Final Exam Solutions

7:00-10:00pm, 10 May 2024

Your First Name: **Your Last Name:**

SIGN Your Name: **Your SID Number:**

Your Exam Room:

Name of Person Sitting on Your Left:

Name of Person Sitting on Your Right:

Name of Person Sitting in Front of You:

Name of Person Sitting Behind You:

Instructions:

(a) *As soon as the exam starts,* **please write your student ID in the space provided at the top of every page!** *(We will remove the staple when scanning your exam.)*

(b) *There are 15 pages (2-sided) on the exam. Notify a proctor immediately if a page is missing.*

(c) *We will* **not** *grade anything outside of the space provided for a question (i.e., either a designated box if it is provided, or otherwise the white space immediately below the question).* **Be sure to write your full answer in the box or space provided!** *Scratch paper is provided on request; however, please bear in mind that nothing you write on scratch paper will be graded!*

(d) **The questions vary in difficulty, so if you get stuck on any question you are strongly advised to leave it and return to it later. In particular, you may find some of the later questions easier than some of the earlier ones.** *There is no penalty for incorrect answers.*

(e) *In the interests of fairness, proctors will* **not** *answer clarifying questions during the exam. If you believe there is an ambiguity or error in a problem, write down what you think is the most appropriate response and move on. Any legitimate such issues will be handled during grading.*

(f) *You may consult* **2 two-sided "cheat sheets"** *of notes. Apart from that, you may not look at any other materials. Calculators, phones, computers, and other electronic devices are* **not** *permitted.*

(g) *You may use, without proof, theorems and lemmas that were proved in the notes and/or in lecture.*

(h) *You have 180 minutes: there are 14 questions on this exam worth a total of 140 points.*

[exam starts on next page]

1. Multiple Choice [21 points].

Answer by shading the **single** *correct bubble in each case; no justifications; no penalty for incorrect answers.*

(a) TRUE or FALSE: Suppose $\forall x \geq 0\,(\neg P(x))$. Then, $\forall x \geq 0\,\big(P(x) \implies (\exists y\; y \geq x \wedge Q(y))\big)$. *2pts*

● TRUE ○ FALSE

The predicate to the left of the implication is FALSE for all $x \geq 0$, so the formula is TRUE.

(b) Recall that a propositional form is a statement that combines propositional variables with standard connectives $\wedge$, $\vee$, and $\neg$. For this problem, assume there are no quantifiers. *2pts*
The number of unique (up to logical equivalence) propositional forms on n propositional variables $P_1, P_2, \ldots, P_n$ is

○ n ○ $2n$ ○ n^2 ○ 2^n ● 2^{2^n}

Consider the truth table for any propositional form F on $P_1, P_2, \ldots, P_n$. It has 2^n rows, where the value of F on each row is TRUE or FALSE (2 choices). So the total number of possible truth tables is 2^{2^n}, which is also the number of unique propositional forms up to logical equivalence.

(c) Let n and p be integers where p is an odd prime. If $n \equiv -n \pmod p$ then $n \pmod p$ must be *2pts*

● 0 ○ 1 ○ 2 ○ $p/2$ ○ $p-1$

Adding n to both sides, we get $2n \equiv 0 \pmod p$, which means that there exists an integer d s.t. $2n = dp$. Since p is odd, it means that p divides n.

(d) TRUE or FALSE: There exists a computer program which takes as inputs a program P, an input x for P, and a line number n, and determines whether the nth line of code in P is executed when P is run on input x. *2pts*

○ TRUE ● FALSE

If such a program R existed, then we could use it to solve the halting problem by taking the input program P' to the halting problem, adding a line after P', and then querying R whether that line is executed (reached) in the resulting program. Therefore, such a program R cannot exist.

(e) The number of polynomials of degree at most k with integer coefficients (in $\mathbb{Z}$) is *2pts*

○ Finite ● Countably Infinite ○ Uncountably Infinite

Each polynomial can be specified as a tuple of k integer coefficients $(a_{k-1}, a_{k-2}, \ldots, a_1, a_0) \in \mathbb{Z}^k$.

(f) Suppose we wish to send a message comprising 10 packets, each of which is an integer between 0 and 15, over a channel which corrupts up to 6 packets. Using the Reed-Solomon Code, the polynomial we construct must be over $GF(p)$ where the minimum value of prime p is *2pts*

○ 11 ○ 17 ● 23 ○ 29

Number of packets $n = 10$, number of errors $k = 6$, Reed-Solomon needs to transmit $n + 2k = 22$ packets, so p must be the minimum prime larger than 22, i.e., 23.

(g) Let G be an undirected graph on n vertices that consists of k trees on disjoint sets of vertices. The number of edges in G is *2pts*

○ $n-1$ ○ $k(n-1)$ ● $n-k$ ○ $n-k+1$ ○ not determined

Label the trees $T_1, T_2, \ldots, T_k$. We know that tree T_i has $n_i - 1$ edges, where n_i is the number of vertices in T_i. Thus the total number of edges in G is $\sum_{i=1}^{k}(n_i - 1) = (\sum_{i=1}^{k} n_i) - k = n - k$.

(h) If n, m, k are positive integers where $k \leq n, k \leq m$, then $\sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i}$ is equal to *2pts*

● $\binom{n+m}{k}$ ○ $\binom{nm}{k}$ ○ 2^{n+m} ○ $\frac{n!m!}{(k!)^2}$ ○ none of these

We can provide a combinatorial proof for the identity $\sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k}$. The RHS is the number of ways of selecting a subset of k elements from a set of $n+m$ elements with n of Type 1 and m of Type 2. The LHS expresses this as the sum of all ways of splitting the selection of k into i elements of Type 1 from n and $k-i$ of Type 2 from m.

(i) Let $u = 0000000$ and $v = 1011001$ be two vertices in the 7-dimensional hypercube. The number of *shortest* (in the sense of having fewest edges) paths between u and v is *2pts*

○ 1 ○ 4 ○ 16 ● 24 ○ 128

u and v differ in exactly 4 coordinates, so any shortest path between u and v corresponds to "flipping" these four coordinates (from 0 to 1) in some order. Since there are $4! = 24$ orderings of four items, this is also the number of shortest paths.

(j) Suppose an undirected simple graph (i.e., no self-loops or multiple edges) has n vertices and m edges. For which of the following combinations of n and m is it **possible** that the graph contains an Eulerian tour? For each combination choose TRUE if such a tour **may** exist, and FALSE if such a tour **cannot** exist.

TRUE FALSE

● ○ $n = 7,\ m = 7$ *1pt*

The cycle graph on 7 vertices is clearly Eulerian as the entire graph is an Eulerian tour.

○ ● $n = 4,\ m = 5$ *1pt*

Note that the average degree in this case is $\frac{2m}{n} = 2.5$. Therefore, there must be at least one vertex of degree 3 (larger degree is not possible as there are only 4 vertices). This immediately implies that the graph has no Eulerian tour, since there is an odd-degree vertex.

● ○ $n = 5,\ m = 7$ *1pt*

Consider a graph with 5 vertices, two of which (call them u, v) are connected to all the other vertices. This graph has $m = 7$ edges, and is Eulerian because all vertices have even degree. (u, v each have degree 4, and the other three vertices have degree 2.)

2. Short Answers [9 points].

(a) The value of $3^{42} \pmod{55}$ is *2pts*

Write your answer in the box; no explanation needed.

ANS: 9.
We use the identity $a^{(p-1)(q-1)} \equiv 1 \pmod{pq}$ for $a = 3, p = 5, q = 11$, and then multiply both sides by $3^2 = 9$ to get the answer.

(b) If $17x \equiv 7 \pmod{42}$, then x is *2pts*

Write your answer in the box; no explanation needed.

ANS: 35.
5 is the inverse of $17 \pmod{42}$, so multiplying the equation by 5 on both sides, we get $x \equiv 35 \pmod{42}$.

(c) Suppose we are using the standard polynomial secret sharing scheme over $GF(5)$ and $k = 3$ shares suffice to reconstruct the secret $P(0)$. Let the shares be $P(1) = 1$, $P(2) = 1$, and $P(4) = 2$. Then the value of the secret is *3pts*

Write your answer in the box; no explanation needed.

ANS: 3. The polynomial obtained after solving a system of equations is $x^2 + 2x + 3$.

(d) Consider the function $f(x) = x^2 \pmod 7$ mapping $\{0, 1, 2, 3, 4, 5, 6\}$ to $\{0, 1, 2, 3, 4, 5, 6\}$. *2pts*
Is f a bijection? If yes, give its inverse. If not, give values of $x, f(x)$ justifying why.

Write your answer by shading the appropriate circle and writing in the box only; no further explanation needed.

● Not a bijection

It suffices to find a pair (x, x') for which $f(x) = f(x')$. There are three possible such pairs: namely, $f(1) = f(6) \equiv 1 \pmod 7$, $f(2) = f(5) \equiv 4 \pmod 7$ and $f(3) = f(4) \equiv 2 \pmod 7$. [In fact, for *any* modulus $n > 2$ (not necessarily prime), and any $x \in \{1, 2, \ldots, n-1\}$, we always have $x^2 \equiv (n-x)^2 \pmod n$, so $f(x) = x^2 \pmod n$ is *never* a bijection.]

[exam continues on next page]

3. Proofs [10 points].

(a) Prove by induction that, for any $x \neq 1$, the following identity holds for all $n \geq 1$. *5pts*

$$\frac{x^n - 1}{x - 1} = \sum_{i=0}^{n-1} x^i$$

Clearly label your base case and induction step.
(**Note:** No points will be awarded for proofs that do not use induction!)

We proceed by induction.

Base Case: For $n = 1$ we see

$$\frac{x^1 - 1}{x - 1} = 1 = x^0 = \sum_{i=0}^{0} x^i.$$

Inductive Hypothesis: Suppose the statement is true for some $n = k \geq 1$.
Inductive Step: We observe

$$\begin{aligned} x^{k+1} - 1 &= (x^k - 1)x + (x - 1) \\ &= (x - 1)\left(\sum_{i=1}^{k} x^i\right) + (x - 1) \\ &= (x - 1)\sum_{i=0}^{k} x^i \end{aligned}$$

and dividing both sides by $x - 1$ (since $x \neq 1$) gives us the desired expression. □

(b) Prove that if $2^m - 1$ is prime for some natural number m, then m is also prime. [HINT: Try a proof by contraposition, and use the previous part.] *5pts*

We proceed by contraposition. Suppose m is composite. This means that there exists some distinct integers a and b not equal to 1 and m so that $m = ab$. Then,

$$2^m - 1 = 2^{ab} - 1 = (2^a)^b - 1.$$

If we denote $x = 2^a$, then our expression is equivalent to $x^b - 1$. From the previous part, we then have

$$2^m - 1 = x^b - 1 = (x - 1)\left(\sum_{i=0}^{b-1} x^i\right)$$

which shows that $2^m - 1$ is composite, since $x \neq 1$. □

4. Stable Matching [6 points]

Consider the following sets of matching preferences for four jobs 1, 2, 3, 4 and four candidates A, B, C, D.

Job	Preferences
1	$A > B > C > D$
2	$B > C > D > A$
3	$C > D > A > B$
4	$B > A > C > D$

Candidate	Preferences
A	$2 > 3 > 1 > 4$
B	$3 > 1 > 2 > 4$
C	$1 > 2 > 4 > 3$
D	$2 > 4 > 3 > 1$

For each of the matchings in (a) and (b) below, indicate whether the matching is stable or not stable. If it is stable, indicate whether it is job-optimal or candidate-optimal. If it is not stable, identify a rogue couple.

Shade the appropriate circle to indicate your answer; if applicable, write the rogue couple in the box given. No explanation needed.

(a) $\{(1, B), (2, C), (3, A), (4, D)\}$. *3pts*

● Stable

○ Job-Optimal

● Candidate-Optimal

The matching is stable. To see this, iterate over the jobs and only the candidates of higher preferences than the one the job ends up with. We can see that none of these "higher-preference" candidates form a rogue couple with the job since they prefer who they are currently with.

Moreover, the matching is candidate-optimal because it's the result we get from running the propose-and-reject algorithm with the candidates proposing to the jobs with the preferences above.

(b) $\{(1, A), (2, B), (3, C), (4, D)\}$. *3pts*

● Not Stable

Rogue Couple: $(4, C)$ form a rogue couple, as each of them prefers the other to D and 3 respectively.

5. [4 points] Suppose there is a bag containing two fair four-sided dice, one with the numbers $1, 2, 3, 4$ on its faces and the other with the numbers $1, 2, 2, 4$ on its faces. A die is randomly drawn from the bag and rolled twice.

(a) Given that the first roll is a 3, what is the probability that the second roll is a 2? *2pts*

ANS: $\frac{1}{4}$. If the first roll is a 3, then we must have chosen the first die.

(b) Given that the first roll is a 2, what is the probability that the second roll is a 2? *2pts*

ANS: $\frac{5}{12}$. For $i = 1, 2$, let A_i denote the event that the ith roll is a 2, and D_i the event that the ith die was chosen. Then

$$\mathbb{P}[A_2|A_1] = \frac{\mathbb{P}[A_1 \cap A_2]}{\mathbb{P}[A_1]} = \frac{\mathbb{P}[A_1 \cap A_2|D_1]\,\mathbb{P}[D_1] + \mathbb{P}[A_1 \cap A_2|D_2]\,\mathbb{P}[D_2]}{\mathbb{P}[A_1|D_1]\,\mathbb{P}[D_1] + \mathbb{P}[A_1|D_2]\,\mathbb{P}[D_2]} = \frac{\frac{1}{2}(\frac{1}{16} + \frac{1}{4})}{\frac{1}{2}(\frac{1}{4} + \frac{1}{2})} = \frac{5}{12}.$$

6. [6 points] Let P and Q be polynomials of degree at most d, whose coefficients are chosen uniformly and independently at random from $GF(p)$ for a prime $p > d$.

(a) What is $\mathbb{P}[P = Q]$, where $P = Q$ denotes the fact that $P(x) \equiv Q(x) \pmod p$ for all $x \in GF(p)$? *2pts*

ANS: $\frac{1}{p^{d+1}}$. For any fixed P, there is a unique polynomial that agrees with P at all points $x \in GF(p)$. We are choosing Q uniformly among all p^{d+1} polynomials of degree at most d, so the probability of this event is $\frac{1}{p^{d+1}}$. Since this holds for each fixed P, it also holds when P itself is chosen randomly.

(b) What is $\mathbb{P}[P(0) \equiv Q(0) \pmod p]$? *2pts*

ANS: $\frac{1}{p}$. The event $P(0) = Q(0)$ holds if and only if the last (constant) coefficients of P, Q agree. This happens with probability $\frac{1}{p}$.

(c) Given an integer $k \in \{0, \ldots, d-1\}$, what is $\mathbb{P}[P(i) \equiv Q(i) \pmod p \text{ for all } i \in \{0, \ldots, k\}]$? *2pts*

ANS: $\frac{1}{p^{k+1}}$. Fix P. This also fixes the value of Q at the $k+1$ points $i \in \{0, \ldots, k\}$. We require $d+1$ points to uniquely specify Q, so the number of remaining possibilities for Q is exactly $p^{d+1-(k+1)} = p^{d-k}$. Hence the probability of the given event is $\frac{\#\text{ valid choices for } Q}{\#\text{ total choices for } Q} = \frac{p^{d-k}}{p^{d+1}} = \frac{1}{p^{k+1}}$.

[exam continues on next page]

7. [4 points] You are building a cloud computing startup and you have N compute nodes. The failure time for a single node is given by an exponential random variable with parameter λ and is independent of other nodes. You choose to distribute these nodes across ℓ locations, so that each location gets the same number of nodes $k = N/\ell$ (assume that ℓ divides N). A location fails if *some node* at that location fails. The whole system fails if *all locations* fail. *4pts*

What is the probability that the system fails before time t?

ANS: $(1 - \exp(-k\lambda t))^{\ell}$. The time to failure at each location is given by an exponential r.v. with parameter $k\lambda$. Hence the probability that any given location fails by time t is $1 - \exp(-k\lambda t)$. Since the locations are independent, the probability that all ℓ locations fail by time t is $(1 - \exp(-k\lambda t))^{\ell}$.

8. [8 points] The continuous random variable X has pdf

$$f(x) = \begin{cases} \alpha x & \text{for } 0 \le x \le 1; \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine the value of α. *2pts*

ANS: 2. Since f is a pdf, we must have $\int_{-\infty}^{+\infty} f(x)dx = 1$. But $\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 \alpha x dx = \frac{\alpha}{2}x^2\Big|_0^1 = \frac{\alpha}{2}$. Therefore, $\alpha = 2$.

(b) Compute $\mathbb{P}\left[X \le \frac{1}{2}\right]$. **[Note: Write your answer in terms of α.]** *2pts*

ANS: $\frac{\alpha}{8} = \frac{1}{4}$. We have $\mathbb{P}\left[X \le \frac{1}{2}\right] = \int_{-\infty}^{1/2} f(x)dx = \int_0^{1/2} \alpha x dx = \frac{\alpha}{2}x^2\Big|_0^{1/2} = \frac{\alpha}{8}$.

(c) Compute $\mathbb{E}[X]$. **[Note: Write your answer in terms of α.]** *2pts*

ANS: $\frac{\alpha}{3} = \frac{2}{3}$. We have $\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 \alpha x^2 dx = \frac{\alpha}{3}x^3\Big|_0^1 = \frac{\alpha}{3}$.

(d) Compute $\text{Var}(X)$. **[Note: Write your answer in terms of α and μ, where $\mu = \mathbb{E}[X]$.]** *2pts*

ANS: $\frac{\alpha}{4} - \mu^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$. We have $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mu^2$. And $\mathbb{E}\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^1 \alpha x^3 dx = \frac{\alpha}{4}x^4\Big|_0^1 = \frac{\alpha}{4}$.

9. [8 points] Let X, Y, Z be random variables with the following properties:

- X, Y are independent and Y, Z are independent
- $\mathbb{E}[X] = 0$; $\mathbb{E}[Y] = 1$; $\mathbb{E}[Z] = 2$.
- $\text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = 10$.

For each of the following statements, determine if it **must** be true, **must** be false, or is undetermined (i.e., may be either true or false). Shade one bubble for each statement.

TRUE FALSE UNDET.

○ ○ ● X, Z are independent. *1pt*

Clearly X, Z could be independent, but they don't need to be: e.g., it could be the case that $X = Z$.

● ○ ○ $\mathbb{E}[X + 2Y - Z] = 0$. *1pt*

By linearity of expectation, $\mathbb{E}[X + 2Y - Z] = \mathbb{E}[X] + 2\mathbb{E}[Y] - \mathbb{E}[Z] = 0 + 2 - 2 = 0$.

○ ○ ● $\mathbb{P}[Y \geq 10] \leq \frac{1}{10}$. *1pt*

This conclusion would follow from Markov's inequality if we knew that Y is a non-negative r.v., but we are not given that fact.

○ ● ○ $\mathbb{E}[XY] = 2$. *1pt*

Since X, Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$.

● ○ ○ $\text{Var}(X - Y) = 20$. *1pt*

Since X, Y are independent, $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = 20$.

○ ● ○ $\text{Var}(2X) = 20$. *1pt*

$\text{Var}(2X) = 4\text{Var}(X) = 40$.

● ○ ○ $\mathbb{P}[|X| \geq 5] \leq \frac{2}{5}$. *1pt*

By Chebyshev, $\mathbb{P}[|X| \geq 5] = \mathbb{P}[|X - \mathbb{E}[X]| \geq 5] \leq \frac{\text{Var}(X)}{25} = \frac{10}{25} = \frac{2}{5}$.

● ○ ○ $\mathbb{P}\left[|X| \geq \sqrt{10}\right] > 0$. *1pt*

Suppose for contradiction that $|X| < \sqrt{10}$ with probability 1. Then we would have $\text{Var}(X) = \mathbb{E}\left[X^2\right] < 10$, which contradicts $\text{Var}(X) = 10$.

10. [2 points] Let $X_1, \ldots, X_n$ be i.i.d. random variables that are uniform over the set of values $\{1, 2, 3\}$, and let $S_n = X_1 + \cdots X_n$. Determine parameters a and b such that the random variable $\frac{S_n - a}{b}$ converges to the standard normal distribution $\mathcal{N}(0, 1)$ as $n \to \infty$. *2pts*

ANS: $a = 2n$, $b = \sqrt{\frac{2}{3}n}$. By the Central Limit Theorem, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to $\mathcal{N}(0, 1)$ as $n \to \infty$, where $\mu = \mathbb{E}[X_i]$ and $\sigma = \sqrt{\text{Var}(X_i)}$. Here by direct calculation $\mu = 2$ and $\sigma = \sqrt{\frac{2}{3}}$.

[exam continues on next page]

11. Musical Counting [18 points].

In preparation for the CS70 musical, we want to understand the various ways we can arrange and combine musical notes. Throughout this problem, we will use the following terminology and assumptions. (Note that this may deviate from standard musical practice: you should disregard any prior musical knowledge you have for this problem!)

- A *piano* has 88 keys in total, 52 of which are white and the other 36 black.
- A *chord* is any set of k distinct keys played simultaneously (i.e., order does not matter), where $2 \leq k \leq 10$.
- A *melody* is a sequence of any number of individual keys played one after the other (here order matters, and keys can repeat unless stated otherwise).

(a) How many chords of $k = 5$ keys containing exactly 2 white keys and 3 black keys are there? *3pts*

ANS: $\binom{52}{2}\binom{36}{3}$. There are $\binom{52}{2}$ ways to choose the white keys, and $\binom{36}{3}$ ways to choose the black keys.

(b) Given a fixed set of 10 distinct keys, how many chords can be played from these keys? *3pts*

ANS: $2^{10} - 11$. A chord is any subset of the 10 keys (2^{10} choices), except for the empty subset and the subsets of size 1 ($1 + 10 = 11$ choices).

(c) How many melodies consisting of exactly 12 keys (repetitions allowed) are there? *3pts*

ANS: 88^{12}. This is just the number of ways of choosing an ordered sequence of 12 items out of 88, with replacement.

(d) How many melodies are there that consist of 40 white keys and 10 black keys (with repetitions allowed), such that no two black keys are played one after the other (i.e., there has to be at least one white key between any two black keys)? *3pts*

ANS: $52^{40} \times 36^{10} \times \binom{41}{10}$. The first two factors choose the 40 white keys and the 10 black keys. Note that these two sets of keys are ordered, so to finish we just need to multiply by the number of ways of interleaving the 10 black keys into the gaps between white keys. There are 41 such gaps (including the first and last positions), 10 of which have to contain one black key; this gives the final factor of $\binom{41}{10}$.

(e) How many ways are there to choose 200 keys? Each of the 88 piano keys can be used any number of times (or not at all), and the order does not matter. *3pts*

ANS: $\binom{287}{87}$. Here we are choosing 200 unordered items with replacement from a set of size 88. This is equivalent to a stars-and-bars problem with 200 stars and 87 bars (the positions of the bars tell us how many copies of each key we are choosing), giving the desired answer.

(f) Suppose now that you are given a melody of 200 *ordered* keys. In how many ways can this melody be split into exactly 4 consecutive parts, such that each part contains at least 20 keys? *3pts*

ANS: $\binom{123}{3}$. Think of the ordered sequence of 200 keys. We need to place 3 "dividers" into the 199 "gaps" between them. However, we can't choose any set of 3 gaps: because each part has to contain at least 20 keys, there have to be at least 19 gaps before the first chosen gap, another 19 between the first and second chosen gaps, etc. This constraint rules out $4 \times 19 = 76$ gaps, leaving a total of $199 - 76 = 123$ allowed gaps. Thus the final number is $\binom{123}{3}$. [To see this argument more formally, consider any particular choice (i, j, k) of three gaps from the final set of 123. This corresponds to the following choice of gaps from the full set of 199 gaps: $(i + 19, j + 38, k + 57)$. This is a bijection between the two sets of gaps.]

12. Isolated Vertices on a Cycle Network [16 points]

Suppose we have a network of $n \geq 3$ processors connected in the form of a single cycle (see the left-hand figure below for the example $n = 5$). Each connection (edge) in the network fails and is removed independently with probability $\frac{1}{2}$. We call a processor (vertex) *isolated* if it has no remaining edges incident to it. (In the $n = 5$ example in the right-hand figure below, three edges have failed and only vertex 2 is isolated.)

(a) For any vertex i, what is $\mathbb{P}[i \text{ is isolated}]$? *2pts*

ANS: $\frac{1}{4}$. Vertex i is isolated if and only if its two incident edges fail.

(b) Let X denote the number of isolated vertices. Compute the expectation $\mathbb{E}[X]$. *2pts*

ANS: $\frac{n}{4}$. Write $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator r.v. for the event that vertex i is isolated. Then $\mathbb{E}[X_i] = \mathbb{P}[i \text{ is isolated}] = \frac{1}{4}$, and by linearity of expectation $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{4}$.

(c) For two *adjacent* vertices i, j in the cycle, what is $\mathbb{P}[i \text{ and } j \text{ are both isolated}]$? *2pts*

ANS: $\frac{1}{8}$. There are three edges incident on one or both of i, j; and i, j are both isolated if and only if all these three edges fail.

(d) For two *non-adjacent* vertices i, j, what is $\mathbb{P}[i \text{ and } j \text{ are both isolated}]$? *2pts*

ANS: $\frac{1}{16}$. In this case there are four edges incident on either of i, j; and i, j are both isolated if and only if all these four edges fail. Alternatively we can note that, when i, j are not adjacent, X_i, X_j are independent, so $\mathbb{P}[X_i = X_j = 1] = \mathbb{P}[X_i = 1]\,\mathbb{P}[X_j = 1] = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$.

(e) Compute $\text{Var}(X)$. [**Note**: You will need to use your answers to parts (b), (c) and (d). Indicate clearly where you use these answers!] *4pts*

ANS: $\frac{5}{16}n$. $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. To compute $\mathbb{E}[X^2]$, we write $X = \sum_{i=1}^{n} X_i$ as in part (b) and get

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \mathbb{E}\Big[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j, i \sim j} X_i X_j + \sum_{i \neq j, i \not\sim j} X_i X_j\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] + \sum_{i \neq j, i \sim j} \mathbb{E}[X_i X_j] + \sum_{i \neq j, i \not\sim j} \mathbb{E}[X_i X_j] \\
&= \mathbb{E}[X] + (2n \times \frac{1}{8}) + (n(n-1) - 2n) \times \frac{1}{16} = \frac{5}{16}n + \frac{n^2}{16}.
\end{aligned}$$

In the first line, we just expanded the square $(X_1 + \cdots + X_n)^2$, splitting the cross terms $X_i X_j$ into those with $i \sim j$ (i adjacent to j) and the rest; in the second line we used linearity of expectation; in the third line we used parts (c) and (d) respectively to obtain that $\mathbb{E}[X_i X_j] = \frac{1}{8}$ when $i \sim j$ and

[exam continues on next page]

$\mathbb{E}[X_iX_j] = \frac{1}{16}$ when $i \not\sim j$; we also observed that there are $2n$ ordered pairs i, j with $i \sim j$, and $n(n-1)$ ordered pairs i, j with $i \neq j$ in total.

Finally, subtracting $\mathbb{E}[X]^2 = \frac{n^2}{16}$, as obtained in part (b), gives the answer $\frac{5}{16}n$.

An alternative approach to the above calculation is to write

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \\ &= \sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{i\neq j, i\sim j}\mathrm{Cov}(X_i, X_j) + \sum_{i\neq j, i\not\sim j}\mathrm{Cov}(X_i, X_j) \\ &= (n \times \frac{3}{16}) + (2n \times \frac{1}{16}) + 0 = \frac{5}{16}n,\end{aligned}$$

where we've used the fact that $\mathrm{Var}(Y+Z) = \mathrm{Var}(Y) + \mathrm{Var}(Z) + \mathrm{Cov}(Y, Z)$ for any r.v.'s Y, Z, and calculated $\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$ and $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = \frac{1}{8} - \frac{1}{16} = \frac{1}{16}$ when $i \sim j$, and also $\mathrm{Cov}(X_i, X_j) = 0$ when $i \not\sim j$ (since in this case X_i, X_j are independent).

(f) Use Chebyshev's inequality to get an upper bound on the probability $\mathbb{P}\left[X \geq \frac{n}{2}\right]$ that there are at least $\frac{n}{2}$ isolated vertices. [**Note:** You will need to use your answers to parts (b) and (e). Indicate clearly where you use these answers!] *4pts*

ANS: $\frac{5}{n}$. Since $\mathbb{E}[X] = \frac{n}{4}$, we have by Chebyshev's inequality that

$$\mathbb{P}\left[X \geq \frac{n}{2}\right] \leq \mathbb{P}\left[|X - \mathbb{E}[X]| \geq \frac{n}{4}\right] \leq \frac{\mathrm{Var}(X)}{(n/4)^2} = \frac{(5/16)n}{n^2/16} = \frac{5}{n}.$$

13. Dolphin Watching [16 points]

You are going on a boat tour to observe dolphins off the Northern California coast. You are given the following information about the dolphins' behavior:

- On average, a dolphin will approach your boat every 20 minutes.
- When a dolphin approaches your boat, it will swim along with you. For each minute until it leaves, it will perform a trick with probability p, otherwise do nothing.
- At the end of each minute, a dolphin at the boat will leave with probability q, otherwise it will stay for at least another minute.
- All dolphin arrivals and all dolphin behaviors are mutually independent.

Let's do some dolphin stats. [**Note: Your answers may include binomial coefficients and fractions, but should not include any summations.**]

(a) Let the r.v. N denote the number of dolphins that approach your boat **per hour**. What is the distribution of N? (Give its name and parameter.) *2pts*

ANS: $N \sim \text{Poisson}(3)$.

(b) As the i-th dolphin approaches your boat, let X_i denote the number of minutes the dolphin will swim alongside your boat. What is the distribution of X_i? (Give its name and parameter.) *2pts*

ANS: $X_i \sim \text{Geometric}(q)$.

(c) What is the probability that dolphin i will swim alongside your boat for exactly 5 minutes? *2pts*

ANS: $(1-q)^4 q$. The probability that a $\text{Geometric}(q)$ r.v. is equal to 5.

(d) How long will a dolphin swim alongside your boat on average? *2pts*

ANS: $1/q$ minutes. This is the expectation of $\text{Geometric}(q)$.

(e) Let Z_i denote the number of tricks that dolphin i will perform while swimming alongside your boat. Compute the probability that a dolphin will perform exactly k tricks, given that it swims alongside for exactly 20 minutes. *2pts*

ANS: $\binom{20}{k} p^k (1-p)^{20-k}$. This is because, given dolphin i swims alongside for 20 minutes, $Z_i \sim \text{Binomial}(20, p)$.

(f) Given that a dolphin swims alongside your boat for 20 minutes, what is the expected number of tricks it performs? *2pts*

ANS: $20p$. The expectation of a $\text{Binomial}(20, p)$ r.v.

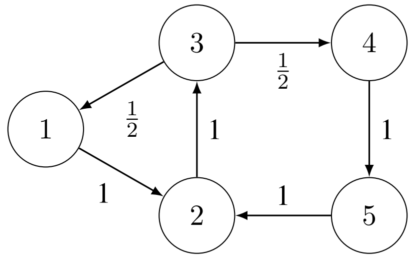
(g) Compute $\mathbb{E}[Z_i]$. *4pts*

ANS: p/q. By the Law of Total Probability (applied to expectations), we have

$$\begin{aligned}
\mathbb{E}[Z_i] &= \sum_{x=1}^{\infty} \mathbb{E}[Z_i \mid X_i = x]\,\mathbb{P}[X_i = x] \\
&= \sum_{x=1}^{\infty} xp\mathbb{P}[X_i = x] \\
&= p\sum_{x=1}^{\infty} x\mathbb{P}[X_i = x] \\
&= p\mathbb{E}[X_i] = \frac{p}{q}.
\end{aligned}$$

14. Five-State Markov Chain [12 points].

Consider the following Markov chain with states $1, \ldots, 5$.



(a) Is this chain irreducible? Briefly explain your answer. *2pts*

ANS: Yes: every state is reachable from every other state by a path of transitions. To check this, we can easily see (e.g.) that we can get to state 2 from every other state, and from every other state to state 2.

(b) Is this chain aperiodic? Briefly explain your answer. *2pts*

ANS: Yes: the gcd of all path lengths from any state to any other state is 1. To see this, note that there are paths of length 3 and 4 from state 2 to itself. Now for any pair of states i, j, take a path from i to j via 2; say this path has length ℓ. By inserting either the path of length 3 or the path of length 4 from 2 back to 2 into this path, we get paths of lengths $\ell + 3$ and $\ell + 4$ from i to j. Hence the gcd of these path lengths is 1.

(c) Let π denote the invariant distribution of the chain. Given that $\pi(1) = \pi(5) = \frac{1}{7}$, compute $\pi(2)$. *3pts*

ANS: $\frac{2}{7}$. We note from the balance equations that $\pi(2) = \pi(1) + \pi(5) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}$.

(d) For $i = 1, \ldots, 5$, let $\beta(i)$ denote the expected number of steps to reach state 1 starting from state i. Given that $\beta(3) = 5$, compute $\beta(5)$. *3pts*

ANS: 7. From the first-step equations, we have $\beta(5) = 1 + \beta(2)$ and $\beta(2) = 1 + \beta(3)$. Therefore, $\beta(5) = 2 + \beta(3) = 7$.

(e) Suppose the transition from state 4 to state 5 is removed and replaced by a self-loop with probability 1 on state 4. What is now the expected number of steps to reach state 1 starting from state 5? *2pts*

ANS: infinite. Note that state 4 is now an absorbing state, so any path that reaches it will never reach state 1. (Formally, we have that $\beta(4) = \infty$.) Starting from state 5, there is a non-zero probability that we reach state 4 (e.g., we get there in 3 steps with probability $\frac{1}{2}$). Hence the expected time to reach state 1 must be infinite.