高等统计学方法

Advanced Statistical Methods

By Xifeng Zhang • 自由未知量 a

• 二〇二四年十月二十六日 •

高等统计学方法

Xifeng Zhang · 自由未知量 a

二〇二四年十月二十六日

Bilibili：自由未知量 a

目录

第一部分

实分析基础

第一章　集合论基础

§1.1　集合

1.1.1　集合及其表示

定义 1.1.1 (**集合**) 在一定范围内的个体事物的全体，当将它们看作一个整体时，我们把这个整体称为一个集合，即一个给定的集合（或称为集）是指具有某种性质的事物的全体. 组成集合的每个事物称为该集合的元素（或简称元）. 设 A 是一个集合，a 是 A 的元素，则一般记作 $a \in A$（读作"a 属于 A"）或 $A \ni a$（读作"A 包含 a"）. 若 a 不是 A 中的元素，一般记为 $a \notin A$.

注 1.1.1 集合的表示方法有两种

1. 列举法：对于一个具体集合 A 可以通过列举其元素 $a, b, c, \ldots$ 来定义，可记为
$$A = \{a, b, c, \ldots\} \tag{1.1}$$

2. 描述法：对于一些集合，我们也可以通过描述集合中元素必须且只需满足的条件 p 来加以定义，可记为
$$A = \{x : x \text{ 满足条件 } p\} \text{或} \{x | x \text{ 满足条件 } p\} \tag{1.2}$$

注 1.1.2 有一些的集合有特殊的含义，我们对它们赋予特殊的记号

名称	含义	记号表示
全集	包含我们所研究问题中涉及的所有元素的集合	Ω（或 U）
空集	不包含任何元素的集合	$\varnothing$
自然数集	全体自然数构成的集合	$\mathbb{N}$（或 $\mathbf{N}$）
整数集	全体整数构成的集合	$\mathbb{Z}$（或 $\mathbf{Z}$）
正整数集	全体正整数构成的集合	$\mathbb{Z}_+$（或 $\mathbf{Z}_+$）
有理数集	全体有理数构成的集合	$\mathbb{Q}$（或 $\mathbf{Q}$）
实数集	全体实数构成的集合	$\mathbb{R}$（或 $\mathbf{R}$）
复数集	全体复数构成的集合	$\mathbb{C}$（或 $\mathbf{C}$）

另外，对于自然数，早期有些定义认为 0 不是自然数，后来经过国家和国际标准的规范，目前

的主流定义方法是认为自然数包括正整数和 0（即非负整数），故正整数集有时也定义为 $\mathbb{N}_+$（或 $\mathbb{N}^*$）. 为了不引起混乱，下文涉及正整数，统一记为 $\mathbb{Z}_+$.

注 1.1.3 若集合 A 的元素只含有有限个，则称 A 为有限集，否则称为无限集.

1.1.2　集合的关系

设 A、B 是两个集合，它们之间的关系有如下几种：

1. **包含（包含于）**：若对一切元素 $x \in B$，必有 $x \in A$，则称集合 B 包含于 A（记为 $B \subset A$），或 A 包含 B（记为 $A \supset B$）. 此时可以称 B 为 A 的子集.
2. **相等**：若 $A \subset B$ 且 $A \supset B$，则称集合 A 和 B 相等（记为 $A = B$）.
3. **真包含（真包含于）**：若对一切元素 $x \in B$，必有 $x \in A$，且满足 $A \neq B$，则称集合 B 真包含于 A（记为 $B \subseteq A$），或 A 真包含 B（记为 $A \supseteq B$）. 此时可以称 B 为 A 的真子集.

注 1.1.4 对于给定的集合 A，空集和全集都是它的子集，即有 $\varnothing \subset A$，$A \subset A$.

注 1.1.5 对于给定集合 A，它的全部子集构成的集合（或可称之为一个集族）称为集合 A 的幂集（Power Set），记为 $\mathcal{P}(A)$.

1.1.3　集合的基本运算

设 A、B 是两个集合，I 为一指标集（可以有限也可以无限），它们之间的运算有如下几种：

1. $A \cup B := \{x : x \in A或x \in B\}$，称为集合 A 与集合 B 的并集；
2. $\bigcup\limits_{\alpha \in I} A_\alpha := \{x : \exists \alpha \in I, x \in A_\alpha\}$，$I$ 为有限集时，称为有限并；无限可列时，称为可列并；
3. $A \cap B := \{x : x \in A且x \in B\}$，称为集合 A 与集合 B 的交集；
4. $\bigcap\limits_{\alpha \in I} A_\alpha := \{x : \forall \alpha \in I, x \in A_\alpha\}$，$I$ 为有限集时，称为有限交；无限可列时，称为可列交；
5. $A \backslash B := \{x : x \in A且x \notin B\}$，称为集合 A 与集合 B 的差集（部分资料也记作 $A - B$）；
6. $A^c := \{x : x \in \Omega \backslash A\}$，称为集合 A 的余集（补集）；

7. $A \triangle B := \{x : x \in A且x \notin B，或x \notin A且x \in B\} = (A\backslash B) \cup (B\backslash A)$，称为集合 A 与集合 B 的对称差.

集合的交、并、差运算具有如下性质：

性质 1.1.1 (**交、并运算的性质**) 集合的交、并运算具有如下性质：

1. $A \cup A = A$，$A \cap A = A$；
2. $A \cup B = B \cup A$，$A \cap B = B \cap A$；
3. $(A \cup B) \cup C = A \cup (B \cup C)$，$(A \cap B) \cap C = A \cap (B \cap C)$；
4. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$，$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$；
5. $(\cup A_i) \cap B = \cup(A_i \cap B)$，$(\cap A_i) \cup B = \cap(A_i \cup B)$.

性质 1.1.2 (**差运算的性质**) 集合的差运算具有如下性质：

1. $(A\backslash B) \cap C = (A \cap C)\backslash(B \cap C)$；
2. $(A\backslash B)\backslash C = A\ (B \cup C)$；
3. 若 $A, B \subset \Omega$，则 $A\backslash B = A \cap B^c$.

性质 1.1.3 (De Morgan **法则**) 设 Ω 是任意一个集合，$\{A_\alpha, \alpha \in I\}$ 是 Ω 的一个子集族，I 为一个指标集，则有

$$\left(\bigcup_{\alpha \in I} A_\alpha\right)^c = \bigcap_{\alpha \in I} (A_\alpha^c), \tag{1.3}$$

$$\left(\bigcap_{\alpha \in I} A_\alpha\right)^c = \bigcup_{\alpha \in I} (A_\alpha^c). \tag{1.4}$$

注 1.1.6 集合的交、并、补运算性质表明，集合的交、并运算本质上是一样的.

1.1.4 集合（序列）的极限

定义 1.1.2 (**集序列的极限**) 设 $\{A_n : n \in \mathbb{Z}_+\}$ 为一集序列.

- 称 $\overline{\lim_{n\to\infty}} A_n = \limsup_{n\to\infty} A_n = \inf \sup_{n\to\infty} A_n := \{x : x属于A_n, n \in \mathbb{Z}_+中的无穷多个\} = \{\forall n \geq 1, \exists k \geq n, 使得\xi \in A_k\}$ 为此集序列的上极限，有时简记为 A_n, i.o.（occurs infinitly often）；
- 称 $\underline{\lim_{n\to\infty}} A_n = \liminf_{n\to\infty} A_n = \sup \inf_{n\to\infty} A_n := \{x : x不属于A_n, n \in \mathbb{Z}_+中的有限多个\} = \{\exists n \geq 1, \forall k \geq n, 使得\xi \in A_k\}$ 为此集序列的下极限；

- 若 $\overline{\lim\limits_{n\to\infty}} A_n = \underline{\lim\limits_{n\to\infty}} A_n \triangleq \lim\limits_{n=\infty} A_n$，则称 $\{A_n\}_{n=1}^{\infty}$ 的极限存在，为 $\lim\limits_{n=\infty} A_n$.

注 1.1.7 由定义可得

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k, \tag{1.5}$$

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k, \tag{1.6}$$

注 1.1.8 若 $\{A_n\}$ 两两互不相交，则

$$\lim_{n=\infty} A_n = \varnothing \tag{1.7}$$

注 1.1.9 借助 A_n 的上、下极限可分别定义上限事件和下限事件.A_n 的上限事件表示 A_n 发生无穷多次，下限事件表示 A_n 只有有限次不发生. 容易得到

$$\underline{\lim_{n\to\infty}} A_n \subset \overline{\lim_{n\to\infty}} A_n. \tag{1.8}$$

定义 1.1.3 (**单调集列**) 设 $\{A_n : n \in \mathbb{Z}_+\}$ 为一集序列.

- 若满足 $A_n \subset A_{n+1}, n \in \mathbb{Z}_+$，则称 $\{A_n\}_{n=1}^{\infty}$ 是单调增的，记作 $A_n \uparrow$;
- 若满足 $A_n \supset A_{n+1}, n \in \mathbb{Z}_+$，则称 $\{A_n\}_{n=1}^{\infty}$ 是单调减的，记作 $A_n \downarrow$;
- 单调增和单调减集序列统称为单调序列.

注 1.1.10 若 $\{A_n\}$ 单调，则 $\lim\limits_{n\to\infty} A_n$ 存在，且

- 若 $\{A_n\}$ 单调增，则

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \tag{1.9}$$

- 若 $\{A_n\}$ 单调减，则

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \tag{1.10}$$

§1.2 映射、对等与势

1.2.1 映射

定义 1.2.1 (**映射**) 对一切 $x \in A$,存在唯一的 $y \in B$,使得 $f(x) = y$,则称 f 为从 A 到 B 上的映射,记为 $f : A \mapsto B$. 称 A 为 f 的定义域. 对 $F \subset A$,称 $f(F) := \{\in B : \exists x \in F, 使得f(x) = y\}$ 为 F 在 f 之下的象,而称 $f(A)$ 为 f 的值域.

注 1.2.1 映射的定义要求 A 到 B 的对应关系只能是一对一或多对一,不能是一对多.

注 1.2.2 显然成立 $f(A) \subset B$.

定义 1.2.2 (**满射、单射和双射**) 对映射 $f : A \mapsto B$:

1. 若满足 $f(A) = B$,则称 f 为满射;
2. 若满足 $f(x_1) = f(x_2) \Leftrightarrow x_1 = x_2$,则称 f 为单射;
3. 若 f 既是单射又是满射,则称其为双射,又称为一一映射.

定义 1.2.3 (**复合映射**) 设 $f : A \mapsto B$,$g : B \mapsto C$,则下式:

$$g \circ f(x) := g(f(x)), \quad x \in A \tag{1.11}$$

称为 g,f 的复合映射,记为 $g \circ f : A \mapsto C$.

注 1.2.3 映射的复合不具有交换率,即 $f \circ g$ 和 $g \circ f$ 是不同的复合映射,这一点可以从定义上看出,且在 $g \circ f$ 成立的情形下,$f \circ g$ 未必有定义.

定义 1.2.4 (**逆象**) 设 $f : A \mapsto B$,则可定义:

$$f^{-1}(\{y\}) := \{x \in A : f(x) = y\}, y \in B \tag{1.12}$$

$$f^{-1}(F) := \{x \in A : f(x) = F\}, F \subset B \tag{1.13}$$

称 $f^{-1}(F)$ 为 F 在 f 下的逆象.

定义 1.2.5 (**逆映射**) 若 f 为 A 到 B 的单射,则可定义 $f(A)(\subset B)$ 到 A 上的逆映射 f^{-1}:

$$f^{-1}(y) := x, \quad 若f(x) = y, \forall y \in f(A). \tag{1.14}$$

注 1.2.4 逆映射定义中 f 为 A 到 B 的单射的要求是合理且有必要的，因为逆映射本质上也是一种映射，所以它需要满足映射的基本要求（对应关系只能是一对一或多对一）. 若 f 不是单射，则可能同时存在不同的 x_1 和 x_2（或更多个）满足 $f(x_1) = f(x_2) = y$. 则此时逆象 $f^{-1}(\{y\})$ 可能为 x_1，也可能为 x_2，就产生了一对多的对应关系.

注 1.2.5 由上注，逆象 $f^{-1}(\{y\})$ 与逆映射 $f^{-1}(y)$ 是两个不同的概念.

性质 1.2.1 设 $A_0, A_1, A_2 \subset A$，$B_1, B_2 \subset B$，象与逆象具有如下性质：

1. 若 $A_1 \subset A_2$，则 $f(A_1) \subset f(A_2)$;
2. $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$;
3. $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$;
4. $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$;
5. $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$;
6. $f^{-1}(f(A_0)) \supset A_0$;
7. $(f^{-1}(B))^c = f^{-1}(B^c)$.

注 1.2.6 上述性质表明，f^{-1} 与集合的交、并、补运算可交换，而 f 与交运算不可交换.

性质 1.2.2 设 f 是由 Ω 到 E 的映射，Γ 为一指标集（不必可数），$B_i \subset E, i = 1, 2, \ldots$ 则下列公式成立：

1. $f^{-1}(E) = \Omega$;
2. $f^{-1}(\varnothing) = \varnothing$;
3. $(f^{-1}(B))^c = f^{-1}(B^c)$;
4. $f^{-1}\left(\bigcup\limits_{\gamma \in \Gamma} B_\gamma\right) = \bigcup\limits_{\gamma \in \Gamma} f^{-1}(B_\gamma)$;
5. $f^{-1}\left(\bigcap\limits_{\gamma \in \Gamma} B_\gamma\right) = \bigcap\limits_{\gamma \in \Gamma} f^{-1}(B_\gamma)$;
6. $f^{-1}(B_1 \backslash B_2) = f^{-1}(B_1) \backslash f^{-1}(B_2)$.

1.2.2 对等

定义 1.2.6 (**对等**) 给定非空集合 A，B，若存在 A 到 B 的一一映射，则称 A 与 B 对等，记作 $A \sim B$. 另外规定 $\varnothing \sim \varnothing$.

注 1.2.7 集合对等是集合之间的一种新的关系，注意区分它与集合的相等关系，二者不是一个概念.

注 1.2.8 对等关系满足三点性质：

1. 自反性：对任何集合 A，成立 $A \sim A$；
2. 传递性：若 $A \sim B$，$B \sim C$，则 $A \sim C$；
3. 对称性：若 $A \sim B$，则 $B \sim A$.

注 1.2.9 无限集能和它的真子集对等，而有限集不行. 这也是一种定义有限集和无限集的角度.

定理 1.2.1 (Bernstein **定理**) 给定集合 A，B，若存在 $A^* \subset A$，$B^* \subset B$，且 $A^* \sim B$，$B^* \sim A$，则 $A \sim B$.

引理 1.2.1 设 I 为以指标集，$\{A_\alpha, \alpha \in I\}$，$\{B_\alpha, \alpha \in I\}$ 是两个集类，且对 $\forall x \in I$，$A_\alpha \sim B_\alpha$，且 $A_\alpha, \alpha \in I$ 两两不交，$B_\alpha, \alpha \in I$ 两两不交，则

$$\bigcup_{\alpha \in I} A_\alpha \sim \bigcup_{\alpha \in I} B_\alpha. \tag{1.15}$$

推论 1.2.1 若 $A \sim B$，$B \sim C$，则 $A \sim B \sim C$.

例 1.2.1 取映射 $f(x) = \dfrac{x}{1-x^2}$，$x \in (-1,1)$，则 f 是 $(-1,1)$ 到 $\mathbb{R}$ 上的一一映射，从而 $(-1,1) \sim \mathbb{R}$.

例 1.2.2 取映射 $f(x) = \arctan x$，$x \in \left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right)$，则 f 是 $\left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right)$ 到 $\mathbb{R}$ 上的一一映射，从而 $\left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right) \sim \mathbb{R}$.

例 1.2.3 对 $\mathbb{R}$ 上的区间 (a,b)，取

$$f : f(x) = \frac{x-a}{b-a}, \quad x \in (a,b) \tag{1.16}$$

则 f 为 (a,b) 到 $(0,1)$ 的一一映射. 又取

$$g : g(y) = \tan\left(\pi y - \frac{\pi}{2}\right), \quad y \in (0,1) \tag{1.17}$$

则 g 为 $(0,1)$ 到 $\mathbb{R}$ 上的一一映射. 故 $g \circ f$ 为 (a,b) 到 $\mathbb{R}$ 上的一一映射，从而 $\mathbb{R} \supset (a,b) \sim \mathbb{R}$.

例 1.2.4 注意到

$$A \triangleq (-1,1]\backslash\left(\left\{1-\frac{1}{n}: n \in \mathbb{Z}_{+}\right\} \bigcup\{1\}\right)=(-1,1)\backslash\left(\left\{1-\frac{1}{n}: n \in \mathbb{Z}_{+}\right\}\right) \tag{1.18}$$

记为

$$A \triangleq (-1,1]\backslash B_1=(-1,1)\backslash B_2 \tag{1.19}$$

取映射

$$f: f(1)=0, f\left(1-\frac{1}{n}\right)=1-\frac{1}{n+1}, n \geq 1 \tag{1.20}$$

则 f 为 B_1 到 B_2 上的一一映射，从而 $B_1 \sim B_2$，又 $A \sim A$，进一步得到 $A \cup B_1=(-1,1] \sim (-1,1)=A \cup B_2$.

例 1.2.5 对于 $\mathbb{R}$ 上的区间 (a,b) 和 $(a,b]$. 从 (a,b) 中取出集合 $B=\{a_n, n=1,2,\ldots\}$，其中

$$a_n=a+\left(\frac{1}{b-a}+n\right)^{-1}, \quad n=1,2,\ldots \tag{1.21}$$

则有

$$(a,b)\backslash B=(a,b]\backslash(B \cup\{b\}) \tag{1.22}$$

取 B 到 $B \cup\{b\}$ 上的一一映射满足

$$a_1 \leftrightarrow b ; \quad a_n \leftrightarrow a_{n-1}, n \geq 2 \tag{1.23}$$

则 $B \sim B \cup\{b\}$，进一步得到 $(a,b) \sim(a,b]$.

例 1.2.6 用 $\mathbb{R}$ 表示全体实数组成的集合，则 $\mathbb{R}$ 中的任何有限区间及无限区间都与 $\mathbb{R}$ 对等，即 $\forall a, b \in \mathbb{R}, a<b$,

$$[a,b] \sim(a,b] \sim[a,b) \sim(a,b) \sim(-\infty,b) \sim(a,+\infty) \sim \mathbb{R}. \tag{1.24}$$

推论 1.2.2 设 $I \in \mathbb{R}$，且 I 包含某个开区间，则 $I \sim \mathbb{R}$.

1.2.3　势

定义 1.2.7 (**势（基数）**) 集合的势，也称为集合的基数，是集合中“元素个数”的概念的推广. 若给定的两个集合 A 和 B 对等，则称它们具有相同的势，记为 $\overline{\overline{A}}=\overline{\overline{B}}$.

定义 1.2.8 设 A，B 是两个集合，如果 A 与 B 不对等，但存在 B 的真子集 B^*，有 $A \sim B^*$，则称 A 比 B 有更小的势 B 比 A 有更大的势)，记为 $\overline{\overline{A}}<\overline{\overline{B}}$（或 $\overline{\overline{B}}>\overline{\overline{A}}$）

定义 1.2.9 设 $\overline{\overline{A}}=\mu$，则 A 的全体子集组成的集合的势定义为 $2^\mu>\mu$.

注 1.2.10 根据势的定义，对等实际上是说两个集合的“元素个数”相同，或者说两个集合的“大小”一样.

注 1.2.11 从势的角度来说，Bernstein 定理实际上是一个很自然的事实：

$$\overline{\overline{A}} \geq \overline{\overline{B}}, \overline{\overline{A}} \leq \overline{\overline{B}} \Rightarrow \overline{\overline{A}} = \overline{\overline{B}} \tag{1.25}$$

注 1.2.12 势无最大.

§1.3 可数集与不可数集

1.3.1 可数集

定义 1.3.1 (**可数集**) 与全体正整数 $\mathbb{Z}_+$ 对等的集合称为可数集合或可列集合，并记可数集合的势为 $\aleph_0$，读作“阿列夫零（*Aleph 0*）”.

注 1.3.1 $\mathbb{Z}_+$ 可按大小顺序排成一个无穷序列，因此，一个集合是可数集合的充分必要条件是：A 可以排成一个无穷序列. 简单理解，可数集合就是其中元素可以一个一个列举出来的集合. 可数集中的元素可以用自然数编号，但不一定能够按大小次序排列.

注 1.3.2 可数集合是一个无穷集合，由定义及上一节的推论1.2.1知：全体可数集合对等.

定理 1.3.1 任何无限集合必包含（至少一个）可数子集.

注 1.3.3 这个定理说明：可数集在所有无限集中有最小的势.

定理 1.3.2 可数集合的任何无限子集必然可数，从而可数集合的子集要么有限，要么可数，即至多可数.

定理 1.3.3 设 A 为可数集，B 为有限或可数集，则 $A \cup B$ 为可数集.

定理 1.3.4 设 $A_i(i = 1, 2, \ldots, n)$ 是有限集或可数集，则 $\bigcup\limits_{i=1}^{n} A_i, n \in \mathbb{Z}_+$ 是有限的或可数的，但若至少有一个 A_i 是可数的，则 $\bigcup\limits_{i=1}^{n} A_i$ 必可数.

定理 1.3.5 设 $A_i(i=1,2,\ldots)$ 至多可数，则 $\bigcup\limits_{i=1}^{\infty} A_i$ 是可数的.

例 1.3.1 偶数集、奇数集、整数集 $\mathbb{Z}$、$\mathbb{Z}^2=\{(n,m):n,m\in\mathbb{Z}\}$ 和有理数集 $\mathbb{Q}$ 都可数.

定理 1.3.6 若 $A_i(i=1,2,\ldots,n)$ 是可数集，则 $A_1,A_2,\ldots,A_n$ 的直乘积

$$A_1\times A_2\times\cdots\times A_n:=\{(a_1,\ldots,a_n):a_i\in A_i,i=1,2,\ldots,n\} \tag{1.26}$$

是可数集.

例 1.3.2 全体整系数多项式组成的集合为可数集合. 它们的根称为代数数，全体代数数组成的集合亦为可数集合.

1.3.2　不可数集

定义 1.3.2 (**不可数集**) 不是可数集的无限集称为不可数集合，并记不可数集合的势为 $\aleph$，读作“阿列夫（$Aleph$）”.

注 1.3.4 直线上任一非退化区间都是不可数的【有一个比较常用的区间是 $(0,1)$】.

注 1.3.5 全体实数列成的集合不可数.

注 1.3.6 全体无限二进制小数组成的集合是 $(0,1]$，因而

$$X:=\{(x_1,x_2,\ldots):x_n\in\{0,1\},n=1,2,\ldots\} \tag{1.27}$$

的势为 $\aleph$.

第二章　点集拓扑基础

§2.1　* 拓扑

2.1.1　拓扑与拓扑空间

定义 2.1.1 (**拓扑与拓扑空间**) 对非空集合 X 的一个子集族 $\mathcal{T}$，设 I 为一指标集，若满足

1. $\varnothing \in \mathcal{T}$，$X \in \mathcal{T}$；
2. 若 $\{A_\alpha, \alpha \in I\} \subset T$，则 $\bigcup\limits_{\alpha \in I} A_\alpha \in \mathcal{T}$【对任意并封闭】；
3. 若 $A_1, A_2 \in \mathcal{T}$，则 $A_1 \cap A_2 \in \mathcal{T}$【对有限交封闭】.

则称其为 X 上的一个拓扑，称二元组 $(X, \mathcal{T})$ 为拓扑空间. 在 $\mathcal{T}$ 不致引起混淆的情况下，有时也简记 X 为拓扑空间.

定义 2.1.2 设 X 为一非空集合，$\mathcal{T}_1, \mathcal{T}_2$ 是 X 上的两个拓扑，若 $\mathcal{T}_1 \subset \mathcal{T}_2$，则称 $\mathcal{T}_1$ 粗于 $\mathcal{T}_2$，或 $\mathcal{T}_2$ 细于 $\mathcal{T}_1$.

定义 2.1.3 (**平凡拓扑**) 对于非空集合 X，$\mathcal{T}_1 = \{X, \varnothing\}$ 是其上的一个拓扑，称之为平凡拓扑，它是 X 上最粗的拓扑.

定义 2.1.4 (**离散拓扑**) 对于非空集合 X，$\mathcal{T}_2 = \mathcal{P}(X)$ 是其上的一个拓扑，称之为离散拓扑，它是 X 上最细的拓扑.

2.1.2　一般拓扑空间中开集、闭集和邻域的定义

定义 2.1.5 (**开集**) $(X, \mathcal{T})$ 为拓扑空间，则拓扑 $\mathcal{T}$ 中的元称为开集.

定义 2.1.6 (**闭集**) 闭集一般定义为开集的补集.

注 2.1.1 开集和闭集是一对对偶的概念.

注 2.1.2 X 和 $\varnothing$ 既是开集又是闭集.

注 2.1.3 与拓扑的定义相吻合：任意个开集之并是开集；有限个开集之交是开集.

注 2.1.4 根据 De Morgen 法则：任意个闭集之交是闭集，有限个闭集之并是闭集.

注 2.1.5 注意实际说法与符号表达的等价：$U \in \mathcal{T}$ 说明 U 是开集.

注 2.1.6 离散空间的每一个子集都既是开集又是闭集.

定义 2.1.7 (**邻域**) 设 $(X, \mathcal{T})$ 为拓扑空间，$x \in N \subset X$. 若存在 $U \in \mathcal{T}$，使得 $x \in U \subset N$，则称 N 为 x 的一个邻域，可记作 $N(x)$. 称包含点 x 的开集为开邻域.x 的所有邻域组成的集族为 x 的邻域系，记作 $\mathcal{N}(x)$.

注 2.1.7 由定义，包含 x 的任一开集为 x 的邻域.

注 2.1.8 简单理解，x 的邻域就是 x 和它附近的点.

定理 2.1.1 设 $(X, \mathcal{T})$ 为拓扑空间，$G \subset X$，则

$$G \in \mathcal{T} \Leftrightarrow \forall x \in G, G \in \mathcal{N}(x), \tag{2.1}$$

即 G 为开集的充分必要条件是 G 是它每一个点的邻域.

注 2.1.9 G 是它每一个点的邻域是指：$\forall x \in G$，$\exists U_x \in \mathcal{T}$，成立

$$x \in U_x \subset G. \tag{2.2}$$

从而

$$G = \bigcup_{x \in G} \{x\} \subset \bigcup_{x \in G} U_x \subset G, \tag{2.3}$$

即

$$G = \bigcup_{x \in G} U_x. \tag{2.4}$$

U_x 是邻域，故均为开集，又根据任意开集之并为开集可得 G 为开集.

2.1.3 一般拓扑空间的基

定义 2.1.8 (**基**) 设 $(X, \mathcal{T})$ 为拓扑空间，$\mathcal{B} \subset \mathcal{T}$，如果 $\forall G \in \mathcal{T}$，$\exists \mathcal{B}_G \subset \mathcal{B}$，满足

$$G = \bigcup_{B \in \mathcal{B}_G} B, \tag{2.5}$$

则称 $\mathcal{B}$ 是 X 的一个基.

定义 2.1.9 (**可数基**) 设 $\mathcal{B}$ 是某拓扑空间的一个基，若 $\mathcal{B}$ 是可数的，则称这个拓扑空间具有可数基.

定理 2.1.2 设 $(X,\mathcal{T})$ 为拓扑空间，$\mathcal{B}\subset\mathcal{T}$，则 $\mathcal{B}$ 是 X 的一个基，当且仅当 $\forall G\in\mathcal{T}$ 及 $x\in G$，存在 $B_x\in\mathcal{B}$，使得 $x\in\mathcal{B}_x\subset G$.

定义 2.1.10 (**由基生成的拓扑**) 设 X 为非空集合，$\mathcal{B}\subset\mathcal{P}(X)$.

1. 若 $\mathcal{B}$ 是 X 上某拓扑的基，则
   (a) $\bigcup\limits_{B\in\mathcal{B}} B = X$；
   (b) 若 $B_1,B_2\in\mathcal{B}$ 且 $x\in B_1\cap B_2$，则 $\exists B_x\in\mathcal{B}$，满足 $x\in B_x\subset B_1\cap B_2$.
2. 若上一条成立，则存在唯一的拓扑

$$\mathcal{T}=\left\{G\in X:\exists\mathcal{B}_G\subset\mathcal{B},\ s.t.\ G=\bigcup_{B\in\mathcal{B}_G} B\right\} \tag{2.6}$$

   以 $\mathcal{B}$ 为基，并称这个拓扑是以 $\mathcal{B}$ 为基生成的拓扑.

注 2.1.10 拓扑的交为拓扑.

注 2.1.11 $\mathcal{C}\subset\mathcal{P}(X)$ 为 X 的任意子集族，我们可以定义包含 $\mathcal{C}$ 的最小拓扑

$$\mathcal{T}(\mathcal{C})=\bigcap_{\mathcal{C}\subset\mathcal{P}(X)}\mathcal{T}, \tag{2.7}$$

称其为由 $\mathcal{C}$ 生成的拓扑.

定义 2.1.11 (**子基**) 设 $(X,\mathcal{T})$ 为拓扑空间，$\mathcal{S}\subset\mathcal{T}$，如果 $\mathcal{S}$ 中一切“有限交”构成的集族

$$\mathcal{S}^*=\{S_1\cap S_2\cap\cdots\cap S_n : S_i\in\mathcal{S}, i=1,2,\ldots,n, n\in\mathbb{Z}_+\} \tag{2.8}$$

是 $\mathcal{T}$ 的一个基，即是说 $\mathcal{T}$ 的每个元都是 $\mathcal{S}$ 中的元的有限交的并，则称 $\mathcal{S}$ 为拓扑 $\mathcal{T}$ 的一个子基.

定理 2.1.3 设 X 为非空集合，$\mathcal{S}\subset\mathcal{P}(X)$ 且 $X=\bigcup\limits_{S\in\mathcal{S}} S$，则 X 上有唯一拓扑基以 $\mathcal{S}$ 为子基，称这个拓扑为以 $\mathcal{S}$ 为子基生成的拓扑.

注 2.1.12 为非空集合赋予拓扑的方法有三种：

1. 从开集出发，按照定义赋予；
2. 给出非空集合的基；
3. 给出非空集合的子基.

2.1.4 一般拓扑空间中的点

定义 2.1.12 (**内点**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，$x\in A$为一定点. 如果存在 $N\in\mathcal{N}(x)$，即 N 为 x 的一个邻域，使得 $N\subset A$，则称 x 为 A 的内点.

注 2.1.13 A 内点是说，x 附近全是 A 的点.

注 2.1.14 A 的内点一定属于 A.

定义 2.1.13 (**每一个点都是内点的集合是开集**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$. 若 A 的全部点都是内点，则说明对于 $\forall x\in A$，都 $\exists N(x)\in\mathcal{N}(x)$，使得 $N(x)\subset A$，这是开集的另一种定义.

定义 2.1.14 (**内核**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，称 A 的所有内点组成的集合称为 A 的内核（或称为内部、开核），记为 A°.

注 2.1.15 显然，若 A 中全部点都为内点，则 $A=A^\circ$，若 A 中的点除了内点还有其他点，则 $A^\circ\subset A$. 实际上，A° 是 A 的最大开子集.

注 2.1.16 A° 是包含于 A 的所有开集的并.

注 2.1.17 A 为开集的充分必要条件是 $A=A^\circ$.

注 2.1.18 有限集无内点.

性质 2.1.1 设 $(X,\mathcal{T})$ 为拓扑空间，$\forall B\subset A\subset X$，则 $B^\circ\subset A^\circ$（B 的全部点都是 A 的点，自然有 B 的全部内点都是 A 的内点）.

性质 2.1.2 设 $(X,\mathcal{T})$ 为拓扑空间，$\forall A,B\in X$，则 $(A\cap B)^\circ=A^\circ\cap B^\circ$.

性质 2.1.3 设 $(X,\mathcal{T})$ 为拓扑空间，$\forall A,B\in X$，则 $A^\circ\cup B^\circ\subset(A\cup B)^\circ$.

定义 2.1.15 (**外点**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，则把 $A^c=X\backslash A$ 的内点称为 A 的外点，即对 $x\in A^c$，存在 $N\in\mathcal{N}(x)$，使得 $N\subset A^c$.

注 2.1.19 A 的外点是说，x 附近全是 A^c 的点，即 x 附近全都不是 A 的点.

注 2.1.20 A 的外点一定不属于 A.

定义 2.1.16 (**外部**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，称 A 的所有外点组成的集合称为 A 的外部.

注 2.1.21 根据外点和外部的定义，A 的外部就是 A^c 的内核 $(A^c)^\circ$.

定义 2.1.17 (**界点**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，$x\in X$为一定点. 如果 x 既非 A 的内点，又非 A 的外点，即 $\forall N\in\mathcal{N}(x)$，满足 $N\cap A\neq\varnothing$，$N\cap A^c\neq\varnothing$，则称 x 为 A 的边界点或界点.

定理 2.1.4 设 $(X,\mathcal{T})$ 为拓扑空间，$A\neq\varnothing$，$A\neq X$，则 A 至少有一个界点.

注 2.1.22 A 的界点是说，x 的任一邻域内都既有属于 A 的点，又有不属于 A 的点，即 x 附近既有 A 的点，又有不属于 A 的点.

注 2.1.23 A 的界点不一定属于 A.

注 2.1.24 在同一个拓扑空间中，A 的界点和 A^c 的界点是一样的.

定义 2.1.18 (**边界**) A 的全部界点组成的集合称为 A 的边界，记为 ∂A.

注 2.1.25 有了内点、外点和界点的定义后，我们可以来理解一下集合 A 中的点，对于 $x\in A$ 来说，它要么是内点，要么是界点.A 的内点一定是 A 中的点，我们需要对界点的归属情况仅讨论：

- 若 A 的全部界点都不属于 A，则 A 中的点就全都是内点，从而 A 为开集；

- 若 A 的全部界点都属于 A，即 $\partial A \subset A$，从而 $A = A^\circ \cup \partial A$，此时 A 为闭集，这是因为此时 A^c 的界点全都不属于 A^c，从而 A^c 中全部点都是内点，A^c 为开集；
- 若 A 的一些界点属于 A，而另外一些不属于 A，则此时 $A \subset A^\circ \cup \partial A$.

定义 2.1.19 (**极限点**) 设 $(X, \mathcal{T})$ 为拓扑空间，$A \subset X$，$x \in X$为一定点. 若在 x 的任一邻域内，至少含有一个属于 A 而异于 x 的点，即 $\forall N \in \mathcal{N}(x)$，$A \cap (N \backslash \{x\}) \neq \varnothing$，则称 x 为 A 的极限点，又称聚点.

定理 2.1.5 设 $(X, \mathcal{T})$ 为拓扑空间，$A \subset X$，$x \in X$为一定点，以下说法等价：

- x 是 A 的聚点；
- x 的任一邻域内都含有无穷多个属于 A 的点；
- 存在 A 中互异的点所成的点列 $\{x_n\}$，使得 $x_n \to x(n \to \infty)$.

注 2.1.26 聚点是说，x 周围“聚集”了很多点.

注 2.1.27 由定义知，有限集无聚点.

注 2.1.28 内点一定是聚点.

注 2.1.29 界点可能是聚点，故聚点不一定是内点，且 A 的聚点可能不属于 A.

定义 2.1.20 (**导集**) 设 $(X, \mathcal{T})$ 为拓扑空间，$A \subset X$，$x \in X$为一定点. 称 A 的全部极限点组成的集合为 A 的导集，可记为 A'.

性质 2.1.4 设 $(X, \mathcal{T})$ 为拓扑空间，$\forall B \subset A \subset X$，则 $B' \subset A'$.

性质 2.1.5 设 $(X, \mathcal{T})$ 为拓扑空间，$\forall A, B \subset X$，则 $(A \cup B)' = A' \cup B'$.

性质 2.1.6 设 $(X, \mathcal{T})$ 为拓扑空间，$\forall A, B \subset X$，则 $(A \cap B)' \subset A' \cap B'$.

注 2.1.30 注意到 A 的聚点只有两种情况：A 的内点和 A 的界点. 内点一定是聚点，我们需要对界点的情况进行讨论：

- 若 A 的界点全都是 A 的聚点，则此时有 $A' = A^\circ \cup \partial A$；
- 若 A 的界点全都是 A 的孤立点，则此时有 $A' = A^\circ$；
- 若 A 的界点既有聚点又有孤立点，则此时有 $A' \subset A^\circ \cup \partial A$.

定理 2.1.6 (Bolzano-Weierstrass **定理**) 设 E 是一个有界的无限集合，则 E 至少有一个聚点.

定义 2.1.21 (**孤立点**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，$x\in A$为一定点. 若 $\forall N\in\mathcal{N}(x)$，$N\cap(A\backslash\{x\})=\varnothing$，则称 x 为 A 的孤立点.

注 2.1.31 A 的孤立点是说，x 附近虽然没有“聚集”很多点，但 x 本身属于 A.

注 2.1.32 A 的孤立点一定属于 A.

注 2.1.33 界点不是聚点就是孤立点.

定义 2.1.22 (**闭包**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$，则将包含 A 的最小闭集称为 A 的闭包，一般记为 $\overline{A}$.

定理 2.1.7 A 为闭集当且仅当 $A=\overline{A}$.

性质 2.1.7 设 $(X,\mathcal{T})$ 为拓扑空间，若 $B\subset A\subset X$，则 $\overline{B}\subset\overline{A}$.

性质 2.1.8 设 $(X,\mathcal{T})$ 为拓扑空间，$\forall A,B\in X$，则 $\overline{A\cup B}=\overline{A}\cup\overline{B}$.

性质 2.1.9 设 $(X,\mathcal{T})$ 为拓扑空间，$\forall A,B\in X$，则 $\overline{A\cap B}\subset\overline{A}\cap\overline{B}$.

注 2.1.34 根据定义，显然成立 $A\subset\overline{A}$.

注 2.1.35 A 的外部 $(A^c)^\circ$ 实际上是不包含 A 的最大的开集（A^c 的最大开子集），取其余集 $((A^c)^\circ)^c$ 即为闭包.

注 2.1.36 结合前述定义，我们可以得到：

$$\overline{A}=A\cup\partial A=A^\circ\cup\partial A=A\cup A' \tag{2.9}$$

注 2.1.37 注意到若 A 为闭集，则 $A=\overline{A}$，由 $\overline{A}=A\cup A'$ 可得 $A'\subset A$，A 的任一聚点属于 A. 即 A 为闭集当且仅当 $A'\subset A$.

注 2.1.38 借助闭包的定义，我们可以发现：

$$\partial A=\overline{A}\cap\overline{(A^c)}=\overline{A}\cap(A^\circ)^c=\overline{A}\backslash A^\circ \tag{2.10}$$

2.1.5　一般拓扑空间中序列的收敛性

定义 2.1.23 (**收敛性**) 设 $(X,\mathcal{T})$ 为拓扑空间，$\{x,x_n,n\geq 1\}\subset X$，如果 $\forall N\in\mathcal{N}(x)$，$\{x_n\}$ 始终落在 N 中，即 $\exists n_0\in\mathbb{Z}_+$，当 $n>n_0$ 时，$x_n\in N$，则称 $\{x_n\}$ 收敛于 x，记为 $x_n\to x$ 或 $\lim\limits_{n\to\infty}x_n=x$.

注 2.1.39 设 $(X,\mathcal{T})$ 为拓扑空间，$x\in X$，$A\subset X$，若在 $A\backslash\{x\}$ 中有序列收敛于 x，则 $x\in A'$.

2.1.6　一般拓扑空间的可分性、列紧性与紧性

为了给出可分性与可分空间的概念，首先介绍稠密、稠密集和疏朗集.

定义 2.1.24 (**稠密**) 设 $(X,\mathcal{T})$ 是拓扑空间，$B\subset A\subset X$. 若 A 中的每个点的任意邻域中都含有 B 的点，即 $A\subset\overline{B}$，则称 B 在 A 中稠密.

定义 2.1.25 (**稠密集**) 设 $(X,\mathcal{T})$ 是拓扑空间，$A\subset X$，若 $\overline{A}=X$，则称 A 为稠密集.

定义 2.1.26 (**疏朗集**) 设 $(X,\mathcal{T})$ 是拓扑空间，$A\subset X$，若 $\overline{A}$ 不含任何非空开集，则称 A 为疏朗集，或称为无处稠密集合.

下面给出可分性与可分空间的定义.

定义 2.1.27 (**可分性与可分空间**) 设 $(X,\mathcal{T})$ 是拓扑空间，$A\subset X$，若 A 有可数的稠子集，则称 A 是可分的. 若 X 自身是可分的，则称该拓扑空间为可分空间.

注 2.1.40 为了研究可分空间的某个性质，可以首先考虑该空间上的一个可数稠密子集，然后利用稠密性将结果推广到整个空间上去. 这是可分性的重要用处.

为了给出列紧性和紧性，首先介绍覆盖的概念.

定义 2.1.28 (**覆盖**) 设 $(X,\mathcal{T})$ 为拓扑空间，$C\subset X$，$\mathcal{A}$ 为 X 的子集族，如果满足：

$$C\subset\bigcup_{A\in\mathcal{A}}A, \tag{2.11}$$

则称 $\mathcal{A}$ 为 C 的覆盖. 若覆盖中的元都是 X 中的开集（闭集），则称该覆盖是开覆盖（闭覆盖）；若覆盖中的元的个数有限（可数），则称该覆盖是有限覆盖（可数覆盖）.

定义 2.1.29 (**子覆盖**) 设 $(X,\mathcal{T})$ 为拓扑空间，$C\subset X$，若 X 的子集族 $\mathcal{A}$ 和 $\mathcal{B}$ 都是 C 的覆盖，且满足 $\mathcal{A}\subset\mathcal{B}$，则称 $\mathcal{A}$ 是 $\mathcal{B}$ 的子覆盖.

下面给出紧集的概念.

定义 2.1.30 (**紧集与紧空间**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$. 若 A 的任何开覆盖都有一个有限的子覆盖，则称 A 为紧集；若 X 自身是紧集，则称该拓扑空间为紧空间.

引理 2.1.1 (E,d) 为紧空间的充分必要条件是：若它的每一个闭集族的任意有限子族的交不空，则此闭集族的交不空.

引理 2.1.2 紧空间的闭子集是紧集.

列紧性是弱于紧性的一种性质.

定义 2.1.31 (**列紧集与列紧空间**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$. 若 A 中任一序列都有收敛的子序列（不要求极限点在 A 中），则称 A 为列紧集. 若 X 本身是列紧的，则称该拓扑空间是列紧空间.

定义 2.1.32 (**自列紧集**) 设 $(X,\mathcal{T})$ 为拓扑空间，$A\subset X$. 若 A 是列紧闭集，则称 A 为子列紧集，即 A 中任何序列总有收敛于 A 中某点的子序列.

2.1.7 一般拓扑空间上的连续映射

定义 2.1.33 设 $(X,\mathcal{T})$ 和 $(Y,\mathcal{U})$ 是两个拓扑空间，$f:X\to Y$.

1. 如果 Y 中的点 $f(x)$ 的每个邻域在 f 下的逆像都是 X 中点 x 的邻域，则称 f 在点 $x\in X$ 是 $\mathcal{T}-\mathcal{U}$ 连续的.
2. 如果 Y 的每个开集在 f 下的逆象都是 X 的开集，则称 f 是 $\mathcal{T}-\mathcal{U}$ 连续的.

§2.2 度量空间

2.2.1 距离与度量空间的定义

定义 2.2.1 (**距离与度量空间**) 设 X 为一非空集合，对于 X 中的任意两个元素 x 和 y，都有唯一确定的实数 $d(x,y)$（有时也记为 $\rho(x,y)$）与之对应（即存在一个映射或称为双变量实值函数 $d:X\times X\mapsto\mathbb{R}$)，若 $\forall x,y,z\in X$，满足：

1. $d(x,y)\geq 0$;
2. $d(x,y)=0\Leftrightarrow x=y$;
3. $d(x,y)=d(y,x)$;
4. $d(x,y)\leq d(x,z)+d(z,y)$.

则称 d 为 X 上的一个距离，或称之为一个度量. 称 (X,d) 为度量空间或距离空间.X 中的元素称为点. 在距离 d 不必具体指出或不易引起混淆时，也简称 X 为度量空间.

注 2.2.1 度量空间是一类特殊的拓扑空间，它在拓扑空间中定义了距离.

注 2.2.2 借助距离这个概念，我们可以重新认识一下度量空间中“收敛”这个概念. 对于度量空间 (X,d) 中的一个点列 $\{x_n\}$，它收敛到 x_0 是指

$$\lim_{n\to\infty} d(x_n,x_0)=0 \tag{2.12}$$

记作 $\lim\limits_{n\to\infty} x_n=x_0$，或简记为 $x_n\to x_0$.

引理 2.2.1 设 (E,d) 为度量空间，则 $\forall x,y,z\in E$，有

$$|d(x,z)-d(y,z)|\leq d(x,y). \tag{2.13}$$

2.2.2 一些度量空间及其上的距离

例 2.2.1 (**离散度量空间**) 给定非空集合 E，$\forall x,y\in E$，定义

$$d(x,y)=\begin{cases}0, & x=y\\ 1, & x\neq y\end{cases}, \tag{2.14}$$

则 d 是一个距离，(E,d) 称为离散度量空间.

例 2.2.2 (**有界函数空间**) 设 A 是给定的集合，$B(A)$ 表示 A 上有界实值（或复值）函数全体 k，对 $B(A)$ 中任意两点（实际上是两个函数）x，y，定义

$$d(x,y)=\sup_{t\in A}|x(t)-y(t)| \tag{2.15}$$

, 则 d 是一个距离，(A,d) 称为有界函数空间.

例 2.2.3 (**连续函数空间**) 设 $C[a\ b]$ 表示闭区间 $[a\ b]$ 上的实值（或复值）连续函数全体，对 $C[a\ b]$ 中任意两点 x，y 定义

$$d(x,y)=\max_{t\in[a,b]}|x(t)-y(t)|, \tag{2.16}$$

则 d 是一个距离，$(C[a,b],d)$ 称为连续函数空间.

例 2.2.4 (**实数空间**) 在实数集 $\mathbb{R}$ 上，对 $\forall x,y\in\mathbb{R}$，常定义

$$d(x,y)=|x-y|, \tag{2.17}$$

则 d 是一个距离，$(\mathbb{R},d)$ 称为实数空间.

例 2.2.5 (n **维欧氏空间**) 在 $\mathbb{R}^n=\{(x_1,\dots,x_n):x_i\in\mathbb{R},i=1,2,\dots,n\}$ 中，对 $\forall x,y\in\mathbb{R}^n$，常定义

$$d_p(x,y)=\left[\sum_{k=1}^{n}|x_k-y_k|^p\right]^{\frac{1}{p}},\quad p\geq 1,\tag{2.18}$$

则 d_p 是一个距离. 若取 $p=2$，则

$$d_2(x,y)=\sqrt{\sum_{k=1}^{n}|x_k-y_k|^2}\tag{2.19}$$

是通常的欧氏距离，称 $(\mathbb{R}^n,d_2)$ 为 n 维欧氏空间.

注 2.2.3 在式 (2.18) 的 d_p 的定义下，易见当 $n=1$，$p=1$ 时，空间 $(\mathbb{R}^1,d_1)$ 既为实数空间.

注 2.2.4 之所以说式 (2.17) 和 (2.18) 是常定义的距离，是因为在 $\mathbb{R}$ 和 $\mathbb{R}^n$ 上，距离的定义方法有很多，如在 $\mathbb{R}^n$ 上还可以定义如下距离

$$d_\infty(x,y)=\max_{1\leq k\leq n}|x_k-y_k|,\tag{2.20}$$

$$d'(x,y)=\sum_{k=1}^{n}|x_k-y_k|,\tag{2.21}$$

它们也满足距离的定义. 距离定义的不同会影响到空间的完备性.

注 2.2.5 可以验证

$$\lim_{p\to\infty}d_p(x,y)=d_\infty(x,y).\tag{2.22}$$

注 2.2.6 往后提到实数空间和 n 维欧式空间时，如无特殊说明，默认采用 d_1 和通常欧氏距离 d_2 的定义.

注 2.2.7 不止是 $\mathbb{R}^n$ 上的距离可以有多种定义，其他空间亦可定义不同的距离，但一般都有一个默认的距离定义，不妨称之为通常距离.

2.2.3 Lévy 距离与分布函数空间

定义 2.2.2 (Lévy **距离**) 设 $\mathscr{D}_F$ 表示一切按右连续定义的概率分布函数组成的集合.$\forall F,G\in\mathscr{D}_F$，定义

$$L(F,G)=\inf\{\varepsilon>0:F(x-\varepsilon)-\varepsilon\leq G(x)\leq F(x+\varepsilon)+\varepsilon,\quad x\in\mathbb{R}\},\tag{2.23}$$

则 L 是 $\mathscr{D}_F$ 上的一个距离，称为 Lévy 距离.

定理 2.2.1 (**分布函数弱收敛**) 概率分布函数列 $\{F_n(x)\}$ 弱收敛于概率分布函数 $F(x)$ 的充分必要条件是当 $n \to 0$ 时，$L(F_n, F) \to 0$.

2.2.4　度量空间中的点集

度量空间是定义了距离的拓扑空间. 故一般拓扑空间中成立的开集、闭集、邻域的概念和性质同样适用于度量空间. 但我们可以借助距离将一些概念特殊化.

定义 2.2.3 (**开球、闭球和球面**) 设 (X, d) 是度量空间，$x \in X$，$\varepsilon > 0$，则分别称

$$B(x, \varepsilon) := B_d(x, \varepsilon) := \{y \in X : d(x, y) < \varepsilon\}, \tag{2.24}$$

$$\overline{B}(x, \varepsilon) := \overline{B}_d(x, \varepsilon) := \{y \in X : d(x, y) \leq \varepsilon\} \tag{2.25}$$

为以 x 为中心，ε 为半径的开球、闭球. 特别地，称

$$S(x, \varepsilon) := \{y \in X : d(x, y) = \varepsilon\} \tag{2.26}$$

为球面.

定义 2.2.4 (**度量空间上的开集**) 设 (X, d) 是度量空间，设 $G \subset X$，如果 $\forall x \in G$，$\varepsilon > 0$，满足 $B(x, \varepsilon) \subset G$，则称 G 为 d-开集，简称开集.

注 2.2.8 度量空间上的闭集，我们同样将其定义为开集的余集.

注 2.2.9 任何开球是开集，但不是所有的开集都是开球.

注 2.2.10 任何闭球是闭集，但不是所有的闭集都是闭球.

定义 2.2.5 (**度量拓扑**) 设 (X, d) 是度量空间，用 $\mathcal{T}_d$ 表示 X 中全体 d-开集，则

1. $\varnothing, X \in \mathcal{T}_d$，即空集和全集为开集；
2. 若 $\mathcal{A} \subset \mathcal{T}_d$，则 $\bigcup\limits_{A \in \mathcal{A}} A \in \mathcal{T}_d$，即开集对任意并封闭；
3. 若 $A_1, A_2 \in \mathcal{T}_d$，则 $A_1 \cap A_2 \in \mathcal{T}_d$，即开集对有限交封闭.

从而

$$\mathcal{T}_d = \{G \in X : G\text{是 } d\text{-开集}\} \tag{2.27}$$

满足拓扑的定义，即它是 X 上的一个拓扑，称之为由 d 诱导的度量拓扑. 特别地，由通常距离诱导产生的拓扑称为通常拓扑.

注 2.2.11 我们约定把度量空间 (X,d) 看作拓扑空间时，指得就是 $(X,\mathcal{T}_d)$.

定义 2.2.6 (**度量空间上的邻域**) 设 (X,d) 是度量空间，$x\in X$，$N\subset X$，若存在 X 上的 d-开集 G，使得 $x\in G\subset N$，则称 N 为 x 的邻域.

注 2.2.12 这个定义本质上和一般拓扑空间中的定义是等价的.

定义 2.2.7 (**球形邻域**) 以 x 为中心，ε 为半径的开球 $B(x,\varepsilon)$ 是邻域，亦称为 x 的球形邻域，或称为 x 的 ε-邻域.

注 2.2.13 球形邻域是其中每一个点的邻域.

注 2.2.14 度量空间的所有球形邻域组成的集族

$$\{\mathcal{B}(x,\varepsilon):x\in X,\varepsilon>0\} \tag{2.28}$$

是它的一个基.

性质 2.2.1 球形邻域具有如下基本性质：

1. x 至少有一个球形邻域，并且 $x\in B(x,\varepsilon)$，$\forall\varepsilon>0$.
2. 对于 x 的任意两个不同的球形邻域 $B(x,\varepsilon)$ 和 $B(x,\varepsilon_2)$，存在一个球形邻域 $B(x,\varepsilon_3)$，使得
$$B(x,\varepsilon_3)\subset B(x,\varepsilon_1)\cap B(x,\varepsilon_2). \tag{2.29}$$
3. 如果 $y\in B(x,\varepsilon)$，则存在 $B(y,\varepsilon')\subset B(x,\varepsilon)$.
4. 对于 $x\neq y$，存在 $B(x,\varepsilon)$ 和 $B(y,\varepsilon')$，使 $B(x,\varepsilon)\cap B(y,\varepsilon')=\varnothing$.

利用球形邻域的概念，将一般拓扑空间中的内点、内核、外点、外部、界点、边界、聚点、导集、孤立点以及闭包的定义引申过来就能得到度量空间中的相关定义，它们在一般拓扑空间中成立的性质当然也在度量空间中成立，此处不重复表述.

需要注意的是，度量空间的上述概念会成立一些特殊的性质，这些性质在一般拓扑空间中未必成立，如：

性质 2.2.2 设 (X,d) 为度量空间，$E\subset X$，则它的导集 E' 一定是闭集.

2.2.5　度量空间中集合的有界性

定义 2.2.8 (**点集间的距离**) 两个非空点集 A 和 B 之间的距离定义为

$$d(A,B)=\inf\{d(x,y):x\in A,y\in B\}. \tag{2.30}$$

注 2.2.15 若 $A=\{x\}$，即 A 为单点集，则把点 x 与集合 B 的距离定义为：

$$d(x,B):=d(\{x\},B)=\inf\{d(x,y):y\in B\} \tag{2.31}$$

定义 2.2.9 (**点集的直径**) 一个非空点集 E 的直径定义为

$$\delta(E)=\sup_{x\in E,y\in E}d(x,y). \tag{2.32}$$

定义 2.2.10 (**有界点集**) 设 E 为度量空间上的一个点集，如果 $\delta(E)\le\infty$，则称 E 为有界点集. 特别地，把空集也作为有界点集.

定理 2.2.2 设 A 为度量空间 (E,d) 中的一个点集，则 A 为有界集当且仅当 $\exists a\in X$，$r\in(0,\infty)$，使 $A\subset B(a,r)$.

注 2.2.16 如下性质亦成立：

1. 若 $\varnothing\neq A\subset E$，则 $\forall x,y\in E$，有
$$|d(x,A)-d(y,A)|\le d(x,y) \tag{2.33}$$
2. 对任意的球 $B(a,r)$ 都有
$$\delta(B(a,r))\le\delta(\overline{B}(a,r))\le 2r \tag{2.34}$$
3. 有界集的子集仍然有界.
4. (X,d) 的任何两个有界集 A,B 的并有界，而且有
$$\delta(A\cup B)\le d(A,B)+\delta(A)+\delta(B). \tag{2.35}$$

2.2.6　度量空间上的完备性

定义 2.2.11 (**度量空间上的序列收敛**) 设 (X,d) 是度量空间，$\{x,x_n,n\ge 1\}\subset X$，称如果 $\forall\varepsilon>0$，$\exists n_0\in\mathbb{Z}_+$，当 $n>n_0$ 时，有

$$d(x_n,x)<\varepsilon, \tag{2.36}$$

则称 $\{x_n\}$ 收敛于 x，记为 $x_n \to x$ 或 $\lim\limits_{n\to\infty} x_n = x$.

定义 2.2.12 (**基本列**) 设 (X,d) 是度量空间，$\{x_n, n \geq 1\} \subset X$，若 $\forall \varepsilon > 0$，$\exists N \in \mathbb{Z}_+$，当 $m,n > N$ 时，有

$$d(x_m, x_n) < \varepsilon, \tag{2.37}$$

即 $\lim\limits_{m,n\to\infty} d(x_m,x_n) = 0$，则称序列 $\{x_n\}$ 为基本列.

定理 2.2.3 在度量空间上，收敛列都是基本列.

定义 2.2.13 (**完备性**) 设 (X,d) 是度量空间，若其中的任何基本列都在 (X,d) 中收敛，则称该空间是完备度量空间.

注 2.2.17 完备性定义实际上要求度量空间中的基本列都收敛，且收敛点仍落于该度量空间当中.

引理 2.2.2 在度量空间 (E,d) 中，基本列收敛当且仅当它有一收敛子列.$a \in \overline{A} \subset E$ 当且仅当存在序列 $\{a_n\} \subset A$，$a_n \to a$.

定义 2.2.14 (**自密集**) 设 (X,d) 是度量空间，$E \subset X$，若 $E \subset E'$，则称 E 为自密集.

注 2.2.18 自密集实际上就是集合中的点全都是聚点（亦即无孤立点）的集合.

定义 2.2.15 (**完备集**) 设 (X,d) 是度量空间，$E \subset X$，若 $E = E'$，则称 E 为完备集（或完全集）.

注 2.2.19 完备集就是自密闭集（即没有孤立点的闭集）.

2.2.7 度量空间上的紧性特征

度量空间上的稠密、疏朗、可分、紧和列紧的定义与一般拓扑空间上的定义是一致的，只需将前面介绍的定义中的拓扑空间 $(X,\mathcal{T})$ 的设定改为度量空间 (X,d) 即可，此处不再重复表述. 下面讨论一些度量空间上的紧性特征.

我们在前面提到了度量空间上的有界点集，现引入可分性的概念，可以得到一种更强的有界性：完全有界.

定义 2.2.16 (ε-**网**) 设 (X,d) 是度量空间，$A,C\subset X$. 如果存在 $\varepsilon>0$，使得

$$\bigcup_{x\in C}B(x,\varepsilon)\supset A, \tag{2.38}$$

则称 B 是 A 的 ε-网.

定义 2.2.17 (**完全有界**) 设 (X,d) 是度量空间，$A\subset X$. 如果 $\forall\varepsilon>0$，A 总有有限的 ε-网 $\{x_1,x_2,\dots,x_n\}\subset A$（点的个数 n 可以随 ε 变化），即 $\exists n=n(\varepsilon)\in\mathbb{Z}_+$，使

$$\bigcup_{i=1}^{n}B(x_i,\varepsilon)\supset A, \tag{2.39}$$

那么称 A 是完全有界的.

定理 2.2.4 设 A 是度量空间的完全有界集，则 A 是有界的且是可分的.

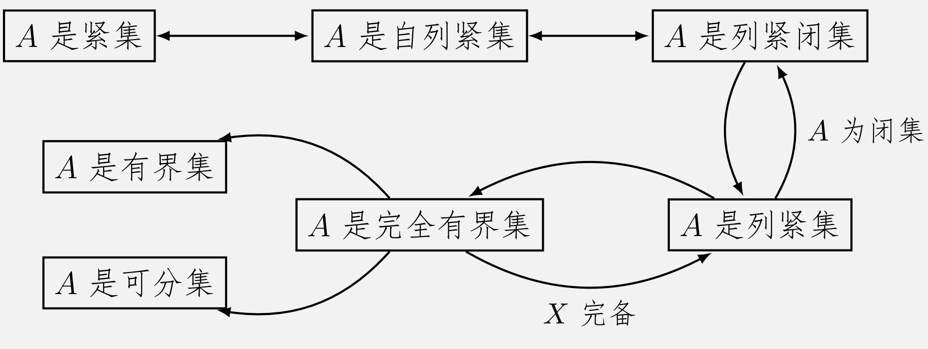
定理 2.2.5 设 (X,d) 是度量空间，$A\subset X$. 则 A 是完全有界的，当且仅当 A 中任何一个序列 $\{x_n,n\geq 1\}$ 都包含一个基本子序列.

注 2.2.20 注意到，列紧性强于完全有界性，即若 A 是列紧的，则它是完全有界的.

定理 2.2.6 设 (X,d) 是完备度量空间，$A\subset X$. 则 A 是列紧的，当且仅当 A 是完全有界的.

定理 2.2.7 设 (X,d) 是度量空间，$A\subset X$. 则 A 是紧集，当且仅当 A 是列紧闭集（即自列紧集）.

注 2.2.21 设 (X,d) 为度量空间，$A\subset X$. 则 A 是紧集、自列紧集、列紧闭集、列紧集、完全有界集、有界集和可分集之间的关系如下图所示：



定义 2.2.18 (**相对紧集**) 设 (X,d) 是度量空间，$A\subset X$. 若 $\overline{A}$ 是紧集，则称 A 为相对紧集.

定义 2.2.19 (**局部紧空间**) 设 (X,d) 是度量空间. 若 X 中的每一点都有一个相对紧的邻域，则称该空间为局部紧空间.

注 2.2.22 下列三个命题等价：

1. E 可分；
2. E 有可数拓扑基；
3. E 的子集的任一覆盖都有可数子覆盖.

引理 2.2.3 度量空间的紧子集是闭集.

引理 2.2.4 设 K 是 (E,d) 的一个子集，则下列三个命题等价：

1. K 列紧；
2. $\overline{K}$ 紧；
3. $\overline{K}$ 完备且 K 全有界.

2.2.8 度量空间上的连续映射

定义 2.2.20 (**度量空间上的连续映射**) 设 (X,d) 和 (Y,ρ) 是两个度量空间，$f:X\to Y$.

1. 如果 $\forall\varepsilon>0$，$\exists\delta>0$，只要 $x,x'\in X$，$d(x,x')<\delta$，就有
$$\rho(f(x),f(x'))<\varepsilon, \tag{2.40}$$
就称 f 在点 x 处是 d-ρ 连续的.

2. 如果 $\forall\varepsilon>0$，$\exists\delta>0$，$\forall x\in X$，只要 $x'\in X$，$d(x,x')<\delta$，就有
$$\rho(f(x),f(x'))<\varepsilon, \tag{2.41}$$
就称 f 是 d-ρ 连续的，简称连续的（即要求 f 在 X 的每一点都连续）.

定理 2.2.8 度量空间 X 到 Y 中的映射 f 是 X 上连续映射的充要条件为 Y 中任一开集 M 的原象 $f^{-1}(M)$ 是 X 中的开集.

§2.3 n 维欧氏空间

n 维欧式空间 $\mathbb{R}^n$ 及其特殊情况实数空间 $\mathbb{R}$（下面亦称直线）是两个最基本的度量空间，故前面对一般度量空间成立的诸多定义、性质，对它们也是成立的. 本节将它们在 $\mathbb{R}^n$ 上进行特殊讨论，并阐述一些在 n 维欧式空间 $\mathbb{R}^n$ 上成立但在一般度量空间上未必成立的事实.

2.3.1 n 维欧氏空间上的基本概念

首先，将度量空间上的邻域的概念对应到 n 维欧氏空间中，就得到了一些我们可能之前已经见过的概念：在 $\mathbb{R}$ 中，邻域对应成一个开区间；在 $\mathbb{R}^2$ 中，邻域对应成一个开圆；在 $\mathbb{R}^3$ 及更高维空间中，邻域对应成一个开球.

其次，$\mathbb{R}^n$ 中有一个特殊的点：$\mathbf{0}=(0,0,\ldots,0)$，称为 n 维实空间的原点. 借助原点的概念，$\mathbb{R}^n$ 上的有界点集有如下结论：

定理 2.3.1 设 E 为 $\mathbb{R}^n$ 中的一个点集，则它是有界点集，当且仅当存在 $K>0$，使对于所有的 $x=(x_1,\ldots,x_n)\in E$，都有 $|x_i|\leq K(i=1,2,\ldots,n)$. 即 $\exists K>0$，$\forall x\in E$，$d(x,0)\leq K$.

另外，需要对 $\mathbb{R}^d$ 上的开区间和闭区间给出定义

定义 2.3.1 (**开区间**) 给定 $a_k:=(a_1^{(k)},\cdots,a_d^{(k)})\in\mathbb{R}^d,k=1,2$，称

$$(a_1,a_2):=\{(x_1,\ldots,x_d)\in\mathbb{R}^d:a_i^{(1)}<x_i<a_i^{(2)},i=1,2,\ldots,d\},\tag{2.42}$$

为开区间.

定义 2.3.2 (**闭区间**) 给定 $a_k:=(a_1^{(k)},\cdots,a_d^{(k)})\in\mathbb{R}^d,k=1,2$，称

$$[a_1,a_2]:=\{(x_1,\ldots,x_d)\in\mathbb{R}^d:a_i^{(1)}\leq x_i\leq a_i^{(2)},i=1,2,\ldots,d\}\tag{2.43}$$

为闭区间.

注 2.3.1 对 n 维欧氏空间而言，我们把由通常欧氏距离 d_2（由（2.19）定义）诱导的度量拓扑 $\mathcal{T}_{d_2}$ 称为 $\mathbb{R}^n$ 上的通常拓扑.

注 2.3.2 $\mathbb{R}^n$ 中所有有限开区间组成的集族

$$\{(\mathbf{a},\mathbf{b}):\mathbf{a},\mathbf{b}\in\mathbb{R}^d,\mathbf{a}<\mathbf{b}\}\tag{2.44}$$

是 $\mathbb{R}^d$ 上通常拓扑的一个基.

2.3.2 直线上的开集、闭集和完备集的构造

$\mathbb{R}^n$ 上的开集的结构性质及其构造方法，对后面我们在这个空间上分配测度十分重要. 这里首先从一维直线出发进行讨论. 首先给出构造区间、余区间的定义.

定义 2.3.3 (**构成区间**) 设 G 是直线上的开集，如果开区间 $(\alpha,\beta)\subset G$，且 $\alpha,\beta\notin G$，则称 (α,β) 为 G 的构成区间.

定义 2.3.4 (**余区间**) 设 F 是直线上的闭集，成 F 的余集 F^c 为 F 的余区间，或称邻接区间.

下面给出直线上的开集、闭集和完备集的构造

定理 2.3.2 (**直线上的开集的构造**) 直线上任一个非空开集可以表示成有限个或可数个不相交的构成区间的和集（并）. 即若设 G 是直线上任一非空开集，则存在开集族

$$\{(\alpha_k, \beta_k) : -\infty \leq \alpha_k < \beta_k \leq \infty, k \in I\}, \tag{2.45}$$

其中 $I = \{1, 2, \ldots, n\}$ 或 $\mathbb{Z}_+$，$(\alpha_k, \beta_k), k \in I$ 两两不交，使

$$G = \bigcup_{k \in I} (\alpha_k, \beta_k). \tag{2.46}$$

定理 2.3.3 (**直线上的闭集的构造**) 直线上的闭集 F 或者是全直线，或者是从直线上挖掉有限个或可数个互不相交的开区间（即挖去 F 的余区间）所得到的集合.

定理 2.3.4 (**直线上的完备集的构造**) 由孤立点的定义，直线上点集 A 的孤立点必是 A^c 中的某两个开区间的公共端点. 因此，闭集的孤立点一定是它的两个余区间的公共端点. 而完备集是没有孤立点的闭集，从而完备集就是没有相邻接的余区间的闭集.

2.3.3 $\mathbb{R}^n$ 上的开集结构

将直线上的结构向 $\mathbb{R}^n$ 推广，首先定义左开右闭、左闭右开区间

定义 2.3.5 (**左开右闭区间、左闭右开区间**) 给定 $a_k := (a_1^{(k)}, \cdots, a_d^{(k)}) \in \mathbb{R}^d, k = 1, 2$，分别称

$$(a_1, a_2] := \{(x_1, \ldots, x_d) \in \mathbb{R}^d : a_i^{(1)} < x_i \leq a_i^{(2)}, i = 1, 2, \ldots, d\}, \tag{2.47}$$

$$[a_1, a_2) := \{(x_1, \ldots, x_d) \in \mathbb{R}^d : a_i^{(1)} \leq x_i < a_i^{(2)}, i = 1, 2, \ldots, d\} \tag{2.48}$$

为左开右闭区间和左闭右开区间.

$\mathbb{R}^n$ 上的开集有如下结果：

定理 2.3.5 (**$\mathbb{R}^n$ 上的开集结构**) $\mathbb{R}^n$ 中任一非空开集都能表成可列个两两不交的左开右闭（或左闭右开）的正方体之并.

2.3.4 $\mathbb{R}^n$ 上的完备性

我们前面提到过，$\mathbb{R}^n$ 上的距离的定义方法有很多，但距离定义的不同会影响空间的完备性. 根据上文定义的 $d_p(x, y)$（2.18）和 $d_\infty(x, y)$（2.20）两个距离而言，$(\mathbb{R}^n, d_p)$ 和 $(\mathbb{R}^n, d_\infty)$ 都是完备的. 此外直线上（按（2.17）定义距离）的任一闭区间以及全直线都是完备的.

2.3.5　$\mathbb{R}^n$ 上的紧性特征

首先给出一个有用的有限覆盖定理

定理 2.3.6 (Heine-Borel **有限覆盖定理**) 在 $\mathbb{R}^n$ 中，设 F 是一个有界闭集，Λ 为一指标集，$\mathscr{U}$ 是一族开集 $\{U_i\}_{i\in\Lambda}$，它覆盖了 F，即

$$F \subset \bigcup_{i\in\Lambda} U_i, \tag{2.49}$$

则 $\mathscr{U}$ 中一定存在有限多个开集 $U_1, U_2, \ldots, U_m$，它们同样覆盖了 F，即

$$F \subset \bigcup_{i\in 1}^{m} U_i. \tag{2.50}$$

注 2.3.3 上述定理是在 $\mathbb{R}^n$ 中成立的定理，结合紧集的定义我们可以发现：$\mathbb{R}^n$ 中的有界闭集就是紧集.

注 2.3.4 上注的结果在一般拓扑空间或一般度量空间中是不成立的. 但是根据前面的讨论我们发现：一般度量空间上的紧集一定是有界闭集，逆不真.

注 2.3.5 设 $\mathbb{R}^n$ 为 n 维欧氏空间，$A \subset \mathbb{R}^n$. 由于 $\mathbb{R}^n$ 具有许多良好的性质，它不仅完备而且其上的集合都是可分的，从而 A 一定是可分集，进一步 A 是紧集、自列紧集、列紧闭集、列紧集、完全有界集、有界集之间的关系如下图所示：

$$\begin{tikzcd}
\boxed{A\text{ 是紧集}} \arrow[r, leftrightarrow] & \boxed{A\text{ 是自列紧集}} \arrow[r, leftrightarrow] & \boxed{A\text{ 是列紧闭集}} \arrow[d, bend right] \\
\boxed{A\text{ 是有界集}} \arrow[r, leftrightarrow] & \boxed{A\text{ 是完全有界集}} \arrow[r, leftrightarrow] & \boxed{A\text{ 是列紧集}} \arrow[u, bend right, "A\text{ 为闭集}"']
\end{tikzcd}$$

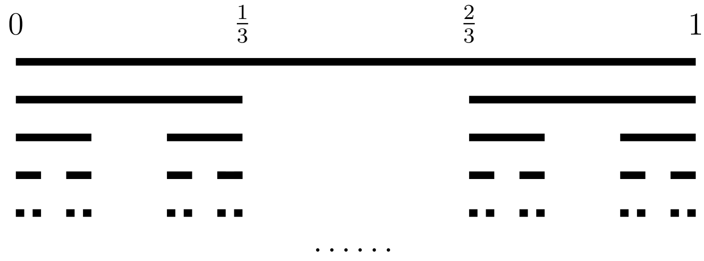
2.3.6　Cantor 集

将闭区间 $[0,1]$ 三等分，去掉中间的开区间 $\left(\dfrac{1}{3}, \dfrac{2}{3}\right)$，共去掉（$2^0 = 1$）个开区间. 按照同样的方法，将闭区间 $\left[0, \dfrac{1}{3}\right]$ 和 $\left[\dfrac{1}{3}, \dfrac{2}{3}\right]$ 各自三等分并去掉中间的开区间 $\left(\dfrac{1}{3^2}, \dfrac{2}{3^2}\right)$ 和 $\left(\dfrac{2\cdot 3+1}{3^2}, \dfrac{2\cdot 3+2}{3^2}\right)$，共去掉（$2^1 = 2$）个开区间. 依此进行下去，第 n 次去掉 2^{n-1} 个开区

间，形如

$$\left(\frac{1}{3^n}\left(\sum_{k=1}^{n-1} a_{n-k}3^k+1\right), \frac{1}{3^n}\left(\sum_{k=1}^{n-1} a_{n-k}3^k+2\right)\right), \quad a_k\in\{0,2\}, k=1,2,\dots,n-1. \tag{2.51}$$

上述构造方法如下图所示



定义 2.3.6 (Cantor **集**) 令

$$G_0 := \bigcup_{n=1}^{\infty} \bigcup_{\substack{a_k=0,2\\ k=1,\dots,n-1}} \left(\frac{1}{3^n}\left(\sum_{k=1}^{n-1} a_{n-k}3^k+1\right), \frac{1}{3^n}\left(\sum_{k=1}^{n-1} a_{n-k}3^k+2\right)\right). \tag{2.52}$$

称 $P_0=[0,1]\backslash G_0$ 为 Cantor 集.

注 2.3.6 Cantor 集是不可数集（它的势为 $\aleph$）. 可将 $[0,1]$ 中的实数表示为三进制的小数，则

$$G_0 = \bigcup_{n=1}^{\infty} \bigcup_{\substack{a_k=0,2\\ k=1,\dots,n-1}} (0.a_1a_2\cdots a_n1, 0.a_2a_2\cdots a_n2). \tag{2.53}$$

其构成区间为 $(0.a_1a_2\cdots a_n1, 0.a_2a_2\cdots a_n2)$，区间内的实数形如 $0.a_1a_2\cdots a_n1a_{n+1}a_{n+2}\cdots$，即小数部分至少有一位是 1，从而

$$A := \{a\in[0,1] : a=0.a_1a_2\cdots a_n\cdots, a_k\in 0,2, k=1,2,\dots\} \tag{2.54}$$

中一定没有 G_0 的元素. 故 $A\subset P_0\subset[0,1]$. 设

$$B := \{(b_1,b_2,\dots) : b_k\in\{0,1\}, k=1,2,\dots\}\sim[0,1]. \tag{2.55}$$

取一一映射 $\phi : A\mapsto B$ 为

$$\phi(0.a_1a_2\cdots) = (b_1,b_2,\dots), \quad 0.a_1a_2\cdots\in A, \tag{2.56}$$

其中

$$b_k = \begin{cases} 0, & a_k=0 \\ 1, & a_k=2 \end{cases}. \tag{2.57}$$

从而 $A\sim B\sim[0,1]$，A 的势为 $\aleph$，所以 P_0 的势也是 $\aleph$.

注 2.3.7 Cantor 集无内点和孤立点，且是闭集，从而它是疏朗集、完备集.

定义 2.3.7 (Cantor **函数**) 设 $F:[0,1]\mapsto[0,1]$，$[0,1]=G_0\cup P_0$，$\forall x\in G_0$，存在唯一的 $n\geq 1$ 和唯一的 $a=\{a_1,\dots,a_{n-1}\}\in\{0,2\}^{n-1}$. 使

$$x'\in\left(\frac{1}{3^n}\left(\sum_{k=1}^{n-1}a_{n-k}3^k+1\right),\frac{1}{3^n}\left(\sum_{k=1}^{n-1}a_{n-k}3^k+2\right)\right)=:\mathbf{1}_{n,a} \tag{2.58}$$

令

$$F(x')=\begin{cases}0, & x'\leq 0\\ \dfrac{1}{2^n}\left(\sum\limits_{k=1}^{n-1}\dfrac{a_{n-k}}{2}\cdot 2^k+1\right), & x'\in\mathbf{1}_{n,a}\\ 1, & x'\geq 1\end{cases} \tag{2.59}$$

则 $\forall x\in P_0$，称

$$F(x)=\inf_{\substack{x'\in G_0\\ x<x'}}F(x') \tag{2.60}$$

为 Canotr 函数. 它是奇异型分布函数的典例.

第二部分

矩阵代数基础

第三章　向量与线性空间

§3.1　向量及其运算

3.1.1　向量的定义

定义 3.1.1 (**向量**) 由 n 个实数 $a_1, a_2, \ldots, a_n$ 组成的一个数组称为 n 维实数向量，记为

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \tag{3.1}$$

也可记成行向量

$$\boldsymbol{a} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}^\top. \tag{3.2}$$

注 3.1.1 n 维向量在几何上可表示为一个有方向的线段.

注 3.1.2 元素全为零的向量 $\mathbf{0} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}^\top$ 称为零向量.

3.1.2　向量的基本运算

定义 3.1.2 (**加法与数乘**) 向量可以进行加法和数乘运算：

1. 设 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^\top$，$\boldsymbol{b} = (b_1, b_2, \cdots, b_n)^\top$，则向量加法定义为

$$\boldsymbol{a} + \boldsymbol{b} := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}. \tag{3.3}$$

2. 设 c 为一常数，$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^\top$，则向量的数乘定义为

$$c\boldsymbol{a} := \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix} \tag{3.4}$$

3.1.3　向量的线性相关与线性无关

定义 3.1.3 (**线性相关**) 一组 n 维向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$, 如果存在不全为零的常数 $c_1, c_2, \ldots, c_p$ 使

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_p\boldsymbol{a}_p = 0, \tag{3.5}$$

则称这组向量线性相关.

定义 3.1.4 (**线性无关**) 一组 n 维向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$，如果当且仅当 $c_1 = c_2 = \ldots = c_p = 0$ 时，可以使

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_p\boldsymbol{a}_p = 0, \tag{3.6}$$

则称这组向量线性无关.

注 3.1.3 从线性相关与线性无关的定义可以看出

1. 任意含有零向量的向量组总是线性相关的.
2. 线性相关意味着这组向量中至少有一个向量能写成其余向量的线性组合.

§3.2　线性空间与子空间

3.2.1　线性空间的定义

定义 3.2.1 (**线性空间**) 线性空间 S 是向量的一个集合，该集合对向量的加法和数乘封闭，即若 $\boldsymbol{a}, \boldsymbol{b} \in S$ 是两个向量，c 为一常数，则 $\boldsymbol{a} + \boldsymbol{b} \in S$ 且 $c\boldsymbol{a} \in S$.

性质 3.2.1 线性空间中的向量运算满足加法结合律和交换律、数乘结合律和分配律,即若 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in S$ 为三个向量，c_0, c_1, c_2 为常数，则

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= \boldsymbol{b} + \boldsymbol{a}, \\ (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} &= \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}), \\ c_0(\boldsymbol{a} + \boldsymbol{b}) &= c_0\boldsymbol{a} + c_0\boldsymbol{b}, \\ (c_1 + c_2)\boldsymbol{a} &= c_1\boldsymbol{a} + c_2\boldsymbol{a}, \\ c_1(c_2\boldsymbol{a}) &= c_1c_2\boldsymbol{a}, \\ 1 \cdot \boldsymbol{a} &= \boldsymbol{a}. \end{aligned} \tag{3.7}$$

注 3.2.1 记全体 n 维实向量组成的集合为 R_n，它是一个线性空间. 本章的讨论也仅限于 R_n.

3.2.2 子空间

定义 3.2.2 (**子空间**) 给定 R_n 中某些向量 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$，考虑这些向量的所有可能的线性组合所构成的集合

$$\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)=\left\{\sum_{i=1}^{k}c_i\boldsymbol{a}_i : c_1,\dots,c_k\text{均为实数}\right\}. \tag{3.8}$$

则 $\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)$ 对加法和数乘封闭，它是一个线性空间，称其为由 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ 生成的 R_n 的子空间. 若记

$$A_{n\times k}=\begin{pmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\dots & \boldsymbol{a}_k\end{pmatrix},\quad \boldsymbol{c}=\begin{pmatrix}c_1 & c_2 & \dots & c_k\end{pmatrix}^\top \tag{3.9}$$

则 $\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)$ 可重写为

$$\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)=A\boldsymbol{c}. \tag{3.10}$$

称其为由矩阵 A 的列向量张成的子空间，可记为 $\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)=\mathcal{M}(A)$.

3.2.3 基与维数

定义 3.2.3 (**基**) 设 $\mathscr{L}$ 是 R_n 中的一个子空间，如果存在 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ 使 $\mathscr{L}=\mathscr{L}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_k)$，且 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ 线性无关，则称 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ 是 $\mathscr{L}$ 的一组基.

定义 3.2.4 (**维数**) 可以证明子空间 $\mathscr{L}$ 如果有两组基，那么两组基中的向量的个数一定相同，我们把子空间 $\mathscr{L}$ 中的一组基所含向量的个数称为 $\mathscr{L}$ 的维数.

注 3.2.2 记第 i 个位置上的元素为 1 其余元素全为 0 的 n 维向量

$$\boldsymbol{e}_i=\begin{pmatrix}0 & \cdots & 0 & 1 & 0 & \cdots & 0\end{pmatrix},\quad i=1,2,\dots,n \tag{3.11}$$

则它们是线性无关的，且 $R_n=\mathscr{L}(\boldsymbol{e}_1,\dots,\boldsymbol{e}_n)$，从而 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ 是 R_n 的一组基，R_n 的维数是 n. 有时就称其为 n 维线性空间. 由 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ 中任意 m 个向量生成的子空间是 R_n 的 m 维子空间且这些子空间以对应的 m 个向量为基.

注 3.2.3 若 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ 是一组基，则空间上的任意向量都能被这组基线性表出，且表示法唯一.

3.2.4 直和

定义 3.2.5 (**直和**) 设 $\mathscr{L}$ 是一个子空间，如果存在 k 个子空间 $\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_k$，使 $\forall \boldsymbol{a} \in \mathscr{L}$，有

$$\boldsymbol{a} = \boldsymbol{a}_1 + \ldots + \boldsymbol{a}_k. \tag{3.12}$$

则称 $\mathscr{L}$ 是 $\mathscr{L}_1, \ldots, \mathscr{L}_k$ 的直和，记为

$$\mathscr{L} = \mathscr{L}_1 \oplus \mathscr{L}_2 \oplus \cdots \oplus \mathscr{L}_k \tag{3.13}$$

注 3.2.4 沿用上面的记法，则

$$R_n = \mathscr{L}(\boldsymbol{e}_1) \oplus \mathscr{L}(\boldsymbol{e}_2) \oplus \cdots \oplus \mathscr{L}(\boldsymbol{e}_k) \tag{3.14}$$

其中 $\mathscr{L}(\boldsymbol{e}_i), i = 1, 2, \ldots, k$ 为由 $\boldsymbol{e}_i$ 生成的 R_n 上的一维子空间.

§3.3 内积与投影

3.3.1 内积

定义 3.3.1 (**内积**) R_n 中任给两个向量 $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^\top$ 和 $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)^\top$，定义二者的内积为

$$(\boldsymbol{a}, \boldsymbol{b}) := \sum_{i=1}^{n} a_i b_i. \tag{3.15}$$

注 3.3.1 内积满足如下性质：

1. $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$;
2. $(\boldsymbol{a}, \boldsymbol{a}) \geq 0$ 且 $(\boldsymbol{a}, \boldsymbol{a}) = 0 \Leftarrow \boldsymbol{a} = 0$;
3. $(c\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a}, c\boldsymbol{b}) = c(\boldsymbol{a}, \boldsymbol{b})$，$c \in \mathbb{R}$;
4. $(\boldsymbol{a}, \boldsymbol{h} + \boldsymbol{g}) = (\boldsymbol{a}, \boldsymbol{h}) + (\boldsymbol{a}, \boldsymbol{g})$，$(\boldsymbol{h} + \boldsymbol{g}, \boldsymbol{b}) = (\boldsymbol{h} + \boldsymbol{b}) + (\boldsymbol{g} + \boldsymbol{b})$.

定义 3.3.2 (**长度**) 我们把 $(\boldsymbol{a}, \boldsymbol{a})$ 的算术平方根称为 $\boldsymbol{a}$ 的长度，记作 $\|\boldsymbol{a}\|$.

注 3.3.2 长度为 1 的向量称为单位向量，也称为标准化向量.

注 3.3.3 长度满足如下两个不等式：

1. $|(\boldsymbol{a},\boldsymbol{b})|^2 \leq \|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2$；
2. $\|\boldsymbol{a}+\boldsymbol{b}\| \leq \|\boldsymbol{a}\|+\|\boldsymbol{b}\|$.

定义 3.3.3 (**正交**) R_n 中任给两个向量 $\boldsymbol{a}=(a_1,a_2,\dots,a_n)^\top$ 和 $\boldsymbol{b}=(b_1,b_2,\dots,b_n)^\top$，若满足

$$(\boldsymbol{a},\boldsymbol{b})=0, \tag{3.16}$$

则称 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 正交.

定义 3.3.4 (**标准正交基**) 如果 R_n 中的子空间 $\mathscr{L}$ 的基 $\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_k$，满足

1. $(\boldsymbol{a}_i,\boldsymbol{a}_i)=1$，$i=1,2,\dots,k$；
2. $(\boldsymbol{a}_i,\boldsymbol{a}_j)=0$，$i=1,2,\dots,k; j=1,2,\dots,k; i\neq j$；

则称 $\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_k$ 是 $\mathscr{L}$ 的一组标准正交基.

3.3.2 投影

定义 3.3.5 (**投影**) 在 R_n 中，给定一个向量 $\boldsymbol{a}$ 及子空间 $\mathscr{L}$，若 $\mathscr{L}$ 中存在 $\boldsymbol{b}$ 使

$$\|\boldsymbol{a}-\boldsymbol{b}\| = \inf_{\boldsymbol{x}\in\mathscr{L}} \|\boldsymbol{a}-\boldsymbol{x}\|, \tag{3.17}$$

则称 $\boldsymbol{b}$ 是 $\boldsymbol{a}$ 在 $\mathscr{L}$ 上的投影.

注 3.3.4 在 R_n 中，给定两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，则 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影为

$$\boldsymbol{b}^* = \frac{(\boldsymbol{a},\boldsymbol{b})}{(\boldsymbol{b},\boldsymbol{b})}\boldsymbol{b}. \tag{3.18}$$

注 3.3.5 投影是存在且唯一的.

定理 3.3.1 $\boldsymbol{b}$ 是 $\boldsymbol{a}$ 在 $\mathscr{L}$ 上的投影，当且仅当 $(\boldsymbol{a}-\boldsymbol{b},\boldsymbol{x})=0,\forall\boldsymbol{x}\in\mathscr{L}$.

3.3.3 Gram-Schmidt 正交化方法

方法 3.3.1 (Gram-Schmidt **正交化**) 对于 R_n 上的一组线性无关的向量 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$，可以按照如下方法进行正交化（得到一组正交的向量）：

1. 取 $\boldsymbol{b}_1=\boldsymbol{a}_1$；

2. 取 $\boldsymbol{b}_2=\boldsymbol{a}_2-\dfrac{(\boldsymbol{a}_1,\boldsymbol{b}_1)}{(\boldsymbol{b}_1,\boldsymbol{b}_1)}\boldsymbol{b}_1$. 从而 $(\boldsymbol{b}_2,\boldsymbol{b}_1)=0$；

3. 取 $\boldsymbol{b}_3=\boldsymbol{a}_3-\dfrac{(\boldsymbol{a}_3,\boldsymbol{b}_2)}{(\boldsymbol{b}_2,\boldsymbol{b}_2)}\boldsymbol{b}_2-\dfrac{(\boldsymbol{a}_3,\boldsymbol{b}_1)}{(\boldsymbol{b}_1,\boldsymbol{b}_1)}\boldsymbol{b}_1$. 从而 $(\boldsymbol{b}_3,\boldsymbol{b}_2)=(\boldsymbol{b}_3,\boldsymbol{b}_1)=0$；

4. 依此进行下去，可以得到一组新的向量

$$\boldsymbol{b}_i=\boldsymbol{a}_i-\sum_{j=1}^{i-1}\frac{(\boldsymbol{a}_i,\boldsymbol{b}_j)}{(\boldsymbol{b}_j,\boldsymbol{b}_j)}\boldsymbol{b}_j,\quad i=1,2,\ldots,k. \tag{3.19}$$

从而

$$(\boldsymbol{b}_i,\boldsymbol{b}_j)=0,\quad i\neq j;i=1,2,\ldots,k;j=1,2,\ldots,k. \tag{3.20}$$

注 3.3.6 对正交化后的向量组进行标准化，得到标准化的向量组

$$\boldsymbol{\beta}_i=\frac{\boldsymbol{b}_i}{\|\boldsymbol{b}_i\|},\quad i=1,2,\ldots,k. \tag{3.21}$$

则该向量组 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_k$ 构成了子空间 $\mathscr{L}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)$ 上的一组标准正交基.

第四章　矩阵

§4.1　矩阵及一些常见形式

4.1.1　定义

定义 4.1.1 (**矩阵**) 将 $p \times q$ 个实数按照一定顺序排成如下 p 行，q 列的长方形表

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots a_{pq} & \end{pmatrix}. \tag{4.1}$$

则称其为 $p \times q$ 矩阵，常记为 $A = (a_{ij})_{p\times q}$，a_{ij} 表示矩阵中第 i 行、第 j 列的元素.A 的每一行可以看作一个 q 维行向量，称其为矩阵 A 的行向量；A 的每一列可以看作一个 p 维列向量，称其为矩阵 A 列向量.

4.1.2　一些常见的特殊矩阵

在上述定义下，我们有

定义 4.1.2 (**零矩阵**) 若矩阵 A 的所以元素全为零,则称其为零矩阵,记作 $A = O_{p\times q}$,或 $A = O$.

定义 4.1.3 (**方阵**) 若矩阵 A 函数和列数相同，即 $p = q = n$，则称其为 n 阶方阵. 称 $a_{ii}, i = 1, 2, \ldots, n$ 为其主对角线元素，其他元素称为非主对角线元素. 另外，该方阵另外一条对角线上的元素称为副对角线元素.

定义 4.1.4 (**对称矩阵**) 若 n 阶方阵 A 的非主对角线元素关于主对角线对应相等，即 $a_{ij} = a_{ji}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, n; i \neq j$，则称其为（主对角线）对称矩阵.

定义 4.1.5 (**反对称矩阵**) 若 n 阶方阵 A 中的元素关于它的副对角线对应相等，则称其为 n 阶反对称矩阵.

定义 4.1.6 (**上（下）三角矩阵**) 若 n 阶方阵 A 的主对角线下方的元素全为零，即 $a_{ij} = 0$ 且 $i > j$，则称其为上三角矩阵；反之，若方阵的主对角线上方的元素全为零，即 $a_{ij} = 0$ 且 $i < j$，则称其为下三角矩阵.

定义 4.1.7 (**对角矩阵**) 若 n 阶方阵 A 除主对角线元素以外，其余元素全为零，则称其为（主对角线）对角矩阵，有时记为 $A=\operatorname{diag}(a_{11},\dots,a_{nn})$. 此外，若若 n 阶方阵 A 出副对角线元素以外，其余元素全为零，则称其为副对角线对角矩阵.

定义 4.1.8 (**单位矩阵**) 若 n 阶对角矩阵 A 的主对角线元素全为 1，即 $a_{ii}=1, i=1,2,\dots,n$，则称其为 n 单位矩阵，有时记为 $A=I_n$.

§4.2 矩阵的基本运算及其性质

4.2.1 代数运算

方法 4.2.1 (**矩阵加法**) 设 A，B 都是 $p\times q$ 矩阵，则

$$A+B \triangleq (a_{ij}+b_{ij})_{p\times q} \tag{4.2}$$

称为 A 与 B 的和.

方法 4.2.2 (**数乘矩阵**) 设 A 是 $p\times q$ 矩阵，c 为一常数，数乘运算定义如下

$$cA \triangleq (ca_{ij})_{p\times q} \tag{4.3}$$

方法 4.2.3 (**矩阵乘法**) 设 A 是 $p\times q$ 矩阵，B 是 $q\times r$ 矩阵，则

$$AB \triangleq \left(\sum_{k=1}^{q} a_{ik}b_{kj}\right)_{p\times r} \tag{4.4}$$

称为 A 与 B 的积.

注 4.2.1 两个矩阵相加需要注意各相加矩阵需要同形

注 4.2.2 两个矩阵相乘需要注意左侧矩阵的行数要与右侧矩阵的列数相同，故 AB 有意义时，BA 不一定有意义. 并且 AB 和 BA 都有意义时，二者形状也未必相同，即使形状也相同了，也未必有 $AB=BA$.

注 4.2.3 从上述定义中容易得出如下的运算规律

1. $A+B=B+A$;
2. $(A+B)+C=A+(B+C)$;
3. $(c+d)A=cA+dA$;
4. $c(A+B)=cA+cB$;

5. $(AB)C = A(BC)$；

6. $A(B+C) = AB + AC$；

7. $(A+B)C = AC + BC$；

8. $A + O = A$；

9. $I_p A_{p\times q} = A_{p\times q} I_q = A_{p\times q}$.

定义 4.2.1 (**幂等矩阵**) 若方阵 A 满足 $A^2 = AA = A$，则乘 A 为幂等矩阵.

定义 4.2.2 (**投影矩阵**) 若方阵 A 是对称的幂等矩阵，则称 A 为投影矩阵.

4.2.2 行列式

定义 4.2.3 (**逆序数**) 给定一组排列 $i_1, i_2, \ldots, i_n$，排列中一对数的前后位置与自然大小顺序相反的情况（即排列中前面的数大于后面的）为逆序，称该组排列中逆序的总数称为这个排列的逆序数，记为 $\tau(i_1, i_2, \ldots, i_n)$. 若逆序数为奇数，则称该排列为奇排列，若逆序数为偶数，则称该排列为偶排列.

定义 4.2.4 (**行列式**) n 阶行列式是一个方形数表，满足如下运算规律

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{p2} & \cdots a_{nn} & \end{vmatrix} = \sum_{j_1, j_2, \ldots, j_p} (-1)^{\tau(j_1 j_2 \ldots j_p)} a_{1j_1} a_{2j_2} \cdots a_{pj_p}. \tag{4.5}$$

$\sum\limits_{j_1, j_2, \ldots, j_p}$ 表示对 $1, 2, \ldots, p$ 的所以排列求和.

注 4.2.4 低阶行列式运算规则如下

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{4.6}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11}. \tag{4.7}$$

注 4.2.5 可以对方阵作用行列式运算，所谓矩阵 A 的行列式是指，将矩阵 A 的各元素放入行列式中，按上述运算法则进行计算. 由于行列式是方形数表，故只有方阵可以计算行列式，实际上任何一个行列式都可以简记为 $|A|$，A 为以行列式中元素为元素的方阵.

定义 4.2.5 (**余子式与代数余子式**) 设 A 为 n 阶方阵，将其元素 a_{ij} 所在的第 i 行与第 j 列划去所得 $n-1$ 阶矩阵的行列式，称为元素 a_{ij} 的子式，记为 M_{ij}. 称

$$A_{ij} = (-1)^{i+j} M_{ij} \tag{4.8}$$

为元素 a_{ij} 的代数余子式.

注 4.2.6 行列式运算满足如下规律

1. 若方阵 A 的某行（或列）为零，则 $|A| = 0$;
2. 若方阵 A 的某一行（或列）是其他一些行（列）的线性组合（这蕴含了某一行（或列）是另一行（或列）的倍数），则 $|A| = 0$;
3. 若将 A 某一行（或）列乘以常数 c，则所得矩阵的行列式为 $c|A|$;
4. A 是一个 n 阶方阵，c 为一常数，则 $|cA| = c^p|A|$;
5. 任意互换 A 的两行（或列），则行列式变号;
6. 若将 A 的某一行（或列）的倍数加到另一行（或列），则所得行列式不变;
7. 若 n 阶方阵为上三角矩阵、下三角矩阵或对角矩阵，则 $|A| = \prod\limits_{i=1}^{n} a_{ii}$;
8. 若 AB 都是 n 方阵，则 $|AB| = |A||B|$;
9. A 为 n 阶方阵，则 $|A| = \sum\limits_{j=1}^{n} a_{ij}A_{ij} = \sum\limits_{i=1}^{n} a_{ij}A_{ij}$;
10. A 为 n 阶方阵，则 $\sum\limits_{j=1}^{n} a_{kj}A_{ij}(k \neq i)$;
11. A 为 n 阶方阵，则 $\sum\limits_{i=1}^{n} a_{ik}A_{ij}(k \neq j)$.

4.2.3 转置

定义 4.2.6 (**矩阵转置**) A 为 $p \times q$ 的矩阵，把 A 的行改成列，列改成行后得到的 $q \times p$ 矩阵称为矩阵 A 的转置，一般记为 $A^\top$（有些教材采用 A' 的形式，本文均采用“⊤”的记法）. 即若

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}, \tag{4.9}$$

则

$$A^\top = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{p1} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & \vdots & & \vdots \\ a_{1q} & a_{2q} & \cdots & a_{pq} \end{pmatrix}. \tag{4.10}$$

注 4.2.7 矩阵转置满足如下运算规律

1. $(A^\top)^\top = A$;
2. $(A+B)^\top = A^\top + B^\top$;
3. $(AB)^\top = B^\top A^\top$;
4. $|A^\top| = |A|$.

注 4.2.8 一个 n 维列向量可以看作一个 $n \times 1$ 的矩阵，故转置运算亦可作用于向量. 在本部分的介绍中，如无特殊说明，我们谈及向量时，均默认其为 n 维列向量，若要表示行向量，则一般加转置.

注 4.2.9 设两个 n 维向量分别为

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{4.11}$$

则有

$$\boldsymbol{a}^\top \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = (\boldsymbol{a}, \boldsymbol{b}). \tag{4.12}$$

$$\boldsymbol{a}\boldsymbol{b}^\top = \begin{pmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & & \vdots \\ a_nb_1 & a_nb_2 & \cdots & a_nb_n \end{pmatrix}. \tag{4.13}$$

定义 4.2.7 (**正交矩阵**) 若方阵 A 满足 $AA^\top = I$，则称 A 为正交矩阵.

注 4.2.10 实际上，引入矩阵行列向量以及转置符号，$p\times q$ 的矩阵 A 和 $q\times r$ 的矩阵 B 可写成

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}=\begin{pmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_p^\top \end{pmatrix}, \tag{4.14}$$

$$B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & & \vdots \\ b_{q1} & b_{q2} & \cdots & b_{qr} \end{pmatrix}=\begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_r \end{pmatrix}. \tag{4.15}$$

则有

$$AB=\begin{pmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_p^\top \end{pmatrix}\begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_r \end{pmatrix}=\begin{pmatrix} \boldsymbol{a}_1^\top\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_1^\top\boldsymbol{b}_r \\ \vdots & & \vdots \\ \boldsymbol{a}_p^\top\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_p^\top\boldsymbol{b}_r \end{pmatrix} \tag{4.16}$$

故对 n 阶方阵 A 有

$$AA^\top=(\boldsymbol{a}_i^\top\boldsymbol{a}_j)_{n\times n}. \tag{4.17}$$

$AA^\top=I$ 意味着

$$\boldsymbol{a}_i^\top\boldsymbol{a}_j=(\boldsymbol{a}_i,\boldsymbol{a}_j)=\begin{cases} 1 & ,i=j \\ 0 & ,i\neq j \end{cases},\quad i=1,2,\ldots,n;j=1,2,\ldots,n. \tag{4.18}$$

即正交矩阵 A 的行（列）向量均为单位向量，且行（列）向量之间两两正交.

注 4.2.11 如下构造的 n 阶方阵是一个正交矩阵

$$\begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\cdot 1}} & \frac{-1}{\sqrt{2\cdot 1}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & \frac{-2}{\sqrt{3\cdot 2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix} \tag{4.19}$$

矩阵中的第一行可以放在最后.

4.2.4 逆

定义 4.2.8 (**矩阵的逆**) 如果 A 是方阵，且存在与 A 同形的方阵 B 使

$$AB = BA = I, \tag{4.20}$$

则称 A 是可逆的，或非退化的、非奇异的（若 A 不可逆，则称其维奇异矩阵），此时称 B 是 A 的逆矩阵，记作 A^{-1}.

注 4.2.12 给定方阵 A，若它可逆，则它的逆存在且唯一.

注 4.2.13 矩阵求逆满足如下运算规律

1. $(A^{-1})^{-1} = A$;
2. $AA^{-1} = A^{-1}A = I$;
3. $(AB)^{-1} = B^{-1}A^{-1}$;
4. $(A^\top)^{-1} = (A^{-1})^\top$;
5. $|A^{-1}| = |A|^{-1}$;
6. 若 $A = \mathrm{diag}(a_{11}, \dots, a_{pp})$ 可逆，则
$$A^{-1} = \mathrm{diag}(a_{11}^{-1}, \dots, a_{pp}^{-1}). \tag{4.21}$$
7. 若 A 为正交矩阵，则 $A^{-1} = A^\top$.

4.2.5 秩

定义 4.2.9 (**矩阵的秩**) 设 A 为 $p \times q$ 矩阵，若存在 A 的一个 r 阶子方阵的行列式不为零，而 A 的一切 $r+1$ 阶子方阵的行列式全为零，则称 A 的秩为 r，记作 $\mathrm{rank}(A) = r$. 若 $\mathrm{rank}(A) = p$ 则称其为行满秩，若 $\mathrm{rank}(A) = q$ 则称其为列满秩，若 n 阶方阵的秩为 n，称其为满秩的.

注 4.2.14 矩阵的秩具有下述基本性质

1. $\mathrm{rank}(A) = 0$，当且仅当 $A = O_{p\times q}$;
2. 若 A 为 $p \times q$ 矩阵，且 $A \neq O_{p\times q}$，则
$$1 \le \mathrm{rank}(A) \le \min\{p, q\}. \tag{4.22}$$

3. $\text{rank}(A) = \text{rank}(A^\top)$；

4. 若 A 为 $p \times q$ 矩阵，B 为 $p \times r$ 矩阵，则

$$\max\{\text{rank}(A), \text{rank}(B)\} \leq \text{rank}(AB) \leq \min\{p, \text{rank}(A) + \text{rank}(B)\}. \tag{4.23}$$

5. $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$；

6. $\text{rank}(A+B) \leq \min\{\text{rank}(A) + \text{rank}(B)\}$；

7. 若 A 和 C 为非退化方阵，则 $\text{rank}(ABC) = \text{rank}(B)$；

8. n 阶方阵 A 是非退化的，当且仅当 $\text{rank}(A) = n$.

4.2.6　迹

定义 4.2.10 (**迹**) 设 A 为 n 阶方阵，则它的对角线元素之和称为 A 的迹，记作 $\text{tr}(A)$，即

$$\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} \tag{4.24}$$

注 4.2.15 矩阵的迹具有下述基本性质

1. $\text{tr}(AB) = \text{tr}(BA)$；

2. $\text{tr}(A) = \text{tr}(A^\top)$；

3. $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$；

4. $\text{tr}\left(\sum\limits_{\alpha=1}^{k} A_\alpha\right) = \sum\limits_{\alpha=1}^{k} \text{tr}(A_\alpha)$；

5. 若 A 为投影矩阵，则 $\text{tr}(A) = \text{rank}(A)$.

§4.3 分块矩阵

§4.4 特征值与特征向量

§4.5 对称矩阵、正定矩阵

§4.6 矩阵微商

§4.7 广义逆

§4.8 拉直运算与 Kronecker 积

4.8.1 拉直运算

定义 4.8.1 (**拉直运算**) 所谓拉直运算，也称为向量化运算，是将矩阵拉成一个长向量，并通过这个长向量来建立矩阵与向量之间联系的一种运算. 即若设随机矩阵 X 是一个 $n \times p$ 矩阵：

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \triangleq \begin{pmatrix} X_{(2)}^\top \\ \vdots \\ X_{(n)}^\top \end{pmatrix} \triangleq \begin{pmatrix} X_1 & X_2 & \cdots & X_p \end{pmatrix}. \tag{4.25}$$

用 X 的列向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ 组成一个 np 维向量，记为

$$\mathrm{Vec}(X) = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{n1} & x_{12} & \cdots & x_{n2} & \cdots & x_{1p} & \cdots & x_{np} \end{pmatrix}^\top \tag{4.26}$$

则称此运算为拉直运算.

注 4.8.1 用 X 的行向量 $X_{(1)}, \ldots, X_{(n)}$ 也可以进行拉直运算：

$$\mathrm{Vec}(X^\top) = \begin{pmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(p)} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} & x_{21} & \cdots & x_{2p} & \cdots & x_{n1} & \cdots & x_{np} \end{pmatrix}^\top \tag{4.27}$$

注 4.8.2 矩阵转置具有如下运算规律

1. $(A+B)^\top = A^\top + B^\top$；

2. $(AB)^\top = B^\top A^\top$.

设 S 是 p 维对称矩阵，则其中只包含 $\dfrac{p(p+1)}{2}$ 个不同的随机变量. 将其拉直为 p^2 维向量并不合适，应拉直为 $\dfrac{p(p+1)}{2}$ 维向量.

定义 4.8.2 (**对称矩阵的拉直运算**) 设 $S=(s_{ij})_{p\times p}$ 为 p 维对称矩阵，令

$$\operatorname{Svec}(S)=\begin{pmatrix} s_{11} & \cdots & s_{p1} & s_{22} & \cdots & s_{p2} & \cdots & s_{pp}\end{pmatrix}^\top \tag{4.28}$$

为 $\dfrac{p(p+1)}{2}$ 维向量. 称该运算为对称矩阵的拉直运算.

4.8.2 Kronecker 积

定义 4.8.3 (Kronecker **积**) 设 $A=(a_{ij})$ 和 B 分别为 $n\times p$ 和 $m\times q$ 的矩阵，A 和 B 的 Kronecker 积 $A\otimes B$ 定义为

$$A\otimes B=(a_{ij}B)=\begin{pmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{np}B\end{pmatrix}. \tag{4.29}$$

这是一个 $mn\times pq$ 矩阵.Kronecker 积有时也称为矩阵的直积或叉积.

第三部分

泛函分析基础

第四部分

高等概率论

第五章　集类

§5.1　基本术语

定义 5.1.1 (**集类**) 所谓集类，就是由集合构成的集合（类）. 在概率论中一般谈论集类是指全集 Ω 上的集类，它是 Ω 的一些子集构成的类.

定义 5.1.2 (**有限交封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A_1, A_2, \ldots, A_n \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{n} A_i, \tag{5.1}$$

则称 $\mathcal{A}$ 对有限交封闭.

定义 5.1.3 (**有限并封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A_1, A_2, \ldots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{n} A_i, \tag{5.2}$$

则称 $\mathcal{A}$ 对有限并封闭.

注 5.1.1 对有限运算封闭的定义有时也可以写作只保留两项的特殊形式，如对有限交封闭可写作：$A, B \in \mathcal{A} \in \mathcal{A} \Rightarrow A \cup B$.

定义 5.1.4 (**可列交封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i, \tag{5.3}$$

则称 $\mathcal{A}$ 对可列交封闭.

定义 5.1.5 (**可列并封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i, \tag{5.4}$$

则称 $\mathcal{A}$ 对可列并封闭.

定义 5.1.6 (**差封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A, B \in \mathcal{A} \Rightarrow A \backslash B \in \mathcal{A}, \tag{5.5}$$

则称 $\mathcal{A}$ 对差封闭.

定义 5.1.7 (**补封闭**) 设 $\mathcal{A}$ 是 Ω 上的一个集类. 若

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \tag{5.6}$$

则称 $\mathcal{A}$ 对补封闭.

§5.2 常见集类

5.2.1 半环

定义 5.2.1 (**半环**) $\mathcal{E}$ 是 Ω 上的一个集类，如果满足

1. $\varnothing \in \mathcal{E}$；
2. $\mathcal{E}$ 对有限交封闭；
3. $\mathcal{E}$ 中任两元素之差都能表示成 $\mathcal{E}$ 中元素的有限不交并，即

$$A, B \in \mathcal{E} \Rightarrow A \backslash B = \biguplus_{i=1}^{n} C_i, \quad C_i \in \mathcal{E}, i = 1, 2, \ldots, n. \tag{5.7}$$

则称 $\mathcal{E}$ 是 Ω 上的一个半环.

注 5.2.1 半环定义的第三点要求有另一种表述如下：$\mathcal{E}$ 中任意两个有包含关系的元素之差都能表示成 $\mathcal{E}$ 中元素的有限不交并，即若 $A, A_1 \in \mathcal{E}$，$A_1 \subset A$，则 $\exists \{A_2, \ldots, A_n\} \in \mathcal{E}$，$A_1, \ldots, A_n$ 两两互不相交，使得

$$A = \bigcup_{k=1}^{n} A_k. \tag{5.8}$$

性质 5.2.1 设 $A_1, \ldots, A_n, B_1, \ldots, B_m$ 都是半环 $\mathcal{E}$ 中的元素，则

$$\bigcup_{i=1}^{n} A_i, \quad \left(\bigcup_{i=1}^{n} A_i\right) \Bigg\backslash \left(\bigcup_{j=1}^{m} B_j\right) \tag{5.9}$$

都可以表示成 $\mathcal{E}$ 的有限不交并.

5.2.2 半（集）代数

定义 5.2.2 (**半集代数**) $\mathscr{S}$ 是 Ω 上的一个集类，如果满足

1. $\varnothing \in \mathscr{S}$，$\Omega \in \mathscr{S}$；

2. $\mathscr{S}$ 对有限交封闭，即若 $A, B \in \mathscr{A}$，则 $A \cap B \in \mathscr{A}$；

3. $\mathscr{S}$ 中任两元素之差都能表示成 $\mathscr{S}$ 中元素的有限不交并，即

$$A, B \in \mathscr{S} \Rightarrow A \backslash B = \biguplus_{i=1}^{n} C_i, \quad C_i \in \mathscr{S}, i = 1, 2, \ldots, n. \tag{5.10}$$

则称 $\mathscr{S}$ 为 Ω 上的一个半集代数，有些资料也直接称其为半代数.

注 5.2.2 半集代数定义的第三点要求有另一种表述如下：$\mathscr{S}$ 中任意两个有包含关系的元素之差都能表示成 $\mathscr{S}$ 中元素的有限不交并，即若 $A, A_1 \in \mathscr{S}$，$A_1 \subset A$，则 $\exists \{A_2, \ldots, A_n\} \in \mathscr{S}$，$A_1, \ldots, A_n$ 两两互不相交，使得

$$A = \bigcup_{k=1}^{n} A_k. \tag{5.11}$$

注 5.2.3 注意到半集代数和半环的定义基本一致. 若一个集类 $\mathscr{S}$ 为半环，且满足 $\Omega \in \mathscr{S}$，那么这个半环就是 $\mathscr{S}$ 在 Ω 上的一个半集代数.

注 5.2.4 半集代数的交不一定是半集代数.

引理 5.2.1 (**集类是半集代数的充要条件**) Ω 的子集类 $\mathscr{S}$ 是半集代数的充分必要条件是

1. $\varnothing \in \mathscr{S}$，$\Omega \in \mathscr{S}$；

2. $\mathscr{S}$ 对有限交封闭，即若 $A, B \in \mathscr{A}$，则 $A \cap B \in \mathscr{A}$；

3. $A \in \mathscr{S}$，则 $\exists \{A_1, \ldots, A_n\} \subset \mathscr{S}$ 两两不交，使得

$$A^c = \Omega \backslash A = \bigcup_{k=1}^{n} A_k. \tag{5.12}$$

注 5.2.5 需要注意，$\mathscr{S}$ 是 Ω 的某些子集构成的集类，$A \in \mathscr{S}$ 时，A^c 未必属于 $\mathscr{S}$.

5.2.3 （集）代数

定义 5.2.3 (**集代数**) $\mathscr{A}$ 是 Ω 上的一个集类，如果满足

1. $\Omega \in \mathscr{A}$；
2. $\mathscr{A}$ 对有限交和有限并都封闭，即若 $A, B \in \mathscr{A}$，则 $A \cap B \in \mathscr{A}$，$A \cup B \in \mathscr{A}$；
3. $\mathscr{A}$ 对补封闭，即若 $A \in \mathscr{A}$，则 $A^c = \Omega \backslash A \in \mathscr{A}$.

则称 $\mathscr{A}$ 为 Ω 上的一个集代数，或布尔代数，有些资料也直接称其为代数.

注 5.2.6 在满足对补封闭的要求后，根据 De Morgen 法则，集代数定义第二点要求中的有限交封闭和有限并封闭只需满足其一，则另一个亦可成立.

注 5.2.7 集代数一定是半集代数.

注 5.2.8 任意多集代数的交一定是集代数.

引理 5.2.2 (**集类是集代数的充要条件**) Ω 的子集类 $\mathscr{A}$ 是集代数的充分必要条件是如下三组条件之一成立：

1. 满足
   (a) $\Omega \in \mathscr{A}$；
   (b) $\mathscr{A}$ 对有限交封闭，即若 $A, B \in \mathscr{A}$，则 $A \cap B \in \mathscr{A}$；
   (c) $\mathscr{A}$ 对补封闭，即若 $A \in \mathscr{A}$，则 $A^c = \Omega \backslash A \in \mathscr{A}$.
2. 满足
   (a) $\Omega \in \mathscr{A}$；
   (b) $\mathscr{A}$ 对有限并封闭，即若 $A, B \in \mathscr{A}$，则 $A \cup B \in \mathscr{A}$；
   (c) $\mathscr{A}$ 对补封闭，即若 $A \in \mathscr{A}$，则 $A^c = \Omega \backslash A \in \mathscr{A}$.
3. 满足
   (a) $\Omega \in \mathscr{A}$；
   (b) $\mathscr{A}$ 对差封闭，即若 $A, B \in \mathscr{A}$，则 $A \backslash B \in \mathscr{A}$.

引理 5.2.3 (**由半集代数生成的集代数**) 设 $\mathscr{S}$ 是 Ω 的半集代数，则

$$\begin{aligned}\mathscr{A} &:= \left\{\bigcup_{k=1}^{n} A_k : A_1,\ldots,A_n \in \mathscr{S}\text{两两不交}, n\in\mathbb{Z}_+\right\} \\ &= \left\{\bigcup_{k=1}^{n} A_k : A_1,\ldots,A_n \in \mathscr{S}, n\in\mathbb{Z}_+\right\}\end{aligned} \tag{5.13}$$

是包含 $\mathscr{S}$ 的最小的集代数（Ω 的任一包含了 $\mathscr{S}$ 的集代数必然包含按上述定义的 $\mathscr{A}$），称之为由 $\mathscr{S}$ 生成的集代数，有时记作 $\mathscr{A}(\mathscr{S})$.

5.2.4 单调类

定义 5.2.4 (**单调类**) 设 $\mathcal{M}\subset\mathcal{P}(\Omega)$，若 $\mathcal{M}$ 对单调序列的极限封闭，即 $\mathcal{M}$ 即对单调上升的极限封闭，又对单调下降的极限封闭，则称其为单调类.

5.2.5 σ 代数

定义 5.2.5 (**σ 代数**) $\mathscr{F}$ 是 Ω 上的一个集类，如果满足

1. $\Omega\in\mathscr{F}$；
2. $\mathscr{F}$ 对补封闭，即若 $A\in\mathscr{F}$，则 $A^c\in\mathscr{F}$；
3. $\mathscr{F}$ 对可列并封闭，即若 $A_n\in\mathscr{F}, n\in\mathbb{Z}_+$，则 $\bigcup\limits_{n=1}^{\infty} A_n\in\mathscr{F}$.

注 5.2.9 σ 代数一定是集代数.

引理 5.2.4 (**集类是 σ 代数的充分必要条件**) Ω 上的一个集类 $\mathscr{F}$ 是 σ 代数的充分必要条件是成立：

1. $\Omega\in\mathscr{F}$；
2. $\mathscr{F}$ 对补封闭，即若 $A\in\mathscr{F}$，则 $A^c\in\mathscr{F}$；
3. $\mathscr{F}$ 对可列交封闭，即若 $A_n\in\mathscr{F}, n\in\mathbb{Z}_+$，则 $\bigcap\limits_{n=1}^{\infty} A_n\in\mathscr{F}$.

注 5.2.10 任意 σ 代数的交一定是 σ 代数.

定义 5.2.6 (**平凡 σ 代数和离散 σ 代数**) $\{\varnothing,\Omega\}$ 是 Ω 上最小的 σ 代数，称其为平凡 σ 代数；$\mathcal{P}(\Omega)$ 是 Ω 上最大的 σ 代数，称其为离散 σ 代数.

定义 5.2.7 (**子 σ 代数**) Ω 上的两个 σ 代数 $\mathscr{F}_1$ 和 $\mathscr{F}_2$，满足 $\mathscr{F}_1 \subset \mathscr{F}_2$，则称 $\mathscr{F}_1$ 是 $\mathscr{F}_2$ 的子 σ 代数.

注 5.2.11 平凡 σ 代数之所以是“最小”的，是因为任何 Ω 上的 σ 代数都包含它，即平凡 σ 代数是 Ω 上所有 σ 代数的子代数.

注 5.2.12 离散 σ 代数之所以是“最大”的，是因为任何 Ω 上的 σ 代数都包含于它，即 Ω 上所有 σ 代数都是离散 σ 代数的子代数.

定理 5.2.1 若 $\mathscr{F}$ 是集代数，则下列各条等价

1. $\mathscr{F}$ 是 σ 代数；
2. $\mathscr{F}$ 对单调下降序列的极限封闭，即 $A_n \in \mathscr{F}, n \geq 1$，$A_n \downarrow$，则 $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{F}$；
3. $\mathscr{F}$ 对单调上升序列的极限封闭，即 $A_n \in \mathscr{F}, n \geq 1$，$A_n \uparrow$，则 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$；
4. $\mathscr{F}$ 对可列不交并封闭.

注 5.2.13 集类是 σ 代数的充分必要条件是它是代数且是单调类.

5.2.6　π 系与 λ 系

定义 5.2.8 (π **系**) Π 是 Ω 上的一个集类，若它对有限交封闭，则称它是 π 系，也可称为 π 类.

定义 5.2.9 (λ **系**) Λ 是 Ω 上的一个集类，如果满足

1. $\Omega \in \Lambda$；
2. Λ 对真差封闭，即 $A, B \in \Lambda$，$A \subset B$，$B \backslash A \in \Lambda$；
3. Λ 对不降序列的极限封闭，即 $\{A_n, n \in \mathbb{Z}_+\} \subset \Lambda$，$A_n \uparrow$，则 $\bigcup\limits_{n=1}^{\infty} A_n \in \Lambda$.

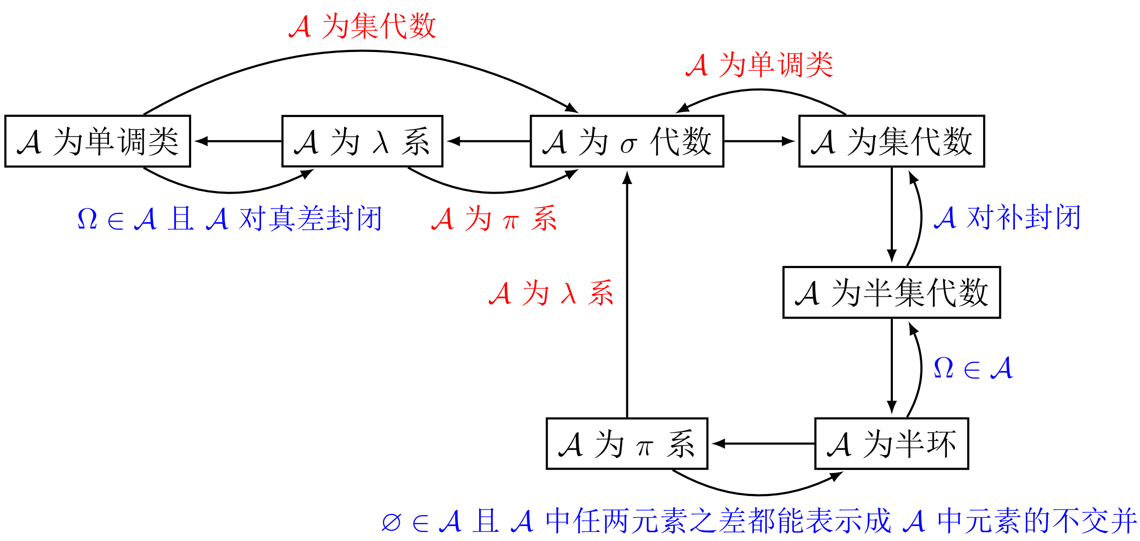
则称其为 λ 系，或称 Dynkin 类，也可称为 λ 类.

注 5.2.14 λ 系是 Dynkin 对单调类的推广.

注 5.2.15 集类是 σ 代数的充分必要条件是它是 π 系且是 λ 系.

5.2.7　常见集类的关系

Ω 的子集类 $\mathcal{A}$ 为半环、半集代数、集代数、单调类、σ 代数、π 系和 λ 系关系如下：



§5.3　单调类定理与 π-λ 定理

5.3.1　生成元

引理 5.3.1 若一族集类 $\{\mathcal{A}_t, t \in T\}$ 中的每个成员 $\mathcal{A}_t$ 都对某种集合运算封闭，则集类 $\bigcap\limits_{t \in T} \mathcal{A}_t$ 也对该运算封闭.

注 5.3.1 Ω 上任意一族 σ 代数、代数、λ 系、单调类的交仍然分别为 σ 代数、代数、λ 系、单调类.

定理 5.3.1 (**由集类生成的集类**) 设 $\mathcal{C}$ 是 Ω 的一个子集类，则分别存在一个唯一的 Ω 的 σ 代数、λ 系和单调类

$$\sigma(\mathcal{C}) = \cap\{\mathscr{F} : \mathscr{F}\text{是 } \sigma \text{ 代数且} \mathscr{F} \supset \mathcal{C}\}, \tag{5.14}$$

$$\lambda(\mathcal{C}) = \cap\{\Lambda : \Lambda\text{是 } \lambda \text{ 系且} \Lambda \supset \mathcal{C}\} \tag{5.15}$$

和

$$M(\mathcal{C}) = \cap\{\mathcal{M} : \mathcal{M}\text{是单调类且} \mathcal{M} \supset \mathcal{C}\}. \tag{5.16}$$

各自满足如下两点：

1. $\sigma(\mathcal{C}) \supset \mathcal{C}$（对应地，$\lambda(\mathcal{C}) \supset \mathcal{C}$ 或 $M(\mathcal{C}) \supset \mathcal{C}$）；

2. 若 $\mathcal{B}$ 也是 $\mathcal{C}$ 的一个 σ 代数（对应地，λ 系或单调类），则 $\sigma(\mathcal{C})\subset\mathcal{B}$（对应地，$\lambda(\mathcal{C})\subset\mathcal{B}$ 或 $M(\mathcal{C})\subset\mathcal{B}$）.

称其为由 $\mathcal{C}$ 生成的 σ 代数（对应地，λ 系或单调类）或包含 $\mathcal{C}$ 的最小的 σ 代数（对应地，λ 系或单调类）. 另称 $\mathcal{C}$ 为 $\sigma(\mathcal{C})$（对应地，$\lambda(\mathcal{C})$ 或 $M(\mathcal{C})$）的生成元.

注 5.3.2 对任意非空集类 $\mathcal{C}$，有 $M(\mathcal{C})\subset\sigma(\mathcal{C})$ 和 $\lambda(\mathcal{C})\subset\sigma(\mathcal{C})$.

定义 5.3.1 (Borel σ **代数**) 设 $(X,\mathcal{T})$ 为拓扑空间，则称开集族 $\mathcal{T}$ 生成的 σ 代数 $\sigma(\mathcal{T})$ 为 X 上的 Borel σ 代数. 记为 $\mathscr{B}(X)$，其中的每个集合被称为 Borel 集.

注 5.3.3 开集族是 Borel σ 代数的生成元，闭集族、基和子基亦是其生成元.

5.3.2　单调类定理

定理 5.3.2 (**单调类定理**) 若 $\mathcal{C}$ 为集代数，则 $M(\mathcal{C})=\sigma(\mathcal{C})$，$M(\mathcal{C})$ 和 $\sigma(\mathcal{C})$ 分别按（5.16）和（5.14）定义. 即由集代数生成的单调类和 σ 代数相等.

定理 5.3.3 (**集合的单调类方法**) 设 $\mathcal{C},\mathcal{A}$ 为两个集类，$\mathcal{C}\subset\mathcal{A}$. 若 $\mathcal{C}$ 为集代数，$\mathcal{A}$ 为单调类，则 $\sigma(\mathcal{C})\subset\mathcal{A}$. 即任何包含 $\mathcal{C}$ 的单调类包含由集代数 $\mathcal{C}$ 生成的 σ 代数.

5.3.3　π-λ 定理

定理 5.3.4 (π-λ **定理**) 若 $\mathcal{C}$ 为 π 系，则 $\lambda(\mathcal{C})=\sigma(\mathcal{C})$，$\lambda(\mathcal{C})$ 和 $\sigma(\mathcal{C})$ 分别按（5.15）和（5.14）定义. 即由 π 系生成的 λ 系和 σ 代数相等.

定理 5.3.5 (**集合的** π-λ **方法**) 设 $\mathcal{C},\mathcal{A}$ 为两个集类，$\mathcal{C}\subset\mathcal{A}$. 若 $\mathcal{C}$ 为 π 系，$\mathcal{A}$ 为 λ 系，则 $\sigma(\mathcal{C})\subset\mathcal{A}$. 即任何包含 $\mathcal{C}$ 的 λ 系包含由 π 系 $\mathcal{C}$ 生成的 σ 代数.

§5.4　与 $\mathbb{R}$ 相关的 Borel σ 代数的结构

5.4.1　$\mathbb{R}^n$ 上的 Borel σ 代数

定义 5.4.1 ($\mathbb{R}^n$ **上的** Borel σ **代数**) $\mathbb{R}^n$ 上的 Borel σ 代数（记为 $\mathscr{B}(\mathbb{R}^n)$）是包含了 $\mathbb{R}^n$ 的所有开集的最小的 σ 代数，以 $\mathcal{E}_1$ 表示 $\mathbb{R}^n$ 中的开集族，则有

$$\mathscr{B}(\mathbb{R}^n)=\mathscr{B}(\mathcal{E}_1) \tag{5.17}$$

即 $\mathcal{E}_1$ 是 $\mathscr{B}(\mathbb{R}^n)$ 的生成元.

定理 5.4.1 下列诸 $\mathcal{E}_i$ 均为 $\mathscr{B}(\mathbb{R}^n)$ 的生成元

$$
\begin{aligned}
\mathcal{E}_2 &= \{B(\boldsymbol{x}, r) : \boldsymbol{x} \in \mathbb{R}^n, r > 0\},\\
\mathcal{E}_3 &= \{(\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \le \boldsymbol{b}\},\\
\mathcal{E}_4 &= \{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \le \boldsymbol{b}\},\\
\mathcal{E}_5 &= \{[\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \le \boldsymbol{b}\},\\
\mathcal{E}_6 &= \{[\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \le \boldsymbol{b}\},\\
\mathcal{E}_7 &= \{(-\infty, \boldsymbol{b}] : \boldsymbol{b} \in \mathbb{R}^n\},\\
\mathcal{E}_8 &= \{(-\infty, \boldsymbol{b}) : \boldsymbol{b} \in \mathbb{R}^n\},\\
\mathcal{E}_9 &= \{[\boldsymbol{a}, \infty) : \boldsymbol{a} \in \mathbb{R}^n\},\\
\mathcal{E}_{10} &= \{(\boldsymbol{a}, \infty) : \boldsymbol{a} \in \mathbb{R}^n\},\\
\mathcal{E}_{11} &= \{(-\infty, \boldsymbol{r}] : \boldsymbol{r} \in \mathbb{Q}^n\},\\
\mathcal{E}_{12} &= \{(-\infty, \boldsymbol{r}) : \boldsymbol{r} \in \mathbb{Q}^n\},\\
\mathcal{E}_{13} &= \{[\boldsymbol{r}, \infty) : \boldsymbol{r} \in \mathbb{Q}^n\},\\
\mathcal{E}_{14} &= \{(\boldsymbol{r}, \infty) : \boldsymbol{r} \in \mathbb{Q}^n\}.
\end{aligned}
\tag{5.18}
$$

其中 $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$，$\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$，$\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathbb{Q}^n$.

5.4.2 广义实数空间 $\overline{\mathbb{R}}$ 上的 Borel σ 代数

定义 5.4.2 (**广义实数空间**) 一般实数空间 $\mathbb{R}$ 不包含 $-\infty$ 和 ∞，而一些函数的取值可能会取到这两个值（广义实值函数），因此有必要将这二者纳入进来，故定义 $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\} = [-\infty, \infty]$，称之为广义实数空间. 称 $-\infty$，∞ 为广义实数. 进一步可以定义广义 n 维欧氏空间 $\overline{\mathbb{R}}^n := \{x : x = (x_1, \ldots, x_n), x_i \in \overline{\mathbb{R}}, i = 1, 2, \ldots, n\}$.

注 5.4.1 尽管 $-\infty$，∞ 本身都不是实数，作为规定，它们与实数 x 之间存在如下序关系及代数运算.

- 序：$-\infty < x < \infty$；
- 加法：$(\pm\infty) + (\pm\infty) = x + (\pm\infty) = (\pm\infty + x) = \pm\infty$；
- 减法：$x - (\mp\infty) = (\pm\infty) - x = (\pm\infty) - (\mp\infty) = \pm\infty$；
- 乘法：$(\pm\infty)\cdot(\pm\infty) = +\infty$，$(\pm\infty)\cdot(\mp\infty) = -\infty$，$x\cdot(\pm\infty) = (\pm\infty)\cdot x = \begin{cases} \pm\infty, & x > 0, \\ 0, & x = 0, \\ \mp\infty, & x < 0; \end{cases}$
- 除法：$\dfrac{x}{\pm\infty} = 0$，$\begin{cases} \pm\infty, & x > 0 \\ \mp\infty, & x < 0; \end{cases}$

- 绝对值：$|\pm\infty| = +\infty$.

注 5.4.2 $(\pm\infty)+(\mp\infty)$，$(\pm\infty)-(\pm\infty)$，$\dfrac{\pm\infty}{0}$，$\dfrac{x}{0}$，$\dfrac{\pm\infty}{\pm\infty}$，$\dfrac{\pm\infty}{\mp\infty}$ 都没有意义.

定义 5.4.3 ($\overline{\mathbb{R}}$ **上的** Borel σ **代数**) $\overline{\mathbb{R}}$ 上的 Borel σ 代数为

$$\mathscr{B}(\overline{\mathbb{R}}) = \{B \cup C : B \in \mathscr{B}(\mathbb{R}), C \subset \{-\infty, \infty\}\}. \tag{5.19}$$

即 $\mathscr{B}(\overline{\mathbb{R}})$ 中的每个 Borel 集是形如

$$B, \quad B \cup \{-\infty\}, \quad B \cup \{\infty\}, \quad B \cup \{-\infty, \infty\} \tag{5.20}$$

之一的集合，其中 $B \in \mathscr{B}(\mathbb{R})$.

注 5.4.3 全空间上的 Borel σ 代数与其子集上的 Borel σ 代数有一般关系式. 这里 $\mathbb{R}$ 是 $\overline{\mathbb{R}}$ 的子集，有

$$\mathscr{B}(\mathbb{R}) = \mathbb{R} \cap \mathscr{B}(\overline{\mathbb{R}}). \tag{5.21}$$

定理 5.4.2 下列诸 $\overline{\mathcal{E}}_i$ 均为 $\mathscr{B}(\overline{\mathbb{R}})$ 的生成元

$$\begin{aligned}
\overline{\mathcal{E}}_1 &= \mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}, \\
\overline{\mathcal{E}}_2 &= \{[-\infty, b], b \in \mathbb{R}\}, \\
\overline{\mathcal{E}}_3 &= \{[-\infty, b), b \in \mathbb{R}\}, \\
\overline{\mathcal{E}}_4 &= \{[a, b], a, b \in \overline{\mathbb{R}}, a < b\}, \\
\overline{\mathcal{E}}_5 &= \{(a, b], a, b \in \overline{\mathbb{R}}, a < b\}, \\
\overline{\mathcal{E}}_6 &= \{[a, \infty], a \in \overline{\mathbb{R}}\}, \\
\overline{\mathcal{E}}_7 &= \{(a, \infty], a \in \overline{\mathbb{R}}\}, \\
\overline{\mathcal{E}}_8 &= \{[r, \infty], r \in \mathbb{Q}\}, \\
\overline{\mathcal{E}}_9 &= \{(r, \infty], r \in \mathbb{Q}\}, \\
\overline{\mathcal{E}}_{10} &= \{[-\infty, b), b \in \mathbb{R}\} \cup \{(a, \infty], a \in \mathbb{R}\}.
\end{aligned} \tag{5.22}$$

第六章　测度空间与概率空间

§6.1　测度与测度空间

6.1.1　基本术语

定义 6.1.1 (**有限可加性**) 设 $\mathcal{C}$ 是 Ω 上的一个集类，$\mu : \mathcal{C} \to \overline{\mathbb{R}}$ 为 $\mathcal{C}$ 上的广义实值集函数. 如果 $A_1, A_2, \dots, A_n \in \mathcal{C}$ 两两不交，满足

$$\biguplus_{i=1}^{n} A_i \in \mathcal{C} \Rightarrow \mu\left(\biguplus_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i), \tag{6.1}$$

则称 μ 具有有限可加性.

注 6.1.1 当 $\mathcal{C}$ 对有限并封闭时，μ 具有有限可加性当且仅当

$$A_1, A_2 \in \mathcal{C}, A_1 \cap A_2 = \varnothing \Rightarrow \mu(A_1 \uplus A_2) = \mu(A_1) + \mu(A_2). \tag{6.2}$$

定义 6.1.2 (**可列可加性**) 设 $\mathcal{C}$ 是 Ω 上的一个集类，$\mu : \mathcal{C} \to \overline{\mathbb{R}}$ 为 $\mathcal{C}$ 上的广义实值集函数. 如果 $A_1, A_2, \dots \in \mathcal{C}$ 两两不交，满足

$$\biguplus_{i=1}^{\infty} A_i \in \mathcal{C} \Rightarrow \mu\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i), \tag{6.3}$$

则称 μ 具有可列可加性，也叫可数可加性或 σ 可加性.

定义 6.1.3 (**有限次可加性**) 设 $\mathcal{C}$ 是 Ω 上的一个集类，$\mu : \mathcal{C} \to \overline{\mathbb{R}}$ 为 $\mathcal{C}$ 上的广义实值集函数. 如果 $A, A_1, A_2, \dots, A_n \in \mathcal{C}$，满足

$$A \subset \bigcup_{i=1}^{n} A_i \Rightarrow \mu(A) \le \sum_{i=1}^{n} \mu(A_i), \tag{6.4}$$

则称 μ 具有有限次可加性.

定义 6.1.4 (**可列次可加性**) 设 $\mathcal{C}$ 是 Ω 上的一个集类，$\mu : \mathcal{C} \to \overline{\mathbb{R}}$ 为 $\mathcal{C}$ 上的广义实值集函数. 如果 $A, A_1, A_2, \dots \in \mathcal{C}$，满足

$$A \subset \bigcup_{i=1}^{\infty} A_i \Rightarrow \mu(A) \le \sum_{i=1}^{\infty} \mu(A_i), \tag{6.5}$$

则称 μ 具有可列次可加性.

6.1.2　测度的定义及其基本性质

定义 6.1.5 (**可加测度**) 设 $\mathcal{C}$ 是 Ω 的一个子集类，$\mu:\mathcal{C}\to\overline{\mathbb{R}}$ 是 $\mathcal{C}$ 上的广义实值集函数，如果满足

1. $A\in\mathcal{C}\Rightarrow\mu(A)\geq 0$（即非负性，这说明 $\mu:\mathcal{C}\to[0,+\infty]$）；
2. 至少存在一个 $A\in\mathcal{C}$，使得 $\mu(A)<\infty$；
3. $\forall A,B\in\mathcal{C}$，$A\cup B\in\mathcal{C}$，$A\cap B=\varnothing$ 都有 $\mu(A\cup B)=\mu(A)+\mu(B)$.

则称 μ 为 $\mathcal{C}$ 上的有限可加测度.

定义 6.1.6 (**有限可加测度**) 设 $\mathcal{C}$ 是 Ω 的一个子集类，$\mu:\mathcal{C}\to\overline{\mathbb{R}}$ 是 $\mathcal{C}$ 上的广义实值集函数，如果满足

1. $A\in\mathcal{C}\Rightarrow\mu(A)\geq 0$（即非负性，这说明 $\mu:\mathcal{C}\to[0,+\infty]$）；
2. 至少存在一个 $A\in\mathcal{C}$，使得 $\mu(A)<\infty$；
3. μ 具有有限可加性.

则称 μ 为 $\mathcal{C}$ 上的有限可加测度.

定义 6.1.7 (**测度**) 设 $\mathcal{C}$ 是 Ω 的一个子集类，$\mu:\mathcal{C}\to\overline{\mathbb{R}}$ 是 $\mathcal{C}$ 上的广义实值集函数，如果满足

1. $A\in\mathcal{C}\Rightarrow\mu(A)\geq 0$（即非负性，这说明 $\mu:\mathcal{C}\to[0,+\infty]$）；
2. 至少存在一个 $A\in\mathcal{C}$，使得 $\mu(A)<\infty$；
3. μ 具有可列可加性.

则称 μ 为 $\mathcal{C}$ 上的测度，也称为 σ 可加测度或可列（可数）可加测度.

注 6.1.2 若 $\varnothing\in\mathcal{C}$，则 $\mu(\varnothing)=0$.

定义 6.1.8 (**测度有限**) 若 $\forall A\in\mathcal{C}$，$\mu(A)<\infty$，即 $\mu(A)\in[0,\infty)$，则称上文定义的各种测度是有限的.

定义 6.1.9 (**测度 σ 有限**) 若 $\forall A\in\mathcal{C}$，$\exists\{A_n:n\in\mathbb{Z}_+\}\subset\mathcal{C}$，使 $\bigcup\limits_{n=1}^{\infty}A_n=A$ 且 $\mu(A_n)<\infty,n\in\mathbb{Z}_+$，则称上文定义的各种测度是 σ 有限的.

注 6.1.3 $\mathcal{C}$ 为集代数或 σ 代数时，若 $\exists\{A_n : n \in \mathbb{Z}_+\} \subset \mathcal{C}$，使 $\bigcup\limits_{n=1}^{\infty} A_n = \Omega$ 且 $\mu(A_n) < \infty, n \in \mathbb{Z}_+$，则 μ 就是 $\mathcal{C}$ 上的 σ 有限测度；若 $\mu(\Omega) < \infty$，则 μ 就是 $\mathcal{C}$ 上的有限测度.

6.1.3 可测空间、测度空间和概率空间

定义 6.1.10 (**可测空间**) 若 $\mathcal{F}$ 是 Ω 上的一个 σ 代数，则称 $(\Omega, \mathcal{F})$ 为可测空间，并称 $A \in \mathcal{F}$ 是 Ω 中关于 $\mathcal{F}$ 的可测集.

定义 6.1.11 (**测度空间**) 若 $\mathcal{F}$ 是 Ω 上的一个 σ 代数，μ 为其上的一个测度，则称 $(\Omega, \mathcal{F}, \mu)$ 为测度空间.

定义 6.1.12 (**概率与概率空间**) 若 $\mathcal{F}$ 是 Ω 上的一个 σ 代数，μ 为其上的一个测度，且满足 $\mu(\Omega) = 1$，则称 μ 为 $\mathcal{F}$ 上的概率（或概率测度），通常可以用 P 替代 μ 作为特殊表示，并称 $(\Omega, \mathcal{F}, \mu)$（或记 $(\Omega, \mathcal{F}, P)$）为概率空间.

定义 6.1.13 (Borel **测度**) $\mathscr{B}(X) = \sigma(\mathcal{T})$ 是定义在拓扑空间 $(X, \mathcal{T})$ 上的 Borel σ 代数. 则称 $(X, \mathscr{B}(X), \mathcal{T})$ 为可测拓扑空间，其上的测度 μ 称为 X 上的 Borel 测度.

定义 6.1.14 (**可测度量空间**) $\mathscr{B}(X)$ 是定义在度量空间 (X, ρ) 上的 Borel σ 代数. 则称三元组 $(X, \mathscr{B}(X), \rho)$ 为可测度量空间.

6.1.4 外测度

定义 6.1.15 (**外测度**) 称 $\mu^* : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ 是 Ω 上的一个广义实值集函数，如果满足

1. $\mu^*(\varnothing) = 0$；
2. μ^* 具有单调性：若 $A \subset B$，则 $\mu^*(A) \subset \mu^*(B)$；
3. μ^* 具有可列次可加性：$\mu^*\left(\bigcup\limits_{n=1}^{\infty} A_n\right) \leq \sum\limits_{n=1}^{\infty} \mu^*(A_n)$，$\forall A_1, A_2, \ldots \in \mathcal{P}(\Omega)$.

则称 μ^* 是 Ω 上的一个外测度.

性质 6.1.1 根据定义可以发现，外测度具有

- 非负性：$\mu^*(A) \geq 0$，$\forall A \in \mathcal{P}(A)$；
- 有限次可加性：$\mu^*\left(\bigcup\limits_{i=1}^{n} A_i\right) \leq \sum\limits_{i=1}^{n} \mu^*(A_i)$，$\forall A_1, A_2, \ldots, A_n \in \mathcal{P}(\Omega)$，$n \geq 1$.

注 6.1.4 测度的定义域一般不必是 $\mathcal{P}(\Omega)$，故测度未必是外测度；而外测度一般不满足可列可加性，故外测度也未必是测度.

定理 6.1.1 设 μ^* 为 Ω 的外测度，$A \subset \Omega$，如果 $\forall D \subset \Omega$，有

$$\mu^*(D) = \mu^*(A \cap D) + \mu^*(A^c \cap D). \tag{6.6}$$

则称 A 为 μ^* 可测集.

注 6.1.5 由于外测度具有有限次可加性，且 $\mu^*(\varnothing) = 0$，容易得知：

$$\mu^*(D) \leq \mu^*(A \cap D) + \mu^*(A^c \cap D) + \mu^*(\varnothing) + \mu^*(\varnothing) + \cdots = \mu^*(A \cap D) + \mu^*(A^c \cap D). \tag{6.7}$$

引理 6.1.1 $A \subset \Omega$ 为 μ^* 可测集当且仅当 $\forall D \subset \Omega$，有

$$\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D). \tag{6.8}$$

§6.2　测度的构造

6.2.1　测度构造思路

定义 6.2.1 (**扩张与限制**) 设 $\mathcal{C}_1$，$\mathcal{C}_2$ 都是 Ω 的子集类，$\mathcal{C}_1 \subset \mathcal{C}_2$，$\mu_i$ 是 $\mathcal{C}_i$ 上的测度或有限可加测度，$i = 1, 2$. 如果 $\forall A \in \mathcal{C}_1$，有 $\mu_1(A) = \mu_2(A)$，则称 μ_2 是 μ_1 的一个扩张，或 μ_1 为 μ_2 的一个限制.

想要直接在 σ 代数上构造测度，除非特别简单的场合，否则在实际当中几乎是不可能的. 但注意到 σ 代数一定是半集代数，借助半集代数构造测度是有希望的，一个思路是先在半集代数上构造测度，然后将其扩张成 σ 代数上的测度，但想要刚好扩张到 σ 代数上也有一定的困难，这时需要借助外测度另辟蹊径.

方法 6.2.1 (**测度构造思路**) 一般都是先在半集代数上构造测度，然后将它扩张成外测度，最后将这个外测度限制在由半集代数生成的 σ 代数上.

6.2.2　半集代数上的测度的性质

定理 6.2.1 设 μ 是半集代数 $\mathscr{S}$ 上的测度（或有限可加测度），则 μ 在由 $\mathscr{S}$ 生成的集代数 $\mathscr{A}(\mathscr{S})$ 上存在唯一扩张 $\tilde{\mu}$.

定理 6.2.2 设 μ 是半集代数 $\mathscr{S}$ 上的有限可加测度，若 $\{A_1,\dots,A_n\}\subset\mathscr{S}$，$\bigcup\limits_{k=1}^{n}A_k\subset A$ 且 $A_k, k=1,2,\dots,n$ 两两不交，则

$$\sum_{k=1}^{n}\mu(A_k)\le\mu(A). \tag{6.9}$$

引理 6.2.1 如下两点成立

1. 若 μ 是半集代数 $\mathscr{S}$ 上的有限可加测度，$\{A_1,\dots,A_n\}\subset\mathscr{S}$ 且 $A\subset\bigcup\limits_{k=1}^{n}A_k$，则

$$\mu(A)\le\sum_{k=1}^{n}\mu(A_k). \tag{6.10}$$

2. 若 μ 是半集代数 $\mathscr{S}$ 上的测度，$A, A_n\in\mathscr{S}$，$n\in\mathbb{Z}_+$，且 $A\subset\bigcup\limits_{k=n}^{\infty}A_n$，则

$$\mu(A)\le\sum_{n=1}^{\infty}\mu(A_n). \tag{6.11}$$

引理 6.2.2 μ 是半集代数 $\mathscr{S}$ 上的 σ 有限测度，则它在集代数 $\mathscr{A}(\mathscr{S})$ 上的扩张也是 σ 有限测度.

6.2.3　外测度的性质

引理 6.2.3 设 μ 是 Ω 上的半集代数 $\mathscr{S}$ 上的测度，$\forall A\subset\Omega$，令

$$\mu^*(A)=\inf\left\{\sum_{n=1}^{\infty}\mu(A_n):A\subset\bigcup_{n=1}^{\infty}A_n, A_n\in\mathscr{S}, n\in\mathbb{Z}_+\right\}, \tag{6.12}$$

则 μ^* 是一个外测度（称之为由 μ 引出的外测度）且它在 $\mathscr{S}$ 上与 μ 是一致的，即

$$\mu^*(A)=\mu(A),\quad\forall A\in\mathscr{S}. \tag{6.13}$$

定理 6.2.3 设 μ^* 是 Ω 的一个外测度，令

$$\mathscr{A}_{\mu^*}:=\left\{A:A\subset\Omega\text{为 }\mu^*\text{ 可测集}\right\}, \tag{6.14}$$

则

1. $\mathscr{A}_{\mu^*}$ 为 σ 代数.

2. 设 $A_n\in\mathscr{A}_{\mu^*}$，$n\in\mathbb{Z}_+$ 两两不交，$A=\bigcup\limits_{n=1}^{\infty}A_n$，则

$$\mu^*(A\cap D)=\bigcup_{n=1}^{n}\mu^*(A_n\cap D),\quad\forall D\subset\Omega. \tag{6.15}$$

3. μ^* 在 $\mathscr{A}_{\mu^*}$ 上的限制（仍记作 $\mathscr{A}_{\mu^*}$）是 $\mathscr{A}_{\mu^*}$ 上的测度.

6.2.4 测度扩张定理

依据上述事实，我们可以给出测度扩张定理.

定理 6.2.4 (**测度扩张定理**) 设 μ 为半集代数 $\mathscr{S}$ 上的测度，则

1. （扩张的存在性）μ 在 $\mathscr{S}$ 上生成的 σ 代数 $\sigma(\mathscr{S})$ 上存在一个扩张.
2. （扩张的唯一性）若 μ 在 $\mathscr{S}$ 上 σ 有限，则 μ 在 $\sigma(\mathscr{S})$ 上的扩张唯一，即若 μ_1，μ_2 是 μ 上的扩张，则 $\forall A \in \sigma(\mathscr{S})$ 都有 $\mu_1(A) = \mu_2(A)$. 此外，所得扩张的测度在 $\sigma(\mathscr{S})$ 上也是 σ 有限的.

注 6.2.1 测度扩张定理是说从一个半集代数上的测度出发，是可以唯一的扩张到由此半集代数生成的 σ 代数上的测度的. 在此约定，如无特殊说明，下文谈及测度默认指可测空间的 σ 代数上的测度.

§6.3　n 维欧氏空间中的测度

6.3.1 连续函数与单调函数

定义 6.3.1 (**单侧极限**) 用 $t \uparrow x$ 表示“$t < x, t \to x$”; 用 $t \downarrow x$ 表示“$t > x, t \to x$”. 定义左极限为

$$\lim_{t \uparrow x} f(t) = f(x-), \tag{6.16}$$

右极限为

$$\lim_{t \downarrow x} f(t) = f(x+). \tag{6.17}$$

定义 6.3.2 (**连续函数与跳跃间断点**) $\forall x \in \mathbb{R}$，函数 $f : \mathbb{R} \to \mathbb{R}$ 在 x 处连续当且仅当 $f(x-) = f(x+)$. 若单侧极限 $f(x-)$ 和 $f(x+)$ 存在但不相等，则称 f 在 x 处有一跳跃，x 为其跳跃间断点. 称 $|f(x-) - f(x+)|$ 为 f 在 x 处的跃度.

定义 6.3.3 (**增函数**) 设函数 $f : \mathbb{R} \to \mathbb{R}$. 如果 $\forall x_1, x_2 \in \mathbb{R}$，$x_1 < x_2$，有 $f(x_1) \leq f(x_2)$，则称 f 为增函数（或不降函数）. 若不等号严格成立，则称其为严格增函数.

性质 6.3.1 由以上内容可知，增函数具有如下性质：

1. $\forall x \in \mathbb{R}$，$f(x-)$ 及 $f(x+)$ 存在且有极限，并有
$$f(x-) \leq f(x) \leq f(x+). \tag{6.18}$$
2. $f(-\infty) := \lim\limits_{t \downarrow -\infty} f(t)$ 及 $f(\infty) := \lim\limits_{t \uparrow \infty} f(t)$ 存在，但前者可能是 $-\infty$，后者可能是 ∞.

3. 增函数的不连续点都是跳跃点，且其不连续点最多有可数个.

4. $\mathbb{R}$ 上的增函数除某些跳跃点的函数值外，由该函数在一稠密集上的值唯一确定. 即设 f_1 和 f_2 是 $\mathbb{R}$ 上的两个增函数，D 在 $\mathbb{R}$ 上稠，且

$$f_1(x) = f_2(x), \quad x \in D. \tag{6.19}$$

则 f_1 和 f_2 具有相同的跳跃点，且在同一跳跃点处的跃度相同. 在除了某些跳跃间断点外，f_1 和 f_2 的值都相等.

5. 若 f 为 $\mathbb{R}$ 上的增函数，则 $\forall x \in \mathbb{R}$，$\tilde{f} := f(x+)$ 是处处右连续的增函数，$\tilde{f} := f(x-)$ 是处处左连续的增函数.

6. 若 D 在 $\mathbb{R}$ 中稠，f 是 D 上的增函数，则

$$\tilde{f} : \tilde{f}(x) := \inf_{x<t\in D} f(t) \tag{6.20}$$

为处处右连续的增函数.

6.3.2 Lebesgue-Stieltjes（L-S）函数与 L-S 测度

定理 6.3.1 设 $F : \mathbb{R} \to \mathbb{R}$ 为 $\mathbb{R}$ 上的右连续增函数，则在半集代数

$$\mathscr{S} := \{(a, b] : a, b \in \mathbb{R}, a \leq b\} \tag{6.21}$$

上有唯一测度 $\mu = \mu_F$ 存在使得

$$\mu((a, b]) = F(b) - F(a), \quad a, b \in \mathbb{R}, a \leq b, \tag{6.22}$$

且 μ 在有限区间上的值有限（因而也是 σ 有限的）.

推论 6.3.1 设 $F : \mathbb{R} \to \mathbb{R}$ 为 $\mathbb{R}$ 上的右连续增函数，则在 $\mathscr{B}(\mathbb{R})$ 上有唯一的 σ 有限测度 μ_F，满足

$$\mu_F((a, b]) = F(b) - F(a), \quad a, b \in \mathbb{R}, a \leq b. \tag{6.23}$$

定义 6.3.4 (Lebesgue-Stieltjes **函数**) 设 $F : \mathbb{R}^n \to \mathbb{R}$ 为定义在 $\mathbb{R}^n$ 上的有限实值函数，满足

1. $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$，$\boldsymbol{a} \leq \boldsymbol{b}$，有

$$\Delta_{\boldsymbol{b},\boldsymbol{a}} F := \sum_{k=1}^{n} (-1)^k \sum_{1\leq i_1<\cdots<i_k\leq n} F\left(\boldsymbol{b} - \sum_{l=1}^{k} (b_{i_l} - a_{i_l}) \boldsymbol{e}_{i_l}\right) \geq 0. \tag{6.24}$$

其中 e_i 表示除第 i 个坐标上的元素为 1 以外，其余坐标元素均为 0 的单位向量，即

$$\boldsymbol{e}_{i_l} = \begin{pmatrix} 0 & 0 & \cdots & \underset{\text{第 } i_l \text{ 个}}{\underset{\uparrow}{1}} & \cdots & 0 \end{pmatrix}. \tag{6.25}$$

2. $F(\boldsymbol{x}) = F(x_1, \ldots, x_n), \boldsymbol{x} \in \mathbb{R}^n$，对每一自变量右连续.

则称 F 为定义在 $\mathbb{R}^n$ 上的 Lebesgue-Stieltjes 函数，简称 L-S 函数.

注 6.3.1 二维场合下，$\boldsymbol{a} = (a_1, a_2)$，$\boldsymbol{b} = (b_1, b_2)$，则

$$\Delta_{\boldsymbol{b},\boldsymbol{a}} F = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \tag{6.26}$$

注 6.3.2 定义6.3.4中的第一点要求中 $\boldsymbol{b} - \sum\limits_{l=1}^{k} (b_{i_l} - a_{i_l}) \boldsymbol{e}_{i_l}$ 实际上表示将 $\boldsymbol{b}$ 中第 $i_1, \ldots, i_k$ 个坐标换成 $\boldsymbol{a}$ 的相应坐标而得到的点. 将 $\Delta_{\boldsymbol{b},\boldsymbol{a}} F$ 展开可写作

$$\begin{aligned} \Delta_{\boldsymbol{b},\boldsymbol{a}} F =& F(b_1, \ldots, b_n) - [F(a_1, b_2, \ldots, b_n) + \cdots + F(b_1, \ldots, b_{n-1}, a_n)] \\ &+ [F(a_1, a_2, b_3, \ldots, b_n) + \cdots + F(b_1, \ldots, b_{n-2}, a_{n-1}, a_n)] \\ &- \cdots + (-1)^n F(a_1, \ldots, a_n) \geq 0. \end{aligned} \tag{6.27}$$

定理 6.3.2 设 $F : \mathbb{R}^n \to \mathbb{R}$ 为定义在 $\mathbb{R}^n$ 上的 L-S 函数，则在 $\mathbb{R}^n$ 的半集代数

$$\mathscr{S}^n := \{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a} \leq \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n\} \tag{6.28}$$

上有唯一的测度 $\mu = \mu_F$ 存在，使得

$$\mu((\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{\boldsymbol{b},\boldsymbol{a}} F, \quad \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \leq \boldsymbol{b}. \tag{6.29}$$

定义 6.3.5 (Lebesgue-Stieltjes **测度**) 设 μ 为 $\mathscr{B}(\mathbb{R}^n)$ 上的 σ 有限测度，如果对 $\mathbb{R}^n$ 中任意的有限区间 I，都有 $\mu(I) < \infty$，则称 μ 为 Lebesgue-Stieltjes 测度，简称 L-S 测度.

推论 6.3.2 设 $F : \mathbb{R}^n \to \mathbb{R}$ 为定义在 $\mathbb{R}^n$ 上的 L-S 函数，则在 $\mathscr{B}(\mathbb{R}^n)$ 上存在唯一的 σ 有限测度 μ_F，满足

$$\mu_F((\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{\boldsymbol{b},\boldsymbol{a}} F, \quad \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{a} \leq \boldsymbol{b}. \tag{6.30}$$

这是一个 L-S 测度，称之为由 F 诱导（或说成决定）的 L-S 测度.

6.3.3　Lebesgue 外测度

定义 6.3.6 (Lebesgue **函数**) 取 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1, \ldots, x_n) = \int_0^{x_1} \cdots \int_0^{x_n} 1 \, dx_1 \cdots dx_n = x_1 \cdots x_n. \tag{6.31}$$

运算可得

$$\Delta_{\boldsymbol{b},\boldsymbol{a}} F = \prod_{i=1}^{n} (b_i - a_i). \tag{6.32}$$

则 F 是 L-S 函数，称之为 Lebesgue 函数，简称 L 函数.

定义 6.3.7 (Lebesgue **测度**) 由 n 维 L 函数在 $\mathscr{B}(\mathbb{R}^n)$ 上诱导的 L-S 测度，称为 $\mathscr{B}(\mathbb{R}^n)$ 上的 Lebesgue 测度，简称 L 测度，记为 λ，即有：

$$\lambda((\boldsymbol{a},\boldsymbol{b}]) = \Delta_{\boldsymbol{b},\boldsymbol{a}}F = \prod_{i=1}^{n}(b_i - a_i), \quad \boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n, \boldsymbol{a}\le\boldsymbol{b}. \tag{6.33}$$

定义 6.3.8 (Lebesgue **外测度**) 设 $\forall A\subset\mathbb{R}^n$，$I_n$ 为 $\mathbb{R}^n$ 的开区间，用 $|I_n|$ 表示 I_n 的体积，令

$$\lambda^*(A) := \inf\left\{\sum_{n=1}^{\infty}|I_n| : A\subset\bigcup_{n=1}^{\infty}I_n\right\}. \tag{6.34}$$

对 $\mathbb{R}^n$ 的半集代数 $\mathscr{S}^n$ 上的 L 测度 λ，成立

$$\lambda^*(A) = \inf\left\{\sum_{n=1}^{\infty}\lambda(A_n) : A\subset\bigcup_{n=1}^{\infty}A_n, A_n\in\mathscr{S}^n, n\in\mathbb{Z}_+\right\}. \tag{6.35}$$

这说明 λ^* 是 $\mathbb{R}^n$ 的一个外测度，称为 Lebesgue 外测度，简称 L 外测度.

推论 6.3.3 由 $\Omega=\mathbb{R}^n(n\in\mathbb{Z}_+)$ 上的 Lebesgue 外测度 λ^* 决定的测度空间 $(\mathbb{R}^n,\mathscr{A}_{\lambda^*},\lambda^*)$ 是 $\mathscr{S}^n$ 上的体积测度 λ，即

$$\lambda(\boldsymbol{a},\boldsymbol{b}] := \begin{cases} \prod_{k=1}^{n}(b_k-a_k), & \boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n \\ \lim_{N\to\infty}\prod_{k=1}^{n}(b_k\wedge N - a_k\vee(-N)), & \exists a_k=-\infty \ or\ b_k=\infty \end{cases} \tag{6.36}$$

的一个扩张，$\mathscr{A}_{\lambda^*}$ 的元称为 n 维 Lebesgue 可测集，$\mathscr{A}_{\lambda^*}$ 上的测度 λ^* 就是 n 维 Lebesgue 测度，并把 λ^* 在 $\mathscr{B}^n=\sigma(\mathscr{S}^n)$ 上的限制称为 n 维 Borel-Lebesgue 测度，也称为 Borel 测度.

6.3.4 定义在 $\mathbb{R}^n$ 上的分布函数与准分布函数

定义 6.3.9 (**分布函数**) 设 $F:\mathbb{R}^n\to\mathbb{R}$ 为定义在 $\mathbb{R}^n$ 上的 L-S 函数，又满足

1. F 单调不减，即 $\boldsymbol{x}<\boldsymbol{y}\Rightarrow F(\boldsymbol{x})\le F(\boldsymbol{y})$.

2. 对每个 $i=1,2,\ldots,n$，满足

$$F(x_1,x_2,\ldots,-\infty,x_{i+1},\ldots,x_n) := \lim_{x_i\to-\infty}F(x_1,\ldots,x_i,\ldots,x_n)=0, \tag{6.37}$$

以及

$$F(\boldsymbol{\infty}) = F(\infty,\ldots,\infty) = \lim_{\boldsymbol{x}\to\boldsymbol{\infty}}F(\boldsymbol{x}) = 1. \tag{6.38}$$

则称 F 为分布函数.

定义 6.3.10 (**准分布函数**) 设 $G(\boldsymbol{x})$ 为定义在 $\mathbb{R}^n$ 上的分布函数，$F:\mathbb{R}^n\to\mathbb{R}$ 是定义在 $\mathbb{R}^n$ 上的有限实值函数，又 $\exists\sigma^2>0$，使得

$$F(\boldsymbol{x}) = \sigma^2 G(\boldsymbol{x}). \tag{6.39}$$

则称 F 为准分布函数.

注 6.3.3 若 F 为准分布函数，则可以发现

$$F(\infty) = \sigma^2. \tag{6.40}$$

§6.4　测度空间的性质

6.4.1　基本运算性质

根据前文定义，$\mathcal{F}$ 是 Ω 上的一个 σ 代数，$(\Omega, \mathcal{F})$ 是可测空间，其上的测度 μ 是 $\mathcal{F}$ 上的广义实值函数，首先需要满足定义的要求，即

1. $A \in \mathcal{F} \Rightarrow \mu(A) \geq 0$（即非负性，这说明 $\mu : \mathcal{F} \to [0, +\infty]$）；
2. 至少存在一个 $A \in \mathcal{F}$，使得 $\mu(A) < \infty$；
3. μ 具有可列可加性. 即若 $A_n \in \mathcal{F}, n \in \mathbb{Z}_+$ 两两不交，则

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{6.41}$$

注 6.4.1 若还满足 $\mu(\Omega) = 1$，则 μ 就是概率测度.

根据上述事实，测度空间上的测度具有如下性质：

性质 6.4.1 $\mu(\varnothing) = 0$.

性质 6.4.2 (Poincaré **公式**) 若 $A, B \in \mathcal{F}$，则

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B) \tag{6.42}$$

若还满足 $\mu(A) < \infty$ 或 $\mu(B) < \infty$，则

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \tag{6.43}$$

一般情况下，$A_k \in \mathcal{F}$，$k = 1, 2, \ldots, n$，若 $\mu(A_k) < \infty$，则

$$\mu(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mu(A_{i_1} A_{i_2} \cdots A_{i_k}). \tag{6.44}$$

这个公式称为 Poincaré 公式.

注 6.4.2 将上述性质中的一般情形展开会更好理解：

$$\begin{aligned}\mu(A_1\cup A_2\cup\cdots\cup A_n)=&\sum_{k=1}^{n}\mu(A_k)-\sum_{1\le i_1<i_2\le n}\mu(A_{i_1}A_{i_2})+\sum_{1\le i_1<i_2<i_3\le n}\mu(A_{i_1}A_{i_2}A_{i_3})\\&+\cdots+(-1)^{n-1}\mu(A_1A_2\cdots A_n).\end{aligned}\tag{6.45}$$

性质 6.4.3 若 $A,B\in\mathcal{F}$，$A\subset B$，则

$$\mu(B)=\mu(A)+\mu(B\backslash A),\tag{6.46}$$

且 $\mu(A)\le\mu(B)$；若还有 $\mu(A)<\infty$，则

$$\mu(B\backslash A)=\mu(B)-\mu(A).\tag{6.47}$$

注 6.4.3 对于概率测度而言，$\mu(A^c)=1-\mu(A)$，且 $\mu(A)\le 1$.

性质 6.4.4 μ 具有次可列可加性，也称次 σ 可加性.

6.4.2 连续性

定义 6.4.1 (**下方连续**) 设 $(\Omega,\mathcal{F},\mu)$ 是测度空间，若 $A_n\in\mathcal{F},n\in\mathbb{Z}_+$，$A_n\uparrow$，有

$$\lim_{n\to\infty}\mu(A_n)=\mu\left(\bigcup_{n=1}^{\infty}A_n\right),\tag{6.48}$$

则 μ 在 $\mathcal{F}$ 上是下方连续的.

定义 6.4.2 (**上方连续**) 设 $(\Omega,\mathcal{F},\mu)$ 是测度空间，若 $A_n\in\mathcal{F},n\in\mathbb{Z}_+$，$A_n\downarrow$，且 $\exists m\in\mathbb{Z}_+$，使 $\mu(A_m)<\infty$，有

$$\lim_{n\to\infty}\mu(A_n)=\mu\left(\bigcap_{n=1}^{\infty}A_n\right),\tag{6.49}$$

则 μ 在 $\mathcal{F}$ 上是上方连续的.

定理 6.4.1 设 $(\Omega,\mathcal{F},\mu)$ 是测度空间，则 μ 既是下方连续又是上方连续的.

定理 6.4.2 μ 是 Ω 的集代数 $\mathscr{A}$ 上的可加测度，若 μ 满足如下两个条件之一：

1. μ 在 $\mathscr{A}$ 上是下方连续的，即对任何满足 $\{A_n\}\subset\mathscr{A}$，$A_n\uparrow\bigcup\limits_{n=1}^{\infty}A_n\in\mathscr{A}$ 的集列，都有

$$\lim_{n\to\infty}\mu(A_n)=\mu\left(\bigcup_{n=1}^{\infty}A_n\right).\tag{6.50}$$

2. $\mu(\Omega)<\infty$ 且在 $\varnothing$ 处上方连续，即对任何满足 $\{A_n\}\subset\mathscr{A}$，$A_n\downarrow\bigcap\limits_{n=1}^{\infty}A_n=\varnothing$ 的集列，都有

$$\lim_{n\to\infty}\mu(A_n)=0=\mu(\varnothing). \tag{6.51}$$

则 μ 是 $\mathscr{A}$ 上的测度.

6.4.3　测度的逼近

命题 6.4.1 设 μ 是半集代数 $\mathscr{S}$ 上的测度，$\mathscr{A}(\mathscr{S})$ 是 $\mathscr{S}$ 生成的集代数，μ^* 是由 μ 引出的外测度，则 $\forall A\in\mathscr{A}_{\mu^*}$，$\mu^*(A)<\infty$ 及 $\varepsilon>0$，存在 $A_\varepsilon\in\mathscr{A}(\mathscr{S})$ 使

$$\mu^*(A\triangle A_\varepsilon)=\mu^*(\{\omega:|I_A(\omega)-I_{A_\varepsilon}(\omega)|=1\})<\varepsilon. \tag{6.52}$$

其中 $I_A(\omega)$ 表示示性函数：

$$I_A(\omega)=\begin{cases}1 & ,\omega\in A\\ 0 & ,\omega\notin A\end{cases}. \tag{6.53}$$

命题 6.4.2 设 A 为 $\mathbb{R}^n$ 的一个 Lebesgue 可测子集，则

$$\begin{aligned}\lambda(A)&=\inf\{\lambda(G):A\subset G\subset\mathbb{R}^n,G\text{是开集}\},\\ &=\sup\{\lambda(K):A\supset K,\text{是紧集}\}.\end{aligned} \tag{6.54}$$

因而，若 $\lambda(A)<\infty$，则 $\forall\varepsilon>0$，存在一个开集 G 及一个紧集 K（此处也是有界闭集）使得

$$K\subset A\subset G\ \text{且}\lambda(G)-\varepsilon\le\lambda(A)\le\lambda(K)+\varepsilon. \tag{6.55}$$

§6.5　测度空间的完备化

6.5.1　零测集与 μ 零集

定义 6.5.1 (**零测集**) 设 $(\Omega,\mathcal{F},\mu)$ 为测度空间，若 $M\in\mathcal{F}$ 且 $\mu(M)=0$，则称 M 为零测集.

定义 6.5.2 (μ **零集**) 设 $(\Omega,\mathcal{F},\mu)$ 为测度空间，若 $N\subset M\in\mathcal{F}$ 且 $\mu(M)=0$（即 M 为零测集），则称 M 为 μ 零集，也称为可略集.Ω 中全体 μ 零集组成的集类称为 μ 零集类，可记为 $\mathcal{N}_\mu$.

6.5.2　完备测度空间与完备测度

定义 6.5.3 (**完备测度空间与完备测度**) 设 $(\Omega,\mathcal{F},\mu)$ 为测度空间，$\mathcal{N}_\mu$ 为 μ 零集类，若 $\mathcal{N}_\mu\subset\mathcal{F}$，即每一 μ 零集都属于 $\mathcal{F}$，则称 $(\Omega,\mathcal{F},\mu)$ 为完备（或称完全）测度空间，称 μ 为完备（或称完全）测度.

定理 6.5.1 设 $(\Omega, \mathcal{F}, \mu)$ 为一测度空间，令

$$\begin{aligned}\overline{\mathcal{F}} :&= \{A \triangle N : A \in \mathcal{F}, N \text{为 } \mu \text{ 零集}\} \\ &= \{A \cup N : A \in \mathcal{F}, N \text{为 } \mu \text{ 零集}\},\end{aligned} \tag{6.56}$$

$$\overline{\mu}(A \triangle N) := \mu(A), A \in \mathcal{F}, N \text{为 } \mu \text{ 零集}, \tag{6.57}$$

则 $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ 为一完备测度空间，称它是 $(\Omega, \mathcal{F}, \mu)$ 的完备化.

定理 6.5.2 设 μ 是半集代数 $\mathscr{S}$ 上的 σ 有限测度，μ^* 是由 μ 引出的外测度，$\mathscr{A}_{\mu^*}$ 是一切 μ^* 可测集组成的 σ 代数，把 μ 在 $\sigma(\mathscr{S})$ 上的扩张仍记为 μ，则测度空间 $(\Omega, \mathscr{A}_{\mu^*}, \mu^*)$ 是 $(\Omega, \sigma(\mathscr{S}), \mu)$ 的完备化. 此时 μ 在 $\sigma(\mathscr{S})$ 上的扩张是唯一的.

推论 6.5.1 $\mathbb{R}^n$ 上的 Lebesgue 测度是 Borel 测度的完备化.

第七章　可测函数与随机变量

§7.1　逆像与可测映射

7.1.1　集类的逆像

定义 7.1.1 设 f 是 Ω 到 Ω' 上的映射，若 $\mathcal{C}' \subset \mathcal{P}(\Omega')$，则定义

$$f^{-1}(\mathcal{C}') := \left\{ f^{-1}(C') : C' \in \mathcal{C}' \right\} \tag{7.1}$$

为集类 $\mathcal{C}'$ 对 f 的逆像.

定理 7.1.1 设 f 是 Ω 到 E 上的映射，如下两条成立：

1. 若 $\mathcal{E}$ 是 E 上的一个 σ 代数，则 $f^{-1}(\mathcal{E})$ 是 Ω 的 σ 代数.
2. 若 $\mathcal{C} \subset \mathcal{P}(E)$ 非空，则
$$f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C})). \tag{7.2}$$
即由集类 $\mathcal{C}$ 生成的 σ 代数的逆像与 $\mathcal{C}$ 的逆像生成的 σ 代数相同.

7.1.2　可测映射

定义 7.1.2 (**可测映射**) 设 $(\Omega, \mathcal{F})$ 及 $(E, \mathcal{E})$ 为两个可测空间，若映射 $f : \Omega \to E$ 满足

$$f^{-1}(\mathcal{E}) \subset \mathcal{F}, \tag{7.3}$$

这意味着

$$f^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{E}. \tag{7.4}$$

即 $\mathcal{E}$ 中的每个集合在 f 下的逆像都是 $\mathcal{F}$ 中的集合，则称 f 是从 $(\Omega, \mathcal{F})$ 到 $(E, \mathcal{E})$ 的可测映射，或称 f 是 $\mathcal{F}\backslash\mathcal{E}$ 可测的，通常记为 $f \in \mathcal{F}\backslash\mathcal{E}$.

注 7.1.1 谈及可测映射时，必须明确 $\mathcal{F}$ 和 $\mathcal{E}$，否则可能造成含义不明确. 若 $\mathcal{F}$ 和 $\mathcal{E}$ 都不言自明时，简称 f 从 Ω 到 E 的可测映射.

注 7.1.2 通过可测映射的定义可以知道，其本质是：可测集的逆像仍是可测集.

注 7.1.3 $f^{-1}(B)$ 有时也记作 $\{f \in B\}$，若 $\mathcal{F}$ 上的测度为 μ，则以下记法等价

$$\mu(\{\omega : f(\omega) \in B\}), \quad \mu_f(B), \quad \mu(\{f \in B\}), \quad \mu(f^{-1}(B)), \quad \mu \circ f^{-1}(B). \tag{7.5}$$

定义 7.1.3 (Borel **可测映射**) 当 Ω 和 E 都是拓扑空间，且 $\mathcal{F} = \mathscr{B}(\Omega)$，$\mathcal{E} = \mathscr{B}(E)$ 时，称可测映射 $f \in \mathcal{F}\backslash\mathcal{E}$ 为 Borel 可测映射.

定理 7.1.2 (**可测映射的复合仍然可测**) 若 X 是从 $(\Omega, \mathcal{F})$ 到 $(E, \mathcal{E})$ 的可测映射，Y 是从 $(E, \mathcal{E})$ 到 $(G, \mathcal{G})$ 的可测映射，则复合映射 $Y \circ X$ 是可测映射.

定理 7.1.3 (**连续映射可测**) 设 $(\Omega, \mathcal{F})$ 和 $(E, \mathcal{E})$ 是两个可测空间，若映射 $f : \Omega \to E$ 是连续的，则 $f \in \mathcal{F}\backslash\mathcal{E}$，即 f 是可测的.

§7.2　可测函数与随机变量

7.2.1　定义

定义 7.2.1 (**可测函数**) 设 $(\Omega, \mathcal{F}, \mu)$ 是一测度空间，若函数 $f : \Delta(\in \mathcal{F}) \to \overline{\mathbb{R}}$ 使 $\forall B \in \overline{\mathscr{B}}$，有

$$f^{-1}(B) := \{\omega \in \Delta : f(\omega) \in B\} \in \{\Delta \cap F : F \in \mathcal{F}\} := \Delta \cap \mathcal{F}, \tag{7.6}$$

则称 f 为定义在 Δ 上的可测函数. 若 $E = \mathbb{R}$，则称 f 为 Δ 上的有限实值可测函数.

注 7.2.1 可测函数是定义在 Δ 上，取值于 $\overline{\mathbb{R}}$ 的可测映射. 而有限实值可测函数是定义在 Δ 上，取值于 $\mathbb{R}$ 的可测映射.

注 7.2.2 此外还有，Δ 上的有限复值可测函数是定义在 $\Delta \in \mathcal{F}$ 上而取值于复平面的可测映射.

定义 7.2.2 (**随机元**) 设 $(\Omega, \mathcal{F})$ 及 $(E, \mathcal{E})$ 为两个可测空间，f 为 $\Omega \to E$ 上的可测映射，若 $\mathcal{F}$ 上有概率测度 $\mathbf{P}$，则称可测映射 f 为 $(\Omega, \mathcal{F}, \mathbf{P})$ 到 $(E, \mathcal{E})$ 上的随机元.

定义 7.2.3 (**随机变量**) 设 $(\Omega, \mathcal{F}, \mu)$ 是一测度空间，若函数 $f : \Delta(\in \mathcal{F}) \to \overline{\mathbb{R}}$ 为定义在 Δ 上的可测函数，且 μ 为概率测度，则可测函数 f 是定义在 Δ 上的广义随机变量. 若 $E = \mathbb{R}$，则称 f 为 Δ 上的有限实值随机变量，也简称随机变量.

注 7.2.3 一个 Δ 上的有限复值可测函数在满足上述两个定义的背景下，也会对应有限复值随机变量. 它的实部和虚部都是有限实值可测函数，从而我们处理复值随机变量时可按实部和虚部化成实的情况进行讨论.

根据前面介绍过的诸多事实，设 $(\Omega,\mathcal{F},\mathbf{P})$ 为概率空间，$(E,\mathcal{E})$ 为可测空间，$B\in\mathcal{E}$，以下内容成立：

1. $\mathbf{P}$ 称为 $\mathcal{F}$ 上的概率测度；
2. $f:\Omega\to E$ 称为 $(\Omega,\mathcal{F},\mathbf{P})$ 到 $(E,\mathcal{E})$ 上的随机元；
3. $\mathbf{P}(f^{-1}(B))$ 称为 f 的分布测度；
4. 若 $E=\mathbb{R}$，则 f 就是随机变量，令
$$F_f(x)=\mathbf{P}_f((-\infty,x]),\tag{7.7}$$
则 F_f 就是该随机变量的分布函数.

定理 7.2.1 X 是 $(\Omega,\mathcal{F})$ 到 $(E,\mathcal{E})$ 的可测映射（或随机元）的充要条件是存在 $\mathcal{E}$ 的一个子集类 $\mathcal{C}$ 满足：

1. $\sigma(\mathcal{C})=\mathcal{E}$；
2. $\forall A\in\mathcal{C}$，$X^{-1}(A)\in\mathcal{F}$.

定理 7.2.2 X 是可测空间 $(\Omega,\mathcal{F})$ 上的有限实值可测函数（或随机变量）当且仅当 $\forall x\in\mathbb{R}$，有
$$X^{-1}((-\infty,x])=\{x\in\Omega:X(\omega)\leq x\}\in\mathcal{F}.\tag{7.8}$$

7.2.2 可测函数与随机变量的扩展

根据前文阐述，可测映射的复合仍然为可测映射且连续映射可测，而可测函数是可测映射的特例，随机变量是可测函数的特例. 我们有推论是成立的.

性质 7.2.1 若 X 为 n 维有限值随机变量，f 是 $\mathbb{R}^n$ 到 $\mathbb{R}$ 上的连续函数（可以是复值函数），则 $f(X)$ 是随机变量. 特别地，若 X 和 Y 都是有限值随机变量，则
$$\begin{gathered}X^r(r\in\mathbb{Z}_+),\quad |X|^r(r\in\mathbb{R}_+),\quad e^{-\lambda X}(\lambda\in\mathbb{R}),\quad e^{itX}(t\in\mathbb{R}),\\ X\vee Y:=\max\{X,Y\},\quad X\wedge Y:=\min\{X,Y\},\\ X+Y,\quad X-Y,\quad XY,\quad X/Y(Y(\omega)\neq 0,\forall\omega\in\Omega)\end{gathered}\tag{7.9}$$
都是随机变量.

性质 7.2.2 若 X 为 n 维广义随机变量，f 是 $\overline{\mathbb{R}}^n$ 到 $\overline{\mathbb{R}}$ 上的连续函数，则 $f(X)$ 是广义可测函数. 特别地，若 X 和 Y 都是随机变量，则

$$\begin{gathered} X^r(r\in\mathbb{Z}_+),\quad |X|^r(r\in\mathbb{R}_+),\quad e^{-\lambda X}(\lambda\in\mathbb{R}),\\ X\vee Y:=\max\{X,Y\},\quad X\wedge Y:=\min\{X,Y\},\quad XY \end{gathered} \tag{7.10}$$

都是广义随机变量. 若

$$X+Y,\quad X-Y,\quad X/Y \tag{7.11}$$

有意义，则它们也都是可测函数.

7.2.3 可测函数的极限

定义 7.2.4 (**上下确界、上下极限**) 设 $\{f_n:n\in\mathbb{Z}_+\}$ 是定义在 Ω 上的广义实值可测函数列，$\forall\omega\in\Omega$，定义

$$\left(\sup_n f_n\right)(\omega):=\sup_n f_n(\omega), \tag{7.12}$$

$$\left(\inf_n f_n\right)(\omega):=\inf_n f_n(\omega), \tag{7.13}$$

$$\left(\varlimsup_{n\to\infty} f_n\right)(\omega)=\inf_n\left(\sup_{n\geq k} f_k(\omega)\right), \tag{7.14}$$

$$\left(\varliminf_{n\to\infty} f_n\right)(\omega)=\sup_n\left(\inf_{n\geq k} f_k(\omega)\right). \tag{7.15}$$

则 $\sup\limits_n f_n$，$\inf\limits_n f_n$，$\varlimsup\limits_{n\to\infty} f_n$ 和 $\varliminf\limits_{n\to\infty} f_n$ 分别称为序列 $\{f_n,n\in\mathbb{Z}_+\}$ 的上确界、下确界、上极限和下极限.

定义 7.2.5 (**极限**) 定义 f_n 的极限函数 $\lim\limits_{n\to\infty} f_n$ 为

$$\Delta:=\left\{\omega:\varlimsup_{n\to\infty} f_n(\omega)=\varliminf_{n\to\infty} f_n(\omega)\right\} \tag{7.16}$$

上的广义实值函数

$$\left(\lim_{n\to\infty} f_n\right)(\omega)\ =\lim_{n\to\infty} f_n(\omega). \tag{7.17}$$

定理 7.2.3 若 $\{f_n:n\in\mathbb{Z}_+\}$ 是 $(\Omega,\mathcal{F})$ 上的（广义）实可测函数序列，

$$\Delta:=\left\{\omega:\varlimsup_{n\to\infty} f_n(\omega)=\varliminf_{n\to\infty} f_n(\omega)\right\} \tag{7.18}$$

则

1. $\sup\limits_n f_n$，$\inf\limits_n f_n$，$\varlimsup\limits_{n\to\infty} f_n$ 和 $\varliminf\limits_{n\to\infty} f_n$ 都是 $(\Omega,\mathcal{F})$ 上的广义实可测函数.

2. $\lim\limits_{n\to\infty} f_n$ 是 Δ 上的广义实可测函数.

7.2.4 简单函数与初等函数

定义 7.2.6 (**简单函数**) 给定 $(\Omega,\mathcal{F})$，若 $A_k\in\mathcal{F}$，$k=1,2,\dots,n$ 两两不交且 $\bigcup\limits_{k=1}^{n}A_k=\Omega.a_k$，$k=1,2,\dots,n$ 为实数或 $\pm\infty$（或复数），称函数

$$f:=\sum_{k=1}^{n}a_kI_{A_k},\quad \text{即}f(\omega):=\sum_{k=1}^{n}a_kI_{A_k}(\omega),\omega\in\Omega \tag{7.19}$$

为 $\mathcal{F}$ 简单函数.

定义 7.2.7 (**初等函数**) 给定 $(\Omega,\mathcal{F})$，若 $A_n\in\mathcal{F}$，$n\in\mathbb{Z}_+$ 两两不交且 $\bigcup\limits_{n=1}^{\infty}A_n=\Omega.a_n$，$n\in\mathbb{Z}_+$ 为广义实数（或复数），称函数

$$f:=\sum_{n=1}^{\infty}a_nI_{A_n} \tag{7.20}$$

为 $\mathcal{F}$ 初等函数.

注 7.2.4 由定义可知，$\mathcal{F}$ 简单函数是 $\mathcal{F}$ 初等函数的特例.

定理 7.2.4 设 $(\Omega,\mathcal{F})$ 是可测空间，则 $\mathcal{F}$ 简单函数和 $\mathcal{F}$ 初等函数都是 $\mathcal{F}$ 可测函数.

例 7.2.1 设 $(\Omega,\mathcal{F},\mathbf{P})$ 是概率空间，$A_n\in\mathcal{F}$，$n\in\mathbb{Z}_+$ 两两不交，且 $\bigcup\limits_{n=1}^{\infty}A_n=\Omega$. 当 $\omega\in A_n$ 时，令 $X(\omega)=x_n$，即

$$X:=\sum_{n=1}^{\infty}x_nI_{A_n}. \tag{7.21}$$

则 X 为取值为 $\{x_n,n\in\mathbb{Z}_+\}$ 的离散随机变量，其分布列为

$$p_n:=\mathbf{P}(\{X=x_n\})=\mathbf{P}(A_n). \tag{7.22}$$

引理 7.2.1 设 $(\Omega,\mathcal{F})$ 是可测空间，f 和 g 是任意两个 $\mathcal{F}$ 简单函数（或 $\mathcal{F}$ 初等函数），则

1. $f+ig$ 是 $\mathcal{F}$ 简单函数（或 $\mathcal{F}$ 初等函数）;
2. $f+g$ 是 $\mathcal{F}$ 简单函数（或 $\mathcal{F}$ 初等函数）.

§7.3 可测函数的构造性质

7.3.1 简单函数逼近可测函数

定理 7.3.1 (**简单函数逼近可测函数**) 设 $(\Omega,\mathcal{F})$ 是可测空间，则

1. 任一 $\mathcal{F}$ 可测函数是 $\mathcal{F}$ 简单函数序列的极限;

2. 任一 $\mathcal{F}$ 可测函数是 $\mathcal{F}$ 初等函数序列的一致极限；

3. 任一有界 $\mathcal{F}$ 可测函数是 $\mathcal{F}$ 简单函数序列的一致极限；

4. 任一非负 $\mathcal{F}$ 可测函数是非负不降 $\mathcal{F}$ 简单函数（$\mathcal{F}$ 初等函数）序列的极限（一致极限）.

定义 7.3.1 设 $f:\Omega\to\overline{\mathbb{R}}$，$f$ 的正部和负部分别定义为

$$f^+=f\vee 0,\quad f^-=(-f)\vee 0=-(f\wedge 0). \tag{7.23}$$

注 7.3.1 显然有 $f=f^+-f^-$，$|f|=f^++f^-$.

注 7.3.2 f 为有限值函数时，

$$f^+=\frac{|f|+f}{2},\quad f^-=\frac{|f|-f}{2}. \tag{7.24}$$

定理 7.3.2 可测函数的正部和负部都是可测函数，从而任一可测函数可以表示成两个非负可测函数之差.

注 7.3.3 由上述定理7.3.2，我们在处理不是非负可测函数时，可以将其转化为正部和负部之差进行处理. 再根据定理7.3.1的第 4 点，我们可以用简单函数逼近一切可测函数.

7.3.2　函数形式单调类定理与 $\mathscr{L}$ 系方法

定义 7.3.2 ($\mathscr{L}$ **系**) 设 $\mathscr{L}$ 时定义在 Ω 上的广义实值函数类，满足条件：

$$f\in\mathscr{L}\Rightarrow f^+,f^-\in\mathscr{L}. \tag{7.25}$$

如果函数族 L 满足

1. $1\in L$；

2. L 中有限个函数的线性组合（如果有意义）属于 L；

3. 若 $f_n\in L$，$n\in\mathbb{Z}_+$，$0\geq f_n\uparrow f$，且 f 有界或 $f\in\mathscr{L}$，则 $f\in L$.

则称 L 为 $\mathcal{L}$ 系.

定理 7.3.3 (**函数形式单调类定理**) 若 $\mathscr{L}$ 系 L 包含一 π 系 $\mathcal{C}$ 中任一集合的示性函数，则 L 包含一切属于 $\mathscr{L}$ 的 $\sigma(\mathcal{C})$ 可测函数. 即

$$\{I_A,A\in\mathcal{C}\}\subset L\Rightarrow\{\sigma(\mathcal{C})\text{可测函数}\}\subset L. \tag{7.26}$$

方法 7.3.1 ($\mathscr{L}$ **系方法**) 要想证明某一函数族 F 具有某种性质 A_0，可以引入一个函数族 $\mathscr{L}$（满足 $f \in \mathscr{L} \Rightarrow f^+, f^- \in \mathscr{L}$）使

$$L := \{f : \text{函数} f \text{具有性质} A_0\} \tag{7.27}$$

是一个 $\mathscr{L}$ 系，再引入一个 π 系 $\mathcal{C}$，使得 $\mathscr{L}$ 中的 $\sigma(\mathcal{C})$ 可测函数类包含 F. 则只需证明 $\forall A \in \mathcal{C}$，$I_A \in L$ 即可证明 F 具有性质 A_0. 这种方法称为 $\mathscr{L}$ 系方法.

定理 7.3.4 设 Ω 为一集合，$(E, \mathcal{E})$ 是可测空间，设映射 $f : \Omega \to E$，令 $\sigma(f) = f^{-1}(\mathcal{E})$（它是 Ω 上的 σ 代数）. 则 φ 是 Ω 到 $\overline{\mathbb{R}}$ 上的 $\sigma(f)$ 可测函数的充分必要条件是：存在 $(E, \mathcal{E})$ 上的可测函数 g，使得

$$\varphi = g \circ f. \tag{7.28}$$

且若 φ 有限（有界），则可取 g 有限（有界）.

定理 7.3.5 设 $\mathscr{L}$ 是定义在 $\overline{\mathbb{R}}^n$ 上的实函数类，L 是 $\overline{\mathbb{R}}^n$ 上包含一切有界连续函数的 $\mathscr{L}$ 系，则 L 包含 $\overline{\mathbb{R}}^n$ 上一切 Borel 可测函数.

第五部分

高等数理统计

第六部分

多元统计分析

第八章　多元正态分布及其参数估计

§8.1　随机向量

8.1.1　随机向量与样本数据阵

定义 8.1.1 (**随机向量**) 多元统计分析中，我们讨论的是多变量总体，把 p 个随机变量放在一起得到的就是一个 p 维随机向量：

$$X=\begin{pmatrix}X_1\\X_2\\\vdots\\X_p\end{pmatrix}\tag{8.1}$$

定义 8.1.2 (**样本数据阵**) 如果同时对上述随机向量中的 p 个变量进行依次观测，得到的观测值：

$$X_{(1)}^{\top}=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1p}\end{pmatrix}\tag{8.2}$$

称为一个样品. 观察 n 次可以得到 n 个样品，记为：

$$X_{(i)}^{\top}=\begin{pmatrix}x_{i1} & x_{i2} & \cdots & x_{ip}\end{pmatrix},\quad i=1,2,\ldots,n\tag{8.3}$$

它们就构成了一个样本. 为了方便我们的研究，常把上述 n 个样品排列成一个 $n\times p$ 的矩阵，称为样本数据阵（或样本资料阵），记为：

$$X=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1p}\\x_{21} & x_{22} & \cdots & x_{2p}\\\vdots & \vdots & \ddots & \vdots\\x_{n1} & x_{n2} & \cdots & x_{np}\end{pmatrix}=\begin{pmatrix}X_{(1)}^{\top}\\X_{(2)}^{\top}\\\vdots\\X_{(n)}^{\top}\end{pmatrix}=(X_1,X_2,\ldots,X_p)\tag{8.4}$$

这是一个随机阵.

注 8.1.1 在多元统计分析之后的介绍中，如果不做特殊说明，我们谈及的向量默认为列向量，欲表示行向量，一般表示为列向量的转置的形式.

注 8.1.2 上述矩阵的第 i 行：

$$X_{(i)}^{\top} = (x_{i1}, x_{i2}, \cdots, x_{ip})(i = 1, 2, \cdots, n) \tag{8.5}$$

表示对第 i 个样本的观测值，在具体观测之前，它是一个 p 维随机向量.

注 8.1.3 矩阵的第 j 列：

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix} (j = 1, 2, \cdots, p) \tag{8.6}$$

表示对第 j 个变量的 n 次观测，在具体观测之前，它是一个 n 维随机向量.

8.1.2 随机向量的分布与独立性

定义 8.1.3 (**联合分布函数**) 设 $X = (X_1, X_2, \ldots, X_p)^{\top}$ 是 p 维随机向量，称 p 元函数

$$F(x_1, \ldots, x_p) = P\left\{X_1 \le x_1, \ldots, X_p \le x_p\right\} \tag{8.7}$$

为 X 的联合分布函数.

定义 8.1.4 (**连续型随机向量及其联合概率密度**) 设 $X = (X_1, X_2, \ldots, X_p)^{\top}$ 是 p 维随机向量，若存在非负函数 $f(x_1, \ldots, x_p)$，使得随机向量 X 的联合分布函数对一切 $(x_1, \ldots, x_p) \in \mathbb{R}^p$ 均可表示为：

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(x_1, \ldots, x_p)\, dx_1 \cdots dx_p \tag{8.8}$$

则称 X 为连续性随机向量，$f(x_1, \ldots, x_p)$ 称为 X 的联合概率密度函数，简称为多元密度函数.

定义 8.1.5 (**边缘分布**) 设 $X = (X_1, X_2, \ldots, X_p)^{\top}$ 是 p 维随机向量,称其部分分量 $(X_{i_1}, \ldots, X_{i_m})^{\top}(1 \le m < p)$ 的分布为边缘分布.

注 8.1.4 设 $X^{(1)}$ 为 r 维随机向量，$X^{(2)}$ 为 $p-r$ 维随机向量. 由二者组成的 p 维随机向量 $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$，则可求得 $X^{(1)}$ 的边缘分布为：

$$f_1(x^{(1)}) = f_1(x_1, \ldots, x_r) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_p)\, dx_{r+1} \cdots dx_p \tag{8.9}$$

$X^{(2)}$ 的边缘分布为：

$$f_2(x^{(2)}) = f_1(x_{r+1}, \ldots, x_p) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_p)\, dx_1 \cdots dx_r \tag{8.10}$$

定义 8.1.6 (**条件分布**) 设 $X^{(1)}$ 为 r 维随机向量，$X^{(2)}$ 为 $p-r$ 维随机向量. 由二者组成的 p 维随机向量 $X=\begin{pmatrix}X^{(1)}\\X^{(2)}\end{pmatrix}$，记 X 的分布为 $f(x^{(1)},x^{(2)})$，$X^{(2)}$ 的密度函数为 $f_2(x^{(2)})$，当给定 $X^{(2)}$ 时，称 $X^{(1)}$ 的分布为条件分布，记为：

$$f_1(x^{(1)}|x^{(2)})=\frac{f(x^{(1)},x^{(2)})}{f_2(x^{(2)})} \tag{8.11}$$

定义 8.1.7 (**随机向量的独立性**) 设 $X_1,\dots,X_p$ 时 p 维随机变量,X_i 的分布函数记为 $F_i(x_i)(i=1,\dots,p)$，它们的联合分布函数维 $F(x_1,\dots,x_p)$，则若对一切实数 $x_1,\dots,x_p$，均成立

$$F(x_1,\dots,x_p)=F_1(x_1)\dots F_p(x_p) \tag{8.12}$$

则称 $X_1,\dots,X_p$ 相互独立.

注 8.1.5 在连续型随机变量的场合下，$X_1,\dots,X_p$ 相互独立，当且仅当对一切实数 $x_1,\dots,x_p$，均成立

$$f(x_1,\dots,x_p)=f_1(x_1)\dots f_p(x_p) \tag{8.13}$$

$f(x_1,\dots,x_p)$ 为联合密度函数，$f_i(x_i)$ 为 X_i 的密度函数.

8.1.3 随机向量的数字特征

设 $X=(X_1,X_2,\cdots,X_p)^\top$ 和 $Y=(Y_1,Y_2,\cdots,Y_q)^\top$ 是两个随机向量. 则可以定义它们的数字特征如下：

定义 8.1.8 (**均值向量**) 若 $E(X_i)=\mu_i$ 存在，则称：

$$E(X)=\begin{pmatrix}E(X_1)\\E(X_2)\\\vdots\\E(X_p)\end{pmatrix}=\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_p\end{pmatrix} \tag{8.14}$$

为随机向量 X 的均值向量.

定义 8.1.9 (**随机向量 X 的协方差阵**) 若随机变量 X_i 和 X_j 的协方差 $Cov(X_i,X_j)$ 存在（$i,j=1,2,\cdots,p$），则称：

$$\begin{aligned}D(X)=&E\left\{[X-E(X)][X-E(X)]^\top\right\}\\=&\begin{pmatrix}Cov(X_1,X_1)&Cov(X_1,X_2)&\cdots&Cov(X_1,X_p)\\Cov(X_2,X_1)&Cov(X_2,X_2)&\cdots&Cov(X_2,X_p)\\\vdots&\vdots&\ddots&\vdots\\Cov(X_p,X_1)&Cov(X_p,X_2)&\cdots&Cov(X_p,X_p)\end{pmatrix}=(\sigma_{ij})_{p\times p}=\Sigma\end{aligned} \tag{8.15}$$

为随机向量 X 的协方差阵.

注 8.1.6 上述定义中 $Var(X_i) = Cov(X_i, X_i) = \sigma_{ii}$ 为 X_i 的方差.

注 8.1.7 当随机向量 X 的协方差阵中除主对角线上的元素外，其余元素均为 0 时，称随机向量 X 的各分量之间互不相关.

注 8.1.8 若随机向量 X 的协方差阵中的主对角线上的元素出现 0，则该元素对应的 X 的分量将退化为一常数，若主对角线上的元素全为零，则此时的 X 不再是随机向量，而退化为一个常数向量.

性质 8.1.1 随机向量 X 的协方差阵 $D(X) = \Sigma$ 满足：

- Σ 是对称的非负定矩阵；
- $\Sigma = L^2$，其中 L 为非负定矩阵.

定义 8.1.10 (**随机向量 X 与 Y 的协方差阵**) 若 X_i 和 Y_j 的协方差 $Cov(X_i, Y_j)$ 存在（$i = 1, \cdots, p; j = 1, \cdots, q$），则称

$$Cov(X, Y) = E\left\{[X - E(X)][Y - E(Y)]^\top\right\} = \begin{pmatrix} Cov(X_1, Y_1) & Cov(X_1, Y_2) & \cdots & Cov(X_1, Y_q) \\ Cov(X_2, Y_1) & Cov(X_2, Y_2) & \cdots & Cov(X_2, Y_q) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, Y_1) & Cov(X_p, Y_2) & \cdots & Cov(X_p, Y_q) \end{pmatrix} \tag{8.16}$$

为随机向量 X 与 Y 的协方差阵.

注 8.1.9 若满足 $Cov(X, Y) = O$，则称 X 与 Y 不相关. 与一元数理统计中相同，在多元统计分析中，若两个随机向量相互独立，则可以推出二者的协方差阵为 $O_{p\times q}$，但反之不成立.

性质 8.1.2 设 X 和 Y 是随机向量，A 和 B 是常数矩阵，则有：

- $E(AX) = AE(X)$
- $E(AXB) = AE(X)B$
- $D(AX) = AD(X)A^\top$
- $Cov(AX, BY) = ACov(X, Y)B^\top$

定义 8.1.11 (**随机向量 X 的相关阵**) 若 X_i 和 X_j 的协方差 $Cov(X_i, X_j)$ 存在（$i, j = 1, 2, \cdots, p$），则称：

$$R = (r_{ij})_{p\times p} \tag{8.17}$$

为 X 的相关阵，其中：

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad (i, j = 1, 2, \cdots, p) \tag{8.18}$$

§8.2　多元正态分布的定义与基本性质

8.2.1　多元正态分布的定义

定义 8.2.1 (**由标准正态分布定义**) 设为 $U = (U_1, \ldots, U_q)^\top$ 随机向量，$U_1, \ldots, U_q$ 相互独立且同分布于标准正态分布；设 μ 为 p 维常数向量，A 为常数矩阵，则称 $X = AU + \mu$ 的分布为 p 元正态分布，或称 X 为 p 维正态随机向量，记为 $X \sim N_p(\mu, AA^\top)$.

定义 8.2.2 (**由特征函数定义**) 若 p 维随机向量的特征函数满足：

$$\Phi_X(t) = \exp\left\{it^\top\mu - \frac{1}{2}t^\top\Sigma t\right\}, \quad \Sigma > 0 \tag{8.19}$$

则称其服从 p 元正态分布，记为 $X \sim N_p(\mu, \Sigma)$.

定义 8.2.3 (**由密度函数定义**) 若随机向量 $X = (X_1, X_2, \cdots, X_p)^\top$ 的联合密度函数为：

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right\} \tag{8.20}$$

其中 μ 为 p 维常数向量，Σ 为 p 阶对称正定矩阵，则称 X 服从（非退化的）p 维正态分布，也称其为 p 维正态随机向量.

定义 8.2.4 若 p 维随机向量 X 各分量的任意线性组合均服从一元正态分布，则称 X 为 p 维正态随机向量.

8.2.2　多元正态分布的基本性质

性质 8.2.1 设 $U = (U_1, U_2, \ldots, U_q)^\top$ 为随机变量，$U_1, U_2, \ldots, U_q$ 相互独立且同 $N(0,1)$ 分布；令 $X = AU + \mu$，则 X 的特征函数为：

$$\Phi_X(t) = \exp\left\{it^\top\mu - \frac{1}{2}t^\top AA^\top t\right\} \tag{8.21}$$

性质 8.2.2 正态分布的任意线性组合仍然为正态分布：

1. 设 $X \sim N_p(\mu, \sum)$，B 为 $s \times p$ 维常数矩阵，d 为 s 维的常数向量，令 $Z = BX + d$，则 $Z \sim N_s(B\mu + d, B\sum B^\top)$.

2. 将多为正态分布的均值 μ 和协方差 Σ 进行对应分块，则所得每个小块对应的随机变量分量亦服从正态分布：

$$X=\begin{pmatrix}X^{(1)}\\X^{(2)}\end{pmatrix},\quad \mu=\begin{pmatrix}\mu^{(1)}_{(r\times 1)}\\\mu^{(2)}_{(p-r\times 1)}\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma^{11}_{(r\times r)} & \Sigma^{12}_{(r\times p-r)}\\\Sigma^{21}_{(p-r\times r)} & \Sigma^{22}_{(p-r\times p-r)}\end{pmatrix} \tag{8.22}$$

则 $X^{(1)}\sim N_r(\mu^{(1)},\Sigma^{11})$，$X^{(2)}\sim N_{p-r}(\mu^{(2)},\Sigma^{22})$.

3. 多元正态分布的边缘分布仍为正态分布，但反之未必成立.

性质 8.2.3 设 $X\sim N_p(\mu,\Sigma)$，则 $E(X)=\mu,D(X)=\Sigma$.

性质 8.2.4 设 $X=(X_1,\cdot\cdot\cdot,X_p)^\top$ 为 p 维随机向量，则 X 服从 p 维正态分布的充分必要条件是：对任一 p 维实向量 a，$\xi=a^\top X$ 是一维正态分布随机变量.

注 8.2.1 讨论二维正态随机向量 $X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim N_2(\mu,\Sigma)$，又记：

$$\mu=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}\quad \Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \tag{8.23}$$

$\mu_1,\mu_2,|\rho|<1,\sigma_1>0,\sigma_2>0$ 均为常数，则其联合密度函数为：

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\times[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\right\} \tag{8.24}$$

并且可以得到：$X_1\sim N(\mu_1,\sigma_1^2)$，$X_2\sim N(\mu_2,\sigma_2^2)$，$\rho$ 是 X_1 和 X_2 的相关系数. 显然：

1. 当 $\rho=0$ 时，X_1 和 X_2 相互独立；
2. 当 $\rho>0$ 时，X_1 和 X_2 存在正相关；
3. 当 $\rho<0$ 时，X_1 和 X_2 存在负相关；
4. 当 $|\rho|=1$ 时，X_1 和 X_2 以概率 1 存在线性相关关系，此时 Σ 退化（$|\Sigma|=0$）.

§8.3　多元正态分布的条件分布和独立性

8.3.1　独立性

定理 8.3.1 设 p 维随机向量 $X\sim N_p(\mu,\Sigma)$，

$$X=\begin{pmatrix}X^{(1)}\\X^{(2)}\end{pmatrix}\sim N_p\left(\begin{pmatrix}\mu^{(1)}\\\mu^{(2)}\end{pmatrix},\begin{pmatrix}\Sigma^{11} & \Sigma^{12}\\\Sigma^{21} & \Sigma^{22}\end{pmatrix}\right) \tag{8.25}$$

则 $X^{(1)}$ 与 $X^{(2)}$ 相互独立的充分必要条件是 $\Sigma^{12}=\Sigma^{21}=O$.（即 $X^{(1)}$ 与 $X^{(2)}$ 不相关）.

引理 8.3.1 设 $r_i \geq 1(i=1,\cdots,k)$，且 $r_1+r_2+\cdots+r_k=p$，有：

$$X=\begin{pmatrix} X_{r_1}^{(1)} \\ X_{r_2}^{(2)} \\ \vdots \\ X_{r_k}^{(k)} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_{r_1}^{(1)} \\ \mu_{r_2}^{(2)} \\ \vdots \\ \mu_{r_k}^{(k)} \end{pmatrix}, \begin{pmatrix} \Sigma^{11} & \Sigma^{12} & \cdots & \Sigma^{1k} \\ \Sigma^{21} & \Sigma^{22} & \cdots & \Sigma^{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma^{k1} & \Sigma^{k2} & \cdots & \Sigma^{kk} \end{pmatrix}\right) \tag{8.26}$$

则 $X^{(1)},X^{(2)},\cdots,X^{(k)}$ 相互独立的充分必要条件是 $\Sigma^{ij}=O(i\neq j)$.

引理 8.3.2 设 $X=(X_1,X_2,\cdots,X_p)^\top \sim N_p(\mu,\Sigma)$，若 Σ 为对角阵，则 $X_1,X_2,\cdots,X_p$ 相互独立.

8.3.2 条件分布

二维场合下，设 $X=(X_1,X_2)^\top$ 为二维随机向量，则有：

$$(X_1|X_2) \sim N_1\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2),\sigma_1^2(1-\rho^2)\right) \tag{8.27}$$

利用 Σ^{-1} 的分块求逆公式，将其推广到 p 元情况有：

$$\Sigma^{-1}=\begin{pmatrix} \Sigma_{11\cdot 2}^{-1} & -\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot 2}^{-1} & \Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix} \quad (\Sigma_{11\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \tag{8.28}$$

故有如下结论成立：

定理 8.3.2 设随机向量满足：

$$X=\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \tag{8.29}$$

则当 $X^{(2)}$ 给定时，$X^{(1)}$ 的条件分布为：

$$(X^{(1)}|X^{(2)}) \sim N_r(\mu_{1\cdot 2},\Sigma_{11\cdot 2}) \tag{8.30}$$

其中：

$$\mu_{1\cdot 2}=\mu^{(1)}+\Sigma_{12}{\Sigma_{22}}^{-1}(X^{(2)}-\mu^{(2)}) \tag{8.31}$$

$$\Sigma_{11\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22-1}\Sigma_{21} \tag{8.32}$$

注 8.3.1 在上述定理的条件下可以得到：

1. $X^{(2)}$ 与 $X^{(1)} - \Sigma_{12}\Sigma_{22}{}^{-1}X^{(2)}$ 相互独立；
2. $X^{(1)}$ 与 $X^{(2)} - \Sigma_{21}\Sigma_{11}{}^{-1}X^{(1)}$ 相互独立；
3. $(X^{(2)}|X^{(1)}) \sim N_{p-r}(\mu_{2\cdot 1}, \Sigma_{22\cdot 1})$，其中：

$$\mu_{2\cdot 1} = \mu^{(2)} + \Sigma_{21}\Sigma_{11}{}^{-1}(X^{(1)} - \mu^{(1)}) \tag{8.33}$$

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}{}^{-1}\Sigma_{12} \tag{8.34}$$

§8.4　多元正态总体的参数估计

8.4.1　多元正态总体样本的数字特征

首先引入多元正态总体样本的相关量：

定义 8.4.1 (**样本均值向量**) 样本均值向量定义为：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_{(i)} = (\overline{x}_1, \cdots, \overline{x}_p)^\top = \frac{1}{n}X^\top \mathbf{1}_n \tag{8.35}$$

其中：

$$\overline{x}_i = \frac{1}{n}\sum_{a=1}^{n} x_{ai}(i = 1, 2, \cdots, p) \tag{8.36}$$

定义 8.4.2 (**样本离差阵**) 样本离差阵定义为：

$$A = \sum_{a=1}^{n}(X_{(a)} - \overline{X})(X_{(a)} - \overline{X})^\top = X^\top X - n\overline{X}\overline{X}^\top = X^\top\left[I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right]X = (a_{ij})_{p\times p} \tag{8.37}$$

定义 8.4.3 (**样本协方差阵**) 样本协方差阵定义为：

$$S = \frac{1}{n-1}A = (s_{ij})_{p\times p} \tag{8.38}$$

或有另外一种定义方式：

$$S^* = \frac{1}{n}A \tag{8.39}$$

一般地：

$$s_{ii} = \frac{1}{n-1}\sum_{a=1}^{n}(x_{ai} - \overline{x}_i)^2 \quad (i = 1, 2, \cdots, p) \tag{8.40}$$

称为变量 X_i 的样本方差；样本方差的平方根 $\sqrt{s_{ii}}$ 称为 X_i 样本标准差.

定义 8.4.4 (**样本相关阵**) 样本相关阵定义为：

$$R = (r_{ij})_{p \times p} \tag{8.41}$$

其中：

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = \frac{a_{ij}}{\sqrt{a_{ii}}\sqrt{a_{jj}}} \quad (i, j = 1, 2, \cdots, p) \tag{8.42}$$

8.4.2 多元正态总体参数的极大似然估计

定理 8.4.1 设 $X_{(i)}(i = 1, \cdots, n)$ 是多元正态总体 $N_p(\mu, \Sigma)$ 的随机样本，$n > p$，则 μ, Σ 的极大似然估计为：

$$\hat{\mu} = \overline{X}, \quad \hat{\Sigma} = \frac{1}{n}A \tag{8.43}$$

定理 8.4.2 设 $\overline{X}$ 和 A 分别为 p 元正态总体 $N_p(\mu, \Sigma)$ 的样本均值向量和样本离差阵，则：

1. $\overline{X} \sim N_p\left(\mu, \dfrac{1}{n}\Sigma\right)$
2. $A = \sum\limits_{t=1}^{n}(X_{(t)} - \overline{X})(X_{(t)} - \overline{X})^\top = \sum\limits_{t=1}^{n-1} Z_t Z_t^\top$，其中 $Z_1, \cdots, Z_{n-1}$ 独立同分布于 $N_p(0, \Sigma)$.
3. $\overline{X}$ 与 A 相互独立
4. $P(A > 0) = 1 \Longleftrightarrow n > p$

性质 8.4.1 根据估计量好坏的评判准则，多元正态总体参数的极大似然估计具有如下性质：

1. 无偏性：$\hat{\mu}$ 是 μ 的无偏估计，但 $\hat{\Sigma}$ 不是 Σ 的无偏估计. 为寻找 Σ 的无偏估计，可考虑 $S = \dfrac{1}{n-p}A$，它是 Σ 的无偏估计.
2. 有效性：上述极大似然估计是 μ 和 Σ 的最小方差无偏估计，即是有效估计量.
3. 相合性（一致性）：当 $n \to \infty$ 时，上述极大似然估计是 μ 和 Σ 的强相合估计量.
4. 还可以证明上述极大似然估计的充分统计量，均值的极大似然估计是极小极大似然估计.
5. 估计量还具有渐近正态性.

§8.5 习题

习题 8.5.1 设三维随机向量 $X \sim N_3(\mu, 2I_3)$，已知：

$$\mu = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, A = \begin{pmatrix} 0.5 & -1 & 0.5 \\ -0.5 & 0 & -0.5 \end{pmatrix}, d = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{8.44}$$

试求 $Y = AX + d$ 的分布.

习题 8.5.2 设 $X = (X_1, X_2)^\top \sim N_2(\mu, \Sigma)$，其中

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{8.45}$$

1. 试证明 $X_1 + X_2$ 和 $X_1 - X_2$ 相互独立；
2. 试求 $X_1 + X_2$ 和 $X_1 - X_2$ 的分布.

习题 8.5.3 设 $X^{(1)}$ 和 $X^{(2)}$ 均为 p 维随机向量，已知

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_1 \end{pmatrix} \right) \tag{8.46}$$

其中 $\mu^{(i)}(i = 1, 2)$ 为 p 维向量，$\Sigma_i(i = 1, 2)$ 维 p 阶矩阵.

1. 试证明 $X^{(1)} + X^{(2)}$ 和 $X^{(1)} - X^{(2)}$ 相互独立；
2. 试求 $X^{(1)} + X^{(2)}$ 和 $X^{(1)} - X^{(2)}$ 的分布.

习题 8.5.4 设 $X \sim N_3(\mu, \Sigma)$，其中

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \mu = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix} \tag{8.47}$$

1. 试求 $3X_1 - 2X_2 + X3$ 的分布；
2. 求二维向量 $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$，使 X_3 与 $X_3 - a^\top \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ 相互独立.

习题 8.5.5 设 $X \sim N_3(\mu, \Sigma)$，其中

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \mu = \begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix} \tag{8.48}$$

试判断以下五对随机变量中哪些是相互独立的：

- X_1 与 $2X_2$；
- X_2 与 X_3；
- $(X_1, X_2)^\top$ 与 X_3；
- $\dfrac{1}{2}(X_1 + X_2)$ 与 X_3；
- X_2 与 $X_2 - \dfrac{5}{2}X_1 - X_3$

习题 8.5.6 设

$$A = \begin{pmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{pmatrix} \tag{8.49}$$

1. 试证明 A 是一个正交矩阵；

2. 已知 $X \sim N_4(\mu \mathbf{1}_4, \sigma^2 I_4)$，设 $Y = (Y_1, Y_2, Y_3, Y_4)^\top = AX$，试证明：

   - $Y_2^2 + Y_3^2 + Y_4^2 = \sum_{i=1}^{4}(X_i - \overline{X})^2$，其中 $\overline{X} = \frac{1}{4}\sum_{j=1}^{4} X_j$；
   - $Y_i(i = 1, 2, 3, 4)$ 相互独立；
   - $Y_1 \sim N(2\mu, \sigma^2)$，$Y_i \sim N(0, \sigma^2)(i = 2, 3, 4)$.

习题 8.5.7 设 $X_1 \sim N(0, 1)$，令

$$X_2 = \begin{cases} -X_1 & , -1 \le X_1 \le 1 \\ X_1 & , others \end{cases} \tag{8.50}$$

1. $X_2 \sim N(0, 1)$；

2. (X_1, X_2) 不是二元正态分布.

习题 8.5.8 已知样本数据阵为

$$X = \begin{pmatrix} 7 & 12 \\ 6 & 13 \\ 8 & 12 \\ 7 & 11 \end{pmatrix} \tag{8.51}$$

试求样本均值和样本协方差矩阵.

习题 8.5.9 正态随机向量 $X = (X_1, X_2, X_3)^\top$ 的协方差矩阵为

$$\Sigma = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 13 & -1 \\ 0 & -1 & 25 \end{pmatrix} \tag{8.52}$$

请给出三对独立的随机变量.

第九章　多元正态总体参数的假设检验

§9.1　几个重要统计量的分布

9.1.1　三大抽样分布的一般情形

数理统计中提到了三大抽样分布：χ^2 分布、t 分布和 F 分布. 这里提到的三个分布均为中心化的分布. 为了给出多元场合下的重要统计量的分布，本节首先将三大抽样分布作一下非中心化推广.

我们首先讨论分量独立的 n 维正态随机向量的二次型的分布，设 $X_i \sim N(\mu_i, \sigma^2)(i = 1, \ldots, n)$，且相互独立，记

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \tag{9.1}$$

则 $X \sim N_n(\mu, \sigma^2 I_n)$，其中 $\mu = (\mu_1, \ldots, \mu_n)^\top$，我们可以发现：

1. 当 $\mu_i = 0(i = 1, \ldots, n)$，$\sigma^2 = 1$ 时，有：

$$\xi = X^\top X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n) \tag{9.2}$$

2. 当 $\mu_i = 0(i = 1, \ldots, n)$，$\sigma^2 \neq 1$ 时，有：

$$\frac{1}{\sigma^2}\xi = \frac{1}{\sigma^2}X^\top X = \frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 \sim \chi^2(n) \quad \text{或记} \quad X^\top X \sim \sigma^2\chi^2(n) \tag{9.3}$$

3. 当 $\mu_i \neq 0(i = 1, \ldots, n)$ 时，$X^\top X$ 的分布一般称为非中心化的 χ^2 分布.

定义 9.1.1 (非中心 χ^2 分布) 设 n 维随机向量 $X \sim N_n(\mu, I_n)(\mu \neq 0)$，则称随机变量 $\xi = X^\top X$ 为服从 n 个自由度，非中心参数 $\delta = \mu^\top\mu = \sum_{i=1}^{n}\mu_i^2$ 的 χ^2 分布，记为 $X^\top X \sim \chi^2(n, \delta)$ 或 $X^\top X \sim \chi_n^2(\delta)$.

此时对于 n 维随机向量 $X \sim N_n(\mu, I_n\sigma^2)$，当 $\mu \neq 0, \sigma^2 \neq 1$ 时，可令：

$$Y_i = \frac{1}{\sigma}X_i \sim N\left(\frac{\mu_i}{\sigma}, 1\right)(i = 1, \ldots, n) \tag{9.4}$$

则此时随机变量 $Y^\top Y$ 为服从 n 个自由度，非中心参数 $\delta=\dfrac{1}{\sigma^2}\mu^\top\mu$ 的 χ^2 分布，可记为

$$Y^\top Y=\frac{1}{\sigma^2}X^\top X\sim\chi_n^2(\delta) \tag{9.5}$$

进一步，对于一般的 p 维随机向量，设 $X\sim N_p(\mu,\Sigma)$，$\Sigma>0$，则 $X^\top\Sigma^{-1}X\sim\chi^2(p,\delta)$，其中 $\delta=\mu^\top\Sigma^{-1}\mu$. 这个结论是最一般的结论，上面的各种情形均可视作该结论的特殊情况.

定义 9.1.2 (**非中心** t **分布**) 设 $X\sim N(\delta,1)$ 与 $Y\sim\chi^2(n)$ 相互独立，令

$$T=\frac{X\sqrt{n}}{\sqrt{Y}} \tag{9.6}$$

则称 T 的分布维具有 n 个自由度、非中心参数为 δ 的非中心 t 分布，记为 $T\sim t(n,\delta)$.

定义 9.1.3 (**非中心** F **分布**) 设 $X\sim\chi^2(m,\delta)$ 与 $Y\sim\chi^2(n)$ 相互独立，令

$$F=\frac{X/m}{Y/n} \tag{9.7}$$

则称 F 的分布为具有自由度 m，n 和非中心参数为 δ 的 F 分布，记为 $F\sim F(m,n,\delta)$.

9.1.2　威沙特分布

威沙特分布是一元统计中 χ^2 分布的推广. 设 $X_{(\alpha)}(\alpha=1,2,\dots,n)$ 为来自总体 $N_p(0,\Sigma)$ 的随机样本，记 $X=(X_{(1)},\dots,X_{(n)})^\top$ 为 $n\times p$ 样本数据阵. 考虑如下随机阵的分布：

$$W=\sum_{i=1}^n X_{(i)}X_{(i)}^\top=\left(X_{(1)},\dots,X_{(n)}\right)\begin{pmatrix}X_{(1)}^\top\\ \vdots\\ X_{(n)}^\top\end{pmatrix}=X^\top X \tag{9.8}$$

当 $p=1$ 时，对应一元正态分布的情况，有 $W\sim\sigma^2\chi^2(n)$. 该结果推广到 p 元正态总体时，即得到威沙特分布.

定义 9.1.4 (**威沙特（**Wishart**）分布**) 设 $X_{(\alpha)}\sim N(0,\Sigma)(\alpha=1,\dots,\Sigma)$ 相互独立，记 $X=(X_{(1)},\dots,X_{(n)})^\top$ 为 $n\times p$ 矩阵，则称：

$$W=\sum_{\alpha=1}^n X_{(\alpha)}X_{(\alpha)}^\top=X^\top X \tag{9.9}$$

的分布为威沙特分布，记为 $W\sim W_p(n,\Sigma)$.

取 $p=1$ 时，有：

$$W=\sum_{\alpha=1}^n X_{(\alpha)}^2\sim\sigma^2\chi^2(n) \tag{9.10}$$

即 $W_1(n,\sigma^2)$ 就是 $\sigma^2\chi^2(n)$. 进一步，取 $p=1,\sigma^2=1$ 时，$W_1(n,1)$ 就是 $\chi^2(n)$

注 9.1.1 设 $X_{(\alpha)} \sim N_p(\mu,\Sigma)(\alpha=1,\ldots,n)$ 相互独立，则样本离差阵 A 服从威沙特分布，即：

$$A=\sum_{\alpha=1}^{n}\left(X_{(\alpha)}-\overline{X}\right)\left(X_{(\alpha)}-\overline{X}\right)^{\top} \sim W_p(n-1,\Sigma) \tag{9.11}$$

注 9.1.2 威沙特分布关于自由度 n 具有可加性：设 $W_i \sim W_p(n_i,\Sigma)(i=1,\ldots,k)$ 相互独立，则：

$$\sum_{i=1}^{k} W_i \sim W_p(n,\Sigma), \quad \text{其中} n=n_1+\cdots+n_k \tag{9.12}$$

注 9.1.3 设 p 阶随机阵 $W \sim W_p(n,\Sigma)$，C 是 $m\times p$ 常数矩阵，则 m 阶随机阵 $CWC^{\top}$ 也服从威沙特分布，即：

$$CWC^{\top} \sim W_m(n, C\Sigma C^{\top}) \tag{9.13}$$

注 9.1.4 设随机阵 $W \sim W_p(n,\Sigma)$，则 $E(W)=n\Sigma$.

9.1.3 霍特林分布

定义 9.1.5 (**霍特林（**Hotelling**）**T^2 **分布**) 设总体 $X \sim N_p(0,\Sigma)$，随机阵 $W \sim W_p(n,\Sigma)$，$\Sigma > 0, n \geq p$，且 X 与 W 相互独立，则称统计量 $T^2 = nX^{\top}W^{-1}X$ 为霍特林 T^2 统计量，其分布称为服从 n 个自由度的 T^2 分布，记为：

$$T^2 \sim T^2(p,n) \tag{9.14}$$

更一般地，若 $X \sim N_p(\mu,\Sigma)(\mu \neq 0)$，则称 T^2 的分布为非中心霍特林 T^2 分布，记为 $T^2 \sim T^2(p,n,\mu)$.

注 9.1.5 设 $X_{(\alpha)}(\alpha=1,\ldots,n)$ 是来自 p 元总体 $N_p(\mu,\Sigma)$ 的随机样本，$\overline{X}$ 和 A 分别是正态总体 $N_p(\mu,\Sigma)$ 的样本均值向量和样本离差阵，则有：

$$\begin{aligned} T^2 &= (n-1)[\sqrt{n}(\overline{X}-\mu)]^{\top}A^{-1}[\sqrt{n}(\overline{X}-\mu)] \\ &= n(n-1)(\overline{X}-\mu)^{\top}A^{-1}(\overline{X}-\mu) \\ &\sim T^2(p,n-1) \end{aligned} \tag{9.15}$$

该形式可记作：

$$(\text{多元正态随机向量})^{\top}\left(\frac{\text{威沙特随机矩阵}}{\text{自由度}}\right)^{-1}(\text{多元正态随机向量}). \tag{9.16}$$

这实际上是一元统计中 $t^2=\sqrt{n}(\overline{x}-\mu)(s^2)^{-1}\sqrt{n}(\overline{x}-\mu)$ 的多元推广结论.

注 9.1.6 设 $T^2 \sim T^2(p,n)$，则 T^2 与 F 分布存在如下关系：

$$\frac{n-p+1}{np}T^2 \sim F(p,n-p+1) \tag{9.17}$$

注 9.1.7 设 $X_{(\alpha)}(\alpha=1,\ldots,n)$ 为来自 p 元总体 $N_p(\mu,\Sigma)$ 的随机样本.$\overline{X}$，A 分别为样本均值向量和样本离差阵，记

$$T^2 = n(n-1)\overline{X}^\top A^{-1}\overline{X} \tag{9.18}$$

则有：

$$\frac{n-p}{p}\frac{T^2}{n-1} \sim F(p,n-p,\delta) \tag{9.19}$$

其中 $\delta = n\mu^\top\Sigma\mu$.

注 9.1.8 T^2 统计量的分布只与 p,n 有关，而与 Σ 无关.

9.1.4　威尔克斯 Λ 统计量及其分布

在 p 元总体 $N_p(\mu,\Sigma)$ 中，协方差 Σ 的估计量为：

$$\hat{\Sigma} = \frac{1}{n-1}A\left(\text{或}\frac{1}{n}A\right) \tag{9.20}$$

我们可以用矩阵的行列式、迹或特征值等数量指标来描述总体的分散程度.

定义 9.1.6 (**样本广义方差**) 设 $X \sim N_P(\mu,\Sigma)$，则称协方差阵的行列式 $|\Sigma|$ 为 X 的广义方差.若 $X_{(\alpha)}(\alpha=1,\ldots,n)$ 为 p 元总体 X 的随机样本，A 为样本离差阵，则称 $\left|\frac{1}{n}A\right|$ 或 $\left|\frac{1}{n-1}A\right|$ 为样本广义方差.

有了广义方差的概念之后，在多元统计的协方差阵齐性检验中，类似一元统计，可考虑两个广义方差之比构成的统计量——威尔克斯统计量的分布.

定义 9.1.7 (**威尔克斯统计量**) 设 $A_1 \sim W_p(n_1,\Sigma)$，$A_2 \sim W_p(n_2,\Sigma)$，其中 $\Sigma>0$，$n_1 \geq p$，且 A_1 与 A_2 独立，则称广义方差之比：

$$\Lambda = \frac{|A_1|}{|A_1+A_2|} \tag{9.21}$$

为威尔克斯统计量或 Λ 统计量，其分布称为威尔克斯分布，记为：

$$\Lambda \sim \Lambda(p,n_1,n_2) \tag{9.22}$$

注 9.1.9 当 $p=1$ 时，Λ 统计量的分布便是一元统计中的参数为 $\frac{n_1}{2}$、$\frac{n_2}{2}$ 的 β 分布.

注 9.1.10 Λ 统计量与 T^2 或 F 统计量之间有如下关系：

1. 当 $n_2=1$ 时，设 $n_1=n>p$，则：
$$\Lambda(p,n,1)=\frac{1}{1+\frac{1}{n}T^2(p,n)} \tag{9.23}$$
或写作：
$$T^2(p,n)=n\cdot\frac{1-\Lambda(p,n,1)}{\Lambda(p,n,1)} \tag{9.24}$$
$$\frac{n-p+1}{np}T^2=\frac{n-p+1}{p}\frac{1-\Lambda}{\Lambda}=F(p,n-p+1) \tag{9.25}$$

2. 当 $n_2=2$ 时，设 $n_1=n>p$，则：
$$\frac{n-p+1}{p}\frac{1-\sqrt{\Lambda(p,n,2)}}{\sqrt{\Lambda(p,n,2)}}=F(2p,2(n-p+1)) \tag{9.26}$$

3. 当 $p=1$ 时，则：
$$\frac{n_1}{n_2}\frac{1-\sqrt{\Lambda(1,n_1,n_2)}}{\sqrt{\Lambda(1,n_1,n_2)}}=F(n_2,n_1) \tag{9.27}$$

4. 当 $p=2$ 时，则：
$$\frac{n_1-1}{n_2}\cdot\frac{1-\sqrt{\Lambda(2,n_1,n_2)}}{\sqrt{\Lambda(2,n_1,n_2)}}=F(2n_2,2(n_1-1)) \tag{9.28}$$

5. 当 $n_2>2,p>2$ 时，可用 χ^2 统计量或 F 统计量近似. 设 $\Lambda\sim\Lambda(p,n_1,n_2)$，则当 $n\to\infty$ 时，$-r\ln\Lambda\sim\chi^2(pn_2)$，其中 $r=n_1-\frac{1}{2}(p-n_2+1)$.

6. 当 n 不太大时，也有一些近似分布.

注 9.1.11 若 $\Lambda\sim\Lambda(p,n_1,n_2)$，则存在 $B_k\sim Be\left(\frac{n_1-p+k}{2},\frac{n_2}{2}\right)(k=1,\ldots,p)$ 相互独立，使得
$$\Lambda=B_1B_2\cdots B_p \tag{9.29}$$

注 9.1.12 若 $n_2<p$，则
$$\Lambda(p,n_1,n_2)=\Lambda(n_2,p,n_1+n_2-p) \tag{9.30}$$
该结论是一元统计中 $F(n,m)=\frac{1}{F(m,n)}$ 的推广.

§9.2　均值向量的检验

检验类型	检验情形	检验统计量
单正态总体均值向量检验	$\Sigma=\Sigma_0$ 已知	$T_0^2=n(\overline{X}-\mu_0)^\top\Sigma_o^{-1}(\overline{X}-\mu_0)\sim\chi^2(p)$
	Σ 未知	$T^2=n(\overline{X}-\mu_0)^\top\left(\frac{1}{n-1}A\right)^{-1}(\overline{X}-\mu_0)\sim T^2(p,n-1)$ $F=\frac{(n-1)-p+1}{(n-1)p}T^2\sim F(p,n-p)$
两正态总体均值向量检验	两总体协方差阵未知但相等	$T^2=\frac{nm}{n+m}(\overline{X}-\overline{Y})^\top\left(\frac{A_1+A_2}{n+m-2}\right)^{-1}(\overline{X}-\overline{Y})\sim T^2(p,n+m-2)$ 其中 $\begin{cases}A_1=\sum\limits_{\alpha=1}^{n}(X_{(\alpha)}-\overline{X})(X_{(\alpha)}-\overline{X})^\top\sim W_p(n-1,\Sigma)\\ A_2=\sum\limits_{\alpha=1}^{m}(Y_{(\alpha)}-\overline{Y})(Y_{(\alpha)}-\overline{Y})^\top\sim W_p(m-1,\Sigma)\end{cases}$ $F=\frac{(n+m-2)-p+1}{(n+m-2)p}T^2\sim F(p,n+m-p-1)$
	两总体协方差阵不等 （样本量 $n=m$）	构造成对数据 $Z_{(i)}=X_{(i)}-Y_{(i)}$，进行单总体检验
	两总体协方差阵不等 （样本量 $n\neq m$ 但相差不大，假设 $n<m$）	构造$Z_{(i)}=X_{(i)}-\sqrt{\frac{n}{m}}Y_{(i)}+\frac{1}{\sqrt{nm}}\sum\limits_{j=1}^{n}Y_{(j)}-\frac{1}{m}\sum\limits_{j=1}^{m}Y_{(j)}$ $(i=1,2,\ldots,n)$，进行单总体检验
	两总体协方差阵不等 （样本量 $n\neq m$ 且相差较大）	可构造近似检验统计量进行检验
多个正态总体均值向量的检验 （多元方差分析）		$\Lambda=\frac{\lvert A\rvert}{\lvert A+B\rvert}=\frac{\lvert A\rvert}{\lvert T\rvert}\sim\Lambda(p,n-k,k-1)$ $A=\sum\limits_{i=1}^{k}A_i$ 称为组内离差阵 $B=\sum\limits_{i=1}^{k}n_i(\overline{X}^{(i)}-\overline{X})(\overline{X}^{(i)}-\overline{X})^\top$ 称为组间离差阵 $T=\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i}\left(X_{(j)}^{(i)}-\overline{X}\right)\left(X_{(j)}^{(i)}-\overline{X}\right)^\top=A+B$ 为总离差阵

§9.3 协方差阵的检验

检验类型	检验情形	检验统计量
单个 p 元正态总体协方差阵的检验	$\Sigma = I_p$ 时 检验 $H_0 : \Sigma = I_p \ vs.\ H_1 : \Sigma \neq I_p$	构造似然比统计量 $\lambda_1 = \exp\left\{-\frac{1}{2}tr(A)\right\}\lvert A\rvert^{n/2}\left(\frac{e}{n}\right)^{np/2}$ n 较大且原假设成立时，$-2\ln\lambda_1$ 近似服从 $\chi^2\left(\frac{p(p+1)}{2}\right)$
	$\Sigma \neq I_p$ 检验 $H_0 : \Sigma = \Sigma_0 \ vs.\ H_1 : \Sigma \neq \Sigma_0$	构造 $Y_{(\alpha)} = DX_{(\alpha)} \sim N_p(D\mu, D\Sigma D^\top) = N_p(\mu^*, \Sigma^*)(\alpha = 1, 2, \ldots, n)$， 其中 $D_{p\times p}$ 为非退化矩阵且 $D\Sigma_o D^\top = I_p$，可从新样本出发作 $\Sigma^* = I_p$ 的检验 $\lambda_2 = \exp\left\{-\frac{1}{2}tr(A^*)\right\}\lvert A^*\rvert^{n/2}\left(\frac{e}{n}\right)^{np/2}, A = DAD^\top$ $-2\ln\lambda_2$ 近似服从 $\chi^2\left(\frac{p(p+1)}{2}\right)$
	检验 $H_0 : \Sigma = \sigma^2\Sigma_0$（$\sigma^2$ 未知）	构造似然比统计量 $\lambda_3 = \frac{\lvert\Sigma_0^{-1}A\rvert^{n/2}}{[tr(\Sigma_0^{-1}A/p)^{np/2}]}, W = (\lambda_3)^{2/n}$ 样本量 n 很大时，$-\left((n-1) - \frac{2p^2+p+2}{6p}\right)\ln W$ 近似服从 $\chi^2\left(\frac{p(p+1)}{2} - 1\right)$ 【注】$\Sigma_0 = I_p$ 时，该检验称为球度检验
多总体协方差阵的检验		构造 $\lambda_4 = \left\lvert\frac{A}{n}\right\rvert^{-n/2} \Big/ \prod\limits_{t=1}^{k}\left\lvert\frac{A_t}{n_t}\right\rvert^{-n_t/2}$，用 $n_i - 1$ 替代 n_i，用 $n-k$ 替代 n，修正 λ_4 为无偏的 λ_4^* 取检验统计量为 $M = -2\ln\lambda_4^* = (n-k)\ln\left\lvert\frac{A}{n-k}\right\rvert - \sum\limits_{t=1}^{k}(n_t-1)\ln\left\lvert\frac{A_t}{n_t-1}\right\rvert$ 样本量 n 很大，在原假设为真时，M 有近似分布：$(1-d)M \sim \chi^2(f), f = \frac{1}{2}p(p+1)(k-1)$ $d = \begin{cases} \frac{2p^2+3p-1}{6(p+1)(k-1)}\left[\sum\limits_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n-k}\right], & n_i\text{不全相等} \\ \frac{(2p^2+3p-1)(k+1)}{6(p+1)(n-k)}, & n_i\text{全相等} \end{cases}$
多个正态总体的均值向量 和协方差阵同时检验		构造 $\lambda_5^* = \frac{\prod\limits_{t=1}^{k}\lvert A_t\rvert^{(n_t-1)/2}}{\lvert T\rvert^{(n-k)/2}} \cdot \frac{(n-k)^{(n-k)p/2}}{\prod\limits_{t=1}^{k}(n_t-1)^{(n_t-1)p/2}}$ $A_t = \sum\limits_{j=1}^{n_t}\left(X_{(j)}^{(t)} - \overline{X}^{(t)}\right)\left(X_{(j)}^{(t)} - \overline{X}^{(t)}\right)^\top, A = \sum\limits_{t=1}^{k}A_t, T = A + \sum\limits_{t=1}^{k}n_t\left(\overline{X}^{(t)} - \overline{X}\right)\left(\overline{X}^{(t)} - \overline{X}\right)^\top$ 样本容量 n 很大，在原假设为真时成立 $-2(1-b)\ln\lambda_5^* \sim \chi^2(f), f = \frac{1}{2}p(p+3)(k-1)$ $b = \left(\sum\limits_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n-k}\right)\left(\frac{2p^2+3p-1}{6(p+3)(k-1)}\right) - \frac{p-k+2}{(n-k)(p+3)}$

§9.4　独立性检验和正态性检验

9.4.1　独立性检验

设总体 $X \sim N_p(\mu, \Sigma)$，将 X 剖分为 k 个子向量，而 μ 和 Σ 也相应剖分为：

$$X = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(k)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \vdots \\ \mu^{(k)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \dots & \Sigma_{1l} \\ \vdots & & \vdots \\ \Sigma_{k1} & \dots & \Sigma_{kk} \end{pmatrix} \tag{9.31}$$

其中 $X^{(t)} \sim N_{p_t}(\mu^{(t)}, \Sigma_{tt})(t = 1, \dots, k)$ 为 p_t 维子向量. 检验它们是否相互独立在于检验其中任意两个子向量的协方差阵是否等于 O（对一切 $i \neq j$）. 构造检验统计量：

$$\lambda = \prod_{i=1}^{k} \left| \frac{A_{ii}}{n} \right|^{-n/2} \Bigg/ \left| \frac{1}{n} A \right|^{-n/2} = \left(\frac{|A|}{\prod\limits_{i=1}^{k} |A_{ii|}} \right)^{n/2} = V^{n/2} \tag{9.32}$$

$$\ln \lambda = \frac{n}{2} \ln V \tag{9.33}$$

在原假设成立，当 $n \to \infty$ 时，有：

$$-b \ln V \sim \chi^2(f) \tag{9.34}$$

其中

$$b = n - \frac{3}{2} - \frac{p^3 - \sum\limits_{\alpha=1}^{k} p_\alpha^3}{3\left(p^2 - \sum\limits_{\alpha=1}^{k} p_\alpha^2\right)} \tag{9.35}$$

$$f = \frac{1}{2}\left[p(p+1) - \sum_{\alpha=1}^{k} p_\alpha(p_\alpha + 1) \right] \tag{9.36}$$

9.4.2　正态性检验

设 p 维随机向量 $X = (X_1, \dots, X_p)^\top$，对于多元数据的正态性检验问题，常可以转化为多个一元或二元数据的正态性检验.

方法 9.4.1 (**一维边缘分布的正态性检验**) 检验分量 $X_i \sim N(\mu_i, \sigma_i^2)(i = 1, \dots, p)$. 常用的检验方法有：

1. χ^2 检验法：适用于连续性或离散型随机变量分布的拟合优度检验方法，也成为皮尔逊 χ^2 检验法；

2. 科尔莫戈洛夫检验法：适用于连续性分布的拟合优度检验方法；

3. 偏峰检验法；

4. W（Wiks）检验和 D 检验；

5. Q-Q（Quantile-Quantile）图检验法；

6. P-P（Probability-Probability）图检验法；

7. 3σ 原则检验法；

8. A^2 和 W^2 统计量检验法.

前两种方法也适用于非正态分布的拟合优度检验，后六种方法仅适用于正态性检验.

方法 9.4.2 (**二元数据的正态性检验**) X 的任意两个分量的 n 次观测数据记为 $X_{(i)} = (X_{i1}, X_{i2})^\top (i = 1, \ldots, n)$，检验二元观测数据是否来自二元正态分布的方法有：

1. 等概椭圆检验法；

2. 二元数据的 χ^2 图检验法；

方法 9.4.3 (p **元数据的正态性检验**) 设 $X_{(\alpha)} = (X_{\alpha 1}, \ldots, X_{\alpha p})^\top (\alpha = 1, \ldots, n)$ 为来自 p 元总体 X 的随机样本，检验 $H_0 : X \sim N_p(\mu, \Sigma)$，$H_1 : X$不服从$N_p(\mu, \Sigma)$，可以使用如下方法：

1. χ^2 统计量的 $Q - Q$ 图检验法（或 $P - P$ 图检验法)；

2. 主成分检验法.

第七部分

多因变量多元线性回归

第十章　经典单变量多元线性回归

§10.1　模型与假设

假定因变量 Y 与 $x_1, x_2, \ldots, x_m$ 线性相关，收集到 n 组数据 $(y_t, x_{t1}, x_{t2}, \ldots, x_{tm})(t = 1, 2, \ldots, n)$，记：

$$C = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n & X \end{pmatrix}, \tag{10.1}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{10.2}$$

则 C 是已知数据矩阵（常称为设计矩阵），且设 $n > m$，$\mathrm{rank}(C) = m + 1$，Y 是可观测的随机向量，β 是未知参数向量，ε 是不可观测的随机向量，σ^2 是未知参数

经典单变量多元线性回归满足以下回归模型

$$\begin{cases} Y = C\beta + \varepsilon \\ E(\varepsilon) = \mathbf{0} \\ D(\varepsilon) = \sigma^2 I_n \end{cases} \tag{10.3}$$

或

$$\begin{cases} Y = C\beta + \varepsilon \\ \varepsilon \sim N_n(\mathbf{0}, \sigma^2 I_n) \end{cases} \tag{10.4}$$

§10.2　参数估计

我们需要对模型中的未知参数向量 β 和未知参数 σ^2 给出估计.

10.2.1　β 的最小二乘估计

在上述模型中，定义误差平方和为

$$Q=\sum_{t=1}^{n}\varepsilon_t^2=(Y-C\beta)^\top(Y-C\beta). \tag{10.5}$$

则在前述假设条件下，$\operatorname{rank}(C)=m+1\leq n$，取

$$\hat{\beta}=(C^\top C)^{-1}C^\top Y, \tag{10.6}$$

则

$$Q(\hat{\beta})=\min_{\text{一切 }\beta}Q(\beta). \tag{10.7}$$

即 $\hat{\beta}$ 使误差平方和达到最小，我们便用 $\hat{\beta}$ 对 β 进行估计，称其为 β 的最小二乘估计量，使误差平方和达到最小的这种估计方法称为最小二乘法.

注 10.2.1 $\hat{\beta}$ 是 β 的最小方差线性无偏估计量.

注 10.2.2 $\hat{\beta}\sim N_{m+1}(\beta,\sigma^2(C^\top C)^{-1})$.

注 10.2.3 在 $\varepsilon\sim N_n(0,\sigma^2 I_n)$ 的假定下，$\hat{\beta}$ 是一切无偏估计中方差最小的估计.

注 10.2.4 注意到 $\hat{\beta}$ 正好是

$$C^\top C\beta=C^\top Y \tag{10.8}$$

的解，常称以上方程组为正规方程.

设 $\hat{Y}=C\hat{\beta}=C(C^\top C)^{-1}C^\top Y\overset{\triangle}{=}HY$ 为 Y 的预测值，称

$$H_{n\times n}=C(C^\top C)^{-1}C^\top \tag{10.9}$$

为帽子矩阵，此时残差向量可表示为

$$\hat{\varepsilon}=Y-\hat{Y}=(I_n-H)Y=(Y-C\hat{\beta})^\top(Y-C\hat{\beta}), \tag{10.10}$$

残差平方和为

$$Q(\hat{\beta})=\hat{\varepsilon}^\top\hat{\varepsilon}=Y^\top(I_n-H)Y=Y^\top Y-Y^\top C\hat{\beta}. \tag{10.11}$$

注 10.2.5 关于帽子矩阵，在运算上有几条显然的结论

$$HC = C(C^\top C)^{-1}C^\top C = C, \tag{10.12}$$

$$\mathrm{tr}(H) = \mathrm{tr}(C(C^\top C)^{-1}C^\top) = \mathrm{tr}((C^\top C)^{-1}C^\top C) = \mathrm{tr}(I_{m+1}) = m+1, \tag{10.13}$$

$$(I_n - H)C = C - C(C^\top C)^{-1}C^\top C = O_{n\times(m+1)}, \tag{10.14}$$

$$\mathrm{tr}(I_n - H) = \mathrm{tr}(I_n) - \mathrm{tr}(H) = n - m - 1. \tag{10.15}$$

10.2.2 β 和 σ^2 的极大似然估计

最小二乘法没有给出 σ^2 的估计，利用极大似然比原理可以得到 β 的极大似然估计与其最小二乘估计相同，此时还可以给出 σ^2 的极大似然估计为

$$\hat{\sigma}^2 = \frac{1}{n}(Y - C\hat{\beta})^\top(Y - C\hat{\beta}) = \frac{1}{n}Y^\top(I_n - H)Y = \frac{1}{n}Q(\hat{\beta}). \tag{10.16}$$

这是有偏的估计量，通常可以取

$$s^2 = \frac{1}{n-m-1}Q(\hat{\beta}). \tag{10.17}$$

它是 σ^2 的无偏估计量.

§10.3 显著性检验

10.3.1 回归方程的显著性检验

我们需要进行两项显著性检验，首先要看因变量与全部自变量之间是否显著的存在着线性关系，即检验方程整体的显著性，若不存在显著的线性关系，所建回归方程便没有了意义，此时方程各系数应该全为 0，检验原假设为

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0. \tag{10.18}$$

对任意给定的观测数据矩阵

$$\begin{pmatrix} y_1 & x_{11} & x_{12} & \cdots & x_{1m} \\ y_2 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{10.19}$$

建立回归方程，用最小二乘法估计 β，则恒有

$$\sum_{t=1}^{n}(y_t - \overline{y})^2 = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2 + \sum_{t=1}^{n}(\hat{y}_t - \overline{y})^2 \tag{10.20}$$

其中

$$\overline{y}=\frac{1}{n}\sum_{t=1}^{n}y_t \tag{10.21}$$

$$\hat{y}_t=\hat{\beta}_0+\hat{\beta}_0x_{t1}+\cdots+\hat{\beta}_0x_{tm} \tag{10.22}$$

分别记总偏差平方和、残差（剩余）平方和、回归平方和为

$$l_{yy}=TSS=\sum_{t=1}^{n}(y_t-\overline{y})^2,\quad Q=ESS=\sum_{t=1}^{n}(y_t-\hat{y}_t)^2,\quad U=RSS=\sum_{t=1}^{n}(\hat{y}_t-\overline{y})^2 \tag{10.23}$$

则有平方和分解

$$l_{yy}(TSS)=Q(ESS)+U(RSS) \tag{10.24}$$

注 10.3.1 总偏差平方和反映 Y 的观测值的总波动大小；残差平方和反映自变量和因变量除了线性关系以外的一切其他因素（包括非线性关系及随机误差）引起波动；回归平方和反映了因变量与自变量之间确有线性关系并通过模型涉及到的自变量引起的 Y 的估计值的波动大小.

由平方和分解公式，若回归平方和比残差平方和大得多，因为此时总偏差主要是由自变量的变化引起的，说明自变量对因变量的线性影响是显著的，则更倾向于拒绝原假设，故可利用回归平方和与残差平方和的比值构建检验统计量，注意到

1. $\hat{\beta}\sim N_{m+1}(\beta,\sigma^2(C^\top C)^{-1})$；
2. $\dfrac{1}{\sigma^2}Q=\dfrac{1}{\sigma^2}ESS\sim\chi^2_{n-m-1}$；
3. $\hat{\beta}$ 与 Q 相互独立.

另外，原假设成立时，$\dfrac{U}{\sigma^2}=\dfrac{RSS}{\sigma^2}\sim\chi^2_m$.
从而构造统计量

$$F=\frac{RSS/m}{ESS/(n-m-1)}\overset{H_0}{\sim}F(m,n-m-1). \tag{10.25}$$

检验统计量越大越容易拒绝原假设，设统计量的实现值为 F_0，则检验的 p 值为

$$p=P(F\geq F_0) \tag{10.26}$$

p 小于显著性水平 α 时拒绝原假设.

10.3.2 回归系数的显著性检验

当方程整体的显著性通过时，仅代表 $\beta_1, \ldots, \beta_m$ 不全为 0，即全体自变量对因变量的线性影响不都是不显著的，所以还需要检验方程中的每一个自变量自身与因变量之间是否存在显著的相关性，即要逐个进行下面的 m 个检验

$$H_0^{(i)} : \beta_i = 0, (i = 1, 2, \ldots, m) \tag{10.27}$$

我们需要考虑每个自变量对因变量的纯粹的影响，设 $U_{(i)}$ 为去掉 x_i 后余下的 $m-1$ 个自变量对因变量的回归平方和（即用除了 x_i 以外的其余变量与因变量进行回归并计算回归平方和），则去掉该自变量后回归平方和的减少量 $P_i = U - U_{(i)}$（或残差平方和的增加量 $P_i = Q_{(i)} - Q$）可以用来衡量 x_i 对因变量的影响程度. 注意到原假设成立时，$P_i \sim \chi^2(1)$，从而可以构造统计量

$$F_i = \frac{P_i}{Q/(n-m-1)} \overset{H_0}{\sim} F(1, n-m-1) \tag{10.28}$$

该统计量的值越大，说明剔除 x_i 后回归平方和的减少量越小（或残差平方和的增加量越大），x_i 对因变量的作用越倾向于是显著的，此时更倾向拒绝原假设，设统计量的实现值为 F_0，则检验的 p 值为

$$p = P(F_i \geq F_0). \tag{10.29}$$

另外，记 $L = (C^\top C)^{-1} = (l^{ij})$ 我们注意到

$$\hat{\beta} \sim N_{m+1}(\beta, \sigma^2 L) \tag{10.30}$$

从而

$$\hat{\beta}_i \sim N(\beta_i, \sigma l^{ii}). \tag{10.31}$$

可构造统计量

$$t_i = \frac{\hat{\beta}_i/\sqrt{l^{ii}}}{\sqrt{Q/(n-m-1)}} \overset{H_0}{\sim} t(n-m-1) \tag{10.32}$$

也可以用该统计量进行检验，同样当统计量的值偏大时拒绝原假设，设统计量的实现值为 t_0，则检验的 p 值为

$$p = P(t_i \geq t_0). \tag{10.33}$$

注 10.3.2 上述两种检验实际上是等价的.

注 10.3.3 观测数据中心化处理后，记

$$\tilde{C} = \begin{pmatrix} 1 & x_{11} - \overline{x}_1 & \cdots & x_{1m} - \overline{x}_m \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} - \overline{x}_1 & \cdots & x_{nm} - \overline{x}_m \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n & \tilde{X} \end{pmatrix} \tag{10.34}$$

$$\tilde{Y}=\begin{pmatrix}y_1-\overline{y}\\y_2-\overline{y}\\\vdots\\y_n-\overline{y}\end{pmatrix},\quad\tilde{\beta}=\begin{pmatrix}\beta_0^*\\\beta_1\\\vdots\\\beta_m\end{pmatrix}:=\begin{pmatrix}\beta_0^*\\B\end{pmatrix},\beta_0^*=\beta_0+\sum_{i=1}^{m}\beta_i\overline{x}_i+\overline{y}. \tag{10.35}$$

原模型可重写为

$$\begin{cases}\tilde{Y}=\tilde{C}\tilde{\beta}+\varepsilon\\\varepsilon\sim N_n(\mathbf{0},\sigma^2I_n)\end{cases} \tag{10.36}$$

对应正规方程为

$$\tilde{C}^\top\tilde{C}\tilde{\beta}=\tilde{X}^\top\tilde{Y}. \tag{10.37}$$

再记

$$\tilde{L}=\tilde{X}^\top\tilde{X}=(\tilde{l}^{ij}) \tag{10.38}$$

从而

$$\tilde{C}^\top\tilde{C}=\begin{pmatrix}n&O_{1\times m}\\O_{m\times 1}&\tilde{L}\end{pmatrix} \tag{10.39}$$

又可算得

$$\tilde{C}^\top\tilde{Y}=\begin{pmatrix}0\\\tilde{X}^\top\tilde{Y}\end{pmatrix} \tag{10.40}$$

从而由正规方程可知 $\hat{\beta}_0^*=0$，正规方程等价于

$$\tilde{L}B=\tilde{X}^\top\tilde{Y} \tag{10.41}$$

即中心化后的模型等价于

$$\begin{cases}\tilde{Y}=\tilde{X}B+\varepsilon\\\varepsilon\sim N_n(\mathbf{0},\sigma^2I_n)\end{cases}. \tag{10.42}$$

B 的最小二乘估计为

$$\hat{B}=(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{Y}\sim N_m(B,\sigma^2(\tilde{X}^\top\tilde{X})^{-1}). \tag{10.43}$$

此时回归平方和满足

$$U=(\hat{Y}-\overline{y}\mathbf{1}_n)^{-1}(\hat{Y}-\overline{y}\mathbf{1}_n)=(\tilde{X}\hat{B})^\top(\tilde{X}\hat{B})=\hat{B}^\top\tilde{X}^\top\tilde{Y} \tag{10.44}$$

可以证明

$$P_i=\frac{\hat{\beta}_i^2}{\tilde{l}^{ii}}. \tag{10.45}$$

第八部分

统计计算

第九部分

附录与参考文献

附录 A　数学工具

§A.1　常用不等式

定义 A.1.1 (Young **不等式**) 设 $p,q>1$，$\dfrac{1}{p}+\dfrac{1}{q}=1$，则对任意正数 a,b 成立

$$a^{\frac{1}{p}}b^{\frac{1}{q}} \le \frac{a}{p}+\frac{b}{q}. \tag{A.1}$$

定义 A.1.2 (**（离散场合下的）**Hölder **不等式**) 设 $p>1$，$\dfrac{1}{p}+\dfrac{1}{q}=1$，则对任意正数 $a_i,b_i,i=1,2,\ldots,n$ 成立

$$\sum_{i=1}^n a_ib_i \le \left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}. \tag{A.2}$$

定义 A.1.3 (**（离散场合下的）**Minkowski **不等式**) 设 $p>1$,$\dfrac{1}{p}+\dfrac{1}{q}=1$,则对任意正数 $a_i,b_i,i=1,2,\ldots,n$ 成立

$$\left[\sum_{i=1}^n (a_i+b_i)^p\right]^{\frac{1}{p}} = \left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}. \tag{A.3}$$

定义 A.1.4 (Cauchy-Schwarz **不等式**) 对 $a_1,\ldots,a_n,b_1,\ldots,b_n$，成立

$$\left(\sum_{i=1}^n a_ib_i\right)^2 \le \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right). \tag{A.4}$$

对于闭区间 $[a,b]$ 上的连续函数 f 和 g，成立

$$\left(\int_a^b f(x)g(x)\,dx\right)^2 \le \left(\int_a^b f^2(x)\,dx\right)\left(\int_a^b g^2(x)\,dx\right). \tag{A.5}$$

在终极的分析中，一切知识都是历史；

All knowledge is, in the final analysis, history;

在抽象的意义下，一切科学都是数学；

All sciences are, in the abstract, mathematics;

在理性的世界里，所有的判断都是统计学。

All methods of acquiring knowledge are, essentially, through statistics.

——Calyampudi Radhakrishna Rao