人工智能数学基础：作业 11 参考解答

约定. 除非特别说明，默认随机放球/取点均指各对象独立且均匀地分配到指定集合；对数 log 以 e 为底。更换对数底数只会引入常数因子，不影响概率不等式与渐进估计。

2.1

有 m 个球独立均匀放入 n 个篮子。设

$$X_{m,n} = \#\{\{i,j\} : 1 \le i < j \le m,\ 第 i\ 球与第 j\ 球在同一篮子\}.$$

等价地，若第 k 个篮子里共有 Y_k 个球，则 $X_{m,n} = \sum_{k=1}^{n} \binom{Y_k}{2}$.

(1) 期望 $\mu_{m,n}$

对每一对 $1 \le i < j \le m$ 定义指示变量

$$I_{ij} = \mathbf{1}\{球 i\ 与球 j\ 在同一篮子\}.$$

则 $X_{m,n} = \sum_{1\le i<j\le m} I_{ij}$。由于两球独立均匀落入 n 个篮子，

$$\mathbb{P}(I_{ij}=1) = \frac{1}{n}, \qquad \mathbb{E}[I_{ij}] = \frac{1}{n}.$$

故

$$\mu_{m,n} := \mathbb{E}[X_{m,n}] = \sum_{1\le i<j\le m} \mathbb{E}[I_{ij}] = \binom{m}{2}\cdot\frac{1}{n} = \frac{m(m-1)}{2n}.$$

(2) 方差 $\mathrm{Var}(X_{m,n})$

先计算单项方差：

$$\mathrm{Var}(I_{ij}) = \mathbb{E}[I_{ij}^2] - \mathbb{E}[I_{ij}]^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{1}{n}\Big(1-\frac{1}{n}\Big).$$

关键在于不同对的协方差为 0。

引理 1 (不同球对的指示变量两两独立)**.** 对任意两对不同的下标 $(i,j) \ne (k,\ell)$（允许共享一个球），有 I_{ij} 与 $I_{k\ell}$ 独立，从而 $\mathrm{Cov}(I_{ij}, I_{k\ell}) = 0$。

证明. 分两类讨论。

(1) 若 $\{i,j\}\cap\{k,\ell\}=\varnothing$，则两事件分别只涉及互不相交的四个球的落点，在独立投掷假设下显然独立，且 $\mathbb{P}(I_{ij}=1,\ I_{k\ell}=1)=\frac{1}{n^2}=\mathbb{P}(I_{ij}=1)\mathbb{P}(I_{k\ell}=1)$。

(2) 若两对共享一个球，不妨设为 (i,j) 与 (i,k)，则

$$\mathbb{P}(I_{ij}=1,\ I_{ik}=1)=\mathbb{P}(\text{球}j\text{ 与}k\text{ 同时落在球}i\text{ 的篮子})=\frac{1}{n}\cdot\frac{1}{n}=\frac{1}{n^2}.$$

另一方面 $\mathbb{P}(I_{ij}=1)\mathbb{P}(I_{ik}=1)=\frac{1}{n^2}$，故亦独立。

综上，两两独立成立。 □

因此

$$\mathrm{Var}(X_{m,n})=\sum_{1\le i<j\le m}\mathrm{Var}(I_{ij})=\binom{m}{2}\cdot\frac{1}{n}\Big(1-\frac{1}{n}\Big)=\mu_{m,n}\Big(1-\frac{1}{n}\Big).$$

(3)

题目要求证明对任意 $c>0$，

$$\Pr\Big(\big|X_{m,n}-\mu_{m,n}\big|\ge c\sqrt{\mu_{m,n}}\Big)\le\frac{1}{c^2}.$$

由切比雪夫不等式，任意 $t>0$ 有

$$\Pr(|X_{m,n}-\mu_{m,n}|\ge t)\le\frac{\mathrm{Var}(X_{m,n})}{t^2}.$$

取 $t=c\sqrt{\mu_{m,n}}$，并用 $\mathrm{Var}(X_{m,n})=\mu_{m,n}(1-\frac{1}{n})\le\mu_{m,n}$，得到

$$\Pr\Big(\big|X_{m,n}-\mu_{m,n}\big|\ge c\sqrt{\mu_{m,n}}\Big)\le\frac{\mu_{m,n}}{c^2\mu_{m,n}}=\frac{1}{c^2},$$

即证。

2.2

定理 1. 设 λ 为有限随机变量，$\alpha>0$，$s>0$。则

$$\Pr(\lambda\ge\alpha)\le e^{-s\alpha}\,\mathbb{E}\big(e^{s\lambda}\big),\qquad \Pr(\lambda\le\alpha)\le e^{s\alpha}\,\mathbb{E}\big(e^{-s\lambda}\big).$$

证明. 第一条由马尔可夫不等式直接推出：对非负随机变量 $Z=e^{s\lambda}$，

$$\Pr(\lambda\ge\alpha)=\Pr(e^{s\lambda}\ge e^{s\alpha})\le\frac{\mathbb{E}[e^{s\lambda}]}{e^{s\alpha}}=e^{-s\alpha}\mathbb{E}(e^{s\lambda}).$$

第二条同理，对 $Z=e^{-s\lambda}$，

$$\Pr(\lambda\le\alpha)=\Pr(e^{-s\lambda}\ge e^{-s\alpha})\le\frac{\mathbb{E}[e^{-s\lambda}]}{e^{-s\alpha}}=e^{s\alpha}\mathbb{E}(e^{-s\lambda}).$$

□

2.3

设 $d = 1000$，单位球为 $B_d(1) = \{x \in \mathbb{R}^d : \|x\| \le 1\}$。由于 d 维体积满足缩放律 $\mathrm{Vol}(B_d(r)) = r^d \,\mathrm{Vol}(B_d(1))$，故半径不超过 r 的体积占比为 r^d。

厚度为 ε 的壳层

$$\{x : 1 - \varepsilon \le \|x\| \le 1\}$$

的体积占比为

$$1 - (1 - \varepsilon)^d.$$

要求该占比至少为 0.99，即

$$1 - (1 - \varepsilon)^{1000} \ge 0.99 \quad \Longleftrightarrow \quad (1 - \varepsilon)^{1000} \le 0.01.$$

因此

$$1 - \varepsilon \le 0.01^{1/1000} \quad \Longleftrightarrow \quad \varepsilon \ge 1 - 0.01^{1/1000}.$$

数值上

$$\varepsilon = 1 - 0.01^{1/1000} = 1 - \exp\Bigl(\frac{\ln 0.01}{1000}\Bigr) \approx 0.004595.$$

（例如取 $\varepsilon \approx 4.6 \times 10^{-3}$ 即可满足 99% 的要求。）

2.4

在立方体 $[-\frac{1}{2}, \frac{1}{2}]^{100}$ 中独立均匀生成 30 个点

$$X^{(1)}, \ldots, X^{(30)} \in \mathbb{R}^{100}.$$

对任意点对 $1 \le i < j \le 30$，定义

$$D_{ij} = \|X^{(i)} - X^{(j)}\|_2, \qquad \Theta_{ij} = \arccos\Bigl(\frac{\langle X^{(i)}, X^{(j)} \rangle}{\|X^{(i)}\|_2 \, \|X^{(j)}\|_2}\Bigr) \in [0, \pi].$$

共有 $\binom{30}{2} = 435$ 个距离与夹角数据。

使用随机种子 20251227 生成样本并绘制直方图，得到如下数值摘要：

$$\overline{D} \approx 4.0303, \quad \mathrm{sd}(D) \approx 0.2328; \qquad \overline{\Theta} \approx 89.972^\circ, \quad \mathrm{sd}(\Theta) \approx 6.044^\circ.$$

对应图形如下。

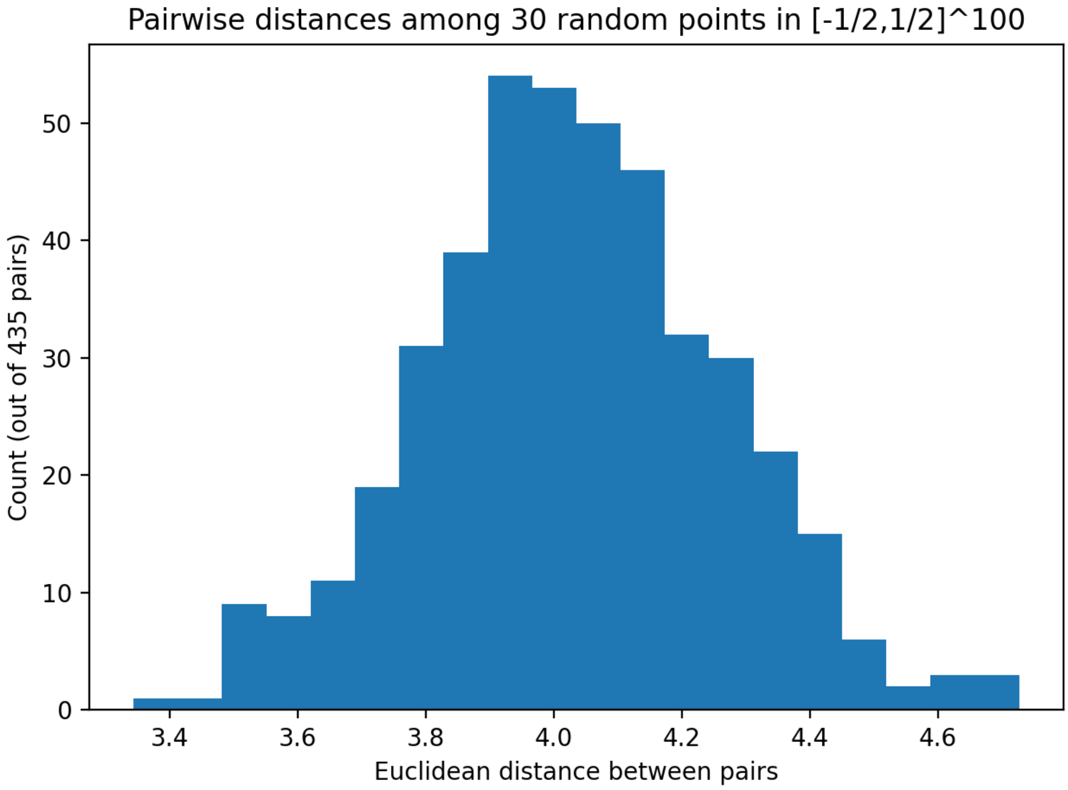

图 1: 30 个随机点的所有点对距离分布（435 个样本）

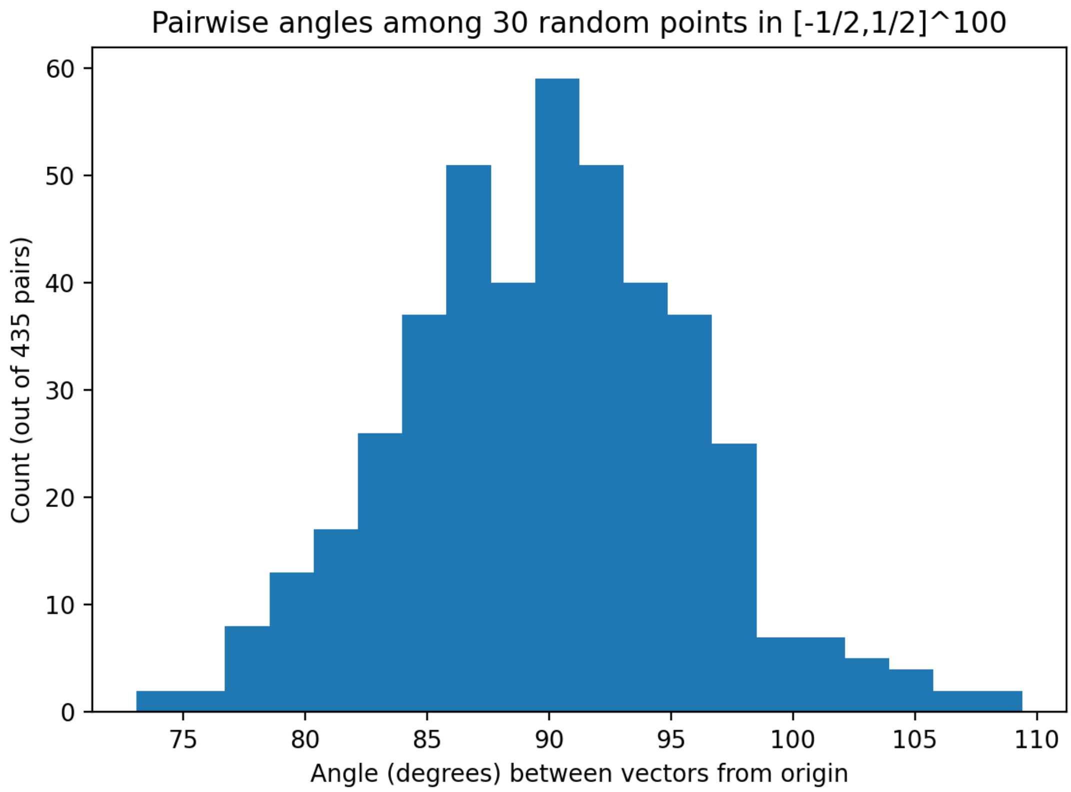

图 2: 从原点指向随机点的向量之间夹角分布（单位：度，435 个样本）

说明. 在高维情形下，D_{ij} 与 Θ_{ij} 会呈现显著集中现象：距离集中在其均值附近，夹角集中在 90° 附近。上述图形与摘要数值体现了这一点。

2.5 报警贝叶斯网络：精确计算 $\Pr(E \mid J = \textbf{True}, M = \textbf{False})$（20 分）

本题对应课件中的 “Alarm” 贝叶斯网络：

$$B \to A \leftarrow E, \qquad A \to J, \qquad A \to M,$$

其中 B=Burglary，E=Earthquake，A=Alarm，J=JohnCalls，M=MaryCalls，均为二值变量。

$$\Pr(B) = 0.001, \qquad \Pr(E) = 0.002.$$

报警结点的 CPT 为

$$\Pr(A = \mathrm{T} \mid B = \mathrm{T}, E = \mathrm{T}) = 0.70, \quad \Pr(A = \mathrm{T} \mid B = \mathrm{T}, E = \mathrm{F}) = 0.01,$$

$$\Pr(A = \mathrm{T} \mid B = \mathrm{F}, E = \mathrm{T}) = 0.70, \quad \Pr(A = \mathrm{T} \mid B = \mathrm{F}, E = \mathrm{F}) = 0.01.$$

由此可见 A **的分布只依赖** E**，与** B **无关**，即

$$\Pr(A=\mathrm{T}\mid E=\mathrm{T})=0.70,\qquad \Pr(A=\mathrm{T}\mid E=\mathrm{F})=0.01.$$

电话结点 CPT 为

$$\Pr(J=\mathrm{T}\mid A=\mathrm{T})=0.90,\ \Pr(J=\mathrm{T}\mid A=\mathrm{F})=0.05;$$

$$\Pr(M=\mathrm{T}\mid A=\mathrm{T})=0.70,\ \Pr(M=\mathrm{T}\mid A=\mathrm{F})=0.01.$$

计算

$$\Pr(E=\mathrm{T}\mid J=\mathrm{T},\,M=\mathrm{F}).$$

在该网络中（给定父结点）子结点条件独立，因此

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid A)=\Pr(J=\mathrm{T}\mid A)\Pr(M=\mathrm{F}\mid A).$$

先算两种 A 取值下的证据概率：

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid A=\mathrm{T})=0.90\cdot(1-0.70)=0.27,$$

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid A=\mathrm{F})=0.05\cdot(1-0.01)=0.0495.$$

因此对任意 $e\in\{\mathrm{T},\mathrm{F}\}$，

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=e)=\sum_{a\in\{\mathrm{T},\mathrm{F}\}}\Pr(A=a\mid E=e)\Pr(J=\mathrm{T},M=\mathrm{F}\mid A=a).$$

代入 $\Pr(A=\mathrm{T}\mid E=\mathrm{T})=0.70$ 与 $\Pr(A=\mathrm{T}\mid E=\mathrm{F})=0.01$ 得

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=\mathrm{T})=0.70\cdot 0.27+0.30\cdot 0.0495=0.20385,$$

$$\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=\mathrm{F})=0.01\cdot 0.27+0.99\cdot 0.0495=0.051705.$$

由 Bayes 公式，

$$\begin{gathered}\Pr(E=\mathrm{T}\mid J=\mathrm{T},M=\mathrm{F})=\\ \frac{\Pr(E=\mathrm{T})\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=\mathrm{T})}{\Pr(E=\mathrm{T})\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=\mathrm{T})+\Pr(E=\mathrm{F})\Pr(J=\mathrm{T},M=\mathrm{F}\mid E=\mathrm{F})}.\end{gathered}\tag{1}$$

代入 $\Pr(E=\mathrm{T})=0.002$、$\Pr(E=\mathrm{F})=0.998$ 以及上式两项：

$$\Pr(E=\mathrm{T}\mid J=\mathrm{T},M=\mathrm{F})=\frac{0.002\cdot 0.20385}{0.002\cdot 0.20385+0.998\cdot 0.051705}\approx 0.007839.$$

因此后验概率约为 0.7839%。

由于本题给定的报警 CPT 使得 A 与 B 无关，故 B 在上述计算中可直接边缘化掉而不影响结果。

2.6 开放性问答

略。

附录：2.4 的 Python 代码（可复现生成图形）

```
import numpy as np, math
import matplotlib.pyplot as plt

rng = np.random.default_rng(20251227)
d = 100
n_points = 30
pts = rng.uniform(-0.5, 0.5, size=(n_points, d))

dist, ang = [], []
for i in range(n_points):
    for j in range(i+1, n_points):
        u, v = pts[i], pts[j]
        dist.append(np.linalg.norm(u - v))
        nu, nv = np.linalg.norm(u), np.linalg.norm(v)
        cos = float(np.dot(u, v) / (nu * nv))
        cos = max(-1.0, min(1.0, cos))
        ang.append(math.degrees(math.acos(cos)))

dist = np.array(dist)
ang = np.array(ang)

plt.figure()
plt.hist(dist, bins=20)
plt.xlabel("Euclidean␣distance␣between␣pairs")
plt.ylabel("Count␣(out␣of␣435␣pairs)")
plt.title("Pairwise␣distances␣among␣30␣random␣points␣in␣[-1/2,1/2]^100")
plt.tight_layout()
plt.savefig("hw11_2_4_distances.png", dpi=200)
plt.close()

plt.figure()
plt.hist(ang, bins=20)
plt.xlabel("Angle␣(degrees)␣between␣vectors␣from␣origin")
```

```
plt.ylabel("Count (out of 435 pairs)")
plt.title("Pairwise angles among 30 random points in [-1/2,1/2]^100")
plt.tight_layout()
plt.savefig("hw11_2_4_angles.png", dpi=200)
plt.close()
```