CS 6550: Randomized Algorithms February 8, 2019

Problem Set 2

Instructor: Eric Vigoda *STUDENT NAME*

Problem 1 Impeachment

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

$$\Pr\left(|X - p| \le \varepsilon p\right) \ge 1 - \delta$$

for a given ε and δ with $0 < \varepsilon, \delta < 1$.

We query N people chosen independently and uniformly at random from the community (with replacement) and output the fraction of them who want the president impeached. How large should N be for our result to be a suitable estimate of p? Use Chernoff bounds, and express N in terms of p, ε, and δ.

So, say that each person i, defines a random variable Y_i which is defined as follows: $Y_i = 1$ with probability p and $Y_i = 0$ with probability $1-p$. So, here p is the actual fraction of people who want the president impeached. So, this is saying that the ith person wants him impeached with probability p. Now, for a sample of size N, say I (so, I is a subset of N indices of all the people),we define $Y = \sum_{i\in I} Y_i$. Now, Chernoff says that $P(Y - E[Y] \ge \epsilon E[Y]) \le e^{\frac{-\epsilon^2}{2+\epsilon}E[Y]}$ and $P(Y - E[Y] \le -\epsilon E[Y]) \le e^{\frac{-\epsilon^2 E[Y]}{2}}$. So, by the union bound, $P(Y - E[Y] \ge \epsilon E[Y] \text{ OR } Y - E[Y] \le -\epsilon E[Y]) \le e^{\frac{-\epsilon^2}{2+\epsilon}E[Y]} + e^{\frac{-\epsilon^2 E[Y]}{2}}$. Namely, in terms of our problem $P(Y - pN \ge \epsilon pN \text{ OR } Y - pN \le -\epsilon pN) = P(|Y - pN| \ge \epsilon pN) \le e^{\frac{-\epsilon^2}{2+\epsilon}pN} + e^{\frac{-\epsilon^2 pN}{2}} \le 2e^{\frac{-\epsilon^2 pN}{3}}$. So, $P(|Y - pN| \le \epsilon pN) \ge 1 - 2e^{\frac{-\epsilon^2 pN}{3}}$. Then, note that $|Y - pN| \le \epsilon pN$ exactly when $|X - p| \le \epsilon p$. So, $P(|Y - pN| \le \epsilon pN) = P(|X - p| \le \epsilon p) \ge 1 - 2e^{\frac{-\epsilon^2 pN}{3}}$. So, if $\delta = 2e^{\frac{-\epsilon^2 pN}{3}}$, we see that the desired bound holds whenever $N \ge \frac{3ln(\frac{\delta}{2})}{-p\epsilon^2}$.

Problem 2 Median of Means

Suppose that we can obtain independent samples $X_1, X_2, \ldots$ of a random variable X and that we want to use these samples to estimate $\mathbb{E}[X]$. Given t independent samples, we use

$$\hat{X} = \frac{\sum_{i=1}^{t} X_i}{t}$$

for our estimate of $\mathbb{E}[X]$. Let ε and δ be given and $0 < \varepsilon, \delta < 1$. We want the estimate $\hat{X}$ to be within $\varepsilon\mathbb{E}[X]$ from the true value of $\mathbb{E}[X]$ with probability at least $1-\delta$; namely,

$$\Pr\left(\left|\hat{X} - \mathbb{E}[X]\right| \le \varepsilon\mathbb{E}[X]\right) \ge 1 - \delta.$$

We may not be able to use Chernoff's bound directly to bound how good our estimate $\hat{X}$ is if X is not a 0-1 random variable, and we do not know the moment generating function of X. We develop an alternative approach that requires only having a bound on the variance of X. Let

$$r = \frac{\sqrt{\mathrm{Var}(X)}}{\mathbb{E}[X]}.$$

(a) Show using Chebyshev's inequality that $O\left(\frac{r^2}{\varepsilon^2\delta}\right)$ samples are sufficient to solve the problem. Chebyshev says that $Pr(|\hat{X} - E[X]| > A) \le \frac{Var(X)}{A^2}$. So, $Pr(|\hat{X} - E[\hat{X}]| > \epsilon E[\hat{X}]) = Pr(|\hat{X} - E[X]| > \epsilon E[X]) = Pr(|t\hat{X} - tE[X]| \le t\epsilon E[X]) \le \frac{Var[X]}{t^2\epsilon^2 E[X]^2}$. Then, $Pr(|\hat{X} - E[X]| \le \epsilon E[X]) = Pr(|t\hat{X} - tE[X]| \le t\epsilon E[X]) \ge 1 - \frac{Var(X)}{t^2\epsilon^2 E[X]^2}$. So, let $\delta := \frac{Var(X)}{t^2\epsilon^2 E[X]^2}$. If $T = t^2 \ge \frac{Var[X]}{\delta\epsilon^2 E[X]^2}$, then the desired bound holds. So, t samples suffices, which means of course that $T = t^2$ (even more samples) suffice.

(b) Suppose that we need only a weak estimate $\hat{X}$ that is within $\varepsilon\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least 3/4. Argue that $O(r^2/\varepsilon^2)$ samples are enough for this weak estimate.

So, we plug in $\delta = \frac{1}{4}$. Whenever, $t^2 \ge \frac{4Var[X]}{\epsilon^2 E[X]^2}$. So, whenever we have $T \ge \frac{4Var[X]}{\epsilon^2 E[X]^2} = O(\frac{r^2}{\epsilon^2})$ samples, the desired bound certainly holds.

(c) Show that, by taking the median of $O(\log(1/\delta))$ independent weak estimates $\hat{X}$'s, we can obtain an estimate within $\varepsilon\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least $1-\delta$. Conclude that we need only $O\left(\frac{r^2\log(1/\delta)}{\varepsilon^2}\right)$ samples.

So, say we use N weak estimates, then take the median. What is the probability that the median fails to be in the desired range? This happens if $\ge \lceil\frac{N}{2}\rceil + 1$ of these weak estimates fall below $(1-\epsilon)E[X]$. This also happens if $\ge \lceil\frac{N}{2}\rceil + 1$ of these weak estimates fall above $(1-\epsilon)E[X]$. So,

$$P(FAILURE) \le P(\#\{\hat{X}_i | \hat{X}_i < (1-\epsilon)E[X]\} \ge \lceil\frac{N}{2}\rceil + 1) + P(\#\{\hat{X}_i | \hat{X}_i > (1+\epsilon)E[X]\} \ge \lceil\frac{N}{2}\rceil + 1) \tag{1}$$

$$\le \prod_{i=1}^{\lceil\frac{N}{2}\rceil+1} P(\hat{X}_i < (1-\epsilon)E[X]) + \prod_{i=1}^{\lceil\frac{N}{2}\rceil+1} P(\hat{X}_i > (1+\epsilon)E[X]) \tag{2}$$

$$\le \prod_{i=1}^{\lceil\frac{N}{2}\rceil+1} \frac{1}{4} + \prod_{i=1}^{\lceil\frac{N}{2}\rceil+1} \frac{1}{4} \tag{3}$$

$$\le 2(\frac{1}{4})^{\frac{N}{2}+2} \tag{4}$$

So, setting $\delta := 2(\frac{1}{4})^{\frac{N}{2}+2} = (\frac{1}{2^{N+3}})$ gives us that $ln(\delta) = -N - 3$ so that $N = -ln(\delta) - 3 = ln(\frac{1}{\delta}) - 3 = O(ln(\frac{1}{\delta}))$ samples suffice. Thus, we only need $O(\frac{r^2 log(1/\delta)}{\epsilon^2})$ samples because we take $O(r^2/\epsilon^2)$ samples $O(ln(1/\delta))$ times. Each set of samples gives us a mean. So we have $O(ln(1/\delta))$ sample means. Then, we take the median.

Problem 3 Geometric Distribution

A random variable X has geometric distribution if X takes value from $\mathbb{N}^+$ and has probability density

$$\Pr(X = k) = (1-p)^{k-1}p, \qquad \forall\, k \in \mathbb{N}^+$$

where p is the parameter of the distribution and $0 < p < 1$.

(a) Suppose we have a fair coin. Let X be the number of tosses till you get a HEAD for the first time. Prove that X has geometric distribution with parameter $p = 1/2$.
Say it takes $X = k$ tosses until I get a head for the first time. This happens exactly when the first $k-1$ tosses were tails and the kth toss is a head. This happens with probability $\frac{1}{2}^{p-1}\frac{1}{2} = (1-\frac{1}{2})^{k-1}(\frac{1}{2})$.

(b) Consider a collection $X_1, \dots, X_n$ of n independent geometrically distributed random variables with parameter $p = 1/2$. Let $X = \sum_{i=1}^{n} X_i$ and $\delta > 0$. Derive an upper bound on

$$\Pr(X \geq 2(1+\delta)n)$$

by applying the Chernoff bound to a sequence of $2(1+\delta)n$ fair coin tosses. (You may assume that $2(1+\delta)n$ is an integer.)
Apparently, Chernoff says $P(X \geq (1+\epsilon)E[X]) \leq (\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}})^{E[X]}$ for any $\epsilon > 0$ (`https://en.wikipedia.org/wiki/Chernoff_bound`). Now, let $\epsilon = (3+4\delta)$. Then, $P(X \geq (1+\epsilon)E[X]) = P(X \geq (1+\epsilon)\frac{n}{2}) = P(X \geq (1+(3+4\delta))\frac{n}{2}) = P(X \geq 2(1+\delta)n) \leq (\frac{e^{(3+4\delta)}}{(4+4\delta)^{(4+4\delta)}})^{\frac{n}{2}}$.

Problem 4 Pairwise Independence

A fair coin is flipped n times. Let X_{ij} with $1 \leq i < j \leq n$ be 1 if the ith and jth flip landed on the same side; let $X_{ij} = 0$ otherwise. Show that the X_{ij}'s are pairwise independent but not mutually independent.

The variables $X_{i_1 j_1}$ and $X_{i_2 j_2}$ are pairwise independent by definition if and only if $P(X_{i_1 j_1}) = P(X_{i_1 j_j} | X_{i_2 j_2})$ or equivalently if $P(X_{i_1 j_1} = a \text{ AND } X_{i_2 j_2} = b) = P(X_{i_1 j_1} = a) P(X_{i_2 j_2} = b)$. We consider 2 cases: either $\{i_1, j_1\} \cap \{i_2, j_2\} = \emptyset$ or $|\{i_1, j_1\} \cap \{i_2, j_2\}| = 1$ (the case in which $\{i_1, j_1\} = \{i_2, j_2\}$ is degenerate). So, if $\{i_1, j_1\} \cap \{i_2, j_2\} = \emptyset$, we calculate $P(X_{i_1 j_1} = 1) = P(X_{i_1} = 1 \text{ AND } X_{j_1} = 1) + P(X_{i_1} = 0 \text{ AND } X_{j_1} = 0) = P(X_{i_1} = 1) * P(X_{j_1} = 1) + P(X_{i_1} = 0) * P(X_{j_1} = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Also, $P(X_{i_1 j_1} = 0) = P(X_{i_1} = 0) * P(X_{j_1} = 1) + P(X_{i_1} = 1) * P(X_{j_1} = 0) = \frac{1}{2}$. Also, $P(X_{i_2 j_2} = 1) = P(X_{i_2} = 1 \text{ AND } X_{j_2} = 1) + P(X_{i_2} = 0 \text{ AND } X_{j_2} = 0) = P(X_{i_2} = 1) * P(X_{j_2} = 1) + P(X_{i_2} = 0) * P(X_{j_2} = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Also, $P(X_{i_2 j_2} = 0) = P(X_{i_2} = 0) * P(X_{j_2} = 1) + P(X_{i_2} = 1) * P(X_{j_2} = 0) = \frac{1}{2}$. Now, we calculate $P(X_{i_1 j_1} = 0 \text{ AND } X_{i_2 j_2} = 0) = P(X_{i_1} = 0) * P(X_{j_1} = 1) * (P(X_{i_2} = 0) * P(X_{j_2} = 1) + P(X_{i_2} = 0) * P(X_{j_2} = 1)) + P(X_{i_1} = 1) * P(X_{j_1} = 0) * (P(X_{i_2} = 0) * P(X_{j_2} = 1) + P(X_{i_2} = 0) * P(X_{j_2} = 1)) = \frac{1}{4}(\frac{1}{4} + \frac{1}{4}) + \frac{1}{4}(\frac{1}{4} + \frac{1}{4}) = \frac{1}{4}\frac{1}{2} * 2 = \frac{1}{4} = \frac{1}{2} * \frac{1}{2} = P(X_{i_1 j_1} = 0) * P(X_{i_2 j_2} = 0)$. The same goes for other values of a,b in the calculation of $P(X_{i_1 j_1} = a \text{ AND } X_{i_2 j_2} = b)$. So, we are done with the case in which $\{i_1, j_1\} \cap \{i_2, j_2\} = \emptyset$. Now, say that $|\{i_1, j_1\} \cap \{i_2, j_2\}| = 1$ and without loss of generality say $i_1 = i_2$. So, we wish to calculate $P(X_{i j_1} = 0 \text{ AND } X_{i j_2} = 0) = P(X_i = 0) * P(X_{j_1} = X_{j_2} = 1) + P(X_i = 1) * P(X_{j_1} = X_{j_2} = 0) = \frac{1}{2}(P(X_{j_1} = 1) * P(X_{j_2} = 1)) + \frac{1}{2}(P(X_{j_1} = 0) * P(X_{j_2} = 0)) = \frac{1}{2}(\frac{1}{4}) + \frac{1}{2}(\frac{1}{4}) = \frac{1}{4} = \frac{1}{2} * \frac{1}{2} = P(X_{i j_1} = 0) * P(X_{i j_2} = 0)$. The same calculation can be done for other $a, b \neq 0, 0$ So, this pair of random variables is independent.

Now, we show that this set of random variables is not mutually independent. Namely, take $X_{1,2}, X_{1,3}$ and $X_{2,3}$. I claim that $P(X_{1,3} = 1) \neq P(X_{1,3} = 1 | X_{1,2} = 1, X_{2,3} = 1)$. In particular, $P(X_{1,3} = 1) = P(X_1 = 1) * P(X_3 = 1) + P(X_1 = 0) * P(X_3 = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. However, if ones knows that $X_{1,2} = 1$ and $X_{2,3} = 1$, then one knows that $X_1 = X_2$ and $X_2 = X_3$, which by the transitive property implies that $X_1 = X_3$. So, $P(X_{1,3} = 1 | X_{1,2} = 1, X_{2,3} = 1) = 1 \neq \frac{1}{2}$ and we see that in this case, these variables are not mutually independent.

Problem 5 k-wise Independence

(a) Let X and Y be numbers that are chosen independently and uniformly at random from $\{0, 1, \ldots, n\}$. Let Z be their sum modulo $n + 1$. Show that X, Y and Z are pairwise independent but not mutually independent.
We know that X and Y are independent since they are chosen independently. Then, we notice that $Z = X + Y$ implies that $Y = Z - X$ and that in particular the map ϕ_{-X} : $\{0, \ldots, n\} \to \{0, \ldots, n\}$ defined by $Z \mapsto Z - X$ is a bijection. We now wish to show that $P(Z|X) = P(Z)$. Say we consider $P(Z = b|X = a) = P(Y = (b - a) \bmod (n + 1))$. We then note that for any $c \in \{0, \ldots, n\}$ we have that $P(Y = c \bmod (n + 1)) = \frac{1}{n+1}$. So, $P(Z = b|X = a) = P(Y = (b-a) \bmod (n+1)) = \frac{1}{n+1}$. We then compare this value to the probability $P(Z = b) = \sum_{i=0}^{n} P(Z = b|X = i)P(X = i) = \sum_{i=0}^{n} (\frac{1}{n+1})(\frac{1}{n+1}) = (n + 1)(\frac{1}{n+1})^2 = \frac{1}{n+1}$ and we see that our probabilities $P(Z = b|X = a)$ and $P(Z = b)$ are equal. (The case for Z,Y is analogous). To see that these are not mutually independent we note that $P(Z = 3|X = 1, Y = 1) \neq P(Z = 3)$. Namely, $P(Z = 3|X = 1, Y = 1) = 0$ and $P(Z = 3) = \frac{1}{n+1}$.

(b) Extend this example to give a collection of random variables that are k-wise independent but not $(k + 1)$-wise independent.
Let $X_1, \ldots, X_k$ be numbers that are chosen independently and uniformly at random from $\{0, 1, \ldots, n\}$. Then, let X_{k+1} be their sum modulo $(n+1)$. The set of variables $\{X_1, X_2, \ldots, X_k, X_{k+1}\}$ is a set of random variables that is k-wise independent but not $(k+1)$-wise independent. This can be shown by the principle of deferred decisions (though the question did not ask us to prove our example). I guess I can though. Clearly, the set $\{X_1, X_2, \ldots, X_k\}$ is independent. I also show that the set $\{X_1, X_2, \ldots, X_{k+1}\} \setminus \{X_j\}$ is independent. By definition this set is independent if and only if $P(\bigwedge_{i \in \{1,\ldots,k+1\} \setminus \{j\}} X_i = a_i) = \prod_{i \in \{1,\ldots,k+1\} \setminus \{j\}} P(X_i = a_i)$. We compute the left hand side as $P(\bigwedge_{i \in \{1,\ldots,k+1\} \setminus \{j\}} X_i = a_i) = P(X_{k+1} = a_{k+1}|X_i = a_i \forall i \in \{1, \ldots, k\} \setminus \{j\}) * P(X_k = a_k|X_i = a_i \forall i \in \{1, \ldots, k - 1\} \setminus \{j\}) * \cdots * P(X_2 = a_2|X_1 = a_1) * P(X_1 = a_1)$. Next, we define a set of partial sums by $S_i = \sum_{r \in \{1,\ldots,i\} \setminus \{j\}} X_i$. Then, we note $P(\bigwedge_{i \in \{1,\ldots,k+1\} \setminus \{j\}} X_i = a_i) = P(S_k + X_j + X_{k+1} = a_{k+1}|X_i = a_i \forall i \in \{0, \ldots, k\} \setminus \{j\}) * P(X_k = a_k) * \cdots * P(X_2 = a_2) * P(X_1 = a_1)$. Note, that we have removed the conditions on all but the first term in the product because the set $\{X_1, \ldots, X_k\}$ is independent which means that conditional probabilities equal their unconditional equivalents. So, continuing on, we get $P(\bigwedge_{i \in \{1,\ldots,k+1\} \setminus \{j\}} X_i = a_i) = P(S_k + X_j = a_{k+1}|X_i = a_i \forall i \in \{1, \ldots, k\} \setminus \{j\}) * \prod_{i=1}^{k-1} (\frac{1}{n+1}) = (\frac{1}{n+1})^{k-1} P(S_k + X_j = a_{k+1}|X_i = a_i \forall i \in \{1, \ldots, k\} \setminus \{j\}) = (\frac{1}{n+1})^{k-1} P(S_k + X_j = a_{k+1}|X_i = a_i \forall i \in \{1, \ldots, k\} \setminus \{j\}) = (\frac{1}{n+1})^{k-1} P(X_j + \sum_{i \in \{1,\ldots,k\} \setminus \{j\}} a_i = a_{k+1} \bmod n + 1)$. Letting $c := \sum_{i \in \{1,\ldots,k\} \setminus \{j\}} a_i \bmod (n + 1)$, we get $P(\bigwedge_{i \in \{1,\ldots,k+1\} \setminus \{j\}} X_i = a_i) = (\frac{1}{n+1})^{k-1} P(X_j + c = a_{k+1} \bmod (n + 1)) = (\frac{1}{n+1})^k = \prod_{i \in \{1,\ldots,k+1\} \setminus \{j\}} P(X_i = a_i)$ (where $R_a := a_{k+1} - c - a \bmod (n + 1)$) and we are done. Finally note that $P(X_{k+1} = S) \neq P(X_{k+1} = S|X_i = a_i \forall i \in \{1, \ldots, k\})$ (where $S := \sum_{i=1}^{k} a_i$). Namely, $P(X_{k+1} = S) = \sum_{Y \in \{0,1\}^{(k-1)}} P(X_k = S - \sum_{l=1}^{(k-1)} Y_l \bmod (n + 1)) * P((X_1, \ldots, X_{k-1}) = Y) = \sum_{Y \in \{0,1\}^{(k-1)}} (\frac{1}{n+1}) * (\frac{1}{2})^{k-1} = (2^{(k-1)}) * (\frac{1}{n+1}) * (\frac{1}{2})^{(k-1)} = \frac{1}{n+1}$ (where Y_l is the lth component of Y). However, we see that $P(X_{k+1} = S|X_i = a_i \forall i \in \{1, \ldots, k\}) = 1$ if $S = \sum_{i=1}^{k} a_i$ and $= 0$ if $S \neq \sum_{i=1}^{k} a_i$. So, we see that $P(X_{k+1} = S) \neq P(X_{k+1} = S|X_i = a_i \forall i \in \{1, \ldots, k\})$ which means that this set is not $(k + 1)$-wise independent.