CS 174: Combinatorics and Discrete Probability Fall 2012

Homework 10 Solution

Problem 1. *(Exercise 10.6 from MU – 8 points)* The problem of counting the number of solutions to a knapsack instance can be defined as follows: Given items with sizes $a_1, \ldots, a_n > 0$ and an integer $b > 0$, find the number of vectors $(x_1, x_2, \ldots, x_n) \in \{0,1\}^n$, such that $\sum_{i=1}^n a_i x_i \leq b$. The number b can be thought of as the size of a knapsack, and the x_i denote whether or not each item is put into the knapsack. Counting solutions corresponds to counting the number of different sets of items that can be placed in the knapsack without exceeding its capacity.

(a) A naïve way of counting the number of solutions to this problem is to repeatedly choose $(x_1, \ldots, x_n) \in \{0,1\}^n$ uniformly at random. If f is the fraction of valid solutions, then return $f \cdot 2^n$. Argue why this is not a good strategy in general; in particular, argue that it will work poorly when each a_i is 1 and $b = \sqrt{n}$.

Solution: The problem is that the solution set may not be dense enough in the sample space. Hence the uniform distribution over the whole sample space might not put enough weight on the solution set, and we may need to wait a long time before obtaining a good enough estimate of the number of solutions.

The expected number of 1's chosen this way is $n/2$. If $a_i = 1$ and $b = \sqrt{n}$, then by the Chernoff bound (Eqn 4.5 in chapter 4)

$$\Pr\left(\sum a_i X_i \leq \sqrt{n}\right) = \Pr\left(\sum a_i X_i \leq \left(1 - \frac{2}{\sqrt{n}}\right)\frac{n}{2}\right) \leq \exp\left(-\frac{n}{2}(\sqrt{n} - 2\sqrt{n})^2/2\right) = O(e^{-n})$$

(b) Consider a Markov chain, $X_0, X_1, \ldots$, on vectors $(x_1, \ldots, x_n) \in \{0,1\}^n$. Suppose that X_j is $(x_1, \ldots, x_n)$. At each time step, the Markov chain chooses $i \in \{1, \ldots, n\}$ uniformly at random. If $x_i = 1$, then X_{j+1} is obtained from X_j by setting x_i to 0. If $x_i = 0$, then X_{j+1} is obtained from X_j by setting x_i to 1 if doing so maintains the restriction $\sum_{i=1}^n a_i x_i \leq b$. Otherwise, $X_{j+1} = X_j$.

Argue that this Markov chain has a uniform stationary distribution whenever $\sum_{i=1}^n a_i > b$. Be sure to argue that the chain is irreducible and aperiodic.

Solution: We first show that the chain is irreducible and aperiodic over the states of all valid solutions of the knapsack problem. The chain is therefore ergodic and has a unique stationary distribution by Theorem 7.7. We then demonstrate that the uniform distribution satisfies the stationarity criterion.

Irreducibility: from any solution vector x, there is a positive probability of going back down to the all zero vector by zeroing out the non-zero x_i's one by one. On the other hand, if x is a solution, then it is possible to reach x starting from the all zero vector. Hence the chain is irreducible.

Aperiodicity: suppose $\sum_{i=1}^n a_i > b$, then there must exist $j \in [1, n]$ and a vector $x = (x_1, \ldots, x_j = 0, \ldots, x_n) \in \{0,1\}^n$ such that $\sum_i a_i x_i \leq b$ but $\sum_i a_i x_i + a_j > b$. This means that in the Markov chain, the self-loop probability $P_{x,x} > 0$. Hence the chain is aperiodic.

Let M be the number of solutions and let x and y be two solutions that differ by one bit. Then $P_{x,y} = 1/n$ and $P_{y,x} = 1/n$. If $\pi_x = \pi_y = 1/M$, then $\pi_x P_{x,y} = \frac{1}{Mn} = \pi_y P_{y,x}$. This satisfies the time-reversibility condition and proves that the stationary distribution is uniform.

(c) Argue that, if we have an FPAUS for the knapsack problem, then we can derive an FPRAS for the problem. To set up the problem properly, assume without loss of generality that $a_1 \leq a_2 \leq \cdots \leq a_n$. Let $b_0 = 0$ and $b_i = \sum_{j=1}^{i} a_i$. Let $\Omega(b_i)$ be the set of vectors $(x_1, \ldots, x_n) \in \{0,1\}^n$ that satisfy $\sum_{i=1}^{n} a_i x_i \leq b_i$. Let k be the smallest integer such that $b_k \geq b$. Consider the equation

$$|\Omega(b)| = \frac{|\Omega(b)|}{|\Omega(b_{k-1})|} \times \frac{|\Omega(b_{k-1}|}{|\Omega(b_{k-2})|} \times \cdots \times \frac{|\Omega(b_1)|}{|\Omega(b_0)|} \times |\Omega(b_0)|$$

You will need to argue that $|\Omega(b_{i-1})|/|\Omega(b_i)|$ is not too small. Specifically, argue that $|\Omega(b_i)| \leq (n+1)|\Omega(b_{i-1})|$.

Solution: The transformation of an FPAUS to an FPRAS is very similar to the example in the book. Once we show that $|\Omega(b_i)| \leq (n+1)|\Omega(b_{i-1})|$, we'll be able to get a run time bound that is polynomial in n. Consider the sets $A = \Omega(b_i)$ and $B = \Omega(b_{i-1})$. Since $b_i > b_{i-1}$, we must have $B \subset A$. We'll show that each point in $A \setminus B$ can be mapped to a point in B, and that using this mapping, at most n solutions in $A \setminus B$ will be mapped to the same point in B. The desired bound of $|A| \leq (n+1)|B|$ then follows.

Consider vector x. If $x \in A$ but $x \notin B$, then by definition, $\sum_{j=1}^{i-1} a_j = b_{i-1} < \sum_{j=1}^{n} a_j x_j \leq b_i = \sum_{j=1}^{i} a_j$. Hence there must be $k \geq i$ such that $x_k = 1$. Take the largest such k, set $x_k = 0$ and call the new vector $\tilde{x}$. Then $\tilde{x} \in B$ because

$$\sum_{j=1}^{n} a_j \tilde{x}_j = \sum_{j=1}^{n} a_j x_j - a_k \leq b_i - a_k \leq b_i - a_i = b_{i-1},$$

where we used the fact that $a_i \leq a_k$ for $k \geq i$. Since there are at most n choices for k, at most n solutions can be mapped to $\tilde{x}$.

To get an actual lower bound for M, notice that the only place that needs changing is lowerbound for $\mathbb{E}[\tilde{r}_i]$ on page 262. Most of the constants remain unchanged, except that the input size is now represented by n. To estimate $r_i = |\Omega(b_i)|/|\Omega(b_{i+1})|$, we will need a $\frac{\varepsilon}{6n}$-uniform sampler of $\Omega(b_{i+1})$. We just proved that $r_i \geq 1/(n+1)$, hence

$$\mathbb{E}[\tilde{r}_i] \geq r_i - \frac{\varepsilon}{6n} \geq \frac{1}{n+1} - \frac{1}{6n} = \frac{5n-1}{6n(n+1)}.$$

Applying Theorem 10.1, we see that we need the number of samples M to satisfy

$$M \geq \frac{3 \ln \frac{2n}{\delta}}{\left(\frac{\varepsilon}{12n}\right)^2} \cdot \frac{6n(n+1)}{5n-1} = 2592\varepsilon^{-2} \frac{n^3(n+1)}{5n-1} \ln \frac{2n}{\delta}.$$

The rest of the proof goes through without any other modifications.

Problem 2. *(Exercise 10.7 from MU – 6 points)* An alternative definition of an ϵ-uniform sample of Ω is as follows: A sampling algorithm generates an ϵ-uniform sample w if, for all $x \in \Omega$,

$$\frac{|\Pr(w = x) - 1/|\Omega||}{1/|\Omega|} \leq \epsilon.$$

Show that an ϵ-uniform sample under this definition yields an ϵ-uniform sample as given in Definition 10.3.

Solution: Suppose for all $x \in \Omega$,

$$\begin{aligned} \frac{|\Pr(w = x) - 1/|\Omega||}{1/|\Omega|} &\leq \epsilon \\ \Rightarrow |\Pr(w = x) - 1/|\Omega|| &\leq \frac{\epsilon}{|\Omega|} \end{aligned}$$

So

$$\begin{aligned} \left|\Pr(w \in S) - \frac{|S|}{|\Omega|}\right| &= \left|\sum_{x \in S} \left(\Pr(w = x) - \frac{1}{|\Omega|}\right)\right| \\ &\leq \sum_{x \in S} \left|\Pr(w = x) - \frac{1}{|\Omega|}\right| \\ &\leq \sum_{x \in S} \frac{\epsilon}{|\Omega|} \\ &= \epsilon \end{aligned}$$

Problem 3. *(Exercise 10.12 from MU – 6 points)* The following generalization of the Metropolis algorithm is due to Hastings. Suppose that we have a Markov chain on a state space Ω given by the transition matrix $\mathbf{Q}$ and that we want to construct a Markov chain on this state space with a stationary distribution $\pi_x = b(x)/B$, where for all $x \in \Omega$, $b(x) > 0$, and $B = \sum_{x \in \Omega} b(x)$ is finite. Define a new Markov chain as follows: When $X_n = x$, generate a random variable Y with $\Pr(Y = y) = Q_{x,y}$. Notice that Y can be generated by simulating one step of the original Markov chain. Set X_{n+1} to Y with probability

$$\min\left(\frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}}, 1\right),$$

and otherwise set X_{n+1} to X_n. Argue that, if this chain is aperiodic and irreducible, then it is also time reversible and has a stationary distribution given by the π_x.

Solution: Note that the transition probability of the new Markov chain is

$$P_{x,y} = Q_{x,y} \min\left(\frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}}, 1\right).$$

P is aperiodic since it has a positive self-loop probability. It is irreducible if Q is irreducible. Hence it must have a unique stationary distribution. We just need to check that π_x satisfies the time-reversibility condition $\pi_x P_{x,y} = \pi_y P_{y,x}$ for any x, y.

$$\pi_x P_{x,y} = \pi_x Q_{x,y} \min\left(\frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}}, 1\right) = \pi_y Q_{y,x} \quad \text{if } \pi_y Q_{y,x} < \pi_x Q_{x,y}$$

$$\text{else} \quad \pi_y P_{y,x} = \pi_y Q_{y,x} \min\left(\frac{\pi_x Q_{x,y}}{\pi_y Q_{y,x}}, 1\right) = \pi_x Q_{x,y}$$

Since π_x satisfies time-reversibility, it must be the stationary distribution of P.

Problem 4. *(10 points)* In this problem we will use a different fingerprinting technique to solve the pattern matching problem. The idea is to map any bit string s into a 2×2 matrix $\mathbf{M}(s)$ as follows:

- For the empty string ϵ, $\mathbf{M}(\epsilon) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
- $\mathbf{M}(0) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.
- $\mathbf{M}(1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.
- For non-empty strings x and y, $\mathbf{M}(xy) = \mathbf{M}(x) \times \mathbf{M}(y)$.

Show that this fingerprint function has the following properties.

1. $\mathbf{M}(x)$ is well-defined for all $x \in \{0,1\}^*$.

   Solution: Well defined means that for a string s, the fingerprint $\mathbf{M}(s)$ is unique. Suppose s is a string of length n made up of the bits $s = s_0 s_1 s_2 s_3 \ldots s_n$. We prove by induction that

   $$\mathbf{M}(s) = \mathbf{M}(s_0)\mathbf{M}(s_1)...\mathbf{M}(s_n)$$

   (or $\mathbf{M}(\epsilon)$ if $n = 0$). This is trivially true for $n = 0$. For $n = k$, assume the above formula holds for $n < k$. Then by the definition, for any split into smaller strings $s = s_a s_b$,

   $$\mathbf{M}(s) = \mathbf{M}(s_a)\mathbf{M}(s_b) = \mathbf{M}(s_0)\mathbf{M}(s_1)...\mathbf{M}(s_k)$$

   using the inductive assumption and the fact that matrix multiplication is associative, so it doesn't matter where the split is.

2. $\mathbf{M}(x) = \mathbf{M}(y) \Rightarrow x = y$.

   Solution: Here consider the top two elements of the fingerprint matrix for $s_0...s_{n-1}$. Call them a, b. Then the fingerprint matrix of $s_0 \ldots s_{n-1}$"1" will have top two elements a, $a+b$, while the fingerprint matrix of $s_0 \ldots s_{n-1}$"0" will have top two elements $a+b$, b. Therefore by looking at which of the top two elements is larger we can determine the last bit of the string. Then note that $\mathbf{M}(s_0 \ldots s_{n-1}) = \mathbf{M}(s_0...s_n) \cdot \mathbf{M}(s_n)^{-1}$ which follows from part (1). Of course this requires that $\mathbf{M}(1)$ and $\mathbf{M}(0)$ were invertible, which they are. So we can go through the fingerprint $\mathbf{M}(A)$ and recover A, which is all we need.

3. For $x \in \{0,1\}^n$, the entries in $\mathbf{M}(x)$ are bounded by Fibonacci number, F_n. (Where the Fibonacci numbers are defined by the recurrence, $F_0 = F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$. You may have to use a slightly clever induction to prove this.)

   Solution: for convenience define $F_{-1} = 0$ (note this still satisfies the recurrence). Let our inductive hypothesis be that $\mathbf{M}(x^{(n)})$ (where $x^{(n)} \in \{0,1\}^n$ is bounded elementwise by at least one of the following matrices:

$$\begin{bmatrix} F_n & F_{n-1} \\ F_{n-1} & F_{n-2} \end{bmatrix}$$
$$\begin{bmatrix} F_{n-1} & F_n \\ F_{n-2} & F_{n-1} \end{bmatrix}$$
$$\begin{bmatrix} F_{n-2} & F_{n-1} \\ F_{n-1} & F_n \end{bmatrix}$$
$$\begin{bmatrix} F_{n-1} & F_{n-2} \\ F_n & F_{n-1} \end{bmatrix}$$

   This is true for $n = 1$. Suppose it's true for $n = k$. Then one of the following is true:

$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_k & F_{k-1} \\ F_{k-1} & F_{k-2} \end{bmatrix} \mathbf{M}(0) = \begin{bmatrix} F_{k+1} & F_{k-1} \\ F_k & F_{k-2} \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-1} & F_k \\ F_{k-2} & F_{k-1} \end{bmatrix} \mathbf{M}(0) = \begin{bmatrix} F_{k+1} & F_k \\ F_k & F_{k-1} \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-2} & F_{k-1} \\ F_{k-1} & F_k \end{bmatrix} \mathbf{M}(0) = \begin{bmatrix} F_k & F_{k-1} \\ F_{k+1} & F_k \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-1} & F_{k-2} \\ F_k & F_{k-1} \end{bmatrix} \mathbf{M}(0) = \begin{bmatrix} F_k & F_{k-2} \\ F_{k+1} & F_{k-1} \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_k & F_{k-1} \\ F_{k-1} & F_{k-2} \end{bmatrix} \mathbf{M}(1) = \begin{bmatrix} F_k & F_{k+1} \\ F_{k-1} & F_k \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-1} & F_k \\ F_{k-2} & F_{k-1} \end{bmatrix} \mathbf{M}(1) = \begin{bmatrix} F_{k-1} & F_{k+1} \\ F_{k-2} & F_k \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-2} & F_{k-1} \\ F_{k-1} & F_k \end{bmatrix} \mathbf{M}(1) = \begin{bmatrix} F_{k-2} & F_k \\ F_{k-1} & F_{k+1} \end{bmatrix}$$
$$\mathbf{M}(x^{(k+1)}) < \begin{bmatrix} F_{k-1} & F_{k-2} \\ F_k & F_{k-1} \end{bmatrix} \mathbf{M}(1) = \begin{bmatrix} F_{k-1} & F_k \\ F_k & F_{k+1} \end{bmatrix}$$

   which proves the induction hypothesis for all natural numbers.

By considering the matrices $\mathbf{M}(x)$ modulo a suitable prime p, show how you would perform efficient randomized pattern matching.

Solution: using a similar argument to fingerprinting a binary number, take the matrix mod p for p a random prime up to T. Then an error can only occur if all 4 elements in the difference matrix

$\mathbf{M}(x) - \mathbf{M}(y)$ are divisible by p, which occurs with probability at most $\pi(\ln F(n))/\pi(T))$, where $F(n)$ is the n^{th} Fibonacci number.