CIS 1600

Recitation Guide - Week 9

Topics Covered: Variance, Markov's Inequality, Bipartite Graphs

Problem 1:

A 10 digit number with no zeroes is chosen by independently and randomly selecting each digit (1 - 9).

a) Let N be the number of digits missing from the 10 digit number. For example, if the number is 1231452832, then we are missing the digits 6, 7, 9 so $N = 3$. Find $\mathbb{E}[N]$ and $\text{Var}[N]$.

b) Using Markov's Inequality, what is the lower bound of the probability that less than 6 digits are missing?

Solution:

a) Define $\Omega = \{x_1x_2...x_{10} \mid x_i \in [1..9]\}$. Note that we have a uniform probability distribution, and $|\Omega| = 9^{10}$, as we have 9 choices for each x_i.

Let N be a random variable that represents the number of digits missing from the 10-digit number. Let N_i be an indicator random variable that is 1 if digit i is missing and 0 otherwise, for $1 \leq i \leq 9$. Notice that $N = \sum_{i=1}^{9} N_i$.

Additionally, note that $\Pr[N_i = 1] = \left(\frac{8}{9}\right)^{10}$, because the 10 digits are selected independently, and for each, there is an $\frac{8}{9}$ chance that i is NOT the digit selected.

$$\begin{aligned}
\mathbb{E}[N] &= \mathbb{E}\left[\sum_{i=1}^{9} N_i\right] \\
&= \sum_{i=1}^{9} \mathbb{E}[N_i] && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^{9} \Pr[N_i = 1] \\
&= 9 \cdot \left(\frac{8}{9}\right)^{10} \\
&\approx 2.772
\end{aligned}$$

In order to calculate the variance, we have to compute $\mathbb{E}[N^2]$. Notice that

$$\begin{aligned}
\mathbb{E}[N^2] &= \mathbb{E}[(N_1 + N_2 + \ldots + N_9)^2] \\
&= \mathbb{E}\left[\sum_{i=1}^{9} N_i^2 + \sum_{i \neq j} N_i \cdot N_j\right] \\
&= \sum_{i=1}^{9} \mathbb{E}[N_i^2] + \sum_{i \neq j} \mathbb{E}[N_i \cdot N_j] && \text{(by Linearity of Expectation)}
\end{aligned}$$

where there are $9 \cdot 8 = 72$ terms of the form $N_i \cdot N_j$, $i \neq j$.

This expansion will yield $9^2 = 81$ total terms because each term consists of an indicator from the first sum and an indicator from the second sum. Observe that each term of the form $N_i^2, i \in [1..9]$ comes from choosing N_i from the first sum and then choosing N_i from the second sum. Because there are 9 indicators, there must be 9 of these terms. This leaves $81 - 9 = 72$ terms of the form $N_i \cdot N_j$ where $i, j \in [1..9], i \neq j$.

We can once again apply independence of the 10 digits to argue that $\Pr[N_i \cdot N_j = 1] = \left(\frac{7}{9}\right)^{10}$, since each digit can be any of the 7 digits that aren't i or j.

Note that

$$\mathbb{E}[N_i^2] = \mathbb{E}[N_i] = \Pr[N_i = 1] = \left(\frac{8}{9}\right)^{10}$$

and further

$$\mathbb{E}[N_i \cdot N_j] = \Pr[N_i = 1 \cap N_j = 1] = \left(\frac{7}{9}\right)^{10}$$

Thus,

$$\mathbb{E}[N^2] = 9\left(\frac{8}{9}\right)^{10} + 72\left(\frac{7}{9}\right)^{10}$$

Finally,

$$\begin{aligned}\mathrm{Var}[N] &= \mathbb{E}[N^2] - \mathbb{E}[N]^2 \\ &= 9\left(\frac{8}{9}\right)^{10} + 72\left(\frac{7}{9}\right)^{10} - 81\left(\frac{8}{9}\right)^{20} \approx 0.9232\end{aligned}$$

b) We are looking to lower-bound $\Pr[N < 6]$. Note that Markov's Inequality gives information about upper bounds on the probability that N is large. However, we also know that $\Pr[N < 6] = 1 - \Pr[N \geq 6]$. Also, keep in mind we can apply Markov's Inequality because N represents the number of missing digits, so N is a non-negative random variable. We begin from the information that Markov's Inequality guarantees us:

$$\begin{aligned}\Pr[N \geq a] &\leq \frac{\mathbb{E}[N]}{a} && \text{(Markov's Inequality)} \\ \Pr[N \geq 6] &\leq \frac{\mathbb{E}[N]}{6} && (a = 6) \\ \Pr[N \geq 6] &\leq \frac{2.772}{6} && (\mathbb{E}[N] \approx 2.772)\end{aligned}$$

Solving for lower bound,

$$\begin{aligned}\Pr[N < 6] &= 1 - \Pr[N \geq 6] \\ &\geq 1 - \frac{2.772}{6} \\ &= 0.5381\end{aligned}$$

Problem 2:

Prove that a graph is bipartite if and only if it has no odd length cycles.

Solution:

($\Longrightarrow$)

First let us prove that if a graph is bipartite, then it has no odd cycles. Let $G = (U, V, E)$ be a bipartite graph, where U is the set of vertices being assigned one color, V is the set of vertices of the other color, and E is the set of edges of G . Suppose for the sake of contradiction that it has some odd cycle C of length $2k + 1$ and let $C = x_1, x_2, \ldots x_{2k+1}, x_1$, where x_i is the i^{th} vertex in C. WLOG, let x_1 be in U.

Note that the i^{th} vertex in C is in U if i is odd, and in V if i is even, by the nature of bipartite graphs. Thus, x_{2k+1} must be in U. However, there is an edge between x_{2k+1} and x_1, which is also in U, by definition of the cycle. This is a contradiction to the fact that G is bipartite, since two vertices in U share an edge.

($\Longleftarrow$)

Now let us prove that if a graph has no odd cycles, then it is a bipartite graph. Let $P(m)$ be the following property:

> If G is a graph with m edges and no odd cycles, then it is bipartite.

We wish to prove $P(m)$, for $m \in \mathbb{N}$. We proceed by induction on m.

Base Case: $m = 0$. The graph is bipartite – any partition of the vertices U, V will suffice.

Induction Hypothesis: Assume that $P(k)$ holds for some $k \in \mathbb{N}$.

Induction Step: Let G be a graph with $k + 1$ edges and no odd length cycles. Let us consider two different cases.

Case 1: The graph is acyclic

We have a forest! Let us select an arbitrary edge $e = \{x, y\}$ where y is a leaf. Let $G' = G \setminus \{e\}$. Note that G' is also a forest, so by the induction hypothesis, we have that G' is bipartite. Let $G' = (U', V', E')$.

We want to now show that we can express $G = (U, V, E)$. Now let us consider what happens when we add e to G'. Assume W.L.O.G. that $x \in U'$. Then let $U = U' \setminus \{y\}$ and $V = V' \cup \{y\}$. In other words, keep the partitions from G' and fix y into V'. Since y was an isolated vertex in G', the change in the partition that it belongs to does not affect the bipartiteness of the graph. Further, e crosses the partition U, V as needed.

Case 2: The graph has at least one even cycle

Select an arbitrary edge $e = \{x, y\}$ that belongs to a cycle C. Note that C is an even-length cycle because there are no odd-length cycles in G. Let $G' = G \setminus \{e\}$. Since G had no odd cycles, and we only removed an edge to create G', G' has no odd cycles. So by the induction hypothesis, we have that G' is bipartite. Let $G' = (U', V', E')$.

We want to now show that we can express $G = (U, V, E)$. Now let us consider what happens when we add e to G'. Note that when we removed e from C, we created an odd length path P from x to y in G'. Therefore, x and y must be in different partition in G'. W.L.O.G. let $x \in U'$ and $y \in V'$.

Therefore we can simply let $U = U'$ and $V = V'$, and observe that e crosses the partition U, V as required, so G is bipartite.

Alternate Solution ($\Longleftarrow$):

We first prove a lemma.

Lemma 1. *Any closed walk of odd length contains an odd length cycle.*

Proof. We prove this by induction on the length of the closed walk l. As we have that there are no self-loops, we prove for $l \geq 3$. When $l = 3$, we have that our closed loop must be an odd cycle, as we have no self-loops. Now, assume for a fixed odd integer k that for all odd integers j such that $3 \leq j \leq k$, any odd length closed walk of length j includes an odd length cycle. Consider an odd length closed walk of length $k + 2$, which we denote $v_1 - v_2 - \cdots - v_{k+2} - v_1$. If this closed walk is a cycle, we are done. Otherwise, there is some repeated vertex $v_i = v_w$, where $1 < i < w < k + 2$, Then, note that we can split this walk into two closed walks $v_1 - \cdots - v_i - v_{w+1} - \cdots - v_{k+2} - v_1$ and $v_i - v_{i+1} - \ldots v_{w-1} - v_w$. We know one of these must have odd length, and both must have length at least 3. Therefore, by our inductive hypothesis, we know that it must contain an odd cycle. $\square$

Now, consider a graph $G = (V, E)$ with no odd length closed walks. Note that this is at least as strong as assuming we have no odd length cycles. Note that it suffices to show that each connected component of G is bipartite, so without loss of generality we can assume G is connected. Fix a vertex $x \in V$. Then, we notice that we can partition the vertex set V into disjoint sets R, B, where $R = \{y \in V | d(x, y) \text{ is even}\}$ and $B = \{y \in V | d(x, y) \text{ is odd}\}$, where $d(u, v)$ denotes the length of the shortest path between u and v. We now show that R, B is a valid coloring.

Note that $x \in R$, as $d(x, x) = 0$. We note that x can share no edge with any other vertex in R, as that would yield an odd length shortest path. Now, assume for contradiction that there exist $y, z \in R$ such that $\{y, z\} \in E$. Then, we consider the shortest paths $x \rightsquigarrow y$ and $x \rightsquigarrow z$, both of which we know have even length. Note that the path $x \rightsquigarrow y - z \rightsquigarrow x$ constitutes a closed walk of odd length, a contradiction, as per our lemma. Similarly, assume that there exist $y, z, \in B$ such that $\{y, z\} \in E$. Note that, again, the path $x \rightsquigarrow y - z \rightsquigarrow x$ is a closed walk of odd length, another contradiction. As such, coloring all vertices in R red and all vertices in B blue is a valid coloring, and thus our graph is bipartite.