# 21-301 Combinatorics
# Section B
## Assignment 3 - Solutions

1.*

(a) Prove the linearity of expectation, namely that for a finite collection of discrete random variables $X_1, X_2, X_3, \ldots, X_n$ on a single probability space, we have that

$$\mathbb{E}\left(\sum X_i\right) = \sum \mathbb{E}(X_i).$$

**Solution** Let $\Omega$ be our probability space and $X_i(\omega)$ be the value that $X_i$ takes when the event $\omega \in \Omega$ occurs. We note that $\left(\sum_i (X_i)\right)(\omega)$, the value that the sum of the $X_i$ takes when $\omega$ occurs is equal to $\sum_i ((X_i)(\omega))$ since the $X_i$ are variables on the same probability space. Using this, we have

$$\begin{aligned}
\mathbb{E}\left(\sum_i X_i\right) &= \sum_{\omega\in\Omega}\left(\sum_i (X_i)\right)(\omega)Pr(\omega)\\
&= \sum_{\omega\in\Omega}\left(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)\right)Pr(\omega)\\
&= \sum_{\omega\in\Omega}\left(\sum_i X_i(\omega)\right)Pr(\omega)\\
&= \sum_i \sum_{\omega\in\Omega}\left(X_i(\omega)Pr(\omega)\right) \text{ (since both sums are finite.)}\\
&= \sum_i \mathbb{E}(X_i).
\end{aligned}$$

(b) Prove the first moment method, namely that if $X$ is a random variable taking values in $\{0, 1, 2, 3, \dots\}$, then

$$Pr[X \geq 1] \leq \mathbb{E}(X).$$

**Solution**

$$\mathbb{E}(X) = (X|X=0)Pr(X=0) + (X|X=1)Pr(X=1) + (X|X=2)Pr(X=2) + \dots$$

The value of $(X|X=i)$ is precisely $i$ and so the above is equal to

$$\begin{aligned}\mathbb{E}(X) =& 0Pr(X=0) + 1Pr(X=1) + 2Pr(X=2) + 3Pr(X=3) + \dots \\ \geq& Pr(X=1) + Pr(X=2) + Pr(X=3) + \dots \\ =& Pr(X \geq 1)\end{aligned}$$

2.* Prove that for a graph with $2n$ vertices and $m > 0$ edges, it is possible to partition the vertices of the graph into two sets of size $n$ such that more than $m/2$ edges go between these two sets.

**Solution** Partition $V$, the vertex set of the graph $G$ into two sets $V_1$ and $V_2$ by choosing $V_1$ uniformly at random from all possible subsets of $V$ of size $n$. Let $V_2 = V \setminus V_1$. There are $\binom{2n}{n}$ ways of choosing $V_1$. let $E(G)$ be the edge set of $G$.

Let $X$ be the random variable denoting the number of edges of $G$ that go between these two sets. For each edge $e$, the $X_e$ be the indicator variable for $e$ going between $V_1$ and $V_2$, i.e. if $e = \{x, y\}$, then

$$X_e = \begin{cases} 1 & \text{if } \{x \in V_1 \text{ and } y \in V_2\} \text{ or if } \{y \in V_1 \text{ and } x \in V_2\} \\ 0 & \text{otherwise.} \end{cases}$$

From this we have $X = \sum_{e \in E(G)} X_e$ and so $\mathbb{E}X = \sum_{e \in E(G)} \mathbb{E}X_e$ by linearity of expectation. For each of the $\binom{2n}{n}$ choices for $V_1$ and for $e = \{x, y\}$, we know that $x$ lies in $V_1$ or in $V_2$. If we require $y$ to be in the other set, then the remaining $2n-2$ vertices can be put into $V_1$ or $V_2$ freely until an extra $n-1$ vertices have been added to each set. There are $\binom{2n-2}{n-1}$ ways of doing this and since each outcome is equally likely the expected value of $X_e$

satisfies,

$$\begin{aligned}
\mathbb{E}(X_e) =& 0Pr(x, y \in V_1 \text{ or } x, y \in V_2) + 1Pr(x \in V_1, y \in V_2) + 1Pr(x \in V_2, y \in V_1)\\
=& 1Pr(x \in V_1, y \in V_2) + 1Pr(x \in V_2, y \in V_1)\\
=& \frac{\binom{2n-2}{n-1}}{\binom{2n}{n}} + \frac{\binom{2n-2}{n-1}}{\binom{2n}{n}} = 2\frac{(2n-2)(2n-3)\dots n}{(n-1!)}\frac{n!}{2n(2n-1)\dots(n+1)}\\
=& \frac{2n^2}{2n(2n-1)} = \frac{n}{2n-1} > \frac{1}{2}.
\end{aligned}$$

Using linearity of expectation, gives us that

$$\mathbb{E}(X) = \sum_{e \in E(G)} X_e > |E(G)|\frac{1}{2},$$

and since there must exist a choice of $V_1$ such that the value of $X$ is at least $E(X)$ we have that there must exist a choice of $V_1$ with $X$ which is the number of edges going from $V_1$ to $V_2$, greater than half the number of edges, as required.

3.* Prove that if $\binom{n}{k}3^{1-\binom{k}{2}} < 1$ then it is possible to colour the edges of $K_n$, the complete graph on $n$ vertices, with three colours such that there is no monochromatic $K_k$.

**Solution** Randomly colour the edges of $K_n$ with each edge receiving one of three colours, uniformly at random, independently of the choice for any other edge. For each subgraph of $K_n$ of size $k$, it contains $\binom{k}{2}$ edges and so the probability that this clique is entirely coloured with the first colour is $\frac{1}{3^{\binom{k}{2}}}$. This is equally true for either of the other two colours, so the probability that it is mono-coloured is $\frac{3}{3^{\binom{k}{2}}} = 3^{1-\binom{k}{2}}$. There are $\binom{n}{k}$ choices for a subset of size $k$ and so the expected number of mono-coloured $K_k$ is

$$\binom{n}{k}3^{1-\binom{k}{2}} < 1$$

by our initial assumption. Since the expectation is strictly less than 1, there must exist some colouring where the number of mono-coloured $K_k$ is less than 1. Since this can only take integer values, there must exist a colouring with no mono-coloured $K_k$ as required.

4. * In an $n \times n$ array, each of the numbers $1, 2, \ldots, n$ appears exactly $n$ times (not randomly). Let $X$ be the random variable determined by counting the number of distinct numbers in a randomly chosen row or column. Use $X$ to prove that there must exist a row or column containing at least $\sqrt{n}$ distinct numbers.

**Solution** Choose a random row or column, where each of the $2n$ choices are equally likely. For each $i \in \{1, 2, \ldots n\}$ let $I_i$ be the indicator variable for $i$ appearing in your randomly chosen row, i.e. $I_i = 1$ if $i$ is present in your row (whether it appears once or multiple times) and 0 otherwise. clearly $X = \sum_{i=1}^{n} I_i$. We note that the expectation of $I_i$ is equal to the number of rows and columns that $i$ appears in, divided by the total number of rows and columns.

We find a lower bound on $\mathbb{E}(I_i)$ by noting that the smallest number of rows and columns it can be present in is if all of the $n$ values of $i$ appear in a single $\sqrt{n} \times \sqrt{n}$ sub array. This is because if $i$ appears in distinct positions $(j, k)$ and $(x, y)$, then putting an extra $i$ in positions $(j, y)$ or $(x, k)$ in the array does not increase the number of rows and columns it appears in and so minimises the expectation. Continuing this way, we say that if we must add extra values of $i$ (which we have to do until there are $n$ present in the array) we can minimise the expectation by building up a square block of all $i$ values.

This tells us that the expectation is minimised when $i$ appears in $2\sqrt{n}$ of the total $2n$ possible rows and columns.

Therefore,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(I_i) \geq \sum_{i=1}^{n} \frac{2\sqrt{n}}{2n} = n\frac{\sqrt{n}}{n} = \sqrt{n}.$$

Since the expected number of distinct $i$ in our randomly chosen row or column is $\sqrt{n}$, there must be a row or column with at least this many distinct elements in it.