概率论笔记

崔嘉祺
华东师范大学 20 级数学强基拔尖班

2023 年 12 月 8 日

摘要

这是华东师范大学数学专业研究生的基础课“概率论”的课程笔记。
每周三交作业。

目录

1 测度与概率空间

定义 1.1. 对集合 Ω, 子集族 $\mathcal{F} \subset 2^{\Omega}$ 称为 Ω 上的一个 σ-代数，若

(1) $\Omega \in \mathcal{F}$.

(2) 若 $A \in \mathcal{F}$, 则 $A^c \in \mathcal{F}$.（补封闭性）

(3) 若 $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, 则 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.（可列并封闭）

注记 1.2. 由定义可推出：

1. $\varnothing \in \mathcal{F}$.
2. 若 $A, B \in \mathcal{F}$, 则 $A \cap B \in \mathcal{F}$.（交封闭性）
3. 若 $A, B \in \mathcal{F}$, 则 $A \cup B \in \mathcal{F}$.（有限并封闭性）

注记 1.3. 若把定义中的 *(3)* 改为有限并，则称其为一个 *Borel* 代数。

定义 1.4. 对集合 Ω, 子集族 $\mathcal{F} \subset 2^{\Omega}$ 是 Ω 上的一个 σ-代数，$\mu : \mathcal{F} \to [0, +\infty]$, 满足

(1) $\mu(\varnothing) = 0$.

(2) 若 $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, $A_i \cap A_j = \varnothing, i \neq j$, 则有

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

（σ 可加性/可数可加性）

则称三元组 $(\Omega, \mathcal{F}, \mu)$ 是一个测度空间。

例 1.5. $\mathbb{R}$ 上的 *Lebesgue* 测度：$\Omega = \mathbb{R}$, $\mathcal{F}$ 是 *Lebesgue* 可测集，μ 是 *Lebesgue* 测度。

定义 1.6. 称一个测度空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间，若 $\mathbb{P} : \mathcal{F} \to [0, 1]$, 且 $\mathbb{P}(\Omega) = 1$. 其中称 Ω 是样本空间，$\omega \in \Omega$ 是基本事件，$\mathcal{F}$ 是事件域，$A \in \mathcal{F}$ 是事件，$\mathbb{P}(A)$ 是 A 的概率。

定理 1.7. 若 $(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间，则有

(1) 若 $A, B \in \mathcal{F}$, $A \subset B$, 则 $\mathbb{P}(A) \le \mathbb{P}(B)$.（单调性）

(2) 若 $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, 则有

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

（次可加性）

(3) 设 $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, 若

- $A_1 \supset A_2 \supset A_3 \supset \cdots$, 则记 $\lim_{i\to\infty} A_i = \bigcap_{i=1}^{\infty} A_i = A$, 记为 $A_i \searrow A$.
- $A_1 \subset A_2 \subset A_3 \subset \cdots$, 则记 $\lim_{i\to\infty} A_i = \bigcup_{i=1}^{\infty} A_i = A$, 记为 $A_i \nearrow A$.

则 $A \in \mathcal{F}$, 且 $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$.

(4) $\forall A \in \mathcal{F}$, $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$.

证明. (1) 由 $B = A \cup (B \cap A^c)$, 则 $A \cap (B \cap A^c) = \varnothing$. 由可加性即得。

(2) 记 $B_i = A_i \cap (A_1 \cup \cdots \cup A_{i-1})^c$, 则 $\{B_i\}$ 两两不交，从而由可数可加性与单调性，

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

(3) 只证 $A_i \nearrow A$. 由 $A = \cup_{i=1}^{\infty} A_i$, 由 σ-代数的可数并封闭性，$A \in \mathcal{F}$. 记 $B_1 = A_1$, $B_n = A_n \cap A_{n-1}^c$. 则

$$A_n = B_1 \cup B_2 \cup \cdots \cup B_n,$$

且 $\{B_i\}$ 两两不交，

$$A = B_1 \cup B_2 \cup \cdots .$$

由可数可加性，

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(4) $A \cap A^c = \varnothing$, 从而由可加性，

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1.$$

$\square$

例 1.8. $\Omega = [0, 1]$, $\mathcal{F}$ 是 *Lebesgue* 可测集，$\mathbb{P}$ 是 *Lebesgue* 测度。

$$\mathbb{P}\left(x \in [0, \frac{1}{3}]\right) = \mathbb{P}\left([0, \frac{1}{3}]\right) = \frac{1}{3}.$$

例 1.9 (古典概型)**.** Ω 是有限集，$\mathcal{F} = 2^{\Omega}$, $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

例 1.10 (离散概率模型)**.** Ω 是可数集，$\mathcal{F} = 2^{\Omega}$, $p : \Omega \to [0, +\infty]$, 满足 $\sum_{\omega\in\Omega} p(w) = 1$. 对 $A \in \mathcal{F}$, 令 $\mathbb{P}(A) = \sum_{\omega\in A} p(w)$. 则 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间。

例 1.11 (投均匀硬币)**.** 记 H 是正，T 是反。

- 投一次：$\Omega = \{H, T\}$, $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$.

- 投 n 次：$\Omega=\{H,T\}^n=\{(x_1,\cdots,x_n)\mid x_i\in\{H,T\},1\le i\le n\}$, $\mathcal{F}=2^\Omega$, $\forall\omega\in\Omega$, $\mathbb{P}(\{\omega\})=\frac{1}{2^n}$.
  - A_1 是第一次投正面，$\mathbb{P}(A_1)=\frac{|A_1|}{|\Omega|}=\frac{2^{n-1}}{2^n}=\frac{1}{2}$. A_1^c 是第一次投正面，$\mathbb{P}(A_1^c)=\frac{1}{2}$.
  - A_i 是第 i 次投正面，$\mathbb{P}(A_i)=\frac{1}{2}$. 若 $1\le i_1\le i_2\le\cdots\le i_k\le n$, $B=B_{i_1}\cap\cdots\cap B_{i_k}$, 其中 $B_{i_j}\in\{A_{i_j},A_{i,j}^c\}$, 则有 $\mathbb{P}(B)=\frac{1}{2^k}$.

例 1.12 (Ramsey 数的应用)**.**

定理 1.13. 任意六个人中必有三个人互相认识或互相不认识。

证明. 任取一个人 A，考察与其余五个人的关系。有三个人与 A 认识或不认识，不妨设 B,C,D 与 A 认识（不认识的情况类似）。若 B,C,D 中有两人认识（如 B,C），则 A,B,C 互相认识；若 B,C,D 没有两人认识，则三者满足互相不认识。 □

注记 1.14. 图论形式：6 阶完全图，边用红蓝染色，则必有一个红色或蓝色三角形。

问题 1.15 (一般的 Ramsey 问题)**.** 给定正整数 k,l, 是否存在 n, 满足将 n 阶完全图 K_n 的边红蓝染色后，一定有一个红的 K_k 或蓝的 K_l. 满足这样的条件的最小的 n 记为 $R(k,l)$.

命题 1.16. 一些结论：

- $R(3,3)=6$.
- $R(2,l)=l$.
- $R(k,l)=R(l,k)$

定理 1.17. 对 $k,l\ge 3$, 有

$$R(k,l)\le R(k-1,l)+R(k,l-1).$$

归纳可证 $R(k,l)$ 是有限数。

证明. 取出一个点 A, 剩下的图记为 W. 记 $n = R(k-1,l) + R(k,l-1)$. 由

$$|W| = n-1 = R(k-1,l) + R(k,l-1) - 1 \geq R(k-1,l),$$

从而

- 要么有 $k-1$ 个点组成红色完全图，此时再把它们与 A 用红色相连即得。
- 要么有 l 个点组成蓝色完全图，

□

定理 1.18. 对 $k \geq 3$, 有

$$R(k,k) \geq \left\lfloor 2^{\frac{k}{2}} \right\rfloor.$$

证明. 对 K_n 的边随机染色，所有可能的染色方案是样本空间 Ω, $|\Omega| = 2^{C_n^2}$. 对指定的 k 个顶点 $v_1, v_2, \cdots, v_k$,

$$\mathbb{P}(v_1, \cdots, v_k \text{ 之间的是红边}) = \frac{1}{2^{C_k^2}}.$$

从而

$$\begin{aligned}\mathbb{P}(\text{存在 } k \text{ 个点之间的是红边}) &= \mathbb{P}(\bigcup_{v_1,\cdots,v_k} v_1, \cdots, v_k \text{ 之间的是红边}) \\ &\leq \sum_{v_1,\cdots,v_k} \mathbb{P}(v_1, \cdots, v_k \text{ 之间的是红边}) = \frac{C_n^k}{2^{C_k^2}}.\end{aligned}$$

同理，

$$\mathbb{P}(\text{存在 } k \text{ 个点之间的是蓝边}) \leq \frac{C_n^k}{2^{C_k^2}}.$$

从而

$$
\begin{aligned}
&\mathbb{P}(\text{存在 } k \text{ 个点之间的是红边或 } k \text{ 个点之间的是蓝边})\\
\leq&\mathbb{P}(\text{存在 } k \text{ 个点之间的是红边}) + \mathbb{P}(\text{存在 } k \text{ 个点之间的是蓝边})\\
\leq&C_n^k \cdot 2^{1-C_k^2}.
\end{aligned}
$$

而其中，

$$C_n^k \cdot 2^{1-C_k^2} = \frac{n!}{k!(n-k)!} \cdot 2^{1+\frac{k}{2}} \cdot 2^{-\frac{k^2}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{\frac{k^2}{2}}}.$$

易知对 $n \geq 3$, 有 $2^{1+\frac{k}{2}} < k!$, 故

$$C_n^k \cdot 2^{1-C_k^2} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{\frac{k^2}{2}}} < \frac{n^k}{2^{\frac{k^2}{2}}} = \left(\frac{n}{2^{\frac{k}{2}}}\right)^k.$$

故只要 $n = \left\lfloor 2^{\frac{k}{2}} \right\rfloor$ 时，就有

$$\mathbb{P}(\text{存在 } k \text{ 个点之间的是红边或 } k \text{ 个点之间的是蓝边}) \leq C_n^k \cdot 2^{1-C_k^2} < \left(\frac{n}{2^{\frac{k}{2}}}\right)^k \leq 1.$$

此时，存在一种染色方案，使得不存在 k 个点之间的是红边，也不存在 k 个点之间的是蓝边。 □

例 1.19 (投掷硬币无穷多次)**.** $\Omega = \{H, T\}^{\mathbb{N}} = \{(x_1, x_2, \cdots) \mid x_i \in \{H, T\}, i = 1, 2, \cdots\}$. 记 A_i 是第 i 次是 H. 定义

$$\sigma(A_1) = \mathcal{F}_1 = \{\varnothing, A_1, A_1^c, \Omega\},$$

它只能分辨第一次的结果。

定义 1.20. 设 Ω 是集合，A 是它的一些子集，记 $\sigma(A)$ 是由 A 生成的 σ-代数，是包含 A 的最小 σ-代数。

定义

$$\sigma(A_1, \cdots, A_n) = \mathcal{F}_n, |\mathcal{F}_n| = 2^{2^n}.$$

它只能分辨前 n 次结果。

问题 1.21. 无法处理无穷次的情况。怎么办？

最初的想法：

$$\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n.$$

但 $\mathcal{A}$ 不是 σ-代数，它只对有限并封闭。

目标 1.22. 找到一个 σ-代数 $\mathcal{B}$ 包含 $\mathcal{A}$, $\mathbb{P} : \mathcal{B} \to [0,1]$, 使得 $\mathbb{P}|_{\mathcal{F}_n} = \mu_n$.

注意到有

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \cdots,$$

对 $i > j$, $\mu_i : \mathcal{F}_i \to [0,1]$, $\mu_i|_{\mathcal{F}_j} = \mu_j$, 拼起来得到 $\mu : \mathcal{A} \to [0,1]$.

定义 1.23. 若 $\mathcal{A}$ 是集合 Ω 的子集族，满足

1. $\Omega \in \mathcal{A}$.
2. 补封闭。
3. 有限并封闭。

则称 $\mathcal{A}$ 是一个 *Boole* 代数。

定义 1.24. 对 *Boole* 代数 $\mathcal{A}$, 函数 $m : \mathcal{A} \to \mathbb{R}$ 满足

1. $\forall A \in \mathcal{A}$, 有 $m(A) \geq 0$.
2. $\forall A, B \in \mathcal{A}$, 满足 $A \cap B = \varnothing$, 则有 $m(A \cup B) = m(A) + m(B)$.
3. $\forall \{A_n\} \subset \mathcal{A}$, 满足 $A_n \searrow \varnothing$, 则 $m(A_n) \to 0$.（连续性）

则称 m 是 *Boole* 代数 $\mathcal{A}$ 上的测度。

定义 1.25. 对 *Boole* 代数 $\mathcal{A}$ 以及它上面的测度 m. $\forall A \subset \Omega$,

$$m^*(A) = \inf\left\{\sum_i m(S_i) \mid S_i \in \mathcal{A}, A \subset \bigcup_i S_i\right\}$$

称为 A 的外测度。

注记 1.26. *1.* 在外测度的定义中可以假设 S_i 互不相交：令 $T_1 = S_1, T_i = S_i - (S_1 \cup \cdots \cup S_{i-1})$, 有 $T_i \in \mathcal{A}$, $A \subset \bigcup_i T_i$, $m(T_i) \le m(S_i)$, 从而 $\sum m(T_i) \le \sum m(S_i)$.

2. 若 $A \in \mathcal{A}$, 则 $m^*(A) = m(A)$.

定义 1.27. 对 $A \subset \Omega$, 称 A 是 m^*-可测的，若 $\forall E \subset \Omega$, 有

$$m^*(A \cap E) + m^*(A^c \cap E) = m^*(E). \tag{1}$$

其中 *(1)* 式称为可测性条件。

定理 1.28 (m^* 的次可加性)**.** $\forall A, A_1, A_2, \cdots \subset \Omega$, 满足 $A \subset \bigcup_i A_i$, 则 $m^*(A) \le \sum_i m^*(A_i)$.

证明. $\forall \varepsilon > 0$, 由 inf 的定义，可取 $\{A_{ij}\}_j \subset \mathcal{A}$, 满足 $A_i \subset \bigcup_j A_{ij}$, 且 $\sum_j m(A_{ij}) \le m^*(A_i) + \frac{\varepsilon}{2^i}$, 从而有 $A \subset \bigcup_i \bigcup_j A_{ij}$. 此时，

$$m^*(A) \le \sum_i \sum_j m(A_{ij}) \le \sum_i \left(m^*(A_i) + \frac{\varepsilon}{2^i}\right) = (m^*(A_i)) + \varepsilon.$$

令 $\varepsilon \to 0$ 即得 $m^*(A) \le \sum_i m^*(A_i)$. □

注记 1.29. 由 m^* 次可加性，$\forall E \subset \Omega$,

$$m^*(E) = m^*((A \cap E) \cup (A^c \cap E)) \le m^*(A \cap E) + m^*(A^c \cap E).$$

故可测性条件只需要验证

$$m^*(A \cap E) + m^*(A^c \cap E) \le m^*(E).$$

定理 1.30 (Caratheodory 测度扩张定理)**.** 设 $\mathcal{A}$ 是 Ω 上的 *Boole* 代数，m 是 $\mathcal{A}$ 上的测度，m^* 是外测度。记 $\mathcal{M}(m^*)$ 是所有 m^*-可测集构成的子集族。则 $\mathcal{M}(m^*)$ 是包含 $\mathcal{A}$ 的一个 σ-代数，$m^* : \mathcal{M}(m^*) \to \mathbb{R}$ 是一个测度，且 $m^*|_{\mathcal{A}} = m$.

证明. 先证包含关系

引理 1.31. $\mathcal{A} \subset \mathcal{M}(m^*)$.

证明. $\forall A \in \mathcal{A}, \forall E \subset \Omega, \forall \varepsilon > 0$, 由 inf 的定义，可以选取一组 $\{A_i\} \subset \mathcal{A}$, 使得 $E \subset \bigcup_i A_i$, 且

$$\sum_i m(A_i) \le m^*(E) + \varepsilon.$$

注意到有

$$A \cap E \subset \bigcup_i (A \cap A_i),\ A^c \cap E \subset \bigcup_i (A^c \cap A_i).$$

由 Boole 代数的交封闭，有 $\{A \cap A_i\}, \{A^c \cap A_i\} \subset \mathcal{A}$, 从而由 m^* 的定义，

$$m^*(A \cap E) \le \sum_i m(A \cap A_i),\ m^*(A^c \cap E) \le \sum_i m(A^c \cap A_i).$$

从而，由于 $(A \cap E) \cap (A^c \cap E) = \varnothing$, 有

$$\begin{aligned} m^*(A \cap E) + m^*(A^c \cap E) &\le \sum_i \left(m(A \cap A_i) + m(A^c \cap A_i)\right) \\ &= \sum_i m((A \cap A_i) \cup (A^c \cap A_i)) \\ &= \sum_i m(A_i) \\ &\le m^*(E) + \varepsilon. \end{aligned}$$

由 ε 的任意性，得

$$m^*(A \cap E) + m^*(A^c \cap E) \le m^*(E).$$

从而 A 满足可测性条件，故 $A \in \mathcal{M}(m^*)$. 最后再由 A 的任意性，即得 $\mathcal{A} \subset \mathcal{M}(m^*)$. □

再验证$\mathcal{M}(m^*)$ 是一个 σ-代数

引理 1.32. $\mathcal{M}(m^*)$ 是一个 σ-代数。

证明. (1) $\forall E \subset \Omega$,

$$m^*(\Omega \cap E) + m^*(\Omega^c \cap E) = m^*(E) + m^*(\varnothing) = m^*(E \cup \varnothing) = m^*(E).$$

即 Ω 满足可测性条件，有 $\Omega \in \mathcal{M}(m^*)$.

(2) $\forall A \in \mathcal{M}(m^*)$, 由 $(A^c)^c = A$, 在可测性条件式中交换 A 与 A^c 的位置，即知 A^c 也满足可测性条件。事实上，$\forall E \subset \Omega$,

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

故 $A^c \in \mathcal{M}(m^*)$. $\mathcal{M}(m^*)$ 满足补封闭性。

(3) 设 $A, B \in \mathcal{M}(m^*), \forall E \subset \Omega$,

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

这说明 $A \cap B \in \mathcal{M}(m^*)$. 即 $\mathcal{M}(m^*)$ 对交封闭。再由对补封闭，

$$A \cup B = (A^c \cap B^c)^c \in \mathcal{M}(m^*).$$

从而 $\mathcal{M}(m^*)$ 对有限并封闭。此时知 $\mathcal{M}(m^*)$ 是一个 Boole 代数。

设 $\{A_i\}_{i=1}^{\infty} \subset \mathcal{M}(m^*)$. 令 $B_1 = A_1, B_i = A_i - (A_1 \cup \cdots \cup A_{i-1})$, 由补封闭性与有限并封闭性，得到 $\{B_i\}_{i=1}^{\infty} \subset \mathcal{M}(m^*)$, 且 $\{B_i\}_{i=1}^{\infty}$ 两两不交，记 $\bigcup_i A_i = \bigcup_i B_i = A$.

$\forall E \subset \Omega, \forall n,$

$$\begin{aligned}
m^*(E) &= m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right) \\
&= m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right) \cap B_n\right) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right) \cap B_n^c\right) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right) \\
&= m^*(E \cap B_n) + m^*\left(E \cap \left(\bigcup_{i=1}^{n-1} B_i\right)\right) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right) \\
&= m^*(E \cap B_n) + m^*(E \cap B_{n-1}) + m^*\left(E \cap \left(\bigcup_{i=1}^{n-2} B_i\right)\right) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right) \\
&= \cdots \\
&= \sum_{i=1}^{n} m^*(E \cap B_i) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right).
\end{aligned}$$

因为

$$\bigcup_{i=1}^{n} B_i \subset \bigcup_{i=1}^{\infty} B_i = A,$$

所以

$$A^c \subset \left(\bigcup_{i=1}^{n} B_i\right)^c.$$

于是

$$\begin{aligned}
m^*(E) &= \sum_{i=1}^{n} m^*(E \cap B_i) + m^*\left(E \cap \left(\bigcup_{i=1}^{n} B_i\right)^c\right) \\
&\geq \sum_{i=1}^{n} m^*(E \cap B_i) + m^*(E \cap A^c).
\end{aligned}$$

令 $n\to\infty$, 由次可加性，

$$
\begin{aligned}
m^*(E) &\geq \sum_{i=1}^{\infty} m^*(E\cap B_i)+m^*(E\cap A^c)\\
&\geq m^*\left(\bigcup_{i=1}^{\infty}(E\cap B_i)\right)+m^*\left(E\cap A^c\right)\\
&= m^*(E\cap A)+m^*(E\cap A^c).
\end{aligned}
$$

即 A 满足可测性条件，是 m^*-可测的，即 $A\in\mathcal{M}(m^*)$. 从而 $\mathcal{M}(m^*)$ 对可列并封闭。

综上，$\mathcal{M}(m^*)$ 是一个 σ-代数。 □

最后验证<u>m^* 是一个测度</u>

引理 1.33. m^* 是 σ-代数 $\mathcal{M}(m^*)$ 上的一个测度。

证明. (1) 由定义显然有 $m^*\geq 0$, $m^*(\varnothing)=0$.

(2) $\forall A,B\in\mathcal{M}(m^*)$, $A\cap B=\varnothing$, 取 $E=A\cup B$, 对 A 使用可测性条件，

$$
m^*(A\cup B)=m^*(A\cap(A\cup B))+m^*(A^c\cap(A\cup B))=m^*(A)+m^*(B).
$$

即 m^* 有有限可加性。

$\forall\{A_n\}\subset\mathcal{M}(m^*)$, $\{A_n\}$ 两两不交。由次可加性，单调性与有限可加性，

$$
\sum_{n=1}^{\infty}m^*(A_n)\geq m^*\left(\bigcup_{n=1}^{\infty}A_n\right)\geq m^*\left(\bigcup_{n=1}^{N}A_n\right)=\sum_{n=1}^{N}m^*(A_n).
$$

令 $N\to\infty$, 得上式的等号都成立，从而有

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

这就证明了 m^* 的可数可加性。

综上，m^* 是 σ-代数 $\mathcal{M}(m^*)$ 上的一个测度。 □

□

定义 1.34. 设 $\mathcal{C}$ 是集合 Ω 的子集族。若 $\forall\{A_n\}\subset\mathcal{C}$, 满足 $A_n\nearrow A$ 或 $A_n\searrow A$, 则有 $A\in\mathcal{C}$, 则称 $\mathcal{C}$ 是单调类。

定理 1.35 (单调类定理)**.** 设 $\mathcal{A}$ 是集合 Ω 上的 *Boole* 代数，$\mathcal{M}^*$ 是包含 $\mathcal{A}$ 的最小单调类，则 $\mathcal{M}^*=\sigma(\mathcal{A})$.

证明. 由定义，σ-代数是单调类，从而 $\sigma(\mathcal{A})\supseteq\mathcal{M}^*$. 只要证 $\sigma(\mathcal{A})\subseteq\mathcal{M}^*$.

(1) $\forall A\in\mathcal{A}$, 考虑

$$\mathcal{M}_A=\{B\in\mathcal{M}^*\mid A\cap B,A-B,B-A\in\mathcal{M}^*\}\subseteq\mathcal{M}^*.$$

由于 $\mathcal{A}$ 是 Boole 代数，从而若 $B\in\mathcal{M}_A$, 则 $A\subset\mathcal{M}_A$.

由于 $\mathcal{M}_A$ 是单调类，$\forall\{B_n\}\subset\mathcal{M}_A$, $B_n\nearrow B$（或 $B_n\searrow B$），则有

- $A\cap B_n\nearrow(\searrow)A\cap B$.
- $A-B_n\searrow(\nearrow)A-B$.
- $B_n-A\nearrow(\searrow)B-A$.

由 $\mathcal{M}^*$ 是单调类，从而 $A\cap B,A-B,B-A\in\mathcal{M}^*$. 于是 $B\in\mathcal{A}$. 这说明 $\mathcal{M}_A$ 是单调类。由 $\mathcal{M}^*$ 的最小性，$\mathcal{M}_A=\mathcal{M}^*$.

(2) $\forall A\in\mathcal{M}^*$, 定义

$$\mathcal{M}_A=\{B\in\mathcal{M}^*\mid A\cap B,A-B,B-A\in\mathcal{M}^*\}\subseteq\mathcal{M}^*.$$

由 (1), 有 $\mathcal{A}\subseteq\mathcal{M}_A$. 类似 (1) 的证明，有 $\mathcal{M}_A$ 是单调类，从而 $\mathcal{M}_A=\mathcal{M}^*$. 于是 $\forall A,B\in\mathcal{M}^*$, 有 $A\cap B,A-B,B-A\in\mathcal{M}^*$, 即 $\mathcal{M}^*$ 是 Boole 代数。从而 $\mathcal{M}^*$ 既是 Boole 代数又是单调类，从而它是 σ-代数（习题）。从而 $\sigma(\mathcal{A})\subseteq\mathcal{M}^*$.

□

定理 1.36 (扩张的唯一性定理)**.** 设 $\mathcal{A}$ 是 Ω 上的 *Boole* 代数，m 是 $\mathcal{A}$ 上的测度，m^*, μ 是 $\sigma(\mathcal{A})$ 上的两个测度，且 $m^*|_{\mathcal{A}} = \mu|_{\mathcal{A}} = m$, 则 $m^* = \mu$.

证明. 考虑 $\mathcal{M} = \{A \in \sigma(\mathcal{A}) \mid \mu(A) = m^*(A)\} \subset \sigma(\mathcal{A})$, 则 $\mathcal{A} \subseteq \mathcal{M}$. 由测度的连续性，若 $\{A_n\} \subset \mathcal{M}$, $A_n \nearrow A$ （或 $A_n \searrow A$），由单调类定理（定理 1.35），$A \in \sigma(\mathcal{A})$. 则

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} m^*(A_n) = m^*(A),$$

从而 $A \in \mathcal{M}$. 这说明 $\sigma(\mathcal{A}) \subseteq \mathcal{M}$. 从而 $\sigma(\mathcal{A}) = \mathcal{M}$. 即 $\forall A \in \mathcal{M} = \sigma(\mathcal{A})$, $m^*(A) = \mu(A)$. □

回到无穷次投硬币。$\Omega = \{H, T\}^{\mathbb{N}} = \{(x_1, x_2, \cdots) \mid x_i \in \{H, T\}, i = 1, 2, \cdots\}$. 记 A_i 是第 i 次是 H. 定义

$$\sigma(A_1) = \mathcal{F}_1 = \{\varnothing, A_1, A_1^c, \Omega\},$$

$$\sigma(A_1, \cdots, A_n) = \mathcal{F}_n,\ |\mathcal{F}_n| = 2^n.$$

有

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots .$$

$$\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$$

是 Boole 代数。定义 $\mu : \mathcal{A} \to [0, 1]$. 设 $A \in \mathcal{A}$, 则 $\exists n$, 使得 $A \in \mathcal{F}_n$. 令 $\mu(A) = \mu_n(A)$, 其中 μ_n 就是投掷有限次时古典概型中的测度。

$$\mu_n : \mathcal{F}_n \to [0, 1]$$

对 $m < n$,$\mu_n|_{\mathcal{F}_m} = \mu_m$. μ 符合 Boole 代数上测度的定义：

(1) $\mu(A) \geq 0$.

(2) $\forall A \cap B = \varnothing$, $\mu(A \cup B) = \mu(A) + \mu(B)$.

(3) 若 $A_n \searrow \varnothing$, 则 $\mu(A_n) \to 0$.

验证 (3): 设有 $\{A_n\} \subset \mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$, $A_n \searrow \varnothing$. 事实上，对充分大的 n, 有 $A_n = \varnothing$.

引理 1.37 (Tychonoff 定理)**.** 任意紧拓扑空间的乘积空间是紧空间。

由引理，$\Omega = \{H,T\}^{\mathbb{N}} = \prod_{i=1}^{\infty}\{H,T\}$ 是紧空间，开集与闭集就是 $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$. 从而由 $A_n \searrow \varnothing$, 就有 $\exists N, \forall n \geq N$, $A_n = \varnothing$. 从而我们得到了包含 $\mathcal{A}$ 的 σ-代数 $\sigma(\mathcal{A})$ 上的概率测度 $\mathbb{P} = \mu$, 使得 $\mathbb{P}|_{\mathcal{F}_n} = \mu_n$, 就达成了目标 1.22, 解决了问题 1.21.

2 独立性

定义 2.1. 设 $(\Omega,\mathcal{F},\mathbb{P})$ 是概率空间，$C\in\mathcal{F}$, $\mathbb{P}(C)>0$. $\forall A\in\mathcal{F}$,

$$\mathbb{P}(A\mid C)=\frac{\mathbb{P}(A\cap C)}{\mathbb{P}(C)}=\mathbb{P}_C(A)$$

称为在条件 C 下事件 A 的概率。

注记 2.2.

- $\mathbb{P}_C(A)$ 也满足概率条件，$(\Omega,\mathcal{F},\mathbb{P}_C)$ 也是一个概率空间。
- 乘法公式：$\mathbb{P}(A\cap C)=\mathbb{P}(A\mid C)\cdot\mathbb{P}(C)$.
- 若 $\{C_i\}$ 是两两不相交的事件，$\mathbb{P}(C_i)>0$, $A\subset\bigcup_i C_i$, 则有如下的全概率公式：

$$\mathbb{P}(A)=\mathbb{P}\left(A\cap\left(\bigcup_i C_i\right)\right)=\mathbb{P}\left(\bigcup_i(A\cap C_i)\right)=\sum_i\mathbb{P}\left(A\cap C_i\right)=\sum_i\mathbb{P}(A\mid C_i)\cdot\mathbb{P}(C_i).$$

定义 2.3 (独立性)**.** $A,B\in\mathcal{F}$, 若

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\cdot\mathbb{P}(B),$$

则称事件 A,B 独立。若有 $A_1,\cdots,A_n\in\mathcal{F}$, $\forall\{i_1,\cdots,i_k\}\subset\{1,\cdots,n\}$, 都有

$$\mathbb{P}(A_{i_1}\cdots A_{i_k})=\mathbb{P}(A_{i_1})\cdots\mathbb{P}(A_{i_k}),$$

则称事件 $A_1,\cdots,A_n$ 互相独立。

注记 2.4. n 个事件独立强于两个事件独立，取 $k=2$ 即得。但反之不一定对：$\Omega=\{a,b,c,d\}$, $A_1=\{a,b\}$, $A_2=\{a,c\}$, $A_3=\{a,d\}$. 它们两两独立但三者不独立。

定义 2.5. 设 $\mathcal{A},\mathcal{B}\subset\mathcal{F}$, 称事件集 $\mathcal{A},\mathcal{B}$ 独立，若 $\forall A\in\mathcal{A},B\in\mathcal{B}$, 有 A,B 独立。

注记 2.6. 若 A, B 独立，则 $\sigma(A) = \{\varnothing, \Omega, A, A^c\}$ 与 $\sigma(B) = \{\varnothing, \Omega, B, B^c\}$ 事件集独立，从而 A, B^c 独立。

问题 2.7. 事件集的独立性能否得出它们生成的 σ-代数的独立性？

定义 2.8 (π-系)**.** $\mathcal{A}$ 是一个集族，满足

(1) $\varnothing \in \mathcal{A}$.

(2) 若 $A, B \in \mathcal{A}$, 则 $A \cap B \in \mathcal{A}$.

则称 $\mathcal{A}$ 是一个 π-系。

定义 2.9 (λ-系)**.** $\mathcal{A}$ 是一个集族，满足

(1) $\Omega \in \mathcal{A}$.

(2) 若 $A, B \in \mathcal{A}$, $A \subset B$, 则 $B - A \in \mathcal{A}$.

(3) 若 $\{A_n\} \subset \mathcal{A}$, $A_n \nearrow A$, 则 $A \in \mathcal{A}$.

则称 $\mathcal{A}$ 是一个 λ-系。

注记 2.10. 任意多个 π-系（λ-系）的交仍然是 π-系（λ-系）。

定理 2.11 (π-λ 定理)**.** 若 $\mathcal{P}$ 是一个 π-系，$\mathcal{L}$ 是包含 $\mathcal{P}$ 的 λ-系，则 $\sigma(\mathcal{P}) \subset \mathcal{L}$.

证明. 可以不妨设 $\mathcal{L} = \lambda(\mathcal{P})$.

1. 首先证明 $\mathcal{L}$ 是 Boole 代数。由定义，$\varnothing, \Omega \in \mathcal{L}$, 且对补封闭。只要证关于交封闭。$\forall A \in \mathcal{P}$, 定义

$$\mathcal{L}_A = \{B \in \mathcal{L} \mid A \cap B \in \mathcal{L}\} \subset \mathcal{L}.$$

   - 由于 $\mathcal{P}$ 是 π-系，$\forall B \in \mathcal{P}$,$A \cap B \in \mathcal{P}$. 从而 $\mathcal{P} \subset \mathcal{L}_A$.

- $\forall B \in \mathcal{L}_A, B \subset C$. 则 $A \cap B, A \cap C \in \mathcal{L}, A \cap C \subset A \cap C$. 由于 $\mathcal{L}$ 是 λ-系,有 $(A \cap C)-(A \cap B) \in \mathcal{L}$. 又 $A \cap (C-B) = (A \cap C)-(A \cap B) \in \mathcal{L}$, 从而 $C - B \in \mathcal{L}_A$.
- 设 $\{B_n\} \subset \mathcal{L}_A$, $B_n \nearrow B$. 有 $A \cap B_n \in \mathcal{L}$, $A \cap B_n \nearrow A \cap B$. 由于 $\mathcal{L}$ 是 λ-系，$A \cap B \in \mathcal{L}$, 从而 $B \in \mathcal{L}_A$.

这说明 $\mathcal{L}_A \subset \mathcal{L}$ 也是 λ-系。由 $\mathcal{L} = \lambda(\mathcal{P})$ 的极小性，必须有 $\mathcal{L}_A = \mathcal{L}$. 即 $\forall A \in \mathcal{P}, B \in \mathcal{L}$, 有 $A \cap B \in \mathcal{L}$. 从而 $\mathcal{L}$ 对交封闭，是一个 Boole 代数。

2. $\forall \{A_i\} \subset \mathcal{L}$, 令 $B_i = \bigcup_{j=1}^{i}$. 有 $B_i \nearrow \bigcup_{i=1}^{\infty} A_i$. 由于 $\mathcal{L} = \lambda(\mathcal{P})$ 是一个 λ-系，$\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

综上，$\mathcal{L}$ 是一个 σ-代数。最后，由于 $\sigma(\mathcal{P})$ 是包含 $\mathcal{P}$ 的最小 σ-代数，有 $\sigma(\mathcal{P}) \subset \mathcal{L}$.

□

推论 2.12. 若 $\mathcal{P}$ 是一个 π-系，则 $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$.

定理 2.13. $(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间，$\mathcal{B} \subset \mathcal{F}$ 是一个 π-系，$A \in \mathcal{F}$. 若 A 与 $\mathcal{B}$ 独立。则 A 与 $\sigma(\mathcal{B})$ 独立。

证明. 考虑

$$\mathcal{L} = \{B \in \mathcal{F} \mid \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}.$$

是与 A 独立的事件的集合。

下面证明 $\mathcal{L}$ 是一个 λ-系。

1. 全空间与任何事件独立，即 $\Omega \in \mathcal{L}$.
2. 若 $B, C \in \mathcal{L}, B \subset C$.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C).$$

从而

$$\mathbb{P}(A \cap (C-B)) = \mathbb{P}((A \cap C) - (A \cap B)) = \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\left(\mathbb{P}(C) - \mathbb{P}(B)\right) = \mathbb{P}(A)\mathbb{P}(C-B).$$

即 $C - B \in \mathcal{L}$.

3. 若在 $\mathcal{L}$ 中有 $B_n \nearrow B$.

$$\mathbb{P}(A \cap B_n) = \mathbb{P}(A)\mathbb{P}(B_n), \forall n.$$

则有 $A \cap B_n \nearrow A \cap B$. 由概率的连续性，令 $n \to \infty$, 就有

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

从而 $B \in \mathcal{L}$.

综上，$\mathcal{L}$ 是一个 λ-系。由 π-λ 定理（定理 2.11），$\sigma(\mathcal{B}) \subset \mathcal{L}$. 即 A 与 $\sigma(\mathcal{B})$ 独立。 □

推论 2.14. $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ 是两个 π-系，$\mathcal{A}$ 与 $\mathcal{B}$ 独立。则 $\sigma(\mathcal{A})$ 与 $\sigma(\mathcal{B})$ 独立。

证明. 习题 □

定义 2.15. 设 $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$. 令

$$B_n = \bigcup_{i=n}^{\infty} A_i, \ C_n = \bigcap_{i=n}^{\infty} A_i.$$

则记 $B_n \searrow B = \limsup_{n\to\infty} A_n$, $C_n \nearrow C = \liminf_{n\to\infty} A_n$. 则有

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} A_i\right), \ \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\left(\bigcap_{i=n}^{\infty} A_i\right).$$

$\omega \in \limsup_{n\to\infty} A_n$ 当且仅当 ω 属于无穷多个 A_n, $\omega \in \liminf_{n\to\infty} A_n$ 当且仅当不含 ω 的 A_n 只有有限个。

由

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\left(\bigcap_{i=n}^{\infty} A_i\right)\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} A_i\right) \le \liminf_{n\to\infty} \mathbb{P}(A_n),$$

由 de Morgan 律，

$$\left(\limsup_{n\to\infty} A_n\right)^c = \left[\bigcap_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} A_i\right)\right]^c = \bigcup_{n=1}^{\infty}\left(\bigcap_{i=n}^{\infty} A_i^c\right) = \liminf_{n\to\infty} A_n^c.$$

从而

$$1-\mathbb{P}\left(\limsup_{n\to\infty} A_n\right)=\mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right)\le \liminf_{n\to\infty}(1-\mathbb{P}(A_n))=1-\limsup_{n\to\infty}\mathbb{P}(A_n).$$

即

$$\limsup_{n\to\infty}\mathbb{P}(A_n)\le\mathbb{P}\left(\limsup_{n\to\infty} A_n\right).$$

定理 2.16 (Borel-Cantelli)**.** 设 $\{A_n\}_{n=1}^\infty\subset\mathcal{F}$.

1. 若 $\sum_{n=1}^\infty\mathbb{P}(A_n)<\infty$, 则 $\mathbb{P}\left(\limsup_{n\to\infty}A_n\right)=0$.

2. 若 $\{A_n\}_{n=1}^\infty$ 独立，且 $\sum_{n=1}^\infty\mathbb{P}(A_n)=\infty$, 则 $\mathbb{P}\left(\limsup_{n\to\infty}A_n\right)=1$.

证明. 1. $\forall n$,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right)\le\mathbb{P}\left(\bigcup_{i=n}^\infty A_i\right)\le\sum_{i=n}^\infty\mathbb{P}(A_i).$$

由于 $\sum_{n=1}^\infty\mathbb{P}(A_n)<\infty$ 收敛，从而 $\sum_{i=n}^\infty\mathbb{P}(A_i)\to 0$. 令 $n\to\infty$, 得 $\mathbb{P}\left(\limsup_{n\to\infty}A_n\right)\le 0$, 从而 $\mathbb{P}\left(\limsup_{n\to\infty}A_n\right)=0$.

2. 只要证 $\mathbb{P}\left(\left(\limsup_{n\to\infty}A_n\right)^c\right)=0$. 由 de Morgan 律, $\left(\limsup_{n\to\infty}A_n\right)^c=\left[\bigcap_{n=1}^\infty\left(\bigcup_{i=n}^\infty A_i\right)\right]^c=\bigcup_{n=1}^\infty\left(\bigcap_{i=n}^\infty A_i^c\right)$, 故只要证 $\mathbb{P}\left(\bigcup_{n=1}^\infty\left(\bigcap_{i=n}^\infty A_i^c\right)\right)=0$. 而这只要证 $\forall n$, $\mathbb{P}\left(\bigcap_{i=n}^\infty A_i^c\right)=0$, 从而由次可加性,

$$\mathbb{P}\left(\bigcup_{n=1}^\infty\left(\bigcap_{i=n}^\infty A_i^c\right)\right)\le\sum_{n=1}^\infty\mathbb{P}\left(\bigcap_{i=n}^\infty A_i^c\right)=\sum_{n=1}^\infty 0=0.$$

由于 $\{A_n\}_{n=1}^\infty$ 独立，它们的补也独立（注记 2.6）。从而 $\forall m>n$,

$$\mathbb{P}\left(\bigcap_{i=n}^m A_i^c\right)=\prod_{i=n}^m\mathbb{P}\left(A_i^c\right)=\prod_{i=n}^m\left(1-\mathbb{P}\left(A_i\right)\right)\le \mathrm{e}^{-\sum_{i=n}^m\mathbb{P}(A_i)}.$$

其中用到了 $\forall x\ge 0$, $1-x\le\mathrm{e}^{-x}$. 由于 $\sum_{i=1}^\infty\mathbb{P}\left(A_i\right)=\infty$, 故当 $m\to\infty$ 时，$\sum_{i=n}^m\mathbb{P}\left(A_i\right)\to\infty$. 此时，由概率的连续性,

$$\mathbb{P}\left(\bigcap_{i=n}^\infty A_i^c\right)=\lim_{m\to\infty}\mathbb{P}\left(\bigcap_{i=n}^m A_i^c\right)\le\lim_{m\to\infty}\mathrm{e}^{-\sum_{i=n}^m\mathbb{P}(A_i)}=\mathrm{e}^{-\infty}=0.$$

即有 $\mathbb{P}\left(\bigcap_{i=n}^{\infty} A_i^c\right) = 0.$

$\square$

推论 2.17 (Borel 0-1 律)**.** 设 $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$. 则 $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \in \{0,1\}$.

注记 2.18. 在概率论中，有很多某事件的概率是 0 或 1 的定理，这说明该事件要么必然发生，要么不可能发生，称为 0-1 律。

应用 2.19. 回到无穷次投硬币的情况。$\Omega = \{H,T\}^{\mathbb{N}}$. A_n 是第 n 次为 T. $\mathbb{P}(A_n) = \frac{1}{2}$. $\{A_n\}_{n=1}^{\infty}$ 独立。由 *Borel-Cantelli* 定理（定理 *2.16*），$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$. 这说明投无穷次硬币出现 T 的概率为 1, 它必然发生。

令 $\varphi : (0,1] \to \Omega = \{H,T\}^{\mathbb{N}}$. $\forall x \in (0,1]$, 将 x 写成二进制小数：$x = 0.a_1a_2a_3\cdots$. 该表示不唯一，因为有限小数也可以写成以 1 循环小数。例如

$$0.10011 = 0.10010\dot{1}$$

规定此时用以 1 循环来表示。

$$x \mapsto \varphi(x) = (a_1, a_2, a_3, \cdots)$$

将 0 换成 H, 1 换成 T. 则 φ 是单射。$\varphi\left((0,1]\right) = \limsup_{n\to\infty} A_n$.

$$\varphi^{-1}(A_1) = \left(\frac{1}{2}, 1\right],\ \varphi^{-1}(A_2) = \left(\frac{1}{4}, \frac{1}{2}\right] \cup \left(\frac{3}{4}, 1\right], \cdots, \varphi^{-1}(A_n) = \bigcup_{j=1}^{2^{n-1}} \left(\frac{2j-1}{2^n}, \frac{2j}{2^n}\right], \cdots$$

设 μ 是 $(0,1]$ 上的 *Lebesgue* 测度，$\mathbb{P}(A_n) = \frac{1}{2} = \mu\left(\varphi^{-1}(A_n)\right)$, 即在 Ω 上取均匀概率就是 $(0,1]$ 上的 *Lebesgue* 测度。

$$\Omega = \{H,T\}^{\mathbb{N}} \supset \Omega' = \limsup_{n\to\infty} \rightleftarrows (0,1].$$

$$\mathbb{P}(\Omega - \Omega') = 0,\ \mathbb{P}(\Omega') = 1.$$

$$\mathbb{P}(A_n \cap \Omega') = \mu(\varphi^{-1}(A_n)).$$

$$\mathcal{F} = \sigma(A_1, A_2, \cdots) \leftrightarrow \textit{Borel}\ \text{代数}.$$

即 $\{A_n\}_{n=1}^{\infty}$ 生成的 σ-代数 $\mathcal{F} = \sigma(A_1, A_2, \cdots)$ 就是 $(0,1]$ 上的 *Borel* 代数。

定理 2.20 (Kolmogorov 0-1 律)**.** 设 $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$. 令 $\mathcal{T}_n = \sigma(A_{n+1}, A_{n+2}, \cdots)$. 有 $\mathcal{T}_1 \supset \mathcal{T}_2 \supset \cdots$. 记 $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$. $\mathcal{T}$ 中的事件称为尾事件。若 $\{A_n\}_{n=1}^{\infty}$ 独立。则 $\forall T \in \mathcal{T}$, 有 $\mathbb{P}(T) \in \{0, 1\}$.

证明. 记 $\mathcal{F}_n = \sigma(A_1, \cdots, A_n)$, $\mathcal{F} = \sigma(A_1, A_2, \cdots)$, $\mathcal{B}_{n,k} = \sigma(A_{n+1}, \cdots .A_{n+k})$.

$\forall n$, $\bigcup_{k=1}^{\infty} \mathcal{B}_{n,k}$ 是 Boole 代数，是 π-系，$\mathcal{T}_n$ 是 $\bigcup_{k=1}^{\infty} \mathcal{B}_{n,k}$ 生成的 σ-代数。由条件，$\mathcal{F}_n$ 与 $\mathcal{B}_{n,k}$ 独立，从而与 $\bigcup_{k=1}^{\infty} \mathcal{B}_{n,k}$ 独立。由 π-λ 定理（定理 2.11），$\mathcal{F}_n$ 与 $\sigma(\bigcup_{k=1}^{\infty} \mathcal{B}_{n,k}) = \mathcal{T}_n$ 独立。从而 $\mathcal{F}_n$ 与 $\mathcal{T} = \bigcap_{k=1}^{\infty} \mathcal{T}_n$ 独立，从而 $\bigcup_{k=1}^{\infty} \mathcal{F}_n$ 与 $\mathcal{T} = \bigcap_{k=1}^{\infty} \mathcal{T}_n$ 独立。由于 $\bigcup_{k=1}^{\infty} \mathcal{F}_n$ 是 Boole 代数，是 π-系，从而 $\mathcal{F} = \sigma(\bigcup_{k=1}^{\infty} \mathcal{F}_n)$ 与 $\mathcal{T}$ 独立。

特别地，$\forall T \in \mathcal{T}$, T 与 T 独立，

$$\mathbb{P}(T) = \mathbb{P}(T \cap T) = \mathbb{P}(T)\mathbb{P}(T),$$

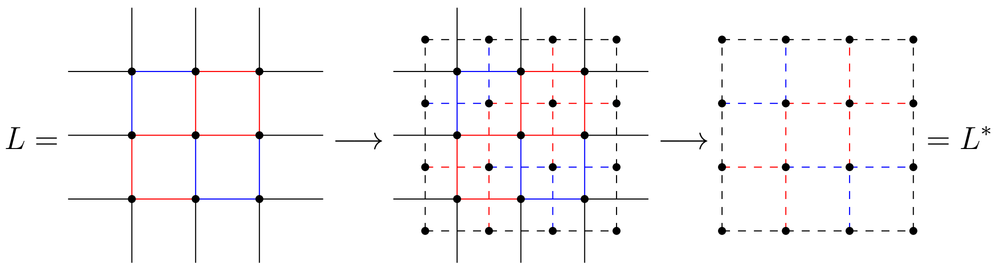
从而 $\mathbb{P}(T) \in \{0, 1\}$. □

注记 2.21. 有另一种形式的 *Kolmogorov* 0-1 律：$\{\mathcal{A}_n\}_{n=1}^{\infty}$ 是一列独立的 σ-代数，$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\mathcal{A}_{n+1}, \mathcal{A}_{n+2}, \cdots)$, 则 $\forall T \in \mathcal{T}$, 有 $\mathbb{P}(T) \in \{0, 1\}$.

证明是完全类似的：把 A 换成 $\mathcal{A}$ 即可。

应用 2.22 (边渗流模型)**.** 考虑平面上的格 $L = \mathbb{Z}^2$. 每一条边随机地，相互独立地，以概率 p 染成红色，以概率 $1-p$ 染成蓝色。L 的边是可数的。记红边构成的图是 G_r, 蓝边构成的图是 G_b. 我们称一个连通分支是巨连通分支，如果它有无穷多个顶点。记事件 $T_{r,p}$ 是 G_r 中有一个巨连通分支，它是一个尾事件：改变有限条边不会影响是否含有巨连通分支的情况，$T_{r,p} \in \bigcap_{n=1}^{\infty} \sigma(A_{n+1}, A_{n+2}, \cdots)$. 由 *Kolmogorov* 0-1 律（定理 *2.20*），$\mathbb{P}(T_{r,p}) \in \{0, 1\}$.

具体地，考虑 L 的对偶图 L^*. 则 L 上的边渗流模型与 L^* 上的边渗流模型有一一对应：



断言：若 L 上无红色巨连通分支，则 L^* 上有蓝色巨连通分支。即 $T_{r,p}^c \subset T_{b,1-p}$. 从而有

$$1 - \mathbb{P}(T_{r,p}) = \mathbb{P}(T_{r,p}^c) \leq \mathbb{P}(T_{b,1-p}) = \mathbb{P}(T_{r,1-p}).$$

这说明

1. 若 $\mathbb{P}(T_{r,p}) = 0$, 则 $\mathbb{P}(T_{b,1-p}) \geq 1 - 0 = 1$, 即 $\mathbb{P}(T_{b,1-p}) = 1$.
2. 当 $p = \frac{1}{2}$ 时，有 $1 - \mathbb{P}(T_{r,\frac{1}{2}}) \leq \mathbb{P}(T_{r,\frac{1}{2}})$, 即 $\mathbb{P}(T_{r,\frac{1}{2}}) \geq \frac{1}{2}$. 由 *Kolmogorov* 0-1 律（定理 *2.20*），只能有 $\mathbb{P}(T_{r,\frac{1}{2}}) = 1$. 且直观上有 $\mathbb{P}(T_{r,p})$ 关于 p 递增，从而 $\forall p \geq \frac{1}{2}, \mathbb{P}(T_{r,p}) = 1$.

3 随机变量

定义 3.1. 设两个可测空间 $(\Omega,\mathcal{F})$ 与 $(\Gamma,\mathcal{G})$, 映射 $f:\Omega\to\Gamma$ 称为可测映射, 若 $\forall A\in\mathcal{G}$, 有 $f^{-1}(A)\in\mathcal{F}$. 特别地, 若 $(\Gamma,\mathcal{G})=(\mathbb{R}, Borel$ 可测集$\mathcal{B}(\mathbb{R}))$, 则称 f 是 Ω 上的可测函数。

性质 3.2.

- 可测映射的复合还是可测映射。
- $\{f^{-1}(A)\mid A\in\mathcal{G}\}$ 是 $\mathcal{F}$ 的 σ-子代数:
  - $f^{-1}(\varnothing)=\varnothing$, $f^{-1}(\Gamma)=\Omega$.
  - $f^{-1}(\Gamma-A)=\Omega-f^{-1}(A)$.
  - $f^{-1}(\bigcup_{n=1}^{\infty}A_n)=\bigcup_{n=1}^{\infty}f^{-1}(A)$.

  记为 $f^{-1}\mathcal{G}$, 称为 f 在 Γ 上的拉回。它是使 f 可测的最小 σ-代数。
- $f_*\mathcal{F}=\{A\subset\Gamma\mid f^{-1}(A)\in\mathcal{F}\}$ 是 σ-代数, 称为 f 在 Ω 上的推出。它是使 f 可测的最大 σ-代数。
- $f:(\Omega,\mathcal{F})\to(\Gamma,\mathcal{G})$ 可测, 当且仅当 $f^{-1}\mathcal{G}\subset\mathcal{F}$, 当且仅当 $f_*\mathcal{F}\supset\mathcal{G}$.
- $f:(\Omega,\mathcal{F})\to(\Gamma,\mathcal{G})$. 如果 $\mathcal{G}=\sigma(\{A_\alpha\mid\alpha\in\Lambda\})$, 则 f 可测当且仅当 $f^{-1}(A_\alpha)\in\mathcal{F}$, $\forall\alpha\in\Lambda$.
- 可测函数的四则运算保持可测性。

定义 3.3. 可测函数的极限:

- 点点收敛: $f_n\xrightarrow{p.w.}f$: $\forall\omega\in\Omega$, $f_n(\omega)\to f(\omega)$.
- 依测度收敛: $f_n\xrightarrow{m}f$: $\forall\sigma>0$, $m(\{\omega\mid |f_n-f|\geq\sigma\})\to 0$.

定理 3.4. 假设 $\mu(\Omega)<\infty$. 若 $f_n\xrightarrow{p.w.}f$, 则 $f_n\xrightarrow{m}f$; 反之, 若 $f_n\xrightarrow{m}f$, 则存在子列 $f_{n_k}\xrightarrow{p.w.}f$.

定义 3.5. 若一个可测函数只取有限个值，则称它是简单函数：$\Omega = \bigsqcup_{i=1}^n A_i$, $A_i \in \mathcal{F}$, $f = \sum_{i=1}^n a_i \chi_{A_i}$.

定理 3.6. $(\Omega, \mathcal{F})$ 上的函数 f 是可测的，当且仅当存在一列 $(\Omega, \mathcal{F})$ 上的简单函数 $\{f_n\}_{n=1}^\infty$, 使得 $f_n \xrightarrow{p.w.} f$.

定义 3.7. $(\Omega, \mathcal{F})$ 是一个概率空间，Ω 上的随机变量 X 就是可测函数 $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. 由 X 的拉回确定了一个 σ-子代数 $X^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{F}$, 记为 $\sigma(X)$.

例 3.8. $A \subset \Omega$ 是一个事件，特征函数 χ_A 是一个随机变量。$\mathbb{P}(\chi_A = 1) = \mathbb{P}(A)$, $\mathbb{P}(\chi_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. $\sigma(\chi_A) = \{\varnothing, A, A^c, \Omega\}$.

定义 3.9. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 是一个随机变量。由 X 可以在 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上定义一个测度 μ_X: $\forall A \in \mathcal{B}(\mathbb{R})$,

$$\mu_X\,(A) = \mathbb{P}(x \in A) = \mathbb{P}(X^{-1}(A)).$$

容易验证 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ 是一个概率空间。μ_X 称为 X 的分布。

$$\Phi_X(x) = \mu_X((\infty, x]) = \mathbb{P}(X \leq x)$$

称为 X 的分布函数。

性质 3.10. 分布函数的性质：

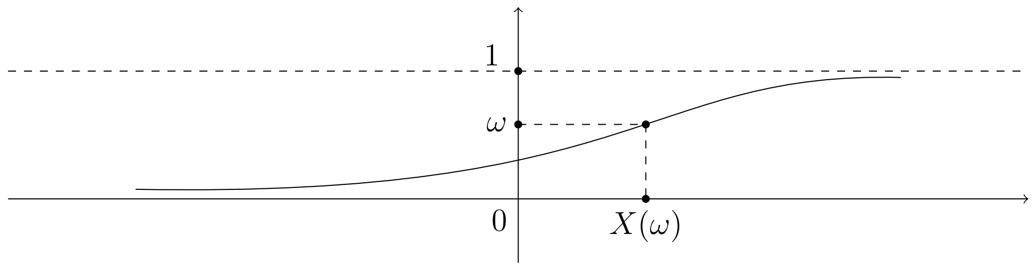
- 单调递增。
- $\lim_{x\to-\infty} \Phi_X(x) = 0$, $\lim_{x\to+\infty} \Phi_X(x) = 1$.
- 右连续性：
$$\lim_{x \to x_0^+} \Phi_X(x) = \Phi_X(x_0).$$

定理 3.11. 若 $\mathbb{R}$ 上的函数 $F(x)$ 满足性质 *3.10*的全部三条，则存在 $((0,1], \mathscr{L} = \textit{Lebesgue}\text{ 可测集}, \mathrm{d}\,\omega))$ 上的随机变量 X, 使得 $\Phi_X = F$.

证明. 简单情形：F 连续且严格单调递增。此时可以将 F 看成同胚 $\mathbb{R} \to (0,1)$. 定义

$$\begin{aligned} X : (0,1) &\to \mathbb{R} \\ \omega &\mapsto F^{-1}(\omega). \end{aligned}$$

则 $X(\omega) \le x$ 当且仅当 $\omega \le F(x)$.



从而

$$\Phi_X(x) = \mathbb{P}(X(\omega) \le x) = \mathrm{d}\,\omega(0, F(x)] = F(x).$$

而 $X(1)$ 可以任意定义，不影响分布。

一般情形：$\forall \omega \in (0,1)$, 定义广义逆元

$$F^{-1}(\omega) = \inf\{x \mid F(x) \ge \omega\}.$$

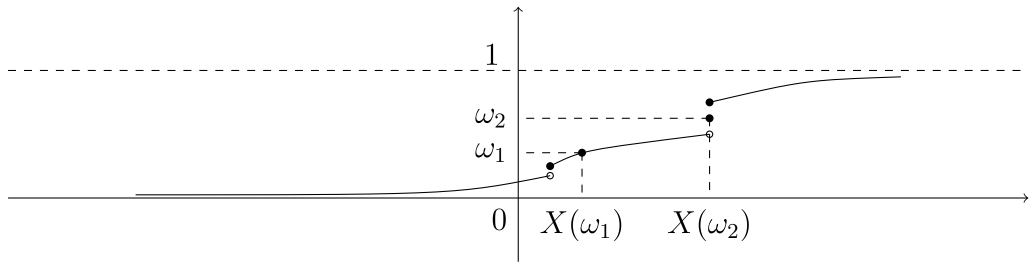
由 F 的右连续性，

$$F^{-1}(\omega) = \inf\{x \mid F(x) \ge \omega\} = \min\{x \mid F(x) \ge \omega\}.$$

定义

$$\begin{aligned} X : (0,1) &\to \mathbb{R} \\ \omega &\mapsto F^{-1}(\omega). \end{aligned}$$

则仍有 $X(\omega) \le x$ 当且仅当 $\omega \le F(x)$.



而 $X(\omega) \le x$ 能得到 $F(X(\omega)) \le F(x)$. 由 $X(\omega) = \min\{x \mid F(x) \ge \omega\}$ 的定义，$X(\omega) \in \{x \mid F(x) \ge \omega\}$. 从而有 $\omega \le F(X(\omega)) \le F(x)$. 反之，若 $F(x) \ge \omega$, 仍由定义，$x \in \{x \mid F(x) \ge \omega\}$. 故 $x \ge \min\{x \mid F(x) \ge \omega\} = X(\omega)$. 从而

$$\Phi_X(x) = \mathbb{P}(X(\omega) \le x) = \mathrm{d}\,\omega(0, F(x)] = F(x).$$

总结：若 F 满足性质 3.10的全部三条，在 $((0,1], \mathscr{L}, \mathrm{d}\,\omega))$ 上，X 是均匀分布的随机变量。$A \in \mathscr{L}$, $\mathbb{P}(x \in A) = \mathrm{d}\,\omega(A)$.

$$\mathbb{P}(X \le x) = \begin{cases} 0, & x \le 0; \\ x, & x \in (0,1); \\ 1, & x \ge 1. \end{cases}$$

$F^{-1}(X) \le x$ 当且仅当 $X \le F(x)$, 从而 $\Phi_{F^{-1}(X)}(x) = \mathbb{P}(F^{-1}(X) \le x) = \mathbb{P}(X \le F(x)) = F(x)$. 故 $F^{-1}(X)$ 的分布函数就是 F: $\Phi_{F^{-1}(X)} = F$. $\square$

例 3.12. $\Omega = \{H, T\}^{\mathbb{N}}$, $\Omega' \subset \Omega$.

$$\Omega' \leftrightarrow ((0,1], \mathscr{L}, \mathrm{d}\,\omega)$$

$$A_n \leftrightarrow B_n = \bigcup_{i=1}^{2^{n-1}} \left(\frac{2i-1}{2^n}, \frac{2i}{2^n} \right].$$

定义 $R_n = \chi_{B_n}$. 记

$$R(t) = \begin{cases} 1, & t \in \left(\frac{1}{2}, 1\right) \cup \{0\}; \\ 0, & t \in \left(0, \frac{1}{2}\right]. \end{cases}$$

称为 *Rademacher* 函数。$\omega \in (0, 1]$.

$$\chi_{B_1}(\omega) = R(\omega),\ \chi_{B_n} = R(2^{n-1}\omega).$$

定义随机变量

$$X = \sum_{i=1}^{\infty} \frac{\chi_{A_n}}{2^i}.$$

它是 $(0,1]$ 上的均匀分布。

$$\mathbb{P}\left(x \in \left(\frac{i}{2^n}, \frac{i+1}{2^n}\right]\right) = \frac{1}{2^n}.$$

$\left(\frac{l_j}{2^{n_j}}, \frac{k_j}{2^{m_j}}\right] \nearrow (a, b]$, $\frac{l_j}{2^{n_j}} \searrow a$, $\frac{k_j}{2^{m_j}} \nearrow b$. 例如：$X \in \left(\frac{3}{8}, \frac{4}{8}\right]$ 当且仅当 $x \in \overline{A_1} \cap A_2 \cap A_3$.

任取分布函数 F,

$$\Omega \xrightarrow{X} (0, 1] \xrightarrow{F^{-1}} \mathbb{R}.$$

F^{-1} 服从 $(0,1]$ 上的均匀分布，$F^{-1}(X)$ 服从 F 分布。$F^{-1}(X)$ 看作是 ω 上的随机变量。

定义 3.13. 称两个随机变量 X, Y 依分布相等，若从而 $\mu_X = \mu_Y$ 能得到 $\Phi_X = \Phi_Y$. 记为 $X \overset{d}{=} Y$.

注记 3.14. 依分布相等不要求 X, Y 定义在同一个概率空间上。

定义 3.15. $X, \{X_n\}_{n=1}^{\infty}$ 是随机变量。称 $\{X_n\}_{n=1}^{\infty}$ 依分布收敛到 X, 若对任意 Φ_X 的连续点 a, 有 $\Phi_{X_n}(a) \to \Phi_X(a)$. 记为 $X_n \xrightarrow{d} X$.

定理 3.16. 若 $X_n \xrightarrow{m} X$, 则 $X_n \xrightarrow{d} X$.

证明. $\forall \varepsilon > 0$,

$$\mathbb{P}(X_n \le a) \le \mathbb{P}(X \le a + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).$$

$$\mathbb{P}(X_n \le a-\varepsilon) \le \mathbb{P}(X \le a) + \mathbb{P}(|X_n - X| > \varepsilon).$$

$$\Phi_X(a-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \le \Phi_{X_n}(a) \le \Phi_X(a+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).$$

令 $n \to \infty$ 有

$$\Phi_X(a-\varepsilon) \le \liminf_{n\to\infty} \Phi_{X_n}(a) \le \limsup_{n\to\infty} \Phi_{X_n}(a) \le \Phi_X(a+\varepsilon).$$

令 $\varepsilon \to 0^+$, 在 Φ_X 的连续点 a 处，$\Phi_X(a) = \Phi_X(a^-) = \Phi_X(a^+)$, 故上式等号同时成立，即有

$$\lim_{n\to\infty} \Phi_{X_n}(a) = \Phi_X(a).$$

即 $X_n \xrightarrow{d} X$. □

应用 3.17. 回到上例。

$$X = \sum_{i=1}^{\infty} \frac{\chi_{A_n}}{2^i}.$$

选取有限子列 $n_1 < n_2 < \cdots < n_k < \cdots$.

$$Y_k = \sum_{i=1}^{k} \frac{\chi_{A_{n_i}}}{2^i}.$$

将 $A_{n_1}, A_{n_2}, \cdots, A_{n_k}$ 的投掷结果重排到第 $a, 2, \cdots, k$ 位上看。

$$\Phi_{Y_k}(x) = \begin{cases} 0, & x \le 0; \\ 2^{-k}\left(1 + \lfloor 2^k x \rfloor\right), & 0 < x \le 1; \\ 1, & x > 1. \end{cases}$$

$Y_k \to Y$ 服从 $(0,1]$ 上的均匀分布。

定义 3.18. 称随机变量 X 是（绝对）连续的，若它的分布函数 Φ_X 是（绝对）连续的。

我们从实分析中知道，绝对连续函数几乎处处可以求导。于是有定义

定义 3.19. 对一个绝对连续型随机变量 X, 存在几乎处处连续函数 f, 使得

$$\Phi_X(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}\,x.$$

$f(x)$ 称为 X 的密度函数。

由于 $\mathbb{P}(X = x) = \Phi_X(x) - \Phi_X(x^-) = f(x)$ 几乎处处成立, 所以 $D = \{x \mid \mathbb{P}(X = x) > 0\}$ 至多是可数集。于是有定义:

定义 3.20. $\forall x \in \mathbb{R}$,

$$\Phi_X(x) = \sum_{y \leq x, y \in D} \mathbb{P}(X = y)$$

称为离散型随机变量。

例 3.21. $X \in \{0, 1\}$, $\mathbb{P}(X = 0) = p$, $\mathbb{P}(X = 1) = 1 - p$. 称为 *Bernoulli* 随机变量。

定义 3.22. 设 $X_1, X_2, \cdots, X_n$ 是 n 个随机变量。$X = (X_1, X_2, \cdots, X_n)$ 称为一个随机向量。

$$X : (\Omega, \mathcal{D}, \mathbb{P}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$$

是可测映射，成为联合分布。定义

$$\Phi_X(x_1, x_2, \cdots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$$

称为 X 的联合分布函数。

性质 3.23. 对联合分布 Φ_X,

1. 对每个变量连续。
2. $\Phi_X(\infty, \infty, \cdots, \infty) = 1$; 只要某个 $x_i = -\infty$, 就有 $\Phi_X(x_1, x_2, \cdots, x_n) = 0$.

3. 右连续性：$y_1 \to x_1^+, y_1 \to x_1^+, \cdots, y_n \to x_n^+$，则 $\Phi_X(y_1, y_2, \cdots, y_n) \to \Phi_X(x_1, x_2, \cdots, x_n)$.

4. 对任意 $a_1 < b_1, a_2 < b_2, \cdots, a_n < b_n$，要有 $\mathrm{D}^1_{a_1,b_1} \cdots \mathrm{D}^{n-1}_{a_{n-1},b_{n-1}} \mathrm{D}^n_{a_n,b_n} \Phi_X \geq 0$. 其中 $\mathrm{D}^i_{a_i,b_i}$ 是差分：

$$\mathrm{D}^i_{a_i,b_i} \Phi_X = \Phi_X(x_1, \cdots, x_{i-1}, b_i, x_{i+1}, \cdots, x_n) = \Phi_X(x_1, \cdots, x_{i-1}, a_i, x_{i+1}, \cdots, x_n).$$

注记 3.24.

$$\mathrm{D}^1_{a_1,b_1} \cdots \mathrm{D}^{n-1}_{a_{n-1},b_{n-1}} \mathrm{D}^n_{a_n,b_n} \Phi_X = \mathbb{P}(X_1 \in (a_1, b_1], X_2 \in (a_2, b_2], \cdots, X_n \in (a_n, b_n]).$$

定理 3.25. 若 $\mathbb{R}^n$ 上的函数 Φ 满足性质 *3.23*的全部四条，则存在 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathrm{d}\,\omega))$ 上的随机变量 X，使得 $\Phi_X = \Phi$. 事实上，记 X_i 是 X 在第 i 个分量上的投影，则 $X = (X_1, \cdots, X_n)$ 即为所求。

证明. $\mathcal{B}(\mathbb{R}^n)$ 是由 $\prod_{i=1}^n (a_i, b_i]$ 生成的 σ-代数。由长方体生成的 Boole 代数，是所有由有限个这样的长方体的并构成的集合。定义 $m : \mathcal{B}(\mathbb{R}^n) \to \mathbb{R}^+$: $\forall A = \prod_{i=1}^n (a_i, b_i]$,

$$m(A) = m\left(\prod_{i=1}^n (a_i, b_i]\right) = \mathrm{D}^1_{a_1,b_1} \cdots \mathrm{D}^{n-1}_{a_{n-1},b_{n-1}} \mathrm{D}^n_{a_n,b_n} \Phi_X.$$

由性质 4, $m(A) \geq 0$. 对长方体不同的细分（将长方体分解成不交的小长方体有许多种方式）：$A = A_1 \sqcup \cdots A_k$, 要验证定义合理：$m(A) = m(A_1) + \cdots + m(A_k)$.

要验证 m 是 $\mathcal{B}(\mathbb{R}^n)$ 上的测度：若 $A_n \searrow \varnothing$, 则 $m(A_n) \to 0$. 由测度扩张定理，在 $\mathcal{B}(\mathbb{R}^n)$ 上的测度 $m^* = \mathbb{P}$.

$$\mathbb{P}(X_1 \leq a_1, X_2 \leq a_2, \cdots, X_n \leq a_n) = m\left(\prod_{i=1}^n (-\infty, a_i]\right) = \Phi_X(a_1, a_2, \cdots, a_n).$$

从而 $\Phi_X = \Phi$. □

有时候我们要研究定义在两个不同的空间上的随即变量 X, Y。此时我们无法讨论它们的独立性。我们希望找到 $\mathbb{R}$ 上的两个随即变量 $\widehat{X}, \widehat{Y}$, 使得 $\Phi_{\widehat{X}} = \Phi_X$, $\Phi_{\widehat{Y}} = \Phi_Y$, 这样我们就可以通过研究 $\widehat{X}, \widehat{Y}$ 的独立性得到 X, Y 的独立性。

定理 3.26. 设 $\{F_n\}_{n=1}^\infty$ 是一列分布函数。则在 $((0,1], \mathcal{B}((0,1]), \mathrm{d}\,\omega)$ 上，存在独立的随机变量序列 $X_1, X_2, \cdots$，满足 $\Phi_{X_n} = F_n$, $n = 1, 2, \cdots$.

证明. 把自然数集 $\mathbb{N}$ 做分化：令 $N_j = \{(2j-1)\cdot 2^k \mid k = 1, 2, \cdots\}$, 则 $\mathbb{N} = N_1 \sqcup N_2 \sqcup \cdots$. 令

$$Y_j = \sum_{k=1}^{\infty} \frac{\chi_{A_{(2j-1)2^k}}}{2^k},$$

则 $Y_1, Y_2, \cdots$ 相互独立，都是 $(0,1]$ 上的均匀分布。令 $X_n = F_n^{-1}(Y_n)$, 则 $X_1, X_2, \cdots$ 也是相互独立的，且有 $\Phi_{X_n} = F_n$. $\square$

注记 3.27. 证明的灵感来自于无穷次投硬币的模型。

例 3.28 (二项分布)**.** 设 $X_1, \cdots, X_n$ 是独立同分布的 *Bernoulli* 随机变量，$\mathbb{P}(X_n = 0) = 1-p$, $\mathbb{P}(X_n = 1) = p$, $n = 1, \cdots, n$. 令 $X = X_1 + \cdots + X_n$. $\forall 0 \le k \le n$,

$$\mathbb{P}(X=k) = \mathbb{P}\left(x_1, \cdots, X_n \text{ 中恰好有 } k \text{ 个 } 1, n-k \text{ 个 } 0\right) = \sum_{1 \le i_1 \le \cdots \le i_k \le n} \mathbb{P}\left(X_{i_1} = \cdots = X_{i_k} = 1, X_j = 0, j \ne i_1, \cdots, i_k\right) = C_n^k \cdot p^k \cdot (1-p)^{n-k}.$$

称 X 服从二项分布，记为 $X \sim B(n,p)$.

例 3.29 (Poisson 分布)**.** 设 $\lambda > 0$ 是参数，定义随机变量 X:

$$\mathbb{P}(X=k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!},\ k = 0, 1, 2, \cdots.$$

则称 X 服从 *Poisson* 分布，记为 $X \sim P(\lambda)$.

若两个独立的 *Poisson* 分布 $X = P(\lambda_1), Y = P(\lambda_2)$. 则 $Z = X + Y \sim P(\lambda_1 + \lambda_2)$:

$$\begin{aligned}\mathbb{P}(Z=k) &= \sum_{i=0}^{k} \mathbb{P}(X=i, Y=k=i) = \sum_{i=0}^{k} \mathbb{P}(X=i) \cdot \mathbb{P}(Y=k-i) \\ &= \sum_{i=0}^{k} \mathrm{e}^{-\lambda_1} \cdot \frac{{\lambda_1}^i}{i!} \cdot \mathrm{e}^{-\lambda_2} \cdot \frac{{\lambda_1}^{k-i}}{(k-i)!} = \frac{\mathrm{e}^{-\lambda_1-\lambda_2}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} {\lambda_1}^i {\lambda_1}^{k-i} = \frac{\mathrm{e}^{-\lambda_1-\lambda_2}}{k!} \sum_{i=0}^{k} C_k^i {\lambda_1}^i {\lambda_1}^{k-i} = \mathrm{e}^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1+\lambda_2)^k}{k!}.\end{aligned}$$

例 3.30 (Bernoulli 分布与 Poisson 分布的比较)**.** 设 $Y \sim P(\lambda)$.

$$F_\lambda(x) = \mathbb{P}(Y \le x) = \mathrm{e}^{-\lambda} \sum_{i \le x} \frac{\lambda^i}{i!}$$

是 Y 的分布函数。设 $X \sim (0,1]$ 是均匀分布。

$$Y(\omega) = F_\lambda^{-1}(X)(\omega) = \begin{cases} 0, & \omega \le \mathrm{e}^{-\lambda} \\ 1, & \omega \in [\mathrm{e}^{-\lambda}, \mathrm{e}^{-\lambda}(1+\lambda)) \\ 2, & \omega \in [\mathrm{e}^{-\lambda}(1+\lambda), \mathrm{e}^{-\lambda}(2+\lambda)) \\ \cdots. \end{cases}$$

令 $G_\lambda = \chi_{[\mathrm{e}^{-\lambda}, \mathrm{e}^{-\lambda}+\lambda)}$, 令 $Z = G_\lambda(X)$. 则 $Z \sim B(1, \lambda)$. 此时,

$$\mathbb{P}(Y \ne Z) = \mathbb{P}\left(\omega \in [\mathrm{e}^{-\lambda}(1+\lambda), 1]\right) = 1 - \mathrm{e}^{-\lambda} - \lambda\,\mathrm{e}^{-\lambda} \le \lambda^2.$$

一般地, 设 $T_1, \cdots, T_n \sim P\left(\frac{\lambda}{n}\right)$ 互相独立, 则 $T = T_1 + \cdots + T_n \sim P(\lambda)$. 设 $W_1, \cdots, W_n \sim B\left(1, \frac{\lambda}{n}\right)$ 互相独立, 则 $W = W_1 + \cdots + W_n \sim B\left(n, \frac{\lambda}{n}\right)$. 设 $X_1, \cdots, X_n$ 是互相独立的 $(0,1]$ 上的均匀分布。

$$T = T_1 + \cdots + T_n = F_{\frac{\lambda}{n}}^{-1}(X_1) + \cdots + F_{\frac{\lambda}{n}}^{-1}(X_n) \sim P(\lambda).$$

$$W = W_1 + \cdots + W_n = G_{\frac{\lambda}{n}}(X_1) + \cdots + G_{\frac{\lambda}{n}}(X_n) \sim B\left(n, \frac{\lambda}{n}\right).$$

由于 $\{T \ne W\} \subset \bigcup_{i=1}^n \{T_i \ne W_i\}$,

$$\mathbb{P}(T \ne W) = \mathbb{P}(\exists i, T_i \ne W_i) \le \sum_{i=1}^n \mathbb{P}(T_i \ne W_i) \le n \cdot \left(\frac{\lambda}{n}\right)^2 = \frac{\lambda^2}{n}.$$

从而

$$|\mathbb{P}(T \le x) - \mathbb{P}(W \le x)| = |\Phi_T(x) - \Phi_W(x)| \le \mathbb{P}(T \ne W) \le \frac{\lambda^2}{n}.$$

将 W 记为 S_n, 是 n 个独立服从 $B(1, \frac{\lambda}{n})$ 的独立随机变量的和, 就有 $\forall x$, $\Phi_{S_n}(x) \to \Phi_T(x)$, 即 $S_n \xrightarrow{d} T$.

注记 3.31. 我们用 $B(n, \frac{\lambda}{n})$ 逼近了 $P(\lambda)$. 这种方法成为耦合。

定义 3.32. 设 X, Y 是两个随机变量，若存在某个概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上存在随机变量 $\widehat{X}, \widehat{Y}$, 使得 $\Phi_{\widehat{X}} = \Phi_X, \Phi_{\widehat{Y}} = \Phi_Y$. 则称 $(\widehat{X}, \widehat{Y})$ 是 (X, Y) 的耦合。

注记 3.33. 通过耦合研究随机变量 X, Y 的关系，只可能得到关于 X, Y 的分布函数的关系。

定义 3.34. 设 X, Y 是两个随机变量，称 X 随机小于 Y, 若存在 (X, Y) 的耦合 $(\widehat{X}, \widehat{Y})$, 使得 $\mathbb{P}(\widehat{X} \leq \widehat{Y}) = 1$, 记为 $X \preceq Y$.

定义 3.35. 记 $\mathbb{P}(X \leq Y) = 1$ 为 $X \leq Y(a.e.)$*(almost everywhere)* 或 $X \leq Y(a.c.)$*(almost certainly)*.

定理 3.36. $X \preceq Y$, 当且仅当 $\forall x$, $\Phi_Y(x) \leq \Phi_X(x)$.

证明. 必要性：若 $X \preceq Y$, 则存在耦合 $(\widehat{X}, \widehat{Y})$, 使得 $\widehat{X} \leq \widehat{Y}(a.c.)$. 注意到由 $\widehat{X} \leq \widehat{Y}(a.c.)$ 可得，$\forall x$, $(\widehat{Y} \leq x) \subset (\widehat{X} \leq x)(a.c.)$. 从而

$$\Phi_Y(x) = \Phi_{\widehat{Y}}(x) = \mathbb{P}(\widehat{Y} \leq x) \leq \mathbb{P}(\widehat{X} \leq x) = \Phi_X(x) = \Phi_{\widehat{X}}(x).$$

充分性：若 $\forall x$, $\Phi_Y(x) \leq \Phi_X(x)$. $\forall \omega \in (0,1]$, 令 $\widehat{X} = \Phi_X^{-1}(\omega)$, $\widehat{Y} = \Phi_Y^{-1}(\omega)$. 则 $\widehat{X}, \widehat{Y}$ 是 $((0,1], \mathcal{B}((0,1]), \mathrm{d}\,\omega)$ 上的随机变量，$\Phi_{\widehat{X}} = \Phi_X, \Phi_{\widehat{Y}} = \Phi_Y$. 从而 $(\widehat{X}, \widehat{Y})$ 是 (X, Y) 的耦合。而 $\forall x$, $\widehat{Y}(\omega) = \Phi_Y^{-1}(\omega) \leq x$ 等价于 $\omega \leq \Phi_Y(x)$. 而这可以推出 $\omega \leq \Phi_Y(x) \leq \Phi_X(x)$, 这等价于 $\widehat{X}(\omega) = \Phi_X^{-1}(\omega) \leq x$. 从而 $\widehat{X} \leq \widehat{Y}$ 在 $(0,1]$ 上处处成立，从而 $\mathbb{P}(\widehat{X} \leq \widehat{Y}) = 1$, 即 $X \preceq Y$. $\square$

4 期望和方差

定义 4.1. 设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，X 是 Ω 上的随机变量。则 X 的期望 $\mathbb{E}[X]$ 定义为

$$\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\,\mathbb{P}.$$

$\mathbb{E}[X]$ 存在，可能是 $\pm\infty$, 故 X 的期望存在。若 $-\infty < \mathbb{E}[X] < +\infty$, 则称 X 可积。

回顾实分析：如何定义可测函数的积分？

1. 定义简单函数的积分：$f = \sum_{i=1}^m a_i \chi_{A_i}$, 则定义 $\int_\Omega f \,\mathrm{d}\,\mathbb{P} = \sum_{i=1}^m a_i \mathbb{P}(A_i)$. 要验证定义与简单函数的表示无关。
2. 定义非负函数的积分：用一列简单函数 $\{f_n\}_{n=1}^\infty$ 逼近 f，定义 $\int_\Omega f \,\mathrm{d}\,\mathbb{P} = \lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\,\mathbb{P}$. 要验证定义与简单函数列的选取无关。$\int_\Omega f \,\mathrm{d}\,\mathbb{P} \in [0, +\infty]$.
3. 定义一般函数的积分：将 f 分成正负部分：$f = f^+ - f^-$. 则 f^+, f^- 都是非负函数，$\int_\Omega f^+ \,\mathrm{d}\,\mathbb{P}, \int_\Omega f^- \,\mathrm{d}\,\mathbb{P} \in [0, +\infty]$. 如果 $\int_\Omega f^+ \,\mathrm{d}\,\mathbb{P}, \int_\Omega f^- \,\mathrm{d}\,\mathbb{P}$ 都不是 $+\infty$, 则称 f 的积分存在，$\int_\Omega f \,\mathrm{d}\,\mathbb{P} = \int_\Omega f^+ \,\mathrm{d}\,\mathbb{P} - \int_\Omega f^- \,\mathrm{d}\,\mathbb{P} \in [-\infty, +\infty]$. 若 $\int_\Omega f^+ \,\mathrm{d}\,\mathbb{P}, \int_\Omega f^- \,\mathrm{d}\,\mathbb{P}$ 都有限，则 $\int_\Omega f \,\mathrm{d}\,\mathbb{P}$ 有限，此时称 f 可积。事实上，只需要其中一个有限，避免 $\infty - \infty$ 的情况，即可定义积分。

注意到 $|f| = f^+ + f^-$, 从而 $\int_\Omega |f| \,\mathrm{d}\,\mathbb{P} = \int_\Omega f^+ \,\mathrm{d}\,\mathbb{P} + \int_\Omega f^- \,\mathrm{d}\,\mathbb{P}$. 故 f 可积等价于 $|f|$ 可积。可测函数的积分有线性性质。

性质 4.2. 记 $\mathcal{M}_B(\Omega, \mathcal{F})$ 是 $(\Omega, \mathcal{F})$ 上的所有有界可测函数构成的实线性空间，

$$\mathbb{E} : \mathcal{M}_B(\Omega, \mathcal{F}) \to \mathbb{R}$$

是 $\mathbb{R}$-线性映射。有性质：

1. 归一化条件：$\mathbb{E}[1] = \mathbb{E}[\chi_\Omega] = 1$.

2. 单调性：若 $X \geq 0$, 则 $\mathbb{E}[X] \geq 0$. 这等价于若 $X \geq Y$, 则 $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

3. 连续性：设 $X \in \mathcal{M}_B(\Omega, \mathcal{F})$, $\{X_n\}_{n=1}^{\infty} \subset \mathcal{M}_B(\Omega, \mathcal{F})$. 若 $\{X_n\}_{n=1}^{\infty}$ 处处单调收敛到 X: $X_n \nearrow X(p.w.)$, 则 $\mathbb{E}[X_n] \to \mathbb{E}[X]$. 这就是单调收敛定理。

4. 由 $\mathbb{E}$ 也唯一确定了 $\mathbb{P}$. 这是因为 $\forall A \in \mathcal{F}$, $\mathbb{P}(A) = \mathbb{E}[\chi_A]$.

5. 线性：$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

例 4.3 (Bernoulli 随机变量). $X = \chi_A$, 则

$$\mathbb{E}[X] = \int_\Omega \chi_A \, \mathrm{d}\,\mathbb{P} = \mathbb{P}(A) = \mathbb{P}(X=1).$$

例 4.4 (离散随机变量). X 可以取可数个值 $a_1, a_2, \cdots \geq 0$, 有负数就将正负分开。记 $A_i = \{\omega : X = a_i\}$, $X_n = \sum_{i=1}^n a_i \cdot \chi_{A_i}$. 则有 $X_n \nearrow X$, 从而

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n] = \sum_{n=1}^{\infty} a_i \cdot \mathbb{P}(A_i) = \sum_{n=1}^{\infty} a_i \cdot \mathbb{P}(X = a_i).$$

这事实上就是取每个值的加权平均。

例 4.5 (投 n 次硬币的 Bernoulli 实验). 不均匀硬币：$\mathbb{P}(T) = p$, $\mathbb{P}(H) = 1-p$. $\Omega = \{H, T\}^n$. 令 X 是朝上 (T) 的次数。记 A_i 是第 i 次为 T. 则 $X = \chi_{A_1} + \cdots + \chi_{A_n}$. 从而

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[\chi_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^n p = np.$$

定义 4.6. 设 X 是 $(\Omega, \mathcal{F}, \mathbb{P})$ 上一个随机变量，$f : \mathbb{R} \to \mathbb{R}$ 是 *Borel* 可测函数，则 $f(X)$ 仍然是 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的可测函数，从而是 Ω 上的一个随机变量。$\mathbb{E}[f(X)]$ 称为 X 的一个数量特征。

例 4.7.

- $f(x) = x^k$. $\mathbb{E}[X^k]$ 称为 X 的 k 阶矩。

- $f(x) = (x - \mathbb{E}[X])^2$. $\mathbb{E}[(x = \mathbb{E}[X])^2]$ 称为 X 的方差，记为 $\sigma^2(X)$. 由 $(x - \mathbb{E}[X])^2 = x^2 - 2x\mathbb{E}[X] + \mathbb{E}[X]^2$, 由期望的线性，

$$\sigma^2(X) = \mathbb{E}\left[(x = \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

例 4.8. 设 $\mathbb{P}(A)=p$, 由例 *4.3*, 则

$$\sigma^2(\chi_A)=\mathbb{E}\left[\chi_A^2\right]-\mathbb{E}[\chi_A]^2=\mathbb{E}\left[\chi_A\right]-\mathbb{E}[\chi_A]^2=p-p^2.$$

定理 4.9. 设 X 是 $(\Omega,\mathcal{F},\mathbb{P})$ 上一个随机变量，$f:\mathbb{R}\to\mathbb{R}$ 是 *Borel* 可测函数，若 $f\geq 0$ 或 $\mathbb{E}\left[|f(X)|\right]<+\infty$, 则

$$\mathbb{E}\left[f(X)\right]=\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_X.$$

其中，$\mu_X=X_*\mathbb{P}$ 是 $\mathbb{R}$ 上的 *Borel* 测度：$\mu_X(A)=\mathbb{P}(X\in A)$, $A\subset\mathbb{R}$ 是 *Borel* 可测集。特别地，$\mathbb{E}\left[X\right]=\int_{\mathbb{R}}x\,\mathrm{d}\,\mu_X$.

证明. 只在 $f\geq 0$ 时证明，否则若 $\mathbb{E}\left[|f(X)|\right]<+\infty$ 时，考虑 $f=f^+-f^-$ 即可。

取一列递增的非负简单函数 $f_n\nearrow f$, 其中 $f_n=\min\left\{2^{-n}\left\lfloor 2^n f\right\rfloor,n\right\}$, 从而 $f_n(X)\nearrow f(X)$, 这是因为

$$f_n(x)=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\chi_{B_i},\ f_n(X)=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\chi_{X^{-1}(B_i)}.$$

由单调收敛定理, $\int_\Omega f(X)\,\mathrm{d}\,\mathbb{P}=\lim_{n\to\infty}\int_\Omega f_n(X)\,\mathrm{d}\,\mathbb{P}$, 而且 $\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_X=\lim_{n\to\infty}\int_{\mathbb{R}}f_n(x)\,\mathrm{d}\,\mu_X$. 注意到 f_n 是简单函数, 取值为 $\{0,\frac{1}{2^n},\cdots,n-\frac{1}{2^n},n\}$, 从而 $\int_\Omega f_n(X)\,\mathrm{d}\,\mathbb{P}=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\mathbb{P}\left(f_n(X)=i\right)$. 记 $B_i=\left[\frac{i}{2^n},\frac{i+1}{2^n}\right)$, $0\leq i<n\cdot 2^n$, $B_{n\cdot 2^n}=[n,+\infty)$. 从而

$$\int_\Omega f_n(X)\,\mathrm{d}\,\mathbb{P}=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\mathbb{P}\left(f_n(X)=i\right)=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\mathbb{P}\left(X\in B_i\right)=\sum_{i=0}^{n\cdot 2^n}\frac{i}{2^n}\cdot\mu_X(B_i)=\int_{\mathbb{R}}f_n(x)\,\mathrm{d}\,\mu_X.$$

从而

$$\mathbb{E}\left[f(X)\right]=\int_\Omega f(X)\,\mathrm{d}\,\mathbb{P}=\lim_{n\to\infty}\int_\Omega f_n(X)\,\mathrm{d}\,\mathbb{P}=\lim_{n\to\infty}\int_{\mathbb{R}}f_n(x)\,\mathrm{d}\,\mu_X=\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_X.$$

□

注记 4.10. $\mathbb{E}[f(X)]=\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_X$ 由 f 和 μ_X 确定，而 μ_X 由 Φ_X 确定：$\mu_X((a,b])=\Phi_X(b)-\Phi_X(a)$, 从而 $\mathbb{E}[f(X)]$ 由 f 和 Φ_X 确定。这说明数量特征只与随机变量的分布函数有关。

定义 4.11. 设 $\{X_n\}_{n=1}^\infty,X$ 是概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量。

- 称 $\{X_n\}_{n=1}^{\infty}$ 几乎处处收敛到 X, 若 $\mathbb{P}(\{\omega|X_n(\omega)\to X(\omega)\})=1$, 记为 $X_n \xrightarrow{a.e./a.c.} X$.
- 称 $\{X_n\}_{n=1}^{\infty}$ 依测度收敛到 X, 若 $\forall\varepsilon>0$, $\mathbb{P}(|X_n-X|\geq\varepsilon)\to 0$, 记为 $X_n \xrightarrow{m} X$.
- 称 $\{X_n\}_{n=1}^{\infty}$ 依分布收敛到 X, 若对任意 Φ_X 的连续点 a, $\Phi_{X_n}(a)\to\Phi_X(a)$, 记为 $X_n \xrightarrow{d} X$.

定理 4.12. 设 $\{X_n\}_{n=1}^{\infty}, X$ 是随机变量。则 $X_n \xrightarrow{d} X$, 当且仅当对任意有界连续函数 $f:\mathbb{R}\to\mathbb{R}$, 都有 $\mathbb{E}[f(X_n)]\to\mathbb{E}[f(X)]$.

证明. 必要性: $\forall\varepsilon>0$, 取定 Φ_X 的连续点 $a<b$, 满足 $\Phi_X(a)\leq\varepsilon$, $1-\Phi_X(b)\leq\varepsilon$. 而 f 在紧区间 $[a,b]$ 上一致连续, $\exists\delta>0$, 使得 $\forall x,y\in[a,b]$, 满足 $|x-y|\leq\delta$, 都有 $|f(x)-f(y)|\leq\varepsilon$. f 的上界记为 M. 将 $[a,b]$ 分成长度小于 δ 的区间: $a=a_0<a_1<\cdots<a_m=b$, $a_{i+1}-a_i<\delta$, 且 Φ_X 在 $a_1,\cdots,a_{m-1}$ 处都连续。由于

$$\mathbb{E}[f(X)]=\int_{\mathbb{R}} f(x)\,\mathrm{d}\,\mu_X=\int_{(-\infty,a]} f(x)\,\mathrm{d}\,\mu_X+\sum_{i=1}^{m}\int_{(a_{i-1},a]} f(x)\,\mathrm{d}\,\mu_X+\int_{[b,+\infty)} f(x)\,\mathrm{d}\,\mu_X.$$

从而

$$\begin{aligned}
\left|\mathbb{E}[f(X)]-\sum_{i=1}^{m} f(a_i)\left(\Phi_X(a_i)-\Phi_X\left(a_{i-1}\right)\right)\right| &= \left|\int_{(-\infty,a]} f(x)\,\mathrm{d}\,\mu_X+\sum_{i=1}^{m}\int_{(a_{i-1},a]} f(x)\,\mathrm{d}\,\mu_X+\int_{[b,+\infty)} f(x)\,\mathrm{d}\,\mu_X-\sum_{i=1}^{m} f(a_i)\left(\Phi_X(a_i)-\Phi_X\left(a_{i-1}\right)\right)\right| \\
&\leq \left|\int_{(-\infty,a]} f(x)\,\mathrm{d}\,\mu_X\right|+\left|\int_{[b,+\infty)} f(x)\,\mathrm{d}\,\mu_X\right|+\left|\sum_{i=1}^{m}\int_{(a_{i-1},a]}\left(f(x)-f(a_i)\right)\mathrm{d}\,\mu_X\right| \\
&\leq \left|\int_{(-\infty,a]} f(x)\,\mathrm{d}\,\mu_X\right|+\left|\int_{[b,+\infty)} f(x)\,\mathrm{d}\,\mu_X\right|+\sum_{i=1}^{m}\int_{(a_{i-1},a]}\left|f(x)-f(a_i)\right|\mathrm{d}\,\mu_X \\
&\leq M\Phi_X(a)+M(1-\Phi_X(b))+\varepsilon\sum_{i=1}^{m}\left(\Phi_X(a_i)-\Phi_X(a_{i-1})\right) \\
&\leq M\Phi_X(a)+M(1-\Phi_X(b))+\varepsilon\left(\Phi_X(b)-\Phi_X(a)\right) \\
&\leq (2M+1)\varepsilon.
\end{aligned}$$

由于 $\Phi_{X_n}(a_i) \to \Phi_X(a_i)$. 不妨设 $\forall n$, $\Phi_{X_n}(a) \le 2\varepsilon$, $1-\Phi_{X_n}(b) \le 2\varepsilon$. 同理可得，

$$\left|\mathbb{E}[f(X_n)] - \sum_{i=1}^{m} f(a_i)\left(\Phi_{X_n}(a_i) - \Phi_{X_n}(a_{i-1})\right)\right| \le (4M+1)\varepsilon.$$

综上，

$$\begin{aligned}
&|\mathbb{E}[f(X)] - \mathbb{E}[f(X_n)]| \\
\le&\left|\mathbb{E}[f(X)] - \sum_{i=1}^{m} f(a_i)\left(\Phi_X(a_i) - \Phi_X(a_{i-1})\right)\right| + \left|\mathbb{E}[f(X_n)] - \sum_{i=1}^{m} f(a_i)\left(\Phi_{X_n}(a_i) - \Phi_{X_n}(a_{i-1})\right)\right| \\
&+\left|\sum_{i=1}^{m} f(a_i)\left[(\Phi_{X_n}(a_i) - \Phi_X(a_i)) - (\Phi_{X_n}(a_{i-1}) - \Phi_X(a_{i-1}))\right]\right| \\
\le&(6M+2)\varepsilon + \left|\sum_{i=1}^{m} f(a_i)\left[(\Phi_{X_n}(a_i) - \Phi_X(a_i)) - (\Phi_{X_n}(a_{i-1}) - \Phi_X(a_{i-1}))\right]\right|.
\end{aligned}$$

由于 $\Phi_{X_n}(a_i) \to \Phi_X(a_i)$. 从而

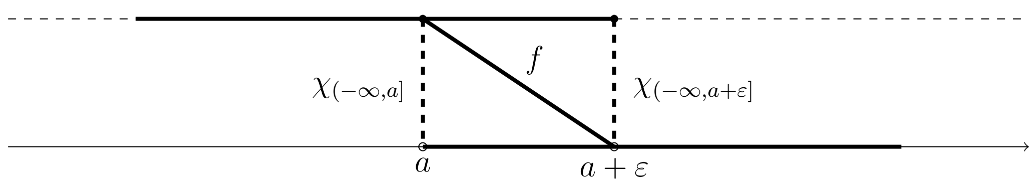
$$\left|\sum_{i=1}^{m} f(a_i)\left[(\Phi_{X_n}(a_i) - \Phi_X(a_i)) - (\Phi_{X_n}(a_{i-1}) - \Phi_X(a_{i-1}))\right]\right| \to 0.$$

于是有

$$0 \le \limsup_{n\to\infty} |\mathbb{E}[f(X)] - \mathbb{E}[f(X_n)]| \le (6M+2)\varepsilon.$$

由 ε 的任意性，$\limsup_{n\to\infty} |\mathbb{E}[f(X)] - \mathbb{E}[f(X_n)]| = 0$, 从而 $\lim_{n\to\infty} |\mathbb{E}[f(X)] - \mathbb{E}[f(X_n)]| = 0$. 即 $\mathbb{E}[f(X)] \to \mathbb{E}[f(X_n)]$.

充分性：$\forall a \in \mathbb{R}$, $\forall \varepsilon > 0$. 取连续函数 f, 使得 $\chi_{(-\infty,a]} \le f \le \chi_{(-\infty,a+\varepsilon]}$.



从而

$$\mathbb{E}\left[f(X_n)\right]=\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_{X_n}\geq\int_{\mathbb{R}}\chi_{(-\infty,a]}\,\mathrm{d}\,\mu_{X_n}=\int_{-\infty}^{a}\mathrm{d}\,\mu_{X_n}=\mu_{X_n}((-\infty,a])=\Phi_{X_n}(a)-\Phi_{X_n}(-\infty)=\Phi_{X_n}(a).$$

$$\mathbb{E}\left[f(X)\right]=\int_{\mathbb{R}}f(x)\,\mathrm{d}\,\mu_{X}\leq\int_{\mathbb{R}}\chi_{(-\infty,a+\varepsilon]}\,\mathrm{d}\,\mu_{X}=\int_{-\infty}^{a+\varepsilon}\mathrm{d}\,\mu_{X}=\mu_{X}((-\infty,a+\varepsilon])=\Phi_{X}(a+\varepsilon)-\Phi_{X}(-\infty)=\Phi_{X}(a+\varepsilon).$$

由假设，$\mathbb{E}[f(X_n)]\to\mathbb{E}[f(X)]$. 从而

$$\limsup_{n\to\infty}\Phi_{X_n}(a)\leq\lim_{n\to\infty}\mathbb{E}[f(X_n)]=\mathbb{E}[f(X)]\leq\Phi_X(a+\varepsilon).$$

由 ε 的任意性，$\limsup_{n\to\infty}\Phi_{X_n}(a)\leq\Phi_X(a^+)$. 再取连续函数 f, 使得 $\chi_{(-\infty,a-\varepsilon]}\leq f\leq\chi_{(-\infty,a]}$. 同理可得，$\Phi_X(a^-)\leq\liminf_{n\to\infty}\Phi_{X_n}(a)$. 综上，

$$\Phi_X(a^-)\leq\liminf_{n\to\infty}\Phi_{X_n}(a)\leq\limsup_{n\to\infty}\Phi_{X_n}(a)\leq\Phi_X(a^+).$$

而分布函数 Φ_X 右连续，从而 $\Phi_X(a^+)=\Phi_X(a)$. 对任意 Φ_X 的连续点 a, 有 $\Phi_X(a^-)=\Phi_X(a)$. 此时，

$$\Phi_X(a)=\Phi_X(a^-)\leq\liminf_{n\to\infty}\Phi_{X_n}(a)\leq\limsup_{n\to\infty}\Phi_{X_n}(a)\leq\Phi_X(a^+)=\Phi_X(a).$$

故 $\lim_{n\to\infty}\Phi_{X_n}(a)$ 存在且等于 $\Phi_X(a)$. 即对任意 Φ_X 的连续点 a, 有 $\Phi_{X_n}(a)\to\Phi_X(a)$, 即 $X_n\xrightarrow{d}X$. □

推论 4.13. 若 $X_n\xrightarrow{d}X$, $g:\mathbb{R}\to\mathbb{R}$ 连续，则 $g(X_n)\xrightarrow{d}g(X)$.

证明. 设 f 是任意有界连续函数，则 $f\circ g$ 也是有界连续函数。而 $X_n\xrightarrow{d}X$, 从而 $\mathbb{E}[(f\circ g)(X_n)]\to\mathbb{E}[(f\circ g)(X)]$. 这等价于 $\mathbb{E}[f(g(X_n))]\to\mathbb{E}[f(g(X))]$. 即对任意有界连续函数 $f:\mathbb{R}\to\mathbb{R}$, 都有 $\mathbb{E}[f(g(X_n))]\to\mathbb{E}[f(g(X))]$. 即 $g(X_n)\xrightarrow{d}g(X)$. □

定义 4.14. 设 $X=(X_1,\cdots,X_n):\Omega\to\mathbb{R}^n$ 是随机向量，$f:\mathbb{R}^n\to\mathbb{R}$ 是 *Borel* 可测函数。$\mathbb{E}[f(X_1,\cdots,X_n)]=\int_{\mathbb{R}^n}f(X_1,\cdots,X_n)\,\mathrm{d}\,\mu_{X_1\cdots X_n}$. 令 $\sigma_{ij}(X_i,X_j)=\mathbb{E}[(X_i-\mathbb{E}[X_i])(X_j-\mathbb{E}[X_j])]$ 称为 X_i 与 X_j 的协方差，$\Sigma=[\sigma_{ij}]$ 称为 $X_1,\cdots,X_n$ 的协方差矩阵。

定理 4.15. 协方差矩阵半正定。

证明. $\forall (z_1, \cdots, z_n) \in \mathbb{R}^n$,

$$\sum_{i,j} \sigma_{ij} z_i z_j = \sum_{i,j} \mathbb{E}\left[z_i z_j \left(X_i - \mathbb{E}[X_i]\right)\left(X_j - \mathbb{E}[X_j]\right)\right] = \mathbb{E}\left[\left(\sum_{i=1}^n z_i \left(X_i - \mathbb{E}[X_i]\right)\right)^2\right] = \sigma^2\left(\sum_{i=1}^n z_i x_i\right) \geq 0.$$

□

定理 4.16. 设 $f: \mathbb{R}^k \to \mathbb{R}$ $g: \mathbb{R}^l \to \mathbb{R}$ 是 *Borel* 可测函数。$X_1, \cdots, X_k, Y_1, \cdots, Y_l$ 是随机变量，且 $(X_1, \cdots, X_k)$ 与 $(Y_1, \cdots, Y_l)$ 独立。则 $f(X_1, \cdots, X_k)$ 与 $g(Y_1, \cdots, Y_l)$ 独立，且 $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)]$.

证明. 由推论 2.14, $\sigma(X_1, \cdots, X_k)$ 与 $\sigma(Y_1, \cdots, Y_l)$ 独立。而 $\sigma(f(X)) \subset \sigma(X_1, \cdots, X_k)$, $\sigma(g(Y)) \subset \sigma(Y_1, \cdots, Y_l)$. 从而 $\sigma(f(X))$ 与 $\sigma(g(Y))$ 独立，更有 $f(X_1, \cdots, X_k)$ 与 $g(Y_1, \cdots, Y_l)$ 独立。

由 Fubini 定理，

$$\begin{aligned}
\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}k \times \mathbb{R}^l} f(x_1, \cdots, x_k) g(y_1, \cdots, y_l) \,\mathrm{d}\, \mu_{X_1 \cdots X_k Y_1 \cdots Y_l} = \int_{\mathbb{R}^k \times \mathbb{R}^l} f(x_1, \cdots, x_k) g(y_1, \cdots, y_l) \,\mathrm{d}\, \mu_{X_1 \cdots X_k} \cdot \mathrm{d}\, \mu_{Y_1 \cdots Y_l} \\
&= \left(\int_{\mathbb{R}^k} f(x_1, \cdots, x_k) \,\mathrm{d}\, \mu_{X_1 \cdots X_k}\right) \cdot \left(\int_{\mathbb{R}^l} g(y_1, \cdots, y_l) \,\mathrm{d}\,\mathrm{d}\, \mu_{Y_1 \cdots Y_l}\right) = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)].
\end{aligned}$$

□

推论 4.17. 若 $X_1, \cdots, X_n$ 是一组两两独立的随机变量，则 $\sigma^2\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sigma^2\left(X_i\right)$

证明.

$$\begin{aligned}
\sigma^2\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n \left(X_i - \mathbb{E}[X_i]\right)\right)^2\right] = \sum_{i,j} \mathbb{E}\left[\left(X_i - \mathbb{E}[X_i]\right)\left(X_j - \mathbb{E}[X_j]\right)\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\left(X_i - \mathbb{E}[X_i]\right)^2\right] + \sum_{i \neq j} \mathbb{E}\left[X_i - \mathbb{E}[X_i]\right] \cdot \mathbb{E}\left[X_j - \mathbb{E}[X_j]\right] = \sum_{i=1}^n \mathbb{E}\left[\left(X_i - \mathbb{E}[X_i]\right)^2\right] = \sum_{i=1}^n \sigma^2\left(X_i\right).
\end{aligned}$$

其中，在 $i \neq j$ 的求和式中, 由期望的线性,

$$\mathbb{E}\left[X_i - \mathbb{E}[X_i]\right] = \mathbb{E}\left[X_i\right] - \mathbb{E}\left[\mathbb{E}[X_i]\right] = \mathbb{E}\left[X_i\right] - \mathbb{E}\left[X_i\right] = 0.$$

故 $\sum_{i \neq j} \mathbb{E}\left[X_i - \mathbb{E}[X_i]\right] \cdot \mathbb{E}\left[X_j - \mathbb{E}[X_j]\right] = 0.$ □

定义 4.18.

- $\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\,zX}] = \int_\Omega \mathrm{e}^{\mathrm{i}\,zX}\,\mathrm{d}\,\mathbb{P}$ 称为 X 的特征函数/示性函数。
- $M_X(z) = \mathbb{E}[\mathrm{e}^{zX}] = \int_\Omega \mathrm{e}^{zX}\,\mathrm{d}\,\mathbb{P}$ 称为 X 的矩生成函数。
  若 X 是连续型随机变量，有密度函数 $f(x)$. 则 $\varphi_X(z) = \int_\mathbb{R} \mathrm{e}^{\mathrm{i}\,zX}\,f(x)\,\mathrm{d}\,x$, $M_X(z) = \int_\mathbb{R} \mathrm{e}^{zX}\,f(x)\,\mathrm{d}\,x$.
  若 $\varphi_X(z)$, $M_X(z)$ 在原点附近有定义，且可以展开成 z 的幂级数。有

$$\mathrm{e}^{zX} = \sum_{k=0}^{\infty} \frac{(zX)^k}{k!},\ \mathrm{e}^{\mathrm{i}\,zX} = \sum_{k=0}^{\infty} \frac{(\mathrm{i}\,zX)^k}{k!}.$$

从而

$$\varphi_X(z) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\,zX}] = \int_\Omega \mathrm{e}^{\mathrm{i}\,zX}\,\mathrm{d}\,\mathbb{P} = \int_\Omega \sum_{k=0}^{\infty} \frac{(\mathrm{i}\,zX)^k}{k!}\,\mathrm{d}\,\mathbb{P} = \sum_{k=0}^{\infty} \frac{(\mathrm{i}\,z)^k \int_\Omega X^k\,\mathrm{d}\,\mathbb{P}}{k!} = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]\cdot(\mathrm{i}\,z)^k}{k!}.$$

$$M_X(z) = \mathbb{E}[\mathrm{e}^{zX}] = \int_\Omega \mathrm{e}^{zX}\,\mathrm{d}\,\mathbb{P} = \int_\Omega \sum_{k=0}^{\infty} \frac{(zX)^k}{k!}\,\mathrm{d}\,\mathbb{P} = \sum_{k=0}^{\infty} \frac{z^k \int_\Omega X^k\,\mathrm{d}\,\mathbb{P}}{k!} = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]\cdot z^k}{k!}.$$

由于 $\left|\mathrm{e}^{\mathrm{i}\,zX}\right| = 1$, 故 $\varphi_X(z)$ 总存在。而 $M_X(z)$ 在除 $z=0$ 之外的点处可能不可积。

5 大数定律

动机：投硬币，正面概率为 p, 投 N 次，正面次数 X.

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次正,} \\ 0, & \text{第 } i \text{ 次反。} \end{cases}$$

则 $X = X_1 + \cdots + X_N$. 则 $N \to \infty$ 时，$\frac{X}{N} \xrightarrow{m} p$.

定理 5.1 (Markov 不等式)**.** $X \geq 0$ 是非负随机变量。则

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}[X].$$

证明.

$$\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\,\mathbb{P} \geq \int_{\{X \geq 1\}} X \,\mathrm{d}\,\mathbb{P} \geq \int_{\{X \geq 1\}} 1 \,\mathrm{d}\,\mathbb{P} = \mathbb{P}(X \geq 1).$$

□

推论 5.2. $X \geq 0$ 是非负随机变量，$\forall \varepsilon > 0$. 则

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{1}{\varepsilon} \cdot \mathbb{E}[X].$$

证明. 由期望的线性与 Markov 不等式（定理 5.1），

$$\mathbb{P}(X \geq \varepsilon) = \mathbb{P}(\frac{X}{\varepsilon} \geq 1) \leq \mathbb{E}[\frac{X}{\varepsilon}] = \frac{1}{\varepsilon} \cdot \mathbb{E}[X].$$

□

定理 5.3 (Chebyshev 不等式)**.** X 是随机变量，$\mathbb{E}[X]$ 存在。$\forall \varepsilon > 0$. 则

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \varepsilon\right) \leq \frac{\sigma^2(X)}{\varepsilon^2}.$$

证明. $(X-\mathbb{E}[X])^2$ 是非负随机变量，由 Markov 不等式（定理 5.1），

$$\mathbb{P}\left(|X-\mathbb{E}[X]| \geq \varepsilon\right) = \mathbb{P}\left((X-\mathbb{E}[X])^2 \geq \varepsilon^2\right) \leq \frac{1}{\varepsilon^2} \cdot \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] = \frac{\sigma^2(X)}{\varepsilon^2}.$$

□

定理 5.4 (L^2 弱大数定律)**.** 设 $X_1, X_2, \cdots$ 是一列两两独立的随机变量。$\forall i \geq 1$, $\mathbb{E}[X_i] = \mu$, 且 $\exists C > 0$, 使得 $\sigma^2(X_i) \leq C$. 令 $S_n = \frac{1}{n} \cdot (X_1 + \cdots + X_n)$. 则 $S_n \xrightarrow{m} \mu$.

证明. 由期望的线性，

$$\mathbb{E}[S_n] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu.$$

$\forall \varepsilon > 0$, 由 Chebyshev 不等式（定理 5.3），

$$\mathbb{P}\left(|S_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2(S_n)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \cdot \sigma^2\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{\varepsilon^2} \cdot \sum_{i=1}^{n} \sigma^2\left(\frac{1}{n} X_i\right) = \frac{1}{n^2 \cdot \varepsilon^2} \cdot \sum_{i=1}^{n} \sigma^2\left(X_i\right) \leq \frac{1}{n^2 \cdot \varepsilon^2} \cdot \sum_{i=1}^{n} C = \frac{1}{n^2 \cdot \varepsilon^2} \cdot nC = \frac{C}{n \cdot \varepsilon^2}.$$

从而 $\lim_{n\to\infty} \mathbb{P}\left(|S_n - \mu| \geq \varepsilon\right) = 0$. 即 $S_n \xrightarrow{m} \mu$. □

定理 5.5. $S_n \xrightarrow{L^2} \mu$. 即 $\lim_{n\to\infty} \mathbb{E}\left[|S_n - \mu|^2\right] = 0$.

证明. 作业。 □

应用 5.6 (Bernside 多项式逼近定理)**.** f 是 $[0,1]$ 上的实连续函数，$\forall n > 0$, 定义 *Bernside* 多项式：

$$B(f,n)(x) = \sum_{k=0}^{n} C_n^k \cdot x^k (1-x)^{n-k} \cdot f\left(\frac{k}{n}\right).$$

则在 $[0,1]$ 上，$B(f,n) \rightrightarrows f$.

证明. $\forall p \in [0,1]$, 定义 Bernoulli 随机变量 X_i:

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

令 $X_1, X_2, \cdots$ 互相独立。$\mathbb{E}[X_i] = p$. 由 L^2 弱大数定律（定理 5.4），$S_n = \frac{1}{n} \cdot \sum_{i=1}^n X_i \xrightarrow{m} p$. 由定理 4.9, $\mathbb{E}[f(S_n)] \to f(p)$. 而

$$\mathbb{E}[f(S_n)] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \cdot \mathbb{P}\left(S_n = \frac{k}{n}\right) = \sum_{k=0}^{n} C_n^k \cdot p^k (1-p)^{n-k} \cdot f\left(\frac{k}{n}\right) = B(f,n)(p).$$

即 $B(f,n) \xrightarrow{p.w.} f$.

再证一致收敛: 由于连续函数在紧区间上一致连续, 且能取到最值。故 f 在 $[0,1]$ 上一致连续, 有界。即 $\forall \varepsilon > 0, \exists \delta > 0$, 使得 $\forall |x-y| < \delta$, 都有 $|f(x) - f(y)| < \varepsilon$, 且设 $M = \sup_{x \in [0,1]} |f(x)|$. $\forall p \in [0,1]$, 由 Jensen 不等式，$|\mathbb{E}[f(S_n)] - f(p)| \leq \mathbb{E}[|f(S_n) - f(p)|]$. 从而，

$$\begin{aligned}
|B(n,f)(p) - f(p)| = |\mathbb{E}[f(S_n)] - f(p)| \leq \mathbb{E}[|f(S_n) - f(p)|] &= \int_\Omega |f(S_n) - f(p)| \,\mathrm{d}\,\mathbb{P} \\
&= \int_{\{x \mid |S_n - p| < \delta\}} |f(S_n) - f(p)| \,\mathrm{d}\,\mathbb{P} + \int_{\{x \mid |S_n - p| \geq \delta\}} |f(S_n) - f(p)| \,\mathrm{d}\,\mathbb{P} \\
&\leq \int_{\{x \mid |S_n - p| < \delta\}} \varepsilon \,\mathrm{d}\,\mathbb{P} + \int_{\{x \mid |S_n - p| \geq \delta\}} 2M \,\mathrm{d}\,\mathbb{P} \\
&= \varepsilon \cdot \mathbb{P}(|S_n - p| < \delta) + 2M \cdot \mathbb{P}(|S_n - p| \geq \delta) \\
&\leq \varepsilon \cdot 1 + 2M \cdot \varepsilon \\
&= (1 + 2M)\varepsilon.
\end{aligned}$$

这与 p 无关。故在 $[0,1]$ 上，$B(f,n) \rightrightarrows f$. □

定理 5.7 (一般弱大数定律). 设 $X_1, X_2, \cdots$ 是独立同分布的随机变量。若 $\lim_{n\to\infty} n \cdot \mathbb{P}(|X_1| \geq n) = 0$. 令 $\mu_n = \mathbb{E}\left[X_1 \cdot \chi_{\{|X_1| \leq n\}}\right]$. 则 $S_n - \mu_n \xrightarrow{m} 0$.

证明. 记 $S_n^* = \frac{1}{n} \cdot \sum_{i=1}^n X_i \cdots \chi_{\{|X_i| \leq n\}}$. 从而

$$\mathbb{E}[S_n^*] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}\left[X_i \cdots \chi_{\{|X_i| \leq n\}}\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu_n = \mu_n.$$

$\forall \varepsilon > 0$, 由于

$$\{|S_n - \mu_n| \geq \varepsilon\} \subset \{S_n \neq S_n^*\} \cup \{|S_n^* - \mu_n| \geq \varepsilon\}.$$

从而

$$\mathbb{P}(|S_n - \mu_n| \geq \varepsilon) \leq \mathbb{P}(S_n \neq S_n^*) + \mathbb{P}(|S_n^* - \mu_n| \geq \varepsilon).$$

其中,

$$\mathbb{P}(S_n \neq S_n^*) = \mathbb{P}\left(\bigcup_{i=1}^n \left\{X_i \neq X_1 \cdot \chi_{|X_i| \leq n}\right\}\right) \leq \sum_{i=1}^n \mathbb{P}\left(X_i \neq X_1 \cdot \chi_{|X_i|}\right) = \sum_{i=1}^n \mathbb{P}(|X_i| > n) = n \cdot \mathbb{P}(|X_1| > n) \to 0.$$

由 Chebyshev 不等式（定理 5.3),

$$\begin{aligned}
\mathbb{P}(|S_n^* - \mu_n| \geq \varepsilon) \leq \frac{\sigma^2(S_n^*)}{\varepsilon^2} &= \frac{1}{\varepsilon^2} \cdot \sigma^2\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i \cdot \chi_{\{|X_i| \leq n\}}\right) \\
&= \frac{1}{n^2 \cdot \varepsilon^2} \cdot \sum_{i=1}^n \sigma^2\left(X_i \cdot \chi_{\{|X_i| \leq n\}}\right) \\
&= \frac{1}{n^2 \cdot \varepsilon^2} \cdot \sum_{i=1}^n \sigma^2\left(X_1 \cdot \chi_{\{|X_1| \leq n\}}\right) = \frac{n \cdot \sigma^2\left(X_1 \cdot \chi_{\{|X_1| \leq n\}}\right)}{n^2 \cdot \varepsilon^2} = \frac{\sigma^2\left(X_1 \cdot \chi_{\{|X_1| \leq n\}}\right)}{n \cdot \varepsilon^2}.
\end{aligned}$$

从而只需要证 $\frac{\sigma^2\left(X_1 \cdot \chi_{\{|X_1| \leq n\}}\right)}{n} \to 0$. 注意到 $\sigma^2(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$. 故只需要证 $\frac{\mathbb{E}\left[X_1^2 \cdot \chi_{\{|X_1| \leq n\}}\right]}{n} \to 0$.

引理 5.8. 设 $X \geq 0$ 是一个非负随机变量, $k \geq 0$. 则

$$\mathbb{E}\left[X^k\right] = \int_0^\infty k \cdot y^{k-1} \cdot \mathbb{P}(X \geq y) \,\mathrm{d}\, y.$$

证明. 由 Fubini 定理,

$$\begin{aligned}
\int_0^\infty k \cdot y^{k-1} \cdot \mathbb{P}(X \geq y) \,\mathrm{d}\, y = \int_0^\infty k \cdot y^{k-1} \cdot \left(\int_\Omega \chi_{\{X \geq y\}} \,\mathrm{d}\,\mathbb{P}\right) \mathrm{d}\, y &= \int_0^\infty \int_\Omega k \cdot y^{k-1} \cdot \chi_{\{X \geq y\}} \,\mathrm{d}\,\mathbb{P} \,\mathrm{d}\, y \\
&= \int_\Omega \left(\int_0^\infty k \cdot y^{k-1} \cdot \chi_{\{X \geq y\}} \,\mathrm{d}\, y\right) \mathrm{d}\,\mathbb{P} = \int_\Omega \left(\int_0^X k \cdot y^{k-1} \,\mathrm{d}\, y\right) \mathrm{d}\,\mathbb{P} = \int_\Omega X^k \,\mathrm{d}\,\mathbb{P} = \mathbb{E}\left[X^k\right].
\end{aligned}$$

□

从而由引理 5.8,

$$\mathbb{E}\left[X_1^2\cdot\chi_{\{|X_1|\leq n\}}\right]=\int_0^\infty 2y\cdot\mathbb{P}\left(|X_1|\cdot\chi_{\{|X_1\leq n|\}}\geq y\right)\mathrm{d}\,y.$$

而其中,

$$\left\{x\mid |X_1|\cdot\chi_{\{|X_1\leq n|\}}\geq y\right\}=\{x\mid y\leq|X_1|\leq n\}.$$

故

$$\begin{aligned}\mathbb{E}\left[X_1^2\cdot\chi_{\{|X_1|\leq n\}}\right]&=\int_0^\infty 2y\cdot\mathbb{P}\left(|X_1|\cdot\chi_{\{|X_1\leq n|\}}\geq y\right)\mathrm{d}\,y\\&=\int_0^\infty 2y\cdot\mathbb{P}\left(y\leq|X_1|\leq n\right)\mathrm{d}\,y\\&=\int_0^\infty 2y\left(\mathbb{P}\left(|X_1|\geq y\right)-\mathbb{P}\left(|X_1|>n\right)\right)\mathrm{d}\,y=\int_0^n 2y\cdot\mathbb{P}\left(|X_1|\geq y\right)\mathrm{d}\,y-\int_0^n 2y\cdot\mathbb{P}\left(|X_1|>n\right)\mathrm{d}\,y\leq\int_0^n 2y\cdot\mathbb{P}\left(|X_1|\geq y\right)\mathrm{d}\,y.\end{aligned}$$

其中，记 $g(y)=2y\cdot\mathbb{P}\left(|X_1|\geq y\right)$. 由于 $\lim_{n\to\infty}n\cdot\mathbb{P}\left(|X_1|\geq n\right)=0$. 故 $g(y)$ 有界，记为 M. $\forall n$, 令 $g_n(y)=g(ny)$. 从而 $g_n(y)=g(ny)=2ny\cdot\mathbb{P}\left(|X_1|\geq ny\right)\to 0$. 即 $g_n\xrightarrow{p.w.}0$. 由变量代换与控制收敛定理,

$$\lim_{n\to\infty}\frac{1}{n}\cdot\int_0^n g(y)\,\mathrm{d}\,y=\lim_{n\to\infty}\int_0^1 g_n(y)\,\mathrm{d}\,y=\int_0^1\lim_{n\to\infty}g_n(y)\,\mathrm{d}\,y=\int_0^1 0\,\mathrm{d}\,y=0.$$

从而

$$\frac{1}{n}\cdot\mathbb{E}\left[X_1^2\cdot\chi_{\{|X_1|\leq n\}}\right]\leq\frac{1}{n}\cdot\int_0^n 2y\cdot\mathbb{P}\left(|X_1|\geq y\right)\mathrm{d}\,y=\frac{1}{n}\cdot\int_0^n g(y)\,\mathrm{d}\,y\to 0.$$

综上，$\forall\varepsilon>0$,

$$\mathbb{P}\left(|S_n-\mu_n|\geq\varepsilon\right)\leq\mathbb{P}\left(S_n\neq S_n^*\right)+\mathbb{P}\left(|S_n^*-\mu_n|\geq\varepsilon\right)\leq\mathbb{P}\left(S_n\neq S_n^*\right)+\frac{\sigma^2\left(X_1\cdot\chi_{\{|X_1|\leq n\}}\right)}{n\cdot\varepsilon^2}\leq\mathbb{P}\left(S_n\neq S_n^*\right)+\frac{\mathbb{E}\left[X_1^2\cdot\chi_{\{|X_1|\leq n\}}\right]}{n\cdot\varepsilon^2}\to 0+0\cdot\frac{1}{\varepsilon^2}=0.$$

即 $S_n-\mu_n\xrightarrow{m}0$. □

注记 5.9. μ_n 一定存在。$\mathbb{E}\left[|X_1|\right]$ 未必存在。

推论 5.10 (L^1 弱大数定律)**.** 设 $X_1, X_2, \cdots$ 是独立同分布的随机变量。$\mu = \mathbb{E}\left[|X_1|\right] < \infty$ 存在。则 $S_n \xrightarrow{m} \mu$.

命题 5.11. X 是连续型随机变量，密度函数

$$f_X(x) = \begin{cases} \frac{1}{x^2 \log|x|}, & |x| \geq 2, \\ 0, & |x| < 2. \end{cases}$$

则

- $\mathbb{E}\left[|X|\right] = +\infty$.
- $n \cdot \mathbb{P}\left(|X| \leq n\right) \to 0$.

证明. 作业。 □

定理 5.12 (L^1 强大数定律)**.** 设 $X_1, X_2, \cdots$ 是一列独立同分布的随机变量，$\mu = \mathbb{E}\left[|X_i|\right] < +\infty$. 则 $S_n = \frac{1}{n} \cdot (X_1 + \cdots + X_n) \xrightarrow{a.e.} \mu$.

先看一个特殊情形：

定理 5.13 (L^4 强大数定律)**.** 设 $X_1, X_2, \cdots$ 是一列独立同分布的随机变量，$\mu = \mathbb{E}\left[|X_i|\right] < +\infty$, $\mathbb{E}\left[(X_i - \mu)^4\right]$ 有界。则 $S_n \xrightarrow{a.e.} \mu$.

证明. 不妨设 $\mu = 0$, 否则用 $X_i - \mu$ 代替 X_i. 设 $\mathbb{E}\left[X_i^4\right] \leq C$.

由均值不等式，$\forall i \neq j$,

$$\mathbb{E}\left[X_i^2 X_j^2\right] \leq \mathbb{E}\left[\frac{1}{2} \cdot \left(X_i^4 + X_j^4\right)\right] = \frac{1}{2} \cdot \left(\mathbb{E}\left[X_i^4\right] + \mathbb{E}\left[X_j^4\right]\right) \leq C.$$

考虑

$$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] = \sum_{i,j,k,l=1}^n \mathbb{E}\left[X_i X_j X_k X_l\right].$$

由于 $X_1, X_2, \cdots$ 相互独立,

$$\mathbb{E}\left[X_iX_jX_kX_l\right] = \mathbb{E}\left[X_i\right]\cdot\mathbb{E}\left[X_j\right]\cdot\mathbb{E}\left[X_k\right]\cdot\mathbb{E}\left[X_l\right] = 0.$$

$$\mathbb{E}\left[X_i^2X_jX_k\right] = \mathbb{E}\left[X_i^2\right]\cdot\mathbb{E}\left[X_j\right]\cdot\mathbb{E}\left[X_k\right] = 0.$$

$$\mathbb{E}\left[X_i^3X_j\right] = \mathbb{E}\left[X_i^3\right]\cdot\mathbb{E}\left[X_j\right] = 0.$$

从而,

$$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] = \sum_{i,j,k,l=1}^n \mathbb{E}\left[X_iX_jX_kX_l\right] = \sum_{i=1}^n \mathbb{E}\left[X_i^4\right] + C_4^2 \sum_{1\le i<j\le n} \mathbb{E}\left[X_i^2X_j^2\right] \le n\cdot C + C_4^2\cdot C_n^2\cdot C \le 3n^2C.$$

由 Chebyshev 不等式(定理 5.3),

$$\mathbb{P}\left(|S_n|\ge\varepsilon\right) = \mathbb{P}\left(|S_n|^4\ge\varepsilon^4\right) \le \frac{\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)\right]^4}{n^4\cdot\varepsilon^4} \le \frac{3n^2C}{n^4\cdot\varepsilon^4} = \frac{3C}{n^2\cdot\varepsilon^4}.$$

由 Borel-Cantelli 定理(定理 2.16),由于

$$\sum_{n=1}^\infty \mathbb{P}\left(|S_n|\ge\varepsilon\right) \le \frac{3C}{\varepsilon^4}\cdot\sum_{i=1}^n\frac{1}{n^2} < \infty.$$

从而 $\forall\varepsilon>0$,

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{|S_n|\ge\varepsilon\right\}\right) = 0.$$

这等价于 $S_n \xrightarrow{a.e.} 0$. □

L^1 强大数定律的证明. 由于 $X_i = X_i^+ - X_i^-$, 故只需考虑 $X_i \ge 0$ 的情况。

做截断:

$$Y_n = X_n\cdot\chi_{\{|X_n|\le n\}} = \begin{cases} X_n, & |X_n|\le n,\\ 0, & \text{其它}.\end{cases}$$

记 $T_n = \frac{1}{n}\cdot(Y_1+\cdots+Y_n)$, $\mu_n = \mathbb{E}\left[Y_n\right]$.

引理 5.14. 若 $Y_n - \mu_n \xrightarrow{a.e.} 0$. 则 $S_n \xrightarrow{a.e.} \mu$.

证明. $\widehat{Y_n} = X_1 \cdot \chi_{\{|X_1| \le n\}}$ 与 Y_n 同分布，从而 $\mu_n = \mathbb{E}[Y_n] = \mathbb{E}\left[\widehat{Y_n}\right] \to \mu$. 又 $\widehat{Y_n} \nearrow X_1$, 由单调收敛定理，$\mathbb{E}\left[\widehat{Y_n}\right] \to \mathbb{E}[X_1] = \mu$. 从而

$$\mu_n = \mathbb{E}[T_n] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}[Y_n] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}\left[\widehat{Y_n}\right] \to \mu.$$

又

$$\mathbb{E}[X_1] = \int_{\mathbb{R}} x \,\mathrm{d}\,\mu_{X_1} \ge \int_{\mathbb{R}} \lfloor x \rfloor \,\mathrm{d}\,\mu_{X_1} = \sum_{i=1}^{\infty} i \cdot \mathbb{P}(i \le X_1 < i+1) = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \mathbb{P}(i \le X_1 < i+1) = \sum_{j=1}^{\infty} \mathbb{P}(X_1 \ge j).$$

从而

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \ne Y_i) = \sum_{i=1}^{\infty} \mathbb{P}(X_i > i) = \sum_{i=1}^{\infty} \mathbb{P}(X_1 > i) \le \sum_{i=1}^{\infty} \mathbb{P}(X_1 \ge i) \le \mathbb{E}[X_1] < \infty.$$

由 Borel-Cantelli 定理（定理 2.16），

$$\mathbb{P}\left(\limsup_{n\to\infty} \{X_n \ne Y_n\}\right) = 0.$$

这说明对几乎所有 ω, $\exists n(\omega)$, 使得 $n \ge n(\omega)$ 时，有 $X_n(\omega) = Y_n(\omega)$. 故此时，

$$S_n(\omega) - T_n(\omega) = \frac{1}{n} \cdot \sum_{n=1}^{n(\omega)} (X_n(\omega) - Y_n(\omega)) \to 0.$$

即 $S_n \xrightarrow{a.e.} T_n$.

综上，$S_n \xrightarrow{a.e.} T_n$, $T_n \xrightarrow{a.e.} \mu$. 故 $S_n \xrightarrow{a.e.} \mu$. □

$\forall \varepsilon > 0$, 取 $n_k = \left\lfloor (1+\varepsilon)^k \right\rfloor \ge \frac{1}{2}(1+\varepsilon)^k$.（不等式是因为，对 $x > 1$, 有 $\lfloor x \rfloor \ge \frac{x}{2}$.）从而 $\forall \delta > 0$, 由 Chebyshev 不等式（定理 5.3），

$$\sum_{i=1}^{\infty} \mathbb{P}(|T_{n_i} - \mu_{n_i}| \ge \delta) = \sum_{i=1}^{\infty} \mathbb{P}\left(|T_{n_i} - \mu_{n_i}|^2 \ge \delta^2\right) \le \frac{1}{\delta^2} \cdot \sum_{i=1}^{\infty} \sigma^2(T_{n_i}) = \frac{1}{\delta^2} \cdot \sum_{i=1}^{\infty} \left(\frac{1}{n_i^2} \cdot \sum_{k=1}^{n_i} \sigma^2(Y_k)\right) = \frac{1}{\delta^2} \cdot \left(\sum_{k=1}^{\infty} \sigma^2(Y_k)\right) \cdot \left(\sum_{i=n_i \ge k}^{n_i} \frac{1}{n_i^2}\right).$$

其中,

$$\sum_{i=n_i\geq k}^{n_i}\frac{1}{n_i^2}=\sum_{(1+\varepsilon)^i\geq k}\frac{1}{\lfloor(1+\varepsilon)^i\rfloor^2}\leq 4\cdot\sum_{(1+\varepsilon)^{2i}\geq k}\frac{1}{(1+\varepsilon)^{2i}}\leq\frac{4}{k^2}\cdot\frac{1}{1-(1+\varepsilon)^2}.$$

从而

$$\sum_{i=1}^{\infty}\mathbb{P}\left(|T_{n_i}-\mu_{n_i}|\geq\delta\right)\leq\frac{1}{\delta^2}\cdot\left(\sum_{k=1}^{\infty}\sigma^2(Y_k)\right)\cdot\left(\sum_{i=n_i\geq k}^{n_i}\frac{1}{n_i^2}\right)\leq\frac{1}{\delta^2}\cdot\frac{4}{1-(1+\varepsilon)^2}\cdot\sum_{k=1}^{\infty}\frac{\sigma^2(Y_k)}{k^2}.$$

引理 5.15. $\sum_{k=1}^{\infty}\frac{\sigma^2(Y_k)}{k^2}<\infty$.

证明. 由于 $\{Y_k\geq y\}=\{y\leq x_k\leq k\}$, 从而

$$\sigma^2(Y_k)=\mathbb{E}\left[Y_k^2\right]-\mathbb{E}\left[Y_k\right]^2\leq\mathbb{E}\left[Y_k^2\right]=\int_0^{+\infty}2y\cdot\mathbb{P}\left(Y_k\geq y\right)\mathrm{d}\,y=\int_0^{k}2y\cdot\mathbb{P}\left(Y_k\geq y\right)\mathrm{d}\,y\leq\int_0^{k}2y\cdot\mathbb{P}\left(X_k\geq y\right)\mathrm{d}\,y=\int_0^{k}2y\cdot\mathbb{P}\left(X_1\geq y\right)\mathrm{d}\,y.$$

故

$$\sum_{k=1}^{\infty}\frac{\sigma^2(Y_k)}{k^2}\leq\sum_{i=1}^{\infty}\frac{1}{k^2}\int_0^{k}2y\cdot\mathbb{P}\left(X_1\geq y\right)\mathrm{d}\,y=\int_0^{\infty}\left(\sum_{k=1}^{\infty}\chi_{\{y\leq k\}}\frac{1}{k^2}\cdot 2y\cdot\mathbb{P}\left(X_1\geq y\right)\right)\mathrm{d}\,y.$$

记 $f(y)=\sum_{k=1}^{\infty}\chi_{\{y\leq k\}}\frac{1}{k^2}\cdot 2y$.

- 若 $y\geq 2$, 设 $m-1\leq y\leq m$, 则

$$f(y)=\sum_{k=1}^{\infty}\chi_{\{y\leq k\}}\frac{1}{k^2}\cdot 2y\leq 2y\sum_{k=m}^{\infty}\frac{1}{k(k-1)}=2y\sum_{k=m}^{\infty}\left(\frac{1}{k}-\frac{1}{k+1}\right)=\frac{2y}{m-1}\leq\frac{2m}{m-1}\leq 4.$$

- 若 $t\leq 2$, 则

$$f(y)=\sum_{k=1}^{\infty}\chi_{\{y\leq k\}}\frac{1}{k^2}\cdot 2y\leq 4\cdot\sum_{k=1}^{\infty}\frac{1}{k^2}=4\cdot\frac{\pi^2}{6}<7.$$

总之，$f(y) < 7$. 从而

$$\sum_{k=1}^{\infty} \frac{\sigma^2(Y_k)}{k^2} \leq \int_0^{\infty} \left(\sum_{k=1}^{\infty} \chi_{\{y \leq k\}} \frac{1}{k^2} \cdot 2y \cdot \mathbb{P}(X_1 \geq y) \right) \mathrm{d}\, y = \int_0^{\infty} f(y) \cdot \mathbb{P}(X_1 \geq y) \,\mathrm{d}\, y \leq 7 \cdot \mathbb{P}(X_1 \geq y) \,\mathrm{d}\, y = 7 \cdot \mathbb{E}[X_1] < \infty.$$

□

从而

$$\sum_{i=1}^{\infty} \mathbb{P}(|T_{n_i} - \mu_{n_i}| \geq \delta) \leq \frac{1}{\delta^2} \cdot \frac{4}{1-(1+\varepsilon)^2} \cdot \sum_{k=1}^{\infty} \frac{\sigma^2(Y_k)}{k^2} < \infty.$$

从而由 Borel-Cantelli 定理（定理 2.16），$\forall \delta > 0$,

$$\mathbb{P}\left(\limsup_{n\to\infty} \{|T_{n_i} - \mu_{n_i}| \geq \delta\} \right) = 0.$$

即 $T_{n_i} - \mu_{n_i} \xrightarrow{a.e.} 0$.

对 $n_k < n \leq n_{k+1}$,

$$\frac{n_k}{n_{k+1}} T_{n_k} = \frac{1}{n_{k+1}} \cdot \sum_{i=1}^{n_k} Y_i < \frac{1}{n} \sum_{i=1}^{n} Y_i = T_n < \frac{n_{k+1}}{n_k} T_{n_{k+1}}.$$

其中，$\frac{n_k}{n_{k+1}} T_{n_k} \xrightarrow{a.e.} \frac{1}{1+\varepsilon} \mu_{n_k}$, $\frac{n_{k+1}}{n_k} T_{n_{k+1}} \xrightarrow{a.e.} (1+\varepsilon)\mu_{n_{k+1}}$. 又 $\mu_n \xrightarrow{a.e.} \mu$. 于是 $\frac{n_k}{n_{k+1}} T_{n_k} \xrightarrow{a.e.} \frac{1}{1+\varepsilon}\mu$, $\frac{n_{k+1}}{n_k} T_{n_{k+1}} \xrightarrow{a.e.} (1+\varepsilon)\mu_n$. 从而几乎处处都有

$$\frac{1}{1+\varepsilon}\mu \leq \liminf_{n\to\infty} T_n \leq \limsup_{n\to\infty} T_n \leq (1+\varepsilon)\mu$$

令 $\varepsilon \to 0^+$. 即得 $T_n \xrightarrow{a.e.} \mu$. 又 $\mu_n \xrightarrow{a.e.} \mu$, 从而 $T_n - \mu_n \xrightarrow{a.e.} 0$. 由引理 5.14, $S_n \xrightarrow{a.e.} \mu$. □

应用 5.16 (Monte Carlo 积分)**.** 设 f 是 $[0,1]$ 上的连续函数。想要计算 $\int_0^1 f(x) \,\mathrm{d}\, x$.

设 $X_1, X_2, \cdots$ 是独立同分布的随机变量，服从 $[0,1]$ 上的均匀分布。则 $f(X_1), f(X_2), \cdots$ 也是独立同分布的随机变量。

$$\mathbb{E}[f(X_1)] = \int_{\mathbb{R}} f(x) \,\mathrm{d}\, \mu_X = \int_0^1 f(x) \,\mathrm{d}\, x.$$

由大数定律，$S_n = \frac{1}{n} \cdot \sum_{i=1}^{n} f(X_i) \xrightarrow{a.e.} \int_0^1 f(x)\,\mathrm{d}\,x = I$. 并且我们有误差估计：由 *Chebyshev* 不等式（定理 *5.3*），

$$\mathbb{P}\left(|S_n - I| \geq \varepsilon\right) \leq \frac{\sigma^2\left(f(X_1)\right)}{n \cdot \varepsilon^2}.$$

其中，$\sigma^2\left(f(X_1)\right)$ 是一个有界的常数。故可以让计算机重复模拟一定的次数，以得到想要的精度。

事实上，*Monte Carlo* 积分要更简单：设 $f : [0,1] \to [0,1]$ 连续。记 $A = \{(x,y) \in [0,1] \mid y \leq f(x)\}$, $\Omega = [0,1] \times [0,1]$. 则 $\mathbb{E}\left[\chi_A\right] = I$. 若 X, Y 是 $[0,1]$ 上的均值分布，且独立，则 (X,Y) 是 Ω 上的均匀分布。设 $(X_1,Y_1),(X_2,Y_2),\cdots$ 是 Ω 上独立同分布的均匀随机变量，则

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \chi_A(X_i, Y_i) \xrightarrow{a.e.} I.$$

并且，也可以做类似的误差估计。

6 随机变量级数的收敛性

设 $X_1, X_2, \cdots$ 是互相独立的随机变量，$\sum_{i=1}^{\infty} X_i$ 是否几乎处处收敛?

令 $A = \{\omega \in \Omega | \sum_{i=1}^{\infty} X_i(\omega)$ 收敛$\}$, 则 A 是一个尾事件。由 Kolmogorov 0-1 律（定理 2.20），$\mathbb{P}(A) \in \{0, 1\}$.

定理 6.1 (Kolmogorov 不等式)**.** 设 $X_1, X_2, \cdots, X_n$ 是两两独立的随机变量，$\mathbb{E}[X_i] = 0$, $1 \le i \le n$. 记 $S_n = X_1 + \cdots + X_n$. 则 $\forall \varepsilon > 0$,

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}\sigma^2(S_n).$$

注记 6.2. 注意到 $\{\max_{1\le k\le n} |S_k| \ge \varepsilon\} \supset \{|S_n| \ge \varepsilon\}$, 故 *Kolmogorov* 不等式比 *Chebyshev* 不等式：$\mathbb{P}(|S_n| \ge \varepsilon) \le \frac{1}{\varepsilon^2}\sigma^2(S_n)$ 强。

证明. 记 $A = \{\max_{1\le k\le n} |S_k| \ge \varepsilon\}$, $A_k = \{|S_k| \ge \varepsilon, \max_{1\le i\le k} |S_i| \le \varepsilon\}$. 于是 $A = A_1 \sqcup \cdots \sqcup A_n$, 且 $A_k \in \sigma(X_1, \cdots, X_k)$, $S_n - S_k = X_{k+1} + \cdots + X_n$ 是 $\sigma(X_{k+1}, \cdots, X_n)$ 中的可测集，$\sigma(X_1, \cdots, X_k)$ 与 $\sigma(X_{k+1}, \cdots, X_n)$ 独立。

$$\begin{aligned}
\sigma^2(S_n) = \mathbb{E}\left[S_n^2\right] - \mathbb{E}[S_n]^2 = \mathbb{E}\left[S_n^2\right] - \left(\sum_{i=1}^{n} \mathbb{E}[X_i]\right)^2 = \mathbb{E}\left[S_n^2\right] \ge \mathbb{E}\left[S_n^2 \cdot \chi_A\right] &= \mathbb{E}\left[S_n \cdot \sum_{k=1}^{n} \chi_{A_k}\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[S_n^2 \cdot \chi_{A_k}\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\left(S_k^2 + 2(S_n - S_k) S_k + (S_n - S_k)^2\right)\chi_{A_k}\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[S_k^2 \cdot \chi_{A_k}\right] + \sum_{k=1}^{n} \mathbb{E}\left[2(S_n - S_k) S_k \cdot \chi_{A_k}\right] + \sum_{k=1}^{n} \mathbb{E}\left[(S_n - S_k)^2 \cdot \chi_{A_k}\right].
\end{aligned}$$

其中，由 $S_n - S_k$ 与 $S_k \cdot \chi_{A_k}$ 的独立性，从而 $\mathbb{E}[2(S_n - S_k) S_k \cdot \chi_{A_k}] = 0$, $\mathbb{E}\left[(S_n - S_k)^2 \cdot \chi_{A_k}\right] = 0$. 从而，由 Chebyshev 不等式（定理 5.3），

$$\mathbb{P}(A_n) \le \frac{\sigma^2(A_n)}{\varepsilon^2} = \frac{\mathbb{E}[S_n^2] - \mathbb{E}[S_n]^2}{\varepsilon^2} \le \frac{\mathbb{E}[S_n^2]}{\varepsilon^2} = \frac{\mathbb{E}[S_n^2 \cdot \chi_{A_n}]}{\varepsilon^2}.$$

从而 $\mathbb{E}[S_n^2\cdot\chi_{A_n}]\geq\varepsilon^2\cdot\mathbb{P}(A_n)$. 从而，由于 $A=A_1\sqcup\cdots\sqcup A_n$,

$$\sigma^2(S_n)\geq\sum_{k=1}^n\mathbb{E}\left[S_k^2\cdot\chi_{A_k}\right]+\sum_{k=1}^n\mathbb{E}\left[2\left(S_n-S_k\right)S_k\cdot\chi_{A_k}\right]+\sum_{k=1}^n\mathbb{E}\left[\left(S_n-S_k\right)^2\cdot\chi_{A_k}\right]=\sum_{k=1}^n\mathbb{E}\left[S_k^2\cdot\chi_{A_k}\right]\geq\sum_{k=1}^n\varepsilon^2\cdot\mathbb{P}\left(A_k\right)=\varepsilon^2\cdot\mathbb{P}\left(A\right).$$

即

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k|\geq\varepsilon\right)=\mathbb{P}\left(A\right)\leq\frac{1}{\varepsilon^2}\sigma^2(S_n).$$

□

定理 6.3. 设 $X_1,X_2,\cdots$ 是互相独立的随机变量，$\mathbb{E}[X_i]=0$, $\forall i$. 记 $S_n=\sum_{i=1}^n X_i$. 则 $\forall\varepsilon>0$,

$$\frac{|S_n|}{n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}}\xrightarrow{a.e.}0.$$

证明. 只要证 $\forall\delta>0$,

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\frac{|S_n|}{n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}}\geq\delta\right\}\right)=0.$$

记 $T_n=\max_{1\leq k\leq n}|S_k|$, $l_n=n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}$.

先证 $\frac{T_{2^n}}{l_{2^n}}\xrightarrow{a.e.}0$: $\forall\delta>0$, 由 Kolmogorov 不等式（定理 6.1），

$$\mathbb{P}\left(\frac{T_{2^n}}{l_{2^n}}\geq\delta\right)=\mathbb{P}\left(\max_{1\leq k\leq 2^n}|S_k|\geq\delta\cdot l_{2^n}\right)\leq\frac{1}{\delta^2\cdot {l_{2^n}}^2}\sigma^2\left(S_{2^n}\right)=\frac{1}{\delta^2\cdot {l_{2^n}}^2}\sum_{i=1}^{2^n}\sigma^2(X_i)=\frac{2^nC}{\delta^2\cdot 2^n\cdot n^{1+2\varepsilon}}=\frac{C}{\delta^2\cdot n^{1+2\varepsilon}}.$$

从而，

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{T_{2^n}}{l_{2^n}}\geq\delta\right)\leq\sum_{n=1}^{\infty}\frac{C}{\delta^2\cdot n^{1+2\varepsilon}}<\infty.$$

由 Borel-Cantelli 定理（定理 2.16），

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\frac{T_{2^n}}{l_{2^n}}\geq\delta\right\}\right)=0.$$

即 $\frac{T_{2^n}}{l_{2^n}} \xrightarrow{a.e.} 0$. 又 $T_{2^n} \geq |S_{2^n}|$, 从而 $\frac{|S_{2^n}|}{l_{2^n}} \xrightarrow{a.e.} 0$.

对 $2^n < m < 2^{n+1}$,

$$\frac{|S_m|}{l_m} \leq \frac{T_{2^{n+1}}}{l_{2^n}} = \frac{T_{2^{n+1}}}{l_{2^{n+1}}} \cdot \frac{l_{2^{n+1}}}{l_{2^n}} = \frac{T_{2^{n+1}}}{l_{2^{n+1}}} \cdot \frac{2^{\frac{1}{2}(n+1)} \cdot [(n+1)\log 2]^{1+\varepsilon}}{2^{\frac{1}{2}n} \cdot [n \log 2]^{1+\varepsilon}} = \frac{T_{2^{n+1}}}{l_{2^{n+1}}} \cdot 2^{\frac{1}{2}} \left[\frac{n+1}{n}\right]^{1+\varepsilon} \xrightarrow{a.e.} 0.$$

综上，$\frac{|S_n|}{n} \xrightarrow{a.e.} 0$. □

定理 6.4 (Kolmogorov 收敛准则)**.** 设 $X_1, X_2, \cdots$ 是互相独立的随机变量，$\sum_{n=1}^{\infty} \sigma^2(X_n) < \infty$. 则 $S_n = \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k])$ 几乎处处收敛。

注记 6.5. 大部分时候，我们应用 $\mathbb{E}[X_k] = 0$ 的情况。

证明. 不妨设 $\mathbb{E}[X_k] = 0$. 由 Kolmogorov 不等式（6.1），

$$\mathbb{P}\left(\max_{N \leq n \leq M} |S_n - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sigma^2(S_M - S_N) = \frac{1}{\varepsilon^2} \sum_{k=N+1}^{M} \sigma^2(X_k).$$

其中，集合 $\{\max_{N \leq n \leq M} |S_n - S_N| \geq \varepsilon\}$ 随着 M 的增加而递增。令 $M \to \infty$,

$$\mathbb{P}\left(\max_{n \geq N} |S_n - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=N+1}^{\infty} \sigma^2(X_k) < \infty.$$

再令 $N \to \infty$. 由三角不等式，$\{\max_{m,n \geq N} |S_n - S_m| \geq 2\varepsilon\} \subset \{\max_{n \geq N} |S_n - S_N| \geq \varepsilon\}$, 从而，

$$\lim_{N \to \infty} \mathbb{P}\left(\max_{m,n \geq N} |S_n - S_m| \geq 2\varepsilon\right) \leq \lim_{N \to \infty} \mathbb{P}\left(\max_{n \geq N} |S_n - S_N| \geq \varepsilon\right) \leq \lim_{N \to \infty} \frac{1}{\varepsilon^2} \sum_{k=N+1}^{\infty} \sigma^2(X_k) = 0.$$

这说明集合 $\{\max_{m,n \geq N} |S_n - S_m| \geq 2\varepsilon\}$ 随着 $N \to \infty$ 单调递减收敛于一个零测集，记为 A_ε. 从而 $A = \bigcup_{\varepsilon > 0} = \bigcup_{n \geq 1} A_{\frac{1}{n}}$ 是零测集。

$\omega \in A$, 当且仅当 $\exists \varepsilon > 0$, 使得 $\omega \in A_\varepsilon$, 当且仅当 $\exists \varepsilon > 0$, 使得 $\forall N > 0$, $\exists m, n > N$, 使得 $|S_n - S_m| > \varepsilon$, 由 Cauchy 收敛准则，这当且仅当 $S_n(\omega)$ 不收敛。即使得 $S_n(\omega)$ 不收敛的 ω 的集合就是 A, 是零测集，即 S_n 几乎处处收敛。 □

定理 6.6 (Kolmogorov 三级数)**.** 设 $X_1, X_2, \cdots$ 是独立的随机变量序列，记 $S_n = X_1 + \cdots + X_n$. 则 S_n 几乎处处收敛，当且仅当存在 $A > 0$, 令

$$Y_n = X_n \cdot \chi_{\{|X_n| \le A\}} = \begin{cases} X_n, & |X_n| \le A, \\ 0, & |X_n| > A, \end{cases}$$

使得

1. $\sum_{i=1}^{\infty} \mathbb{P}(|X_n| > A) < \infty$.
2. $\sum_{i=1}^{\infty} \mathbb{E}[Y_n] < \infty$.
3. $\sum_{i=1}^{\infty} \sigma^2(Y_n) < \infty$.

注记 6.7. *Kolmogorov 三级数定理告诉我们，独立随机变量序列的收敛性等价于三个数项级数的收敛性。*

证明. 充分性: 由 3 与 Kolmogorov 收敛准则 (定理 6.4), $\sum_{n=1}^{\infty} (Y_n - \mathbb{E}[Y_n])$ 几乎处处收敛。由 2, $\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty} (Y_n - \mathbb{E}[Y_n]) + \sum_{n=1}^{\infty} \mathbb{E}[Y_n]$ 几乎处处收敛。由 1, $\sum_{n=1}^{\infty} \mathbb{P}(X_n \ne Y_n) < \infty$. 由 Borel-Cantelli 定理（定理 2.16), $\mathbb{P}(\limsup_{n\to\infty} \{X_n \ne Y_n\}) = 0$. 这说明这样的 ω: 存在无穷多 n, 使得 $X_n(\omega) \ne Y_n(\omega)$ 几乎处处不存在。即有以概率 1 成立：$\forall \omega$, 只有有限个 n, 使得 $X_n(\omega) \ne Y_n(\omega)$. 从而 $\sum_{n=1}^{\infty} X_n(\omega)$ 与 $\sum_{n=1}^{\infty} Y_n(\omega)$ 的收敛性相同，从而 $S_n = \sum_{n=1}^{\infty} X_n$ 几乎处处收敛。

必要性:

1. 若否，$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \ge A) = \infty$. 由 Borel-Cantelli 定理（定理 2.16), $\mathbb{P}(\limsup_{n\to\infty} \{|S_n| \ge A\}) = 1$. 即几乎处处的 ω, 存在无穷多 n, 使得 $|X_n(\omega)| \ge A$. 从而 $S_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega)$ 不收敛。即 $S_n = \sum_{n=1}^{\infty} X_n$ 几乎处处不收敛，矛盾。

2. 由于 1 成立，可以用充分性中的证明，得到 $\sum_{n=1}^{\infty} X_n(\omega)$ 与 $\sum_{n=1}^{\infty} Y_n(\omega)$ 的收敛性相同。即 $\sum_{n=1}^{\infty} X_n$ 几乎处处收敛，当且仅当 $\sum_{n=1}^{\infty} Y_n$ 几乎处处收敛。不妨设初始时 X_n 一致有界。若 3 成立，由 Kolmogorov 收敛准则（定理 6.4), $\sum_{n=1}^{\infty} (Y_n - \mathbb{E}[Y_n])$ 几乎处处收敛。从而由 $\sum_{n=1}^{\infty} Y_n$ 几乎处处收敛，则 $\sum_{n=1}^{\infty} \mathbb{E}[Y_n] = \sum_{n=1}^{\infty} Y_n - \sum_{n=1}^{\infty} (Y_n - \mathbb{E}[Y_n])$ 几乎处处收敛。从而只要证 3.

3. 不妨设 X_n 一致有界。

定义 6.8. 称 X 是对称随机变量，若 X 与 $-X$ 同分布。此时，$\Phi_X = \Phi_{-X}$. 从而 $\forall x \in \mathbb{R}$,

$$\Phi_X(x) = \Phi_{-X}(x) = \mathbb{P}(-X \le x) = \mathbb{P}(X \ge -x) = 1 - \Phi_X(-x^-).$$

引理 6.9. 若 X, Y 独立同分布，则 $X - Y$ 是对称随机变量。

设 X 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量。令

$$\begin{tikzcd} & \Omega \arrow[dr, "X"] & \\ \overline{\Omega} = \Omega \times \Omega \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & & \mathbb{R} \\ & \Omega \arrow[ur, "X"'] & \end{tikzcd}$$

在概率空间 $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ 上，$p_1 \cdot X = X_1$, $p_2 \cdot X = X_2$. 则 X_1, X_2 独立，与 X 同分布。定义 $X^s = X_1 - X_2$, 则它是对称随机变量。它满足

(a) $\mathbb{E}[X^s] = \mathbb{E}[X_1] - \mathbb{E}[X_2] = 0$.

(b) $\sigma^2(X^s) = \sigma^2(X_1) + \sigma^2(-X_2) = 2\sigma^2(X)$. 从而由 $\sum_{i=1}^{\infty} \sigma^2(X_i) < \infty$, 有 $\sum_{i=1}^{\infty} \sigma^2(X_i^s) = 2\sum_{i=1}^{\infty} \sigma^2(X_i) < \infty$.

(c) 由 X_i 一致有界，设 $|X_i| \le A$. 则 $|X_i^s| = |X_{i,1} - X_{i,2}| \le 2A$. 即 X_i^s 一致有界。

(d) 由 $\sum_{i=1}^{\infty} X_i$ 几乎处处收敛知，$\sum_{i=1}^{\infty} X_i^s = \sum_{i=1}^{\infty} X_i - \sum_{i=2}^{\infty} X_i$ 也几乎处处收敛。

(e) 由 $X_1, X_2, \cdots$ 互相独立知，$X_1^s, X_2^s, \cdots$ 也相互独立。

从而就划归到 $X_1, X_2, \cdots$ 是相互独立的一致有界的对称随机变量的情形。

引理 6.10 (Kolmogorov)**.** 设 $X_1, X_2, \cdots$ 是相互独立的随机变量，一致有界：$|X_i| \le A$, $\mathbb{E}[X_i] = 0$. 记 $S_n = \sum_{i=1}^{n} X_i$. 则 $\forall \varepsilon > 0$, 有

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge \varepsilon\right) \ge 1 - \frac{(\varepsilon + A)^2}{\sigma^2(S_n)}.$$

证明. 令 $A_k=\{|S_k|\geq\varepsilon\}\cap\{\max_{i<k}|S_i|<\varepsilon\}$, $B_k=\{\max_{i\leq k}|S_i|<\varepsilon\}$. 注意到 $A_k\subset B_{k-1}$, 且 $\forall l\geq k$, 有 $A_k\cap B_l=\varnothing$, 且 $\bigcup_{i=1}^k A_i=\{\max_{1\leq i\leq k}|S_i|\geq\varepsilon\}=B_k^c$. 从而 $B_{k-1}=B_k\sqcup A_k$. 从而

$$S_{k-1}\cdot\chi_{B_{k-1}}+X_k\cdot\chi_{B_{k-1}}=S_k\cdot\chi_{B_{k-1}}=S_k\cdot\left(\chi_{B_k-1}+\chi_{A_k}\right).$$

从而左边看，注意到 X_k 与 $S_{k-1},\chi_{B_{k-1}}$ 独立，且 $\mathbb{E}\left[X_i\right]=0$, 从而

$$\begin{aligned}
\mathbb{E}\left[S_k^2\cdot\chi_{B_{k-1}}\right]=\mathbb{E}\left[\left(S_{k-1}+X_k\right)^2\cdot\chi_{B_{k-1}}\right]&=\mathbb{E}\left[S_{k-1}^2\cdot\chi_{B_{k-1}}+2\cdot S_{k-1}\cdot X_k\cdot\chi_{B_{k-1}}+X_k^2\cdot\chi_{B_{k-1}}\right]\\
&=\mathbb{E}\left[S_{k-1}^2\cdot\chi_{B_{k-1}}\right]+2\cdot\mathbb{E}\left[X_k\right]\cdot\mathbb{E}\left[S_{k-1}\cdot\chi_{B_{k-1}}\right]+\mathbb{E}\left[X_k^2\right]\cdot\mathbb{E}\left[\chi_{B_{k-1}}\right]\\
&=\mathbb{E}\left[S_{k-1}^2\cdot\chi_{B_{k-1}}\right]+0+\left(\mathbb{E}\left[X_k^2\right]+\mathbb{E}\left[X_k\right]^2\right)\cdot\mathbb{E}\left[\chi_{B_{k-1}}\right]\\
&=\mathbb{E}\left[S_{k-1}^2\cdot\chi_{B_{k-1}}\right]+\sigma^2(X_k)\cdot\mathbb{P}\left(B_{k-1}\right).
\end{aligned}$$

从右边看，由 X_i 的独立性，且 $\mathbb{E}\left[X_i\right]=0$, 有估计

$$\begin{aligned}
\mathbb{E}\left[S_k^2\cdot\chi_{B_{k-1}}\right]&=\mathbb{E}\left[S_k^2\cdot\left(\chi_{B_{k-1}}+\chi_{A_k}\right)\right]\\
&=\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]+\mathbb{E}\left[\left(S_{k-1}+X_k\right)^2\cdot\chi_{A_k}\right]\\
&=\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]+\mathbb{E}\left[S_{k-1}\cdot\chi_{A_k}\right]+2\cdot\mathbb{E}\left[X_k\right]\cdot\mathbb{E}\left[S_{k-1}\cdot\chi_{A_k}\right]+\mathbb{E}\left[X_k^2\cdot\chi_{A_k}\right]\\
&\leq\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]+\mathbb{E}\left[\varepsilon^2\cdot\chi_{A_k}\right]+0+\mathbb{E}\left[A^2\cdot\chi_{A_k}\right]\\
&=\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]+\varepsilon^2\cdot\mathbb{P}\left(A_k\right)+A^2\cdot\mathbb{P}\left(A_k\right)\\
&\leq\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]+\left(\varepsilon+A\right)^2\cdot\mathbb{P}\left(A_k\right).
\end{aligned}$$

二式相减，得到

$$\mathbb{P}\left(B_{k-1}\right)\cdot\sigma^2\left(X_k\right)\leq\mathbb{E}\left[S_k^2\cdot\chi_{B_k}\right]-\mathbb{E}\left[S_n^2\cdot\chi_{B_{k-1}}\right]+\left(\varepsilon+A\right)^2\cdot\mathbb{P}\left(A_k\right).$$

另一方面，$\forall k \le n$, $B_n \subset B_k$, 且 $A_i \cap A_j = \varnothing$, 从而

$$\begin{aligned}
\mathbb{P}\left(B_n\right) \cdot \sum_{k=1}^{n} \sigma^2(X_k) \le \sum_{k=1}^{n} \mathbb{P}\left(B_{k-1}\right) \cdot \sigma^2(X_k) &\le \sum_{k=1}^{n} \left(\mathbb{E}\left[S_k^2 \cdot \chi_{B_k}\right] - \mathbb{E}\left[S_n^2 \cdot \chi_{B_{k-1}}\right]\right) + \sum_{k=1}^{n} \left(\varepsilon + A\right)^2 \cdot \mathbb{P}\left(A_k\right) \\
&= \mathbb{E}\left[S_n^2 \cdot \chi_{B_n}\right] + \left(\varepsilon + A\right)^2 \cdot \mathbb{P}\left(\bigsqcup_{k=1}^{n} A_k\right) \\
&\le \mathbb{E}\left[\varepsilon^2 \cdot \chi_{B_n}\right] + \left(\varepsilon + A\right)^2 \cdot \mathbb{P}\left(\bigsqcup_{k=1}^{n} A_k\right) \\
&= \varepsilon^2 \cdot \mathbb{P}\left(B_n\right) + \left(\varepsilon + A\right)^2 \cdot \mathbb{P}\left(B_n^c\right) \\
&= \varepsilon^2 \cdot \left(\mathbb{P}\left(B_n\right) + \mathbb{P}\left(B_n^c\right)\right) + 2\varepsilon A \cdot \mathbb{P}\left(B_n^c\right) + A^2 \cdot \mathbb{P}\left(B_n^c\right) \\
&\le \varepsilon^2 + 2\varepsilon A + A^2 \\
&= \left(\varepsilon + A\right)^2.
\end{aligned}$$

最后，

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge \varepsilon\right) = \mathbb{P}\left(B_n^c\right) = 1 - \mathbb{P}\left(B_n\right) \ge 1 - \frac{\left(\varepsilon + A\right)^2}{\sum_{k=1}^{n} \sigma^2\left(X_k\right)} = 1 - \frac{\left(\varepsilon + A\right)^2}{\sigma^2\left(S_n\right)}.$$

□

由此，记 $S_n^s = \sum_{i=1}^{n} X_i^s$, 由 $|X_i^s| \le 2A$, $\mathbb{E}\left[X_i^s\right] = 0$, 则有

$$\mathbb{P}\left(\max_{N \le k \le M} |S_k^s - S_N^s| \ge \varepsilon\right) \ge 1 - \frac{\left(\varepsilon + 2A\right)^2}{\sigma^2\left(S_M^s - S_N^s\right)} = 1 - \frac{\left(\varepsilon + 2A\right)^2}{\sum_{i=N+1}^{M} \sigma^2\left(X_i^s\right)}.$$

从而若 $\sum_{n=1}^{\infty} \sigma^2(X_i^s) = \infty$, 则 $\forall N$, $\sum_{i=N+1}^{\infty} \sigma^2(X_i^s) = \infty$. 在上式中令 $M \to \infty$, 有 $\mathbb{P}\left(\max_{N \le k} |S_k^s - S_N^s| \ge \varepsilon\right) = 1$. 这说明 S_k^s 几乎处处不收敛，与假设矛盾。

□

注记 6.11. 令 $X_n' = \frac{X_n}{A}$, $Y_n' = \frac{Y_n}{A} = X_n' \cdot \chi_{\{|X_n| \le A\}} = X_n' \cdot \chi_{\{|X_n'| \le 1\}}$. 注意到

$$
\begin{array}{ccc}
\sum X_i < \infty & \Longleftrightarrow & \sum X_n' = \frac{1}{A} \sum X_i < \infty \\
\Big\Updownarrow{\scriptstyle \exists A} & & \Big\Updownarrow \\
\begin{cases} \sum \mathbb{P}(X_n > A) < \infty. \\ \sum \mathbb{E}[Y_n] < \infty. \\ \sum \sigma^2(Y_n) < \infty. \end{cases} & \Longleftrightarrow & \begin{cases} \sum \mathbb{P}(X_n' > 1) = \sum \mathbb{P}(X_n > A) < \infty. \\ \sum \mathbb{E}[Y_n'] = \frac{1}{A} \sum \mathbb{E}[Y_n] < \infty. \\ \sum \sigma^2(Y_n) = \frac{1}{A^2} \sum \sigma^2(Y_n) < \infty. \end{cases}
\end{array}
$$

从而不妨设 $A = 1$. 这也说明 *Kolmogorov* 三级数定理中的 A 可以是任意取的。

定理 6.12. 设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，记 $S_n = X_1 + \cdots + X_n$. 则以下三条等价:

1. $S_n \xrightarrow{a.e.} S$.
2. $S_n \xrightarrow{m} S$.
3. $S_n \xrightarrow{d} S$.

证明. $1 \Rightarrow 2 \Rightarrow 3$ 是显然的，只要证 $3 \Rightarrow 2 \Rightarrow 1$.

引理 6.13 (Skorohod)**.** 设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，记 $S_n = X_1 + \cdots + X_n$. $\forall \varepsilon > 0$, 若 $\beta = \max_{1 \le k \le n} \mathbb{P}\left(|S_n - S_k| > \varepsilon\right) < 1$, 则

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge 2\varepsilon\right) \le \frac{1}{1-\beta} \cdot \mathbb{P}\left(|S_n| \ge \varepsilon\right).$$

证明. $\forall \omega \in \{\max_{1 \le k \le n} |S_k| \ge 2\varepsilon\}$, 记

$$\tau(\omega) = \min\left\{k \middle| 1 \le k \le n, |S_k| \ge 2\varepsilon\right\}.$$

$$A_k = \left\{\omega \in \left\{\max_{1 \le k \le n} |S_k| \ge 2\varepsilon\right\} \middle| \tau(\omega) = k\right\} = \left\{|S_1| < 2\varepsilon, \cdots, |S_{k-1}| < 2\varepsilon, |S_n| \ge 2\varepsilon\right\}.$$

则

$$\{|S_n - S_k| \leq \varepsilon\} \cap \{|S_k| \geq 2\varepsilon\} \subset \{|S_n| = |S_n - S_k + S_k| \geq |S_k| - |S_n - S_k| = 2\varepsilon - \varepsilon = \varepsilon\} = \{|S_n| \geq \varepsilon\}.$$

从而由于 $A_1, \cdot, A_n$ 相互独立,

$$\mathbb{P}(|S_n| \geq \varepsilon) \geq \sum_{k=1}^{n} \mathbb{P}(\{|S_n \geq \varepsilon|\} \cap A_k) \geq \sum_{k=1}^{n} \mathbb{P}(\{|S_n - S_k| \leq \varepsilon\} \cap A_k).$$

注意到 $\{|S_n - S_k| \leq \varepsilon\}$ 与 A_k 独立，从而 $\mathbb{P}(\{|S_n - S_k| \leq \varepsilon\} \cap A_k) = \mathbb{P}(\{|S_n - S_k| \leq \varepsilon\}) \cdot \mathbb{P}(A_k)$, 从而

$$\begin{aligned}\mathbb{P}(|S_n| \geq \varepsilon) \geq \sum_{k=1}^{n} \mathbb{P}(\{|S_n - S_k| \leq \varepsilon\} \cap A_k) = \sum_{k=1}^{n} \mathbb{P}(\{|S_n - S_k| \leq \varepsilon\}) \cdot \mathbb{P}(A_k) &= \sum_{k=1}^{n} (1 - \mathbb{P}(\{|S_n - S_k| > \varepsilon\}))\, \mathbb{P}(A_k) \\ &\geq \sum_{k=1}^{n} \left(1 - \max_{1 \leq k \leq n} \mathbb{P}(\{|S_n - S_k| > \varepsilon\})\right) \mathbb{P}(A_k) = \sum_{k=1}^{n} (1 - \beta) \cdot \mathbb{P}(A_k).\end{aligned}$$

又注意到 $A_1, \cdots, A_n$ 相互独立，且 $A_1 \sqcup \cdots \sqcup A_n = \{\max_{1 \leq k \leq n} |S_k| \geq 2\varepsilon\}$, 从而

$$\mathbb{P}(|S_n| \geq \varepsilon) \geq \sum_{k=1}^{n} (1 - \beta) \cdot \mathbb{P}(A_k) = (1 - \beta) \cdot \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq 2\varepsilon\right).$$

即

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq 2\varepsilon\right) \leq \frac{1}{1 - \beta} \cdot \mathbb{P}(|S_n| \geq \varepsilon).$$

□

$2 \Rightarrow 1$: $\forall \delta > 0$, $\forall 0 \leq \varepsilon \leq \frac{1}{2}$, $\exists N$, 使得 $\forall m, n \geq N$, 有 $\mathbb{P}(|S_n - S_m| > \delta) \leq \varepsilon$. 对 $M > N$, 有 $\beta = \max_{N \leq k \leq M} \mathbb{P}(|S_M - S_k| \geq \delta) \leq \varepsilon < \frac{1}{2} < 1$. 由 Skorohod 不等式（引理 6.13）,

$$\mathbb{P}\left(\max_{N \leq k \leq M} |S_k - S_N| \geq 2\varepsilon\right) \leq \frac{1}{1 - \beta} \cdot \mathbb{P}(|S_M - S_N| \geq \delta) \leq \frac{\varepsilon}{1 - \beta} \leq \frac{\varepsilon}{1 - \frac{1}{2}} = 2\varepsilon.$$

令 $M \to \infty$, 有 $\mathbb{P}(\max_{k \geq N} |S_k - S_N| \geq 2\varepsilon) \leq 2\varepsilon$. 从而

$$\mathbb{P}\left(\max_{k, l \geq N} |S_k - S_l| \geq 4\varepsilon\right) \leq \mathbb{P}\left(\max_{k \geq N} |S_k - S_N| \geq 2\varepsilon\right) \leq 2\varepsilon.$$

这就是 $S_n \xrightarrow{a.e.} S$ 的 Cauchy 收敛准则。

注记 6.14. *Skorohod 不等式（引理 6.13）的作用就是把概率测度 $\mathbb{P}$ 外的 $\max$ 移到了里面。*

$3 \Rightarrow 2$: 考虑示性函数 $\varphi_{S_n}(z) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,zS_n}\right]$(定义 4.18). 在紧集上，$\varphi_{S_n}(z)$ 一致收敛到 $\varphi_S(z)$(作业). 且 $\varphi_S(0) = 1$. 从而 $\forall \varepsilon > 0, \exists a > 0, N > 0$, 使得 $\forall m > n \geq N, \forall |z| \leq a$, 都有 $|\varphi_{S_n}(z)| \geq \frac{1}{2}$, $|\varphi_{S_m}(z) - \varphi_{S_n}(z)| \leq \frac{1}{2}\varepsilon$. 又注意到若随机变量 X, Y 相互独立，则 $\varphi_{X+Y}(z) = \varphi_X(z) \cdot \varphi_Y(z)$(性质 7.1), 从而 $\varphi_{S_m}(z) = \varphi_{(S_m - S_n)+S_n}(z) = \varphi_{S_m - S_n}(z) \cdot \varphi_{S_n}(z)$, 此时，

$$|1 - \varphi_{S_m - S_n}(z)| = \left|\frac{\varphi_{S_n}(z) - \varphi_{S_n}(z)\varphi_{S_m - S_n}(z)}{\varphi_{S_n}(z)}\right| = \left|\frac{\varphi_{S_n}(z) - \varphi_{S_m}(z)}{\varphi_{S_n}(z)}\right| \leq \frac{\frac{1}{2}\varepsilon}{\frac{1}{2}} = \varepsilon.$$

另一方面，由 Fubini 定理，

$$1 - \varphi_{S_m - S_n}(z) = \int_{\mathbb{R}} \left(1 - \mathrm{e}^{\mathrm{i}\,zx}\right) \mathrm{d}\,\mu_{S_m - S_n} = \frac{1}{2a}\int_{-a}^{a} \left(1 - \mathrm{e}^{\mathrm{i}\,zx}\right) \mathrm{d}\,\mu_{S_m - S_n}\,\mathrm{d}\,z = \int_{\mathbb{R}} \frac{1}{2a}\int_{-a}^{a} \left(1 - \mathrm{e}^{\mathrm{i}\,zx}\right) \mathrm{d}\,z\,\mathrm{d}\,\mu_{S_m - S_n} = \int_{\mathbb{R}} \left(1 - \frac{\sin ax}{ax}\right) \mathrm{d}\,\mu_{S_m - S_n}.$$

其中，函数 $1 - \frac{\sin ax}{ax}$ 在原点附近有展开：$1 - \frac{\sin ax}{ax} = \frac{(ax)^2}{6} + \mathrm{o}(x^3)$. 且 $x \to \infty$ 时，$1 - \frac{\sin ax}{ax} \to 1$. 从而函数

$$f(x) = \frac{1 - \frac{\sin ax}{ax}}{\frac{(ax^2)}{6+(ax)^2}}$$

在 $x \to \pm\infty$ 时趋于 1. 从而 $\beta = \min_{x\in\mathbb{R}} f(x) > 0$ 是有定义的。即存在 $\beta > 0$, 使得 $\forall x \in \mathbb{R}$, $f(x) \geq \beta$, 即 $\forall x \in \mathbb{R}$,

$$1 - \frac{\sin ax}{ax} \geq \frac{\beta(ax)^2}{6 + (ax)^2}.$$

于是 $\forall \delta>0$, 注意到 $\forall |x|\geq\delta$, 由于函数 $\frac{6+a^2x^2}{\beta a^2x^2}=\frac{6}{\beta a^2x^2}+\frac{1}{\beta}$ 是单调递减的，从而 $\frac{6+a^2x^2}{\beta a^2x^2}\leq\frac{6+a^2\delta^2}{\beta a^2\delta^2}$, 即 $\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot\frac{6+(ax)^2}{\beta(ax)^2}\geq 1$. 综上，

$$\begin{aligned}\mathbb{P}\left(|S_m-S_n|\geq\delta\right)=\int_{|x|\geq\delta}\mathrm{d}\,\mu_{S_m-S_n}&=\int_{|x|\geq\delta}1\cdot\mathrm{d}\,\mu_{S_m-S_n}\\&\leq\int_{|x|\geq\delta}\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot\frac{6+(ax)^2}{\beta(ax)^2}\,\mathrm{d}\,\mu_{S_m-S_n}=\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot\int_{|x|\geq\delta}\frac{6+(ax)^2}{\beta(ax)^2}\,\mathrm{d}\,\mu_{S_m-S_n}\\&\leq\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot\int_{|x|\geq\delta}\left(1-\frac{\sin ax}{ax}\right)\mathrm{d}\,\mu_{S_m-S_n}\leq\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot\int_{\mathbb{R}}\left(1-\frac{\sin ax}{ax}\right)\mathrm{d}\,\mu_{S_m-S_n}\\&=\frac{6+a^2\delta^2}{\beta a^2\delta^2}\cdot|1-\varphi_{S_m-S_n}(z)|\\&\leq\frac{(6+a^2\delta^2)\varepsilon}{\beta a^2\delta^2}.\end{aligned}$$

这就是 S_n 依概率测度收敛的 Cauchy 收敛准则，即得 $S_n\xrightarrow{m}S$.

□

引理 6.15 (Kronecker)**.** 设 $a_n\nearrow\infty$, $a_n>0$, $\{x_n\}\subset\mathbb{R}$, 使得 $\sum_{i=1}^{\infty}\frac{x_i}{a_i}$ 收敛。则数列 $\frac{1}{a_n}\sum_{i=1}^{n}x_i\to 0$.

证明. 记 $b_n=\sum_{i=1}^{n}\frac{x_i}{a_i}$, $a_0=b_0=0$, $b=\sum_{i=1}^{\infty}\frac{x_i}{a_i}$, 有 $b_n\to b$. 有 $b_n-b_{n-1}=\frac{x_n}{a_n}$, 即 $x_n=a_n(b_n-b_{n-1})$. 从而

$$\frac{1}{a_n}\sum_{i=1}^{n}x_i=\frac{1}{a_n}\cdot\sum_{i=1}^{n}a_i(b_i-b_{i-1})=\frac{1}{a_n}\cdot\left(\sum_{i=1}^{n}a_ib_i-\sum_{i=1}^{n}a_ib_{i-1}\right)=\frac{1}{a_n}\cdot\left(a_nb_n+\sum_{i=1}^{n}b_{i-1}\left(a_{i-1}-a_i\right)\right)=b_n-\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}.$$

从而只要证 $\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}\to b$, 这样就有 $\frac{1}{a_n}\sum_{i=1}^{n}x_i=b_n-\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}\to b-b=0$.

由于 $b_n\to b$, 从而 $\forall\varepsilon>0$, $\exists N$, 使得 $\forall n\geq N$, 都有 $|b_n-b|<\varepsilon$. 并且 $\{b_n\},b$ 是有界的，设 $B<\infty$ 是一个上界. 注意到 $\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}=$

$\frac{a_n-a_0}{a_n}=1$. 此时，

$$\begin{aligned}
\left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}-b\right| &= \left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}-1\cdot b\right| = \left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}-\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b\right| \\
&= \left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}(b_{i-1}-b)\right| \\
&\le \left|\sum_{i=1}^{N}\frac{a_i-a_{i-1}}{a_n}(b_{i-1}-b)\right| + \left|\sum_{i=N+1}^{n}\frac{a_i-a_{i-1}}{a_n}(b_{i-1}-b)\right| \le \sum_{i=1}^{N}\frac{a_i-a_{i-1}}{a_n}\left|(b_{i-1}-b)\right| + \sum_{i=N+1}^{n}\frac{a_i-a_{i-1}}{a_n}\left|(b_{i-1}-b)\right| \\
&\le 2B\cdot\sum_{i=1}^{N}\frac{a_i-a_{i-1}}{a_n}+\varepsilon\cdot\sum_{i=N+1}^{n}\frac{a_i-a_{i-1}}{a_n} = 2B\cdot\frac{a_N}{a_n}+\varepsilon\cdot\frac{a_n-a_N}{a_n}.
\end{aligned}$$

其中，由于 $a_n\to\infty$, 从而 $\frac{a_N}{a_n}\to 0$, $\frac{a_n-a_N}{a_n}=1-\frac{a_N}{a_n}\to 1$. 从而

$$\left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}-b\right| \le 2B\cdot\frac{a_N}{a_n}+\varepsilon\cdot\frac{a_n-a_N}{a_n}\to 2B\cdot 0+\varepsilon\cdot 1=\varepsilon.$$

由 ε 的任意性，$\left|\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}-b\right|\to 0$, 即 $\sum_{i=1}^{n}\frac{a_i-a_{i-1}}{a_n}b_{i-1}\to b$. □

定理 6.16. 设 $X_1, X_2, \cdots$ 是互相独立的随机变量，$\mathbb{E}[X_i]=0$, $\sigma^2(X_i)\le C<\infty$, $\forall i$. 则 $\forall\varepsilon>0$,

$$\frac{\sum_{i=1}^{n}X_i}{n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}}\xrightarrow{a.e.}0.$$

证明. 令 $a_n=n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}$, 考虑相互独立的随机变量序列 $\frac{X_n}{a_n}$. 有

$$\sum_{i=1}^{\infty}\sigma^2\left(\frac{X_n}{a_n}\right)=\sum_{i=1}^{\infty}\left(\mathbb{E}\left[\left(\frac{X_n}{a_n}\right)^2\right]-\mathbb{E}\left[\frac{X_n}{a_n}\right]^2\right)=\sum_{i=1}^{\infty}\mathbb{E}\left[\left(\frac{X_n}{a_n}\right)^2\right]=\sum_{i=1}^{\infty}\frac{1}{a_n^2}\mathbb{E}\left[X_n^2\right]\le\sum_{i=1}^{\infty}\frac{C}{n(\log n)^{1+2\varepsilon}}<\infty.$$

从而由 Kolmogorov 收敛准则(定理 6.4), $\sum_{i=1}^{n}\frac{X_i}{a_i}$ 几乎处处收敛。由 Kronecker 引理(引理 6.15), 在这些收敛的点 ω 处, 有 $\frac{1}{a_n}\cdot\sum_{i=1}^{n}X_i(\omega)\to 0$, 即

$$\frac{1}{a_n}\cdot\sum_{i=1}^{n}X_i=\frac{\sum_{i=1}^{n}X_i}{n^{\frac{1}{2}}\cdot(\log n)^{\frac{1}{2}+\varepsilon}}\xrightarrow{a.e.}0$$

□

定理 6.17. 设 $X_1,X_2,\cdots$ 是一列独立同分布的随机变量, 记 $Y_n=\max_{1\le i\le n}|X_i|$. $\{b_n\}\subset\mathbb{R}$, $b_n\nearrow b>0$. 则 $\frac{Y_n}{b_n}\xrightarrow{m}0$, 当且仅当 $\forall\varepsilon>0$, $n\cdot\mathbb{P}\left(|X_1|\ge b_n\varepsilon\right)\to 0$.

证明. 充分性: 注意到 $\{Y_n\ge b_n\varepsilon\}=\bigcup_{i=1}^{n}\{|X_i|\ge b_n\varepsilon\}$, 从而

$$\mathbb{P}\left(\frac{Y_n}{b_n}\ge\varepsilon\right)=\mathbb{P}\left(\bigcup_{i=1}^{n}\{|X_i|\ge b_n\varepsilon\}\right)\le\sum_{i=1}^{n}\mathbb{P}\left(|X_n|\ge b_n\varepsilon\right)=n\cdot\mathbb{P}\left(|X_1|\ge b_n\varepsilon\right)\to 0.$$

即 $\frac{Y_n}{b_n}\xrightarrow{m}0$.

必要性:

$$\mathbb{P}\left(\frac{Y_n}{b_n}\ge\varepsilon\right)=1-\mathbb{P}\left(|X_1|<b_n\varepsilon,\cdots,|X_n|<b_n\varepsilon\right)=1-\prod_{i=1}^{n}\mathbb{P}\left(|X_i|<b_n\varepsilon\right)=1-\prod_{i=1}^{n}\left(1-\mathbb{P}\left(|X_i|\ge b_n\varepsilon\right)\right)=1-\left(1-\mathbb{P}\left(|X_1|\ge b_n\varepsilon\right)\right)^n\ge 1-\mathrm{e}^{-n\mathbb{P}(|X_i|\ge b_n\varepsilon)}\ge 0.$$

由于 $\frac{Y_n}{b_n}\xrightarrow{m}0$, 从而 $\mathbb{P}\left(\frac{Y_n}{b_n}\ge\varepsilon\right)\to 0$, 从而 $1-\mathrm{e}^{-n\mathbb{P}(|X_i|\ge b_n\varepsilon)}\le\mathbb{P}\left(\frac{Y_n}{b_n}\ge\varepsilon\right)\to 0$, 从而 $n\cdot\mathbb{P}\left(|X_1|\ge b_n\varepsilon\right)\to 0$. □

注记 6.18. 容易推广到独立不一定同分布的随机变量序列的情况:

定理 6.19. 设 $X_1,X_2,\cdots$ 是一列相互独立的随机变量, 记 $Y_n=\max_{1\le i\le n}|X_i|$. $\{b_n\}\subset\mathbb{R}$, $b_n\nearrow b>0$. 则 $\frac{Y_n}{b_n}\xrightarrow{m}0$, 当且仅当 $\forall\varepsilon>0$, $\sum_{i=1}^{n}\mathbb{P}\left(|X_i|\ge b_n\varepsilon\right)\to 0$.

7 示性函数与中心极限定理

历史上，问题来源于 Bernoulli 实验：投硬币，正的概率是 p, 反的概率是 $1-p$. 记 S_n 是投 n 次正的次数，则 $\mathbb{E}\left[S_n\right]=p^n$. 但只有期望我们得不到 S_n 的分布情况。重复实验发现，越靠近均值取值的概率越高。我们将证明，S_n 会趋向一个正态分布，即 $\frac{S_n-p^n}{\sigma^2(S_n)} \xrightarrow{d} \mathcal{N}(0,1)$. 最后我们将这个结果推广到一般的独立同分布的随机变量序列：设 $X_1, X_2, \cdots$ 是独立同分布的随机变量序列，$\mathbb{E}\left[X_1\right]=\mu$, $\sigma^2(X_1)=\sigma^2$, 则 $\frac{\sum_{i=1}^n X_i-n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$.

性质 7.1. X 是随机变量，$\varphi_X(z)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,zX}\right]=\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,zx}\,\mathrm{d}\,\mu_X$ 是它的示性函数（定义 *4.18*），它是一个实变元的复值函数。它只与分布 Φ_X 或者说 μ_X 有关，与具体的概率空间无关。则

1. $\varphi_X(0)=1$.
2. $|\varphi_X| \leq 1$.
3. $\varphi_X(-z)=\varphi_{-X}(z)=\overline{\varphi_X(z)}$.
4. $\forall a, b \in \mathbb{R}$, $\varphi_{aX+b}(z)=\mathrm{e}^{\mathrm{i}\,zb} \cdot \varphi_X(az)$.
5. 若 X, Y 相互独立，则 $\varphi_{X+Y}(z)=\varphi_X(z) \cdot \varphi_Y(z)$.
6. 若 X 对称（定义 *6.8*），则 $\varphi_X(z)=\varphi_{-X}(z)=\overline{\varphi_X(z)}$, 从而 $\varphi_X(z)$ 是实值函数。反之也对：若 $\varphi_X(z)$ 是实值函数，则 X 是对称随机变量。
7. $\varphi_X(z)$ 是一致连续的。
8. 示性函数的凸组合也是示性函数：设 φ_X 与 φ_Y 是两个示性函数，$\lambda_1, \lambda_2 \geq 0$, $\lambda_1+\lambda_2=1$, 则 $\lambda_1\varphi_X+\lambda_2\varphi_Y$ 也是示性函数。
9. 若 X 是连续型随机变量，有密度函数 $f_X(x)$, 则 $\mathrm{d}\,\mu_X=f_X(x)\,\mathrm{d}\,x$, 从而 $\varphi_X(z)=\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,zx} f_X(x)\,\mathrm{d}\,x$ 是 $f_X(x)$ 的 *Fourier* 逆变换。

证明. 只证 7,8, 其它几条由定义和简单计算即得（除了 6 的反方向，我们将在之后证明）。

7:

$$|\varphi_X(z+h)-\varphi_X(z)| = \left|\int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}(z+h)x}-\mathrm{e}^{\mathrm{i}\,zx}\right) \mathrm{d}\,\mu_X\right| \leq \int_{\mathbb{R}} \left|\mathrm{e}^{\mathrm{i}\,zx}\right| \cdot \left|\mathrm{e}^{\mathrm{i}\,hx}-1\right| \mathrm{d}\,\mu_X = \int_{\mathbb{R}} \left|\mathrm{e}^{\mathrm{i}\,hx}-1\right| \mathrm{d}\,\mu_X.$$

由控制收敛定理，$h\to 0$ 时，$\mathrm{e}^{\mathrm{i}\,hx}-1\to 0$, 从而 $|\varphi_X(z+h)-\varphi_X(z)| \leq \int_{\mathbb{R}} \left|\mathrm{e}^{\mathrm{i}\,hx}-1\right| \mathrm{d}\,\mu_X \to 0$. 即 $\varphi_X(z)$ 一致收敛。

8: 令 $\mu_Z=\lambda_1\mu_X+\lambda_2\mu_Y$, 验证它是 $\mathbb{R}$ 上的概率测度，从而 $\varphi_Z=\lambda_1\varphi_X+\lambda_2\varphi_Y$ 就是随机变量 Z 的示性函数。 □

例 7.2.

- 若 $X\sim\mathcal{N}(0,1)$, $f_X(x)=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$, 从而

$$\varphi_X(z)=\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,zx}\cdot\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}\,x=\mathrm{e}^{-\frac{z^2}{2}}.$$

  一般地，若 $Y\sim\mathcal{N}(\mu,\sigma^2)$, 即 $Y=\sigma X+\mu$, 则 $\varphi_Y(z)=\frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}}$.

- 若 $X\sim(0,1]$ 均匀分布，则

$$\varphi_X(z)=\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,zx}\,\mathrm{d}\,\mu_X=\int_0^1 \mathrm{e}^{\mathrm{i}\,zx}\,\mathrm{d}\,x=\frac{2}{z}\cdot\mathrm{e}^{\mathrm{i}\,\frac{z}{2}}\cdot\sin\left(\frac{z}{2}\right).$$

引理 7.3 (Dirichlet 积分). 记 $\operatorname{sgn}(a)=\begin{cases}1, & a>0;\\ 0, & a=0;\\ -1, & a<0\end{cases}$ 是符号函数，则 $\forall y>0$,

1.

$$0\leq \operatorname{sgn}(a)\int_0^y \frac{\sin az}{z}\,\mathrm{d}\,z\leq\int_0^\pi\frac{\sin z}{z}\,\mathrm{d}\,z.$$

2.

$$\int_0^\infty\frac{\sin az}{z}\,\mathrm{d}\,z=\operatorname{sgn}(a)\cdot\frac{\pi}{2}.$$

证明. 做变量代换 $u=az$. 则 $\int_0^y \frac{\sin az}{z}\,\mathrm{d}\,z=\int_0^{ay}\frac{\sin u}{u}\,\mathrm{d}\,u$. 由 y 的任意性，若 $a>0$, 则 $0\le \operatorname{sgn}(a)\int_0^y \frac{\sin az}{z}\,\mathrm{d}\,z\le \int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$ 当且仅当 $0\le \operatorname{sgn}(1)\int_0^{ay}\frac{\sin u}{u}\,\mathrm{d}\,u\le \int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$, $\int_0^{\infty}\frac{\sin az}{z}\,\mathrm{d}\,z=\operatorname{sgn}(a)\cdot\frac{\pi}{2}$ 当且仅当 $\int_0^{\infty}\frac{\sin u}{u}\,\mathrm{d}\,u=\operatorname{sgn}(1)\cdot\frac{\pi}{2}$; 若 $a<0$, 则 $0\le \operatorname{sgn}(a)\int_0^y \frac{\sin az}{z}\,\mathrm{d}\,z\le \int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$ 当且仅当 $0\le -\operatorname{sgn}(a)\int_0^{-ay}\frac{\sin u}{u}\,\mathrm{d}\,u\le\int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$, 当且仅当 $0\le \operatorname{sgn}(1)\int_0^{-ay}\frac{\sin u}{u}\,\mathrm{d}\,u\le \int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$; $\int_0^{\infty}\frac{\sin az}{z}\,\mathrm{d}\,z=\operatorname{sgn}(a)\cdot\frac{\pi}{2}$ 当且仅当 $\int_0^{-\infty}\frac{\sin u}{u}\,\mathrm{d}\,u=\operatorname{sgn}(a)\cdot\frac{\pi}{2}$, 当且仅当 $\int_0^{\infty}\frac{\sin u}{u}\,\mathrm{d}\,u=\operatorname{sgn}(1)\cdot\frac{\pi}{2}$. 于是只要证 $a=1$ 的情况即可。

1. 对变限积分 $\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z$ 求导，使它为 0: $\left(\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z\right)'=\frac{\sin y}{y}=0$, 计算极值点为 $\{n\pi\}$. 于是 $\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z$ 的所有极值也是可能的最值为 $\int_0^{n\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$. 注意到
$$\int_0^{n\pi}\frac{\sin z}{z}\,\mathrm{d}\,z=\sum_{k=1}^{n}\int_{(k-1)\pi}^{k\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$$
是一个首项为正，绝对值递减的交错级数，于是只有第一项为最大。从而函数 $\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z$ 的可能的极大值 $\int_0^{n\pi}\frac{\sin z}{z}\,\mathrm{d}\,z\le\int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$. 并且求和项两两配对即得函数 $\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z$ 的可能的极小值 $\int_0^{n\pi}\frac{\sin z}{z}\,\mathrm{d}\,z$ 都恒正。即有
$$0\le\int_0^y\frac{\sin z}{z}\,\mathrm{d}\,z\le\int_0^{\pi}\frac{\sin z}{z}\,\mathrm{d}\,z.$$

2. 直接计算：由 Fubini 定理，
$$\int_0^{\infty}\frac{\sin z}{z}\,\mathrm{d}\,z=\int_0^{\infty}\sin z\left(\int_0^{\infty}\mathrm{e}^{-uz}\,\mathrm{d}\,u\right)\mathrm{d}\,z=\int_0^{\infty}\sin z\int_0^{\infty}\mathrm{e}^{-uz}\,\mathrm{d}\,u\,\mathrm{d}\,z=\int_0^{\infty}\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z\,\mathrm{d}\,u.$$
其中，
$$\begin{aligned}\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z&=-\frac{1}{u}\cdot\int_0^{\infty}\sin z\,\mathrm{d}\,\mathrm{e}^{-uz}=-\frac{1}{u}\cdot\sin z\cdot\mathrm{e}^{-uz}\Big|_0^{\infty}+\frac{1}{u}\cdot\int_0^{\infty}\cos z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z\\&=-\frac{1}{u^2}\cdot\int_0^{\infty}\cos z\,\mathrm{d}\,\mathrm{e}^{-uz}=-\frac{1}{u^2}\cdot\cos z\cdot\mathrm{e}^{-uz}\Big|_0^{\infty}-\frac{1}{u^2}\cdot\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z\\&=\frac{1}{u^2}-\frac{1}{u^2}\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z.\end{aligned}$$
即 $\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}=\frac{1}{1+u^2}$. 从而
$$\int_0^{\infty}\frac{\sin z}{z}\,\mathrm{d}\,z=\int_0^{\infty}\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z\,\mathrm{d}\,u=\int_0^{\infty}\left(\int_0^{\infty}\sin z\cdot\mathrm{e}^{-uz}\,\mathrm{d}\,z\right)\mathrm{d}\,u=\int_0^{\infty}\frac{1}{1+u^2}\,\mathrm{d}\,u=\arctan u|_0^{\infty}=\frac{\pi}{2}.$$

□

定理 7.4 (逆转公式)**.** 设 X 是一个随机变量，$\varphi_X(z)$ 是它的示性函数，$\Phi_X(x)$ 是它的分布函数，设 $a<b$, 则

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\,az}-\mathrm{e}^{-\mathrm{i}\,bz}}{\mathrm{i}\,z}\varphi_X(z)\,\mathrm{d}\,z=\Phi_X(b)-\Phi_X(a)+\frac{1}{2}\cdot\mathbb{P}\left(X=a\right)-\frac{1}{2}\cdot\mathbb{P}\left(X=b\right).$$

注记 7.5. 逆转公式告诉我们，如果两个随机变量示性函数相同，则它们有相同的分布。

证明. 由 Fubini 定理与奇偶性，记

$$\begin{aligned}
I_T&=\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\,az}-\mathrm{e}^{-\mathrm{i}\,bz}}{\mathrm{i}\,z}\varphi_X(z)\,\mathrm{d}\,z\\
&=\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\,az}-\mathrm{e}^{-\mathrm{i}\,bz}}{\mathrm{i}\,z}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,zx}\,\mathrm{d}\,\mu_X\,\mathrm{d}\,z=\frac{1}{2\pi}\int_{\mathbb{R}}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}(x-a)z}-\mathrm{e}^{-\mathrm{i}(x-b)z}}{\mathrm{i}\,z}\,\mathrm{d}\,z\,\mathrm{d}\,\mu_X\\
&=\frac{1}{2\pi}\int_{\mathbb{R}}\int_{-T}^{T}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z\,\mathrm{d}\,\mu_X\\
&=\frac{1}{\pi}\int_{\mathbb{R}}\int_{0}^{T}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z\,\mathrm{d}\,\mu_X=\int_{\mathbb{R}}\left(\frac{1}{\pi}\int_{0}^{T}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z\right)\mathrm{d}\,\mu_X\\
&=\mathbb{E}\left[\frac{1}{\pi}\int_{0}^{T}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z\right].
\end{aligned}$$

记

$$H(T,x,a,b)=\frac{1}{\pi}\int_{0}^{T}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z.$$

由 Dirichlet 积分（引理 7.3），

$$
\begin{aligned}
\lim_{T\to\infty} H(T,x,a,b) &= \frac{1}{\pi}\cdot\int_0^{\infty}\frac{\sin(x-a)z-\sin(x-b)z}{z}\,\mathrm{d}\,z = \frac{1}{\pi}\cdot\left(\int_0^{\infty}\frac{\sin(x-a)z}{z}-\int_0^{\infty}\frac{\sin(x-b)z}{z}\,\mathrm{d}\,z\right)\\
&= \frac{1}{\pi}\cdot\frac{\pi}{2}\cdot(\operatorname{sgn}(x-a)-\operatorname{sgn}(x-b)) = \frac{1}{2}\cdot\left(\begin{cases}1, & x>a;\\ 0, & x=a;\\ -1, & x<a\end{cases}-\begin{cases}1, & x>b;\\ 0, & x=b;\\ -1, & x<b\end{cases}\right)\\
&= \begin{cases}0, & x<a;\\ \frac{1}{2}, & x=a;\\ 1, & a<x<b;\\ \frac{1}{2}, & x=b;\\ 0, & x>b.\end{cases}
\end{aligned}
$$

由控制收敛定理，

$$
\begin{aligned}
\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}az}-\mathrm{e}^{-\mathrm{i}bz}}{\mathrm{i}z}\varphi_X(z)\,\mathrm{d}z &= \lim_{T\to\infty} I_T = \lim_{T\to\infty}\mathbb{E}\left[H(T,x,a,b)\right] \\
&= \mathbb{E}\left[\lim_{T\to\infty} H(T,x,a,b)\right] = \mathbb{E}\left[\begin{cases} 0, & x<a; \\ \frac{1}{2}, & x=a; \\ 1, & a<x<b; \\ \frac{1}{2}, & x=b; \\ 0, & x>b. \end{cases}\right] \\
&= 0\cdot\mathbb{P}\left(x<a\right)+\frac{1}{2}\cdot\mathbb{P}\left(X=a\right)+1\cdot\mathbb{P}\left(a<X<b\right)+\frac{1}{2}\cdot\mathbb{P}\left(X=b\right)+0\cdot\mathbb{P}\left(x>b\right) \\
&= \frac{1}{2}\cdot\mathbb{P}\left(X=a\right)+\mathbb{P}\left(a<X<b\right)+\frac{1}{2}\cdot\mathbb{P}\left(X=b\right) = \frac{1}{2}\cdot\mathbb{P}\left(X=a\right)+\mathbb{P}\left(a<X\le b\right)-\mathbb{P}\left(X=b\right)+\frac{1}{2}\cdot\mathbb{P}\left(X=b\right) \\
&= \frac{1}{2}\cdot\mathbb{P}\left(X=a\right)+\mathbb{P}\left(a<X\le b\right)-\frac{1}{2}\cdot\mathbb{P}\left(X=b\right) \\
&= \Phi_X(b)-\Phi_X(a)+\frac{1}{2}\cdot\mathbb{P}\left(X=a\right)-\frac{1}{2}\cdot\mathbb{P}\left(X=b\right).
\end{aligned}
$$

□

命题 7.6 (性质 7.1 6 的反向). 若 X 是对称随机变量，则 $\varphi_X(z)$ 是实值函数。

证明. 由于 X 是对称随机变量，从而 $\varphi_X(z)=\varphi_{-X}(z)$. 又由性质 7.1 3, $\varphi_{-X}(z)=\overline{\varphi_X(z)}$. 综上，

$$\varphi_X(z)=\varphi_{-X}(z)=\overline{\varphi_X(z)}.$$

这就说明 $\varphi_X(z)$ 是实值函数。 □

定理 7.7. X 是一个随机变量。则 $\forall a\in\mathbb{R}$,

$$\mathbb{P}\left(X=a\right)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}az}\cdot\varphi_X(z)\,\mathrm{d}z.$$

证明. 直接计算:

$$\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}\,az}\cdot\varphi_X(z)\,\mathrm{d}\,z=\frac{1}{2T}\int_{-T}^{T}\left(\mathrm{e}^{-\mathrm{i}\,az}\cdot\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,zx}\,\mathrm{d}\,\mu_X\right)\mathrm{d}\,z=\frac{1}{2T}\int_{\mathbb{R}}\left(\int_{-T}^{T}\mathrm{e}^{\mathrm{i}\,z(x-a)}\,\mathrm{d}\,z\right)\mathrm{d}\,\mu_X=\int_{\mathbb{R}}g(x,T)\,\mathrm{d}\,\mu_X.$$

其中，记

$$g(x,T)=\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{\mathrm{i}\,z(x-a)}\,\mathrm{d}\,z=\begin{cases}1, & x=a;\\ \frac{1}{2T}\int_{-T}^{T}\cos z\,(x-a)\,\mathrm{d}\,z, & x\neq a.\end{cases}$$

其中,

$$\lim_{T\to\infty}\left|\frac{1}{2T}\int_{-T}^{T}\cos z\,(x-a)\,\mathrm{d}\,z\right|=\lim_{T\to\infty}\frac{\left|\int_{-T}^{T}\cos z\,(x-a)\,\mathrm{d}\,z\right|}{2T}\leq\lim_{T\to\infty}\frac{2}{2T}=0.$$

故

$$\lim_{T\to\infty}g(x,T)=\begin{cases}1, & x=a;\\ 0, & x\neq a\end{cases}=\chi_{\{a\}}.$$

且关于 T 是一致有界的。由控制收敛定理,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}\,az}\cdot\varphi_X(z)\,\mathrm{d}\,z=\lim_{T\to\infty}\int_{\mathbb{R}}g(x,T)\,\mathrm{d}\,\mu_X=\int_{\mathbb{R}}\lim_{T\to\infty}g(x,T)\,\mathrm{d}\,\mu_X=\int_{\mathbb{R}}\chi_{\{a\}}\,\mathrm{d}\,\mu_X=\mathbb{P}\,(X=a)\,.$$

□

定理 7.8. 设 X 是一个随机变量。若 $\int_{\mathbb{R}}|\varphi_X(z)|\,\mathrm{d}\,z<\infty$, 则 $\Phi_X(x)$ 是绝对连续的。此时，X 有密度函数 $f_X(x)=\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}\,xz}\cdot\varphi_X(z)\,\mathrm{d}\,z$.

注记 7.9. 若 X 有密度函数 $f_X(x)$, 则 $\varphi_X(z)=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,xz}\cdot f_X(x)\,\mathrm{d}\,x$. 从而 $f_X(x)$ 是 $\varphi_X(z)$ 的 *Fourier* 变换，$\varphi_X(z)$ 是 $f_X(x)$ 的 *Fourier* 逆变换。

证明. $\forall a\in\mathbb{R}$,

$$\left|\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}\,az}\cdot\varphi_X(z)\,\mathrm{d}\,z\right|\leq\int_{-T}^{T}\left|\mathrm{e}^{-\mathrm{i}\,az}\cdot\varphi_X(z)\right|\mathrm{d}\,z=\int_{-T}^{T}|\varphi_X(z)|\,\mathrm{d}\,z<\int_{\mathbb{R}}|\varphi_X(z)|\,\mathrm{d}\,z.$$

这说明 $\int_{-T}^{T} \mathrm{e}^{-\mathrm{i}\, az} \cdot \varphi_X(z)\,\mathrm{d}\, z$ 关于 T 一致有界，记 M 是一个上界。由定理 7.7,

$$\mathbb{P}\,(X=a)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}\, az}\cdot\varphi_X(z)\,\mathrm{d}\, z\le\lim_{T\to\infty}\left|\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{-\mathrm{i}\, az}\cdot\varphi_X(z)\,\mathrm{d}\, z\right|\le\lim_{T\to\infty}\frac{|M|}{2T}=0.$$

从而

$$\lim_{x\to a^-}\left(\Phi_X(a)-\Phi_X(x)\right)=\lim_{x\to a^-}\mathbb{P}\,(X=a)=0.$$

这说明 $\Phi_X(x)$ 是左连续的。又由性质 3.10 3, $\Phi_X(x)$ 还是右连续的。从而这就说明 $\Phi_X(x)$ 连续。

又 $\forall a,h\in\mathbb{R}$, 由逆转公式（定理 7.4），由于已证 $\Phi_X(x)$ 连续，再由 Fubini 定理，

$$\begin{aligned}\Phi_X(a+h)-\Phi_X(a)&=\Phi_X(a+h)-\Phi_X(a)+\frac{1}{2}\cdot\mathbb{P}\,(X=a)-\frac{1}{2}\cdot\mathbb{P}\,(X=a+h)\\&=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\, az}-\mathrm{e}^{-\mathrm{i}(a+h)z}}{\mathrm{i}\, z}\varphi_X(z)\,\mathrm{d}\, z=\int_{\mathbb{R}}\frac{\mathrm{e}^{-\mathrm{i}\, az}-\mathrm{e}^{-\mathrm{i}(a+h)z}}{\mathrm{i}\, z}\varphi_X(z)\,\mathrm{d}\, z\\&=\int_{\mathbb{R}}\left(\int_a^{a+h}\mathrm{e}^{-\mathrm{i}\, xz}\,\mathrm{d}\, x\right)\cdot\varphi_X(z)\,\mathrm{d}\, z\\&=\int_a^{a+h}\left(\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}\, xz}\cdot\varphi_X(z)\,\mathrm{d}\, z\right)\mathrm{d}\, x=\int_a^{a+h}f_X(x)\,\mathrm{d}\, x.\end{aligned}$$

由控制收敛定理，$f_X(x)=\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}\, xz}\cdot\varphi_X(z)\,\mathrm{d}\, z$ 是连续的。由 a,h 的任意性，这说明 $\Phi_X(z)$ 有连续的微分 $f_X(x)$, 这就说明 $\Phi_X(x)$ 是绝对连续的。 □

定理 7.10. 设 $X_1,X_2,\cdots$ 是随机变量序列，X 是一个随机变量。若 $X_n\xrightarrow{d}X$, 则在 $\mathbb{R}$ 上有 $\varphi_{X_n}(z)\rightrightarrows\varphi_X(z)$.

证明. $\forall\varepsilon>0$, $\exists A>0$, 使得 $-A$ 与 A 都是 Φ_X 的连续点，且 $\Phi_X(-A)=1-\Phi(A)<\varepsilon$. 此时有 $\Phi_{X_n}(-A)\to\Phi_X(-A)$, $\Phi_{X_n}(A)\to\Phi_X(A)$. 于是 $\exists N>0$, 使得 $\forall n\ge N$, 有 $\Phi_{X_n}(-A)=1-\Phi_{X_n}(A)\le 2\varepsilon$. 此时，$\forall h$,

$$|\varphi_{X_n}(z+h)-\varphi_{X_n}(z)|=\left|\int_{\mathbb{R}}\left(\mathrm{e}^{\mathrm{i}(z+h)x}-\mathrm{e}^{\mathrm{i}\, zx}\right)\mathrm{d}\,\mu_{X_n}\right|\le\int_{\mathbb{R}}\left|\mathrm{e}^{\mathrm{i}(z+h)x}-\mathrm{e}^{\mathrm{i}\, zx}\right|\mathrm{d}\,\mu_{X_n}=\int_{\mathbb{R}}\left|\mathrm{e}^{\mathrm{i}\, xh}-1\right|\mathrm{d}\,\mu_{X_n}=\int_{|z|>A}\left|\mathrm{e}^{\mathrm{i}\, xh}-1\right|\mathrm{d}\,\mu_{X_n}+\int_{|z|\le A}\left|\mathrm{e}^{\mathrm{i}\, xh}-1\right|\mathrm{d}\,\mu_{X_n}$$

其中,

$$\int_{|z|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|\mathrm{d}\,\mu_{X_n}\le\int_{|z|\le A}\max_{|x|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|\mathrm{d}\,\mu_{X_n}=\mathbb{P}\left(|X_n|\le A\right)\cdot\max_{|x|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|\le\max_{|x|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|.$$

于是 $\lim_{h\to0}\int_{|z|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|\mathrm{d}\,\mu_{X_n}\le\lim_{h\to0}\max_{|x|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|=0$, 且这个收敛与 z 无关。

$$\begin{aligned}\int_{|z|>A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|\mathrm{d}\,\mu_{X_n}&\le\int_{|z|>A}\left(\left|\mathrm{e}^{\mathrm{i}\,xh}\right|+1\right)\mathrm{d}\,\mu_{X_n}=\int_{|z|>A}2\,\mathrm{d}\,\mu_{X_n}\\&=2\cdot\mathbb{P}\left(|X|>A\right)=2\cdot\left(\mathbb{P}\left(X>A\right)+\mathbb{P}\left(X<-A\right)\right)\\&=2\cdot\left(1-\Phi_X(A)+\Phi_X(-A)\right)\\&\le2\cdot(2\varepsilon+2\varepsilon)=8\varepsilon.\end{aligned}$$

综上，$\forall z\in\mathbb{R},\forall n\ge N$,

$$\lim_{h\to0}\left|\varphi_{X_n}(z+h)-\varphi_{X_n}(z)\right|\le8\varepsilon+\lim_{h\to0}\max_{|x|\le A}\left|\mathrm{e}^{\mathrm{i}\,xh}-1\right|=8\varepsilon+0=8\varepsilon.$$

这说明 $\varphi_{X_n}(z)$ 在 $\mathbb{R}$ 上等度连续。由定理 4.12, $\forall z\in\mathbb{R}$, 取 $f(x)=\mathrm{e}^{\mathrm{i}\,zx}$, 有 $\varphi_{X_n}(z)=\mathbb{E}\left[f(X_n)\right]\to\mathbb{E}\left[f(X)\right]=\varphi_X(z)$. 这说明 $\varphi_{X_n}(z)\xrightarrow{p.w.}\varphi_X(z)$. 由 Arzela-Ascoli 引理，点点收敛与等度连续就得到一致收敛：在 $\mathbb{R}$ 上有 $\varphi_{X_n}(z)\rightrightarrows\varphi_X(z)$. □

定理 7.11 (Lévy 连续性定理). $X_1,X_2,\cdots$ 是随机变量序列，若有某个函数 $\varphi(z)$, 使得 $\varphi_{X_n}(z)\xrightarrow{p.w.}\varphi(z)$, 且 $\varphi(z)$ 在 0 处连续。则存在随机变量 X, 使得 $\varphi_X(z)=\varphi(z)$, 且 $X_n(z)\xrightarrow{d}X$.

证明. 证明需要用到两个引理:

引理 7.12 (Helly 选择定理). 对任意一族分布函数序列 $\{F_n(x)\}$, 一定存在子列 $\{F_{n_k}(x)\}$ 和一个右连续的单调递增函数 $F(x)$, 使得在 $F(x)$ 的每个连续点 x_0 上有 $\lim_{k\to\infty}F_{n_k}(x_0)=F(x_0)$.

证明. 由于 $\mathbb{Q}$ 是可数的，可记 $\mathbb{Q}=\{x_1,x_2,\cdots\}$.

对 x_1, 由于 $0\le F_n(x_1)\le1$ 有界，从而有子列 $N_1\subset\mathbb{N}$, 使得 $\{F_n(x_1)\}_{n\in N_1}$ 收敛;

对 x_2, 取 $\{F_n(x_2)\}_{n\in N_1}$ 的子列 $N_2\subset N_1$, 使得 $\{F_n(x_2)\}_{n\in N_2}$ 收敛;

$\cdots$

令 $N=\bigcap_{k=1}^{\infty}N_k$, 于是子列 $\{F_n\}_{n\in N}$ 在每个有理点处收敛。记 $\{F_n\}_{n\in N}=\{F_{n_k}\}_{k=1}^{\infty}$. 对任意有理点 x, 记 $F_{n_k}(x)\to F(x)$.

由分布函数的单调性，任意有理数 $x_i<x_j$, $F(x_i)=\lim_{k\to\infty}F_{n_k}(x_i)\le\lim_{k\to\infty}F_{n_k}(x_j)=F(x_j)$. 于是我们可以将 $\mathbb{Q}$ 上的函数 F 延拓成一个 $\mathbb{R}$ 上的右连续的单调递增函数，仍记为 F.

对任意 F 的连续点 x_0, $\forall\varepsilon>0$, 可以找到有理数 $r_1<x_0<r_2$, 使得

$$F(x_0)-\varepsilon\le F(r_1)\le F(r_2)\le F(x_0)+\varepsilon.$$

又 $F_{n_k}(r_1)\to F(r_1)$, $F_{n_k}(r_2)\to F(r_2)$, 从而存在 $K>0$, 使得 $\forall k\ge K$, 有 $F_{n_k}(x_0)\ge F_{n_k}(r_1)\ge F(r_1)-\varepsilon$, $F_{n_k}(x_0)\le F_{n_k}(r_2)\le F(r_2)+\varepsilon$. 综上，此时

$$F_{n_k}(x_0)-F(x_0)=F_{n_k}(x_0)-F(r_2)+F(r_2)-F(x_0)=(F_{n_k}(x_0)-F(r_2))+(F(r_2)-F(x_0))\le\varepsilon+\varepsilon=2\varepsilon.$$

$$F_{n_k}(x_0)-F(x_0)=F_{n_k}(x_0)-F(r_1)+F(r_1)-F(x_0)=(F_{n_k}(x_0)-F(r_1))+(F(r_1)-F(x_0))\ge(-\varepsilon)+(-\varepsilon)=-2\varepsilon.$$

这说明 $|F_{n_k}(x_0)-F(x_0)|\le 2\varepsilon$. 即子列 $\{F_{n_k}(x)\}$ 在任意 $F(x)$ 的连续点 x_0 处都有 $F_{n_k}(x_0)\to F(x_0)$. □

注记 7.13. 此时，$F(x)$ 不一定是一个分布函数。例如

$$F_n(x)=\frac{1}{2}\cdot\chi_{[-n,\infty)}(x)+\frac{1}{2}\cdot\chi_{[n,\infty)}\to\frac{1}{2}.$$

我们想让 F 也是一个分布函数。这要对 $\{F_n(x)\}$ 附加条件：

定义 7.14. 称分布函数序列 $\{F_n(x)\}$ 是紧凑（*tight*）的，若 $\forall\varepsilon>0$, $\exists A,N>0$, 使得 $\forall n\ge N$, 有

$$F_n(-A)+(1-F_n(A))\le\varepsilon.$$

若 $\{F_n(x)\}$ 是紧凑的，则 $F_{n_k}(-A)+(1-F_{n_k}(A))\le\varepsilon$ 就蕴含了 $F(-A)+(1-F(A))\le\lim_{k\to\infty}(F_{n_k}(-A)+(1-F_{n_k}(A)))\le\varepsilon$. 由 ε 的任意性，就有 $F(-\infty)=0$, $F(+\infty)=1$. 否则，$\forall A>0$,

$$F(-A)+(1-F(A))\ge F(-\infty)+(1-F(\infty))\neq 0.$$

除非 $F(-\infty)=F(\infty)=\frac{1}{2}$. 而这由 F 单调递增可以得到 $F\equiv\frac{1}{2}$. 从而 $F_{n_k}\to\frac{1}{2}$. 这与 $\{F_n(x)\}$ 是紧凑的矛盾。于是取 $\varepsilon_0<F(-\infty)+(1-F(\infty))$, 则 $\forall A>0$,

$$F(-A)+(1-F(A))\ge F(-\infty)+(1-F(\infty))>\varepsilon_0.$$

就得到了矛盾。从而由定理 *3.11*, $F(x)$ 是一个分布函数。

引理 7.15. 设 $\varphi_X(z)$ 是随机变量 X 的示性函数。则 $\forall h>0$, 有

$$\Phi_X\left(-\frac{2}{h}\right)+\left(1-\Phi_X\left(\frac{2}{h}\right)\right)=\mathbb{P}\left(|X|\ge\frac{2}{h}\right)\le\frac{1}{h}\cdot\int_{|z|\le h}(1-\varphi_X(z))\,\mathrm{d}\,z.$$

证明. 由 Fubini 定理，

$$\begin{aligned}\frac{1}{h}\cdot\int_{|z|\le h}(1-\varphi_X(z))\,\mathrm{d}\,z&=\frac{1}{h}\cdot\int_{|z|\le h}\int_{\mathbb{R}}\left(1-\mathrm{e}^{\mathrm{i}\,zx}\right)\mathrm{d}\,\mu_X\,\mathrm{d}\,z\\&=\int_{\mathbb{R}}\frac{1}{h}\cdot\int_{-h}^{h}\left(1-\mathrm{e}^{\mathrm{i}\,zx}\right)\mathrm{d}\,z\,\mathrm{d}\,\mu_X=\int_{\mathbb{R}}\frac{1}{h}\cdot\int_{-h}^{h}(1-\cos zx)\,\mathrm{d}\,z\,\mathrm{d}\,\mu_X=2\cdot\int_{\mathbb{R}}\left(1-\frac{\sin hx}{hx}\right)\mathrm{d}\,\mu_X.\end{aligned}$$

又 $|u|\ge 2$ 时，$1-\frac{\sin u}{u}\ge 1-\left|\frac{\sin u}{u}\right|=1-\frac{|\sin u|}{|u|}\ge 1-\frac{1}{2}=\frac{1}{2}$. 从而

$$\begin{aligned}\frac{1}{h}\cdot\int_{|z|\le h}(1-\varphi_X(z))\,\mathrm{d}\,z&=2\cdot\int_{\mathbb{R}}\left(1-\frac{\sin hx}{hx}\right)\mathrm{d}\,\mu_X\\&\ge 2\cdot\int_{|x|\ge\frac{2}{h}}\left(1-\frac{\sin hx}{hx}\right)\mathrm{d}\,\mu_X\\&\ge 2\cdot\int_{|x|\ge\frac{2}{h}}\frac{1}{2}\,\mathrm{d}\,\mu_X=\int_{|x|\ge\frac{2}{h}}\mathrm{d}\,\mu_X=\mathbb{P}\left(|X|\ge\frac{2}{h}\right)=\Phi_X\left(-\frac{2}{h}\right)+\left(1-\Phi_X\left(\frac{2}{h}\right)\right).\end{aligned}$$

□

下面我们回到 Lévy 连续性定理的证明：先证明 φ 也是某个随机变量的示性函数。

首先证明 $\{\Phi_{X_n}(x)\}$ 是紧凑的。$\forall \varepsilon > 0$,

$$\frac{1}{h}\cdot\int_{|z|\le h}(1-\varphi_{X_n}(z))\,\mathrm{d}\,z \le \frac{1}{h}\cdot\int_{|z|\le h}|(1-\varphi(z))+(\varphi(z)-\varphi_{X_n}(z))|\,\mathrm{d}\,z \le \frac{1}{h}\cdot\int_{|z|\le h}|1-\varphi(z)|\,\mathrm{d}\,z+\frac{1}{h}\cdot\int_{|z|\le h}|\varphi_{X_n}(z)-\varphi(z)|\,\mathrm{d}\,z.$$

由于 $\varphi_{X_n}(z) \xrightarrow{p.w.} \varphi_X(z)$, 于是 $\varphi(0) = \lim_{n\to\infty}\varphi_{X_n}(0) = \lim_{n\to\infty}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,0X_n}\right] = 1$. 且 $\varphi(z)$ 在 0 处连续，从而存在 $h > 0$, 使得 $\forall |z| \le h$, 有 $|1-\varphi(z)| = |\varphi(0)-\varphi(z)| \le \varepsilon$. 从而此时

$$\frac{1}{h}\cdot\int_{|z|\le h}|1-\varphi(z)|\,\mathrm{d}\,z \le \frac{1}{h}\cdot\int_{|z|\le h}\varepsilon\,\mathrm{d}\,z = \frac{1}{h}\cdot 2h\cdot\varepsilon = 2\varepsilon.$$

由于 $\varphi_{X_n}(z) \xrightarrow{p.w.} \varphi_X(z)$, 且示性函数在 $|z| \le h$ 时有界，由控制收敛定理，$\lim_{n\to\infty}\frac{1}{h}\cdot\int_{|z|\le h}|\varphi_{X_n}(z)-\varphi(z)|\,\mathrm{d}\,z \to 0$. 从而存在 $N > 0$, 使得 $\forall n \ge N$, 有 $\frac{1}{h}\cdot\int_{|z|\le h}|\varphi_{X_n}(z)-\varphi(z)|\,\mathrm{d}\,z \le \varepsilon$. 由引理 7.15,

$$\Phi_{X_n}\left(-\frac{2}{h}\right)+\left(1-\Phi_{X_n}\left(\frac{2}{h}\right)\right) \le \frac{1}{h}\cdot\int_{|z|\le h}(1-\varphi_{X_n}(z))\,\mathrm{d}\,z \le \frac{1}{h}\cdot\int_{|z|\le h}|1-\varphi(z)|\,\mathrm{d}\,z+\frac{1}{h}\cdot\int_{|z|\le h}|\varphi_{X_n}(z)-\varphi(z)|\,\mathrm{d}\,z \le 2\varepsilon+\varepsilon = 3\varepsilon.$$

取 $A = \frac{2}{h}$, 于是 $\forall \varepsilon > 0$, 我们找到了 $A, N > 0$, 使得 $\forall n \ge N$, 都有

$$\Phi_{X_n}(-A)+(1-\Phi_{X_n}(A)) = \Phi_{X_n}\left(-\frac{2}{h}\right)+\left(1-\Phi_{X_n}\left(\frac{2}{h}\right)\right) \le 3\varepsilon.$$

即 $\{\Phi_{X_n}(x)\}$ 是紧凑的（定义 7.14）。

由 Helly 选择定理（引理 7.12），存在 $\{\Phi_{X_n}(x)\}$ 的子列 $\left\{\Phi_{X_{n_k}}(x)\right\}$ 与某个随机变量 X, 使得在 $\Phi_X(x)$ 的连续点处，都有 $\left\{\Phi_{X_{n_k}}(x)\right\}$ 收敛到 $\Phi_X(x)$. 由于示性函数只与分布有关，所以这蕴含了 $\varphi_{X_{n_k}}(z) \xrightarrow{p.w.} \varphi_X(z)$. 而 $\varphi_{X_n}(z) \xrightarrow{p.w.} \varphi(z)$ 说明 $\varphi_{X_{n_k}}(z) \xrightarrow{p.w.} \varphi(z)$. 从而 $\varphi_X(z) = \varphi(z)$.

只剩下证 $X_n \xrightarrow{d} X$. 在实分析中我们知道，$X_n \xrightarrow{d} X$ 当且仅当 $\{X_n\}$ 的任意子列都有子列依分布收敛到 X. 于是任意 $\{X_n\}$ 的子列 $\{X_{n_k}\}$, 它仍是紧凑的，由 Helly 选择定理（引理 7.12），存在 $\left\{\Phi_{X_{n_k}}(x)\right\}$ 的子列 $\left\{\Phi_{X_{n_{k_i}}}(x)\right\}$ 与某个随机变量 Y, 使得在 $\Phi_Y(x)$ 的连续点处，都有 $\left\{\Phi_{X_{n_{k_i}}}(x)\right\}$ 收敛到 $\Phi_Y(x)$, 即 $X_{n_{k_i}} \xrightarrow{d} Y$. 由于示性函数只与分布有关，所以这蕴含了 $\varphi_{X_{n_{k_i}}}(z) \xrightarrow{p.w.} \varphi_Y(z)$. 而 $\varphi_{X_n}(z) \xrightarrow{p.w.} \varphi(z)$ 说明 $\varphi_{X_{n_{k_i}}}(z) \xrightarrow{p.w.} \varphi(z) = \varphi_X(z)$. 从而 $\varphi_Y(z) = \varphi(z) = \varphi_X(z)$. 而逆转公式（定理 7.4）告诉我们，随机变量的示性函数也决定了分布（注记 7.5），这说明 $Y \overset{d}{=} X$. 从而 $X_{n_{k_i}} \xrightarrow{d} Y \overset{d}{=} X$, 即 $X_{n_{k_i}} \xrightarrow{d} X$. 由本段开始的等价命题的断言，这就证明了 $X_n \xrightarrow{d} X$. □

定理 7.16 (中心极限定理)**.** 设 $X_1, X_2, \cdots$ 是一列独立同分布的随机变量。$\mathbb{E}[X_i] = \mu$, $\sigma^2(X_i) = \sigma^2$. 令 $S_n = \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^n X_i - n\mu\right)$. 则 $S_n \xrightarrow{d} \mathcal{N}(0,1)$.

证明. 通过平移，不妨设 $\mu = 0$, $\sigma^2 = 1$, 否则用 $\frac{X_i - \mu}{\sigma}$ 代替 X_i. 此时，$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. 不妨记 $X_1, X_2, \cdots$ 同分布于 X.

计算 S_n 的示性函数。由性质 7.1 5,

$$\varphi_{S_n}(z) = \prod_{i=1}^n \varphi_{\frac{X_i}{\sqrt{n}}}(z) = \left(\varphi_{\frac{X}{\sqrt{n}}}(z)\right)^n.$$

由 Laurent 展开，

$$\mathrm{e}^{\mathrm{i}\,x} = 1 + \mathrm{i}\,x + \frac{\mathrm{i}^2 x^2}{2!} + \cdots = 1 + \mathrm{i}\,x - \frac{x^2}{2} + R(x).$$

其中，$R(x)$ 是余项，满足 $|R(x)| \le \min\left\{\frac{|x|^3}{6}, |x|^2\right\}$. (习题)

$$\varphi_{\frac{X}{\sqrt{n}}}(z) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,z\frac{X}{\sqrt{n}}}\,\mathrm{d}\,\mu_X = \int_{\mathbb{R}} \left(1 + \mathrm{i}\,\frac{zx}{\sqrt{n}} - \frac{z^2x^2}{2n} + R\left(\frac{zx}{\sqrt{n}}\right)\right)\mathrm{d}\,\mu_X.$$

其中，$\int_{\mathbb{R}} 1\,\mathrm{d}\,\mu_X = 1$, $\int_{\mathbb{R}} x\,\mathrm{d}\,\mu_X = 0$, $\int_{\mathbb{R}} x^2\,\mathrm{d}\,\mu_X = \sigma^2(X) + \mathbb{E}[X] = 1 + 0 = 1$. 从而，

$$\varphi_{\frac{X}{\sqrt{n}}}(z) = 1 - \frac{z^2}{2n} + \int_{\mathbb{R}} R\left(\frac{zx}{\sqrt{n}}\right)\mathrm{d}\,\mu_X.$$

我们来估计最后一项积分。$\forall \varepsilon > 0$, $\exists A > 0$, 使得 $\int_{|X|\ge A} x^2\,\mathrm{d}\,\mu_X \le \varepsilon$. 记 $r_n = 2n\int_{\mathbb{R}} R\left(\frac{zx}{\sqrt{n}}\right)\mathrm{d}\,\mu_X$.

$$\begin{aligned}
r_n = 2n\int_{\mathbb{R}} R\left(\frac{zx}{\sqrt{n}}\right)\mathrm{d}\,\mu_X &= 2n\left(\int_{|X|<A} R\left(\frac{zx}{\sqrt{n}}\right)\mathrm{d}\,\mu_X + \int_{|X|\ge A} R\left(\frac{zx}{\sqrt{n}}\right)\mathrm{d}\,\mu_X\right) \\
&\le 2n\int_{|X|<A} \frac{|z|^3 \cdot |x|^3}{6n\sqrt{n}}\,\mathrm{d}\,\mu_X + 2n\int_{|X|\ge A} \frac{z^2x^2}{n}\,\mathrm{d}\,\mu_X \\
&\le 2n\int_{|X|<A} \frac{|z|^3 \cdot A^3}{6n\sqrt{n}}\,\mathrm{d}\,\mu_X + 2n\int_{|X|\ge A} \frac{z^2x^2}{n}\,\mathrm{d}\,\mu_X = 2n \cdot \frac{|z|^3 \cdot A^3}{6n\sqrt{n}} \cdot \int_{|X|<A} \mathrm{d}\,\mu_X + 2n \cdot \frac{z^2}{n} \cdot \int_{|X|\ge A} x^2\,\mathrm{d}\,\mu_X \\
&\le \frac{|z|^3 \cdot A^3}{3\sqrt{n}} \cdot 1 + 2z^2\varepsilon.
\end{aligned}$$

对固定的 $z \in \mathbb{R}$, $\limsup_{n\to\infty} |r_n| \le 2z^2\varepsilon$. 由 ε 的任意性，$\lim_{n\to\infty}|r_n| = 0$. 从而由 $\mathrm{e}^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$, 从而

$$\left(\varphi_{\frac{X}{\sqrt{n}}}(z)\right)^n = \left(1 + \frac{-z^2 + r_n}{2n}\right)^n \to \mathrm{e}^{\frac{1}{2}\left(-z^2+\lim_{n\to\infty} r_n\right)} = \mathrm{e}^{-\frac{1}{2}z^2}.$$

即 $\varphi_{S_n}(z) \xrightarrow{p.w.} \mathrm{e}^{-\frac{1}{2}z^2}$. 由 Lévy 连续性定理（定理 7.11），存在随机变量 Y, 使得 $\varphi_Y(z) = \mathrm{e}^{-\frac{1}{2}z^2}$, 且 $S_n \xrightarrow{d} Y$. 由于示性函数决定了分布函数（注记 7.5），于是 $Y \sim \mathcal{N}(0,1)$ 服从标准正态分布。这就证明了 S_n 依分布收敛于一个标准正态分布 $Y \sim \mathcal{N}(0,1)$. □

8 随机过程

定义 8.1. 固定概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$. 参数集 $T \subset [0, +\infty)$. $\forall t \in T$, 有一个随机变量 X_t. $\{X_t | t \in T\}$ 称为一个随即过程。若 $T = \mathbb{N}$, 则称其为一个离散过程；否则称为连续过程。

定义 8.2. 对离散过程 $X_1, X_2, \cdots$, $\forall \omega \in \Omega$, 称 $X_1(\omega), X_2(\omega), \cdots$ 是一个样本轨道。记 $S_n = X_1 + \cdots X_n$. 则 $S_1, S_2, \cdots$ 也是一个离散过程，称为一个随机游动。

例 8.3. $\Omega = \{-1, 1\}^{\mathbb{N}}$. 取 $\omega = (1, 1, -1, -1, 1, \cdots)$. 则 $S_1(\omega) = 1, S_2(\omega) = 2, S_3(\omega) = 1, \cdots$. 可以问：样本轨道是否无限次接近 0? 是否无限次接近无穷？

定义 8.4. 由于 X_t 给出了 $\mathbb{R}$ 上的测度 $\mu_{X_t} = \mu_t$. 于是我们不妨在概率空间 $(\mathbb{R}, \mu_t, \mathbb{P}_t), t \in T$ 上考虑问题。即 $\Omega = \mathbb{R}^T = \{\lambda : T \to \mathbb{R}\}$, $\forall t \in T$, $\mathbb{P}_t : \Omega \to \mathbb{R}$, $\mathbb{P}_t(\lambda) = \lambda(t)$.

定义 8.5. 可以在有限个时间上做：设 $t_1, \cdots, t_n \in T$.

$$\begin{aligned} \mathbb{P}_{t_1, \cdots, t_n} : \Omega &\to \mathbb{R}^n \\ \lambda &\mapsto (\lambda(t_1), \cdots, \lambda(t_n)) . \end{aligned}$$

取长方体 $\prod_{i=1}^n (a_i, b_i] \subset \mathbb{R}^n$, $\mathbb{P}^{-1}_{t_1, \cdots, t_n} \left(\prod_{i=1}^n (a_i, b_i]\right) \subset \Omega$ 称为 Ω 中的长方体。

性质 8.6. *1.* 长方体的交还是长方体。从而 Ω 上的所有长方体构成了一个 π-系。

2. 所有有限个长方体的并集的集合构成一个 *Boole*-代数 $\mathcal{A}$. 在 $\mathcal{A}$ 上定义测度 m:

$$m\left(\mathbb{P}^{-1}_{t_1, \cdots, t_n}\left(\prod_{i=1}^n (a_i, b_i]\right)\right) = \prod_{i=1}^n \mu_{t_i}\left((a_i, b_i]\right).$$

证明. 1. 是容易的，只证 2. Boole-代数上的测度 m 要满足三个条件:

(a) $m(A) \geq 0$.

(b) 若 $A \cap B = \varnothing$, 则 $m(A \cup B) = m(A) + m(B)$.

(c) 若 $A_n \searrow \varnothing$, 则 $m(A_n) \to 0$.

前两条只涉及有限维，容易验证。只证 (c). 同样地，每个 A_n 只涉及有限个 $t \in T$, 于是全部 A_n 只涉及可数个 $t \in T$. 于是不妨设 $T = \mathbb{N}$. 记 $\Omega = \mathbb{R}^{\mathbb{N}}$, 设 $\mathcal{F}_n$ 是 $\mathbb{R}^n$ 上所有长方体 $\prod_{i=1}^n (a_i, b_i]$ 生成的 Boole-代数，则 $\left\{\mathbb{P}_{t_1,\cdots,t_n}^{-1}(A^*) \middle| A^* \in \mathcal{F}_n\right\} = \mathcal{A}_n \subset \mathcal{A}$ 是所有 $\mathbb{P}_{t_1,\cdots,t_n}^{-1}\left(\prod_{i=1}^n (a_i, b_i]\right)$ 生成的 Boole-代数，有 $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$.

用反证法：若否，设 $m(A_n) \to 2\delta > 0$. 可以看出，重复插入一些 A_n, 可以假设 $A_n = \mathbb{P}_{t_1,\cdots,t_n}^{-1}(A_n^*) \in \mathcal{A}_n$, $A_n^* \in \mathcal{F}_n$. $\forall n$, 存在 $B_n^* \in \mathcal{F}_n$, 使得它的闭包 $\mathrm{cl}(B_n^*) \subset A_n^*$ 是紧集，且 $m(A_n^* - B_n^*) \le \frac{\delta}{2^n}$, 从而 $m\left(A_n - \mathbb{P}_{t_1,\cdots,t_n}^{-1}(B_n^*)\right) \le \frac{\delta}{2^n}$.

记 $B_n = \mathbb{P}_{t_1,\cdots,t_n}^{-1}(B_n^*)$. 令 $C_n = \bigcap_{i=1}^n B_i$. 从而有 $C_n = \mathbb{P}_{t_1,\cdots,t_n}^{-1}(C_n^*)$, 其中 $C_n^* = \bigcap_{i=1}^n (B_n^* \times \mathbb{R}^{n-i}) \subset B_n^*$. 则 $\mathrm{cl}(C_n^*) \subset \mathrm{cl}(C_n^*) \subset A_n^*$ 是紧的，从而 $\mathrm{cl}(C_n) \subset A_n$, 从而 $\mathrm{cl}(C_n) \searrow \varnothing$. 注意到有 $A_n - C_n \subset \bigcup_{i=1}^n (A_i - B_i)$, 于是有

$$m(A_n - C_n) \le m\left(\bigcup_{i=1}^n (A_i - B_i)\right) \le \sum_{i=1}^n m(A_i - B_i) \le \sum_{i=1}^n \frac{\delta}{2^n} < \delta.$$

从而 $m(A_n) - m(C_n) \le m(A_n - C_n) \le \delta$, 从而 $m(C_n) \ge m(A_n) - \delta \ge 2\delta - \delta = \delta > 0$. 这说明 $\mathrm{cl}(C_n)$ 非空.

这允许我们对任意 n, 取 $\lambda_n : \mathbb{N} \to \mathbb{R} \in C_n$. 有 $\lambda_n(1) \in C_n^* \subset C_1^* \subset \mathrm{cl}(C_1^*)$. 由于 $\mathrm{cl}(C_1^*)$ 是紧集，于是 $\{\lambda_n(1)\}$ 有一个收敛子列 $\lambda_{n_j}(1) \to \theta(1) \in \mathrm{cl}(C_1^*)$. 记 $\left\{\lambda_j^2\right\} = \left\{\lambda_{n_j}\right\}$. 同理有 $\{\lambda_n^2(1), \lambda_n^2(2)\} \in C_2^*$ 是一个紧集，从而有收敛子列，设这个收敛子列收敛到 $(\theta(1), \theta(2)) \in \mathrm{cl}(C_2^*)$. 重复这个操作，由对角线法则，取对角线子列记为 $\{\lambda_{n_k}\}$，就得到了事件 $\theta : \mathbb{N} \to \mathbb{R} \in \mathbb{R}^{\mathbb{N}}$, 使得 $\forall t \in \mathbb{N}$, $\lambda_{n_k}(t) \to \theta(t)$. 从而 $\forall n$, 有 $\mathbb{P}_{1,\cdots,n}(\theta) = (\theta(1), \cdots, \theta(n)) \leftarrow (\lambda_{n_k}(1), \cdots, \lambda_{n_k}(n)) \in \mathrm{cl}(C_n^*)$, 由于 $\mathrm{cl}(C_n^*)$ 是紧集，从而 $\mathbb{P}_{1,\cdots,n}(\theta) \in \mathrm{cl}(C_n^*)$, 从而 $\theta \in \mathrm{cl}(C_n)$. 从而 $\theta \in \bigcap_{n=1}^\infty \mathrm{cl}(C_n)$. 这说明 $C_n \searrow \bigcap_{n=1}^\infty \mathrm{cl}(C_n) \neq \varnothing$, 就与 $C_n \searrow \varnothing$ 矛盾.。 □

定理 8.7 (Carathéodory 扩张)**.** *$(\mathcal{A}, m)$ 可以扩张到 $(\sigma(\mathcal{A}), m^*)$.*

定义 8.8. *由定理 8.7, 我们得到了概率空间 $(\mathbb{R}^T, \sigma(\mathcal{A}), m^*)$. 此时 X_t 可以看成这个概率空间上的随机变量.*

定义 8.9. *有时我们不必要有 $X_t : \Omega \to \mathbb{R}$, 可以把 $\mathbb{R}$ 换成任意一个度量空间 S, 比如 $\mathbb{R}^d$. 此时 X_t 称为一个广义随机变量.*

定义 8.10. 同理定理 *2.20* 中的定义，设 $X_1, X_2, \cdots$ 是一列独立的随机变量序列，定义

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\left(X_n, X_{n+1}, \cdots\right),$$

称为尾事件代数，其中的事件称为尾事件。

我们有类似的 Kolmogorov 0-1 律：

定理 8.11. $\forall A \in \mathcal{T}, \mathbb{P}\left(A\right) \in \{0, 1\}$.

它有缺陷：

例 8.12. 设 $\Omega = \{1, -1\}^{\mathbb{N}}$，$X_1, X_2, \cdots$ 是一列独立的随机变量序列，$\mathbb{P}\left(X_n = 1\right) = p$, $\mathbb{P}\left(X_n = -1\right) = 1 - p$. 记 $S_n = X_1 + \cdots + X_n$. 则 $A = \{\omega|$ 有无穷多个 n, 使得 $S_n(\omega) \neq 0\}$ 不是一个尾事件，但我们也想得到 $\mathbb{P}\left(A\right) \in \{0, 1\}$.

定义 8.13. 设 $S = \mathbb{R}^d$, $\Omega = S^{\mathbb{N}}$, $X_1, X_2, \cdots$ 是一列独立同分布的取值在 S 中的随机变量序列。设 $\pi : \mathbb{N} \to \mathbb{N}$ 是一个置换, 称它是有限置换, 若只有有限多个 $i \in \mathbb{N}$, 使得 $\pi(i) \neq i$. 有限置换 π 诱导了两个映射

$$\begin{aligned} \pi : \Omega &\to \Omega \\ \omega = \left(\omega_1, \omega_2, \cdots\right) &\mapsto \omega \circ \pi = \left(\omega_{\pi(1)}, \omega_{\pi(2)}, \cdots\right) \end{aligned}$$

与

$$\begin{aligned} \pi^{-1} : \mathcal{F} &\to \mathcal{F} \\ A &\mapsto \left\{\omega \circ \pi^{-1} \middle| \omega \in A\right\}. \end{aligned}$$

称一个事件 $A \in \mathcal{F}$ 是可置换的，若对任意有限置换 $\pi : \mathbb{N} \to \mathbb{N}$, 有 $\pi^{-1}(A) = A$.

例 8.14. $A = \{\omega|$ 有无穷多个 n, 使得 $S_n(\omega) \neq 0\}$ 是一个可置换事件。对任意有限置换 π, $\omega \in A$ 当且仅当 $\omega \circ \pi \in A$.

注记 8.15. 尾事件一定是可置换事件。

定理 8.16 (Hewitt-Savage 0-1 律)**.** 设 $X_1, X_2, \cdots$ 是一列独立同分布的取值在 $S = \mathbb{R}^d$ 中的随机变量序列，$\Omega = S^{\mathbb{N}}$, $\mathcal{F}$ 是 Ω 上的 σ-代数。若 $A \in \mathcal{F}$ 是可置换事件，则 $\mathbb{P}(A) \in \{0, 1\}$.

证明. 同理 Kolmogorov 0-1 律（定理 2.20）的证明，我们将证明 $\mathbb{P}(A) = \mathbb{P}(A)^2$.

$\forall \varepsilon > 0$, 可以取一列 $\{A_i\} \subset \bigcup_{n=1}^{\infty} \mathcal{F}_n$, 使得 $A \subset \bigcup_{i=1}^{\infty} A_i$, 且满足 $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i - A) \leq \varepsilon$, 其中 $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. 记 $E = \bigcup_{i=1}^{\infty} A_i$, 有 $\mathbb{P}(E) = \mathbb{P}((E - A) \sqcup A) \leq \mathbb{P}(E - A) + \mathbb{P}(A) \leq \varepsilon + \mathbb{P}(A)$. 并且可以取充分大的 k, 使得对事件 $B = \bigcup_{i=1}^{k} A_i \in \mathcal{F}_n$, 有 $\mathbb{P}(E - B) \leq \varepsilon$. 此时，有

$$\mathbb{P}(B - A) \leq \mathbb{P}(E - A) \leq \varepsilon, \qquad \mathbb{P}(A - B) \leq \mathbb{P}(E - B) \leq \varepsilon.$$

从而

$$\mathbb{P}(B \triangle A) = \mathbb{P}((B - A) \cup (A - B)) \leq \mathbb{P}(B - A) + \mathbb{P}(A - B) \leq \varepsilon + \varepsilon = 2\varepsilon.$$

这说明我们可以找到一个事件序列 $\{B_n\}$, $B_n \in \mathcal{F}_n$, 使得 $\mathbb{P}(B_n \triangle A) \to 0$. 又由于

$$|\mathbb{P}(B_n) - \mathbb{P}(A)| \leq \mathbb{P}(B_n - A) \leq \mathbb{P}(B_n \triangle A) \leq 2\varepsilon,$$

从而此时也有 $\mathbb{P}(B_n) \to \mathbb{P}(A)$.

构造一个有限置换

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n & n+1 & \cdots & 2n & 2n+1 & 2n+2 & \cdots \\ n+1 & n+2 & \cdots & 2n & 1 & \cdots & n & 2n+1 & 2n+2 & \cdots \end{pmatrix}.$$

记 $B_n' = \pi(B_n)$, 则 B_n 与 B_n' 相互独立. 并且我们还注意到，由于 A 是可置换事件，有 $\pi(B_n \triangle A) = \pi(B_n) \triangle \pi(A) = B_n' \triangle A$. 由于 π 保持概率测度不变，从而也有 $\mathbb{P}(B_n' \triangle A) = \mathbb{P}(\pi(B_n \triangle A)) = \mathbb{P}(B_n \triangle A) \to 0$. 同理由 $|\mathbb{P}(B_n') - \mathbb{P}(A)| \leq \mathbb{P}(B_n' \triangle A)$, 有 $\mathbb{P}(B_n') \to \mathbb{P}(A)$.

最后，注意到

$$\begin{aligned} |\mathbb{P}(B_n) - \mathbb{P}(B_n \cap B_n')| = \mathbb{P}(B_n - B_n') &\leq \mathbb{P}(B_n \triangle B_n') \\ &= \mathbb{P}((B_n \triangle A) \triangle (B_n' \triangle A)) = \mathbb{P}([(B_n \triangle A) - (B_n' \triangle A)] \cup [(B_n' \triangle A) - (B_n \triangle A)]) \\ &\leq \mathbb{P}((B_n \triangle A) - (B_n' \triangle A)) + \mathbb{P}((B_n' \triangle A) - (B_n \triangle A)) \\ &\leq \mathbb{P}(B_n \triangle A) + \mathbb{P}(B_n' \triangle A) \to 0. \end{aligned}$$

从而 $\mathbb{P}(B_n \cap B'_n) \to \mathbb{P}(B_n)$. 从而由 B_n 与 B'_n 相互独立，就有

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}(B_n \cap B'_n) = \lim_{n\to\infty} \mathbb{P}(B_n)\mathbb{P}(B'_n) = \mathbb{P}(A)^2 .$$

这就说明 $\mathbb{P}(A) \in \{0,1\}$. □

定理 8.17 ($\mathbb{R}^1$ 上的随机游动). *设 $X_1, X_2, \cdots$ 是取值于 $\mathbb{R}^1$ 的独立同分布的随机变量，记 $S_n = x_1 + \cdots + X_n$, 即 $\{S_n\}$ 是 $\mathbb{R}$ 上的随机游动。则以下事件中恰有一个概率为 1:*

(1) $S_n \overset{a.e.}{\equiv} 0$.

(2) $S_n \xrightarrow{a.e.} \infty$.

(3) $S_n \xrightarrow{a.e.} -\infty$.

(4) $-\infty = \liminf_{n\to\infty} S_n < \limsup_{n\to\infty} S_n = \infty$.

证明. 注意到 $\forall c \in [-\infty, +\infty]$, $\{\limsup_{n\to\infty} S_n \geq c\}$ 与 $\{\limsup_{n\to\infty} S_n \leq c\}$ 都是可置换事件,从而 $\{\limsup_{n\to\infty} S_n = c\} = \{\limsup_{n\to\infty} S_n \geq c\} \cap \{\limsup_{n\to\infty} S_n \leq c\}$ 也是可置换事件。于是若 $c \in (-\infty, +\infty)$, 使得 $\mathbb{P}(\limsup_{n\to\infty} S_n = c) = 1$, 注意到 $S'_n = S_{n+1} - X_1$ 与 S_n 同分布，从而 $\limsup_{n\to\infty} S'_n \overset{a.e.}{=} c$, $\limsup_{n\to\infty}(X_1 + S'_n) \overset{a.e.}{=} c$. 从而 $\limsup_{n\to\infty} X_1 \overset{a.e.}{=} 0$. 同理，将 $= c$ 改为 $\geq c$, $\leq c$, 将 $\limsup$ 改为 $\liminf$ 也都成立。这就说明 $X_1 \overset{a.e.}{0}$. 从而 $S_n \overset{a.e.}{=} 0$. 这就是 (1).

剩下的就是 $\mathbb{P}(\limsup_{n\to\infty} S_n = \infty) = 1$ 或 $\mathbb{P}(\limsup_{n\to\infty} S_n = -\infty) = 1$ 以及 $\mathbb{P}(\liminf_{n\to\infty} S_n = \infty) = 1$ 或 $\mathbb{P}(\liminf_{n\to\infty} S_n = -\infty) = 1$ 的情况。

- 若 $\mathbb{P}(\limsup_{n\to\infty} S_n = \infty) = 1$ 且 $\mathbb{P}(\liminf_{n\to\infty} S_n = \infty) = 1$. 从而 $\mathbb{P}(\lim_{n\to\infty} S_n = \infty) = 1$, 这就是 (2).
- 若 $\mathbb{P}(\limsup_{n\to\infty} S_n = \infty) = 1$ 且 $\mathbb{P}(\liminf_{n\to\infty} S_n = -\infty) = 1$. 这就是 (4).
- 若 $\mathbb{P}(\limsup_{n\to\infty} S_n = -\infty) = 1$ 且 $\mathbb{P}(\liminf_{n\to\infty} S_n = \infty) = 1$, 此时有 $\liminf_{n\to} > \limsup_{n\to}$, 这不可能。
- 若 $\mathbb{P}\big(\limsup_{n\to-\infty} S_n = \infty\big) = 1$ 且 $\mathbb{P}(\liminf_{n\to\infty} S_n = -\infty) = 1$ 从而 $\mathbb{P}(\lim_{n\to\infty} S_n = -\infty) = 1$, 这就是 (3).

□

例 8.18 (简单随机游动)**.** 设 $X_1, X_2, \cdots$ 是取值于 $\mathbb{R}^1$ 的独立同分布的随机变量，$\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = -1) = 1 - p$. 记 $S_n = x_1 + \cdots + X_n$, 即 $\{S_n\}$ 是 $\mathbb{R}$ 上的随机游动。

- 若 $p = \frac{1}{2}$, 则称为对称随机游动。
- 若 $p \neq \frac{1}{2}$, 则称为非对称随机游动。

此时定理 *8.17* 中的 *(1)* 此时不可能发生。

- 若 $p < \frac{1}{2}$. 设 $u > 0$ 是一个待定参数。由 *Markov* 不等式（定理 *5.1*）与 X_i 之间的独立性，

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}\left(\mathrm{e}^{uS_n} \geq 1\right) \leq \mathbb{E}\left[\mathrm{e}^{uS_n}\right] = \mathbb{E}\left[\mathrm{e}^{uX_1}\right]^n = \left(p\,\mathrm{e}^u + (1-p)\,\mathrm{e}^{-u}\right)^n.$$

  取 $u > 0$ 使得 $p\,\mathrm{e}^u + (1-p)\,\mathrm{e}^{-u}$ 取到最小，此时有 $\mathrm{e}^u = \sqrt{\frac{1-p}{p}} > 1$. 于是

$$\mathbb{P}(S_n \geq 0) \leq \left(2\sqrt{p(q-p)}\right)^n.$$

  由于 $p < \frac{1}{2}$, 于是 $2\sqrt{p(q-p)} < 2\sqrt{\frac{1}{4}} = 1$. 从而

$$\sum_{n=1}^{\infty} \mathbb{P}(S_n \geq 0) \leq \sum_{n=1}^{\infty} \left(2\sqrt{p(q-p)}\right)^n < \infty.$$

  由 *Borel-Cantelli* 定理（定理 *2.16*），$\mathbb{P}(\limsup_{n\to\infty} \{S_n \geq 0\}) = 0$. 只能是定理 *8.17* 中的 *(3)*.

- 若 $p > \frac{1}{2}$. 同理，只能是定理 *8.17* 中的 *(2)*.
- 若 $p = \frac{1}{2}$. 此时是对称随机游动，S_n 与 $-S_n$ 同分布，从而不可能是定理 *8.17* 中的 *(2)(3)*, 只能是 *(4)*. 这说明此时 S_n 可以无穷次取到任意一个整数。

定义 8.19 (停时)**.** 设 $X_1, X_2, \cdots$ 是取值于 $S=\mathbb{R}^d$ 的独立同分布的随机变量，记 $S_n = x_1 + \cdots + X_n$, 即 $\{S_n\}$ 是 $\mathbb{R}$ 上的随机游动。设 $\tau : \Omega = S^{\mathbb{N}} \to \mathbb{N} \cap \{\infty\}$ 是一个随机变量，满足 $\forall n \in \mathbb{N}$, 事件 $\{\tau = n\} \in \mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, 则称随机变量 τ 是一个停时。

例 8.20. 取一个 *Borel* 可测集 $B \subset \mathbb{R}^d$, 令 $\tau_B = \inf\{n | S_n \in B\}$. 这个停时称为 B 的首达。有

$$\{\tau_B = n\} = \{S_n \in B\} \cap \left(\bigcap_{i<n} \{S_i \in B^c\}\right) \in \mathcal{F}_n.$$

它确实是 $\mathcal{F}_n$ 中的一个事件。

例 8.21 (赌徒模型)**.** 设赌徒的资金为 m, 每注 1, 赢的概率为 p, 输的概率为 $1-p$. 赌 n 次后赢的总数就是一个简单随机游动（例 *8.18*）。则赌徒输光的时刻 $\tau_{-m} = \inf\{S_n | S_n = -m\}$ 就是一个停时。

当 $p \le \frac{1}{2}$ 时，由例 *8.18*, $\mathbb{P}(\liminf_{n\to\infty} S_n = -\infty) = 1$, 于是 $\mathbb{P}(\exists n, S_n = -m) = 1$, 即 $\mathbb{P}(\tau_{-m} < \infty) = 1$. 这说明此时赌徒一定会输光。

考虑公平赌博 $p = \frac{1}{2}$. 我们设定 $M > 0$, 当赌徒赢到这个值的时候就停止赌博，这也是一个停时 τ_M。此时有 $\mathbb{P}(\limsup_{n\to\infty} S_n = \infty) = 1$, 从而 $\mathbb{P}(\tau_M) = 1$. 这说明如果赌徒可以借钱的话，他总能赢到 M.

我们不让赌徒借钱。考虑 $B = \{-m, M\}$, 此时 $\tau = \tau_{-m,M}$. 此时有 $\mathbb{P}(\tau < \infty) = 1$. 我们想计算 $\mathbb{P}(S_\tau = -m)$ 与 $\mathbb{P}(S_\tau = M)$.

定理 8.22 (Wald 第一方程)**.** 设 $X_1, X_2, \cdots$ 是取值在 $\mathbb{R}^1$ 上的独立同分布的随机变量，$\mathbb{E}[|X_1|] < \infty$. 设 τ 是一个停时，$\mathbb{E}[\tau] < \infty$. 则

$$\mathbb{E}[S_\tau] = \mathbb{E}[X_1] \cdot \mathbb{E}[\tau].$$

证明. 注意到事件 $\{\tau \ge i\} = \{\tau \le i-1\}^c \in \mathcal{F}_{n-1}$. 从而 $\{\tau \ge i\}$ 与 X_i 独立。从而

$$\mathbb{E}[S_\tau] = \mathbb{E}\left[\sum_{i=1}^{\infty} X_i \chi_{\{i \ge \tau\}}\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[X_i \chi_{\{i \ge \tau\}}\right] = \sum_{i=1}^{\infty} \mathbb{E}[X_i] \cdot \mathbb{E}\left[\chi_{\{i \ge \tau\}}\right] = \mathbb{E}[X_1] \cdot \sum_{i=1}^{\infty} \mathbb{P}(\tau \ge i) = \mathbb{E}[X_1] \cdot \mathbb{E}[\tau].$$

□

我们想使用这个公式，要先检验 $\mathbb{E}[\tau] < \infty$. 注意到 $\forall k \ge 1$,

$$\mathbb{P}(\tau \ge k + m + M | \tau \ge k) \le 1 - 2^{-(m+M)}.$$

这是因为 $\tau \geq k$ 时，$X_{k+1}, \cdots, X_{k+m+M}$ 不能都取 1. 记 $\lambda = 1 - 2^{-(m+M)} < 1$. 从而

$$\mathbb{P}(\tau \geq k+m+M) = \mathbb{P}(\tau \geq k) \cdot \mathbb{P}(\tau \geq k+m+M | \tau \geq k) \leq \mathbb{P}(\tau \geq k) \cdot \lambda.$$

从而

$$\begin{aligned}
\mathbb{E}[\tau] = \sum_{i=1}^{\infty} \mathbb{P}(\tau \geq i) &= \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \sum_{i=m+M+1}^{\infty} \mathbb{P}(\tau \geq i) = \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \sum_{i=1}^{\infty} \mathbb{P}(\tau \geq i+m+M) \\
&\leq \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \lambda \sum_{i=1}^{\infty} \mathbb{P}(\tau \geq i) \\
&= (1+\lambda) \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \lambda \sum_{i=1+m+M}^{\infty} \mathbb{P}(\tau \geq i+m+M) = (1+\lambda) \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \lambda \sum_{i=1}^{\infty} \mathbb{P}(\tau \geq i+m+M) \\
&\leq (1+\lambda) \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \lambda^2 \sum_{i=1}^{\infty} \mathbb{P}(\tau \geq i) \\
&\leq (1+\lambda+\lambda^2) \sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) + \lambda^3 \sum_{i=1+m+M}^{\infty} \mathbb{P}(\tau \geq i) \\
&\cdots \\
&= \left(\sum_{n=0}^{\infty} \lambda^n \right) \cdot \left(\sum_{i=1}^{m+M} \mathbb{P}(\tau \geq i) \right) \\
&\leq \left(\sum_{n=0}^{\infty} \lambda^n \right) \cdot \left(\sum_{i=1}^{m+M} 1 \right) \\
&= (m+M) \sum_{n=0}^{\infty} \lambda^n < \infty.
\end{aligned}$$

由于 $\mathbb{E}[X_1] = 1 \cdot \mathbb{P}(X_1 = 1) + (-1) \cdot \mathbb{P}(X_1 = -1) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$. 由 *Wald* 第一方程（定理 *8.22*），

$$\mathbb{E}[S_\tau] = \mathbb{E}[X_1] \cdot \mathbb{E}[\tau] = 0.$$

另一方面

$$\begin{cases}(-m)\cdot\mathbb{P}\left(S_\tau=-m\right)+M\cdot\mathbb{P}\left(S_\tau=M\right)=\mathbb{E}\left[S_\tau\right]=0,\\ \mathbb{P}\left(S_\tau=-m\right)+\mathbb{P}\left(S_\tau=M\right)=1.\end{cases}$$

解得

$$\mathbb{P}\left(S_\tau=-m\right)=\frac{M}{m+M},\qquad \mathbb{P}\left(S_\tau=M\right)=\frac{m}{m+M}.$$

这符合我们的直观：想赢的钱越多，赢钱走人的概率越小。

我们还想计算 $\mathbb{E}\left[\tau\right]$:

定理 8.23 (Wald 第二方程)**.** 设 $X_1,X_2,\cdots$ 是独立同分布的随机变量序列，$\mathbb{E}\left[X_1\right]=0$, $\sigma^2(X_1)<\infty$. 若 τ 是一个停时，$\mathbb{E}\left[\tau\right]<\infty$, 则

$$\mathbb{E}\left[S_\tau^2\right]=\sigma^2(X_1)\cdot\mathbb{E}\left[\tau\right].$$

证明. 注意到 $\forall i\le n$, $\chi_{\{i\le\tau\}}\chi_{\{n\le\tau\}}=\chi_{\{n\le\tau\}}$, 从而

$$S_{\min\{\tau,n\}}=\sum_{i=1}^{n}X_i\chi_{\{i\le\tau\}}=\sum_{i=1}^{n-1}X_i\chi_{\{i\le\tau\}}+X_n\chi_{\{n\le\tau\}}=S_{\min\{\tau,n-1\}}+X_n\chi_{\{n\le\tau\}}.$$

从而

$$S^2_{\min\{\tau,n\}}=\left(S_{\min\{\tau,n-1\}}+X_n\chi_{\{n\le\tau\}}\right)^2=S^2_{\min\{\tau,n-1\}}+\left(2X_nS_{n-1}+X_n^2\right)\chi_{\{n\le\tau\}}.$$

注意到其中，X_n 分别与 S_{n-1} 和 $\chi_{\{n\le\tau\}}$ 独立，从而

$$\mathbb{E}\left[S^2_{\min\{\tau,n\}}\right]=\mathbb{E}\left[S^2_{\min\{\tau,n-1\}}+\left(2X_nS_{n-1}+X_n^2\right)\chi_{\{n\le\tau\}}\right]=\mathbb{E}\left[S^2_{\min\{\tau,n-1\}}\right]+\left(2\mathbb{E}\left[X_n\right]\mathbb{E}\left[S_{n-1}\right]+\mathbb{E}\left[X_n^2\right]\right)\mathbb{E}\left[\chi_{\{n\le\tau\}}\right].$$

由于 $\mathbb{E}\left[X_1\right]=0$, 从而 $\mathbb{E}\left[S_n-1\right]=0$, $\sigma^2(X_n)=\mathbb{E}\left[X_n^2\right]$. 从而

$$\mathbb{E}\left[S^2_{\min\{\tau,n\}}\right]=\mathbb{E}\left[S^2_{\min\{\tau,n-1\}}\right]+\left(2\mathbb{E}\left[X_n\right]\mathbb{E}\left[S_{n-1}\right]+\mathbb{E}\left[X_n^2\right]\right)\mathbb{E}\left[\chi_{\{n\le\tau\}}\right]=\mathbb{E}\left[S^2_{\min\{\tau,n-1\}}\right]+\sigma^2(X_1)\cdot\mathbb{P}\left(\tau\ge n\right)=\sigma^2(X_1)\cdot\sum_{i=1}^{n}\mathbb{P}\left(\tau\ge i\right).$$

从而极限与期望可交换。令 $n\to\infty$ 就有

$$\mathbb{E}\left[S_\tau\right]=\lim_{n\to\infty}\mathbb{E}\left[S^2_{\min\{\tau,n\}}\right]=\lim_{n\to\infty}\sigma^2(X_1)\cdot\sum_{i=1}^{n}\mathbb{P}\left(\tau\geq i\right)=\sigma^2(X_1)\cdot\sum_{i=1}^{\infty}\mathbb{P}\left(\tau\geq i\right)=\sigma^2(X_1)\cdot\mathbb{E}\left[\tau\right].$$

□

由于 $\sigma^2(X_1)=\mathbb{E}\left[X_1^2\right]-\mathbb{E}\left[X_1\right]^2=1^2\cdot\mathbb{P}\left(X_1=1\right)+(-1)^2\cdot\mathbb{P}\left(X_1=-1\right)-0^2=1^2\cdot\frac{1}{2}+(-1)^2\cdot\frac{1}{2}=1$. 从而由 *Wald* 第二方程（定理 *8.23*），

$$\mathbb{E}\left[\tau\right]=\sigma^2(X_1)\cdot\mathbb{E}\left[\tau\right]=\mathbb{E}\left[S_\tau^2\right]=(-m)^2\cdot\mathbb{P}\left(S_\tau=-m\right)+M^2\cdot\mathbb{P}\left(S_\tau=M\right)=m^2\cdot\frac{M}{m+M}+M^2\cdot\frac{m}{m+M}=mM.$$

由于 $\tau=\tau_{-m,M}\leq\tau_M$, 令 $m\to\infty$, 就有

$$\mathbb{E}\left[\tau_M\right]\geq\mathbb{E}\left[\tau_{-m,M}\right]=\mathbb{E}\left[\tau\right]=mM\to\infty.$$

这说明如果赌场的资本是无穷大，虽然赌徒一定会输光，但是他输光花费的事件也会趋于无穷大。

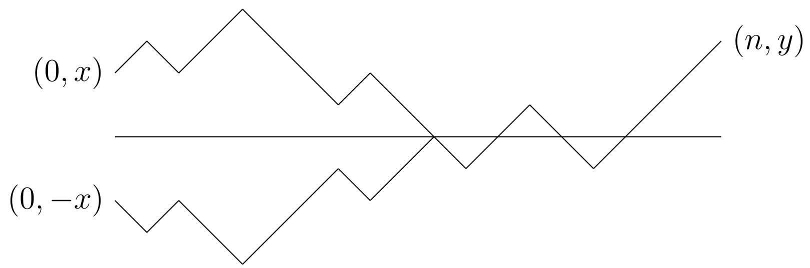
命题 8.24. 对于简单随机游动，我们可以计算 τ_m 的分布。注意到，每一条样本轨道 $S_1,\cdots,S_n$ 可以被表示成由 $(k-1,S_{k-1})$ 到 (k,S_k) 的线段构成的折线。我们要对 $(0,0)$ 到 (n,x) 的轨道进行计数。注意到 $k-1+S_{k-1}\equiv k+S_k \mod 2$, 从而 n,x 有相同的奇偶性。记 $a=\frac{n+x}{2}$, $b=\frac{n-x}{2}$. 则从 $(0,0)$ 到 (n,x) 的折线共有 $N_{n,x}=C_n^a$ 条。

定理 8.25 (反射原理)**.** 若 $x,y>0$. 则中间碰到 x-轴的从 $(0,x)$ 到 (n,y) 的道路的数目等于从 $(0,-x)$ 到 (n,y) 的道路的数目。

证明. 设 $(0,s_0),(1,s_1),\cdots,(n,s_n)$ 是一条中间碰到 x-轴的从 $(0,x)$ 到 (n,y) 的道路。设 $K=\inf\{k|\,s_k=0\}$, 令

$$s_k'=\begin{cases}-s_k, & k\leq K,\\ s_k, & K<k\leq n.\end{cases}$$

则 $(0,s'_0),(1,s'_1),\cdots,(n,s'_n)$ 就是一条从 $(0,-x)$ 到 (n,y) 的道路。



反之，设 $(0,t_0),(1,t_1),\cdots,(n,t_n)$ 是从 $(0,-x)$ 到 (n,y) 的道路。此时，一定存在 $t_i=0$. 设 $K=\inf\{k|\,t_k=0\}$, 令

$$t'_k=\begin{cases}-t_k, & k\le K,\\ t_k, & K<k\le n.\end{cases}$$

则 $(0,t'_0),(1,t'_1),\cdots,(n,t'_n)$ 就是一条中间碰到 x-轴的从 $(0,x)$ 到 (n,y) 的道路。

综上，我们建立了中间碰到 x-轴的从 $(0,x)$ 到 (n,y) 的道路与从 $(0,-x)$ 到 (n,y) 的道路的一一对应，从而中间碰到 x-轴的从 $(0,x)$ 到 (n,y) 的道路的数目等于从 $(0,-x)$ 到 (n,y) 的道路的数目。 □

定理 8.26 (选票定理)**.** *设整数 $\alpha>\beta\ge 0$. A,B 两人选举，得票数分别为 α,β. 则在唱票过程中，A 始终领先 B 的概率是 $\frac{\alpha-\beta}{\alpha+\beta}$.*

证明. 令 A 得票就向上，B 得票向下。记 $n=\alpha+\beta$, $x=\alpha-\beta$. 也就是说折线的终点是 (n,x). 并且 A 始终领先 B 就要求折线从 $(1,1)$ 开始不能碰到 x-轴。这样的折线的数目就是从 $(1,1)$ 到 (n,x) 的折线的数目减去碰到 x-轴的从 $(1,1)$ 到 (n,x) 的折线的数目，由反射原理（定理 8.25），它就是从 $(1,1)$ 到 (n,x) 的折线的数目减去从 $(1,-1)$ 到 (n,x) 的折线的数目。从 $(1,1)$ 到 (n,x) 的折线的数目就是在剩下 $n-1$ 次向右中取 β 次向下，为 C_{n-1}^{β}. 而从 $(1,-1)$ 到 (n,x) 的折线要比从 $(1,1)$ 到 (n,x) 的折线多一次向上，少一次向下，为 $C_{n-1}^{\beta-1}$. 综上，A 始终领先 B 的概率是

$$\frac{C_{n-1}^{\beta}-C_{n-1}^{\beta-1}}{C_n^{\alpha}}=\frac{\frac{(n-1)!}{\beta!(\alpha-1)!}-\frac{(n-1)!}{(\beta-1)!\alpha!}}{\frac{n!}{\alpha!\beta!}}=\frac{\alpha-\beta}{n}=\frac{\alpha-\beta}{\alpha+\beta}.$$

□

例 8.27. 用选票定理我们来计算随机游动 S_n 碰到 0 的首达时 $\tau_0 = \inf\{n|\, S_{2n} = 0\}$ 的分布。

引理 8.28. $\mathbb{P}(S_1 \neq 0, \cdots, S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0)$.

证明.

$$\mathbb{P}(S_1 > 0, \cdots, S_{2n} > 0) = \sum_{r=1}^{\infty} \mathbb{P}(S_1 > 0, \cdots, S_{2n-1} > 0, S_{2n} = 2r).$$

这事实上是一个有限和。并且由选票定理（定理 8.26），满足 $S_1 > 0, \cdots, S_{2n-1} > 0, S_{2n} = 2r$ 的折线就是从 $(1,1)$ 出发到 (n,x) 且不碰到 x-轴的折线，数目为 $C_{2n-1}^{n-r} - C_{2n-1}^{n-r-1}$. 所有折线的数目为 2^{2n}. 于是

$$\mathbb{P}(S_1 > 0, \cdots, S_{2n} > 0) = \sum_{r=1}^{\infty} \mathbb{P}(S_1 > 0, \cdots, S_{2n-1} > 0, S_{2n} = 2r) = \sum_{r=1}^{\infty} \frac{1}{2^{2n}} \left(C_{2n-1}^{n-r} - C_{2n-1}^{n-r-1}\right) = \frac{1}{2^{2n}} C_{2n-1}^{n-1} = \frac{1}{2} \cdot \frac{1}{2^{2n-1}} C_{2n-1}^{n-1}.$$

注意到 $S_{2n-1} = 1$ 是从而 $(0,0)$ 到 $(2n-1, 1)$ 的折线，它有 n 次向上，$n-1$ 次向下，从而数目为 C_{2n-1}^{n-1}, 从而 $\mathbb{P}(S_{2n-1} = 1) = \frac{1}{2^{2n-1}} C_{2n-1}^{n-1}$. 于是

$$\mathbb{P}(S_1 > 0, \cdots, S_{2n} > 0) = \frac{1}{2} \cdot \frac{1}{2^{2n-1}} C_{2n-1}^{n-1} = \frac{1}{2} \cdot \mathbb{P}(S_{2n-1} = 1).$$

同理有

$$\mathbb{P}(S_1 < 0, \cdots, S_{2n} < 0) = \frac{1}{2} \cdot \mathbb{P}(S_{2n-1} = -1).$$

从而

$$\begin{aligned}
\mathbb{P}(S_1 \neq 0, \cdots, S_{2n} \neq 0) &= \mathbb{P}(S_1 > 0, \cdots, S_{2n} > 0) + \mathbb{P}(S_1 < 0, \cdots, S_{2n} < 0) \\
&= \frac{1}{2} \cdot \mathbb{P}(S_{2n-1} = 1) + \frac{1}{2} \cdot \mathbb{P}(S_{2n-1} = -1) \\
&= \mathbb{P}(X_{2n} = -1) \cdot \mathbb{P}(S_{2n-1} = 1) + \mathbb{P}(X_{2n} = 1) \cdot \mathbb{P}(S_{2n-1} = -1) \\
&= \mathbb{P}(S_{2n} = 0|\, S_{2n-1} = 1) \cdot \mathbb{P}(S_{2n-1} = 1) + \mathbb{P}(S_{2n} = 0|\, S_{2n-1} = -1) \cdot \mathbb{P}(S_{2n-1} = -1) \\
&= \mathbb{P}(S_{2n} = 0).
\end{aligned}$$

□

由 *Stirling* 公式:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \mathrm{e}^{\lambda_n},\ \ \frac{1}{12n+1} < \lambda_n < \frac{1}{2n}.$$

就有

$$\mathbb{P}\left(\tau_0 > 2n\right) = \mathbb{P}\left(S_{2n} = 0\right) = \frac{C_{2n}^n}{2^{2n}} = \frac{\mathrm{e}^{\lambda_{2n} - 2\lambda_n}}{\sqrt{\pi n}} = \frac{1}{\sqrt{\pi n}}\left(1 + \mathrm{O}\left(\frac{1}{n^2}\right)\right).$$