École Polytechnique Fédérale de Lausanne
Eisenbrand, Moldenhauer

October 25, 2011

Randomized Algorithms (Fall 2011)

Solutions to Assignment 2

0 points

Note: The purpose of these notes is to give a sketch of one possible solution. We do not guarantee correctness, nor completeness. It is your task to find and report mistakes.

Solution to Problem 1:
We use the following Chernoff bound proved in class for $0 < \delta < 1$:

$$\Pr(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} .$$

We need now to show that

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \le e^{-\mu\delta^2/2} .$$

This is equivalent with showing that

$$-\delta - (1-\delta)\ln(1-\delta) + \delta^2/2 \le 0 .$$

Let $f(\delta) := -\delta - (1-\delta)\ln(1-\delta) + \delta^2/2$. We have $f'(\delta) = \ln(1-\delta) + \delta$ and $f''(\delta) = -1/(1-\delta) + 1$. Since $f'(0) = 0$ and $f''(\delta) < 0$ for $0 < \delta < 1$ we have $f'(\delta) \le 0$. Since $f(0) = 0$ we therefore know that $f(\delta) \le 0$ for $0 < \delta < 1$.
It remains to show the statement for $\delta = 1$. Then, we need to show

$$\Pr(X \le (1-\delta)\mu) = \Pr(X \le 0) = \Pr(X = 0) \le e^{\frac{-\mu}{2}}$$

Now, $X = 0$ only happens when all poisson trials $X_i = 0$. Thus,

$$\Pr(X = 0) = \prod_{i=1}^{n}(1-p_i) \le \prod_{i=1}^{n} e^{-p_i} = e^{-\sum_{i=1}^{n} p_i} = e^{-\mu} \le e^{-\mu/2}.$$

Solution to Problem 2 (due to Anastasios Kyrillidis):
Easy solution: write $q = \frac{a}{b}$ for integers a and b, use the procedure from the last assignment to generate a random integer r from the interval $[0, b-1]$ and return 1 if $r \le a-1$ and otherwise 0. This solution has the problem that it requires unbiased random bits. The following is a solution even if the source is biased with some fixed probability.
Again, write $q = \frac{a}{b}$. Create the following set:

$$S = \{000\ldots001, 000\ldots010, 000\ldots100, \ldots, 001\ldots000, 010\ldots000, 100\ldots000\},$$

where each element $s \in S$ has b bits thereof exactly one is 1 and the rest 0. Pick a elements of S (it does not matter which ones) to obtain $T \subseteq S$. Set $\bar{T} = S \setminus T$.
Now consider the stream of bits from the source. We sample b bits. If this sequence is in T, we return 1. If the sequence is in $\bar{T}$, we return 0. Otherwise, we sample again. Continue until success. Since all the elements of S have same probability of occurrence in the stream of bits of the source, this gives the wanted sample probability. Furthermore, even though this process is not very efficient, the expected number of needed bits is bounded (probability that the sample is neither in T nor $\bar{T}$ is strictly smaller than 1 by some ϵ).

Solution to Problem 3 (due to Mingfu Shao):
Let Y denote the random variable that counts the number of people that want the president to be impeached; $Y = \sum_{i=1}^{N} Y_i$ where N is the number of people queried. We have $\mathbb{E}[Y_i] = p$ and $\mathbb{E}[Y] = Np$.
Let $X = Y/N$ be the estimator for p. Then,

$$\begin{aligned}\Pr(|X - p| \leq \epsilon p) &= \Pr(|Y/N - p| \leq \epsilon p) = \Pr((1-\epsilon)Np \leq Y \leq (1+\epsilon)Np) \\ &= 1 - \Pr(Y \leq (1-\epsilon)Np) - \Pr((1+\epsilon)Np \leq Y) \; .\end{aligned}$$

Applying Chernoff bounds (the Y_i are all independent) we get

$$\begin{aligned}\Pr(Y \leq (1-\epsilon)Np) &\leq e^{\frac{-Np\epsilon^2}{2}} \\ \Pr(Y \geq (1+\epsilon)Np) &\leq e^{\frac{-Np\epsilon^2}{3}}\end{aligned}$$

which gives

$$\Pr(|X - p| \leq \epsilon p) \geq 1 - 2e^{\frac{-Np\epsilon^2}{3}} \; .$$

Our goal is to lower bound this probability by $1 - \delta$. Hence, we want

$$2e^{\frac{-Np\epsilon^2}{3}} < \delta$$

which gives

$$N > \frac{-3 \ln \frac{\delta}{2}}{p\epsilon^2} \; .$$

Thus, by setting $N = \left\lfloor \frac{-3\ln(\delta/2)}{p\epsilon^2} \right\rfloor + 1$ we obtain the desired property for X.

Solution to Problem 4 (due to Alexandre Duc):

(a) Let X_i be a random variable which is 1 if vote i is misrecorded and 0 else for $1 \leq i \leq N = 1\,000\,000$. Then, $\Pr(X_i = 1) = p$ for $1 \leq i \leq N$. Set $X = \sum_{i=0}^{N} X_i$. We have $\mu := \mathbb{E}[X] = Np = 20000$. Hence,

$$\Pr\left(X \geq \frac{4}{100}N\right) = \Pr(X \geq 40000) = \Pr(X \geq (1+1)\mu) \leq \left(\frac{e^1}{2^2}\right)^{\mu} = \left(\frac{e}{4}\right)^{20000} .$$

(b) Let X be the number of votes for A that are misrecorded. Let Y be the number of votes for B that are correctly recorded. Candidate B wins the election if $Y + X > 500\,000$. Note that both X and Y are the sum of Poisson trials with $\Pr[X_i = 1] = 0.02$ for $X = \sum_{i=1}^{510\,000} X_i$ and $\Pr[Y_j = 1] = 0.98$ for $Y = \sum_{j=1}^{490\,000} Y_j$. Note also that all these trials are independent. We can, thus, apply the Chernoff bound on $X + Y$.

$$\mathbb{E}[X+Y] = 510\,000p + 490\,000(1-p) = 490\,400 \; .$$

Then,

$$\begin{aligned}\Pr(X + Y > 500\,000) &= \Pr\left(X + Y > (1 + \frac{9600}{490400})490400\right) \\ &= \Pr\left(X + Y > (1 + \frac{12}{613})\mathbb{E}[X+Y]\right) \; .\end{aligned}$$

By the Chernoff bound seen in class, we get

$$\Pr(X+Y > 500\,000) \le \exp\left\{-\frac{\mathbb{E}[X+Y]\left(\frac{12}{613}\right)^2}{3}\right\} = \exp\left\{-\frac{38\,400}{613}\right\}.$$

Solution to Problem 5:
The proof is done as in the lecture. Let $Y_i = a_i X_i$. The only difference is when finding a bound on $\mathbb{E}[e^{tY_i}]$:

$$\mathbb{E}[e^{tY_i}] = p_i e^{ta_i} + (1-p_i) = 1 + p_i(e^{ta_i} - 1) \le 1 + a_i p_i (e^t - 1) \le e^{a_i p_i (e^t - 1)}.$$

The first inequality follows by considering $f(a) = a(e^t - 1) - e^{ta} + 1$ and showing that $f(a) \ge 0$ for $0 \le a \le 1$. The proof of the other weighted Chernoff inequality is analogous.

Solution to Problem 6 (due to Alexandre Duc):

(a) By definition of d, there is at least one packet which needs to travel a distance of d. This packet will take at least d steps to reach its destination. Hence, the schedule is $\Omega(d)$. Similarly, there is at least one edge in G which is traversed by c packets. Since only one packet can traverse this edge at the same time, it takes at least c steps for all the packets to traverse this edge. Hence, the schedule is $\Omega(c)$. Thus, the schedule is $\Omega(d+c)$.

(b) Fix an edge e and a time step t. Consider all the packets that traverse e. The probability that one specific such packet traverses e at time t is $\frac{1}{\lceil \alpha c/\log(Nd) \rceil}$. Denote this by p. The probability that some fixed $\log(Nd)$ packets traverse e at time t is thus $p^{\log(Nd)}$. Since at most c packets traverse through e, the probability that (any) $\log(Nd)$ packets traverse e at time t is at most

$$\binom{c}{\log(Nd)} p^{\log(Nd)}.$$

(c) We can now set α sufficiently large. For easier calculations we omit the $\lceil\rceil$. Then,

$$\binom{c}{\log(Nd)} \left(\frac{\log(Nd)}{\alpha c}\right)^{\log(Nd)} \le \left(\frac{ec}{\log(Nd)}\right)^{\log(Nd)} \left(\frac{\log(Nd)}{\alpha c}\right) = \left(\frac{e}{\alpha}\right)^{\log Nd},$$

where the inequality follows due to Stirling's inequality ($n! > n^n e^{-n}$).
Now, we would like to apply the union bound over all time steps and all edges. There are at most Nd edges that are getting traversed by some packet. There are at most $d + \frac{\alpha c}{\log(Nd)}$ time steps in which packets can travel. Hence, the probability, that any edge at any time step is traversed by more than $\log(Nd)$ packets is at most

$$\left(d + \frac{\alpha c}{\log(Nd)}\right) \; Nd \; \left(\frac{e}{\alpha}\right)^{\log(Nd)}.$$

Set $\alpha = 8e$ to obtain

$$\left(d + \frac{8ec}{\log(Nd)}\right) Nd \frac{1}{(Nd)^3} < \frac{2}{Nd}$$

since $c \le N$ and N sufficiently large.

(d) We convert the unconstrained schedule of the previous question into a constrained schedule by simply extending each time step into $\log(Nd)$ time steps. Since we proved that with high probability we do not have more than $\log(Nd)$ packets traversing one edge, with this "extended time" we can let each packet take one arbitrary extended slot. The queue sizes are of size $O(\log(Nd))$ since at most $O(\log(Nd))$ packets cross an edge at the same time in the unconstrained schedule (with high probability). We showed before that the maximal length of the unconstrained schedule is $\frac{\alpha c}{\log(Nd)}+d$. Thus, the maximal length of the new schedule is

$$\left(\frac{\alpha c}{\log(Nd)} + d\right) \log(Nd) = \alpha c + d\log(Nd) \in O(c + d\log(Nd)) \; .$$

Solution to Problem 7 (due to Alexandre Duc):
We will use the following Chernoff type bound (Probability and Computing, Exercise 4.15): Let $X_1, \ldots, X_n$ be independent random variables such that

$$\Pr(X_i = 1 - p) = p_i \qquad \text{and} \qquad \Pr(X_i = -p_i) = 1 - p_i \; .$$

Let $X = \sum_{i=1}^{n} X_i$. Then,

$$\Pr(|X| \geq a) \leq 2e^{-2a^2/n} \; .$$

To show that there exists a q such that $\|A(p-q)\|_\infty = O(\sqrt{n\log n})$ we use the probablistic method, i.e., we devise a randomized method to generate a q and show that the probability $\Pr(\|\ldots\| = O(\ldots)) > 0$.
Let p_i (resp. q_i) denote the ith component of p (resp. q). Set each q_i to 1 with probability p_i. Now, consider any row i of A. It is easy to see that $\mathbb{E}[A_i(p-q)] = 0$. We have

$$\Pr\left(|A_i(p-q)| \geq c\sqrt{n\log n}\right) = \Pr\left(|A_{i,1}(q_1 - p_1) + \cdots + A_{i,n}(q_n - p_n)| \geq c\sqrt{n\log n}\right) \; .$$

Let k_i be the number of non-zero elements (1s) in A_i. We can consider the terms $(q_j - p_j)$ for all non-zero $A_{i,j}$ as k_i independent variables that obey the above conditions. Hence, we can apply the above bound to obtain

$$\Pr\left(|A_i(q-p)| \geq c\sqrt{n\log n}\right) \leq 2e^{-2c^2 n\log n/k_i} \leq 2e^{-2c^2 n\log n/n} = 2e^{-2c^2\log n} \leq \frac{2}{n^2}$$

for an appropriate choice of c. By the union bound we finally get

$$\Pr\left(\|A(p-q)\|_\infty \geq c\sqrt{n\log n}\right) \leq \frac{2}{n} \; .$$

which gives us the wanted result

$$\Pr\left(\|A(p-q)\|_\infty < c\sqrt{n\log n}\right) > 0 \; .$$