(Rewritten by Alec Li, as the original solutions have mistakes and unclear answers.)

1 True or False?

For each of the questions below, answer TRUE or FALSE. No need to justify your answer.

Please fill in the appropriate bubble!

1. $(P \implies (R \wedge \neg R)) \implies \neg P$

● **True** ○ False

Answer: This is the form of a proof by contradiction of $\neg P$. Specifically, if by assuming P is true, we get a contradiction (i.e. $R \wedge \neg R \equiv F$), then it must be the case that P is false (i.e. $\neg P$ is true).

2. Let $\mathbb{Z}$ be the integers, and $P(i)$ be a predicate on integers;

$$P(0) \wedge ((\exists i \in \mathbb{Z}) P(i) \wedge P(i+1)) \implies (\forall i \in \mathbb{Z})((i \geq 0) \implies P(i)).$$

○ True ● **False**

Answer: Just because a particular i makes $P(i) \wedge P(i+1)$ true does not mean that $P(i)$ is true for every i.

For example, suppose we have a P such that $P(0)$, $P(4)$, and $P(5)$ are all true, and false everywhere else. This predicate satisfies the LHS but not the RHS; it is true for *some* $P(i)$ and $P(i+1)$, but is not true for all $P(i)$.

3. Let $\mathbb{R}$ be the real numbers; $(\forall x, y \in \mathbb{R})\big((x < y) \implies \big((\exists z \in \mathbb{R})(x < z < y)\big)\big)$

● **True** ○ False

Answer: This statement claims that there always exists another real number between any pair of real numbers, which is true.

4. Let $\mathbb{Q}$ be the rational numbers; $(\forall x, y \in \mathbb{R})\big((x < y) \implies \big((\exists z \in \mathbb{R})(x < z < y)\big)\big)$

● **True** ○ False

Answer: This statement claims that there always exists another rational number between any pair of rational numbers, which is true.

5. Any stable pairing that is optimal for one man is optimal for all men.

○ True ● **False**

Answer: Consider a stable matching instance constructed by taking the union of two separate instances. In particular, we take the union of the two sets of men, and the union of the two sets of women, and extend each preference list, where the people in the other instance are disliked (i.e. appended to the end).

Here, the only stable pairings will pair people within their original instances. Now, consider the male-pessimal pairing in one instance and the male-optimal pairing in the other instance; together, these pairings are stable, as there is no rogue couple between the two different instances. Further, this pairing is optimal for half of the men, and pessimal for the other half.

As an example, consider the following, where x, y, z, and w are men;

x	$1 > 2 > 3 > 4$	1	$y > x > z > w$
y	$2 > 1 > 3 > 4$	2	$x > y > z > w$
z	$3 > 4 > 1 > 2$	3	$w > z > x > y$
w	$4 > 3 > 1 > 2$	4	$z > w > x > y$

Here, $\{x, y, 1, 2\}$ are people in one stable matching instance, and $\{z, w, 3, 4\}$ are the people in the other stable matching instance. Notice how everyone prefers those in different instances the least.

Now, consider the stable pairing $(x, 1)$, $(y, 2)$, $(w, 3)$, $(z, 4)$. This is male optimal in the first instance (i.e. the instance involving $\{x, y, 1, 2\}$), but male pessimal in the second instance (i.e. the instance involving $\{z, w, 3, 4\}$).

In particular, although this pairing is optimal for x, it is not optimal for z (as z would rather be paired with 3, and can be paired with 3 in another stable pairing).

6. Any graph with no triangles is two colorable.

○ True ● **False**

Answer: Consider a single cycle of length 5. Since this is a cycle of odd length, it is not 2-colorable; it also does not contain any triangles, as it is a single cycle.

7. There is a graph with average degree 2 that does not have a cycle.

○ True ● **False**

Answer: Consider the connected components. Since the average degree of the entire graph is 2, there must be a connected component of average degree at least 2.

By the handshaking lemma, the sum of all degrees in a component must be twice the number of edges, i.e. $\sum_{v \in V} \deg(v) = 2e$; this means the average degree is $\frac{1}{n} \sum_{v \in V} \deg(v) = \frac{2}{n} e \geq 2$ for this connected component. This implies that $e \geq n$; in a tree, the number of edges must be exactly $n - 1$, meaning this connected component is connected but not a tree, i.e. it contains a cycle.

8. The length of any Eulerian tour of a graph is even.

○ True ● **False**

Answer: Consider a triangle; the Eulerian tour of the graph has length 3. In particular, the length of a Eulerian tour of a graph must be the total number of edges, which has no guarantee of being even. Only the degrees of all vertices must be even in order for a Eulerian tour to exist.

9. There is a program that takes a program P and input x and number of steps, s and returns YES if and only if P run on x halts in s steps.

● **True** ○ False

Answer: One can simulate the program P run on x for s steps.

10. If one can write a program that solves a problem P using the halting problem as a subroutine, then the problem P is undecidable.

○ True ● **False**

Answer: This is the wrong direction for the reduction; in order to show that a problem P is undecidable, one must solve the halting problem, using a program solving P as a subroutine.

In particular, P can be very easy to solve, and one can still use the halting problem as a subroutine.

11. There is a bijection between the powerset of rational numbers and the real numbers. (The powerset of a set S is the set of all subsets of S).

● **True** ○ False

Answer: The powerset of an infinitely countable set is uncountable. In particular, it is sufficient to show that the powerset of the natural numbers is uncountable. Using Cantor diagonalization, we can list out all possible sets of natural numbers by assigning a 1 or a 0 to each natural number for each set; we write a 1 if the number is included in the set, and a 0 if it is not included.

This diagonalization would look like the following:

	0	1	2	3	4	$\cdots$
S_1	1	1	1	1	1	$\cdots$
S_2	1	0	1	0	1	$\cdots$
S_3	0	0	0	0	0	$\cdots$
S_4	0	1	1	0	1	$\cdots$
S_5	1	1	1	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

If by contradiction we assume this list is complete, we can construct a subset of the natural numbers not in the list by taking the diagonal entries and flip their values, i.e. if we wrote a 1, we do not include the number, and if we wrote a 0, we do include the number in the set. Since this new set differs from each listed set in at most one location, it does not appear in the list, and thus the list is incomplete.

This shows that the powerset of the natural numbers is uncountable; since the set of rational numbers is also countable, there exists a bijection between the rational numbers and the natural numbers, and thus also a bijection between the powerset of rational numbers and the powerset of the natural numbers. This suggests that the powerset of the rational numbers is also uncountable.

12. If $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$ then A and B are independent.

○ True ● **False**

Answer: A and B are independent if and only if $\mathbb{P}[A \cap B] = \mathbb{P}[A] \times \mathbb{P}[B]$.

If A and B have nonzero probability, this implies that $\mathbb{P}[A \cap B] = 0$ (by inclusion-exclusion, as $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$), which is not equal to $\mathbb{P}[A] \times \mathbb{P}[B] > 0$. This means that A and B cannot be independent.

In particular, this statement is only true if A or B have zero probability.

13. Given n balls being thrown into n bins, the event "the first bin is empty" and the event "the second bin is empty" are independent.

○ True ● **False**

Answer: If the first bin is empty, the second bin is less likely to be empty, as we have the same number of balls to be thrown in fewer bins. As such, the two events are not independent.

2 Quick Proof

Prove that $\sum_{i=1}^{n} \frac{1}{i(i+1)} = \frac{n}{n+1}$.

Answer: There are two possible approaches; one with induction, and one with telescoping series (through partial fractions).

Solution 1 (induction): We proceed by induction on n.

Base case ($n = 1$): For $n = 1$, we have $\frac{1}{1(1+1)} = \frac{1}{2} = \frac{1}{1+1}$, and the claim holds.

Inductive Hypothesis: Suppose the claim holds for some $n = k$; that is, $\sum_{i=1}^{k} \frac{1}{i(i+1)} = \frac{k}{k+1}$.

Inductive step: Notice that

$$\sum_{i=1}^{k} \frac{1}{i(i+1)} = \frac{k}{k+1} \quad \text{(IH)}$$

$$\sum_{i=1}^{k} \frac{1}{i(i+1)} + \frac{1}{(k+1)(k+2)} = \frac{k}{k+1} + \frac{1}{(k+1)(k+2)}$$

$$\begin{aligned}\sum_{i=1}^{k+1} \frac{1}{i(i+1)} &= \frac{k(k+2)}{(k+1)(k+2)} + \frac{1}{(k+1)(k+2)} \\ &= \frac{1+k(k+2)}{(k+1)(k+2)} \\ &= \frac{k^2+2k+1}{(k+1)(k+2)} \\ &= \frac{(k+1)(k+1)}{(k+1)(k+2)} \\ &= \frac{k+1}{k+2} = \frac{(k+1)}{(k+1)+1}\end{aligned}$$

This is exactly the claim for $n = k+1$, and as such the claim holds for $n = k+1$.

By the principles of induction, the claim must then hold for all $n \geq 1$.

Solution 2 (telescoping series): Notice that

$$\frac{1}{i(i+1)} = \frac{1}{i} - \frac{1}{i+1}.$$

This can be derived through partial fractions, solving for A and B in $\frac{1}{i(i+1)} = \frac{A}{i} + \frac{B}{i+1}$.

Expanding the summation, we have

$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i+1}\right) = \frac{1}{1} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{n} - \frac{1}{n+1}.$$

Everything cancels out in the middle to leave only

$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = 1 - \frac{1}{n+1} = \frac{n}{n+1},$$

as desired.

3 Short Answer: Discrete Math

1. What is the number of faces in a planar drawing of a planar graph with n vertices where every vertex has degree 3?

$$\frac{n}{2} + 2$$

Answer: By the handshaking lemma, the sum of all degrees is exactly twice the number of edges; that is, $\sum_{u \in V} \deg(u) = 2e$.

Since every vertex has degree three, the LHS becomes $3n$, i.e. three times the number of vertices; this means that $2e = 3n \implies e = \frac{3n}{2}$. By Euler's formula, we have $n + f = e + 2$ (here n is the number of vertices), and plugging in our value of e, we have

$$n + f = \frac{3n}{2} + 2 \implies f = \frac{n}{2} + 2.$$

2. Given a graph $G = (V, E)$ with k connected components, what is the minimum number of edges one needs to add to ensure that the resulting graph is connected?

$k-1$

Answer: We can add edges between connected components, decreasing the number of connected components by 1 each time. To get to just one connected component from an original k connected components, we need to add $k-1$ such edges.

3. The hypercube graph for dimension d has an Eulerian tour when $d =$ ____________ (mod 2).

0

Answer: In order for a graph to have a Eulerian tour, all vertices must have even degree. This means that we must have $d \equiv 2 \pmod 2$

4. For a dimension d hypercube with a Eulerian tour of length L and a Hamiltonian cycle of length ℓ, what is L/ℓ?

$\frac{d}{2}$

Answer: A hypercube of dimension d has a total of $d \cdot 2^{d-1}$ edges, and 2^d vertices.

The length of a Eulerian tour will be the total number of edges, and the length of a Hamiltonian cycle will be the total number of vertices. This means that $L = d \cdot 2^{d-1}$ and $\ell = 2^d$. The ratio is thus

$$\frac{L}{\ell} = \frac{d \cdot 2^{d-1}}{2^d} = \frac{d}{2}.$$

5. What is the minimum number of odd degree vertices in a connected acyclic graph?

2

Answer: There must be at least one odd degree vertex (if all vertices had degree at least 2, then the graph must be cyclic; see question 1.7), and there must be an even number of odd degree vertices (the handshaking lemma implies that the total degree must be even, as it is exactly $2e$).

This means that the minimum number of odd degree vertices in a connected acyclic graph must be 2.

6. What is $2^{10} \pmod{11}$?

1

Answer: FLT says that $a^{p-1} \equiv 1 \pmod{p}$, and here we have $a = 2$ and $p = 11$.

7. For distinct primes p, q, r and $N = pqr$, how many elements of $\{0, 1, \ldots, N-1\}$ are relatively prime to N?

$(p-1)(q-1)(r-1)$

Answer: We can use inclusion-exclusion here, through complementary counting. We can find the number of elements *not* relatively prime to N, i.e. the number of elements that share at least one prime factor with N. The set of such elements is exactly the union of the multiples of p, multiples of q, and multiples of r.

There are a total of qr multiples of p (i.e. $\{0, p, 2p, \ldots, (qr-1)p\}$, and the next multiple of p is $pqr = N$, which is not in the set). Similarly, there are a total of pr multiples of q, and a total of pq multiples of r.

However, just adding $qr + pr + pq$ overcounts the elements that are multiples of two of the primes; we need to subtract these overcounts. In a similar manner as prior, there are a total of r multiples of pq, a total of q multiples of pr, and a total of p multiples of qr.

Lastly, we have now subtracted too much; multiples of pqr (i.e. 0) is subtracted too many times, and is now not counted at all. This means we need to add 1 at the very end to compensate.

In total, we have a total of $qr + pr + pq - p - q - r + 1$ elements that are *not* relatively prime with N. Taking the complement, we have a total of

$$pqr - qr - pr - pq + p + q + r - 1 = (p-1)(q-1)(r-1)$$

elements in $\{0, 1, \ldots, N-1\}$ relatively prime to N.

An alternate solution is to consider numbers mod p, mod q, and mod r. Suppose we look at a number x. In order for x to be relatively prime to N, it must not be equal to $0 \pmod{p}$, it must not be equal to $0 \pmod{q}$, and it must not be equal to $0 \pmod{r}$. This means that it has $p-1$ choices for its value $\pmod{p}$, $q-1$ choices for its value $\pmod{q}$, and $r-1$ choices for its value $\pmod{r}$.

Once we've chosen the residues mod p, q, and r, CRT tells us that this system of modular equations gives a unique solution for $x \in \{0, 1, \ldots, N-1\}$. This means that any combination of residues gives a unique element, and there are a total of $(p-1)(q-1)(r-1)$ such combinations.

8. Consider N and the set $S = \{x \in \{0, \ldots, N-1\} : \gcd(x, N) = 1\}$, where $k = |S|$. For $a \in S$, we define $T = \{ax \pmod{n} : x \in S\}$. What is $|T|$? Answer may include N and k.

k

Answer: Since a is relatively prime to N (i.e. $a \in S$, and as such $\gcd(a, N) = 1$), $f(x) = ax \pmod{N}$ is a bijection.

To see why, we can show that it is both injective and surjective. To show that f is injective, we need to show that $f(x) = f(y) \implies x \equiv y \pmod{N}$:

$$\begin{aligned} f(x) = f(y) &\implies ax \equiv ay \pmod{N} \\ &\implies x \equiv a^{-1}ay \pmod{N} \\ &\implies x \equiv y \pmod{N} \end{aligned}$$

Here, we know that $a^{-1} \pmod{N}$ exists, since $\gcd(a, N) = 1$.

To show that f is surjective, we need to show that for all $y \in S$, there exists some x such that $f(x) = y$. In particular, we have $x = a^{-1}y \pmod{N}$, since $f(a^{-1}y) = aa^{-1}y \equiv y \pmod{N}$.

Now that we've shown that f is bijective, $T = f(S)$ must contain exactly the same number of elements as S, and as such $|T| = |S| = k$.

9. For a prime p, what is a positive integer x that guarantees $a^x \equiv 1 \pmod{p^2}$ for all a relatively prime to p? Answer may include p.

$p(p-1)$

Answer: This follows from a proof similar to that of the proof of FLT.

Consider the function $f(x) = ax \pmod{p^2}$. Since $\gcd(a, p) = 1$, it must be the case that $\gcd(a, p^2) = 1$, as p^2 has only p as a prime factor.

Consider the set $S = \{x \in \{0, 1, \ldots, p^2 - 1\} : \gcd(x, p^2) = 1\}$, i.e. the set of integers between 0 and $p^2 - 1$ that are relatively prime to p^2. In a similar fashion as in the previous part, we can see that $f: S \to S$ is bijective.

Therefore, the set S and the set $f(S)$ contain exactly the same numbers, and the products of the elements in each set must also be equal. That is,

$$\prod_{x \in S} x \equiv \prod_{x \in S} ax = a^{|S|} \prod_{x \in S} x \pmod{p^2}.$$

Since $\prod_{x \in S} x$ contains no factors of p, its value is relatively prime to p^2, and an inverse exists. This means that the above equation implies that $a^{|S|} \equiv 1 \pmod{p^2}$.

To find $|S|$, we can see that there are p multiples of p (i.e. $\{0, p, \ldots, (p-1)p\}$) that we must subtract from the total of p^2 possible values, giving us $|S| = p^2 - p = p(p-1)$.

10. For distinct primes p, q, r, what is $a^{(p-1)(q-1)(r-1)} \pmod{pqr}$, where a is relatively prime to pqr? Answer may include p, q, r.

$1 \pmod{pqr}$

Answer: By FLT, we have

$$\begin{aligned} a^{(p-1)(q-1)(r-1)} &= \left(a^{p-1}\right)^{(q-1)(r-1)} \equiv 1 \pmod{p} \\ a^{(p-1)(q-1)(r-1)} &= \left(a^{q-1}\right)^{(p-1)(r-1)} \equiv 1 \pmod{q} \\ a^{(p-1)(q-1)(r-1)} &= \left(a^{r-1}\right)^{(p-1)(q-1)} \equiv 1 \pmod{r} \end{aligned}$$

By CRT, we have $a^{(p-1)(q-1)(r-1)} \equiv 1 \pmod{pqr}$ as the unique solution to the system $\pmod{pqr}$.

11. Jonathan wants to tell Emaan how many chicken nuggets he ate today, which we will call c. He doesn't want the world to know, so he encrypts it with Emaan's public key (N, e), which yields the ciphertext x. Jerry intercepts the message, and wants to make it look like Jonathan actually ate 5 times as many chicken nuggets. What message should she send to Emaan? Answer may include x, N, and e. You may not include c.

$5^e \cdot x \pmod{N}$

Answer: We want to send some value y such that $y \equiv (5c)^e \pmod{N}$, i.e. y is the ciphertext for the message $5c$. Simplifying, we get $y \equiv 5^d c^d \equiv 5^e x \pmod{N}$.

For the following parts consider two non-zero polynomials $P(x)$ and $Q(x)$ of degree d over $\mathrm{GF}(p)$ (modulo p), with r_p roots and r_q roots respectively.

12. What is the maximum number of roots for the polynomial $P(x)Q(x)$? Answer may include d, r_p, and r_q. (Your answer should be achievable for any valid d, r_p and r_q.)

$\min(p, r_p + r_q)$

Answer: The roots of $P(x)Q(x)$ must also be roots of $P(x)$ or $Q(x)$, which yields $r_p + r_q$ as an upper bound. However, since we are working in $\mathrm{GF}(p)$, any polynomial can only have at most p roots, whereas $r_p + r_q$ can exceed this value. Taking the minimum gives $\min(p, r_p + r_q)$ as the maximum number of roots for $P(x)Q(x)$.

13. What is the minimum number of roots for the polynomial $P(x)Q(x)$? Answer may include d, r_p, and r_q.

$\max(r_p, r_q)$

Answer: To get the least amount of roots possible in $P(x)Q(x)$, we want to have the maximum overlap between the roots of P and the roots of Q. This means that the total number of roots in $P(x)Q(x)$ must be the larger of r_p and r_q, i.e. $\max(r_p, r_q)$.

14. Let $S = \{(x_1, y_1), \ldots, (x_{n+2k}, y_{n+2k})\}$ be a set of $n+2k$ points where the x_i are distinct. If $P(x)$ and $Q(x)$ are polynomials where $P(x_i) = y_i$) for at least $n+k$ points in S and $Q(x_j) = y_j$ for at least $n+k$ points in S, what is the minimum number of points that $P(x)$ and $Q(x)$ must agree on in S? Answer may include n and k.

n

Answer: Between $P(x)$ and $Q(x)$, there are a total of $2n+2k$ points contained in S, but S only contains $n+2k$ points. This means that the overlap must be at least $(2n+2k)-(n+2k) = n$ points.

This fact comes from error correction; specifically, from the derivation of why we need to send $n+2k$ points in order to protect against k general errors for a message of length n. In particular, we can show through contradiction that if we have two distinct polynomials $P(x)$ and $Q(x)$ that pass through at least $n+k$ of the received points, they must have at least n points in common; since P and Q are of degree at most $n+1$, this uniquely defines both P and Q, meaning they must not be distinct.

15. Working over GF(5), describe a degree *exactly* 2 polynomial where $P(1) = 1$ and $P(2) = 2$.

$c(x-1)(x-2)+x$ for any $c \neq 0$

Answer: Notice that $P(x)-x$ has roots at both $x = 1$ and $x = 2$, so we have $P(x)-x = c(x-1)(x-2)$ for some nonzero c. Here, c must be nonzero, otherwise this polynomial has degree 1.

16. Let $P(x)$ be a degree $d = n-1$ polynomial over GF(p) (p is prime) that contains all but $\ell \leq k$ of $n+2k$ points which are given. In this situation, recall that the Berlekamp–Welsh procedure can reconstruct $P(x)$ be assuming the existence of an error polynomial $E(x)$ of degree exactly k and leading coefficient of 1, and a polynomial $Q(x) = P(x)E(x)$. How many possible pairs of $Q(x)$ and $E(x)$ are consistent with the Berlekamp–Welsh procedure? Answer may include ℓ, k, d, n, and p.

$p^{k-\ell}$

Answer: Here, Berlekamp–Welsh is run with the assumption that we are protecting against at most k general errors. This means that the error polynomial is of degree k, with roots at the k errors; here, we have fewer than k errors (specifically, only $\ell \leq k$ errors), so we have freedom in choosing the $k-\ell$ extra error locations. Each extra error location has p choices under GF(p), so we have a total of $p^{k-\ell}$ possible error polynomials, each one of them associated with a different $Q(x) = \frac{P(x)}{E(x)}$ to give the same $P(x)$.

Put another way, $E(x)$ must be of the form $(x-e_1)(x-e_2)\cdots(x-e_k)$, to identify the locations of the k errors. Of these, only ℓ are true error locations, and the rest can vary freely. This means that the latter $(x-e_{\ell+1})\cdots(x-e_k)$ can be any polynomial, and there are a total of $p^{k-\ell}$ ways to define these last few terms.

4 Short Answer: Counting

1. What is the number of ways to place n distinguishable balls into k distinguishable bins?

k^n

Answer: Each of the n balls has k possibilities for which bin it should go in. This gives us k^n possible ways to place the n balls into k bins.

2. What is the number of ways to place n distinguishable balls into k distinguishable bins where no two balls are placed in the same bin? You may assume that $n \leq k$.

$$\frac{k!}{(k-n)!}$$

Answer: The first ball has k choices for the corresponding bin, the second ball has $k-1$ choices (as it can't be the same as the first), etc. This is equivalent to $\frac{k!}{(k-n)!}$.

3. What is the number of ways to divide d dollar bills among p people? Assume dollar bills are indistinguishable and people are distinguishable.

$$\binom{d+p-1}{p-1} = \binom{d+p-1}{d}$$

Answer: Using stars and bars, we have d stars (bills) and $p-1$ bars (p people). This gives us $\binom{d+p-1}{p-1} = \binom{d+p-1}{d}$ ways to split the bills.

4. How many $(x_1, \ldots, x_k, y_1, y_2, \ldots, y_k)$ are there such that all x_i, y_i are non-negative integers, $\sum_{i=1}^{k} x_i = n$, and $y_i \le x_i$ for $1 \le i \le k$? Answer may *not* include any summations.

$$\binom{n+2k-1}{2k-1} = \binom{n+2k-1}{n}$$

Answer: Here, while the equation $\sum_{i=1}^{k} x_i = n$ suggests that this is a stars and bars problem, we also want to pick the y_i's, which have the constraint $y_i \le x_i$ for $1 \le i \le k$. This constraint is difficult to deal with, since it depends entirely on the value of x_i.

Instead, notice that if we pick our y_i's, the x_i's can be defined entirely based on the *distance* between the corresponding y_i's. This difference can vary without bound, and does not depend on the value of the y_i's.

Formally, we can define $z_i = x_i - y_i$ for each i; this represents the distance between the x_i's and y_i's. Notice that $z_i \ge 0$ and we have

$$\sum_{i=1}^{k} x_i - \sum_{i=1}^{k} y_i + \sum_{i=1}^{k} y_i = \sum_{i=1}^{k} z_i + \sum_{i=1}^{k} y_i = n.$$

The last summation sums $2k$ non-negative integers with no other constraints; this means that we can now use stars and bars. We have n stars (i.e summing n 1's) and $2k-1$ bars (i.e. $2k$ variables to group the 1's into), giving us $\binom{n+2k-1}{2k-1} = \binom{n+2k-1}{n}$ choices.

5 Short Answer: Probability

1. Given two tosses of a fair coin, what is $\mathbb{P}$[heads on the second coin | at least one heads in the two tosses]?

$$\frac{2}{3}$$

Answer: With two tosses of a fair coin, HH, TH, HT, and TT are all equally likely. Out of these four outcomes, three have heads: HH, TH, HT. Of these outcomes, exactly two have heads on the second coin.

2. Consider two events, A and B with $\mathbb{P}[A \cup B] = \frac{2}{3}$, $\mathbb{P}[A] = \frac{1}{2}$, $\mathbb{P}[B] = \frac{4}{5}$. What is $\mathbb{P}[A \cap B]$?

$$\frac{11}{20}$$

Answer: By inclusion-exclusion, we have

$$\begin{aligned}\mathbb{P}[A \cup B] &= \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] \\ \frac{3}{4} &= \frac{1}{2} + \frac{4}{5} - \mathbb{P}[A \cap B] \\ \mathbb{P}[A \cap B] &= \frac{1}{2} + \frac{4}{5} - \frac{3}{4} = \frac{11}{20}\end{aligned}$$

3. Alice and Bob both try to climb a rope. Alice and Bob will get to the top of the rope with probability $\frac{1}{3}$ and $\frac{1}{4}$ respectively. Given that exactly one person got to the top, what is the probability that the person is Alice?

$$\frac{3}{5}$$

Answer: With Bayes' rule, we have

$$\mathbb{P}\left[\text{Alice got to the top} \mid \text{exactly one person got to the top}\right] = \frac{\mathbb{P}\left[\text{Alice was the only one that got to the top}\right]}{\mathbb{P}\left[\text{exactly one person got to the top}\right]}.$$

Looking at the numerator, in order for Alice to be the only person to get to the top, Alice must have gotten to the top and Bob must not have gotten to the top. This happens with probability $\frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$.

Looking at the denominator, there are two ways for exactly one person to get to the top: either only Alice got to the top, or only Bob got to the top. This happens with probability $\frac{1}{3} \cdot \frac{3}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$.

Plugging these in, we have

$$\mathbb{P}\left[\text{Alice got to the top} \mid \text{exactly one person got to the top}\right] = \frac{\left(\frac{1}{4}\right)}{\left(\frac{5}{12}\right)} = \frac{3}{5}.$$

4. Given $X \sim \text{Geom}(p)$, what is $\mathbb{P}\left[X = i \mid X > j\right]$? Assume $i > j$.

$$p(1-p)^{(i-j-1)}$$

Answer: Given that $X > j$, we know that we've already had j failed trials. This means that we have $i - j$ trials left, and of these trials, we must have $i - j - 1$ failures followed by one success. This occurs with probability $p(1-p)^{(i-j-1)}$.

Formally, the memoryless property gives that $\mathbb{P}\left[X = i \mid X > j\right] = \mathbb{P}\left[X = i - j\right]$; in particular, the first j failed trials do not change the probabilities of the next $i - j$ trials. This probability is what we just computed above.

5. Given independent $X, Y \sim \text{Bin}(n, p)$, what is $\mathbb{P}[X + Y = i]$?

$$\binom{2n}{i} p^i (1-p)^{2n-i}$$

Answer: Since X and Y are independent, we can imagine that we have a total of $2n$ trials, each with probability p. $X + Y$ is the total number of successful trials among all of these trials. This means that $X + Y \sim \text{Bin}(2n, p)$, and

$$\mathbb{P}[X+Y=i] = \binom{2n}{i} p^i (1-p)^{2n-i}.$$

6. Consider a random variable X where $\mathbb{E}[X^4] = 5$. Give as good upper bound on $\mathbb{P}[X \ge 5]$ as you can.

$$\frac{1}{125}$$

Answer: We have

$$\mathbb{P}[X \ge 5] \le \mathbb{P}[X^4 \ge 5^4] \le \frac{\mathbb{E}[X^4]}{5^4} = \frac{5}{5^4} = \frac{1}{125}.$$

We know that this is the best upper bound we can give, since the following random variable

$$X = \begin{cases} 5 & \text{w.p. } \frac{1}{125} \\ 0 & \text{w.p. } \frac{124}{125} \end{cases}$$

has expectation $\mathbb{E}[X^4] = 5^4 \cdot \frac{1}{125} = 5$ and $\mathbb{P}[X \ge 5] = \frac{1}{125}$.

6 Concepts through balls in bins

Consider throwing n balls into n bins uniformly at random. Let X be the number of balls in the first bin.

1. What is the expected value of X?

$$1$$

Answer: Let X_i be the indicator that ball i lands into the first bin. With linearity of expectation, we have

$$X = \sum_{i=1}^{n} X_i \implies \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

For any given ball, it has a $\frac{1}{n}$ probability of landing in the first bin, so $\mathbb{E}[X_i] = \frac{1}{n}$. This means that we have

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

2. Use Markov's inequality to give an upper bound on $\mathbb{P}[X \ge k]$.

$$\frac{1}{k}$$

Answer: In the previous part, we found that $\mathbb{E}[X] = 1$, meaning we have

$$\mathbb{P}[X \ge k] \le \frac{\mathbb{E}[X]}{k} = \frac{1}{k}.$$

3. What is the variance of X?

$$1-\frac{1}{n}$$

Answer: Similar to the first part, suppose X_i is the indicator that ball i lands in the first bin. This means we have $X = \sum_{i=1}^{n} X_i$. Further, each X_i is independent from the others, as the balls do not affect one another. Since the variance of the indicator is $\text{Var}(X_i) = \frac{1}{n}\left(1-\frac{1}{n}\right)$, we have

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} \frac{1}{n}\left(1-\frac{1}{n}\right) = 1-\frac{1}{n}.$$

4. Use Chebyshev's inequality to give an upper bound on $\mathbb{P}[X \geq k]$.

$$\frac{1-\frac{1}{n}}{(k-1)^2}$$

Answer: From the previous part, we found that $\text{Var}(X) = 1-\frac{1}{n}$, and in the first part, we found that $\mathbb{E}[X] = 1$. Chebyshev's inequality gives

$$\begin{aligned}
\mathbb{P}[X \geq k] &= \mathbb{P}[X - \mathbb{E}[X] \geq k - \mathbb{E}[X]] \\
&\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq k - \mathbb{E}[X]] \\
&= \mathbb{P}[|X-1| \geq k-1] \\
&\leq \frac{\text{Var}(X)}{(k-1)^2} \\
&= \frac{1-\frac{1}{n}}{(k-1)^2}
\end{aligned}$$

5. Now let Y be the number of balls in the second bin. What is the joint distribution of X, Y, i.e. what is $\mathbb{P}\left[X = i, Y = j\right]$?

$$\binom{n}{i}\binom{n-i}{j}\left(\frac{1}{n}\right)^{i+j}\left(1-\frac{2}{n}\right)^{n-i-j}$$

Answer: We want to compute the probability that we have exactly i balls in the first bin and exactly j balls in the second bin.

We have $\binom{n}{i}$ ways to choose the balls to go into the first bin. After choosing these i balls, we have $n-i$ balls left, of which we need to choose j to go into the second bin; this gives $\binom{n-i}{j}$ ways to choose the balls to go into the second bin.

Now that we know the balls that go into the first and second bins, we need to find the probability that these balls actually go into their desired bins. The probability that the selected i balls go into the first bin is $\frac{1}{n^i}$, and the probability that the selected j balls go into the second bin is $\frac{1}{n^j}$. The probability that the $n-i-j$ other balls go into some other bin is $\left(1-\frac{2}{n}\right)^{n-i-j}$, since any given ball has a $1-\frac{2}{n}$ probability of not going into either the first or second bin.

Multiplying all of these terms together, we have a $\binom{n}{i}\binom{n-i}{j}\left(\frac{1}{n}\right)^{i+j}\left(1-\frac{2}{n}\right)^{n-i-j}$ probability of i balls falling in the first bin and j balls falling in the second bin.

6. What is $\mathbb{P}\left[X=i \mid Y=j\right]$?

$$\binom{n-j}{i}\left(\frac{1}{n-1}\right)^i\left(\frac{n-2}{n-1}\right)^{n-i-j}$$

Answer: Given that $Y=j$, we know that j balls will be in the second bin. This means that we essentially have j fewer balls and one fewer bin to consider when looking at X.

As such, out of $n-j$ balls, we need i to land in the first bin, giving $\binom{n-j}{i}$ choices for these i balls. The probability that exactly these i balls land in the first bin is $\left(\frac{1}{n-1}\right)^i$, and the probability that the other $n-i-j$ balls do not land in the first bin is $\left(\frac{n-2}{n-1}\right)^{n-i-j}$.

Multiplying all of these terms together, we have a $\binom{n-j}{i}\left(\frac{1}{n-1}\right)^i\left(\frac{n-2}{n-1}\right)^{n-i-j}$ probability of i balls falling in the first bin, given the j balls have fallen in the second bin.

7 Lots of chicken nuggets

We will model the number of customers going into Emaan's and Jonathan's favorite McDonalds within an hour as a random Poisson variable, i.e. $X \sim \text{Pois}(\lambda)$.

1. Given that $\lambda=5$, what is the probability that 5 people come in during the hour that Emaan and Jonathan are eating chicken nuggets?

$$e^{-5}\frac{5^5}{5!}$$

Answer: We have $X \sim \text{Pois}(5)$, with

$$\mathbb{P}[X=5]=e^{-5}\frac{5^5}{5!}.$$

2. If λ is unknown but is definitely at most 10, how many hours do Emaan and Jonathan need to be at McDonalds to be able to construct a 95% confidence interval for λ that is of width 2. (You should use Chebyshev's inequality here. Recall for $X \sim \text{Pois}(\lambda)$ that $\text{Var}(X)=\lambda$.)

200 hours

Answer: With $X \sim \text{Pois}(\lambda)$, we have $\mathbb{E}[X]=\text{Var}(X)=\lambda$.

In each hour, we take a sample X_i for the number of people entering during the ith hour. The sample average is $Y=\frac{1}{n}\sum_{i=1}^n X_i$, with $\mathbb{E}[Y]=\lambda$ and $\text{Var}(Y)=\frac{1}{n^2}\sum_{i=1}^n \text{Var}(X_i)=\frac{\lambda}{n}$, since each X_i are independent. This means that our unbiased estimate of λ is Y itself.

Using Chebyshev's inequality, we have

$$\mathbb{P}[|Y-\mathbb{E}[Y]|\geq k]\leq\frac{\text{Var}(Y)}{k}.$$

Here, we have $k = 1$, since we want an interval width of 2. This means the RHS simplifies to $\frac{\lambda}{n}$. We want this probability to be at most $\frac{1}{20}$, since we want 95% confidence that the true value of λ lands in this interval.

This means we have

$$\frac{\lambda}{n} \le \frac{1}{20} \implies n \ge 20\lambda \implies n \ge 200.$$

Here, the last inequality comes from the fact that λ is bounded above by 10; $n = 200$ is the smallest number of hours to give the desired confidence no matter what the true λ is.

That's a lot of time to eat chicken nuggets.

3. Solve the previous problem but now assume you can use the Central Limit Theorem. (*Hint*: You may want to use the table in the back of the exam.)

40 hours, or $\lceil 10 \cdot 1.96^2 \rceil = 39$ hours

Answer: By the central limit theorem, we can assume that $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ follows a normal distribution. In particular,

$$\mathbb{E}[Y] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$
$$\text{Var}(Y) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \lambda = \frac{\lambda}{n}$$

This means that $Y \sim \mathcal{N}(\lambda, \frac{\lambda}{n})$, and standardizing Y, we have

$$Z = \frac{Y - \lambda}{\sqrt{\lambda/n}} \sim \mathcal{N}(0,1).$$

We want to find a w such that the probability $\mathbb{P}[-w \le Z \le w] = 0.95$. That is, we want an interval centered around the mean such that Z lies in the interval with 95% confidence. Looking at the standard normal table, we can find that $w \approx 1.96$.

This means that we have

$$0.95 \approx \mathbb{P}\left[-1.96 \le \frac{Y-\lambda}{\sqrt{\lambda/n}} \le 1.96\right] = \mathbb{P}\left[-1.96 \cdot \sqrt{\lambda/n} \le Y - \lambda \le 1.96 \cdot \sqrt{\lambda/n}\right].$$

Here, the width of the confidence interval is in the units of Y, not in the units of Z, so we have to rewrite the fraction to compute the width. After doing so, the width is the distance between the lower bound and upper bound of the interval, or $2 \cdot 1.96 \cdot \sqrt{\lambda/n}$, which we want to be 2. Solving for n, we have

$$2 = 2 \cdot 1.96 \cdot \sqrt{\frac{\lambda}{n}}$$
$$1 = 1.96 \cdot \sqrt{\frac{\lambda}{n}}$$
$$= 1.96^2 \cdot \frac{\lambda}{n}$$
$$n = 1.96^2 \lambda$$

In the worst case, we need to wait $n = 1.96^2 \cdot 10 \approx 39$ hours, since we know λ is at most 10.

8 Not so dense density functions

1. Consider a continuous random variable whose probability density function is cx^{-3} for $x \geq 1$, and 0 outside this range. What is c?

   2

   Answer: We know that the density function integrates to 1. This means that we have

$$\int_1^\infty cx^{-3}\,\mathrm{d}x = -\frac{c}{2}x^{-2}\Big|_1^\infty = \frac{c}{2}.$$

   Since we know this is also equal to 1, we have $c = 2$.

2. Consider random variables X, Y with joint density function $f(x, y) = cxy$ for $x, y \in [0, 1]$, and 0 outside that range.

   (a) What is c?

   4

   Answer: We know that the joint density function integrates to 1. This means that we have

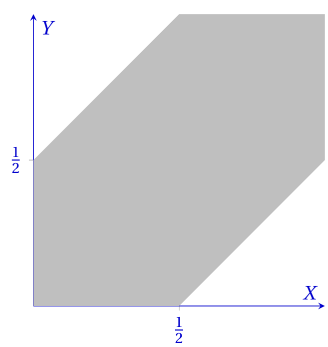
$$\int_0^1\int_0^1 cxy\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \left(\frac{1}{2}cx^2y\right)\Big|_{x=0}^1 \mathrm{d}y = \int_0^1 \frac{1}{2}cy\,\mathrm{d}y = \left(\frac{1}{4}cy^2\right)\Big|_{y=0}^1 = \frac{c}{4}.$$

   Since we know that this is also equal to 1, we have $c = 4$.

   (b) What is $\mathbb{P}\left[|X - Y| \leq \frac{1}{2}\right]$?

$$\frac{41}{48}$$

   Answer: It is perhaps clearest to draw out the region $|X - Y| \leq \frac{1}{2}$:



   The shaded area is the probability that we want to compute. Notice that it is also easier to compute the complement of the probability, i.e. the two unshaded triangles. Since $f(x, y) = 4xy$ is symmetric about the line $x = y$, we need only integrate over one triangle, and multiply by two to get the final probability.

In particular, the integral becomes

$$
\begin{aligned}
1-\mathbb{P}\left[|X-Y|\le \frac{1}{2}\right] &= 2\int_{\frac{1}{2}}^{1}\int_{0}^{x-\frac{1}{2}} 4xy\,\mathrm{d}y\,\mathrm{d}x \\
&= 2\int_{\frac{1}{2}}^{1} (2xy^2)\Big|_{y=0}^{x-\frac{1}{2}}\,\mathrm{d}x \\
&= 2\int_{\frac{1}{2}}^{1} 2x\left(x-\frac{1}{2}\right)^2\mathrm{d}x \\
&= 2\int_{\frac{1}{2}}^{1} 2x\left(x^2-x+\frac{1}{4}\right)\mathrm{d}x \\
&= 2\int_{\frac{1}{2}}^{1} 2x^3-2x^2+\frac{1}{2}x\,\mathrm{d}x \\
&= 2\left(\frac{1}{2}x^4-\frac{2}{3}x^3+\frac{1}{4}x^2\right)\Big|_{\frac{1}{2}}^{1} \\
&= 2\left(\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right)-2\left(\frac{1}{32}-\frac{1}{12}+\frac{1}{16}\right) \\
&= 2\cdot\frac{1}{12}-2\cdot\frac{1}{96}=\frac{7}{48}
\end{aligned}
$$

Taking the complement, we have $\mathbb{P}\left[|X-Y|\le \frac{1}{2}\right]=\frac{41}{48}$.

9 This is Absolutely Not Normal!

Consider a standard Gaussian random variable Z whose PDF is

$$\forall z\in\mathbb{R},\quad f_Z(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}.$$

Define another random variable X such that $X=|Z|$.

(a) Determine a reasonably simple expression for $f_X(x)$, the PDF of X. It may be helpful to draw a plot. Place your final expression in the box below.

$$\begin{cases} 0 & \text{if } x<0 \\ \sqrt{\frac{2}{\pi}}e^{-x^2/2} & \text{if } x\ge 0 \end{cases}$$

Answer: We'll provide two solutions. The first is probably more intuitive, but the second is also worth mentioning.

Solution 1: It's perhaps easiest to consider the CDF first, then differentiate to find the PDF. Notice that since $X=|Z|$, X is always positive. This means that $F_X(x)=f_X(x)=0$ for $x<0$.

Looking at $F_X(x)$ for $x\ge 0$, we have

$$
\begin{aligned}
F_X(x) &= \mathbb{P}[X\le x] \\
&= \mathbb{P}[-x\le Z\le x] \\
&= \Phi(x)-\Phi(-x) \\
&= 2\Phi(x)-1
\end{aligned}
$$

Here, we make use of the CDF of a standard Gaussian $\Phi(x)$. In particular, in the second to last line, $\Phi(x)$ computes $\mathbb{P}[Z \le x]$ and $\Phi(-x)$ computes $\mathbb{P}[Z \le -x]$; subtracting the two gives the probability we want. Equivalently, $\Phi(-x) = 1 - \Phi(x)$ by symmetry, which allows us to simplify to $F_X(x) = 2\Phi(x) - 1$.

Differentiating, we have $f_X(x) = 2f_Z(x)$, where $f_Z(x)$ is the PDF of the standard normal. Putting everything together (and substituting the value of $f_Z(x)$ in), we have

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sqrt{\frac{2}{\pi}} e^{-x^2/2} & \text{if } x \ge 0 \end{cases}$$

Solution 2: Observe that

$$X = \begin{cases} Z & \text{if } Z \ge 0 \\ -Z & \text{if } Z < 0 \end{cases}$$

By the Law of Total Probability, we can write the PDF for X as a *mixture PDF*; in particular,

$$f_X(x) = f_{X|Z\ge 0}(x)\,\mathbb{P}[Z \ge 0] + f_{X|Z<0}(x)\,\mathbb{P}[Z < 0].$$

Note that $\mathbb{P}[Z \ge 0] = \mathbb{P}[Z < 0] = \frac{1}{2}$. Further, for $x < 0$, it must be the case that $f_X(x) = 0$, since X is always positive.

For $x \ge 0$, we have

$$\begin{aligned} f_X(x) &= f_{X|Z\ge 0}(x) \\ &= \frac{f_{X,Z\ge 0}(x)}{\mathbb{P}[Z \ge 0]} && \text{(def. conditional probability)} \\ &= \frac{f_Z(x)}{\mathbb{P}[Z \ge 0]} && (X = Z \text{ if } Z \ge 0) \\ &= \sqrt{\frac{2}{\pi}} e^{-x^2/2} \end{aligned}$$

Here, in the last equality we plug in $\mathbb{P}[Z \ge 0] = \frac{1}{2}$ and $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ as given.

This means that

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sqrt{\frac{2}{\pi}} e^{-x^2/2} & \text{if } x \ge 0 \end{cases}$$

(b) Determine a reasonably simple expression for $\mathbb{E}[X]$, the mean of X. Place your final answer in the box below.

$$\boxed{\sqrt{\frac{2}{\pi}}}$$

Answer: The mean of X is given by

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x f_X(x)\,\mathrm{d}x \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty x e^{-x^2/2}\,\mathrm{d}x \end{aligned}$$

Letting $u = x^2/2$, we have $\mathrm{d}u = x\,\mathrm{d}x$; as such

$$\begin{aligned} &= \sqrt{\frac{2}{\pi}} \int_0^\infty e^{-u}\,\mathrm{d}u \\ &= \sqrt{\frac{2}{\pi}} \end{aligned}$$

More specifically, the last integral evaluates to

$$\int_0^\infty e^{-u}\,\mathrm{d}u = \left(-e^{-u}\right)\Big|_0^\infty = 0-(-1) = 1,$$

allowing us to eliminate it.

10 Joint Distributions with Kyle and Lara

Kyle and Lara arrive in Saint Petersburg randomly and independently, on any one of the first five (5) days of May 2019. Let K be the day that Kyle arrives, and let L be the day that Lara arrives. (Note that K and L will both be in $\{1,2,3,4,5\}$).

Whoever arrives first must wait for the other to arrive before going on any kind of excursion in the city.

(a) Determine $\mathbb{E}[|K-L|]$, the expected wait time in days.

$$\frac{8}{5}$$

Answer: If Kyle and Lara arrive in Saint Petersburg randomly, then their PMFs are uniform over the set $\{1,2,3,4,5\}$:

$$\mathbb{P}[K=k] = \begin{cases} \frac{1}{5} & \text{if } k \in \{1,2,3,4,5\} \\ 0 & \text{otherwise} \end{cases} \qquad \mathbb{P}[L=\ell] = \begin{cases} \frac{1}{5} & \text{if } k \in \{1,2,3,4,5\} \\ 0 & \text{otherwise} \end{cases}$$

Since K and L are independent, the joint PMF will be the uniform PMF over the range $k \in \{1,2,3,4,5\}$ and $\ell \in \{1,2,3,4,5\}$:

$$\mathbb{P}[K=k, L=\ell] = \begin{cases} \frac{1}{25} & \text{if } k,\ell \in \{1,2,3,4,5\} \\ 0 & \text{otherwise} \end{cases}$$

Computing the summation directly, we have

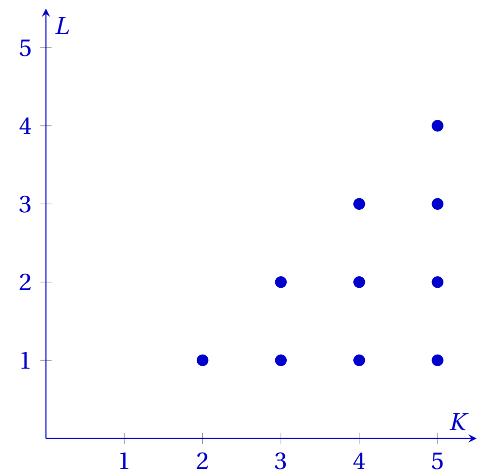
$$\begin{aligned}
\mathbb{E}[|K-L|] &= \sum_{k=1}^{5}\sum_{\ell=1}^{5} \mathbb{P}[K=k, L=\ell]\cdot|k-\ell| \\
&= \sum_{k=1}^{5}\sum_{\ell=1}^{5} \left(\frac{1}{25}\right)|k-\ell| \\
&= 0\left(\frac{5}{25}\right) + 1\left(\frac{8}{25}\right) + 2\left(\frac{6}{25}\right) + 3\left(\frac{4}{25}\right) + 4\left(\frac{2}{25}\right) \\
&= \frac{8}{5}
\end{aligned}$$

(b) Given that Kyle arrives *at least a day later* than Lara:

(i) Determine the conditional probability mass function for Kyle's arrival day, $p_{K|(K>L)}(k)$.

$$\begin{cases} 1/10 & k=2 \\ 2/10 & k=3 \\ 3/10 & k=4 \\ 4/10 & k=5 \\ 0 & \text{otherwise} \end{cases}$$

Answer: Conditioning on the event $K > L$, we have the following restricted sample space:

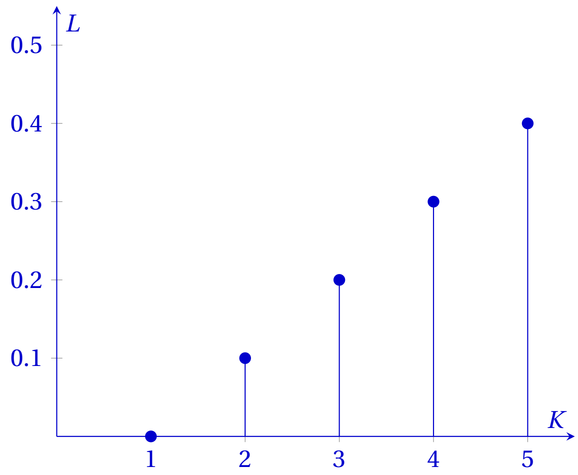


The PMF is still uniform over these sample points, so the conditional PMF of K is the sum along each column of the graph (i.e. along the ℓ direction). In particular,

$$\mathbb{P}[K=k \mid K>L] = \sum_{\ell} \mathbb{P}[K=k, L=l \mid K>L] = \begin{cases} \frac{1}{10} & k=2 \\ \frac{2}{10} & k=3 \\ \frac{3}{10} & k=4 \\ \frac{4}{10} & k=5 \\ 0 & \text{otherwise} \end{cases}$$

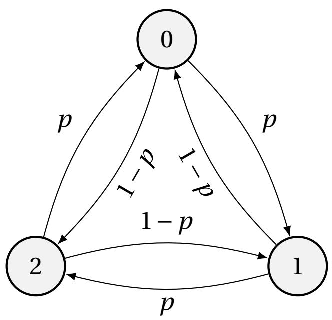
(ii) Provide a well-labeled plot of $p_{K|(K>L)}(k)$.

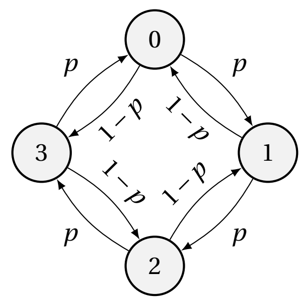
Answer: We have the following stem plot:



11 Markov Chains

Consider the two Markov Chains represented by the following state transition diagrams.





(a) For Markov Chain I:

(i) Do the n-step transition probabilities—defined by $r_{ij}(n) = \mathbb{P}[X_n = j \mid X_0 = i]$—converge as $n \to \infty$?

● **Converges** ○ Does not converge

(ii) If so, determine the corresponding limit to which each transition probability converges, and explain whether and why the limit depends on the initial state (i.e. the state at which the walker was stationed initially). If you assert that the transitional probabilities do not converge, explain why no limit exists.

Answer: This chain is aperiodic since we can return to the same state in either 2 or 3 steps, so the n-step transition probabilities converge.

By symmetry, the limiting values are $\frac{1}{3}$ for all i and j; all transitions are identical in structure. Because all of the limiting values are the same, they do not depend on the initial state.

(b) For Markov Chain II:

(i) Do the n-step transition probabilities—defined by $r_{ij}(n) = \mathbb{P}[X_n = j \mid X_0 = i]$—converge as $n \to \infty$?

○ Converges ● **Does not converge**

(ii) If so, determine the corresponding limit to which each transition probability converges, and explain whether and why the limit depends on the initial state (i.e. the state at which the walker was stationed initially). If you assert that the transitional probabilities do not converge, explain why no limit exists.

Answer: This chain is periodic with period 2, so the transition probabilities do not converge. In particular, we only ever return to the starting state after an even number of steps.

For example, $r_{00}(1) = 0$, $r_{00}(2) \neq 0$, $r_{00}(3) = 0$, $r_{00}(4) \neq 0$, etc. This sequence converges if and only if $r_{00}(n) \to 0$; this is impossible though, as this means we must stay in states 1, 2, and 3 without ever going to state 0—the symmetry of the Markov chain prevents this from occurring.

(c) Consider Markov Chain I above. Determine t_0^*, the *mean recurrence time* for state 0.

The mean recurrence time for a state s is the expected number of steps up to the first return to state s, starting from state s. In other words,

$$t_s^* = \mathbb{E}[\min(n \geq 1 \text{ such that } X_n = s) \mid X_0 = s].$$

In particular,

$$t_s^* = 1 + \sum_i p_{si} t_i,$$

where t_i, which denotes the *mean first passage time* from state i to state s, is given by

$$t_i = \mathbb{E}[\min(n \geq 0 \text{ such that } X_n = s) \mid X_0 = i].$$

(i) Write the system of equations that you would solve in the box below. Use t_0^*, t_1, t_2, and p.

$$t_0^* = 1 + p t_1 + (1-p) t_2$$
$$t_1 = 1 + p t_2$$
$$t_2 = 1 + (1-p) t_1$$

Answer: Writing the first step equations, we have

$$t_0^* = 1 + p t_1 + (1-p) t_2$$
$$t_1 = 1 + p t_2$$
$$t_2 = 1 + (1-p) t_1$$

In particular, if we start at state 0, we can either go to state 1 with probability p, or go to state 2 with probability $1-p$. In either case, we have taken an additional step, and we can recurse to compute the value of t_0^*. For t_1, we can either go to t_2 with probability p, or go back to t_0 with probability $1-p$; in the latter case, we take one step and we're done, so no additional terms are needed. Similar reasoning holds for t_2.

(ii) Set p to $\frac{1}{2}$ and write your final answer for the value of t_0^* in the box below.

3

Answer: With $p = \frac{1}{2}$, we can solve the system of equations for t_1 and t_2 first to give

$$\begin{aligned} t_1 &= 1 + \frac{1}{2}t_2 \\ &= 1 + \frac{1}{2}\left(1 + \frac{1}{2}t_1\right) \\ &= \frac{3}{2} + \frac{1}{4}t_1 \\ \frac{3}{4}t_1 &= \frac{3}{2} \\ t_1 &= \frac{3}{2} \cdot \frac{4}{3} = 2 \end{aligned}$$

This means $t_2 = 1 + \frac{1}{2}t_1 = 1 + 1 = 2$ as well, making

$$t_0^* = 1 + \frac{1}{2}t_1 + \frac{1}{2}t_2 = 1 + 1 + 1 = 3.$$

12 Derive Magic from a Uniform PDF

A random-number generator produces sample values of a continuous random variable U that is uniformly distributed between 0 and 1.

In this problem you'll explore a method that uses the generated values of U to produce another random variable X that follows a desired probability law distinct from the uniform.

(a) Let $g : \mathbb{R} \to [0,1]$ be a function that satisfies all the properties of a CDF. Furthermore, assume that g is invertible, i.e. for every $y \in (0,1)$ there exists a unique $x \in \mathbb{R}$ such that $g(x) = y$.

Let random variable X be given by $X = g^{-1}(U)$, where g^{-1} denotes the inverse of g. Prove that the CDF of X is $F_X(x) = g(x)$.

Answer: The CDF of X is

$$F_X(x) = \mathbb{P}[X \le x] = \mathbb{P}\left[g^{-1}(U) \le x\right].$$

Since g is strictly increasing, we know that $g^{-1}(U) \le x$ if and only if $g(g^{-1}(U)) = U \le g(x)$. This means that

$$F_X(x) = \mathbb{P}\left[g^{-1}(U) \le x\right] = \mathbb{P}\left[U \le g(X)\right].$$

Since U is a uniform between 0 and 1, $\mathbb{P}\left[U \le g(x)\right] = g(x)$, and we can conclude that the CDF of X is $F_X(x) = g(x)$, as desired.

As a side note, if we want to simulate a random variable X that obeys a desired CDF $F_X(x)$ that is invertible over a range $S = \{x \mid 0 < g(x) < 1\}$, we can generate random variable U randomly distributed in $[0,1)$, and let $X = F_X^{-1}(U)$.

(b) A random variable X follows a *double-exponential* PDF given by

$$f_X(x) = \frac{\lambda}{2}e^{-\lambda|x|},$$

for all $x \in \mathbb{R}$, where $\lambda > 0$ is a fixed parameter.

Using the random-number generator described above (which samples U), we want to generate sample values of X. Derive the explicit function that expresses X in terms of U. In other words, determine the expression on the right-hand side of

$$X = g^{-1}(U).$$

To do this, you must first determine the function g. From part (a) you know that $g(x) = F_X(x)$, so you must first determine $F_X(x)$. It might help you to sketch the PDF of X first. Place your expression for g^{-1} in the box below.

$$\begin{cases} \frac{1}{\lambda}\ln(2U) & \text{if } 0 \le U < \frac{1}{2} \\ -\frac{1}{\lambda}\ln(2-2U) & \text{if } \frac{1}{2} \le U < 1 \end{cases}$$

Answer: We can first determine the CDF $F_X(x)$. For $x < 0$, we have $f_X(t) = \frac{\lambda}{2}e^{\lambda x}$, so we can compute

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \le x] \\ &= \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \\ &= \frac{\lambda}{2}\int_{-\infty}^{x} e^{\lambda t}\,\mathrm{d}t \\ &= \frac{\lambda}{2}\left(\frac{1}{\lambda}e^{\lambda t}\right)\Bigg|_{-\infty}^{x} \\ &= \frac{\lambda}{2}\cdot\left(\frac{1}{\lambda}e^{\lambda t} - 0\right) \\ &= \frac{1}{2}e^{\lambda t} \end{aligned}$$

For $x \ge 0$, we have $f_X(t) = \frac{\lambda}{2}e^{-\lambda x}$. Additionally, we know that $F_X(0) = \frac{1}{2}$, so we can compute

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \le 0] + \mathbb{P}[0 \le X \le x] \\ &= \frac{1}{2} + \int_{0}^{x} f_X(t)\,\mathrm{d}t \\ &= \frac{1}{2} + \frac{\lambda}{2}\int_{0}^{x} e^{-\lambda t}\,\mathrm{d}t \\ &= \frac{1}{2} + \frac{\lambda}{2}\left(-\frac{1}{\lambda}e^{-\lambda t}\right)\Bigg|_{0}^{x} \\ &= \frac{1}{2} + \frac{\lambda}{2}\cdot\left(-\frac{1}{\lambda}e^{-\lambda t} + \frac{1}{\lambda}\right) \\ &= \frac{1}{2} + \frac{1}{2}\left(1 - e^{-\lambda t}\right) \\ &= 1 - \frac{1}{2}e^{-\lambda t} \end{aligned}$$

Together, we have

$$g(x) = F_X(x) = \begin{cases} \frac{1}{2}e^{\lambda x} & \text{if } x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x} & \text{if } x \ge 0 \end{cases}$$

To find $X = g^{-1}(U)$, we can look at two different halves of the function; specifically, for $0 \le U < \frac{1}{2}$ and $\frac{1}{2} \le U < 1$ in the inverse function. These two halves correspond exactly with the two halves of the piecewise function we defined earlier.

In particular, for $0 \le U < \frac{1}{2}$, solving for X, we have

$$\begin{aligned} U &= F_X(X) \\ U &= \frac{1}{2}e^{\lambda X} \\ 2U &= e^{\lambda X} \\ \ln(2U) &= \lambda X \\ \frac{1}{\lambda}\ln(2U) &= X \end{aligned}$$

For $\frac{1}{2} \le U < 1$, solving for X, we have

$$\begin{aligned} U &= F_X(X) \\ U &= 1 - \frac{1}{2}e^{-\lambda X} \\ 1 - U &= \frac{1}{2}e^{-\lambda X} \\ 2 - 2U &= e^{-\lambda X} \\ \ln(2-2U) &= -\lambda X \\ -\frac{1}{\lambda}\ln(2-2U) &= X \end{aligned}$$

Together, we have

$$X = g^{-1}(U) = \begin{cases} \frac{1}{\lambda}\ln(2U) & \text{if } 0 \le U < \frac{1}{2} \\ -\frac{1}{\lambda}\ln(2-2U) & \text{if } \frac{1}{2} \le U < 1 \end{cases}.$$