Contents

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 ($\mathbb{R}$ 上的 Borel σ-域)**.** 记 $\mathcal{C}_1$ 为 $\mathbb{R}$ 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): $\mathbb{R}$ 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 $\mathbb{R}$ 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

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

(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

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

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

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

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

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

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 1.23 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.

定义 1.20. 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 1.21. $\mathcal{P}(E)$ 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 1.22 (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 1.23. 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

$$\begin{aligned}\mu(A) &= \mu(A \cap E) = \mu\Big(\bigcup_{n\in\mathbb{N}}(A \cap E_n)\Big) = \lim_{n\to\infty}\uparrow \mu(A \cap E_n)\\ &= \lim_{n\to\infty}\uparrow \mu_n(A) = \lim_{n\to\infty}\uparrow \nu_n(A) = \lim_{n\to\infty}\uparrow \nu(A \cap E_n)\\ &= \nu\Big(\bigcup_{n\in\mathbb{N}}(A \cap E_n)\Big) = \nu(A \cap E) = \nu(A),\end{aligned}$$

这就表明 $\mu = \nu$. $\square$

推论 1.23 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 1.23 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 $\boldsymbol{f}$ **相对于 $\boldsymbol{\mu}$ 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 $\boldsymbol{f}$ **相对于** $\boldsymbol{\mu}$ **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

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

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

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

这就表明 μ 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

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

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 $f = g,\ \mu$ 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

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

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

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

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

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

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

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

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

$$\int \liminf f_n \, d\mu \le \liminf \int f_n \, d\mu.$$

Proof. 只需证明

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

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

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

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

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

令 $n \to \infty$, 即可得到

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

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

$$\int_E h(\varphi(x))\mu(dx) = \int_F h(y)\nu(dy).$$

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

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，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

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.

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.

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

Proof. (1) 显然 $\lambda^*(\emptyset) = 0$ 并且 $A \subseteq B$ 表明 $\lambda^*(A) \le \lambda^*(B)$. 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 $\big(a_i^{(n)}, b_i^{(n)}\big)$ 使得

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

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

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

由于 ε 的任意性, 所以 λ^* 满足 σ-次可加性.

(2) $\square$

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

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

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.

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

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

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

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

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

Part II

概率论

5

概率论基础

5.1 一般定义

5.1.1 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 5.1. 一些常见的概率模型.

1. 考虑扔两次骰子这一实验, 那么
$$\Omega = \{1, 2, \dots, 6\}^2, \quad \mathcal{A} = \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \frac{\mathrm{card}(A)}{36}.$$
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 $\mathbb{P}(\{\omega\}) = 1/\,\mathrm{card}(\Omega)$ 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:
$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}}.$$
Ω 上的 σ-域 $\mathcal{A}$ 被定义为包含形如
$$\{\omega \in \Omega \mid \omega_1 = i_1, \dots, \omega_n = i_n\}$$

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.

5.1.2 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 5.2. 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 5.3. 回顾 (5.1) 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 5.4. 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

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

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 5.5. 我们考虑 (5.1) 中的第二个例子, 随机变量为 $X(\omega) = \inf\{j \mid \omega_j = 6\}$. 那么

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

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 $\mathbb{P}_X(\mathrm{d}x) = p(x)\lambda(\mathrm{d}x) = p(x)\,\mathrm{d}x$.

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

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

5.1.3 数学期望

定义 5.6. 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:

- $X \ge 0$, 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 $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$, 此时我们定义 $\mathbb{E}[X] = \big(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d]\big)$. 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 5.7. 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 5.8. 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

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

命题 5.8 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 $f : E \to \mathbb{R}$ 是任意示性函数, 此时对于任意 E 的可测子集 A, 有

$$\mathbb{P}_X(A) = \int \mathbf{1}_A \, d\mathbb{P}_X = \mathbb{E}[\mathbf{1}_A(X)] = \int \mathbf{1}_A \, d\nu = \nu(A),$$

所以分布 $\mathbb{P}_X = \nu$. 下面的命题应用了这样的思想.

命题 5.9. 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

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_d) \, dx_1 \cdots dx_{j-1} \, dx_{j+1} \cdots dx_d.$$

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

$$\begin{aligned}
\mathbb{E}[f(X_j)] &= \mathbb{E}[f \circ \pi_j(X)] \\
&= \int_{\mathbb{R}^d} f(\pi_j(x)) \mathbb{P}_X(dx) \\
&= \int_{\mathbb{R}^d} f(x_j) p(x_1, \ldots, x_d) \, dx_1 \cdots dx_d \\
&= \int_{\mathbb{R}} f(x_j) \left(\int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_d) \, dx_1 \cdots dx_{j-1} \, dx_{j+1} \cdots dx_d \right) dx_j \\
&= \int_{\mathbb{R}} f(x_j) p_j(x_j) \, dx_j = \int_{\mathbb{R}} f(x_j) \mathbb{P}_{X_j}(dx_j),
\end{aligned}$$

这就表明对于任意 Borel 子集 A 有

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

即 X_j 有密度函数 p_j. □

如果 $X = (X_1, \ldots, X_d)$ 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 $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_j}$, 也不能确定 $\mathbb{P}_X$.

5.1.4 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 $\boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$ 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 $\mathcal{B}(\boldsymbol{n}, \boldsymbol{p})$ ($\boldsymbol{n} \in \mathbb{N}, \boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$)**. 如果值在 $\{0,1,\ldots,n\}$ 中的随机变量 X 满足

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

那么我们说 X 是 E 上的二项分布.

4. **参数 $\boldsymbol{p} \in (\mathbf{0}, \mathbf{1})$ 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 $\boldsymbol{\lambda}>\mathbf{0}$ 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. **$[\boldsymbol{a}, \boldsymbol{b}]$ 上的均匀分布**:

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

2. **参数 $\boldsymbol{\lambda}>\mathbf{0}$ 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

$$\mathbb{P}(X \geq a+b) = \mathbb{P}(X \geq a)\mathbb{P}(X \geq b). \tag{5.1}$$

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

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

这是指数分布的推广, $a = 1$ 时即指数分布.

4. **参数 $\boldsymbol{a > 0}$ 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 $\boldsymbol{\mathcal{N}(m, \sigma^2)}$ ($\boldsymbol{m \in \mathbb{R}, \sigma > 0}$)**:

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

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 $\mathcal{N}(am+b, a^2\sigma^2)$.

5.1.5 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

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

根据 推论 1.23, 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? 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

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

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.

5.2 随机变量的矩

5.2.1 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 $X\ge 0$ 或者 $\mathbb{E}[|X|^p]<\infty$ 的时候有定义.

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

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

- **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

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

  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

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

取 $Y=1$，我们得到 $\|X\|_1 \le \|X\|_p$. 此外，如果 $1 \le p < q \le \infty$，有 $\|X\|_p \le \|X\|_q$，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 5.10. 令 $X \in L^2(\Omega,\mathcal{A},\mathbb{P})$，定义 X 的**方差**为

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

X 的**标准差**为

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