概率论笔记

三生万物

林晓烁

https://xiaoshuo-lin.github.io

2024 年 3 月 26 日

前言

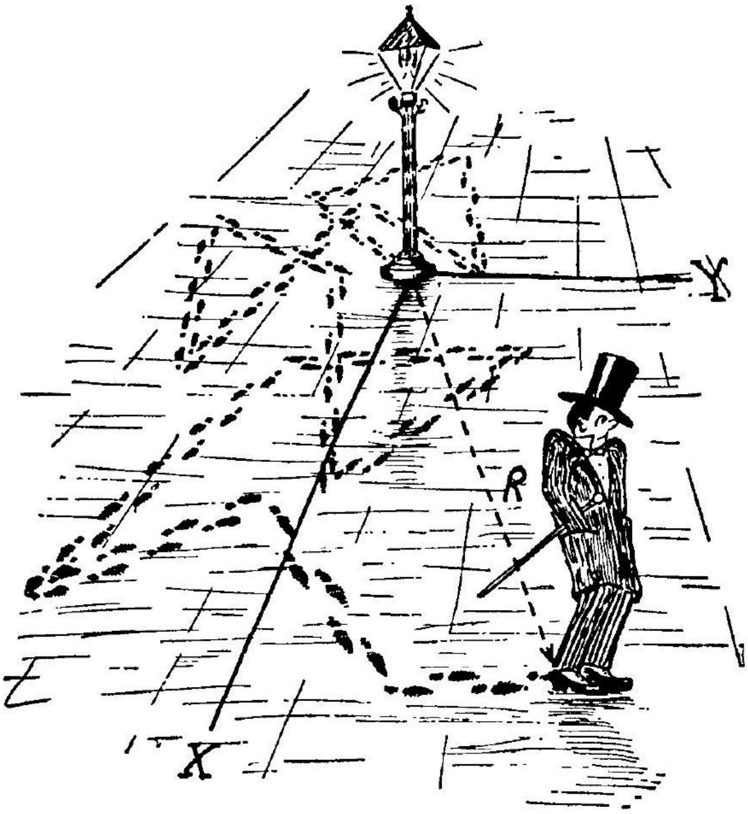
1. 第一部分是 2023 秋刘党政老师的概率论课堂笔记.

2. 第二部分是对 Probability and Random Processes 一书中部分习题及刘党政老师补充习题的解答.

3. 刘党政老师说, 学好概率论的一个充分条件是 $\sqrt{\text{干掉教材的习题}}$.

林晓烁

2024 年 3 月 26 日



目录

第一部分　课堂笔记

1　概率空间

例 1.1.1 (1) 掷硬币：$\Omega = \{\mathrm{H}, \mathrm{T}\}$, $A = \{\mathrm{H}\}$.

(2) 掷骰子：$\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{5, 6\}$.

(3) 电子自旋：$\Omega = \{\uparrow, \downarrow\}$, $A = \{\uparrow\}$.

定义 1.1.2 样本点指随机试验中出现的基本结果, 记为 ω. 样本空间指样本点全体构成的集合, 记为 Ω. 事件指样本空间的某个子集, 记为 A.

例 1.1.3 (Dow Jones 指数) $\mathcal{C}([0, T])$ 构成样本空间.

注 1.1.4 例1.1.1中样本点个数均有限, 而例1.1.3中样本点个数无穷. 这个不平凡的例子是随机过程中的一类研究对象.

我们可以用集合论语言与集合运算描述事件.

表 1.1: 集合术语与概率论术语对照

集合术语	概率论术语
随机试验结果 $\omega \in A$	A 发生
Ω	必然事件
$\varnothing$	不可能事件
A^{c}	事件 A 的补/余/对立事件
$A \cap B$(或简记为 AB)	事件交 (同时发生)
$A \cup B$	事件并 (A 发生或 B 发生)
$A \subset B$	A 发生时 B 亦发生
$A \cap B = \varnothing$	A 与 B 互不相容

$A_1,\cdots,A_n$ 两两不交	$A_1,\cdots,A_n$ 互不相容

我们知道, 事件都是 Ω 的子集, 但是否 Ω 的所有子集都是事件?

例 1.1.5 掷硬币至 H 出现的时刻, $\Omega=\{1,2,3,\cdots\}$, 样本空间为可列无穷集, 我们关心事件 $A=\{2,4,6,\cdots\}$. 这就要求事件域对可列并封闭.

定义 1.1.6 $\mathcal{F}\subset\{0,1\}^{\Omega}$ 称为一个 σ 代数 (或 σ 域, 事件域), 若

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

(2) $A\in\mathcal{F}\implies A^{\mathrm{c}}\in\mathcal{F}$.

(3) $A_1,A_2,\cdots\in\mathcal{F}\implies\displaystyle\bigcup_{n=1}^{\infty}A_n\in\mathcal{F}$.

并称二元组 $(\Omega,\mathcal{F})$ 为一个可测空间.

例 1.1.7 (1) 关于 Ω 的最小 σ 代数是 $\mathcal{F}=\{\varnothing,\Omega\}$.

(2) 设 $A\subset\Omega$, 则 $\mathcal{F}=\{\varnothing,A,A^{\mathrm{c}},\Omega\}$ 是一个 σ 域.

(3) 有时 Ω“不太大”, Ω 的幂集 $\{0,1\}^{\Omega}$ 也是一个 σ 域.

定义概率的直观想法是频率稳定性, 即重复试验 N 次, 看 A 发生次数 N_A. 经验表明, $\displaystyle\lim_{n\to\infty}\frac{N_A}{N}=$ 常数 $=:\mathbb{P}(A)$. 此时显然有如下性质:

(1) 当 $AB=\varnothing$ 时, $N_{A\cup B}=N_A+N_B$, 进而 $\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B)$.

(2) $\mathbb{P}(\Omega)=1$, $\mathbb{P}(\varnothing)=0$.

定义 1.1.8 $\mathbb{P}:\mathcal{F}\to\mathbb{R}$ 称为一个概率测度, 若

(1) (非负性) $\forall A\in\mathcal{F}$, $\mathbb{P}(A)\geqslant 0$.

(2) (规范性/归一化) $\mathbb{P}(\Omega)=1$.

(3) (可列可加性) 当 $\{A_n\}$ 互不相容时, $\mathbb{P}\left(\displaystyle\bigcup_n A_n\right)=\displaystyle\sum_n\mathbb{P}(A_n)$.

并称三元组 $(\Omega,\mathcal{F},\mathbb{P})$ 为一个概率空间.

注 1.1.9 可列可加性蕴含着有限可加性.

例 1.1.10 掷硬币, $\Omega=\{\mathrm{H},\mathrm{T}\}$, $\mathcal{F}=\{0,1\}^{\Omega}$. 一个可能的概率测度 $\mathbb{P}:\mathcal{F}\to[0,1]$ 为

$$\mathbb{P}(\varnothing)=0,\quad\mathbb{P}(\mathrm{H})=p,\quad\mathbb{P}(\mathrm{T})=1-p,\quad\mathbb{P}(\Omega)=1,$$

其中 $p\in[0,1]$. 若 $p=\dfrac{1}{2}$, 则称硬币是“均匀”的.

例 1.1.11 均匀骰子, $\Omega=\{1,2,3,4,5,6\}$, $\mathcal{F}=\{0,1\}^{\Omega}$, $\mathbb{P}(A)=\dfrac{|A|}{6}$.

注 1.1.12 例1.1.10与1.1.11均属于“有限等可能”情形, 称为古典概型.

引理 1.1.13 (1) $\mathbb{P}\left(A^{c}\right)+\mathbb{P}(A)=1$.

(2) 若 $A\subset B$, 则 $\mathbb{P}(B)=\mathbb{P}(A)+\mathbb{P}(B\setminus A)\geqslant\mathbb{P}(A)$.

(3) $\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B\setminus AB)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(AB)$.

(4) (Jordan 公式) $\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}\right)$.

引理 1.1.14 ($\mathbb{P}$ 的连续性) 设有单调增事件列 $A_1\subset A_2\subset A_3\subset\cdots$, 记

$$A=\bigcup_{i=1}^{\infty}A_i=\lim_{i\to\infty}A_i,$$

则 $\mathbb{P}(A)=\lim_{i\to\infty}\mathbb{P}(A_i)$. 类似地, 设有单调减事件列 $B_1\supset B_2\supset B_3\supset\cdots$, 则

$$B=\bigcap_{i=1}^{\infty}B_i=\lim_{i\to\infty}B_i$$

满足 $\mathbb{P}(B)=\lim_{i\to\infty}\mathbb{P}(B_i)$.

证明 只需证单调增事件列情形, 利用 De Morgan 法则可得单调减事件列情形.

可将 A 写成不交并 $A=A_1\cup(A_2\setminus A_1)\cup(A_3\setminus A_2)\cup\cdots$, 由定义1.1.8得

$$\begin{aligned}\mathbb{P}(A)&=\mathbb{P}(A_1)+\sum_{i=1}^{\infty}\mathbb{P}\left(A_{i+1}\setminus A_i\right)\\&=\mathbb{P}(A_1)+\lim_{n\to\infty}\sum_{i=1}^{n-1}\left[\mathbb{P}(A_{i+1})-\mathbb{P}(A_i)\right]\\&=\lim_{n\to\infty}\mathbb{P}(A_n).\end{aligned}$$

□

2　条件概率和独立性

直观想法：重复试验 N 次, 看在 B 发生的条件下 A 发生的次数：

$$\frac{N_{AB}}{N_B}=\frac{\frac{N_{AB}}{N}}{\frac{N_B}{B}}\to\frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

定义 1.2.1 对 $B, \mathbb{P}(B) > 0$, B 发生条件下 A 发生的条件概率

$$\mathbb{P}(A \mid B) \coloneqq \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

注 1.2.2 由定义验证可知, 给定 B 时, 条件概率 $\mathbb{P}(\cdot \mid B)$ 也是概率测度.

定理 1.2.3 (乘法规则) $\mathbb{P}(AB) = \mathbb{P}(B)\mathbb{P}(A \mid B)$.

定义 1.2.4 若 $B_1, \cdots, B_n$ 满足 $\bigcup_{i=1}^{n} B_i = \Omega$, 且 $\{B_i\}$ 互不相容, 则称之为 Ω 的一个划分, 其中 n 可以为 ∞, 即可列无穷. 特别地, B 与 B^{c} 为 Ω 的一个划分.

引理 1.2.5 (全概公式) 若 $\{B_i\}_{i=1}^{n}$ 为 Ω 的划分, 且 $\mathbb{P}(B_i) > 0, \forall i$. 则

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(B_i)\mathbb{P}(A \mid B_i).$$

注 1.2.6 全概公式可以将较难求概率的事件化为较简单的事件进行研究.

引理 1.2.7 (Bayes 公式) 若 $\{A_i\}_{i=1}^{n}$ 为 Ω 的划分, 且 $\mathbb{P}(A_i) > 0, \forall i$. 则当 $\mathbb{P}(B) > 0$ 时, 有

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B \mid A_i)}{\displaystyle\sum_{j=1}^{n} \mathbb{P}(A_j)\mathbb{P}(B \mid A_j)}.$$

注 1.2.8 Bayes 公式可以理解为“已知结果 (B) 寻找原因 (A_i)”, 是“不全概”的公式.

定义 1.2.9 称事件 A 与 B 独立, 若

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

更一般地, 称 $\{A_i\}_{i \in I}$ 相互独立, 若对任意有限子集 $J \subset I$ 均有

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

注 1.2.10 (1) 这里的“独立”是通过等式规范的“统计独立性”, 与生活中对“独立”的理解不完全相同.

(2) $\{A_i\}_{i \in I}$ 两两独立是指对任意 $i, j \in I$, A_i 与 A_j 相互独立. 注意与 $\{A_i\}_{i \in I}$ 相互独立区分 (相互独立的阶为指数级, 两两独立的阶为 n^2 级).

(3) 由“独立”带来的“分离性”在计算中有重要意义 (可类比多元微积分).

例 1.2.11 (两两独立 $\neq$ 相互独立) 设 $\Omega=\{1,2,3,4\}$, $A=\{1,2\}$, $B=\{1,3\}$, $C=\{1,4\}$. 则 $\mathbb{P}(A)=\mathbb{P}(B)=\mathbb{P}(C)=\dfrac{1}{2}$, $\mathbb{P}(AB)=\mathbb{P}(BC)=\mathbb{P}(CA)=\dfrac{1}{4}$, 因此 A,B,C 两两独立. 但 $\mathbb{P}(ABC)=\mathbb{P}(\{1\})=\dfrac{1}{4}\neq\left(\dfrac{1}{2}\right)^3$, 因此 A,B,C 非相互独立.

引理 1.2.12 若 A 与 B 独立, 则 A 与 B^{c}、A^{c} 与 B、A^{c} 与 B^{c} 均独立.

证明 $\mathbb{P}(AB^{\mathrm{c}})=-\mathbb{P}(AB)+\mathbb{P}(A)=\mathbb{P}(A)\left(1-\mathbb{P}(B)\right)=\mathbb{P}(A)\mathbb{P}(B^{\mathrm{c}})$. 余下结论由此可得. □

注 1.2.13 若 $\{A_i\}_{i=1}^n$ 相互独立, 则 $A_1^{\mathrm{c}},A_2,\cdots,A_n$ 也相互独立.

例 1.2.14 (重复独立试验, 小概率事件必然发生) 记 $A_k=\{A\text{在第 }k\text{ 次发生}\}$, $\mathbb{P}(A_k)=\varepsilon\in(0,1)$. 则 $\{A_k\}_{k=1}^N$ 相互独立. 因此在前 N 次中 A 发生的概率为

$$\mathbb{P}\left(\bigcup_{k=1}^N A_k\right)=1-\mathbb{P}\left(\bigcap_{k=1}^N A_k^{\mathrm{c}}\right)=1-(1-\varepsilon)^N\to 1,\quad N\to\infty.$$

3　概率模型

例 1.3.1 (生日问题) 设 $A=\{n\text{个人中至少有两人同一天生日}\}$, 求 $\mathbb{P}(A)$.

解 $\mathbb{P}(A^{\mathrm{c}})=\dfrac{\mathrm{A}_{365}^n}{365^n}\implies\mathbb{P}(A)=1-\dfrac{\mathrm{A}_{365}^n}{365^n}$. □

注 1.3.2 实际上, 对于不那么大的 n, $\mathbb{P}(A)$ 也可以十分接近 1.

n	40	45	50	55
$\mathbb{P}(A)$	0.87	0.94	0.97	0.99

例 1.3.3 (计数问题) 从 n 个不同对象中取 m 个, 按有序、无序及是否允许重复分类讨论方式数如下:

	不重复	可重复
有序	A_n^m	n^m
无序	C_n^m	C_{n-1+m}^m

关于无序、可重复情形的说明: 用 (无侧边) 匣子表示, 第 k 个小格表示第 k 个对象, 每小格放入小球数表示该对象选取重复次数. 从所需小球 (m) 和挡板 $(n-1)$ 共 $n-1+m$ 个位置中选取挡板位置 $(\mathrm{C}_{n-1+m}^{n-1})$ 即可确定小球位置, 进而确定取法.

例 1.3.4 (三种统计) 将 n 个小球投入 N ($\geqslant n$) 个盒中, 每种投法等可能. 记

$$A=\{\text{前 } n \text{ 个盒子中各含一个球}\},$$

求 $\mathbb{P}(A)$.

解 分球是否可分辨、盒子容量是否有限进行讨论：

情形 1：可分辨、无限制 $\mathbb{P}(A)=\dfrac{n!}{N^n}$.

情形 2：不可辨、无限制 同例1.3.3中无序、可重复情形知 $\mathbb{P}(A)=\dfrac{1}{\mathrm{C}_{N-1+n}^n}$.

情形 3：不可辨、每盒至多 1 个球 $\mathbb{P}(A)=\dfrac{1}{\mathrm{C}_N^n}$. □

注 1.3.5 以上三种情形分别对应 Maxwell-Boltzmann 统计、Bose-Einstein 统计、Fermi-Dirac 统计.

例 1.3.6 (配对问题) n 对夫妇随机坐在长桌的两侧, 先生全部坐在其中一侧. 求至少有一对夫妇面对面的概率.

解 设 $B=\{\text{至少有一对夫妇面对面}\}$, $A_k=\{\text{第 } k \text{ 位先生与他的妻子面对面}\}$. 则 $B=\bigcup\limits_{k=1}^{n}A_k$, 由 Jordan 公式得

$$\mathbb{P}(B)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}(A_{i_1}\cdots A_{i_k}),$$

又

$$\mathbb{P}(A_{i_1}\cdots A_{i_k})=\frac{(n-k)!}{n!},$$

因此

$$\begin{aligned}\mathbb{P}(B)&=\sum_{k=1}^{n}(-1)^{k+1}\frac{(n-k)!}{n!}\mathrm{C}_n^k\\&=\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!}\to 1-\frac{1}{\mathrm{e}},\quad n\to\infty.\end{aligned}$$

□

例 1.3.7 (赌徒破产问题) 设玩家财富为 k, 庄家财富为 $N-k$, 掷硬币, 若正面朝上则玩家财富增加 1, 否则减少 1. 双方赌至一方破产, 求玩家破产的概率.

解 设 $A_k=\{\text{初始财富为 } k \text{ 最后破产}\}$, $B=\{\text{第一次掷硬币正面朝上}\}$, 记 $p_k=\mathbb{P}(A_k)$. 则由全概公式,

$$\mathbb{P}(A_k)=\mathbb{P}(B)\mathbb{P}(A_k\mid B)+\mathbb{P}(B^{\mathrm{c}})\mathbb{P}(A_k\mid B^{\mathrm{c}}),$$

即

$$p_k = \frac{1}{2}p_{k+1} + \frac{1}{2}p_{k-1}, \quad k = 1, 2, \cdots.$$

再由边界条件 $p_0 = 1$, $p_N = 0$ 得 $p_k = 1 - \dfrac{k}{N}$. □

例 1.3.8 (Pólya 坛子模型) 坛子里有 b 个黑球和 r 个红球, 每次从中取一个后放回, 再放入 c 个同色球. 记 $B_n = \{$第 n 次抽取取到黑球$\}$, 求 $\mathbb{P}(B_n)$.

解 1 容易验证, 在 n 次抽取中抽到 k 个黑球和 $n-k$ 个红球的概率与两种颜色出现的次序无关, 每种次序的概率均为

$$D_k(b) = \frac{b(b+c)\cdots(b+(k-1)c)r(r+c)\cdots(r+(n-k-1)c)}{(b+r)(b+r+c)\cdots(b+r+(n-1)c)}.$$

记 $A_k = \{$前 n 次抽取共抽中 k 个黑球$\}$, 则 $\{A_k\}$ 为样本空间的一个划分, 由全概公式得

$$\begin{aligned}
\mathbb{P}(B_{n+1}) &= \sum_k \mathbb{P}(A_k)\mathbb{P}(B_{n+1} \mid A_k) \\
&= \sum_k D_k(b)\mathrm{C}_n^k \frac{b+kc}{b+r+nc} \\
&= \frac{b}{b+r}\sum_k D_k(b+c)\mathrm{C}_n^k \\
&= \frac{b}{b+r}.
\end{aligned}$$

其中 $\displaystyle\sum_k D_k(b+c)\mathrm{C}_n^k = 1$ 可由其对应的事件恰为整个样本空间解释. □

解 2 记 $p_n = \mathbb{P}(B_n)$. 考虑第 $n-1$ 次抽取的结果可得递推关系

$$p_n = p_{n-1}\frac{[b+r+(n-2)c]p_{n-1}+c}{b+r+(n-1)c} + (1-p_{n-1})\frac{[b+r+(n-2)c]p_{n-1}}{b+r+(n-1)c}.$$

化简可得 $p_n = p_{n-1}$. 于是 $p_n = p_1 = \dfrac{b}{b+r}$. □

注 1.3.9 当 $n = -1$ 时即无放回抽取, 对应于抽奖; 当 $n = 0$ 时即有放回抽取.

4　随机变量

定义 1.4.1 设概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$. 若函数 $X: \Omega \to \mathbb{R}$ 满足

$$\{X \leqslant x\} \coloneqq \{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F}, \quad \forall x \in \mathbb{R},$$

则 X 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量.

注 1.4.2 定义1.4.1中的 $(\Omega, \mathcal{F}, \mathbb{P})$ 亦可换为 $(\Omega, \mathcal{F})$.

例 1.4.3 $\Omega = \{\mathrm{H}, \mathrm{T}\}$. $X(\mathrm{H}) = 1$, $X(\mathrm{T}) = -1$. 则 X 为随机变量.

定义 1.4.4 设 X 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上随机变量, 称 $F(x) = \mathbb{P}(X \leqslant x)$ 为 X 的概率分布函数.

例 1.4.5 设例1.4.3中 $\mathbb{P}(\mathrm{H}) = p$, $\mathbb{P}(\mathrm{T}) = q$, 则 X 的分布函数为

$$F(x) = \begin{cases} 1, & x \geqslant 1, \\ q, & -1 \leqslant x < 1, \\ 0, & x < -1. \end{cases}$$

定理 1.4.6 (概率分布函数 F 的性质)

(1) 单调增, 即若 $x < y$, 则 $F(x) \leqslant F(y)$.

(2) $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

(3) 右连续, 即 $\lim\limits_{h \to 0^+} F(x+h) = F(x)$.

证明 (1) 由 $\{X \leqslant x\} \subset \{X \leqslant y\}$ 可知.

(2) 令 $A_n = \{X \leqslant n\}$, $n = 1, 2, \cdots$. 则 $F(n) = \mathbb{P}(A_n)$ 且 $\{A_n\}$ 单调升. 由 $\mathbb{P}$ 的连续性,

$$\lim_{n \to \infty} F(n) = \lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_n A_n\right) = \mathbb{P}(\Omega) = 1.$$

再由 (1) 单调增可知 $F(x) \to 1,\ x \to \infty$. 另一部分类似.

(3) 取 $B_n = \left\{X \leqslant x + \dfrac{1}{n}\right\}$, 则 $\{B_n\}$ 单调降且 $\bigcap\limits_n B_n = \{X \leqslant x\}$. 因此 $F\left(x + \dfrac{1}{n}\right) = \mathbb{P}(B_n) \to \mathbb{P}\left(\bigcap\limits_n B_n\right) = F(x),\ n \to \infty$. □

注 1.4.7 (1) 测度论表明满足定理1.4.6(1)(2)(3) 的函数 F 必为某概率空间上某随机变量的概率分布函数. 因此将这样的 F 称为分布函数.

(2) 有时定义 $G(x) \coloneqq \mathbb{P}(X < x)$ 为 X 的分布函数, 这时定理1.4.6(3) 的“右连续”应改为“左连续”.

(3) 分布函数“忘却”了样本空间 Ω 的信息.

命题 1.4.8 (1) $\mathbb{P}(X > x) = 1 - F(x)$.

(2) $\mathbb{P}(x < X \leqslant y) = F(y) - F(x)$.

(3) $\mathbb{P}(X = y) = F(y) - F(y^-)$.

例 1.4.9 设常值随机变量 $X=c$. 则 $F(x)=\begin{cases}1, & x\geqslant c,\\ 0, & x<c.\end{cases}$

更一般地, 若 $\mathbb{P}(X=c)=1$, 即几乎处处常值随机变量, 则 $F(x)$ 也同上.

例 1.4.10 (Bernoulli 两点分布) 设 $\mathbb{P}(X=1)=p$, $\mathbb{P}(X=0)=q$, 其中 $p+q=1$. 则 $F(x)=\begin{cases}1, & x\geqslant 1,\\ q, & 0\leqslant x<1,\\ 0, & x<0.\end{cases}$

例 1.4.11 (示性函数) 对 $A\in\mathcal{F}$, 定义 $I_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A.\end{cases}$ 则 $\mathbb{P}(I_A=1)=\mathbb{P}(A)$.

定义 1.4.12 (1 维 Borel 域) 所有形如 $(a,b]$ 的区间生成的 $\mathbb{R}$ 上最小 σ 域称为 1 维 Borel 域, 记为 $\mathcal{B}(\mathbb{R})$.

注 1.4.13 “最小”指的是 $\mathcal{B}(\mathbb{R})$ 是所有包含形如 $(a,b]$ 的区间的 σ 域之交.

定义 1.4.14 (d 维 Borel 域) 所有形如 $(a_1,b_1]\times\cdots\times(a_d,b_d]$ 的区域生成的 $\mathbb{R}^d$ 上最小 σ 域称为 d 维 Borel 域, 记为 $\mathcal{B}\left(\mathbb{R}^d\right)$.

注 1.4.15 易知, $\{b\}=\bigcap\limits_{n=1}^{\infty}\left(b-\dfrac{1}{n},b\right]\in\mathcal{B}(\mathbb{R})$, $(a,b)=(a,b]\setminus\{b\}\in\mathcal{B}(\mathbb{R})$. 同理, $[a,b),[a,b]\in\mathcal{B}(\mathbb{R})$.

定义 1.4.16 Borel 域中的集合称为 Borel 集.

定理 1.4.17 设 X 为 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量, 则对任意 Borel 集 $B\in\mathcal{B}(\mathbb{R})$, 有

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

证明 令 $\mathcal{A}=\{A\subset\mathbb{R}: X^{-1}(A)\in\mathcal{F}\}$. 我们断言：$\mathcal{A}$ 为 σ 域.

- $X^{-1}(\mathbb{R})=\Omega\in\mathcal{F}\implies\mathbb{R}\in\mathcal{A}$.
- 若 $A\in\mathcal{A}$, 即 $X^{-1}(A)\in\mathcal{F}$, 则 $X^{-1}(A^{\mathrm{c}})=\left(X^{-1}(A)\right)^{\mathrm{c}}\in\mathcal{F}\implies A^{\mathrm{c}}\in\mathcal{A}$.
- 若 $A_n\in\mathcal{A}$, 即 $X^{-1}(A_n)\in\mathcal{F}$, 则 $X^{-1}\left(\bigcup\limits_n A_n\right)=\bigcup\limits_n X^{-1}(A_n)\in\mathcal{F}\implies\bigcup\limits_n A_n\in\mathcal{A}$.

由断言, 根据 X 的定义知 $(-\infty,x]\in\mathcal{A}$, $\forall x\in\mathbb{R}$. 进而 $(a,b]=(-\infty,b]\setminus(-\infty,a]\in\mathcal{A}$, $\forall a<b$. 再由 $\mathcal{B}(\mathbb{R})$ 的最小性知 $\mathcal{B}(\mathbb{R})\subset\mathcal{A}$. $\square$

注 1.4.18 随机变量可等价地定义为 $X:\Omega\to\mathbb{R}$ 可测 $[(\Omega,\mathcal{F})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))]$, 使得 $X^{-1}(B)\in\mathcal{F},\forall B\in\mathcal{B}(\mathbb{R})$.

命题 1.4.19 若 X,Y 为 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量, 则 $X+Y$ 也是 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量.

证明 只需证

$$\{X+Y\leqslant x\}=\bigcap_{r\in\mathbb{Q}}\left(\{X\leqslant r\}\cup\{Y\leqslant x-r\}\right),\quad \forall x\in\mathbb{R}.$$

LHS $\subset$ RHS 是显然的. 下证 RHS $\subset$ LHS. 若 $\omega\notin$ LHS, 即 $X(\omega)>-Y(\omega)+x$, 由 $\mathbb{Q}$ 在 $\mathbb{R}$ 中稠密可取 $r\in\mathbb{Q}$, 使得 $X(\omega)>r>-Y(\omega)+x$, 也即

$$X(\omega)>r\quad\text{且}\quad Y(\omega)>x-r.$$

故 $\omega\notin$ RHS. □

5　随机向量

定义 1.5.1 $X_1,\cdots,X_n$ 为 $(\Omega,\mathcal{F},\mathbb{P})$ 上随机变量, 则称 $\mathbf{X}=(X_1,\cdots,X_n)$ 为 n 维随机向量 (或 n 维随机变量), 且称 $F(x_1,\cdots,x_n)=\mathbb{P}(X_1\leqslant x_1,\cdots,X_n\leqslant x_n)$ 为 $\mathbf{X}$ 的联合分布函数.

为了叙述简便, 下面先考虑 2 维随机向量 (X,Y) 及其联合分布函数 $F(x,y)=\mathbb{P}(X\leqslant x,Y\leqslant y)$.

定理 1.5.2 (1) F 分别关于 x,y 单调增.

(2) F 分别关于 x,y 右连续.

(3) $\lim\limits_{x\to-\infty}F(x,y)=\lim\limits_{y\to-\infty}F(x,y)=0$, $\lim\limits_{x,y\to+\infty}F(x,y)=1$.

(4) 对任意 $x_1<x_2,y_1<y_2$,

$$\begin{aligned}\mathbb{P}\left(X\in(x_1,x_2],Y\in(y_1,y_2]\right)&=\mathbb{P}\left(X\leqslant x_2,Y\in(y_1,y_2]\right)-\mathbb{P}\left(X\leqslant x_1,Y\in(y_1,y_2]\right)\\&=[F(x_2,y_2)-F(x_2,y_1)]-[F(x_1,y_2)-F(x_1,y_1)]\geqslant 0.\end{aligned}$$

注 1.5.3 在定理1.5.2中, 由 (2)(3)(4) 可推出 (1), 但由 (1)(2)(3) 不能推出 (4). 反例如下：设

$$F(x,y)=\begin{cases}1, & x+y\geqslant 0,\\ 0, & x+y<0\end{cases}.$$

则 (1)(2)(3) 成立, 但对 $x_1=y_1=-1,x_2=y_2=1$, (4) 不成立. 一般地, 若一个函数满足 (2)(3)(4), 则它必为某概率空间上某 2 维随机向量的联合分布函数.

定义 1.5.4 $\mathbf{X}=(X_1,\cdots,X_n)$ 只取 $\mathbb{R}^n$ 中至多可数个点, 则称 $\mathbf{X}$ 为离散型随机变量, 并称 $f(x_1,\cdots,x_n)=\mathbb{P}(X_1=x_1,\cdots,X_n=x_n)$ 为 $\mathbf{X}$ 的 (联合) 分布列 (或联合质量函数).

注 1.5.5 此时

$$F(x_1,\cdots,x_n)=\sum_{u_i\leqslant x_i,\ \forall i} f(u_1,\cdots,u_n).$$

由于 F 有跳跃, 又称此分布为原子分布.

定义 1.5.6 若存在 n 元函数 $f:\mathbb{R}^n\to\mathbb{R}$ 非负可积, 使得 $\mathbf{X}$ 的联合分布函数

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

则称 $\mathbf{X}$ 为连续型随机向量, 并称 f 为联合密度函数.

注 1.5.7 定理1.5.6中 f 可积指的是 Lebesgue 意义下的可积, 但可在 Riemann 积分意义下理解.

例 1.5.8 有界区域 $G\subset\mathbb{R}^n$, 其体积 $|G|<\infty$. 均匀分布的密度函数

$$f(\mathbf{x})=\frac{1}{|G|},\quad \forall \mathbf{x}\in G.$$

定义 1.5.9 对 $\mathbf{X}=(X_1,\cdots,X_n)$, $1\leqslant k\leqslant n$, 称 $(X_1,\cdots,X_k)$ 为 $\mathbf{X}$ 的边缘分布, 并称 $\mathbb{P}(X_1\leqslant x_1,\cdots,X_k\leqslant x_k)$ 为 $F(x_1,\cdots,x_n)$ 的边缘分布函数. 易知,

$$\mathbb{P}(X_1\leqslant x_1,\cdots,X_k\leqslant x_k)=\lim_{x_{k+1},\cdots,x_n\to+\infty} F(x_1,\cdots,x_n).$$

注 1.5.10 在 1 维情形下作几点说明:

(1) $\dfrac{F(x_0+\Delta x)-F(x_0)}{\Delta x}=\dfrac{1}{\Delta x}\displaystyle\int_{x_0}^{x_0+\Delta x} f(u)\,\mathrm{d}u$. 当 x_0 是 f 的连续点时,

$$\mathbb{P}\left(x\in(x_0,x_0+\Delta x]\right)\sim\Delta x\cdot f(x_0).$$

这是密度函数的直观意义.

(2) $F(x)=\displaystyle\int_{-\infty}^{x} f(u)\,\mathrm{d}u$. 由此可见密度函数不唯一, f 在有限个点上改变取值不影响 F 的取值.

(3) $\mathbb{P}(X=a)=\displaystyle\int_{a-\frac{1}{n}}^{a} f(u)\,\mathrm{d}u\to 0,\quad n\to\infty$, 即 $\mathbb{P}(X=a)=0$.

(4) 若 $F(x)$ 连续, 且在有限个点之外 $F'(x)$ 存在且连续, 则 F 为连续型随机变量的分布函数, 且 F' 为密度函数.

(5) $F(x)$ 的不连续点至多可数.[1]

(6) 有 Lebesgue 分解：$F(x)=c_1F_1+c_2F_2+c_3F_3$, 其中 $c_1+c_2+c_3=1$, $c_i\geqslant 0$, F_1 为离散型随机变量的分布函数, F_2 为连续型随机变量的分布函数, F_3 奇异.

例 1.5.11 (钟表指针) $\Omega=[0,2\pi)$, $\mathcal{F}=\{A\subset\Omega : A\in\mathcal{B}(\mathbb{R})\}$.$\mathbb{P}(A):=\dfrac{|A|}{2\pi}$, 其中 $|\cdot|$ 为 Lebesgue 测度. 定义随机变量 $X(\omega)=\omega$, $Y(\omega)=\omega^2$, 可知

$$F_X(x)=\mathbb{P}(X\leqslant x)=\begin{cases}1, & x\geqslant 2\pi,\\ \dfrac{x}{2\pi}, & x\in[0,2\pi),\\ 0, & x<0.\end{cases}$$

从而

$$f_X(x)=\begin{cases}\dfrac{1}{2\pi}, & x\in[0,2\pi),\\ 0, & \text{其他}.\end{cases}$$

同样地,

$$F_Y(y)=\mathbb{P}(Y\leqslant y)=\begin{cases}1, & y\geqslant 4\pi^2,\\ \dfrac{\sqrt{y}}{2\pi}, & y\in[0,4\pi^2),\\ 0, & y<0.\end{cases}$$

从而

$$f_Y(y)=\begin{cases}\dfrac{1}{4\pi\sqrt{y}}, & y\in[0,4\pi^2),\\ 0, & \text{其他}.\end{cases}$$

例 1.5.12 (既非离散型也非连续型) 随机变量 X 有密度函数 $f(x)=\begin{cases}1, & x\in[0,1],\\ 0, & \text{其他}.\end{cases}$. 掷一枚均匀的硬币, 每次结果相互独立. 若出现 H 则令 $Y=0$, 若出现 T 则令 $Y=X$. 则

- 当 $y\in[0,1]$ 时,

$$\begin{aligned}F_Y(y)&=\mathbb{P}(Y\leqslant y)=\mathbb{P}(\mathrm{H})\mathbb{P}(Y\leqslant y\mid \mathrm{H})+\mathbb{P}(\mathrm{T})\mathbb{P}(Y\leqslant y\mid \mathrm{T})\\&=\frac{1}{2}+\frac{1}{2}\mathbb{P}(X\leqslant y\mid \mathrm{T})=\frac{1+y}{2}.\end{aligned}$$

[1]回忆数学分析中的如下定理：区间 (a,b) 上的递增 (减) 函数的间断点一定是跳跃点, 且跳跃点集是至多可数的.

- 当 $y>1$ 时, $\mathbb{P}(y)=1$.
- 当 $y<0$ 时, $\mathbb{P}(y)=0$.

画出 $F_Y(y)$ 图像可见 Y 既非离散型也非连续型随机变量.

6　离散型随机变量

不要小看离散情形　阿伏伽德罗常数 $\sim 6.022\times 10^{23}$、地球上原子数 $\sim 10^{50}$、宇宙间原子数 $\sim 10^{80}$、围棋有效棋局 (约占 1.2%)$\sim 2\times 10^{170}$.

离散型随机变量的性质　$\mathbb{P}(X=x_k)=p_k\ (k=1,2,\cdots),\ \sum\limits_k p_k=1,\ p_k\geqslant 0, \forall k.$

例 1.6.1 (二项分布) 背景：掷硬币 n 次, 求正面出现次数.

$$\mathbb{P}(X=k)=\mathrm{C}_n^k p^k q^{n-k},\quad k=0,1,\cdots,n,\quad p+q=1.$$

称 X 服从参数 (n,p) 的二项分布, 记为 $X\sim B(n,p)$.

例 1.6.2 (几何分布) 背景：掷硬币到首次出现 H 的等待时间.

$$\mathbb{P}(X=k)=pq^{k-1},\quad k=1,2,\cdots,\quad p\in(0,1).$$

易知, $\mathbb{P}(X>k)=q^k\ (k=0,1,\cdots)$. 若前 m 次 H 未出现, 设新的等待时间为 X', 则

$$\mathbb{P}(X'=k)=\mathbb{P}(X=m+k\mid X>m)=\frac{\mathbb{P}(X=m+k)}{\mathbb{P}(X>m)}=\frac{pq^{m+k-1}}{q^m}=pq^{k-1},\quad k=1,2,\cdots.$$

表明 X' 亦服从同样几何分布, 称其“无记忆性”或“永远年轻”.

命题 1.6.3 设 X 是取正整数值的随机变量, 若 $\mathbb{P}(X=m+1\mid X>m)$ 与 m 无关, 则 X 服从几何分布.

证明　令 $p=\mathbb{P}(X=m+1\mid X>m)$, 则

$$p=\frac{\mathbb{P}(X=m+1)}{\mathbb{P}(X>m)}=\frac{r_m-r_{m+1}}{r_m},$$

这里 $r_m=\mathbb{P}(X>m)$. 又 $r_0=1$, 可知 $r_m=(1-p)^m$. 进而 $\mathbb{P}(X=m)=r_{m-1}-r_m=p(1-p)^{m-1}$. □

例 1.6.4 (Poisson 分布) 背景：网站访问量、电话总台呼唤数、放射性物质放出粒子数.

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

记为 $X\sim P(\lambda)$.

例 1.6.5 将体积为 V 的物块分为 n 等份, $\Delta V=\dfrac{V}{n}$, 并假设

(1) 每小块 7.5s 内放出 1 个 α 粒子的概率为 $p=\mu\Delta V$ $(\mu>0)$, 放出 2 个及以上 α 粒子的概率为 0.

(2) 各小块是否放出 α 粒子相互独立.

用 X 表示放出粒子数, 设 $\lambda=\mu V$, 则 $p=\dfrac{\lambda}{n}$, 从而

$$\mathbb{P}(X=k)=\mathrm{C}_n^k p^k q^{n-k}=\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\to\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad n\to\infty.$$

定义 1.6.6 设 $X_1,\cdots,X_n$ 为 $(\Omega,\mathcal{F},\mathbb{P})$ 随机变量, 若 $\forall x_1,\cdots,x_n\in\mathbb{R}$ 均有

$$\mathbb{P}(X_1=x_1,\cdots,X_n=x_n)=\mathbb{P}(X_1=x_1)\cdots\mathbb{P}(X_n=x_n),$$

则称 $X_1,\cdots,X_n$ 相互独立.

引理 1.6.7 $X_1,\cdots,X_n$ 相互独立 $\iff$ $F(x_1,\cdots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n)$, $\forall x_1,\cdots,x_n\in\mathbb{R}$.

证明 只需证 $n=2$ 的情形. 设 X,Y 为随机变量, 则

$$F_X(x)=\sum_{u\leqslant x}f_X(u),\quad f_X(x)=F_X(x)-F_X(x^-),$$
$$F_Y(y)=\sum_{u\leqslant y}f_Y(u),\quad f_Y(y)=F_Y(y)-F_Y(y^-).$$

$\Rightarrow$: 由分离性可得

$$F(x,y)=\sum_{x_i\leqslant x,y_j\leqslant y}\mathbb{P}(X=x_i,Y=y_j)=\sum_{x_i\leqslant x,y_j\leqslant y}\mathbb{P}(X=x_i)\mathbb{P}(Y=y_j)=F_X(x)F_Y(y).$$

$\Leftarrow$: 我们有

$$F(x,y)=F_X(x)F_Y(y),\tag{1}$$

$$F(x^-,y)=F_X(x^-)F_Y(y),\tag{2}$$

$$F(x,y^-) = F_X(x)F_Y(y^-), \tag{3}$$

$$F(x^-,y^-) = F_X(x^-)F_Y(y^-). \tag{4}$$

由 $[(1)-(2)]-[(3)-(4)]$ 即得 $f(x,y)=f_X(x)f_Y(y)$. □

例 1.6.8 掷 1 次硬币, $\mathbb{P}(\mathrm{H})=p\in(0,1)$, 记 X,Y 为 H,T 出现的次数. 则 X 与 Y 不独立, 如 $\mathbb{P}(X=1,Y=1)=0\neq\mathbb{P}(X=1)\mathbb{P}(Y=1)$. 但若改成掷 N 次硬币, $N\sim P(\lambda)$, 则 X 与 Y 独立, 这是因为

$$\begin{aligned}\mathbb{P}(X=x,Y=y) &= \mathbb{P}(X=x,Y=y\mid N=x+y)\mathbb{P}(N=x+y)\\ &= \mathrm{C}_{x+y}^{x}p^xq^y\frac{\lambda^{x+y}}{(x+y)!}\,\mathrm{e}^{-\lambda}=\frac{(\lambda p)^x\,\mathrm{e}^{-\lambda p}}{x!}\cdot\frac{(\lambda q)^y\,\mathrm{e}^{-\lambda q}}{y!},\end{aligned}$$

从而

$$\mathbb{P}(X=x)=\sum_y\mathbb{P}(X=x,Y=y)\xlongequal{\text{Taylor 级数}}\frac{(\lambda p)^x\,\mathrm{e}^{-\lambda p}}{x!}.$$

同理,

$$\mathbb{P}(Y=y)=\frac{(\lambda q)^y\,\mathrm{e}^{-\lambda q}}{y!}.$$

这表明 X 与 Y 独立.

7　数学期望

定义 1.7.1 若 $\sum\limits_{x:f(x)>0}|x|f(x)<+\infty$, 则称级数 $\sum\limits_{x:f(x)>0}xf(x)$ 为随机变量 X 的 (数学) 期望, 记为 $\mathbb{E}[X]$.

注 1.7.2 (1) 绝对收敛避免了级数重排后出现两个不同和的问题.

(2) $\mathbb{E}[X]=\sum\limits_{k=1}^{\infty}x_kp_k$. 若存在 $j\neq k$ 使得 $x_j=x_k$, 此求和仍是良定的.

设 X 是一个随机变量, $g:\mathbb{R}\to\mathbb{R}$ 是一个 Borel 可测函数. 令 $Y=g(X)$, 则 $Y(\omega)=g(X(\omega))$. 设 X 有分布列 $f(x)$. 我们有如下定理.

定理 1.7.3 (1 维佚名统计学家公式) $\mathbb{E}[g(X)]=\sum\limits_x g(x)f(x)$, 这里右边级数绝对收敛.

证明 设 $Y=g(X)$, 则

$$f_Y(y)=\mathbb{P}(g(X)=y)=\mathbb{P}\left(\bigcup_{x:g(x)=y}\{X=x\}\right)=\sum_{x:g(x)=y}f(x),$$

则由 Fubini 定理,

$$\mathbb{E}[Y]=\sum_{y} y \sum_{x: g(x)=y} f(x)=\sum_{y} \sum_{x: g(x)=y} g(x) f(x)=\sum_{x} g(x) f(x).$$

$\square$

定义 1.7.4 k 阶矩 $m_k=\mathbb{E}\left[X^k\right]$, 期望 $\mu=\mathbb{E}[X]$, 方差 $\operatorname{Var}(X)=\mathbb{E}\left[(X-\mu)^2\right]$, 标准差 $\sigma=\sqrt{\operatorname{Var}(X)}$, k 阶中心矩 $\sigma_k=\mathbb{E}\left[(X-\mu)^k\right]$.

注 1.7.5 方差可以写成“二阶矩减去一阶矩的平方”:

$$\operatorname{Var}(X)=\sum_{x}(x-\mu)^2 f(x)=\mathbb{E}\left[X^2\right]-2 \mu \mathbb{E}[X]+\mu^2=\mathbb{E}\left[X^2\right]-(\mathbb{E}[X])^2 \leqslant \mathbb{E}\left[X^2\right].$$

下面求常见分布的数字特征.

例 1.7.6 (二项分布) 若 $X \sim B(n, p)$, 记 $q=1-p$, 则有

$$\begin{aligned}
\mathbb{E}[X] & =\sum_{k=1}^{n} k \mathrm{C}_n^k p^k q^{n-k}=\sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \xlongequal{k \rightarrow k+1} \sum_{k=0}^{n-1} \frac{n!}{k!(n-1-k)!} p^{k+1} q^{n-1-k} \\
& =n p \sum_{k=0}^{n-1} \mathrm{C}_{n-1}^k p^k q^{n-k-1} \xlongequal{\text{分布列} \rightarrow \text{求和为 } 1} n p,
\end{aligned}$$

以及

$$\begin{aligned}
\mathbb{E}[X(X-1)] & =\sum_{k=1}^{n} k(k-1) \mathrm{C}_n^k p^k q^{n-k}=\sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k q^{n-k} \\
& =n(n-1) p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-2-k)!} p^k q^{n-2-k}=n(n-1) p^2 .
\end{aligned}$$

进而可得二阶中心矩

$$\mathbb{E}\left[X^2\right]=\mathbb{E}[X]+\mathbb{E}[X(X-1)]=n p[1+(n-1) p]$$

以及方差

$$\operatorname{Var}(X)=\mathbb{E}\left[X^2\right]-(\mathbb{E}[X])^2=n p(1-p)=n p q .$$

定理 1.7.7 ($\mathbb{E}$ 可视作线性算子)

(1) (非负性) $X \geqslant 0 \Longrightarrow \mathbb{E}[X] \geqslant 0$.

(2) (归一性) $\mathbb{E}[1]=1$.

(3) (线性性) 对任意 $a, b \in \mathbb{R}$, $\mathbb{E}[a X+b Y]=a \mathbb{E}[X]+b \mathbb{E}[Y]$.

证明 非负性与归一性是显然的, 下面验证线性性 (不平凡).

令 $A_x=\{X=x\}, B_y=\{Y=y\}$, 则用示性函数有 $X=\sum\limits_x xI_{A_x}, Y=\sum\limits_y yI_{B_y}$,

$$\begin{aligned} aX+bY &= a\sum_x xI_{A_x}+b\sum_y yI_{B_y}=a\sum_{x,y} xI_{A_x}I_{B_y}+b\sum_{x,y} yI_{A_x}I_{B_y} \\ &= a\sum_{x,y} xI_{A_xB_y}+b\sum_{x,y} yI_{A_xB_y}=\sum_{x,y}(ax+by)I_{A_xB_y}, \end{aligned}$$

这里用到了示性函数满足 $I_{A_xB_y}=I_{A_x}I_{B_y}$ 及 $\sum\limits_x I_{A_x}=\sum\limits_y I_{B_y}=I_\Omega\equiv 1$ 的良好性质 (A_xB_y 表示交集). 通过将随机变量写成示性函数线性组合的形式, 我们获得了描述期望的新观点[2]:

$$\mathbb{E}[X]=\sum_x x\mathbb{P}(A_x).$$

由此表示可知

$$\mathbb{E}[aX+bY]=\sum_{x,y}(ax+by)\mathbb{P}(A_xB_y)=a\sum_{x,y}x\mathbb{P}(A_xB_y)+b\sum_{x,y}y\mathbb{P}(A_xB_y)=a\mathbb{E}[X]+b\mathbb{E}[Y],$$

这里第一个等号后的 $ax+by$ 可能出现重复取值, 但根据注1.7.2(2), 这不影响结果; 最后一个等号是由加号两边分别先对 y,x 求和得到的. □

定理 1.7.8 *若 X 与 Y 是相互独立的两个随机变量, 且 $\mathbb{E}[|X|]<+\infty, \mathbb{E}[|Y|]<+\infty$, 则 $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$.*

证明 沿用定理1.7.7证明中的记号, 我们有 $XY=\sum\limits_{x,y} xyI_{A_xB_y}$, 因此

$$\mathbb{E}[XY]=\sum_{x,y}xy\mathbb{P}(A_xB_y)\xlongequal{\text{独立}}\sum_{x,y}xy\mathbb{P}(A_x)\mathbb{P}(B_y)=\mathbb{E}[X]\mathbb{E}[Y].$$

□

定理 1.7.9 (1) $\operatorname{Var}(aX+b)=a^2\operatorname{Var}(X), \forall a,b\in\mathbb{R}$.

(2) $\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)+2\left(\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\right)$. 特别地, 当 X 与 Y 相互独立时, $\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)$.

证明 (1) 我们有

$$\begin{aligned} \operatorname{Var}(aX+b) &= \mathbb{E}\left[(aX+b-\mathbb{E}[aX+b])^2\right]=\mathbb{E}\left[(aX+b-a\mathbb{E}[X]-b)^2\right] \\ &= a^2\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]=a^2\operatorname{Var}(X). \end{aligned}$$

[2] 也就是说, 我们由原先对 x 轴 (即样本空间) 进行分割转为对 y 轴 (即随机变量的取值) 进行分割.

(2) 我们有

$$\begin{aligned}\mathrm{Var}(X+Y)&=\mathbb{E}\left[(X+Y-\mathbb{E}[X]-\mathbb{E}[Y])^2\right]=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]\\&=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\left(\mathbb{E}[XY]-\mu_X\mu_Y\right).\end{aligned}$$

□

例 1.7.10 (不存在期望的随机变量) 设 $x_k=(-1)^k\dfrac{2^k}{k}, \mathbb{P}(X=x_k)=\dfrac{1}{2^k}$ $(k=1,2,\cdots)$. 这时 $\displaystyle\sum_{k=1}^{\infty}x_kp_k=\sum_{k=1}^{\infty}\frac{(-1)^k}{k}=-\ln 2$, 但 $\displaystyle\sum_{k=1}^{\infty}|x_k|p_k=\sum_{k=1}^{\infty}\frac{1}{k}=+\infty$. 故 X 的期望不存在.

8　概率方法

本节内容参考书目：

The Probabilistic Method, Noga Alon, Joel H. Spencer, John Wiley & Sons, Inc, 2016.

例 1.8.1 (概率与期望的联系) $\mathbb{E}[I_A]=\mathbb{P}(A)$.

例 1.8.2 (随机置换) 从 n 阶置换群 $\mathfrak{S}_n$ 中均匀随机选取一个置换 σ, 记 $N(\sigma)$ 为置换 σ 的不动点个数. 求 $\mathbb{P}(N=r)$.

解 令 $A_i=\{\sigma(i)=i\}$, 记 $I_i=I_{A_i}$. 则

$$X=\sum_{\substack{i_1<\cdots<i_r\\ i_{r+1}<\cdots<i_n}} I_{i_1}\cdots I_{i_r}\left(1-I_{i_{r+1}}\right)\cdots\left(1-I_{i_n}\right)$$

为 $\{N=r\}$ 的示性函数. 利用例1.8.1中概率与期望的联系就有

$$\begin{aligned}\mathbb{P}(N=r)&=\mathbb{E}[X]\\&=\mathrm{C}_n^r\mathbb{E}\left[I_1\cdots I_r(1-I_{r+1})\cdots(1-I_n)\right]\\&=\mathrm{C}_n^r\sum_{s=0}^{n-r}(-1)^s\mathrm{C}_{n-r}^s\mathbb{E}\left[I_1\cdots I_rI_{r+1}\cdots I_{r+s}\right]\\&=\mathrm{C}_n^r\sum_{s=0}^{n-r}(-1)^s\mathrm{C}_{n-r}^s\frac{(n-r-s)!}{n!}\\&=\frac{1}{r!}\sum_{s=0}^{n-r}\frac{(-1)^s}{s!}.\end{aligned}$$

□

例 1.8.3 求例1.8.2中随机变量 N 的期望和方差.

解 由 $N=\sum_{k=1}^{n} I_k$ 可得

$$\mathbb{E}[N]=n\mathbb{E}[I_1]=n\cdot\frac{(n-1)!}{n!}=1,$$

以及

$$\begin{aligned}\mathbb{E}\left[N^2\right]&=\mathbb{E}\left[\sum_k I_k\sum_j I_j\right]=\sum_{j,k}\mathbb{E}[I_kI_j]=n\mathbb{E}\left[I_1^2\right]+n(n-1)\mathbb{E}[I_1I_2]\\&=1+n(n-1)\frac{(n-2)!}{n!}=2.\end{aligned}$$

故

$$\operatorname{Var}(N)=\mathbb{E}\left[N^2\right]-(\mathbb{E}[N])^2=1.$$

□

例 1.8.4 (Erdős 概率方法) 正十七边形 17 个顶点中恰有 5 个是红色的. 证明存在 7 个相邻顶点, 其中至少 3 个为红色.

证明 $\Omega=\{1,\cdots,17\}$. 随机取 1 个顶点, 记 $a_i=\begin{cases}1, & i\text{为红色},\\0, & \text{其他}.\end{cases}$ 定义随机变量 X, $X(k)=a_{k+1}+\cdots+a_{k+7}$ (下标取模 17 后落在 Ω 中的数). 则

$$\mathbb{E}[X]=\sum_{k=1}^{17}\frac{1}{17}\left(a_{k+1}+\cdots+a_{k+7}\right)=\frac{5\times 7}{17}>2.$$

我们断言 $\mathbb{P}(X>2)>0$. 否则, $\mathbb{P}(X\leqslant 2)=1$, 进而由期望的非负性得 $\mathbb{E}[X]\leqslant 2$, 矛盾. 这表明 $\{X>2\}$ 非空, 即存在 k, 使得 $X(k)>2$, 亦即 $X(k)\geqslant 3$. □

例 1.8.5 (概率数论) $\Omega_N=\{1,2,\cdots,N\}$, 均匀测度. 令 $X_q(n)=\begin{cases}1, & q\mid n,\\0, & q\nmid n,\end{cases}$ $(n\in\Omega_N)$. 则

$$\mathbb{E}[X_q]=\frac{1}{N}\left\lfloor\frac{N}{q}\right\rfloor\sim\frac{1}{q}.$$

对互素的两个数 p,q,

$$\operatorname{Cov}(X_p,X_q)=\mathbb{E}[X_pX_q]-\mathbb{E}[X_p]\mathbb{E}[X_q]=\frac{1}{N}\left\lfloor\frac{N}{pq}\right\rfloor-\frac{1}{N}\left\lfloor\frac{N}{p}\right\rfloor\cdot\frac{1}{N}\left\lfloor\frac{N}{q}\right\rfloor\sim 0.$$

9 协方差与条件期望

定义 1.9.1 协方差 $\mathrm{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$. 当 $\mathrm{Var}(X)\,\mathrm{Var}(Y)\neq 0$ 时, 相关系数 $\rho(X,Y)=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$.

注 1.9.2 (1) 相关系数是对协方差的规范化处理 (如 X 以 kg 为单位而 Y 以 m 为单位).

(2) 更一般, 对 $\mathbf{X}=(X_1,\cdots,X_n)$, 定义协方差矩阵 $\Sigma=(\sigma_{ij})_{n\times n}$, 其中 $\sigma_{ij}=\mathrm{Cov}(X_i,X_j)$. 我们有 $\Sigma\geqslant 0$(半正定), 这是因为实对称方阵的特征值均非负, 这也可以由以下推导看出. 对 $t_i\in\mathbb{R}$ $(i=1,\cdots,n)$,

$$\begin{aligned}\sum_{i,j=1}^{n}t_it_j\sigma_{ij}&=\sum_{i,j=1}^{n}t_it_j\mathbb{E}\left[(X_i-\mu_i)(X_j-\mu_j)\right]\\&=\mathbb{E}\left[\sum_i t_i(X_i-\mu_i)\sum_j t_j(X_j-\mu_j)\right]\\&=\mathbb{E}\left[\left(\sum_i t_i(X_i-\mu_i)\right)^2\right]\geqslant 0.\end{aligned}$$

这种协方差可理解为向量各个分量之间的关系.

(3) 当 $\mathrm{Cov}(X,Y)=0$ 时, 称 X 与 Y 不相关. 易知, “独立”蕴含了“不相关”.

引理 1.9.3 (2 维佚名统计学家公式) 设 (X,Y) 有联合分布列 f, $g:\mathbb{R}^2\to\mathbb{R}$ 是一个 Borel 可测函数, 则

$$\mathbb{E}\left[g(X,Y)\right]=\sum_{x,y}g(x,y)f(x,y).$$

注 1.9.4 借助 2 维佚名统计学家公式, 即使未知单个随机变量分布列, 也可以求对应期望.

引理 1.9.5 (Cauchy-Schwarz 不等式) $|\mathbb{E}[XY]|\leqslant\sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}$, 等号成立当且仅当存在不全为零的 $a,b\in\mathbb{R}$, 使得 $\mathbb{P}(aX=bY)=1$.

证明 ① 若 $\mathbb{E}\left[X^2\right]=0$, 即 $\sum_x x^2f(x)=0$, 则对任意 x, $x^2f(x)=0$. 进而当 $x\neq 0$ 时 $f_X(x)=0$, 于是 $f_X(0)=1$ 即 $\mathbb{P}(X=0)=1$. 又 $f_X(x)=\sum_y f(x,y)$, 因此当 $x\neq 0$ 时必有 $f(x,y)=0$. 故由 2 维佚名统计学家公式得 $\mathbb{E}[XY]=\sum_{x,y}xyf(x,y)=0$.

② 若 $\mathbb{E}\left[X^2\right]\neq 0$, 由

$$\mathbb{E}\left[(Y-tX)^2\right]=t^2\mathbb{E}\left[X^2\right]-2t\mathbb{E}[XY]+\mathbb{E}\left[Y^2\right]\geqslant 0,\quad\forall t\in\mathbb{R}$$

可得判别式 $\Delta = 4\left(\mathbb{E}[XY]\right)^2 - 4\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right] \leqslant 0$. 等号成立当且仅当存在 $t_0 \in \mathbb{R}$, 使得 $\mathbb{E}\left[(Y-t_0X)^2\right] = 0$, 再由①知这等价于 $\mathbb{P}(Y = t_0X) = 1$. □

定理 1.9.6 (1) $|\rho(X,Y)| \leqslant 1$.

(2) 当 X 与 Y 独立或不相关时, $\rho(X,Y) = 0$.

(3) $\rho(X,Y) = \pm 1 \iff$ 存在 $a, b \in \mathbb{R}$, 使得 $\mathbb{P}(aX+b=Y) = 1$ (即 X 与 Y 几乎处处具有线性关系).

证明 将引理1.9.5中的 X, Y 分别替换成 $X-\mu_X, Y-\mu_Y$ 易证. □

例 1.9.7 (多项分布) $\mathbf{X} = (X_1, \cdots, X_r)$, $\mathbb{P}(X_1 = k_1, \cdots, X_r = k_r) = \dfrac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}$, $\sum\limits_{i=1}^{r} p_i = 1$, $\sum\limits_{i=1}^{r} k_i = n$, $p_i > 0$, $\forall i$.【背景：独立重复 n 次, 每次有 r 种结果, 发生概率分别为 $p_1, \cdots, p_r$.[3]】计算 $\operatorname{Cov}(X_i, X_j), \rho(X_i, X_j)$ $(i \neq j)$.

解 联系概率背景可观察到 $X_i \sim B(n, p_i)$[4], 以及对 $i \neq j$ 有 $X_i + X_j \sim B(n, p_i + p_j)$[5]. 利用二项分布的期望与方差公式 (例1.7.6) 可得

$$\begin{aligned}
\operatorname{Cov}(X_i, X_j) &\xlongequal{\text{定理1.7.9(2)}} \frac{1}{2}\left[\operatorname{Var}(X_i + X_j) - \operatorname{Var}(X_i) - \operatorname{Var}(X_j)\right] \\
&= \frac{1}{2}\left[n(p_i+p_j)(1-p_i-p_j) - np_i(1-p_i) - np_j(1-p_j)\right] \\
&= -np_ip_j,
\end{aligned}$$

以及

$$\rho(X_i, X_j) = \frac{-np_ip_j}{\sqrt{np_i(1-p_i)np_j(1-p_j)}} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}.$$

□

定义 1.9.8 设 (X,Y) 是离散型随机向量. 当 $f_X(x) > 0$ 时, 给定 $X = x$ 下 Y 的条件分布列 $f_{Y|X}(y \mid x) \coloneqq \mathbb{P}(Y = y \mid X = x)$, 条件分布函数 $F_{Y|X}(y \mid x) \coloneqq \mathbb{P}(Y \leqslant y \mid X = x)$ (可验证由此定义的 $F_{Y|X}(y \mid x)$ 当 y 变动时满足定理1.4.6中的 3 条性质, 因此是概率分布函数). 由定义, $f_{Y|X}(y \mid x) = \dfrac{f(x,y)}{f_X(x)}$.

---

[3]我们有多项式展开恒等式

$$\sum_{k_1+\cdots+k_r=n} \frac{n!}{k_1! \cdots k_r!} x_1^{k_1} \cdots x_r^{k_r} = (x_1 + \cdots + x_r)^n.$$

[4]考虑第 i 种结果与其余 $r-1$ 种结果 (把后者视作一个整体).

[5]考虑第 i 种结果和第 j 种结果的并事件与其余 $r-2$ 种结果.

注 1.9.9 当 $f_X(X)=0$ 时, $f(x,y)=0$. 故总可以把 $f_{Y|X}(y\mid x)$ 视作有界量.

定义 1.9.10 给定 $X=x$ 下, Y 关于 X 的条件期望 $\psi(x)=\mathbb{E}\left[Y\mid X=x\right]\coloneqq\sum_y yf_{Y|X}(y\mid x)$, 并称 $\psi(X)$ 为 Y 关于 X 的条件期望[6], 记为 $\mathbb{E}[Y\mid X]$.

注 1.9.11 (1) 虽然使 $\psi(x)$ 有定义的 x 只有至多可列个, 我们仍将 ψ 视作 $\mathbb{R}\to\mathbb{R}$ 的函数.

(2) 两个随机变量的条件期望是一个随机变量. 下面的定理1.9.12就对其求期望.

定理 1.9.12 $\mathbb{E}\left[\mathbb{E}[Y\mid X]\right]=\mathbb{E}[Y]$.

证明

$$\begin{aligned}\mathrm{LHS}&=\mathbb{E}\left[\psi(X)\right]\\&=\sum_x\psi(x)f_X(x)\\&=\sum_x f_X(x)\sum_y yf_{Y|X}(y\mid x)\\&=\sum_y\sum_x yf(x,y)\\&=\sum_y yf_Y(y)\\&=\mathbb{E}[Y].\end{aligned}$$

□

定理1.9.12也可以写成如下的全期望公式.

定理 1.9.13 (全期望公式) $\mathbb{E}[Y]=\sum_x f_X(x)\mathbb{E}\left[Y\mid X=x\right]$.

进一步地, 可对全期望公式进行如下推广.

定理 1.9.14 设 $\psi(X)=\mathbb{E}[Y\mid X]$, 则对“好”[7]函数 $g:\mathbb{R}\to\mathbb{R}$ 可测, 有

$$\mathbb{E}\left[g(X)\psi(X)\right]=\mathbb{E}[Yg(X)].$$

[6] $\psi(X)$ 表示当 x 遍历所有可能值后 $\psi(x)$ 的所有取值.

[7] “好”的标准即这样的函数使等式两边的期望都有意义.

证明

$$
\begin{aligned}
\mathrm{LHS} &= \mathbb{E}\left[g(X)\psi(X)\right] \\
&= \sum_x g(x)\psi(x)f_X(x) \\
&= \sum_x f_X(x)g(x)\sum_y y f_{Y|X}(y \mid x) \\
&= \sum_y \sum_x y f(x,y)g(x) \\
&= \mathbb{E}[Yg(X)].
\end{aligned}
$$

□

注 1.9.15 高等概率论中将此恒等式作为条件期望的定义.

例 1.9.16 鸟下 N 枚蛋, $N \sim P(\lambda)$, 每枚蛋独立地以概率 p 变成小鸟, 记 K 为小鸟总数. 计算 $\mathbb{E}\left[K \mid N\right]$, $\mathbb{E}[K]$, $\mathbb{E}[N \mid K]$.

解 记 $q = 1-p$. 由 $f_{K|N}(k \mid n) = \mathrm{C}_n^k p^k q^{n-k}$ 可得 $\mathbb{E}\left[K \mid N=n\right] = np$, 进而 $\mathbb{E}[K \mid N] = pN$. 由定理1.9.12, $\mathbb{E}[K] = \mathbb{E}\left[\mathbb{E}[K \mid N]\right] = \mathbb{E}[pN] = p\lambda$ (用到了服从以 λ 为参数的 Poisson 分布的随机变量的期望为 λ).

下面先求 $f_{N|K}(n \mid k)$.

$$
\begin{aligned}
f_{N|K}(n \mid k) &= \frac{\mathbb{P}(N=n, K=k)}{\mathbb{P}(K=k)} \\
&= \frac{\mathbb{P}(K=k \mid N=n)\mathbb{P}(N=n)}{\sum\limits_{m=k}^{\infty}\mathbb{P}(K=k \mid N=m)\mathbb{P}(N=m)} \\
&= \frac{\mathrm{C}_n^k p^k q^{n-k}\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}}{\sum\limits_{m=k}^{\infty}\mathrm{C}_m^k p^k q^{m-k}\frac{\lambda^m}{m!}\,\mathrm{e}^{-\lambda}} \\
&\xlongequal{m \to m+k} \frac{\mathrm{C}_n^k p^k q^{n-k}\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}}{\frac{1}{k!}\sum\limits_{m=0}^{\infty}\frac{p^k q^m}{m!}\lambda^{m+k}\,\mathrm{e}^{-\lambda}} \\
&= \frac{\mathrm{C}_n^k p^k q^{n-k}\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}}{\frac{(\lambda p)^k}{k!}\,\mathrm{e}^{-\lambda p}} \\
&= \frac{(\lambda q)^{n-k}\,\mathrm{e}^{-\lambda q}}{(n-k)!}.
\end{aligned}
$$

故

$$\begin{aligned}
\mathbb{E}[N \mid K=k] &= \sum_{n=k}^{\infty} n \frac{(\lambda q)^{n-k}\,\mathrm{e}^{-\lambda q}}{(n-k)!} \\
&= \sum_{n=0}^{\infty} (n+k) \frac{(\lambda q)^{n}\,\mathrm{e}^{-\lambda q}}{n!} \\
&= k\,\mathrm{e}^{-\lambda q} \sum_{n=0}^{\infty} \frac{(\lambda q)^{n}}{n!} + \lambda q\,\mathrm{e}^{-\lambda q} \sum_{n=1}^{\infty} \frac{(\lambda q)^{n-1}}{(n-1)!} \\
&= k + \lambda q.
\end{aligned}$$

于是 $\mathbb{E}[N \mid K] = \lambda q + K$. □

10　随机游走

A drunk man will find his way home but a drunk bird may get lost forever.

——Shizuo Kakutani

$\{S_n\}, S_0 = a \in \mathbb{Z}^d, S_n = S_{n-1} + X_n = a + \sum_{k=1}^{n} X_k, \{X_k\}$ 独立同分布. 当 $d=1$ 时, $\mathbb{P}(X_k = 1) = p, \mathbb{P}(X_k = -1) = q, p+q=1$, 称为直线上的简单随机游走. 若 $p = \frac{1}{2}$, 则称为对称简单随机游走.

定理 1.10.1 设 $\{S_n\}$ 为 $\mathbb{Z}$ 上的简单随机游走, 则有

(1) (空齐性) $\mathbb{P}(S_n = j+b \mid S_0 = a+b) = \mathbb{P}(S_n = j \mid S_0 = a)$.

(2) (时齐性) $\mathbb{P}(S_{n+m} = j \mid S_m = a) = \mathbb{P}(S_n = j \mid S_0 = a)$.

(3) (Markov 性[8]) $\mathbb{P}(S_{n+m} = j \mid S_0 = j_0, \cdots, S_m = j_m) = \mathbb{P}(S_{n+m} = j \mid S_m = j_m)$, 这里等式只考虑两边有意义的情形[9].

证明 (1) 由 $\{X_k\}$ 独立,

$$\begin{aligned}
\mathrm{LHS} &= \frac{\mathbb{P}(S_n = j+b, S_0 = a+b)}{\mathbb{P}(S_0 = a+b)} = \frac{\mathbb{P}\left(\sum_{k=1}^{n} X_k = j-a, S_0 = a+b\right)}{\mathbb{P}(S_0 = a+b)} \\
&= \mathbb{P}\left(\sum_{k=1}^{n} X_k = j-a\right) = \mathrm{RHS}.
\end{aligned}$$

[8] 立足现在, 未来与过去无关.

[9] 因为条件概率要求分母非零, RHS 有意义时 LHS 未必有意义.

(2) 由 $\{X_k\}$ 同分布,

$$\begin{aligned}\text{LHS} &= \frac{\mathbb{P}\left(S_{n+m}=j, S_m=a\right)}{\mathbb{P}(S_m=a)} = \frac{\mathbb{P}\left(\displaystyle\sum_{k=m+1}^{m+n} X_k = j-a, S_m=a\right)}{\mathbb{P}(S_m=a)} \\ &= \mathbb{P}\left(\sum_{k=m+1}^{m+n} X_k = j-a\right) = \mathbb{P}\left(\sum_{k=1}^{n} X_k = j-a\right) = \text{RHS}.\end{aligned}$$

(3) 我们有

$$\begin{aligned}\text{LHS} &= \frac{\mathbb{P}\left(S_{n+m}=j, S_0=j_0, \cdots, S_m=j_m\right)}{\mathbb{P}\left(S_0=j_0, \cdots, S_m=j_m\right)} \\ &= \frac{\mathbb{P}\left(\displaystyle\sum_{k=m+1}^{m+n} X_k = j-j_m, S_0=j_0, \cdots, S_m=j_m\right)}{\mathbb{P}(S_0=j_0, \cdots, S_m=j_m)} \\ &= \mathbb{P}\left(\sum_{k=m+1}^{m+n} X_k = j-j_m\right) = \text{RHS}.\end{aligned}$$

□

一个简单问题　若 $S_0=0$, 则 $\mathbb{P}\left(S_{2n}=0\right)=\mathrm{C}_{2n}^{n} p^n q^n$.

轨道计数　平面表示：$\{(n, S_n) : n=0,1,\cdots\}$. 引入记号：

$$N_n(a,b) = \sharp\{(0,a) \to (n,b)\},$$

$$N_n^0(a,b) = \sharp\{(0,a) \to (n,b)\text{且与 } x \text{ 轴有交点}\}.$$

引理 1.10.2 $N_n(a,b)=\mathrm{C}_n^{\frac{1}{2}(n+b-a)}$.

引理 1.10.3 (反射原理) 若 $a,b>0$, 则 $N_n^0(a,b)=N_n(-a,b)$.

证明 每一条满足 $(0,a)\to(n,b)$ 且与 x 轴有交点的路径都可以通过将从出发到第一次与 x 轴相交的部分关于 x 轴作反射与 $(0,-a)\to(n,b)$ 的路径建立 1-1 对应. □

定理 1.10.4 (投票定理) 若 $b>0$, 则 $\sharp\{(0,0)\to(n,b)\text{且不再过 } x \text{ 轴}\} = \dfrac{b}{n} N_n(0,b)$.

证明 第一步只能向右, 到达 $(1,1)$, 因此所求为

$$\begin{aligned}
\sharp\{(1,1)\to(n,b)\text{且不过 } x \text{ 轴}\} &= \sharp\{(0,1)\to(n-1,b)\text{且不过 } x \text{ 轴}\}\\
&= N_{n-1}(1,b) - N_{n-1}^0(1,b)\\
&\xlongequal{\text{反射原理}} N_{n-1}(1,b) - N_{n-1}(-1,b)\\
&= \binom{n-1}{\frac{n+b-2}{2}} - \binom{n-1}{\frac{n+b}{2}}\\
&= \frac{b(n-1)!}{\left(\frac{n+b}{2}\right)!\left(\frac{n-b}{2}\right)!}\\
&= \frac{b}{n}N_n(0,b).
\end{aligned}$$

□

例 1.10.5 (竞选问题) A 最终得票 a, B 最终得票 b, $a>b$. 求 A 得票始终多于 B 的概率.

解 问题可转化为求 $(0,0)\to(a+b,a-b)$ 的轨道中不再过 x 轴的轨道数占比, 结合投票定理即

$$\frac{\frac{a-b}{a+b}N_{a+b}(0,a-b)}{N_{a+b}(0,a-b)} = \frac{a-b}{a+b}.$$

□

定理 1.10.6 若 $S_0=0$, 则对 $n\geqslant 1$, 有

(1) $\mathbb{P}\left(S_1\cdots S_n\neq 0, S_n=b\right) = \frac{|b|}{n}\mathbb{P}(S_n=b)$.

(2) $\mathbb{P}\left(S_1\cdots S_n\neq 0\right) = \frac{1}{n}\mathbb{E}\left[|S_n|\right]$.

证明 (1) 设 $b>0$, 由投票定理,

$$\mathbb{P}\left(S_1\cdots S_n\neq 0, S_n=b\right) = \frac{b}{n}N_n(0,b)p^{\frac{n+b}{2}}q^{\frac{n-b}{2}} = \frac{b}{n}\mathbb{P}(S_n=b).$$

(2) 利用 (1) 即得

$$\mathbb{P}\left(S_1\cdots S_n\neq 0\right) = \sum_{b\neq 0}\mathbb{P}\left(S_1,\cdots S_n\neq 0, S_n=b\right) = \sum_b \frac{|b|}{n}\mathbb{P}(S_n=b) = \frac{1}{n}\mathbb{E}\left[|S_n|\right].$$

□

11 母函数

A generating function is a clothesline on which we hang up a sequence of numbers for display.

——Herbert Wilf

数列的母函数 $\{a_n\}_{n=0}^{\infty}, G_a(s)=\sum_{n=0}^{\infty} a_n s^n$.

例 1.11.1 $a_k=\mathrm{C}_n^k, k=0,1,\cdots,n, G_a(s)=(1+s)^n$.

例 1.11.2 $\{a_n\}$ 与 $\{b_n\}$ 的卷积 $\{c_n\}$ 定义为 $c_n=\sum_{i+j=n} a_i b_j$, 记为 $a*b$. 我们有 $G_c(s)=G_a(s)G_b(s)$.

例 1.11.3 对称随机游走 $\{S_n\}, S_0=0$. 求 $\mathbb{P}(S_0=S_{2n}=0, S_i \geqslant 0, i=1,\cdots,2n-1)$.

解 记 c_n 为满足条件的轨道数, 则所求为 $\left(\frac{1}{2}\right)^{2n} c_n$. 设游走在 $t=2k\ (k>0)$ 时与 x 轴首次相交, 考虑第 1 步与第 $2k$ 步可知 $0 \to 2k$ 的轨道数为 c_{k-1}, 于是 $c_n=\sum_{k=1}^{n} c_{k-1}c_{n-k}=\sum_{k=0}^{n-1} c_k c_{n-1-k}$, 又 $c_0=1$. 令 $G(s)=\sum_{n=0}^{\infty} c_n s^n$, 则 $\frac{G(s)-1}{s}=G(s)G(s)$. 解得 $G(s)=\frac{1\pm\sqrt{1-4s}}{2s}$. 通过考虑 $G(s)$ 在 $s\to 1$ 时的极限可知应取 $G(s)=\frac{1-\sqrt{1-4s}}{2s}$. 作 Taylor 展开, 有

$$G(s)=\frac{1}{2s}\left[1-1-\sum_{n=1}^{\infty}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-n+1\right)}{n!}(-4s)^n\right]=\sum_{n=0}^{\infty}\frac{(2n-1)!!}{(n+1)!}2^n s^n=\sum_{n=0}^{\infty}\frac{\mathrm{C}_{2n}^n}{n+1}s^n.$$

故 $c_n=\frac{1}{n+1}\mathrm{C}_{2n}^n$, 称为 Catalan 数. □

非负整值随机变量

定义 1.11.4 $G_X(s)=\mathbb{E}\left[s^X\right]$ 称为随机变量 X 的 (概率) 母函数.

注 1.11.5 由佚名统计学家公式, $G_X(s)=\sum_{k=0}^{\infty}\mathbb{P}(X=k)s^k$. 由 $G_X(1)=\sum_{k=0}^{\infty}\mathbb{P}(X=k)=1$ 知 $G_X(s)$ 的收敛半径 $R\geqslant 1$; 又 $G_X(s)$ 系数非负、求和为 1. 这两个性质完全刻画了母函数的性质.

例 1.11.6 (典型分布)

(1) (二项分布) $X\sim B(n,p)$, $G(s)=\sum_{k=0}^{n}\mathrm{C}_n^k p^k q^{n-k}s^k=(ps+q)^n$.

(2) (几何分布) $\mathbb{P}(X=k)=q^{k-1}p, k=1,2,\cdots$, $G(s)=\sum_{k=1}^{\infty}pq^{k-1}s^k=\frac{ps}{1-qs}$.

(3) (Poisson 分布) $X\sim P(\lambda)$, $G(s)=\sum_{k=0}^{\infty}\frac{s^k\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda+\lambda s}$.

母函数的性质

定理 1.11.7 (1) $\mathbb{E}[X]=G'(1)$.

(2) $\mathbb{E}\left[X(X-1)\cdots(X-k+1)\right]=G^{(k)}(1)$.

(3) $\operatorname{Var}(X)=G''(1)+G'(1)\left[1-G'(1)\right]$.

注 1.11.8 这里 $G^{(k)}(1)\coloneqq\lim\limits_{s\to1^-}G^{(k)}(s)$, Abel 第二定理保证它在期望存在时是良定的.

定义 1.11.9 当 X 与 Y 独立时, 称 $Z=X+Y$ 为 X 与 Y 的卷积, f_{X+Y} 为 f_X 与 f_Y 的卷积.

定理 1.11.10 设 $X_1,\cdots,X_n$ 相互独立, 则 $G_{S_n}(s)=\prod\limits_{k=1}^{n}G_{X_k}(s)$, 这里 $S_n=\sum\limits_{k=1}^{n}X_k$.

证明 我们用到以下事实：若随机变量 X 与 Y 独立, 则 $g(X)$ 与 $h(Y)$ 亦独立. 由此,

$$G_{S_n}(s)=\mathbb{E}\left[s^{S_n}\right]=\mathbb{E}\left[s^{X_1}\cdots s^{X_n}\right]=\prod_{k=1}^{n}\mathbb{E}\left[s^{X_k}\right]=\prod_{k=1}^{n}G_{X_k}(s).$$

□

定理 1.11.11 设 $\{X_k\}$ 相互独立同分布, 且 N 与 $\{X_k\}$ 独立, 则对 $S\coloneqq\sum\limits_{k=1}^{N}X_k$ 有

$$G_S(s)=G_N\left(G_{X_1}(s)\right).$$

证明 由定理1.9.12(或直接由全期望公式),

$$\begin{aligned}G_S(s)&=\mathbb{E}\left[s^S\right]=\mathbb{E}\left[\mathbb{E}\left[s^S\mid N\right]\right]=\sum_{n=0}^{\infty}f_N(n)\mathbb{E}\left[s^S\mid N=n\right]\\&=\sum_{n=0}^{\infty}f_N(n)\left(G_{X_1}(s)\right)^n=G_N\left(G_{X_1}(s)\right).\end{aligned}$$

□

注 1.11.12 由定理1.11.11, 我们在一定前提下给函数间的复合运算赋予了概率意义.

定义 1.11.13 随机向量 (X,Y) 的联合母函数 $G(X,Y)=\mathbb{E}\left[s^Xt^Y\right]$.

定理 1.11.14 X 与 Y 独立 $\iff G_{X,Y}(s,t)=G_X(s)G_Y(t)$.

证明 两边作 Taylor 展开, 即证

$$\sum_{i,j=0}^{\infty}\mathbb{P}(X=i,Y=j)s^it^j=\sum_{i=0}^{\infty}\mathbb{P}(X=i)s^i\sum_{j=0}^{\infty}\mathbb{P}(Y=j)t^j.$$

比较 $s^i t^j$ 前系数, 即证

$$\mathbb{P}(X=i,Y=j)=\mathbb{P}(X=i)\mathbb{P}(Y=j),\quad \forall i,j.$$

这即是 X 与 Y 独立. □

例 1.11.15 掷 $k=3$ 枚均匀骰子, 求点数之和为 9 的概率.

解 记 X_i 为第 i 枚骰子的点数, 令 $S_k=\sum_{i=1}^{k}X_i$, 则

$$G_{S_k}(s)=(G_{X_1}(s))^k=\left(\sum_{i=1}^{6}\frac{s^i}{6}\right)^k=\left[\frac{1}{6}\frac{s\left(1-s^6\right)}{1-s}\right]^k.$$

对 $G_{S_3}(s)$ 作 Taylor 展开,

$$G_{S_3}(s)=\frac{s^3}{6^3}\left(1-3s^6+3s^{12}-s^{18}\right)\sum_{n=0}^{\infty}\binom{-3}{n}(-s)^n,$$

因此所求概率即 s^9 前系数

$$\frac{1}{6^3}\left[\binom{-3}{6}-3\right]=\frac{1}{6}\left[\frac{(-3)(-4)(-5)(-6)(-7)(-8)}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}-3\right]=\frac{25}{216}.$$

□

定义 1.11.16 矩母函数 $M_X(t)=\mathbb{E}\left[\mathrm{e}^{tX}\right]$.

注 1.11.17 注意这只是形式上的定义, 实际上此期望未必存在. 特别地, 若存在 $\delta>0$, 使得当 $t\in(-\delta,\delta)$ 时 $M_X(t)$ 良定, 则可以使用许多分析手段.

12　连续型随机变量

God doesn’t play dice.

——Albert Einstein

密度函数 $f:\mathbb{R}\to\mathbb{R}$ 非负可积且 $\int_{-\infty}^{+\infty}f(x)\,\mathrm{d}x=1$.

分布函数 $F(x)=\int_{-\infty}^{x}f(u)\,\mathrm{d}u$.

$\mathbb{P}(X\in B)=\int_B f(x)\,\mathrm{d}x$, 这里 $B\in\mathcal{B}(\mathbb{R})$.

例 1.12.1 (均匀分布) $X \sim U[a,b], f(x)=\dfrac{1}{b-a}, x\in[a,b]$.【背景：概率和区间长度成正比.】

例 1.12.2 (指数分布) $X \sim \operatorname{Exp}(\lambda), f(x)=\lambda\,\mathrm{e}^{-\lambda x}, x>0$. 易知，$F(x)=1-\mathrm{e}^{-\lambda x}, x\geqslant 0$.【背景：百科新词条的时间间隔、旅客进入候机厅的时间间隔、电子产品的寿命等. 推导见例1.12.3.】

例 1.12.3 某产品使用 t 时间后, 在 Δt 时间内失效的概率为 $\lambda\Delta t+o(\Delta t)$. 设其寿命为 X, 则 $\mathbb{P}(X\leqslant t+\Delta t \mid X>t)=\lambda\Delta t+o(\Delta t)$, 即

$$\frac{F(t+\Delta t)-F(t)}{1-F(t)}=\Delta t\,(\lambda+o(1))\,.$$

令 $\Delta t\to 0$, 有

$$F'(t)=\lambda(1-F(t)),$$

又 $F(0)=0$, 解得 $F(t)=1-\mathrm{e}^{-\lambda t}$.

注 1.12.4 若 $X\sim\operatorname{Exp}(\lambda)$, 则

$$\mathbb{P}(X>t+s \mid X>t)=\frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}}=\mathrm{e}^{-\lambda s}\,.$$

与几何分布 (例1.6.2) 类似, 我们称其“无记忆性”或“永远年轻”.

例 1.12.5 (正态分布) $X\sim N\left(\mu,\sigma^2\right), f(x)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. 当 $\mu=0,\sigma=1$ 时称为标准正态分布. 有以下事实：①$x=\mu$ 为对称轴、最大值点.②$x=\mu\pm\sigma$ 为两拐点.【背景：学生成绩、测量误差; $f(x)$ 在分析、方程、数论、几何中很重要, 例如热方程 $\begin{cases}\dfrac{\partial u}{\partial t}=\dfrac{1}{2}\dfrac{\partial^2 u}{\partial x^2},\\ u(0,x)=g(x)\end{cases}$ 的解为 $u(t,x)=\dfrac{1}{\sqrt{2\pi t}}\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{e}^{-\frac{1}{2t}(y-x)^2}\,\mathrm{d}y$.】

例 1.12.6 (Wigner 半圆律) $f(x)=\dfrac{1}{2\pi\sigma^2}\sqrt{4\sigma^2-x^2}, |x|\leqslant 2\sigma$. 由

$$\int\sqrt{4\sigma^2-x^2}\,\mathrm{d}x \xlongequal[\theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]]{x=2\sigma\sin\theta} 4\sigma^2\int\cos^2\theta\,\mathrm{d}\theta=2\sigma^2\int[1+\cos(2\theta)]\,\mathrm{d}\theta=2\sigma^2\left[\theta+\frac{\sin(2\theta)}{2}\right]+C$$

可得

$$\mathbb{P}(X\in(0,\sigma))=\frac{2\sigma^2}{2\pi\sigma^2}\left[\theta+\frac{\sin(2\theta)}{2}\right]\Bigg|_0^{\frac{\pi}{6}}=\frac{1}{6}+\frac{\sqrt{3}}{4\pi}.$$

【背景：在随机矩阵与自由概率中扮演正态分布角色.】

定义 1.12.7 设 $X_1,\cdots,X_n$ 为概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 中随机变量. 称 $X_1,\cdots,X_n$ 相互独立, 若

$$F(x_1,\cdots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n),\quad \forall x_1,\cdots,x_n\in\mathbb{R}.$$

定理 1.12.8 $X_1,\cdots,X_n$ 相互独立当且仅当

$$\mathbb{P}\left(X_1\in B_1,\cdots,X_n\in B_n\right)=\prod_{i=1}^{n}\mathbb{P}\left(X_i\in B_i\right),\quad \forall B_1,\cdots,B_n\in\mathcal{B}(\mathbb{R}).$$

证明 $\Leftarrow$：显然.

$\Rightarrow$：不显然, 参见高等概率论. □

定理 1.12.9 设函数 $g_i:\mathbb{R}\to\mathbb{R}$ $(i=1,\cdots,n)$Borel 可测. 若 $X_1,\cdots,X_n$ 相互独立, 则 $g_1(X_1),\cdots,g_n(X_n)$ 亦独立.

证明 令 $Y_i=g_i(X_i)$, 则

$$\begin{aligned}\mathbb{P}\left(Y_i\in(-\infty,y_i],i=1,\cdots,n\right)&=\mathbb{P}\left(X_i\in\underbrace{g_i^{-1}\left((-\infty,y_i]\right)}_{B_i\in\mathcal{B}(\mathbb{R})},i=1,\cdots,n\right)\\&\xlongequal{\text{定理1.12.8}}\prod_{i=1}^{n}\mathbb{P}(X_i\in B_i)=\prod_{i=1}^{n}\mathbb{P}(Y_i\leqslant y_i).\end{aligned}$$

□

定理 1.12.10 设 $X_1,\cdots,X_n$ 分别有密度函数 $f_1,\cdots,f_n$, 则 $X_1,\cdots,X_n$ 相互独立当且仅当联合密度函数 $f(x_1,\cdots,x_n)=\prod\limits_{i=1}^{n}f_i(x_i)$.

命题 1.12.11 设 $(Y_1,Y_2)=T(X_1,X_2)$, (X_1,X_2) 有联合密度函数 $f(x_1,x_2)$. 当映射

$$T:D\to T(D),\quad (x_1,x_2)\mapsto(y_1,y_2)$$

为 1-1 对应时, 可求出反函数 $x_1=g_1(y_1,y_2),x_2=g_2(y_1,y_2)$. 再设 g_1,g_2 有连续偏导数, 则 (Y_1,Y_2) 的联合密度函数为

$$f\left(g_1(y_1,y_2),g_2(y_1,y_2)\right)\cdot|J|\cdot I_{T(D)},$$

其中 $J=\begin{vmatrix}\dfrac{\partial g_1}{\partial y_1}&\dfrac{\partial g_1}{\partial y_2}\\\dfrac{\partial g_2}{\partial y_1}&\dfrac{\partial g_2}{\partial y_2}\end{vmatrix}$ 为 T^{-1} 的 Jacobi 行列式.

证明 设变量替换 $T: A \to B = T(A)$, 则 $(Y_1, Y_2) \in B \iff (X_1, X_2) \in A$. 因此

$$\begin{aligned}\mathbb{P}\left((Y_1, Y_2) \in B\right) = \mathbb{P}\left((X_1, X_2) \in A\right) &= \iint_A f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \iint_B f\left(x_1(y_1, y_2), x_2(y_1, y_2)\right) \cdot |J|\,\mathrm{d}y_1\,\mathrm{d}y_2.\end{aligned}$$

取 $B = (-\infty, y_1] \times (-\infty, y_2] \cap T(D)$ 即可. □

注 1.12.12 (零测集不影响积分结果) 若 $D_0 \subset D, \mathbb{P}((X_1, X_2) \in D_0) = 1$, T 在 D_0 上是 1-1 映射 (而不要求在 D 上是 1-1 映射), 则命题1.12.11仍成立.

例 1.12.13 设 $X, Y \sim N(0,1)$ 独立, 则 $f(x,y) = \dfrac{1}{2\pi}\,\mathrm{e}^{-\frac{1}{2}(x^2+y^2)}$. 令 $X = R\cos\Theta, Y = R\sin\Theta$, 则 (R, Θ) 的密度函数为 $\dfrac{1}{2\pi} r\,\mathrm{e}^{-\frac{1}{2}r^2}$.

13　数学期望与条件期望

定义 1.13.1 设连续型随机变量 X 有密度函数 f, 当 $\displaystyle\int_{\mathbb{R}} |x| f(x)\,\mathrm{d}x < +\infty$ 时, 称 $\displaystyle\int_{\mathbb{R}} x f(x)\,\mathrm{d}x$ 为 X 的期望, 记为 $\mathbb{E}[X]$.

定义 1.13.2 k 阶矩 $m_k = \mathbb{E}\left[X^k\right]$, 方差 $\operatorname{Var}(X) = \mathbb{E}\left[(X - \mu_X)^2\right]$, 标准差 $\sigma = \sqrt{\operatorname{Var}(X)}$, 协方差 $\operatorname{Cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, 当 $\operatorname{Var}(X)\operatorname{Var}(Y) \neq 0$ 时, 相关系数 $\rho(X, Y) = \dfrac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$.

引理 1.13.3 设连续型随机变量 X 有分布函数 F, 则

$$\mathbb{E}[X] = \int_0^{+\infty} (1 - F(x))\,\mathrm{d}x - \int_{-\infty}^0 F(x)\,\mathrm{d}x.$$

证明 我们有

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{+\infty} x f(x)\,\mathrm{d}x + \int_{-\infty}^0 x f(x)\,\mathrm{d}x \\ &= \int_0^{+\infty} \left(\int_0^x 1\,\mathrm{d}t\right) f(x)\,\mathrm{d}x - \int_{-\infty}^0 \left(\int_x^0 1\,\mathrm{d}t\right) f(x)\,\mathrm{d}x \\ &\xlongequal{\text{Fubini 定理}} \int_0^{+\infty} \int_t^{+\infty} f(x)\,\mathrm{d}x\,\mathrm{d}t - \int_{-\infty}^0 \int_{-\infty}^t f(x)\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^{+\infty} (1 - F(t))\,\mathrm{d}t - \int_{-\infty}^0 F(t)\,\mathrm{d}t.\end{aligned}$$

□

定理 1.13.4 设 $g:\mathbb{R}\to\mathbb{R}$ 是一个 Borel 可测函数, X 和 $g(X)$ 为连续型随机变量且期望存在, 则

$$\mathbb{E}[g(X)]=\int_{\mathbb{R}} g(x)f_X(x)\,\mathrm{d}x.$$

证明 由引理1.13.3,

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_0^{+\infty}\mathbb{P}\left(g(X)>t\right)\,\mathrm{d}t-\int_{-\infty}^0\mathbb{P}\left(g(X)\leqslant t\right)\,\mathrm{d}t\\
&= \int_0^{+\infty}\int_{\{x|g(x)>t\}} f_X(x)\,\mathrm{d}x\,\mathrm{d}t-\int_{-\infty}^0\int_{\{x|g(x)\leqslant t\}} f_X(x)\,\mathrm{d}x\,\mathrm{d}t\\
&\xlongequal{\text{Fubini 定理}} \int_{\{x|g(x)>0\}}\left(\int_0^{g(x)}1\,\mathrm{d}t\right)f_X(x)\,\mathrm{d}x-\int_{\{x|g(x)\leqslant 0\}}\left(\int_{g(x)}^0 1\,\mathrm{d}t\right)f_X(x)\,\mathrm{d}x\\
&= \int_{\mathbb{R}} g(x)f_X(x)\,\mathrm{d}x.
\end{aligned}$$

□

定理 1.13.5 设 (X,Y) 是连续型随机向量, $g:\mathbb{R}\to\mathbb{R}$ 是 Borel 可测函数, 且连续型随机向量 $g(X,Y)$ 的期望存在, 则

$$\mathbb{E}\left[g(X,Y)\right]=\iint_{\mathbb{R}^2} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

注 1.13.6 特别地, 利用积分的线性性, 可得 $\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$.

定理 1.13.7 (Cauchy-Schwarz 不等式) $|\mathbb{E}[XY]|\leqslant\sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}$.

注 1.13.8 将 X 与 Y 换成 $X-\mu_X$ 与 $Y-\mu_Y$ 即得 $|\rho(X,Y)|\leqslant 1$.

例 1.13.9 (正态分布的数字特征) 设 $X\sim N\left(\mu,\sigma^2\right)$.

$$\mathbb{E}[X]=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}(x-\mu+\mu)\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x\xlongequal{x-\mu\to x}\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}x\,\mathrm{e}^{-\frac{1}{2\sigma^2}x^2}\,\mathrm{d}x+\mu=\mu.$$

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}(x-\mu)^2\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x\xlongequal{x-\mu\to x}\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}x^2\,\mathrm{e}^{-\frac{1}{2\sigma^2}x^2}\,\mathrm{d}x\\
&\xlongequal{x=\sigma u}\frac{\sigma^2}{\sqrt{2\pi}}\int_{\mathbb{R}}u^2\,\mathrm{e}^{-\frac{1}{2}u^2}\,\mathrm{d}u\xlongequal{\text{分部积分}}\frac{\sigma^2}{\sqrt{2\pi}}\left(\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2}u^2}\,\mathrm{d}u-u\,\mathrm{e}^{-\frac{1}{2}u^2}\Big|_{-\infty}^{+\infty}\right)\\
&= \frac{\sigma^2}{\sqrt{2\pi}}\left(\sqrt{2\pi}+0\right)=\sigma^2.
\end{aligned}$$

例 1.13.10 (Cauchy 分布) 概率密度函数 $f(x)=\dfrac{1}{\pi}\dfrac{1}{1+x^2}$, 但 $\displaystyle\int_{\mathbb{R}}\frac{|x|}{1+x^2}\,\mathrm{d}x=\int_0^{+\infty}\frac{1}{1+x^2}\,\mathrm{d}x^2=+\infty$, 说明期望不存在. Cauchy 分布的概率密度函数在无穷处衰减速度比正态分布慢.

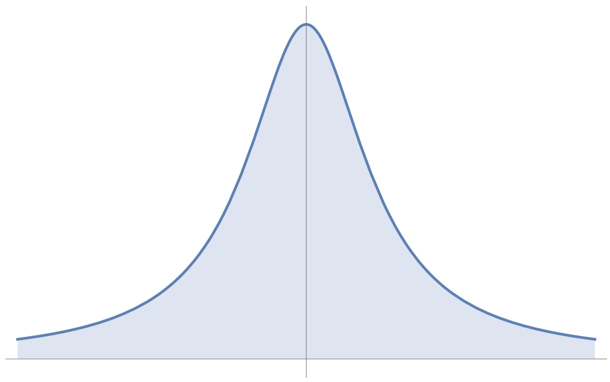


图 1.1: Cauchy 分布的概率密度函数

定义 1.13.11 设连续型随机变量 (X,Y) 有密度函数 $f(x,y)$. 当 $f_X(x)>0$ 时, 给定 $X=x$ 下 Y 的条件密度函数 $f_{Y|X}(y\mid x):=\dfrac{f(x,y)}{f_X(x)}$, 条件分布函数 $F_{Y|X}(y\mid x)=\displaystyle\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}\,\mathrm{d}v$, 条件期望 $\psi(x):=\mathbb{E}[Y\mid X=x]=\displaystyle\int_{\mathbb{R}}yf_{Y|X}(y\mid x)\,\mathrm{d}y$, 并称 $\psi(X)$ 为 Y 关于 X 的条件期望, 记为 $\mathbb{E}[Y\mid X]$.

注 1.13.12 直观上看,

$$\begin{aligned}
\mathbb{P}(Y\leqslant y\mid x<X\leqslant x+\Delta x)&=\frac{\mathbb{P}(Y\leqslant y,x<X\leqslant x+\Delta x)}{\mathbb{P}(x<X\leqslant x+\Delta x)}\\
&=\frac{\displaystyle\int_{-\infty}^{y}\int_{x}^{x+\Delta x}f(u,v)\,\mathrm{d}u\,\mathrm{d}v}{\displaystyle\int_{x}^{x+\Delta x}f_X(u)\,\mathrm{d}u}\\
&\approx\frac{\displaystyle\int_{-\infty}^{y}\Delta x\cdot f(x,v)\,\mathrm{d}v}{\Delta x\cdot f_X(x)}\\
&=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}\,\mathrm{d}v.
\end{aligned}$$

定理 1.13.13 $\mathbb{E}[\mathbb{E}[Y\mid X]]=\mathbb{E}[Y]$. 也可将其写成如下的全期望公式:

$$\mathbb{E}[Y]=\int_{\mathbb{R}}f_X(x)\mathbb{E}[Y\mid X=x]\,\mathrm{d}x.$$

证明 我们证明第二种形式 (全期望公式):

$$\mathrm{RHS}=\int_{\mathbb{R}}f_X(x)\int_{\mathbb{R}}\frac{yf(x,y)}{f_X(x)}\,\mathrm{d}y\,\mathrm{d}x=\iint_{\mathbb{R}^2}yf(x,y)\,\mathrm{d}x\,\mathrm{d}y=\int_{\mathbb{R}}yf_Y(y)\,\mathrm{d}y=\mathbb{E}[Y].$$

□

与离散型随机变量的情形相同, 若对一般地“好”函数 $g:\mathbb{R}\to\mathbb{R}$, 成立

$$\mathbb{E}\left[g(X)\psi(X)\right]=\mathbb{E}\left[Yg(X)\right],$$

可将 $\psi(X)$ 定义为 Y 关于 X 的条件期望.

例 1.13.14 (二元标准正态分布) $f(x,y)=\dfrac{1}{2\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}, x,y\in\mathbb{R}, -1<\rho<1.$

我们先来解释为什么称之为“标准”的. 由 $f(x,y)$ 可求边缘密度函数:

$$\begin{aligned}f_X(x)&=\int_{\mathbb{R}}f(x,y)\,\mathrm{d}y=\frac{1}{\sqrt{1-\rho^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2}x^2-\frac{(y-\rho x)^2}{2(1-\rho^2)}}\,\mathrm{d}(y-\rho x)\\&=\frac{\mathrm{e}^{-\frac{1}{2}x^2}}{\sqrt{1-\rho^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{y^2}{2(1-\rho^2)}}\,\mathrm{d}y=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2},\end{aligned}$$

即 $X\sim N(0,1)$. 同理, $Y\sim N(0,1)$.

再来求 X 与 Y 的协方差.

$$\begin{aligned}\mathrm{Cov}(X,Y)&=\iint_{\mathbb{R}^2}(x-0)(y-0)f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\iint_{\mathbb{R}^2}\left[x(y-\rho x)+\rho x^2\right]f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=\frac{1}{2\pi\sqrt{1-\rho^2}}\underbrace{\int_{\mathbb{R}}x\,\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x}_{0}\int_{\mathbb{R}}(y-\rho x)\,\mathrm{e}^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}\,\mathrm{d}(y-\rho x)+\rho\iint_{\mathbb{R}^2}x^2f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=\frac{\rho}{2\pi\sqrt{1-\rho^2}}\int_{\mathbb{R}}x^2\,\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x\int_{\mathbb{R}}\mathrm{e}^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}\,\mathrm{d}(y-\rho x)=\rho.\end{aligned}$$

由于 $\mathrm{Var}(X)=\mathrm{Var}(Y)=1$, 此时 ρ 即相关系数. 且 $\rho=0\iff X$ 与 Y 独立 (“$\Rightarrow$”由 $f(x,y)=f_X(x)f_Y(y)$ 得到).

例 1.13.15 例1.13.14中 $f_{Y|X}(y\mid x)=\dfrac{1}{\sqrt{2\pi(1-\rho^2)}}\,\mathrm{e}^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$. 我们有

$$\mathbb{E}[Y\mid X=x]=\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{\mathbb{R}}(y-\rho x+\rho x)\,\mathrm{e}^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}\,\mathrm{d}(y-\rho x)=\rho x,$$

因此 $\mathbb{E}[Y\mid X]=\rho X$.

14　多元正态分布

定义 1.14.1 称随机向量 $\mathbf{X}=(X_1,\cdots,X_n)$ 服从多元正态分布, 若它有联合密度函数

$$f(\mathbf{x})=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\,\mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^\mathsf{T}},$$

其中 $\Sigma\in\mathbb{R}^{n\times n}$ 是正定对称矩阵. 记作 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma)$.

注 1.14.2 当 $n=2$ 时, 可写出 $\Sigma=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}, \Sigma^{-1}=\dfrac{1}{1-\rho^2}\begin{pmatrix}\dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2}\\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2}\end{pmatrix}$.

定理 1.14.3 $\mathbb{E}[\mathbf{X}]=\boldsymbol{\mu}, \mathbb{E}\left[(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}(\mathbf{X}-\boldsymbol{\mu})\right]=\Sigma$, 按分量即 $\mathbb{E}[X_i]=\mu_i, \operatorname{Cov}(X_i,X_j)=\sigma_{ij}$.

证明 由于 Σ 是实对称矩阵, 可设 $\Sigma=B^\mathsf{T}\Lambda B$, 其中 B 是正交方阵, $\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$. 令 $\mathbf{y}=(\mathbf{x}-\boldsymbol{\mu})B^\mathsf{T}$, 即 $\mathbf{x}=\boldsymbol{\mu}+\mathbf{y}B$. 先验证 $f(\mathbf{x})$ 为概率密度函数.

$$\begin{aligned}\int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \frac{|\det(B)|}{\sqrt{(2\pi)^n\det(\Sigma)}}\int_{\mathbb{R}^n}\mathrm{e}^{-\frac{1}{2}\mathbf{y}\Lambda^{-1}\mathbf{y}^\mathsf{T}}\,\mathrm{d}\mathbf{y}=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\prod_{k=1}^n\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2\lambda_k}y_k^2}\,\mathrm{d}y_k\\ &= \frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\prod_{k=1}^n\sqrt{2\pi\lambda_k}=1.\end{aligned}$$

再计算期望和方差.

$$\mathbb{E}[X_i]=\int_{\mathbb{R}^n}x_if(\mathbf{x})\,\mathrm{d}\mathbf{x}=\int_{\mathbb{R}^n}\left(\mu_i+\sum_{j=1}^n y_jb_{ji}\right)\frac{\mathrm{e}^{-\frac{1}{2}\mathbf{y}\Lambda^{-1}\mathbf{y}^\mathsf{T}}}{\sqrt{(2\pi)^n\det(\Lambda)}}\,\mathrm{d}\mathbf{y}=\mu_i.$$

$$\begin{aligned}\operatorname{Cov}(X_i,X_j) &= \int_{\mathbb{R}^n}(x_i-\mu_i)(x_j-\mu_j)f(\mathbf{x})\,\mathrm{d}\mathbf{x}=\int_{\mathbb{R}^n}\sum_{k,l=1}^n y_kb_{ki}y_lb_{lj}\frac{\mathrm{e}^{-\frac{1}{2}\mathbf{y}\Lambda^{-1}\mathbf{y}^\mathsf{T}}}{\sqrt{(2\pi)^n\det(\Lambda)}}\,\mathrm{d}\mathbf{y}\\ &= \sum_{k=1}^n\int_{\mathbb{R}}y_k^2b_{ki}b_{kj}\frac{\mathrm{e}^{-\frac{y_k^2}{2\lambda_k}}}{\sqrt{2\pi\lambda_k}}\,\mathrm{d}y_k \xlongequal{\text{分部积分}} \sum_{k=1}^n b_{ki}\lambda_kb_{kj}=\left(B^\mathsf{T}\Lambda B\right)_{ij}=(\Sigma)_{ij}.\end{aligned}$$

□

定理 1.14.4 (线性变换下的不变性) 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma)$, $D\in\mathbb{R}^{n\times n}$ 可逆, 则 $\mathbf{Y}=\mathbf{X}D\sim N\left(\boldsymbol{\mu}D, D^\mathsf{T}\Sigma D\right)$.

证明 记 $B=\{\mathbf{x}\mid x_i\in(a_i,b_i],\forall i\}, A=\{\mathbf{x}\mid \mathbf{x}D\in B\}$, 则

$$\begin{aligned}\mathbb{P}(\mathbf{Y}\in B) &\xlongequal{D\text{可逆}}\mathbb{P}(\mathbf{X}\in A)=\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x}\xlongequal{\mathbf{y}=\mathbf{x}D}\int_B f\left(\mathbf{y}D^{-1}\right)\left|\det\left(D^{-1}\right)\right|\,\mathrm{d}\mathbf{y}\\ &= \int_B\frac{1}{|\det(D)|\sqrt{(2\pi)^n\det(\Sigma)}}\mathrm{e}^{-\frac{1}{2}\left(\mathbf{y}D^{-1}-\boldsymbol{\mu}\right)\Sigma^{-1}\left(\mathbf{y}D^{-1}-\boldsymbol{\mu}\right)^\mathsf{T}}\,\mathrm{d}\mathbf{y}\\ &= \int_B\frac{1}{\sqrt{(2\pi)^n\det\left(D^\mathsf{T}\Sigma D\right)}}\mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}D)\left(D^\mathsf{T}\Sigma D\right)^{-1}(\mathbf{y}-\boldsymbol{\mu}D)^\mathsf{T}}\,\mathrm{d}\mathbf{y}.\end{aligned}$$

故 $\mathbf{Y}\sim N(\boldsymbol{\mu}D, D^\mathsf{T}\Sigma D)$. □

引理 1.14.5 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma), \Sigma=\begin{pmatrix}\Sigma_{11} & O\\ O & \Sigma_{22}\end{pmatrix}$, 其中 Σ_{11},Σ_{22} 均为方阵. 将 $\mathbf{X}$ 及 $\boldsymbol{\mu}$ 对应分段为 $\mathbf{X}=\left(\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right)$ 与 $\left(\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}\right)$, 则 $\mathbf{X}^{(i)}\sim N\left(\boldsymbol{\mu}^{(i)},\Sigma_{ii}\right), i=1,2$.

引理 1.14.6 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma),\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}$, 其中 Σ_{11},Σ_{22} 均为方阵. 将 $\mathbf{X}$ 及 $\boldsymbol{\mu}$ 对应分段为 $\mathbf{X}=\left(\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right)$ 与 $\left(\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}\right)$, 则 $\mathbf{X}^{(1)}\sim N\left(\boldsymbol{\mu}^{(1)},\Sigma_{11}\right)$.

证明 进行分块初等变换

$$\underbrace{\begin{pmatrix}I&O\\-\Sigma_{21}\Sigma_{11}^{-1}&I\end{pmatrix}}_{D^{\mathsf{T}}}\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}\underbrace{\begin{pmatrix}I&-\Sigma_{11}^{-\mathsf{T}}\Sigma_{21}^{\mathsf{T}}\\O&I\end{pmatrix}}_{D}=\begin{pmatrix}\Sigma_{11}&O\\O&*\end{pmatrix}.$$

令 $\mathbf{Y}=\mathbf{X}D$, 则 $\mathbf{Y}\sim N\left(\boldsymbol{\mu}D,D^{\mathsf{T}}\Sigma D\right)$, 但

$$\mathbf{Y}=\mathbf{X}D=\left(\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right)\begin{pmatrix}I&-\Sigma_{11}^{-\mathsf{T}}\Sigma_{21}^{\mathsf{T}}\\O&I\end{pmatrix}=\left(\mathbf{X}^{(1)},\ *\ \right),$$

$$\boldsymbol{\mu}D=\left(\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}\right)\begin{pmatrix}I&-\Sigma_{11}^{-\mathsf{T}}\Sigma_{21}^{\mathsf{T}}\\O&I\end{pmatrix}=\left(\boldsymbol{\mu}^{(1)},\ *\ \right),$$

由引理1.14.5, $\mathbf{X}^{(1)}\sim N\left(\boldsymbol{\mu}^{(1)},\Sigma_{11}\right)$. □

定理 1.14.7 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma)$, $A\in\mathbb{R}^{n\times m}$ 列满秩 (即 $\operatorname{rank}(A)=m$), 则 $\mathbf{Y}=\mathbf{X}A\sim N\left(\boldsymbol{\mu}A,A^{\mathsf{T}}\Sigma A\right)$.

证明 ① 若 $m=n$, 则 A 是可逆方阵, 这即是定理1.14.4.

② 若 $m<n$, 取 B 使得 $D=(A,B)$ 可逆, 则 $\mathbf{X}D=(\mathbf{X}A,\mathbf{X}B)$, 且 $D^{\mathsf{T}}\Sigma D=\begin{pmatrix}A^{\mathsf{T}}\\B^{\mathsf{T}}\end{pmatrix}\Sigma\begin{pmatrix}A&B\end{pmatrix}=\begin{pmatrix}A^{\mathsf{T}}\Sigma A&A^{\mathsf{T}}\Sigma B\\B^{\mathsf{T}}\Sigma A&B^{\mathsf{T}}\Sigma B\end{pmatrix}$. 由定理1.14.4及引理1.14.6, $\mathbf{X}A\sim N\left(\boldsymbol{\mu}A,A^{\mathsf{T}}\Sigma A\right)$. □

注 1.14.8 特别地, 若 $\mathbf{X}$ 的各个分量是独立同标准正态分布时, $\mathbf{X}\sim N(\mathbf{0},I_n)$. 此时对任意 n 阶正交方阵 Q, 有 $\mathbf{X}Q\sim N(\mathbf{0},I_n)$, 这体现出一种对称性 (旋转不变性).

定理 1.14.9 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma)$, 则 $\mathbf{X}$ 各分量独立 $\iff$ Σ 是对角方阵.

15　再谈期望

All epistemologic value of the theory of probability is based on this: that large scale random phenomena in their collective action create strict, non random regularity.

——Andrey Nikolaevich Kolmogorov

本节内容参考书目：

- A Course in Probability Theory, Kai Lai Chung, Academic Press, 2001.
- 实变函数论, 周民强, 北京大学出版社, 2016.

一、记号准备

对于随机变量 X 与其分布函数 F、概率密度函数 f, 我们知道

$$\mathbb{E}[X]=\begin{cases}\displaystyle\sum_x xf(x), & \text{若 } X \text{ 是离散型随机变量},\\ \displaystyle\int_{\mathbb{R}} xf(x)\,\mathrm{d}x, & \text{若 } X \text{ 是连续型随机变量}.\end{cases}$$

现引入记号

$$\mathrm{d}F(x)=\begin{cases}F(x)-F(x^-), & \text{若 } X \text{ 是离散型随机变量},\\ f(x)\,\mathrm{d}x, & \text{若 } X \text{ 是连续型随机变量}.\end{cases}$$

则可以给出期望的统一形式：

$$\mathbb{E}[X]=\int_{\mathbb{R}} x\,\mathrm{d}F(x).$$

而佚名统计学家公式也可以统一为

$$\mathbb{E}[g(X)]=\int_{\mathbb{R}} g(x)\,\mathrm{d}F(x).$$

二、抽象积分

问题：对于一般的概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 及随机变量 X、分布函数 F, 如何定义期望 $\mathbb{E}[X]$？

第一步：对简单随机变量 (即只取有限个值)

可记 $X=\displaystyle\sum_{i=1}^n x_iI_{A_i}$, 其中 $\{A_i\}_{i=1}^n$ 为 Ω 的一个划分, 则 $\mathbb{E}[X]=\displaystyle\sum_{i=1}^n x_i\mathbb{P}(A_i)$.

第二步：对非负随机变量

由 $X\geqslant 0$, 存在简单随机变量列 $\{X_n\},X_n\geqslant 0$, 使得 $X_n\uparrow X$(单调上升收敛到 X). 例如, $X_n=nI_{A_n}+\displaystyle\sum_{j=1}^{n2^n}\frac{j-1}{2^n}I_{A_{n_j}}$, 其中 $A_n=\{X\geqslant n\}$, $A_{n_j}=\left\{\dfrac{j-1}{2^n}\leqslant x<\dfrac{j}{2^n}\right\}$, $j=1,\cdots,n2^n$. 由此定义 $\mathbb{E}[X]=\lim\limits_{n\to\infty}\mathbb{E}[X_n]$. 良定性的保证：根据 Lévy 渐升积分定理, 若 $X_n\uparrow X,Y_n\uparrow X$, 则 $\lim\limits_{n\to\infty}\mathbb{E}[X_n]=\lim\limits_{n\to\infty}\mathbb{E}[Y_n]$.

第三步：对一般随机变量

对一般随机变量 X 进行正负部分解：$X = X^+ - X^-$, 其中 $X^+ = \max\{X, 0\}, X^- = \max\{-X, 0\}$. 当 $\mathbb{E}\left[X^+\right] < +\infty$ 或 $\mathbb{E}\left[X^-\right] < +\infty$(即二者不同为 $+\infty$) 时, 可定义 $\mathbb{E}[X] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$. 我们使用如下统一记号：

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P} \quad 或 \quad \mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega).$$

特别地, 当 $\mathbb{E}[X] \coloneqq \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right] < +\infty$ 时, 称 X 的期望存在.

三、期望算子的性质

定理 1.15.1 (期望算子的基本性质)

(1) (非负性) $X \geqslant 0 \xRightarrow{\because X^- \equiv 0} \mathbb{E}[X] \geqslant 0$.

(2) (规范性) $\mathbb{E}[1] = 1$.

(3) (线性性) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \forall a, b \in \mathbb{R}$.

注 1.15.2 以更高的观点来看, 若一个作用在元素为函数的线性空间上的算子满足以上三条性质, 则它必对应于某一概率空间.

定理 1.15.3 (期望算子的连续性) 设随机变量列 $\{X_n\}$ 逐点收敛于 X：$X_n(\omega) \to X(\omega)$, $\forall \omega \in \Omega$. 在以下三种情形下极限和期望可换序：

(1) 单调收敛定理：若 $X_{n+1}(\omega) \geqslant X_n(\omega) \geqslant 0, \forall n, \omega$, 则 $\mathbb{E}[X_n] \to \mathbb{E}[X], n \to \infty$.

(2) 控制收敛定理：若 $|X_n| \leqslant Y, \forall n, \mathbb{E}[Y] < +\infty$, 则 $\mathbb{E}[X_n] \to \mathbb{E}[X], n \to \infty$.

(3) 有界收敛定理：若存在 $c > 0$ 使得 $|X_n| \leqslant c, \forall n$, 则 $\mathbb{E}[X_n] \to \mathbb{E}[X], n \to \infty$.

注 1.15.4 (1) 由期望算子的规范性, (2) 蕴含着 (3).

(2) 若不是逐点收敛, 即 $X_n(\omega) \to X(\omega), \forall \omega \in \Omega_0$, 但 $\mathbb{P}(\Omega_0) = 1$, 上述结论仍成立.

定理 1.15.5 (Fatou 引理) 设随机变量列 $\{X_n\}$ 满足 $X_n \overset{\text{a.s.}}{\geqslant} 0$, $\forall n$[10], 则 $\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n]$.

注 1.15.6 若 $X_n \overset{\text{a.s.}}{\geqslant} -c$, 其中 $c > 0$, 则上述结论仍成立, 因为可对 $X_n + c$ 应用上述定理.

[10]a.s. 即 almost sure. 意即此事件发生的概率为 1.

四、Lebesgue-Stieltjes 积分

对随机变量 X 与其分布函数 F, 引入 $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ 上的概率测度[11]$\mu_F\left((a,b]\right) := F(b)-F(a)$. 更一般地, $\mu_F\left(\bigsqcup_i (a_i,b_i]\right)=\sum_i [F(b_i)-F(a_i)]$, 且可扩展到 $\mathcal{B}(\mathbb{R})$ 上 (但过于复杂, 此处略去不表). 由此 $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu_F)$ 构成概率空间, 其上的随机变量 $g:\mathbb{R}\to\mathbb{R}$ 就是 Borel 可测 (任一 Borel 集的原象是 Borel 集) 函数, 抽象积分 $\int g\,\mathrm{d}\mu_F$ 或 $\int g\,\mathrm{d}F$ 称为 Lebesgue-Stieltjes 积分. 我们有结论：

$$\mathbb{E}[g(X)]=\int_{\mathbb{R}} g\,\mathrm{d}F.$$

若不是在整个 $\mathbb{R}$ 上积分, 可借助示性函数表示成

$$\int_B g\,\mathrm{d}F := \int_{\mathbb{R}} I_B g\,\mathrm{d}F.$$

对多元情形, 我们也有

$$\mathbb{E}\left[g(X,Y)\right]=\iint_{\mathbb{R}^2} g(x,y)\,\mathrm{d}F(x,y).$$

值得注意的是, 这里积分中集合 B 是否包含“边界点”并不是一件无关紧要的事 (对比 Riemann 积分). 例如离散型随机变量会出现“跳跃点”, 此时是否包含边界显然很重要.

五、独立随机变量之积

定理 1.15.7 设随机变量 X 与 Y 独立且期望均存在, 则 $\mathbb{E}[|XY|]<+\infty$ 且 $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$.

证明 按照定义期望时的“三步走”验证：

① 对简单随机变量, 我们在离散型随机变量中已验证.

② 对非负随机变量, 可取简单随机变量列 $X_n\uparrow X, Y_n\uparrow Y$ 且 X_n 与 Y_n 独立, 则

$$\mathbb{E}[XY]:=\lim_{n\to\infty}\mathbb{E}[X_nY_n]\overset{①}{=\!=}\lim_{n\to\infty}\mathbb{E}[X_n]\mathbb{E}[Y_n]=\mathbb{E}[X]\mathbb{E}[Y].$$

③ 对一般随机变量, 作正负部分解：$X=X^+-X^-, Y=Y^+-Y^-$, 则

$$XY=\left(X^+Y^++X^-Y^-\right)-\left(X^+Y^-+X^-Y^+\right).$$

而两个二元组 $\left\{X^+,X^-\right\}$ 与 $\left\{Y^+,Y^-\right\}$ 独立, 进而

$$\mathbb{E}[XY]=\left(\mathbb{E}\left[X^+Y^+\right]+\mathbb{E}\left[X^-Y^-\right]\right)-\left(\mathbb{E}\left[X^+Y^-\right]+\mathbb{E}\left[X^-Y^+\right]\right)$$

[11]注意概率测度与 Lebesgue 测度等其他测度具有显著差异, 如 $\mu_F(\mathbb{R})=1$.

$$
\begin{aligned}
&= \left(\mathbb{E}\left[X^{+}\right]-\mathbb{E}\left[X^{-}\right]\right)\left(\mathbb{E}\left[Y^{+}\right]-\mathbb{E}\left[Y^{-}\right]\right) \\
&= \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}
$$

$\square$

16　几种收敛

Mathematics consists in proving the most obvious thing in the least obvious way.

——George Pólya

定义 1.16.1 设 $X, X_1, \cdots, X_n$ 为概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量.

(1) 几乎处处收敛/以概率 1 收敛：$\mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\}) = 1$. 记为 $X_n \xrightarrow{\text{a.s.}} X$.

(2) r 阶收敛 $(r \geqslant 1)$：$\mathbb{E}\left[|X_n|^r\right] < +\infty, \forall n$ 且 $\mathbb{E}\left[|X_n - X|^r\right] \to 0, n \to \infty$. 记为 $X_n \xrightarrow{r} X$.

(3) 依概率收敛：$\mathbb{P}(|X_n - X| > \varepsilon) \to 0, \forall \varepsilon > 0$. 记为 $X_n \xrightarrow{\text{P}} X$.

(4) 依分布收敛/弱收敛：$F_{X_n}(x) \to F_X(x), \forall x \in \mathcal{C}_{F_X}$, 其中 $\mathcal{C}_{F_X}$ 为 X 的分布函数 F_X 的全体连续点构成的集合. 记为 $X_n \xrightarrow{\text{D}} X$.

注 1.16.2 (1) 设 $X_n = \dfrac{1}{n}$, 则 $F_{X_n}(x) = \begin{cases} 1, & x \geqslant \dfrac{1}{n}, \\ 0, & x < \dfrac{1}{n}. \end{cases}$ 我们有 $\lim\limits_{n\to\infty} F_{X_n}(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$ 但 $X := \lim\limits_{n\to\infty} X_n \equiv 0$. 我们看到即使是常值随机变量列 $\{X_n\}$ 的分布函数 F_{X_n} 也可能不逐点收敛于 X 的分布函数 F_X, 由此希望定义分布函数弱收敛：$F_n(x) \to F(x), \forall x \in \mathcal{C}_F$. 记为 $F_n \xrightarrow{\text{w}} F$. 故依分布收敛有时也称为弱收敛.

(2) 依分布收敛与样本空间无关 (F_{X_n} 可以不在同一概率空间上), 而其他三种收敛都涉及两个随机变量作差, 不能脱离样本空间来谈.

定理 1.16.3 四种收敛有如下蕴含关系：

$$
\begin{array}{ccc}
X_n \xrightarrow{\text{a.s.}} X & & \\
 & \searrow\!\!\Rightarrow & \\
 & & X_n \xrightarrow{\text{P}} X \Longrightarrow X_n \xrightarrow{\text{D}} X \\
 & \nearrow\!\!\Rightarrow & \\
X_n \xrightarrow{r} X & &
\end{array}
$$

当 $r > s \geqslant 1$ 时, $X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X$.

例 1.16.4 设 $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\dfrac{1}{2}$. 令 $X_n=X, Y=1-X$, 则 X_n, Y 都与 X 同分布. 但 $|X_n-Y|=|2X-1|\equiv 1$, 因此 X_n 依分布收敛但在其他三种意义下均不收敛.

引理 1.16.5 $X_n \xrightarrow{\mathrm{P}} X \implies X_n \xrightarrow{\mathrm{D}} X$.

证明 对任意 $\varepsilon>0$,

$$\begin{aligned} F_n(x) &= \mathbb{P}\left(X_n \leqslant x, X \leqslant x+\varepsilon\right)+\mathbb{P}\left(X_n \leqslant x, X > x+\varepsilon\right) \\ &\leqslant F(x+\varepsilon)+\mathbb{P}\left(|X-X_n|>\varepsilon\right). \end{aligned}$$

对换上面不等式中的 X_n 与 X 并将 x 替换成 $x-\varepsilon$, 又有

$$F(x-\varepsilon) \leqslant F_n(x) \leqslant F(x+\varepsilon)+\mathbb{P}\left(|X_n-X|>\varepsilon\right).$$

结合以上两式就有

$$F(x-\varepsilon)-\mathbb{P}\left(|X_n-X|>\varepsilon\right) \leqslant F_n(x) \leqslant F(x+\varepsilon)+\mathbb{P}\left(|X_n-X|>\varepsilon\right).$$

令 $n\to\infty$ 并取上下极限, 利用依概率收敛就有

$$F(x-\varepsilon) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+\varepsilon).$$

对任意 $x\in\mathcal{C}_F$, 令 $\varepsilon\to 0$ 可得

$$\lim_{n\to\infty} F_n(x)=F(x).$$

□

引理 1.16.6 (重要不等式)

(1) 记 $\|X\|_p=\left(\mathbb{E}\left[|X|^p\right]\right)^{\frac{1}{p}}$, 其中 $p\geqslant 1$.

- Hölder 不等式：$\|XY\|_1 \leqslant \|X\|_p\|Y\|_q$, 其中 p, q 满足共轭关系 $\dfrac{1}{p}+\dfrac{1}{q}=1$.
- Minkowski 不等式：$\|X+Y\|_p \leqslant \|X\|_p+\|Y\|_p$.
- Lyapunov 不等式：$\|X\|_r \geqslant \|X\|_s, \forall r>s\geqslant 1$.

(2) Markov 不等式：$\mathbb{P}(|X|\geqslant a) \leqslant \dfrac{\mathbb{E}[|X|]}{a}, \forall a>0$.[12]

(3) Chebyshev 不等式：$\mathbb{P}\left(|X-\mu_X|\geqslant a\right) \leqslant \dfrac{\mathrm{Var}(X)}{a^2}, \forall a>0$.

[12] $\mathbb{P}(|X|\geqslant a)$ 换成 $\mathbb{P}(|X|>a)$ 不等式仍成立. 这是因为证明中可将 $|X|$ 另分解为 $|X|I_{\{|X|>a\}}+|X|I_{\{|X|\leqslant a\}}$. Markov 不等式可用来估计概率密度函数在无穷远处的收敛速度.

证明 (1) 见数学分析.

(2) 对

$$|X| = |X|I_{\{|X|\geqslant a\}} + |X|I_{\{|X|<a\}}$$

两端取期望即得

$$\mathbb{E}[|X|] \geqslant \mathbb{E}\left[|X|I_{\{|X|\geqslant a\}}\right] \geqslant a\mathbb{P}(|X| \geqslant a).$$

(3) 由 Markov 不等式,

$$\mathbb{P}\left(|X-\mu_X| \geqslant a\right) = \mathbb{P}\left((X-\mu_X)^2 \geqslant a^2\right) \leqslant \frac{\mathbb{E}\left[(X-\mu_X)^2\right]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}.$$

□

注 1.16.7 利用 Markov 不等式证明 Chebyshev 不等式的思想很常用, 例如当 $\mathbb{E}\left[\mathrm{e}^{X^2}\right]$ 存在时, 我们有

$$\mathbb{P}(|X| \geqslant a) = \mathbb{P}\left(\mathrm{e}^{X^2} \geqslant \mathrm{e}^{a^2}\right) \leqslant \frac{\mathbb{E}\left[\mathrm{e}^{X^2}\right]}{\mathrm{e}^{a^2}}, \quad \forall a > 0.$$

引理 1.16.8 (1) 当 $r > s \geqslant 1$ 时, $X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X$.

(2) 当 $r \geqslant 1$ 时, $X_n \xrightarrow{r} X \implies X_n \xrightarrow{\mathrm{P}} X$.

证明 (1) 由 Lyapunov 不等式,

$$\|X_n - X\|_s \leqslant \|X_n - X\|_r \to 0, \quad n \to \infty.$$

(2) 由 Markov 不等式,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}\left(|X_n - X|^r > \varepsilon^r\right) \leqslant \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\varepsilon^r} \to 0, \quad n \to \infty.$$

□

例 1.16.9 设 $\Omega = (0,1]$, $\mathbb{P}$ 为其上的 Lebesgue 测度. 令 $X_n(\omega) = \begin{cases} n^{\frac{1}{r}}, & 0 \leqslant \omega \leqslant \dfrac{1}{n}, \\ 0, & \text{其他}. \end{cases}$ 再令 $X \equiv 0$. 则 $\mathbb{P}(|X_n - X| > \varepsilon) \leqslant \dfrac{1}{n} \to 0, n \to \infty$. 但 $\mathbb{E}\left[|X_n - X|^r\right] = n \cdot \dfrac{1}{n} = 1$, 非 r 阶收敛.

定理 1.16.10 (1) 若 $X_n \xrightarrow{\mathrm{D}} c \in \mathbb{R}$, 则 $X_n \xrightarrow{\mathrm{P}} c$.

(2) 若存在 $K > 0$, 使得 $|X_n| \overset{\text{a.s.}}{\leqslant} K$, 则当 $X_n \xrightarrow{\mathrm{P}} X$ 时有 $X_n \xrightarrow{r} X$.

证明 (1) 由 $X_n \xrightarrow{\mathrm{D}} c$ 可知, 对任意 $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(|X_n - c| > \varepsilon\right) &= \mathbb{P}\left(X_n > c + \varepsilon\right) + \mathbb{P}\left(X_n \leqslant c - \varepsilon\right)\\ &\leqslant \underbrace{\left[1 - \mathbb{P}(X_n \leqslant c + \varepsilon)\right]}_{\to 1-1=0} + \underbrace{\mathbb{P}\left(X_n \leqslant c - \frac{\varepsilon}{2}\right)}_{\to 0} \to 0.\end{aligned}$$

(2) 我们断言：$|X| \overset{\text{a.s.}}{\leqslant} K$. 这是因为, 由

$$\{|X| \leqslant K + \varepsilon\} \quad \supset \quad \underbrace{\{|X_n - X| \leqslant \varepsilon\}}_{\text{当 } n \text{ 充分大时可认为即全空间}} \quad \cap \quad \underbrace{\{|X_n| \leqslant K\}}_{\text{可认为即全空间}}$$

可知 $\mathbb{P}\left(|X| \leqslant K + \varepsilon\right) = 1$, 再令 $\varepsilon \to 0^+$, 利用分布函数右连续性质即得 $\mathbb{P}(|X| \leqslant K) = 1$.

记 $A = \{|X_n - X| \geqslant \varepsilon\}$, 则由

$$|X_n - X|^r = |X_n - X|^r I_A + |X_n - X|^r I_{A^c}$$

两端取期望即得

$$\mathbb{E}\left[|X_n - X|^r\right] \leqslant (2K)^r \mathbb{P}\left(|X_n - X| \geqslant \varepsilon\right) + \varepsilon^r \cdot 1.$$

在上式中先令 $n \to \infty$ 取上极限, 再令 $\varepsilon \to 0$ 即得 $X_n \xrightarrow{r} 0$. □

随机变量之和

定理 1.16.11 以下用 代表 a.s. 或 r 或 P.

(1) 若 $X_n \xrightarrow{} X, X_n \xrightarrow{} Y$, 则 $\mathbb{P}(X = Y) = 1$.

(2) 若 $X_n \xrightarrow{} X, Y_n \xrightarrow{} Y$, 则 $X_n + Y_n \xrightarrow{} X + Y$.

(3) 若将 换成 D, 则 (1)(2) 一般不成立.

证明 (1) 只证 $= r$ 情形. 由 Minkowski 不等式,

$$\|X - Y\|_r \leqslant \|X - X_n\|_r + \|Y - X_n\|_r \to 0, \quad n \to \infty.$$

这表明 $\mathbb{E}\left[|X - Y|^r\right] = 0$. 只需再证明如下事实：若 $X \geqslant 0$ 且 $\mathbb{E}\left[X^r\right] = 0$, 则 $\mathbb{P}(X = 0) = 1$. 由 Markov 不等式,

$$\mathbb{P}(X > \varepsilon) = \mathbb{P}\left(X^r > \varepsilon^r\right) \leqslant \frac{\mathbb{E}\left[X^r\right]}{\varepsilon^r} = 0 \implies \mathbb{P}(X \leqslant \varepsilon) = 1, \quad \forall \varepsilon > 0.$$

再由分布函数的右连续性, 令 $\varepsilon \to 0^+$ 即可证明如上事实.

(2) 只证 $\xrightarrow{}$= P 情形. 由三角不等式,

$$\begin{aligned}\mathbb{P}\left(|(X_n+Y_n)-(X+Y)|>\varepsilon\right) &\leqslant \mathbb{P}\left(\{|X_n-X|+|Y_n-Y|>\varepsilon\}\right)\\ &\leqslant \mathbb{P}\left(\left\{|X_n-X|>\frac{\varepsilon}{2}\right\}\cup\left\{|Y_n-Y|>\frac{\varepsilon}{2}\right\}\right)\\ &\leqslant \mathbb{P}\left(\left\{|X_n-X|>\frac{\varepsilon}{2}\right\}\right)+\mathbb{P}\left(\left\{|Y_n-Y|>\frac{\varepsilon}{2}\right\}\right)\\ &\to 0+0=0.\end{aligned}$$

(3) 设 $\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\dfrac{1}{2}$, 令 X_n 与 X 同分布, 则 $X_n\xrightarrow{\mathrm{D}}X, X_n\xrightarrow{\mathrm{D}}-X$, 但 $\mathbb{P}(X=-X)=0$, 且 X_n+X_n 不依分布收敛于 $X-X=0$. □

注 1.16.12 上面没有给出 $\xrightarrow{}$= a.s. 的证明, 但可以利用两个概率为 1 的事件的交事件概率仍为 1(习题 (1.3.5)), 在交事件上推导 (这时已经转化为函数列逐点收敛问题).

17　几乎处处收敛与 Borel-Cantelli 引理

在定义1.16.1中, 我们是通过“收敛的样本点的集合概率测度为 1”来定义几乎处处收敛的, 现在我们希望换一种方式定义.

由

$$\left\{\omega\in\Omega:\lim_{n\to\infty}X_n(\omega)=X(\omega)\right\}=\underset{\text{对任意 }k\in\mathbb{N}}{\bigcap_{k=1}^{\infty}}\;\underset{\text{存在 }m\in\mathbb{N}}{\bigcup_{m=1}^{\infty}}\;\overset{\text{当 }n\geqslant m\text{ 时}}{\bigcap_{n=m}^{\infty}}\left\{|X_n-X|\leqslant\frac{1}{k}\right\}$$

立即得到 $X_n\xrightarrow{\text{a.s.}}X$ 等价于下面两条之一：

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{|X_n-X|\leqslant\frac{1}{k}\right\}\right)=1,$$
$$\mathbb{P}\left(\bigcup_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\left\{|X_n-X|>\frac{1}{k}\right\}\right)=0.$$

注意到上面第二式对 k 指标对应的事件求并后概率为 0, 因此每一个 k 对应的事件概率均为 0. 由此得到如下引理.

引理 1.17.1 (1) $X_n \xrightarrow{\text{a.s.}} X \iff \mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right) = 0, \forall\varepsilon > 0.$

$$\xLeftrightarrow[\text{引理1.1.14}]{\mathbb{P}\text{ 的连续性}} \lim_{m\to\infty}\mathbb{P}\left(\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right) = 0, \forall\varepsilon > 0.$$

(2) $X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\text{P}} X.$

(3) 若 $\sum_{n=1}^{\infty}\mathbb{P}(|X_n - X| > \varepsilon) < +\infty, \forall\varepsilon > 0$, 则 $X_n \xrightarrow{\text{a.s.}} X$.

对概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 中的事件列 $\{A_n\}$, $\bigcup_{m=n}^{\infty} A_m$ 表示 $A_n, A_{n+1}, \cdots$ 至少 1 个发生, $\bigcap_{m=n}^{\infty} A_m$ 表示 $A_n, A_{n+1}, \cdots$ 同时发生.

定义 $\{A_n\}$ 的上极限事件

$$\limsup_{n\to\infty} A_n \coloneqq \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m = \left\{\omega \in \Omega : \text{有无穷多个 } A_n \text{ 使 } \omega \in A_n\right\},$$

含义为 $\{A_n\}$ 中**有无穷多个发生**, 也常记作 $\{A_n \text{ i.o.}\}$[13].

定义 $\{A_n\}$ 的下极限事件

$$\liminf_{n\to\infty} A_n \coloneqq \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m = \left\{\omega \in \Omega : \text{只有有限多个 } A_n \text{ 使 } \omega \notin A_n\right\},$$

含义为 $\{A_n\}$ 中**只有有限多个不发生**.

定理 1.17.2 (Borel-Cantelli 引理)

(1) 若 $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < +\infty$, 则 $\mathbb{P}(A_n \text{ i.o.}) = 0$.

(2) 若 $\{A_n\}$ 相互独立, 且 $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = +\infty$, 则 $\mathbb{P}(A_n \text{ i.o.}) = 1$.

证明 (1) $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\right) \leqslant \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leqslant \sum_{m=n}^{\infty}\mathbb{P}(A_m) \xrightarrow{n\to\infty} 0.$

(2) 为证 $\mathbb{P}(A_n \text{ i.o.}) = 1$, 只需证

$$\mathbb{P}\left(\left(\limsup_{n\to\infty} A_n\right)^{\mathrm{c}}\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^{\mathrm{c}}\right) = 0,$$

[13] i.o. 即 infinitely often.

进而只需证 $\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^{\mathrm{c}}\right)=0, \forall n$. 但 $\{A_n^{\mathrm{c}}\}$ 相互独立, 利用不等式 $1-t \leqslant \mathrm{e}^{-t}$, 有

$$\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^{\mathrm{c}}\right)=\prod_{m=n}^{\infty}\left[1-\mathbb{P}(A_m)\right] \leqslant \prod_{m=n}^{\infty} \mathrm{e}^{-\mathbb{P}(A_m)}=\exp\left(-\sum_{m=n}^{\infty} \mathbb{P}(A_m)\right) \rightarrow 0.$$

□

注 1.17.3 若 $\{A_n\}$ 相互独立, 则 $\mathbb{P}(A_n \text{ i.o.})$ 要么为 0 要么为 1, 这是一种 0-1 律. 下面给出 Borel-Cantelli 引理的一个直观例子. 掷一枚均匀硬币无穷多次, 设 $A_n=\{$第 n 次正面朝上$\}$, 则 $\{A_n\}$ 相互独立, 且 $\sum_{n=1}^{\infty} \mathbb{P}(A_n)=+\infty$, 根据 Borel-Cantelli 引理, $\mathbb{P}(A_n \text{ i.o.})=1$, 而

$$\begin{aligned} A_n \text{ i.o.} &= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty}\{\text{第 } n \text{ 次正面朝上}\}=\bigcap_{n=1}^{\infty}\{\text{从第 } n \text{ 次往后有正面出现}\} \\ &= \big\{\text{有无穷多次正面朝上}\big\}, \end{aligned}$$

这说明出现无穷多次正面朝上的概率为 1, 进而只出现有限多次正面朝上的概率为 0.

引理 1.17.4 设随机变量列 $\{X_n\}$ 同分布, 且 $\mathbb{E}[|X_1|]<+\infty$, 则

(1) $\sum_{n=1}^{\infty} \mathbb{P}\left(|X_1| \geqslant n\right) \leqslant \mathbb{E}[|X_1|] \leqslant \sum_{n=0}^{\infty} \mathbb{P}\left(|X_1| \geqslant n\right)$.

(2) 设 $Y_n=X_n I_{\{|X_n| \leqslant n\}}$, $\{a_n\}$ 是发散到 $+\infty$ 的正项数列, 则 $\sum_{k=1}^{n} \frac{1}{a_n}\left(X_k-Y_k\right) \xrightarrow{\text{a.s.}} 0$.

证明 (1) 我们有

$$\begin{aligned} \mathbb{E}[|X_1|] &= \mathbb{E}\left[\sum_{m=0}^{\infty}|X_1| I_{\{m \leqslant |X_1|<m+1\}}\right] \leqslant \sum_{m=0}^{\infty}(m+1) \mathbb{P}\left(m \leqslant |X_1|<m+1\right) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{m} \mathbb{P}\left(m \leqslant |X_1|<m+1\right)=\sum_{n=0}^{\infty} \sum_{m=n}^{\infty} \mathbb{P}\left(m \leqslant |X_1|<m+1\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}\left(|X_1| \geqslant n\right), \end{aligned}$$

以及

$$\begin{aligned} \mathbb{E}[|X_1|] &= \mathbb{E}\left[\sum_{m=0}^{\infty}|X_1| I_{\{m \leqslant |X_1|<m+1\}}\right] \geqslant \sum_{m=0}^{\infty} m \mathbb{P}\left(m \leqslant |X_1|<m+1\right) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{m-1} \mathbb{P}\left(m \leqslant |X_1|<m+1\right)=\sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} \mathbb{P}\left(m \leqslant |X_1|<m+1\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}\left(|X_1| \geqslant n+1\right)=\sum_{n=1}^{\infty} \mathbb{P}\left(|X_1| \geqslant n\right). \end{aligned}$$

(2) 因为

$$\sum_{k=1}^{\infty}\mathbb{P}(X_k\neq Y_k)=\sum_{k=1}^{\infty}\mathbb{P}\left(|X_k|\geqslant k\right)=\sum_{k=1}^{\infty}\mathbb{P}\left(|X_1|\geqslant k\right)\leqslant\mathbb{E}[|X_1|]<+\infty,$$

由 Borel-Cantelli 引理, $\mathbb{P}\left(X_k\neq Y_k \text{ i.o.}\right)=0$, 即 $\mathbb{P}\Big(\{X_k\neq Y_k\}\text{只发生有限次}\Big)=1$. 于是 $\sum_{k=1}^{n}\frac{1}{a_n}\left(X_k-Y_k\right)\xrightarrow{\text{a.s.}}0$. □

注 1.17.5 由引理1.17.4(1), 非负随机变量 X 的期望存在 $\Longleftrightarrow \sum_{n=0}^{\infty}\mathbb{P}(X\geqslant n)<+\infty \Longleftrightarrow \sum_{n=0}^{\infty}\mathbb{P}(X>n)<+\infty$. 又注意到若 X 为非负整值随机变量, $\mathbb{E}[X]=\sum_{n=0}^{\infty}\mathbb{P}(X>n)$.

18　大数定律

定理 1.18.1 (弱大数律) 设 $\{X_n\}$ 相互独立且同分布, 期望 $\mu=\mathbb{E}[X_k]$、方差 $\mathrm{Var}(X_k)=\sigma^2<+\infty$ 均存在. 记 $S_n=\sum_{k=1}^{n}X_k$, 则 $\frac{S_n}{n}\xrightarrow{2}\mu$ 且 $\frac{S_n}{n}\xrightarrow{\text{P}}\mu$.

证明 我们有

$$\mathrm{Var}\left(\frac{S_n}{n}\right)=\frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k)=\frac{\sigma^2}{n}\to 0,$$

由 Chebyshev 不等式, 还有

$$\mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\leqslant\frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2}\to 0,\quad \forall\varepsilon>0.$$

□

问题　方差不存在时当如何?

截尾术　设 $Y_n(\omega)=\begin{cases}X_n(\omega), & |X_n(\omega)|\leqslant n,\\ 0, & |X_n(\omega)|>n.\end{cases}$ 令 $S_n=\sum_{k=1}^{n}X_k$, $T_n=\sum_{k=1}^{n}Y_k$, 则对任一发散到无穷的数列 $\{a_n\}$, 有 $\frac{S_n}{a_n}$ 与 $\frac{T_n}{a_n}$ 收敛于同一个数或同时发散到 $\pm\infty$(这是引理1.17.4(2)). 从习题 (5.6.4) 可以看到, 随机变量矩的信息与其尾部收敛速度有密切联系, 因此这里截断时与 n 作比较是“恰到好处”的. 另外, 当 $\{X_n\}$ 相互独立时 $\{Y_n\}$ 也相互独立.

定理 1.18.2 (Khinchine 弱大数律) 设 $\{X_n\}$ 相互独立且同分布, 期望 $\mu=\mathbb{E}[X_k]$ 存在. 记 $S_n=\sum_{k=1}^{n}X_k$, 则 $\frac{S_n}{n}\xrightarrow{\text{P}}\mu$.

证明 根据截尾术原理, 只需对 $\{Y_n\}$ 证明.

$$\begin{aligned}\mathbb{P}\left(\left|\frac{T_n}{n}-\frac{\mathbb{E}[T_n]}{n}\right|>\varepsilon\right)&\leqslant\frac{1}{\varepsilon^2n^2}\operatorname{Var}(T_n)=\frac{1}{\varepsilon^2n^2}\sum_{k=1}^{n}\operatorname{Var}(Y_k)\\&=\frac{1}{\varepsilon^2n^2}\sum_{k=1}^{n}\left[\mathbb{E}\left[Y_k^2\right]-(\mathbb{E}\left[Y_k\right])^2\right]\leqslant\frac{1}{\varepsilon^2n^2}\underbrace{\sum_{k=1}^{n}\mathbb{E}\left[X_k^2I_{\{|X_k|\leqslant k\}}\right]}_{Q_n}.\end{aligned}$$

取 $a_n=n^\delta$, 其中 $\delta\in(0,1)$, 则

$$Q_n=\sum_{k\leqslant a_n}\mathbb{E}\left[X_k^2I_{\{|X_k|\leqslant k\}}\right]+\sum_{a_n<k\leqslant n}\mathbb{E}\left[X_k^2I_{\{|X_k|\leqslant k\}}\right]=:Q_n^{(1)}+Q_n^{(2)}.$$

我们有如下估计：

$$Q_n^{(1)}\leqslant a_n\sum_{k\leqslant a_n}\mathbb{E}\left[|X_k|I_{\{|X_k|\leqslant a_n\}}\right]=a_n\sum_{k\leqslant a_n}\mathbb{E}\left[|X_1|I_{\{|X_1|\leqslant a_n\}}\right],$$

$$\begin{aligned}Q_n^{(2)}&=\sum_{a_n<k\leqslant n}\mathbb{E}\left[X_1^2\left(I_{\{|X_1|\leqslant a_n\}}+I_{\{a_n<|X_1|\leqslant k\}}\right)\right]\\&\leqslant a_n\sum_{a_n<k\leqslant n}\mathbb{E}\left[|X_1|I_{\{|X_1|\leqslant a_n\}}\right]+n\sum_{a_n<k\leqslant n}\mathbb{E}\left[|X_1|I_{\{|X_1|>a_n\}}\right].\end{aligned}$$

于是

$$Q_n\leqslant na_n\mathbb{E}\left[|X_1|I_{\{|X_1|\leqslant a_n\}}\right]+n^2\mathbb{E}\left[|X_1|I_{|X_1|>a_n}\right].$$

因为 $|X_1|I_{\{|X_1|>a_n\}}\leqslant|X_1|$, $\mathbb{E}[X_1]$ 存在, 由控制收敛定理知 $\lim\limits_{n\to\infty}\mathbb{E}\left[|X_1|I_{\{|X_1|>a_n\}}\right]=0$. 从而

$$\frac{T_n-\mathbb{E}[T_n]}{n}\xrightarrow{\mathrm{P}}0.$$

另一方面, 再由控制收敛定理,

$$\mathbb{E}[Y_k]=\mathbb{E}\left[X_1I_{\{|X_1|\leqslant k\}}\right]\to\mu,\quad n\to\infty.$$

于是由 Stolz 定理,

$$\frac{1}{n}\mathbb{E}[T_n]=\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[Y_k]\to\mu,\quad n\to\infty.$$

进而 $\dfrac{T_n}{n}\xrightarrow{\mathrm{P}}\mu$ 即 $\dfrac{S_n}{n}\xrightarrow{\mathrm{P}}\mu$. □

注 1.18.3 由于二阶矩至多同时涉及两个随机变量的乘积, 相互独立可改为两两独立.

定理 1.18.4 设 $\{X_k\}$ 相互独立, 期望 $\mu=\mathbb{E}[X_k]$ 存在, 方差一致有界：$\operatorname{Var}(X_k)\leqslant C$, $\forall k$. 记 $S_n=\sum_{k=1}^{n}X_k$, 则 $\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mu$.

证明 不妨设 $\mu=0$. 由 Chebyshev 不等式,

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{|S_{n^2}|}{n^2}>\varepsilon\right)\leqslant\sum_{n=1}^{\infty}\frac{\operatorname{Var}(S_{n^2})}{\varepsilon^2n^4}\leqslant\sum_{n=1}^{\infty}\frac{n^2C}{\varepsilon^2n^4}=\frac{C}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{n^2}<+\infty,\quad\forall\varepsilon.$$

由 Borel-Cantelli 引理, 子列 $\frac{S_{n^2}}{n^2}\xrightarrow{\text{a.s.}}0$. 记 $M_n\coloneqq\max_{n^2<k\leqslant(n+1)^2}|S_k-S_{n^2}|$, 对任意 $k\in\left(n^2,(n+1)^2\right]$,

$$\begin{aligned}\mathbb{E}\left[|S_k-S_{n^2}|^2\right]&=\mathbb{E}\left[|X_{n^2+1}+X_{n^2+2}+\cdots+X_k|^2\right]\\&=\mathbb{E}\left[X_{n^2+1}^2\right]+\mathbb{E}\left[X_{n^2+2}^2\right]+\cdots+\mathbb{E}\left[X_k^2\right]+2\mathrm{C}_{k-n^2}^2\mu^2\\&=\operatorname{Var}\left(X_{n^2+1}\right)+\operatorname{Var}\left(X_{n^2+2}\right)+\cdots+\operatorname{Var}\left(X_k\right)\end{aligned}$$

随 k 单调递增, 因此

$$\mathbb{E}\left[M_n^2\right]\leqslant\sum_{k=n^2+1}^{(n+1)^2}\mathbb{E}\left[|S_k-S_{n^2}|^2\right]\leqslant(2n+1)\mathbb{E}\left[\left|S_{(n+1)^2}-S_{n^2}\right|^2\right]\leqslant(2n+1)^2C.$$

由 Markov 不等式,

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{M_n}{n^2}>\varepsilon\right)=\sum_{n=1}^{\infty}\mathbb{P}\left(M_n^2>\varepsilon^2n^4\right)\leqslant\sum_{n=1}^{\infty}\frac{\mathbb{E}\left[M_n^2\right]}{\varepsilon^2n^4}\leqslant\sum_{n=1}^{\infty}\frac{(2n+1)^2C}{\varepsilon^2n^4}<+\infty,\quad\forall\varepsilon>0,$$

由 Borel-Cantelli 引理, $\frac{M_n}{n^2}\xrightarrow{\text{a.s.}}0$. 因此对任意 $k\in\left(n^2,(n+1)^2\right]$,

$$\left|\frac{S_k}{k}\right|=\left|\frac{S_k-S_{n^2}}{k}+\frac{S_{n^2}}{k}\right|\leqslant\frac{M_n}{n^2}+\frac{|S_{n^2}|}{n^2}\xrightarrow{\text{a.s.}}0.$$

□

推论 1.18.5 设随机变量 $\{X_k\}$ 独立且均服从 Bernoulli 两点分布, 则 $\frac{1}{n}\sum_{k=1}^{n}X_k\to p\coloneqq\mathbb{E}[X_1]$.

定理 1.18.6 (Kolmogorov 强大数律) 设 $\{X_k\}$ 相互独立且同分布, 记 $S_n=\sum_{k=1}^{n}X_k$. 则 $\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mu\in\mathbb{R}\iff\mathbb{E}[|X_1|]<+\infty$ 且 $\mu=\mathbb{E}[X_1]$.

证明 $\Leftarrow$: 对 X_k 作正负部分解 $X_k=X_k^+-X_k^-$, 其中 $X_k^+=\max\{0,X_k\}$, $X_k^-=\max\{0,-X_k\}$. 则 $\{X_k^+\}$ 相互独立且同分布, $\{X_k^-\}$ 相互独立且同分布, 且期望分别为 $\mu^+=\max\{0,X_k\}$ 和

$\mu^- = \max\{0, -\mu\}$. 因此只需对非负随机变量证明 “$\Leftarrow$”, 进而可推广至一般随机变量. 下设 $\{X_k\}$ 非负, 由截尾术原理, 只需对截断后的随机变量列 $\{Y_k\}$ 证明. 记 $T_n = \sum_{k=1}^{n} Y_k$.

对 $\alpha > 1$, 令 $\beta_k = \lfloor \alpha^k \rfloor$, 则 $\alpha^k - 1 < \beta_k \leqslant \alpha^k$, 且存在只依赖于 α 的常数 C_α, 使得

$$\sum_{k=m}^{\infty} \frac{1}{\beta_k^2} \leqslant \frac{C_\alpha}{\beta_m^2}, \quad \forall m \geqslant 1.$$

由 Chebyshev 不等式,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{|T_{\beta_n} - \mathbb{E}[T_{\beta_n}]|}{\beta_n} > \varepsilon\right) &\leqslant \sum_{n=1}^{\infty} \frac{\operatorname{Var}(T_{\beta_n})}{\varepsilon^2 \beta_n^2} = \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 \beta_n^2} \sum_{k=1}^{\beta_n} \operatorname{Var}(Y_k) \\
&\leqslant \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 \beta_n^2} \sum_{k=1}^{\beta_n} \mathbb{E}\left[Y_k^2\right] \xlongequal{\text{求和换序}} \sum_{k=1}^{\infty} \frac{1}{\varepsilon^2} \mathbb{E}\left[Y_k^2\right] \sum_{n:\beta_n \geqslant k} \frac{1}{\beta_n^2} \\
&\leqslant \frac{C_\alpha}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbb{E}\left[Y_k^2\right].
\end{aligned}$$

我们断言最后的级数收敛. 记 $B_{kj} = \{j-1 < X_k \leqslant j\}$, 则

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbb{E}\left[Y_k^2\right] &= \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{j=1}^{k} \mathbb{E}\left[X_k^2 I_{B_{kj}}\right] \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{j=1}^{k} j^2 \mathbb{P}(B_{kj}) \\
&\leqslant \sum_{j=1}^{\infty} j^2 \sum_{k=j}^{\infty} \frac{1}{k^2} \mathbb{P}(B_{1j}) \leqslant 2 \sum_{j=1}^{\infty} j \mathbb{P}(B_{1j}) \\
&= \left(\sum_{j=1}^{\infty} (j-1) \mathbb{P}(B_{1j}) + \sum_{j=1}^{\infty} \mathbb{P}(B_{1j})\right) \leqslant 2(\mathbb{E}[X_1] + 1) < +\infty.
\end{aligned}$$

“$\leqslant$” 放缩方式如下：当 $j = 1$ 时, $\sum_{k=j}^{\infty} \frac{j}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} < 2$; 当 $j \geqslant 2$ 时,

$$\sum_{k=j}^{\infty} \frac{j}{k^2} = \sum_{k=j}^{\infty} \frac{j}{k(k-1)} = j \sum_{k=j}^{\infty} \left(\frac{1}{k-1} - \frac{1}{k}\right) = \frac{j}{j-1} \leqslant 2.$$

这就证明了断言. 于是由 Borel-Cantelli 引理, 子列

$$\frac{T_{\beta_n} - \mathbb{E}[T_{\beta_n}]}{\beta_n} \xrightarrow{\text{a.s.}} 0.$$

另一方面, 由控制收敛定理, $\lim_{k\to\infty} \mathbb{E}\left[X_1 I_{\{X_1 \leqslant k\}}\right] = \mathbb{E}[X_1]$. 再由 Stolz 定理得 $\lim_{n\to\infty} \frac{\mathbb{E}[T_n]}{n} = \mathbb{E}[X_1]$. 故 $\lim_{n\to\infty} \frac{\mathbb{E}[T_{\beta_n}]}{\beta_n} = \mathbb{E}[X_1]$, 进而 $\frac{1}{\beta_n} T_{\beta_n} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$.

对任意正整数 k, 设 $k\in[\beta_n,\beta_{n+1})$, 由

$$\frac{T_{\beta_n}}{\beta_{n+1}}\leqslant\frac{T_k}{k}\leqslant\frac{T_{\beta_{n+1}}}{\beta_n}$$

即

$$\frac{\beta_n}{\beta_{n+1}}\frac{T_{\beta_n}}{\beta_n}\leqslant\frac{T_k}{k}\leqslant\frac{T_{\beta_{n+1}}}{\beta_{n+1}}\frac{\beta_{n+1}}{\beta_n}$$

取上下极限就得到

$$\frac{\mu}{\alpha}\leqslant\liminf_{k\to\infty}\frac{T_k}{k}\leqslant\limsup_{k\to\infty}\frac{T_k}{k}\leqslant\alpha\mu.$$

再令 $\alpha\to1^+$ 即得 $\lim\limits_{k\to\infty}\frac{T_k}{k}=\mu$.

$\Rightarrow$：由 $\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mu\in\mathbb{R}$ 可得

$$\frac{X_n}{n}=\frac{S_n-S_{n-1}}{n}=\frac{S_n}{n}-\frac{S_{n-1}}{n-1}\cdot\frac{n-1}{n}\xrightarrow{\text{a.s.}}0,$$

因此

$$\sum_{n=1}^{\infty}\mathbb{P}\left(|X_n|\geqslant n\right)<+\infty.$$

否则, 由 Borel-Cantelli 引理, $\mathbb{P}\left(|X_n|\geqslant n\text{ i.o.}\right)=1$, 与 $\frac{X_n}{n}\xrightarrow{\text{a.s.}}0$ 矛盾. 由注1.17.5即知 $\mathbb{E}[|X_1|]<+\infty$ 即 X_1 期望存在. 利用“$\Leftarrow$”即知 $\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mathbb{E}[X_1]$, 即 $\mu=\mathbb{E}[X_1]$. □

注 1.18.7 以上是 Kolmogorov 强大数律的一个较初等证明, 利用 Kolmogorov 不等式还可以给出另一种证明.

定理 1.18.8 (Khinchine 重对数律) 设 $\{X_k\}$ 相互独立且同分布, $\mathbb{E}[X_k]=0$, $\mathrm{Var}(X_k)=1$, $\forall k$. 记 $S_n=\sum\limits_{k=1}^{n}X_k$, 则

$$\limsup_{n\to\infty}\frac{S_n}{\sqrt{2n\ln\ln n}}\overset{\text{a.s.}}{=\!=}1,\quad\liminf_{n\to\infty}\frac{S_n}{\sqrt{2n\ln\ln n}}\overset{\text{a.s.}}{=\!=}-1.$$

注 1.18.9 (1) 钟开莱评价此结果为 “a growing achievement in classical probability”. 由重对数律可知, 对任意 $\delta>0$, $\lim\limits_{n\to\infty}\frac{S_n}{n^{\frac{1}{2}+\delta}}=0$.

(2) 若 $\mathbb{E}[X_k]=\mu,\mathrm{Var}(X_k)=\sigma^2$, 则上述定理变为

$$\limsup_{n\to\infty}\frac{S_n-n\mu}{\sigma\sqrt{2n\ln\ln n}}\overset{\text{a.s.}}{=\!=}1,\quad\liminf_{n\to\infty}\frac{S_n-n\mu}{\sigma\sqrt{2n\ln\ln n}}\overset{\text{a.s.}}{=\!=}-1.$$

19　特征函数

What we know is not much. What we do not know is immense.

——Pierre-Simon Laplace

定义 1.19.1 设 X,Y 为概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量, 称 $Z=X+\mathrm{i}\,Y$ 为复随机变量, 并定义其期望为 $\mathbb{E}[Z]=\mathbb{E}[X]+\mathrm{i}\,\mathbb{E}[Y]$.

注 1.19.2 (1) 复随机变量即 2 维随机向量.

(2) $Z_1=X_1+\mathrm{i}\,Y_1$ 与 $Z_2=X_2+\mathrm{i}\,Y_2$ 独立是指 (X_1,Y_1) 与 (X_2,Y_2) 独立, 也即

$$\mathbb{P}\left(X_1\leqslant x_1,Y_1\leqslant y_1,X_2\leqslant x_2,Y_2\leqslant y_2\right)=\mathbb{P}\left(X_1\leqslant x_1,Y_1\leqslant y_1\right)\mathbb{P}\left(X_2\leqslant x_2,Y_2\leqslant y_2\right).$$

(3) 当 Z_1 与 Z_2 独立时, 有 $\mathbb{E}[Z_1Z_2]=\mathbb{E}[Z_1]\mathbb{E}[Z_2]$.【利用 (2) 中等式求边缘分布可知 X_1 与 X_2、Y_1 与 Y_2、X_1 与 Y_2、X_2 与 Y_1 均独立, 再根据定义即可验证.】

定义 1.19.3 随机变量 X 的特征函数 $\phi_X(t)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX}\right]$, $t\in\mathbb{R}$.

注 1.19.4 (1) $\phi_X(t)=\mathbb{E}[\cos(tX)]+\mathrm{i}\,\mathbb{E}[\sin(tX)]$.

(2) 由于 $\left|\mathrm{e}^{\mathrm{i}\,tX}\right|\equiv 1$ 是有界量, $\phi_X(t)$ 总存在.

(3) $\phi_X(t)=\displaystyle\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_X$. 特别地, 当 X 是连续型随机变量时, $\phi_X(t)=\displaystyle\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx}f(x)\,\mathrm{d}x$.

下面在不引起混淆的前提下将 $\phi_X(t)$ 简记为 $\phi(t)$.

定理 1.19.5 (1) $\phi(0)=1$, $\overline{\phi(t)}=\phi(-t)$, $|\phi(t)|\leqslant 1$.

(2) $\phi(t)$ 在 $\mathbb{R}$ 上一致连续.

(3) $\phi(t)$半正定, 即

$$\sum_{j,k=1}^{n}z_j\overline{z_k}\phi\left(t_j-t_k\right)\geqslant 0,\quad\forall z_j\in\mathbb{C}\ (j=1,\cdots,n).$$

证明 (1) $\phi(0)=\mathbb{E}[1]=1$, $\overline{\phi(t)}=\overline{\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tx}\right]}=\mathbb{E}\left[\mathrm{e}^{-\mathrm{i}\,tx}\right]=\phi(-t)$, $|\phi(t)|\leqslant\mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}\,tX}\right|\right]=\mathbb{E}[1]=1$.

(2) 由于

$$|\phi(t_0+h)-\phi(t_0)|=\left|\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t_0X}\left(\mathrm{e}^{\mathrm{i}\,hX}-1\right)\right]\right|\leqslant\mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}\,t_0X}\right|\cdot\left|\mathrm{e}^{\mathrm{i}\,hX}-1\right|\right]=\mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}\,hX}-1\right|\right],$$

而 $\left|\mathrm{e}^{\mathrm{i}\,hX}-1\right|\leqslant\left|\mathrm{e}^{\mathrm{i}\,hX}\right|+1=2$, 根据有界收敛定理,

$$\lim_{h\to 0}\mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}\,hX}-1\right|\right]=\mathbb{E}\left[0\right]=0.$$

注意到 $\mathbb{E}\left[\left|\mathrm{e}^{\mathrm{i}\,hX}-1\right|\right]$ 不含 t_0, 因此 $\phi(t)$ 在 $\mathbb{R}$ 上一致连续.

(3) 我们有

$$\begin{aligned}
\sum_{j,k=1}^{n} z_j\overline{z_k}\phi\left(t_j-t_k\right) &= \sum_{j,k=1}^{n} \mathbb{E}\left[z_j\overline{z_k}\,\mathrm{e}^{\mathrm{i}(t_j-t_k)X}\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{n} z_j\,\mathrm{e}^{\mathrm{i}\,t_jX}\right)\left(\sum_{k=1}^{n}\overline{z_k}\,\mathrm{e}^{-\mathrm{i}\,t_kX}\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{n} z_j\,\mathrm{e}^{\mathrm{i}\,t_jX}\right)\overline{\left(\sum_{k=1}^{n} z_k\,\mathrm{e}^{\mathrm{i}\,t_kX}\right)}\right] \\
&= \mathbb{E}\left[\left|\sum_{j=1}^{n} z_j\,\mathrm{e}^{\mathrm{i}\,t_jX}\right|^2\right] \geqslant 0.
\end{aligned}$$

□

注 1.19.6 根据 Bochner 定理, 以上 3 个性质完整刻画了特征函数的性质, 即满足这 3 个性质的函数必为某随机变量的特征函数.

定理 1.19.7 若 $\mathbb{E}\left[|X|^k\right]<+\infty$, 则对任意 $j\leqslant k$, 有 $\phi^{(j)}(0)=\mathrm{i}^j\,\mathbb{E}\left[X^j\right]$. 进而有 Taylor 展开

$$\phi(t)=\sum_{j=0}^{k}\frac{(\mathrm{i}\,t)^j}{j!}\mathbb{E}\left[X^j\right]+o\left(t^k\right).$$

证明 由习题 (5.6.4)[14]可知 $\mathbb{E}\left[|X|^j\right]<+\infty,\ \forall j\leqslant k$. 因此

$$\left|\frac{\mathrm{d}^j}{\mathrm{d}t^j}\,\mathrm{e}^{\mathrm{i}\,tx}\right|=\left|(\mathrm{i}\,x)^j\,\mathrm{e}^{\mathrm{i}\,tx}\right|=|x|^j,\quad j\leqslant k$$

仍可积. 故积分与求导可交换次序, 即当 $j\leqslant k$ 时,

$$\phi^{(j)}(t)=\frac{\mathrm{d}^j}{\mathrm{d}t^j}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_X=\int_{\mathbb{R}}\frac{\mathrm{d}^j}{\mathrm{d}t^j}\,\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_X=\int_{\mathbb{R}}(\mathrm{i}\,x)^j\,\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_X.$$

因此 $\phi^{(j)}(0)=\displaystyle\int_{\mathbb{R}}(\mathrm{i}\,x)^j\,\mathrm{d}F_X=\mathrm{i}^j\,\mathbb{E}\left[X^j\right]$. 进而可作 Taylor 展开. □

定理 1.19.8 (1) $\phi_{aX+b}(t)=\mathrm{e}^{\mathrm{i}\,tb}\,\phi_X(at)$.

(2) 当 X 与 Y 独立时, $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)$.

定义 1.19.9 随机向量 $\mathbf{X}=(X_1,\cdots,X_n)$ 的特征函数 $\phi_{\mathbf{X}}(\mathbf{t})=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,\mathbf{t}\cdot\mathbf{X}}\right]$.

定理 1.19.10 X 与 Y 独立 $\iff \phi_{X,Y}(s,t)=\phi_X(s)\phi_Y(t)$.

[14]或参考：实变函数论, 周民强, 北京大学出版社, 2016: 248-256.

证明 $\Rightarrow$：$\phi_{X,Y}(s,t)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}(sX+tY)}\right]\xlongequal{\text{注1.19.2(3)}}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,sX}\right]\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tY}\right]=\phi_X(s)\phi_Y(t)$.

$\Leftarrow$：需用反转公式, 见注1.20.4. □

例 1.19.11 (Bernoulli 两点分布) $\phi(t)=q+p\,\mathrm{e}^{\mathrm{i}\,t}$.

例 1.19.12 (指数分布) 由 $f(x)=\lambda\,\mathrm{e}^{-\lambda x}$, $x>0$ 可知

$$\phi(t)=\int_0^{+\infty}\mathrm{e}^{\mathrm{i}\,tx}\cdot\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x=\lambda\int_0^{+\infty}\mathrm{e}^{-(\lambda-\mathrm{i}\,t)x}\,\mathrm{d}x=\frac{\lambda}{\lambda-\mathrm{i}\,t}.$$

例 1.19.13 (标准正态分布) 由 $f(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$ 可知

$$\phi(t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx-\frac{1}{2}x^2}\,\mathrm{d}x\xlongequal{s=\mathrm{i}\,t}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2}(x-s)^2+\frac{1}{2}s^2}\,\mathrm{d}x=\mathrm{e}^{-\frac{1}{2}t^2}.$$

注 1.19.14 (1) 积分计算的另一种方式：

$$\phi(t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\cos(tx)\,\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x+\mathrm{i}\,\frac{1}{\sqrt{2\pi}}\underbrace{\int_{\mathbb{R}}\sin(tx)\,\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x}_{0},$$

求导得

$$\phi'(t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sin(tx)\,\mathrm{d}\,\mathrm{e}^{-\frac{1}{2}x^2}\xlongequal{\text{分部积分}}-t\phi(t).$$

解初值问题

$$\begin{cases}\phi'(t)+t\phi(t)=0,\\ \phi(0)=1\end{cases}$$

即得 $\phi(t)=\mathrm{e}^{-\frac{1}{2}t^2}$.

(2) 设 $X\sim N(0,1)$, $Y\sim N\left(\mu,\sigma^2\right)$, 则由 $Y=\sigma X+\mu$ 及定理1.19.8(1) 可得 $\phi(t)=\mathrm{e}^{\mathrm{i}\,\mu t-\frac{1}{2}\sigma^2t^2}$.

例 1.19.15 (多元正态分布) 设 $\mathbf{X}\sim N(\boldsymbol{\mu},\Sigma)$, 欲求 $\phi_{\mathbf{X}}(\mathbf{t})=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,\mathbf{X}\mathbf{t}^\top}\right]$. 令 $Y=\mathbf{X}\mathbf{t}^\top$, 则由定理1.14.7, 当 $\mathbf{t}\neq\mathbf{0}$ 时, 将 $Y=\mathbf{X}\mathbf{t}^\top$ 视作满秩线性变换就有 $Y\sim N\left(\boldsymbol{\mu}\mathbf{t}^\top,\mathbf{t}\Sigma\mathbf{t}^\top\right)$. 进而由注1.19.14(2) 可知

$$\phi_Y(s)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,sY}\right]=\mathrm{e}^{\mathrm{i}\,s\boldsymbol{\mu}\mathbf{t}^\top-\frac{1}{2}s^2\mathbf{t}\Sigma\mathbf{t}^\top}.$$

令 $s=1$ 即得

$$\phi_{\mathbf{X}}(\mathbf{t})=\mathrm{e}^{\mathrm{i}\,\boldsymbol{\mu}\mathbf{t}^\top-\frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^\top}.$$

注 1.19.16 注意到此时未出现 Σ^{-1}, 因此可不对 Σ 作正定要求, 即 $\mathbf{X}$ 可以服从“退化的”多元正态分布, 此时无密度函数. 表达式的形式体现了一种对偶.

20　反转公式与连续性定理

定理 1.20.1 (反转公式) 给定 $-\infty < a < b < +\infty$, 则

$$\frac{1}{2}\left[F(b)+F(b^-)\right]-\frac{1}{2}\left[F(a)+F(a^-)\right]=\lim_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\,at}-\mathrm{e}^{-\mathrm{i}\,bt}}{\mathrm{i}\,t}\phi(t)\,\mathrm{d}t.$$

证明 记 $I_T=\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\,at}-\mathrm{e}^{-\mathrm{i}\,bt}}{\mathrm{i}\,t}\phi(t)\,\mathrm{d}t$, 则由 $\phi(t)=\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F$ 知

$$I_T=\frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{\mathrm{i}\,t(x-a)}-\mathrm{e}^{\mathrm{i}\,t(x-b)}}{\mathrm{i}\,t}\,\mathrm{d}F\,\mathrm{d}t.$$

而[15]

$$\left|\frac{\mathrm{e}^{\mathrm{i}\,t(x-a)}-\mathrm{e}^{\mathrm{i}\,t(x-b)}}{\mathrm{i}\,t}\right|\leqslant\frac{\left|\mathrm{e}^{\mathrm{i}\,t(x-a)}-1\right|+\left|\mathrm{e}^{\mathrm{i}\,t(x-b)}-1\right|}{|t|}\leqslant|x-a|+|x-b|$$

有界, 由 Fubini 定理,

$$I_T=\int_{-\infty}^{+\infty}g_T(x)\,\mathrm{d}F,$$

其中

$$g_T(x)=\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{\mathrm{i}\,t(x-a)}-\mathrm{e}^{\mathrm{i}\,t(x-b)}}{\mathrm{i}\,t}\,\mathrm{d}t.$$

又

$$\begin{aligned}g_T(x)&=\frac{1}{2\pi}\int_{-T}^{T}\frac{[\cos(x-a)t-\cos(x-b)t]+\mathrm{i}\,[\sin(x-a)t-\sin(X-b)t]}{\mathrm{i}\,t}\,\mathrm{d}t\\&=\frac{1}{\pi}\int_{0}^{T}\frac{\sin(x-a)t-\sin(x-b)t}{t}\,\mathrm{d}t.\end{aligned}$$

利用

$$\frac{1}{\pi}\int_0^{+\infty}\frac{\sin(xt)}{t}\,\mathrm{d}t=\begin{cases}\frac{1}{2}, & x>0,\\ 0, & x=0,\\ -\frac{1}{2}, & x<0\end{cases}$$

可知 $|g_T(x)|$ 有界, 且

$$\lim_{T\to+\infty}g_T(x)=\begin{cases}1, & x\in(a,b),\\ \frac{1}{2}, & x=a\ \text{或}\ b,\\ 0, & \text{其他}.\end{cases}$$

[15]这里用到了 $\left|\mathrm{e}^{\mathrm{i}\,\alpha}-1\right|\leqslant|\alpha|,\ \forall\alpha\in\mathbb{R}$, 可由几何意义得到, 也可由 $\left|\int_0^1\mathrm{e}^{\mathrm{i}\,\alpha t}\,\mathrm{d}t\right|\leqslant\int_0^1\left|\mathrm{e}^{\mathrm{i}\,\alpha t}\right|\,\mathrm{d}t=1$ 得到.

再由控制收敛定理, 将积分与极限换序得

$$\begin{aligned}\lim_{T\to+\infty} I_T &= \frac{1}{2}\mu_F(\{a,b\}) + \mu_F((a,b)) \\ &= \frac{1}{2}\left[F(b)-F(b^-)+F(a)-F(a^-)\right] + [F(b)-F(a)] - \left[F(b)-F(b^-)\right] \\ &= \frac{1}{2}\left[F(b)+F(b^-)\right] - \frac{1}{2}\left[F(a)+F(a^-)\right].\end{aligned}$$

□

从特征函数的定义可知, X 与 Y 同分布 $\implies \phi_X(t)=\phi_Y(t)$. 反过来, 我们有

推论 1.20.2 (唯一性定理) $\phi_X(t)=\phi_Y(t) \implies X$ 与 Y 同分布.

证明 要说明 F 被 $\phi(t)$ 唯一确定. 对 $a,b\in\mathcal{C}_F$, $F(b)-F(a)$ 由 ϕ 确定, 再令 $\mathcal{C}_F \ni a\to-\infty$(由分布函数的不连续点至多可数知这可以实现), 于是 $F(b)$ 由 ϕ 确定. 对 $b\in\mathbb{R}\setminus\mathcal{C}_F$, 取 $\mathcal{C}_F\ni b_n\downarrow b$, 由 F 的右连续性, $F(b)=\lim\limits_{n\to\infty}F(b_n)$ 也由 ϕ 确定. □

定理 1.20.3 (多元反转公式) 记 $\phi(\mathbf{t})=\int_{\mathbb{R}^n}\mathrm{e}^{\mathrm{i}(t_1x_1+\cdots+t_nx_n)}\,\mathrm{d}F$, $R=(a_1,b_1]\times\cdots\times(a_n,b_n]$. 若 $\mu_F(\partial R)=0$, 则

$$\mu_F(R)=\lim_{T_1,\cdots,T_n\to+\infty}\frac{1}{(2\pi)^n}\int_{-T_1}^{T_1}\cdots\int_{-T_n}^{T_n}\prod_{k=1}^n\frac{\mathrm{e}^{-\mathrm{i}\,t_ka_k}-\mathrm{e}^{-\mathrm{i}\,t_kb_k}}{\mathrm{i}\,t_k}\phi(\mathbf{t})\,\mathrm{d}t_1\cdots\mathrm{d}t_n.$$

注 1.20.4 特别地, 当 $n=2$ 时, 若 $\phi(t_1,t_2)=\phi_{X_1}(t_1)\phi_{X_2}(t_2)$, 则 X_1 与 X_2 独立.

例 1.20.5 求 $\cos t$ 对应的分布函数.

证明 直接构造 $\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\dfrac{1}{2}$, 验证知 $\phi_X(t)=\dfrac{\mathrm{e}^{\mathrm{i}\,t}+\mathrm{e}^{-\mathrm{i}\,t}}{2}=\cos t$. □

问题 $\{X_n\},\{F_n\},\{\phi_n\}$ 三者极限之间有何联系?

定理 1.20.6 (Lévy-Cramér 连续性定理) 记 $F_n(x)=\mathbb{P}(X_n\leqslant x)$, $\phi_n(t)=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_n$.

(1) 若 $F_n\xrightarrow{\mathrm{w}}F$, F 为分布函数, 则 $\phi_n(t)\to\phi(t):=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F$, 且此收敛为 $\mathbb{R}$ 上的内闭一致收敛.

(2) 若 $\phi(t):=\lim\limits_{n\to\infty}\phi_n(t)$, 且 $\phi(t)$ 在 $t=0$ 处连续, 则 ϕ 必为某分布函数 F 的特征函数, 且 $F_n\xrightarrow{\mathrm{w}}F$.

定理1.20.6的证明此处从略, 不过我们可以通过下面的反例体会 (2) 中 $\phi(t)$ 在 $t=0$ 处连续这一条件的关键性.

例 1.20.7 设 $X\sim U[-n,n]$, 则 $\phi_n(t)=\dfrac{1}{2n}\displaystyle\int_{-n}^{n}\mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}x=\dfrac{\sin nt}{nt}$, 且 $\lim\limits_{n\to\infty}\phi_n(t)=\begin{cases}0, & t\neq 0,\\ 1, & t=0\end{cases}$ 在 $t=0$ 处不连续, 由定理1.19.5(2) 知此极限函数不是特征函数.

补充：依分布收敛与几乎处处收敛的联系

定理 1.20.8 (Skorokhod 表示定理) 设 $X_n\xrightarrow{\mathrm{D}}X$, 则存在概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 及其上的随机变量 $\{Y_n\},Y$, 满足

(1) Y_n 与 X_n 同分布, Y 与 X 同分布.

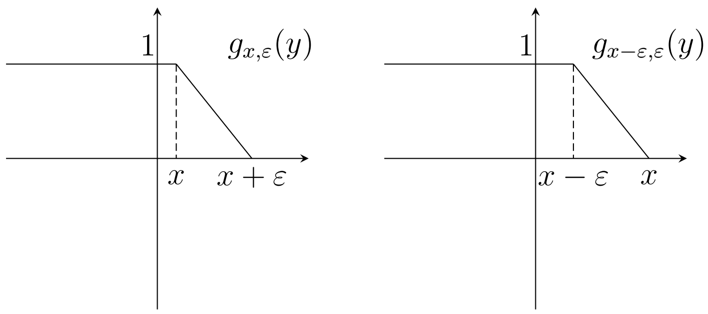
(2) $Y_n\xrightarrow{\text{a.s.}}Y$.

我们可以从以下定理中体会弱收敛中“弱”的意味.

定理 1.20.9 $X_n\xrightarrow{\mathrm{D}}X\iff\forall g\in\mathcal{C}_{\mathrm{b}}(\mathbb{R}),\ \mathbb{E}[g(X_n)]\to\mathbb{E}[g(X)]$.

证明 $\Rightarrow$:由 Skorokhod 表示定理, 可取 $\{Y_n\}$ 使 Y_n 与 X_n 同分布, Y 与 X 同分布, 且 $Y_n\xrightarrow{\text{a.s.}}Y$. 结合 g 的连续性即知 $g(Y_n)\xrightarrow{\text{a.s.}}g(Y)$. 又因为 g 有界, 由控制收敛定理, $\mathbb{E}[g(Y_n)]\to\mathbb{E}[g(Y)]$, 即 $\mathbb{E}[g(X_n)]\to\mathbb{E}[g(X)]$.

$\Leftarrow$: 为了构造连续函数 g, 将示性函数 $I_{(-\infty,x]}$ 分别磨光为下图两个折线形函数.



于是

$$F_n(x)=\mathbb{E}\left[I_{(-\infty,x]}\circ X_n\right]\leqslant\mathbb{E}\left[g_{x,\varepsilon}\circ X_n\right],$$
$$F_n(x)=\mathbb{E}\left[I_{(-\infty,x]}\circ X_n\right]\geqslant\mathbb{E}\left[g_{x-\varepsilon,\varepsilon}\circ X_n\right].$$

分别取上下极限就得到

$$\limsup_{n\to\infty}F_n(x)\leqslant\lim_{n\to\infty}\mathbb{E}\left[g_{x,\varepsilon}\circ X_n\right]=\mathbb{E}\left[g_{x,\varepsilon}\circ X\right]\leqslant F(x+\varepsilon),$$
$$\liminf_{n\to\infty}F_n(x)\geqslant\lim_{n\to\infty}\mathbb{E}\left[g_{x-\varepsilon,\varepsilon}\circ X_n\right]=\mathbb{E}\left[g_{x-\varepsilon,\varepsilon}\circ X\right]\geqslant F(x-\varepsilon).$$

于是得到

$$F(x-\varepsilon) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+\varepsilon).$$

对任意 $x \in \mathcal{C}_F$, 令 $\varepsilon \to 0$, 即得 $\lim\limits_{n\to\infty} F_n(x) = F(x)$. 故 $X_n \xrightarrow{\mathrm{D}} X$. □

注 1.20.10 利用定理1.20.9的"⇒", 我们可以证明 Lévy-Cramér 连续性定理的 (1):

设 $X_n \xrightarrow{\mathrm{D}} X$. 对任意 $t \in \mathbb{R}$, 令 $f(x) = \cos(tx), g(x) = \sin(tx) \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$. 由定理1.20.9,

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)], \quad \mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)],$$

因此

$$\phi_n(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX_n}\right] = \mathbb{E}[f(X_n)] + \mathrm{i}\,\mathbb{E}[g(X_n)] \to \mathbb{E}[f(X)] + \mathrm{i}\,\mathbb{E}[g(X)] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX}\right].$$

21 极限定理

Nature is a mutable cloud, which is always and never the same.

——Ralph Waldo Emerson

一、问题来源

设随机变量 $\{X_k\}$ 相互独立且同分布, $S_n = \sum\limits_{k=1}^{n} X_k$, $\mu = \mathbb{E}[X_k]$. 由 Kolmogorov 强大数律, $\dfrac{S_n - n\mu}{n} \xrightarrow{\text{a.s.}} 0$. 若还已知 $\mathrm{Var}(X_k) < +\infty$, 我们尝试更精细地估计随机变量收敛的速度. 由 $\mathrm{Var}\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n^2}\sum\limits_{k=1}^{n}\mathrm{Var}(X_k) = \dfrac{\mathrm{Var}(X_1)}{n}$, 结合 Khinchine 重对数律 (定理1.18.8与注1.18.9), 我们可以尝试研究随机变量 $\dfrac{S_n - n\mu}{\sqrt{n}\sigma}$ 的收敛情况.[16]

二、大数定律与中心极限定理

利用特征函数证明弱大数律 (定理1.18.1)

$$\begin{aligned}\phi_{\frac{S_n}{n}}(t) &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t\frac{S_n}{n}}\right] = \left(\phi_{\frac{X_1}{n}}(t)\right)^n = \left(\phi_{X_1}\left(\frac{t}{n}\right)\right)^n \\ &\xlongequal[\text{Taylor 展开}]{\text{定理1.19.7}} \left(1 + \mathrm{i}\,\frac{t}{n}\mu + o\left(\frac{t}{n}\right)\right)^n \to \mathrm{e}^{\mathrm{i}\,\mu t}, \quad n \to \infty.\end{aligned}$$

注意到 $\mathrm{e}^{\mathrm{i}\,\mu t}$ 是常值随机变量 μ 的特征函数, 它在 $t = 0$ 处连续, 由 Lévy-Cramér 连续性定理, $\dfrac{S_n}{n} \xrightarrow{\mathrm{D}} \mu$. 再由定理1.16.10(1) 得 $\dfrac{S_n}{n} \xrightarrow{\mathrm{P}} \mu$. □

[16]注意此时 $\dfrac{S_n - n\mu}{\sqrt{n}\sigma}$ 是经规范化的, 即期望为 0, 方差为 1.

定理 1.21.1 (中心极限定理) 设随机变量 $\{X_k\}$ 相互独立且同分布, $\mu = \mathbb{E}[X_k]$, $\sigma^2 = \operatorname{Var}(X_k)$, $\sigma \in (0,+\infty)$, 则

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathrm{D}} Z,$$

这里 $Z \sim N(0,1)$.

证明 不妨设 $\mu = 0, \sigma = 1$. 我们有

$$\begin{aligned}
\phi_{\frac{S_n}{\sqrt{n}}}(t) &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t\frac{S_n}{\sqrt{n}}}\right] = \left(\phi_{\frac{X_1}{\sqrt{n}}}(t)\right)^n = \left(\phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&\xlongequal[\text{Taylor 展开}]{\text{定理1.19.7}} \left(1 + \mathrm{i}\,\frac{t}{\sqrt{n}}\mathbb{E}[X_1] + \frac{1}{2}\left(\mathrm{i}\,\frac{t}{\sqrt{n}}\right)^2 \mathbb{E}\left[X_1^2\right] + o\left(\frac{t^2}{n}\right)\right)^n \\
&\xlongequal{\mu=0,\sigma^2=1} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to \mathrm{e}^{-\frac{1}{2}t^2}, \quad n \to \infty.
\end{aligned}$$

因为 $\mathrm{e}^{-\frac{1}{2}t^2}$ 是标准正态分布的特征函数 (自然在 $t=0$ 处连续), 由 Lévy-Cramér 连续性定理, $\dfrac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathrm{D}} Z$. □

注 1.21.2 我们有时会直接记作 $\dfrac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathrm{D}} N(0,1)$, 尽管这有滥用记号之嫌.

三、Lindeberg 条件

下面探讨独立但不要求同分布的随机变量的中心极限定理.

对 $X_1, \cdots, X_n$, 不妨设 $\mathbb{E}[X_k] = 0$, $\forall k$. 记 $b_k^2 = \operatorname{Var}(X_k)$, 总方差 $B_n^2 = \sum\limits_{k=1}^n b_k^2$, 并设 $B_n > 0$ 为总标准差. 称以下条件为 Lindeberg 条件：

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^n \mathbb{E}\left[X_k^2 I_{\{|X_k| > \varepsilon B_n\}}\right] = 0, \quad \forall \varepsilon > 0. \tag{L}$$

定理 1.21.3 (Lindeberg-Feller 中心极限定理) 设随机变量 $\{X_k\}$ 相互独立, 且满足 Lindeberg 条件(L), 则

$$\frac{S_n}{B_n} \xrightarrow{\mathrm{D}} N(0,1) \tag{LF}$$

且

$$\lim_{n\to\infty} \frac{1}{B_n^2} \max_{1\leqslant k\leqslant n} b_k^2 = 0. \tag{F}$$

注 1.21.4 (1) 我们称(F)为 Feller 条件. 它意味着 $\dfrac{b_k^2}{B_n^2} \to 0$ 对 k 一致成立.

(2) 称以下条件为 3 阶矩条件：

$$\lim_{n\to\infty}\frac{1}{B_n^3}\sum_{k=1}^n\mathbb{E}\left[|X_k|^3\right]=0. \tag{T}$$

3 阶矩条件可推出 Lindeberg 条件(L), 进而可使中心极限定理成立. 证明如下：

$$\begin{aligned}\frac{1}{B_n^2}\sum_{k=1}^n\mathbb{E}\left[X_k^2I_{\{|X_k|>\varepsilon B_n\}}\right]&=\frac{1}{B_n^2}\sum_{k=1}^n\mathbb{E}\left[X_k^2\frac{|X_k|}{|X_k|}I_{\{|X_k|>\varepsilon B_n\}}\right]\\&\leqslant\frac{1}{\varepsilon B_n^3}\sum_{k=1}^n\mathbb{E}\left[|X_k|^3\right]\xrightarrow{3\text{ 阶矩条件}}0,\quad n\to\infty.\end{aligned}$$

(3) Lindeberg 条件(L)的概率含义：

$$\begin{aligned}\frac{1}{B_n^2}\sum_{k=1}^n\mathbb{E}\left[X_k^2I_{\{|X_k|>\varepsilon B_n\}}\right]&\geqslant\varepsilon^2\sum_{k=1}^n\mathbb{E}\left[I_{\{|X_k|>\varepsilon B_n\}}\right]\\&=\varepsilon^2\sum_{k=1}^n\mathbb{P}\left(|X_k|>\varepsilon B_n\right)\\&\geqslant\varepsilon^2\mathbb{P}\left(\bigcup_{k=1}^n\{|X_k|>\varepsilon B_n\}\right)\\&=\varepsilon^2\mathbb{P}\left(\max_{1\leqslant k\leqslant n}\frac{|X_k|}{B_n}>\varepsilon\right).\end{aligned}$$

由此可知, Lindeberg 条件(L)意味着, 对任意 $\varepsilon>0$, $\lim\limits_{n\to\infty}\mathbb{P}\left(\max\limits_{1\leqslant k\leqslant n}\dfrac{|X_k|}{B_n}>\varepsilon\right)=0$, 即“相对偏差 $\dfrac{|X_k|}{B_n}$ 一致小”的概率接近于 1.

(4) 当 $\{X_k\}$ 相互独立时, (LF)+(F) $\Longrightarrow$ (L).

(5) Lyapunov 条件：存在 $\delta>0$, 使得

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^n\mathbb{E}\left[|X_k|^{2+\delta}\right]=0.$$

Lyapunov 条件可推出 Lindeberg 条件(L), 进而可使中心极限定理成立.

(6) (L) $\Longrightarrow$ (F). 证明如下：

$$\begin{aligned}\frac{b_k^2}{B_n^2}&=\frac{1}{B_n^2}\mathbb{E}\left[X_k^2\left(I_{\{|X_k|\leqslant\varepsilon B_n\}}+I_{\{|X_k|>\varepsilon B_n\}}\right)\right]\\&\leqslant\varepsilon^2+\frac{1}{B_n^2}\sum_{k=1}^n\mathbb{E}\left[X_k^2I_{\{|X_k|>\varepsilon B_n\}}\right],\end{aligned}$$

再结合 Lindeberg 条件(L), 取上极限得

$$\limsup_{n\to\infty}\frac{b_k^2}{B_n^2}\leqslant\varepsilon^2,$$

且这关于 k 是一致的. 再令 $\varepsilon\to 0$ 即得证.

四、二项分布的正态逼近

设随机变量 $\{X_k\}$ 相互独立且同分布, $\mathbb{P}(X_1=1)=p$, $\mathbb{P}(X_1=0)=q$, $p+q=1$. 设 $S_n=\sum\limits_{k=1}^{n}X_k$, 则 $\mathbb{E}[S_n]=np$, $\mathrm{Var}(S_n)=npq$. 引入规范化的 $x_k:=\dfrac{k-np}{\sqrt{npq}}$ $(k=0,1,\cdots,n)$.

定理 1.21.5 (局部中心极限定理) 设 $p\in(0,1)$, $x_k:=\dfrac{k-np}{\sqrt{npq}}$ $(k=0,1,\cdots,n)$, 则

$$\mathbb{P}(S_n=k)\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{1}{2}x_k^2},\quad n\to\infty$$

对所有满足 $|x_k|\leqslant A$ 的 k 一致成立.

证明 由 $\begin{cases}k=np+\sqrt{npq}x_k,\\ n-k=nq+\sqrt{npq}x_k\end{cases}$ 可知对 $k=0,1,\cdots,n$ 一致地有

$$k\sim np,\quad n-k\sim nq,\quad n\to\infty.$$

利用 Stirling 公式 $n!\sim\left(\dfrac{n}{\mathrm{e}}\right)^n\sqrt{2\pi n}$ 可得

$$\begin{aligned}\mathrm{C}_n^kp^kq^{n-k}&\sim\frac{\left(\frac{n}{\mathrm{e}}\right)^n\sqrt{2\pi n}p^kq^{n-k}}{\left(\frac{k}{\mathrm{e}}\right)^k\sqrt{2\pi k}\left(\frac{n-k}{\mathrm{e}}\right)^{n-k}\sqrt{2\pi(n-k)}}\\&=\sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}\\&\sim\frac{1}{\sqrt{2\pi npq}}\underbrace{\left(1-\frac{\sqrt{npq}}{k}x_k\right)^k\left(1+\frac{\sqrt{npq}}{n-k}x_k\right)^{n-k}}_{\varphi_{n,k}}.\end{aligned}$$

我们有

$$\begin{aligned}\ln\varphi_{n,k}&=k\ln\left(1-\frac{\sqrt{npq}}{k}x_k\right)+(n-k)\ln\left(1+\frac{\sqrt{npq}}{n-k}x_k\right)\\&=-\sqrt{npq}x_k-\frac{npq}{2k}x_k^2+O\left(\frac{1}{\sqrt{n}}\right)+\sqrt{npq}x_k-\frac{npq}{2(n-k)}x_k^2+O\left(\frac{1}{\sqrt{n}}\right)\\&=-\frac{x_k^2}{2}+O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

□

定理 1.21.6 (积分形式的中心极限定理) 设 $S_n\sim B(n,p)$, 则

$$\mathbb{P}\left(a<\frac{S_n-n\mu}{\sqrt{npq}}\leqslant b\right)\to\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x,\quad n\to\infty.$$

证明 同定理1.21.5取 $x_k := \dfrac{k-np}{\sqrt{npq}}$ $(k=0,1,\cdots,n)$, 则

$$\begin{aligned}\text{LHS} &= \sum_{k:x_k\in(a,b]} \mathbb{P}(S_n=k) \underbrace{\sim}_{\text{对 } k \text{ 一致地}} \sum_{k:x_k\in(a,b]} \frac{1}{\sqrt{2\pi npq}} \mathrm{e}^{-\frac{1}{2}x_k^2} \\ &= \sum_{k:x_k\in(a,b]} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x_k^2}(x_{k+1}-x_k) \xrightarrow{*} \int_a^b \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x, \quad n\to\infty.\end{aligned}$$

当 $n\to\infty$ 时“$*$”处可由 Riemann 和转化为定积分的理由：

(1) 对于固定的 n, 每个 x_k 与 k 一一对应.

(2) $x_0,x_1,\cdots,x_n$ 为区间 $\left[-\sqrt{\dfrac{np}{q}},\sqrt{\dfrac{nq}{p}}\right]$ 的一个等间隔划分, 且当 $n\to\infty$ 时此区间包含 $(a,b]$, 每个子区间长度 $\dfrac{1}{\sqrt{npq}}\to 0$. □

五、矩方法

标准正态分布的 k 阶矩 在本部分中我们引入记号

$$\gamma_k := \int_{\mathbb{R}} x^k \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = \begin{cases} (2m-1)!!, & k=2m, \\ 0, & k=2m-1. \end{cases}$$

其中 $(2m-1)!!$ 的组合意义为 $1,2,\cdots,2m$ 两两配对的方法数 (每次均从未配对的第一个开始考虑).

定理 1.21.7 设随机变量相互独立, 且满足：

(1) $\mathbb{E}[X_k]=0$, $\mathrm{Var}(X_k)=1$, $\forall k$.

(2) 一致有界高阶矩：$C_m := \sup\limits_k \mathbb{E}\left[|X_k|^m\right] < +\infty$, $\forall m\geqslant 3$.

则对 $S_n = \sum\limits_{k=1}^n X_k$ 有

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] \to \gamma_k, \quad n\to\infty.$$

证明 我们知道

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] = \frac{1}{n^{\frac{k}{2}}} \sum_{i_1,\cdots,i_k=1}^n \mathbb{E}\left[X_{i_1}\cdots X_{i_k}\right]. \tag{$*$}$$

下面先对 $k=0,1,2,3,4$ 进行观察, 再处理一般的 k.

① $k=1$ 时, 由中心极限定理, 结论平凡.

② $k=2$ 时, 由(∗)式得

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]=\frac{1}{n}\sum_{i_1}\mathbb{E}\left[X_{i_1}^2\right]+\frac{1}{n}\sum_{i_1\neq i_2}\mathbb{E}\left[X_{i_1}X_{i_2}\right]=\frac{1}{n}\cdot n\operatorname{Var}(X_1)+\frac{1}{n}\sum_{i_1\neq i_2}\left(\mathbb{E}[X_1]\right)^2=1.$$

③ $k=3$ 时, 由(∗)式得

$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]&=\frac{1}{n^{\frac{3}{2}}}\sum_{i_1}\mathbb{E}\left[X_{i_1}^3\right]+\frac{3}{n^{\frac{3}{2}}}\sum_{i_1\neq i_2}\mathbb{E}\left[X_{i_1}^2X_{i_2}\right]+\frac{1}{n^{\frac{3}{2}}}\sum_{i_1\neq i_2\neq i_3}\mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}\right]\\&=\frac{1}{n^{\frac{3}{2}}}\cdot O(n)+\frac{3}{n^{\frac{3}{2}}}\sum_{i_1\neq i_2}\operatorname{Var}(X_1)\mathbb{E}[X_1]+\frac{1}{n^{\frac{3}{2}}}\sum_{i_1\neq i_2\neq i_3}\left(\mathbb{E}[X_1]\right)^3\\&=O\left(\frac{1}{\sqrt{n}}\right),\quad n\to\infty.\end{aligned}$$

④ $k=4$ 时, 由(∗)式得

$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]=&\frac{1}{n^2}\sum_{i_1}\mathbb{E}\left[X_{i_1}^4\right]+\frac{3}{n^2}\sum_{i_1\neq i_2}\mathbb{E}\left[X_{i_1}^2X_{i_2}^2\right]+\frac{4}{n^2}\sum_{i_1\neq i_2}\mathbb{E}\left[X_{i_1}X_{i_2}^3\right]\\&+\frac{6}{n^2}\sum_{i_1\neq i_2\neq i_3}\mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}^2\right]+\frac{1}{n^2}\sum_{i_1\neq i_2\neq i_3\neq i_4}\mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}X_{i_4}\right]\\=&\frac{1}{n^2}\cdot O(n)+\frac{3}{n^2}\cdot n(n-1)\left(\operatorname{Var}(X_1)\right)^2+\frac{4}{n^2}\sum_{i_1\neq i_2}\mathbb{E}[X_1]\mathbb{E}[X_2^3]\\&+\frac{6}{n^2}\sum_{i_1\neq i_2\neq i_3}\left(\mathbb{E}[X_1]\right)^2\operatorname{Var}(X_1)+\frac{1}{n^2}\sum_{i_1\neq i_2\neq i_3\neq i_4}\left(\mathbb{E}[X_1]\right)^4\\=&O\left(\frac{1}{n}\right)+\frac{3}{n^2}\cdot n(n-1)\to 3,\quad n\to\infty.\end{aligned}$$

由上面的观察可知, 对一般的 k, (∗)式右端非零项必形如

$$\mathbb{E}\left[X_{i_1}^{a_1}X_{i_2}^{a_2}\cdots X_{i_m}^{a_m}\right],$$

其中 $i_1\neq i_2\neq\cdots\neq i_m$, $a_1,a_2,\cdots,a_m\geqslant 2$ 且 $\sum_{i=1}^m a_i=k$. 于是 $m\leqslant\left\lfloor\frac{k}{2}\right\rfloor$.

- 若 k 为奇数, 则 $m\leqslant\frac{k-1}{2}$, 从而 $\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]=\frac{1}{n^{\frac{k}{2}}}\cdot O\left(n^{\frac{k-1}{2}}\right)=O\left(\frac{1}{\sqrt{n}}\right)$, $n\to\infty$.
- 若 k 为偶数, 则 $m\leqslant\frac{k}{2}$, 渐近展开式逐项在 $m=\frac{k}{2}$ 处取到. 而 $k=2m$ 时, 只能 $a_1=a_2=\cdots=a_m=2$, 因此主项为

$$\frac{1}{n^m}\sum_{i_1\neq i_2\neq\cdots\neq i_m}(2m-1)!!\cdot\mathbb{E}\left[X_{i_1}^2X_{i_2}^2\cdots X_{i_m}^2\right]\to\gamma_{2m},\quad n\to\infty.$$

□

定理 1.21.8 (矩收敛定理) 假设

(1) $\forall k \in \mathbb{N}$, $\gamma_{k,n} := \int_{\mathbb{R}} x^k \,\mathrm{d}F_n$ 存在, $\forall n$.

(2) $\lim\limits_{n\to\infty} \gamma_{k,n} = \gamma_k$, $\forall k$.

(3) $\gamma_k = \int_{\mathbb{R}} x^k \,\mathrm{d}F$ 满足 Riesz 条件[17]:

$$\limsup_{k\to\infty} \frac{1}{k}\gamma_{2k}^{\frac{1}{2k}} < +\infty \tag{R}$$

或 Carleman 条件:

$$\sum_{k=1}^{\infty} \gamma_{2k}^{-\frac{1}{2k}} = +\infty. \tag{C}$$

则 $F_n \xrightarrow{\mathrm{w}} F$, $n \to \infty$.

注 1.21.9 定理1.21.7在一定的条件下得到了 $\dfrac{S_n}{\sqrt{n}}$ 的 k 阶矩在 $n \to \infty$ 时收敛于标准正态分布的 k 阶矩. 现在我们看到, 只需验证 Riesz 条件或 Carleman 条件即可得到 $\dfrac{S_n}{\sqrt{n}} \xrightarrow{\mathrm{D}} N(0,1)$. 可参考Moment problem及Carleman's condition.

[17]Riesz 条件 (以及后面的 Carleman 条件) 保证了 F 的唯一性 (determinacy), 即若 Riesz 条件(R)或 Carleman 条件(C)成立, 则至多存在一个分布函数 F, 使得 $\gamma_k = \int_{\mathbb{R}} x^k \,\mathrm{d}F$, $\forall k \in \mathbb{N}$. 这里 γ_k 不局限于标准正态分布的 k 阶矩.

第二部分　课后习题

习题 (1.2.2) 设 $\mathcal{F}$ 是 σ 域, $A, B \in \mathcal{F}$. 证明：$A \cap B, A \setminus B, A \triangle B \in \mathcal{F}$.

证明 ① 由 $A \cap B = (A^{\mathrm{c}} \cup B^{\mathrm{c}})^{\mathrm{c}}$ 可知 $A \cap B \in \mathcal{F}$.

② 由 $A \setminus B = A \cap B^{\mathrm{c}}$ 及①即知 $A \setminus B \in \mathcal{F}$.

③ 由 $A \triangle B = (A \setminus B) \cup (B \setminus A)$ 及②即知 $A \triangle B \in \mathcal{F}$. □

习题 (1.3.1) 设事件 A, B 的概率分别为 $\mathbb{P}(A) = \dfrac{3}{4}, \mathbb{P}(B) = \dfrac{1}{3}$. 证明：$\dfrac{1}{12} \leqslant \mathbb{P}(A \cap B) \leqslant \dfrac{1}{3}$, 并举例说明等号可取到. 对 $\mathbb{P}(A \cup B)$ 找出相应的上下界.

证明 由 $A \cap B \subset B$ 得 $\mathbb{P}(A \cap B) \leqslant \mathbb{P}(B) = \dfrac{1}{3}$. 又

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geqslant \mathbb{P}(A) + \mathbb{P}(B) - 1 = \frac{3}{4} + \frac{1}{3} - 1 = \frac{1}{12}.$$

等号成立的例子：掷一个正十二面体匀质骰子 (各面依次标记为 $1, \cdots, 12$), 样本空间 $\Omega = \{1,2,3,4,5,6,7,8,9,10,11,12\}$. 若设 $A = \{1,2,3,4,5,6,7,8,9\}$, $B = \{9,10,11,12\}$, 则 $\mathbb{P}(A \cap B) = \mathbb{P}(\{9\}) = \dfrac{1}{12}$; 若设 $A = \{1,2,3,4,5,6,7,8,9\}$, $B = \{1,2,3,4\}$, 则 $\mathbb{P}(A \cap B) = \mathbb{P}(B) = \dfrac{1}{3}$.

对 $\mathbb{P}(A \cup B)$, 由于 $\mathbb{P}(A) + \mathbb{P}(B) = \dfrac{13}{12} > 1$, 因此 $\mathbb{P}(A \cup B) \leqslant 1$. 又有

$$\mathbb{P}(A \cup B) \geqslant \max(\mathbb{P}(A), \mathbb{P}(B)) = \frac{3}{4}.$$

□

习题 (1.3.4) 对事件 $A_1, A_2, \cdots, A_n$ $(n \geqslant 2)$, 证明：

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}\left(A_i \cap A_j\right) + \sum_{i<j<k} \mathbb{P}\left(A_i \cap A_j \cap A_k\right)$$

$$-\cdots+(-1)^{n+1}\mathbb{P}\left(A_1\cap A_2\cap\cdots\cap A_n\right).$$

每包麦片中可能含有剑桥大学过去五任副校长的塑料半身像, 每包含有每位副校长半身像的概率都是 $\frac{1}{5}$, 且与其他包独立. 求证: 购买 6 包麦片后, 得到过去三任副校长半身像的概率为 $1-3\left(\frac{4}{5}\right)^6+3\left(\frac{3}{5}\right)^6-\left(\frac{2}{5}\right)^6$.

证明 当 $n=2$ 时, 由 $A_1\cup A_2=A_1\cup(A_2\setminus A_1A_2)$ 是无交并得

$$\mathbb{P}(A_1\cup A_2)=\mathbb{P}(A_1)+\mathbb{P}(A_2\setminus A_1A_2)=\mathbb{P}(A_1)+\mathbb{P}(A_2)-\mathbb{P}(A_1A_2),$$

结论成立. 设结论对 $n-1$ $(n\geqslant 3)$ 成立, 则

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=&\mathbb{P}\left(\bigcup_{i=1}^{n-1}A_i\right)+\mathbb{P}(A_n)-\mathbb{P}\left(\left(\bigcup_{i=1}^{n-1}A_i\right)\cap A_n\right)\\
=&\mathbb{P}\left(\bigcup_{i=1}^{n-1}A_i\right)+\mathbb{P}(A_n)-\mathbb{P}\left(\bigcup_{i=1}^{n-1}(A_i\cap A_n)\right)\\
=&\sum_{k=1}^{n-1}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}\right)+\mathbb{P}(A_n)\\
&-\sum_{k=1}^{n-1}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(\left(A_{i_1}\cap A_n\right)\cap\cdots\cap\left(A_{i_k}\cap A_n\right)\right)\\
=&\sum_{k=1}^{n-1}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}\right)+\mathbb{P}(A_n)-\sum_{k=1}^{n-1}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}A_n\right)\\
=&\sum_{k=1}^{n}(-1)^{k+1}\sum_{i_1<\cdots<i_k}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}\right).
\end{aligned}$$

由数学归纳法知结论成立.

用 VC_i $(i=1,2,3)$ 表示过去三任副校长, 设事件 A_i 表示购买 6 包麦片不含 VC_i 的半身像, 则欲求事件的对立事件概率为

$$\begin{aligned}
\mathbb{P}\left(A_1\cup A_2\cup A_3\right)&=\sum_{k=1}^{3}(-1)^{k+1}\sum_{i_1<i_2<i_3}\mathbb{P}\left(A_{i_1}\cdots A_{i_k}\right)\\
&\xlongequal{\mathbb{P}(A_1)=\mathbb{P}(A_2)=\mathbb{P}(A_3)}3\mathbb{P}(A_1)-3\mathbb{P}\left(A_1\cap A_2\right)+\mathbb{P}\left(A_1\cap A_2\cap A_3\right)\\
&=3\left(\frac{4}{5}\right)^6-3\left(\frac{3}{5}\right)^6+\left(\frac{2}{5}\right)^6.
\end{aligned}$$

故欲求事件概率如题目所述. $\square$

习题 (1.3.5) 设事件列 A_r $(r \geqslant 1)$ 满足 $\mathbb{P}(A_r) = 1$, $\forall r$. 证明：$\mathbb{P}\left(\bigcap_{r=1}^{\infty} A_r\right) = 1$.

证明 由概率测度的连续性及 De Morgan 法则可得

$$\begin{aligned}1 \geqslant \mathbb{P}\left(\bigcap_{r=1}^{\infty} A_r\right) &= \mathbb{P}\left(\lim_{n\to\infty}\bigcap_{r=1}^{n} A_r\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{r=1}^{n} A_r\right) = \lim_{n\to\infty}\left[1-\mathbb{P}\left(\left(\bigcap_{r=1}^{n} A_r\right)^{\mathrm{c}}\right)\right] \\ &= \lim_{n\to\infty}\left[1-\mathbb{P}\left(\bigcup_{r=1}^{n} A_r^{\mathrm{c}}\right)\right] \geqslant 1-\sum_{r=1}^{\infty}\mathbb{P}\left(A_r^{\mathrm{c}}\right) = 1.\end{aligned}$$

于是 $\mathbb{P}\left(\bigcap_{r=1}^{\infty} A_r\right) = 1$. □

习题 (1.4.7) 设有 n 个缸, 其中第 r 个缸中含有 $r-1$ 个红球和 $n-r$ 个洋红色球. 随机选取一个缸并不放回地取出两个球, 试求以下事件的概率：

(a) 第 2 个球是洋红色球;

(b) 在第 1 个球是洋红色球的前提下第 2 个球是洋红色球.

解 (a) 因为两种球总数相同, 由对称性知第 2 个球是洋红色球的概率为 $\dfrac{1}{2}$.

(b) 同 (a) 分析可知第 1 个球是洋红色球的概率 $\mathbb{P}(A) = \dfrac{1}{2}$. 所取 2 个球均为洋红色球的概率

$$\mathbb{P}(B) = \sum_{r=1}^{n}\frac{1}{n}\cdot\frac{\mathrm{C}_{n-r}^2}{\mathrm{C}_{n-1}^2} = \sum_{r=1}^{n}\frac{(n-r)(n-r-1)}{n(n-1)(n-2)} = \sum_{r=0}^{n}\frac{r(r-1)}{n(n-1)(n-2)} = \frac{1}{3}.$$

因此在第 1 个球是洋红色球的前提下第 2 个球是洋红色球的概率为

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

□

习题 (1.5.2) 掷 n 次骰子, 设第 i 次和第 j 次结果相同为事件 A_{ij}. 证明：事件 $\{A_{ij} : 1 \leqslant i < j \leqslant n\}$ 两两独立但不相互独立.

证明 对任意 $1 \leqslant i < j \leqslant n$, $\mathbb{P}(A_{ij}) = \dfrac{6}{6\times 6} = \dfrac{1}{6}$. 考虑 $i_1 < j_1$ 与 $i_2 < j_2$.

- 若 $\{i_1, j_1\} \cap \{i_2, j_2\} \neq \varnothing$, 不妨设 $i_1 < j_1 = i_2 < j_2$, 则

$$\mathbb{P}(A_{i_1j_1}A_{i_2j_2}) = \mathbb{P}\left(\text{第 } i_1, j_1, j_2 \text{ 次结果相同}\right) = \frac{6}{6\times 6\times 6} = \frac{1}{36} = \mathbb{P}(A_{i_1j_1})\mathbb{P}(A_{i_2j_2}).$$

- 若 $\{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing$, 则

$$\mathbb{P}(A_{i_1j_1}A_{i_2j_2})=\frac{6\times 6}{6\times 6\times 6\times 6}=\frac{1}{36}=\mathbb{P}(A_{i_1j_1})\mathbb{P}(A_{i_2j_2}).$$

综上可知总有 $\mathbb{P}(A_{i_1j_1}A_{i_2j_2})=\mathbb{P}(A_{i_1j_1})\mathbb{P}(A_{i_2j_2})$, 即事件 $\{A_{ij}:1\leqslant i<j\leqslant n\}$ 两两独立.

设 $1\leqslant i<j<k\leqslant n$. 则

$$\mathbb{P}(A_{ij}A_{jk}A_{ik})=\mathbb{P}(\text{第 } i,j,k \text{ 次结果相同})=\frac{6}{6\times 6\times 6}=\frac{1}{36}\neq\mathbb{P}(A_{ij})\mathbb{P}(A_{jk})\mathbb{P}(A_{ik}),$$

即事件 $\{A_{ij}:1\leqslant i<j\leqslant n\}$ 不相互独立. □

习题 (1.5.4) 设 $\Omega=\{1,2,\cdots,p\}$, 其中 p 是素数, $\mathcal{F}=\{0,1\}^{\Omega}$ 是 Ω 的幂集, $\mathbb{P}(A)=\dfrac{|A|}{p}$, $\forall A\in\mathcal{F}$. 证明：若 A 与 B 相互独立, 则 A 与 B 中至少有一者为 $\varnothing$ 或 Ω.

证明 由 A 与 B 相互独立得 $\dfrac{|AB|}{p}=\dfrac{|A||B|}{p^2}$, 即 $p|AB|=|A||B|$. 若 A 与 B 均不为 $\varnothing$ 或 Ω, 则 $|A|,|B|\in\{1,2,\cdots,p-1\}$, 从而 $p\nmid|A|$, $p\nmid|B|$, 进而 $p\nmid|A||B|$, 但这与 $p|AB|=|A||B|$ 矛盾. 故 A 与 B 中至少有一者为 $\varnothing$ 或 Ω. □

习题 (1.5.7) Jane 怀有 3 个孩子, 他们的性别为男或女的概率均等且相互独立. 设有以下事件：

$$A=\{\text{所有孩子性别相同}\},$$
$$B=\{\text{至多有一个男孩}\},$$
$$C=\{\text{既有男孩又有女孩}\}.$$

(a) 证明 A 与 B 相互独立、B 与 C 相互独立.

(b) A 与 C 是否相互独立？

(c) 若两种性别出现的概率不相等, 上述叙述是否仍成立？

(d) 若 Jane 怀有 4 个孩子, 上述叙述是否仍成立？

解 (a) 因为 $\mathbb{P}(A)=\dfrac{2}{2^3}=\dfrac{1}{4}$, $\mathbb{P}(B)=\dfrac{1+3}{2^3}=\dfrac{1}{2}$, $\mathbb{P}(AB)=\dfrac{1}{2^3}=\dfrac{1}{8}=\mathbb{P}(A)\mathbb{P}(B)$, 所以 A 与 B 相互独立. 又 $\mathbb{P}(C)=1-2\times\dfrac{1}{2^3}=\dfrac{3}{4}$, $\mathbb{P}(BC)=\dfrac{3}{2^3}=\dfrac{3}{8}=\mathbb{P}(B)\mathbb{P}(C)$, 所以 B 与 C 相互独立.

(b) 因为 $\mathbb{P}(AC)=0\neq\mathbb{P}(A)\mathbb{P}(C)$, 所以 A 与 C 非相互独立.

(c) 设男孩、女孩出现的概率分别为 p,q , $p+q=1$ 且 $p\neq q$, 则 $\mathbb{P}(A)=p^3+q^3$, $\mathbb{P}(B)=q^3+3pq^2$, $\mathbb{P}(C)=1-p^3-q^3$, $\mathbb{P}(AB)=q^3$, $\mathbb{P}(BC)=3pq^2$, $\mathbb{P}(AC)=0$. 由此计算知 A 与 B 相互独立 $\iff$ $p=0$, B 与 C 相互独立 $\iff$ $p=0$ 或 $p=1$, A 与 C 相互独立 $\iff$ $p=0$ 或 $p=1$.

(d) 此时 $\mathbb{P}(A)=\dfrac{2}{2^4}=\dfrac{1}{8}$, $\mathbb{P}(B)=\dfrac{1+4}{2^4}=\dfrac{5}{16}$, $\mathbb{P}(C)=1-2\times\dfrac{1}{2^4}=\dfrac{7}{8}$, $\mathbb{P}(AB)=\dfrac{1}{2^4}=\dfrac{1}{16}$, $\mathbb{P}(BC)=\dfrac{4}{2^4}=\dfrac{1}{4}$, $\mathbb{P}(AC)=0$, 所以 A 与 B、B 与 C、A 与 C 均非相互独立. □

习题 (1.7.1) 从 A 到 B 和从 B 到 C 各有两条路, 每条路被大雪封住的概率均为 p 且彼此独立. 求在从 A 到 C 没有走得通的路的条件下从 A 到 B 走得通的概率.

若还有一条直接从 A 到 C 的路, 这条路被大雪封住的概率为 p 且与其他路的独立, 求此时如上的条件概率.

解 从 A 到 B 走得通与从 B 到 C 走得通的概率均为 $p_1=1-p^2$, 因此所求概率

$$p_2=\frac{\mathbb{P}(\text{从 } A \text{ 到 } B \text{ 走得通但从 } B \text{ 到 } C \text{ 走不通})}{\mathbb{P}(\text{从 } A \text{ 到 } C \text{ 走不通})}=\frac{(1-p^2)p^2}{1-(1-p^2)^2}=\begin{cases}0, & p=0,\\ \dfrac{1-p^2}{2-p^2}, & p\neq 0.\end{cases}$$

若还有一条直接从 A 到 C 的路, 上式中分子、分母均各乘上 p, 因此结果不变 (要求 $p\neq 0$, 否则此条件概率无意义), 即 $\dfrac{1-p^2}{2-p^2}$. □

习题 (1.7.3) 设整数 $0,1,2,\cdots,N$ 上的对称随机游走带吸收壁 0 和 N, 起点为 k. 证明该随机游走永远不被吸收的概率为 0.

证明 设起点为 k 的随机游走在 0 处被吸收的概率为 p_k, 在 N 处被吸收的概率为 q_k. 则由

$$\begin{cases}p_k=\dfrac{1}{2}\left(p_{k-1}+p_{k+1}\right)\\ p_0=1\\ p_N=0\end{cases} \qquad \text{与} \qquad \begin{cases}q_k=\dfrac{1}{2}\left(q_{k-1}+q_{k+1}\right)\\ q_0=0\\ q_N=1\end{cases}$$

可得

$$p_k=1-\frac{k}{N},\quad q_k=\frac{k}{N}.$$

于是以 k 为起点的随机游走永远不被吸收的概率为

$$1-p_k-q_k=1-\left(1-\frac{k}{N}\right)-\frac{k}{N}=0.$$

□

习题 (1.8.20) 重复掷一枚不均匀的硬币, 每次人像面朝上的概率为 p. 设 n 次试验后人像面朝上次数为偶数的概率为 p_n (0 也算偶数). 证明

$$p_n = \begin{cases} 1, & n = 0, \\ p(1-p_{n-1}) + (1-p)p_{n-1}, & n \geqslant 1. \end{cases}$$

并解这个差分方程.

解 由定义知 $p_0 = 1$. 对 $n \geqslant 1$, 若前 $n-1$ 次试验后人像面朝上次数为偶数, 则第 n 次试验结束后人像面朝上次数为偶数的概率为 $1-p$; 若前 $n-1$ 次试验后人像面朝上次数为奇数, 则第 n 次试验结束后人像面朝上次数为偶数的概率为 p. 因此由全概公式即得

$$p_n = p(1-p_{n-1}) + (1-p)p_{n-1}.$$

于是 $p_n - \dfrac{1}{2} = (1-2p)\left(p_{n-1} - \dfrac{1}{2}\right)$, 其中 $p \neq \dfrac{1}{2}$. 由 $p_0 - \dfrac{1}{2} = \dfrac{1}{2}$ 得 $p_n - \dfrac{1}{2} = \dfrac{1}{2}(1-2p)^n$, 即 $p_n = \dfrac{1}{2} + \dfrac{1}{2}(1-2p)^n$. □

补充题 1 掷一枚均匀的硬币. 甲掷 $n+1$ 次, 乙掷 n 次. 求甲试验得到的正面数比乙试验得到的正面数多的概率.

解 设在甲掷 $n+1$ 次、乙掷 n 次后甲、乙得到的正面数分别为 a, b, 则由

$$a \leqslant b \iff n+1-a \geqslant n+1-b \iff n+1-a > n-b$$

可知 $\mathbb{P}(\text{甲正} \leqslant \text{乙正}) = \mathbb{P}(\text{甲反} > \text{乙反})$. 而由对称性可知 $\mathbb{P}(\text{甲反} > \text{乙反}) = \mathbb{P}(\text{甲正} > \text{乙正}) = 1 - \mathbb{P}(\text{甲正} \leqslant \text{乙正})$, 因此 $\mathbb{P}(\text{甲正} > \text{乙正}) = \dfrac{1}{2}$. □

习题 (2.1.2) 设随机变量 X 有分布函数 F. 求 $Y = aX + b$ $(a, b \in \mathbb{R})$ 的分布函数.

解 ① 当 $a = 0$ 时, $\mathbb{P}(Y \leqslant y) = \begin{cases} 1, & y \geqslant b, \\ 0, & y < b. \end{cases}$

② 当 $a > 0$ 时, $\mathbb{P}(Y \leqslant y) = \mathbb{P}\left(Y \leqslant \dfrac{y-b}{a}\right) = F\left(\dfrac{y-b}{a}\right)$.

③ 当 $a < 0$ 时, $\mathbb{P}(Y \leqslant y) = \mathbb{P}\left(Y \geqslant \dfrac{y-b}{a}\right) = 1 - F\left(\left(\dfrac{y-b}{a}\right)^-\right)$. □

习题 (2.1.3) 掷一枚均匀的硬币 n 次. 证明：在公平的假设下, 恰出现 k 次人像面的概率为 $\binom{n}{k}\left(\frac{1}{2}\right)^n$. 若人像面每次出现的概率为 p, 这个概率变为多少？

证明 样本数 $|\Omega| = 2^n$, 事件数 $|A| = \mathrm{C}_n^k$, 故所求概率为 $\dfrac{|A|}{|\Omega|} = \binom{n}{k}\left(\frac{1}{2}\right)^n$. 若人像面每次出现的概率为 p, 则该概率为 $\binom{n}{k}p^k(1-p)^{n-k}$. □

习题 (2.1.4) 设 F 和 G 是分布函数, $0 \leqslant \lambda \leqslant 1$. 证明：$\lambda F + (1-\lambda)G$ 是分布函数. 乘积函数 FG 是分布函数吗？

证明 记 $H = \lambda F + (1-\lambda)G$, 由 $\lim\limits_{x\to-\infty} H(x) = 0$, $\lim\limits_{x\to+\infty} H(x) = \lambda + (1-\lambda) = 1$, 以及 $H(x)$ 单调增可知 H 是分布函数.

因为 $\lim\limits_{x\to-\infty} F(x)G(x) = 0$, $\lim\limits_{x\to+\infty} F(x)G(x) = 1$, $F(x)G(x)$ 单调增, 所以 FG 也是分布函数. □

习题 (2.1.5(b)(d)) 设 F 是分布函数, r 是一个正整数. 证明以下函数为分布函数：

(b) $1 - \{1 - F(x)\}^r$;

(d) $\{F(x) - 1\}\,\mathrm{e} + \exp\{1 - F(x)\}$.

证明 (b) 记 $H(x) = 1 - \{1 - F(x)\}^r$, 由 $\lim\limits_{x\to-\infty} H(x) = 0$, $\lim\limits_{x\to+\infty} H(x) = 1$, 以及 $H(x)$ 单调增可知 H 的分布函数.

(d) 记 $H(x) = \{F(x) - 1\}\,\mathrm{e} + \exp\{1 - F(x)\}$, 由 $\lim\limits_{x\to-\infty} H(x) = 0$, $\lim\limits_{x\to+\infty} H(x) = 1$, 以及 $H(x)$ 单调增可知 H 的分布函数. 其中 H 单调增可由函数 $f(x) = \mathrm{e}(x-1) + \mathrm{e}^{1-x}$ 的导数 $f'(x) = \mathrm{e} - \mathrm{e}^{1-x} \geqslant 0$ $(x \in [0,1])$ 得到. □

习题 (2.3.3) 设随机变量 X 的分布函数为

$$\mathbb{P}(X \leqslant x) = \begin{cases} 0, & x \leqslant 0, \\ x, & 0 < x \leqslant 1, \\ 1, & x > 1, \end{cases}$$

设 F 是连续且严格单调增的分布函数. 证明：$Y = F^{-1}(X)$ 是以 F 为分布函数的随机变量. 这里 F 连续和/或严格单调增的要求是必要的吗？

证明 由 $\{Y \leqslant y\} = \{F^{-1}(X) \leqslant Y\} = \{X \leqslant F(y)\}$ 知 Y 是随机变量. 再由

$$\mathbb{P}(Y \leqslant y) = \mathbb{P}(F^{-1}(X) \leqslant y) = \mathbb{P}(X \leqslant F(y)) \xlongequal{F(y)\in[0,1]} F(y)$$

即知 F 是 Y 的分布函数. 若没有连续性或严格单调性, 则无法确保 F^{-1} 的存在性. □

习题 (2.3.5) 下列哪些函数是密度函数? 求 c 以及相应的分布函数 F.

(a) $f(x) = \begin{cases} cx^{-d}, & x > 1, \\ 0, & \text{其他}. \end{cases}$

(b) $f(x) = c\,\mathrm{e}^x\,(1+\mathrm{e}^x)^{-2}$.

证明 (a) 若 $d > 1$, 则由 $\displaystyle\int_1^{+\infty} cx^{-d}\,\mathrm{d}x = \frac{c}{d-1}$ 知当 $c = d-1$ 时 f 是密度函数, 此时分布函数 $F(x) = 1 - x^{1-d}$. 若 $d \leqslant 1$, 则在 $c \neq 0$ 时积分 $\displaystyle\int_1^{+\infty} cx^{-d}\,\mathrm{d}x$ 不收敛, f 不可能是密度函数.

(b) 由 $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = c$ 知当 $c = 1$ 时 f 是密度函数, 此时分布函数 $F(x) = \displaystyle\int_{-\infty}^{x} f(u)\,\mathrm{d}u = 1 - \frac{1}{1+\mathrm{e}^x} = \frac{\mathrm{e}^x}{1+\mathrm{e}^x}$. □

习题 (2.4.2) 设随机变量 X 的分布函数为 F, $a < b$. 画出以下"截断"随机变量 Y 和 Z 分布函数的草图.

$$Y = \begin{cases} a, & X < a, \\ X, & a \leqslant X \leqslant b, \\ b, & X > b, \end{cases} \qquad Z = \begin{cases} X, & |X| \leqslant b, \\ 0, & |X| > b. \end{cases}$$

并指出这些分布函数在 $a \to -\infty, b \to +\infty$ 时的性态.

解 ① Y 的分布函数为

$$F_Y(y) = \begin{cases} 0, & y < a, \\ F(y), & a \leqslant x < b, \\ 1, & x \geqslant b. \end{cases}$$

② 若 $b \leqslant 0$, 则 $Z \equiv 0$, 因此 Z 的分布函数为

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ 1, & z \geqslant 0. \end{cases}$$

若 $b>0$, 注意到当 $-b\leqslant z<0$ 时,

$$\{Z\leqslant z\}=\{|X|\leqslant b\}\cap\{X\leqslant z\}=\{-b\leqslant X\leqslant z\},$$

而当 $0\leqslant z<b$ 时,

$$\{Z\leqslant z\}=(\{|X|\leqslant b\}\cap\{X\leqslant z\})\cup\{|X|>b\}=\{X\leqslant z\}\cup\{X>b\},$$

因此 Z 的分布函数为

$$F_Z(z)=\begin{cases}0, & z<-b,\\ F(z)-F(-b^-), & -b\leqslant z<0,\\ 1-F(b)+F(z), & 0\leqslant z<b,\\ 1, & z\geqslant b.\end{cases}$$

③ 当 $a\to-\infty,b\to+\infty$ 时 Y 和 Z 的分布函数 (逐点) 收敛于 X 的分布函数 F. □

习题 (2.5.2) 设 X 是 Bernoulli 随机变量, 满足 $\mathbb{P}(X=0)=1-p$, $\mathbb{P}(X=1)=p$. 设 $Y=1-X$, $Z=XY$. 求 $\mathbb{P}(X=x,Y=y)$ 与 $\mathbb{P}(X=x,Z=z)$, 其中 $x,y,z\in\{0,1\}$.

解 由题即得

$$\mathbb{P}(X=x,Y=y)=\begin{cases}1-p, & (x,y)=(0,1),\\ p, & (x,y)=(1,0),\\ 0, & \text{其他}.\end{cases}$$

以及

$$\mathbb{P}(X=x,Z=z)=\begin{cases}1-p, & (x,z)=(0,0),\\ p, & (x,z)=(1,0),\\ 0, & \text{其他}.\end{cases}$$

□

习题 (2.5.5) 设 X,Y 是取值为整数的离散型随机变量, 它们的联合分布列为 f. 证明：对 $x,y\in\mathbb{Z}$, 有

$$\begin{aligned}f(x,y)=&\mathbb{P}(X\geqslant x,Y\leqslant y)-\mathbb{P}(X\geqslant x+1,Y\leqslant y)\\&-\mathbb{P}(X\geqslant x,Y\leqslant y-1)+\mathbb{P}(X\geqslant x+1,Y\leqslant y-1).\end{aligned}$$

由此求掷 r 次均匀骰子得到的最小数和最大数的联合分布列.

证明 $\text{RHS}=\mathbb{P}(x\leqslant X<x+1,y-1<Y\leqslant y)$, 再由 X,Y 取值均为整数知这即是 $\mathbb{P}(X=x,Y=y)=\text{LHS}$. 对于所给具体情境, 分别用 X,Y 表示得到的最小数和最大数, 则对 $1\leqslant x\leqslant y\leqslant 6$, 有

$$f(x,y)=\begin{cases}\left(\dfrac{y-x+1}{6}\right)^r-2\left(\dfrac{y-x}{6}\right)^r+\left(\dfrac{y-x-1}{6}\right)^r, & x<y,\\ \left(\dfrac{1}{6}\right)^r, & x=y.\end{cases}$$

□

习题 (2.7.9) 将下列分布函数表示成 X 的分布函数 F 的形式：

$$X^+=\max\{0,X\},\quad X^-=-\min\{0,X\},\quad |X|=X^++X^-,\quad -X.$$

解 ① $\mathbb{P}\left(X^+\leqslant x\right)=\begin{cases}F(x), & x\geqslant 0,\\ 0, & x<0.\end{cases}$

② 由 $-\min\{0,X\}=\max\{0,-X\}$ 及①得 $\mathbb{P}\left(X^+\leqslant x\right)=\begin{cases}\mathbb{P}(-X\leqslant x), & x\geqslant 0,\\ 0, & x<0.\end{cases}$ 再由 $\mathbb{P}(-X\leqslant x)=\mathbb{P}(X\geqslant -x)=1-F(-x^-)$ 得 $\mathbb{P}(X^-\leqslant x)=\begin{cases}1-F(-x^-), & x\geqslant 0,\\ 0, & x<0.\end{cases}$

③ 由 $|X|=X^++X^-=\max\{0,X\}+\max\{0,-X\}=\max\{0,X,-X\}$ 知, 当 $x\geqslant 0$ 时, $\mathbb{P}(|X|\leqslant x)=\mathbb{P}(-x\leqslant X\leqslant x)$, 所以

$$\mathbb{P}(|X|\leqslant x)=\begin{cases}F(x)-F(-x^-), & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

④ $\mathbb{P}(-X\leqslant x)=\mathbb{P}(X\geqslant -x)=1-\mathbb{P}(X<-x)=1-F(-x^-)$. □

补充题 2 设 X,Y 是随机变量. 证明：$\min\{X,Y\}$ 和 $\max\{X,Y\}$ 也是随机变量.

证明 1 注意到 $\min\{X,Y\}=\dfrac{X+Y}{2}-\dfrac{|X-Y|}{2}$, $\max\{X,Y\}=\dfrac{X+Y}{2}+\dfrac{|X-Y|}{2}$. 由 X,Y 是随机变量可知 $X+Y$ 也是随机变量, 再由 $\left\{\dfrac{X+Y}{2}\leqslant x\right\}=\{X+Y\leqslant 2x\}$, $\forall x\in\mathbb{R}$ 知 $\dfrac{X+Y}{2}$ 是随机变量. 因此只需证 $|X-Y|$ 是随机变量. 这可由

$$\{|X-Y|\leqslant x\}=\{-x\leqslant X-Y\leqslant x\}=\{X-Y\leqslant x\}\cap\{Y-X\leqslant x\},\quad \forall x\in\mathbb{R}$$

及 σ 域对集合交运算封闭 (习题 (1.2.2)) 得到. □

证明 2 只需注意到

$$\{\min\{X,Y\}\leqslant x\}=\{X\leqslant x\}\cap\{Y\leqslant x\},\quad \{\max\{X,Y\}\leqslant x\}=\{X\leqslant x\}\cup\{Y\leqslant x\}.$$

□

习题 (3.1.5) (a) 证明:若随机变量 X 服从二项分布或 Poisson 分布, 则分布列 $f(k)=\mathbb{P}(X=k)$ 满足 $f(k-1)f(k+1)\leqslant f(k)^2$.

(b) 证明:若 $f(k)=\dfrac{90}{(\pi k)^4}$, $k\geqslant 1$, 则 $f(k-1)f(k+1)\geqslant f(k)^2$.

(c) 举例:分布列 f 满足 $f(k)^2=f(k-1)f(k+1)$, $k\geqslant 1$.

解 (a) ① 若 $X\sim B(n,p)$, 则 $f(k)=\mathrm{C}_n^k p^k(1-p)^{n-k}$. 欲证 $f(k-1)f(k+1)\leqslant f(k)^2$, 即证

$$\mathrm{C}_n^{k-1}\mathrm{C}_n^{k+1}\leqslant \mathrm{C}_n^k\mathrm{C}_n^k,$$

展开整理即证

$$k(n-k)\leqslant (n-k+1)(k+1),$$

这在 $n\geqslant k+1$ 时是显然的.

② 若 $X\sim P(\lambda)$, 则 $f(k)=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$. 欲证 $f(k-1)f(k+1)\leqslant f(k)^2$, 即证

$$(k!)^2\leqslant (k-1)!(k+1)!,$$

这等价于 $k\leqslant k+1$.

(b) 即证 $k^8\geqslant (k-1)^4(k+1)^4$, 这由 $k^2\geqslant k^2-1$ 立得.

(c) 设 X 服从几何分布, $f(k)=p(1-p)^{k-1}$, 则 $f(k)^2=f(k-1)f(k+1)$, $k\geqslant 1$. □

习题 (3.2.3) 设随机变量 X_1,X_2,X_3 相互独立且取值均为正整数, 它们的分布列为 $\mathbb{P}(X_i=x)=(1-p_i)p_i^{x-1}$, $x=1,2,\cdots$, $i=1,2,3$.

(a) 证明:

$$\mathbb{P}(X_1<X_2<X_3)=\frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)}.$$

(b) 求 $\mathbb{P}(X_1\leqslant X_2\leqslant X_3)$.

证明 (a) 直接计算得

$$\begin{aligned}\mathbb{P}(X_1 < X_2 < X_3) &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{k=j+1}^{\infty}(1-p_1)(1-p_2)(1-p_3)p_1^{i-1}p_2^{j-1}p_3^{k-1} \\ &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}(1-p_1)(1-p_2)p_1^{i-1}p_2^{j-1}p_3^{j} \\ &= \sum_{i=1}^{\infty}\frac{p_3(1-p_1)(1-p_2)}{1-p_2p_3}p_1^{i-1}p_2^{i}p_3^{i} \\ &= \frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)}.\end{aligned}$$

(b) 直接计算得

$$\begin{aligned}\mathbb{P}(X_1 \leqslant X_2 \leqslant X_3) &= \sum_{i=1}^{\infty}\sum_{j=i}^{\infty}\sum_{k=j}^{\infty}(1-p_1)(1-p_2)(1-p_3)p_1^{i-1}p_2^{j-1}p_3^{k-1} \\ &= \sum_{i=1}^{\infty}\sum_{j=i}^{\infty}(1-p_1)(1-p_2)p_1^{i-1}p_2^{j-1}p_3^{j-1} \\ &= \sum_{i=1}^{\infty}\frac{(1-p_1)(1-p_2)}{1-p_2p_3}(p_1p_2p_3)^{i-1} \\ &= \frac{(1-p_1)(1-p_2)}{(1-p_2p_3)(1-p_1p_2p_3)}.\end{aligned}$$

□

习题 (3.2.5) 设随机变量 X_r $(1 \leqslant r \leqslant n)$ 相互独立且关于 0 对称, 即 X_r 与 $-X_r$ 有相同的分布列. 证明：对任意 x, $\mathbb{P}(S_n \geqslant x) = \mathbb{P}(S_n \leqslant -x)$, 其中 $S_n = \sum_{r=1}^{n} X_r$. 若去掉相互独立这一条件, 结论还一定成立吗？

证明 先证 $n=2$ 的情形. 设随机变量 X, Y 如题设所述, 则

$$\begin{aligned}\mathbb{P}(X+Y \geqslant x) &= \sum_k \mathbb{P}(X=k, Y \geqslant x-k) \xlongequal{\text{独立性}} \sum_k \mathbb{P}(X=k)\mathbb{P}(Y \geqslant x-k) \\ &= \sum_k \mathbb{P}(-X=k)\mathbb{P}(-Y \geqslant x-k) = \sum_k \mathbb{P}(X=-k)\mathbb{P}(Y \leqslant k-x) \\ &\xlongequal{\text{独立性}} \sum_k \mathbb{P}(X=-k, Y \leqslant k-x) = \mathbb{P}(X+Y \leqslant -x).\end{aligned}$$

现设结论对 $n-1$ 个随机变量成立, 则对 n 个如题设所述的随机变量 X_r $(1 \leqslant r \leqslant n)$, 有

$$\mathbb{P}(S_n \geqslant x) = \mathbb{P}(X_n + S_{n-1} \geqslant x) = \sum_k \mathbb{P}(X_n = k, S_{n-1} \geqslant x-k)$$

$$\xlongequal{\text{独立性}} \sum_k \mathbb{P}(X_n = k)\mathbb{P}(S_{n-1} \geqslant x - k) = \sum_k \mathbb{P}(-X_n = k)\mathbb{P}(S_{n-1} \leqslant k - x)$$
$$= \sum_k \mathbb{P}(X_n = -k)\mathbb{P}(S_{n-1} \leqslant k - x) = \mathbb{P}(S_n \leqslant -x).$$

由数学归纳法知结论成立.

若去掉相互独立这一条件, 结论不一定成立, 反例如下. 从 $\{1,2,3\}$ 中随机抽取一个数字 (等可能), 定义随机变量 X, Y 如下：

抽中数字	X	Y
1	-1	-1
2	0	1
3	1	0

则 X, Y 均关于 0 对称, 但

$$\mathbb{P}(X+Y \geqslant 1) = \mathbb{P}(X=1, Y=0) + \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=1, Y=1) = 0,$$
$$\mathbb{P}(X+Y \leqslant -1) = \mathbb{P}(X=-1, Y=0) + \mathbb{P}(X=0, Y=-1) + \mathbb{P}(X=-1, Y=-1) = \frac{1}{3}.$$

□

补充题 3 先掷一个均匀的骰子, 得到一个点数 k, 再抛 k 个均匀硬币, 求最终得到的正面个数的分布列.

解 用 X 表示最终得到的正面的个数, 则对 $n \in \{0,1,2,3,4,5,6\}$, 有

$$\mathbb{P}(X=n) = \sum_{i=\max\{1,n\}}^{6} \frac{1}{6} \mathrm{C}_i^n \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{i-n}.$$

计算即得如下分布列：

n	0	1	2	3	4	5	6
$\mathbb{P}(X=n)$	$\frac{21}{128}$	$\frac{5}{16}$	$\frac{33}{128}$	$\frac{1}{6}$	$\frac{29}{384}$	$\frac{1}{48}$	$\frac{1}{384}$

□

习题 (3.3.3(a)) 一个小组有 n 位玩家, 每人掷一个骰子. 每有两位玩家掷出同一点数, 小组就得 1 分. 求小组总分的期望和方差.

解 任意两位玩家掷出同一点数的概率

$$p = 6 \times \left(\frac{1}{6}\right)^2 = \frac{1}{6}.$$

因此小组总分的期望

$$\mu = \mathrm{C}_n^2 \cdot 1 \cdot p = \frac{n(n-1)}{12}.$$

设事件 $A_{ij} = \left\{\text{第 } i \text{ 位玩家与第 } j \text{ 位玩家掷出点数相同}\right\}$, 则由习题 (1.5.2) 可知事件 $\{A_{ij} : 1 \leqslant i < j \leqslant n\}$ 两两独立. 故

$$\begin{aligned}\operatorname{Var}\left(\sum_{i<j} I_{A_{ij}}\right) &= \sum_{i<j} \operatorname{Var}\left(I_{A_{ij}}\right) = \mathrm{C}_n^2 \mathbb{E}\left[\left(I_{A_{ij}} - p\right)^2\right] = \mathrm{C}_n^2\left[(1-p)^2 p + p^2(1-p)\right] \\ &= \frac{5n(n-1)}{72}.\end{aligned}$$

□

习题 (3.3.5) 设随机变量 X 有分布列

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)}, & x = 1, 2, \cdots, \\ 0, & \text{其他}, \end{cases}$$

并设 $\alpha \in \mathbb{R}$. 问 α 取何值时有 $\mathbb{E}\left[X^\alpha\right] < +\infty$?

解 我们有

$$\mathbb{E}\left[X^\alpha\right] = \sum_{k=1}^{\infty} \frac{k^\alpha}{k(k+1)},$$

由于 $\dfrac{k^\alpha}{k(k+1)} \sim k^{\alpha-2}$ $(k \to \infty)$, 该正项级数收敛 $\iff$ $\alpha < 1$. □

补充题 4 设 $X \sim B(n,p)$, 求 $\mathbb{E}\left[X^3\right]$.

解 我们有

$$\begin{aligned}\mathbb{E}\left[X(X-1)(X-2)\right] &= \sum_{k=0}^{n} k(k-1)(k-2)\mathrm{C}_n^k p^k(1-p)^{n-k} \\ &= \sum_{k=3}^{n} \frac{n!}{(k-3)!(n-k)!} p^k (1-p)^{n-k}\end{aligned}$$

$$
\begin{aligned}
&= n(n-1)(n-2)\sum_{k=3}^{n}\frac{(n-3)!}{(k-3)!(n-k)!}p^k(1-p)^{n-k}\\
&= n(n-1)(n-2)p^3\sum_{k=0}^{n-3}\frac{(n-3)!}{k!(n-3-k)!}p^k(1-p)^{n-3-k}\\
&= n(n-1)(n-2)p^3.
\end{aligned}
$$

于是

$$
\begin{aligned}
\mathbb{E}\left[X^3\right] &= \mathbb{E}\left[X(X-1)(X-2)\right] + 3\mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\\
&= n(n-1)(n-2)p^3 + 3np\left[1+(n-1)p\right] - 2np\\
&= np\left[(n-1)(n-2)p^2 + 3(n-1)p + 1\right].
\end{aligned}
$$

□

习题 (3.4.2) 一个缸中有标号为 $1,2,\cdots,n$ 的 n 个球. 从中 (不放回地) 随机取出 k 个球并把它们的标号相加. 求这个和数的期望和方差.

解 设第 i 个球的标号为 X_i, 则所求和数 $S=\sum\limits_{i=1}^{k}X_i$, 从而

$$
\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{k}X_i\right] = \sum_{i=1}^{k}\mathbb{E}[X_i] = k\mathbb{E}[X_1] = k\cdot\frac{1+\cdots+n}{n} = \frac{k(n+1)}{2}.
$$

又

$$
\begin{aligned}
\mathbb{E}\left[S^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{k}X_i\right)^2\right]\\
&= \sum_{i=1}^{k}\mathbb{E}\left[X_i^2\right] + 2\sum_{1\leqslant i<j\leqslant k}\mathbb{E}[X_iX_j]\\
&= k\mathbb{E}\left[X_1^2\right] + k(k-1)\mathbb{E}[X_1X_2]\\
&= k\cdot\frac{n(n+1)(2n+1)}{6n} + k(k-1)\cdot\frac{1}{\mathrm{C}_n^2}\sum_{1\leqslant i<j\leqslant k} ij\\
&= \frac{k(n+1)(2n+1)}{6} + \frac{2k(k-1)}{n(n-1)}\sum_{i=1}^{n-1} i\sum_{j=i+1}^{n} j\\
&= \frac{k(n+1)(2n+1)}{6} + \frac{k(k-1)}{n(n-1)}\sum_{i=1}^{n-1} i(n+i+1)(n-i)
\end{aligned}
$$

$$= \frac{k(n+1)(2n+1)}{6} + k(k-1)\left(\frac{1}{4}n^2 + \frac{5}{12}n + \frac{1}{6}\right).$$

故

$$\mathrm{Var}(S) = \mathbb{E}\left[S^2\right] - (\mathbb{E}[S])^2 = \frac{k(n+1)(n-k)}{12}.$$

□

习题 (3.4.4) R 缸有 n 个红球, B 缸有 n 个蓝球. 每次从两缸中各选一个球并交换. 证明：进行 k 次操作后缸 R 中的红球数的期望为 $\dfrac{n\left[1+\left(1-\frac{2}{n}\right)^k\right]}{2}$. Daniel Bernoulli 在 1769 年描述了这个“扩散模型”.

证明 设一个红球在 k 次操作后仍留在 R 缸中的概率为 p_k, 则

$$p_k = \left(1 - \frac{1}{n}\right)p_{k-1} + \frac{1}{n}(1 - p_{k-1}),$$

也即

$$p_k - \frac{1}{2} = \left(1 - \frac{2}{n}\right)\left(p_{k-1} - \frac{1}{2}\right).$$

结合 $p_1 = 1 - \dfrac{1}{n}$ 可解得

$$p_k = \frac{1}{2}\left[1 + \left(1 - \frac{2}{n}\right)^k\right].$$

对于固定的 k, 缸 R 中红球数服从二项分布, 从而红球数的期望为 np_k, 此即欲证. □

习题 (3.4.7) 设 $G = (V, E)$ 是有限图. 对任意顶点集 W 和任一边 $e \in E$, 定义示性函数

$$I_W(e) = \begin{cases} 1, & e \text{ 连接 } W \text{ 与 } W^{\mathrm{c}}, \\ 0, & \text{其他}. \end{cases}$$

设 $N_W = \sum_{e \in E} I_W(e)$. 证明：存在 $W \subset V$ 使得 $N_W \geqslant \dfrac{|E|}{2}$.

证明 1 通过有限求和交换次序可得

$$\sum_{W \in \{0,1\}^V} \sum_{e \in E} I_W(e) = \sum_{e \in E} \sum_{W \in \{0,1\}^V} I_W(e) = \sum_{e \in E} 2 \cdot 2^{|V|-2} = |E| \cdot 2^{|V|-1}.$$

若对任意 $W \subset V$ 都有 $N_W < \dfrac{|E|}{2}$, 则

$$\sum_{W \in \{0,1\}^V} \sum_{e \in E} I_W(e) < 2^{|V|} \cdot \frac{|E|}{2} = |E| \cdot 2^{|V|-1},$$

矛盾. 故存在 $W \subset V$ 使得 $N_W \geqslant \dfrac{|E|}{2}$. □

证明 2 取 $W \subset V$ 使得 $\mathbb{P}(v \in W) = \dfrac{1}{2}$, 且不同顶点是否在 W 中是相互独立的. 则

$$I_W(e) = 2 \times \left(\frac{1}{2}\right)^2 = \frac{1}{2}, \quad \forall e \in E.$$

于是

$$\mathbb{E}[N_W] = \mathbb{E}\left[\sum_{e \in E} I_W(e)\right] = \sum_{e \in E} \mathbb{E}\left[I_W(e)\right] = \frac{|E|}{2}.$$

我们断言 $\mathbb{P}\left(N_W \geqslant \dfrac{|E|}{2}\right) > 0$, 否则 $\mathbb{P}\left(N_W < \dfrac{|E|}{2}\right) = 1$, 进而由期望的非负性得 $\mathbb{E}[N_W] < \dfrac{|E|}{2}$, 矛盾. 这表明存在 $W \subset V$ 使得 $N_W \geqslant \dfrac{|E|}{2}$. □

习题 (3.5.2) 设口袋中有 N 个硬币, 其中随机变量 $N \sim P(\lambda)$. 现将每枚硬币都掷一次, 设每次正面朝上的概率均为 p. 证明：出现的正面总数服从以 λp 为参数的 Poisson 分布.

证明 设出现的正面总数为 X, 则对 $r \in \mathbb{N}$, 有

$$\begin{aligned}
\mathbb{P}(X = r) &= \sum_{k=r}^{\infty} \mathbb{P}(N = k) \mathrm{C}_k^r p^r (1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \frac{k!}{r!(k-r)!} p^r (1-p)^{k-r} \\
&= \frac{\mathrm{e}^{-\lambda}(\lambda p)^r}{r!} \sum_{k=0}^{\infty} \frac{[\lambda(1-p)]^k}{k!} \\
&= \frac{\mathrm{e}^{-\lambda}(\lambda p)^r}{r!} \cdot \mathrm{e}^{\lambda(1-p)} \\
&= \frac{(\lambda p)^r}{r!} \mathrm{e}^{-\lambda p}.
\end{aligned}$$

故 $X \sim P(\lambda p)$. □

补充题 5 设 $\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{R}^n$, 其中 $\|\mathbf{v}_i\|_2 \leqslant 1$, $\forall i$. 令 $\mathbf{w} = \sum\limits_{i=1}^{n} p_i \mathbf{v}_i$, 其中 $p_i \in [0,1]$, $\forall i$. 证明：存在 $\varepsilon_1, \cdots, \varepsilon_n \in \{0,1\}$, 使得

$$\left\|\mathbf{w} - \sum_{i=1}^{n} \varepsilon_i \mathbf{v}_i\right\|_2 \leqslant \frac{\sqrt{n}}{2}.$$

证明 设 $\mathbb{P}(\varepsilon_i=1)=p_i$, $\mathbb{P}(\varepsilon_i=0)=1-p_i$, 不同的 ε_i 之间的选取相互独立. 设随机变量 $X=\left\|\mathbf{w}-\sum_{i=1}^{n}\varepsilon_i\mathbf{v}_i\right\|_2^2$, 则

$$X=\left\|\sum_{i=1}^{n}(p_i-\varepsilon_i)\mathbf{v}_i\right\|_2^2=\sum_{i=1}^{n}\sum_{j=1}^{n}\langle\mathbf{v}_i,\mathbf{v}_j\rangle(p_i-\varepsilon_i)(p_j-\varepsilon_j).$$

于是 X 的期望

$$\begin{aligned}\mathbb{E}[X]&=\sum_{i=1}^{n}\sum_{j=1}^{n}\langle\mathbf{v}_i,\mathbf{v}_j\rangle\,\mathbb{E}\left[(p_i-\varepsilon_i)(p_j-\varepsilon_j)\right]\\&=\sum_{i\neq j}\langle\mathbf{v}_i,\mathbf{v}_j\rangle\,\mathbb{E}\left[(p_i-\varepsilon_i)(p_j-\varepsilon_j)\right]+\sum_{i=1}^{n}\|\mathbf{v}_i\|_2^2\mathbb{E}\left[(p_i-\varepsilon_i)^2\right]\\&=\sum_{i\neq j}\langle\mathbf{v}_i,\mathbf{v}_j\rangle\underbrace{\mathbb{E}[p_i-\varepsilon_i]}_{0}\underbrace{\mathbb{E}[p_j-\varepsilon_j]}_{0}+\sum_{i=1}^{n}\|\mathbf{v}_i\|_2^2\left[\mathbb{E}\left[\varepsilon_i^2\right]-2p_i\mathbb{E}[\varepsilon_i]+p_i^2\right]\\&=\sum_{i=1}^{n}\|\mathbf{v}_i\|_2^2\left(p_i-p_i^2\right)\\&\leqslant\frac{1}{4}\sum_{i=1}^{n}\|\mathbf{v}_i\|_2^2\\&\leqslant\frac{n}{4}.\end{aligned}$$

我们断言 $\mathbb{P}\left(X\leqslant\frac{n}{4}\right)>0$, 否则 $\mathbb{P}\left(X>\frac{n}{4}\right)=1$, 进而由期望的非负性得 $\mathbb{E}[X]>\frac{n}{4}$, 矛盾. 这表明 $\left\{X\leqslant\frac{n}{4}\right\}$ 非空, 即存在 $\varepsilon_1,\cdots,\varepsilon_n\in\{0,1\}$, 使得 $\left\|\mathbf{w}-\sum_{i=1}^{n}\varepsilon_i\mathbf{v}_i\right\|_2\leqslant\frac{\sqrt{n}}{2}$. □

习题 (3.6.4) 设离散型随机变量 X,Y 的期望均为 0, 方差均为 1, 协方差为 ρ. 证明:

$$\mathbb{E}\left[\max\left\{X^2,Y^2\right\}\right]\leqslant1+\sqrt{1-\rho^2}.$$

证明 由已知, 我们有

$$\rho\coloneqq\operatorname{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]=\mathbb{E}[XY],$$
$$\mathbb{E}\left[X^2\right]=\operatorname{Var}(X)+(\mathbb{E}[X])^2=1,\quad\mathbb{E}\left[Y^2\right]=\operatorname{Var}(Y)+(\mathbb{E}[Y])^2=1.$$

由 Cauchy-Schwarz 不等式,

$$\begin{aligned}\mathbb{E}\left[\max\left\{X^2,Y^2\right\}\right]&=\frac{1}{2}\mathbb{E}\left[\left(X^2+Y^2\right)+\left|X^2-Y^2\right|\right]\\&=\frac{1}{2}\mathbb{E}\left[X^2+Y^2\right]+\frac{1}{2}\mathbb{E}\left[\left|(X+Y)(X-Y)\right|\right]\end{aligned}$$

$$
\begin{aligned}
&\leqslant \frac{1}{2}\left(\mathbb{E}\left[X^2\right]+\mathbb{E}\left[Y^2\right]\right)+\frac{1}{2}\sqrt{\mathbb{E}\left[(X+Y)^2\right]\mathbb{E}\left[(X-Y)^2\right]}\\
&=\frac{1+1}{2}+\frac{\sqrt{(1+1+2\rho)(1+1-2\rho)}}{2}\\
&=1+\sqrt{1-\rho^2}.
\end{aligned}
$$

□

习题 (3.6.5) 设离散型随机变量 X, Y 有联合分布列 f.

(a) 证明：$\mathbb{E}[\ln f_X(X)] \geqslant \mathbb{E}[\ln f_Y(X)]$.

(b) 证明：互信息

$$I=\mathbb{E}\left[\ln\left(\frac{f(X,Y)}{f_X(X)f_Y(Y)}\right)\right]$$

满足 $I \geqslant 0$, 等号成立当且仅当 X 与 Y 相互独立.

证明 (a) 由期望的非负性及佚名统计学家公式,

$$
\begin{aligned}
\mathrm{RHS}-\mathrm{LHS}&=\mathbb{E}\left[\ln\frac{f_Y(X)}{f_X(X)}\right]\leqslant\mathbb{E}\left[\frac{f_Y(X)}{f_X(X)}-1\right]\\
&=\sum_x\left(\frac{f_Y(x)}{f_X(x)}-1\right)f_X(x)=\sum_x\left[f_Y(x)-f_X(x)\right]=0.
\end{aligned}
$$

(b) 我们有

$$
\begin{aligned}
I&=-\mathbb{E}\left[\ln\left(\frac{f_X(X)f_Y(Y)}{f(X,Y)}\right)\right]\geqslant\mathbb{E}\left[1-\frac{f_X(X)f_Y(Y)}{f(X,Y)}\right]\\
&=\sum_{x,y}\left(1-\frac{f_X(x)f_Y(y)}{f(x,y)}\right)f(x,y)=\sum_{x,y}\left[f(x,y)-f_X(x)f_Y(y)\right]\\
&=\sum_{x,y}f(x,y)-\sum_x f_X(x)\sum_y f_Y(y)=0.
\end{aligned}
$$

等号成立当且仅当

$$f_X(X)f_Y(Y)=f_X(X)f_Y(Y),$$

也即 X 与 Y 相互独立. □

习题 (3.7.1(a)(b)(c)(d)(e)) 完成以下证明：

(a) $\mathbb{E}[aY+bZ \mid X]=a\mathbb{E}[Y\mid X]+b\mathbb{E}[Z\mid X], \forall a,b\in\mathbb{R}$.

(b) 若 $Y\geqslant 0$, 则 $\mathbb{E}[Y\mid X]\geqslant 0$.

(c) $\mathbb{E}[1 \mid X]=1$.

(d) 若 X 与 Y 相互独立, 则 $\mathbb{E}[Y \mid X]=\mathbb{E}[Y]$.

(e) $\mathbb{E}[Yg(X) \mid X]=g(X)\mathbb{E}[Y \mid X]$, 其中函数 g 使得等式两边的表达式均有意义.

证明 (a) 对任意 $x \in \mathbb{R}$, 有

$$\begin{aligned}\mathbb{E}[aY+bZ \mid X=x] &= \sum_{y,z}(ay+bz)\mathbb{P}(Y=y, Z=z \mid X=x)\\ &= a\sum_{y,z} y\mathbb{P}(Y=y, Z=z \mid X=x)+b\sum_{y,z} z\mathbb{P}(Y=y, Z=z \mid X=x)\\ &= a\sum_{y} y\mathbb{P}(Y=y \mid X=x)+b\sum_{z} z\mathbb{P}(Z=z \mid X=x).\end{aligned}$$

(b) 对任意 $x \in \mathbb{R}$, 有

$$\mathbb{E}[Y \mid X=x]=\sum_{y} y\mathbb{P}(Y=y \mid X=x) \geqslant 0.$$

(c) 记 $Y=1$, 则对任意 $x \in \mathbb{R}$, 有

$$\mathbb{E}[1 \mid X=x]=\sum_{y} y f_{Y|X}(y \mid x)=\mathbb{P}(Y=1 \mid X=x)=1.$$

(d) 对任意 $x \in \mathbb{R}$, 有

$$\begin{aligned}\mathbb{E}[Y \mid X=x] &= \sum_{y} y\mathbb{P}(Y=y \mid X=x)=\sum_{y} y\frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)}\\ &= \sum_{y} y\frac{\mathbb{P}(Y=y)\mathbb{P}(X=x)}{\mathbb{P}(X=x)}=\sum_{y} y\mathbb{P}(Y=y)\\ &= \mathbb{E}[Y].\end{aligned}$$

(e) 对任意 $\overline{x} \in \mathbb{R}$, 由 2 维佚名统计学家公式, 有

$$\begin{aligned}\mathbb{E}\left[Yg(X) \mid X=\overline{x}\right] &= \sum_{x,y} yg(x)\mathbb{P}(X=x, Y=y \mid X=\overline{x})\\ &= \sum_{y} yg\left(\overline{x}\right)\mathbb{P}(Y=y \mid X=\overline{x})\\ &= g\left(\overline{x}\right)\mathbb{E}[Y \mid X=\overline{x}].\end{aligned}$$

□

习题 (3.7.4) 如何定义 Y 关于 X 的条件方差 $\mathrm{Var}(Y \mid X)$？并证明：

$$\mathrm{Var}(Y)=\mathbb{E}[\mathrm{Var}(Y \mid X)]+\mathrm{Var}(\mathbb{E}[Y \mid X]).$$

解 Y 关于 X 的条件方差定义为

$$\mathrm{Var}(Y \mid X)=\mathbb{E}\left[(Y-\mathbb{E}[Y \mid X=x])^2 \mid X=x\right].$$

由

$$\mathbb{E}[\mathrm{Var}(Y \mid X)]=\mathbb{E}\left[\mathbb{E}\left[Y^2 \mid X\right]-(\mathbb{E}[Y \mid X])^2\right]=\mathbb{E}\left[Y^2\right]-\mathbb{E}\left[(\mathbb{E}[Y \mid X])^2\right]$$

以及

$$\mathrm{Var}(\mathbb{E}[Y \mid X])=\mathbb{E}\left[(\mathbb{E}[Y \mid X])^2\right]-(\mathbb{E}[\mathbb{E}[Y \mid X]])^2=\mathbb{E}\left[(\mathbb{E}[Y \mid X])^2\right]-(\mathbb{E}[Y])^2$$

可得

$$\mathbb{E}[\mathrm{Var}(Y \mid X)]+\mathrm{Var}(\mathbb{E}[Y \mid X])=\mathbb{E}\left[Y^2\right]-(\mathbb{E}[Y])^2=\mathrm{Var}(Y).$$

□

习题 (3.7.10) 一枚硬币正面朝上的概率为 p. 设 X_n 为得到连续 n 个正面朝上结果所需的抛掷数. 证明：$\mathbb{E}[X_n]=\sum_{k=1}^{n} p^{-k}$.

证明 利用 $\mathbb{E}[X_n]=\mathbb{E}[\mathbb{E}[X_n \mid X_{n-1}]]$ 可将问题转化为求条件期望 $\mathbb{E}[X_n \mid X_{n-1}]$. 由

$$\begin{aligned}
&\mathbb{E}[X_n \mid X_{n-1}=b]\\
=&\mathbb{P}\left(\text{第 } b+1 \text{ 次正面朝上}\right)(b+1)+\mathbb{P}\left(\text{第 } b+1 \text{ 次反面朝上}\right)\sum_{k=n}^{\infty}(b+1+k)\mathbb{P}(X_n=k)\\
=&p(b+1)+(1-p)\sum_{k=n}^{\infty}(b+1+k)\mathbb{P}(X_n=k)
\end{aligned}$$

可得

$$\begin{aligned}
\mathbb{E}[X_n \mid X_{n-1}]&=p(X_{n-1}+1)+(1-p)\sum_{k=n}^{\infty}(X_{n-1}+1+k)\mathbb{P}(X_n=k)\\
&=p(X_{n-1}+1)+(1-p)\left[(X_{n-1}+1)\underbrace{\sum_{k=n}^{\infty}\mathbb{P}(X_n=k)}_{1}+\underbrace{\sum_{k=n}^{\infty}k\mathbb{P}(X_n=k)}_{\mathbb{E}[X_n]}\right]
\end{aligned}$$

$$= p\,(X_{n-1}+1) + (1-p)\,[X_{n-1}+1+\mathbb{E}[X_n]]\,.$$

于是

$$\mathbb{E}[X_n] = \mathbb{E}\,[\mathbb{E}\,[X_n \mid X_{n-1}]] = p\,[\mathbb{E}[X_{n-1}]+1] + (1-p)\,[\mathbb{E}[X_{n-1}]+1+\mathbb{E}[X_n]]\,,$$

即

$$\mathbb{E}[X_n] = \frac{\mathbb{E}[X_{n-1}]+1}{p}.$$

又

$$\begin{aligned}\mathbb{E}[X_1] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = -p\sum_{k=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}p}(1-p)^k \\ &= -p\frac{\mathrm{d}}{\mathrm{d}p}\sum_{k=1}^{\infty}(1-p)^k = -p\frac{\mathrm{d}}{\mathrm{d}p}\left(\frac{1}{p}-1\right) \\ &= \frac{1}{p},\end{aligned}$$

故可解得

$$\mathbb{E}[X_n] = \sum_{k=1}^{n} p^{-k}.$$

□

习题 (3.8.6) 设 $N \sim P(\lambda)$. 证明：$\mathbb{E}[Ng(N)] = \lambda\mathbb{E}[g(N+1)]$, 其中 g 为使得等号两边期望均有意义的任一函数. 更一般地, 若 $S = \sum_{r=1}^{N} X_r$, 其中 $\{X_r : r \geqslant 0\}$ 是独立同分布非负整值随机变量, 且与 N 独立, 证明：

$$\mathbb{E}[Sg(S)] = \lambda\mathbb{E}[g(S+X_0)X_0].$$

证明 ① 利用佚名统计学家公式计算得

$$\begin{aligned}\mathbb{E}[Ng(N)] &= \sum_{k=0}^{\infty} kg(k)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \sum_{k=1}^{\infty} g(k)\frac{\lambda^k}{(k-1)!}\,\mathrm{e}^{-\lambda} = \lambda\sum_{k=1}^{\infty} g(k)\frac{\lambda^{k-1}}{(k-1)!}\,\mathrm{e}^{-\lambda} \\ &\xlongequal{k\to k+1} \lambda\sum_{k=0}^{\infty} g(k+1)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \lambda\mathbb{E}[g(N+1)].\end{aligned}$$

② 由全期望公式,

$$\begin{aligned}\mathbb{E}\,[Sg(S)] &= \mathbb{E}\,[\mathbb{E}\,[Sg(S) \mid N]] \\ &= \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}\,\mathbb{E}\,[Sg(S) \mid N=n]\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} n\mathbb{E}\left[X_1 g\left(\sum_{r=1}^{n} X_r\right)\right] \\
&= \lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda}\, \mathbb{E}\left[X_1 g\left(\sum_{r=1}^{n} X_r\right)\right] \\
&\xlongequal{n\to n+1} \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda}\, \mathbb{E}\left[X_1 g\left(\sum_{r=1}^{n+1} X_r\right)\right] \\
&= \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda}\, \mathbb{E}\left[g(S+X_0)X_0 \mid N=n\right] \\
&= \lambda\mathbb{E}\left[\mathbb{E}\left[g(S+X_0)X_0 \mid N\right]\right] \\
&= \lambda\mathbb{E}\left[g(S+x_0)X_0\right].
\end{aligned}
$$

□

习题 (3.9.4) 重复掷一枚硬币, 每次正面朝上的概率为 p. 若 m 次正面结果比 n 次反面结果先达到, 则玩家 A 获胜; 反之玩家 B 获胜. 设玩家 A 获胜的概率为 p_{mn}. 对 p_{mn} 建立差分方程. 它的边界条件是什么?

解 可建立差分方程

$$
p_{mn} = pp_{m-1,n} + (1-p)p_{m,n-1},
$$

边界条件为

$$
p_{m0} = 0, \quad p_{0n} = 1.
$$

□

习题 (3.10.1) 考虑对称简单随机游走 S, $S_0 = 0$. 设 $T = \min\{n \geqslant 1 : S_n = 0\}$ 为第一次回到出发点的时刻. 证明:

$$
\mathbb{P}(T=2n) = \frac{1}{2n-1}\binom{2n}{n}2^{-2n}.
$$

并证明 $\mathbb{E}\left[T^\alpha\right] < +\infty \iff \alpha < \dfrac{1}{2}$.

证明 考虑 $2n-1$ 时刻的两种情形, 有

$$
\begin{aligned}
&\mathbb{P}(T=2n) \\
=&\frac{1}{2}\mathbb{P}\left((0,0)\to(2n-1,1)\text{且不再过 } x \text{ 轴}\right) + \frac{1}{2}\mathbb{P}\left((0,0)\to(2n-1,-1)\text{且不再过 } x \text{ 轴}\right) \\
=&2\cdot\frac{1}{2}\cdot\frac{1}{2n-1}N_{2n-1}(0,1)
\end{aligned}
$$

$$
\begin{aligned}
&=\frac{\mathrm{C}_{2n-1}^{n}}{2n-1}\cdot\frac{1}{2^{2n-1}}\\
&=\frac{1}{2n-1}\binom{2n}{n}2^{-2n}.
\end{aligned}
$$

因为

$$\mathbb{E}\left[T^{\alpha}\right]=\sum_{n=1}^{\infty}(2n)^{\alpha}\mathbb{P}(T=2n)=\sum_{n=1}^{\infty}\frac{(2n)^{\alpha}}{2n-1}\binom{2n}{n}2^{-2n}=\sum_{n=1}^{\infty}\frac{(2n)^{\alpha}(2n-1)!}{(n!)^{2}2^{2n}},$$

由 Stirling 公式, 当 $n\to\infty$ 时,

$$\frac{(2n)^{\alpha}(2n-1)!}{(n!)^{2}2^{2n}}=\frac{(2n)^{\alpha}\sqrt{(4n-2)\pi}}{2n(2n-1)\pi}\,\mathrm{e}\left(1-\frac{1}{2n}\right)^{2n}\sim\frac{(2n)^{\alpha}\sqrt{(4n-2)\pi}}{2n(2n-1)\pi}=O\left(n^{\alpha-\frac{3}{2}}\right),$$

因此 $\mathbb{E}\left[T^{\alpha}\right]<+\infty\iff\alpha<\dfrac{1}{2}$. □

补充题 6 考虑直线上简单随机游走 S, $S_0=0$, $\mathbb{P}(X_i=1)=p$. 求 $\mathbb{E}[S_n],\mathrm{Var}(S_n),\mathrm{Cov}(S_n,S_m)$.

解 利用期望的线性性, 有

$$\mathbb{E}\left[S_n\right]=\mathbb{E}\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\mathbb{E}\left[X_i\right]=n\left[p-(1-p)\right]=n(2p-1).$$

又

$$
\begin{aligned}
\mathbb{E}\left[S_n^2\right]&=n\mathbb{E}\left[X_1^2\right]+\left(n^2-n\right)\mathbb{E}\left[X_1X_2\right]=n\cdot1+\left(n^2-n\right)\left[\left(2p^2-2p+1\right)-\left(2p-2p^2\right)\right]\\
&=n+\left(n^2-n\right)\left(4p^2-4p+1\right),
\end{aligned}
$$

所以

$$\mathrm{Var}(S_n)=\mathbb{E}\left[S_n^2\right]-\left(\mathbb{E}\left[S_n\right]\right)^2=4np(1-p).$$

不妨设 $n\geqslant m$, 则

$$
\begin{aligned}
\mathbb{E}\left[S_nS_m\right]=\mathbb{E}\left[\left(\sum_{i=1}^{n}X_i\right)\left(\sum_{j=1}^{m}X_j\right)\right]&=m\mathbb{E}\left[X_1^2\right]+(nm-m)\,\mathbb{E}\left[X_1X_2\right]\\
&=m+(nm-m)\left(4p^2-4p+1\right),
\end{aligned}
$$

因此协方差

$$\mathrm{Cov}\left(S_n,S_m\right)=\mathbb{E}\left[S_nS_m\right]-\mathbb{E}\left[S_n\right]\mathbb{E}\left[S_m\right]=4mp(1-p)=4\min\{m,n\}p(1-p).$$

□

习题 (5.1.2) 设随机变量 X ($\geqslant 0$) 有概率母函数 G, 并记 $t(n)=\mathbb{P}(X>n)$ 为 X 的“尾”概率. 证明：

(1) 数列 $\{t(n):n\geqslant 0\}$ 的母函数是 $T(s)=\dfrac{1-G(s)}{1-s}$.

(2) $\mathbb{E}[X]=T(1),\mathrm{Var}(X)=2T'(1)+T(1)-T(1)^2$.

证明 (1) 数列 $\{t(n):n\geqslant 0\}$ 的母函数

$$\begin{aligned}T(s)&=\sum_{n=0}^{\infty}\mathbb{P}\left(X>n\right)s^n=\mathbb{E}\left[\sum_{n=0}^{\infty}I_{\{X>n\}}s^n\right]=\mathbb{E}\left[\sum_{n=0}^{X-1}s^n\right]=\mathbb{E}\left[\frac{1-s^X}{1-s}\right]\\&=\frac{1-\mathbb{E}\left[s^X\right]}{1-s}=\frac{1-G(s)}{1-s}.\end{aligned}$$

(2) 由 L'Hospital 法则,

$$\begin{aligned}T(1)&=\lim_{s\to 1}\frac{1-G(s)}{1-s}=\lim_{s\to 1}\frac{G'(s)}{1}=G'(1)=\mathbb{E}[X],\\T'(1)&=\lim_{s\to 1}\frac{G'(s)(s-1)-G(s)+1}{(s-1)^2}=\lim_{s\to 1}\frac{G''(s)(s-1)}{2(s-1)}=\frac{G''(1)}{2},\end{aligned}$$

因此

$$2T'(1)+T(1)-T(1)^2=G''(1)+\mathbb{E}[X]-\left(\mathbb{E}[X]\right)^2=\mathrm{Var}(X).$$

□

习题 (5.1.7) 证明

$$G(x,y,z,w)=\frac{1}{8}\left(xyzw+xy+yz+zw+wx+yw+xz+1\right)$$

是 4 个两两独立、三三独立但不相互独立的随机变量的联合母函数.

证明 由联合母函数可求得 4 个母函数

$$\begin{aligned}G_X(x)&=G(x,1,1,1)=\frac{x+1}{2},\quad G_Y(y)=G(1,y,1,1)=\frac{y+1}{2},\\G_Z(z)&=G(1,1,z,1)=\frac{z+1}{2},\quad G_W(w)=G(1,1,1,w)=\frac{w+1}{2}.\end{aligned}$$

不失一般性地, 从

$$\begin{aligned}G_{X,Y}(x,y)&=G(x,y,1,1)=\frac{xy+x+y+1}{4}=G_X(x)G_Y(y),\\G_{X,Y,Z}(x,y,z)&=G(x,y,z,1)=\frac{xyz+xy+yz+zx+x+y+z+1}{8}=G_X(x)G_Y(y)G_Z(z)\end{aligned}$$

可知随机变量 X, Y, Z, W 两两独立、三三独立. 但

$$G_X(x)G_Y(y)G_Z(z)G_W(w) \neq G(x, y, z, w),$$

故 X, Y, Z, W 不相互独立. □

习题 (5.1.9) 设 G_1, G_2 是概率母函数, $0 \leqslant \alpha \leqslant 1$. 证明 G_1G_2 和 $\alpha G_1 + (1-\alpha)G_2$ 也是概率母函数. 问：$\dfrac{G(\alpha s)}{G(\alpha)}$ 一定是概率母函数吗？

证明 三者皆为概率母函数, 因为它们的表达式中 s 的所有幂次前系数之和均非负且求和为 1(非负性显然, 而求和为 1 可以通过取 $s = 1$ 得出). □

习题 (4.1.1) 当参数 C 取何值时下列函数是概率密度函数？

(a) (反正弦律的密度函数)$f(x) = \dfrac{C}{\sqrt{x(1-x)}}, 0 < x < 1$.

(b) (极值分布的密度函数)$f(x) = C\exp\left(-x - \mathrm{e}^{-x}\right), x \in \mathbb{R}$.

(c) $f(x) = \dfrac{C}{(1+x^2)^m}, x \in \mathbb{R}$.

解 (a) 因为

$$\int_0^1 \frac{1}{\sqrt{x(1-x)}}\,\mathrm{d}x \xlongequal[\theta\in\left(0,\frac{\pi}{2}\right)]{x=\sin^2\theta} 2\int_0^{\frac{\pi}{2}}\,\mathrm{d}\theta = \pi,$$

所以 $C = \dfrac{1}{\pi}$.

(b) 因为

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-x-\mathrm{e}^{-x}}\,\mathrm{d}x = \mathrm{e}^{-\mathrm{e}^{-x}}\Big|_{-\infty}^{+\infty} = 1,$$

所以 $C = 1$.

(c) 因为

$$\int_{-\infty}^{+\infty} \frac{1}{(1+x^2)^m}\,\mathrm{d}x \xlongequal[\theta\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)]{x=\tan\theta} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{2m-2}\theta\,\mathrm{d}\theta = 2\int_0^{\frac{\pi}{2}} \cos^{2m-2}\theta\,\mathrm{d}\theta = \frac{(2m-3)!!}{(2m-2)!!}\pi,$$

所以 $C = \dfrac{(2m-2)!!}{(2m-3)!!\pi}$. □

习题 (4.1.2) 设随机变量 $Y = aX$, 其中 $a > 0$. 用 X 的密度函数表示 Y 的密度函数. 证明：连续性随机变量 X 与 $-X$ 有相同的分布函数当且仅当 $f_X(x) = f_X(-x), \forall x \in \mathbb{R}$.

证明 由

$$F_Y(y)=\mathbb{P}(Y\leqslant y)=\mathbb{P}(aX\leqslant y)=\mathbb{P}\left(X\leqslant\frac{y}{a}\right)=F_X\left(\frac{y}{a}\right)$$

对 y 求导即得 $f_Y(y)=\dfrac{1}{a}f_X\left(\dfrac{y}{a}\right)$.

由

$$F_{-X}(x)=\mathbb{P}\left(-X\leqslant x\right)=\mathbb{P}\left(X\geqslant -x\right)=1-\mathbb{P}(X\leqslant -x)=1-F_X(-x)$$

求导即得 $f_{-X}(x)=f_X(-x)$. 因此若 X 与 $-X$ 有相同的分布函数, 则 $f_X(x)=f_X(-x)$. 反过来, 若 $f_X(x)=f_X(-x)$, 则

$$\begin{aligned}F_{-X}(x)&=1-F_X(-x)=1-\int_{-\infty}^{-x}f_X(t)\,\mathrm{d}t=1-\int_{-\infty}^{-x}f_X(-t)\,\mathrm{d}t\\&\xlongequal{u=-t}1-\int_x^{\infty}f_X(u)\,\mathrm{d}u=\int_{-\infty}^{x}f_X(u)\,\mathrm{d}u=F_X(x).\end{aligned}$$

□

习题 (4.2.2) 设随机变量 X,Y 独立且有相同的分布函数 F 与密度函数 f. 证明：$V=\max\{X,Y\}$ 有分布函数 $\mathbb{P}(V\leqslant x)=F(x)^2$ 与密度函数 $f_V(x)=2f(x)F(x),x\in\mathbb{R}$. 求 $U=\min\{X,Y\}$ 的密度函数.

解 我们有

$$\mathbb{P}(V\leqslant x)=\mathbb{P}(X\leqslant x,Y\leqslant x)=\mathbb{P}(X\leqslant x)\mathbb{P}(Y\leqslant x)=F(x)^2.$$

对 x 求导即得

$$f(x)=2F(x)F'(x)=2f(x)F(x).$$

类似地, 有

$$\mathbb{P}(U\leqslant x)=1-\mathbb{P}(U>x)=1-\mathbb{P}(X>x,Y>x)=1-\mathbb{P}(X>x)\mathbb{P}(Y>x)=1-\left[1-F(x)\right]^2.$$

对 x 求导即得

$$f_U(x)=2\left[1-F(x)\right]F'(x)=2\left[1-F(x)\right]f(x).$$

□

习题 (4.2.4) 设随机变量 $\{X_r\mid r\geqslant 1\}$ 相互独立且同分布, 该分布函数 F 满足 $F(y)<1,\forall y$. 设 $Y(y)=\min\{k\mid X_k>y\}$. 证明：

$$\lim_{y\to+\infty}\mathbb{P}\left(Y(y)\leqslant\mathbb{E}\left[Y(y)\right]\right)=1-\frac{1}{\mathrm{e}}.$$

证明 因为

$$\mathbb{P}(Y(y)>k)=\mathbb{P}\left(X_i\leqslant y, i=1,\cdots,k\right)=\prod_{i=1}^{k}\mathbb{P}(X_i\leqslant y)=F(y)^k,$$

所以

$$\mathbb{E}\left[Y(y)\right]=\sum_{k=1}^{\infty}k\left[F(y)^{k-1}-F(y)^k\right]=\frac{1}{1-F(y)}.$$

于是

$$\begin{aligned}\mathbb{P}\left(Y(y)\leqslant\mathbb{E}\left[Y(y)\right]\right)&=1-\mathbb{P}\left(Y(y)>\mathbb{E}\left[Y(y)\right]\right)=1-(F(y))^{\lfloor\mathbb{E}[Y(y)]\rfloor}\\&=1-F(y)^{\left\lfloor\frac{1}{1-F(y)}\right\rfloor},\end{aligned}$$

而当 $y\to+\infty$ 时, $F(y)\to1^-$, 故

$$\lim_{y\to+\infty}\mathbb{P}\left(Y(y)\leqslant\mathbb{E}\left[Y(y)\right]\right)=\lim_{x\to1^-}1-x^{\frac{1}{1-x}}\xlongequal{t=\frac{1}{1-x}}\lim_{t\to+\infty}1-\left(1-\frac{1}{t}\right)^t=1-\frac{1}{\mathrm{e}}.$$

□

习题 (4.4.6) 设 $X\sim N\left(\mu,\sigma^2\right)$. 证明：$\mathbb{E}\left[(X-\mu)g(X)\right]=\sigma^2\mathbb{E}\left[g'(X)\right]$(设等号两边均有意义).

证明 我们有

$$\begin{aligned}\mathrm{LHS}&=\int_{\mathbb{R}}(x-\mu)g(x)\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x\\&\xlongequal{\text{分部积分}}\sigma^2\int_{\mathbb{R}}g'(x)\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{1}{2\sigma}(x-\mu)^2}\,\mathrm{d}x-\frac{\sigma^2}{\sqrt{2\pi\sigma^2}}g(x)\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}\bigg|_{-\infty}^{+\infty}\\&=\sigma^2\mathbb{E}\left[g'(X)\right]-0=\mathrm{RHS}.\end{aligned}$$

□

习题 (4.5.4) 设随机变量 X 与 Y 相互独立且服从 $[0,1]$ 上的均匀分布. 记 $U=\min\{X,Y\},V=\max\{X,Y\}$. 求 $\mathbb{E}[U]$ 与 $\mathrm{Cov}(U,V)$.

解 由 $F_U(u)=1-\mathbb{P}\left(X>u,Y>u\right)=1-(1-u)(1-u)=2u-u^2$ $(u\in[0,1])$ 可得 $f_U(u)=\begin{cases}2-2u, & u\in[0,1],\\ 0, & u\notin[0,1].\end{cases}$, 从而 $\mathbb{E}[U]=\int_{\mathbb{R}}uf_U(u)\,\mathrm{d}u=\int_0^1u(2-2u)\,\mathrm{d}u=\dfrac{1}{3}$. 因此 $\mathbb{E}[V]=\mathbb{E}[X+Y-U]=2\mathbb{E}[X]-\mathbb{E}[U]=2\cdot\dfrac{1}{2}-\dfrac{1}{3}=\dfrac{2}{3}$.

由 $U=\dfrac{X+Y}{2}-\dfrac{|X-Y|}{2}$ 及 $V=\dfrac{X+Y}{2}+\dfrac{|X-Y|}{2}$ 可知 $UV=\dfrac{(X+Y)^2}{4}-\dfrac{(X-Y)^2}{4}=XY$. 由 X,Y 独立得 $\mathbb{E}[UV]=\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]=\dfrac{1}{4}$. 于是

$$\operatorname{Cov}(U,V)=\mathbb{E}[UV]-\mathbb{E}[U]\mathbb{E}[V]=\frac{1}{4}-\frac{1}{3}\left(1-\frac{1}{3}\right)=\frac{1}{36}.$$

□

习题 (4.5.7) 设随机变量列 $\{X_r \mid 1\leqslant r\leqslant n\}$ 独立同分布且方差存在, 定义 $\overline{X}=\dfrac{1}{n}\sum_{r=1}^{n}X_r$. 证明：$\operatorname{Cov}\left(\overline{X},X_r-\overline{X}\right)=0$.

证明 $\operatorname{Cov}\left(\overline{X},X_r-\overline{X}\right)=\mathbb{E}\left[\overline{X}\right]\mathbb{E}\left[X_r-\overline{X}\right]-\mathbb{E}\left[\overline{X}\left(X_r-\overline{X}\right)\right]$, 而由 $\{X_r\}$ 独立同分布可知 $\mathbb{E}\left[X_r-\overline{X}\right]=0$, 又

$$\begin{aligned}
\mathbb{E}\left[\overline{X}\left(X_r-\overline{X}\right)\right]&=\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}X_i\left(X_r-\frac{1}{n}\sum_{j=1}^{n}X_j\right)\right]\\
&=\frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}\left[X_iX_r\right]-\frac{1}{n}\sum_{i,j=1}^{n}\mathbb{E}\left[X_iX_j\right]\right]\\
&=\frac{1}{n}\left[(n-1)\mathbb{E}[X_1X_2]+\mathbb{E}\left[X_1^2\right]-\frac{1}{n}\left[n\mathbb{E}\left[x_1^2\right]+\left(n^2-n\right)\mathbb{E}[X_1X_2]\right]\right]\\
&=\frac{1}{n}\left[(n-1)\left(\mathbb{E}[X_1]\right)^2+\mathbb{E}\left[X_1^2\right]-\mathbb{E}\left[X_1^2\right]-(n-1)\mathbb{E}[X_1]\mathbb{E}[X_2]\right]\\
&=\frac{n-1}{n}\left[\left(\mathbb{E}[X_1]\right)^2-\mathbb{E}[X_1]\mathbb{E}[X_2]\right]\\
&=0,
\end{aligned}$$

因此 $\operatorname{Cov}\left(\overline{X},X_r-\overline{X}\right)=0$. □

习题 (4.6.6) 设 $\{X_r \mid r\geqslant 1\}$ 独立且服从 $[0,1]$ 上的均匀分布. 设 $0<x<1$ 并定义

$$N=\min\left\{n\geqslant 1 \mid X_1+X_2+\cdots+X_n>x\right\}.$$

证明：$\mathbb{P}(N>n)=\dfrac{x^n}{n!}$. 再求 $\mathbb{E}[N]$ 与 $\operatorname{Var}(N)$.

解 由 N 定义知, 对 $x\in(0,1)$, 有

$$\mathbb{P}(N>n)=\mathbb{P}\left(X_1+\cdots+X_n\leqslant x\right)$$

$$
\begin{aligned}
&= \int\limits_{\substack{x_1+\cdots+x_n\leqslant x,\\(x_1,\cdots,x_n)\in[0,1]^n}} 1\,\mathrm{d}x_1\cdots\mathrm{d}x_n\\
&= x^n\cdot n\text{ 维单形测度}\\
&= \frac{x^n}{n!}.
\end{aligned}
$$

因此

$$\mathbb{E}[N]=\sum_{n=0}^{\infty}\mathbb{P}(N>n)=\sum_{n=0}^{\infty}\frac{x^n}{n!}=\mathrm{e}^x\,.$$

N 的概率母函数

$$G_N(s)=\sum_{n=1}^{\infty}\left(\frac{x^{n-1}}{(n-1)!}-\frac{x^n}{n!}\right)s^n=s\sum_{n=1}^{\infty}\frac{(sx)^{n-1}}{(n-1)!}-\sum_{n=1}^{\infty}\frac{(sx)^n}{n!}=s\,\mathrm{e}^{xs}-(\mathrm{e}^{xs}-1)\,.$$

因此

$$\operatorname{Var}(N)=G_N''(1)+G_N'(1)\left[1-G_N'(1)\right]=2x\,\mathrm{e}^x+\mathrm{e}^x\left(1-\mathrm{e}^x\right)=2x\,\mathrm{e}^x+\mathrm{e}^x-\mathrm{e}^{2x}\,.$$

□

习题 (4.8.6) 对独立同分布随机变量 X 和 Y, 证明：

(1) $U=X+Y$ 与 $V=X-Y$ 不相关但未必独立.

(2) 若 $X,Y\sim N(0,1)$, 则 U 与 V 独立.

证明 (1) 由 $\mathbb{E}[U]=\mathbb{E}[X+Y]=2\mathbb{E}[X],\mathbb{E}[V]=\mathbb{E}[X-Y]=0,\mathbb{E}[UV]=\mathbb{E}\left[X^2-Y^2\right]=0$ 即得 $\operatorname{Cov}(U,V)=\mathbb{E}[UV]-\mathbb{E}[U]\mathbb{E}[V]=0$, U 与 V 不相关.

U 与 V 不独立的例子：设 $\mathbb{P}(X=-1)=\mathbb{P}(X=1)=\mathbb{P}(Y=-1)=\mathbb{P}(Y=1)=\dfrac{1}{2}$, 则 $\mathbb{P}(U=2)=\mathbb{P}(V=2)=\dfrac{1}{4}$, 但 $\mathbb{P}(U=2,V=2)=0\neq\mathbb{P}(U=2)\mathbb{P}(V=2)$.

(2) 若 $X,Y\sim N(0,1)$, 则

$$
\begin{aligned}
f_U(u)=\mathbb{P}\left(X+Y=u\right)&=\int_{\mathbb{R}}f_X(t)f_Y(u-t)\,\mathrm{d}t\\
&=\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\frac{t^2+(u-t)^2}{2}}\,\mathrm{d}t\\
&=\frac{\mathrm{e}^{-\frac{u^2}{4}}}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\left(t-\frac{u}{2}\right)^2}\,\mathrm{d}\left(t-\frac{u}{2}\right)\\
&=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{u^2}{4}},
\end{aligned}
$$

$$
\begin{aligned}
f_V(v) = \mathbb{P}(X-Y=v) &= \int_{\mathbb{R}} f_X(t+v) f_Y(t)\,\mathrm{d}t \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\frac{(t+v)^2+t^2}{2}}\ \mathrm{d}t \\
&= \frac{\mathrm{e}^{-\frac{v^2}{4}}}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\left(t-\frac{v}{2}\right)^2}\ \mathrm{d}\left(t-\frac{v}{2}\right) \\
&= \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{v^2}{4}},
\end{aligned}
$$

$$
\begin{aligned}
f_{U,V}(u,v) = \mathbb{P}(X+Y=u, X-Y=v) &= \mathbb{P}\left(X=\frac{u+v}{2}, Y=\frac{u-v}{2}\right) \\
&= \mathbb{P}\left(X=\frac{u+v}{2}\right)\mathbb{P}\left(Y=\frac{u-v}{2}\right) \\
&= \frac{1}{2\pi}\,\mathrm{e}^{-\frac{\left(\frac{u+v}{2}\right)^2+\left(\frac{u-v}{2}\right)^2}{2}} = \frac{1}{2\pi}\,\mathrm{e}^{-\frac{u^2+v^2}{4}} \\
&= f_U(u) f_V(v).
\end{aligned}
$$

由 $f_{U,V}(u,v)=f_U(u)f_V(v)$, U 与 V 独立. □

习题 (4.9.4) 设 X 与 Y 服从二元正态分布, 且它们的期望均为 0、方差均为 1、相关系数为 ρ. 求 $X+Y$ 与 $X-Y$ 的联合密度函数及边缘密度函数.

解 设 $U=X+Y, V=X-Y$, 作变量替换 $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$, 则

$$
\begin{aligned}
f_{U,V}(u,v) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{(1-\rho)u^2+(1+\rho)v^2}{4(1-\rho^2)}}\cdot\frac{1}{2} \\
&= \frac{1}{4\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{u^2}{4(1+\rho)}-\frac{v^2}{4(1-\rho)}}.
\end{aligned}
$$

从而可求边缘密度函数

$$
\begin{aligned}
f_U(u) &= \frac{1}{4\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{u^2}{4(1+\rho)}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{v^2}{4(1-\rho)}}\ \mathrm{d}v \\
&= \frac{\sqrt{2(1-\rho)}}{4\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{u^2}{4(1+\rho)}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\left(\frac{v}{\sqrt{2(1-\rho)}}\right)^2}\ \mathrm{d}\left(\frac{v}{\sqrt{2(1-\rho)}}\right) \\
&= \frac{1}{\sqrt{2\pi\cdot 2(1+\rho)}}\,\mathrm{e}^{-\frac{1}{2\left(\sqrt{2(1-\rho)}\right)^2}u^2},
\end{aligned}
$$

$$
f_V(v) = \frac{1}{4\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{v^2}{4(1-\rho)}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{u^2}{4(1+\rho)}}\ \mathrm{d}u
$$

$$
\begin{aligned}
&= \frac{\sqrt{2(1+\rho)}}{4\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{v^2}{4(1-\rho)}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{1}{2}\left(\frac{u}{\sqrt{2(1+\rho)}}\right)^2}\,\mathrm{d}\left(\frac{u}{\sqrt{2(1+\rho)}}\right)\\
&= \frac{1}{\sqrt{2\pi\cdot 2(1-\rho)}}\,\mathrm{e}^{-\frac{1}{2\left(\sqrt{2(1+\rho)}\right)^2}v^2}.
\end{aligned}
$$

□

习题 (4.14.37) 设 ϕ 是 $N(0,1)$ 的密度函数, 定义函数列 $\{H_n\}_{n\geqslant 0}$, $H_0=1$, $(-1)^nH_n\phi=\phi^{(n)}$. 证明:

(a) $H_n(x)$ 是 n 次首一多项式, 且

$$
\int_{\mathbb{R}} H_m(x)H_n(x)\phi(x)\,\mathrm{d}x = \begin{cases} 0, & m\neq n,\\ n!, & m=n. \end{cases}
$$

(b) $\displaystyle\sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!} = \mathrm{e}^{tx-\frac{1}{2}t^2}$.

证明 (a) 由 $\phi(x)=\dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2}$ 知 $\phi'(x)=-x\phi(x)$. 又 $\phi'(x)=-H_1(x)\phi(x)$, 故 $H_1(x)=x$. 对 $(-1)^nH_n(x)\phi(x)=\phi^{(n)}(x)$ 两边求导得

$$
(-1)^nH_n'(x)\phi(x)+(-1)^nH_n(x)\phi'(x)=\phi^{(n+1)}(x)=(-1)^{n+1}H_{n+1}(x)\phi(x),
$$

化简得

$$
H_{n+1}(x)=xH_n(x)-H_n'(x).
$$

由此递推式及 $H_0=1, H_1=x$ 归纳即得 $H_n(x)$ 是 n 次首一多项式. 利用 $H_m(x)\phi^{(n)}(x)\Big|_{-\infty}^{+\infty}=0$, $\forall m,n\in\mathbb{N}$, 由分部积分, 对 $m\leqslant n$, 有

$$
\begin{aligned}
\int_{\mathbb{R}} H_m(x)H_n(x)\phi(x)\,\mathrm{d}x &= (-1)^n\int_{\mathbb{R}} H_m(x)\phi^{(n)}(x)\,\mathrm{d}x\\
&= (-1)^{n+1}\int_{\mathbb{R}} H_m'(x)\phi^{(n-1)}(x)\,\mathrm{d}x\\
&= \cdots\\
&= (-1)^{n+m}\int_{\mathbb{R}} H_m^{(m)}(x)\phi^{(n-m)}(x)\,\mathrm{d}x\\
&= (-1)^{n+m}m!\int_{\mathbb{R}} \phi^{(n-m)}(x)\,\mathrm{d}x
\end{aligned}
$$

$$=\begin{cases}(-1)^{n+m}m!\phi^{(n-m-1)}(x)\Big|_{-\infty}^{+\infty}=0, & m\neq n,\\ n!, & m=n.\end{cases}$$

(b) 直接计算得

$$\begin{aligned}\sum_{n=0}^{\infty}H_n(x)\frac{t^n}{n!}&=\frac{1}{\phi(x)}\sum_{n=0}^{\infty}(-1)^n\phi^{(n)}(x)\frac{t^n}{n!}=\frac{1}{\phi(x)}\sum_{n=0}^{\infty}\frac{\phi^{(n)}(x)}{n!}(-t)^n\\&=\frac{\phi(x-t)}{\phi(x)}=\mathrm{e}^{tx-\frac{1}{2}t^2}.\end{aligned}$$

□

补充题 7 设 X 与 Y 为随机变量, 称 $Z=X+\mathrm{i}Y$ 为复随机变量. 称 Z 服从复 Gauss 分布 (或复正态分布), 若它有密度函数

$$f(z)=\frac{1}{\pi\sigma^2}\mathrm{e}^{-\frac{1}{\sigma^2}|z-\mu|^2},\quad z\in\mathbb{C},$$

其中 $\mu\in\mathbb{C},\sigma^2>0$. 记为 $Z\sim N_{\mathbb{C}}\left(\mu,\sigma^2\right)$. 定义其期望为 $\mathbb{E}[Z]\coloneqq\mathbb{E}[X]+\mathrm{i}\,\mathbb{E}[Y]$. 证明：若 $Z\sim N_{\mathbb{C}}(0,1)$, 则

$$\mathbb{E}\left(Z^k\overline{Z}^l\right)=\begin{cases}k!, & k=l,\\ 0, & k\neq l.\end{cases}$$

证明 若 $Z\sim N_{\mathbb{C}}(0,1)$, 则 $f_Z(z)=\dfrac{1}{\pi}\mathrm{e}^{-|z|^2}$, 用实部、虚部改写即联合密度函数 $f_{X,Y}(x,y)=\dfrac{1}{\pi}\mathrm{e}^{-\left(x^2+y^2\right)}$, 再作极坐标换元 $\begin{cases}X=R\cos\Theta,\\ Y=R\sin\Theta,\end{cases}$ 得到 $f_{R,\Theta}(r,\theta)=\dfrac{1}{\pi}r\,\mathrm{e}^{-r^2}$. 由

$$Z^k\overline{Z}^l=R^{k+l}\left[\cos(k\Theta)+\mathrm{i}\sin(k\Theta)\right]\left[\cos(l\Theta)-\mathrm{i}\sin(l\Theta)\right]=R^{k+l}\left[\cos(k-l)\Theta+\mathrm{i}\sin(k-l)\Theta\right]$$

得

$$\begin{aligned}\mathbb{E}\left[Z^k\overline{Z}^l\right]=&\mathbb{E}\left[R^{k+l}\cos(k-l)\Theta\right]+\mathrm{i}\,\mathbb{E}\left[R^{k+l}\sin(k-l)\Theta\right]\\=&\frac{1}{\pi}\int_0^{+\infty}r^{k+l+1}\mathrm{e}^{-r^2}\,\mathrm{d}r\int_0^{2\pi}\cos(k-l)\theta\,\mathrm{d}\theta\\&+\mathrm{i}\,\frac{1}{\pi}\int_0^{+\infty}r^{k+l+1}\mathrm{e}^{-r^2}\,\mathrm{d}r\int_0^{2\pi}\sin(k-l)\theta\,\mathrm{d}\theta\\=&\frac{1}{\pi}\int_0^{+\infty}r^{k+l+1}\mathrm{e}^{-r^2}\,\mathrm{d}r\int_0^{2\pi}\cos(k-l)\theta\,\mathrm{d}\theta.\end{aligned}$$

若 $k \neq l$, 则由 $\int_0^{2\pi} \cos(k-l)\theta\,\mathrm{d}\theta = 0$ 得上述积分为 0. 若 $k=l$, 上述积分即

$$2\int_0^{+\infty} r^{2k+1}\,\mathrm{e}^{-r^2}\,\mathrm{d}r = \int_0^{+\infty} \left(r^2\right)^k \mathrm{e}^{-r^2}\,\mathrm{d}r^2 = \Gamma(k+1) = k!.$$

□

补充题 8 设 $\phi_n(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}}\,\mathrm{e}^{-\frac{1}{2}\sum\limits_{j=1}^{n} x_j^2}, \Delta_n(\mathbf{x}) = \prod\limits_{1\leqslant i<j\leqslant n}(x_j - x_i)$. 证明：对任意 $m \in \mathbb{N}$, 有

$$\int_{\mathbb{R}^n} \left|\Delta_n(\mathbf{x})\right|^{2m} \phi_n(\mathbf{x})\,\mathrm{d}\mathbf{x} = \prod_{j=1}^{n} \frac{\Gamma(1+jm)}{\Gamma(1+m)}.$$

证明 这是 Selberg 积分公式的推论, 见 An Introduction to Random Matrices, Greg W. Anderson, Alice Guionnet, Ofer Zeitouni, Cambridge University Press, 2009: 59. □

注 2.0.1 事实上, 此结论对任意 $m \in \mathbb{R}_{\geqslant 0}$ 均成立.

习题 (5.6.1) (a) (Jensen 不等式) 称函数 $u:\mathbb{R}\to\mathbb{R}$ 为凸的, 若对任意 $a\in\mathbb{R}$, 存在 $\lambda=\lambda(a)$, 使得 $u(x) \geqslant u(a) + \lambda(x-a), \forall x$. 称凸函数 u 是严格凸的, 若 $\lambda(a)$ 关于 a 严格单调递增.

① 证明：对于凸函数 u 与期望存在的随机变量 X, 有 $\mathbb{E}[u(X)] \geqslant u(\mathbb{E}[X])$.

② 证明：若 u 是严格凸的且 $\mathbb{E}[u(X)] = u(\mathbb{E}[X])$, 则 X 是常值以概率 1 成立.

(b) 对于概率密度函数 f, 定义它的熵函数为 $H(f) = -\int_{\mathbb{R}} f(x)\ln f(x)\,\mathrm{d}x$, 它的支撑集为 $S(f)=\{x\in\mathbb{R} \mid f(x)>0\}$. 证明：在支撑集为 $\mathbb{R}$、期望为 μ、方差 $\sigma^2>0$ 的概率密度函数中, 只有正态分布 $N\left(\mu,\sigma^2\right)$ 的概率密度函数熵是最大的.

证明 (a) ① 取 $a=\mathbb{E}[X]$ 代入题干式中可得存在 λ 使得

$$u(X) \geqslant u\left(\mathbb{E}[X]\right) + \lambda\left(X - \mathbb{E}[X]\right),$$

对上式两边取期望, 由期望的线性性与非负性即得

$$\mathbb{E}[u(X)] \geqslant u\left(\mathbb{E}[X]\right) + \lambda\left(\mathbb{E}[X] - \mathbb{E}[X]\right) = u\left(\mathbb{E}[X]\right).$$

② 由①可知

$$u(X) \xlongequal{\text{a.s.}} u\left(\mathbb{E}[X]\right) + \lambda\left(X - \mathbb{E}[X]\right).$$

若 $\mathbb{P}(X\text{为常值}) \neq 1$, 则由上式可见 u 必在某个局部上为一次函数, 这与严格凸性矛盾.

(b) 设 f 为 $N\left(\mu,\sigma^2\right)$ 的密度函数, g 为任一满足题意的密度函数, 则由已知,

$$\int_{\mathbb{R}} xf(x)\,\mathrm{d}x = \int_{\mathbb{R}} xg(x)\,\mathrm{d}x = \mu,$$
$$\int_{\mathbb{R}} (x-\mu)^2 f(x)\,\mathrm{d}x = \int_{\mathbb{R}} (x-\mu)^2 g(x)\,\mathrm{d}x = \sigma^2.$$

于是

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\ln f(x)\,\mathrm{d}x &= \int_{\mathbb{R}} f(x)\left[-\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{(x-\mu)^2}{2\sigma^2}\right]\,\mathrm{d}x \\
&= -\frac{1}{2}\ln\left(2\pi\sigma^2\right)\int_{\mathbb{R}} f(x)\,\mathrm{d}x - \frac{1}{2\sigma^2}\int_{\mathbb{R}} (x-\mu)^2 f(x)\,\mathrm{d}x \\
&= -\frac{1}{2}\ln\left(2\pi\sigma^2\right)\int_{\mathbb{R}} g(x)\,\mathrm{d}x - \frac{1}{2\sigma^2}\int_{\mathbb{R}} (x-\mu)^2 g(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} g(x)\ln f(x)\,\mathrm{d}x.
\end{aligned}$$

由 $\ln x$ 是下凸函数, 利用凸函数的 Jensen 不等式, 我们有

$$\begin{aligned}
H(g) - H(f) &= -\int_{\mathbb{R}} g(x)\ln g(x)\,\mathrm{d}x + \int_{\mathbb{R}} f(x)\ln f(x)\,\mathrm{d}x \\
&= -\int_{\mathbb{R}} g(x)\ln g(x)\,\mathrm{d}x + \int_{\mathbb{R}} g(x)\ln f(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} g(x)\ln\frac{f(x)}{g(x)}\,\mathrm{d}x \\
&\leqslant \ln\left(\int_{\mathbb{R}} g(x)\cdot\frac{f(x)}{g(x)}\,\mathrm{d}x\right) \\
&= \ln 1 = 0.
\end{aligned}$$

□

注 2.0.2 以下利用 Lagrange 乘子法给出 (b) 的一个不完备的证明.

设 $p(x)$ 是一个期望为 μ、方差为 $\sigma^2>0$ 的随机变量的概率密度函数, 考虑

$$\begin{aligned}
F(p,\lambda_1,\lambda_2,\lambda_3) &= -\int_{\mathbb{R}} p(x)\ln p(x)\,\mathrm{d}x + \lambda_1\left(\int_{\mathbb{R}} p(x)\,\mathrm{d}x - 1\right) + \lambda_2\left(\int_{\mathbb{R}} xp(x)\,\mathrm{d}x - \mu\right) \\
&\quad + \lambda_3\left(\int_{\mathbb{R}} (x-\mu)^2 p(x)\,\mathrm{d}x - \sigma^2\right) \\
&= \int_{\mathbb{R}} \left[-p(x)\ln p(x) + \lambda_1 p(x) + \lambda_2 x p(x) + \lambda_3 (x-\mu)^2 p(x)\right]\,\mathrm{d}x \\
&\quad - \lambda_1 - \mu\lambda_2 - \sigma^2\lambda_3 \\
&= \int_{\mathbb{R}} \mathcal{L}(x,p(x),\lambda_1,\lambda_2,\lambda_3)\,\mathrm{d}x - \lambda_1 - \mu\lambda_2 - \sigma^2\lambda_3,
\end{aligned}$$

其中

$$\mathcal{L}(x,p,\lambda_1,\lambda_2,\lambda_3) = -p\ln p + \lambda_1 p + \lambda_2 xp + \lambda_3 (x-\mu)^2 p.$$

将 p 视为未定元, 由

$$\frac{\partial \mathcal{L}}{\partial p} = -1 - \ln p + \lambda_1 + \lambda_2 x + \lambda_3 (x-\mu)^2 = 0$$

可得

$$p(x) = \mathrm{e}^{\lambda_1 - 1 + \lambda_2 x + \lambda_3 (x-\mu)^2}.$$

由 $\left|\int_{\mathbb{R}} p(x)\,\mathrm{d}x\right| < +\infty$ 可知 $\lambda_2 = 0, \lambda_3 < 0$. 因此可设 $p(x) = a\,\mathrm{e}^{-b(x-\mu)^2}$, 其中 a,b 为待定常数且 $b>0$. 再由 $\int_{\mathbb{R}} p(x)\,\mathrm{d}x = 1$ 及 $\int_{\mathbb{R}} (x-\mu)^2 p(x)\,\mathrm{d}x = \sigma^2$ 即可确定 $a = \frac{1}{\sqrt{2\pi\sigma^2}}, b = \frac{1}{2\sigma^2}$.

习题 (5.6.4) 设 $\mathbb{E}[|X^r|] < +\infty$, 其中 $r>0$. 证明: $\lim\limits_{x\to+\infty} x^r \mathbb{P}(|X| \geqslant x) = 0$. 反过来, 设 $\lim\limits_{x\to+\infty} x^r \mathbb{P}(|X| \geqslant x) = 0$, 其中 $r>0$, 证明: $\mathbb{E}[|X^s|] < +\infty, \forall 0 \leqslant s < r$.

证明 由 $\mathbb{E}[|X^r|] < +\infty$, 对任意 $x>0$,

$$x^r \mathbb{P}(|X| \geqslant x) \leqslant \int_x^{+\infty} u^r \,\mathrm{d}F_{|X|}(u) \to 0, \quad x \to +\infty.$$

若 $\lim\limits_{x\to+\infty} x^r \mathbb{P}(|X| \geqslant x) = 0$, 则对任意 $s \in [0,r)$, 由分部积分,

$$\begin{aligned}
\int_0^x u^s \,\mathrm{d}F_{|X|}(u) &= -\int_0^x u^s \,\mathrm{d}\left(1 - F_{|X|}(u)\right) \\
&= \left[-u^s \left(1 - F_{|X|}(u)\right)\right]\Big|_0^x + \int_0^x s u^{s-1} \left(1 - F_{|X|}(u)\right) \,\mathrm{d}u \\
&= \left[-u^s \left(1 - F_{|X|}(u)\right)\right]\Big|_0^x + \int_0^x s u^{s-1} \mathbb{P}(|X| > u) \,\mathrm{d}u \\
&\leqslant s \int_0^x u^{s-1} \mathbb{P}(|X| > u) \,\mathrm{d}u,
\end{aligned}$$

对充分大的 u 有 $u^{s-1}\mathbb{P}(|X|>u) \leqslant u^{s-1}u^{-r} = u^{s-r-1}$, 而 $s-r-1<-1$, 因此 $\int_0^x u^s \,\mathrm{d}F_{|X|}(u)$ 在 $[0,+\infty)$ 上有界, 从而 $\mathbb{E}[|X^s|] = \lim\limits_{x\to+\infty} \int_0^x u^s \,\mathrm{d}F_{|X|}(u) < +\infty$. □

习题 (7.1.1) 设 $r \geqslant 1$, 定义 $\|X\|_r = (\mathbb{E}[|X|^r])^{\frac{1}{r}}$. 证明:

(a) $\|cX\|_r = |c| \cdot \|X\|_r, \forall c \in \mathbb{R}$.

(b) $\|X+Y\|_r \leqslant \|X\|_r + \|Y\|_r$.

(c) $\|X\|_r = 0$ 当且仅当 $\mathbb{P}(X=0)=1$.

也就是说, $\|\cdot\|$ 是给定概率空间上具有 r 阶矩的随机变量等价类上的一个范数, 其中等价关系为 $X \sim Y \iff \mathbb{P}(X=Y)=1$.

证明 (a) $\|cX\|_r = (\mathbb{E}[|cX|^r])^{\frac{1}{r}} = (|c|^r \mathbb{E}[|X|^r])^{\frac{1}{r}} = |c| \cdot \|X\|_r$.

(b) 设 r, s 为一对共轭参数, 即 $\dfrac{1}{r}+\dfrac{1}{s}=1$. 由 Hölder 不等式,

$$\begin{aligned}
\|X+Y\|_r^r = \mathbb{E}[|X+Y|^r] &= \mathbb{E}\left[|X+Y|^{r-1}|X+Y|\right] \\
&\leqslant \mathbb{E}\left[|X+Y|^{r-1}(|X|+|Y|)\right] = \||X+Y|^{r-1}|X|\|_1 + \||X+Y|^{r-1}|Y|\|_1 \\
&\overset{\text{Hölder}}{\leqslant} \||X+Y|^{r-1}\|_s (\|X\|_r + \|Y\|_r) = \left(\mathbb{E}\left[|X+Y|^{(r-1)s}\right]\right)^{\frac{1}{s}} (\|X\|_r + \|Y\|_r) \\
&\xlongequal{(r-1)s=r} (\|X+Y\|_r)^{\frac{r}{s}} (\|X\|_r + \|Y\|_r) \\
&\xlongequal{\frac{r}{s}=r-1} \|X+Y\|_r^{r-1} (\|X\|_r + \|Y\|_r),
\end{aligned}$$

于是

$$\|X+Y\|_r \leqslant \|X\|_r + \|Y\|_r.$$

(c) 只需证若 $X \geqslant 0$ 且 $\mathbb{E}[X^r]=0$, 则 $\mathbb{P}(X=0)=1$. 证明如下：由 Markov 不等式,

$$\mathbb{P}(X>\varepsilon) = \mathbb{P}(X^r > \varepsilon^r) \leqslant \frac{\mathbb{E}[X^r]}{\varepsilon^r} = 0 \implies \mathbb{P}(X \leqslant \varepsilon) = 1, \quad \forall \varepsilon > 0.$$

再由分布函数的右连续性, 令 $\varepsilon \to 0^+$ 即得证. □

习题 (7.2.1) (a) 设 $X_n \xrightarrow{r} X$, 其中 $r \geqslant 1$. 证明：$\mathbb{E}[|X_n|^r] \to \mathbb{E}[|X|^r]$.

(b) 设 $X_n \xrightarrow{1} X$. 证明：$\mathbb{E}[X_n] \to \mathbb{E}[X]$. 逆命题成立吗？

(c) 设 $X_n \xrightarrow{2} X$. 证明：$\mathrm{Var}(X_n) \to \mathrm{Var}(X)$.

证明 (a) 由 Minkowski 不等式,

$$\|X_n\|_r \leqslant \|X_n - X\|_r + \|X\|_r, \quad \|X\|_r \leqslant \|X - X_n\|_r + \|X_n\|_r.$$

在以上两式中令 $n \to \infty$ 并取上极限, 利用 $\lim\limits_{n\to\infty} \|X - X_n\|_r = 0$, 就得到

$$\limsup_{n\to\infty} \|X_n\|_r \leqslant \limsup_{n\to\infty} \|X\|_r, \quad \limsup_{n\to\infty} \|X\|_r \leqslant \limsup_{n\to\infty} \|X_n\|_r.$$

故 $\|X_n\|_r \to \|X\|_r$ 即 $\mathbb{E}[|X_n|^r] \to \mathbb{E}[|X|^r]$.

(b) 由

$$|\mathbb{E}[X_n]-\mathbb{E}[X]|=|\mathbb{E}[X_n-X]|\leqslant\mathbb{E}\left[|X_n-X|\right]\to 0,\quad n\to\infty$$

可知 $\mathbb{E}[X_n]\to\mathbb{E}[X]$. 逆命题不成立, 如设 $\mathbb{P}\left(X_n=-1\right)=\mathbb{P}\left(X_n=1\right)=\dfrac{1}{2}$, $\mathbb{P}(X=0)=1$, 则 $\mathbb{E}[X_n]=0=\mathbb{E}[X]$, $\forall n$, 但 $\mathbb{E}[|X_n-X|]=\mathbb{E}[|X_n|]=1\neq 0$.

(c) 由 2 阶收敛可推出 1 阶收敛, 又由 (a) 知 $\mathbb{E}\left[X_n^2\right]\to\mathbb{E}\left[X^2\right]$. 因此

$$\operatorname{Var}(X_n)=\mathbb{E}\left[X_n^2\right]-\left(\mathbb{E}\left[X_n\right]\right)^2\to\mathbb{E}\left[X^2\right]-\left(\mathbb{E}[X]\right)^2=\operatorname{Var}(X).$$

□

习题 (7.2.5(a)) 设 $X_n\xrightarrow{\mathrm{D}}X$ 且 $Y_n\xrightarrow{\mathrm{P}}c$, 其中 c 为常数. 证明：$X_nY_n\xrightarrow{\mathrm{D}}cX$, 且当 $c\neq 0$ 时有 $\dfrac{X_n}{Y_n}\xrightarrow{\mathrm{D}}\dfrac{X}{c}$.

证明 ① 若 $c\neq 0$, 不妨设 $c>0$, 进而可不妨设 $Y_n\geqslant 0$. 对任意 $\varepsilon\in(0,c)$,

$$\begin{aligned}\mathbb{P}\left(X_nY_n\leqslant x\right)&=\mathbb{P}\left(X_nY_n\leqslant x,|Y_n-c|>\varepsilon\right)+\mathbb{P}\left(X_nY_n\leqslant x,|Y_n-c|\leqslant\varepsilon\right)\\&\leqslant\mathbb{P}\left(|Y_n-c|>\varepsilon\right)+\mathbb{P}\left(X_nY_n\leqslant x,|Y_n-c|\leqslant\varepsilon\right)\\&\leqslant\mathbb{P}\left(|Y_n-c|>\varepsilon\right)+\mathbb{P}\left(X_n\leqslant\frac{x}{c-\varepsilon}\right).\end{aligned}$$

取 $\varepsilon\in(0,c)$ 使 $\dfrac{x}{c-\varepsilon}\in\mathcal{C}_{F_X}$, 令 $n\to\infty$ 并取上极限就得到

$$\limsup_{n\to\infty}\mathbb{P}\left(X_nY_n\leqslant x\right)\leqslant\mathbb{P}\left(X\leqslant\frac{x}{c-\varepsilon}\right).$$

再令 $\varepsilon\to 0^+$(注意到分布函数的不连续点至多可数, 因此此操作与上一步相容), 就有

$$\limsup_{n\to\infty}\mathbb{P}\left(X_nY_n\leqslant x\right)\leqslant\mathbb{P}\left(X\leqslant\frac{x}{c}\right)=\mathbb{P}\left(cX\leqslant x\right).$$

又

$$\begin{aligned}\mathbb{P}\left(X_nY_n\leqslant x\right)&\geqslant\mathbb{P}\left(X_nY_n\leqslant x,|Y_n-c|\leqslant\varepsilon\right)\\&\geqslant\mathbb{P}\left(X_n\leqslant\frac{x}{c+\varepsilon},|Y_n-c|\leqslant\varepsilon\right)\\&=\mathbb{P}\left(X_n\leqslant\frac{x}{c+\varepsilon}\right)-\mathbb{P}\left(X_n\leqslant\frac{x}{c+\varepsilon},|Y_n-c|>\varepsilon\right)\\&\geqslant\mathbb{P}\left(X_n\leqslant\frac{x}{c+\varepsilon}\right)-\mathbb{P}\left(|Y_n-c|>\varepsilon\right).\end{aligned}$$

取 $\varepsilon \in (0,c)$ 使 $\dfrac{x}{c+\varepsilon} \in \mathcal{C}_{F_X}$, 令 $n \to \infty$ 并取下极限就得到

$$\liminf_{n\to\infty} \mathbb{P}\left(X_n Y_n \leqslant x\right) \geqslant \mathbb{P}\left(X \leqslant \frac{x}{c+\varepsilon}\right).$$

再令 $\varepsilon \to 0^+$, 就有

$$\liminf_{n\to\infty} \mathbb{P}\left(X_n Y_n \leqslant x\right) \geqslant \mathbb{P}\left(X \leqslant \frac{x}{c}\right) = \mathbb{P}\left(cX \leqslant x\right).$$

故 $X_n Y_n \xrightarrow{\mathrm{D}} cX$.

若 $c = 0$, 对任意 $\varepsilon, \delta > 0$,

$$\begin{aligned}\mathbb{P}\left(|X_n Y_n| > \varepsilon\right) &= \mathbb{P}\left(|X_n Y_n| > \varepsilon, |Y_n| > \delta\right) + \mathbb{P}\left(|X_n Y_n| > \varepsilon, |Y_n| \leqslant \delta\right) \\ &\leqslant \mathbb{P}\left(|Y_n| > \delta\right) + \mathbb{P}\left(|X_n| > \frac{\varepsilon}{\delta}\right) \\ &\leqslant \mathbb{P}\left(|Y_n| > \delta\right) + 1 - \mathbb{P}\left(X_n \leqslant \frac{\varepsilon}{\delta}\right) + \mathbb{P}\left(X_n \leqslant -\frac{\varepsilon}{\delta}\right).\end{aligned}$$

取 $\delta > 0$ 使 $\pm\dfrac{\varepsilon}{\delta} \in \mathcal{C}_{F_X}$, 令 $n \to \infty$ 并取上极限就得到

$$\limsup_{n\to\infty} \mathbb{P}\left(|X_n Y_n| > \varepsilon\right) \leqslant 1 - F_X\left(\frac{\varepsilon}{\delta}\right) + F_X\left(-\frac{\varepsilon}{\delta}\right).$$

再令 $\delta \to 0^+$, 利用分布函数在 $\pm\infty$ 的性质就有

$$\limsup_{n\to\infty} \mathbb{P}\left(|X_n Y_n| > \varepsilon\right) \leqslant 0.$$

故 $\lim\limits_{n\to\infty} \mathbb{P}(X_n Y_n = 0) = 1$, 即 $X_n Y_n \xrightarrow{\mathrm{D}} 0$.

② 只需证 $\dfrac{1}{Y_n} \xrightarrow{\mathrm{P}} \dfrac{1}{c}$. 对任意 $\varepsilon \in (0,c)$, 由于 $Y_n \xrightarrow{\mathrm{P}} c$, 对充分大的 n, $|Y_n - c| \overset{\text{a.s.}}{\leqslant} \varepsilon$, 从而

$$\mathbb{P}\left(\left|\frac{1}{Y_n} - \frac{1}{c}\right| < \varepsilon\right) = \mathbb{P}\left(\frac{|c - Y_n|}{|cY_n|} < \varepsilon\right) \leqslant \mathbb{P}\left(\frac{|c - Y_n|}{c(c-\varepsilon)} < \varepsilon\right) = \mathbb{P}\left(|c - Y_n| < c\varepsilon(c-\varepsilon)\right).$$

而当 $n \to \infty$ 时, RHS $\to 0$, 故 $\dfrac{1}{Y_n} \xrightarrow{\mathrm{P}} \dfrac{1}{c}$. 再由①即知 $\dfrac{X_n}{Y_n} \xrightarrow{\mathrm{D}} \dfrac{X}{c}$. □

习题 (7.2.7) 设 $\{X_n\}$ 为一随机变量列, $\{c_n\}$ 是一收敛于 c 的实数列. 对几乎处处收敛、r 阶收敛、依概率收敛、依分布收敛, 证明 $X_n \to X$ 的必要条件是 $c_n X_n \to cX$.

证明 ① 几乎处处收敛：对任意 $\omega \in \Omega$, 由 $X_n(\omega) \to X(\omega)$ 及 $c_n \to c$ 即得 $c_n X_n(\omega) \to cX(\omega)$.

② r 阶收敛：由 Minkowski 不等式,

$$\|c_n X_n - cX\|_r = \|c_n(X_n - X) + (c_n - c)X\|_r \leqslant \|c_n(X_n - X)\|_r + \|(c_n - c)X\|_r \to 0.$$

③ 依概率收敛：对任意 $\varepsilon>0$, 对于充分大的 n 有 $|c_n-c|<\varepsilon$ 且 $||c_n|-|c||<\varepsilon$, 此时

$$\begin{aligned}\mathbb{P}\left(|c_nX_n-cX|>\varepsilon\right)&=\mathbb{P}\left(|c_n(X_n-X)+(c_n-c)X|>\varepsilon\right)\\&\leqslant\mathbb{P}\left(\left\{|c_n(X_n-X)|>\frac{\varepsilon}{2}\right\}\cup\left\{|(c_n-c)X|>\frac{\varepsilon}{2}\right\}\right)\\&\leqslant\mathbb{P}\left(|c_n(X_n-X)|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(|(c_n-c)X|>\frac{\varepsilon}{2}\right)\\&\leqslant\mathbb{P}\left((|c|+\varepsilon)|X_n-X|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(\varepsilon|X|>\frac{\varepsilon}{2}\right)\\&\leqslant\mathbb{P}\left(|X_n-X|>\frac{\varepsilon}{2(|c|+\varepsilon)}\right)+\mathbb{P}\left(|X|>\frac{1}{2}\right)\\&\to 0,\quad n\to\infty.\end{aligned}$$

④ 依分布收敛：由 Skorokhod 表示定理, 存在随机变量 Y_n,Y, 使得 Y_n 与 X_n 同分布, X 与 Y 同分布, 且 $Y_n\xrightarrow{\text{a.s.}}Y$. 则 $c_nY_n\xrightarrow{\text{a.s.}}cY$, 进而 $c_nY_n\xrightarrow{\text{D}}cY$, 从而 $c_nX_n\xrightarrow{\text{D}}cX$. □

习题 (7.2.9) 设离散型随机变量 X_n 的分布列为 f_n. 称 X_n 全变差收敛于分布列为 f 的随机变量 X, 若

$$\sum_x|f_n(x)-f(x)|\to 0,\quad n\to\infty.$$

设 X_n 全变差收敛于 X, $u:\mathbb{R}\to\mathbb{R}$ 为有界函数, 证明：$\mathbb{E}[u(X_n)]\to\mathbb{E}[u(X)]$.

证明 设 $|u(x)|<M,\forall x\in\mathbb{R}$. 由佚名统计学家公式,

$$|\mathbb{E}\left[u(X_n)\right]-\mathbb{E}[u(X)]|=\left|\sum_x u(x)\left[f_n(x)-f(x)\right]\right|\leqslant M\left|\sum_x\left[f_n(x)-f(x)\right]\right|\to 0.$$

故 $\mathbb{E}[u(X_n)]\to\mathbb{E}[u(X)]$. □

习题 (7.3.6) (Weierstrass 逼近定理) 设 $f:[0,1]\to\mathbb{R}$ 连续, 随机变量 $S_n\sim B(n,x)$. 利用等式 $\mathbb{E}[Z]=\mathbb{E}\left[ZI_A\right]+\mathbb{E}\left[ZI_{A^c}\right]$, 其中 $Z=f(x)-f\left(\frac{S_n}{n}\right)$, $A=\left\{\left|\frac{S_n}{n}-x\right|>\delta\right\}$, 证明：

$$\lim_{n\to\infty}\sup_{0\leqslant x\leqslant 1}\left|f(x)-\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}\right|=0.$$

证明 因为 f 在 $[0,1]$ 上连续, 所以 f 在 $[0,1]$ 上有界且一致连续, 即

- 存在 $M>0$, 使得 $|f(x)|<M,\forall x\in[0,1]$.
- 对任意 $\varepsilon>0$, 存在 $\delta>0$, 使得对任意 $x_1,x_2\in[0,1]$, 只要 $|x_1-x_2|\leqslant\delta$, 就有 $|f(x_1)-f(x_2)|<\varepsilon$.

对这样的 ε 与 δ, 有 $|\mathbb{E}[ZI_{A^c}]| < \varepsilon$. 又由 Chebyshev 不等式,

$$\begin{aligned}|\mathbb{E}[ZI_A]| &\leqslant 2M\mathbb{P}(A) = 2M \cdot \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \delta\right) \leqslant 2M \cdot \frac{\operatorname{Var}\left(\frac{S_n}{n}\right)}{\delta^2} \\ &= 2M \cdot \frac{\operatorname{Var}(S_n)}{n^2\delta^2} = 2M \cdot \frac{x(1-x)}{n\delta^2} \leqslant \frac{M}{2n\delta^2}.\end{aligned}$$

于是

$$|\mathbb{E}[Z]| = |\mathbb{E}[ZI_A] + \mathbb{E}[ZI_{A^c}]| \leqslant \varepsilon + \frac{M}{2n\delta^2},$$

当 $n > \dfrac{M}{2\varepsilon\delta^2}$ 时, 就有 $|\mathbb{E}[Z]| < 2\varepsilon$, 故对任意 $x \in [0,1]$, 只要 $n > \dfrac{M}{2\varepsilon\delta^2}$, 就有

$$\left|f(x) - \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]\right| = \left|f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}\right| < \varepsilon.$$

□

习题 (7.3.10) 设随机变量 $X_n \sim N(0,1)$ $(n \geqslant 1)$. 证明:

(a) $\mathbb{P}\left(\limsup\limits_{n\to\infty} \dfrac{|X_n|}{\sqrt{\ln n}} = \sqrt{2}\right) = 1.$

(b) $\mathbb{P}(X_n > a_n \text{ i.o.}) = \begin{cases} 0, & 若 \sum\limits_n \mathbb{P}(X_1 > a_n) < +\infty, \\ 1, & 若 \sum\limits_n \mathbb{P}(X_1 > a_n) = +\infty. \end{cases}$

证明 (a) 设 $f(x)$ 与 $F(x)$ 分别为 $N(0,1)$ 的概率密度函数与分布函数. 由 Mills 比率, 当 $|x| \to +\infty$ 时, $1 - F(x) \sim \dfrac{f(x)}{x}$. 故对 $|\varepsilon| < 1$, 有

$$\mathbb{P}\left(|X_n| \geqslant \sqrt{2\ln n}(1+\varepsilon)\right) = 2\left[1 - F\left(\sqrt{2\ln n}(1+\varepsilon)\right)\right] \sim \frac{1}{\sqrt{\pi\ln n}(1+\varepsilon)n^{(1+\varepsilon)^2}}.$$

- 若 $\varepsilon \in (0,1)$, $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi\ln n}(1+\varepsilon)n^{(1+\varepsilon)^2}} < +\infty$. 由 Borel-Cantelli 引理,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{|X_n|}{\sqrt{\ln n}} \leqslant \sqrt{2}\right) = 1.$$

- 若 $\varepsilon \in (-1,0]$, $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi\ln n}(1+\varepsilon)n^{(1+\varepsilon)^2}} = +\infty$. 由 Borel-Cantelli 引理,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{|X_n|}{\sqrt{\ln n}} \geqslant \sqrt{2}\right) = 1.$$

故

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{|X_n|}{\sqrt{\ln n}}=\sqrt{2}\right)=1.$$

(b) 由于 $\{X_n\}_{n\geqslant 1}$ 同分布, 根据 Borel-Cantelli 引理即得证. □

习题 (7.3.13) 设随机变量列 $\{X_r \mid 1\leqslant r\leqslant n\}$ 独立同分布, 且有期望 μ 与方差 σ^2. 设 $\overline{X}=\dfrac{1}{n}\sum\limits_{r=1}^{n}X_r$. 证明:

$$\frac{\sum\limits_{r=1}^{n}(X_r-\mu)}{\sqrt{\sum\limits_{r=1}^{n}\left(X_r-\overline{X}\right)^2}}\xrightarrow{\mathrm{D}}N(0,1).$$

证明 由中心极限定理,

$$\frac{\sum\limits_{r=1}^{n}(X_r-\mu)}{\sqrt{n}\sigma}\xrightarrow{\mathrm{D}}X,$$

其中 $X\sim N(0,1)$. 又

$$\sqrt{\frac{1}{n\sigma^2}\sum_{r=1}^{n}\left(X_r-\overline{X}\right)^2}=\sqrt{\frac{1}{n\sigma^2}\sum_{r=1}^{n}(X_r-\mu)^2-\frac{2}{n\sigma^2}\left(\overline{X}-\mu\right)\sum_{r=1}^{n}(X_r-\mu)+\frac{1}{\sigma^2}\left(\overline{X}-\mu\right)^2},$$

由弱大数律,

$$\frac{1}{n\sigma^2}\sum_{r=1}^{n}(X_r-\mu)^2\xrightarrow{\mathrm{P}}1,\quad \frac{2}{n\sigma^2}\left(\overline{X}-\mu\right)\sum_{r=1}^{n}(X_r-\mu)\xrightarrow{\mathrm{P}}0,\quad \frac{1}{\sigma^2}\left(\overline{X}-\mu\right)^2\xrightarrow{\mathrm{P}}0,$$

因此

$$\sqrt{\sum_{r=1}^{n}\left(X_r-\overline{X}\right)^2}\xrightarrow{\mathrm{P}}1.$$

再由 Slutsky 定理 (习题 7.2.5(a)) 得

$$\frac{\sum\limits_{r=1}^{n}(X_r-\mu)}{\sqrt{\sum\limits_{r=1}^{n}\left(X_r-\overline{X}\right)^2}}\xrightarrow{\mathrm{D}}X,$$

其中 $X\sim N(0,1)$. □

习题 (7.4.1) 设独立随机变量 $X_2, X_3, \cdots$ 满足

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n\ln n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n\ln n}.$$

证明：这列随机变量符合弱大数律但不符合强大数律, 即 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 依概率收敛于 0 但非几乎处处收敛于 0.

证明 记 $S_n = \sum_{i=1}^{n} X_i$. 由 Markov 不等式,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = \mathbb{P}\left(S_n^2 > \varepsilon^2 n^2\right) \leqslant \frac{\mathbb{E}\left[S_n^2\right]}{\varepsilon^2 n^2}, \quad \forall \varepsilon > 0.$$

而

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[X_1^2\right] + \sum_{i=2}^{n} \mathbb{E}\left[X_i^2\right] = \mathbb{E}\left[X_1^2\right] + \sum_{i=2}^{n} i^2 \cdot \frac{1}{i\ln i} = \mathbb{E}\left[X_1^2\right] + \sum_{i=2}^{n} \frac{i}{\ln i}.$$

当 $i \geqslant 3$ 时, $\left\{\frac{i}{\ln i}\right\}$ 单调递增, 对 $n \geqslant i \geqslant 3$ 有 $\frac{i}{\ln i} \leqslant \frac{n}{\ln n}$. 因此存在常数 C, 使得

$$\mathbb{E}\left[S_n^2\right] \leqslant C\frac{n^2}{\ln n}.$$

故

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leqslant \frac{C}{\varepsilon^2 \ln n} \xrightarrow{n\to\infty} 0, \quad \forall \varepsilon > 0,$$

即 $\frac{S_n}{n} \xrightarrow{\mathrm{P}} 0$. 又

$$\sum_{n=2}^{\infty} \mathbb{P}\left(|X_n| \geqslant n\right) = \sum_{n=2}^{\infty} \frac{1}{n\ln n} = +\infty,$$

由 Borel-Cantelli 引理, $\mathbb{P}\left(|X_n| \geqslant n \text{ i.o.}\right) = 1$. 假设 $\frac{S_n}{n}$ 几乎处处收敛, 则

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{S_{n-1}}{n-1}\right| \geqslant 1 \text{ i.o.}\right) = 0 \implies \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{S_{n-1}}{n}\right| \geqslant 1 \text{ i.o.}\right) = 0,$$

但 $\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{S_{n-1}}{n}\right| \geqslant 1 \text{ i.o.}\right) = \mathbb{P}\left(|S_n - S_{n-1}| \geqslant n \text{ i.o.}\right) = \mathbb{P}\left(|X_n| \geqslant n \text{ i.o.}\right) = 1$, 矛盾. 故 $\frac{S_n}{n}$ 不几乎处处收敛. □

习题 (7.5.1) 设将区间 $[0,1]$ 划分为不交的子区间, 长度分别为 $p_1,p_2,\cdots,p_n$, 定义此划分的熵为

$$h=-\sum_{i=1}^{n}p_i\ln p_i.$$

设随机变量 $\{X_k\}$ 相互独立且服从 $[0,1]$ 上的均匀分布, 用随机变量 $Z_m(i)$ 表示前 m 个随机变量中取值在第 i 个子区间者的个数. 证明:

$$R_m=\prod_{i=1}^{n}p_i^{Z_m(i)}$$

满足 $\dfrac{\ln R_m}{m}\xrightarrow{\text{a.s.}}-h$.

证明 设 $A_{ij}=\{X_j$ 取值在第 i 个子区间中$\}$, 令 $I_{ij}=I_{A_{ij}}$. 则

$$\ln R_m=\sum_{i=1}^{n}Z_m(i)\ln p_i=\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}\ln p_i=\sum_{j=1}^{m}\sum_{i=1}^{n}I_{ij}\ln p_i.$$

设 $Y_j=\sum\limits_{i=1}^{n}I_{ij}\ln p_i$, 则

$$\mathbb{E}[Y_j]=\sum_{i=1}^{n}\mathbb{P}(A_{ij})\ln p_i=\sum_{i=1}^{n}p_i\ln p_i=-h.$$

因为 $\{Y_j\}$ 相互独立且同分布, 由 Kolmogorov 强大数律,

$$\frac{\ln R_m}{m}=\frac{1}{m}\sum_{j=1}^{m}Y_j\xrightarrow{\text{a.s.}}-h.$$

□

习题 (7.11.4) 设随机变量 $\{Y_k\}$ 相互独立且同分布, 均在 $\{0,1,\cdots,9\}$ 中等概率取值. 设 $X_n=\sum\limits_{i=1}^{n}Y_i10^{-i}$. 利用特征函数证明 X_n 依分布收敛于 $[0,1]$ 上的均匀分布, X_n 几乎处处收敛于 $[0,1]$ 上的均匀分布.

证明 X_n 的特征函数

$$\begin{aligned}\phi_{X_n}(t)&=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX_n}\right]=\prod_{k=1}^{n}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tY_k10^{-k}}\right]=\prod_{k=1}^{n}\sum_{j=0}^{9}\frac{1}{10}\,\mathrm{e}^{\mathrm{i}tj10^{-k}}\\&=\prod_{k=1}^{n}\left[\frac{1}{10}\sum_{j=0}^{9}\left(\mathrm{e}^{\mathrm{i}t10^{-k}}\right)^{j}\right]=\prod_{k=1}^{n}\left(\frac{1}{10}\cdot\frac{1-\mathrm{e}^{\mathrm{i}t10^{-k+1}}}{1-\mathrm{e}^{\mathrm{i}t10^{-k}}}\right)\end{aligned}$$

$$= \frac{1}{10^n} \cdot \frac{1-\mathrm{e}^{\mathrm{i}\,t}}{1-\mathrm{e}^{\mathrm{i}\,t10^{-n}}} \to \frac{\mathrm{e}^{\mathrm{i}\,t}-1}{\mathrm{i}\,t}, \quad n \to \infty.$$

设 $Y \sim U[0,1]$, 则 Y 的特征函数

$$\phi_Y(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tY}\right] = \int_0^1 \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}x = \frac{\mathrm{e}^{\mathrm{i}\,t-1}}{\mathrm{i}\,t} = \lim_{n\to\infty} \phi_{X_n}(t).$$

又对任意 $\omega \in \Omega$, $\{X_n(\omega)\}$ 单调递增且有上界 1, 所以 $X_n \xrightarrow{\text{a.s.}} Y$, 进而 $X_n \xrightarrow{\text{D}} Y$. □

习题 (7.11.6) 设随机变量 $\{X_k\}$ 相互独立且同分布, 期望为 0, 四阶矩 $\mathbb{E}\left[X_1^4\right] < +\infty$. 证明: $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} 0$.【本题不得使用强大数律!】

证明 记 $S_n = \sum_{i=1}^n X_i$. 由 Markov 不等式,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = \mathbb{P}\left(S_n^4 > \varepsilon^4 n^4\right) \leqslant \frac{\mathbb{E}\left[S_n^4\right]}{\varepsilon^4 n^4}, \quad \forall \varepsilon > 0.$$

而

$$\begin{aligned}\mathbb{E}\left[S_n^4\right] =& \mathrm{C}_n^1\mathbb{E}\left[X_1^4\right] + \mathrm{C}_n^2\mathrm{C}_2^1\mathrm{C}_4^1\mathbb{E}\left[X_1^3X_2\right] + \mathrm{C}_n^2\mathrm{C}_4^2\mathbb{E}\left[X_1^2X_2^2\right]\\ &+ \mathrm{C}_n^3\mathrm{C}_3^1\mathrm{C}_4^2\mathrm{C}_2^1\mathbb{E}\left[X_1^2X_2X_3\right] + \mathrm{C}_n^4\mathrm{A}_4^4\mathbb{E}\left[X_1X_2X_3X_4\right]\\ =& n\mathbb{E}\left[X_1^4\right] + 3n(n-1)\left(\mathbb{E}\left[X_1^2\right]\right)^2,\end{aligned}$$

最后一步用到了若高阶矩存在则低阶矩也存在 (习题 (5.6.4)). 于是存在常数 C 使得 $\mathbb{E}\left[S_n^4\right] \leqslant Cn^2$, 从而

$$\sum_{n=1}^\infty \mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leqslant \sum_{n=1}^\infty \frac{C}{\varepsilon^4 n^2} < +\infty, \quad \forall \varepsilon > 0.$$

由 Borel-Cantelli 引理, $\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon \text{ i.o.}\right) = 0$, $\forall \varepsilon > 0$, 即 $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$. □

补充题 9 设非负随机变量 $\{X_k\}$ 相互独立且同分布, $\mathbb{E}[X_1] = +\infty$. 证明: $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\text{a.s.}} +\infty$.

证明 记 $S_n = \sum_{k=1}^n X_k$. 对任意 $M > 0$, 设 $X_k^{(M)} = \min\{X_k, M\}$, 则 $\left\{X_k^{(M)}\right\}$ 独立同分布且 $\mathbb{E}\left[X_1^{(M)}\right] < +\infty$. 记 $S_n^{(M)} = \sum_{k=1}^n X_k^{(M)}$, 由强大数律,

$$\frac{S_n^{(M)}}{n} \xrightarrow{\text{a.s.}} \mathbb{E}\left[X_1^{(M)}\right], \quad n \to \infty.$$

而 $X_k \geqslant X_k^{(M)}$, 因此

$$\liminf_{n\to\infty}\frac{S_n}{n} \geqslant \lim_{n\to\infty}\frac{S_n^{(M)}}{n} \quad 即 \quad \liminf_{n\to\infty}\frac{S_n}{n} \overset{\text{a.s.}}{\geqslant} \mathbb{E}\left[X_1^{(M)}\right], \quad \forall M>0.$$

又 $X_1^{(M+1)} \geqslant X_1^{(M)} \geqslant 0$, $\lim\limits_{M\to+\infty} X_1^{(M)} = X_1$, 由单调收敛定理, $\lim\limits_{M\to+\infty}\mathbb{E}\left[X_1^{(M)}\right] = \mathbb{E}[X_1] = +\infty$. 故在上式中令 $M\to+\infty$, 就得到

$$\liminf_{n\to\infty}\frac{S_n}{n} \xlongequal{\text{a.s.}} +\infty \quad 即 \quad \frac{S_n}{n} \xrightarrow{\text{a.s.}} +\infty.$$

□

补充题 10 记

$$I_n = \int_{[0,1]^n} \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n.$$

证明：$\lim\limits_{n\to\infty} I_n$ 存在.

证明 1 对 $p\in(-1,0)$, 设

$$I_n^{(p)} = \int_{[0,1]^n}\left(\frac{x_1^p+x_2^p+\cdots+x_n^p}{n}\right)^{\frac{1}{p}}\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n.$$

设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且同分布, $X_k\sim U[0,1]$, 则 $X_1^p,X_2^p\cdots,X_n^p$ 独立同分布. 由于

$$\mathbb{E}\left[X_1^p\right] = \int_0^1 x^p\,\mathrm{d}x = \frac{1}{p+1},$$

根据强大数律,

$$\frac{X_1^p+X_2^p+\cdots+X_n^p}{n} \xrightarrow{\text{a.s.}} \mathbb{E}\left[X_1^p\right] = \frac{1}{p+1}.$$

因为 $f(x)=x^{\frac{1}{p}}$ 在 $x=\dfrac{1}{p+1}$ 处连续, 所以

$$\left(\frac{X_1^p+X_2^p+\cdots+X_n^p}{n}\right)^{\frac{1}{p}} \xrightarrow{\text{a.s.}} \left(\frac{1}{p+1}\right)^{\frac{1}{p}}.$$

由 $g(x)=x^p$ $(p<0)$ 是 $[0,1]$ 上的凸函数可得

$$\frac{1}{n}\sum_{k=1}^n g(X_k) \geqslant g\left(\frac{1}{n}\sum_{k=1}^n X_k\right),$$

从而

$$\left|\left(\frac{X_1^p+X_2^p+\cdots+X_n^p}{n}\right)^{\frac{1}{p}}\right| \leqslant \frac{1}{n}\sum_{k=1}^n X_k \leqslant 1,$$

由控制收敛定理,

$$\mathbb{E}\left[\left(\frac{X_1^p+X_2^p+\cdots+X_n^p}{n}\right)^{\frac{1}{p}}\right]\to\mathbb{E}\left[\left(\frac{1}{p+1}\right)^{\frac{1}{p}}\right]=\left(\frac{1}{p+1}\right)^{\frac{1}{p}},\quad n\to\infty.$$

由 $X_1,X_2,\cdots,X_n$ 独立同分布, 这即是

$$\lim_{n\to\infty}I_n^{(p)}=\left(\frac{1}{p+1}\right)^{\frac{1}{p}}.$$

故

$$\lim_{n\to\infty}I_n=\lim_{p\to-1^+}\lim_{n\to\infty}I_n^{(p)}=\lim_{p\to-1^+}\left(\frac{1}{p+1}\right)^{\frac{1}{p}}=\exp\left(-\lim_{p\to-1^+}\frac{\ln(p+1)}{p}\right)=0.$$

□

证明 2 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且同分布, $X_k\sim U[0,1]$, 则 $\mathbb{E}\left[\frac{1}{X_1}\right]=+\infty$. 由补充题9结论, $\frac{1}{n}\sum\limits_{k=1}^{n}\frac{1}{X_k}\xrightarrow{\text{a.s.}}+\infty$. 因为 $\left|\frac{n}{\sum\limits_{k=1}^{n}\frac{1}{X_k}}\right|\leqslant 1$, 所以由控制收敛定理,

$$I_n=\mathbb{E}\left[\frac{1}{\frac{1}{n}\sum\limits_{k=1}^{n}\frac{1}{X_k}}\right]\to 0,\quad n\to\infty.$$

□

习题 (5.7.7) 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立, $X_i\sim N(\mu_i,1)$, 并设 $Y=X_1^2+X_2^2+\cdots+X_n^2$. 证明：$Y$ 的特征函数为

$$\phi_Y(t)=\frac{1}{(1-2\mathrm{i}\,t)^{\frac{n}{2}}}\exp\left(\frac{\mathrm{i}\,t\theta}{1-2\mathrm{i}\,t}\right),$$

其中 $\theta=\mu_1^2+\mu_2^2+\cdots+\mu_n^2$. 我们称随机变量 Y 服从自由度为 n、非中心参数为 θ 的非中心卡方分布, 记作 $Y\sim\chi^2(n;\theta)$.

证明 由

$$\phi_{X_k^2}(t)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX_k^2}\right]=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx^2}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\mu_k)^2}\,\mathrm{d}x\xlongequal{s=\mathrm{i}\,t}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathrm{e}^{\left(s-\frac{1}{2}\right)x^2+\mu_kx-\frac{1}{2}\mu_k^2}\,\mathrm{d}x$$

$$\xlongequal{u=\sqrt{\frac{1}{2}-s}x}\frac{1}{\sqrt{\pi(1-2s)}}\int_{\mathbb{R}}\exp\left(-\left(u-\frac{\mu_k}{2\sqrt{\frac{1}{2}-s}}\right)^2+\frac{\mu_k^2}{4\left(\frac{1}{2}-s\right)}-\frac{1}{2}\mu_k^2\right)\,\mathrm{d}u$$

$$= \frac{1}{\sqrt{1-2s}} \exp\left(\frac{\mu_k^2 s}{1-2s}\right) = \frac{1}{(1-2\,\mathrm{i}\,t)^{\frac{1}{2}}} \exp\left(\frac{\mu_k^2\,\mathrm{i}\,t}{1-2\,\mathrm{i}\,t}\right)$$

及 $X_1^2, \cdots, X_n^2$ 独立可得

$$\phi_Y(t) = \prod_{k=1}^n \phi_{X_k^2}(t) = \prod_{k=1}^n \left[\frac{1}{(1-2\,\mathrm{i}\,t)^{\frac{1}{2}}} \exp\left(\frac{\mu_k^2\,\mathrm{i}\,t}{1-2\,\mathrm{i}\,t}\right)\right] = \frac{1}{(1-2\,\mathrm{i}\,t)^{\frac{n}{2}}} \exp\left(\frac{\mathrm{i}\,t\theta}{1-2\,\mathrm{i}\,t}\right).$$

□

习题 (5.8.7) 设 $X, Y \sim N(0,1)$ 独立, U, V 与 X, Y 独立. 证明：$Z := \dfrac{UX+VY}{\sqrt{U^2+V^2}} \sim N(0,1)$. 推广这一结论到 (X,Y) 服从期望为 0、方差为 1、协方差为 ρ 的二元标准正态分布的情形.

证明 ① 对任意 $\theta \in \mathbb{R}$,

$$\begin{aligned}\phi_{\cos\theta X+\sin\theta Y}(t) &= \phi_{\cos\theta X}(t)\phi_{\sin\theta Y}(t) = \phi_X(\cos\theta t)\phi_Y(\sin\theta t) \\ &= \mathrm{e}^{-\frac{1}{2}\cos^2\theta t^2} \cdot \mathrm{e}^{-\frac{1}{2}\sin^2\theta t^2} = \mathrm{e}^{-\frac{1}{2}t^2}.\end{aligned}$$

因此

$$\phi_Z(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tZ}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tZ} \mid U, V\right]\right] = \mathbb{E}\left[\mathrm{e}^{-\frac{1}{2}t^2}\right] = \mathrm{e}^{-\frac{1}{2}t^2}.$$

由唯一性定理知 $Z \sim N(0,1)$.

② 对任意 $u, v \in \mathbb{R}$, $uX+vY$ 的特征函数

$$\phi(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t(uX+vY)}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t(uX+vY)} \mid X\right]\right] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tuX}\,\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tvY} \mid X\right]\right].$$

而

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tvY} \mid X = x\right] &= \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,tvy} \frac{\frac{1}{2\pi\sqrt{1-\rho^2}}\,\mathrm{e}^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}}{\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2}}\,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{(\rho x-y)^2}{2(1-\rho^2)}+\mathrm{i}\,tvy}\,\mathrm{d}y \\ &\xlongequal{w=\frac{y-\rho x}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}w^2+\mathrm{i}\,tv\left(\sqrt{1-\rho^2}w+\rho x\right)}\,\mathrm{d}w \\ &\xlongequal{s=\mathrm{i}\,tv\sqrt{1-\rho^2}} \frac{\mathrm{e}^{\mathrm{i}\,tv\rho x}}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}w^2+sw}\,\mathrm{d}w \\ &= \frac{\mathrm{e}^{\mathrm{i}\,tv\rho x}}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(w-s)^2+\frac{1}{2}s^2}\,\mathrm{d}w \\ &= \mathrm{e}^{\mathrm{i}\,tv\rho x+\frac{1}{2}s^2} = \mathrm{e}^{\mathrm{i}\,tv\rho x-\frac{t^2v^2(1-\rho^2)}{2}},\end{aligned}$$

因此

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tvY} \mid X\right] = \mathrm{e}^{\mathrm{i}\,tv\rho X - \frac{t^2v^2(1-\rho^2)}{2}}.$$

故 $uX+vY$ 的特征函数

$$\begin{aligned}\phi(t) &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t(u+\rho v)X - \frac{t^2v^2(1-\rho^2)}{2}}\right] = \mathrm{e}^{-\frac{t^2v^2(1-\rho^2)}{2}}\,\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t(u+\rho v)X}\right] \\ &= \frac{\mathrm{e}^{-\frac{t^2v^2(1-\rho^2)}{2}}}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\,t(u+\rho v)x - \frac{1}{2}x^2}\,\mathrm{d}x \\ &\overset{m=\mathrm{i}\,t(u+\rho v)}{=\!=\!=\!=} \frac{\mathrm{e}^{-\frac{t^2v^2(1-\rho^2)-m^2}{2}}}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x-m)^2}\,\mathrm{d}x \\ &= \mathrm{e}^{-\frac{t^2v^2(1-\rho^2)+t^2(u+\rho v)^2}{2}} = \mathrm{e}^{-\frac{u^2+2\rho uv+v^2}{2}t^2}.\end{aligned}$$

由此可知, 若记 $W \coloneqq \dfrac{uX+vY}{\sqrt{u^2+2\rho uv+v^2}}$, 则 W 的特征函数

$$\phi_W(t) = \mathrm{e}^{-\frac{1}{2}t^2}.$$

由唯一性定理知 $W \sim N(0,1)$. 再设 $Z \coloneqq \dfrac{UX+VY}{\sqrt{U^2+2\rho UV+V^2}}$, 则 Z 的特征函数

$$\phi_Z(t) = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tZ} \mid U,V\right]\right] = \mathbb{E}\left[\mathrm{e}^{-\frac{1}{2}t^2}\right] = \mathrm{e}^{-\frac{1}{2}t^2},$$

由唯一性定理知 $Z \sim N(0,1)$. □

习题 (5.8.8) 设 $X \sim \mathrm{Exp}(\lambda)$. 证明：$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX}\right] = \dfrac{\lambda}{\lambda - \mathrm{i}\,t}$.

证明 利用不定积分

$$\begin{aligned}\int \cos(tx)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x &= \frac{t\sin(tx)-\lambda\cos(tx)}{\lambda^2+t^2}\,\mathrm{e}^{-\lambda x}+C, \\ \int \sin(tx)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x &= \frac{-\lambda\sin(tx)-t\cos(tx)}{\lambda^2+t^2}\,\mathrm{e}^{-\lambda x}+C\end{aligned}$$

直接计算得

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX}\right] &= \lambda\int_0^{+\infty} \mathrm{e}^{(\mathrm{i}\,t-\lambda)x}\,\mathrm{d}x \\ &= \lambda\int_0^{+\infty}\cos(tx)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x + \lambda\,\mathrm{i}\int_0^{+\infty}\sin(tx)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \\ &= \frac{\lambda^2}{\lambda^2+t^2} + \frac{\mathrm{i}\,\lambda t}{\lambda^2+t^2} \\ &= \frac{\lambda}{\lambda-\mathrm{i}\,t}.\end{aligned}$$

□

习题 (5.8.9) 求下列概率密度函数的特征函数：

(a) $f(x)=\dfrac{1}{2}\,\mathrm{e}^{-|x|}$, $x\in\mathbb{R}$.

(b) $f(x)=\dfrac{1}{2}|x|\,\mathrm{e}^{-|x|}$, $x\in\mathbb{R}$.

证明 (a) 我们有

$$\phi(t)=\frac{1}{2}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\,tx-|x|}\,\mathrm{d}x=\frac{1}{2}\int_{-\infty}^{0}\mathrm{e}^{(\mathrm{i}\,t+1)x}\,\mathrm{d}x+\frac{1}{2}\int_{0}^{+\infty}\mathrm{e}^{(\mathrm{i}\,t-1)x}\,\mathrm{d}x=\frac{1}{t^2+1}.$$

(b) 我们有

$$\begin{aligned}\phi(t)&=\frac{1}{2}\int_{\mathbb{R}}|x|\,\mathrm{e}^{\mathrm{i}\,tx-|x|}\,\mathrm{d}x=-\frac{1}{2}\int_{-\infty}^{0}x\,\mathrm{e}^{(\mathrm{i}\,t+1)x}\,\mathrm{d}x+\frac{1}{2}\int_{0}^{+\infty}x\,\mathrm{e}^{(\mathrm{i}\,t-1)x}\,\mathrm{d}x\\&=\frac{1}{2(\mathrm{i}\,t+1)}\int_{-\infty}^{0}\mathrm{e}^{(\mathrm{i}\,t+1)x}\,\mathrm{d}x+\frac{1}{2(1-\mathrm{i}\,t)}\int_{0}^{+\infty}\mathrm{e}^{(\mathrm{i}\,t-1)x}\,\mathrm{d}x=\frac{1-t^2}{(1+t^2)^2}.\end{aligned}$$

□

习题 (5.8.10) 设 $U\sim U[0,1]$, 问是否存在同分布随机变量 X,Y,Z, 其中 Y,Z 相互独立且均与 U 独立, 使得 $X=U(Y+Z)$?

解 由 $M_{U(Y+Z)}(t)=\mathbb{E}\left[\mathrm{e}^{tU(Y+Z)}\right]=\mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{tU(Y+Z)}\mid U\right]\right]=\mathbb{E}\left[M_{Y+Z}(tU)\right]=\int_0^1 M_{Y+Z}(tu)\,\mathrm{d}u$ 及 Y 与 Z 独立可知, 若 $X=U(Y+Z)$, 则需有

$$M_X(t)=\int_0^1 M_{Y+Z}(tu)\,\mathrm{d}u=\int_0^1\left(M_X(tu)\right)^2\,\mathrm{d}u=\frac{1}{t}\int_0^t\left(M_X(v)\right)^2\,\mathrm{d}v,$$

也即 X 的矩母函数 $M_X(t)$ 需满足积分方程

$$tM_X(t)=\int_0^t\left(M_X(v)\right)^2\,\mathrm{d}v.$$

两边对 t 求导化为常微分方程

$$M_X(t)+tM_X'(t)=\left(M_X(t)\right)^2.$$

求得通解为

$$M_X(t)=\frac{1}{1+ct},$$

其中 c 为任意常数. 注意到若 $X\sim\mathrm{Exp}(\lambda)$, 则 $M_X(t)=\dfrac{\lambda}{\lambda-t}$, 因此这样的 X 满足要求. □

习题 (5.9.2) 设 X_n 的分布函数为

$$F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, \quad 0 \leqslant x \leqslant 1.$$

(a) 证明：F_n 的确是分布函数, 且 X_n 有概率密度函数.

(b) 证明：当 $n \to \infty$ 时, F_n 收敛于 $[0,1]$ 上均匀分布的分布函数, 但 F_n 对应的密度函数不收敛于均匀分布的密度函数.

证明 (a) $f_n(x) \coloneqq F_n'(x) = 1 - \cos(2n\pi x) \geqslant 0$, 且 $F_n(0) = 0, F_n(1) = 1$, 因此 F_n 的确实分布函数, 且 f_n 即为 X_n 的密度函数.

(b) 由 $\lim\limits_{n\to\infty} F_n(x) = x$ 知 F_n 收敛于 $[0,1]$ 上均匀分布的分布函数. 而当 $n \to \infty$ 时, f_n 极限不存在. □

习题 (5.9.5) 利用反转公式证明：

$$\int_{-\infty}^{+\infty} \frac{\sin(at)\sin(bt)}{t^2}\,\mathrm{d}t = \pi\min\{a,b\}, \quad \forall a,b > 0.$$

证明 设 $X \sim U[-a,a]$ 与 $Y \sim U[-b,b]$ 独立. 则

$$\phi_X(t) = \frac{1}{2a}\int_{-a}^{a} \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}x = \frac{\sin(at)}{at},$$
$$\phi_Y(t) = \frac{1}{2b}\int_{-b}^{b} \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}x = \frac{\sin(bt)}{bt}.$$

对 $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \dfrac{\sin(at)\sin(bt)}{abt^2}$ 作 Fourier 反变换得

$$f_{X+Y}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{i}\,tx}\,\phi_{X+Y}(t)\,\mathrm{d}t = \frac{1}{2\pi ab}\int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{i}\,tx}\,\frac{\sin(at)\sin(bt)}{t^2}\,\mathrm{d}t.$$

记 $c = \min\{a,b\} > 0$, 则 f_{X+Y} 在 $(-c,c)$ 上可微, 从而

$$f_{X+Y}(0) = \frac{1}{2\pi ab}\int_{-\infty}^{+\infty} \frac{\sin(at)\sin(bt)}{t^2}\,\mathrm{d}t.$$

另一方面,

$$f_{X+Y}(0) = \mathbb{P}(X+Y=0) = \int_{-c}^{c} \frac{1}{2a}\cdot\frac{1}{2b}\,\mathrm{d}x = \frac{c}{2ab}.$$

故

$$\int_{-\infty}^{+\infty} \frac{\sin(at)\sin(bt)}{t^2}\,\mathrm{d}t = \pi\cdot c = \pi\min\{a,b\}.$$

□

补充题 11 求 $\cos^2 t$ 对应的的分布函数.

解 设相互独立的随机变量 X 与 Y 满足 $\mathbb{P}(X=-1)=\mathbb{P}(X=1)=\mathbb{P}(Y=-1)=\mathbb{P}(Y=1)=\dfrac{1}{2}$. 我们已知 $\phi_X(t)=\phi_Y(t)=\cos t$, 于是 $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)=\cos^2 t$. 由 $X+Y$ 的分布列

n	-2	0	2
$\mathbb{P}(X+Y=n)$	$\dfrac{1}{4}$	$\dfrac{1}{2}$	$\dfrac{1}{4}$

即得 $\cos^2 t$ 对应的分布函数为 $F_{X+Y}(x)=\begin{cases}1, & x\geqslant 2,\\ \dfrac{3}{4}, & 0\leqslant x<2,\\ \dfrac{1}{4}, & -2\leqslant x<0,\\ 0, & x<-2.\end{cases}$ □

习题 (5.10.1(b)) 证明：对 $x\geqslant 0$, 当 $n\to\infty$ 时,

$$\sum_{k:|k-n|\leqslant x\sqrt{n}}\frac{n^k}{k!}\sim \mathrm{e}^n\int_{-x}^{x}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}u^2}\,\mathrm{d}u.$$

证明 设随机变量 $\{X_k\}$ 相互独立且同分布, $X_k\sim P(\lambda)$, $S_n=\sum\limits_{k=1}^{n}X_k$. 则由 X_1 的特征函数

$$\phi_{X_1}(t)=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX_1}\right]=\sum_{k=0}^{\infty}\mathrm{e}^{\mathrm{i}\,tk}\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda\,\mathrm{e}^{\mathrm{i}\,t}\right)^k}{k!}=\mathrm{e}^{\lambda\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)}$$

以及 S_n 的特征函数

$$\phi_{S_n}(t)=(\phi_{X_1}(t))^n=\mathrm{e}^{n\lambda\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)}$$

可知 $S_n\sim P(n\lambda)$. 现取 $\lambda=1$, 则 $\mathbb{P}(S_n=k)=\dfrac{n^k}{k!}\,\mathrm{e}^{-n}$. 又 $\mathbb{E}[X_1]=\lambda=1, \mathrm{Var}(X_1)=\lambda=1$, 由中心极限定理,

$$\frac{S_n-n}{\sqrt{n}}\xrightarrow{\mathrm{D}}N(0,1),$$

从而

$$\mathbb{P}\left(\left|\frac{S_n-n}{\sqrt{n}}\right|\leqslant x\right)\to\int_{-x}^{x}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}u^2}\,\mathrm{d}u,\quad n\to\infty.$$

而

$$\mathbb{P}\left(\left|\frac{S_n-n}{\sqrt{n}}\right|\leqslant x\right)=\mathbb{P}\left(|S_n-n|\leqslant x\sqrt{n}\right)=\sum_{k:|k-n|\leqslant x\sqrt{n}}\frac{n^k}{k!}\,\mathrm{e}^{-n},$$

这就完成了证明. □

习题 (5.10.3) 设 $X \sim \Gamma(1,s)$. 对给定的 $X=x$, 设 $Y \sim P(x)$. 求 Y 的特征函数, 并证明

$$\frac{Y-\mathbb{E}[Y]}{\sqrt{\operatorname{Var}(Y)}} \xrightarrow{\mathrm{D}} N(0,1), \quad s \to +\infty.$$

解释它与中心极限定理的联系.

注：Gamma 分布 $\Gamma(\lambda,t)$ 的概率密度函数为 $f(x)=\frac{1}{\Gamma(t)}\lambda^t x^{t-1}\,\mathrm{e}^{-\lambda x}$, $x \geqslant 0$, 其中 $\Gamma(t)=\int_0^{+\infty} x^{t-1}\,\mathrm{e}^{-x}\,\mathrm{d}x$.

解 Y 的特征函数

$$\begin{aligned}
\phi_Y(t) &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tY}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tY} \mid X\right]\right] = \mathbb{E}\left[\mathrm{e}^{X(\mathrm{e}^{\mathrm{i}\,t}-1)}\right] = \int_0^{+\infty} \mathrm{e}^{(\mathrm{e}^{\mathrm{i}\,t}-1)x}\,\frac{1}{\Gamma(s)}x^{s-1}\,\mathrm{e}^{-x}\,\mathrm{d}x \\
&= \frac{1}{\Gamma(s)}\int_0^{+\infty} x^{s-1}\,\mathrm{e}^{(\mathrm{e}^{\mathrm{i}\,t}-2)x}\,\mathrm{d}x \\
&= \frac{1}{(2-\mathrm{e}^{\mathrm{i}\,t})^s\,\Gamma(s)}\int_0^{+\infty}\left[\left(2-\mathrm{e}^{\mathrm{i}\,t}\right)x\right]^{s-1}\mathrm{e}^{-\left(2-\mathrm{e}^{\mathrm{i}\,t}\right)x}\,\mathrm{d}\left(2-\mathrm{e}^{\mathrm{i}\,t}\right)x \\
&= \frac{1}{(2-\mathrm{e}^{\mathrm{i}\,t})^s}.
\end{aligned}$$

由此可知

$$\mathbb{E}[Y]=\frac{1}{\mathrm{i}}\phi_Y'(0)=s, \quad \mathbb{E}\left[Y^2\right]=\frac{1}{\mathrm{i}^2}\phi_Y''(0)=s(s+2),$$

进而

$$\operatorname{Var}(Y)=\mathbb{E}\left[Y^2\right]-(\mathbb{E}[Y])^2=s(s+2)-s^2=2s.$$

于是 $Z := \frac{Y-\mathbb{E}[Y]}{\sqrt{\operatorname{Var}(Y)}}=\frac{1}{\sqrt{2s}}Y-\sqrt{\frac{s}{2}}$ 的特征函数

$$\phi_Z(t)=\mathrm{e}^{-\mathrm{i}\,t\sqrt{\frac{s}{2}}}\,\phi_Y\left(\frac{t}{\sqrt{2s}}\right).$$

当 $s \to +\infty$ 时,

$$\begin{aligned}
\phi_Y\left(\frac{t}{\sqrt{2s}}\right) &= \exp\left(-s\log\left(2-\mathrm{e}^{\mathrm{i}\,\frac{t}{\sqrt{2s}}}\right)\right)=\exp\left(-s\log\left[1-\left(\mathrm{e}^{\mathrm{i}\,\frac{t}{\sqrt{2s}}}-1\right)\right]\right) \\
&= \exp\left(s\left[\left(\mathrm{e}^{\mathrm{i}\,\frac{t}{\sqrt{2s}}}-1\right)+\frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\,\frac{t}{\sqrt{2s}}}-1\right)^2\right]+o(1)\right) \\
&= \exp\left(\frac{s}{2}\left(\mathrm{e}^{\mathrm{i}\,\frac{2t}{\sqrt{2s}}}-1\right)+o(1)\right) \\
&= \exp\left(\frac{s}{2}\left(\mathrm{i}\,\frac{2t}{\sqrt{2s}}+\frac{1}{2}\left(\mathrm{i}\,\frac{2t}{\sqrt{2s}}\right)^2\right)+o(1)\right) \\
&= \exp\left(\mathrm{i}\,t\sqrt{\frac{s}{2}}-\frac{1}{2}t^2+o(1)\right).
\end{aligned}$$

故

$$\phi_Z(t) \to \mathrm{e}^{-\frac{1}{2}t^2}, \quad s \to +\infty,$$

因为 $\mathrm{e}^{-\frac{1}{2}t^2}$ 是标准正态分布的特征函数 (自然在 $t=0$ 处连续), 由 Lévy-Cramér 连续性定理, 当 $s \to +\infty$ 时, $Z \xrightarrow{\mathrm{D}} N(0,1)$.

与中心极限定理的联系：若将 s 和 X 视作整数, 则 $X \sim \Gamma(1,s)$ 意味着 X 是 s 个服从参数为 1 的指数分布的独立随机变量之和, 故当 $s \to +\infty$ 时 $X \xrightarrow{\text{a.s.}} +\infty$. 而 Y 服从参数为 X 的 Poisson 分布, 当 $X \to +\infty$ 时, 由中心极限定理, Y 规范化后接近标准正态分布. □

习题 (5.12.33) (a) 设 $X \sim P(\lambda)$, 证明：$Y_\lambda \coloneqq \dfrac{X-\mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}} \xrightarrow{\lambda\to+\infty} N(0,1)$.

(b) 设 $X \sim \Gamma(1,\lambda)$, 证明：$Y_\lambda \coloneqq \dfrac{X-\mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}} \xrightarrow{\lambda\to+\infty} N(0,1)$.

(c) 证明：

$$\mathrm{e}^{-n}\left(1+n+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\right) \to \frac{1}{2}, \quad n \to \infty.$$

证明 (a) $Y_\lambda = \dfrac{1}{\sqrt{\lambda}}X - \sqrt{\lambda}$ 的特征函数

$$\phi_{Y_\lambda} = \mathrm{e}^{-\mathrm{i}\sqrt{\lambda}t}\phi_X\left(\frac{t}{\sqrt{\lambda}}\right) = \exp\left(\lambda\,\mathrm{e}^{\mathrm{i}\frac{t}{\sqrt{\lambda}}} - \lambda - \mathrm{i}\sqrt{\lambda}t\right) \to \mathrm{e}^{-\frac{1}{2}t^2}, \quad \lambda \to +\infty,$$

$\mathrm{e}^{-\frac{1}{2}t^2}$ 是标准正态分布的特征函数, 由 Lévy-Cramér 连续性定理, $Y_\lambda \xrightarrow{\lambda\to+\infty} N(0,1)$.

(b) 由 $X \sim \Gamma(1,\lambda)$ 知 X 的密度函数 $f(x) = \dfrac{1}{\Gamma(\lambda)}x^{\lambda-1}\mathrm{e}^{-x}$, 从而

$$\begin{aligned}
\phi_X(t) &= \frac{1}{\Gamma(\lambda)}\int_0^{+\infty} x^{\lambda-1}\mathrm{e}^{(\mathrm{i}t-1)x}\,\mathrm{d}x \\
&= \frac{1}{(1-\mathrm{i}t)^\lambda\Gamma(\lambda)}\int_0^{+\infty}[(1-\mathrm{i}t)x]^{\lambda-1}\mathrm{e}^{-(1-\mathrm{i}t)x}\,\mathrm{d}(1-\mathrm{i}t)x \\
&= \frac{1}{(1-\mathrm{i}t)^\lambda},
\end{aligned}$$

进而

$$\phi_{Y_\lambda} = \mathrm{e}^{-\mathrm{i}\sqrt{\lambda}t}\phi_X\left(\frac{t}{\sqrt{\lambda}}\right) = \mathrm{e}^{-\mathrm{i}\sqrt{\lambda}t}\left(1-\mathrm{i}\frac{t}{\sqrt{\lambda}}\right)^{-\lambda}.$$

再由

$$\begin{aligned}
\log\phi_{Y_\lambda}(t) &= -\mathrm{i}\sqrt{\lambda}t - \lambda\log\left(1-\mathrm{i}\frac{t}{\sqrt{\lambda}}\right) \\
&= -\mathrm{i}\sqrt{\lambda}t - \lambda\left(-\mathrm{i}\frac{t}{\sqrt{\lambda}} - \frac{1}{2}\left(\mathrm{i}\frac{t}{\sqrt{\lambda}}\right)^2 + o\left(\frac{1}{\lambda}\right)\right)
\end{aligned}$$

$$= -\frac{1}{2}t^2 + o(1), \quad \lambda \to +\infty$$

即知 $\phi_{Y_\lambda}(t) \xrightarrow{\lambda\to+\infty} \mathrm{e}^{-\frac{1}{2}t^2}$. 由 Lévy-Cramér 连续性定理, $Y_\lambda \xrightarrow{\lambda\to+\infty} N(0,1)$.

(c) 设 $X_n \sim P(n)$, 由 (a) 知

$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{n\to\infty} N(0,1),$$

进而

$$\mathbb{P}\left(\frac{X_n - n}{\sqrt{n}} \leqslant 0\right) \to \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = \frac{1}{2}, \quad n \to \infty.$$

再由

$$\mathbb{P}\left(\frac{X_n - n}{\sqrt{n}} \leqslant 0\right) = \mathbb{P}\left(X_n \leqslant n\right) = \mathrm{e}^{-n}\sum_{k=1}^{n}\frac{n^k}{k!}$$

即得证. □

习题 (5.12.39) 利用 Lévy-Cramér 连续性定理证明, 当 $n \to \infty$ 时,

(a) 若 $X_n \sim B\left(n, \dfrac{\lambda}{n}\right)$, 则 X_n 依分布收敛于一个 Poisson 分布随机变量.

(b) 若 Y_n 服从参数为 $p = \dfrac{\lambda}{n}$ 的几何分布, 则 $\dfrac{Y_n}{n}$ 依分布收敛于一个指数分布随机变量.

证明 (a) X_n 的特征函数

$$\begin{aligned}\phi_{X_n}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tX_n}\right] &= \sum_{k=0}^{n}\mathrm{e}^{\mathrm{i}\,tk}\,\mathrm{C}_n^k\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \left(1-\frac{\lambda}{n}+\frac{\lambda}{n}\mathrm{e}^{\mathrm{i}\,t}\right)^n \\ &= \left[1+\frac{\lambda}{n}\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)\right]^n \to \mathrm{e}^{\lambda\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)}, \quad n\to\infty.\end{aligned}$$

注意到 $\mathrm{e}^{\lambda\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)}$ 是 Poisson 分布 $P(\lambda)$ 的特征函数, 由 Lévy-Cramér 连续性定理, 结论得证.

(b) Y_n 的特征函数

$$\phi_{Y_n}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,tY_n}\right] = \sum_{k=1}^{\infty}\mathrm{e}^{\mathrm{i}\,tk}\,p(1-p)^{k-1} = p\,\mathrm{e}^{\mathrm{i}\,t}\sum_{k=1}^{\infty}\left[(1-p)\,\mathrm{e}^{\mathrm{i}\,t}\right]^{k-1} = \frac{p\,\mathrm{e}^{\mathrm{i}\,t}}{1-(1-p)\,\mathrm{e}^{\mathrm{i}\,t}},$$

因此 $\dfrac{Y_n}{n}$ 的特征函数

$$\phi_{\frac{Y_n}{n}}(t) = \phi_{Y_n}\left(\frac{t}{n}\right) = \frac{p\,\mathrm{e}^{\mathrm{i}\,\frac{t}{n}}}{1-(1-p)\,\mathrm{e}^{\mathrm{i}\,\frac{t}{n}}} = \frac{\lambda}{\lambda - n\left(1-\mathrm{e}^{-\mathrm{i}\,\frac{t}{n}}\right)} \to \frac{\lambda}{\lambda-\mathrm{i}\,t}, \quad n\to\infty.$$

注意到 $\dfrac{\lambda}{\lambda-\mathrm{i}\,t}$ 是指数分布 $\mathrm{Exp}(\lambda)$ 的特征函数, 由 Lévy-Cramér 连续性定理, 结论得证. □

习题 (5.12.41) 设随机变量 $\{X_k\}$ 相互独立且同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\dfrac{1}{2}$. 证明:

$$\sqrt{\frac{3}{n^3}}\sum_{k=1}^{n}kX_k\xrightarrow{\mathrm{D}}N(0,1),\quad n\to\infty.$$

证明 设 $Y_k=kX_k$, 则 $\{Y_k\}$ 相互独立, 且 $\mathbb{E}[Y_k]=0,\operatorname{Var}(Y_k)=k^2,\mathbb{E}\left[|Y_k|^3\right]=k^3$. 设 $B_n^2=\sum_{k=1}^{n}(\operatorname{Var}(Y_k))^2=\sum_{k=1}^{n}k^2=\dfrac{n(n+1)(2n+1)}{6}$, 令 $B_n=\sqrt{B_n^2}$, 则

$$\frac{1}{B_n^3}\sum_{k=1}^{n}\mathbb{E}\left[|Y_k|^3\right]=\left[\frac{6}{n(n+1)(2n+1)}\right]^{\frac{3}{2}}\frac{n^2(n+1)^2}{4}\sim\frac{3\sqrt{3}}{4\sqrt{n}}\to 0,\quad n\to\infty,$$

即 $\{Y_k\}$ 满足 3 阶矩条件(T), 进而满足 Lindeberg 条件(L). 由 Lindeberg-Feller 中心极限定理,

$$\sqrt{\frac{3}{n^3}}\sum_{k=1}^{n}kX_k\sim\frac{1}{B_n}\sum_{k=1}^{n}Y_k\xrightarrow{\mathrm{D}}N(0,1),\quad n\to\infty.$$

□

习题 (5.12.42) 设随机变量 $\{X_k\}$ 相互独立且同分布, $X_1\sim N\left(\mu,\sigma^2\right)$. 设 $\overline{X}=\dfrac{1}{n}\sum_{k=1}^{n}X_k$, $Z_k=X_k-\overline{X}$. 求 $\overline{X},Z_1,Z_2,\cdots,Z_n$ 的联合特征函数, 并由此证明 $\overline{X}$ 与 $S^2=\dfrac{1}{n-1}\sum_{k=1}^{n}\left(X_k-\overline{X}\right)^2$ 独立.

证明 设 $\mathbf{Y}=\left(\overline{X},Z_1,Z_2,\cdots,Z_n\right)$, $\mathbf{t}=(t_0,t_1,t_2,\cdots,t_n)\in\mathbb{R}^{n+1}$, $\bar{t}=\dfrac{1}{n}\sum_{k=1}^{n}t_k$, 则 $\mathbf{Y}$ 的特征函数

$$\begin{aligned}\phi_{\mathbf{Y}}(\mathbf{t})&=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,\mathbf{t}\cdot\mathbf{Y}}\right]=\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,t_0\overline{X}}\prod_{k=1}^{n}\mathrm{e}^{\mathrm{i}\,t_k\left(X_k-\overline{X}\right)}\right]=\mathbb{E}\left[\prod_{k=1}^{n}\mathrm{e}^{\mathrm{i}\left(\frac{t_0}{n}+t_k-\bar{t}\right)X_k}\right]\\&\xlongequal{\{X_k\}\text{ 独立}}\prod_{k=1}^{n}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\left(\frac{t_0}{n}+t_k-\bar{t}\right)X_k}\right]=\prod_{k=1}^{n}\mathrm{e}^{\mathrm{i}\,\mu\left(\frac{t_0}{n}+t_k-\bar{t}\right)-\frac{1}{2}\sigma^2\left(\frac{t_0}{n}+t_k-\bar{t}\right)^2}\\&=\exp\left(\mathrm{i}\,\mu t_0-\frac{1}{2}\sigma^2\sum_{k=1}^{n}\left(\frac{t_0}{n}+t_k-\bar{t}\right)^2\right).\end{aligned}$$

而

$$\sum_{k=1}^{n}\left(\frac{t_0}{n}+t_k-\bar{t}\right)^2=\sum_{k=1}^{n}\left(\frac{t_0}{n}\right)^2+\sum_{k=1}^{n}\left(t_k-\bar{t}\right)^2+\frac{2t_0}{n}\sum_{k=1}^{n}\left(t_k-\bar{t}\right)$$

$$= \frac{t_0^2}{n} + \sum_{k=1}^{n} \left(t_k - \bar{t}\right)^2,$$

因此

$$\begin{aligned}
\phi_{\mathbf{Y}}(\mathbf{t}) &= \exp\left(\mathrm{i}\,\mu t_0 - \frac{\sigma^2 t_0^2}{2n} - \frac{\sigma^2}{2}\sum_{k=1}^{n}\left(t_k - \bar{t}\right)^2\right) \\
&= \exp\left(\mathrm{i}\,\mu t_0 - \frac{\sigma^2 t_0^2}{2n}\right)\exp\left(-\frac{\sigma^2}{2}\sum_{k=1}^{n}\left(t_k - \bar{t}\right)^2\right) \\
&= \phi_{\overline{X}}(t_0)\prod_{k=1}^{n} \mathrm{e}^{\mathrm{i}\,\mu\left(t_k-\bar{t}\right) - \frac{1}{2}\sigma^2\left(t_k-\bar{t}\right)^2} \\
&= \phi_{\overline{X}}(t_0)\prod_{k=1}^{n} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\left(t_k-\bar{t}\right)X_k}\right] \\
&= \phi_{\overline{X}}(t_0)\mathbb{E}\left[\exp\left(\mathrm{i}\sum_{k=1}^{n}\left(t_k-\bar{t}\right)\left(Z_k+\overline{X}\right)\right)\right] \\
&= \phi_{\overline{X}}(t_0)\mathbb{E}\left[\exp\left(\mathrm{i}\sum_{k=1}^{n} t_k Z_k\right)\right] \\
&= \phi_{\overline{X}}(t_0)\phi_{Z_1,\cdots,Z_n}\left(t_1,\cdots,t_n\right).
\end{aligned}$$

由此可知 $\overline{X}$ 与 $(Z_1,\cdots,Z_n)$ 独立, 进而 $\overline{X}$ 与 S^2 独立. □

补充题 12 证明：标准正态分布被其矩序列决定.

证明 由 Wallis 公式与 Stirling 公式可知, 当 $k\to\infty$ 时,

$$\frac{1}{k}\left[(2k-1)!!\right]^{\frac{1}{2k}} \sim \frac{\left(\frac{2^k k!}{\sqrt{k\pi}}\right)^{\frac{1}{2k}}}{k} = \frac{\sqrt{2}\left(\frac{k!}{\sqrt{k\pi}}\right)^{\frac{1}{2k}}}{k} \sim \frac{\sqrt{2}\left[\frac{\sqrt{2k\pi}\left(\frac{k}{\mathrm{e}}\right)^k}{\sqrt{k\pi}}\right]^{\frac{1}{2k}}}{k} = \frac{2^{\frac{1}{4k}+\frac{1}{2}}}{\sqrt{k\,\mathrm{e}}} \to 0,$$

这说明标准正态分布的矩序列满足 Riesz 条件(R), 因此标准正态分布被其矩序列决定. □

补充题 13 求半圆律 $\rho(x) = \dfrac{1}{2\pi}\sqrt{4-x^2}$, $x\in[-2,2]$ 的 k 阶矩, 并验证其决定 $\rho(x)$.

解 当 k 为奇数时, $\displaystyle\int_{-2}^{2} x^k\sqrt{4-x^2}\,\mathrm{d}x = 0$. 当 $k=2m$ 为偶数时,

$$\begin{aligned}
\gamma_k \coloneqq \int_{-2}^{2} x^{2m}\sqrt{4-x^2}\,\mathrm{d}x \xlongequal[\theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]]{x=2\sin\theta} &\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (2\sin\theta)^{2m}(2\cos\theta)^2\,\mathrm{d}\theta \\
&= 2^{2m+2}\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m}\theta\,\mathrm{d}\theta - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m+2}\theta\,\mathrm{d}\theta\right)
\end{aligned}$$

$$
\begin{aligned}
&= 2^{2m+2}\pi\left[\frac{(2m-1)!!}{(2m)!!}-\frac{(2m+1)!!}{(2m+2)!!}\right]\\
&= 2^{k+2}\pi\left[\frac{(k-1)!!}{k!!}-\frac{(k+1)!!}{(k+2)!!}\right]\\
&= 2^{k+2}\pi\frac{(k-1)!!}{(k+2)!!}.
\end{aligned}
$$

由 Wallis 公式, 当 $k\to\infty$ 时,

$$
\frac{1}{k}\gamma_{2k}^{\frac{1}{2k}}\sim\frac{2}{k}\left[\frac{(2k-1)!!}{(2k+2)!!}\right]^{\frac{1}{2k}}\sim\frac{2}{k\left[(2k+2)\sqrt{k\pi}\right]^{\frac{1}{2k}}}\sim\frac{2}{k}\to 0.
$$

这说明矩序列 $\{\gamma_k\}$ 满足 Riesz 条件(R), 因此其决定了 $\rho(x)$. □

补充题 14 设非负随机变量 $\{X_k\}$ 相互独立且同分布, $\mathbb{E}[X_1]=1, \mathrm{Var}(X_1)=\sigma^2$, 其中 $\sigma\in(0,+\infty)$, 令 $S_n=\sum\limits_{k=1}^{n}X_k$. 证明:

$$
\frac{2}{\sigma}\left(\sqrt{S_n}-\sqrt{n}\right)\xrightarrow{\mathrm{D}}N(0,1).
$$

证明 注意到 $\frac{2}{\sigma}\left(\sqrt{S_n}-\sqrt{n}\right)=\frac{1}{\sqrt{n}\sigma}\cdot\frac{S_n-n}{\sqrt{S_n}+\sqrt{n}}\cdot 2\sqrt{n}$. 由中心极限定理,

$$
\frac{S_n-n}{\sqrt{n}\sigma}\xrightarrow{\mathrm{D}}N(0,1).
$$

而由强大数律,

$$
\frac{S_n-n}{\sqrt{S_n}+\sqrt{n}}\cdot 2\sqrt{n}=\frac{2}{\sqrt{\frac{S_n}{n}}+1}\xrightarrow{\text{a.s}}\frac{2}{1+1}=1,
$$

进而

$$
\frac{S_n-n}{\sqrt{S_n}+\sqrt{n}}\cdot 2\sqrt{n}\xrightarrow{\mathrm{P}}1.
$$

由 Slutsky 定理 (习题 7.2.5(a)) 即得

$$
\frac{2}{\sigma}\left(\sqrt{S_n}-\sqrt{n}\right)\xrightarrow{\mathrm{D}}N(0,1).
$$

□

补充题 15 设随机变量 X 期望为 μ, 标准差 $\sigma>0$. 证明:

$$
\mathbb{P}(X\geqslant\mu+a)\leqslant\frac{\sigma^2}{\sigma^2+a^2},\quad\forall a>0.
$$

证明 由欲证形式可不妨设 $\mu = 0$. 则由 Cauchy-Schwarz 不等式,

$$\begin{aligned} a &= \mathbb{E}[a - X] = \mathbb{E}\left[(a-X)I_{\{X \geqslant a\}} + (a-X)I_{\{X<a\}}\right] \\ &= \underbrace{\mathbb{E}\left[(a-X)I_{\{X \geqslant a\}}\right]}_{\leqslant 0} + \mathbb{E}\left[(a-X)I_{\{X<a\}}\right] \\ &\leqslant \mathbb{E}\left[(a-X)I_{\{X<a\}}\right] \leqslant \sqrt{\mathbb{E}\left[(a-X)^2\right]\mathbb{P}(X<a)} \\ &= \sqrt{(a^2+\sigma^2)\,\mathbb{P}(X<a)}. \end{aligned}$$

于是

$$\mathbb{P}(X \geqslant \mu + a) = 1 - \mathbb{P}(X < \mu + a) \leqslant 1 - \frac{a^2}{\sigma^2 + a^2} = \frac{\sigma^2}{\sigma^2 + a^2}.$$

□