CS70 Discrete Mathematics and Probability Theory, Fall 2019

Final Exam

7:00-10:00pm, 20 December

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 **7 double-sided** *sheets (14 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 3-8, 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 three 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.

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

(i) You have 3 hours: there are 8 questions on this exam worth a total of 185 points.

[exam starts on next page]

1. True/False [*No justification; answer by shading the correct bubble. 2 points per answer unless otherwise stated; total of 39 points. No penalty for incorrect answers.*]

(a) Let $P(x), Q(x), R(x)$ be propositions involving a variable x belonging to a universe $\mathcal{U}$. Suppose you are asked to prove the following statement: $(\forall x \in \mathcal{U})(P(x) \Rightarrow (\neg Q(x) \vee \neg R(x))$. Which of the following would constitute a valid proof strategy? Answer **YES** or **NO** for each by shading the appropriate bubble.

YES	**NO**		
○	●	Find some $x \in \mathcal{U}$ for which $P(x)$ holds and $Q(x)$ does not hold.	*2pts*
●	○	Show that $P(x)$ is false for all $x \in \mathcal{U}$.	*2pts*
●	○	Show that, for all $x \in \mathcal{U}$ for which $P(x)$ and $Q(x)$ both hold, $R(x)$ does not hold.	*2pts*
●	○	Show that $R(x)$ is false for all $x \in \mathcal{U}$.	*2pts*
●	○	For all $x \in \mathcal{U}$, show that if $Q(x)$ and $R(x)$ both hold, then $P(x)$ does not hold.	*2pts*

(b) Classify each of the following functions $f : \mathbb{Z} \to \mathbb{Z}$ as (i) neither 1-1 nor onto; (ii) 1-1 but not onto; (iii) onto but not 1-1; (iv) both 1-1 and onto (a bijection).

(i) Neither	**(ii)** 1-1	**(iii)** Onto	**(iv)** Both		
○	●	○	○	$f(n) = 3n - 4$	*1pt*
●	○	○	○	$f(n) = -3n^2 + 7$	*1pt*
○	○	○	●	$f(n) = n + 17$	*1pt*
○	○	●	○	$f(n) = \lfloor \frac{n}{2} \rfloor$. [Note: for any rational number r, $\lfloor r \rfloor$ denotes the largest integer less than or equal to r.]	*1pt*
●	○	○	○	$f(n) = n \bmod 1000$	*1pt*

[Q1 continued on next page]

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

TRUE	FALSE		
●	○	A stable marriage instance has a *unique* stable pairing if and only if the male-optimal pairing is the same as the female-optimal pairing.	*2pts*
○	●	In a stable marriage instance, if every man has a different favorite woman, and every woman has a different favorite man, then there is a *unique* stable pairing.	*2pts*
○	●	There exists a tree with 7 vertices whose degrees are respectively $(1, 1, 1, 2, 2, 3, 4)$.	*2pts*
●	○	In any simple, undirected graph G with at least two vertices, there must be at least two vertices with the same degree.	*2pts*
●	○	If (N, e) is a valid RSA public key with private key d, then (N, d) is also a valid public key with private key e.	*2pts*
●	○	Let A be an event with $\mathbb{P}[A] = 1$ and let B be any other event. Then, A and B are independent.	*2pts*
○	●	There exist random variables X, Y with $\text{Cov}(X, Y) > 0$ and $\text{Var}[X + Y] < \text{Var}[X] + \text{Var}[Y]$.	*2pts*
○	●	For some $0 < p < 1$, let W_1 and W_2 be independent Geometric(p) random variables. Then, $\mathbb{P}[W_1 + W_2 = n] = \binom{n}{2} p^2 (1-p)^{n-2}$ for all integers $n \geq 2$.	*2pts*
●	○	Suppose $X \sim \text{Exp}(\lambda)$ for arbitrary $\lambda > 0$. Then, for all $t \in \mathbb{R}^+$, $\mathbb{P}[X > t] > \mathbb{P}[X > 3t \mid X > t]$.	*2pts*
●	○	For a continuous random variable $X \sim \text{Uniform}(0, 1)$, $\mathbb{P}[X \in S] = 0$ for every countable subset S of the unit interval $(0, 1)$.	*2pts*
○	●	Let X be a continuous random variable with p.d.f. $f(x)$. Then, for all intervals $(a, b) \subseteq \mathbb{R}$, $\mathbb{P}[X \notin (a, b)] = \int_a^b [1 - f(x)]dx$.	*2pts*
●	○	For X a random variable with finite mean $\mathbb{E}[X]$, and for all constants a and $\varepsilon > 0$, the generalized Markov inequality implies $\mathbb{P}[\lvert X - a \rvert \geq \varepsilon] \leq \frac{\text{Var}[X] + (\mathbb{E}[X] - a)^2}{\varepsilon^2}$.	*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 unless otherwise stated; total of 53 points. No penalty for incorrect answers.*]

(a) **Note: The next four questions all concern the following equation in the integer variable n:**

$$133^5 + 110^5 + 84^5 + 27^5 = n^5.$$

(i) What is the value of n mod 2? *2pts*

$n \equiv 0 \pmod 2$. [Clearly $133^5 \equiv 27^5 \equiv 1 \pmod 2$ while $110^5 \equiv 84^5 \equiv 0 \pmod 2$. Hence the sum is $1 + 0 + 0 + 1 \equiv 0 \pmod 2$. Hence $n^5 \equiv 0 \pmod 2$, which implies $n \equiv 0 \pmod 2$.]

(ii) What is the value of n mod 3? *2pts*

$n \equiv 0 \pmod 3$. [$133 \equiv 1 \pmod 3$, $110 \equiv 2 \pmod 3$, $84 \equiv 27 \equiv 0 \pmod 3$. Hence the sum is $1 + 32 + 0 + 0 \equiv 0 \pmod 3$. Hence $n^5 \equiv 0 \pmod 3$, which implies $n \equiv 0 \pmod 3$.]

(iii) What is the value of n mod 5? [Hint: Use Fermat's Little Theorem.] *2pts*

$n \equiv 4 \pmod 5$. [By FLT we have $a^5 \equiv a \pmod 5$ for all a. Hence the equation becomes $n \equiv 133 + 110 + 84 + 27 \equiv 3 + 0 + 4 + 2 \equiv 4 \pmod 5$.]

(iv) Assuming that n exists and is less than 170, what is n? *2pts*

$n = 144$. [Parts (a)–(c) imply that n is even, a multiple of 3, and equal to 4 mod 5. Obviously $n > 133$, and we are also given that $n < 170$. The even numbers in this range that are equal to 4 mod 5 are: 134, 144, 154, 164. Of these, only 144 is a multiple of 3.]
[**Note:** You may be interested to know that the above equation was in fact discovered in 1967 in order to disprove a conjecture of Euler, which stated that the sum of the fifth powers of four positive integers cannot itself be a fifth power.]

(b) What is the inverse of 7 mod 60? [Your answer should be an integer in $\{0, 1, \ldots, 59\}$.] *3pts*

43. [The sequence of recursive calls to the extended Euclidean algorithm is: $(60, 7) \to (7, 4) \to (4, 3) \to (3, 1) \to (1, 0)$. The corresponding sequence of returned triples is: $(1, 1, 0) \to (1, 0, 1) \to (1, 1, -1) \to (1, -1, 2) \to (1, 2, -17)$. Hence the inverse is $-17 \equiv 43 \pmod{60}$.]

(c) A dial on a piece of equipment has a circular scale with integer markings $\{0, 1, \ldots, 19\}$ arranged clockwise in increasing order. Whenever it detects an event, the dial jumps a distance of 7 clockwise on the scale. If it starts at position 0, after how many jumps will the dial first reach position 5 on the scale? *3pts*

15. [Let z denote the number of jumps. Then z satisfies the modular equation $7z \equiv 5 \pmod{20}$. By inspection we see that $7^{-1} \equiv 3 \pmod{20}$. Multiplying the equation by 3 gives $z \equiv 15 \pmod{20}$.]

(d) Polly has chosen a degree-13 polynomial $P(x)$ over $GF(19)$, but has forgotten one of its coefficients. Fortunately, however, she did write down the value of $P(x)$ at a few points $x > 0$. How many of these values does she need in order to reconstruct the missing coefficient? *3pts*

1. [Suppose Polly knows that $P(x) = y$ for some $x > 0$. Substituting x into $P(x)$ and setting the result equal to y gives a simple linear equation (mod 19) for the missing coefficient.]

[Q2 continued on next page]

(e) Alice wants to send a message to Bob over an expensive, noisy channel, which may corrupt up to 10% of the packets sent. If Alice's budget allows her to send only 100 packets in total, and she uses the Berlekamp-Welch scheme, how many message packets can she send? *3pts*

80. [Alice sends 100 packets (the maximum possible), of which $k = 10$ may be corrupted. Thus if the number of message packets is n we have $n + 2k = 100$, and hence $n = 80$.]

(f) Consider a 9×9 regular grid consisting of the vertices (i, j), where $i, j \in \{0, 1, \ldots, 9\}$. Your goal is to move from the $(0, 0)$ corner of the grid to the $(9, 9)$ corner while obeying the following rule: from any given position (i, j), you are allowed to move to either $(i, j + 1)$ or $(i + 1, j)$, provided that you stay inside the grid. **For the following two questions, leave your answers in terms of binomial coefficients.**

(i) How many paths from $(0, 0)$ to $(4, 5)$ are there? [Hint: Note that all of these paths are of length 9.] *3pts*

$\binom{9}{4}$ or $\binom{9}{5}$. [As given in the hint, all such paths must consist of 9 steps. Exactly 4 of these steps must be to the right while exactly 5 of them must be upwards. Choosing which 4 (or 5) of the 9 steps are to the right (or upwards) uniquely specifies a path, and there are $\binom{9}{4}$ (or $\binom{9}{5}$) such choices.]

(ii) How many paths from $(0, 0)$ to $(9, 9)$ pass through $(3, 3)$ or $(6, 6)$? *3pts*

$2\binom{6}{3}\binom{12}{6} - \binom{6}{3}^3$. [Any path from $(0, 0)$ to $(9, 9)$ passing through $(3, 3)$ consists of a path from $(0, 0)$ to $(3, 3)$ concatenated with a path from $(3, 3)$ to $(9, 9)$. By the same reasoning as in part (i), there exist $\binom{6}{3}\binom{12}{6}$ such paths. Similarly, there are $\binom{12}{6}\binom{6}{3}$ paths passing through $(6, 6)$, and $\binom{6}{3}^3$ paths passing through both $(3, 3)$ and $(6, 6)$. Inclusion-exclusion gives the final answer.]

(g) For $0 < p < 1$, let $W_1, W_2, \ldots, W_n$ be i.i.d. Geometric(p) random variables and define $S_n := W_1 + \cdots + W_n$.

(i) Let m be a positive integer $\geq n$. For how many different configurations $(a_1, a_2, \ldots, a_n)$ is the conditional probability $\mathbb{P}[W_1 = a_1, \ldots, W_n = a_n \mid S_n = m]$ non-zero? *3pts*

$\binom{m-1}{n-1}$. [Since $W_i \in \{1, 2, \ldots\}$, we need to count all configurations $(a_1, \ldots, a_n)$, where $a_i \in \{1, 2, \ldots\}$, such that $\sum_{i=1}^n a_i = m$. This is the same as distributing m balls over n bins, where each bin contains at least one ball.]

(ii) Find $\mathbb{P}[W_1 = a_1, W_2 = a_2 \mid S_2 = m]$ for (a_1, a_2) such that $a_1 + a_2 = m \geq 2$. *3pts*

$\frac{1}{m-1}$. [$\mathbb{P}[W_1 = a_1, W_2 = a_2 \mid S_2 = m] = \frac{\mathbb{P}[W_1=a_1, W_2=a_2]}{\mathbb{P}[W_1+W_2=m]} = \frac{\mathbb{P}[W_1=a_1]\mathbb{P}[W_2=a_2]}{\sum_{j=1}^{m-1}\mathbb{P}[W_1=j]\mathbb{P}[W_2=m-j]} = \frac{(1-p)^{a_1+a_2-2}p^2}{\sum_{j=1}^{m-1}(1-p)^{j+m-j-2}p^2} = \frac{(1-p)^{m-2}p^2}{\sum_{j=1}^{m-1}(1-p)^{m-2}p^2} = \frac{1}{\sum_{j=1}^{m-1}1} = \frac{1}{m-1}$.

Alternative solution: $\{S_2 = m\}$ is the event that the second heads in a sequence of coin tosses appears on the m^{th} toss. There are $(m-1)$ positions for the first heads, and each such sequence has probability $p^2(1-p)^{m-2}$, for a total of $\mathbb{P}[S_2 = m] = (m-1)(1-p)^{m-2}p^2$. Therefore, $\mathbb{P}[W_1 = a_1, W_2 = a_2 \mid S_2 = m] = \frac{(1-p)^{m-2}p^2}{(m-1)(1-p)^{m-2}p^2} = \frac{1}{m-1}$.]

(h) For a continuous random variable $X \sim$ Uniform(0,2), find $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, and $\mathrm{Var}[X]$. *3pts*

$\mathbb{E}[X] = 1, \mathbb{E}[X^2] = 4/3, \mathrm{Var}[X] = 1/3$. [The p.d.f. of X is $f(x) = 1/2$, and so $\mathbb{E}[X] = \int_0^2 \frac{x}{2}\,dx = 1, \mathbb{E}[X^2] = \int_0^2 \frac{x^2}{2}\,dx = 8/6 = 4/3$ and $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4/3 - 1 = 1/3$.]

[Q2 continued on next page]

(i) Suppose $X \sim \text{Normal}(2,2)$ and $Y \sim \text{Normal}(0,1)$ are independent random variables, and define $Z = X - 2Y - 3$. Find $\mathbb{E}[Z]$ and $\text{Var}[Z]$. *3pts*

$\mathbb{E}[Z] = -1, \text{Var}[Z] = 6$. [By linearity of expectation, we have $\mathbb{E}[Z] = \mathbb{E}[X - 2Y - 3] = \mathbb{E}[X] - 2\mathbb{E}[Y] - \mathbb{E}[3] = 2 - 2 \cdot 0 - 3 = -1$, and since X and Y are independent, it follows that $\text{Var}[Z] = \text{Var}[X - 2Y - 3] = \text{Var}[X] + \text{Var}[-2Y] + \text{Var}[-3] = 2 + (-2)^2\text{Var}[Y] + 0 = 2 + 4 \cdot 1 = 6$.]

(j) Let X be a continuous random variable with the following probability density function (p.d.f.): *3pts*

$$f(x) = \begin{cases} 0, & x < 1, \\ e^{1-x}, & x \geq 1. \end{cases}$$

Find the cumulative distribution function (c.d.f.) F of X.

$$F(x) = \begin{cases} 0, & \text{if } x < 1, \\ 1 - e^{1-x}, & \text{if } x \geq 1. \end{cases}$$

$$[F(x) = \mathbb{P}[X \leq x] = \textstyle\int_{-\infty}^{x} f(x)\,\mathrm{d}x = \begin{cases} 0, & \text{if } x < 1, \\ \int_1^x e^{1-y}\,\mathrm{d}y = 1 - e^{1-x}, & \text{if } x \geq 1. \end{cases}$$

Alternative solution: $f(x) = f_{\mathcal{E}}(x-1)$, where $f_{\mathcal{E}}$ is the p.d.f. of an $\text{Exp}(1)$ variable $\mathcal{E}$. That is, $X = \mathcal{E} + 1$, and so $F(x) = \mathbb{P}[\mathcal{E} \leq x - 1] = 1 - e^{-(x-1)}$ if $x \geq 1$, and $F(x) = 0$ otherwise.]

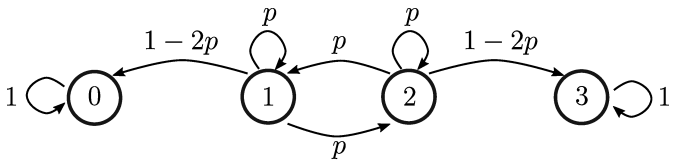
(k) Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables and define $S_n = X_1 + \cdots + X_n$. If $\mathbb{P}[X_i = +1] = \frac{3}{4}$ and $\mathbb{P}[X_i = -1] = \frac{1}{4}$ for $\forall i \in \mathbb{Z}^+$, what is $\lim_{n\to\infty} \mathbb{P}[2S_n < (1 + 10^{-100})n]$? *3pts*

1. [The X_i are i.i.d. with mean $1/2$ and so S_n/n converges by the law of large numbers to $1/2$; that is, for any $\varepsilon > 0$, we have $\mathbb{P}[|\frac{S_n}{n} - \frac{1}{2}| < \varepsilon] \to 1$. Therefore, in particular $\mathbb{P}[2S_n < (1 + 10^{-100})n] = \mathbb{P}[\frac{S_n}{n} - \frac{1}{2} < \frac{10^{-100}}{2}] \geq \mathbb{P}[|\frac{S_n}{n} - \frac{1}{2}| < \frac{10^{-100}}{2}]$ converges to 1.]

(l) Let $X_1, \ldots, X_n$ be i.i.d. random variables and let F denote their c.d.f. Find $\mathbb{P}[\min\{X_1, \ldots, X_n\} \leq a]$ in terms of F. *3pts*

$1 - [1 - F(a)]^n$. [First, note that $\mathbb{P}[\min\{X_1, \ldots, X_n\} \leq a] = 1 - \mathbb{P}[\min\{X_1, \ldots, X_n\} > a] = 1 - \mathbb{P}[X_1 > a, \ldots, X_n > a]$. Then, by independence, we have $\mathbb{P}[X_1 > a, \ldots, X_n > a] = \mathbb{P}[X_1 > a] \cdots \mathbb{P}[X_n > a] = (1 - \mathbb{P}[X_1 \leq a]) \cdots (1 - \mathbb{P}[X_n \leq a]) = [1 - F(a)] \cdots [1 - F(a)] = [1 - F(a)]^n$.]

(m) Consider a 4-state Markov Chain $\{X_n, n \in \mathbb{N}\}$ with the following allowed transitions, where $0 < p < \frac{1}{2}$:



(i) Find $\mathbb{P}[X_3 = 1 \mid X_0 = 1]$. *3pts*

$4p^3$. [There are four paths going from state 1 to state 1 in three steps: $(X_0, X_1, X_2, X_3) = (1,1,1,1), (1,1,2,1), (1,2,2,1)$ and $(1,2,1,1)$. Each of these happens with probability p^3.]

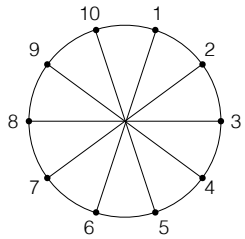
(ii) Find $\mathbb{E}[\text{Time to hit either state 0 or state 3} \mid X_0 = 1]$. *3pts*

$\frac{1}{1-2p}$. [Defining $\tau_i = \mathbb{E}[\text{time to hit either state 0 or state 3} \mid X_0 = i]$, the system of equations is $\tau_0 = \tau_3 = 0, \tau_1 = (1-2p)(\tau_0 + 1) + p(\tau_1 + 1) + p(\tau_2 + 1) = 1 + p(\tau_1 + \tau_2), \tau_2 = (1-2p)(\tau_3 + 1) + p(\tau_1 + 1) + p(\tau_2 + 1) = 1 + p(\tau_1 + \tau_2)$, from which it follows immediately that $\tau_1 = \tau_2$, and hence $\tau_1 = 1/(1-2p)$.]

[exam continued on next page]

3. Möbius Ladders [*Total of 15 points.*]

Consider the family of graphs G_n ($n \geq 2$) known as *Möbius ladders*. G_n has $2n$ vertices arranged in a single cycle, with an additional edge for each vertex connecting it to the "opposite" vertex on the cycle. The figure below shows the graph G_5. [Note: the point in the center where edges cross is *not* a vertex!]



In parts (a)–(c) below, indicate whether the claimed property holds for ALL values of n, for no (NONE) values of n, only for EVEN values of n, or only for ODD values of n, by shading the appropriate bubble.

(a) For which values of n (if any) does G_n have an Eulerian tour? *2pts*

NONE. [A graph has an Eulerian tour if and only if all vertex degrees are even. In this case, all vertex degrees are odd.]

(b) For which values of n (if any) does G_n have a Hamiltonian cycle? *2pts*

ALL. [The outer cycle is a Hamiltonian cycle.]

(c) For which values of n (if any) is G_n bipartite? *2pts*

ODD. [Because of the outer cycle, if G_n is bipartite then the two parts must consist of the odd-numbered vertices and the even-numbered vertices, respectively. Now each chord (edge crossing the cycle) connects vertex i to vertex $i + n$ (for $1 \leq i \leq n$). For G_n to be bipartite, these edges must connect vertices of *opposite* parities, which happens if and only if n is odd.]

In parts (d)–(f), you may use without proof results from class, provided you state them clearly.

(d) Is G_2 planar? Shade the correct bubble and **justify your answer.** *3pts*

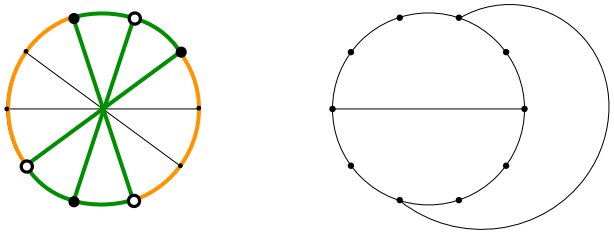
YES. [G_2 is just K_4, which was easily seen in class to be planar.]

(e) Is G_3 planar? Shade the correct bubble and **justify your answer.** *3pts*

NO. [G_3 is just $K_{3,3}$, which we know from class to be non-planar.]

(f) For all $n > 3$, show that G_n is non-planar. *3pts*

The first picture below shows how to find a copy of $K_{3,3}$ in G_5: the vertices on each side of $K_{3,3}$ are shown in bold black and white respectively, while each edge of $K_{3,3}$ is represented by a single edge of G_n (in green), except for two edges represented as paths (in gold). This obviously generalizes to any G_n for $n \geq 3$. Hence by Kuratowski's Theorem G_n is non-planar.



Alternatively, we can observe that any planar drawing of G_n must preserve the outer cycle (though of course it may no longer be on the exterior face). But once this is fixed, we can add at most two of the chords—one inside the cycle and the other outside the cycle. Once these two chords have been added, any further chord must cross one of them, as can be seen in the second picture above.

[exam continued on next page]

4. An Inductive Proof of Fermat's Little Theorem [*All parts to be justified. Total of 10 points.*]

Recall Fermat's Little Theorem (FLT): for any prime p, and all $a \in \{1, \ldots, p-1\}$, we have $a^{p-1} \equiv 1 \pmod{p}$. In class we gave a proof of FLT using a bijection between integers mod p. In this problem we'll look at a different, inductive proof based on the binomial theorem, which says that

$$(a+1)^p = a^p + \binom{p}{1}a^{p-1} + \binom{p}{2}a^{p-2} + \cdots + \binom{p}{p-1}a + 1. \qquad (*)$$

(a) Fix an arbitrary prime p. We will actually prove the following statement by induction. *2pts*

Claim: *For all natural numbers* a, $a^p \equiv a \pmod{p}$.

Explain why this Claim implies FLT.

Let a be any integer in $\{1, \ldots, p-1\}$. Then, since p is prime, a has an inverse mod p. Multiplying both sides of the given congruence $a^p \equiv a \pmod{p}$ by this inverse gives $a^{p-1} \equiv 1 \pmod{p}$, which is the statement of Fermat's Little Theorem.

(b) For any prime p, prove that p divides every binomial coefficient $\binom{p}{k}$ for $1 \leq k \leq p-1$. *3pts*

By definition, $\binom{p}{k} = \frac{p!}{k!(p-k)!}$. Rearranging, and writing $N = \binom{p}{k}$, we have $Nk!(p-k)! = p!$. Clearly p divides the right-hand side, so it must also divide the left-hand side. But $k!(p-k)!$ contains only factors that are strictly less than p, and p is prime, so $\gcd(p, k!(p-k)!) = 1$. Hence p must divide N, as required.

(c) Prove the Claim in part (a) by induction on a, using the binomial theorem $(*)$ and part (b) for the inductive step. *5pts*

Fix a prime p. We will prove by induction on a that, for every natural number a, $a^p \equiv a \pmod{p}$.

Base Case: For $a = 0$, the claim says that $0^p \equiv 0 \pmod{p}$, which is trivially true.

Inductive Step: Assume, for some arbitrary $a \geq 0$, that $a^p \equiv a \pmod{p}$ holds. We need to prove that $(a+1)^p \equiv a+1 \pmod{p}$ also holds. Using the binomial theorem as stated gives

$$(a+1)^p = a^p + \binom{p}{1}a^{p-1} + \binom{p}{2}a^{p-2} + \cdots + \binom{p}{p-1}a + 1. \qquad (*)$$

Now by part (b) all terms on the right-hand side of $(*)$, except for the first and the last term, are zero mod p. Hence if we take $(*)$ mod p we get

$$(a+1)^p \equiv a^p + 1 \pmod{p}.$$

By the induction hypothesis, we know also that $a^p \equiv a \pmod{p}$. Hence we conclude

$$(a+1)^p \equiv a + 1 \pmod{p},$$

which completes the proof of the inductive step.

[exam continued on next page]

5. Testing Equality of Polynomials [*All parts to be justified unless stated otherwise. Total of 14 points.*]

Let $P(x), Q(x)$ be polynomials of degree at most d over $GF(q)$, where $d \leq q/2$. We do not know the coefficients of P or Q, but instead we are given a black box for each of them that, when given as input a point $x \in GF(q)$, outputs the value of $P(x)$ (respectively, $Q(x)$).

We want to use these black boxes to test whether $P = Q$ (i.e., whether they are the same polynomial).

(a) If $P \neq Q$, what is the maximum possible number of points $x \in GF(q)$ for which $P(x) = Q(x)$? *2pts*

Write your answer in the box; no justification required.

d. [If $P \neq Q$ then $P(x) - Q(x)$ is a non-zero polynomial of degree at most d and hence can have at most d zeros. But zeros of $P(x) - Q(x)$ correspond precisely to values of x for which $P(x) = Q(x)$.]

(b) Explain how you would use the black boxes to test whether $P = Q$, and specify how many evaluations of each black box you would need. **Justify your answer.** *4pts*

For each $x \in \{0, 1, \ldots, d\}$ (or any set of $d+1$ distinct values $x \in GF(q)$), use the black boxes to test whether $P(x) = Q(x)$. If one of these tests detects a difference, then output "not the same"; otherwise, output "same".

If $P = Q$ then this test will always output "same", since it only outputs "not the same" when it finds a specific x for which $P(x) \neq Q(x)$. If $P \neq Q$ then by part (a) there can be at most d values of x for which $P(x) = Q(x)$; hence at least one of the $d+1$ test values of x must satisfy $P(x) \neq Q(x)$, so the procedure will output "not the same", as required. The test uses $d+1$ evaluations of each black box.

(c) Suppose now that you are given a random number generator that outputs independent uniform samples from $\{0, 1, \ldots, q-1\}$. Explain how to use the generator and just *one* evaluation of each black box to design a randomized test with the following behavior: *4pts*

(i) if $P = Q$ then the test always outputs "same";

(ii) if $P \neq Q$ then the test outputs "not the same" with probability at least $1/2$.

Justify your answer. [Hint: Use the fact that $d \leq q/2$.]

Use the generator to pick a uniformly random element $x \in \{0, 1, \ldots, q-1\}$. Then use the black boxes to test whether $P(x) = Q(x)$. If yes, then output "same", else output "not the same". This procedure uses just one evaluation of each black box.

If $P = Q$ then, as in part (b), this procedure will always output "same". If $P \neq Q$, it will output "same" only if the random x chosen happens to satisfy $P(x) = Q(x)$, i.e., x happens to be a zero of the non-zero polynomial $P(x) - Q(x)$. By part (a) there are at most d such zeros x, and x is picked uniformly from a set of size q, so the probability this happens is at most $\frac{d}{q} \leq \frac{1}{2}$, as required.

(d) How could you increase the success probability in case (ii) of part (c) to $1 - 2^{-t}$ for any desired positive integer t? **Justify your answer.** [Note: You may make additional uses of the generator and black boxes.] *4pts*

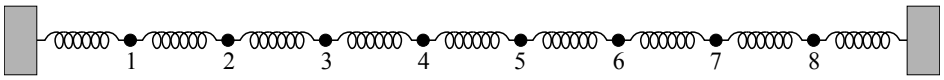
Repeat the test in part (c) t times, using an independent random sample x each time. If one or more of these tests outputs "not the same" then output "not the same", else output "same".

Once again, if $P = Q$ then this procedure will always output the correct answer "same". On the other hand, if $P \neq Q$ then the procedure will only output the incorrect answer "same" if every one of the t tests fails to find an x for which $P(x) \neq Q(x)$. Since, by part (c), this failure happens with probability at most 1/2 on each test, and the tests are independent, the probability of an incorrect answer overall is at most 2^{-t}, as required.

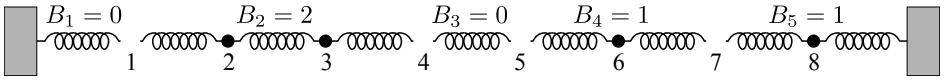
[exam continued on next page]

6. Sampling without Replacement [*Write your answer in the box provided. Total of 16 points.*]

Suppose the Physics 7A class is doing an experiment involving n beads connected by springs in a linear chain, as illustrated below ($n = 8$ in the example). The beads are labeled $1, \ldots, n$.



The instructor brings a well-shuffled deck of n cards numbered $1, \ldots, n$, and draws $k < n$ cards from the deck *without replacement*. She then removes the beads corresponding to the numbers drawn, thereby producing $k + 1$ connected components of bead-spring chains. Let B_i denote the number of beads in connected component i. For example, if $k = 4$ and the set of numbers drawn are $\{1, 4, 5, 7\}$, then the resulting configuration with 5 connected components is:



(**NOTE:** Whenever possible, express your answers in terms of binomial coefficients.)

(a) Are $B_1, \ldots, B_{k+1}$ independent random variables? Shade the correct bubble. ○ Yes ● No *2pts*

(b) How many distinct configurations $(B_1, \ldots, B_{k+1})$ are possible? No justification required. *3pts*

$\binom{n}{k}$. [The configuration $(B_1, \ldots, B_{k+1})$ is completely determined by the k beads that get removed, and there are $\binom{n}{k}$ ways to choose k beads from $1, 2, \ldots, n$.]

(c) For $i \in \{1, \ldots, k+1\}$, what is the range of B_i? No justification required. *2pts*

$\{0, 1, \ldots, n-k\}$. [$n - k$ beads remain, so $\sum_{i=1}^{k+1} B_i = n - k$, while $B_i \in \{0, \ldots, n-k\}$ for all $i = 1, \ldots, k+1$.]

(d) For $i \in \{1, \ldots, k+1\}$, find $\mathbb{P}[B_i = b]$, where b is in the range found in part (c). Write your final answer in the box below, **and justify your answer in the space provided**. *5pts*

$\dfrac{\binom{n-b-1}{k-1}}{\binom{n}{k}}$. [There are $\binom{n}{k}$ configurations of $(B_1, \ldots, B_{k+1})$ and every configuration is equally likely, so $\mathbb{P}[B_1 = b_1, \ldots, B_{k+1} = b_{k+1}] = \frac{1}{\binom{n}{k}}$, which implies

$$\mathbb{P}[B_i = b] = \frac{\text{number of configurations with } B_i = b}{\binom{n}{k}}.$$

The numerator is equal to the number of ways to divide up $n - k - b$ beads among the remaining k connected components. Equivalently, it is equal to the number of solutions to $x_1 + \cdots + x_k = n - k - b$, where $x_1, \ldots, x_k$ are non-negative integers. There are $\binom{n-k-b+(k-1)}{k-1} = \binom{n-b-1}{k-1}$ such solutions.]

(e) For $i \in \{1, \ldots, k+1\}$, find $\mathbb{E}[B_i]$ in terms of n and k. Your answer should not involve any summation signs. No justification required. [Hint: Do not try to solve this problem using the formula for $\mathbb{P}[B_i = b]$ found in part (d). There is a way to find $\mathbb{E}[B_i]$ without explicitly using $\mathbb{P}[B_i = b]$.] *4pts*

$\dfrac{n-k}{k+1}$. [$B_1, \ldots, B_{k+1}$ are identically distributed, so $\mathbb{E}[B_1] = \mathbb{E}[B_2] = \cdots = \mathbb{E}[B_{k+1}]$. Furthermore, since $B_1 + \cdots + B_{k+1} = n - k$ and $\mathbb{E}[B_1 + \cdots + B_{k+1}] = (k+1)\mathbb{E}[B_i]$ for any $i = 1, \ldots, k+1$, we obtain $\mathbb{E}[B_i] = \frac{n-k}{k+1}$.]

[exam continued on next page]

7. Random Hash Function [*Write your answer in the box provided. Total of 18 points.*]

Suppose a hash function is defined by mapping m keys independently and uniformly at random to the n bins of a hash table. Two different keys may get mapped to the same bin, and when that happens we say that a "collision" has occurred in that bin.

(a) Let C_1 denote the event that there is at least one collision in bin 1. Find $\mathbb{P}[C_1]$. No justification required. *4pts*

$1 - \left(1 - \frac{1}{n}\right)^m - \frac{m}{n}\left(1 - \frac{1}{n}\right)^{m-1}$ or $\sum_{k=2}^{m} \binom{m}{k} \frac{(n-1)^{m-k}}{n^m}$. [Note that $\mathbb{P}[C_1] = 1 - \mathbb{P}[\overline{C_1}]$, while $\mathbb{P}[\overline{C_1}] = \mathbb{P}[\text{bin 1 is empty}] + \mathbb{P}[\text{exactly 1 key gets mapped to bin 1}] = \left(1-\frac{1}{n}\right)^m + \frac{m}{n}\left(1-\frac{1}{n}\right)^{m-1}$. Alternatively, $\mathbb{P}[C_1] = \sum_{k=2}^{m} \mathbb{P}[\text{exactly } k \text{ keys get mapped to bin 1}] = \sum_{k=2}^{m} \binom{m}{k} \frac{1}{n^k}\left(1 - \frac{1}{n}\right)^{m-k} = \sum_{k=2}^{n} \binom{m}{k} \frac{(n-1)^{m-k}}{n^m}$.]

(b) Let p denote the answer to part (a), and let N denote the number of bins with collisions. Use Markov's Inequality to obtain an upper bound on $\mathbb{P}[N \geq \frac{n}{2}]$ in terms of p. No justification required. *3pts*

$2p$. [For $i = 1, \ldots, n$, let C_i denote the event that there is at least one collision in bin i. Then, $\mathbb{P}[C_1] = \cdots = \mathbb{P}[C_n] = p$ and $\mathbb{E}[N] = \sum_{i=1}^{n} \mathbb{P}[C_i] = np$. Hence, Markov's Inequality gives $\mathbb{P}[N \geq \frac{n}{2}] \leq \frac{\mathbb{E}[N]}{n/2} = 2p$.]

(c) Let K_i denote the number of keys assigned to bin i, where $i = 1, \ldots, n$. Find $\mathrm{Var}[K_i]$. Your answer should not involve any summation sign. No justification required. *3pts*

$\frac{m}{n}\left(1 - \frac{1}{n}\right)$. [As explained below, $K_i \sim \text{Binomial}(m, \frac{1}{n})$, so $\mathrm{Var}[K_i] = m \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)$.]

(d) Let v denote the answer to part (c). Use Chebyshev's Inequality to obtain an upper bound on $\mathbb{P}[K_i \geq \frac{3m}{n}]$ in terms of m, n, and v. Write your final answer in the box below, **and justify your answer in the space provided**. *4pts*

$\left(\frac{n}{2m}\right)^2 v$. [Each key gets mapped to bin i with probability $\frac{1}{n}$, independently of all other keys, so $K_i \sim \text{Binomial}(m, \frac{1}{n})$. Therefore, $\mathbb{E}[K_i] = \frac{m}{n}$ and

$$\mathbb{P}\left[K_i \geq \frac{3m}{n}\right] = \mathbb{P}\left[K_i - \mathbb{E}[K_i] \geq \frac{2m}{n}\right] \leq \mathbb{P}\left[\left|K_i - \mathbb{E}[K_i]\right| \geq \frac{2m}{n}\right] \leq \frac{\mathrm{Var}[K_i]}{(2m/n)^2} = \left(\frac{n}{2m}\right)^2 v.]$$

(e) For $k \leq m, n$, find $\mathbb{P}[\text{Exactly } k \text{ bins are non-empty}]$ in terms of m, n, k, and $S(a, b)$ for suitable values of a, b, where $S(a, b)$ denotes the number of surjections from $\{1, \ldots, a\}$ to $\{1, \ldots, b\}$. No justification required. [Note: You found a formula for $S(a, b)$ in Homework 8, but you do not need it here.] *4pts*

$\frac{\binom{n}{k} S(m, k)}{n^m}$. [There are n^m ways to map the m keys to n bins and they are all equally likely, so $\mathbb{P}[\text{Exactly } k \text{ bins are non-empty}] = \frac{1}{n^m} \times (\text{number of ways to get exactly } k \text{ non-empty bins})$. The latter factor can be computed as follows. There are $\binom{n}{k}$ ways to choose k distinct bins from $\{1, \ldots, n\}$, and there are $S(m, k)$ ways to map the m keys $\{1, \ldots, m\}$ surjectively to those k chosen bins. Multiplying these factors gives $\binom{n}{k} S(m, k)$.]

[exam continued on next page]

8. Competing Poisson Arrival Processes [*Write your answer in the box provided. Total of 20 points.*]

Suppose spam calls arrive at a call center according to a Poisson Arrival Process at rate $\lambda > 0$ per minute, while non-spam calls arrive according to a Poisson Arrival Process at rate 1 per minute, independently of spam calls. In this problem, all times are measured in minutes.

(a) Suppose you reset your timer to 0 exactly at noon and let W denote the waiting time (starting at noon) until either a spam or a non-spam call arrives. What is $\mathbb{P}[W \leq t]$? No justification required. *3pts*

$1 - e^{-(1+\lambda)t}$. [The two Poisson Arrival Processes are independent, so the number of calls (either spam or non-spam) follows a Poisson Arrival Process with rate $1+\lambda$, which means $W \sim \text{Exp}(1+\lambda)$. Hence, $\mathbb{P}[W \leq t] = 1 - e^{-(1+\lambda)t}$.]

(b) Define W as in part (a) and let E denote the event that exactly 1 call arrives in the time interval $(0, s)$, for $s > t$. Find $\mathbb{P}[W \leq t \mid E]$. Write your final answer in the box below, **and justify your answer in the space provided**. *4pts*

$\frac{t}{s}$. [For ease of notation, define $\rho := 1 + \lambda$. Via the definition of conditional probability, $\mathbb{P}[W \leq t \mid E] = \frac{\mathbb{P}[\{W \leq t\} \cap E]}{\mathbb{P}[E]}$, where we note that $\mathbb{P}[E] = (\rho s)e^{-\rho s}$. The numerator is equal to $\mathbb{P}[(\text{Exactly 1 arrival in } (0,t]) \cap (\text{No arrival in } (t,s))]$, which factorizes as $\mathbb{P}[\text{Exactly 1 arrival in } (0,t]] \times \mathbb{P}[\text{No arrival in } (t,s))] = [\rho t e^{-\rho t}][e^{-\rho(s-t)}] = (\rho t)e^{-\rho s}$, since arrival counts in disjoint intervals are independent in a Poisson Arrival Process. Therefore, $\frac{\mathbb{P}[\{W \leq t\} \cap E]}{\mathbb{P}[E]} = \frac{(\rho t)e^{-\rho s}}{(\rho s)e^{-\rho s}} = \frac{t}{s}$.

Alternative solution 1: To compute the numerator $\mathbb{P}[\{W \leq t\} \cap E]$, we let $V \sim \text{Exp}(\rho)$ be the time between the first and the second calls, and observe that $\{W \leq t\} \cap E = \{W \leq t\} \cap \{W + V \geq s\}$. That is, writing $R = \{(w, v) : w \leq t, w + v \geq s\}$, we have $\mathbb{P}[\{W \leq t\} \cap E] = \mathbb{P}[(W, V) \in R] = \int\int_R f(w,v)\, \mathrm{d}v\mathrm{d}w = \int_0^t \rho e^{-\rho w}[\int_{s-w}^{\infty} \rho e^{-\rho v}\, \mathrm{d}v]\mathrm{d}w = (\rho t)e^{-\rho s}$.

Alternative solution 2: Instead of performing the double integration, we can compute the numerator as $\mathbb{P}[(W \leq t) \cap E] = \int_0^t e^{-x\rho}e^{-\rho(s-x)}\rho dx$, where $e^{-x\rho}$ (respectively, $e^{-\rho(s-x)}$) corresponds to the probability of there being no arrivals in the time intervals $(0, x)$ (respectively, (x, s)), while ρdx corresponds to the probability of there being an arrival in the infinitesimal interval of size dx. Carrying out the integral, we obtain $\mathbb{P}[(W \leq t) \cap E] = \int_0^t e^{-x\rho}e^{-\rho(s-x)}\rho dx = \rho e^{-\rho s}\int_0^t dx = (\rho t)e^{-\rho s}$.

Alternative solution 3: Instead of performing any integration, we can observe that the Poisson Arrival process on $[0, s]$ is the limit of tossing sn independent coins of bias ρ/n as $n \to \infty$. The desired probability then is the same as the chance of seeing the first heads before the $\lfloor tn \rfloor^{\text{th}}$ coin toss, conditional on the event E_n of there being only one heads among the total $\lfloor sn \rfloor$ tosses. But all $\lfloor sn \rfloor$ sequences in E_n have the same probability $\left(\frac{\rho}{n}\right)\left(1 - \frac{\rho}{n}\right)^{\lfloor sn \rfloor - 1}$ of happening, and so the chance of a heads within the first $\lfloor tn \rfloor$ tosses is $\frac{\lfloor tn \rfloor (\rho/n)(1-\rho/n)^{\lfloor sn \rfloor - 1}}{\lfloor sn \rfloor (\rho/n)(1-\rho/n)^{\lfloor sn \rfloor - 1}} = \frac{\lfloor tn \rfloor}{\lfloor sn \rfloor}$, which converges to $\frac{t}{s}$ as $n \to \infty$.]

(c) Given that a call arrives, what is the probability that it is a spam call? No justification required. *3pts*

$\frac{\lambda}{1+\lambda}$. [Let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(1)$ denote the time of the first spam call and the time of the first non-spam call, respectively. The desired probability is $\mathbb{P}(X < Y) = \int\int_S \lambda e^{-\lambda x}e^{-y}\, \mathrm{d}y\mathrm{d}x$, where $S = \{(x, y) : x \leq y\}$. More explicitly, $\mathbb{P}(X < Y) = \int_0^{\infty} \lambda e^{-\lambda x}\int_x^{\infty} e^{-y}\, \mathrm{d}y\mathrm{d}x = \int_0^{\infty} \lambda e^{-x(\lambda+1)}\, \mathrm{d}x = \lambda/(1+\lambda)$.

Alternative solution: This problem is closely related to Problem 5 of Discussion 12. Let X denote the number of spam calls, and Y the number of non-spam calls, in any time interval $(0, t)$. Then, $\mathbb{P}[X = 1 \mid X + Y = 1] = \frac{\mathbb{P}[X=1] \times \mathbb{P}[Y=0]}{\mathbb{P}[X+Y=1]} = \frac{\lambda e^{-\lambda t} \times e^{-t}}{(1+\lambda)e^{-(1+\lambda)t}} = \frac{\lambda}{1+\lambda}$.]

[Q8 continued on next page]

(d) Let p denote the probability found in part (c). Let N denote the number of non-spam calls received before a spam call arrives. For $k \in \mathbb{N}$, find $\mathbb{P}[N = k]$ in terms of p and k. No justification required. *3pts*

$(1-p)^k p$. [When a call arrives, it is a non-spam call with probability $1-p$, independently of all other calls. So, the probability of receiving k non-spam calls before a spam call arrives is $(1-p)^k p$.]

(e) For $i \in \mathbb{Z}^+$, let X_i denote the number of non-spam calls received in the time interval $[i-1, i)$ and define $S_n = X_1 + X_2 + \cdots + X_n$. For $k \in \mathbb{N}$, find $\mathbb{P}[S_n = k]$. Your answer should not involve any summation signs. No justification required. *3pts*

$\frac{1}{k!} n^k e^{-n}$. [$X_i \sim$ Poisson(1) for all $i = 1, \ldots, n$. Furthermore, since they are independent, $S_n = X_1 + \cdots + X_n \sim$ Poisson(n), so $\mathbb{P}[S_n = k] = \frac{1}{k!} n^k e^{-n}$.]

(f) Let S_n be defined as in part (e), and let c and ε be some constants. For $\varepsilon < \frac{1}{2}$, what is $\lim_{n\to\infty} \mathbb{P}[S_n < c\, n^{\varepsilon} + n]$? Write your final answer in the box below, **and justify your answer in the space provided**. *4pts*

$\frac{1}{2}$. [Since $S_n = X_1 + \cdots + X_n \sim$ Poisson(n), $\mathbb{E}[S_n] = n$ and $\mathrm{Var}(S_n) = n$. Therefore, by the Central Limit Theorem, the distribution of $\frac{S_n - n}{\sqrt{n}}$ converges to Normal$(0,1)$ as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \mathbb{P}[S_n < cn^{\varepsilon} + n] = \lim_{n\to\infty} \mathbb{P}\Big[\frac{S_n - n}{\sqrt{n}} < cn^{\varepsilon - \frac{1}{2}}\Big] = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{2},$$

where the last equality follows from $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1$ and that $e^{-\frac{x^2}{2}}$ is symmetric about 0.]

[End of Exam!]