Contents

Part II 概率论

Part III 随机过程

Part I

测度论

1

可测空间

1.1 可测集

定义 1.1. 集合 E 上的 σ-域 $\mathcal{A}$ 指的是 E 的一个子集族，其满足下面的性质：

1. $E \in \mathcal{A}$；
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$；
3. 如果一列子集 $A_n \in \mathcal{A}$，那么 $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$。

$\mathcal{A}$ 的元素被称为**可测集**, $(E, \mathcal{A})$ 被称为**可测空间**. 根据定义, 我们很容易得出下面的结果:

- $\emptyset = E^c \in \mathcal{A}$.
- 如果一列子集 $A_n \in \mathcal{A}$, 那么

$$\bigcap_{n\in\mathbb{N}} A_n = \left(\bigcup_{n\in\mathbb{N}} A_n\right)^c \in \mathcal{A}.$$

- $\mathcal{A}$ 对有限并和有限交也是封闭的, 只需要从某一项 A_n 开始全部取空集即可.

例 1.2. 根据可测集的定义, 很容易构造出一些最简单的例子:

1. $\mathcal{A} = \mathcal{P}(E)$, 当 E 是有限集或者可数集的时候我们通常会使用这样的 σ-域, 其他情况则很少使用.
2. $\mathcal{A} = \{\emptyset, E\}$, 平凡 σ-域.
3. E 的所有至多可数的子集以及所有补集至多可数的子集构成 E 上的一个 σ-域.

为了产生更多的例子, 我们注意到 E 上任意 σ-域的交集仍然是 σ-域, 这导出了下面的定义.

定义 1.3. 令 $\mathcal{C}$ 是 $\mathcal{P}(E)$ 的子集，E 上包含 $\mathcal{C}$ 的最小的 σ-域被记为 $\sigma(\mathcal{C})$，不难看出其是所有包含 $\mathcal{C}$ 的 σ-域的交集。我们称 $\sigma(\mathcal{C})$ 是由 $\mathcal{C}$ 生成的 σ-域。

定义 1.4. 设 $(E, \mathcal{O})$ 是拓扑空间，所有开集 $\mathcal{O}$ 生成的 σ-域 $\sigma(\mathcal{O})$ 被称为 E 上的 Borel σ-域，记为 $\mathcal{B}(E)$。

E 上的 Borel σ-域是包含所有开集的最小的 σ-域. $\mathcal{B}(E)$ 的元素被称为 E 的 **Borel 子集**. 显然, E 中的闭集也都是 Borel 子集.

例 1.5 (ℝ 上的 Borel σ-域)**.** 记 $\mathcal{C}_1$ 为 ℝ 中开区间的集合:

$$\mathcal{C}_1 = \{(a, b) \mid a, b \in \mathbb{R}, a < b\},$$

显然有 $\mathcal{C}_1 \subseteq \mathcal{B}(\mathbb{R})$, 于是 $\sigma(\mathcal{C}_1) \subseteq \mathcal{B}(\mathbb{R})$. 下面我们说明 $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C}_1)$. 我们不加证明地使用一个结论 (Lindelöf 定理): ℝ 的任意开子集 U 都是开区间的可数并. 那么根据 σ-域的定义, 任意开区间都在 $\sigma(\mathcal{C}_1)$ 中, 故 $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C}_1)$. 这表明 $\mathcal{B}(\mathbb{R})$ 可以由所有开区间生成.

此外, 如果注意到

$$(a, b) = (-\infty, b) \cap (-\infty, a)^c,$$

还可以证明 $\mathcal{B}(\mathbb{R})$ 由 $\mathcal{C}_2$ 生成, 其中

$$\mathcal{C}_2 = \{(-\infty, a) \mid a \in \mathbb{R}\}.$$

最后, 不难证明这里的开区间都可以换成闭区间.

在后文中, 每当我们考虑拓扑空间 (例如 ℝ 或者 $\mathbb{R}^d$) 时, 除非有特别说明, 否则我们总是假设它们配备 Borel σ-域.

下一个非常重要的 σ-域是乘积 σ-域.

定义 1.6. 令 $(E_1, \mathcal{A}_1)$ 和 $(E_2, \mathcal{A}_2)$ 是可测空间，定义 $E_1 \times E_2$ 上的 σ-域 $\mathcal{A}_1 \otimes \mathcal{A}_2$ 为

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

引理 1.7. 设 E 和 F 是可分 (有可数的稠密子集) 的拓扑空间，$E \times F$ 配备积拓扑，那么 $\mathcal{B}(E \times F) = \mathcal{B}(E) \otimes \mathcal{B}(F)$.

1.2 正测度

令 $(E, \mathcal{A})$ 是可测空间.

定义 1.8. $(E, \mathcal{A})$ 上的正测度指的是一个映射 $\mu : \mathcal{A} \to [0, \infty]$，其满足下面的性质:

1. $\mu(\emptyset) = 0$;
2. (σ-可加性) 对于任意可数个不相交的可测集序列 $(A_n)_{n \in \mathbb{N}}$，有

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

此时，三元组 $(E, \mathcal{A}, \mu)$ 被称为**测度空间**. 值 $\mu(E)$ 被称为测度 μ 的总质量.

需要注意的是, 我们允许 μ 的值为 $+\infty$, 此时级数 $\sum_{n \in \mathbb{N}} \mu(A_n)$ 作为正向级数在 $[0, \infty]$ 中总是有意义的. 根据 σ-可加性, 如果我们令 $n > n_0$ 开始 $A_n = \emptyset$, 便可以得到有限可加性.

命题 1.9 (**测度的性质**)**.** 根据定义，测度 μ 满足下面的性质:

1. 如果 $A \subseteq B$，那么 $\mu(A) \le \mu(B)$. 此外，如果还满足 $\mu(A) < \infty$，那么

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

2. 如果 $A, B \in \mathcal{A}$，那么

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

3. 如果 $A_n \in \mathcal{A}$ 且 $A_n \subseteq A_{n+1}$，那么

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

4. 如果 $B_n \in \mathcal{A}$ 且 $B_{n+1} \subseteq B_n$，$\mu(B_1) < \infty$，那么

$$\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \lim_{n\to\infty} \mu(B_n).$$

5. 如果 $A_n \in \mathcal{A}$，那么

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

Proof. (1) 若 $A \subseteq B$, 那么 $B = A \bigcup (B \setminus A)$ 是无交并, 所以

$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$

(2) 若 $\mu(A), \mu(B)$ 中有至少一个为无穷, 那么根据 (1), $\mu(A \cup B)$ 为无穷, 所以结论成立. 下面假设 $\mu(A), \mu(B)$ 均有限, 记 $C = A \cap B$, 那么 $A \cup B = (A \setminus C) \cup C \cup (B \setminus C)$ 是无交并, 所以

$$\mu(A \cup B) = \mu(A \setminus C) + \mu(C) + \mu(B \setminus C) = \mu(A) + \mu(B) - \mu(C),$$

结论 (2) 成立.

(3) 令 $C_1 = A_1$, 对于 $n \ge 2$ 的时候, 令

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

那么 $A_n = \bigcup_{k\le n} C_k$ 是无交并, 所以

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} C_n\right) = \sum_{n\in\mathbb{N}} \mu(C_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(C_k) = \lim_{n\to\infty} \mu(A_n).$$

(4) 令 $A_n = B_1 \setminus B_n$, 那么 $A_n \subseteq A_{n+1}$, 此时

$$\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \mu(B_1) - \mu\left(B_1 \setminus \bigcap_{n\in\mathbb{N}} B_n\right) = \mu(B_1) - \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right),$$

再根据 (3), 就有

$$\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \mu(B_1) - \lim_{n\to\infty}\mu(A_n) = \lim_{n\to\infty}\mu(B_1 \setminus A_n) = \lim_{n\to\infty}\mu(B_n).$$

(5) 令 $C_1 = A_1$, 对于 $n \geq 2$ 的时候, 令

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

那么 C_n 之间互不相交, 所以

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

□

例 1.10 (常见的测度).

1. 令 $E = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$, 定义**计数测度**为

$$\mu(A) = \operatorname{card}(A).$$

2. 如果 A 是 E 的子集, 定义 A 的示性函数 $\mathbf{1}_A : E \to \{0,1\}$ 为

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

令 $(E, \mathcal{A})$ 是可测空间, 固定 $x \in E$. 对于每个 $A \in \mathcal{A}$, 令 $\delta_x(A) = \mathbf{1}_A(x)$, 这给出了 $(E, \mathcal{A})$ 上的一个测度, 被称为 **x 处的 Dirac 测度**. 更一般的, 如果 $(x_n)_{n\in\mathbb{N}}$ 是 E 中的点列, $(\alpha_n)_{n\in\mathbb{N}}$ 是 $[0,\infty]$ 中的点列, 我们可以考虑测度 $\sum_{n\in\mathbb{N}}\alpha_n\delta_{x_n}$ 为

$$\left(\sum_{n\in\mathbb{N}}\alpha_n\delta_{x_n}\right)(A) = \sum_{n\in\mathbb{N}}\alpha_n\delta_{x_n}(A) = \sum_{n\in\mathbb{N}}\alpha_n\mathbf{1}_A(x_n),$$

这个测度被称为 E 上的**点测度**.

3. 可以证明, 在 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上存在唯一的正测度 λ 使得: 对于每个闭区间 $[a,b]$, 有 $\lambda\big([a,b]\big) = b - a$. 这个测度 λ 被称为 **Lebesgue 测度**. Lebesgue 测度的唯一性可以由推论 1.24 保证, 存在性由? 保证.

如果 μ 是 $(E, \mathcal{A})$ 上的正测度, $C \in \mathcal{A}$, 那么可以定义 μ 在 C 上的**限制** ν 为:

$$\nu(A) = \mu(A \cap C), \quad \forall A \in \mathcal{A}.$$

不难验证 ν 还是 $(E, \mathcal{A})$ 上的正测度.

定义 1.11.

- 如果 $\mu(E) < \infty$, 那么我们说测度 μ 是**有限的**.
- 如果 $\mu(E) = 1$, 那么我们说测度 μ 是**概率测度**, $(E, \mathcal{A}, \mu)$ 是**概率空间**.
- 如果存在一列可测集 $(E_n)_{n\in\mathbb{N}}$ 使得 $E = \bigcup_n E_n$ 以及每个 $\mu(E_n) < \infty$, 那么我们说测度 μ 是 **σ-有限的**.

- 如果 $x \in E$ 使得单点集 $\{x\} \in \mathcal{A}$ 并且 $\mu(\{x\}) > 0$，那么我们说 x 是测度 μ 的一个**原子**.
- 如果测度 μ 没有原子，那么我们说 μ 是**扩散测度**.

如果 $(A_n)_{n\in\mathbb{N}}$ 是一列可测集, 类比数列的上下极限, 我们可以定义集合列的上下极限分别为:

$$\limsup A_n = \bigcap_{n=1}^{\infty}\left(\bigcup_{k=n}^{\infty} A_k\right), \quad \liminf A_n = \bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty} A_k\right).$$

注意到对于任意 m, 都有

$$\bigcup_{n=1}^{m}\left(\bigcap_{k=n}^{\infty} A_k\right) = \bigcap_{k=m}^{\infty} A_k, \quad \bigcap_{n=1}^{m}\left(\bigcup_{k=n}^{\infty} A_k\right) = \bigcup_{k=m}^{\infty} A_k,$$

所以显然有 $\liminf A_n \subseteq \limsup A_n$.

引理 1.12. 令 μ 是 $(E,\mathcal{A})$ 上的测度，那么

$$\mu(\liminf A_n) \le \liminf \mu(A_n).$$

如果 μ 是有限测度，或者更一般地，$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) < \infty$，那么

$$\mu(\limsup A_n) \ge \limsup \mu(A_n).$$

Proof. 对于任意的 n, 有

$$\mu\left(\bigcap_{k=n}^{\infty} A_k\right) \le \inf_{k\ge n} \mu(A_k),$$

所以

$$\mu(\liminf A_n) = \lim_{n\to\infty} \mu\left(\bigcap_{k=n}^{\infty} A_k\right) \le \lim_{n\to\infty}\inf_{k\ge n} \mu(A_k) = \liminf \mu(A_n).$$

第二个结论同理. □

1.3 可测函数

定义 1.13. 令 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$ 是两个可测空间，如果映射 $f: E \to F$ 满足：

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

那么我们说 f 是**可测映射**. 当 E, F 是两个配备了 Borel σ-域的拓扑空间时，我们说 f 是 **Borel 可测的**.

显然, 可测映射的复合是可测映射.

命题 1.14. 令 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$ 是两个可测空间，映射 $f: E \to F$. f 可测当且仅当对于某个生成 $\mathcal{B}$ 的子集族 $\mathcal{C}$ (即 $\mathcal{B} = \sigma(\mathcal{C})$)，有 $f^{-1}(B) \in \mathcal{A}$ ($\forall B \in \mathcal{C}$).

Proof. 只需证明充分性. 记

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

直接验证可知 $\mathcal{G}$ 是一个 σ-域, 又因为 $\mathcal{C} \subseteq \mathcal{G}$, 所以 $\mathcal{B} = \sigma(\mathcal{C}) \subseteq \mathcal{G} \subseteq \mathcal{B}$, 所以 $\mathcal{G} = \mathcal{B}$, 这就表明 f 是可测的. □

例 1.15. 若 $(F, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 要证明 f 是可测的, 只需说明集合 $f^{-1}((a,b))$ 是可测的, 或者 $f^{-1}((-\infty, a))$ 是可测的.

推论 1.16. 设 E, F 是两个配备 Borel σ-域的拓扑空间,那么连续映射 $f : E \to F$ 都是可测的.

引理 1.17. 令 $(E, \mathcal{A}), (F_1, \mathcal{B}_1)$ 和 $(F_2, \mathcal{B}_2)$ 是可测空间，乘积 $F_1 \times F_2$ 配备乘积 σ-域 $\mathcal{B}_1 \otimes \mathcal{B}_2$，令映射 $f_1 : E \to F_1$ 和 $F_2 : E \to F_2$，定义 $f : E \to F_1 \times F_2$ 为 $f(x) = (f_1(x), f_2(x))$，那么 f 可测当且仅当 f_1, f_2 都可测.

推论 1.18. 令 $(E, \mathcal{A})$ 是可测空间，f, g 是从 E 到 $\mathbb{R}$ 的可测函数，那么函数

$$f + g, fg, \min(f, g), \max(f, g)$$

都是可测的.

记扩充实数 $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, 其拓扑为序拓扑. 与 $\mathbb{R}$ 类似, $\bar{\mathbb{R}}$ 的 Borel σ-域由区间 $[-\infty, a)$ 生成.

命题 1.19. 令 $(f_n)_{n \in \mathbb{N}}$ 是 $E \to \bar{\mathbb{R}}$ 的可测函数列，那么

$$\sup f_n, \quad \inf f_n, \quad \limsup_{n\to\infty} f_n, \quad \liminf_{n\to\infty} f_n$$

都是可测函数. 特别地，如果 (f_n) 逐点收敛，那么极限 $\lim f_n$ 是可测函数.

下述技巧性引理通常是有用的.

引理 1.20. 令 (f_n) 是一列从 E 到 $\mathbb{R}$ 的可测函数. 所有使得 $f_n(x)$ $(n \to \infty)$ 收敛的 $x \in \mathbb{R}$ 的集合 A 是可测的. 此外，定义函数 $h : E \to \mathbb{R}$ 为

$$h(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & x \in A, \\ 0 & x \notin A. \end{cases}$$

那么 h 是可测的.

定义 1.21. 令 $(E, \mathcal{A})$ 和 $(F, \mathcal{B})$ 是可测空间，$\varphi : E \to F$ 是可测映射，μ 是 $(E, \mathcal{A})$ 上的测度，定义 $(F, \mathcal{B})$ 上的测度 ν 为

$$\nu(B) = \mu(\varphi^{-1}(B)), \quad \forall B \in \mathcal{B}.$$

ν 被称为 **$\boldsymbol{\mu}$ 在 $\boldsymbol{\varphi}$ 下的推前**，记为 $\varphi(\mu)$，有时也记为 $\varphi_* \mu$.

1.4 单调类

本节我们陈述单调类定理, 这是测度论甚至概率论中的一个基本工具.

定义 1.22. $\mathcal{P}(E)$ 的一个子集 $\mathcal{M}$ 如果满足:

1. $E \in \mathcal{M}$;
2. 对于任意 $A, B \in \mathcal{M}$ 且 $A \subseteq B$, 有 $B \setminus A \in \mathcal{M}$;
3. 如果一列子集 $A_n \subseteq \mathcal{M}$ 且 $A_n \subseteq A_{n+1}$, 那么 $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$,

那么我们说 $\mathcal{M}$ 是一个**单调类**.

显然, σ-域都是单调类. 反之, 一个单调类是 σ-域当且仅当其对有限交封闭. 这很容易证明, 若单调类 $\mathcal{M}$ 对有限交封闭, 那么任取一列子集 $A_n \subseteq \mathcal{M}$, 对于任意的 n, 有

$$\bigcup_{k=1}^{n} A_k = E \setminus \bigcap_{k=1}^{n} A_k^c \in \mathcal{M},$$

所以

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} \left(\bigcup_{k=1}^{n} A_k \right) \in \mathcal{M},$$

这就表明 $\mathcal{M}$ 是一个 σ-域.

与 σ-域类似, 显然单调类的任意交仍然是单调类. 如果 $\mathcal{C} \subseteq \mathcal{P}(E)$, 那么我们可以定义由 $\mathcal{C}$ 生成的单调类 $\mathcal{M}(\mathcal{C})$, 即包含 $\mathcal{C}$ 的最小的单调类, 其可以通过对所有包含 $\mathcal{C}$ 的单调类取交集得到.

定理 1.23 (单调类定理). 令 $\mathcal{C} \subseteq \mathcal{P}(E)$ 对有限交封闭, 那么 $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$. 因此, 如果 $\mathcal{M}$ 是包含 $\mathcal{C}$ 的任意单调类, 那么 $\sigma(\mathcal{C}) \subseteq \mathcal{M}$.

Proof. 显然有 $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$. 要证明 $\sigma(\mathcal{C}) \subseteq \mathcal{M}(\mathcal{C})$, 只需要说明 $\mathcal{M}(\mathcal{C})$ 是 σ-域. 根据上面的叙述, 这只需要说明 $\mathcal{M}(\mathcal{C})$ 对有限交封闭.

对于 $A \in \mathcal{P}(E)$, 记

$$\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{C}) \mid A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

直接验证可知 $\mathcal{M}_A$ 是一个单调类. 下面任取 $A \in \mathcal{C}$, 由于 $\mathcal{C}$ 对有限交封闭, 所以 $\mathcal{C} \subseteq \mathcal{M}_A$, 这就表明 $\mathcal{M}(\mathcal{C}) \subseteq \mathcal{M}_A$.

接下来任取 $D \in \mathcal{M}(\mathcal{C})$, 上面的叙述告诉我们 $\mathcal{C} \subseteq \mathcal{M}_D$, 所以 $\mathcal{M}(\mathcal{C}) \subseteq \mathcal{M}_D$. 这就表明 $\mathcal{M}(\mathcal{C})$ 对有限交封闭, 所以 $\mathcal{M}(\mathcal{C})$ 是 σ-域. □

单调类定理最重要的应用是证明某些测度的唯一性.

推论 1.24. 令 μ, ν 是 $(E, \mathcal{A})$ 上的两个测度. 假设存在一个子集族 $\mathcal{C} \subseteq \mathcal{A}$ 满足 $\mathcal{C}$ 对有限交封闭且 $\mathcal{A} = \sigma(\mathcal{C})$, 并且对于任意 $A \in \mathcal{C}$ 都有 $\mu(A) = \nu(A)$.

1. 如果 $\mu(E) = \nu(E) < \infty$, 那么 $\mu = \nu$.
2. 如果存在一列 $\mathcal{C}$ 中的递增序列 $(E_n)_{n\in\mathbb{N}}$ 使得 $E = \bigcup_{n\in\mathbb{N}} E_n$, 并且 $\mu(E_n) = \nu(E_n) < \infty$, 那么 $\mu = \nu$.

Proof. (1) 令
$$\mathcal{G} = \{A \in \mathcal{A} \mid \mu(A) = \nu(A)\},$$
那么 $\mathcal{C} \subseteq \mathcal{G}$ 且不难验证 $\mathcal{G}$ 是单调类, 根据单调类定理, 有 $\mathcal{A} = \sigma(\mathcal{C}) \subseteq \mathcal{G}$, 即 $\mu = \nu$.

(2) 记 μ_n 为 μ 在 E_n 上的限制, ν_n 同理. 那么
$$\mu_n(E) = \mu(E \cap E_n) = \mu(E_n) = \nu(E_n) = \nu(E \cap E_n) = \nu_n(E),$$
根据 (1), 有 $\mu_n = \nu_n$. 于是任取 $A \in \mathcal{A}$, 有
$$\begin{aligned}
\mu(A) = \mu(A \cap E) = \mu\left(\bigcup_{n\in\mathbb{N}} (A \cap E_n)\right) &= \lim_{n\to\infty} \uparrow \mu(A \cap E_n) \\
&= \lim_{n\to\infty} \uparrow \mu_n(A) = \lim_{n\to\infty} \uparrow \nu_n(A) = \lim_{n\to\infty} \uparrow \nu(A \cap E_n) \\
&= \nu\left(\bigcup_{n\in\mathbb{N}} (A \cap E_n)\right) = \nu(A \cap E) = \nu(A),
\end{aligned}$$
这就表明 $\mu = \nu$. □

推论 1.24 表明了 Lebesgue 测度的唯一性. 即若 λ 是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的正测度, 且使得 $\lambda\big([a,b]\big) = b - a$, 那么这样的测度 λ 是唯一的. 这是因为我们可以取
$$\mathcal{C} = \big\{[a,b] \mid a, b \in \mathbb{R}, a < b\big\},$$
此时 $\mathcal{C}$ 对有限交封闭并且 $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. 取 $E_n = [-n, n] \in \mathcal{C}$, 那么 $\mathbb{R} = \bigcup_{n\in\mathbb{N}} E_n$ 且 $\lambda(E_n) < \infty$, 应用推论 1.24 的 (2) 即可表明唯一性.

2

可测函数的积分

2.1 非负函数的积分

在本章中, 我们考虑配备正测度 μ 的可测空间 $(E, \mathcal{A})$.

简单函数　如果可测函数 $f : E \to \mathbb{R}$ 的值域是有限集, 那么我们说 f 的**简单函数**. 假设 f 的所有可能的取值为 $\alpha_1, \ldots, \alpha_n$, 不妨假设 $\alpha_1 < \alpha_2 < \cdots < \alpha_n$. 那么 f 可以表示为

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

其中 $A_i = f^{-1}(\{\alpha_i\}) \in \mathcal{A}$.注意到 E 是 $A_1, \ldots, A_n$ 的无交并.上述公式 $f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$ 被称为 f 的标准表示.

定义 2.1. 令 f 是取值在 $\mathbb{R}_+$ 中的简单函数，标准表示为 $f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$. 定义 f **相对于 μ 的积分**为

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

在 $\alpha_i = 0$ 和 $\mu(A_i) = \infty$ 的情况下，约定 $0 \times \infty = 0$.

注意上述定义中 $\sum_{i=1}^{n} \alpha_i \mu(A_i)$ 的取值为 $[0, \infty]$. 所以在上述定义中我们只考虑非负的简单函数, 这是为了避免出现 $\infty - \infty$ 之类的表达式.

值得注意的是, 如果简单函数 f 有表达

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

其中 B_j 仍然构成 E 的一个划分, 但是 β_j 不再是两两不同的. 此时 f 的积分仍然为

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

这是因为对于每个 A_i, 某些 B_j 构成了 A_i 的划分, 即

$$A_i = \bigcup_{\{j \mid \beta_j = \alpha_i\}} B_j,$$

那么

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

非负简单函数的积分满足下面的一些基本的性质.

命题 2.2. **令** f, g **是** E **上的非负简单函数.**

1. 对于每个 $a, b \in \mathbb{R}_+$, 有

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

2. 如果 $f \le g$, 那么

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

Proof. (1) 设 f, g 的标准表示分别为

$$f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}, \quad g = \sum_{j=1}^{m} \beta_j \mathbf{1}_{B_j}.$$

那么每个 A_i 都是某些 $A_i \cap B_j$ 的无交并, 同理, 每个 B_j 都是某些 $A_i \cap B_j$ 的无交并, 于是我们可以使用一个新的划分 $\{C_1, \ldots, C_p\}$ 使得

$$f = \sum_{k=1}^{p} \gamma_k \mathbf{1}_{C_k}, \quad g = \sum_{k=1}^{p} \theta_k \mathbf{1}_{C_k},$$

此时 γ_k 不一定互不相同, θ_k 也不一定互不相同, 根据命题前面的叙述, 我们有

$$\begin{aligned}
\int (af + bg)\,\mathrm{d}\mu &= \sum_{k=1}^{p} (a\gamma_k + b\theta_k)\mu(C_k) \\
&= a \sum_{k=1}^{p} \gamma_k \mu(C_k) + b \sum_{k=1}^{p} \theta_k \mu(C_k) \\
&= a \int f\,\mathrm{d}\mu + b \int g\,\mathrm{d}\mu.
\end{aligned}$$

(2) 由 (1), 有

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

我们用 $\mathcal{E}_+$ 来表示 E 上的非负简单函数的集合.

定义 2.3. 令 $f : E \to [0, \infty]$ 是可测函数, 定义 **f 相对于 μ 的积分**为

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

f 相对于 μ 的积分通常有很多写法, 下面的表达

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

表示的含义是完全相同的. 此外, 如果 A 是 E 的可测子集, 我们定义

$$\int_A f\,\mathrm{d}\mu = \int f\mathbf{1}_A\,\mathrm{d}\mu.$$

从现在开始, 我们用非负可测函数表示 $E\to[0,\infty]$ 的可测函数 (值可以为无穷). 需要注意的是, 我们前面定义的非负简单函数值必须有限.

命题 2.4. 令 f,g 是 E 上的非负可测函数.

1. 如果 $f\le g$, 那么 $\int f\,\mathrm{d}\mu\le\int g\,\mathrm{d}\mu$.
2. 如果 $\mu(\{x\in E\mid f(x)>0\})=0$, 那么 $\int f\,\mathrm{d}\mu=0$.

Proof. (1) 显然

$$\{h\in\mathcal{E}_+\mid h\le f\}\subseteq\{h\in\mathcal{E}_+\mid h\le g\},$$

根据定义即得 $\int f\,\mathrm{d}\mu\le\int g\,\mathrm{d}\mu$.

(2) 设 $h\in\mathcal{E}_+$ 且 $h\le f$, 设 h 的标准表示为 $h=\sum_{i=1}^n\alpha_i\mathbf{1}_{A_i}$, 若 $\alpha_i>0$, 那么

$$\mu(A_i)\le\mu(\{x\in E\mid h(x)>0\})\le\mu(\{x\in E\mid f(x)>0\})=0,$$

所以

$$\int h\,\mathrm{d}\mu=\sum_{\{i\mid\alpha_i=0\}}\alpha_i\mu(A_i)+\sum_{\{i\mid\alpha_i>0\}}\alpha_i\mu(A_i)=0+0=0,$$

故 $\int f\,\mathrm{d}\mu=0$. □

下面的单调收敛定理是测度论中的一个极为重要的基本定理, 其表明对于一列递增的非负可测函数, 极限和积分可以交换次序.

定理 2.5 (单调收敛定理). 令 $(f_n)_{n\in\mathbb{N}}$ 是 E 上的一列递增的非负可测函数, 即 $f_n\le f_{n+1}$, 记 $f=\lim\uparrow f_n$, 那么

$$\int f\,\mathrm{d}\mu=\lim_{n\to\infty}\uparrow\int f_n\,\mathrm{d}\mu.$$

Proof. 由于 $f_n\le f$, 所以 $\int f_n\,\mathrm{d}\mu\le\int f\,\mathrm{d}\mu$, 所以 $\lim\uparrow\int f_n\,\mathrm{d}\mu\le\int f\,\mathrm{d}\mu$, 于是我们只需要证明反向的不等式.

假设非负可测函数 $h=\sum_{i=1}^k\alpha_i\mathbf{1}_{A_i}$ 满足 $h\le f$, 任取 $a\in[0,1)$, 定义一列可测集

$$E_n=\{x\in E\mid ah(x)\le f_n(x)\},$$

此时对于任意的 $x\in E$, 都有 $ah(x)<h(x)\le f(x)$, 而 $f=\lim\uparrow f_n$, 所以总存在足够大的 n, 使得 $ah(x)\le f_n(x)$, 这表明 $E=\bigcup_{n\in\mathbb{N}}E_n$. 此外, $f_n\le f_{n+1}$ 表明 $E_n\subseteq E_{n+1}$.

显然 $f_n \geq ah\mathbf{1}_{E_n}$, 所以

$$\int f_n \,\mathrm{d}\mu \geq a\int_{E_n} h \,\mathrm{d}\mu = a\sum_{i=1}^{k}\alpha_i \mu(A_i \cap E_n),$$

由于 $A_i = A_i \cap E = \bigcup_{n\in\mathbb{N}}(A_i \cap E_n)$, 所以

$$\mu(A_i) = \mu\left(\bigcup_{n\in\mathbb{N}}(A_i \cap E_n)\right) = \lim_{n\to\infty}\uparrow \mu(A_i \cap E_n),$$

于是

$$\lim_{n\to\infty}\uparrow \int f_n \,\mathrm{d}\mu \geq a\sum_{i=1}^{k}\alpha_i \lim_{n\to\infty}\uparrow \mu(A_i \cap E_n) = a\sum_{i=1}^{k}\alpha_i\mu(A_i) = a\int h\,\mathrm{d}\mu,$$

由于 a 可以任意接近 1, 所以

$$\lim_{n\to\infty}\uparrow \int f_n \,\mathrm{d}\mu \geq \int h\,\mathrm{d}\mu,$$

所以

$$\lim_{n\to\infty}\uparrow \int f_n \,\mathrm{d}\mu \geq \int f\,\mathrm{d}\mu = \sup_{h\in\mathcal{E}_+, h\leq f}\int h\,\mathrm{d}\mu. \qquad \square$$

命题 2.6.

1. 设 f 是 E 上的非负可测函数，那么存在一列递增的非负简单函数 $(f_n)_{n\in\mathbb{N}}$ 使得 $f = \lim \uparrow f_n$. 如果 f 有界，那么 $f_n \to f$ 一致收敛.
2. 令 f, g 是两个 E 上的非负可测函数，$a, b \in \mathbb{R}_+$，那么

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

3. 令 $(f_n)_{n\in\mathbb{N}}$ 是一列 E 上的非负可测函数，那么

$$\int\left(\sum_{n\in\mathbb{N}} f_n\right)\mathrm{d}\mu = \sum_{n\in\mathbb{N}}\int f_n\,\mathrm{d}\mu.$$

Proof. (1) 令 $d_n : [0,\infty] \to \mathbb{R}_+$ 为

$$d_n = \sum_{k=1}^{n2^n}\frac{k-1}{2^n}\mathbf{1}_{[\frac{k-1}{2^n},\frac{k}{2^n})} + n\mathbf{1}_{[n,\infty]},$$

显然 d_n 是非负简单函数. 直观上来看, d_n 将区间 $[0,n]$ 等分为了 $n2^n$ 份, 即将 $[0,1]$ 等分为了 2^n 份. 那么对于 $x \in [0,n)$, 总存在唯一的 k_n 使得 $(k_n-1)/2^n \leq x < k_n/2^n$, 此时 $k_{n+1} = 2k_n$ 或者 $k_{n+1} = 2k_n - 1$, 所以

$$d_{n+1}(x) = \frac{k_{n+1}-1}{2^{n+1}} \geq \frac{k_n - 1}{2^n} = d_n(x),$$

这表明 $d_n \le d_{n+1}$. 此外, 不难看出 $\lim d_n(x) = x$.

令 $f_n = d_n \circ f$, 由于 f_n 只有有限多个取值, 所以 f_n 是非负简单函数. $d_n \le d_{n+1}$ 表明 $f_n \le f_{n+1}$. 且 $\lim f_n = \lim d_n \circ f = f$, 所以 f_n 就是一列递增的非负简单函数且 $f = \lim \uparrow f_n$. f 有界表明在 n 足够大的时候有 $0 \le f - f_n \le 2^{-n}$, 即 $f_n \to f$ 一致收敛.

(2) 由 (1), 设 $f = \lim \uparrow f_n$, $g = \lim \uparrow g_n$, 其中 $(f_n), (g_n)$ 均为一列递增的简单函数, 那么

$$\int (af_n + bg_n)\,\mathrm{d}\mu = a\int f_n\,\mathrm{d}\mu + b\int g_n\,\mathrm{d}\mu,$$

令 $n \to \infty$, 再根据单调收敛定理, 就有

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

(3) 根据 (2), 有

$$\int \left(\sum_{n=1}^{m} f_n\right)\mathrm{d}\mu = \sum_{n=1}^{m}\int f_n\,\mathrm{d}\mu,$$

令 $m \to \infty$, 再根据单调收敛定理, 就有

$$\int \left(\sum_{n\in\mathbb{N}} f_n\right)\mathrm{d}\mu = \sum_{n\in\mathbb{N}}\int f_n\,\mathrm{d}\mu.$$

□

注释 2.7. 命题 2.6 和单调收敛定理 2.5 给出了证明关于非负可测函数积分的命题的一种基本范式，即根据 命题 2.6 的 (1)，假设一列非负简单函数逼近原函数，先证明命题对非负简单函数成立，这通常是非常容易的，再使用单调收敛定理证明命题对所有的非负可测函数成立.

下面的推论在概率论中十分有用, 其对应于随机变量的概率密度函数. 其证明是上述注释中技巧的一个典型运用.

推论 2.8. 令 g 是非负可测函数，对于 $A \in \mathcal{A}$，令

$$\nu(A) = \int_A g\,\mathrm{d}\mu = \int g\mathbf{1}_A\,\mathrm{d}\mu,$$

那么 ν 是 E 上的正测度，被称为密度 g 相对于 μ 的测度，记为 $\nu = g\cdot\mu$. 此外，对于非负可测函数 f，有

$$\int f\,\mathrm{d}\nu = \int fg\,\mathrm{d}\mu.$$

Proof. 显然 $\nu(\emptyset) = 0$. 任取一列不相交的 $A_n \in \mathcal{A}$, 那么

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

这就表明 ν 是 E 上的正测度.

对于任意示性函数 $\mathbf{1}_A$, 有

$$\int \mathbf{1}_A \,\mathrm{d}\nu = \nu(A) = \int \mathbf{1}_A g \,\mathrm{d}\mu,$$

进一步的, 令 $f = \lim \uparrow f_n$, 其中 f_n 是非负简单函数, 对于每个 f_n, 根据积分的线性性, 都有

$$\int f_n \,\mathrm{d}\nu = \int f_n g \,\mathrm{d}\mu,$$

令 $n \to \infty$, 根据单调收敛定理, 就有

$$\int f \,\mathrm{d}\nu = \int fg \,\mathrm{d}\mu. \qquad \square$$

注释 2.9. 在实际中，我们通常也会写作 $\nu(\mathrm{d}x) = g(x)\mu(\mathrm{d}x)$，或者 $g = \mathrm{d}\nu/\,\mathrm{d}\mu$.

在测度论中, 命题通常在**几乎处处** (almost everywhere) 的意义下成立, 也就是说, 对于不满足该命题的所有 $x \in E$ 的集合, 这个集合的 μ-测度为 0, 我们使用简写 μ a.e. 来表示这个意思. 也就是说, 当我们写到 $f = g$, μ a.e. 的时候, 我们表示的意思实际上是

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

命题 2.10. 令 f 是非负可测函数.

1. 对于每个 $a \in (0, \infty)$，有

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

2. 我们有

$$\int f \,\mathrm{d}\mu < \infty \Rightarrow f < \infty,\ \mu \text{ a.e.}$$

3. 我们有

$$\int f \,\mathrm{d}\mu = 0 \Leftrightarrow f = 0,\ \mu \text{ a.e.}$$

4. 如果 g 是非负可测函数，

$$f = g,\ \mu \text{ a.e.} \Rightarrow \int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu.$$

Proof. (1) 令可测集 $A = \{x \in E \mid f(x) \geq a\}$, 那么 $f \geq a\mathbf{1}_A$, 所以

$$\int f \,\mathrm{d}\mu \geq a \int \mathbf{1}_A \,\mathrm{d}\mu = a\mu(A).$$

(2) 令可测集 $A_n = \{x \in E \mid f(x) \geq n\}$ 以及 $A_\infty = \{x \in E \mid f(x) = \infty\}$, 那么 $A_{n+1} \subseteq A_n$ 且 $A_\infty = \bigcap_{n \in \mathbb{N}} A_n$. 根据 (1), 有

$$\mu(A_1) \leq \int f \,\mathrm{d}\mu < \infty,$$

所以

$$\mu(A_\infty) = \lim_{n\to\infty} \downarrow \mu(A_n) \le \lim_{n\to\infty} \downarrow \frac{1}{n}\int f\,\mathrm{d}\mu = 0,$$

所以 $\mu(A_\infty) = 0$，即 $f < \infty$，μ a.e..

(3) 充分性由命题 2.6 的 (2) 保证．下证必要性．令可测集 $A_n = \{x \in E \mid f(x) \ge 1/n\}$ 以及 $A_\infty = \{x \in E \mid f(x) \ne 0\}$，那么 $A_n \subseteq A_{n+1}$ 且 $A_\infty = \bigcup_{n\in\mathbb{N}} A_n$．根据 (1)，有

$$\mu(A_\infty) = \lim_{n\to\infty} \uparrow \mu(A_n) \le \lim_{n\to\infty} \uparrow n\int f\,\mathrm{d}\mu = 0,$$

所以 $\mu(A_\infty) = 0$．

(4) 记 $f \wedge g = \min(f, g)$ 及 $f \vee g = \max(f, g)$，那么 $f = g$，μ a.e. 表明 $f \vee g = f \wedge g$，μ a.e.．根据 (3)，有

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

又因为 $\int f \wedge g\,\mathrm{d}\mu \le \int f\,\mathrm{d}\mu \le \int f \vee g\,\mathrm{d}\mu$，对于 g 类似，所以

$$\int f\,\mathrm{d}\mu = \int g\,\mathrm{d}\mu. \qquad \square$$

定理 2.11 (Fatou 引理). 令 $(f_n)_{n\in\mathbb{N}}$ 是一列非负可测函数，那么

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

Proof. 只需证明

$$\int \lim_{n\to\infty} \inf_{k\ge n} f_k\,\mathrm{d}\mu \le \lim_{n\to\infty} \inf_{k\ge n} \int f_k\,\mathrm{d}\mu,$$

令 $g_n = \inf_{k\ge n} f_k$，那么 $g_n \le g_{n+1}$，根据单调收敛定理，有

$$\int \lim_{n\to\infty} g_n\,\mathrm{d}\mu = \lim_{n\to\infty} \uparrow \int g_n\,\mathrm{d}\mu.$$

对于任意 n 和 $k \ge n$，有 $\int g_n\,\mathrm{d}\mu \le \int f_k\,\mathrm{d}\mu$，所以

$$\int \lim_{n\to\infty} g_n\,\mathrm{d}\mu = \lim_{n\to\infty} \uparrow \int g_n\,\mathrm{d}\mu \le \inf_{k\ge n} \int f_k\,\mathrm{d}\mu,$$

令 $n \to \infty$，即可得到

$$\int \liminf f_n\,\mathrm{d}\mu \le \liminf \int f_n\,\mathrm{d}\mu. \qquad \square$$

可以用一个简单的例子理解 Fatou 引理．考虑 $f_n = \mathbf{1}_{[n,n+1]}$，那么对于任意 $x \in \mathbb{R}$，有 $\liminf f_n(x) = 0$，此时显然有

$$0 = \int \liminf f_n\,\mathrm{d}\mu \le \liminf \int f_n\,\mathrm{d}\mu = 1.$$

直观上来看，就是把函数压扁（取极限）再求和（积分），得到的总和通常会变小，因为在这个过程中，有些原本存在的面积可能逃逸或者坍缩掉了．

命题 2.12. 令 $(F,\mathcal{B})$ 是可测空间，$\varphi: E \to F$ 是可测映射. 令 ν 是 μ 在 φ 下的推前. 那么，对于任意 F 上的非负可测函数 h，我们有

$$\int_E h(\varphi(x))\mu(\mathrm{d}x) = \int_F h(y)\nu(\mathrm{d}y).$$

Proof. 若 $h = \mathbf{1}_B$ 是示性函数, 那么

$$\int_E h(\varphi(x))\mu(\mathrm{d}x) = \mu(\varphi^{-1}(B)) = \nu(B) = \int_F h(y)\nu(\mathrm{d}y).$$

若 $h = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ 是非负简单函数, 那么根据积分的线性性, 结论也成立. 若 h 是一般的非负可测函数, 设 $(h_n)_{n\in\mathbb{N}}$ 是一列递增的非负简单函数且 $h = \lim \uparrow h_n$, 根据单调收敛定理, 即可证明结论. □

2.2 可积函数

本节我们讨论可变号的可测函数. 如果 $f: E \to \mathbb{R}$ 是可测函数, 记 f 正部分 $f^+ = \max(f, 0)$, 负部分 $f^- = \max(-f, 0)$, 需要注意 f^+ 和 f^- 此时都是非负可测函数并且 $f = f^+ - f^-$, $|f| = f^+ + f^-$.

定义 2.13. 令 $f: E \to \mathbb{R}$ 是可测函数，如果

$$\int |f|\,\mathrm{d}\mu < \infty,$$

那么我们说 f 相对于 μ **可积**. 在这种情况下，我们定义

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

如果 $A \in \mathcal{A}$，记

$$\int_A f\,\mathrm{d}\mu = \int f\mathbf{1}_A\,\mathrm{d}\mu.$$

我们使用 $\mathcal{L}^1(E,\mathcal{A},\mu)$ 来表示所有可积函数 $f: E \to \mathbb{R}$ 构成的空间. $\mathcal{L}^1_+(E,\mathcal{A},\mu)$ 来表示所有非负可积函数构成的空间.

命题 2.14 (可积函数的性质).

1. 对于任意 $f \in \mathcal{L}^1(E,\mathcal{A},\mu)$，有 $\left|\int f\,\mathrm{d}\mu\right| \le \int |f|\,\mathrm{d}\mu$.
2. $\mathcal{L}^1(E,\mathcal{A},\mu)$ 是 $\mathbb{R}$-向量空间.
3. 如果 $f, g \in \mathcal{L}^1(E,\mathcal{A},\mu)$ 且 $f \le g$，那么 $\int f\,\mathrm{d}\mu \le \int g\,\mathrm{d}\mu$.
4. 如果 $f \in \mathcal{L}^1(E,\mathcal{A},\mu)$，$g: E \to [0,\infty]$ 是非负可测函数使得 $f = g$, μ a.e.，那么 $g \in \mathcal{L}^1(E,\mathcal{A},\mu)$ 且 $\int f\,\mathrm{d}\mu = \int g\,\mathrm{d}\mu$.
5. 令 $(F,\mathcal{B})$ 是可测空间，$\varphi: E \to F$ 是可测映射. 令 ν 是 μ 在 φ 下的推前. 那么，对于任意可测函数 $h: F \to \mathbb{R}$，h 是 ν-可积的当且仅当 $h \circ \varphi$ 是 μ-可积的，并且我们有

$$\int_E h(\varphi(x))\mu(\mathrm{d}x) = \int_F h(y)\nu(\mathrm{d}y).$$

延拓到复值函数 令 $f : E \to \mathbb{C}$ 是可测函数 (这等价于实部 $\mathrm{Re}\, f$ 和虚部 $\mathrm{Im}\, f$ 都是可测的), 如果

$$\int |f|\, \mathrm{d}\mu < \infty,$$

那么我们说 f 是可积的. 在这种情况下, 我们定义

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

记复值可积函数空间为 $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{A}, \mu)$, 那么上述命题对复值可积函数依然成立. 并且 $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{A}, \mu)$ 是一个复向量空间.

定理 2.15 (控制收敛定理). 令 $(f_n)_{n\in\mathbb{N}}$ 是 $\mathcal{L}^1(E, \mathcal{A}, \mu)$ 中的一列函数 (对于 $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{A}, \mu)$ 也成立), 如果:

1. 存在可测函数 $f : E \to \mathbb{R}$ 使得
$$f_n(x) \to f(x), \quad \mu \text{ a.e.}$$
2. 存在非负可测函数 g 使得 $\int g\, \mathrm{d}\mu < \infty$, 并且对于每个 $n \in \mathbb{N}$, 都有
$$|f_n(x)| \le g(x), \quad \mu \text{ a.e.}$$

那么 $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ 且我们有

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

Proof. 我们首先将两个条件中的几乎处处去掉, 证明结论成立. 由于 $|f_n| \le g$, 所以 $|f| = \lim |f_n| \le g$, 所以 $\int |f|\, \mathrm{d}\mu \le \int g\, \mathrm{d}\mu < \infty$, 故 $f \in L^1(E, \mathcal{A}, \mu)$. 由于 $|f - f_n| \le 2g$ 以及 $\lim |f - f_n| = 0$, 根据 Fatou 引理, 有

$$\liminf \int \big(2g - |f - f_n|\big)\, \mathrm{d}\mu \ge \int \big(2g - \limsup |f - f_n|\big)\, \mathrm{d}\mu = \int 2g\, \mathrm{d}\mu,$$

再根据积分的线性性, 有

$$\int 2g\, \mathrm{d}\mu - \limsup \int |f - f_n|\, \mathrm{d}\mu \ge \int 2g\, \mathrm{d}\mu,$$

这表明

$$\limsup \int |f - f_n|\, \mathrm{d}\mu = 0,$$

所以 $\lim \int |f - f_n|\, \mathrm{d}\mu$ 存在且为 0. 最后, 我们有

$$\left| \int f\, \mathrm{d}\mu - \int f_n\, \mathrm{d}\mu \right| \le \int |f - f_n|\, \mathrm{d}\mu \to 0,$$

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

现在我们证明几乎处处的情况. 记

$$A = \big\{ x \in E \mid f_n(x) \to f(x), |f_n(x)| \le g(x) \big\},$$

那么 A 可测且条件表明 $\mu(A^c)=0$. 定义

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

于是在几乎处处的意义下有 $f_n=\tilde{f}_n$ 以及 $f=\tilde{f}$, 所以 $\int f_n\,\mathrm{d}\mu=\int\tilde{f}_n\,\mathrm{d}\mu$, $\int f\,\mathrm{d}\mu=\int\tilde{f}\,\mathrm{d}\mu$ 以及 $\int|f-f_n|\,\mathrm{d}\mu=\int|\tilde{f}-\tilde{f}_n|\,\mathrm{d}\mu$. 对 $\tilde{f}_n$ 和 $\tilde{f}$ 应用上面的结论即可. □

2.3 含参积分

我们考虑带有一个参数的函数的积分. 设 (U,d) 是一个度量空间, 参数位于这个空间中.

定理 2.16 (含参积分的连续性). 令 $f:U\times E\to\mathbb{R}$ (or $\mathbb{C}$), $u_0\in U$. 假设:

1. 对于每个 $u\in U$, 函数 $x\mapsto f(u,x)$ 可测;
2. $\mu(\mathrm{d}x)$ a.e., 函数 $u\mapsto f(u,x)$ 在 u_0 处连续;
3. 存在函数 $g\in\mathcal{L}^1_+(E,\mathcal{A},\mu)$ 使得任取 $u\in U$ 有

$$|f(u,x)|\le g(x)\quad \mu(\mathrm{d}x)\text{ a.e.}$$

那么函数 $F(u)=\int f(u,x)\mu(\mathrm{d}x)$ 是良好定义的且在 u_0 处连续.

Proof. 条件 (1) 保证了 $x\mapsto f(u,x)$ 是可测的, 所以 $F(u)$ 是良好定义的. 设 $(u_n)_{n\in\mathbb{N}}$ 是任意趋于 u_0 的点列, 那么对于几乎处处的 x, $u\mapsto f(u,x)$ 连续表明 $f(u_n,x)\to f(u,x)$, 再根据条件 (3) 和控制收敛定理, 就有

$$F(u_0)=\int\lim_{n\to\infty}f(u_n,x)\mu(\mathrm{d}x)=\lim_{n\to\infty}\int f(u_n,x)\mu(\mathrm{d}x)=\lim_{n\to\infty}F(u_n),$$

这就表明 $F(u)$ 在 u_0 处连续. □

例 2.17.

1. 令 μ 是 $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ 上的扩散测度并且 $\varphi\in\mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu)$. 定义函数 $F:\mathbb{R}\to\mathbb{R}$ 为

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

   令 $f(u,x)=\mathbf{1}_{(-\infty,u]}(x)\varphi(x)$, 那么任取 $u,x\in\mathbb{R}$ 有 $|f(u,x)|\le|\varphi(x)|$. 此外, 对于任意 $u_0\in\mathbb{R}$, 当 $x\in\mathbb{R}\setminus\{u_0\}$ 时, 函数 $u\mapsto f(u,x)$ 在 u_0 处连续, 由于 μ 是扩散测度, 即 $\mu(\{u_0\})=0$, 所以满足条件 (2). 根据上述定理, F 在 $\mathbb{R}$ 上连续.
2. **Fourier 变换**. 令 λ 表示 $\mathbb{R}$ 上的 Lebesgue 测度. 如果 $\varphi\in\mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$, 定义函数 $\hat{\varphi}:\mathbb{R}\to\mathbb{C}$ 为:

$$\hat{\varphi}(u)=\int e^{iux}\varphi(x)\lambda(\mathrm{d}x),$$

   根据上面的定理, $\hat{\varphi}$ 是连续函数. 函数 $\hat{\varphi}$ 被称为 φ 的 **Fourier 变换**. 在概率论中, 我们经常会考虑有限测度的 Fourier 变换. 如果 μ 是 $\mathbb{R}$ 上的有限测度, 定义 μ 的 Fourier 变换为

$$\hat{\mu}(u)=\int e^{iux}\mu(\mathrm{d}x)\quad u\in\mathbb{R}.$$

此时 $|e^{iux}| \le \mathbf{1}_{\mathbb{R}}$ 是可积函数, 所以 $\hat{\mu}$ 是连续函数.

3. **卷积**. 令 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, $h : \mathbb{R} \to \mathbb{R}$ 是有界连续函数, 那么定义函数 $h * \varphi$ 为

$$h * \varphi(u) = \int h(u - x)\varphi(x)\lambda(\mathrm{d}x),$$

这是一个连续函数.

下面我们叙述含参积分的可微性. 令 $I \subseteq \mathbb{R}$ 是开区间.

定理 2.18. 考虑函数 $f : I \times E \to \mathbb{R}$，$u_0 \in I$，假设

1. 对于每个 $u \in I$，函数 $x \mapsto f(u, x)$ 是可积函数；
2. $\mu(\mathrm{d}x)$ a.e.，函数 $u \mapsto f(u, x)$ 在 u_0 处可导，导数记为
$$\frac{\partial f}{\partial u}(u_0, x);$$
3. 存在函数 $g \in \mathcal{L}^1_+(E, \mathcal{A}, \mu)$ 使得对于任意 $u \in I$ 有
$$|f(u, x) - f(u_0, x)| \le g(x)|u - u_0|, \quad \mu(\mathrm{d}x) \text{ a.e.}$$

那么函数 $F(u) = \int f(u, x)\mu(\mathrm{d}x)$ 在 u_0 处可导，并且

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

注释 2.19. (2) 中的 $\partial f / \partial u(u_0, x)$ 只在某个 μ-零可测集 H 的补集上有定义. 当 $x \in H$ 的时候，我们定义其为 0，这样就可以保证 $x \mapsto \partial f / \partial u(u_0, x)$ 在整个 E 上都有定义. 利用下面证明中的 φ_n 和 引理 1.20，$x \mapsto \partial f / \partial u(u_0, x)$ 是可测的. 再利用条件 (3)，这个函数也是可积的，这使得 $F'(u_0)$ 有意义. 我们也可以把这个定理推广到复数情形，只需要对实部和虚部分别应用该定理即可.

Proof. 令 $(u_n)_{n\ge1}$ 是 $I \setminus \{u_0\}$ 中收敛到 u_0 的序列, 令

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

那么对于几乎处处的 x, 有

$$\lim_{n\to\infty} \varphi_n(x) = \frac{\partial f}{\partial u}(u_0, x).$$

根据条件 (3), 对于任意 n, 都有

$$|\varphi_n(x)| \le g(x), \quad \mu(\mathrm{d}x) \text{ a.e.}.$$

根据控制收敛定理, 有

$$\lim_{n\to\infty} \int \varphi_n(x)\mu(\mathrm{d}x) = \int \frac{\partial f}{\partial u}(u_0, x)\mu(\mathrm{d}x).$$

注意到
$$\int \varphi_n(x)\mu(\mathrm{d}x) = \frac{F(u_n) - F(u_0)}{u_n - u_0},$$
所以
$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x)\mu(\mathrm{d}x).$$
□

在许多应用中, 条件 (2) 和 (3) 通常以下更易验证的形式出现:

(2′) $\mu(\mathrm{d}x)$ a.e., 函数 $u \mapsto f(u,x)$ 在 I 上可微;

(3′) 存在函数 $g \in \mathcal{L}^1_+(E, \mathcal{A}, \mu)$ 使得对于任意 $u \in I$ 有
$$\left|\frac{\partial f}{\partial u}(u,x)\right| \le g(x), \quad \mu(\mathrm{d}x) \text{ a.e.}.$$

例 2.20.

1. 令 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ 使得
$$\int |x\varphi(x)|\lambda(\mathrm{d}x) < \infty.$$
那么 Fourier 变换 $\hat{\varphi}(u)$ 在 $\mathbb{R}$ 中可微, 并且
$$\hat{\varphi}'(u) = \int ixe^{iux}\varphi(x)\lambda(\mathrm{d}x).$$
2. 令 $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, $h : \mathbb{R} \to \mathbb{R}$ 是连续可微函数, 并且 h 和 h' 都有界, 那么卷积 $h * \varphi$ 在 $\mathbb{R}$ 上可微, 并且
$$(h * \varphi)' = h' * \varphi.$$
这个论证可以递归下去. 例如, 如果 h 是无穷可微且紧支的, 那么 $h * \varphi$ 也是无穷可微的.

Exercise 2.1. 计算
$$\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x}\, \mathrm{d}x.$$
令 $\alpha \in \mathbb{R}$, 证明极限
$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n x^{\alpha-1}\, \mathrm{d}x$$
在 $[0, \infty]$ 上存在, 且极限值有限当且仅当 $\alpha > 0$.

Proof. 令
$$f_n(x) = \mathbf{1}_{[0,n]}(x) \left(1 + \frac{x}{n}\right)^n e^{-2x},$$
那么对于任意 $x \ge 0$, 都有 $\lim_{n\to\infty} f_n(x) = e^{-x}$. 此外, 对于任意 $x \ge 0$, 还有
$$|f_n(x)| \le (e^{x/n})^n e^{-2x} = e^{-x},$$
所以可以选控制函数为 $g(x) = e^{-x}$. 根据控制收敛定理, 有
$$\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x}\, \mathrm{d}x = \int_0^\infty e^{-x}\, \mathrm{d}x = 1.$$
□

3

测度的构造

3.1 外测度

定义 3.1. 令 E 是集合，映射 $\mu^* : \mathcal{P}(E) \to [0,\infty]$ 如果满足：

1. $\mu^*(\emptyset) = 0$；
2. $A \subseteq B \Rightarrow \mu^*(A) \le \mu^*(B)$；
3. (σ-次可加性) 对于 $\mathcal{P}(E)$ 中的一列子集 $(A_k)_{k\in\mathbb{N}}$，有

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

那么我们说 μ^* 是一个**外测度**.

外测度的要求不如测度严格, 首先 σ-可加性被替换为 σ-次可加性, 其次外测度是在幂集 $\mathcal{P}(E)$ 上定义的, 而测度只能在 σ-域上定义.

我们本节的目标是从外测度 μ^* 开始, 在某个 σ-域 $\mathcal{M}(\mu^*)$ 上构造一个测度. 从现在开始, 我们固定一个外测度 μ^*.

定义 3.2. 对于 E 的子集 B，如果任取 $A \subseteq E$，都有

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

那么我们说 B 是 **μ^*-可测的**. 用 $\mathcal{M}(\mu^*)$ 表示所有 μ^*-可测的子集构成的子集族.

注释 3.3. 根据 σ-次可加性，总是有

$$\mu^*(A) \le \mu^*(A\cap B) + \mu^*(A\cap B^c),$$

所以要验证子集 B 是 μ^*-可测的，只需要说明反向的不等式即可.

外测度的定义来源于 $\mathbb{R}$ 中定义集合长度的想法，也即用一堆开区间覆盖这个集合，然后取这些开区间长度之和的下确界，所以叫“外”测度，有从外面逼近的含义. μ^*-可测的定义看似奇怪, 其实可以把 A 理解为一个“测试集”, 用集合 B 把 A 切成两部分 $A\cap B$ 和 $A\cap B^c$, 如果这两部分的测度之和等于 A 的测度, 那么就说 B 是可测的, 也就是比较好的集合. 如果 B 的形状非常复杂, 那么切出来的两部分可能会产生很多复杂的碎片, 导致体积之和增大, 这时说 B 是不可测的.

定理 3.4.

1. $\mathcal{M}(\mu^*)$ 是 σ-域，并且其包含所有的满足 $\mu^*(B)=0$ 的子集 $B\subseteq E$.
2. μ^* 在 $\mathcal{M}(\mu^*)$ 上的限制是一个测度.

Proof. (1) 如果 $\mu^*(B)=0$, 那么对于任意 $A\subseteq E$, 由于 $A\cap B\subseteq B$, 所以 $\mu^*(A\cap B)\le\mu^*(B)=0$, 所以

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

这就表明 $B\in\mathcal{M}(\mu^*)$.

显然 $\emptyset\in\mathcal{M}(\mu^*)$, 并且 $B\in\mathcal{M}(\mu^*)$ 蕴含 $B^c\in\mathcal{M}(\mu^*)$. 下面证明 $\mathcal{M}(\mu^*)$ 对可数并运算封闭. 我们首先证明 $\mathcal{M}(\mu^*)$ 对有限并封闭. 令 $B_1,B_2\in\mathcal{M}(\mu^*)$, 任取 $A\subseteq E$, 则 $B_1\in\mathcal{M}(\mu^*)$ 表明

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

再利用 $B_2\in\mathcal{M}(\mu^*)$, 就得到

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

所以 $B_1\cup B_2\in\mathcal{M}(\mu^*)$. 因为 $\mathcal{M}(\mu^*)$ 对有限并和补运算封闭, 所以对有限交也封闭. 因此, 如果 $B,B'\in\mathcal{M}(\mu^*)$, 那么 $B'\setminus B=B'\cap B^c\in\mathcal{M}(\mu^*)$.

由于 $\mathcal{M}(\mu^*)$ 对差集运算封闭, 所以要证明可数并封闭, 只需要证明 $\mathcal{M}(\mu^*)$ 中一列不相交的集合 $(B_k)_{k\in\mathbb{N}}$ 的并在 $\mathcal{M}(\mu^*)$ 中. 采用归纳法. 对于任意整数 $m\in\mathbb{N}$ 和任意 $A\subseteq E$, 我们证明

$$\mu^*(A)=\sum_{k=1}^{m}\mu^*(A\cap B_k)+\mu^*\left(A\cap\bigcap_{k=1}^{m}B_k^c\right).\tag{3.1}$$

当 $m=1$ 时, 这就是 $B_1\in\mathcal{M}(\mu^*)$ 的定义. 假设 m 步成立, B_k 互不相交表明对于 $1\le k\le m$ 有 $B_{m+1}\subseteq B_k^c$, 那么

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

再根据归纳假设, 就得到式 (3.1) 对 $m+1$ 也成立. 利用式 (3.1), 对于任意 $A\subseteq E$, 都有

$$\mu^*(A)\ge\sum_{k=1}^{m}\mu^*(A\cap B_k)+\mu^*\left(A\cap\bigcap_{k=1}^{\infty}B_k^c\right).$$

然后令 $m \to \infty$, 就得到

$$\begin{aligned}\mu^*(A) &\geq \sum_{k=1}^{\infty}\mu^*(A\cap B_k)+\mu^*\left(A\cap\bigcap_{k=1}^{\infty}B_k^c\right) \\ &\geq \mu^*\left(A\cap\bigcup_{k=1}^{\infty}B_k\right)+\mu^*\left(A\cap\bigcap_{k=1}^{\infty}B_k^c\right),\end{aligned} \tag{3.2}$$

这就表明 $\bigcup_{k=1}^{\infty}B_k\in\mathcal{M}(\mu^*)$.

(2) 显然 $\mu^*(\emptyset)=0$. 下面证明 σ-可加性. 令 $(B_k)_{k\in\mathbb{N}}$ 是 $\mathcal{M}(\mu^*)$ 中的一列互不相交的集合, 那么根据 (3.2), 取 $A=\bigcup_{k\in\mathbb{N}}B_k$, 就得到

$$\mu^*\left(\bigcup_{k\in\mathbb{N}}B_k\right)\geq\sum_{k=1}^{\infty}\mu^*(B_k).$$

结合外测度的 σ-次可加性, 就得到等号成立, 所以 μ^* 在 $\mathcal{M}(\mu^*)$ 上是一个测度. □

3.2 Lebesgue 测度

对于任意子集 $A\subseteq\mathbb{R}$, 定义

$$\lambda^*(A)=\inf\left\{\sum_{i\in\mathbb{N}}(b_i-a_i)\mid A\subseteq\bigcup_{i\in\mathbb{N}}(a_i,b_i)\right\}.$$

注意这个下确界的取值范围为 $[0,\infty]$: 如果 A 无界, 那么将会得到 ∞.

定理 3.5.

1. λ^* 是 $\mathbb{R}$ 上的一个外测度.
2. σ-域 $\mathcal{M}(\lambda^*)$ 包含 $\mathcal{B}(\mathbb{R})$.
3. 对于任意实数 $a\leq b$, $\lambda^*([a,b])=\lambda^*((a,b))=b-a$.

注释 3.6. λ^* 在 $\mathcal{B}(\mathbb{R})$ 上的限制被称为 $\mathbb{R}$ 上的 **Lebesgue 测度**, 记为 λ. 根据单调类定理的推论 1.24, 这是 $\mathcal{B}(\mathbb{R})$ 上唯一一个满足 $\lambda([a,b])=b-a$ 的测度.

Proof. (1) 显然 $\lambda^*(\emptyset)=0$ 并且 $A\subseteq B$ 表明 $\lambda^*(A)\leq\lambda^*(B)$. 下面证明 σ-次可加性. 任取 $\mathbb{R}$ 的一列子集 $(A_n)_{n\in\mathbb{N}}$, 不妨假设每个 $\lambda^*(A_n)<\infty$. 任取 $\varepsilon>0$, 对于每个 A_n, 都存在一列开区间 $(a_i^{(n)},b_i^{(n)})$ 使得

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

注意到所有的开区间 $(a_i^{(n)},b_i^{(n)})$ $(i,n\in\mathbb{N})$ 构成了 $\bigcup_{n\in\mathbb{N}}A_n$ 的一个可数开覆盖, 所以

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

由于 ε 的任意性, 所以 λ^* 满足 σ-次可加性.

(2) 因为 $\mathcal{M}(\lambda^*)$ 是一个 σ-域, 所以只需要证明其包含生成 Borel σ-域的某个集合族即可, 我们选择所有区间 $(-\infty, \alpha]$. 固定 $\alpha \in \mathbb{R}$, 记 $B = (-\infty, \alpha]$. 我们需要证明对于任意 $A \subseteq \mathbb{R}$, 都有

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

令 $\big((a_i, b_i)\big)_{i \in \mathbb{N}}$ 是 A 的一个开覆盖, $\varepsilon > 0$. 那么区间 $\big(a_i \wedge \alpha, (b_i \wedge \alpha) + \varepsilon 2^{-i}\big)$ 覆盖 $A \cap B$, 并且区间 $(a_i \vee \alpha, b_i \vee \alpha)$ 覆盖 $A \cap B^c$, 所以

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

将上面两式相加, 就有

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

因为 ε 的任意性, 就有

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

由于 $\lambda^*(A)$ 是所有开覆盖长度之和的下确界, 所以就得到需要的结论.

(3) 根据定义, 立马得到

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

假设 $[a, b] \subseteq \bigcup_{i \in \mathbb{N}} (a_i, b_i)$. 根据紧性, 存在 N 使得 $[a, b] \subseteq \bigcup_{i=1}^{N} (a_i, b_i)$. 那么

$$b - a \leq \sum_{i=1}^{N} (b_i - a_i) \leq \sum_{i \in \mathbb{N}} (b_i - a_i).$$

这就表明 $b - a \leq \lambda^*([a, b])$. 最后, 注意到 $\lambda^*(\{a\}) = \lambda^*(\{b\}) = 0$ 就得到 $\lambda^*((a, b)) = \lambda^*([a, b]) = b - a$. □

下面我们将 Lebesgue 测度推广到 $\mathbb{R}^d$ 上. 我们说 $\mathbb{R}^d$ 中的一个开矩形 P 是形如以下形式的集合:

$$P = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d),$$

对应的, 如果每个区间都是闭区间, 那么我们说 P 是一个闭矩形. 定义 P 的体积为

$$\operatorname{vol}(P) = \prod_{i=1}^{d} (b_i - a_i).$$

然后, 对于每个子集 $A \subseteq \mathbb{R}^d$, 定义

$$\lambda^*(A) = \inf \left\{ \sum_{i \in \mathbb{N}} \operatorname{vol}(P_i) \mid A \subseteq \bigcup_{i \in \mathbb{N}} P_i \right\},$$

即考虑所有覆盖 A 的开矩形体积之和的下确界. 现在我们可以推广 定理 3.5 的结论到 $\mathbb{R}^d$ 上.

定理 3.7.

1. λ^* 是 $\mathbb{R}^d$ 上的一个外测度.
2. σ-域 $\mathcal{M}(\lambda^*)$ 包含 $\mathcal{B}(\mathbb{R}^d)$.
3. 对于任意开或者闭矩形 P, 有 $\lambda^*(P) = \mathrm{vol}(P)$.

Proof. (1) 这一点和 $d = 1$ 时的证明完全相同.

(2) 只需要证明, 如果 $A \subseteq \mathbb{R}^d$ 形如

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

其中 $a \in \mathbb{R}$, 那么 $A \in \mathcal{M}(\lambda^*)$. 证明和 $d = 1$ 时的做法也完全类似.

(3) 根据定义, 立马得到 $\lambda^*(P) \le \mathrm{vol}(P)$. 下面证明反向不等式. 同样利用 P 的紧性, 只需要证明: 如果 P 是闭矩形, 并且 $P \subseteq \bigcup_{i=1}^n P_i$, 其中 P_i 是开矩形, 那么

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

记

$$C_n^{k_1 \dots k_n} = [k_1 2^{-n}, (k_1+1)2^{-n}] \times \cdots \times [k_d 2^{-n}, (k_d+1)2^{-n}], \quad k_1, \dots, k_d \in \mathbb{Z}.$$

那么 $\{C_n^{k_1 \dots k_d} \mid k_1, \dots, k_d \in \mathbb{Z}\}$ 构成了 $\mathbb{R}^d$ 的一个划分, 并且每个 $C_n^{k_1 \dots k_d}$ 的体积为 2^{-nd}. 固定 n, 假设 P 和 c_n 个 $C_n^{k_1 \dots k_d}$ 相交, $\bigcup_{i=1}^n P_i$ 和 c_n' 个 $C_n^{k_1 \dots k_d}$ 相交, 那么 $c_n \le c_n'$, 所以

$$c_n 2^{-nd} \le c_n' 2^{-nd}.$$

取 $n \to \infty$, 就得到 $\mathrm{vol}(P) \le \sum_{i=1}^n \mathrm{vol}(P_i)$. □

通常, 我们用

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\lambda(\mathrm{d}x)$$

表示 f 相对于 Lebesgue 测度的积分. 当 $d = 1$ 以及 $a \le b$ 时, 我们也用

$$\int_a^b f(x)\,\mathrm{d}x = \int_{[a,b]} f(x)\lambda(\mathrm{d}x).$$

一个自然的问题是 $\mathcal{M}(\lambda^*)$ 比 $\mathcal{B}(\mathbb{R}^d)$ 大多少. 从某种意义上, 这两个 σ-域的差距不算很大.

命题 3.8. 令 $(E, \mathcal{A}, \mu)$ 是测度空间. 定义 μ-可忽略集的集合为

$$\mathcal{N} = \left\{A \in \mathcal{P}(E) \;\middle|\; \exists B \in \mathcal{A}, A \subseteq B \text{ and } \mu(B) = 0\right\}.$$

我们把包含 $\mathcal{A}$ 和 $\mathcal{N}$ 的最小 σ-域称为 $\mathcal{A}$ 相对于 μ 的**完备化**, 记作 $\bar{\mathcal{A}}$. 此时, $(E, \bar{\mathcal{A}})$ 上存在唯一的测度使得其在 $\mathcal{A}$ 上的限制等于 μ.

Proof. 令

$$\mathcal{B} = \{A \in \mathcal{P}(E) \mid \exists B, B' \in \mathcal{A}, B \subseteq A \subseteq B',\ \mu(B' \setminus B) = 0\}.$$

首先我们证明 $\bar{\mathcal{A}} = \mathcal{B}$. 首先, 直接验证可知 $\mathcal{B}$ 是一个 σ-域. 显然还有 $\mathcal{A} \subseteq \mathcal{B}$ 以及 $\mathcal{N} \subseteq \mathcal{B}$, 所以 $\bar{\mathcal{A}} \subseteq \mathcal{B}$. 反之, 任取 $A \in \mathcal{B}$, 那么存在 $B, B' \in \mathcal{A}$ 使得 $B \subseteq A \subseteq B'$ 且 $\mu(B' \setminus B) = 0$. 此时 $A \setminus B \subseteq B' \setminus B$ 且 $\mu(B' \setminus B) = 0$, 所以 $A \setminus B \in \mathcal{N}$, 所以 $A = B \cup (A \setminus B) \in \bar{\mathcal{A}}$.

下面我们把 μ 延拓到 $\mathcal{B}$ 上. 如果 $A \in \bar{\mathcal{A}} = \mathcal{B}$, 且 $B, B' \in \mathcal{A}$ 是满足 $B \subseteq A \subseteq B'$ 且 $\mu(B' \setminus B) = 0$ 的集合, 那么定义 $\mu(A) = \mu(B) = \mu(B')$. 这种定义不依赖于 B 和 B' 的选取: 如果 C, C' 也有这样的性质, 那么同时有 $\mu(C) \le \mu(B')$ 以及 $\mu(B) \le \mu(C')$, 所以必须有 $\mu(B) = \mu(B') = \mu(C) = \mu(C')$. 最后只需要验证 μ 在 $\mathcal{B}$ 上确实是一个测度. 假设 (A_n) 是 $\mathcal{B}$ 中一列不相交的集合, 那么对于每个 n 都可以选取一个 $B_n \in \mathcal{A}$ 使得 $B_n \subseteq A_n$ 并且 $A_n \setminus B_n$ 是可忽略集. 此时, 有

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \mu\Big(\bigcup_{n\in\mathbb{N}} B_n\Big) = \sum_{n\in\mathbb{N}} \mu(B_n) = \sum_{n\in\mathbb{N}} \mu(A_n). \qquad \square$$

注释 3.9. 令 f, g 是两个定义在 E 上的实值函数. 假设 g 是 $\mathcal{A}$-可测的并且 $f = g$, μ a.e., 这里指的是集合 $\{x \in E \mid f(x) \neq g(x)\}$ 是 μ-可忽略集. 那么 f 是 $\bar{\mathcal{A}}$-可测的. 根据定义, 存在 $C \in \mathcal{A}$ 使得 $\mu(C) = 0$ 并且 $f(x) = g(x)$ 对于 $x \notin C$. 此时, 任取 Borel 集 $H \subseteq \mathbb{R}$, 有 $g^{-1}(H) \setminus C \subseteq f^{-1}(H) \subseteq g^{-1}(H) \cup C$, 所以 $f^{-1}(H) \in \bar{\mathcal{A}}$.

命题 3.10. σ-域 $\mathcal{M}(\lambda^*)$ 等于 $\mathcal{B}(\mathbb{R}^d)$ 相对于 Lebesgue 测度的完备化 $\overline{\mathcal{B}}(\mathbb{R}^d)$.

Proof. $\overline{\mathcal{B}}(\mathbb{R}^d) \subseteq \mathcal{M}(\lambda^*)$ 是简单的. 如果 A 是可忽略集, 那么存在 $B \in \mathcal{B}(\mathbb{R}^d)$ 使得 $\lambda(B) = 0$ 且 $A \subseteq B$. 于是 $\lambda^*(A) \le \lambda^*(B) = 0$, 根据定理 3.4, $A \in \mathcal{M}(\lambda^*)$.

反之, 令 $A \in \mathcal{M}(\lambda^*)$, 我们需要证明 $A \in \overline{\mathcal{B}}(\mathbb{R}^d)$. 不失一般性, 我们假设存在某个 $K > 0$ 使得 $A \subseteq (-K, K)^d$, 否则我们可以把 A 写为可数个集合 $A \cap (-n, n)^d$ 的并集. 那么 $\lambda^*(A) < \infty$, 并且对于每个 $n \ge 1$, 我们可以找到可数个开矩形 $(P_i^n)_{i\in\mathbb{N}}$ 使得

$$A \subseteq \bigcup_{i\in\mathbb{N}} P_i^n, \quad \sum_{i\in\mathbb{N}} \mathrm{vol}(P_i^n) \le \lambda^*(A) + \frac{1}{n}.$$

我们还可以假设每个 $P_i^n \subseteq (-K, K)^d$, 否则将 P_i^n 替换为 $P_i^n \cap (-K, K)^d$ (两个开矩形的交依然是开矩形) 即可. 令

$$B_n = \bigcup_{i\in\mathbb{N}} P_i^n, \quad B = \bigcap_{n=1}^{\infty} B_n.$$

那么 $B \in \mathcal{B}(\mathbb{R}^d)$ 并且 $A \subseteq B$. 对于每个 n, 还有

$$\lambda(B) \le \lambda(B_n) \le \sum_{i\in\mathbb{N}} \mathrm{vol}(P_i^n) \le \lambda^*(A) + \frac{1}{n},$$

这就表明 $\lambda(B) \le \lambda^*(A)$, 同时还有 $\lambda(B) = \lambda^*(B) \ge \lambda^*(A)$, 所以 $\lambda(B) = \lambda^*(A)$. 然后把 A 替换为 $(-K,K)^d \setminus A$, 同样的论证表明存在 $B' \in \mathcal{B}(\mathbb{R}^d)$ 使得 $(-K,K)^d \setminus A \subseteq B'$ 且 $\lambda(B') = \lambda^*((-K,K)^d \setminus A)$. 再令 $C = (-K,K)^d \setminus B'$, 那么 $C \subseteq A \subseteq B$ 且 $\lambda(C) = \lambda^*(A) = \lambda(B)$. 由于 $\lambda(B \setminus C) = 0$, 所以 $A \in \overline{\mathcal{B}}(\mathbb{R}^d)$. □

命题 3.11. $\mathbb{R}^d$ 上的 Lebesgue 测度是平移不变的：对于任意 $A \in \mathcal{B}(\mathbb{R}^d)$ 和 $x \in \mathbb{R}^d$，有 $\lambda(A+x) = \lambda(A)$. 反过来，如果 μ 是 $\mathcal{B}(\mathbb{R}^d)$ 上的一个测度，其在有界集上的值有限并且平移不变，那么存在常数 $c \ge 0$ 使得 $\mu = c\lambda$.

Proof. 对于 $x \in \mathbb{R}^d$, 记 σ_x 是平移映射 $\sigma_x(y) = y - x$. 那么推前测度 $\sigma_x(\lambda)$ 满足

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

首先, 对于任意矩形 A, 肯定有 $\sigma_x(\lambda)(A) = \lambda(A)$, 所以根据单调类定理的推论 1.24, 对于任意 $A \in \mathcal{B}(\mathbb{R}^d)$, 都有 $\sigma_x(\lambda)(A) = \lambda(A)$.

□

3.3 $\mathbb{R}$ 上的有限测度和 Stieltjes 积分

下面的定理描述了 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的所有有限测度.

定理 3.12.

1. 令 μ 是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的一个有限测度. 对于每个 $x \in \mathbb{R}$，令

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

   函数 F_μ 是递增的，有界的，右连续的，并且使得 $F_\mu(-\infty) = 0$ (这意味着 $F_\mu(x)$ 在 $x \to -\infty$ 时为零).
2. 反之，若函数 $F : \mathbb{R} \to \mathbb{R}_+$ 是递增的，有界的，右连续的，并且使得 $F(-\infty) = 0$，那么唯一存在 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上的一个有限测度 μ 使得 $F = F_\mu$.

Proof. (1) 显然 F_μ 是递增的并且有界的. 下面证明右连续性. 令 (x_n) 是一个递减到 x 的数列, 那么

$$F_\mu(x_n) = \mu((-\infty, x_n]) \xrightarrow[n\to\infty]{} \mu\left(\bigcap_{n=1}^{\infty}(-\infty, x_n]\right) = \mu((-\infty, x]) = F_\mu(x).$$

类似地, 如果 (x_n) 是递减趋于 $-\infty$ 的数列, 由于 $\bigcap_n(-\infty, x_n] = \emptyset$, 所以 $F_\mu(x_n) \to 0$.

(2) 因为集合族 $\mathcal{C} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ 对有限交封闭并且生成了 $\mathcal{B}(\mathbb{R})$, 所以 μ 的唯一性由单调类定理的推论 1.24 保证. 下面证明存在性. 对于每个 $A \subseteq \mathbb{R}$, 令

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

注意这里使用左开右闭区间覆盖更加方便. 类似于 Lebesgue 测度的证明, 可以验证 μ^* 是 $\mathbb{R}$ 上的一个外测度并且区间 $(-\infty,\alpha]$ 在 $\mathcal{M}(\mu^*)$ 中, 所以 $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{M}(\mu^*)$. 于是我们可以把 μ^* 限制到 $\mathcal{B}(\mathbb{R})$ 上, 得到一个测度 μ. 显然, μ 是有限测度.

最后只需要验证 $\mu((-\infty,x]) = F(x)$. 只需要证明 $a < b$ 的时候有 $\mu((a,b]) = F(b) - F(a)$. 根据 μ^* 的定义, 立马得到 $\mu((a,b]) \le F(b) - F(a)$.

反过来, 令 $((x_i, y_i])$ 是 $(a,b]$ 的可数覆盖, $\varepsilon \in (0, b-a)$. 对于每个 $i \in \mathbb{N}$, 我们可以找到 $y_i' > y_i$ 使得 $F(y_i') \le F(y_i) + \varepsilon 2^{-i}$. 根据紧性, 我们可以找到足够大的 N_ε 使得 $[a+\varepsilon, b]$ 可以被有限个 $((x_i, y_i'))_{i\in\{1,\dots,N_\varepsilon\}}$ 覆盖. 我们有

$$\begin{aligned} F(b) - F(a+\varepsilon) &\le \sum_{i=1}^{N_\varepsilon} (F(y_i') - F(x_i)) \le \sum_{i=1}^{\infty} (F(y_i') - F(x_i)) \\ &\le \sum_{i=1}^{\infty} (F(y_i) - F(x_i)) + \varepsilon. \end{aligned}$$

因为 ε 是任意的并且 $\varepsilon \to 0$ 的时候有 $F(a+\varepsilon) \to F(a)$, 所以就得到

$$F(b) - F(a) \le \sum_{i=1}^{\infty} (F(y_i) - F(x_i)).$$

这就表明 $\mu((a,b]) \ge F(b) - F(a)$. □

令 F 是满足上述条件的函数, μ 是对应的使得 $F = F_\mu$ 的有限测度. 对于任意有界的 Borel 可测函数 $f : \mathbb{R} \to \mathbb{R}$, 我们定义

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

此时把 $\int f(x)\,\mathrm{d}F(x)$ 称为 f 关于 F 的 **Stieltjes 积分**. 这个记号来源于下面的等式, 对于每个 $a \le b$, 有

$$\int \mathbf{1}_{(a,b]}(x)\,\mathrm{d}F(x) = F(b) - F(a).$$

同时, 我们还有

$$\int \mathbf{1}_{[a,b]}(x)\,\mathrm{d}F(x) = \lim_{n\to\infty} \int \mathbf{1}_{(a-\frac{1}{n},b]}(x)\,\mathrm{d}F(x) = F(b) - F(a-).$$

其中 $F(a-)$ 表示左极限. 特别的, 取 $b = a$, 函数 F 连续当且仅当 μ 是扩散的.

例 3.13. 当 F 连续可导的时候, Stieltjes 积分和 Riemann 积分 $\int f(x)F'(x)\,\mathrm{d}x$ 一致. Stieltjes 积分的威力在于可以处理性质更奇怪的函数. 例如, 考虑 $f : \mathbb{R} \to \mathbb{R}$ 是连续函数, 定义

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

那么显然有 $\mu(\{0\}) = 1$, 并且当 $x \notin A$ 时有 $\mu(A) = 0$, $x \in A$ 时有 $\mu(A) = 1$. 此时, 有

$$\int f(x)\,\mathrm{d}F(x) = f(0).$$

这实际上严格定义了 Dirac δ 函数 δ_0，即 $\mathrm{d}F(x)$ 在 $x=0$ 处放置了一个质量为 1 的点质量.

3.4 Riesz-Markov-Kakutani 表示定理

回顾一个拓扑空间 X 被称为局部紧的，如果 X 中的每个点都有一个紧邻域. 我们考虑局部紧的度量空间 X. X 上的所有紧支实值连续函数空间记为 $C_c(X)$. $C_c(X)$ 上的一个线性泛函 J 指的是从 $C_c(X)$ 到 $\mathbb{R}$ 的线性映射. 如果当 $f\geq 0$ 的时候 $J(f)\geq 0$，那么说 J 是**正定的**.

$(X,\mathcal{B}(X))$ 上的测度 μ 如果对于 X 的每个紧集 K 都有 $\mu(K)<\infty$，那么说 μ 是 **Radon 测度**. 如果 μ 是 Ranon 测度，那么可以定义 $C_c(X)$ 上的线性泛函

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

注意到这个定义是良好的，因为存在紧集 K 和常数 C 使得 $|f|\leq C\mathbf{1}_K$ 并且 $\mu(K)<\infty$. 此外，显然 J 还是正定的.

Riesz-Markov-Kakutani 表示定理给出了反向的结论. 在合适的条件下，任意正定的线性泛函都可以来源于某个 Radon 测度.

定理 3.14. 令 X 是可分的局部紧的度量空间，J 是 $C_c(X)$ 上的一个正定的线性泛函. 那么存在唯一的 Radon 测度 μ 使得

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

测度 μ 还是正则的，也即，对于每个 $A\in\mathcal{B}(X)$，有

$$\begin{aligned}\mu(A)&=\inf\{\mu(U)\mid A\subseteq U,\ U\text{ 是开集}\},\\&=\sup\{\mu(K)\mid K\subseteq A,\ K\text{ 是紧集}\}.\end{aligned}$$

此外，对于 X 的任意开集 U，有

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

例 3.15. 如果 $X=\mathbb{R}$，我们可以取 $J(f)=I(f)$，其中 $I(f)$ 是 Riemann 积分. 可以验证 J 是 $C_c(\mathbb{R})$ 上的一个正定的线性泛函，并且对应的 Radon 测度就是 Lebesgue 测度. Riesz-Markov-Kakutani 表示定理给出了 Lebesgue 测度的另一个构造.

4

L^p 空间

4.1 定义与 Hölder 不等式

在本章中, 我们考虑测度空间 $(E, \mathcal{A}, \mu)$. 对于实数 $p \geq 1$, 我们令 $\mathcal{L}^p(E, \mathcal{A}, \mu)$ 表示所有满足

$$\int |f|^p \,\mathrm{d}\mu < \infty$$

的可测函数 $f : E \to \mathbb{R}$ 构成的空间. 此外, 我们引入 $\mathcal{L}^\infty(E, \mathcal{A}, \mu)$ 表示所有几乎处处有界的可测函数 $f : E \to \mathbb{R}$ 构成的空间, 即存在常数 $C \in \mathbb{R}_+$ 使得

$$|f| \leq C, \ \mu \text{ a.e.}$$

同样, 我们可以引入复值函数空间 $\mathcal{L}^p_{\mathbb{C}}(E, \mathcal{A}, \mu)$ 以及 $\mathcal{L}^\infty_{\mathbb{C}}(E, \mathcal{A}, \mu)$.

对于每个 $p \in [1, \infty]$, 我们可以定义 $\mathcal{L}^p$ 上的一个等价关系:

$$f \sim g \Leftrightarrow f = g, \ \mu \text{ a.e.}$$

于是我们可以考虑商空间

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

也就是说, 我们只考虑几乎处处相等意义上的函数, 如果两个函数几乎处处相等, 那么我们认为这是同一个函数. 通常情况下, 我们把 $L^p(E, \mathcal{A}, \mu)$ 中的元素视为其在 $\mathcal{L}^p(E, \mathcal{A}, \mu)$ 中的任意代表元: 当我们谈及 $L^p(E, \mathcal{A}, \mu)$ 中的函数时, 这意味着我们在 $\mathcal{L}^p(E, \mathcal{A}, \mu)$ 中选取了一个代表元, 并且得到的结论和这个代表元的选取无关.

在没有歧义的情况下, 我们使用 $L^p(\mu)$ 或者 L^p 表示 $L^p(E, \mathcal{A}, \mu)$. 注意到 L^1 就是所有可积函数构成的空间.

对于可测函数 $f : E \to \mathbb{R}$ 和 $p \in [1, \infty)$, 我们定义

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

约定 $\infty^{1/p} = \infty$. 定义

$$\|f\|_\infty = \inf \left\{ C \in [0, \infty] \,\middle|\, |f| \leq C, \ \mu \text{ a.e.} \right\}.$$

显然有 $|f| \le \|f\|_\infty$ 并且 $\|f\|_\infty$ 是 $[0,\infty]$ 中满足这个性质的最小值. 如果 f, g 是两个几乎处处相等的可测函数, 那么有 $\|f\|_p = \|g\|_p$, 所以我们可以针对 $f \in L^p(E, \mathcal{A}, \mu)$ 良好的定义 $\|f\|_p$.

对于 $p, q \in [1,\infty]$, 我们说 p 和 q 是**共轭指数**, 如果

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

特别地, 1 和 ∞ 是共轭的.

定理 4.1 (Hölder 不等式). 令 p, q 是共轭指数，f, g 是两个 $E \to \mathbb{R}$ 的可测函数，那么

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

特别地，如果 $f \in L^p$ 以及 $g \in L^q$，那么 $fg \in L^1$.

Proof. 若 $\|f\|_p = 0$, 那么 $|f| = 0$ μ a.e., 这表明 $\int |fg| \,\mathrm{d}\mu = 0$, 结论显然成立, 所以我们不妨假设 $\|f\|_p > 0$ 以及 $\|g\|_p > 0$. 进一步的, 我们还可以假设 $f \in L^p$ 以及 $g \in L^q$, 否则右边为 ∞ 显然成立.

先假设 $p = 1$ 和 $q = \infty$, 那么

$$\int |fg| \,\mathrm{d}\mu \le \|g\|_\infty \int |f| \,\mathrm{d}\mu = \|f\|_1 \|g\|_\infty .$$

下面假设 $1 < p, q < \infty$.

设 $\alpha \in (0,1)$, 那么对于 $x \in [0,\infty)$ 有不等式

$$x^\alpha - \alpha x \le 1 - \alpha,$$

取 $x = u/v$ $(u \ge 0, v > 0)$, 我们有

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

该不等式在 $v = 0$ 时也成立. 取 $\alpha = 1/p$, $1 - \alpha = 1/q$, 以及

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

那么

$$\frac{|fg|}{\|f\|_p^p \|g\|_q^q} \le \frac{1}{p}\frac{|f|^p}{\|f\|_p^p} + \frac{1}{q}\frac{|g|^q}{\|g\|_q^q},$$

两边积分, 即得

$$\int |fg| \,\mathrm{d}\mu \le \|f\|_p^p \|g\|_q^q .$$

□

推论 4.2 (Cauchy-Schwarz 不等式). 取 $p = q = 2$，即得

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

推论 4.3. 假设 μ 是有限测度，p, q 是共轭指数且 $p > 1$，那么对于任意可测函数 $f : E \to \mathbb{R}$，有

$$\|f\|_1 \leq \mu(E)^{1/q} \|f\|_p ,$$

因此，对于任意 $p \in (1, \infty]$，有 $L^p \subseteq L^1$. 更一般地，对于任意 $1 \leq r < r' < \infty$，有

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

因此，对于任意 $1 \leq p < q \leq \infty$，有 $L^q \subseteq L^p$，特别地，当 μ 是概率测度的时候，还有 $\|f\|_p \leq \|f\|_q$.

Proof. 取 $g = \mathbf{1}_E$, 即得

$$\int |f| \, \mathrm{d}\mu = \int |f \mathbf{1}_E| \, \mathrm{d}\mu \leq \|f\|_p \|\mathbf{1}_E\|_q = \mu(E)^{1/q} \|f\|_p .$$

用 f^r 替代 f, 取 $p = r'/r$, $1/q = 1 - r/r'$, 那么

$$\|f\|_r \leq \mu(E)^{1/r - 1/r'} \|f^r\|_{r'/r}^{1/r} = \mu(E)^{1/r - 1/r'} \|f\|_{r'} .$$ □

定理 4.4 (Jensen 不等式). 假设 μ 是概率测度，$\varphi : \mathbb{R} \to \mathbb{R}_+$ 是凸函数，那么对于每个 $f \in L^1(E, \mathcal{A}, \mu)$，有

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

4.2 Banach 空间 $L^p(E, \mathcal{A}, \mu)$

定理 4.5 (Minkowski 不等式). 令 $p \in [1, \infty]$，$f, g \in L^p(E, \mathcal{A}, \mu)$，那么 $f + g \in L^p(E, \mathcal{A}, \mu)$ 并且

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

Proof. $p = 1$ 时就是绝对值不等式 $|f + g| \leq |f| + |g|$. $p = \infty$ 时也是显然的. 假设 $1 < p < \infty$. 由于

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

所以 $f + g \in L^p(E, \mathcal{A}, \mu)$. 然后, 考虑不等式

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

两边积分, 并且利用 Hölder 不等式, 取 p 和 $q = p/(p-1)$, 即得

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

如果 $\int |f+g|^p \, \mathrm{d}\mu = 0$, 那么定理自然成立, 否则两边同时除以 $\left(\int |f + g|^p \, \mathrm{d}\mu\right)^{(p-1)/p}$, 即得所需不等式. □

定理 4.6 (Reisz). 对于每个 $p \in [1,\infty]$，空间 $L^p(E,\mathcal{A},\mu)$ 配备范数 $f \mapsto \|f\|_p$ 是一个 Banach 空间 (即完备的赋范向量空间).

Proof. 首先考虑 $1 \le p < \infty$. 先验证 $f \mapsto \|f\|_p$ 是一个范数. 我们有

$$\|f\|_p = 0 \Rightarrow \int |f|^p \,\mathrm{d}\mu = 0 \Rightarrow |f| = 0,\ \mu \text{ a.e.}$$

这意味着在 L^p 中有 $f = 0$. 线性性 $\|\lambda f\|_p = |\lambda|\,\|f\|_p$ 是显然的. 三角不等式由 Minkowski 不等式保证.

然后我们说明 L^p 空间是完备的. 令 $(f_n)_{n\ge1}$ 是 L^p 中的一个 Cauchy 列. 我们可以找到一个严格递增的序列 (k_n), 使得对于每个 $n \ge 1$, 有

$$\|f_{k_{n+1}} - f_{k_n}\|_p < 2^{-n}.$$

令 $g_n = f_{k_n}$, 那么有 $\|g_{n+1} - g_n\|_p < 2^{-n}$. 利用单调收敛定理和 Minkowski 不等式, 我们有

$$\begin{aligned}
\int \left(\sum_{n=1}^{\infty} |g_{n+1} - g_n|\right)^p \mathrm{d}\mu &= \lim_{N\to\infty} \uparrow \int \left(\sum_{n=1}^{N} |g_{n+1} - g_n|\right)^p \mathrm{d}\mu \\
&= \lim_{N\to\infty} \uparrow \left\|\sum_{n=1}^{N} (g_{n+1} - g_n)\right\|_p^p \\
&\le \lim_{N\to\infty} \uparrow \left(\sum_{n=1}^{N} \|g_{n+1} - g_n\|_p\right)^p \\
&< \infty,
\end{aligned}$$

这表明

$$\sum_{n=1}^{\infty} |g_{n+1} - g_n| < \infty,\ \mu \text{ a.e.},$$

对于每个使得 $\sum_{n=1}^{\infty} |g_{n+1}(x) - g_n(x)| < \infty$ 的 x, 我们有

$$h(x) = g_1(x) + \sum_{n=1}^{\infty} \big(g_{n+1}(x) - g_n(x)\big) = \lim_{n\to\infty} g_n(x).$$

否则令 $h(x) = 0$. 根据引理 1.20, h 是可测函数. 因为 g_n 几乎处处收敛到 h, 所以几乎处处有 $|h| = \liminf |g_n|$, 利用 Fatou 引理, 我们有

$$\int |h|^p \,\mathrm{d}\mu \le \liminf \int |g_n|^p \,\mathrm{d}\mu \le \sup_{n\ge1} \int |g_n|^p \,\mathrm{d}\mu < \infty,$$

最后一个小于号是因为 Cauchy 列 (f_n) 一定有界. 因此, $h \in L^p$. 再次利用 Fatou 引理, 对于每个 $n \ge 1$, 有

$$\begin{aligned}
\|h - g_n\|_p^p &= \int |h - g_n|^p \,\mathrm{d}\mu \le \liminf_{m\to\infty} \int |g_m - g_n|^p \,\mathrm{d}\mu \\
&= \liminf_{m\to\infty} \|g_m - g_n\|_p^p \le (2^{-n+1})^p,
\end{aligned}$$

最后一个不等号使用了估计

$$\|g_m - g_n\|_p \le \|g_{n+1} - g_n\|_p + \cdots + \|g_m - g_{m-1}\|_p < 2^{-n+1}.$$

于是, 序列 (g_n) 在 L^p 中收敛到 h. 由于 Cauchy 列有收敛子列则一定收敛, 所以 (f_n) 也收敛到 h, 这就完成了 $1 \le p < \infty$ 时的证明.

当 $p = \infty$ 时, 验证 $f \mapsto \|f\|_\infty$ 是一个范数的方法是一样的. 令 $(f_n)_{n\ge1}$ 是 L^∞ 中的 Cauchy 列. 根据 L^∞-范数的定义, 对于每个 $m > n \ge 1$, 存在一个零可测集 $N_{m,n}$ 使得对于 $x \in E \setminus N_{m,n}$, 有 $|f_m(x) - f_n(x)| \le \|f_m - f_n\|_\infty$. 令 N 是所有 $N_{m,n}$ 的可数并, 那么仍然有 $\mu(N) = 0$, 并且对于每个 $x \in E \setminus N$ 有 $|f_n(x) - f_m(x)| \le \|f_n - f_m\|_\infty$. 此时, 对于每个 $x \in E \setminus N$, 序列 $(f_n(x))_{n\ge1}$ 是 $\mathbb{R}$ 中的 Cauchy 列, 因此有一个极限, 记为 $g(x)$. 若 $x \in N$ 则令 $g(x) = 0$. 此时 g 是可测函数, 并且

$$\sup_{x\in E\setminus N} |f_n(x) - g(x)| \le \sup_{m\in\{n+1,n+2,\dots\}} \|f_n - f_m\|_\infty,$$

右端在 $n \to \infty$ 时趋于 0, 因此 f_n 在 L^∞ 中收敛到 g. □

例 4.7. 如果 $E = \mathbb{N}$ 并且 μ 是计数测度,那么对于每个 $p \in [1,\infty)$,空间 $L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ 就是所有实数列 $a = (a_n)_{n\in\mathbb{N}}$ 的空间, 并且满足

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

空间 L^∞ 则是所有有界实数列构成的空间, 并且范数 $\|a\|_\infty = \sup_{n\in\mathbb{N}} |a_n|$. 注意到这个测度空间中没有非空的零测集, 所以此时 L^p 和 $\mathcal{L}^p$ 是重合的. 这个空间通常记为 $\ell^p = \ell^p(\mathbb{N})$, 是一个非常重要的 Banach 空间.

从上述定理的证明中我们还可以得到以下命题.

命题 4.8. 令 $p \in [1,\infty)$,$(f_n)_{n\in\mathbb{N}}$ 是 $L^p(E, \mathcal{A}, \mu)$ 中的收敛序列,有极限 f. 那么存在一个子列 $(f_{k_n})_{n\in\mathbb{N}}$ 几乎处处逐点收敛到 f.

注释 4.9. 这个命题对 $p = \infty$ 时也成立,但是这种情况下不需要提取子列了,因为 L^∞ 中的范数在除开零测集的意义上等价于一致收敛.

需要注意的是, 函数空间 L^p 中的收敛并不意味着逐点收敛. 下面的例子说明了这两种收敛之间的区别. 我们构造一个在 $L^p\big([0,1], \mathcal{B}([0,1]), \lambda\big)$ 中收敛到 0 但是逐点处处发散的序列 (f_n). 首先, 我们把区间 $[0,1]$ 不断二等分、四等分、八等分、……, 按照下面的顺序依次编号为 I_n:

$$[0,1], \left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right], \left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, 1\right], \dots.$$

令 $f_n = \mathbf{1}_{I_n}$. 对于任意 $p \in [1,\infty)$, 有

$$\|f_n - 0\|_p = \lambda(I_n)^{1/p} \to 0,$$

因此 $f_n \xrightarrow{L^p} 0$. 但是任取 $x \in [0,1]$, 总存在无穷多个 n 使得 $x \in I_n$, 于是有无穷多个 $f_n(x) = 1$, 因此序列 $(f_n(x))$ 发散. 所以为了构造 (f_n) 在 L^p 中的收敛函数, 直接取逐点的极限是不行的, 必须要从中挑选收敛行为一致的子列. 例如, 在这个例子中, 我们可以取 $g_n = f_{2^n} = \mathbf{1}_{[0,1/2^n]}$, 那么对于任意 $x \in [0,1]$ 有

$$\sum_{n=1}^{\infty} |g_{n+1}(x) - g_n(x)| < \infty.$$

此时再取极限

$$h(x) = g_1(x) + \sum_{n=1}^{\infty}(g_{n+1}(x) - g_n(x)) = \lim_{n\to\infty} g_n(x) = \begin{cases} 1 & x = 0, \\ 0 & x \in (0,1], \end{cases}$$

就能得到 L^p 空间中的极限函数.

不过, 根据命题 4.8, 如果已知 $L^p(E,\mathcal{A},\mu)$ 空间中的函数列 (f_n) 有极限 f, 并且几乎处处有逐点收敛 $f_n(x) \to g(x)$, 那么可以得到 $f = g$, μ a.e..

对于 $p \in [1,\infty)$, 我们可能会问什么时候反过来成立: 如果 $L^p(E,\mathcal{A},\mu)$ 中的函数列 (f_n) 几乎处处逐点收敛到 f, 那么是否在 L^p 中收敛? 也即函数空间 L^p 中的逐点收敛是否意味着 L^p 中的收敛? 一般来说这也是错误的, 但是根据控制收敛定理, 如果满足下面的条件,

1. $f_n \to f$, μ a.e.,
2. 存在一个非负可测函数 g 使得 $\int g^p \,\mathrm{d}\mu < \infty$ 并且对于每个 n 有 $|f_n| \le g$, μ a.e.,

那么函数 $f \in L^p$ 且 $f_n \xrightarrow{L^p} f$.

$p = 2$ 的时候尤其重要, 因为 L^2 空间有一个 Hilbert 空间的结构.

定理 4.10. 空间 $L^2(E,\mathcal{A},\mu)$ 配备内积

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

是一个实 Hilbert 空间.

Proof. Cauchy-Schwarz 不等式保证了 fg 是可积的, 因此内积是良定义的. 在这个内积中有 $\langle f, f\rangle = \|f\|_2^2$, 完备性在上面已经证明. □

经典的 Hilbert 空间理论可以应用到 L^2 空间中. 例如, 如果 $\Phi : L^2(E,\mathcal{A},\mu) \to \mathbb{R}$ 是连续线性映射, 那么存在唯一的 $g \in L^2(E,\mathcal{A},\mu)$, 使得对于每个 $f \in L^2(E,\mathcal{A},\mu)$, 都有 $\Phi(f) = \langle f, g\rangle$. 这个结果非常有用.

4.3 L^p 空间中的密度定理

令 (E,d) 是度量空间. 回顾一个函数 $f : E \to \mathbb{R}$ 被称为 Lipschitz 的, 如果存在常数 $K > 0$ 使得

$$\forall x, y \in E, \quad |f(x) - f(y)| \le K d(x,y).$$

$(E, \mathcal{B}(E))$ 上的测度 μ 被称为外正则的, 如果对于每个 $A \in \mathcal{B}(E)$, 都有

$$\mu(A) = \inf\left\{\mu(U) \,\middle|\, U \supseteq A,\ U \text{ 是开集}\right\}.$$

当 μ 是有限测度时, 外正则性总是成立的.

考虑一般的可测空间 $(E, \mathcal{A}, \mu)$, 我们在前面的章节引入了简单函数. 如果 μ 是 $(E, \mathcal{A})$ 上的测度, 对于可积的简单函数和任意 $p \in [1, \infty)$, 它们也都是 $L^p(\mu)$ 中的元素.

定理 4.11. 对于 $p \in [1, \infty)$.

1. 如果 $(E, \mathcal{A}, \mu)$ 是测度空间, 那么所有可积的简单函数集合在 $L^p(E, \mathcal{A}, \mu)$ 中稠密.
2. 如果 (E, d) 是度量空间, μ 是 $(E, \mathcal{B}(E))$ 上的外正则测度, 那么所有有界 Lipschitz 函数的集合是 $L^p(E, \mathcal{B}(E), \mu)$ 的稠密子集.
3. 如果 (E, d) 是可分的局部紧的度量空间, μ 是 E 上的 Radon 测度, 那么所有紧支的 Lipschitz 函数的集合在 $L^p(E, \mathcal{B}(E), \mu)$ 中稠密.

Proof. (1) 任取函数 $f \in L^p$, 有分解 $f = f^+ - f^-$, 所以不妨假设 f 是非负函数. 此时, 根据命题 2.6, f 是一列递增简单函数 (φ_n) 的逐点极限, 记作

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

那么 $\int |\varphi_n|^p \,\mathrm{d}\mu \leq \int |f|^p \,\mathrm{d}\mu < \infty$, 所以 $\varphi_n \in L^p$ (对于简单函数而言, 这等价于 $\varphi_n \in L^1$). 因为 $|f - \varphi_n|^p \leq f^p$, 根据控制收敛定理, 有

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

这就表明 $\varphi_n \xrightarrow{L^p} f$.

(2) 令 $f \in L^p(E, \mathcal{B}(E), \mu)$, 根据 (1), 存在一列 □

4.4 Radon-Nikodym 定理

定义 4.12. 令 μ 和 ν 是 $(E, \mathcal{A})$ 上的两个测度.

1. 如果
$$\forall A \in \mathcal{A}, \quad \mu(A) = 0 \Rightarrow \nu(A) = 0,$$
那么说 ν 相对于 μ 是**绝对连续**的, 记作 $\nu \ll \mu$.
2. 如果存在 $N \in \mathcal{A}$ 使得 $\mu(N) = 0$ 并且 $\nu(N^c) = 0$, 那么说 ν 相对于 μ 是**奇异**的, 记作 $\nu \perp \mu$.

例 4.13. 令 f 是 E 上的非负可测函数. 那么测度 $\nu = f \cdot \mu$ 相对于 μ 是绝对连续的. 实际上, 如果 $A \in \mathcal{A}$ 且 $\mu(A) = 0$, 那么

$$\nu(A) = \int_A f \,\mathrm{d}\mu = \int \mathbf{1}_A f \,\mathrm{d}\mu = 0.$$

回顾 μ 是 σ-有限的, 如果 E 是可数个 μ-有限测度集的并.

定理 4.14 (Radon-Nikodym). 假设 μ 是 σ-有限的，并且 ν 是另一个 σ-有限测度. 那么，存在唯一的 σ-有限测度对 (ν_a, ν_s)，使得

1. $\nu = \nu_a + \nu_s$.
2. $\nu_a \ll \mu$ 并且 $\nu_s \perp \mu$.

此外，存在非负可测函数 $g : E \to \mathbb{R}_+$ 使得 $\nu_a = g \cdot \mu$，也即

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

并且 g 在以下意义下唯一: 如果 $\tilde{g}$ 也满足上述性质，那么有 $g = \tilde{g}$, μ a.e..

注释 4.15. 这个定理的第一部分，是说 ν 可以分解为绝对连续部分和奇异部分的和，被称为 **Lebesgue 分解**. 如果 $\nu \ll \mu$，这个定理表明存在非负可测函数 g 使得 $\nu = g \cdot \mu$，这个函数 g 被称为 ν 相对于 μ 的 **Radon-Nikodym 导数**.

Proof. 我们证明 μ 和 ν 都是有限测度的情况, 最后解释如何推广到 σ-有限测度的情况.

Step 1. 假设 μ 和 ν 是有限的且 $\nu \leq \mu$. 也就是说, 我们假设对于每个 $A \in \mathcal{A}$, 都有 $\nu(A) \leq \mu(A)$, 特别的, 这当然表明 $\nu \ll \mu$. 并且, 对于任意非负可测函数 g, 有 $\int g \,\mathrm{d}\nu \leq \int g \,\mathrm{d}\mu$. 考虑线性映射 $\Phi : L^2(E, \mathcal{A}, \mu) \to \mathbb{R}$ 为

$$\Phi(f) = \int f \,\mathrm{d}\nu.$$

注意到这个积分是良定义的, 因为 $\int |f| \,\mathrm{d}\nu \leq \int |f| \,\mathrm{d}\mu$, 并且 μ 是有限测度表明 $L^2(\mu) \subseteq L^1(\mu)$. 此外 Φ 也不依赖于代表元 f 的选取, 这是因为

$$f = \tilde{f},\ \mu \text{ a.e.} \Rightarrow f = \tilde{f},\ \nu \text{ a.e.} \Rightarrow \int f \,\mathrm{d}\nu = \int \tilde{f} \,\mathrm{d}\nu.$$

那么 Cauchy-Schwarz 不等式表明

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

这表明 Φ 是连续线性映射, 因此根据 Riesz 表示定理, 存在唯一的 $g \in L^2(E, \mathcal{A}, \mu)$, 使得对于每个 $f \in L^2(E, \mathcal{A}, \mu)$, 都有

$$\Phi(f) = \langle f, g \rangle = \int fg \,\mathrm{d}\mu.$$

特别的, 取 $f = \mathbf{1}_A$, 就有

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

我们还可以注意到 $0 \le g \le 1$, μ a.e.. 这是因为, 任取 $\varepsilon > 0$, 有

$$\begin{aligned}\mu\big(\{x \in E \mid g(x) \ge 1+\varepsilon\}\big) &\ge \nu\big(\{x \in E \mid g(x) \ge 1+\varepsilon\}\big) \\ &= \int_{\{x|g(x)\ge 1+\varepsilon\}} g \,\mathrm{d}\mu \\ &\ge (1+\varepsilon)\mu\big(\{x \in E \mid g(x) \ge 1+\varepsilon\}\big),\end{aligned}$$

这表明 $\mu(\{x|g(x) \ge 1+\varepsilon\}) = 0$, 由于 ε 是任意的, 所以 $g \le 1$, μ a.e.. 类似的论述可以得到 $g \ge 0$, μ a.e.. 将 g 替换为 $(g \vee 0) \wedge 1$ 之后, 我们可以假设对于任意 $x \in E$ 有 $0 \le g(x) \le 1$. 于是, 我们得到了定理在 $\nu_a = g \cdot \mu$ 以及 $\nu_s = 0$ 的情况下的结论. g 的唯一性在第二步证明.

Step 2. 假设 μ 和 ν 是有限测度. 首先我们用 $\mu + \nu$ 替代 μ, 根据第一步的结论, 存在满足 $0 \le h \le 1$ 的可测函数 h, 使得对于任意 $f \in L^2(\mu+\nu)$, 都有

$$\int f \,\mathrm{d}\nu = \int f h \,\mathrm{d}(\mu+\nu).$$

特别的, 对于任意有界可测函数 f, 都有

$$\int f \,\mathrm{d}\nu = \int f h \,\mathrm{d}\mu + \int f h \,\mathrm{d}\nu.$$

这表明

$$\int f(1-h) \,\mathrm{d}\nu = \int f h \,\mathrm{d}\mu. \tag{4.1}$$

使用单调收敛定理, 我们可以将上式推广到任意非负可测函数 f.

令 $N = \{x \in E \mid h(x) = 1\}$. 那么, 在 (4.1) 中取 $f = \mathbf{1}_N$, 我们得到 $\mu(N) = 0$. 考虑测度 $\nu_s = \mathbf{1}_N \cdot \nu$, 那么对于任意 $A \in \mathcal{A}$, 都有

$$\nu_s(A) = \nu(A \cap N),$$

于是 $\nu_s(N^c) = 0$, 因此 ν_s 相对于 μ 是奇异的. 另一方面, 在 (4.1) 中取 $f = \mathbf{1}_{N^c}(1-h)^{-1} f$, 我们得到, 对于每个非负可测函数 f, 都有

$$\int_{N^c} f \,\mathrm{d}\nu = \int_{N^c} f \frac{h}{1-h} \,\mathrm{d}\mu = \int f g \,\mathrm{d}\mu,$$

其中 $g = \mathbf{1}_{N^c} \frac{h}{1-h}$. 令

$$\nu_a = \mathbf{1}_{N^c} \cdot \nu = g \cdot \mu.$$

此时定理的 (1) 和 (2) 就得到了满足. 注意到 $\int g \,\mathrm{d}\mu = \nu_a(E) < \infty$.

下面验证 (ν_a, ν_s) 的唯一性. 假设 $(\tilde{\nu}_a, \tilde{\nu}_s)$ 是另一个满足 (1) 和 (2) 的测度对. 那么对于每个 $A \in \mathcal{A}$, 都有

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

由于 $\nu_s \perp \mu$ 并且 $\tilde{\nu}_s \perp \mu$, 所以可以找到两个可测集 N 和 $\tilde{N}$ 使得 $\nu_s(N^c) = 0$ 并且 $\tilde{\nu}_s(\tilde{N}^c) = 0$. 那么, 对于每个 $A \in \mathcal{A}$, 就有

$$\begin{aligned}\nu_s(A) - \tilde{\nu}_s(A) &= \nu_s(A \cap (N \cup \tilde{N})) - \tilde{\nu}_s(A \cap (N \cup \tilde{N})) \\ &= \tilde{\nu}_a(A \cap (N \cup \tilde{N})) - \nu_a(A \cap (N \cup \tilde{N})) = 0.\end{aligned}$$

第一个等号是因为 $\nu_s(N^c \cap \tilde{N}^c) = 0$, 所以 $\nu_s(N \cup \tilde{N}) = \nu_s(E)$. 第三个等号是因为 $\mu(N \cup \tilde{N}) = 0$, 同时 $\nu_a \ll \mu$ 以及 $\tilde{\nu}_a \ll \mu$. 所以我们证明了 $\nu_s = \tilde{\nu}_s$, 那么自然也有 $\nu_a = \tilde{\nu}_a$.

最后, 再说明 g 的唯一性. 假设存在另一个函数 $\tilde{g}$ 使得 $\nu_a = \tilde{g} \cdot \mu$. 令 $\{\tilde{g} \geq g\}$ 表示集合 $\{x \in E \mid \tilde{g}(x) \geq g(x)\}$. 那么

$$\int_{\{\tilde{g} \geq g\}} \tilde{g} \, \mathrm{d}\mu = \nu_a(\{\tilde{g} \geq g\}) = \int_{\{\tilde{g} \geq g\}} g \, \mathrm{d}\mu,$$

这表明

$$\int_{\{\tilde{g} \geq g\}} (\tilde{g} - g) \, \mathrm{d}\mu = 0,$$

即 $\mathbf{1}_{\{\tilde{g} \geq g\}}(\tilde{g} - g) = 0$, μ a.e., 即 $\tilde{g} \leq g$, μ a.e.. 交换 g 和 $\tilde{g}$ 的角色, 我们也可以得到 $\tilde{g} \geq g$, μ a.e.. 综上所述, $g = \tilde{g}$, μ a.e..

Step 3. 现在考虑一般情况. 如果 μ 和 ν 都是 σ-有限的, 那么可以找到 E 的一列不相交的可测集 $(E_n)_{n \in \mathbb{N}}$ 使得每个 $\mu(E_n) < \infty$ 和 $\nu(E_n) < \infty$, 并且 $E = \bigcup_n E_n$. 令 μ_n 是 μ 在 E_n 上的限制, ν_n 是 ν 在 E_n 上的限制. 根据第二步的结论, 对于每个 $n \in \mathbb{N}$, 有分解

$$\nu_n = \nu_a^n + \nu_s^n,$$

其中 $\nu_s^n \perp \mu_n$ 并且 $\nu_a^n = g_n \cdot \mu_n$, 我们还可以假设 g_n 在 E_n^c 上为零. 令

$$\nu_a = \sum_{n \in \mathbb{N}} \nu_a^n, \quad \nu_s = \sum_{n \in \mathbb{N}} \nu_s^n, \quad g = \sum_{n \in \mathbb{N}} g_n.$$

对于每个 $x \in E$, 由于 E_n 互不相交, 所以只有一个 n 使得 $g_n(x) > 0$. (ν_a, ν_s, g) 的唯一性与有限测度的情况类似. □

例 4.16.

1. 取 $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 假设 $\mu = f \cdot \lambda$, 其中 λ 是 Lebesgue 测度. 假设 f 在 $(0, \infty)$ 上是正的, 在 $(-\infty, 0]$ 上为零. 令 $\nu = g \cdot \lambda$ 是另一个相对于 λ 绝对连续的测度, 此时 ν 相对于 μ 的 Lebesgue 分解为

$$\nu = h \cdot \mu + \theta,$$

其中 $h(x) = \mathbf{1}_{(0,\infty)} \cdot g(x)/f(x)$ 并且 $\theta(\mathrm{d}x) = \mathbf{1}_{(-\infty,0]}(x)\nu(\mathrm{d}x)$.

2. 取 $E = [0, 1)$ 和 $\mathcal{A} = \mathcal{B}([0, 1))$, 对于每个 $n \in \mathbb{N}$, 令 σ-域

$$\mathcal{F}_n = \sigma\left(\left\{\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right) \,\middle|\, i \in \{1, 2, \ldots, 2^n\}\right\}\right),$$

不难发现任意 $A \in \mathcal{F}_n$ 都是区间 $\big[(i-1)2^{-n}, i2^{-n}\big)$ 的有限并. 记 λ 是 $[0,1)$ 上的 Lebesgue 测度, ν 是 $\mathcal{B}([0,1))$ 上的有限测度. 把 λ 和 ν 都限制到 $\mathcal{F}_n$ 上, 可以把 λ 和 ν 都视为 $([0,1),\mathcal{F}_n)$ 上的测度. 显然, 此时 ν 相对于 λ 是绝对连续的, 因为只有 $\emptyset$ 使得 $\lambda(\emptyset)=0$. 此外, 还可以验证 ν 相对于 λ 的 Radon-Nikodym 导数为

$$f_n(x) = \sum_{i=1}^{2^n} \frac{\nu\big(\big[(i-1)2^{-n}, i2^{-n}\big)\big)}{2^{-n}} \mathbf{1}_{[(i-1)2^{-n}, i2^{-n})}(x).$$

后面,我们将使用鞅论证明存在一个 Borel 可测函数 f 使得 $f_n(x) \to f(x)$, λ a.e., 并且将 ν 和 λ 视为 $([0,1),\mathcal{B}([0,1)))$ 上的测度时, ν 相对于 λ 的 Lebesgue 分解的绝对连续部分正是 $f \cdot \lambda$.

5

积测度

5.1 积 σ-域

令 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$ 是两个可测空间. 回顾第一章, 我们定义 $E\times F$ 上的乘积 σ-域

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

不难验证 $\mathcal{A}\otimes\mathcal{B}$ 是使得两个投影映射 $\pi_1:E\times F\to E$ 和 $\pi_2:E\times F\to F$ 都可测的最小的 σ-域.

如果 $C\subseteq E\times F$, $x\in E$, 记

$$C_x=\{y\in F\,|\,(x,y)\in C\}\subseteq F,$$

如果 $y\in F$, 记

$$C^y=\{x\in E\,|\,(x,y)\in C\}\subseteq E.$$

如果 f 是 $E\times F$ 上的函数, $x\in E$, 我们记 f_x 表示 F 上的函数 $f_x(y)=f(x,y)$. 类似地, 如果 $y\in F$, 我们记 $f^y(x)=f(x,y)$ 表示 E 上的函数.

命题 5.1.

1. 令 $C\in\mathcal{A}\otimes\mathcal{B}$，那么对于任意 $x\in E$，$C_x\in\mathcal{B}$，对于任意 $y\in F$，$C^y\in\mathcal{A}$.
2. 令 $(G,\mathcal{G})$ 是可测空间，$f:E\times F\to G$ 是可测函数，那么对于任意 $x\in E$，$f_x:F\to G$ 是可测的，对于任意 $y\in F$，$f^y:E\to G$ 是可测的.

Proof. (1) 对于 $x\in E$, 令

$$\mathcal{C}=\{C\in\mathcal{A}\otimes\mathcal{B}\,|\,C_x\in\mathcal{B}\},$$

那么不难验证 $\mathcal{C}$ 是一个 σ-域且包含所有的可测矩形, 于是 $\mathcal{C}=\mathcal{A}\otimes\mathcal{B}$, 即表明对于任意 $C\in\mathcal{A}\otimes\mathcal{B}$ 都有 $C_x\in\mathcal{B}$. $C^y\in\mathcal{A}$ 同理.

(2) 对于 $x\in E$, 任取 $D\in\mathcal{G}$, 有

$$f_x^{-1}(D)=\left(f^{-1}(D)\right)_x\in\mathcal{B}.$$ □

5.2 积测度

定理 5.2. 令 μ 和 ν 分别是 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$ 上的 σ-有限测度，那么

1. 存在唯一的 $(E\times F,\mathcal{A}\otimes\mathcal{B})$ 上的测度 m，使得对于每个 $A\in\mathcal{A}, B\in\mathcal{B}$，都有

$$m(A\times B)=\mu(A)\nu(B),$$

约定 $0\times\infty=0$. 测度 m 也是 σ-有限的，记为 $m=\mu\otimes\nu$.

2. 对于每个 $C\in\mathcal{A}\otimes\mathcal{B}$，函数 $x\mapsto\nu(C_x)$ 是 $\mathcal{A}$-可测的，$y\mapsto\mu(C^y)$ 是 $\mathcal{B}$-可测的，并且我们有

$$\mu\otimes\nu(C)=\int_E\nu(C_x)\mu(\mathrm{d}x)=\int_F\mu(C^y)\nu(\mathrm{d}y).$$

Proof. 我们首先说明这样的测度 m 一定是唯一的. 我们使用 推论 1.24 说明唯一性. 若测度 m' 也满足性质 (1). 首先所有可测矩形对有限交封闭且生成 $\mathcal{A}\otimes\mathcal{B}$, 并且 m 和 m' 在可测矩形上取值相同. μ 是 σ-有限的表明存在递增的可测子集 $(A_n)_{n\in\mathbb{N}}$ 使得 $\bigcup_{n\in\mathbb{N}}A_n=E$ 并且 $\mu(A_n)$ 有限.同理存在递增的可测子集 $(B_n)_{n\in\mathbb{N}}$ 使得 $\bigcup_{n\in\mathbb{N}}B_n=F$ 并且 $\nu(B_n)$ 有限. 令 $G_n=A_n\times B_n$, 那么 $G_n\subseteq G_{n+1}$ 并且 $E\times F=\bigcup_{n\in\mathbb{N}}G_n$, 此时

$$m'(G_n)=\mu(A_n)\nu(B_n)=m(G_n)<\infty,$$

所以 推论 1.24 表明 $m=m'$.

然后我们说明存在性. 对于 $C\in\mathcal{A}\otimes\mathcal{B}$, 定义

$$m(C)=\int_E\nu(C_x)\mu(\mathrm{d}x). \tag{5.1}$$

对于任意 $x\in E$, 有 $C_x\in\mathcal{B}$, 所以 $\nu(C_x)$ 是良好定义的. 下面我们证明 $x\mapsto\nu(C_x)$ 是可测函数.

(1) 首先假设 ν 是有限测度. 令 $\mathcal{G}=\{C\in\mathcal{A}\otimes\mathcal{B}\mid x\mapsto\nu(C_x)\text{ 可测}\}$. 那么对于可测矩形 $A\times B\in\mathcal{A}\otimes\mathcal{B}$, 有 $\nu((A\times B)_x)=\mathbf{1}_A(x)\nu(B)$, 此时 $x\mapsto\mathbf{1}_A(x)\nu(B)$ 当然是可测函数. 故 $\mathcal{G}$ 包含所有的可测矩形. 其次, 我们证明 $\mathcal{G}$ 是一个单调类. 如果 $C,C'\in\mathcal{G}$ 且 $C\subseteq C'$, 那么利用 ν 的有限性, 就有

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

所以 $x\mapsto\nu\big((C'\setminus C)_x\big)$ 是可测函数, 即 $C'\setminus C\in\mathcal{G}$. 如果 $(C_n)_{n\in\mathbb{N}}$ 是 $\mathcal{G}$ 中的一个递增序列, 那么

$$\nu\left(\left(\bigcup_{n\in\mathbb{N}}C_n\right)_x\right)=\nu\left(\bigcup_{n\in\mathbb{N}}(C_n)_x\right)=\lim_{n\to\infty}\uparrow\nu((C_n)_x),$$

而 $x\mapsto\lim\uparrow\nu((C_n)_x)$ 是可测函数, 所以 $\bigcup_{n\in\mathbb{N}}C_n\in\mathcal{G}$. 这就表明 $\mathcal{G}$ 是单调类. 由于 $\mathcal{G}$ 包含可测矩形, 可测矩形对有限交封闭, 根据单调类定理, 所以 $\mathcal{G}$ 是 σ-域, 所以 $\mathcal{G}=\mathcal{A}\otimes\mathcal{B}$, 这表明对于任意 $C\in\mathcal{A}\otimes\mathcal{B}$, $x\mapsto\nu(C_x)$ 都是可测函数.

(2) 然后假设 ν 是 σ-有限测度. 此时存在 $\mathcal{B}$ 的一列递增子集 $(B_n)_{n\in\mathbb{N}}$ 使得 $F = \bigcup_{n\in\mathbb{N}} B_n$ 以及 $\nu(B_n) < \infty$. 令 ν_n 表示测度 ν 在 B_n 上的限制, 那么根据上面的叙述, 任取 $C \in \mathcal{A}\otimes\mathcal{B}$, 函数 $x \mapsto \nu_n(C_x)$ 是可测函数. 注意到

$$\nu(C_x) = \nu\left(\bigcup_{n\in\mathbb{N}} (C_x \cap B_n)\right) = \lim_{n\to\infty} \uparrow \nu_n(C_x),$$

所以 $x \mapsto \lim \uparrow \nu(C_x)$ 是可测函数.

于是我们证明了 $x \mapsto \nu(C_x)$ 是可测函数, 这表明定义 (5.1) 式是有意义的. 下面我们验证 m 满足测度的条件. 显然 $m(\emptyset) = 0$. 任取 $(C_n)_{n\in\mathbb{N}}$ 是 $\mathcal{A}\otimes\mathcal{B}$ 中的一列不相交的子集, 那么

$$\begin{aligned} m\left(\bigcup_{n\in\mathbb{N}} C_n\right) &= \int_E \nu\left(\bigcup_{n\in\mathbb{N}} (C_n)_x\right)\mu(\mathrm{d}x) = \int_E \sum_{n\in\mathbb{N}} \nu((C_n)_x)\mu(\mathrm{d}x) \\ &= \sum_{n\in\mathbb{N}} \int_E \nu((C_n)_x)\mu(\mathrm{d}x) = \sum_{n\in\mathbb{N}} m(C_n), \end{aligned}$$

其中第三个等号利用了单调收敛定理. 这就表明 m 确实是一个测度.

对于可测矩形 $A \times B \in \mathcal{A}\otimes\mathcal{B}$, 有

$$m(A\times B) = \int_E \nu((A\times B)_x)\mu(\mathrm{d}x) = \int_E \nu(B)\mathbf{1}_A(x)\mu(\mathrm{d}x) = \mu(A)\nu(B),$$

所以这样的 m 是唯一的. 对于 $C \in \mathcal{A}\otimes\mathcal{B}$, 定义

$$m'(C) = \int_F \mu(C^y)\nu(\mathrm{d}y),$$

重复上面的过程, 可以证明 m' 满足和 m 相同的性质, 所以 $m = m'$. □

例 5.3. 如果 $(E,\mathcal{A}) = (F,\mathcal{B}) = (\mathbb{R},\mathcal{B}(\mathbb{R}))$, 并且 $\mu = \nu = \lambda$, 可以验证 $\lambda\otimes\lambda$ 就是 $\mathbb{R}^2$ 上的 Lebesgue 测度 (只需要在矩形 $[a,b]\times[c,d]$ 上验证, 然后利用 推论 1.24).

5.3 Fubini 定理

考虑可测空间 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$.

定理 5.4 (Fubini-Tonelli). 令 μ 和 ν 分别是 $(E,\mathcal{A})$ 和 $(F,\mathcal{B})$ 上的两个 σ-有限的测度. 令 $f : E\times F \to [0,\infty]$ 是可测函数.

1. 函数

$$E \ni x \mapsto \int_F f(x,y)\nu(\mathrm{d}y), \quad F \ni y \mapsto \int_E f(x,y)\mu(\mathrm{d}x)$$

是值在 $[0,\infty]$ 中的可测函数.

2. 我们有

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

Proof. (1) 设 $f = \lim \uparrow f_n$, $(f_n)_{n\in\mathbb{N}}$ 是一列递增的非负简单函数, 那么根据单调收敛定理, 有

$$\int_F f(x,y)\nu(\mathrm{d}y) = \lim_{n\to\infty} \uparrow \int_F f_n(x,y)\nu(\mathrm{d}y),$$

所以我们只需要说明对于任意非负简单函数 h, $x \mapsto \int_F h(x,y)\nu(\mathrm{d}y)$ 可测即可. 对于示性函数 $\mathbf{1}_C$, 有 $\int_F \mathbf{1}_C(x,y)\nu(\mathrm{d}y) = \nu(C_x)$, 定理 5.2 表明 $x \mapsto \nu(C_x)$ 是可测的, 再根据线性性, 这就说明了 $x \mapsto \int_F h(x,y)\nu(\mathrm{d}y)$ 可测. 对于 $x \mapsto \int_E f(x,y)\mu(\mathrm{d}x)$ 同理.

(2) 设 $f = \lim \uparrow f_n$, $(f_n)_{n\in\mathbb{N}}$ 是一列递增的非负简单函数, 那么根据单调收敛定理, 有

$$\int_{E\times F} f \,\mathrm{d}\mu\otimes\nu = \lim_{n\to\infty} \uparrow \int_{E\times F} f_n(x,y)\nu(\mathrm{d}y),$$

所以只需要证明结论对于非负简单函数成立即可. 根据线性性, 只需要证明结论对示性函数成立即可. 任取示性函数 $\mathbf{1}_C$, 根据 定理 5.2, 有

$$\int_{E\times F} \mathbf{1}_C \,\mathrm{d}\mu\otimes\nu = \mu\otimes\nu(C) = \int_E \nu(C_x)\mu(\mathrm{d}x) = \int_E \left(\int_F \mathbf{1}_C(x,y)\nu(\mathrm{d}y)\right)\mu(\mathrm{d}x),$$

另一个等式同理. □

定理 5.4 也可以推广到任意符号的版本.

定理 5.5 (Fubini-Lebesgue). 令 $f \in \mathcal{L}^1(E\times F, \mathcal{A}\otimes\mathcal{B}, \mu\otimes\nu)$，那么

1. $\mu(\mathrm{d}x)$ a.e.，函数 $y \mapsto f(x,y)$ 属于 $\mathcal{L}^1(F,\mathcal{B},\nu)$. $\nu(\mathrm{d}y)$ a.e.，函数 $x \mapsto f(x,y)$ 属于 $\mathcal{L}^1(E,\mathcal{A},\mu)$.
2. 函数 $x \mapsto \int_F f(x,y)\nu(\mathrm{d}y)$ 属于 $\mathcal{L}^1(E,\mathcal{A},\mu)$. 函数 $y \mapsto \int_E f(x,y)\mu(\mathrm{d}x)$ 属于 $\mathcal{L}^1(F,\mathcal{B},\nu)$.
3. 我们有

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

注释 5.6. (1) 函数 $x \mapsto \int_F f(x,y)\nu(\mathrm{d}y)$ 仅仅在某个 μ-零测集之外有定义，在这个零测集上，我们通常让这个函数为零. (2) 对于 $f \in \mathcal{L}^1_{\mathbb{C}}(E\times F, \mathcal{A}\otimes\mathcal{B}, \mu\otimes\nu)$ 有同样的结论.

Proof. (1) 将 定理 5.4 应用于 $|f|$, 有

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

这表明 $\int_F |f(x,y)|\nu(\mathrm{d}y) < \infty$, μ a.e.. 令

$$N = \left\{x \in E \,\middle|\, \int_F |f(x,y)|\nu(\mathrm{d}y) = \infty\right\}$$

是 $\mathcal{A}$-可测集, 那么 $\mu(N) = 0$. 对于任意 $x \in E\setminus N$, 函数 $y \mapsto f(x,y)$ 属于 $\mathcal{L}^1(F,\mathcal{B},\nu)$. 对于 $x \mapsto f(x,y)$ 有同样的论述.

(2) 如果 $x \in N^c$, 那么

$$\int_E \left| \int_F f(x,y)\nu(\mathrm{d}y) \right| \mu(\mathrm{d}x) \le \int_E \left(\int_F |f(x,y)|\nu(\mathrm{d}y) \right) \mu(\mathrm{d}x)$$
$$= \int_{E\times F} |f| \,\mathrm{d}(\mu \otimes \nu) < \infty.$$

如果 $x \in N$, 注意我们定义 $\int_F f(x,y)\nu(\mathrm{d}y) = 0$. 此时

$$x \mapsto \int_F f(x,y)\nu(\mathrm{d}y) = \mathbf{1}_{N^c}(x) \int_F f^+(x,y)\nu(\mathrm{d}y) - \mathbf{1}_{N^c}(x) \int_F f^-(x,y)\nu(\mathrm{d}y),$$

这个函数是 $\mathcal{A}$-可测的，并且同样有上面的不等式，所以 $x \mapsto \int_F f(x,y)\nu(\mathrm{d}y)$ 在 $\mathcal{L}^1(E,\mathcal{A},\mu)$ 中. 对于 $y \mapsto \int_E f(x,y)\mu(\mathrm{d}x)$ 有同样的论述.

(3) 将定理 5.4 应用于 f^+ 和 f^- 即可. $\square$

Part II

概率论

6

概率论基础

6.1 一般定义

6.1.1 概率空间

令 $(\Omega, \mathcal{A})$ 是可测空间, $\mathbb{P}$ 是 $(\Omega, \mathcal{A})$ 上的概率测度, 我们说 $(\Omega, \mathcal{A}, \mathbb{P})$ 是**概率空间**.

因此, 概率空间是测度空间的一个特例. 然而, 概率论的观点与测度论有很大不同. 在概率论中, 我们的目标是一个“随机实验”的数学模型:

- Ω 表示实验的所有可能的结果的集合.
- $\mathcal{A}$ 是所有“事件”的集合. 这里的事件指的是 Ω 的一个子集, 其概率可以被计算 (也就是可测集). 我们应当把事件 A 视为满足某一属性的所有 $\omega \in \Omega$ 构成的子集.
- 对于每个 $A \in \mathcal{A}$, $\mathbb{P}(A)$ 表示事件 A 发生的概率.

当然, 一个自然的疑问是, 为什么需要考虑事件域 $\mathcal{A}$? 换句话说, 为什么不能对 Ω 的任意子集都计算一个概率? 原因在于, 一般不可能在 Ω 的幂集 $\mathcal{P}(\Omega)$ 上定义我们感兴趣的概率测度 (除开 Ω 是可数集这一简单情况). 例如, 取 $\Omega = [0, 1]$, 配备 Borel σ-域和 Lebesgue 测度, 但是, 可以证明不可能将 Lebesgue 测度扩展到 $[0, 1]$ 的任意子集上使得其仍然满足测度的定义.

例 6.1. 一些常见的概率模型.

1. 考虑扔两次骰子这一实验, 那么
$$\Omega = \{1, 2, \ldots, 6\}^2, \quad \mathcal{A} = \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \frac{\operatorname{card}(A)}{36}.$$
这里概率 $\mathbb{P}$ 的选取意味着让所有结果都有相同的概率. 更一般地, 如果 Ω 是有限集, $\mathcal{A} = \mathcal{P}(\Omega)$, 概率测度 $\mathbb{P}(\{\omega\}) = 1/\operatorname{card}(\Omega)$ 被称为 Ω 上的**均匀概率测度**.
2. 现在我们考虑实验: 扔骰子, 直到出现 6 为止. 由于得到 6 所需的投掷次数是无界的 (即使你扔了 1000 次骰子, 仍有可能没有得到 6), 所以 Ω 的正确选择是想象我们扔了无限次骰子:
$$\Omega = \{1, 2, \ldots, 6\}^{\mathbb{N}}.$$
Ω 上的 σ-域 $\mathcal{A}$ 被定义为包含形如
$$\{\omega \in \Omega \mid \omega_1 = i_1, \ldots, \omega_n = i_n\}$$

的最小的 σ-域. 这个形式的集合代表了“只观测有限次”的事件, 例如假设 $n=1, i_1=6$, 那么这个形式的集合包含了所有以 6 开头的无限序列. 虽然 ω 是无限长的, 但是这个集合的性质只由前几项决定. 这样的 σ-域被称为**圆柱 σ-域**. 最后, 令 $\mathbb{P}$ 是有限测度, 满足对于每个 n 和 $i_1,\dots,i_n$ 有

$$\mathbb{P}\big(\{\omega\in\Omega \mid \omega_1=i_1,\dots,\omega_n=i_n\}\big)=\left(\frac{1}{6}\right)^n.$$

与测度论类似, 零测集也会出现在概率论的很多叙述中, 如果某个命题对于某个概率为 1 的事件中的每个 $\omega\in\Omega$ 都成立, 那么我们说这个命题**几乎肯定**成立, 用缩写 a.s. 表示.

6.1.2 随机变量

在本章的剩余部分, 我们都考虑一个概率空间 $(\Omega,\mathcal{A},\mathbb{P})$, 并且所有随机变量都将在这个概率空间上定义.

定义 6.2. 令 $(E,\mathcal{E})$ 是可测空间，值在 E 中的**随机变量**指的是一个可测映射 $X:\Omega\to E$.

例 6.3. 回顾 (6.1) 中的模型.

1. $X((i,j))=i+j$ 定义了值在 $\{2,3,\dots,12\}$ 中的随机变量.
2. $X(\omega)=\inf\{j\mid \omega_j=6\}$, 约定 $\inf\emptyset=\infty$, 定义了值在 $\bar{\mathbb{N}}=\mathbb{N}\cup\{\infty\}$ 中的随机变量. 为了验证 X 的可测性, 只需要注意到

$$X^{-1}(\{k\})=\{\omega\in\Omega\mid \omega_1\neq 6,\dots,\omega_{k-1}\neq 6,\omega_k=6\}.$$

定义 6.4. 令 X 是值在 $(E,\mathcal{E})$ 中的随机变量，定义随机变量 X 的 **分布律** $\mathbb{P}_X$ 是概率测度 $\mathbb{P}$ 在 X 下的推前. 也就是说，$\mathbb{P}_X$ 是 $(E,\mathcal{E})$ 上的概率测度，满足

$$\mathbb{P}_X(B)=\mathbb{P}(X^{-1}(B)),\quad \forall B\in\mathcal{E}.$$

两个值在 $(E,\mathcal{E})$ 中的随机变量 Y,Y' 如果有相同的分布 $\mathbb{P}_Y=\mathbb{P}_{Y'}$，那么我们说 Y 和 Y' 是**同分布**的.

在概率论中, 我们通常将 $\mathbb{P}_X(B)$ 写为 $\mathbb{P}(X\in B)$ 而不是 $\mathbb{P}(X^{-1}(B))$. 这里 $X\in B$ 是集合 $\{\omega\in\Omega\mid X(\omega)\in B\}$ 的简写, 这是一个一般性的简写规则, 在概率论中参数 ω 通常被隐藏.

离散型随机变量 当 E 是有限或者可数 ($\mathcal{E}=\mathcal{P}(E)$) 的时候, X 的分布是点测度, 这是因为

$$\mathbb{P}_X(B)=\mathbb{P}(X\in B)=\mathbb{P}\Big(\bigcup_{x\in B}\{X=x\}\Big)=\sum_{x\in B}\mathbb{P}(X=x)=\sum_{x\in E}p_x\delta_x(B),$$

其中 $p_x = \mathbb{P}(X = x)$. 这就表明

$$\mathbb{P}_X = \sum_{x \in E} p_x \delta_x$$

是 E 上的点测度.

例 6.5. 我们考虑 (6.1) 中的第二个例子, 随机变量为 $X(\omega) = \inf\{j \mid \omega_j = 6\}$. 那么

$$\begin{aligned}\mathbb{P}(X = k) &= \mathbb{P}\left(\bigcup_{1 \le i_1, \dots, i_k \le 5} \{\omega \mid \omega_1 = i_1, \dots, \omega_{k-1} = i_{k-1}, \omega_k = 6\}\right) \\ &= 5^{k-1} \left(\frac{1}{6}\right)^k = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}.\end{aligned}$$

注意到

$$\sum_{k=1}^{\infty} \mathbb{P}(X = k) = \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 1$$

并且 $\{X = \infty\} \cup \bigcup_{k=1}^{\infty} \{X = k\} = \Omega$, 所以

$$\mathbb{P}(X = \infty) = 1 - \sum_{k=1}^{\infty} \mathbb{P}(X = k) = 0,$$

但是 $\{X = \infty\} \neq \emptyset$.

具有密度的随机变量 $\mathbb{R}^d$ 上的密度函数是一个非负的 Borel 函数 $p : \mathbb{R}^d \to \mathbb{R}_+$, 其满足

$$\int_{\mathbb{R}^d} p(x)\,\mathrm{d}x = 1.$$

对于一个值在 $\mathbb{R}^d$ 中的随机变量 X, 如果存在密度 p 使得

$$\mathbb{P}_X(B) = \int_B p(x)\,\mathrm{d}x$$

对于任意 Borel 子集 B 都成立, 那么我们说 X 有密度函数 p. 换句话说, p 是 $\mathbb{P}_X$ 相对于 Lebesgue 测度 λ 的密度 (推论 2.8), 也记为 $\mathbb{P}_X(\mathrm{d}x) = p(x)\lambda(\mathrm{d}x) = p(x)\,\mathrm{d}x$. 根据 Radon-Nikodym 定理, 随机变量 X 有密度函数的充分必要条件是 $\mathbb{P}_X$ 相对于 Lebesgue 测度 λ 是绝对连续的. 此时 p 在相差一个 Lebesgue 零测集的意义下是唯一确定的.

注意到密度 p 实际上是在相差一个 Lebesgue 零测集的意义下由 $\mathbb{P}_X$ 确定的. 在我们遇到的大多数例子中, p 在 $\mathbb{R}^d$ 上连续, 在这种情况下, p 由 $\mathbb{P}_X$ 唯一确定.

在 $d = 1$ 的时候, 我们有

$$\mathbb{P}(\alpha \le X \le \beta) = \int_\alpha^\beta p(x)\,\mathrm{d}x.$$

6.1.3 数学期望

定义 6.6. 令 X 是定义在 $(\Omega, \mathcal{A}, \mathbb{P})$ 上的实随机变量，我们定义

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega) = \int X \,\mathrm{d}\mathbb{P},$$

只要上述积分有意义，我们就说 $\mathbb{E}[X]$ 是 X 的**期望**.

根据前面的内容, 上述积分有意义的条件为下列二者之一:

- $X \geq 0$, 此时 $\mathbb{E}[X] \in [0, \infty]$.
- X 符号任意, 但是 $\mathbb{E}[|X|] = \int |X| \,\mathrm{d}\mathbb{P} < \infty$.

上面的定义可以拓展到多元随机变量 $X = (X_1, \dots, X_d) \in \mathbb{R}^d$, 此时我们定义 $\mathbb{E}[X] = \big(\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]\big)$. 类似的, 如果 M 是随机矩阵 (值在实矩阵空间中的随机变量), 我们可以定义矩阵 $\mathbb{E}[M]$ 为对 M 的每个分量求期望构成的矩阵.

注意到若 $X = \mathbf{1}_B$, 那么

$$\mathbb{E}[X] = \int \mathbf{1}_B \,\mathrm{d}\mathbb{P} = \mathbb{P}(B).$$

对于一些特殊的随机变量, 下面的命题被频繁地使用.

命题 6.7. 令 X 是值在 $[0, \infty]$ 中的随机变量，那么

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)\,\mathrm{d}x.$$

令 Y 是值在 $\mathbb{Z}_+$ 中的随机变量，那么

$$\mathbb{E}[Y] = \sum_{k=0}^\infty k\,\mathbb{P}(X = k) = \sum_{k=1}^\infty \mathbb{P}(Y \geq k).$$

Proof. 根据 Fubini 定理, 我们有

$$\mathbb{E}[X] = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{x \leq X\}}\,\mathrm{d}x\right] = \int_0^\infty \mathbb{E}[\mathbf{1}_{\{x \leq X\}}]\,\mathrm{d}x = \int_0^\infty \mathbb{P}(X \geq x)\,\mathrm{d}x.$$

对于随机变量 Y, 我们有

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{k=0}^\infty k\mathbf{1}_{\{Y=k\}}\right] = \int \left(\sum_{k=0}^\infty k\mathbf{1}_{\{Y=k\}}\right)\mathrm{d}\mathbb{P} = \sum_{k=0}^\infty k\,\mathbb{P}(Y = k).$$

对于第二个等式, 只需注意到

$$Y = \sum_{k=1}^\infty \mathbf{1}_{\{Y \geq k\}}.$$

□

下面的命题是命题 2.12 的特例, 由于其结果十分重要, 所以我们再次叙述一遍.

命题 6.8. 令 X 是值在 $(E, \mathcal{E})$ 中的随机变量，对于任意可测函数 $f : E \to [0, \infty]$，我们有

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\mathbb{P}(\mathrm{d}\omega) = \int_E f(x)\mathbb{P}_X(\mathrm{d}x).$$

如果可测函数 $f : E \to \mathbb{R}$, 上面的命题在两端有意义的情况下也是成立的, 即 $\mathbb{E}[|f(X)|] < \infty$ 的时候. 特别地, 如果 X 是实值随机变量且使得 $\mathbb{E}[|X|] < \infty$, 那么有

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} x\,\mathbb{P}_X(\mathrm{d}x).$$

如果 X 有密度 p, 也就是说 $\mathbb{P}_X = p \cdot \lambda$, 那么还有

$$\mathbb{E}[X] = \int_\mathbb{R} x\,\mathbb{P}_X(\mathrm{d}x) = \int_\mathbb{R} x p(x)\,\mathrm{d}x.$$

分布律的刻画 根据推论 1.24, $\mathbb{R}^d$ 上的概率测度 μ 完全由其在开矩形上的值确定, 也即由所有 $\mu((a_1,b_1)\times\cdots\times(a_d,b_d))$ 确定 (当然, 也可以把区间换成闭区间, 甚至要求所有的 a_i, b_i 都是有理数). 因为开矩形上的示性函数是一列紧支连续函数的递增极限, 所以概率测度 μ 也可以由所有 $\int \varphi(x)\mu(\mathrm{d}x)$ 的值确定, 其中 $\varphi \in C_c(\mathbb{R}^d)$ 是从 $\mathbb{R}^d$ 到 $\mathbb{R}$ 的紧支连续函数. 特别地, 对于值在 $\mathbb{R}^d$ 中的随机向量 $X = (X_1, \ldots, X_d)$, 分布律 $\mathbb{P}_X$ 由所有

$$\mathbb{P}_X((a_1,b_1)\times\cdots\times(a_d,b_d)) = \mathbb{P}(a_1 < X_1 < b_1, \ldots, a_d < X_d < b_d)$$

确定, 其中 $a_1 < b_1, \ldots, a_d < b_d$. 或者, 由所有

$$\int_{\mathbb{R}^d} \varphi(x)\mathbb{P}_X(\mathrm{d}x) = \mathbb{E}[\varphi(X)]$$

确定, 其中 $\varphi \in C_c(\mathbb{R}^d)$.

命题 6.8 告诉我们可以使用分布 $\mathbb{P}_X$ 来计算 $f(X)$ 的期望. 实际上这个过程可以倒过来, 如果我们能找到 E 上的测度 ν 使得

$$\mathbb{E}[f(X)] = \int f\,\mathrm{d}\nu,$$

其中 $f : E \to \mathbb{R}$ 属于某个足够大的函数类 (如 $E = \mathbb{R}^d$ 时 f 取变所有紧支的连续函数), 此时就一定有 $\mathbb{P}_X = \nu$. 下面的命题应用了这样的思想.

命题 6.9. 令 $X = (X_1, \ldots, X_d)$ 是值在 $\mathbb{R}^d$ 中的随机变量, 假设 X 有密度 $p(x_1, \ldots, x_d)$. 那么, 对于任意 $1 \le j \le d$, X_j 的密度为

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_d)\,\mathrm{d}x_1 \cdots \mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1} \cdots \mathrm{d}x_d.$$

Proof. 记 π_j 是投影函数 $\pi_j(x_1, \ldots, x_d) = x_j$. 对于任意的 Borel 函数 $f : \mathbb{R} \to \mathbb{R}_+$, 根

据 Fubini 定理, 有

$$\begin{aligned}
\mathbb{E}[f(X_j)] &= \mathbb{E}[f \circ \pi_j(X)] \\
&= \int_{\mathbb{R}^d} f(\pi_j(x))\mathbb{P}_X(\mathrm{d}x) \\
&= \int_{\mathbb{R}^d} f(x_j)p(x_1,\dots,x_d)\,\mathrm{d}x_1\cdots\mathrm{d}x_d \\
&= \int_{\mathbb{R}} f(x_j)\left(\int_{\mathbb{R}^{d-1}} p(x_1,\dots,x_d)\,\mathrm{d}x_1\cdots\mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1}\cdots\mathrm{d}x_d\right)\mathrm{d}x_j \\
&= \int_{\mathbb{R}} f(x_j)p_j(x_j)\,\mathrm{d}x_j = \int_{\mathbb{R}} f(x_j)\mathbb{P}_{X_j}(\mathrm{d}x_j),
\end{aligned}$$

这就表明对于任意 Borel 子集 A 有

$$\mathbb{P}_{X_j}(A) = \int_A p_j(x_j)\,\mathrm{d}x_j,$$

即 X_j 有密度函数 p_j. □

如果 $X = (X_1,\dots,X_d)$ 是值在 $\mathbb{R}^d$ 中的随机变量, 那么概率测度 $\mathbb{P}_{X_j}$ 被称为 X 的**边缘分布**, 分布律 $\mathbb{P}_{X_j}$ 由 $\mathbb{P}_X$ 完全决定: $\mathbb{P}_{X_j}$ 就是 $\mathbb{P}_X$ 在投影 π_j 下的推前. 需要注意反之不是正确的, 也就是说即使确定了所有的边缘分布 $\mathbb{P}_{X_1},\dots,\mathbb{P}_{X_j}$, 也不能确定 $\mathbb{P}_X$.

6.1.4 经典分布

本小节我们列举一些重要的概率分布.

离散分布

1. **均匀分布**. 如果 E 是有限集, 值在 E 中的随机变量 X 如果满足

$$\mathbb{P}(X = x) = \frac{1}{\mathrm{card}(E)}, \quad \forall x \in E,$$

那么我们说 X 是 E 上的均匀分布.

2. **参数 $p \in [0,1]$ 的 Bernoulli 分布**. 如果值在 $\{0,1\}$ 中的随机变量 X 满足

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p,$$

那么我们说 X 是 E 上参数 p 的 Bernoulli 分布.

3. **二项分布 $\mathcal{B}(n,p)$ ($n \in \mathbb{N}, p \in [0,1]$)**. 如果值在 $\{0,1,\dots,n\}$ 中的随机变量 X 满足

$$\mathbb{P}(X = k) = \binom{n}{k} p^k(1-p)^{n-k}, \quad \forall k \in \{0,1,\dots,n\},$$

那么我们说 X 是 E 上的二项分布.

4. **参数 $p \in (0,1)$ 的几何分布**. 如果值在 $\mathbb{Z}_+$ 中的随机变量 X 使得

$$\mathbb{P}(X=k)=(1-p)p^k, \quad k \in \mathbb{Z}_+,$$

那么我们说 X 是 E 上参数 p 的几何分布.

5. **参数 $\lambda > 0$ 的 Poisson 分布**. 如果值在 $\mathbb{Z}_+$ 中的随机变量 X 使得

$$\mathbb{P}(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad \forall k \in \mathbb{Z}_+,$$

那么我们说 X 是 E 上参数 λ 的 Poisson 分布. 容易计算

$$\mathbb{E}[X]=\sum_{k=0}^{\infty} k\,\mathbb{P}(X=k)=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\lambda,$$

Poisson 分布在实际应用中非常重要, 通常被用于建模某个“罕见事件”在长时间段内发生的次数. 准确的数学叙述是 Poisson 分布是二项分布的近似. 对于每个 $n \geq 1$, 记 X_n 为服从二项分布 $\mathcal{B}(n,p_n)$ 的随机变量, 如果在 $n \to \infty$ 的时候有 $np_n \to \lambda$, 那么对于每个 $k \in \mathbb{N}$, 有

$$\lim_{n\to\infty}\mathbb{P}(X_n=k)=\frac{\lambda^k}{k!}e^{-\lambda}.$$

这可以解释为, 如果每天有很小的概率 $p_n \approx \lambda/n$ 发生地震, 那么地震在 n 天内发生的次数将近似服从泊松分布.

连续分布 在下面的五个例子中, X 都指的是一个有密度 p 的实值随机变量.

1. **$[a,b]$ 上的均匀分布**:

$$p(x)=\frac{1}{b-a}\mathbf{1}_{[a,b]}(x).$$

2. **参数 $\lambda > 0$ 的指数分布**:

$$p(x)=\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x),$$

此时对于 $a \geq 0$, 有

$$\mathbb{P}(X \geq a)=\int_a^{\infty} p(x)\,\mathrm{d}x=e^{-\lambda a}.$$

这表明指数分布有下面的重要性质: 对于 $a,b \geq 0$, 有

$$\mathbb{P}(X \geq a+b)=\mathbb{P}(X \geq a)\mathbb{P}(X \geq b). \tag{6.1}$$

3. **Gamma 分布 $\Gamma(a,\lambda)$ $(a>0,\lambda>0)$**:

$$p(x)=\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x),$$

这是指数分布的推广, $a=1$ 时即指数分布.

4. **参数 $a > 0$ 的 Cauchy 分布**:

$$p(x) = \frac{1}{\pi}\frac{a}{a^2 + x^2},$$

注意到服从 Cauchy 分布的随机变量的数学期望是不存在的, 因为

$$\mathbb{E}[|X|] = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{a|x|}{a^2 + x^2}\,\mathrm{d}x = \infty.$$

5. **正态分布 $\mathcal{N}(m, \sigma^2)$ ($m \in \mathbb{R}, \sigma > 0$)**:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

正态分布与 Poisson 分布一起成为概率论中最重要的两个分布. 正态分布的密度曲线呈著名的钟形曲线. 按定义很容易验证

$$m = \mathbb{E}[X], \quad \sigma^2 = \mathbb{E}[(X-m)^2].$$

对于 $a, b \in \mathbb{R}$, 考虑随机变量 $Y = aX + b$, 那么对于任意的 Borel 函数 $f : \mathbb{R} \to \mathbb{R}_+$, 有

$$\begin{aligned}\mathbb{E}[f(Y)] = \mathbb{E}[f(aX+b)] &= \int_{\mathbb{R}} f(ax+b)\mathbb{P}_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} f(ax+b)p(x)\,\mathrm{d}x = \frac{1}{a}\int_{\mathbb{R}} f(y)p\left(\frac{y-b}{a}\right)\mathrm{d}y \\ &= \int_{\mathbb{R}} f(y)\frac{1}{a}p\left(\frac{y-b}{a}\right)\mathrm{d}y,\end{aligned}$$

这表明

$$p(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-(am+b))^2}{2(a\sigma)^2}\right),$$

即 $aX + b$ 服从分布 $\mathcal{N}(am+b, a^2\sigma^2)$.

6.1.5 实值随机变量的分布函数

令 X 是实值随机变量, 定义 X 的**分布函数**为 $F_X : \mathbb{R} \to [0,1]$, 其满足

$$F_X(t) = \mathbb{P}(X \le t) = \mathbb{P}_X((-\infty, t]), \quad \forall t \in \mathbb{R}.$$

根据推论 1.24, F_X 实际上完全刻画了分布 $\mathbb{P}_X$. 确切的说, 如果知道了 F_X, 即相当于知道了所有 $\mathbb{P}_X((-\infty, t])$ 的值, 而所有区间 $(-\infty, t]$ 构成的子集族对有限交封闭, 又因为 $\mathbb{P}_X$ 为有限测度, 所以 $\mathbb{P}_X$ 在所有区间 $(-\infty, t]$ 上的值可以完全确定 $\mathbb{P}_X$ 在 $\mathcal{B}(\mathbb{R})$ 上的值.

显然函数 F_X 是递增的、右连续的并且在 $-\infty$ 处极限为 0、在 $+\infty$ 处极限为 1. 反之, 如果 $F : \mathbb{R} \to [0,1]$ 满足上面的性质, 定理 3.12 表明存在 (唯一的) $\mathbb{R}$ 上的概率测

度 μ 使得 $\mu((-\infty, t]) = F(t)$. 即这样的函数 F 总能解释为某个实值随机变量的分布函数.

令 $F_X(a-)$ 表示 F_X 在 $a \in \mathbb{R}$ 处的左极限. 那么容易验证

$$\mathbb{P}(a \le X \le b) = F_X(b) - F_X(a-),$$
$$\mathbb{P}(a < X < b) = F_X(b-) - F_X(a).$$

特别的, $\mathbb{P}(X = a) = F_X(a) - F_X(a-)$. 这表明 F_X 的间断点的个数恰为 $\mathbb{P}_X$ 的原子个数.

6.1.6 由随机变量生成的 σ-域

定义 6.10. 令 X 是值在 $(E, \mathcal{E})$ 中的随机变量，定义由 X 生成的 σ-**域**为

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{E}\} \subseteq \mathcal{A}.$$

换句话说，这是使得 $\omega \mapsto X(\omega)$ 可测的最小 σ-域，记为 $\sigma(X)$.

$\sigma(X)$ 的定义可以延拓到任意族随机变量 $(X_i)_{i\in I}$ 上, 其中 X_i 是值在 $(E_i, \mathcal{E}_i)$ 中的随机变量. 这种情况下, 定义

$$\sigma\big((X_i)_{i\in I}\big) = \sigma\left(\left\{X_i^{-1}(B_i) \;\middle|\; B_i \in \mathcal{E}_i\right\}\right).$$

下面的命题表明一个实值随机变量是 $\sigma(X)$-可测的当且仅当其是 X 的一个可测函数. 这对于研究条件非常重要.

命题 6.11. 令 X 是值在 $(E, \mathcal{E})$ 中的随机变量，Y 是实值随机变量，那么下面的说法等价：

1. Y 是 $\sigma(X)$-可测的；
2. 存在 $(E, \mathcal{E})$ 到 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 的可测函数 f 使得 $Y = f(X)$.

Proof. (2) $\Rightarrow$ (1) 是显然的, 因为 X 是 $\sigma(X)$-可测的, 所以 $Y = f(X)$ 也是 $\sigma(X)$-可测的.

(1) $\Rightarrow$ (2) 假设 Y 是 $\sigma(X)$-可测的. 首先考虑 Y 是简单函数, 即

$$Y = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{A_i},$$

其中 λ_i 是不同的实数, $A_i = Y^{-1}(\lambda_i) \in \sigma(X)$. 那么, 对于每个 i, 存在 $B_i \in \mathcal{E}$ 使得 $A_i = X^{-1}(B_i)$. 定义函数 $f : E \to \mathbb{R}$ 为

$$f(x) = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{B_i}(x).$$

显然 f 是 $\mathcal{E}$-可测的函数, 并且

$$Y = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{A_i} = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{X^{-1}(B_i)} = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{B_i} \circ X = f \circ X.$$

一般情况下, 假设 Y 是一列 $\sigma(X)$-可测的简单函数 $(Y_n)_{n\in\mathbb{N}}$ 的极限. 对于每个 n, 根据上面的讨论, 存在 $\mathcal{E}$-可测函数 $f_n : E \to \mathbb{R}$ 使得 $Y_n = f_n(X)$. 对于每个 $x \in E$, 定义

$$f(x) = \begin{cases} \lim\limits_{n\to\infty} f_n(x), & \text{如果极限存在;} \\ 0, & \text{否则.} \end{cases}$$

对于每个 $\omega \in \Omega$, 由于 $Y(\omega) = \lim Y_n(\omega) = \lim f_n(X(\omega))$, 所以 $X(\omega)$ 使得上述极限存在, 所以 $Y(\omega) = f(X(\omega))$. □

6.2 随机变量的矩

6.2.1 矩和方差

令 X 是实值随机变量, $p \in \mathbb{N}$. 定义 X 的 p-**阶矩**为 $\mathbb{E}[X^p]$, 其仅在 $X \geq 0$ 或者 $\mathbb{E}[|X|^p] < \infty$ 的时候有定义. 如果 X 满足 $\mathbb{E}[X] = 0$, 那么我们说 X 是**中心化**的.

因为期望是相对于测度 $\mathbb{P}_X$ 的一种积分, 所以我们有下面的结果. 如果 X 是值在 $[0, \infty]$ 中的随机变量, 那么我们有

- $\mathbb{E}[X] < \infty \Rightarrow X < \infty,\ \mathbb{P}_X$ a.s.
- $\mathbb{E}[X] = 0 \Rightarrow X = 0\ \mathbb{P}_X$ a.s.

此外, 各种极限与积分交换次序的定理也可以直接改写为期望的形式:

- **单调收敛定理**. 如果 $(X_n)_{n\in\mathbb{N}}$ 是一列值在 $[0, \infty]$ 中递增的随机变量, 那么

$$\lim_{n\to\infty} \uparrow \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} \uparrow X_n\right].$$

- **Fatou 引理**. 如果 $(X_n)_{n\in\mathbb{N}}$ 是一列值在 $[0, \infty]$ 中的随机变量, 那么

$$\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n].$$

- **控制收敛定理**. 如果 $(X_n)_{n\in\mathbb{N}}$ 是一列实值随机变量, 并且存在值在 $[0, \infty]$ 中的随机变量 Z 使得

$$|X_n| \leq Z, \quad \mathbb{E}[Z] < \infty, \quad X_n \to X,\ \mathbb{P}_X \text{ a.s.}$$

  那么

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \mathbb{E}[X], \quad \lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0.$$

对于每个 $p \in [1, \infty]$, 考虑空间 $L^p(\Omega, \mathcal{A}, \mathbb{P})$. Hölder 不等式表明对于任意实值随机变量 X, Y, 如果 $p, q \in (1, \infty)$ 使得 $1/p + 1/q = 1$, 那么

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y|^q]^{1/q}.$$

取 $Y = 1$, 我们得到 $\|X\|_1 \leq \|X\|_p$. 此外, 如果 $1 \leq p < q \leq \infty$, 有 $\|X\|_p \leq \|X\|_q$, 这也表明 $L^q(\Omega, \mathcal{A}, \mathbb{P}) \subseteq L^p(\Omega, \mathcal{A}, \mathbb{P})$.

Hilbert 空间 $L^2(\Omega,\mathcal{A},\mathbb{P})$ 上的内积定义为 $\langle X,Y\rangle_{L^2}=\mathbb{E}[XY]$, Cauchy-Schwarz 不等式表明

$$\mathbb{E}[|XY|]\le\mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}.$$

特别地, 我们有

$$\mathbb{E}[|X|]^2\le\mathbb{E}[X^2].$$

如果 $X\in L^1(\Omega,\mathcal{A},\mathbb{P})$, $f:\mathbb{R}\to\mathbb{R}_+$ 是凸函数, 那么 Jensen 不等式表明

$$\mathbb{E}[f(X)]\ge f(\mathbb{E}[X]).$$

定义 6.12. 令 $X\in L^2(\Omega,\mathcal{A},\mathbb{P})$，定义 X 的**方差**为

$$\operatorname{var}(X)=\mathbb{E}\big[(X-\mathbb{E}[X])^2\big]\ge 0,$$

X 的**标准差**为

$$\sigma_X=\sqrt{\operatorname{var}(X)}.$$

命题 6.13. 令 $X\in L^2(\Omega,\mathcal{A},\mathbb{P})$，方差 $\operatorname{var}(X)=\mathbb{E}[X^2]-\big(\mathbb{E}[X]\big)^2$. 对于任意的 $a\in\mathbb{R}$，有

$$\mathbb{E}[(X-a)^2]=\operatorname{var}(X)+\big(\mathbb{E}[X]-a\big)^2.$$

因此，还有

$$\operatorname{var}(X)=\inf_{a\in\mathbb{R}}\mathbb{E}[(X-a)^2].$$

下面的两个不等式是非常重要的工具.

Markov 不等式 如果 X 是非负随机变量并且 $a>0$, 那么根据命题 2.10, 有

$$\mathbb{P}(X\ge a)\le\frac{1}{a}\mathbb{E}[X].$$

Bienaymé-Chebyshev 不等式 如果 $X\in L^2(\Omega,\mathcal{A},\mathbb{P})$ 并且 $a>0$, 对 $(X-\mathbb{E}[X])^2$ 和 a^2 应用 Markov 不等式, 得到

$$\mathbb{P}\big(\big|X-\mathbb{E}[X]\big|\ge a\big)\le\frac{1}{a^2}\operatorname{var}(X).$$

定义 6.14. 令 $X,Y\in L^2(\Omega,\mathcal{A},\mathbb{P})$. 定义 X 和 Y 的**协方差**为

$$\operatorname{cov}(X,Y)=\mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big]=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y].$$

如果 $Z=(Z_1,\dots,Z_d)$ 是值在 $\mathbb{R}^d$ 中的随机变量并且所有分量都属于 $L^2(\Omega,\mathcal{A},\mathbb{P})$ (等价的说 $\mathbb{E}[|Z|^2]<\infty$)，定义 Z 的**协方差矩阵**为

$$K_Z=\Big(\operatorname{cov}(Z_i,Z_j)\Big)_{1\le i,j\le d}.$$

直觉上, X 和 Y 的协方差衡量了 X 和 Y 之间的相关性. 可以注意到 $\mathrm{cov}(X,X)=\mathrm{var}(X)$, 还有 Cauchy-Schwarz 不等式表明

$$|\mathrm{cov}(X,Y)|\le\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}.$$

映射 $(X,Y)\mapsto\mathrm{cov}(X,Y)$ 给出了 $L^2(\Omega,\mathcal{A},\mathbb{P})$ 上的一个对称的双线性映射.

对于一个随机向量 $Z=(Z_1,\dots,Z_d)$, 矩阵 K_Z 是对称的半正定矩阵: 对于每个 $\lambda_1,\dots,\lambda_d\in\mathbb{R}$, 有

$$\begin{aligned}\sum_{i,j=1}^{d}\lambda_i\lambda_jK_Z(i,j)&=\sum_{i,j=1}^{d}\lambda_i\lambda_j\left(\mathbb{E}[Z_iZ_j]-\mathbb{E}[Z_i]\mathbb{E}[Z_j]\right)\\&=\mathbb{E}\Bigg[\sum_{i,j=1}^{d}\lambda_i\lambda_jZ_iZ_j\Bigg]-\sum_{i,j=1}^{d}\lambda_i\lambda_j\mathbb{E}[Z_i]\mathbb{E}[Z_j]\\&=\mathbb{E}\Bigg[\Bigg(\sum_{i=1}^{d}\lambda_iZ_i\Bigg)^2\Bigg]-\Bigg(\mathbb{E}\Bigg[\sum_{i=1}^{d}\lambda_iZ_i\Bigg]\Bigg)^2\\&=\mathrm{var}\Bigg(\sum_{i=1}^{d}\lambda_iZ_i\Bigg)\ge0.\end{aligned}$$

令 $\widetilde{Z}=Z-\mathbb{E}[Z]$. 如果我们把 $\widetilde{Z}$ 视为列向量, 那么我们可以把协方差矩阵写成 $K_Z=\mathbb{E}[\widetilde{Z}\widetilde{Z}^T]$. 因此, 如果 A 是一个 $n\times d$ 实矩阵, 并且 $Z'=AZ$, 那么

$$K_{Z'}=\mathbb{E}\big[\widetilde{Z}'\widetilde{Z}'^T\big]=\mathbb{E}\big[A\widetilde{Z}\widetilde{Z}A^T\big]=AK_ZA^T. \tag{6.2}$$

作为特殊情况, 如果 $\xi\in\mathbb{R}^d$, 点积 $\xi\cdot Z$ 可以写为 ξ^TZ, 于是

$$\mathrm{var}(\xi\cdot Z)=\xi^TK_Z\xi. \tag{6.3}$$

6.3 线性回归

令 $X,Y_1,\dots,Y_n$ 是 $L^2(\Omega,\mathcal{A},\mathbb{P})$ 中的实值随机变量. 记 $Y=(Y_1,\dots,Y_n)$. 我们来寻找最接近 X 的 $Y_1,\dots,Y_n$ 的仿射函数. 准确地说, 我们需要寻找实数 $(\beta_0,\beta_1,\dots,\beta_n)$ 最小化

$$\mathbb{E}\big[\big(X-(\beta_0+\beta_1Y_1+\cdots+\beta_nY_n)\big)^2\big].$$

命题 6.15. 我们有

$$\inf_{\beta_0,\beta_1,\dots,\beta_n\in\mathbb{R}}\mathbb{E}\big[\big(X-(\beta_0+\beta_1Y_1+\cdots+\beta_nY_n)\big)^2\big]=\mathbb{E}\big[(X-Z)^2\big],$$

其中

$$Z=\mathbb{E}[X]+\sum_{j=1}^{n}\alpha_j(Y_j-\mathbb{E}[Y_j]), \tag{6.4}$$

系数 α_j 是下述线性方程组的任意解：

$$\sum_{j=1}^{n} \alpha_j \operatorname{cov}(Y_j, Y_k) = \operatorname{cov}(X, Y_k), \quad 1 \le k \le n.$$

特别的，如果 K_Y 是可逆的，那么我们有 $\alpha = \operatorname{cov}(X,Y)K_Y^{-1}$，这里 $\operatorname{cov}(X,Y)$ 表示向量 $(\operatorname{cov}(X,Y_j))_{1\le j\le n}$.

6.4 特征函数

定义 6.16. 令 X 是值在 $\mathbb{R}^d$ 中的随机变量，$t \in \mathbb{R}^d$. 定义 X 的**特征函数** $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ 为

$$\Phi_X(\xi) = \mathbb{E}\left[e^{i\xi\cdot X}\right] = \int_{\mathbb{R}^d} e^{i\xi\cdot x}\,\mathbb{P}_X(\mathrm{d}x).$$

由于 $|e^{i\xi\cdot x}| = 1$, 所以任意随机变量 X 的特征函数 Φ_X 都是良定义的. 可以发现 Φ_X 就是分布律 $\mathbb{P}_X$ 的 Fourier 变换, 所以我们也时常记作 $\Phi_X(\xi) = \widehat{\mathbb{P}}_X(\xi)$. 利用控制收敛定理, 可以验证 Φ_X 在 $\mathbb{R}^d$ 上连续.

我们的目标是证明特征函数确定了分布律 $\mathbb{P}_X$. 这等价于说明 Fourier 变换 $\mathbb{P}_X \mapsto \widehat{\mathbb{P}}_X$ 是单射的映射.

引理 6.17. 令 X 是服从正态分布 $\mathcal{N}(m,\sigma^2)$ 的实值随机变量，那么

$$\Phi_X(\xi) = \exp\left(im\xi - \frac{\sigma^2\xi^2}{2}\right), \quad \forall \xi \in \mathbb{R}.$$

Proof. 不妨设 $\sigma > 0$ 且用 $X - m$ 替代 X, 也即 $m = 0$. 那么

$$\Phi_X(\xi) = \int_{\mathbb{R}} e^{i\xi x}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x.$$

通过换元, 不妨设 $\sigma = 1$. 并且扔掉奇函数的部分, 只需要计算含参积分

$$f(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\cos(\xi x)\,\mathrm{d}x.$$

由于 $|xe^{-x^2/2}\sin(\xi x)| \le |x|e^{-x^2/2}$ 是可积函数, 所以

$$f'(\xi) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2}\sin(\xi x)\,\mathrm{d}x.$$

再通过分部积分, 有

$$f'(\xi) = -\xi\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\cos(\xi x)\,\mathrm{d}x = -\xi f(\xi).$$

且初值条件 $f(0) = 1$, 解这个微分方程, 就得到 $f(\xi) = e^{-\xi^2/2}$. □

定理 6.18. 值在 $\mathbb{R}^d$ 中的随机变量 X 的特征函数确定了它的分布律. 换句话说，$\mathbb{R}^d$ 上概率测度的 Fourier 变换算子是单射.

Proof. 考虑 $d=1$ 的情况. 对于每个 $\sigma>0$, 令 g_σ 是正态分布 $\mathcal{N}(0,\sigma^2)$ 的密度

$$g_\sigma(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right).$$

令 μ 是 $\mathbb{R}$ 上的概率测度, 令

$$\begin{aligned}f_\sigma(x)&=\int_{\mathbb{R}}g_\sigma(x-y)\mu(\mathrm{d}y)=g_\sigma*\mu(x),\\ \mu_\sigma(\mathrm{d}x)&=f_\sigma(x)\,\mathrm{d}x.\end{aligned}$$

记 $C_b(\mathbb{R})$ 是所有有界连续函数 $\varphi:\mathbb{R}\to\mathbb{R}$ 构成的空间. 回顾 $\mathbb{R}$ 上的概率测度 ν 由所有 $\int\varphi(x)\nu(\mathrm{d}x)$ 刻画, 其中 $\varphi\in C_c(\mathbb{R})\subseteq C_b(\mathbb{R})$. 我们证明下面两点:

1. μ_σ 被 $\hat{\mu}$ 确定.
2. 对于每个 $\varphi\in C_b(\mathbb{R})$, 当 $\sigma\to0$ 时有 $\int\varphi(x)\mu_\sigma(\mathrm{d}x)\to\int\varphi(x)\mu(\mathrm{d}x)$.

根据引理 6.17, 对于任意 $x\in\mathbb{R}$, 有

$$\sigma\sqrt{2\pi}g_\sigma(x)=\exp\left(-\frac{x^2}{2\sigma^2}\right)=\int_{\mathbb{R}}e^{i\xi x}g_{1/\sigma}(\xi)\,\mathrm{d}\xi.$$

因此, 有

$$\begin{aligned}f_\sigma(x)=\int_{\mathbb{R}}g_\sigma(x-y)\mu(\mathrm{d}y)&=\left(\sigma\sqrt{2\pi}\right)^{-1}\int_{\mathbb{R}}\left(\int_{\mathbb{R}}e^{i\xi(x-y)}g_{1/\sigma}(\xi)\,\mathrm{d}\xi\right)\mu(\mathrm{d}y)\\&=\left(\sigma\sqrt{2\pi}\right)^{-1}\int_{\mathbb{R}}e^{i\xi x}g_{1/\sigma}(\xi)\left(\int_{\mathbb{R}}e^{-i\xi y}\mu(\mathrm{d}y)\right)\mathrm{d}\xi\\&=\left(\sigma\sqrt{2\pi}\right)^{-1}\int_{\mathbb{R}}e^{i\xi x}g_{1/\sigma}(\xi)\hat{\mu}(-\xi)\,\mathrm{d}\xi.\end{aligned}$$

注意我们使用了 Fubini-Tonelli 定理来交换积分次序. 这就表明 f_σ 完全由 $\hat{\mu}$ 确定, 所以 μ_σ 也由 $\hat{\mu}$ 确定.

下面证明 (2). 对于每个 $\varphi\in C_b(\mathbb{R})$, 有

$$\begin{aligned}\int\varphi(x)\mu_\sigma(\mathrm{d}x)=\int\varphi(x)\left(\int g_\sigma(x-y)\mu(\mathrm{d}y)\right)\mathrm{d}x&=\int\left(\int\varphi(x)g_\sigma(x-y)\,\mathrm{d}x\right)\mu(\mathrm{d}y)\\&=\int g_\sigma*\varphi(y)\mu(\mathrm{d}y).\end{aligned}$$

这里再次使用了 Fubini-Tonelli 定理来交换积分次序. 注意到

$$\begin{aligned}\int g_\sigma(x)\,\mathrm{d}x&=1,\\ \lim_{\sigma\to0}\int_{\{|x|>\varepsilon\}}g_\sigma(x)\,\mathrm{d}x&=0,\quad\forall\varepsilon>0.\end{aligned}$$

这表明对于任意 $y\in\mathbb{R}$, 有

$$\lim_{\sigma\to0}g_\sigma*\varphi(y)=\varphi(y).$$

根据控制收敛, 注意到 $|g_\sigma * \varphi| \le \sup\{|\varphi(x)| \mid x \in \mathbb{R}\}$, 我们有

$$\lim_{\sigma\to 0}\int \varphi(x)\mu_\sigma(\mathrm{d}x) = \lim_{\sigma\to 0}\int g_\sigma * \varphi(y)\mu(\mathrm{d}y) = \int \varphi(x)\mu(\mathrm{d}x).$$

这就证明了 (2).

对于一般 d 维的情况, 使用辅助函数

$$g_\sigma^{(d)}(x_1,\ldots,x_d) = \prod_{j=1}^{d} g_\sigma(x_j)$$

即可. 此时, 对于 $\xi = (\xi_1,\ldots,\xi_d) \in \mathbb{R}^d$ 有

$$\int_{\mathbb{R}^d} g_\sigma^{(d)}(x)e^{i\xi\cdot x}\,\mathrm{d}x = \prod_{j=1}^{d}\int_{\mathbb{R}} g_\sigma(x_j)e^{i\xi_j x_j}\,\mathrm{d}x_j = (2\pi/\sigma^2)^{d/2} g_{1/\sigma}^{(d)}(\xi),$$

然后采取同样的论述即可. □

7

独立性

7.1 独立事件

在本章中, 我们考虑概率空间 $(\Omega, \mathcal{A}, \mathbb{P})$. 如果 $A, B \in \mathcal{A}$ 且

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

那么我们说 A 和 B 是**独立的**. 如果 $\mathbb{P}(B) > 0$, 我们定义条件概率

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

此时 A 和 B 独立等价于 $\mathbb{P}(A|B) = \mathbb{P}(A)$.

定义 7.1. 如果对于 $\{1, \ldots, n\}$ 的任意子集 $\{j_1, \ldots, j_p\}$ 都有

$$\mathbb{P}(A_{j_1} \cap \cdots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \cdots \mathbb{P}(A_{j_p}),$$

那么我们说 n 个事件 $A_1, \ldots, A_n$ 是独立的.

命题 7.2. n 个事件 $A_1, \ldots, A_n$ 独立当且仅当

$$\mathbb{P}(B_1 \cap \cdots \cap B_n) = \mathbb{P}(B_1) \cdots \mathbb{P}(B_n),$$

其中 $B_i \in \sigma(A_i) = \{\emptyset, A, A^c, \Omega\}$.

7.2 σ-域和随机变量的独立性

如果 $\mathcal{B} \subseteq \mathcal{A}$ 是一个 σ-域, 那么我们说 $\mathcal{B}$ 是 $\mathcal{A}$ 的**子 σ-域**. 我们可以认为子 σ-域 $\mathcal{B}$ 反映了概率空间的部分信息, 即 $\mathcal{B}$ 中发生的事件. 例如, 如果 $\mathcal{B} = \sigma(X)$, X 是随机变量, 那么 $\mathcal{B}$ 反映了 X 的值的信息. 这暗示了子 σ-域的独立性的概念: 我们希望两个子 σ-域 $\mathcal{B}$ 和 $\mathcal{B}'$ 是独立的当且仅当它们中的任意两个事件都是独立的.

定义 7.3. 令 $\mathcal{B}_1, \ldots, \mathcal{B}_n$ 是 $\mathcal{A}$ 的 n 个 σ-子域, 我们说 $\mathcal{B}_1, \ldots, \mathcal{B}_n$ 是独立的, 如果对于任意 $A_1 \in \mathcal{B}_1, \ldots, A_n \in \mathcal{B}_n$, 都有

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

令 $X_1, \dots, X_n$ 分别是值在 $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ 中的随机变量, 我们说 $X_1, \dots, X_n$ 是独立的当且仅当 $\sigma(X_1), \dots, \sigma(X_n)$ 是独立的. 这等价于任取 $F_1 \in \mathcal{E}_1, \dots, F_n \in \mathcal{E}_n$ 有

$$\mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) = \mathbb{P}(X_1 \in F_1) \cdots \mathbb{P}(X_n \in F_n).$$

有时候考虑一族随机变量之间的独立性也是有用的. 对于两族随机变量 $(X_i)_{i \in I}$ 和 $(Y_j)_{j \in J}$, 如果 σ-域 $\sigma((X_i)_{i \in I})$ 和 $\sigma((Y_j)_{j \in J})$ 是独立的, 那么我们说这两族随机变量是独立的. 注意到这比任意 X_i 与 Y_j 独立要更强.

注释 7.4. (1) 如果 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 是 $\mathcal{A}$ 的 n 个独立的 σ-子域, 并且 X_i 是 $\mathcal{B}_i$-可测的随机变量, 那么 $X_1, \dots, X_n$ 是独立的. 这是因为 X_i 是 $\mathcal{B}_i$-可测的就表明 $\sigma(X_i) \subseteq \mathcal{B}_i$.

(2) n 个事件 $A_1, \dots, A_n$ 是独立的当且仅当 σ-域 $\sigma(A_1), \dots, \sigma(A_n)$ 是独立的 (命题 7.2).

定理 7.5. 令 $X_1, \dots, X_n$ 分别是值在 $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ 中的随机变量. 那么 $X_1, \dots, X_n$ 是独立的当且仅当 $(X_1, \dots, X_n)$ 的分布是 $X_1, \dots, X_n$ 的分布的乘积测度, 即

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

此外, 我们有

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)],$$

其中 f_i 是 $(E_i, \mathcal{E}_i)$ 上的非负可测函数.

Proof. 令 $F_i \in \mathcal{E}_i$, $X_1, \dots, X_n$ 独立当且仅当

$$\mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) = \mathbb{P}(X_1 \in F_1) \cdots \mathbb{P}(X_n \in F_n),$$

这表明

$$\mathbb{P}_{(X_1, \dots, X_n)}(F_1 \times \cdots \times F_n) = \mathbb{P}_{X_1}(F_1) \cdots \mathbb{P}_{X_n}(F_n).$$

所以 $\mathbb{P}_{(X_1, \dots, X_n)}$ 就是乘积测度 $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$.

记 $\pi_i : E_1 \times \cdots \times E_n \to E_i$ 为投影, 根据 Fubini 定理, 有

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] &= \int_{E_1 \times \cdots \times E_n} \prod_{i=1}^n f_i \circ \pi_i \, \mathrm{d}\mathbb{P}_{(X_1, \dots, X_n)} \\
&= \prod_{i=1}^n \int_{E_i} f_i(x_i) \mathbb{P}_{X_i}(\mathrm{d}x_i) \\
&= \prod_{i=1}^n \mathbb{E}[f_i(X_i)].
\end{aligned}$$

□

注释 7.6. 该定理中 f_i 也可以是任意符号, 此时如果 $\mathbb{E}[|f_i(X_i)|] < \infty$, 也即 $f_i \in \mathcal{L}^1(E_i, \mathcal{E}_i, \mathbb{P}_{X_i})$, 那么定理的结论依然成立. 只需要注意到

$$\mathbb{E}\left[\prod_{i=1}^n |f_i(X_i)|\right] = \prod_{i=1}^n \mathbb{E}[|f_i(X_i)|] < \infty.$$

特别地，如果 $X_1,\dots,X_n$ 是独立的实值随机变量且 $X_i\in L^1$，那么 $X_1\cdots X_n\in L^1$ 且

$$\mathbb{E}[X_1\cdots X_n]=\prod_{i=1}^n \mathbb{E}[X_i].$$

需要注意一般情况下 L^1 中的随机变量的积不一定还在 L^1 中.

定理 7.5 展示了如何构造有限多个独立的随机变量. 考虑实值随机变量的情况. 令 $\mu_1,\dots,\mu_n$ 是 $\mathbb{R}$ 上的 n 个概率测度. 首先我们可以构造一个值在 $\mathbb{R}^n$ 中的随机变量 $Y=(Y_1,\dots,Y_n)$ 使得其有分布律 $\mu_1\otimes\cdots\otimes\mu_n$ (比如令 $\Omega=\mathbb{R}^n$, $\mathcal{A}=\mathcal{B}(\mathbb{R}^n)$, $\mu_1\otimes\cdots\otimes\mu_n$ 是概率测度, 那么随机变量 $Y(\omega)=\omega$ 即符合要求). 然后定理 7.5 表明分量 $Y_1,\dots,Y_n$ 是分别拥有分布律 $\mu_1,\dots,\mu_n$ 的独立实值随机变量.

推论 7.7. 如果 X_1,X_2 是 L^2 中的两个独立实值随机变量，那么 $\mathrm{cov}(X_1,X_2)=0$.

推论 7.8. 令 $X_1,\dots,X_n$ 是实值随机变量.

1. 假设 X_i 有密度 p_i 并且 $X_1,\dots,X_n$ 是独立的，那么 $(X_1,\dots,X_n)$ 有密度

$$p(x_1,\dots,x_n)=\prod_{i=1}^n p_i(x_i).$$

2. 反之，假设 $(X_1,\dots,X_n)$ 有密度 p，并且 p 可以表达为

$$p(x_1,\dots,x_n)=\prod_{i=1}^n q_i(x_i),$$

其中 q_i 是 $\mathbb{R}$ 上的非负可测函数. 那么 $X_1,\dots,X_n$ 是独立的并且 X_i 有密度 $p_i=C_iq_i$，其中 $C_i>0$ 为常数.

Proof. (1) X_i 有密度 p_i 表明 $\mathbb{P}_{X_i}(\mathrm{d}x)=p_i(x)\,\mathrm{d}x$, $X_1,\dots,X_n$ 独立表明

$$\begin{aligned}\mathbb{P}_{(X_1,\dots,X_n)}(A)=\mathbb{P}_{X_1}\otimes\cdots\otimes\mathbb{P}_{X_n}(A)&=\int_{\mathbb{R}^n}\mathbf{1}_A\,\mathrm{d}\mathbb{P}_{X_1}\otimes\cdots\otimes\mathbb{P}_{X_n}\\&=\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\mathbf{1}_A(x_1,\dots,x_n)\mathbb{P}_{X_1}(\mathrm{d}x_1)\cdots\mathbb{P}_{X_n}(\mathrm{d}x_n)\\&=\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\mathbf{1}_A(x_1,\dots,x_n)\prod_{i=1}^n p_i(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_n\\&=\int_{\mathbb{R}^n}\mathbf{1}_A\prod_{i=1}^n p_i\,\mathrm{d}\lambda=\int_A\prod_{i=1}^n p_i(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_n,\end{aligned}$$

这就表明

$$p(x_1,\dots,x_n)=\prod_{i=1}^n p_i(x_i).$$

(2) 根据命题 6.9, 有

$$p_i(x_i)=\int_{\mathbb{R}^{n-1}}p(x_1,\dots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\cdots\mathrm{d}x_n=q_i(x_i)\prod_{j\neq i}\int_{\mathbb{R}}q_j(x_j)\,\mathrm{d}x_j,$$

故 $p_i = C_i q_i$. 此时

$$p(x_1,\ldots,x_n) = \prod_{i=1}^n q_i(x_i) = \prod_{i=1}^n \frac{1}{C_i} p_i(x_i),$$

两边积分可知 $\prod_{i=1}^n C_i = 1$, 所以

$$p(x_1,\ldots,x_n) = \prod_{i=1}^n p_i(x_i),$$

这就表明 $\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \mathbb{P}_{X_n}$, 即 $X_1,\ldots,X_n$ 独立. □

例 7.9. 令 U 是服从参数 1 的指数分布的随机变量, V 是服从 $[0,1]$ 上的均匀分布的随机变量, 假设 U, V 是独立的, 记

$$X = \sqrt{U}\cos(2\pi V), \quad Y = \sqrt{U}\sin(2\pi V),$$

证明 X, Y 是独立的随机变量.

Proof. 任取非负可测函数 $\varphi : \mathbb{R}^2 \to \mathbb{R}$, 有

$$\begin{aligned}
\mathbb{E}[\varphi(X,Y)] &= \mathbb{E}\Big[\varphi\big(\sqrt{U}\cos(2\pi V), \sqrt{U}\sin(2\pi V)\big)\Big] \\
&= \int_{\mathbb{R}^2} \varphi\big(\sqrt{u}\cos(2\pi v), \sqrt{u}\sin(2\pi v)\big)\,\mathrm{d}\mathbb{P}_{(U,V)} \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi\big(\sqrt{u}\cos(2\pi v), \sqrt{u}\sin(2\pi v)\big)e^{-u}\mathbf{1}_{[0,\infty)}(u)\mathbf{1}_{[0,1]}(v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \int_0^\infty\int_0^1 \varphi\big(\sqrt{u}\cos(2\pi v), \sqrt{u}\sin(2\pi v)\big)e^{-u}\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{1}{\pi}\int_0^\infty\int_0^{2\pi} \varphi(r\cos\theta, r\sin\theta)re^{-r^2}\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \frac{1}{\pi}\int_{\mathbb{R}^2} \varphi(x,y)e^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

这表明 (X,Y) 有概率密度 $p(x,y) = \pi^{-1}\exp(-x^2-y^2) = \pi^{-1}\exp(-x^2)\exp(-y^2)$, 根据 推论 7.7, 这表明 X, Y 是独立的. □

7.3 Borel-Cantelli 引理

回顾集合极限的定义: 如果 $(A_n)_{n\in\mathbb{N}}$ 是一列集合, 我们记

$$\limsup A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k,$$

不难发现点 $\omega \in \limsup A_n$ 当且仅当存在无限多个 n 使得 $\omega \in A_n$. 注意到 1.12 告诉我们 $\mathbb{P}(\limsup A_n) \geq \limsup \mathbb{P}(A_n)$.

引理 7.10. 令 $(A_n)_{n\in\mathbb{N}}$ 是一列事件.

1. 如果 $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$，那么
$$\mathbb{P}(\limsup A_n) = 0,$$
2. 如果 $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$，事件 A_n 是独立的，那么
$$\mathbb{P}(\limsup A_n) = 1.$$

两个应用 (2) 我们令
$$(\Omega, \mathcal{A}, \mathbb{P}) = \big([0,1), \mathcal{B}([0,1)), \lambda\big),$$
其中 λ 表示 Lebesgue 测度. 对于每个 $n \in \mathbb{N}$, 令
$$X_n(\omega) = \lfloor 2^n\omega\rfloor - 2\lfloor 2^{n-1}\omega\rfloor,$$
其中 $\lfloor x\rfloor$ 表示向下取整. 那么 $X_n(\omega) \in \{0,1\}$ 并且容易验证对于任意 $\omega \in [0,1)$ 有
$$0 \le \omega - \sum_{k=1}^{n} X_k(\omega)2^{-k} < 2^{-n}.$$
这表明
$$\omega = \sum_{k=1}^{\infty} X_k(\omega)2^{-k}.$$

7.4 独立随机变量的和

命题 7.11 (强大数定律). 令 $(X_n)_{n\in\mathbb{N}}$ 是一列独立同分布的实值随机变量，若 $\mathbb{E}[X_1^4] < \infty$，那么我们几乎肯定有
$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{} \mathbb{E}[X_1].$$

7.5 Poisson 过程

在本节, 我们固定一个参数 $\lambda > 0$, 令 $U_1, U_2, \ldots$ 是一列独立同分布的随机变量, 它们都服从参数 λ 的指数分布, 即有概率密度 $\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}$. 令
$$T_n = U_1 + U_2 + \cdots + U_n.$$
对于任意实数 $t \ge 0$, 令
$$N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \le t\}} = \sup\{n \in \mathbb{N} \mid T_n \le t\}.$$

约定 $\sup\emptyset = 0$. 命题 7.10 告诉我们在 $n\to\infty$ 的时候, $T_n\to\infty$ a.s.. 因此, 使得 $(T_n)_{n\in\mathbb{N}}$ 有界的 $\omega\in\Omega$ 的集合构成一个零测集, 除开这个零测集, 对于任意 $t\geq 0$, 我们都有 $N_t<\infty$. 类似地, 因为随机变量 U_i 几乎处处是正值, 我们可以假设对于每个 $\omega\in\Omega$ 都有 $0<T_1(\omega)<T_2(\omega)<\cdots$.

固定 ω, 函数 $t\mapsto N_t(\omega)$ 在 0 处为零、单调递增且右连续, 此外其每次以大小为 1 的幅度增加. 这个函数我们称为计数函数. 在 $t\to\infty$ 的时候有 $N_t\to\infty$.

定义 7.12. 随机变量族 $(N_t)_{t\geq 0}$ 被称为参数 λ 的 Poisson 过程.

泊松过程经常用于应用概率模型中, 例如在排队论中, N_t 表示在时间 t 之前到达服务器的客户数量. 选择指数分布来模拟两个连续到达的客户之间的时间段与指数分布缺乏记忆的特性有关. 粗略地说, 该属性表示, 对于任何给定时间 $t\geq 0$, t 与客户下一次到达之间的时间始终具有相同的分布, 与时间 t 之前发生的情况无关.

命题 7.13. 对于每个 $n\geq 1$, T_n 服从 Gamma 分布 $\Gamma(n,\lambda)$, 密度为

$$p(x)=\frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x).$$

对于每个 $t>0$, N_t 服从参数 λt 的 Poisson 分布:

$$\mathbb{P}(N_t=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad \forall k\in\mathbb{N}.$$

Proof. 注意到参数 λ 的指数分布就是 Gamma 分布 $\Gamma(1,\lambda)$. 我们首先证明若 X 服从分布 $\Gamma(a,\lambda)$, Y 服从分布 $\Gamma(b,\lambda)$, 且 X,Y 独立, 那么 $X+Y$ 服从分布 $\Gamma(a+b,\lambda)$. 那么

$$\begin{aligned}
\mathbb{E}[\varphi(X+Y)] &= \int_{\mathbb{R}^2}\varphi(x+y)p_a(x)p_b(y)\,\mathrm{d}x\,\mathrm{d}y\\
&= \int_{\mathbb{R}}\varphi(z)\left(\int_{\mathbb{R}}p_a(x)p_b(z-x)\,\mathrm{d}x\right)\mathrm{d}z\\
&= \int_0^\infty\varphi(z)\left(\int_0^z\frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)}x^{a-1}(z-x)^{b-1}e^{-\lambda z}\,\mathrm{d}x\right)\mathrm{d}z\\
&= \int_0^\infty\frac{\lambda^{a+b}e^{-\lambda z}z^{a+b-1}}{\Gamma(a)\Gamma(b)}\varphi(z)\left(\int_0^1x^{a-1}(1-x)^{b-1}\,\mathrm{d}x\right)\mathrm{d}z\\
&= \int_0^\infty\varphi(z)\frac{\lambda^{a+b}}{\Gamma(a+b)}z^{a+b-1}e^{-\lambda z}\,\mathrm{d}z,
\end{aligned}$$

这就表明 $X+Y$ 服从分布 $\Gamma(a+b,\lambda)$. 由于 $T_n=U_1+\cdots+U_n$, 所以 T_n 服从分布 $\Gamma(n,\lambda)$.

对于 $k\geq 1$, 有

$$\begin{aligned}
\mathbb{P}(N_t=k) &= \mathbb{P}(T_k\leq t<T_{k+1})\\
&= \mathbb{P}(T_k\leq t)-\mathbb{P}(T_{k+1}\leq t)\\
&= \int_0^t\frac{\lambda^k}{(n-1)!}x^{k-1}e^{-\lambda x}\,\mathrm{d}x-\int_0^t\frac{\lambda^{k+1}}{k!}x^ke^{-\lambda x}\,\mathrm{d}x
\end{aligned}$$

$$= \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

对于 $k = 0$ 的时候，有 $\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$. 这就表明 N_t 服从参数 λt 的 Poisson 分布. □

我们现在将陈述有关 Poisson 过程的第一个重要结果. 我们需要引入给定事件的条件概率的概念（更多关于条件的内容将在 chapter 9 中找到）. 如果 $B \in \mathcal{A}$ 使得 $\mathbb{P}(B) > 0$, 我们定义 $(\Omega, \mathcal{A})$ 上的一个新的概率测度: 已知 B 的条件概率, 记为 $\mathbb{P}(\cdot \mid B)$. 对于每个 $A \in \mathcal{A}$, 其满足

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

对于每个非负随机变量 X, X 在 $\mathbb{P}(\cdot \mid B)$ 下的期望记为 $\mathbb{E}[X|B]$, 容易看出

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \int_A \frac{\mathbf{1}_B}{\mathbb{P}(B)} \,\mathrm{d}\mathbb{P},$$

所以 $\mathbb{P}(\cdot \mid B)$ 相对于 $\mathbb{P}$ 有密度 $\mathbf{1}_B / \mathbb{P}(B)$, 故

$$\mathbb{E}[X|B] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega|B) = \int_\Omega X(\omega)\frac{\mathbf{1}_B(\omega)}{\mathbb{P}(B)}\mathbb{P}(\mathrm{d}\omega) = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}.$$

命题 7.14. 令 $t > 0$，$n \in \mathbb{N}$. 在条件概率 $\mathbb{P}(\cdot \mid N_t = n)$ 下，随机变量 $(T_1, \ldots, T_n)$ 有密度

$$\frac{n!}{t^n}\mathbf{1}_{\{0<s_1<s_2<\cdots<s_n<t\}}.$$

此外，在条件概率 $\mathbb{P}(\cdot \mid N_t = n)$ 下，随机变量 $T_{n+1} - t$ 服从参数 λ 的指数分布并且独立于 $(T_1, \ldots, T_n)$.

现在我们陈述关于 Poisson 过程的一个非常重要的定理.

定理 7.15. 令 $t > 0$，对于每个 $r \geq 0$，令

$$N_r^{(t)} = N_{t+r} - N_t.$$

随机变量族 $(N_r^{(t)})_{r\geq 0}$ 仍然是参数 λ 的 Poisson 过程，并且与 $(N_r)_{0\leq r\leq t}$ 独立.

注释 7.16 (直观解释). 如果我们将 Poisson 过程的跳跃时间解释为客户到达服务器的时间，则该定理意味着如果有一个在时间 $t > 0$ 到达的观察者，其记录 t 之后客户的到达时间，看到（在分布的意义下）的情况与他在时间 0 到达的时候一样，并且了解时间 0 和 t 之间客户的到达时间不会给他提供有关时间 t 之后客户到达情况的信息. 这可以被视为所谓 “Markov 性质” 的一个方面.

推论 7.17. 令 $t_0 = 0 \leq t_1 \leq \cdots \leq t_k$，随机变量 $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_k} - N_{t_{k-1}}$ 是独立的，并且，对于每个 $1 \leq j \leq k$，$N_{t_j} - N_{t_{j-1}}$ 服从参数 $\lambda(t_j - t_{j-1})$ 的 Poisson 分布.

8

随机变量的收敛

9

条件

9.1 离散条件

本章中考虑概率空间 $(\Omega, \mathcal{A}, \mathbb{P})$. 我们已经在 section 7.5 中提到了, 如果 $B \in \mathcal{A}$ 是一个概率为正的事件, 我们可以定义 $(\Omega, \mathcal{A})$ 上的一个新的概率测度: 对于 $A \in \mathcal{A}$, 定义

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

概率测度 $A \mapsto \mathbb{P}(A|B)$ 被称为给定 B 下的条件概率. 类似地, 对于每个非负随机变量 X, 或者 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, 定义给定 B 下的 X 的**条件期望**为

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}.$$

现在我们定义已知一个离散随机变量下的条件期望. 考虑一个离散随机变量 Y, 取值在可数空间 E 中（σ-域为幂集）. 令 $E' = \{y \in E \mid \mathbb{P}(Y = y) > 0\}$. 如果 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, 那么对于每个 $y \in E'$, 有

$$\mathbb{E}[X|Y = y] = \frac{\mathbb{E}[X\mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}.$$

定义 9.1. 令 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，定义已知 Y 下的 X 的条件期望为一个**随机变量**

$$\mathbb{E}[X|Y] = \varphi(Y),$$

其中 $\varphi : E \to \mathbb{R}$ 为

$$\varphi(y) = \begin{cases} \mathbb{E}[X|Y = y] & y \in E', \\ 0 & y \in E \setminus E'. \end{cases}$$

注释 9.2. 定义中 $y \in E \setminus E'$ 时 $\varphi(y)$ 的取值是无关紧要的，因为我们可以证明这样的情况仅仅构成一个零测集：

$$\mathbb{P}(Y \in E \setminus E') = \sum_{y \in E \setminus E'} \mathbb{P}(Y = y) = 0.$$

与相对于一个事件的条件期望相比, 需要注意 $\mathbb{E}[X|Y]$ 是一个 $\Omega \to \mathbb{R}$ 的**随机变量**, 其几乎肯定满足

$$\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = Y(\omega)].$$

命题 9.3. 令 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，我们有 $\mathbb{E}\big[|\mathbb{E}[X|Y]|\big] \le \mathbb{E}[|X|]$，这表明 $\mathbb{E}[X|Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. 此外，对于任意有界的 $\sigma(Y)$-可测的实值随机变量 Z，有

$$\mathbb{E}[ZX] = \mathbb{E}\big[Z\mathbb{E}[X|Y]\big].$$

9.2 条件期望的定义

9.2.1 可积随机变量

下面的定理提供了关于一个子 σ-域的可积随机变量的条件期望的定义.

定理 9.4. 令 $\mathcal{B}$ 是 $\mathcal{A}$ 的一个子 σ-域，$X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. 那么 $L^1(\Omega, \mathcal{A}, \mathbb{P})$ 中存在唯一的随机变量 $\mathbb{E}[X|\mathcal{B}]$，使得

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X\mathbf{1}_B] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]\mathbf{1}_B\big].$$

更一般地，对于每个有界的 $\mathcal{B}$-可测的实值随机变量 Z，有

$$\mathbb{E}[XZ] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]Z\big].$$

如果 $X \ge 0$，那么我们几乎肯定有 $\mathbb{E}[X|\mathcal{B}] \ge 0$.

特别地, 如果 $\mathcal{B} = \sigma(Y)$ 是随机变量 Y 生成的 σ-域, 我们在写法上不区分

$$\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X|\sigma(Y)] = \mathbb{E}[X|Y].$$

这与前文离散情况下的定义是不冲突的.

命题 9.5 (条件期望的性质).

1. 如果 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ 并且 X 是 $\mathcal{B}$-可测的，那么 $\mathbb{E}[X|\mathcal{B}] = X$.
2. $L^1(\Omega, \mathcal{A}, \mathbb{P})$ 上的映射 $X \mapsto \mathbb{E}[X|\mathcal{B}]$ 是线性映射.
3. 如果 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，那么 $\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]\big] = \mathbb{E}[X]$.
4. 如果 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，那么 $|\mathbb{E}[X|\mathcal{B}]| \le \mathbb{E}\big[|X| \bigm| \mathcal{B}\big]$ a.s.，因此 $\mathbb{E}\big[|\mathbb{E}[X|\mathcal{B}]|\big] \le \mathbb{E}[|X|]$. 因此映射 $X \mapsto \mathbb{E}[X|\mathcal{B}]$ 是 $L^1(\Omega, \mathcal{A}, \mathbb{P})$ 上的压缩映射.
5. 如果 $X, X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ 并且 $X \ge X'$，那么 $\mathbb{E}[X|\mathcal{B}] \ge \mathbb{E}[X'|\mathcal{B}]$ a.s..

9.3 条件期望的具体性质

命题 9.6 (嵌套 σ-域). 令 $\mathcal{B}_1, \mathcal{B}_2$ 是 $\mathcal{A}$ 的两个子 σ-域且 $\mathcal{B}_1 \subseteq \mathcal{B}_2$. 那么对于任意非负（可积）随机变量 X，我们有

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}_2] \bigm| \mathcal{B}_1\big] = \mathbb{E}[X|\mathcal{B}_1].$$

9.4 条件期望的计算

9.4.1 离散条件

如果 Y 是值在可数空间 E 中的随机变量, 令 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. 那么我们有

$$\mathbb{E}[X|Y] = \varphi(Y),$$

其中 $y \in E$ 且 $\mathbb{P}(Y = y) > 0$ 的时候有

$$\varphi(y) = \frac{\mathbb{E}[X\mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}.$$

9.4.2 带有密度的随机变量

令 X, Y 分别是值在 $\mathbb{R}^m$ 和 $\mathbb{R}^n$ 中的随机变量, 假设 (X, Y) 有相对于 Lebesgue 测度的密度, 记为 $p(x, y)$. 那么对于任意 Borel 可测函数 $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}_+$, 有

$$\mathbb{E}[f(X, Y)] = \int_{\mathbb{R}^m \times \mathbb{R}^n} f(x, y)p(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

注意此时 Y 有密度

$$q(y) = \int_{\mathbb{R}^m} p(x, y)\,\mathrm{d}x.$$

令 $h : \mathbb{R}^m \to \mathbb{R}_+$ 是可测函数, 我们计算 $\mathbb{E}[h(X)|Y]$. 对于任意可测函数 $g : \mathbb{R}^n \to \mathbb{R}_+$, 我们有

$$\begin{aligned}
\mathbb{E}[h(X)g(Y)] &= \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x)g(y)p(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} h(x)p(x, y)\,\mathrm{d}x \right) g(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} h(x)p(x, y)\,\mathrm{d}x \right) g(y)\mathbf{1}_{\{q(y)>0\}}\,\mathrm{d}y,
\end{aligned}$$

最后一个等式是因为若 y 使得 $q(y) = 0$, 那么对于 x 而言几乎处处有 $p(x, y) = 0$, 所以 $\int h(x)p(x, y)\,\mathrm{d}x = 0$. 于是

$$\begin{aligned}
\mathbb{E}[h(X)g(Y)] &= \int_{\mathbb{R}^n} \frac{\int_{\mathbb{R}^m} h(x)p(x, y)\,\mathrm{d}x}{q(y)} g(y)q(y)\mathbf{1}_{\{q(y)>0\}}\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \varphi(y)g(y)q(y)\mathbf{1}_{\{q(y)>0\}}\,\mathrm{d}y \\
&= \mathbb{E}[\varphi(Y)g(Y)],
\end{aligned}$$

其中

$$\varphi(y) = \begin{cases} \dfrac{1}{q(y)} \displaystyle\int_{\mathbb{R}^m} h(x)p(x, y)\,\mathrm{d}x & q(y) > 0, \\ h(0) & q(y) = 0. \end{cases}$$

由于 g 的任意性, 根据 命题 6.11, 这就表明对于每个有界的 $\sigma^{-1}(Y)$-可测的实值随机变量 Z 有

$$\mathbb{E}[h(X)Z] = \mathbb{E}[\varphi(Y)Z],$$

即

$$\mathbb{E}[h(X)|Y] = \varphi(Y).$$

故我们得到了下面的命题.

命题 9.7. 对于每个 $y \in \mathbb{R}^n$，令 $\nu(y, \mathrm{d}x)$ 为 $\mathbb{R}^m$ 上的概率测度，定义为

$$\nu(y, \mathrm{d}x) = \begin{cases} \dfrac{1}{q(y)} p(x, y)\,\mathrm{d}x & q(y) > 0, \\ \delta_0(\mathrm{d}x) & q(y) = 0, \end{cases}$$

那么对于任意可测函数 $h : \mathbb{R}^m \to \mathbb{R}_+$，我们有

$$\mathbb{E}[h(X)|Y] = \int h(x)\nu(Y, \mathrm{d}x).$$

对于使得 $q(y) > 0$ 的 y, 我们有

$$\mathbb{E}[h(X)|Y = y] = \int h(x)\nu(y, \mathrm{d}x) = \frac{1}{q(y)} \int h(x) p(x, y)\,\mathrm{d}x,$$

我们通常说

$$x \mapsto \frac{p(x, y)}{q(y)}$$

是已知 $Y = y$ 下 X 的条件密度函数.

9.5 转移概率和条件分布

定义 9.8. 令 $(E, \mathcal{E})$ 和 $(F, \mathcal{F})$ 是两个可测空间，E 到 F 的一个**转移概率**指的是映射

$$\nu : E \times \mathcal{F} \to [0, 1],$$

其满足:

1. 对于每个 $x \in E$，$A \mapsto \nu(x, A)$ 是 $(F, \mathcal{F})$ 上的概率测度.
2. 对于每个 $A \in \mathcal{F}$，$x \mapsto \nu(x, A)$ 是 $\mathcal{E}$-可测的.

从直观上, 每固定一个“起点”$x \in E$, 概率测度 $\nu(x, \cdot)$ 给出了一种随机的选择一个“到达点”$y \in F$ 的方法, 这个概念将在 Markov 链中发挥重要作用.

定义 9.9. 令 X 和 Y 分别是值在 $(E, \mathcal{E})$ 和 $(F, \mathcal{F})$ 中的随机变量. 如果 E 到 F 的转移概率 ν 使得: 对于任意 F 上的非负可测函数 h，有

$$\mathbb{E}[h(Y)|X] = \int \nu(X, \mathrm{d}y)h(y), \quad \text{a.s.},$$

那么我们说 ν 是已知 X 下 Y 的条件分布.

根据定义, 如果 ν 是已知 X 下 Y 的条件分布, 那么对于每个 $A \in \mathcal{F}$, 有

$$\mathbb{P}(Y \in A|X) = \mathbb{E}[\mathbf{1}_{\{Y \in A\}}|X] = \nu(X, A), \quad \text{a.s.}$$

也可以写为对于任意 $x \in E$, 有

$$\mathbb{P}(Y \in A|X = x) = \nu(x, A).$$

Part III

随机过程

10

Markov 链

10.1 定义和首要性质

在本章中, E 表示一个有限或者可数空间, 配备 σ-域 $\mathcal{P}(E)$. E 上的随机矩阵指的是一族实数 $(Q(x,y))_{(x,y)\in E\times E}$, 其满足:

1. 对于任意 $x,y\in E$, $0\le Q(x,y)\le 1$;
2. 对于每个 $x\in E$, $\sum_{y\in E} Q(x,y)=1$.

这个概念等价于从 E 到 E 的转移概率. 实际上, 如果我们令

$$\nu(x,A)=\sum_{y\in A} Q(x,y),\quad x\in E, A\subseteq E,$$

那么 ν 就是一个从 E 到 E 的转移概率. 反之, 给定一个转移概率 ν, 定义 $Q(x,y)=\nu(x,\{y\})$, 这就给出了 E 上的一个随机矩阵.

对于每个 $n\ge 1$, 我们定义 $Q_n=Q^n$ 为通常的矩阵乘法: 定义 $Q_1=Q$, 然后递归定义

$$Q_{n+1}(x,y)=\sum_{z\in E} Q_n(x,z)Q(z,y).$$

不难验证 Q_n 仍然是 E 上的一个随机矩阵. 我们同时定义 $Q_0(x,y)=\mathbf{1}_{\{x=y\}}$ (类似单位阵). 那么对于任意 $m,n\ge 0$, 我们有 $Q_{m+n}=Q_mQ_n$:

$$Q_{m+n}(x,y)=\sum_{z\in E} Q_m(x,z)Q_n(z,y).$$

对于每个函数 $f:E\to\mathbb{R}_+$, 用 Qf 表示 E 上的函数, 满足

$$Qf(x)=\sum_{y\in E} Q(x,y)f(y),$$

取值在 $[0,\infty]$ 中. 如果 ν 是 E 上的测度, 我们同时定义

$$\nu Q(y)=\sum_{x\in E}\nu(x)Q(x,y),\quad y\in E.$$

定义 10.1. 令 Q 是 E 上的随机矩阵, $(X_n)_{n\in\mathbb{Z}_+}$ 是定义在 $(\Omega,\mathcal{A},\mathbb{P})$ 上值在 E 中的随机过程. 如果对于每个 $n\ge 0$, 已知 $(X_0,X_1,\dots,X_n)$ 下的 X_{n+1} 的条件分布是

$Q(X_n,\cdot)$，那么我们说 $(X_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 是 Markov 链．这等价于说对于每个 $x_0,\ldots,x_n,y\in E$ 使得 $\mathbb{P}(X_0=x_0,\ldots,X_n=x_n)>0$，都有

$$\mathbb{P}(X_{n+1}=y|X_0=x_0,\ldots,X_n=x_n)=Q(x_n,y).$$

已知 $(X_0,X_1,\ldots,X_n)$ 下的 X_{n+1} 的条件分布是 $Q(X_n,\cdot)$ 还意味着对于每个 $y\in E$, 有

$$\mathbb{P}(X_{n+1}=y|X_0,\ldots,X_n)=Q(X_n,y).$$

注意这是随机变量的等式.

利用 命题 9.6, 对于 $\{0,1,\ldots,n-1\}$ 的任意子集 $\{i_1,\ldots,i_k\}$, 我们有

$$\begin{aligned}&\mathbb{P}(X_{n+1}=y|X_{i_1},\ldots,X_{i_k},X_n)\\&=\mathbb{E}\big[\mathbb{P}(X_{n+1}=y|X_0,X_1,\ldots,X_n)\;\big|\;X_{i_1},\ldots,X_{i_k},X_n\big]\\&=\mathbb{E}[Q(X_n,y)|X_{i_1},\ldots,X_{i_k},X_n]\\&=Q(X_n,y)\end{aligned}$$

特别地, 有

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

这等价于, 对于每个使得 $\mathbb{P}(X_n=x)>0$ 的 $x\in E$, 有

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

这个式子提供了一种直观理解, 即如果 Markov 链在 n 时间处于点 $x\in E$ 处, 那么可以根据概率测度 $Q(x,\cdot)$ 衡量其在 $n+1$ 时间到达点的位置.

注释 10.2.

1. 对于一般的随机过程 $(X_n)_{n\in\mathbb{Z}_+}$，已知 $(X_0,X_1,\ldots,X_n)$ 下 X_{n+1} 的条件分布可以写为 $\nu((X_0,X_1,\ldots,X_n),\cdot)$ 的形式，其中 ν 表示从 E^{n+1} 到 E 的转移概率，此时条件分布依赖于所有 $X_0,X_1,\ldots,X_n$ 而不仅仅是 X_n．而 Markov 链的定义要求已知 $(X_0,X_1,\ldots,X_n)$ 下 X_{n+1} 的条件分布仅仅依赖于 X_n，这被称为 **Markov 属性**：为了预测 X_{n+1}，过去的所有 $(X_0,X_1,\ldots,X_n)$ 的知识并不能提供比 X_n 更多的信息．
2. 函数 $Q(x,\cdot)$ 给出了在不依赖 n 的情况下，已知 $X_n=x$ 下 X_{n+1} 的条件分布．这就是“转移机制”的**时齐性**．人们还可以考虑非齐次 Markov 链，其中时间 n 和 $n+1$ 之间的转移机制取决于 n，但在本课程中我们仅考虑齐次 Markov 链．

命题 10.3. 一个值在 E 中的随机过程 $(X_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 的 Markov 链当且仅当对于每个 $n\geq 0$ 和 $x_0,x_1,\ldots,x_n\in E$，我们有

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

Proof. 若 $(X_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 的 Markov 链, 那么依据

$$\begin{aligned}&\mathbb{P}(X_0=x_0,X_1=x_1,\ldots,X_n=x_n,X_{n+1}=x_{n+1})\\&=\mathbb{P}(X_0=x_0,\ldots,X_n=x_n)\mathbb{P}(X_{n+1}=x_{n+1}|X_0=x_0,\ldots,X_n=x_n)\end{aligned}$$

进行归纳即可.

反之, 假设上式成立, 那么

$$
\begin{aligned}
&\mathbb{P}(X_{n+1}=y|X_0=x_0,\ldots,X_n=x_n)\\
&=\frac{\mathbb{P}(X_0=x_0,\ldots,X_n=x_n,X_{n+1}=y)}{\mathbb{P}(X_0=x_0,\ldots,X_n=x_n)}=Q(x_n,y),
\end{aligned}
$$

即 $(X_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 的 Markov 链. □

注释 10.4. 这表明对于 Markov 链 $(X_n)_{n\in\mathbb{Z}_+}$，$(X_0,X_1,\ldots,X_n)$ 的分布完全由初始分布和转移矩阵 Q 确定.

命题 10.5. $(X_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 的 Markov 链.

1. 对于每个 $n\geq 0$ 和可测函数 $f:E\to\mathbb{R}_+$，

   $$\mathbb{E}[f(X_{n+1})|X_0,X_1,\ldots,X_n]=\mathbb{E}[f(X_{n+1})|X_n]=Qf(X_n).$$

   更一般地，对于 $\{0,1,\ldots,n-1\}$ 的任意子集 $\{i_1,\ldots,i_k\}$，我们有

   $$\mathbb{E}[f(X_{n+1})|X_{i_1},\ldots,X_{i_k},X_n]=Qf(X_n).$$

2. 对于每个 $n\geq 0$，$p\geq 1$ 以及 $y_1,\ldots,y_p\in E$，有

   $$
   \begin{aligned}
   &\mathbb{P}(X_{n+1}=y_1,\ldots,X_{n+p}=y_p|X_0,\ldots,X_n)\\
   &=Q(X_n,y_1)Q(y_1,y_2)\cdots Q(y_{p-1},y_p).
   \end{aligned}
   $$

   因此，有

   $$\mathbb{P}(X_{n+p}=y_p|X_n)=\mathbb{P}(X_{n+p}=y_p|X_0,\ldots,X_n)=Q_p(X_n,y_p).$$

   如果我们固定 $n\geq 0$，对于每个 $p\in\mathbb{Z}_+$，令 $Y_p=X_{n+p}$，那么随机过程 $(Y_p)_{p\in\mathbb{Z}_+}$ 仍然是带有转移矩阵 Q 的 Markov 链.

Proof. (1) 注意到 $f(X_{n+1})=\sum_{y\in E}f(y)\mathbf{1}_{\{X_{n+1}=y\}}$, 那么

$$
\begin{aligned}
\mathbb{E}[f(X_{n+1})|X_0,X_1,\ldots,X_n]&=\sum_{y\in E}f(y)\mathbb{P}(X_{n+1}=y|X_0,X_1,\ldots,X_n)\\
&=\sum_{y\in E}f(y)Q(X_n,y)\\
&=Qf(X_n).
\end{aligned}
$$

对于 $\{0,1,\ldots,n-1\}$ 的任意子集 $\{i_1,\ldots,i_k\}$, 我们有

$$
\begin{aligned}
&\mathbb{E}[f(X_{n+1})|X_{i_1},\ldots,X_{i_k},X_n]\\
&=\mathbb{E}\big[\mathbb{E}[f(X_{n+1})|X_0,X_1,\ldots,X_n]\bigm| X_{i_1},\ldots,X_{i_k},X_n\big]\\
&=\mathbb{E}[Qf(X_n)|X_{i_1},\ldots,X_{i_k},X_n]\\
&=Qf(X_n).
\end{aligned}
$$

(2) 根据命题 10.3, 有

$$\begin{aligned}&\mathbb{P}(X_{n+1}=y_1,\dots,X_{n+p}=y_p|X_0=x_0,\dots,X_n=x_n)\\&=\frac{\mathbb{P}(X_0=x_0)Q(x_0,x_1)\cdots Q(x_n,y_1)\cdots Q(y_{p-1},y_p)}{\mathbb{P}(X_0=x_0)Q(x_0,x_1)\cdots Q(x_{n-1},x_n)}\\&=Q(x_n,y_1)Q(y_1,y_2)\cdots Q(y_{p-1},y_p),\end{aligned}$$

即

$$\begin{aligned}&\mathbb{P}(X_{n+1}=y_1,\dots,X_{n+p}=y_p|X_0,\dots,X_n)\\&=Q(X_n,y_1)Q(y_1,y_2)\cdots Q(y_{p-1},y_p).\end{aligned}$$

对于 $p=2$ 的情况, 有

$$\begin{aligned}\mathbb{P}(X_{n+2}=y_2|X_0,\dots,X_n)&=\sum_{y_1\in E}\mathbb{P}(X_{n+1}=y_1,X_{n+2}=y_2|X_0,\dots,X_n)\\&=\sum_{y_1\in E}Q(X_n,y_1)Q(y_1,y_2)\\&=Q_2(X_n,y_2),\end{aligned}$$

一般的情况由此归纳即可. 于是

$$\begin{aligned}\mathbb{P}(X_{n+p}=y_p|X_n)&=\mathbb{E}\big[\mathbb{P}(X_{n+p}=y_p|X_0,\dots,X_n)\bigm| X_n\big]\\&=\mathbb{E}[Q_p(X_n,y_p)|X_n]=Q_p(X_n,y_p).\end{aligned}$$

最后我们验证 $(Y_p)_{p\in\mathbb{Z}_+}$ 是带有转移矩阵 Q 的 Markov 链. 根据定义, 有

$$\begin{aligned}&\mathbb{P}(Y_{p+1}=y_{p+1}|Y_0,\dots,Y_p)\\&=\mathbb{P}(X_{n+p+1}=y_{p+1}|X_n,\dots,X_{n+p})\\&=\mathbb{E}\big[\mathbb{P}(X_{n+p+1}=y_{p+1}|X_0,\dots,X_{n+p})\bigm| X_n,\dots,X_{n+p}\big]\\&=\mathbb{E}[Q(X_{n+p},y_{p+1})|X_n,\dots,X_{n+p}]\\&=Q(X_{n+p},y_{p+1})=Q(Y_p,y_{p+1}),\end{aligned}$$

这就表明 $(Y_p)_{p\in\mathbb{Z}_+}$ 是带有转移矩阵 Q 的 Markov 链. □

10.2 例子

10.2.1 $\mathbb{Z}^d$ 上的随机游走

10.3 典范的 Markov 链

在本节中, 我们解释如何对概率空间 (和随机过程) 进行典范的选择, 以获得具有给定转移矩阵的 Markov 链. 这与前面的章节形成对比, 前面的章节没有指定潜在的概率空间. 在 Markov 链的背景下, 这种典范的选择将有几个优点. 特别是, 它将使我们能够同时考虑给定的 Markov 链和所有可能的初始点.

命题 10.6. 令 Q 是 E 上的随机矩阵，我们可以找到一个概率空间 $(\Omega, \mathcal{A}, \mathbb{P})$ 使得对于每个 $x \in E$，都可以构建一个带有转移矩阵 Q 的 Markov 链 $(X_n^x)_{n\in\mathbb{Z}_+}$ 且 $X_0^x = x$.

下面我们解释这个概率空间的选择，我们令

$$\mathbf{\Omega} = E^{\mathbb{Z}_+},$$

即 $\mathbf{\Omega}$ 的元素形如 $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$. 定义坐标映射 $\boldsymbol{X}_n$ 为 $\boldsymbol{X}_n(\omega) = \omega_n$. 我们给 $\mathbf{\Omega}$ 配备使得所有坐标映射 $\boldsymbol{X}_n$ 都可测的最小的 σ-域，记为 $\mathfrak{F}$，等价地说，$\mathfrak{F}$ 由所有的柱集生成:

$$\{x_0\} \times \cdots \times \{x_n\} \times E \times E \times \cdots,$$

其中 $n \in \mathbb{Z}_+, x_0, \ldots, x_n \in E$.

引理 10.7. 令 $(G, \mathcal{G})$ 是可测空间，映射 $\psi : G \to \mathbf{\Omega}$，那么 ψ 可测当且仅当 $\boldsymbol{X}_n \circ \psi$ 可测.

定理 10.8. 令 Q 是 E 上的随机矩阵，对于每个 $x \in E$，存在唯一的 $(\mathbf{\Omega}, \mathfrak{F})$ 上的概率测度 $\mathbf{P}_x$ 使得坐标过程 $(\boldsymbol{X}_n)_{n\in\mathbb{Z}_+}$ 是带有转移矩阵 Q 的 Markov 链并且 $\mathbf{P}_x(\boldsymbol{X}_0 = x) = 1$.

注释 10.9. 我们将说在概率空间 $(\mathbf{\Omega}, \mathfrak{F}, \mathbf{P}_x)$ 上的坐标过程 $(\boldsymbol{X}_n)_{n\in\mathbb{Z}_+}$ 是具有转移矩阵 Q 和初值 x 的**典范 Markov 链**.

注释 10.10. 令 $(Y_n)_{n\in\mathbb{Z}_+}$ 是概率测度 $\mathbb{P}$ 下具有转移矩阵 Q 的 Markov 链且 $Y_0 = x$. 那么对于 $\mathbf{\Omega} = E^{\mathbb{Z}_+}$ 的任意可测子集 B，有

$$\mathbb{P}\big((Y_n)_{n\in\mathbb{Z}_+} \in B\big) = \mathbf{P}_x(B).$$

事实上，当 B 是柱集 (cylinder set) 时，这个等式成立，然后依据单调类给出一般情况. 这个等式表明，我们后面为典范 Markov 链建立的所有结果都可以转移到具有相同转移矩阵的任何 Markov 链上 (无论它是在哪个概率空间上定义的).

10.4 状态分类

从现在开始, 除非另有说明, 否则我们将考虑上一节中针对 E 上给定随机矩阵 Q 引入的典范 Markov 链. 如 注释 10.10 所述, 我们可以将得出的所有结果都转移到在任何概率空间上定义的具有相同转移矩阵的 Markov 链.

回顾记号

$$H_x = \inf\{n \geq 1 | \boldsymbol{X}_n = x\}, N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{\boldsymbol{X}_n = x\}}.$$

在 Markov 链中, H_x 可以解释为状态第一次经过点 x 时需要的步数, N_x 可以解释为状态经过点 x 的次数.

命题 10.11. 令 $x \in E$，那么

- 要么 $\mathbf{P}_x(H_x < \infty) = 1$，此时

$$N_x = \infty, \ \mathbf{P}_x \text{ a.s.}$$

这种情况下 x 被称为**常返的**.

- 要么 $\mathbf{P}_x(H_x < \infty) < 1$，此时

$$N_x < \infty,\ \mathbf{P}_x \text{ a.s.}$$

更准确地，对于每个 $k \geq 1$ 有 $\mathbf{P}_x(N_x = k) = \mathbf{P}_x(H_x = \infty)\mathbf{P}_x(H_x < \infty)^{k-1}$，并且 $\mathbf{E}_x[N_x] = 1/\mathbf{P}_x(H_x = \infty) < \infty$，在这种情况下 x 被称为是**暂留的**.

注释 10.12. 直观上来说，$\mathbf{P}_x(H_x < \infty) = 1$ 的情况表明如果状态几乎肯定能够经过点 x，那么其几乎肯定能够无数次经过点 x. $\mathbf{P}_x(H_x < \infty) < 1$ 的情况表明如果状态不一定能经过点 x，那么其经过点 x 的次数几乎肯定是有限次，且期望也是有限次.

定义 10.13. Markov 链的潜在核定义为 $E \times E$ 上的函数

$$U(x,y) = \mathbf{E}_x[N_y].$$

注意到 $U(x,y) > 0$ 当且仅当从 x 出发的链能够到达 y 的概率是正的.

命题 10.14.

1. 对于每个 $x, y \in E$，有
$$U(x,y) = \sum_{n=0}^{\infty} Q_n(x,y).$$
2. $U(x,x) = \infty$ 当且仅当 x 是常返的.
3. 对于每个 $x, y \in E$ 且 $x \neq y$，有
$$U(x,y) = \mathbf{P}_x(H_y < \infty)U(y,y).$$

因此，如果 y 是暂留的，那么对于每个 $x \in E$，我们有 $\mathbf{P}_x(N_y < \infty) = 1$.

Proof. (1) 直接计算可知

$$U(x,y) = \mathbf{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = y\}}\right] = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y) = \sum_{n=0}^{\infty} Q_n(x,y).$$

(2) 根据 U 的定义, 我们有 $U(x,x) = \infty$ 当且仅当 $\mathbf{P}_x(N_x = \infty) = 1$, 即 x 是常返的.

(3) □

注释 10.15. 这个命题告诉我们

$$U(x,y) = \sum_{n=0}^{\infty} Q_n(x,y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_{m+n} = y | X_m = x),$$

所以 $U(x,y) > 0$ 当且仅当存在某个 n 使得 $\mathbf{P}_x(X_{m+n} = y | X_m = x) > 0$. 有些教材会把 $\mathbf{P}_x(X_{m+n} = y | X_m = x)$ 记为 $p_{xy}^{(n)}$，称作 n **步转移概率**. 若 $U(x,y) > 0$（等价地说，存在 $n \geq 0$ 使得 $p_{xy}^{(n)} > 0$），那么我们说 x **可达** y，记为 $x \to y$. 如果同时有 $y \to x$，那么我们说 x 和 y **互通**，记为 $x \leftrightarrow y$. 不难验证互通是一个等价关系.

我们记所有常返点的集合为 R.

引理 10.16. 假设 $x \in R$ 且 $y \in E \setminus \{x\}$ 使得 $U(x,y) > 0$ $(x \to y)$，那么 $y \in R$ 且 $\mathbf{P}_y(H_x < \infty) = 1$，因此 $U(y,x) > 0$ $(y \to x)$.