MATH 340 – SPRING 2026 – HOMEWORK 4

Due Thursday, February 5, 2026 at 8am on Gradescope. You must justify all of your answers for full credit.

You are encouraged to collaborate with other students, but you must write up your solutions individually, without reference to notes from the collaboration. You may not search the internet or ask AI for solutions to the homework problems.

Problem 1. In a certain community (which I make no claim actually exists, or corresponds to any real community), families desire to have as few children as possible, but they want to have at least one female child. Let us assume that there are no multiple births, and that each child is assigned either male or female at birth, with equal probability, and that the sexes assigned to all births are independent. Each family has children until a girl is born, and then stops.

(a) What is the expected number of male and female children in each family? How does this reproduction strategy affect the balance of males to females in the community?

(b) Conditional on the total number of children in a family being n, what is the conditional distribution of the number of boys in the family?

In a different community, after having each child, families decide to keep having children with probability $p \in (0, 1)$, independent of everything else (i.e. they flip an unfair coin after each birth).

(c) Conditional on the total number of children in a family being n, what is the conditional distribution of the number of boys in the family?

(d) What is the expected number of male and female children in each family? How does this reproduction strategy affect the balance of males to females in the community? [*Hint*: if N is the total number of children and M is the number of boys, start by computing $E[M \mid N = n]$.]

Let's now generalize the previous examples. Suppose in another community, there is some reproduction strategy that we do not know. All we know is that this community does not practice sex-selective abortion, so after having each child, parents decide whether or not to (try to) have another child without knowing what the gender of that child would be. We can model this situation as follows. Let N be the total number of children that the family has. Assume that $N \le 20$ almost surely. Let ξ_i be the indicator of the event that the ith child would be female if that child were actually born, so $\xi_1, \dots, \xi_{20}$ are independent.

(e) Explain why, based on the description given, we should make the modeling assumption that, for each i, the random variables $\mathbf{1}\{N \ge i\}, \xi_i, \xi_{i+1}, \dots, \xi_{20}$ are independent. Also, explain why the total number of girls that the family has is

$$G := \sum_{i=1}^{20} \mathbf{1}\{N \ge i\}\xi_i$$

and the total number of boys that the family has is

$$B := \sum_{i=1}^{20} \mathbf{1}\{N \ge i\}(1 - \xi_i).$$

(f) Compute $E[G - B]$. How does the strategy affect the balance of males to females in the community?

Solution.

(a) The expected number of female children in each family is 1, since each family has exactly one girl. If B denotes the number of male children, then $B+1 \sim \text{Geom}(1/2)$, so $E[B+1] = 2$, so $E[B] = 1$. Thus the expected numbers of boys and girls born in the community are equal.

(b) Conditional on the total number of children N being n, there are always exactly $n-1$ boys, so $P(B = k \mid N = n) = \mathbf{1}_{k=n-1}$.

(c) Let B_k be the number of boys in the first k (hypothetical) children. (I say "hypothetical" since we include boys who weren't actually born.) Then $B_k \sim \text{Bin}(k, 1/2)$ is independent of the number of children N who were actually born. So we have

$$\begin{aligned} P(B = m \mid N = n) &= \frac{P(B = m \text{ and } N = n)}{P(N = n)} = \frac{P(B_n = m \text{ and } N = n)}{P(N = n)} \\ &= \frac{P(B_n = m)P(N = n)}{P(N = n)} = P(B_n = m) = \binom{n}{m}2^{-n}. \end{aligned}$$

(d) Since, conditional on $N = n$, the distribution of B is $\text{Bin}(n, 1/2)$, we have $E[B \mid N = n] = n/2$, so $E[B \mid N] = N/2$. Also, we have $E[N] = 2$ since $N \sim \text{Geom}(1/2)$. Therefore, we have

$$E[B] = E[E[B \mid N]] = E[N/2] = 1.$$

By the same logic, $E[G] = 1$ where G is the number of girls born. So again the expected numbers of boys and girls are the same.

(e) The event $\{N \geq i\}$ is the event that the parents decided to continue having children after child $i-1$ was born, and they had to make this determination without knowing the sexes of the children after i. Since the sexes of the children after i are independent of each other and of the sexes of the previous children, it makes sense that we should model them as independent from $\mathbf{1}\{N \geq i\}$. The formula for G is just the sums of the indicators that each child is a girl, but multiplied the indicator of the event that that child is actually born. The formula for B is obtained similarly.

(f) We have

$$\begin{aligned} E[G - B] &= E\left[\sum_{i=1}^{20} \mathbf{1}\{N \geq i\}(1 - 2\xi_i)\right] = \sum_{i=1}^{20} E\left[\mathbf{1}\{N \geq i\}(1 - 2\xi_i)\right] \\ &= \sum_{i=1}^{20} E[\mathbf{1}\{N \geq i\}]E\left[1 - 2\xi_i\right] = \sum_{i=1}^{20} E[\mathbf{1}\{N \geq i\}]\left(1 - 2E[\xi_i]\right) = 0 \end{aligned}$$

since $1 - 2E[\xi_i] = 1 - 2 \cdot (1/2) = 0$. So the expected numbers of girls and boys in the community are still equal regardless of the strategy.

Problem 2 (Airplane problem again)**.** Recall again the chaotic airplane passengers of Homework 1, Problem 3. Compute the expected number of people who sit in their assigned seats. [*Hint*: use linearity of expectation and don't work too hard.]

Solution. Let ξ_i be the indicator of the event that person i sits in seat i. Then $E[\xi_i]$ is the probability that person i sits in seat i, which is $\frac{(n-1)!}{n!} = \frac{1}{n}$ since there are $(n-1)!$ configurations in which person i sits in the correct seat and $n!$ total configurations, all with the same probability. So the expected number of people who sit in their assigned seats is

$$E\left[\sum_{i=1}^{n} \xi_i\right] = \sum_{i=1}^{n} E[\xi_i] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Problem 3. Let $\xi_1, \dots, \xi_n$ be i.i.d. Rademacher random variables. Compute

$$R_n \coloneqq E\left[\left(\sum_{i=1}^n \xi_i\right)^4\right].$$

Find an α such that R_n/n^α approaches a finite limit as $n \to \infty$. [*Hint*: expand the sum as $(\sum_{i=1}^n \xi_i)^4 = \sum_{i_1,i_2,i_3,i_4=1}^n \xi_{i_1}\xi_{i_2}\xi_{i_3}\xi_{i_4}$, and then separate the right hand side into different parts depending on which of the i_js are the same or different.]

Proof. We start by writing

$$E\left[\left(\sum_{i=1}^n \xi_i\right)^4\right] = \sum_{i_1,i_2,i_3,i_4=1}^n E\left[\xi_{i_1}\xi_{i_2}\xi_{i_3}\xi_{i_4}\right].$$

There are a total of n^4 terms in this sum. We collect them into several different categories:

- If $i_1 = i_2 = i_3 = i_4$ (n terms), then $E\left[\xi_{i_1}\xi_{i_2}\xi_{i_3}\xi_{i_4}\right] = E[1] = 1$.
- If $i_1 = i_2$ and $i_3 = i_4$, but $i_2 \neq i_3$ ($n(n-1)$ terms), then $E\left[\xi_{i_1}\xi_{i_2}\xi_{i_3}\xi_{i_4}\right] = E[1] = 1$.
- Similarly, if $i_1 = i_3 \neq i_2 = i_4$ ($n(n-1)$ terms) or $i_1 = i_4 \neq i_3 = i_3$ ($n(n-1)$ terms), then $E\left[\xi_{i_1}\xi_{i_2}\xi_{i_3}\xi_{i_4}\right] = 1$.
- Otherwise, there is an i_j that is not equal to any of the other i_ks, and then the expectation is 0 by independence since $E[\xi_{i_j}] = 0$.

So altogether the expectation is $R_n = n + 3n(n-1)$, and we can choose $\alpha = 2$ so that $R_n/n^2 \to 3$ as $n \to \infty$. □

Problem 4 (second moment method)**.** Suppose that X is a random variable such that $X \geq 0$ almost surely. Show that

$$P(X = 0) \leq \frac{\mathrm{Var}(X)}{(E[X])^2}.$$

[*Hint*: start with the inequality $P(X = 0) \leq P(|X - E[X]| \geq E[X])$.]

Solution. If $X = 0$, then certainly $|X - E[X]| = |E[X]| \leq E[X]$, so we have

$$P(X = 0) \leq P(|X - E[X]| \geq E[X]) \leq \frac{\mathrm{Var}(X)}{(E[X])^2}$$

by Chebyshev's inequality.

Problem 5. A *graph* consists of a set $\mathcal{V}$ of *vertices* and a set of *edges* $\mathcal{E} \subseteq \binom{V}{2}$ (so an edge is an ordered pair of distinct vertices). We think of the edge set as the set of vertices that are "adjacent" to each other in the graph. For $n \in \mathbb{N}$, let $\mathcal{G}$ be a *random* graph with vertex set $\mathcal{V} = \{1, \dots, n\}$ and each edge present independently with probability p. For $i \neq j \in \mathcal{V}$, let $\xi_{i,j}$ denote the indicator of the event that $\{i,j\} \in E$, i.e. that i and j are adjacent. This means that the family $(\xi_{i,j})_{\{i,j\} \in \binom{\{1,\dots,n\}}{2}}$ is a family of $\binom{n}{2}$ independent $\mathrm{Ber}(p)$ random variables.

(a) Compute $E|\mathcal{E}|$, the expected number of edges in the graph, in terms of n and p.

(b) A *triangle* in a graph is a set of distinct vertices $i, j, k \in \mathcal{V}$ such that $\{i,j\}, \{j,k\}, \{i,k\} \in \mathcal{V}$ (i.e. all three vertices are connected). Let T be the number of triangles in $\mathcal{G}$. Compute $E[T]$. [*Hint*: for i, j, k, let $\eta_{i,j,k}$ be the indicator of the event that i, j, k form a triangle, and start by computing $E[\eta_{i,j,k}]$.]

(c) Suppose that $p = p_n$ is such that $np_n \to 0$ as $n \to \infty$. Show that the probability that $\mathcal{G}$ contains at least one triangle goes to 0 as $n \to \infty$. [*Hint*: previous part and Markov's inequality.]

(d) Compute $\mathrm{Var}(T)$. [*Hint*: this is somewhat similar to Problem 3. The $\eta_{i,j,k}$s are not all independent, but many of them are.]

(e) Suppose that $p = p_n$ is such that $np_n \to \infty$ as $n \to \infty$. Show that the probability that $\mathcal{G}$ contains at least one triangle goes to 1 as $n \to \infty$. [*Hint*: use Problems 4, 5(b), and 5(d).]

Solution.

(a) We have

$$E|\mathcal{E}| = E\left[\sum_{\{i,j\}\in\binom{\mathcal{V}}{2}} \xi_{i,j}\right] = \sum_{\{i,j\}\in\binom{\mathcal{V}}{2}} E[\xi_{i,j}] = p\binom{|\mathcal{V}|}{2} = p\binom{n}{2}.$$

(b) We have $\eta_{ijk} = \xi_{ij}\xi_{jk}\xi_{ik}$, so

$$E[\eta_{ijk}] = E[\xi_{ij}\xi_{jk}\xi_{ik}] = E[\xi_{ij}]E[\xi_{jk}]E[\xi_{ik}] = p^3.$$

Therefore, we have

$$E[T] = E\left[\sum_{\{i,j,k\}\in\binom{\mathcal{V}}{3}} \eta_{ijk}\right] = \sum_{\{i,j,k\}\in\binom{\mathcal{V}}{3}} E[\eta_{ijk}] = p^3\binom{n}{3}.$$

(c) We have by Markov's inequality that

$$P(T \geq 1) \leq E[T] = p_n^3\binom{n}{3} = \frac{p_n^3 n(n-1)(n-2)}{6} = \frac{(np_n)^3(1-1/n)(1-2/n)}{6} \to 0$$

as $n \to \infty$ since $np_n \to 0$ and $1 - 1/n, 1 - 2/n \to 1$.

(d) We can compute

$$E[T^2] = E\left[\left(\sum_{\{i,j,k\}\in\binom{\mathcal{V}}{3}} \eta_{ijk}\right)^2\right] = \sum_{\{i,j,k\}\in\binom{\mathcal{V}}{3}} \sum_{\{i',j',k'\}\in\binom{\mathcal{V}}{3}} E[\eta_{ijk}\eta_{i'j'k'}].$$

Now we again break up the sum into different pieces:

- If $\{i,j,k\} = \{i',j',k'\}$ ($\binom{n}{3} = \frac{1}{6}n(n-1)(n-2)$ terms), then $E[\eta_{ijk}\eta_{i'j'k'}] = E[\eta_{ijk}] = p^3$.
- If $\{i,j,k\}$ and $\{i',j',k'\}$ have two elements in common, then $E[\eta_{ijk}\eta_{i'j'k'}] = p^5$ since five edges need to be present for both triangles to be present. The number of such terms is the number of ways of picking the two common elements times the number of ways of picking the other element of $\{i,j,k\}$ times the number of ways of picking the other element of $\{i',j',k'\}$ which is $\binom{n}{2}(n-2)(n-3) = \frac{1}{2}n(n-1)(n-2)(n-3)$.
- If $\{i,j,k\}$ and $\{i',j',k'\}$ have one or zero elements in common, then $E[\eta_{ijk}\eta_{i'j'k'}] = p^6$ since six edges need to be present for both triangles to be present. The number of such terms with one element in common is the number of ways of picking the common element times the number of ways of choosing the other two elements of $\{i,j,k\}$ times the number of ways of choosing the other two elements of $\{i',j',k'\}$, which is $n\binom{n-1}{2}\binom{n-3}{2} = \frac{1}{4}n(n-1)(n-2)(n-3)(n-4)$. The number of such terms with no elements in common is $\binom{n}{3}\binom{n-3}{3} = \frac{1}{36}n(n-1)(n-2)(n-3)(n-4)(n-5)$.

Adding up all of these contributions, we get

$$E[T^2] = \frac{p^3}{6}n(n-1)(n-2) + \frac{p^5}{2}n(n-1)(n-2)(n-3) + p^6\left[\frac{1}{4}n(n-1)(n-2)(n-3)(n-4) + \frac{1}{36}n(n-1)(n-2)(n-3)(n-4)(n-5)\right],$$

so

$$\begin{aligned}
\mathrm{Var}(T) &= E[T^2] - (E[T])^2 \\
&= \frac{p^3}{6} n(n-1)(n-2) + \frac{p^5}{2} n(n-1)(n-2)(n-3) \\
&\quad + \frac{p^6}{4} n(n-1)(n-2)(n-3)(n-4) \\
&\quad + \frac{p^6}{36} \left\{ n(n-1)(n-2)(n-3)(n-4)(n-5) - n^2(n-1)^2(n-2)^2 \right\} \\
&= \frac{p^3}{6} n(n-1)(n-2) + \frac{p^5}{2} n(n-1)(n-2)(n-3) + \frac{p^6}{4} n(n-1)(n-2)(n-3)(n-4) \\
&\quad + \frac{p^6}{36} n(n-1)(n-2) \left\{ (n-3)(n-4)(n-5) - n(n-1)(n-2) \right\}
\end{aligned}$$

(e) We have

$$\begin{aligned}
\frac{\mathrm{Var}(T)}{(E[T])^2} &= \frac{\frac{p^3}{6} n(n-1)(n-2)}{\frac{p^6}{36} n^2(n-1)^2(n-2)^2} + \frac{\frac{p^5}{2} n(n-1)(n-2)(n-3)}{\frac{p^6}{36} n^2(n-1)^2(n-2)^2} + \frac{\frac{p^6}{4} n(n-1)(n-2)(n-3)(n-4)}{\frac{p^6}{36} n^2(n-1)^2(n-2)^2} \\
&\quad + \frac{\frac{p^6}{36} n(n-1)(n-2) \left\{ (n-3)(n-4)(n-5) - n(n-1)(n-2) \right\}}{\frac{p^6}{36} n^2(n-1)^2(n-2)^2} \\
&= \frac{1}{\frac{p^3}{6} n(n-1)(n-2)} + \frac{n-3}{\frac{p}{18} n(n-1)(n-2)} + \frac{9(n-3)(n-4)}{n(n-1)(n-2)} \\
&\quad + \frac{(n-3)(n-4)(n-5) - n(n-1)(n-2)}{n(n-1)(n-2)} \\
&\sim \frac{6}{p^3 n^3} + \frac{18}{pn^2} + \frac{9}{n} + \frac{(n-3)(n-4)(n-5) - n(n-1)(n-2)}{n^3}.
\end{aligned}$$

Now as $n \to \infty$, $p^3 n^3 = (pn)^3 \to \infty$ by assumption, so the first term goes to 0. Similarly, $pn^2 = (np)n \to \infty$, so the second term goes to 0. The third term clearly goes to 0 and the fourth term goes to 0 because the highest power of n in the numerator is n^2 (since the n^3 terms cancel). Therefore, we have

$$P(T = 0) \leq \frac{\mathrm{Var}(T)}{(E[T])^2} \to 0$$

by the second moment method, so $P(T \geq 1) \to 1$.