CMSC 498K : Homework 2

Noah Luck Easterly

March 6, 2008

Problem 1.

Suppose that a document contains 300 pages and contains 20 misprints. What is the probability that there is more than one misprint on a particular page?

Assuming that all the misprints are independently randomly distributed, we have

$$\begin{aligned}
Prob(\text{ misprint } i \text{ is on page } j\ ) &= \frac{1}{300} \\
Prob(\text{ page } j \text{ has exactly } k \text{ misprints }) &= \binom{20}{k}\left(\frac{1}{300}\right)^k\left(\frac{299}{300}\right)^{20-k} \\
Prob(\text{ page } j \text{ has } m \text{ or more misprints }) &= \sum_{k=m}^{20} Prob(\text{ page } j \text{ has exactly } k \text{ misprints }) \\
&= 1 - \sum_{k=1}^{m-1} Prob(\text{ page } j \text{ has exactly } k \text{ misprints })
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
&Prob(\text{ page } j \text{ has more than one misprint }) \\
&= 1 - [Prob(\text{ page } j \text{ has exactly 0 misprints }) + Prob(\text{ page } j \text{ has exactly 1 misprint })] \\
&= 1 - \left[\binom{20}{0}\left(\frac{1}{300}\right)^0\left(\frac{299}{300}\right)^{20} + \binom{20}{1}\left(\frac{1}{300}\right)^1\left(\frac{299}{300}\right)^{19}\right] \\
&= 1 - \left[\left(\frac{299}{300}\right)^{20} + 20 * \frac{299^{19}}{300^{20}}\right] \\
&= 1 - \frac{299^{19}}{300^{20}}\left[299 + 20\right] \\
&= 1 - \frac{299^{19}}{300^{20}} * 319 \approx 2.03 * 10^{-3}
\end{aligned}$$

Problem 2.

Suppose 4% of the parts made in a factory are defective. Suppose we ship out a batch of 25 parts. What is the probability that there are no defective parts in this batch?

Assuming that all occurrences of parts are independent, we have

$$Prob(\text{ any particular part is defective }) = \frac{1}{25}$$

$\Rightarrow$

$$\begin{aligned} Prob(\text{ in a batch of } n \text{ parts, exactly } k \text{ are defective }) &= \binom{n}{k}\left(\frac{1}{25}\right)^k \left(\frac{24}{25}\right)^{n-k} \\ Prob(\text{ in a batch of 25 parts, exactly 0 are defective }) &= \binom{25}{0}\left(\frac{1}{25}\right)^0 \left(\frac{24}{25}\right)^{25} \\ &= \left(\frac{24}{25}\right)^{25} \approx 0.360 \end{aligned}$$

Problem 3.

Suppose we toss 1000 fair coins. Use Chernoff bounds to derive an upper bound on the probability that we either get less than 400 heads or more than 600 heads is small.

A fair coin means that

$$Prob(\text{head}) = Prob(\text{tail}) = \frac{1}{2}$$

Each fair coin is independent. Define 1000 random variables

$$X_i = \begin{cases} 1, & \text{if the } i\text{th coin is a head;} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$Prob(X_i = 1) = \frac{1}{2}$$

Let

$$X = \sum_{i=1}^{1000} X_i$$

Then

$$\mu = E[X] = E[\sum_{i=1}^{1000} X_i] = \sum_{i=1}^{1000} E[X_i] = \sum_{i=1}^{1000} \frac{1}{2} = 500$$

So

$$Prob(\text{less than 400 heads or more than 600 heads}) = Prob(|X - \mu| \geq 100)$$

The corollary to Chernoff's bounds states:

$$Prob(|X - \mu| \geq \delta\mu) \leq 2e^{(-\mu\delta^2)/2}$$

So, with $\delta = \frac{1}{5}$, we have,

$$Prob(\text{less than 400 heads or more than 600 heads}) \leq 2e^{(-500*\frac{1}{5}^2)/2} = 2e^{-10} \approx 4.54 * 10^{-5}$$

Problem 4.

Suppose we randomly label n nodes of a graph using labels in $\{1,\ldots,n\}$. Suppose we now select a node if its label value is smaller than the labels of all its neighbors. What is the probability that a node v is selected? Also prove that two adjacent nodes cannot be selected.

Let v_k be the node with label k. Due to the selection method

$$Prob(\text{ } v_k \text{ is selected }) = Prob(\text{ } v_k \text{ is connected only to nodes later in labeled order })$$

Suppose we know the degree d_k of each node v_k. There are $\binom{n-1}{d_k}$ different sets of size d_k among the other $n-1$ nodes. If we assume that each node's edges are evenly distributed among the other edges, the number of ways of choosing d_k nodes from the $n-k$ nodes later in labeled order than v_k is $\binom{n-k}{d_k}$. Therefore,

$$Prob(\text{ } v_k \text{ is selected }) = \frac{\binom{n-k}{d_k}}{\binom{n-1}{d_k}}$$

If we don't know the degree of each node, but only know the total number of edges m, we approximate d_k by assuming the edges are evenly distributed through the nodes, so that each node has degree, $d = \frac{2m}{n}$. This gives

$$Prob(\text{ } v_k \text{ is selected}) = \frac{\binom{n-k}{d}}{\binom{n-1}{d}}$$

v_1 is always selected (as it has the minimum label), so $Prob(\text{ } v_1 \text{ is selected }) = 1$. Under our approximation, for $k > n-d$, v_k is never selected, as at least one of the d nodes v_k is connected to must have label $\leq n-d$, so $Prob(\text{ } v_k \text{ is selected }, k > n-d) = 0$.

For $k \in (1, n-d]$, we have

$$\begin{aligned}
Prob(\text{ } v_k \text{ is selected}) &= \frac{\binom{n-k}{d}}{\binom{n-1}{d}} = \frac{\frac{(n-k)!}{d!(n-k-d)!}}{\frac{(n-1)!}{d!(n-1-d)!}} \\
&= \frac{\frac{(n-k)!}{(n-k-d)!}}{\frac{(n-1)!}{(n-1-d)!}} = \frac{\frac{(n-1-d)!}{(n-k-d)!}}{\frac{(n-1)!}{(n-k)!}} \\
&= \prod_{i=1}^{k-1} \frac{n-i-d}{n-i} \\
\Rightarrow & \\
Prob(\text{ } v_k \text{ is selected}) &= \left(\frac{n-k+1-d}{n-k+1}\right) \prod_{i=1}^{k-2} \frac{n-i-d}{n-i} \\
&= \left(\frac{n-k+1-d}{n-k+1}\right) Prob(\text{ } v_{k-1} \text{ is selected})
\end{aligned}$$

Which gives a natural sense of how the probability changes as k increases in this scenario.

To see that no two adjacent nodes can both be selected, consider adjacent nodes v_j, v_k. If v_j is selected, this means that $j < i, \forall v_i$ adjacent to v_j. In particular, this means that $j < k$. But this means $\exists v_i = v_j$, s.t. v_k adjacent to $v_i, i < k$, so v_k cannot be selected.

Problem 5.

Suppose we have a collection of sensor that monitor some targets. Assume that each target can be monitored by exactly two sensors. Construct a graph $G = (V, E)$ where nodes in V correspond to sensors, and corresponding to each target there is an edge in the graph connecting the sesnsor nodes. Our goal is to put the sensors into an on-off cycle to save batter power. Lets assume that we partition the sensors into two groups A and B, such that each sensor is in exactly one group and the groups are disjoint. Sensors in group A are switched on for some time, and then they are shut off and sensors in group B are switched on. Some targets are always monitored since one of the two sensors is always on.

* Give an example to show that it may be the case that no matter how we come up with a partitioning of the sensors in some cases there is no schedule that always monitors all targets.

The simple example is the triangle graph, with three edges connecting three nodes. In any partition, one group must contain at least two edges, which means that the other partition cannot cover the point where those two edges meet.

* If the graph has the property that such a schedule exists then develop an algorithm to find it.

Unfortunately, I was unable to think of any algorithm better than an exhaust, which takes $O((|V| + |E|)2^{|E|-1})$.

Input: Graph $G = (V, E)$
Output: Partition (A, B)
Choose $e \in E$
foreach $A \in \mathbb{P}(E - \{e\})$ **do**
  $B \leftarrow E - A$
  foreach $v \in V$ **do**
    $countA[v] \leftarrow 0$
    $countB[v] \leftarrow 0$
  end
  foreach $(u, v) \in A$ **do**
    $countA[u] \leftarrow countA[u] + 1$
    $countA[v] \leftarrow countA[v] + 1$
  end
  foreach $(u, v) \in B$ **do**
    $countB[u] \leftarrow countB[u] + 1$
    $countB[v] \leftarrow countB[v] + 1$
  end
  $done \leftarrow$ true
  foreach $v \in V$ **do**
    if $countA[v] = 0$ *or* $countB[v] = 0$ **then**
      $done \leftarrow$ false
    end
  end
  if $done$ **then**
    return (A, B)
  end
end

Since this iterates over all partitions where $e \in B$, and the labeling of the partitions is arbitrary, it's easy to see that this checks all possibile partitions. So if one exists in which both groups are edge covers of the graph, it will be found.

There exists a polynomial algorithm to find a single, minimal edge cover, but I don't know it. I suspect it could be adapted to find a partition s.t. each side is an edge cover, though.

* Suppose we partition the sensors randomly. In other words, each sensor joins partition A with probability 0.5 and B with probability 0.5. What is the expected number of targets that are covered in both time slots?

If a target v is covered by d_v sensors, then the probability that all those sensors are in group A is $\left(\frac{1}{2}\right)^{d_v}$. The same for group B. Therefore, the probability that it is covered by sensors from both groups is $1 - 2\left(\frac{1}{2}\right)^{d_v} = 1 - 2^{1-d_v}$.

If we define the indicator random variable

$$X_v = \begin{cases} 1, & v \text{ is covered by both } A \text{ and } B; \\ 0, & \text{otherwise.} \end{cases}$$

So X, the total number of sensors covered by both halves of the partition, is $X = \sum_v X_v$, then

$$E[X] = E[\sum_{v \in V} X_v] = \sum_{v \in V} (1 - 2^{1-d_v})$$

If we approximate d_v by $d = \frac{2|E|}{|V|}$, the average number of sensors per target, then

$$E[X] = \sum_{v \in V} (1 - 2^{1-d}) = |V|(1 - 2^{1-d})$$