CS584
HW#5

Due: November 19

1. (12 points) In class we presented an algorithm to represent a set *S* using a Bloom Filter of *m* bits and *k* hash functions. Show how to use two Bloom filters to estimate, for any sets *S* and *T*, the cardinality of *S-T* (the number of elements in *S* which are not in *T*). Note that you may only examine the Bloom Filters, not the original sets *S* and *T*. You should assume that they each use the same sequence of hash functions.

2. (5 points) In class it was stated that providing a stream of random bits is expensive. Suppose you have access to a stream of bits $b_1, b_2, b_3, \ldots$ such that for each *i*, $\Pr\{b_i = 0\} = p$ for some $0 < p < 1$ and for each pair *i, j*, $i \neq j$, $\Pr\{b_i = 0\}$ is independent of $\Pr\{b_j = 0\}$. Show how to convert this to another stream of bits $a_1, a_2, a_3, \ldots$ such that for each *i*, $\Pr\{a_i = 0\} = 1/2$ and for each pair $i, j, 1 \le i < j$, $\Pr\{a_i = 0\}$ is independent of $\Pr\{a_j = 0\}$. Make sure that your new stream is infinite for any $0 < p < 1$.

3. (12 points) Strassen's Algorithm in **Section *28.2*** of our text accepts as input $n \times n$ matrices *A* and *B* and contructs *C=AB* in time $\Theta\left(n^{\lg 7}\right) = O\left(n^{2.81}\right)$. In this problem we're trying to solve an easier problem. We are accepting as input $n \times n$ matrices *A, B* and *C* and we want to **test** if *C=AB*. We want a randomized algorithm which may be wrong, but our algorithm should work in time in $O\left(n^2\right)$. And we want to be able to make the probability of being wrong arbitrarily small.

a Let *D* be any nonzero $n \times n$ matrix and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be any *n*-dimensional vector for which each entry x_i is randomly chosen to be 0 or 1 with probability ½, independent of all x_j, $j \neq i$. Show that $\Pr\left\{ Dx = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \right\} \le$ ½.

b Use the result of ***a*** to construct a $O\left(n^2\right)$ time randomized test of $AB \neq C$.

4. (10 points) There are *n* agents which are identical except that each agent has a unique name. In order to select a leader the agents agree to use a hash function $h : \{\text{names}\} \to [0 \ldots b]$ to distribute names independently and uniformly over $[0 \ldots b]$. The agent with the highest hash value becomes the leader. What is the probability that a unique leader is chosen? You needn't find a closed form for your answer.

CS584
HW#5 SOLUTIONS

1. We assume *S* is stored in Bloom Filter $A[0..m-1]$ and *T* is stored in $B[0..m-1]$. We let $\omega(A)$ (respectively $\omega(B)$) denote the number of 1's in *A* (resp. *B*), and we define array boolean array $A \wedge B[0..m-1]$ by

$$A \wedge B[i] = \begin{cases} 1, \text{ if } A[i] = B[i] = 1 \\ 0, \text{ otherwise} \end{cases}, \quad 0 \le i < m \quad .$$

We assume that $\frac{\omega(A)}{m}$ and $\frac{\omega(B)}{m}$ are small, and hence $\Pr\{A[i] = B[i] = 1\} \ll 1$ for all $0 \le i \le m-1$. With every bit $A[i], 0 \le i \le m-1$ (respectively $B[i]$), we associate random variable $X_i = A[i]$ (respectively Y_i), and the number of 1's in *A* (resp. in *B*) is $X = \sum_{0 \le i \le m-1} X_i$ $\left(Y = \sum_{0 \le i \le m-1} Y_i \right)$. By linearity of expectation,

$$E[X] = E\left[\sum_{0 \le i \le m-1} X_i \right] = \sum_{0 \le i \le m-1} E[X_i] = 1 - \left(1 - \frac{1}{m}\right)^{k|S|} \approx 1 - e^{-k|S|/m}$$

so an estimate of $|S|$ is obtained solving $\omega(A) = 1 - e^{k|S|/m}$ for $|S|$, which yields $k|S|/m = \ln(1-\omega(A))$, or $|S| = \frac{m \ln(1-\omega(A))}{k}$.

Since $S = (S-T) \cup (S \cap T)$, it follows that $|S-T| = |S| - |S \cap T|$, and we estimate $|S-T|$ to be $\frac{m \ln(1-\omega(A))}{k} - \frac{m \ln(1-\omega(A \wedge B))}{k} = \frac{m}{k}\left(\ln(1-\omega(A)) - \ln(1-\omega(A \wedge B))\right)$.

2. You want to decompose stream $b_1, b_2, b_3, \ldots$ into an infinite sequence of events (bits) each with probability ½.

$i \leftarrow 1$
repeat forever
  if $b_i = 0 \wedge b_{i+1} = 1$ **then return** 0
  if $b_i = 1 \wedge b_{i+1} = 0$ **then return** 1
  $i \leftarrow i + 2$

To see that the stream $a_1, a_2, a_3, \ldots$ is indeed infinite, we note that $\Pr\{b_i \ne b_{i+1}\} = 2p(1-p)$. So the number of *a*'s generated from $b_1, b_2, b_3, \ldots, b_n$ is binomially distributed, with expectation $p(1-p)(n-1)$, which clearly goes to infinity as *n* goes to infinity.

3. ***a*** From our class notes, **THEOREM** (Zippel, Schwartz): If $P(x_1,...,x_n)$ is a nonzero polynomial of degree *d* over field *F* and $S \subseteq F$ and $(s_1,...,s_n)$ is a random element of S^n, then $\Pr\{P(s_1,...,s_n)=0\} \le \frac{d}{|S|}$.

For each nonzero row $M_i = (m_{i,1},...,m_{i,n})$ of *M* (there must be at least one), $M_i x$ is a nonzero polynomial over $\{0,1\}^n$, and $\Pr\{M_i x = 0\} \le 2^{-n}$ where *x* is a random element of $\{0,1\}^n$.

b To check if $AB \stackrel{?}{=} C$, we check if $AB - C \stackrel{?}{=} 0$. A necessary condition for $AB - C = 0$ is for $(AB-C)x = 0$ for all $x \in \{0,1\}^n$. By distributivity and associativity, this is equivalent to $A(Bx) - Cx = 0$, which can be evaluated for any $x \in \{0,1\}^n$ in time in $O(n^2)$. By the result of ***a***, $\Pr\{A(Bx) - Cx = 0\} \le$ ½, and we can repeat this test independently as often as necessary to make the probability of error arbitrarily small.

4. A unique leader is chosen if for some bucket b^*, some agent gets hashed to b^* and every other agent gets mapped to a bucket in $[0 \ldots b^* - 1]$. The probability of this happening for some fixed b^* is the probability that exactly one agent gets hashed to b^* times the probability that all the others get mapped to $[0 \ldots b^* - 1]$. There are $\binom{n}{1} = n$ ways to choose the agent to be mapped to b^*, and the probability of taking that agent to b^* is 1/*n*, and the probability of taking all the other agents to $[0 \ldots b^* - 1]$ is $\left(\frac{b^*}{b}\right)^{n-1}$. So the probability of a unique agent becoming a leader in b^* is $n\frac{1}{n}\left(\frac{b^*}{b}\right)^{n-1} = \left(\frac{b^*}{b}\right)^{n-1}$. To find the probability of a unique agent becoming a leader we sum these probabilities over all b^*, yielding $\sum_{0 \le b^* \le b} \left(\frac{b^*}{b}\right)^{n-1} = \left(\frac{1}{b}\right)^{n-1} \sum_{0 \le b^* \le b} (b^*)^{n-1}$.