CS70 Discrete Mathematics and Probability Theory, Fall 2018

Midterm 2

8:00-10:00pm, 31 October

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 6 **double-sided** *sheets (11 numbered pages) on the exam. Notify a proctor immediately if a sheet is missing. Do* **not** *write any answers on page 12; it will not be scanned.*

(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 must 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.

(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, without proof, use theorems and lemmas that were proven in the notes and/or in lecture.

(i) You have 120 minutes: there are 6 questions on this exam worth a total of 128 points.

[exam starts on next page]

1. True/False [*No justification; answer by shading the correct bubble. 2 points per answer; total of 26 points. No penalty for incorrect answers.*]

(a) Are the following sets countable? Answer **YES** or **NO** for each set by shading the appropriate bubble.

YES	NO		
○	●	The set of all functions from $\mathbb{N}$ to $\{0,1\}$.	*2pts*
●	○	The set of all *finite* subsets of $\mathbb{N}$.	*2pts*
○	●	The set of all *irrational* numbers in the interval $[0,1]$.	*2pts*
●	○	The set of all possible Google search terms.	*2pts*
●	○	The set of all undirected graphs with a finite number of vertices.	*2pts*

(b) Answer each of the following questions **TRUE** or **FALSE** by shading the appropriate bubble.

TRUE	FALSE		
●	○	Any two distinct polynomials, each of degree at most d, can have at most d points in common.	*2pts*
●	○	Every function over $\mathrm{GF}(p)$ can be represented as a polynomial of degree at most $p-1$.	*2pts*
●	○	There exists a *fixed* computer program P such that no program can decide for any given input x whether P halts on x.	*2pts*
○	●	For any two sets A and B, if there exists an injection $f : A \to B$ and a surjection $g : B \to A$, then $\lvert A\rvert = \lvert B\rvert$.	*2pts*
●	○	For any two sets A and B, if there exists an injection $f : A \to B$ and an injection $g : B \to A$, then $\lvert A\rvert = \lvert B\rvert$.	*2pts*
●	○	For any two finite sets A and B, if there exists a surjection $f : A \to B$ such that $\lvert f^{-1}(x)\rvert = m$ for all $x \in B$, then $\lvert A\rvert = m\lvert B\rvert$. (Here $f^{-1}(x)$ denotes the preimage of x.)	*2pts*
●	○	Let A and B be events on the same probability space, and suppose $\mathbb{P}[A] = \frac{3}{4}$ and $\mathbb{P}[B] = \frac{1}{2}$. Then it must be the case that $\frac{1}{4} \le \mathbb{P}[A \cap B] \le \frac{1}{2}$.	*2pts*
○	●	Consider a fair coin and a biased coin. One of the two coins is chosen at random and tossed twice. The outcomes of the two tosses are independent.	*2pts*

2. Short Answers [*Answer is a single number or expression;* **write it in the box provided: anything outside the box will not be graded;** *no justification necessary. 3 points per answer; total of 45 points. No penalty for incorrect answers.*]

(a) Suppose that $P(x)$ is a real polynomial of degree 2 that has zeros at $x = 0$ and $x = 2$ and also passes through the point $(1, 1)$. What is the value of $P(3)$? *3pts*

-3. [Since P is of degree 2 and has zeros at 0 and 2, it must be of the form $P(x) = cx(x-2)$. Furthermore, $P(1) = 1$ implies $c = -1$. Hence, $P(3) = -3(3-2) = -3$.]

(b) How many polynomials over $\mathrm{GF}(17)$ of degree at most 5 pass through the points $(0, 0)$, $(1, 3)$ and $(2, 5)$? (You may leave your answer as an unevaluated expression.) *3pts*

17^3. [A polynomial of degree at most five is fully determined by 6 points. Three are given, so we need to specify three more. For each of these three points, there are 17 possible values our polynomial can take.]

(c) A message consisting of $n = 3$ packets, each of which is an integer mod 11, is transmitted over an unreliable channel. We use the Berlekamp-Welch encoding scheme (over $\mathrm{GF}(11)$) to protect against $k = 1$ error. The $n + 2k = 5$ packets received are $(1, 1)$, $(2, 2)$, $(3, 0)$, $(4, 2)$, $(5, 8)$. After running the interpolation procedure, we recover the error polynomial $E(x) = x - 1$ and the product polynomial $Q(x) = P(x)E(x) = 2x^3 + 8x^2 + 8x + 4$. Answer the following two questions.

(i) Which one of the five received packets was (possibly) corrupted? *3pts*

$(1, 1)$. [By definition, $E(x) = x - e_1$ where e_1 is the x-value of the possibly corrupted packet.]

(ii) What was the original value of the corrupted packet? *3pts*

$(1, 8)$. [From polynomial division, we find that $P(x) = Q(x)/E(x) \equiv 2x^2 - x + 7 \pmod{11}$, and hence $P(1) \equiv 2 - 1 + 7 \equiv 8 \pmod{11}$.]

(d) Alice and Bob are playing poker using a standard deck of cards. How many ways are there to deal 5 cards each to Alice and Bob (where the order of the cards does not matter)? *3pts*

$\binom{52}{5}\binom{47}{5}$, $\binom{52}{10}\binom{10}{5}$, or the so-called multinomial coefficient $\binom{52}{42,5,5} = \frac{52!}{42!5!5!}$. [We first choose 5 cards for Alice, and then choose 5 out of the remaining cards for Bob. Alternatively, we first choose 10 cards from the deck, and then choose five of those selected cards for Alice and give the remaining five to Bob.]

(e) How many permutations of $\{1, \ldots, n\}$ are there with exactly k fixed points, where $1 \leq k < n$? Express your answer in terms of D_m, the number of derangements of $\{1, \ldots, m\}$. *3pts*

$\binom{n}{k} D_{n-k}$. [We first need to choose the k fixed points, and then permute the remaining $n - k$ points so that none is mapped to itself.]

[Q2 continued on next page]

(f) Suppose a random number generator returns a number in $\{0, 1, \ldots, 9\}$ with uniform probability, and you run it 100 times to generate a 100-digit number (possibly with leading zeros).

(i) What is the probability that the 100-digit number contains the digit 7 more than 50 times? (You should leave your answer as a summation.) *3pts*

$\sum_{k=51}^{100} \binom{100}{k} \left(\frac{1}{10}\right)^k \left(\frac{9}{10}\right)^{100-k}$. [Each run of the random number generator samples 7 with probability $\frac{1}{10}$ and some number other than 7 with probability $\frac{9}{10}$, so the total number of 7's in 100 trials is distributed as a Binomial$(100, \frac{1}{10})$ random variable.]

(ii) What is the probability that the 100-digit number contains either no 0-digits or no 1-digits? (You should leave your answer as an unevaluated expression.) *3pts*

$2 \cdot \left(\frac{9}{10}\right)^{100} - \left(\frac{8}{10}\right)^{100}$. [Let A and B respectively denote the event of observing no 0-digits and no 1-digits. Then, by the inclusion-exclusion principle, we obtain $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$. The answer now follows from $\mathbb{P}[A] = \mathbb{P}[B] = \left(\frac{9}{10}\right)^{100}$ and $\mathbb{P}[A \cap B] = \left(\frac{8}{10}\right)^{100}$.]

(g) You have a fair 6-sided die, and also a loaded 6-sided die that shows 6 with probability $1/2$ and $1, 2, 3, 4, 5$ with probability $1/10$ each. One of the two dice is chosen uniformly at random and rolled once. Let A be the event that 6 is observed.

(i) What is $\mathbb{P}[A]$? *3pts*

$\frac{1}{3}$. [Let B be the event that the biased die is chosen. Then, $\mathbb{P}[A] = \mathbb{P}[A|B]\mathbb{P}[B] + \mathbb{P}[A|\overline{B}]\mathbb{P}[\overline{B}] = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{3}$.]

(ii) What is $\mathbb{P}[\,\text{Loaded die was chosen} \mid A\,]$? *3pts*

$\frac{3}{4}$. $\left[\mathbb{P}[B \mid A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]} = \frac{\frac{1}{2}\frac{1}{2}}{\frac{1}{3}} = \frac{3}{4}.\right]$

(h) Let $X \sim$ Bernoulli$(\frac{1}{2})$ and let Y be a random variable with probability distribution $\mathbb{P}[Y = 1] = \frac{1}{2}$ and $\mathbb{P}[Y = 2] = \frac{1}{2}$. Assuming that X, Y are independent, find $\mathbb{P}[X < Y]$. *3pts*

$\frac{3}{4}$. [The event $(X < Y)$ is the union of the event $(Y = 2)$ and the event $(Y = 1) \cap (X = 0)$. By independence, $\mathbb{P}[(Y = 1) \cap (X = 0)] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Also, the events $(Y = 2)$ and $(Y = 1) \cap (X = 0)$ are disjoint. Hence, $\mathbb{P}[X < Y] = \mathbb{P}[Y = 2] + \mathbb{P}[(Y = 1) \cap (X = 0)] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.]

(i) Let X be a random variable with probability distribution $\mathbb{P}[X = -8] = \frac{1}{4}$, $\mathbb{P}[X = -4] = \frac{1}{2}$, $\mathbb{P}[X = 4] = \frac{1}{4}$. Find $\mathbb{E}[\,2X]$. *3pts*

-6. $\left[\mathbb{E}[2X] = 2\mathbb{E}[X] = 2 \cdot \left(-8 \cdot \frac{1}{4} - 4 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4}\right) = -6.\right]$

[Q2 continued on next page]

(j) Suppose m balls are thrown uniformly at random into n bins (one ball at a time). What is the expected number of bins that have exactly one ball in them? *3pts*

$m\left(1-\frac{1}{n}\right)^{m-1}$. [Let I_k be the indicator that bin k has exactly one ball. Then, $\mathbb{E}[\sum_{k=1}^{n} I_k] = \sum_{k=1}^{n} \mathbb{P}[I_k = 1] = \sum_{k=1}^{n} m \cdot \frac{1}{n}\left(1-\frac{1}{n}\right)^{m-1}$.]

(k) The problem *Finite* takes as input a program P and decides whether the set of inputs on which P loops forever is finite (or empty). The following pseudo-code gives a reduction from the Halting Problem, *Halt*, to *Finite*. Fill in the blanks to make the reduction behave correctly. *3pts*

```
Test-Halt(P,x)
      let P' be a program that, on every input, runs P on x
      if Test-Finite(P') then return "yes"
            else return "no"
```

[P' halts on all inputs if and only if $P(x)$ halts, so if `Test-Finite(P')` returns "yes", then $P(x)$ halts, and if `Test-Finite(P')` returns "no", then $P(x)$ does not halt.]

(l) We say that programs P_1, P_2 *differ* on input x if one of the programs halts and the other loops forever on x. The problem *InfDiff* takes as input two programs, P_1 and P_2, and decides whether there are infinitely many inputs x on which P_1, P_2 differ. The following pseudo-code gives a reduction from the Halting Problem, *Halt*, to *InfDiff*. Fill in the blank to make the reduction behave correctly. *3pts*

```
Test-Halt(P,x)
      let P1 be a program that, on every input, runs P on x
      let P2 be a program that, on every input, loops forever
      if Test-InfDiff(P1,P2) then return "yes" else return "no"
```

[If $P(x)$ loops forever then P_1 loops forever on all inputs, and if $P(x)$ halts then P_1 halts on all inputs. Since P_2 loops forever on all inputs, P_1 and P_2 will differ on infinitely many inputs if and only if $P(x)$ halts.]

3. Modifying Secrets [*All answers to be justified. Total of 12 points.*]

Alice sets up a secret sharing scheme with her n friends $\text{Bob}_1, \ldots, \text{Bob}_n$, in which each Bob_i gets a point (x_i, y_i) where $y_i = P(x_i)$ for a fixed polynomial P over a field $\text{GF}(q)$, where $q > 2n$. The secret is kept at $P(0) = s$. When Alice is distributing these points to the Bobs, an adversary Eve can tamper with the points, and thus change the value of the secret that will be recovered. For each scenario in (a)–(c) below, give the value of the new secret that will be recovered; your answers may depend on s or on P. In each case, prove that your answer is correct.

(a) Eve replaces each point (x_i, y_i) with $(x_i, 2y_i + 1)$. *4pts*

The recovered secret is $\boxed{2s + 1 \pmod q}$: In the original scheme, a subset of k of the Bobs would reconstruct the unique polynomial $P(x)$ of degree at most $k - 1$ passing through k of the points (x_i, y_i), for some k. After tampering, the Bobs reconstruct instead the unique polynomial $Q(x)$ (over $\text{GF}(q)$) of degree at most $k - 1$ that passes through the points $(x_i, 2y_i + 1)$; and we can see that $Q(x) = 2P(x) + 1$ since both of these polynomials have degree at most $k - 1$ and agree on k points: $Q(x_i) = 2y_i + 1 = 2P(x_i) + 1$; so they must be the same polynomial. To recover the secret, the Bobs compute $Q(0) = 2P(0) + 1 = 2s + 1$.

An alternative (more complicated) approach is to use Lagrange interpolation. Assuming w.l.o.g. that the Bobs who are doing the reconstruction are labeled $i = 1, \ldots, k$, they will reconstruct a polynomial Q of degree at most $k - 1$ via

$$Q(x) = \sum_{i=1}^{k} (2y_i + 1)\Delta_i(x), \tag{1}$$

where as usual the basis polynomials are given by $\Delta_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}$. Now notice that (1) can be written

$$Q(x) = 2\sum_{i=1}^{k} y_i \Delta_i(x) + \sum_{i=1}^{k} \Delta_i(x) = 2P(x) + \sum_{i=1}^{k} \Delta_i(x), \tag{2}$$

where we have noticed that the first sum is exactly the Lagrange interpolation formula for P itself. Finally, notice that the polynomial $\sum_{i=1}^{k} \Delta_i(x)$ has degree at most $k - 1$ (since Q has), and takes the value 1 at all k points $x_1, \ldots, x_k$ (since at those points $Q(x_i) = 2P(x_i) + 1$). Hence this polynomial must be the constant polynomial 1. We can now deduce from (2) that $Q(x) = 2P(x) + 1$, from which the Bobs get the new secret $Q(0) = 2s + 1$.

(b) Eve replaces each point (x_i, y_i) with $(2x_i, y_i)$. *4pts*

The recovered secret is $\boxed{s}$: Write the polynomial $P(x) = a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \cdots + a_1 x + a_0$, and note that $a_0 = P(0) = s$. Recall that this polynomial passes through the points (x_i, y_i), so $P(x_i) = y_i$. Now consider the polynomial $Q(x) = 2^{-(k-1)}a_{k-1}x^{k-1} + 2^{-(k-2)}a_{k-2}x^{k-2} + \cdots + 2^{-1}a_1 x + a_0$, where all inverses are mod q. We claim that this polynomial passes through Eve's new points $(2x_i, y_i)$: to see this, just substitute $x = 2x_i$ to get $Q(2x_i) = P(x_i) = y_i$. Hence this is the unique polynomial of the same degree as P through these points, and so it must be the polynomial reconstructed by the Bobs. The new secret value is then $Q(0) = a_0 = s$.

Again, it is possible (but messier) to approach this via Lagrange interpolation. Unlike in part (a), where the basis polynomials Δ_i are unchanged from those for the original polynomial $P(x)$ (because the x_i values are the same), now the Bobs are using new basis polynomials as follows:

$$\Delta_i(x) = \frac{\prod_{j \neq i}(x - 2x_j)}{\prod_{j \neq i}(2x_i - 2x_j)}.$$

These polynomials are quite complicated, but when they are evaluated at $x = 0$ (the secret point) they become much simpler:

$$\Delta_i(0) = \frac{\prod_{j \neq i}(-2x_j)}{\prod_{j \neq i}(2x_i - 2x_j)} = \frac{\prod_{j \neq i}(-x_j)}{\prod_{j \neq i}(x_i - x_j)}.$$

And these are exactly the same as the basis polynomials for $P(x)$ itself, when evaluated at $x = 0$! Moreover, the values y_i are unchanged. Hence, at the point $x = 0$, the new polynomial $Q(x)$ obtained by the Bobs will have exactly the same value as $P(x)$, i.e., $Q(0) = P(0) = s$.

(c) Eve replaces each point (x_i, y_i) with $(x_i - 1, y_i)$. [You may assume that $x_i \neq 1$ for all $i = 1, \ldots, n$.] *4pts*

The new secret is $\boxed{P(1)}$: We claim that now the polynomial reconstructed by the Bobs is $Q(x) = P(x + 1)$. To see this, note that $Q(x)$ is a polynomial of degree at most $k - 1$ (since P is), and that it passes through k points of the form $(x_i - 1, y_i)$ held by the Bobs, since $Q(x_i - 1) = P(x_i) = y_i$. Hence this $Q(x)$ must be the polynomial reconstructed by the Bobs, and so they evaluate the secret as $Q(0) = P(1)$.

4. Breaking RSA [*All parts to be briefly justified. Total of 16 points.*]

(a) Alice sends the same message, $m < N$, to two friends, Bob and Carol using the standard RSA protocol discussed in class. Bob's public key is (N, e_1) and Carol's is (N, e_2), where $\gcd(e_1, e_2) = 1$. Explain how an eavesdropper, Eve, can decrypt m by observing the encrypted messages that Alice sends to Bob and Carol. [Hint: Use the extended gcd algorithm.] *4pts*

Since $\gcd(e_1, e_2) = 1$, Eve can use the extended gcd algorithm to efficiently find $a, b \in \mathbb{Z}$ such that $ae_1 + be_2 = 1$. Observing m^{e_1} and m^{e_2}, she can then proceed to compute $(m^{e_1})^a \cdot (m^{e_2})^b \equiv m^{ae_1} \cdot m^{be_2} \equiv m^{ae_1+be_2} \equiv m \pmod{N}$.

(b) Dennis decides to simplify the RSA cryptosystem as follows. Instead of choosing the usual type of public key $(N = pq, e)$, he instead chooses a key (N, e) where N is a prime and e is an integer in $\{2, \ldots, N-1\}$ with $\gcd(N-1, e) = 1$. To send a message m to Dennis, Alice sends the encrypted message $m^e \pmod{N}$. Is Dennis' scheme secure? Explain your answer. *4pts*

No, Dennis' scheme is not secure. Since e is relatively prime to $N-1$, Eve can efficiently compute the inverse $e^{-1} \bmod (N-1)$ (using the extended gcd algorithm). Furthermore, since $ee^{-1} \equiv 1 \pmod{N-1}$, it must be the case that $ee^{-1} = k(N-1) + 1$ for some $k \in \mathbb{Z}$, and consequently Eve can compute $(m^e)^{e^{-1}} \equiv m^{ee^{-1}} \equiv m^{k(N-1)+1} \equiv \left(m^{N-1}\right)^k \cdot m \equiv m \pmod{N}$, where the last congruence follows from Fermat's Little Theorem.

[Q4 continued on next page]

(c) Frank has published his RSA public key $(N = pq, e)$. Gina wants to construct her own public key, and has found one large prime $p' \neq p$; being too lazy to find another one, she asks if she can use one of Frank's; since Frank trusts Gina, he gives her his prime q. Gina then publishes her key $(N' = p'q, e')$. Explain how Eve is able to break both Frank's and Gina's RSA schemes. *4pts*

Since $p' \neq p$, Eve can run Euclid's algorithm to compute $\gcd(N, N') = q$, and then obtain $p = N/q$ and $p' = N'/q$. Knowing p, p' and q, she can then proceed to compute Frank's and Gina's private keys d and d' through $d \equiv e^{-1} \pmod{(p-1)(q-1)}$ and $d' \equiv (e')^{-1} \pmod{(p'-1)(q-1)}$ (again using the extended gcd algorithm).

(d) Harry, Imogen and Jasper have RSA public keys $(N_H, 3)$, $(N_I, 3)$ and $(N_J, 3)$ respectively, where N_H, N_I, N_J are all distinct. Alice sends the same message m (where m is less than all of N_H, N_I, N_J) to all three of them using their respective keys. Explain how Eve is able to decrypt this message by observing the three encrypted messages. [Hints: You may use without proof the Chinese Remainder Theorem, which says the following: *Let n be a natural number. Given the values $c_i = n \bmod r_i$, for $1 \leq i \leq k$, where the r_i are coprime, we can efficiently compute the value $c = n \bmod (r_1 r_2 \dots r_k)$.* You may also use the fact that the cube root of an integer can be found efficiently.] *4pts*

If N_I, N_H and N_L are not pairwise coprime, then Eve can apply her results from part (c) to find m, so let us assume that N_I, N_H, N_L are pairwise coprime. Letting m_I, m_H, m_L denote the three encrypted messages Eve observes, she may apply the Chinese remainder theorem on

$$\begin{aligned} m_I &= m^3 \mod N_I, \\ m_H &= m^3 \mod N_H, \\ m_L &= m^3 \mod N_L, \end{aligned}$$

to efficiently obtain $m^3 \mod (N_I \cdot N_H \cdot N_L)$. Since we are given that $m < \min(N_I, N_H, N_L)$, we know that $m^3 < N_I \cdot N_H \cdot N_L$, and hence $m^3 \mod (N_I \cdot N_H \cdot N_L)$ is actually equal to m^3 (as integers—no mod!). So, all Eve needs to do is take a cube root over the integers, which we are told can be done efficiently. (This is actually not too hard to see: e.g., we can perform a binary search, at each iteration checking whether the current element cubed is m^3 or not; the binary search takes at most $\log(N_I N_H N_L)$ iterations, and each multiplication takes $O(\log^2(N_I N_H N_L))$ steps. Note in contrast that computing cube roots mod p has no known efficient algorithm!)

5. Counting Team Compositions [*All parts to be briefly justified. Total of 16 points.*]

UC Berkeley has n students who are interested in participating in both the CS Programming and the Putnam Mathematical competitions. In parts (a) and (b), count the number of ways to choose a CS team with c members and a Putnam team with p members under the specified constraint, assuming that $n \geq c + p$.

Whenever possible, express your answers in terms of binomial coefficients and **clearly indicate how each coefficient arises.**

(a) No restriction; any student can be on both teams. *4pts*

There are exactly $\binom{n}{c}$ ways of choosing the CS team, and $\binom{n}{p}$ ways of choosing the Putnam team. Since neither choice affects the other, we have an overall number of

$$\binom{n}{c}\binom{n}{p}$$

possibilities to choose both a CS team and a Putnam team.

(b) No student can be on both teams. *4pts*

There are $\binom{n}{c}$ possibilities to assemble the CS team. The Putnam team now has to be formed from the remaining $n - c$ students, which we can do in $\binom{n-c}{p}$ ways. Consequently, there are a total of

$$\binom{n}{c}\binom{n-c}{p} \qquad \text{or equivalently} \qquad \binom{n}{p}\binom{n-p}{c}$$

ways of choosing a CS team and a Putnam team without any overlap. Other correct solutions include $\binom{n}{c+p}\binom{c+p}{c}$ and $\binom{n}{c+p}\binom{c+p}{p}$, which correspond to first selecting $c + p$ students, and then dividing them into a CS team and a Putnam team.

(c) Let $A_{n,c,p}$ denote the number of ways to choose the teams as in part (a), and let $B_{n,c,p}$ denote the number of ways to choose the teams as in part (b). Fill in the blank in the equation below to produce a valid combinatorial identity, **and briefly explain your reasoning**. *4pts*

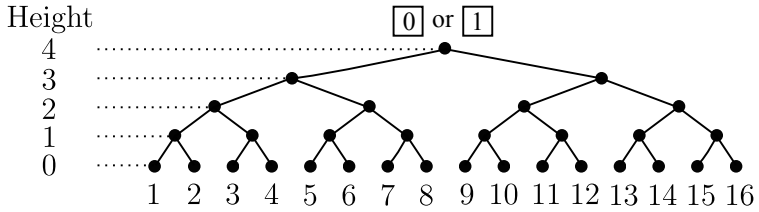
$$A_{n,c,p} = \sum_{j=0}^{\min(c,p)} \binom{n}{j} B_{n-j,c-j,p-j}$$

In order to choose CS and Putnam teams with overlap, we may first choose j students who should be on both teams, and then out of the remaining $n - j$ students assign $c - j$ and $p - j$ to the respective teams without overlap.

(d) The CS team wins an international programming contest and receives k gold coins as a prize. How many ways are there to divide up the coins among the c team members, under the condition that each student should receive at least three coins? (Assume that $k > 3c$.) *4pts*

After we have assigned each member 3 coins, we are left with $k - 3c$ coins that we have to distribute among the students on the CS team. This is a balls and bins problem with $n = k - 3c$ balls and $k = c$ bins, for which we know the answer to be $\binom{k-2c-1}{c-1}$.

6. Noisy Transmission on a Binary Tree [*All parts to be justified. Total of 13 points.*]



Consider a binary tree of depth D, which has 2^D leaves; in the example shown above, $D = 4$. At the root of the tree is a single bit (0 or 1). This bit is transmitted separately to each of the two children of the root, but in each case the value of the bit is *flipped* (i.e., $0 \to 1$ or $1 \to 0$) independently with probability p. This process continues, with each vertex of the tree transmitting its bit value (flipped with probability p, independently of all other transmissions) to its two children. The process stops at the leaves.

Suppose two distinct leaves a, b are chosen uniformly at random. Let T denote the height of the lowest common ancestor (LCA) of a and b; in the example above, $T = 1$ for $(a, b) = (1, 2)$, while $T = 3$ for $(a, b) = (1, 8)$.

(a) Find a formula for the probability $\mathbb{P}[T = t]$, where $t = 1, \ldots, D$, for a general binary tree of depth D. *5pts*

Let Ω be the set of all pairs of leaves and $E_t \subset \Omega$ the event that $T = t$. Then $\mathbb{P}[T = t] = \mathbb{P}[E_t] = |E_t|/|\Omega|$. Now in order for E_t to happen, a and b must have an LCA at height t. Let us denote by $\mathcal{H}_t$ the set of vertices at height t, and by A_v the event that a and b have LCA $v \in \mathcal{H}_t$. Then $E_t = \cup_{v \in \mathcal{H}_t} A_v$, and since the A_v are mutually disjoint, $|E_t| = \sum_{v \in \mathcal{H}_t} |A_v| = |\mathcal{H}_t||A_w| = 2^{D-t}|A_w|$, where w is any vertex in $\mathcal{H}_t$ and the second equality follows from the fact that $|A_v| = |A_{v'}|$ for any $v, v' \in \mathcal{H}_D$ by symmetry. To compute $|A_w|$ we note that w subtends two sub-trees of size 2^{t-1} each, from which it follows that there are exactly $2^{t-1} \cdot 2^{t-1} = 2^{2t-2}$ pairs of leaves whose LCA is w. So $|E_t| = 2^{D-t} \cdot 2^{2t-2} = 2^{D+t-2}$, which combined with $|\Omega| = 2^{D-1}\left(2^D - 1\right)$ yields

$$\mathbb{P}[T = t] = \frac{2^{D-1} \cdot 2^{t-1}}{2^{D-1}\left(2^D - 1\right)} = \frac{2^{t-1}}{2^D - 1}.$$

Alternative Solution: Instead of thinking of $\{a, b\}$ as an unordered pair like we did above, we may think of it as the ordered pair (a, b). Then conditional on fixing $a = i \in \{1, \ldots, 2^D\}$, looking at the unique vertex $v \in \mathcal{H}_t$ that subtends a, we see that in order for $T = t$, leaf b must be inside the sub-tree subtended by v that does not contain a. That sub-tree has exactly 2^{t-1} leaves, and so $\mathbb{P}[T = t \mid a = i] = \frac{2^{t-1}}{2^D - 1}$. Then by the law of total probability:

$$\mathbb{P}[T = t] = \sum_{i=1}^{2^D} \mathbb{P}[(T = t) \cap (a = i)] = \sum_{i=1}^{2^D} \mathbb{P}[T = t | a = i]\mathbb{P}[a = i] = \frac{2^{t-1}}{2^D - 1} \sum_{i=1}^{2^D} \mathbb{P}[a = i] = \frac{2^{t-1}}{2^D - 1}.$$

(b) Let M denote the total number of times the bit is flipped when traversing the tree from the LCA to leaf a **plus** the analogous number of bit flips from the LCA to b. Write down a formula for the conditional probability $\mathbb{P}[M = m \mid T = t]$. Be sure to specify what values of m are possible. *3pts*

On the event E_t, a and b are each connected to their LCA by a path of length exactly t. That is, there are a total of $2t$ distinct edges along which bit flips may contribute to M. Let us denote the set of these edges by $\mathcal{E}_{a,b}$, and call $F_e, e \in \mathcal{E}_{a,b}$ the indicator variable that is 1 if a bit flip occurred across edge e and 0 otherwise. Then $M = \sum_{e \in \mathcal{E}_{a,b}} F_e$, and since the F_e are independent Bernoulli(p) variables,

the conditional distribution of M given $T = t$ is Binomial$(2t, p)$. So,

$$\mathbb{P}[M = m \mid T = t] = \binom{2t}{m} p^m (1-p)^{2t-m}, \tag{3}$$

for $m = 0, 1, \ldots, 2t$.

(c) Now let the random variables B_a and B_b respectively denote the bits observed at the sampled leaves a and b. What condition on M guarantees that $B_a = B_b$? *2pts*

If $B_a = B_b$, then either B_a and B_b both differ from the bit at $\mathrm{LCA}(a, b)$, or they are both equal to the bit at $\mathrm{LCA}(a, b)$. The former case can happen if only if there are an odd number of flips on the path between a and $\mathrm{LCA}(a, b)$, and also an odd number of flips on the path between b and $\mathrm{LCA}(a, b)$. The latter case occurs if and only if there are an even number of bit flips on both paths. Therefore $B_a = B_b$ if and only if M is even.

(d) Find a formula for $\mathbb{P}[B_a = B_b]$. You may leave your answer as an unevaluated expression. [Hints: Use parts (b) and (c) to compute $\mathbb{P}[B_a = B_b \mid T = t]$, then combine with part (a).] *3pts*

By the law of total probability and part (c), we know that

$$\mathbb{P}[B_a = B_b] = \sum_{t=1}^{D} \mathbb{P}[T = t] \cdot \mathbb{P}[M \text{ is even } \mid T = t].$$

Plugging in our results from part (a) for $\mathbb{P}[T = t]$ and Equation (3) for $\mathbb{P}[M = m \mid T = t]$, we thus obtain

$$\mathbb{P}[B_a = B_b] = \sum_{t=1}^{D} \frac{2^{t-1}}{2^D - 1} \left[\sum_{k=0}^{t} \binom{2t}{2k} p^{2k} (1-p)^{2(t-k)} \right]. \tag{4}$$

Aside (**For interested students only; not needed to receive full credit**): We can actually compute $\mathbb{P}[B_a = B_b]$ in closed form. First, noting

$$[(1-p) + p]^{2t} = \sum_{k=0}^{t} \binom{2t}{2k} p^{2k} (1-p)^{2(t-k)} + \sum_{k=0}^{t-1} \binom{2t}{2k+1} p^{2k+1} (1-p)^{2(t-k)-1}, \tag{5}$$

$$[(1-p) - p]^{2t} = \sum_{k=0}^{t} \binom{2t}{2k} p^{2k} (1-p)^{2(t-k)} - \sum_{k=0}^{t-1} \binom{2t}{2k+1} p^{2k+1} (1-p)^{2(t-k)-1}, \tag{6}$$

we see that adding (5) and (6), and then dividing by 2 gives

$$\mathbb{P}[M \text{ is even } \mid T = t] = \frac{1}{2}[1 + (1-2p)^{2t}].$$

Then, using this result, (4) can be simplified as

$$\mathbb{P}[B_a = B_b] = \frac{1}{2}\left[1 + \frac{x(x^D - 1)}{2(x-1)(2^D - 1)}\right],$$

where $x := 2(1-2p)^2$. See Figure 1 for illustration of $\mathbb{P}[B_a = B_b]$ as a function of p. For all D, note that $\mathbb{P}[B_a = B_b] = \frac{1}{2}$ at $p = \frac{1}{2}$, while $\mathbb{P}[B_a = B_b] = 1$ at $p = 0$ and $p = 1$. As D increases, the interior

region with $\mathbb{P}[B_a = B_b] = \frac{1}{2}$ enlarges, and transitions from $\mathbb{P}[B_a = B_b] = \frac{1}{2}$ to $\mathbb{P}[B_a = B_b] = 1$ near $p = 0$ and $p = 1$ become sharper and sharper. This makes intuitive sense, since T grows linearly with D in expectation $\left(\text{more precisely, } \mathbb{E}[T] = D - 1 - \frac{D}{2^{D-1}}\right)$, and when tossing a coin sufficiently many times, regardless of how biased it is, the chance of observing an even number of heads is roughly the same as the chance of observing an odd number of them.

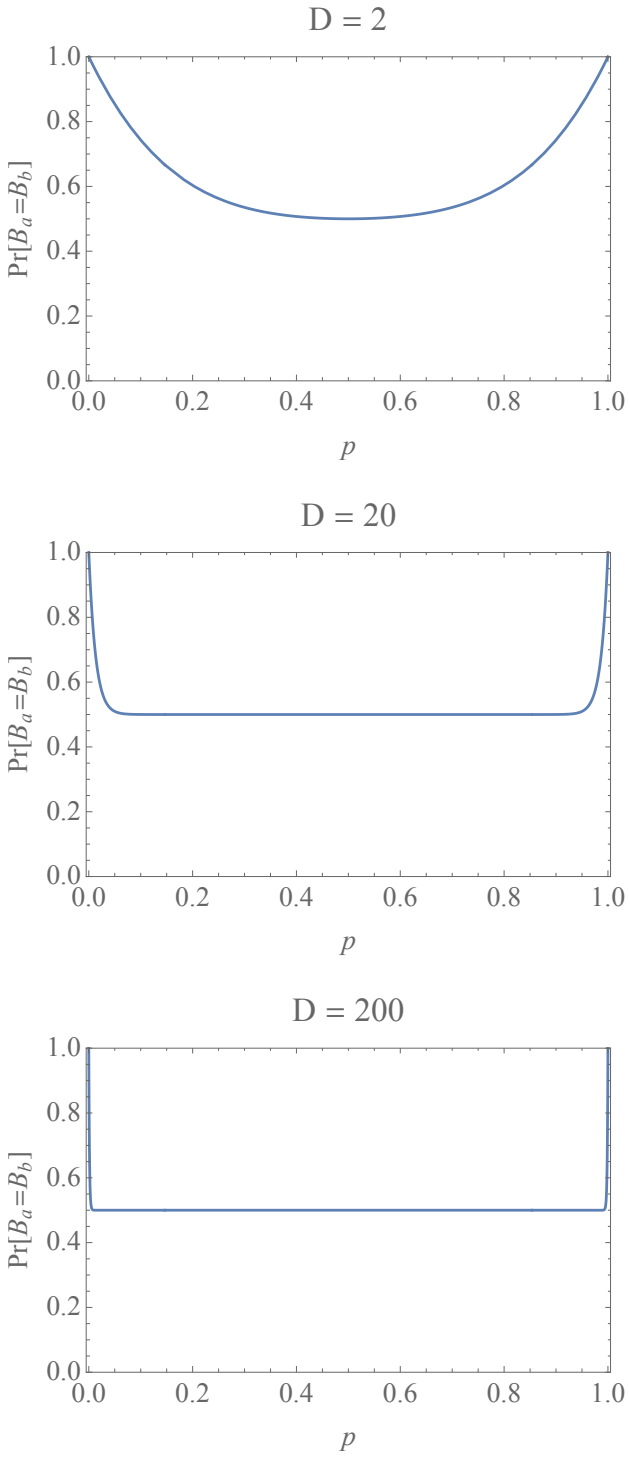


Figure 1: Plots of $\mathbb{P}[B_a = B_b]$ as a function of p for $D = 2, 20$, and 200.