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

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 acknowlege that:

- I alone am taking this exam. Other than with the instructor and GSI, 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.

Signed: _____

2 Long Ago

1. $A \vee \neg(B \wedge C) \equiv A \vee (\neg B \vee \neg C)$

   Answer: True

   Using De Morgan's law, we have $\neg(B \wedge C) \equiv \neg B \vee \neg C$, meaning $A \vee \neg(B \wedge C) \equiv A \vee (\neg B \vee \neg C)$, as claimed.

2. $\neg \forall x \exists y Q(x, y) \equiv \exists x \exists y \neg Q(x, y)$

   Answer: False

   Distributing the negation, we must negate $\forall$ to become $\exists$, and vice versa. This means that the LHS is actually equivalent to $\exists x \forall y \neg Q(x, y)$.

3. $P \implies Q$ is logically equivalent to $\neg Q \vee P$.

   Answer: False

   The actual equivalence is $P \implies Q \equiv \neg P \vee Q$, as shown in lecture and in the notes.

4. Consider a stable matching instance where S is the job optimal stable pairing. Consider a run of the job-propose matching algorithm, where one candidate c rejects a job j that they should not have in one step. That is, c receives an offer from j and j' chooses j' instead of j, when j was ahead of j' in their preference list.

   Let P be the resulting pairing.

   (a) For every instance, job j cannot do better in P than in S. (Better means get a partner who they prefer more.)

   Answer: True

   There are two cases: $(j, c) \in S$ or $(j, c) \notin S$.

   Case 1: $(j, c) \in S$. Since S is job-optimal, c is the best candidate in any stable matching for j; since at some point both j and j' propose to c, with c rejecting j's offer, j can only do worse by proposing to a next (worse) candidate. This means that j cannot do better in P than in S.

   Case 2: $(j, c) \notin S$. Since j was not paired with c in the job-optimal pairing, there are now two other choices: either j is paired with a candidate they like more than c, or j is paired with a candidate they like less than c.

If in the job-optimal matching j is paired with a candidate c' they like *more* than c, then in the job-propose algorithm, j would have proposed to c' before c, and as such never would have gotten to propose to c in the first place, for c to make the incorrect rejection. This means that this case cannot possibly happen.

If in the job-optimal matching j is paired with a candidate c' they like *less* than c, then it must be the case that c rejects j for another job at some point in the future anyways—the incorrect rejection sorts itself out in the future. This means that S and P would end up being the same pairing, and j does exactly the same in P as it does in S.

With all of this together, we've shown that j cannot possibly do better in P than in S.

(b) Every candidate other than c does as well or better in P than in S.

Answer: False

Since c rejects j, it could be the case that whatever job c ends up with was first on another candidate's preference list. In this case, c stole the other candidate's first preference, and the other candidate does worse.

For a concrete example, consider the following instance:

c	$j > j'$
c'	$j' > j$

j	$c > c'$
j'	$c > c'$

The correct job-propose algorithm would proceed as follows:

Day 1	
c	$j, \cancel{j'}$
c'	—

$\Longrightarrow$

Day 2	
c	j
c'	j'

With c incorrectly rejecting j, the algorithm would proceed as follows:

Day 1	
c	$\cancel{j}, j'$
c'	—

$\Longrightarrow$

Day 2	
c	j'
c'	j

Here, c' ends up with a less preferred job in the second case.

(c) Every job other than j does as well or better in P than in S.

Answer: False

Consider the following instance:

c	$j'' > j > j'$
c'	$j > j'' > j'$
c''	$j > j' > j''$

j	$c > c'' > c'$
j'	$c > c' > c''$
j''	$c'' > c' > c$

The correct job-propose algorithm would proceed as follows:

Day 1	
c	$j, \cancel{j'}$
c'	—
c''	j''

$\Longrightarrow$

Day 2	
c	j
c'	j'
c''	j''

With c incorrectly rejecting j, the algorithm would proceed as follows:

Day 1	
c	$\cancel{j}, j'$
c'	—
c''	j''

$\Longrightarrow$

Day 2	
c	j'
c'	—
c''	$\cancel{j''}, j$

$\Longrightarrow$

Day 3	
c	j'
c'	j''
c''	j

Looking at j'', it now ends up with c', which it prefers less than its optimal c''.

Crucially here, while the specific set of preference lists may seem very arbitrary, we can notice that when c rejects j for j', job j now has to re-propose to a new candidate c''. If this new candidate finds that j is more desirable, we've now kicked off another job that would have ended up with c'' (namely, j''). If j'' has c'' first on their preference list, then this means that j'' is forced to take a worse candidate.

This is the key idea when constructing the example here, and should be the main focus; other details in the preference lists are not as important.

3 Graphs

1. Consider a graph on n vertices with exactly one cycle and m edges. What is the number of connected components? (Hint: $m \leq n$.)

   Answer: $n - m + 1$

   To find the number of connected components, it can be helpful to relate this graph to a tree, as a tree has exactly one connected component. Since we are given that the graph contains exactly one cycle, we can remove one edge from the cycle to eliminate the cycle from the graph. This will not change the number of connected components.

   Next, we want to see how many edges we need to add in order to turn the graph into a tree. Since a tree with n vertices will always have exactly $n-1$ edges, we must add $n-m$ new edges to our graph: $(m-1)+(n-m) = n-1$, keeping in mind that we had to remove one edge to eliminate the cycle earlier.

   Since each edge must connect two different components together, each new edge reduces the number of components by 1. This means that we removed $n-m$ components and ended up with 1 component, so we originally had $n-m+1$ components.

2. For a tree on n vertices, what is the expected number of connected components if each edge is deleted with probability $\frac{1}{3}$?

   Answer: $1 + \frac{n-1}{3}$

   There are a total of $n-1$ edges in a tree with n vertices, and each edge is removed with probability $\frac{1}{3}$. We can thus model the number of deleted edges as $\text{Bin}(n-1, \frac{1}{3})$, with expected value $\frac{n-1}{3}$.

   Each removed edge adds a new component, and since we started with one component, we end up with an expected $1 + \frac{n-1}{3}$ connected components.

3. If we delete every edge with probability $\frac{1}{2}$ from an Eulerian graph on n vertices, what is the expected number of odd degree vertices in the remaining graph?

   Answer: $\frac{n}{2}$

   A preliminary identity we need to establish is that

   $$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots.$$

   A nice proof is by expanding $(1+(-1))^n$ using the binomial theorem:

   $$(1+(-1))^n = \sum_{k=0}^{n} \binom{n}{k} (1)^k (-1)^{n-k}.$$

   The positive values in the summation are values where $n-k$ are even, and the negative values are where $n-k$ is odd. Since n is fixed, this splits the binomial coefficients by whether k is even or odd. Since this sum is 0, it must be the case that the identity claimed above is true.

Moving on to the question on hand, we can split up the expected value using indicators X_i denoting whether the ith vertex has an odd degree after edges are removed. Now, looking at a specific vertex, the number of edges removed is distributed as $\text{Bin}(d, \frac{1}{2})$, where d is the initial degree of the vertex.

Further, the final degree of the vertex is even if the number of incident edges removed is even, and similarly, the final degree of the vertex is odd if the number of incident edges removed is odd. This means that we have

$$\begin{aligned}
\mathbb{P}(\text{even degree}) &= \binom{d}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{d-0} + \binom{d}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{d-2} + \binom{d}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{d-4} + \cdots \\
&= \left(\frac{1}{2}\right)^d\left(\binom{d}{0} + \binom{d}{2} + \binom{d}{4} + \cdots\right) \\
\mathbb{P}(\text{odd degree}) &= \binom{d}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{d-1} + \binom{d}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{d-3} + \binom{d}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{d-5} + \cdots \\
&= \left(\frac{1}{2}\right)^d\left(\binom{d}{1} + \binom{d}{3} + \binom{d}{5} + \cdots\right)
\end{aligned}$$

Using the identity stated and proven earlier, we can arrive at the conclusion that both of these probabilities are equal—and in fact, since a vertex degree can only be even or odd, these probabilities are both $\frac{1}{2}$.

This means that $\mathbb{E}[X_i] = \frac{1}{2}$, and the final expectation is $\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}$.

Note: It doesn't actually matter whether the graph is Eulerian or not; the combinatorial identity holds regardless of whether n is even or odd, and we never used the fact that the degrees d were even in the solution either.

4. Every simple cycle is 2-colorable.

   Answer: False

   Any cycle of odd length is not 2-colorable. For example, consider the cycle on 3 vertices (i.e. a triangle)—it is impossible to color this graph with 2 colors (this cycle is also K_3, the complete graph on three vertices, and any complete graph K_n can only be colored with n colors).

4 Mostly Modular

1. For all nonzero $x, y \in \mathbb{N}$, $\gcd(x, y \bmod x) = \gcd(x, y)$.

   Answer: True

   This is an identity that the Euclidean algorithm relies upon for finding GCDs efficiently. To see how this holds, suppose we write $y = qx + r$, where $r = y \bmod x$ and q is an integer.

   Looking at the forward direction of the equality, suppose $d = \gcd(x, y \bmod x)$; this means that d divides x and r. As such, d must also divide $qx + r$, which is equal to y—hence, d divides x and y.

   Looking at the backward direction of the equality, suppose $d = \gcd(x, y)$; this means that d divides x and y. As such, d must also divide qx, and since $r = y - qx$, it must also divide r—hence, d divides x and $y \bmod x$.

   Combined, the claim that $\gcd(x, y \bmod x) = \gcd(x, y)$ must be true.

2. For all nonzero $x, y \in \mathbb{N}$, $\gcd(x, x \bmod y) = \gcd(x, y)$.

   Answer: False

   This is the incorrect form of the identity in the previous part. As a counterexample, take $x = 8$ and $y = 3$. We have $\gcd(x, x \bmod y) = \gcd(8, 2) = 2$, while $\gcd(x, y) = \gcd(8, 3) = 1$.

3. Give an example of positive integers for a and n where

$$(1 \cdot 2 \cdot \cdots \cdot (n-1))a^{n-1} \neq (1 \cdot 2 \cdot \cdots \cdot (n-1)) \pmod{n}.$$

A proof of the equality is given in the proof of FLT, but the important thing is the requirement that $\gcd(a,n) = 1$. When a and n are not relatively prime, this does not hold—an example could be for $a = 2$, $n = 4$, among others.

4. Let $S = \{x : x \in \{1,\ldots,34\} \text{ and } \gcd(x,35) = 1\}$

(a) What is the size of the set S?

Answer: 24

Out of the 34 integers in the set, we have 4 multiples of 7 (from $7 \cdot 1$ to $7 \cdot 4$ and 6 multiples of 5 (from $5 \cdot 1$ to $5 \cdot 6$). Subtracting these, we have $34 - 4 - 6 = 24$.

(b) What is $a^{|S|+1} \pmod{35}$ for $a \in S$?

Answer: a

Recalling the proof of RSA with $N = pq$, and $ed \equiv 1 \pmod{(p-1)(q-1)}$, we know that

$$a^{ed} \equiv a^{1+k(p-1)(q-1)} \equiv a \pmod{pq}.$$

Here, notice that $24 = 6 \cdot 4 = (7-1)(5-1)$ from our calculation in part (a), so what we're actually trying to find here is

$$a^{|S|+1} \equiv a^{24+1} \equiv a^{(7-1)(5-1)+1} \pmod{7 \cdot 5},$$

which we know to be equal to a.

(c) What is $a^{|S|+1} \pmod{35}$ for $a \notin S$?

Answer: a

Note that in the previous part, we made no assumptions on whether a was in the set S or not—RSA works for any message a, as it will always be mapped to a number in $\{0,\ldots,34\}$ under mod 35.

5. What is $18^{-1} \pmod{13}$?

Answer: 8

We can first simplify $18^{-1} \equiv 5^{-1} \pmod{13}$, and noting that $5 \cdot 8 = 40 \equiv 1 \pmod{13}$, we find that $5^{-1} \equiv 8 \pmod{13}$.

One can also arrive at the same answer through EGCD, though with small numbers like these, it can be a lot quicker to iterate through possible multiples of 5.

6. If $x \equiv 1 \pmod{13}$ and $x \equiv 0 \pmod{18}$ then what is $x \pmod{234}$? (Note: $234 = 18 \times 13$)

Answer: 144

We can apply CRT directly here; with a system of two equations $x \equiv a \pmod{p}$ and $x \equiv b \pmod{q}$, the solution is $x \equiv aq(q^{-1} \bmod p) + bp(p^{-1} \bmod q) \pmod{pq}$.

One thing to note here is that one of the coefficients is 0, so we do not need to calculate the second term. The first (and only) term is thus $1 \cdot 18 \cdot (18^{-1} \bmod 13) \equiv 1 \cdot 18 \cdot 8 \equiv 144 \pmod{234}$, using our result from the previous question.

7. For primes p and q, where $e \equiv d^{-1} \pmod{(p-1)(q-1)}$?

(a) What is $a^{ed} \pmod{q}$? (Answer cannot use e or d, but may use numbers, a, p, or q.)

Answer: a

This is just RSA; encryption is $E(x) = x^e$, and decryption is $D(x) = x^d$, meaning the decrypted message is $a^{ed} \equiv a \pmod{pq}$. This would then imply that $a^{ed} \equiv a \pmod{q}$ as well (this is an intermediate step in the proof of RSA).

(b) Find an $x \leq pq$, where $p \mid (a^{ed} - x)$. (Answer is an expression that may use a, p, and q.)

Answer: a

Similarly, this is also just RSA; rewriting the expression in modular arithmetic, we have $a^{ed} - x \equiv 0 \pmod{p}$, or $a^{ed} \equiv x \pmod{p}$. We can now apply the same logic as previous part—the result after encryption and decryption is the original message a.

5 Polynomials

1. Given a polynomial, $x^3 + a_2x^2 + a_1x + a_0$ modulo 7 with roots at 3, 1, and 6. What is a_0? (Notice that the coefficient of x^3 is 1.)

   Answer: $3 \pmod 7$

   Expressed as a product of factors, the polynomial is $(x-3)(x-1)(x-6)$. To find a_0, all we need to do is plug in $x = 0$ to eliminate all other terms—this gives us

   $$(0-3)(0-1)(0-6) = (-3)(-1)(-6) = -18 \equiv 3 \pmod 7.$$

2. Working mod 5, find a polynomial modulo 5 of degree 2 that has roots at 0 and 3, and goes through the point $(2,3)$.

   Answer: $(x-0)(x-3)$

   We know that the polynomial must have the factor $(x-0)(x-3)$ because of its two known zeroes. In order for us to ensure that the polynomial goes through the point $(2,3)$, we can plug in $x = 2$ and rescale to match. We have $(2-0)(2-3) = -2 \equiv 3 \pmod 5$, so we do not need to do any rescaling, and our original polynomial is our final answer.

   Alternatively, we could have directly used Lagrange interpolation on the points $(0,0)$, $(3,0)$, and $(2,3)$. One thing to notice here is that our final polynomial will be of the form

   $$y_1\Delta_{x_1}(x) + y_2\Delta_{x_2}(x) + y_3\Delta_{x_3}(x) = 0 \cdot \Delta_0(x) + 0 \cdot \Delta_3(x) + 3 \cdot \Delta_2(x) = 3\Delta_2(x).$$

   Thus, the only intermediate polynomial we need to calculate is

   $$\Delta_2(x) = \frac{(x-0)(x-3)}{(2-0)(2-3)} = (-2)^{-1}(x-0)(x-3) \equiv 2(x-0)(x-3) \pmod 5.$$

   Noting that $3 \cdot 2 = 6 \equiv 1 \pmod 5$, the coefficients cancel out and we're just left with $(x-0)(x-3)$.

3. Consider that one encodes a message of n numbers $\bmod m$ by forming a degree $n-1$ polynomial using the numbers as coefficients, and sending $2n-1$ points. If each point is erased with probability $\frac{1}{2}$, what is the probability that the original message can be reconstructed? (Hint: each pattern of erasures is equally probable.)

   Answer: $\frac{1}{2}$

   We need at least n points in order to reconstruct a degree $n-1$ polynomial. This means that we need at least a majority of the points to be successfully sent through—that is, at the minimum we can have $n-1$ erasures and n valid points.

   With a symmetry argument (explained in a bit), we can see that this occurs with probability $\frac{1}{2}$. The symmetry arises from the presence of a bijection between the outcomes where a majority of points are erased and the outcomes where a majority of points are sent. More specifically, for each outcome where k points are erased, there will be an outcome with $2n-1-k$ points erased, just by swapping whether each point is erased or successfully sent.This means that the final answer is $\frac{1}{2}$.

6 Countability/Computability

1. For every pair of distinct rational numbers there is a rational number in between them.

   Answer: True

   Simply taking the average of two rational numbers will give another rational number between them. Formally, if $x = \frac{a}{b}$ and $y = \frac{c}{d}$, where a, b, c, d are all integers, we have

   $$\frac{x+y}{2} = \frac{1}{2}\left(\frac{a}{b} + \frac{c}{d}\right) = \frac{ad+bc}{2bd},$$

   which is also a rational number.

2. The rational numbers are uncountable.

   Answer: False

   It was shown in the notes and in lecture that the rational numbers are countable—we can enumerate all rational numbers by enumerating all pairs (a, b) of natural numbers, each pair forming the rational number $\frac{a}{b}$. It is guaranteed that every single (nonnegative) rational number is listed through this process (as we are going through all possibilities of numerators and denominators); to include all negative rational numbers as well, we can just list $\frac{a}{b}$ and $-\frac{a}{b}$ at once as we enumerate.

3. There is a program that takes a program P, an input x, and a number n and determines whether P run on input x ever writes to memory location n.

   Answer: False

   Suppose the program does exist, called `Writes`. We can then argue that we can reduce this to the Halting problem as follows:

```
def TestHalt(P, x):
    def Q(y):
        P(x)  # run P on input x from outer function; we can shift all memory accesses
         ↪  up by 1
        write_to_memory_address(0)  # if Q halts, write to memory address 0
    return Writes(Q, 0, 0)  # check whether Q writes to memory address 0
```

   This program defines an inner function `Q` that first runs `P(x)` (redirecting all memory accesses to avoid writing to memory address 0), then writes to memory address 0. Note that this write only occurs if `P(x)` halts; if `P(x)` loops forever, it will never write to the memory address. The memory address redirection makes it so that we never have a false memory write.

   This means that we can use `Writes` to check whether any program halts or not, just by passing in this inner function. Since we already know that `TestHalt` can never exist, and we've just shown how we can construct `TestHalt` using `Writes`, we know that `Writes` cannot exist in the first place.

4. There is a program that takes a program P, an input x, and a number n and determines whether P run on input x ever writes to any memory location $i \geq n$.

   Answer: True

   The difference between this question and the one previous is that we only have a finite range of memory locations that a program can write to if it loops forever. We can keep track of this, along with the program counter, to confidently tell whether or not the program will ever write to a memory address $\geq n$—we can stop early once we've detected a write to an address $\geq n$, or if we've detected a loop consisting of a subset of only the first n memory addresses.

5. A program "knows" a real number if it takes an integer n and outputs the nth bit of the real number. (Note: positive values of n signify to the left of the decimal point, and negative ones to the right.)

(a) There is a program that knows π.

Answer: True

The key here is that we can always get the nth bit of π in finite time—many approximation methods exist that can do this.

(b) For every real number x, there is a program that knows x.

Answer: False

There are more real numbers than there are programs, as the reals are uncountable while the set of all programs are countable. This means that it is impossible to map programs to real numbers, and there will always exist real numbers which a program cannot know.

7 A little counting

1. What is the number of ways to have k strictly positive numbers that add up to n?

   Answer: $\binom{n-1}{k-1} = \binom{n-1}{n-k}$

   This is a classic stars and bars problem. Since we have k strictly positive numbers $x_1, x_2, \ldots, x_k$, we have

   $$\begin{aligned} x_1 + x_2 + \cdots + x_k &= n \\ (y_1 + 1) + (y_2 + 1) + \cdots + (y_k + 1) &= n \\ y_1 + y_2 + \cdots + y_k &= n - k \end{aligned}$$

   Here, each $x_i = y_i + 1$ to remove the strictly positive constraint. Using stars and bars directly now, we have a total of $n-k$ stars and $k-1$ bars, giving us a total of $\binom{n-k+k-1}{k-1} = \binom{n-1}{k-1} = \binom{n-1}{n-k}$ ways to define these y_i's, each corresponding to an x_i.

2. What is the number of ways to produce a sequence of numbers $0 < x_1 < x_2 < \cdots < x_k < n$?

   Answer: $\binom{n-1}{k}$

   In order for all x_i's to be between 0 and n exclusive, we have a total of $n-1$ possibilities, and we cannot have any duplicates—with any selection of k distinct numbers, we have exactly one arrangement of x_i's that fit the constraints.

   This means that this is equivalent to choosing k items out of $n-1$ total, where order does not matter and with no replacement: $\binom{n-1}{k}$.

3. What is the number of ways to produce a sequence of numbers $0 \le x_1 \le x_2 \le \cdots \le x_k < n$?

   Answer: $\binom{k+n-1}{n-1} = \binom{k+n-1}{k}$

   Similar to the previous part, we can model this as choosing k items from n total, where order does not matter but with replacement. Note that order still does not matter here because any selection corresponds to a unique arrangement that fits the constraints.

   This is then just stars and bars—we have k stars and $n-1$ bars, giving us a total of $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}$ ways of selecting our x_i's.

4. What is the number of poker hands that have at least 1 ace? (Recall that a poker hand is 5 cards from a 52 card deck.)

   Answer: $\binom{52}{5} - \binom{48}{5}$

   It is easier in this case to count the complement of our desired event, i.e. the number of poker hands that have no aces.

This quantity is just $\binom{48}{5}$, as we can remove the 4 aces from the deck, then choose our 5 cards for our hand. We now have to subtract this from the total number of poker hands, $\binom{52}{5}$, to arrive at our final answer of $\binom{52}{5} - \binom{48}{5}$.

8 Probability

1. Consider rolling two six sided fair dice.

   (a) What is the probability that exactly one die is 6?

   Answer: $\frac{10}{36}$

   There are two cases: the first die is a 6, and the second is not, or the second die is a 6, and the first is not. Each case has 5 possible outcomes, giving us 10 possibilities where one die is a 6, out of a total of 36.

   (b) What is the probability that the sum of the two dice is 6?

   Answer: $\frac{5}{36}$

   For each value of the first die from 1 through 5 inclusive, there is exactly one value for the second die that makes their sum equal to 6. This means that there are 5 possibilities where the sum of the pair of dice is 6, out of a total of 36.

   (c) What is the probability that the sum is 6 given that at least one die is at least 3?

   Answer: $\frac{5}{32}$

   In all outcomes where the sum is 6, at least one die is at least 3—this means that the condition just restricts our sample space. The outcomes that are excluded by the condition are outcomes where both dice are less than 3; there are 4 outcomes that fit this (both dice must be either 1 or 2), meaning we now have 32 outcomes in the conditioned sample space.

   (d) The event of rolling a 6 on the first die is independent of the event that the dice sum to 7.

   Answer: True

   The probability that we roll a 6 on the first die is $\frac{1}{6}$, and the probability that the dice sum to 7 is also $\frac{1}{6}$ (no matter what the first die is, the second die has exactly one possible outcome that makes it sum to 7). Together, the probability that we roll a 6 on the first die *and* the dice sum to 7 is $\frac{1}{36}$; we must get the pair $(6,1)$.

   Since $\mathbb{P}(\text{roll 6 on first die} \cap \text{sum to 7}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(\text{roll 6 on first die}) \cdot \mathbb{P}(\text{sum to 7})$, the two events are in fact independent.

   (e) The event of rolling at least one 6 is independent of the event that the dice sum to 7.

   Answer: False

   The probability of rolling at least one 6 is $\frac{11}{36}$, since we can have $(6,?)$ or $(?,6)$, and subtracting the one overcounting of $(6,6)$.

   The probability of rolling at least one 6 *given* that the dice sum to 7 is $\frac{1}{3}$, since the only possibilities are $(1,6)$, $(6,1)$, $(2,5)$, $(5,2)$, $(3,4)$, or $(4,3)$.

   As such, we've shown that $\mathbb{P}(\text{at least one 6}) \neq \mathbb{P}(\text{at least one 6} \mid \text{sum to 7})$, so the two events are not independent.

2. Flip a coin until you repeat either heads or tails 2020 times. We will derive the probability that the first coin is the same as the last coin in the entire sequence of flips.

   (a) If the process stops after 2020 tosses, what is the probability that the first and last coin are the same?

Answer: 1

If we have stopped after 2020 tosses, all 2020 tosses must have been the same, meaning the first and last coin are always the same.

(b) If the process stops after 2021 tosses, what is the probability that the first and last coin are the same?

Answer: 0

If we have stopped after 2021 tosses, the last 2020 tosses must have been the same, and the first toss must have been different—otherwise, we would have stopped after 2020 tosses. This means that the first and last toss will always be different.

(c) What is the probability that the first coin is the same as the last coin in the entire sequence of flips?

Answer: $\frac{2^{2019}}{2^{2020}-1}$

For ease, let p be the probability that the first and last flips are the same.

Without loss of generality, suppose the first flip was heads (the same reasoning holds if we assumed the first flip was tails).

Looking at the first 2020 flips, the probability that all of the first 2020 flips are heads is $\frac{1}{2^{2019}}$ (as we assumed the first flip is heads, and we want 2019 more heads). Equivalently, the probability that we do not get 2020 heads in a row in the first 2020 flips is $1-\frac{1}{2^{2019}}$.

In the case that we do not end after the first 2020 flips, i.e. we flipped tails somewhere in middle, let's focus on the moment in which we flip that tails. Since each flip is independent of any previous flips, we can think of this moment as if we just started flipping coins, and our first flip was a tails—we've essentially just reset our counter.

At this point, now that we've broken our streak of heads, what's the new probability of the last flip being heads? We're right back where we started in the first place, except starting with tails, so the probability that the last flip is also a tails (whenever we stop) is again p. Hence, the probability that the last flip is a heads is $1-p$ (to match our true first flip).

Written out formally, with the total probability rule we have

$$\underbrace{\mathbb{P}(\text{first} = \text{last})}_{p} = \underbrace{\mathbb{P}(\text{first} = \text{last} \mid 2020 \text{ flips})}_{1}\underbrace{\mathbb{P}(2020 \text{ flips})}_{\frac{1}{2^{2019}}} + \underbrace{\mathbb{P}(\text{first} = \text{last} \mid > 2020 \text{ flips})}_{(1-p)}\underbrace{\mathbb{P}(> 2020 \text{ flips})}_{\left(1-\frac{1}{2^{2019}}\right)}.$$

If we now solve this resulting equation for p, we have

$$\begin{aligned}
p &= \frac{1}{2^{2019}} + (1-p)\left(1-\frac{1}{2^{2019}}\right) \\
p &= \cancel{\frac{1}{2^{2019}}} + 1 - \cancel{\frac{1}{2^{2019}}} - p + \frac{p}{2^{2019}} \\
2p - \frac{p}{2^{2019}} &= 1 \\
p\left(2-\frac{1}{2^{2019}}\right) &= 1 \\
p &= \frac{1}{2-\frac{1}{2^{2019}}} = \frac{1}{\frac{2^{2020}}{2^{2019}} - \frac{1}{2^{2019}}} = \frac{2^{2019}}{2^{2020}-1}
\end{aligned}$$

3. Which of the following are always true?

(a) $\mathbb{E}[10X] = 10\mathbb{E}[X]$

Answer: True

This is linearity of expectation (that is, $\mathbb{E}[aX+b] = a\mathbb{E}[X]+b$), which always holds.

(b) $\mathbb{E}\left[X^2\right] = \mathbb{E}[X]^2$

Answer: False

This is not true in general. In fact, if this were true, then we must have $\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \text{Var}(X) = 0$.

(c) $\mathbb{E}\left[(X-Y)^2\right] = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[X]\mathbb{E}[Y]$

Answer: False

If we were to expand the LHS and use linearity on each expanded term, we have

$$\mathbb{E}\left[(X-Y)^2\right] = \mathbb{E}\left[X^2 - 2XY + Y^2\right] = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[XY].$$

We cannot simplify this any further with linearity. The problem statement has additionally assumed $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, which is not always true—it is only true if X and Y are independent.

(d) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

Answer: False

This identity is only true if X and Y are independent.

If X and Y are dependent, this may not be true—consider the simple case where $X = Y$; $\text{Var}(X+Y) = \text{Var}(2X) = 4\,\text{Var}(X)$, while $\text{Var}(X) + \text{Var}(Y) = 2\,\text{Var}(X)$, which are not equal.

9 Marbles

Consider two bags of marbles; the "majority red" bag has 6 red marbles and 4 blue marbles, and the "majority blue" bag has 3 red marbles and 7 blue marbles, and each bag is chosen with probability $\frac{1}{2}$.

1. If you draw a blue marble where each marble in the bag is equally likely, what is the probability that the bag is the "majority blue" bag?

   Answer: $\frac{7}{11}$

   Let A be the event that we chose the "majority blue' bag, and let B be the event that we get a blue marble.

   Note that since each bag can be chosen with equal probability, each marble in each bag is chosen with equal probability, and there are an equal number of marbles in each bag, any given marble among the two bags is also equally likely to be chosen. This means that $\mathbb{P}(B) = \frac{11}{20}$ because there are a total of 11 blue marbles out of 20 total marbles.

   We also have $\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A) = \frac{7}{10} \cdot \frac{1}{2} = \frac{7}{20}$, or in another perspective, 7 of the blue marbles are from the "majority blue" bag, out of the total of 20 marbles.

   This then means that

   $$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\left(\frac{7}{20}\right)}{\left(\frac{11}{20}\right)} = \frac{7}{11}.$$

2. What is the probability that the next marble is blue?

   Answer: $\frac{6}{11}$

   Let N be the event that the next marble we choose is blue. For ease, let us also define $C = A \mid B$ (that is, the event that we drew from the "majority blue" bag given that we drew a blue marble), and let $\overline{C} = \overline{A} \mid B$ (that is, the event that we drew from the "majority red" bag given that we drew a blue marble). Note that $\mathbb{P}(C) = \frac{7}{11}$ from the previous question, and $\overline{C} = 1 - C = \frac{4}{11}$ as well.

   With the Law of Total Probability, we have two cases: either we have drawn the first blue marble from the "majority blue" bag (i.e. A happened), or we have drawn the first blue marble from the "majority red" bag

(i.e. $\overline{A}$ happened). In either case, the bag we drew from now has 9 marbles, and one less blue marble:

$$\begin{aligned}\mathbb{P}(N) &= \mathbb{P}(N \mid C)\,\mathbb{P}(C) + \mathbb{P}\left(N \mid \overline{C}\right)\mathbb{P}\left(\overline{C}\right)\\ &= \frac{6}{9}\cdot\frac{7}{11} + \frac{3}{9}\cdot\frac{4}{11}\\ &= \frac{54}{99} = \frac{6}{11}\end{aligned}$$

10 Variance, covariance, tail bounds

1. If $\mathbb{E}[X] = 4$, and $\mathbb{E}[Y] = 5$, and $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, what is $\mathrm{Cov}(X, Y)$?

   Answer: 0

   Using the definition of covariance, we have $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$; knowing that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ means the RHS is 0.

2. A student earns one standard deviation above the mean on both exam 1 and exam 2. We define random variables X and Y as the score of a randomly chosen student on exam 1 and exam 2 respectively. If $\mathrm{Var}(X) = 1$, $\mathrm{Var}(Y) = 1$, and $\mathrm{Cov}(X, Y) = 0.5$, how many standard deviations above the mean did the student get on the sum of her two scores? (Recall $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.)

   Answer: $\frac{2}{\sqrt{3}}$

   Since the student scored one standard deviation above the mean on both exams, and both standard deviations are $\sqrt{1} = 1$, the student scored 1 point above the mean on both exams. This means that the student also scored 2 points above the mean on the sum of the two scores.

   Computing the variance of the sum of the scores, we have $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y) = 1 + 1 + 2 \cdot 0.5 = 3$, utilizing the hint given. This means that the standard deviation of the sum of the two scores is $\sqrt{3}$.

   To see how many standard deviations above the mean the student scored on the sum, we divide the raw point score by the standard deviation—this means that they scored $\frac{2}{\sqrt{3}}$ standard deviations above the mean.

3. For a random variable X, where $X \geq -1$, $\mathbb{E}[X] = 5$, and $\mathbb{E}\left[X^2\right] = 26$, give an upper bound on $\mathbb{P}(X \geq 6)$. (It should be tight with respect to the appropriate inequality.)

   Answer: $\frac{6}{7}$

   Using Markov's Inequality on $Y = X + 1$, we have

   $$\mathbb{P}(X \geq 6) = \mathbb{P}(Y - 1 \geq 6) = \mathbb{P}(Y \geq 7) \leq \frac{\mathbb{E}[Y]}{7}.$$

   We also have $\mathbb{E}[Y] = \mathbb{E}[X + 1] = \mathbb{E}[X] + 1 = 5 + 1 = 6$, so the upper bound is $\frac{6}{7}$.

   If we were to try this with Chebyshev's Inequality, we have

   $$\mathbb{P}(X \geq 6) = \mathbb{P}(X - 5 \geq 1) \leq \mathbb{P}(|X - 5| \geq 1) \leq \frac{\mathrm{Var}(X)}{1} = \mathrm{Var}(X).$$

   Since $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X] = 26 - 5^2 = 1$, Chebyshev's gives us a quite useless upper bound of 1. In this case, Markov's Inequality is the best choice of an upper bound.

4. For a random variable X, where $\mathbb{E}[X] = 5$ and $\mathbb{E}\left[X^2\right] = 26$, give an upper bound on $\mathbb{P}(X \geq 9)$. (It should be tight with respect to the appropriate inequality.)

Answer: $\frac{1}{16}$

Using Chebyshev's Inequality, we have

$$\mathbb{P}(X \geq 9) = \mathbb{P}(X - 5 \geq 4) \geq \mathbb{P}(|X-5| \geq 4) \leq \frac{\text{Var}(X)}{16}.$$

Since we know that $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = 26 - 5^2 = 26 - 25 = 1$, we have an upper bound of $\frac{1}{16}$.

We cannot use Markov's Inequality here, because we are not guaranteed that X is nonnegative; thus, Chebyshev's Inequality is the only concentration inequality we can use.

5. Let X be a random variable such that $\mathbb{E}[X] = 10$ and $\text{Var}(X) = \sigma^2$. Let $Y = \frac{X_1 + X_2 + \cdots + X_n}{n}$ where X_i are i.i.d samples of X. For what value of n is $\mathbb{P}(|Y - \mathbb{E}[X]| \geq 1) \leq 0.05$? (Provide a bound that is as tight as possible using Chebyshev's inequality.)

Answer: $n \geq 20\sigma^2$

Firstly, note that

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X].$$

This means that we can use Chebyshev's directly on the desired probability:

$$\mathbb{P}(|Y - \mathbb{E}[X]| \geq 1) = \mathbb{P}(|Y - \mathbb{E}[Y]| \geq 1) \leq \frac{\text{Var}(Y)}{1} = \text{Var}(Y).$$

We know that

$$\text{Var}(Y) = \text{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{\sigma^2}{n}.$$

This means that the probability is upper bounded by $\frac{\sigma^2}{n}$. We want this probability to be upper bounded by 0.05, so we want to find n such that $\frac{\sigma^2}{n} \leq 0.05$. Rearranging, we have $n \geq 20\sigma^2$.

11 Continuous: warmup

Consider a continuous random variable X with pdf $f(x)$. (Answers below are a number or possibly expressions that involve the random variable and $\mathbb{E}[\cdot]$ or $\text{Var}(\cdot)$)

1. What is $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$?

Answer: 1

This follows by the definition of a valid probability density function; the integral over all possible values must equal 1.

2. What is $\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$?

Answer: $\mathbb{E}[X]$

This follows by the definition of the expected value of a random variable.

3. What is $\int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$?

Answer: $\mathbb{E}\left[X^2\right]$

This follows by the Law of the Unconscious Statistician (LOTUS), which says that for a continuous RV X with pdf $f(x)$,

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x.$$

Here, $g(x) = x^2$.

4. Consider $Y = 2X$ where $X \sim \text{Exp}(\lambda)$, $Y \sim \text{Exp}(\lambda')$. What is λ'?

Answer: $\frac{\lambda}{2}$

A property of exponential distributions is that the rate scales inversely with a constant coefficient. Specifically, if $X \sim \text{Exp}(\lambda)$, $aX \sim \text{Exp}(\frac{\lambda}{a})$. In this case, $Y \sim \text{Exp}(\frac{\lambda}{2})$, meaning $\lambda' = \frac{\lambda}{2}$.

We can formally prove this by looking at the CDF:

$$\mathbb{P}(aX \le t) = \mathbb{P}\left(X \le \frac{t}{a}\right) = 1 - e^{-\lambda \frac{t}{a}} = 1 - e^{-\frac{\lambda}{a}t}.$$

The last expression is the CDF of an $\text{Exp}(\frac{\lambda}{a})$ RV.

5. Recall that for choosing a random point in a unit square, the pdf is $f(x,y) = 1$ in the unit square and zero elsewhere. What is the pdf for choosing a uniform point in a 2×2 square? (Answer need only state the pdf inside the 2×2 square as outside it is zero.

Answer: $\frac{1}{4}$

Since we're multiplying the area by 4, we should divide the density by 4 to keep the integral of the pdf constant at 1. This means the pdf becomes a constant $\frac{1}{4}$.

More formally, we can derive this pdf by solving for its constant density (set to an unknown c):

$$1 = \int_0^2 \int_0^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^2 \int_0^2 c\,\mathrm{d}x\,\mathrm{d}y = 4c \implies c = \frac{1}{4}.$$

12 Distributions: continuous and discrete

1. Given $X, Y \sim \text{Bin}(n,p)$, what is the variance of $X+Y$?

Answer: $2np(1-p)$

The question was missing the assumption that X and Y are independent; the intended solution utilizes the fact that $X+Y \sim \text{Bin}(2n,p)$ if X and Y are independent. The reasoning is that each RV describes n independent trials, so together, they describe $2n$ independent trials, each with probability p of success. This means that the variance of $X+Y$ is $2np(1-p)$.

However, the question statement as is does not have that assumption, so we must also add $2\,\text{Cov}(X,Y) = 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) = 2\big(\mathbb{E}[XY] - (np)^2\big)$, giving an answer of $2np(1-p) + 2\big(\mathbb{E}[XY] - (np)^2\big)$.

2. What is $\mathbb{E}[\min(X,Y,Z)]$ where $X, Y, Z \sim \text{Geom}(p)$?

Answer: $\frac{1}{1-(1-p)^3}$

One thing to note is that

$$\begin{aligned}\mathbb{P}(\min(X,Y,Z) \ge k) &= \mathbb{P}(X \ge k \cap Y \ge k \cap Z \ge k)\\ &= \mathbb{P}(X \ge k)\,\mathbb{P}(Y \ge k)\,\mathbb{P}(Z \ge k)\\ &= \left((1-p)^{k-1}\right)^3 = (1-p)^{3(k-1)}\end{aligned}$$

Using the tail sum formula for expectation, we have

$$\mathbb{E}[\min(X,Y,Z)] = \sum_{k=1}^{\infty} \mathbb{P}(\min(X,Y,Z) \ge k) = \sum_{k=1}^{\infty} (1-p)^{3(k-1)} = \sum_{k=0}^{\infty} (1-p)^{3k}.$$

This is an infinite geometric series with common ratio $r = (1-p)^3$, and initial term of $a_0 = 1$, so the sum is $\frac{a_0}{1-r} = \frac{1}{1-(1-p)^3}$.

3. What is $\mathbb{E}[\min(X,Y,Z)]$ where $X, Y, Z \sim \text{Exp}(\lambda)$?

Answer: $\frac{1}{3\lambda}$

Following a similar initial thought process for the previous question, we have

$$\begin{aligned}\mathbb{P}(\min(X,Y,Z) \geq t) &= \mathbb{P}(X \geq t \cap Y \geq t \cap Z \geq t) \\ &= \mathbb{P}(X \geq t)\,\mathbb{P}(Y \geq t)\,\mathbb{P}(Z \geq t) \\ &= \left(e^{-\lambda t}\right)^3 = e^{-3\lambda t}\end{aligned}$$

One thing that is different in this case is that the last expression is the CCDF of an Exp(3λ) RV. This means that we can immediately find the mean to be $\frac{1}{3\lambda}$.

4. Let $Z \sim \text{Exp}(\lambda)$ and $Y = \lceil Z \rceil$, where $\lceil x \rceil$ is the lowest integer of value at least x. Note that the variable $Y \sim \text{Geom}(p)$. What is the value of p in terms of λ?

   Answer: $1 - e^{-\lambda}$

   In order for $\lceil Z \rceil = k$, we must have $k-1 < Z \leq k$. This means that we have

$$\begin{aligned}\mathbb{P}(\lceil Z \rceil = k) &= \mathbb{P}(k-1 < Z \leq k) \\ &= \mathbb{P}(Z > k-1) - \mathbb{P}(Z > k) \\ &= e^{-\lambda(k-1)} - e^{-\lambda k} \\ &= e^{-\lambda(k-1)} - e^{-\lambda(k-1)} e^{-\lambda} \\ &= (e^{-\lambda})^{k-1}\left(1 - e^{-\lambda}\right)\end{aligned}$$

   This is a geometric distribution with parameter $1 - e^{-\lambda}$.

   Alternatively, one can arrive at the PDF through one integral:

$$\begin{aligned}\mathbb{P}(k-1 < Z \leq k) &= \int_{k-1}^{k} f_Z(t)\,\mathrm{d}t \\ &= \int_{k-1}^{k} \lambda e^{-\lambda t}\,\mathrm{d}t \\ &= (-e^{-\lambda t})\Big|_{k-1}^{k} \\ &= e^{-\lambda(k-1)} - e^{-\lambda k}\end{aligned}$$

   This is the same intermediate result from prior, and we can simplify in the same way.

5. Let $Y \sim \text{Exp}(\lambda)$. What is the conditional probability density function of Y if $Y \in [i, i+1]$ for a natural number i?

   [A] $\dfrac{\lambda e^{-\lambda x}}{e^{-\lambda i}}$ [B] $\dfrac{\lambda e^{-\lambda x}}{1 - e^{-\lambda}}$ [C] $\dfrac{\lambda e^{-\lambda(x-i)}}{1 - e^{-\lambda}}$ [D] $\dfrac{\lambda e^{-\lambda(x-i)}}{(1 - e^{-\lambda})^i}$

   Answer: $\frac{\lambda e^{-\lambda(x-i)}}{1-e^{-\lambda}}$

   Notice that by the memoryless property, $f_{Y|Y\in[i,i+1]}(y) = f_{Y|Y\in[0,1]}(y-i)$ for $y \in [i, i+1]$; it doesn't matter if we start observing at time 0 or time i if we know that we have to wait at most another 1 unit of time.

   This means that we can rewrite this conditional pdf as

$$f_{Y|Y\in[0,1]}(y-i) = \frac{f_Y(y-i)}{\mathbb{P}(Y \in [0,1])} = \frac{\lambda e^{-\lambda(y-i)}}{1 - e^{-\lambda}}.$$

   The answer choice just has y replaced with x.

6. For $X \sim \text{Geom}(p)$ and $Y \sim \text{Pois}(X)$, what is $\mathbb{E}[Y]$?

Answer: $\frac{1}{p}$

Using the law of total expectation, we can condition on all possible values of X:

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{x=1}^{\infty} \mathbb{E}[Y \mid X = x]\,\mathbb{P}(X = x) \\ &= \sum_{x=1}^{\infty} \mathbb{E}[Y \sim \text{Pois}(x)]\,\mathbb{P}(X = x) \\ &= \sum_{x=1}^{\infty} x\,\mathbb{P}(X = x) \\ &= \mathbb{E}[X] = \frac{1}{p}\end{aligned}$$

In the third line, we used the fact that the expected value of a Pois(λ) RV is just λ; after conditioning on the value of X, the parameter is now a constant. After making that substitution, we notice that the summation just becomes the expected value of X, which we know to be $\frac{1}{p}$.

7. Consider a random variable $X = 2\ln Y$ where $Y \sim \text{Uniform}[0,1]$.

   (a) What is the range of X? (The range is where the pdf of X is positive.)

   Answer: $[-\infty, 0]$

   Since Y can only take on values in $[0,1]$, it suffices to determine the range of $2\ln Y$ on the interval $Y \in [0,1]$. At $Y = 0$, we know that $2\ln Y \to -\infty$, and at $Y = 1$, we know that $2\ln Y = 0$. Since $\ln Y$ is an increasing function, we know that the range of X is $[-\infty, 0]$.

   (b) What is the pdf of X on the range defined above? (Hint: $\mathbb{P}(X \in [x, x + \mathrm{d}x]) = \mathbb{P}\left(Y \in [e^{x/2}, e^{(x+\mathrm{d}x)/2}]\right)$ and $e^{x+\mathrm{d}x} \approx e^x(1 + \mathrm{d}x)$.)

   Answer: $\frac{1}{2}e^{x/2}$

   It's actually easiest to work with the cdf rather than the pdf of X here; since we know that $X = 2\ln Y$, we can plug this into the cdf probability and rearrange:

$$\begin{aligned}\mathbb{P}(X \le x) &= \mathbb{P}(2\ln Y \le x) \\ &= \mathbb{P}\left(Y \le e^{x/2}\right) \\ &= e^{x/2}\end{aligned}$$

   The last simplification is because the cdf of $Y \sim \text{Uniform}[0,1]$ is $F_Y(y) = y$. Taking the derivative of the cdf of X, we get a pdf of $\frac{\mathrm{d}}{\mathrm{d}x}\left[e^{x/2}\right] = \frac{1}{2}e^{x/2}$.

   An alternate solution (which the hint guides you toward) is to work directly with the pdfs. Remember that when working with pdfs as probabilities, we need to include a factor of $\mathrm{d}x$ in order to convert these densities into true probabilities.

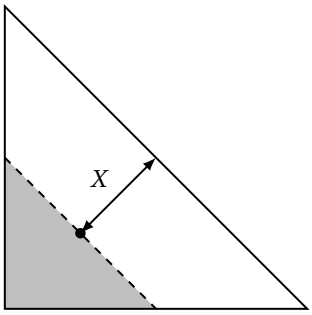
   Using the hint, we have that

$$\begin{aligned}f_X(x)\,\mathrm{d}x = \mathbb{P}(X \in [x, x + \mathrm{d}x]) &= \mathbb{P}\left(Y \in \left[e^{x/2}, e^{(x+\mathrm{d}x)/2}\right]\right) \\ &= \mathbb{P}\left(Y \in \left[e^{x/2}, e^{x/2+\mathrm{d}x/2}\right]\right) \\ &\approx \mathbb{P}\left(Y \in \left[e^{x/2}, e^{x/2}\left(1 + \frac{1}{2}\,\mathrm{d}x\right)\right]\right) \\ &= e^{x/2}\left(1 + \frac{1}{2}\,\mathrm{d}x\right) - e^{x/2} \\ &= \frac{1}{2}e^{x/2}\,\mathrm{d}x\end{aligned}$$

   Cancelling out the $\mathrm{d}x$ from both sides, we get that $f_X(x) = \frac{1}{2}e^{x/2}$.

13 Continuous: triangle

Consider a right equilateral triangle of side lengths 1, 1, and $\sqrt{2}$. Given a random point in the triangle, we define the random variable X as the distance from the hypotenuse as shown in the figure below.



1. What is the joint density function $f(x,y)$ for points inside the triangle? (Again, the point is chosen uniformly inside the entire triangle. Ignore the shading in the figure for now.)

   Answer: 2

   We want the integral of the density function over the triangle to be equal to 1. Since the point is chosen uniformly, we have a constant density function to solve for:

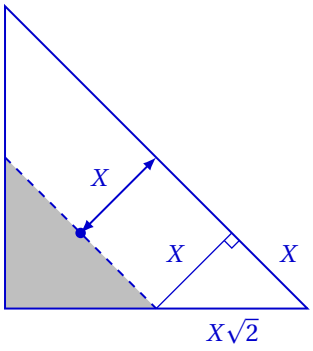
   $$1 = \iint_A f(x,y)\,dx\,dy = \iint_A c\,dx\,dy = c\iint_A dx\,dy \implies c = \frac{1}{\text{area}} = 2.$$

   Here, we used the fact that the area of the triangle is $\frac{1}{2}\cdot 1\cdot 1 = \frac{1}{2}$.

2. What is the area of the shaded triangle in terms of X? (Hint: the range of X is $\left[0, \frac{\sqrt{2}}{2}\right]$.)

   Answer: $\frac{(1-x\sqrt{2})^2}{2}$

   The best way to derive this is to draw it out:

   The main thing to notice here is that since the triangle is a 45–45–90 triangle, we have the same ratios in the smaller triangle in the bottom right (or top left) of the figure. This allows us to calculate the side length of the shaded triangle as $1 - X\sqrt{2}$.

   Now that we know the side length of the shaded triangle, we can directly find the area to be $\frac{1}{2}\left(1 - X\sqrt{2}\right)^2$.

3. What is the cdf of X for the range $x \in \left[0, \frac{\sqrt{2}}{2}\right]$?

   Answer: $1 - \left(1 - x\sqrt{2}\right)^2$

   The cdf is the probability $\mathbb{P}(X \le x)$; it describes the probability that the point lands at most x units away from the hypotenuse. This is just the complement of the shaded area found in the previous part

(scaled by the density). Hence, the probability $\mathbb{P}(X > x) = 2 \cdot (\text{shaded area}) = (1 - x\sqrt{2})^2$, and our desired probability is $1 - (1 - x\sqrt{2})^2$.

4. What is the pdf of X in the range $x \in \left[0, \frac{\sqrt{2}}{2}\right]$?

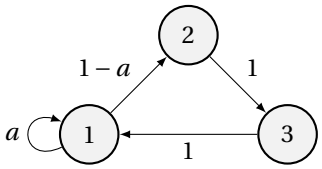
   Answer: $2\sqrt{2}\left(1 - x\sqrt{2}\right)$

   The pdf is the derivative of the cdf, so

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[1 - (1 - x\sqrt{2})^2\right] = -2(1 - x\sqrt{2}) \cdot -\sqrt{2} = 2\sqrt{2}\left(1 - x\sqrt{2}\right).$$

14 Markov Chain

Consider the following Markov chain.



1. For what value of a does the chain have a unique invariant distribution but does not always converge to it?

   Answer: $a = 0$

   The chain will always have a unique invariant distribution if it is irreducible, but if it is periodic, then not all initial distributions will converge to the invariant distribution.

   In order to make the chain periodic, we need $a = 0$ to eliminate the self loop. The resulting chain has period 3.

2. For $a = \frac{1}{2}$, what is the stationary distribution?

   Answer: $\pi(1) = \frac{1}{2}$, $\pi(2) = \pi(3) = \frac{1}{4}$

   The transition matrix is

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

   The system $\pi = \pi\mathbf{P}$ along with the restriction on the sum gives us the system of equations

$$\begin{aligned} \pi(1) &= \frac{1}{2}\pi(1) + \pi(3) \\ \pi(2) &= \frac{1}{2}\pi(1) \\ \pi(3) &= \pi(2) \\ 1 &= \pi(1) + \pi(2) + \pi(3) \end{aligned}$$

   We know that $\pi(3) = \pi(2) = \frac{1}{2}\pi(1)$, so substituting this into the last equation, we have

$$1 = \pi(1) + \pi(2) + \pi(3) = \pi(1) + \frac{1}{2}\pi(1) + \frac{1}{2}\pi(1) = 2\pi(1) \implies \pi(1) = \frac{1}{2}.$$

   This means that $\pi(2) = \pi(3) = \frac{1}{4}$ as well.

15 Small Faces

Given a planar graph with minimum degree 3 with e edges, v vertices, and f faces, we will prove there is a face of length at most 5. (The length of a face is the number of edges along it.)

1. What is the sum of the face lengths, $\sum_{i=1}^{f} s_i$, where s_i is the size of face i, in terms of e?

   Answer: $2e$

   The key here is to notice that each edge is counted twice; each edge is adjacent to two different faces. This means that summing up all of the face lengths will double count every edge, giving us a total value of $2e$.

2. Give a lower bound on e in terms of v. (Hint: the minimum degree is 3.)

   Answer: $e \geq \frac{3}{2}v$

   Recall the handshaking lemma, which says that $2e = \sum_v \deg(v)$. Since we know that the minimum degree is 3, the sum of all degrees must be at least $3v$, meaning $2e \geq 3v \implies e \geq \frac{3}{2}v$.

3. Prove that there is a face of size at most 5. (Recall Euler's formula is $v + f = e + 2$.)

   Answer: Suppose for contradiction that there does not exist any face of length at most 5. This means that every single face has length at least 6. Computing the sum of all face lengths and using the result from the first part, we have

   $$2e = \sum_{i=1}^{f} s_i \geq 6f.$$

   This means that $2e \geq 6f \implies f \leq \frac{1}{3}e$. Using Euler's formula, we arrive at

   $$v + f = e + 2 \implies v + \frac{1}{3}e \geq e + 2 \implies v \geq \frac{2}{3}e + 2 \implies \frac{2}{3}e \leq v - 2.$$

   But we've already arrived at the conclusion that $e \geq \frac{3}{2}v \implies \frac{2}{3}e \geq v$, which is a contradiction.

16 Balls and Bins

Consider placing $5n$ balls into $3n$ bins uniformly at random. (Careful, the constants in front of the n's are important.)

1. What is the expected number of empty bins?

   Answer: $3n\left(1 - \frac{1}{3n}\right)^{5n}$

   Let X represent the number of empty bins, and X_i represent whether the ith bin is empty or not. We then have

   $$\mathbb{E}[X] = \sum_{i=1}^{3n} \mathbb{E}[X_i] = 3n\mathbb{P}(X_i = 1)$$

   by linearity of expectation and since X_i is an indicator. The probability that the ith bin is empty is $\left(1 - \frac{1}{3n}\right)^{5n}$; we have $5n$ total balls, each of which has a probability $1 - \frac{1}{3n}$ of not landing in bin i.

   This means that our final answer is $3n\left(1 - \frac{1}{3n}\right)^{5n}$.

2. What is the variance of the number of empty bins?

   Answer: $3n\left(1 - \frac{1}{3n}\right)^{5n} + 2\binom{3n}{2}\left(1 - \frac{2}{3n}\right)^{5n} - \left(3n\left(1 - \frac{1}{3n}\right)^{5n}\right)^2$

With the same variables as in the previous part, the variance is

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[\left(\sum_{i=1}^{3n} X_i\right)^2\right] - \mathbb{E}[X]^2 \\ &= \sum_{i=1}^{3n} \mathbb{E}\left[X_i^2\right] + 2\sum_{i<j}^{3n} \mathbb{E}\left[X_i X_j\right] - \left(3n\left(1-\frac{1}{3n}\right)^{5n}\right)^2\end{aligned}$$

Since X_i is an indicator, $\mathbb{E}\left[X_i^2\right]$ is the same as $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1)$, as it can only be 1 or 0. Similarly, the only case where $X_i X_j = 1$ is if both $X_i = 1$ and $X_j = 1$.

$$= \sum_{i=1}^{3n} \mathbb{P}(X_i = 1) + 2\sum_{i<j}^{3n} \mathbb{P}\left(X_i = 1 \cap X_j = 1\right) - \left(3n\left(1-\frac{1}{3n}\right)^{5n}\right)^2$$

The first probability is the same as what we calculated in the previous part; there are $5n$ balls, each of which has a $1-\frac{1}{3n}$ probability of not landing in bin i. The second probability $\mathbb{P}\left(X_i = 1 \cap X_j = 1\right)$ is calculated similarly; we have $5n$ balls, each of which has a $1-\frac{2}{3n}$ probability of not landing in either bin i or bin j.

$$= 3n\left(1-\frac{1}{3n}\right)^{5n} + 2\binom{3n}{2}\left(1-\frac{2}{3n}\right)^{5n} - \left(3n\left(1-\frac{1}{3n}\right)^{5n}\right)^2$$

The factor of $\binom{3n}{2}$ is because we have $\binom{3n}{2}$ different combinations of $X_i X_j$ where $i < j$; we multiply by 2 to account for the cases where $i > j$.

3. What is the expected number of non-empty bins?

   Answer: $3n\left(1-\left(1-\frac{1}{3n}\right)^{5n}\right)$

   The number of non-empty bins is $3n - X$, where X is the number of empty bins as in the first part—each bin is either empty or non-empty. This means that we have

$$\mathbb{E}[3n - X] = 3n - \mathbb{E}[X] = 3n - 3n\left(1-\frac{1}{3n}\right)^{5n} = 3n\left(1-\left(1-\frac{1}{3n}\right)^{5n}\right).$$

4. What is the variance of the number of non-empty bins (in terms of the answer for part (1), (2), and/or (3).)

   Answer: Same as (2).

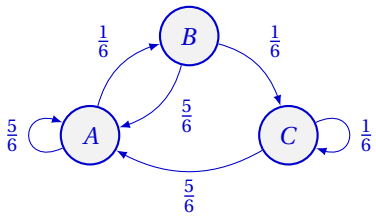
   Similar to the previous part, the number of non-empty bins is $3n - X$, so $\text{Var}(3n - X) = \text{Var}(X)$ because $3n$ is a constant and inverting an RV does not change the variance.

17 Sequential Dice

Consider rolling a die repeatedly until one gets two 6's in a row.

1. Draw a three state Markov chain where the states are labelled A, B, and C. Your chain should have a state C which is the "goal": the previous two rolls were a 6. State A should indicate that one has not rolled any die or that the previous die is not 6.

   Answer:



Here, state A represents the time in which we have not started rolling, or if we have made no progress toward two 6's in a row. State B represents the time in which we have just rolled a 6, and state C represents the time in which we have just rolled two 6's.

Transitioning from A to B or from B to C or from C to C has probability $\frac{1}{6}$, as we need to roll a 6. Otherwise, all other transitions have a probability of $\frac{5}{6}$, for every other value.

2. What is the expected number of rolls to roll two 6's in a row?

Answer: 42

Let $\beta(i)$ represent $\mathbb{E}[\text{rolls until two 6's} \mid \text{at state } i]$. This means that we have the system of equations:

$$\begin{aligned}\beta(A) &= 1 + \frac{1}{6}\beta(B) + \frac{5}{6}\beta(A)\\ \beta(B) &= 1 + \frac{5}{6}\beta(A) + \frac{1}{6}\beta(C)\\ \beta(C) &= 0\end{aligned}$$

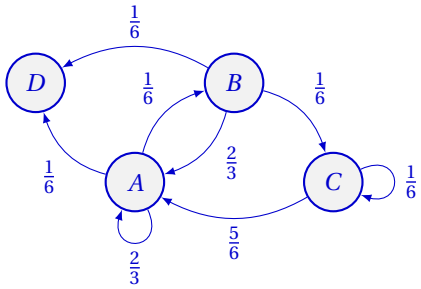
Plugging in the equation for $\beta(B)$ into the equation for $\beta(A)$, and eliminating $\beta(C)$, we have

$$\begin{aligned}\beta(A) &= 1 + \frac{1}{6}\left(1 + \frac{5}{6}\beta(A)\right) + \frac{5}{6}\beta(A)\\ &= 1 + \frac{1}{6} + \frac{5}{36}\beta(A) + \frac{5}{6}\beta(A)\\ \frac{25}{36}\beta(A) &= \frac{7}{6}\\ \beta(A) &= 42\end{aligned}$$

3. What is the probability of rolling two 6's in a row prior to rolling a 6? (Hint: add a state to your previous Markov Chain and do a computation.)

Answer: $\frac{1}{8}$

Suppose we add a state D, representing the time in which we have rolled a 5. We can transition to state D from state A and state B with probability $\frac{1}{6}$, meaning the transitions that had probability $\frac{5}{6}$ prior now have probability $\frac{4}{6} = \frac{2}{3}$:



Now, let $\alpha(i)$ represent $\mathbb{P}(C \text{ before } D \mid \text{at state } i)$. This means that we have the system of equations:

$$\begin{aligned}\alpha(A) &= \frac{2}{3}\alpha(A) + \frac{1}{6}\alpha(B) + \frac{1}{6}\alpha(D)\\ \alpha(B) &= \frac{2}{3}\alpha(A) + \frac{1}{6}\alpha(C) + \frac{1}{6}\alpha(D)\\ \alpha(C) &= 1\\ \alpha(D) &= 0\end{aligned}$$

Keeping in mind the values of $\alpha(C)$ and $\alpha(D)$, we can plug in the expression for $\alpha(B)$ into the equation for $\alpha(A)$:

$$\begin{aligned}\alpha(A) &= \frac{2}{3}\alpha(A) + \frac{1}{6}\left(\frac{2}{3}\alpha(A) + \frac{1}{6}\right)\\ &= \frac{2}{3}\alpha(A) + \frac{1}{9}\alpha(A) + \frac{1}{36}\\ \frac{2}{9}\alpha(A) &= \frac{1}{36}\\ \alpha(A) &= \frac{1}{8}\end{aligned}$$

18 Bayes Rule

A doctor has information that 80% of the sick children in a neighborhood have the flu and the other 20% of sick children have measles. The further knows that the probability of a rash with measles is 0.95, and that the probability of a rash with flu is 0.10. If a sick child has a rash, what is the probability the child has measles? (Show your work here, and use the box for your final answer.)

Answer: $\frac{19}{27}$

Let M denote the event that a child has measles, let F denote the event that a child has the flue, and let R denote the event that a child has a rash.

Converting the sentences into probabilities, we have

$$\mathbb{P}(F) = 0.8 \qquad \mathbb{P}(R \mid M) = 0.95$$
$$\mathbb{P}(M) = 0.2 \qquad \mathbb{P}(R \mid F) = 0.1$$

Our goal is to find $\mathbb{P}(M \mid R)$; we can use Bayes' rule with the Law of Total Probability to switch the condition and evaluate the probability:

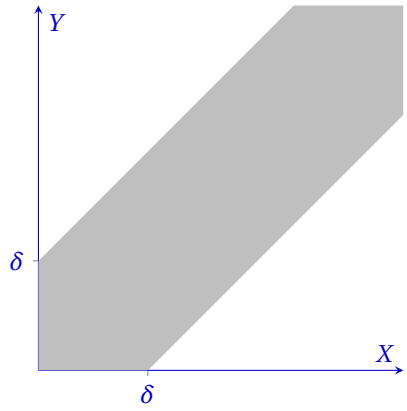
$$\begin{aligned}\mathbb{P}(M \mid R) &= \frac{\mathbb{P}(R \mid M)\mathbb{P}(M)}{\mathbb{P}(R \mid M)\mathbb{P}(M) + \mathbb{P}(R \mid F)\mathbb{P}(F)}\\ &= \frac{0.95 \cdot 0.2}{0.95 \cdot 0.2 + 0.1 \cdot 0.8}\\ &= \frac{0.19}{0.19 + 0.08} = \frac{0.19}{0.27} = \frac{19}{27}\end{aligned}$$

19 Close enough!

Given a circle (dartboard) of radius 1, choose two points at random on the dartboard uniformly, and let X and Y define the distance to the center. What is the probability that $|X - Y| \leq \delta$? (Recall that the pdf of both variables is $f(x) = 2x$ for $x < 1$.)

Answer: $\frac{1}{3}\delta^4 - 2\delta^2 + \frac{8}{3}\delta$

It is easiest to look at this on a graph:



X and Y are defined only on the unit square, and the probability we are trying to find is $\mathbb{P}(|X-Y| \le \delta)$, or the shaded area above. While we could integrate the joint pdf over this area, it is easiest to calculate the complement probability—that is, we can integrate over the two unshaded triangles.

Another thing to notice is that we only have to integrate over one triangle, as we can multiply by two (because the joint pdf is symmetric about $y = x$). Taking the lower right triangle and noting that our joint pdf is $f(x,y) = f(x) \cdot f(y) = 4xy$, our integral would be

$$\begin{aligned}
2\int_0^{1-\delta}\int_{y+\delta}^{1} 4xy\,\mathrm{d}x\,\mathrm{d}y &= 2\int_0^{1-\delta} \left(2x^2y\right)\Big|_{y+\delta}^{1}\mathrm{d}y \\
&= 2\int_0^{1-\delta} 2y - 2y(y+\delta)^2\,\mathrm{d}y \\
&= 4\int_0^{1-\delta} y - y^3 - 2\delta y^2 - \delta^2 y\,\mathrm{d}y \\
&= 4\left(\frac{1}{2}y^2 - \frac{1}{4}y^4 - \frac{2\delta}{3}y^3 - \frac{\delta^2}{2}y^2\right)\Big|_0^{1-\delta} \\
&= 2(1-\delta)^2 - (1-\delta)^4 - \frac{8\delta}{3}(1-\delta)^3 - 2\delta^2(1-\delta)^2
\end{aligned}$$

The calculations after this point are just expanding and simplifying the expression. The calculation is left here as a possible way of simplifying(it is quite tedious).

$$\begin{aligned}
&= (1-\delta)^2\left(2 - (1-\delta)^2 - \frac{8\delta}{3}(1-\delta) - 2\delta^2\right) \\
&= (1-\delta)^2\left(2 - 1 + 2\delta - \delta^2 - \frac{8}{3}\delta + \frac{8}{3}\delta^2 - 2\delta^2\right) \\
&= \left(1 - 2\delta + \delta^2\right)\left(1 - \frac{2}{3}\delta - \frac{1}{3}\delta^2\right) \\
&= -\frac{1}{3}\delta^4 - \frac{2}{3}\delta^3 + \delta^2 + \frac{2}{3}\delta^3 + \frac{4}{3}\delta^2 - 2\delta - \frac{1}{3}\delta^2 - \frac{2}{3}\delta + 1 \\
&= -\frac{1}{3}\delta^4 + 2\delta^2 - \frac{8}{3}\delta + 1
\end{aligned}$$

We still need to take the complement of this result, as this integral is for the two unshaded triangles. Taking the complement, we have our final answer of $\frac{1}{3}\delta^4 - 2\delta^2 + \frac{8}{3}\delta$.

20 Puzzler

Consider the following game on an $n \times m$ grid, with two cooperating players. A key is hidden under a grid square and on each square there is a single coin that is either heads or tails. Player 1 knows the key location and *must flip exactly one coin.*

Player 2 should observe the pattern of heads and tails and produce the key location.

To reiterate, from an arbitrary initial setup of heads and tails on the grid, player 1 should flip exactly one coin to make a setup where player 2 can determine the location of the hidden key.

1. What is a strategy for the players to win on a 2×1 grid? (Hint: Think about $2x + y \pmod 2$ for $x, y \in \{0, 1\}$ and think of heads as 1 and 0 as tails.)

   Answer: Let $x = 1$ denote that the left square is heads, and $y = 1$ denote that the right square is heads. Further, let $z = 2x + y \pmod 2$; this means that z can either be 0 or 1. Suppose the players agree beforehand that $z = 1$ means that the key is in the left square, and $z = 0$ means that the key is in the right square.

   The key insight here is that changing x does not change the value of $2x + y \pmod 2$, whereas changing y flips z. This means that if the key is already in the correct location based on the initial arrangement of the coins, player 1 can just flip the left coin (i.e. flip x), leaving the value of z unchanged. If the key is in the incorrect location, then player 1 can flip the right coin (i.e. flip y), flipping the value of z and allowing player 2 to find the right location.

2. What is a strategy for the players to win on a $2^k \times 2^k$ grid? (Hint: use induction to find the column and row of the coin to flip. (Notice $2^k = 2 \times 2^{k-1}$.)

   Answer: See the following videos for an explanation better than anything I could give (it was a collaboration between 3Blue1Brown and Matt Parker):

   3Blue1Brown: `https://youtu.be/wTJI_WuZSwE`

   Matt Parker: `https://youtu.be/as7Gkm7Y7h4`

   The essence of it is that we can use something similar to a Hamming code to encode the position of the keys, using parities of smaller subgrids. Identifying the flipped coin is just the same process of error identification for any typical Hamming code.