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。

证明.

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

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？

证明. 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 $\mathbb{E}[X^3]$。

证明. 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）

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

【习题 3.9】 (2.2.4) 对非负整值随机变量 X，证明

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

<u>**证明.**</u> 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 $\mathrm{Var}(X)$。

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 $N_W \geq |E|/2$。

<u>**证明.**</u> 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。

<u>**证明.**</u> 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。

证明. 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

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

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

证明. 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：$n!\sim n^n e^{-n}\sqrt{2\pi n}$。

证明. 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：

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

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

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

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？

证明. 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
$$G(s)=\sum_{n=0}^{\infty}p_ns^n.$$
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 $\lim_{x\to+\infty} F(x) = 1$。

<u>证明.</u> 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

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

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，
$$\mathbb{E}\big[(X_n - n)^{2m+1}\big] = 0.$$

3. 令
$$X_n^* := \frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathrm{Var}(X_n)}},$$
证明
$$1 \le \mathbb{E}\big[(X_n^*)^4\big] < 3.$$

<u>**解.**</u> **(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
$$G(z) = \lambda \prod_{k=1}^{n} (z^2 - a_k z + 1), \qquad a_k \in [-2,2].$$
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
$$\mathbb{E}\big[(X_n - n)^{2m+1}\big] = 0.$$

(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 $0 < c_k \leq 4 < 6$，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

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

另一方面，由 Jensen 不等式（或 Cauchy 不等式）

$$\mathbb{E}\left[(X_n^*)^4\right] \geq \left(\mathbb{E}[(X_n^*)^2]\right)^2 = 1.$$

综上，

$$1 \leq \mathbb{E}\left[(X_n^*)^4\right] < 3.$$

□

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

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

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 $1 \le i < j \le n$，定义示性随机变量

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

则

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

由期望的线性性，

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

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

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

$\square$