



分析与概率

咕咕嘎嘎

目录

第 1 章 测度和积分	1
1.1 σ -代数	1
1.2 测度	1
1.3 外测度	2
1.4 可测函数	4
1.5 函数积分和收敛定理	4
1.6 乘积测度与 Fubini-Tonelli 定理	4
第 2 章 符号测度和微分	5
第 3 章 度量空间	6
3.1 度量空间	6
3.1.1 度量空间	6
3.1.2 赋范线性空间	7
3.1.3 内积空间和 Hilbert 正交系	8
3.2 度量空间中的拓扑性质	10
3.2.1 基础概念	10
3.2.2 稠密性和可分性	11
3.2.3 完备性	11
3.2.4 列紧性和紧度量空间	12
第 4 章 线性泛函	14
4.1 赋范线性空间上的线性算子	14
4.1.1 线性算子	14
4.1.2 线性算子的有界性、连续性	15
4.2 有界线性泛函	16
4.2.1 线性泛函和核	16
4.2.2 特定空间上的连续线性泛函的表示形式和空间的同构	16
4.2.3 有限维的例子	17
4.3 Hahn-Banach 延拓定理	18
4.4 二次对偶空间和自反	19
4.5 弱拓扑和弱 * 拓扑	19
第 5 章 有界算子的基本定理	21
5.1 Baire 纲定理	21
5.2 开映射定理, 逆算子定理, 闭图像定理, 共鸣定理	22
5.2.1 开映射定理、逆算子定理	22
5.2.2 闭图像定理	22
5.2.3 共鸣定理	23
第 6 章 概率公理系统	25
6.1 概率空间	25
6.2 随机变量、分布函数、随机变量的数字特征	27

6.2.1 随机变量	27
6.2.2 分布函数	28
6.2.3 随机向量	28
6.2.4 随机变量的数字特征	29
6.2.5 独立性	29
6.3 随机变量的收敛性	30
第 7 章 条件期望	32
7.1 $L^2(\Omega, \mathcal{F}, P)$ 空间	32
7.2 条件期望的形式	32
7.3 推广条件下的条件期望	33
7.4 条件期望的性质	34
第 8 章 随机过程	38
8.1 随机过程的基本概念	38
8.1.1 基本概念	38
8.1.2 基本类型	40
8.2 计数过程和泊松过程	42
8.2.1 泊松过程	42
8.2.2 泊松过程衍生分布	44
8.2.3 泊松过程的推广过程	46
8.3 马尔可夫过程	47
第 9 章 鞅	48
第 10 章 布朗运动	49
第 11 章 ITO 公式和随机微积分	50

第1章 测度和积分

内容提要

□ 测度与外测度

□ 可测函数和抽象积分

测度起源于测量图形面积（体积）的需求。我们很容易就能定义、计算简单图形的面积，比如矩形、三角形等。但是对于人工构造出来的极其复杂的图形，测量面积这个任务就显得很棘手。比如说 Contor 集这样涉及到无穷多步分割、丢弃后产生的图形，它的长度是多少？传统几何无法回答这些问题。

我们将在 σ -代数这一具有强大性质的代数结构上构造出测度，并用外测度延拓测度的定义范围。

1.1 σ -代数

集族是元素为集合的集合。

定义 1.1 (代数和 σ -代数)

设集族 \mathcal{A} 非空，如果 \mathcal{A} 对取补集和有限并集封闭，也就是说若有 $E \in \mathcal{A}$ ，则有 $E^c \in \mathcal{A}$ ；若有 $E_j \in \mathcal{A}, j = 1, 2, \dots, n$ ，则 $\bigcup_{j=1}^n E_j \in \mathcal{A}$ ，则说 \mathcal{A} 是一个代数。

如果把代数定义中的对有限并封闭改为对可数并封闭，那么就得到了 σ -代数。



最“小”的 σ -代数就是 $\{X, \emptyset\}$ ，其中 X 是任意集合。最“大”的 σ -代数就是 $\mathcal{P}(X)$ ，这就是 X 的幂集，它由集合 X 的所有子集构成（有些时候写成 2^X ，代表其中有 $2^{|X|}$ 个元素）。

给定一个集族 \mathcal{E} ，我们能找到很多个 σ -代数，使得这个 σ -代数包含 \mathcal{E} ，把这些 σ -代数取交集，由于 σ -代数的可数交还是 σ -代数，因此我们就得到了一个包含了集族 \mathcal{E} 的最小 σ -代数，称为集族 \mathcal{E} 生成的 σ -代数，记作 $\mathcal{M}(\mathcal{E})$ 。显然我们有下面结论：

命题 1.1 ((L1.1))

若集族 $\mathcal{E} \subset \mathcal{M}(\mathcal{F})$ ，则集族 \mathcal{E} 生成的 σ -代数 $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}(\mathcal{F})$ 。



证明 证明只需注意 $\mathcal{M}(\mathcal{F})$ 这个 σ -代数包含了 \mathcal{E} ，因此也就包含了 $\mathcal{M}(\mathcal{E})$ 。

(P1.2) 一个重要的例子是拓扑空间 X 中的开集族生成的 σ -代数，我们称其为 Borel σ -代数，记作 \mathcal{B}_X . \mathcal{B}_X 内的所有元素称为 Borel 集。

在拓扑学中，我们知道开集对可数并封闭，闭集对可数交封闭，其他情况则不然。于是我们拓展开集和闭集：我们称开集的可数交为 G_δ 集，闭集的可数并为 F_σ 集。后面我们将看到这两种集合和零测集构建了任意的可测集 (δ, σ 这两个表示来源于德语中的交集和并集 (Durchschnitt, Summe))。 \mathcal{B}_X 的范围很广：包括了开集、闭集、 G_δ 集、 F_σ 集。

在各种 Borel σ -代数中，我们最关心的是实直线上的 Borel σ -代数 $\mathcal{B}_{\mathbb{R}}$ 。实际上 $\mathcal{B}_{\mathbb{R}}$ 可以由开区间、闭区间、左开右闭区间、左闭右开区间、到正负无穷的开闭区间族生成。

1.2 测度

测度实际上就是在 σ -代数上以集合为自变量的非负函数（集函数）。

定义 1.2 (测度)

设集合 X 上有 σ -代数 \mathcal{M} , 我们称集函数 μ 为测度, 如果:

1. $\mu(\emptyset) = 0$;
2. 可数可加性: 设有一列互不相交的集合列 $\{E_n\}_{n \geq 1}$, 则

$$\mu\left(\bigcup_{n \geq 1} E_n\right) = \sum_{n \geq 1} \mu(E_n).$$

容易由测度的可数可加性得到有限可加性: 只要取 $E_m = \emptyset, m \geq n$, 再利用 $\mu(\emptyset) = 0$.

定义 1.3 (测度空间)

设集合 X 非空, σ -代数 $\mathcal{M} \subset \mathcal{P}(X)$, 则二元体 (X, \mathcal{M}) 就称为可测空间, \mathcal{M} 中的元素称为可测集。若可测空间上还定义了测度 μ , 则称三元体 (X, \mathcal{M}, μ) 为测度空间。



笔记 实际上概率空间 (Ω, \mathcal{F}, P) 就是一个测度空间。

若所有测度为 0 的集合 (零测集) 都包含于 \mathcal{M} , 那么说测度 μ 是完备的。我们有下面的完备化定理:

定理 1.1 (完备化)

设有测度空间 (X, \mathcal{M}, μ) , 设 $\mathcal{N} = \{N \in \mathcal{M} : \mu(N) = 0\}$, $\overline{\mathcal{M}} = \{E \cup F : E \in \mathcal{M}, F \in \mathcal{N}\}$, 则 $\overline{\mathcal{M}}$ 也是一个 σ -代数, 而且存在测度 $\bar{\mu}$ 在 $\overline{\mathcal{M}}$ 上的唯一延拓 $\bar{\mu}$, 使得 $\bar{\mu}$ 是 $\overline{\mathcal{M}}$ 上的完备测度。



测度有下面的优良性质 (T1.8):

性质 设有测度空间 (X, \mathcal{M}, μ) ,

1. 单调性: 若 $E, F \in \mathcal{M}, E \subset F$, 则 $\mu(E) < \mu(F)$.
2. 次可加性: 设有一列集合列 $\{E_n\}_{n \geq 1}$, 则 $\mu(\bigcup_{n \geq 1} E_n) \leq \sum_{n \geq 1} \mu(E_n)$. 这里没有要求集合列互不相交。
3. (从) 下连续性: 设有一列集合列 $\{E_n\}_{n \geq 1} \subset \mathcal{M}$, 且 $E_n \subset E_{n+1}, n \geq 1$, 则 $\mu(\bigcup_{n \geq 1} E_n) = \lim_{n \rightarrow \infty} \mu(E_n)$.
4. (从) 上连续性: 设有一列集合列 $\{E_n\}_{n \geq 1} \subset \mathcal{M}$, 且 $E_{n+1} \subset E_n, n \geq 1, \mu(E_1) < \infty$, 则 $\mu(\bigcup_{n \geq 1} E_n) = \lim_{n \rightarrow \infty} \mu(E_n)$. (注意这里的 $\mu(E_1) < \infty$ 不能去掉)

证明 证明只要利用测度的可列可加性, 并作出互不相交的集列

$$F_n = E_n - \bigcup_{k=1}^{n-1} E_k$$

即可。

测度还可以配备一些描述测度“大小”的性质:

1. 若 $\mu(X) < \infty$, 则称 μ 为有限的。
2. 若空间 X 可以被分割成无交并, 即 $X = \bigcup_{n \geq 1} E_n$, $\{E_n\}_{n \geq 1}$ 互不相交, 那么如果对于任意的 n , 都有 $\mu(E_n) < \infty$, 则称 μ 为 σ -有限的。

1.3 外测度

这一节我们从最基本的集合开始, 从代数的结构出发得到预测度, 再延拓预测度到更广泛的集类上, 最后再引出测度的构造。

定义 1.4 (预测度)

设集族 $\mathcal{A} \subset \mathcal{P}(X)$ 是一个代数, 我们称 $\mu_0 : \mathcal{A} \rightarrow [0, \infty]$ 为预测度, 如果

1. $\mu_0(\emptyset) = 0$;

2. 可数可加性：设有一列互不相交的集合列 $\{E_n\}_{n \geq 1}$, 则

$$\mu(\bigcup_{n \geq 1} E_n) = \sum_{n \geq 1} \mu(E_n).$$



笔记 预测度和测度的区别就在于预测度是定义在代数上而测度是定义在 σ -代数上的。和测度同样，预测度也有有限可加性。

然而在应用预测度来测量复杂图形（如 Contor 集）的面积时我们遇到了不可避免的困难，因为这样的图形的构造步骤就是无穷多步的，而预测度无法确认可数个区间构成的图形是否能被测量。因此我们干脆用更几何化的视角来看待图形的测量问题：从外部逼近这个图形，取逼近的下确界为这个图形的“面积”，这就是外测度的出发点。

定义 1.5 (预测度诱导的外测度)

我们称

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu_0(A_n) : A_j \in \mathcal{A}, E \subset \bigcup_{n=1}^{\infty} A_n \right\}$$

为集合 E 的外测度，其中 μ_0 是定义在代数 \mathcal{A} 上的预测度。



笔记 条件中的 $E \subset \bigcup_{n=1}^{\infty} A_n$ 就是“从外部逼近”这个想法的数学体现，我们并不在意 $\bigcup_{n=1}^{\infty} A_n$ 是否属于 \mathcal{A} 。

实际上，上面由预测度诱导的外测度的下确界定义并不是外测度的正式定义，但是我们可以验证上面的定义满足下面外测度的定义 (P1.10)。

定义 1.6 (外测度)

外测度就是 X 上满足以下条件的集函数 $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$

1. $\mu^*(\emptyset) = 0$;
2. (单调性) 若 $A \subset B$, 则 $\mu^*(A) \leq \mu^*(B)$;
3. (次可加性) $\mu^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.



现在我们用可测性的概念揭示外测度和测度之间的联系。

定义 1.7 (μ^* 可测性)

设外测度 μ^* 定义在空间 X 上, 取 X 的子集 A , 我们说 A 是 μ^* 可测的, 如果对于任意的 $E \subset X$ 都有

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E - A).$$



笔记 我们由外测度的次可加性可以马上得到 $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E - A)$, 因此如果要判断集合 A 是否 μ^* 可测, 只要验证反向不等式即可。而验证反向不等式时由于 $\mu^*(E) = \infty$ 时的情况是平凡的, 因此也只用考虑 $\mu^*(E) < \infty$ 时的情况。

外测度并不具有可列可加性, 但是由外测度我们可以衍生出测度和 σ -代数结构。

定理 1.2 (Carathéodory(T1.11))

设 μ^* 是 X 上的一个外测度, 则 μ^* 可测集形成的集类 \mathcal{M} 是一个 σ -代数, 并且 $\mu^*|_{\mathcal{M}}$ 是完备的测度。



笔记 Carathéodory 将预测度诱导的外测度扩张到测度, 并构造了 σ -代数结构, 是测度论中的一个基础且重要的定理。

下面的命题说明了外测度是预测度的延拓。

命题 1.2 (P1.13)

设 μ^* 是代数 \mathcal{A} 上的预测度 μ_0 诱导出的外测度，则

1. $\mu^*|_{\mathcal{A}} = \mu_0$;
2. 代数 \mathcal{A} 中的集合都是 μ^* 可测的。

最后 Carathéodory 扩张定理说明预测度、外测度、测度三者之间的关系。

定理 1.3 (Carathéodory 扩张定理)

设 $\mathcal{A} \subset \mathcal{P}(X)$ 是配有预测度 μ_0 的代数，代数 \mathcal{A} 生成的 σ -代数为 \mathcal{M} 。则 \mathcal{M} 上存在测度 μ ，使得 μ 是预测度 μ_0 在 \mathcal{M} 上的延拓，即 $\mu_0 = \mu|_{\mathcal{A}}$ ，且外测度 μ^* 是测度 μ 的延拓，即 $\mu = \mu^*|_{\mathcal{M}}$ 。另外，如果 μ_0 是 σ -有限的，则 μ 是 \mathcal{M} 上 μ 唯一的测度延拓。



1.4 可测函数

可测函数沟通了不同的 σ -代数，保留了 σ -代数的优良性质。可测性是积分论中的重要理论基础。我们先看逆映射：逆映射具有非常良好的性质，它对取补集、取交集、取并集的运算封闭，也就是对于 σ -代数 $\mathcal{M} \subset \mathcal{P}(Y)$ 中的任意集合 E, F ，映射 $f : \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ 都有

$$\begin{aligned} f^{-1}(E \cup F) &= f^{-1}(E) \cup f^{-1}(F) \in \mathcal{N}, \\ f^{-1}(E \cap F) &= f^{-1}(E) \cap f^{-1}(F) \in \mathcal{N}, \\ f^{-1}(E^c) &= (f^{-1}(E))^c \in \mathcal{N}. \end{aligned}$$

其中 $\mathcal{N} \subset \mathcal{P}(X)$ 。

于是我们可以定义函数的可测性：

定义 1.8 (可测函数)

设有可测空间 $(X, \mathcal{M}), (Y, \mathcal{N})$ ，函数 $f : X \rightarrow Y$ 满足对于任意的 $E \in \mathcal{N}$ ，都有 $f^{-1}(E) \in \mathcal{M}$ ，则称函数 f 是 $(\mathcal{M}, \mathcal{N})$ 可测的。

若函数 f 是 $(\mathcal{M}, \mathcal{B}_{\mathbb{C}}(\mathcal{B}_{\mathbb{R}}))$ 可测的，则简称函数 f 是 \mathcal{M} 可测的。



笔记 显然若 f 是 $(\mathcal{M}, \mathcal{N})$ 可测的， g 是 $(\mathcal{N}, \mathcal{O})$ 可测的，那么 $g \circ f$ 是 $(\mathcal{M}, \mathcal{O})$ 可测的。

实际上， f 是 $(\mathcal{M}, \mathcal{N})$ 可测的当且仅当 \mathcal{N} 由集族 \mathcal{E} 生成，且对于任意的 $E \in \mathcal{E}$ 都有 $f^{-1}(E) \in \mathcal{M}$ 。

例题 1.1 Lebesgue 可测函数是 $(\mathcal{L}, \mathcal{B}_{\mathbb{C}})$ 可测的，Borel 可测函数是 $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{C}})$ 可测的。注意：若 f, g 都是 Lebesgue 可测的，不一定能得到 $f \circ g$ 也是 Lebesgue 可测的。但是若 f Borel 可测， g Lebesgue (Borel) 可测，则 $f \circ g$ Lebesgue (Borel) 可测。

下面我们讨论由函数列生成的 σ -代数。在这之前，先引进坐标映射 (coordinate map) 的概念。

定义 1.9 (坐标映射)

1.5 函数积分和收敛定理

1.6 乘积测度与 Fubini-Tonelli 定理

第2章 符号测度和微分

内容提要

□ 符号测度

□ 微分

第3章 度量空间

内容提要

□ 度量空间及其衍生空间

□ 度量空间中的拓扑性质

3.1 度量空间

3.1.1 度量空间

度量空间是集合和度量函数 $d(\cdot, \cdot)$ 构成的有序对 (X, d) :

定义 3.1 (度量空间)

设集合 X 非空, 若有 $d(x, y) : X \times X \rightarrow \mathbb{R}$, 且满足下面的条件:

1. (非负性和正定性): $\forall x, y \in X, d(x, y) \geq 0$, 且 $d(x, y) = 0$ 当且仅当 $x = y$;
2. (对称性): $d(x, y) = d(y, x)$;
3. (三角不等式): $d(x, z) \geq d(x, y) + d(y, z)$

满足上面条件的有序对 (X, d) 构成度量空间, d 成为 X 上的度量.



在度量空间中, 我们得以用实数从度量的角度拓展极限的含义。

定义 3.2 (度量空间中的极限)

设 (X, d) 是度量空间, $\{x_n\}$ 是 X 中的点列, $x_0 \in X$, 若

$$\lim_{n \rightarrow \infty} d(x_n, x_0) = 0$$

则称点列 $\{x_n\}$ 按照度量 d 收敛于 x_0 , 记作 $x_n \rightarrow x_0$.

有了度量下的极限概念, 我们还可以平行定义 Cauchy 列: $\forall \epsilon > 0, \exists N \in \mathbb{N}_+, s.t. \forall m, n > N$, 都有

$$d(x_m, x_n) < \epsilon,$$

则称点列 $\{x_n\}$ 是 X 中的 Cauchy 列.



性质 度量的性质

1. 若点列 $\{x_n\}$ 收敛, 则极限唯一. (三角不等式)

2. (度量的连续性) 若 $x_n \rightarrow x_0, y_n \rightarrow y_0$, 则

$$\lim_{n \rightarrow \infty} d(x_n, y_n) = d(x_0, y_0).$$

(三角不等式及对称性)

度量空间的完备性和可分性将在度量空间中的拓扑性质中涉及。

例题 3.1 不同的度量可能诱导了相同的拓扑. 如在 n 维线性空间 \mathbb{R}^n 中定义度量:

$$d_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}, p > 1;$$
$$d_\infty(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|.$$

可知 $\lim_{n \rightarrow \infty} d_p(x_n, x_0) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} d_\infty(x_n, x_0) = 0$, 他们都等价于 $\{x_n\}$ 按坐标收敛, 即等价于

$$\lim_{n \rightarrow \infty} |x_n^{(i)} - x_0^{(i)}| = 0, 1 \leq i \leq n.$$

例题 3.2 当然不同的度量也有可能诱导不同的拓扑, 如在集合 $C[0, 1]$ 中, 定义度量

$$d_c(x, y) = \max_{0 \leq t \leq 1} |x(t) - y(t)|;$$

$$d_{L^1}(x, y) = \int_0^1 |x(t) - y(t)| dt;$$

考虑函数列

$$g_n(x) = n^3 \left(\frac{1}{n^2} - x \right) \chi_{[0, \frac{1}{n^2}]},$$

$g_n(x)$ 在 d_{L^1} 下收敛于 0, 但是在 d_c 下发散至 $+\infty$.

3.1.2 赋范线性空间

具有线性的线性空间上如果再定义“距离”范数, 就成为了赋范线性空间. 完备的赋范线性空间称为 Banach 空间.

定义 3.3 (赋范线性空间)

记 \mathbb{K} 为实或复数域, 对 \mathbb{K} 上的线性空间 X , 若有 $\|\cdot\| : X \mapsto \mathbb{R}$, 且满足以下条件:

1. $\forall x \in X, \|x\| \geq 0$;
2. $\forall \alpha \in \mathbb{K}, \|\alpha x\| = |\alpha| \cdot \|x\|$;
3. $\|x + y\| \leq \|x\| + \|y\|$;
4. $\|x\| = 0$ 当且仅当 $x = 0$;

则称 $\|\cdot\|$ 为 X 上的范数, $(X, \|\cdot\|)$ 成为赋范线性空间.



例题 3.3 在几乎处处相等等价类下, 证明 $L^p(E)$ 空间的完备性.

证明 容易验证范数 $\|\cdot\|_p$ 满足范数性质的前两条, 而三角不等式由 Minkovski 不等式给出, 于是验证了范数 $\|\cdot\|_p$ 是 $L^p(E)$ 的范数. 而 $\|f\|_p = 0$ 推出 $f = 0$ 只有在几乎处处意义下才成立, 因此等价类只能选择几乎处处相等, 才能保证 $(X, \|\cdot\|_p)$ 成为赋范线性空间. 下面证明由范数导出的度量 $d(x, y)$ 也是完备的.

先取 $L^p(E)$ 中的 Cauchy 列 $\{f_n\}$, 则对任意正数 ϵ , 存在正整数 N , 使得任意 $m, n \geq N$, 都有

$$\|f_m - f_n\|_p^p = \int_E |f_m - f_n|^p < \epsilon^p.$$

于是由 Chebyshev 不等式:

$$m(|f_m - f_n|^p \geq \epsilon) \leq \frac{1}{\epsilon} \int_E |f_m - f_n|^p \leq \epsilon^{p-1}.$$

即 $\{f_n\}$ 是按测度 Cauchy 列.

由于 Lebesgue 可测函数在测度度量下完备, 因此存在 $f \in \mathcal{M}(E)$, s.t. $f_n \rightarrow f, n \rightarrow \infty$. 再由 Fatou:

$$\|f_n - f\|_p^p = \int_E |f_n - f|^p \leq \liminf_{m \rightarrow \infty} \int_E |f_n - f_m|^p < \epsilon^p.$$

于是得出 $f_n - f \in L^p(E)$. 又由 $L^p(E)$ 的线性性知:

$$f = (f - f_n) + f_n \in L^p(E).$$



笔记 赋范线性空间的范数性质不仅和空间本身有关, 还和等价关系有关.

例题 3.4 证明 l^p 空间完备.

证明 由练习 1.1 可知 d_p 和 d_∞ 诱导了相同的拓扑, 因此取 l^p 空间中的 Cauchy 列 $\{x_n\}$, 任给正数 ϵ , 存在正整数 N , 使得任意 $m, n \geq N$, 都有

$$\|x_m - x_n\|_p < \epsilon.$$

于是有

$$\|x_m - x_n\|_\infty < \epsilon.$$

由实数的完备性, $\{x_n\}$ 按坐标收敛于 x_0 , 于是 $x_n \rightarrow x_0, n \rightarrow \infty$.

如上例, 由 Fatou 知:

$$\|x_n - x_0\|_p = \left(\sum_{i \geq 1} |x_n^{(i)} - x_0^{(i)}|^p \right)^{1/p} \leq \liminf_{m \rightarrow \infty} \left(\sum_{i \geq 1} |x_n^{(i)} - x_m^{(i)}|^p \right)^{1/p} \leq \epsilon.$$

于是知 $x_n - x_0 \in l^p$, 故

$$x_0 = (x_0 - x_n) + x_n \in l^p.$$

3.1.3 内积空间和 Hilbert 正交系

我们将线性空间中的内积运算推广到任意向量空间中, 具有内积运算的向量空间就是内积空间. 正交性是内积空间中重要的性质.

定义 3.4 (内积空间)

设 \mathbb{K} 为数域, 对于 \mathbb{K} 上的线性空间 H , 有 $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{K}$, 且满足条件:

1. (共轭对称性) $\langle x, y \rangle = \overline{\langle y, x \rangle}$
2. (第一变元线性性) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
3. (正定性) $\langle x, x \rangle \geq 0$, 且 $\langle x, x \rangle = 0$ 当且仅当 $x = 0$.

则称 $\langle \cdot, \cdot \rangle$ 是 H 中的内积, H 是内积空间.

内积导出的度量是完备的内积空间称为 Hilbert 空间.



命题 3.1

内积可以导出范数:

$$\sqrt{\langle x, x \rangle} = \|x\|,$$

范数可以导出度量:

$$d(x, y) = \|x - y\|.$$



命题 3.2 (平行四边形公式 (L1.2.5))

设 H 是内积空间, $\|\cdot\|$ 是内积导出的范数, 则对 H 中的任意 x, y , 都有:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$



笔记 只有满足平行四边形公式, 空间中的范数才能被内积诱导.(T1.2.3)

性质 [内积的连续性] 若 $x_n \rightarrow x, y_n \rightarrow y$, 则

$$\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle.$$

例题 3.5 $l^p, L^p[a, b]$ 空间中只有 $p = 2$ 时构成 Hilbert 空间.

证明 只要证明 $p \neq 2$ 时不满足平行四边形公式即可.

l^p :

当 $p \neq 2$, 取 $x = (1, 1, 0, \dots), y = (1, -1, 0, \dots)$, 则 $\|x + y\|_p = 2, \|x - y\|_p = 2, \|x\|_p = \|y\|_p = 2^{1/p}$,

若满足平行四边形公式, 则

$$2^{2/p} = 2,$$

也即 $p \neq 2$ 时无解.

$L^p[a, b]$:

不妨设 $a = 0, b = 1$. 取 $f = \chi_{[0, \frac{1}{n}]}, g = \chi_{[\frac{1}{n}, \frac{2}{n}]}$, 则有 $\|f\|_p = \|g\|_p = (\frac{2}{n})^{1/p}$, 而 $f + g = 2\chi_{[0, \frac{1}{n}]}, f - g = 2\chi_{[\frac{1}{n}, \frac{2}{n}]}$, $\|f + g\|_p = \|f - g\|_p = \frac{2}{n^{1/p}}$.

若平行四边形公式成立，则有

$$\left(\frac{2}{n}\right)^{2/p} = 2\left(\frac{1}{n}\right)^{2/p},$$

$p \neq 2$ 时无解.

定义 3.5 (正交和正交系)

1. 若 $\langle x, y \rangle = 0$, 则称 x, y 正交.
2. 设 \mathfrak{F} 是 Hilbert 空间的一族非零向量, 若 \mathfrak{F} 中任意两个向量都正交, 且所有向量范数为 1, 则称 \mathfrak{F} 是 H 的一组标准正交系.



定理 3.1 (标准正交基的充要条件 (T1.3.2))

设 $\mathfrak{F} = \{e_\lambda : \lambda \in \Lambda\}$ 是 Hilbert 空间 H 中的标准正交系, 下面的命题相互等价:

1. 对 H 中任意点 x , 都有 $x = \sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle e_\lambda$, 即任意 H 中点都能被 Fourier 展开.
2. $\overline{\text{span } \mathfrak{F}} = H$, 即 \mathfrak{F} 在 H 中稠密.
3. \mathfrak{F} 完全, 即若 $x \in H$ 有 $\langle x, e_\lambda \rangle = 0 (\forall \lambda \in \Lambda)$, 则 $x = 0$;
4. \mathfrak{F} 完备, 即任意 $x \in H$, $\|x\|^2 = \sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2$ 成立.



笔记 上面 (4) 中等式为 Parseval 等式, 上面条件给出了是 Bessel 不等式成立的条件.

Bessel 不等式 (T1.3.1):

$$\sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2 \leq \|x\|^2.$$

同时 H 中任意 x 的 Fourier 系数最多有可数个不为零. Bessel 不等式给出了 $x \in H$ 的必要条件: x 的 Fourier 系数平方和级数收敛:

$$\sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2 < \infty.$$

例题 3.6 证明 $\mathfrak{F} = \{1, \dots, \sqrt{2}\sin nx, \sqrt{2}\cos nx, \dots\}$ 是空间 $L^p[0, 2\pi]$ 的一组标准正交基, 进而得出 Parseval 等式

$$\|f\|^2 = \sum_{e \in \mathfrak{F}} |\langle f, e \rangle|^2,$$

从而得出 Fourier 级数的 Riemann-Lebesgue 引理:

$$\lim_{n \rightarrow \infty} \int_0^{2\pi} f(x) \sin nx dx = 0, \quad \lim_{n \rightarrow \infty} \int_0^{2\pi} f(x) \cos nx dx = 0,$$

证明 易知 \mathfrak{F} 是标准正交系, 下面验证 $\overline{\text{span } \mathfrak{F}} = L^2[0, 2\pi]$ 即可. 只要证明 $\overline{\text{span } \mathfrak{F}}$ 在 C_c^∞ 中稠密, 又由 C_c^∞ 在 $L^p[a, b]$ 中稠密的事实 (见后例), 再由稠密性的传递性即可得出 $\overline{\text{span } \mathfrak{F}}$ 在 $L^p[a, b]$ 中稠密.

任取 $f \in C_c^\infty$, 对 Fourier 系数 $a_n(f)$ 分部积分可得

$$a_n(f) = -\frac{a_n(f'')}{n^2}.$$

于是由 Bessel 不等式:

$$|a_n(f'')|^2 = |\langle f'', \sqrt{2}\cos nx \rangle|^2 \leq \|f''\|^2$$

则有

$$|a_n(f)| \leq \frac{\|f''\|}{n^2}.$$

同理可得 $|b_n(f)| \leq \frac{\|f''\|}{n^2}$. 于是部分和

$$(S_n f)(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) dt + \sum_{k=1}^n (a_k(f) \sqrt{2}\cos kx + b_k(f) \sqrt{2}\sin kx)$$

满足 $(S_n f)(x) \leq C + \sup_{x \in [0, 2\pi]} \sum_{n \geq 1} \frac{1}{n^2} < \infty$, 即 $(S_n f)(x)$ 在 $[0, 2\pi]$ 上一致收敛. 再由 Dini-Lipschitz 知道 $S_n f$ 点态收敛至 f . 也就是说 $\|S_n f - f\| \rightarrow 0$, 这就证明了 $\overline{\text{span } \mathfrak{F}}$ 在 C_c^∞ 中稠密. 由于 $\overline{\text{span } \mathfrak{F}} = L^2[0, 2\pi]$, 因此由上述

定理的充要条件得出, Parseval 等式成立, 因此该级数的余项趋于零, 即

$$\lim_{\lambda \rightarrow \infty} |\langle f, e_\lambda \rangle|^2 = \lim_{n \rightarrow \infty} |a_n(f)|^2 = \lim_{n \rightarrow \infty} |b_n(f)|^2 = 0,$$

这就证明了 Fourier 级数的 Riemann-Lebesgue 引理.

实际上, 由 Weierstrass 第二逼近定理, $\{z^n = e^{inx} : n \in \mathbb{Z}\}$ 张成的三角多项式全体 \mathcal{P} 在 $C(\mathbb{T})$ 中按照 L^2 范数稠密. 又由 $C(\mathbb{T})$ 按照 L^2 范数在 $L^2(\mathbb{T})$ 中稠密, 由稠密性的传递性, \mathcal{P} 在 $L^2(\mathbb{T})$ 中稠密. 这提供了上面的另一个证明.

定义 3.6 (Hilbert 空间的同构)

设 H, \overline{H} 都是数域 \mathbb{K} 上的 Hilbert 空间, 若存在线性双射 $T : H \rightarrow \overline{H}$, 对任意的 $x, y \in H$ 满足内积不变:

$$\langle x, y \rangle = \langle Tx, Ty \rangle$$

则称 H, \overline{H} 是保持内积不变同构的 Hilbert 空间.



笔记 映射内积不变的性质表明 T 还是一个等距.

定理 3.2 (Hilbert 空间同构的充要条件 [3.6-5])

Hilbert 空间由标准正交基的势唯一决定, H_1 和 H_2 保持内积线性空间同构的充要条件是 $|H_1| = |H_2|$.



笔记 对于每一个 Hilbert 空间, 恰好有一个抽象的实 Hilbert 空间和一个抽象的复 Hilbert 空间. 换句话说, 同一个域上的两个抽象的 Hilbert 空间只有在势上有所差别. 这就把欧几里得空间的情形做了推广.

定理 3.3 (投影定理 (T1.3.4))

设 M 是 Hilbert 空间 H 的一个闭线性子空间, 则对于任意的 $x \in H$, 存在唯一的 $x_0 \in M$, 使得 $d(x, M) = \|x - x_0\|$, 且 $x - x_0 \perp x_0$.



笔记 投影定理说明了在用子空间 M 的点逼近 H 中的点时, 最佳逼近点就是 H 向 M 的投影点, 此时投影点距离垂足的距离最小.

3.2 度量空间中的拓扑性质

3.2.1 基础概念

定义 3.7

1. 开集: 由内点构成的集合, 集合的内部是包含于集合的最小开集.
2. 闭集: 包含自身所有极限点的集合, 即闭集 X 满足:

$$\forall x_0 \in X, \exists \{x_n\} \in X, x_n \rightarrow x_0.$$

3. 导集: 集合的所有极限点构成的集合.
4. 闭包: $\overline{A} = A \cup A'$, 是包含 A 的最小闭集.
5. 连续映射 f : 若 $x_n \rightarrow x_0$, 则 $f(x_n) \rightarrow f(x_0)$. 若 $f : X \rightarrow Y$, 则 f 连续的充要条件是: Y 中任意开集 (闭集) O 的原像 $f^{-1}(O)$ 在 X 中是开集 (闭集).



笔记

1. 导集运算符不是幂等的, 但是闭包运算符幂等, 即 $\overline{\overline{A}} = \overline{A}$.
2. \overline{A}, A' 都是闭集.

3.2.2 稠密性和可分性

定义 3.8

1. 稠密性：设 X 的子集 A, B , 若 $A \subset B$ 且 $B \subset \overline{A}$, 则称 A 在 B 中稠密.
2. 可分性：设 X 为度量空间，若 X 有可数子集在 X 中稠密，则称 X 是可分的.



性质 [子空间可分性] 若 (X, d) 可分，则任意 X 的子集 A 有： (A, d) 可分.

笔记 稠密性具有传递性：若 A 在 B 中稠密， B 在 C 中稠密，则 A 在 C 中稠密（按照稠密的范数相同的情况下）

例题 3.7 $l^\infty, L^\infty[a, b]$ 都是不可分的.

例题 3.8 由 Weiersreass 第一、第二逼近定理，多项式全体 \mathcal{P}_1 在 $C[a, b]$ 中稠密，三角多项式全体 \mathcal{P}_2 在 Banach 空间 $C(\mathbb{T})$ 中稠密.

例题 3.9

1. 对直线上的闭区间 $[a, b]$ 及 $1 \leq p < \infty, L^p[a, b]$ 中的简单函数全体，阶梯函数全体， $C_c^\infty, C[a, b]$ 和多项式全体 \mathcal{P} 都是 $L^p[a, b]$ 的稠密子集. 并且成立关系 ($\|\cdot\|_p$ 范数下)：

$$L^p[a, b] = \overline{\text{SimF}} \subset \overline{\text{StaF}} \subset \overline{C_c^\infty} \subset \overline{C[a, b]} \subset \overline{\mathcal{P}}.$$

2. $L^p(\mathbb{R})(1 \leq p < \infty)$ 中的简单函数全体，阶梯函数全体， $C_c^\infty(\mathbb{R})$ 都是 $L^p(\mathbb{R})$ 的稠密子集.
3. $L^\infty[a, b]$ 只有简单函数全体是稠密子集.

3.2.3 完备性

定义 3.9

1. (Cauchy 列) 对度量空间 X 中的点列 $\{x_n\}$, 任给正数 ϵ , 存在正整数 N , 对任意 $m, n \geq N$, 都有 $d(x_m, x_n) < \epsilon$, 就称 $\{x_n\}$ 是 X 的一个 Cauchy 列.
2. 度量空间 X 的完备性是指, X 中任意的 Cauchy 列 $\{x_n\}$ 都收敛于 X 中的一点 x_0 .



性质

1. 若点列 $\{x_n\}$ 有子列 $\{x_{n_k}\}$ 收敛于 x_0 , 则点列 $\{x_n\}$ 也收敛于 x_0 .
2. 度量空间中的收敛点列是 Cauchy 列, 而若度量空间完备则 Cauchy 列收敛于空间中的点.
3. 设 A 为度量空间 (X, d) 的一个子集, 子空间 (A, d) 完备的充要条件是 A 是 X 的一个闭子集.(L1.2.1)
4. 不完备空间可以具有完备子空间 (如 \mathbb{Q} 的子空间 \mathbb{Z} 完备).

推论 3.1 (T1.4.2)

有限维赋范线性空间完备.



定理 3.4 (闭球套定理和度量空间的完备性等价 (T1.4.4, T1.4.5))

闭球套定理成立的度量空间完备，完备度量空间成立闭球套定理.

设 X 是完备的度量空间，设 $B_n = \{x \in X : d(x, x_n) \leq \epsilon_n\}$ 是一列单调下降的闭球：

$$B_{n+1} \subset B_n,$$

若球的半径 $\epsilon_n \rightarrow 0$, 则存在唯一的 $x \in X$, 使得 $x \in B_n, \forall n \in \mathbb{N}$.



推论 3.2

1. 等价范数导出相同的拓扑.
2. 等价范数下的同一个度量空间中的 Cauchy 列等价.

3. 对于一个有限维的赋范线性空间，该空间上的所有范数都是等价的.



证明

- 设范数 $\|\cdot\|_X, \|\cdot\|_Y$ 等价，即对任意的 $x \in X$, 存在正数 c, C 使得

$$c\|x\|_X \leq \|x\|_Y \leq C\|x\|_X.$$

则当 $\|x_n - x_0\|_X \rightarrow 0$, 就有 $\|x_n - x_0\|_Y \rightarrow 0$. 反之也有当 $\|x_n - x_0\|_Y \rightarrow 0$, 就有 $\|x_n - x_0\|_X \rightarrow 0$. 这就证明了等价范数诱导了相同的拓扑.

- 由上面的证明可知两个等价范数下的 Cauchy 列等价.

- 有限维的赋范线性空间完备，因此任取 Cauchy 列 $\{x_n\}$ 都收敛于 $x \in X$, 也即 Cauchy 列等价，因此所有范数等价.

例题 3.10 证明 $C[a, b]$ 完备.

证明 设 $\{f_n\}$ 是 $C[a, b]$ 中的 Cauchy 列，于是对任意正数 ϵ , 存在正整数 N , 使得任意 $m, n \geq N$ 时有

$$\max_{a \leq x \leq b} |f_n(x) - f_m(x)| < \epsilon.$$

也就是说 $\{f_n(x)\}$ 是 \mathbb{R} 中的 Cauchy 列. 由实数的完备性，知道 $\{f_n(x)\}$ 点态收敛于函数 $f(x)$.

下面证明 $f(x) \in C[a, b]$. 只要证明 $\{f_n(x)\}$ 一致收敛于 $f(x)$ 即可. 由前式，令 $m \rightarrow \infty$, 由于点态收敛性，只要 $n \geq N$, 就有

$$|f_n(x) - f(x)| < \epsilon, \forall x \in [a, b].$$

此即 $\{f_n(x)\}$ 一致收敛于 $f(x)$, 故有 $f(x) \in C[a, b]$.

笔记 证明度量空间完备性分为两步：第一步写出 Cauchy 列，找出 Cauchy 列的极限点（本例通过实数的完备性将 $C[a, b]$ 的 Cauchy 列转化为实数 Cauchy 列处理，并得出其极限）；第二部证明该极限点仍在度量空间内（本例由一致连续性得出结论）.

例题 3.11 证明 $(c_0, \|\cdot\|_\infty)$ 完备.

证明 c_0 是 l^∞ 的线性子空间，并且 $\|\cdot\|_\infty$ 也给出了 c_0 的一个范数，因此由本节性质 3，只要证明 c_0 的闭性，即证明 c_0 中的极限点还在 c_0 中即可.

设 c_0 中有 Cauchy 列 $\{x_n\}, x_n = (x_1^{(n)}, \dots, x_k^{(n)}, \dots)$, 易知存在 $x_0 \in l^\infty$, 使得 $\|x_n - x_0\|_\infty \rightarrow 0$. 只要证明 $x_0 \in c_0$ 即可.

对于 Cauchy 列 $\{x_n\} \in c_0, x_n \rightarrow 0, n \rightarrow \infty$. 对任意正数 ϵ , 存在正整数 j , 使得任意 $m \geq j$, 就有 $|x_m| < \epsilon$. 同时有 $\|x_m - x_0\| < \epsilon$ 于是对于 x_0 的任意坐标分量 $x_j^{(0)}$ 有：

$$|x_j^{(0)}| \leq |x_j^{(m)} - x_j^{(0)}| + |x_j^{(m)}| \leq \|x_m - x_0\|_\infty + \epsilon < 2\epsilon.$$

由于按坐标收敛在 d_∞ 下等价于按范数 $\|\cdot\|_\infty$ 收敛，于是知道 $x_0 \in c_0$.

3.2.4 列紧性和紧度量空间

定义 3.10 ((相对) 列紧集)

- 相对列紧集：设有度量空间 X, A 是 X 中的相对列紧集，当任意 A 中的点列都有收敛于 X 中的子列.
- 列紧集：闭的相对列紧集，即相对列紧集的任意点列都有子列收敛于 A .



性质 [(T1.6.2,T1.6.3,E1.6.3)]

- A 完全有界且 X 完备 $\Rightarrow A$ 相对列紧 (Hausdorff)
- A 相对列紧且 X 闭 $\Rightarrow A$ 列紧 $\Leftrightarrow A$ 紧 $\Leftrightarrow A$ 的开覆盖族存在有限开覆盖
- A 相对列紧 $\Rightarrow A$ 完全有界 $\Rightarrow A$ 有界

笔记 度量空间中的列紧集等价于紧集.

定理 3.5 (有限维和无穷维空间的本质区别 (T1.6.1))

1. 无穷维空间 X 的单位闭球非相对列紧.
2. 有限维的赋范线性空间的有界集完全有界.



笔记 有限维赋范线性空间完备，因此通过其有界集完全有界可以得出有界集相对列紧，又由空间的完备性得知空间的闭性，于是有界集列紧. 在度量空间中列紧性和紧性相同，于是知道 Banach 空间中的有界集等价于紧集. 因此得知有限维赋范线性空间的单位闭球相对列紧.

定理 3.6 (紧集上的连续映射性质 (T1.6.8,C1.6.3,C1.6.4,,T1.6.9))

1. 设 K 为紧集， f 为 K 上的连续映射，则 $f(K)$ 紧.
2. (紧集上的连续映射有界) 紧集上的连续映射像集有界，并且上下确界可达.
3. 紧集上的连续双射 f 若满足逆 f^{-1} 连续，则 f 为拓扑同胚.
4. 设 X 为紧度量空间，若 f 为 X 上的连续函数，则 f 一致连续.
5. (Arzela-Ascoli) $C(X)$ 中的子集 A 是相对列紧集的充要条件是 A 为等度连续的有界集.



第4章 线性泛函

内容提要

- | | |
|--------------------|-----------------------|
| □ 线性算子 | □ 对偶空间和几种空间下连续泛函的表现形式 |
| □ 有界性、连续性 | □ 二次对偶空间和自反性 |
| □ 连续泛函 | □ 弱拓扑和弱*拓扑 |
| □ Hahn-Banach 延拓定理 | |

4.1 赋范线性空间上的线性算子

4.1.1 线性算子

保持线性性（群同态）的映射就是线性算子。

定义 4.1 (线性算子)

设 \mathbb{K} 是数域, X, Y 是 \mathbb{K} 上的线性空间, \mathcal{D} 是 X 的子集, \mathcal{R} 是 Y 的子集。若有映射 $T : \mathcal{D} \rightarrow \mathcal{R}$, 满足线性性:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y), \forall \alpha, \beta \in \mathcal{D},$$

就称 T 是线性算子, 其中 \mathcal{D}, \mathcal{R} 分别是 T 的定义域和值域。值域 $\mathcal{R} = T(\mathcal{D})$ 。又称 $\ker T = \{x \in \mathcal{D} : T(x) = 0\}$ 为 T 的核(零空间)。

特别地, 若 $\mathcal{R} = \mathbb{K}$ 时, 称 T 为 X 上的线性泛函, 当 $\mathcal{R} = X$ 时, 称 T 为 X 上的线性算子。



容易看出 $T0 = 0$. 线性算子本质就是映射, 因此也继承了映射的几条性质:

性质 [2.6-9] 设 $T : X \rightarrow Y$, 定义域 X 是线性空间。则:

1. 值域 $T(X)$ 是线性空间;
2. 零空间 $\ker T$ 是线性空间;
3. 若 $\dim X = n < \infty$, 则 $\dim TX \leq n$.

逆算子的性质 (2.6-10):

1. 逆算子 T^{-1} 存在的充要条件是: $Tx = 0$ 蕴含 $x = 0$;
2. 若逆算子 T^{-1} 存在, 则 T^{-1} 也是线性算子;
3. 若逆算子 T^{-1} 存在, 则由于 $\dim X \leq \dim TX$, 且由对称性, 有 $\dim TX \leq \dim X$, 故有 $\dim X = \dim TX$.

例题 4.1 可以验证 Fourier 系数映射 $\tau : L^1(\mathbb{T}) \rightarrow c_0(\mathbb{Z})$, $\tau(f) = \{\hat{f}(n)\}$, 是线性算子, 其中 $\hat{f}(n) = \langle f, e^{inx} \rangle$. 再由 Riemann-Lebesgue 引理, $\lim_{n \rightarrow \infty} \langle f, e^{inx} \rangle = 0$ 可知, $\lim_{n \rightarrow \infty} \hat{f}(n) = 0$, 即 $\tau(f) \in c_0(\mathbb{Z})$.

例题 4.2 Fourier 变换 $\mathcal{F} : L^1(\mathbb{R}) \rightarrow c_0(\mathbb{R})$ 形为

$$\mathcal{F}(f)(t) = \int_{\mathbb{R}} f(x) e^{-i2\pi tx} dx, f \in L^1(\mathbb{R}).$$

当 $t_n \rightarrow t$, 有 $f(x)e^{-i2\pi t_n x} \rightarrow e^{-i2\pi tx}$ (点态收敛). 而 $|f(x)e^{-i2\pi tx}| = |f(x)| \in L^1(\mathbb{R})$, 故由 Lebesgue 控制收敛:

$$\int_{\mathbb{R}} f(x) e^{-i2\pi t_n x} dx \rightarrow \int_{\mathbb{R}} f(x) e^{-i2\pi tx} dx,$$

即 $\mathcal{F}(f)(t_n) \rightarrow \mathcal{F}(f)(t)$, 故 $\mathcal{F}(f)(t)$ 连续. 而由 Riemann-Lebesgue, 当 $|t| \rightarrow \infty$, $\mathcal{F}(f)(t) \rightarrow 0$. 故 $\mathcal{F}(f) \in c_0(\mathbb{R})$.

4.1.2 线性算子的有界性、连续性

定理 4.1 (连续性和有界性 (2.1.1))

设 $T : X \rightarrow Y$ 是线性算子，则下列命题等价：

1. T 是连续映射，即 $x_n \rightarrow x$ 蕴含 $Tx_n \rightarrow Tx$ 或其拓扑表述。
2. T 在 X 任意一点连续。
3. (有界性) 存在正数 M ，使得 $\|Tx\|_Y \leq M\|x\|_X$.



笔记 有界性一条说明，对于有界的 X ，其值域 $T(X)$ 仍是有界的。

并不是所有算子都是有界的，比如求导算子就是无界的。

例题 4.3 设 $X = C^{(1)}[a, b]$ ，范数定义为 $C[a, b]$ 的范数。即 $\|x\| = \max_{a \leq t \leq b} x(t)$ 。定义求导算子 $D : X \rightarrow C[a, b]$ ：

$$D(x)(t) = \frac{d}{dt}x(t), \forall x \in X.$$

则 D 无界。

证明 只要取 $x_n(t) = e^{-n(t-a)}$, $x_n(t) \in X$, 可以知道 $\|x\| = 1$, 但是 $D(x_n)(t) = -ne^{-n(t-a)}$, 有 $\|D(x_n)(t)\| = n \rightarrow \infty$, 这就说明了 D 是无界算子。

Think outside of the box, 实际上 X 上的有界算子全体也能形成一个线性空间，记为 X' ，称其为 X 的对偶空间。实际上 X' 也成为线性空间，而且可以在上面定义范数，使其成为赋范线性空间。特别地，当 T 是有界线性泛函时，记 X' 为 X^* ，称其为 X 的（代数）对偶空间。我们后面提到的对偶空间都是指有界线性泛函构成的空间。

定理 4.2 (对偶空间 (2.1.2))

设 X, Y 是赋范线性空间， X' 是 X 到 Y 的有界线性算子全体。若

1. 对 $T \in X'$ 定义范数

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}$$

2. 设 $A, B \in X'$, $\alpha \in \mathbb{K}$, 定义算子运算：

- (a). $(A + B)(x) = Ax + Bx$;
- (b). $(\alpha A)x = \alpha Ax$.

则 X' 成为赋范线性空间。

同时，若 Y 是 Banach 空间，则 X' 也是 Banach 空间。



笔记 T 的范数 $\|T\|$ 还能表示为

$$\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|<1} \|Tx\|.$$

由上确界性质易知 $\|Tx\| \leq \|T\| \cdot \|x\|$, $\forall x \in X$. 我们可以把 $\|T\|$ 看作在单位球 $\|x\| = 1$ 上，像 $\|Tx\|$ 的最大扩展系数。

同时， T 有界的充要条件是 $\|T\| < \infty$.

证明 先证明 $\|\cdot\|$ 满足范数定义。

正定性显然。而 $\|\alpha T\| = \sup_{\|x\|=1} \|\alpha Tx\|_Y = |\alpha| \sup_{\|x\|=1} \|Tx\|$.

三角不等式： $\|T_1 + T_2\| = \sup_{\|x\|=1} \|(T_1 + T_2)x\|_Y = \sup_{\|x\|=1} \|T_1x + T_2x\|_Y \leq \sup_{\|x\|=1} \|T_1x\|_Y + \sup_{\|x\|=1} \|T_2x\|_Y = \|T_1\| + \|T_2\|$.

再证明 X' 是线性空间设 $A, B \in X'$.

$\|(A + B)x\| \leq \|Ax\| + \|Bx\| < \infty$; $\|\alpha Ax\| = |\alpha| \cdot \|Ax\| < \infty$. 因此 $A + B \in X'$, $\alpha A \in X'$. 线性得证。

下面证明当 Y 是 Banach 空间时， X' 也是 Banach 空间。

取 X' 中的 Cauchy 列 $\{T_n\}$ ，则对于任意正数 ϵ ，存在正整数 N ，使得任意的 $m, n \geq N$ ，都有 $\|T_m - T_n\| < \epsilon$ 。因此 $\|T_m x - T_n x\| = \|(T_m - T_n)x\| \leq \|T_m - T_n\| \cdot \|x\| \leq \epsilon \|x\|$. 于是知道 $T_n x$ 是 Y 中的 Cauchy 列。由于 Y 完

备, 因此 $\{T_n x\}$ 存在极限, 记为 Tx . 下面证明 $T \in X'$.

由于 $\|(T-T_n)x\| = \|Tx-T_nx\| = \lim_{m \rightarrow \infty} \|T_m x - T_n x\| \leq \epsilon \|x\|$, 因此 $T-T_n \in X'$, 故 $T = T_n - (T_n - T) \in X'$, 即 X' 完备.

笔记

1. X^* 的完备性只和 Y 有关, 而和 X 无关.
2. 由实数完备性, 元素为线性泛函的对偶空间 X^* 是 Banach 空间.

除了线性运算, 我们还能再为算子定义乘法运算, 即设有赋范线性空间 $X, Y, Z, A \in X^*, B \in Y^*$, 规定

$$(BA)x = B(Ax),$$

则 $BA : X \rightarrow Z$, 且 $\|BA\| \leq \|B\| \cdot \|A\|$.

如果有 $T : X \rightarrow X$, 则由上面定义的乘法, 我们有 $\|T^n\| \leq \|T\|^n$.

4.2 有界线性泛函

4.2.1 线性泛函和核

线性泛函是最简单的线性算子, 具有很多特殊的性质. 线性泛函的很多性质都由它的核决定.

设 A 是 B 的子空间, 则 A 的余维数定义为: $\text{codim } A = \dim B/A$. 其中 B/A 是商空间.

定理 4.3 (线性泛函和核的余维数 (2.2.0))

设 f 是线性空间 V 上的线性泛函, 则 $\text{codim } \ker f = 1$.

反之, V' 是 V 的一个余维数为 1 的线性子空间, 则存在 V 上的线性泛函 f 使得 $\ker f = V'$.

笔记 定理揭示了余维数为 1 的子空间和线性泛函的核的一一对应关系.

定理 4.4 (具有闭核的线性泛函有界)

设线性算子 $T : X \rightarrow Y$. 若 T 有界, 则 $\ker T$ 闭, 但是反之则不然.

证明 线性算子有界等价于连续, 故 $\ker T = f^{-1}\{0\}$ 仍为闭集, 因为 $\{0\}$ 闭. 反之考虑求导算子, $\ker D\}$ 为常数集, 显然是闭的, 但是 D 无界.

然而, 对于线性泛函 f , f 有界的充要条件是 $\ker f$ 闭.

证明 必要性在上面已经证过, 下面证充分性.

反证法. 设 $\ker f$ 闭, 但是 f 无界, 即 $\|f\| = \infty$. 于是可以取点列 $\{x_n\}$, $\|x_n\| = 1$, $|f(x_n)| > n$. 此时设

$$y_n = \frac{x_n}{f(x_n)} - \frac{x_1}{f(x_1)},$$

可知 $y_n \in \ker f$. 然而由于 $|\frac{x_n}{f(x_n)}| \rightarrow 0$, 于是有 $y_n \rightarrow -\frac{x_1}{f(x_1)}$, 即

$$\lim_{n \rightarrow \infty} y_n = -\frac{x_1}{f(x_1)},$$

但是由于 $f(-\frac{x_1}{f(x_1)}) = -1 \neq 0$, 故 $y_n \rightarrow y \notin \ker f$, 这与 $\ker f$ 的闭性矛盾. 故 f 有界.

4.2.2 特定空间上的连续线性泛函的表示形式和空间的同构

实际上, 在特定的空间 (比如 Hilbert 空间, $L^p[a, b]$) 上定义的连续线性泛函具有固定的形式, 而且每个形式都与另一个空间中的某个元素形成一一对应. 于是我们着重研究什么样的连续线性泛函能和另一个空间中的元素形成一一对应关系.

定义 4.2 (同构)

设有赋范线性空间 X, Y 和映射 $T : X \rightarrow Y$.

1. 若 $\|Tx\|_Y = \|x\|_X, \forall x \in X$, 则称 T 是保范算子.
2. 若 T 既是保范的, 又是线性映射, 而且还是 X, Y 上的一一对应, 则称 T 是 X, Y 上的(保范)同构映射.
3. 若 X, Y 上存在映射 $T : X \rightarrow Y$ 是保范同构映射, 就称 X, Y 是同构的记为 $X \cong Y$.

下面来看几种特定赋范线性空间上的连续线性泛函的表现形式.

定理 4.5 (F.Riesz 连续线性泛函表示 (2.2.2-))

1. Hilbert 空间: 设 H 是一个 Hilbert 空间, F 是 H 上的连续线性泛函, 则存在唯一的 $y \in H$, 使得

$$F_y(x) = \langle x, y \rangle, \forall x \in H,$$

且有 $\|y\| = \|F_y\|$.

2. l^p 空间: 设 $1 \leq p < \infty, \frac{1}{p} + \frac{1}{q} = 1$. 若 f 是 l^p 上的线性连续泛函, 则存在唯一的 $\eta \in l^q$, 使得

$$f_\eta(x) = \sum_{i=1}^{\infty} x_i \eta_i, \forall x \in l^p.$$

且有 $\|\eta\|_q = \|f_\eta\|$. 这说明了 $(l^p)^* \cong l^q$.

3. $L^p[a, b]$: 设 $1 \leq p < \infty, \frac{1}{p} + \frac{1}{q} = 1$. 若 f 是 $L^p[a, b]$ 上的线性连续泛函, 则存在唯一的 $\beta \in L^q[a, b]$, 使得

$$f_\beta(x) = \int_a^b x(t) \beta(t) dt, \forall x(t) \in L^p[a, b],$$

且有 $\|\beta\|_q = \|f_\beta\|$. 这说明 $(L^p[a, b])^* \cong L^q[a, b]$.

4. $C[a, b]$: 若 f 是 $C[a, b]$ 上的线性连续泛函, 则存在唯一的 $g \in V_0[a, b]$, 使得

$$f_g(x) = \int_a^b x(t) dg(t),$$

且 $\|g\|_{V_0} = \|f_g\|$. 这说明 $C[a, b] \cong V_0[a, b]$.



笔记 我们发现, 设赋范线性空间 X, Y 和连续线性泛函 $f : X \rightarrow Y$, 这些泛函具有固定的表现形式, 而且还和另一个空间中的点形成了一一对应, 并且泛函的范数和这些点的范数相等, 这就给出了对偶空间到另一个空间的同构.

4.2.3 有限维的例子

定理 4.6 (有限维线性算子和矩阵)

有限维线性空间上选定基后, 线性算子和矩阵一一对应. 由矩阵的运算性质, 有限维赋范线性空间中的线性算子连续.

**定理 4.7 (对偶空间维数 (2.9-1))**

设 X 是 n 维向量空间, 其基为 $\{e_i\}_{i=1}^n$, 则 $\{f_i\}_{i=1}^n$ 是对偶空间 X^* 的一组基. 因此 $\dim X = \dim X^* = n$.

其中 $f_i(e_j) = \delta_{ij}$, $\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$



证明 先证明线性无关性: 设 $f(x) = \sum_{i=1}^n k_i f_i(x) = 0$, 代入 $x = e_j$, 则有 $f(e_j) = \sum_{i=1}^n k_i f_i(e_j) = k_j = 0$.

再证明任意 $f \in X^*$ 都能被 $\{f_i\}_{i=1}^n$ 线性表示. 设 $f(x) = \sum_{i=1}^n k_i f_i(x)$, 则 $f_j(x) = \sum_{i=1}^n k_i f_i(e_j) = k_j$, 即

$$f(x) = \sum_{i=1}^n f_i(x) f(e_i).$$

 **笔记** 注意这里 $f, f_i \in X^*$, 而 $f(e_j), f(x) \in \mathbb{K}$.

我们还有下列引理:

引理 4.1 (2.9-2)

设 X 是有限维向量空间. 若 $x_0 \in X$ 满足对任意的 $f \in X^*$ 都有 $f(x_0) = 0$, 则 $x_0 = 0$. 

证明 设 $x_0 = \sum_{i=1}^n k_i e_i$, 则 $f(x_0) = \sum_{i=1}^n k_i f(e_i) = \sum_{i=1}^n k_i \sum_{j=1}^n l_j f_j(e_j) = 0$ 对于任何 $f \in X^*$ 成立, 因此 $k_i = 0, x_0 = 0$.

4.3 Hahn-Banach 延拓定理

一般来说, 在延拓问题中考虑的数学对象(比如映射)本来定义在给定集合 X 的一个子集 Z 上, 我们希望把它从 Z 延拓到整个 X 上, 并且要求原对象的某些性质在延拓后能够继续保留. 在 Hahn-Banach 延拓定理中, 被延拓的对象是定义在向量空间 X 的子空间 Z 上的线性泛函 f , 还要求这个泛函具有一定的有界性质, 而这个有界性质是用次线性泛函来描述的. 线性空间 L 上的次线性函数 p 满足次可加性和正齐次性:

$$p(x+y) \leq p(x) + p(y), \forall x, y \in L;$$

$$p(tx) = tp(x), \forall x \in L, t \geq 0.$$

这个定理保证了赋范空间充分地配备了有界线性泛函, 并且使我们有可能得到足够的对偶空间理论. 这里只介绍原定理的推论.

定理 4.8 (Hahn-Banach(2.3.1))

设实赋范线性空间 L 有线性子空间 L_0 , 若 f_0 是 L_0 上的有界线性泛函, 则存在 L 上的线性泛函 f , 使得 f 是 f_0 的延拓, 满足 $f|_{L_0} = f_0, \|f\| = \|f_0\|$. 

下面来看看 Hahn-Banach 延拓定理的应用.

推论 4.1

对赋范线性空间 X 中的非零元素 x_0 , 存在 $f \in X^*$, 使得 $\|f\| = 1, f(x_0) = \|x_0\|$. 

证明 作 x_0 张开的一维线性空间 $\mathbb{K}x_0$ 上的泛函 f_0 :

$$f_0(ax_0) = |a| \cdot \|x_0\|,$$

则 $\|f_0\| = \sup_{\|x\|=1} |f_0(x)| = 1$. 故由 Hahn-Banach 延拓定理: 存在 $f \in X^*$, 满足

$$f|_{\mathbb{K}x_0}(x_0) = f_0(x_0) = \|x_0\|, \|f\| = \|f_0\|.$$

推论 4.2 (分离性 1)

对赋范线性空间 X 中的不同两个元素 x_1, x_2 , 存在 $f \in X^*$, 使得 $f(x_1) \neq f(x_2)$, 即 X^* 分离 X 中的点. 

证明 由于 $x_1 - x_2 \neq 0$, 由上一条推论知道 $f(x_1) - f(x_2) = f(x_1 - x_2) = \|x_1 - x_2\| \neq 0$.

推论 4.3 (对偶)

对赋范线性空间 X 中的元素 x , 存在 $f \in X^*$ 满足

$$\|x\| = \max_{\|f\|=1} |f(x)|.$$

证明 易知 $|f(x)| \leq \|f\| \cdot \|x\| = \|x\|$. 而由上述推论知这样的 f 是存在的, 故等号成立.

推论 4.4 (分离性 2)

设 M 是赋范线性空间 X 的闭线性子空间, 若有 $x \notin M$, 则存在 $f \in X^*$, 使得 $\|f\| = 1$, $f(M) = 0$, $f(x) = d(x, M)$. 这说明了 X^* 可以分离 X 中的闭线性子空间和点.



4.4 二次对偶空间和自反

X 的对偶空间 X^* 的对偶空间 X^{**} 称为 X 的二次对偶空间, 它与 X 具有很大的关系, 特别是当 X 自反时. 固定 z , 我们可以定义映射 $J : X \rightarrow X^{**}$, $x \mapsto g_x(f) = f(x)$ 称为典范映射 (自然嵌入). 可以证明 J 是线性泛函. 事实上, $J(\alpha x + \beta y) = g_{\alpha x + \beta y}(f) = f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha J(x) + \beta J(y)$.

同时, J 还是保范的, 即 $\|J(x)(f)\| = \|g_x(f)\| = \|x\|$. 因为 $\|g_x\| = \sup_{f \neq 0, f \in X^*} \frac{|g_x(f)|}{\|f\|} = \sup_{\|f\|=1} |f(x)| = \|x\|$ (推论 2.3).

笔记 典范映射 $J : X \rightarrow X^{**}$, $x \mapsto g_x(f) = f(x)$ 中, 注意 $x \in X$ 是固定的量, 而变量是 $f \in X^*$. 典范映射 $J(x) : X^* \rightarrow \mathbb{K}$, $J(x)(f) = f(x)$.(其实写成 $x(f)$ 可能会更好些)

定义 4.3 (自反)

若典范映射 $J : X \rightarrow X^{**}$ 是同构, 称 X 自反. 此时以等价类 “ \cong ” 代替 “ $=$ ”.



笔记 事实上存在同构 $J : X \rightarrow X^{**}$ 不是满射.

推论 4.5

若 X 自反, 则 X^* 自反. 事实上 $(X^*)^{**} = (X^{**})^* = X^*$.



例题 4.4 设 $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. 则有 $(L^p[a, b])^{**} = L^p[a, b]$, 存在 $J = I$ 为同构, 因此 $L^p[a, b]$ 自反. 同理有 l^p 自反.

不过也有非自反的例子, 比如 $L^1[a, b]$ 就不是自反的. 为了证明这一点, 我们要证明下面的定理:

定理 4.9 (2.5.5)

设 X 是赋范线性空间, 若 X^* 可分, 则 X 可分.



有了这个定理, 我们来证明 $L^1[a, b]$ 不是自反的. 反证法: 若 $L^1[a, b]$ 自反, 即 $(L^1[a, b])^{**} = L^1[a, b]$, 故 $(L^1[a, b])^{**}$ 可分, 于是 $(L^1[a, b])^* = L^\infty[a, b]$ 可分, 矛盾.

推论 4.6 (对偶)

若 X 自反, 对任意的 $f \in X^*$, 都存在 $x \in X$, 使得 $\|x\| = 1$, $f(x) = J(x)(f) = \|f\|$.



笔记 若 X 自反, 设 $f_1 \neq f_2$, 则存在 $x \in X$, $\|x\| = 1$, 使得 $f_1(x) - f_2(x) = (f_1 - f_2)(x) = J(x)(f_1 - f_2) = \|f_1 - f_2\| \neq 0$, 即 X^{**} 分离 X^* 的点.

推论 4.7

1. 自反的线性空间完备. 事实上, 设赋范线性空间 X 自反, 则 $X = X^{**} = (X^*)^*$ 完备.
2. Hilbert 空间自反.



4.5 弱拓扑和弱*拓扑

我们知道有限维空间和无限维空间的本质区别是, 在赋范线性空间 X 的范数诱导的拓扑下, 有限维空间中的闭单位球是紧的, 而无限维的闭单位球不是. 为了解决这个问题, 我们引入比范数诱导的拓扑更弱的条件, 也

就是弱拓扑的概念.

定义 4.4 (弱收敛和弱*收敛)

设有赋范线性空间 X 和其对偶空间 X^* .

1. 若点列 $\{x_n\} \in X$ 对于任何 $f \in X^*$ 都有 $f(x_n) \rightarrow f(x)$, 则称 $\{x_n\}$ 弱收敛于 x , $x \in X$. 记作 $x_n \xrightarrow{w} x$.
2. 若点列 $\{f_n\} \in X^*$ 对于任何 $x \in X$ 都有 $f_n(x) \rightarrow f(x)$, 则称 $\{f_n\}$ 弱*收敛于 $f \in X^*$. 记作 $f_n \xrightarrow{w^*} f$.



笔记 由于 X^* 分离 X 中的点, 因此弱收敛极限唯一; 又由于 X^{**} 分离 X^* 中的点, 因此弱*收敛极限唯一.

命题 4.1 (依范数收敛(强收敛)和弱收敛的关系)

1. 强收敛蕴含弱收敛, 反之不然. 事实上若 $\{x_n\}$ 强收敛于 $x \in X$, 则对任意 $f \in X^*$, $|f(x_n) - f(x)| \leq \|f\| \cdot \|x_n - x\| \rightarrow 0$, 反之考虑 $X = L^2[0, 2\pi]$, 由 Riemann-Lebesgue: $f(e^{inx}) - f(0) = \int_0^{2\pi} f(x)e^{inx} dx \rightarrow 0$, $n \rightarrow \infty$. 即 $e^{inx} \xrightarrow{w} 0$. 然而 $\|e^{inx}\|_2 - \|0\|_2 = \sqrt{2\pi} \neq 0$, 即 e^{inx} 不强收敛于 0.
2. 强*收敛蕴含弱*收敛, 反之不然. 若 f_n 依 X^* 范数收敛于 $f \in X^*$, 则 $|f_n(x) - f(x)| = |(f_n - f)x| \leq \|f_n - f\| \cdot \|x\| \rightarrow 0$, $n \rightarrow \infty$. 反之考虑 $\{e_n\} \in l^\infty$, $x \in l^1$, 则 $|e_n \cdot x - 0| = |x_n| \rightarrow 0$, 即 $e_n \xrightarrow{w^*} 0$. 但 $\|e_n\|_{l^\infty} - 0 = 1 \neq 0$, 因此 $\{e_n\}$ 不强收敛于 0.
3. 赋范线性空间 X 中点列 $x_n \xrightarrow{w} x$ 等价于二次对偶空间中 X^{**} 点列 $Jx_n \xrightarrow{w^*} Jx$. 实际上二次对偶空间 X^{**} 的点列 $Jx_n \xrightarrow{w^*} Jx$ 等价于任意 $f \in X^*$, $Jx_n(f) \rightarrow Jx(f)$, 而这就是 $f(x_n) \rightarrow f(x)$. 例如, $x^{(n)} \xrightarrow[c_0]{w} x$ 和 $x^{(n)} \xrightarrow{l^\infty} x$ 都是指对于任意的 $\eta \in l^1$, $x^{(n)} \cdot \eta \rightarrow x \cdot \eta$.
4. 弱收敛蕴含弱*收敛.
5. 若 X 自反, 则 X^* 的弱收敛和弱*收敛等价.



笔记 有限维赋范线性空间中, 弱收敛等价于强收敛.

这里省略弱拓扑、弱*拓扑、弱紧、弱*紧的定义, 见定义 2.5.3.

下面介绍弱拓扑中的重要定理.

定理 4.10 (Mazur(2.5.7))

设 M 是赋范线性空间 X 中的凸子集, 则 M 的范数拓扑闭包等于其弱拓扑闭包.



定理 4.11 (Banach-Alaoglu(2.5.8))

赋范线性空间 X 的对偶空间 X^* 中的闭单位球是弱*紧的.



第5章 有界算子的基本定理

内容提要

□ Baire 纲定理

□ 几个有界算子的基本定理

5.1 Baire 纲定理

相对于稠密集的概念，我们引入疏朗集的概念来衡量一个集合中子集元素数量的多少。

定义 5.1 (疏朗集)

设 A 是度量空间 X 的一个子集，若 \overline{A} 不含有 X 的非空开集，即 \overline{A} 没有内点，则称 A 是 X 中的疏朗集。
 A 是疏朗集当且仅当 \overline{A} 是疏朗集。这就等价于 $\overline{A}^c = X/\overline{A}$ 在 X 中稠密。



笔记

1. 稠密性的定义：

A 在 X 中稠密 $\Leftrightarrow X \subset \overline{A} \Leftrightarrow$ 任意 X 中一点 x_0 , 存在 x_0 的一个邻域 V_0 使得 $V_0 \cap A \neq \emptyset$.

2. 由定义看出，一个集合是不是疏朗集和这个集合上采取的拓扑有关。

有了疏朗集的定义，我们就能描述子集中元素的多少情况了。

定义 5.2 (纲)

度量空间 X 中，如果子集 A 可以写成可数个疏朗集的并，则称 A 是第一纲的，否则称 A 为第二纲的。



笔记

设 A_i 是第一纲集族，则 $\bigcup_{i=0}^{\infty} A_i$ 也是第一纲的。

例题 5.1 在 $(\mathbb{R}, \|\cdot\|)$ 中，单点集 $\{q\}$ 是疏朗集，因此整数集 \mathbb{Z} 是第一纲的；然而在 $(\mathbb{Z}, \|\cdot\|)$ 这个平凡的度量空间中，由于每一个点 $\{p\}$ 都是开集，因此不是疏朗集。故此在这个拓扑下 \mathbb{Z} 是第二纲的。

定理 5.1 (强 Baire 纲定理 (3.1.2))

设度量空寂 (X, d) 完备，若 $\{U_n\}$ 是 X 中一族稠密开集，则 $\bigcap_n U_n$ 是第二纲的稠密集。



定理 5.2 (Baire(3.1.1))

完备度量空间是第二纲的。



证明 反证法：若完备度量空间 X 是第一纲的，则存在疏朗集列 A_i ，使得 $X = \bigcup_i A_i$ 。因此 $X/\bigcup_i A_i = \bigcap_i A_i^c$ ，其中 A_i^c 是稠密开集，因此 $X/\bigcup_i A_i \neq \emptyset$ ，这和 X 是第一纲的矛盾。

例题 5.2 无理数 \mathbb{R}/\mathbb{Q} 是第二纲的。若不然，则它是第一纲的，又由于 \mathbb{Q} 是第一纲的，因此 $\mathbb{R}/\mathbb{Q} \cup \mathbb{Q} = \mathbb{R}$ 是第一纲的，这与 \mathbb{R} 的完备性矛盾。

例题 5.3 见例 3.1.4 和 3.1.5，存在 Lebesgue 测度为 0 的第二纲集，也存在 Lebesgue 测度大于 0 的第一纲集。

5.2 开映射定理, 逆算子定理, 闭图像定理, 共鸣定理

5.2.1 开映射定理、逆算子定理

定义 5.3

设 X, Y 是两个赋范线性空间, $T \in \mathfrak{B}(X, Y)$.

1. 若 T 是一对一的, 即 $x_1 \neq x_2$ 蕴含 $Tx_1 \neq Tx_2$, 则称 T 是单射.
2. 设 T 的值域 $\mathfrak{R}(T) = Y$, 则称 T 是满射, 或者 T 是到上的.
3. 若 T 既是单射又是满射, 且 $T^{-1} \in \mathfrak{B}(Y, X)$, 则称 T 是可逆的.



笔记

1. T 是可逆的, 当且仅当存在 $J \in \mathfrak{B}(Y, X)$ 使得 $TJ = I_Y, JT = I_X$.

2. 有限维空间中, 左逆或右逆存在即可推出逆算子存在, 但是无限维空间就不行.

设 X, Y 是两个赋范线性空间, $T \in \mathfrak{B}(X, Y)$, 且 T 是一一对应, 则 T^{-1} 存在且也是线性算子, 但是不一定有界(如求导算子). 然而, 逆算子定理给出了 T^{-1} 有界的条件.

连续算子的像若是开集, 则原像也是开集. 与其相对应, 把开集映为开集的映射是开映射. 注意开映射和连续算子是不同的概念. 开映射的例子有如商映射, 全纯函数等.

定理 5.3 (开映射定理 (3.2.1))

设 X, Y 是两个 Banach 空间, $T \in \mathfrak{B}(X, Y)$. 若 $TX = Y$, 则 T 是开映射.



利用开映射定理, 我们可以得到逆算子定理.

定理 5.4 (逆算子定理 (3.2.2))

设 X, Y 是两个 Banach 空间, $T \in \mathfrak{B}(X, Y)$, 且 T 是一一对应, 则逆算子 T^{-1} 有界.



证明 由开映射定理, T 是开映射. 任取 X 中开集 U , $TU = (T^{-1})^{-1}U$ 是开集, 因此 T^{-1} 连续.

下面是逆算子定理在范数等价方面的一个应用.

定理 5.5 (范数等价定理)

设线性空间 X 对其上定义的范数 $\|\cdot\|_1, \|\cdot\|_2$, 都成为 Banach 空间, 且存在正数 c 使得

$$\|x\|_1 \leq c\|x\|_2, \forall x \in X,$$

则 $\|\cdot\|_1, \|\cdot\|_2$ 等价.



证明 把恒等算子 I 看作 $(X, \|\cdot\|_1) \rightarrow (X, \|\cdot\|_2)$ 的线性算子, 则

$$\|Ix\|_1 = \|x\|_1 \leq c\|x\|_2,$$

即 I 有界, 又 I 是一一对应, 因此由逆算子定理, $I^{-1} = I$ 有界, 因此存在正数 c' 满足

$$\|x\|_2 = \|I^{-1}x\|_2 \leq c'\|x\|_1.$$

即 $\|\cdot\|_1$ 弱于 $\|\cdot\|_2$, 因此两个范数等价.

5.2.2 闭图像定理

先定义映射的图像.

定义 5.4 (映射的图像)

设有集合 X, Y , 映射 $T : \mathfrak{D}(T) \rightarrow Y$. 做 X, Y 的乘积集 $X \times Y = \{(x, y) : x \in X, y \in Y\}$, 称 $X \times Y$ 的子集

$$\mathfrak{G}(T) = \{(x, Tx) : x \in \mathfrak{D}(T)\}$$

为映射 T 的图像.



引进 X, Y 中的度量 d_X, d_Y , 使其称为度量空间, 则 $(X \times Y, d)$ 也是度量空间, 其中 $d((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}$.

定义 5.5 (闭算子)

设 $(X, d_X), (Y, d_Y)$ 是两个度量空间, $T : \mathfrak{D}(T) \rightarrow Y$, 若 T 的图像 $\mathfrak{G}(T)$ 是 $(X \times Y, d)$ 中的闭集, 则称 T 是闭算子.



闭算子的判定充要条件如下:

引理 5.1

设有度量空间 X, Y 和映射 $T : \mathfrak{D}(T) \rightarrow Y$, 则 T 是闭算子的充要条件是, 对任何 $\mathfrak{D}(T)$ 中点列 $\{x_n\}$, 当 $x_n \rightarrow x_0, y_n = Tx_n \rightarrow y_0$ 时, 就有 $x_0 \in \mathfrak{D}(T)$, 且 $y_0 = Tx_0$.

**引理 5.2**

定义域是闭集的连续算子是闭算子. 但是闭算子不一定是连续算子.

**定理 5.6 (闭图像定理 (3.2.5))**

设 X, Y 是两个 Banach 空间, $T : \mathfrak{D}(T) \rightarrow Y$ 是闭线性算子, 若 $\mathfrak{D}(T)$ 是 X 中的闭线性子空间, 则 T 连续.



例题 5.4 设 $f \in \mathcal{M}[0, 1]$, 若对于任意的 $g \in L^2[0, 1]$ 都有 $fg \in L^1[0, 1]$, 则 $f \in L^2[0, 1]$.

证明 定义乘法算子 $T(g) = fg$, 则 T 是 $L^2[0, 1] \rightarrow L^1[0, 1]$ 的闭算子. 由闭图像定理, 由于 $L^2[0, 1]$ 是闭线性子空间, 因此 T 连续, 即 T 有界. 取 $g_n = f\chi_{|f| < n} \in L^2[0, 1]$, 则 $\|g_n\|_2 = \frac{\|fg_n\|_1}{\|g_n\|_2} \leq \|T\|$, 再令 $n \rightarrow \infty$, 就得到 $\|f\|_2 \leq \|T\|$, $f \in L^2[0, 1]$.

下面证明 T 是闭算子. 对于 $g_n \xrightarrow{L^2} g, fg_n \xrightarrow{L^1} h$, 由 Chebyshev 知道函数列 $\{g_n\}, \{fg_n\}$ 分别依测度收敛于 g, h . 再由 Riesz, 存在子列 $\{g_{n_k}\}, \{fg_{n_k}\}$ 几乎处处收敛于 g, h . 于是知道当 $g_n \rightarrow g$, 就有 $Tg_n = fg_n \rightarrow h = fg = Tg, g \in L^2[0, 1]$. 即 T 是闭算子.

5.2.3 共鸣定理

定理 5.7 (共鸣定理 (一致有界原理) (3.2.6))

设 X 是 Banach 空间, Y 是赋范线性空间, $\{T_\lambda : \lambda \in \Lambda\} \subset \mathfrak{B}(X, Y)$, 若对于任意的 $x \in X$, 都有

$$\sup_{\lambda \in \Lambda} \|T_\lambda x\| < \infty,$$

则数集 $\{\|T_\lambda\| : \lambda \in \Lambda\}$ 有界.

用 Baire 纲定理, 可以给出共鸣定理的另一种形式:

设赋范线性空间 $X, Y, \{T_\lambda : \lambda \in \Lambda\} \subset \mathfrak{B}(X, Y)$. 若

$$\{x \in X : \sup_{\lambda \in \Lambda} \|T_\lambda x\| < \infty\}$$

是 X 的第二纲集, 则数集 $\{\|T_\lambda\| : \lambda \in \Lambda\}$ 有界.



例题 5.5 设 $p \geq 1$, $\alpha(t)$ 是实直线 \mathbb{R} 中闭区间 $[a, b]$ 上的可测函数. 若对任意 $x \in L^p[a, b]$, 积分

$$\int_a^b \alpha(t)x(t)dt$$

存在, 则 $\alpha(t) \in L^q[a, b]$, $\frac{1}{p} + \frac{1}{q} = 1$.

证明 取 $x(t) \equiv 1$, 则 $\alpha(t) \in L^1[a, b]$. 因此 α 几乎处处有限. 作函数列 $\alpha_n = \alpha \chi_{|\alpha| < n}$, 则 α_n 几乎处处收敛于 α . 作 $L^p[a, b]$ 上的连续线性泛函

$$F_n(x) = \int_a^b x(t)\alpha_n(t)dt, \forall x \in L^p[a, b],$$

由于 $|x(t)\alpha_n(t)| \leq |x(t)\alpha(t)| \in L^1[a, b]$, 由 Lebesgue 控制收敛定理:

$$\lim_{n \rightarrow \infty} F_n(x) = \int_a^b \alpha(t)x(t)dt, \forall x \in L^p[a, b].$$

即 $\sup_n |F_n(x)| < \infty$. 由共鸣定理: 存在数 M 使得 $\|F_n\| = \|\alpha_n\|_q < M$, 令 $n \rightarrow \infty$, 知道 $\|\alpha\|_q < M$. 即 $\alpha \in L^q[a, b]$.

第 6 章 概率公理系统

内容提要

- 概率空间 (Ω, \mathcal{F}, P)
- 随机变量的收敛性
- 随机变量与期望

我们无法预知随机现象的结果是什么，但是我们希望或多或少掌握一些随机现象的规律。为了用数学语言描述随机现象，我们引进了事件和事件的概率的概念。然而，事件和概率并没有严格的定义，而且朴素的概率观念在发展过程中遇到了很多阻碍，比如贝特朗悖论等。即使 19 世纪拉普拉斯建立了古典概率的理论基础，但其限制条件太多，特别是无法处理“无限”的情形，概率论的发展受到了阻碍。

不过，在 Lebesgue 测度和积分的研究日渐成熟的情况下，类比集合和测度，事件和概率建立了其严格的数学概念。这样的背景下，Kolmogorov 提出了概率的公理化系统，才为概率论作为数学的一个分支奠定了坚实基础。

6.1 概率空间

假定随机试验可以在相同条件下重复地进行，每次试验的结果可能不止一个，并且能事先确定试验的所有可能结果，但每次试验的结果事先又不可预测。在概率中我们对这些问题感兴趣：

1. 事件和事件对应的概率如何定义？
2. 事件能否可以被最基础的“原子”通过运算得到？

定义 6.1 (样本点和样本空间)

把随机试验每一个可能的结果称为一个样本点，通常用 ω 表示；所有可能的结果组成的集合称为样本空间，通常用 Ω 表示。

 **笔记** 我们用样本点可以组合出各种各样的“事件”，但是并不是所有的所有组合都能被称为事件。

定义 6.2 (事件)

我们把事件定义为样本点的某个集合。称某事件发生当且仅当它所包含的某个样本点出现。事件是样本空间的一个子集。

 **笔记** 虽然试验的样本点在试验前就很明确，但是只有试验之后，才能确定某个给定的事件是否发生。

和其他建立在严格符号系统和推理系统的数学相比，概率论具有明显的现实色彩。什么叫做“样本点出现”？这句话就让概率论和现实世界有了联系。

以上就给出了“事件”的定义。然而，单单有这样的定义还不够。我们希望事件作为集合的类比，也具有集合运算的性质。本质上事件之间的关系和运算就是集合之间的关系与运算，我们在此需要强调的是用概率论的语言来解释这些关系及运算。事件运算的详细内容，见 [1] 第 2 到 3 页。

定义完了事件，我们还想知道概率的定义。伊藤清说，概率就是测度。我们在概率论中的做法是对事件这个样本点的集合赋予一个取值在 $[0, 1]$ 之间的实数，也就是类比赋予可测集测度。因此概率是一个集函数，同时不可能事件赋予概率 0，必然事件赋予概率 1。同时，我们也赋予概率以类比测度定义中的可数可加性。然而，最重要的一点是集合的代数结构：测度是定义在 σ 代数上的，因此我们要求概率也定义在 σ 代数上，我们在概率论中称其为事件域，记为 \mathcal{F} 。

实际上，为概率测度引入 σ 代数结构是很自然的——当我们对事件 A 感兴趣， A 发生当然 A^c 不会发生，我们自然会对这个事件的对立事件 A^c 感兴趣；当我们对一列事件 $\{A_n\}_{n=1}^\infty$ 感兴趣，这一列事件中的任何一个发

生，并事件 $\cup_{n=1}^{\infty} A_n$ 也发生，我们也对 $\cup_{n=1}^{\infty} A_n$ 感兴趣。最后，把所有感兴趣的事件组成一个集族，其结构恰好就是 σ 代数的结构。

事实上，我们如果再将所有零概率集纳入事件域中，我们就称事件域是完备的。

定义 6.3 (概率)

定义在事件域 \mathcal{F} 上的一个集合函数 P 称为概率（或概率测度），如果它满足：

1. 非负性：对于任意的 $A \in \mathcal{F}$, $P(A) \geq 0$;
2. 规范性： $P(\Omega) = 1$;
3. 可列可加性或完全可加性：若 $A_n \in \mathcal{F}$ 互不相容，则

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



笔记 概率只对事件域中的元素，也就是样本空间的某些子集有定义。

推论来自 [2] 第 17 页。

推论 6.1

概率有如下性质：

1. $P(\emptyset) = 0$;
2. 若 $A, B \in \mathcal{F}, A \subset B$, 则 $P(B - A) = P(B) - P(A)$;
3. 若 $A, B \in \mathcal{F}$, 则 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
4. 单调性、次可加性、下连续性、上连续性可以直接从概率的测度本质得到。

第二条用可数可加性，取其余集合为空集即可；第三条只要注意 $A \cup B = A \cup (B - A)$, $B - A = B - (A \cap B)$.



于是类比测度空间 $(X, \sigma(X), \mu)$, 我们建立起概率空间，他是一个由样本空间、事件域和概率测度组成的三元组 (Ω, \mathcal{F}, P) .

定义 6.4 (概率空间)

称三元组 (Ω, \mathcal{F}, P) 是一个概率空间。其中 Ω 是样本空间， \mathcal{F} 是关于 Ω 的事件域， P 是 \mathcal{F} 上的概率测度。



这是 Kolmogorov 在 1933 年建立的概率论公理化结构。一个概率空间就是描述一个随机试验的数学模型。对于具体的随机试验如何构建概率空间，需要视具体问题而定。我们特别强调：此后，当概率空间确定好，只有事件域 \mathcal{F} 中的元素才称为事件，也就是说当表述中涉及到事件时，一定要明确它是哪个事件域的元素。并不是随意一个样本点的集合都是事件。当我们不能确定一个样本点 ω 是否属于 A ，就不能说 A 是一个事件。借用测度论的语言，我们说一个集合是事件，当且仅当它是关于事件域可测的。

概率论中常用的 Borel σ 代数 $\mathcal{B}(\mathbb{R})$ 是全体左开右闭区间 \mathcal{C} 生成的 σ 代数。它是一个范围很大的集类，包含了半开半闭区间（2 种）、点、闭区间、开区间。应用测度扩张定理，我们能把 \mathcal{C} 上的测度扩张到 $\mathcal{B}(\mathbb{R})$ 上。 n 维半开半闭矩体生成的 σ 代数记作 $\mathcal{B}(\mathbb{R}^n)$ 。Borel σ 代数中的元素称为 Borel 集。

下面给出和 Borel σ 代数息息相关的概念：Borel 函数。

定义 6.5 (Borel 函数)

设 $f : \mathbb{R}^n \rightarrow \mathbb{R}^d$. 若对任意的 $A \in \mathcal{B}(\mathbb{R}^d)$ 都有 $\{x : f(x) \in A\} \in \mathcal{B}(\mathbb{R}^n)$, 则称 f 是 Borel 函数。我们形式地把它写成 $f^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$.



6.2 随机变量、分布函数、随机变量的数字特征

6.2.1 随机变量

对于复杂的试验，我们总是希望利用简单事件的概率来推算出复杂事件的概率。这样，需要使用更多统一的数学方法来研究，比如将试验结果“数字化”，用一个数字 ξ 来表示。也就是说 $\xi : \Omega \rightarrow \mathbb{R}$ ，而且对任意的 $A \in \mathcal{B}(\mathbb{R})$ 满足

$$\xi^{-1}(A) \in \mathcal{F},$$

这种函数我们称为关于 \mathcal{F} 的随机变量。事实上，这个定义就是说随机变量 ξ 关于事件域 \mathcal{F} 可测。

 **笔记** 随机变量的定义依赖于事件域 \mathcal{F} 。设有 Borel 函数 f ，我们有些时候会写出 $f(\xi)$ 的符号，这个符号的意思是 $f \circ \xi(\omega)$ ，也就是 f 和 ξ 的复合。

如果 ξ 是关于 \mathcal{F} 的随机变量，则 $f(\xi)$ 也是关于 \mathcal{F} 的随机变量。

随机变量定义不依赖于概率测度的选取。

我们类比实变函数中“几乎处处相等”的概念来定义概率空间中的等价类。即随机变量 X, Y 相等当且仅当

$$P(\omega : X(\omega) \neq Y(\omega)) = 0.$$

对于有划分的样本空间上的随机变量，我们有下面重要的结论：

命题 6.1

设 $\{A_n\}_{n=1}^{\infty}$ 是一个 Ω 的划分，而 $\mathcal{F} = \sigma(\{A_n\}_{n=1}^{\infty})$ 。则 (Ω, \mathcal{F}) 上的任意一个随机变量 X 都能被分解为 $\{A_n\}_{n=1}^{\infty}$ 上的示性函数的线性组合

$$\sum_{n=1}^{\infty} a_n \chi_{A_n}, \quad a_n \in \mathbb{R}.$$



证明 由于 $\{A_n\}_{n=1}^{\infty}$ 是一个 Ω 的划分，且 $\{A_n\}_{n=1}^{\infty}$ 生成了 \mathcal{F} ，因此任取 $A \in \mathcal{F}$ ，都有 $A = \bigcup_{n \in L} A_n$ 。其中 L 是指标集。现在证明在 A_n 上任意随机变量的取值都是常数。如若不然，则存在 $\omega_1, \omega_2 \in A_n$ ，但是存在常数 t 使得 $X(\omega_1) \leq t < X(\omega_2)$ 。于是集合 $B = \{\omega : X(\omega) \leq t\} \in \mathcal{F}$ ，即 $B = \bigcup_{n \in J} A_n$ 。则 $\omega_1 \in A_n \cap B$ 但是 $\omega_2 \in A_n \cap B^c$ ，也就是说 A_n 可以继续划分下去，矛盾。于是在 A_n 上，任意随机变量的取值都为常数。

随机变量把事件映成 Borel 集。我们能由随机变量 ξ 得到多少信息？实际上，所有 Borel 函数和随机变量的复合 $f(\xi)$ 就“提取”出了 ξ 所有的信息，而 $f(\xi)$ 仍关于事件域可测。由实变函数知识，我们知道可测函数可以由简单函数逼近，也就是由示性函数的线性组合逼近。也就是说，样本空间上的示性函数就蕴含了随机变量的所有信息。而示性函数 χ_A 由等价于 $\{\xi^{-1}(A) : \forall A \in \mathcal{B}(\mathbb{R})\}$ 。可以验证，这是一个事件域。于是我们把 $\{\xi^{-1}(A) : \forall A \in \mathcal{B}(\mathbb{R})\}$ 称为由 ξ 生成的事件域，它记录了随机变量的所有信息。因此如果我们只考虑 ξ 上的事件，我们只用考虑概率空间 $(\Omega, \sigma(\xi), P)$ 。

关于随机变量生成的事件域，我们有如下性质：

性质 在概率空间 $(\Omega, \sigma(\xi), P)$ 上，

1. 任意一个随机变量 F ，都存在 Borel 函数 f ，使得 $F = f(\xi)$ 。
2. 对于任意的 Borel 函数 f ， $f(\xi)$ 都是 $(\Omega, \sigma(\xi), P)$ 上的随机变量。

$$\begin{array}{ccc} (\Omega, \sigma(X)) & \xrightarrow{X} & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ & \searrow F=f(X) & \downarrow f \\ & & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{array}$$

6.2.2 分布函数

定义 6.6 (概率分布)

设 ξ 是概率空间 (Ω, \mathcal{F}, P) 上的随机变量。若对于任意的 Borel 集 A , $P(\xi^{-1}(A))$ 都构成 $\mathcal{B}(\mathbb{R})$ 上的一个概率, 就说这是 ξ 的概率分布。

若随机变量 η, ξ 有相同的概率分布, 我们说它们同分布。



定义 6.7 (分布函数)

设 ξ 是概率空间 (Ω, \mathcal{F}, P) 上的随机变量。对于任意实数 x , 称

$$F(x) = P(\omega : \xi(\omega) \leq x)$$

为 ξ 的分布函数。

形式上, $F(x) = P(\xi^{-1}(A))$, A 是形如 $(-\infty, x]$ 的 Borel 集全体。



性质

1. 单调不减, 右连续。
2. $F(x) \rightarrow 1, x \rightarrow \infty; F(x) \rightarrow 0, x \rightarrow -\infty.$

证明

1. 单调不减性由概率的单调性直接得出。右连续性: 对于任意正数 ϵ , 有 $F(x + \epsilon) - F(x) = P(\omega : \xi(\omega) \leq x + \epsilon) - P(\omega : \xi(\omega) \leq x) \leq P(\omega : x < \xi(\omega) \leq x + \epsilon) \rightarrow 0$. 其中最后一个式子用了概率的上连续性。
2. 分别利用概率的下、上连续性。

反过来, 如果一个函数满足分布函数的两条性质, 则存在概率空间 (Ω, \mathcal{F}, P) 和上面的随机变量 ξ 使得 $F(x)$ 是 ξ 的分布函数。一个平凡的例子是 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$, 其中 $F(x) = P((-\infty, x])$, $\xi(\omega) = \omega$.

我们可以把随机变量分成三种。

1. 离散型: 取值只有有限个或者可列个的随机变量 ξ .
2. 连续型: 分布函数 $F(x)$ 是连续的, 可以写成 $\int_{-\infty}^x p(t)dt$ 的随机变量 ξ , 其中 $p(x)$ 非负。称为密度函数, 且满足 $\frac{d}{dx}F(x) = p(x)$ 和 $\int_{\mathbb{R}} p(t)dt = 1$.
3. 奇异型: 分布函数是连续的, 但不存在密度函数使其可以表示成不定积分的形式.

Lebesgue 分解定理也告诉我们任何随机变量的分布都是这三种类型随机变量分布的混合.

6.2.3 随机向量

定义 6.8 (随机向量)

设有 (Ω, \mathcal{F}) 上的随机变量 $\xi_1, \xi_2, \dots, \xi_n$, 则称 $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ 为 n 维随机向量。



 **笔记** 实际上, n 维随机向量 ξ 是 Ω 到 $\mathcal{B}(\mathbb{R}^n)$ 上的函数。对任意的 $A \in \mathcal{B}(\mathbb{R}^n)$ 有 $\xi^{-1}(A) \in \mathcal{F}$.

我们知道 Lebesgue 可测函数在加法、数乘、乘法、上下极限运算下封闭, 类比之下得出随机变量也在加法、数乘、乘法、上下极限运算下封闭。

将一元随机变量概率分布定义中的 $\mathcal{B}\mathbb{R}$ 改为 $\mathcal{B}\mathbb{R}^n$, 就得到了随机向量概率分布的概念。

随机向量也有分布函数。

定义 6.9 (联合分布函数)

设有 (Ω, \mathcal{F}) 上的随机向量 $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, 对任意的 $(\xi_1, \xi_2, \dots, \xi_n) \in \mathbb{R}^n$, 称 \mathbb{R}^n 上的 n 元函数

$$F(x_1, x_2, \dots, x_n) = P(\omega : \xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n)$$

为随机向量 ξ 的联合分布函数。



和一元随机变量生成的事件域类比，随机向量也能生成事件域，这个事件域为 $\sigma(\{\xi^{-1}(A); \forall A \in \mathcal{B}(\mathbb{R})\})$ 。与一元随机变量相比，随机向量多了边际分布的概念。详见 [1]8 到 9 页。

6.2.4 随机变量的数字特征

定义 6.10 (数学期望)

设 ξ 是概率空间 (Ω, \mathcal{F}, P) 上的随机变量。若 $\int_{\mathbb{R}} |\xi(\omega)| dP(\omega) < \infty$, 即 $\xi(\omega) \in L^1(\Omega, \mathcal{F}, P)$, 则称随机变量 ξ 的数学期望存在。 ξ 的数学期望定义为

$$\int_{\mathbb{R}} \xi(\omega) dP(\omega).$$

这里的积分是 Lebesgue 积分。



笔记 Lebesgue 积分可以理解成加权平均：当 $\xi(\omega) \in (x - \epsilon, x + \epsilon]$, 对应的概率权重是 $P(\omega : x - \epsilon < \xi(\omega) \leq x + \epsilon)$.

作加权平均，得到 $\sum_i x_i(\omega) P(\omega : \xi(\omega) \in (x - \epsilon, x + \epsilon])$, 再将分割取极限，就得到了期望的定义

我们还能得到个重要的推论：设 f 是 Borel 函数，若 $f(\xi(\omega)) \in L^1$, 则 $f(\xi(\omega))$ 期望存在，其期望为 $\int_{\mathbb{R}} f(\xi(\omega)) dP(\omega)$.

从抽象积分定义出发，我们得到离散型随机变量和连续型随机变量的期望计算式：设 f 是 Borel 函数， ξ 是 (Ω, \mathcal{F}, P) 上的随机变量且期望存在，

1. 若 ξ 是离散型随机变量，取值为 $x_1, x_2, \dots, x_n, \dots$ 则

$$E(\xi) = \sum_i f(x_i)p_i.$$

2. 若 ξ 是连续型随机变量，密度函数为 $p(x)$, 则

$$E(\xi) = \int_{\mathbb{R}} f(x)p(x)dx.$$

数学期望有如下性质。

性质

1. $E\chi_A = P(A)$;
2. 若 $\xi \geq 0$, 则 $E\xi \geq 0$. 从而我们有：若 $\eta \geq \xi$, 则 $E\eta \geq E\xi$ 。
3. 线性性：对任意实数 a, b , 有 $E(a\eta + b\xi) = aE\eta + bE\xi$;
4. $|E\xi| \leq E|\xi|$;
5. (Cauchy-Schwarz) 若 $\eta, \xi \in L^2(\Omega, \mathcal{F}, P)$, 则 $[E\eta\xi]^2 \leq E\xi^2 \cdot E\eta^2$. 当且仅当存在实数 a, b 使得 $a\xi + b\eta = 0$, $a.s.\omega \in \Omega$ 时成立。

其他数字特征见 [1] 第 13-14 页。

6.2.5 独立性

独立性是随机变量之间一种的重要关系。

定义 6.11 (独立性)

设有随机变量 $\xi_1, \xi_2, \dots, \xi_n$. 若对于任意的 Borel 集 $\{A_k\}_{k=1}^n$, 都有

$$P(\omega : \xi_1(\omega) \in A_1, \dots, \xi_n(\omega) \in A_n) = P(\omega : \xi_1(\omega) \in A_1) \cdots P(\omega : \xi_n(\omega) \in A_n)$$

则称 $\xi_1, \xi_2, \dots, \xi_n$ 相互独立。



笔记 若随机变量的期望存在，则 $\xi_1, \xi_2, \dots, \xi_n$ 相互独立的另一种判定方法是 $E[\prod_{k=1}^n \xi_k] = \prod_{k=1}^n E[\xi_k]$.

6.3 随机变量的收敛性

我们在分析中学习过可测函数的几种收敛方式：几乎处处收敛、按测度收敛、 L^p 收敛、弱 * 收敛。类比这些收敛方式，随机变量同样也有几乎必然收敛、按概率收敛、 L^p 收敛、按分布收敛。下面就来介绍这几种收敛方式。

定义 6.12 (几乎必然收敛 (a.s.))

设有随机变量列 $\{\xi_n\}_{n=1}^\infty$, 若存在随机变量 ξ 使得

$$P(\omega : \lim_{n \rightarrow \infty} \xi_n \neq \xi) = 0,$$

则称随机变量列 $\{\xi_n\}_{n=1}^\infty$ 几乎必然收敛于随机变量 ξ , 记作 $\xi_n \rightarrow \xi, a.s.$.



定义 6.13 (按概率收敛)

设有随机变量列 $\{\xi_n\}_{n=1}^\infty$, 若对任意正数 ϵ , 存在随机变量 ξ 使得

$$\lim_{n \rightarrow \infty} P(\omega : |\xi_n - \xi| \geq \epsilon) = 0,$$

则称随机变量列 $\{\xi_n\}_{n=1}^\infty$ 按概率收敛于随机变量 ξ , 记作 $\xi_n \xrightarrow{P} \xi$.



定义 6.14 (L^p 收敛)

设有随机变量列 $\{\xi_n\}_{n=1}^\infty$, 存在随机变量 ξ 使得

$$\lim_{n \rightarrow \infty} E[|\xi_n - \xi|^p] = 0,$$

则称随机变量列 $\{\xi_n\}_{n=1}^\infty$ L^p 收敛于随机变量 ξ , 记作 $\xi_n \xrightarrow{L^p} \xi$.

这是一种范数收敛, 即强收敛。 $p = 2$ 时称为均方收敛。



定义 6.15 (弱收敛 (按分布收敛))

设 $\{F_n(x)\}$ 是一列分布函数, 若存在单调不减函数 $F(x)$, 使得在 $F(x)$ 的所有连续点上都有 $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, 则说 $F_n(x)$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$.

$F_n(x) \xrightarrow{W} F(x)$ 的充要条件是 $\psi_n(x)$ 点态收敛于 $\psi(x)$, 其中 $\psi_n(x), \psi(x)$ 分别是 $F_n(x), F(x)$ 的特征函数。



Helly-Bray 说明了弱收敛实际上就是弱 * 收敛。

定理 6.1 (Helly-Bray)

设有分布函数 F, F_n . 则 $F_n \xrightarrow{W} F$ 当且仅当对任意的 $f \in C_b(\mathbb{R})$ 有

$$\int f dF \rightarrow \int f dF_n.$$

其中 $\int f dF$ 是分布函数 F 诱导的有界线性泛函。



这几种收敛方式的关系也类似于几乎处处收敛、按测度收敛、 L^p 收敛、弱 * 收敛之间的关系, 也就是由 Egorov 定理知道几乎必然收敛蕴含按概率收敛, 由 Riesz 定理知道按概率收敛蕴含随机变量有子列几乎必然收敛。另外, 按概率收敛蕴含弱收敛。然而, 几乎必然收敛和 L^p 收敛没有必然关系, 但是 L^p 收敛蕴含按概率收敛。

下面类比可测函数的性质给出随机变量的一些可测性质。

性质

1. 类比 Beppo-Levi, 设有单调递增随机变量列 $\{\xi_n\}$, 且 $\xi_n \xrightarrow{a.s.} \xi$, 则

$$\lim_{n \rightarrow \infty} E[\xi_n] = E[\lim_{n \rightarrow \infty} \xi_n] = E[\xi].$$

这可以得出可列线性：设 $\{\xi_n\}$ 是一列非负随机变量，则

$$\sum_{n \in \mathbb{N}} E[\xi_n] = E\left[\sum_{n \in \mathbb{N}} \xi_n\right].$$

证明只要设 $S_n = \sum_{k=1}^n \xi_k$, $S = \sum_{n=1}^{\infty} \xi_n$, 并注意到 $S_n \xrightarrow{a.s.} S$ 且 S_n 上升。

2. 类比 Beppo-Levi 的单调递减版，设有单调递减随机变量列 $\{\xi_n\}$ ，且 $\xi_n \xrightarrow{a.s.} \xi$, $E[\xi_1] < \infty$, 则

$$\lim_{n \rightarrow \infty} E[\xi_n] = E\left[\lim_{n \rightarrow \infty} \xi_n\right] = E[\xi].$$

3. (Fatou) 设 $\{\xi_n\}$ 是一列非负随机变量，则

$$E\left[\liminf_{n \rightarrow \infty} \xi_n\right] \leq \liminf_{n \rightarrow \infty} E[\xi_n] \leq \limsup_{n \rightarrow \infty} E[\xi_n] \leq E\left[\limsup_{n \rightarrow \infty} \xi_n\right].$$

4. (Lebesgue 控制收敛定理) 设 $\{\xi_n\}$ 是一列随机变量，且 $|\xi_n| < X$, $X \in L^1(\Omega, \mathcal{F}, P)$, 且 $\xi_n \xrightarrow{a.s.} \xi$ 或 $\xi_n \xrightarrow{P} \xi$, 则

$$E\left[\lim_{n \rightarrow \infty} \xi_n\right] = E[\xi] = \lim_{n \rightarrow \infty} E[\xi_n].$$

上面可列线性中的非负条件可以去掉，只要注意 $\sum_{k=1}^n \xi \leq \sum_{n=1}^{\infty} |\xi_n| < \infty$, 且有 $\xi_n \xrightarrow{a.s.} \xi$.

第7章 条件期望

内容提要

□ 空间 $L^2(\Omega, \mathcal{F}, P)$ 与其性质

□ 条件期望与其表达式

我们在初等概率论中学过“静态”的条件期望： $E[Y|X = x_i]$ 。这个式子的意思是，当固定随机变量 X 的值于 x_i 时，随机变量 Y 的期望，这相当于函数在一个点 x_i 的取值。然而如果我们从动态的角度来看待，让 X 的取值“动起来”——静态的条件期望变成 $g(x) = E[Y|X = x]$ ，这是一个以 x 为自变量的函数。带入随机变量 X ，条件期望就变为 $g(X) = E[Y|X]$ ，这同样是一个随机变量。下面我们将从集合角度来看条件期望的种种性质。

7.1 $L^2(\Omega, \mathcal{F}, P)$ 空间

定义 7.1 ($L^2(\Omega, \mathcal{F}, P)$ 空间)

满足 $E\xi < \infty$ 的随机变量的全体构成了一个赋范线性空间，并且可以在上面定义内积，从而使它是一个 Hilbert 空间，记作 $L^2(\Omega, \mathcal{F}, P)$ 。



证明 线性由期望的线性得出：对于 $\xi, \eta \in L^2(\Omega, \mathcal{F}, P)$ ，任取实数 a, b ， $E[a\xi + b\eta] = aE[\xi] + bE[\eta] < \infty$ 。

我们在 $L^2(\Omega, \mathcal{F}, P)$ 上定义度量 $d(\xi, \eta) = \sqrt{E(\xi - \eta)^2}$ ，于是范数 $||\xi|| = \sqrt{E\xi^2}$ 。可以验证 $\sqrt{E\xi^2}$ 满足范数的四个条件。同时我们还可以定义内积： $\langle \xi, \eta \rangle = E[\xi\eta]$ 。也容易验证内积的几条性质成立。

我们想用已知的信息 $Y = (Y_1, Y_2, \dots, Y_n)$ 来预测随机变量 X 。也就是说我们想用 Y 生成的 σ 代数中的元素来最佳近似 X 。而我们知道，最佳近似就是 (Ω, \mathcal{F}, P) 到 $(\Omega, \sigma(Y), P)$ 上的投影。这个投影就是我们讨论的条件期望。

实际上， $(\Omega, \sigma(Y), P)$ 是 (Ω, \mathcal{F}, P) 的子空间，而且是一个 Hilbert 空间。它的内积继承于 $L^2(\Omega, \mathcal{F}, P)$ 。

严格地说，条件期望的定义如下。

定义 7.2 (条件期望)

$X \in L^2(\Omega, \mathcal{F}, P), Y$ 是随机向量。则随机变量 $f(Y)$ 被称为 X 关于 Y 的条件期望，如果：

1. $f(Y) \in L^2(\Omega, \mathcal{F}, P)$;
2. 对任意的 $g(Y) \in L^2(\Omega, \mathcal{F}, P)$ ，成立

$$E[f(Y)g(Y)] = E[Xg(Y)].$$

我们把 $f(Y)$ 记为 $E[X|Y]$ 。



笔记 实际上定义中的第二条就是 $E[(f(Y) - X)g(Y)] = 0$ ，也即 $\langle f(Y) - X, g(Y) \rangle = 0$ ，这就是 $f(Y) - X \perp g(Y)$ 。这就是说， $f(Y)$ 是 X 在 $(\Omega, \sigma(Y), P)$ 上的投影。我们知道投影是内积意义上的距离最近逼近，因此条件期望 $E[X|Y]$ 是在已知 Y 时对 X 最好的逼近或预测。

7.2 条件期望的形式

这一节中我们给出条件期望在离散型随机变量和连续性随机变量的形式。

例题 7.1 设有离散型随机变量 (X, Y) ，其概率分布为 $P(X = x_i, Y = y_j) = p_{ij}$ 。假定 $p_{ij} > 0$ ，求 $E[X|Y]$ 。

证明 设 $E[X|Y] = f(Y(\omega))$ ，而 Y 是离散型随机变量，因此 $Y = y_j$ 是样本空间的一个划分。于是我们知道 $f(Y)$

可以写成示性函数的线性组合

$$f(Y(\omega)) = \sum_{j \geq 1} f(y_j) \chi_{\{y_j\}}(Y(\omega)).$$

下面我们通过条件期望的定义 $E[f(Y)g(Y)] = E[Xg(Y)]$, $g(Y)$ 为任意 Borel 函数。由于

$$E[f(Y)g(Y)] = \sum_{j \geq 1} f(y_j)g(y_j) \sum_{i \geq 1} p_{ij} = \sum_{i,j} f(y_j)g(y_j)p_{ij} = E[Xg(Y)] = \sum_{i,j} x_i g(y_j)p_{ij},$$

我们取 $g(Y) = \chi_{\{y_j\}}(Y(\omega))$ 来得出条件期望的形式。于是有

$$\sum_{j \geq 1} f(y_j) \chi_{\{y_j\}}(Y(\omega)) p_{ij} = f(y_j) \sum_{i \geq 1} p_{ij} = \sum_{i,j} x_i \chi_{\{y_j\}}(Y(\omega)) p_{ij} = \sum_{i \geq 1} x_i p_{ij}.$$

于是得到

$$f(y_j) = \frac{\sum_{i \geq 1} x_i p_{ij}}{\sum_{i \geq 1} p_{ij}} = E[X|Y = y_j].$$

所以

$$E[X|Y] = \sum_{j \geq 1} E[X|Y = y_j] \chi_{\{y_j\}}(Y(\omega)).$$

其中 $E[X|Y = y_j]$ 是初等概率论中给定 $Y = y_j$ 时的条件期望。

例题 7.2 设有连续性随机变量 X, Y , 密度函数为 $p(x, y) > 0, \forall (x, y) \in \mathbb{R}^2, X \in L^2(\Omega, \mathcal{F}, P)$. 求 $E[X|Y]$.

证明 依旧由定义式 $E[f(Y)g(Y)] = E[Xg(Y)]$, $g(Y)$ 为任意 Borel 函数出发, 我们有

$$E[f(Y)g(Y)] = \int_{\mathbb{R}} g(y)(f(y) \int_{\mathbb{R}} p(x, y) dx) dy,$$

而

$$E[Xg(Y)] = \iint_{\mathbb{R}^2} x g(y) p(x, y) dx dy = \int_{\mathbb{R}} g(y) (\int_{\mathbb{R}} x p(x, y) dx) dy,$$

由实变函数知识, 有

$$f(y) = \frac{\int_{\mathbb{R}} x p(x, y) dx}{\int_{\mathbb{R}} p(x, y) dx} = E[X|Y = y].$$

因此

$$E[X|Y] = E[X|Y = y]|_{y=Y} = \frac{\int_{\mathbb{R}} x p(x, Y) dx}{\int_{\mathbb{R}} p(x, Y) dx}$$

 **笔记** 上面的例子只考虑了 p_{ij} 或者 $p(x, y)$ 都大于零的情况。如果没有大于零的假设, 结果将会是:

1. 离散型:

$$E[X|Y] = \sum_{j \geq 1} E[X|Y = y_j] \chi_{\{y_j\}}(Y(\omega)) \chi_{p_{ij} > 0}(Y(\omega)) + c_k \chi_{\{p_{ij} = 0\}}(Y(\omega)).$$

其中 c_k 是任意常数。我们发现条件期望并不是唯一的, 但它在几乎必然相等等价类下是唯一的。

2. 连续型:

$$E[X|Y] = E[X|Y = y]|_{y=Y} = \frac{\int_{\mathbb{R}} x p(x, Y) dx}{\int_{\mathbb{R}} p(x, Y) dx} \chi_{\{p(x, y) > 0\}}(Y(\omega)) + c(y) \chi_{\{p(x, y) = 0\}}(Y(\omega)).$$

其中 $c(y)$ 是任意 Borel 函数。

7.3 推广条件下的条件期望

由于高阶矩存在可以推出低阶矩存在, 我们知道 $L^2(\Omega, \mathcal{F}, P) \subset L^1(\Omega, \mathcal{F}, P)$, 上面条件期望的假设是 $X, f(Y) \in L^2(\Omega, \mathcal{F}, P)$, 现在我们把这个条件放宽到 $X, f(Y) \in L^1(\Omega, \mathcal{F}, P)$ 来定义条件期望。

定义 7.3 (条件期望 (L^1))

设随机变量 $X \in L^1(\Omega, \mathcal{F}, P)$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是 Borel 函数, 我们称 $f(Y)$ 是 X 关于 n 维随机向量 Y 的条件期望, 如果

1. $f \in L^1(\Omega, \mathcal{F}, P)$;
2. $E[f(Y)g(Y)] = E[Xg(Y)]$, 其中 $g(Y)$ 是任意的有界 Borel 函数。



我们自然要问, 设 $X \in L^2(\Omega, \mathcal{F}, P)$, 那么对于任意的 $g(Y) \in L^2(\Omega, \mathcal{F}, P)$, 若

$$\|X - f_1(Y)\|_{L^1} = \inf_{g(Y) \in L^1(\Omega, \mathcal{F}, P)} \|X - g(Y)\|_{L^1},$$

$$\|X - f_2(Y)\|_{L^2} = \inf_{g(Y) \in L^2(\Omega, \mathcal{F}, P)} \|X - g(Y)\|_{L^2},$$

那么 f_1 和 f_2 有什么关系? 实际上, L^1 , L^2 范数下的投影分别为条件中位数和条件期望, 它们在分布是对称的时候是一致的。我们来看两个例子。

例题 7.3 对于连续型随机变量 X , 求 v , 使得 $E[|X - v|] = \inf_{c \in \mathbb{R}} E[|X - c|]$.

证明 我们直接最小化 $f(c) = E[|X - c|]$. 由于

$$f(c) = \int_{\mathbb{R}} |x - c| p(x) dx = \int_{-\infty}^c (c - x) p(x) dx + \int_c^{\infty} (x - c) p(x) dx,$$

两边同时求导, 得到

$$f'(c) = \int_{-\infty}^c p(x) dx - \int_c^{\infty} p(x) dx = F(c) - (1 - F(c)) = 0,$$

解得驻点 $c = F^{-1}(\frac{1}{2})$, 也就是说 c 是分布 F 的中位数。由于 $f''(c) = 2p(c) > 0$, 因此 $v = F^{-1}(\frac{1}{2})$ 最小化 $f(c)$.

例题 7.4 设随机变量 $X \in L^2(\Omega, \mathcal{F}, P)$, 求 v 使得 $E[(X - v)^2] = \inf_{c \in \mathbb{R}} E[(X - c)^2]$.

证明 这里没有做 X 是什么类型的随机变量的假设, 我们可以用期望的线性性来做。设 $f(c) = E[(X - c)^2] = E[X^2] - 2cE[X] + c^2$, 这是一个简单的二次函数求最小值问题, 解得最小值点为 $v = E[X]$.

我们再来看看条件期望和条件概率的关系。

例题 7.5 设随机变量 $X = \chi_B(\omega)$, $Y = \chi_A(\omega)$, 于是

$$E[X|Y](\omega) = E[X|Y=0]\chi_{A^c}(\omega) + E[X|Y=1]\chi_A(\omega)$$

运用静态条件期望公式, 我们知道

$$E[X|Y=0] = P(X=1|Y=0) = \frac{P(A^c \cap B)}{P(B)} = P(B|A^c),$$

$$E[X|Y=1] = P(X=1|Y=1) = \frac{P(A \cap B)}{P(B)} = P(B|A).$$

于是

$$E[X|Y](\omega) = P(B|A^c)\chi_{A^c}(\omega) + P(B|A)\chi_A(\omega).$$

A 发生时, 这个条件期望给出 $= P(B|A)$; A 不发生时, 这个条件期望给出 $= P(B|A^c)$.

7.4 条件期望的性质

作为条件期望定义式的直接应用, 我们有全期望公式:

定理 7.1 (全期望公式)

$$E[E[X|Y]] = E[X].$$



证明 在条件期望定义式 $E[f(Y)g(Y)] = E[Xg(Y)]$ 中取 $g(Y)$ 恒为 1 即可。

笔记 实际上全期望公式可以推出全概率公式。只要取 $X = \chi_A(\omega)$, 设离散型随机变量 Y 的分布列为 $P(Y = y_i) =$

p_i , 于是就有

$$E[X|Y] = \sum_{i \geq 1} E[X|Y=y_i] \chi_{\{y_i\}}(Y(\omega)) = \sum_{i \geq 1} P(A|Y=y_i) \chi_{\{y_i\}}(Y(\omega)),$$

于是

$$E[X] = P(A) = E[E[X|Y]] = \sum_{i \geq 1} P(A|Y=y_i)P(Y=y_i).$$

下面我们再给出连续型随机变量的全期望公式的计算式。

命题 7.1 (连续型随机变量的全期望公式)

设 (X, Y) 是连续性随机变量, 密度函数 $p(x, y) > 0$, 设边缘密度函数 $p_X(x) = \int_{\mathbb{R}} p(x, y) dy$, $p_Y(y) = \int_{\mathbb{R}} p(x, y) dx$. 于是条件密度函数 $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$. 则

$$E[X] = \int_{\mathbb{R}} E[X|Y=y] p_Y(y) dy.$$



证明

$$E[X|Y=y] = \int_{\mathbb{R}} x p_{X|Y}(x|y) dx,$$

$$E[X|Y] = \int_{\mathbb{R}} x p_{X|Y}(x|Y) dx.$$

因此

$$E[X] = E[E[X|Y]] = \int_{\mathbb{R}} E[X|Y=y] p_Y(y) dy.$$

条件期望具有线性和单调性。

性质

1. 设 $X_1, X_2 \in L^1(\Omega, \mathcal{F}, P)$, Y 是随机向量, 对任意实数 a, b , 有

$$E[aX_1 + bX_2|Y] = aE[X_1|Y] + bE[X_2|Y];$$

2. 若 X 非负, 则

$$E[X|Y] \geq 0.$$

3. 若 $h \in L^1(\Omega, \mathcal{F}, P)$ 为 n 维 Borel 函数, 则

$$E[h(Y)|Y] = h(Y).$$

4. (独立性) 若随机向量 X, Y 相互独立, 则对于 m 维 Borel 函数 $h \in L^1(\Omega, \mathcal{F}, P)$ 有

$$E[h(X)|Y] = E[h(X)].$$

笔记 设 $\psi : L^1(\Omega, \mathcal{F}, P) \rightarrow L^1(\Omega, \sigma(Y), P)$, 第一条实际上说明了条件期望 (投影 ψ) 保留了 $L^1(\Omega, \mathcal{F}, P)$ 上的线性结构, 因此两个空间线性同构, 即

$$L^1(\Omega, \mathcal{F}, P) /_{ker \psi} \simeq L^1(\Omega, \sigma(Y), P).$$

容易知道, 投影 ψ 幂等, 即 $\psi^2 = \psi$.

第三条性质容易理解: $h(Y)$ 在 $\sigma(Y)$ 上的投影还是 $h(Y)$. 当然, 我们也能理解成当我们取定一个 $Y = y$ 时 $h(Y) = h(y)$, 再让 y “动起来” 成为 Y , 就有了性质三。

我们再给出一条重要的性质:

性质 设 $g, h \in L^1(\Omega, \mathcal{F}, P)$ 分别为 $n.m$ 维 Borel 函数, 且 $gh \in L^1(\Omega, \mathcal{F}, P)$, 则

$$E[h(Y)g(X)|Y] = h(Y)E[g(X)|Y].$$

笔记 我们可以这样理解这条性质: 先从静态的角度出发, $E[h(Y)g(x)|Y=y]$ 就相当于固定 $Y = y$ 后对 $h(Y)g(X)$ 取期望, 这个期望当然就是 $h(y)E[g(X)|Y=y]$. 然后再让 $Y = y$ 动起来, 也就是说用 Y 代替 y , 就得到了

$$E[h(Y)g(X)|Y=y]_{y=Y} = h(Y)E[g(X)|Y=y]_{y=Y} = h(Y)E[g(X)|Y].$$

例题 7.6 设 $X \sim [0, 1]$, 在 $X = x$ 的条件下, $Y \sim N(x, x^2)$. 求 $EY, VY, \text{cov}(X, Y)$.

证明 由 $p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{(y-x)^2}{2x^2}\right)$ 知道 $E[Y|X=x] = x$, 于是 $E[Y|X] = X$. 由全期望公式知道 $EY = E[E[Y|X]] = E[X] = \frac{1}{2}$.

另一方面, $VY = E[(Y - EY)^2] = E[E[(Y - EY)^2|X]] = \int_0^1 dx \int_{\mathbb{R}} (y - \frac{1}{2})^2 \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{(y-x)^2}{2x^2}\right) dy = \frac{5}{12}$.

对于 $\text{cov}(X, Y) = E[XY] - EX \cdot EY$, 我们知道

$$E[XY] = E[E[XY|X]], E[XY|X] = X \cdot E[Y|X] = X^2,$$

因此

$$E[XY] = E[X^2] = \frac{1}{3}.$$

故 $\text{cov}(X, Y) = \frac{1}{12}$.

我们来推广上面的性质。

命题 7.2

设有 n, m 维随机向量 X, Y , 设有 Borel 函数 $h(x, y)$ 使得 $h(X, Y) \in L^1(\Omega, \mathcal{F}, P)$, 且对于任意实数 y , 都有 $h(X, y) \in L^1(\Omega, \mathcal{F}, P)$, 则有

$$E[h(X, Y)|Y] = E[h(X, y)|Y=y]_{y=Y}.$$

如果 X, Y 相互独立, 则

$$E[h(X, Y)|Y] = E[h(X, y)]_{y=Y}.$$

例题 7.7 设随机变量 X, Y 独立, 分别服从参数为 λ, μ 的指数分布。设随机变量 $Z = \chi_{\{X \leq Y\}}$, 求 EZ .

证明 设 $h(x, y) = \chi_{\{x \leq y\}}$, 于是由上一条性质知道

$$E[h(X, Y|Y=y)] = \int_0^y p_X(x)dx \cdot \chi_{\{y \geq 0\}} = \int_0^y \lambda e^{-\lambda x} dx \cdot \chi_{\{y \geq 0\}}$$

即

$$E[h(X, Y|Y=y)] = (1 - e^{-\lambda y}) \cdot \chi_{\{y \geq 0\}}.$$

由于 X, Y 独立, 因此 $E[h(X, Y)|Y] = E[h(X, Y|Y=y)]_{y=Y} = (1 - e^{-\lambda Y}) \cdot \chi_{\{Y \geq 0\}}$. 由全期望公式:

$$E[h(X, Y)] = E[E[h(X, Y)|Y]] = \int_0^\infty (1 - e^{-\lambda y}) \mu e^{-\mu y} dy = \frac{\lambda}{\lambda + \mu}.$$

例题 7.8 续例 4.6, 证明 $\frac{Y}{X}$ 和 X 相互独立。

证明 要证明 $\frac{Y}{X}$ 和 X 相互独立, 可以证明对于任意的 Borel 函数 f, g 有 $E[f(\frac{Y}{X})g(X)] = E[f(\frac{Y}{X})]E[g(X)]$. 由全期望公式和上述性质, 我们有

$$E[f(\frac{Y}{X})g(X)] = E[E[f(\frac{Y}{X})g(X)|X]] = E[g(X)E[f(\frac{Y}{X})|X]],$$

于是我们先求 $E[f(\frac{Y}{X})|X]$, 由上面的性质, 我们要求 $E[f(\frac{Y}{X})|X=z]$. 我们记 $\phi(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

于是 $E[f(\frac{Y}{X})|X=z] = \int_{\mathbb{R}} f(\frac{y}{z})\phi(y; z, z^2)dy$, 因此 $E[f(\frac{Y}{X})|X] = \int_{\mathbb{R}} f(\frac{y}{X})\phi(y; X, X^2)dy$. 由上面的性质,

$$E[\frac{Y}{X}|X] = E[\frac{Y}{X}|X]_{x=X} = \int_{\mathbb{R}} f(\frac{y}{X})\phi(y; X, X^2)dy,$$

做变量代换, 设 $z = \frac{y}{X}$, 则上式变为

$$\int_{\mathbb{R}} f(z)\phi(z, 1, 1)dz = E[f(Z)], Z \sim N(1, 1).$$

显然有 $E[g(X)E[f(Z)]] = E[g(X)]E[f(Z)]$, 这就证明了 $\frac{Y}{X}$ 和 X 相互独立。

实际上, 若采用 4.3 中条件期望的定义, 如果进一步有 $h \in L^2(\Omega, \mathcal{F}, P)$, 那么这里第条件期望和 4.2 中定义的条件期望等价。因为对于任意 $g(Y) \in L^2(\Omega, \mathcal{F}, P)$, $\psi(h(X)) = f(Y)$, 于是 $E[f(Y)g(Y)] = E[h(X)g(Y)]$. 如果 $g(Y)$ 有界, 那么等号右边可以写成 $E[g(Y)E[h(X)|Y]]$, 因此 $E[g(Y)(f(Y) - E[X|Y])] = 0$, 由 $g(Y)$ 的任意性知

道 $f(Y) = E[X|Y]$. 最后我们给出一个关于投影的结论。

命题 7.3

设随机变量 $X \in L^1(\Omega, \mathcal{F}, P)$, 有随机向量 Y, Z . 则

$$E[Z] = E[E[X|Y, Z]|Z] = E[E[X|Z]|Y, Z].$$



 **笔记** 实际上这就是说从 \mathcal{F} 先后向 $\sigma(Y, Z), \sigma(Z)$ 上投影的结果就是先后向 $\sigma(Z), \sigma(Y, Z)$ 上投影的结果, 两者都等于 $E[Z]$.

第8章 随机过程

内容提要

- 随机过程的基本概念
- 马氏过程（离散、连续时间）
- 泊松过程
- 布朗运动

8.1 随机过程的基本概念

随机向量是把一列随机变量放在一起时产生的组，我们推广这个概念，把一族随机变量放在一起，就产生了随机过程。

8.1.1 基本概念

定义 8.1 (随机过程)

设 (Ω, \mathcal{F}, P) 是概率空间， $T \subset \mathbb{R}$ 是参数集。如果对任意的 $t \in T$ ，都存在随机变量 $X_t(\omega)$ ，则称 $\{X_t(\omega), t \in T\}$ 为定义在 (Ω, \mathcal{F}, P) 上的随机过程。如果 $T = \{0, 1, 2, \dots\}$ 或 $T = \{t_0, t_1, t_2, \dots\}$ ，则称为离散时间随机过程；如果 T 是实轴的子集则称为连续时间随机过程。

$X_t(\omega)$ 关于 t, ω 的所有取值集合称为状态空间 S 。如果 S 是可数集，则称为离散状态随机过程；若 S 是实轴的子集，称为连续状态随机过程。



笔记 实际上随机过程 $\{X_t(\omega), t \in T\}$ 是一个 $(\Omega, T) \rightarrow S$ 的二元函数，我们不妨记作 $X(t, \omega)$ 。 $(\Omega, T) \rightarrow S$ 的所有映射就是随机过程。

我们仅要求 $X(t, \omega)$ 对 ω 可测，而不要求对 t 可测。当我们固定 t ，就得到了随机变量 $X_t(\omega)$ ；当我们固定 ω ，就得到了函数 $X(\cdot, \omega)$ ，我们称它为随机过程对于 ω 的轨道。

既然说随机过程是拓展的随机向量，那么随机变量的有限联合分布和随机过程有什么关系？Kolmogorov 定理回答了这个问题。我们先定义有限维分布。

定义 8.2 (有限维分布)

对任意有限个 $t_1, t_2, \dots, t_n \in T$ 定义随机过程的 n 维分布函数

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n).$$

随机过程的有限维分布全体就是

$$\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) : n \geq 1\}.$$

称为随机过程的有限分布族。



知道了随机过程的有限分布族，就知道了随机过程所有的有限分布。然后我们给出有限维分布族的联合矩母函数的定义。

定义 8.3 (联合矩母函数)

设随机过程 $\{X_t(\omega), t \in T\}$ 的有限维分布族为 $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) : n \geq 1\}$ 。如果对任意的 $n \geq 1$ 和 $t_1, t_2, \dots, t_n \in T$ ，都有 $E[e^{\sum_{i=1}^n \theta_i X_{t_i}}] < \infty$, $\theta_i \in \mathbb{R}$ ，则称

$$M_{t_1, t_2, \dots, t_n}(\theta_1, \theta_2, \dots, \theta_n) = E[e^{\sum_{i=1}^n \theta_i X_{t_i}}]$$

为随机过程的联合矩母函数。



我们可以借矩母函数刻画随机过程的独立性。

命题 8.1 (随机过程的独立性)

设随机过程 $\{X_t(\omega), t \in T\}$ 的有限维分布族为 $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) : n \geq 1\}$. 如果对任意的 $n \geq 1$ 和 $t_1, t_2, \dots, t_n \in T$, 都有 $E[e^{\sum_{i=1}^n \theta_i X_{t_i}}] < \infty$, $\theta_i \in \mathbb{R}$, 则随机变量族 $\{X_t, t \in T\}$ 相互独立的充分必要条件是对任意的 $n \geq 1$ 和 $t_1, t_2, \dots, t_n \in T$, 都有

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_{t_i}(x_i).$$

或者

$$M_{t_1, t_2, \dots, t_n}(\theta_1, \theta_2, \dots, \theta_n) = \prod_{i=1}^n M_{t_i}(\theta_i).$$

或者 $n \geq 1, \forall t_i \in T, X_i$ 相互独立。

下面给出一个矩母函数的例子。

例题 8.1 设独立随机序列 $\{X_n, n \geq 1\}$, 每个 X_i 的矩母函数 $M_{X_i}(t)$ 都存在, 设广义随机游动 $S_n = \sum_{i=1}^n X_i$. 求 $\{S_n, n \geq 1\}$ 的 n 维矩母函数。

证明

$$\begin{aligned} M_{S_1, \dots, S_n}(t_1, \dots, t_n) &= \exp\left(\sum_{i=1}^n t_i S_i\right) \\ &= \exp\left((t_1 + \dots + t_n)X_1 + (t_2 + \dots + t_n)X_2 + \dots + t_n X_n\right) \\ &= M_{X_1}(t_1 + \dots + t_n) \cdot M_{X_2}(t_2 + \dots + t_n) \cdots M_{X_n}(t_n). \end{aligned}$$

定理 8.1 (Kolmogorov 一致性定理)

设 $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) : n \geq 1\}$ 是随机变量有限维分布族, 且满足:

- 对称性: 对 $(1, 2, \dots, n)$ 的任意排列 (j_1, j_2, \dots, j_n) , 都有

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_{j_1}, t_{j_2}, \dots, t_{j_n}}(x_{j_1}, x_{j_2}, \dots, x_{j_n});$$

- 相容性: 对任意的 n 和 $t_1, t_2, \dots, t_n, t_{n+1} \in T$, 都有

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \lim_{x_{n+1} \rightarrow \infty} F_{t_1, t_2, \dots, t_n, t_{n+1}}(x_1, x_2, \dots, x_n, x_{n+1}).$$

则存在定义在某概率空间上的随机过程 $\{X_t(\omega), t \in T\}$, 使得对于任意的 n 和 $t_1, t_2, \dots, t_n \in T$, 都有 $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) : t_i \in T, n \geq 1\}$ 是随机过程的有限分布族。



Kolmogorov 说明了一个随机过程和一个有限维分布族一一对应。因此我们可以通过有限维分布族来研究随机过程。然而, 我们不可能把所有有限维分布都研究清楚, 因此我们需要考虑随机过程的其他因素, 比如类比随机变量的数字特征, 我们引入随机过程的数字特征: 均值函数和协方差函数。

定义 8.4 (均值函数和协方差函数)

设随机过程 $\{X_t(\omega), t \in T\}$ 满足 $X_t \in L^1(\Omega, \mathcal{F}, P), \forall t \in T$. 则称

$$m(t) = EX_t, t \in T$$

为随机过程的均值函数; 若进一步有 $X_t \in L^2(\Omega, \mathcal{F}, P)$, 称

$$R(s, t) = cov(X_s, X_t) = E[(X_s - m(s))(X_t - m(t))], s, t \in T$$

为随机过程的协方差函数。特别地, 我们称 $cov(X_t, X_t) = Var(X_t)$ 为随机过程的方差函数, 并且 $E[X_s X_t]$ 称为自相关函数。



 **笔记** 均值函数和协方差函数都是指标 t 的函数，或者说是时间的函数。

例题 8.2 设随机过程 $X(t) = X_0 + tV, t \in [a, b]$, 其中 X_0, V 独立同分布于 $N(0, 1)$. 求均值函数和协方差函数。

证明 由于 X_0, V 独立同分布于 $N(0, 1)$, 因此 $EX_0 = EV = 0, Var(X_0) = Var(V) = 1$. 于是

$$m(t) = E[X_0 + tV] = EX_0 + tEV = 0.$$

又由于 X_0, V 独立, 因此

$$R(s, t) = E[X_s X_t] = E[X_0^2 + tsV^2 + tX_0V + sX_0V] = 1 + ts.$$

8.1.2 基本类型

对于随机过程, 我们对其两个角度感兴趣, 一个是增量有没有趋势可言, 随机变量会不会产生突变? 另一个增量之间是否存在关联? 基于这两个角度, 我们主要关注随机过程的平稳性和独立增量性。

定义 8.5 (严平稳)

我们首先看比较严格的平稳性定义。若对任意的实数 h , 随机过程 $\{X_t, t \in T\}$ 满足有限分布族 $(X(t_1), X(t_2), \dots, X(t_n))$ 和 $(X(t_1 + h), X(t_2 + h), \dots, X(t_n + h))$ 同分布, 则说随机过程 $\{X_t, t \in T\}$ 是严格平稳的。



严平稳过程的有限分布族随时间推移不变。

如上面关于有限维分布和数字特征的讨论, 严平稳的要求实在太强。因此我们又引入宽平稳的概念。

定义 8.6 (宽平稳)

设随机过程 $\{X_t, t \in T\}$ 满足 $X_t \in L^2(\Omega, \mathcal{F}, P), \forall t \in T$. 如果对任意的 $t, s, t+h, s+h \in T$, 都有

$$m(t) = m(t+h), R(t, s) = R(t+h, s+h),$$

则称随机过程 $\{X_t, t \in T\}$ 是宽平稳的。



 **笔记** 宽平稳过程只针对均值函数和协方差函数的平稳性: 均值函数是常数, 不随时间变化; 协方差函数只与时间间隔有关, 而与具体时间无关。因此宽平稳过程的方差函数也是常数。

例题 8.3 白噪声序列 设一列两两互不相关序列 $\{X_t, t \in T\}, EX_t = 0, Var(X_t) = \sigma^2 < \infty$. 且有 $E[X_n X_m] = \sigma^2 \delta_{n,m}$. 我们称 $\{X_t, t \in T\}$ 为白噪声序列, 记作 $WN(0, \sigma^2)$. 证明白噪声序列宽平稳。

证明 由于 $EX_t = 0$, 因此均值函数 $m(t) = 0$. 又由于

$$R(t, s) = E[X_t X_s] = \sigma^2 \delta_{t,s},$$

也就是说协方差函数仅于时间差有关。因此白噪声序列宽平稳。

时间序列分析里常用 ARMA 模型。

例题 8.4ARMA

1. 线性序列: 设 $\{\epsilon(n), n \in \mathbb{Z}\}$ 为 $WN(0, \sigma^2)$, 数列 $\{a_n\}$ 满足 $\sum_{n=-\infty}^{\infty} |a_n| < \infty$, 定义

$$X(t) = \sum_{j=-\infty}^{\infty} a_j \epsilon(t-j), t \in \mathbb{Z},$$

则级数 $X(t)$ a.s. 收敛, 且称序列为线性序列。若级数只有有限项, 则称 $\{X(t)\}$ 为滑动平均序列。

2. AR 模型: 设 $\{\epsilon(n), n \in \mathbb{Z}\}$ 为 $WN(0, \sigma^2)$, 实数 a 满足 $|a| < 1$, 定义

$$X(t) = aX(t-1) + \epsilon(t), t \in \mathbb{Z},$$

则 $\{X(t)\}$ 称为一阶自回归序列。

3. MA 模型: 设 $\{\epsilon(n), n \in \mathbb{Z}\}$ 为 $WN(0, \sigma^2)$, 实数 b 满足 $|b| < 1$, 定义

$$X(t) = \mu + \epsilon(t) + b\epsilon(t-1), t \in \mathbb{Z},$$

则 $\{X(t)\}$ 称为一阶滑动平均序列。

现在我们来看随机过程的增量性质。

定义 8.7 (独立增量和平稳增量)

设随机过程 $\{X_t, t \in T\}$ 满足 $X_t \in L^2(\Omega, \mathcal{F}, P), \forall t \in T$. 如果对任意的 $n \geq 1$ 和分割 $t_0 < t_1 < \dots < t_n \in T$, 都有增量 $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ 相互独立, 则称随机过程 $\{X_t, t \in T\}$ 具有独立增量性。

同时若对任意的 $t_1 < t_2 \in T$ 和实数 h , 都有 $X(t_2) - X(t_1)$ 和 $X(t_2 + h) - X(t_1 + h)$ 同分布, 则称随机过程 $\{X_t, t \in T\}$ 是平稳增量过程。同时平稳增量和独立增量的过程称为平稳独立增量过程。



笔记 如果随机过程 $\{X_t, t \in T\}$ 是平稳独立增量过程, 则对于任意的分割 $0 = t_0 \leq t_1 < \dots < t_n, h > 0, (X(t_1) - X(t_0), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1}))$ 和 $(X(t_1 + h) - X(t_0 + h), X(t_2 + h) - X(t_1 + h), \dots, X(t_n + h) - X(t_{n-1} + h))$ 具有相同的分布。

我们给出平稳独立增量过程的一个性质。

命题 8.2

设随机过程 $\{X_t, t \in T\}$ 具有独立增量性, 且 $X(0) = 0$, 则 $\{X_t, t \in T\}$ 具有平稳增量的充要条件是: 其特征函数具有可乘性, 即

$$\psi_{X(t+s)}(u) = \psi_{X(t)}(u)\psi_{X(s)}(u).$$

其中 $\psi_{X(t)}(u) = E[\exp(uX(t))]$.



后面介绍的泊松过程和布朗运动都是平稳独立增量过程。

命题 8.3

设随机过程 $\{X_t, t \in T\}$ 具有独立增量性, 且 $X(0) = 0, m(t) = E[X(t)]$, 则对于任意正整数 k 有

$$m(k) = E[X(k) - X(0)] = \sum_{i=1}^k E[X(i) - X(i-1)] = km(1),$$

对有理数 r 也有 $m(r) = rm(1)$. 如果 $m(t)$ 连续, 则对任意实数 t 都有 $m(t) = tm(1)$.



高斯过程是一种重要的随机过程。

定义 8.8 (高斯过程)

设随机过程 $\{X_t, t \in T\}$ 满足对任意的 $n \geq 1$ 和 $t_1, t_2, \dots, t_n \in T$, 都有 n 元随机变量组 $(X(t_1), X(t_2), \dots, X(t_n))$ 服从 n 维正态分布, 则称随机过程 $\{X_t, t \in T\}$ 为高斯过程。



性质

- 设有随机过程 $\{X_t, t \in T\}$, 其为高斯过程的充要条件是对于任意的 $n \geq 1, t_1, t_2, \dots, t_n \in T$ 和实数 a_1, a_2, \dots, a_n , 随机变量的线性组合

$$\sum_{i=1}^n a_i X(t_i)$$

服从一维正态分布。

- 宽平稳的高斯过程必为严平稳过程。

8.2 计数过程和泊松过程

8.2.1 泊松过程

我们考虑一种普遍的随机现象：我们记 $N(t)$ 为在某一时间段 $[0, t]$ 内某种事件发生的次数。如果 $\{N(t), t \geq 0\}$ 满足以下条件，则称其为计数过程。

1. $N(t)$ 为非负整数值随机变量，且 $N(0) = 0$ ；
2. $N(t)$ 随时间单调不减；
3. $N(t)$ 轨道右连续，且左极限存在。

若 $\{N(t), t \geq 0\}$ 是一个计数过程，那么 $N(t) - N(s)$ 就代表了 $(s, t]$ 这段时间内发生事件的总数。计数过程是一个随机过程的框架，我们向其增加一些定量定性的假设，就可以得到更新过程、泊松过程等重要的随机过程。我们先考虑泊松过程。

定义 8.9 (泊松过程 1)

设 $\{N(t), t \geq 0\}$ 为计数过程，若其满足以下条件，我们称其为泊松过程。

1. 独立增量性；
2. 对任意的 $s, t \geq 0$ ，都有 $N(t) - N(s) \sim P(\lambda(t-s))$ ，其中 $\lambda > 0$ 为常数，称为泊松过程的参数。



笔记

1. 泊松过程的增量服从参数为 $\lambda(t-s)$ 的泊松分布，因此泊松过程具有平稳增量性。
2. 泊松过程的均值函数 $m(t) = E[N(t)] = E[N(t) - N(0)] = \lambda t$. 协方差函数 $R(s, t) = E[N(s)N(t)] - E[N(s)]E[N(t)] = E[N(s)(N(t) - N(s)) + N^2(s)] - E[N(s)]E[N(t)] = \lambda s$. 这里应用了增量的独立性。

泊松过程还有很多种定义方式。

定义 8.10 (泊松过程 2)

设 $\{N(t), t \geq 0\}$ 为计数过程，若其满足以下条件，我们称其为泊松过程。

1. 过程有平稳独立增量。
2. 存在正数 λ ，使得

$$\lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) = 1)}{h} = \lambda, \quad \lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) \geq 2)}{t} = 0.$$



笔记 第二条条件实际上说明在很短的时间间隔内只可能发生一次事件（所谓稀有事件），发生两个或两个以上事件的概率可以忽略不计。下面来验证这两个定义的等价性。先证明 1 可以推出 2.

证明 由泊松过程的定义， $N(t+h) - N(t) \sim P(\lambda h)$ ，因此

$$P(N(t+h) - N(t) = 1) = \lambda h e^{-\lambda h}, \quad P(N(t+h) - N(t) \geq 2) = 1 - e^{-\lambda h} - \lambda h e^{-\lambda h}.$$

于是

$$\lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) = 1)}{h} = \lim_{h \rightarrow 0} \lambda h e^{-\lambda h} = \lambda,$$

$$\lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) \geq 2)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h} - \lambda h e^{-\lambda h}}{h} = 0.$$

然后来验证 2 能推出 1. 我们只要证明 $N(t) - N(s) \sim P(\lambda(t-s))$ 即可. 我们知道 $(t, t+h]$ 这段时间内不发生事件的概率为 $1 - \lambda h + o(h)$ ，发生一次事件的概率为 $\lambda h + o(h)$ ，发生两个或两个以上事件的概率为 $o(h)$. 我们

记 $P_n(t) = P(N(t) = n)$, 于是由独立增量性:

$$\begin{aligned} P_0(t+h) &= P(N(t+h) - N(t) = 0, N(t) = 0) \\ &= P(N(t+h) - N(t) = 0)P(N(t) = 0) \\ &= (1 - \lambda h + o(h))P_0(t). \end{aligned}$$

于是

$$\frac{P_0(t+h) - P_0(t)}{h} = 1 - \lambda + o(1),$$

令 $h \rightarrow 0$,

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t),$$

解这个常微分方程就得到了 $P_0(t) = \exp(-\lambda t)$.

现在我们用数学归纳法得出一般情形。假设

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

对于 $0, 1, 2, \dots, n-1$ 都成立, 我们来验证 n 也成立。由全概率公式和独立增量性

$$P_n(t+h) = \sum_{k=0}^{\infty} P(N(t+h) - N(t) = k, N(t) = n-k) = \sum_{k=0}^{\infty} P(N(t+h) - N(t) = k)P(N(t) = n-k)$$

我们知道在短时间段 $(t, t+h]$ 内发生至少两次事件的概率为 $o(h)$, 因此

$$\sum_{k=2}^{\infty} P(N(t+h) - N(t) = k)P(N(t) = n-k) = o(h).$$

而 $P(N(t+h) - N(t) = 0)P(N(t) = n) = (1 - \lambda h + o(h))P_n(t)$, $P(N(t+h) - N(t) = 1)P(N(t) = n-1) = \lambda h P_{n-1}(t)$. 综合几个式子就得到了

$$\lim_{h \rightarrow 0} \frac{P(t+h) - P(t)}{h} = \lambda(P_{n-1}(t) - P_n(t)),$$

代入假设 $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$, 移项就有了

$$e^{\lambda t}(P_n(t)' + \lambda P_n(t)) \frac{d}{dt}(e^{\lambda t}P_n(t)) = \frac{\lambda^n t^{n-1}}{(n-1)!},$$

结合边界条件 $P_n(0) = P(N(0) = n) = 0$ 就可验证假设在 $k = n$ 时成立。

我们给出一条泊松过程的渐近性质。

性质 设 $\{N(t), t \geq 0\}$ 为参数为 λ 的泊松过程, 则对于任意的 $t > 0$, 都有

$$\frac{N(t)}{t} \xrightarrow{L^2} \lambda.$$

证明 证明很简单: 空间 $L^2(\Omega, \mathcal{F}, P)$ 中的收敛就是均方收敛, 因此式子改写为

$$\lim_{t \rightarrow \infty} E\left[\left(\frac{N(t)}{t} - \lambda\right)^2\right] = 0.$$

通分就得

$$E\left[\left(\frac{N(t)}{t} - \lambda\right)^2\right] = E\left[\left(\frac{N(t)-\lambda t}{t}\right)^2\right] = \frac{1}{t^2} E[(N(t) - \lambda t)] = \frac{1}{t^2} V[N(t)] = \frac{\lambda}{t} \rightarrow 0.$$

实际上这就给出了泊松过程速率 λ 的另一种解释: 在很长一段时间内时间发生数和时间的比值。

若事件 A 的发生服从参数为 λ 的泊松过程, 若每次事件发生时都有概率 p 记录事件 A 的发生, 并以 $M(t), t \geq 0$ 记在时点 t 已经记录的事件数, 则 $M(t), t \geq 0$ 也是一个泊松过程, 且其参数为 λp . 我们称其为原泊松过程的稀释过程。

事实上, 泊松过程的稀释过程可以这样解释其平稳增量性: 事件的记录与否和这件事发生的时点无关, 因此有独立增量性; 又因为事件的发生概率恒为 p , 与时间无关, 因此就有平稳增量性。下面只要证明 $M(t), t \geq 0$

的速率为 λp . 这只要借助全概率公式就能说明:

$$\begin{aligned} P(M(t) = m) &= \sum_{n=0}^{\infty} P(M(t) = m | N(t) = m+n) P(N(t) = m+n) \\ &= \sum_{n=0}^{\infty} \binom{m+n}{m} p^m (1-p)^n \frac{(\lambda t)^{m+n}}{(m+n)!} e^{-\lambda t} \\ &= \frac{(\lambda p t)^m}{m!} e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t(1-p))^n}{n!} \\ &= \frac{(\lambda p t)^m}{m!} e^{-\lambda p t}. \end{aligned}$$

8.2.2 泊松过程衍生分布

我们已经知道在一小段时间内，泊松过程描述的事件最多能发生一次，因此 $N(t)$ 是一个跳跃高度为 1 的阶梯函数。我们用 T_n 表示第 n 次事件发生的时刻，那么第 $n, n-1$ 次事件发生的时间间隔 X_n 为 $X_n = T_n - T_{n-1}$ 。我们知道 $\{N(t), t \geq 0\}, \{T(n), n \in \mathbb{Z}\}$ 分别是连续时间离散状态空间、离散时间连续状态空间的随机过程。这两者一一对应，即：

$$N(t) \geq n \Leftrightarrow T(n) \leq t;$$

$$T(n) = \inf\{t : N(t) \geq n\}, N(t) = \sup\{n : T(n) \leq t\}.$$

下面我们来讨论 $X_n, T(n)$ 的分布。

定理 8.2

$X_n, n \geq 1$ 相互独立，且 $X_n \sim \text{Exp}(\lambda)$.



证明 首先注意到事件 $\{X_1 > t\}$ 等价于事件 $\{N(t) = 0\}$, 于是

$$P(\{X_1 > t\}) = P(\{N(t) = 0\}) = e^{-\lambda t}.$$

也就是说 $X_1 \sim \text{Exp}(\lambda)$.

再来看 X_2 . 注意到事件 $\{X_2 > t | X_1 = s\}$ 等价于事件 $\{N(s+t) - N(s) = 0 | X_1 = s\}$, 因此由独立增量性得出

$$P(\{X_2 > t | X_1 = s\}) = P(\{N(s+t) - N(s) = 0 | X_1 = s\}) = e^{-\lambda t}.$$

于是我们知道 $X_2 \sim \text{Exp}(\lambda)$ 且与 X_1 独立。同理证明其余情况。



笔记 上述结果实际上来源于泊松分布的平稳独立增量性。在事件发生后一切都“重新开始”，与之前发生过的事件无关。

再来看 $T(n)$ 的分布。

定理 8.3

$T(n)$ 服从参数为 n, λ 的 Γ 分布，即

$$f_{T(n)}(t) = \frac{\lambda(\lambda t)^{n-1}}{\Gamma(n)} e^{-\lambda t}, t > 0.$$



证明 只要注意到 $T(n) = \sum_{i=1}^n X_i$, 利用独立同分布的指数分布之和与 Γ 分布的关系即可。

然后我们可以得到泊松分布的第三种定义：

定义 8.11 (泊松过程 3)

设计数过程 $\{N(t), t \geq 0\}$ 满足每次事件发生的间隔独立同分布于 $\text{Exp}(\lambda)$, 则其为参数为 λ 的泊松过程。



关于泊松过程的一个有趣的事实在于一段时间内发生事件的时点和均匀分布有关。下面的证明都应用了泊松过程的独立增量性。

命题 8.4

设有泊松过程 $\{N(t), t \geq 0\}$, 若在时间段 $(0, t]$ 内只发生一次事件, 则发生事件的时间服从分布 $U[0, t]$.

证明 对任意的 $s \leq t$:

$$\begin{aligned} P(T_1 \leq s | N(t) = 1) &= \frac{P(T_1 \leq s, T_2 \geq t)}{P(N(t) = 1)} \\ &= \frac{P(N(s) = 1)P(N(t) - N(s) = 0)}{e^{-\lambda t}\lambda t} \\ &= \frac{e^{-\lambda s}\lambda s \cdot e^{-\lambda(t-s)}}{e^{-\lambda t}\lambda t} \\ &= \frac{s}{t}. \end{aligned}$$

这就是说, 在 $N(t) = 1$ 的条件下, $T_1 \sim U[0, t]$.

命题 8.5

设有泊松过程 $\{N(t), t \geq 0\}$, 在已知 $N(t) = n$ 时, 发生 n 次事件的时刻 T_1, \dots, T_n 的联合密度函数为

$$f(t_1, \dots, t_n) = \frac{n!}{t^n}, t_1 < \dots < t_n.$$

即 n 个独立同分布 $(U[0, t])$ 的随机变量的次序统计量的联合分布。

证明 我们取 $(0, t]$ 的分割 $0 < t_1 < \dots < t_n = t$, 并取小数 h_i 使得 $t_i + h_i < t_{i+1}$. 于是

$$\begin{aligned} P(t_i \leq T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n) &= P(N(t_i + h_i) - N(t_i) = 1, N(t_{i+1}) - N(t_i + h_i) = 0, i = 1, 2, \dots, n | N(t) = n) \\ &= \frac{P(N(t_i + h_i) - N(t_i) = 1)P(N(t_{i+1}) - N(t_i + h_i) = 0)}{P(N(t) = n)} \\ &= \frac{\prod_{i=1}^n e^{-\lambda h_i} e^{-\lambda(t_{i+1} - t_i - h_i)} \lambda h_i}{e^{-\lambda t} (\lambda t)^n / n!} \\ &= \frac{n!}{t^n} \prod_{i=1}^n h_i. \end{aligned}$$

其中 $t_{n+1} = t$. 于是联合密度函数

$$\begin{aligned} f(t_1, \dots, t_n) &= \lim_{h_i \rightarrow 0} \frac{P(t_i \leq T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n)}{\prod_{i=1}^n h_i} \\ &= \frac{n!}{t^n}. \end{aligned}$$

例题 8.5 设乘客到达火车站是一个强度为 λ 的泊松过程, 设火车在 t 时刻出发, 求乘客等候时间总和的均值。

证明 实际上就是求

$$\sum_{k=1}^{N(t)} E[t - T_k].$$

我们先求条件期望。

$$\begin{aligned} \sum_{k=1}^{N(t)} E[t - T_k | N(t) = n] &= \sum_{k=1}^n E[t - T_k] \\ &= nt - \sum_{k=1}^n E[T_k] \\ &= \frac{nt}{2}. \end{aligned}$$

于是由全期望公式,

$$\sum_{k=1}^{N(t)} E[t - T_k] = E\left[\frac{N(t)t}{2}\right] = \frac{\lambda t^2}{2}.$$

8.2.3 泊松过程的推广过程

泊松过程的假设中, 参数 λ 是常数, 和时间无关。我们将泊松过程的参数换为和时间有关的函数 $\lambda(t)$, 就得到了非齐次泊松过程。在实际应用中, 由于设备的老化, 设备出问题这种“稀有事件”的发生频率显然和时间有关, 此时齐次泊松过程已经不适用。另外, 我们还考虑复合泊松过程: 设有一列独立同分布的随机变量 $Y_i, i = 1, 2, \dots, N(t)$, $\{N(t), t \geq 0\}$ 是参数为 λ 的泊松过程。那么称 $X(t) = \sum_{i=1}^{N(t)} Y_i$ 为复合泊松过程。复合泊松过程并不是一个计数过程, 因此它也不是一个泊松过程。但是它在实践中有重要的应用, 比如保险精算中, Y_i 表示第 i 次事故的损失金额, $N(t)$ 表示在时间段 $[0, t]$ 内发生事故的次数, 那么 $X(t)$ 表示在时间段 $[0, t]$ 内的总损失金额。我们类比泊松过程定义 1 非齐次泊松过程也有等价定义。

定义 8.12 (非齐次泊松过程 1)

计数过程 $\{N(t), t \geq 0\}$ 称为强度函数为 $\lambda(t)$ 的非齐次泊松过程, 如果它满足

1. 过程独立增量;
2. 对任意实数 $s, t \geq 0, N(t+s) - N(s) \sim P\left(\int_s^{t+s} \lambda(u)du\right)$.



定义 8.13 (非齐次泊松过程 2)

我们能证明沿用泊松过程的定义 2, 也有非齐次泊松过程的等价定义。计数过程 $\{N(t), t \geq 0\}$ 称为强度函数为 $\lambda(t)$ 的非齐次泊松过程, 如果它满足

1. 过程独立增量;
2. 存在函数 $\lambda(t)$, 使得

$$\lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) = 1)}{h} = \lambda(t), \quad \lim_{h \rightarrow 0} \frac{P(N(t+h) - N(t) \geq 2)}{t} = 0.$$



除了非齐次泊松过程, 我们还看复合泊松过程。复合泊松过程不一定是计数过程, 我们看看保险公司理赔的例子。保险公司理赔的总金额就是事件发生的次数和每次事件损失的金额的加权和, 而这里事件的发生次数 $\{N(t), t \geq 0\}$ 是泊松过程, 而每次赔偿金额独立同分布。满足这样的条件的随机过程就是复合泊松过程。

定义 8.14 (复合泊松过程)

设 $\{N(t), t \geq 0\}$ 为参数为 λ 的泊松过程, 且 $\{Y_i, i = 1, 2, \dots\}$ 为一列独立同分布的随机变量, 且与 $\{N(t), t \geq 0\}$ 独立。定义

$$X(t) = \sum_{i=1}^{N(t)} Y_i.$$

则称 $\{X(t), t \geq 0\}$ 为复合泊松过程。



笔记 复合泊松过程不一定是泊松过程, 因为虽然复合泊松过程是独立增量的, 但不一定是平稳增量的。但是若上式的 Y_i 恒为常数, 则复合泊松过程就是泊松过程。

下面我们来看复合泊松过程的独立增量性和数字特征函数。

定理 8.4

设 $\{X(t) = \sum_{i=1}^{N(t)}, t \geq 0\}$ 为复合泊松过程, 其中 $\{N(t), t \geq 0\}$ 为参数为 λ 的泊松过程, 且 $E|Y_i| <$

$\infty, \text{Var}(Y_i) < \infty$. 则 $\{X(t), t \geq 0\}$ 独立增量，且

$$\begin{aligned} m(t) &= E[X(t)] = \lambda t EY_i, \\ R(s, t) &= \text{cov}(X(s), X(t)) = \lambda \min(s, t) E[Y_i^2]. \end{aligned}$$



证明 独立增量性容易证明：取分割 $T = \{t_0 = 0 < t_1 < \dots < t_n\}$, 则

$$X(t_i) - X(t_{i-1}) = \sum_{j=N(t_{i-1})+1}^{N(t_i)} Y_j.$$

由于 Y_i 独立同分布，而 $\{N(t), t \geq 0\}$ 独立增量，因此 $\{X(t_i) - X(t_{i-1}), i = 1, 2, \dots, n\}$ 相互独立。

然后我们计算均值函数：这里我们使用全期望公式。

$$m(t) = E[X(t)] = E \left[E \left(\sum_{i=1}^n Y_i | N(t) = n \right) \right] = E[N(t) EY_i] = E[N(t)] EY_i = \lambda t EY_i.$$

再来看协方差函数（不妨设 $s < t$ ）：

$$\begin{aligned} R(s, t) &= E[X(s)X(t)] - E[X(s)]E[X(t)] \\ &= E \left[E \left(\sum_{i=1}^{N(s)} Y_i \sum_{j=1}^{N(t)} Y_j | N(s), N(t) \right) \right] - m(s)m(t) \\ &= E \left[E \left(\sum_{i=1}^{N(s)} Y_i^2 | N(s), N(t) \right) + E \left(\sum_{i \neq j}^{N(s), N(t)} Y_i Y_j | N(s), N(t) \right) \right] - m(s)m(t) \\ &= E[N(s)E[Y_i^2] + (N(s)N(t) - N(s))(EY_i)^2] - m(s)m(t) \\ &= \lambda s E[Y_i^2]. \end{aligned}$$

这里我们使用了增量的独立性： $E[N(s)N(t)] = E[N(s)(N(t) - N(s) + N(s))] = E[N(s^2)] + E[(N(s) - N(0))(N(t) - N(s))] = E[N(s^2)] + E[N(s)]E[N(t) - N(s)] = \lambda s + \lambda^2 s^2 + \lambda^2 s(t-s) = \lambda s + \lambda^2 st$.

笔记 这就是说，复合泊松过程的均值函数就是泊松过程的均值函数乘以每次事件的均值，协方差函数也是类似，只不过乘的是每次事件的二阶矩。

8.3 马尔可夫过程

第9章 鞍

第 10 章 布朗运动

第 11 章 ITO 公式和随机微积分