MATH 541a Homework 1*

Qilin Ye

February 18, 2022

Problem 3

† Two people take turns throwing darts at a board. Person A goes first, and each of their throws has a probability of $1/4$ of hitting the bullseye. Person B goes next, and each of their throws has a probability of $1/3$ of hitting the bullseye. Then Person A goes, and so on. With what probability will Person A hit the bullseye before Person B does?

Solution. Let S_n be the event in which the n^{th} shot is the first to hit the bullseye. (For example, in the event S_4, A misses, then B misses, then A misses again, and finally B hits.) It is clear that the $S_n's$ are pairwise disjoint and that the events in which A hits the bullseye before B does is

$$S := \bigcup_{n=1}^{\infty} S_{2n-1}.$$

Therefore,

$$\mathbb{P}(S) = \sum_{n=1}^{\infty} P(S_{2n-1}) = \sum_{n=1}^{\infty} \frac{1}{4}\left(\frac{3}{4}\cdot\frac{2}{3}\right)^{n-1} = \frac{1}{2}.$$

Problem 4

† Two people are flipping fair coins. Let n be a positive integer. Person 1 flips $n+1$ coins and person 2 flips n coins. Prove that the following event has probability $1/2$: "person 1 has more heads than person 2."

Proof. Let each person flip n flips first. Define

$$E_A := \{\text{events where person 1 has more heads}\}$$
$$E_B := \{\text{events where person 2 has more heads}\}$$
$$E_0 := \{\text{event where both have same number of heads}\}.$$

Clearly $\mathbb{P}(E_A) + \mathbb{P}(E_B) + \mathbb{P}(E_0) = 1$. Note that if E_A happens then person 1 is guaranteed to have more heads after an additional flip; if E_0 happens, person 1 has probability 0.5 to; and if E_B happens, person 1 cannot have

*The solutions to questions marked with † are copied verbatim from MATH 408.

more heads. Also, since the coin is fair, $\mathbb{P}(E_A) = \mathbb{P}(E_B)$. Thus, after an additional flip,

$$\begin{aligned}\mathbb{P}(\text{person 1 has more heads}) &= \mathbb{P}(E_A) + \mathbb{P}(E_0)/2 \\ &= \mathbb{P}(E_A) + \frac{1 - \mathbb{P}(E_A) - \mathbb{P}(E_B)}{2} \\ &= \mathbb{P}(E_A) + \frac{1 - 2\mathbb{P}(E_A)}{2} = \frac{1}{2}.\end{aligned}$$

Problem 5

† Suppose a test for a disease is 99.9% accurate. That is, if you have the disease, the test will be positive with 99.9% probability. And if you do not have the disease, the test will be negative with 99.9% probability. Suppose also the disease is fairly rare, so that roughly 1 in 20,000 people have the disease. If you test positive for the disease, with what probability do you actually have the disease?

Solution. We apply Bayes' Theorem.

	Disease (D)	No Disease (ND)
Positive (+)	$0.999 \cdot 1/20000$	$0.001 \cdot 19999/20000$
Negative (-)	$0.001 \cdot 1/20000$	$0.999 \cdot 19999/20000$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.999 \cdot 1/20000}{0.999/20000 + 0.001 \cdot 19999/20000} \approx 0.048.$$

Problem 6: Inclusion-Exclusion Formula

Let Ω be a discrete sample space and let $\mathbb{P}$ be a probability law on Ω. Prove that if $A_1, \ldots, A_n \subset \Omega$ then

$$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Proof. Since Ω is discrete, it suffices to show that each $x \in \Omega$ is "counted" exactly once by the RHS, as it is by the LHS. Suppose $x \in \Omega$ and WLOG assume $x \in A_1 \cap A_2 \cap \cdots \cap A_m$ but $x \notin A_{m+1} \cup \cdots \cup A_n$, where $1 \leqslant m \leqslant n$. Then x is counted exactly

$$1 - \binom{m}{2} + \binom{m}{3} - \cdots + (-1)^{m-1} = -\sum_{k=1}^m (-1)^k \binom{m}{k}$$

times. Using binomial expansion on $(1-1)^m = \sum_{k=0}^m (-1)^k \binom{m}{k}$ we see $-\sum_{k=1}^m (-1)^k \binom{m}{k} = 1$, proving the claim. □

Problem 7

† A community has m facilities. Each family has at least one child. The largest family has $k > 0$ children. For each $i \in \{1, \cdots, k\}$, there are n_i families with i children so $n_1 + \cdots + n_k = m$. Choose a child randomly in the following ways.

(1) First choose one of the families uniformly at random among all the families. Then, in the chosen family, choose one of the children uniformly at random.

(2) Among all $n_1 + 2n_2 + \cdots + kn_k$ children, choose one uniformly at random.

what is the probability that the chosen child is the first-born in their family if you use method (1)? What about (2)?

Solution. For method 1, there is a probability of n_i/m to choose a family of i children. Then there is a probability of $1/i$ that the children picked is the first-born. Thus, the total probability is $m^{-1}\sum_{i=1}^{k} n_i/i$.
For the second method, we simply need to compute the number of first-born children and decide it by the total number of children. Clearly m family correspond to m first-born children, and there are $\sum_{i=1}^{k} i \cdot n_i$ children. Thus the total probability is $m/\sum_{i=1}^{k}(i \cdot n_i)$.

Problem 8

† You are trapped in a maze. Your starting point is a room with three doors. The first door will lead you to a corridor which lets you exist the maze after 3 hours of walking. The second dour leads you through a corridor which puts you back to the starting point of the maze after seven hours of walking. The third door leads you through a corridor which puts you back to the starting point of the maze after nine hours of walking. Each time at the starting point you choose one of the doors with equal probability. Let X be the number of hours it takes for you to exist the maze and let Y be the number of door that you initially choose.

- Compute $\mathbb{E}(X \mid Y = i)$, $i \in \{1, 2, 3\}$, in terms of $\mathbb{E}X$.
- Compute $\mathbb{E}X$.

Solution.

$$\begin{aligned}
\mathbb{E}(X \mid Y = 1) &= 3 \\
\mathbb{E}(X \mid Y = 2) &= 7 + \mathbb{E}X \\
\mathbb{E}(X \mid Y = 3) &= 9 + \mathbb{E}X.
\end{aligned}$$

Then, since $\mathbb{E}X = \sum_{i=1}^{3} \mathbb{P}(Y = i)\mathbb{E}(X \mid Y = i)$, we obtain the function

$$\frac{3 + 7 + \mathbb{E}X + 9 + \mathbb{E}X}{3} = \mathbb{E}X \implies \mathbb{E}X = 19.$$

Problem 9: Stein Identity

Let X be a standard Gaussian random variable and let $g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable with g, g' having polynomial volume growth, i.e., there exist $a, b > 0$ such that $|g(x)|, |g'(x)| \leqslant a(1+|x|)^b$. Prove the **Stein identity**

$$\mathbb{E}Xg(X) = Eg'(X)$$

and use it to recursively compute $\mathbb{E}X^k$ for any positive integer k.

Proof.

$$\begin{aligned}
\mathbb{E}Xg(X) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} xg(x)e^{-x^2/2}\,\mathrm{d}x \\
&= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x)\cdot(-xe^{-x^2/2})\,\mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x)\frac{\mathrm{d}}{\mathrm{d}x}(e^{-x^2/2})\,\mathrm{d}x \\
&= -g(x)e^{-x^2/2}\Big|_{x=-\infty}^{\infty} + \int_{\mathbb{R}} g'(x)e^{-x^2/2}\,\mathrm{d}x = \int_{\mathbb{R}} g'(x)e^{-x^2/2}\,\mathrm{d}x = \mathbb{E}g'(X).
\end{aligned}$$

The first term vanishes: $\lim_{x\to\pm\infty} g(x)e^{-x^2/2} = 0$ as exponential growth dominates polynomial growth.
Using this identity recursively and the fact that $\mathbb{E}X^1 = \mathbb{E}X^0 = \mathbb{E}1 = 1$,

$$\mathbb{E}X^k = \mathbb{E}XX^{k-1} = (k-1)\mathbb{E}X^{k-2} = \cdots = (k-1)!!.$$ □

Problem 10

Let $G = (V, E)$ be an undirected graph on the vertices $V = \{1, \ldots, n\}$. Using MAX-CUT, prove that there exists a cut (S, S^c) of the graph such that the number of edges going between S and S^c is at least $|E|/2$.
Hint: define a random $S \subset V$ such that, for every $i \in V$, $\mathbb{P}(i \in S) = 1/2$, and the events $1 \in S, 2 \in S, \ldots, n \in S$ are all independent.

Proof. Since the events of form $i \in S$ are independent, $\mathbb{P}(i \in S, j \notin S) = \mathbb{P}(i \in S)\mathbb{P}(j \notin S) = (1/2)(1-1/2) = 1/4$ for $i \neq j$ and in particular for $\{i, j\} \in E$. Since i, j are symmetric,

$$\mathbb{P}(i \in S \oplus j \in S) = \mathbb{P}(i \in S, j \notin S) + \mathbb{P}(i \notin S, j \in S) = \frac{1}{2}.$$

($\oplus$ denotes "exclusive or".) Therefore, summing over all $\{i, j\} \in S$, the expected value of number of edges going between S and S^c is $|S|/2$, and the remainder of the claim follows from MAX-CUT. □

Problem 11

Let $n \geqslant 2$ and let S^{n-1} be the boundary of the n-dimensional ball. Let $x \in S^{n-1}$ be fixed and let v be a random vector uniformly distributed in S^{n-1}. Prove that

$$\mathbb{E}|\langle x, v\rangle| \geqslant \frac{1}{10\sqrt{n}}.$$

Proof. First we reduce the claim to a much simpler case. Since the uniform distribution on S^{n-1} is invariant under rotations about the origin, and since inner product is also preserved under rotations, i.e., $\langle a, b\rangle = \langle Ra, Rb\rangle$, we have, for any rotation $R : S^{n-1} \to S^{n-1}$,

$$\mathbb{E}|\langle x, v\rangle| = \mathbb{E}|\langle x, Rv\rangle| = \mathbb{E}|\langle R^{-1}x, R^{-1}Rv\rangle| = \mathbb{E}|\langle R^{-1}x, v\rangle|.$$

For any $x \in S^{n-1}$, letting R be such that $R^{-1}x = u := (1, 0, \cdots, 0)$, we have

$$\mathbb{E}|\langle x, v\rangle| = \mathbb{E}|\langle u, v\rangle| = \frac{1}{\text{Area}(S^{n-1})} \int_{S^{n-1}} |v_1| \,\mathrm{d}V. \tag{Q9.1}$$

Note that, under spherical coordinates with parameters $r, \varphi_1, \varphi_2, \ldots, \varphi_{n-1}$, the first component v_1 can be expressed as $r \cos\varphi_1$, and the Jacobian is

$$r^{n-1} \prod_{i=1}^{n-2} \sin^{n-1-i}(\varphi_i) = r^{n-1} \sin^{n-2}(\varphi - 1) \sin^{n-3}(\varphi_2) \cdots \sin(\varphi_{n-2}).$$

In this case $r \equiv 1$ on S^{n-1} so we get two simpler $(n-1)$-fold integrals:

$$\int_{S^{n-1}} |v_1| \,\mathrm{d}V = \int_{\varphi_{n-1}=0}^{2\pi} \int_{\varphi_{n-2}=0}^{\pi} \cdots \int_{\varphi_1=0}^{\pi} |\cos\varphi_1| \prod_{i=1}^{n-2} \sin^{n-1-i}(\varphi_i) \,\mathrm{d}\varphi_1 \cdots \mathrm{d}\varphi_{n-2} \,\mathrm{d}\varphi_{n-1} \tag{Q9.2}$$

and

$$\text{Area}(S^{n-1}) = \int_{S^{n-1}} 1 \,\mathrm{d}V = \int_{\varphi_{n-1}=0}^{2\pi} \int_{\varphi_{n-2}=0}^{\pi} \cdots \int_{\varphi_1=0}^{\pi} \prod_{i=1}^{n-2} \sin^{n-1-i}(\varphi_i) \,\mathrm{d}\varphi_1 \cdots \mathrm{d}\varphi_{n-2} \,\mathrm{d}\varphi_{n-1}. \tag{Q9.3}$$

Division gives $\mathbb{E}|\langle x, v\rangle| = (\text{Q9.2})/(\text{Q9.3}) = \int_0^\pi |\cos\varphi| \sin^{n-2}\varphi \,\mathrm{d}\varphi \Big/ \int_0^\pi \sin^{n-2}\varphi \,\mathrm{d}\varphi$. Since both integrals satisfy $\int_0^\pi = 2\int_0^{\pi/2}$, the ratio further equals $\int_0^{\pi/2} \cos\varphi \sin^{n-2}\varphi \,\mathrm{d}\varphi \Big/ \int_0^{\pi/2} \sin^{n-2}\varphi \,\mathrm{d}\varphi$. The numerator is $1/(n-1)$ by a simple u-substitution with $u := \sin\varphi$, and for $n \geqslant 3$, the denominator is bounded by 0 and $\sqrt{\pi/2(n-2)}$ since $\cos x \leqslant \exp(-x^2/2)$ on $[0, \pi/2]$ and[1]

$$\begin{aligned}\int_0^{\pi/2} \sin^{n-2}\varphi \,\mathrm{d}\varphi = \int_0^{\pi/2} \cos^{n-2}\varphi \,\mathrm{d}\varphi &\leqslant \int_0^{\pi/2} \exp(-(n-2)x^2/2) \,\mathrm{d}x \\ &< \int_0^{\infty} \exp(-(n-2)x^2/2) \,\mathrm{d}x = \frac{1}{2} \cdot \sqrt{2\pi/(n-2)}.\end{aligned}$$

For $n = 2$, it is immediate that

$$\int_0^{\pi/2} \sin^2\varphi \,\mathrm{d}\varphi = \frac{1}{2} \int_0^{\pi/2} \sin^0\varphi \,\mathrm{d}\varphi = \pi/4$$

using the well-known reduction formula

$$\int_0^{\pi/2} \sin^k\varphi \,\mathrm{d}\varphi = \frac{k-1}{k} \int_0^{\pi/2} \sin^{k-2}\varphi \,\mathrm{d}\varphi.$$

Therefore, for $n = 2$, $\mathbb{E}|\langle x, v\rangle| = 4/\pi > 1/(10\sqrt{2})$ and for $n \geqslant 3$,

$$10\sqrt{n} \cdot \mathbb{E}|\langle x, v\rangle| \geqslant \frac{20}{\sqrt{\pi}} \cdot \left(\frac{n^2 - 2n}{n^2 - 2n + 1}\right)^{1/2} \geqslant \frac{20\sqrt{3/4}}{\sqrt{\pi}} > 1.$$

(Note that $(n^2 - 2n)/(n^2 - 2n + 1)$ is monotone on $[3, \infty)$ and equals $3/4$ at 3.) This proves the claim. □

[1] Without dominating $\cos x$ by $\exp(-x^2/2)$ on $[0, \pi/2]$, I was quite stuck. But one day when browsing Zhihu, a Chinese version of Quora, I accidentally stumbled into a related question (hyperlinked in PDF), providing me with exactly what I need.

Problem 12: The Power Method

Let A be an $n \times n$ real symmetric matrix. Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the unknown eigenvalues of A and let $v_1, \ldots, v_n \in \mathbb{R}^n$ be the corresponding normalized unknown eigenvectors of A.
Given A, our first goal is to first find λ_1 and v_1. For simplicity assume $1/2 < \lambda_1 < 1$ and $0 \leqslant \lambda_n \leqslant \cdots \leqslant \lambda_2 < 1/4$. Suppose we have found a vector $v \in \mathbb{R}^n$ with $\|v\| = 1$ and $|\langle v, v_1\rangle| > 1/n$. Show that $A^k v$ approximates v_1 as k becomes large. More specifically, show that for $k \geqslant 1$,

$$\|A^k v - \langle v, v_1\rangle \lambda_1^k v_1\|^2 \leqslant \frac{n-1}{16^k}.$$

Hint: the spectral theorem.
Since $|\langle v, v_1\rangle \lambda_1^k| > 2^{-k}/n$, the inequality implies $A^k v$ is approximate an eigenvector with eigenvalue λ_1. That is, by the triangle inequality

$$\|A(A^k v) - \lambda_1(A^k v)\| \leqslant \|A^{k+1}v - \langle v, v_1\rangle \lambda_1^{k+1} v_1\| + \lambda_1 \| \langle v, v_1\rangle \lambda_1^k v_1 - A^k v\| \leqslant \frac{2\sqrt{n-1}}{4^k}$$

and by reverse triangle inequality

$$\|A^k v\| = \|A^k v - \langle v, v_1\rangle \lambda_1^k v_1 + \langle v, v_1\rangle \lambda_1^k v_1\| \geqslant n^{-1}2^{-k} - 4^{-k}\sqrt{n-1}.$$

In conclusion, if we take k large, say $k > 10\log n$, and if we define $z := (A^k v)/\|A^k v\|$, then

$$\|Az - \lambda_1 z\| \leqslant 4n^{3/2}2^{-k} < 4n^{-4},$$

and the corresponding λ_1 is simply $z^T Az/z^T z$.

Proof. By the spectral theorem, A admits n eigenvalues $\lambda_1, \cdots, \lambda_n$ corresponding to n independent eigenvectors $v_1, \cdots, v_n$. Let v be given as stated; we can express it as a linear combination $v = \sum_{i=1}^n \langle v, v_i\rangle v_i$. Then,

$$A^k v = \sum_{i=1}^n A^k \langle v, v_i\rangle v_i = \sum_{i=1}^n \langle v, v_i\rangle \lambda_i^k v_i.$$

Therefore,

$$\begin{aligned}
\|A^k v - \langle v, v_1\rangle \lambda_1^k v_1\|^2 &= \left\| \sum_{i=1}^n \langle v, v_i\rangle \lambda_i^k v_i - \langle v, v_1\rangle \lambda_1^k v_1 \right\|^2 \\
&= \left\| \sum_{i=2}^n \langle v, v_i\rangle \lambda_i^k v_i \right\|^2 = \sum_{i=2}^n (|\langle v, v_i\rangle||\lambda_i|^k)^2 \\
&\leqslant \sum_{i=2}^n \|v\|\|v_i\||\lambda_i|^{2k} = 2^{-2k}(n-2) < 16^{-k}(n-1).
\end{aligned}$$

□

Problem 13

† Let X_1, Y_1 be random variables with joint PDF f_{X_1,Y_1}. Let X_2, Y_2 be random variables with joint PDF f_{X_2,Y_2}. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be inverses of each other. Let $J(x, y)$ denote the determinant of the

Jacobian of S at (x, y). Assume that $(X_2, Y_2) = T(X_1, Y_1)$. Show that

$$f_{X_2,Y_2}(x, y) = f_{X_1,Y_1}(S(x, y))|J(x, y)|.$$

Proof. Using change of variable formula,

$$\begin{aligned}\int_U f_{X_2,Y_2}(x, y)\,\mathrm{d}x\mathrm{d}y &= \mathbb{P}((X_2, Y_2) \in U)\\ &= \mathbb{P}((X_1, Y_1) \in S(U))\\ &= \int_{S(U)} f_{X_1,Y_1}(x, y)\,\mathrm{d}x\mathrm{d}y\\ &= \int_U f_{X_1,Y_1}(S(x, y))|J(x, y)|\,\mathrm{d}x\mathrm{d}y\end{aligned}$$

for all measurable U. The equality stated in the problem follows. □

Problem 14

† Problems 11 and 12 skipped. One possible way is to use sample mean and variance.

Problem 16

† n people are about to be interviewed, each having a distinct rank $1 \leqslant a_i \leqslant n$, $1 \leqslant i \leqslant n$. For each $1 \leqslant i \leqslant n$, upon interviewing the i^{th} person, if $a_i > a_j$ for all $1 \leqslant j < i$ then the i^{th} person is hired. That is, if the person currently being interviewed is better than all previous candidates, they will be hired. What is the expected number of hiring that will be made?

Hint: let $X_i = 1$ if the i^{th} person to arrive is hired and let $X_i = 0$ otherwise. Consider $\sum_{i=1}^{n} X_i$.

Solution. Let X_i be the indicator variable that evaluates to 1 if the i^{th} person to arrive is hired. Then, among the first i people to arrive, since each permutation is equally likely to occur, $\mathbb{E}X_i = 1/i$. Clearly $\sum_{i=1}^{n} X_i$ is the random variable whose output represents the number of people hired, so it remains to compute its expectation:

$$\mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}X_i = \sum_{i=1}^{n} \frac{1}{i}.$$