École Polytechnique Fédérale de Lausanne
Eisenbrand, Moldenhauer

October 7, 2011

Randomized Algorithms (Fall 2011)

Assignment 2

Due date: 10:00am, October 18, 2011 — 25 points

Problem 1 (Chernoff Bounds) — 3 points

Show that for independent random variables $X_i \in \{0,1\}$ with $\Pr(X_i = 1) = p_i$, $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$ and for $0 < \delta \leq 1$ we have

$$\Pr(X \leq (1-\delta)\mu) \leq e^{\frac{-\mu\delta^2}{2}}.$$

Use the first Chernoff bound, proved in the lecture.

Problem 2 (Random Variables) — 2 points

Suppose you have access to a source of random bits. How can you simulate a Bernoulli trial $X = \{0,1\}$ with $\Pr(X = 1) = q$ where q is a rational number?

Problem 3 (Prob. and Computing, Problem 4.5) — 3 points

We plan to conduct an opinion poll to find out the percentage of people that want its president impeached. Assume that every person eithers yes or no. If the actual fraction of people who wants the president impeached is p, we want to find an estimate X of p such that

$$\Pr(|X - p| \leq \epsilon p) > 1 - \delta,$$

for a given ϵ and δ with $0 < \epsilon, \delta < 1$.
We query N people chosen independently and uniformly at random and output the fraction who want the president impeached. How large should N be for our result to be a suitable estimator of p? Use Chernoff bounds and express N in terms of p, ϵ and δ.

Problem 4 (Prob. and Computing, Problem 4.6) — 3 points

(a) In an election with two candidates using paper ballots, each vote is independently misrecorded with probability $p = 0.02$. Use a Chernoff bound to give an upper bound on the probability that more than 4% of the votes are misrecorded in an election of $1,000,000$ ballots.

(b) Assume that a misrecorded ballot always counts as a vote for the other candidate. Suppose that candidate A received $510,000$ votes and that candidate B received $490,000$ votes. Use Chernoff bounds to upper bound the probability that candidate B wins the election owing to misrecorded ballots. Specifically, let X be the number of votes for candidate A that are misrecorded and let Y be the number of votes for candidate B that are misrecorded. Bound $\Pr((X > k) \cup (Y < \ell))$ for suitable choices of k and ℓ.

Problem 5 (Prob. and Computing, Problem 4.14) — 4 points

Let $X_1, \ldots, X_n$ be independent Poisson trails such that $\Pr(X_i = 1) = p_i$ and let $a_1, \ldots, a_n$ be real numbers in $[0,1]$. Let $X = \sum_{i=1}^{n} a_i X_i$ and $\mu = \mathbb{E}[X]$. Prove the following Chernoff bound for any $\delta > 0$

$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

Prove a similar bound for $\Pr(X \leq (1-\delta)\mu)$ for $0 < \delta < 1$.

Problem 6 (Prob. and Computing, Problem 4.25) 6 points

We design a randomized algorithm for the following packet routing problem. We are given a network that is an undirected connected graph G, where nodes represent processors and the edges between the nodes represent wires. We are also given a set of N packets to route. For each packet we are given a source node, a destination node, and the exact route (path in the graph) that the packet should take from the source to its destination. (We may assume that there are no loops in the path). In each time step, at most one packet can traverse an edge. A packet can wait at any node during any time step, and we assume unbounded queue sizes at each node.
A *schedule* for a set of packets specifies the timing for the movement of packets along their respective routes. That is, it specifies which packet should move and which should wait at each time step. Our goal is to produce a schedule for the packets that tries to minimize the total time and the maximum queue size needed to route all the packets to their destinations.

(a) The dilation d is the maximum distance traveled by any packet. The congestion c is the maximum number of packets that must traverse a single edge during the entire course of the routing. Argue that the time required for any schedule should be at least $\Omega(c+d)$.

(b) Consider the following unconstrained schedule, where many packets may traverse an edge during a single time step. Assign each packet an integral delay chosen randomly, independently, and uniformly from the interval $[1, \lceil \alpha c/\log(Nd) \rceil]$, where α is a constant. A packet that is assigned a delay of x waits in its source node for x time steps; then moves on to its final destination through its specified route without ever stopping. Give an upper bound on the probability that more than $O(\log(Nd))$ packets use a particular edge e at a particular time step t.

(c) Again using the unconstrained schedule of part (b), show that the probability that more than $O(\log(Nd))$ packets pass through any edge at any time step is at most $1/(Nd)$ for a sufficiently large α.

(d) Use the constrained schedule to devise a simple randomized algorithm that, with high probability, produces a schedule of length $O(c + d\log(Nd))$ using queues of size $O(\log(Nd))$ and following the constraint that at most one packet crosses an edge per time step.

Problem 7 (Randomized Rounding) 4 points

Consider a matrix $A \in \{0,1\}^{n\times n}$ and a vector $p \in \mathbb{R}^n$ with all entries from the interval $[0,1]$. We wish to find a vector $q \in \{0,1\}^n$ that minimizes $\|A(p-q)\|_\infty$. We can think of q as an integer approximation to p in the sense that Aq is close to Ap in every component.
Prove that there exists a q such that $\|A(p-q)\|_\infty = O(\sqrt{n\log n})$.

Hint: Use randomized rounding to obtain q. Show that $\Pr(\|A(p-q)\|_\infty \leq c\sqrt{n\log n}) > 0$. Consider first $\Pr(|A_i(p-q)| \geq c\sqrt{n\log n})$ for a fixed row i and then use the union bound.