2026春概率论习题课讲义

李函哲，刘景寒，胡洁洋

2026 年 5 月 28 日

目录

Chapter 1

第一次习题课

§1.1 习题1.1

【习题 1.1】 掷两枚均匀硬币，详细写出概率空间三要素。

证明. $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$，$\mathcal{F} = 2^{\Omega}$，$\mathbb{P}(A) = \dfrac{|A|}{4}$。 □

【习题 1.2】 比较“掷 3 次均匀骰子至少出现 1 个 6 点”和“掷 6 次均匀骰子至少出现 2 个 6 点”的概率大小.

证明. $\mathbb{P}(\text{掷 3 次均匀骰子至少出现 1 个 6 点}) = 1-(\frac{5}{6})^3$，$\mathbb{P}(\text{掷 6 次均匀骰子至少出现 2 个 6 点}) = 1-(\frac{5}{6})^6 - 6\cdot\frac{1}{6}\cdot(\frac{5}{6})^5$。比较知前者更大。 □

【习题 1.3】 设 F 是σ-代数, $A, B \in F$, 证明 $A \cap B$, $A \setminus B \in F$。

证明. 由$A \cap B = (A^c \cup B^c)^c$以及 $A \backslash B = A \cap B^c$即可。 □

【习题 1.4】 证明 Jordan 公式。

证明. 归纳，n=1 时显然成立，那么：

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{k=1}^{n-1}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k\le n-1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})+\mathbb{P}(A_n)\\ &\quad-\sum_{k=1}^{n-1}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k\le n-1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_n)\\ &=\sum_{k=1}^{n-1}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k\le n-1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})+\mathbb{P}(A_n)\\ &\quad+\sum_{k=1}^{n-1}(-1)^{k}\sum_{1\le i_1<\cdots<i_k\le n-1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_n).\\ &=\sum_{k=1}^{n-1}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k<n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})+\mathbb{P}(A_n)\\ &\quad+\sum_{k=2}^{n}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k=n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}).\\ &=RHS\end{aligned}$$

□

【习题 1.5】 证明$|P(A)B)-P(A)P(B)|\le 1/4$,并讨论等号成立的条件.

证明. 事实上是 $Cov(\mathbb{1}_A,\mathbb{1}_B)\le\sqrt{Var(\mathbb{1}_A)}\sqrt{Var(\mathbb{1}_B)}$取等条件是 $Cauchy-schwarz$的条件，不过，我们也可以直接说明：不妨设 $P(A)=x, P(B)=y$.

从而

$$\max\{0,x+y-1\}\le P(A\cap B)\le\min\{x,y\}.$$

只需 $P(A\cap B)$ 取最大值与最小值时均满足不等式即可.

不妨设 $x\le y$，从而：

$$0\le x-xy=x(1-y)\le x(1-x)\le\frac{1}{4}.$$

$P(A\cap B)$ 取最大值时不等式成立.

当 $x+y\le 1$ 时，$\min P(A\cap B)=0$，此时：

$$0<xy\le x(1-x)\le\frac{1}{4}.$$

当 $x+y\ge 1$ 时，$\min P(A\cap B)=x+y-1$，此时：

$$|x+y-1-xy|=(1-x)(1-y).$$

且 $(1-x)+(1-y)\le 1$，由基本不等式，

$$(1-x)(1-y)\le\frac{1}{4}.$$

$P(A\cap B)$ 取最小值时不等式成立.

综上，知原不等式成立，取等当且仅当

$$P(A)=P(B)=\frac{1}{2},\quad P(A\cap B)=\frac{1}{2}\text{ 或 }0.$$

【习题 1.6】 设事件列 $\{A_k\}_{k=1}^{\infty}$ 满足 $\mathbb{P}(A_k)=1,\ \forall k\in\mathbb{N}^*$. 证明

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty}A_k\right)=1.$$

<u>**证明.**</u> 考虑对立事件，由德摩根定律，

$$\bigcap_{k=1}^{\infty}A_k\text{ 的对立事件为 }\bigcup_{k=1}^{\infty}A_k^c.$$

已知 $\mathbb{P}(A_k)=1$，故 $\mathbb{P}(A_k^c)=0$ 对每个 $k\in\mathbb{N}^*$ 成立。由可列可加性，有

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty}A_k^c\right)\le\sum_{k=1}^{\infty}\mathbb{P}(A_k^c)=0.$$

因此 $\mathbb{P}\left(\bigcup_{k=1}^{\infty}A_k^c\right)=0$，从而

$$1\ge\mathbb{P}\left(\bigcap_{k=1}^{\infty}A_k\right)=1-\mathbb{P}\left(\bigcup_{k=1}^{\infty}A_k^c\right)\ge1-0=1.$$

□

□

§1.2

习题1.2

【习题 1.7】 掷 n 次均匀硬币, 事件 A_{ij}记为第 i 次投掷结果和第 j 次投掷结果相同. 证明事件$\{A_{ij},1\le i<j\le n\}$ 两两独立但不相互独立.。

<u>**证明.**</u>

$$\mathbb{P}(A_{ij}\cap A_{kl})=\begin{cases}\mathbb{P}(A_{ij})\mathbb{P}(A_{kl}) & (i,j)\ne(k,l)\text{独立性}\\ \mathbb{P}(X_i=X_j=X_k) & i\ne j=k\ne l,i\ne l\end{cases}$$

而 $\mathbb{P}(X_i=X_j=X_k)=2\times\dfrac{1}{8}=\dfrac{1}{2}\times\dfrac{1}{2}=\mathbb{P}(X_i=X_j)\mathbb{P}(X_j=X_k)$。但是对于多个事件，如 $\mathbb{P}(X_1=X_2,X_2=X_3,X_3=X_1)=\dfrac{1}{4}$， 但是 $\mathbb{P}(X_1=X_2)\mathbb{P}(X_2=X_3)\mathbb{P}(X_3=X_1)=\dfrac{1}{8}$。 □

【习题 1.8】 对素数 p, 取Ω = {1, 2, · · · , p} 上的古典概型 (即选取每一点的概率相同). 若事件 A 与事件 B 相互独立, 则 A 与 B 中至少其一为□或Ω.

<u>**证明.**</u> 如不然，假设A，B均不是空集或者全空间，但是满足 $\dfrac{|A\cap B|}{p}=\dfrac{|A||B|}{p^2}$，记 $|A|=a,|B|=b,|A\cap B|=C$，那么 $pc=ab\Rightarrow p|ab\Rightarrow p|a$或$p|b$。这就是题设条件。 □

【习题 1.9】 现共有 n 个缸, 其中第 r 个缸中含有 r - 1 个红球和 n-r 个蓝色球. 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:(1) 第 2 个球是蓝色球;(2) 在第 1 个球是蓝色球的前提下第 2 个球是蓝色球.

<u>**证明.**</u>

$$\begin{aligned}\mathbb{P}(\text{第 2 个球是蓝色球}) &= \sum_{k=1}^{n}\mathbb{P}(\text{第 2 个球是蓝色球} \mid \text{选择第k个罐子})\mathbb{P}(\text{选择第k个罐子})\\ &= \sum_{k=1}^{n}\frac{n-k}{n(n-1)} = \frac{1}{2}\end{aligned}$$

记 $B_1 = \mathbb{P}(\text{第一个是蓝球}), B_2 = \mathbb{P}(\text{第二个是蓝球})$ 同样地：

$$\mathbb{P}(B_1) = \frac{1}{2} \quad \mathbb{P}(B_1 \cap B_2) = \frac{1}{n}\sum_{r=1}^{n}\frac{(n-r)(n-r-1)}{(n-1)(n-2)} = \frac{1}{3}$$

进而 $\mathbb{P}(B_2|B_1) = \dfrac{2}{3}$。 □

【习题 1.10】 100 名乘客登上一架正好有 100 个座位的飞机, 每名乘客对应一个座位. 第一位乘客先随机选择一个座位坐. 第二位乘客如果自己的座位空着就坐自己的座位, 否则就在其他空余的座位中随机选择一个座位坐. 第三位乘客如果自己的座位空着就坐自己的座位, 否则就在其他空余的座位中随机选择一个座位坐. 这个过程一直持续到所有的 100 名乘客都登机为止. 求最后一名乘客坐自己的座位的概率.

<u>**证明.**</u> 如果1坐到自己的位置，那么n一定能坐到自己的位置。如果不然，假设1坐到k_1位置，那么2到k_1-1都坐到原位。k_1如果坐1位，n仍旧能坐回原位，如果k_1坐k_2位置，那么k_1+1到k_2-1都坐回原位置。此时k_2的选择影响和k_1是相同的。因此，关键的是$X = \{$第一个坐到1或n位置的
而$X=k$的条件下，选择1或者n的概率是相同的。那么：

$$\mathbb{P}(\text{n坐回原位}) = \sum_{k=1}^{100}\mathbb{P}(\text{n坐回原位}|X=k)\mathbb{P}(X=k) = \frac{1}{2}\sum_{k=1}^{100}\mathbb{P}(X=k) = \frac{1}{2}$$

□

§1.3 习题1.3

【习题 1.11】 甲、乙两个乒乓球运动员相互对决比赛, 已知甲的胜率高于乙. 现有两个备选的竞赛规则: “3 局 2 胜制” 或 “5 局 3 胜制”. 试讨论哪一种竞赛规则对甲更有利.

<u>**证明.**</u> 三局两胜下，$\mathbb{P}(\text{甲胜}) = p^2 + 2p^2(1-p)$，五局三胜下$\mathbb{P}(\text{甲胜}) = p^3 + \binom{3}{1}p_1^3p_2 + \binom{4}{2}p_1^3p_2^2 = p^3(10-15p+6p_2)$。进而：

$$\mathbb{P}(\text{五局三胜}) - \mathbb{P}(\text{三局二胜}) = 3p^2(p-1)^2(2p-1)$$

在 $p > \dfrac{1}{2}$时大于0. □

【习题 1.12】 小朋友 ξ 掷 $n+1$ 枚均匀硬币, 小朋友 δ 掷 n 枚均匀硬币. 求 ξ 掷出硬币的正面数比 δ 掷出硬币的正面数多的概率

证明. 记X_1是掷 n 枚硬币正面朝上的数量，X_2是掷 1 枚硬币正面朝上的数量,Y是 δ 掷 n 枚硬币正面朝上的数量，所求事件为A。考虑全概率公式：

$$\begin{aligned}\mathbb{P}(A) &= \frac{1}{2}\mathbb{P}(X_1+X_2>Y|X_2=1)+\frac{1}{2}\mathbb{P}(X_1+X_2>Y|X_2=0)\\ &= \frac{1}{2}\mathbb{P}(X_1>Y-1)+\frac{1}{2}\mathbb{P}(X_1>Y)\\ &= \frac{1}{2}(\mathbb{P}(X_1>Y)+\mathbb{P}(X_1=Y))+\frac{1}{2}\mathbb{P}(X_1>Y)\\ &= \mathbb{P}(X_1>Y)+\frac{1}{2}\mathbb{P}(X_1=Y)\end{aligned}$$

由对称性

$$2\mathbb{P}(A)=2\mathbb{P}(X_1>Y)+\mathbb{P}(X_1=Y)=\mathbb{P}(X_1>Y)+\mathbb{P}(X_1=Y)+\mathbb{P}(X_1<Y)=1$$

进而为 $\frac{1}{2}$。不过，你也可以直接计算：

$$\begin{aligned}P(A) &= \sum_{k=1}^{n+1}\binom{n+1}{k}\left(\frac{1}{2}\right)^{n+1}\cdot\sum_{i=0}^{k-1}\binom{n}{i}\left(\frac{1}{2}\right)^{n}\\ &= 2^{-2n-1}\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}\\ &= 2^{-2n-2}\left(\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}+\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}\right)\\ &= 2^{-2n-2}\left(\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}+\sum_{i=0}^{n}\binom{n}{i}\sum_{k=0}^{n-i}\binom{n+1}{k}\right)\\ &= 2^{-2n-2}\left(\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}+\sum_{j=0}^{n}\binom{n}{n-j}\sum_{k=0}^{j}\binom{n+1}{k}\right)\\ &= 2^{-2n-2}\left(\sum_{k=1}^{n+1}\binom{n+1}{k}\sum_{i=0}^{k-1}\binom{n}{i}+\sum_{j=0}^{n}\binom{n}{j}\sum_{k=0}^{j}\binom{n+1}{k}\right)\\ &= 2^{-2n-2}\cdot 2^{2n+1}\\ &= \frac{1}{2}\end{aligned}$$

倒数第二个等式是因为里面的求和是两人抛硬币的所有等可能结果的数量（前者是ξ比δ多，后者是δ不比ξ少）。

□

【习题 1.13】 甲、乙两人轮流抛掷一枚均匀骰子. 甲先掷, 一直到掷出了 1 点, 交给乙掷, 而到乙掷出了1 点, 再交给甲掷, 并如此一直下去. 求第 n 次抛掷时由甲掷的概率.

证明. 记 a_n为第n次甲抛，b_n为第n次乙抛那么：

$$a_{n+1}=\frac{5}{6}a_n+\frac{1}{6}b_n,\quad b_{n+1}=\frac{1}{6}a_n+\frac{5}{6}b_n,\quad a_n+b_n=1,\quad a_1=1$$

递推公式为：$a_{n+1}=\frac{2}{3}a_n+\frac{1}{6}$.即 $a_n=\frac{1}{2}+\frac{1}{2}(\frac{2}{3})^{n-1}$。 □

§1.4 习题1.4

【习题 1.14】 设 F(x), G(x) 是分布函数.(1) 对 0 ⩽ λ ⩽ 1, 证明 λ F(x) + (1 - λ)G(x), F(x)G(x) 是分布函数.(2) 证明$1-\{1-F(x)\}^n$, $(F(x)-1)e+exp\{1-F(x)\}$ 是分布函数

证明. 只需验证单调性，右连续性以及归一化三条性质即可。特别地$1-\{1-F(x)\}^n$是n个独立同分布随机变量最小次序的分布函数。设最小值设最小值 $m_n=\min(X_1,\ldots,X_n)$，则其分布函数为

$$\mathbb{P}(m_n\leq x)=1-\mathbb{P}(m_n>x)=1-\mathbb{P}(X_1>x,\ldots,X_n>x)=1-[1-F(x)]^n.$$

□

【习题 1.15】 设 X 为随机变量, 令函数 G(x) = P(X < x), 证明 G(x) 在 R 上左连续, 并用 G(x) 来表示事件 {y ⩽ X ⩽ x} 的概率.

证明. $\lim_{n\to\infty}P(X<x-1/n)=P(\bigcup_n\{X<x-1/n\})=P(X<x)$。这里用了概率测度的连续性。$\mathbb{P}(X=x)=G(x+0)-G(x)$。因此$\{y\leq X\leq x\}=\{y\leq X\}-\{X\geq x\}+\{X=x\}$，这就说明 $\mathbb{P}(y\leq X\leq x)=G(x+0)-G(y)$。 □

【习题 1.16】 设 X, Y 是随机变量.(1) 证明 min {X, Y } 和 max {X, Y } 也是随机变量.(2) 证明 |X|, X^2 也是随机变量

证明. 注意到：$\{\min\{X,Y\}>x\}=\{X>x\}\cap\{Y>x\}$，$\{\max\{X,Y\}\leq x\}=\{X\leq x\}\cap\{Y\leq x\}$，$\{|X|\leq x\}=\{-x\leq X\leq x\}$（x<0时为空集），$\{X^2\leq x\}=\{-\sqrt{x}\leq X\leq\sqrt{x}\}$（x<0时为空集）即可。 □

【习题 1.17】 若随机变量 X 的分布函数为 F(x) = A arctan x + B, x ∈ R. 求常数 A, B.

证明. 必要地：$\lim_{x\to+\infty}F(x)=\frac{A\pi}{2}+B=1$，$\lim_{x\to-\infty}F(x)=-\frac{A\pi}{2}+B=0$。解之：$B=\frac{1}{2},A=1/\pi$。 □

§1.5 问题

【习题 1.18】 证明 Bonferroni 不等式：

$$\mathbb{P}\left(\bigcup_{r=1}^{n}A_r\right)\geq\sum_{r=1}^{n}\mathbb{P}(A_r)-\sum_{1\leq r<k\leq n}\mathbb{P}(A_r\cap A_k).$$

证明. 考虑概率是示性函数的期望，即：$\mathbb{E}(\mathbb{1}_A(w))=1\cdot\mathbb{P}(\omega\in A)+0\cdot\mathbb{P}(\omega\notin A)=\mathbb{P}(A)$。因此我们考虑 $f(\omega)=\mathbb{1}_{\cup_{i=1}^n A_i}$，$g(\omega)=\sum_{i=1}^n\mathbb{1}_{A_i}-\sum_{1\leq i<k\leq n}\mathbb{1}_{A_i\cap A_k}$。记 $r=|\{\omega:\omega\in A_i\}|$是给定 ω，在 $\{A_i\}_{i=1}^n$中出现的次数，那么：

- $r=0$时，$f=g=0$
- $r=1$时，$f=g=1$
- $r>1$时，$f=1$，$g=r-\binom{r}{2}$

因此 $f\geq g$。再对它们取期望，就能得到不等式。同样地，用这个方法，还能得到更一般地：当 m 为奇数时：

$$\mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) \leq \sum_{k=1}^{m}(-1)^{k-1}\sum_{1\leq i_1<\cdots<i_k\leq n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}).$$

当 m 为偶数时：

$$\mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) \geq \sum_{k=1}^{m}(-1)^{k-1}\sum_{1\leq i_1<\cdots<i_k\leq n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}).$$

□

【习题 1.19】 对正整数集 $\Omega=\{1,2,\cdots\}$，引入概率测度

$$\mathbb{P}(\{n\})=\frac{1}{\zeta(s)\cdot n^s},\quad \zeta(s)=\sum_{n=1}^{\infty}n^{-s},\quad s>1.$$

对正整数 q, $A_q=\{mq:m\in\Omega\}$。

(1) 对任意不同的素数 $p_1,p_2,\cdots,p_t$，证明 $A_{p_1},A_{p_2},\cdots,A_{p_t}$ 相互独立。

(2) 用概率方法证明欧拉公式：

$$\zeta(s)=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^s}\right)^{-1}.$$

其中所有素数 $p_1<p_2<\cdots$。

(3) 在概率测度 $\mathbb{P}$ 下，从所有正整数中独立地随机选取两数 a,b，证明 a,b 互素的概率为

$$\frac{1}{\zeta(2s)}.$$

注.

$$\sum_{n=1}^{+\infty}n^{-s}$$

收敛保证了概率测度 $\mathbb{P}$ 的良定。

<u>**证明.**</u> 由素数的最大公倍数的性质：

$$\mathbb{P}(A_{p_1}\cap A_{p_2}\cdots\cap A_{p_t})=\mathbb{P}(A_{p_1p_2\cdots p_t})=\frac{1}{\zeta(s)\prod_{i=1}^{t}p_i^s}\sum\frac{1}{n^s}=\prod_{i=1}^{t}\frac{1}{p_i^s}=\prod_{i=1}^{t}\mathbb{P}(A_{p_i})$$

得到 $A_{p_1},A_{p_2},\cdots,A_{p_t}$ 相互独立。由此：

$$\frac{1}{\zeta(s)}=\mathbb{P}(\{1\})=\mathbb{P}(\cap_{i=1}^{\infty}A_{p_i}^c)=\prod_{i=1}^{\infty}(1-\frac{1}{p_i^s})$$

最后，考虑X, Y是独立同上分布的随机变量，那么：

$$\begin{aligned}\mathbb{P}(\text{X,Y互素}) &= 1 - \mathbb{P}(\cup_i(X \in A_{p_i} \cap Y \in A_{p_i})) \\ &= \prod_i \mathbb{P}(X \notin A_{p_i} \cap Y \notin A_{p_i}) \\ &= \prod_i (1 - \frac{1}{p_i^{2s}}) \\ &= \frac{1}{\zeta(2s)}\end{aligned}$$

□

【习题 1.20】 7. 定义实数 m 为分布函数 F 的中位数，若

$$F(m-0) \leq \frac{1}{2} \leq F(m).$$

证明分布函数至少有一个中位数，且中位数集合构成一个闭区间。

证明. 定义： $a := \sup\{x : F(x) \leq \frac{1}{2}\}, b := \inf\{x : F(x) \geq \frac{1}{2}\}$，claim：$a \leq b$且中位数集合 $= [a, b]$。
如果存在 $b < c < a$，那么： $\frac{1}{2} \leq F(c) \leq \frac{1}{2}$。进而 $a = b = c$矛盾。对 $\forall c \in [a, b]$，满足：$F(c) \geq \frac{1}{2} \quad (c \geq a)$以及 $F(c-0) \leq F(c) \leq \frac{1}{2} \quad (c \leq b)$。而对 $c' < a$，如果 c'是中位数，那么 $F(c') \geq \frac{1}{2}$，这说明 $a > c' \geq b$，不成立。同样地： $c'' > b$也不是中位数。综上所述：中位数集合 $= [a, b]$。 □

Chapter 2

第二次习题课

§2.1 习题1.5

【习题 2.1】 哪些函数是密度函数？若是，求 C 及分布函数 $F(x)$. (1) $f(x) = \begin{cases} Cx^{-d}, & x > 1 \\ 0, & x < 1 \end{cases}$.
(2) $f(x) = Ce^{-x-e^{-x}}, -\infty < x < \infty$.

解. (1) 要使 $f(x)$ 成为密度函数，必须满足规范性 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$。

$$\int_1^\infty Cx^{-d}\mathrm{d}x = \lim_{t\to\infty}\left[\frac{C}{1-d}x^{1-d}\right]_1^t$$

为了使广义积分收敛，必须要求 $1-d<0$，即 $d>1$。此时积分值为 $\frac{C}{d-1}$。令 $\frac{C}{d-1}=1$，得到 $C=d-1$。此时，分布函数 $F(x)$ 为：当 $x \le 1$ 时，$F(x)=0$。当 $x>1$ 时，$F(x) = \int_1^x (d-1)t^{-d}\mathrm{d}t = 1-x^{-(d-1)}$。

(2) 验证规范性：

$$\int_{-\infty}^{\infty} Ce^{-x-e^{-x}}\mathrm{d}x$$

令 $u=e^{-x}$，则 $\mathrm{d}u = -e^{-x}\mathrm{d}x$。当 $x\to-\infty$ 时 $u\to\infty$；$x\to\infty$ 时 $u\to 0$。

$$\int_\infty^0 Ce^{-u}(-\mathrm{d}u) = C\int_0^\infty e^{-u}\mathrm{d}u = C$$

要成为密度函数，必有 $C=1$。其分布函数为：

$$F(x) = \int_{-\infty}^x e^{-t-e^{-t}}\mathrm{d}t = \int_{e^{-x}}^\infty e^{-u}\mathrm{d}u = e^{-e^{-x}}, \quad -\infty < x < \infty$$

□

【习题 2.2】 设 U 为 $(0,1)$ 上某概率空间上均匀分布，F 为严格单调递增的分布函数，定义新随机变量 $Y = F^{-1}(U)$, $Y(\omega) := F^{-1}(U(\omega))$，证明 Y 的分布函数为 F.

证明. 已知 $U \sim U(0,1)$，且 F 为严格单调递增的分布函数，因此其反函数 F^{-1} 存在且也是严格单调递增的。求 $Y = F^{-1}(U)$ 的分布函数 $F_Y(y)$：

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F^{-1}(U) \leq y)$$

因为 F 严格单调递增，对不等式两边同时作用 F 函数不改变不等号方向：

$$\mathbb{P}(F^{-1}(U) \leq y) = \mathbb{P}(U \leq F(y))$$

由于 U 在 $(0,1)$ 上均匀分布，且对于任意实数 y，都有 $0 \leq F(y) \leq 1$。根据均匀分布的性质：

$$\mathbb{P}(U \leq F(y)) = F(y)$$

即 $F_Y(y) = F(y)$，得证。 □

【习题 2.3】 设 (X,Y) 是取值整数值的随机向量, 它们的联合分布列为 $f(x,y)$. 证明对 $x,y \in \mathbb{Z}$, 有

$$\begin{aligned} f(x,y) = &\mathbb{P}(X \geq x, Y \leq y) - \mathbb{P}(X \geq x+1, Y \leq y) \\ &- \mathbb{P}(X \geq x, Y \leq y-1) + \mathbb{P}(X \geq x+1, Y \leq y-1). \end{aligned}$$

并求掷 r 次均匀骰子中最小值 $X_{\min}$ 和最大值 $X_{\max}$ 的联合分布列.

解. **第一部分：** 令事件 $A = \{X \geq x, Y \leq y\}$。该事件可以分解为：

$$\{X \geq x, Y \leq y\} = \{X = x, Y \leq y\} \cup \{X \geq x+1, Y \leq y\}$$

这是两个互斥事件，因此：

$$\mathbb{P}(X \geq x, Y \leq y) = \mathbb{P}(X = x, Y \leq y) + \mathbb{P}(X \geq x+1, Y \leq y)$$

整理得：$\mathbb{P}(X = x, Y \leq y) = \mathbb{P}(X \geq x, Y \leq y) - \mathbb{P}(X \geq x+1, Y \leq y)$ 同理，对于 $y-1$，有：

$$\mathbb{P}(X = x, Y \leq y-1) = \mathbb{P}(X \geq x, Y \leq y-1) - \mathbb{P}(X \geq x+1, Y \leq y-1)$$

又因为：

$$\{X = x, Y \leq y\} = \{X = x, Y = y\} \cup \{X = x, Y \leq y-1\}$$

所以：

$$f(x,y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x, Y \leq y) - \mathbb{P}(X = x, Y \leq y-1)$$

将前面推导的两式代入上式，即得证所需等式。

第二部分： 对于掷 r 次骰子，$X_{\min} \geq i$ 且 $X_{\max} \leq j$ 意味着 r 次掷出的点数全部落在区间 $[i,j]$ 内。当 $1 \leq i \leq j \leq 6$ 时，该区间共有 $j-i+1$ 个可能的结果。

$$\mathbb{P}(X_{\min} \geq i, X_{\max} \leq j) = \left(\frac{j-i+1}{6}\right)^r$$

利用刚才证明的公式，求 $f(i,j) = \mathbb{P}(X_{\min} = i, X_{\max} = j)$：当 $1 \leq i < j \leq 6$ 时：

$$f(i,j) = \frac{(j-i+1)^r - 2(j-i)^r + (j-i-1)^r}{6^r}$$

当 $1 \le i = j \le 6$ 时，由于必须每次都掷出 i:

$$f(i,i) = \frac{1}{6^r}$$

其余情况 $f(i,j) = 0$。 □

【习题 2.4】 二元函数

$$F(x,y) = \begin{cases} 1 - e^{-x-y}, & x, y \ge 0 \\ 0, & \text{否则} \end{cases}$$

是否为某随机向量 (X,Y) 的联合分布函数? 若是, 求出 X 和 Y 的分布函数, 若不是请说明理由.

解. 不是。我们可以通过验证联合分布函数的“矩形不等式”性质（即分配给任意矩形区域的概率必须非负）来证明。对于任意 $x_1 < x_2, y_1 < y_2$，必须满足:

$$\mathbb{P}(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \ge 0$$

取 $x_1 = 0, x_2 = 1, y_1 = 0, y_2 = 1$，代入给定的函数:

$$\begin{aligned}
&F(1,1) - F(0,1) - F(1,0) + F(0,0) \\
&= (1-e^{-2}) - (1-e^{-1}) - (1-e^{-1}) + (1-e^{0}) \\
&= 1 - e^{-2} - 1 + e^{-1} - 1 + e^{-1} + 0 \\
&= -1 + 2e^{-1} - e^{-2} = -(1-e^{-1})^2
\end{aligned}$$

因为 $1 - e^{-1} \ne 0$，所以 $-(1-e^{-1})^2 < 0$。由于该区域的概率算出来为负数，违背了概率非负性公理，因此它**不是**任何随机向量的联合分布函数。 □

【习题 2.5】 X_1 和 X_2 是两个独立的随机变量且有相同的分布函数 $F(x)$. 记

$$U = \max\{X_1, X_2\}, \quad V = \min\{X_1, X_2\},$$

(1) 求 U 和 V 的分布函数. (2) 求 (U,V) 的联合分布函数.

解. (1) 求 U 和 V 的分布函数：对于 $U = \max\{X_1, X_2\}$:

$$F_U(u) = \mathbb{P}(U \le u) = \mathbb{P}(X_1 \le u, X_2 \le u)$$

由独立性得:

$$F_U(u) = \mathbb{P}(X_1 \le u)\mathbb{P}(X_2 \le u) = F(u)^2$$

对于 $V = \min\{X_1, X_2\}$:

$$F_V(v) = \mathbb{P}(V \le v) = 1 - \mathbb{P}(V > v) = 1 - \mathbb{P}(X_1 > v, X_2 > v)$$

由独立性得:

$$F_V(v) = 1 - \mathbb{P}(X_1 > v)\mathbb{P}(X_2 > v) = 1 - (1-F(v))^2 = 2F(v) - F(v)^2$$

(2) 求 (U,V) 的联合分布函数：$F_{U,V}(u,v)=\mathbb{P}(U\le u,V\le v)$。由于 $V\le U$ 是必然事件：**情形一**：当 $u\le v$ 时，若 $U\le u$，则必有 $V\le u\le v$ 成立。

$$F_{U,V}(u,v)=\mathbb{P}(U\le u)=F(u)^2$$

情形二：当 $u>v$ 时，使用事件的差集：

$$\mathbb{P}(U\le u,V\le v)=\mathbb{P}(U\le u)-\mathbb{P}(U\le u,V>v)$$

其中，事件 $\{U\le u,V>v\}$ 等价于 $\{v<X_1\le u,v<X_2\le u\}$。由独立性：

$$\mathbb{P}(v<X_1\le u,v<X_2\le u)=[F(u)-F(v)]^2$$

因此：

$$F_{U,V}(u,v)=F(u)^2-(F(u)-F(v))^2=2F(u)F(v)-F(v)^2$$

综上所述，(U,V) 的联合分布函数为：

$$F_{U,V}(u,v)=\begin{cases}F(u)^2, & u\le v\\ 2F(u)F(v)-F(v)^2, & u>v\end{cases}$$

□

§2.2 习题1.6

【习题 2.6】 对 (Borel 可测) 函数 $g,h:\mathbb{R}\to\mathbb{R}$, 若离散型随机变量 X,Y 独立, 不利用定理 1.6.4 而直接证明 $g(X),h(Y)$ 亦独立.

证明. 令 $U=g(X),V=h(Y)$。因为 X,Y 是离散型随机变量，所以 U,V 也是离散型随机变量。对于 U 和 V 能取到的任意一对值 u 和 v，根据定义有：

$$\mathbb{P}(U=u,V=v)=\mathbb{P}(g(X)=u,h(Y)=v)=\mathbb{P}(X\in g^{-1}(u),Y\in h^{-1}(v))$$

其中 $g^{-1}(u)=\{x\mid g(x)=u\}$，$h^{-1}(v)=\{y\mid h(y)=v\}$。由于 X,Y 相互独立，其联合概率满足边缘概率的乘积：

$$\begin{aligned}
\mathbb{P}(X\in g^{-1}(u),Y\in h^{-1}(v)) &= \sum_{x\in g^{-1}(u)}\sum_{y\in h^{-1}(v)}\mathbb{P}(X=x,Y=y)\\
&= \sum_{x\in g^{-1}(u)}\sum_{y\in h^{-1}(v)}\mathbb{P}(X=x)\mathbb{P}(Y=y)\\
&= \left(\sum_{x\in g^{-1}(u)}\mathbb{P}(X=x)\right)\left(\sum_{y\in h^{-1}(v)}\mathbb{P}(Y=y)\right)\\
&= \mathbb{P}(X\in g^{-1}(u))\mathbb{P}(Y\in h^{-1}(v))\\
&= \mathbb{P}(g(X)=u)\mathbb{P}(h(Y)=v)\\
&= \mathbb{P}(U=u)\mathbb{P}(V=v)
\end{aligned}$$

因此，$g(X)$ 与 $h(Y)$ 相互独立。 □

【习题 2.7】 设取值为正整数随机变量 X_1,X_2,X_3 相互独立, 它们的分布列为

$$\mathbb{P}(X_i=x)=(1-p_i)p_i^{x-1}, i=1,2,3.$$

(1) 证明

$$\mathbb{P}(X_1<X_2<X_3)=\frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)}.$$

(2) 求 $\mathbb{P}(X_1\le X_2\le X_3)$.

解. (1) 证明：由于 X_1,X_2,X_3 相互独立，有：

$$\mathbb{P}(X_1<X_2<X_3)=\sum_{x_1=1}^{\infty}\sum_{x_2=x_1+1}^{\infty}\sum_{x_3=x_2+1}^{\infty}\mathbb{P}(X_1=x_1)\mathbb{P}(X_2=x_2)\mathbb{P}(X_3=x_3)$$

先计算最内层的求和：

$$\sum_{x_3=x_2+1}^{\infty}(1-p_3)p_3^{x_3-1}=p_3^{x_2}$$

代入第二层求和：

$$\sum_{x_2=x_1+1}^{\infty}(1-p_2)p_2^{x_2-1}p_3^{x_2}=(1-p_2)p_3\sum_{x_2=x_1+1}^{\infty}(p_2p_3)^{x_2-1}=(1-p_2)p_3\frac{(p_2p_3)^{x_1}}{1-p_2p_3}$$

代入最外层求和：

$$\begin{aligned}\mathbb{P}(X_1<X_2<X_3)&=\sum_{x_1=1}^{\infty}(1-p_1)p_1^{x_1-1}\frac{(1-p_2)p_3(p_2p_3)^{x_1}}{1-p_2p_3}\\&=\frac{(1-p_1)(1-p_2)p_2p_3^2}{1-p_2p_3}\sum_{x_1=1}^{\infty}(p_1p_2p_3)^{x_1-1}\\&=\frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)}\end{aligned}$$

得证。

(2) 求 $\mathbb{P}(X_1\le X_2\le X_3)$：同理，改变求和的下界：

$$\mathbb{P}(X_1\le X_2\le X_3)=\sum_{x_1=1}^{\infty}\sum_{x_2=x_1}^{\infty}\sum_{x_3=x_2}^{\infty}\mathbb{P}(X_1=x_1)\mathbb{P}(X_2=x_2)\mathbb{P}(X_3=x_3)$$

最内层求和：

$$\sum_{x_3=x_2}^{\infty}(1-p_3)p_3^{x_3-1}=p_3^{x_2-1}$$

第二层求和：

$$\sum_{x_2=x_1}^{\infty}(1-p_2)p_2^{x_2-1}p_3^{x_2-1}=(1-p_2)\frac{(p_2p_3)^{x_1-1}}{1-p_2p_3}$$

最外层求和：

$$\begin{aligned}\mathbb{P}(X_1\le X_2\le X_3)&=\sum_{x_1=1}^{\infty}(1-p_1)p_1^{x_1-1}\frac{(1-p_2)(p_2p_3)^{x_1-1}}{1-p_2p_3}\\&=\frac{(1-p_1)(1-p_2)}{1-p_2p_3}\sum_{x_1=1}^{\infty}(p_1p_2p_3)^{x_1-1}\\&=\frac{(1-p_1)(1-p_2)}{(1-p_2p_3)(1-p_1p_2p_3)}\end{aligned}$$

□

【习题 2.8】 设连续型随机变量 X_1,X_2,X_3,X_4,X_5 相互独立且有相同的分布函数 F, 令

$$I=\mathbb{P}(X_1<X_2<X_3<X_4<X_5).$$

证明 I 与 F 无关, 并求 I 的值.

证明. 由于 X_1,X_2,X_3,X_4,X_5 是相互独立且同分布的连续型随机变量，它们取值相等的概率为 0。这 5 个随机变量的大小排列共有 $5!=120$ 种可能，且由对称性可知，每种排列出现的概率是相等的。$X_1<X_2<X_3<X_4<X_5$ 只是这 120 种排列中的特定一种。因此：

$$I=\mathbb{P}(X_1<X_2<X_3<X_4<X_5)=\frac{1}{5!}=\frac{1}{120}$$

显然，该结果是一个常数，与具体的分布函数 F 无关。 □

【习题 2.9】 在线段 $[0,1]$ 上任意投掷 3 个点, 各个点独立且均服从均匀分布, 求: (1) 中间点的分布函数. (2) 最左边点和最右边点的联合密度.

解. 设这三个点的位置分别为 $X_1, X_2, X_3 \sim U(0,1)$，且相互独立。将其排序得到次序统计量 $X_{(1)} \leq X_{(2)} \leq X_{(3)}$。

(1) 中间点 $X_{(2)}$ 的分布函数：对于 $x \in [0,1]$，中间点小于等于 x 等价于这三个点中至少有 2 个点小于等于 x。这等价于进行 3 次独立的伯努利试验，每次成功的概率为 x。

$$F_{(2)}(x) = \mathbb{P}(X_{(2)} \leq x) = \binom{3}{2} x^2(1-x) + \binom{3}{3} x^3 = 3x^2 - 2x^3$$

(当 $x < 0$ 时 $F = 0$；当 $x > 1$ 时 $F = 1$)

(2) 最左边点与最右边点的联合密度：令 $U = X_{(1)}, V = X_{(3)}$。对于 $0 \leq u \leq v \leq 1$:

$$\mathbb{P}(U > u, V \leq v) = \mathbb{P}(\text{三个点均落在 } (u,v] \text{ 内}) = (v-u)^3$$

另一方面，利用联合分布函数：

$$\mathbb{P}(U > u, V \leq v) = \mathbb{P}(V \leq v) - \mathbb{P}(U \leq u, V \leq v) = F_V(v) - F_{U,V}(u,v)$$

因此联合分布函数为 $F_{U,V}(u,v) = F_V(v) - (v-u)^3$。对其求混合偏导数得到联合密度函数：

$$f_{U,V}(u,v) = \frac{\partial^2 F_{U,V}(u,v)}{\partial u \partial v} = \frac{\partial^2}{\partial u \partial v}[-(v-u)^3] = \frac{\partial}{\partial u}[-3(v-u)^2] = 6(v-u)$$

所以，当 $0 \leq u \leq v \leq 1$ 时，联合密度为 $f(u,v) = 6(v-u)$；其余区域为 0。 □

§2.3
补充内容：机器学习基础

1. 问题与动机

在机器学习的回归任务中，我们将特征（Feature）和标签（Label）分别抽象为同一概率空间上的随机变量 X 和 Y。核心目标是寻找一个最优的预测函数 $g(\cdot)$，使得预测值 $g(X)$ 与真实值 Y 之间的差异最小。

数学上，我们通常采用均方误差（Mean Squared Error, MSE）作为风险函数。为保证几何结构良好，假设 Y 具有二阶矩（即 $Y \in L^2$），且预测函数 g 满足 $\mathbb{E}[g(X)^2] < \infty$。寻找最优预测模型即求解以下极值问题：

$$g^* = \arg\min_g \mathbb{E}[(Y - g(X))^2]$$

2. 最优预测的求解

我们将通过代数与几何两种视角证明：在均方误差准则下，最优预测函数正是条件期望，即 $g^*(X) = \mathbb{E}[Y|X]$。

视角一：代数配方法 利用期望的线性性质，我们在均方误差中"加上并减去"条件期望 $\mathbb{E}[Y|X]$，进行配方展开：

$$\begin{aligned}\mathbb{E}[(Y - g(X))^2] &= \mathbb{E}\left[\left((Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - g(X))\right)^2\right] \\ &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))] + \mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2]\end{aligned}$$

根据全期望公式（平滑定理），交叉项恒为 0：

$$\mathbb{E}\left[\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X)) \mid X]\right] = \mathbb{E}\left[(\mathbb{E}[Y|X] - g(X))\underbrace{\mathbb{E}[Y - \mathbb{E}[Y|X] \mid X]}_{=0}\right] = 0$$

极值问题化简为：

$$\mathbb{E}[(Y - g(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2]$$

要使整体均方误差最小，必须让非负的第二项为 0，即最优预测为 $g^*(X) = \mathbb{E}[Y|X]$ (a.s.)。

视角二：几何投影法（Hilbert 空间） 既然同学们正在学习实分析与线性代数，我们可以将概率论问题转化为泛函分析中的几何问题。所有满足二阶矩有限的随机变量构成了一个 Hilbert 空间 $L^2(\Omega, \mathcal{F}, \mathbb{P})$。在该空间中，内积定义为 $\langle X, Y\rangle = \mathbb{E}[XY]$，距离的平方即均方误差 $\|X - Y\|^2 = \mathbb{E}[(X - Y)^2]$。所有形如 $g(X)$ 的可测函数构成了一个闭子空间 $\mathcal{H}_X$。求解 $g^*(X)$ 本质上是在 $\mathcal{H}_X$ 中寻找距离 Y 最近的点 $\hat{Y}$。

第一步：距离最小等价于正交。
记误差向量为 $e = Y - \hat{Y}$。假设存在某个 $V \in \mathcal{H}_X$ 使得 $\langle e, V \rangle \neq 0$。我们构造受微扰的预测 $\hat{Y} + tV \in \mathcal{H}_X$，其距离平方为：

$$f(t) = \|Y - (\hat{Y} + tV)\|^2 = \|e - tV\|^2 = \|e\|^2 - 2t\langle e, V \rangle + t^2\|V\|^2$$

这是一个关于 t 的二次函数，在 $t = 0$ 处的导数为 $f'(0) = -2\langle e, V \rangle \neq 0$。这意味着我们总能沿着梯度的反方向稍微移动一点（取充分小的 t），使得 $f(t) < \|e\|^2$，但这与 $\hat{Y}$ 是最近点相矛盾！因此，误差向量必须与子空间正交：$\langle Y - \hat{Y}, V \rangle = 0$。

第二步：条件期望即为正交投影。
令 $\hat{Y} = \mathbb{E}[Y|X]$，我们验证其误差是否与任意 $V = g(X)$ 正交。同样利用平滑定理：

$$\langle Y - \mathbb{E}[Y|X], g(X) \rangle = \mathbb{E}[(Y - \mathbb{E}[Y|X])g(X)] = \mathbb{E}\left[g(X)(\mathbb{E}[Y|X] - \mathbb{E}[Y|X])\right] = 0$$

这从几何上严格证明了：条件期望 $\mathbb{E}[Y|X]$ 就是 Y 在由 X 生成的信息空间上的正交投影。

3. 从概率到统计：现实世界中的预测

概率论与统计学的核心分界 在前文的推导中，我们得到了理论上的最优预测函数 $g^*(x) = \mathbb{E}[Y \mid X = x]$。请注意，计算条件期望需要完全知晓 X 和 Y 的联合概率分布。这正是**概率论**的典型范式（正向演绎）：假设数据生成的底层物理法则（分布与参数）已知，去研究随机变量的性质和最优解。

然而，在**统计学与机器学习**面临的现实问题中（逆向），真实的联合分布永远是个黑盒。我们拥有的仅仅是从该分布中独立同分布（i.i.d.）采样得到的一组有限数据 $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, ($
我们只能利用这组样本训练出一个估计模型 $\hat{f}_{\mathcal{D}}(x)$，去逼近那个未知的、完美的条件期望 $f(x) = \mathbb{E}[Y \mid X = x]$。

既然训练集 $\mathcal{D}$ 是随机采样的，那么我们训练出来的模型 $\hat{f}_{\mathcal{D}}(x)$ 本质上也是一个**随机变量**。评估一个机器学习模型的好坏，不仅要看它在某一个特定数据集上的表现，更要看它在所有可能抽取的数据集上的“期望表现”。

4. Bias-Variance Decomposition

假设真实的客观规律为 $Y = f(x) + \epsilon$，其中 $f(x) = \mathbb{E}[Y \mid X = x]$ 是真实的最佳预测，ϵ 是不可预测的客观随机噪声，满足 $\mathbb{E}[\epsilon] = 0$ 且 $\mathrm{Var}(\epsilon) = \sigma^2$。

对于给定的测试点 x，我们用基于随机数据集 $\mathcal{D}$ 训练出的模型 $\hat{f}(x)$ 去预测全新的观测值 Y。我们来考察这个预测过程的期望泛化误差（对所有可能的数据集 $\mathcal{D}$ 和噪声 ϵ 求期望）：

$$\mathrm{Err}(x) = \mathbb{E}_{\mathcal{D},\epsilon}\left[(Y - \hat{f}(x))^2\right]$$

为了看清误差的来源，我们使用数学中经典的“加一项减一项”配方技巧。在括号内同时加上并减去模型预测的期望值 $\mathbb{E}_{\mathcal{D}}[\hat{f}(x)]$ 以及真实函数 $f(x)$：

$$\begin{aligned}
\mathrm{Err}(x) &= \mathbb{E}\left[(f(x) + \epsilon - \hat{f}(x))^2\right] \\
&= \mathbb{E}\left[\left((f(x) - \mathbb{E}[\hat{f}(x)]) + (\mathbb{E}[\hat{f}(x)] - \hat{f}(x)) + \epsilon\right)^2\right]
\end{aligned}$$

将其展开为三项平方和与三个交叉项。由于噪声 ϵ 独立于模型 $\hat{f}$，且 $\mathbb{E}[\epsilon] = 0$；同时 $\mathbb{E}[\mathbb{E}[\hat{f}(x)] - \hat{f}(x)] = 0$，所有的交叉乘积项在取期望后全部消掉（同学们可作为课后练习自行验证正交性）。最终，误差被极其优美地分解为三个纯粹的部分：

$$\mathrm{Err}(x) = \underbrace{(f(x) - \mathbb{E}[\hat{f}(x)])^2}_{\text{偏差的平方 (Bias}^2)} + \underbrace{\mathbb{E}\left[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2\right]}_{\text{方差 (Variance)}} + \underbrace{\sigma^2}_{\text{不可约误差 (Irreducible Error)}}$$

数学与物理直觉解读：

- **偏差 (Bias)**：$\mathbb{E}[\hat{f}(x)] - f(x)$。衡量了我们模型的“平均预测”与“真实客观规律”之间的距离。偏差大，说明模型本身表达能力不足（例如用线性方程去拟合非线性曲线），这在机器学习中被称为**欠拟合 (Underfitting)**。
- **方差 (Variance)**：$\mathrm{Var}(\hat{f}(x))$。衡量了模型在不同训练集 $\mathcal{D}$ 之间跳跃的剧烈程度。方差大，说明模型过于敏感，把某次采样中的偶然噪声也当成了规律学了进去，这被称为**过拟合 (Overfitting)**。
- **不可约误差 (σ^2)**：这是客观世界固有的随机性（例如测量仪器的物理精度极限）。无论你设计多么精妙的算法，均方误差都不可能低于这个下界 σ^2。

在机器学习中，如果我们把模型变复杂，偏差会减小，但方差会急剧增大；反之亦然。寻找最优模型的本质，就是在无穷维的函数空间中，寻找使得 $\mathrm{Bias}^2 + \mathrm{Variance}$ 达到极小的平衡点。

Chapter 3

第三次习题课

§3.1 习题2.1

【习题 3.1】(2.1.1) 先掷一个均匀骰子，记下点数后再掷同样个数的均匀硬币，令 X 表示正面朝上的硬币个数，求 X 的分布列。

<u>**证明.**</u> 设掷骰子得到点数N，则X, N独立。 X的所有可能取值为$0, 1, \cdots, 6$，且

$$\begin{aligned}\mathbb{P}(X=x) &= \sum_{n=x}^{6} \mathbb{P}(X=x \mid N=n)\mathbb{P}(N=n) \\ &= \frac{1}{6}\sum_{n=x}^{6}\binom{n}{x}2^{-n}\end{aligned}$$

□

【习题 3.2】(2.1.2) 微信朋友圈单位时间分享的讯息条数服从参数为 λ 的泊松分布，若在相邻时间间隔内新增讯息条数是相互独立的，求在两个单位的间隔时间内发现 k 条讯息的概率。

<u>**证明.**</u> 记两个单位时间发送条数为X。

$$\mathbb{P}(X=k) = \sum_{m+n=k}\frac{\lambda^m}{m!}e^{-\lambda}\frac{\lambda^n}{n!}e^{-\lambda} = \frac{\lambda^k e^{-2\lambda}}{k!}\sum_{m=0}^{k}\binom{k}{m} = \frac{(2\lambda)^k e^{-2\lambda}}{k!}$$

□

【习题 3.3】(2.1.3) 设离散型随机变量 $X_1, \cdots, X_n$ 相互独立且关于 0 对称，即 X_i 与 $-X_i$ 有相同的分布列。证明对任意 $x \in \mathbb{R}$，

$$\mathbb{P}(S_n \geq x) = \mathbb{P}(S_n \leq -x),$$

其中

$$S_n = X_1 + \cdots + X_n.$$

若去掉相互独立这一条件，结论还一定成立吗？请说明理由。

<u>证明.</u>

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &= \sum_{x_1+\cdots+x_n \geq x} \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) \\
&= \sum_{x_1+\cdots+x_n \geq x} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\
&= \sum_{x_1+\cdots+x_n \geq x} \mathbb{P}(X_1 = -x_1) \cdots \mathbb{P}(X_n = -x_n) \\
&= \sum_{x_1+\cdots+x_n \leq -x} \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) \\
&= \mathbb{P}(S_n \leq -x),
\end{aligned}$$

但若不独立，令$n = 2$，取以下分布，

$$\mathbb{P}\big((X_1, X_2) = (-1, 0)\big) = \mathbb{P}\big((X_1, X_2) = (0, -1)\big) = \mathbb{P}\big((X_1, X_2) = (1, 1)\big) = \frac{1}{3}.$$

则 X_1, X_2 都服从 $\{-1, 0, 1\}$ 上的均匀分布，因而都关于 0 对称，但

$$\mathbb{P}(S_2 \geq 2) = \frac{1}{3} \neq 0 = \mathbb{P}(S_2 \leq -2).$$

□

【习题 3.4】 **(2.1.4)** 随机变量 X 的分布列 $\mathbb{P}(X = x_k) = p_k, k = 1, 2, \cdots, n$, Shannon 信息熵定义为

$$H(X) = -\sum_{k=1}^{n} p_k \ln p_k.$$

给定 n，X 服从什么样的分布时信息熵 $H(X)$ 最大？

<u>证明.</u> 由Jensen不等式，

$$H(X) = -\sum_{k=1}^{n} p_k \log p_k \leq \log n,$$

等号成立时，$p_1 = p_2 = \cdots = p_n = \frac{1}{n}$，即为离散型均匀分布。 □

§3.2 习题2.2

【习题 3.5】 **(2.2.1)** 对 $X \sim B(n, p)$，求 $\mathbb{E}[X^3]$。

<u>证明.</u> 设$X = \sum_{k=1}^{n} I_k$, $I_1, \cdots, I_n$ i.i.d.是示性随机变量，且$\mathbb{P}(I_k = 1) = p$，则

$$\begin{aligned}
\mathbb{E}(X^3) = \mathbb{E}&\left(\left(\sum_{k=1}^{n} I_k\right)^3\right) \\
&= n\mathbb{E}(I_1^3) + 6\binom{n}{2}\mathbb{E}(I_1^2 I_2) + 6\binom{n}{3}\mathbb{E}(I_1 I_2 I_3) \\
&= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np
\end{aligned}$$

□

【习题 3.6】 (2.2.2) 离散型随机变量 X 的分布列

$$f(x)=\begin{cases}\dfrac{1}{x(x+1)}, & x=1,2,\cdots,\\ 0, & \text{否则}.\end{cases}$$

讨论实数 α 取何值时 α 阶矩（当 α 非整数时称为分数阶矩）$\mathbb{E}[X^\alpha]<+\infty$?

<u>证明.</u> 当$\alpha<1$时，

$$\mathbb{E}[X^\alpha]=\sum_{k=1}^{\infty}\frac{1}{k^{1-\alpha}(k+1)}\le\sum_{k=1}^{\infty}\frac{1}{k^{2-\alpha}}<+\infty;$$

当$\alpha\ge 1$时，

$$\mathbb{E}[X^\alpha]\ge\sum_{k=1}^{\infty}\frac{1}{k+1}=+\infty.$$

故取值范围为$\alpha<1$。 □

【习题 3.7】 (2.2.3) 无人驾驶网约车是当今社会的科技结晶，设一辆无人驾驶网约车一天内穿过的路口总数为 X，且

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

每个路口的红绿灯是独立工作的，该车在每个路口遇到红灯的概率是 p。

(1) 求此出租车穿过路口总数的期望和方差。

(2) 求此出租车一天内遇到红灯数的期望。

<u>证明.</u> 记 $q=1-p$，则由

$$\sum_{k=0}^{\infty}q^k=\frac{1}{1-q}$$

求导得

$$\sum_{k=1}^{\infty}kq^{k-1}=\frac{1}{(1-q)^2},\qquad\sum_{k=2}^{\infty}k(k-1)q^{k-2}=\frac{2}{(1-q)^3}.$$

于是

$$\mathbb{E}[X]=\sum_{k=1}^{\infty}kpq^{k-1}=p\cdot\frac{1}{(1-q)^2}=\frac{1}{p}.$$

又

$$\mathbb{E}[X(X-1)]=\sum_{k=2}^{\infty}k(k-1)pq^{k-1}=pq\sum_{k=2}^{\infty}k(k-1)q^{k-2}=\frac{2q}{p^2}.$$

故

$$\mathbb{E}[X^2]=\mathbb{E}[X(X-1)]+\mathbb{E}[X]=\frac{2q}{p^2}+\frac{1}{p}=\frac{2-p}{p^2},$$

从而

$$\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}.$$

设一天内遇到红灯数为 Y，则在 $X=n$ 的条件下，$Y \mid X=n \sim B(n,p)$，故

$$\mathbb{E}[Y \mid X=n]=np.$$

于是

$$\mathbb{E}[Y]=\mathbb{E}\big(\mathbb{E}[Y \mid X]\big)=p\mathbb{E}[X]=1.$$

□

【注 3.8】 上面的 X 实际上服从参数为 p 的几何分布。有结论（可直接引用）

$$\mathbb{E}[X]=\frac{1}{p}, \qquad \operatorname{Var}(X)=\frac{1-p}{p^2}.$$

【习题 3.9】 **(2.2.4)** 对非负整值随机变量 X，证明

$$\mathbb{E}[X]=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

证明. 由 $X \geq 0$ 且取整数值，

$$X=\sum_{n=0}^{\infty}\mathbf{1}_{\{X>n\}}.$$

两边取期望，并交换求和与期望，得

$$\mathbb{E}[X]=\sum_{n=0}^{\infty}\mathbb{E}\left[\mathbf{1}_{\{X>n\}}\right]=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

□

【注 3.10】 对非负实值随机变量 X，由

$$\lfloor X\rfloor \leq X \leq \lceil X\rceil$$

并对 $\lfloor X\rfloor, \lceil X\rceil$ 分别应用上题，可得

$$\mathbb{E}[\lfloor X\rfloor]=\sum_{n=0}^{\infty}\mathbb{P}(\lfloor X\rfloor>n)=\sum_{n=1}^{\infty}\mathbb{P}(X \geq n),$$

$$\mathbb{E}[\lceil X\rceil]=\sum_{n=0}^{\infty}\mathbb{P}(\lceil X\rceil>n)=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

故

$$\sum_{n=1}^{\infty}\mathbb{P}(X \geq n) \leq \mathbb{E}[X] \leq \sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

这个不等式非常重要，它给出了随机变量数学期望的一个简单而且比较紧的估计。

【习题 3.11】 **(2.2.6)** 随机图模型 $G(n,p)$ 指 n 个顶点 $V=\{1,2,\cdots,n\}$ 的图，两个顶点以概率 p 连边，且每两个顶点是否连边相互独立。顶点 i 的度 D_i 定义为与 i 相连的边数。

(1) 求 D_i 的分布列与期望 $\mathbb{E}[D_i]$。

(2) 若 X 表示 $G(n,p)$ 中“三角形”个数，试求“三角形”期望数 $\mathbb{E}[X]$ 和方差 $\operatorname{Var}(X)$。

证明. 对固定顶点 i，D_i 是其余 $n-1$ 条边中出现的条数，故

$$D_i \sim B(n-1,p), \qquad \mathbb{E}[D_i] = (n-1)p.$$

记所有三角形的集合为 $\mathcal{T}$，对每个 $T \in \mathcal{T}$，记 I_T 为“三角形 T 出现”的示性函数，则

$$X = \sum_{T\in\mathcal{T}} I_T, \qquad |\mathcal{T}| = \binom{n}{3}.$$

于是

$$\mathbb{E}[X] = \sum_{T\in\mathcal{T}} \mathbb{E}[I_T] = \binom{n}{3}p^3.$$

再看方差，

$$\mathrm{Var}(X) = \sum_{T\in\mathcal{T}} \mathrm{Var}(I_T) + 2\sum_{T<S} \mathrm{Cov}(I_T, I_S).$$

其中

$$\mathrm{Var}(I_T) = p^3(1-p^3).$$

若两个不同三角形没有公共边，则对应边集独立，协方差为 0；若共用两条边时，

$$\mathbb{E}[I_T I_S] = p^5, \qquad \mathrm{Cov}(I_T, I_S) = p^5 - p^6 = p^5(1-p).$$

而共用一条边的三角形对数为

$$\binom{n}{2}\binom{n-2}{2} = 6\binom{n}{4}.$$

故

$$\mathrm{Var}(X) = \binom{n}{3}p^3(1-p^3) + 12\binom{n}{4}p^5(1-p).$$

□

§3.3 习题2.3

【习题 3.12】**(2.3.1)** Daniel Bernoulli 在 1769 年描述了“扩散模型”：A 瓶有 n 个红球，B 瓶有 n 个蓝球，每次从两瓶中各选一个球并相互交换。求进行 k 次操作后 A 瓶中的红球数的期望。

证明. 对于A瓶初始的每一个球，先求第k次交换后仍在A瓶的概率p_k，则$p_0 = 1$，且

$$p_{k+1} = \frac{n-1}{n}p_k + \frac{1}{n}(1-p_k),$$

得

$$p_k = \left(\left(\frac{n-2}{n}\right)^k + 1\right)\cdot\frac{1}{2},$$

因此对A瓶原来的n个球编号，第i个球第k次交换后在A瓶个数为I_i，则$N = I_1 + I_2 + \cdots + I_n$，

$$\mathbb{E}(N) = \sum_{i=1}^{n} \mathbb{E}(I_i) = np_k = \frac{n}{2}\left(\left(\frac{n-2}{n}\right)^k + 1\right).$$

□

【习题 3.13】 (2.3.2) 设 $G = (V, E)$ 是有限图，其中 V 是 G 的顶点集，E 是 G 的边集。对任意顶点集 W 和任一边 $e \in E$，定义示性函数

$$\mathbf{1}_W(e) = \begin{cases} 1, & e \text{ 连接 } W \text{ 和 } W^c, \\ 0, & \text{否则}. \end{cases}$$

设

$$N_W = \sum_{e \in E} \mathbf{1}_W(e).$$

利用概率方法证明存在 $W \subset V$ 使得 $N_W \geq |E|/2$。

证明. 我们独立地取V的每个点，取每个点概率为$\frac{1}{2}$，记取出的顶点集W，随机变量$N = N_W$，则

$$\mathbb{E}(N_W) = \sum_{e \in E} \mathbb{E}(I_W(e)) = |E| \cdot 2\frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{|E|}{2},$$

故存在一种取法使得$N_W \geq \frac{|E|}{2}$。 □

【习题 3.14】 (2.3.3) 一个盒子里有标号为 $1, 2, \cdots, n$ 的 n 个球。现从中不放回地随机取出 k 个球并把它们的标号相加得到和数。求该和数的期望和方差。

证明. 记和数为 X，对每个 $i = 1, 2, \cdots, n$，记

$$I_i = \mathbf{1}_{\{\text{第 } i \text{ 号球被取到}\}},$$

则

$$X = \sum_{i=1}^{n} i I_i.$$

由于每个球被取到的概率都为$\frac{k}{n}$，故

$$\mathbb{E}(X) = \sum_{i=1}^{n} i\mathbb{E}(I_i) = \frac{k}{n}\sum_{i=1}^{n} i = \frac{k(n+1)}{2}.$$

再求方差. 由

$$\mathrm{Var}(X) = \sum_{i=1}^{n} i^2 \,\mathrm{Var}(I_i) + 2\sum_{1 \leq i < j \leq n} ij \,\mathrm{Cov}(I_i, I_j),$$

先有

$$\mathrm{Var}(I_i) = \frac{k}{n}\left(1 - \frac{k}{n}\right) = \frac{k(n-k)}{n^2}.$$

对$i \neq j$，有

$$\mathbb{P}(I_i = 1, I_j = 1) = \frac{\binom{n-2}{k-2}}{\binom{n}{k}} = \frac{k(k-1)}{n(n-1)},$$

故

$$\mathrm{Cov}(I_i, I_j) = \frac{k(k-1)}{n(n-1)} - \frac{k^2}{n^2} = -\frac{k(n-k)}{n^2(n-1)}.$$

于是

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{k(n-k)}{n^2}\sum_{i=1}^{n} i^2 - \frac{2k(n-k)}{n^2(n-1)}\sum_{1\le i<j\le n} ij \\
&= \frac{k(n-k)}{n^2}\cdot\frac{n(n+1)(2n+1)}{6} - \frac{k(n-k)}{n^2(n-1)}\left[\left(\sum_{i=1}^{n} i\right)^2 - \sum_{i=1}^{n} i^2\right] \\
&= \frac{k(n-k)}{n^2}\cdot\frac{n(n+1)(2n+1)}{6} - \frac{k(n-k)}{n^2(n-1)}\left[\frac{n^2(n+1)^2}{4} - \frac{n(n+1)(2n+1)}{6}\right] \\
&= \frac{k(n-k)(n+1)}{12}.
\end{aligned}$$

□

【习题 3.15】**(2.3.6)** 设 n 个向量 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n \in \mathbb{R}^n$ 满足 $|\mathbf{v}_i| \le 1$, $i = 1, 2, \cdots, n$。令

$$\mathbf{w} = \sum_{i=1}^{n} p_i \mathbf{v}_i, \quad p_i \in [0,1].$$

利用概率方法证明存在 $\varepsilon_i \in \{0,1\}$，使得

$$\left|\sum_{i=1}^{n} \varepsilon_i \mathbf{v}_i - \mathbf{w}\right| \le \frac{\sqrt{n}}{2}.$$

证明. 独立地取$\varepsilon_i \in \{0,1\}$其中$\varepsilon_i$取1的概率为$p_i$, 考虑随机变量

$$X := \left|\sum_{i=1}^{n} \varepsilon_i v_i - w\right|^2 = \sum_{i=1}^{n} (\varepsilon_i - p_i)^2 |v_i|^2 + 2\sum_{1\le i<j\le n} (\varepsilon_i - p_i)(\varepsilon_j - p_j) v_i \cdot v_j,$$

有

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[(\varepsilon_i - p_i)^2]|v_i|^2 + 2\sum_{1\le i<j\le n} \mathbb{E}[(\varepsilon_i - p_i)(\varepsilon_j - p_j)] v_i \cdot v_j \\
&= \sum_{i=1}^{n} \mathbb{E}[(\varepsilon_i - p_i)^2]|v_i|^2 + 2\sum_{1\le i<j\le n} \mathbb{E}(\varepsilon_i - p_i)\mathbb{E}(\varepsilon_j - p_j) v_i \cdot v_j \\
&= \sum_{i=1}^{n} \mathbb{E}[(\varepsilon_i - p_i)^2]|v_i|^2 \\
&= \sum_{i=1}^{n} p_i(1-p_i)|v_i|^2 \\
&\le \frac{n}{4},
\end{aligned}$$

故存在一种选取方式，使得$X \le \frac{n}{4}$，即

$$\left|\sum_{i=1}^{n} \varepsilon_i v_i - w\right| \le \frac{\sqrt{n}}{2}.$$

□

§3.4 习题2.4

【习题 3.16】**(2.4.1)** 证明条件期望的如下性质：

(1) $\mathbb{E}[aY+bZ \mid X] = a\mathbb{E}[Y \mid X] + b\mathbb{E}[Z \mid X]$, $\forall a, b \in \mathbb{R}$。

(2) 若 $Y \geq 0$，则 $\mathbb{E}[Y \mid X] \geq 0$。

(3) $\mathbb{E}[1 \mid X] = 1$。

(4) 若 X 与 Y 相互独立，则 $\mathbb{E}[Y \mid X] = \mathbb{E}[Y]$。

(5) $\mathbb{E}[Yg(X) \mid X] = g(X)\mathbb{E}[Y \mid X]$，其中函数 g 使得等式两边的表达式均有意义。

<u>**证明.**</u> 对任意满足 $\mathbb{P}(X = x) > 0$ 的 x，按定义

$$\mathbb{E}[Y \mid X = x] = \sum_y y\mathbb{P}(Y = y \mid X = x).$$

只需对每个这样的x分别证明。

(1)

$$\begin{aligned}\mathbb{E}[aY+bZ \mid X=x] &= \sum_{y,z}(ay+bz)\mathbb{P}(Y=y, Z=z \mid X=x) \\ &= a\sum_{y,z} y\mathbb{P}(Y=y, Z=z \mid X=x) + b\sum_{y,z} z\mathbb{P}(Y=y, Z=z \mid X=x) \\ &= a\mathbb{E}[Y \mid X=x] + b\mathbb{E}[Z \mid X=x].\end{aligned}$$

(2) 若 $Y \geq 0$，则

$$\mathbb{E}[Y \mid X = x] = \sum_y y\mathbb{P}(Y = y \mid X = x) \geq 0.$$

(3)

$$\mathbb{E}[1 \mid X = x] = 1.$$

(4) 若 X 与 Y 相互独立，则

$$\mathbb{P}(Y = y \mid X = x) = \mathbb{P}(Y = y),$$

故

$$\mathbb{E}[Y \mid X = x] = \sum_y y\mathbb{P}(Y = y) = \mathbb{E}[Y].$$

(5) 由于在条件 $X = x$ 下，$g(X) = g(x)$ 是常数，故

$$\mathbb{E}[Yg(X) \mid X = x] = \mathbb{E}[Yg(x) \mid X = x] = g(x)\mathbb{E}[Y \mid X = x].$$

上面各式对任意 x 都成立，故结论成立。 □

【习题 3.17】**(2.4.2)** 设 X 和 Y 相互独立，分别服从参数为 λ_1 和 λ_2 的 Poisson 分布。求条件期望 $\mathbb{E}[X \mid X+Y]$。

证明. 记 $S=X+Y$，则对 $0\le k\le n$，

$$\begin{aligned}\mathbb{P}(X=k\mid S=n)&=\frac{\mathbb{P}(X=k,Y=n-k)}{\mathbb{P}(S=n)}\\&=\frac{\dfrac{\lambda_1^k e^{-\lambda_1}}{k!}\dfrac{\lambda_2^{n-k}e^{-\lambda_2}}{(n-k)!}}{\dfrac{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}}\\&=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.\end{aligned}$$

故在 $S=n$ 条件下，X 服从参数为 $\left(n,\dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)$ 的二项分布，从而

$$\mathbb{E}[X\mid S=n]=n\frac{\lambda_1}{\lambda_1+\lambda_2}.$$

即

$$\mathbb{E}[X\mid X+Y]=\frac{\lambda_1}{\lambda_1+\lambda_2}(X+Y).$$

□

【习题 3.18】 **(2.4.3)** 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

$$\mathbb{E}[\max\{X^2,Y^2\}]\le 1+\sqrt{1-\rho^2}.$$

证明. 注意到

$$\max\{X^2,Y^2\}=\frac{X^2+Y^2+|X^2-Y^2|}{2}=\frac{X^2+Y^2+|X-Y||X+Y|}{2},$$

故

$$\mathbb{E}[\max\{X^2,Y^2\}]=1+\frac{1}{2}\mathbb{E}[|X-Y||X+Y|].$$

由 Cauchy 不等式，

$$\mathbb{E}[|X-Y||X+Y|]\le\sqrt{\mathbb{E}[(X-Y)^2]\mathbb{E}[(X+Y)^2]}.$$

又

$$\begin{aligned}\mathbb{E}[(X-Y)^2]&=\mathrm{Var}(X-Y)=1+1-2\rho=2(1-\rho),\\\mathbb{E}[(X+Y)^2]&=\mathrm{Var}(X+Y)=1+1+2\rho=2(1+\rho).\end{aligned}$$

于是

$$\mathbb{E}[|X-Y||X+Y|]\le\sqrt{2(1-\rho)\cdot 2(1+\rho)}=2\sqrt{1-\rho^2},$$

从而

$$\mathbb{E}[\max\{X^2,Y^2\}]\le 1+\sqrt{1-\rho^2}.$$

□

【习题 3.19】 **(2.4.5)** 通常定义 Y 关于 X 的条件方差 $\mathrm{Var}(Y \mid X)$ 为条件分布 $Y \mid X$ 的方差，由常用公式

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2,$$

我们也可直接定义，

$$\mathrm{Var}(Y \mid X) = \mathbb{E}[Y^2 \mid X] - \mathbb{E}[Y \mid X]^2.$$

根据上述定义，证明

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X]).$$

证明. 由定义，

$$\mathbb{E}[\mathrm{Var}(Y \mid X)] = \mathbb{E}[\mathbb{E}[Y^2 \mid X]] - \mathbb{E}[\mathbb{E}[Y \mid X]^2].$$

而

$$\mathbb{E}[\mathbb{E}[Y^2 \mid X]] = \mathbb{E}[Y^2], \qquad \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y].$$

故

$$\mathbb{E}[\mathrm{Var}(Y \mid X)] = \mathbb{E}[Y^2] - \mathbb{E}[\mathbb{E}[Y \mid X]^2].$$

另一方面，

$$\mathrm{Var}(\mathbb{E}[Y \mid X]) = \mathbb{E}[\mathbb{E}[Y \mid X]^2] - \mathbb{E}[\mathbb{E}[Y \mid X]]^2 = \mathbb{E}[\mathbb{E}[Y \mid X]^2] - \mathbb{E}[Y]^2.$$

两式相加即得

$$\mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X]) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathrm{Var}(Y).$$

□

【习题 3.20】 **(2.4.8)** 2024 年诺贝尔物理学奖授予 Hopfield 和 Hinton，表彰他们利用人工神经网络进行机器学习的基础性发现和发明。Hinton 在 Hopfield 网络想法基础上引入了玻尔兹曼机：给定连接两点间权重 $w_{ij} = w_{ji}$, $w_{ii} = 0$，定义取值于 $\{0,1\}^n$ 的 n 维随机向量

$$X = (X_1, \cdots, X_n)$$

的联合概率

$$\mathbb{P}(X = x) = \frac{1}{Z_n} \exp\left\{ \sum_{1 \le i < j \le n} w_{ij} x_i x_j + \sum_{1 \le i \le n} b_i x_i \right\},$$

这里配分函数为

$$Z_n = \sum_{x \in \{0,1\}^n} \exp\left\{ \sum_{1 \le i < j \le n} w_{ij} x_i x_j + \sum_{1 \le i \le n} b_i x_i \right\}.$$

$X^{(k)}$ 表示 X 去掉第 k 个分量后的向量，试证明条件期望

$$\mathbb{E}[X_k \mid X^{(k)}] = \frac{\exp\left\{ b_k + \sum_{i \ne k} w_{ki} X_i \right\}}{1 + \exp\left\{ b_k + \sum_{i \ne k} w_{ki} X_i \right\}}.$$

<u>**证明.**</u> 对任意给定的 $x^{(k)}$，记

$$\eta = b_k + \sum_{i \neq k} w_{ki} x_i.$$

当 $X^{(k)} = x^{(k)}$ 固定时，联合概率中与 x_k 无关的部分可并入常数 C，从而

$$\mathbb{P}(X_k = x_k, X^{(k)} = x^{(k)}) = C \exp\{x_k \eta\}, \qquad x_k = 0, 1.$$

于是

$$\mathbb{P}(X_k = 1 \mid X^{(k)} = x^{(k)}) = \frac{Ce^{\eta}}{C + Ce^{\eta}} = \frac{e^{\eta}}{1 + e^{\eta}}.$$

又因为 X_k 只取 $0, 1$ 两个值，故

$$\mathbb{E}[X_k \mid X^{(k)} = x^{(k)}] = \mathbb{P}(X_k = 1 \mid X^{(k)} = x^{(k)}) = \frac{e^{\eta}}{1 + e^{\eta}}.$$

即

$$\mathbb{E}[X_k \mid X^{(k)}] = \frac{\exp\left\{b_k + \sum_{i \neq k} w_{ki} X_i\right\}}{1 + \exp\left\{b_k + \sum_{i \neq k} w_{ki} X_i\right\}}.$$

□

§3.5 习题2.5

【习题 3.21】 (2.5.1) 直线上简单随机游走

$$S_n = \sum_{k=1}^{n} X_k, \quad S_0 = 0,$$

这里

$$P(X_i = 1) = p, \quad P(X_i = -1) = 1 - p, \quad 0 < p < 1.$$

求 $E(S_n)$, $Var(S_n)$, $Cov(S_m, S_n)$, $E[S_n \mid S_m]$.

<u>**证明.**</u> 先记

$$E(X_1) = p - (1 - p) = 2p - 1, \qquad \mathrm{Var}(X_1) = 1 - (2p - 1)^2 = 4p(1 - p).$$

因此

$$E(S_n) = \sum_{k=1}^{n} E(X_k) = n(2p - 1),$$

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k) = 4np(1 - p).$$

又由独立性，

$$\mathrm{Cov}(S_m, S_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{k=1}^{m \wedge n} \mathrm{Var}(X_k) = 4p(1 - p)(m \wedge n).$$

最后求条件期望。

若 $n \geq m$，则

$$S_n = S_m + \sum_{k=m+1}^{n} X_k,$$

而后面这段与 S_m 独立，故

$$E[S_n \mid S_m] = S_m + (n-m)(2p-1).$$

若 $n \leq m$，则在给定 S_m 的条件下，$X_1, \ldots, X_m$ 的地位完全对称，故

$$E[X_1 \mid S_m] = \cdots = E[X_m \mid S_m].$$

又

$$\sum_{k=1}^{m} E[X_k \mid S_m] = E\left[\sum_{k=1}^{m} X_k \mid S_m\right] = E[S_m \mid S_m] = S_m,$$

所以

$$E[X_k \mid S_m] = \frac{S_m}{m}, \qquad 1 \leq k \leq m.$$

从而

$$E[S_n \mid S_m] = \sum_{k=1}^{n} E[X_k \mid S_m] = \frac{n}{m} S_m.$$

综上，

$$E[S_n \mid S_m] = \begin{cases} \dfrac{n}{m} S_m, & n \leq m, \\ S_m + (n-m)(2p-1), & n \geq m. \end{cases}$$

□

【习题 3.22】 **(2.5.2)** 在一次只有两个候选人的选举中，每次投票只投给一位候选人且不能弃票。已知最后投票结果 A 得 α 张选票，B 得 β 张选票，且 $\alpha \geq \beta$，投票过程中出现的各种情况可能性相同。

1. 求计票过程中出现两人票数相等的概率。
2. 证明计票过程中 A 从不落后于 B 的概率为

$$\frac{\alpha - \beta + 1}{\alpha + 1}.$$

证明. 仿照课本例2.5.4，构造随机游走，令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 票给 } A, \\ -1, & \text{第 } i \text{ 票给 } B, \end{cases} \qquad S_k = \sum_{i=1}^{k} X_i.$$

则前 k 张票计完后，A 比 B 多出的票数就是 S_k。每一种计票次序都对应一条从 $(0,0)$ 到 $(\alpha+\beta, \alpha-\beta)$ 的轨道，且这些轨道等可能；轨道总数为

$$N_{\alpha+\beta}(0, \alpha-\beta) = \binom{\alpha+\beta}{\alpha}.$$

(1) “计票过程中出现两人票数相等”就是轨道在出发后再次经过 x 轴。

若 $\alpha=\beta$，则终点就在 x 轴上，故所求概率为 1。

若 $\alpha>\beta$，则其对立事件是“轨道不再过 x 轴”，也就是计票过程中 A 始终领先于 B。由投票定理，

$$\#\{\text{从 } (0,0) \text{ 到 } (\alpha+\beta,\alpha-\beta) \text{ 不再过 } x \text{ 轴的轨道}\}=\frac{\alpha-\beta}{\alpha+\beta}N_{\alpha+\beta}(0,\alpha-\beta).$$

因此

$$P(\text{出现票数相等})=1-\frac{\alpha-\beta}{\alpha+\beta}=\frac{2\beta}{\alpha+\beta}.$$

当 $\alpha=\beta$ 时，此式也仍为 1。

(2) “A 从不落后于 B”就是对一切 k 都有 $S_k\geq 0$。

把每一条这样的轨道最前面补上一条向上的边，就得到一条从 $(0,0)$ 到 $(\alpha+\beta+1,\alpha-\beta+1)$ 且不再过 x 轴的轨道；反过来，删去第一步也可恢复原轨道，所以这是一个一一对应。

于是由投票定理，

$$\#\{A \text{ 从不落后于 } B \text{ 的轨道}\}=\frac{\alpha-\beta+1}{\alpha+\beta+1}N_{\alpha+\beta+1}(0,\alpha-\beta+1).$$

又

$$N_{\alpha+\beta+1}(0,\alpha-\beta+1)=\binom{\alpha+\beta+1}{\alpha+1},$$

故

$$\#\{A \text{ 从不落后于 } B \text{ 的轨道}\}=\frac{\alpha-\beta+1}{\alpha+\beta+1}\binom{\alpha+\beta+1}{\alpha+1}=\frac{\alpha-\beta+1}{\alpha+1}\binom{\alpha+\beta}{\alpha}.$$

再除以总轨道数 $\binom{\alpha+\beta}{\alpha}$，得

$$P(A \text{ 从不落后于 } B)=\frac{\alpha-\beta+1}{\alpha+1}.$$

□

【习题 3.23】 **(2.5.3)** 直线上简单对称随机游走 S_n, $S_0=0$。设

$$T=\min\{n\geq 1: S_n=0\}$$

为第一次回到出发点的时刻。证明

$$P(T=2n)=\frac{1}{2n-1}\binom{2n}{n}2^{-2n},$$

并讨论 α 取何值时 $E[T^\alpha]<\infty$。

注. 可以利用 Stirling 公式：$n!\sim n^ne^{-n}\sqrt{2\pi n}$。

<u>**证明.**</u> 显然 T 只能取偶数。记

$$A_n^+=\{T=2n,\ X_1=1\},\qquad A_n^-=\{T=2n,\ X_1=-1\}.$$

由对称性，

$$P(T=2n)=P(A_n^+)+P(A_n^-)=2P(A_n^+).$$

现在来数满足 A_n^+ 的轨道。若 $T=2n$ 且第一步走到 1，则

$$S_1,S_2,\cdots,S_{2n-1}>0,\qquad S_{2n}=0.$$

把这条轨道倒过来读，便得到一条从 $(0,0)$ 到 $(2n-1,1)$ 且在出发后不再过 x 轴的轨道；反过来也可以恢复原轨道，所以这是一个一一对应。

由投票定理，

$$\#\{\text{满足 } A_n^+ \text{ 的轨道}\}=\frac{1}{2n-1}N_{2n-1}(0,1)=\frac{1}{2n-1}\binom{2n-1}{n}.$$

每条长为 $2n$ 的轨道概率都是 2^{-2n}，故

$$\begin{aligned}P(T=2n)&=2\cdot\frac{1}{2n-1}\binom{2n-1}{n}2^{-2n}\\&=\frac{1}{2n-1}\binom{2n}{n}2^{-2n}.\end{aligned}$$

下面讨论 $E[T^\alpha]$。由 Stirling 公式，

$$\binom{2n}{n}\sim\frac{4^n}{\sqrt{\pi n}},$$

从而

$$P(T=2n)=\frac{1}{2n-1}\binom{2n}{n}2^{-2n}\sim\frac{1}{2\sqrt{\pi}}n^{-3/2}.$$

因此

$$E[T^\alpha]=\sum_{n=1}^{\infty}(2n)^\alpha P(T=2n)\asymp\sum_{n=1}^{\infty}n^{\alpha-3/2}.$$

而幂级数 $\sum n^{\alpha-3/2}$ 收敛当且仅当

$$\alpha-\frac{3}{2}<-1,$$

即

$$\alpha<\frac{1}{2}.$$

故

$$E[T^\alpha]<\infty\iff\alpha<\frac{1}{2}.$$

□

【习题 3.24】 **(2.5.4)** 考虑一质点，它沿着按一个圆周排列的标以 $0,1,\cdots,m$ 的 $m+1$ 个节点移动。在每一步质点等概率按顺时针或逆时针方向移动至下一个位置。现在质点从 0 出发按上述规则移动，直到节点 $1,2,\cdots,m$ 均被访问过为止。

1. 证明质点以概率 1 访问所有点 $1,2,\cdots,m$。

2. 求最后一个被访问的节点是 $i(1 \le i \le m)$ 的概率。

<u>**证明.**</u> (1) 对每个固定的 $i \in \{1, 2, \cdots, m\}$，记$A_i = \{$质点最终访问到节点 $i\}$，再对每个 $r \ge 1$，记

$$B_r = \{\text{在第 } (r-1)m+1, (r-1)m+2, \cdots, rm \text{ 步中，节点 } i \text{ 从未被访问}\}.$$

无论第 $(r-1)m$ 步末质点在哪个节点，总可以选定一个方向，使其在接下来的至多 m 步内到达 i；这一特定走法的条件概率至少为 2^{-m}。因此

$$P(B_r \mid \text{前 } (r-1)m \text{ 步的一切结果}) \le 1 - 2^{-m}.$$

从而对任意 $N \ge 1$，

$$P(B_1 \cap \cdots \cap B_N) \le (1 - 2^{-m})^N.$$

若 A_i^c 发生，则每一段长为 m 的时间里都不会访问到 i，故对任意 $N \ge 1$，

$$A_i^c \subseteq B_1 \cap \cdots \cap B_N.$$

从而

$$P(A_i^c) \le P(B_1 \cap \cdots \cap B_N) \le (1 - 2^{-m})^N.$$

令 $N \to \infty$，得 $P(A_i) = 1$。这对每个 $i = 1, 2, \cdots, m$ 都成立。由于只有有限个点，

$$P\Big(\bigcap_{i=1}^{m} A_i\Big) = 1.$$

故质点以概率 1 访问所有点 $1, 2, \cdots, m$。

(2) 设

$$p_i = P(\text{最后一个被访问的节点是 } i), \qquad 1 \le i \le m.$$

由(1)知$\sum_{i=1}^{m} p_i = 1$。

对 $2 \le i \le m-1$，对第一步用全概率公式：

$$\begin{aligned} p_i = &\frac{1}{2} P(\text{第一步到 } 1 \text{ 后，最后一个被访问的是 } i) \\ &+ \frac{1}{2} P(\text{第一步到 } m \text{ 后，最后一个被访问的是 } i). \end{aligned}$$

若第一步走到 1，又最后一个被访问的是 i，则在到达 i 之前，节点 0 必已再次被访问；否则质点不可能“跨过” i 去访问另一侧的节点。于是此时“最后一个被访问的是 i”这件事，与“从 1 出发，把其余 m 个点都看作尚未访问时，最后一个被访问的是 i”是同一事件。再把节点重标为

$$1 \mapsto 0, \quad 2 \mapsto 1, \quad \cdots, \quad m \mapsto m-1, \quad 0 \mapsto m,$$

便回到原问题的同一形式，所以

$$P(\text{第一步到 } 1 \text{ 后，最后一个被访问的是 } i) = p_{i-1}.$$

同理，

$$P(\text{第一步到 } m \text{ 后，最后一个被访问的是 } i) = p_{i+1}.$$

故

$$p_i = \frac{p_{i-1} + p_{i+1}}{2}, \qquad 2 \le i \le m-1.$$

这说明 $p_1, \cdots, p_m$ 成等差数列。

再由关于节点 0 的对称性，得 $p_1 = p_m$。而等差数列首末项相等，只能是常数列，故

$$p_1 = p_2 = \cdots = p_m.$$

结合 $\sum_{i=1}^m p_i = 1$，即得

$$p_i = \frac{1}{m}, \qquad 1 \le i \le m.$$

□

§3.6 习题2.6

【习题 3.25】(2.6.1) 设 G_1, G_2 是概率母函数，$0 \le \alpha \le 1$。证明 G_1G_2 和 $\alpha G_1 + (1-\alpha)G_2$ 也是概率母函数。问

$$\frac{G(\alpha s)}{G(\alpha)}$$

是否依然是概率母函数？

<u>**证明.**</u> 设

$$G_i(s) = \sum_{n=0}^{\infty} p_n^{(i)} s^n, \qquad p_n^{(i)} \ge 0, \qquad \sum_{n=0}^{\infty} p_n^{(i)} = 1, \quad i = 1, 2.$$

则

$$G_1(s)G_2(s) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} p_k^{(1)} p_{n-k}^{(2)} \right) s^n.$$

各项系数非负，且

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} p_k^{(1)} p_{n-k}^{(2)} = \left(\sum_{n=0}^{\infty} p_n^{(1)} \right) \left(\sum_{n=0}^{\infty} p_n^{(2)} \right) = 1,$$

故 G_1G_2 是概率母函数。

又

$$\alpha G_1(s) + (1-\alpha)G_2(s) = \sum_{n=0}^{\infty} \left(\alpha p_n^{(1)} + (1-\alpha) p_n^{(2)} \right) s^n,$$

其系数也都非负，且和为

$$\sum_{n=0}^{\infty} \left(\alpha p_n^{(1)} + (1-\alpha) p_n^{(2)} \right) = \alpha + (1-\alpha) = 1,$$

故 $\alpha G_1 + (1-\alpha)G_2$ 也是概率母函数。

再设

$$G(s)=\sum_{n=0}^{\infty} p_n s^n.$$

当 $G(\alpha)>0$ 时，

$$\frac{G(\alpha s)}{G(\alpha)}=\sum_{n=0}^{\infty}\frac{p_n\alpha^n}{G(\alpha)}s^n.$$

其系数非负，且

$$\sum_{n=0}^{\infty}\frac{p_n\alpha^n}{G(\alpha)}=\frac{G(\alpha)}{G(\alpha)}=1,$$

故它仍是概率母函数。特别地，当 $\alpha\in(0,1]$ 时总成立；若 $\alpha=0$，则只有在 $G(0)>0$ 时该式才有意义，此时它恒等于 1，也仍是概率母函数。 □

【习题 3.26】**(2.6.3)** 设 X 服从参数为 $p\ (0<p<1)$ 的几何分布，即

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

又设非负整值随机变量 Y 的概率母函数为 $G(s)$，且 Y 与 X 独立。证明

$$\mathbb{P}(X>Y)=G(1-p).$$

证明. 由全概率公式和独立性，

$$\mathbb{P}(X>Y)=\sum_{n=0}^{\infty}\mathbb{P}(X>n,Y=n)=\sum_{n=0}^{\infty}\mathbb{P}(X>n)\mathbb{P}(Y=n).$$

而

$$\mathbb{P}(X>n)=\sum_{k=n+1}^{\infty}(1-p)^{k-1}p=(1-p)^n.$$

故

$$\mathbb{P}(X>Y)=\sum_{n=0}^{\infty}(1-p)^n\mathbb{P}(Y=n)=G(1-p).$$

□

【习题 3.27】**(2.6.4)** 证明

$$G(x,y,z,w)=\frac{1}{8}(xyzw+xy+yz+zw+xw+yw+xz+1)$$

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

$$G(x,y,z,w)=\frac{1}{8}(xyzw+xy+yz+zw+xw+yw+xz+1)$$

可见各项系数都非负，且系数和为 1，故它确是某个四维随机向量的联合母函数。

对边缘母函数，有

$$G_X(x)=G(x,1,1,1)=\frac{1+x}{2},$$

其余三个也一样。

再看二维联合母函数，

$$G_{X,Y}(x,y) = G(x,y,1,1) = \frac{(1+x)(1+y)}{4} = G_X(x)G_Y(y).$$

由对称性，任意两个随机变量都独立。

再看三维联合母函数，

$$G_{X,Y,Z}(x,y,z) = G(x,y,z,1) = \frac{(1+x)(1+y)(1+z)}{8} = G_X(x)G_Y(y)G_Z(z).$$

由对称性，任意三个随机变量也都独立。

但若四个随机变量相互独立，则其联合母函数应为

$$G_X(x)G_Y(y)G_Z(z)G_W(w) = \frac{(1+x)(1+y)(1+z)(1+w)}{16},$$

这显然不等于 $G(x,y,z,w)$，例如右边含有 x 项而左边没有，故它们不相互独立。 □

§3.7 补充内容

课程拾遗

【定理 3.28】(分布函数的刻画) 设 $F:\mathbb{R}\to\mathbb{R}$ 是一个函数，则 F 是某个随机变量的分布函数，当且仅当它满足以下三个性质：

1. **单调不减性**：对于任意 $x_1 < x_2$，有 $F(x_1) \le F(x_2)$；
2. **右连续性**：对于任意 $x\in\mathbb{R}$，有 $\lim_{y\to x^+} F(y) = F(x)$；
3. **规范性**：$\lim_{x\to-\infty} F(x) = 0$，且 $\lim_{x\to+\infty} F(x) = 1$。

证明. 必要性略。下面证明充分性。

设 F 满足以上三条性质。取一个服从 $U(0,1)$ 的随机变量 U，定义

$$X = \inf\{t\in\mathbb{R} : F(t) \ge U\}.$$

由 $\lim_{x\to-\infty} F(x) = 0$ 与 $\lim_{x\to+\infty} F(x) = 1$ 可知，上面的集合非空且有下界，因此 X 的定义是有意义的。

对任意 $x\in\mathbb{R}$，若 $U \le F(x)$，则 $x\in\{t : F(t)\ge U\}$，从而 $X \le x$。故

$$\{U \le F(x)\} \subseteq \{X \le x\}.$$

反过来，若 $X \le x$，则对每个 $n\ge 1$ 都可取某个 $t_n < x + \frac{1}{n}$ 使得 $F(t_n) \ge U$。由 F 单调不减，

$$U \le F(t_n) \le F\left(x + \frac{1}{n}\right).$$

令 $n \to \infty$，再用 F 的右连续性，就得 $U \le F(x)$。因此

$$\{X \le x\} \subseteq \{U \le F(x)\}.$$

于是

$$\{X \le x\} = \{U \le F(x)\},$$

从而

$$\mathbb{P}(X \le x) = \mathbb{P}(U \le F(x)) = F(x).$$

所以 F 正是随机变量 X 的分布函数。 □

好题共赏

【习题 3.29】(简单随机游走的吸收时间) 设 $\{S_n\}_{n\ge 0}$ 是状态空间 $\{0,1,\dots,L\}$ 上的简单随机游走，且 0 与 L 都是吸收态。若 $S_0=1$，求在第 n 步恰好被吸收的概率。

<u>**证明.**</u> 记吸收时刻为

$$\tau = \inf\{n \ge 0 : S_n \in \{0, L\}\}.$$

对 $x = 1,2,\dots,L-1$，记

$$P(x,n) = \mathbb{P}(S_n = x,\ \tau > n),$$

则

$$P(x,n) = \frac{1}{2}P(x-1,n-1) + \frac{1}{2}P(x+1,n-1),$$

并满足边界条件

$$P(0,n) = P(L,n) = 0,$$

以及初值

$$P(x,0) = \mathbf{1}_{\{x=1\}}.$$

由于边界为零，对空间变量作正弦展开

$$P(x,n) = \sum_{m=1}^{L-1} a_m(n) \sin\frac{m\pi x}{L},$$

代入递推式得

$$a_m(n) = a_m(n-1)\cos\frac{m\pi}{L},$$

故

$$a_m(n) = a_m(0)\cos^n\frac{m\pi}{L}.$$

再由初值可得

$$a_m(0) = \frac{2}{L}\sin\frac{m\pi}{L}.$$

于是

$$P(x,n) = \sum_{m=1}^{L-1} \frac{2}{L}\sin\frac{m\pi}{L}\cos^n\frac{m\pi}{L}\sin\frac{m\pi x}{L}.$$

从而

$$\mathbb{P}(\tau > n) = \sum_{x=1}^{L-1} P(x, n),$$

所以

$$\mathbb{P}(\tau = n) = \mathbb{P}(\tau > n-1) - \mathbb{P}(\tau > n).$$

这就把题目化成了上面的显式表达式。 □

【注 3.30】 若从任意初始位置 $i \in \{1, 2, \ldots, L-1\}$ 出发，只需把初值改成

$$P(x, 0) = \mathbf{1}_{\{x=i\}},$$

其余推导完全相同。

【习题 3.31】(首次连续成功的分布) 设独立重复掷一枚硬币，每次出现正面的概率为 p，反面的概率为 $q = 1 - p$。记 N 为首次连续 m 次出现正面所需的抛掷次数，求 N 的生成函数。

证明. 记

$$P_n = \mathbb{P}(N = n), \qquad n \geq m.$$

显然

$$P_n = 0 \quad (n < m), \qquad P_m = p^m.$$

对 $n > m$，按前 m 次中第一次出现反面的位置分类，可得递推

$$P_n = q \sum_{k=1}^{m} p^{k-1} P_{n-k}, \qquad n > m.$$

设生成函数

$$G(z) = \sum_{n=m}^{\infty} P_n z^n.$$

对上式求和得

$$G(z) - p^m z^m = q \sum_{n=m+1}^{\infty} \sum_{k=1}^{m} p^{k-1} P_{n-k} z^n.$$

整理得

$$G(z) - p^m z^m = qz\big(1 + pz + \cdots + (pz)^{m-1}\big)G(z).$$

于是

$$G(z) = \frac{p^m z^m}{1 - qz\big(1 + pz + \cdots + (pz)^{m-1}\big)}.$$

再用等比数列求和公式

$$1 + pz + \cdots + (pz)^{m-1} = \frac{1 - (pz)^m}{1 - pz},$$

化简得

$$G(z) = \frac{(pz)^m (1 - pz)}{1 - z + q p^m z^{m+1}}.$$

这就是 N 的生成函数。 □

§3.8 期中复习

复习建议是：是把书上的概念、作业里的题型，以及反复讲过的方法重新理一遍。大体上可以按“概念 → 作业 → 典型难点”这条线来复习。

一、基本概念

考试里最容易失分的，往往不是最难的计算，而是概念模糊、定义写不清楚、性质不会用。下面这些内容至少要做到“能自己说清楚定义，能判断一道题该用哪个概念”：

1. 概率空间三要素：样本空间、事件域、概率测度；
2. 随机变量与分布函数的定义，以及分布函数的基本性质；
3. 离散型与连续型随机变量，分布列、密度函数、分布函数之间的关系；
4. 二维随机变量的联合分布、边缘分布、条件分布、独立性；
5. 数学期望、方差、协方差、相关系数的定义与基本性质；
6. 条件期望的含义，以及“先条件、后取期望”的思想；
7. 常见分布的特点：Bernoulli，Binomial，Geometric，Poisson。

【注 3.32】 这一部分建议大家不要只“看着眼熟”，而要真的能脱离讲义自己复述出来。尤其是“什么叫独立”“什么叫条件期望”“什么叫分布函数”，最好都能用一句完整的话说明白。

举两例：

【习题 3.33】 (24秋,1) 掷两枚均匀硬币，详细写出概率空间三要素，并说明其上存在两个独立的随机变量。

解. 可以取

$$\Omega = \{HH, HT, TH, TT\},$$

其中第一个字母表示第一枚硬币的结果，第二个字母表示第二枚硬币的结果。事件域取为

$$\mathcal{F} = 2^{\Omega},$$

概率测度 P 由

$$P(\{\omega\}) = \frac{1}{4}, \qquad \omega \in \Omega$$

给出。

这就写出了概率空间三要素 $(\Omega, \mathcal{F}, P)$。

再定义两个随机变量

$$X = \mathbf{1}_{\{\text{第一枚硬币为正面}\}}, \qquad Y = \mathbf{1}_{\{\text{第二枚硬币为正面}\}}.$$

则 X, Y 都只取 $0,1$ 两个值，且

$$P(X=1)=P(Y=1)=\frac{1}{2}.$$

进一步，

$$P(X=i, Y=j)=\frac{1}{4}=P(X=i)P(Y=j), \qquad i,j\in\{0,1\}.$$

所以 X 与 Y 相互独立。 □

【习题 3.34】 **(19秋,2)** 在 $[0,1]$ 上给出一个概率空间，并问

$$A_1=[a_1,b_1], \qquad A_2=[a_2,b_2]$$

何时独立？

解. 取

$$\Omega=[0,1], \qquad \mathcal{F}=\mathcal{B}([0,1]), \qquad P=\mu,$$

其中 μ 表示 $[0,1]$ 上由区间长度给出的 Borel 概率测度，即对任意闭区间 $[a,b]\subset[0,1]$ 有

$$P([a,b])=b-a.$$

记

$$\ell_1=b_1-a_1, \qquad \ell_2=b_2-a_2.$$

则 A_1, A_2 独立当且仅当

$$P(A_1\cap A_2)=P(A_1)P(A_2)=\ell_1\ell_2.$$

不妨设 $a_1\le a_2$。

(1) 若 $b_1\le a_2$，则两区间至多在一个点相交，所以

$$P(A_1\cap A_2)=0.$$

此时独立当且仅当 $\ell_1\ell_2=0$，也就是至少有一个区间退化成一点。

(2) 若 $a_2\le b_2\le b_1$，则 $A_2\subset A_1$，于是

$$P(A_1\cap A_2)=P(A_2)=\ell_2.$$

独立要求

$$\ell_2=\ell_1\ell_2,$$

所以或者 $\ell_2=0$，或者 $\ell_1=1$。也就是说，或者 A_2 是单点，或者 $A_1=[0,1]$。

(3) 若 $a_2<b_1<b_2$，则是部分重叠但互不包含。记

$$x=a_2-a_1, \qquad y=b_1-a_2, \qquad z=b_2-b_1,$$

则 $x,y,z>0$，并且

$$\ell_1=x+y, \qquad \ell_2=y+z, \qquad P(A_1\cap A_2)=y.$$

若独立，则应有

$$y=(x+y)(y+z)=y^2+y(x+z)+xz>y,$$

矛盾。因此这种情形不可能独立。

综上，在这个概率空间上，两个闭区间独立当且仅当至少有一个区间的概率是 0 或 1；也就是说，至少有一个区间要么退化成单点，要么就是整个 $[0,1]$。 □

二、作业回顾

作业题本身就是最重要的复习资料。很多考点会以相似(甚至相同)形式反复出现，所以必须所有的作业题都会做。

其中建议重点回看的例子包括：

1. 若 X, Y 独立，且 $X \sim \text{Poisson}(\lambda_1)$，$Y \sim \text{Poisson}(\lambda_2)$，求
$$\mathbb{E}[X \mid X+Y].$$

2. 若 $N \sim \text{Poisson}(\lambda)$，在给定 N 的条件下抛 N 次硬币，设得到的正面数为 X，求
$$\mathbb{E}[N \mid X].$$

3. 设直线上简单随机游走
$$S_n = \sum_{k=1}^{n} X_k, \qquad S_0 = 0,$$
其中
$$\mathbb{P}(X_i = 1) = p, \qquad \mathbb{P}(X_i = -1) = 1-p, \qquad 0 < p < 1.$$
对 $m \le n$，求
$$\text{Cov}(S_n, S_m) \quad 与 \quad \text{Var}(S_n \mid S_m).$$

需要掌握的方法和技术有：

1. Cauchy 不等式，以及它在估计、证方差非负、控制期望量级中的基本用法；
2. Markov 不等式，以及由“期望控制尾概率”这一思路出发的简单估计；
3. 全期望公式(“匿名统计学家公式”)；
4. 随机变量拆成示性 r.v. 之和，方便期望、方差甚至高阶矩计算。

三、典例分析

1. 母函数与矩母函数

母函数/矩母函数的作用：

1. 普通母函数可以把非负整值随机变量的分布列编码成一个函数，从整体上用分析手段处理分布；
2. 通过求导可以计算期望、方差以及更高阶矩；
3. 对独立随机变量，和的母函数/矩母函数等于各自母函数/矩母函数之积，因此便于求和的分布；

4. 母函数特别适合处理递推关系、首次出现时间等问题；

5. 矩母函数在存在时常可用来刻画分布，并方便比较不同分布的矩。

【习题 3.35】(对称随机游走与 Catalan 数) 设 $\{S_n\}_{n\geq 0}$ 为直线上的对称随机游走，$S_0 = 0$，每一步以概率 $\frac{1}{2}$ 向右走一步，以概率 $\frac{1}{2}$ 向左走一步。求

$$\mathbb{P}(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0, S_{2n} = 0).$$

解. 记

$$C_n := \#\{(S_1, \ldots, S_{2n}) : S_i \geq 0,\ 1 \leq i \leq 2n,\ S_{2n} = 0\}, \qquad C_0 = 1.$$

则所求概率为

$$\mathbb{P}(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0, S_{2n} = 0) = \frac{C_n}{2^{2n}},$$

因为每条长度为 $2n$ 的轨道出现的概率都等于 2^{-2n}。

设这条轨道第一次回到原点的时刻是 $2k$，其中 $1 \leq k \leq n$。那么在时刻 $1, 2, \ldots, 2k-1$ 必有 $S_i \geq 1$。把这一段轨道整体下移 1，便得到一条从 0 出发、长为 $2k-2$、始终不低于 0、并在末时刻回到 0 的轨道，因此这样的前段共有 C_{k-1} 条。第一次回到原点以后，后面的 $2(n-k)$ 步又是一条同类轨道，共有 C_{n-k} 条。所以

$$C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}, \qquad n \geq 1.$$

引入母函数

$$F(z) := \sum_{n=0}^{\infty} C_n z^n.$$

由上式得

$$\begin{aligned} F(z) &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} C_{k-1} C_{n-k} z^n \\ &= 1 + zF(z)^2. \end{aligned}$$

因此

$$zF(z)^2 - F(z) + 1 = 0,$$

解得

$$F(z) = \frac{1 - \sqrt{1-4z}}{2z},$$

这里取使 $F(0) = 1$ 的那一支。于是

$$C_n = \frac{1}{n+1} \binom{2n}{n}.$$

因此

$$\mathbb{P}(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0, S_{2n} = 0) = \frac{1}{2^{2n}} \cdot \frac{1}{n+1} \binom{2n}{n}.$$

□

【习题 3.36】 **(25春,6)** 设非常值随机变量 X_n 取值于 $\{0,1,\ldots,2n\}$，其母函数$G(z) = \mathbb{E}[z^{X_n}]$ 为 $2n$ 次多项式，且满足 Lee–Yang 性质：复变量 z 的方程 $G(z)=0$ 的所有根均在单位圆上。

1. 写出一个随机变量，其母函数具有 Lee–Yang 性质。
2. 证明对所有非负整数 m，
$$\mathbb{E}\left[(X_n-n)^{2m+1}\right]=0.$$
3. 令
$$X_n^* := \frac{X_n-\mathbb{E}[X_n]}{\sqrt{\mathrm{Var}(X_n)}},$$
证明
$$1 \le \mathbb{E}\left[(X_n^*)^4\right] < 3.$$

解. **(i)** 一个典型例子是二项分布
$$X_n \sim \mathrm{Bin}\left(2n,\frac{1}{2}\right).$$
此时
$$G(z)=\left(\frac{1+z}{2}\right)^{2n},$$
它的全部零点都是 $z=-1$，位于单位圆上，因此满足 Lee–Yang 性质。

(ii) 设
$$Y := X_n - n.$$
由于 G 的系数都是实数，且所有根都在单位圆上，因此这些根成共轭对出现，从而可写成
$$G(z)=\lambda\prod_{k=1}^{n}(z^2-a_kz+1), \qquad a_k\in[-2,2].$$
于是
$$M_Y(t) := \mathbb{E}[e^{tY}] = e^{-nt}G(e^t)=\lambda\prod_{k=1}^{n}(e^t+e^{-t}-a_k).$$
右端是 t 的偶函数，所以 $M_Y(t)$ 是偶函数。因而对一切非负整数 m，
$$\mathbb{E}[Y^{2m+1}] = M_Y^{(2m+1)}(0)=0.$$
也就是
$$\mathbb{E}\left[(X_n-n)^{2m+1}\right]=0.$$

(iii) 由上问取 $m=0$ 可知
$$\mathbb{E}[X_n]=n,$$
因此 $Y=X_n-\mathbb{E}[X_n]$。

记

$$M_Y(t) = \mathbb{E}[e^{tY}], \qquad f(t) = \log M_Y(t).$$

由于奇数阶矩都为零，

$$M_Y(t) = 1 + \frac{\mathbb{E}[Y^2]}{2}t^2 + \frac{\mathbb{E}[Y^4]}{24}t^4 + o(t^4),$$

从而

$$f(t) = \frac{\mathbb{E}[Y^2]}{2}t^2 + \frac{\mathbb{E}[Y^4] - 3(\mathbb{E}[Y^2])^2}{24}t^4 + o(t^4).$$

另一方面，令 $c_k := 2 - a_k \in (0,4]$，则

$$f(t) = \sum_{k=1}^{n} \log\left(c_k + t^2 + \frac{t^4}{12} + o(t^4)\right) + \log\lambda,$$

故

$$\begin{aligned}\log\left(c_k + t^2 + \frac{t^4}{12} + o(t^4)\right) &= \log c_k + \log\left(1 + \frac{t^2}{c_k} + \frac{t^4}{12c_k} + o(t^4)\right) \\ &= \log c_k + \frac{t^2}{c_k} + \left(\frac{1}{12c_k} - \frac{1}{2c_k^2}\right)t^4 + o(t^4).\end{aligned}$$

因此

$$f(t) = C + \sum_{k=1}^{n}\left[\frac{t^2}{c_k} + \left(\frac{1}{12c_k} - \frac{1}{2c_k^2}\right)t^4\right] + o(t^4),$$

其中 C 为常数。又因为 $0 < c_k \le 4 < 6$，所以

$$\frac{1}{12c_k} - \frac{1}{2c_k^2} = \frac{c_k - 6}{12c_k^2} < 0.$$

若改写成四阶导数，则

$$f^{(4)}(0) = 24\sum_{k=1}^{n}\left(\frac{1}{12c_k} - \frac{1}{2c_k^2}\right) = \sum_{k=1}^{n}\left(\frac{2}{c_k} - \frac{12}{c_k^2}\right) < 0.$$

因此 $f(t)$ 的 t^4 系数为负，故

$$\mathbb{E}[Y^4] - 3(\mathbb{E}[Y^2])^2 < 0,$$

即

$$\mathbb{E}[Y^4] < 3(\mathbb{E}[Y^2])^2.$$

标准化后得到

$$\mathbb{E}\big[(X_n^*)^4\big] = \frac{\mathbb{E}[Y^4]}{(\mathbb{E}[Y^2])^2} < 3.$$

另一方面，由 Jensen 不等式（或 Cauchy 不等式）

$$\mathbb{E}\big[(X_n^*)^4\big] \ge \big(\mathbb{E}[(X_n^*)^2]\big)^2 = 1.$$

综上，

$$1 \le \mathbb{E}\big[(X_n^*)^4\big] < 3.$$

$\square$

2. 简单随机游走及其常见变式

请优先掌握讲义内的所有内容，可阅读前面作业讲解的部分，那几题都需要重点掌握。一定要掌握课本定理2.5.2到定理2.5.5，考试时反射原理、投票定理都可以直接引用。

【习题 3.37】 (24秋,5) 在只有两位候选人的选举中，每张选票只投给其中一人且不能弃票。已知最终计票结果为 T 有 α 张选票，H 有 β 张选票，其中 $\alpha \geq \beta$。若按随机顺序计票，求计票过程中 T 至多落后 H 一票的概率。

解. 仿照前面的做法，构造随机游走，令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 票给 } T, \\ -1, & \text{第 } i \text{ 票给 } H, \end{cases} \qquad S_k = \sum_{i=1}^{k} X_i.$$

则前 k 张票计完后，T 比 H 多出的票数就是 S_k。题目要求的是

$$S_k \geq -1, \qquad 1 \leq k \leq \alpha + \beta.$$

现在在每一种计票次序最前面补上一张投给 T 的选票。这样便得到一个新的计票次序：其中 T 有 $\alpha + 1$ 张票，H 有 β 张票，而且新序列的每一步满足

$$1 + S_k \geq 0.$$

也就是说，原问题恰好化为：在新的选举中，T 在计票过程中从不落后于 H。

反过来，任何一个"T 从不落后于 H"的新计票次序，第一票必为 T；删去这一票后，就恢复为原问题中的一个合法次序。因此这是一个一一对应。

于是由投票定理（或上面习题2.5.2(2)的结论），合法次序数为

$$\frac{(\alpha+1) - \beta + 1}{(\alpha+1)+1} \binom{\alpha+\beta+1}{\alpha+1} = \frac{\alpha - \beta + 2}{\alpha + 2} \binom{\alpha+\beta+1}{\alpha+1}.$$

而原问题的总计票次序数为

$$\binom{\alpha+\beta}{\alpha}.$$

故所求概率为

$$\begin{aligned} &\frac{\frac{\alpha-\beta+2}{\alpha+2}\binom{\alpha+\beta+1}{\alpha+1}}{\binom{\alpha+\beta}{\alpha}} \\ =&\frac{\alpha-\beta+2}{\alpha+2} \cdot \frac{\alpha+\beta+1}{\alpha+1}. \end{aligned}$$

即

$$P(\text{计票过程中 } T \text{ 至多落后 } H \text{ 一票}) = \frac{(\alpha+\beta+1)(\alpha-\beta+2)}{(\alpha+1)(\alpha+2)}.$$

□

3. 概率论与其他学科的交叉

【习题 3.38】 (25春,5) 概率论与线性代数的结合可能催生有趣的数学问题与方法，且看一例。令$X_n = (X_{ij})$为 $n \times n$ 矩阵，n^2 个矩阵元 $\{X_{ij}\}$ 为相互独立且同分布的对称伯努利随机变量，即

$$\mathbb{P}(X_{ij} = 0) = \mathbb{P}(X_{ij} = 1) = \frac{1}{2}.$$

定义$p_n = \mathbb{P}(\det(X_n)$为奇数)，试回答 (i) 计算 p_2, p_3；(ii) 猜测 p_n 的一般公式并证明之。

解. 关键观察是：一个整数是奇数，当且仅当它模 2 余 1。因此

$$\det(X_n)\text{ 为奇数} \iff \det(X_n) \not\equiv 0 \pmod 2.$$

也就是说，把 X_n 看成 $\mathbf{F}_2 = \{0, 1\}$ 上的矩阵时，问题就变成了：

$$p_n = \mathbb{P}(X_n \text{ 在 } \mathbf{F}_2 \text{ 上可逆}).$$

现在从“行向量是否线性无关”来计算这个概率。把 X_n 的各行记为

$$R_1, R_2, \ldots, R_n \in \mathbf{F}_2^n.$$

由于各个矩阵元独立且都以概率 $\frac{1}{2}$ 取 $0, 1$，所以每个 R_i 都在 $\mathbf{F}_2^n$ 中等概率取值，并且彼此独立。

第一行非零的概率为

$$\frac{2^n - 1}{2^n} = 1 - 2^{-n}.$$

若前 k 行已经线性无关，则它们张成的子空间有 2^k 个向量，所以第 $k+1$ 行落在这个子空间外的条件概率为

$$\frac{2^n - 2^k}{2^n} = 1 - 2^{k-n}.$$

于是

$$p_n = \prod_{k=0}^{n-1} \left(1 - 2^{k-n}\right) = \prod_{j=1}^{n} \left(1 - 2^{-j}\right).$$

因此

$$p_2 = \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{4}\right) = \frac{3}{8},$$

$$p_3 = \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{8}\right) = \frac{21}{64}.$$

所以一般公式为

$$p_n = \prod_{j=1}^{n} \left(1 - \frac{1}{2^j}\right).$$

□

【习题 3.39】 **(20秋,4)** 记对称群 S_n 为从 $\{1,2,\cdots,n\}$ 到 $\{1,2,\cdots,n\}$ 的所有一一映射（共 $n!$ 个），从 S_n 中均匀等概率选取一个映射 σ，记其不动点数

$$X(\sigma)=|\{k \mid \sigma(k)=k\}|,$$

对换数

$$Y(\sigma)=|\{(i,j) \mid i<j,\, \sigma(i)=j,\, \sigma(j)=i\}|.$$

1. 详细给出有关概率空间。
2. X,Y 是否独立。
3. 计算 X 的分布列。
4. 计算 $\mathbb{E}[Y]$。

解. **(1)** 概率空间可取为

$$\Omega=S_n,\qquad \mathcal{F}=2^{S_n},\qquad P(A)=\frac{|A|}{n!}\quad (A\subset S_n).$$

这里样本点就是一个排列 σ，而 X,Y 都是定义在 Ω 上的随机变量。

(2) 当 $n\geq 2$ 时，X,Y 不独立。事实上，

$$P(X=n)=\frac{1}{n!}>0,$$

而

$$P(Y>0)>0$$

因为例如排列 $(1\,2)$ 就有一个对换。另一方面，若 $X=n$，则 σ 只能是恒等排列，此时必有 $Y=0$。所以

$$P(X=n,Y>0)=0\neq P(X=n)P(Y>0).$$

故 X,Y 不独立。

(3) 对 $k=0,1,\ldots,n$，先选出哪 k 个点是不动点，有

$$\binom{n}{k}$$

种选法。剩下的 $n-k$ 个点必须都不是不动点，因此对应的是一个错排。记 D_m 为 m 个元素的错排数，则

$$P(X=k)=\frac{\binom{n}{k}D_{n-k}}{n!},\qquad k=0,1,\ldots,n.$$

再由容斥原理，

$$D_m=m!\sum_{j=0}^{m}\frac{(-1)^j}{j!}.$$

于是

$$P(X=k)=\frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!},\qquad k=0,1,\ldots,n.$$

这就是 X 的分布列。

(4) 对每个 $1 \leq i < j \leq n$，定义示性随机变量

$$I_{ij} = \mathbf{1}_{\{\sigma(i)=j,\ \sigma(j)=i\}}.$$

则

$$Y = \sum_{1\leq i<j\leq n} I_{ij}.$$

由期望的线性性，

$$\mathbb{E}[Y] = \sum_{1\leq i<j\leq n} \mathbb{E}[I_{ij}] = \sum_{1\leq i<j\leq n} P(\sigma(i) = j, \sigma(j) = i).$$

固定一对 $i < j$ 后，使 $\sigma(i) = j, \sigma(j) = i$ 的排列共有 $(n-2)!$ 个，所以

$$P(\sigma(i) = j, \sigma(j) = i) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

因此

$$\mathbb{E}[Y] = \binom{n}{2}\frac{1}{n(n-1)} = \frac{1}{2}.$$

□

Chapter 4

第四次习题课

§4.1 习题3.1

【习题 4.1】 求以下分布的归一化常数

1) $f(x) = \frac{1}{C}\frac{1}{\sqrt{x(1-x)}}, \quad 0 < x < 1$

2) $f(x) = \frac{1}{C}e^{-x-e^{-x}}, \quad x \in \mathbb{R}$

证明. 1)

$$f(x) = \frac{1}{C}\frac{1}{\sqrt{x(1-x)}}, \quad 0 < x < 1$$

由 $\int_0^1 f(x)\,dx = 1$，得

$$\frac{1}{C}\int_0^1 \frac{dx}{\sqrt{x(1-x)}} = 1$$

计算积分：令 $x = \sin^2 t$，则 $dx = 2\sin t\cos t\,dt$，

$$\sqrt{x(1-x)} = \sin t\cos t, \quad \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = \int_0^{\pi/2} 2\,dt = \pi$$

所以

$$\frac{\pi}{C} = 1 \quad \Rightarrow \quad C = \pi.$$

2) 注意到 $\int e^{-x-e^{-x}}\,dx = e^{-e^{-x}} + C$，计算立得 $C = 1$。

□

【习题 4.2】 求下面积分的归一化常数

$$\int_{\mathbb{R}^2} |x_1 - x_2| e^{-\frac{1}{2}(x_1^2+x_2^2)}\,dx_1dx_2$$

证明. 令 $u = x_1 - x_2$, $v = x_1 + x_2$

$$|J| = \frac{1}{2}$$

$$\text{上式} = \int_{\mathbb{R}^2} |u| e^{-\frac{1}{4}(u^2+v^2)} \cdot \frac{1}{2} dudv$$

$$= 2\int_0^{+\infty} ue^{-\frac{1}{4}u^2}\, du \cdot \int_0^{+\infty} e^{-\frac{1}{4}v^2}\, dv$$

$$= 4\sqrt{\pi}.$$

从而，$C = 4\sqrt{\pi}$. □

【习题 4.3】 随机变量 X, Y 相互独立且服从参数为 1 的指数分布, 求 U = X + Y 的密度.

<u>**证明.**</u> 直接计算 $f_U(u) = \int_0^u e^{-x}e^{x-u}\, dx = ue^{-u}\mathbb{1}_{u\geq 0}$. □

【习题 4.4】 4. 设 X, Y, Z 相互独立, 且均服从 [0, 1] 上的均匀分布.
(1) 求 $-\log X$ 的密度函数;
(2) 证明 $W = XYZ$ 也服从 $[0,1]$ 上的均匀分布.

<u>**证明.**</u> (1) $\mathbb{P}(-\log X \leq x) = \mathbb{P}(X \geq e^{-x}) = 1 - e^{-x}$，故密度为 $e^{-x}\mathbb{1}_{x\geq 0}$。

(2) $-\log W = Z(-\log X - \log Y)$, 故只需证明 $Z(-\log X - \log Y)$ 满足参数为1的指数分布. 记 $T = -\log X - \log Y$, 由上题,

$$f_T(t) = te^{-t}\mathbb{1}_{t\geq 0}.$$

$$f_{T,Z}(t,z) = te^{-t}\mathbb{1}_{t\geq 0}\mathbb{1}_{0\leq z\leq 1}.$$

只需证 $W = TZ$ 服从参数为1的指数分布.

作变换 $w = tz, s = t$,

$$J = \begin{pmatrix} \frac{\partial z}{\partial w} & \frac{\partial z}{\partial s} \\ \frac{\partial w}{\partial t} & \frac{\partial w}{\partial s} \end{pmatrix} = \begin{pmatrix} \frac{1}{s} & -\frac{w}{s^2} \\ 0 & 1 \end{pmatrix}.$$

$$|\det(J)| = \frac{1}{s},$$

由联合密度函数变换公式，得

$$f_{W,S}(w,s) = e^{-s}\mathbb{1}_{0\leq w\leq s}.$$

$$f_W(w) = \int_{\mathbb{R}} e^{-s}\mathbb{1}_{0\leq w\leq s}\, ds = \mathbb{1}_{w\geq 0}\int_w^{+\infty} e^{-s}\, ds = e^{-w}\mathbb{1}_{w\geq 0}.$$

□

§4.2 习题3.2

【习题 4.5】 X 服从标准正态分布, Y 服从标准 Wigner 半圆律, 求它们的各阶矩

证明. 因为二者都是对称的随机变量，只需证明2k阶矩。

$$\begin{aligned}
\mathbb{E}[X^{2k}] &= \int_{-\infty}^{+\infty} x^{2k}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx \\
&= \frac{2^{k+\frac{1}{2}}}{\sqrt{2\pi}}\int_0^{+\infty} u^{k-\frac{1}{2}}e^{-u}\,du \quad (\text{令 } u=\frac{x^2}{2}) \\
&= \frac{2^{k+\frac{1}{2}}}{\sqrt{2\pi}}\Gamma\left(k+\frac{1}{2}\right) \\
&= \frac{2^{k+\frac{1}{2}}}{\sqrt{2\pi}}\cdot\left(\frac{\sqrt{\pi}\,(2k-1)!!}{2^k}\right) \quad (\text{利用 } \Gamma(n+1)=n\Gamma(n) \text{ 与 } \Gamma(\tfrac{1}{2})=\sqrt{\pi}) \\
&= (2k-1)!!
\end{aligned}$$

对于半圆律：

$$\begin{aligned}
\mathbb{E}[Y^{2k}] &= 2\int_0^2 \frac{1}{2\pi}y^{2k}\sqrt{4-y^2}\,dy \\
&\overset{\text{令 } y=2\sin\theta}{=} \frac{2^{2k+2}}{\pi}\int_0^{\frac{\pi}{2}}(\sin\theta)^{2k}\cos^2\theta\,d\theta \\
&= \frac{2^{2k+2}}{\pi}\int_0^{\frac{\pi}{2}}\left[(\sin\theta)^{2k}-(\sin\theta)^{2k+2}\right]d\theta.
\end{aligned}$$

数学分析里已经算过：

$$\int_0^{\frac{\pi}{2}}\sin^n x\,dx = \begin{cases} \dfrac{n-1}{n}\dfrac{n-3}{n-2}\cdots\dfrac{1}{2}\cdot\dfrac{\pi}{2}, & n \text{ 为正偶数} \\ \dfrac{n-1}{n}\dfrac{n-2}{n-3}\cdots\dfrac{1}{3}\cdot 1, & n \text{ 为正奇数} \end{cases}$$

代入得到最后的结果为 $\dfrac{1}{k+1}\dbinom{2k}{k}$。 □

【习题 4.6】 (X,Y) 联合密度为

$$f(x,y) = C^{-1}x(y-x)e^{-y}, \quad 0 \le x \le y < \infty.$$

求常数 C，条件密度 $f_{X|Y}$ 与条件期望 $\mathbb{E}[Y|X]$。

证明.

$$\begin{aligned}
&\int_0^{+\infty}\int_0^y Cx(y-x)e^{-y}\,dx\,dy \\
&= \int_0^{+\infty} Ce^{-y}\,dy\int_0^y x(y-x)\,dx \\
&= \int_0^{+\infty} C\cdot\frac{1}{6}y^3e^{-y}\,dy \\
&= C\cdot\frac{1}{6}\cdot 3! \\
&= C.
\end{aligned}$$

故 $C=1$.

$$\begin{aligned}
f_{X|Y}(x) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{x(y-x)e^{-y}\mathbb{1}_{0\le x\le y}}{\frac{1}{6}y^3e^{-y}} \\
&= \frac{6x(y-x)}{y^3}\mathbb{1}_{0\le x\le y}.
\end{aligned}$$

类似地,

$$\begin{aligned}
f_{Y|X}(y) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{x(y-x)e^{-y}\mathbb{1}_{0\le x\le y}}{xe^{-x}} \\
&= (y-x)e^{-(y-x)}\mathbb{1}_{y\ge x}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y\mid X] &= \int_x^\infty y(y-x)e^{-(y-x)}\,dy \\
&= X+2.
\end{aligned}$$

□

【习题 4.7】 4. 设 g 为 $\mathbb{R}$ 上连续可微函数，并且 g 和其导数 g' 均有界，证明

$$X\sim N(\mu,\sigma^2) \implies \mathbb{E}[(X-\mu)g(X)] = \sigma^2\mathbb{E}[g'(X)].$$

证明. 不妨令 $Z \sim N(0,1)$。

$$
\begin{aligned}
\mathbb{E}f'(Z) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-t^2/2} f'(t)dt \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} f'(t)\int_t^{\infty} we^{-w^2/2}dwdt - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 f'(t)\int_{-\infty}^t we^{-w^2/2}dwdt \\
&\overset{Fubini}{=} \frac{1}{\sqrt{2\pi}}\int_0^{\infty} we^{-w^2/2}\left[\int_0^w f'(t)dt\right]dw - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 we^{-w^2/2}\left[\int_w^0 f'(t)dt\right]dw \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} we^{-w^2/2}f(w)dw + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 we^{-w^2/2}f(w)dw \\
&= \mathbb{E}[Zf(Z)].
\end{aligned}
$$

反之，如果 Z 满足方程 $\mathbb{E}(Zg(Z)) = \mathbb{E}(g'(Z))$，考虑 Stein 方程

$$xf(x) - f'(x) = h(x) - \mathbb{E}(h(G))$$

（其中 $G \sim N(0,1)$）。固定 h，这是一个关于 f 的常微分方程，具有唯一解。令 $h(x) = \mathbb{1}_{x \le z}$。把 $x = Z$ 代入，两边求期望得到：

$$0 = \mathbb{E}(Zf(Z)) - \mathbb{E}(f'(Z)) = \mathbb{P}(Z \le z) - \mathbb{P}(G \le z)$$

这就说明了 $Z \overset{\text{law}}{=} G$。 □

【习题 4.8】 设 $\{X_r : 1 \le r \le n\}$ 是独立同分布且方差有限的随机变量列，记

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

求协方差 $\mathrm{Cov}(\overline{X}, X_k - \overline{X})$。

证明.

$$
\begin{aligned}
&\mathrm{Cov}\,(\bar{X}, X_k - \bar{X}) \\
=&\mathrm{Cov}\,(\bar{X}, X_k) - \mathrm{Cov}\,(\bar{X}, \bar{X}) \\
=&\mathrm{Cov}\left(\frac{X_k}{n}, X_k\right) - \mathrm{Var}(\bar{X}) \\
=&\frac{\mathrm{Var}(X_k)}{n} - \frac{1}{n^2}\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) \\
=&\frac{\mathrm{Var}(X_k)}{n} - \frac{1}{n}\mathrm{Var}(X_k) \\
=&0.
\end{aligned}
$$

□

【习题 4.9】 设 X 为非负随机变量，证明对任意 $r > 0$ 都有

$$\mathbb{E}[X^r] = \int_0^{\infty} rt^{r-1}\mathbb{P}(X \ge t)\,dt.$$

证明.

$$\begin{aligned}\mathbb{E}[X^r] &= \mathbb{E}\left[\int_0^X rt^{r-1}\,dt\right]\\ &= \mathbb{E}\left[\int_0^{+\infty} rt^{r-1}\mathbb{1}_{t\le X}\,dt\right]\\ &= \int_0^{\infty} rt^{r-1}\mathbb{E}[\mathbb{1}_{t\le X}]\,dt \qquad \text{(Fubini)}\\ &= \int_0^{\infty} rt^{r-1}\mathbb{P}(X\ge t)\,dt.\end{aligned}$$

□

【习题 4.10】 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)

$$\mathbb{E}[(X+Y)(X-Y)] = \mathbb{E}X^2 - \mathbb{E}Y^2 = 0.$$

$$\mathbb{E}[X+Y]\mathbb{E}[X-Y] = \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 = 0.$$

故不相关.
当X, Y独立同分布于$B(1, \frac{1}{2})$时,$X+Y$与$X-Y$不独立.
(2)

$$\begin{aligned}&f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-\frac{1}{2}x^2-\frac{1}{2}y^2}\\ &\text{令 } u = x+y \quad v = x-y\\ &J = \begin{pmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{pmatrix}\\ &|\det J| = \frac{1}{2}\\ &f_{U,V}(u,v) = \frac{1}{4\pi}e^{-\frac{1}{4}u^2-\frac{1}{4}v^2}\\ &\qquad\qquad = \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}}e^{-\frac{1}{2}\left(\frac{u}{\sqrt{2}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\cdot\sqrt{2}}e^{-\frac{1}{2}\left(\frac{v}{\sqrt{2}}\right)^2}.\end{aligned}$$

因此 U, V独立。 □

§4.3 习题3.3

【习题 4.11】 设 (X, Y) 服从二元标准正态分布, 求:(1) X + Y 与 X - Y 的联合密度函数及边缘密度函数;(2) E[X - Y | X + Y] 和 Var(X - Y | X + Y).

证明. 记$U = X + Y, V = X - Y$,并设

$$(X, Y) \sim N(0, \Sigma).$$

由于二维标准正态,记$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.记$D = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.则,

$$(U, V) = (X, Y)D.$$

由讲义定理3.3.3,

$$(U, V) \sim N(0, D^T \Sigma D) := N(0, \Sigma').$$

经计算,$\Sigma' = \begin{pmatrix} 2+2\rho & 0 \\ 0 & 2-2\rho \end{pmatrix}$. 故$U, V$独立,$U \sim N(0, 2+2\rho), V \sim N(0, 2-2\rho)$.

(2) 既然U, V独立,

$$\mathbb{E}[V|U] = \mathbb{E}[V] = 0. Var(V|U) = Var(V) = 2 - 2\rho.$$

□

【习题 4.12】 3. 设 $X = (X_1, X_2, \cdots, X_n)$ 服从多元正态 $N(0, \Sigma)$，这里正定矩阵 $\Sigma = (\sigma_{ij})_{i,j=1}^n$。证明

$$U = \sum_{k=1}^n a_k X_k \text{ 与 } V = \sum_{k=1}^n b_k X_k \text{ 独立当且仅当 } \sum_{j,k=1}^n a_j b_k \sigma_{jk} = 0,$$

其中 $a_1, \cdots, a_n, b_1, \cdots, b_n$ 为实数。并在 $b_1, \cdots, b_n$ 不全为零时，求条件期望 $\mathbb{E}[U \mid V]$。

证明. (1)多元正态分布的分量的线性组合仍构成多元正态分布.故(U, V)满足二元正态分布, 显然

$$\mathbb{E}[U] = \mathbb{E}[V] = 0.$$

$$\begin{aligned}
&U, V\text{独立} \\
\Leftrightarrow &\operatorname{Cov}(U, V) = 0 \\
\Leftrightarrow &\mathbb{E}[UV] = 0 \\
\Leftrightarrow &\sum_{j,k=1}^n a_j b_k \mathbb{E}[X_j X_k] = 0 \\
\Leftrightarrow &\sum_{j,k=1}^n a_j b_k \sigma_{jk} = 0
\end{aligned}$$

(2) 对于正态分布，因为不相关和独立性等价，所以

$$\operatorname{Cov}(U - \frac{\operatorname{Cov}(U, V)}{\operatorname{Var}(V)} V, V) = 0$$

进而有 $U-\frac{\mathrm{Cov}(U,V)}{\mathrm{Var}(V)}V$与 V独立。于是 $\mathbb{E}(U-\frac{\mathrm{Cov}(U,V)}{\mathrm{Var}(V)}V|V)=\mathbb{E}(U-\frac{\mathrm{Cov}(U,V)}{\mathrm{Var}(V)}V)=0$
由条件期望的线性性，得到以下

$$\mathbb{E}[U\mid V]=\frac{\mathrm{Cov}(U,V)}{\mathrm{Var}(V)}V.$$

代入计算即可得：

$$\mathbb{E}[U\mid V]=\frac{\sum_{j,k}a_jb_k\sigma_{jk}}{\sum_{j,k}b_jb_k\sigma_{jk}}V$$

□

【习题 4.13】 设 $X_1,X_2,\cdots,X_n$ 独立同 $N(\mu,\sigma^2),\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$, 求 $\rho(X_1,\overline{X})$。

<u>**证明.**</u> 由题意，$X_1,X_2,\ldots,X_n$ 独立同分布，$E(X_i)=\mu$，$\mathrm{Var}(X_i)=\sigma^2$。

$$\overline{X}=\frac{1}{n}\sum_{j=1}^n X_j$$

首先计算协方差：

$$\mathrm{Cov}(X_1,\overline{X})=\mathrm{Cov}\left(X_1,\frac{1}{n}\sum_{j=1}^n X_j\right)=\frac{1}{n}\sum_{j=1}^n\mathrm{Cov}(X_1,X_j)$$

由于独立性，当 $j\neq1$ 时，$\mathrm{Cov}(X_1,X_j)=0$；当 $j=1$ 时，$\mathrm{Cov}(X_1,X_1)=\mathrm{Var}(X_1)=\sigma^2$。

因此：

$$\mathrm{Cov}(X_1,\overline{X})=\frac{1}{n}\cdot\sigma^2=\frac{\sigma^2}{n}$$

再计算方差：

$$\mathrm{Var}(\overline{X})=\frac{1}{n^2}\sum_{j=1}^n\mathrm{Var}(X_j)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}$$

$$\mathrm{Var}(X_1)=\sigma^2$$

故相关系数：

$$\rho(X_1,\overline{X})=\frac{\mathrm{Cov}(X_1,\overline{X})}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(\overline{X})}}=\frac{\frac{\sigma^2}{n}}{\sqrt{\sigma^2\cdot\frac{\sigma^2}{n}}}=\frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{\sqrt{n}}}=\frac{1}{\sqrt{n}}$$

□

【习题 4.14】 若 e 为 $\mathbb{R}^n(n\geq2)$ 中一个固定的单位向量，随机向量 X 服从 n 元正态分布 $N(0,I_n)$，这里 I_n 为单位矩阵。记 Z 为 e 在 X 所在直线上投影长度的平方，试求 Z 的密度函数。

<u>**证明.**</u> 投影向量长度的平方为：

$$Z=\frac{(e^\top X)^2}{\|X\|^2}$$

由于 $X\sim N(0,I_n)$，作正交变换（这样变换是可行的因为X球对称）使 e 变为第一个坐标轴方向 $(1,0,\ldots,0)^\top$。记 $Y=(Y_1,Y_2,\ldots,Y_n)^\top$ 为变换后的向量，则 $Y\sim N(0,I_n)$，且

$$e^\top X=Y_1,\quad\|X\|^2=\sum_{i=1}^n Y_i^2.$$

于是：

$$Z = \frac{Y_1^2}{Y_1^2 + Y_2^2 + \cdots + Y_n^2}.$$

令 $U = Y_1^2 \sim \chi_1^2$，$V = Y_2^2 + \cdots + Y_n^2 \sim \chi_{n-1}^2$，且 U 与 V 独立。则：

$$Z = \frac{U}{U+V}.$$

对于 $U \sim \chi_a^2$，$V \sim \chi_b^2$ 独立，有：

$$\frac{U}{U+V} \sim \text{Beta}\left(\frac{a}{2}, \frac{b}{2}\right).$$

这里 $a = 1$，$b = n-1$，故：

$$Z \sim \text{Beta}\left(\frac{1}{2}, \frac{n-1}{2}\right).$$

其密度函数为：

$$f_Z(z) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} z^{\frac{1}{2}-1}(1-z)^{\frac{n-1}{2}-1}, \quad 0 < z < 1.$$

即：

$$f_Z(z) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} z^{-1/2}(1-z)^{\frac{n-3}{2}}, \quad 0 < z < 1,$$

其他处为 0。

□

§4.4 习题3.4

【习题 4.15】 若 $Z \sim N_{\mathbb{C}}(0,1)$，证明

$$\mathbb{E}[Z^k \bar{Z}^l] = \begin{cases} k! & , \quad k = l, \\ 0 & , \quad k \neq l. \end{cases}$$

证明.

$$\begin{aligned}
&\text{利用 } f(z) = \frac{1}{\pi}\mathrm{e}^{-|z|^2}, \text{有} \\
&\mathbb{E}\left[Z^k\bar{Z}^l\right] = \int_{\mathbb{C}} \frac{1}{\pi}\mathrm{e}^{-|z|^2} z^k \bar{z}^l \,\mathrm{d}z \\
&\overset{z=re^{i\theta}}{=} \int_0^{2\pi}\int_0^{+\infty} \frac{1}{\pi}\mathrm{e}^{-r^2}\left(r\mathrm{e}^{\mathrm{i}\theta}\right)^k\left(r\mathrm{e}^{-\mathrm{i}\theta}\right)^l r\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \int_0^{2\pi}\int_0^{+\infty} \frac{1}{\pi} r^{k+l+1}\mathrm{e}^{i(k-l)\theta}\mathrm{e}^{-r^2}\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \begin{cases} 2\int_0^{+\infty} r^{2k+1}\mathrm{e}^{-r^2}\,\mathrm{d}r = \Gamma(k+1) = k! & ,k = l \\ 0 & ,k \neq l. \end{cases}
\end{aligned}$$

□

【习题 4.16】 设 $Z_1, Z_2 \sim N_{\mathbb{C}}(0,1)$ 且独立，
(1) 求 Z_1/Z_2 的密度函数。
(2) 利用习题 3.4.1，对正整数 n 求期望 $\mathbb{E}[|Z_1 - Z_2|^{2n}]$。

证明. (1) 已知 $Z_1, Z_2 \overset{\text{i.i.d.}}{\sim} N_{\mathbb{C}}(0,1)$，即

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\pi^2} e^{-|z_1|^2 - |z_2|^2}, \quad z_1, z_2 \in \mathbb{C}.$$

令

$$W = \frac{Z_1}{Z_2}, \quad V = Z_2.$$

则变换为

$$Z_1 = WV, \quad Z_2 = V.$$

该变换的 Jacobian 行列式为 $|v|^2$。因此联合密度为

$$f_{W,V}(w,v) = f_{Z_1,Z_2}(wv,v) \cdot |v|^2 = \frac{1}{\pi^2} e^{-|wv|^2 - |v|^2} \cdot |v|^2 = \frac{|v|^2}{\pi^2} e^{-|v|^2(1+|w|^2)}.$$

对 v 积分得 W 的边际密度：

$$f_W(w) = \int_{\mathbb{C}} f_{W,V}(w,v)\, dv.$$

令 $r = |v|$，则 $dv = r\,dr\,d\theta$，积分与 θ 无关：

$$f_W(w) = \frac{1}{\pi^2} \int_0^\infty \int_0^{2\pi} r^2 e^{-r^2(1+|w|^2)} \cdot r\,d\theta\,dr = \frac{2\pi}{\pi^2} \int_0^\infty r^3 e^{-r^2(1+|w|^2)}\,dr.$$

计算积分：令 $a = 1 + |w|^2$，$t = r^2$，则 $r^3 dr = \frac{t}{2} dt$，

$$\int_0^\infty r^3 e^{-ar^2} dr = \frac{1}{2} \int_0^\infty t e^{-at} dt = \frac{1}{2} \cdot \frac{1}{a^2} = \frac{1}{2a^2}.$$

代入得

$$f_W(w) = \frac{2\pi}{\pi^2} \cdot \frac{1}{2(1+|w|^2)^2} = \frac{1}{\pi(1+|w|^2)^2}.$$

(2) 由于 $Z_1, Z_2 \sim N_{\mathbb{C}}(0,1)$ 独立，有 $Z_1 - Z_2 \sim N_{\mathbb{C}}(0,2)$。令 $Y = \frac{Z_1 - Z_2}{\sqrt{2}} \sim N_{\mathbb{C}}(0,1)$。则：

$$|Z_1 - Z_2|^{2n} = (2)^n |Y|^{2n} = 2^n (Y\bar{Y})^n.$$

由习题 3.4.1（取 $k = n, l = n$）：

$$\mathbb{E}[(Y\bar{Y})^n] = \mathbb{E}[Y^n \bar{Y}^n] = n!.$$

因此：

$$\mathbb{E}[|Z_1 - Z_2|^{2n}] = 2^n \cdot n!.$$

□

【习题 4.17】 4. (GOE 的矩) 设 H 服从 GOE_n 分布，令 $a_k=\mathbb{E}[\mathrm{tr}(H^k)]$。计算出前 6 阶矩 $a_k,k=1,\ldots,6$。

<u>证明.</u> 由对称性，只考虑偶数阶矩。一般随机矩阵中 $\mathrm{tr}=\frac{1}{n}\mathrm{Tr}$，其中 Tr是迹运算，$n$是矩阵维数,由对称性，只考虑偶数阶矩。由Wick公式

$$\begin{aligned}\mathbb{E}\,\mathrm{Tr}(H^{2k}) &= \sum_{i_1,\cdots,i_{2k}}\mathbb{E}(h_{i_1i_2}h_{i_2i_3}\cdots h_{i_{2k}i_1})\\ &= \sum_{i_1,\cdots,i_{2k}}\sum_{\pi\in\mathcal{P}(2k)}\prod_{(p,q)\in\pi}\mathbb{E}(h_{i_pi_{p+1}}h_{i_qi_{q+1}})\\ &= \sum_{i_1,\cdots,i_{2k}}\sum_{\pi\in\mathcal{P}(2k)}\prod_{(p,q)\in\pi}\mathbb{E}(\delta_{i_pi_q}\delta_{i_{p+1}i_{q+1}}+\delta_{i_pi_{q+1}}\delta_{i_{p+1}i_q})\end{aligned}$$

k=1：

$$a_2=\frac{1}{n}\sum_{i_1,i_2}\mathbb{E}[h_{i_1i_2}h_{i_2i_1}]=\frac{1}{n}\sum_{i_1,i_2}\mathbb{E}[h_{i_1i_2}^2]=\frac{1}{n}(2n+(n^2-n))=n+1$$

k=2：

$$a_4=\frac{1}{n}\sum_{i_1,i_2,i_3,i_4}\mathbb{E}[h_{i_1i_2}h_{i_2i_3}h_{i_3i_4}h_{i_4i_1}].$$

Wick 公式给出 3 种配对

- $(1,2)(3,4)$：仅在 $i_1=i_3$时不为0，$i_1=i_2=i_3=i_4$时贡献4，$i_1=i_2\neq i_4$和 $i_1=i_4\neq i_2$时贡献2，$i_1\neq i_2,i_1\neq i_4$时贡献1。因此总贡献：$4n+2n(n-1)+2n(n-1)+n(n-1)^2$
- $(1,3)(2,4)$：交错配对，当仅当 $i_1=i_3,i_2=i_4$或者 $i_2=i_3,i_1=i_4$时不为0，总贡献 $4n+n(n-1)=n^2+3n$
- $(1,4)(2,3)$：同第一种，贡献也是 n^3+2n^2+n

  计算得

$$a_4=2n^2+5n+5.$$

k=3：由轮换对称性，可以分成5种情况，对应的代表配对分别是：$(1,2)(3,4)(5,6)$，$(1,2)(3,6)(4,5)$，$(1,2)(3,5)(4,6)$，$(1,3)(2,5)(4,6)$，$(1,4)(2,5)(3,6)$。每种情况有2,3,6,3,1种配对。下面，针对 $(1,2)(3,4)(5,6)$详细过程：

$\mathbb{E}h_{i_1i_2}h_{i_2i_3}\neq0$当仅当 $i_1=i_2,i_2=i_3$或者 $i_1=i_3$；
$\mathbb{E}h_{i_3i_4}h_{i_4i_5}\neq0$当仅当 $i_3=i_4,i_4=i_5$或者 $i_3=i_5$
$\mathbb{E}h_{i_5i_6}h_{i_6i_1}\neq0$当仅当 $i_5=i_6,i_6=i_1$或者 $i_1=i_5$
共有八种情况，每种情况贡献1，例如都选后者算出的关系为：$i_1=i_3=i_5,i_2,i_4,i_6$，共n^4种情况。同样算下去，结果为 $n^4+3n^2+3n^2+n$。
第二种配对：$n^4+3n^3+3n^2+n$
第三种配对：n^3+4n^2+3n
第四种配对：$3n^2+5n$
第五种配对：n^3+4n^2+3n

最后的合为：$5n^4+22n^3+52n^2+41n$。
另外M.Ledoux在A recursion formula for the moments of the Gaussian orthogonal ensemble给出了计算GOE精确矩的五项递推公式：记$\mathbb{E}(Tr(H^{2p}))=b_p^N=\sum_{k\geq 1}\eta_k(p)N^k$，

$$\begin{aligned}(p+1)\eta_k(p)=&(8p-2)\eta_{k-1}(p-1)-(4p-1)\eta_k(p-1)\\&+p(2p-3)(10p-9)\eta_k(p-2)-8(2p-3)\eta_{k-2}(p-2)\\&+8(2p-3)\eta_{k-1}(p-2)-10(2p-3)(2p-4)(2p-5)\eta_{k-1}(p-3)\\&+5(2p-3)(2p-4)(2p-5)\eta_k(p-3)\\&-2(2p-3)(2p-4)(2p-5)(2p-6)(2p-7)\eta_k(p-4)\end{aligned}$$

□

【习题 4.18】 假设 $\{x_{ij}\}_{i,j=1}^n$ 独立同分布 $N(0,1)$，令

$$X_n=(x_{ij})_{i,j=1}^n,$$

构造对称矩阵

$$H=\frac{1}{\sqrt{2}}(X_n+X_n^t),$$

试证明 H 服从 GOE 分布。

<u>**证明.**</u> 由于 $H=\frac{1}{\sqrt{2}}(X_n+X_n^t)$，其矩阵元为

$$H_{ij}=\frac{1}{\sqrt{2}}(x_{ij}+x_{ji}).$$

由 $\{x_{ij}\}$ 独立同分布于 $N(0,1)$，可计算 H 各矩阵元的分布。对角元 H_{ii}：

$$H_{ii}=\frac{1}{\sqrt{2}}(x_{ii}+x_{ii})=\sqrt{2}\,x_{ii}.$$

因为 $x_{ii}\sim N(0,1)$，所以

$$H_{ii}\sim N(0,2).$$

非对角元 H_{ij}（$i<j$）：

$$H_{ij}=\frac{1}{\sqrt{2}}(x_{ij}+x_{ji}).$$

由于 x_{ij} 与 x_{ji} 独立且均服从 $N(0,1)$，它们的和服从 $N(0,2)$，乘以 $1/\sqrt{2}$ 后得

$$H_{ij}\sim N(0,1).$$

独立元共 n 个对角元与 $\frac{n(n-1)}{2}$ 个非对角元，总计 $\frac{n(n+1)}{2}$ 个。
联合密度为各独立元密度的乘积：

$$\begin{aligned}f(H)&=\prod_{i=1}^n\frac{1}{\sqrt{2\pi\cdot 2}}\exp\left(-\frac{H_{ii}^2}{4}\right)\times\prod_{i<j}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{H_{ij}^2}{2}\right)\\&=2^{-\frac{n}{2}}(2\pi)^{-\frac{n(n+1)}{4}}\exp\left(-\frac{1}{4}\sum_{i=1}^nH_{ii}^2-\frac{1}{2}\sum_{i<j}H_{ij}^2\right).\end{aligned}$$

注意到

$$\frac{1}{4}\sum_{i=1}^{n} H_{ii}^2 + \frac{1}{2}\sum_{i<j} H_{ij}^2 = \frac{1}{4}\operatorname{tr}(H^2),$$

故

$$f(H) = 2^{-\frac{n}{2}}(2\pi)^{-\frac{n(n+1)}{4}} e^{-\frac{1}{4}\operatorname{tr}(H^2)}.$$

$\square$

【习题 4.19】 (GOE 的正交不变性) 设 H 服从 GOE 分布，证明任给正交矩阵 Q，QHQ^{-1} 也服从 GOE 分布。

证明. 设 $H \sim \mathrm{GOE}_n$，其联合密度为

$$f(H) = C_n \cdot e^{-\frac{1}{4}\operatorname{tr}(H^2)},$$

其中 $C_n = 2^{-\frac{n}{2}}(2\pi)^{-\frac{n(n+1)}{4}}$。

令 $\tilde{H} = QHQ^{-1} = QHQ^t$（因 Q 为正交矩阵）。

对称性：

$$\tilde{H}^t = (QHQ^t)^t = QH^tQ^t = QHQ^t = \tilde{H},$$

所以 $\tilde{H}$ 仍是实对称矩阵。

迹的不变性：

$$\operatorname{tr}(\tilde{H}^2) = \operatorname{tr}(QHQ^tQHQ^t) = \operatorname{tr}(QH^2Q^t) = \operatorname{tr}(H^2Q^tQ) = \operatorname{tr}(H^2).$$

因此密度函数中的指数部分不变：

$$\exp\left(-\frac{1}{4}\operatorname{tr}(\tilde{H}^2)\right) = \exp\left(-\frac{1}{4}\operatorname{tr}(H^2)\right).$$

变换的 Jacobian 行列式：$\tilde{H} = QHQ^t$ 是一个线性变换，因此存在矩阵 A 使得 $\operatorname{Vec}(\tilde{H}) = A\operatorname{Vec}(H)$，其中 Vec 是把矩阵拉长为 n^2 维向量。一个矩阵正交当且仅当其保持任意向量的 2-范数，因此

$$\|\operatorname{Vec}(\tilde{H})\|_2^2 = \operatorname{tr}(\tilde{H}^2) = \operatorname{tr}(H^2) = \|\operatorname{Vec}(H)\|_2^2$$

因此 A 正交。故其 Jacobian 行列式的绝对值为 1。

变换后 $\tilde{H}$ 的密度为

$$f(\tilde{H}) = f(H)\cdot|\det(J)|^{-1} = f(H) = C_n \cdot e^{-\frac{1}{4}\operatorname{tr}(\tilde{H}^2)}.$$

因此 $\tilde{H} = QHQ^{-1}$ 同样服从 GOE 分布。$\square$ $\square$

§4.5 习题4.1

【习题 4.20】 对非负随机变量 X，证明

$$\sum_{n=1}^{\infty}\mathbb{P}(X \geq n) \leq \mathbb{E}[X] \leq \sum_{n=1}^{\infty}\mathbb{P}(X \geq n) + 1.$$

证明. 注意,对非负随机变量X,我们有,

$$X = \sum_{i=0}^{\infty} X\mathbb{1}_{i \le X < i+1}.$$

同时显然我们有,

$$i\mathbb{1}_{i \le X < i+1} \le X\mathbb{1}_{i \le X < i+1} \le (i+1)\mathbb{1}_{i \le X < i+1}.$$

故,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} \mathbb{E}[X\mathbb{1}_{i \le X < i+1}] \\
&\le \sum_{i=0}^{\infty} (i+1)\mathbb{E}[\mathbb{1}_{i \le X < i+1}] \\
&\le 1 + \sum_{i=0}^{\infty} i\mathbb{E}[\mathbb{1}_{i \le X < i+1}] \\
&\le 1 + \sum_{i=1}^{\infty} \sum_{j=1}^{i} \mathbb{E}[\mathbb{1}_{i \le X < i+1}] \\
&\quad \text{累和号换序.} \\
&\le 1 + \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \mathbb{E}[\mathbb{1}_{i \le X < i+1}] \\
&\le 1 + \sum_{j=1}^{\infty} \mathbb{P}(X \ge j).
\end{aligned}$$

另一侧同理. □

【习题 4.21】 **(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

$$u(x) \ge u(a) + \lambda_a(x - a), \quad \forall x \in \mathbb{R}.$$

称凸函数 u 是严格凸的，若 λ_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 $\mathbb{E}[u(X)] \ge u(\mathbb{E}[X])$。
2. 证明若 u 是严格凸的且 $\mathbb{E}[u(X)] = u(\mathbb{E}[X])$，则 X 以概率 1 为常数。

证明. (1)取$a = \mathbb{E}[X]$,由凸函数定义,

$$\mathbb{E}[u(X)] \ge \mathbb{E}[u(a) + \lambda_a(X - a)] = u(a) + \lambda_a(\mathbb{E}[X] - a) = u(a) = u(\mathbb{E}[X]).$$

(2)不等号取等当且仅当$X = \mathbb{E}[X] \quad a.e.$,即$X$以概率1为常数. □

【习题 4.22】 X 为非负随机变量，证明对任意 $r > 0$ 均有

$$\mathbb{E}[X^r] = \int_0^{\infty} r x^{r-1} \mathbb{P}(X > x)\, dx.$$

证明.

$$\begin{aligned}
\mathbb{E}[X^r] &= \mathbb{E}\int_0^X rt^{r-1}dt \\
&= \mathbb{E}\int_0^\infty rt^{r-1}\mathbb{1}_{X>t}dt \\
&\text{使用Fubini定理交换积分顺序.} \\
&= \int_0^\infty rt^{r-1}\mathbb{E}[\mathbb{1}_{X>t}]dt \\
&= \int_0^\infty rt^{r-1}\mathbb{P}(X>t)dt.
\end{aligned}$$

□

【习题 4.23】 给定 $r>0$。

1. 当 $\mathbb{E}[|X|^r]<\infty$ 时，证明
$$\lim_{x\to+\infty} x^r\mathbb{P}(|X|\geq x)=0.$$
2. 若
$$\lim_{x\to+\infty} x^r\mathbb{P}(|X|\geq x)=0,$$
证明 $\mathbb{E}[|X|^s]<\infty$，$\forall s\in(0,r)$。并问 $\mathbb{E}[|X|^r]<\infty$ 是否成立？请说明理由或给出反例。

证明. (1)易知,

$$x^r\mathbb{P}(|X|\geq x)=x^r\int_{\mathbb{R}}\mathbb{1}_{|X|\geq x}d\mathbb{P}\leq\int_{\mathbb{R}}|X|^r\mathbb{1}_{|X|\geq x}d\mathbb{P}.$$

另一方面,$|X|^r\mathbb{1}_{|X|\geq x}\leq|X|^r$,同时$|X|^r$可积($\mathbb{E}[|X|^r]<\infty$),由 DCT.

$$\lim_{n\to\infty}\int_{\mathbb{R}}|X|^r\mathbb{1}_{|X|\geq n}d\mathbb{P}=\int_{\mathbb{R}}\lim_{n\to\infty}|X|^r\mathbb{1}_{|X|\geq n}d\mathbb{P}=\int_{\mathbb{R}}0d\mathbb{P}=0.$$

综上,证毕. (2)$\forall\varepsilon>0$,记M使得$\forall x>M, x^r\mathbb{P}(|X|\geq x)<\varepsilon$. 由上题结论,

$$\begin{aligned}
\mathbb{E}[|X|^s] &= \int_0^\infty st^{s-1}\mathbb{P}(|X|>t)dt \\
&= \int_0^M st^{s-1}\mathbb{P}(|X|>t)dt+\int_M^\infty st^{s-1}\mathbb{P}(|X|>t)dt \\
&= C_1+\int_M^\infty st^{s-1}\frac{\varepsilon}{t^r}dt \\
&= C_1+C_2\varepsilon.
\end{aligned}$$

故$\mathbb{E}[|X|^s]<\infty$.
此时并不一定有$\mathbb{E}[|X|^r]<\infty$,取$\mathbb{P}(|X|\geq x)\sim\frac{1}{x^r\log x}$,利用上题公式验证即知$\mathbb{E}[|X|^r]$发散. □

§4.6
习题3.5

【习题 4.24】 标准正态分布的尾概率估计，对于 $X \sim N(0,1)$，那么

$$\frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \le \mathbb{P}(X \ge x) \le \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

证明. 右边的估计来自于 $\mathbb{E}(\mathbb{1}_{X\ge x}) \le \mathbb{E}(\frac{X}{x}\mathbb{1}_{X\ge x})$进而有：

$$\mathbb{P}\{X > x\} \le \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\frac{u}{x}e^{-u^2/2}\,du = \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

对于左边的等式，定义

$$f(x) = xe^{-x^2/2} - (x^2+1)\int_x^{\infty} e^{-u^2/2}\,du.$$

由 $f(0) < 0$ 和 $\lim_{x\to\infty} f(x) = 0$. 以及

$$f'(x) = (1-x^2+x^2+1)e^{-x^2/2} - 2x\int_x^{\infty} e^{-u^2/2}\,du = -2x\left(\int_x^{\infty} e^{-u^2/2}\,du - \frac{e^{-x^2/2}}{x}\right),$$

$x > 0$时 $f' > 0$，进而 $f(x) \le 0$。

□

(1) 定义函数 $H_n, n \ge 0, H_0 = 1, (-1)^n H_n\phi = \phi^{(n)}$，证明 $H_n(x)$ 是主项为 x^n 的 n 次多项式，且满足正交关系：

$$\int_{-\infty}^{+\infty} H_m(x)H_n(x)\phi(x)\,dx = \begin{cases} m!, & m = n, \\ 0, & m \ne n. \end{cases}$$

并证明

$$\sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!} = e^{xt-\frac{1}{2}t^2}.$$

解答. 由定义 $(-1)^n H_n(x)\phi(x) = \phi^{(n)}(x)$，其中 $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$。计算 $\phi' = -x\phi$，$\phi'' = -\phi + x^2\phi'$。通过归纳法可以得到：$H_n(x) = -H'_{n-1}(x) + xH_{n-1}(x)$。可以看出$H_n$是最高项为 x^n的首一多项式。

正交性：设 $m \ge n$，

$$\int H_m H_n \phi\,dx = (-1)^n \int H_m \phi^{(n)}\,dx.$$

分部积分 n 次，若 $n \ge m$，边界项为零得

$$(-1)^n \int H_m \phi^{(n)}\,dx = \int H_m^{(n)} \phi\,dx.$$

$H_m^{(n)}$ 是非零常数当 $m=n$ 时为 $n!$，且 $\int \phi = 1$，故得证。

由 Taylor 展开

$$\phi(x-t) = \sum_{n=0}^{\infty} \frac{(-t)^n}{n!} \phi^{(n)}(x).$$

又 $\phi(x-t) = \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} = \phi(x) e^{xt-t^2/2}$。代入 $\phi^{(n)}(x) = (-1)^n H_n(x)\phi(x)$，消去 $\phi(x)$即可。 □

(2) 对任意正整数 m,n，计算相关系数 $\rho(H_m(X), H_n(Y))$。

<u>**解答.**</u> 由(1)中母函数展开

$$e^{Xt-\frac{t^2}{2}} = \sum_{i=0}^{\infty} H_i(X) \frac{t^i}{i!}, \quad e^{Ys-\frac{s^2}{2}} = \sum_{j=0}^{\infty} H_j(Y) \frac{s^j}{j!}.$$

首先，取期望得

$$\mathbb{E}\big(e^{Xt-\frac{t^2}{2}}\big) = e^{-\frac{t^2}{2}} \mathbb{E}\big(e^{tX}\big) = e^{-\frac{t^2}{2}} \cdot e^{\frac{t^2}{2}} = 1,$$

对照展开式 $\mathbb{E}\big(\sum_i H_i(X) \frac{t^i}{i!}\big) = \sum_i \mathbb{E}(H_i(X)) \frac{t^i}{i!} = 1$，比较系数得 $\mathbb{E}(H_0(X)) = 1$，且对 $n \geq 1$ 有 $\mathbb{E}(H_n(X)) = 0$。

其次，考虑联合母函数

$$\mathbb{E}\big(e^{Xt-\frac{t^2}{2}} \cdot e^{Ys-\frac{s^2}{2}}\big) = e^{-\frac{t^2+s^2}{2}} \mathbb{E}\big(e^{tX+sY}\big).$$

由于 (X,Y) 服从标准二元正态，$tX+sY \sim N(0, t^2+s^2+2\rho ts)$，故

$$\mathbb{E}\big(e^{tX+sY}\big) = e^{\frac{1}{2}(t^2+s^2+2\rho ts)},$$

因此

$$\mathbb{E}\big(e^{Xt-\frac{t^2}{2}} \cdot e^{Ys-\frac{s^2}{2}}\big) = e^{-\frac{t^2+s^2}{2}} \cdot e^{\frac{1}{2}(t^2+s^2+2\rho ts)} = e^{\rho ts}.$$

另一方面，将母函数展开代入：

$$\mathbb{E}\left(\sum_{i=0}^{\infty} H_i(X) \frac{t^i}{i!} \sum_{j=0}^{\infty} H_j(Y) \frac{s^j}{j!}\right) = \sum_{i,j=0}^{\infty} \mathbb{E}[H_i(X)H_j(Y)] \frac{t^i s^j}{i!j!} = e^{\rho ts} = \sum_{n=0}^{\infty} \frac{(\rho ts)^n}{n!} = \sum_{n=0}^{\infty} \rho^n \frac{t^n s^n}{n!}.$$

比较 $t^i s^j$ 的系数，得到

$$\mathbb{E}[H_i(X)H_j(Y)] = \begin{cases} \rho^n \cdot n!, & i=j=n, \\ 0, & i \neq j. \end{cases}$$

结合 $\mathbb{E}(H_n(X)) = \mathbb{E}(H_n(Y)) = 0$ $(n \geq 1)$ 以及方差

$$\mathrm{Var}(H_n(X)) = \mathbb{E}[H_n^2(X)] - 0^2 = n!,$$

可得相关系数

$$\rho(H_m(X), H_n(Y)) = \frac{\mathbb{E}[H_m(X)H_n(Y)]}{\sqrt{\mathrm{Var}(H_m(X))\,\mathrm{Var}(H_n(Y))}} = \begin{cases} \rho^n, & m = n \geq 1, \\ 0, & m \neq n. \end{cases}$$

□

(3) 设 $P(x), Q(y)$ 为非常数多项式，试证明

$$|\rho(P(X), Q(Y))| \leq |\rho(X, Y)|.$$

[解答] 将 P, Q 用 Hermite 多项式展开：

$$P(x) = \sum_{i=1}^{k} a_i H_i(x), \quad Q(y) = \sum_{j=1}^{l} b_j H_j(y).$$

常数项不影响协方差。由 (2) 知

$$\mathrm{Cov}(P(X), Q(Y)) = \sum_{i=1}^{\min(k,l)} a_i b_i i! \rho^i.$$

方差

$$\mathrm{Var}(P(X)) = \sum_{i=1}^{k} a_i^2 i!, \quad \mathrm{Var}(Q(Y)) = \sum_{j=1}^{l} b_j^2 j!.$$

需证

$$\left|\sum_i a_i b_i i! \rho^i\right| \leq |\rho| \sqrt{\sum_i a_i^2 i!} \sqrt{\sum_i b_i^2 i!}.$$

由 Cauchy-Schwarz 不等式与 $|\rho|^i \leq |\rho|$（对 $i \geq 1, |\rho| < 1$）：

$$\begin{aligned} \left|\sum_{i=1} a_i b_i i! \rho^i\right| &\leq \sum_{i=1} |a_i||b_i| i! |\rho|^i \\ &\leq |\rho| \sum_{i=1} (|a_i|\sqrt{i!})(|b_i|\sqrt{i!}) \\ &\leq |\rho| \sqrt{\sum_{i=1} a_i^2 i!} \sqrt{\sum_{i=1} b_i^2 i!}. \end{aligned}$$

即得 $|\rho(P, Q)| \leq |\rho|$（这里 $|\rho| = |\rho(X, Y)|$）。

Chapter 5

第五次习题课

§5.1 习题4.2

【习题 5.1】(4.2.1) 证明如下两个不等式。

(1)（Lyapunov 不等式）对 $0 < r < s$，有

$$\big(\mathbb{E}[|X|^r]\big)^{1/r} \le \big(\mathbb{E}[|X|^s]\big)^{1/s}.$$

(2)（C_r 不等式）对 $r > 0$，有

$$\mathbb{E}[|X+Y|^r] \le C_r\big(\mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r]\big),$$

其中

$$C_r = \begin{cases} 1, & 0 < r < 1, \\ 2^{r-1}, & r \ge 1. \end{cases}$$

<u>**证明.**</u> (1) 令 $\alpha = \frac{r}{s} \in (0,1)$。由于函数 $x \mapsto x^\alpha$ 在 $[0,\infty)$ 上凹，故由 Jensen 不等式，

$$\mathbb{E}[|X|^r] = \mathbb{E}\big[(|X|^s)^\alpha\big] \le \big(\mathbb{E}[|X|^s]\big)^\alpha.$$

两边取 $1/r$ 次方即得

$$\big(\mathbb{E}[|X|^r]\big)^{1/r} \le \big(\mathbb{E}[|X|^s]\big)^{1/s}.$$

(2) 当 $0 < r < 1$ 时，对任意 $a, b \ge 0$ 有

$$(a+b)^r \le a^r + b^r,$$

因而

$$|X+Y|^r \le (|X| + |Y|)^r \le |X|^r + |Y|^r.$$

取期望得

$$\mathbb{E}[|X+Y|^r] \le \mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r].$$

当 $r \geq 1$ 时，对任意 $a, b \geq 0$，由凸性或 Jensen 不等式可得

$$(a+b)^r = 2^r \left(\frac{a+b}{2}\right)^r \leq 2^{r-1}(a^r + b^r).$$

因此

$$|X+Y|^r \leq 2^{r-1}\big(|X|^r + |Y|^r\big),$$

再取期望即得结论。 □

【习题 5.2】 **(4.2.2)** 已知 $\{X_n\}$ 为随机变量列，实数列 $\{c_n\}$ 收敛于常数 c。在几乎处处收敛、L^p 收敛、依概率收敛和依分布收敛意义下分别证明

$$X_n \to X \Longrightarrow c_n X_n \to cX.$$

证明. 若 $X_n \xrightarrow{\text{a.s.}} X$，则对几乎处处的 ω，有

$$c_n X_n(\omega) \to cX(\omega),$$

故 $c_n X_n \xrightarrow{\text{a.s.}} cX$。

若 $X_n \xrightarrow{L^p} X$，则 $X \in L^p$，且 $\{c_n\}$ 有界。由前面的 C_r 不等式，存在只依赖于 p 的常数 $C_p > 0$，使得

$$|c_n X_n - cX|^p \leq C_p\big(|c_n|^p |X_n - X|^p + |c_n - c|^p |X|^p\big).$$

两边取期望，便得

$$\mathbb{E}[|c_n X_n - cX|^p] \to 0,$$

即 $c_n X_n \xrightarrow{L^p} cX$。

若 $X_n \xrightarrow{P} X$，则

$$c_n X_n - cX = c_n(X_n - X) + (c_n - c)X.$$

由于 $\{c_n\}$ 有界，第一项依概率收敛于 0；第二项因 $c_n - c \to 0$ 为常数，故 a.s. 收敛于 0，从而也依概率收敛于 0。因此

$$c_n X_n \xrightarrow{P} cX.$$

若 $X_n \xrightarrow{D} X$，则把 c_n 看成常值随机变量，有 $c_n \xrightarrow{P} c$。由后面的 Slutsky 定理可得

$$c_n X_n \xrightarrow{D} cX.$$

□

【习题 5.3】 **(4.2.3)** 证明当 $n \to \infty$ 时，

$$X_n \xrightarrow{P} 0 \quad \Longleftrightarrow \quad \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] \to 0.$$

证明. 若 $X_n \xrightarrow{P} 0$，则对任意 $\varepsilon > 0$，

$$\begin{aligned}\mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] &= \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}; |X_n| < \varepsilon\right] + \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}; |X_n| \geq \varepsilon\right] \\ &\leq \varepsilon + \mathbb{P}(|X_n| \geq \varepsilon).\end{aligned}$$

令 $n\to\infty$ 后得

$$\limsup_{n\to\infty}\mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right]\le\varepsilon.$$

再令 $\varepsilon\downarrow 0$，便知

$$\mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right]\to 0.$$

反过来，若

$$\mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right]\to 0,$$

则对任意 $\varepsilon>0$,

$$\mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right]\ge\mathbb{E}\left[\frac{|X_n|}{1+|X_n|};|X_n|\ge\varepsilon\right]\ge\frac{\varepsilon}{1+\varepsilon}\mathbb{P}(|X_n|\ge\varepsilon).$$

因而 $\mathbb{P}(|X_n|\ge\varepsilon)\to 0$，即 $X_n\xrightarrow{P}0$。 □

【习题 5.4】(4.2.4) 随机变量列 $\{X_n\}$、$\{Y_n\}$ 满足 $X_n\xrightarrow{D}X$ 且 $Y_n\xrightarrow{P}c$，其中 X 是随机变量，c 为常数。证明

(1) $X_n+Y_n\xrightarrow{D}X+c$.

(2) $X_nY_n\xrightarrow{D}cX$，且当 $c\ne 0$ 时有

$$\frac{X_n}{Y_n}\xrightarrow{D}\frac{X}{c}.$$

<u>证明.</u> (1) 与 (2) 的前半部分正是下述 Slutsky 定理在 $Z_n\equiv 0$ 时的结论，因此

$$X_n+Y_n\xrightarrow{D}X+c,\qquad X_nY_n\xrightarrow{D}cX.$$

当 $c\ne 0$ 时，由函数 $x\mapsto 1/x$ 在 c 处连续可得

$$\frac{1}{Y_n}\xrightarrow{P}\frac{1}{c}.$$

再对随机变量列 $\{X_n\}$ 与 $\{1/Y_n\}$ 应用 Slutsky 定理，得到

$$\frac{X_n}{Y_n}=X_n\cdot\frac{1}{Y_n}\xrightarrow{D}\frac{X}{c}.$$

□

Remark（Slutsky 定理及其证明）. 设随机变量列 $\{X_n\}$、$\{Y_n\}$、$\{Z_n\}$ 满足

$$X_n\xrightarrow{D}X,\qquad Y_n\xrightarrow{P}b,\qquad Z_n\xrightarrow{P}c,$$

其中 X 为随机变量，b,c 为常数，则

$$X_nY_n+Z_n\xrightarrow{D}bX+c.$$

特别地，

$$X_n+Y_n\xrightarrow{D}X+c,\qquad X_nY_n\xrightarrow{D}bX,$$

且当 $b \neq 0$ 时，

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{b}.$$

证明. 先证一个常用引理：若

$$U_n - V_n \xrightarrow{P} 0, \qquad V_n \xrightarrow{D} V,$$

则 $U_n \xrightarrow{D} V$。

事实上，对任意 $\varepsilon > 0$ 及 V 的任一分布函数连续点 x，有

$$\{V_n \le x - \varepsilon\} \cap \{|U_n - V_n| \le \varepsilon\} \subset \{U_n \le x\}$$

与

$$\{U_n \le x\} \subset \{V_n \le x + \varepsilon\} \cup \{|U_n - V_n| > \varepsilon\}.$$

因此

$$\mathbb{P}(V_n \le x - \varepsilon) - \mathbb{P}(|U_n - V_n| > \varepsilon) \le \mathbb{P}(U_n \le x)$$

以及

$$\mathbb{P}(U_n \le x) \le \mathbb{P}(V_n \le x + \varepsilon) + \mathbb{P}(|U_n - V_n| > \varepsilon).$$

令 $n \to \infty$，得到

$$F_V(x - \varepsilon) \le \liminf_{n\to\infty} \mathbb{P}(U_n \le x) \le \limsup_{n\to\infty} \mathbb{P}(U_n \le x) \le F_V(x + \varepsilon).$$

再令 $\varepsilon \downarrow 0$，由 x 的连续性可知

$$\mathbb{P}(U_n \le x) \to F_V(x),$$

故 $U_n \xrightarrow{D} V$。

下面证明 Slutsky 定理。先看加法。由连续映射定理，

$$X_n + c \xrightarrow{D} X + c.$$

又因为

$$(X_n + Y_n) - (X_n + c) = Y_n - c \xrightarrow{P} 0,$$

由上面的引理立得

$$X_n + Y_n \xrightarrow{D} X + c.$$

再看乘法。由 $X_n \xrightarrow{D} X$ 可知 $\{X_n\}$ 是紧的。于是对任意 $\varepsilon, \eta > 0$，可取 $M > 0$ 使得对充分大的 n,

$$\mathbb{P}(|X_n| > M) < \eta.$$

从而

$$\mathbb{P}\big(|X_n(Y_n - b)| > \varepsilon\big) \le \mathbb{P}(|X_n| > M) + \mathbb{P}\left(|Y_n - b| > \frac{\varepsilon}{M}\right).$$

令 $n \to \infty$ 得

$$X_n(Y_n - b) \xrightarrow{P} 0.$$

另一方面，由连续映射定理，

$$bX_n \xrightarrow{D} bX.$$

又

$$X_nY_n - bX_n = X_n(Y_n - b) \xrightarrow{P} 0,$$

故再由引理，

$$X_nY_n \xrightarrow{D} bX.$$

最后，由 $X_nY_n \xrightarrow{D} bX$ 以及 $Z_n \xrightarrow{P} c$，将刚证明的加法情形应用于 $\{X_nY_n\}$ 与 $\{Z_n\}$，便得

$$X_nY_n + Z_n \xrightarrow{D} bX + c.$$

若 $b \neq 0$，则函数 $x \mapsto 1/x$ 在 b 处连续，所以

$$\frac{1}{Y_n} \xrightarrow{P} \frac{1}{b}.$$

再将上面的乘法结论应用于 X_n 与 $1/Y_n$，便得到

$$\frac{X_n}{Y_n} = X_n \cdot \frac{1}{Y_n} \xrightarrow{D} \frac{X}{b}.$$

□

§5.2 习题4.3

【习题 5.5】(4.3.1) 设 $\{X_n\}$ 相互独立且服从标准正态分布，利用第 3 章问题第 14 题 (1) 的结论证明

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{X_n}{\sqrt{\log n}} = \sqrt{2}\right) = 1.$$

证明. 对任意 $a > 0$，记

$$A_n(a) = \left\{X_n \geq \sqrt{2a\log n}\right\}.$$

由第 3 章问题第 14 题 (1) 的标准正态尾概率估计，存在正常数 C_1, C_2，使得对充分大的 n，

$$C_1\frac{n^{-a}}{\sqrt{\log n}} \leq \mathbb{P}(A_n(a)) \leq C_2\frac{n^{-a}}{\sqrt{\log n}}.$$

若 $0 < a < 1$，则

$$\sum_{n=2}^{\infty} \mathbb{P}(A_n(a)) = \infty.$$

由于 $\{A_n(a)\}$ 相互独立，第二 Borel-Cantelli 引理给出

$$\mathbb{P}(A_n(a) \text{ i.o.}) = 1.$$

这说明

$$\limsup_{n\to\infty}\frac{X_n}{\sqrt{\log n}}\geq\sqrt{2a}\qquad\text{a.s.}$$

若 $a>1$，则

$$\sum_{n=2}^{\infty}\mathbb{P}(A_n(a))<\infty,$$

故由第一 Borel-Cantelli 引理，

$$\mathbb{P}(A_n(a)\text{ i.o.})=0,$$

即

$$\limsup_{n\to\infty}\frac{X_n}{\sqrt{\log n}}\leq\sqrt{2a}\qquad\text{a.s.}$$

因此对任意 $0<a<1<b$，几乎处处都有

$$\sqrt{2a}\leq\limsup_{n\to\infty}\frac{X_n}{\sqrt{\log n}}\leq\sqrt{2b}.$$

令 $a\uparrow 1$、$b\downarrow 1$，便得

$$\limsup_{n\to\infty}\frac{X_n}{\sqrt{\log n}}=\sqrt{2}\qquad\text{a.s.}$$

□

【习题 5.6】(4.3.6) 设随机变量 $X_1,\cdots,X_n$ 独立同分布，且服从 $[0,a]$ 上的均匀分布，其中 $a>0$。记

$$M_n=\max\{X_1,\cdots,X_n\},$$

分别在 a.s.、p 阶收敛的意义下证明当 $n\to\infty$ 时 $M_n\to a$。

证明. 对任意 $0<\varepsilon<a$，有

$$\mathbb{P}(|M_n-a|>\varepsilon)=\mathbb{P}(M_n<a-\varepsilon)=\left(\frac{a-\varepsilon}{a}\right)^n.$$

因为

$$\sum_{n=1}^{\infty}\left(\frac{a-\varepsilon}{a}\right)^n<\infty,$$

由第一 Borel-Cantelli 引理可知

$$|M_n-a|>\varepsilon$$

只会发生有限次。对有理数 $\varepsilon>0$ 取可数交，即得

$$M_n\xrightarrow{\text{a.s.}}a.$$

又因为 $0\leq M_n\leq a$，故

$$|M_n-a|^p\leq a^p.$$

结合 $M_n\xrightarrow{\text{a.s.}}a$，由 DCT 可得

$$\mathbb{E}[|M_n-a|^p]\to 0.$$

因而

$$M_n \xrightarrow{L^p} a.$$

□

【习题 5.7】(4.3.7) 随机变量列 $\{X_n\}$ 满足 $X_n \xrightarrow{P} X$。证明存在子列 $\{X_{n_k}\}$ 满足

$$X_{n_k} \xrightarrow{\text{a.s.}} X.$$

证明. 由于 $X_n \xrightarrow{P} X$，对每个 $k \in \mathbb{N}^*$ 都可取 $n_k > n_{k-1}$，使得

$$\mathbb{P}\left(|X_{n_k} - X| > 2^{-k}\right) < 2^{-k}.$$

于是

$$\sum_{k=1}^{\infty} \mathbb{P}\left(|X_{n_k} - X| > 2^{-k}\right) < \infty.$$

由第一 Borel-Cantelli 引理，事件

$$|X_{n_k} - X| > 2^{-k}$$

只会发生有限次。故几乎处处存在 $K(\omega)$，使得当 $k \geq K(\omega)$ 时，

$$|X_{n_k}(\omega) - X(\omega)| \leq 2^{-k}.$$

于是 $X_{n_k}(\omega) \to X(\omega)$，即

$$X_{n_k} \xrightarrow{\text{a.s.}} X.$$

□

【习题 5.8】(4.3.8) (1) 设 $\{X_n\}$ 是相互独立的实值随机变量列且满足 $X_n \xrightarrow{P} 0$，$\{a_n\}$ 为单调递增至 $+\infty$ 的正实数列。问

$$\frac{X_n}{a_n} \xrightarrow{\text{a.s.}} 0$$

是否成立？

(2) 设 $\{X_n\}$ 是实值随机变量列，试构造正实数列 $\{c_n\}$，使得

$$\frac{X_n}{c_n} \xrightarrow{\text{a.s.}} 0.$$

证明. (1) 结论不一定成立。对给定的 $\{a_n\}$，定义独立随机变量

$$\mathbb{P}(X_n = a_n) = \frac{1}{n+1}, \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n+1}.$$

由于 $a_n \to \infty$，对任意 $\varepsilon > 0$，当 n 充分大时 $a_n > \varepsilon$，故

$$\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{n+1} \to 0,$$

即 $X_n \xrightarrow{P} 0$。但是

$$\mathbb{P}\left(\frac{X_n}{a_n} = 1\right) = \frac{1}{n+1}, \qquad \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty.$$

由第二 Borel-Cantelli 引理，

$$\frac{X_n}{a_n} = 1$$

会发生无穷多次，故 $\frac{X_n}{a_n}$ 不 a.s. 收敛于 0。

(2) 对每个 n，由 $\mathbb{P}(|X_n| > t) \downarrow 0$ $(t \to \infty)$，可取 $c_n > 0$ 使得

$$\mathbb{P}(|X_n| > 2^{-n} c_n) < 2^{-n}.$$

记

$$A_n = \{|X_n| > 2^{-n} c_n\}.$$

则

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

由第一 Borel-Cantelli 引理，A_n 只会发生有限次。故几乎处处存在 $N(\omega)$，使得当 $n \geq N(\omega)$ 时，

$$\left| \frac{X_n}{c_n} \right| \leq 2^{-n}.$$

因此

$$\frac{X_n}{c_n} \xrightarrow{\text{a.s.}} 0.$$

□

§5.3 习题4.4

【习题 5.9】(4.4.1) $\{X_n\}$ 为非负独立同分布随机变量列，$\mathbb{E}[X_1] = +\infty$，证明

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} +\infty.$$

证明. 对每个 $M > 0$，令

$$Y_k^{(M)} = X_k \wedge M.$$

则 $\{Y_k^{(M)}\}$ 仍是非负独立同分布随机变量列，且 $\mathbb{E}[Y_1^{(M)}] < \infty$。由强大数定律，

$$\frac{1}{n} \sum_{k=1}^{n} Y_k^{(M)} \xrightarrow{\text{a.s.}} \mathbb{E}[Y_1^{(M)}].$$

又因 $X_k \geq Y_k^{(M)}$，故

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k \geq \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k^{(M)} = \mathbb{E}[Y_1^{(M)}] \qquad \text{a.s.}$$

由于 $Y_1^{(M)} \uparrow X_1$，由 MCT，

$$\mathbb{E}[Y_1^{(M)}] \uparrow \mathbb{E}[X_1] = +\infty.$$

因而对任意 $L>0$，可取 M 充分大使得 $\mathbb{E}[Y_1^{(M)}]\geq L$。于是

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k\geq L\qquad \text{a.s.}$$

由于 L 任意，得到

$$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{\text{a.s.}}+\infty.$$

□

【习题 5.10】(4.4.2) （Weierstrass 逼近定理）任给连续函数 $f:[0,1]\to\mathbb{R}$，随机变量 S_n 服从二项分布 $B(n,x)$，证明

$$\lim_{n\to+\infty}\sup_{0\leq x\leq 1}\left|f(x)-\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}\right|=0.$$

证明. 对固定的 $x\in[0,1]$，令 $S_n\sim B(n,x)$。则

$$\mathbb{P}(S_n=k)=\binom{n}{k}x^k(1-x)^{n-k},$$

因而

$$\sum_{k=0}^{n}f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}=\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right].$$

于是只需证明

$$\sup_{0\leq x\leq 1}\left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]-f(x)\right|\to 0.$$

由于 f 在 $[0,1]$ 上连续，故一致连续。任给 $\varepsilon>0$，存在 $\delta>0$，使得当 $|u-v|<\delta$ 时，

$$|f(u)-f(v)|<\varepsilon.$$

记 $M=\sup_{0\leq y\leq 1}|f(y)|$。则

$$\begin{aligned}\left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]-f(x)\right|&\leq \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right)-f(x)\right|;\left|\frac{S_n}{n}-x\right|<\delta\right]\\&\quad+\mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right)-f(x)\right|;\left|\frac{S_n}{n}-x\right|\geq\delta\right]\\&\leq\varepsilon+2M\,\mathbb{P}\left(\left|\frac{S_n}{n}-x\right|\geq\delta\right).\end{aligned}$$

由 Chebyshev 不等式，

$$\mathbb{P}\left(\left|\frac{S_n}{n}-x\right|\geq\delta\right)\leq\frac{\mathrm{Var}(S_n/n)}{\delta^2}=\frac{x(1-x)}{n\delta^2}\leq\frac{1}{4n\delta^2}.$$

因而

$$\sup_{0\leq x\leq 1}\left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]-f(x)\right|\leq\varepsilon+\frac{M}{2n\delta^2}.$$

令 $n\to\infty$ 后再令 $\varepsilon\downarrow 0$，即得结论。 □

【习题 5.11】(4.4.3) 随机变量 $X_1, \cdots, X_n$ 独立同分布，满足 $\mathbb{E}[X_1] = 0$、$\mathbb{E}[X_1^4] < \infty$。不利用强大数定律的结论，直接证明

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} 0.$$

证明. 记

$$S_n = \sum_{k=1}^{n} X_k.$$

由于 $\mathbb{E}[X_1] = 0$，由独立性展开四阶矩可得

$$\mathbb{E}[S_n^4] = n\mathbb{E}[X_1^4] + 6\binom{n}{2}\left(\mathbb{E}[X_1^2]\right)^2 = O(n^2).$$

因而存在常数 $C > 0$，使得对所有 n,

$$\mathbb{E}[S_n^4] \le Cn^2.$$

由 Markov 不等式，

$$\mathbb{P}\big(|S_n| > n\varepsilon\big) \le \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4} \le \frac{C}{n^2\varepsilon^4}.$$

所以

$$\sum_{n=1}^{\infty} \mathbb{P}\big(|S_n| > n\varepsilon\big) < \infty.$$

由第一 Borel-Cantelli 引理，

$$\mathbb{P}\big(|S_n| > n\varepsilon \text{ i.o.}\big) = 0.$$

由于 $\varepsilon > 0$ 任意，得到

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0.$$

□

【习题 5.12】(4.4.4) 设 $\{X_n\}$ 相互独立且服从参数为 1 的指数分布。

(1) 证明 $(X_1 \cdots X_n)^{1/n}$ 几乎处处收敛，并求出收敛值。

(2) 探索

$$\frac{n}{\frac{1}{X_1} + \cdots + \frac{1}{X_n}}$$

的极限分布。

证明. (1) 令 $Y_n = \log X_n$。由于 $X_n \sim \operatorname{Exp}(1)$，有

$$\mathbb{E}[|Y_1|] < \infty, \qquad \mathbb{E}[Y_1] = \int_0^{\infty} (\log x)e^{-x}\,dx = -\gamma,$$

其中 γ 为 Euler 常数。由强大数定律，

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{\text{a.s.}} -\gamma.$$

因而

$$(X_1\cdots X_n)^{1/n}=\exp\left(\frac{1}{n}\sum_{k=1}^{n}Y_k\right)\xrightarrow{\text{a.s.}}e^{-\gamma}.$$

(2) 令

$$Z_k=\frac{1}{X_k}.$$

则 $Z_k\geq 0$ 且 $\{Z_k\}$ 独立同分布。又

$$\mathbb{E}[Z_1]=\int_0^\infty\frac{1}{x}e^{-x}\,dx=+\infty.$$

由上一题结论，

$$\frac{1}{n}\sum_{k=1}^{n}Z_k\xrightarrow{\text{a.s.}}+\infty.$$

于是

$$\frac{n}{\frac{1}{X_1}+\cdots+\frac{1}{X_n}}=\left(\frac{1}{n}\sum_{k=1}^{n}Z_k\right)^{-1}\xrightarrow{\text{a.s.}}0.$$

因而它的极限分布为退化分布 δ_0。 □

【习题 5.13】 **(4.4.5)** 区间 $[0,1]$ 被划分成 n 个互不相交的子区间之并，子区间长度分别为 $p_1,p_2,\cdots,p_n$，定义该划分的熵为

$$h=-\sum_{i=1}^{n}p_i\log p_i.$$

设 $X_1,X_2,\cdots,X_m$ 是相互独立且均服从 $[0,1]$ 上均匀分布的随机变量，定义 $Z_m(i)$ 是 $X_1,\cdots,X_m$ 中位于第 i 个区间的总数，

$$R_m=\prod_{i=1}^{n}p_i^{Z_m(i)}.$$

证明当 $m\to\infty$ 时，

$$\frac{\log R_m}{m}\xrightarrow{\text{a.s.}}-h.$$

证明. 对每个 k，定义

$$Y_k=\sum_{i=1}^{n}(\log p_i)\mathbf{1}_{\{X_k\text{ 落在第 }i\text{ 个区间}\}}.$$

则 $\{Y_k\}$ 独立同分布，且

$$\mathbb{P}(Y_k=\log p_i)=p_i,\qquad 1\leq i\leq n.$$

于是

$$\mathbb{E}[Y_1]=\sum_{i=1}^{n}p_i\log p_i=-h.$$

另一方面，

$$\log R_m = \sum_{i=1}^{n} Z_m(i) \log p_i = \sum_{k=1}^{m} Y_k.$$

由强大数定律，

$$\frac{\log R_m}{m} = \frac{1}{m}\sum_{k=1}^{m} Y_k \xrightarrow{\text{a.s.}} \mathbb{E}[Y_1] = -h.$$

□

【习题 5.14】(4.4.7) 随机变量列 $\{X_k : k \geq 2\}$ 相互独立且满足

$$\mathbb{P}(X_k = 2k) = \mathbb{P}(X_k = -2k) = \frac{1}{2k\log k}, \qquad \mathbb{P}(X_k = 0) = 1 - \frac{1}{k\log k}.$$

记

$$S_n = X_2 + \cdots + X_n,$$

证明

$$\frac{S_n}{n} \xrightarrow{P} 0, \qquad \frac{S_n}{n(n-1)} \xrightarrow{\text{a.s.}} 0,$$

但

$$\frac{S_n}{n}$$

不 a.s. 收敛于 0。

证明. 先注意到

$$\mathbb{E}[X_k] = 0,$$

且

$$\mathbb{E}[X_k^2] = 4k^2 \cdot \frac{1}{k\log k} + 1 - \frac{1}{k\log k} \leq C\frac{k}{\log k}.$$

因此

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{k=2}^{n} \mathbb{E}[X_k^2].$$

又有

$$\sum_{k=2}^{n} \frac{k}{\log k} \leq \sum_{k\leq\sqrt{n}} \frac{k}{\log 2} + \sum_{k>\sqrt{n}} \frac{2k}{\log n} = O\left(\frac{n^2}{\log n}\right),$$

故

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = O\left(\frac{1}{\log n}\right) \to 0.$$

由 Chebyshev 不等式，

$$\frac{S_n}{n} \xrightarrow{P} 0.$$

再看几乎处处收敛。由上面的估计，

$$\operatorname{Var}(S_n) = \sum_{k=2}^{n} \mathbb{E}[X_k^2] = O\left(\frac{n^2}{\log n}\right).$$

故对任意 $\varepsilon > 0$，由 Chebyshev 不等式，

$$\mathbb{P}\left(\left|\frac{S_n}{n(n-1)}\right| > \varepsilon\right) \le \frac{\mathrm{Var}(S_n)}{\varepsilon^2 n^2 (n-1)^2} = O\left(\frac{1}{n^2 \log n}\right).$$

因而

$$\sum_{n=2}^{\infty} \mathbb{P}\left(\left|\frac{S_n}{n(n-1)}\right| > \varepsilon\right) < \infty.$$

由第一 Borel-Cantelli 引理，

$$\frac{S_n}{n(n-1)} \xrightarrow{\text{a.s.}} 0.$$

最后证明 $\frac{S_n}{n}$ 不 a.s. 收敛于 0。记

$$A_n = \{X_n = 2n\}.$$

则 $\{A_n\}$ 相互独立，且

$$\sum_{n=2}^{\infty} \mathbb{P}(A_n) = \sum_{n=2}^{\infty} \frac{1}{2n \log n} = \infty.$$

由第二 Borel-Cantelli 引理，A_n a.s. 无穷多次发生。若假设

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0,$$

则

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

但在 A_n 上，

$$\frac{S_n}{n} = \frac{S_{n-1}}{n} + 2.$$

由于 A_n 无穷多次发生，这与 $\frac{S_n}{n} \to 0$ 矛盾。故

$$\frac{S_n}{n}$$

不 a.s. 收敛于 0。 □

§5.4 习题5.1

【习题 5.15】(5.1.1) 随机变量 X 的密度

$$f(x) = \frac{1}{2} e^{-|x|}, \qquad -\infty < x < \infty,$$

求 X 的特征函数。

证明.

$$\begin{aligned}\phi_X(t) = \mathbb{E}[e^{itX}] &= \frac{1}{2}\int_{-\infty}^{\infty} e^{itx-|x|}\,dx \\ &= \int_0^{\infty} e^{-x}\cos(tx)\,dx = \frac{1}{1+t^2}.\end{aligned}$$

□

【习题 5.16】(5.1.2) 已知 $\{U,V\}$ 与 $\{X,Y\}$ 独立，令

$$Z = \frac{UX+VY}{\sqrt{U^2+V^2}}.$$

证明若 X 与 Y 独立且均服从 $N(0,1)$，则 $Z \sim N(0,1)$。若 (X,Y) 服从标准二元正态分布，上述结论是否成立？

证明. 若 X,Y 独立且都服从 $N(0,1)$，则对任意固定的 $(u,v) \in \mathbb{R}^2$，

$$uX+vY \sim N(0,u^2+v^2).$$

因此在 $(U,V)=(u,v)$ 条件下，

$$Z \mid (U,V)=(u,v) \sim N(0,1).$$

也就是说，对任意 $t \in \mathbb{R}$,

$$\mathbb{E}\left[e^{itZ} \mid U,V\right] = e^{-t^2/2}.$$

再取期望，得到

$$\mathbb{E}[e^{itZ}] = e^{-t^2/2},$$

故 $Z \sim N(0,1)$。

若 (X,Y) 只服从标准二元正态分布而不要求独立，则结论一般不成立。设

$$\mathrm{Cov}(X,Y) = \rho \neq 0,$$

并取 $U=V=1$（常数），则

$$Z = \frac{X+Y}{\sqrt{2}}.$$

于是

$$\mathrm{Var}(Z) = \frac{1}{2}\mathrm{Var}(X+Y) = \frac{1}{2}(1+1+2\rho) = 1+\rho \neq 1.$$

所以此时 $Z \not\sim N(0,1)$。因此一般情形下结论不成立。 □

【习题 5.17】(5.1.3) 记

$$\phi(t) = \left(\frac{\sin t}{t}\right)^2.$$

试用概率方法证明对实数 $t_1,\cdots,t_n$，矩阵

$$H_n = \big(\phi(t_i-t_j)\big)_{i,j=1}^n$$

非负定。

证明. 取独立同分布随机变量 $X, Y \sim U[-1,1]$。则

$$\phi_X(t) = \phi_Y(t) = \frac{\sin t}{t}.$$

因而

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \left(\frac{\sin t}{t}\right)^2 = \phi(t).$$

这说明 ϕ 是某个随机变量（即 $X+Y$）的特征函数。

于是对任意复数 $c_1, \cdots, c_n$，有

$$\begin{aligned}
\sum_{i,j=1}^{n} c_i \overline{c_j} \phi(t_i - t_j) &= \sum_{i,j=1}^{n} c_i \overline{c_j}\, \mathbb{E}\left[e^{i(t_i - t_j)(X+Y)}\right] \\
&= \mathbb{E}\left[\left|\sum_{j=1}^{n} c_j e^{it_j(X+Y)}\right|^2\right] \\
&\geq 0.
\end{aligned}$$

故矩阵 H_n 非负定。 □

【习题 5.18】**(5.1.5)** 设 $X_1, X_2, \cdots, X_n$ 为一族相互独立的随机变量，令

$$Y_n = X_1^2 + X_2^2 + \cdots + X_n^2.$$

(1) 假设 $X_i \sim N(i, 1)$，试求 Y_n 的特征函数。

(2) 假设 $X_i \sim N(1,1)$。若随机变量 $N \sim P(\lambda)$，且 N 与 X_i（$i = 1, 2, \cdots$）均独立，试求 Y_N 的特征函数。

证明. 若 $X \sim N(\mu, 1)$，则

$$\begin{aligned}
\mathbb{E}[e^{itX^2}] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(itx^2 - \frac{(x-\mu)^2}{2}\right) dx \\
&= \frac{1}{\sqrt{1-2it}} \exp\left(\frac{i\mu^2 t}{1-2it}\right).
\end{aligned}$$

(1) 由独立性，

$$\phi_{Y_n}(t) = \prod_{k=1}^{n} \mathbb{E}[e^{itX_k^2}] = (1-2it)^{-n/2} \exp\left(\frac{it}{1-2it} \sum_{k=1}^{n} k^2\right).$$

即

$$\phi_{Y_n}(t) = (1-2it)^{-n/2} \exp\left(\frac{it}{1-2it} \cdot \frac{n(n+1)(2n+1)}{6}\right).$$

(2) 此时

$$\phi_{X_1^2}(t) = (1-2it)^{-1/2} \exp\left(\frac{it}{1-2it}\right).$$

条件于 $N = m$ 时，

$$\phi_{Y_N | N=m}(t) = \phi_{X_1^2}(t)^m.$$

因此

$$\phi_{Y_N}(t) = \mathbb{E}\left[\phi_{X_1^2}(t)^N\right] = \exp\left\{\lambda(\phi_{X_1^2}(t) - 1)\right\}.$$

即

$$\phi_{Y_N}(t) = \exp\left\{\lambda\left((1-2it)^{-1/2}\exp\left(\frac{it}{1-2it}\right) - 1\right)\right\}.$$

□

【习题 5.19】(5.1.7) 设 $X_1, \cdots, X_n$ 独立同分布，记

$$S_n = X_1 + \cdots + X_n.$$

(1) 若矩母函数 $M(t) = \mathbb{E}[e^{tX_1}]$ 存在，证明尾概率估计

$$\mathbb{P}(X_1 \geq a) \leq \inf_{t>0}\{e^{-at}M(t)\}.$$

(2) 若 $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$，试证明对任意 $a > 0$ 均有

$$\mathbb{P}(S_n \geq a) \leq e^{-\frac{a^2}{2n}}.$$

<u>**证明.**</u> (1) 对任意 $t > 0$，由 Markov 不等式，

$$\mathbb{P}(X_1 \geq a) = \mathbb{P}(e^{tX_1} \geq e^{ta}) \leq e^{-ta}\mathbb{E}[e^{tX_1}] = e^{-ta}M(t).$$

再对 $t > 0$ 取下确界即得

$$\mathbb{P}(X_1 \geq a) \leq \inf_{t>0}\{e^{-at}M(t)\}.$$

(2) 对任意 $t > 0$，由 (1) 作用于 S_n 得

$$\mathbb{P}(S_n \geq a) \leq e^{-at}\mathbb{E}[e^{tS_n}] = e^{-at}\left(\mathbb{E}[e^{tX_1}]\right)^n.$$

又

$$\mathbb{E}[e^{tX_1}] = \frac{e^t + e^{-t}}{2} = \cosh t.$$

并且

$$\cosh t = \sum_{m=0}^{\infty}\frac{t^{2m}}{(2m)!} \leq \sum_{m=0}^{\infty}\frac{(t^2/2)^m}{m!} = e^{t^2/2}.$$

因而

$$\mathbb{P}(S_n \geq a) \leq \exp\left(-at + \frac{nt^2}{2}\right).$$

取 $t = \frac{a}{n}$，得到

$$\mathbb{P}(S_n \geq a) \leq e^{-a^2/(2n)}.$$

□

【习题 5.20】(5.1.8) 若随机变量 X 的尾部概率对某正常数 K 满足

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/K^2}, \qquad \forall t \geq 0,$$

则称 X 为次高斯随机变量。证明

(1) 若 X 的矩母函数满足

$$\mathbb{E}[e^{sX}] \leq e^{s^2/2}, \qquad \forall s \in \mathbb{R},$$

则 X 为次高斯随机变量。

(2) 次高斯随机变量的矩满足不等式

$$\mathbb{E}[|X|^p] \leq (K_1\sqrt{p})^p, \qquad \forall p \geq 1.$$

这里 K_1 为不依赖 p 的正常数。提示：必要时可以利用 Stirling 公式

$$n! \sim n^n e^{-n}\sqrt{2\pi n}.$$

<u>**证明.**</u> (1) 对任意 $s, t > 0$，由 Markov 不等式，

$$\mathbb{P}(X \geq t) = \mathbb{P}(e^{sX} \geq e^{st}) \leq e^{-st}\mathbb{E}[e^{sX}] \leq e^{-st+s^2/2}.$$

取 $s = t$，得

$$\mathbb{P}(X \geq t) \leq e^{-t^2/2}.$$

同理对 $-X$ 也有

$$\mathbb{P}(X \leq -t) \leq e^{-t^2/2}.$$

因而

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/2},$$

故 X 为次高斯随机变量。

(2) 由矩的尾积分公式，

$$\mathbb{E}[|X|^p] = \int_0^\infty pt^{p-1}\mathbb{P}(|X| > t)\,dt \leq 2p\int_0^\infty t^{p-1}e^{-t^2/K^2}\,dt.$$

作代换 $u = t^2/K^2$，得

$$\mathbb{E}[|X|^p] \leq pK^p\int_0^\infty u^{p/2-1}e^{-u}\,du = pK^p\Gamma(p/2) = 2K^p\Gamma(p/2+1).$$

由 Stirling 公式，存在常数 $C > 0$，使得对所有 $p \geq 1$,

$$\Gamma(p/2+1) \leq C^p p^{p/2}.$$

因而

$$\mathbb{E}[|X|^p] \leq (K_1\sqrt{p})^p$$

对某个与 p 无关的常数 K_1 成立。 □

§5.5 习题5.2

【习题 5.21】(5.2.2) 若 X_n, Y_n 独立，X, Y 也独立，且 $X_n \xrightarrow{D} X$、$Y_n \xrightarrow{D} Y$，证明

$$X_n + Y_n \xrightarrow{D} X + Y.$$

证明. 由独立性，

$$\phi_{X_n+Y_n}(t) = \phi_{X_n}(t)\phi_{Y_n}(t).$$

又因为 $X_n \xrightarrow{D} X$、$Y_n \xrightarrow{D} Y$，故对任意 $t \in \mathbb{R}$,

$$\phi_{X_n}(t) \to \phi_X(t), \qquad \phi_{Y_n}(t) \to \phi_Y(t).$$

再由 X, Y 独立，

$$\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t).$$

因而

$$\phi_{X_n+Y_n}(t) \to \phi_{X+Y}(t).$$

由 Lévy 连续性定理，

$$X_n + Y_n \xrightarrow{D} X + Y.$$

□

【习题 5.22】(5.2.3) 随机变量 $X_1, \cdots, X_n$ 相互独立且服从柯西分布，证明

$$\frac{1}{n}\sum_{k=1}^{n} X_k$$

亦服从柯西分布。

证明. 先计算标准柯西分布的特征函数。若 X 服从密度

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

则

$$\phi_X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx.$$

当 $t > 0$ 时，考虑

$$g(z) = \frac{e^{itz}}{1+z^2},$$

并取上半平面的半圆轮廓。由 Jordan 引理，半圆弧上的积分趋于 0。轮廓内只有极点 $z = i$，且

$$\mathrm{Res}(g, i) = \lim_{z\to i} \frac{e^{itz}}{z+i} = \frac{e^{-t}}{2i}.$$

故由留数定理，

$$\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx = 2\pi i \cdot \frac{e^{-t}}{2i} = \pi e^{-t}.$$

因而

$$\phi_X(t) = e^{-t}, \qquad t > 0.$$

又由于密度 f 是偶函数，

$$\phi_X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\cos(tx)}{1+x^2}\,dx,$$

从而 ϕ_X 是偶函数。故当 $t < 0$ 时，

$$\phi_X(t) = \phi_X(-t) = e^{-(-t)} = e^t.$$

再加上 $\phi_X(0) = 1$，综上

$$\phi_X(t) = e^{-|t|}, \qquad t \in \mathbb{R}.$$

因而

$$\phi_{X_k/n}(t) = \phi_{X_k}\left(\frac{t}{n}\right) = e^{-|t|/n}.$$

由独立性，

$$\phi_{\frac{1}{n}\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \phi_{X_k/n}(t) = \left(e^{-|t|/n}\right)^n = e^{-|t|}.$$

这与标准柯西分布的特征函数相同，故

$$\frac{1}{n}\sum_{k=1}^n X_k$$

仍服从柯西分布。 □

【习题 5.23】(5.2.5) 设 $\phi_n(t) = \cos^n t$，$t \in \mathbb{R}$。

(1) 求特征函数 $\phi_2(t)$ 对应的分布函数；

(2) 对一般的正整数 n，$\phi_n(t)$ 是否为特征函数？回答并给出理由。

证明. (1) 若定义随机变量 X 满足

$$\mathbb{P}(X=-2) = \frac{1}{4}, \qquad \mathbb{P}(X=0) = \frac{1}{2}, \qquad \mathbb{P}(X=2) = \frac{1}{4},$$

则

$$\phi_X(t) = \frac{1}{4}e^{-2it} + \frac{1}{2} + \frac{1}{4}e^{2it} = \cos^2 t.$$

因而 ϕ_2 对应的分布函数为

$$F_2(x) = \begin{cases} 0, & x < -2, \\ \frac{1}{4}, & -2 \le x < 0, \\ \frac{3}{4}, & 0 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

(2) 对任意正整数 n，令 $Y_1, \cdots, Y_n$ 独立同分布，且

$$\mathbb{P}(Y_k = 1) = \mathbb{P}(Y_k = -1) = \frac{1}{2}.$$

则

$$\phi_{Y_k}(t)=\frac{1}{2}(e^{it}+e^{-it})=\cos t.$$

于是由独立性，

$$\phi_{Y_1+\cdots+Y_n}(t)=\prod_{k=1}^{n}\phi_{Y_k}(t)=\cos^n t=\phi_n(t).$$

因而对任意正整数 n，$\phi_n(t)$ 都是特征函数。 □

§5.6 习题5.3

【习题 5.24】(5.3.1) 试选择合适的数列 $\{\mu_n\}$、$\{\sigma_n\}$ 证明

$$\frac{X_n-\mu_n}{\sigma_n}\xrightarrow{D}N(0,1).$$

(1) X_n 服从参数为正整数 n 的泊松分布；

(2) X_n 服从密度为

$$f(x)=\frac{x^{n-1}e^{-x}}{\Gamma(n)}\mathbf{1}_{x\geq 0}$$

的 Γ 分布。

证明. (1) 若 $Y_1,\cdots,Y_n$ 独立同分布且 $Y_i\sim P(1)$，则

$$X_n':=Y_1+\cdots+Y_n\sim P(n),$$

即 X_n' 与 X_n 同分布。由 i.i.d. CLT，

$$\frac{X_n'-n}{\sqrt{n}}\xrightarrow{D}N(0,1).$$

因而取

$$\mu_n=n,\qquad \sigma_n=\sqrt{n},$$

就有

$$\frac{X_n-\mu_n}{\sigma_n}\xrightarrow{D}N(0,1).$$

(2) 若 $Z_1,\cdots,Z_n$ 独立同分布且每个 Z_i 服从参数为 1 的指数分布，则

$$X_n':=Z_1+\cdots+Z_n$$

的密度正是

$$f(x)=\frac{x^{n-1}e^{-x}}{\Gamma(n)}\mathbf{1}_{x\geq 0}.$$

于是 X_n' 与 X_n 同分布。再次由 i.i.d. CLT，

$$\frac{X_n'-n}{\sqrt{n}}\xrightarrow{D}N(0,1).$$

故同样取

$$\mu_n = n, \qquad \sigma_n = \sqrt{n},$$

即得

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{D} N(0,1).$$

□

【习题 5.25】(5.3.3) 随机变量 $X_1, \cdots, X_n$ 独立同分布且满足

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2},$$

证明

$$\frac{\sqrt{3}}{n^{3/2}} \sum_{k=1}^{n} kX_k \xrightarrow{D} N(0,1).$$

证明. 记

$$Y_{n,k} = kX_k, \qquad 1 \le k \le n.$$

则 $\{Y_{n,k}\}_{k=1}^n$ 相互独立，且

$$\mathbb{E}[Y_{n,k}] = 0, \qquad \mathrm{Var}(Y_{n,k}) = k^2.$$

令

$$B_n^2 = \sum_{k=1}^{n} \mathrm{Var}(Y_{n,k}) = \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

对任意 $\varepsilon > 0$，当 n 充分大时，由于 $B_n \asymp n^{3/2}$，便有

$$|Y_{n,k}| = k \le n < \varepsilon B_n, \qquad 1 \le k \le n.$$

从而

$$\sum_{k=1}^{n} \mathbb{E}\left[Y_{n,k}^2; |Y_{n,k}| > \varepsilon B_n\right] = 0,$$

所以 Lindeberg 条件显然成立。由 Lindeberg-Feller CLT，

$$\frac{\sum_{k=1}^n kX_k}{B_n} \xrightarrow{D} N(0,1).$$

又

$$\frac{B_n}{n^{3/2}} = \sqrt{\frac{(n+1)(2n+1)}{6n^2}} \longrightarrow \frac{1}{\sqrt{3}},$$

因而

$$\frac{\sqrt{3}}{n^{3/2}} \sum_{k=1}^{n} kX_k \xrightarrow{D} N(0,1).$$

□

§5.7 习题5.5

【习题 5.26】(5.5.12) Slutsky 定理的叙述如下：设随机变量 $\{X_n\}$、$\{Y_n\}$、$\{Z_n\}$ 满足

$$X_n \xrightarrow{D} X, \qquad Y_n \xrightarrow{P} b, \qquad Z_n \xrightarrow{P} c,$$

其中 X 为随机变量，b, c 为常数，则

$$X_nY_n + Z_n \xrightarrow{D} bX + c.$$

试利用 Slutsky 定理回答如下问题：

(1) 设随机变量列 $\{X_n\}$ 独立同分布，$\mathbb{E}[X_1] = 0$ 且二阶矩有限，令

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

证明

$$\frac{\sum_{k=1}^{n} X_k}{\sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2}} \xrightarrow{D} N(0,1).$$

(2) 随机变量列 $\{X_n\}$ 相互独立且满足

$$\mathbb{P}(X_n = \pm 2^n) = \frac{1}{2^{n+1}}, \qquad \mathbb{P}(X_n = \pm 1) = \frac{1}{2} - \frac{1}{2^{n+1}},$$

证明

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \xrightarrow{D} N(0,1).$$

(3) 随机变量列 $\{X_n\}$ 独立同分布，且满足 $\mathbb{E}[X_1] = \mathrm{Var}(X_1) = 1$。记

$$S_n = \sum_{k=1}^{n} X_k,$$

证明

$$\frac{S_n^{3/2} - n^{3/2}}{\frac{3}{2}n} \xrightarrow{D} N(0,1).$$

证明. (1) 记 $\sigma^2 = \mathrm{Var}(X_1)$。由 CLT，

$$\frac{\sum_{k=1}^{n} X_k}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1).$$

又

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}^2.$$

由弱大数定律，

$$\frac{1}{n}\sum_{k=1}^{n} X_k^2 \xrightarrow{P} \mathbb{E}[X_1^2] = \sigma^2, \qquad \overline{X} \xrightarrow{P} 0.$$

因而

$$\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2 \xrightarrow{P} \sigma^2, \qquad \frac{\sigma}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2}} \xrightarrow{P} 1.$$

于是由 Slutsky 定理，

$$\frac{\sum_{k=1}^{n}X_k}{\sqrt{\sum_{k=1}^{n}(X_k-\overline{X})^2}} = \frac{\sum_{k=1}^{n}X_k}{\sigma\sqrt{n}} \cdot \frac{\sigma}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2}} \xrightarrow{D} N(0,1).$$

(2) 不妨设

$$X_k = (1-B_k)\varepsilon_k + B_k 2^k \eta_k,$$

其中 $\{B_k\}$、$\{\varepsilon_k\}$、$\{\eta_k\}$ 相互独立，且

$$\mathbb{P}(B_k=1)=2^{-k}, \qquad \mathbb{P}(\varepsilon_k=\pm 1)=\mathbb{P}(\eta_k=\pm 1)=\frac{1}{2}.$$

这样定义的 X_k 正好具有题中的分布。记

$$T_n=\sum_{k=1}^{n}\varepsilon_k, \qquad R_n=\sum_{k=1}^{n}B_k(2^k\eta_k-\varepsilon_k).$$

则

$$\sum_{k=1}^{n}X_k = T_n+R_n.$$

又因为

$$\sum_{k=1}^{\infty}\mathbb{P}(B_k=1)=\sum_{k=1}^{\infty}2^{-k}<\infty,$$

由第一 Borel-Cantelli 引理，事件 $\{B_k=1\}$ 只会发生有限次。故 R_n a.s. 最终为常数，从而

$$\frac{R_n}{\sqrt{n}} \xrightarrow{\text{a.s.}} 0.$$

另一方面，由 CLT，

$$\frac{T_n}{\sqrt{n}} \xrightarrow{D} N(0,1).$$

因此由 Slutsky 定理，

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k = \frac{T_n}{\sqrt{n}}+\frac{R_n}{\sqrt{n}} \xrightarrow{D} N(0,1).$$

(3) 记

$$T_n=\frac{S_n-n}{\sqrt{n}}, \qquad U_n=\frac{S_n}{n}.$$

由 CLT，

$$T_n \xrightarrow{D} N(0,1),$$

由弱大数定律，

$$U_n \xrightarrow{P} 1.$$

又

$$\frac{S_n^{3/2}-n^{3/2}}{\frac{3}{2}n}=T_n\cdot\frac{2}{3}\cdot\frac{U_n^{3/2}-1}{U_n-1}.$$

记

$$g(u)=\frac{2}{3}\cdot\frac{u^{3/2}-1}{u-1}\quad(u\neq 1),\qquad g(1)=1.$$

则 g 在 $u=1$ 处连续，故

$$g(U_n)\xrightarrow{P}1.$$

再由 Slutsky 定理，

$$\frac{S_n^{3/2}-n^{3/2}}{\frac{3}{2}n}\xrightarrow{D}N(0,1).$$

□

§5.8 习题5.4

【习题 5.27】 **(5.4.1)** 随机变量$X_1,X_2,\ldots$独立同分布，$\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\frac{1}{2}$，证明对任意$\delta>0$均有

$$\frac{1}{n^{1/2+\delta}}\sum_{k=1}^{n}X_k\xrightarrow{\text{a.s.}}0.$$

证明. 如果只用Chebyshev不等式，你会发现只能做到$\delta>1/2$。我们尝试使用更高阶矩。待定正整数 m。记

$$S_n=\sum_{k=1}^{n}X_k.$$

对任意 $\varepsilon>0$，由 Markov 不等式，

$$\mathbb{P}\left(\left|\frac{S_n}{n^{1/2+\delta}}\right|>\varepsilon\right)=\mathbb{P}\left(\left|\frac{S_n}{n^{1/2+\delta}}\right|^{2m}>\varepsilon^{2m}\right)\leq\frac{\mathbb{E}|S_n|^{2m}}{\varepsilon^{2m}n^{m+2m\delta}}.$$

下面估计 $\mathbb{E}|S_n|^{2m}$。展开得

$$\mathbb{E}S_n^{2m}=\sum_{i_1,\ldots,i_{2m}=1}^{n}\mathbb{E}(X_{i_1}\cdots X_{i_{2m}}).$$

由于 X_i 独立且 $\mathbb{E}X_i=0$，若某个指标只出现奇数次，则该项期望为 0。因此非零项中出现的不同指标个数至多为 m。于是存在只依赖于 m 的常数 C_m，使得

$$\mathbb{E}S_n^{2m}\leq C_m n^m.$$

从而

$$\mathbb{P}\left(\left|\frac{S_n}{n^{1/2+\delta}}\right|>\varepsilon\right)\leq\frac{C_m}{\varepsilon^{2m}}n^{-2m\delta}.$$

选取正整数 m 使得

$$2m\delta > 1.$$

则

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_n}{n^{1/2+\delta}}\right| > \varepsilon\right) < \infty.$$

由 Borel-Cantelli 引理，

$$\mathbb{P}\left(\left|\frac{S_n}{n^{1/2+\delta}}\right| > \varepsilon \text{ i.o.}\right) = 0.$$

因此对任意固定的 $\varepsilon > 0$，几乎必然存在 $N(\omega)$，使得当 $n \geq N(\omega)$ 时，

$$\left|\frac{S_n}{n^{1/2+\delta}}\right| \leq \varepsilon.$$

再令 ε 取所有正有理数，即得

$$\frac{S_n}{n^{1/2+\delta}} \xrightarrow{\text{a.s.}} 0.$$

□

【习题 5.28】(5.4.4) 设 $\{X_k\}$ 为独立同分布随机变量列，

$$\mathbb{E}X_1 = 0, \qquad \operatorname{Var}(X_1) = 1, \qquad \mathbb{E}|X_1|^3 < \infty.$$

试用 Lindeberg 替换法证明 CLT 的收敛速度

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \leq t\right) - \Phi(t)\right| = O(n^{-1/8}).$$

这里 $\Phi(t)$ 表示标准正态分布函数。

证明. 记

$$S_n = \sum_{k=1}^{n} X_k, \qquad W_n = \frac{S_n}{\sqrt{n}}.$$

令 $Y_1, \ldots, Y_n$ 为独立同分布的标准正态随机变量，并且与 $X_1, \ldots, X_n$ 独立。记

$$Z_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k.$$

则 $Z_n \sim N(0,1)$，所以

$$\mathbb{P}(Z_n \leq t) = \Phi(t).$$

取 $\varepsilon > 0$。存在光滑函数 $f_{t,\varepsilon} \in C^3(\mathbb{R})$，使得

$$1_{\{x\leq t\}} \leq f_{t,\varepsilon}(x) \leq 1_{\{x\leq t+\varepsilon\}},$$

且

$$\|f_{t,\varepsilon}^{(3)}\|_\infty \leq C\varepsilon^{-3},$$

其中常数 C 与 t, ε, n 无关。

下面估计

$$|\mathbb{E}f_{t,\varepsilon}(W_n)-\mathbb{E}f_{t,\varepsilon}(Z_n)|.$$

逐个将 X_k 替换为 Y_k。令

$$T_k=\frac{1}{\sqrt{n}}\left(Y_1+\cdots+Y_{k-1}+X_{k+1}+\cdots+X_n\right).$$

则 T_k 与 X_k,Y_k 独立。由 Taylor 展开，

$$f_{t,\varepsilon}\left(T_k+\frac{X_k}{\sqrt{n}}\right)=f_{t,\varepsilon}(T_k)+\frac{X_k}{\sqrt{n}}f'_{t,\varepsilon}(T_k)+\frac{X_k^2}{2n}f''_{t,\varepsilon}(T_k)+R_{k,X},$$

其中

$$|R_{k,X}|\le\frac{\|f_{t,\varepsilon}^{(3)}\|_\infty}{6}\frac{|X_k|^3}{n^{3/2}}.$$

同理，

$$f_{t,\varepsilon}\left(T_k+\frac{Y_k}{\sqrt{n}}\right)=f_{t,\varepsilon}(T_k)+\frac{Y_k}{\sqrt{n}}f'_{t,\varepsilon}(T_k)+\frac{Y_k^2}{2n}f''_{t,\varepsilon}(T_k)+R_{k,Y},$$

且

$$|R_{k,Y}|\le\frac{\|f_{t,\varepsilon}^{(3)}\|_\infty}{6}\frac{|Y_k|^3}{n^{3/2}}.$$

因为

$$\mathbb{E}X_k=\mathbb{E}Y_k=0,\qquad \mathbb{E}X_k^2=\mathbb{E}Y_k^2=1,$$

且 T_k 与 X_k,Y_k 独立，所以一阶项和二阶项在取期望后相消。因此

$$\left|\mathbb{E}f_{t,\varepsilon}\left(T_k+\frac{X_k}{\sqrt{n}}\right)-\mathbb{E}f_{t,\varepsilon}\left(T_k+\frac{Y_k}{\sqrt{n}}\right)\right|\le C\varepsilon^{-3}n^{-3/2}.$$

将 $k=1,\ldots,n$ 相加，得到

$$|\mathbb{E}f_{t,\varepsilon}(W_n)-\mathbb{E}f_{t,\varepsilon}(Z_n)|\le C\varepsilon^{-3}n^{-1/2}.$$

于是

$$\mathbb{P}(W_n\le t)\le\mathbb{E}f_{t,\varepsilon}(W_n)\le\mathbb{E}f_{t,\varepsilon}(Z_n)+C\varepsilon^{-3}n^{-1/2}.$$

又因为

$$f_{t,\varepsilon}(x)\le 1_{\{x\le t+\varepsilon\}},$$

所以

$$\mathbb{E}f_{t,\varepsilon}(Z_n)\le\mathbb{P}(Z_n\le t+\varepsilon)=\Phi(t+\varepsilon).$$

因此

$$\mathbb{P}(W_n\le t)-\Phi(t)\le\Phi(t+\varepsilon)-\Phi(t)+C\varepsilon^{-3}n^{-1/2}.$$

由于标准正态密度有界，

$$\Phi(t+\varepsilon)-\Phi(t)\le C\varepsilon,$$

故
$$\mathbb{P}(W_n \le t) - \Phi(t) \le C\varepsilon + C\varepsilon^{-3}n^{-1/2}.$$

另一方面，同理取光滑函数满足
$$1_{\{x \le t-\varepsilon\}} \le g_{t,\varepsilon}(x) \le 1_{\{x \le t\}}, \qquad \|g_{t,\varepsilon}^{(3)}\|_\infty \le C\varepsilon^{-3}.$$

重复上述 Lindeberg 替换，得到
$$|\mathbb{E}g_{t,\varepsilon}(W_n) - \mathbb{E}g_{t,\varepsilon}(Z_n)| \le C\varepsilon^{-3}n^{-1/2}.$$

因此
$$\mathbb{P}(W_n \le t) \ge \mathbb{E}g_{t,\varepsilon}(W_n) \ge \mathbb{E}g_{t,\varepsilon}(Z_n) - C\varepsilon^{-3}n^{-1/2}.$$

又
$$\mathbb{E}g_{t,\varepsilon}(Z_n) \ge \mathbb{P}(Z_n \le t-\varepsilon) = \Phi(t-\varepsilon),$$

所以
$$\Phi(t) - \mathbb{P}(W_n \le t) \le \Phi(t) - \Phi(t-\varepsilon) + C\varepsilon^{-3}n^{-1/2} \le C\varepsilon + C\varepsilon^{-3}n^{-1/2}.$$

综上，对任意 $t \in \mathbb{R}$，
$$|\mathbb{P}(W_n \le t) - \Phi(t)| \le C\varepsilon + C\varepsilon^{-3}n^{-1/2}.$$

取
$$\varepsilon = n^{-1/8},$$

得
$$|\mathbb{P}(W_n \le t) - \Phi(t)| \le Cn^{-1/8}.$$

因此
$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \le t \right) - \Phi(t) \right| = O(n^{-1/8}).$$

□

【注 5.29】 核心是：先用光滑函数近似指标函数，再逐个把 X_i 替换成正态变量 Y_i。因为 X_i 与 Y_i 的前两阶矩相同，Taylor 展开中的一阶、二阶项会相消，只剩三阶余项。平滑误差为 $O(\varepsilon)$，替换误差为 $O(\varepsilon^{-3}n^{-1/2})$，取 $\varepsilon = n^{-1/8}$ 即得结论。

【习题 5.30】**(5.4.5)** $\{X_n\}$ 为独立同分布的随机变量列，
$$\mathbb{E}X_1 = 0, \qquad \mathbb{E}X_1^2 = 1,$$

且对所有 $l \ge 3$ 均有
$$\mathbb{E}|X_1|^l < \infty.$$

令
$$S_n = X_1 + \cdots + X_n.$$

令 $H_k(x)$ 是 k 阶 Hermite 多项式，即满足

$$H_0 = 1, \qquad (-1)^k H_k(x)\phi(x) = \phi^{(k)}(x),$$

其中 ϕ 是标准正态分布的密度函数。求证：

$$\lim_{n\to\infty} \mathbb{E}\left[H_k\left(\frac{S_n}{\sqrt{n}}\right)\right] = 0, \qquad \forall k \geq 1.$$

证明. 记

$$W_n = \frac{S_n}{\sqrt{n}}.$$

我们先证明：对任意固定的正整数 j，有

$$\lim_{n\to\infty} \mathbb{E}W_n^j = \mathbb{E}Z^j,$$

其中 $Z \sim N(0,1)$。

展开得

$$\mathbb{E}W_n^j = n^{-j/2} \sum_{i_1,\ldots,i_j=1}^{n} \mathbb{E}(X_{i_1}\cdots X_{i_j}).$$

由于 $\mathbb{E}X_1 = 0$ 且 X_i 独立，若某个指标只出现一次，则对应项期望为 0。

因此非零项中，每个出现的指标至少出现两次。若出现了 r 个不同指标，则

$$r \leq \frac{j}{2}.$$

当 $r < j/2$ 时，这类项的总贡献至多为

$$O(n^r)n^{-j/2} = o(1).$$

所以极限只可能来自 $r = j/2$ 的情形。这要求 j 为偶数，并且每个出现的指标恰好出现两次。设 $j = 2m$，则这种配对方式共有

$$(2m-1)!!$$

种，且每一项的期望为

$$\mathbb{E}X_1^2 \cdots \mathbb{E}X_m^2 = 1.$$

因此

$$\lim_{n\to\infty} \mathbb{E}W_n^{2m} = (2m-1)!!.$$

若 j 为奇数，则不存在 $r = j/2$ 的情形，故

$$\lim_{n\to\infty} \mathbb{E}W_n^j = 0.$$

这正是标准正态随机变量 Z 的各阶矩，因此

$$\lim_{n\to\infty} \mathbb{E}W_n^j = \mathbb{E}Z^j.$$

由于 $H_k(x)$ 是一个 k 次多项式，可写为

$$H_k(x) = \sum_{j=0}^{k} a_j x^j.$$

于是由上面的矩收敛，

$$\lim_{n\to\infty} \mathbb{E}H_k(W_n) = \sum_{j=0}^{k} a_j \lim_{n\to\infty} \mathbb{E}W_n^j = \sum_{j=0}^{k} a_j \mathbb{E}Z^j = \mathbb{E}H_k(Z).$$

最后计算 $\mathbb{E}H_k(Z)$。由 $Z \sim N(0,1)$，

$$\mathbb{E}H_k(Z) = \int_{-\infty}^{\infty} H_k(x)\phi(x)\, dx.$$

根据 Hermite 多项式的定义，

$$H_k(x)\phi(x) = (-1)^k \phi^{(k)}(x).$$

因此

$$\mathbb{E}H_k(Z) = (-1)^k \int_{-\infty}^{\infty} \phi^{(k)}(x)\, dx.$$

当 $k \geq 1$ 时，

$$\int_{-\infty}^{\infty} \phi^{(k)}(x)\, dx = \phi^{(k-1)}(\infty) - \phi^{(k-1)}(-\infty) = 0.$$

故

$$\mathbb{E}H_k(Z) = 0, \qquad k \geq 1.$$

从而

$$\lim_{n\to\infty} \mathbb{E}\left[H_k\left(\frac{S_n}{\sqrt{n}}\right)\right] = 0, \qquad \forall k \geq 1.$$

□

【注 5.31】 本题想法是先证明 $S_n/\sqrt{n}$ 的固定阶矩收敛到标准正态矩。展开矩时，由于 $\mathbb{E}X_i = 0$，只有指标成对出现的项在极限中保留下来，这正对应正态分布的矩。又因为 H_k 是多项式，所以可由矩收敛推出 $\mathbb{E}H_k(S_n/\sqrt{n}) \to \mathbb{E}H_k(Z)$。最后利用 Hermite 多项式在标准正态下满足 $\mathbb{E}H_k(Z) = 0$。

【习题 5.32】**(5.4.8)** （Stein 方法）试证明

$$X \sim N(0,1)$$

当且仅当对任意有界连续函数 g，若其导数 g' 也有界连续，则总有

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)].$$

提示：对标准正态分布 Z 和有界连续函数 h，构造一个新的函数

$$g_0(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2}\big(h(y) - \mathbb{E}h(Z)\big)\, dy.$$

证明. 先证必要性。若 $X \sim N(0,1)$，其密度为

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

因为

$$\phi'(x) = -x\phi(x),$$

所以

$$\mathbb{E}[Xg(X)] = \int_{-\infty}^{\infty} xg(x)\phi(x)\,dx = -\int_{-\infty}^{\infty} g(x)\phi'(x)\,dx.$$

对右边分部积分，得

$$-\int_{-\infty}^{\infty} g(x)\phi'(x)\,dx = -\Big[g(x)\phi(x)\Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} g'(x)\phi(x)\,dx.$$

由于 g 有界且 $\phi(x) \to 0$，边界项为 0。因此

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)].$$

下面证充分性。设对任意有界连续 g，且 g' 有界连续时，都有

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)].$$

我们要证明 $X \sim N(0,1)$。

令 $Z \sim N(0,1)$。任取有界连续函数 h，定义

$$g_0(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2} \big(h(y) - \mathbb{E}h(Z)\big)\,dy.$$

由于

$$\int_{-\infty}^{\infty} e^{-y^2/2} \big(h(y) - \mathbb{E}h(Z)\big)\,dy = 0,$$

也可写为

$$g_0(x) = -e^{x^2/2} \int_{x}^{\infty} e^{-y^2/2} \big(h(y) - \mathbb{E}h(Z)\big)\,dy.$$

由标准正态尾部估计可知，g_0 有界连续，且其导数也有界连续。

对 g_0 求导：

$$g_0'(x) = xe^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2} \big(h(y) - \mathbb{E}h(Z)\big)\,dy + h(x) - \mathbb{E}h(Z).$$

因此

$$g_0'(x) = xg_0(x) + h(x) - \mathbb{E}h(Z),$$

即

$$g_0'(x) - xg_0(x) = h(x) - \mathbb{E}h(Z).$$

由假设，将 $g = g_0$ 代入，得到

$$\mathbb{E}[Xg_0(X)] = \mathbb{E}[g_0'(X)].$$

因此

$$\mathbb{E}\left[g_0'(X) - Xg_0(X)\right] = 0.$$

由上面的 Stein 方程，

$$g_0'(X) - Xg_0(X) = h(X) - \mathbb{E}h(Z).$$

所以

$$\mathbb{E}h(X) - \mathbb{E}h(Z) = 0.$$

即

$$\mathbb{E}h(X) = \mathbb{E}h(Z)$$

对任意有界连续函数 h 成立。

因此 X 与 Z 分布相同，即

$$X \sim N(0,1).$$

综上，命题得证。 □

【注 5.33】 本题的核心是 Stein 方法中的一个基本刻画：

$$X \sim N(0,1) \quad \Longleftrightarrow \quad \mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)]$$

对足够多的测试函数 g 成立。

必要性来自正态密度的特殊性质

$$\phi'(x) = -x\phi(x),$$

因此可以通过分部积分把 $\mathbb{E}[Xg(X)]$ 转化为 $\mathbb{E}[g'(X)]$。

充分性更有技巧。我们想证明 X 和标准正态 Z 分布相同，只需证明对任意有界连续函数 h，都有

$$\mathbb{E}h(X) = \mathbb{E}h(Z).$$

为此构造一个函数 g_0，使它满足 Stein 方程

$$g_0'(x) - xg_0(x) = h(x) - \mathbb{E}h(Z).$$

然后将 g_0 代入假设

$$\mathbb{E}[Xg_0(X)] = \mathbb{E}[g_0'(X)],$$

就得到

$$\mathbb{E}h(X) = \mathbb{E}h(Z).$$

所以 X 必为标准正态分布。这个方法的强大之处在于：它把“证明分布接近正态”转化成了“估计 Stein 方程两边的误差”。

不光是正态分布，其他一些经典分布也有类似的 Stein 刻画，比如指数分布、泊松分布等。通过构造合适的 Stein 方程，可以得到这些分布的刻画，从而在证明极限定理时提供了一个强有力的工具。举个例子，值得同学们思考：

（指数分布的 Stein 刻画）设 $\lambda > 0$，W 是取值于 $(0, \infty)$ 的连续型随机变量，具有密度 q。证明：在适当正则条件下，

$$W \sim \mathrm{Exp}(\lambda)$$

当且仅当对任意 $f \in C_c^1(0, \infty)$，均有

$$\mathbb{E}f'(W) = \lambda \mathbb{E}f(W).$$