 但由正均值和负均值序列组合而成，例如 $P(X_n=1 \text{ 对所有} n)=P(X_n=-1 \text{ 对所有} n)=1/2$.

证明. 若 $E(X_1|\mathcal{I})=0$ 则由遍历定理知 $S_n/n\to 0$ 几乎必然. 现

$$\limsup_{n\to\infty}\max_{1\leqslant k\leqslant n}|S_k|/n=\limsup_{n\to\infty}\max_{K\leqslant k\leqslant n}|S_k|/n\leqslant\max_{k\geqslant K}|S_k|/k$$

对任意 K 成立，且右边当 $K\uparrow\infty$ 时 $\downarrow 0$. 最后一个结论容易推出

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}|S_k|/n=0.$$

由于 $R_n\leqslant 1+2\max_{1\leqslant k\leqslant n}|S_k|$，故 $R_n/n\to 0$，由定理 6.3.1 知 $P(A)=0$.

记 $F_j=\{S_i\neq 0 \text{ 对} 1\leqslant i<j, S_j=0\}$ 以及

$$G_{j,k}=\{S_{j+i}-S_j\neq 0 \text{ 对} 1\leqslant i<k, S_{j+k}-S_j=0\}.$$

$P(A)=0$ 蕴含 $\sum_j P(F_j)=1$. 平稳性蕴含 $P(G_{j,k})=P(F_k)$，且对固定的 j，$G_{j,k}$ 不交，故 $\bigcup_k G_{j,k}=\Omega$ 几乎必然. 由此

$$\sum_k P(F_j\cap G_{j,k})=P(F_j)\quad\text{且}\quad\sum_{j,k}P(F_j\cap G_{j,k})=1.$$

在 $F_j\cap G_{j,k}$ 上，$S_j=0$ 且 $S_{j+k}=0$，故已证 $P(S_n=0 \text{ 至少两次})=1$. 重复上述论证得 $P(S_n=0 \text{ 至少} k \text{ 次})=1$ 对所有 k 成立，证毕. □

推广定理 6.3.2(ii) 证明中的推理可得 Kac (1947b) 的一个结果. 设 $X_0,X_1,\ldots$ 为取值于 $(S,\mathcal{S})$ 的平稳序列. 设 $A\in\mathcal{S}$，$T_0=0$，对 $n\geqslant 1$，记 $T_n=\inf\{m>T_{n-1}:X_m\in A\}$ 为第 n 次返回 A 的时间.

定理 6.3.3. 若 $P(X_n\in A \text{ 至少一次})=1$，则在 $P(\cdot|X_0\in A)$ 下，$t_n=T_n-T_{n-1}$ 是平稳序列且 $E(T_1|X_0\in A)=1/P(X_0\in A)$.

注释. 若 X_n 是可数状态空间 S 上的不可约 Markov 链，始于其平稳分布 π，且 $A=\{x\}$，则定理 6.3.3 表明 $E_xT_x=1/\pi(x)$，即定理 5.5.11. 定理 6.3.3 将该结果推广到任意 $A\subset S$，并去掉了 X_n 是 Markov 链的假设.

证明. 首先证明在 $P(\cdot|X_0\in A)$ 下，$t_1,t_2,\ldots$ 是平稳的. 为减少省略号，仅证明

$$P(t_1=m,t_2=n|X_0\in A)=P(t_2=m,t_3=n|X_0\in A).$$

同样的证明适用于任意有限维分布.

第一步是将 $\{X_n, n \geqslant 0\}$ 利用定理 6.1.4 延拓为双边平稳序列 $\{X_n, n \in \mathbb{Z}\}$. 记 $C_k = \{X_{-k} \notin A, \ldots, X_{-1} \notin A, X_0 \in A\}$. 则

$$\bigcup_{k=1}^{K} C_k = \{X_k \notin A \text{ 对 } -K \leqslant k \leqslant -1\}.$$

最后一个事件与 $\{X_k \notin A \text{ 对} 1 \leqslant k \leqslant K\}$ 概率相同，故令 $K \to \infty$ 得 $P(\bigcup_{k=1}^{\infty} C_k) = 1$. 为证所需的平稳性，记 $I_{j,k} = \{i \in [j,k] : X_i \in A\}$ 并注意到

$$\begin{aligned} P(t_2 = m, t_3 = n, X_0 \in A) &= \sum_{\ell=1}^{\infty} P(X_0 \in A, t_1 = \ell, t_2 = m, t_3 = n) \\ &= \sum_{\ell=1}^{\infty} P(C_\ell, X_0 \in A, t_1 = m, t_2 = n). \end{aligned}$$

为完成证明，计算

$$\begin{aligned} E(t_1 | X_0 \in A) &= \sum_{k=1}^{\infty} P(t_1 \geqslant k | X_0 \in A) \\ &= P(X_0 \in A)^{-1} \sum_{k=1}^{\infty} P(t_1 \geqslant k, X_0 \in A) \\ &= P(X_0 \in A)^{-1} \sum_{k=1}^{\infty} P(C_k) = 1/P(X_0 \in A) \end{aligned}$$

因为 C_k 不交且其并概率为 1. □

习题

习题 6.3.1. 记 $g_n = P(S_1 \neq 0, \ldots, S_n \neq 0)$ 对 $n \geqslant 1$ 且 $g_0 = 1$. 证明 $ER_n = \sum_{m=1}^{n} g_{m-1}$.

习题 6.3.2. 仿照定理 6.3.2(i) 的证明，证明若假设 $P(X_i > 1) = 0$，$EX_i > 0$ 且序列 X_i 遍历，则 $P(A) = EX_i$.

习题 6.3.3. 证明若 $P(X_n \in A \text{ 至少一次}) = 1$ 且 $A \cap B = \varnothing$ 则

$$E\left(\sum_{1 \leqslant m \leqslant T_1} 1_{(X_m \in B)} \,\middle|\, X_0 \in A \right) = \frac{P(X_0 \in B)}{P(X_0 \in A)}.$$

当 $A = \{x\}$ 且 X_n 是 Markov 链时，这就是定义平稳测度的 “循环技巧”. 见定理 5.5.7.

习题 6.3.4. 考虑 $X_n \in \{0,1\}$ 的特殊情形，记 $\bar{P} = P(\cdot | X_0 = 1)$. 这里 $A = \{1\}$，故 $T_1 = \inf\{m > 0 :$

$X_m = 1\}$. 证明 $\bar{P}(T_1 = n) = \bar{P}(T_1 \geqslant n)/\bar{E}T_1$. 当 $t_1, t_2, \ldots$ 独立同分布时，这退化为平稳更新过程中首次等待时间的公式.

第 4 节 次可加遍历定理

本节证明 Liggett (1985) 版本的 Kingman (1968) 定理.

定理 6.4.1. 设 $X_{m,n}$（$0 \leqslant m < n$）满足:

(i) $X_{0,m} + X_{m,n} \geqslant X_{0,n}$;

(ii) 对每个 k，$\{X_{nk,(n+1)k}, n \geqslant 1\}$ 是平稳序列;

(iii) $\{X_{m,m+k}, k \geqslant 1\}$ 的分布不依赖于 m;

(iv) $EX_{0,1}^{+} < \infty$ 且对每个 n，$EX_{0,n} \geqslant \gamma_0 n$，其中 $\gamma_0 > -\infty$.

则

(a) $\lim\limits_{n\to\infty} EX_{0,n}/n = \inf\limits_{m\geqslant 1} EX_{0,m}/m \equiv \gamma$;

(b) $X = \lim\limits_{n\to\infty} X_{0,n}/n$ 几乎必然存在且 L^1 收敛，故 $EX = \gamma$;

(c) 若 (ii) 中所有平稳序列都是遍历的，则 $X = \gamma$ 几乎必然.

注释. Kingman 假设了 (iv)，但用 $\ell < m < n$ 时 $X_{\ell,m} + X_{m,n} \geqslant X_{\ell,n}$ 以及 $\{X_{m+k,n+k}, 0 \leqslant m < n\}$ 的分布不依赖于 k 代替 (i)–(iii). 在下节的四个应用中有两个，这些更强的条件不成立.

在给出证明之前（证明颇为冗长），先考虑若干例子以获得直观理解. 由于 (ii) 和 (iii) 的有效性在每种情形下都显然，只需检验 (i) 和 (iv). 第一个例子表明定理 6.4.1 包含 Birkhoff 遍历定理作为特例.

例题 6.4.2. 平稳序列. 设 $\xi_1, \xi_2, \ldots$ 为平稳序列且 $E|\xi_k| < \infty$，记 $X_{m,n} = \xi_{m+1} + \cdots + \xi_n$. 则 $X_{0,n} = X_{0,m} + X_{m,n}$，且 (iv) 成立.

例题 6.4.3. 随机游走范围. 设 $\xi_1, \xi_2, \ldots$ 为平稳序列且记 $S_n = \xi_1 + \cdots + \xi_n, X_{m,n} = |\{S_{m+1}, \ldots, S_n\}|$. 显然 $X_{0,m} + X_{m,n} \geqslant X_{0,n}$. $0 \leqslant X_{0,n} \leqslant n$，故 (iv) 成立. 应用定理 6.4.1 得 $X_{0,n}/n \to X$ 几乎必然且 L^1 收敛，但并未告诉我们极限是什么.

例题 6.4.4. 最长公共子序列. 给定遍历平稳序列 $X_1, X_2, X_3, \ldots$ 和 $Y_1, Y_2, Y_3, \ldots$. 记 $L_{m,n} = \max\{K : X_{i_k} = Y_{j_k}$ 对$1 \leqslant k \leqslant K$, 其中$m < i_1 < i_2 \cdots < i_K \leqslant n$ 且$m < j_1 < j_2 \cdots < j_K \leqslant n\}$. 显然

$$L_{0,m} + L_{m,n} \leqslant L_{0,n}$$

故 $X_{m,n} = -L_{m,n}$ 是次可加的. $0 \leqslant L_{0,n} \leqslant n$ 故 (iv) 成立. 应用定理 6.4.1 得

$$L_{0,n}/n \to \gamma = \sup_{m\geqslant 1} E(L_{0,m}/m).$$

上述例子已提供了足够的动机. 下节将给出定理 6.4.1 的四个更多应用.

定理 6.4.1 的证明. 证明分四步. 第一、二、四步追溯至 Kingman (1968). 现存的六种次可加遍历定理的证明都以不同的方式处理关键的第三步. 这里采用 S. Leventhal (1988) 的方法，他反过来又基于 Katznelson 和 Weiss (1982).

第一步. 首先检验 $E|X_{0,n}| \leqslant Cn$. 为此注意 (i) 蕴含 $X^+_{0,m} + X^+_{m,n} \geqslant X^+_{0,n}$. 反复使用最后一个不等式并引用 (iii) 得 $EX^+_{0,n} \leqslant nEX^+_{0,1} < \infty$. 由于 $|x| = 2x^+ - x$，由 (iv) 得

$$E|X_{0,n}| \leqslant 2EX^+_{0,n} - EX_{0,n} \leqslant Cn < \infty.$$

记 $a_n = EX_{0,n}$. (i) 和 (iii) 蕴含

$$a_m + a_{n-m} \geqslant a_n. \tag{6.4.1}$$

由此易得

$$a_n/n \to \inf_{m\geqslant 1} a_m/m \equiv \gamma. \tag{6.4.2}$$

为证此，注意下极限显然 $\geqslant \gamma$，故只需证明上极限 $\leqslant a_m/m$ 对任意 m. 最后这个事实容易，因若写 $n = km + \ell$ 且 $0 \leqslant \ell < m$，则反复使用 (6.4.1) 得 $a_n \leqslant ka_m + a_\ell$. 除以 $n = km + \ell$ 得

$$\frac{a_n}{n} \leqslant \frac{km}{km+\ell} \cdot \frac{a_m}{m} + \frac{a_\ell}{n}.$$

令 $n \to \infty$ 并注意 $0 \leqslant \ell < m$ 得 (6.4.2)，从而证明了定理 6.4.1 的 (a).

第二步. 反复使用 (i) 得

$$\begin{aligned} X_{0,n} &\leqslant X_{0,km} + X_{km,n}, \\ X_{0,n} &\leqslant X_{0,(k-1)m} + X_{(k-1)m,km} + X_{km,n} \end{aligned}$$

如此继续直到右边第一项为 $X_{0,m}$. 除以 $n = km + \ell$ 得

$$\frac{X_{0,n}}{n} \leqslant \frac{k}{km+\ell} \cdot \frac{X_{0,m} + \cdots + X_{(k-1)m,km}}{k} + \frac{X_{km,n}}{n}. \tag{6.4.3}$$

利用 (ii) 和遍历定理得

$$\frac{X_{0,m} + \cdots + X_{(k-1)m,km}}{k} \to A_m \quad \text{几乎必然且} L^1 \text{收敛}$$

其中 $A_m = E(X_{0,m}|\mathcal{I}_m)$ 且下标表示 $\mathcal{I}_m$ 是序列 $X_{(k-1)m,km}$ $(k \geqslant 1)$ 的移位不变 σ-域. 极限的精确公式不重要，但后面需要知道 $EA_m = EX_{0,m}$.

若固定 ℓ 且令 $\varepsilon > 0$，则 (iii) 蕴含

$$\sum_{k=1}^{\infty} P(X_{km,km+\ell} > (km+\ell)\varepsilon) \leqslant \sum_{k=1}^{\infty} P(X_{0,\ell} > k\varepsilon) < \infty$$

因 $EX_{0,\ell}^{+} < \infty$ 由第一步开始的结果. 最后两个观察蕴含

$$\bar{X} \equiv \limsup_{n\to\infty} X_{0,n}/n \leqslant A_m/m. \tag{6.4.4}$$

取期望值得 $E\bar{X} \leqslant E(X_{0,m}/m)$，对 m 取下确界，有 $E\bar{X} \leqslant \gamma$. 注意若 (ii) 中所有平稳序列都是遍历的，则有 $\bar{X} \leqslant \gamma$.

注释. 若 (i)–(iii) 成立，$EX_{0,1}^{+} < \infty$ 且 $\inf_m EX_{0,m}/m = -\infty$，则由最后论证知当 $n \to \infty$ 时 $X_{0,n}/n \to -\infty$ 几乎必然.

第三步. 下一步记

$$\underline{X} = \liminf_{n\to\infty} X_{0,n}/n$$

并证明 $E\underline{X} \geqslant \gamma$. 由于 $\infty > EX_{0,1} \geqslant E\underline{X} \geqslant \gamma > -\infty$，且第二步已证 $E\bar{X} \leqslant \gamma$，由此 $\bar{X} = \underline{X}$，即 $X_{0,n}/n$ 的极限几乎必然存在. 记

$$\underline{X}_m = \liminf_{n\to\infty} X_{m,m+n}/n.$$

(i) 蕴含

$$X_{0,m+n} \leqslant X_{0,m} + X_{m,m+n}.$$

两边除以 n 并令 $n \to \infty$ 得 $\underline{X}_m \leqslant \underline{X}$ 几乎必然. 然而 (iii) 蕴含 $\underline{X}_m$ 与 $\underline{X}$ 同分布故 $\underline{X}_m = \underline{X}$ 几乎必然.

设 $\varepsilon > 0$ 且 $Z_{m,n} = \varepsilon + (\underline{X} \vee -M)$. 由于 $\underline{X} \leqslant \bar{X}$ 且 $E\bar{X} \leqslant \gamma < \infty$ 由第二步，$E|Z_{0,1}| < \infty$. 记

$$Y_{m,n} = X_{m,n} - (n-m)Z_{m,n}.$$

$Y_{m,n}$ 满足 (i)–(iv)，因 $Z_{m,n} = -(n-m)Z_{m,n}$ 满足，且

$$\underline{Y} \equiv \liminf_{n\to\infty} Y_{0,n}/n \leqslant -\varepsilon. \tag{6.4.5}$$

记 $T_m = \min\{n \geqslant 1 : Y_{m,m+n} \leqslant 0\}$. (iii) 蕴含 $T_m \stackrel{d}{=} T_0$ 且

$$E(Y_{m,m+1}; T_m > N) = E(Y_{0,1}; T_0 > N).$$

(6.4.5) 蕴含 $P(T_0<\infty)=1$，故可取 N 足够大使得

$$E(Y_{0,1};T_0>N)\leqslant\varepsilon.$$

记

$$S_m=\begin{cases}T_m & \text{在}\{T_m\leqslant N\}\text{ 上},\\ 1 & \text{在}\{T_m>N\}\text{ 上}.\end{cases}$$

这不是停时，但对平稳序列而言停时并无特殊之处. 记

$$\xi_m=\begin{cases}0 & \text{在}\{T_m\leqslant N\}\text{ 上},\\ Y_{m,m+1} & \text{在}\{T_m>N\}\text{ 上}.\end{cases}$$

由于 $Y(m,m+T_m)\leqslant 0$ 恒成立且 $S_m=1$，$Y_{m,m+1}>0$ 在 $\{T_m>N\}$ 上，有 $Y(0,m+S_m)\leqslant\xi_m$ 且 $\xi_m\geqslant 0$. 记 $R_0=0$，对 $k\geqslant 1$，记 $R_k=R_{k-1}+S(R_{k-1})$. 记 $K=\max\{k:R_k\leqslant n\}$. 由 (i) 得

$$Y(0,n)\leqslant Y(R_0,R_1)+\cdots+Y(R_{K-1},R_K)+Y(R_K,n).$$

由于 $\xi_m\geqslant 0$ 且 $n-R_K\leqslant N$，最后一个量

$$\leqslant\sum_{m=0}^{n-1}\xi_m+\sum_{j=1}^{N}|Y_{n-j,n-j+1}|.$$

这里在 $Y(R_K,n)$ 上使用了 (i). 两边除以 n，取期望值，并令 $n\to\infty$ 得

$$\limsup_{n\to\infty}EY_{0,n}/n\leqslant E\xi_0\leqslant E(Y_{0,1};T_0>N)\leqslant\varepsilon.$$

由 (a) 和 $Y_{0,n}$ 的定义得

$$\gamma=\lim_{n\to\infty}EX_{0,n}/n\leqslant 2\varepsilon+E(\underline{X}\vee -M).$$

由于 $\varepsilon>0$ 和 M 任意，故 $E\underline{X}\geqslant\gamma$，第三步完成.

第四步. 只需证 L^1 收敛. 记 $\Gamma_m=A_m/m$ 为 (6.4.4) 中的极限，回忆 $E\Gamma_m=E(X_{0,m}/m)$，记 $\underline{\Gamma}_m=\inf_{\ell\geqslant m}\Gamma_\ell$. 注意 $|x|=2x^+-x$（分 $z\geqslant 0$ 和 $z<0$ 两种情形），可写

$$E|X_{0,n}/n-\Gamma|=2E(X_{0,n}/n-\Gamma)^+-E(X_{0,n}/n-\Gamma)\leqslant 2E(X_{0,n}/n-\Gamma)^+$$

因

$$E(X_{0,n}/n)\geqslant\gamma=\inf_m E\Gamma_m\geqslant E\underline{\Gamma}_m.$$

利用平凡不等式 $(x+y)^+ \leqslant x^+ + y^+$ 并注意 $\Gamma_m \geqslant \underline{\Gamma}_m$ 得

$$E(X_{0,n}/n - \Gamma)^+ \leqslant E(X_{0,n}/n - \Gamma_m)^+ + E(\Gamma_m - \underline{\Gamma}_m).$$

现 $E\Gamma_m \to \gamma$ 当 $m \to \infty$ 且 $E\underline{\Gamma}_m \geqslant E\underline{X} \geqslant E\bar{X} \geqslant \gamma$ 由第二、三步，故 $E\underline{\Gamma}_m = \gamma$，于是若 m 大则 $E(\Gamma_m - \underline{\Gamma}_m)$ 小. 为估计另一项，注意 (i) 蕴含

$$E(X_{0,n}/n - \Gamma_m)^+ \leqslant E\left(\frac{X_{0,m} + \cdots + X_{(k-1)m,km}}{km+\ell} - \Gamma_m\right)^+ + E\left(\frac{X_{km,n}}{n}\right)^+.$$

第二项 $= E(X_{0,\ell}^+/n) \to 0$ 当 $n \to \infty$. 对第一项，注意 $y^+ \leqslant |y|$，遍历定理蕴含

$$E\left|\frac{X_{0,m} + \cdots + X_{(k-1)m,km}}{k} - \Gamma_m\right| \to 0$$

故定理 6.4.1 的证明完成. □

第 5 节 应用

本节给出次可加遍历定理 6.4.1 的三个应用. 这些例子相互独立, 可按任意顺序阅读. 在例 6.5.5 中, 会遇到 Liggett 版本适用但 Kingman 版本不适用的情况.

例题 6.5.1. 随机矩阵乘积. 设 $A_1, A_2, \ldots$ 为 $k \times k$ 矩阵的正元素平稳序列且记

$$\alpha_{m,n}(i,j) = (A_{m+1} \cdots A_n)(i,j),$$

即乘积中第 i 行第 j 列的元素. 显然

$$\alpha_{0,m}(1,1)\alpha_{m,n}(1,1) \leqslant \alpha_{0,n}(1,1)$$

故若记 $X_{m,n} = -\log \alpha_{m,n}(1,1)$ 则 $X_{0,m} + X_{m,n} \geqslant X_{0,n}$. 为检验 (iv), 注意

$$\prod_{m=1}^{n} A_m(1,1) \leqslant \alpha_{0,n}(1,1) \leqslant k^{n-1} \sup_{i,j} \prod_{m=1}^{n} A_m(i,j)$$

或取对数

$$-\sum_{m=1}^{n} \log A_m(1,1) \geqslant X_{0,n} \geqslant -(n \log k) - \sum_{m=1}^{n} \log \sup_{i,j} A_m(i,j).$$

故若 $E \log A_m(1,1) > -\infty$ 则 $EX_{0,1}^{+} < \infty$, 且若

$$E \log \sup_{i,j} A_m(i,j) < \infty$$

则 $EX_{0,n}^{-} \leqslant \gamma_0 n$. 若注意到

$$P\left(\log \sup_{i,j} A_m(i,j) \geqslant x\right) \leqslant \sum_{i,j} P(\log A_m(i,j) \geqslant x)$$

则可见只需假设

$$E|\log A_m(i,j)| < \infty \quad \text{对所有} i,j. \tag{$*$}$$

当 $(*)$ 成立时, 应用定理 6.4.1 得 $X_{0,n}/n \to X$ 几乎必然. 利用元素的严格正性, 容易将此结果改进为

$$\frac{1}{n} \log \alpha_{0,n}(i,j) \to -X \quad \text{几乎必然对所有} i,j, \tag{6.5.1}$$

这是 Furstenberg 和 Kesten (1960) 首次证明的结果.

另一种方法是记

$$\|A\|_1 = \max_i \sum_j |A(i,j)| = \max\{\|xA\|_1 : \|x\|_1 = 1\},$$

其中 $(xA)_j = \sum_i x_i A(i,j)$ 且 $\|x\|_1 = |x_1| + \cdots + |x_k|$. 显然 $\|AB\|_1 \leqslant \|A\|_1 \cdot \|B\|_1$，故若记

$$\beta_{m,n} = \|A_{m+1} \cdots A_n\|_1$$

且 $Y_{m,n} = \log \beta_{m,n}$，则 $Y_{m,n}$ 是次可加的. 容易用次可加遍历定理得

$$\frac{1}{n} \log \|A_{m+1} \cdots A_n\|_1 \to -X \quad \text{几乎必然},$$

其中 X 是 $X_{0,n}/n$ 的极限. 由此可见

$$\sup_{m \geqslant 1} (E \log \alpha_{0,m})/m = -X = \inf_{m \geqslant 1} (E \log \beta_{0,m})/m.$$

例题 6.5.2. 随机排列中的递增序列. 设 π 为 $\{1, 2, \ldots, n\}$ 的排列，$\ell(\pi)$ 为 π 中最长递增序列的长度. 即最大的 k 使得存在整数 $i_1 < i_2 \cdots < i_k$ 满足 $\pi(i_1) < \pi(i_2) < \cdots < \pi(i_k)$. Hammersley (1970) 通过在平面上放置单位速率 Poisson 过程来研究此问题，且对 $s < t \in [0, \infty)$，记 $Y_{s,t}$ 表示位于正方形 $R_{s,t}$（顶点为 $(s,s),(s,t),(t,t),(t,s)$）中最长递增路径的长度. 即最大的 k 使得 Poisson 过程中存在点 (x_i, y_i) 满足 $s < x_1 < \cdots < x_k < t$ 且 $s < y_1 < \cdots < y_k < t$. 显然 $Y_{0,m} + Y_{m,n} \leqslant Y_{0,n}$. 对 $-Y_{0,n}$ 应用定理 6.4.1 得

$$Y_{0,n}/n \to \gamma \equiv \sup_{m \geqslant 1} EY_{0,m}/m \quad \text{几乎必然}.$$

对每个 k，$Y_{nk,(n+1)k}$（$n \geqslant 0$）独立同分布，故极限为常数. 习题 6.5.4 将证明 $\gamma < \infty$.

为从 Poisson 过程的结果回到随机排列问题，记 $\tau(n)$ 为 $R_{0,t}$ 中有 n 个点的最小 t 值. 将 $R_{0,\tau(n)}$ 中的 n 个点写为 (x_i, y_i)，其中 $0 < x_1 < x_2 \cdots < x_n \leqslant \tau(n)$，记 π_n 为 $\{1, 2, \ldots, n\}$ 的唯一排列使得 $y_{\pi_n(1)} < y_{\pi_n(2)} \cdots < y_{\pi_n(n)}$. 显然 $Y_{0,\tau(n)} = \ell(\pi_n)$. 一个简单的论证表明：

引理 6.5.3. $\tau(n)/\sqrt{n} \to 1$ 几乎必然.

证明. 记 S_n 为 $R_{0,\sqrt{n}}$ 中的点数. $S_n - S_{n-1}$ 是均值为 1 的独立 Poisson 随机变量，故强大数定律蕴含 $S_n/n \to 1$ 几乎必然. 若 $\varepsilon > 0$ 则对大 n，$S_{n(1-\varepsilon)} < n < S_{n(1+\varepsilon)}$，故 $(1-\varepsilon)\sqrt{n} \leqslant \tau(n) \leqslant (1+\varepsilon)\sqrt{n}$. □

由引理 6.5.3 和 $m \mapsto Y_{0,m}$ 的单调性得

$$n^{-1/2} \ell(\pi_n) \to \gamma \quad \text{几乎必然}.$$

Hammersley (1970) 证明了 $\pi/2 \leqslant \gamma \leqslant e$，Kingman (1973) 表明 $1.59 < \gamma < 2.49$. 见习题 6.5.3 和 6.5.4. 关于随机排列问题的后续工作，见 Logan 和 Shepp (1977) 以及 Vershik 和 Kerov (1977)，已证明 $\gamma = 2$.

例题 6.5.4. 首次通过渗流. 将 $\mathbb{Z}^d$ 视为图，边连接每个 $x, y \in \mathbb{Z}^d$ 满足 $|x - y| = 1$. 给每条边独立赋予非负随机变量 $\tau(e)$，表示沿任一方向通过该边所需的时间. 若 e 是连接 x 和 y 的边，记 $\tau(x, y) = \tau(y, x) = \tau(e)$. 若 $x = x_0, x_1, \ldots, x_n = y$ 是从 x 到 y 的路径，即满足 $|x_m - x_{m-1}| = 1$ $(1 \leqslant m \leqslant n)$ 的序列，定义该路径的通行时间为 $\tau(x_0, x_1) + \cdots + \tau(x_{n-1}, x_n)$. 定义从 x 到 y 的通过时间 $t(x, y)$ 为所有从 x 到 y 路径的通行时间之下确界. 设 $z \in \mathbb{Z}^d$ 且 $X_{m,n} = t(mu, nu)$，其中 $u = (1, 0, \ldots, 0)$.

显然 $X_{0,m} + X_{m,n} \geqslant X_{0,n}$. $X_{0,n} \geqslant 0$，故若 $E\tau(x, y) < \infty$ 则 (iv) 成立，定理 6.4.1 蕴含 $X_{0,n}/n \to X$ 几乎必然. 为证极限为常数，按某种顺序枚举边 $e_1, e_2, \ldots$ 并注意 X 关于独立同分布序列 $\tau(e_1), \tau(e_2), \ldots$ 的尾 σ-域可测.

不难看到有限一阶矩的假设可以减弱. 若 τ 有分布 F 满足

$$\int_0^\infty (1 - F(x))^{2d} \mathrm{d}x < \infty \tag{$*$}$$

即 $2d$ 个独立复制的最小值有有限均值，则通过找到 $2d$ 条从 0 到 $u = (1, 0, \ldots, 0)$ 的不交路径，可推出 $E\tau(0, u) < \infty$ 且可应用 (6.1). 条件 $(*)$ 也是 $X_{0,n}/n$ 收敛到有限极限的必要条件. 若 $(*)$ 不成立且 Y_n 是 ν 处所有边 $t(e)$ 的最小值，则

$$\limsup_{n\to\infty} X_{0,n}/n \geqslant \limsup_{n\to\infty} Y_n/n = \infty \quad \text{几乎必然.}$$

例题 6.5.5. 年龄依赖分支过程. 这是分支过程的一个变体，其中每个个体存活一段时间（服从分布 F）后以概率 p_k 产生 k 个后代. 过程的描述通过假设第 0 代有一个个体在时刻 0 出生，且当这个粒子死亡时，其后代启动原过程的独立复制来完成.

设 $p_0 = 0$，记 $X_{0,m}$ 为第 m 代第一个成员的出生时间，$X_{m,n}$ 为该个体在第 n 代有后代所需的时间滞后. 有平局时，从第 m 代在时刻 $X_{0,m}$ 出生的个体中随机选取一个. 显然 $X_{0,n} \leqslant X_{0,m} + X_{m,n}$. 由于 $X_{0,n} \geqslant 0$，若假设 F 有有限均值则 (iv) 成立. 应用定理 6.4.1 得

$$X_{0,n}/n \to \gamma \quad \text{几乎必然.}$$

极限为常数因为序列 $\{X_{nk,(n+1)k}, n \geqslant 0\}$ 独立同分布.

注释. 当 $\ell > 0$ 时不等式 $X_{\ell,m} + X_{m,n} \geqslant X_{\ell,n}$ 不成立，因为若称 i_m 为 $n > m$ 时决定 $X_{m,n}$ 值的个体，则 i_m 可能不是 i_ℓ 的后代.

与通常一样，需用其他方法确定常数. 设 $t_1, t_2, \ldots$ 独立同分布服从分布 F，$T_n = t_1 + \cdots + t_n$，

$\mu = \sum_k kp_k$. 记 $Z_n(an)$ 为时刻 an 之前出生的第 n 代个体数. 第 n 代的每个个体在时刻 an 之前出生的概率为 $P(T_n \leqslant an)$，且时间与后代数独立故

$$EZ_n(an) = E(E(Z_n(an)|Z_n)) = E(Z_nP(T_n \leqslant an)) = \mu^n P(T_n \leqslant an).$$

由 2.6 节的结果，$n^{-1}\log P(T_n \leqslant an) \to -c(a)$ 当 $n \to \infty$. 若 $\log\mu - c(a) < 0$ 则由 Chebyshev 不等式和 Borel-Cantelli 引理得 $P(Z_n(an) \geqslant 1 \text{ i.o.}) = 0$. 反之，若对某个 n 有 $EZ_n(an) > 1$，则可定义一个超临界分支过程 Y_m，它由 Y_{m-1} 中个体在 an 单位时间内出生的后代中的第 mn 代后代组成. 这表明以正概率，$X_{0,mn} \leqslant mna$ 对所有 m 成立. 结合最后两个观察以及 $c(a)$ 严格递增得

$$\gamma = \inf\{a : \log\mu - c(a) > 0\}.$$

最后一个结果来自 Biggins (1977). 见他的 (1978) 和 (1979) 论文以了解推广和改进.

习题

习题 6.5.1. 为说明定理 6.4.1(a) 中的收敛可以任意慢，记 $X_{m,m+k} = f(k) \geqslant 0$，其中 $f(k)/k$ 递减，并检验 $X_{m,m+k}$ 是次可加的.

习题 6.5.2. 考虑最长公共子序列问题，例 6.4.4，当 $X_1, X_2, \ldots$ 和 $Y_1, Y_2, \ldots$ 独立同分布且以概率 $1/2$ 分别取值 0 和 1.

(a) 计算 EL_1 和 $EL_2/2$ 以得到 γ 的下界.

(b) 通过计算长度为 $K = an$ 的 i 和 j 序列具有所需性质的期望数来证明 $\gamma < 1$. Chvatal 和 Sankoff (1975) 证明了 $0.727273 \leqslant \gamma \leqslant 0.866595$.

习题 6.5.3. 给定 $[0,\infty) \times [0,\infty)$ 中的单位速率 Poisson 过程，记 (X_1, Y_1) 为最小化 $x + y$ 的点. 记 (X_2, Y_2) 为 $[X_1,\infty) \times [Y_1,\infty)$ 中最小化 $x + y$ 的点，依此类推. 用此构造证明在例 6.5.2 中 $\gamma \geqslant (8/\pi)^{1/2} > 1.59$.

习题 6.5.4. 设 π_n 为 $\{1,\ldots,n\}$ 的随机排列，J_n^k 为 $\{1,\ldots,n\}$ 的大小为 k 的子集数使得对应的 $\pi_n(j)$ 形成递增子序列. 计算 EJ_n^k 并取 $k \sim \alpha n^{1/2}$ 以推出在例 6.5.2 中 $\gamma \leqslant e$.

习题 6.5.5. 记 $\varphi(\theta) = E\exp(-\theta t_i)$ 且

$$Y_n = (\mu\varphi(\theta))^{-n} \sum_{i=1}^{Z_n} \exp(-\theta T_n(i))$$

其中求和取遍第 n 代个体且 $T_n(i)$ 是第 i 个人的出生时间. 证明 Y_n 是非负鞅并用此推出：如果 $\exp(-\theta a)/\mu\varphi(\theta) > 1$ 则 $P(X_{0,n} \leqslant an) \to 0$. 稍加思考即知此界与上题答案相同.

第七章
布朗运动

布朗运动是一个具有巨大实际和理论意义的过程. 它起源于 (a) Robert Brown 在 1828 年观察到的 “悬浮在水中的花粉粒持续进行群集运动” 现象之模型，以及 (b) Bachelier (1900) 关于股票市场的工作. 这只是布朗运动被用来建模的众多系统中的两个. 在理论方面，布朗运动是具有平稳独立增量的高斯 Markov 过程. 它位于三类重要过程的交集中，且在每种理论中都是一个基本例子.

本章第一部分发展布朗运动的性质. 7.1 节定义布朗运动并研究其路径的连续性. 7.2 节证明 Markov 性质和一个相关的 0-1 律.

第 1 节　定义与构造

一维**布朗运动**是满足如下性质的一个实值过程 B_t $(t \geqslant 0)$：

(a) 若 $t_0 < t_1 < \cdots < t_n$，则 $B(t_0), B(t_1) - B(t_0), \ldots, B(t_n) - B(t_{n-1})$ 相互独立.

(b) 若 $s, t \geqslant 0$，则

$$P(B(s+t) - B(s) \in A) = \int_A (2\pi t)^{-1/2} \exp(-x^2/2t)\, \mathrm{d}x.$$

(c) 以概率 1，$t \mapsto B_t$ 连续.

(a) 表明 B_t 具有**独立增量**. (b) 表明增量 $B(s+t) - B(s)$ 服从均值为 0、方差为 t 的正态分布. (c) 是自明的.

就布朗的花粉粒而言，(c) 当然是合理的. 花粉粒的运动由数百万个水分子的轰击的净效应引起，因此由中心极限定理，任一区间内的位移应服从正态分布，且两个不交区间内的位移应相互独立，这说明了 (a) 和 (b) 的合理性.

由定义直接可得两个今后多次有用的推论：

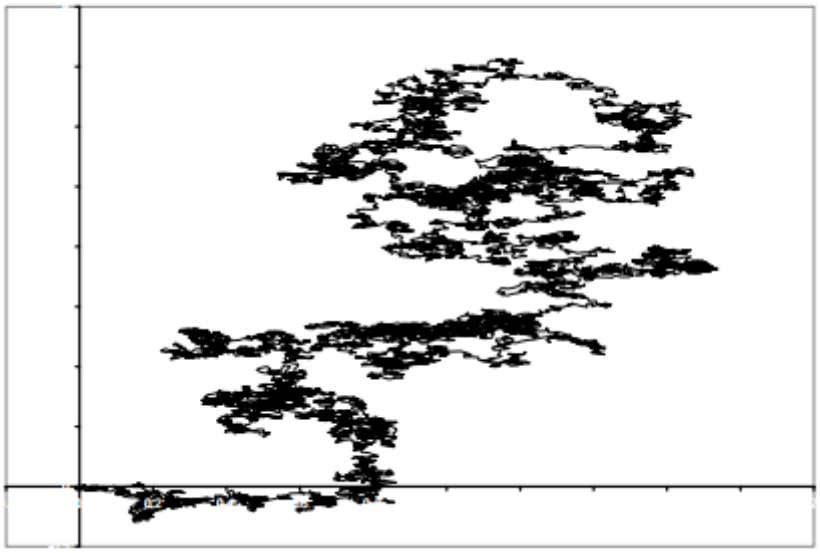

图 7.1: 二维布朗运动的模拟

平移不变性. $\{B_t - B_0, t \geqslant 0\}$ 独立于 B_0，且与从 0 出发的布朗运动同分布.

证明. 令 $A_1 = \sigma(B_0)$，A_2 为形如

$$\{B(t_1) - B(t_0) \in A_1, \ldots, B(t_n) - B(t_{n-1}) \in A_n\}$$

的事件. A_i 是独立的 π-系，故所需结果由 $\pi - \lambda$ 定理 2.1.6 推出. □

布朗尺度变换. 若 $B_0 = 0$，则对任意 $t > 0$，

$$\{B_{st}, s \geqslant 0\} \overset{\mathrm{d}}{=} \{t^{1/2} B_s, s \geqslant 0\}. \tag{7.1.1}$$

更确切地说，这两族随机变量具有相同的**有限维分布**，即若 $s_1 < \cdots < s_n$，则

$$(B_{s_1 t}, \ldots, B_{s_n t}) \overset{\mathrm{d}}{=} (t^{1/2} B_{s_1}, \ldots, t^{1/2} B_{s_n}).$$

证明. $n = 1$ 时，注意 $t^{1/2}$ 乘以均值为 0、方差为 s 的正态变量是均值为 0、方差为 st 的正态变量. $n > 1$ 的结果由独立增量推出. □

从 $B_0 = 0$ 出发的布朗运动还有如下等价定义，有时很有用，即 B_t（$t \geqslant 0$）是满足如下条件的实值过程：

(a′) $B(t)$ 是**高斯过程**（即其所有有限维分布都是多元正态的）.

(b′) $EB_s = 0$ 且 $EB_s B_t = s \wedge t$.

(c′) 以概率 1，$t \mapsto B_t$ 连续.

容易看出 (a) 和 (b) 蕴含 (a′). 为从 (a) 和 (b) 得到 (b′)，设 $s<t$ 并写

$$EB_sB_t = E(B_s^2) + E(B_s(B_t - B_s)) = s.$$

反过来更容易. (a′) 和 (b′) 确定了 B_t 的有限维分布，由上述计算可知它们必须与 (a) 和 (b) 中定义的分布一致.

任何关于布朗运动的处理中，需解决的首要问题是：是否存在满足这些性质的过程？答案当然是肯定的，否则本章就不存在. 出于教学原因，下面先追求一种导致困境的方法，然后稍作修正以解决困难.

固定 $x\in\mathbb{R}$，并对每个 $0<t_1<\cdots<t_n$，在 $\mathbb{R}^n$ 上定义一个测度：

$$\mu_{x,t_1,\dots,t_n}(A_1\times\cdots\times A_n) = \int_{A_1}\mathrm{d}x_1\cdots\int_{A_n}\mathrm{d}x_n\prod_{m=1}^{n}p_{t_m-t_{m-1}}(x_{m-1},x_m) \tag{7.1.2}$$

其中 $A_i\in\mathcal{R}$，$x_0=x$，$t_0=0$，且

$$p_t(a,b) = (2\pi t)^{-1/2}\exp(-(b-a)^2/2t).$$

由上述公式，容易看出对固定的 x，测度族 μ 是一组相容的**有限维分布**（f.d.d.）. 也就是说，如果 $\{s_1,\dots,s_{n-1}\}\subset\{t_1,\dots,t_n\}$ 且 $t_j\notin\{s_1,\dots,s_{n-1}\}$，则

$$\mu_{x,s_1,\dots,s_{n-1}}(A_1\times\cdots\times A_{n-1}) = \mu_{x,t_1,\dots,t_n}(A_1\times\cdots\times A_{j-1}\times\mathbb{R}\times A_j\times\cdots\times A_{n-1}).$$

当 $j=n$ 时这是显然的. 为验证 $1\leqslant j<n$ 时的等式，只需证明

$$\int p_{t_j-t_{j-1}}(x,y)p_{t_{j+1}-t_j}(y,z)\,\mathrm{d}y = p_{t_{j+1}-t_{j-1}}(x,z).$$

由平移不变性，不妨设 $x=0$，但这无非是说均值为 0、方差分别为 t_j-t_{j-1} 和 $t_{j+1}-t_j$ 的独立正态变量之和服从均值为 0、方差为 $t_{j+1}-t_{j-1}$ 的正态分布.

验证有限维分布的相容性后，即得布朗运动的第一个构造：

定理 7.1.1. 令 $\Omega_o=\{\text{函数}\omega:[0,\infty)\to\mathbb{R}\}$，$\mathcal{F}_o$ 为由有限维集 $\{\omega:\omega(t_i)\in A_i\text{ 对}1\leqslant i\leqslant n\}$ 生成的 σ-域，其中 $A_i\in\mathcal{R}$. 对每个 $x\in\mathbb{R}$，存在唯一的概率测度 ν_x 在 $(\Omega_o,\mathcal{F}_o)$ 上使得 $\nu_x\{\omega:\omega(0)=x\}=1$，且当 $0<t_1<\cdots<t_n$ 时

$$\nu_x\{\omega:\omega(t_i)\in A_i\} = \mu_{x,t_1,\dots,t_n}(A_1\times\cdots\times A_n). \tag{7.1.3}$$

这由 **Kolmogorov 扩张定理**的一个推广（附录中的 (7.1)）推出. 由此陷入上述困境，不再纠缠

于细节. 若 $C=\{\omega: t\mapsto\omega(t)\text{ 连续}\}$，则 $C\notin\mathcal{F}_o$，即 C 不可测. 证明 $C\notin\mathcal{F}_o$ 的最简单方法是做习题 7.1.4.

上述问题很容易解决. 令 $\mathbb{Q}_2=\{m2^{-n}: m,n\geqslant 0\}$ 为**二进有理数**. 若 $\Omega_q=\{\omega:\mathbb{Q}_2\to\mathbb{R}\}$，$\mathcal{F}_q$ 为由有限维集生成的 σ-域，则枚举有理数 $q_1,q_2,\ldots$ 并应用 Kolmogorov 扩张定理表明，由 Kolmogorov 扩张定理，可在 $(\Omega_q,\mathcal{F}_q)$ 上构造概率 ν_x 使得 $\nu_x\{\omega:\omega(0)=x\}=1$ 且当 $t_i\in\mathbb{Q}_2$ 时 (7.1.3) 成立. 为将 B_t 延拓为定义在 $[0,\infty)$ 上的过程，下面证明：

定理 7.1.2. 令 $T<\infty$ 且 $x\in\mathbb{R}$. ν_x 赋予路径 $\omega:\mathbb{Q}_2\to\mathbb{R}$ 在 $\mathbb{Q}_2\cap[0,T]$ 上**一致连续**的概率为 1.

注释. 证明定理 7.1.2 需要相当多工作. 在着手这项任务之前，先处理最后一个测度论细节：通过将概率测度转移到 $(C,\mathcal{C})$ 来整理局面，其中 $C=\{\text{连续}\omega:[0,\infty)\to\mathbb{R}\}$，$\mathcal{C}$ 是由坐标映射 $t\mapsto\omega(t)$ 生成的 σ-域. 为此，注意到将 Ω_q 中一致连续的点映到 C 中其唯一连续延拓的映射 ψ 可测，并令

$$P_x=\nu_x\circ\psi^{-1}.$$

上述构造保证了 $B_t(\omega)=\omega_t$ 对 $t\in\mathbb{Q}_2$ 具有正确的有限维分布. 路径的连续性和一个简单的极限论证表明，这对 $t\in[0,\infty)$ 也成立. 最后，读者应注意，如同 Markov 链的情形，有一族随机变量 $B_t(\omega)=\omega(t)$ 和一族概率测度 P_x（$x\in\mathbb{R}$），使得在 P_x 下，B_t 是满足 $P_x(B_0=x)=1$ 的布朗运动.

证明. 由平移不变性及尺度变换 (7.1.1)，不妨设 $B_0=0$ 并对 $T=1$ 证明结果. 在此情形下，定义的 (b) 部分和尺度关系蕴含

$$E_0(|B_t-B_s|^4)=E_0|B_{t-s}|^4=C(t-s)^2$$

其中 $C=E_0|B_1|^4<\infty$. 由最后一个观察，通过使用 Kolmogorov 的如下结果得到所需的**一致连续性**.

感谢 Robin Pemantle 简化证明以及 Timo Seppäläinen 纠正简化. □

定理 7.1.3. 假设 $E|X_s-X_t|^\beta\leqslant K|t-s|^{1+\alpha}$，其中 $\alpha,\beta>0$. 若 $\gamma<\alpha/\beta$，则以概率 1 存在常数 $C(\omega)$ 使得

$$|X(q)-X(r)|\leqslant C|q-r|^\gamma\quad\text{对所有}q,r\in\mathbb{Q}_2\cap[0,1].$$

证明. 令 $G_n=\{|X(i/2^n)-X((i-1)/2^n)|\leqslant 2^{-\gamma n}\text{ 对所有}0<i\leqslant 2^n\}$. **Chebyshev 不等式**蕴含 $P(|Y|>a)\leqslant a^{-\beta}E|Y|^\beta$，故若令 $\lambda=\alpha-\beta\gamma>0$，则

$$P(G_n^c)\leqslant 2^n\cdot 2^{n\beta\gamma}\cdot E|X(j2^{-n})-X(i2^{-n})|^\beta=K2^{-n\lambda}.$$

引理 7.1.4. 在 $H_N=\bigcap_{n=N}^\infty G_n$ 上，对满足 $|q-r|<2^{-N}$ 的 $q,r\in\mathbb{Q}_2\cap[0,1]$ 有

$$|X(q)-X(r)|\leqslant\frac{3}{1-2^{-\gamma}}|q-r|^\gamma.$$

引理 7.1.4 的证明. 令 $q,r\in\mathbb{Q}_2\cap[0,1]$ 且 $0<r-q<2^{-N}$. 令 $I_i^k=[(i-1)/2^k,i/2^k)$，令 m 是使得 q 和 r 位于不同区间 I_i^k 的 k 的最小值. 由于它们在 $k-1$ 层位于同一区间，$q\in I_i^m$ 且 $r\in I_{i+1}^m$. 可写

$$r=i2^{-m}+2^{-r(1)}+\cdots+2^{-r(\ell)},$$
$$q=i2^{-m}-2^{-q(1)}-\cdots-2^{-q(k)}$$

其中 $N<r(1)<\cdots<r(\ell)$ 且 $N<q(1)<\cdots<q(k)$. 为理解这一点，注意当进一步细分时，r 将位于左半或右半. 当它位于右半时添加一项.

在 H_N 上

$$|X(q)-X((i-1)2^{-m})|\leqslant\sum_{h=1}^{k}(2^{-q(h)})^\gamma\leqslant\sum_{h=m}^{\infty}(2^{-\gamma})^h=\frac{2^{-\gamma m}}{1-2^{-\gamma}}$$

$$|X(r)-X(i2^{-m})|\leqslant\frac{2^{-\gamma m}}{1-2^{-\gamma}}.$$

将最后三个不等式与 $2^{-m}\leqslant|q-r|$ 和 $1-2^{-\gamma}>1$ 结合，引理 7.1.4 得证. □

为证明定理 7.1.3，注意

$$P(H_N^c)\leqslant\sum_{n=N}^{\infty}P(G_n^c)\leqslant K\sum_{n=N}^{\infty}2^{-n\lambda}=K2^{-N\lambda}/(1-2^{-\lambda}).$$

由于 $\sum_{N=1}^{\infty}P(H_N^c)<\infty$，**Borel-Cantelli 引理**（定理 2.3.1）蕴含

$$|X(q)-X(r)|\leqslant A|q-r|^\gamma$$

对满足 $|q-r|<\delta(\omega)$ 的 $q,r\in\mathbb{Q}_2$.

为将此延拓到 $q,r\in\mathbb{Q}_2\cap[0,1]$，令 $s_0=q<s_1<\cdots<s_n=r$ 且 $|s_i-s_{i-1}|<\delta(\omega)$，并利用三角不等式推出 $|X(q)-X(r)|\leqslant C(\omega)|q-r|^\gamma$，其中 $C(\omega)=1+\delta(\omega)^{-1}$. □

尺度关系 (7.1.1) 蕴含

$$E|B_t-B_s|^{2m}=C_m|t-s|^m$$

其中 $C_m=E|B_1|^{2m}$，故在定理 7.1.3 中取 $\beta=2m$，$\alpha=m-1$ 并令 $m\to\infty$，得到 Wiener (1923) 的一个结果.

定理 7.1.5. 布朗路径对任意指数 $\gamma<1/2$ 是 **Hölder 连续**的.

容易证明：

定理 7.1.6. 以概率 1，布朗路径在任何点都不是 Lipschitz 连续的（因此不可微）.

注释. 布朗路径的不可微性是由 Paley, Wiener 和 Zygmund (1933) 发现的. Paley 于 1933 年在论文付印期间的一次滑雪事故中去世，年仅 26 岁. 下面给出的证明归功于 Dvoretsky, Erdös 和 Kakutani (1961).

证明. 固定常数 $C<\infty$，令 $A_n=\{\omega:$ 存在$s\in[0,1]$ 使得$|B_t-B_s|\leqslant C|t-s|$ 当$|t-s|\leqslant 3/n\}$. 对 $1\leqslant k\leqslant n-2$，令

$$Y_{k,n}=\max\left\{B\left(\frac{k+j}{n}\right)-B\left(\frac{k+j-1}{n}\right):j=0,1,2\right\}$$

$$B_n=\{\text{至少一个}Y_{k,n}\leqslant 5C/n\}.$$

三角不等式蕴含 $A_n\subset B_n$. 最坏的情形是 $s=1$. 取 $k=n-2$ 并注意

$$\left|B\left(\frac{n-2}{n}\right)-B\left(\frac{n-3}{n}\right)\right|\leqslant\left|B\left(\frac{n-3}{n}\right)-B(1)\right|+\left|B(1)-B\left(\frac{n-2}{n}\right)\right|\leqslant C(3/n+2/n).$$

利用 $A_n\subset B_n$ 和尺度关系 (7.1.1) 得

$$P(A_n)\leqslant P(B_n)\leqslant nP(|B(1/n)|\leqslant 5C/n)^3=nP(|B(1)|\leqslant 5C/n^{1/2})^3\leqslant n\{(10C/n^{1/2})\cdot(2\pi)^{-1/2}\}^3$$

由于 $\exp(-x^2/2)\leqslant 1$. 令 $n\to\infty$ 表明 $P(A_n)\to 0$. 注意到 $n\mapsto A_n$ 递增，表明对所有 n 有 $P(A_n)=0$，证毕. □

注释. 令 $H_\gamma(\omega)$ 为路径 $\omega\in C$ 是 γ 阶 Hölder 连续的时刻集. 定理 7.1.5 表明对 $\gamma<1/2$ 有 $P(H_\gamma=[0,\infty))=1$. 习题 7.1.5 表明对 $\gamma>1/2$ 有 $P(H_\gamma=\emptyset)=1$. 习题 7.2.4 表明对每个 t 有 $P(t\in H_{1/2})=0$，但 B. Davis (1983) 证明了 $P(H_{1/2}\neq\emptyset)=1$. Perkins (1983) 计算了

$$\left\{t\in(0,1):\limsup_{h\downarrow 0}\frac{|B_{t+h}-B_t|}{h^{1/2}}\leqslant c\right\}$$

的 Hausdorff 维数.

多维布朗运动

本节的所有结果都是关于一维布朗运动的. 为了定义从 $x\in\mathbb{R}^d$ 出发的 d 维布朗运动，令 $B_t^1,\ldots,B_t^d$ 为独立的布朗运动且 $B_0^i=x_i$. 如同 $d=1$ 的情形，它们被实现为概率测度 P_x 在 $(C,\mathcal{C})$ 上，其中 $C=\{$连续$\omega:[0,\infty)\to\mathbb{R}^d\}$，$\mathcal{C}$ 是由坐标映射生成的 σ-域. 由于坐标是独立的，容易看出有限维分布满足 (7.1.2)，转移概率为

$$p_t(x,y)=(2\pi t)^{-d/2}\exp(-|y-x|^2/2t). \tag{7.1.4}$$

习题

习题 7.1.1. 给定 $s < t$，求 $P(B(s) > 0, B(t) > 0)$. 提示：用 $B(s)$ 和 $B(t) - B(s)$ 表示该概率.

习题 7.1.2. 求 $E(B_1^2 B_2 B_3)$.

习题 7.1.3. 令 $W = \int_0^t B_s \, \mathrm{d}s$. 求 EW 和 EW^2. W 的分布是什么?

习题 7.1.4. $A \in \mathcal{F}_o$ 当且仅当存在时间序列 $t_1, t_2, \ldots$ 于 $[0, \infty)$ 中和 $B \in \mathcal{R}^{\{1,2,\ldots\}}$ 使得 $A = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in B\}$. 换言之，$\mathcal{F}_o$ 中的所有事件仅依赖于可数个坐标.

习题 7.1.5. 仔细审视定理 7.1.6 的证明可知，若 $\gamma > 5/6$，则 B_t 在 $[0, 1]$ 中任何点都不是 γ 阶 Hölder 连续的. 通过考虑 k 个增量而非 3 个，证明最后一个结论对所有 $\gamma > 1/2 + 1/k$ 成立.

习题 7.1.6. 固定 t 并令 $\Delta_{m,n} = B(tm2^{-n}) - B(t(m-1)2^{-n})$. 计算

$$E\left(\sum_{m \leqslant 2^n} \Delta_{m,n}^2 - t\right)^2$$

并用 Borel-Cantelli 推出 $\sum_{m \leqslant 2^n} \Delta_{m,n}^2 \to t$ 几乎必然当 $n \to \infty$.

注释. 最后一个结果对任何满足 $\Pi_1 \subset \Pi_2 \subset \cdots$ 且 mesh $\to 0$ 的分划序列都成立. 见 Freedman (1971a) p. 42–46. 然而，真正的二**次变差**，定义为对所有分划的上确界，是 ∞.

第 2 节 Markov 性质，Blumenthal 0-1 律

直观而言，**Markov 性质**表明“若 $s \geqslant 0$，则 $B(t+s)-B(s)$ $(t \geqslant 0)$ 是一个独立于时刻 s 之前发生之事的布朗运动.” 将其精确化的第一步是解释“时刻 s 之前发生的事” 的含义. 首先想到的是

$$\mathcal{F}_s^o = \sigma(B_r : r \leqslant s).$$

随着深入，原因将变得清晰，用

$$\mathcal{F}_{s+} = \bigcap_{t>s} \mathcal{F}_t^o$$

代替 $\mathcal{F}_s^o$ 更方便. 域 $\mathcal{F}_{s+}$ 更好，因为它们是右连续的：

$$\bigcap_{t>s} \mathcal{F}_{t+} = \bigcap_{t>s} \left(\bigcap_{u>t} \mathcal{F}_u^o \right) = \bigcap_{u>s} \mathcal{F}_u^o = \mathcal{F}_{s+}.$$

换言之，$\mathcal{F}_{s+}$ 允许“无穷小地窥视未来”，即 $A \in \mathcal{F}_{s+}$ 若对任意 $\varepsilon > 0$ 有 $A \in \mathcal{F}_{s+\varepsilon}^o$. 若对所有 $u > 0$ 有 $f(u) > 0$，则在 $d = 1$ 时随机变量

$$\limsup_{t \downarrow s} \frac{B_t - B_s}{f(t-s)}$$

关于 $\mathcal{F}_{s+}$ 可测但不对 $\mathcal{F}_s^o$ 可测. 下面将看到没有有趣的例子，即 $\mathcal{F}_{s+}$ 和 $\mathcal{F}_s^o$ 是相同的（在零集意义下）.

为陈述 Markov 性质，需要一些记号. 回顾有一族测度 P_x $(x \in \mathbb{R}^d)$ 在 $(C, \mathcal{C})$ 上使得在 P_x 下，$B_t(\omega) = \omega(t)$ 是从 x 出发的布朗运动. 对 $s \geqslant 0$，定义**移位变换** $\theta_s : C \to C$ 为

$$(\theta_s \omega)(t) = \omega(s+t) \quad \text{对} t \geqslant 0.$$

换言之，切除时刻 s 之前的路径部分，然后将路径移位使得时刻 s 变成时刻 0.

定理 7.2.1. Markov 性质. 若 $s \geqslant 0$ 且 Y 有界且 $\mathcal{C}$ 可测，则对所有 $x \in \mathbb{R}^d$

$$E_x(Y \circ \theta_s | \mathcal{F}_{s+}) = E_{B_s} Y$$

其中右端是函数 $\varphi(x) = E_x Y$ 在 $x = B_s$ 处的取值.

证明. 由条件期望的定义，需要证明

$$E_x(Y \circ \theta_s; A) = E_x(E_{B_s} Y; A) \quad \text{对所有} A \in \mathcal{F}_{s+}. \tag{7.2.1}$$

下面先对一个精心选择的特殊情形证明结果，然后利用**单调类定理**（MCT）得到一般情形.

假设 $Y(\omega)=\prod_{1\leqslant m\leqslant n} f_m(\omega(t_m))$，其中 $0<t_1<\cdots<t_n$ 且 f_m 有界可测. 令 $0<h<t_1$，$0<s_1,\ldots,s_k\leqslant s+h$，令 $A=\{\omega:\omega(s_j)\in A_j,1\leqslant j\leqslant k\}$，其中 $A_j\in\mathcal{R}$ 对 $1\leqslant j\leqslant k$. 由布朗运动的定义，

$$E_x(Y\circ\theta_s;A)=\int_{A_1}\mathrm{d}x_1\,p_{s_1}(x,x_1)\int_{A_2}\mathrm{d}x_2\,p_{s_2-s_1}(x_1,x_2)\cdots$$
$$\int_{A_k}\mathrm{d}x_k\,p_{s_k-s_{k-1}}(x_{k-1},x_k)\int\mathrm{d}y\,p_{s+h-s_k}(x_k,y)\varphi(y,h)$$

其中

$$\varphi(y,h)=\int\mathrm{d}y_1\,p_{t_1-h}(y,y_1)f_1(y_1)\cdots\int\mathrm{d}y_n\,p_{t_n-t_{n-1}}(y_{n-1},y_n)f_n(y_n).$$

更多细节见 (5.2.3) 的证明，它不加修改地适用于此处. 在右端利用该恒等式，有

$$E_x(Y\circ\theta_s;A)=E_x(\varphi(B_{s+h},h);A). \tag{7.2.2}$$

最后一个等式对所有有限维集 A 成立，由 $\pi-\lambda$ 定理（定理 2.1.6）知它对 $A\in\mathcal{F}^o_{s+h}\supset\mathcal{F}_{s+}$ 有效. 由归纳法容易看出

$$\psi(y_1)=f_1(y_1)\int\mathrm{d}y_2\,p_{t_2-t_1}(y_1,y_2)f_2(y_2)\cdots$$
$$\int\mathrm{d}y_n\,p_{t_n-t_{n-1}}(y_{n-1},y_n)f_n(y_n)$$

是有界可测的. 令 $h\downarrow 0$ 并利用控制收敛定理表明，若 $x_h\to x$，则

$$\varphi(x_h,h)=\int\mathrm{d}y_1\,p_{t_1-h}(x_h,y_1)\psi(y_1)\to\varphi(x,0)$$

当 $h\downarrow 0$. 利用 (7.2.2) 和控制收敛定理得

$$E_x(Y\circ\theta_s;A)=E_x(\varphi(B_s,0);A)$$

对所有 $A\in\mathcal{F}_{s+}$. 这表明 (7.2.1) 对 $Y=\prod_{1\leqslant m\leqslant n} f_m(\omega(t_m))$ 且 f_m 有界可测成立.

所需结论由单调类定理 5.2.2 推出. 令 $\mathcal{H}$ 为使 (7.2.1) 成立的有界函数全体. $\mathcal{H}$ 显然具有性质 (ii) 和 (iii). 令 $\mathcal{A}$ 为形如 $\{\omega:\omega(t_j)\in A_j\}$ 之集，其中 $A_j\in\mathcal{R}$. 上面处理的特殊情形表明 (i) 成立，所需结论得证. □

下面将看到 Markov 性质的诸多应用，因此现在把注意力转向一个“平凡性”，它有令人惊讶的推论. 由于

$$E_x(Y\circ\theta_s|\mathcal{F}_{s+})=E_{B(s)}Y\in\mathcal{F}^o_s$$

由定理 4.1.12 得

$$E_x(Y\circ\theta_s|\mathcal{F}_{s+})=E_x(Y\circ\theta_s|\mathcal{F}_s^o).$$

由最后一个方程，短步骤即可得到：

定理 7.2.2. 若 $Z\in\mathcal{C}$ 有界，则对所有 $s\geqslant 0$ 和 $x\in\mathbb{R}^d$，

$$E_x(Z|\mathcal{F}_{s+})=E_x(Z|\mathcal{F}_s^o).$$

证明. 如在定理 7.2.1 的证明中，只需对

$$Z=\prod_{m=1}^{n}f_m(B(t_m))$$

且 f_m 有界可测时证明结果. 在此情形下，Z 可写为 $X(Y\circ\theta_s)$，其中 $X\in\mathcal{F}_s^o$ 且 Y 是 $\mathcal{C}$ 可测的，故

$$E_x(Z|\mathcal{F}_{s+})=XE_x(Y\circ\theta_s|\mathcal{F}_{s+})=XE_{B_s}Y\in\mathcal{F}_s^o$$

证毕. □

若令 $Z\in\mathcal{F}_{s+}$，则定理 7.2.2 蕴含 $Z=E_x(Z|\mathcal{F}_s^o)\in\mathcal{F}_s^o$，故两个 σ-域在零集意义下相同. 乍看之下，这一结论并不令人兴奋. 乐趣从在定理 7.2.2 中取 $s=0$ 开始：

定理 7.2.3. Blumenthal 0-1 律. 若 $A\in\mathcal{F}_{0+}$，则对所有 $x\in\mathbb{R}^d$，

$$P_x(A)\in\{0,1\}.$$

证明. 利用 $A\in\mathcal{F}_{0+}$，定理 7.2.2，以及 $\mathcal{F}_0^o=\sigma(B_0)$ 在 P_x 下是平凡的，得

$$1_A=E_x(1_A|\mathcal{F}_{0+})=E_x(1_A|\mathcal{F}_0^o)=P_x(A)\quad P_x\text{ a.s.}$$

这表明指示函数 1_A 几乎必然等于数 $P_x(A)$，结果得证. □

换言之，最后一个结果说**芽 σ-域** $\mathcal{F}_{0+}$ 是平凡的. 这一结果在研究布朗路径的局部行为时非常有用. 在本节其余部分，将注意力限制在 $d=1$.

定理 7.2.4. 若 $\tau=\inf\{t\geqslant 0:B_t>0\}$，则 $P_0(\tau=0)=1$.

证明. $P_0(\tau\leqslant t)\geqslant P_0(B_t>0)=1/2$，由于正态分布关于 0 对称. 令 $t\downarrow 0$，得

$$P_0(\tau=0)=\lim_{t\downarrow 0}P_0(\tau\leqslant t)\geqslant 1/2$$

故由定理 7.2.3 得 $P_0(\tau = 0) = 1$. □

一旦布朗运动从 0 出发必须立即击中 $(0, \infty)$，它也必须立即击中 $(-\infty, 0)$. 由于 $t \mapsto B_t$ 连续，这迫使：

定理 7.2.5. 若 $T_0 = \inf\{t > 0 : B_t = 0\}$，则 $P_0(T_0 = 0) = 1$.

定理 7.2.3 涉及 B_t 当 $t \to 0$ 时的行为. 借助一个技巧，可利用这一结果得到关于 $t \to \infty$ 时行为的信息.

定理 7.2.6. 若 B_t 是从 0 出发的布朗运动，则由 $X_0 = 0$ 和 $X_t = tB(1/t)$ $(t > 0)$ 定义的过程也是布朗运动.

证明. 这里将检验布朗运动的第二个定义. 为此，注意到：

(i) 若 $0 < t_1 < \cdots < t_n$，则 $(X(t_1), \ldots, X(t_n))$ 服从均值为 0 的多元正态分布.

(ii) $EX_s = 0$ 且若 $s < t$ 则

$$E(X_s X_t) = stE(B(1/s)B(1/t)) = s.$$

对 (iii)，注意到 X 显然在 $t \neq 0$ 处连续.

为处理 $t = 0$，首先观察到强大数定律蕴含 $B_n/n \to 0$ 当 $n \to \infty$ 通过整数. 为处理整数之间的值，注意到 Kolmogorov 不等式（定理 2.5.5）蕴含

$$P\left(\sup_{0<k\leqslant 2^m} |B(n + k2^{-m}) - B_n| > n^{2/3}\right) \leqslant n^{-4/3} E(B_{n+1} - B_n)^2.$$

令 $m \to \infty$，有

$$P\left(\sup_{u\in[n,n+1]} |B_u - B_n| > n^{2/3}\right) \leqslant n^{-4/3}.$$

由于 $\sum_n n^{-4/3} < \infty$，Borel-Cantelli 引理蕴含 $B_u/u \to 0$ 当 $u \to \infty$. 取 $u = 1/t$，有 $X_t \to 0$ 当 $t \to 0$. □

定理 7.2.6 使能够将 B_t 当 $t \to \infty$ 和 $t \to 0$ 时的行为联系起来. 将这一思想与 Blumenthal 0-1 律结合，得到一个非常有用的结果. 令

$$\mathcal{F}_t^0 = \sigma(B_s : s \geqslant t) = \text{时刻}t\text{ 的将来}$$

$$\mathcal{T} = \bigcap_{t\geqslant 0} \mathcal{F}_t^0 = \text{尾 } \sigma\text{-域}.$$

定理 7.2.7. 若 $A \in \mathcal{T}$，则要么 $P_x(A) \equiv 0$，要么 $P_x(A) \equiv 1$.

注释. 注意这比 Blumenthal 0-1 律的结论更强. 例子 $A = \{\omega : \omega(0) \in D\}$ 表明，对 $\mathcal{F}_{0+}$ 中的芽 σ-域，$P_x(A)$ 的值（在此情形为 $1_D(x)$）可能依赖于 x.

证明. 由于 B 的尾 σ-域与 X 的芽 σ-域相同，$P_0(A) \in \{0, 1\}$. 为改进到所给的结论，观察到 $A \in \mathcal{F}_1^0$，故 1_A 可写为 $1_D \circ \theta_1$. 应用 Markov 性质得

$$P_x(A) = E_x(1_D \circ \theta_1) = E_x(E_x(1_D \circ \theta_1 | \mathcal{F}_1)) = E_x(E_{B_1} 1_D) = \int (2\pi)^{-1/2} \exp(-(y-x)^2/2) P_y(D)\, \mathrm{d}y.$$

取 $x = 0$，若 $P_0(A) = 0$，则 $P_y(D) = 0$ 对关于 Lebesgue 测度几乎处处的 y，再利用公式表明 $P_x(A) = 0$ 对所有 x. 为处理 $P_0(A) = 1$ 的情形，观察到 $A^c \in \mathcal{T}$ 且 $P_0(A^c) = 0$，故最后一个结果蕴含 $P_x(A^c) = 0$ 对所有 x. □

下一个结果是定理 7.2.7 的一个典型应用.

定理 7.2.8. 令 B_t 为从 0 出发的一维布朗运动，则以概率 1，

$$\limsup_{t\to\infty} B_t/\sqrt{t} = \infty, \quad \liminf_{t\to\infty} B_t/\sqrt{t} = -\infty.$$

证明. 令 $K < \infty$. 由习题 2.3.1 和尺度变换

$$P_0(B_n/\sqrt{n} \geqslant K \text{ i.o.}) \geqslant \limsup_{n\to\infty} P_0(B_n \geqslant K\sqrt{n}) = P_0(B_1 \geqslant K) > 0$$

故定理 7.2.7 中的 0-1 律蕴含概率为 1. 由于 K 任意，这证明了第一个结果. 第二个由对称性推出. □

由定理 7.2.8、平移不变性和布朗运动的连续性，得到：

定理 7.2.9. 令 B_t 为一维布朗运动并令 $A = \bigcap_n \{B_t = 0 \text{ 对某个} t \geqslant n\}$. 则对所有 x 有 $P_x(A) = 1$.

换言之，一维布朗运动是常返的. 对任何出发点 x，它将 “无穷多次” 返回 0，即存在时间序列 $t_n \uparrow \infty$ 使得 $B_{t_n} = 0$. 必须小心解释引号中的短语，因为从 0 出发，B_t 将在时间 $\varepsilon > 0$ 之前无穷多次击中 0.

结束语. 随着关于 Blumenthal 0-1 律的讨论完成，$\mathcal{F}_{s+}$ 和 $\mathcal{F}_s^o$ 之间的区别不再重要，因此将对 σ-域作最后一次改进并去掉上标. 令

$$\mathcal{N}_x = \{A : A \subset D \text{ 且} P_x(D) = 0\}$$

$$\mathcal{F}_s^x = \sigma(\mathcal{F}_{s+} \cup \mathcal{N}_x)$$

$$\mathcal{F}_s = \bigcap_x \mathcal{F}_s^x.$$

$\mathcal{N}_x$ 是零集，$\mathcal{F}_s^x$ 是 P_x 的完备化 σ-域. 由于不希望 σ-域流依赖于初始状态，取所有 σ-域的交. 由定义可知 $\mathcal{F}_s$ 是右连续的.

习题

习题 7.2.1. 令 $T_0 = \inf\{s > 0 : B_s = 0\}$，$R = \inf\{t > 1 : B_t = 0\}$. R 代表右或返回. 利用时刻 1 的 Markov 性质得到

$$P_x(R > 1 + t) = \int p_1(x, y) P_y(T_0 > t)\, \mathrm{d}y. \tag{7.2.3}$$

习题 7.2.2. 令 $T_0 = \inf\{s > 0 : B_s = 0\}$，$L = \sup\{t \leqslant 1 : B_t = 0\}$. L 代表左或最后. 利用时刻 $0 < t < 1$ 的 Markov 性质推出

$$P_0(L \leqslant t) = \int p_t(0, y) P_y(T_0 > 1 - t)\, \mathrm{d}y. \tag{7.2.4}$$

习题 7.2.3. 若 $a < b$，则以概率 1，a 是 B_t 在 (a, b) 中局部极大值的极限. 因此 B_t 的局部极大值集几乎必然是一个稠密集. 然而，与零点集不同，它是可数的.

习题 7.2.4. (i) 假设对所有 $t > 0$ 有 $f(t) > 0$. 利用定理 7.2.3 推出 $\limsup_{t \downarrow 0} B(t)/f(t) = c$，$P_0$ a.s.，其中 $c \in [0, \infty]$ 是常数. (ii) 证明若 $f(t) = \sqrt{t}$，则 $c = \infty$，故以概率 1，布朗路径在 0 处不是 1/2 阶 Hölder 连续的.

第 3 节 停时，强 Markov 性质

推广离散时间中的定义，称取值于 $[0,\infty]$ 的随机变量 S 为**停时**，若对任意 $t\geqslant 0$ 有 $\{S<t\}\in\mathcal{F}_t$.

上述定义在 $\{S<t\}$ 和 $\{S\leqslant t\}$ 之间做出了选择. 这在离散时间中差别很大，但在连续时间中（对于右连续 σ-域流 $\mathcal{F}_t$）二者等价：若 $\{S\leqslant t\}\in\mathcal{F}_t$，则 $\{S<t\}=\bigcup_n\{S\leqslant t-1/n\}\in\mathcal{F}_t$. 若 $\{S<t\}\in\mathcal{F}_t$，则 $\{S\leqslant t\}=\bigcap_n\{S<t+1/n\}\in\mathcal{F}_t$.

第一个结论只要求 $t\mapsto\mathcal{F}_t$ 是递增的. 第二个依赖于 $t\mapsto\mathcal{F}_t$ 是右连续的这一事实. 下面的定理 7.3.2 和 7.3.3 表明，当验证某随机变量是停时时，知道这两种定义等价是很有用的.

定理 7.3.1. 若 G 为开集且 $T=\inf\{t\geqslant 0:B_t\in G\}$，则 T 是停时.

证明. 由于 G 是开集且 $t\mapsto B_t$ 连续，$\{T<t\}=\bigcup_{q<t}\{B_q\in G\}$，其中并集取遍所有有理数 q，故 $\{T<t\}\in\mathcal{F}_t$. 这里我们需要使用有理数来获得可数并，从而得到可测集. □

定理 7.3.2. 若 T_n 是一列停时且 $T_n\downarrow T$，则 T 是停时.

证明. $\{T<t\}=\bigcup_n\{T_n<t\}$. □

定理 7.3.3. 若 T_n 是一列停时且 $T_n\uparrow T$，则 T 是停时.

证明. $\{T\leqslant t\}=\bigcap_n\{T_n\leqslant t\}$. □

定理 7.3.4. 若 K 为闭集且 $T=\inf\{t\geqslant 0:B_t\in K\}$，则 T 是停时.

证明. 记 $B(x,r)=\{y:|y-x|<r\}$，令 $G_n=\bigcup_{x\in K}B(x,1/n)$，且令 $T_n=\inf\{t\geqslant 0:B_t\in G_n\}$. 由于 G_n 是开集，由定理 7.3.1 知 T_n 是停时. 断言当 $n\uparrow\infty$ 时，$T_n\uparrow T$. 为证明这一点，注意对所有 n 有 $T\geqslant T_n$，故 $\lim T_n\leqslant T$. 为证 $T\leqslant\lim T_n$，可设 $T_n\uparrow t<\infty$. 由于对所有 n 有 $B(T_n)\in\bar{G}_n$ 且 $B(T_n)\to B(t)$，可得 $B(t)\in K$，从而 $T\leqslant t$. □

定理 7.3.1 和 7.3.4 涵盖我们将要考虑的所有击中时. 利用 7.3.2 和第二个结果可以证明，闭集之并的击中时是停时. Borel 集的击中时是停时，但证明这一点非常困难.

下述结果将在强 Markov 性质的证明中用到：

定理 7.3.5. 设 S 为停时，令 $S_n=([2^nS]+1)/2^n$，其中 $[x]$ 表示不超过 x 的最大整数. 即

$$S_n=(m+1)2^{-n}\quad 若m2^{-n}\leqslant S<(m+1)2^{-n}.$$

换言之，我们在 S 之后（即 $>S$）的第一个形如 $k2^{-n}$ 的时刻停止. 从文字描述看，S_n 应该是停时.

证明. 若 $m/2^n<t\leqslant(m+1)/2^n$，则 $\{S_n<t\}=\{S\leqslant m/2^n\}\in\mathcal{F}_{m/2^n}\subset\mathcal{F}_t$. 不可能有 $S_n=0$. □

下面叙述并证明强 Markov 性质. 为此，需推广 4.1 节中的两个定义. 给定非负随机变量 $S(\omega)$，我们定义随机移位 θ_S，它 “切掉 ω 在 $S(\omega)$ 之前的部分，然后将路径平移使得时刻 $S(\omega)$ 变成时刻 0”. 用符号表示，我们令

$$(\theta_S\omega)(t) = \begin{cases} \omega(S(\omega)+t) & \text{在}\{S<\infty\}\text{ 上,} \\ \Delta & \text{在}\{S=\infty\}\text{ 上,} \end{cases}$$

其中 Δ 是我们添加到 C 的一个额外点. 如 6.3 节，我们通常明确地将注意力限制在 $\{S<\infty\}$ 上，因此读者不必担心定义的后半部分.

第二个量 $\mathcal{F}_S$，“时刻 S 已知的信息”，则更为微妙. 模仿 4.1 节中的离散时间定义，我们令

$$\mathcal{F}_S = \{A : A\cap\{S\leqslant t\}\in\mathcal{F}_t \text{ 对所有} t\geqslant 0\}.$$

换言之，这一要求是说 A 位于 $\{S\leqslant t\}$ 中的部分关于时刻 t 可用的信息可测. 同样，这里在 $\leqslant t$ 和 $<t$ 之间做出了选择，但与停时的情况一样，二者等价.

注释. 当 $\mathcal{F}_t$ 右连续时，若将 $\{S\leqslant t\}$ 替换为 $\{S<t\}$，上述定义不变.

我们将推导 $\mathcal{F}_S$ 的一些性质. 第一个是直观上显然的：在较晚时刻我们有更多信息.

定理 7.3.6. 若 $S\leqslant T$ 为停时，则 $\mathcal{F}_S\subset\mathcal{F}_T$.

证明. 若 $A\in\mathcal{F}_S$，则 $A\cap\{T\leqslant t\} = (A\cap\{S\leqslant t\})\cap\{T\leqslant t\}\in\mathcal{F}_t$. □

定理 7.3.7. 若 $T_n\downarrow T$ 为停时，则 $\mathcal{F}_T = \bigcap\mathcal{F}(T_n)$.

证明. 定理 7.3.6 蕴含 $\mathcal{F}(T_n)\supset\mathcal{F}_T$ 对所有 n 成立. 为证明另一包含关系，令 $A\in\bigcap\mathcal{F}(T_n)$. 由于 $A\cap\{T_n<t\}\in\mathcal{F}_t$ 且 $T_n\downarrow T$，可得 $A\cap\{T<t\}\in\mathcal{F}_t$. □

下一个结果显然但需要一些工作.

定理 7.3.8. $B_S\in\mathcal{F}_S$，即 B_S 的值关于时刻 S 已知的信息可测.

证明. 为证明这一点，令 $S_n = ([2^nS]+1)/2^n$ 为定理 7.3.5 中定义的停时. 若 A 为 Borel 集，

$$\{B(S_n)\in A\} = \bigcup_{m=1}^{\infty}\{S_n = m/2^n, B(m/2^n)\in A\}\in\mathcal{F}_{S_n}.$$

现令 $n\to\infty$ 并应用定理 7.3.7. □

下述结果即为强 Markov 性质，它表明 Markov 性质在停时处成立. 值得注意的是，布朗运动的概念可追溯到 20 世纪初，但强 Markov 性质的首次证明由 Hunt(1956) 以及 Dynkin 和 Yushkevich(1956) 独立给出. Hunt 写道 “尽管数学家们使用这一推广的 Markov 性质，至少作为启发性原理，但我在任何地方都未见到它被严格讨论.”

定理 7.3.9. 强 Markov 性质. 设 $(s,\omega)\mapsto Y_s(\omega)$ 为有界且 $\mathbb{R}\times C$ 可测的函数. 若 S 为停时，则对所有 $x\in\mathbb{R}^d$

$$E^x(Y_S\circ\theta_S|\mathcal{F}_S)=E^{B(S)}Y_S\quad \text{在}\{S<\infty\}\text{ 上,}$$

其中右边是在 $x=B(S),t=S$ 处计算的函数 $\varphi(x,t)=E^xY_t$.

注释. 这里使用的关于布朗运动的唯一事实是 (i) 它是 Markov 过程，(ii) 若 f 有界连续，则 $x\mapsto E^xf(B_t)$ 连续. 在 Markov 过程理论中，(ii) 称为 **Feller 性质**. 虽然 Hunt 的证明只适用于布朗运动，但 Dynkin 和 Yushkevich 在这一一般性下证明了这一结果.

证明. 我们首先假设存在一列时刻 $t_n\uparrow\infty$，使得 $P^x(S<\infty)=\sum P^x(S=t_n)$. 在这一情形下，证明基本上与定理 5.2.5 的证明相同. 我们按 S 的取值分解，应用 Markov 性质，然后将各部分重新组合. 若令 $Z_n=Y_{t_n}(\omega)$ 且 $A\in\mathcal{F}_S$，则

$$E^x(Y_S\circ\theta_S;A\cap\{S<\infty\})=\sum_{n=1}^{\infty}E^x(Z_n\circ\theta_{t_n};A\cap\{S=t_n\}).$$

现若 $A\in\mathcal{F}_S$，则 $A\cap\{S=t_n\}=(A\cap\{S\leqslant t_n\})-(A\cap\{S\leqslant t_{n-1}\})\in\mathcal{F}_{t_n}$，故由 Markov 性质，上式等于

$$=\sum_{n=1}^{\infty}E^x(E^{B(t_n)}Z_n;A\cap\{S=t_n\})=E^x(E^{B(S)}Y_S;A\cap\{S<\infty\}).$$

为证明一般情形的结果，我们令 $S_n=([2^nS]+1)/2^n$ 为习题 7.3.5 中定义的停时. 为能够令 $n\to\infty$，我们将注意力限制于如下形式的 Y：

$$Y_s(\omega)=f_0(s)\prod_{m=1}^{n}f_m(\omega(t_m)) \tag{7.3.1}$$

其中 $0<t_1<\cdots<t_n$ 且 $f_0,\dots,f_n$ 有界连续. 若 f 有界连续，则由控制收敛定理，

$$x\mapsto\int \mathrm{d}y\,p_t(x,y)f(y)$$

连续. 由此及归纳法，可得

$$\varphi(x,s)=E^xY_s=f_0(s)\int \mathrm{d}y_1\,p_{t_1}(x,y_1)f_1(y_1)\cdots\int \mathrm{d}y_n\,p_{t_n-t_{n-1}}(y_{n-1},y_n)f_n(y_n)$$

有界连续.

有了这些必要的准备，下面完成证明. 令 $A \in \mathcal{F}_S$. 由于 $S \leqslant S_n$，定理 7.3.6 蕴含 $A \in \mathcal{F}(S_n)$. 将上面证明的特殊情形应用于 S_n，并注意 $\{S_n < \infty\} = \{S < \infty\}$，得

$$E^x(Y_{S_n} \circ \theta_{S_n}; A \cap \{S < \infty\}) = E^x(\varphi(B(S_n), S_n); A \cap \{S < \infty\}).$$

现当 $n \to \infty$ 时，$S_n \downarrow S$，$B(S_n) \to B(S)$，$\varphi(B(S_n), S_n) \to \varphi(B(S), S)$，且 $Y_{S_n} \circ \theta_{S_n} \to Y_S \circ \theta_S$，故由有界收敛定理，当 Y 具有 (7.3.1) 给出的形式时，结果成立.

为完成证明，我们将应用单调类定理. 如定理 7.2.1 的证明中，令 $\mathcal{H}$ 为满足

$$E^x(Y_S \circ \theta_S; A) = E^x(E^{B(S)}Y_S; A) \quad \text{对所有} A \in \mathcal{F}_S$$

的 Y 的集合，易见 (ii) 和 (iii) 成立. 然而，这次我们取 $\mathcal{A}$ 为形如 $A = G_0 \times \{\omega : \omega(s_j) \in G_j, 1 \leqslant j \leqslant k\}$ 的集合，其中 G_j 为开集. 为验证 (i)，我们注意若 $K_j = G_j^c$ 且 $f_j^n(x) = 1 \wedge n\rho(x, K_j)$，其中 $\rho(x, K) = \inf\{|x-y| : y \in K\}$，则 f_j^n 是连续函数且当 $n \uparrow \infty$ 时 $f_j^n \uparrow 1_{G_j}$. 由 $Y_s^n(\omega) = f_0^n(s) \prod_{j=1}^k f_j^n(\omega(s_j)) \in \mathcal{H}$ 以及 (iii) 对 $\mathcal{H}$ 成立，可得 $1_A \in \mathcal{H}$. 这验证了单调类定理中的 (i) 并完成了证明. □

习题

习题 7.3.1. 设 A 为 F_σ 集，即可数闭集的并. 证明 $T_A = \inf\{t : B_t \in A\}$ 是停时.

习题 7.3.2. 若 S 和 T 是停时，则 $S \wedge T = \min\{S, T\}$，$S \vee T = \max\{S, T\}$，以及 $S + T$ 也是停时. 特别地，若 $t \geqslant 0$，则 $S \wedge t$，$S \vee t$ 和 $S + t$ 是停时.

习题 7.3.3. 设 T_n 为一列停时. 证明 $\sup_n T_n$，$\inf_n T_n$，$\limsup_n T_n$，$\liminf_n T_n$ 是停时.

习题 7.3.4. 设 S 为停时，$A \in \mathcal{F}_S$，且令 $R = S$ 在 A 上，$R = \infty$ 在 A^c 上. 证明 R 是停时.

习题 7.3.5. 设 S 和 T 为停时. $\{S < T\}$，$\{S > T\}$，以及 $\{S = T\}$ 属于 $\mathcal{F}_S$（且属于 $\mathcal{F}_T$）.

习题 7.3.6. 设 S 和 T 为停时. $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}$.

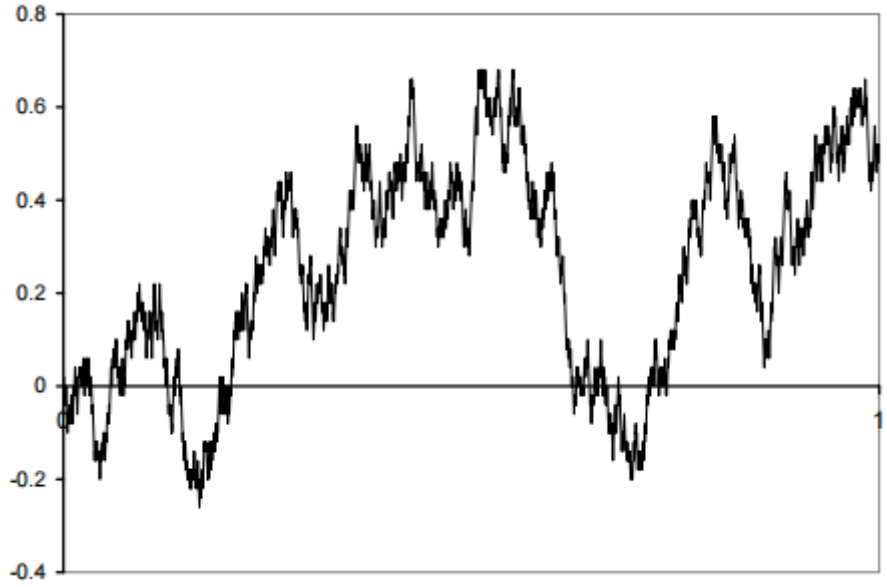

图 7.2: 一维布朗运动的模拟

第 4 节 路径性质

本节利用强 Markov 性质推导零点集 $\{t : B_t = 0\}$、击中时 $T_a = \inf\{t : B_t = a\}$，以及 $\max_{0\leqslant s\leqslant t} B_s$ 对于一维布朗运动的性质.

7.4.1 布朗运动的零点

令 $R_t = \inf\{u > t : B_u = 0\}$ 且 $T_0 = \inf\{u > 0 : B_u = 0\}$. 由定理 7.2.9 知 $P^x(R_t < \infty) = 1$，故 $B(R_t) = 0$，再由强 Markov 性质及定理 7.2.5 得

$$P^x(T_0 \circ \theta_{R_t} > 0 | \mathcal{F}_{R_t}) = P^0(T_0 > 0) = 0.$$

对最后一个方程取期望得

$$P^x(T_0 \circ \theta_{R_t} > 0 \text{ 对某个有理数} t) = 0.$$

由此可得，若点 $u \in Z(\omega) \equiv \{t : B_t(\omega) = 0\}$ 在左侧孤立（即存在有理数 $t < u$ 使得 $(t, u) \cap Z(\omega) = \emptyset$），则它以概率 1 是 $Z(\omega)$ 中点的递减极限. 这表明闭集 $Z(\omega)$ 无孤立点，因此必为不可数集. 最后一步见 Hewitt 与 Stromberg(1965)，第 72 页.

若令 $|Z(\omega)|$ 表示 $Z(\omega)$ 的 Lebesgue 测度，则由 Fubini 定理

$$E^x(|Z(\omega)| \cap [0, T]) = \int_0^T P^x(B_t = 0)\, \mathrm{d}t = 0.$$

故 $Z(\omega)$ 是零测集.

以上四点表明，Z 类似于 Cantor 集：从 $[0,1]$ 中移除 $(1/3,2/3)$，然后反复从剩余区间中移除中间的三分之一. 然而 Cantor 集更大. 其 Hausdorff 维数为 $\log 2/\log 3$，而 Z 的维数为 $1/2$.

7.4.2 击中时

定理 7.4.1. 在 P^0 下，$\{T_a : a \geqslant 0\}$ 具有平稳独立增量.

证明. 第一步是注意若 $0 < a < b$，则

$$T_b \circ \theta_{T_a} = T_b - T_a,$$

故若 f 有界可测，由强 Markov 性质 7.3.9 及平移不变性，

$$E^0(f(T_b - T_a)|\mathcal{F}_{T_a}) = E^0(f(T_b) \circ \theta_{T_a}|\mathcal{F}_{T_a}) = E^a f(T_b) = E^0 f(T_{b-a}).$$

为证明增量独立，令 $a_0 < a_1 \cdots < a_n$，令 f_i （$1 \leqslant i \leqslant n$）有界可测，且令 $F_i = f_i(T_{a_i} - T_{a_{i-1}})$. 对 $\mathcal{F}_{T_{a_{n-1}}}$ 取条件并用上述计算，有

$$E^0\left(\prod_{i=1}^{n} F_i\right) = E^0\left(\prod_{i=1}^{n-1} F_i \cdot E^0(F_n|\mathcal{F}_{T_{a_{n-1}}})\right) = E^0\left(\prod_{i=1}^{n-1} F_i\right) E^0 F_n.$$

由归纳法，$E^0 \prod_{i=1}^{n} F_i = \prod_{i=1}^{n} E^0 F_i$，这蕴含所需结论. □

尺度关系 (7.1.1) 蕴含

$$T_a \overset{d}{=} a^2 T_1. \tag{7.4.1}$$

结合定理 7.4.1 和 (7.4.1)，可见 $t_k = T_k - T_{k-1}$ 独立同分布且

$$\frac{t_1 + \cdots + t_n}{n^2} \to T_1,$$

故由定理 3.8.8，T_a 具有稳定律. 由于我们用 n^2 作除法且 $T_a \geqslant 0$，指数 $\alpha = 1/2$，偏度参数 $\kappa = 1$，见 (3.8.11).

不借助上一段提到的理论，也可直接确定 Laplace 变换

$$\varphi_a(\lambda) = E^0 \exp(-\lambda T_a) \quad 对 a \geqslant 0$$

并得出同样的结论. 为此，我们首先观察到定理 7.4.1 蕴含

$$\varphi_x(\lambda)\varphi_y(\lambda) = \varphi_{x+y}(\lambda).$$

由此易得

$$\varphi_a(\lambda) = \exp(-ac(\lambda)). \tag{7.4.2}$$

证明. 令 $c(\lambda) = -\log \varphi_1(\lambda)$，故 (7.4.2) 对 $a = 1$ 成立. 用上一恒等式取 $x = y = 2^{-m}$ 并由归纳法，得结果对 $a = 2^{-m}$ $(m \geqslant 1)$ 成立. 然后，令 $x = k2^{-m}$ 和 $y = 2^{-m}$，得结果对 $a = (k+1)2^{-m}$ $(k \geqslant 1)$ 成立. 最后，为延拓到 $a \in [0, \infty)$，注意 $a \mapsto \varphi_a(\lambda)$ 是递减的. □

为确定 $c(\lambda)$，我们观察到 (7.4.1) 蕴含

$$E \exp(-T_a) = E \exp(-a^2 T_1),$$

故 $ac(1) = c(a^2)$，即 $c(\lambda) = c(1)\sqrt{\lambda}$. 由于所有论证也适用于 σB_t，我们无法指望计算出 $c(1)$. 定理 7.5.7 将表明

$$E^0(\exp(-\lambda T_a)) = \exp(-a\sqrt{2\lambda}). \tag{7.4.3}$$

下面计算击中时 T_a 的分布. 强 Markov 性质的这一应用表明为何允许应用于移位路径的函数 Y 依赖于停时 S.

例题 7.4.2. 反射原理. 设 $a > 0$ 且 $T_a = \inf\{t : B_t = a\}$. 则

$$P^0(T_a < t) = 2P^0(B_t \geqslant a). \tag{7.4.4}$$

直观证明. 我们观察到若 B_s 在某一时刻 $s < t$ 击中 a，则由强 Markov 性质，$B_t - B(T_a)$ 独立于时刻 T_a 之前发生的事情. 正态分布的对称性及 $P^a(B_u = a) = 0$（对 $u > 0$）蕴含

$$P^0(T_a < t, B_t > a) = \frac{1}{2} P^0(T_a < t). \tag{7.4.5}$$

重排最后一个方程并用 $\{B_t > a\} \subset \{T_a < t\}$，得

$$P^0(T_a < t) = 2P^0(T_a < t, B_t > a) = 2P^0(B_t > a).$$

证明. 为使直观证明严格，我们只需证明 (7.4.5). 为从强 Markov 性质（定理 7.3.9）中提取这一点，我们令

$$Y_s(\omega) = \begin{cases} 1 & \text{若} s < t, \omega(t-s) > a, \\ 0 & \text{其他}. \end{cases}$$

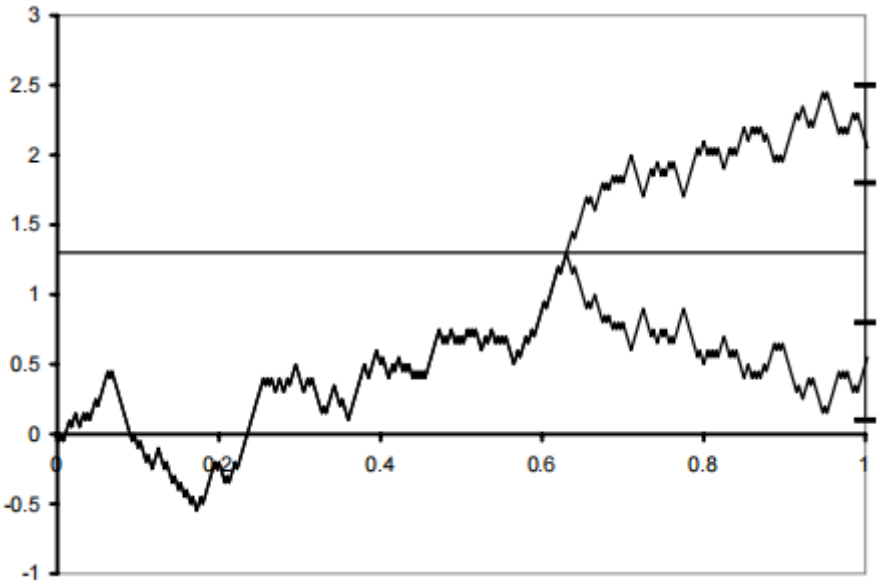

图 7.3: 反射原理的图片证明

我们这样做是为了若令 $S = \inf\{s < t : B_s = a\}$（约定 $\inf\emptyset = \infty$），则

$$Y_S(\theta_S\omega) = \begin{cases} 1 & \text{若} S < t, B_t > a, \\ 0 & \text{其他}, \end{cases}$$

强 Markov 性质蕴含

$$E^0(Y_S \circ \theta_S | \mathcal{F}_S) = \varphi(B_S, S) \quad \text{在} \{S < \infty\} = \{T_a < t\} \text{ 上},$$

其中 $\varphi(x, s) = E^x Y_s$. 在 $\{S < \infty\}$ 上 $B_S = a$ 且 $\varphi(a, s) = 1/2$（若 $s < t$），故取期望值得

$$P^0(T_a < t, B_t \geqslant a) = E^0(Y_S \circ \theta_S; S < \infty) = E^0(E^0(Y_S \circ \theta_S | \mathcal{F}_S); S < \infty) = E^0(1/2; T_a < t),$$

这证明了 (7.4.5). □

利用 (7.4.4)，我们可以计算 T_a 的概率密度. 我们首先注意

$$P(T_a \leqslant t) = 2P^0(B_t \geqslant a) = 2\int_a^\infty (2\pi t)^{-1/2} \exp(-x^2/2t)\, \mathrm{d}x,$$

然后换元 $x=(t^{1/2}a)/s^{1/2}$，得

$$\begin{aligned} P^0(T_a\leqslant t) &= 2\int_0^t (2\pi t)^{-1/2}\exp(-a^2/2s)\left(-\frac{t^{1/2}a}{2s^{3/2}}\right)\mathrm{d}s \\ &= \int_0^t (2\pi s^3)^{-1/2}a\exp(-a^2/2s)\,\mathrm{d}s. \end{aligned} \tag{7.4.6}$$

利用最后一个公式，可计算：

例题 7.4.3. $L=\sup\{t\leqslant 1: B_t=0\}$ 的分布. 由 (7.2.4)，

$$\begin{aligned} P^0(L\leqslant s) &= \int_{-\infty}^{\infty} p_s(0,x)P^x(T_0>1-s)\,\mathrm{d}x \\ &= 2\int_0^{\infty}(2\pi s)^{-1/2}\exp(-x^2/2s)\int_{1-s}^{\infty}(2\pi r^3)^{-1/2}x\exp(-x^2/2r)\,\mathrm{d}r\,\mathrm{d}x \\ &= \frac{1}{\pi}\int_{1-s}^{\infty}(sr^3)^{-1/2}\int_0^{\infty}x\exp(-x^2(r+s)/2rs)\,\mathrm{d}x\,\mathrm{d}r \\ &= \frac{1}{\pi}\int_{1-s}^{\infty}(sr^3)^{-1/2}\frac{rs}{r+s}\,\mathrm{d}r. \end{aligned}$$

下一步是令 $t=s/(r+s)$，将关于 $r\in[1-s,\infty)$ 的积分转化为关于 $t\in[0,s]$ 的积分. $\mathrm{d}t=-s/(r+s)^2\mathrm{d}r$，故为简化计算，我们首先将积分重写为

$$=\frac{1}{\pi}\int_{1-s}^{\infty}\left(\frac{r+s}{rs}\right)^{1/2}\frac{s}{(r+s)^2}\,\mathrm{d}r,$$

然后换元得

$$P^0(L\leqslant s)=\frac{1}{\pi}\int_0^s (t(1-t))^{-1/2}\,\mathrm{d}t=\frac{2}{\pi}\arcsin(\sqrt{s}). \tag{7.4.7}$$

由于 L 是时刻 1 之前最后零点的时刻，其密度关于 $1/2$ 对称且在 0 处发散，这令人惊讶. arcsin 可能使读者想起定理 4.9.5 中给出的 $L_{2n}=\sup\{m\leqslant 2n: S_m=0\}$ 的极限定理. 我们将在 8.6 节中看到，新结果是旧结果的一个推论.

习题

习题 7.4.1. 设 $B_t=(B_t^1,B_t^2)$ 为从 0 出发的二维布朗运动. 令 $T_a=\inf\{t: B_t^2=a\}$. 证明 (i) $B^1(T_a)$ $(a\geqslant 0)$ 具有独立增量且 $B^1(T_a)\overset{d}{=}aB^1(T_1)$. 利用这一点得出 $B^1(T_a)$ 具有 Cauchy 分布.

习题 7.4.2. 利用 (7.2.3) 证明 $R=\inf\{t>1: B_t=0\}$ 具有概率密度

$$P^0(R=1+t)=1/(\pi t^{1/2}(1+t)).$$

习题 7.4.3. (a) 推广 (7.4.5) 的证明以得出结论：若 $u < v \leqslant a$，则

$$P^0(T_a < t, u < B_t < v) = P^0(2a - v < B_t < 2a - u). \tag{7.4.8}$$

(b) 在 (7.4.8) 中令 (u, v) 收缩到 x，得出对 $a < x$

$$P^0(T_a < t, B_t = x) = p_t(0, 2a - x),$$

即在 $\{T_a < t\}$ 上 B_t 的（次概率）密度. (c) 利用事实 $\{T_a < t\} = \{M_t > a\}$ 并对 a 求导，得联合密度

$$f_{(M_t, B_t)}(a, x) = \frac{2(2a - x)}{\sqrt{2\pi t^3}} e^{-(2a-x)^2/2t}.$$

习题 7.4.4. 设 $A_{s,t}$ 为布朗运动在 $[s, t]$ 中至少有一个零点的事件. 证 $P^0(A_{s,t}) = (2/\pi) \arccos(\sqrt{s/t})$. 提示：开始时注意

$$P^0(A_{s,t}) = 2 \int_0^\infty (2\pi s)^{-1/2} e^{-x^2/2} P^x(T_0 \leqslant t - s) \, \mathrm{d}x,$$

并利用 T_x 的密度函数的公式. 最后你可能需要一些三角公式将 arctan 转化为 arccos.

注释. 利用上一习题，容易看到若 $s \geqslant 1$，则 $P(A_{s,s+h}) \leqslant Ch^{1/2}$. 于是若将区间 $[1, 2]$ 分成 n 等份，则包含零点的区间的期望个数 $\leqslant Cn^{-1/2}$. 由此可得零点集的 Hausdorff 维数至多为 $1/2$.

第 5 节 鞅

5.7 节末尾曾用鞅研究随机游走的击中时. 同样的方法亦可应用于布朗运动，一旦证得以下结果:

定理 7.5.1. 设 X_t 为关于右连续 σ-域流的右连续鞅. 若 T 为有界停时，则 $EX_T = EX_0$.

证明. 设 n 为整数使得 $P(T \leqslant n-1) = 1$. 如强 Markov 性质之证，令 $T_m = ([2^m T] + 1)/2^m$. 则 $Y_k^m = X(k2^{-m})$ 关于 $\mathcal{F}_k^m = \mathcal{F}(k2^{-m})$ 为鞅，且 $S_m = 2^m T_m$ 为 $(Y_k^m, \mathcal{F}_k^m)$ 的停时，故由习题 4.4.4 知

$$X(T_m) = Y_{S_m}^m = E(Y_{n2^m}^m | \mathcal{F}_{S_m}^m) = E(X_n | \mathcal{F}(T_m))$$

当 $m \uparrow \infty$ 时，由右连续性 $X(T_m) \to X(T)$，且由定理 7.3.7 知 $\mathcal{F}(T_m) \downarrow \mathcal{F}(T)$，故由定理 4.7.3，得

$$X(T) = E(X_n | \mathcal{F}(T))$$

取期望得 $EX(T) = EX_n = EX_0$，因 X_n 为鞅. □

定理 7.5.2. B_t 是关于第 8.2 节中定义的 σ-域 $\mathcal{F}_t$ 的鞅.

注释. 所有鞅结果均使用这些 σ-域，但在陈述中将不再明确提及.

证明. Markov 性质给出

$$E_x(B_t | \mathcal{F}_s) = E_{B_s}(B_{t-s}) = B_s$$

因由对称性，对所有 $u \geqslant 0$ 有 $E_y B_u = y$. □

由定理 7.5.2 立得:

定理 7.5.3. 若 $a < x < b$，则 $P_x(T_a < T_b) = (b-x)/(b-a)$.

证明. 令 $T = T_a \wedge T_b$. 定理 7.2.8 给出 $T < \infty$ 几乎必然. 利用定理 7.5.1 和 7.5.2，有 $x = E_x B(T \wedge t)$. 令 $t \to \infty$ 并利用有界收敛定理，得

$$x = aP_x(T_a < T_b) + b(1 - P_x(T_a < T_b))$$

解出 $P_x(T_a < T_b)$ 即得结果. □

定理 7.5.4. $B_t^2 - t$ 是鞅.

证明. 写 $B_t^2=(B_s+B_t-B_s)^2$，则

$$\begin{aligned}E_x(B_t^2|\mathcal{F}_s)&=E_x(B_s^2+2B_s(B_t-B_s)+(B_t-B_s)^2|\mathcal{F}_s)\\&=B_s^2+2B_sE_x(B_t-B_s|\mathcal{F}_s)+E_x((B_t-B_s)^2|\mathcal{F}_s)\\&=B_s^2+0+(t-s)\end{aligned}$$

因 B_t-B_s 独立于 $\mathcal{F}_s$ 且均值为 0、方差为 $t-s$ □

定理 7.5.5. 设 $T=\inf\{t:B_t\notin(a,b)\}$，其中 $a<0<b$. 则 $E_0T=-ab$.

证明. 定理 7.5.1 和 7.5.4 蕴含 $E_0(B^2(T\wedge t))=E_0(T\wedge t)$. 令 $t\to\infty$ 并利用单调收敛定理得 $E_0(T\wedge t)\uparrow E_0T$. 由有界收敛定理和定理 7.5.3，有

$$E_0B^2(T\wedge t)\to E_0B_T^2=a^2\frac{b}{b-a}+b^2\frac{-a}{b-a}=ab\frac{a-b}{b-a}=-ab$$

□

定理 7.5.6. $\exp(\theta B_t-(\theta^2t/2))$ 是鞅.

证明. 将 $\exp(\theta B_s)$ 提出，有

$$\begin{aligned}E_x(\exp(\theta B_t)|\mathcal{F}_s)&=\exp(\theta B_s)E(\exp(\theta(B_t-B_s))|\mathcal{F}_s)\\&=\exp(\theta B_s)\exp(\theta^2(t-s)/2)\end{aligned}$$

因 B_t-B_s 独立于 $\mathcal{F}_s$ 且服从均值为 0、方差为 $t-s$ 的正态分布. □

定理 7.5.7. 若 $T_a=\inf\{t:B_t=a\}$，则 $E_0\exp(-\lambda T_a)=\exp(-a\sqrt{2\lambda})$.

证明. 由定理 7.5.1 和 7.5.6，$1=E_0\exp(\theta B(T_a\wedge t)-\theta^2(T_a\wedge t)/2)$. 取 $\theta=\sqrt{2\lambda}$，令 $t\to\infty$ 并利用有界收敛定理，得

$$1=E_0\exp(a\sqrt{2\lambda}-\lambda T_a).$$

□

定理 7.5.8. 若 $u(t,x)$ 为 t 和 x 的多项式且满足

$$\frac{\partial u}{\partial t}+\frac{1}{2}\frac{\partial^2u}{\partial x^2}=0 \tag{7.5.1}$$

则 $u(t,B_t)$ 是鞅.

证明. 令 $p_t(x,y)=(2\pi)^{-1/2}t^{-1/2}\exp(-(y-x)^2/2t)$. 首先验证 p_t 满足热方程 $\partial p_t/\partial t=(1/2)\partial^2 p_t/\partial y^2$.

$$\begin{aligned}\frac{\partial p}{\partial t}&=-\frac{1}{2}t^{-1}p_t+\frac{(y-x)^2}{2t^2}p_t\\ \frac{\partial p}{\partial y}&=-\frac{y-x}{t}p_t\\ \frac{\partial^2 p}{\partial y^2}&=-\frac{1}{t}p_t+\frac{(y-x)^2}{t^2}p_t\end{aligned}$$

交换 $\partial/\partial t$ 与 $\int$，并利用热方程，得

$$\begin{aligned}\frac{\partial}{\partial t}E_x u(t,B_t)&=\int\frac{\partial}{\partial t}(p_t(x,y)u(t,y))\,\mathrm{d}y\\ &=\int\frac{1}{2}\frac{\partial^2}{\partial y^2}p_t(x,y)u(t,y)+p_t(x,y)\frac{\partial}{\partial t}u(t,y)\,\mathrm{d}y\end{aligned}$$

分部积分两次，得

$$=\int p_t(x,y)\left(\frac{\partial}{\partial t}+\frac{1}{2}\frac{\partial^2}{\partial y^2}\right)u(t,y)\,\mathrm{d}y=0$$

因 $u(t,y)$ 是多项式，积分收敛无问题，分部积分时边界项亦无贡献.

至此已证 $t\mapsto E_x u(t,B_t)$ 为常数. 为验证鞅性质，只需证

$$E(u(t,B_t)|\mathcal{F}_s)=u(s,B_s)$$

为此，注意 $v(r,x)=u(s+r,x)$ 满足 $\partial v/\partial r=(1/2)\partial^2 v/\partial x^2$，于是利用 Markov 性质

$$\begin{aligned}E(u(t,B_t)|\mathcal{F}_s)&=E(v(t-s,B_{t-s})\circ\theta_s|\mathcal{F}_s)\\ &=E_{B(s)}v(t-s,B_{t-s})=v(0,B_s)=u(s,B_s)\end{aligned}$$

其中倒数第二步利用了已证的期望不变性. □

满足 (7.5.1) 的多项式例子有 $x,x^2-t,x^3-3tx,x^4-6x^2t+3t^2,\ldots$ 结果也可推广到 $\exp(\theta x-\theta^2 t/2)$. 使用 $u(t,x)$ 为多项式这一事实的唯一目的是验证积分收敛，以保证分部积分的合理性.

定理 7.5.9. 若 $T=\inf\{t:B_t\notin(-a,a)\}$，则 $ET^2=5a^4/3$.

证明. 定理 7.5.1 蕴含

$$E(B(T\wedge t)^4-6(T\wedge t)B(T\wedge t)^2)=-3E(T\wedge t)^2.$$

由定理 7.5.5，知 $ET = a^2 < \infty$. 令 $t \to \infty$，左边用控制收敛定理，右边用单调收敛定理，得

$$a^4 - 6a^2 ET = -3E(T^2)$$

代入 $ET = a^2$ 即得所求结果. □

习题

习题 7.5.1. 设 $T = \inf\{B_t \notin (-a, a)\}$. 证明

$$E_0 \exp(-\lambda T) = 1/\cosh(a\sqrt{2\lambda}).$$

习题 7.5.2. 本习题的目的是获取带漂移 $-b$ 的布朗运动 $X_t \equiv B_t - bt$ 击中水平 a 所需时间的信息. 令 $\tau = \inf\{t : B_t = a + bt\}$，其中 $a > 0$.

(i) 用鞅 $\exp(\theta B_t - \theta^2 t/2)$，取 $\theta = b + (b^2 + 2\lambda)^{1/2}$，证明

$$E_0 \exp(-\lambda\tau) = \exp(-a\{b + (b^2 + 2\lambda)^{1/2}\})$$

令 $\lambda \to 0$ 得 $P_0(\tau < \infty) = \exp(-2ab)$.

习题 7.5.3. 设 $\sigma = \inf\{t : B_t \notin (a, b)\}$ 且 $\lambda > 0$.

(i) 利用强 Markov 性质证明

$$E_x \exp(-\lambda T_a) = E_x(e^{-\lambda\sigma}; T_a < T_b) + E_x(e^{-\lambda\sigma}; T_b < T_a) E_b \exp(-\lambda T_a)$$

(ii) 交换 a 和 b 的角色得到第二个方程，利用定理 7.5.7 并解之得

$$E_x(e^{-\lambda\sigma}; T_a < T_b) = \sinh(\sqrt{2\lambda}(b - x))/\sinh(\sqrt{2\lambda}(b - a))$$
$$E_x(e^{-\lambda\sigma}; T_b < T_a) = \sinh(\sqrt{2\lambda}(x - a))/\sinh(\sqrt{2\lambda}(b - a))$$

习题 7.5.4. 若 $T = \inf\{t : B_t \notin (a, b)\}$，其中 $a < 0 < b$ 且 $a \neq -b$，则 T 与 B_T^2 不独立，故不能像定理 7.5.9 的证明中那样计算 ET^2. 用 Cauchy-Schwarz 不等式估计 $E(TB_T^2)$ 并推出

(i) $ET^2 \leqslant 4E(B_T^4)$

(ii) $EB_T^4 \leqslant 36ET^2$

习题 7.5.5. 找一个形如 $B_t^6 - c_1 t B_t^4 + c_2 t^2 B_t^2 - c_3 t^3$ 的鞅，并用它计算 $T = \inf\{t : B_t \notin (-a, a)\}$ 的三阶矩.

习题 7.5.6. 证明 $(1+t)^{-1/2}\exp(B_t^2/2(1+t))$ 是鞅，并用此结论证明

$$\limsup_{t\to\infty} B_t/((1+t)\log(1+t))^{1/2} \leqslant 1 \quad \text{几乎必然.}$$

第 6 节 Itô 公式 *

本节目标是证明三个版本的 Itô 公式. 第一个是关于一维布朗运动的:

定理 7.6.1. 设 $f \in C^2$，即 f 具有二阶连续导数. 则以概率 1，对所有 $t \geqslant 0$ 有

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s \tag{7.6.1}$$

当然，作为推导的一部分，必须解释右边第一项的含义. 作为对比，若 A_s 连续且在每一有限区间上有界变差，且 $f \in C^1$:

$$f(A_t) - f(A_0) = \int_0^t f'(A_s)\,\mathrm{d}A_s$$

其中右边的 Riemann-Stieltjes 积分为

$$\lim_{n\to\infty} \sum_{i=1}^{k(n)} f'(A_{t_{i-1}^n})[A(t_i^n) - A(t_{i-1}^n)]$$

这里 $0 = t_0^n < t_1^n < \cdots < t_{k(n)}^n = t$ 为一列分割，其网格 $\max_i |t_i^n - t_{i-1}^n| \to 0$.

证明. 下证 (7.6.1). 令 $t_i^n = ti/2^n$ （$0 \leqslant i \leqslant k(n) = 2^n$）. 由微积分知，对任意 a 和 b，存在介于 a 和 b 之间的 $c(a,b)$ 使得

$$f(b) - f(a) = (b-a)f'(a) + \frac{1}{2}(b-a)^2 f''(c(a,b)) \tag{7.6.2}$$

由此

$$\begin{aligned} f(B_t) - f(B_0) &= \sum_i [f(B_{t_{i+1}^n}) - f(B_{t_i^n})] \\ &= \sum_i f'(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n}) + \frac{1}{2}\sum_i g_i^n(\omega)(B_{t_{i+1}^n} - B_{t_i^n})^2 \end{aligned} \tag{7.6.3}$$

其中 $g_i^n(\omega) = f''(c(B_{t_i^n}, B_{t_{i+1}^n}))$. 比较 (7.6.3) 与 (7.6.1)，只需证

$$\sum_i f'(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n}) \to \int_0^t f'(B_s)\,\mathrm{d}B_s \tag{7.6.4}$$

$$\frac{1}{2}\sum_i g_i(\omega)(B_{t_{i+1}^n} - B_{t_i^n})^2 \to \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s \tag{7.6.5}$$

依概率收敛（当 $n \to \infty$）.

先在假设 $|f'(x)|, |f''(x)| \leqslant K$ 下证之. 令

$$I_m^1 \equiv \sum_i f'(B_{t_i^m})(B_{t_{i+1}^m} - B_{t_i^m})$$

为比较 I_m^1 与 I_n^1 $(n \geqslant m)$，将长为 $t2^{-m}$ 的区间进一步细分为长 $t2^{-n}$ 的区间. 为保持和不变，需重新定义. 令 $i(n,m,j) = [2^m j/2^n]/2^m$ 且 $H_j^m = f'(B_{i(n,m,j)})$. 若令 $K_j^n = f'(B_{t_j^n})$，则

$$I_m^1 - I_n^1 = \sum_j (H_j^m - K_j^n)(B_{t_{j+1}^n} - B_{t_j^n})$$

平方得

$$\begin{aligned}(I_m^1 - I_n^1)^2 = &\sum_j (H_j^m - K_j^n)^2 (B_{t_{j+1}^n} - B_{t_j^n})^2 \\ &+ 2\sum_{j<k} (H_j^m - K_j^n)(H_k^m - K_k^n)(B_{t_{j+1}^n} - B_{t_j^n})(B_{t_{k+1}^n} - B_{t_k^n})\end{aligned}$$

下一步取期望. 第二项和的期望为零，因对 j,k 项关于 $\mathcal{F}_{t_k^n}$ 取期望时，前三项可提出期望外

$$(H_j^m - K_j^n)(H_k^m - K_k^n)(B_{t_{j+1}^n} - B_{t_j^n}) E(B_{t_{k+1}^n} - B_{t_k^n} | \mathcal{F}_{t_k^n}) = 0$$

对第一项和关于 $\mathcal{F}_{t_j^n}$ 取条件期望，得

$$E(I_m^1 - I_n^1)^2 = \sum_j E(H_j^m - K_j^n)^2 \cdot 2^{-n} t \tag{7.6.6}$$

由于 $|i(n,m,j)/2^m - j/2^n| \leqslant 1/2^m$ 且 $|f''(x)| \leqslant K$，有

$$\sup_{j \leqslant 2^n} |H_j^m - K_j^n| \leqslant K \sup\{|B_s - B_r| : 0 \leqslant r \leqslant s \leqslant t, |s-r| \leqslant 2^{-m}\}$$

由于布朗运动在 $[0,t]$ 上一致连续且 $E \sup_{s \leqslant t} |B_s|^2 < \infty$，由控制收敛定理

$$\lim_{m,n \to \infty} E(I_m^1 - I_n^1)^2 = 0$$

这表明 I_m^1 是 L^2 中的 Cauchy 列，故存在 I_∞^1 使得 $I_m^1 \to I_\infty^1$ 于 L^2. 该极限定义 $\int_0^t f'(B_s)\,\mathrm{d}B_s$ 为逼近 Riemann 和的极限.

下证 (7.6.5). 令 $G_s^n = g_i^n(\omega) = f''(c(B_{t_i^n}, B_{t_{i+1}^n}))$（当 $s \in (t_i^n, t_{i+1}^n]$），并令

$$A_s^n = \sum_{t_{i+1}^n \leqslant s} (B_{t_{i+1}^n} - B_{t_i^n})^2$$

则

$$\sum_i g_i^n(\omega)(B_{t_{i+1}^n} - B_{t_i^n})^2 = \int_0^t G_s^n \,\mathrm{d}A_s^n$$

只需证

$$\int_0^t G_s^n \,\mathrm{d}A_s^n \to \int_0^t f''(B_s)\,\mathrm{d}s \tag{7.6.7}$$

为此，首先注意 f'' 的连续性蕴含：若 $s_n \to s$（当 $n \to \infty$），则 $G_{s_n}^n \to f''(B_s)$，而习题 7.1.6 蕴含 A_s^n 几乎必然收敛于 s. 此时，可固定 ω 并由以下简单结果推出 (7.6.7)：

引理 7.6.2. 若

(i) $[0,t]$ 上的测度 μ_n 弱收敛于有限测度 μ_∞，

(ii) g_n 为一列函数满足 $|g_n| \leqslant K$，且具有性质：只要 $s_n \in [0,t] \to s$ 就有 $g_n(s_n) \to g(s)$，

则当 $n \to \infty$

$$\int g_n \,\mathrm{d}\mu_n \to \int g \,\mathrm{d}\mu_\infty$$

证明. 令 $\mu_n'(A) = \mu_n(A)/\mu_n([0,t])$，可设所有 μ_n 为概率测度. 标准构造（见定理 3.2.8）表明存在随机变量 X_n 服从分布 μ_n 使得 $X_n \to X_\infty$ 几乎必然（当 $n \to \infty$）. g_n 到 g 的收敛蕴含 $g_n(X_n) \to g(X_\infty)$，故结果由有界收敛定理得证 □

引理 7.6.2 是在附加假设 $|f'(x)|, |f''(x)| \leqslant M$ 下完成证明的最后一步. 回溯证明过程，引理 7.6.2 蕴含 (7.6.7)，而 (7.6.7) 又完成 (7.6.5) 的证明. 于是结合 (7.6.4) 并利用 (7.6.3) 得

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s \quad \text{几乎必然.}$$

为去掉对导数的有界性假设，取 M 充分大并定义 f_M 使得 $f_M = f$ 于 $[-M, M]$ 且 $|f_M'|, |f_M''| \leqslant M$. 公式对 f_M 成立，但在 $\{\sup_{s\leqslant t} |B_s| \leqslant M\}$ 上 f_M 与 f 无异. □

例题 7.6.3. 在 (7.6.1) 中取 $f = x^2$，有

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

故若 $B_0 = 0$，则

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

这与微积分公式 $\int_0^a 2x\,\mathrm{d}x = a^2$ 不同.

直接证明. 利用前面引入的分割，可写出

$$B_t^2 = \left(\sum_{i=0}^{2^n-1}(B(t_{i+1}^n) - B(t_i^n))\right)^2 + \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^n-1}(B(t_{i+1}^n) - B(t_i^n))(B(t_{j+1}^n) - B(t_j^n))$$

若对 $j < i$ 求和，则有

$$\sum_{i=0}^{2^n-1} B(t_i^n)(B(t_{i+1}^n) - B(t_i^n)) \to \int_0^t B_s\,\mathrm{d}B_s$$

对 $i < j$ 求和得到同样的极限当对 $i = j$ 求和时，有

$$\sum_{i=0}^{2^n-1}(B(t_{i+1}^n) - B(t_i^n))^2 \to t$$

为说明答案不能是 $B_t^2/2$，注意

定理 7.6.4. 若 g 连续且 $E\int_0^t |g(B_s)|^2\,\mathrm{d}s < \infty$，则 $\int_0^t g(B_s)\,\mathrm{d}B_s$ 是连续鞅.

若 g 在 ∞ 处增长过快，可积性条件可能不成立. 为处理这一问题，引入一个较弱的概念：M_t 是局部鞅，若存在递增停时序列 $T_n \uparrow \infty$ 使得 $M(t \wedge T_n)$ 为鞅. 若 g 连续，则随机积分总是局部鞅.

证明. 先在假设 $|g| \leqslant K$ 下证之. 令 $t_i^n = ti/2^n$ $(0 \leqslant i \leqslant 2^n)$，$J = \max\{i : t_{i+1}^n \leqslant s\}$，并令

$$I_n^1(s) = \sum_{i=0}^{J} g(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n}) + g(B_{t_{J+1}^n})(B_s - B_{t_{J+1}^n})$$

易验证 $I_n^1(s)$ $(s \leqslant t)$ 为连续鞅，且

$$EI_n^1(t)^2 = \sum_{0\leqslant i<2^n} E[g(B_{t_i^n})^2]\cdot t2^{-n}.$$

L^2 极大不等式蕴含

$$E\left(\sup_{s\leqslant t}|I_m^1(s) - I_n^1(s)|^2\right) \leqslant 4E(I_m^1 - I_n^1)^2$$

右边 $\to 0$（由前面的证明），故 $I_m^1(s) \to I_\infty^1(s)$ 在 $[0,t]$ 上一致收敛. 令 $Y \in \mathcal{F}_r$ 满足 $EY^2 < \infty$. 由于 $I_n^1(s)$ 为鞅，若 $r < s < t$ 有

$$E[Y\{I_n^1(s) - I_n^1(r)\}] = 0$$

由此及 Cauchy-Schwarz 不等式，得

$$E[Y\{I_\infty^1(s)-I_\infty^1(r)\}]=0$$

由于这对 $Y=1_A$（$A\in\mathcal{F}_r$）成立，在假设 $|g|\leqslant K$ 下已证得所需结果。

用 $g_M=g$ 于 $[-M,M]$ 且 g_M 有界来逼近 g，并停在时刻 $T_M=\inf\{t:|B_t|\geqslant M\}$，则当 g 连续时

$$\int_0^{s\wedge T_M} g(B_r)\,\mathrm{d}B_r$$

为连续鞅，且

$$E\left(\int_0^{s\wedge T_M} f'(B_r)\,\mathrm{d}B_r\right)^2=E\int_0^{s\wedge T_M} f'(B_r)^2\,\mathrm{d}r$$

令 $M\to\infty$，若 $E\int_0^t|f'(B_s)|^2\,\mathrm{d}s<\infty$，则 $\int_0^t f'(B_s)\,\mathrm{d}B_s$ 为连续鞅. □

Itô 公式的第二个版本允许 f 同时依赖于 t

定理 7.6.5. 设 $f\in C^2$，即 f 具有到二阶的连续偏导数. 则以概率 1，对所有 $t\geqslant 0$

$$\begin{aligned}f(t,B_t)-f(0,B_0)=&\int_0^t\frac{\partial f}{\partial t}(s,B_s)\,\mathrm{d}s+\int_0^t\frac{\partial f}{\partial x}(s,B_s)\,\mathrm{d}B_s\\&+\frac{1}{2}\int_0^t\frac{\partial^2 f}{\partial x^2}(s,B_s)\,\mathrm{d}s\end{aligned}\tag{7.6.8}$$

证明. 给出此结果（及下一个结果）完全严格证明的最简方法或许是：先对 t 和 x 的多项式 f 证之，再注意对任意 $f\in C^2$ 和 $M<\infty$，可找到多项式 φ_n 使得 φ_n 及其直到二阶的所有偏导数在 $[0,t]\times[-M,M]$ 上一致收敛于 f 的相应偏导数. 这里仅解释此结果为何成立.

展开到二阶，且对误差项不作精细估计

$$\begin{aligned}f(t,b)-f(s,a)\approx&\frac{\partial f}{\partial t}(s,a)(t-s)+\frac{\partial f}{\partial x}(s,a)(b-a)\\&+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,a)(b-a)^2+\frac{\partial^2 f}{\partial t\partial x}(s,a)(t-s)(b-a)+\frac{1}{2}\frac{\partial^2 f}{\partial t^2}(s,a)(t-s)^2\end{aligned}$$

令 t 为固定正数，$t_k^n=tk/2^n$. 如前，先在所有一阶和二阶偏导数有界（界为 K）的假设下证之. 前三项收敛到其极限的细节类似于前述证明. 为说明第五项在极限中消失，利用 $t_{i+1}^n-t_i^n=2^{-n}$ 得

$$\sum_i\left|\frac{\partial^2 f}{\partial x^2}(t_i^n,B(t_i^n))\right|(t_{i+1}^n-t_i^n)^2\leqslant Kt\cdot 2^{-n}\to 0$$

为处理第四项，首先考虑

$$V_n^{1,1}(t)=\sum_i (t_{i+1}^n-t_i^n)|B(t_{i+1}^n)-B(t_i^n)|$$

$EV_n^{1,1}(t)=\sum_i C_1(t_{i+1}^n-t_i^n)^{3/2}$，其中 $C_1=E\chi^{1/2}$，χ 为标准正态如前论证，

$$E\sum_i \left|\frac{\partial^2 f}{\partial x^2}(t_i^n,B(t_i^n))\right|(t_{i+1}^n-t_i^n)|B(t_{i+1}^n)-B(t_i^n)|\leqslant KC_1t\cdot 2^{-n/2}\to 0$$

现由 Markov 不等式，若 $\varepsilon>0$

$$P\left(\sum_i \left|\frac{\partial^2 f}{\partial x^2}(t_i^n,B(t_i^n))\right|(t_{i+1}^n-t_i^n)|B(t_{i+1}^n)-B(t_i^n)|>\varepsilon\right)\leqslant KC_1t2^{-n/2}/\varepsilon\to 0$$

故该和依概率收敛于 0。

为去掉偏导数有界的假设，令 $T_M=\min\{t:|B_t|>M\}$ 并利用前面的计算，即得公式在时刻 $T_M\wedge t$ 前成立. □

例题 7.6.6. 利用 (7.6.8) 可证明定理 7.5.8：若 $u(t,x)$ 为 t 和 x 的多项式且 $\partial u/\partial t+(1/2)\partial^2 u/\partial x^2=0$，则 Itô 公式蕴含

$$u(t,B_t)-u(0,B_0)=\int_0^t \frac{\partial u}{\partial x}(s,B_s)\,\mathrm{d}B_s$$

由于 $\partial u/\partial x$ 是多项式，它满足定理 7.6.4 中的可积性条件，故 $u(t,B_t)$ 为鞅. 为得到新结论，令 $u(t,x)=\exp(\mu t+\sigma x)$ 指数布朗运动 $X_t=u(t,B_t)$ 常被用作股票价格的模型. 由于

$$\frac{\partial u}{\partial t}=\mu u,\quad \frac{\partial u}{\partial x}=\sigma u,\quad \frac{\partial^2 u}{\partial x^2}=\sigma^2 u$$

Itô 公式给出

$$X_t-X_0=\int_0^t\left(\mu-\frac{\sigma^2}{2}\right)X_s\,\mathrm{d}s+\int_0^t \sigma X_s\,\mathrm{d}B_s$$

由此可见若 $\mu=\sigma^2/2$ 则 X_t 为鞅，但该公式提供了关于 X_t 行为的更多信息.

Itô 公式的第三个也是最后一个版本是针对多维布朗运动的. 为简化记号，记 $D_if=\partial f/\partial x_i$，$D_{ij}=\partial^2 f/\partial x_i\partial x_j$

定理 7.6.7. 设 $f \in C^2$. 则以概率 1，对所有 $t \geqslant 0$

$$f(t, B_t) - f(0, B_0) = \int_0^t \frac{\partial f}{\partial t}(s, B_s)\,\mathrm{d}s + \sum_{i=1}^{d} \int_0^t D_i f(B_s)\,\mathrm{d}B_s^i + \frac{1}{2}\sum_{i=1}^{d} \int_0^t D_{ii} f(B_s)\,\mathrm{d}s \tag{7.6.9}$$

证明. 令 $x_0 = t$，记 $D_0 f = \partial f / \partial x_0$ 展开到二阶，且对误差项不作精细估计：

$$f(y) - f(x) \approx \sum_i D_i f(x)(y_i - x_i) + \frac{1}{2}\sum_i D_{ii} f(x)(y_i - x_i)^2 + \sum_{i<j} D_{ij} f(x)(y_i - x_i)(y_j - x_j)$$

令 t 为固定正数，$t_i^n = ti/2^n$ 前两项收敛到其极限的细节类似于前两个证明为处理第三项，令

$$F_{i,j}(t_k^n) = D_{ij} f(B(t_k^n))$$

由坐标的独立性，有

$$\begin{aligned} &E(F_{i,j}(t_k^n)(B_i(t_{k+1}^n) - B_i(t_k^n)) \cdot (B_j(t_{k+1}^n) - B_j(t_k^n))|\mathcal{F}_{t_k^n}) \\ &\quad = F_{i,j}(t_k^n) E((B_i(t_{k+1}^n) - B_i(t_k^n)) \cdot (B_j(t_{k+1}^n) - B_j(t_k^n))|\mathcal{F}_{t_k^n}) = 0 \end{aligned}$$

固定 $i < j$ 并令 $G_k = F_{i,j}(t_k^n)(B_i(t_{k+1}^n) - B_i(t_k^n)) \cdot (B_j(t_{k+1}^n) - B_j(t_k^n))$. 上述计算表明

$$E\left(\sum_k G_k\right) = 0$$

为计算和的方差，注意

$$E\left(\sum_k G_k\right)^2 = E\sum_k G_k^2 + 2E\sum_{k<\ell} G_k G_\ell$$

首先在假设 $|\partial^2 f / \partial x_i \partial x_j| \leqslant C$ 下完成证明此时，对第一项关于 $\mathcal{F}_{t_k^n}$ 取条件以将 $F_{i,j}^2(t_k^n)$ 提出，

$$\begin{aligned} &E(F_{i,j}^2(t_k^n)(B_i(t_{k+1}^n) - B_i(t_k^n))^2 \cdot (B_j(t_{k+1}^n) - B_j(t_k^n))^2|\mathcal{F}_{t_k^n}) \\ &\quad = F_{i,j}^2(t_k^n) E((B_i(t_{k+1}^n) - B_i(t_k^n))^2 \cdot (B_j(t_{k+1}^n) - B_j(t_k^n))^2|\mathcal{F}_{t_k^n}) = F_{i,j}^2(t_k^n)(t_{k+1}^n - t_k^n)^2 \end{aligned}$$

故

$$E\sum_k G_k^2 \leqslant K^2\sum_k (t_{k+1}^n - t_k^n)^2 \leqslant K^2 t 2^{-n} \to 0$$

为处理交叉项，对 $\mathcal{F}_{t_\ell^n}$ 取条件，得

$$E(G_k G_\ell|\mathcal{F}_{t_\ell^n}) = G_k F_{i,j}(t_\ell^n) E((B_i(t_{\ell+1}^n) - B_i(t_\ell^n))(B_j(t_{\ell+1}^n) - B_j(t_\ell^n))|\mathcal{F}_{t_\ell^n}) = 0$$

至此已证 $E(\sum_k G_k)^2 \to 0$，所需结果由 Markov 不等式得证二阶导数有界的假设可如前去掉. □

习题

习题 7.6.1. 设 $f \in C^2$. 证明 $f(B_t)$ 是鞅当且仅当 $f(x) = a + bx$.

习题 7.6.2. 证明 $B_t^3 - 3tB_t$ 和 $B_t^3 - \int_0^t 3B_s \,\mathrm{d}s$ 是鞅.

习题 7.6.3. 令 $\beta_{2k}(t) = E_0 B_t^{2k}$. 用 Itô 公式建立 $\beta_{2k}(t)$ 与 $\beta_{2k-2}(t)$ 的关系，并用此关系导出 $\beta_{2k}(t)$ 的公式.

习题 7.6.4. 对 $|B_t|$ 应用 Itô 公式.

习题 7.6.5. 对 $|B_t|^2$ 应用 Itô 公式. 用此证明 $E_0|B_t|^2 = td$.

第八章
随机游走的应用

本节内容的核心在于，我们可以将均值为零的独立随机变量之和（见第 8.1 节）以及鞅（见第 8.2 节）嵌入到布朗运动中，并以此推导出中心极限定理. 基于这些研究结果，第 8.3 节讨论了平稳序列的中心极限定理，第 8.4 节则推导了经缩放的经验分布向布朗桥的收敛性. 最后，在第 8.5 节中，我们利用这种嵌入方法证明了重对数律.

第 1 节　Donsker 定理

本节证明 Donsker (1951) 的定理. 下述结果是证明的关键：

定理 8.1.1. Skorokhod 表示定理. 若 $EX = 0$ 且 $EX^2 < \infty$，则存在 Brownian 运动的停时 T 使得 $B_T \overset{d}{=} X$ 且 $ET = EX^2$.

注释. 定理陈述中的 Brownian 运动以及本节所有 Brownian 运动均满足 $B_0 = 0$.

证明. 首先设 X 的支撑集为 $\{a, b\}$，其中 $a < 0 < b$. 由于 $EX = 0$，必须有

$$P(X = a) = \frac{b}{b-a}, \quad P(X = b) = \frac{-a}{b-a}.$$

设 $T = T_{a,b} = \inf\{t : B_t \notin (a, b)\}$，则定理 7.5.3 蕴含 $B_T \overset{d}{=} X$，而定理 7.5.5 告诉我们

$$ET = -ab = EB_T^2.$$

为处理一般情形，将 $F(x)=P(X\leqslant x)$ 写成均值为 0 的两点分布之混合. 令

$$c=\int_{-\infty}^{0}(-u)\,\mathrm{d}F(u)=\int_{0}^{\infty}v\,\mathrm{d}F(v).$$

若 φ 有界且 $\varphi(0)=0$，则利用 c 的两个公式可得

$$\begin{aligned}c\int_{0}^{\infty}\varphi(x)\,\mathrm{d}F(x)&=\int_{0}^{\infty}\varphi(v)\,\mathrm{d}F(v)\int_{-\infty}^{0}(-u)\,\mathrm{d}F(u)\\&\quad+\int_{-\infty}^{0}\varphi(u)\,\mathrm{d}F(u)\int_{0}^{\infty}v\,\mathrm{d}F(v)\\&=\int_{0}^{\infty}\mathrm{d}F(v)\int_{-\infty}^{0}\mathrm{d}F(u)(v\varphi(u)-u\varphi(v)).\end{aligned}$$

于是我们有

$$\int\varphi(x)\,\mathrm{d}F(x)=c^{-1}\int_{0}^{\infty}\mathrm{d}F(v)\int_{-\infty}^{0}\mathrm{d}F(u)(v-u)\left\{\frac{-u}{v-u}\varphi(u)+\frac{v}{v-u}\varphi(v)\right\}.$$

最后一个等式给出了所求的混合. 设 $(U,V)\in\mathbb{R}^2$ 满足

$$\begin{aligned}P\{(U,V)=(0,0)\}&=F(\{0\})\\P((U,V)\in A)&=c^{-1}\iint_{(u,v)\in A}\mathrm{d}F(u)\,\mathrm{d}F(v)(v-u)\end{aligned}$$

对 $A\subset(-\infty,0)\times(0,\infty)$，并定义概率测度 $\mu_{0,0}(\{0\})=1$ 且

$$\mu_{u,v}(\{u\})=\frac{v}{v-u},\quad \mu_{u,v}(\{v\})=\frac{-u}{v-u},\quad \text{对}u<0<v,$$

则

$$\int\varphi(x)\,\mathrm{d}F(x)=E\int\varphi(x)\mu_{U,V}(\mathrm{d}x).$$

上面证明了当 $\varphi(0)=0$ 时最后一个公式成立，但易见它对一般情形也成立. 将 $\varphi\equiv1$ 代入最后一个等式可见 (8.1.1) 定义的测度总质量为 1.

由上述计算可见，若在同一空间上有服从 (8.1.1) 分布的 (U,V) 以及独立的 Brownian 运动，则 $B(T_{U,V})\overset{d}{=}X$. 细心的读者会注意到 $T_{U,V}$ 不是 B_t 的停时，因为 (U,V) 独立于 Brownian 运动. 但这并不构成严重问题，因为若关于 $U=u$ 和 $V=v$ 取条件，则 $T_{u,v}$ 是停时，且这对下面的所有计算已足够. 例如，为计算 $E(T_{U,V})$，注意到

$$E(T_{U,V})=E\{E(T_{U,V}|(U,V))\}=E(-UV)$$

由定理 7.5.5. (8.1.1) 蕴含

$$\begin{aligned} E(-UV) &= \int_{-\infty}^{0} \mathrm{d}F(u)(-u) \int_{0}^{\infty} \mathrm{d}F(v)v(v-u)c^{-1} \\ &= \int_{-\infty}^{0} \mathrm{d}F(u)(-u) \left\{ -u + \int_{0}^{\infty} \mathrm{d}F(v)c^{-1}v^2 \right\} \end{aligned}$$

因为

$$c = \int_{0}^{\infty} v \,\mathrm{d}F(v) = \int_{-\infty}^{0} (-u) \,\mathrm{d}F(u).$$

由 c 的第二个表达式得

$$E(T_{U,V}) = E(-UV) = \int_{-\infty}^{0} u^2 \,\mathrm{d}F(u) + \int_{0}^{\infty} v^2 \,\mathrm{d}F(v) = EX^2.$$

□

由定理 8.1.1，只需一小步即可得到：

定理 8.1.2. 设 $X_1, X_2, \ldots$ 为独立同分布，具有分布 F，均值为 0 且方差为 1，并令 $S_n = X_1 + \cdots + X_n$. 存在停时序列 $T_0 = 0, T_1, T_2, \ldots$ 使得 $S_n \overset{d}{=} B(T_n)$ 且 $T_n - T_{n-1}$ 独立同分布.

证明. 设 $(U_1, V_1), (U_2, V_2), \ldots$ 为独立同分布且具有 (8.1.1) 给出的分布，并设 B_t 为独立的 Brownian 运动. 令 $T_0 = 0$，且对 $n \geqslant 1$，令

$$T_n = \inf\{t \geqslant T_{n-1} : B_t - B(T_{n-1}) \notin (U_n, V_n)\}.$$

□

作为定理 8.1.2 的推论，我们得到：

定理 8.1.3. 中心极限定理. 在定理 8.1.2 的假设下，$S_n/\sqrt{n} \Rightarrow \chi$，其中 χ 具有标准正态分布.

证明. 设 $W_n(t) = B(nt)/\sqrt{n} \overset{d}{=} B_t$ 由 Brownian 尺度变换，则

$$S_n/\sqrt{n} \overset{d}{=} B(T_n)/\sqrt{n} = W_n(T_n/n).$$

弱大数定律蕴含 $T_n/n \to 1$ 依概率. 由此应可清楚地看出 $S_n/\sqrt{n} \Rightarrow B_1$. 为补充细节，令 $\varepsilon > 0$，选取 δ 使得

$$P(|B_t - B_1| > \varepsilon \text{ 对某个} t \in (1-\delta, 1+\delta)) < \varepsilon/2$$

然后选取 N 足够大使得对 $n \geqslant N$，$P(|T_n/n - 1| > \delta) < \varepsilon/2$. 最后两个估计蕴含对 $n \geqslant N$

$$P(|W_n(T_n/n) - W_n(1)| > \varepsilon) < \varepsilon.$$

由于 ε 任意，$W_n(T_n/n) - W_n(1) \to 0$ 依概率. 应用收敛合并引理（引理 3.2.13），取 $X_n = W_n(1)$ 和 $Z_n = W_n(T_n/n)$，即得所需结果. □

令 $\mathbb{N}$ 为非负整数集并令

$$S(u) = \begin{cases} S_k & \text{若} u = k \in \mathbb{N}, \\ \text{在}[k, k+1] \text{ 上线性} & \text{对} k \in \mathbb{N}. \end{cases}$$

本节稍后证明：

定理 8.1.4. Donsker 定理. 在定理 8.1.2 的假设下，

$$S(n\cdot)/\sqrt{n} \Rightarrow B(\cdot),$$

即 $C[0,1]$ 上相应的测度弱收敛.

为引出证明的准备工作，先推导几个推论. 每个推论的关键都源于定理 3.10.1 的如下结果.

定理 8.1.5. 若 $\psi : C[0,1] \to \mathbb{R}$ 具有性质：它 P_0-几乎必然连续，则

$$\psi(S(n\cdot)/\sqrt{n}) \Rightarrow \psi(B(\cdot)).$$

例题 8.1.6. 极大值. 令 $\psi(\omega) = \max\{\omega(t) : 0 \leqslant t \leqslant 1\}$. 此时 $\psi : C[0,1] \to \mathbb{R}$ 连续. 定理 8.1.5 这次蕴含

$$\max_{0 \leqslant m \leqslant n} S_m/\sqrt{n} \Rightarrow M_1 \equiv \max_{0 \leqslant t \leqslant 1} B_t.$$

为完善此结果，由 (7.4.4)，右边的分布为

$$P_0(M_1 \geqslant a) = P_0(T_a \leqslant 1) = 2P_0(B_1 \geqslant a).$$

例题 8.1.7. 时刻 n 前的最后一个零点. 令 $\psi(\omega) = \sup\{t \leqslant 1 : \omega(t) = 0\}$. 此时 ψ 不连续，因若 ω_ε 满足 $\omega_\varepsilon(0) = 0$ 且分段线性，斜率在 $[0, 1/3 + \varepsilon]$ 上为 1，在 $[1/3 + \varepsilon, 2/3]$ 上为 -1，在 $[2/3, 1]$ 上为 1，则 $\psi(\omega_0) = 2/3$ 但 $\psi(\omega_\varepsilon) = 0$ 对 $\varepsilon > 0$.

易见若 $\psi(\omega) < 1$ 且 $\omega(t)$ 在每个区间 $(\psi(\omega) - \delta, \psi(\omega) + \delta)$ 中取正值和负值，则 ψ 在 ω 处连续. 由 8.4.1 小节的论证，最后一个集合具有 P_0 测度 1.（若零点在左边孤立，则它在右边不孤立.）由

此可得

$$\sup\{m \leqslant n : S_{m-1} \cdot S_m \leqslant 0\}/n \Rightarrow L = \sup\{t \leqslant 1 : B_t = 0\}.$$

L 的分布由 (7.4.7) 给出. 最后一个结果说明反正弦律（定理 4.9.5），对简单随机游走已证明，当均值为 0 且方差有限时也成立.

例题 8.1.8. 半直线的占位时间. 令

$$\psi(\omega) = |\{t \in [0,1] : \omega(t) > a\}|.$$

点 $\omega \equiv a$ 说明 ψ 不连续，但易见 ψ 在满足 $|\{t \in [0,1] : \omega(t) = a\}| = 0$ 的路径 ω 处连续. Fubini 定理蕴含

$$E_0|\{t \in [0,1] : B_t = a\}| = \int_0^1 P_0(B_t = a)\,\mathrm{d}t = 0$$

所以 ψ 是 P_0-几乎必然连续的. 由定理 8.1.5，可得

$$|\{u \leqslant n : S(u) > a\sqrt{n}\}|/n \Rightarrow |\{t \in [0,1] : B_t > a\}|.$$

如下所示，稍加处理即可将其转化为更自然的结果

$$|\{m \leqslant n : S_m > a\sqrt{n}\}|/n \Rightarrow |\{t \in [0,1] : B_t > a\}|.$$

证明. 定理 8.1.5 的应用给出对任意 a，

$$|\{t \in [0,1] : S(nt) > a\sqrt{n}\}| \Rightarrow |\{t \in [0,1] : B_t > a\}|.$$

为将此转化为关于 $|\{m \leqslant n : S_m > a\sqrt{n}\}|$ 的结果，注意在 $\{\max_{m\leqslant n} |X_m| \leqslant \varepsilon\sqrt{n}\}$ 上，由 Chebyshev 不等式其概率 $\to 1$，有

$$\begin{aligned}|\{t \in [0,1] : S(nt) > (a+\varepsilon)\sqrt{n}\}| &\leqslant \frac{1}{n}|\{m \leqslant n : S_m > a\sqrt{n}\}| \\ &\leqslant |\{t \in [0,1] : S(nt) > (a-\varepsilon)\sqrt{n}\}|.\end{aligned}$$

将此与证明的第一个结论以及 $b \mapsto |\{t \in [0,1] : B_t > b\}|$ 在 $b = a$ 处几乎必然连续的事实结合，易得所需结论. □

为计算 $|\{t \in [0,1] : B_t > 0\}|$ 的分布，注意在定理 4.9.6 中证明了对称简单随机游走时左边收敛到反正弦律，故右边具有该分布，且它是任何均值为 0 方差有限的随机游走的极限. 最后的论证使用了一个称为 “不变原理” 的思想，源于 Erdős 与 Kac (1946, 1947)：只要中心极限定理适用，S_n 的泛函的渐近行为就应相同.

定理 8.1.4 的证明. 为简化证明并给下一节的推广做准备，令 $X_{n,m}$，$1 \leqslant m \leqslant n$，为三角阵列的随机变量，$S_{n,m} = X_{n,1} + \cdots + X_{n,m}$ 并假设 $S_{n,m} = B(\tau_m^n)$. 令

$$S_{n,(u)} = \begin{cases} S_{n,m} & \text{若} u = m \in \{0, 1, \ldots, n\}, \\ \text{当} m \in \{1, \ldots, n\} \text{ 时对} u \in [m-1, m] \text{ 线性.} \end{cases}$$

引理 8.1.9. 若 $\tau_{[ns]}^n \to s$ 依概率对每个 $s \in [0,1]$ 成立，则

$$\|S_{n,(n\cdot)} - B(\cdot)\|_\infty \to 0 \text{ 依概率.}$$

为与原问题建立联系，令 $X_{n,m} = X_m/\sqrt{n}$ 并定义 $\tau_1^n, \ldots, \tau_n^n$ 使得

$$(S_{n,1}, \ldots, S_{n,n}) \stackrel{d}{=} (B(\tau_1^n), \ldots, B(\tau_n^n)).$$

由定理 8.1.3 证明中定义的 $T_1, T_2, \ldots$ 为停时，Brownian 尺度变换知 $\tau_m^n \stackrel{d}{=} T_m/n$，故引理 8.1.9 的假设满足.

证明. B 具有连续路径（因而在 $[0,1]$ 上一致连续）的事实蕴含若 $\varepsilon > 0$ 则存在 $\delta > 0$ 使得 $1/\delta$ 为整数且

$$\text{(a) } P(|B_t - B_s| < \varepsilon \text{ 对所有} 0 \leqslant s \leqslant 1, |t-s| < 2\delta) > 1 - \varepsilon.$$

引理 8.1.9 的假设蕴含若 $n \geqslant N_\delta$ 则

$$P(|\tau_{[nk\delta]}^n - k\delta| < \delta \text{ 对} k = 1, 2, \ldots, 1/\delta) \geqslant 1 - \varepsilon.$$

由于 $m \mapsto \tau_m^n$ 递增，若 $s \in ((k-1)\delta, k\delta)$，则

$$\begin{aligned} \tau_{[ns]}^n - s &\geqslant \tau_{[n(k-1)\delta]}^n - k\delta, \\ \tau_{[ns]}^n - s &\leqslant \tau_{[nk\delta]}^n - (k+1)\delta, \end{aligned}$$

故若 $n \geqslant N_\delta$，

$$\text{(b) } P\left(\sup_{0 \leqslant s \leqslant 1} |\tau_{[ns]}^n - s| < 2\delta\right) \geqslant 1 - \varepsilon.$$

当 (a) 和 (b) 中的事件发生时

$$\text{(c) } |S_{n,m} - B_{m/n}| < \varepsilon \text{ 对所有} m \leqslant n.$$

为处理 $t=(m+\theta)/n$ 且 $0<\theta<1$，注意到

$$|S_{n,(nt)}-B_t|\leqslant(1-\theta)|S_{n,m}-B_{m/n}|+\theta|S_{n,m+1}-B_{(m+1)/n}|$$
$$+(1-\theta)|B_{m/n}-B_t|+\theta|B_{(m+1)/n}-B_t|.$$

对前两项用 (c), 对后两项用 (a), 可见若 $n\geqslant N_\delta$ 且 $1/n<2\delta$, 则以概率 $\geqslant 1-2\varepsilon$ 有 $\|S_{n,(n\cdot)}-B(\cdot)\|_\infty<2\varepsilon$. 由于 ε 任意，引理 8.1.9 的证明完成. □

为得到定理 8.1.4，我们必须证明：

引理 8.1.10. 若 φ 有界且连续，则 $E\varphi(S_{n,(n\cdot)})\to E\varphi(B(\cdot))$.

证明. 固定 $\varepsilon>0$，令 $G_\delta=\{\omega:$ 若$\|\omega-\omega_0\|_\infty<\delta$ 则$|\varphi(\omega)-\varphi(\omega_0)|<\varepsilon\}$. 由于 φ 连续，$G_\delta\uparrow C[0,1]$ 当 $\delta\downarrow 0$. 令 $\Delta_n=\|S_{n,(n\cdot)}-B(\cdot)\|_\infty$. 由引理 8.1.9 和如下平凡不等式即得所需结果

$$|E\varphi(S_{n,(n\cdot)})-E\varphi(B(\cdot))|\leqslant\varepsilon+(2\sup_\omega|\varphi(\omega)|)\{P(G_\delta^c)+P(\Delta_n\geqslant\delta)\}.$$

□

为容纳最后一个例子，我们需要定理 8.1.4 的一个平凡推广. 令 $C[0,\infty)=\{$连续$\omega:[0,\infty)\to\mathbb{R}\}$ 并令 $C[0,\infty)$ 为由有限维集生成的 σ-域. 给定 $C[0,\infty)$ 上的概率测度 μ, 存在 $C[0,M]=\{$连续$\omega:[0,M]\to\mathbb{R}\}$（以有限维集生成的 $C[0,M]$ 为 σ-域）上相应的测度 $\pi_M\mu$，由 “在时间 M 截断路径” 得到. 令 $(\psi_M\omega)(t)=\omega(t)$ 对 $t\in[0,M]$ 并令 $\pi_M\mu=\mu\circ\psi_M^{-1}$. 称 $C[0,\infty)$ 上的概率测度序列 μ_n 弱收敛于 μ 若对所有 M，$\pi_M\mu_n$ 在 $C[0,M]$ 上弱收敛于 $\pi_M\mu$，最后一个概念由 $M=1$ 情形的定义的平凡扩张定义. 有了这些定义，易得上述推广成立，故定理 8.1.4 得证. □

定理 8.1.11. $S(n\cdot)/\sqrt{n}\Rightarrow B(\cdot)$，即 $C[0,\infty)$ 上相应的测度弱收敛.

证明. 由定义，只需证明对所有 $M<\infty$，弱收敛发生在 $C[0,M]$ 上. 定理 8.1.4 的证明在 1 换为 M 时同样有效. □

例题 8.1.12. 令 $N_n=\inf\{m:S_m\geqslant n\}$ 且 $T_1=\inf\{t:B_t\geqslant 1\}$. 由于 $\psi(\omega)=T_1(\omega)\wedge 1$ 在 $C[0,1]$ 上是 P_0-几乎必然连续的且 T_1 的分布连续，由定理 8.1.5 得对 $0<t<1$

$$P(N_n\leqslant nt)\to P(T_1\leqslant t).$$

用 M 代替 1 重复最后一个论证，并用定理 8.1.11 表明最后一个结论对所有 t 成立.

习题

习题 8.1.1. 利用习题 7.5.4 得出 $E(T_{U,V}^2) \leqslant 4EX^4$.

习题 8.1.2. 设 S_n 为一维简单随机游走并令

$$R_n = 1 + \max_{m \leqslant n} S_m - \min_{m \leqslant n} S_m$$

为时刻 n 前访问的点数. 证明 $R_n/\sqrt{n} \Rightarrow$ 一个极限.

第 2 节 鞅的中心极限定理

本节证明鞅的中心极限定理. 关键在于 Strassen (1967) 的一个嵌入定理. 见他的定理 4.3.

定理 8.2.1. 若 S_n 为平方可积鞅且 $S_0 = 0$，B_t 为 Brownian 运动，则存在 Brownian 运动的停时序列 $0 = T_0 \leqslant T_1 \leqslant T_2 \ldots$ 使得

$$(S_0, S_1, \ldots, S_k) \overset{d}{=} (B(T_0), B(T_1), \ldots, B(T_k)) \text{ 对所有} k \geqslant 0.$$

证明. 包含 $S_0 = 0 = B(T_0)$ 仅为启动归纳法. 设对某个 $k \geqslant 1$ 已有

$$(S_0, \ldots, S_{k-1}) \overset{d}{=} (B(T_0), \ldots, B(T_{k-1})).$$

强 Markov 性质蕴含 $\{B(T_{k-1} + t) - B(T_{k-1}) : t \geqslant 0\}$ 是独立于 $\mathcal{F}(T_{k-1})$ 的 Brownian 运动. 令 $\mu_k(s_0, \ldots, s_{k-1}; \cdot)$ 为 $S_k - S_{k-1}$ 给定 $S_j = s_j$，$0 \leqslant j \leqslant k-1$ 的正则条件分布，即对每个 Borel 集 A

$$P(S_k - S_{k-1} \in A | S_0, \ldots, S_{k-1}) = \mu_k(S_0, \ldots, S_{k-1}; A)$$

由习题 4.1.18 和 4.1.16，这存在，且我们有

$$0 = E(S_k - S_{k-1} | S_0, \ldots, S_{k-1}) = \int x \mu_k(S_0, \ldots, S_{k-1}; \mathrm{d}x)$$

故条件分布的均值为 0 几乎必然. 对几乎每一个 $\hat{S} \equiv (S_0, \ldots, S_{k-1})$ 利用定理 8.1.1，我们看到存在停时 $\tau_{\hat{S}}$（对 $\{B(T_{k-1} + t) - B(T_{k-1}) : t \geqslant 0\}$）使得

$$B(T_{k-1} + \tau_{\hat{S}}) - B(T_{k-1}) \overset{d}{=} \mu_k(S_0, \ldots, S_{k-1}; \cdot)$$

令 $T_k = T_{k-1} + \tau_{\hat{S}}$，则 $(S_0, \ldots, S_k) \overset{d}{=} (B(T_0), \ldots, B(T_k))$，由归纳法得证. □

注释. 在证明细节尚留在读者脑海之际，指出若 $E(S_k - S_{k-1})^2 < \infty$，则

$$E(\tau_{\hat{S}} | S_0, \ldots, S_{k-1}) = \int x^2 \mu_k(S_0, \ldots, S_{k-1}; \mathrm{d}x)$$

因为 $B_t^2 - t$ 是鞅且 $\tau_{\hat{S}}$ 是从随机选取的区间 $(S_{k-1} + U_k, S_{k-1} + V_k)$ 的退出时.

在进入中心极限定理的细节之前，下面说明嵌入的存在使我们能给出定理 4.5.2 的一个简单证明.

定理 8.2.2. 令 $\mathcal{F}_m = \sigma(S_0, S_1, \ldots S_m)$. 在 $\sum_{m=1}^{\infty} E((S_m - S_{m-1})^2 | \mathcal{F}_{m-1}) < \infty$ 上，$\lim_{n\to\infty} S_n$ 存在且有限.

证明. 记 B_t 生成的 σ-域流为 $\mathcal{B}_t$，并令 $t_m = T_m - T_{m-1}$. 由构造有

$$E((S_m - S_{m-1})^2|\mathcal{F}_{m-1}) = E(t_m|\mathcal{B}(T_{m-1})).$$

令 $M = \inf\{n : \sum_{m=1}^{n+1} E(t_m|\mathcal{B}(T_{m-1})) > A\}$. M 是停时，且由定义知

$$\sum_{m=1}^{M} E(t_m|\mathcal{B}(T_{m-1})) \leqslant A.$$

由 $\{M \geqslant m\} \in \mathcal{F}_{m-1}$ 知

$$\begin{aligned} E\sum_{m=1}^{M} t_m &= E\sum_{m=1}^{\infty}(t_m; M \geqslant m) \\ &= \sum_{m=1}^{\infty} E(E(t_m|\mathcal{B}(T_{m-1}); M \geqslant m) = E\sum_{m=1}^{M} E(t_m|\mathcal{B}(T_{m-1})) \leqslant A. \end{aligned}$$

故 $\sum_{m=1}^{M} t_m < \infty$. 当 $A \to \infty$ 时 $P(M < \infty) \downarrow 0$，故 $T_\infty = \sum_{m=1}^{\infty} t_m < \infty$. 由 $B(T_n) \to B(T_\infty)$（当 $n \to \infty$）即得所需结果. □

迈向所承诺的中心极限定理的第一步是证明 Freedman (1971a) 的一个结果. 称 $X_{n,m}$，$\mathcal{F}_{n,m}$，$1 \leqslant m \leqslant n$，为鞅差阵列，若 $X_{n,m} \in \mathcal{F}_{n,m}$ 且 $E(X_{n,m}|\mathcal{F}_{n,m-1}) = 0$ 对 $1 \leqslant m \leqslant n$ 成立，其中 $\mathcal{F}_{n,0} = \{\emptyset, \Omega\}$. 令 $S_{n,m} = X_{n,1} + \cdots + X_{n,m}$. 记 $\mathbb{N} = \{0, 1, 2, \ldots\}$. 本节中，$S_{n,(n\cdot)}$ 表示 $S_{n,m}$ 的线性插值，定义为

$$S_{n,(u)} = \begin{cases} S_k & \text{若} u = k \in \mathbb{N}, \\ \text{当} k \in \mathbb{N} \text{ 时对} u \in [k, k+1] \text{ 线性}, \end{cases}$$

且 $B(\cdot)$ 为满足 $B_0 = 0$ 的 Brownian 运动. 令

$$V_{n,k} = \sum_{1 \leqslant m \leqslant k} E(X_{n,m}^2|\mathcal{F}_{n,m-1}).$$

定理 8.2.3. 设 $\{X_{n,m}, \mathcal{F}_{n,m}\}$ 为鞅差阵列. 若

(i) 对每个 t，$V_{n,[nt]} \to t$ 依概率，

(ii) $|X_{n,m}| \leqslant \varepsilon_n$ 对所有 m 成立且 $\varepsilon_n \to 0$，

则 $S_{n,(n\cdot)} \Rightarrow B(\cdot)$.

证明. 由 (i) 知 $V_{n,n} \to 1$ 依概率. 通过在 $V_{n,k}$ 首次超过 2 时停止每个序列并将后面的 $X_{n,m}$ 设为 0，不失一般性可假设 $V_{n,n} \leqslant 2 + \varepsilon_n^2$ 对所有 n 成立. 由定理 8.2.1，可找到停时 $T_{n,1}, \ldots, T_{n,n}$ 使得 $(S_{n,1}, \ldots, S_{n,n}) \stackrel{d}{=} (B(T_{n,1}), \ldots, B(T_{n,n}))$. 由引理 8.1.9，只需证明 $T_{n,[nt]} \to t$ 依概率对每个 $t \in [0, 1]$.

为此令 $t_{n,m} = T_{n,m} - T_{n,m-1}$（其中 $T_{n,0} = 0$），由定理 8.2.1 证明后的注记知 $E(t_{n,m}|\mathcal{F}_{n,m-1}) = E(X_{n,m}^2|\mathcal{F}_{n,m-1})$. 最后一个观察和假设 (i) 蕴含

$$\sum_{m=1}^{[nt]} E(t_{n,m}|\mathcal{F}_{n,m-1}) \to t \text{ 依概率.}$$

为由此得到 $T_{n,[nt]} \to t$ 依概率，注意到

$$\begin{aligned} &E\left(\sum_{m=1}^{[nt]} t_{n,m} - E(t_{n,m}|\mathcal{F}_{n,m-1})\right)^2 \\ &\quad = E\sum_{m=1}^{[nt]} \{t_{n,m} - E(t_{n,m}|\mathcal{F}_{n,m-1})\}^2 \end{aligned}$$

由鞅增量的正交性，定理 4.4.7. 现在

$$\begin{aligned} &E(\{t_{n,m} - E(t_{n,m}|\mathcal{F}_{n,m-1})\}^2|\mathcal{F}_{n,m-1}) \\ &\quad \leqslant E(t_{n,m}^2|\mathcal{F}_{n,m-1}) \\ &\quad \leqslant CE(X_{n,m}^4|\mathcal{F}_{n,m-1}) \leqslant C\varepsilon_n^2 E(X_{n,m}^2|\mathcal{F}_{n,m-1}) \end{aligned}$$

由习题 7.5.4 和假设 (ii). 对 n 求和，取期望，并回忆已假设 $V_{n,n} \leqslant 2 + \varepsilon_n^2$，得

$$E\left(\sum_{m=1}^{[nt]} t_{n,m} - E(t_{n,m}|\mathcal{F}_{n,m-1})\right)^2 \to C\varepsilon_n^2 EV_{n,n} \to 0.$$

解开定义，已证 $E(T_{n,[nt]} - V_{n,[nt]})^2 \to 0$，故 Chebyshev 不等式蕴含 $P(|T_{n,[nt]} - V_{n,[nt]}| > \varepsilon) \to 0$，利用 (ii) 即完成证明. □

有了定理 8.2.3，一个截断论证给出我们：

定理 8.2.4. 鞅的 Lindeberg-Feller 定理. 设 $X_{n,m}$，$\mathcal{F}_{n,m}$，$1 \leqslant m \leqslant n$ 为鞅差阵列. 若

(i) $V_{n,[nt]} \to t$ 依概率对所有 $t \in [0,1]$，且

(ii) 对所有 $\varepsilon > 0$，$\sum_{m \leqslant n} E(X_{n,m}^2 1_{(|X_{n,m}|>\varepsilon)}|\mathcal{F}_{n,m-1}) \to 0$ 依概率，

则 $S_{n,(n\cdot)} \Rightarrow B(\cdot)$.

注释. 若每行 $X_{n,m}$ 独立，则这退化为鞅中心极限定理.

为证明此定理，第一步是截断以便我们能应用定理 8.2.3. 令

$$\hat{V}_n(\varepsilon_n) = \sum_{m=1}^{n} E(X_{n,m}^2 1_{(|X_{n,m}|>\varepsilon_n)}|\mathcal{F}_{n,m-1}).$$

引理 8.2.5. 若 $\varepsilon_n \to 0$ 足够慢，则 $\varepsilon_n^{-2}\hat{V}_n(\varepsilon_n) \to 0$ 依概率.

注释. 前面的 ε_n^{-2} 是为了使我们能得出

$$\sum_{m\leqslant n} P(|X_{n,m}| > \varepsilon_n|\mathcal{F}_{n,m-1}) \to 0 \text{ 依概率.}$$

证明. 选取 N_m 使得 $P(m^2\hat{V}_n(1/m) > 1/m) \leqslant 1/m$ 对 $n \geqslant N_m$. 令 $\varepsilon_n = 1/m$ 对 $n \in [N_m, N_{m+1})$ 且 $\varepsilon_n = 1$ 对 $n < N_1$. 若 $\delta > 0$ 且 $1/m < \delta$，则对 $n \in [N_m, N_{m+1})$

$$P(\varepsilon_n^{-2}\hat{V}_n(\varepsilon_n) > \delta) \leqslant P(m^2\hat{V}_n(1/m) > 1/m) \leqslant 1/m. \quad \square$$

记 $\hat{X}_{n,m} = X_{n,m}1_{(|X_{n,m}|>\varepsilon_n)}$，$\bar{X}_{n,m} = X_{n,m}1_{(|X_{n,m}|\leqslant\varepsilon_n)}$，$\tilde{X}_{n,m} = \bar{X}_{n,m} - E(\bar{X}_{n,m}|\mathcal{F}_{n,m-1})$.

引理 8.2.6. 若以明显的方式定义 $\tilde{S}_{n,(n\cdot)}$，则定理 8.2.3 蕴含 $\tilde{S}_{n,(n\cdot)} \Rightarrow B(\cdot)$.

证明. 由 $|\tilde{X}_{n,m}| \leqslant 2\varepsilon_n$，只需验证定理 8.2.3 中的 (ii). 为此，由条件方差公式（定理 4.4.8）知

$$E(\tilde{X}_{n,m}^2|\mathcal{F}_{n,m-1}) = E(\bar{X}_{n,m}^2|\mathcal{F}_{n,m-1}) - E(\bar{X}_{n,m}|\mathcal{F}_{n,m-1})^2.$$

对第一项，注意到

$$E(\bar{X}_{n,m}^2|\mathcal{F}_{n,m-1}) = E(X_{n,m}^2|\mathcal{F}_{n,m-1}) - E(\hat{X}_{n,m}^2|\mathcal{F}_{n,m-1}).$$

对第二项，由 $E(X_{n,m}|\mathcal{F}_{n,m-1}) = 0$ 知

$$E(\bar{X}_{n,m}|\mathcal{F}_{n,m-1})^2 = E(\hat{X}_{n,m}|\mathcal{F}_{n,m-1})^2 \leqslant E(\hat{X}_{n,m}^2|\mathcal{F}_{n,m-1})$$

由 Jensen 不等式，故由 (a) 和 (i) 得

$$\sum_{m=1}^{[nt]} E(\tilde{X}_{n,m}^2|\mathcal{F}_{n,m-1}) \to t \text{ 对所有} t \in [0,1]. \quad \square$$

证明了上述引理，还需估计 $\tilde{S}_{n,(n\cdot)}$ 和 $S_{n,(n\cdot)}$ 之间的差异. 在 $\{|X_{n,m}| \leqslant \varepsilon_n \text{ 对所有} 1 \leqslant m \leqslant n\}$

上，有

$$\|\tilde{S}_{n,(n\cdot)} - S_{n,(n\cdot)}\|_\infty \leqslant \sum_{m=1}^{n} |E(\bar{X}_{n,m}|\mathcal{F}_{n,m-1})|. \tag{8.2.1}$$

为处理右边，注意到

$$\begin{aligned}\sum_{m=1}^{n} |E(\bar{X}_{n,m}|\mathcal{F}_{n,m-1})| &= \sum_{m=1}^{n} |E(\hat{X}_{n,m}|\mathcal{F}_{n,m-1})| \\ &\leqslant \sum_{m=1}^{n} E(|\hat{X}_{n,m}||\mathcal{F}_{n,m-1}) \\ &\leqslant \varepsilon_n^{-1} \sum_{m=1}^{n} E(\hat{X}_{n,m}^2|\mathcal{F}_{n,m-1}) \to 0\end{aligned} \tag{8.2.2}$$

依概率由引理 8.2.5. 为完成证明，只需证明

$$P(|X_{n,m}| > \varepsilon_n \text{ 对某个} m \leqslant n) \to 0. \tag{8.2.3}$$

因为由 (8.2.2) 和 (8.2.1) 这蕴含 $\|\tilde{S}_{n,(n\cdot)} - S_{n,(n\cdot)}\|_\infty \to 0$ 依概率. 定理 8.2.3 的证明构造了 Brownian 运动满足 $\|\tilde{S}_{n,(n\cdot)} - B(\cdot)\|_\infty \to 0$，故所需结果由三角不等式和引理 8.1.10 得证.

为证明 (8.2.3)，我们将使用 Dvoretsky (1972) 的引理 3.5.

引理 8.2.7. 若 A_n 适应于 $\mathcal{G}_n$ 则对任何非负 $\delta \in \mathcal{G}_0$

$$P(\cup_{m=1}^{n} A_m|\mathcal{G}_0) \leqslant \delta + P\left(\sum_{m=1}^{n} P(A_m|\mathcal{G}_{m-1}) > \delta \Big| \mathcal{G}_0\right).$$

证明. 下用归纳法. 当 $n = 1$ 时，结论说

$$P(A_1|\mathcal{G}_0) \leqslant \delta + P(P(A_1|\mathcal{G}_0) > \delta|\mathcal{G}_0).$$

这在 $\Omega_- \equiv \{P(A_1|\mathcal{G}_0) \leqslant \delta\}$ 上显然成立，在 $\Omega_+ \equiv \{P(A_1|\mathcal{G}_0) > \delta\} \in \mathcal{G}_0$ 上也成立，因为在 Ω_+ 上

$$P(P(A_1|\mathcal{G}_0) > \delta|\mathcal{G}_0) = 1 \geqslant P(A_1|\mathcal{G}_0).$$

为证明 n 个集的结果，由上一论证知不等式在 Ω_+ 上平凡成立. 令 $B_m = A_m \cap \Omega_-$. 由于 $\Omega_- \in \mathcal{G}_0 \subset \mathcal{G}_{m-1}$，$P(B_m|\mathcal{G}_{m-1}) = P(A_m|\mathcal{G}_{m-1})$ 在 Ω_- 上.（见习题 4.1.2.）应用 $n-1$ 个集的结果，取

$\gamma = \delta - P(B_1|\mathcal{G}_0) \geqslant 0$,

$$P(\cup_{m=2}^{n} B_m|\mathcal{G}_1) \leqslant \gamma + P\left(\sum_{m=2}^{n} P(B_m|\mathcal{G}_{m-1}) > \gamma \Big| \mathcal{G}_1\right).$$

关于 $\mathcal{G}_0$ 取条件期望并由 $\gamma \in \mathcal{G}_0$ 知

$$P(\cup_{m=2}^{n} B_m|\mathcal{G}_0) \leqslant \gamma + P\left(\sum_{m=2}^{n} P(B_m|\mathcal{G}_{m-1}) > \delta \Big| \mathcal{G}_0\right).$$

$\cup_{2\leqslant m\leqslant n} B_m = (\cup_{2\leqslant m\leqslant n} A_m) \cap \Omega_-$ 且习题 4.1.2 的另一应用表明

$$\sum_{1\leqslant m\leqslant n} P(B_m|\mathcal{G}_{m-1}) = \sum_{1\leqslant m\leqslant n} P(A_m|\mathcal{G}_{m-1})$$

在 Ω_- 上. 故在 Ω_- 上,

$$P(\cup_{m=2}^{n} A_m|\mathcal{G}_0) \leqslant \delta - P(A_1|\mathcal{G}_0) + P\left(\sum_{m=1}^{n} P(A_m|\mathcal{G}_{m-1}) > \delta \Big| \mathcal{G}_0\right).$$

结果由

$$P(\cup_{m=1}^{n} A_m|\mathcal{G}_0) \leqslant P(A_1|\mathcal{G}_0) + P(\cup_{m=2}^{n} A_m|\mathcal{G}_0)$$

得证. 为见此，令 $C = \cup_{2\leqslant m\leqslant n} A_m$，观察到 $1_{A_1\cup C} \leqslant 1_{A_1} + 1_C$，并利用条件期望的单调性. □

(8.2.3) 的证明. 令 $A_m = \{|X_{n,m}| > \varepsilon_n\}$，$\mathcal{G}_m = \mathcal{F}_{n,m}$ 并令 δ 为正数. 引理 8.2.7 蕴含

$$P(|X_{n,m}| > \varepsilon_n \text{ 对某个} m \leqslant n) \leqslant \delta + P\left(\sum_{m=1}^{n} P(|X_{n,m}| > \varepsilon_n|\mathcal{F}_{n,m-1}) > \delta\right).$$

为估计右边，注意 “Chebyshev 不等式”（第 4 章习题 1.3）蕴含

$$\sum_{m=1}^{n} P(|X_{n,m}| > \varepsilon_n|\mathcal{F}_{n,m-1}) \leqslant \varepsilon_n^{-2} \sum_{m=1}^{n} E(\hat{X}_{n,m}^2|\mathcal{F}_{n,m-1}) \to 0$$

故 $\limsup_{n\to\infty} P(|X_{n,m}| > \varepsilon_n \text{ 对某个} m \leqslant n) \leqslant \delta$. 由于 δ 任意，(8.2.3) 的证明从而定理 8.2.4 的证明完成. □

对应用而言，有一个关于单个序列的结果是有用的.

定理 8.2.8. 鞅中心极限定理. 设 X_n，$\mathcal{F}_n$，$n \geqslant 1$，为鞅差序列并令 $V_k = \sum_{1\leqslant n\leqslant k} E(X_n^2|\mathcal{F}_{n-1})$. 若

(i) $V_k/k \to \sigma^2 > 0$ 依概率，且

(ii) $n^{-1}\sum_{m\leqslant n}E(X_m^2 1_{(|X_m|>\varepsilon\sqrt{n})})\to 0$

则 $S_{(n\cdot)}/\sqrt{n}\Rightarrow\sigma B(\cdot)$.

证明. 令 $X_{n,m}=X_m/\sigma\sqrt{n}$，$\mathcal{F}_{n,m}=\mathcal{F}_m$. 改变记号并令 $k=nt$，第一个假设变为定理 8.2.4 的 (i). 为验证 (ii)，注意到

$$\sum_{m=1}^{n}E(E(X_{n,m}^2 1_{(|X_{n,m}|>\varepsilon)}|\mathcal{F}_{n,m-1}))=\sigma^{-2}n^{-1}\sum_{m=1}^{n}E(X_m^2 1_{(|X_m|>\varepsilon\sigma\sqrt{n})})\to 0.$$

□

第 3 节 平稳序列的中心极限定理

先考虑鞅的情形.

定理 8.3.1. 设 X_n,$n \in \mathbb{Z}$ 为遍历平稳的平方可积鞅差序列, 即 $\sigma^2 = EX_0^2 < \infty$ 且 $E(X_n|\mathcal{F}_{n-1}) = 0$, 其中 $\mathcal{F}_n = \sigma(X_m, m \leqslant n)$. 令 $S_n = X_1 + \cdots + X_n$. 则

$$S_{(n\cdot)}/n^{1/2} \Rightarrow \sigma B(\cdot)$$

注释. 这一结果由 Billingsley (1961) 和 Ibragimov (1963) 独立发现.

证明. $u_n \equiv E(X_n^2|\mathcal{F}_{n-1})$ 可表示为 $\phi(X_{n-1}, X_{n-2}, \ldots)$ 的形式, 因此第 6 章的 (1.3) 蕴含 u_n 是平稳遍历的, 且遍历定理给出

$$n^{-1}\sum_{m=1}^{n} u_m \to Eu_0 = EX_0^2 \quad \text{a.s.}$$

上述结论表明定理 8.2.8 的 (i) 成立. 为验证 (ii), 注意到

$$n^{-1}\sum_{m=1}^{n} E(X_m^2 1_{(|X_m|>\varepsilon\sqrt{n})}) = E(X_0^2 1_{(|X_0|>\varepsilon\sqrt{n})}) \to 0$$

由控制收敛定理可得. □

通过约化到鞅的情形, 下证关于平稳序列的第一个中心极限定理. 下述证明基于 Scott (1973), 但关键在于 Gordin (1969) 的一个思想. 在应用这一结果的某些场合 (例如见习题 8.3.1), 使用比 $\mathcal{F}_m = \sigma(X_n, n \leqslant m)$ 更丰富的 σ-域是方便的. 故采用满足鞅论和遍历论需求的抽象条件:

(i) $n \mapsto \mathcal{F}_n$ 是递增的, 且 $X_n \in \mathcal{F}_n$;

(ii) $\theta^{-1}\mathcal{F}_m = \mathcal{F}_{m+1}$.

定理 8.3.2. 设 X_n,$n \in \mathbb{Z}$ 为满足 $EX_n = 0$ 的遍历平稳序列. 设 $\mathcal{F}_n$ 满足 (i),(ii) 以及

$$\sum_{n\geqslant 1} \|E(X_0|\mathcal{F}_{-n})\|_2 < \infty$$

其中 $\|Y\|_2 = (EY^2)^{1/2}$. 令 $S_n = X_1 + \cdots + X_n$. 则 $S_{(n\cdot)}/\sqrt{n} \Rightarrow \sigma B(\cdot)$, 其中

$$\sigma^2 = EX_0^2 + 2\sum_{n=1}^{\infty} EX_0X_n$$

且定义中的级数绝对收敛.

证明. 设 X_n,$n \in \mathbb{Z}$ 定义在序列空间 $(\mathbb{R}^{\mathbb{Z}}, \mathcal{R}^{\mathbb{Z}}, P)$ 上, 满足 $X_n(\omega) = \omega_n$ 且 $(\theta^n\omega)(m) = \omega(m+n)$. 令

$$H_n = \{Y \in \mathcal{F}_n \text{ 满足} EY^2 < \infty\},$$
$$K_n = \{Y \in H_n \text{ 满足} E(YZ) = 0 \text{ 对所有} Z \in H_{n-1}\}.$$

从几何上看,$H_0 \supset H_{-1} \supset H_{-2} \ldots$ 为 L^2 的子空间序列, 而 K_n 是 H_{n-1} 在 H_n 中的正交补. 若 Y 为随机变量, 令 $(\theta^n Y)(\omega) = Y(\theta^n\omega)$. 从例子 $Y = f(X_{-j}, \ldots, X_k)$ 推广, 它有 $\theta^n Y = f(X_{n-j}, \ldots, X_{n+k})$, 容易看出若 $Y \in H_k$, 则 $\theta^n Y \in H_{k+n}$, 因此若 $Y \in K_j$, 则 $\theta^n Y \in K_{n+j}$.

若 X_0 恰好在 K_0 中, 则情形较为简单, 因对所有 n 都有 $X_n = \theta^n X_0 \in K_n$, 且取 $Z = 1_A \in H_{n-1}$, 有 $E(X_n 1_A) = 0$ 对所有 $A \in \mathcal{F}_{n-1}$ 成立, 因此 $E(X_n|\mathcal{F}_{n-1}) = 0$. 次佳情形是 X_0 有如下分解:

$$X_0 = Y_0 + Z_0 - \theta Z_0 \tag{8.3.1}$$

其中 $Y_0 \in K_0$ 且 $Z_0 \in L^2$, 因为此时若令

$$S_n = \sum_{m=1}^{n} X_m = \sum_{m=1}^{n} \theta^m X_0 \quad \text{且} \quad T_n = \sum_{m=1}^{n} \theta^m Y_0$$

则 $S_n = T_n + \theta Z_0 - \theta^{n+1} Z_0$.$\theta^m Y_0$ 是平稳遍历鞅差序列 (遍历性由第 6 章的 (1.3) 得出), 因此定理 8.3.1 给出

$$T_{(n\cdot)}/\sqrt{n} \Rightarrow \sigma B(\cdot) \quad \text{其中} \quad \sigma^2 = EY_0^2.$$

为消去另一项, 注意 $\theta Z_0/\sqrt{n} \to 0$ a.s. 且

$$P\left(\sup_{1 \leqslant m \leqslant n} \theta^{m+1} Z_0 > \varepsilon\sqrt{n}\right) \leqslant nP(Z_0 > \varepsilon\sqrt{n}) \leqslant \varepsilon^{-2} E(Z_0^2; Z_0 > \varepsilon\sqrt{n}) \to 0$$

由控制收敛定理. 为形式地求解 (8.3.1), 令

$$Z_0 = \sum_{j=0}^{\infty} E(X_j|\mathcal{F}_{-1}),$$
$$\theta Z_0 = \sum_{j=0}^{\infty} E(X_{j+1}|\mathcal{F}_0),$$
$$Y_0 = \sum_{j=0}^{\infty} \{E(X_j|\mathcal{F}_0) - E(X_j|\mathcal{F}_{-1})\}$$

并验证

$$Y_0 + Z_0 - \theta Z_0 = E(X_0|\mathcal{F}_0) = X_0.$$

为证明最后一个计算, 需验证 Z_0 和 Y_0 定义中级数的收敛性. 由假设和平移不变性可知

$$\sum_{j=0}^{\infty} \|E(X_j|\mathcal{F}_{-1})\|_2 < \infty,$$

因此三角不等式蕴含 Z_0 的级数在 L^2 中收敛. 由于

$$\|E(X_j|\mathcal{F}_0) - E(X_j|\mathcal{F}_{-1})\|_2 \leqslant \|E(X_j|\mathcal{F}_0)\|_2$$

(左边是到 $K_0 \subset H_0$ 的投影),Y_0 的级数也在 L^2 中收敛. 将各部分综合起来, 已证得带有 $\sigma^2 = EY_0^2$ 的定理 8.3.2. 为得到所示的 σ^2 公式, 注意到由条件化并使用 Cauchy-Schwarz 不等式,

$$|EX_0X_n| = |E(X_0E(X_n|\mathcal{F}_0))| \leqslant \|X_0\|_2\|E(X_n|\mathcal{F}_0)\|_2.$$

平移不变性蕴含 $\|E(X_n|\mathcal{F}_0)\|_2 = \|E(X_0|\mathcal{F}_{-n})\|_2$, 因此级数绝对收敛.

$$ES_n^2 = \sum_{j=1}^{n}\sum_{k=1}^{n} EX_jX_k = nEX_0^2 + 2\sum_{m=1}^{n-1}(n-m)EX_0X_m.$$

由此易知

$$n^{-1}ES_n^2 \to EX_0^2 + 2\sum_{m=1}^{\infty} EX_0X_m.$$

为完成证明, 令 $T_n = \sum_{m=1}^{n} \theta^m Y_0$, 注意到 $\sigma^2 = EY_0^2$ 且

$$n^{-1}E(S_n - T_n)^2 = n^{-1}E(\theta Z_0 - \theta^{n+1}Z_0)^2 \leqslant 4EZ_0^2/n \to 0$$

因为 $(a-b)^2 \leqslant (2a)^2 + (2b)^2$. □

下面转向例子. 第一个例子验证定理 8.3.2 的假设是平凡的.

例题 8.3.3. M-相依序列. 设 X_n,$n \in \mathbb{Z}$ 为满足 $EX_n = 0$,$EX_n^2 < \infty$ 的平稳序列, 且设 $\{X_j, j \leqslant 0\}$ 与 $\{X_k, k > M\}$ 独立. 在此情形下, 对 $n > M$ 有 $E(X_0|\mathcal{F}_{-n}) = 0$, 因此定理 8.3.2 给出 $S_{n,(n\cdot)}/\sqrt{n} \Rightarrow \sigma B(\cdot)$, 其中

$$\sigma^2 = EX_0^2 + 2\sum_{m=1}^{M} EX_0X_m.$$

注释. 关于此问题的直接处理方法, 见 Chung (1974) 的定理 7.3.1.

习题 8.3.1. 考虑例 8.3.3 的特例, 其中 ξ_i,$i \in \mathbb{Z}$ 为独立同分布, 以相等概率取值 H 和 T. 令 $\eta_n = f(\xi_n, \xi_{n+1})$, 其中 $f(H,T) = 1$ 且 $f(i,j) = 0$ 对其他情形.$S_n = \eta_1 + \cdots + \eta_n$ 计数 $\xi_1, \ldots, \xi_{n+1}$ 中 head

runs 的数目. 应用定理 8.3.2 于 $\mathcal{F}_n=\sigma(X_m:m\leqslant n+1)$, 证明存在常数 μ 和 σ 使得 $(S_n-n\mu)/\sigma n^{1/2}\Rightarrow\chi$. 证明过程中构造的随机变量 Y_0 在此情形下是什么?

例题 8.3.4. Markov 链. 设 ζ_n,$n\in\mathbb{Z}$ 为可数状态空间 S 上的不可约 Markov 链, 每个 ζ_n 具有平稳分布 π. 令 $X_n=f(\zeta_n)$, 其中 $\sum f(x)\pi(x)=0$ 且 $\sum f(x)^2\pi(x)<\infty$. 第 6 章的结果蕴含 X_n 是遍历平稳序列. 若令 $\mathcal{F}_{-n}=\sigma(\zeta_m,m\leqslant -n)$, 则

$$E(X_0|\mathcal{F}_{-n})=\sum_y p_n(\zeta_{-n},y)f(y)$$

其中 $p_n(x,y)$ 是 n 步转移概率, 因此

$$\|E(X_0|\mathcal{F}_{-n})\|_2^2=\sum_x\pi(x)\left(\sum_y p_n(x,y)f(y)\right)^2.$$

当 f 有界时, 可用 $\sum f(x)\pi(x)=0$ 得到如下估计:

$$\|E(X_0|\mathcal{F}_{-n})\|_2^2\leqslant\|f\|_\infty^2\sum_x\pi(x)\|p_n(x,\cdot)-\pi(\cdot)\|^2$$

其中 $\|f\|_\infty=\sup|f(x)|$ 且 $\|\cdot\|$ 是全变差范数.

当 S 有限时, 所有 f 均有界. 若链非周期, 第 5 章的习题 5.6 蕴含

$$\sup_x\|p_n(x,\cdot)-\pi(\cdot)\|\leqslant Ce^{-\varepsilon n}$$

且定理 8.3.2 的假设得以满足. 为说明此情形下极限方差 σ^2 可能为 0, 考虑习题 7.1 的修正, 其中 $X_n=(\xi_n,\xi_{n+1})$,$f(H,T)=1$,$f(T,H)=-1$ 且 $f(H,H)=f(T,T)=0$. 在此情形下,$\sum_{m=1}^n f(X_m)\in\{-1,0,1\}$, 因此不存在中心极限定理.

最后一个例子用直接方法处理很简单, 但有助于评估定理 8.3.2 及后续定理中条件的强弱.

例题 8.3.5. 滑动平均过程. 设

$$X_m=\sum_{k\geqslant 0}c_k\xi_{m-k}\quad\text{其中}\quad\sum_{k\geqslant 0}c_k^2<\infty$$

且 ξ_i,$i\in\mathbb{Z}$ 为独立同分布, 满足 $E\xi_i=0$ 且 $E\xi_i^2=1$. 若 $\mathcal{F}_{-n}=\sigma(\xi_m;m\leqslant -n)$ 则

$$\|E(X_0|\mathcal{F}_{-n})\|_2=\left\|\sum_{k\geqslant n}c_k\xi_{-k}\right\|_2=\left(\sum_{k\geqslant n}c_k^2\right)^{1/2}.$$

若, 例如,$c_k = (1+k)^{-p}$, 则 $\|E(X_0|\mathcal{F}_{-n})\|_2 \sim n^{(1/2)-p}$, 且当 $p > 3/2$ 时定理 8.3.2 适用.

注释. Hall 和 Heyde (1980) 的定理 5.3 表明

$$\sum_{j\geqslant 0} \|E(X_0|\mathcal{F}_{-j}) - E(X_0|\mathcal{F}_{-(j-1)})\|_2 < \infty$$

对中心极限定理是充分的. 利用这一结果可知 $\sum_k |c_k| < \infty$ 对例 7.3 中的中心极限定理是充分的.

改进结果中的条件接近最优. 设 ξ_i 以相等概率取值 1 和-1 且令 $c_k = (1+k)^{-p}$, 其中 $1/2 < p < 1$. Lindeberg-Feller 定理可用于证明 $S_n/n^{3/2-p} \Rightarrow \sigma\chi$. 为检验标准化, 注意

$$\sum_{m=1}^{n} X_m = \sum_{j\leqslant n} a_{n,j}\xi_j$$

若 $j \geqslant 0$ 则 $a_{n,j} = \sum_{i=0}^{n-j} c_k \approx (n-j)^{1-p}$, 因此利用 $0 \leqslant j \leqslant n/2$, 方差至少为 n^{3-2p}. 更多细节留给读者.

8.3.1 混合性质

最后三个例子表明, 在许多情况下直接验证定理 8.3.2 的假设极易. 为将定理 8.3.2 与文献中的其他结果联系起来, 下面引入以 σ-域间依赖性定量测度表述的充分条件. 第一个是

$$\alpha(\mathcal{G},\mathcal{H}) = \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{G}, B \in \mathcal{H}\}.$$

若 $\alpha = 0$,$\mathcal{G}$ 和 $\mathcal{H}$ 独立, 因此 α 度量 σ-域间的依赖性.

引理 8.3.6. 设 $p,q,r \in (1,\infty]$ 满足 $1/p + 1/q + 1/r = 1$, 且设 $X \in \mathcal{G}$,$Y \in \mathcal{H}$ 满足 $E|X|^p, E|Y|^q < \infty$. 则

$$|EXY - EXEY| \leqslant 8\|X\|_p\|Y\|_q(\alpha(\mathcal{G},\mathcal{H}))^{1/r}.$$

这里, 当 $x > 0$ 时我们解释 $x^0 = 1$ 且 $0^0 = 0$.

证明. 若 $\alpha = 0$,X 和 Y 独立且结论成立, 因此可设 $\alpha > 0$. 我们分三步建立结果, 从 $r = \infty$ 的情形开始.

$$|EXY - EXEY| \leqslant 2\|X\|_p\|Y\|_q. \tag{8.3.2}$$

证明. Hölder 不等式蕴含 $|EXY| \leqslant \|X\|_p\|Y\|_q$, 且 Jensen 不等式蕴含

$$\|X\|_p\|Y\|_q \geqslant ||E|X|E|Y|| \geqslant |EXEY|,$$

因此结果由三角不等式得出. □

$$|EXY - EXEY| \leqslant 4\|X\|_\infty \|Y\|_\infty \alpha(\mathcal{G}, \mathcal{H}). \tag{8.3.3}$$

证明. 令 $\eta = \text{sgn}\{E(Y|\mathcal{G}) - EY\} \in \mathcal{G}.EXY = E(XE(Y|\mathcal{G}))$, 因此

$$|EXY - EXEY| = |E(X\{E(Y|\mathcal{G}) - EY\})| \leqslant \|X\|_\infty E|E(Y|\mathcal{G}) - EY|.$$

后者等于

$$= \|X\|_\infty E(\eta\{E(Y|\mathcal{G}) - EY\}) = \|X\|_\infty \{E(\eta Y) - E\eta EY\}.$$

对 $X = Y$ 和 $Y = \eta$ 应用最后一个结果给出

$$|E(Y\eta) - EYE\eta| \leqslant \|Y\|_\infty |E(\zeta\eta) - E\zeta E\eta|$$

其中 $\zeta = \text{sgn}\{E(\eta|\mathcal{H}) - E\eta\}$. 现在 $\eta = 1_A - 1_B$ 且 $\zeta = 1_C - 1_D$, 因此

$$\begin{aligned} |E(\zeta\eta) - E\zeta E\eta| &= |P(A \cap C) - P(B \cap C) - P(A \cap D) + P(B \cap D) \\ &\quad - P(A)P(C) + P(B)P(C) + P(A)P(D) - P(B)P(D)| \\ &\leqslant 4\alpha(\mathcal{G}, \mathcal{H}). \end{aligned}$$

结合最后三个等式给出所要结果. □

$$|EXY - EXEY| \leqslant 6\|X\|_p \|Y\|_\infty \alpha(\mathcal{G}, \mathcal{H})^{1-1/p}. \tag{8.3.4}$$

证明. 令 $C = \alpha^{-1/p}\|X\|_p$,$X_1 = X1_{(|X| \leqslant C)}$ 且 $X_2 = X - X_1$. 则

$$|EXY - EXEY| \leqslant |EX_1Y - EX_1EY| + |EX_2Y - EX_2EY| \leqslant 4\alpha C\|Y\|_\infty + 2\|Y\|_\infty E|X_2|$$

由 (8.3.3) 和 (8.3.2). 现在

$$E|X_2| \leqslant C^{-(p-1)}E(|X|^p 1_{(|X|>C)}) \leqslant C^{-p+1}E|X|^p.$$

结合最后两个不等式并使用 C 的定义给出

$$|EXY - EXEY| \leqslant 4\alpha^{1-1/p}\|X\|_p \|Y\|_\infty + 2\|Y\|_\infty \alpha^{1-1/p}\|X\|_p^{-p+1+p}$$

这证明了 (8.3.4). □

最后, 为证明引理 8.3.6, 令 $C = \alpha^{-1/q}\|Y\|_q$,$Y_1 = Y1_{(|Y|\leqslant C)}$, 且 $Y_2 = Y - Y_1$.

$$|EXY - EXEY| \leqslant |EXY_1 - EXEY_1| + |EXY_2 - EXEY_2| \leqslant 6C\|X\|_p\alpha^{1-1/p} + 2\|X\|_p\|Y_2\|_\theta$$

其中 $\theta = (1 - 1/p)^{-1}$ 由 (8.3.4) 和 (8.3.2). 现在

$$E|Y_2|^\theta \leqslant C^{-q+\theta}E(|Y|^q 1_{(|Y|>C)}) \leqslant C^{-q+\theta}E|Y|^q.$$

两边取 $1/\theta$ 次方并回顾 C 的定义,

$$\|Y_2\|_\theta \leqslant C^{-(q-\theta)/\theta}\|Y\|_q^{q/\theta} \leqslant \alpha^{(q-\theta)/\theta q}\|Y\|_q.$$

因此我们有

$$|EXY - EXEY| \leqslant 6\alpha^{-1/q}\|Y\|_q\|X\|_p\alpha^{1-1/p} + 2\|X\|_p\alpha^{1/\theta-1/q}\|Y\|_q^{1/\theta+1/q}$$

证毕. □

注释. 上述证明来自 Hall 和 Heyde (1980) 的附录 III. 他们将 (b) 归于 Ibragimov (1962),(c) 和引理 8.3.6 归于 Davydov (1968).

定理 8.3.7. 设 X_n,$n \in \mathbb{Z}$ 为满足 $EX_n = 0$,$E|X_0|^{2+\delta} < \infty$ 的遍历平稳序列. 令 $\alpha(n) = \alpha(\mathcal{F}_{-n}, \sigma(X_0))$, 其中 $\mathcal{F}_{-n} = \sigma(X_m : m \leqslant -n)$, 且设

$$\sum_{n=1}^{\infty} \alpha(n)^{\delta/2(2+\delta)} < \infty.$$

若 $S_n = X_1 + \cdots + X_n$, 则 $S_{(n\cdot)}/\sqrt{n} \Rightarrow \sigma B(\cdot)$, 其中

$$\sigma^2 = EX_0^2 + 2\sum_{n=1}^{\infty} EX_0X_n.$$

注释. 若加强混合条件, 则可减弱可求和条件. 令 $\bar{\alpha}(n) = \alpha(\mathcal{F}_{-n}, \mathcal{F}_0^0)$, 其中 $\mathcal{F}_0^0 = \sigma(X_k, k \geqslant 0)$. 当 $\bar{\alpha}(n) \downarrow 0$ 时, 称序列是强混合的. Rosenblatt (1956) 引入此概念作为获得平稳序列中心极限定理的条件. Ibragimov (1962) 在假设 $\sum_{n=1}^{\infty} \bar{\alpha}(n)^{\delta/(2+\delta)} < \infty$ 下证明了 $S_n/\sqrt{n} \Rightarrow \sigma\chi$, 其中 $\sigma^2 = \lim_{n\to\infty} ES_n^2/n$. 证明见 Ibragimov 和 Linnik (1971), 定理 18.5.3, 或 Hall 和 Heyde (1980), 推论 5.1.

证明. 为用引理 8.3.6 估计定理 8.3.2 中的量, 我们从

$$\|E(X|\mathcal{F})\|_2 = \sup\{E(XY) : Y \in \mathcal{F}, \|Y\|_2 = 1\}. \tag{8.3.5}$$

(8.3.5) 的证明. 若 $Y \in \mathcal{F}$ 满足 $\|Y\|_2 = 1$, 则使用条件期望的熟知性质和 Cauchy-Schwarz 不等式,

$$EXY = E(E(XY|\mathcal{F})) = E(YE(X|\mathcal{F})) \leqslant \|E(X|\mathcal{F})\|_2 \|Y\|_2.$$

当 $Y = E(X|\mathcal{F})/\|E(X|\mathcal{F})\|_2$ 时等号成立. □

在引理 8.3.6 中令 $p = 2+\delta$ 且 $q = 2$, 注意到

$$\frac{1}{r} = 1 - \frac{1}{p} - \frac{1}{q} = 1 - \frac{1}{2+\delta} - \frac{1}{2} = \frac{4+2\delta-2-(2+\delta)}{(2+\delta)2} = \frac{\delta}{(2+\delta)2}$$

且回忆 $EX_0 = 0$, 表明若 $Y \in \mathcal{F}_{-n}$ 则

$$|EX_0 Y| \leqslant 8\|X_0\|_{2+\delta}\|Y\|_2 \alpha(n)^{\delta/2(2+\delta)}.$$

结合 (8.3.5) 给出

$$\|E(X_0|\mathcal{F}_{-n})\|_2 \leqslant 8\|X_0\|_{2+\delta}\alpha(n)^{\delta/2(2+\delta)}$$

且由此可知定理 8.3.2 的假设得到满足. □

在 M 相依情形 (例 8.3.3), 对 $n > M$ 有 $\alpha(n) = 0$, 因此定理 8.3.7 适用. 对于 Markov 链 (例 8.3.4), 在此情形下

$$\begin{aligned}\alpha(n) &= \sup_{A,B} |P(X_{-n} \in A, X_0 \in B) - \pi(A)\pi(B)| \\ &= \sup_{A,B} \left| \sum_{x\in A, y\in B} \pi(x)p_n(x,y) - \pi(x)\pi(y) \right| \\ &\leqslant \sum_x \pi(x)^2 \|p_n(x,\cdot) - \pi(\cdot)\|,\end{aligned}$$

因此若我们对收敛到平衡的速度有足够了解, 则可检验定理 8.3.7 的假设.

最后, 为检验定理 8.3.7 中条件的好坏, 考虑滑动平均过程 (例 8.3.5) 的特殊情形, 其中 ξ_i 为独立同分布标准正态. 令

$$\rho(\mathcal{G},\mathcal{H}) = \sup\{\text{corr}(X,Y) : X \in \mathcal{G}, Y \in \mathcal{H}\}$$

其中

$$\text{corr}(X,Y) = \frac{EXY - EXEY}{\|X - EX\|_2 \|Y - EY\|_2}.$$

显然,$\alpha(\mathcal{G},\mathcal{H}) \leqslant \rho(\mathcal{G},\mathcal{H})$.Kolmogorov 和 Rozanov (1960) 证明了当 $\mathcal{G}$ 和 $\mathcal{H}$ 由 Gauss 随机变量生成时,$\rho(\mathcal{G},\mathcal{H}) \leqslant 2\pi\alpha(\mathcal{G},\mathcal{H})$. 他们通过证明 $\rho(\mathcal{G},\mathcal{H})$ 是 $L^2(\mathcal{G})$ 和 $L^2(\mathcal{H})$ 之间的夹角来证明这一点. 利用几何解释, 我们看到若 $|c_k|$ 递减, 则 $\alpha(n) \leqslant \bar{\alpha}(n) \leqslant |c_n|$, 因此定理 8.3.7 要求 $\sum_n |c_n|^{(1/2)-\varepsilon} < \infty$, 但 Ibragimov 的结果在 $\sum_n |c_n|^{1-\varepsilon} < \infty$ 时适用. 然而如前所述, 若 $\sum_n |c_n| < \infty$ 则中心极限定理成立.

第 4 节 经验分布与布朗桥

设 $X_1, X_2, \ldots$ 为独立同分布, 分布函数为 F. 令

$$\hat{F}_n(x) = \frac{1}{n}|\{m \leqslant n : X_m \leqslant x\}|$$

为经验分布. 在定理 2.4.7 中我们证明了

定理 8.4.1. Glivenko-Cantelli 定理. 当 $n \to \infty$,

$$\sup_x |\hat{F}_n(x) - F(x)| \to 0 \quad \text{a.s.}$$

本节中, 将研究 F 连续时收敛的速度. 施加这一限制以便通过令 $Y_n = F(X_n)$ 约化到 $(0,1)$ 上均匀分布的情形. 由于 $x \mapsto F(x)$ 是非减连续的, 且没有观测落在 F 的常数区间上, 容易看出若令

$$\hat{G}_n(y) = \frac{1}{n}|\{m \leqslant n : Y_m \leqslant y\}|$$

则

$$\sup_x |\hat{F}_n(x) - F(x)| = \sup_{0<y<1} |\hat{G}_n(y) - y|.$$

对本节其余部分, 假设 $Y_1, Y_2, \ldots$ 为 $(0,1)$ 上独立同分布均匀分布. 为应用 Donsker 定理, 将问题作变换. 将观测 $Y_1, \ldots, Y_n$ 按递增顺序排列:$U_1^n < U_2^n < \ldots < U_n^n$. 断言

$$\begin{aligned} \sup_{0<y<1} \hat{G}_n(y) - y &= \sup_{1\leqslant m\leqslant n} \frac{m}{n} - U_m^n, \\ \inf_{0<y<1} \hat{G}_n(y) - y &= \inf_{1\leqslant m\leqslant n} \frac{m-1}{n} - U_m^n. \end{aligned} \tag{8.4.1}$$

因为上确界在 $\hat{G}_n$ 的跳跃处达到而下确界恰在跳跃前达到. 为说明这一点, 参见图 8.1. 下证

$$D_n \equiv n^{1/2} \sup_{0<y<1} |\hat{G}_n(y) - y|$$

有极限, 因此下确界中的额外 $-1/n$ 没有影响.

第三个也是最后一个技巧是给出顺序统计量 $U_1^n < U_2^n < \ldots < U_n^n$ 的一个特殊构造. 设 $W_1, W_2, \ldots$ 为独立同分布, 满足 $P(W_i > t) = e^{-t}$, 且令 $Z_n = W_1 + \cdots + W_n$.

引理 8.4.2. $\{U_k^n : 1 \leqslant k \leqslant n\} \overset{d}{=} \{Z_k/Z_{n+1} : 1 \leqslant k \leqslant n\}$.

证明. 作变量替换 $v = r(t)$, 其中 $v_i = t_i/t_{n+1}$ 对 $i \leqslant n$,$v_{n+1} = t_{n+1}$. 反函数为

$$s(v) = (v_1 v_{n+1}, \ldots, v_n v_{n+1}, v_{n+1})$$

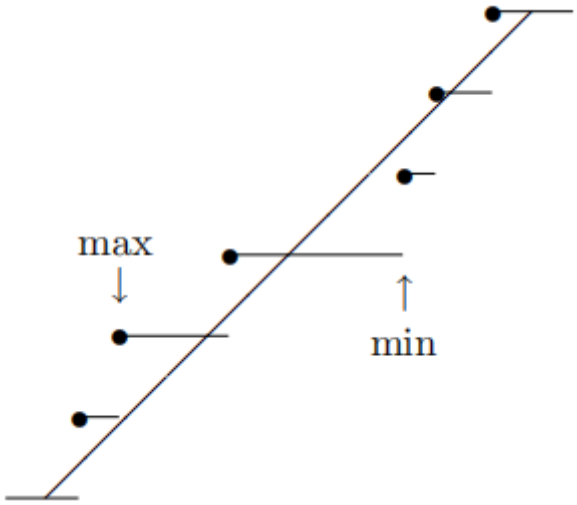

图 8.1: (8.4.1) 的图片证明

其偏导数矩阵 $\partial s_i/\partial v_j$ 由

$$\begin{pmatrix} v_{n+1} & 0 & \cdots & 0 & v_1 \\ 0 & v_{n+1} & \cdots & 0 & v_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & v_{n+1} & v_n \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

给出. 该矩阵的行列式为 v_{n+1}^n, 因此若令 $W = (V_1, \ldots, V_{n+1}) = r(Z_1, \ldots, Z_{n+1})$, 变量替换公式蕴含 W 有联合密度

$$f_W(v_1, \ldots, v_n, v_{n+1}) = \prod_{m=1}^{n} \lambda e^{-\lambda v_{n+1}(v_m - v_{m-1})} \cdot \lambda e^{-\lambda v_{n+1}(1-v_n)} v_{n+1}^n.$$

为求 $V = (V_1, \ldots, V_n)$ 的联合密度, 简化上述公式并对最后一个坐标积分得

$$f_V(v_1, \ldots, v_n) = \int_0^\infty \lambda^{n+1} v_{n+1}^n e^{-\lambda v_{n+1}} dv_{n+1} = n!$$

对 $0 < v_1 < v_2 < \ldots < v_n < 1$, 这就是所求的联合密度. □

现在转向 D_n 的极限律. 如上所述, 只需考虑

$$\begin{aligned} D_n^0 &= n^{1/2} \max_{1\leqslant m\leqslant n} \left| \frac{Z_m}{Z_{n+1}} - \frac{m}{n} \right| \\ &= \max_{1\leqslant m\leqslant n} \left| \frac{n}{Z_{n+1}} \cdot \frac{Z_m - m}{n^{1/2}} - \frac{m}{n} \cdot \frac{Z_{n+1} - n}{n^{1/2}} \right| \\ &= \max_{1\leqslant m\leqslant n} \left| \frac{Z_m - m}{n^{1/2}} - \frac{m}{n} \cdot \frac{Z_{n+1} - Z_n}{n^{1/2}} \right|. \end{aligned} \tag{8.4.2}$$

若令

$$B_n(t) = \begin{cases} (Z_m - m)/n^{1/2} & \text{若} t = m/n \text{ 且} m \in \{0, 1, \ldots, n\}, \\ \text{在} [(m-1)/n, m/n] \text{ 上线性} \end{cases}$$

则

$$D_n^0 = \frac{n}{Z_{n+1}} \max_{0\leqslant t\leqslant 1} \left| B_n(t) - t \cdot \frac{B_n(1) + Z_{n+1} - Z_n}{n^{1/2}} \right|.$$

强大数定律蕴含 $Z_{n+1}/n \to 1$ a.s., 因此第一个因子在极限中消失. 为求第二个因子的极限, 注意到 Donsker 定理 (定理 8.1.4) 蕴含 $B_n(\cdot) \Rightarrow B(\cdot)$, 一个布朗运动, 且计算二阶矩表明

$$(Z_{n+1} - Z_n)/n^{1/2} \to 0 \quad \text{依概率.}$$

$\psi(\omega) = \max_{0\leqslant t\leqslant 1} |\omega(t) - t\omega(1)|$ 是从 $C[0,1]$ 到 $\mathbb{R}$ 的连续函数, 因此由 Donsker 定理可得:

定理 8.4.3. $D_n \Rightarrow \max_{0\leqslant t\leqslant 1} |B_t - tB_1|$, 其中 B_t 是从 0 出发的布朗运动.

注释. Doob (1949) 建议用这一方法来推导 Kolmogorov 和 Smirnov 的结果, 后来由 Donsker (1952) 加以严格化. 我们的证明遵循 Breiman (1968).

为确定定理 8.4.3 中极限的分布, 先证明

$$\{B_t - tB_1, 0 \leqslant t \leqslant 1\} \stackrel{d}{=} \{B_t, 0 \leqslant t \leqslant 1 \mid B_1 = 0\} \tag{8.4.3}$$

一个我们将记为 B_t^0 并称之为**布朗桥**的过程. 事件 $B_1 = 0$ 的概率为 0, 但容易看出条件概率应如何理解. 若 $0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$,$x_0 = 0$,$x_{n+1} = 0$, 且 $x_1, \ldots, x_n \in \mathbb{R}$, 则

$$P(B(t_1) = x_1, \ldots, B(t_n) = x_n \mid B(1) = 0) = \frac{1}{p_1(0,0)} \prod_{m=1}^{n+1} p_{t_m - t_{m-1}}(x_{m-1}, x_m) \tag{8.4.4}$$

其中 $p_t(x,y) = (2\pi t)^{-1/2} \exp(-(y-x)^2/2t)$.

(8.4.3) 的证明. 公式 (8.4.4) 表明 B_t^0 的有限维分布是多元正态的且均值为 0. 由于 $B_t - tB_1$ 也有此

性质, 只需证明协方差相等. 我们从较容易的计算开始. 若 $s<t$ 则

$$E((B_s - sB_1)(B_t - tB_1)) = s - st - st + st = s(1-t). \tag{8.4.5}$$

对另一过程,$P(B_s^0 = x, B_t^0 = y)$ 为

$$\begin{aligned}&\frac{\exp(-x^2/2s)}{(2\pi s)^{1/2}} \cdot \frac{\exp(-(y-x)^2/2(t-s))}{(2\pi(t-s))^{1/2}} \cdot \frac{\exp(-y^2/2(1-t))}{(2\pi(1-t))^{1/2}} \cdot (2\pi)^{1/2} \\ &= (2\pi)^{-1}(s(t-s)(1-t))^{-1/2} \exp(-(ax^2 + 2bxy + cy^2)/2)\end{aligned} \tag{8.4.6}$$

其中

$$a = \frac{1}{s} + \frac{1}{t-s} = \frac{t}{s(t-s)}, \quad b = -\frac{1}{t-s},$$
$$c = \frac{1}{t-s} + \frac{1}{1-t} = \frac{1-s}{(t-s)(1-t)}.$$

回顾第 3.9 节结尾的讨论并注意到

$$\begin{pmatrix} \frac{t}{s(t-s)} & \frac{-1}{t-s} \\ \frac{-1}{t-s} & \frac{1-s}{(1-s)(1-t)} \end{pmatrix}^{-1} = \begin{pmatrix} s(1-s) & s(1-t) \\ s(1-t) & t(1-t) \end{pmatrix}$$

(矩阵相乘!) 表明 (8.4.3) 成立. □

注释. 为使用联合密度进行计算, 知道若 (X,Y) 有均值为 0 的二元正态分布,$EX^2 = \sigma_X^2$,$EY^2 = \sigma_Y^2$, 相关系数为 ρ, 则联合密度由

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_X^2} - \frac{2xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right]\right)$$

给出是有用的. 在 (8.4.6) 中若 $X = B_s^0$ 且 $Y = B_t^0$, 则 $\sigma_X^2 = s(1-s)$,$\sigma_Y^2 = t(1-t)$ 且

$$\rho = \frac{s(1-t)}{\sqrt{s(1-s)t(1-t)}} = \sqrt{\frac{s(1-t)}{(1-s)t}}$$

因此我们有

$$1 - \rho^2 = \frac{(1-s)t - s(1-t)}{(1-s)t} = \frac{t-s}{(1-s)t}.$$

下面将验证以下熟知事实的工作留给读者: 条件分布

$$(Y \mid X = x) = \text{normal}(x\rho\sigma_Y/\sigma_X, (1-\rho^2)\sigma_Y^2)$$

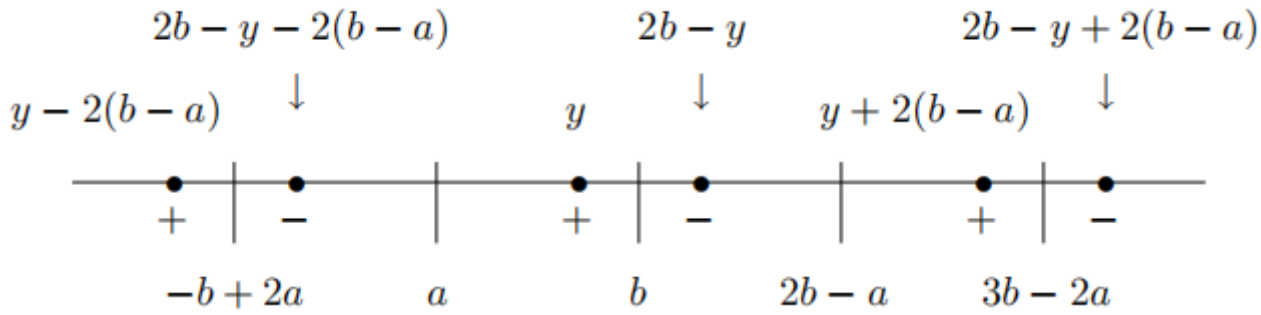

图 8.2: (8.4.8) 中的无穷级数，注意：从 a 或 b 点看去，正负符号（+ 和 −）的排列呈反对称分布

在我们的特殊情形下给出

$$(B_t^0 \mid B_s^0 = x) = \text{normal}(x(1-t)/(1-s), (1-t)(1-s)/(1-s)). \tag{8.4.7}$$

研究 D_n 极限分布的最后一步是计算 $\max_{0\leqslant t\leqslant 1}|B_t^0|$ 的分布. 为此, 先证明

定理 8.4.4. 在 $\{T_a \wedge T_b > t\}$ 上 B_t 的密度函数为

$$P_x(T_a \wedge T_b > t, B_t = y) = \sum_{n=-\infty}^{\infty} P_x(B_t = y + 2n(b-a)) - P_x(B_t = 2b - y + 2n(b-a)) \tag{8.4.8}$$

证明. 首先注意到若 $A \subset (a, b)$,

$$P_x(T_a \wedge T_b > t, B_t \in A) = P_x(B_t \in A) - P_x(T_a < T_b, T_a < t, B_t \in A) - P_x(T_b < T_a, T_b < t, B_t \in A). \tag{8.4.9}$$

若令 $\rho_a(y) = 2a - y$ 为关于 a 的反射且注意到 $\{T_a < T_b\}$ 是 $\mathcal{F}(T_a)$ 可测的, 则由 (7.4.5) 的证明可得

$$P_x(T_a < T_b, T_a < t, B_t \in A) = P_x(T_a < T_b, B_t \in \rho_a A)$$

其中 $\rho_a A = \{\rho_a(y) : y \in A\}$. 为消去 $T_a < T_b$, 注意到

$$P_x(T_a < T_b, B_t \in \rho_a A) = P_x(B_t \in \rho_a A) - P_x(T_b < T_a, B_t \in \rho_a A).$$

注意到 $B_t \in \rho_a A$ 和 $T_b < T_a$ 蕴含 $T_a < t$, 并再次使用反射原理给出

$$\begin{aligned} P_x(T_b < T_a, B_t \in \rho_a A) &= P_x(T_b < T_a, B_t \in \rho_b \rho_a A) \\ &= P_x(B_t \in \rho_b \rho_a A) - P_x(T_a < T_b, B_t \in \rho_b \rho_a A). \end{aligned}$$

重复最后两个计算 n 次更多给出

$$P_x(T_a < T_b, B_t \in \rho_a A) = \sum_{m=0}^{n} P_x(B_t \in \rho_a(\rho_b\rho_a)^m A) - P_x(B_t \in (\rho_b\rho_a)^{m+1}A) + P_x(T_a < T_b, B_t \in (\rho_b\rho_a)^{n+1}A).$$

每对反射将 A 推离 0 更远, 因此令 $n \to \infty$ 表明

$$P_x(T_a < T_b, B_t \in \rho_a A) = \sum_{m=0}^{\infty} P_x(B_t \in \rho_a(\rho_b\rho_a)^m A) - P_x(B_t \in (\rho_b\rho_a)^{m+1}A).$$

交换 a 和 b 的角色给出

$$P_x(T_b < T_a, B_t \in \rho_b A) = \sum_{m=0}^{\infty} P_x(B_t \in \rho_b(\rho_a\rho_b)^m A) - P_x(B_t \in (\rho_a\rho_b)^{m+1}A).$$

结合最后两个表达式与 (8.4.9) 并使用 $\rho_c^{-1} = \rho_c$,$(\rho_a\rho_b)^{-1} = \rho_b^{-1}\rho_a^{-1}$ 给出

$$P_x(T_a \wedge T_b > t, B_t \in A) = \sum_{m=-\infty}^{\infty} P_x(B_t \in (\rho_b\rho_a)^n A) - P_x(B_t \in \rho_a(\rho_b\rho_a)^n A).$$

为准备应用, 令 $A = (u, v)$ 其中 $a < u < v < b$, 注意 $\rho_b\rho_a(y) = y + 2(b - a)$, 并对第二个和式作变量替换得

$$P_x(T_a \wedge T_b > t, u < B_t < v) = \sum_{n=-\infty}^{\infty} \{P_x(u + 2n(b-a) < B_t < v + 2n(b-a)) - P_x(2b - v + 2n(b-a) < B_t < 2b - u + 2n(b-a))\}. \tag{8.4.10}$$

令 $u = y - \varepsilon$,$v = y + \varepsilon$, 两边除以 2ε, 并令 $\varepsilon \to 0$（验证控制收敛定理适用留给读者）即得所要结果. □

令 $x = y = 0, t = 1$, 除以 $(2\pi)^{-1/2} = P_0(B_1 = 0)$, 得到关于布朗桥 B_t^0 的结果:

$$P\left(a < \min_{0\leqslant t\leqslant 1} B_t^0 < \max_{0\leqslant t\leqslant 1} B_t^0 < b\right) = \sum_{n=-\infty}^{\infty} e^{-(2n(b-a))^2/2} - e^{-(2b+2n(b-a))^2/2}. \tag{8.4.11}$$

取 $a = -b$, 有

$$P\left(\max_{0\leqslant t\leqslant 1} |B_t^0| < b\right) = \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2b^2}. \tag{8.4.12}$$

这一公式给出 Kolmogorov-Smirnov 统计量的分布, 可用于检验独立同分布序列 $X_1, \ldots, X_n$ 是否有分布 F. 为此, 将数据变换为 $F(X_n)$ 并考察经验分布与均匀分布之间的最大差异.(8.4.12) 给出当 X_i 有分布 F 时误差的分布.

习题

习题 8.4.1. 利用习题 7.4.8 和导致 (8.4.11) 的推理得出

$$P\left(\max_{0\leqslant t\leqslant 1} B_t^0 > b\right) = \exp(-2b^2).$$

习题 8.4.2. 设 B_t 为从 0 出发的布朗运动. 证明 $X_t = (1-t)B(t/(1-t))$ 是布朗桥.

习题 8.4.3. 证明布朗桥是 Markov 过程, 通过检验若 $s_1 < \ldots < s_n < s < t < 1$

$$P(B(t) = y \mid B(s) = x, B(s_n) = x_n, \ldots, B(s_1) = x_1) = P(B(t) = y \mid B(s) = x).$$

习题 8.4.4. 设 B_t^0 为布朗桥. 证明 $X_t = B_t^0/(1-t)$ 是鞅但不是 L^2 有界的.

第 5 节 重对数律

本节首先证明如下结论：

定理 8.5.1. 布朗运动的重对数律.

$$\limsup_{t\to\infty}\frac{B_t}{(2t\log\log t)^{1/2}}=1\quad\text{a.s.}$$

这里 LIL 是 “law of the iterated logarithm”（重对数律）的缩写，这一名称源于分母中的 $\log\log t$. 一旦定理 8.5.1 得证，即可利用 Skorokhod 表示来证明均值为零且方差有限的随机游走的相应结果.

证明. 证明的关键在于公式 (7.4.4)：

$$P_0\left(\max_{0\leqslant s\leqslant 1}B_s>a\right)=P_0(T_a\leqslant 1)=2P_0(B_1\geqslant a). \tag{8.5.1}$$

为控制右端，使用引理 1.2.6 中的结果. 当 $x>0$ 时，

$$(x^{-1}-x^{-3})\exp(-x^2/2)\leqslant\int_x^\infty\exp(-y^2/2)\,\mathrm{d}y\leqslant x^{-1}\exp(-x^2/2), \tag{8.5.2}$$

因此

$$\int_x^\infty\exp(-y^2/2)\,\mathrm{d}y\sim\frac{1}{x}\exp(-x^2/2),\quad x\to\infty. \tag{8.5.3}$$

由 (8.5.3) 和布朗尺度变换可得

$$P_0(B_t>(tf(t))^{1/2})\sim\kappa f(t)^{-1/2}\exp(-f(t)/2),$$

其中 $\kappa=(2\pi)^{-1/2}$ 是一个常数，下文将尽量忽略它. 上述结果表明，若 $\epsilon>0$，则

$$\sum_{n=1}^\infty P_0(B_n>(nf(n))^{1/2})\begin{cases}<\infty & \text{当}f(n)=(2+\epsilon)\log n,\\ =\infty & \text{当}f(n)=(2-\epsilon)\log n,\end{cases}$$

因此由 Borel-Cantelli 引理可得

$$\limsup_{n\to\infty}\frac{B_n}{(2n\log n)^{1/2}}\leqslant 1\quad\text{a.s.}$$

为将 $\log n$ 替换为 $\log\log n$，需沿指数增长的序列考察. 设 $t_n=\alpha^n$，其中 $\alpha>1$. 由 (8.5.1) 和

(8.5.2) 可得

$$P_0\left(\max_{t_n\leqslant s\leqslant t_{n+1}} B_s > (t_n f(t_n))^{1/2}\right) \leqslant P_0\left(\max_{0\leqslant s\leqslant t_{n+1}} \frac{B_s}{t_{n+1}^{1/2}} > \left(\frac{f(t_n)}{\alpha}\right)^{1/2}\right)$$
$$\leqslant 2\kappa(f(t_n)/\alpha)^{-1/2}\exp(-f(t_n)/2\alpha).$$

若 $f(t)=2\alpha^2\log\log t$，则

$$\log\log t_n = \log(n\log\alpha) = \log n + \log\log\alpha,$$

因此 $\exp(-f(t_n)/2\alpha)\leqslant C_\alpha n^{-\alpha}$，其中 C_α 是仅依赖于 α 的常数，故

$$\sum_{n=1}^{\infty} P_0\left(\max_{t_n\leqslant s\leqslant t_{n+1}} B_s > (t_n f(t_n))^{1/2}\right) < \infty.$$

由于 $t\mapsto (tf(t))^{1/2}$ 递增且 $\alpha>1$ 任意，可得

$$\limsup_{t\to\infty} \frac{B_t}{(2t\log\log t)^{1/2}} \leqslant 1. \tag{8.5.4}$$

为证定理 8.5.1 的另一半，再次设 $t_n=\alpha^n$，但这次 α 将取较大的值，因为为得到独立事件，将考察

$$P_0\left(B(t_{n+1})-B(t_n) > (t_{n+1}f(t_{n+1}))^{1/2}\right) = P_0\left(B_1 > (\beta f(t_{n+1}))^{1/2}\right),$$

其中 $\beta = t_{n+1}/(t_{n+1}-t_n) = \alpha/(\alpha-1) > 1$. 当 n 充分大时，由 (8.5.3)，上述量

$$\geqslant \frac{\kappa}{2}(\beta f(t_{n+1}))^{-1/2}\exp(-\beta f(t_{n+1})/2).$$

若 $f(t)=(2/\beta^2)\log\log t$，则 $\log\log t_n = \log n+\log\log\alpha$，因此

$$\exp(-\beta f(t_{n+1})/2) \geqslant C_\alpha n^{-1/\beta},$$

其中 C_α 是仅依赖于 α 的常数，故

$$\sum_{n=1}^{\infty} P_0\left(B(t_{n+1})-B(t_n) > (t_{n+1}f(t_{n+1}))^{1/2}\right) = \infty.$$

由于所考虑的事件相互独立，由第二 Borel-Cantelli 引理可得

$$B(t_{n+1})-B(t_n) > \left((2/\beta^2)t_{n+1}\log\log t_{n+1}\right)^{1/2} \quad \text{i.o.} \tag{8.5.5}$$

由 (8.5.4) 可得

$$\limsup_{n\to\infty} \frac{B(t_n)}{(2t_n \log\log t_n)^{1/2}} \leqslant 1. \tag{8.5.6}$$

由于 $t_n = t_{n+1}/\alpha$ 且 $t \mapsto \log\log t$ 递增，结合 (8.5.5) 和 (8.5.6)，并回忆 $\beta = \alpha/(\alpha-1)$ 可得

$$\limsup_{n\to\infty} \frac{B(t_{n+1})}{(2t_{n+1} \log\log t_{n+1})^{1/2}} \geqslant \frac{\alpha-1}{\alpha} - \frac{1}{\alpha^{1/2}}.$$

令 $\alpha \to \infty$ 即得所需的下界，定理 8.5.1 证毕. □

转向随机游走，下面证明 Hartman 与 Wintner (1941) 的一个结果：

定理 8.5.2. 若 $X_1, X_2, \ldots$ 独立同分布，满足 $EX_i = 0$ 且 $EX_i^2 = 1$，则

$$\limsup_{n\to\infty} \frac{S_n}{(2n \log\log n)^{1/2}} = 1 \quad \text{a.s.}$$

证明. 由定理 8.1.2 可知，可写 $S_n = B(T_n)$，其中 $T_n/n \to 1$ a.s. 如同 Donsker 定理的证明，下面仅需这一事实. 一旦证明

$$\frac{S_{[t]} - B_t}{(t \log\log t)^{1/2}} \to 0 \quad \text{a.s.}, \tag{8.5.7}$$

定理 8.5.2 将可由定理 8.5.1 推出.

为此，首先注意到若 $\epsilon > 0$ 且 $t \geqslant t_0(\omega)$，则

$$T_{[t]} \in [t/(1+\epsilon), t(1+\epsilon)]. \tag{8.5.8}$$

为估计 $S_{[t]} - B_t$，设

$$M(t) = \sup\{|B(s) - B(t)| : t/(1+\epsilon) \leqslant s \leqslant t(1+\epsilon)\}.$$

为控制上述量，设 $t_k = (1+\epsilon)^k$ 并注意到若 $t_k \leqslant t \leqslant t_{k+1}$，则

$$\begin{aligned} M(t) &\leqslant \sup\{|B(s) - B(t)| : t_{k-1} \leqslant s, t \leqslant t_{k+2}\} \\ &\leqslant 2\sup\{|B(s) - B(t_{k-1})| : t_{k-1} \leqslant s \leqslant t_{k+2}\}. \end{aligned}$$

注意 $t_{k+2}-t_{k-1}=\delta t_{k-1}$，其中 $\delta=(1+\epsilon)^3-1$，尺度变换蕴含

$$\begin{aligned}
&P\left(\max_{t_{k-1}\leqslant s\leqslant t_{k+2}}|B(s)-B(t)|>(3\delta t_{k-1}\log\log t_{k-1})^{1/2}\right)\\
&=P\left(\max_{0\leqslant r\leqslant 1}|B(r)|>(3\log\log t_{k-1})^{1/2}\right)\\
&\leqslant 2\kappa(3\log\log t_{k-1})^{-1/2}\exp(-3\log\log t_{k-1}/2)
\end{aligned}$$

由 (8.5.1) 和 (8.5.2) 的熟悉应用. 对 k 求和并使用 Borel-Cantelli 引理可得

$$\limsup_{t\to\infty}\frac{S_{[t]}-B_t}{(t\log\log t)^{1/2}}\leqslant(3\delta)^{1/2}.$$

由 $\delta=(1+\epsilon)^3-1$ 并令 $\epsilon\downarrow 0$，(8.5.7) 得证，证毕. □

Strassen (1965) 证明了精确的逆命题：若定理 8.5.2 成立，则 $EX_i=0$ 且 $EX_i^2=1$. 他对这一课题的另一贡献是：

定理 8.5.3. Strassen (1964) 不变原理. 设 $X_1,X_2,\ldots$ 独立同分布，满足 $EX_i=0$ 且 $EX_i^2=1$，令 $S_n=X_1+\cdots+X_n$，并令 $S(n\cdot)$ 为通常的线性插值. 对 $n\geqslant 3$，

$$Z_n(\cdot)=(2n\log\log n)^{-1/2}S(n\cdot)$$

的极限集（即收敛子列的极限全体）为

$$\mathcal{K}=\left\{f:f(x)=\int_0^x g(y)\,\mathrm{d}y,\int_0^1 g(y)^2\,\mathrm{d}y\leqslant 1\right\}.$$

Jensen 不等式蕴含 $f(1)^2\leqslant\int_0^1 g(y)^2\,\mathrm{d}y\leqslant 1$，等号成立当且仅当 $f(t)=t$，因此定理 8.5.3 包含定理 8.5.2 作为特例，并提供了关于 S_n 的大值如何产生的信息.

习题

习题 8.5.1. 证明若 $E|X_i|^\alpha=\infty$ 对某个 $\alpha<2$ 成立，则

$$\limsup_{n\to\infty}\frac{|X_n|}{n^{1/\alpha}}=\infty\quad\text{a.s.},$$

因此重对数律不成立.

习题 8.5.2. 在定理 8.5.3 的假设下，直接证明 $\{S_n/(2n\log\log n)^{1/2}\}$ 的极限集为 $[-1,1]$.

第九章
多维布朗运动

本章旨在建立布朗运动与各种抛物型和椭圆型偏微分方程之间的联系. 第一节用鞅来计算多维布朗运动的相关量.

第 1 节　鞅

记 $\Delta f = \sum_{i=1}^{d} \partial^2 f/\partial x_i^2$ 为 f 的**拉普拉斯算子**. 将 Itô 公式 (7.6.9) 与定理 7.6.4 结合，得

定理 9.1.1. 假设 $v \in C^2$, 即 v 的所有一、二阶偏导数存在且连续, 且对所有 t 有 $E\int_0^t |\nabla v(B_s)|^2 \,\mathrm{d}s < \infty$. 则

$$v(B_t) - \int_0^t \frac{1}{2}\Delta v(B_s)\,\mathrm{d}s$$

是连续鞅.

记 $S_r = \inf\{t : |B_t| = r\}$ $(r < R)$，且 $\tau = S_r \wedge S_R$. 首先需要指出的是，若 $|x| < R$ 则 $P_x(S_R < \infty) = 1$. 确定这一点后，便可得出

定理 9.1.2. 若 $|x| < R$ 则 $E_x S_R = (R^2 - |x|^2)/d$.

证明. 由定理 7.5.4 可知 $|B_t|^2 - dt = \sum_{i=1}^{d}(B_t^i)^2 - t$ 是鞅. 定理 7.5.1 蕴含 $|x|^2 = E|B_{S_R\wedge t}|^2 - dE(S_R \wedge t)$. 令 $t \to \infty$ 即得所需结果. □

为进一步研究布朗运动，引入两个新函数：

$$\phi(x) = \begin{cases} \log|x| & d = 2, \\ |x|^{2-d} & d \geqslant 3. \end{cases}$$

读者可自行验证，在每种情形下都有 $\Delta\phi = 0$.

引理 9.1.3. $\phi(x) = E_x\phi(B_\tau)$.

证明. 取 $\psi(x) = g(|x|)$ 为 C^2 函数且具有紧支集，且当 $r < |x| < R$ 时 $\psi(x) = \phi(x)$. 定理 9.1.1 蕴含 $\psi(x) = E_x\psi(B_{t\wedge\tau})$. 令 $t \to \infty$ 即得所需结果. □

引理 9.1.3 蕴含

$$\phi(x) = \phi(r)P_x(S_r < S_R) + \phi(R)(1 - P_x(S_r < S_R))$$

其中 $\phi(r)$ 是 $\phi(x)$ 在 $\{x : |x| = r\}$ 上的值的简写. 解此方程可得

$$P_x(S_r < S_R) = \frac{\phi(R) - \phi(x)}{\phi(R) - \phi(r)}. \tag{9.1.1}$$

当 $d = 2$ 时，上式变为

$$P_x(S_r < S_R) = \frac{\log R - \log|x|}{\log R - \log r}. \tag{9.1.2}$$

若固定 r 并令 $R \to \infty$，则 (9.1.2) 右端趋于 1. 故对任意 x 和任意 $r > 0$ 有

$$P_x(S_r < \infty) = 1.$$

由此可见，二维布朗运动在下述意义下是常返的：若 G 为任意开集，则 $P_x(B_t \in G \text{ i.o.}) \equiv 1$.

若固定 R，令 $r \to 0$ 于 (9.1.2) 中，并记 $S_0 = \inf\{t > 0 : B_t = 0\}$，则对 $x \neq 0$ 有

$$P_x(S_0 < S_R) \leqslant \lim_{r\to 0} P_x(S_r < S_R) = 0.$$

由于这对所有 R 成立，且布朗路径的连续性蕴含 $S_R \uparrow \infty$（当 $R \uparrow \infty$），故对所有 $x \neq 0$ 有 $P_x(S_0 < \infty) = 0$. 为将这一结果推广到 $x = 0$ 的情形，注意 Markov 性质蕴含

$$P_0(B_t = 0 \text{ 对某个} t \geqslant \epsilon) = E_0[P_{B_\epsilon}(T_0 < \infty)] = 0$$

对所有 $\epsilon > 0$ 成立，故 $P_0(B_t = 0 \text{ 对某个} t > 0) = 0$. 根据 $S_0 = \inf\{t > 0 : B_t = 0\}$ 的定义，有

$$P_x(S_0 < \infty) = 0 \quad \text{对所有} x. \tag{9.1.3}$$

因此，当 $d \geqslant 2$ 时，布朗运动即使在起点也不会在正时刻击中 0. 作为推论，三维布朗运动不会击中一条直线.

对 $d \geqslant 3$，公式 (9.1.1) 给出

$$P_x(S_r < S_R) = \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}. \tag{9.1.4}$$

此处固定 R 并令 $r \to 0$ 没有意义. 二维布朗运动不击中 0 的事实蕴含三维布朗运动也不击中 0，实际上也不会击中直线 $\{x : x_1 = x_2 = 0\}$. 若固定 r 并令 $R \to \infty$ 于 (9.1.4) 中，得

$$P_x(S_r < \infty) = (r/|x|)^{d-2} < 1 \quad 若|x| > r. \tag{9.1.5}$$

由这一结果易知，当 $d \geqslant 3$ 时，布朗运动是暂态的，即其不会无穷多次返回任何有界集.

定理 9.1.4. 当 $t \to \infty$ 时，$|B_t| \to \infty$ a.s.

证明. 记 $A_n = \{|B_t| > n^{1-\epsilon}$ 对所有$t \geqslant S_n\}$. 强 Markov 性质蕴含

$$P_x(A_n^c) = E_x(P_{B(S_n)}(S_{n^{1-\epsilon}} < \infty)) = (n^{1-\epsilon}/n)^{d-2} \to 0$$

（当 $n \to \infty$）. 现 $\limsup A_n = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n$ 满足

$$P(\limsup A_n) \geqslant \limsup P(A_n) = 1.$$

故无穷多次地，布朗路径在时刻 S_n 之后永不返回 $\{x : |x| \leqslant n^{1-\epsilon}\}$，这蕴含所需结果. □

尺度关系 (7.1.1) 蕴含 $S_n \overset{d}{=} n^2 S_1$，故定理 9.1.4 的证明表明

$$|B_t|/t^{(1-\epsilon)/2} \to \infty.$$

Dvoretsky 与 Erdős (1951) 证明了关于 $d \geqslant 3$ 时布朗运动趋于 ∞ 的速度的下述结果.

定理 9.1.5. 设 $g(t)$ 为正递减函数. 则

$$P_0(|B_t| \leqslant g(t)\sqrt{t} \text{ i.o. 当} t \uparrow \infty) = 1 \text{ 或} 0$$

视 $\int^{\infty} g(t)^{d-2}/t \, \mathrm{d}t = \infty$ 或 $< \infty$ 而定.

此处积分下限的省略意味着只关心积分 “在 ∞ 附近” 的行为. 简单计算表明

$$\int^{\infty} t^{-1}(\log t)^{-\alpha} \, \mathrm{d}t = \infty \text{ 或} < \infty$$

视 $\alpha \leqslant 1$ 或 $\alpha > 1$ 而定，故 B_t 趋于 ∞ 的速度快于 $t^{1/2}(\log t)^{-\alpha/(d-2)}$（对任意 $\alpha > 1$）.

习题

习题 9.1.1. 将尺度变换与 0-1 律结合，证明

$$\liminf_{t\to\infty}|B_t|/t^{1/2}=0 \quad \text{a.s.}$$

习题 9.1.2. 设 $f\in C^2$ 非负且上调和. 证明在一维和二维情形，f 必为常数. 在 $d\geqslant 3$ 时给出一个形如 $f(x)=g(|x|)$ 的例子.

习题 9.1.3. 设 $f\in C^2$ 有界且调和. 则 f 必为常数（在任何维数下）.

第 2 节 热方程

本节考虑如下方程：

(a) $u \in C^{1,2}$ 且在 $(0,\infty)\times\mathbb{R}^d$ 中 $u_t = \frac{1}{2}\Delta u$.

(b) u 在 $\{0\}\times\mathbb{R}^d$ 的每点连续，且 $u(0,x) = f(x)$，其中 f 有界连续.

此方程得名于如下事实：若适当选取度量单位，则解 $u(t,x)$ 给出时刻 0 温度分布为 $f(x)$ 时，$\mathbb{R}^d$ 中点 x 在时刻 t 的温度.

显然，除非 f 连续，(b) 不可能成立. 以下结果可推广到无界 f 的情形. 为简化起见，假设 f 有界. 在本节及接下来五节中，开头陈述的假设贯穿全节.

解此方程的第一步（如下文六次所做）是构造局部鞅 M_s（$s\in[0,t]$），即存在递增停时列使 $M(s\wedge T_n)$ 为鞅且 $P(T_n<t)\to 0$ 的过程.

定理 9.2.1. 若 u 满足 (a)，则 $M_s = u(t-s,B_s)$ 是 $[0,t]$ 上的局部鞅.

证明. 应用 Itô 公式 (7.6.9) 得

$$u(t-s,B_s)-u(t,B_0) = \int_0^s -u_t(t-r,B_r)\,\mathrm{d}r + \int_0^s \nabla u(t-r,B_r)\cdot \mathrm{d}B_r + \frac{1}{2}\int_0^s \Delta u(t-r,B_r)\,\mathrm{d}r.$$

结果随之得出，因为 $-u_t+\frac{1}{2}\Delta u = 0$ 且右端第二项是局部鞅（由定理 7.6.4）. 此处需要停时以确保可积性条件成立. □

下一步是证明唯一性定理.

定理 9.2.2. 若存在有界解，则它必为

$$v(t,x) \equiv E_x f(B_t).$$

此处 $\equiv$ 表示最后一式定义 v. 下面总用 u 表示方程的通解，用 v 表示特解.

证明. 设 u 有界，则 M_s（$0\leqslant s<t$）是有界鞅. 鞅收敛定理蕴含

$$M_t \equiv \lim_{s\uparrow t} M_s$$

几乎必然存在.

若 u 满足 (b)，此极限必为 $f(B_t)$. 由于 M_s 一致可积，故

$$u(t,x) = E_x M_0 = E_x M_t = v(t,x).$$

□

既然定理 9.2.2 已给出解的形式，下一步逻辑上便是寻找 v 为解的条件. 此处及后面五个例子中，容易证明若 v 足够光滑则它是解.

定理 9.2.3. 若 $v \in C^{1,2}$，则它满足 (a).

证明. 考虑运行时空布朗运动 $(t-s, B_s)$ 来构造解，可得

$$v(t+h,x) = E_x v(t, B_h).$$

两边减去 $v(t,x)$ 并用 Taylor 级数

$$\begin{aligned} v(t+h,x) - v(t,x) = E_x \Bigg(& \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}(t,x)(B_i(h) - x_i) \\ & + \frac{1}{2} \sum_{1 \leqslant i,j \leqslant d} \frac{\partial^2 v}{\partial x_i \partial x_j}(t,x)(B_i(h) - x_i)(B_j(h) - x_j) \Bigg) + o(h). \end{aligned}$$

由于当 $i \neq j$ 时，$B_i(t)$，$B_i^2(t) - t$ 和 $B_i(t)B_j(t)$ 都是鞅

$$v(t+h,x) - v(t,x) = \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 v}{\partial x_i^2}(t,x)h + o(h).$$

除以 h 并令 $h \to 0$ 即得所需结果. □

定理 9.2.4. v 满足 (b).

证明. $(B_t - B_0) \overset{d}{=} t^{1/2}\chi$，其中 χ 服从均值为 0、方差为 1 的正态分布，故当 $t_n \to 0$ 且 $x_n \to x$ 时，有界收敛定理蕴含

$$v(t_n, x_n) = Ef(x_n + t_n^{1/2}\chi) \to f(x),$$

这就证明了所需结果. □

证明 v 为解的最后一步是寻找保证它光滑的条件.

定理 9.2.5. $v \in C^{1,2}$，故满足 (a).

证明. 我们的解是

$$v(t,x) = E_x f(B_t) = \int (2\pi t)^{-d/2} e^{-|x-y|^2/2t} f(y)\, \mathrm{d}y.$$

用微积分和分析的一些基本事实可证明所需结果. 细节枯燥乏味，留给读者. □

第 3 节 非齐次热方程

本节考虑当给热方程添加有界连续函数 $g(x)$ 时会发生什么. 即研究

(a) $u \in C^{1,2}$ 且在 $(0,\infty)\times\mathbb{R}^d$ 中 $u_t = \frac{1}{2}\Delta u + g$.

(b) u 在 $\{0\}\times\mathbb{R}^d$ 的每点连续，且 $u(0,x) = f(x)$ 有界连续.

我们在上一节已观察到，除非 f 连续，(b) 不能成立. 此处 $g = u_t - (1/2)\Delta u$，故当 $u \in C^{1,2}$ 时，(a) 中的方程不能成立，除非 g 连续.

处理新方程的第一步是注意到：若 u_1 为上一节研究的方程（$f = f_0$，$g = 0$）的解，u_2 为方程（$f = 0$，$g = g_0$）的解，则 $u_1 + u_2$ 为方程（$f = f_0$，$g = g_0$）的解，故只需考虑 $f \equiv 0$ 的情形.

做了这一简化后，下面用上一节的方法研究上述方程.

定理 9.3.1. 若 u 满足 (a)，则

$$M_s = u(t-s, B_s) + \int_0^s g(B_r)\,\mathrm{d}r$$

是 $[0,t]$ 上的局部鞅.

证明. 如定理 9.2.1 的证明，应用 Itô 公式得

$$\begin{aligned} u(t-s,B_s) - u(t,B_0) = &\int_0^s (-u_t + \tfrac{1}{2}\Delta u)(t-r,B_r)\,\mathrm{d}r \\ &+ \int_0^s \nabla u(t-r,B_r)\cdot \mathrm{d}B_r, \end{aligned}$$

这就证明了结果，因为 $-u_t + \frac{1}{2}\Delta u = -g$ 且右端第二项是局部鞅（由定理 7.6.4）. □

下一步仍是唯一性结果.

定理 9.3.2. 若存在解在 $[0,T]\times\mathbb{R}^d$ 上有界（对任意 $T < \infty$），则它必为

$$v(t,x) \equiv E_x\left(\int_0^t g(B_s)\,\mathrm{d}s\right).$$

证明. 在关于 g 和 u 的假设下，定理 9.3.1 中定义的 M_s（$0 \leqslant s < t$）是有界鞅且 $u(0,x) \equiv 0$，故

$$M_t \equiv \lim_{s\uparrow t} M_s = \int_0^t g(B_s)\,\mathrm{d}s.$$

由于 M_s 一致可积，$u(t,x) = E_x M_0 = E_x M_t = v(t,x)$. □

同样，容易证明若 v 足够光滑则它是解.

定理 9.3.3. 若 $v \in C^{1,2}$，则它在 $(0,\infty)\times\mathbb{R}^d$ 中满足 (a).

证明. 考虑运行时空布朗运动 $(t-s, B_s)$ 来构造解，可得

$$v(t+h,x) = E_x\left(v(t,B_h) + \int_0^h g(B_s)\,\mathrm{d}s\right).$$

两边减去 $v(t,x)$ 并用 Taylor 级数

$$\begin{aligned} v(t+h,x) - v(t,x) = E_x\Bigg(&\sum_{i=1}^d \frac{\partial v}{\partial x_i}(t,x)(B_i(h)-x_i) \\ &+ \frac{1}{2}\sum_{1\leqslant i,j\leqslant d} \frac{\partial^2 v}{\partial x_i \partial x_j}(t,x)(B_i(h)-x_i)(B_j(h)-x_j) \\ &+ \int_0^h g(B_s)\,\mathrm{d}s\Bigg) + o(h). \end{aligned}$$

由于当 $i \neq j$ 时，$B_i(t)$，$B_i^2(t)-t$ 和 $B_i(t)B_j(t)$ 都是鞅

$$v(t+h,x) - v(t,x) = \frac{1}{2}\sum_{i=1}^d \frac{\partial^2 v}{\partial x_i^2}(t,x)h + E_x\int_0^h g(B_s)\,\mathrm{d}s + o(h).$$

除以 h 并令 $h \to 0$ 得 $v_t = \frac{1}{2}\Delta v + g(x)$，这就是所需结果. □

定理 9.3.4. v 满足 (b).

证明. 若 $|g| \leqslant M$，则当 $t \to 0$ 时，

$$|v(t,x)| \leqslant E_x\int_0^t |g(B_s)|\,\mathrm{d}s \leqslant Mt \to 0,$$

这就证明了所需结果. □

证明 v 为解的最后一步是验证 $v \in C^{1,2}$. 计算繁琐，故仅陈述结果.

Hölder 连续性. 此处及后文中，说函数是 **Hölder 连续的**是指 $|g(x)-g(y)| \leqslant C|x-y|^\alpha$ 对某个 $\alpha > 0$ 成立.

定理 9.3.5. 若 g 是 Hölder 连续的，则 $v \in C^{1,2}$.

注意到 $v(t,x) = \int G_t(x,y)g(y)\,\mathrm{d}y$，其中

$$G_t(x,y) = \int_0^t (2\pi s)^{-d/2} e^{-|y-x|^2/2s}\,\mathrm{d}s.$$

考察满足 $|y-x|^2 \leqslant s$ 的 y，可得下界

$$G_t(x,y) \geqslant C \int_{|y-x|^2}^{t} s^{-d/2}\,\mathrm{d}s = O(|y-x|^{2-d}).$$

由此可知一阶导数像 δ^{1-d} 那样发散，二阶像 δ^{-d}，在 0 附近不可积. 然而，若 g 是 Hölder 连续的，其光滑性可以挽救局面. 细节见 Friedman (1964) 的第 12-13 页.

第 4 节 Feynman-Kac 公式

本节考虑当给热方程右端添加 cu 时会发生什么，其中 $c(x)$ 有界连续. 即研究

(a) $u \in C^{1,2}$ 且在 $(0,\infty)\times\mathbb{R}^d$ 中 $u_t = \frac{1}{2}\Delta u + cu$.

(b) u 在 $\{0\}\times\mathbb{R}^d$ 的每点连续，且 $u(0,x) = f(x)$，其中 f 有界连续.

若 $c(x) \leqslant 0$，则此方程描述带冷却的热流. 即 $u(t,x)$ 给出 $\mathbb{R}^d$ 中点 x 在时刻 t 的温度，其中 x 处时刻 t 的热量以速率 $-c(x)$ 耗散.

第一步仍是构造局部鞅.

定理 9.4.1. 令 $c_s = \int_0^s c(B_r)\,\mathrm{d}r$. 若 u 满足 (a)，则

$$M_s = u(t-s, B_s)\exp(c_s)$$

是 $[0,t]$ 上的局部鞅.

证明. 为此，首先推广 Itô 公式的第三版

$$\begin{aligned} &u(t-s,B_s)\exp(c_s) - u(t,B_0) \\ &\quad= \int_0^s -u_t(t-r,B_r)\exp(c_r)\,\mathrm{d}r + \int_0^s \exp(c_r)\nabla u(t-r,B_r)\cdot \mathrm{d}B_r \\ &\qquad+ \int_0^s u(t-r,B_r)c(B_r)\exp(c_r)\,\mathrm{d}r + \frac{1}{2}\int_0^s \Delta u(t-r,B_r)\exp(c_r)\,\mathrm{d}r. \end{aligned} \tag{9.4.1}$$

(9.4.1) 的证明. 与 Itô 公式的前三个证明一样，将区间 $[0,s]$ 细分，$s_i^n = si/2^n$. 将差写为

$$\begin{aligned} &[u(t-s_{i+1}^n, B(s_{i+1}^n)) - u(t-s_i^n, B(s_i^n))]\exp(c(s_{i+1}^n)) \\ &\quad+ u(t-s_i^n, B(s_i^n))[\exp(c(s_{i+1}^n)) - \exp(c(s_i^n))]. \end{aligned}$$

第一行的极限给出第一、第二、第四项. 利用

$$\exp(c(s_{i+1}^n)) - \exp(c(s_i^n)) = \exp(c(s_i^n))\left(\exp\left(\int_{s_i^n}^{s_{i+1}^n} c(B_r)\,\mathrm{d}r\right) - 1\right)$$

可见第二行的极限给出第三项.

重排 (9.4.1) 的右端可得

$$\begin{aligned} &= \int_0^s \left(-u_t + cu + \frac{1}{2}\Delta u\right)(t-r,B_r)\exp(c_r)\,\mathrm{d}r \\ &\quad+ \int_0^s \exp(c_r)\nabla u(t-r,B_r)\cdot \mathrm{d}B_r, \end{aligned}$$

这就证明了所需结果，因为 $-u_t + cu + \frac{1}{2}\Delta u = 0$，故由定理 7.6.4 知第二项是局部鞅. □

下一步，仍是唯一性结果.

定理 9.4.2. 若存在解在 $[0,T]\times\mathbb{R}^d$ 上有界（对任意 $T<\infty$），则它必为

$$v(t,x) \equiv E_x\{f(B_t)\exp(c_t^t)\}. \tag{9.4.2}$$

证明. 在关于 c 和 u 的假设下，M_s（$0\leqslant s<t$）是有界鞅且 $M_t \equiv \lim_{s\to t} M_s = f(B_t)\exp(c_t^t)$. 由于 M_s 一致可积，故

$$u(t,x) = E_x M_0 = E_x M_t = v(t,x),$$

这就证明了所需结果. □

如前，容易证明若 v 足够光滑则它是解.

定理 9.4.3. 若 $v\in C^{1,2}$，则它满足 (a).

证明. 将 (9.4.2) 中的解写出时，考虑从时刻 0 到 t 运行时空布朗运动 $(t-s,B_s)$，可得

$$v(t+h,x) = E_x[v(t,B_h)\exp(c^h)].$$

两边减去 $v(t,x)$

$$v(t+h,x)-v(t,x) = E_x[v(t,B_h)\exp(c^h)-v(t,x)].$$

在右端加减 $v(t,B_h)$ 并用 Taylor 定理

$$\begin{aligned} v(t,B_h)\exp(c^h)-v(t,x) \approx &\sum_{i=1}^d \frac{\partial v}{\partial x_i}(t,x)(B_h^i-x) \\ &+\frac{1}{2}\sum_{1\leqslant i,j\leqslant d}\frac{\partial^2 v}{\partial x_i\partial x_j}(t,x)(B_h^i-x)(B_h^j-x)+v(t,B_h)(\exp(c^h)-1)+o(h). \end{aligned}$$

由于当 $i\neq j$ 时，$W_i(t)$，$W_i^2(t)=t$ 和 $W_i(t)W_j(t)$ 均为鞅，

$$E_x[v(t,B_h)\exp(c^h)-v(t,x)] \approx \sum_{i=1}^d \frac{\partial^2 v}{\partial x_i^2}h + E_x[v(t,B_h)(\exp(c^h)-1)]+o(h).$$

两边除以 h 并令 $h\to 0$ 得

$$\frac{\partial v}{\partial t} = \frac{1}{2}\Delta v(t,x)+c(x)v(t,x),$$

这就证明了所需结果. □

定理 9.4.4. v 满足 (b).

证明. 若 $|c| \leqslant M$，则 $e^{-Mt} \leqslant \exp(c_t^t) \leqslant e^{Mt}$，故当 $t \to 0$ 时 $\exp(c_t^t) \to 1$. 记 $\|f\|_\infty = \sup_x |f(x)|$，此结果蕴含

$$|E_x \exp(c_t^t) f(B_t) - E_x f(B_t)| \leqslant \|f\|_\infty E_x |\exp(c_t^t) - 1| \to 0.$$

定理 9.2.4 蕴含 $(t,x) \to E_x f(B_t)$ 在 $\{0\} \times \mathbb{R}^d$ 的每点连续，所需结果随之得出. □

这就引出何时 v 足够光滑以成为解的问题.

定理 9.4.5. 若 f 与 c 均 Hölder 连续，则 $v \in C^{1,2}$，故满足 (a).

为解此情形的问题，下面用一个技巧将结果约化到前一情形. 首先注意

$$c_s^t = \int_0^s c(B_r)\,\mathrm{d}r$$

连续且局部有界变差. 故若 $h \in C^1$，用普通变量替换公式有

$$h(c_t^t) - h(c_0^t) = \int_0^t h'(c_s^t) c(B_s)\,\mathrm{d}s.$$

取 $h(x) = e^{-x}$ 可得

$$\exp(-c_t^t) - 1 = -\int_0^t \exp(-c_s^t) c(B_s)\,\mathrm{d}s.$$

乘以 $-\exp(c_t^t)$ 得

$$\exp(c_t^t) - 1 = \int_0^t c(B_s) \exp(c_t^t - c_s^t)\,\mathrm{d}s.$$

代入 c_t^t 和 c_s^t 的定义有

$$\exp\left(\int_0^t c(B_r)\,\mathrm{d}r\right) = 1 + \int_0^t c(B_s) \exp\left(\int_s^t c(B_r)\,\mathrm{d}r\right)\,\mathrm{d}s.$$

乘以 $f(B_t)$，取期望，并用 Fubini 定理（由一切有界知合理）得

$$v(t,x) = E_x f(B_t) + \int_0^t E_x \left\{ c(B_s) \exp\left(\int_s^t c(B_r)\,\mathrm{d}r\right) f(B_t) \right\}\,\mathrm{d}s.$$

对 $\mathcal{F}_s$ 条件化，注意 $c(B_s) \in \mathcal{F}_s$，并用 Markov 性质

$$E_x \left(c(B_s) \exp\left(\int_s^t c(B_r)\,\mathrm{d}r\right) f(B_t) \,\middle|\, \mathcal{F}_s \right) = c(B_s) v(t-s, B_s).$$

取最后一式的期望并代入前一式，得

$$\begin{aligned} v(t,x) &= E_x f(B_t) + \int_0^t E_x\{c(B_s)v(t-s,B_s)\}\,\mathrm{d}s \\ &\equiv v_1(t,x) + v_2(t,x). \end{aligned} \tag{9.4.3}$$

右端第一项 $v_1(t,x)$ 由定理 9.2.3 知是 $C^{1,2}$ 的. 第二项 $v_2(t,x)$ 具有上节考虑的形式，其中 $g(r,x) = c(x)v(r,x)$. 若从 c 和 f 有界则 v 在 $[0,N]\times\mathbb{R}^d$ 上有界这一平凡观察出发，并应用 (2.6b)，可见

$$|v_2(t,x) - v_2(t,y)| \leqslant C_N|x-y| \quad \text{当} t \leqslant N.$$

为得到 v_1 的类似估计，令 $\bar{B}_t$ 为从 0 出发的布朗运动，并注意由于 f 是 Hölder 连续的

$$\begin{aligned} |v_1(t,x) - v_1(t,y)| &= |E\{f(x+\bar{B}_t) - f(y+\bar{B}_t)\}| \\ &\leqslant E|f(x+\bar{B}_t) - f(y+\bar{B}_t)| \leqslant C|x-y|^\alpha. \end{aligned}$$

结合最后两个估计可见 $v(r,x)$ 关于 t 局部 Hölder 连续.

由于 c 和 v 有界，且已假设 $c(\cdot,x)$ 是 Hölder 连续的，三角不等式蕴含 $g(r,x)$ 关于 t 局部 Hölder 连续. 由非齐次热方程的光滑性结果即得.

第 5 节 Dirichlet 问题

本节及 9.7 节内容遵循 Port 和 Stone (1978) 的论述.

本节考虑 **Dirichlet 问题**. 即，给定开集 G，寻找满足如下条件的函数 u:

(a) $u \in C^2$ 且在 G 中 $\Delta u = 0$.

(b) 在 ∂G 的每一点，u 连续且 $u = f$，其中 f 为有界函数.

物理上，u 表示当 ∂G 保持固定温度分布 f 时的**平衡温度分布**.

记 $\tau = \inf\{t > 0 : B_t \notin G\}$. 本节始终假设对所有 $x \in G$ 有

$$P_x(\tau < \infty) = 1 \tag{9.5.1}$$

若 u 满足 (a)，则 $M_t = u(B_t)$ 是 $[0, \tau)$ 上的局部鞅. 也就是说，存在递增的停时序列 T_n 使得 $u(B(t \wedge T_n))$ 是鞅. 为将局部鞅转化为真正的鞅，引入函数 $\gamma : [0, \infty) \to [0, \tau)$，当取值达到 T_n 时，它在单位时间内保持常数. 即，$\gamma(t) = t$ 对 $0 \leqslant t \leqslant T_1$，$\gamma(t) = T_1$ 对 $T_1 \leqslant t \leqslant T_1 + 1$，且一般地

$$\begin{aligned} \gamma(t) &= t - k \quad \text{对} T_k + k \leqslant t \leqslant T_{k+1} + k \\ \gamma(t) &= T_{k+1} \quad \text{对} T_{k+1} + k \leqslant t \leqslant T_{k+1} + k + 1 \end{aligned}$$

由于 $\gamma(t)$ 是停时且 $\gamma(n) \leqslant T_n$.

引理 9.5.1. $M_t = u(B(\gamma(t)))$ 是鞅.

定理 9.5.2. 若 Dirichlet 问题存在有界解，则它必为

$$v(x) \equiv E_x f(B_\tau)$$

证明. 由于 u 有界，M_t 是有界鞅，因此 $\lim_{t\to\infty} M_t$ 存在. 边界条件蕴含 $M_\infty = f(B_\tau)$. 由于该鞅一致可积，故 $u(x) = E_x M_\infty = E_x f(B_\tau)$. □

定理 9.5.3. 若 $v \in C^2$，则它满足 (a).

证明. 此论证由 Liggett (2010) 给出，它避免了随机微积分的使用. 设 $x \in D$ 且 $B = B(x, r) \subset D$. 若 τ_B 是从 B 的退出时，则由强 Markov 性质

$$v(x) = E_x v(B(\tau_B)).$$

利用 Taylor 定理

$$E_x v(B(\tau_B)) - v(x) = E_x \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}(x)[B_i(\tau_B) - x_i] + \frac{1}{2} E_x \sum_{1 \leqslant i,j \leqslant d} \frac{\partial^2 v}{\partial x_i \partial x_j}(x)[B_i(\tau_B) - x_i][B_j(\tau_B) - x_j] + o(r^2)$$

由于对 $i \neq j$，$B_i(t)$、$B_i^2(t) - t$ 和 $B_i(t)B_j(t)$ 都是鞅，有

$$\begin{aligned} E_x[B_i(\tau_B) - x_i] &= 0 \\ E_x[B_i(\tau_B) - x_i][B_j(\tau_B) - x_j] &= 0 \quad \text{当} i \neq j \\ E_x[B_i(\tau_B) - x_i]^2 &= E_0[B_i^2(\tau_B)] = E_x \tau_B \end{aligned}$$

将这些结果用于前述公式

$$0 = E_x v(B(\tau_B)) - v(x) = \frac{1}{2} \Delta v(x) E_x \tau_B + o(r^2)$$

这就证明了所需结果. □

定理 9.5.4. $v \in C^\infty$.

证明. 设 $x \in D$ 且 $B = B(x, r) \subset D$. 若 τ_B 是从 B 的退出时，则由强 Markov 性质

$$v(x) = E_x v(B(\tau_B)) = \int_{\partial D(x,r)} \bar{v}(y)\, \pi(\mathrm{d}y).$$

其中 π 是边界 $\partial D(x, r)$ 上的表面测度，已归一化为概率测度.

上述结果是调和函数的**均值性质**.

下面给出 $v \in C^\infty$ 的简单分析论证. 为此，令 ψ 为非负无穷可微函数，它在 $[\delta^2, \infty)$ 上为零但不恒为零. 容易证明

$$h(x) = \int_{D(x,\delta)} \psi(|y - x|^2) v(y)\, \mathrm{d}y \in C^\infty$$

变换到球坐标并利用均值性质，容易看出 $h(x)$ 是 $v(x)$ 的常数倍，故 $v \in C^\infty$. □

如下例所示，v 并不总是满足 (b).

例题 9.5.5. 穿孔圆盘. 设 $d = 3$ 且 $G = D - K$, 其中 $D = \{x : |x| < 1\}$, $K = \{x : x_1 = x_2 = 0, x_3 \geqslant 0\}$. 唯一性结果以及三维 Brownian 运动从不击中直线的事实蕴含解必为 $v(x) = E_x f(B_{\tau(D)})$, 其中 $\tau(D)$ 是从 D 的退出时. 若令 f 在 D 的边界上为 0，在 K 上为 $1 - x_3$，则解 $v \equiv 0$ 于 G 中，边界条件不被满足.

例题 9.5.6. Lebesgue 刺. 设 $d \geqslant 3$ 且

$$G = (-1,1)^d - \bigcup_{1 \leqslant n \leqslant \infty} \{[2^{-n}, 2^{-n+1}] \times [-a_n, a_n]^{d-1}\}$$

其中 $n = \infty$ 项为单点集 $\{0\}$. 下面证明若 $a_n \downarrow 0$ 足够快，则 $P_0(\tau = 0) = 0$. 由此可得，若定义 f 在 $(-1,1)^d$ 的边界上为 0，在其余边界上为 $1 - x_1$，则 $E_0 f(B_\tau) < 1$，边界条件不被满足.

证明. $P_0((B_t^2, B_t^3) = (0,0) \text{ 对某} t > 0) = 0$，因此以概率 1，起始于 0 的 Brownian 运动 B_t 不会击中

$$I_n = \{x : x_1 \in [2^{-n}, 2^{-n+1}], x_2 = x_3 = \cdots = x_d = 0\}$$

由于 B_t 在 $d \geqslant 3$ 时是暂态的，故对几乎必然 ω，$\{B_s : 0 \leqslant s < \infty\}$ 与 I_n 的距离为正. 由最后这一观察，若令 $T_n = \inf\{t : B_t \in [2^{-n}, 2^{-n+1}] \times [a_n, a_n]^{d-1}\}$ 并取 a_n 足够小，则 $P_0(T_n < \infty) \leqslant 3^{-n}$. 现在 $\sum_{n=1}^{\infty} 3^{-n} = 3^{-1}(3/2) = 1/2$，因此若令 $\tau = \inf\{t > 0 : B_t \notin G\}$ 且 $\sigma = \inf\{t > 0 : B_t \notin (-1,1)^d\}$，则有

$$P_0(\tau < \sigma) \leqslant \sum_{n=1}^{\infty} P_0(T_n < \infty) \leqslant \frac{1}{2}$$

于是 $P_0(\tau > 0) \geqslant P_0(\tau = \sigma) \geqslant 1/2$. 由于 $\tau = 0$ 关于 $\mathcal{F}_0^+$ 可测，有 $P_0(\tau = 0) = 1$.（译者注：原文此处似有误，应为 $P_0(\tau = 0) = 0$） □

称点 $y \in \partial G$ 为**正则点**，若 $P_y(\tau = 0) = 1$.

定理 9.5.7. 设 G 为任意开集. 假设 f 有界连续且 y 是 ∂G 的正则点. 若 $x_n \in G$ 且 $x_n \to y$，则 $v(x_n) \to f(y)$.

第一步是证明

引理 9.5.8. 若 $t > 0$，则 $x \mapsto P_x(\tau \leqslant t)$ 是下半连续的. 即，若 $x_n \to x$，则

$$\liminf_{n \to \infty} P_{x_n}(\tau \leqslant t) \geqslant P_x(\tau \leqslant t)$$

证明. 由 Markov 性质

$$P_x(B_s \in G^c \text{ 对某} s \in (\epsilon, t]) = \int p_\epsilon(x,y) P_y(\tau \leqslant t - \epsilon)\, \mathrm{d}y$$

由于 $y \mapsto P_y(\tau \leqslant t - \epsilon)$ 有界可测且

$$p_\epsilon(x,y) = (2\pi\epsilon)^{-d/2} e^{-|x-y|^2/2\epsilon}$$

由控制收敛定理，对每个 $\epsilon > 0$

$$x \mapsto P_x(B_s \in G^c \text{ 对某} s \in (\epsilon, t])$$

是连续的. 令 $\epsilon \downarrow 0$ 表明 $x \mapsto P_x(\tau \leqslant t)$ 是连续函数的递增极限，因此是下半连续的. □

若 y 对 G 正则且 $t > 0$，则 $P_y(\tau \leqslant t) = 1$，因此由引理 9.5.8，若 $x_n \to y$，则对所有 $t > 0$

$$\liminf_{n\to\infty} P_{x_n}(\tau \leqslant t) \geqslant 1$$

有了这一结论，容易完成证明. 由于 f 有界连续，只需证明若 $D(y,\delta) = \{x : |x - y| < \delta\}$，则

引理 9.5.9. 若 y 对 G 正则且 $x_n \to y$，则对所有 $\delta > 0$

$$P_{x_n}(\tau < \infty, B_\tau \in D(y,\delta)) \to 1$$

证明. 设 $\epsilon > 0$ 并取 t 足够小使得

$$P_0\left(\sup_{0\leqslant s\leqslant t} |B_s| > \frac{\delta}{2}\right) < \epsilon$$

由于 $P_{x_n}(\tau \leqslant t) \to 1$（当 $x_n \to y$），由上述选取知

$$\begin{aligned}
\liminf_{n\to\infty} P_{x_n}(\tau < \infty, B_\tau \in D(y,\delta)) &\geqslant \liminf_{n\to\infty} P_{x_n}\left(\tau \leqslant t, \sup_{0\leqslant s\leqslant t} |B_s - x_n| \leqslant \frac{\delta}{2}\right) \\
&\geqslant \liminf_{n\to\infty} P_{x_n}(\tau \leqslant t) - P_0\left(\sup_{0\leqslant s\leqslant t} |B_t| > \frac{\delta}{2}\right) \\
&> 1 - \epsilon
\end{aligned}$$

由于 ϵ 任意，这证明了引理，从而证明了定理 9.5.7. □

下面结果表明，若 G^c 在 y 附近不太小，则 y 是正则点.

定理 9.5.10. 锥条件. 若存在以 y 为顶点的锥 V 及 $r > 0$ 使得 $V \cap D(y,r) \subset G^c$，则 y 是正则点.

证明. 首先要定义以 y 为顶点、指向方向 v、开口为 a 的锥：

$$V(y,v,a) = \{x : x = y + \theta(v + z) \text{ 其中} \theta \in (0,\infty), z \perp v, \text{且} |z| < a\}$$

定义了锥之后，其余就容易了. 由于正态分布是球对称的，

$$P_y(B_t \in V(y,v,a)) = \epsilon_a > 0$$

其中 ϵ_a 是仅依赖于开口 a 的常数. 设 $r>0$ 使得 $V(y,v,a)\cap D(y,r)\subset G^c$. Brownian 路径的连续性蕴含

$$\lim_{t\to 0} P_y\left(\sup_{s\leqslant t}|B_s-y|>r\right)=0$$

结合最后两个结果与一简单不等式有

$$\epsilon_a\leqslant \liminf_{t\downarrow 0} P_y(B_t\in G^c)\leqslant \lim_{t\downarrow 0} P_y(\tau\leqslant t)\leqslant P_y(\tau=0)$$

由 Blumenthal 0-1 律知 $P_y(\tau=0)=1$. □

9.5.1 退出分布

半空间. 令 $\tau=\inf\{t:B_t\notin H\}$, 其中 $H=\{(x,y):x\in\mathbb{R}^{d-1},y>0\}$. 对 $\theta\in\mathbb{R}^{d-1}$ 令

$$h_\theta(x,y)=\frac{C_d y}{(|x-\theta|^2+y^2)^{d/2}} \tag{9.5.2}$$

其中 C_d 的选取使得 $\int \mathrm{d}\theta\, h_\theta(0,1)=1$.

引理 9.5.11. 在 H 中 $\Delta h_\theta=0$.

证明. 略去 C_d 并求导得

$$\begin{aligned}
D_{x_i}h_\theta&=-\frac{d}{2}\cdot\frac{2(x_i-\theta_i)y}{(|x-\theta|^2+y^2)^{(d+2)/2}}\\
D_{x_ix_i}h_\theta&=-\frac{dy}{(|x-\theta|^2+y^2)^{(d+2)/2}}+\frac{d(d+2)(x_i-\theta)^2y}{(|x-\theta|^2+y^2)^{(d+4)/2}}\\
D_yh_\theta&=\frac{1}{(|x-\theta|^2+y^2)^{d/2}}-d\cdot\frac{y^2}{(|x-\theta|^2+y^2)^{(d+2)/2}}\\
D_{yy}h_\theta&=-d\cdot\frac{y+2y}{(|x-\theta|^2+y^2)^{(d+2)/2}}+\frac{d(d+2)y^3}{(|x-\theta|^2+y^2)^{(d+4)/2}}
\end{aligned}$$

相加可见

$$\begin{aligned}
\sum_{i=1}^{d-1}D_{x_ix_i}h_\theta+D_{yy}h_\theta&=\frac{\{(-d)(d-1)+(-d)\cdot 3\}y}{(|x-\theta|^2+y^2)^{(d+2)/2}}\\
&\quad+\frac{d(d+2)y(|x-\theta|^2+y^2)}{(|x-\theta|^2+y^2)^{(d+4)/2}}=0
\end{aligned}$$

这就完成了证明. □

定理 9.5.12. 设 f 有界连续且

$$u(x,y)=\int \mathrm{d}\theta\, h_\theta(x,y)f(\theta,0)$$

u 满足 H 中以 f 为边界值的 Dirichlet 问题. 由此可得 $h_\theta(x,y)$ 给出概率密度 $P_{x,y}(B(\tau_H)=(\theta,0))$.

证明. 交换导数与积分可见 $\Delta u=0$. 下一步证明它满足边界条件.

显然 $\int \mathrm{d}\theta\, h_\theta(x,y)$ 与 x 无关. 为证明它与 y 无关, 令 $x=0$ 并作变量替换 $\theta_i=y\phi_i\,(1\leqslant i\leqslant d-1)$ 得

$$\int \mathrm{d}\theta\, h_\theta(0,y)=\int \frac{C_d y}{(y^2|\phi|^2+y^2)^{d/2}}\cdot y^{d-1}\mathrm{d}\phi=\int \mathrm{d}\phi\, h_\phi(0,1)=1$$

作变量替换 $\theta_i=x_i+r_i y$ 并利用控制收敛

$$\int_{D(x,\epsilon)^c}\mathrm{d}\theta\, h_\theta(x,y)=\int_{D(0,\epsilon/y)^c}\mathrm{d}r\, h_r(0,1)\to 0$$

由于 $P_{x,y}(\tau<\infty)=1$ 对所有 $x\in H$，最后结论由定理 9.5.2 推出. □

当 $d=2$ 时

$$h_\theta(0,y)=\frac{y}{\pi(|\theta|^2+y^2)}$$

是 Cauchy 密度. 为理解为何如此, 从 $(0,0)$ 启动 Brownian 运动并令 $T_y=\inf\{t:B_t^2=y\}$. 强 Markov 性质蕴含 $B_1(T_y)$ 有独立增量. Brownian 尺度变换蕴含 $B_1(T_y)\overset{d}{=}yB_1(T_1)$. 由于 $B_1(T_y)\overset{d}{=}-B_1(T_y)$, 分布必为 $\alpha=1$ 的对称稳定律.

球的退出分布. 令 $D=\{x:|x|<1\}$，$\tau_D=\inf\{t:B_t\notin D\}$，并定义 **Poisson 核**

$$k_y(x)=\frac{1-|x|^2}{|x-y|^d} \tag{9.5.3}$$

定理 9.5.13. 若 f 有界连续，则

$$E_x f(B_\tau)=u(x)=\int_{\partial D}k_y(x)f(y)\,\pi(\mathrm{d}y) \tag{9.5.4}$$

其中 π 是 ∂D 上的表面测度，已归一化为概率测度.

证明技术与半空间中的结果相同. 只需证明 u 满足以 f 为边界条件的 Dirichlet 问题. 固定 y 证明 $\Delta k_y(x)=0$ 留给读者，然后如上完成证明.

习题

习题 9.5.1. 设 h 在 $\mathbb{R}^d$ 上有界调和. 证明 h 是常数.

习题 9.5.2. 设 $f \in C^2$ 是上调和的，取值于 $[0,\infty]$ 且连续. 证明在 1 维和 2 维，f 必为常数. 在 $d \geqslant 3$ 中给出一个形如 $f(x) = g(|x|)$ 的例子.

习题 9.5.3. 设 B_t 是 d 维 Brownian 运动且 $\tau = \inf\{t : B_t \notin G\}$. 设 G 是连通开集，所有边界点都是正则的，且在 G 中 $P_x(\tau < \infty) < 1$. 证明 Dirichlet 问题以 0 为边界值的所有有界解都有形式 $u(x) = cP_x(\tau = \infty)$. 提示：在 G 中定义 $\bar{u} = u + \|u\|_\infty P_x(\tau = \infty)$，在 G^c 上 $\bar{u} = 0$，并证明 $u(B_y)$ 是下鞅.

习题 9.5.4. 令 $\tau = \inf\{t : B_t \notin G\}$，考虑 $d \geqslant 3$ 中连通开集 G 的 Dirichlet 问题，其中所有边界点都是正则的且在 G 中 $P_x(\tau < \infty) < 1$. 假设边界条件 $f \equiv 0$. 推广前一结果，证明所有有界解都有形式 $u(x) = cP_x(\tau = \infty)$.

习题 9.5.5. 定义平锥 $\bar{V}(y,v,a)$ 为 $V(y,v,a)$ 与包含直线 $\{y + \theta v : \theta \in \mathbb{R}\}$ 的 $d-1$ 维超平面的交. 证明定理 9.5.10 在" 锥" 替换为" 平锥" 时仍然成立.

第 6 节 Green 函数与位势核

令 $D = B(0,r) = \{y : |y| < r\}$ 为以 0 为中心、半径为 r 的球. 在 $d \leqslant 2$ 时 B_t 将无穷多次返回 D，但在 $d \geqslant 3$ 时不会. 本节研究占据时间 $\int_0^\infty 1_D(B_t)\,\mathrm{d}t$. 第一步是证明对任意 x

$$P_x\left(\int_0^\infty 1_D(B_t)\,\mathrm{d}t = \infty\right) = 1 \qquad \text{在} d \leqslant 2, \tag{9.6.1}$$

$$E_x \int_0^\infty 1_D(B_t)\,\mathrm{d}t < \infty \qquad \text{在} d \geqslant 3. \tag{9.6.2}$$

证明. 令 $T_0 = 0$ 且 $G = B(0,2r)$. 对 $k \geqslant 1$，令

$$S_k = \inf\{t > T_{k-1} : B_t \in D\}$$
$$T_k = \inf\{t > S_k : B_t \in G\}$$

记 τ 为 T_1，利用强 Markov 性质，对 $k \geqslant 1$ 有

$$P_x\left(\int_{S_k}^{T_k} 1_D(B_t)\,\mathrm{d}t \geqslant s \,\middle|\, \mathcal{F}_{S_k}\right) = P_{B(S_k)}\left(\int_0^\tau 1_D(B_t)\,\mathrm{d}t \geqslant s\right) = H(s)$$

由强 Markov 性质，$\int_{S_k}^{T_k} 1_D(B_t)\,\mathrm{d}t$ 是独立同分布的.

由于这些随机变量有正均值，由强大数定律

$$\int_0^\infty 1_D(B_t)\,\mathrm{d}t \geqslant \lim_{n\to\infty} \sum_{k=1}^{n} \int_{S_k}^{T_k} 1_D(B_t)\,\mathrm{d}t = \infty \quad \text{a.s.}$$

证明了 (9.6.1). □

若 f 是非负函数，则 Fubini 定理蕴含

$$E_x \int_0^\infty f(B_t)\,\mathrm{d}t = \int_0^\infty E_x f(B_t)\,\mathrm{d}t = \int_0^\infty \int p_t(x,y) f(y)\,\mathrm{d}y\,\mathrm{d}t = \int \int_0^\infty p_t(x,y)\,\mathrm{d}t\, f(y)\,\mathrm{d}y$$

其中 $p_t(x,y) = (2\pi t)^{-d/2} e^{-|x-y|^2/2t}$ 是 Brownian 运动的转移密度. 当 $t \to \infty$ 时，$p_t(x,y) \sim (2\pi t)^{-d/2}$，

因此若 $d \leqslant 2$ 则 $\int p_t(x,y)\mathrm{d}t = \infty$. 当 $d \geqslant 3$ 时，作变量替换 $t = |x-y|^2/2s$ 得

$$\begin{aligned}\int_0^\infty p_t(x,y)\,\mathrm{d}t &= \int_0^\infty \frac{1}{(2\pi t)^{d/2}} e^{-|y-x|^2/2t}\,\mathrm{d}t \\ &= \int_\infty^0 \left(\frac{s}{\pi|x-y|^2}\right)^{d/2} e^{-s}\left(-\frac{|x-y|^2}{2s^2}\right)\mathrm{d}s \\ &= \frac{|x-y|^{2-d}}{2\pi^{d/2}}\int_0^\infty s^{(d/2)-2}e^{-s}\,\mathrm{d}s \\ &= \frac{\Gamma(d/2-1)}{2\pi^{d/2}}|x-y|^{2-d} \end{aligned} \tag{9.6.3}$$

其中 $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}\,\mathrm{d}s$ 是通常的伽马函数. 若定义

$$G(x,y) = \int_0^\infty p_t(x,y)\,\mathrm{d}t$$

则在 $d \geqslant 3$ 时，$G(x,y) < \infty$（当 $x \neq y$），且

$$E_x \int_0^\infty f(B_t)\,\mathrm{d}t = \int G(x,y)f(y)\,\mathrm{d}y$$

为完成证明，注意取 $f = 1_D$（$D = B(0,r)$）并变换到极坐标

$$\int_D G(0,y)\,\mathrm{d}y = \int_0^r s^{d-1}C_d s^{2-d}\,\mathrm{d}s = \frac{C_d}{2}r^2 < \infty$$

为将最后结论推广到 $x \neq 0$，注意对起始于 0 的 Brownian 运动在从 $B(0,|x|)$ 的退出时 τ 应用强 Markov 性质并利用旋转对称性，有

$$E_x\int_0^\infty 1_D(B_s)\,\mathrm{d}s = E_0\int_\tau^\infty 1_D(B_s)\,\mathrm{d}s \leqslant E_0\int_0^\infty 1_D(B_s)\,\mathrm{d}s$$

这就完成了证明.

称 $G(x,y)$ 为**位势核**，因为 $G(\cdot,y)$ 是 y 处单位电荷的**静电位势**. 详见 Port 和 Stone (1978) 的第三章. 在 $d \leqslant 2$ 时，$\int_0^\infty p_t(x,y)\,\mathrm{d}t \equiv \infty$，因此需要采用另一种方法定义一个有用的 G:

$$G(x,y) = \int_0^\infty (p_t(x,y) - a_t)\,\mathrm{d}t \tag{9.6.4}$$

其中 a_t 的选取使积分收敛（至少当 $x \neq y$ 时）. 为理解这一定义为何有用，注意若 $\int f(y)\,\mathrm{d}y = 0$ 则

$$\int G(x,y)f(y)\,\mathrm{d}y = \lim_{t\to\infty}\int_0^t E_x f(B_s)\,\mathrm{d}s$$

当 $d=1$ 时令 $a_t=p_t(0,0)$. 如此选取，

$$G(x,y)=\frac{1}{\sqrt{2\pi}}\int_0^\infty (e^{-(y-x)^2/2t}-1)t^{-1/2}\,\mathrm{d}t$$

且积分收敛，因为被积函数 $\leqslant 0$ 且当 $t\to\infty$ 时 $\sim -(y-x)^2/2t^{3/2}$. 作变量替换 $t=(y-x)^2/2u$ 得

$$\begin{aligned}
G(x,y)&=\frac{1}{\sqrt{2\pi}}\int_\infty^0 (e^{-u}-1)\left(\frac{2u}{(y-x)^2}\right)^{1/2}\frac{-(y-x)^2}{2u^2}\,\mathrm{d}u\\
&=-\frac{|y-x|}{2\sqrt{\pi}}\int_0^\infty\left(\int_0^u e^{-s}\,\mathrm{d}s\right)u^{-3/2}\,\mathrm{d}u\\
&=-\frac{|y-x|}{\sqrt{\pi}}\int_0^\infty \mathrm{d}s\,e^{-s}\int_s^\infty\frac{u^{-3/2}}{2}\,\mathrm{d}u\\
&=-\frac{|y-x|}{\sqrt{\pi}}\int_0^\infty \mathrm{d}s\,e^{-s}s^{-1/2}=-|y-x|
\end{aligned}\tag{9.6.5}$$

因为

$$\int_0^\infty \mathrm{d}s\,e^{-s}s^{-1/2}=\int_0^\infty \mathrm{d}r\,e^{-r^2/2}\sqrt{2}=\frac{1}{2}\sqrt{2}\cdot\sqrt{2\pi}=\sqrt{\pi}$$

$d=2$ 的计算几乎相同. 唯一改变的是 a_t 的选取. 若再试 $a_t=p_t(0,0)$，则对 $x\neq y$ 被积函数当 $t\to 0$ 时 $\sim -t^{-1}$，积分发散，因此令 $a_t=p_t(0,e_1)$，其中 $e_1=(1,0)$. 如此选取 a_t，得

$$\begin{aligned}
G(x,y)&=\frac{1}{2\pi}\int_0^\infty (e^{-|x-y|^2/2t}-e^{-1/2t})t^{-1}\,\mathrm{d}t\\
&=\frac{1}{2\pi}\int_0^\infty\left(\int_{|x-y|^2/2t}^{1/2t} e^{-s}\,\mathrm{d}s\right)t^{-1}\,\mathrm{d}t\\
&=\frac{1}{2\pi}\int_0^\infty \mathrm{d}s\,e^{-s}\int_{|x-y|^2/2s}^{1/2s} t^{-1}\,\mathrm{d}t\\
&=\frac{1}{2\pi}\left(\int_0^\infty \mathrm{d}s\,e^{-s}\right)(-\log(|x-y|^2))\\
&=-\frac{1}{\pi}\log(|x-y|)
\end{aligned}\tag{9.6.6}$$

总之，位势核由下式给出

$$G(x,y)=\begin{cases}\dfrac{\Gamma(d/2-1)}{2\pi^{d/2}}\cdot|x-y|^{2-d} & d\geqslant 3\\[2ex] -\dfrac{1}{\pi}\cdot\log(|x-y|) & d=2\\[2ex] -1\cdot|x-y| & d=1\end{cases}\tag{9.6.7}$$

读者应注意，在每种情况下 $G(x,y)=C\phi(|x-y|)$，其中 ϕ 是 9.1 节中使用的调和函数. 当然，这并非偶然. $x\mapsto G(x,0)$ 显然是球对称的，且如将看到的，对 $x\neq 0$ 满足 $\Delta G(x,0)=0$，因此上述结果蕴含 $G(x,0)=A+B\phi(|x|)$.

上述公式对应于 $A=0$，这简洁优美. 但那些看起来奇怪的 B 呢？它们有什么特别之处？答案很简单：它们的选取使得 $\frac{1}{2}\Delta G(x,0)=-\delta_0$（0 处的点质量）在分布意义下成立. 在 $d=1$ 时容易看出这一点. 此时 $\phi(x)=|x|$，则

$$\phi'(x)=\begin{cases}1 & x>0\\ -1 & x<0\end{cases}$$

因此若 $B=-1$ 则 $B\phi''(x)=-2\delta_0$. 更高深的读者可以验证在 $d\geqslant 2$ 时这也是成立的.（见 F. John (1982)，第 96-97 页.）

第 7 节 Poisson 方程

本节研究有界开集 G 中的如下方程：

(a) $u \in C^2$ 且在 G 中 $\frac{1}{2}\Delta u = -g$，其中 g 有界连续.

(b) 在 ∂G 的每一点，u 连续且 $u = 0$.

可以加上 Dirichlet 问题的解以将 (b) 中的 $u = 0$ 替换为 $u = f$. 由定理 9.1.1 可知

定理 9.7.1. 令 $\tau = \inf\{t > 0 : B_t \notin G\}$. 若 u 满足 (a)，则

$$M_t = u(B_t) + \int_0^t g(B_s)\,\mathrm{d}s$$

是 $[0, \tau)$ 上的局部鞅.

下一步证明唯一性结果.

定理 9.7.2. 若 (a) 存在有界解，则它必为

$$v(x) \equiv E_x\left(\int_0^\tau g(B_t)\,\mathrm{d}t\right)$$

注意若 $g \equiv 1$，解为 $v(x) = E_x\tau$.

证明. 若 u 满足 (a), 则如上定义的 M_t 是 $[0, \tau)$ 上的局部鞅. 若 G 有界, 则对所有 $x \in G$ 有 $E_x\tau < \infty$. 若 u 和 g 有界，则对 $t < \tau$

$$|M_t| \leqslant \|u\|_\infty + \tau\|g\|_\infty$$

由于 M_t 被可积随机变量控制，故

$$M_\tau \equiv \lim_{t\uparrow\tau} M_t = \int_0^\tau g(B_t)\,\mathrm{d}t$$

且 $u(x) = E_x(M_\tau)$. □

如同 Dirichlet 问题的处理：

定理 9.7.3. 若 $v \in C^2$，则它满足 (a).

证明. 设 $x \in D$ 且 $B(x, r) \subset D$. 强 Markov 性质蕴含

$$v(x) = E_x\int_0^{\tau_B} g(B_s)\,\mathrm{d}s + E_x v(B(\tau_B))$$

利用 Taylor 定理

$$E_x v(B(\tau_B)) - v(x) = E_x \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}(x)[B_i(\tau_B) - x_i] + \frac{1}{2} E_x \sum_{1 \leqslant i,j \leqslant d} \frac{\partial^2 v}{\partial x_i \partial x_j}(x)[B_i(\tau_B) - x_i][B_j(\tau_B) - x_j] + o(r^2)$$

由于对 $i \neq j$，$B_i(t)$、$B_i^2(t) - t$ 和 $B_i(t)B_j(t)$ 都是鞅，有

$$E_x v(B(\tau_B)) - v(x) = \frac{1}{2} \Delta v(x) E_x \tau_B + o(r^2)$$

另一方面 $E_x \int_0^{\tau_B} g(B_s)\,\mathrm{d}s = [g(x) + o(1)]E_x \tau_B$，令 $r \to 0$ 得 $\frac{1}{2}\Delta v = -g$. □

经过上节对 Dirichlet 问题的详细讨论，保证边界条件成立所需的条件应不足为奇.

定理 9.7.4. 假设 G 和 g 有界. 设 y 是 ∂G 的正则点. 若 $x_n \in G$ 且 $x_n \to y$，则 $v(x_n) \to 0$.

证明. 首先注意：

(i) 由引理 9.5.9，若 $\epsilon > 0$，则 $P_{x_n}(\tau > \epsilon) \to 0$.

(ii) 若 G 有界，则 $C = \sup_x E_x \tau < \infty$，因此 $\|v\|_\infty \leqslant C\|g\|_\infty < \infty$.

设 $\epsilon > 0$. 从一些基本不等式开始，然后利用 Markov 性质有

$$\begin{aligned} |v(x_n)| &\leqslant E_{x_n}\left(\int_0^{\tau \wedge \epsilon} |g(B_s)|\,\mathrm{d}s\right) + E_{x_n}\left(\left|\int_\epsilon^\tau g(B_s)\,\mathrm{d}s\right|; \tau > \epsilon\right) \\ &\leqslant \epsilon\|g\|_\infty + E_{x_n}(|v(B_\epsilon)|; \tau > \epsilon) \leqslant \epsilon\|g\|_\infty + \|v\|_\infty P_{x_n}(\tau > \epsilon) \end{aligned}$$

令 $n \to \infty$，利用 (i) 和 (ii) 即得所需结果，因为 ϵ 任意. □

最后，证明 $v \in C^2$ 的问题. 为此，假设 g 定义在 $\mathbb{R}^d$ 上而不仅在 G 中，且有紧支集. 回忆假设 G 有界，注意 g 在 G^c 上的值无关紧要，因此若以后想假设 $\int g(x)\,\mathrm{d}x = 0$ 也不失一般性. 先从 $d \geqslant 3$ 的情形开始，因为此时

$$\bar{w}(x) = E_x \int_0^\infty |g(B_t)|\,\mathrm{d}t < \infty$$

且是 x 的有界函数. 这意味着可以定义

$$w(x) = E_x \int_0^\infty g(B_t)\,\mathrm{d}t,$$

利用强 Markov 性质得

$$w(x) = E_x \int_0^\tau g(B_t)\,\mathrm{d}t + E_x w(B_\tau),$$

改变记号得

$$v(x) = w(x) - E_x w(B_\tau). \tag{9.7.1}$$

定理 9.5.4 表明第二项在 G 中是 C^∞ 的，因此为验证 $v \in C^2$ 只需证明 w 是，这在 $d \geqslant 3$ 时由于 (9.6.3) 给出如下显式公式而变得更简单：

$$w(x) = C_d \int |x-y|^{2-d} g(y)\,\mathrm{d}y$$

细节繁琐，下面只陈述结果.

定理 9.7.5. 在 $d \geqslant 2$ 时若 g 是 Hölder 连续的，则 w 是 C^2 的. 在 $d = 1$ 时假设 g 连续就足够了.

最后这个结果解决了 $d \geqslant 3$ 时光滑性的问题. 为将结果推广到 $d \leqslant 2$，使用 (9.6.4) 中定义的位势核. 利用 w 的这一解释，容易看出 (9.7.1) 成立，因此问题再次归结为证明 w 是 C^2 的，这是一个微积分问题.

所有计算完成后，发现定理 9.7.5 在 $d = 2$ 时成立，且在 $d = 1$ 时假设 g 连续就足够了. 读者可在 Port 和 Stone (1978) 第 116-117 页或 Gilbarg 和 Trudinger (1977) 第 53-55 页找到证明细节.

9.7.1 占据时间

半空间

为解释为何要定义位势核，回到半空间 $H = \{y : y_d > 0\}$ 中的 Brownian 运动. 令 $\tau = \inf\{t : B_t \notin H\}$，且令

$$\bar{y} = (y_1, \ldots, y_{d-1}, -y_d)$$

为 y 关于平面 $\{y \in \mathbb{R}^d : y_d = 0\}$ 的反射.

定理 9.7.6. 若 $x \in H$，$f \geqslant 0$ 有紧支集，且 $\{x : f(x) > 0\} \subset H$，则

$$E_x\left(\int_0^\tau f(B_t)\,\mathrm{d}t\right) = \int G(x,y) f(y)\,\mathrm{d}y - \int G(x,\bar{y}) f(y)\,\mathrm{d}y$$

证明. 利用 Fubini 定理（由 $f \geqslant 0$ 保证合理），然后利用反射原理以及 $\{x : f(x) > 0\} \subset H$ 有

$$\begin{aligned} E_x \int_0^\tau f(B_t)\,\mathrm{d}t &= \int_0^\infty E_x(f(B_t); \tau > t)\,\mathrm{d}t \\ &= \int_0^\infty \int_H (p_t(x,y) - p_t(x,\bar{y}))f(y)\,\mathrm{d}y\,\mathrm{d}t \\ &= \int_0^\infty \int_H (p_t(x,y) - a_t)f(y)\,\mathrm{d}y\,\mathrm{d}t - \int_0^\infty \int_H (p_t(x,\bar{y}) - a_t)f(y)\,\mathrm{d}y\,\mathrm{d}t \\ &= \int G(x,y)f(y)\,\mathrm{d}y - \int G(x,\bar{y})f(y)\,\mathrm{d}y \end{aligned}$$

f 的紧支集与 G 的公式蕴含 $\int |G(x,y)f(y)|\,\mathrm{d}y$ 和 $\int |G(x,\bar{y})f(y)|\,\mathrm{d}y$ 有限，因此最后两个等式成立. □

$d \geqslant 3$ 时上述证明大大简化；然而，上述证明的部分意义在于，利用在常返情形下为 G 选择的定义，公式和证明可以对所有 d 相同.

令 $G_H(x,y) = G(x,y) - G(x,\bar{y})$. 将 $G_H(x,y)$ 理解为" 起始于 x 的 Brownian 运动在离开 H 前在 y 处的期望占据时间（密度）." 这一解释的合理性在于（对适当的 f）

$$E_x\left(\int_0^\tau f(B_t)\,\mathrm{d}t\right) = \int G_H(x,y)f(y)\,\mathrm{d}y$$

$d = 1$ 时答案非常简单.

$$G_H(x,y) = 2(x \wedge y) \quad \text{对所有} x, y > 0 \tag{9.7.2}$$

有点令人惊讶的是 $y \mapsto G_H(x,y)$ 对 $y \geqslant x$ 恒为 $2x$，即所有点 $y > x$ 有相同的期望占据时间!

证明. $G(x,y) = -|x-y|$，因此 $G_H(x,y) = -|x-y| + |x+y|$. 分情况讨论，可见

$$G_H(x,y) = \begin{cases} -(x-y) + (x+y) = 2y & \text{当} 0 < y < x \\ -(y-x) + (x+y) = 2x & \text{当} x < y \end{cases}$$

这就给出了所需公式. □

球. 令 $G(x,y)$ 为 (9.6.7) 中定义的 Green 函数.

定理 9.7.7. 若 g 有界可测，则

$$E_x \int_0^\tau g(B_t)\,\mathrm{d}t = \int G_D(x,y)g(y)\,\mathrm{d}y$$

其中

$$G_D(x,y)=G(x,y)-\int\frac{1-|x|^2}{|x-z|^d}G(z,y)\,\pi(\mathrm{d}z) \tag{9.7.3}$$

这由 (9.7.1) 和 (9.5.4) 推出.

这一公式可以大大简化. 查阅 Folland (1976) 第 123 页，发现

定理 9.7.8. 在 $d\geqslant 3$ 时，若 $0<|y|<1$ 则

$$G_D(x,y)=G(x,y)-|y|^{2-d}G(x,y/|y|^2)$$

证明. 由 (9.7.1)，只需证明 (i) 第二项在 D 中调和，(ii) 在边界上等于 $G(x,y)$. $y\mapsto y/|y|^2$ 将 $y\in D$ 中的点映到 D^c，因此 (i) 成立. 为证 (ii)，注意

$$\begin{aligned}G(x,y)-|y|^{2-d}G(x,y/|y|^2)&=\frac{C_d}{|x-y|^{d-2}}-\frac{C_d}{|y|^{d-2}}\left|x-\frac{y}{|y|^2}\right|^{-(d-2)}\\&=\frac{C_d}{|x-y|^{d-2}}-\frac{C_d}{|x|y|-y|y|^{-1}|^{d-2}}=0\end{aligned}$$

最后等式由对下一证明有用的事实推出: □

引理 9.7.9. 若 $|x|=1$ 则 $|x|y|-y|y|^{-1}|=|x-y|$.

证明. 利用 $|z|^2=z\cdot z$ 及 $|x|^2=1$ 有

$$|x|y|-y|y|^{-1}|^2=|x|^2|y|^2-2x\cdot y+1=|y|^2-2x\cdot y+|x|^2=|x-y|^2$$

证毕. □

再次查阅 Folland (1976) 第 123 页，”猜测”

定理 9.7.10. 在 $d=2$ 时，若 $0<|y|<1$ 则

$$G_D(x,y)=-\frac{1}{\pi}\ln|x-y|-\ln|x|y|-y|y|^{-1}|$$

证明留给读者.

在 $d=1$ 时，$D=(-1,1)$，(9.7.3) 表明

$$\begin{aligned}G_D(x,y)&=G(x,y)-\frac{x+1}{2}G(1,y)-\frac{1-x}{2}G(-1,y)\\&=-|x-y|+\frac{x+1}{2}(1-y)+\frac{1-x}{2}(y+1)\\&=-|x-y|+1-xy\end{aligned}$$

考虑 $x \geqslant y$ 和 $x \leqslant y$ 两种情形得

$$G_D(x,y) = \begin{cases} (1-x)(1+y) & -1 \leqslant y \leqslant x \leqslant 1 \\ (1-y)(1+x) & -1 \leqslant x \leqslant y \leqslant 1 \end{cases} \tag{9.7.4}$$

几何上，若固定 y，则 $x \mapsto G_D(x,y)$ 由以下条件确定：它在 $[0,y]$ 和 $[y,1]$ 上线性，$G_D(0,y) = G_D(1,y) = 0$，且 $\frac{1}{2}G_D''(y,y) = -\delta_y$（$y$ 处的点质量）. 即，G_D' 在 y 处不连续且 $G_D'(y-,y) - G_D'(y+,y) = 2$.

习题

习题 9.7.1. 设 G 是有界开集且 $\tau = \inf\{t : B_t \notin G\}$. 为求 $w(x) = E_x\tau^2$，应解什么方程?

第 8 节 薛定谔方程

本节内容遵循 Chung 与 Zhao (1995) 的阐述. 本节中，我们将考虑在 Dirichlet 问题左侧添加 cu（其中 c 为有界连续函数）后的情形. 即，给定 ∂G 上的有界连续函数 f，研究

(a) $u \in C^2$ 且在 G 内满足 $\frac{1}{2}\Delta u + cu = 0$.

(b) 在 ∂G 的每一点处，u 连续且 $u = f$.

一如既往，求解该偏微分方程的第一步是找到一个局部鞅.

定理 9.8.1. 设 $\tau = \inf\{t > 0 : B_t \notin G\}$. 若 u 满足 (a)，则

$$M_t = u(B_t) \exp\left(\int_0^t c(B_s)\,\mathrm{d}s\right)$$

是 $[0, \tau)$ 上的局部鞅.

证明. 记 $c_t = \int_0^t c(B_s)\,\mathrm{d}s$. 对 (9.4.1) 版本的 Itô 公式应用得

$$u(B_t)\exp(c_t) - u(B_0) = \int_0^t \exp(c_s)\nabla u(B_s)\cdot \mathrm{d}B_s + \int_0^t u(B_s)\exp(c_s)c(B_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t \Delta u(B_s)\exp(c_s)\,\mathrm{d}s$$

对 $t < \tau$ 成立. 这即证明了结论，因为 $\frac{1}{2}\Delta u + cu = 0$，且由定理 7.6.4 知右端第一项是 $[0, \tau)$ 上的局部鞅. □

至此，读者或许会预期下一步——如同此前五次所做那样——假设一切有界并得出：若存在有界解，则必为

$$v(x) \equiv E_x(f(B_\tau)\exp(c_\tau))$$

但本节不采用这一方法，因为下例表明该结论不成立.

例题 9.8.2. 设 $d = 1$，$G = (-a, a)$，$c \equiv \gamma$，$f \equiv 1$. 我们所考虑的方程为

$$\frac{1}{2}u'' + \gamma u = 0, \quad u(a) = u(-a) = 1$$

通解为 $A\cos bx + B\sin bx$，其中 $b = \sqrt{2\gamma}$. 因此，若要满足边界条件，必须有

$$1 = A\cos ba + B\sin ba$$
$$1 = A\cos(-ba) + B\sin(-ba) = A\cos ba - B\sin ba$$

将两式相加再相减，得

$$2 = 2A\cos ba, \quad 0 = 2B\sin ba$$

由此可知，$B=0$ 恒为解，而 A 是否有解则不确定.

若 $\cos ba=0$，则不存在解. 若 $\cos ba\neq 0$，则 $u(x)=\cos bx/\cos ba$ 是一个解. 稍后可见，若 $ab<\pi/2$，则

$$v(x)=\frac{\cos bx}{\cos ba}$$

但当 $ab>\pi/2$ 时这不可能成立，因为 $v(x)\geqslant 0$ 而右端对某些 x 值却小于 0.

上例的问题在于当 $ab>\pi/2$ 时 $c\equiv\gamma$ 过大，或者更准确地说，令

$$w(x)=E_x\exp\left(\int_0^\tau c(B_s)\,\mathrm{d}s\right)$$

则 w 在 $(-a,a)$ 中恒等于 ∞. 本节余下部分致力于证明：若 $w\not\equiv\infty$，则 “一切正常”. 为此需分几步进行. 首先证明

引理 9.8.3. 设 $\theta>0$. 存在 $\mu>0$ 使得若 H 为开集且 Lebesgue 测度 $|H|\leqslant\mu$，$\tau_H=\inf\{t>0:B_t\notin H\}$，则

$$\sup_x E_x(\exp(\theta\tau_H))\leqslant 2$$

证明. 取 $\gamma>0$ 使得 $e^{\theta\gamma}\leqslant 4/3$. 显然，

$$P_x(\tau_H>\gamma)\leqslant\int_H\frac{1}{(2\pi\gamma)^{d/2}}e^{-|x-y|^2/2\gamma}\,\mathrm{d}y\leqslant\frac{|H|}{(2\pi\gamma)^{d/2}}\leqslant\frac{1}{4}$$

只需取 μ 足够小. 利用 Markov 性质，可推出

$$P_x(\tau_H>k\gamma)=E_x(P_{B_\gamma}(\tau_H>(k-1)\gamma);\tau_H>\gamma)\leqslant\frac{1}{4}\sup_y P_y(\tau_H>(k-1)\gamma)$$

因此由归纳法，对所有整数 $k\geqslant 0$ 有

$$\sup_x P_x(\tau_H>k\gamma)\leqslant\frac{1}{4^k}$$

由于 e^x 递增，且由假设 $e^{\theta\gamma}\leqslant 4/3$，有

$$\begin{aligned}E_x\exp(\theta\tau_H)&\leqslant\sum_{k=1}^\infty\exp(\theta\gamma k)P_x((k-1)\gamma<\tau_H\leqslant k\gamma)\\&\leqslant\sum_{k=1}^\infty\left(\frac{4}{3}\right)^k\frac{1}{4^{k-1}}=\frac{4}{3}\sum_{k=1}^\infty\frac{1}{3^{k-1}}=\frac{4}{3}\cdot\frac{1}{1-\frac{1}{3}}=2\end{aligned}$$

细心的读者或许已经注意到，上述期望遗漏了 $\tau_H=0$ 的情形. 但由 Blumenthal 0-1 律，要么 $P_x(\tau_H=0)=0$（此时计算有效），要么 $P_x(\tau_H=0)=1$（此时 $E_x\exp(\theta\tau_H)=1$）. □

令 $c^* = \sup_x |c(x)|$. 由引理 9.8.3，可取 r_0 充分小，使得若 $T_r = \inf\{t : |B_t - B_0| > r\}$ 且 $r \leqslant r_0$，则对所有 x 有 $E_x \exp(c^* T_r) \leqslant 2$.

引理 9.8.4. 设 $2\delta \leqslant r_0$. 若 $D(x, 2\delta) \subset G$ 且 $y \in D(x, \delta)$，则

$$w(y) \leqslant 2^{d+2} w(x)$$

注意这一性质表明：若 $w(x) < \infty$，则当 $y \in D(x, \delta)$ 时 $w(y) < \infty$.

证明. 若 $D(y, r) \subset G$ 且 $r \leqslant r_0$，则强 Markov 性质蕴含

$$w(y) = E_y[\exp(c_{T_r}) w(B(T_r))] \leqslant E_y[\exp(c^* T_r) w(B(T_r))]$$

$$= E_y[\exp(c^* T_r)] \int_{\partial D(y,r)} w(z) \pi(\mathrm{d}z) \leqslant 2 \int_{\partial D(y,r)} w(z) \pi(\mathrm{d}z) \tag{9.8.1}$$

其中 π 是 $\partial D(y, r)$ 上规范化为概率测度的表面测度，因为离开时间 T_r 与离开位置 $B(T_r)$ 独立.

若 $\delta \leqslant r_0$ 且 $D(y, \delta) \subset G$，将最后一个不等式乘以 r^{d-1} 并从 0 到 δ 积分，得

$$\frac{\delta^d}{d} w(y) \leqslant 2 \cdot \frac{1}{\sigma_d} \int_{D(y,\delta)} w(z) \, \mathrm{d}z$$

其中 σ_d 是 $\{x : |x| = 1\}$ 的表面积. 整理得

$$\int_{D(y,\delta)} w(z) \, \mathrm{d}z \geqslant 2^{-1} \frac{\delta^d}{C_o} w(y) \tag{9.8.2}$$

其中 $C_o = d/\sigma_d$ 是仅依赖于 d 的常数.

在 (9.8.1) 中以 $y = x$ 重复上述推导，并利用 $c_{T_r} \geqslant -c^* T_r$，得

$$w(x) = E_x[\exp(c_{T_r}) w(B(T_r))] \geqslant E_x[\exp(-c^* T_r) w(B(T_r))]$$

$$= E_x[\exp(-c^* T_r)] \int_{\partial D(x,r)} w(z) \pi(\mathrm{d}z)$$

由于 $1/x$ 为凸函数，由 Jensen 不等式得

$$E_x[\exp(-c^* T_r)] \geqslant 1/E_x[\exp(c^* T_r)] \geqslant 1/2$$

结合最后两式，乘以 r^{d-1} 并从 0 到 2δ 积分，得下界

$$\frac{(2\delta)^d}{d} w(x) \geqslant 2^{-1} \cdot \frac{1}{\sigma_d} \int_{D(x,2\delta)} w(z) \, \mathrm{d}z$$

整理并利用 $D(x,2\delta)\supset D(y,\delta)$ 以及 $w\geqslant 0$，有

$$w(x)\geqslant 2^{-1}\frac{C_o}{(2\delta)^d}\int_{D(y,\delta)}w(z)\,\mathrm{d}z \tag{9.8.3}$$

其中 $C_o=d/\sigma_d$. 结合 (9.8.3) 与 (9.8.2)，得

$$\frac{C_o}{(2\delta)^d}\int_{D(y,\delta)}w(z)\,\mathrm{d}z\geqslant\frac{C_o}{(2\delta)^d}\cdot 2^{-1}C_o\delta^d w(y)=2^{-(d+2)}w(y)$$

即得所需结果. □

引理 9.8.4 与简单的覆盖论证导出

定理 9.8.5. 设 G 为连通开集. 若 $w\not\equiv\infty$，则对所有 $x\in G$ 有 $w(x)<\infty$.

证明. 由引理 9.8.4 可见，若 $w(x)<\infty$，$2\delta\leqslant r_0$ 且 $D(x,2\delta)\subset G$，则 w 在 $D(x,\delta)$ 上有限. 由此，$G_0=\{x:w(x)<\infty\}$ 是 G 的开子集. 为论证 G_0 也是闭集（视为 G 的子集），注意若 $2\delta<r_0$，$D(y,3\delta)\subset G$ 且 $x_n\in G_0$ 满足 $x_n\to y\in G$，则对充分大的 n，有 $y\in D(x_n,\delta)$ 且 $D(x_n,2\delta)\subset G$，故 $w(y)<\infty$. □

在讨论唯一性之前，先加强这一结论.

定理 9.8.6. 设 G 为具有有限 Lebesgue 测度 $|G|<\infty$ 的连通开集. 若 $w\not\equiv\infty$，则

$$\sup_x w(x)<\infty$$

证明. 设 $K\subset G$ 为紧集使得 $|G-K|<\mu$，其中 μ 为引理 9.8.3 中 $\theta=c^*$ 时的常数. 对每个 $x\in K$，可取 δ_x 使得 $2\delta_x\leqslant r_0$ 且 $D(x,2\delta_x)\subset G$. 开集 $D(x,\delta_x)$ 覆盖 K，故存在有限子覆盖 $D(x_i,\delta_{x_i})$，$1\leqslant i\leqslant I$. 显然

$$\sup_{1\leqslant i\leqslant I}w(x_i)<\infty$$

引理 9.8.4 蕴含当 $y\in D(x_i,\delta_{x_i})$ 时 $w(y)\leqslant 2^{d+2}w(x_i)$，故

$$M=\sup_{y\in K}w(y)<\infty$$

若 $y\in H=G-K$，则由 (6.3a) 有 $E_y(\exp(c^*\tau_H))\leqslant 2$，利用强 Markov 性质

$$\begin{aligned}w(y)&=E_y(\exp(c_\tau);\tau_H=\tau)+E_y(\exp(c_{\tau_H})w(B_{\tau_H});B_{\tau_H}\in K)\\&\leqslant 2+ME_y(\exp(c_{\tau_H});B_{\tau_H}\in K)\leqslant 2+2M\end{aligned}$$

即得所需结果. □

确立定理 9.8.6 后，下面证明唯一性结果. 为简化叙述，列出本节余下部分采用的假设:
(A1) G 为有界连通开集.
(A2) f 与 c 有界且连续.
(A3) $w \not\equiv \infty$.

定理 9.8.7. 若存在有界解，则它必为

$$v(x) \equiv E_x(f(B_\tau)\exp(c_\tau))$$

证明. 定理 9.8.1 蕴含 $M_s = u(B_s)\exp(c_s)$ 是 $[0,\tau)$ 上的局部鞅. 由于 f，c，u 有界，令 $s \uparrow \tau \wedge t$ 并利用有界收敛定理得

$$u(x) = E_x(f(B_\tau)\exp(c_\tau); \tau \leqslant t) + E_x(u(B_t)\exp(c_t); \tau > t)$$

由 f 有界且 $w(x) = E_x \exp(c_\tau) < \infty$，由控制收敛定理，$t \to \infty$ 时第一项收敛于 $E_x(f(B_\tau)\exp(c_\tau))$. 为证第二项趋于 0，首先注意因 $\{\tau > t\} \in \mathcal{F}_t$，由条件期望的定义与 Markov 性质得

$$\begin{aligned} E_x(u(B_t)\exp(c_\tau); \tau > t) &= E_x(E_x(u(B_t)\exp(c_\tau)|\mathcal{F}_t); \tau > t) \\ &= E_x(u(B_t)\exp(c_t)w(B_t); \tau > t) \end{aligned}$$

现在我们断言对所有 $y \in G$ 有

$$w(y) \geqslant \exp(-c^*)P_y(\tau \leqslant 1) \geqslant \varepsilon > 0$$

第一个不等式是平凡的. 最后两个不等式容易由 (A1) 推出. 将 $w(B_t)$ 替换为 ε，

$$\begin{aligned} E_x(|u(B_t)|\exp(c_t); \tau > t) &\leqslant \varepsilon^{-1}E_x(|u(B_t)|\exp(c_\tau); \tau > t) \\ &\leqslant \varepsilon^{-1}\|u\|_\infty E_x(\exp(c_\tau); \tau > t) \to 0 \end{aligned}$$

当 $t \to \infty$ 时，由控制收敛定理，因为 $w(x) = E_x \exp(c_\tau) < \infty$ 且 $P_x(\tau < \infty) = 1$. 回到证明中的第一个等式，已证 $u(x) = v(x)$. 证毕. □

唯一性的讨论至此结束. 接下来讨论存在性，这相对容易. 以下均假设 (A1)–(A3) 成立.

定理 9.8.8. 若 $v \in C^2$，则它在 G 内满足 (a).

证明. 设 $x \in D$ 且 $B(x,r) \subset D$. 设 σ 为从 $B(x,r)$ 离开的时间. 由强 Markov 性质

$$v(x) = E_x v(B_\sigma) \exp(c_\sigma)$$

由 Taylor 定理

$$\begin{aligned} E_x v(B(\sigma)) \exp(c_\sigma) - v(x) = &\sum_{i=1}^{d} E_x \frac{\partial v}{\partial x_i}(x)[B_i(\tau_D) - x_i] \\ &+ \frac{1}{2} \sum_{1 \leqslant i,j \leqslant d} E_x \frac{\partial^2 v}{\partial x_i \partial x_j}(x)[B_i(\tau_D) - x_i][B_j(\tau_D) - x_j] + o(r^2) \end{aligned}$$

由于对 $i \neq j$，$B_i(t)$，$B_i^2(t) - t$ 与 $B_i(t)B_j(t)$ 均为鞅，有

$$E_x v(B(\tau_D)) - v(x) = \frac{1}{2} \Delta v(x) E_x \tau_B + o(r^2)$$

另一方面，$E_x \int_0^{\tau_B} c(B_s) v(B_s) \, \mathrm{d}s = [c(x)v(x) + o(1)] E_x \tau_B$，故令 $r \to 0$ 得 $\frac{1}{2}\Delta v + cv = 0$. □

证明定理 9.8.8 后，下一步讨论边界条件. 如同前两节，需要边界是正则的.

定理 9.8.9. v 在 ∂G 的每个正则点处满足 (b).

证明. 设 y 为 ∂G 的正则点. 由 (4.5a) 与 (4.5b) 已证，若 $x_n \to y$，则对所有 $\delta > 0$ 有 $P_{x_n}(\tau \leqslant \delta) \to 1$ 且 $P_{x_n}(B_\tau \in D(y,\delta)) \to 1$. 由于 c 有界且 f 有界连续，由有界收敛定理得

$$E_{x_n}(\exp(c_\tau) f(B_\tau); \tau \leqslant 1) \to f(y)$$

为控制其余空间的贡献，我们观察到若 $|c| \leqslant M$，则利用 Markov 性质与定理 9.8.6 所确立的 w 的有界性，有

$$\begin{aligned} E_{x_n}(\exp(c_\tau) f(B_\tau); \tau > 1) &\leqslant e^M \|f\|_\infty E_{x_n}(w(B_1); \tau > 1) \\ &\leqslant e^M \|f\|_\infty \|w\|_\infty P_{x_n}(\tau > 1) \to 0 \end{aligned}$$

证毕. □

下面讨论 v 何时足够光滑从而成为解.

定理 9.8.10. 若在 (A1)–(A3) 之外，c 还是 Hölder 连续的，则 $v \in C^2$，因而满足 (a).

回到例 9.8.2. 回忆 $d = 1$，$G = (-a, a)$，$c(x) \equiv \gamma > 0$，$f \equiv 1$.

定理 9.8.11. 设 $\tau = \inf\{t : B_t \notin (-a, a)\}$. 若 $0 < \gamma < \pi^2/8a^2$，则

$$E_x e^{\gamma\tau} = \frac{\cos(x\sqrt{2\gamma})}{\cos(a\sqrt{2\gamma})}$$

若 $\gamma \geqslant \pi^2/8a^2$，则 $E_x e^{\gamma\tau} \equiv \infty$.

证明. 由例 9.8.2 的结果

$$u(x) = \frac{\cos(x\sqrt{2\gamma})}{\cos(a\sqrt{2\gamma})}$$

是非负解，满足

$$\frac{1}{2}u'' + \gamma u = 0, \quad u(-a) = u(a) = 1$$

为验证 $w \not\equiv \infty$，注意定理 9.8.1 蕴含 $M_t = u(B_t)e^{\gamma t}$ 是 $[0, \tau)$ 上的局部鞅. 若令 T_n 为递减 M_t 的停时序列，则

$$u(x) = E_x u(B_{T_n \wedge n}) e^{\gamma(T_n \wedge n)}$$

令 $n \to \infty$ 并注意 $T_n \wedge n \uparrow \tau$，由 Fatou 引理得

$$u(x) \geqslant E_x e^{\gamma\tau}$$

最后一个不等式蕴含 $w \not\equiv \infty$，故由定理 9.8.8 得 $\gamma < \pi^2/8a^2$ 时的结果.

为证 $\gamma \geqslant \pi^2/8a^2$ 时 $w(x) \equiv \infty$，假设不然. 在此情形 (6.3) 蕴含 $v(x) = E_x(f(B_\tau)\exp(c_\tau))$ 是该方程的唯一解，但例 9.8.2 中的计算表明此时不存在非负解. □

附录 A
测度论细节

本附录给出正文中陈述但未证明的测度论结果的证明.

第 1 节　Carathéodory 扩张定理

本节旨在证明:

定理 A.1.1. 设 $\mathcal{S}$ 为半代数且 μ 定义在 $\mathcal{S}$ 上满足 $\mu(\varnothing)=0$. 假设

(i) 若 $S\in\mathcal{S}$ 是有限个不交集合 $S_i\in\mathcal{S}$ 的并, 则 $\mu(S)=\sum_i\mu(S_i)$;

(ii) 若 $S_i,S\in\mathcal{S}$ 且 $S=+_{i\geq1}S_i$, 则 $\mu(S)\leq\sum_i\mu(S_i)$.

则 μ 可唯一延拓为 $\bar{\mathcal{S}}$ 上的测度 $\bar{\mu}$, 其中 $\bar{\mathcal{S}}$ 是由 $\mathcal{S}$ 生成的代数. 若该延拓是 σ-有限的, 则存在唯一的延拓 ν 为 $\sigma(\mathcal{S})$ 上的测度.

证明. 引理 1.1.7 表明 $\bar{\mathcal{S}}$ 是 $\mathcal{S}$ 中集合的有限不交并的全体. 当 $A=+_iS_i$ 时, 定义 $\bar{\mu}$ 在 $\bar{\mathcal{S}}$ 上为 $\bar{\mu}(A)=\sum_i\mu(S_i)$. 为验证 $\bar{\mu}$ 定义良好, 假设 $A=+_jT_j$ 并注意到

$$S_i=+_j(S_i\cap T_j),\quad T_j=+_i(S_i\cap T_j),$$

故 (i) 蕴含

$$\sum_i\mu(S_i)=\sum_{i,j}\mu(S_i\cap T_j)=\sum_j\mu(T_j).$$

第 1.1 节已证:

引理 A.1.2. 假设 (i) 成立.

(a) 若 $A,B_i\in\bar{\mathcal{S}}$ 且 $A=+_{i=1}^nB_i$, 则 $\bar{\mu}(A)=\sum_i\bar{\mu}(B_i)$.

(b) 若 $A, B_i \in \bar{\mathcal{S}}$ 且 $A \subset \cup_{i=1}^{n} B_i$, 则 $\bar{\mu}(A) \leq \sum_i \bar{\mu}(B_i)$.

为将可加性性质推广到 $A \in \bar{\mathcal{S}}$ 是可数不交并 $A = +_{i\geq 1} B_i$（其中 $B_i \in \bar{\mathcal{S}}$）的情形, 注意到每个 $B_i = +_j S_{i,j}$ 且 $S_{i,j} \in \mathcal{S}$, 而 $\sum_{i\geq 1} \bar{\mu}(B_i) = \sum_{i\geq 1, j} \mu(S_{i,j})$, 故将 B_i 替换为 $S_{i,j}$ 后, 不失一般性可设 $B_i \in \mathcal{S}$. 现 $A \in \bar{\mathcal{S}}$ 蕴含 $A = +_j T_j$（有限不交并）且 $T_j = +_{i\geq 1} T_j \cap B_i$, 故 (ii) 蕴含

$$\mu(T_j) \leq \sum_{i\geq 1} \mu(T_j \cap B_i).$$

对 j 求和并注意到非负数可以任意顺序求和,

$$\bar{\mu}(A) = \sum_j \mu(T_j) \leq \sum_{i\geq 1} \sum_j \mu(T_j \cap B_i) = \sum_{i\geq 1} \mu(B_i),$$

最后等式由 (i) 得到. 为证反向不等式, 令 $A_n = B_1 + \cdots + B_n$ 且 $C_n = A \cap A_n^c$. $C_n \in \bar{\mathcal{S}}$, 因为 $\bar{\mathcal{S}}$ 是代数, 故 $\bar{\mu}$ 的有限可加性蕴含

$$\bar{\mu}(A) = \bar{\mu}(B_1) + \cdots + \bar{\mu}(B_n) + \bar{\mu}(C_n) \geq \bar{\mu}(B_1) + \cdots + \bar{\mu}(B_n),$$

令 $n \to \infty$, 得 $\bar{\mu}(A) \geq \sum_{i\geq 1} \bar{\mu}(B_i)$. □

在代数 $\bar{\mathcal{S}}$ 上定义测度后, 现通过建立以下定理来完成证明:

定理 A.1.3. Carathéodory 扩张定理. 设 μ 为代数 $\mathcal{A}$ 上的 σ-有限测度. 则 μ 可唯一延拓到 $\sigma(\mathcal{A})$（包含 $\mathcal{A}$ 的最小 σ-代数）上.

唯一性. 先证明延拓的唯一性, 再处理存在性问题. 唯一性证明的关键是 Dynkin 的 $\pi - \lambda$ 定理, 这是本书将多次使用的结果. 照例, 需要一些定义才能陈述该结果. 若 $\mathcal{P}$ 对交封闭, 即若 $A, B \in \mathcal{P}$ 则 $A \cap B \in \mathcal{P}$, 则称 $\mathcal{P}$ 为 π-系. 例如, 矩形 $(a_1, b_1] \times \cdots \times (a_d, b_d]$ 的全体构成一个 π-系. 若 $\mathcal{L}$ 满足: (i) $\Omega \in \mathcal{L}$; (ii) 若 $A, B \in \mathcal{L}$ 且 $A \subset B$ 则 $B - A \in \mathcal{L}$; (iii) 若 $A_n \in \mathcal{L}$ 且 $A_n \uparrow A$ 则 $A \in \mathcal{L}$, 则称 $\mathcal{L}$ 为 λ-系. 读者将立即看到下一结果正是证明延拓唯一性所需.

定理 A.1.4. $\pi - \lambda$ 定理. 若 $\mathcal{P}$ 为 π-系且 $\mathcal{L}$ 为包含 $\mathcal{P}$ 的 λ-系, 则 $\sigma(\mathcal{P}) \subset \mathcal{L}$.

证明. 下面证明

(a) 若 $\ell(\mathcal{P})$ 为包含 $\mathcal{P}$ 的最小 λ-系, 则 $\ell(\mathcal{P})$ 是 σ-域.

所需结果由 (a) 推出. 为看出这一点, 注意因 $\sigma(\mathcal{P})$ 是包含 $\mathcal{P}$ 的最小 σ-域而 $\ell(\mathcal{P})$ 是包含 $\mathcal{P}$ 的最小 λ-系, 有

$$\sigma(\mathcal{P}) \subset \ell(\mathcal{P}) \subset \mathcal{L}.$$

为证 (a), 首先注意对交封闭的 λ-系是 σ-域, 因为

$$\begin{aligned}&\text{若}A\in\mathcal{L}\text{ 则}A^c=\Omega-A\in\mathcal{L},\\&A\cup B=(A^c\cap B^c)^c,\\&\cup_{i=1}^n A_i\uparrow\cup_{i=1}^\infty A_i\text{ 当}n\uparrow\infty.\end{aligned}$$

因此, 只需证明

(b) $\ell(\mathcal{P})$ 对交封闭.

为证 (b), 令 $G_A=\{B:A\cap B\in\ell(\mathcal{P})\}$ 并证明

(c) 若 $A\in\ell(\mathcal{P})$ 则 G_A 是 λ-系.

验证如下: (i) $\Omega\in G_A$ 因 $A\in\ell(\mathcal{P})$. (ii) 若 $B,C\in G_A$ 且 $B\supset C$, 则 $A\cap(B-C)=(A\cap B)-(A\cap C)\in\ell(\mathcal{P})$ 因 $A\cap B,A\cap C\in\ell(\mathcal{P})$ 且 $\ell(\mathcal{P})$ 是 λ-系. (iii) 若 $B_n\in G_A$ 且 $B_n\uparrow B$, 则 $A\cap B_n\uparrow A\cap B\in\ell(\mathcal{P})$ 因 $A\cap B_n\in\ell(\mathcal{P})$ 且 $\ell(\mathcal{P})$ 是 λ-系.

由 (c) 到 (b), 注意因 $\mathcal{P}$ 是 π-系,

$$\text{若}A\in\mathcal{P}\text{ 则}G_A\supset\mathcal{P},\text{ 故 (c) 蕴含}G_A\supset\ell(\mathcal{P}),$$

即若 $A\in\mathcal{P}$ 且 $B\in\ell(\mathcal{P})$ 则 $A\cap B\in\ell(\mathcal{P})$. 在最后一句中交换 A 和 B: 若 $A\in\ell(\mathcal{P})$ 且 $B\in\mathcal{P}$ 则 $A\cap B\in\ell(\mathcal{P})$, 但这蕴含

$$\text{若}A\in\ell(\mathcal{P})\text{ 则}G_A\supset\mathcal{P},\text{ 故 (c) 蕴含}G_A\supset\ell(\mathcal{P}).$$

此结论蕴含若 $A,B\in\ell(\mathcal{P})$ 则 $A\cap B\in\ell(\mathcal{P})$, 即证 (b). □

为证明定理 A.1.3 中的延拓是唯一的, 下面证明:

定理 A.1.5. 设 $\mathcal{P}$ 为 π-系. 若 ν_1 和 ν_2 为（σ-域 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 上的）测度且在 $\mathcal{P}$ 上一致, 且存在序列 $A_n\in\mathcal{P}$ 满足 $A_n\uparrow\Omega$ 且 $\nu_i(A_n)<\infty$, 则 ν_1 和 ν_2 在 $\sigma(\mathcal{P})$ 上一致.

证明. 设 $A\in\mathcal{P}$ 满足 $\nu_1(A)=\nu_2(A)<\infty$. 令

$$\mathcal{L}=\{B\in\sigma(\mathcal{P}):\nu_1(A\cap B)=\nu_2(A\cap B)\}.$$

现验证 $\mathcal{L}$ 构成 λ-系. 因 $A\in\mathcal{P},\nu_1(A)=\nu_2(A)$ 且 $\Omega\in\mathcal{L}$. 若 $B,C\in\mathcal{L}$ 且 $C\subset B$, 则

$$\nu_1(A\cap(B-C))=\nu_1(A\cap B)-\nu_1(A\cap C)=\nu_2(A\cap B)-\nu_2(A\cap C)=\nu_2(A\cap(B-C)).$$

这里用到 $\nu_i(A)<\infty$ 以保证减法合理. 最后, 若 $B_n\in\mathcal{L}$ 且 $B_n\uparrow B$, 则定理 1.1.1 的 (iii) 部分蕴含

$$\nu_1(A\cap B)=\lim_{n\to\infty}\nu_1(A\cap B_n)=\lim_{n\to\infty}\nu_2(A\cap B_n)=\nu_2(A\cap B).$$

因 $\mathcal{P}$ 由假设对交封闭, $\pi-\lambda$ 定理蕴含 $\mathcal{L}\supset\sigma(\mathcal{P})$, 即若 $A\in\mathcal{P}$ 满足 $\nu_1(A)=\nu_2(A)<\infty$ 且 $B\in\sigma(\mathcal{P})$, 则 $\nu_1(A\cap B)=\nu_2(A\cap B)$. 令 $A_n\in\mathcal{P}$ 满足 $A_n\uparrow\Omega$, $\nu_1(A_n)=\nu_2(A_n)<\infty$, 利用最后结果和定理 1.1.1 的 (iii) 部分, 即得所需结论. □

习题 A.1.1. 给出 $\mathcal{F}=\{1,2,3,4\}$ 的所有子集构成的集类上两个概率测度 $\mu\neq\nu$ 的例子, 它们在一集合族 $\mathcal{C}$ 上一致且 $\sigma(\mathcal{C})=\mathcal{F}$, 即包含 $\mathcal{C}$ 的最小 σ-代数是 $\mathcal{F}$.

存在性. 下面证明定义在代数 $\mathcal{A}$ 上的（不必 σ-有限的）测度可延拓到由 $\mathcal{A}$ 生成的 σ-代数上. 若 $E\subset\Omega$, 令 $\mu^*(E)=\inf\sum_i\mu(A_i)$, 其中下确界取遍所有来自 $\mathcal{A}$ 的序列使得 $E\subset\cup_iA_i$. 直观上, 若 ν 是与 μ 在 $\mathcal{A}$ 上一致的测度, 则定理 1.1.1 的 (ii) 部分蕴含

$$\nu(E)\le\nu(\cup_iA_i)\le\sum_i\nu(A_i)=\sum_i\mu(A_i),$$

故 $\mu^*(E)$ 是 E 的测度的上界. 直观上, 可测集就是上界是紧的那些集合. 形式上, 若对所有 $F\subset\Omega$ 有

$$\mu^*(F)=\mu^*(F\cap E)+\mu^*(F\cap E^c)\tag{A.1.1}$$

则称 E 为可测的.

最后这个定义不太直观, 但下面的证明将表明它行之有效.

由定义立即有 μ^* 具有以下性质:

(i) **单调性.** 若 $E\subset F$ 则 $\mu^*(E)\le\mu^*(F)$.

(ii) **次可加性.** 若 $F\subset\cup_iF_i$（可数并）, 则 $\mu^*(F)\le\sum_i\mu^*(F_i)$.

任何满足 $\mu^*(\varnothing)=0$ 和 (i), (ii) 的集函数称为外测度. 利用 (ii) 并取 $F_1=F\cap E$, $F_2=F\cap E^c$（其余 $F_i=\varnothing$）, 可见为证明集合可测, 只需证明

$$\mu^*(F)\ge\mu^*(F\cap E)+\mu^*(F\cap E^c).\tag{A.1.2}$$

先证明新定义推广了旧定义.

引理 A.1.6. 若 $A\in\mathcal{A}$ 则 $\mu^*(A)=\mu(A)$ 且 A 可测.

证明. 定理 1.1.1 的 (ii) 部分蕴含若 $A\subset\cup_iA_i$ 则

$$\mu(A)\le\sum_i\mu(A_i),$$

故 $\mu(A) \le \mu^*(A)$. 当然, 总可取 $A_1 = A$ 且其余 $A_i = \varnothing$, 故 $\mu^*(A) \le \mu(A)$.

为证任何 $A \in \mathcal{A}$ 可测, 先注意当 $\mu^*(F) = \infty$ 时不等式 (A.1.2) 显然成立, 故不妨设 $\mu^*(F) < \infty$. 为证 $E = A$ 时 (A.1.2) 成立, 注意到因 $\mu^*(F) < \infty$ 存在序列 $B_i \in \mathcal{A}$ 使得 $\cup_i B_i \supset F$ 且

$$\sum_i \mu(B_i) \le \mu^*(F) + \varepsilon.$$

因 μ 在 $\mathcal{A}$ 上可加且 $\mu = \mu^*$ 在 $\mathcal{A}$ 上, 有

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

对 i 求和并利用 μ^* 的次可加性得

$$\mu^*(F) + \varepsilon \ge \sum_i \mu^*(B_i \cap A) + \sum_i \mu^*(B_i \cap A^c) \ge \mu^*(F \cap A) + \mu^*(F^c \cap A),$$

因 ε 任意, 即得所需结果. □

引理 A.1.7. 可测集类 $\mathcal{A}^*$ 是 σ-域, 且 μ^* 在 $\mathcal{A}^*$ 上的限制是测度.

注释 A.1.8. 此结果对任何外测度都成立.

证明. 由定义显然:

(a) 若 E 可测则 E^c 可测.

下面证明非平凡结论:

(b) 若 E_1 和 E_2 可测则 $E_1 \cup E_2$ 和 $E_1 \cap E_2$ 可测.

(b) 的证明. 为证第一个结论, 令 G 为 Ω 的任意子集. 利用次可加性、E_2 的可测性（在 (A.1.1) 中令 $F = G \cap E_1^c$）和 E_1 的可测性, 得

$$\begin{aligned}
&\mu^*(G \cap (E_1 \cup E_2)) + \mu^*(G \cap (E_1^c \cap E_2^c)) \\
&\quad \le \mu^*(G \cap E_1) + \mu^*(G \cap E_1^c \cap E_2) + \mu^*(G \cap E_1^c \cap E_2^c) \\
&\quad = \mu^*(G \cap E_1) + \mu^*(G \cap E_1^c) = \mu^*(G).
\end{aligned}$$

为证 $E_1 \cap E_2$ 可测, 注意到 $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$ 并利用 (a).

(c) 设 $G \subset \Omega$ 且 $E_1, \dots, E_n$ 为不交可测集. 则

$$\mu^*(G \cap \cup_{i=1}^n E_i) = \sum_{i=1}^n \mu^*(G \cap E_i).$$

(c) 的证明. 令 $F_m = \cup_{i\le m} E_i$. E_n 可测, $F_n \supset E_n$ 且 $F_{n-1} \cap E_n = \varnothing$, 故

$$\mu^*(G \cap F_n) = \mu^*(G \cap F_n \cap E_n) + \mu^*(G \cap F_n \cap E_n^c) = \mu^*(G \cap E_n) + \mu^*(G \cap F_{n-1}).$$

由归纳法即得所需结果.

(d) 若集 E_i 可测则 $E = \cup_{i=1}^{\infty} E_i$ 可测.

(d) 的证明. 令 $E_i' = E_i \cap \cap_{j<i} E_j^c$. (a) 和 (b) 蕴含 E_i' 可测, 故不失一般性可设 E_i 两两不交. 令 $F_n = E_1 \cup \cdots \cup E_n$. F_n 由 (b) 可测, 故利用单调性和 (c) 有

$$\mu^*(G) = \mu^*(G \cap F_n) + \mu^*(G \cap F_n^c) \ge \mu^*(G \cap F_n) + \mu^*(G \cap E^c) = \sum_{i=1}^{n} \mu^*(G \cap E_i) + \mu^*(G \cap E^c).$$

令 $n \to \infty$ 并利用次可加性

$$\mu^*(G) \ge \sum_{i=1}^{\infty} \mu^*(G \cap E_i) + \mu^*(G \cap E^c) \ge \mu^*(G \cap E) + \mu^*(G \cap E^c),$$

此即 (A.1.2).

证明定理 A.1.7 的最后一步是:

(e) 若 $E = \cup_i E_i$ 其中 $E_1, E_2, \ldots$ 不交且可测, 则

$$\mu^*(E) = \sum_{i=1}^{\infty} \mu^*(E_i).$$

(e) 的证明. 令 $F_n = E_1 \cup \cdots \cup E_n$. 由单调性和 (c)

$$\mu^*(E) \ge \mu^*(F_n) = \sum_{i=1}^{n} \mu^*(E_i).$$

令 $n \to \infty$ 并利用次可加性即得所需结论. □

第 2 节　哪些集合是可测的？

第 A.1 节中定理 A.1.3 的证明将测度延拓到 $\mathcal{A}^* \supset \sigma(\mathcal{A})$. 下一个目标是描述这两个 σ-代数之间的关系. 令 $\mathcal{A}_\sigma$ 表示 $\mathcal{A}$ 中集合的可数并的全体, 令 $\mathcal{A}_{\sigma\delta}$ 表示 $\mathcal{A}_\sigma$ 中集合的可数交的全体. 第一个目标是证明每个可测集几乎就是 $\mathcal{A}_{\sigma\delta}$ 中的一个集合.

定义对称差为 $A\Delta B = (A - B) \cup (B - A)$.

引理 A.2.1. 设 E 为任意满足 $\mu^*(E) < \infty$ 的集合.

(i) 对任意 $\varepsilon > 0$, 存在 $A \in \mathcal{A}_\sigma$ 满足 $A \supset E$ 且 $\mu^*(A) \leq \mu^*(E) + \varepsilon$.

(ii) 对任意 $\varepsilon > 0$, 存在 $B \in \mathcal{A}$ 满足 $\mu(B\Delta E) \leq 2\varepsilon$.

(iii) 存在 $C \in \mathcal{A}_{\sigma\delta}$ 满足 $C \supset E$ 且 $\mu^*(C) = \mu^*(E)$.

证明. 由 μ^* 的定义, 存在序列 A_i 使得 $A \equiv \cup_i A_i \supset E$ 且 $\sum_i \mu(A_i) \leq \mu^*(E) + \varepsilon$. 由 μ^* 的定义蕴含 $\mu^*(A) \leq \sum_i \mu(A_i)$, 即得 (i).

对于 (ii), 注意到存在有限并 $B = \cup_{i=1}^n A_i$ 使得 $\mu(A - B) \leq \varepsilon$, 故 $\mu(E - B) \leq \varepsilon$. 因 $\mu(B - E) \leq \mu(A - E) \leq \varepsilon$, 即得所需结果.

对于 (iii), 令 $A_n \in \mathcal{A}_\sigma$ 满足 $A_n \supset E$ 且 $\mu^*(A_n) \leq \mu^*(E) + 1/n$, 并令 $C = \cap_n A_n$. 显然, $C \in \mathcal{A}_{\sigma\delta}$, $B \supset E$, 故由单调性 $\mu^*(C) \geq \mu^*(E)$. 为证另一不等式, 注意 $B \subset A_n$ 故对任意 n 有 $\mu^*(C) \leq \mu^*(A_n) \leq \mu^*(E) + 1/n$. □

定理 A.2.2. 设 μ 在 $\mathcal{A}$ 上是 σ-有限的. $B \in \mathcal{A}^*$ 当且仅当存在 $A \in \mathcal{A}_{\sigma\delta}$ 和满足 $\mu^*(N) = 0$ 的集合 N 使得 $B = A - N (= A \cap N^c)$.

证明. 由引理 A.1.6 和 A.1.7, 若 $A \in \mathcal{A}_{\sigma\delta}$ 则 $A \in \mathcal{A}^*$. A.1.2 节和单调性蕴含 $\mu^*(N) = 0$ 的集合可测, 故再利用引理 A.1.7 得 $A \cap N^c \in \mathcal{A}^*$.

为证另一方向, 令 Ω_i 为不交集合列满足 $\mu(\Omega_i) < \infty$ 且 $\Omega = \cup_i \Omega_i$. 令 $B_i = B \cap \Omega_i$ 并利用引理 A.2.1 找 $A_n^i \in \mathcal{A}_\sigma$ 使得 $A_n^i \supset B_i$ 且 $\mu(A_n^i) \leq \mu^*(E_i) + 1/n2^i$. 令 $A_n = \cup_i A_n^i$. $B \subset A_n$ 且

$$A_n - B \subset \sum_{i=1}^{\infty} (A_n^i - B_i),$$

故由次可加性,

$$\mu^*(A_n - B) \leq \sum_{i=1}^{\infty} \mu^*(A_n^i - B_i) \leq 1/n.$$

因 $A_n \in \mathcal{A}_\sigma$, 集合 $A = \cap_n A_n \in \mathcal{A}_{\sigma\delta}$. 显然, $A \supset B$. 因 $N \equiv A - B \subset A_n - B$ 对所有 n 成立, 单调性蕴含 $\mu^*(N) = 0$, 证毕. □

若测度空间 $(\Omega, \mathcal{F}, \mu)$ 包含所有零测集的子集, 则称其为完备的. 在定理 A.2.2 的证明中, 证明了 $(\Omega, \mathcal{A}^*, \mu^*)$ 是完备的. 下一个结果表明 $(\Omega, \mathcal{A}^*, \mu^*)$ 是 $(\Omega, \sigma(\mathcal{A}), \mu)$ 的完备化.

定理 A.2.3. 若 $(\Omega, \mathcal{F}, \mu)$ 为测度空间, 则存在完备测度空间 $(\Omega, \bar{\mathcal{F}}, \bar{\mu})$（称为 $(\Omega, \mathcal{F}, \mu)$ 的完备化）使得: (i) $E \in \bar{\mathcal{F}}$ 当且仅当 $E = A \cup B$, 其中 $A \in \mathcal{F}$ 且 $B \subset N \in \mathcal{F}$ 满足 $\mu(N) = 0$; (ii) $\bar{\mu}$ 在 $\mathcal{F}$ 上与 μ 一致.

证明. 第一步是验证 $\bar{\mathcal{F}}$ 是 σ-代数. 若 $E_i = A_i \cup B_i$ 其中 $A_i \in \mathcal{F}$ 且 $B_i \subset N_i$ 满足 $\mu(N_i) = 0$, 则 $\cup_i A_i \in \mathcal{F}$ 且次可加性蕴含 $\mu(\cup_i N_i) \leq \sum_i \mu(N_i) = 0$, 故 $\cup_i E_i \in \bar{\mathcal{F}}$. 至于补集, 若 $E = A \cup B$ 且 $B \subset N$, 则 $B^c \supset N^c$ 故

$$E^c = A^c \cap B^c = (A^c \cap N^c) \cup (A^c \cap B^c \cap N).$$

$A^c \cap N^c$ 在 $\mathcal{F}$ 中且 $A^c \cap B^c \cap N \subset N$, 故 $E^c \in \bar{\mathcal{F}}$.

以明显的方式定义 $\bar{\mu}$: 若 $E = A \cup B$ 其中 $A \in \mathcal{F}$ 且 $B \subset N$ 满足 $\mu(N) = 0$, 则令 $\bar{\mu}(E) = \mu(A)$. 首先要证明的是 $\bar{\mu}$ 定义良好, 即若 $E = A_i \cup B_i$ $(i = 1, 2)$ 是两个分解, 则 $\mu(A_1) = \mu(A_2)$. 令 $A_0 = A_1 \cap A_2$ 且 $B_0 = B_1 \cup B_2$. $E = A_0 \cup B_0$ 是第三个分解且 $A_0 \in \mathcal{F}$ 和 $B_0 \subset N_1 \cup N_2$, 并具有如下良好性质: 若 $i = 1$ 或 2

$$\mu(A_0) \leq \mu(A_i) \leq \mu(A_0) + \mu(N_1 \cup N_2) = \mu(A_0).$$

最后一个细节是验证 $\bar{\mu}$ 是测度, 但这很容易. 若 $E_i = A_i \cup B_i$ 不交, 则 $\cup_i E_i$ 可分解为 $\cup_i A_i \cup (\cup_i B_i)$ 且 $A_i \subset E_i$ 不交, 故

$$\bar{\mu}(\cup_i E_i) = \mu(\cup_i A_i) = \sum_i \mu(A_i) = \sum_i \bar{\mu}(E_i).$$

□

定理 1.1.11 允许我们在 $(\mathbb{R}^d, \mathcal{R}^d)$ 上构造 Lebesgue 测度 λ. 利用定理 A.2.3, 可以将 λ 延拓为 $(\mathbb{R}, \bar{\mathcal{R}}^d)$ 上的测度, 其中 $\bar{\mathcal{R}}^d$ 是 $\mathcal{R}^d$ 的完备化. 完成这些后, 很自然地（尽管有些乐观地）问: 是否存在不在 $\bar{\mathcal{R}}^d$ 中的集合? 答案是" 是的", 下面给出 $\mathbb{R}$ 中不可测集 B 的一个例子.

$[0, 1)$ 的不可测子集

构造的关键是观察到 λ 具有平移不变性: 即若 $A \in \bar{\mathcal{R}}$ 且 $x + A = \{x + y : y \in A\}$, 则 $x + A \in \bar{\mathcal{R}}$ 且 $\lambda(A) = \lambda(x + A)$. 若 $x - y$ 是有理数, 我们说 $x, y \in [0, 1)$ 等价并记 $x \sim y$. 由选择公理, 存在集合 B 恰好包含每个等价类中的一个元素. B 即为所求的不可测集, 即,

定理 A.2.4. $B \notin \bar{\mathcal{R}}$.

证明. 关键是以下引理:

引理 A.2.5. 若 $E \subset [0, 1)$ 在 $\bar{\mathcal{R}}$ 中, $x \in (0, 1)$, 且 $x +_0 E = \{(x + y) \mod 1 : y \in E\}$, 则 $\lambda(E) = \lambda(x +_0 E)$.

证明. 令 $A = E \cap [0, 1-x)$ 和 $B = E \cap [1-x, 1)$. 令 $A' = x + A = \{x + y : y \in A\}$ 和 $B' = x - 1 + B$. $A, B \in \bar{\mathcal{R}}$, 故由平移不变性 $A', B' \in \bar{\mathcal{R}}$ 且 $\lambda(A) = \lambda(A')$, $\lambda(B) = \lambda(B')$. 因 $A' \subset [x, 1)$ 和 $B' \subset [0, x)$ 不交,

$$\lambda(E) = \lambda(A) + \lambda(B) = \lambda(A') + \lambda(B') = \lambda(x +_0 E). \qquad \square$$

由引理 A.2.5, 易见 B 不可测; 若它可测, 则 $q +_0 B$ （$q \in \mathbb{Q} \cap [0, 1)$）将是 $[0, 1)$ 中可数不交的可测子集列, 都具有相同的测度 α 且满足

$$\cup_{q \in \mathbb{Q} \cap [0,1)} (q +_0 B) = [0, 1).$$

若 $\alpha > 0$ 则 $\lambda([0, 1)) = \infty$, 若 $\alpha = 0$ 则 $\lambda([0, 1)) = 0$. 两个结论都与 $\lambda([0, 1)) = 1$ 矛盾, 故 $B \notin \bar{\mathcal{R}}$. $\square$

习题 A.2.1. 设 B 为定理 A.2.4 中构造的不可测集.

(i) 令 $B_q = q +_0 B$ 并证明若 $D_q \subset B_q$ 可测, 则 $\lambda(D_q) = 0$.

(ii) 利用 (i) 得出若 $A \subset \mathbb{R}$ 满足 $\lambda(A) > 0$, 则存在不可测集 $S \subset A$.

令 $B' = B \times [0, 1]^{d-1}$ 其中 B 是 $(0, 1)$ 的不可测子集, 则当 $d > 1$ 时得到一个不可测集. 在 $d = 3$ 时, 有一个更有趣得多的例子, 但需要读者做一些预备工作. 在欧几里得几何中, 若 $\mathbb{R}^d$ 的一个子集可通过平移和旋转映射到另一个子集上, 则称它们全等.

断言. 两个全等的可测集必具有相同的 Lebesgue 测度.

习题 A.2.2. 通过证明 (i) 若 B 是矩形 A 的旋转则 $\lambda^*(B) = \lambda(A)$; (ii) 若 C 与 D 全等则 $\lambda^*(C) = \lambda^*(D)$ 来证明 $d = 2$ 时的断言.

Banach-Tarski 定理

Banach 和 Tarski (1924) 利用选择公理证明可以将 $\mathbb{R}^3$ 中的球 $\{x : |x| \leq 1\}$ 分割为有限个集合 $A_1, \ldots, A_n$ 并找到全等的集合 $B_1, \ldots, B_n$, 它们的并是两个不交的单位球! 因全等的集合具有相同的 Lebesgue 测度, 至少有一个集合 A_i 必须不可测. 该构造依赖于 $\mathbb{R}^3$ 中由旋转生成的群不是 Abel 群这一事实. Lindenbaum (1926) 证明这在 $\mathbb{R}^2$ 中对任何有界集都不可能. 关于 Banach-Tarski 定理的通俗介绍, 参见 French (1988).

Solovay 定理

选择公理在前两个不可测集的构造中起了重要作用. Solovay (1970) 证明其使用是不可避免的. 用他自己的话说, " 我们证明, 如果不允许使用选择公理, 则在 Zermelo-Frankel 集合论中无法证明非 Lebesgue 可测集的存在." 这应该使读者相信, 所有" 在实际中" 出现的 $\mathbb{R}^d$ 的子集都在 $\bar{\mathcal{R}}^d$ 中.

第 3 节　符号测度

本节证明 Radon-Nikodym 定理. 为发展该结果, 从一个初看起来似乎无关的主题开始. 设 $(\Omega, \mathcal{F})$ 为可测空间. 若 α 满足 (i) α 取值于 $(-\infty, \infty]$; (ii) $\alpha(\varnothing) = 0$; (iii) 若 $E = +_i E_i$ 为不交并则 $\alpha(E) = \sum_i \alpha(E_i)$（如下意义: 若 $\alpha(E) < \infty$ 则级数绝对收敛且等于 $\alpha(E)$; 若 $\alpha(E) = \infty$ 则 $\sum_i \alpha(E_i)^- < \infty$ 且 $\sum_i \alpha(E_i)^+ = \infty$）, 则称 α 为 $(\Omega, \mathcal{F})$ 上的符号测度. 显然, 符号测度不能同时取值 ∞ 和 $-\infty$, 因 $\alpha(A) + \alpha(B)$ 可能无意义. 在大多数表述中, 符号测度允许取值于 $(-\infty, \infty]$ 或 $[-\infty, \infty)$. 为简化后面的陈述, 将忽略第二种可能. 照例, 转向例子以帮助解释定义.

例题 A.3.1. 设 μ 为测度, f 为满足 $\int f^- \,\mathrm{d}\mu < \infty$ 的函数, 并令 $\alpha(A) = \int_A f \,\mathrm{d}\mu$. 习题 5.8 蕴含 α 是符号测度.

例题 A.3.2. 设 μ_1 和 μ_2 为测度且 $\mu_2(\Omega) < \infty$, 并令 $\alpha(A) = \mu_1(A) - \mu_2(A)$.

下面的 Jordan 分解（定理 A.4.6）将表明例 A.4.2 是一般情形. 为推导该结果, 从两个定义开始. 若每个可测集 $B \subset A$ 满足 $\alpha(B) \geq 0$, 则称集 A 为正集. 若每个可测集 $B \subset A$ 满足 $\alpha(B) \leq 0$, 则称集 A 为负集.

习题 A.3.1. 在例 A.4.1 中, A 为正集当且仅当 $\mu(A \cap \{x : f(x) < 0\}) = 0$.

引理 A.3.3.　(i) 正集的可测子集是正集.
(ii) 若集 A_n 为正集则 $A = \cup_n A_n$ 也是正集.

证明. (i) 是平凡的. 为证 (ii), 注意到

$$B_n = A_n \cap \cap_{m=1}^{n-1} A_m^c \subset A_n$$

是正、不交的且 $\cup_n B_n = \cup_n A_n$. 令 $E \subset A$ 可测且 $E_n = E \cap B_n$. 因 B_n 为正, $\alpha(E_n) \geq 0$, 故 $\alpha(E) = \sum_n \alpha(E_n) \geq 0$. □

引理 A.4.3 中结论对负集亦成立. 下一结果是定理 A.4.5 证明的关键.

引理 A.3.4. 设 E 为满足 $\alpha(E) < 0$ 的可测集. 则存在负集 $F \subset E$ 满足 $\alpha(F) < 0$.

证明. 若 E 是负的, 结论成立. 若否, 令 n_1 为最小的正整数使得存在 $E_1 \subset E$ 满足 $\alpha(E_1) \geq 1/n_1$. 令 $k \geq 2$. 若 $F_k = E - (E_1 \cup \cdots \cup E_{k-1})$ 是负的, 则证毕. 若否, 继续构造, 令 n_k 为最小的正整数使得存在 $E_k \subset F_k$ 满足 $\alpha(E_k) \geq 1/n_k$. 若构造对任何 $k < \infty$ 不停止, 令

$$F = \cap_k F_k = E - (\cup_k E_k).$$

因 $0 > \alpha(E) > -\infty$ 且 $\alpha(E_k) \geq 0$, 由符号测度的定义得

$$\alpha(E) = \alpha(F) + \sum_{k=1}^{\infty} \alpha(E_k).$$

$\alpha(F) \leq \alpha(E) < 0$ 且和有限. 由最后的观察和构造, 可见 F 没有子集 G 满足 $\alpha(G) > 0$, 因若不然则 $\alpha(G) \geq 1/N$ 对某个 N, 产生矛盾. □

定理 A.3.5. Hahn 分解. 设 α 为符号测度. 则存在正集 A 和负集 B 使得 $\Omega = A \cup B$ 且 $A \cap B = \varnothing$.

证明. 令 $c = \inf\{\alpha(B) : B \text{ 为负集}\} \leq 0$. 令 B_i 为负集满足 $\alpha(B_i) \downarrow c$. 由引理 A.4.3, $B = \cup_i B_i$ 是负的, 由 c 定义知 $\alpha(B) \geq c$. 为证 $\alpha(B) \leq c$, 注意到 $\alpha(B) = \alpha(B_i) + \alpha(B - B_i) \leq \alpha(B_i)$ 因 B 是负的, 并令 $i \to \infty$. 最后两个不等式表明 $\alpha(B) = c$, 由符号测度的定义知 $c > -\infty$. 令 $A = B^c$. 为证 A 为正, 注意到若 A 包含满足 $\alpha(E) < 0$ 的集, 则由引理 A.4.4, 它包含满足 $\alpha(F) < 0$ 的负集 F, 但此时 $B \cup F$ 将是满足 $\alpha(B \cup F) = \alpha(B) + \alpha(F) < c$ 的负集, 矛盾. □

Hahn 分解不唯一. 在例 A.4.1 中, A 可以是满足

$$\{x : f(x) > 0\} \subset A \subset \{x : f(x) \geq 0\} \quad \text{a.e.}$$

的任何集, 其中 $B \subset C$ a.e. 表示 $\mu(B \cap C^c) = 0$. 最后一个例子是一般情形的典型. 设 $\Omega = A_1 \cup B_1 = A_2 \cup B_2$ 是两个 Hahn 分解. $A_2 \cap B_1$ 既是正的又是负的, 故它是零集: 它的所有子集测度为 0. 类似地, $A_1 \cap B_2$ 是零集.

若存在集合 A 满足 $\mu_1(A) = 0$ 且 $\mu_2(A^c) = 0$, 则称两测度 μ_1 和 μ_2 相互奇异. 此时, 也称 μ_1 关于 μ_2 奇异并记作 $\mu_1 \perp \mu_2$.

习题 A.3.2. 证明 Cantor 集上的均匀分布（例 1.2.7）关于 Lebesgue 测度奇异.

定理 A.3.6. Jordan 分解. 设 α 为符号测度. 存在相互奇异的测度 α_+ 和 α_- 使得 $\alpha = \alpha_+ - \alpha_-$. 而且, 这样的对是唯一的.

证明. 设 $\Omega = A \cup B$ 为 Hahn 分解. 令

$$\alpha_+(E) = \alpha(E \cap A) \quad \text{且} \quad \alpha_-(E) = -\alpha(E \cap B).$$

因 A 为正且 B 为负, α_+ 和 α_- 是测度. $\alpha_+(A^c) = 0$ 且 $\alpha_-(A) = 0$, 故它们相互奇异. 为证唯一性, 设 $\alpha = \nu_1 - \nu_2$ 且 D 为满足 $\nu_1(D) = 0$ 和 $\nu_2(D^c) = 0$ 的集. 若令 $C = D^c$, 则 $\Omega = C \cup D$ 是 Hahn 分解, 由 D 的选取

$$\nu_1(E) = \alpha(C \cap E) \quad \text{且} \quad \nu_2(E) = -\alpha(D \cap E).$$

Hahn 分解的唯一性结果表明 $A\cap D = A\cap C^c$ 和 $B\cap C = A^c\cap C$ 是零集, 故 $\alpha(E\cap C) = \alpha(E\cap(A\cup C)) = \alpha(E\cap A)$ 且 $\nu_1 = \alpha_+$. □

习题 A.3.3. 证明 $\alpha_+(E) = \sup\{\alpha(F) : F \subset E\}$.

注释 A.3.7. 设 α 为 $(\mathbb{R}, \mathcal{R})$ 上的有限符号测度 (即不取值 ∞ 或 $-\infty$) . 令 $\alpha = \alpha_+ - \alpha_-$ 为其 Jordan 分解. 令 $A(x) = \alpha((-\infty, x])$, $F(x) = \alpha_+((-\infty, x])$, 和 $G(x) = \alpha_-((-\infty, x])$. $A(x) = F(x) - G(x)$ 故有限符号测度的分布函数可写为两个有界递增函数之差. 由例 A.4.2, 逆命题也成立. 令 $|\alpha| = \alpha_+ + \alpha_-$. $|\alpha|$ 称为 α 的全变差, 因在此例中 $|\alpha|((a, b])$ 是分析教材中定义的 A 在 $(a, b]$ 上的全变差. 参见例如 Royden (1988), 第 103 页. 排除区间的左端点因那里的跳跃对 $[a, b]$ 上的全变差无贡献, 但它确实出现在 $|\alpha|$ 中.

第三个也是最后一个分解是:

定理 A.3.8. Lebesgue 分解. 设 μ, ν 为 σ-有限测度. 则 ν 可写为 $\nu_r + \nu_s$, 其中 ν_s 关于 μ 奇异且

$$\nu_r(E) = \int_E g \,\mathrm{d}\mu.$$

证明. 通过分解 $\Omega = +_i\Omega_i$, 不妨设 μ 和 ν 为有限测度. 令 $\mathcal{G}$ 为满足对所有 E 有 $\int_E g \,\mathrm{d}\mu \le \nu(E)$ 的 $g \ge 0$ 的集合.

(a) 若 $g, h \in \mathcal{G}$ 则 $g \vee h \in \mathcal{G}$.

(a) 的证明. 令 $A = \{g > h\}$, $B = \{g \le h\}$.

$$\int_E g \vee h \,\mathrm{d}\mu = \int_{E\cap A} g \,\mathrm{d}\mu + \int_{E\cap B} h \,\mathrm{d}\mu \le \nu(E\cap A) + \nu(E\cap B) = \nu(E).$$

令 $\kappa = \sup\{\int g \,\mathrm{d}\mu : g \in \mathcal{G}\} \le \nu(\Omega) < \infty$. 选取 g_n 使得 $\int g_n \,\mathrm{d}\mu > \kappa - 1/n$, 并令 $h_n = g_1 \vee \cdots \vee g_n$. 由 (a), $h_n \in \mathcal{G}$. 当 $n \uparrow \infty$, $h_n \uparrow h$. 由 κ 的定义、单调收敛定理和 g_n 的选取,

$$\kappa \ge \int h \,\mathrm{d}\mu = \lim_{n\to\infty} \int h_n \,\mathrm{d}\mu \ge \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu = \kappa.$$

令 $\nu_r(E) = \int_E h \,\mathrm{d}\mu$ 和 $\nu_s(E) = \nu(E) - \nu_r(E)$. 最后要证明的是:

(b) ν_s 关于 μ 奇异.

(b) 的证明. 令 $\varepsilon > 0$ 且 $\Omega = A_\varepsilon \cup B_\varepsilon$ 为 $\nu_s - \varepsilon\mu$ 的 Hahn 分解. 利用 ν_r 的定义和 A_ε 对 $\nu_s - \varepsilon\mu$ 为正的事实 (故 $\varepsilon\mu(A_\varepsilon \cap E) \le \nu_s(A_\varepsilon \cap E)$) ,

$$\int_E (h + \varepsilon 1_{A_\varepsilon}) \,\mathrm{d}\mu = \nu_r(E) + \varepsilon\mu(A_\varepsilon \cap E) \le \nu(E).$$

这对所有 E 成立, 故 $k = h + \varepsilon 1_{A_\varepsilon} \in \mathcal{G}$. 由此得 $\mu(A_\varepsilon) = 0$, 因若不然则 $\int k\,\mathrm{d}\mu > \kappa$, 矛盾. 令 $A = \cup_n A_{1/n}$, 有 $\mu(A) = 0$. 为见 $\nu_s(A^c) = 0$, 注意到若 $\nu_s(A^c) > 0$, 则对小 ε 有 $(\nu_s - \varepsilon\mu)(A^c) > 0$, 矛盾因 $A^c \subset B_\varepsilon$ 为负集. □

习题 A.3.4. 证明 Lebesgue 分解是唯一的. 注意可不妨设 μ 和 ν 为有限.

终于准备好本节的主题. 若 $\mu(A) = 0$ 则 $\nu(A) = 0$, 则称测度 ν 关于 μ 绝对连续（记作 $\nu \ll \mu$）.

习题 A.3.5. 若 $\mu_1 \ll \mu_2$ 且 $\mu_2 \perp \nu$ 则 $\mu_1 \perp \nu$.

定理 A.3.9. Radon-Nikodym 定理. 若 μ, ν 为 σ-有限测度且 ν 关于 μ 绝对连续, 则存在 $g \geq 0$ 使得 $\nu(E) = \int_E g\,\mathrm{d}\mu$. 若 h 是另一这样的函数则 $g = h$ μ a.e.

证明. 令 $\nu = \nu_r + \nu_s$ 为任意 Lebesgue 分解. 选取 A 使得 $\nu_s(A^c) = 0$ 和 $\mu(A) = 0$. 因 $\nu \ll \mu$, $0 = \nu(A) \geq \nu_s(A)$ 故 $\nu_s \equiv 0$. 为证唯一性, 注意到若对所有 E 有 $\int_E g\,\mathrm{d}\mu = \int_E h\,\mathrm{d}\mu$, 则令 $E \subset \{g > h, g \leq n\}$ 为任意有限测度子集, 得 $\mu(g > h, g \leq n) = 0$ 对所有 n, 故 $\mu(g > h) = 0$, 类似地 $\mu(g < h) = 0$. □

例题 A.3.10. 若 μ 不 σ-有限则定理 A.4.8 可能失效. 令 $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{R})$, $\mu =$ 计数测度且 $\nu =$Lebesgue 测度.

定理 A.4.8 中证明存在的函数 g 常记为 $\mathrm{d}\nu/\mathrm{d}\mu$. 这一记号暗示以下性质, 其证明留给读者.

习题 A.3.6. 若 $\nu_1, \nu_2 \ll \mu$ 则 $\nu_1 + \nu_2 \ll \mu$ 且

$$\mathrm{d}(\nu_1 + \nu_2)/\mathrm{d}\mu = \mathrm{d}\nu_1/\mathrm{d}\mu + \mathrm{d}\nu_2/\mathrm{d}\mu.$$

习题 A.3.7. 若 $\nu \ll \mu$ 且 $f \geq 0$ 则

$$\int f\,\mathrm{d}\nu = \int f \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\mu.$$

习题 A.3.8. 若 $\pi \ll \nu \ll \mu$ 则 $\mathrm{d}\pi/\mathrm{d}\mu = (\mathrm{d}\pi/\mathrm{d}\nu) \cdot (\mathrm{d}\nu/\mathrm{d}\mu)$.

习题 A.3.9. 若 $\nu \ll \mu$ 且 $\mu \ll \nu$ 则 $\mathrm{d}\mu/\mathrm{d}\nu = (\mathrm{d}\nu/\mathrm{d}\mu)^{-1}$.

第 4 节　积分号下求导

正文中多处需要交换求和或积分内的微分. 本节给出可用于证明这些计算合理性的结果.

定理 A.4.1. 设 $(S,\mathcal{S},\mu)$ 为测度空间. 设 f 为 $\mathbb{R}\times S$ 上定义的复值函数. 设 $\delta>0$, 且假设对 $x\in(y-\delta,y+\delta)$ 有

(i) $u(x)=\int_S f(x,s)\,\mu(\mathrm{d}s)$ 满足 $\int_S|f(x,s)|\,\mu(\mathrm{d}s)<\infty$;
(ii) 对固定的 s, $\partial f/\partial x(x,s)$ 存在且是 x 的连续函数;
(iii) $v(x)=\int_S\frac{\partial f}{\partial x}(x,s)\,\mu(\mathrm{d}s)$ 在 $x=y$ 连续;
(iv) $\int_S\int_{-\delta}^{\delta}\left|\frac{\partial f}{\partial x}(y+\theta,s)\right|\,\mathrm{d}\theta\,\mu(\mathrm{d}s)<\infty$,

则 $u'(y)=v(y)$.

证明. 令 $|h|\le\delta$ 并利用 (i)、(ii)、(iv) 及习题 1.7.4 给出的 Fubini 定理形式, 有

$$\begin{aligned}u(y+h)-u(y)&=\int_S f(y+h,s)-f(y,s)\,\mu(\mathrm{d}s)\\&=\int_S\int_0^h\frac{\partial f}{\partial x}(y+\theta,s)\,\mathrm{d}\theta\,\mu(\mathrm{d}s)\\&=\int_0^h\int_S\frac{\partial f}{\partial x}(y+\theta,s)\,\mu(\mathrm{d}s)\,\mathrm{d}\theta.\end{aligned}$$

最后一个等式蕴含

$$\frac{u(y+h)-u(y)}{h}=\frac{1}{h}\int_0^h v(y+\theta)\,\mathrm{d}\theta.$$

因 v 在 y 连续由 (iii), 令 $h\to0$ 即得所需结果. □

例题 A.4.2. 对于第 3.3 节中的一个结果, 需要知道可以在

$$u(x)=\int\cos(xs)e^{-s^2/2}\,\mathrm{d}s$$

中在积分号下求导. 为方便, 省略了因子 $(2\pi)^{-1/2}$ 并作了变量替换以匹配定理 A.5.1. 显然, (i) 和 (ii) 成立. 控制收敛定理蕴含 (iii)

$$x\mapsto\int-s\sin(sx)e^{-s^2/2}\,\mathrm{d}s$$

连续. 对于 (iv), 注意到

$$\int\left|\frac{\partial f}{\partial x}(x,s)\right|\,\mathrm{d}s=\int|s|e^{-s^2/2}\,\mathrm{d}s<\infty$$

且值不依赖于 x, 故 (iv) 成立.

对某些例子, 以下形式更方便:

定理 A.4.3. 设 $(S,\mathcal{S},\mu)$ 为测度空间. 设 f 为 $\mathbb{R}\times S$ 上定义的复值函数. 设 $\delta>0$, 且假设对 $x\in(y-\delta,y+\delta)$ 有

(i) $u(x)=\int_S f(x,s)\,\mu(\mathrm{d}s)$ 满足 $\int_S|f(x,s)|\,\mu(\mathrm{d}s)<\infty$;

(ii) 对固定的 s, $\partial f/\partial x(x,s)$ 存在且是 x 的连续函数;

(iii)′ $\int_S \sup_{\theta\in[-\delta,\delta]}\left|\frac{\partial f}{\partial x}(y+\theta,s)\right|\,\mu(\mathrm{d}s)<\infty$,

则 $u'(y)=v(y)$.

证明. 鉴于定理 A.5.1, 只需证明该结果的 (iii) 和 (iv) 成立. 因

$$\int_{-\delta}^{\delta}\left|\frac{\partial f}{\partial x}(y+\theta,s)\right|\,\mathrm{d}\theta\le 2\delta\sup_{\theta\in[-\delta,\delta]}\left|\frac{\partial f}{\partial x}(y+\theta,s)\right|,$$

显然 (iv) 成立. 为验证 (iii), 注意到

$$|v(x)-v(y)|\le\int_S\left|\frac{\partial f}{\partial x}(x,s)-\frac{\partial f}{\partial x}(y,s)\right|\,\mu(\mathrm{d}s).$$

(ii) 蕴含被积函数当 $x\to y$ 时 $\to 0$. 由 (iii)′ 和控制收敛定理即得所需结果. □

为表明新结果的用途, 下面证明:

例题 A.4.4. 若 $\phi(\theta)=Ee^{\theta Z}<\infty$ 对 $\theta\in[-\varepsilon,\varepsilon]$ 成立, 则 $\phi'(0)=EZ$.

证明. 这里 θ 扮演 x 的角色, 取 μ 为 Z 的分布. 令 $\delta=\varepsilon/2$. $f(x,s)=e^{xs}\ge 0$, 故 (i) 由假设成立. $\partial f/\partial x=se^{xs}$ 显然是连续函数, 故 (ii) 成立. 为验证 (iii)′, 注意到存在常数 C 使得若 $x\in(-\delta,\delta)$, 则 $|s|e^{xs}\le C(e^{-\varepsilon s}+e^{\varepsilon s})$. □

在定理 A.5.3 中取 $S=\mathbb{Z}$, $\mathcal{S}=S$ 的所有子集, 和 $\mu=$ 计数测度, 得:

定理 A.4.5. 设 $\delta>0$. 假设对 $x\in(y-\delta,y+\delta)$ 有

(i) $u(x)=\sum_{n=1}^{\infty}f_n(x)$ 满足 $\sum_{n=1}^{\infty}|f_n(x)|<\infty$;

(ii) 对每个 n, $f_n'(x)$ 存在且是 x 的连续函数;

(iii)′ $\sum_{n=1}^{\infty}\sup_{\theta\in(-\delta,\delta)}|f_n'(y+\theta)|<\infty$,

则 $u'(x)=v(x)$.

例题 A.4.6. 在第 2.6 节中想证明若 $p\in(0,1)$ 则

$$\left(\sum_{n=1}^{\infty}(1-p)^n\right)'=-\sum_{n=1}^{\infty}n(1-p)^{n-1}.$$

证明. 令 $f_n(x) = (1-x)^n$, $y = p$, 并选取 δ 使得 $[y-\delta, y+\delta] \subset (0,1)$. 显然 (i) $\sum_{n=1}^{\infty} |(1-x)^n| < \infty$ 和 (ii) $f_n'(x) = n(1-x)^{n-1}$ 在 $[y-\delta, y+\delta]$ 中对 x 连续. 为验证 (iii), 注意到若令 $2\eta = y - \delta$, 则存在常数 C 使得若 $x \in [y-\delta, y+\delta]$ 且 $n \geq 1$, 则

$$n(1-x)^{n-1} = \frac{n(1-x)^{n-1}}{(1-\eta)^{n-1}} \cdot (1-\eta)^{n-1} \leq C(1-\eta)^{n-1}. \qquad \square$$

附录 B
正态分布表

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\,\mathrm{d}y$$

下述例子说明本表的用法：$\Phi(0.36)=0.6406, \Phi(1.34)=0.9099.$

	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

参考文献

[1] Aldous, D., & Diaconis, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly*, 93, 333–348.

[2] Andersen, E. S., & Jessen, B. (1984). On the introduction of measures in infinite product spaces. *Danske Vid. Selsk. Mat.-Fys. Medd.*, 25, No. 4.

[3] Athreya, K., & Ney, P. (1972). *Branching Processes*. Springer-Verlag, New York.

[4] Athreya, K., & Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. AMS*, 245, 493–501.

[5] Athreya, K., McDonald, D., & Ney, P. (1978). Coupling and the renewal theorem. *Amer. Math. Monthly*, 85, 809–814.

[6] Bachelier, L. (1900). Théorie de la spéculation. *Ann. Sci. École Norm. Sup.*, 17, 21–86.

[7] Banach, S., & Tarski, A. (1924). Sur la décomposition des ensembles de points en parties respectivements congruent. *Fund. Math.*, 6, 244–277.

[8] Baum, L. E., & Billingsley, P. (1966). Asymptotic distributions for the coupon collector's problem. *Ann. Math. Statist.*, 36, 1835–1839.

[9] Benford, F. (1938). The law of anomalous numbers. *Proc. Amer. Phil. Soc.*, 78, 552–572.

[10] Biggins, J. D. (1977). Chernoff's theorem in branching random walk. *J. Appl. Probab.*, 14, 630–636.

[11] Biggins, J. D. (1978). The asymptotic shape of branching random walk. *Adv. in Appl. Probab.*, 10, 62–84.

[12] Biggins, J. D. (1979). Growth rates in branching random walk. *Z. Warsch. verw. Gebiete*, 48, 17–34.

[13] Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proc. AMS*, 12, 788–792.

[14] Billingsley, P. (1979). *Probability and measure*. John Wiley and Sons, New York.

[15] Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17, 656–660.

[16] Blackwell, D., & Freedman, D. (1964). The tail σ-field of a Markov chain and a theorem of Orey. *Ann. Math. Statist.*, 35, 1291–1295.

[17] Blumenthal, R. M., & Getoor, R. K. (1968). *Markov processes and their potential theory*. Academic Press, New York.

[18] Borel, E. (1909). Les probabilités dénombrables et leur applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27, 247–271.

[19] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

[20] Chung, K. L. (1974). *A Course in Probability Theory*, second edition. Academic Press, New York.

[21] Chung, K. L., Erdős, P., & Sirao, T. (1959). On the Lipschitz's condition for Brownian motion. *J. Math. Soc. Japan*, 11, 263–274.

[22] Chung, K. L., & Fuchs, W. H. J. (1951). On the distribution of values of sums of independent random variables. *Memoirs of the AMS*, No. 6.

[23] Chung, K. L., & Zhao, Z. (1995). *From Brownian Motion to Schrödinger's Equation*. Springer, New York.

[24] Chvátal, V., & Sankoff, D. (1975). Longest common subsequences of two random sequences. *J. Appl. Probab.*, 12, 306–315.

[25] Cohen, J., Kesten, H., & Newman, C. (1985). Random matrices and their applications. *AMS Contemporary Math.*, 50, Providence, RI.

[26] Cox, J. T., & Durrett, R. (1981). Limit theorems for percolation processes with necessary and sufficient conditions. *Ann. Probab.*, 9, 583–603.

[27] Davis, B. (1983). On Brownian slow points. *Z. Warsch. verw. Gebiete*, 64, 359–367.

[28] Diaconis, P., & Freedman, D. (1980). Finite exchangeable sequences. *Ann. Prob.*, 8, 745–764.

[29] Dieudonné, J. (1948). Sur la théorème de Lebesgue-Nikodym, II. *Ann. Univ. Grenoble*, 23, 25–53.

[30] Donsker, M. (1951). An invariance principle for certain probability limit theorems. *Memoirs of the AMS*, No. 6.

[31] Donsker, M. (1952). Justification and extension of Doob's heurisitc approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.*, 23, 277–281.

[32] Doob, J. L. (1949). A heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.*, 20, 393–403.

[33] Doob, J. L. (1953). *Stochastic Processes*. John Wiley and Sons, New York.

[34] Dubins, L. E. (1968). On a theorem of Skorkhod. *Ann. Math. Statist.*, 39, 2094–2097.

[35] Dubins, L. E., & Freedman, D. A. (1965). A sharper form of the Borel-Cantelli lemma and the strong law. *Ann. Math. Statist.*, 36, 800–807.

[36] Dubins, L. E., & Freedman, D. A. (1979). Exchangeable processes need not be distributed mixtures of independent and identically distributed random variables. *Z. Warsch. verw. Gebiete*, 48, 115–132.

[37] Dudley, R. M. (1989). *Real Analysis and Probability*. Wadsworth Pub. Co., Pacific Grove, CA.

[38] Durrett, R., & Resnick, S. (1978). Functional limit theorems for dependent random variables. *Ann. Probab.*, 6, 829–846.

[39] Dvoretsky, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proc. 6th Berkeley Symp.*, Vol. II, 513–535.

[40] Dvoretsky, A., & Erdős, P. (1951). Some problems on random walk in space. In *Proc. 2nd Berkeley Symp.*, 353–367.

[41] Dvoretsky, A., Erdős, P., & Kakutani, S. (1961). Nonincrease everywhere of the Brownian motion process. In *Proc. 4th Berkeley Symp.*, Vol. II, 103–116.

[42] Dynkin, E. B. (1965). *Markov processes*. Springer-Verlag, New York.

[43] Dynkin, E. B., & Yushkevich, A. A. (1956). Strong Markov processes. *Theory of Probability and its Applications*, 1, 134–139.

[44] Erdős, P., & Kac, M. (1946). On certain limit theorems of the theory of probability. *Bull. AMS*, 52, 292–302.

[45] Erdős, P., & Kac, M. (1947). On the number of positive sums of independent random variables. *Bull. AMS*, 53, 1011–1020.

[46] Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Z. Warsch. verw. Gebiete*, 55, 119–122.

[47] Ethier, S. N., & Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley and Sons.

[48] Faden, A. M. (1985). The existence of regular conditional probabilities: necessary and sufficient conditions. *Ann. Probab.*, 13, 288–298.

[49] Feller, W. (1946). A limit theorem for random variables with infinite moments. *Amer. J. Math.*, 68, 257–262.

[50] Feller, W. (1961). A simple proof of renewal theorems. *Comm. Pure Appl. Math.*, 14, 285–293.

[51] Feller, W. (1968). *An introduction to probability theory and its applications*, Vol. I, third edition. John Wiley and Sons, New York.

[52] Feller, W. (1971). *An introduction to probability theory and its applications*, Vol. II, second edition. John Wiley and Sons, New York.

[53] Freedman, D. (1965). Bernard Friedman's urn. *Ann. Math. Statist.*, 36, 956–970.

[54] Freedman, D. (1971a). *Brownian motion and diffusion*. Originally published by Holden Day, San Francisco, CA. Second edition by Springer-Verlag, New York.

[55] Freedman, D. (1971b). *Markov chains*. Originally published by Holden Day, San Francisco, CA. Second edition by Springer-Verlag, New York.

[56] Freedman, D. (1980). A mixture of independent and identically distributed random variables need not admit a regular conditional probability given the exchangeable σ-field. *Z. Warsch. verw. Gebiete*, 51, 239–248.

[57] French, R. M. (1988). The Banach-Tarski theorem. *Math. Intelligencer*, 10, No. 4, 21–28.

[58] Friedman, B. (1949). A simple urn model. *Comm. Pure Appl. Math.*, 2, 59–70.

[59] Furstenberg, H. (1970). Random walks in discrete subgroups of Lie Groups. In *Advances in Probability* edited by P. E. Ney, Marcel Dekker, New York.

[60] Furstenberg, H., & Kesten, H. (1960). Products of random matrices. *Ann. Math. Statist.*, 31, 451–469.

[61] Garsia, A. (1965). A simple proof of E. Hopf's maximal ergodic theorem. *J. Math. Mech.*, 14, 381–382.

[62] Glasser, M. L., & Zucker, I. J. (1977). Extended Watson integrals for the cubic lattice. *Proc. Nat. Acad. Sci.*, 74, 1800–1801.

[63] Gnedenko, B. V. (1943). Sur la distribution limité du terme maximum d'une série aléatoire. *Ann. Math.*, 44, 423–453.

[64] Gnedenko, B. V., & Kolmogorov, A. V. (1954). *Limit distributions for sums of independent random variables*. Addison-Wesley, Reading, MA.

[65] Gordin, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Doklady*, 10, 1174–1176.

[66] Hall, P. (1982). *Rates of convergence in the central limit theorem*. Pitman Pub. Co., Boston, MA.

[67] Hall, P., & Heyde, C. C. (1976). On a unified approach to the law of the iterated logarithm for martingales. *Bull. Austral. Math. Soc.*, 14, 435–447.

[68] Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press, New York.

[69] Halmos, P. R. (1950). *Measure theory*. Van Nostrand, New York.

[70] Hammersley, J. M. (1970). A few seedlings of research. In *Proc. 6th Berkeley Symp.*, Vol. I, 345–394.

[71] Hardy, G. H., & Littlewood, J. E. (1914). Some problems of Diophantine approximation. *Acta Math.*, 37, 155–239.

[72] Hardy, G. H., & Wright, E. M. (1959). *An introduction to the theory of numbers*, fourth edition. Oxford University Press, London.

[73] Harris, T. E. (1956). The existence of stationary measures for certain Markov processes. In *Proc. 3rd Berkeley Symp.*, Vol. II, 113–124.

[74] Hartman, P., & Wintner, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.*, 63, 169–176.

[75] Hewitt, E., & Savage, L. J. (1956). Symmetric measures on Cartesian products. *Trans. AMS*, 80, 470–501.

[76] Hewitt, E., & Stromberg, K. (1965). *Real and abstract analysis*. Springer-Verlag, New York.

[77] Heyde, C. C. (1963). On a property of the lognormal distribution. *J. Royal. Stat. Soc. B*, 29, 392–393.

[78] Heyde, C. C. (1967). On the influence of moments on the rate of convergence to the normal distribution. *Z. Warsch. verw. Gebiete*, 8, 12–18.

[79] Hodges, J. L., Jr., & Le Cam, L. (1960). The Poisson approximation to the binomial distribution. *Ann. Math. Statist.*, 31, 737–740.

[80] Hunt, G. (1956). Some theorems concerning Brownian motion. *Trans. AMS*, 81, 294–319.

[81] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.*, 7, 349–382.

[82] Ibragimov, I. A. (1963). A central limit theorem for a class of dependent random variables. *Theory Probab. Appl.*, 8, 83–89.

[83] Ibragimov, I. A., & Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen.

[84] Ishitani, H. (1977). A central limit theorem for the subadditive process and its application to products of random matrices. *RIMS, Kyoto*, 12, 565–575.

[85] Jacod, J., & Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Second edition. Springer-Verlag, Berlin.

[86] Itô, K., & McKean, H. P. (1965). *Diffusion processes and their sample paths*. Springer-Verlag, New York.

[87] Jacod, J., & Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, New York.

[88] Kac, M. (1947a). Random walk and the theory of Brownian motion. *Amer. Math. Monthly*, 54, 369–391.

[89] Kac, M. (1947b). On the notion of recurrence in discrete stochastic processes. *Bull. AMS*, 53, 1002–1010.

[90] Kac, M. (1959). *Statistical independence in probability, analysis, and number theory*. Carus Monographs, Math. Assoc. of America.

[91] Katznelson, Y., & Weiss, B. (1982). A simple proof of some ergodic theorems. *Israel J. Math.*, 42, 291–296.

[92] Keeler, E., & Spencer, J. (1975). Optimal doubling in backgammon. *Operations Research*, 23, 1063–1071.

[93] Kesten, H. (1986). Aspects of first passage percolation. In *École d'été de probabilités de Saint-Flour XIV*. Lecture Notes in Math 1180, Springer-Verlag, New York.

[94] Kesten, H. (1987). Percolation theory and first passage percolation. *Ann. Probab.*, 15, 1231–1271.

[95] Kingman, J. F. C. (1968). The ergodic theory of subadditive processes. *J. Roy. Stat. Soc. B*, 30, 499–510.

[96] Kingman, J. F. C. (1973). Subadditive ergodic theory. *Ann. Probab.*, 1, 883–909.

[97] Kingman, J. F. C. (1975). The first birth problem for age dependent branching processes. *Ann. Probab.*, 3, 790–801.

[98] Kolmogorov, A. N., & Rozanov, Y. A. (1964). On strong mixing conditions for stationary Gaussian processes. *Theory Probab. Appl.*, 5, 204–208.

[99] Kondo, K., & Hara, T. (1987). Critical exponent of susceptibility for a general class of ferromagnets in $d > 4$ dimensions. *J. Math. Phys.*, 28, 1206–1208.

[100] Krengel, U. (1985). *Ergodic theorems*. deGruyter, New York.

[101] Leventhal, S. (1988). A proof of Liggett's version of the subadditive ergodic theorem. *Proc. AMS*, 102, 169–173.

[102] Lévy, P. (1937). *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris.

[103] Liggett, T. M. (1985). An improved subadditive ergodic theorem. *Ann. Probab.*, 13, 1279–1285.

[104] Liggett, T. M. (2010). *Continuous Time Markov Processes. An Introduction*. American Mathematical Society, Providence, RI.

[105] Lindenbaum, A. (1926). Contributions à l'étude de l'espace metrique. *Fund. Math.*, 8, 209–222.

[106] Lindvall, T. (1977). A probabilistic proof of Blackwell's renewal theorem. *Ann. Probab.*, 5, 482–485.

[107] Logan, B. F., & Shepp, L. A. (1977). A variational problem for random Young tableaux. *Adv. in Math.*, 26, 206–222.

[108] McKean, H. P. (1969). *Stochastic integrals*. Academic Press, New York.

[109] Motoo, M. (1959). Proof of the law of the iterated logarithm through diffusion equation. *Ann. Inst. Stat. Math.*, 10, 21–28.

[110] Neveu, J. (1965). *Mathematical foundations of the calculus of probabilities*. Holden-Day, San Francisco, CA.

[111] Neveu, J. (1975). *Discrete parameter martingales*. North Holland, Amsterdam.

[112] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.*, 4, 39–40.

[113] Ornstein, D. (1969). Random walks. *Trans. AMS*, 138, 1–60.

[114] Oseledeč, V. I. (1968). A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19, 197–231.

[115] Paley, R. E. A. C., Wiener, N., & Zygmund, A. (1933). Notes on random functions. *Math. Z.*, 37, 647–668.

[116] Pitman, E. J. G. (1956). On derivatives of characteristic functions at the origin. *Ann. Math. Statist.*, 27, 1156–1160.

[117] Perkins, E. (1983). On the Hausdorff dimension of the Brownian slow points. *Z. Warsch. verw. Gebiete*, 64, 369–399.

[118] Port, S. C., & Stone, C. J. (1969). Potential theory of random walks on abelian groups. *Acta Math.*, 122, 19–114.

[119] Port, S. C., & Stone, C. J. (1978). *Brownian Motion and Classical Potential Theory*. Academic Press, New York.

[120] Ragunathan, M. S. (1979). A proof of Oseledeč's multiplicative ergodic theorem. *Israel J. Math.*, 32, 356–362.

[121] Raimi, R. (1976). The first digit problem. *Amer. Math. Monthly*, 83, 521–538.

[122] Resnick, S. (1987). *Extreme values, regular variation, and point processes*. Springer-Verlag, New York.

[123] Revuz, D. (1984). *Markov Chains*, second edition. North Holland, Amsterdam.

[124] Root, D. H. (1969). The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40, 715–718.

[125] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, 42, 43–47.

[126] Royden, H. (1988). *Real Analysis*, third edition. McMillan, New York.

[127] Ruelle, D. (1979). Ergodic theory of differentiable dynamical systems. *IHES Pub. Math.*, 50, 275–306.

[128] Ryll-Nardzewski, C. (1951). On the ergodic theorems, II. *Studia Math.*, 12, 74–79.

[129] Savage, L. J. (1972). *The foundations of statistics*, second edition. Dover, New York.

[130] Scott, D. J. (1973). Central limit theorems for martingales and for processes with stationary independent increments using a Skorokhod representation approach. *Adv. Appl. Probab.*, 5, 119–137.

[131] Shepp, L. A. (1964). Recurrent random walks may take arbitrarily large steps. *Bull. AMS*, 70, 540–542.

[132] Sheu, S. (1986). Representing a distribution by stopping Brownian motion: Root's construction. *Bull. Austral. Math. Soc.*, 34, 427–431.

[133] Smirnov, N. V. (1949). Limit distributions for the terms of a variational series. *AMS Transl. Series*, 1, No. 67.

[134] Smythe, R., & Wierman, J. C. (1978). *First passage percolation on the square lattice*. Lecture Notes in Math 671, Springer-Verlag, New York.

[135] Solovay, R. M. (1970). A model of set theory in which every set of reals is Lebesgue measurable. *Ann. Math.*, 92, 1–56.

[136] Spitzer, F. (1964). *Principles of random walk*. Van Nostrand, Princeton, NJ.

[137] Stein, C. (1987). *Approximate computation of expectations*. IMS Lecture Notes Vol. 7.

[138] Stone, C. J. (1969). On the potential operator for one dimensional recurrent random walks. *Trans. AMS*, 136, 427–445.

[139] Stoyanov, J. (1987). *Counterexamples in probability*. John Wiley and Sons, New York.

[140] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Warsch. verw. Gebiete*, 3, 211–226.

[141] Strassen, V. (1965). A converse to the law of the iterated logarithm. *Z. Warsch. verw. Gebiete*, 4, 265–268.

[142] Strassen, V. (1967). Almost sure behavior of the sums of independent random variables and martingales. In *Proc. 5th Berkeley Symp.*, Vol. II, 315–343.

[143] Thorisson, H. (1987). A complete coupling proof of Blackwell's renewal theorem. *Stoch. Proc. Appl.*, 26, 87–97.

[144] van Beek, P. (1972). An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Z. Warsch. verw. Gebiete*, 23, 187–196.

[145] Vershik, A. M., & Kerov, S. V. (1977). Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of random Young tableau. *Dokl. Akad. Nauk SSR*, 233, 1024–1027.

[146] Wegner, H. (1973). On consistency of probability measures. *Z. Warsch. verw. Gebiete*, 27, 335–338.

[147] Weiss, L. (1955). The stochastic convergence of a function of sample successive differences. *Ann. Math. Statist.*, 26, 532–536.

[148] Wiener, N. (1923). Differential space. *J. Math. Phys.*, 2, 131–174.

[149] Yosida, K., & Kakutani, S. (1939). Birkhoff's ergodic theorem and the maximal ergodic theorem. *Proc. Imp. Acad. Tokyo*, 15, 165–168.

术语表

外文术语	中文术语
0-1 law	0-1 律.
a.s. convergence	几乎必然收敛.
absolutely continuous	绝对连续的.
absorbing state	吸收态.
adapted	适应的.
adapted process	适应过程.
adapted to	适应于.
additive functional	加性泛函.
age	年龄.
age (at time t)	在时刻 t 的年龄.
age-dependent branching process	年龄依赖分支过程.
algebra	代数.
algebra generated by	由... 生成的代数.
almost everywhere	几乎处处.
almost invariant	几乎不变的.
almost sure convergence	几乎必然收敛.
almost surely	几乎必然.
alphabet	字母表.
alternating renewal process	交替更新过程.
aperiodic	非周期的.
aperiodic chain	非周期链.
aperiodic Harris chain	非周期 Harris 链.
aperiodicity	非周期性.
appendix	附录.
arbitrary	任意的.

外文术语	中文术语
Arcsine law	反正弦律.
arcsine law	反正弦律.
arcsine laws	反正弦律.
arithmetic distribution	算术分布.
Arzelà-Ascoli theorem	Arzelà-Ascoli 定理.
assign	指定.
associated with	与... 关联的.
assume	假设.
asymmetric simple random walk	非对称简单随机游走.
asymptotic behavior	渐近行为.
Asymptotic equipartition property	渐近等分性.
asymptotic property	渐近性质.
atom	原子.
augmented chain	扩充链.
averaging property	均值性质.
axiom of choice	选择公理.
backward martingale	向后鞅.
ball drawn	抽出的球.
ballot	投票.
Ballot Theorem	投票定理.
ballot theorem	投票定理.
Banach-Tarski theorem	Banach-Tarski 定理.
basic	基本的.
Bayes' formula	Bayes 公式.
Benford's law	Benford 定律.
Bernard Friedman's urn	Bernard Friedman 罐子模型.
Bernoulli distribution	伯努利分布.
Bernstein polynomial	Bernstein 多项式.
Berry-Esseen theorem	Berry-Esseen 定理.
best guess	最佳猜测.
bet	赌注.
beta distribution	贝塔分布.
Biggins	Biggins.
binary expansion	二进制展开.
binomial coefficient	二项式系数.
Birkhoff's ergodic theorem	Birkhoff 遍历定理.
birth and death chain	生灭链.
birth time	出生时间.

外文术语	中文术语
Blackwell's renewal theorem	Blackwell 更新定理.
Blackwell and Freedman	Blackwell 与 Freedman.
Blumenthal's 0-1 law	Blumenthal 0-1 律.
Blumenthal 0-1 law	Blumenthal 0-1 律.
boldface	粗体.
Bonferroni inequalities	Bonferroni 不等式.
Borel's paradox	Borel 悖论.
Borel measure	Borel 测度.
Borel set	Borel 集.
Borel sets	Borel 集.
Borel subsets	Borel 子集.
Borel-Cantelli lemma	Borel-Cantelli 引理.
boundary	边界.
boundary condition	边界条件.
boundary value	边值.
bounded	有界的.
bounded and continuous	有界连续的.
bounded convergence theorem	有界收敛定理.
bounded increments	有界增量.
bounded martingale	有界鞅.
bounded solution	有界解.
bounded stopping time	有界停时.
bounded variation	有界变差.
branching process	分支过程.
branching processes	分支过程.
Brownian bridge	布朗桥.
Brownian motion	布朗运动.
Brownian motion with drift	带漂移的布朗运动.
Brownian scaling	布朗尺度变换.
burn out	烧坏.
C[0,1]	C[0,1] 空间.
C[0,∞)	C[0,∞) 空间.
canonical measure	典范测度.
canonical probability space	典则概率空间.
Cantor distribution	Cantor 分布.
Cantor set	Cantor 集.
capacity	容量.
Carathéodory's extension theorem	Carathéodory 扩张定理.

外文术语	中文术语
Cauchy distribution	柯西分布.
Cauchy sequence	Cauchy 序列.
Cauchy-Schwarz inequality	Cauchy-Schwarz 不等式.
centering constant	中心化常数.
central limit theorem	中心极限定理.
Cesaro sense	Cesaro 意义.
change of variables formula	变量替换公式.
Chapman-Kolmogorov equation	Chapman-Kolmogorov 方程.
chapter	章.
characteristic function	特征函数.
Chebyshev's inequality	Chebyshev 不等式.
chi-square distribution	卡方分布.
chi-squared distribution	卡方分布.
Chung-Fuchs	Chung-Fuchs.
Chung-Fuchs theorem	Chung-Fuchs 定理.
Chvatal	Chvatal.
claims	索赔.
class property	类性质.
classical solution	经典解.
closed interval	闭区间.
closed set	闭集.
closed set of recurrent states	常返态闭集.
closed subspace	闭子空间.
closed under	对... 封闭.
closed under countable intersections	对可数交封闭.
closed under intersection	对交封闭.
coefficient	系数.
coefficients	系数.
collection	集合族.
collection of random variables	随机变量族.
combinatorial	组合的.
combinatorial arguments	组合论证.
Combinatorial Calculations	组合计算.
Combinatorics	组合学.
combinatorics	组合学.
compact support	紧支集.
complement	补集.
complete separable metric space	完备可分度量空间.

外文术语	中文术语
complete space	完备空间.
completed σ-field	完备化 σ-域.
completed σ-fields	完备化 σ-域.
completion	完备化.
complex conjugate	复共轭.
compound Poisson distribution	复合 Poisson 分布.
compound Poisson process	复合 Poisson 过程.
concave	凹的.
concept	概念.
Conditional Expectation	条件期望.
conditional expectation	条件期望.
conditional probability density	条件概率密度.
conditional variance	条件方差.
Conditional variance formula	条件方差公式.
conditioning	条件化.
cone	锥.
Cone Condition	锥条件.
congruent (sets)	全等的（集合）.
consequence	推论.
consistent	相容的.
consistent probability measures	相容的概率测度.
Constructing Independent Random Variables	构造独立随机变量.
continued fraction	连分数.
continued fractions	连分数.
continuity from above	上连续性.
continuity from below	下连续性.
continuity point	连续点.
continuous	连续的.
continuous density	连续密度.
continuous distribution	连续分布.
continuous function space	连续函数空间.
continuous mapping theorem	连续映射定理.
continuous martingale	连续鞅.
continuous path	连续路径.
continuous paths	连续路径.
continuum hypothesis	连续统假设.
contraction	压缩.
contraction in Lp	Lp 压缩.

外文术语	中文术语
converge almost surely	几乎必然收敛.
converge in distribution	依分布收敛.
converge in probability	依概率收敛.
converge weakly	弱收敛.
convergence almost surely	几乎必然收敛.
convergence in L1	L1 收敛.
convergence in Lp	Lp 收敛.
convergence in probability	依概率收敛.
convergence of random series	随机级数的收敛.
convergence of types theorem	收敛类型定理.
convergence theorem	收敛定理.
convergent subsequence	收敛子列.
converging together lemma	收敛合并引理.
converse	逆命题.
convex	凸的.
convex function	凸函数.
convolution	卷积.
coordinate maps	坐标映射.
countable	可数的.
countable dense set	可数稠密集.
countable intersection	可数交.
countable sequence	可数序列.
countable state space	可数状态空间.
countable union	可数并.
countably additive	可数可加的.
countably infinite	可数无限的.
counterexample	反例.
counterexamples	反例.
counting measure	计数测度.
coupled	耦合的.
coupling	耦合.
coupling time	耦合时间.
coupon collector’s problem	优惠券收集问题.
covariance matrix	协方差矩阵.
covariation	协变差.
Cramér’s estimates of ruin	Cramér 破产估计.
Cramér’s theorem	Cramér 定理.
Cramér-Wold device	Cramér-Wold 方法.

外文术语	中文术语
cross-section	截面.
cycle	循环.
cycle structure	循环结构.
cycle trick	循环技巧.
cyclic class	循环类.
cyclic decomposition	循环分解.
d-dimensional Brownian motion	d 维布朗运动.
De Moivre-Laplace theorem	De Moivre-Laplace 定理.
decomposition theorem	分解定理.
decreasing in m	关于 m 递减.
decreasing limit	递减极限.
decreasing sequence	递减序列.
definition	定义.
degenerate	退化的.
degenerate (distribution)	退化（分布）.
delay distribution	延迟分布.
delayed renewal process	延迟更新过程.
denote	记为.
dense discontinuities	稠密不连续点.
dense set	稠密集.
density function	密度函数.
dependence between σ-fields	σ-域之间的依赖性.
detailed balance	细致平衡.
detailed balance condition	细致平衡条件.
diagonal	对角线.
difference	差集.
differentiating under the integral	积分号下求导.
diffusion process	扩散过程.
directly Riemann integrable	直接黎曼可积.
Dirichlet problem	Dirichlet 问题.
discontinuity	不连续点.
discrete	离散的.
discrete Ornstein-Uhlenbeck process	离散 Ornstein-Uhlenbeck 过程.
discrete probability space	离散概率空间.
discrete probability spaces	离散概率空间.
discrete time	离散时间.
disjoint	不交的.
disjoint collections	不相交集合族.

外文术语	中文术语
disjoint sets	不交集合.
disjoint union	不交并.
distribution	分布.
distribution function	分布函数.
distribution function of the sum	和的分布函数.
Doeblin's condition	Doeblin 条件.
domain	定义域.
domain of attraction	吸引域.
dominated by	被... 控制.
dominated convergence	控制收敛.
dominated convergence theorem	控制收敛定理.
Donsker's theorem	Donsker 定理.
Donsker theorem	Donsker 定理.
Doob's decomposition	Doob 分解.
Doob's inequality	Doob 不等式.
Doob's martingale inequality	Doob 鞅不等式.
double exponential distribution	双指数分布.
doubly stochastic	双随机的.
drift	漂移.
dual transition probability	对偶转移概率.
Dubins' inequality	Dubins 不等式.
Dvoretsky's lemma	Dvoretsky 引理.
Dvoretsky-Erdős theorem	Dvoretsky-Erdős 定理.
dyadic rational	二进有理数.
dyadic rationals	二进有理数.
Dynkin's π-λ Theorem	Dynkin 的 π-λ 定理.
Ehrenfest chain	Ehrenfest 链.
electrostatic potential	静电位势.
elliptic equation	椭圆型方程.
embedding	嵌入.
embedding theorem	嵌入定理.
empirical distribution	经验分布.
Empirical distribution function	经验分布函数.
empirical distribution function	经验分布函数.
Empirical distribution functions	经验分布函数.
empty set	空集.
Entropy	熵.
entropy	熵.

外文术语	中文术语
equal in distribution	同分布的.
equicontinuous	等度连续.
equidistribution theorem	等分布定理.
equilibrium	平衡.
equilibrium temperature distribution	平衡温度分布.
Erdős-Kac theorem	Erdős-Kac 定理.
ergodic	遍历的.
ergodic stationary sequence	遍历平稳序列.
ergodic theorem	遍历定理.
Ergodic Theorems	遍历定理.
error term	误差项.
Etemadi	Etemadi.
Euclidean space	欧几里得空间.
event	事件.
events	事件.
example	例子.
exchangeable	可交换的.
exchangeable field	可交换域.
exchangeable σ-field	可交换 σ-域.
excursion	游程.
exit distribution	退出分布.
exit point	出口点.
exit time	退出时.
expectation	期望.
expected occupation time	期望占据时间.
expected value	期望值.
exponential Brownian motion	指数布朗运动.
exponential convergence	指数收敛.
exponential distribution	指数分布.
exponential martingale	指数鞅.
exponentially growing	指数增长的.
extended real line	扩充实数直线.
extension	延拓.
extension theorem	扩张定理.
extreme value distribution	极值分布.
F_sigma set	Fσ 集.
f.d.d.	有限维分布.
fair game	公平博弈.

外文术语	中文术语
Fatou’s lemma	Fatou 引理.
favorable game	有利博弈.
Feller condition	Feller 条件.
Feller property	Feller 性质.
Feynman-Kac formula	Feynman-Kac 公式.
Feynman-Kac theorem	Feynman-Kac 定理.
field	域.
filtration	σ-域流.
finite	有限的.
finite collection	有限集合.
finite dimensional distribution	有限维分布.
finite dimensional distributions	有限维分布.
finite dimensional sets	有限维集.
finite disjoint union	有限不交并.
finite measure	有限测度.
finite or countably infinite	有限或可数无限的.
finite permutation	有限置换.
finite rectangle	有限矩形.
finite subcollection	有限子集族.
finite variation	有限变差.
first derivative	一阶导数.
first digit	首位数字.
first entrance decomposition	首次进入分解.
first excursion	首次游程.
first exit time	首次退出时.
first passage percolation	首次通过渗流.
first return	首次返回.
first return time	首次返回时间.
fixation	固定.
fixed point (of permutation)	不动点.
F-measurable	F-可测的.
follow	推出.
Fourier analysis	Fourier 分析.
Fourier transform	Fourier 变换.
fraction of green balls	绿球比例.
fraction of time	时间比例.
Fubini’s Theorem	Fubini 定理.
Fubini’s theorem	Fubini 定理.

外文术语	中文术语
function	函数.
functional central limit theorem	泛函中心极限定理.
fundamental solution	基本解.
Furstenberg	Furstenberg.
Furstenberg-Kesten theorem	Furstenberg-Kesten 定理.
Fσ set	Fσ 集.
g.c.d.	最大公约数.
Galton-Watson process	Galton-Watson 过程.
gambler’s ruin	赌徒破产.
gambler	赌徒.
gambling system	赌博系统.
Gamma density	伽马密度.
gamma density	伽马密度.
Gamma distribution	伽马分布.
gamma distribution	伽马分布.
Gaussian process	高斯过程.
generalize	推广.
generated σ-field	生成的 σ-域.
generates	生成.
generating function	生成函数.
generator	生成元.
geometric Brownian motion	几何布朗运动.
geometric distribution	几何分布.
geometric interpretation	几何解释.
geometric series	几何级数.
germ field	芽 σ-域.
germ σ-field	芽 σ-域.
Glivenko-Cantelli theorem	Glivenko-Cantelli 定理.
great circle	大圆.
greatest common divisor	最大公约数.
Green’s function	Green 函数.
Green’s Functions	Green 函数.
green balls	绿球.
Gumbel distribution	Gumbel 分布.
Hahn decomposition	Hahn 分解.
half-line	半直线.
half-open interval	半开区间.
half-space	半空间.

外文术语	中文术语
Hamburger moment problem	Hamburger 矩问题.
Hammersley	Hammersley.
harmonic function	调和函数.
harmonic measure	调和测度.
Harris chain	Harris 链.
Hartman-Wintner theorem	Hartman-Wintner 定理.
Hausdorff dimension	Hausdorff 维数.
head runs	连续正面.
heat equation	热方程.
Helly's selection theorem	Helly 选择定理.
Hewitt-Savage 0-1 law	Hewitt-Savage 0-1 律.
higher order	高阶.
Hilbert space	Hilbert 空间.
hitting probability	击中概率.
hitting time	击中时.
hitting time of the diagonal	击中对角线的时间.
Hölder's inequality	Hölder 不等式.
Hölder continuity	Hölder 连续性.
Hölder continuous	Hölder 连续的.
Holtsmark distribution	Holtsmark 分布.
h-path	h-路径.
hypothesis	假设.
i.i.d.	独立同分布.
i.i.d. sequences	独立同分布序列.
i.o.	无穷多次.
identical parity	相同奇偶性.
identically distributed	同分布的.
imaginary part	虚部.
in distribution	依分布.
in measure	依测度.
in probability	依概率.
inclusion-exclusion formula	容斥公式.
increasing convex function	递增凸函数.
increasing in m	关于 m 递增.
increasing path	递增路径.
increasing process	增过程.
increasing sequence	递增序列.
increasing sequences in random permutations	随机排列中的递增序列.

外文术语	中文术语
Independence	独立性.
independent	独立的.
independent and identically distributed	独立同分布.
independent events	独立事件.
independent increments	独立增量.
independent random variables	独立随机变量.
independent σ-fields	独立的 σ-域.
index	指数.
index set	指标集.
index α	指数 α.
indicator function	指示函数.
indicator random variable	示性随机变量.
individual ergodic theorem	个体遍历定理.
inf	下确界.
infimum	下确界.
infinite collection	无限集合.
infinite mean	无限均值.
infinite measure	无限测度.
infinite product measures	无穷乘积测度.
infinite product space	无穷乘积空间.
infinitely divisible	无穷可分.
infinitely divisible distribution	无穷可分分布.
infinitely often	无穷多次.
information source	信息源.
inhomogeneous heat equation	非齐次热方程.
inhomogeneous Markov chain	非齐次 Markov 链.
initial condition	初始条件.
initial distribution	初始分布.
insurance company	保险公司.
integers	整数.
integrable	可积的.
integral	积分.
integrand	被积函数.
integration	积分.
integration by parts	分部积分.
integration to the limit	极限积分.
integrator	积分算子.
interarrival time	到达间隔时间.

外文术语	中文术语
intersection	交.
intersections	交.
interval of constancy	常数区间.
intuition	直观.
invariance principle	不变原理.
invariant	不变的.
invariant event	不变事件.
invariant events	不变事件.
invariant random variable	不变随机变量.
invariant set	不变集.
invariant under	在... 下不变.
inverse	逆（函数）.
inversion formula	反演公式.
irreducible	不可约的.
isolated point	孤立点.
isolated points	孤立点.
iterate	迭代.
iterated logarithm	重对数.
iterated logarithm law	重对数律.
Itô's formula	Itô 公式.
Itô's lemma	Itô 引理.
Itô calculus	Itô 微积分.
Itô integral	Itô 积分.
Itô process	Itô 过程.
janitor	看门人.
Jensen's inequality	Jensen 不等式.
Jensen inequality	Jensen 不等式.
join	并.
joint density	联合密度.
joint distribution	联合分布.
Jordan decomposition	Jordan 分解.
Kac's theorem	Kac 定理.
Kac	Kac.
Kakutani dichotomy	Kakutani 二分法.
Katznelson	Katznelson.
Kerov	Kerov.
Kesten	Kesten.
killed	被杀死的（过程）.

外文术语	中文术语
Kingman's theorem	Kingman 定理.
Kingman	Kingman.
Kolmogorov's 0-1 law	Kolmogorov 0-1 律.
Kolmogorov's cycle condition	Kolmogorov 循环条件.
Kolmogorov's extension theorem	Kolmogorov 扩张定理.
Kolmogorov's maximal inequality	Kolmogorov 极大不等式.
Kolmogorov's theorem (on infinitely divisible distributions)	Kolmogorov 定理（关于无穷可分分布）.
Kolmogorov's three-series theorem	Kolmogorov 三级数定理.
Kolmogorov-Smirnov statistic	Kolmogorov-Smirnov 统计量.
Kronecker's lemma	Kronecker 引理.
kth moment	k 阶矩.
kth return	第 k 次返回.
Ky Fan metric	Ky Fan 度量.
L∞ norm	L∞ 范数.
L2 convergence	L2 收敛.
L2 maximal inequality	L2 极大不等式.
L2 weak law	L2 弱大数定律.
lack of memory property	无记忆性.
ladder height	阶梯高度.
ladder point	阶梯点.
ladder variable	阶梯变量.
ladder variables	阶梯变量.
Laplace transform	Laplace 变换.
Laplacian	拉普拉斯算子.
large deviations	大偏差.
last visit to 0	最后访问 0.
last zero	最后零点.
latitude	纬度.
lattice distribution	格子点分布.
law of rare events	稀有事件定律.
law of the iterated logarithm	重对数律.
Laws of Large Numbers	大数定律.
Lebesgue's Thorn	Lebesgue 刺.
Lebesgue decomposition	Lebesgue 分解.
Lebesgue measure	Lebesgue 测度.
left limit	左极限.
Legendre transform	Legendre 变换.

外文术语	中文术语
lemma	引理.
length of the longest increasing sequence	最长递增序列的长度.
Leventhal	Leventhal.
Lévy's 0-1 law	Lévy 0-1 律.
Lévy measure	Lévy 测度.
Lévy metric	Lévy 度量.
Lévy-Khinchin theorem	Lévy-Khinchin 定理.
Liggett's version	Liggett 版本.
Liggett	Liggett.
light bulb	灯泡.
LIL	重对数律.
lim inf	下极限.
lim sup	上极限.
liminf	下极限.
limit	极限.
limit from the left	左极限.
limit from the right	右极限.
limit inferior	下极限.
limit set	极限集.
limit superior	上极限.
limiting variance	极限方差.
limsup	上极限.
Lindeberg condition	Lindeberg 条件.
Lindeberg-Feller theorem	Lindeberg-Feller 定理.
Lindeberg-Feller theorem for martingales	鞅的 Lindeberg-Feller 定理.
linear interpolation	线性插值.
linear martingale	线性鞅.
linearity	线性性.
Lipschitz continuity	Lipschitz 连续性.
Lipschitz continuous	Lipschitz 连续.
local limit theorem	局部极限定理.
local martingale	局部鞅.
local maxima	局部极大值.
local maximum	局部极大值.
local time	局部时.
location parameter	位置参数.
Logan	Logan.
logarithmic potential	对数位势.

外文术语	中文术语
lognormal distribution	对数正态分布.
longest common subsequence	最长公共子序列.
longest increasing path	最长递增路径.
longest increasing sequence	最长递增序列.
longitude	经度.
lower bound	下界.
lower semicontinuous	下半连续.
Lp convergence	Lp 收敛.
Lp maximal inequality	Lp 极大不等式.
Lp maximum inequality	Lp 极大不等式.
Lp̂ norm	Lp̂ 范数.
Lyapunov exponent	Lyapunov 指数.
M/G/1 queue	M/G/1 排队系统.
M/M/∞ queue	M/M/∞ 排队系统.
Marcinkiewicz	Marcinkiewicz.
marginal distribution	边缘分布.
Markov's inequality	Markov 不等式.
Markov chain	Markov 链.
Markov property	Markov 性质.
martingale	鞅.
martingale CLT	鞅中心极限定理.
martingale convergence	鞅收敛.
martingale convergence theorem	鞅收敛定理.
martingale difference	鞅差.
martingale difference array	鞅差阵列.
martingale difference sequence	鞅差序列.
martingale differences	鞅差.
martingale increments	鞅增量.
martingale property	鞅性质.
martingale transform	鞅变换.
matching problem	匹配问题.
matrix norm	矩阵范数.
maxima	极大值.
maximal ergodic lemma	极大遍历引理.
maximal inequality	极大不等式.
maximum	极大值.
maximum inequality	极大不等式.
M-dependent sequence	M-相依序列.

外文术语	中文术语
M-dependent sequences	M-相依序列.
mean	均值.
mean measure	均值测度.
mean square error	均方误差.
mean zero implies recurrence	均值为零蕴含常返性.
measurable	可测的.
measurable map	可测映射.
measurable set	可测集.
measurable space	可测空间.
measure	测度.
measure of dependence	依赖性测度.
measure on an algebra	代数上的测度.
Measure Theory	测度论.
measure-preserving	保测的.
measure-preserving transformation	保测变换.
measures on the real line	实直线上的测度.
meet	交.
mesh	网格.
metric	度量.
Minkowski's inequality	Minkowski 不等式.
minorization condition	小化条件.
mixing properties	混合性质.
mixing property	混合性质.
mixing time	混合时间.
mixture	混合.
mixture distribution	混合分布.
mod 1	模 1.
modulus	模.
modulus of continuity	连续性模.
moment	矩.
moment problem	矩问题.
moment-generating function	矩母函数.
monkey at a typewriter	打字机上的猴子.
monotone class theorem	单调类定理.
monotone convergence theorem	单调收敛定理.
monotonicity	单调性.
moving average process	滑动平均过程.
multi-dimensional Brownian motion	多维布朗运动.

外文术语	中文术语
multidimensional Brownian motion	多维布朗运动.
multiplicative functional	乘性泛函.
multivariate normal distribution	多元正态分布.
mutually singular	相互奇异的.
n dimensional Euclidean space	n 维欧几里得空间.
n-dimensional Euclidean space	n 维欧几里得空间.
negative part	负部.
negative set	负集.
negative steps	负步.
net amount	净额.
new maximum	新极大值.
Newtonian potential	Newton 位势.
nice space	好空间.
nonarithmetic distribution	非算术分布.
nondecreasing	非减的.
nondecreasing function	非减函数.
nondifferentiability	不可微性.
nonempty	非空的.
nonhomogeneous Poisson process	非齐次 Poisson 过程.
nonlattice	非格子点的.
nonmeasurable set	不可测集.
nonnegative	非负的.
norm	范数.
normal	正态的.
normal distribution	正态分布.
nth iterate	第 n 次迭代.
null recurrent	零常返.
null set	零集.
number of points visited	访问点数.
number of visits	访问次数.
observation	观测.
occupancy problem	占位问题.
occupation time	占位时间.
occupation times	占位时间.
offspring distribution	后代分布.
omit	省略.
open interval	开区间.
open set	开集.

外文术语	中文术语
open sets	开集.
opposite color	相反颜色.
Optional Stopping Theorem	可选停时定理.
optional stopping theorem	可选停时定理.
Optional Stopping Theorems	可选停时定理.
order relation	序关系.
order statistic	顺序统计量.
order statistics	顺序统计量.
Orey's theorem	Orey 定理.
Ornstein's coupling	Ornstein 耦合.
Ornstein-Uhlenbeck process	Ornstein-Uhlenbeck 过程.
orthogonal	正交的.
orthogonality	正交性.
Orthogonality of martingale increments	鞅增量的正交性.
outcome	结果.
outcomes	结果.
pairwise independence	两两独立.
pairwise independent	两两独立的.
parabolic equation	抛物型方程.
parameter	参数.
parity	奇偶性.
Parseval relation	Parseval 关系.
partial derivative	偏导数.
partial differential equation	偏微分方程.
passage time	通过时间.
path	路径.
path by path	逐路径的.
path properties	路径性质.
paths	路径.
pedestrian delay	行人延迟.
period	周期.
periodic chain	周期链.
permutable	可置换的.
permutable event	可置换事件.
piecewise linear	分段线性.
point mass	点质量.
pointwise convergence	逐点收敛.
pointwise ergodic theorem	点态遍历定理.

外文术语	中文术语
Poisson’s Equation	Poisson 方程.
Poisson convergence	Poisson 收敛.
Poisson distribution	泊松分布.
Poisson equation	Poisson 方程.
Poisson kernel	Poisson 核.
Poisson process	Poisson 过程.
Poisson process in the plane	平面上的 Poisson 过程.
polar set	极集.
Pollaczek-Khintchine formula	Pollaczek-Khintchine 公式.
Polya’s criterion	Polya 准则.
Polya’s urn	Polya 罐子模型.
Polya’s Urn Scheme	Polya 罐子模型.
polygonal line	折线.
polynomial approximation	多项式逼近.
positive entries	正元素.
positive part	正部.
positive recurrent	正常返.
positive steps	正步.
possible value	可能值.
posterior distribution	后验分布.
Potential Kernel	位势核.
potential kernel	位势核.
potential kernels	位势核.
potential theory	位势理论.
predictable	可料的.
predictable increasing sequence	可料递增序列.
predictable process	可料过程.
predictable sequence	可料序列.
premiums	保费.
presentable	可表示的.
prime divisor	素因子.
prior distribution	先验分布.
probability measure	概率测度.
probability space	概率空间.
product measure	乘积测度.
Product Measures	乘积测度.
product σ-field	乘积 σ-代数.
products of random matrices	随机矩阵乘积.

外文术语	中文术语
progressively measurable	循序可测的.
projection	投影.
projection onto L2	到 L2 的投影.
proof	证明.
Punctured Disc	穿孔圆盘.
purpose	目的.
quadratic covariation	二次协变差.
quadratic martingale	二次鞅.
quadratic variation	二次变差.
quantitative measure of dependence	依赖性定量测度.
queueing system	排队系统.
r.c.d.	正则条件分布.
radius of convergence	收敛半径.
Radon-Nikodym derivative	Radon-Nikodym 导数.
Radon-Nikodym theorem	Radon-Nikodym 定理.
random element	随机元.
random index central limit theorem	随机指标中心极限定理.
random permutation	随机排列.
random permutations	随机排列.
random series	随机级数.
random shift	随机移位.
random variable	随机变量.
random vector	随机向量.
random walk	随机游走.
random walk on a tree	树上的随机游走.
random walk with stationary increments	具有平稳增量的随机游走.
range	值域.
rank	秩.
rate	速率.
rate function	速率函数.
rate of convergence	收敛速度.
rate of escape	逃逸速度.
rate one Poisson process	单位速率 Poisson 过程.
ratio limit theorem	比值极限定理.
real line	实直线.
real numbers	实数.
real part	实部.
real valued function	实值函数.

外文术语	中文术语
realization	实现.
recall	回顾.
recurrent	常返的.
recurrent Harris chain	常返 Harris 链.
recurrent state	常返态.
recurrent value	常返值.
reduced word	约化字.
reducible	可约的.
Reflection Principle	反射原理.
Reflection principle	反射原理.
reflection principle	反射原理.
regular conditional distribution	正则条件分布.
regular conditional probability	正则条件概率.
regular point	正则点.
regular subdivision	正则细分.
regular variation	正则变化.
regularity	正则性.
remainder term	余项.
remote future	尾 σ-域.
renewal chain	更新链.
renewal density	更新密度.
renewal equation	更新方程.
renewal measure	更新测度.
renewal process	更新过程.
Renewal theorem	更新定理.
Renewal theory	更新理论.
residual lifetime	剩余寿命.
result	结果.
return time	返回时.
return to 0	返回 0.
reversed martingale	反向鞅.
reversible measure	可逆测度.
Riemann sum	Riemann 和.
Riemann-Lebesgue lemma	Riemann-Lebesgue 引理.
Riemann-Stieltjes integral	Riemann-Stieltjes 积分.
right continuous	右连续的.
right continuous filtration	右连续 σ-域流.
right continuous martingale	右连续鞅.

外文术语	中文术语
right-continuous	右连续的.
rotation of the circle	圆周旋转.
row sum	行和.
row sums	行和.
ruin	破产.
run to infinity	跑向无穷.
sample path	样本路径.
sample space	样本空间.
Sankoff	Sankoff.
satisfy	满足.
scale parameter	尺度参数.
scaling constant	缩放常数.
scaling property	尺度性质.
scaling relation	尺度关系.
Scheffé's theorem	Scheffé 定理.
Second Borel-Cantelli Lemma	第二 Borel-Cantelli 引理.
second Borel-Cantelli lemma	第二 Borel-Cantelli 引理.
second derivative	二阶导数.
second order	二阶.
semi-algebra	半代数.
semialgebra	半代数.
separable metric space	可分度量空间.
service time	服务时间.
set	集合.
set function	集函数.
set of outcomes	结果集.
Shannon's theorem	Shannon 定理.
shares	股份.
Shepp	Shepp.
shift invariant	移位不变的.
shift invariant σ-field	移位不变 σ-域.
shift operator	移位算子.
shift operators	移位算子.
shift transformation	移位变换.
shifted sequence	移位序列.
signed measure	符号测度.
simple function	简单函数.
simple process	简单过程.

外文术语	中文术语
simple random walk (on $\hat{Zd}$)	简单随机游走（$\hat{Zd}$ 上）.
singular	奇异的.
singular with respect to	关于... 奇异.
skewness parameter	偏度参数.
skewness parameter κ	偏度参数 κ.
Skorokhod's representation theorem	Skorokhod 表示定理.
slowly varying function	慢变函数.
smaller σ-field wins	较小的 σ-域取胜.
smallest σ-field	最小 σ-域.
smooth solution	光滑解.
sojourn time	逗留时间.
source term	源项.
space-time Brownian motion	时空布朗运动.
space-time harmonic	时空调和的.
space-time harmonic function	时空调和函数.
spacings	间隔.
span (of distribution)	跨度.
spherical mean	球面平均.
square integrable	平方可积的.
square integrable martingale	平方可积鞅.
square integrable martingale differences	平方可积鞅差.
Square Integrable Martingales	平方可积鞅.
St. Petersburg paradox	St. Petersburg 悖论.
stable distribution	稳定分布.
stable law	稳定律.
standard normal distribution	标准正态分布.
state space	状态空间.
stationary distribution	平稳分布.
stationary increments	平稳增量.
stationary measure	平稳测度.
stationary renewal process	平稳更新过程.
stationary sequence	平稳序列.
step function	阶梯函数.
Stieltjes measure function	Stieltjes 测度函数.
Stieltjes measure functions	Stieltjes 测度函数.
Stieltjes moment problem	Stieltjes 矩问题.
Stirling's formula	Stirling 公式.
stochastic calculus	随机微积分.

外文术语	中文术语
stochastic differential equation	随机微分方程.
stochastic integral	随机积分.
stock	股票.
stopping time	停时.
storage model	存储模型.
Strassen's invariance principle	Strassen 不变原理.
Strassen's theorem	Strassen 定理.
Strassen embedding	Strassen 嵌入.
Strassen invariance principle	Strassen 不变原理.
strict positivity	严格正性.
strict sense	严格意义.
strong law of large numbers	强大数定律.
strong Markov property	强 Markov 性质.
strong mixing	强混合.
subadditive	次可加的.
subadditive ergodic theorem	次可加遍历定理.
subadditivity	次可加性.
subdivision	细分.
submartingale	下鞅.
subsequence	子序列.
subset	子集.
subsets	子集.
success probability	成功概率.
Sufficient Conditions for Independence	独立的充分条件.
Sums of Independent Random Variables	独立随机变量之和.
sup	上确界.
sup norm	上确界范数.
supercritical branching process	超临界分支过程.
superharmonic	上调和的.
superharmonic function	上调和函数.
supermartingale	上鞅.
supremum	上确界.
surface measure	表面测度.
switching principle	切换原理.
symmetric about	关于... 对称.
symmetric difference	对称差.
symmetric simple random walk	对称简单随机游走.
symmetrization	对称化.

外文术语	中文术语
symmetry	对称性.
tail σ-field	尾 σ-域.
Taylor’s theorem	Taylor 定理.
Taylor expansion	Taylor 展开.
temporally homogeneous	时齐的.
terminating renewal process	终止更新过程.
theorem	定理.
thinning	稀疏.
three-series theorem	三级数定理.
throughout	贯穿.
tight	胎紧的.
tightness	胎紧性.
time above 0	位于 0 之上的时间.
time lag	时滞.
time of gambler’s ruin	赌徒破产时间.
time of the last zero	最后零点的时刻.
topological space	拓扑空间.
topology	拓扑.
total assets	总资产.
total mass	总质量.
total variance formula	总方差公式.
total variance in the path	路径上的总方差.
total variation	全变差.
total variation distance	全变差距离.
total variation norm	全变差范数.
tower property	塔性质.
trajectory	轨道.
transient	暂态的.
transient Harris chain	暂态 Harris 链.
transient state	暂态.
transition density	转移密度.
transition kernel	转移核.
transition probability	转移概率.
translation invariance	平移不变性.
travel time	通行时间.
triangular array	三角阵列.
triangular arrays	三角阵列.
triangular distribution	三角分布.

外文术语	中文术语
trichotomy	三分法.
triple	三元组.
trivial	平凡的.
trivial σ-field	平凡 σ-域.
truncating	截断.
truncation	截断.
two-sided stationary sequence	双边平稳序列.
uncorrelated	不相关的.
uncorrelated random variables	不相关随机变量.
uncountable	不可数的.
unfavorable game	不利博弈.
uniform convergence	一致收敛.
uniform distribution	均匀分布.
Uniform Integrability	一致可积性.
uniform norm	一致范数.
uniform on (0,1)	(0,1) 上的均匀分布.
uniformly continuous	一致连续.
uniformly convergent	一致收敛的.
uniformly distributed	均匀分布.
uniformly integrable	一致可积的.
union	并.
unions	并.
unique	唯一的.
unique extension	唯一延拓.
unique measure	唯一测度.
uniqueness theorem	唯一性定理.
unit charge	单位电荷.
up to constant multiples	在常数倍意义下.
upcrossing	上穿.
upcrossing inequality	上穿不等式.
upcrossings	上穿.
upper bound	上界.
upper semicontinuous	上半连续.
usual topology	通常拓扑.
vague convergence	淡收敛.
vanishes	为零的.
variance	方差.
variance at time n	时刻 n 的方差.

外文术语	中文术语
variance decomposition	方差分解.
vectors	向量.
Vershik	Vershik.
version	版本.
vertex	顶点.
vertices	顶点.
w.r.t.	关于.
waiting time	等待时间.
Wald's equation	Wald 等式.
Wald's second equation	Wald 第二等式.
water reservoir	水库.
weak convergence	弱收敛.
weak convergence in $R\hat{d}$	$R\hat{d}$ 中的弱收敛.
weak law	弱大数定律.
weak law for triangular arrays	三角阵列的弱大数定律.
weak law of large numbers	弱大数定律.
weak law of small numbers	小数弱定律.
Weak Laws of Large Numbers	弱大数定律.
weakly converge	弱收敛.
Weiss	Weiss.
Wiener process	维纳过程.
winnings	盈利.
with respect to	关于.
without loss of generality	不失一般性.
word length	字长.
Wright-Fisher model	Wright-Fisher 模型.
zero set	零点集.
Zygmund	Zygmund.
λ-system	λ-系.
π-system	π-系.
π-λ theorem	π-λ 定理.
σ-algebra	σ-代数.
σ-algebra generated by	由... 生成的 σ-代数.
σ-field	σ-域.
σ-field generated by	由... 生成的 σ-域.
σ-finite	σ-有限的.
σ-finite measure space	σ-有限测度空间.
σ-finite measure spaces	σ-有限测度空间.