CIS 160

Recitation Guide - Week 10

Topics Covered: Graphs, Variance

Problem 1:

We've proven that in any graph, the number of vertices of odd degree is even with the Handshake Lemma. Prove this again, using induction.

Solution:

We will prove this statement by induction on the number of edges, m. Let $P(m)$ be defined as:

> In any graph with m edges, the number of vertices of odd degree is even.

<u>Base Case:</u> $P(0)$ holds, because a graph with no edge has only isolated vertices, that is, vertices of degree 0. Hence, there are an even number (0) of vertices with odd degree.

<u>Induction Step:</u> Assume $P(k)$ is true, for an arbitrary $k \in \mathbb{N}$. Now, we want to prove $P(k+1)$ is true.

Let G be a graph with $k+1$ edges. Remove an arbitrary edge $e = \{u, v\}$ from G (note that it could be any edge), so that we now have a graph G' with k edges. By the Induction Hypothesis, the number of vertices with odd degree in G' is even. Denote the number of vertices with odd degree in G' to be $2a$, where $a \in \mathbb{N}$. Now put back the edge e that we removed earlier. Observe that doing so increases the degree of vertices u and v by one each. We consider the following three cases:

Case 1: Both u and v have odd degree in G'. Adding e back would make the degree of both u and v even. Hence, the number of vertices with odd degree becomes $2a - 2$.

Case 2: Both u and v have even degree in G'. Adding e back would make the degree of both u and v odd. Hence, the number of vertices with odd degree becomes $2a + 2$.

Case 3: Exactly one of u and v has odd degree in G'. WLOG, assume u has an odd degree and v has an even degree in G'. Adding e back would result in u with an even degree and v with an odd degree. Hence, the number of vertices with odd degree would stay unchanged ($2a$).

In all cases, the number of odd degree vertices in G is even. Thus, we have shown our claim is true when $m = k + 1$, concluding our Induction Step and completing our proof.

ALTERNATE:

We will prove this statement by induction on the number of vertices, n. Let $P(n)$ be defined as:

> In any graph with n vertices, the number of vertices of odd degree is even.

Base Case: $P(1)$ holds, because a graph with 1 vertex is an edgeless graph, that is, the graph contains one vertex of degree 0. Hence, there are an even number (0) of vertices with odd degree.

Induction Step: Assume $P(k)$ is true, for an arbitrary $k \in \mathbb{Z}^+$. Now, we want to prove $P(k+1)$ is true.

Let G be a graph with $k+1$ vertices. We partition V into the following sets:

$X =$ the neighbors of v with even degree in G.
$Y =$ the neighbors of v with odd degree in G.
$R =$ vertices not adjacent to v with even degree in G.
$S =$ vertices not adjacent to v with odd degree in G.

Since Y and S partition the set of odd-degree vertices in G, want to show $|Y| + |S|$ is even.

Remove an arbitrary vertex v with degree α from G, so that we now have a graph G' with k vertices.

Case 1: α is even. Then $\alpha = 2k, k \in \mathbb{N}$. This means that $|X| + |Y| = 2k$, so $|X|$ and $|Y|$ have the same parity.

The degree of each neighbor of v decreased by 1, so all $x \in X$ in G became odd degree vertices in G'. By the Induction Hypothesis, the number of vertices with odd degree in G' is even. Denote the number of vertices with odd degree in G' to be $2a$, where $a \in \mathbb{N}$. In G' this gives $|X| + |S| = 2a$, since $|X|$ now represents the new odd vertices in G', and $|S|$ represents the odd-degree vertices unaffected by the removal of v. Note that $|X|$ and $|S|$ must have the same parity (think why this is true!).

Now put back the vertex v that we removed earlier. Observe that doing so increases the degree of each of the neighbors of v by 1. All vertices in X now have even degree, and all vertices in Y have odd degree again.

Since $|X|$ and $|S|$ must have the same parity, we have $|X|, |Y|, |S|$ all have the same parity. This means $|Y| + |S|$ is even (since odd+odd is even, and even + even is odd).

Case 2: α is odd. (We include v in S). Then $\alpha = 2k+1, k \in \mathbb{N}$. This means that $|X|+|Y| = 2k+1$, so $|X|$ and $|Y|$ have the opposite parity.

The degree of each neighbor of v decreased by 1, so all $x \in X$ in G became odd degree vertices in G'. By the Induction Hypothesis, the number of vertices with odd degree in G' is even. Denote the number of vertices with odd degree in G' to be $2a$, where $a \in \mathbb{N}$. In G' this gives $|X|+|S| = 2a$, since $|X|$ now represents the new odd vertices in G', and $S' = S\backslash\{v\}$ since v had odd degree, removing it means $|S'| = |S| - 1$ since all other vertices in S are unaffected. Note that $|X| + |S'| = 2a$, or $|X| + |S| = 2a + 1$. This means $|X|$ and $|S|$ have opposite parity.

Now put back the vertex v that we removed earlier. Observe that doing so increases the degree of each of the neighbors of v by 1. All vertices in X now have even degree, and all vertices in Y have odd degree again.

Since $|X|$ and $|S|$ must have the opposite parity, and $|X|$ and $|Y|$ have opposite parity, we have note $|Y|$ and $|S|$ have the same parity. (Can you prove this?). This means $|Y| + |S|$ is even (since odd+odd is even, and even + even is odd).

In all cases, the number of odd degree vertices in G is even. Thus, we have shown our claim is true when $m = k + 1$, concluding our Induction Step and completing our proof.

Problem 2:

Prove that a graph $G = (V, E)$ is connected iff for every partition of V into two disjoint, non-empty sets S and T, there exists an edge between some vertex in S and some vertex in T.

Solution:

($\implies$): We first show that if a graph $G = (V, E)$ is connected, then for every partition of V into two disjoint, non-empty sets S and T, there exists an edge between some vertex in S and some vertex in T. Consider an arbitrary partition $V = S \cup T$ into two disjoint, non-empty sets S and T. Let $x \in S$ and $y \in T$; since G is connected, there must be a path $x \rightsquigarrow y$, say:

$$P = x - v_1 - v_2 - \ldots - v_{k-1} - y$$

We claim that there must be some edge from S to T in this path. Suppose towards contradiction that all edges are between two vertices in S or two vertices in T. Since $x \in S$, we must have $v_1 \in S$. Similarly, we must then have $v_2 \in S$. We may continue this process to show that $v_{k-1} \in S$ (see if you can formally prove this with induction!), and $y \in S$, a contradiction.

($\impliedby$): We now show that, given a graph $G = (V, E)$, if for every partition of V into two disjoint, non-empty sets S and T, there exists an edge between some vertex in S and some vertex in T, then G is connected. We proceed by proving the contrapositive, namely, that if G is not connected, then there exists a partition of V into disjoint nonempty sets S and T with no edges between the two.

Since G is not connected, it must have at least two connected components. Let S be a connected component of G and let $T = V \setminus S$. By definition of connected component, there is no edge from a vertex of S to one in T (if there were, we would violate the maximality condition). This gives us our desired partition.

Problem 3:

Let X, Y be two random variables defined on the same probability space. The covariance of X and Y is defined to be

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$$

Note that by Proposition 17.5 if $X \perp Y$ then $\mathrm{Cov}(X,Y) = 0$. Two random variables are uncorrelated if $\mathrm{Cov}(X,Y) = 0$ (equivalently, such that $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$). Therefore independence implies uncorrelation.

(a) Give an example of two random variables that are uncorrelated yet they are not independent

(b) Prove that $\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$. Conclude that $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

(c) Then show that $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)$.

Solution:

(a) Let our probability space consist of the outcomes $\{(0,1),(0,-1),(1,0),(-1,0)\}$ with a uniform probability distribution. Define the random variable X to denote the first value in the ordered pair and the random variable Y to denote the second value.

$$\begin{aligned}
\mathrm{E}[X] = \mathrm{E}[Y] &= \sum_{x \in X} x \Pr[X = x] \\
&= (0)(1/2) + (-1)(1/4) + (-1)(1/4) \\
&= 0 \\
E[XY] &= \sum_{z \in XY} z \Pr[XY = z] \\
&= (0)(1) = 0 \\
\mathrm{Cov}(X,Y) &= 0 - 0 = 0
\end{aligned}$$

Thus, X and Y are uncorrelated. However, they are not independent:

$$\begin{aligned}
\Pr[X = 1 \cap Y = 1] &= 0 \\
\Pr[X = 1] &= \frac{1}{4} \\
\Pr[Y = 1] &= \frac{1}{4} \\
\Pr[X = 1] \times \Pr[Y = 1] &= \frac{1}{16} \neq 0
\end{aligned}$$

(b) Applying the Linearity of Expectation (LOE), we see:

$$\begin{aligned}
\mathrm{E}\Big[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\Big] &= \mathrm{E}\Big[XY - X \cdot \mathrm{E}[Y] - \mathrm{E}[X] \cdot Y + \mathrm{E}[X]\,\mathrm{E}[Y]\Big] \\
&= \mathrm{E}\big[XY\big] - \mathrm{E}\big[X \cdot \mathrm{E}[Y]\big] - \mathrm{E}\big[\mathrm{E}[X] \cdot Y\big] + \mathrm{E}\big[\mathrm{E}[X]\,\mathrm{E}[Y]\big] \\
&\qquad \text{(by LOE)}
\end{aligned}$$

Observing that $\mathrm{E}[S]$ is a constant for any random variable S, and noting that $\mathrm{E}[c] = c$ for all constants $c \in \mathbb{R}$, we see:

$$\begin{aligned} &= \mathrm{E}[XY] - \mathrm{E}[Y]\,\mathrm{E}[X] - \mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= \mathrm{Cov}(X, Y) && \text{(by definition)} \end{aligned}$$

Now applying this result gives:

$$\begin{aligned} \mathrm{Cov}(X, X) &= \mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])] \\ &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{Var}(X) \end{aligned}$$

(c) We begin by expanding the definition of variance:

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathrm{E}\left[(X + Y)^2\right] - \mathrm{E}[X + Y]^2 \\ &= \mathrm{E}\left[X^2 + 2XY + Y^2\right] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 && \text{(by LOE)} \\ &= \mathrm{E}\left[X^2\right] + 2\mathrm{E}[XY] + \mathrm{E}\left[Y^2\right] - \left(\mathrm{E}[X]^2 + 2\mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[Y]^2\right) && \text{(by LOE)} \end{aligned}$$

Rearranging this gives:

$$\begin{aligned} &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 + \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2 + 2\mathrm{E}[XY] - 2\mathrm{E}[X]\,\mathrm{E}[Y] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) \end{aligned}$$

Problem 4:

Waley is prepping for an Easter Egg hunt. He has a basket of n eggs, and independently for each egg, paints it either red with probability $\frac{1}{3}$ or blue otherwise. He defines an equivalence relation ρ where $x \; \rho \; y$ if and only if x and y are the same color. What is $\mathrm{E}[|\rho|]$?

Solution:

Call the set of eggs E. Consider the sample space:

$$\Omega = \{R, B\}^n$$

where each outcome lists the colors of the eggs in a fixed order. We define:

Let R_x be the event that egg x is colored red, $x \in E$.
Let B_x be the event that egg x is colored blue, $x \in E$.
Let $S_{x,y}$ be the event that eggs x and y have the same color, $x, y \in E$.
Let $|\rho|$ be the random variable denoting the cardinality of ρ.
Let $I_{S_{x,y}}$ be an indicator random variable for $S_{x,y}$, $x, y \in E$.

We seek $\mathrm{E}[|\rho|]$. All elements of ρ are ordered pairs (x, y) where $x, y \in E$, and x and y have the same color. We thus see that:

$$|\rho| = \sum_{x,y} I_{S_{x,y}} = \sum_{x} I_{S_{x,x}} + \sum_{x \neq y} I_{S_{x,y}}$$

By the Linearity of Expectation, we have that:

$$\begin{aligned}
\mathrm{E}[|\rho|] &= \mathrm{E}\left[\sum_{x} I_{S_{x,x}} + \sum_{x \neq y} I_{S_{x,y}}\right] \\
&= \sum_{x} \mathrm{E}\left[I_{S_{x,x}}\right] + \sum_{x \neq y} \mathrm{E}\left[I_{S_{x,y}}\right] \\
&= \sum_{x} \Pr[S_{x,x}] + \sum_{x \neq y} \Pr[S_{x,y}]
\end{aligned}$$

Note that $\Pr[S_{x,x}] = 1$, $\forall x \in E$, since an egg is always the same color as itself. For the case where $x \neq y$, we see that $S_{x,y} = (R_x \cap R_y) \cup (B_x \cap B_y)$. Since these events are disjoint, we can add their probabilities with the Sum Rule:

$$= \sum_{x} 1 + \sum_{x \neq y} \Pr[R_x \cap R_y] + \Pr[B_x \cap B_y]$$

Observing that $R_x \perp R_y$ and $B_x \perp B_y$ when $x \neq y$, since Waley colors eggs independently:

$$\begin{aligned}
&= n + \sum_{x \neq y} \left(\Pr[R_x] \times \Pr[R_y] + \Pr[B_x] \times \Pr[B_y]\right) \\
&= n + \sum_{x \neq y} \left(\frac{1}{3} \times \frac{1}{3} + \frac{2}{3} \times \frac{2}{3}\right) \\
&= n + \frac{5}{9} n(n-1) = \frac{5n^2 + 4n}{9}
\end{aligned}$$