CIS 160

Recitation Guide - Week 11

Topics Covered: Bipartite Graphs, Variance, Tail Bounds

Problem 1:

A 10 digit number with no zeroes is chosen by independently and randomly selecting each digit (1 - 9). 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 $N = 3$.

a) Find $\mathbf{E}[N]$ and $\mathrm{Var}[N]$.

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

c) How can you improve the bound you obtained above?

Solution:

a) Define $\Omega = \{x_1 x_2 ... x_{10} \mid x_i \in [1..9]\}$. 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.

Using Linearity of Expectation, we have

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

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

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

where there are $9 \cdot 8 = 72$ terms of the form $N_i \cdot N_j$, $i \neq j$. (Note that $\sum_{i \neq j} \mathbf{E}[N_i \cdot N_j] = 2\sum_{i<j} \mathbf{E}[N_i \cdot N_j]$)

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

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

and further

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

Thus,

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

Finally,

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

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:

$$\begin{aligned}
\Pr[N \geq a] &\leq \frac{\mathbf{E}[N]}{a} && \text{(Markov's Inequality)} \\
\Pr[N \geq 6] &\leq \frac{\mathbf{E}[N]}{6} && (a = 6) \\
\Pr[N \geq 6] &\leq \frac{9 \cdot \left(\frac{8}{9}\right)^{10}}{6} && \text{(from part a)} \\
1 + \Pr[N \geq 6] &\leq 1 + \frac{9 \cdot \left(\frac{8}{9}\right)^{10}}{6} && \text{(adding one to both sides)} \\
1 - \frac{9 \cdot \left(\frac{8}{9}\right)^{10}}{6} &\leq 1 - \Pr[N \geq 6] && \text{(rearranging)} \\
1 - \frac{9 \cdot \left(\frac{8}{9}\right)^{10}}{6} &\leq \Pr[N < 6] \\
0.5381 &\leq \Pr[N < 6] && \text{(rounding)}
\end{aligned}$$

c) We can use Chebyshev's inequality:

$$\Pr[|N - \mathbf{E}[N]| \geq a] \leq \frac{\mathrm{Var}[N]}{a^2} \qquad \text{(Chebyshev's Inequality)}$$

since we have $\mathrm{Var}[N] \approx 0.9232$ from part a). Also from part a), we have that $\mathbf{E}[N] \approx 2.772$. Hence, we choose $a = 6 - 2.772 = 3.228$. We have

$$\Pr[|N - 2.772| \geq 3.228] \leq \frac{0.9232}{3.228^2} \approx 0.0886$$

As N is non-negative, we have that

$$\Pr[N \geq 6] \leq 0.0886$$

Using the same rearranging of terms from part b), we get

$$\Pr[N < 6] \geq 0.9114$$

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. 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. 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.