CS70 Discrete Mathematics and Probability Theory, Fall 2019

Midterm 2

8:00-10:00pm, 14 November

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* **5 double-sided** *sheets (10 numbered pages) on the exam. Notify a proctor immediately if a sheet 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 it may help to leave it and return to it later.*

(e) *On questions 1-2, you need only give the answer in the format requested (e.g., True/False, an expression, a statement.) An expression may simply be a number or an expression with a relevant variable in it. For short answer questions, correct, clearly identified answers will receive full credit with no justification. Incorrect answers may receive partial credit.*

(f) *On questions 3-6, you should give arguments, proofs or clear descriptions if requested. If there is a box you must use it for your answer: answers written outside the box may not be graded!*

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

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

(i) *You have 120 minutes: there are 6 questions on this exam worth a total of 140 points.*

[exam starts on next page]

1. **True/False** [*No justification; answer by shading the correct bubble. 2 points per answer; total of 28 points. No penalty for incorrect answers.*]

   (a) For each of the following sets, indicate whether it is finite, countably infinite (CTBL INF) or uncountable (UNC) by shading the appropriate bubble.

FINITE	**CTBL INF**	**UNC**		
○	●	○	The set $\{\frac{2^i}{3^j} \mid i, j \in \mathbb{N}\}$.	*2pts*
○	○	●	The set of all subsets of the prime numbers.	*2pts*
○	●	○	The set of all finite sets of rational numbers.	*2pts*
○	○	●	The set of real numbers in the interval $[0, 1]$ whose decimal expansion involves only the digits 3 and 5.	*2pts*
○	●	○	The set of finite arithmetic expressions over the integers, using symbols $+, -, \times, \div, (,)$.	*2pts*
●	○	○	The set of all movies with file size at most 4 GB.	*2pts*

   (b) Indicate which of the following statements is **TRUE** or **FALSE** by shading the appropriate bubble.

TRUE	**FALSE**		
●	○	Every infinite set has a countably infinite subset.	*2pts*
●	○	There are uncountably many real polynomials of degree 2 that pass through the points $(0, 0)$ and $(1, 1)$.	*2pts*
●	○	Let A and B be finite sets. If $f : A \to B$ is a surjection such that $\lvert f^{-1}(\{b\})\rvert \leq m$ for all $b \in B$, where f^{-1} denotes the preimage of f, then $\lvert A\rvert \geq \lvert B\rvert \geq \lvert A\rvert/m$.	*2pts*
○	●	For events A and B on the same probability space such that $A \subseteq B$, it is possible to have $\mathbb{P}[A] > \mathbb{P}[B]$.	*2pts*
○	●	For any event B on the probability space $(\Omega, \mathbb{P})$, $\mathbb{P}[B] = \sum_{k=1}^{n} \mathbb{P}[B \cap A_k]$ if $A_1, A_2, \ldots, A_n$ are any subsets of the same sample space Ω satisfying $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$.	*2pts*
○	●	For events A, B on the same probability space, if $A \cap B = \varnothing$, then A and B are independent.	*2pts*
●	○	Let X and Y be Bernoulli random variables. If $\mathbb{P}[X = 0, Y = 1] = \mathbb{P}[X = 0]\,\mathbb{P}[Y = 1]$, then X and Y are independent.	*2pts*
●	○	For all random variables X and Y defined on the same probability space, $\mathbb{E}[XY] \leq \frac{1}{2}(\mathbb{E}[X^2] + \mathbb{E}[Y^2])$.	*2pts*

[exam continued on next page]

2. Short Answers [*Answer is a single number or expression; write it in the box provided; no justification necessary. 3 points per answer; total of 54 points. No penalty for incorrect answers.*]

(a) Two real polynomials $P(x), Q(x)$ have degrees exactly 3 and 2 respectively. What is the maximum number of points x for which $P(x) = Q(x)$? *3pts*

3. [$P(x) = Q(x)$ if and only if $P(x) - Q(x) = 0$. But $P(x) - Q(x)$ is a polynomial of degree exactly 3, and so can have at most 3 zeros. $P(x) = x^3 + 2x$ and $Q(x) = 3x^2$ is an example where $P(x) - Q(x)$ has exactly three roots.]

(b) Two real polynomials $P(x), Q(x)$ have exactly 2 and 3 zeros respectively. What is the maximum possible number of zeros of $P(x)Q(x)$? *3pts*

5. [If x is a root of either $P(x)$ or $Q(x)$, then it must also be a root of $P(x)Q(x)$. Furthermore, if x is a root of neither $P(x)$ nor $Q(x)$, then it cannot be a root of $P(x)Q(x)$. If the roots of $P(x)$ and $Q(x)$ are all distinct, then $P(x)Q(x)$ has 5 roots.]

(c) A polynomial $P(x)$ of degree 2 over $GF(11)$ has zeros at $x = 0$ and $x = 2$, and passes through the point $(1, 1)$. What is P? [All coefficients should be integers in $\{0, \dots, 10\}$.] *3pts*

$P(x) = 10x^2 + 2x + 0$. [Since $P(x)$ is quadratic with roots $x = 0$ and $x = 2$, it must be of the form $P(x) \equiv c \cdot x(x - 2) \pmod{11}$, where the constant c is determined by $1 \equiv P(1) \equiv c \cdot 1 \cdot (1 - 2) \pmod{11}$. Namely, $c \equiv -1 \pmod{11}$. That is, $P(x) \equiv -x(x - 2) \equiv -x^2 + 2x \equiv 10x^2 + 2x \pmod{11}$.]

(d) Alice uses the Berlekamp-Welch encoding scheme to send Bob a message consisting of three characters, each an integer mod 11, using a degree-2 polynomial $P(x)$ over $GF(11)$ where the message characters are encoded as $P(1), P(2), P(3)$. It is known that up to two characters may be corrupted during transmission. Answer the following questions.

(i) How many points of P in total must be sent to ensure that the message can be reconstructed? *3pts*

7. [In order to send a message of length $m = 3$ while guarding against $e = 2$ errors, we need a total of $m + 2e = 7$ packets.]

(ii) Following the Berlekamp-Welch scheme, Bob computes the error-locator polynomial as $E(x) = x^2 + 7x + 3$. Which of the characters, $P(1), P(2), P(3)$, were corrupted? *3pts*

$P(1)$ and $P(3)$. [The locations of the corrupted packets are given by the roots of $E(x) \pmod{11}$. $E(x) \equiv (x - 1)(x - 3) \pmod{11}$, so the roots are $x = 1$ and $x = 3$.]

(iii) Bob also computes the polynomial $Q(x) = x^4 + 7x^3 + 4x^2 + 7x + 3$ (with notation as in the class notes). What was the original message sent by Alice? *3pts*

$2, 5, 10$. [$P(x) = Q(x)/E(x) = x^2 + 1$, and $P(1) = 2, P(2) = 5, P(3) = 10$.]

[Q2 continued on next page]

PLEASE NOTE: In the remaining short problems, express your answers in terms of binomial coefficients and factorials whenever possible.

(e) How many distinct 6-digit numbers can you create by rearranging the numbers $\{1, 3, 5, 7, 7, 9\}$ such that no two consecutive numbers are identical? *3pts*

$\frac{6!}{2} - 5! = 2 \cdot 5! = 240$. [There are $\frac{6!}{2}$ ways to rearrange the six numbers (accounting for the fact that the two 7's are indistinguishable), and in 5! of those, the two 7's are located next to each other (treating them as one unit gives us 5 units to permute).]

(f) Suppose you hold 100 stocks for each of five utility companies, including PG&E. You wish to sell a total of 100 stocks while ensuring you sell at least 50 PG&E stocks. In how many different ways can you do this? (Stocks in a given company are indistinguishable.) *3pts*

$\binom{54}{4}$. [This is a balls-and-bins problem (a.k.a. stars-and-bars problem), where we distribute $k = 100$ balls (stocks) over $n = 5$ bins (utility companies), making sure that bin 1 (PG&E) has at least 50 balls. Placing any of the 50 (indistinguishable) balls into bin 1 leaves us with $k - 50 = 50$ balls to distribute over $n = 5$ bins, for a total of $\binom{k-50+n-1}{n-1} = \binom{54}{4}$ possibilities.]

(g) For $k \le n$, how many functions from $\{1, \ldots, k\}$ to $\{1, \ldots, n\}$ are **not** injective? *3pts*

$n^k - \frac{n!}{(n-k)!}$. [As discussed in lecture, the total number of functions f from $K = \{1, \ldots, k\}$ to $N = \{1, \ldots, n\}$ is n^k (each element in K can be assigned to any element in N), while the number of injective functions from K to N is $\frac{n!}{(n-k)!}$ (which corresponds to the number of ways to choose k ordered elements from N without replacement). Subtracting the latter from the former gives the answer.]

(h) What should x be to make the following a valid combinatorial identity? $\sum_{m=k}^{n-k-1} \binom{m}{k}\binom{n-m-1}{k} = \binom{n}{x}$. *3pts*

$2k + 1$. [The RHS counts the number of ways to choose a subset of size $2k + 1$ from $\{1, \ldots, n\}$. The LHS does the same by summing over the possibilities of what the $(k + 1)^{\text{st}}$-smallest element in this subset can be: let $m+1$ be the $(k+1)^{\text{st}}$-smallest element, and choose k elements from $\{1, \ldots, m\}$ and k elements from $\{m + 2, \ldots, n\}$.]

(i) Consider a biased coin which shows heads with probability $2/3$. Suppose Alice tosses the coin n times and then Bob also tosses it n times. You may assume independence between trials.

(i) For $j, k \in \{0, 1, \ldots, n\}$, let A_j denote the event that Alice gets heads j times, and B_k the event that Bob gets heads k times. Find $\mathbb{P}[A_j \cap B_k]$. *3pts*

$\binom{n}{j}\left(\frac{2}{3}\right)^j\left(\frac{1}{3}\right)^{n-j} \cdot \binom{n}{k}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n-k} = \binom{n}{j}\binom{n}{k}\frac{2^{j+k}}{3^{2n}}$. [$A_j$ and B_k are the events that two independent binomial random variables assume values j and k, respectively. That is, if $X \sim \text{Bin}\left(n, \frac{2}{3}\right)$ and $Y \sim \text{Bin}\left(n, \frac{2}{3}\right)$ are independent, then $\mathbb{P}[A_j \cap B_k] = \mathbb{P}[X = j, Y = k] = \mathbb{P}[X = j] \cdot \mathbb{P}[Y = k]$.]

(ii) For $m \in \{0, \ldots, 2n\}$, let S_m denote the event that the sum of the number of heads Alice gets and the number of heads Bob gets is equal to m. Find $\mathbb{P}[S_m]$. (**NOTE**: Your answer should not involve any summation signs.) *3pts*

$\binom{2n}{m}\left(\frac{2}{3}\right)^m\left(\frac{1}{3}\right)^{2n-m} = \binom{2n}{m}\frac{2^m}{3^{2n}}$. [We are throwing the same coin $2n$ times independently, and are interested in the probability of observing m heads. This is $\mathbb{P}[Z = m]$, where $Z \sim \text{Bin}\left(2n, \frac{2}{3}\right)$.]

[Q2 continued on next page]

(j) In the U.S., suppose the proportion of people who have diabetes is $d > 0$, and a fraction $m > 0$ of those people have a particular risk mutation. Also, among those who do **not** have diabetes, a fraction $a > 0$ do **not** have the risk mutation. Suppose a person is selected uniformly at random from the population. Given that the person has the risk mutation, what is the probability that she/he has diabetes? Express your answer in terms of d, m, and a. *3pts*

$\frac{dm}{dm+(1-a)(1-d)}$. [If D and M are the events of having diabetes and carrying the risk mutation, respectively, then by Bayes' rule we have $\mathbb{P}[D \mid M] = \frac{\mathbb{P}[M|D]\mathbb{P}[D]}{\mathbb{P}[M|D]\mathbb{P}[D]+\mathbb{P}[M|\overline{D}]\mathbb{P}[\overline{D}]}$. We are given $\mathbb{P}[D] = d$, $\mathbb{P}[M \mid D] = m$, and $\mathbb{P}[\overline{M} \mid \overline{D}] = a$, so $\mathbb{P}[\overline{D}] = 1 - d$ and $\mathbb{P}[M \mid \overline{D}] = 1 - a$. Plugging in these probabilities into the above formula for $\mathbb{P}[D \mid M]$ yields the answer.]

(k) In a given day, a computer manufacturer produced L laptops each containing 4 memory chips. It turns out that, of the $4L$ chips inserted into the laptops, exactly d chips were defective. Assume that the memory chips were well shuffled before they went into the laptops, and let X denote the number of defective memory chips in a randomly selected laptop. For $k \in \{0, 1, 2, 3, 4\}$, what is $\mathbb{P}[X = k]$?

$\frac{\binom{d}{k}\binom{4L-d}{4-k}}{\binom{4L}{4}}$ or equivalently $\frac{\binom{4}{k}\binom{4L-4}{d-k}}{\binom{4L}{d}}$. [This is the distribution of a Hypergeometric $(4L, d, 4)$, or equivalently Hypergeometric $(4L, 4, d)$, random variable: Sampling 4 chips without replacement from d defective and $4L - d$ functional chips for any given laptop is the same as drawing 4 balls from a bag of d black and $4L - d$ white balls. Alternatively, sampling without replacement d chips to be defective from 4 chips of the laptop of interest and the $4L - 4$ chips contained in the other laptops is the same as sampling d balls from a bag of 4 black and $4L - 4$ white balls.]

(l) Suppose a single card is dealt from a well-shuffled standard deck of 52 cards. What is the probability that the card is either a red suit or an ace? *3pts*

$\frac{28}{52} = \frac{7}{13}$. [The event in question contains a total of 28 cards: The 26 red suits and ♠A and ♣A. Alternatively, $\mathbb{P}[\text{Ace or Red}] = \mathbb{P}[\text{Ace}] + \mathbb{P}[\text{Red}] - \mathbb{P}[\text{Ace and Red}] = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13}$.]

(m) Let X be a random variable with $\mathbb{P}[X = -4] = \frac{1}{4}, \mathbb{P}[X = 0] = \frac{1}{4}$, and $\mathbb{P}[X = 4] = \frac{1}{2}$. Find $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ and $\text{Var}[X]$. *3pts*

$$\mathbb{E}[X] = -4 \cdot \tfrac{1}{4} + 0 \cdot \tfrac{1}{4} + 4 \cdot \tfrac{1}{2} = 1, \qquad \mathbb{E}[X^2] = (-4)^2 \cdot \tfrac{1}{4} + 0 \cdot \tfrac{1}{4} + 4^2 \cdot \tfrac{1}{2} = 12,$$

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 11.$$

(n) Let X and Y be independent random variables with $\text{Var}[X] = 2$ and $\text{Var}[Y] = 3$. Find $\text{Var}[3X - Y]$. *3pts*

21. [Since $\text{Var}[aZ] = a^2\text{Var}[Z]$ for any constant a and random variable Z, and $\text{Var}[Z_1 + Z_2] = \text{Var}[Z_1] + \text{Var}[Z_2]$ if Z_1 and Z_2 are independent random variables, we have $\text{Var}[3X - Y] = \text{Var}[3X + (-Y)] = \text{Var}[3X] + \text{Var}[-Y] = 3^2\text{Var}[X] + (-1)^2\text{Var}[Y] = 9 \cdot 2 + 1 \cdot 3 = 21$.]

(o) Consider two events A and B on the same probability space and suppose $A \subset B$. Let I_A and I_B denote indicator random variables for events A and B, respectively. Find $\text{Cov}(I_A, I_B)$ in terms of $\mathbb{P}[A]$ and $\mathbb{P}[B]$. *3pts*

$\mathbb{P}[A](1 - \mathbb{P}[B])$. [$\text{Cov}(I_A, I_B) = \mathbb{E}[I_A I_B] - \mathbb{E}[I_A] \times \mathbb{E}[I_B] = \mathbb{P}[A] - \mathbb{P}[A] \times \mathbb{P}[B]$.]

[exam continued on next page]

3. Homomorphic Secret Sharing [*Total of 13 points.*]

A group of n students in a charity fundraising group wish to compute the total amount of money they have raised without revealing how much each of them has individually raised. To do this, they enlist the help of $k \geq 2$ administrators at the charity. Under the assumption that individual amounts are natural numbers no larger than $1,000, the students want to design a scheme with the following properties:

(i) The k administrators can work together to compute the sum of the raised amounts.

(ii) The administrators cannot discover anything about the individual amounts raised by the students (other than what can be deduced from the sum), unless all k of them conspire to do that.

Since the students have recently taken CS70, they come up with the following secret sharing scheme to solve this problem:

Each student i, for $1 \leq i \leq n$, constructs a polynomial p_i over $GF(q)$, for a suitable prime q, so that $p_i(0) = m_i$, the amount that i has raised. The student then sends the value $p_i(j) \pmod q$ to administrator j, for all $1 \leq j \leq k$.

Answer the following questions about this scheme.

(a) What should the degree of the polynomials p_i be? Write your answer in the box; no justification required. *2pts*

$k-1$. [Each of the k administrators will have one point on the polynomial.]

(b) How can the administrators compute the total amount of money raised, $\sum_{i=1}^{n} m_i$, without computing the individual amounts m_i? Be sure to specify clearly the computations that the administrators perform. *5pts*

The idea is that the administrators will combine their information in order to compute the polynomial $p(x) := \sum_i p_i(x) \pmod q$. Since $p_i(0) = m_i$ for each i, they will then be able to immediately deduce the desired sum from the fact that $p(0) = \sum_i p_i(0) = \sum_i m_i$.

Since the degree of p is the same as that of the p_i, in order to recover p they need its value at k distinct points. For each $1 \leq j \leq k$, the jth administrator can deduce the value $p(j)$ by computing the sum $\sum_i p_i(j)$ of the values received from each of the n students. The administrators then combine the values $p(1), \ldots, p(k)$ using Lagrange interpolation to deduce the entire polynomial p.

NOTE: It is crucial that the administrators do **not** compute the individual polynomials p_i by sharing all the values $p_i(j)$, as this would reveal the individual amounts m_i. They must work only with the sums.

[Q3 continued on next page]

(c) Give a lower bound on the field size q. Write your answer in the box; no justification required. *2pts*

$q \geq \max\{1000n, k+1\}$. [$1000n$ is the maximum possible value of the sum; if q is smaller than this then information will be lost when doing computation mod q. We also need $q \geq k+1$ in order to get k distinct non-zero points on the polynomials. Of course, in practice $1000n$ will be much larger than $k+1$, so we really only need to ensure that $q \geq 1000n$.]

(d) Explain briefly why this scheme is secure in the sense demanded by the students, i.e., why the administrators cannot learn anything about the individual amounts m_i unless all k of them conspire to do so (as specified in property (ii) above). *4pts*

Suppose that the administrators compute the sum polynomial p as described in part (b), and then $k-1$ of them conspire to try to discover the individual amounts m_i of the students. WLOG, we may assume that administrators $1, \ldots, k-1$ are the conspirators. Intuitively, the security of the scheme follows from the fact that, even if they share all their values $p_i(j)$, these conspirators only have $k-1$ points on each of n polynomials p_i, each of degree k, so they have no information about the missing values $p_i(0) = m_i$. But this intuition is not completely convincing because it doesn't rule out the possibility that the conspirators could combine this information with their knowledge of the sum polynomial $p(x)$ to get more information about the m_i.

To give a convincing argument, we will show that every student i can change his/her amount to any value m_i', subject only to the constraint that $\sum_i m_i' = \sum_i m_i$, *without changing the values given to the $k-1$ conspiring administrators or the sum polynomial p*. This will show that the information possessed by those administrators is consistent with *any* such set of amounts for the students, meaning that the administrators have learned nothing other than the sum polynomial p.

To see the above claim, suppose the students recompute their polynomials as $p_i'(x)$, where $p_i'(j) = p_i(j)$ for $1 \leq j \leq k-1$, and $p_i'(0) = m_i'$. Thus p_i' agrees with p_i at all points $x = 1, \ldots, k-1$, but has a different value at 0 (namely, m_i' instead of m_i). These k points define a unique polynomial p_i' of degree $k-1$. Now, as far as the $k-1$ administrators are concerned, all the polynomials p_i' are indistinguishable from the original polynomials p_i; and, moreover, the sum polynomial $p' := \sum_i p_i'$ is the same as the original sum polynomial $p = \sum_i p_i$, since these two degree-$(k-1)$ polynomials agree on the k points $0, 1, \ldots, k-1$. (They may disagree at the point k, but that is not a problem since we are interested only in how things look to the conspiring administrators.)

[exam continued on next page]

4. **Counting** [*Total of 14 points.*]

Alice has n distinct items on her wish list $W = \{1, \ldots, n\}$ and, after getting her first paycheck, she decides to purchase $m \leq n$ items from the list. Suppose items $1, \ldots, k$, where $k < m$ and $k \leq n - m$, are of special interest to her. (**NOTE:** Whenever possible, express your answers in terms of binomial coefficients.)

(a) In how many ways can Alice choose m distinct items from her wish list W? Write your answer in the box; no justification required. *2pts*

$\binom{n}{m}$.

(b) If all of items $1, \ldots, k$ must be included, in how many ways can Alice choose m distinct items from W? Write your answer in the box; no justification required. *3pts*

$\binom{n-k}{m-k}$. [If all of items $1, \ldots, k$ are included in the selection, the remaining $m - k$ items must be chosen from $\{k + 1, \ldots, n\}$, which can be achieved in $\binom{n-k}{m-k}$ different ways.]

(c) If at least two of items $1, \ldots, k$ must be included, in how many ways can Alice choose m distinct items from W? Write your answer in the box; no justification required. *4pts*

Solution 1: $\binom{n}{m} - \binom{n-k}{m} - k\binom{n-k}{m-1}$. [There are $\binom{n-k}{m}$ ways to choose all m items from $\{k + 1, \ldots, n\}$, while there are $k\binom{n-k}{m-1}$ ways to choose one item from $\{1, \ldots, k\}$ and the remaining $m - 1$ items from $\{k + 1, \ldots, n\}$. Subtracting these possibilities from the answer in part (a) yields the answer.]

Solution 2: $\sum_{j=2}^{k} \binom{k}{j}\binom{n-k}{m-j}$. [An m-subset of $\{1, \ldots, n\}$ consisting of at least two items from $\{1, \ldots, k\}$ can be constructed by choosing j items from $\{1, \ldots, k\}$ and $m - j$ items from $\{k + 1, \ldots, n\}$, where $j = 2, \ldots, n$. There are $\binom{k}{j}$ ways to choose j items from $\{1, \ldots, k\}$, and for each of these choices, there are $\binom{n-k}{m-j}$ ways to choose $m - j$ items from $\{k + 1, \ldots, n\}$. The subsets constructed in this way for different values of j cannot be the same, so the answer is obtained by summing $\binom{k}{j}\binom{n-k}{m-j}$ over $j = 2, \ldots, n$.]

(d) Let A and B denote the answers to parts (a) and (b), respectively. If we let I_i denote the event that item i is included when m distinct items are chosen from W, note that $B = |I_1 \cap \cdots \cap I_k|$. Fill in the blank in the following equation to produce a valid combinatorial identity, **and provide a brief justification below that equation**. Your justification should include a mathematical expression in terms of the I_i's for what the right hand side of the equation is counting. *5pts*

$$A - B = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} \binom{n-j}{m}.$$

Let Ω denote the set of all m-subsets of $\{1, 2, \ldots, n\}$. Since $A = |\Omega|$ and $B = |I_1 \cap \cdots \cap I_k|$, we have $A - B = |\overline{I_1 \cap \cdots \cap I_k}| = |\overline{I}_1 \cup \cdots \cup \overline{I}_k|$, where the second equality follows from De Morgan's Laws. Let S be a j-subset of $\{1, \ldots, k\}$ where $1 \leq j \leq k$. Then, $|\bigcap_{i \in S} \overline{I}_i| = \binom{n-j}{m}$. Furthermore, there are $\binom{k}{j}$ distinct j-subsets of $\{1, \ldots, k\}$. Hence, the inclusion-exclusion formula yields

$$|\overline{I}_1 \cup \cdots \cup \overline{I}_k| = \sum_{j=1}^{k} (-1)^{j-1} \sum_{S \subseteq \{1,\ldots,k\}: |S| = j} |\textstyle\bigcap_{i \in S} \overline{I}_i| = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} \binom{n-j}{m}.$$

[exam continued on next page]

5. **Balls and Bins** [*Total of 14 points.*]

Suppose k balls are thrown independently and uniformly at random into n labeled bins. For $i = 1, \ldots, n$, let E_i denote the event that bin i is empty. (**NOTE:** Your answers should not involve any summation signs.)

(a) Find $\mathbb{P}[E_1]$. Write your answer in the box; no justification required. *3pts*

$\left(1-\frac{1}{n}\right)^k$. [Each ball has probability $1-\frac{1}{n}$ of missing bin 1, so $\mathbb{P}[E_1] = \mathbb{P}[\cap_{i=1}^k(\text{ball } i \text{ misses bin 1})] = \left(1-\frac{1}{n}\right)^k$, by independence. Alternatively, there are n^k ways to assign labeled balls $1, \ldots, k$ to labeled bins $1, \ldots, n$, and $(n-1)^k$ ways to assign the balls to bins $2, \ldots, n$. So, $\mathbb{P}[E_1] = \frac{(n-1)^k}{n^k} = \left(1-\frac{1}{n}\right)^k$.]

(b) Find $\mathbb{P}[E_1 \cap \overline{E_2}]$. Write your answer in the box; no justification required. *3pts*

$\left(1-\frac{1}{n}\right)^k - \left(1-\frac{2}{n}\right)^k$. [Note that $\mathbb{P}[E_1 \cap \overline{E_2}] = \mathbb{P}[E_1] - \mathbb{P}[E_1 \cap E_2]$. From part (a), we have $\mathbb{P}[E_1] = \left(1-\frac{1}{n}\right)^k$. Each ball has probability $1-\frac{2}{n}$ of missing both bins 1 and 2, so $\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[\cap_{i=1}^k(\text{ball } i \text{ misses bins 1 and 2})] = \left(1-\frac{2}{n}\right)^k$, by independence. Alternatively, since there are $(n-2)^k$ ways to assign labeled balls $1, \ldots, k$ to labeled bin $3, \ldots, n$, we again conclude $\mathbb{P}[E_1 \cap E_2] = \frac{(n-2)^k}{n^k} = \left(1-\frac{2}{n}\right)^k$.]

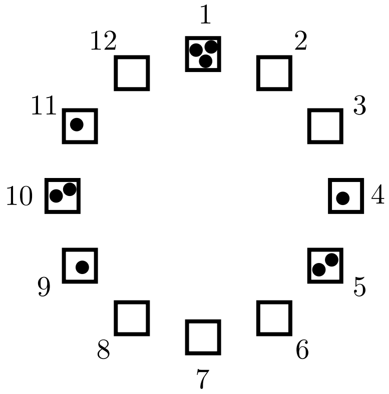
NOTE: For the remaining parts, let p and q denote the probabilities found in parts (a) and (b), respectively. Express your answers to parts (c) and (d) in terms of some or all of n, p, and q.

(c) For this part, assume $k \geq n$. Apply the union bound to $\mathbb{P}[\text{at least one bin is empty}]$ and use this result to find a lower bound on $\mathbb{P}[\text{no bin is empty}]$. Write your answer in the box; no justification required. (**Hint**: Express the event "at least one bin is empty" in terms of $E_1, \ldots, E_n$.) *4pts*

$\mathbb{P}[\text{no bin is empty}] \geq 1 - np$.

[We have $\mathbb{P}[\text{no bin is empty}] = 1 - \mathbb{P}[\text{at least one bin is empty}]$, while $\mathbb{P}[\text{at least one bin is empty}] = \mathbb{P}[E_1 \cup \cdots \cup E_n]$. Since $\mathbb{P}[E_i] = p$ for all $i = 1, \ldots, n$, the union bound yields $\mathbb{P}[E_1 \cup \cdots \cup E_n] \leq \sum_{i=1}^n \mathbb{P}[E_i] = np$. Therefore, $\mathbb{P}[\text{no bin is empty}] = 1 - \mathbb{P}[E_1 \cup \cdots \cup E_n] \geq 1 - np$.

(d) Suppose the n bins are arranged in a circle. A contiguous set of empty bins flanked by non-empty bins is called a *gap*. Below is an example for $k = 10$ and $n = 12$ with $G = 3$ gaps: $\{2, 3\}$, $\{6, 7, 8\}$, and $\{12\}$. Find the expectation $\mathbb{E}[G]$, where G denotes the number of gaps. Write your final answer in the box provided, **and give a brief justification in the space to the right of the figure**. *4pts*



$\mathbb{E}[G] = nq$. Let Ω denote the sample space and, for ease of notation, let bin $n+1$ correspond to bin 1. Then, for $i = 1, \ldots, n$, define an indicator random variable I_i such that $I_i(\omega) = 1$ if bin i is empty and bin $i+1$ is non-empty in $\omega \in \Omega$. Then, $G = I_1 + \cdots + I_n$, and, by linearity of expectation, $\mathbb{E}[G] = \mathbb{E}[I_1 + \cdots + I_n] = \sum_{i=1}^n \mathbb{E}[I_i]$. Noting $\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \mathbb{P}[E_i \cap \overline{E_{i+1}}] = q$ for all $i = 1, \ldots, n$, we obtain $\mathbb{E}[G] = nq$.

[exam continued on next page]

6. Random Permutations [*Total of 17 points.*]

Suppose a permutation is sampled uniformly at random from the set of all permutations of $\{1,\ldots,n\}$. Let Ω denote the sample space, and for $i = 1,\ldots,n$, let X_i be random variables such that $X_i(\omega) \in \{1,\ldots,n\}$ denotes the number that i maps to under permutation $\omega \in \Omega$. (**NOTE:** Your answers should not involve any summation signs.)

(a) What is the cardinality of Ω? Write your answer in the box; no justification required. *2pts*

$n!$

(b) Find $\mathbb{P}[X_1 = 2, X_2 = 1]$. Write your answer in the box; no justification required. *3pts*

$\frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$. [The number of permutations $\omega \in \Omega$ with $X_1(\omega) = 2$ and $X_2(\omega) = 1$ is $(n-2)!$, so $\mathbb{P}[X_1 = 2, X_2 = 1] = \frac{(n-2)!}{|\Omega|} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$.]

(c) Find $\mathbb{P}[X_1 = 2, X_2 \neq 1]$. Write your answer in the box; no justification required. *3pts*

$\frac{(n-1)!-(n-2)!}{n!} = \frac{1}{n} - \frac{1}{n(n-1)}$. [The number of permutations $\omega \in \Omega$ with $X_1(\omega) = 2$ is $(n-1)!$, so $\mathbb{P}[X_1 = 2] = \frac{(n-1)!}{n!}$. Hence, combining this result with the answer from part (b), we obtain $\mathbb{P}[X_1 = 2, X_2 \neq 1] = \mathbb{P}[X_1 = 2] - \mathbb{P}[X_1 = 2, X_2 = 1] = \frac{(n-1)!-(n-2)!}{n!} = \frac{1}{n} - \frac{1}{n(n-1)}$. Alternatively, if $X_2 \neq 1$ and $X_2 \neq 2$ (since $X_1 = 2$), there are $(n-2)$ options for X_2, and then $(n-2)!$ possibilities for $X_3,\ldots,X_n$, giving a total of $(n-2)\cdot(n-2)!$ permutations of interest. Therefore, we have $\mathbb{P}[X_1 = 2, X_2 \neq 1] = \frac{(n-2)\cdot(n-2)!}{n!} = \frac{[(n-1)-1](n-2)!}{n!} = \frac{(n-1)!-(n-2)!}{n!}$.]

(d) Find $\mathbb{P}[X_1 < X_2 - 1]$. Write your final answer in the box below, **and show your work in the space provided**. (**Hint**: Use the Total Probability Rule. Also, recall from earlier in the course that $\sum_{i=1}^{m} i = \frac{m(m+1)}{2}$.) *5pts*

The answer is $\frac{1}{2} - \frac{1}{n}$, which can be derived as follows:

$$\begin{aligned}
\mathbb{P}[X_1 < X_2 - 1] &= \sum_{k=1}^{n} \mathbb{P}[X_1 < X_2 - 1 \mid X_2 = k]\, \mathbb{P}[X_2 = k] \\
&= \sum_{k=3}^{n} \frac{k-2}{n-1}\left(\frac{1}{n}\right) \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n-2} i = \frac{1}{n(n-1)} \frac{(n-2)(n-1)}{2} = \frac{n-2}{2n} = \frac{1}{2} - \frac{1}{n},
\end{aligned}$$

where in the second equality we have used $\mathbb{P}[X_1 < X_2 - 1 \mid X_2 = 1] = \mathbb{P}[X_1 < X_2 - 1 \mid X_2 = 2] = 0$, $\mathbb{P}[X_1 < X_2 - 1 \mid X_2 = k] = \frac{k-2}{n-1}$ for $k = 3,\ldots,n$, and $\mathbb{P}[X_2 = k] = \frac{1}{n}$ for all k. Alternatively, there are $n!/2$ permutations in which $X_1 < X_2$. From these we can subtract the $(n-1)!$ permutations for which $X_1 = X_2 - 1$, giving a probability of $\mathbb{P}[X_1 < X_2 - 1] = \frac{1}{n!}\left[\frac{n!}{2} - (n-1)!\right] = \frac{1}{2} - \frac{1}{n}$.

(e) Let p denote the probability found in part (c), and let D_n denote the number of derangements of $\{1,\ldots,n\}$. Consider the event Δ that the sampled permutation is a derangement. Find $\mathbb{P}[\Delta \mid X_1 = 2, X_2 \neq 1]$ in terms of p, n, and D_k for a suitable k. Write your final answer in the box below; no justification required. *4pts*

$\frac{D_{n-1}}{n!\,p}$. [As discussed in the proof of Theorem 11.2 in Note 11, the number of derangements with $X_1 = 2$ and $X_2 \neq 1$ is D_{n-1}, so $\mathbb{P}[\Delta, X_1 = 2, X_2 \neq 1] = \frac{D_{n-1}}{n!}$. Therefore,

$$\mathbb{P}[\Delta \mid X_1 = 2, X_2 \neq 1] = \frac{\mathbb{P}[\Delta, X_1 = 2, X_2 \neq 1]}{\mathbb{P}[X_1 = 2, X_2 \neq 1]} = \frac{D_{n-1}}{n! \cdot p},$$

which is equal to $\frac{D_{n-1}}{(n-1)!-(n-2)!}$.]

[End of Exam!]