



高等概率论

Advanced Probability Theory

时间: August 16, 2023

目录

第 1 章 测度论	1	2.2 弱大数定律	53
1.1 集合代数	1	2.2.1 L^2 弱定律	53
1.2 测度	3	2.2.2 三角阵列与截断	56
1.2.1 基本概念	3	2.3 Borel-Cantelli 引理	61
1.2.2 Carathéodory 扩张定理	5	2.4 强大数定律	66
1.2.3 Lebesgue 测度的建立	11	2.5 随机级数的收敛	69
1.3 随机变量及其分布	15	第 3 章 中心极限定理	77
1.3.1 随机变量与随机向量	15	3.1 依分布收敛	77
1.3.2 分布函数	18	3.2 紧性	81
1.4 积分	22	3.3 特征函数	83
1.5 Radon-Nikodym 定理	30	3.3.1 基本性质	83
1.6 期望	33	3.3.2 Feller 反演公式	86
1.6.1 期望的计算	33	3.3.3 Lévy-Cramér 连续定理	90
1.6.2 不等式	36	3.4 中心极限定理	92
1.6.3 积分与极限	39	3.4.1 Lindeberg-Feller 定理	92
1.7 乘积测度与 Fubini 定理	40	3.4.2 Berry-Esseen 定理	98
1.8 积分下的微分	44	3.4.3 Poisson 收敛	101
第 2 章 大数定律	47	第 4 章 条件期望与条件概率	103
2.1 独立性	47	4.1 条件期望	103
2.1.1 基本概念	47	4.1.1 存在性与唯一性	103
2.1.2 独立性, 分布与期望	49	4.1.2 示例	104
2.1.3 卷积公式	51	4.1.3 性质	106
2.1.4 独立随机变量的构造	53	4.2 正则条件概率	109

第 1 章 测度论

1.1 集合代数

概率空间 (probability space) 是一个用三元序对 (Ω, \mathcal{F}, P) 表示的测度空间 (measure space), 其中 Ω 为样本空间, \mathcal{F} 为事件空间, $P: \mathcal{F} \rightarrow [0, 1]$ 是将事件指派为概率的函数. 更精确地, 事件集合 \mathcal{F} 是输出空间 Ω 的 σ -代数 (σ -algebra).

定义 1.1

σ -代数 \mathcal{F} 是由 Ω 的子集构成的集族 (collection), 并且满足以下性质

- $\Omega \in \mathcal{F}$.
- 如果 $A \in \mathcal{F}$, 那么 $A^c \in \mathcal{F}$.
- 如果 $A = \bigcup_{i=1}^{\infty} A_i$ 且 $A_i \in \mathcal{F}$, 则 $A \in \mathcal{F}$, 其中 $i = 1, 2, \dots$.

\mathcal{F} 的元素称为可测集 (measurable set).



显然, Ω 的幂集 2^Ω 构成它的一个 σ -代数. 根据 De Morgan 律 $\bigcap_{i \geq 1} A_i = (\bigcup_{i \geq 1} A_i^c)^c$ 可知, \mathcal{F} 在可数交和有限交运算下也是封闭的.

示例 1.1 设集合 $\Omega = \{1, 2, 3\}$, 则 $\mathcal{F}_1 = \{\{1\}, \{2, 3\}, \Omega, \emptyset\}$ 与 $\mathcal{F}_2 = \{\{1, 2\}, \{3\}, \Omega, \emptyset\}$ 均是 Ω 的 σ -代数, 然而 $\mathcal{F}_1 \cup \mathcal{F}_2$ 并不是.

上述的例子表明集合 Ω 的 σ -代数不止一个, 事实上一个集合 Ω 具有大量的 σ -代数, 现用定理表述如下.

定理 1.1

设 \mathcal{A} 是 Ω 的任意子集构成的集族, 则在 Ω 内存在一个最小的 σ -代数 \mathcal{F}^* , 使得 $\mathcal{A} \subset \mathcal{F}^*$. 通常称 \mathcal{F}^* 是由 \mathcal{A} 生成的 σ -代数, 记作 $\sigma(\mathcal{A})$.



证明 令 X 为 Ω 内所有包含 \mathcal{A} 的 σ -代数构成的集族, 显然 Ω 的幂集是具有这种性质的 σ -代数, 故而 X 非空. 设 \mathcal{F}^* 是所有 $\mathcal{F} \in X$ 的交, 显然 $\mathcal{A} \subset \mathcal{F}^*$, 并且 \mathcal{F}^* 包含于每一个包含 \mathcal{A} 的 σ -代数中, 现在需要证明 \mathcal{F}^* 是一个 σ -代数.

根据 σ -代数的定义, 对于每个 $\mathcal{F} \in X$ 都有 $\Omega \in \mathcal{F}$, 故而 $\Omega \in \mathcal{F}^*$.

假设 $A \in \mathcal{F}^*$, 于是对于每个 $\mathcal{F} \in X$ 都有 $A \in \mathcal{F}$, 因为 \mathcal{F} 为 σ -代数, 故而 $A^c \in \mathcal{F}$ 对一切 $\mathcal{F} \in X$ 成立, 从而 $A^c \in \mathcal{F}^*$.

假设当 $i = 1, 2, \dots$ 时有 $A_i \in \mathcal{F}^*$, 同理可证 $A_i \in \mathcal{F}$, 因为 \mathcal{F} 为 σ -代数, 故而 $\bigcup_i A_i \in \mathcal{F}$, 从而 $\bigcup_i A_i \in \mathcal{F}^*$.

设 \mathbb{R}^k 是实向量 (x_1, \dots, x_k) 构成的集合, 根据定理 1.1 可知, 存在一个包含 \mathbb{R}^k 的全体开集的最小 σ -代数 \mathcal{R}^k , 称为 Borel σ -代数, \mathcal{R}^k 的元素称为 Borel 集. 特别地, 闭集也是 Borel 集, 因为它是开集的补集.

定义 1.2

Ω 的子集构成的集族 \mathcal{P} 称为一个 π -系, 如果当 $A, B \in \mathcal{P}$ 时有 $A \cap B \in \mathcal{P}$.



示例 1.2 \mathbb{R} 上所有区间构成的集族 \mathcal{P} 是一个 π -系 (注意空集也是开区间), 而 \mathbb{R}^2 上全体开圆盘构成的集族不是 π -系.

定义 1.3

Ω 的子集构成的集族 \mathcal{L} 称为一个 λ -系, 如果它满足以下条件

- $\Omega \in \mathcal{L}$.
- 如果 $A, B \in \mathcal{L}, A \subset B$, 那么 $B - A \in \mathcal{L}$.
- 如果 $A_n \in \mathcal{L}$ 且 $A_n \uparrow A$, 那么 $A \in \mathcal{L}$.

$\lambda(A)$ 是包含 A 的最小 λ -系.



示例 1.3 集族 \mathcal{F} 是一个 σ -代数, 当且仅当它是 π -系和 λ -系.

证明 必要性: 假设 \mathcal{F} 为 σ -代数, 由于 \mathcal{F} 对交运算封闭, 故而 \mathcal{F} 是一个 π -系. 另一方面, 当 $A \subset B$ 时有 $B - A = B \cap A^c$, 由于 $\Omega \in \mathcal{F}$, 并且 \mathcal{F} 对补运算、交运算和可数并运算封闭, 因此 \mathcal{F} 是一个 λ -系.

充分性: 假设 \mathcal{F} 为 π -系和 λ -系, 容易验证

- $\Omega \in \mathcal{F}$;
- 如果 $A \in \mathcal{F}$, 那么 $A^c = \Omega - A \in \mathcal{F}$

根据引理 1.4 与 λ -系在单调并运算下封闭的性质, 只需证明 \mathcal{F} 是一个代数. 注意到恒等式

$$A \cup B = (A^c \cap B^c)^c$$

因此只需证明 $\lambda(\mathcal{P})$ 在交运算下封闭, 这已经由 \mathcal{F} 是 π -系所保证.

定理 1.2 ($\pi - \lambda$ 定理)

设 \mathcal{P} 是一个 π -系, 那么 $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$.



证明 对于任意集族 \mathcal{P} , 由于 $\sigma(\mathcal{P})$ 是一个包含了 \mathcal{P} 的 λ -系, 故而 $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$, 因此只需证明当 \mathcal{P} 为 π -系时, $\lambda(\mathcal{P})$ 是一个 σ -代数, 根据上述示例, 这等价于证明 $\lambda(\mathcal{P})$ 是一个 π -系, 现在按照以下步骤进行证明.

Step 1: 定义 $\lambda_A(\mathcal{P}) = \{B : A \cap B \in \lambda(\mathcal{P})\}$, 如果 $A \in \lambda(\mathcal{P})$, 那么 $\lambda_A(\mathcal{P})$ 是一个 λ -系.

(1) 由于 $A \cap \Omega = A \in \lambda(\mathcal{P})$, 故而 $\Omega \in \lambda_A(\mathcal{P})$.

(2) 设 $B, C \in \lambda_A(\mathcal{P})$, $B \supset C$, 于是 $A \cap B \supset A \cap C$, 又根据 $A \in \lambda(\mathcal{P})$ 可知 $B, C \in \lambda(\mathcal{P})$, 可以得到

$$(A \cap B) - (A \cap C) = A \cap (B - C) \in \lambda(\mathcal{P})$$

也即 $B - C \in \lambda_A(\mathcal{P})$.

(3) 设 $D_1 \subset D_2 \subset \dots$ 是 $\lambda_A(\mathcal{P})$ 中的单调递增集合序列, $D = \bigcup_{i=1}^{\infty} D_i$, 那么 $A \cap D_1 \subset$

$A \cap D_2 \subset \dots$, 从而

$$A \cap D = \bigcup_{i=1}^{\infty} (A \cap D_i) \in \lambda(\mathcal{P})$$

也即 $D \in \lambda_A(\mathcal{P})$. 因此, 当 $A \in \lambda(\mathcal{P})$ 时, $\lambda_A(\mathcal{P})$ 是一个 λ -系.

Step 2: $A \in \mathcal{P}, B \in \lambda(\mathcal{P}) \Rightarrow A \cap B \in \lambda(\mathcal{P})$.

由 $A \in \mathcal{P} \subset \lambda(\mathcal{P})$ 可知, $\lambda_A(\mathcal{P})$ 是一个 λ -系. 任取集合 $B \in \mathcal{P}$, 因为 \mathcal{P} 是 π -系, 故而 $A \cap B \in \mathcal{P} \subset \lambda(\mathcal{P}) \Rightarrow B \in \lambda_A(\mathcal{P})$, 于是 $\mathcal{P} \subset \lambda_A(\mathcal{P})$, 也即 $\lambda_A(\mathcal{P})$ 是包含了 \mathcal{P} 的一个 λ -系, 根据 $\lambda(\mathcal{P})$ 的最小性可知 $\lambda(\mathcal{P}) \subset \lambda_A(\mathcal{P})$. 根据 $\lambda_A(\mathcal{P})$ 的定义, 对于任意 $A \in \mathcal{P}, B \in \lambda(\mathcal{P})$, 都有 $A \cap B \in \lambda(\mathcal{P})$.

Step 3: $\lambda(\mathcal{P})$ 是一个 π -系.

固定 $B \in \lambda(\mathcal{P})$, 定义 $\lambda_B(\mathcal{P}) = \{A : A \cap B \in \lambda(\mathcal{P})\}$, 根据之前的结论可知 $\lambda_B(\mathcal{P})$ 是一个 λ -系, 同理也有 $\lambda(\mathcal{P}) \subset \lambda_B(\mathcal{P})$, 从而对于一切 $A \in \lambda(\mathcal{P})$ 都有 $A \cap B \in \lambda(\mathcal{P})$. 综上所述, 对于一切 $A, B \in \lambda(\mathcal{P})$, $A \cap B \in \lambda(\mathcal{P})$ 成立.

推论 1.1

设 \mathcal{P} 是一个 π -系, \mathcal{L} 是包含 \mathcal{P} 的一个 λ -系, 那么 $\sigma(\mathcal{P}) \subset \mathcal{L}$.



证明 因为 \mathcal{L} 是包含 \mathcal{P} 的 λ -系, 故而 $\lambda(\mathcal{P}) \subset \mathcal{L}$, 根据 π - λ 定理可知 $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$, 推论得证.

1.2 测度

1.2.1 基本概念

满足一定条件的集函数 $P : \mathcal{F} \rightarrow [0, 1]$ 称为概率测度 (probability measure), 如果在概率空间中去掉 P , 则二元序对 (Ω, \mathcal{F}) 称为可测空间 (measurable space), 意指可以在该空间上添加测度 (measure).

定义 1.4

测度 (measure) 是一个定义在 σ -代数 \mathcal{F} 上的集函数 μ , 并且满足以下性质

- $\mu(\emptyset) = 0$.
- 对于任意 $A \in \mathcal{F}$, $\mu(A) \in [0, \infty]$.
- 如果 $A_1, A_2, \dots \in \mathcal{F}$ 是可数个互不相交的集合且 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, 那么 $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.



以上性质的第三点叫做测度的可数可加性, 又称 σ -可加性. 如果 $\mu(\Omega) = 1$, 则 μ 为前面提到的概率测度, 通常用 P 表示.

注 为避免麻烦, 假设至少对于某个 $A \in \mathcal{F}$ 有 $\mu(A) < \infty$. 如果对于一切 $A \in \mathcal{F}$ 都有 $\mu(A) < \infty$, 则称 μ 为有限测度 (finite measure).

示例 1.4 设 $\Omega = \mathbb{R}$, \mathcal{F} 是全体 $A \subset \mathbb{R}$ 构成的集族并且使得 A 或 A^c 为可数集, 如果 A 为可数集则 $P(A) = 0$, 否则 $P(A) = 1$, 那么 (Ω, \mathcal{F}, P) 为概率空间.

证明 首先 $\emptyset \in \mathcal{F}$, 以及 \mathcal{F} 在补运算下封闭是显然的. 假设 $A_1, A_2, \dots \in \mathcal{F}$, 如果它们均是可数集, 那么 $\bigcup_{i=1}^{\infty} A_i$ 也可数, 故而 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. 若至少存在某个 A_i 为不可数集, 假设它为 A_k , 由 $A_k \in \mathcal{F}$ 可知 A_k^c 可数, 所以

$$\left(\bigcup_{i=1}^{\infty} A_i \right)^c = \bigcap_{i=1}^{\infty} A_i^c \subset A_k^c \in \mathcal{F}$$

这就表明 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, 因此 \mathcal{F} 是一个 σ -代数.

由于 \emptyset 是有限集, 故而 $P(A) \geq P(\emptyset) = 0$. 假设 $A_1, A_2, \dots \in \mathcal{F}$ 是互不相交的可数集, 如果全体 A_i 均为可数集, 那么

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = 0$$

如果 A_k 是不可数集, 那么

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_k) + \sum_{j \neq k} P(A_j) = 1$$

又因为 $P((-\infty, \infty)) = 1$, 因此 (Ω, \mathcal{F}, P) 为概率空间.

定理 1.3

设 μ 是可测空间 (Ω, \mathcal{F}) 上的测度, 以下性质成立:

- (1) 有限可加性: 如果 $A_1, \dots, A_n \in \mathcal{F}$ 两两不相交, 那么 $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$;
- (2) 单调性: 如果 $A \subset B$, 那么 $\mu(A) \leq \mu(B)$;
- (3) σ -次可加性: 如果 $A \subset \bigcup_{m=1}^{\infty} A_m$, 那么 $\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$;
- (4) 自下连续: 如果 $A_i \uparrow A$, 那么 $\mu(A_i) \uparrow \mu(A)$;
- (5) 自上连续: 如果 $A_i \downarrow A$ 且 $\mu(A_1) < \infty$, 那么 $\mu(A_i) \downarrow \mu(A)$.



证明 (1) 令 $A_{n+1} = A_{n+2} = \dots = \emptyset$, 根据测度的可数次可加性即可得到.

(2) 由于 $A \subset B$, 故而有 $B = A \cup (B - A)$ 以及 $A \cap (B - A) = \emptyset$, 根据有限可加性推得 $\mu(B) = \mu(B - A) + \mu(A) \geq \mu(A)$.

(3) 令 $A'_n = A_n \cap A$, $B_1 = A'_1$, 并且对任意 $n > 1$ 令 $B_n = A'_n - \bigcup_{m=1}^{n-1} A'_m$, 易知各 B_n 互不相交, 又因为 $A \subset \bigcup_{m=1}^{\infty} A_m$, 故而有 $\bigcup_n B_n = A$. 根据测度的定义, $B_m \subset A_m$, 以及测度的单调性可知

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

(4) 令 $B_1 = A_1$, 并且对任意 $n > 1$, 令 $B_n = A_n - A_{n-1}$, 由 $A_i \uparrow A$ 可知各 B_n 均不相交, 并且 $\bigcup_{m=1}^{\infty} B_m = A$ 以及 $\bigcup_{m=1}^n B_m = A_n$, 因此

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

也即 $\mu(A_n) \uparrow \mu(A)$.

(5) 由 $A_i \downarrow A$ 可知 $A_1 - A_n \uparrow A_1 - A$, 根据测度的自下连续可知

$$\mu(A_1 - A_n) \uparrow \mu(A_1 - A)$$

因为 $A_1 \supset A$, 所以

$$\mu(A_1) - \mu(A) = \mu(A_1 - A) = \lim_{n \rightarrow \infty} \mu(A_1 - A_n) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n)$$

也即 $\mu(A_n) \downarrow \mu(A)$.

注 $A_i \uparrow A$ 表示 $A_1 \subset A_2 \subset \cdots$ 且 $\bigcup A_i = A$, 而 $A_i \downarrow A$ 表示 $A_1 \supset A_2 \supset \cdots$ 且 $\bigcap A_i = A$.

1.2.2 Carathéodory 扩张定理

直接在 σ -代数 \mathcal{F} 上构造一个测度是很困难的, 因为 \mathcal{F} 中包含了太多元素. 一个直观的想法是在相对有限的集合代数 \mathcal{A} 中定义集函数 μ , 然后将 μ 扩张为 $\sigma(\mathcal{A})$ 上的测度 μ^* , 使得 μ^* 在 \mathcal{A} 上有 $\mu^* = \mu$, 称 μ^* 是 μ 的扩张 (extension).

定义 1.5

Ω 的子集构成的集族 $\mathcal{S} \subset 2^\Omega$ 称为一个半代数 (semi-algebra), 如果它满足以下性质

- $\Omega \in \mathcal{S}$.
- 如果 $S, T \in \mathcal{S}$, 那么 $S \cap T \in \mathcal{S}$.
- 如果 $S \in \mathcal{S}$, 则 S^c 是 \mathcal{S} 中有限个互不相交集的并集.



示例 1.5 $\Omega = \mathbb{R}$, $\mathcal{S} = \{(a, b], (b, \infty) : -\infty \leq a < b < \infty\}$ 是半代数.

定义 1.6

Ω 的子集构成的集族 $\mathcal{A} \subset 2^\Omega$ 称为一个代数 (algebra), 如果它满足以下性质

- $\Omega \in \mathcal{A}$.
- 如果 $A \in \mathcal{A}$, 那么 $A^c \in \mathcal{A}$.
- 如果 $A, B \in \mathcal{A}$, 那么 $A \cup B \in \mathcal{A}$.



根据 De Morgan 律, 代数对有限交运算也是封闭的. 容易验证, σ -代数必然是代数, 代数必然是半代数, 但是反过来不成立.

示例 1.6 $\Omega = \{a, b, c\}$, $\mathcal{S} = \{\Omega, \emptyset, \{a\}, \{b\}, \{c\}\}$ 是半代数但不是代数.

示例 1.7 令 $\Omega = \mathbb{Z}$, 集族 \mathcal{A} 是由 $A \subset \mathbb{Z}$ 构成的, 其中 A 或 A^c 是有限集. 容易验证 \mathcal{A} 是一个代数. 对于 $n = 1, 2, \dots$, 定义集合 $A_n = \{2n - 1\} \in \mathcal{A}$, 于是

$$A = \bigcup_{n=1}^{\infty} A_n = \{1, 3, 5, \dots\}$$

由于 A 和 A^c 均是无限集, 因此 $A \notin \mathcal{A}$, 也即 \mathcal{A} 不是 σ -代数.

引理 1.1

如果 \mathcal{S} 是一个半代数, 定义 $\bar{\mathcal{S}}$ 为 \mathcal{S} 中有限个互不相交集合并构成的集族, 那么

- (1) $\bar{\mathcal{S}}$ 是由 \mathcal{S} 生成的代数;
- (2) $\sigma(\mathcal{S}) = \sigma(\bar{\mathcal{S}})$.



证明 (1) 设 $S_i, T_i \in \mathcal{S}$, 其中下标 i 对应于某个有限指标集, 再定义 $A = \biguplus_i S_i, B = \biguplus_i T_i$, 其中 \biguplus 表示互不相交集合并.

显然有 $A \cap B = \biguplus_{i,j} S_i \cap T_j \in \bar{\mathcal{S}}$, 也即 $\bar{\mathcal{S}}$ 对有限交运算封闭. 根据半代数的定义可知 $S_i^c \in \bar{\mathcal{S}}$, 又因为 $A^c = \bigcap_i S_i^c$, 加上 $\bar{\mathcal{S}}$ 对有限交运算封闭, 故而 $A^c \in \bar{\mathcal{S}}$. 最后由于 $\Omega \in \mathcal{S}$, 故而 $\emptyset \in \bar{\mathcal{S}}$, 根据 $\bar{\mathcal{S}}$ 对补运算封闭可知 $\Omega \in \bar{\mathcal{S}}$, 根据定义即可证明集族 $\bar{\mathcal{S}}$ 是一个代数.

(2) 由于 $\mathcal{S} \subset \bar{\mathcal{S}}$, 故而 $\sigma(\mathcal{S}) \subset \sigma(\bar{\mathcal{S}})$, 只需证明反向不等式即可. 任取 $A \in \bar{\mathcal{S}}$, 根据定义有 $A = \biguplus_{i=1}^k B_i$, 其中 $B_1, \dots, B_k \in \mathcal{S}$, 从而 $A \in \sigma(\mathcal{S})$, 也即 $\bar{\mathcal{S}} \subset \sigma(\mathcal{S})$. 又因为 $\sigma(\bar{\mathcal{S}})$ 是包含 $\bar{\mathcal{S}}$ 的最小 σ -代数, 因此 $\sigma(\bar{\mathcal{S}}) \subset \sigma(\mathcal{S})$.

定义 1.7

定义在半代数 \mathcal{S} 上的集函数 $\mu: \mathcal{S} \rightarrow [0, \infty]$ 称为 \mathcal{S} 上的加性函数, 如果它满足以下性质

- $\mu(\emptyset) = 0$.
- 如果 $\biguplus_{i=1}^{\infty} S_i \in \mathcal{S}$, 那么 $\mu(\biguplus_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} \mu(S_i)$, 其中 $S_1, S_2, \dots \in \mathcal{S}$.



现在可以将半代数 \mathcal{S} 上的加性函数 μ 扩张为 \mathcal{S} 生成的代数 $\bar{\mathcal{S}}$ 上的预测度 (pre-measure), 它的定义如下.

定义 1.8

定义在代数 \mathcal{A} 上的集函数 $\bar{\mu}: \mathcal{A} \rightarrow [0, \infty]$ 称为 \mathcal{A} 上的预测度, 如果它满足以下性质

- $\bar{\mu}(\emptyset) = 0$.
- 如果 $\biguplus_{i=1}^{\infty} A_i \in \mathcal{A}$, 那么 $\bar{\mu}(\biguplus_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \bar{\mu}(A_i)$, 其中 $A_1, A_2, \dots \in \mathcal{A}$.

**定理 1.4**

设 μ 是定义在半代数 \mathcal{S} 上的加性函数, $\bar{\mathcal{S}}$ 是由 \mathcal{S} 生成的代数, 对于每个 $A = \biguplus_{i=1}^k B_i$, $B_i \in \mathcal{S}$, 定义集函数

$$\bar{\mu}(A) = \sum_{i=1}^k \mu(B_i)$$

那么 $\bar{\mu}$ 是定义在 $\bar{\mathcal{S}}$ 上的预测度.



证明 取 $B_1 = B_2 = \dots = \emptyset$, 因为 $\mu(\emptyset) = 0$, 故而 $\bar{\mu}(\emptyset) = 0$. 由于集函数 μ 的值域为 $[0, \infty]$, 根据定义可知 $\bar{\mu}: \bar{\mathcal{S}} \rightarrow [0, \infty]$. 现在只需证明 $\bar{\mu}$ 满足可数可加性, 也即当 $A_1, A_2, \dots \in \bar{\mathcal{S}}$ 且 $\bigcup_{n \geq 1} A_n \in \bar{\mathcal{S}}$ 时有

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

对于每个 $n \geq 1$, 令 $A_n = \bigcup_{j=1}^{k_n} B_{n_j}$, 其中 $B_{n_j} \in \mathcal{S}$, $j = 1, \dots, k_n$. 假设 $\bigcup_{n \geq 1} A_n \in \bar{\mathcal{S}}$, 故存在有限个互不相交的集合 $B_1, \dots, B_k \in \mathcal{S}$, 使得 $\bigcup_{n \geq 1} A_n = \bigcup_{i=1}^k B_i$, 于是

$$B_i = B_i \cap \left(\bigcup_{n=1}^{\infty} A_n \right) = \bigcup_{n=1}^{\infty} (B_i \cap A_n) = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{k_n} (B_i \cap B_{n_j})$$

因为对于每个指标 i , $B_i \in \mathcal{S}$, 且 $B_i \cap B_{n_j} \in \mathcal{S}$ 对所有 n, j 成立, μ 又是定义在半代数 \mathcal{S} 上的加性函数, 所以

$$\mu(B_i) = \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \mu(B_i \cap B_{n_j})$$

进而

$$\begin{aligned} \bar{\mu} \left(\bigcup_{n=1}^{\infty} A_n \right) &= \sum_{i=1}^k \mu(B_i) = \sum_{i=1}^k \sum_{n=1}^{\infty} \left[\sum_{j=1}^{k_n} \mu(B_i \cap B_{n_j}) \right] \\ &= \sum_{n=1}^{\infty} \left[\sum_{i=1}^k \sum_{j=1}^{k_n} \mu(B_i \cap B_{n_j}) \right] = \sum_{n=1}^{\infty} \bar{\mu}(A_n) \end{aligned}$$

其中最后一个等号用到了以下事实

$$A_n = A_n \cap \left(\bigcup_{i=1}^k B_i \right) = \bigcup_{i=1}^k \bigcup_{j=1}^{k_n} (B_i \cap B_{n_j})$$

由此证得 $\bar{\mu}$ 是 $\bar{\mathcal{S}}$ 上的一个预测度.

现在已经将半代数 \mathcal{S} 上的加性函数扩张为了代数 $\bar{\mathcal{S}}$ 上的预测度, 为了进一步将其扩张到 σ -代数上, 还需要一些新的定义和定理.

定义 1.9

定义在 Ω 上的集函数 $\mu^* : 2^\Omega \rightarrow [0, \infty]$ 称为 Ω 上的外测度 (outer measure), 如果它满足以下性质

- $\mu^*(\emptyset) = 0$.
- $A \subset B \Rightarrow \mu^*(A) \leq \mu^*(B)$.
- 对任意的 $A_1, A_2, \dots \in 2^\Omega$, 都有 $\mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.



定理 1.5

设集族 $\mathcal{E} \subset 2^\Omega$, 并且 $\emptyset \in \mathcal{E}$, $\Omega \in \mathcal{E}$, 集函数 $\mu : \mathcal{E} \rightarrow [0, \infty]$ 使得 $\mu(\emptyset) = 0$. 对于任意 $A \subset \Omega$, 定义

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

那么 μ^* 是一个外测度.



证明 对于任意 $A \subset \Omega$, 总存在集族 $\{E_j\}_{j=1}^{\infty} \subset \mathcal{E}$ 使得 $A \subset \bigcup_{j=1}^{\infty} E_j$, 所以 μ^* 是定义良好的. 现在来验证 μ^* 满足外测度的三个条件.

首先对于一切 $j = 1, 2, \dots$, 取 $E_j = \emptyset$, 由于 $\mu(\emptyset) = 0$, 因此 $\mu^*(\emptyset) = 0$.

假设 $A \subset B$, 由于每个可以盖住 B 的集族 $\{E_j\}_{j=1}^\infty$ 也能盖住 A , 因此 $\mu^*(A) \leq \mu^*(B)$.

若对于某个 $j \geq 1$ 有 $\mu^*(A_j) = \infty$, 那么可数次可加性显然成立. 若对于一切 $j \geq 1$ 都有 $\mu^*(A_j) < \infty$, 此时任取 $\epsilon > 0$ 以及集族 $\{A_j\}_{j=1}^\infty \subset 2^\Omega$. 根据 μ^* 的定义, 对于每个 $j = 1, 2, \dots$, 存在集族 $\{E_{j,k}\}_{k=1}^\infty \subset \mathcal{E}$ 使得 $A_j \subset \bigcup_{k=1}^\infty E_{j,k}$, 以及

$$\sum_{k=1}^\infty \mu(E_{j,k}) \leq \mu^*(A_j) + \epsilon 2^{-j}$$

现在令 $A = \bigcup_{j=1}^\infty A_j$, 那么有 $A \subset \bigcup_{j,k=1}^\infty E_{j,k}$, 以及


$$\sum_{j=1}^\infty \sum_{k=1}^\infty \mu(E_{j,k}) \leq \sum_{j=1}^\infty \mu^*(A_j) + \epsilon$$

根据定义可得 $\mu^*(A) \leq \sum_{j=1}^\infty \mu^*(A_j) + \epsilon$, 由于 $\epsilon > 0$ 是任意的, 所以可数次可加性成立.

定义 1.10

设 μ^* 是 Ω 上的外测度, 如果对于任意 $E \subset \Omega$ 都有


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

那么集合 A 是 μ^* -可测的, 上式称为 Carathéodory 条件. 

定义 1.11

测度空间 $(\Omega, \mathcal{F}, \mu)$ 是完备的 (complete), 如果当 $A \in \mathcal{F}$, $\mu(A) = 0$ 且 $B \subset A$ 时有 $B \in \mathcal{F}$. 

引理 1.2


设集族 $\mathcal{F} \subset 2^\Omega$, \mathcal{F} 是 Ω 上的 σ -代数, 当且仅当 \mathcal{F} 是一个代数, 并且对于一切 $n = 1, 2, \dots$, 当 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}$ 时有 $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$. 

证明 充分性是显然的, 下面只证必要性. 设集族 $\{B_n\}_{n=1}^\infty \in \mathcal{F}$, 由于 \mathcal{F} 是一个代数, 那么对于一切 $n = 1, 2, \dots$, 可以构造集合 $A_n = \bigcup_{j=1}^n B_j \in \mathcal{F}$, 显然满足 $A_n \subset A_{n+1}$, 根据 $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$ 可知

$$\bigcup_{n=1}^\infty B_n = \bigcup_{n=1}^\infty A_n \in \mathcal{F}$$

也即 \mathcal{F} 是一个 σ -代数.

定理 1.6

设 μ^* 是 Ω 上的外测度, 那么由 μ^* -可测集构成的集族 \mathcal{F} 是一个 σ -代数, 并且 μ^* 在 \mathcal{F} 上是一个完备测度. 

证明 首先根据 μ^* -可测的定义可知 \mathcal{F} 在补运算下封闭. 现在设 $A, B \in \mathcal{F}$ 以及 $E \subset \Omega$, 于是

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) \end{aligned}$$

注意到 $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, 根据次可加性可知

$$\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \geq \mu^*(E \cap (A \cup B))$$

因此

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

又根据次可加性得到反向不等式, 于是 Carathéodory 条件成立, 由此可知 $A \cup B \in \mathcal{F}$, 故而 \mathcal{F} 是一个代数. 为了证明 \mathcal{F} 是一个 σ -代数, 根据引理 1.2, 只需证明 \mathcal{F} 在单调并运算下封闭, 也即当 $A_n \in \mathcal{F}$ 且 $A_n \uparrow A$ 时有 $A \in \mathcal{F}$.

设 $B_1 = A_1$, $B_n = A_n \cap A_{n-1}^c$ 对一切 $n \geq 2$ 成立. 由于 \mathcal{F} 是一个代数, 因而对于一切 $n \geq 1$ 都有 $B_n \in \mathcal{F}$. 注意到 $A_n = \bigcup_{j=1}^n B_j$, $A = \bigcup_{j=1}^{\infty} B_j$, 于是

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A_n) + \mu^*(E \cap A_n^c) \\ &= \mu^*(E \cap A_n \cap B_n) + \mu^*(E \cap A_n \cap B_n^c) + \mu^*(E \cap A_n^c) \\ &= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n-1}) + \mu^*(E \cap A_n^c) \\ &= \sum_{j=1}^n \mu^*(E \cap B_j) + \mu^*(E \cap A_n^c) \\ &\geq \sum_{j=1}^n \mu^*(E \cap B_j) + \mu^*(E \cap A^c) \end{aligned}$$

现在令 $n \rightarrow \infty$, 利用 $A = \bigcup_{j=1}^{\infty} B_j$ 及外测度的可数次可加性得到

$$\begin{aligned} \mu^*(E) &\geq \sum_{j=1}^{\infty} \mu^*(E \cap B_j) + \mu^*(E \cap A^c) \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \end{aligned}$$

由于反向不等式总成立, 故而 $A \in \mathcal{F}$, 也即 \mathcal{F} 是一个 σ -代数.

现在设集族 $\{D_n\}_{n=1}^{\infty} \subset \mathcal{F}$, 并且对一切 $i \neq j$ 都有 $D_i \cap D_j = \emptyset$, 再设 $A_j = \bigcup_{i=1}^j D_i$, 其中 $j = 1, 2, \dots$. 由于 \mathcal{F} 是一个 σ -代数, 故而 $A_j \in \mathcal{F}$, 因此

$$\begin{aligned} \mu^*(A_1) &= \mu^*(A_1 \cap D_1) + \mu^*(A_1 \cap D_1^c) = \mu^*(D_1) + \mu^*(A_2) \\ &= \sum_{i=1}^n \mu^*(D_i) + \mu^*(A_{n+1}) \geq \sum_{i=1}^n \mu^*(D_i) \end{aligned}$$

令 $n \rightarrow \infty$ 得到 $\mu^*(A_1) \geq \sum_{i=1}^{\infty} \mu^*(D_i)$. 由于反向不等式总是成立, 故而

$$\mu^*(A_1) = \mu^*\left(\bigcup_{i=1}^{\infty} D_i\right) = \sum_{i=1}^{\infty} \mu^*(D_i)$$

因此 μ^* 是 σ -代数 \mathcal{F} 上的测度.

根据测度的单调性, 如果 $\mu^*(A) = 0$ 且 $B \subset A$, 那么 $\mu^*(B) = 0$, 故而对于任意 $E \subset \Omega$, 总有 $\mu^*(E \cap B) = 0$. 由于 $\mu^*(E) \geq \mu^*(E \cap B^c)$, 因此 $\mu^*(E) \geq \mu^*(E \cap B^c) + \mu^*(E \cap B)$. 因为反向不等式总成立, 所以 $B \in \mathcal{F}$, 也即测度 μ^* 是完备的.

根据定理1.5, 如果 $\bar{\mu}$ 是代数 \mathcal{A} 上的预测度, 那么它可以诱导出 Ω 上的外测度

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \bar{\mu}(A_j) : A_j \in \mathcal{A}, E \subset \bigcup_{j=1}^{\infty} A_j \right\} \quad (1.1)$$

下面给出关于由 $\bar{\mu}$ 诱导出的外测度 μ^* 的性质.

定理 1.7

设 $\bar{\mu}$ 是代数 \mathcal{A} 上的预测度, μ^* 由式 (1.1) 定义, 那么

- (1) 限制在 \mathcal{A} 上有 $\mu^* = \bar{\mu}$;
- (2) \mathcal{A} 中的每个集合都是 μ^* -可测的.



证明 (1) 假设 $E \subset \mathcal{A}$, $A_j \in \mathcal{A}$ 并且 $E \subset \bigcup_{j=1}^{\infty} A_j$, 再令 $B_1 = A_1$, 而当 $n \geq 2$ 时有 $B_n = E \cap (A_n - \bigcup_{j=1}^{n-1} A_j)$, 于是 B_n 是 \mathcal{A} 中互不相交的元素并且它们的并为 E , 故而

$$\bar{\mu}(E) = \sum_{j=1}^{\infty} \bar{\mu}(B_j) \leq \sum_{j=1}^{\infty} \bar{\mu}(A_j)$$

由此得到 $\bar{\mu}(E) \leq \mu^*(E)$. 现在令 $A_1 = E$ 并且对于 $j \geq 2$ 都有 $A_j = \emptyset$, 故而反向不等式成立.

(2) 任取 $\epsilon > 0$, 如果 $A \in \mathcal{A}$, $E \subset \Omega$, 那么存在集族 $\{B_j\}_{j=1}^{\infty} \subset \mathcal{A}$, 使得 $E \subset \bigcup_{j=1}^{\infty} B_j$ 以及 $\sum_{j=1}^{\infty} \bar{\mu}(B_j) \leq \mu^*(E) + \epsilon$. 进一步有

$$\begin{aligned} \mu^*(E) + \epsilon &\geq \sum_{j=1}^{\infty} \bar{\mu}(B_j \cap A) + \sum_{j=1}^{\infty} \bar{\mu}(B_j \cap A^c) \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \end{aligned}$$

由于 $\epsilon > 0$ 是任意的且反向不等式成立, 因此 A 是 μ^* -可测的.

定义 1.12

定义在代数 \mathcal{A} 上的预测度 μ 是 σ -有限的 (σ -finite), 如果存在可数集族 $A_1, A_2, \dots \in \mathcal{A}$, 使得对于一切 $n \geq 1$ 都有 $\mu(A_n) < \infty$ 并且 $\bigcup_{n=1}^{\infty} A_n = \Omega$.



注 令 $A'_1 = A_1$, 并且当 $n \geq 2$ 时有 $A'_n = A_n \cap (\bigcup_{m=1}^{n-1} A_m^c)$, 那么在上述定义中可以不失一般性地假设 A_n 是互不相交的集合. 此外, 这一概念同样适用于定义在 σ -代数上的测度 μ .

示例 1.8 计数测度 设 (Ω, \mathcal{F}) 为一个可测空间, 对于任意 $A \in \mathcal{F}$, 定义计数测度 μ 为

$$\mu(A) = \begin{cases} |A|, & |A| < \infty \\ \infty, & |A| = \infty \end{cases}$$

其中 $|A|$ 为集合 A 的基数. 容易证明, 当且仅当 Ω 为可数集时, μ 是 σ -有限测度.

定理 1.8 (Carathéodory 扩张定理)

设定义在代数 \mathcal{A} 上的预测度 $\bar{\mu}$ 是 σ -有限的, 那么 $\bar{\mu}$ 在 $\sigma(\mathcal{A})$ 上有唯一的扩张测度.



证明 首先根据式 (1.1) 从预测度 $\bar{\mu}$ 诱导出外测度 μ^* , 根据定理1.7(2) 可知 \mathcal{A} 中的元素都是 μ^* -可测集, 再由定理1.6可知这些 μ^* -可测集构成了一个 σ -代数 \mathcal{F} , 由于 $\sigma(\mathcal{A})$ 是由 \mathcal{A} 生成的,

因此 $\sigma(\mathcal{A}) \subset \mathcal{F}$. 因为 $\mu = \mu^*$ 在 \mathcal{F} 上是一个完备测度, 所以在 $\sigma(\mathcal{A})$ 上也是. 根据定理 1.7(1) 可知, 在 \mathcal{A} 上有 $\mu = \mu^* = \bar{\mu}$.

令 ν 是 $\sigma(\mathcal{A})$ 上的另一个测度, 并且在 \mathcal{A} 上有 $\nu = \bar{\mu}$. 假设 $\nu(\Omega) = \bar{\mu}(\Omega) < \infty$, 定义集族

$$\mathcal{L} = \{A \in \sigma(\mathcal{A}) : \mu(A) = \nu(A)\}$$

首先证明 \mathcal{L} 是一个 λ -系.

- 注意到 $\Omega \in \mathcal{A}$, 于是 $\mu(\Omega) = \bar{\mu}(\Omega) = \nu(\Omega)$, 也即 $\Omega \in \mathcal{L}$.
- 假设 $A, B \in \mathcal{L}$ 且 $A \subset B$, $B - A \in \mathcal{L}$ 是显然的.
- 设 $A_n \in \mathcal{L}$, 并且 $A_n \uparrow A = \bigcup_{n=1}^{\infty} A_n$, 于是

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n) = \lim_{n \rightarrow \infty} \nu(A_n) = \nu(A)$$

由此可知 \mathcal{L} 是一个 λ -系. 根据 $\pi - \lambda$ 定理可知 $\sigma(\mathcal{A}) \subset \mathcal{L}$, 由于在 \mathcal{A} 上有 $\mu = \bar{\mu}$, 因此在 $\sigma(\mathcal{A})$ 上也有 $\mu = \nu$.

现在假设 $\bar{\mu}$ 是 σ -有限的, 也即存在互不相交的集合 $B_1, B_2, \dots \in \mathcal{A}$, 使得对于一切 $n \geq 1$ 都有 $\bar{\mu}(B_n) < \infty$ 并且 $\bigcup_{n=1}^{\infty} B_n = \Omega$. 定义集族

$$\mathcal{L}_n = \{A \in \sigma(\mathcal{A}) : \mu(A \cap B_n) = \nu(A \cap B_n)\}$$

同样可以证明 \mathcal{L}_n 是一个 λ -系, 根据 $\pi - \lambda$ 定理可知 $\sigma(\mathcal{A}) \subset \mathcal{L}_n$, 也即对于一切 $A \in \mathcal{A}$ 都有 $\mu(A \cap B_n) = \nu(A \cap B_n)$. 注意到 $A = \bigcup_{n=1}^{\infty} (A \cap B_n)$, 因此

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap B_n) = \sum_{n=1}^{\infty} \nu(A \cap B_n) = \nu(A)$$

因此 μ 是 $\bar{\mu}$ 唯一的扩张测度.

注 在上述证明中, 预测度 $\bar{\mu}$ 的 σ -有限性只用于证明扩张测度 μ 的唯一性, 而 μ 的存在性不需要利用这一条件.

根据以上论述, 一旦给定了半代数 \mathcal{S} 上的加性函数 μ , 就存在完备测度空间 $(\Omega, \mathcal{F}, \nu)$ 使得 $\mathcal{S} \subset \mathcal{F}$, 并且在 \mathcal{S} 上有 $\mu = \nu$. 因为 $\sigma(\mathcal{S}) \subset \mathcal{F}$, 故而 $(\Omega, \sigma(\mathcal{S}), \mu^*)$ 也是测度空间, 但它可能不是完备的.

还可以证明, 对于一个不完备的测度空间 $(\Omega, \mathcal{F}, \mu)$, 总可以通过向 \mathcal{F} 添加特定的元素, 使其成为完备测度空间.

1.2.3 Lebesgue 测度的建立

现在我们可以构造 \mathbb{R} 上的 Lebesgue 测度. 假设 $F : \mathbb{R} \rightarrow \mathbb{R}$ 是单调不减函数, 对于任意 $x \in \mathbb{R}$, 令

$$F(x+) = \lim_{y \downarrow x} F(y), \quad F(x-) = \lim_{y \uparrow x} F(y) \quad (1.2)$$

以及

$$F(\infty) = \lim_{x \uparrow \infty} F(x), \quad F(-\infty) = \lim_{x \downarrow -\infty} F(x) \quad (1.3)$$

设集族

$$\mathcal{S} = \{(a, b] : -\infty \leq a \leq b < \infty\} \cup \{(a, \infty) : -\infty \leq a < \infty\} \quad (1.4)$$

定义集函数

$$\mu((a, b]) = F(b+) - F(a+) \quad (1.5)$$

$$\mu((a, \infty)) = F(\infty) - F(a+) \quad (1.6)$$

容易验证集族 \mathcal{S} 是 \mathbb{R} 上的半代数, 现在需要证明 μ 是 \mathcal{S} 上的加性函数.

定理 1.9

给定式 (1.4)–(1.6) 定义的半代数 \mathcal{S} 和集函数 μ , 如果 $F : \mathbb{R} \rightarrow \mathbb{R}$ 单调不减且是右连续的, 那么 μ 是 \mathcal{S} 上的加性函数.



证明 假设 $-\infty < a \leq b < \infty$, 由于 F 是右连续的, 根据以上各式可知

$$\mu((-a, b]) = F(b) - F(a)$$

$$\mu((-\infty, b]) = F(b) - F(-\infty)$$

$$\mu((a, \infty)) = F(\infty) - F(a)$$

$$\mu((-\infty, \infty)) = F(\infty) - F(-\infty)$$

$$\mu(\emptyset) = 0$$

因此集函数 $\mu : \mathcal{S} \rightarrow [0, \infty]$ 是定义良好的.

假设 $I = (a, b]$, $I_n = \bigcup_{i=1}^n (a_i, b_i]$, 如果 $I = I_n$, 令 $a = a_1$, $b = b_n$, 并且当 $2 \leq i \leq n$ 时 $a_i = b_{i-1}$, 显然每个 $(a_i, b_i]$ 都是互不相交的集合, 故而

$$\sum_{i=1}^n \mu((a_i, b_i]) = \sum_{i=1}^n [F(b_i) - F(a_i)] = F(b) - F(a) = \mu((a, b]) \quad (1.7)$$

现在需要证明当 $I = I_\infty$ 时有

$$\sum_{i=1}^{\infty} \mu((a_i, b_i]) = \mu((a, b]) \quad (1.8)$$

进一步, 可以将 $I = (a, b]$ 划分为 $I = I_n \cup (I - I_n)$, 根据式 (1.7) 可知

$$\mu((a, b]) = \mu(I_n) + \mu(I - I_n) \geq \mu(I_n) = \sum_{i=1}^n \mu((a_i, b_i])$$

再令 $n \rightarrow \infty$, 于是

$$\mu((a, b]) \geq \sum_{i=1}^{\infty} \mu((a_i, b_i])$$

现在只需证明反向不等式即可.

任意固定 $\epsilon > 0$, 因为函数 F 是右连续的, 所以存在 $\delta > 0$ 使得

$$F(a + \delta) < F(a) + \epsilon$$

并且还存 $\eta_j > 0$ 使得

$$F(b_j + \eta_j) - F(b_j) < \epsilon 2^{-j}$$

又因为 $[a + \delta, b] \subset \bigcup_j (a_j, b_j + \eta_j)$, 根据 Heine-Borel 定理可知 $[a + \delta, b]$ 是一个紧集, 故而在有限指标 $J < \infty$ 使得

$$[a + \delta, b] \subset \bigcup_{j=1}^J (a_j, b_j + \eta_j)$$

通过重新标记下标 j , 可以假设

$$b_j + \eta_j \in (a_{j+1}, b_{j+1} + \eta_{j+1}), \quad j = 1, \dots, J-1$$

因此

$$\begin{aligned} \mu((a, b]) &\leq F(b) - F(a + \delta) + \epsilon \leq F(b_J + \eta_J) - F(a_1) + \epsilon \\ &= F(b_J + \eta_J) - F(a_J) + \sum_{j=1}^{J-1} [F(a_{j+1}) - F(a_j)] + \epsilon \\ &\leq F(b_J + \eta_J) - F(a_J) + \sum_{j=1}^{J-1} [F(b_j + \eta_j) - F(a_j)] + \epsilon \\ &< \sum_{j=1}^J [F(b_j) - F(a_j) + \epsilon 2^{-j}] + \epsilon \\ &< \sum_{j=1}^{\infty} \mu((a_j, b_j]) + 2\epsilon \end{aligned}$$

由于 ϵ 是任意的, 故而在 a, b 有限的情况下已证得式 (1.8).

最后考虑无穷区间的情况, 假设 $a = -\infty$, 对于任意的 $M < \infty$, 开区间族 $(a_j, b_j + \eta_j)$ 是 $[-M, b]$ 的一个开覆盖, 同理可证

$$F(b) - F(-M) \leq \sum_{j=1}^{\infty} \mu((a_j, b_j]) + 2\epsilon$$

令 $\epsilon \rightarrow 0^+$, $M \rightarrow \infty$, 于是定理最终得证.

定理 1.10

假设 $F: \mathbb{R} \rightarrow \mathbb{R}$ 是单调不减的右连续函数, 那么在可测空间 $(\mathbb{R}, \mathcal{R})$ 上, 存在与 F 相联系的唯一测度 μ^* , 使得 $\mu^*((a, b]) = F(b) - F(a)$, 称 μ^* 为 Lebesgue-Stieltjes 测度.



证明 根据定理 1.9 可知, $\mu((a, b]) = F(b) - F(a)$ 是半代数 \mathcal{S} 上的加性函数, 其中 \mathcal{S} 为式 (1.4) 所定义的. 又根据定理 1.4, μ 可以被扩张为 $\bar{\mathcal{S}}$ 上的预测度 $\bar{\mu}$, 因为 $\mathbb{R} = \bigcup_{j=1}^{\infty} (-j, j]$, 并且 $\bar{\mu}((-j, j]) < \infty$ 对任意 $j \geq 1$ 成立, 所以 $\bar{\mu}$ 在 $\bar{\mathcal{S}}$ 上是 σ -有限的. 最后由 Carathéodory 扩张定理, $\bar{\mu}$ 在 σ -代数 $\sigma(\bar{\mathcal{S}})$ 上具有唯一的扩张测度 μ^* , 而 $\mathcal{R} = \sigma(\bar{\mathcal{S}})$, 定理得证.

特别地, 如果函数 $F(x) = x$, 那么称 μ^* 为 $(\mathbb{R}, \mathcal{R})$ 上的 Lebesgue 测度, 通常用 m 表示. $(\mathbb{R}, \mathcal{R})$ 上的 Lebesgue 测度衡量的是线段长度, 对于更高维的 Euclid 空间, Lebesgue 测度衡量的

是面积、体积等. 下面的定理直接给出 $(\mathbb{R}^k, \mathcal{R}^k)$ 上的 Lebesgue 测度.

定理 1.11

有限维 Euclid 空间 $(\mathbb{R}^k, \mathcal{R}^k)$ 上存在唯一的测度 m 使得

$$m((a, b]) = \prod_{i=1}^k (b_i - a_i), \quad \forall a, b \in \mathbb{R}^k, a < b$$

称 m 为 $(\mathbb{R}^k, \mathcal{R}^k)$ 上的 Lebesgue 测度.



根据 Carathéodory 扩张定理, 在构造 Lebesgue 测度 m 的过程中直接产生了完备测度空间 $(\mathbb{R}^k, \mathcal{L}^k, m)$, 其中 \mathcal{L}^k 是 Borel σ -代数 \mathcal{R}^k 的完备化, 称 \mathcal{L}^k 的元素为 Lebesgue 可测集, 并且 Borel 可测集和 Lebesgue 可测集的 Lebesgue 测度相等.

下面直接给出一些关于定义在 $(\mathbb{R}, \mathcal{R})$ 上的 Lebesgue 测度的性质.

- 对于任意 $x \in \mathbb{R}$, $m(\{x\}) = 0$.
- 如果 $E \in \mathcal{R}$ 为可数集, 那么 $m(E) = 0$.
- 对于任意 $E \in \mathcal{R}$, 以及 $c, \lambda \in \mathbb{R}$

$$m(E + c) = m(E), \quad m(\lambda E) = |\lambda| m(E)$$

其中 $E + c = \{x + c : x \in E\}$ 以及 $\lambda E = \{\lambda x : x \in E\}$.

根据以上性质及选择公理 (axiom of choice), 可以证明在 \mathbb{R} 上存在 Lebesgue 不可测集, 下面给出一个例子.

示例 1.9 Lebesgue 不可测集 任意给定实数 $x, y \in \mathbb{R}$, 如果 $x - y \in \mathbb{Q}$, 则称 x 和 y 是等价的, 将这种等价关系记为 $x \sim y$. 对于每个 $x \in \mathbb{R}$, 存在 \mathbb{R} 的子集 $E = \{y \in \mathbb{R} : x \sim y\}$, 将其称为 x 的等价类. 例如, 有理数集 \mathbb{Q} 自身构成任意有理数的等价类, 集合 $\{\sqrt{2} + 0.1, \sqrt{2} - 0.01, \dots\}$ 构成 $\sqrt{2}$ 的等价类.

两个等价类要么是不相交的, 要么是重合的. 现在将 x 限定在闭区间 $[0, 1]$ 上, 那么 $[0, 1]$ 是该区间上全体不相交的等价类的并, 也即

$$[0, 1] = \bigcup_{\alpha} E_{\alpha}$$

根据选择公理, 可以从每个 E_{α} 中选取唯一一个元素构成集合 $V = \{x_{\alpha}\}$, 那么 V 是 Lebesgue 不可测集.

证明 由于有理数是可数的, 设 $\{r_k\}_{k=1}^{\infty}$ 为 $[-1, 1]$ 上全体有理数构成的集合, 定义

$$V_k = V + r_k = \{x_{\alpha} + r_k : x_{\alpha} \in V\}$$

下面按照以下步骤进行证明.

Step 1: 对于每个 $k \geq 1$, V_k 之间两两不相交.

假设存在指标 k 和 k' 使得 $V_k \cap V_{k'} \neq \emptyset$, 那么存在指标 α, β 以及 $r_k \neq r_{k'}$, 使得 $x_{\alpha} + r_k = x_{\beta} + r_{k'}$, 也即

$$x_{\alpha} - x_{\beta} = r_{k'} - r_k$$

由于 $\alpha \neq \beta$ 且 $x_{\alpha} - x_{\beta}$ 为有理数, 故而 $x_{\alpha} \sim x_{\beta}$, 这与 V 的构造产生矛盾.

Step 2: $[0, 1] \subset \bigcup_{k=1}^{\infty} V_k \subset [-1, 2]$.

右边的包含关系是显然的. 现在任取 $x \in [0, 1]$, 必然存在指标 α, k 使得 $x - x_\alpha = r_k$, 也即 $x \in \bigcup_{k=1}^{\infty} V_k$, 因此左边的包含关系也成立.

Step 3: V 是 Lebesgue 不可测集.

假设 V 是 Lebesgue 可测的, 那么对于所有的 $k \geq 1$, 集合 V_k 也是可测的. 由于 $\bigcup_{k=1}^{\infty} V_k$ 是互不相交的 V_k 构成的并, 因此

$$1 \leq \sum_{k=1}^{\infty} m(V_k) \leq 3$$

根据 Lebesgue 测度的平移不变性可知 $m(V_k) = m(V)$ 恒成立, 故而

$$1 \leq \sum_{k=1}^{\infty} m(V) \leq 3$$

如果 $m(V) = 0$, 则和式为 0; 如果 $m(V) > 0$, 则和式为 ∞ , 产生矛盾.

注 在不带有选择公理的 Zermelo-Fraenkel 集合论中, Lebesgue 不可测集的存在性不可证明.

直觉上看, 之所以将概率测度定义在由样本空间 Ω 子集构成的 σ -代数 \mathcal{F} 上, 就是为了排除样本空间中不可测集 (无法定义概率的事件集合) 的存在. 尽管存在大量的不可测集, 但它们的构造十分麻烦, 因此在实践中几乎不会遇到无法定义概率的事件.

示例 1.10 设 $A \subset [0, 1]$ 是不可测的, 设 $B = \{(x, 0) : x \in A\}$, 则 B 在 \mathbb{R}^2 是 Lebesgue 可测的, 因为它的外测度为 0.

1.3 随机变量及其分布

1.3.1 随机变量与随机向量

定义 1.13

$X : \Omega \rightarrow S$ 是从 (Ω, \mathcal{F}) 到 (S, \mathcal{S}) 的可测映射 (measurable mapping), 如果对于任意 $B \in \mathcal{S}$ 都有

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

如果 $(S, \mathcal{S}) = (\mathbb{R}^k, \mathcal{R}^k)$, 则称 X 为随机向量 (random vector). 当 $k = 1$ 时, 则称 X 为随机变量 (random variable).



映射 $X : \Omega \rightarrow S$ 的可测性又称 $(\mathcal{F}, \mathcal{S})$ -可测或 \mathcal{F} -可测, 可以简写为 $X \in \mathcal{F}$. 特别地, 如果 X 是 (Ω, \mathcal{F}) 上的随机变量, 那么可测性等价于对任意 $a \in \mathbb{R}$ 都有

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$$

上式可简记为 $\{X \leq a\} \in \mathcal{F}$. 根据集合运算可知, 可测性也等价于 $\{X < a\} \in \mathcal{F}$, $\{X \geq a\} \in \mathcal{F}$, $\{X > a\} \in \mathcal{F}$.

定理 1.12

如果对于任意的 $A \in \mathcal{A}$ 都有 $\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$, 并且 \mathcal{S} 是由 \mathcal{A} 生成的 σ -代数, 那么 $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ 是可测的.



证明 定义集合 $\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$, 于是

$$\{X \in \bigcup_i B_i\} = \bigcup_i \{X \in B_i\} \quad (1.9)$$

$$\{X \in B^c\} = \{X \in B^c\} \quad (1.10)$$

再定义集族 $\mathcal{B} = \{B : \{X \in B\} \in \mathcal{F}\}$, 容易验证 \mathcal{B} 是一个 σ -代数. 由于 $\mathcal{A} \subset \mathcal{B}$ 且 $\mathcal{S} = \sigma(\mathcal{A})$, 故而 $\mathcal{S} \subset \mathcal{B}$, 因此对于任意 $B \in \mathcal{S}$ 都有 $\{X \in B\} \in \mathcal{F}$.

根据式 (1.9) 和 (1.10) 还可以得到, 如果 \mathcal{S} 是一个 σ -代数, 那么 $\{\{X \in B\} : B \in \mathcal{S}\}$ 仍然是 σ -代数, 它是 Ω 上的由 X 生成的使得 X 可测的最小 σ -代数, 记作

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

定理 1.13

如果 $f : (\mathbb{R}^m, \mathcal{R}^m) \rightarrow (\mathbb{R}^n, \mathcal{R}^n)$ 是连续函数, 那么 f 是可测的.



证明 由于 \mathcal{R}^n 是 \mathbb{R}^n 内所有开集生成的 σ -代数, 根据定理 1.12 可知, 只需证明对于任意开集 $B \in \mathcal{R}^n$ 都有 $f^{-1}(B) \in \mathcal{R}^m$. 由于 f 是 \mathbb{R}^m 上的连续函数, 因此 $f^{-1}(B)$ 是 \mathbb{R}^m 中的开集, 也即 $f^{-1}(B) \in \mathcal{R}^m$.

注 可测函数不一定连续, 例如 Dirichlet 函数

$$D(x) = \begin{cases} 0, & x \in \mathbb{Q} \cap [0, 1] \\ 1, & x \in \bar{\mathbb{Q}} \cap [0, 1] \end{cases}$$

它在 $[0, 1]$ 上处处不连续, 但它是可测的.

定理 1.14

如果 $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ 和 $f : (S, \mathcal{S}) \rightarrow (T, \mathcal{T})$ 是可测映射, 那么 $f(X)$ 是从 (Ω, \mathcal{F}) 到 (T, \mathcal{T}) 的可测映射.



证明 对于任意的 $B \in \mathcal{T}$, 由于 f 是可测的, 故而

$$f^{-1}(B) = \{x \in S : f(x) \in B\} \in \mathcal{S}$$

又因为 X 是可测的

$$\{\omega \in \Omega : f(X(\omega)) \in B\} = \{\omega \in \Omega : X(\omega) \in f^{-1}(B)\} \in \mathcal{F}$$

也即 $f(X)$ 是可测的.

以上定理表明, 如果 X 是一个随机变量, 那么 $cX, X^2, \sin(X)$ 等也是随机变量, 其中 c 为任意实数.

定理 1.15

如果 X_1, \dots, X_n 是随机变量, 并且 $f: (\mathbb{R}^n, \mathcal{R}^n) \rightarrow (\mathbb{R}, \mathcal{R})$ 是可测的, 那么 $f(X_1, \dots, X_n)$ 也是随机变量.



证明 根据定理 1.14, 只需证明 (X_1, \dots, X_n) 是一个随机向量即可.

假设 B_1, \dots, B_n 为 \mathbb{R} 上的 Borel 集, 因为 X_i 为随机变量, 所以 $\{X_i \in B_i\} \in \mathcal{F}$, 其中 $i = 1, \dots, n$. 又因为 \mathcal{F} 是 σ -代数, 故而

$$\{(X_1, \dots, X_n) \in B_1 \times \dots \times B_n\} = \bigcap_{i=1}^n \{X_i \in B_i\} \in \mathcal{F}$$

由于 $B_1 \times \dots \times B_n$ 生成了 \mathcal{R}^n , 根据定理 1.12 可知结论成立.

推论 1.2

如果 X_1, \dots, X_n 为随机变量, 那么 $X_1 + \dots + X_n$ 也为随机变量.



证明 由于 $f(x_1, \dots, x_n) = x_1 + \dots + x_n$ 是连续函数, 根据定理 1.13 可知它是可测的, 最后由 1.15 即可推得结论.

定理 1.16

如果 X_1, X_2, \dots 为随机变量, 那么

$$\inf_n X_n, \quad \sup_n X_n, \quad \liminf_n X_n, \quad \limsup_n X_n$$

也是随机变量.



证明 由于序列的下确界小于 a 等价于存在序列的某个元素小于 a , 故而

$$\{\inf_n X_n < a\} = \bigcup_n \{X_n < a\} \in \mathcal{F}$$

同理可得 $\{\sup_n X_n > a\} = \bigcup_n \{X_n > a\} \in \mathcal{F}$. 又因为

$$\begin{aligned} \limsup_n X_n &= \inf_{n \rightarrow \infty} \left(\sup_{m \geq n} X_m \right) \\ \liminf_n X_n &= \sup_{n \rightarrow \infty} \left(\inf_{m \geq n} X_m \right) \end{aligned}$$

因此结论成立.

注 以上各随机变量 X_1, \dots, X_n 都是定义在同一个可测空间 (Ω, \mathcal{F}) 上的.

假设 $\lim_{n \rightarrow \infty} X_n$ 存在, 根据定理 1.16 可知

$$\Omega_o \equiv \left\{ \omega : \lim_{n \rightarrow \infty} X_n < \infty \right\} = \left\{ \omega : \limsup_{n \rightarrow \infty} X_n - \liminf_{n \rightarrow \infty} X_n = 0 \right\}$$

是 \mathcal{F} 中的可测集. 如果 $P(\Omega_o) = 1$, 则称 X_n 几乎必然收敛 (converge almost surely), 简记为 a.s.

定义 1.14

设 μ 是可测空间 (Ω, \mathcal{F}) 上的测度, 集合 $E \in \mathcal{F}$, 称性质 P 在 E 上几乎处处 (almost everywhere, a.e.) 成立, 如果不满足性质 P 的集合 $N \subset E$ 构成零测集.



示例 1.11 设 $f: [0, 1] \rightarrow \mathbb{R}$ 几乎处处连续, 那么 f 是可测的.

证明 设 $E \subset [0, 1]$ 是 f 的不连续点构成的集合, 于是 $m(E) = 0$, 并且 f 限制在 $E^c = [0, 1] \setminus E$ 是连续的. 设 $U \subset \mathbb{R}$ 是一个开集, 从而

$$f^{-1}(U) = [f^{-1}(U) \cap E] \cup [f^{-1}(U) \cap E^c]$$

并且 $f^{-1}(U) \cap E^c$ 在 E^c 中是开的, 而 $V = f^{-1}(U) \cap E$ 是 E 的零测子集, 于是 $f^{-1}(U)$ 是两个可测集的并, 因此也是可测的.

示例 1.12 设 $f : [0, 1] \rightarrow \mathbb{R}$, 如果对于任意 $c \in \mathbb{R}$, 集合 $B = \{x \in [0, 1] : f(x) = c\}$ 是可测的, 那么 f 不一定可测.

证明 设 $A \subset [0, 1]$ 是不可测集, 定义单射函数

$$f(x) = \begin{cases} x, & x \in A \\ -x, & [0, 1] \setminus A \end{cases}$$

由此可见, 对于任意 $c \in \mathbb{R}$, B 要么为空集要么为单点集, 在这两种情况下 B 都是可测的, 然而 $f^{-1}([0, 1]) = A$ 是不可测的.

1.3.2 分布函数

定理 1.17

如果 X 为概率空间 (Ω, \mathcal{F}, P) 上的随机变量, 那么 X 诱导出一个定义在 $(\mathbb{R}, \mathcal{R})$ 上的概率测度 $\mu_X = P \circ X^{-1}$, 使得对于任意 Borel 集 $B \in \mathcal{R}$ 都有

$$\mu_X(B) = P(X^{-1}(B)) = P(X \in B)$$



证明 对于任意的 $B \in \mathcal{R}$, 定义集函数

$$\mu_X(B) = P(X^{-1}(B)) = P(X \in B)$$

只需证明它是 $(\mathbb{R}, \mathcal{R})$ 上的概率测度. 首先易知 $\mu_X(\emptyset) = 0$, 再设 $B_1, B_2, \dots \in \mathcal{R}$ 是两两不相交的 Borel 集, 于是

$$\begin{aligned} \mu_X\left(\bigcup_{i=1}^{\infty} B_i\right) &= P\left(X \in \bigcup_{i=1}^{\infty} B_i\right) = P\left(\bigcup_{i=1}^{\infty} \{X \in B_i\}\right) \\ &= \sum_{i=1}^{\infty} P(X \in B_i) = \sum_{i=1}^{\infty} \mu_X(B_i) \end{aligned}$$

因为 $P(\Omega) = 1$, 故而 μ_X 也是概率测度.

通常称 μ_X 是 X 的概率分布. 在初等概率论和统计课程中, 一般使用概率质量函数 (probability mass function, p.m.f.) 或概率密度函数 (probability density function, p.d.f.) 来分别定义离散型随机变量和连续型随机变量对应于事件 $\{X \in [a, b]\}$ 发生的概率, μ_X 不仅在一个统一框架下包含了上述的所有情形, 还包含了“混合型”的情形.

定义 1.15

随机变量 X 的累积分布函数 (cumulative distribution function, c.d.f.) 为

$$F(x) = \mu_X((-\infty, x]) = P(X \leq x), \quad x \in \mathbb{R}$$

**定理 1.18**

对于任意随机变量 X 的累积分布函数 F , 以下性质成立.

- (1) F 是单调不减函数;
- (2) F 是右连续的, 也即 $\lim_{y \downarrow x} F(y) = F(x)$;
- (3) 如果 $F(x-) = \lim_{y \uparrow x} F(y)$, 那么 $F(x-) = P(X < x)$;
- (4) $\lim_{x \rightarrow \infty} F(x) = 1, \lim_{x \rightarrow -\infty} f(x) = 0$;
- (5) $P(X = x) = F(x) - F(x-)$.



证明 (1) 设 $x_1 \leq x_2$, 于是 $(-\infty, x_1] \subset (-\infty, x_2]$, 从而

$$F(x_1) = \mu_X((-\infty, x_1]) \leq \mu_X((-\infty, x_2]) = F(x_2)$$

(2) 设 $x_n \downarrow x$, 于是 $(-\infty, x_n] \downarrow (-\infty, x]$, 从而

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} \mu_X((-\infty, x_n]) = \mu_X((-\infty, x]) = F(x)$$

(3) 设 $x_n \uparrow x$, 于是 $(-\infty, x_n] \uparrow (-\infty, x)$, 从而

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} \mu_X((-\infty, x_n]) = \mu_X((-\infty, x)) = P(X < x)$$

(4) 设 $x_n \downarrow -\infty, y_n \uparrow \infty$, 于是 $(-\infty, x_n] \downarrow \emptyset, (-\infty, y_n] \uparrow (-\infty, \infty)$, 根据测度的自上连续和自下连续可得

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} \mu_X((-\infty, x_n]) = \mu_X(\emptyset) = 0$$

$$\lim_{n \rightarrow \infty} F(y_n) = \lim_{n \rightarrow \infty} \mu_X((-\infty, y_n]) = \mu_X((-\infty, \infty)) = 1$$

(5) $P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x-)$.

定理 1.19

如果 F 满足定理 1.18 中的 (1), (2) 和 (4), 那么 F 是某个随机变量的 c.d.f.



证明 令 $\Omega = (0, 1)$, \mathcal{F} 为 Borel σ -代数, P 为 Lebesgue 测度, 再设

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

由于 P 是 Lebesgue 测度, 故而 $P(\{\omega : \omega \leq F(x)\}) = F(x)$, 因此只需证明

$$\{\omega : X(\omega) \leq x\} = \{\omega : \omega \leq F(x)\} \quad (1.11)$$

如果 $\omega \leq F(x)$, 此时 $x \notin \{y : F(y) < \omega\}$, 故而 $X(\omega) \leq x$. 如果 $\omega > F(x)$, 由于 F 是右连续的, 因此存在 $\epsilon > 0$ 使得 $F(x + \epsilon) < \omega$ 以及 $X(\omega) \geq x + \epsilon > x$.

定义 1.16

如果存在可数集 $A \subset \mathbb{R}$ 使得 $P(X \in A) = 1$, 则称 X 是离散型随机变量.

如果对于任意 $x \in \mathbb{R}$ 都有 $P(X = x) = 0$, 则称 X 是连续型随机变量.



示例 1.13 如果 F 是阶梯函数, 那么 X 是离散型随机变量, 但是它的逆命题不成立. 例如

$$X(\omega) = \begin{cases} 0, & \omega \in \mathbb{Q} \\ 1, & \omega \in \mathbb{Q}^c \end{cases}$$

其中 \mathbb{Q} 为有理数集. 由于 $P(X \in \mathbb{Q}^c) = 1$, 因此 X 是离散型随机变量, 但是 X 的 c.d.f. 却不是阶梯函数.

示例 1.14 如果 X 是连续型随机变量, 那么它的 c.d.f. 一定连续, 但它的逆命题也不成立. 例如 X 的累积分布为 Cantor 函数, 它甚至不存在概率分布 μ_X .

当 X 是连续型随机变量时, X 的累积分布函数 F 具有如下形式

$$F(x) = \int_{-\infty}^x f(y) dy$$

这里称 f 为 X 的密度函数 (density function). 这里的积分严格来说是下一节介绍的 Lebesgue 积分, 而非初等概率论中直接使用的 Riemann 积分, 尽管两者在可积的意义下数值相等.

示例 1.15 均匀分布 如果 X 服从 (a, b) 上的均匀分布 (uniform distribution), 也即 $X \sim U(a, b)$, 那么 X 的 c.d.f. 为

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases}$$

示例 1.16 如果累积分布函数 $F(x) = P(X \leq x)$ 是连续的, 那么 $Y = F(X)$ 服从 $U(0, 1)$ 分布.

证明 根据定理 1.14 可知 $Y = F(X)$ 是随机变量, 设它的累积分布函数为 G , 于是

$$G(y) = P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

从而 $Y \sim U(0, 1)$ 成立.

示例 1.17 指数分布 如果 X 服从参数为 λ 的指数分布 (exponential distribution), 也即 $X \sim E(\lambda)$, 那么 X 的 c.d.f. 为

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

示例 1.18 正态分布 如果 X 服从均值为 μ , 方差为 σ^2 的正态分布 (normal distribution), 也即 $X \sim N(\mu, \sigma^2)$, 那么 X 的 p.d.f. 为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}$$

但它的 c.d.f. 不具有初等表达, 只能写为积分的形式. 特别地, 当 $\mu = 0$ 且 $\sigma^2 = 1$ 时, 称 X 服从标准正态分布 (standard normal distribution).

示例 1.19 对数正态分布 如果 X 服从均值为 μ , 方差为 σ^2 的对数正态分布 (logarithmic normal distribution), 也即 $X \sim LN(\mu, \sigma^2)$, 那么 X 的 p.d.f. 为

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], & x \geq 0 \\ 0, & x < 0 \end{cases}$$

可以证明, 如果 $X \sim LN(\mu, \sigma^2)$, 那么 $Y = \ln X \sim N(\mu, \sigma^2)$.

示例 1.20 Poisson 分布 如果 X 服从参数为 λ 的 Poisson 分布, 也即 $X \sim P(\lambda)$, 那么 X 的分布律为

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots$$

示例 1.21 几何分布 如果 X 服从参数为 p 的几何分布 (geometric distribution), $0 < p < 1$, 那么 X 的分布律为

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

最后来看随机向量的情形, 设 $X = (X_1, \dots, X_k)$ 是定义在可测空间 (Ω, \mathcal{F}) 上的随机向量, 现在令

$$\begin{aligned} F(x_1, \dots, x_k) &= P(X_1(\omega) \leq x_1, \dots, X_k(\omega) \leq x_k) \\ &= P((-\infty, x_1] \times \dots \times (-\infty, x_k]) \end{aligned}$$

其中 $x_i \in \mathbb{R}, i = 1, \dots, k$. 此时称 F 是 X 的联合累计分布函数. 如果对于一切 $A \in \mathcal{R}^k$ 都有

$$\mu_X(A) = P(X^{-1}(A))$$

则称 μ_X 为 X 的联合概率分布. 进一步, 如果 F_i 为随机变量 X_i 的 c.d.f., 那么它又被称为边际 (marginal) 累积分布函数, 具体而言就是

$$F_i(x) = \lim_{x_j \rightarrow \infty, j=1, \dots, j-1, j+1, \dots, k} F(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_k)$$

显然, 边际累积分布由联合累积分布决定, 尤其是在独立性的条件下有

$$F(x_1, \dots, x_k) = F_1(x_1) \cdots F_k(x_k) = \prod_{i=1}^k P(X_i \leq x_i)$$

其中 $(x_1, \dots, x_k) \in \mathbb{R}^k$. 此时与 F 相联系着的概率测度 P 即为后面提到的乘积测度 $P_1 \otimes \dots \otimes P_k$.

1.4 积分

定义 1.17

假设 $(\Omega, \bar{\mathbb{R}})$ 为可测空间, $f: \Omega \rightarrow \bar{\mathbb{R}}$ 称为简单函数 (simple function), 如果存在有限个 $c_1, \dots, c_k \in \bar{\mathbb{R}}$, 以及 $A_1, \dots, A_k \in \mathcal{F}$, 使得

$$f(x) = \sum_{i=1}^k c_i 1_{A_i}(x)$$

其中 1 为示性函数 (indicator function).



定理 1.20

设 $f: \Omega \rightarrow \bar{\mathbb{R}}_+$ 为非负可测函数, 那么存在非负可测的简单函数序列 s_1, s_2, \dots , 使得

- (1) $s_n \uparrow f$;
- (2) 对于任意 $x \in \Omega$, $\lim_{k \rightarrow \infty} s_k(x) = f(x)$.



证明 对于一切 $k = 1, 2, \dots, i = 1, 2, \dots, k2^k$, 规定

$$E_{ki} = \left\{ x \in \Omega : \frac{i-1}{2^k} \leq f(x) < \frac{i}{2^k} \right\}, \quad E_k = \{x \in \Omega : f(x) \geq k\}$$

再定义函数

$$s_k(x) = \begin{cases} \frac{i-1}{2^k}, & x \in E_{ki} \\ k, & x \in E_k \end{cases}$$

将其改写为

$$s_k(x) = k 1_{E_k}(x) + \sum_{i=1}^{k2^k} \frac{i-1}{2^k} 1_{E_{ki}}(x), \quad x \in \Omega$$

显然每个 s_k 都是非负可测简单函数, 且满足 $0 \leq s_1 \leq s_2 \leq \dots \leq f$. 对任意 $x \in \Omega$, 如果存在 $0 < M < \infty$, 使得 $f(x) \leq M$, 那么当 $k > M$ 时有

$$0 \leq f(x) - s_k(x) \leq 2^{-k}, \quad x \in \Omega$$

如果 $f(x) = \infty$, 那么取 $s_k(x) = k$, 其中 $k = 1, 2, \dots$. 从而当 $k \rightarrow \infty$ 时, $s_k(x) \rightarrow f(x)$.

定义 1.18

设 $(\Omega, \mathcal{F}, \mu)$ 为测度空间, $s: \Omega \rightarrow \bar{\mathbb{R}}_+$ 为 Ω 上的非负可测简单函数, 也即

$$s = \sum_{i=1}^n c_i 1_{A_i}, \quad A_1, \dots, A_n \in \mathcal{F}$$

其中每个 $c_i > 0$. 如果 $E \in \mathcal{F}$, 定义 s 关于测度 μ 的积分 (integral) 为

$$\int_E s d\mu = \sum_{i=1}^n c_i \mu(A_i \cap E)$$

若对于某个 i 而言, $c_i = 0$ 且 $\mu(A_i \cap E) = \infty$, 则规定 $0 \cdot \infty = 0$.



定义 1.19

若 $f: \Omega \rightarrow \mathbb{R}_+$ 为非负可测函数, $E \in \mathcal{F}$, 则定义 f 关于测度 μ 的积分为

$$\int_E f \, d\mu = \sup \left\{ \int_E s \, d\mu : 0 \leq s \leq f \right\}$$

其中 s 为简单函数, 该积分是 $[0, \infty]$ 的一个数.



假设 $(\Omega, \mathcal{F}, \mu)$ 是一个测度空间, $A, B, E \in \mathcal{F}$, 并且 f, g 均为 \mathcal{F} -可测函数, 可以得到以下性质

- (1) 如果 $0 \leq f \leq g$, 那么 $\int_E f \, d\mu \leq \int_E g \, d\mu$;
- (2) 如果 $A \subset B$ 且 $f \geq 0$, 那么 $\int_A f \, d\mu \leq \int_B f \, d\mu$;
- (3) 如果 $f \geq 0$ 且 $0 \leq c < \infty$, 那么 $\int_E cf \, d\mu = c \int_E f \, d\mu$.
- (4) 如果对于一切 $x \in E$ 都有 $f(x) = 0$, 即使 $\mu(E) = \infty$, 也有 $\int_E f \, d\mu = 0$;
- (5) 如果 $\mu(E) = 0$, 即使对于每个 $x \in E$ 都有 $f(x) = \infty$, 也有 $\int_E f \, d\mu = 0$;
- (6) 如果 $f \geq 0$, 那么 $\int_E f \, d\mu = \int_\Omega 1_E f \, d\mu$.

定理 1.21 (单调收敛定理 (Monotone Convergence Theorem, MCT))

设 f_1, f_2, \dots 和 f 都是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的非负可测函数, $E \in \mathcal{F}$, 并且 $f_n \uparrow f$ 在 E 上几乎处处成立, 那么

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f_n \, d\mu$$



证明 由于 $f_n \uparrow f$, 故而当 $n \rightarrow \infty$ 时, 存在 $\alpha > 0$ 使得

$$\lim_{n \rightarrow \infty} \int_E f_n \, d\mu = \alpha$$

又因为 $\int f_n \, d\mu \leq \int f \, d\mu$, 显然有

$$\alpha \leq \int_E f \, d\mu$$

另一方面, 设 $0 < c < 1$, s 是 E 上满足 $0 \leq s \leq f$ 的可测简单函数, 再令

$$E_n = \{x \in E : f_n(x) \geq cs(x)\} \quad (n = 1, 2, \dots)$$

于是每个 $E_n \in \mathcal{F}$, 并且 $E_1 \subset E_2 \subset \dots$, 又因为 $f_n \uparrow f$, 故而 $E_n \uparrow E$, 因此对于每个 $n \geq 1$ 都有

$$\int_E f_n \, d\mu \geq \int_{E_n} f_n \, d\mu \geq c \int_{E_n} s \, d\mu$$

令 $n \rightarrow \infty$ 得到

$$\alpha \geq \int_E f \, d\mu$$

由于上式对一切 $0 < c < 1$ 成立, 从而 $\alpha \geq \int_E f \, d\mu$, 也即

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f_n \, d\mu$$

证毕.

推论 1.3

设 f_1, f_2, \dots 是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的非负可测函数, $E \in \mathcal{F}$, 并且

$$f(x) = \sum_{n=1}^{\infty} f_n(x), \quad x \in E$$

那么

$$\int_E f \, d\mu = \sum_{n=1}^{\infty} \int_E f_n \, d\mu \quad (1.12)$$



证明 注意到式 (1.12) 中的部分和构成非负可测的递增函数列, 由 MCT 即可证得结论.

定理 1.22 (Fatou 引理)

设 f_1, f_2, \dots 是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的非负可测函数, $E \in \mathcal{F}$, 定义

$$f(x) = \liminf_{n \rightarrow \infty} f_n(x) \quad (x \in E)$$

那么

$$\int_E f \, d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n \, d\mu$$



证明 对一切 $n = 1, 2, \dots$, 以及 $x \in E$, 令

$$g_n(x) = \inf_{i \geq n} f_i(x), \quad i \geq n$$

那么 g_n 是可测函数, 并且 $g_n \uparrow f$, 根据 MCT 可知

$$\lim_{n \rightarrow \infty} \int_E g_n(x) \, d\mu = \int_E f \, d\mu$$

注意到 $g_n \leq f_n$, 于是

$$\begin{aligned} \int_E f \, d\mu &= \int_E \lim_{n \rightarrow \infty} g_n \, d\mu = \lim_{n \rightarrow \infty} \int_E g_n \, d\mu \\ &= \liminf_{n \rightarrow \infty} \int_E g_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n \, d\mu \end{aligned}$$

由此证得定理.

定义 1.20

设 f 是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的 \mathcal{F} -可测实值函数, $E \in \mathcal{F}$, $f^+ = \max\{f, 0\}$, $f^- = -\min\{f, 0\}$, 如果积分

$$\int_E f^+ \, d\mu, \quad \int_E f^- \, d\mu$$

至少有一个是有限的, 那么定义 f 的积分为

$$\int_E f \, d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu \quad (1.13)$$

如果它们都是有限的, 则式 (1.13) 也有限, 称 f 关于测度 μ 在 E 上可积. 式 (1.13) 左端也可写作为 $\int_{\Omega} f(\omega) \mu(d\omega)$, 或 $\int_{\Omega} f(\omega) \, d\mu(\omega)$.



定义 1.21

设 $(\Omega, \mathcal{F}, \mu)$ 是测度空间, $0 < p < \infty$, 定义

$$L^p(\Omega, \mathcal{F}, \mu) = \left\{ f : \int_{\Omega} |f|^p d\mu < \infty \right\}$$

将其简写为 $L^p(\mu)$, 称为 L^p 空间. 显然, f 可积当且仅当 $f \in L^1(\mu)$.



注 这里的 p 可以为 ∞ , 此时 $L^\infty(\mu) = \{f : \mu(\{|f| > K\}) = 0, K \in (0, \infty)\}$.

示例 1.22 Lebesgue 积分的计算 首先定义集合

$$A_{n,k} = \{2^{-n}(k-1) \leq x < 2^{-n}k\}$$

以及简单函数

$$f_n = \sum_{k=1}^{2^n} \left(\frac{k-1}{2^n} \right)^2 1_{A_{n,k}}$$

可以证明 $f_n \leq f_{n+1}$, 并且在 $[0, 1]$ 上有 $f_n(x) \rightarrow f(x) = x^2$. 选取 μ 为 $\mathcal{R}([0, 1])$ 上的 Lebesgue 测度 m , 根据 MCT 可知

$$\begin{aligned} \int_{[0,1]} f(x) m(dx) &= \lim_{n \rightarrow \infty} \int_{[0,1]} f_n(x) m(dx) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \frac{(k-1)^2}{2^{2n}} \left(\frac{k}{2^n} - \frac{k-1}{2^n} \right) \\ &= \lim_{n \rightarrow \infty} \frac{2^n(2^n-1)(2^{n+1}-1)}{2^{3n} \cdot 6} = \frac{1}{3} \end{aligned}$$

根据微积分基本定理, Riemann 积分 $\int_0^1 x^2 dx = 1/3$, 由此可见 $f(x) = x^2$ 在 $[0, 1]$ 上的 Riemann 积分和 Lebesgue 积分相等.

事实上, 如果函数 f 在 $[a, b]$ 上 Riemann 可积, 那么在 $[a, b]$ 上有 $f \in L^1(m)$, 并且

$$\int_a^b f(x) dx = \int_{[a,b]} f(x) m(dx)$$

由此可见 Lebesgue 积分是 Riemann 积分的推广. 不仅如此, $[a, b]$ 上的有界函数 f 是 Riemann 可积的, 当且仅当 f 在 $[a, b]$ 上几乎处处连续.

然而, Lebesgue 积分不能替代广义 Riemann 积分, 下面提供了一个广义 Riemann 可积但 Lebesgue 不可积的例子.

示例 1.23 在广义 Riemann 积分的意义下有

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$$

然而 $\sin x/x$ 在此区间上是 Lebesgue 不可积的.

证明 首先注意到对于一切 $n = 1, 2, \dots$ 有

$$\int_{(n-1)\pi}^{n\pi} \frac{|\sin x|}{x} dx \geq \frac{1}{n\pi} \int_{(n-1)\pi}^{n\pi} |\sin x| dx = \frac{2}{n\pi}$$

于是

$$\int_0^{n\pi} \frac{|\sin x|}{x} dx \geq \frac{2}{\pi} \sum_{k=1}^n \frac{1}{k}$$

从而

$$\int_0^{\infty} \frac{|\sin x|}{x} dx = \infty$$

也即 $f \notin L^1(m)$.

引理 1.3

设 $f = f_1 + f_2$, 并且 $f_1, f_2 \in L^1(\mu)$, 那么 $f \in L^1(\mu)$, 并且

$$\int f d\mu = \int f_1 d\mu + \int f_2 d\mu \quad (1.14)$$



证明 令 $g = f_1 + f_2$, 于是

$$g^+ - g^- = f_1^+ - f_1^- + f_2^+ - f_2^-$$

也即

$$g^+ + f_1^- + f_2^- = f_1^+ + f_2^+ + g^-$$

根据推论 1.3 可知

$$\int g^+ + \int f_1^- + \int f_2^- = \int f_1^+ + \int f_2^+ + \int g^-$$

移项就可得到式 (1.14).

定理 1.23

若 $f, g \in L^1(\mu)$, 那么对任意 $\alpha, \beta \in \mathbb{R}$, 都有 $\alpha f + \beta g \in L^1(\mu)$, 并且

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$$



证明 令 $h = h_1 - h_2$, 其中 h_1 和 h_2 都是 $L^1(\mu)$ 上的非负函数, 于是 $h \in L^1(\mu)$, 并且

$$\int h d\mu = \int h_1 d\mu - \int h_2 d\mu \quad (1.15)$$

将 $h \equiv \alpha f + \beta g$ 改写为

$$\begin{aligned} & (\alpha^+ - \alpha^-)(f^+ - f^-) + (\beta^+ - \beta^-)(g^+ - g^-) \\ &= (\alpha^+ f^+ + \alpha^- f^- + \beta^+ g^+ + \beta^- g^-) \\ & \quad - (\alpha^+ f^- + \alpha^- f^+ + \beta^+ g^- + \beta^- g^+) \\ &= h_1 - h_2 \end{aligned}$$

可知 $h_1, h_2 \in L^1(\mu)$, 以及

$$\int h d\mu = \int h_1 d\mu - \int h_2 d\mu$$

注意到 h_1, h_2 的非负性, 由积分的性质和引理 1.3 可知

$$\int h \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu$$

定理 1.24

- (1) 如果 $f : \Omega \rightarrow [0, \infty]$ 是可测函数, $E \in \mathcal{F}$, $\int_E f \, d\mu = 0$, 那么 $f = 0$ 在 E 上几乎处处成立;
- (2) 如果 $f \in L^1(\mu)$ 且 $\int_E f \, d\mu = 0$ 对于一切 $E \in \mathcal{F}$ 成立, 那么 $f = 0$ 在 Ω 上几乎处处成立.



证明 (1) 定义 $A_n = \{\omega \in E : f(\omega) > 1/n\}$, $n = 1, 2, \dots$, 那么

$$\frac{1}{n} \mu(A_n) \leq \int_{A_n} f \, d\mu \leq \int_E f \, d\mu = 0$$

故而 $\mu(A_n) = 0$, 因为 $\{f > 0\} = \bigcup_n A_n$, 结论成立.

(2) 定义 $E = \{f \geq 0\}$, 那么 $\int_E f \, d\mu = \int_E f^+ \, d\mu$, 由于 $\int_E f^+ \, d\mu = 0$, 故而 (1) 的结论表明 $f^+ = 0$ a.e., 类似可证 $f^- = 0$ a.e.

引理 1.4

如果 $f \in L^1(\mu)$, 那么

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$$



证明 令 $z = \int f \, d\mu$, 存在 $|\alpha| = 1$ 使得 $\alpha z = |z|$, 再令 $u = \alpha f$, 于是 $u \leq |\alpha f| = |f|$, 因此

$$\left| \int f \, d\mu \right| = \alpha \int f \, d\mu = \int \alpha f \, d\mu = \int u \, d\mu \leq \int |f| \, d\mu$$

由此证得结论.

定理 1.25 (控制收敛定理 (Dominated Convergence Theorem, DCT))

设 f_1, f_2, \dots 是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的可测函数, $E \in \mathcal{F}$, 并且 $f_n \rightarrow f$ 在 E 上几乎处处成立. 若存在 $g \in L^1(\mu)$ 使得 $|f_n| \leq g$, 则 $f \in L^1(\mu)$, 并且

$$\lim_{n \rightarrow \infty} \int_E f_n \, d\mu = \int_E f \, d\mu$$



证明 因为 $|f| \leq g$ 且 f 可测, 因此 $f \in L^1(\mu)$. 又因为 $|f_n - f| \leq |f_n| + |f| \leq 2g$, 将 Fatou 引理应用于函数 $2g - |f_n - f|$ 得

$$\begin{aligned} \int_E 2g \, d\mu &\leq \liminf_{n \rightarrow \infty} \int_E (2g - |f_n - f|) \, d\mu \\ &= \int_E 2g \, d\mu + \liminf_{n \rightarrow \infty} \left(- \int_E |f_n - f| \, d\mu \right) \\ &= \int_E 2g \, d\mu - \limsup_{n \rightarrow \infty} \int_E |f_n - f| \, d\mu \end{aligned}$$

因为 $g \in L^1(\mu)$, 因此

$$\limsup_{n \rightarrow \infty} \int_E |f_n - f| d\mu \leq 0$$

注意到如果非负实序列不收敛于 0, 那么它的上极限为正, 从而上式一定为 0, 这就表明

$$\lim_{n \rightarrow \infty} \int_E |f_n - f| d\mu = 0$$

再由引理 1.4 可知

$$\lim_{n \rightarrow \infty} \int_E f_n d\mu = \int_E f d\mu$$

证毕.

在 Riemann 积分的体系下, 积分和极限可交换顺序的一个充分条件是函数列 $\{f_n\}_{n=1}^{\infty}$ 一致收敛于极限函数 f , 这是一个相当严格的条件, DCT 则在相当程度上放开了这一限制, 但 DCT 仍然只是充分条件而非充要条件.

推论 1.4

如果 f_1, f_2, \dots 是可测函数, 并且满足

$$\sum_{n=1}^{\infty} \int_E |f_n| d\mu < \infty$$

那么

$$\int_E \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu$$



证明 设 $g = \sum_{n=1}^{\infty} |f_n|$, 根据 MCT 可知

$$\int_E g d\mu = \sum_{n=1}^{\infty} \int_E |f_n| d\mu < \infty$$

从而 $g \in L^1(\mu)$, 由于 $|f| \leq g$, 再由 DCT 得到

$$\begin{aligned} \int_E \sum_{n=1}^{\infty} f_n d\mu &= \int_E \lim_{m \rightarrow \infty} \sum_{n=1}^m f_n d\mu = \lim_{m \rightarrow \infty} \int_E \sum_{n=1}^m f_n d\mu \\ &= \lim_{m \rightarrow \infty} \sum_{n=1}^m \int_E f_n d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu \end{aligned}$$

证毕.

示例 1.24 Riemann-Lebesgue 引理 如果 g 是可积的, 那么对于任意 $E \in \mathcal{R}$ 都有

$$\lim_{n \rightarrow \infty} \int_E g(x) \cos nx dx = 0$$

证明 根据积分的定义可知, 如果 g 为非负可测函数, 那么必定存在简单函数 $0 \leq \varphi \leq g$, 使得

$$\int_E g dx - \epsilon < \int_E \varphi dx \Leftrightarrow \int_E |g - \varphi| dx < \epsilon$$

当 $g = g^+ - g^-$ 为一般的可测函数时, 同理可证

$$\int_E |g^+ - \varphi^+| dx < \frac{\epsilon}{2}, \quad \int_E |g^- - \varphi^-| dx < \frac{\epsilon}{2}$$

从而

$$\int_E |g - \varphi| dx \leq \int_E |g^+ - \varphi^+| dx + \int_E |g^- - \varphi^-| dx < \epsilon \quad (1.16)$$

现在具体假设 φ 为

$$\varphi = \sum_{i=1}^k c_i 1_{(a_{i-1}, a_i)}$$

其中 $a_0 < a_1 < \dots < a_k$. 进一步有

$$\begin{aligned} \int_E \varphi(x) \cos nx dx &= \sum_{i=1}^k \int_{a_{i-1}}^{a_i} \cos nx dx \\ &= \sum_{i=1}^k n^{-1} c_i [\sin(na_i) - \sin(na_{i-1})] \leq \sum_{i=1}^k 2n^{-1} |c_i| \end{aligned}$$

于是当 $n \rightarrow \infty$ 时有

$$\int_E \varphi(x) \cos nx dx \rightarrow 0 \quad (1.17)$$

与此同时

$$\begin{aligned} \left| \int_E g(x) \cos nx dx \right| &\leq \int_E |g(x) - \varphi(x)| |\cos nx| dx + \left| \int_E \varphi(x) \cos nx dx \right| \\ &\leq \int_E |g(x) - \varphi(x)| dx + \left| \int_E \varphi(x) \cos nx dx \right| \end{aligned}$$

根据式 (1.16) 和 (1.17) 即可证得结论.

示例 1.25 假设 f 是可积的, $E = \biguplus_{m=1}^{\infty} E_m$, 那么

$$\sum_{m=1}^{\infty} \int_{E_m} f d\mu = \int_E f d\mu$$

证明 定义 $f_n = \sum_{m=1}^n f 1_{E_m} = f 1_{\biguplus_{m=1}^n E_m}$, 于是

$$\lim_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} \sum_{m=1}^n f 1_{E_m} = f 1_{\biguplus_{m=1}^{\infty} E_m} = f 1_E$$

也即 $f_n \rightarrow f 1_E$. 注意到 $|f_n| \leq f$, 根据 DCT 可知

$$\begin{aligned} \lim_{n \rightarrow \infty} \int f_n d\mu &= \lim_{n \rightarrow \infty} \int \sum_{m=1}^n f 1_{E_m} d\mu = \lim_{n \rightarrow \infty} \sum_{m=1}^n \int_{E_m} f d\mu \\ &= \sum_{m=1}^{\infty} \int_{E_m} f d\mu = \int f 1_E d\mu = \int_E f d\mu \end{aligned}$$

示例 1.26 极限 $\lim_{n \rightarrow \infty} \int_0^1 \frac{n \sin x}{1 + n^2 \sqrt{x}} dx = 0$.

证明 注意到当 $x \in (0, 1]$ 且 $n \geq 1$ 时有

$$\frac{n \sin x}{1 + n^2 \sqrt{x}} \leq \frac{n}{1 + n^2 \sqrt{x}} \leq \frac{1}{\sqrt{x}}$$

由于函数 $g(x) = 1/\sqrt{x}$ 在 $[0, 1]$ 上是可积的, 根据 DCT 得到

$$\lim_{n \rightarrow \infty} \int_0^1 \frac{n \sin x}{1 + n^2 \sqrt{x}} dx = \int_0^1 \lim_{n \rightarrow \infty} \frac{n \sin x}{1 + n^2 \sqrt{x}} dx = 0$$

1.5 Radon-Nikodym 定理

定义 1.22

设 (Ω, \mathcal{F}) 是一个可测空间, α 称为 (Ω, \mathcal{F}) 上的符号测度 (signed measure), 如果它满足以下条件

- $\alpha(\emptyset) = 0$.
- $\alpha \in (-\infty, \infty]$.
- α 满足可数可加性.



注 在通常的定义中, α 还可以取值为 $[-\infty, \infty)$, 这里为了简便剔除了这种情况. 如果 α 在实数范围内取值, 那么 α 为有限符号测度.

定义 1.23

如果对于每个 $B \subset A$ 都有 $\alpha(B) \geq 0$, 那么称 A 为一个正集 (positive set); 如果对于每个 $B \subset A$ 都有 $\alpha(B) \leq 0$, 那么称 A 为一个负集 (negative set).



引理 1.5

- (1) 每一个正 (负) 集的子集都是正 (负) 的;
- (2) 如果 A_n 都为正 (负) 集, 那么 $A = \bigcup_n A_n$ 也是正 (负) 的.



证明 (1) 根据定义直接可证.

(2) 只证 A_n 为正集的情况, 负集同理可得. 注意到

$$B_n = A_n \cap \left(\bigcap_{m=1}^{n-1} A_m^c \right) \subset A_n$$

是不相交的正集, 并且 $\bigcup_n B_n = \bigcup_n A_n$. 令 $E \subset A$ 为可测集, 以及 $E_n = E \cap B_n \subset B_n$, 于是 $\alpha(E_n) \geq 0$, 注意到 $E = \bigcup_n E_n$, 故而 $\alpha(E) = \sum_n \alpha(E_n) \geq 0$.

引理 1.6

设 E 为可测集并且 $\alpha(E) < 0$, 那么存在一个负集 $F \subset E$, 使得 $\alpha(F) < 0$.



证明 如果 E 是负集, 那么结论是显然的.

如果 E 不是负集, 令 n_1 为最小正整数, 使得存在 $E_1 \subset E$ 且 $\alpha(E_1) \geq 1/n_1$. 再令 $k \geq 2$, 如果 $F_k = E - (E_1 \cup \cdots \cup E_{k-1})$ 是负的, 那么结论成立. 如果它不是, 那么令 $n(k)$ 为最小正整数,

使得存在 $E_k \subset F_k$ 且 $\alpha(E_k) \geq 1/n(k)$. 如果对于任意 $k < \infty$, 这一构造过程都不停止, 令

$$F = \bigcap_k F_k = E - \bigcup_k E_k$$

因为 $0 > \alpha(E) > -\infty$ 并且 $\alpha(E_k) \geq 0$, 注意到每个 E_k 都是互不相交的, 故而

$$\alpha(E) = \alpha(F) + \sum_{k=1}^{\infty} \alpha(E_k)$$

以及 $\alpha(F) \leq \alpha(E) < 0$, 并且和式有限. 根据 F 的构造过程, 此时不可能有使得 $\alpha(G) > 0$ 的子集 $G \subset F$, 因为此时必然存在某个正整数 N 使得 $\alpha(G) > 1/N$, 产生矛盾.


定理 1.26 (Hahn 分解定理)

设 α 为 (Ω, \mathcal{F}) 上的符号测度, 那么存在正集 A 与负集 B , 使得 $\Omega = A \cup B$ 且 $A \cap B = \emptyset$. 


证明 设 $c = \inf\{\alpha(B) : B \text{ 为负集}\}$, B_i 为负集且 $\alpha(B_i) \downarrow c$, 以及 $B = \bigcup_i B_i$. 根据引理 1.5 可知 B 是负的, 根据 c 的定义可知 $\alpha(B) \geq c$. 注意到 $\alpha(B) = \alpha(B_i) + \alpha(B - B_i) \leq \alpha(B_i)$, 令 $i \rightarrow \infty$ 立即可知 $\alpha(B) \leq c$, 由此得到 $\alpha(B) = c$. 令 $A = B^c$, 假设 A 包含了一个集合 E 且 $\alpha(E) < 0$, 根据引理 1.6 可知存在负集 $F \subset E$ 使得 $\alpha(F) < 0$, 此时 $B \cup F$ 为负集, 并且 $\alpha(B \cup F) = \alpha(B) + \alpha(F) < c$, 产生矛盾, 因此 A 为正集.

Hahn 分解并不是唯一的, 令 $\Omega = A_1 \cup B_1 = A_2 \cup B_2$ 为两个 Hahn 分解, 容易证明 $A_2 \cap B_1$ 是正的和负的, 此时称它是零集 (null set), 它的全体子集的测度为 0. 类似地, $A_1 \cap B_2$ 也是零集.

定义 1.24

设 (Ω, \mathcal{F}) 为可测空间, μ 和 ν 为 (Ω, \mathcal{F}) 上的测度, 称 μ 和 ν 是相互奇异的 (mutually singular), 如果存在集合 $A \in \mathcal{F}$ 使得 $\mu(A) = 0$ 且 $\nu(A^c) = 0$, 记作 $\mu \perp \nu$. 

定理 1.27 (Jordan 分解定理)

设 α 为 (Ω, \mathcal{F}) 上的符号测度, 那么存在测度 $\alpha_+ \perp \alpha_-$ 使得 $\alpha = \alpha_+ - \alpha_-$, 并且这样的分解是唯一的. 

证明 设 $\Omega = A \cup B$ 为一个 Hahn 分解, 定义

$$\alpha_+(E) = \alpha(E \cap A), \quad \alpha_-(E) = -\alpha(E \cap B)$$

因为 A 是正的, B 是负的, 所以 α_+ 和 α_- 是测度. 注意到 $\alpha_+(A^c) = \alpha_-(A) = 0$, 于是 α_+ 和 α_- 是相互奇异的.

为了证明分解的唯一性, 假设 $\alpha = \nu_1 - \nu_2$ 并且 D 是一个集合, 使得 $\nu_1(D) = \nu_2(D^c) = 0$. 现在设 $C = D^c$, 于是 $\Omega = C \cup D$ 为一个 Hahn 分解. 从 D 的选择中可知

$$\nu_1(E) = \alpha(C \cap E), \quad \nu_2(E) = -\alpha(D \cap E)$$

注意到 $A \cap D = A \cap C^c$, $B \cap C = A^c \cap C$ 是零集, 于是 $\alpha(E \cap C) = \alpha(E \cap (A \cup C)) = \alpha(E \cap A)$, 因此 $\nu_1 = \alpha_+$.

定义 1.25

测度 μ 关于 ν 是绝对连续的 (absolutely continuous), 如果 $\nu(A) = 0$ 意味着 $\mu(A) = 0$, 记作 $\mu \ll \nu$.



示例 1.27 设 μ, ν, ν_1 和 ν_2 是 σ -代数 \mathcal{F} 上的测度, 容易证明

- (1) $\nu_1 \perp \mu, \nu_2 \perp \mu \Rightarrow \nu_1 + \nu_2 \perp \mu$;
- (2) $\nu_1 \ll \mu, \nu_2 \ll \mu \Rightarrow \nu_1 + \nu_2 \ll \mu$;
- (3) $\nu_1 \ll \mu, \nu_2 \perp \mu \Rightarrow \nu_1 \perp \nu_2$;
- (4) $\lambda \ll \mu, \lambda \perp \mu \Rightarrow \lambda = 0$.

定理 1.28

设 μ, ν 是定义在 (Ω, \mathcal{F}) 上的 σ -有限测度, 那么

- 存在唯一的 (ν_r, ν_s) , 使得 $\nu = \nu_r + \nu_s, \nu_r \ll \mu, \nu_s \perp \mu$.
- 存在几乎处处相等意义下唯一的非负可测函数 h , 使得 $\nu_r(E) = \int_E h d\mu$ 对于一切 $E \in \mathcal{F}$ 成立, $h = d\nu_r/d\mu$ 称为 Radon-Nikodym 导数.



证明 通过将 Ω 分解为 $\Omega = \biguplus_i \Omega_i$, 可以不失一般性地假设 μ 和 ν 为有限测度.

任取 $E \in \mathcal{F}$, 设 \mathcal{G} 为满足 $\int_E g d\mu \leq \nu(E)$ 的非负可测函数 $g: \Omega \rightarrow [0, \infty]$ 构成的集合, 再令 $\kappa = \sup\{\int g d\mu : g \in \mathcal{G}\}$, 使得 $\kappa \leq \nu(\Omega) < \infty$, 首先需要证明 \mathcal{G} 中包含了一个可以取到上确界的元素.

设 $g_n \in \mathcal{G}$, 并且满足 $\kappa - 1/n \leq \int g_n d\mu \leq \kappa$, 再设 $h_1 = g_1$ 且当 $n > 1$ 时 $h_n = h_{n-1} \vee g_n$, 显然 $\{h_n\}$ 是单调不减的函数列. 定义 $A_n = \{\omega : h_{n-1}(\omega) > g_n(\omega)\}, n = 2, 3, \dots$, 如果 $h_{n-1} \in \mathcal{G}$, 那么

$$\begin{aligned} \int_E h_n d\mu &= \int_{E \cap A_n} h_{n-1} d\mu + \int_{E \cap A_n^c} g_n d\mu \\ &\leq \nu(E \cap A_n) + \nu(E \cap A_n^c) = \nu(E) \end{aligned}$$

从而 $h_n \in \mathcal{G}$. 由于 $h_n \uparrow h$, 根据 MCT 可知

$$\kappa \geq \int h d\mu = \lim_{n \rightarrow \infty} \int h_n d\mu \geq \lim_{n \rightarrow \infty} \int g_n d\mu = \kappa$$

现在可以定义 ν_r 为

$$\nu_r(E) = \int_E h d\mu, \quad E \in \mathcal{F}$$

容易证明 ν_r 是一个测度, 并且 $\nu_r \ll \mu$.

定义 $\nu_s = \nu - \nu_r$, 现在需要证明 $\nu_s \perp \mu$. 任取 $\epsilon > 0$, 设 $\Omega = A_\epsilon \cup B_\epsilon$ 为关于测度 $\nu_s - \epsilon\mu$ 的 Hahn 分解, 因为 A_ϵ 关于 $\nu_s - \epsilon\mu$ 是正的, 故而

$$\int_E (h + \epsilon 1_{A_\epsilon}) d\mu = \nu_r(E) + \epsilon\mu(A_\epsilon \cap E) \leq \nu(E)$$

这对于一切 $E \in \mathcal{F}$ 都成立, 于是 $k = h + \epsilon 1_{A_\epsilon} \in \mathcal{G}$, 这意味着 $\mu(A_\epsilon) = 0$, 否则会得到 $\int k d\mu > \kappa$, 产生矛盾. 再令 $A = \bigcup_n A_{1/n}$, 于是 $\mu(A) = 0$. 为了证明 $\nu_s(A^c) = 0$, 注意到如果 $\nu_s(A^c) > 0$,

那么对某个充分小的 ϵ 有 $(v_s - \epsilon\mu)(A^c) > 0$, 这就产生矛盾, 因为 $A^c \subset B_\epsilon$ 是负集.

现在证明 (v_r, v_s) 的唯一性, 假设存在 (v'_r, v'_s) 使得

$$v = v'_r + v'_s, \quad v'_r \ll \mu, \quad v'_s \perp \mu$$

那么

$$v'_r - v_r = v'_s - v_s$$

由此可得 $v'_r - v_r \ll \mu$, $v'_s - v_s \perp \mu$, 因此上式两端为 0. 最后, h 的唯一性可以根据定理 1.24(2) 立即得到.

以上定理的第一部分称为 Lebesgue 分解定理, 第二部分称为 Radon-Nikodym 定理, 二者统一的证明由 von Neumann 给出, 下面是关于 Radon-Nikodym 导数的运算性质

(1) 如果 $v_1, v_2 \ll \mu$, 那么 $d(v_1 + v_2)/d\mu = dv_1/d\mu + dv_2/d\mu$;

(2) 如果 $v \ll \mu$ 且 $f \geq 0$, 那么 $\int f dv = \int f \frac{dv}{d\mu} d\mu$;

(3) 如果 $\pi \ll v \ll \mu$, 那么 $d\pi/d\mu = (d\pi/dv) \cdot (dv/d\mu)$;

(4) 如果 $v \ll \mu$ 且 $\mu \ll v$, 那么 $d\mu/dv = (dv/d\mu)^{-1}$.

Radon-Nikodym 定理在概率论中具有重要作用, 连续型随机变量的概率密度函数是诱导测度相对于 Lebesgue 测度的 Radon-Nikodym 导数, 绝对连续的概率测度恰好是具有概率密度函数的测度, 并且该定理可以用来证明条件期望的存在性.

1.6 期望

1.6.1 期望的计算

定义 1.26

设 X 是概率空间 (Ω, \mathcal{F}, P) 上的随机变量, 定义 X 的期望 (expectation) 为

$$E(X) = \int_{\Omega} X dP$$

如果 $\int_{\Omega} X^+ dP$ 和 $\int_{\Omega} X^- dP$ 至少有一个是有限的.



定理 1.29

设 X 是概率空间 (Ω, \mathcal{F}, P) 上的 n 维随机向量, $h: (\mathbb{R}^n, \mathcal{R}^n) \rightarrow (\mathbb{R}, \mathcal{R})$ 是可测函数, 如果 $h \geq 0$ 或 $E|h(X)| < \infty$, 那么

$$E[h(X)] = \int_{\mathbb{R}} h(x) \mu(dx) \quad (1.18)$$

其中 μ 是 X 的概率分布.



证明 先设 $h = 1_A$, $A \in \mathcal{R}^n$, 于是式 (1.18) 的左端为 $P(X \in A)$, 右端为 $\mu(A)$, 根据 μ 的定义可

知两端相等. 根据线性性质, 可以将 h 拓展为非负可测简单函数, 也即在

$$h = \sum_{i=1}^k c_i 1_{A_i}$$

的情况下式 (1.18) 也成立. 对于任意非负的可测函数 h , 又存在非负可测简单函数 h_1, h_2, \dots 使得 $h_m \uparrow h$, 并且

$$\int_{\Omega} h_m(X) dP = \int_{\mathbb{R}^n} h_m(x) \mu(dx)$$

当 $m \rightarrow \infty$ 时, 通过 MCT 就证明了 $h \geq 0$ 的情况, 最后根据 $f = f^+ - f^-$ 和 $E|h(X)| < \infty$ 即可证得一般情况.

根据上述定理可以很方便地计算随机变量的期望值, 根据 Radon-Nikodym 定理, 如果随机变量 X 的概率分布 μ (关于 Lebesgue 测度) 绝对连续, 那么存在非负可测函数 f , 使得对于任意 $A \in \mathcal{R}$ 都有

$$\mu(A) = \int_A f(x) dx$$

如果此时 $\int |g(x)| \mu(dx) < \infty$, 那么还可以证明

$$\int_{\Omega} g(x) \mu(dx) = \int_{\mathbb{R}} g(x) f(x) dx \quad (1.19)$$

这就得到了最常见的连续型随机变量的期望计算公式.

另一方面, 如果 X 是离散型随机变量, 那么对于每个 $A_i \in \mathcal{F}$ 都有

$$X(\omega) = \sum_{i \geq 1} c_i 1_{A_i}(\omega)$$

其中 $i = 1, 2, \dots$. 于是 X 的期望为

$$E(X) = \int_{\Omega} X dP = \sum_{i \geq 1} c_i P(A_i)$$

其中 $A_i \equiv \{\omega : X(\omega) = c_i\}$.

定义 1.27

设 $k \in \mathbb{Z}$, 则称 $E(X^k)$ 为 X 的 k 阶矩. 一阶矩 $\mu = E(X)$ 就是期望, 如果二阶矩 $E(X^2) < \infty$, 则称 $E[(X - \mu)^2]$ 为 X 的方差 (variance), 记作 $\sigma^2 = \text{Var}(X)$.



定理 1.30

设 X 和 Y 为随机变量, $a, b \in \mathbb{R}$, 那么

- (1) $\text{Var}(X) = E(X^2) - E(X)^2$;
- (2) $\text{Var}(aX + b) = a^2 \text{Var}(X)$.



证明 (1) 根据定义可知

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2 \end{aligned}$$

(2) 由于 $E(aX + b) = aE(X) + b$, 于是

$$\begin{aligned}\text{Var}(X) &= E\{[aX + b - E(aX + b)]^2\} \\ &= a^2 E[(X - \mu)^2] = a^2 \text{Var}(X)\end{aligned}$$

示例 1.28 假设随机变量 X 服从参数为 $\lambda > 0$ 的指数分布, 那么它的 k 阶矩为

$$E(X^k) = \int_0^\infty x^k e^{-x} dx = k!$$

从而 $E(X) = 1$, 并且 $\text{Var}(X) = E(X^2) - E(X)^2 = 1$.

示例 1.29 随机变量 X 服从 Bernoulli 分布, 如果对于某个参数 $p \in [0, 1]$, $P(X = 1) = p$ 且 $P(X = 0) = 1 - p$, 那么

$$E(X) = p \cdot 1 + (1 - p) \cdot 0 = p$$

又因为 $X^2 = X$, 故而 $E(X^2) = E(X) = p$, 从而

$$\text{Var}(X) = E(X^2) - E(X)^2 = p(1 - p)$$

示例 1.30 随机变量 X 服从参数为 $\lambda > 0$ 的 Poisson 分布, 如果对于一切 $k = 0, 1, \dots$ 都有

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

于是

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

当 $k \geq 1$ 时有

$$\begin{aligned}E[X(X-1)\cdots(X-k+1)] &= \sum_{j=k}^{\infty} j(j-1)\cdots(j-k+1) e^{-\lambda} \frac{\lambda^j}{j!} \\ &= \lambda^k \sum_{j=k}^{\infty} e^{-\lambda} \frac{\lambda^{j-k}}{(j-k)!} = \lambda^k\end{aligned}$$

因此

$$\text{Var}(X) = E(X^2) - E(X)^2 = E[X(X-1)] + E(X) - \lambda^2 = \lambda$$

示例 1.31 随机变量 X 服从标准 Cauchy 分布, 也即它的 p.d.f. 为

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

那么 X 的期望不存在, 这是因为

$$E|X| = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \lim_{x \rightarrow \infty} \frac{\ln(1+x^2)}{\pi} = \infty$$

然而可以证明, 如果 X_1, X_2, \dots 为服从 Cauchy 分布的 i.i.d. 随机变量, 那么

$$\frac{X_1 + \cdots + X_n}{n}$$

仍然服从 Cauchy 分布.

1.6.2 不等式

定理 1.31 (Markov 不等式)

设 X 是测度空间 $(\Omega, \mathcal{F}, \mu)$ 上的非负可测函数, 那么对于任意 $t \in (0, \infty)$ 都有

$$\mu(A) \leq \frac{1}{t} \int_{\Omega} X \, d\mu \quad (1.20)$$

其中 $A = \{\omega \in \Omega : X(\omega) \geq t\}$.



证明 显然有 $\int_{\Omega} X \, d\mu \geq \int_A X \, d\mu \geq \int_A t \, d\mu \geq t\mu(A)$.

推论 1.5 (Chebyshev 不等式)

设 X 是概率空间 (Ω, \mathcal{F}, P) 上的随机变量, 并且满足 $E(X^2) < \infty$, $E(X) = \mu$ 及 $\text{Var}(X) = \sigma^2$, 那么对于任意 $\epsilon > 0$ 都有

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$



证明 定义集合 $\{|X| \geq t\} = \{\omega \in \Omega : |X(\omega)| \geq t\}$, 显然对于任意 $r, t > 0$ 都有

$$\{|X| \geq t\} = \{|X|^r \geq t^r\}$$

由 Markov 不等式可知

$$P(|X| \geq t) \leq \frac{E|X|^r}{t^r} \quad (1.21)$$

将上式的 X 用 $X - \mu$ 替代, 取 $r = 2, t = k\sigma$ 即可证得结论.

定理 1.32 (Jensen 不等式)

假设 (Ω, \mathcal{F}, P) 为概率空间, $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{R})$, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ 为凸函数, 那么

$$\varphi[E(X)] \leq E[\varphi(X)]$$

其中 $E|X| < \infty$, $E|\varphi(X)| < \infty$.



证明 由于 φ 是凸函数, 故而对于每一点 $\omega_0 \in \mathbb{R}$, 存在 $a, b \in \mathbb{R}$ 使得 $\varphi(\omega_0) = a\omega_0 + b$, 并且对一切 $\omega \in \mathbb{R}$ 都有 $\varphi(\omega) \geq a\omega + b$, 于是

$$(\varphi \circ X)(\omega) \geq aX(\omega) + b$$

再设 $\omega_0 = E(X)$, 注意到 $\mu(\Omega) = 1$, 因此

$$\begin{aligned} \int_{\Omega} (\varphi \circ X) \, dP &\geq \int_{\Omega} (aX + b) \, dP = a \int_{\Omega} X \, dP + b \int_{\Omega} dP \\ &= a\omega_0 + b = \varphi(\omega_0) = \varphi\left(\int_{\Omega} X \, dP\right) \end{aligned}$$

也即 $\varphi[E(X)] \leq E[\varphi(X)]$.

引理 1.7 (Young 不等式)

设 $p, q \in (1, \infty)$ 并且满足 $1/p + 1/q = 1$, 那么对于任意 $x, y \in [0, \infty)$ 都有

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$



证明 任意固定 $y \in [0, \infty)$, 定义函数

$$f(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy, \quad x \in [0, \infty)$$

显然 f 在 $(0, \infty)$ 二阶可微, 并且

$$f'(x) = x^{p-1} - y$$

$$f''(x) = (p-1)x^{p-2}$$

易知 f 是严格凸的, 它的最小值点为 $x_0 = y^{1/(p-1)}$, 根据假设可得 $q = p/(p-1)$, 从而 $x_0^p = y^q$, 最终得到 f 的最小值为

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

证毕.

定理 1.33 (Hölder 不等式)

设 $(\Omega, \mathcal{F}, \mu)$ 为测度空间, $1 < p < \infty$ 且 $1/p + 1/q = 1$, $X \in L^p(\mu)$ 而 $Y \in L^q(\mu)$, 那么

$$E|XY| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}$$

也即 $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.



证明 令 $A = \|X\|_p$, $B = \|Y\|_q$, 显然 $0 < A, B < \infty$. 再令 $a = |X|/A$, $b = |Y|/B$, 根据 Young 不等式可知

$$ab = \frac{|XY|}{AB} \leq \frac{|X|^p}{pA^p} + \frac{|Y|^q}{qB^q} = \frac{a^p}{p} + \frac{b^q}{q}$$

也即

$$\frac{1}{AB} \int |XY| d\mu \leq \frac{1}{pA^p} \int |X|^p d\mu + \frac{1}{qB^q} \int |Y|^q d\mu$$

将 $A^p = \int |X|^p d\mu$ 和 $B^q = \int |Y|^q d\mu$ 代入得

$$\frac{1}{\|X\|_p \|Y\|_q} \|XY\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1$$

这就证明了结论.

推论 1.6 (Cauchy-Schwarz 不等式)

设随机变量 $X, Y \in L^2(\mu)$, 则 $E|XY| \leq (E|X|^2)^{1/2} (E|Y|^2)^{1/2}$.



定理 1.34 (Minkowski 不等式)

设随机变量 $X, Y \in L^p(\mu)$, 并且 $1 \leq p < \infty$, 那么

$$(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$$

即 $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.



证明 注意到对任意 $a, b \in \mathbb{R}$ 且 $1 \leq p < \infty$ 都有

$$|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p)$$

令 $h_1 = |X + Y|$, $h_2 = |X + Y|^{p-1}$, 于是

$$|X + Y|^p \leq 2^{p-1}(|X|^p + |Y|^p)$$

故而 $h_1 \in L^p(\mu)$, $h_2 \in L^q(\mu)$, 其中 $q = p/(p-1)$. 因为

$$|X + Y|^p = h_1 h_2 \leq |X| h_2 + |Y| h_2$$

由 Hölder 不等式可知

$$\int |X + Y|^p d\mu \leq \|X\|_p \left(\int h_2^q d\mu \right)^{\frac{1}{q}} + \|Y\|_p \left(\int h_2^q d\mu \right)^{\frac{1}{q}}$$

由于 $\int h_2^q d\mu = \int |X + Y|^p d\mu$, 因此结论成立.

示例 1.32 设随机变量 $Y \geq 0$ 并且 $E(Y^2) < \infty$, 那么 $P(Y > 0) \geq [E(Y)]^2/E(Y^2)$.

证明 首先有

$$E|Y1_{(Y>0)}| = E[Y1_{(Y>0)}] = E(Y)$$

注意到 $E|Y|^2 = E(Y^2)$, 以及 $E|1_{(Y>0)}|^2 = E[1_{(Y>0)}] = P(Y > 0)$, 根据 Cauchy-Schwarz 不等式可知

$$(E|Y1_{(Y>0)}|)^2 \leq E|Y|^2 E|1_{(Y>0)}|^2$$

也即 $[E(Y)]^2 \leq E(Y^2)P(Y > 0)$.

示例 1.33 如果 $E|X|^k < \infty$, 那么对于任意 $0 < j < k$ 都有 $E|X|^j < \infty$, 并且

$$E|X|^j \leq (E|X|^k)^{j/k}$$

证明 定义集合 $A = \{\omega : |X(\omega)| > 1\}$, 于是对于一切 $0 < j < k$, 在 A 上都有 $|X|^j < |X|^k$, 因此

$$\begin{aligned} \int |X|^j dP &= \int_A |X|^j dP + \int_{A^c} |X|^j dP \\ &\leq \int_A |X|^k dP + 1 \\ &\leq \int |X|^k dP + 1 < \infty \end{aligned}$$

再设函数 $f(x) = x^{k/j} 1_{(x \geq 0)}$, 显然它是一个凸函数. 最后令 $Y = |X|^j$, 于是 $|E(Y)| = E|X|^j < \infty$, 并且 $|E[f(Y)]| = E|X|^k < \infty$, 根据 Jensen 不等式可知

$$(E|X|^j)^{k/j} = f[E(Y)] \leq E[f(Y)] = E|X|^k$$

结论成立.

1.6.3 积分与极限

现在使用期望的语言将上一节所介绍的积分与极限可交换的经典结论重新描述一遍, 直接给出以下定理: 假设 X_1, X_2, \dots 及 X 为随机变量, 以下结论成立

- 单调收敛定理: 如果 $0 \leq X_n \uparrow X$, 那么 $E(X_n) \uparrow E(X)$.
- Fatou 引理: 如果 $X_n \geq 0$, 那么 $\liminf_{n \rightarrow \infty} E(X_n) \geq E(\liminf_{n \rightarrow \infty} X_n)$.
- 控制收敛定理: 如果 $X_n \xrightarrow{a.s.} X$, 对于一切 $n \geq 1$ 都有 $|X_n| \leq Y$, 并且 $E(Y) < \infty$, 那么 $E(X_n) \rightarrow E(X)$.

下面给出另一个关于期望收敛的定理.

定理 1.35

如果随机变量 $X_n \xrightarrow{a.s.} X$, 实值函数 g, h 连续并且满足

- (1) $g \geq 0$ 且当 $|x| \rightarrow \infty$ 时有 $g(x) \rightarrow \infty$;
- (2) 当 $|x| \rightarrow \infty$ 时有 $|h(x)|/g(x) \rightarrow 0$;
- (3) 对于一切 $n \geq 1$ 都有 $E[g(X_n)] \leq K < \infty$;

那么 $E[h(X_n)] \rightarrow E[h(X)]$.



证明 由于可以在 h 上减去一个常数, 所以可以不失一般性地假设 $h(0) = 0$.

对于在同一个可测空间上的任意随机变量 Y , 定义 $\bar{Y} = Y1_{(|Y| \leq M)}$, 其中 M 为某个充分大的常数, 使得 $P(|X| = M) = 0$ 且当 $|x| \geq M$ 时有 $g(x) > 0$. 定义

$$\epsilon_M = \sup\{|h(x)|/g(x) : |x| \geq M\}$$

于是

$$|E[h(\bar{Y})] - E[h(Y)]| \leq E|h(\bar{Y}) - h(Y)| \leq E[|h(Y)|; |Y| > M] \leq \epsilon_M E[g(Y)] \quad (1.22)$$

上式第二个不等号是由根据以下事实得到

- $h(0) = 0$.
- $E|X - Y| \leq E|Y| + E|Y|$.
- 当 $|Y| \leq M$ 时, 随机变量 $\bar{Y} = Y$.

最后一个不等式是由 ϵ_M 的定义得到的.

在式 (1.22) 中取 $Y = X_n$, 根据条件 (iii) 可知

$$|E[h(X_n)] - E[h(X)]| \leq K\epsilon_M \quad (1.23)$$

因为 $g \geq 0$ 且 g 连续, 由 Fatou 引理可知

$$E[g(X)] \leq \liminf_{n \rightarrow \infty} E[g(X_n)] \leq K$$

在式 (1.22) 中取 $Y = X$, 由上式可知

$$|E[h(\bar{X})] - E[h(X)]| \leq K\epsilon_M \quad (1.24)$$

由于 $h(\bar{X}_n)$ 有界且 h 连续, 根据 DCT 可知

$$E[h(\bar{X}_n)] \rightarrow E[h(\bar{X})] \quad (1.25)$$

注意到以下三角不等式成立

$$\begin{aligned} |E[h(X_n)] - E[h(X)]| &\leq |E[h(X_n)] - E[h(\bar{X}_n)]| \\ &\quad + |E[h(\bar{X}_n)] - E[h(\bar{X})]| + |E[h(\bar{X})] - E[h(X)]| \end{aligned}$$

对上式取极限并利用式 (1.23), (1.24) 及 (1.25), 可以得到

$$\limsup_{n \rightarrow \infty} |E[h(X_n)] - E[h(X)]| \leq 2K\epsilon_M$$

由于 $K < \infty$ 且当 $M \rightarrow \infty$ 时有 $\epsilon_M \rightarrow 0$, 因此结论成立.

示例 1.34 设 $\Omega = (0, 1)$, 在 Ω 上配置 Borel σ -代数和 Lebesgue 测度 m . 令 $\alpha \in (1, 2)$, 易知当 $n \rightarrow \infty$ 时, 随机变量

$$X_n = n^\alpha 1_{(\frac{1}{n+1}, \frac{1}{n})} \xrightarrow{a.s.} 0$$

再设函数 $g(x) = |x|^{\frac{2}{\alpha}}$ 和 $h(x) = x$, 容易验证定理 1.35 中的条件 (1), (2). 对于条件 (3), 注意到

$$E[g(X_n)] = \int_0^1 |X_n|^{\frac{2}{\alpha}} dm = \frac{n}{n+1} < 1$$

因此定理 1.35 保证了 $E(X_n) \rightarrow E(X)$.

假设存在随机变量 Y 使得 $|X_n| \leq Y$, 于是

$$\int_0^1 Y dm \geq \sum_{n=1}^{\infty} \frac{n^{\alpha-1}}{n+1} = \infty$$

因此无法通过 DCT 得出 $E(X_n) \rightarrow E(X)$, 由此可见控制函数不是收敛的必要条件.

1.7 乘积测度与 Fubini 定理

定义 1.28

设 (X, \mathcal{A}) 和 (Y, \mathcal{B}) 为可测空间, 再定义

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

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

其中 Ω 称为集合 X 与 Y 的乘积, \mathcal{S} 中的元素称为可测矩形 (measurable rectangle).



容易验证 \mathcal{S} 是一个半代数, 这是因为

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

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

其中 $A, C \in \mathcal{A}$ 并且 $B, D \in \mathcal{B}$. 由此可以定义 \mathcal{A} 和 \mathcal{B} 的乘积 σ -测度

$$\mathcal{F} = \sigma(\mathcal{S}) = \mathcal{A} \otimes \mathcal{B}$$

它是由 \mathcal{S} 生成的最小 σ -代数, 此时称 $(\Omega, \mathcal{F}) = (X \times Y, \mathcal{A} \otimes \mathcal{B})$ 为乘积可测空间.

定义 1.29

设 E 为 $X \times Y$ 的子集, 令

$$E_x = \{y \in Y : (x, y) \in E\}$$

$$E_y = \{x \in X : (x, y) \in E\}$$

分别称为 E 的 x -截面 (section) 和 y -截面.

**引理 1.8**

如果 $E \in \mathcal{A} \otimes \mathcal{B}$, 则对于任意 $x \in X, y \in Y$ 都有 $E_x \in \mathcal{B}$ 及 $E_y \in \mathcal{A}$.



证明 令 $\mathcal{S} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$, 显然对于一切 $E \in \mathcal{S}$, 结论成立. 再定义集族

$$\mathcal{L} = \{E \in \mathcal{A} \otimes \mathcal{B} : x \in X, y \in Y, E_x \in \mathcal{B}, E_y \in \mathcal{A}\}$$

易证 \mathcal{L} 是一个 λ -系, 由于 \mathcal{S} 为 π -系且 $\mathcal{S} \subset \mathcal{L}$, 根据 $\pi - \lambda$ 定理可知 $\sigma(\mathcal{S}) = \mathcal{A} \otimes \mathcal{B}$, 从而对于一切 $E \in \mathcal{A} \otimes \mathcal{B}$ 结论成立.

引理 1.9

设 (X, \mathcal{A}, μ_1) 和 (Y, \mathcal{B}, μ_2) 为两个 σ -有限测度空间, 那么对于任意集合 $E \in \mathcal{A} \otimes \mathcal{B}$, 函数 $f_E(x) = \mu_2(E_x)$ 与 $g_E(y) = \mu_1(E_y)$ 分别是 \mathcal{A} -可测与 \mathcal{B} -可测的.



证明 首先根据引理 1.8 可知 f_E 和 g_E 是定义良好的.

设 μ_2 为有限测度, 令 $\mathcal{S} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$, 再令

$$\mathcal{L} = \{E \in \mathcal{A} \otimes \mathcal{B} : f_E \text{ 为 } \mathcal{A}\text{-可测的}\}$$

注意到 $\mu_2((A \times B)_x) = 1_A(x)\mu_2(B)$, 由于 A 是 \mathcal{A} -可测的且 $\mu_2(B) \in [0, \infty]$, 故而 $f_{A \times B}$ 也是 \mathcal{A} -可测的, 由此可知 $\mathcal{S} \subset \mathcal{L}$.

现在证明集族 \mathcal{L} 是一个 λ -系. 令 $E = X \times Y$, 由于对一切 $x \in X$ 都有 $\mu_2(E_x) = \mu_2(Y)$, 故而 $X \times Y \in \mathcal{L}$. 再令 $A, B \in \mathcal{L}$ 且 $A \supset B$, 易证 $(A - B)_x = A_x - B_x$, 因为 μ_2 是有限测度, 故而

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

是 \mathcal{A} -可测的, 也即 $A - B \in \mathcal{L}$. 设 $B_1, B_2, \dots \in \mathcal{L}$ 且满足 $B_n \subset B_{n+1}$, 于是对于任意 $x \in X$ 都有 $(B_n)_x \subset (B_{n+1})_x$, 根据 MCT 可知

$$\lim_{n \rightarrow \infty} \mu_2((B_n)_x) = \mu_2\left(\bigcup_{n=1}^{\infty} (B_n)_x\right) = \mu_2\left(\left(\bigcup_{n=1}^{\infty} B_n\right)_x\right) < \infty$$

对于一切 $x \in X$ 都成立, 因此 $\mu_2((\bigcup_{n=1}^{\infty} B_n)_x)$ 是 \mathcal{A} -可测的, 也即 $\bigcup_{n=1}^{\infty} B_n \in \mathcal{L}$, 因此 \mathcal{L} 是一个 λ -系. 因为 \mathcal{S} 是一个 π -系, $\mathcal{S} \subset \mathcal{L}$, 根据 $\pi - \lambda$ 定理可知 $\mathcal{L} \supset \sigma(\mathcal{S})$, 因此 $\mathcal{L} = \sigma(\mathcal{S}) = \mathcal{A} \otimes \mathcal{B}$, 也即 f_E 是 \mathcal{A} -可测的.

现在设 μ_2 为 σ -有限测度, 故而存在可数个互不相交的集合 $D_1, D_2, \dots \in \mathcal{B}$, 使得 $Y = \bigcup_{n=1}^{\infty} D_n$ 且 $\mu_2(D_n) < \infty$, 其中 $n = 1, 2, \dots$. 再定义有限测度

$$\mu_{2n}(B) = \mu_2(B \cap D_n)$$

由于 μ_2 是 \mathcal{B} 上的测度, 因此

$$\mu_2(E_x) = \sum_{n=1}^{\infty} \mu_{2n}(E_x), \quad E \in \mathcal{A} \otimes \mathcal{B}$$

从而 f_E 是 \mathcal{A} -可测的, 同理可证 g_E 是 \mathcal{B} -可测的.

定理 1.36

设 (X, \mathcal{A}, μ_1) 和 (Y, \mathcal{B}, μ_2) 为两个 σ -有限测度空间, 则在乘积 σ -代数 $\mathcal{A} \otimes \mathcal{B}$ 上存在唯一的测度 $\mu_1 \otimes \mu_2$, 使得

$$(\mu_1 \otimes \mu_2)(A \times B) = \mu_1(A)\mu_2(B) \quad (1.26)$$

其中 $A \in \mathcal{A}$ 且 $B \in \mathcal{B}$.



证明 根据引理 1.9, 可以在 $\mathcal{A} \otimes \mathcal{B}$ 上定义集函数 λ_1 和 λ_2 :

$$\lambda_1(E) = \int_X \mu_2(E_x) \mu_1(dx) \quad (1.27)$$

$$\lambda_2(E) = \int_Y \mu_1(E_y) \mu_2(dy) \quad (1.28)$$

其中 $E \in \mathcal{A} \otimes \mathcal{B}$. 易知 λ_1 和 λ_2 均为测度, 并且满足

$$\lambda_1(A \times B) = \lambda_2(A \times B) = \mu_1(A)\mu_2(B)$$

由于 $\mathcal{S} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$ 是一个半代数, 并且 μ_1 和 μ_2 是 σ -有限的, 故而定义在 \mathcal{S} 上的加性函数

$$\nu(A \times B) = \mu_1(A)\mu_2(B)$$

可以被扩张为代数 $\bar{\mathcal{S}}$ 上的 σ -有限预测度. 根据 Carathéodory 扩张定理, 在 $\sigma(\mathcal{S}) = \mathcal{A} \otimes \mathcal{B}$ 上存在唯一的测度 $\mu_1 \otimes \mu_2$ 使得式 (1.26) 成立. 特别地, $\mu_1 \otimes \mu_2 = \lambda_1 = \lambda_2$.

根据定理 1.36, 对于 $i = 1, \dots, n$, 如果 $(\Omega_i, \mathcal{F}_i, \mu_i)$ 是 σ -有限测度空间, 定义 $\Omega = \Omega_1 \times \dots \times \Omega_n$, 那么在由 \mathcal{S} 生成的 σ -代数 \mathcal{F} 上存在唯一的测度 μ 使得

$$\mu(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \quad (1.29)$$

其中

$$\mathcal{S} = \{A_1 \times \dots \times A_n : A_i \in \mathcal{F}_i, i = 1, \dots, n\}$$

如果对于一切 $i = 1, \dots, n$ 都有 $(\Omega_i, \mathcal{F}_i, \mu_i) = (\mathbb{R}, \mathcal{R}, m)$, 那么 μ 是 n 维 Euclid 空间 \mathbb{R}^n 的 Borel 集上的 Lebesgue 测度.

定理 1.37 (Fubini 定理)

设 (X, \mathcal{A}, μ_1) 和 (Y, \mathcal{B}, μ_2) 为 σ -有限测度空间, $\mu = \mu_1 \otimes \mu_2$ 为 $\mathcal{F} = \mathcal{A} \otimes \mathcal{B}$ 上的乘积测度, 如果 $f \geq 0$ 或 $\int |f| d\mu < \infty$, 那么

$$\int_X \left[\int_Y f(x, y) \mu_2(dy) \right] \mu_1(dx) = \int_{X \times Y} f d\mu = \int_Y \left[\int_X f(x, y) \mu_1(dx) \right] \mu_2(dy)$$



证明 这里只证第一个等号, 因为第二个等号的证明是完全类似的. 根据乘积测度的存在性和式 (1.27) 可知

Case 1: 如果 $E \in \mathcal{F}$ 且 $f = 1_E$, 那么结论自然成立.

Case 2: 如果 f 为简单函数, 根据积分线性于 f 的性质可知结论成立.

Case 3: 假设 $f \geq 0$, 定义 $f_n(x) = ([2^n f(x)]/2^n) \wedge n$, 其中 $a \wedge b = \min(a, b)$, 而 $[x]$ 表示不超过 x 的最大整数. 于是 f_n 为简单函数并且 $f_n \uparrow f$, 根据 MCT 可知结论成立.

Case 4: 一般情况下, 注意到 $f = f^+ - f^-$, 将 Case 3 的结论应用到 f^+ , f^- 及 $|f|$ 即可证得结论成立.

示例 1.35 σ -有限测度的必要性 设 $X = Y = [0, 1]$, \mathcal{A} 为 $[0, 1]$ 的 Borel σ -代数, \mathcal{B} 为 $[0, 1]$ 的幂集, μ_1 为 Lebesgue 测度, μ_2 为计数测度, 定义函数

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

那么对于任意 $x \in X$ 有

$$\int_Y f(x, y) \mu_2(dy) = 1 \Rightarrow \int_X \int_Y f d(\mu_1 \otimes \mu_2) = 1$$

然而对于任意 $y \in Y$ 有

$$\int_X f(x, y) \mu_1(dx) = 0 \Rightarrow \int_Y \int_X f d(\mu_1 \otimes \mu_2) = 0$$

示例 1.36 基于选择公理和连续统假设, 可以在 $(0, 1)$ 上定义一种序关系 $<'$, 使得集合 $\{x : x <' y\}$ 对每个 y 来说都是可数集.

令 $X = Y = (0, 1)$, $\mathcal{A} = \mathcal{B}$ 为 Borel σ -代数, $\mu_1 = \mu_2$ 为 Lebesgue 测度. 定义函数

$$f(x, y) = \begin{cases} 1, & x <' y \\ 0, & \text{其它} \end{cases}$$

于是

$$\begin{aligned} \int_X f(x, y) \mu_1(dx) &= \mu_1(X \cap \{x : x <' y\}) = 0, \quad \forall y \in Y \\ \int_Y f(x, y) \mu_2(dy) &= \mu_2(Y \cap \{y : x <' y\}) = 1, \quad \forall x \in X \end{aligned}$$

此时 Fubini 定理失效.

示例 1.37 设 $g \geq 0$ 是测度空间 (X, \mathcal{A}, μ) 上的可测函数, 那么

$$\int_X g d\mu = (\mu \otimes \lambda)(\{(x, y) : 0 \leq y < g(x)\}) = \int_0^\infty \mu(\{x : g(x) > y\}) dy \quad (1.30)$$

证明 给定测度空间 (Y, \mathcal{B}, ν) , 其中 $Y = [0, \infty)$, \mathcal{B} 是配置在 Y 上的 Borel σ -代数, ν 为 Lebesgue

测度. 注意到

$$\begin{aligned}\int_X \int_Y f \, d\lambda \, d\mu &= \int_X g \, d\mu \\ \int_{X \times Y} f \, d(\mu \otimes \lambda) &= (\mu \otimes \lambda)(\{(x, y) : 0 \leq y < g(x)\}) \\ \int_Y \int_X f \, d\mu \, d\lambda &= \int_0^\infty \mu(\{x : g(x) > y\}) \, dy\end{aligned}$$

因为 $f \geq 0$, 由 Fubini 定理可知式 (1.30) 成立.

示例 1.38 设 μ 是 \mathbb{R} 上的有限测度并且 $F(x) = \mu((-\infty, x])$, 那么

$$\int_{\mathbb{R}} [F(x+c) - F(x)] \, dx = c\mu(\mathbb{R}) \quad (1.31)$$

证明 设函数 $f = 1_{\{(x,y): x < y \leq x+c\}}$, 注意到

$$\begin{aligned}\int_{\mathbb{R}} \int_{\mathbb{R}} f \, d\mu(y) \, dx &= \int_{\mathbb{R}} [F(x+c) - F(x)] \, dx \\ \int_{\mathbb{R}} \int_{\mathbb{R}} f \, dx \, d\mu(y) &= \int_{\mathbb{R}} c \, d\mu(y) = c\mu(\mathbb{R})\end{aligned}$$

于是由 Fubini 定理可知式 (1.31) 成立.

示例 1.39 概率积分 $I = \int_0^\infty e^{-x^2} \, dx = \frac{\sqrt{\pi}}{2}$.

证明 显然 $I > 0$, 故而只需证明 $I^2 = \pi/4$ 即可. 注意到

$$I^2 = \int_0^\infty \int_0^\infty e^{-x^2-y^2} \, dy \, dx$$

置 $y = xu$, 于是

$$I^2 = \int_0^\infty \int_0^\infty e^{-x^2-x^2u^2} x \, du \, dx$$

再置 $v = x^2(1+u^2)$, 根据 Fubini 定理可知

$$I^2 = \int_0^\infty \int_0^\infty \frac{e^{-v}}{2(1+u^2)} \, dv \, du = \frac{1}{2} \int_0^\infty e^{-v} \, dv \int_0^\infty \frac{du}{1+u^2} = \frac{\pi}{4}$$

1.8 积分下的微分

定理 1.38

设 (S, \mathcal{S}, μ) 是一个测度空间, f 是定义在 $\mathbb{R} \times S$ 上的复值函数. 令 $\delta > 0$, 如果以下条件成立

- (1) $u(x) = \int_S f(x, s) \, \mu(ds)$, 且 $\int_S |f(x, s)| \, \mu(ds) < \infty$;
- (2) 对某个固定 s , $\frac{\partial f}{\partial x}(x, s)$ 存在, 并且是关于 x 的连续函数;
- (3) $v(x) = \int_S \frac{\partial f}{\partial x}(x, s) \, \mu(ds)$ 在 $x = y$ 处连续;
- (4) $\int_S \int_{-\delta}^\delta \left| \frac{\partial f}{\partial x}(y + \theta, s) \right| \, d\theta \, \mu(ds) < \infty$;

那么 $u'(y) = v(y)$.



证明 令 $|h| \leq \delta$, 根据条件 (1), (2) 和 (4), 以及 Fubini 定理可知

$$\begin{aligned} u(y+h) - u(y) &= \int_S f(y+h) - f(y, s) \mu(ds) \\ &= \int_S \int_0^h \frac{\partial f}{\partial x}(y+\theta, s) d\theta \mu(ds) \\ &= \int_0^h \int_S \frac{\partial f}{\partial x}(y+\theta, s) \mu(ds) d\theta \end{aligned}$$

于是

$$\frac{u(y+h) - u(y)}{h} = \frac{1}{h} \int_0^h v(y+\theta) d\theta$$

由条件 (3) 可知 v 在 y 处连续, 令 $h \rightarrow 0$ 即可证得结论.

示例 1.40 设 $\varphi(s) = \int_{-\infty}^{\infty} \cos xs e^{-x^2/2} dx$, 那么它的表达式为

$$\varphi(s) = \sqrt{2\pi} e^{-s^2/2}$$

证明 首先设

$$u(x) = \int_{-\infty}^{\infty} e^{-x^2/2} \cos xs dx$$

其中 $f(x, s) = e^{-x^2/2} \cos xs$. 容易验证

$$\int_{-\infty}^{\infty} |e^{-x^2/2} \cos xs| dx \leq \int_{-\infty}^{\infty} e^{-x^2/2} dx < \infty$$

以及

$$\frac{\partial f}{\partial x}(x, s) = -e^{-x^2/2} (s \sin xs + x \cos xs)$$

故而条件 (2) 和 (3) 也满足. 最后注意到

$$\int_{-\infty}^{\infty} \left| \frac{\partial f}{\partial x}(x, s) \right| ds = \int_{-\infty}^{\infty} |s| e^{-s^2/2} ds < \infty$$

因此条件 (4) 成立. 根据定理 1.38 以及分部积分法可知

$$\begin{aligned} \varphi'(s) &= \int_{-\infty}^{\infty} -x e^{-x^2/2} \sin xs dx \\ &= \int_{-\infty}^{\infty} -s e^{-x^2/2} \cos xs dx = -s \varphi(s) \end{aligned}$$

解这个微分方程即可求得 $\varphi(s) = \sqrt{2\pi} e^{-s^2/2}$.

定理 1.39

设 (S, \mathcal{S}, μ) 是一个测度空间, f 是定义在 $\mathbb{R} \times S$ 上的复值函数. 令 $\delta > 0$, 如果以下条件成立

- (1) $u(x) = \int_S f(x, s) \mu(ds)$, 且 $\int_S |f(x, s)| \mu(ds) < \infty$;
- (2) 对某个固定 s , $\frac{\partial f}{\partial x}(x, s)$ 存在, 并且是关于 x 的连续函数;
- (3') $\int_S \sup_{\theta \in [-\delta, \delta]} \left| \frac{\partial f}{\partial x}(y+\theta, s) \right| \mu(ds) < \infty$;

那么 $u'(y) = v(y)$.



证明 根据定理 1.38, 只需证明条件 (3) 和 (4) 成立即可. 注意到

$$\int_{-\delta}^{\delta} \left| \frac{\partial f}{\partial x}(y + \theta, s) \right| d\theta \leq 2\delta \sup_{\theta \in [-\delta, \delta]} \left| \frac{\partial f}{\partial x}(y + \theta, s) \right|$$

这就验证了条件 (4). 再注意到

$$|v(x) - v(y)| \leq \int_S \left| \frac{\partial f}{\partial x}(x, s) - \frac{\partial f}{\partial x}(y, s) \right| \mu(ds)$$

条件 (2) 意味着当 $x \rightarrow y$ 时, 上述积分 $\rightarrow 0$, 于是根据 DCT 和条件 (3') 可知结论成立.

示例 1.41 如果 $\varphi(\theta) = E(e^{\theta Z}) < \infty$ 对 $\theta \in [-\epsilon, \epsilon]$, 那么 $\varphi'(0) = E(Z)$.

证明 设 μ 为随机变量 Z 的概率分布, 取 $\delta = \epsilon/2$, $f(x, s) = e^{xs} \geq 0$. 因此条件 (1) 成立. 此外, $\partial f / \partial x = se^{xs}$ 显然是一个连续函数, 故而条件 (2) 也成立. 最后, 注意到存在常数 C , 使得当 $x \in (-\delta, \delta)$ 时有 $|s|e^{xs} \leq C(e^{-\epsilon s} + e^{\epsilon s})$, 于是条件 (3') 也成立.

根据定理 1.39, 如果令 $S = \mathbb{Z}$, \mathcal{S} 为 S 的全体子集, μ 为 (S, \mathcal{S}) 上的计数测度, 那么可以得到以下定理.

定理 1.40

设 $\delta > 0$, 假设对于 $x \in (y - \delta, y + \delta)$ 以下条件成立

- (1) $u(x) = \sum_{n=1}^{\infty} f_n(x)$ 且 $\sum_{n=1}^{\infty} |f_n(x)| < \infty$;
- (2) 对于每个 $n \geq 1$, $f'_n(x)$ 存在且它是 x 的连续函数;
- (3) $\sum_{n=1}^{\infty} \sup_{\theta \in (-\delta, \delta)} |f'_n(y + \theta)| < \infty$;

那么 $u'(x) = v(x)$.



第2章 大数定律

2.1 独立性

2.1.1 基本概念

定义 2.1

- 事件 A 和 B 是独立的 (independent), 如果 $P(A \cap B) = P(A)P(B)$.
- 随机变量 X 和 Y 是独立的, 如果对于一切 $C, D \in \mathcal{R}$ 都有

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

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

- σ -代数 \mathcal{F} 和 \mathcal{G} 是独立的, 如果对于一切 $A \in \mathcal{F}, B \in \mathcal{G}$, 事件 A 和 B 是独立的.



定义 2.2

设 X 为一个随机变量, 那么由 X 生成的 σ -代数

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{R}\}$$



定理 2.1

- 如果随机变量 X 和 Y 独立, 那么 $\sigma(X)$ 和 $\sigma(Y)$ 独立;
- 如果 \mathcal{F} 和 \mathcal{G} 独立, $X \in \mathcal{F}, Y \in \mathcal{G}$, 那么 X 和 Y 独立.



证明 (1) 设 $A \in \sigma(X), B \in \sigma(Y)$, 根据 $\sigma(X)$ 和 $\sigma(Y)$ 的定义可知, 对于某些 $C \in \mathcal{R}$ 有 $A = \{X \in C\}$, 而对于某些 $D \in \mathcal{R}$ 有 $B = \{Y \in D\}$, 于是

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

(2) 设 $X \in \mathcal{F}, Y \in \mathcal{G}$, 并且 $C, D \in \mathcal{R}$, 根据可测性的定义可知 $\{X \in C\} \in \mathcal{F}, \{Y \in D\} \in \mathcal{G}$. 因为 \mathcal{F} 和 \mathcal{G} 是独立的, 所以 $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$.

定理 2.2

- 如果 A 和 B 是独立的, 那么 A^c 和 B , A 和 B^c , A^c 和 B^c 也都是独立的;
- 事件 A 和 B 是独立的, 当且仅当示性随机变量 1_A 和 1_B 是独立的.



证明 (1) 根据 $P(A \cap B) = P(A)P(B)$ 可知

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B)[1 - P(A)] = P(B)P(A^c)$$

其余几个证明类似.

(2) 显然, 如果 $C, D \in \mathcal{R}$, 那么 $\{1_A \in C\} \in \{\emptyset, A, A^c, \Omega\}$, 并且 $\{1_B \in D\} \in \{\emptyset, B, B^c, \Omega\}$, 故而 $P(1_A \in C, 1_B \in D)$ 有 16 种情况需要讨论. 只要 $\{1_A \in C\}$ 或 $\{1_B \in D\}$ 为 Ω 和 \emptyset 中的某

一个, 那么

$$P(1_A \in C, 1_B \in D) = P(1_A \in C)P(1_B \in D)$$

另一方面, 此时 A 和 B 分别对应于 $\{1_A \in C\}$ 和 $\{1_B \in D\}$, 也为 Ω 和 \emptyset 中的某一个, 因此上式等价于 $P(A \cap B) = P(A)P(B)$. 现在还剩 4 种情况需要考虑, 而它们可以由 (1) 直接证明.

现在考虑一般的情况. σ -代数 $\mathcal{F}_1, \dots, \mathcal{F}_n$ 是独立的, 如果当 $A_i \in \mathcal{F}_i$ 时有

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

随机变量 X_1, \dots, X_n 是独立的, 如果当 $B_i \in \mathcal{R}$ 时有

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

事件 A_1, \dots, A_n 是独立的, 如果当 $I \subset \{1, \dots, n\}$ 时有

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

注意, 仅假定对于一切 $i \neq j$ 都有 $P(A_i \cap A_j) = P(A_i)P(A_j)$ 不足以表明 A_1, \dots, A_n 都是独立的, 此时 A_1, \dots, A_n 两两独立 (pairwise independent). 显然, A_1, \dots, A_n 的独立性蕴含着两两独立, 但是反过来不成立, 下面给出一个例子.

示例 2.1 设 X_1, X_2, X_3 是独立的随机变量, 并且满足

$$P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$$

定义事件 $A_1 = \{X_2 = X_3\}$, $A_2 = \{X_3 = X_1\}$, $A_3 = \{X_1 = X_2\}$, 由于对于一切 $i \neq j$ 都有

$$P(A_i \cap A_j) = P(X_1 = X_2 = X_3) = \frac{1}{4} = P(A_i)P(A_j)$$

故而 A_1, A_2, A_3 两两独立, 但是它们不独立, 这是因为

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$$

定理 2.3

设事件 A_1, A_2, \dots, A_n 是独立的, 那么以下结论成立.

- (1) A_1^c, A_2, \dots, A_n 是独立的;
- (2) $1_{A_1}, 1_{A_2}, \dots, 1_{A_n}$ 是独立的.



证明 (1) 令 $B_1 = A_1^c$, 并且当 $i > 1$ 时 $B_i = A_i$. 如果 $I \subset \{1, 2, \dots, n\}$ 不包含 1, 那么显然有 $P(\bigcap_{i \in I} B_i) = \prod_{i \in I} P(B_i)$. 假设 $1 \in I$, 再令 $J = I - \{1\}$, 于是 $P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$, 从该式中减去 $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$ 即可得到 $P(A_1^c \cap \bigcap_{i \in J} A_i) = P(A_1^c) \prod_{i \in J} P(A_i)$.

(2) 如果 $B_i \in \{A_i, A_i^c\}$, 迭代使用 (i) 的结论即可推得 B_1, B_2, \dots, B_n 是独立的. 当 $C_i \in \{A_i, A_i^c, \Omega\}$ 时也有 $P(\bigcap_{i=1}^n C_i) = \prod_{i=1}^n P(C_i)$, 如果某些 $C_i = \emptyset$, 该式也是显然的. 最后注意到 $1_{A_i} \in \{\emptyset, A_i, A_i^c, \Omega\}$, 因此结论成立.

定义 2.3

集族 $\mathcal{G}_1, \dots, \mathcal{G}_n \subset \mathcal{F}$ 是独立的, 如果当 $A_i \in \mathcal{G}_i$ 且 $I \subset \{1, \dots, n\}$ 时有 $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

**定理 2.4**

设 A 为非空集合, $\alpha \in A$, 并且 $\mathcal{G}_\alpha \subset \mathcal{F}$ 是一个 π -系. 如果 \mathcal{G}_α 是独立的, 那么 $\{\sigma(\mathcal{G}_\alpha) : \alpha \in A\}$ 中的元素是独立的.



证明 固定 $2 \leq k < \infty$, $\{\alpha_1, \dots, \alpha_k\} \subset A$, 再令 $B_i \in \mathcal{G}_{\alpha_i}$, $i = 1, 2, \dots, k-1$. 定义集族

$$\mathcal{L} = \left\{ A \in \sigma(\mathcal{A}_n) : P(A_1 \cap \dots \cap A_{n-1} \cap A) = \left[\prod_{i=1}^{n-1} P(A_i) \right] P(A) \right\}$$

容易验证 \mathcal{L} 是一个 λ -系. 由于 \mathcal{L} 包含了 \mathcal{G}_{α_k} , 而它是一个 π -系, 根据 π - λ 定理可知 $\mathcal{L} = \sigma(\mathcal{G}_{\alpha_k})$, 迭代 k 次后即可证得结论.

定理 2.5

随机变量 X_1, X_2, \dots, X_n 独立, 当且仅当

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

其中 $x_1, x_2, \dots, x_n \in (-\infty, \infty]$.



证明 根据独立性的定义, 必要性的证明是显然的. 令 $\mathcal{A}_i = \{X_i \in (-\infty, x_i] : x_i \in (-\infty, \infty]\}$, 其中 $i = 1, 2, \dots, n$, 易知每个 \mathcal{A}_i 都是一个 π -系, 注意到 $\sigma(\mathcal{A}_i) = \sigma(X_i)$, 利用定理 2.1(2) 和 2.4 即可证得结论.

2.1.2 独立性, 分布与期望**定理 2.6**

如果 X_1, X_2, \dots, X_n 是独立的随机变量, 并且 X_i 的分布为 μ_i , 那么 (X_1, \dots, X_n) 的概率分布为 $\mu_1 \otimes \dots \otimes \mu_n$.



证明 设 $A_1, A_2, \dots, A_n \in \mathcal{R}$, 根据 X_1, X_2, \dots, X_n 的独立性可知

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

故而 (X_1, \dots, X_n) 的概率分布 $\mu_{(X_1, \dots, X_n)}$ 和测度 $\mu_1 \otimes \dots \otimes \mu_n$ 在集族 $\mathcal{P} = \{A_1 \times \dots \times A_n : A_i \in \mathcal{R}\}$ 上相同. 注意到 \mathcal{P} 是一个 π -系并且生成了 \mathcal{R}^n , 定义集族

$$\mathcal{L} = \{A \in \mathcal{R}^n : \mu_{(X_1, \dots, X_n)}(A) = \mu_1 \otimes \dots \otimes \mu_n(A)\}$$

易知 \mathcal{L} 是一个 λ -系且包含了 \mathcal{P} , 根据 $\pi - \lambda$ 定理可知 $\mathcal{L} = \mathcal{R}^n$, 结论成立.

注 由此可知, X_1, \dots, X_k 独立, 当且仅当 $X = (X_1, \dots, X_k)$ 的联合累积分布 F 和 X_i 的边际累积分布 F_i 具有以下关系

$$F(x_1, \dots, x_k) = F_1(x_1) \cdots F_k(x_k)$$

定理 2.7

假设 X 和 Y 是独立的随机变量并且具有分布 μ 和 ν , 函数 $h: (\mathbb{R}^2, \mathcal{R}^2) \rightarrow (\mathbb{R}^2, \mathcal{R}^2)$ 是可测的, 如果 $h \geq 0$ 或 $E|h(X, Y)| < \infty$, 那么

$$E[h(X, Y)] = \iint h(x, y) \mu(dx) \nu(dy)$$

特别地, 如果 $h(x, y) = f(x)g(y)$, 其中 $f, g: (\mathbb{R}, \mathcal{R}) \rightarrow (\mathbb{R}, \mathcal{R})$ 可测, 并且满足 $f, g \geq 0$ 或 $E|f(X)| < \infty, E|g(Y)| < \infty$, 那么

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$



证明 根据期望的计算公式和 Fubini 定理可知

$$E[h(X, Y)] = \int_{\mathbb{R}^2} h \, d(\mu \otimes \nu) = \iint h(x, y) \mu(dx) \nu(dy)$$

为了证明第二个结论, 先假设 $f, g \geq 0$, 根据期望计算公式得到

$$\begin{aligned} E[f(X)g(Y)] &= \iint f(x)g(y) \mu(dx) \nu(dy) = \int g(y) \left[\int f(x) \mu(dx) \right] \nu(dy) \\ &= E[f(X)] \int g(y) \nu(dy) = E[f(X)]E[g(Y)] \end{aligned}$$

将上述结论应用于 $|f|$ 和 $|g|$ 可得 $E|f(X)g(Y)| = E|f(X)|E|g(Y)| < \infty$, 再次套用期望计算公式即可证得结论.

推论 2.1

设随机变量 X_1 和 X_2 是独立的, 并且对于 $i = 1, 2$ 都有 $X_i \geq 0$ 或 $E|X_i| < \infty$, 那么 $E(X_1 X_2) = E(X_1)E(X_2)$.



注 正如初等概率论中提到的, 如果随机变量 X_1 和 X_2 满足 $E(X_1 X_2) = E(X_1)E(X_2)$, 那么 X_1 和 X_2 线性不相关, 但无法说明 X_1 和 X_2 独立.

示例 2.2 设 $\Omega = [0, 1]$, \mathcal{F} 为 Borel σ -代数, P 为 Lebesgue 测度, 那么 $X_n(\omega) = \sin(2\pi n\omega)$ 是不相关也不独立的.

证明 根据期望的定义可知

$$E(X_n) = \int_0^1 \sin(2\pi nx) \, dx = -(2\pi n)^{-1} \cos(2\pi nx) \Big|_0^1 = 0$$

当 $m \neq n$ 时, 根据积化和差可知

$$E(X_m X_n) = \int_0^1 \sin(2\pi mx) \sin(2\pi nx) \, dx = 0$$

由此可知 X_m 和 X_n 是不相关的.

任取 $0 < \epsilon < 1$, 定义 $A = X_m^{-1}((0, \epsilon))$, $B = X_n^{-1}(X_n(A))$, 于是 $A \subset B$. 进一步可知, 集合

A 有 $2n$ 个区间, 设 l_ϵ 为区间长度, 选取 ϵ 使得 $2nl_\epsilon < 1/m$, 这是可以做到的, 因为当 $\epsilon \downarrow 0$ 时有 $P(A) \downarrow 0$, 此时 $P(B) < 1$, 显然有 $P(A \cap B) = P(A) > P(A)P(B)$.

2.1.3 卷积公式

定理 2.8

如果 X 和 Y 是独立的, $F(x) = P(X \leq x)$, $G(y) = P(Y \leq y)$, 那么

$$P(X + Y \leq z) = \int F(z - y) dG(y) \quad (2.1)$$



证明 令 $h(x, y) = 1_{(x+y \leq z)}(x, y)$, μ 和 ν 分别为关于 F 和 G 的概率测度, 固定 y 不变, 于是

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

根据定理 2.7 可知

$$\begin{aligned} P(X + Y \leq z) &= \iint 1_{(x+y \leq z)} \mu(dx) \nu(dy) \\ &= \int F(z - y) \nu(dy) = \int F(z - y) dG(y) \end{aligned}$$

注 称式 (2.1) 等式右端为 F 和 G 的卷积 (convolution), 表示为 $F * G$.

定理 2.9

设随机变量 X 具有概率密度 f , Y 具有累积分布 G , 并且 X 和 Y 是独立的, 那么 $X + Y$ 的概率密度为

$$h(x) = \int f(x - y) dG(y)$$

当 Y 具有概率密度 g 时, 上式变为

$$h(x) = \int f(x - y)g(y) dy$$



证明 根据定理 2.8 和 Fubini 定理, 容易得到

$$\begin{aligned} P(X + Y \leq z) &= \int F(z - y) dG(y) = \int \left[\int_{-\infty}^z f(x - y) dx \right] dG(y) \\ &= \int_{-\infty}^z \left[\int f(x - y) dG(y) \right] dx \end{aligned}$$

再根据式 (1.19) 即可证得第二个式子.

示例 2.3 设随机变量 X 和 Y 独立, 且都服从 $(-1, 1)$ 上的均匀分布, 那么 $X + Y$ 的概率密度为

$$h(x) = \begin{cases} (x + 2)/4, & -2 < x \leq 0 \\ (2 - x)/4, & 0 < x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

证明 根据卷积公式可得

$$h(x) = \int f(x-y)g(y)dy = \int \frac{1}{4} \cdot 1_{(-1 \leq x-y \leq 1, -1 \leq y \leq 1)} dy$$

也即当 $-2 < x \leq 0$ 时

$$h(x) = \int_{-1}^{x+1} \frac{1}{4} dy = \frac{1}{4}x + \frac{1}{2}$$

而当 $0 < x \leq 2$ 时

$$h(x) = \int_{x-1}^1 \frac{1}{4} dy = -\frac{1}{4}x + \frac{1}{2}$$

即为所求.

根据定理2.9和一些微积分的运算, 我们可以处理一些标准例子, 首先考虑参数为 α 和 λ 的 Gamma 概率密度

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

其中 $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. 现在给出以下定理.

定理 2.10

设 $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$, 并且 X 和 Y 是独立的, 那么 $X + Y \sim \Gamma(\alpha + \beta, \lambda)$.



证明 设 f_{X+Y} 为 $X + Y$ 的 p.d.f., 根据定理2.9可知

$$\begin{aligned} f_{X+Y}(x) &= \int_0^x \frac{\lambda^\alpha (x-y)^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda(x-y)} \frac{\lambda^\beta y^{\beta-1}}{\Gamma(\beta)} e^{-\lambda y} dy \\ &= \frac{\lambda^{\alpha+\beta} e^{-\lambda x}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x (x-y)^{\alpha-1} y^{\beta-1} dy \end{aligned}$$

故而只需证明以上积分等于 $x^{\alpha+\beta-1} \Gamma(\alpha)\Gamma(\beta) / \Gamma(\alpha + \beta)$ 即可. 现在令 $y = xu$, 故而 $dy = xdu$, 以及

$$x^{\alpha+\beta-1} \int_0^1 (1-u)^{\alpha-1} u^{\beta-1} du = \int_0^x (x-y)^{\alpha-1} y^{\beta-1} dy$$

再令 $h = \Gamma(\alpha + \beta) \int_0^1 (1-u)^{\alpha-1} u^{\beta-1} du$, 注意到

$$\begin{aligned} h &= \int_0^\infty \int_0^x y^{\beta-1} e^{-y} (x-y)^{\alpha-1} e^{-(x-y)} dy dx \\ &= \int_0^\infty y^{\beta-1} e^{-y} \int_y^\infty (x-y)^{\alpha-1} e^{-(x-y)} dx dy = \Gamma(\alpha)\Gamma(\beta) \end{aligned}$$

于是结论成立.

最后指出, 如果随机变量 $X \sim N(\mu, a)$, $Y \sim N(v, b)$, 那么 $X \pm Y \sim N(\mu \pm v, a + b)$, 它的证明不再赘述.

2.1.4 独立随机变量的构造

我们知道, 给定有限个分别函数 $F_i, 1 \leq i \leq n$, 十分容易构造出独立随机变量 X_1, \dots, X_n , 使得 $P(X_i \leq x) = F_i(x)$. 令 $\Omega = \mathbb{R}^n, \mathcal{F} = \mathcal{R}^n$, 存在 $(\mathbb{R}^n, \mathcal{R}^n)$ 上的概率测度 P , 使得

$$P((a_1, b_1] \times \cdots \times (a_n, b_n]) = \prod_{m=1}^n [F_m(b_m) - F_m(a_m)] \quad (2.2)$$

如果 μ_i 是具有分布函数 F_i 的概率测度, 那么 $P = \mu_1 \otimes \cdots \otimes \mu_n$.

为了构造给定分布函数的无限个独立随机变量 X_1, X_2, \dots , 首先定义无穷乘积空间

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \mathbb{R}\}$$

其中 $\mathbb{N} = \{1, 2, \dots\}$ 为正整数集. 进一步定义 $X_i(\omega) = \omega_i$, 并且在 $\mathbb{R}^{\mathbb{N}}$ 上配置 σ -代数 $\mathcal{R}^{\mathbb{N}}$, 它是由有限维集合

$$\{\omega : \omega_i \in B_i, 1 \leq i \leq n\}, \quad B_i \in \mathcal{R}$$

生成的, 现在我们在上面定义 P , 并断言它在 $\mathbb{R}^{\mathbb{N}}$ 上具有唯一扩张.

定理 2.11 (Kolmogorov 扩张定理)

如果 $(\mathbb{R}^n, \mathcal{R}^n)$ 上的概率测度 μ_n 是一致的, 也即

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

那么在 $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ 上存在唯一的概率测度 P , 使得

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



在式 (2.2) 中, 令 $X_n(\omega) = \omega_n$, 那么 X_1, \dots, X_n 是独立的, 且 μ_n 是一致的, 于是 Kolmogorov 扩张定理保证了 X_1, X_2, \dots 的独立性. 事实上, 无限个随机变量 X_1, X_2, \dots 是独立的, 当且仅当它们的每一个有限子列是独立的.

2.2 弱大数定律

2.2.1 L^2 弱定律

本节的最终目的是推导出弱大数定律 (Weak Law of Large Numbers, WLLN), 为了得到前置定理, 我们需要用到关于方差的公式及 Chebyshev 不等式.

首先拓展之前随机变量不相关的定义, 任意一族满足 $E(X_i^2) < \infty$ 的随机变量 $X_i, i \in I$ 是不相关的, 如果当 $i \neq j$ 时有 $E(X_i X_j) = E(X_i)E(X_j)$. 现在给出一些预备定理.

定理 2.12

设 X_1, \dots, X_n 满足 $E(X_i^2) < \infty$, 并且它们都是不相关的, 那么

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



证明 令 $\mu_i = E(X_i)$, $S_n = \sum_{i=1}^n X_i$. 因为 $E(S_n) = \sum_{i=1}^n \mu_i$, 根据方差的定义可知

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

上式最后等号的第一个式子为 $\text{Var}(X_1) + \cdots + \text{Var}(X_n)$, 故而需要证明第二个式子为 0. 注意到

$$\begin{aligned}E[(X_i - \mu_i)(X_j - \mu_j)] &= E(X_i X_j) - \mu_i E(X_j) - \mu_j E(X_i) + \mu_i \mu_j \\ &= E(X_i X_j) - \mu_i \mu_j = 0\end{aligned}$$

定理得证.

定义 2.4

- 随机变量序列 $\{X_n\}_{n \geq 1}$ 依概率收敛 (converge in probability) 于 X , 如果对于任意的 $\epsilon > 0$, 当 $n \rightarrow \infty$ 时有 $P(|X_n - X| > \epsilon) \rightarrow 0$, 记作 $X_n \xrightarrow{P} X$.
- 随机变量序列 $\{X_n\}_{n \geq 1}$ 依 L^p 收敛 (converge in L^p) 于 X , 如果对于当 $n \rightarrow \infty$ 时有 $E|X_n - X|^p \rightarrow 0$, 记作 $X_n \xrightarrow{L^p} X$.
- 随机变量序列 $\{X_n\}_{n \geq 1}$ 几乎必然收敛 (almost surely converge) 于 X , 如果对于任意的 $\epsilon > 0$ 都有 $P(\lim |X_n - X| < \epsilon) = 1$, 记作 $X_n \xrightarrow{a.s.} X$.



定理 2.13

$$Z_n \xrightarrow{a.s.} Z \Rightarrow Z_n \xrightarrow{P} Z.$$



证明 任取 $\epsilon > 0$, 可以得到

$$P(|Z_n - Z| < \epsilon) \geq P(|Z_n - Z| < \epsilon, \forall j \geq n) \quad (2.3)$$

也即

$$\{\omega : |Z_j(\omega) - Z(\omega)| < \epsilon, \forall j \geq n\} \subset \{\omega : |Z_n(\omega) - Z(\omega)| < \epsilon\}$$

因为 $Z_n \xrightarrow{a.s.} Z$, 根据几乎必然收敛的定义可知, 存在充分大的正整数 N , 使得对于任意给定的 $\delta > 0$, 当 $n > N$ 时就有 $P(|Z_n - Z| < \epsilon, \forall j \geq n) > 1 - \delta$. 在式 (2.3) 两端取极限得到

$$\lim_{n \rightarrow \infty} P(|Z_n - Z| < \epsilon) \geq \lim_{n \rightarrow \infty} P(|Z_n - Z| < \epsilon, \forall j \geq n) = 1$$

也即 $Z_n \xrightarrow{P} Z$.

示例 2.4 依概率收敛但不几乎必然收敛 设 X_1, X_2, \dots 是 i.i.d. 随机变量, 满足

$$P(X_n = 1) = 1/n, \quad P(X_n = 0) = 1 - 1/n$$

对于任意 $0 < \epsilon < 1/2$, 当 $n \rightarrow \infty$ 时有

$$P(|X_n| > \epsilon) = 1/n \rightarrow 0$$

然而根据 Borel-Cantelli 第二引理 (见后文) 可知

$$\sum_{n=1}^{\infty} P(X_n = 1) = \infty$$

从而 $P(X_n = 1 \text{ i.o.}) = 1$.

定理 2.14

$$Z_n \xrightarrow{L^p} Z \Rightarrow Z_n \xrightarrow{p} Z.$$



证明 对于任意的 $\epsilon > 0$, 由 Markov 不等式可知当 $n \rightarrow \infty$ 时有

$$P(|Z_n - Z| > \epsilon) \leq \frac{E|Z_n - Z|^p}{\epsilon^p} = 0$$

由此可知 $Z_n \xrightarrow{p} Z$.

示例 2.5 依概率收敛但不依 L^p 收敛 设随机变量 X_n 的概率分布为

$$P(X_n = 1/n) = 1 - 1/n, \quad P(X_n = n) = 1/n$$

当 $n \rightarrow \infty$ 时有

$$E|X_n|^p = n^{p-1} + \frac{1}{n^p} - \frac{1}{n^{p+1}} \rightarrow \infty$$

并且对任意 $0 < \epsilon < 1/2$ 有

$$P(|X_n| > \epsilon) = P(X_n = n) = 1/n \rightarrow 0$$

定理 2.15 (L^2 弱定律)

设 X_1, X_2, \dots 为不相关随机变量, $E(X_i) = \mu$ 且 $\text{Var}(X_i) \leq C < \infty$, 定义 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时 $S_n/n \xrightarrow{L^2} \mu$ 且 $S_n/n \xrightarrow{p} \mu$.



证明 注意到 $E(S_n/n) = \mu$, 于是

$$E(S_n/n - \mu)^2 = \text{Var}(S_n/n) = \frac{\text{Var}(X_1) + \dots + \text{Var}(X_n)}{n^2} \leq \frac{Cn}{n^2} \rightarrow 0$$

故而 S_n/n 依 L^2 收敛于 μ . 将定理 2.14 施加到 $Z_n = S_n/n - \mu$ 即可证明依概率收敛.

在定理 2.15 中, 最重要的特殊情况为 X_1, X_2, \dots 是独立同分布 (independent and identical distributed, i.i.d.) 随机变量, 此时可以将定理 2.15 中的条件 $\text{Var}(X_i) < \infty$ 替换为 $E|X_i| < \infty$, 而结论保持不变.

示例 2.6 多项式逼近 设 f 是 $[0, 1]$ 上的连续函数, 并且

$$f_n(x) = \sum_{m=0}^n \binom{n}{m} x^m (1-x)^{n-m} f(m/n)$$

为关于 f 的 Bernstein 多项式, 那么当 $n \rightarrow \infty$ 时有

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

证明 设 X_1, \dots, X_n 为独立随机变量, 满足 $P(X_i = 1) = p, P(X_i = 0) = 1 - p$, 于是 $E(X_i) = p$ 且 $\text{Var}(X_i) = p(1 - p)$, 再令 $S_n = X_1 + \dots + X_n$, 故而

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

从而 $E[f(S_n/n)] = f_n(p)$. 根据定理 2.15 可知当 $n \rightarrow \infty$ 时有 $S_n/n \xrightarrow{p} p$.

注意到当 $p \in [0, 1]$ 时有 $p(1 - p) \leq 1/4$, 利用 Chebyshev 不等式可知

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

对任意 $\delta > 0$ 成立. 令 $M = \sup_{x \in [0, 1]} |f(x)|$, 由于 f 在 $[0, 1]$ 上是连续的, 故而也是一致连续的, 因此对于任意的 $\epsilon > 0$, 总存在 $\delta > 0$, 使得当 $|x - y| < \delta$ 时有 $|f(x) - f(y)| < \epsilon$. 根据 Jensen 不等式可知

$$\begin{aligned} |E[f(S_n/n)] - f(p)| &\leq E|f(S_n/n) - f(p)| \\ &\leq \epsilon + 2MP(|S_n/n - p| > \delta) \end{aligned}$$

最后令 $n \rightarrow \infty$, 于是

$$\limsup_{n \rightarrow \infty} |E[f(S_n/n)] - f(p)| \leq \epsilon$$

因为 ϵ 是任意的, 因此结论成立.

2.2.2 三角阵列与截断

定义 2.5

设 $\{X_{n,1}, \dots, X_{n,n}\}$ 是定义在 $(\Omega_n, \mathcal{F}_n, P_n)$ 上的随机变量序列, 并且使得 $X_{n,1}, \dots, X_{n,n}$ 独立, 那么称 $\{X_{n,k} : 1 \leq k \leq n\}_{n \geq 1}$ 为独立随机变量的三角阵列 (triangular array), 每行的和为 $S_n = X_{n,1} + \dots + X_{n,n}$.



定理 2.16

设 $\mu_n = E(S_n)$, $\sigma_n^2 = \text{Var}(S_n)$, 如果 $\sigma_n^2/b_n^2 \rightarrow 0$, 那么当 $n \rightarrow \infty$ 时有

$$\frac{S_n - \mu_n}{b_n} \xrightarrow{p} 0$$



证明 注意到 $E[(S_n - \mu_n)/b_n]^2 = b_n^{-2} \text{Var}(S_n) \rightarrow 0$, 也即 $(S_n - \mu_n)/b_n$ 依 L^2 收敛于 0, 根据定理 2.14 即可推出依概率收敛.

注 上述定理无需用到随机变量 $X_{n,1}, \dots, X_{n,n}$ 独立这一条件.

示例 2.7 卡牌收集问题 设 X_1, X_2, \dots 是 i.i.d. 随机变量且在 $\{1, \dots, n\}$ 上服从均匀分布, 可以将其视为卡池中的不同卡牌, 令 $\tau_k^n = \inf\{m : |X_1, \dots, X_m| = k\}$ 表示首次有 k 张不同卡牌所需的时间, 我们现在对 τ_k^n 的渐近性质感兴趣, 它表示完整收集整套卡牌所需的时间.

易知 $\tau_1^n = 1$, 进一步假定 $\tau_0^n = 0$, 并且对于任意 $1 \leq k \leq n$, 定义 $X_{n,k} = \tau_k^n - \tau_{k-1}^n$, 它表示抽到与之前 $(k-1)$ 张不同卡牌所需时间, 独立于之前等待的时间 $X_{n,j}, 1 \leq j < k$, 并且 $X_{n,k}$

服从 $p = 1 - (k-1)/n$ 的几何分布, 进而

$$E(T_n) = \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-1} = n \sum_{m=1}^n m^{-1} \sim n \log n$$

$$\text{Var}(T_n) \leq \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-2} = n^2 \sum_{m=1}^n m^{-2} \leq n^2 \sum_{m=1}^{\infty} m^{-2}$$

在定理2.16中取 $b_n = n \log n$, 当 $n \rightarrow \infty$ 时有

$$\frac{T_n - n \sum_{m=1}^n m^{-1}}{n \log n} \xrightarrow{p} 0$$

也即 $T_n/n \log n \xrightarrow{p} 1$.

为了将弱定律推广至无需二阶矩条件的情形, 我们需要将随机变量进行截断 (truncation), 具体而言就是

$$\bar{X} = X 1_{(|X| \leq M)} = \begin{cases} X, & |X| \leq M \\ 0, & |X| > M \end{cases}$$

然后再使用 Chebyshev 不等式和一些必需的定理.

定理 2.17 (三角阵列的弱定律)

假设对于每一个 n , 随机变量 $X_{n,k}$ 都是独立的. 再令 $b_n > 0$, 以及 $\bar{X}_{n,k} = X_{n,k} 1_{(|X_{n,k}| \leq b_n)}$, 假设当 $n \rightarrow \infty$ 时有

- $b_n \rightarrow \infty$.
- $\sum_{k=1}^n P(|X_{n,k}| > b_n) \rightarrow 0$.
- $b_n^{-2} \sum_{k=1}^n E(\bar{X}_{n,k}^2) \rightarrow 0$.

再令 $S_n = \sum_{k=1}^n X_{n,k}$, $a_n = \sum_{k=1}^n E(\bar{X}_{n,k})$, 那么 $(S_n - a_n)/b_n \xrightarrow{p} 0$.



证明 设 $\bar{S}_n = \sum_{k=1}^n \bar{X}_{n,k}$, 根据概率的次可加性得到

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

注意到

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

再根据 Chebyshev 不等式可知

$$\begin{aligned} P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right) &\leq \epsilon^{-2} E\left|\frac{\bar{S}_n - a_n}{b_n}\right|^2 = \epsilon^{-2} b_n^{-2} \text{Var}(\bar{S}_n) \\ &= (b_n \epsilon)^{-2} \sum_{k=1}^n \text{Var}(\bar{X}_{n,k}) \\ &\leq (b_n \epsilon)^{-2} \sum_{k=1}^n E(\bar{X}_{n,k}^2) \rightarrow 0 \end{aligned}$$

由此证得结论.

引理 2.1

如果随机变量 $Y \geq 0$ 且实数 $p > 0$, 那么 $E(Y^p) = \int_0^\infty py^{p-1}P(Y > y)dy$.



证明 根据期望的定义和 Fubini 定理可知

$$\begin{aligned}\int_0^\infty py^{p-1}P(Y > y)dy &= \int_0^\infty \int_\Omega py^{p-1}1_{(Y>y)}dPdy \\ &= \int_\Omega \int_0^\infty py^{p-1}1_{(Y>y)}dydP \\ &= \int_\Omega \int_0^Y py^{p-1}dydP = \int_\Omega Y^p dP = E(Y^p)\end{aligned}$$

这就得到了希望的结果.

定理 2.18 (弱大数定律)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, 并且当 $x \rightarrow \infty$ 时有

$$xP(|X_1| > x) \rightarrow 0$$

令 $S_n = X_1 + \dots + X_n$ 及 $\mu_n = E[X_1 1_{(|X_1| \leq n)}]$, 那么当 $n \rightarrow \infty$ 时有 $S_n/n - \mu_n \xrightarrow{P} 0$.



证明 在定理 2.17 中, 令 $X_{n,k} = X_k$ 及 $b_n = n$, 首先注意到

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

又有

$$P(|\bar{X}_{n,k}| > y) = \begin{cases} 0, & y \geq n \\ P(|X_1| > y) - P(|X_1| > n), & y < n \end{cases}$$

于是根据引理 2.1 可知

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

设 $g(y) = 2yP(|X_1| > y)$, 因为 $0 \leq g(y) \leq 2y$, 并且当 $y \rightarrow \infty$ 时有 $g(y) \rightarrow 0$, 故而 $M = \sup g(y) < \infty$. 再令 $g_n(y) = g(ny)$, 因为 g_n 是有界的且当 $n \rightarrow \infty$ 时有 $g_n(y) \rightarrow 0$, 因此由 DCT 可知

$$n^{-1} \int_0^n g(y)dy = \int_0^1 g_n(x)dx \rightarrow 0$$

由此可知

$$\lim_{n \rightarrow \infty} E(\bar{X}_{n,1}^2)/n = 0$$

因此根据定理 2.17 即可证得结论.

注 在引理 2.1 中令 $p = 1 - \epsilon$, 其中 $\epsilon > 0$, 因此 $xP(|X_1| > x) \rightarrow 0$ 意味着 $E|X_1|^{1-\epsilon} < \infty$.

定理 2.19 (Khinchin 弱大数定律)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, 并且 $E|X_i| < \infty$, 令 $S_n = X_1 + \dots + X_n$ 与 $\mu = E(X_1)$, 那么 $S_n/n - \mu \xrightarrow{p} 0$.



证明 令 $Y = |X_1|1_{(|X_1|>x)}$, 显然有 $Y \geq x1_{(|X_1|>x)}$, 两端取期望得到

$$xP(|X_1| > x) \leq E[|X_1|1_{(|X_1|>x)}] = \int_x^\infty P(|X_1| > x) dx$$

于是当 $x \rightarrow \infty$ 时有 $xP(|X_1| > x) \rightarrow 0$. 注意到

$$\lim_{n \rightarrow \infty} X_1 1_{(|X_1| \leq n)} = X_1$$

又因为 $E|X_1| < \infty$, 根据 DCT 可知当 $n \rightarrow \infty$ 时有

$$\mu_n = E[X_1 1_{(|X_1| \leq n)}] \rightarrow E(X_1) = \mu$$

最后根据弱大数定律, 对于任意给定的 $\epsilon > 0$, 当 $n \rightarrow \infty$ 时有 $P(|S_n/n - \mu_n| > \epsilon/2) \rightarrow 0$, 又因为 $\mu_n \rightarrow \mu$, 因此

$$\lim_{n \rightarrow \infty} P(|S_n/n - \mu| > \epsilon) = 0$$

结论成立.

示例 2.8 Monte Carlo 积分 设 f 是 $[0, 1]$ 上的可测函数, $\int_0^1 |f(x)| dx < \infty$. 令 U_1, U_2, \dots 是独立随机变量, 并且在 $[0, 1]$ 上服从均匀分布, 那么当 $n \rightarrow \infty$ 时

$$I_n = n^{-1}[f(U_1) + \dots + f(U_n)]$$

依概率收敛于 $I = \int_0^1 f(x) dx$.

证明 由于 f 在 $[0, 1]$ 上是可测的, 故而 $f(U_1), f(U_2), \dots$ 也是 i.i.d. 随机变量. 根据均匀分布可得 $E[f(U_1)] = \int_0^1 f(x) dx$, 因为 $E|f(U_1)| < \infty$, 所以由 WLLN 可得

$$I_n = \frac{f(U_1) + \dots + f(U_n)}{n} \xrightarrow{p} E[f(U_1)] = \int_0^1 f(x) dx = I$$

示例 2.9 设 X_1, X_2, \dots 为 i.i.d. 随机变量, 并且对于一切 $k \geq 2$ 都有

$$P(X_i = (-1)^k k) = \frac{C}{k^2 \log k}$$

其中 C 为使得概率之和为 1 的那个常数. 此时 $E|X_i| = \infty$, 但存在有限值 μ 使得 $S_n/n \xrightarrow{p} \mu$.

证明 首先注意到

$$\sum_{k=2}^{\infty} \frac{1}{k \log k} \geq \int_2^{\infty} \frac{dx}{x \log x} = \infty$$

于是 $E|X_i| = \sum_{k=2}^{\infty} C/k \log k = \infty$. 又因为

$$P(|X_i| > n) \leq \sum_{k=n+1}^{\infty} \frac{C}{k^2 \log k} \leq \frac{C}{n \log n}$$

从而当 $n \rightarrow \infty$ 时有 $nP(|X_i| > n) \rightarrow 0$. 此外, 还存在 $0 < \mu < \infty$ 使得当 $n \rightarrow \infty$ 时

$$\mu_n = E[X_i 1_{(|X_i| \leq n)}] = \sum_{k=2}^n (-1)^k \frac{C}{k \log k} \rightarrow \mu$$

根据 WLLN 可知结论成立.

示例 2.10 圣彼得堡悖论 设 X_1, X_2, \dots 是独立随机变量, 并且

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

换言之, 假设投出一枚均匀硬币的正面为成功, 一旦成功则游戏结束, 如果你在第 j 次掷币时成功, 那么可以获得 2^j 元. 这个悖论在于, 尽管 $E(X_1) = \infty$, 但是你肯定不想花费大量金钱来玩这个游戏.

现在令 $S_n = X_1 + \dots + X_n$, 表示进行 n 次游戏获得的总收益, 由于使用 WLLN 的矩条件 $E|X| < \infty$ 不成立, 因此无法用它来解释为何人们不愿花太多钱玩这个游戏. 事实上, 可以证明当 $n \rightarrow \infty$ 时, $S_n/(n \log_2 n) \xrightarrow{P} 1$, 也即当 n 相当大时, 玩 n 次这样的游戏, 平均每次花费的金钱应为 $\log_2 n$ 元.

证明 考虑使用定理 2.17 来解决这一问题. 在这个例子中, $X_{n,k} = X_k$, 首先注意到

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

令 $m(n) = \log_2 n + K(n)$, 其中 $K(n)$ 使得 $m(n)$ 为整数, 并且当 $n \rightarrow \infty$ 时有 $K(n) \rightarrow \infty$. 再令 $b_n = 2^{m(n)}$, 显然有 $b_n \rightarrow \infty$, 以及

$$nP(X_1 \geq b_n) = n2^{-m(n)+1} = 2^{-K(n)+1} \rightarrow 0$$

再设 $\bar{X}_{n,k} = X_k 1_{(|X_k| \leq b_n)}$, 于是

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

从而

$$b_n^{-2} \sum_{k=1}^n E(\bar{X}_{n,k}^2) \leq 2n/b_n = 2^{-K(n)+1} \rightarrow 0$$

又因为

$$E(\bar{X}_{n,k}) = \sum_{j=1}^{m(n)} 2^j 2^{-j} = m(n)$$

所以 $a_n = nm(n)$. 如果我们选取 $K(n)$ 使得当 $n \rightarrow \infty$ 时有 $K(n)/\log_2 n \rightarrow 0$, 那么此时 $a_n/n \log_2 n \rightarrow 1$. 根据定理 2.17 可知

$$\frac{S_n - a_n}{n2^{K(n)}} \xrightarrow{P} 0$$

进一步调整 $K(n)$, 使得对于充分大的 n 有 $K(n) \leq \log_2 \log_2 n$, 因此上式的分母可由 $n \log_2 n$ 替代, 也即

$$\frac{S_n}{n \log_2 n} \xrightarrow{P} 1$$

注 这仍然是一个依概率收敛但不几乎必然收敛的例子, 可以通过下一节的 Borel-Cantelli 引理证明.

2.3 Borel-Cantelli 引理

定义 2.6

设 (Ω, \mathcal{F}) 为可测空间, $\{A_n\}_{n \geq 1} \subset \mathcal{F}$, 则它的上极限和下极限分别为

$$\begin{aligned}\limsup_{n \rightarrow \infty} A_n &= \lim_{m \rightarrow \infty} \bigcup_{n=m}^{\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n \\ \liminf_{n \rightarrow \infty} A_n &= \lim_{m \rightarrow \infty} \bigcap_{n=m}^{\infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} A_n\end{aligned}$$



集合上极限 $\limsup A_n$ 是在无限个 A_n 中出现的元素构成的集合, 通常记为 $\limsup A_n = \{\omega : \omega \in A_n \text{ i.o.}\}$, 其中 i.o. 表示 infinitely often. 集合下极限 $\liminf A_n$ 是在有限个 A_n 中未出现的元素构成的集合.

示例 2.11 设 $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{R}$, 定义

$$A_n = \begin{cases} [0, 1/n], & n = 2k + 1, k = 0, 1, \dots \\ [1 - 1/n, 1], & n = 2k, k = 1, 2, \dots \end{cases}$$

那么 $\limsup A_n = \{0, 1\}$, $\liminf A_n = \emptyset$.

定理 2.20 (Borel-Cantelli 第一引理)

如果 $\sum_{n=1}^{\infty} P(A_n) < \infty$, 那么 $P(\limsup A_n) = 0$.



证明 设 $N = \sum_k 1_{A_k}$ 代表事件发生的次数, Fubini 定理意味着 $E(N) = \sum_k P(A_k) < \infty$, 因此必然有 $N < \infty$, 也即 $P(A_n \text{ i.o.}) = 0$.

定理 2.21 (Borel 强大数定律)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = \mu$ 且 $E(X_i^4) < \infty$, 再设 $S_n = X_1 + \dots + X_n$, 那么 $S_n/n - \mu \xrightarrow{a.s.} 0$.



证明 设 $X'_i = X_i - \mu$, 故而不失一般性地假设 $\mu = 0$. 现有

$$E(S_n^4) = E\left(\sum_{1 \leq i, j, k, l \leq n} X_i X_j X_k X_l\right)$$

如果 i, j, k, l 各不相同, 那么形式为 $E(X_i^3 X_j)$, $E(X_i^2 X_j X_k)$, $E(X_i X_j X_k X_l)$ 的和为 0, 因此上式中只有形式为 $E(X_i^4)$ 和 $E(X_i^2 X_j^2) = [E(X_i^2)]^2$ 的项不为 0. 此时

$$E(S_n^4) = nE(X_1^4) + 3(n^2 - n)[E(X_1^2)]^2 \leq Cn^2$$

其中 C 是某个充分大的实数. 根据 Chebyshev 不等式可知

$$P(|S_n| > n\epsilon) \leq (n\epsilon)^{-4} E(S_n^4) \leq C/(n^2 \epsilon^4)$$

让上式在 n 上取和, Borel-Cantelli 第一引理表明 $P(|S_n| > n\epsilon \text{ i.o.}) = 0$, 因为 ϵ 是任意的, 证明


完毕.

注 上述证明用到了几乎必然收敛的等价定义, 也即

$$\begin{aligned}
 X_n \xrightarrow{a.s.} X &\Leftrightarrow P(\lim X_n = X) = 1 \\
 &\Leftrightarrow P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1 \\
 &\Leftrightarrow P(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} \{|X_n - X| > \epsilon\}) = 0 \\
 &\Leftrightarrow P(\limsup \{|X_n - X| > \epsilon\}) = 0 \\
 &\Leftrightarrow P(|X_n - X| > \epsilon \text{ i.o.}) = 0
 \end{aligned}$$

此外, Borel 强大数定律要求 X_i 的四阶矩有限, 同样可以将该条件弱化为 $E|X_i| < \infty$, 这就是 Kolmogorov 强大数定律.

定理 2.22 (Borel-Cantelli 第二引理)

如果事件 A_n 是独立的, $n = 1, 2, \dots$, 并且 $\sum_{n=1}^{\infty} P(A_n) = \infty$, 那么 $P(\limsup A_n) = 1$. 

证明 设 $M < N < \infty$, 根据独立性和 $1 - x \leq e^{-x}$ 可知当 $N \rightarrow \infty$ 时有

$$\begin{aligned}
 P\left(\bigcap_{n=M}^N A_n^c\right) &= \prod_{n=M}^N [1 - P(A_n)] \leq \prod_{n=M}^N \exp[-P(A_n)] \\
 &= \exp\left[-\sum_{n=M}^N P(A_n)\right] \rightarrow 0
 \end{aligned}$$

因此对于所有 M 都有 $P(\bigcup_{n=M}^{\infty} A_n) = 1$, 又因为 $\bigcup_{n=M}^{\infty} A_n \downarrow \limsup A_n$, 故而结论成立.

示例 2.12 圣彼得堡悖论: 续 前面已经提到, $S_n/n \log_2 n$ 依概率收敛于 1, 但却不是几乎必然收敛于 1 的. 为了看到这一点, 只需证明对于任意 $K > 0$ 都有

$$P(X_n/n \log_2 n > K \text{ i.o.}) = 1 \quad (2.4)$$

因为这一结论可以适用于 S_n .

设 $k_n = \lfloor \log_2(Kn \log_2 n) \rfloor + 1$, 于是


$$\begin{aligned}
 P(X_n/n \log_2 n > K) &= \sum_{j=k_n}^{\infty} 2^{-j} = 2^{1-k_n} \\
 &= 2^{-\lfloor \log_2(Kn \log_2 n) \rfloor} \geq (Kn \log_2 n)^{-1}
 \end{aligned}$$

从而

$$\sum_{n=2}^{\infty} P\left(\frac{X_n}{n \log_2 n} > K\right) \geq \sum_{n=2}^{\infty} \frac{1}{Kn \log_2 n} = \infty$$

根据 Borel-Cantelli 第二引理可知式 (2.4) 成立.

引理 2.2

对于任意随机变量 X 都有 $\sum_{n=1}^{\infty} P(|X| > n) \leq E|X| \leq \sum_{n=0}^{\infty} P(|X| > n)$. 

证明 对于 $n \geq 1$, 定义 $A_n = \{\omega : n-1 < |X(\omega)| \leq n\}$, 以及随机变量

$$Y = \sum_{n=1}^{\infty} (n-1)1_{A_n}, \quad Z = \sum_{n=1}^{\infty} n1_{A_n}$$

显然有 $Y \leq |X| \leq Z$, 进而

$$E(Y) \leq E|X| \leq E(Z)$$

注意到

$$\begin{aligned} E(Y) &= \sum_{n=1}^{\infty} (n-1)P(A_n) = \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} P(A_n) \\ &= \sum_{k=1}^{\infty} \sum_{n=k+1}^{\infty} P(n-1 < |X| \leq n) \\ &= \sum_{k=1}^{\infty} P(|X| > k) \end{aligned}$$

类似可以证明 $E(Z) = \sum_{k=0}^{\infty} P(|X| > k)$.

定理 2.23

设 X_1, X_2, \dots 是 i.i.d. 随机变量且 $E|X_i| = \infty$, 那么 $P(|X_n| \geq n \text{ i.o.}) = 1$. 因此若 $S_n = X_1 + \dots + X_n$, 则 $P(\lim S_n/n \in (-\infty, \infty)) = 0$.



证明 根据引理 2.1 和 2.2 可知

$$E|X_1| = \int_0^{\infty} P(|X_1| > x) dx \leq \sum_{n=0}^{\infty} P(|X_1| > n)$$

因为 $E|X_1| = \infty$ 且 X_1, X_2, \dots 是 i.i.d. 随机变量, 根据 Borel-Cantelli 第二引理可知 $P(|X_n| \geq n \text{ i.o.}) = 1$. 注意到

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

再设 $C = \{\omega : \lim S_n/n \in (-\infty, \infty)\}$, 故而当 $n \rightarrow \infty$ 时, 在 C 上有 $S_n/[n(n+1)] \rightarrow 0$, 也即对于充分大的 n 有 $|S_n/[n(n+1)]| < 1/3$, 进而在集合 $C \cap \{\omega : |X_n| \geq n \text{ i.o.}\}$ 上有

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| \geq \left| \frac{X_{n+1}}{n+1} \right| - \left| \frac{S_n}{n(n+1)} \right| > \frac{2}{3} \quad \text{i.o.}$$

这与 $\omega \in C$ 矛盾, 因此

$$\{\omega : |X_n| \geq n \text{ i.o.}\} \cap C = \emptyset$$

因为 $P(|X_n| \geq n \text{ i.o.}) = 1$, 这意味着 $P(C) = 0$.

定理 2.24

设 A_1, A_2, \dots 是两两独立的并且 $\sum_{n=1}^{\infty} P(A_n) = \infty$, 那么当 $n \rightarrow \infty$ 时有

$$\frac{\sum_{m=1}^n 1_{A_m}}{\sum_{m=1}^n P(A_m)} \xrightarrow{a.s.} 1$$



证明 设 $X_m = 1_{A_m}$, $S_n = X_1 + \cdots + X_n$. 因为 A_n 是两两独立的, 故而 X_m 是不相关的, 于是

$$\text{Var}(S_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

并且 $\text{Var}(X_m) \leq E(X_m^2) = E(X_m)$. 又因为 $X_m \in \{0, 1\}$, 于是 $\text{Var}(S_n) \leq E(S_n)$, 再由 Chebyshev 不等式可知当 $n \rightarrow \infty$ 时有

$$P(|S_n - E(S_n)| > \delta E(S_n)) \leq \text{Var}(S_n)/[\delta E(S_n)]^2 \leq [\delta^2 E(S_n)]^{-1} \rightarrow 0 \quad (2.5)$$

其中 $E(S_n) \rightarrow \infty$ 是由 $\sum_{n=1}^{\infty} P(A_n) = \infty$ 得到的.

式 (2.5) 表明 $S_n/E(S_n)$ 依概率收敛于 1, 为了得到几乎必然收敛, 还需要取子列. 令 $n_k = \inf\{n : E(S_n) \geq k^2\}$, $T_k = S_{n_k}$, 再注意到 $E(X_m) \leq 1$ 意味着 $k^2 \leq E(T_k) \leq k^2 + 1$. 在式 (2.5) 中用 n_k 替代 n , 于是

$$P(|T_k - E(T_k)| > \delta E(T_k)) \leq \delta^{-2} k^{-2}$$

从而 $\sum_{k=1}^{\infty} P(|T_k - E(T_k)| > \delta E(T_k)) < \infty$, 根据 Borel-Cantelli 第一引理可知

$$P(|T_k - E(T_k)| > \delta E(T_k) \text{ i.o.}) = 0$$

因为 δ 是任意的, 所以 $T_k/E(T_k) \xrightarrow{\text{a.s.}} 1$.

为了证明 $S_n/E(S_n) \xrightarrow{\text{a.s.}} 1$, 选取 ω 使得 $T_k(\omega)/E(T_k) \rightarrow 1$, 设 $n_k \leq n \leq n_{k+1}$, 于是

$$\frac{T_k(\omega)}{E(T_{k+1})} \leq \frac{S_n(\omega)}{E(S_n)} \leq \frac{T_{k+1}(\omega)}{E(T_k)}$$

现在需要证明上式两端趋近于 1. 将上式改写为

$$\frac{E(T_k)}{E(T_{k+1})} \cdot \frac{T_k(\omega)}{E(T_k)} \leq \frac{S_n(\omega)}{E(S_n)} \leq \frac{T_{k+1}(\omega)}{E(T_{k+1})} \cdot \frac{E(T_{k+1})}{E(T_k)}$$

因此只需证明当 $k \rightarrow \infty$ 时有 $E(T_{k+1})/E(T_k) \rightarrow 1$ 即可. 注意到

$$k^2 \leq E(T_k) \leq E(T_{k+1}) \leq (k+1)^2 + 1$$

以及

$$\frac{(k+1)^2 + 1}{k^2} = 1 + \frac{2}{k} + \frac{2}{k^2} \rightarrow 1$$

因此结论成立.

示例 2.13 设 $\alpha > 2$ 为实数, $p, q \in \mathbb{N}$, 定义

$$E = \left\{ x \in [0, 1] : \left| x - \frac{p}{q} \right| < q^{-\alpha}, \text{ i.o.} \right\}$$

那么 $m(E) = 0$.

证明 定义集合

$$E_{p,q} = \{x \in [0, 1] : |x - p/q| < q^{-\alpha}\}$$

注意到 $m(E_{p,q}) = 2/q^\alpha$, $0 \leq p \leq q$, 于是

$$\sum_{p,q} m(E_{p,q}) \leq \sum_{q=1}^{\infty} \frac{2(q+1)}{q^\alpha} \leq \sum_{q=1}^{\infty} \frac{4}{q^{\alpha-1}} < \infty$$

根据 Borel-Cantelli 第一引理可知 $m(E) = 0$.

示例 2.14 记录值 设 X_1, X_2, \dots 是 i.i.d. 随机变量, 并且具有连续的累计分布函数 $F(x)$, 定义

$A_k = \{X_k > \sup_{j < k} X_j\}$, 那么 $P(A_k) = 1/k$.

证明 因为 F 是连续的, 故而对于任意 $j \neq k$ 都有 $P(X_j = X_k) = 0$. 再令 $Y_1^n > Y_2^n > \cdots > Y_n^n$ 为 X_1, X_2, \dots, X_n 的降序排列, 然后通过函数

$$\pi_n(i) = j, \quad X_i = Y_n^j$$

定义随机排列 $\{1, \dots, n\}$. 由于 (X_1, \dots, X_n) 的分布不受排序的影响, 故而 π_n 在 $n!$ 种排序的每一种上都服从均匀分布, 因此 $P(A_n) = P[\pi_n(n) = 1] = 1/n$.

注 设 $R_n = \sum_{m=1}^n 1_{A_m}$ 为 n 时刻的记录数, 根据 $\sum_{m=1}^n 1/m \sim \log n$ 和定理 2.24 可知, 当 $n \rightarrow \infty$ 时有

$$R_n / \log n \xrightarrow{a.s.} 1$$

示例 2.15 设 A_n 是一列独立事件且 $P(A_n) < 1$ 对所有的 n 成立, 那么 $P(\bigcup_{n=1}^{\infty} A_n) = 1$ 意味着 $\sum_{n=1}^{\infty} P(A_n) = \infty$.

证明 根据条件可知 $P(\bigcup_{n=1}^{\infty} A_n^c) = 0$, 又根据独立性得到

$$P\left(\bigcap_{n=1}^{\infty} A_n^c\right) = \prod_{n=1}^{\infty} [1 - P(A_n)] = 0$$

因为 $P(A_n) < 1$, 故而对于 $m = 2, 3, \dots$ 都有 $\prod_{n=1}^{m-1} [1 - P(A_n)] > 0$, 因此对于一切 $m \geq 1$ 都有

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0$$

于是

$$P(\limsup A_n^c) \leq \sum_{m=1}^{\infty} P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0$$

由 Borel-Cantelli 第一引理可知结论成立.

示例 2.16 设 X_1, X_2, \dots 为独立随机变量, $P(X_n = 1) = p_n$ 且 $P(X_n = 0) = 1 - p_n$, 那么当 $n \rightarrow \infty$ 时

- $X_n \xrightarrow{P} 0$ 当且仅当 $p_n \rightarrow 0$.
- $X_n \xrightarrow{a.s.} 0$ 当且仅当 $\sum_n p_n < \infty$.

证明 首先任取 $0 < \epsilon < 1$, 于是 $P(|X_n| > \epsilon) = p_n$, 显然当且仅当 $p_n \rightarrow 0$ 时有 $X_n \xrightarrow{P} 0$. 再定义事件 $A_n = \{|X_n| > \epsilon\}$, 由 Borel-Cantelli 第一引理可知

$$\sum_n p_n < \infty \Rightarrow P(A_n \text{ i.o.}) = 0$$

反向不等式可由独立性推得.

示例 2.17 设 X_1, X_2, \dots 为 i.i.d. 随机变量, $P(X_i > x) = e^{-x}$, 那么

$$\limsup_{n \rightarrow \infty} X_n / \log n = 1 \quad \text{a.s.}$$

证明 任取 $\epsilon > 0$, 首先注意到 $P(X_n > (1 + \epsilon) \log n) = n^{-(1+\epsilon)}$, 于是

$$\sum_{n=1}^{\infty} P(X_n > (1 + \epsilon) \log n) = \sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon}} < \infty$$

根据 Borel-Cantelli 第一引理可知 $P(X_n/\log n > 1 + \epsilon \text{ i.o.}) = 0$. 另一方面, 由于各个 X_n 是独立的, 并且

$$\sum_{n=1}^{\infty} P(X_n > \log n) = \sum_{n=1}^{\infty} \frac{1}{\log n} = \infty$$

根据 Borel-Cantelli 第二引理可知 $P(X_n/\log n > 1 \text{ i.o.}) = 1$.

2.4 强大数定律

本节主要是在更弱的条件下, 将 Khinchin 弱大数定律的结论强化为几乎必然收敛, 得到 Etemadi 强大数定律 (Strong Law of Large Numbers, SLLN), 它是 Kolmogorov 强大数定律的推广.

定理 2.25 (Etemadi 强大数定律)

设 X_1, X_2, \dots 是两两独立的同分布随机变量, 并且满足 $E|X_i| < \infty$, 令 $\mu = E(X_i)$ 和 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时有 $S_n/n - \mu \xrightarrow{a.s.} 0$.



证明 由于 X_1, X_2, \dots 是两两独立且同分布的, 并且 $E(X_1) < \infty$, 故而非负序列 $\{X_n^+\}_{n \geq 1}$ 和 $\{X_n^-\}_{n \geq 1}$ 中的元素也是如此, 同样满足 $E(X_1^+) < \infty$ 和 $E(X_1^-) < \infty$. 由于

$$\frac{S_n}{n} = \frac{1}{n} \left[\sum_{i=1}^n X_i^+ - \sum_{i=1}^n X_i^- \right]$$

因此只需证明非负情形下的 X_1, X_2, \dots 有 $S_n/n - \mu \xrightarrow{a.s.} 0$.

假设对于每个 $i \geq 1$ 都有 $X_i \geq 0$, 定义 $Y_i = X_i 1_{(X_i \leq i)}$, 于是

$$\begin{aligned} \sum_{i=1}^{\infty} P(X_i \neq Y_i) &= \sum_{i=1}^{\infty} P(X_i > i) \\ &= \sum_{i=1}^{\infty} P(X_1 > i) \leq \sum_{i=1}^{\infty} \int_{i-1}^i P(X_1 > t) dt \\ &= \int_0^{\infty} P(X_1 > t) dt = E(X_1) < \infty \end{aligned}$$

根据 Borel-Cantelli 第一引理可知 $P(X_i \neq Y_i \text{ i.o.}) = 0$, 这表明除了有限多个 i 以外都有 $X_i = Y_i$, 设 $T_n = Y_1 + \dots + Y_n$, 现在只需证明 $T_n/n - \mu \xrightarrow{a.s.} 0$.

注意到 $E(Y_i) = E[X_i 1_{(X_i \leq i)}] = E[X_1 1_{(X_1 \leq i)}]$, 根据 MCT 可知当 $n \rightarrow \infty$ 时有 $E(Y_i) \rightarrow E(X_1)$, 故而此时

$$E(T_n/n) \rightarrow E(X_1) \quad (2.6)$$

假设对于每个固定的 $1 < \rho < \infty$, 当 $k \rightarrow \infty$ 时有

$$T_{n_k}/n_k \xrightarrow{a.s.} E(X_1) \quad (2.7)$$

其中 $n_k = \lfloor \rho^k \rfloor$, $k \in \mathbb{N}$. 由于每个 Y_i 都是非负的, 对于任意满足 $\rho^k \leq n < \rho^{k+1}$ 的 n 和 k 可以

得到

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n_k} &\leq \frac{T_n}{n} \leq \frac{1}{n} \sum_{i=1}^{n_{k+1}} Y_i \\ \Rightarrow \frac{n_k}{n} \cdot \frac{T_{n_k}}{n_k} &\leq \frac{T_n}{n} \leq \frac{n_{k+1}}{n} \cdot \frac{T_{n_{k+1}}}{n_{k+1}} \\ \Rightarrow \rho^{-1} \frac{T_{n_k}}{n_k} &\leq \frac{T_n}{n} \leq \rho \frac{T_{n_{k+1}}}{n_{k+1}} \end{aligned}$$

根据式 (2.7) 可知对于一切 $1 < \rho < \infty$ 都有

$$\rho^{-1} E(X_1) \leq \liminf_{n \rightarrow \infty} T_n/n \leq \limsup_{n \rightarrow \infty} T_n/n \leq \rho E(X_1) \quad \text{a.s.}$$

由于 $\rho > 1$ 是任意的, 所以式 (2.6) 成立.

现在只需证明之前假设的条件 (2.7) 成立, 又因为式 (2.6) 成立, 因此只需证明当 $k \rightarrow \infty$ 时有

$$T_{n_k}/n_k - E(T_{n_k}/n_k) \xrightarrow{a.s.} 0 \quad (2.8)$$

根据两两独立性和 Chebyshev 不等式可知, 对于任意的 $\epsilon > 0$ 都有

$$\begin{aligned} P(|T_{n_k} - E(T_{n_k})| > \epsilon n_k) &\leq \epsilon^{-2} \text{Var}(T_{n_k}) = \epsilon^{-2} n_k^{-2} \sum_{i=1}^{n_k} \text{Var}(Y_i) \\ &\leq \epsilon^{-2} n_k^{-2} \sum_{i=1}^{n_k} E(Y_i^2) \end{aligned}$$

因此

$$\begin{aligned} \sum_{k=1}^{\infty} P(|T_{n_k} - E(T_{n_k})| > \epsilon n_k) &\leq \epsilon^{-2} \sum_{k=1}^{\infty} n_k^{-2} \sum_{i=1}^{n_k} E(Y_i^2) \\ &= \epsilon^{-2} \sum_{i=1}^{\infty} E(Y_i^2) \left(\sum_{k: n_k \geq i} n_k^{-2} \right) \end{aligned}$$

由于 $n_k = \lfloor \rho^k \rfloor > \rho^{k-1}$ 对一切 $\rho > 1$ 成立, 故而存在 $C_1 > 0$ 使得

$$\sum_{k: n_k \geq i} n_k^{-2} \leq \sum_{k: \rho^{k-1} \geq i} \rho^{-2(k-1)} \leq i^{-2} C_1$$

又因为每个 X_i 是同分布的, 存在 $C_2 > 0$ 使得

$$\begin{aligned} \sum_{i=1}^{\infty} i^{-2} E(Y_i^2) &= \sum_{i=1}^{\infty} i^{-2} E(X_1^2) 1_{(0 \leq X_1 \leq i)} \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^i i^{-2} E(X_1^2) 1_{(j-1 < X_1 \leq j)} \\ &= \sum_{j=1}^{\infty} E(X_1^2) 1_{(j-1 < X_1 \leq j)} \sum_{i=j}^{\infty} i^{-2} \\ &\leq \sum_{j=1}^{\infty} [j E(X_1) 1_{(j-1 \leq X_1 < j)}] \cdot C_2 j^{-1} = C_2 E(X_1) < \infty \end{aligned}$$

从而对于任意 $\epsilon > 0$ 都有

$$\sum_{k=1}^{\infty} P(|T_{n_k} - E(T_{n_k})| > \epsilon n_k) < \infty$$

由 Borel-Cantelli 第一引理可知式 (2.7) 成立.

定理 2.26

设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i^+) = \infty$ 且 $E(X_i^-) < \infty$, 如果 $S_n = X_1 + \dots + X_n$, 那么 $S_n/n \xrightarrow{a.s.} \infty$.



证明 令 $M > 0$, 以及 $X_i^M = X_i \wedge M$, 可知 X_i^M 是 i.i.d. 的并且 $E|X_i^M| < \infty$, 再令 $S_n^M = X_1^M + \dots + X_n^M$, 强大数定律表明 $S_n^M/n \xrightarrow{a.s.} E(X_i^M)$. 因为 $X_i \geq X_i^M$, 于是

$$\liminf_{n \rightarrow \infty} S_n/n \geq \lim_{n \rightarrow \infty} S_n^M/n = E(X_i^M)$$

再由 MCT 可知当 $M \uparrow \infty$ 时有 $E(X_i^M)^+ \uparrow E(X_i^+) = \infty$, 进一步有 $E(X_i^M) = E(X_i^M)^+ - E(X_i^M)^- \uparrow \infty$, 以及 $\liminf S_n/n \geq \infty$, 这就得到了希望的结果.

示例 2.18 经验分布函数 设 X_1, X_2, \dots 为 i.i.d. 随机变量, F 是每个随机变量的 c.d.f., 现在定义

$$F_n(x) \equiv n^{-1} \sum_{j=1}^n 1_{(X_j \leq x)}, \quad x \in \mathbb{R}$$

称为 $\{X_1, X_2, \dots\}$ 的经验分布函数.

定理 2.27 (Glivenko-Cantelli 定理)

设以下给定的符号与上例中的一致, 则当 $n \rightarrow \infty$ 时有

$$\tilde{\Delta}_n \equiv \sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$



证明 固定 $x \in \mathbb{R}$, 令 $Y_n = 1_{(X_n \leq x)}$. 因为 $\{Y_n\}_{n \geq 1}$ 是 i.i.d. 随机变量构成的序列且 $E(Y_n) = P(X \leq x) = F(x)$, 根据 SLLN 可知 $F_n(x) = n^{-1} \sum_{j=1}^n Y_j \xrightarrow{a.s.} F(x)$.

现在考虑 F 存在跳跃间断点的情形, 固定 $x \in \mathbb{R}$ 并令 $Z_n = 1_{(X_n < x)}$, 由于每个 Z_n 也是 i.i.d. 且

$$E(Z_n) = P(X_n < x) = F(x-) = \lim_{y \uparrow x} F(y)$$

根据 SLLN 可知

$$F_n(x-) = n^{-1} \sum_{j=1}^n Z_j \xrightarrow{a.s.} F(x-)$$

对于 $1 \leq j \leq k-1$, 令 $x_{j,k} = \inf\{y : F(y) \geq j/k\}$, $F_n(x)$ 和 $F_n(x-)$ 的点态收敛表明, 存在正整数 N_k , 使得当 $n > n_k$ 时就有

$$|F_n(x_{j,k}) - F(x_{j,k})| < k^{-1}, \quad |F_n(x_{j,k}-) - F(x_{j,k}-)| < k^{-1}$$

这里规定 $x_{0,k} = -\infty$ 且 $x_{k,k} = \infty$, 显然以上两个不等式成立. 如果 $x \in (x_{j-1,k}, x_{j,k})$, $1 \leq j \leq$

k 且 $n > n_k$, 注意到 $F(x_{j,k-}) - F(x_{j-1,k}) \leq k^{-1}$, 于是

$$F_n(x) \leq F_n(x_{j,k-}) \leq F(x_{j,k-}) + k^{-1} \leq F(x_{j-1,k}) + 2k^{-1} \leq F(x) + 2k^{-1}$$

$$F_n(x) \geq F_n(x_{j-1,k}) \geq F(x_{j-1,k}) - k^{-1} \geq F(x_{j,k-}) - 2k^{-1} \geq F(x) - 2k^{-1}$$

也即 $\sup_x |F_n(x) - F(x)| \leq 2k^{-1}$, 结论成立.

示例 2.19 Shannon 定理 设 $X_1, X_2, \dots \{1, \dots, r\}$ 是独立的, 并且对 $1 \leq k \leq r$ 都有 $P(X_i = k) = p(k) > 0$. 这里可以将 $1, \dots, r$ 视作一个字母表, X_1, X_2, \dots 为信息源生成的有序字母. 令

$$\pi_n(\omega) = P(X_1(\omega)) \cdots P(X_n(\omega))$$

为前 n 次试验中观测到实现值的概率, 根据 SLLN 可知

$$-n^{-1} \log \pi_n(\omega) \rightarrow H = -\sum_{k=1}^r p(k) \log p(k) \quad \text{a.s.}$$

其中 H 称为信息熵 (information entropy), 用来衡量信息源的随机程度.

2.5 随机级数的收敛

定义 2.7

设 X_1, X_2, \dots 是定义在概率空间 (Ω, \mathcal{F}, P) 上的随机变量, $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$ 是由 n 时刻后的全体 X_m 生成的最小 σ -代数, 称

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}'_n$$

为尾 σ -代数 (tail σ -algebra), 任意 $A \in \mathcal{T}$ 为尾事件.



直观上看, \mathcal{F}'_n 相当于是 n 时刻后的事件, $A \in \mathcal{T}$ 当且仅当改变有限个随机变量的取值不影响事件 A 的发生.

示例 2.20 设 $B_n \in \mathcal{R}$, 那么 $\{X_n \in B_n \text{ i.o.}\} \in \mathcal{T}$.

示例 2.21 设 $S_n = X_1 + \dots + X_n$, 那么

- $\{\lim_{n \rightarrow \infty} S_n \in (-\infty, \infty)\} \in \mathcal{T}$.
- $\{\limsup_{n \rightarrow \infty} S_n > 0\} \notin \mathcal{T}$.
- $\{\limsup_{n \rightarrow \infty} S_n/c_n > x\}$, 如果 $c_n \rightarrow \infty$.

定理 2.28 (Kolmogorov 0-1 律)

如果 X_1, X_2, \dots 是独立随机变量, 并且 $A \in \mathcal{T}$, 那么 $P(A) \in \{0, 1\}$.



证明 假设事件 A 关于它自身是独立的, 也即 $P(A \cap A) = P(A)P(A)$, 那么可以得出 $P(A) = 0$ 或 1, 因此只需证明 \mathcal{T} 关于自身独立即可.

对于任意 $n \geq 1$, 定义 σ -代数 $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$, $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \dots)$, 根据独立性可知对一切 $n > k$, \mathcal{F}_k 和 \mathcal{T}_n 是独立的. 因为 $\mathcal{T} \subset \mathcal{T}_n$, 所以对于一切 $k \geq 1$, \mathcal{F}_k 和 \mathcal{T} 是独立的. 再令 $\mathcal{A} = \bigcup_{k=1}^{\infty} \mathcal{F}_k$, 容易验证它是一个代数, 根据定理 2.4 可知 $\sigma(\mathcal{A})$ 和 \mathcal{T} 独立, 从而

$\sigma(X_1, X_2, \dots) = \sigma(\mathcal{A})$ 和 \mathcal{T} 独立, 也即 \mathcal{T} 独立于它自身.

Kolmogorov 0-1 律表明, 如果事件 A_1, A_2, \dots 是独立的, 那么 $P(A_n \text{ i.o.}) \in \{0, 1\}$. 在示例 2.21 中, $P(\lim S_n \in (-\infty, \infty)) = 0$ 或 1, 为了得到特殊情形下该事件的概率为 1, 需要用到一些新的定理. 首先是 Kolmogorov 极大不等式, 它是 Chebyshev 不等式的推广.

定理 2.29 (Kolmogorov 极大不等式)

设 X_1, \dots, X_n 为独立随机变量, 并且满足 $E(X_i) = 0$ 及 $\text{Var}(X_i) < \infty$, 如果 $S_n = X_1 + \dots + X_n$, 那么

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq x^{-2} \text{Var}(S_n)$$



证明 令 $A = \{\omega : \max_{1 \leq i \leq n} |S_i| \geq x\}$, 以及

$$A_1 = \{\omega : |S_1| \geq x\}$$

$$A_k = \{\omega : |S_1| < x, \dots, |S_{k-1}| < x, |S_k| \geq x\}$$

其中 $k = 2, \dots, n$. 显然 A_1, \dots, A_n 互不相交, $A = \bigcup_{k=1}^n A_k$ 且 $P(A) = \sum_{k=1}^n P(A_k)$. 因为 $E(X_k) = 0$ 对一切 $k \geq 1$ 成立, 所以

$$\begin{aligned} \text{Var}(S_n) &= E(S_n^2) \geq E(S_n^2 1_A) = \sum_{k=1}^n E(S_n^2 1_{A_k}) \\ &= \sum_{k=1}^n E\{[(S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k] 1_{A_k}\} \\ &\geq \sum_{k=1}^n E(S_k^2 1_{A_k}) + 2 \sum_{k=1}^{n-1} E[(S_n - S_k)S_k 1_{A_k}] \end{aligned} \quad (2.9)$$

注意到 X_1, \dots, X_n 是独立的, $S_n - S_k = \sum_{j=k+1}^n X_j$, 并且对一切 $1 \leq k \leq n-1$, $S_k 1_{A_k}$ 也是独立的, 因此

$$E[(S_n - S_k)S_k 1_{A_k}] = E(S_n - S_k)E(S_k 1_{A_k}) = 0$$

又因为在 A_k 上都有 $S_k^2 \geq x^2$, 根据式 (2.9) 可知

$$\text{Var}(S_n) \geq \sum_{k=1}^n x^2 P(A_k) = x^2 P(A)$$

由此证得结论.

示例 2.22 设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = 0$ 且 $\text{Var}(X_i) = C < \infty$, 设 $S_n = X_1 + \dots + X_n$, $p > 1/2$, 那么当 $n \rightarrow \infty$ 时有 $S_n/n^p \xrightarrow{a.s.} 0$.

证明 选取 α 使得 $\alpha(2p-1) > 1$, 令 $n = m^\alpha$, 根据 Kolmogorov 极大不等式可知

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq n^p\right) \leq \frac{C}{m^{\alpha(2p-1)}}$$

由于 $\alpha(2p-1) > 1$, 故而上式右端的级数是收敛的, 根据 Borel-Cantelli 第一引理可知当 n 充分

大时有 $\max_{1 \leq k \leq n} |S_k/n^p| < 1$, 也即

$$\limsup_{n \rightarrow \infty} |S_n/n^p| < 1 \quad \text{a.s.}$$

任意选取 $q > p > 1/2$, 此时必有 $\limsup_{n \rightarrow \infty} S_n/n^q = 0$ a.s., 因为 $q > 1/2$ 也是任意的, 所以结论成立.

注 只用 Chebyshev 不等式无法完成证明, 因为没有给出当 n 不是 m^α 这一形式时的信息.

定理 2.30 (Khinchin-Kolmogorov 级数定理)

设 X_1, X_2, \dots 为独立随机变量并且 $E(X_n) = 0$, 如果级数 $\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty$, 那么 $\sum_{n=1}^{\infty} X_n(\omega)$ 几乎必然收敛.



证明 令 $S_N = \sum_{n=1}^N X_n$, 根据 Kolmogorov 极大不等式可知

$$P\left(\max_{M \leq m \leq N} |S_m - S_M|\right) \leq \epsilon^{-2} \text{Var}(S_N - S_M) = \epsilon^{-2} \sum_{n=M+1}^N \text{Var}(X_n)$$

在上式中令 $N \rightarrow \infty$ 得

$$\lim_{M \rightarrow \infty} P\left(\sup_{m \geq M} |S_m - S_M| > \epsilon\right) \leq \epsilon^{-2} \lim_{M \rightarrow \infty} \sum_{n=M+1}^N \text{Var}(X_n) = 0$$

再令 $w_M = \sup_{n, m \geq M} |S_m - S_n|$, 当 M 递增时 w_M 递减, 当 $M \rightarrow \infty$ 时有

$$P(w_M > 2\epsilon) \leq P\left(\sup_{m \geq M} |S_m - S_M| > \epsilon\right) \rightarrow 0$$

并且此时以概率 1 有 $w_M \downarrow 0$, 这意味着 $S_n(\omega)$ 是一个 Cauchy 序列, 因此 $\lim S_n(\omega)$ 存在.

定理 2.31 (Kolmogorov 三级数定理)

设 X_1, X_2, \dots 是独立随机变量, $A > 0$ 且 $Y_i = X_i 1_{(|X_i| \leq A)}$, 级数 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛, 当且仅当以下条件成立

- (1) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$;
- (2) $\sum_{n=1}^{\infty} E(Y_n)$ 收敛;
- (3) $\sum_{n=1}^{\infty} \text{Var}(Y_n) < \infty$.



证明 必要性的证明放在下一章, 这里只证明充分性.

根据条件 (1) 和 Borel-Cantelli 第一引理可知 $P(X_i \neq Y_i \text{ i.o.}) = P(|X_i| > A \text{ i.o.}) = 0$, 于是只需证明 $\sum_{n=1}^{\infty} Y_n$ 几乎必然收敛即可. 根据 Khinchin-Kolmogorov 级数定理和条件 (3) 可知 $\sum_{n=1}^{\infty} [Y_n - E(Y_n)]$ 几乎必然收敛, 最后由条件 (2) 可知结论成立.

示例 2.23 随机调和级数 考虑随机变量 $X_n \in \{-1, 1\}$ 并且

$$P(X_n = -1) = P(X_n = 1) = \frac{1}{2}$$

那么级数 $\sum_{n=1}^{\infty} X_n/n$ 收敛.

证明 首先令 $A = 1$, 那么 $Y_n = X_n 1_{(|X_n| \leq n)}/n = X_n/n$, 容易验证 $\sum_{n=1}^{\infty} P(|X_n| > n) = 0$,

$\sum_{n=1}^{\infty} E(Y_n) = 0$ 以及

$$\sum_{n=1}^{\infty} \text{Var}(Y_n) = E \left(\sum_{m=1}^{\infty} \frac{X_m}{m} \sum_{n=1}^{\infty} \frac{X_n}{n} \right) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi}{6}$$

根据 Kolmogorov 三级数定理可知 $\sum_{n=1}^{\infty} X_n/n$ 收敛.

示例 2.24 设 $X_n \geq 0$ 是独立的, 那么以下说法是等价的

- (1) $\sum_{n=1}^{\infty} X_n < \infty$;
- (2) $\sum_{n=1}^{\infty} P(X_n > 1) + E[X_n 1_{(X_n \leq 1)}] < \infty$;
- (3) $\sum_{n=1}^{\infty} E[X_n/(1 + X_n)] < \infty$.

证明 (1) \iff (2): 根据 $X_n \geq 0$ 和 Kolmogorov 三级数定理可知 $\sum P(X_n > 1) < \infty$, 取 $A = 1$, 从而 $\sum E[X_n 1_{(|X_n| \leq 1)}] < \infty$, 由此可知 (2) 成立.

另一方面, 设 $Y_n = X_n 1_{(|X_n| \leq 1)}$, 由 (2) 可知 $\sum P(|X_n| > 1) < \infty$, $\sum E(Y_n) < \infty$, 以及

$$\text{Var}(Y_n) \leq E(Y_n^2) \leq E(Y_n)$$

也即 $\sum \text{Var}(Y_n) < \infty$, 根据 Kolmogorov 三级数定理知 (1) 成立.

(2) \iff (3): 注意到

$$\frac{X_n}{1 + X_n} \leq X_n 1_{(X_n \leq 1)} + 1_{(X_n > 1)} \leq \frac{2X_n}{1 + X_n}$$

两端取期望即可证得结论.

示例 2.25 定义函数

$$\psi(x) = \begin{cases} x^2, & |x| \leq 1 \\ |x|, & |x| > 1 \end{cases}$$

如果 X_1, X_2, \dots 是独立的且 $\sum_{n=1}^{\infty} E[\psi(X_n)] < \infty$, 那么 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛.

证明 根据 $\sum_{n=1}^{\infty} E[\psi(X_n)] < \infty$ 可知

$$\sum_{n=1}^{\infty} E[X_n^2 1_{(|X_n| \leq 1)}] + E[|X_n| 1_{(|X_n| > 1)}]$$

从而

$$\sum_{n=1}^{\infty} P(|X_n| > 1) \leq \sum_{n=1}^{\infty} E[|X_n| 1_{(|X_n| > 1)}] < \infty$$

再设 $Y_n = X_n 1_{(|X_n| \leq 1)}$, 于是

$$\sum_{n=1}^{\infty} \text{Var}(Y_n) \leq \sum_{n=1}^{\infty} E[X_n^2 1_{(|X_n| \leq 1)}] < \infty$$

由于 $E(X_n) = 0$, 所以 $E(Y_n) = -E[X_n 1_{(|X_n| > 1)}]$, 从而

$$\sum_{n=1}^{\infty} |E(Y_n)| \leq \sum_{n=1}^{\infty} E(|X_n| 1_{(|X_n| \leq 1)}) < \infty$$

根据 Kolmogorov 三级数定理可知结论成立.

定理 2.32 (Kronecker 引理)

设 $0 < a_n \uparrow \infty$, 并且 $\sum_{n=1}^{\infty} x_n/a_n$ 收敛, 那么当 $n \rightarrow \infty$ 时有

$$a_n^{-1} \sum_{m=1}^n x_m \rightarrow 0$$



证明 令 $a_0 = b_0 = 0$, 对 $m \geq 1$ 定义 $b_m = \sum_{k=1}^m x_k/a_k$, 于是 $x_m = a_m(b_m - b_{m-1})$, 并且

$$\begin{aligned} a_n^{-1} \sum_{m=1}^n x_m &= a_n^{-1} \left(\sum_{m=1}^n a_m b_m - \sum_{m=1}^n a_m b_{m-1} \right) \\ &= a_n^{-1} \left(a_n b_n + \sum_{m=2}^n a_{m-1} b_{m-1} - \sum_{m=1}^n a_m b_{m-1} \right) \\ &= b_n - \sum_{m=1}^n \frac{a_m - a_{m-1}}{a_n} b_{m-1} \end{aligned}$$

根据假设, 当 $n \rightarrow \infty$ 时 $b_n \rightarrow b_\infty$. 现在考虑上式最后一项和式, 任取 $\epsilon > 0$, 令 $B = \sup |b_n|$, 选取正整数 M , 使得当 $m > M$ 时就有 $|b_m - b_\infty| < \epsilon/2$, 再选取正整数 $N > M$, 使得当 $n > N$ 时就有 $a_M/a_n < \epsilon/4B$, 于是当 $n > N$ 时

$$\begin{aligned} \left| \sum_{m=1}^n \frac{a_m - a_{m-1}}{a_n} b_{m-1} - b_\infty \right| &\leq \sum_{m=1}^n \frac{a_m - a_{m-1}}{a_n} |b_{m-1} - b_\infty| \\ &\leq \frac{a_M}{a_n} \cdot 2B + \frac{a_n - a_M}{a_n} \cdot \frac{\epsilon}{2} < \epsilon \end{aligned}$$

因为 ϵ 是任意的, 结论成立.

示例 2.26 设 X_1, X_2, \dots 为独立随机变量, $E(X_n) = 0$, $\text{Var}(X_n) = \sigma_n^2$, 如果 $\sum_n \sigma_n^2/n^2 < \infty$, 那么 $n^{-1} \sum_{m=1}^n X_m \xrightarrow{a.s.} 0$.

证明 定义 $Y_n = X_n/n$, 显然有 $E(Y_n) = 0$, $\sum_n \text{Var}(Y_n) < \infty$, 根据 Khinchin-Kolmogorov 级数定理可知 $\sum_n X_n/n$ 几乎必然收敛, 最后由 Kronecker 引理知 $n^{-1} \sum_{m=1}^n X_m \xrightarrow{a.s.} 0$.

引理 2.3

如果 $y \geq 0$, 那么 $2y \sum_{k>y} k^{-2} \leq 4$.



证明 首先注意到如果 $m \geq 2$, 那么

$$\sum_{k \geq m} k^{-2} \leq \int_{m-1}^{\infty} x^{-2} dx = (m-1)^{-1}$$

当 $y \geq 1$ 时, 上述和式从 $k = [y] + 1 \geq 2$ 开始计算, 于是

$$2y \sum_{k>y} k^{-2} \leq 2y/[y] \leq 4$$

当 $0 \leq y < 1$ 时有

$$2y \sum_{k>y} k^{-2} \leq 2 \left(1 + \sum_{k=2}^{\infty} k^{-2} \right) \leq 4$$

由此证得结论.

定理 2.33 (Kolmogorov 强大数定律)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, 满足 $E|X_i| < \infty$, 令 $\mu = E(X_i)$ 和 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时有 $S_n/n - \mu \xrightarrow{a.s.} 0$.



证明 设 $Y_k = X_k 1_{(|X_k| \leq k)}$, $T_n = \sum_{i=1}^n Y_i$, 只需证明当 $n \rightarrow \infty$ 时有 $T_n/n \xrightarrow{a.s.} \mu$. 再设 $Z_k = Y_k - E(Y_k)$, 于是 $E(Z_k) = 0$, 这里先证明 $\sum_{k=1}^{\infty} \text{Var}(Z_k)/k^2 < \infty$.

首先注意到

$$\begin{aligned} \text{Var}(Y_k) &\leq E(Y_k^2) = \int_0^{\infty} 2yP(|Y_k| > y) dy \\ &\leq \int_0^k 2yP(|X_1| > y) dy \end{aligned}$$

由于每一项非负, 根据 Fubini 定理可知

$$\begin{aligned} \sum_{k=1}^{\infty} E(Y_k^2)/k^2 &\leq \sum_{k=1}^{\infty} k^{-2} \int_0^{\infty} 1_{(y < k)} 2yP(|X_1| > y) dy \\ &= \int_0^{\infty} \left(\sum_{k=1}^{\infty} k^{-2} 1_{(y < k)} \right) 2yP(|X_1| > y) dy \end{aligned}$$

由于 $E|X_1| = \int_0^{\infty} P(|X_1| > y) dy$, 并且引理 2.3 表明

$$2y \sum_{k > y} k^{-2} \leq 4$$

对一切 $y \geq 0$ 成立, 因此 $\sum_{k=1}^{\infty} \text{Var}(Y_k)/k^2$ 存在且有限, 也即

$$\sum_{k=1}^{\infty} \frac{\text{Var}(Z_k)}{k^2} \leq \sum_{k=1}^{\infty} \frac{E(Y_k^2)}{k^2} < \infty$$

根据 Khinchin-Kolmogorov 级数定理可知 $\sum_{k=1}^{\infty} Z_k/k$ 几乎必然收敛, 再根据 Kronecker 引理, 当 $n \rightarrow \infty$ 时, $n^{-1} \sum_{k=1}^n [Y_k - E(Y_k)]$ 几乎必然收敛, 也即

$$\frac{T_n}{n} - n^{-1} \sum_{k=1}^n E(Y_k) \rightarrow 0 \quad \text{a.s.}$$

最后根据 DCT 可知, 当 $k \rightarrow \infty$ 时有 $E(Y_k) \rightarrow \mu$, 这也意味着 $n^{-1} \sum_{k=1}^n E(Y_k) \rightarrow \mu$, 因此 $T_n/n \xrightarrow{a.s.} \mu$.

定理 2.34

设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = 0$ 且 $E(X_i^2) = \sigma^2 < \infty$, 令 $S_n = X_1 + \dots + X_n$, 那么对于任意 $\epsilon > 0$ 都有

$$\frac{S_n}{\sqrt{n(\log n)^{\epsilon}}} \xrightarrow{a.s.} 0$$



证明 假设 $a_1 > 0$, 并且对于任意 $n \geq 2$, 令 $a_n = n^{1/2}(\log n)^{1/2+\epsilon}$, 于是

$$\sum_{n=1}^{\infty} \text{Var}(X_n/a_n) = \sigma^2 \left(\frac{1}{a_1^2} + \sum_{n=2}^{\infty} \frac{1}{n(\log n)^{1+2\epsilon}} \right) < \infty$$

根据 Khinchin-Kolmogorov 级数定理可知 $\sum_{n=1}^{\infty} X_n/a_n$ 几乎必然收敛, 最后由 Kronecker 引理即可证得结论.

定理 2.35 (Marcinkiewicz-Zygmund 强大数定律)

设 X_1, X_2, \dots 是 i.i.d. 随机变量, $E(X_1) = 0$ 且 $E|X_1|^p < \infty$, $1 < p < 2$, 令 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时有 $S_n/n^{1/p} \xrightarrow{a.s.} 0$.



证明 设 $Y_k = 1_{(|X_k| \leq k^{1/p})}$, $T_n = Y_1 + \dots + Y_n$, 于是

$$\sum_{k=1}^{\infty} P(Y_k \neq X_k) = \sum_{k=1}^{\infty} P(|X_k|^p > k) \leq E|X_1|^p < \infty$$

根据 Borel-Cantelli 第一引理可知 $P(Y_k \neq X_k \text{ i.o.}) = 0$, 因此只需证明 $T_n/n^{1/p} \xrightarrow{a.s.} 0$ 即可.

根据 $\text{Var}(X_m) \leq E(X_m^2)$, $P(|Y_m| > y) \leq P(|X_1| > y)$, 引理 2.1 及 Fubini 定理, 可以得到

$$\begin{aligned} \sum_{m=1}^{\infty} \text{Var}(Y_m/m^{1/p}) &\leq \sum_{m=1}^{\infty} E(Y_m^2/m^{2/p}) \\ &\leq \sum_{m=1}^{\infty} \sum_{n=1}^m \int_{(n-1)^{1/p}}^{n^{1/p}} \frac{2y}{m^{2/p}} P(|X_1| > y) dy \\ &= \sum_{n=1}^{\infty} \int_{(n-1)^{1/p}}^{n^{1/p}} \sum_{m=n}^{\infty} \frac{2y}{m^{2/p}} P(|X_1| > y) dy \end{aligned}$$

注意到当 $y \in [(n-1)^{1/p}, n^{1/p}]$ 时有

$$\sum_{m=n}^{\infty} m^{-2/p} \leq \frac{p}{2-p} (n-1)^{\frac{p-2}{p}} \leq Cy^{p-2}$$

以及 $E|X_1|^p = \int_0^{\infty} px^{p-1} P(|X_1| > x) dx < \infty$, 因此

$$\sum_{m=1}^{\infty} \text{Var}(Y_m/m^{1/p}) < \infty$$

令 $\mu_m = E(Y_m)$, 根据 Khinchin-Kolmogorov 级数定理和 Kronecker 引理知

$$n^{-1} \sum_{m=1}^{\infty} (Y_m - \mu_m) \rightarrow 0 \quad \text{a.s.}$$

另一方面

$$\begin{aligned} |\mu_m| &\leq E(|X_i|; |X_i| > m^{1/p}) = m^{1/p} E(|X_i|/m^{1/p}; |X_i| > m^{1/p}) \\ &\leq m^{1/p} E[(|X_i|/m^{1/p})^p; |X_i| > m^{1/p}] \\ &\leq m^{-1+1/p} p^{-1} E(|X_i|^p; |X_i| > m^{1/p}) \end{aligned}$$

现在有 $\sum_{m=1}^n m^{-1+1/p} \leq Cn^{1/p}$, 以及 $E(|X_i|^p; |X_i| > m^{1/p}) \rightarrow 0$, 所以 $n^{-1/p} \sum_{m=1}^n \mu_m \rightarrow 0$, 这就得到了希望的结果.

示例 2.27 上述定理的逆命题 设 X_1, X_2, \dots 是 i.i.d. 随机变量, $p > 0$, 令 $S_n = X_1 + \dots + X_n$, 如果 $S_n/n^{1/p} \xrightarrow{a.s.} 0$, 那么 $E|X_1|^p < \infty$.

证明 假设 $E|X_1|^p = \infty$, 于是 $\sum_{n=1}^{\infty} P(|X_1|^p \geq n) = \infty$, 根据 Borel-Cantelli 第二引理可知

$P(|X_1| \geq n^{1/p} \text{ i.o.}) = 1$. 根据条件可知, 对任意的 $\epsilon > 0$ 都有 $P(|S_n| > \epsilon n^{1/p} \text{ i.o.}) = 0$, 注意到 $2 \max\{|S_{n-1}|, |S_n|\} \geq |X_n|$, 这就产生矛盾.

第3章 中心极限定理

3.1 依分布收敛

定义 3.1

设 F_n 为随机变量 X_n 的累积分布, 称随机变量序列 $\{X_n\}_{n \geq 1}$ 依分布收敛 (converge in distribution) 于 X_∞ , 如果对于一切 $x \in \mathcal{C}(F_\infty)$ 都有

$$\lim_{n \rightarrow \infty} F_n(x) = F_\infty(x) \quad (3.1)$$

其中 $\mathcal{C}(F_\infty)$ 是 F_∞ 的连续点构成的集合.



注 依分布收敛的概念不要求每个 X_n 都定义在同一个概率空间上. 事实上, 对于每个 $n \geq 1$, X_n 可以定义在不同的概率空间 $(\Omega_n, \mathcal{F}_n, P_n)$ 上, 并且可以使得 $\{X_n\}_{n \geq 1}$ 依分布收敛于 X_∞ , 而几乎必然收敛与依概率收敛则无法这样定义.

定义 3.2

设 $\{\mu_n\}_{n \geq 1}$ 为 $(\mathbb{R}, \mathcal{R})$ 上的概率测度序列, 称 $\{\mu_n\}_{n \geq 1}$ 依分布收敛于 μ_∞ , 如果 (3.1) 成立, 并且 $F_n(x) \equiv \mu_n((-\infty, x])$, $x \in \mathbb{R}$, 记作 $\mu_n \xrightarrow{d} \mu$.



示例 3.1 对于 $n \geq 1$, 设 $X_n \sim U[0, 1/n]$, 那么 X_n 的累计概率分布为

$$F_n(x) = \begin{cases} 0, & x \leq 0 \\ nx, & 0 < x < 1/n \\ 1, & x \geq 1/n \end{cases}$$

再设 X_∞ 为以概率 1 取值为 0 的退化随机变量, 它的累积概率分布为

$$F_\infty(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

注意到 F_∞ 唯一的不连续点为 $x = 0$, 因此 $\mathcal{C}(F_\infty) = \mathbb{R} \setminus \{0\}$, 故而对任意的 $x \neq 0$, 当 $n \rightarrow \infty$ 时有

$$F_n(x) \rightarrow F_\infty(x)$$

也即 $X_n \xrightarrow{d} X_\infty$.

示例 3.2 生日问题 设 X_1, X_2, \dots 是独立的且在 $\{1, \dots, N\}$ 上服从均匀分布, 定义 $T_N = \min\{n : X_n = X_m, m < n\}$, 那么

$$P(T_N > n) = \prod_{m=2}^n \left(1 - \frac{m-1}{N}\right)$$

当 $N = 365$ 时, 它表示在一个数量为 n 的人群中, 任何两个人的生日均不相同的概率. 易知对

于任意 $x \geq 0$ 都有

$$P(T_N/N^{1/2} > x) \rightarrow \exp(-x^2/2)$$

取 $N = 365, n = 22$, 注意到 $22/\sqrt{365} = 1.1515$, 以及 $(1.1515)^2/2 = 0.6630$, 根据上式计算的概率为

$$P(T_{365} > 22) \approx e^{-0.6630} \approx 0.515$$

只比真实概率 0.524 低不到 2%.

定理 3.1

若 $X_n \xrightarrow{p} X_\infty$, 则 $X_n \xrightarrow{d} X_\infty$.



证明 设 F_n 为 X_n 的 c.d.f., 固定 $x \in \mathcal{C}(F_\infty)$, 于是对于任意 $\epsilon > 0$ 都有

$$\begin{aligned} P(X_n \leq x) &\leq P(X_n \leq x, X_\infty \leq x + \epsilon) + P(X_n \leq x, X_\infty > x + \epsilon) \\ &\leq P(X_n \leq x) + P(X_n - X_\infty \leq x - X_\infty, x - X_\infty < -\epsilon) \\ &\leq P(X_n \leq x) + P(X_n - X_\infty < -\epsilon) + P(X_n - X_\infty > -\epsilon) \\ &\leq P(X_\infty \leq x + \epsilon) + P(|X_n - X_\infty| > \epsilon) \end{aligned} \quad (3.2)$$

类似地有

$$P(X_n \leq x) \geq P(X_\infty \leq x - \epsilon) - P(|X_n - X_\infty| > \epsilon) \quad (3.3)$$

根据式 (3.2) 和 (3.3) 可知

$$F_\infty(x - \epsilon) - P(|X_n - X_\infty| > \epsilon) \leq F_n(x) \leq F_\infty(x + \epsilon) + P(|X_n - X_\infty| > \epsilon)$$

由于 $X_n \xrightarrow{p} X_\infty$, 令 $n \rightarrow \infty$ 可知对一切 $\epsilon \in (0, \infty)$ 都有

$$F_\infty(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F_\infty(x + \epsilon) \quad (3.4)$$

注意到 $x \in \mathcal{C}(F_\infty)$, 故而 $F_\infty(x-) = F_\infty(x+)$, 在式 (3.4) 中令 $\epsilon \downarrow 0$ 即得 $\lim F_n(x) = F_\infty(x)$.

定理 3.2

如果 $X_n \xrightarrow{d} X_\infty$, 并且存在 $c \in \mathbb{R}$ 使得 $P(X_\infty = c) = 1$, 那么 $X_n \xrightarrow{p} c$.



证明 由于 $P(X_\infty = c) = 1$, 故而

$$F_\infty(x) = P(X \leq x) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}$$

又因为 $X_n \xrightarrow{d} c$, 于是对于任意给定的 $\epsilon > 0$ 都有

$$\lim_{n \rightarrow \infty} F_n(c - \epsilon) = 0, \quad \lim_{n \rightarrow \infty} F_n(c + \epsilon) = 1$$

进而有

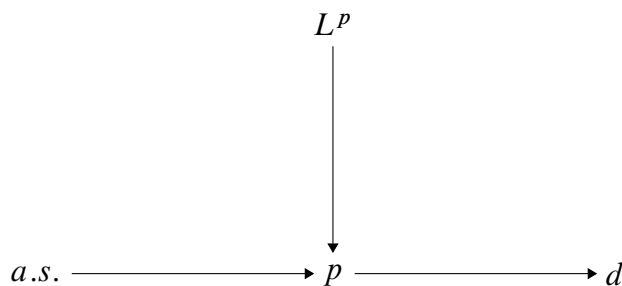
$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) &= \lim_{n \rightarrow \infty} [P(X_n \leq c - \epsilon) + P(X_n > c + \epsilon)] \\
 &= \lim_{n \rightarrow \infty} F_n(c - \epsilon) + \lim_{n \rightarrow \infty} P(X_n > c + \epsilon) \\
 &= 0 + \lim_{n \rightarrow \infty} P(X_n > c + \epsilon) \\
 &\leq 1 - \lim_{n \rightarrow \infty} F_n(c + \epsilon) = 0
 \end{aligned}$$

也即

$$\lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

由此证得结论.

以下示意图描述了几乎必然收敛, 依 L^p 收敛, 依概率收敛和依分布收敛的关系. 可以看出, 依分布收敛是最弱形式的收敛方式, 因此也称弱收敛 (weak convergence), 而几乎必然收敛和 L^p 收敛之间没有互推关系.



定理 3.3 (Slutsky 定理)

设 $\{X_n\}_{n \geq 1}$ 和 $\{Y_n\}_{n \geq 1}$ 为两个随机变量序列, (X_n, Y_n) 均定义在概率空间 $(\Omega_n, \mathcal{F}_n, P_n)$ 上, 若 $X_n \xrightarrow{d} X$, 并且存在 $c \in \mathbb{R}$ 使得 $Y_n \xrightarrow{p} c$, 那么

- (1) $X_n + Y_n \xrightarrow{d} X + c$;
- (2) $X_n Y_n \xrightarrow{d} cX$;
- (3) $X_n / Y_n \xrightarrow{d} X/c$, 其中 $c \neq 0$.



证明 (1) 设 F_∞ 为 X 的 c.d.f., 于是 $X + c$ 的 c.d.f. 由 $F(x) = F_\infty(x - c)$, $x \in \mathbb{R}$ 定义, 固定 $x \in \mathcal{C}(F)$, 于是 $x - c \in \mathcal{C}(F_\infty)$.

对于任意给定的 $\epsilon > 0$, 可以得到

$$\begin{aligned}
 P(X_n + Y_n \leq x) &= P(X_n + Y_n \leq x, |Y_n - c| \leq \epsilon) + P(X_n + Y_n \leq x, |Y_n - c| > \epsilon) \\
 &\leq P(X_n + Y_n \leq x, |Y_n - c| \leq \epsilon) + P(|Y_n - c| > \epsilon) \\
 &\leq P(X_n + c - \epsilon \leq x) + P(|Y_n - c| > \epsilon)
 \end{aligned} \tag{3.5}$$

类似地有

$$P(X_n + Y_n \leq x) \geq P(X_n + c + \epsilon \leq x) - P(|Y_n - c| > \epsilon) \tag{3.6}$$

选取 $x - c, x - c + \epsilon$ 与 $x - c - \epsilon$, 使得它们都是 $\mathcal{C}(F_\infty)$ 中的元素, 那么由式 (3.5) 和 (3.6) 可知

$$\begin{aligned} F_\infty(x - c - \epsilon) &\leq \liminf_{n \rightarrow \infty} P(X_n + Y_n \leq x) \\ &\leq \limsup_{n \rightarrow \infty} P(X_n + Y_n \leq x) \leq F_\infty(x - c + \epsilon) \end{aligned}$$

现在令 $\epsilon \rightarrow 0^+$, 于是

$$\lim_{n \rightarrow \infty} P(X_n + Y_n \leq x) = F_\infty(x - c) = F(x)$$

(2), (3) 若 $c = 0$, 根据定理 3.1, 只需证明 $X_n Y_n \xrightarrow{P} 0$ 即可. 对于任意给定的 $\epsilon, \delta > 0$, 可以得到

$$\begin{aligned} P(|X_n Y_n| > \epsilon) &= P(|X_n Y_n| > \epsilon, |Y_n| \geq \delta) + P(|X_n Y_n| > \epsilon, |Y_n| < \delta) \\ &\leq P(|X_n| \geq \epsilon/\delta) + P(|Y_n| \geq \delta) \end{aligned}$$

可以选取 $\pm\epsilon/\delta$ 为 $\mathcal{C}(F_\infty)$ 中的元素, 从而

$$0 \leq \liminf_{n \rightarrow \infty} P(|X_n Y_n| > \epsilon) \leq \limsup_{n \rightarrow \infty} P(|X_n Y_n| > \epsilon) \leq P(|X_n| > \epsilon/\delta)$$

令 $\epsilon \rightarrow 0^+$, 并且可以将 δ 选得充分小, 使得 ϵ/δ 充分大, 因此 $X_n Y_n \xrightarrow{P} 0$.

若 $c \neq 0$, 那么 $X_n Y_n = c X_n + X_n(Y_n - c)$, 上面已经证明了 $X_n(Y_n - c) \xrightarrow{P} 0$, 因此只需证明 $c X_n \xrightarrow{d} c X$ 即可. 当 $c > 0$ 时, $x/a \in \mathcal{C}(F_\infty)$, 于是

$$P(c X_n \leq x) = P(X_n \leq x/c) \rightarrow P(X \leq x/c) = P(c X \leq x)$$

也即 $c X_n \xrightarrow{d} c X$, $c < 0$ 时的证明类似.

示例 3.3 假设 X_n 和 Y_n 独立, 且均服从标准正态分布, 那么

$$X_n + Y_n \xrightarrow{d} N(0, 2)$$

另一方面, 如果 $X_n = Y_n \sim N(0, 1)$, 则 $X_n + Y_n \sim N(0, 4)$. 由此可见, 如果 X_n 和 Y_n 不独立, 那么

$$X_n \xrightarrow{d} X, Y_n \xrightarrow{d} Y \not\Rightarrow X_n + Y_n \xrightarrow{d} X + Y$$

定理 3.4 (Skorohod 表示定理)

设 F 为 F_n 的极限函数, 那么存在累积分布函数为 F_n 的随机变量序列 $\{Y_n\}_{n \geq 1}$, 使得对一切 $n \geq 1$ 都有 $Y_n \xrightarrow{a.s.} Y$.



证明 设 $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{R}([0, 1])$, P 为 Lebesgue 测度, 定义 $Y_n(x) = \sup\{y : F_n(y) < x\}$, 于是 Y_n 具有累积分布函数 F_n . 现在需要证明除了可数个 $x \in \mathbb{R}$ 外都有 $Y_n(x) \rightarrow Y(x)$.

现在将 $Y_n(x)$ 改写为 $F_n^{-1}(x)$, 令 $a_x = \sup\{y : F(y) < x\}$, $b_x = \inf\{y : F(y) > x\}$, 以及 $\Omega_0 = \{x : (a_x, b_x) = \emptyset\}$, 显然各个 (a_x, b_x) 是不相交的开区间, 并且每个非空区间包含一个有理数, 故而 $\Omega \setminus \Omega_0$ 是可数集. 如果 $x \in \Omega_0$, 那么对于 $y < F^{-1}(x)$ 都有 $F(y) < x$, 对于 $z > F^{-1}(x)$ 都有 $F(z) > x$, 现在需要证明对一切 $x \in \Omega_0$ 都有 $F_n^{-1}(x) \rightarrow F^{-1}(x)$.

现在取 $y < F^{-1}(x)$ 并且使 F 在 y 处连续, 因为 $x \in \Omega_0$, 故而当 n 充分大时有 $F_n(y) < x$,

也即 $F_n^{-1}(x) \geq y$, 这对于一切 $y < F^{-1}(x)$ 成立, 故而

$$\liminf_{n \rightarrow \infty} F_n^{-1}(x) \geq F^{-1}(x)$$

类似可以证明 $\limsup_{n \rightarrow \infty} F_n^{-1}(x) \leq F^{-1}(x)$, 由此证得定理.

定理 3.5 (连续映射定理)

若 $X_n \xrightarrow{d} X$, $f: (\mathbb{R}, \mathcal{R}) \rightarrow (\mathbb{R}, \mathcal{R})$ 为可测函数, $P(X \in D_f) = 0$, 其中 D_f 为 f 的不连续点构成的集合, 那么 $f(X_n) \xrightarrow{d} f(X)$.



证明 根据 Skorohod 表示定理可知, 存在 $([0, 1], \mathcal{R}([0, 1]), m)$ 上的随机变量序列 $\{\tilde{X}_n\}_{n \geq 1}$, 使得 $\tilde{X}_n \xrightarrow{d} \tilde{X}$ 以及 $\tilde{X} \xrightarrow{d} \tilde{X}$, 并且 $\tilde{X}_n \xrightarrow{a.s.} \tilde{X}$. 令

$$A = \{\omega : \tilde{X}_n(\omega) \rightarrow \tilde{X}(\omega)\}$$

$$B = \{\omega : \tilde{X}(\omega) \notin D_f\}$$

于是 $P(A) = P(B) = 1$, 从而 $\omega \in A \cap B$, 因此

$$f(\tilde{X}_n) \rightarrow f(\tilde{X})$$

也即 $f(\tilde{X}_n) \xrightarrow{a.s.} f(\tilde{X}) \Rightarrow f(X_n) \xrightarrow{d} f(X)$.

注 如果将上述的依分布收敛改成依概率收敛或几乎必然收敛, 结论仍然成立.

3.2 紧性

定义 3.3

- 定义在 $(\mathbb{R}, \mathcal{R})$ 上的概率测度序列 $\{\mu_n\}_{n \geq 1}$ 是紧的 (tight), 如果对于任意 $\epsilon > 0$, 总存在 $M \in (0, \infty)$, 使得

$$\sup_{n \geq 1} \mu_n([-M, M]^c) < \epsilon \quad (3.7)$$

- 定义在概率空间 (Ω, \mathcal{F}, P) 上的随机序列 $\{X_n\}_{n \geq 1}$ 是紧的或随机有界的 (stochastically bounded), 如果 $\{X_n\}_{n \geq 1}$ 的概率测度序列 $\{\mu_n\}_{n \geq 1}$ 是紧的. 换言之, 对于任意给定的 $\epsilon > 0$, 总存在 $M \in (0, \infty)$, 使得

$$\sup_{n \geq 1} P(|X_n| > M) < \epsilon \quad (3.8)$$



注 对于上面的第二个定义, X_n 并不需要在同一个概率空间上, 如果 X_n 定义在 $(\Omega_n, \mathcal{F}_n, P_n)$ 上, 则式 (3.8) 变为

$$\sup_{n \geq 1} P_n(|X_n| > M) < \epsilon$$

示例 3.4 设 $X_n \sim U(n, n+1)$, 对于任意给定的 $0 < M < \infty$ 都有

$$P(|X_n| > M) \geq P(X_n > M) = 1, \quad \forall n > M$$


因此 $\sup_{n \geq 1} P(|X_n| > M) = 1$, 这表明 $\{X_n\}_{n \geq 1}$ 不是随机有界的.

示例 3.5 对于 $n \geq 1$, 定义

$$X_n \sim U(a_n, 2 + a_n) \quad (3.9)$$

其中 $a_n = (-1)^n$, 那么 $\{X_n\}$ 是随机有界的, 事实上对于任意 $\epsilon > 0$, 只需取 $M = 3$, 式 (3.7) 就能成立, 然而 $\{X_n\}_{n \geq 1}$ 并不会依分布收敛于某个随机变量 X .

定理 3.6

设 μ 和 $\{\mu_n\}_{n \geq 1}$ 是 $(\mathbb{R}, \mathcal{R})$ 上的概率测度及序列, 如果 $\mu_n \xrightarrow{d} \mu$, 那么 $\{\mu_n\}_{n \geq 1}$ 是紧的. 

证明 任意固定 $\epsilon > 0$, 存在 $a, b \in \mathbb{R}$, 使得 $\mu(\{a\}) = \mu(\{b\}) = 0$ 以及 $\mu((a, b]^c) < \epsilon/2$. 因为 $\mu_n \xrightarrow{d} \mu$, 存在正整数 $N \geq 1$, 使得当 $n \geq N$ 时就有

$$|\mu_n((a, b]) - \mu((a, b])| < \frac{\epsilon}{2}$$

故而

$$\mu_n((a, b]^c) \leq \mu((a, b]^c) + \frac{\epsilon}{2} < \epsilon \quad (3.10)$$

并且对于每个 $n = 1, 2, \dots, N$, 存在 $M_i \in (0, \infty)$ 使得


$$\mu_i([-M_i, M_i]^c) < \epsilon \quad (3.11)$$

现在取 $M = \max\{M_i : 0 \leq i \leq N\}$, 其中 $M_0 = \max\{|a|, |b|\}$, 根据式 (3.10) 和 (3.11) 可知

$$\sup_{n \geq 1} \mu_n([-M, M]^c) < \epsilon$$

因此 $\{\mu_n\}_{n \geq 1}$ 是紧的.

定理 3.7

设 μ 和 $\{\mu_n\}_{n \geq 1}$ 是 $(\mathbb{R}, \mathcal{R})$ 上的概率测度及序列, 那么 $\mu_n \xrightarrow{d} \mu$, 当且仅当 $\{\mu_n\}_{n \geq 1}$ 是紧的, 并且 $\{\mu_n\}_{n \geq 1}$ 所有依分布收敛的子序列都收敛于同一个极限概率测度. 

证明 必要性: 由 $\mu_n \xrightarrow{d} \mu$ 可知它的子序列也依分布收敛于 μ , 而紧性则由定理 3.6 给出.

充分性: 设 $\{\mu_n\}_{n \geq 1}$ 是紧的, 并且 $\{\mu_n\}_{n \geq 1}$ 所有依分布收敛的子序列都收敛于 μ , 再设 $\{F_n\}_{n \geq 1}$ 和 F 分别为 $\{\mu_n\}_{n \geq 1}$ 和 μ 的 c.d.f.

假设 $\mu_n \xrightarrow{d} \mu$ 不成立, 根据定义可知, 存在 $x_0 \in \mathbb{R}$, 使得 $\mu(\{x_0\}) = 0$, 并且当 $n \rightarrow \infty$ 时, $F_n(x_0)$ 不收敛于 $F(x_0)$, 于是存在 $\epsilon_0 \in (0, 1)$ 与一个子序列 $\{n_i\}_{i \geq 1}$, 使得

$$|F_{n_i}(x_0) - F(x_0)| \geq \epsilon_0, \quad \forall i \geq 1 \quad (3.12)$$

因为 $\{\mu_n\}_{n \geq 1}$ 是紧的, 那么存在子序列 $\{m_i\}_{i \geq 1} \subset \{n_i\}_{i \geq 1}$, 以及概率测度 μ_0 , 使得当 $i \rightarrow \infty$ 时就有

$$\mu_{m_i} \xrightarrow{d} \mu_0 \quad (3.13)$$

根据条件可知 $\mu = \mu_0$. 又因为 $\mu_0(\{x_0\}) = \mu(\{x_0\}) = 0$, 根据式 (3.13) 可知当 $i \rightarrow \infty$ 时有

$$F_{m_i}(x_0) \rightarrow F(x_0)$$

这与式 (3.12) 矛盾.

3.3 特征函数

3.3.1 基本性质

定义 3.4

定义在 $(\mathbb{R}, \mathcal{R})$ 上的概率测度 μ 的特征函数 (characteristic function, c.f.) $\varphi_\mu : \mathbb{R} \rightarrow \mathbb{C}$ 为

$$\varphi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx), \quad t \in \mathbb{R}$$

特别地, 若 X 是一个随机变量, 则它的 c.f. 为

$$\varphi(t) = E(e^{itX}) = E(\cos tX) + iE(\sin tX), \quad t \in \mathbb{R}$$

其中 $i = \sqrt{-1}$ 为虚数单位.



定理 3.8

随机变量 X 的特征函数 $\varphi(t)$ 具有以下性质:

- (1) $\varphi(0) = 1$;
- (2) $\varphi(-t) = \overline{\varphi(t)}$;
- (3) $|\varphi(t)| \leq E|e^{itX}| \leq 1$;
- (4) $|\varphi(t+h) - \varphi(t)| \leq E|e^{itX} - 1|$, 因此 $\varphi(t)$ 在 \mathbb{R} 上一致连续;
- (5) $E[e^{it(aX+b)}] = e^{itb}\varphi(at)$.



证明 (1) 是显然的.

(2) 注意到 $\varphi(-t) = E[\cos(-tX)] + iE[\sin(-tX)] = E(\cos tX) - iE(\sin tX)$.

(3) 对任意 $t \in \mathbb{R}$ 都有

$$\begin{aligned} |\varphi(t)| &= \{[E(\cos tX)]^2 + [E(\sin tX)]^2\}^{1/2} \\ &\leq [E(\cos^2 tX) + E(\sin^2 tX)]^{1/2} \leq 1 \end{aligned}$$

(4) 对任意 $t, h \in \mathbb{R}$ 都有

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= |E[e^{i(t+h)X} - e^{itX}]| \\ &\leq E|e^{i(t+h)X} - e^{itX}| = E|e^{itX} - 1| \end{aligned}$$

设 $\Delta(h) \equiv |e^{itX} - 1|$, 显然 $|\Delta(h)| \leq 2$ 且 $\lim_{h \rightarrow 0} \Delta(h) = 0$, 根据有界收敛定理可知当 $h \rightarrow 0$ 时有 $E[\Delta(h)] \rightarrow 0$, 因此

$$\limsup_{h \rightarrow 0} \limsup_{t \in \mathbb{R}} |\varphi(t+h) - \varphi(t)| \leq \lim_{h \rightarrow 0} E[\Delta(h)] = 0$$

因此 $\varphi(t)$ 在 \mathbb{R} 上一致连续.

(5) 显然有 $E[e^{it(aX+b)}] = e^{itb} E[e^{i(ta)X}] = e^{itb} \varphi(at)$.

示例 3.6 设 $P(X=1) = P(X=-1) = 1/2$, 那么 X 的 c.f. 为

$$E(e^{itX}) = (e^{it} + e^{-it})/2 = \cos t$$

示例 3.7 设 X 服从 Poisson 分布 $P(\lambda)$, 它的分布律为

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

它的特征函数为

$$E(e^{itX}) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = e^{\lambda(e^{it}-1)}$$

示例 3.8 设 X 服从标准正态分布 $N(0, 1)$, 它的 p.d.f. 为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in (-\infty, \infty)$$

它的 c.f. 为

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-it)^2/2} dx = e^{-t^2/2}$$

示例 3.9 设 X 服从均匀分布 $U(a, b)$, 它的 p.d.f. 为

$$f(x) = \frac{1}{b-a}, \quad x \in (a, b)$$

它的 c.f. 为

$$\varphi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

示例 3.10 设 X 服从指数分布 $E(\lambda)$, 它的 p.d.f. 为

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

它的 c.f. 为

$$\varphi(t) = \frac{1}{1 - \lambda^{-1}it}$$

示例 3.11 设 X 服从标准 Cauchy 分布, 它的 p.d.f. 为

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty, \infty)$$

它的 c.f. 为

$$\varphi(t) = e^{-|t|}$$

如果 X_1, X_2, \dots 是独立的且都服从标准 Cauchy 分布, 那么

$$\frac{X_1 + \dots + X_n}{n}$$

具有和 X_1 相同的分布.

定理 3.9

设随机变量 X 的 c.d.f. 和 c.f. 分别为 F 和 φ , 如果 F 绝对连续, 那么

$$\lim_{|t| \rightarrow \infty} |\varphi(t)| = 0 \quad (3.14)$$



证明 因为 F 绝对连续, 故而 X 的概率分布 μ_X 存在关于 \mathbb{R} 上的 Lebesgue 测度 m 的概率密度函数 f , 并且

$$\varphi(t) = \int e^{itx} f(x) dx, \quad t \in \mathbb{R}$$

任取 $\epsilon > 0$, 因为 $f \in L^1(\mathbb{R}, \mathcal{R}, m)$, 故而存在阶梯函数 $f_\epsilon = \sum_{j=1}^k c_j I_{(a_j, c_j)}$, 使得

$$\int |f - f_\epsilon| dm < \epsilon/2 \quad (3.15)$$

其中 $a_j, b_j, c_j \in \mathbb{R}, j = 1, \dots, k, 1 \leq k < \infty$. 于是对于任意 $t \neq 0$ 都有

$$\left| \int e^{itx} f_\epsilon(x) dx \right| = \left| \sum_{j=1}^k c_j \int_{a_j}^{b_j} e^{itx} dx \right| \leq \sum_{j=1}^k |c_j| \frac{2}{|t|} \quad (3.16)$$

从而由式 (3.15) 和 (3.16) 可知对于一切 $|t| > |t_\epsilon|$ 都有

$$\begin{aligned} |\varphi(t)| &= \left| \int e^{itx} f(x) dx \right| \\ &\leq \int |f - f_\epsilon| dx + \left| \int e^{itx} f_\epsilon(x) dx \right| \\ &< \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

其中 $|t_\epsilon| = 4 \sum_{j=1}^k |c_j|/\epsilon$. 因此式 (3.14) 成立, 命题得证.

引理 3.1

对于任意 $x \in \mathbb{R}$ 和 $r \in \mathbb{N}$ 有

$$\left| e^{ix} - \sum_{k=0}^{r-1} \frac{(ix)^k}{k!} \right| \leq \min \left\{ \frac{|x|^r}{r!}, \frac{2|x|^{r-1}}{(r-1)!} \right\} \quad (3.17)$$



证明 注意到对任意 $x \in \mathbb{R}$ 和 $r \in \mathbb{N}$ 有

$$\frac{d^r}{dx^r} e^{ix} = \frac{d^r}{dx^r} \cos x + i \frac{d^r}{dx^r} \sin x = i^r e^{ix}$$

根据一阶 Taylor 展开得

$$e^{ix} = \sum_{k=0}^{r-1} \frac{(ix)^k}{k!} + \frac{(ix)^r}{(r-1)!} \int_0^1 (1-u)^r e^{iux} du$$

故而对任意 $x \in \mathbb{R}$ 和 $r \in \mathbb{N}$ 有

$$\left| e^{ix} - \sum_{k=0}^{r-1} \frac{(ix)^k}{k!} \right| \leq \frac{|x|^r}{r!} \quad (3.18)$$

当 $r \geq 2$ 时, 根据式 (3.18) 可得

$$\left| e^{ix} - \sum_{k=0}^{r-1} \frac{(ix)^k}{k!} \right| \leq \left| e^{ix} - \sum_{k=0}^{r-2} \frac{(ix)^k}{k!} \right| + \frac{|x|^{r-1}}{(r-1)!} \quad (3.19)$$

最后由式 (3.18) 和 (3.19) 可知式 (3.17) 对一切 $x \in \mathbb{R}$ 和大于等于 2 的整数 r 成立, 而 $r = 1$ 时的结论是易证的.

定理 3.10

设 X 为随机变量且它的 c.f. 为 φ , 若 $r \in \mathbb{N}$ 且 $E|X|^r < \infty$, 那么 φ 是 r 阶连续可微的, 并且

$$\varphi^{(r)}(t) = E[(iX)^r]e^{itX}, \quad t \in \mathbb{R}$$



证明 设 μ 为 $(\mathbb{R}, \mathcal{R})$ 上的 X 的概率分布, 假设 $E|X| < \infty$, 先证对于任意 $t \in \mathbb{R}$ 都有 $\varphi^{(1)}(t) = E(iXe^{itX})$.

固定 $t \in \mathbb{R}$, 对于任意 $h \in \mathbb{R}$ 且 $h \neq 0$ 都有

$$\begin{aligned} h^{-1}[\varphi(t+h) - \varphi(t)] &= \int_{\mathbb{R}} e^{itx} \left(\frac{e^{ihx} - 1}{h} \right) \mu(dx) \\ &= \int_{\mathbb{R}} \psi_h(x) \mu(dx) + \int_{\mathbb{R}} ix e^{itx} \mu(dx) \end{aligned} \quad (3.20)$$

其中

$$\psi_h(x) = e^{itx} \left(\frac{e^{ihx} - 1}{h} - ix \right)$$

根据引理 3.1 可知

$$|\psi_h(x)| \leq \min \left\{ \frac{h|x|^2}{2}, 2|x| \right\}, \quad \forall x \in \mathbb{R}, h \neq 0 \quad (3.21)$$

因此对一切 $x \in \mathbb{R}$ 都有 $\lim_{h \rightarrow 0} \psi_h(x) = 0$. 除此之外, 还能得到 $|\psi_h(x)| \leq 2|x|$ 及 $\int_{\mathbb{R}} |x| \mu(dx) = E|X| < \infty$, 根据 DCT 可知

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} \psi_h(x) \mu(dx) = 0$$

根据式 (3.20) 即可得到 φ 在 t 处可微, 并且 $\varphi^{(1)}(t) = \int_{\mathbb{R}} ix e^{itx} \mu(dx) = E(iXe^{itX})$.

为了证明 $r+1$ 的情况, 对任意 $t \in \mathbb{R}$ 和 $h \neq 0$ 都有

$$h^{-1}[\varphi^{(r)}(t+h) - \varphi^{(r)}(t)] - E[(iX)^{r+1}]e^{itX} = \int_{\mathbb{R}} (ix)^r \psi_h(x) \mu(dx)$$

根据式 (3.21) 蕴含的 $\psi_h(x)$ 的有界性, 矩条件 $E|X|^{r+1} < \infty$, 由 DCT 可知当 $h \rightarrow 0$ 时, 上式右端也趋近于 0.

3.3.2 Feller 反演公式

定理 3.11 (Feller 反演公式)

设 X 是定义在概率空间 (Ω, \mathcal{F}, P) 上的随机变量, 并且

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} \mu_X(dx)$$

μ_X 为 X 的概率分布, 如果 $a < b$, 那么

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \mu((a, b)) + \frac{1}{2} \mu(\{a, b\})$$



证明 根据 $\varphi(t) = \int_{\mathbb{R}} e^{itx} \mu_X(dx)$, 设

$$I_T = \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{-T}^T \int \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \mu_X(dx) dt$$

注意到

$$\left| \int_a^b e^{-ity} dy \right| = \left| \frac{e^{-ita} - e^{-itb}}{it} \right| \leq b - a$$

以及 μ_X 是概率测度, 区间 $[-T, T]$ 是有限的, 根据 Fubini 定理

$$\begin{aligned} I_T &= \int_{\mathbb{R}} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt \mu_X(dx) \\ &= \int_{\mathbb{R}} \left\{ \int_{-T}^T \frac{\sin[t(x-a)]}{t} dt - \int_{-T}^T \frac{\sin[t(x-b)]}{t} dt \right\} \mu_X(dx) \end{aligned}$$

引入记号 $R(\theta, T) = \int_{-T}^T (\sin \theta t)/t dt$, 可以将上式简写为

$$I_T = \int [R(x-a, T) - R(x-b, T)] \mu_X(dx) \quad (3.22)$$

当 $\theta > 0$ 时, 令 $t = x/\theta$, 于是

$$R(\theta, T) = 2 \int_0^{\theta T} \frac{\sin x}{x} dx = 2S(\theta T)$$

对于 $\theta < 0$ 有 $R(\theta, T) = -R(|\theta|, T)$, 利用符号函数 sgn 可以将 $R(\theta, T)$ 记为

$$R(\theta, T) = 2(\text{sgn } \theta) S(|\theta|T)$$

令 $T \rightarrow \infty$, 此时 $S(T) \rightarrow \pi/2$, $R(\theta, T) \rightarrow \pi \text{sgn } \theta$, 因此

$$R(x-a, T) - R(x-b, T) \rightarrow \begin{cases} 2\pi, & x \in (a, b) \\ \pi, & x \in \{a\} \cup \{b\} \\ 0, & x \in (-\infty, a) \cup (b, \infty) \end{cases}$$

显然 $|R(\theta, T)| \leq 2 \sup_y S(y) < \infty$, 根据有界收敛定理可知当 $T \rightarrow \infty$ 时有

$$(2\pi)^{-1} I_T \rightarrow \mu((a, b)) + \frac{1}{2} \mu(\{a, b\})$$

推论 3.1

设 μ_X 和 μ_Y 分别为随机变量 X 和 Y 的概率分布, 如果 $\varphi_X = \varphi_Y$, 那么 $\mu_X = \mu_Y$. 

证明 根据 Feller 反演公式可知, 对一切 $a, b \in K$ 都有 $\mu_X((a, b)) = \mu_Y((a, b))$, 其中 $K = \{x \in \mathbb{R} : \mu_X(\{x\}) = \mu_Y(\{x\}) = 0\}$. 因为 K^c 是至多可数的, 于是集族 $\mathcal{C} = \{(a, b) : a, b \in K\}$ 是一个生成了 σ -代数 \mathcal{R} 的 π -系, 根据测度扩张的唯一性可知 $\mu_X = \mu_Y$.

注 这个推论表明, 如果两个随机变量具有相同的 c.f., 那么它们有相同的概率分布. 事实上, 这个命题反过来也成立.

推论 3.2

若 φ 为实函数, 那么 X 和 $-X$ 有相同的概率分布. 

证明 显然, 若 φ 为实函数, 则

$$\varphi(t) = \int_{\mathbb{R}} \cos tx \, dF(x)$$

因为 F 为 X 的 c.d.f., 所以

$$\varphi(t) = \varphi(-t) = E(e^{-itX}) = E[e^{it(-X)}]$$

故而 X 和 $-X$ 的概率分布相同.

定理 3.12

设 μ_X 为 X 的概率分布, 那么对于任意 $a \in \mathbb{R}$ 都有

$$\mu_X(\{a\}) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-ita} \varphi(t) \, dt$$



证明 根据特征函数 φ 的定义和 Fubini 定理可知

$$\begin{aligned} I_T &= \frac{1}{2T} \int_{-T}^T e^{-ita} \varphi(t) \, dt \\ &= \frac{1}{2T} \int_{-T}^T \int_{-\infty}^{\infty} e^{-ita} e^{itx} \, dx \, dt \\ &= \frac{1}{2T} \int_{-\infty}^{\infty} \int_{-T}^T e^{it(x-a)} \, dt \, dx \end{aligned}$$

根据 DCT 可知

$$\lim_{T \rightarrow \infty} I_T = \int_{-\infty}^{\infty} \left(\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \cos t(x-a) \, dt \right) dx$$

当 $x \neq a$ 时, 极限为 0; 而当 $x = a$ 时, 极限为 1.

示例 3.12 设 $h > 0$, 如果 $P(X \in h\mathbb{Z}) = 1$, 那么 X 的 c.f. 满足 $\varphi(2\pi/h + t) = \varphi(t)$, 以及

$$P(X = x) = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-itx} \varphi(t) \, dt, \quad x \in h\mathbb{Z} \quad (3.23)$$

进一步, 如果 $P(X \in b + h\mathbb{Z}) = 1$, 那么式 (3.23) 也适用于 $x \in b + h\mathbb{Z}$.

证明 由于 $P(x \in h\mathbb{Z}) = 1$, 并且对于任意整数 n 都有 $e^{2i\pi n} = 1$, 故而

$$\varphi(2\pi/h + t) = \int_{h\mathbb{Z}} e^{i(2\pi/h+t)x} \, dx = \int_{h\mathbb{Z}} e^{itx} \, dx = \varphi(t)$$

根据定理 3.12 可知

$$P(X = x) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi(t) \, dt$$

令 $T_n = \pi/h + 2\pi n/h$, 其中 $n \geq 0$, 于是

$$P(X = x) = \lim_{n \rightarrow \infty} \frac{1}{2T_n} \int_{-T_n}^{T_n} e^{-itx} \varphi(t) \, dt$$

对任意 $k \geq 1$, 注意到

$$\begin{aligned}\int_{T_{k-1}}^{T_k} e^{-itx} \varphi(t) dt &= \int_{-\pi/h}^{\pi/h} e^{-i(2k\pi/h+t)x} \varphi(2k\pi/h+t) dt \\ &= \int_{-T_0}^{T_0} e^{-itx} \varphi(t) dt\end{aligned}$$

由此可知

$$\frac{1}{2T_n} \int_{-T_n}^{T_n} e^{-itx} \varphi(t) dt = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-itx} \varphi(t) dt$$

从而结论成立. 最后, 由于 $P(X \in b + h\mathbb{Z}) = 1$, 故而 $P(Y \in h\mathbb{Z}) = 1$, 套用以上结论即可.

引理 3.2

设 X 是定义在概率空间 (Ω, \mathcal{F}, P) 上的随机变量, 如果 $\int_{\mathbb{R}} |\varphi(t)| dt < \infty$, 那么对任意 $a \in \mathbb{R}$ 都有 $P(X = a) = 0$.



证明 根据 Feller 反演公式可知

$$\begin{aligned}\mu_X((a, b)) + \frac{1}{2}\mu_X(\{a, b\}) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \\ &\leq \limsup_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \left| \frac{e^{-ita} - e^{-itb}}{it} \right| |\varphi(t)| dt \\ &\leq \frac{|b-a|}{2\pi} \int_{\mathbb{R}} |\varphi(t)| dt < \infty\end{aligned}$$

令 $a = b$ 可得 $P(X = a) \leq 0$, 也即 $P(X = a) = 0$.

定理 3.13

如果 $\int_{\mathbb{R}} |\varphi(t)| dt < \infty$, μ_X 为 X 的概率分布, 那么 μ_X 具有连续有界的密度函数

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \varphi(t) dt, \quad x \in \mathbb{R}$$



证明 注意到

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-itx} dx \right| \leq b - a$$

根据 Feller 反演公式可知

$$\begin{aligned}\mu_X((a, b)) + \frac{1}{2}\mu_X(\{a, b\}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \\ &\leq \frac{b-a}{2\pi} \int_{-\infty}^{\infty} |\varphi(t)| dt < \infty\end{aligned}$$

根据引理3.2可知, $\mu_X(\{a, b\}) = 0$, 于是对于任意 $h > 0$, 由 Fubini 定理可知

$$\begin{aligned} F_X(a+h) - F_X(a) &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-itx} - e^{-it(x+h)}}{it} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_a^{a+h} e^{-itx} dx \right) \varphi(t) dt \\ &= \int_a^{a+h} \left(\frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt \right) dx \end{aligned}$$

其中 F_X 为 X 的 c.d.f., 因此

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt$$

示例 3.13 设 X_1, \dots, X_n 是独立随机变量, 并且在 $(-1, 1)$ 上服从均匀分布, 那么对 $X_1 + \dots + X_n$ 具有概率密度

$$f(x) = \frac{1}{\pi} \int_0^\infty \left(\frac{\sin t}{t} \right)^n \cos tx dt$$

证明 首先易知 X_1 的 c.f. 为

$$\varphi(t) = E(e^{itX_1}) = \frac{1}{2it} (e^{it} - e^{-it}) = \frac{\sin t}{t}$$

于是 $X_1 + \dots + X_n$ 的 c.f. 为 $\varphi_n(t) = (\sin t/t)^n$. 由于对于任意 $n \geq 2$ 都有

$$\int_{-\infty}^\infty \left| \frac{\sin t}{t} \right|^n dt < \infty$$

根据定理3.13可知

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{-itx} \left(\frac{\sin t}{t} \right)^n dt = \frac{1}{\pi} \int_0^\infty \left(\frac{\sin t}{t} \right)^n \cos tx dt$$

3.3.3 Lévy-Cramér 连续定理

定理 3.14

设 F_n 和 F , 以及 φ_n 和 φ 分别为随机变量 X_n 和 X 对应的 c.d.f. 与 c.f., 若 $F_n \xrightarrow{d} F$, 那么对一切 $0 < K < \infty$, 当 $n \rightarrow \infty$ 时都有

$$\sup_{|t| < K} |\varphi_n(t) - \varphi(t)| \rightarrow 0$$



证明 根据 Skorohod 表示定理可知, 存在定义在 $([0, 1], \mathcal{R}([0, 1]), m)$ 上的随机变量 X_n 和 X , 使得 $X_n \stackrel{d}{=} F_n$, $X \stackrel{d}{=} F$, 并且 $X_n \xrightarrow{a.s.} X$, 其中 m 为 Lebesgue 测度. 于是对于任意 $t \in \mathbb{R}$ 都有

$$\begin{aligned} |\varphi_n(t) - \varphi(t)| &\leq E|1 - e^{it(X-X_n)}| \\ &\leq E \left[|1 - e^{it(X-X_n)}| 1_{(|X-X_n| \leq \epsilon)} \right] + P(|X - X_n| > \epsilon) \end{aligned}$$

从而

$$\sup_{|t| < K} |\varphi_n(t) - \varphi(t)| \leq \sup_{|u| \leq K\epsilon} |1 - e^{iu}| + P(|X - X_n| > \epsilon)$$

给定 K 和 $\delta > 0$, 选择充分小的 $\epsilon > 0$ 就能使得

$$\sup_{|u| \leq K\epsilon} |1 - e^{iu}| < \delta$$

由于对于任意的 $\epsilon > 0$, 当 $n \rightarrow \infty$ 时有 $P(|X - X_n| > \epsilon) \rightarrow 0$, 因此

$$\lim_{n \rightarrow \infty} \sup_{|t| < K} |\varphi_n(t) - \varphi(t)| = 0$$

引理 3.3

设 μ 是 $(\mathbb{R}, \mathcal{R})$ 上的概率测度, 它的 c.f. 为 φ , 那么对于任意 $\delta > 0$ 都有

$$\mu(\{x : |x|\delta \geq 2\}) \leq \delta^{-1} \int_{-\delta}^{\delta} 1 - \varphi(u) du$$



证明 任意固定 $\delta > 0$, 根据 Fubini 定理可知

$$\begin{aligned} \int_{-\delta}^{\delta} 1 - \varphi(u) du &= \int_{\mathbb{R}} \left[\int_{-\delta}^{\delta} (1 - e^{iux}) du \right] \mu(dx) \\ &= 2\delta \int_{\mathbb{R}} \left(1 - \frac{\sin \delta x}{\delta x} \right) \mu(dx) \\ &\geq 2\delta \int_{\{x: |x|\delta \geq 2\}} \left(1 - \frac{1}{|x|\delta} \right) \mu(dx) \\ &\geq \delta \mu(\{x : |x|\delta \geq 2\}) \end{aligned}$$

因此结论成立.

引理 3.4

设 $\{\mu_n\}_{n \geq 1}$ 为概率测度序列, 其对应的 c.f. 构成序列 $\{\varphi_n\}_{n \geq 1}$, 若对于某个 $\delta_0 > 0$ 而言, 极限

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t), \quad \forall |t| \leq \delta_0$$

存在, 若 $\varphi(t)$ 在 $t = 0$ 处连续, 那么 $\{\mu_n\}_{n \geq 1}$ 是紧的.



证明 对于任意 $0 < \delta < \delta_0$, 根据有界收敛定理可知

$$\delta^{-1} \int_{-\delta}^{\delta} [1 - \varphi_n(t)] dt \rightarrow \delta^{-1} \int_{-\delta}^{\delta} [1 - \varphi(t)] dt$$

根据 $\varphi(t)$ 在 $t = 0$ 处的连续性, 当 $\delta \rightarrow 0$ 时有

$$\delta^{-1} \int_{-\delta}^{\delta} [1 - \varphi(t)] dt \rightarrow 0$$

也即对于任意给定的 $\epsilon > 0$, 总存在 $\delta_\epsilon \in (0, \delta_0)$ 和正整数 $M_\epsilon \in (0, \infty)$, 使得当 $n \geq M_\epsilon$ 时就有

$$\delta_\epsilon^{-1} \int_{-\delta_\epsilon}^{\delta_\epsilon} [1 - \varphi_n(t)] dt < \epsilon$$

根据引理 3.3 可知, 对一切 $n \geq M_\epsilon$ 都有

$$\mu(\{x : |x| \geq 2/\delta_\epsilon\}) < \epsilon$$

选取 $K_\epsilon > 2/\delta_\epsilon$ 使得

$$\mu_j(\{x : |x| \geq K_\epsilon\}) < \epsilon, \quad 1 \leq j \leq M_\epsilon$$

因此

$$\sup_{n \geq 1} \mu_n(\{x : |x| \geq K_\epsilon\}) < \epsilon$$

也即 $\{\mu_n\}_{n \geq 1}$ 是紧的.

定理 3.15 (Lévy-Cramér 连续定理)

设 μ_n 为 $(\mathbb{R}, \mathcal{R})$ 上的概率测度, $1 \leq n \leq \infty$, 并且它的 c.f. 为 φ_n .

- (1) 如果 $\mu_n \xrightarrow{d} \mu_\infty$, 那么对一切 $t \in \mathbb{R}$ 都有 $\varphi_n(t) \rightarrow \varphi_\infty(t)$;
- (2) 若对一切 $t \in \mathbb{R}$ 都有 $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$, 如果 $\varphi(t)$ 在 $t = 0$ 处连续, 那么 φ 是 μ 的 c.f., 并且 $\mu_n \xrightarrow{d} \mu$.



证明 (1) 因为 $f(x) = e^{itx}$ 是连续有界函数, 因为 $\mu_n \xrightarrow{d} \mu_\infty$, 根据连续映射定理可知结论成立.

(2) 根据引理 3.4 可知 $\{\mu_n\}_{n \geq 1}$ 是紧的. 设 $\{\mu_{n_j}\}_{j \geq 1}$ 为 $\{\mu_n\}_{n \geq 1}$ 的子序列且它收敛于 μ . 根据紧性可知 μ 是一个概率测度, 根据定理 3.14 可知, 极限 $\lim_{n \rightarrow \infty} \varphi_{n_j}(t)$ 是 μ 的 c.f., 也即 φ 是 μ 的 c.f., 最后根据定理 3.7 推知 $\mu_n \xrightarrow{d} \mu$.

示例 3.14 设 X_1, X_2, \dots 为独立随机变量, $S_n = X_1 + \dots + X_n$, φ_k 为 X_k 的 c.f., 如果 $S_n \rightarrow S_\infty$, 那么 S_∞ 的 c.f. 为 $\prod_{k=1}^{\infty} \varphi_k(t)$.

示例 3.15 如果 X_n 和 Y_n 相互独立, $1 \leq n \leq \infty$, $X_n \xrightarrow{d} X_\infty$ 且 $Y_n \xrightarrow{d} Y_\infty$, 那么 $X_n + Y_n \xrightarrow{d} X_\infty + Y_\infty$.

证明 设 φ_n 和 ψ_n 分别为 X_n 和 Y_n 的 c.f., 由 Lévy-Cramér 连续定理, 对于任意 $t \in \mathbb{R}$ 都有

$$\varphi_n(t) \rightarrow \varphi_\infty(t), \quad \psi_n(t) \rightarrow \psi_\infty(t)$$

根据独立性可知 $X_n + Y_n$ 的 c.f. 为 $\varphi_n \psi_n$, 进一步有

$$\varphi_n(t) \psi_n(t) \rightarrow \varphi_\infty(t) \psi_\infty(t)$$

因为 φ_∞ 和 ψ_∞ 在 $t = 0$ 处连续, 再由 Lévy-Cramér 连续定理知结论成立.

3.4 中心极限定理

3.4.1 Lindeberg-Feller 定理

引理 3.5

对任意 $n \geq 1$ 和复数 z_1, \dots, z_n 及 w_1, \dots, w_n , 若对一切 $1 \leq j \leq n$ 都有 $|z_j| \leq 1$ 及 $|w_j| \leq 1$, 那么

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n w_j \right| \leq \sum_{j=1}^n |z_j - w_j|$$



证明 只需将 $\prod_{j=1}^n z_j - \prod_{j=1}^n w_j$ 改写为

$$\begin{aligned} \prod_{j=1}^n z_j - \prod_{j=1}^n w_j &= \prod_{j=1}^n z_j - \left(\prod_{j=1}^{n-1} w_j \right) w_n \\ &\quad + \left(\prod_{j=1}^{n-1} z_j \right) w_n - \left(\prod_{j=1}^{n-2} z_j \right) w_{n-1} w_n \\ &\quad + \cdots + z_1 \prod_{j=2}^n w_j - \prod_{j=1}^n w_j \end{aligned}$$

根据三角不等式即可证得引理.

定理 3.16 (Lindeberg-Feller 定理)

假设对于每个 $n \geq 1, 1 \leq m \leq n$, 随机变量 $X_{n,m}$ 是独立的, $E(X_{n,m}) = 0, E(X_{n,m}^2) = \sigma_{n,m}^2 < \infty$, 并且对任意 $\epsilon > 0$ 都满足 Lindeberg 条件

$$\lim_{n \rightarrow \infty} s_n^{-2} \sum_{m=1}^n E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon s_n)}] = 0 \quad (3.24)$$

其中 $s_n^2 = \sum_{m=1}^n \sigma_{n,m}^2$. 再设 $S_n = \sum_{m=1}^n X_{n,m}$, 那么当 $n \rightarrow \infty$ 时有

$$S_n/s_n \xrightarrow{d} N(0, 1)$$



证明 不失一般性, 可以假设对一切 $n \geq 1$ 都有 $s_n^2 = 1$, 因为总可以置 $\tilde{X}_{n,m} = X_{n,m}/s_n$. 根据 Lévy-Cramér 连续定理, 只需证明对于一切 $t \in \mathbb{R}$ 都有

$$\lim_{n \rightarrow \infty} E[\exp(itS_n)] = e^{-t^2/2}$$

对于任意的 $\epsilon > 0$, 注意到

$$\begin{aligned} \sigma_{n,m}^2 &= E[X_{n,m}^2 1_{(|X_{n,m}| \leq \epsilon)}] + E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon)}] \\ &\leq \epsilon^2 + E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon)}] \end{aligned}$$

根据 Lindeberg 条件可知

$$\Delta_n = \max_{1 \leq m \leq n} \sigma_{n,m}^2 \leq \epsilon^2 + \sum_{m=1}^n E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon)}]$$

由于 ϵ 是任意的, 故当 $n \rightarrow \infty$ 时有

$$\Delta_n \rightarrow 0 \quad (3.25)$$

设 $\varphi_{n,m}$ 为随机变量 $X_{n,m}$ 的 c.f., 固定 $t \in \mathbb{R}$, 根据式 (3.25) 可知存在正整数 N , 使得对于一切 $n > N$ 都有 $\max_{1 \leq m \leq n} |1 - t^2 \sigma_{n,m}^2/2| \leq 1$, 注意到 $s_n^2 = \sum_{m=1}^n \sigma_{n,m}^2 = 1$, 由引理 3.5 可知, 对于

$n > N$ 都有

$$\begin{aligned}
 \left| E[\exp(itS_n)] - e^{-t^2/2} \right| &\leq \left| \prod_{m=1}^n \varphi_{n,m}(t) - \prod_{m=1}^n \left(1 - \frac{t^2 \sigma_{n,m}^2}{2} \right) \right| \\
 &\quad + \left| \prod_{m=1}^n \left(1 - \frac{t^2 \sigma_{n,m}^2}{2} \right) - \prod_{m=1}^n e^{-t^2 \sigma_{n,m}^2/2} \right| \\
 &\leq \sum_{m=1}^n \left| \varphi_{n,m}(t) - \left(1 - \frac{t^2 \sigma_{n,m}^2}{2} \right) \right| \\
 &\quad + \sum_{m=1}^n \left| e^{-t^2 \sigma_{n,m}^2/2} - \left(1 - \frac{t^2 \sigma_{n,m}^2}{2} \right) \right| \\
 &= I_{1n} + I_{2n}
 \end{aligned}$$

现在只需证明对于 $k = 1, 2$, 当 $n \rightarrow \infty$ 时有 $I_{kn} \rightarrow 0$.

根据引理 3.1 可知, 对于任意 $x \in \mathbb{R}$ 都有 $|e^{ix} - [1 + ix + (ix)^2/2]| \leq \min\{|x|^3/6, |x|^2\}$. 从而

$$\begin{aligned}
 I_{1n} &= \sum_{m=1}^n \left| E[\exp(itX_{n,m})] - \left[1 + itE(X_{n,m}) + \frac{(it)^2}{2} E(X_{n,m}^2) \right] \right| \\
 &\leq \sum_{m=1}^n E(\min\{|tX_{n,m}|^3/6, |tX_{n,m}|^2\}) \\
 &\leq \sum_{m=1}^n E[|tX_{n,m}|^3 1_{(|X_{n,m}| \leq \epsilon)}] + \sum_{m=1}^n E[(tX_{n,m})^2 1_{(|X_{n,m}| > \epsilon)}] \\
 &\leq |t|^3 \epsilon + t^2 \sum_{m=1}^n E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon)}]
 \end{aligned}$$

由 Lindeberg 条件和 ϵ 的任意性可知当 $n \rightarrow \infty$ 时有 $I_{1n} \rightarrow 0$. 注意到对于任意 $x \in \mathbb{R}$ 都有 $|e^x - x - 1| \leq x^2 e^{|x|}$, 于是

$$I_{2n} \leq \sum_{m=1}^n \left(\frac{t^2 \sigma_{n,m}^2}{2} \right)^2 e^{t^2 \sigma_{n,m}^2/2} \leq t^4 e^{t^2 \Delta_n/2} \Delta_n$$

从而当 $n \rightarrow \infty$ 时也有 $I_{2n} \rightarrow 0$, 由此证得结论.

示例 3.16 设 X_1, X_2, \dots 为独立随机变量, $|X_i| \leq M$, $\sum_n \text{Var}(X_n) = \infty$, 那么当 $n \rightarrow \infty$ 时有

$$\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow{d} N(0, 1)$$

证明 令 $s_n^2 = \sum_{m=1}^n \text{Var}(X_m)$, $Y_{n,m} = [X_m - E(X_m)]/s_n$, 于是 $E(Y_{n,m}) = 0$ 且 $\sum_{m=1}^n E(Y_{n,m}^2) = 1$. 注意到 $|X_i| < M$, 故而 $|X_m - E(X_m)| \leq 2M$, 因为 $\sum_n \text{Var}(X_n) = \infty$, 所以对于任意 $\epsilon > 0$, 存在充分大的 $n \in \mathbb{N}$ 使得 $\epsilon s_n > 2M$, 又因为

$$|Y_{n,m}| > \epsilon \Leftrightarrow |X_m - E(X_m)| > \epsilon s_n$$

从而 Lindeberg 条件

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n E[Y_{n,m}^2 1_{(|Y_{n,m}| > \epsilon)}] = 0$$

成立, 根据 Lindeberg-Feller 定理可知结论成立.

示例 3.17 Kolmogorov 三级数定理的必要条件 设 X_1, X_2, \dots 是独立随机变量, $A > 0$ 且 $Y_i = X_i 1_{(|X_i| \leq A)}$, 如果级数 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛, 那么必有

- (1) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$;
- (2) $\sum_{n=1}^{\infty} E(Y_n)$ 收敛;
- (3) $\sum_{n=1}^{\infty} \text{Var}(Y_n) < \infty$.

证明 由于 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛, 直接可得条件 (1) 成立, 如若不然, 根据 Borel-Cantelli 第二引理可知 $P(|X_n| > A \text{ i.o.}) = 1$, 这与 $\sum_{n=1}^{\infty} X_n$ 的收敛性矛盾.

再假设当条件 (1) 成立时, 条件 (3) 不成立. 令

$$c_n = \sum_{m=1}^n \text{Var}(Y_m), \quad X_{n,m} = [Y_m - E(Y_m)]/c_n^{1/2}$$

于是 $E(X_{n,m}) = 0$ 且 $\sum_{m=1}^n E(X_{n,m}^2) = 1$. 因为当 $n \rightarrow \infty$ 时有 $c_n \rightarrow \infty$, 故而对任意的 $\epsilon > 0$ 和充分大的 $n \in \mathbb{N}$ 有 $2A/c_n^{1/2} < \epsilon$, 也即 $|X_{n,m}| \leq \epsilon$ 对一切 $m = 1, \dots, n$ 成立, 进一步可知

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n E[|X_{n,m}|^2; |X_{n,m}| > \epsilon] = 0$$

设 $S_n = X_{n,1} + \dots + X_{n,n}$, 根据 Lindeberg-Feller 定理可知当 $n \rightarrow \infty$ 时有 $S_n \xrightarrow{d} N(0, 1)$.

级数 $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛和条件 (1) 意味着 $\sum_{n=1}^{\infty} Y_n$ 也是几乎必然收敛的, 并且当 $n \rightarrow \infty$ 时, $T_n = (Y_1 + \dots + Y_n)/c_n^{1/2} \xrightarrow{d} 0$, 于是由定理 3.2 和 Slutsky 定理可知 $(S_n - T_n) \xrightarrow{d} N(0, 1)$. 又因为对于任意 $n \in \mathbb{N}$, $S_n - T_n$ 是非随机的, 这就产生矛盾, 所以条件 (3) 必然成立.

最后假设条件 (1) 和 (3) 成立, 由 Khinchin-Kolmogorov 级数定理可知

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n [Y_m - E(Y_m)]$$

存在, 因为 $\sum_{n=1}^{\infty} Y_n$ 几乎必然收敛, 所以条件 (2) 成立.

示例 3.18 无限方差 假设 X_1, X_2, \dots 是 i.i.d. 随机变量, $P(X_1 > x) = P(X_1 < -x)$ 且对一切 $x \geq 1$ 都有 $P(|X_1| > x) = x^{-2}$, 那么

$$E|X_1|^2 = \int_0^{\infty} 2xP(|X_1| > x) dx = \infty$$

再令 $S_n = X_1 + \dots + X_n$ 是适当标准化的, 那么它收敛于一个正态分布.

证明 首先令 $Y_{n,m} = X_{n,m} 1_{(|X_m| \leq n^{1/2} \log \log n)}$, 其中 $c_n = n^{1/2} \log \log n$ 选得足够大, 可以使得当 $n \rightarrow \infty$ 时有

$$\sum_{m=1}^n P(Y_{n,m} \neq X_m) \leq nP(|X_1| > c_n) \rightarrow 0$$

现在我们要 $Y_{n,m}$ 尽可能的小. 因为 $P(|Y_{n,m}| > x) \leq P(|X_1| > x)$, 并且当 $x > c_n$ 时有 $P(|Y_{n,m}| > x) = 0$, 进而

$$\begin{aligned} E(Y_{n,m}^2) &\leq \int_0^{c_n} 2yP(|X_1| > y) dy = 1 + \int_1^{c_n} \frac{2}{y} dy \\ &= 1 + 2 \log c_n = 1 + \log n + 2 \log \log n \sim \log n \end{aligned}$$

注意到

$$P(|Y_{n,m}| > x) = P(|X_1| > x) - P(|X_1| > c_n)$$

并且当 $x \leq \sqrt{n}$ 时有 $P(|X_1| > c_n) \leq (\log \log n)^{-2} P(|X_1| > x)$, 因此

$$P(|Y_{n,m}| > x) \geq [1 - (\log \log n)^{-2}] P(|X_1| > x)$$

从而

$$E(Y_{n,m}^2) \geq [1 - (\log \log n)^{-2}] \int_1^{\sqrt{n}} \frac{2}{y} dy \sim \log n$$

再令 $T_n = Y_{n,1} + \cdots + Y_{n,n}$, 所以 $\text{Var}(T_n) \sim n \log n$, 因为 $|Y_{n,m}| \leq n^{1/2} \log \log n$, 容易验证 Lindeberg 条件成立, 对 $X_{n,m} = Y_{n,m}/(n \log n)^{1/2}$ 使用 Lindeberg-Feller 定理可得

$$\frac{T_n}{\sqrt{n \log n}} \xrightarrow{d} N(0, 1)$$

因为 c_n 使得 $P(S_n \neq T_n) \rightarrow 0$, 故而这一结果对 S_n 也成立.

定理 3.17 (Lindeberg-Lévy 定理)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_1) = \mu$ 且 $\text{Var}(X_1) = \sigma^2 < \infty$, 那么当 $n \rightarrow \infty$ 时有

$$\sqrt{n}(S_n/n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

其中 $S_n = X_1 + \cdots + X_n$.



证明 令 $X_{n,m} = (X_m - \mu)/\sigma\sqrt{n}$, 于是 $s_n^2 = \sum_{m=1}^n \text{Var}(X_{n,m}) = 1$, 考虑验证 Lindeberg 条件. 当 $n \rightarrow \infty$ 时有

$$\begin{aligned} s_n^{-2} \sum_{m=1}^n E[X_{n,m}^2 1_{(|X_{n,m}| > \epsilon s_n)}] &= \sum_{m=1}^n E \left[\left(\frac{X_m - \mu}{\sigma\sqrt{n}} \right)^2 1_{(|\frac{X_m - \mu}{\sigma\sqrt{n}}| > \epsilon)} \right] \\ &= n \left\{ \frac{1}{n\sigma^2} E[(X_1 - \mu)^2 1_{(|X_1 - \mu| > \epsilon\sigma\sqrt{n})}] \right\} \\ &= \sigma^{-2} E[(X_1 - \mu)^2 1_{(|X_1 - \mu| > \epsilon\sigma\sqrt{n})}] \rightarrow 0 \end{aligned}$$

根据 Lindeberg-Feller 定理可知结论成立.

注 无需验证 Lindeberg 条件也能证明该定理, 首先可以不失一般性地假定 $\mu = 0$, 只需证明当 $n \rightarrow \infty$ 时有

$$E[\exp(itS_n/\sigma n^{1/2})] = \left[1 - \frac{t^2}{2n} + o(n^{-1}) \right]^n \rightarrow e^{-t^2/2}$$

其中 $o(n^{-1})$ 是 n^{-1} 的高阶无穷小量.

推论 3.3 (De Moivre-Laplace 定理)

设 X_1, X_2, \dots 是 i.i.d. 随机变量, $P(X_1 = 1) = P(X_1 = -1) = 1/2$, 令 $S_n = X_1 + \cdots + X_n$, 如果 $a < b$, 那么当 $n \rightarrow \infty$ 时有

$$P(a \leq S_n/\sqrt{n} \leq b) \rightarrow \int_a^b (2\pi)^{-1/2} e^{-x^2/2} dx$$



示例 3.19 Lindeberg 条件的非必要性 设 X_1, X_2, \dots 为独立随机变量, 对于任意 $m \geq 2$ 都有 $P(X_m = m) = P(X_m = -m) = m^{-2}/2$, 并且

$$P(X_m = 1) = P(X_m = -1) = (1 - m^{-2})/2$$

令 $S_n = X_1 + \dots + X_n$, 当 $n \rightarrow \infty$ 时有 $S_n/\sqrt{n} \xrightarrow{d} N(0, 1)$.

证明 定义 $Y_m = 1_{(X_m > 0)} + 1_{(X_m \leq 0)}$, 于是 $E(Y_m) = 0$, $\text{Var}(X_m) = 1$, 并且 Y_1, Y_2, \dots 是 i.i.d. 随机变量, 根据 Lindeberg-Lévy 定理可知

$$S_n/\sqrt{n} \xrightarrow{d} N(0, 1)$$

注意到

$$P(X_m \neq Y_m) = P(X_m = \pm m) = m^{-2}$$

由于 $\sum_{m=1}^{\infty} P(X_m \neq Y_m) < \infty$, 根据 Borel-Cantelli 第一引理可知 $P(X_m \neq Y_m \text{ i.o.}) = 0$, 也即当 $n \rightarrow \infty$ 时有 $S_n - \sum_{m=1}^n Y_m \xrightarrow{p} 0$, 因此 Slutsky 定理表明

$$\frac{S_n}{\sqrt{n}} = \frac{\sum_{m=1}^n Y_m}{\sqrt{n}} + \frac{S_n - \sum_{m=1}^n Y_m}{\sqrt{n}} \xrightarrow{d} N(0, 1)$$

然而如果令 $X_{n,m} = X_m/\sqrt{n}$, 易知 $E(X_m) = 0$ 且 $E(X_m^2) = 2 - m^{-2}$, 于是

$$E(X_{n,m}^2) = n^{-1}(2 - m^{-2})$$

令 $s_n^2 = \sum_{m=1}^n E(X_{n,m}^2)$, 显然 $s_n^2 \leq 2$, 并且

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n \frac{2 - m^{-2}}{n} = 2$$

任取 $0 < \epsilon < 1/2$ 可知 Lindeberg 条件不成立.

示例 3.20 设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = 0$ 且 $E(X_i^2) = \sigma^2 \in (0, \infty)$, 那么当 $n \rightarrow \infty$ 时有

$$\sum_{m=1}^n X_m / \left(\sum_{m=1}^n X_m^2 \right)^{1/2} \xrightarrow{d} N(0, 1)$$

证明 根据 Lindeberg-Lévy 定理可知当 $n \rightarrow \infty$ 时有

$$(X_1 + \dots + X_n)/\sigma\sqrt{n} \rightarrow N(0, 1)$$

再由弱大数定律知

$$\frac{X_1^2 + \dots + X_n^2}{n} \xrightarrow{p} \sigma^2$$

最后由连续映射定理和 Slutsky 定理即可得出结论.

示例 3.21 设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = 0$, $E(X_i^2) = \sigma^2 \in (0, \infty)$, 再设 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时, $\limsup S_n/\sqrt{n} = \infty$ 几乎必然成立.

证明 由于当 $n \rightarrow \infty$ 时有 $S_n/\sigma\sqrt{n} \xrightarrow{d} \chi \sim N(0, 1)$, 对于任意 $a \in \mathbb{R}$ 都有

$$\limsup_{n \rightarrow \infty} P\left(\frac{S_n}{n} > a\right) = P(\chi > a/\sigma) > 0$$

因此

$$\limsup_{n \rightarrow \infty} P\left(\frac{S_n}{\sqrt{n}} > a \text{ i.o.}\right) \geq \limsup_{n \rightarrow \infty} P\left(\frac{S_n}{\sqrt{n}} > a\right) > 0$$

注意到 $\{\limsup_{n \rightarrow \infty} S_n/\sqrt{n} > a\}$ 是一个尾事件, 由 Kolmogorov 0-1 律知结论成立.

示例 3.22 设 X_1, X_2, \dots 为 i.i.d. 随机变量, $X_i \geq 0$, $E(X_i) = 1$ 且 $\text{Var}(X_i) = \sigma^2 \in (0, \infty)$, 记 $S_n = X_1 + \dots + X_n$, 那么当 $n \rightarrow \infty$ 时有 $2(\sqrt{S_n} - \sqrt{n}) \xrightarrow{d} N(0, \sigma^2)$.

证明 根据 Lindeberg-Lévy 定理可知

$$\frac{S_n - E(S_n)}{\sigma \sqrt{n}} = \frac{S_n - n}{\sigma \sqrt{n}} \xrightarrow{d} N(0, 1)$$

注意到

$$\frac{S_n - n}{\sigma \sqrt{n}} = \frac{\sqrt{S_n} + \sqrt{n}}{\sigma \sqrt{n}} (\sqrt{S_n} - \sqrt{n})$$

根据弱大数定律有 $S_n/n \xrightarrow{p} 1$, 再由连续映射定理知

$$\sigma^{-1} \left(\sqrt{\frac{S_n}{n}} + 1 \right) \xrightarrow{p} 2\sigma^{-1}$$

最后由 Slutsky 定理即可推得结论.

3.4.2 Berry-Esseen 定理

中心极限定理表明了一定条件下, S_n/s_n 在 $n \rightarrow \infty$ 时依分布收敛于标准正态分布, 本节的 Berry-Esseen 定理则提供了收敛率的一个估计.

引理 3.6

设 F 和 G 为 c.d.f., 并且 $G'(x) \leq \lambda < \infty$. 令 $\Delta(x) = F(x) - G(x)$, $\eta = \sup |\Delta(x)|$, $\Delta_L = \Delta * H_L$, $\eta_L = \sup |\Delta_L(x)|$, 那么

$$\eta_L \geq \frac{\eta}{2} - \frac{12\lambda}{\pi L}$$

其中 H_L 是概率密度为

$$h_L(x) = \frac{1 - \cos Lx}{\pi Lx^2}$$

的随机变量的 c.d.f.



证明 注意到 Δ 在无穷远点为 0, 并且单侧极限 $\Delta(x+)$ 和 $\Delta(x-)$ 处处存在, 因此存在点 x_0 , 要么 $|\Delta(x_0+)| = \eta$, 要么 $|\Delta(x_0-)| = \eta$, 故而可以假设 $\Delta(x_0) = \eta$.

因为 $G'(x) \leq \lambda$ 并且 F 不减, 所以对任意 $s > 0$ 都有 $\Delta(x_0 + s) \geq \eta - \lambda s$. 现在令 $\delta = \eta/2\lambda$, $t = x_0 + \delta$, 于是

$$\Delta(t - x) \geq \begin{cases} \eta/2 + \lambda x, & |x| \leq \delta \\ -\eta, & |x| > \delta \end{cases}$$

注意到

$$2 \int_{\delta}^{\infty} h_L(x) dx \leq 2 \int_{\delta}^{\infty} 2/(\pi L x^2) dx = 4/(\pi L \delta)$$

并且 $\int_{-\delta}^{\delta} x h_L(x) dx = 0$, 因此

$$\begin{aligned} \eta_L \geq \Delta_L(t) &= \int_{-\infty}^{\infty} \Delta(t-x) h_L(x) dx \\ &\geq \frac{\eta}{2} \int_{-\delta}^{\delta} h_L(x) dx - 2\eta \int_{\delta}^{\infty} h_L(x) dx \\ &= \frac{\eta}{2} - \frac{6\eta}{\pi L \delta} = \frac{\eta}{2} - \frac{12\lambda}{\pi L} \end{aligned}$$

引理 3.7

设 K_1, K_2 为分布函数, 其对应的 c.f. 分别为 κ_1, κ_2 , 那么

$$K_1(x) - K_2(x) = \frac{1}{2\pi} \int -e^{-itx} \frac{\kappa_1(t) - \kappa_2(t)}{it} dt$$



证明 因为 κ_i 是可积的, 根据 Feller 反演公式可知概率密度 $k_i(x)$ 为

$$k_i(y) = \frac{1}{2\pi} \int e^{-ity} \kappa_i(t) dt$$

用 k_2 减去 k_1 , 从 a 到 x 积分, 再令 $\Delta K = K_1 - K_2$ 可得

$$\begin{aligned} \Delta K(x) - \Delta K(a) &= \frac{1}{2\pi} \int_a^x \int e^{-ity} [\kappa_1(t) - \kappa_2(t)] dt dy \\ &= \frac{1}{2\pi} \int (e^{-ita} - e^{-itx}) \frac{\kappa_1(t) - \kappa_2(t)}{it} dt \end{aligned}$$

由于 $\kappa_i(0) = 1$ 且 $\kappa_i'(0) = 0$, 故而被积函数在原点处消失. 最后令 $a \rightarrow -\infty$, 根据 Riemann-Lebesgue 引理即可证得结论.

现在设 φ_F 和 φ_G 分别为 F 和 G 的 c.f., ω_L 为 H_L 的 c.f., 再令 $F_L = F * H_L$ 和 $G_L = G * H_L$, 根据引理 3.7 可知

$$\begin{aligned} |F_L(x) - G_L(x)| &\leq \frac{1}{2\pi} \int \left| \frac{\varphi_F(t)\omega_L(t) - \varphi_G(t)\omega_L(t)}{t} \right| dt \\ &\leq \frac{1}{2\pi} \int_{-L}^L \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt \end{aligned}$$

因为 $|\omega_L(t)| \leq 1$, 由引理 3.6 可知

$$|F(x) - G(x)| \leq \frac{1}{\pi} \int_{-L}^L \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt + \frac{24\lambda}{\pi L} \quad (3.26)$$

称为 Esseen 平滑不等式.

引理 3.8

设 X 为一个随机变量, $E(X) = 0$, $E(X^2) = \sigma^2 < \infty$, $E|X^3| = \gamma < \infty$, 再设 X 的 c.f. 为 $\varphi(t)$, 那么对于任意 $t \in [-1/\sigma, 1/\sigma]$ 都有

$$\left| \log \varphi(t) + \frac{t^2 \sigma^2}{2} \right| \leq \frac{5}{12} |t|^3 \gamma \quad (3.27)$$



证明 根据引理3.1可知, 当 $|t| \leq \sigma^{-1}$ 时有

$$|\varphi(t) - 1| = |E(e^{itX}) - 1 - itX| \leq \frac{t^2\sigma^2}{2} \leq \frac{1}{2} \quad (3.28)$$

又因为对于复数 $z \in \mathbb{C}$ 有

$$\log(1+z) = \sum_{k=1}^{\infty} z^k/k, \quad |z| < 1$$

又由引理3.1可知

$$\begin{aligned} \left| \log \varphi(t) + \frac{t^2\sigma^2}{2} \right| &= \left| \log\{1 + [\varphi(t) - 1]\} + \frac{t^2\sigma^2}{2} \right| \\ &\leq \left| \varphi(t) - \left(1 - \frac{t^2\sigma^2}{2}\right) \right| + \sum_{k=2}^{\infty} |\varphi(t) - 1|^k/k \\ &\leq E \left| (tX)^2 \wedge \frac{|tX|^3}{6} \right| + \frac{1}{2} \left(\frac{t^2\sigma^2}{2} \right)^2 \sum_{k=2}^{\infty} 2^{2-k} \\ &\leq \frac{|t|^3\gamma}{6} + \frac{t^4\sigma^4}{4} \end{aligned}$$

根据 $|t\sigma| \leq 1$, 以及

$$\sigma^3 = [E(Y^2)]^{3/2} \leq E|Y|^3 = \gamma$$

可知式 (3.27) 成立.

定理 3.18 (Berry-Esseen 定理)

设 X_1, X_2, \dots 为 i.i.d. 随机变量, $E(X_i) = 0$, $E(X_i^2) = \sigma^2 < \infty$ 且 $E|X_i|^3 = \gamma < \infty$, 令 $F_n(x)$ 是 $(X_1 + \dots + X_n)/\sigma\sqrt{n}$ 的分布函数, 那么

$$|F_n(x) - \Phi(x)| \leq \frac{C\gamma}{\sigma^3\sqrt{n}}$$

其中 $0 < C < \infty$ 为一个常数.



证明 不失一般性可以假设 $\sigma^2 = 1$, 在式 (3.26) 中取 $\lambda = 1/\sqrt{2\pi}$, 并且分别用 F_n 和 Φ 替代 F 和 Φ , 再取 $L = \sqrt{n}/\gamma$, 于是

$$|F_n(x) - \Phi(x)| \leq \frac{1}{\pi} \int_{-L}^L \left| \frac{\varphi^n(t/\sqrt{n}) - e^{-t^2/2}}{t} \right| + \frac{24\gamma}{\pi\sqrt{2\pi n}} \quad (3.29)$$

其中 φ 为 X_i 的 c.f. 再由引理3.8可知, 对一切 $|t| \leq \sqrt{n}$ 都有

$$\begin{aligned} r_n(t) &= \left| n \log \varphi \left(\frac{t}{\sqrt{n}} \right) + \frac{t^2}{2} \right| \\ &= n \left| \log \varphi \left(\frac{t}{\sqrt{n}} \right) + \frac{\sigma^2}{2} \left(\frac{t}{\sqrt{n}} \right)^2 \right| \\ &\leq \frac{5|t|^3\gamma}{12\sqrt{n}} \end{aligned} \quad (3.30)$$

因为 $\gamma = E|X_1|^3 \geq [E(X_1^2)]^{3/2} = 1$, $|L| \leq \sqrt{n}$, 根据复数不等式 $|e^z - 1| \leq |z|e^{|z|}$ 和式 (3.30)

可知

$$\begin{aligned}
 \left| \varphi^n \left(\frac{t}{\sqrt{n}} \right) - e^{-t^2/2} \right| &= \left| \exp \left[n \log \varphi \left(\frac{t}{\sqrt{n}} \right) + \frac{t^2}{2} \right] - 1 \right| \exp \left(-\frac{t^2}{2} \right) \\
 &\leq |r_n(t)| \exp \left(|r_n(t)| - \frac{t^2}{2} \right) \\
 &\leq \frac{5\gamma}{12\sqrt{n}} |t|^3 \exp \left[-\frac{t^2}{2} \left(1 - \frac{5|t|\gamma}{6\sqrt{n}} \right) \right] \\
 &\leq \frac{5\gamma}{12\sqrt{n}} |t|^3 \exp \left(-\frac{t^2}{12} \right)
 \end{aligned} \tag{3.31}$$

对一切 $|t|\gamma \leq \sqrt{n}$ 成立. 注意到 $\int_{-\infty}^{\infty} t^2 e^{-t^2/12} dt = 6\sqrt{2\pi}$, 于是根据式 (3.29) 和 (3.31) 可知, 定理在

$$C = \sqrt{\frac{2}{\pi}} \left[\frac{5}{2} + \frac{12}{\pi} \right] < 5.05$$

的情况下成立.

注 事实上, 常数 C 的取值范围至少已经缩小到了 (0.4097, 0.7655).

3.4.3 Poisson 收敛

引理 3.9

设 $b \in \mathbb{C}$ 满足 $|z| \leq 1$, 那么 $|e^b - (1+b)| \leq |b|^2$.



证明 根据展开式

$$e^b - (1+b) = \frac{b^2}{2!} + \frac{b^3}{3!} + \cdots$$

当 $|b| \leq 1$ 时有

$$|e^b - (1+b)| \leq \frac{|b|^2}{2} (1 + 1/2 + 1/2^2 + \cdots) = |b|^2$$

结论成立.

定理 3.19

假设对于每个 $n \geq 1$, $1 \leq m \leq n$, 随机变量 $X_{n,m}$ 是独立的, $P(X_{n,m} = 1) = p_{n,m}$, $P(X_{n,m} = 0) = 1 - p_{n,m}$, 并且以下条件成立

(1) $\sum_{m=1}^n p_{n,m} \rightarrow \lambda \in (0, \infty)$;

(2) $\max_{1 \leq m \leq n} p_{n,m} \rightarrow 0$;

再设 $S_n = X_{n,1} + \cdots + X_{n,n}$, 那么当 $n \rightarrow \infty$ 时有 $S_n \xrightarrow{d} P(\lambda)$.



证明 设 $\varphi_{n,m}(t) = E(e^{itX_{n,m}}) = (1 - p_{n,m}) + p_{n,m}e^{it}$, 以及 $S_n = X_{n,1} + \cdots + X_{n,n}$, 于是

$$E(e^{itS_n}) = \prod_{m=1}^n [1 + p_{n,m}(e^{it} - 1)]$$

令 $0 \leq p \leq 1$, 于是

$$\begin{aligned} |\exp[p(e^{it} - 1)]| &= \exp[p\operatorname{Re}(e^{it} - 1)] \leq 1 \\ |1 + p(e^{it} - 1)| &\leq 1 \end{aligned}$$

根据引理3.4和3.9可知, 当 $\max_m p_{n,m} \leq 1/2$ 时有

$$\begin{aligned} &\left| \exp \left[\sum_{m=1}^n p_{n,m}(e^{it}) \right] - \prod_{m=1}^n [1 + p_{n,m}(e^{it} - 1)] \right| \\ &\leq \sum_{m=1}^n |\exp[p_{n,m}(e^{it} - 1)] - [1 + p_{n,m}(e^{it} - 1)]| \\ &\leq \sum_{m=1}^n p_{n,m}^2 |e^{it} - 1|^2 \end{aligned}$$

又根据 $|e^{it} - 1| \leq 2$, 以上最后一个表达式小于等于

$$4 \max_{1 \leq m \leq n} p_{n,m} \sum_{m=1}^n p_{n,m} \rightarrow 0, \quad n \rightarrow \infty$$

又因为 $\sum_{m=1}^n p_{n,m} \rightarrow \lambda$, 故而

$$E(e^{itS_n}) \rightarrow e^{\lambda(e^{it}-1)}$$

根据 Poisson 分布的 c.f. 及 Lévy-Cramér 连续定理知结论成立.

示例 3.23 假设我们掷两枚骰子 36 次, 则掷出两个 1 的概率为 $1/36$, 并且这种情况发生的次数近似服从均值为 1 的 Poisson 分布, 它们的关系如下

k	0	1	2	3
Poisson	0.3678	0.3678	0.1839	0.0613
精确值	0.3627	0.3730	0.1865	0.0604

第4章 条件期望与条件概率

4.1 条件期望

4.1.1 存在性与唯一性

定义 4.1

设 $(\Omega, \mathcal{F}_0, P)$ 为一个概率空间, σ -代数 $\mathcal{F} \subset \mathcal{F}_0$, 随机变量 $X \in \mathcal{F}_0$ 且 $E|X| < \infty$, 那么给定 \mathcal{F} 时 X 的条件期望 (Conditional Expectation, CE) 记作 $E(X|\mathcal{F})$, 它是任意满足以下条件的随机变量 Y

- (1) $Y \in \mathcal{F}$, 也即 Y 是 \mathcal{F} -可测的;
- (2) 对任意 $A \in \mathcal{F}$, $\int_A X dP = \int_A Y dP$.



注 令 $A = \Omega$, 根据定义可知 $E(X) = E[E(X|\mathcal{F})]$, 称为迭代期望定律 (law of iterated expectation).

现在的第一件事是证明条件期望 $E(X|\mathcal{F})$ 的存在性和唯一性, 这里将先证明唯一性, 为此需要一些引理.

引理 4.1

如果随机变量 Y 满足定义 4.1 的条件 (1) 和 (2), 那么 Y 是可积的.



证明 设 $A = \{Y > 0\} \in \mathcal{F}$, 于是

$$\begin{aligned}\int_A Y dP &= \int_A X dP \leq \int_A |X| dP \\ \int_{A^c} -Y dP &= \int_{A^c} -X dP \leq \int_{A^c} |X| dP\end{aligned}$$

因此 $E|Y| \leq E|X| < \infty$.

现在来证明条件期望的唯一性, 假设存在另一个 Y' 满足条件 (1) 和 (2), 那么对于一切 $A \in \mathcal{F}$ 都有

$$\int_A Y dP = \int_A Y' dP$$

任取 $A = \{Y - Y' \geq \epsilon > 0\}$, 于是

$$0 = \int_A X - X dP = \int_A Y - Y' dP \geq \epsilon P(A)$$

也即 $P(A) = 0$. 因为这对任意 $\epsilon > 0$ 成立, 故而 $Y \leq Y'$ a.s., 在上式中交换 Y 与 Y' 的位置就得到 $Y = Y'$ a.s. 技术上来说, 所有形如 $Y = E(X|\mathcal{F})$ 的等式都该写作 $Y = E(X|\mathcal{F})$ a.s., 但是为了简单起见而将 a.s. 省去.

定理 4.1

如果在 $B \in \mathcal{F}$ 上有 $X_1 = X_2$, 那么在 B 上 $E(X_1|\mathcal{F}) = E(X_2|\mathcal{F})$ a.s.



证明 令 $Y_1 = E(X_1|\mathcal{F})$, $Y_2 = E(X_2|\mathcal{F})$, $A = \{Y_1 - Y_2 \geq \epsilon > 0\}$, 于是

$$0 = \int_{A \cap B} X_1 - X_2 dP = \int_{A \cap B} Y_1 - Y_2 dP \geq \epsilon P(A \cap B)$$

因此 $P(A \cap B) = 0$, 根据之前的做法即可证得结论.

继续证明条件期望的存在性, 首先假设 $X \geq 0$, 设 $\mu = P$, 以及

$$\nu(A) = \int_A X dP, \quad \forall A \in \mathcal{F}$$

根据 DCT 可知 ν 是一个测度, 并且由上式可知 $\nu \ll \mu$, 根据 Radon-Nikodym 定理, 存在

$$h = \frac{d\nu}{d\mu} \in \mathcal{F}$$

使得对于任意 $A \in \mathcal{F}$ 都有

$$\int_A X dP = \nu(A) = \int_A \frac{d\nu}{d\mu} dP$$

现在考虑一般情形, 设 $X = X^+ - X^-$, 并且令 $Y_1 = E(X^+|\mathcal{F})$, $Y_2 = E(X^-|\mathcal{F})$, 从而 $Y_1 - Y_2$ 是可积的, 于是

$$\begin{aligned} \int_A X dP &= \int_A X^+ dP - \int_A X^- dP \\ &= \int_A Y_1 dP - \int_A Y_2 dP = \int_A Y_1 - Y_2 dP \end{aligned}$$

从而 $Y_1 - Y_2$ 是 $E(X|\mathcal{F})$ 的一种形式, 也即条件期望 $E(X|\mathcal{F})$ 存在.

4.1.2 示例

直观上看, \mathcal{F} 描述了我们所掌握的信息, 对于任意 $A \in \mathcal{F}$, 我们能知道它发生与否, 于是 $E(X|\mathcal{F})$ 是在给定信息后, 我们能对 X 的值所作的最优猜测 (best guess).

示例 4.1 如果 $X \in \mathcal{F}$, 那么 $E(X|\mathcal{F}) = X$. 换言之, 如果我们已经知道 X , 那么对 X 的最优猜测即为 X 自身. 它的证明是容易的, 因为 X 总是满足条件 (2), 唯一能阻止 X 为 $E(X|\mathcal{F})$ 的是条件 (1), 但这个条件也已被 $X \in \mathcal{F}$ 所满足.

示例 4.2 假设 X 独立于 \mathcal{F} , 也即对于任意 $A \in \mathcal{F}$ 和 $B \in \mathcal{R}$ 都有

$$P(\{X \in B\} \cap A) = P(X \in B)P(A)$$

那么此时 $E(X|\mathcal{F}) = E(X)$. 换言之, 如果你不知道关于 X 的任何信息, 那么对 X 的最优猜测就是它的期望 $E(X)$.

显然, $E(X) \in \mathcal{F}$, 故而条件 (1) 成立. 为了验证条件 (2), 注意到如果 $A \in \mathcal{F}$, 那么 X 和 1_A 是独立的, 根据定理 2.3 可知

$$\int_A X dP = E(X1_A) = E(X)E(1_A) = \int_A E(X) dP$$

示例 4.3 假设 $\Omega_1, \Omega_2, \dots$ 是 Ω 的一个有限或无限的分割, $\bigcup_i \Omega_i = \Omega$, 并且每一个都具有正概

率, 令 $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \dots)$ 为这些集合生成的 σ -代数, 那么在每个 Ω_i 上都有

$$E(X|\mathcal{F}) = \frac{E(X; \Omega_i)}{P(\Omega_i)}$$

注意到上式在每个 Ω_i 上都是常数, 故而是 \mathcal{F} -可测的, 另一方面

$$\int_{\Omega_i} \frac{E(X; \Omega_i)}{P(\Omega_i)} dP = E(X; \Omega_i) = \int_{\Omega_i} X dP$$

当 \mathcal{F} 退化为 $\{\emptyset, \Omega\}$ 时就得到 $E(X|\mathcal{F}) = E(X)$.

为了与初等概率论的符号联系起来, 定义给定 \mathcal{G} 时事件 A 发生的条件概率 (conditional probability)

$$\begin{aligned} P(A|\mathcal{G}) &= E(1_A|\mathcal{G}) \\ P(A|B) &= P(A \cap B)/P(B) \end{aligned} \quad (4.1)$$

此外, 给定 σ -代数的条件期望的定义, 涵盖了以随机变量为条件的这一特殊情形, 因此再定义

$$E(X|Y) = E[X|\sigma(Y)]$$

其中 $\sigma(Y)$ 是由 Y 生成的 σ -代数.

假设 $A_1, A_2 \in \mathcal{G}$ 是两个不相交的集合, 容易证明

$$P(A_1 \cup A_2|\mathcal{G}) = P(A_1|\mathcal{G}) + P(A_2|\mathcal{G}) \quad \text{a.s.}$$

这似乎意味着, 存在集合 $A_0 \in \mathcal{G}$ 使得 $P(A_0) = 0$, 并且对于所有 $\omega \notin A_0$ 都有

$$P\left(\bigcup_{n=1}^{\infty} A_n \middle| \mathcal{G}\right) = \sum_{n=1}^{\infty} P(A_n|\mathcal{G})(\omega)$$

然而这一论断并不成立, 这是因为 A_0 依赖于 $\{A_n\}_{n \geq 1}$, 当这个集族变化时, 异常集可以是不可数的, 并且它们的并集不包含于零概率集之中.

示例 4.4 Bayes 公式 设 $G \in \mathcal{G}$, 那么

$$P(G|A) = \int_G P(A|\mathcal{G}) dP \bigg/ \int_{\Omega} P(A|\mathcal{G}) dP$$

证明 根据定义可知

$$\begin{aligned} \int_G P(A|\mathcal{G}) dP &= \int_G E(1_A|\mathcal{G}) dP = \int_G 1_A dP = P(A \cap G) \\ \int_{\Omega} P(A|\mathcal{G}) dP &= \int_{\Omega} E(1_A|\mathcal{G}) dP = \int_{\Omega} 1_A dP = P(A) \end{aligned}$$

根据式 (4.1) 可知结论成立.

示例 4.5 假设 (X, Y) 具有联合概率密度 $f(x, y)$, 也即

$$P((X, Y) \in B) = \int_B f(x, y) dx dy, \quad \forall B \in \mathcal{R}^2$$

假设对于一切 y 都有 $\int f(x, y) dx > 0$, 如果 $E|g(X)| < \infty$, 那么 $E[g(X)|Y] = h(Y)$, 其中

$$h(y) = \int g(x) f(x, y) dx \bigg/ \int f(x, y) dx$$

证明 首先注意到 $h(Y) \in \sigma(Y)$, 故而条件 (1) 满足. 另一方面, 如果 $A \in \sigma(Y)$, 那么对某些

$B \in \mathcal{R}$ 有 $A = \{\omega : Y(\omega) \in B\}$, 于是

$$\begin{aligned} E[h(Y); A] &= \int_A h(Y) dP = \int_B h(y) f_Y(y) dy \\ &= \int_B \int g(x) f(x, y) dx dy \\ &= E[g(X)1_B(Y)] = E[g(X); A] \end{aligned}$$

故而条件 (2) 也满足. 最后, 去掉 $\int f(x) dx > 0$ 的步骤与之前类似.

示例 4.6 如果 X 和 Y 是独立的, $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ 满足 $E|\varphi(x, y)| < \infty$, 令 $g(x) = E[\varphi(x, Y)]$, 那么

$$E[\varphi(X, Y)|X] = g(X)$$

证明 易知 $g(X) \in \sigma(X)$, 为了验证条件 (2), 注意到如果 $A \in \sigma(X)$, 那么可以将其设为 $A = \{X \in C\}$, 于是

$$\begin{aligned} \int_A \varphi(X, Y) dP &= E[\varphi(X, Y)1_C(X)] \\ &= \int \left[\int \varphi(x, y)1_C(x) \nu(dy) \right] \mu(dx) \\ &= \int 1_C(x)g(x) \mu(dx) = \int_A g(X) dP \end{aligned}$$

这就得到了希望的结果.

4.1.3 性质

定理 4.2

以下论断中的前两个假设了 $E|X|, E|Y| < \infty$

- (1) $E(aX + Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F})$;
- (2) 如果 $X \leq Y$, 那么 $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$;
- (3) 如果 $X_n \geq 0, X_n \uparrow X$ 且 $E(X) < \infty$, 那么 $E(X_n|\mathcal{F}) \uparrow E(X|\mathcal{F})$.



证明 (1) 易知 $aE(X|\mathcal{F}) + E(Y|\mathcal{F})$ 是 \mathcal{F} -可测的, 注意到对任意 $A \in \mathcal{F}$ 都有

$$\begin{aligned} \int_A [aE(X|\mathcal{F}) + E(Y|\mathcal{F})] dP &= a \int_A E(X|\mathcal{F}) dP + \int_A E(Y|\mathcal{F}) dP \\ &= a \int_A X dP + \int_A Y dP = \int_A aX + Y dP \end{aligned}$$

(2) 对于任意 $A \in \mathcal{F}$, 因为

$$\int_A E(X|\mathcal{F}) dP = \int_A X dP \leq \int_A Y dP \leq \int_A E(Y|\mathcal{F}) dP$$

令 $A = \{E(X|\mathcal{F}) - E(Y|\mathcal{F}) \geq \epsilon > 0\}$, 于是对任意 $\epsilon > 0$ 都有 $P(A) = 0$.

(3) 令 $Y_n = X - X_n$, 故而只需证明 $E(Y_n|\mathcal{F}) \downarrow 0$ 即可. 因为 Y_n 是递减的, 根据 (b) 可知 $Z_n = E(Y_n|\mathcal{F}) \downarrow$ 某个极限 Z_∞ . 如果 $A \in \mathcal{F}$, 那么

$$\int_A Z_n dP = \int_A Y_n dP$$

现在令 $n \rightarrow \infty$, 于是 $Y_n \downarrow 0$, 根据 DCT 可知 $\int_A Z_\infty dP = 0$ 对一切 $A \in \mathcal{F}$ 成立, 从而 $Z_\infty = 0$.

注 将 (c) 部分施加到 $Y_1 - Y_n$, 因此如果 $Y_n \downarrow Y$, 那么当 $E|Y_1|, E|Y| < \infty$ 时有 $E(Y_n|\mathcal{F}) \downarrow E(Y|\mathcal{F})$.

定理 4.3

设子 σ -代数 $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_0$, 并且 $E|X| < \infty$, 那么

$$(1) E(X|\mathcal{F}_1) = E[E(X|\mathcal{F}_2)|\mathcal{F}_1];$$

$$(2) E(X|\mathcal{F}_1) = E[E(X|\mathcal{F}_2)|\mathcal{F}_1].$$



证明 (1) 示例 4.1 的直接结果.

(2) 令 $A \in \mathcal{F}_1$, $Z_1 = E(X|\mathcal{F}_1)$, $Z_2 = E(X|\mathcal{F}_2)$, 于是 $\int_A X dP = \int_A Z_1 dP$, 又因为 $\mathcal{F}_1 \subset \mathcal{F}_2$, $A \in \mathcal{F}_2$, 从而 $\int_A X dP = \int_A Z_2 dP$, 也即对任何 $A \in \mathcal{F}_1$ 都有

$$\int_A Z_1 dP = \int_A Z_2 dP$$

又根据 $E(Z_2|\mathcal{F}_1)$ 的定义, 于是 $Z_1 = E(Z_2|\mathcal{F}_1)$, 这就证明了结论.

定理 4.4

如果 $X \in \mathcal{F}$, $E|X|, E|XY| < \infty$, 那么

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F}) \quad (4.2)$$



证明 显然上式右端 $\in \mathcal{F}$, 故而只需验证条件 (2) 即可. 首先令 $X = 1_B$, $B \in \mathcal{F}$, 此时对于任意 $A \in \mathcal{F}$ 都有

$$\int_A 1_B E(Y|\mathcal{F}) dP = \int_{A \cap B} E(Y|\mathcal{F}) dP = \int_{A \cap B} Y dP = \int_A 1_B Y dP$$

根据 CE 的线性性质可知上面的结果对简单函数 X 成立. 如果 $X, Y \geq 0$, 令 X_n 为简单随机变量并且当 $n \rightarrow \infty$ 时有 $X_n \uparrow X$, 从而

$$\int_A X E(Y|\mathcal{F}) dP = \int_A X Y dP$$

最后将 X 和 Y 分割为正负部分即可证得结论.

定理 4.5 (Jensen 不等式)

如果 φ 是凸的, $E|X|, E|\varphi(X)| < \infty$, 那么

$$\varphi[E(X|\mathcal{F})] \leq E[\varphi(X)|\mathcal{F}] \quad (4.3)$$



证明 根据 φ 在 (a, b) 上的凸性, 对于任意 $c, x \in (a, b)$ 都有

$$\varphi(x) - \varphi(c) - (x - c)\varphi'_-(c) \geq 0 \quad (4.4)$$

其中 φ'_- 是 φ 在 c 处的左导数. 取 $c = E(X|\mathcal{F})$, $x = X$, 于是

$$Z = \varphi(Y) - \varphi[E(X|\mathcal{F})] - [X - E(X|\mathcal{F})]\varphi'_-[E(X|\mathcal{F})] \geq 0 \quad (4.5)$$

因为 $E(\varphi[E(X|\mathcal{F})]|\mathcal{F}) = \varphi[E(X|\mathcal{F})]$, 根据式 (4.2) 可知

$$E(\{[X - E(X|\mathcal{F})]\varphi'_-[E(X|\mathcal{F})]\}|\mathcal{F}) = \varphi'_-[E(X|\mathcal{F})]E\{[X - E(X|\mathcal{F})]|\mathcal{F}\} = 0$$

又由于 $E(Z|\mathcal{F}) \geq 0$, 所以式 (4.3) 成立.

定理 4.6 (Chebyshev 不等式)

如果 $a > 0$, 那么

$$P(|X| > a|\mathcal{F}) \leq a^{-2}E(X^2|\mathcal{F})$$




证明 对于任意 $A \in \mathcal{F}$ 有

$$\int_A P(|X| > a|\mathcal{F}) = \int_A 1_{(|X|>a)} dP \leq \int_A \frac{|X|^2}{a^2} dP = a^{-2} \int_A E(|X|^2|\mathcal{F}) dP$$

因为 $A \in \mathcal{F}$ 是任意的, 所以结论成立.

定理 4.7

如果 $E(X^2) < \infty$, 那么 $E(X|\mathcal{F})$ 是一个最小化均方误差 $E(X - Y)^2$ 的随机变量 $Y \in \mathcal{F}$. 

证明 设 $L^2(\mathcal{F}_0) = \{Y \in \mathcal{F}_0 : E(Y^2) < \infty\}$ 为一个 Hilbert 空间, $L^2(\mathcal{F})$ 是它的一个闭的子空间. 如果 $Z \in L^2(\mathcal{F})$, 那么由 Cauchy-Schwarz 不等式可知

$$E|XZ| \leq (E|X|^2)^{1/2}(E|Z|^2)^{1/2} < \infty$$

进而由定理 4.4 可知

$$ZE(X|\mathcal{F}) = E(ZX|\mathcal{F})$$

两端取期望得

$$E[ZE(X|\mathcal{F})] = E[E(ZX|\mathcal{F})] = E(ZX)$$

也即

$$E\{Z[X - E(X|\mathcal{F})]\} = 0$$

如果 $Y \in L^2(\mathcal{F})$, $Z = E(X|\mathcal{F}) - Y$, 那么

$$E(X - Y)^2 = E[X - E(X|\mathcal{F}) + Z]^2 = E[X - E(X|\mathcal{F})]^2 + E(Z^2)$$

易知上式在 $Z = 0$ 时取得最小值.

定理 4.8

如果 $E(X^2) < \infty$, 那么

$$\text{Var}(X) = \text{Var}[E(X|\mathcal{F})] + E[\text{Var}(X|\mathcal{F})]$$



证明 令 $\mu = E(X)$, 于是 $\text{Var}(X) = E(X - \mu)^2$, $X - \mu = X - E(X|\mathcal{F}) + E(X|\mathcal{F}) - \mu$, 进而

$$\begin{aligned} E\{[X - E(X|\mathcal{F})][E(X|\mathcal{F}) - \mu]|\mathcal{F}\} &= [E(X|\mathcal{F}) - \mu]E\{[X - E(X|\mathcal{F})]|\mathcal{F}\} \\ &= [E(X|\mathcal{F}) - \mu]E\{[X - E(X|\mathcal{F})]|\mathcal{F}\} = 0 \end{aligned}$$

以及

$$E(X - \mu)^2 = E[X - E(X|\mathcal{F})]^2 + E[E(X|\mathcal{F}) - \mu]^2 \quad (4.6)$$

注意到 $E[XE(X|\mathcal{F})] = E\{E[X(\mu|\mathcal{F})|\mathcal{F}]\} = E[E(X|\mathcal{F})]^2$, 于是

$$\begin{aligned} E[X - E(X|\mathcal{F})]^2 &= E(X^2) - 2E[XE(X|\mathcal{F})] + E[E(X|\mathcal{F})]^2 \\ &= E(X^2) - E\{[E(X|\mathcal{F})]^2\} \\ &= E[E(X^2|\mathcal{F})] - E\{[E(X|\mathcal{F})]^2\} \\ &= E[\text{Var}(X|\mathcal{F})] \end{aligned}$$

因为 $E[E(X|\mathcal{F})] = \mu$, 从而

$$E[E(X|\mathcal{F}) - \mu]^2 = \text{Var}[E(X|\mathcal{F})]$$

根据式 (4.6) 即可证得结论.

示例 4.7 如果随机变量 X 和 Y 满足 $E(Y|\mathcal{G}) = X$, $E(Y^2) = E(X^2) < \infty$, 那么 $X = Y$ a.s.

证明 设 $\mu = E(X)$, $\nu = E(Y)$, 在定理 4.8 中取 $\mathcal{G} = \{\emptyset, \Omega\}$ 可知

$$E(Y - X)^2 + E(X - \mu)^2 = E(Y - \nu)^2$$

根据 $E(Y|\mathcal{G}) = X$ 可得 $E(Y) = E(X)$, 于是

$$E(X - \mu)^2 = E(X^2) - E(X)^2 = E(Y^2) - E(Y)^2 = E(Y - \nu)^2$$

从而 $E(Y - X)^2 = 0$, 也即 $Y = X$ a.s.

4.2 正则条件概率

定义 4.2

设 (Ω, \mathcal{F}, P) 为概率空间, $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ 为可测映射, 子 σ -代数 $\mathcal{G} \subset \mathcal{F}$, 称 $\mu : \Omega \times \mathcal{S} \rightarrow [0, 1]$ 为给定 \mathcal{G} 时 X 的正则条件分布 (regular conditional distribution, r.c.d.), 如果以下条件成立

- 对于任意 $A \in \mathcal{S}$, 映射 $\omega \rightarrow \mu(\omega, A)$ 以概率 1 等于 $P(X \in A|\mathcal{G})$;
- 对于几乎处处 $\omega \in \Omega$, 映射 $A \rightarrow \mu(\omega, A)$ 是 (S, \mathcal{S}) 上的概率测度.

如果 $S = \Omega$, X 为恒等映射, 那么 μ 称为正则条件概率 (regular conditional probability).



示例 4.8 假设 X 和 Y 具有联合概率密度 $f(x, y) > 0$, 如果

$$\mu(y, A) = \int_A f(x, y) dx / \int f(x, y) dx$$

那么 $\mu(Y(\omega), A)$ 是给定 $\sigma(Y)$ 时 X 的 r.c.d.

定理 4.9

设 $\mu(\omega, A)$ 是给定 \mathcal{F} 时 X 的 r.c.d., 如果 $f : (S, \mathcal{S}) \rightarrow (\mathbb{R}, \mathcal{R})$ 满足 $E|f(X)| < \infty$, 那么

$$E[f(X)|\mathcal{F}] = \int \mu(\omega, dx) f(x) \quad \text{a.s.}$$



证明 如果 $f = 1_A$, 则由定义证得结论, 然后根据线性性质和 MCT 拓展到非负可测函数 f , 最

后由 $f = f^+ - f^-$ 即可证得一般结论.

定义 4.3

可测空间 (S, \mathcal{S}) 是好的 (nice), 如果存在 1-1 映射 $\varphi : S \rightarrow \mathbb{R}$, 使得 φ 和 φ^{-1} 都是可测的.



定理 4.10

设 (Ω, \mathcal{F}, P) 为概率空间, X 和 Y 在好的可测空间 (S, \mathcal{S}) 中取值并且 $\mathcal{G} = \sigma(Y)$, 那么存在 $\mu : S \times S \rightarrow [0, 1]$ 使得

- (1) 对于每一个 $A \in \mathcal{S}$, $\mu(Y(\omega), A)$ 以概率 1 等于 $P(X \in A | \mathcal{G})$;
- (2) 对于几乎处处 $\omega \in \Omega$, $A \rightarrow \mu(Y(\omega), A)$ 是 (S, \mathcal{S}) 上的概率测度.



证明 见 Durrett 所著《Probability: Theory and Examples》第 5 版第 186 – 187 页.