CS 70 Discrete Mathematics and Probability Theory
Fall 2021 Ayazifar and Rao

Final Solutions

PRINT Your Name: Oski Bear

SIGN Your Name: $\mathscr{O}\mathscr{S}\mathscr{K}\mathscr{I}$

Do not turn this page until your instructor tells you to do so.

1. Pledge.

Berkeley Honor Code: As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.

In particular, I acknowledge that:

- I alone am taking this exam. Other than with the instructor and GSIs, I will not have any verbal, written, or electronic communication about the exam with anyone else while I am taking the exam or while others are taking the exam.
- I will not have any other browsers open while taking the exam.
- I will not refer to any books, notes, or online sources of information while taking the exam, other than what the instructor has allowed.
- I will not take screenshots, photos, or otherwise make copies of exam questions to share with others.
- I will make an effort to be polite to people from Stanford. [This is optional.]

Signed:______________

2. Long Ago.

1. $\forall x, y \in \mathbb{N}, x - y \leq \min(x, y)$.
   Answer: False. $x = 5, y = 1$.
2. $(\neg(\exists x \in S, (\neg P(x) \vee Q(x)))) \equiv (\forall x \in S, (P(x) \wedge (\neg Q(x)))$.
   Answer: True. Application of DeMorgan's theorem.
3. $\forall p \in \mathbb{N}, ((\forall n \in \mathbb{N}, (n < 2 \vee n = p \vee \neg(n|p)) \implies (\forall a \in \mathbb{N}, a^p - a = 0 \pmod{p})))$.
   Answer: True. This follows from Fermat's theorem and is used to proved that RSA encryption/decryption is reconstructs original message.
4. If $x|y$ and $x|z$, then $x|\gcd(y,z)$.
   Answer: True. Assume $x > 0$. If $x|y$ and $x|z$, then x must be a common divisor of y and z, and therefore divides any expression $ax + by$ for integers a and b. Moreover, $\gcd(y,z) = ax + by$ by the extended euclid's algorithm, and therefore $x|\gcd(y,z)$.
5. Let A and B be finite sets, and let $f : A \to B$ and $g : B \to A$ be two functions such that
   $$\forall x \in B, \quad f(g(x)) = x.$$
   Fill in the blank with one of $>, \geq, <, \leq, =$ or "Incomparable" in the statement
   $$|A| \;\_\_\; |B|.$$
   Choose the most specific answer that is always true.
   Answer: $\geq$. This is true for any function. If $|A| < |B|$, then g maps at least two elements in B to the same element in A, and a function f cannot map one element of A to different elements of B. It could be a bijection where $|A| = |B|$, and it could not be a bijection so $|A| > |B|$.
6. For natural numbers k and n, any solution to $x^k = n$ must be an integer or irrational.
   Answer: True. Consider reduced form: (a/b) where $b \neq 1$. Since a and b have no common factors, (a^k/b^k) also cannot be reduced if $b \neq 1$. Thus, either $b = 1$ in which case the solution is an integer, or there are no rational solutions.

7. For a natural number n which is *not* a perfect square, then $x^3 = n$ has no rational solutions.
   Answer: False. n could be a perfect cube.
8. For a natural number n which is *not* a perfect square, then $x^4 = n$ has no rational solutions.
   Answer: True. Otherwise x^2 would be a solution to the equation $y^2 = n$ which contradicts the assumption that n is not a perfect square.
9. Consider job optimal and candidate optimal stable pairings, P and P', in a stable matching instance, consider a graph G where vertices consist of jobs and candidates and the pairs in P and P' form edges.
   (a) The resulting graph is simple.
   Answer: False. The job optimal and candidate optimal pairings could be the same.
   (b) If the graph consists of a single simple cycle, then there are at most two stable pairings.
   Answer: False.
   Consider that the job(candidate)-optimal pairing is one where every job (candidate) gets their favorite partner and together these pairings form a simple cycle. Consider that the second favorite candidate on each entities list gives a pairing.
   We construct an example where an instance with n jobs and candidates. Let job i's favorite be candidate i, and candidate i's favorite be job $i+1 \pmod n$. Moreover, job i's second favorite candidate is $i-2 \pmod n$ and candidate i's second favorite candidate is job $i+2 \pmod n$ since $i+2 \pmod n$ is a bijection and its inverse is $i-2 \pmod n$.
   Notice also that matching is stable since for job i and candidate job i prefers $i-2 \pmod n$ to any candidate $j \neq i$, and candidate i prefers job $i+2 \pmod n$ to job i since its favorite is $i+1 \pmod 2$.
10. If no one gets rejected in the job propose stable matching algorithm, then:
   (a) Every job gets the first candidate on its list.
   Answer: True. Each job proposed only to its first candidate when the algorithm terminated.
   (b) The resulting pairing is *both* job optimal and candidate optimal.
   Answer: False. The candidates preferences lists didn't matter at all, so one can construct a different pairing which is candidate optimal by selecting a different job to be each candidate's favorite.
11. There are always at least two distinct stable pairings in any stable matching instance; the job optimal and the candidate optimal pairing.
   Answer: False. The job optimal and candidate optimal pairings could be the same.
12. What is the number of edges in an n-vertex simple graph where every vertex has degree 2?
   Answer: n. The sum of degrees is $2n$.
13. A simple graph with n vertices and c connected components, where each has degree exactly 2, has _______ simple cycles.
   Answer: c. Each component is a simple cycle.
14. Give as tight an upper bound as possible on the number of edges in a simple planar graph with n vertices and c connected components.
   Answer: $3n-6c$. A connected component with v vertices can have at most $3v-6$ edges and summing over components gives $3n-6c$.
15. How many paths of length d are there from the all 1's vertex to the all 0's vertex in a d-dimensional hypercube?
   Answer: $d!$. Every path of length d corresponds to a different ordering for flipping the bits (e.g. $11 \to 10 \to 00$ flips bit 2 then 1, whereas $11 \to 01 \to 00$ flips bit 1 then 2). There are d bits to flip, and in turn $d!$ different orders to flip them in.

3. More and more pi.

Prove *using induction* that the sum of the interior angles of a convex polygon with n sides (and n vertices) is $(n-2)\pi$. You may assume this is true for a triangle. (For a convex polygon, you may assume a line segment between "non-adjacent" vertices u and v splits the polygon into two convex polygons each containing u and v as vertices and are otherwise disjoint.)

Answer:

Base case: A polygon must have at least 3 sides so a triangle is our base case. Since the problem tells us we can the statement is true for a triangle, this concludes the base case.
Inductive Hypothesis: Assume that the sum of the interior angles of a convex polygon with k sides is $(k-2)\pi$ for all $3 \leq n \leq k$
Inductive Step: Let A be an arbitrary convex polygon with $k+1$ sides. We can take an three consecutive vertices in A and connect the two non-adjacent vertices. We've essentially split A into a convex k-polygon and a triangle. From our inductive hypothesis, we know that the sum of the interior angles of each will be $(k-2)\pi$ and π. Hence, the sum of the angles in A will be $(k-2)\pi + \pi = ((k+1)-2)\pi$ proving our claim.

4. Three, what's up with thee?

1. What are all the possible values of perfect squares modulo 3?
   Answer: Since $0^2 \equiv 0 \pmod 3$, $1^2 \equiv 1 \pmod 3$, and $2^2 \equiv 1 \pmod 3$, the perfect squares modulo 3 are 0 and 1.
2. Prove that any integer solution to the equation $x^2 + y^2 = 3xy$ must have $x \equiv y \equiv 0 \pmod 3$.
   Answer: Taking the equation modulo 3, we have that $x^2 + y^2 \equiv 0 \pmod 3$. x^2 and y^2 are either 0 or 1, so for their sum to be $0 \pmod 3$, we need them to be both 0. Thus, $x \equiv y \equiv 0 \pmod 3$.
3. Prove that the equation $x^2 + y^2 = 3xy$ has no solutions for positive integers x and y. (Hint: Suppose for contradiction that there was a solution (a,b), with a and b positive integers, where a was as small as possible. Find a positive integer solution (a',b') where $a' < a$. You may find part (b) helpful.)
   Answer: Suppose for contradiction that there was a solution (a,b), with a and b positive integers, where a was as small as possible. (This is possible because we are considering positive integer solutions). By part (b), we know that a and b are divisible by 3. Thus, $\frac{a}{3}$ and $\frac{b}{3}$ are positive integers, and
   $$\left(\frac{a}{3}\right)^2 + \left(\frac{b}{3}\right)^2 = \frac{1}{9}(a^2+b^2) = \frac{1}{9}(3ab) = 3\left(\frac{a}{3}\right)\left(\frac{b}{3}\right),$$
   so $(\frac{a}{3}, \frac{b}{3})$ is a solution to our equation, contradicting the minimality of a. Thus, our initial assumption was incorrect, and thus there are no positive integer solutions to our equation, as desired.

5. Polynomials.

1. Suppose $P(x)$, $Q(x)$ are distinct degree d polynomials. Then they have at most _____ intersections. Recall an intersection is a value x, where $P(x) = Q(x)$. (Give the tightest upper bound possible.)
   Answer: d. $P(x) - Q(x) = 0$ has at most d solutions.
2. Suppose polynomial $P(x)$ is of degree $2d$ and $Q(x)$ is of degree d. Then they have at most ___ intersections. Recall an intersection is a value x, where $P(x) = Q(x)$. (Give the tightest upper bound possible.)
   Answer: $2d$. $P(x) - Q(x) = 0$ has at most $2d$ solutions since it is a degree $2d$ polynomial.

3. Give a polynomial with roots at r_1, r_2 that has value v at r_3 modulo a prime p.
   Answer: $v(x-r_1)(x-r_2)(r_3-r_1)^{-1}(r_3-r_2)^{-1} \pmod{p}$. This is just a special case of LaGrange interpolation.
4. Every polynomial modulo a prime p is equivalent to a polynomial of degree at most ________. (Your answer should be as small as possible.)
   Answer: $p-1$. One can use the identity that $a^p = a \pmod{p}$ to reduce the degree of any term to less than p.
5. The Berlekamp-Welch algorithm is used to send a message of size $n = 3$ sent over a noisy channel, which possibly corrupts at most 2 packets.
   (a) How many packets does one send in this situation? (Looking for a number.)
   Answer: 7. Its $n+2k$.
   (b) If exactly 2 errors occur somwhere in the set of packets used, how many possible polynomials $Q(x)$ (working modulo a prime p) could one possibly reconstruct?
   Answer: $\binom{7}{2}$. There are $n+2k$ possible places for errors and there are exactly 2 places. Once the positions of the 2 errors are specfied there is a unique error polynomial that corresponds to $Q(x)$.
   (c) If the error polynomial is $E(x) = x^2+4x+2 \pmod 7$, where are the errors? (Give the x value(s) for your answer.)
   Answer: $x = 1$, and $x = 2$ since $x^2+4x+2 = (x-1)(x-2) \pmod 7$.

6. Some counting.

1. Sylvia has found seven different NFTs (non-fungible tokens or digitally stamped pictures), and plans to acquire them for herself by taking screenshots. She wishes to end up with ten screenshots in total. She does not necessarily need a screenshot of every one.
   (a) How many different sets of screenshots can she obtain? The order in which she takes the screenshots does not matter, and screenshots of the same NFT are indistinguishable.
   Answer: By stars and bars, there are $\binom{16}{6}$ ways for Sylvia to get screenshots.
   (b) (5 points) Sylvia decides she does not want to have more than four screenshots of the same NFT. How many ways could she take screenshots now?
   Answer: We subtract out by the situations where Sylvia ends up with at least four screenshots of the same NFT. Here, we proceed with PIE. Let the NFTS be labeled from 1 to 7. The number of situations where NFT i is screenshotted more than four times is $\binom{11}{6}$ by balls and bins, and there are 7 choices for i. However, it is possible that two different NFTs are both screenshotted more than four times. The number of situations where NFT i and j are both screenshotted, with $i \neq j$, is $\binom{6}{6} = 1$ by balls and bins, and there are $\binom{7}{2}$ choices for i and j. Thus, by PIE, the number of situations where Sylvia ends up with at least four screenshots of the same NFT is $7\binom{11}{6} - \binom{7}{2}$.
   Our final answer is thus $\boxed{\binom{16}{6} - 7\binom{11}{6} + \binom{7}{2}}$.
2. (5 points) How many ways are there to express 10 as the sum of positive integers, where the order of the sum matters? For example, $10 = 10$, $10 = 4+3+3$ and $10 = 3+3+4$ count as three different ways.
   Answer: 512. We may model the problem with 10 balls, and in each of the 9 spots between the 10 balls, we may choose whether or not to place a divider. Then the number of balls between dividers (or to the left of the leftmost divider/right of the rightmost divider) is a summand. There are 9 spots, and each has 2 choices, for an answer of $2^9 = \boxed{512}$.

Alternatively, one can do stars and bars where each term has at least 1. This gives the formula $\sum_{i=0} \binom{9}{i}$, which is 2^9 by the binomial theorem.

3. (5 points) How many positive factors of 1008 are also factors of 840? (Hint: Think of factors of gcd(840, 1008).)

   Answer: A factor of both 1008 and 840 is a factor of their greatest common divisor, or 168. The prime factorization of $168 = 2^3 \cdot 3^1 \cdot 7^1$. Any factor of this must be in the form $2^a \cdot 3^b \cdot 7^c$, where $0 \leq a \leq 3, 0 \leq b \leq 1$, and $0 \leq c \leq 1$. We have 4 choices for a, 2 choices for b, and 2 choices for c, for a total of $\boxed{16}$ factors.

7. Countability

1. The set of indices (or locations) of the 1's in the decimal representation of π. (For example, $\pi =$ 3.14159... so the set contains the indices 1 and 3 as there is a 1 at these locations in the decimal representation of π.)

   Answer: Countable. This is just a subset of the integers.

2. The set of subsets of a countably infinite set.

   Answer: Uncountable. This is a bijection with the powerset of the integers.

3. The set of finite sized subsets of a countably infinite set.

   Answer: Countable. One can enumerate the set of subsets of size at most k from the first k elements of the countable set for increasing k. Any finite set has a maximal element m and thus will be output when one is enumerating the subsets of the first m elements.

4. The set of all pairs of elements of two countably infinite sets.

   Answer: Countable. One can use the spiral trick we used to show the rationals are countable. Alternatively, each countable set can be ordered. Then we can enumerate pairs by listing them in an order corresponding to the sum of the indexes of the elements in their respective orderings. Every pair will appear in the list at a finite place.

8. Computability

1. Given a program P, an input x, and a number n, does the program P halt on x in n steps?

   Answer: Computable. Just run P on x for n steps.

2. Given a program P, an input x, and a number n, does the program P touch the memory location numbered n?

   Answer: Uncomputable. We can reduce the halting problem to this problem by modifying a program P to only write to memory locations above 1, unless it exits, in which case it writes to location 0. The old program halts if and only the new program writes to 0.

3. Given a program P, an input x, and a number n, does the program P touch more than n memory locations?

   Answer: Computable. Run the program and if it uses more then n units of memory, say yes, otherwise if the n locations have been touched repeat a configuration and line number say no. The number of states is finite, so if this state repeats if the program never repeats a configuration.

9. Probability: Based!

Given a probability space $(\Omega, \mathbb{P})$ and events A, B and C, fill in $\leq$, $<$, $=$, $\geq$, $>$, or "Incomparable" for questions 1-7 such that the statement always holds. If no inequality always holds use "Incomparable."

1. $\mathbb{P}[A \cup B \cup C]$ ______ $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C]$.

   Answer: $\leq$. This is the union bound. The right hand side accounts for each sample point in the union at least once.

2. $\mathbb{P}[A \cap B \cap C]$ _________ $\mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]$

   Answer: Incomparable. It could be less than the product if they are disjoint or more than the product if they are the same events.

3. $\mathbb{P}[A \cap B]$ _______ $\mathbb{P}[A]$.

   Answer: $\leq$. $A \cap B$ is a subset of A.

4. $\mathbb{P}[A|B]$ _______ $\mathbb{P}[A]$.

   Answer: Incomparable. If $A = B$, then $\mathbb{P}[A|B] = 1$. If A and B are disjoint $\mathbb{P}[A|B] = 0$

5. $\mathbb{P}[A \cup B \cup C]$ _____ $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] - \mathbb{P}[A \cap B] - \mathbb{P}[A \cap C] - \mathbb{P}[A \cap B]$.

   Answer: Incomparable. This is the inclusion-exclusion equality but with $P[B \cap C] - P[A \cap B] - P[A \cap B \cap C]$ added to the RHS, which could be positive or negative.

   Since the question was meant to test understanding of the inclusion-exclusion formula, we also gave credit to $\geq$.

   For the following, let X_A, X_B, and X_C be indicator random variables for events A, B, C.

6. If $\text{Cov}(X_A, X_B) > 0$, then $\mathbb{P}[A \cap B]$_____$\mathbb{P}[A]\mathbb{P}[B]$.

   Answer: $>$. $\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$. If $\text{Cov}(X_A, X_B) > 0$ then $\mathbb{P}[A|B] > \mathbb{P}[A]$.

7. If $\text{Cov}(X_A, X_B) > 0$ and $\text{Cov}(X_B, X_C) > 0$, then $\text{Cov}(X_A, X_C)$_____0.

   Answer: Incomparable. A, B, C could be the same event in which case $\text{Cov}(X_A, X_C) > 0$, or A and C could be disjoint events such that $B = A \cup C$ in which case $\text{Cov}(X_A, X_C) < 0$.

8. If $\mathbb{P}[A|B] = .6$ and $\mathbb{P}[A] = .5$ and $\mathbb{P}[B] = .5$, what is the linear function $f(X_B)$ that minimizes $\mathbb{E}[(X_A - f(X_B))^2]$?

   Answer: $.2X_B + .4$. $\text{Var}X_A = \text{Var}X_B = .25$ and $Cov(X_A, X_B) = (.05)$. Thus we get $f(X_B) = \frac{.05}{.25}(X_B - .5) + .5 = .2X_B + .4$.

   Another way to arrive at this answer is to notice that the MMSE is $E[X_A|X_B] = P[A|X_B]$, which is $P[A|B] = .6$ if $X_B = 1$ and $P[A|B^C] = .4$ if $X_B = 0$. This is the linear function $.2X_B + .4$.

10. More Probability: Cringe?

1. For random variables X and Y, the linear regression line of Y given X goes through the origin if and only __________.

   Answer: $\mathbb{E}[Y] = \frac{Cov(X,Y)}{Var(X)}\mathbb{E}[X]$. This follows by plugging in $x = 0, \hat{y}(x) = 0$ into the LLSE formula.

2. For a random variable X, $\mathbb{E}[X^2] =$ ______. (In terms of $\text{Var}(X)$ and $\mathbb{E}[X]$.)

   Answer: $\text{Var}(X) + \mathbb{E}[X]^2$. This follows from the variance definition.

3. For random variables X and Y, $\text{Cov}(XY) = \mathbb{E}[XY]$ if $\mathbb{E}[X] = \mathbb{E}[Y] =$ ________.

   Answer: 0. $Cov(XY) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Since $\mathbb{E}[X] = \mathbb{E}[Y]$, we need both to be 0.

4. For independent random variables X and Y, what is the best linear estimator for Y given X, i.e., $\hat{y}(X)$? (Fully simplify the answer for this setup.)

   Answer: $\mathbb{E}[Y]$. The covariance is zero since the variables are independent.

5. Consider the process of sampling n people who tested for flu last year to determine the fraction, p, of the population would test positive. To set this up, let X_i be the random indicator variable that i person in the sample got the flu for $i \in \{1, \dots, n\}$.
   (a) What is $\mathbb{E}[X_i]$ in terms of p?
   Answer: p. $\mathbb{E}[X_i]$ is just the probability of person i getting the flu.
   (b) What is the tightest possible upper bound on $\mathrm{Var}(X_i)$, independent of the value of p?
   Answer: $1/4$. The variance of an indicator random variable with expected value p is $p(1-p)$. This is maximized by $p = 1/2$.
   (c) Using Chebyshev, give a 95% confidence interval for p, given that 50 people in your sample of 100 people tested positive for having flu.
   Answer: $[1/2 - 1/\sqrt{20}, 1/2 + 1/\sqrt{20}] \approx [.28, .72]$
   The variance of the proportion who test positive $X = \frac{1}{100}\sum_{i=1}^{100} X_i$ is $Var(X_1)/100 \leq 1/400$. By Chebyshev's inequality $\Pr[|X - \mathbb{E}[X]| \geq c] = \Pr[|X - p| \geq c] \leq \frac{Var(X)}{c^2} \leq \frac{1}{400c^2}$. We want $\frac{1}{400c^2} = \frac{1}{20}$, so we set $c = 1/\sqrt{20}$ and get a 95% confidence interval $[1/2 - c, 1/2 + c]$.

11. The die is cast.

A bag contains a 4-sided die and a 6-sided die. Your friend Lucas pulls a die out of the bag uniformly at random, rolls it, and gets a 1. Conditional on this event, what is the probability they pulled the 4-sided die out of the bag? Show your work.

Answer: Let D be the number of sides of the die pulled out of the bag, and R be the roll. The probability $R = 1$ is:

$$\mathbb{P}[R = 1 | D = 4] \cdot \mathbb{P}[D = 4] + \mathbb{P}[R = 1 | D = 6] \cdot \mathbb{P}[D = 6] = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = \frac{5}{24}.$$

By Bayes' theorem:

$$\mathbb{P}[D = 4 | R = 1] = \frac{\mathbb{P}[R = 1 | D = 4] \cdot \mathbb{P}[D = 4]}{\mathbb{P}[R = 1]} = \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{5}{24}} = \frac{3}{5}.$$

12. Fixed Points.

Recall that the number of fixed points X in a permutation π of the n elements $\{1, 2, \dots, n\}$ is the number of points such that $\pi(i) = i$.

Use the union bound to show that the probability that there are at least k fixed points in a random permutation of $\{1, 2, \dots, n\}$ is at most $\frac{1}{k!}$.

(Hint: What is the probability $1, 2, \dots, k$ are all fixed points?)

Answer: Let A_S be the event that the points in S are all fixed points. The probability there are at least k fixed points is the union over all sets S of size k of A_S, i.e. $\cup_{S:|S|=k} A_S$. For any set S of k points, they are all fixed points with probability $\mathbb{P}[A_S] = (n-k)!/n!$, since the values of these k points are determined, and there are $(n-k)!$ permutations of the remaining $n-k$ elements. There are $\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{1}{k!Pr[A_S]}$ choices for S, so the union bound gives an overall probability bound of $1/k!$.

13. Happier apart.

There are 6 students (Alice, Bob, Catherine, Dustin, Emily, and Frank) in a classroom in which 6 seats are arranged in a circle. When the semester begins, the teacher assigns a seat to each student randomly; that is, all possible seating assignments are equally likely.

Alice and Bob do *not* want to be seated next to each other. If they're assigned to adjacent seats, they will complain of unhappiness. Determine the probability that the teacher's seating assignment does not elicit a complaint.

Answer: $\frac{3}{5}$.

We can sample a seating assignment by first sampling Alice's seat, then Bob's seat, then the remaining students' seats. After sampling Alice's seat, there are 2 empty seats next to her and 3 empty seats not next to her. Bob's seat is chosen uniformly at random from these, so he has a 3/5 chance of not sitting next to Alice.

14. Go big! Or not!

You and a friend play marbles. Each player starts with 10 marbles, and to win you must get all 20 marbles. After deciding on a game, you are given two options of game format. For the game in each part, select the option maximizing your chance of getting all 20 marbles.

- Option A: Play one round; the winner gets all 20 marbles
- Option B: Play a 20-round series (in the event of a 10-to-10 tie, replay until the tie is broken): whoever wins more individual rounds gets all 20 marbles
- Option C: Doesn't matter between A and B. This is the only option that should be selected if A and B yield equal probability of winning.

(a) Odds and Evens, where you have a 35% chance of winning each round.

(b) Rock-Paper-Scissors, where you have a 50% chance of winning each round.

(c) Mancala, where you have a 90% chance of winning each round.

(d) Nim, where you have a 100% chance of winning each round.

Answer:

(a) A. The law of large numbers suggests our chances of winning decrease as we play more rounds.

(b) C. We have a $1/2$ chance of winning in either option.

(c) B. The law of large numbers suggests our chances of winning increases as we play more rounds.

(d) C. We always win regardless of which option we pick.

15. Spam Filter

Jon Byrnu Lee receives many emails each day. Each email is as likely to be *Spam* (S) as it is to be *Not Spam* (S^c)—that is, $\mathbb{P}[S] = \mathbb{P}[S^c] = 1/2$.

To increase his daily productivity, Jon decides to implement a content-based statistical filtering algorithm. By parsing the words in the subject line—perhaps in addition to analyzing other aspects of each email as well—his spam filter directs each email either to his Inbox or to his Spam Folder. Jon *never* reads an email that the filter places in his Spam Folder.

The table below depicts a listing of the probabilities Jon uses to design his spam filter; for ease of notation and grading, the *Symbol* column indicates the symbol you should use to denote each word in your mathematical expressions.

Word	Symbol	$\mathbb{P}[\text{Word}\|S]$	$\mathbb{P}[\text{Word}\|S^c]$
Bonds	B	$\mathbb{P}[B \mid S] = 0.1$	$\mathbb{P}[B \mid S^c] = 0.15$
Reward	R	$\mathbb{P}[R \mid S] = 0.15$	$\mathbb{P}[R \mid S^c] = 0.25$
Winner	W	$\mathbb{P}[W \mid S] = 0.08$	$\mathbb{P}[W \mid S^c] = 0.02$

By way of example, the table specifies that the probability that the word "Winner" appears in the subject line, given that it is *Spam*, is $\mathbb{P}[W \mid S] = 0.08$. The probability that the same word appears in the subject line, given that it is *Not Spam*, is $\mathbb{P}[W \mid S^c] = 0.02$.

(a) Determine $\mathbb{P}[R]$, the probability of the occurrence of the word "Reward" in the subject line of his emails.

Answer: Using the Law of total probability,

$$\mathbb{P}[R] = \mathbb{P}[R \mid S]\mathbb{P}[S] + \mathbb{P}[R \mid S^c]\mathbb{P}[S^c].$$

The emails that Jon Byrnu Lee receives are equally likely to be *Spam* or *Not Spam*, therefore $= \frac{1}{2}$. Using the probabilities in the table, we get that

$$\mathbb{P}[R] \;=\; \left(\frac{15}{100}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) \;=\; \frac{1}{5}.$$

(b) Given that an incoming email has the word "Reward," what is the probability that Jon will read it? Note that Jon will read it if it is *Not Spam*. You may leave your answer here in terms of $\mathbb{P}[R]$.

Answer: The required probability is $\mathbb{P}[S^c \mid R]$. Using the Bayes Rule and the Law of Total Probability, we find that

$$\mathbb{P}[S^c \mid R] \;=\; \frac{\mathbb{P}[R \mid S^c]\mathbb{P}[S^c]}{\mathbb{P}[R]}$$

$$= \frac{\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)}{\mathbb{P}[R]} = \frac{\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{5}\right)} = \frac{5}{8}.$$

You receive full credit if you leave the denominator as $\mathbb{P}[R]$, *provided* your numerator is correct.

(c) Jon receives an email at noon containing the words "Bonds" and "Winner" in the subject line. Determine the probability that this email is *Spam*.

You may *not* assume that the words appearing in the subject line are independent. However, you may assume that the words "Bonds" and "Winner" are independent, *conditioned on Spam* (S). You may also assume that the words "Bonds" and "Winner" are independent, *conditioned on Not Spam* (S^c).

Answer: The probability required is $\mathbb{P}[S \mid B \cap W]$. Using the Bayes Rule, the Law of Total Probability, and Conditional Independence, we get

$$\begin{aligned}
\mathbb{P}[S \mid B \cap W] &= \frac{\mathbb{P}[B \cap W \mid S]\mathbb{P}[S]}{\mathbb{P}[B \cap W]} \\
&= \frac{\mathbb{P}[B \cap W \mid S]\mathbb{P}[S]}{\mathbb{P}[B \cap W \mid S]\mathbb{P}[S] + \mathbb{P}[B \cap W \mid S^c]\mathbb{P}[S^c]} \\
&= \frac{\mathbb{P}[B \mid S]\mathbb{P}[W \mid S]\mathbb{P}[S]}{\mathbb{P}[B \mid S]\mathbb{P}[W \mid S]\mathbb{P}[S] + \mathbb{P}[B \mid S^c]\mathbb{P}[W \mid S^c]\mathbb{P}[S^c]} \\
&= \frac{\left(\frac{10}{100}\right)\left(\frac{8}{100}\right)\left(\frac{1}{2}\right)}{\left(\frac{10}{100}\right)\left(\frac{8}{100}\right)\left(\frac{1}{2}\right) + \left(\frac{15}{100}\right)\left(\frac{2}{100}\right)\left(\frac{1}{2}\right)} \\
\mathbb{P}[S \mid B \cap W] &= \frac{8}{11}.
\end{aligned}$$

16. Going to a party!

Nate is waiting for the bus to go to Professor Rao's social. The time in between buses arriving at Nate's bus stop is an exponential distribution with an average arrival time of 20 minutes. Nate decides that if the next bus arrives within m minutes, he will take the bus to Professor's Rao social, and the bus ride takes 10 minutes. Otherwise, Nate will walk to the social, which takes 30 minutes. Prove that no matter the choice of m, the expected amount of time it will take for Nate to get to the social is 30 minutes.

Answer: Let B be the distribution of the time for the next bus to arrive, which is $\text{Exp}(\lambda = \frac{1}{20})$, and T be the time it takes for Nate to get to the social. Then

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[T|B < m]\mathbb{P}[B < m] + \mathbb{E}[T|B \geq m]\mathbb{P}[B \geq m] \\
&= (\mathbb{E}[B|B < m] + 10)\mathbb{P}[B < m] + (m + 30)\mathbb{P}[B \geq m]
\end{aligned}$$

by the law of total expectation.

To compute $\mathbb{E}[B|B < m]$, we again invoke the Law of Total Expectation:

$$\begin{aligned}
\mathbb{E}[B] &= \mathbb{E}[B|B < m]\mathbb{P}[B < m] + \mathbb{E}[B|B \geq m]\mathbb{P}[B \geq m] \\
&= \mathbb{E}[B|B < m](1 - e^{-m/20}) + \mathbb{E}[B|B \geq m]e^{-m/20}
\end{aligned}$$

We know that $\mathbb{E}[B] = 20$, and by memorylessness, $\mathbb{E}[B|B \geq m] = m + 20$. Thus,

$$\mathbb{E}[B|B < m] = \frac{20 - (m + 20)e^{-m/20}}{1 - e^{-m/20}}.$$

Thus,

$$\begin{aligned}
\mathbb{E}[T] &= \left(\frac{20 - (m + 20)e^{-m/20}}{1 - e^{-m/20}} + 10\right)(1 - e^{-m/20}) + (m + 30)e^{-m/20} \\
&= 30.
\end{aligned}$$

17. Positively Gaussian.

Consider a *zero-mean* Gaussian random variable X whose probability density function (PDF) is given by

$$\forall x \in \mathbb{R}, \qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

where $\sigma > 0$ is the standard deviation of X.

Define another random variable $Y = |X|$.

(a) Let K denote an indicator random variable for the event $Y \leq \sigma$. Evaluate—that is, determine a numerical value for—$\mathbb{E}[K]$, the mean of K. (Use the standard normal table attached to the final page of the exam.)

Answer: We know that the mean of an indicator RV is the probability of its defining event. So,

$$\begin{aligned}\mathbb{E}[K] &= \mathbb{P}[Y \leq \sigma] \\ &= \mathbb{P}[|X| \leq \sigma] \\ &= \mathbb{P}\left[\left|\frac{X}{\sigma}\right| \leq 1\right]\end{aligned}$$

Noting that X/σ is a standardized Gaussian, we have:

$$\mathbb{E}[K] = 2\Phi(1) - 1 = 2(0.8413) - 1,$$

which leads to

$$\mathbb{E}[K] = \mathbb{P}[Y \leq \sigma] = 0.6826.$$

This is simply the probability that the standard Gaussian is within a standard deviation of its mean.

(b) (5 points) Determine a reasonably simple expression for $\mathbb{E}[Y]$. Show your setup.

Note: You do *not* need the PDF $f_Y(y)$ to determine the mean $\mathbb{E}[Y]$.

Answer:

$$\mathbb{E}[Y] = \sigma\sqrt{\frac{2}{\pi}}.$$

We use the Expectation Rule for Functions of a Random Variable. That is,

$$\mathbb{E}[Y] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|\, e^{-x^2/2\sigma^2}\, dx.$$

Noting that the integrand is an even function, we rewrite the right-hand side.

$$= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{+\infty} x\, e^{-x^2/2\sigma^2}\, dx.$$

Let $u = -\frac{x^2}{2\sigma^2}$, so $du = -\frac{x}{\sigma^2}\, dx$, so $x\,dx = -\sigma^2\, du$. Accordingly,

$$\mathbb{E}[Y] = -\frac{2\sigma^2}{\sigma\sqrt{2\pi}} \int_{0}^{-\infty} e^{u}\, du.$$

Flip the limits of the integral to absorb the negative sign in the front.

$$
\begin{aligned}
&= \sigma \sqrt{\frac{2}{\pi}} \int_{-\infty}^{0} e^{u}\, du \\
&= \sigma \sqrt{\frac{2}{\pi}} e^{u} \Big|_{-\infty}^{0} \\
&= \sigma \sqrt{\frac{2}{\pi}} \left(e^{0} - \underbrace{e^{-\infty}}_{=0} \right)
\end{aligned}
$$

This leads us to

$$\mathbb{E}[Y] = \sigma \sqrt{\frac{2}{\pi}}.$$

(c) Determine a reasonably simple expression for the variance of Y.

Note: You do *not* need the PDF $f_Y(y)$ to determine the variance σ_Y^2. You may leave your answer in terms of $\mathbb{E}[Y]$ if you wish.

Answer:

$$\sigma_Y^2 = \sigma^2 \left(1 - \frac{2}{\pi} \right).$$

We know

$$\mathbb{E}[Y] = \sigma \sqrt{\frac{2}{\pi}}.$$

Therefore, all we need before we can determine the variance σ_Y^2 is to determine $\mathbb{E}[Y^2]$, and use the identity

$$\sigma_Y^2 = \mathbb{E}[Y^2] - \mathbb{E}^2[Y].$$

Since $Y = |X|$, we know that $Y^2 = X^2$. But X is a zero-mean random variable, so its variance is equal to its second moment—that is, $\sigma_X^2 = \mathbb{E}[X^2]$. But we know that the variance of X is $\sigma_X^2 = \sigma^2$. So, we have

$$\mathbb{E}[Y^2] = \mathbb{E}[X^2] = \sigma_X^2 = \sigma^2.$$

We can now determine the variance of Y as follows:

$$
\begin{aligned}
\sigma_Y^2 &= \mathbb{E}[Y^2] - \mathbb{E}^2[Y] \\
&= \sigma^2 - \left(\sigma \sqrt{\frac{2}{\pi}} \right)^2 \\
&= \sigma^2 - \sigma^2 \frac{2}{\pi},
\end{aligned}
$$

which leads us to

$$\sigma_Y^2 = \sigma^2 \left(1 - \frac{2}{\pi} \right).$$

If we leave our answer in terms of $\mathbb{E}[Y]$, the answer is $\sigma^2 - E[Y]^2$.

(d) Determine a reasonably simple expression for, and provide a well-labeled plot of, $f_Y(y)$, the PDF of Y. Your plot must be as close to hand-sketched as possible. It must *not* be generated by a computing device or software.

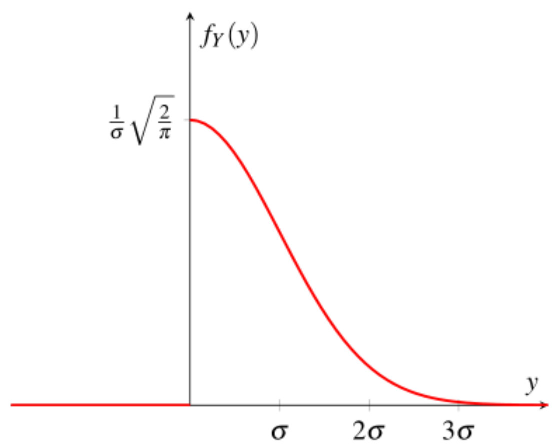
Please note that you must account for the *entire* y axis—that is, your expression must specify what the PDF $f_Y(y)$ is for *every* real y.

Expression:

Well-labeled plot of the $f_Y(Y)$ with y-intercept and units in terms of σ on the x-axis.

Answer: Since $Y = |X| \geq 0$, it's clear that $f_Y(y) = 0$ for $y < 0$.
One can then see that $Pr[Y \in [x, x+\delta]] = Pr[X \in [x, x+\delta]] + Pr[X \in [-x, -x-\delta]]$, and thus that $f_X(y) = 2f_X(x)$ since X is symmetric.
Or that

$$f_Y(y) == \begin{cases} \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\, e^{-y^2/2\sigma^2} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0. \end{cases}$$



18. You complete me.

A *complete graph* K_n is a collection of n nodes where every pair of nodes is connected by an edge. For example, the complete graph K_5 is shown in Figure 1. Throughout this problem, assume that $n \geq 4$.

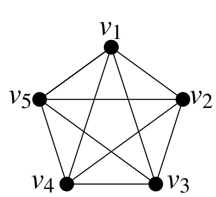


Figure 1

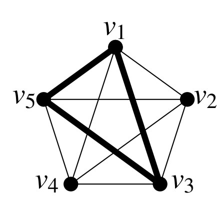


Figure 2

(a) A **triangle** is a set of three nodes, each two of which are connected by an edge. Determine t_n, the number of triangles in the complete graph K_n.

Answer: A triangle is a collection of 3 distinct nodes out of n. The order in which we choose 3 nodes out of n does not matter. So, $t_n = \binom{n}{3}$.

Now assume that we randomly thicken some of the edges in the complete graph K_n. Specifically, each edge is thickened with probability p, or it's left intact (thin) with probability $1-p$, independently of all other edges. Figure 2 depicts such a random thickening of edges as applied to the complete graph K_5 of Figure 1.

A **thick triangle** is a triangle that has three thick edges. For example, in Figure 2, the nodes v_1, v_3, and v_5 form the only thick triangle in the graph.

(b) Determine the probability q that the nodes v_1, v_2 and v_3 form a **thick** triangle.

Answer:

$$\begin{aligned} q &= \mathbb{P}[\{(v_1,v_2) \text{ is thick}\} \cap \{(v_2,v_3) \text{ is thick}\} \cap \{(v_3,v_1) \text{ is thick}\}] \\ &= \mathbb{P}[\{(v_1,v_2) \text{ is thick}\}]\,\mathbb{P}[\{(v_2,v_3) \text{ is thick}\}]\,\mathbb{P}[\{(v_3,v_1) \text{ is thick}\}] \\ &= p^3 \end{aligned}$$

The second equality follows from independence.

(c) Let X_n be the number of thick triangles in K_n. Determine $\mathbb{E}[X_n]$ in terms of possibly t_n, p, q, n.

Hint: Appropriately-chosen indicator random variables may be useful to you.

Answer: Let Y_i be an indicator random variable which takes the value 1 if the i^{th} triangle in K_n is thick. From part (a), there are t_n triangles in K_n.

$$\begin{aligned} X_n &= \sum_{i=1}^{t_n} Y_i \\ E[X_n] &= \sum_{i=1}^{t_n} E[Y_i] \qquad \text{by linearity of expectation} \\ &= \sum_{i=1}^{t_n} P(Y_i = 1) \\ &= \sum_{i=1}^{t_n} q = t_n q \end{aligned}$$

(d) (10 points.) Determine the variance of X_n, the number of thick triangles in K_n. Express your answer in terms of t_n, p, q, n, or a combination of these. Show your work.

Hint: The number of pairs of triangles in K_n that share exactly one edge is $6\binom{n}{4}$.

Answer:

$$\begin{aligned} X_n &= \sum_{i=1}^{t_n} Y_i \\ \mathsf{var}(X_n) &= \sum_{i=1}^{t_n} \mathsf{var}(Y_i) + \sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j) \end{aligned}$$

The variance of Y_i is:

$$\mathsf{var}(Y_i) = E[Y_i^2] - E[Y_i]^2 = E[Y_i] - E[Y_i]^2 = q - q^2$$

If triangles i and j have no common edges, Y_i and Y_j are uncorrelated, and $\mathrm{Cov}(Y_i,Y_j)=0$. If they have one edge in common,

$$\mathrm{cov}(Y_i,Y_j)=E[Y_iY_j]-E[Y_i]E[Y_j]=p^5-q^2.$$

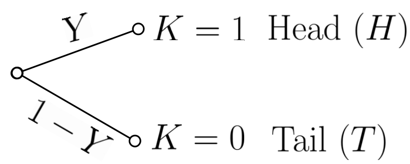
From the hint, there are $6\binom{n}{4}$ pairs of triangles with a common edge. So,

$$\mathrm{var}(X_n)=t_n\left(q-q^2\right)+12\binom{n}{4}\left(p^5-q^2\right)$$

19. What is the bias?

Consider a coin whose bias (probability) for a "Head" is determined by a continuous random variable Y uniformly distributed in the interval $(0,1)$. This is shown in the figure below, with random variable K acting as the indicator variable for the outcome "Head."

$$f_Y(y)=\begin{cases}1 & 0<y<1\\ 0 & \text{elsewhere.}\end{cases}$$



Let H_i denote the event that the outcome of the i^{th} toss is a Head. Similarly, T_i denotes the event that the outcome of the i^{th} toss is a Tail.

Given that $Y=y$, then we have that

$$\mathbb{P}[H_i \mid Y=y]=\mathbb{P}[K_i=1 \mid Y=y]=\mathbb{E}[K_i \mid Y=y]=y,$$

and

$$\mathbb{P}[T_i \mid Y=y]=\mathbb{P}[K_i=0 \mid Y=y]=1-y,$$

Moreover, we have the coin flips are conditionally independent, *given* $Y=y$. For example,

$$\mathbb{P}[H_1\cap H_2 \mid Y=y]=\mathbb{P}[H_1 \mid Y=y]\,\mathbb{P}[H_2 \mid Y=y]$$

and

$$\mathbb{P}\left[\bigcap_{i=1}^{n}H_i \,\middle|\, Y=y\right]=\prod_{i=1}^{n}\mathbb{P}[H_i \mid Y=y],$$

and so on.

However, suppose you have no observation on Y. In this problem, you'll show that the coin flips are *not* independent.

In what follows, it may be useful to know the Law of Total Probability for an event A and a continuous random variable Y:

$$\mathbb{P}[A]=\int_{-\infty}^{+\infty}\mathbb{P}[A \mid Y=y]\,f_Y(y)\,dy.$$

(a) (5 points) Determine a reasonably simple closed-form expression for

$$\mathbb{P}\left[\bigcap_{i=1}^{n} H_i\right] = \mathbb{P}[H_1 \cap H_2 \cap \cdots H_n],$$

in terms of n. Show your setup.

Hint: The events H_i, conditioned on $Y = y$, are independent.

Answer: Using the Law of Total Probability we have

$$\mathbb{P}\left[\bigcap_{i=1}^{n} H_i\right] = \int_{-\infty}^{+\infty} \mathbb{P}\left[\bigcap_{i=1}^{n} H_i \middle| Y = y\right] f_Y(y)\, dy.$$

We now exploit the conditional independence premise, as well as the description of $f_Y(y)$, to rewrite the integral as follows:

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{i=1}^{n} H_i\right] &= \int_0^1 \prod_{i=1}^{n} \mathbb{P}[H_i \mid Y = y]\, dy \\
&= \int_0^1 \prod_{i=1}^{n} y\, dy \\
&= \int_0^1 y^n\, dy \\
&= \left.\frac{y^{n+1}}{n+1}\right|_0^1 \\
\mathbb{P}\left[\bigcap_{i=1}^{n} H_i\right] &= \frac{1}{n+1}.
\end{aligned}$$

(b) Determine a numerical value for $\mathbb{P}[H_j \mid H_i]$ for any pair of distinct positive integers i and j ($i \neq j$). Your result should indicate that H_i and H_j are not independent.

Hint: Recall the events H_i and H_j, conditioned on $Y = y$, are independent.

Answer:

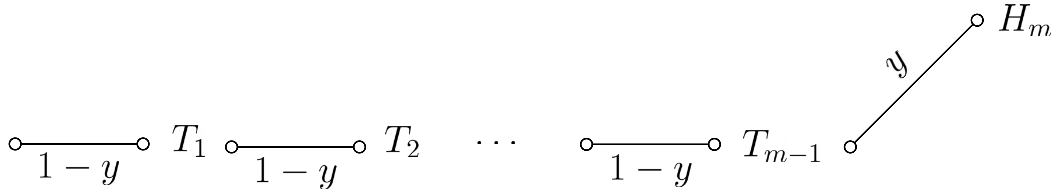
$$\mathbb{P}[H_j \mid H_i] = \frac{\mathbb{P}[H_i \cap H_j]}{\mathbb{P}[H_i]} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \neq \frac{1}{2} = \mathbb{P}[H_j].$$

As we can see, the a priori probability of a Head on the j^{th} toss is $\mathbb{P}[H_j] = 1/2$, whereas if we're told that the outcome of the i^{th} toss is a Head, our belief in the likelihood of H_j changes—it increases. Therefore, we can conclude that H_i and H_j are a priori dependent.

Note that we do not assume any particular ordering on i and j. Either one could be the later toss. It doesn't matter. The point is that if we're told that the outcome of some toss (call it the i^{th}) is a Head, it increases the likelihood of any other toss (call it the j^{th}) being a Head Therefore, H_i and H_j are dependent.

(c) Let M be the random variable denoting the number of tosses up to, and including, the first Head. Determine a reasonably simple expression for the probability mass function (PMF) $p_M(m)$.

A probability tree diagram that might be helpful to you is shown below:



$M = m$ tosses up to, and including, the first Head.

You may find it useful to know the following facts:

- Law of Total Probability involving a discrete and a continuous random variable:

$$p_M(m) = \int_{-\infty}^{+\infty} p_{M|Y}(m \mid y)\, f_Y(y)\, dy.$$

- For $0 < y < 1$ and nonnegative integers α and β,

$$\int_0^1 y^{\alpha}\,(1-y)^{\beta}\, dy = \frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}.$$

Answer: If it's given that $Y = y$, then the conditional PMF

$$p_{M|Y}(m \mid y) = \mathsf{Geom}(y) = \begin{cases} (1-y)^{m-1}\,y & m = 1,2,3,\ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Plugging this and the description of the PDF $f_Y(y)$ into the Law of Total Probability, we have

$$p_M(m) = \int_0^1 (1-y)^{m-1}\, y\, dy,$$

to which we apply the integral identity (Beta Formula) to obtain

$$= \frac{(m-1)!\,1!}{(m+1)!} = \frac{1}{m(m+1)}.$$

Accordingly,

$$p_M(m) = \begin{cases} \dfrac{1}{m(m+1)} & \text{if } m = 1,2,3,\ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Even though you were not asked to do this, you can verify that $p_M(m)$ is a valid PMF. First note that it's nonnegative for all integers m. Then, note that it sums to 1. In particular,

$$\sum_{m=1}^{\infty} p_M(m) = \sum_{m=1}^{\infty} \frac{1}{m(m+1)}$$

Apply a partial-fraction expansion to rewrite the right-hand side as:

$$\begin{aligned}
&= \sum_{m=1}^{\infty} \left(\frac{1}{m} - \frac{1}{m+1} \right) \\
&= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) \cdots . \\
&= 1 \underbrace{- \frac{1}{2} + \frac{1}{2}}_{=0} \underbrace{- \frac{1}{3} + \frac{1}{3}}_{=0} \cdots
\end{aligned}$$

All the terms cancel, except the very first one, which is a 1. Therefore,

$$\sum_{m=1}^{\infty} p_M(m) = 1.$$

(d) Let L denote the number of Heads in a set of n tosses of the coin. In particular,

$$L = K_1 + \cdots + K_n = \sum_{i=1}^{n} K_i,$$

where K_i denotes the value of the indicator random variable K for the i^{th} toss. Our primary goal in this part and the next is estimate the bias Y based on the observed Head count L (pun intended!).
For this part, you may find it useful to know two facts.

- For $0 < y < 1$ and nonnegative integers α and β,

$$\int_0^1 y^{\alpha} (1-y)^{\beta} \, dy = \frac{\alpha! \beta!}{(\alpha + \beta + 1)!}.$$

- The conditional PDF of Y, given L, is of the form

$$f_{Y|L}(y|\ell) = \begin{cases} c(n,\ell) \, y^{\ell} (1-y)^{n-\ell} & \text{if } \ell \in \{0,1,2,\ldots,n\} \text{ and } 0 < y < 1. \\ 0 & \text{otherwise,} \end{cases}$$

  where $c(n,\ell)$ is the normalization factor, dependent on n and ℓ.

(i) What is the normalization factor $c(n,\ell)$?
Answer: The conditional PMF for L, given $Y = y$, is simply the binomial

$$p_{L|Y}(\ell \mid y) = \begin{cases} \binom{n}{\ell} y^{\ell} (1-y)^{n-\ell} & \text{if } \ell \in \{0,1,2,\ldots,n\} \text{ and } 0 < y < 1. \\ 0 & \text{otherwise.} \end{cases}$$

We now plug this into the appropriate version of the Bayes Rule given in the problem, recognizing that $f_Y(y) = 1$ for $0 < y < 1$.

$$\begin{aligned}
f_{Y|L}(y|\ell) &= \frac{\binom{n}{\ell}}{p_L(\ell)} y^{\ell} (1-y)^{n-\ell} \\
&= \begin{cases} c(n,\ell) \, y^{\ell} (1-y)^{n-\ell} & \text{if } \ell \in \{0,1,2,\ldots,n\} \text{ and } 0 < y < 1. \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

where

$$c(n,\ell) = \frac{\binom{n}{\ell}}{p_L(\ell)}.$$

We need not determine $p_L(\ell)$ for this problem, because we'll treat $c(n,\ell)$ as a normalizing factor that forces the area under the conditional PDF $f_{Y|L}(y|\ell)$ to equal 1. In other words, $c(n,\ell)$ is such that

$$\int_{-\infty}^{+\infty} f_{Y|L}(y \mid \ell)\, dy = 1$$

$$c(n,\ell) \int_0^1 y^\ell\, (1-y)^{n-\ell}\, dy = 1.$$

Isolating $c(n,\ell)$, we have

$$c(n,\ell) = \frac{1}{\displaystyle\int_0^1 y^\ell\, (1-y)^{n-\ell}\, dy}.$$

We now use the integral identity given in the problem statement:

$$c(n,\ell) = \frac{1}{\dfrac{\ell!\,(n-\ell)!}{(n+1)!}}.$$

$$c(n,\ell) = \frac{(n+1)!}{\ell!\,(n-\ell)!}.$$

Therefore, the conditional PDF $f_{Y|L}(y|\ell)$ is given by

$$f_{Y|L}(y|\ell) = \begin{cases} \dfrac{(n+1)!}{\ell!\,(n-\ell)!}\, y^\ell\, (1-y)^{n-\ell} & \text{if } \ell \in \{0,1,2,\ldots,n\} \text{ and } 0 < y < 1. \\ 0 & \text{otherwise.} \end{cases}$$

(ii) The minimum mean squared-error estimator of Y, based on L, is defined by

$$\widehat{Y}_{\text{MMSE}}(L=\ell) = \mathbb{E}[Y \mid L=\ell] = \int_{-\infty}^{+\infty} y\, f_{Y|L}(y|\ell)\, dy.$$

Determine a reasonably simply expression for $\widehat{Y}_{\text{MMSE}}(L)$ in terms of L and n. (If you are unsure of your expression for $c(n,\ell)$, you may leave your here in terms of $c(n,\ell)$.)

Answer:

$$\widehat{Y}_{\text{MMSE}}(L) = \mathbb{E}[Y \mid L] = \frac{L+1}{n+2}.$$

$$\begin{aligned} \widehat{Y}_{\text{MMSE}}(\ell) &= \mathbb{E}[Y \mid L=\ell] \\ &= \int_{-\infty}^{+\infty} y\, f_{Y|L}(y|\ell)\, dy \\ &= c(n,\ell) \int_0^1 y\, y^\ell\, (1-y)^{n-\ell}\, dy \\ &= c(n,\ell) \int_0^1 y^{\ell+1}\, (1-y)^{n-\ell}\, dy \end{aligned}$$

We now use the integral identity given in the problem statement:

$$= c(n,\ell)\,\frac{(\ell+1)!\,(n-\ell)!}{(\ell+1+n-\ell+1)!}$$
$$= c(n,\ell)\,\frac{(\ell+1)!\,(n-\ell)!}{(n+2)!}.$$

Now bring in the expression for $c(n,\ell)$ and simplify:

$$= \frac{(n+1)!}{\ell!\,(n-\ell)!}\,\frac{(\ell+1)!\,(n-\ell)!}{(n+2)!}$$
$$= \frac{\ell+1}{n+2}.$$

Therefore, the MMSE estimator is given by

$$\widehat{Y}_{\text{MMSE}}(L) = \mathbb{E}[Y \mid L] = \frac{L+1}{n+2}.$$

(e) Using a formula below, or through some other means, determine a reasonably simple expression for $\widehat{Y}_{\text{LLSE}}(L)$, the *linear* least squares estimator (LLSE) for Y, based on L.
Recall that the LLSE is given by:

$$\widehat{Y}_{\text{LLSE}}(L) = \mathbb{E}[Y] + \frac{\sigma_{LY}}{\sigma_L^2}\Big[L - \mathbb{E}[L]\Big]$$

Note: Please pause and contemplate before you dive into complicated mathematical manipulations. This part can be answered *without* resorting to the LLSE formulas above!

Answer: Among all functions $g(L)$—linear and nonlinear—of the observed random variable L, the optimal estimator of Y, in the MMSE sense, is given by $\widehat{Y}_{\text{MMSE}}(L) = \mathbb{E}[Y \mid L]$, which we found to be

$$\widehat{Y}_{\text{MMSE}}(L) = \mathbb{E}[Y \mid L] = \frac{L+1}{n+2}.$$

But $\widehat{Y}_{\text{MMSE}}(L) = \mathbb{E}[Y \mid L]$ is already of the form $\alpha L + \beta$, where $\alpha = \beta = \dfrac{1}{n+2}$. Therefore, it is also the *linear* MMSE estimator of Y as well. In other words,

$$\widehat{Y}_{\text{LLSE}}(L) = \widehat{Y}_{\text{MMSE}}(L) = \frac{L+1}{n+2}.$$