1 Readings

STUDY IN-DEPTH... ...the posted notes for lectures 8-15 including the supplements. (See below for details about lecture 8.)

STUDY IN-DEPTH... ...the posted guides for recitations 4 and 5.

STUDY IN-DEPTH... ...the posted solutions to homeworks 4 and 5. Compare with your own solutions.

STUDY IN-DEPTH... ...the solutions to the mock exam and the additional problems contained in this document will be posted on Sunday, June 30th, in the morning. Until then, try very hard to solve these on your own.

For this exam, you must review carefully some material from lecture 8, specifically everything about the Monty Hall problem.

2 Not cumulative?

Superficially, Exam 2 is not cumulative. Yet, you cannot solve most problems from part 2 of this course without knowing how to do proofs, knowing counting techniques, knowing the basics of probability and having an understanding of mathematical notation. All these were covered in lecture 1-8 so they do not appear on the readings list above. However, if you still have gaps in understanding the material from part 1 now is a good time to close them.

3 Memorize!

Find and memorize formulas:

- For the expectation and for the variance of a Bernoulli random variable.
- For the expectation and for the variance of a binomial random variable.

4 Mock Exam (120 minutes for 240 points)

1. (20 pts)

   (a) Draw a connected undirected graph with 6 nodes and exactly 2 cut edges.
   (b) Draw an acyclic undirected graph with 5 edges and 7 nodes.
(c) Draw a connected undirected graph in which every node has degree 3.
(d) Draw a digraph that has no sources and no sinks and has strictly more edges than vertices.

**Answer**

(a) Draw a connected graph with 6 nodes and exactly 2 cut edges.

(b) Draw an acyclic graph with 5 edges and 7 nodes.

(c) Draw a connected graph in which every node has degree 3.

2. (35 pts)

Each of the parts in this problem is 5-10pts.

(a) Let $G = (V, E)$ be an undirected graph in which every node has degree 3. Prove that $|E|/3 = |V|/2$.

(b) Prove or disprove: if we add one edge to a tree between any two existing vertices the resulting graph cannot be bipartite.

(c) Let $X, Y, Z$ be random variables on the same probability space such that $Z = X + Y - XY$. Show that if $X$ and $Y$ are Bernoulli random variables, then $Z$ is also a Bernoulli random variable.

(d) Let $A, B$ be events in the same probability space and let $I_A, I_B$ be their indicator random variables. Prove that if $E(I_A + I_B) = 1$ then $P(A) = P(\overline{B})$.

(e) For any positive integer $n \geq 2$, if the complete bipartite graph $K_{2,n}$ has a cycle of length $n + 2$ then $n$ must be even.

**Answer**
(a) **TRUE.** By the Handshaking Lemma, we have

\[ 2|E| = \sum_{v \in V} \deg v \]

\[ = \sum_{v \in V} 3 \quad \text{(because } \deg v = 3 \text{ for all } v) \]

\[ = 3|V|. \]

Dividing through by 6, we have

\[ \frac{|E|}{3} = \frac{|V|}{2}. \]

(b) **FALSE.** Here is our original tree

```
1
  / \  \
2   3
   /\  |
  4  \
```

We can draw an edge between nodes 2 and 4 to get the complete bipartite graph $K_{2,2}$

```
1 2
\|
4 3
```

(c) **TRUE.** It suffices to show that $\text{Val}(Z) = \{0, 1\}$. There are four cases for $(X, Y) : (0, 0), (0, 1), (1, 0), (0, 0)$. When $(X, Y) = (0, 0)$

\[ Z = 0 + 0 - 0 \cdot 0 = 0. \]

When $(X, Y) = (1, 0)$

\[ Z = 1 + 0 - 1 \cdot 0 = 1. \]

When $(X, Y) = (0, 1)$

\[ Z = 0 + 1 - 0 \cdot 1 = 1. \]

When $(X, Y) = (1, 1)$

\[ Z = 1 + 1 - 1 \cdot 1 = 1. \]

Hence, we know that $\text{Val}(Z) = \{0, 1\}$, so $Z$ is a Bernoulli random variable.
(d) TRUE. By LOE and the definition of indicator random variables, we have
\[ \Pr[B] + \Pr[B] = 1 \]  
(def. of complement)
\[ = \mathbb{E}[I_A + I_B] \]  
(given)
\[ = \mathbb{E}[I_A] + \mathbb{E}[I_B] \]  
(by LOE)
\[ = \Pr[A] + \Pr[B] \]  
(Slide 2, Lecture 20.)

Hence, we know \( \Pr[A] = \Pr[B] \).

(e) TRUE. By Slide 13 in Lecture 22, we know that a bipartite graph has no odd length cycles. Then, we know \( n + 2 \) is even, i.e. \( n + 2 = 2k \) for some positive integer \( k \). Solving for \( n \), we see that
\[ n = 2(k - 1). \]

Since \( k - 1 \in \mathbb{Z} \), we know \( n \) must be even.

3. (25pts)
A fair coin is flipped twice. Let \((\Omega, \mathbb{P})\) be the resulting probability space. Let \( X_H \) be random variable defined on \( \Omega \) that returns the number of heads observed and \( X_T \) similarly the number of tails observed.

(a) Describe the probability space \((\Omega, \mathbb{P})\). That is, list the outcomes and their probabilities.

(b) Show that the random variable \( Z \) defined by \( \forall w \in \Omega \quad Z(w) = X_H(w) \cdot X_T(w) \)
is a Bernoulli random variable and find its probability of success.

(c) Show that \( \mathbb{E}[Z] \neq \mathbb{E}[X_H] \mathbb{E}[X_T] \).

Answer

(a) Our sample space is all possible outcomes of two coin tosses. Therefore,
\[ \Omega = \{HH, HT, TH, TT\} \]

Since we are flipping a fair coin, the probability space is uniform, so each outcome has a probability of \( \frac{1}{|\Omega|} = \frac{1}{4} \). Alternatively, since each coin flip has \( \Pr[H] = \Pr[T] = \frac{1}{2} \) and the coin flips are independent, the probability of an outcome in our sample space is simply \( \left(\frac{1}{2}\right)^2 = \frac{1}{4} \).

(b) Observe that \( X_H \) and \( X_T \) are defined as follows:
\[ X_H(w) = \begin{cases} 0 & w = TT \\ 1 & w \in \{HT, TH\} \\ 2 & w = HH \end{cases} \]
\[ X_T(w) = \begin{cases} 0 & w = HH \\ 1 & w \in \{HT, TH\} \\ 2 & w = TT \end{cases} \]

Now in order to find the distribution of \( Z(w) = X_H(w) \cdot X_T(w) \), we plug in each \( w \in \omega \):
\[ Z(HH) = X_H(HH) \cdot X_T(HH) = 2 \cdot 0 = 0 \]
\[ Z(HT) = X_H(HT) \cdot X_T(HT) = 1 \cdot 1 = 1 \]
\[ Z(TH) = X_H(TH) \cdot X_T(TH) = 1 \cdot 1 = 1 \]
\[ Z(TT) = X_H(TT) \cdot X_T(TT) = 0 \cdot 2 = 0 \]
Therefore, we have:

\[ Z(w) = \begin{cases} 
0 & w \in \{HH, TT\} \\
1 & w \in \{HT, TH\} 
\end{cases} \]

Since \( Z \) only takes on the values 1 and 0, we can define success as the outcomes in \( \{HT, TH\} \), and failure as the outcomes in \( \{HH, TT\} \). This means that \( Z \) is a Bernoulli random variable with a success probability \( p \) equal to \( \Pr[Z = 1] \):

\[ p = \Pr[Z = 1] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \]

Therefore, \( Z \) is a Bernoulli random variable with probability success of \( p = \frac{1}{2} \).

(c) We use our answer to part b to calculate the expectation of \( Z, X_H, \) and \( X_T \).

\[
E[Z] = 0 \cdot \Pr[Z = 0] + 1 \cdot \Pr[Z = 1] = 0 + 1 \cdot \frac{1}{2} = \frac{1}{2}
\]

\[
E[X_H] = 0 \cdot \Pr[X_H = 0] + 1 \cdot \Pr[X_H = 1] + 2 \cdot \Pr[X_H = 2] = 0 + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1
\]

\[
E[X_T] = 0 \cdot \Pr[X_T = 0] + 1 \cdot \Pr[X_T = 1] + 2 \cdot \Pr[X_T = 2] = 0 + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1
\]

Therefore, we can conclude that

\[ E[Z] = \frac{1}{2} \neq 1 \cdot 1 = E[X_H] \cdot E[X_T] \]

4. (20 pts) Let \( B \) be a binomial random variable with parameters \( p \) and \( n \). Show that \( R = n - B \) is also a binomial random variable and calculate its parameters.

**Answer**

We prove that \( R \) is a binomial random variable with parameters \( 1 - p \) and \( n \). Note the following:

\[
\text{Val}(R) = \{n - b \mid b \in \text{Val}(B)\} = \{n - b \mid b \in [0..n]\} = \{n - 0, n - 1, \ldots, n - n\} = [0..n],
\]

which is the value-set of a binomial r.v. with parameter \( n \). Additionally, for \( k \in [0..n] \), we have

\[
\Pr[R = k] = \Pr[n - B = k] = \Pr[B = n - k] = \binom{n}{n-k} p^{n-k} (1-p)^{n-(n-k)} \quad \text{(by the formula for the distribution of a binomial r.v.)}
\]

\[
= \binom{n}{k} (1-p)^k p^{n-k},
\]

which is the distribution of a binomial random variable with parameters \( 1 - p \) and \( n \), as desired.

5. (20 pts) Let \( T \) be a tree with at least 3 vertices. Assume that every vertex in \( T \) has either degree 3 or is a leaf. Let \( L \) be the set of leaves of \( T \) and let \( R \) be the set of vertices in \( T \) that have degree 3. Show that

\[ |R| = |L| - 2 \]
The total number of vertices is the number of the leaves (|L|) plus the number of the vertices in T with degree three (|R|). We thus see that:

|V| = |L| + |R|

We can find the total degree of the tree as follows:

\[ \sum_{v \in V} \deg(v) = \sum_{v \in L} \deg(v) + \sum_{v \in R} \deg(v) = \sum_{v \in L} 1 + \sum_{v \in R} 3 = |L| + 3|R| \]

However, we also know by the Handshaking Lemma that:

\[ \sum_{v \in V} \deg(v) = 2|E| = 2(|V| - 1) \quad (|E| = |V| - 1, \text{ as } T \text{ is a tree}) \]

Combining these two facts, we conclude that:

\[ |L| + 3|R| = 2(|V| - 1) \]
\[ |L| + 3|R| = 2(|L| + |R| - 1) \quad \text{(plugging in } |V| = |L| + |R|) \]
\[ |L| + 3|R| = 2|L| + 2|R| - 2 \]
\[ |R| = |L| - 2 \]

6. (20 pts) Prove by induction that any tree with at least 3 vertices must have at least one vertex of degree \( \geq 2 \). (Only proofs by induction will receive credit.)

**Answer**

**(BASE CASE):** \( n = 3 \). We note that there is only one tree with 3 vertices, \( P_3 \). We see that \( P_3 \) has an internal vertex with degree 2.

**(INDUCTION STEP):** Let \( k \in \mathbb{Z}^+, \ k \geq 3 \). Assume (IH) that any tree with \( k \) vertices must have at least one vertex of degree \( \geq 2 \). We now wish to show that any tree with \( k + 1 \) vertices must have at least one vertex of degree \( \geq 2 \).

Let \( T \) be a tree with \( k + 1 \) vertices. Since we know that \( k \geq 3 \), we have that \( T \) has a nonzero number of edges. The lemma on slide 22 of lecture 12 gives us that \( T \) has at least one leaf, call it \( v \). Remove \( v \) from \( T \) to form a new graph \( T' \). By the lemma on slide 24 of lecture 12, \( T' \) is a tree with \( k \) vertices. By the Induction Hypothesis, we know \( T' \) has at least one vertex \( u \) with \( \deg(u) \geq 2 \).

Now consider adding \( v \) back to \( T' \) to form \( T \). Adding back a vertex cannot decrease \( u \)'s degree in \( T \). Therefore, \( \deg(u) \geq 2 \) still holds and \( T \) has at least one vertex with degree \( \geq 2 \) (namely \( u \)). Thus, we have shown our claim is true when \( n = k + 1 \), concluding our Induction Step and completing our proof.
7. (35pts) Alice has a *fair* coin that shows the number 2 on one side and the number 3 on the other. Bob has a *fair* tetrahedral die (a tetradie) that shows the numbers 1, 2, 3, and 4 on its four faces. They play the following game:

- Alice flips the coin showing the number \( a \) and, independently, Bob rolls the tetradie showing the number \( b \)
- If \( a > b \) then Alice wins and Bob pays Alice \( a - b \) dollars. If \( a = b \) then it’s a tie and no money changes hands. If \( b > a \) then Bob wins and Alice pays Bob \( b - a \) dollars.

(a) Draw the tree of possibilities for a single game.

(b) Compute the probability that Alice wins a single game.

(c) Suppose that Alice and Bob play the game 3 times in a row, independently. Assume that Alice starts with 10 dollars. Let \( Z \) be the random variable that returns the amount of dollars that Alice has after these 3 games. Compute \( E[Z] \).

**Answer**

(a) We first note that our sample space consists of ordered pairs, where the first element corresponds to the result of Alice’s coin flip and the second corresponds to the result of Bob’s die roll. More formally, we see that:

\[
\Omega = \{ (a, b) \mid a \in \{2, 3\}, b \in [1..4] \}
\]

We define the event of Alice flipping 2 and 3 to be \( A_2 \), \( A_3 \), respectively, and the event of Bob rolling 1, 2, 3, and 4 to be \( B_1 \), \( B_2 \), \( B_3 \), \( B_4 \), respectively.

The tree of possibilities is thus:
(b) Let the event Alice wins be $W_A$. We seek $\Pr[W_A]$. As seen in the above tree, Alice wins when she flips a 3 and Bob rolls either a 1 or 2, and when she flips a 2 and Bob rolls a 1. These are disjoint events. We find

$$\Pr[W_A] = \Pr[(A_3 \cap B_2) \cup (A_3 \cap B_3)]$$

$$= \Pr[A_3 \cap B_2] + \Pr[A_3 \cap B_3] + \Pr[A_2 \cap B_1]$$

$$= \Pr[A_3] \cdot \Pr[B_2] + \Pr[A_3] \cdot \Pr[B_3] + \Pr[A_2] \cdot \Pr[B_1]$$

$$= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}$$

$$= \frac{3}{8}$$

(c) Let $Y_1, Y_2, Y_3$ be the amount Alice gains in games 1, 2, and 3 respectively. Since each game has the same rules, $E[Y_1] = E[Y_2] = E[Y_3]$. Further, $Z = 10 + Y_1 + Y_2 + Y_3$. By the Linearity of Expectation,

Now we simply apply the definition of expectation:

\[ E[Y_1] = \sum_{w \in \Omega} \Pr[w] Y_1(w) \]

\[ = \Pr[A_2 \cap B_1](2 - 1) + \Pr[A_2 \cap B_2](2 - 2) + \Pr[A_2 \cap B_3](2 - 3) + \Pr[A_2 \cap B_4](2 - 4) \]
\[ + \Pr[A_3 \cap B_1](3 - 1) + \Pr[A_3 \cap B_2](3 - 2) + \Pr[A_3 \cap B_3](3 - 3) + \Pr[A_3 \cap B_4](3 - 4) \]
\[ = \frac{1}{8}(1) + \frac{1}{8}(0) + \frac{1}{8}(-1) + \frac{1}{8}(-2) + \frac{1}{8}(2) + \frac{1}{8}(1) + \frac{1}{8}(0) + \frac{1}{8}(-1) \]
\[ = \frac{1}{8}(1 + 0 - 1 - 2 + 2 + 1 + 0 - 1) \]
\[ = 0 \]

Alternatively, we could have noticed that \( Y_1 = a - b \), and applied linearity again to have

\[ E[Y_1] = E[a] - E[b] \]

\[ = \sum_{w \in \Omega} \Pr[w]a(w) - \sum_{w \in \Omega} \Pr[w]b(w) \]
\[ = \left( \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 3 \right) - \left( \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 3 + \frac{1}{4} \cdot 4 \right) \]
\[ = \frac{5}{2} - \frac{5}{2} \]
\[ = 0 \]

Plugging in to our original equation,

\[ E[Z] = 10 + 3E[Y_1] = 10 + 3 \cdot 0 = 10 \]

8. (20 pts) Let \( n \geq 3 \) be a positive integer. Consider \( K_{3,n} \), the complete bipartite graph with 3 red nodes and \( n \) blue nodes.

(a) Consider a cycle of length 6 in \( K_{3,n} \). How many blue nodes must such a cycle have? Explain your answer.

**Answer**

Since \( K_{3,n} \) is bipartite, every edge in \( K_{3,n} \) must have different-colored endpoints. In other words, the vertices along any walk, including the cycle, must alternate colors. Since there are 6 total vertices in a cycle of length 6, exactly half of the vertices must be blue, for a total of \( \left\lceil \frac{3}{2} \right\rceil = 3 \) blue vertices.

(b) Count the number of paths of length 3 in \( K_{3,n} \).

**Answer**

Note that paths of length 3 contain 4 vertices. Observe that the two endpoints must be different colors. Thus, we can count by starting at the red end, and building up the path. We apply the Multiplication Rule:

*Step 1:* Choose a red endpoint. (3 ways)

*Step 2:* Choose a blue vertex. (\( n \) ways)

*Step 3:* Choose a different red vertex. (2 ways)

*Step 4:* Choose a different blue vertex. (\( n - 1 \) ways)
Observe that each of these steps is valid, since the edges will always exist between selected
vertices in $K_{3,n}$. Additionally, we are not missing any paths, since any path of length 3 must be
of the form red–blue–red–blue (direction does not matter, since we are working in an undirected
graph, so it would be the same if you started by choosing a blue vertex instead of a red one).
By Multiplication Rule, there are $6n(n - 1)$ such paths.

9. (30 pts) Let’s call Peano-digraph a digraph in which every vertex has outdegree 1.

(a) Prove that any Peano-digraph that is strongly connected is, in fact, a directed cycle.

**Answer**

Let $G$ be a strongly connected Peano-digraph. $G$ cannot have a sink because a sink has outdegree
0. Hence, $G$ cannot be a DAG, because it would have to have at least one sink, as proved in
lecture.

Therefore $G$ has a directed cycle $C$ inside it. We will prove by contradiction that all the nodes
must be in $C$ (hence the whole graph coincides with $C$).

Suppose, toward a contradiction that $u$ is a vertex in $G$ that is not in $C$. Let $v$ be a vertex
in $C$. By strong connectivity, there is a path $v \rightarrow u$. Let $x \rightarrow y$ be the first edge along this path
which is not part of the cycle $C$ (this edge exists because $v$ is in $C$ but $u$ is not in $C$). Then $x$
is in $C$ but $y$ is not in $C$. Since $x$ already has a successor in $C$, it must now have two distinct
successors which contradicts the fact that the outdegree of $x$ is 1.

(b) Count the number of different Peano-digraphs whose set of vertices is $[1..n]$, where $n$ is a positive
integer?

**Answer**

Each vertex is the start of exactly 1 edge. We can choose the end to be any of the $n$ vertices
for each of these edges. Thus, there are a total of $n^n$ Peano-digraphs.

Another way to count is to notice that there is a 1-1 correspondence between the set of Peano-
digraphs with vertices $[1..n]$ and the set of functions with domain $[1..n]$ and codomain $[1..n]$.

10. (15 pts) Let $G = (V,E)$ be an undirected graph with $|V| = n \geq 3$ vertices and satisfying the following
property. For any $u, v, w \in V$, distinct vertices, at least one of these three is adjacent to the other
two, that is,

$\quad v-u-w \quad \text{OR} \quad u-v-w \quad \text{OR} \quad u-w-v$

(Note that it is also possible that the 3 vertices are pairwise adjacent.) Prove that $G$ has at least
$\frac{n^2}{2} - n$ edges.

**Answer**

We first show that every vertex has degree at least $n - 2$. Assume otherwise, that is, there exists a
vertex $v \in V$ with at most $n - 3$ neighbors. Consider $v$ and two of its non-neighbors. Since $v$ is not
adjacent to either of these two vertices, the given condition cannot be satisfied for this group of 3.
From this, we use the Handshake Lemma to show that:

$$|E| = \frac{1}{2} \sum_{v \in V} \deg(v)$$

$$\geq \frac{1}{2} \sum_{v \in V} (n - 2)$$

$$= \frac{1}{2} n(n - 2)$$

$$= \frac{n^2}{2} - n$$

Alternate Solution:

Let $P(n)$ be:

Any graph $G = (V, E)$ with $|V| = n$ vertices where, for any three distinct vertices $u, v, w \in V$, at least one of $u, v, w$ is adjacent to the other two, then $G$ has at least $\frac{n^2}{2} - n$ edges.

We proceed with a proof by strong induction on $n$ (see note below why strong induction is needed).

(BASE CASE): $n = 3, 4$.

For $n = 3$, we observe that picking all three vertices means that at least two edges have to be present. $2 \geq \frac{3^2}{2} - 3 = \frac{3}{2}$, so the condition holds.

For $n = 4$, let the four vertices be $a, b, c, d$. We then wish to show that any graph satisfying the condition contains at least $\frac{n^2}{2} - 4$ edges.

Consider the triple $\{a, b, c\}$. By the condition in the problem, at least two edges must occur between these three vertices. First suppose all three edges occur, meaning $a, b, c$ form a triangle. Then consider the triple $\{a, b, d\}$; $d$ must have at least one edge to $a$ or $b$, so there are at least 4 edges in total.

In the case where there are exactly two edges, suppose WLOG these edges are $\{a, b\}$ and $\{a, c\}$. Then consider the triple $\{b, c, d\}$. By the condition in the problem, at least two other edges exist amongst these three vertices, thus showing us that there are at least 4 edges in total.

(INDUCTION STEP): Let $k \in \mathbb{Z}^+$, $k \geq 4$. Assume (IH) that $P(j)$ holds, for all integers $3 \leq j \leq k$. We want to show that $P(k+1)$ holds.

Consider a graph $G = (V, E)$ with $k + 1$ vertices where, for any three distinct vertices $u, v, w \in V$, at least one of $u, v, w$ is adjacent to the other two. We consider the following cases:

Case 1: For any pair of distinct vertices $s, t \in V$, the edge $\{s, t\}$ is in $G$.

In this case, we see that $G \simeq K_{k+1}$. Thus, we know that

$$|E| = \binom{n}{2} = \frac{n(n-1)}{2} = \frac{n^2}{2} - \frac{n}{2} > \frac{n^2}{2} - n$$

as desired.

Case 2: There exists a pair of distinct vertices $s, t \in V$ such that the edge $\{s, t\}$ is not in $G$. 

11
Let $G'$ be the graph formed by removing $s$ and $t$ from $G$. Note that $G'$ is a graph on $k - 1$ vertices where between any three distinct vertices $u, v, w$, at least one of $u, v, w$ is adjacent to the other two. Thus, we can apply our strong IH to determine that $G'$ has at least $\frac{(k - 1)^2}{2} - (k - 1)$ edges.

Now consider vertices $s$ and $t$. In order for our condition to be satisfied for $G'$, we know that every other vertex $w \in V \setminus \{s, t\}$ must have edges to both $s$ and $t$; indeed, if some vertex $x \in V \setminus \{s, t\}$ did not have edges to both $s$ and $t$, then the triple $\{s, t, x\}$ would violate the condition of the problem. Noting there are at least $k - 1$ other vertices in $G$ and that every edge in $G'$ also exists in $G$, we see:

$$|E| \geq \frac{(k - 1)^2}{2} - (k - 1) + 2(k - 1)$$
$$= \frac{k^2}{2} - k + \frac{1}{2} - k + 1 + 2k - 2$$
$$= \frac{k^2}{2} - \frac{1}{2}$$
$$= \frac{k^2 + 2k + 1}{2} - \frac{2k + 2}{2}$$
$$= \frac{(k + 1)^2}{2} - (k + 1)$$

as desired.

Thus, we have shown our claim is true when $n = k + 1$ in both cases, concluding our Induction Step and completing our proof.

**Note:** In order to show the above claim, we needed a strong Induction Hypothesis; in other words, we showed that $\forall k \in \mathbb{Z}^+, k \geq 4, \left( (P(k - 1) \land P(k)) \implies P(k + 1) \right)$. One may wonder if we could have performed the same proof using ordinary induction instead. Say we removed a single vertex, say $v$, from $G$ to get a graph $G'$ on $k$ vertices, which by the IH has at least $\frac{k^2}{2} - k$ edges. In order to complete the proof, we would need to be able to show that $v$ has at least $k - \frac{1}{2}$ edges, since

$$\frac{k^2}{2} - k + \left( k - \frac{1}{2} \right) = \frac{k^2}{2} - \frac{1}{2} = \frac{(k + 1)^2}{2} - (k + 1)$$

similar to above. However, this would mean that we would need to show that $v$ has exactly $k$ edges, since the degree of any vertex must be an integer. However, this is not necessarily always the case – there are examples of graphs satisfying the property in our claim where no vertex has degree $n - 1$. A simple example is $C_4$. Thus, we know that we cannot actually complete this proof, and must use strong induction instead.

5. **Additional Problems**

1. For each statement below, decide whether it is TRUE or FALSE. In each case attach a very brief explanation of your answer.

   (a) Let $X$ be a Bernoulli random variable such that $\text{Var}[X] = 0.2 \cdot E[X]$. Then, the probability of success for $X$ is 0.4.
(b) There exists a random variable $X$ for which $E[X^2] < (E[X])^2$.
(c) Let $A, B$ be events in a probability space such that $\Pr[A] = 0$ and $\Pr[B] \neq 0$. Then, $\Pr[A \mid B] = 0$, true or false?
(d) For any probability space $(\Omega, P)$ and any event $A \subseteq \Omega$ such that $\Pr[A] \neq 0$ we have $\Pr[\Omega \mid A] = \Pr[A \mid \Omega]$, true or false?
(e) If $X_1$ and $X_2$ are Bernoulli random variables with $\Pr[X_1 = 1] = 1/2$ and $\Pr[X_2 = 1] = 1/3$ then $E[X_1 - X_2] = 0$.
(f) For any two events $A, B$ in the same probability space $(\Omega, P)$ such that $\Pr[B] \neq 0$ we have $\Pr[A \cup B \mid B] = 1$.
(g) There exists an undirected graph $G$ with 4 vertices such that both $G$ and its complement, $\overline{G}$, are connected.

**Answer**

(a) **FALSE**. We have $\Var[X] = p - p^2 = 0.2 \cdot p$, so $p = 0.8$.

(b) **FALSE.** We claim that $(E[X])^2 \leq E[X^2]$. Recall that

$$
\Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2,
$$

However, $(X - E[X])^2$ takes only non-negative values so its expectation is non-negative. Hence, $\Var[X]$ is always non-negative. It follows that $(E[X])^2 \leq E[X^2]$.

(c) **TRUE.**

Since $A \cap B \subseteq A$, it follows by monotonicity of probability that $0 \leq \Pr[A \cap B] \leq \Pr[A] = 0$. Hence $\Pr[A \cap B] = 0$. Therefore, we have

$$
\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = 0
$$

(d) **FALSE**. We proceed with a disproof by counterexample.

Let $(\Omega, P)$ be the probability space of one flip of a fair coin. Further, let $A$ be the event that the coin shows heads.

$$
\Pr[\Omega \mid A] = \frac{\Pr[\Omega \cap A]}{\Pr[A]} = \frac{\Pr[A]}{\Pr[A]} = 1
$$

(Because $A \subseteq \Omega$)

$$
\Pr[A \mid \Omega] = \frac{\Pr[A \cap \Omega]}{\Pr[\Omega]} = \frac{\Pr[A]}{\Pr[\Omega]} = \frac{1}{2} \neq 1
$$
(e) FALSE. By linearity of expectation and by the formula for expectation of Bernoulli random variables:

$$E[X_1 - X_2] = E[X_1] - E[X_2] = Pr[X_1 = 1] - Pr[X_2 = 1] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \neq 0$$

(f) TRUE. Since $B \subseteq A \cup B$ we have $(A \cup B) \cap B = B$. Therefore

$$Pr[A \cup B \mid B] = \frac{Pr[(A \cup B) \cap B]}{Pr[B]} = \frac{Pr[B]}{Pr[B]} = 1$$

(g) TRUE. Here is one example

2. Alice has an urn with three marbles labeled 1, 2, and 3. Each of the marbles is equally likely to be extracted. Bob has a separate, similar urn. They play the following game of chance:

1. Alice extracts a marble from her urn and obtains $a \in \{1, 2, 3\}$.
2. Independently, Bob extracts a marble from his urn and obtains $b \in \{1, 2, 3\}$.
3. If $a > b$ then Alice wins. If $b > a$ then Bob wins. If $a = b$ they flip a fair coin and if the coin shows heads, Alice wins. If the coin shows tails, Bob wins.

Solve the problems below. Please do not spend time on the arithmetic. It is OK to leave your results as products and fractions in your calculations.

(a) Draw the “tree of possibilities” diagram for this game, with all the outcomes and their probabilities.

(b) Compute the probability that the game was decided by a coin flip.

(c) Compute the conditional probability that Alice wins given that Bob extracts marble labeled 2.

(d) Alice and Bob put bets on the game. If Alice wins without a coin flip Bob pays her 2$. If Alice wins with a coin flip then Bob pays her 1$. If Bob wins then Alice pays him 1.5$. What is Alice’s expected monetary win/loss (wins are positive, losses are negative) after $n$ such games?

**Answer**

For $i, j \in Z, i, j \in [1, 3]$ let $A_i$ be the event that Alice removes the marble labeled $i$ and $B_j$ be the event that Bob removes the marble labeled $j$. Let $C$ be the event that the coin is flipped. Finally, let $W_A$ be the event that Alice wins.

(a) The tree of possibilities is as follows (on the next page). Note that the extracted value (for marbles) /outcome (for coin flips) is on top of each edge and the probability is listed below each edge in the tree.
We want to find \( \Pr[C] \). Note that we flip a coin exactly when \( a = b \), i.e. we are interested in the event \( \bigcup_{i=1}^{3} A_i \cap B_i \). Since each of these are disjoint, we apply the Sum Rule:

\[
\Pr \left[ \bigcup_{i=1}^{3} A_i \cap B_i \right] = \Pr[A_1 \cap B_1] + \Pr[A_2 \cap B_2] + \Pr[A_3 \cap B_3]
\]

Since the marbles are drawn independently, \( A_i \perp B_j \) for all \( i, j \):

\[
= \Pr[A_1] \Pr[B_1] + \Pr[A_2] \Pr[B_2] + \Pr[A_3] \Pr[B_3]
\]
Since we know each marble is equally likely to be extracted:

\[
\frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} = \frac{1}{3}
\]

(c) We want to find \( \Pr[W_A \mid B_2] \). We can do so using conditional probability formulas and Law of Total Probabilities.

\[
\Pr[W_A \mid B_2] = \frac{\Pr[W_A \cap B_2]}{\Pr[B_2]} \quad \text{By Conditional Probability}
\]

\[
= \frac{\Pr[W_A \cap B_2 \cap C] + \Pr[W_A \cap B_2 \cap \overline{C}]}{\Pr[B_2]} \quad \text{By Law of Total Probabilities}
\]

\[
= \frac{\Pr[W_A \mid C \cap B_2] \Pr[C \mid B_2] \Pr[B_2] + \Pr[W_A \mid C \cap B_2] \Pr[C \mid B_2] \Pr[B_2]}{\Pr[B_2]}
\]

Now let's evaluate the probabilities in the expression above. \( \Pr[W_A \mid C \cap B_2] = \frac{1}{2} \) because event \( C \cap B_i \) means that the outcome of the game (whether or not \( W_A \) occurs) is decided by a coin flip. Since the coin is fair, \( W_A \) occurs with probability \( \frac{1}{2} \). Additionally, \( \Pr[C \mid B_i] = \frac{1}{3} \), since we flip the coin exactly when \( a = b \). We are given \( \Pr[B_2] = \frac{1}{3} \). Next, we see that \( \Pr[W_A \mid \overline{C} \cap B_2] = \frac{1}{2} \), since Alice is equally likely to have drawn a 1 or 3. Finally, we see that \( \Pr[C \mid B_2] \) because this is just the event that we do not flip a coin given that \( B_2 \) has occurred (So Alice chooses either marble 1 or 3), so \( \Pr[C \mid B_2] = \frac{2}{3} \). This simplifies to the following

\[
= \frac{\frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} \times \frac{1}{3}}{\frac{1}{3}} = \frac{1}{2}
\]

(d) Let \( G \) be the random variable denoting Alice’s monetary gain after a single game. We seek \( E[G] \). We express \( G \) as a piecewise function:

\[
G = \begin{cases} 
2 & W_A \cap \overline{C} \\
1 & W_A \cap C \\
-1.5 & W_A
\end{cases}
\]

By the definition of expectation and Law of Total Probabilities, we can express \( E[G] \) as the following:

\[
E[G] = G(W_A \cap \overline{C}) \cdot \Pr[W_A \cap \overline{C}] + G(W_A \cap C) \cdot \Pr[W_A \cap C] + G(W_A) \cdot \Pr[W_A]
\]

\[
= (2) \cdot \Pr[W_A \cap \overline{C}] + (1) \cdot \Pr[W_A \cap C] + (-1.5) \cdot \Pr[W_A]
\]
We proceed by finding the relevant probabilities. We have \( W_A \cap C \) when the coin is flipped and comes up with a head. Then:

\[
\Pr[W_A \cap C] = \Pr[W_A | C] \Pr[C]
\]

\[
= \frac{1}{2} \cdot \frac{1}{3} \quad \text{(from part (b))}
\]

\[
= \frac{1}{6}
\]

We also know the event \( W_A \cap \overline{C} \) occurs exactly when Alice draws a value strictly greater than Bob’s. That is:

\[
\Pr[W_A \cap \overline{C}] = \Pr[A_3 \cap B_2] + \Pr[A_3 \cap B_1] + \Pr[A_2 \cap B_1]
\]

\[
= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} \quad \text{(because } A_i \perp B_j \text{ from part (b))}
\]

\[
= \frac{1}{3}
\]

By the Law of Total Probabilities:

\[
\Pr[W_A] = 1 - \Pr[W_A \cap C] - \Pr[W_A \cap \overline{C}] = 1 - \frac{1}{6} - \frac{1}{3} = \frac{1}{2}
\]

We now plug these values in:

\[
E[G] = (2) \cdot \Pr[W_A \cap C] + (1) \cdot \Pr[W_A \cap \overline{C}] + (-1.5) \cdot \Pr[W_A] = 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{6} - 1.5 \cdot \frac{1}{2}
\]

\[
= \frac{1}{12}
\]

Now, if we play \( n \) games, we can apply linearity of expectation to see that Alice’s expected winnings are \( \frac{n}{12} \).

3. A fair coin is flipped \( 2n \) times \( (n \geq 1) \), independently. Let \( X_H \) the random variable that returns the number of heads that occurred and \( X_T \) the random variable that returns the number of tails that occurred. Compute \( P(X_H > X_T) \).

**Answer**

We first observe that we are working in a uniform probability space of size \( 2^{2n} \), since any series of tosses is equally likely. Additionally, there is a bijection between outcomes where \( X_H > X_T \) and \( X_T > X_H \), by inverting the results of every flip in the sequence. In other words,

\[
\Pr[X_H > X_T] = \Pr[X_T > X_H].
\]

But we know that our sample space can be partitioned into the events \([X_H > X_T]\), \([X_T > X_H]\), and \([X_H = X_T]\). Thus,

\[
1 = \Pr[X_H > X_T] + \Pr[X_T > X_H] + \Pr[X_H = X_T]
\]

\[
= 2 \Pr[X_H > X_T] + \Pr[X_H = X_T]
\]

\[
\Pr[X_H = X_T] = \frac{1}{2^{2n}}
\]

\[
\Pr[X_H > X_T] = \frac{1}{2^{2n}}
\]

\[
\Pr[X_T > X_H] = \frac{1}{2^{2n}}
\]

Therefore,

\[
P(X_H > X_T) = \frac{1}{2^{2n}}.
\]
We can count \( \binom{2n}{n} \) outcomes where \( X_H = X_T \) (choose \( n \) spots of the \( 2n \) for the heads to occur).

Since the probability space is uniform, \( \Pr[X_H = X_T] = \frac{|X_H = X_T|}{|\Omega|} = \frac{\binom{2n}{n}}{2^{2n}} \).

Plugging this into our equation from above gives us:

\[
\Pr[X_H > X_T] = \frac{1 - \frac{\binom{2n}{n}}{2^{2n}}}{2^{2n+1}}
\]

4. Let \((\Omega, P)\) be a probability space and let \( X \) be a random variable defined on \( \Omega \) such that \( \text{Val}(X) = \{a, b\} \) where \( a < b \). We also denote \( \mu = E[X] \).

(a) Express \( P(X \leq (a + b)/2) \) in terms of \( a, b \) and \( \mu \).

(b) Let \( a = -1 \) and \( b = 1 \). Show that if \( E(X) = 0 \) then there exists an event \( A \subseteq \Omega \) such that \( P(A) = 1/2 \).

**Answer**

(a) Since \( a < b \) we have \( a < \frac{a+b}{2} < b \). Then, \( \{\omega \mid X(\omega) \leq \frac{a+b}{2}\} = \{\omega \mid X(\omega) = a\} \). Therefore

\[
\Pr[X \leq \frac{a+b}{2}] = \Pr[X = a] \quad \text{so it suffices to express} \quad \Pr[X = a] \quad \text{in terms of} \quad \mu, a, b.
\]

\[
\mu = E[X] = a \Pr[X = a] + b \Pr[X = b]
\]

But we also have \( \Pr[X = a] + \Pr[X = b] = 1 \):

\[
\mu = a \Pr[X = a] + b(1 - \Pr[X = a])
\]

\[
\mu = (a - b)\Pr[X = a] + b
\]

Hence

\[
\Pr[X = a] = \frac{\mu - b}{a - b} = \frac{b - \mu}{b - a}
\]

We conclude that

\[
\Pr[X \leq \frac{a+b}{2}] = \frac{b - \mu}{b - a}
\]

(b) Plugging \( a = -1, b = 1, \mu = 0 \) in to the result of part (a), we get

\[
\Pr[X \leq \frac{a+b}{2}] = \frac{1 - 0}{1 - (-1)} = \frac{1}{2}
\]

Therefore we can take \( A = \{w \mid X(w) \leq \frac{-1+1}{2}\} = \{w \mid X(w) \leq 0\} \)

5. Weird Al (WAI) is playing with his coins. The game uses two *fair coins* and one *urn*. The result of the game is one of \( H \) (heads) or \( T \) (tails) and is determined as follows:
• WAii places both coins in the urn.
• WAii reaches inside the urn and (a) with probability $2/3$ WAii grabs one of the coins and tosses it, OR (b) with probability $1/3$ WAii grabs both coins, then tosses them separately in some order (doesn’t matter which order).
• If WAii has tossed just one coin then whatever that coin shows is the result of the game. If WAii has tossed both coins then applying the weird $\otimes$ operation to what the two coins show is the result of the game, where $T \otimes T = T$, $T \otimes H = H$, $H \otimes T = H$, and $H \otimes H = T$.

(a) Draw the “tree of possibilities” diagram for WAii’s game.
(b) Calculate the probability that the result of of the game is $H$.
(c) What simpler game could Weird Al play that would give him exactly the same odds?

Answer
We define the following events: $C_1$ indicating the event WAii flips one coin, $H_1$ indicating the event that flip is a head, $C_2$ indicating the event he flips 2 coins, $HH, HT, TH, TT$ indicating the sequence of flips he gets.

(a) The tree of possibilities:

(b) The outcome of the game is $H$ if 1) WAii flips one coin (event $C_1$) and it lands on $H$ (event $H_1$) or 2) WAii flips two coins (event $C_2$) and gets one $H$, one $T$ (event $HT$). If WAii flips two coins, he will get one $H$, one $T$ with probability $1/4$, since both $HT$ and $TH$ will yield a head, and both
occur with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

$$Pr[H] = Pr[C_1 \cap H_1] + Pr[C_2 \cap HT]$$
$$= Pr[C_1]Pr[H_1|C_1] + Pr[C_2]Pr[HT|C_2]$$
$$= \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}$$
$$= \frac{1}{2}$$

(c) Since the result of WAl’s game is H with probability $\frac{1}{2}$ and T otherwise, this is equivalent to simply flipping a fair coin once.

6. Consider $X$ and $Y$, two independent Bernoulli random variables defined on the same probability space. We are given $Pr[X = 1] = 1/3$ and $Pr[Y = 1] = 1/4$. Compute $E[(X + Y)^2]$.

**Answer**

We calculate $E[(X + Y)^2]$ as follows:

$$E[(X + Y)^2] = E[X^2 + 2XY + Y^2]$$
$$= E[X] + 2E[X]E[Y] + E[Y]$$

We use the fact that $X$ and $Y$ are Bernoulli to conclude that $E[X^2] = E[X]$ and $E[Y^2] = E[Y]$. Since $X$ only takes on values of 0 or 1, $X^2$ will take on values of 0 or 1 precisely when $X$ does. The same reasoning applies to $Y^2$.

Further, note that since $X$ and $Y$ are Bernoulli random variables, $E[X] = Pr[X = 1]$ and $E[Y] = Pr[Y = 1]$.

$$E[(X + Y)^2] = Pr[X = 1] + 2Pr[X = 1]Pr[Y = 1] + Pr[Y = 1]$$
$$= \frac{1}{3} + 2 \left( \frac{1}{3} \cdot \frac{1}{4} \right) + \frac{1}{4}$$
$$= \frac{3}{4}$$

7. For each statement below, decide whether it is TRUE or FALSE and circle the right one. In each case attach a very brief explanation of your answer.

(a) There exists a tree in which every node is a leaf.
(b) The complete bipartite graph $K_{5,5}$ has a cycle of length 5.
(c) There exists an undirected graph with 23 vertices such that 11 of them have degree 11 and 12 of them have degree 12.
(d) There exists a connected undirected graph with 100 vertices and 50 edges.
(e) A graph has an edge that is not a cut edge. Then it must have three edges that are not cut edges.
(f) There exists a tree with exactly 3 leaves, in which the length of a longest path is 1000.

(g) Consider an undirected graph with 3 or more vertices and with exactly 3 connected components. In order to make this into a connected graph, we must add at least 2 edges.

**Answer**

(a) **TRUE.** An example of this is a graph $G$ with only two vertices connected by an edge. $G$ is connected and acyclic, so it is a tree. In addition, the two vertices each have degree 1, so each node is a leaf.

(b) **FALSE.** As proved in lecture, a graph is bipartite iff it has only even length cycles (regardless of whether or not the graph is complete). Hence, if $K_{5,5}$ has an odd cycle, it cannot be bipartite.

(c) **FALSE.** We know by the Handshaking Lemma that there must be an even number of vertices with odd degree. Since 11 vertices have degree 11, this would violate the aforementioned property, so it would not be a valid graph.

(d) **FALSE.** We know that a tree is a minimally connected graph with $n$ vertices and $n - 1$ edges. This would mean that any connected graph with $n$ vertices will have at least $n - 1$ edges. $100 - 1 = 99 > 50$. Thus, a graph with 100 vertices and 50 edges would not be connected. Alternatively, we also know a graph has at least $|V| - |E|$ connected components. This graph has at least $100 - 50 = 50 > 1$ CCs which means it is not connected.

(e) **TRUE.** If a graph does not have a cut edge, that would mean it is not minimally connected, so it must have a cycle. The smallest cycle we can create in a graph is a cycle of length 3. Each of the 3 edges in this cycle are not cut edges since removing any of them will still leave the graph connected.

(f) **TRUE.** Consider the distinct vertices

$$a, b, v_1, v_2, \ldots, v_{999}, v_{1000}$$

and the edges

$$a - v_1, b - v_1, v_1 - v_2, \ldots, v_{999} - v_{1000}$$

This a tree with leaves $a, b, v_{1000}$, and there are two longest paths, both of length 1000.

(g) **TRUE.** In order to make the graph connected, we need to find a way to create paths between the vertices in each of the three components. Call the connected components $A, B,$ and $C$. We can add an edge between any vertex in $A$ and any vertex in $B$. This create paths between all of the vertices in $A$ and $B$ via the edge we just added. However, the graph is not connected yet because the vertices in $A$ and $B$ cannot reach the vertices in $C$.

We can add another edge between any vertex in $C$ and any vertex in $B$ (or $A$) to create paths between the two connected components. Now, we have connected all three connected components with only two edges. Note that you can add more edges to the graph, but this would not affect the connectedness of the graph once we add the two edges as described above.

8. Let $G$ be a bipartite graph in which every connected component is a cycle.
(a) Draw the smallest (minimal number of vertices) such $G$. (Just the drawing, no need for explanation)

**Answer**

$G = (\{r_1, r_2, b_1, b_2\}, \{r_1 - b_1, b_1 - r_2, r_2 - b_2, b_2 - r_1\})$

(b) Prove that, not just in the smallest, but in any such $G$ the number of red nodes is equal to the number of blue nodes.

**Answer**

Recall from lecture that a graph $G$ is bipartite if and only if it does not contain any odd length cycles. Then, every connected component in the graph $G$ in this problem must be an even length cycle. If we show that an arbitrary connected component in $G$ has an equal number of blue nodes and red nodes, we can take the sum over all components and get our desired result.

Consider an arbitrary connected component in $G$, call it $C$. We know that $C$ is an even length cycle; suppose that $C$ has $2n$ edges, for some $n \geq 2 \in \mathbb{Z}$. Since $C$ is a cycle with $2n$ edges, we know that $C$ must have $2n$ vertices. Call these vertices $v_1, v_2, ..., v_{2n}$, where $v_i$ is adjacent to $v_{i+1}$ for $i \leq 2n - 1$, and $v_{2n}$ is adjacent to $v_1$. Without loss of generality, suppose that $v_1$ is red. Then, since an edge exists between $v_1$ and $v_2$, $v_2$ must be blue. Continuing, we require that all odd indexed vertices be red and all even indexed vertices be blue for $G$ to be bipartite. Note that there are $n$ odd indexed vertices and $n$ even vertices, so the number of blue vertices in $G$ is $n$ and the number of red vertices in $G$ is also $n$. Then, there is an equal number of blue and red vertices in any given connected component of $G$. Taking the sum over all connected components in $G$ proves our claim.

9. Consider a connected graph $G = (V, E)$ such that $|E| = |V|$. Prove that $G$ contains exactly one cycle.

**Answer**

As the graph $G$ is connected, we know it has at least one spanning tree $T$ that we can arbitrarily select. From the definitions of a tree proved in lecture, we know $T$ to have $n - 1$ edges from the original graph, to be connected, and to be acyclic. From this, we know there is only one edge $e$ that isn’t included in the spanning tree, but is in the original graph. Thus, if there was a cycle to exist in the original graph, it would have to include the edge $e$, as there are no cycles in the spanning tree by definition.

Looking at the two vertices adjacent on $e$, $v_a$ and $v_b$, we know from the connectivity of the spanning tree that there is a path connecting the two vertices. Combining this path with the edge $e$, we can see that a cycle forms as there is a path from a vertex to itself. Furthermore, we know that there can be no more cycles that form, as otherwise, there would have had to have been a different path connecting $v_a$ and $v_b$ (as the new cycle would have to be distinct), which would contradict our definition of a tree as being acyclic (as two unique paths connecting two vertices implies a cycle to already exist without $e$).

As such we have proven the desired claim.

10. Define the complement of a string of bits $w$ of length $n \geq 1$ to be the string obtained by replacing all the 0’s in $w$ with 1 and all the 1’s with 0’s. Clearly, the complement of the complement of $w$ is $w$. Now construct an undirected graph whose nodes are all the strings of bits of length $n$ and such that
there is an edge \( u-v \) exactly when \( v \) is the complement of \( u \) (equivalently, when \( u \) is the complement of \( v \)). Prove that the resulting graph is bipartite.

**Answer**

Let \( G = (V,E) \) be the graph that was constructed given the constraints above. Let \( X \) and \( Y \) be the sets of strings whose first bit is a 0 and whose first bit is a 1, respectively. Since every string starts with a 0 or 1 but not both, we know \( X \) and \( Y \) partition \( V \).

Consider an edge \( e \) in \( G \). Since the complement operation replaces 0's with 1's and 1's with 0, one endpoint of \( e \) starts with a 0 and the other with a 1. Hence, every edge in \( e \) has one endpoint in \( X \) and one endpoint in \( Y \). It follows that \( G \) is bipartite.

11. Let \( X, Y, Z \) be three finite nonempty sets such that \( X \cap Y = \emptyset, Z \cap Y = \emptyset, X \cap Z = \emptyset \) and denote \(|X| = m, |Y| = n, |Z| = p \). Assume that \( m < n < p \). Let also \( f : X \to Y \) and \( g : Y \to Z \) be two functions. Consider the undirected graph \( G = (V,E) \) where \( V = X \cup Y \cup Z \) and

\[
E = \{ \{x, f(x)\} \mid x \in X\} \cup \{ \{y, g(y)\} \mid y \in Y\}
\]

(a) What is \(|V|\) and what is \(|E|\) (in terms of \( m, n, p \))?.

(b) What is the maximum number of nodes of degree 0 that \( G \) can have (in terms of \( m, n, p \))?.

(c) What is the minimum number of nodes of degree 0 that \( G \) can have (in terms of \( m, n, p \))?.

(d) What is the maximum length that a path in \( G \) can have?

(e) Prove that \( G \) is acyclic.

**Answer**

(a) Since \( X, Y, Z \) are disjoint, We know

\[|V| = |X \cup Y \cup Z| = |X| + |Y| + |Z| = m + n + p.\]

Since \( f, g \) are functions there is exactly one edge of the form \( \{x, f(x)\} \) for each \( x \in X \) and exactly one set of the form \( \{y, g(y)\} \) for each \( y \in Y \). Therefore \(|E| = m + n\).

(b) Since every node in \( X \) and \( Y \) is the endpoint of an edge only nodes from \( Z \) can have degree 0. They correspond to elements outside of the range of \( g \). To maximize their number we must minimize the size of \( g(Y) \). This happens when \( g \) maps all the elements of \( Y \) to one element of \( Z \). Hence, the answer is \( p - 1 \).

(c) Reason like in part (b) then maximize the size of \( g(Y) \). This happens when \( g \) is injective and the size of \( g(Y) \) is \( n \). The answer is \( p - n \). (Note that we need \( n \leq p \) or else we cannot define \( g \) to be an injection.)

(d) First we observe that \( X \) is nonempty so either \( m = 1 \) or \( m \geq 2 \). We now consider the maximum length path in each case.

**Case 1** \( m = 1 \):

Let \( X = \{a\} \). Since \( n > m \) we have \( n \geq 2 \) so let \( b_1, b_2 \in Y \) with \( b_1 \neq b_2 \). Define \( f(a) = b_1 \) and \( g(b_1) = g(b_2) \). Then we have a path of length 3 in \( G \):

\[a - f(a) = b_1 - g(b_1) = g(b_2) - b_2.\]
We now prove that there is no path of length 4. Since \( a \notin Y \cup Z \), there is exactly one edge incident to the \( a \). In any path of length 4, therefore, there must be 3 edges between \( Y \)-nodes and \( Z \)-nodes. Then two of these edges would be incident to the same \( Y \)-node. Since \( g \) is a function, the edges would be incident to the same \( Z \)-node. We cannot repeat edges, however, so no such path exists. Hence, when \( m = 1 \), the maximum path length is 3.

**CASE 2** \( m \geq 2 \):

Let \( a_1, a_2 \in X \) with \( a_1 \neq a_2 \). Since also \( n \geq 2 \) let \( b_1, b_2 \in Y \) with \( b_1 \neq b_2 \). Define \( f(a_1) = b_1 \) and \( f(a_2) = b_2 \) and \( g(b_1) = g(b_2) \). Then we have a path of length 4 in \( G \):

\[
a_1 - f(a_1) = b_1 - g(b_1) = g(b_2) - b_2 = f(a_2) - a_2.
\]

We now prove that no path of length 5 can exist. Every edge either connects an \( X \)-node and \( Y \) node or a \( Y \)-node and \( Z \)-node. By PHP, any path \( P \) with 5 edges must have at least three 3 edges between \( X \)-nodes and \( Y \)-nodes or between \( Y \)-nodes and \( Z \)-nodes. In either case, at least 2 of the 3 edges would have to share an \( X \)-node (in the first case) or a \( Y \)-node (in the second case). Since \( f \) and \( g \) are functions, these two edges would also be incident to the same \( Y \) node (in the first case) or \( Z \) node (in the second case). This means the two edges are equal, which is a contradiction. Hence, when \( m \geq 2 \) the maximum path length is 4.

(e) Assume, for the sake of contradiction, that a cycle \( C \) exists. Since \( f \) maps each \( X \)-node to one \( Y \)-node, the degree of each \( X \)-node is 1. Any vertex in a cycle must have at least two incident edges. Thus, cycle \( C \) cannot contain an \( X \)-node.

Now, assume there is a \( Y \)-node \( y \) in \( C \). From above, we know that \( C \) contains no \( X \)-nodes, so we know \( y \) is incident to some \( Z \)-node \( z \) in \( C \). Vertex \( y \) is only incident to one \( Z \)-node because \( g \) is a function. Since \( C \) does contain any \( X \)-nodes, there is no edge that can return to \( y \).

Thus, cycle \( C \) must only contain \( Z \)-nodes. This, however, is impossible because there is no edge between two \( Z \)-nodes. Hence, no cycle exists, and \( G \) must be acyclic.

12. An \( r \)-regular graph is a graph in which the degree of each vertex is exactly \( r \). Show that any 3-regular graph must have an even number of vertices and a number of edges divisible by 3.

**Answer**

Let \( G = (V, E) \) be a 3-regular graph. As we have shown in lecture using the handshake lemma, every graph has an even number of vertices of odd degree. Since, every vertex in \( G \) has degree 3 by the definition of 3-regular, every vertex in \( G \) has odd degree. With this we can say that the graph must have an even number of vertices because there are only odd degree vertices.

Also, by the handshake lemma we know that \( 2|E| = \sum_{v \in V} \text{deg}(v) \). Every vertex in \( G \) has degree 3, substituting into the above formula, we get that \( 2|E| = 3|V| \). We know that \( |E| \) is the number of edges in \( G \), and since 2 is not divisible by 3, we can say that the number of edges must be divisible by 3.

Now, we have show that the for any 3-regular graph \( G \) there are an even number of vertices, and a number of edges that is divisible by 3, and we are done.

13. For each statement below, decide whether it is TRUE or FALSE and circle the right one. In each case attach a *very brief* explanation of your answer.
(a) Let $G$ be a DAG with $n \geq 2$ vertices and let the sequence $\sigma$ be a topological sort of $G$. If $u$ appears before $v$ in $\sigma$ then there exists a directed path from $u$ to $v$ in $G$.

**Answer**

*FALSE.* Counter example:

A topological sort would be 1423. Let $u = 1$ and $v = 4$. However, there is no path from 1 to 4.

(b) A strongly connected digraph with at least two nodes can have neither sources nor sinks.

**Answer**

*TRUE.* A strongly connected digraph with at least two nodes can have neither sources nor sinks.

Assume for contradiction that there is source $v$. Then there must be no edges leading to $v$, and therefore no paths to $v$. The graph cannot be strongly connected, so we have a contradiction. By similar logic, there cannot be any sinks.

14. For the three parts below, use the following definition: for any digraph $G = (V, E)$ without self-loops and without cycles of length 2, define an undirected graph $G^u = (V, E^u)$ that has the same vertices as $G$ and moreover in $G^u$ we have an edge $v \rightarrow w$ whenever we have the edge $v \rightarrow w$ or the edge $w \rightarrow v$ in $G$.

(a) If $G$ is strongly connected then $G^u$ is connected. Prove or disprove.

**Answer**

*TRUE.* Let $v, w$ be two vertices in $G^u$ (hence in $G$). Since $G$ is strongly connected, there exists a directed walk $v \rightarrow \cdots \rightarrow w$ in $G$. Erasing the direction of the edges in this walk gives a walk $v \cdots \rightarrow w$ in $G^u$.

(b) If $G$ is a DAG then $G^u$ is acyclic. Prove or disprove.

**Answer**

*FALSE.* Counterexample: $G = (\{a, b, c\}, \{a \rightarrow b, a \rightarrow c, c \rightarrow b\})$ is a DAG but $G^u$ is a cycle of length 3.

(c) Prove that if $G$ is a DAG in which every sink is reachable from every source then $G^u$ is connected.

**Answer**

Assume $G$ is such a DAG and let’s prove that $G^u$ is connected.

Let $v, w$ be two distinct vertices. We want to show that there is a walk $v \cdots \rightarrow w$ in $G^u$.

CLAIM In $G$ there exists a source $s$ such that $v$ is reachable from $s$.

PROOF OF CLAIM Consider the set $S$ of all paths in $G$ from some vertex to $v$. There is at least on such path, the path of length 0. By the Well-Ordering Principle at least one of these paths has maximum length among the paths in $S$, we call it a *maximal* path.
If this path has length 0 then \( v \) itself must be a source, otherwise we can extend the length of the path by including one of \( v \)'s predecessors. So \( v \) is reachable from a source (itself).

Suppose the maximal path is not of length 0, let it be \( s \rightarrow \cdots \rightarrow v \). Then \( s \) must be a source. Indeed, if \( s \) has a predecessor \( p \) then either \( p \) is among the vertices of the maximal path (and then \( G \) has a cycle, contradiction) or \( p \) is not among the path's vertices and then the path is not maximal (also a contradiction) because we can extend it with the edge \( p \rightarrow s \). So \( v \) is reachable from the source \( s \). This ends the proof of the claim.

Similarly, we prove the claim

CLAIM In \( G \) there exists a sink \( t \) such that \( t \) is reachable from \( w \).

So we have a source \( s \) and a sink \( t \) such that \( s \rightarrow v \) and \( w \rightarrow t \). In addition we know that in \( G \) every sink is reachable from every source therefore \( s \rightarrow t \). Now we erase direction on the walks that give \( s \rightarrow v \), \( w \rightarrow t \) and \( s \rightarrow t \). This gives a walk \( v \rightarrow \cdots \rightarrow s \rightarrow \cdots \rightarrow t \rightarrow \cdots \rightarrow w \) in \( G \). Done.

15. Recall the complete undirected graph on \( n \) vertices, \( K_n \). Prove that for any \( n \geq 4 \) it is possible to assign direction to each of the edges of \( K_n \) such that the resulting digraph has exactly \( n - 2 \) strongly connected components.

**Answer**

**Solution One** We proceed by induction on \( n \).

**Base case:** \( n = 4 \). If our graph is

\[
\{(a, b, c, 1), (a, b), (b, c), (c, a), (a, 1), (b, 1), (c, 1)\},
\]

we can see that the \( n - 2 = 2 \) strongly connected components are \( \{a, b, c\} \) and \( \{1\} \) since the subgraph induced on \( \{a, b, c\} \) is a cycle and since \( a \), \( b \), and \( c \) are not reachable from \( 1 \).

**Induction step:** Let \( k \) be an arbitrary integer strictly greater than 3, and assume that there is a digraph \( G = (V, E) \) with exactly \( k - 2 \) strongly connected components that is the result of assigning direction to each of the edges of \( K_k \) (our induction hypothesis). Now consider the graph \( G' \) formed by adding a vertex \( v \) to \( G \) and directing edges to \( v \) from every other vertex. Since no other vertex is reachable from \( v \), it forms its own strongly connected component, so \( G' \) has \( (k + 1) - 2 \) strongly connected components, and the induction step holds.

**Solution Two** Let’s name the vertices \( 1, 2, \ldots, n - 3, a, b, c \) (check that there are indeed \( n \) vertices).

We have three types of edges in \( K_n \): edges between two numbered vertices, edges between two vertices labeled with letters, and edges between one vertex labeled with a number and one with a letter. For the edges with numbers, we orient the edge so that it points from the lower number to the higher number. For the edges with letters, we use the edges \( (a, b), (b, c), \) and \( (c, a) \). For the edges between letters and numbers, we orient the edge so that it points from the letter to the number.

We claim that each vertex \( 1, 2, \ldots, n - 3 \) forms a distinct strongly connected component and that \( \{a, b, c\} \) forms another. To prove this, we show that no two vertices from different strongly connected components are mutually reachable from each other.

Since every edge that begins in a numbered vertex ends in a numbered vertex, we cannot have a path from a numbered vertex to a non-numbered vertex, so the lettered vertices are not reachable from the numbered vertices and thus cannot be in the same strongly connected component. In addition,
since the only edge from any numbered vertex is to a higher-numbered vertex, we cannot have any paths from a higher-numbered vertex to a lower-numbered vertex, and thus the lower-numbered vertices are not reachable from the higher-numbered vertices; therefore no two numbered vertices are in the same strongly connected component, and since they are all separate from the letters, they each constitute their own strongly connected component. Finally, since \( a, b, \) and \( c \) are part of the cycle \( a \to b \to c \to a \), they are in the same strongly connected component.

We have shown that the strongly connected components are \( \{1\}, \{2\}, \ldots, \{n - 3\}, \{a,b,c\} \). Thus there are exactly \( n - 2 \), and we are done.

16. Let’s call a *slug* a DAG \( G = (V, E) \) with at least 4 vertices, \(|V| \geq 4\), and such that \( G \) has exactly one source \( r \) and exactly one sink \( s \).

(a) Draw two different slugs, both with 4 vertices, one of them with 3 edges and the other one with 5 edges.

**Answer**

This is a slug with 3 edges:

```
1  2  3  4
```

This is a slug with 5 edges:

```
1  2  3  4
```

(b) Prove that in any slug, for every node \( u \) that is not \( r \) or \( s \), there exists a directed path from \( r \) to \( s \) that passes through \( u \).

**Answer**

Consider a maximum-length directed path through \( u \), which we call \( v_0, v_1, \ldots, u, \ldots, v_k \). We claim that the endpoints of this path are \( r \) and \( s \), i.e. \( v_0 = r \) and \( v_k = s \).

We know that \( \text{in}(v_0) = 0 \) or \( \text{in}(v_0) > 0 \). If it is the former, then \( v_0 = r \), since there is only one source in the graph. Otherwise, there must be some vertex \( v \) such that \( v \to v_0 \) is present in the graph. Then \( v \) must already be on the path; otherwise, we could extend it and the path does not have maximum length. Then there is a cycle \( v_0 \to v \to v_0 \) (the first part of the cycle is found...
Let $G = (V, E)$ be a connected graph with at least two distinct spanning trees.

(a) Prove that $|E| \geq |V|$.

**Answer**
Since $G$ is connected, we know that $|E| \geq |V| - 1$. Now, since $|E|, |V| \in \mathbb{Z}$, it suffices to show that $|E| \neq |V| - 1$.

Assume for contradiction that $|E| = |V| - 1$. Then, $G$ is a tree. Therefore, it has exactly one spanning tree, itself. The problem statement tells us that $G$ has at least two distinct spanning trees, so this is a contradiction.

(b) Prove that the graph has at least three distinct spanning trees.

**Answer**
Consider $T$, one of the spanning trees of $G$. $T$ has exactly $n - 1$ edges, and since $G$ has at least $n$ edges (shown in part (a)), we know that there exists an edge $e = (u, v)$ in $G$ that is not in $T$. Thus, consider the graph $G'$ created from adding the edge $e$ into $T$, and so $G'$ has $n$ edges. Because $T$ was a tree, there was a unique path from $u$ to $v$ before adding in $e$, and so the addition of $e$ creates exactly one cycle, $C$. Because $G'$ is undirected, we know $C$ has at least 3 edges. Now, consider the graphs $T_1, T_2,$ and $T_3$ created by removing any one of 3 distinct edges within $C$, and so each of these graphs is distinct. Because we removed an edge belonging to a cycle, these graphs remain connected. Furthermore, since $G'$ had $n$ edges, each of $T_1, T_2,$ and $T_3$ has $n - 1$ edges. Thus, since each of these 3 graphs is connected and contains $n - 1$ edges, they form 3 distinct spanning trees.

**Alternate Solution 1**
Let $T_1$ and $T_2$ be the two distinct spanning trees. Since they are distinct there must exist some $T_1$-edge that is not a $T_2$-edge. Call it $u-v$.

Since $T_2$ is a spanning tree there must exist a path $P = u \cdot \cdot \cdot v$ in $T_2$. At least one edge in $P$ cannot be a $T_1$-edge. Indeed, if all of them were, then together with $u-v$ they would form a cycle in $T_1$ which is impossible. Call this $T_2$-edge in the path $P$ that is not a $T_1$-edge $x-y$.

To the path $P$ we add the edge $v-u$ to form a cycle (in $G$), call this cycle $C$. It contains the two distinct edges $u-v$ and $x-y$. But cycles have at least 3 edges so there must be a third edge in $C$ distinct from both $u-v$ and $x-y$. Call it $s-t$. (By the way saying that these three edges are distinct does not imply that their endpoints are distinct. Some of these edges may be adjacent.)

To recap, in the cycle $C$ we have $v-u$ which is a $T_1$-edge but not a $T_2$-edge, $x-y$ which is a $T_2$-edge but not a $T_1$-edge, and $s-t$ which is a $T_2$-edge (and may or may not be a $T_1$ edge).

Now let $T_3 = T_2 \cup \{u-v\} \setminus \{s-t\}$ (these are the edges; the vertices are $V$). $T_3$ differs from $T_1$ because it still has $x-y$. $T_3$ also differs from $T_2$ because it does not have $s-t$. So if we show $T_3$ is a spanning tree we have our third distinct spanning tree and we are done.

$T_3$ has $|V| - 1$ edges because although we deleted an edge from $T_2$ we added another one. $T_3$ is connected because $T_2$ is connected and any path in $T_2$ that used the edge we deleted, $s-t$ can be
“repaired” to a path in $T_3$ using the rest of the cycle $C$. Therefore $T_3$ is a tree, and a spanning because it has all the vertices.

**Alternate Solution 2**

Let $T_1$ and $T_2$ be the two distinct spanning trees. Since they are distinct and they have the same number of edges ($|V| - 1$) there must exist some $T_1$-edge that is not a $T_2$-edge, call it $u-v$, and a $T_2$-edge that is not a $T_1$-edge, call it $x-y$.

Since $T_2$ is a spanning tree, it must have a path $P = u-\cdots-x$ and a path $Q = v-\cdots-y$. We claim that $P$ and $Q$ are node-disjoint. Indeed, if $P$ and $Q$ have a node in common then that would give a path in $T_2$ between $x$ and $y$ that is not the same as the edge $x-y$. But paths in a tree are unique, contradiction.

By concatenating $u-v$, $Q$, $y-x$, and the reverse of $P$ we get a cycle, $C$.

From here on we proceed as in the alternative solution 1.