CSE523: Randomized Algorithms

Assignment 2 Solutions

Divyajeet Singh (2021529)

Solution 1.

We are tasked with designing a modified Morris counter that increments the counter X with probability $\frac{1}{(1+c)^X}$ for some constant $c > 1$. We first see what our estimator $\tilde{n}$ for the count n would be. The modified algorithm is given in Algorithm 1.

Algorithm 1 Modified Morris Counter

```
1: X ← 0
2: while events occur do
3:     X ← X + 1 with probability 1/(1+c)^X
4: end while
5: return ñ = ((1+c)^X − 1)/c
```

Our modified estimator is $\tilde{n} = \frac{1}{c}((1+c)^X - 1)$. We proceed with the analysis as done in class. Let X_n denote the counter value when we have seen n events. We compute $\mathbb{E}\left[(1+c)^{X_n}\right]$ and $\text{Var}\left[(1+c)^{X_n}\right]$.

Computation of $\mathbb{E}\left[(1+c)^{X_n}\right]$

Note that in any case, $\mathbb{E}\left[(1+c)^{X_0}\right] = 1$, since $X = X_0 = 0$ when no events have occured. Then, for any X_k, we derive the recurrence

$$\begin{aligned}
\mathbb{E}\left[(1+c)^{X_k}\right] &= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \mathbb{E}\left[(1+c)^{X_k} \mid X_{k-1} = j\right] \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \left((1+c)^j \left(1 - \frac{1}{(1+c)^j}\right) + (1+c)^{j+1} \frac{1}{(1+c)^j}\right) \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \left((1+c)^j - 1 + (1+c)\right) \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] (1+c)^j + c \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \\
&= \mathbb{E}\left[(1+c)^{X_{k-1}}\right] + c
\end{aligned} \tag{1}$$

Using (1) and the base case, we can derive the expectation of the counter value when we have seen n events

$$\begin{aligned}
\mathbb{E}\left[(1+c)^{X_n}\right] &= \mathbb{E}\left[(1+c)^{X_{n-1}}\right] + c \\
&= \mathbb{E}\left[(1+c)^{X_{n-2}}\right] + 2c \\
&\ \vdots \\
&= \mathbb{E}\left[(1+c)^{X_0}\right] + nc \\
&= 1 + nc
\end{aligned} \tag{2}$$

Computation of $\textsc{Var}\left[(1+c)^{X_n}\right]$

We derive the variance of the counter value when we have seen n events. We first find the second moment $\mathbb{E}\left[(1+c)^{2X_n}\right]$. We proceed in a similar fashion as above. Note that for $X_0 = 0$, we have $\mathbb{E}\left[(1+c)^{2X_0}\right] = 1$. Then, for any X_k,

$$\begin{aligned}
\mathbb{E}\left[(1+c)^{2X_k}\right] &= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \mathbb{E}\left[(1+c)^{2X_k} \mid X_{k-1} = j\right] \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \left((1+c)^{2j}\left(1 - \frac{1}{(1+c)^j}\right) + (1+c)^{2(j+1)} \frac{1}{(1+c)^j}\right) \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] \left((1+c)^j \left((1+c)^j - 1\right) + (1+c)^{j+2}\right) \\
&= \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] (1+c)^{2j} + ((1+c)^2 - 1) \sum_{j=0}^{\infty} \mathbb{P}\left[X_{k-1} = j\right] (1+c)^j \\
&= \mathbb{E}\left[(1+c)^{2X_{k-1}}\right] + c(c+2)\mathbb{E}\left[(1+c)^{X_{k-1}}\right] \\
&= \mathbb{E}\left[(1+c)^{2X_{k-1}}\right] + c(c+2)(1+(k-1)c) \quad \text{(by (1))}
\end{aligned} \tag{3}$$

Using (3) and the base case, we can derive the second moment of the counter value when we have seen n events

$$\begin{aligned}
\mathbb{E}\left[(1+c)^{2X_n}\right] &= \mathbb{E}\left[(1+c)^{2X_{n-1}}\right] + c(c+2)(1+(n-1)c) \\
&= \mathbb{E}\left[(1+c)^{2X_{n-2}}\right] + c(c+2)\left[(1+(n-2)c) + (1+(n-1)c)\right] \\
&\vdots \\
&= \mathbb{E}\left[(1+c)^{2X_0}\right] + c(c+2) \sum_{i=0}^{n-1} (1+ic) \\
&= 1 + c(c+2)\left[n + \frac{cn(n-1)}{2}\right]
\end{aligned} \tag{4}$$

Now that we have the expectation and second moment of the counter value, we can derive the variance

$$\begin{aligned}
\textsc{Var}\left[(1+c)^{X_n}\right] &= \mathbb{E}\left[(1+c)^{2X_n}\right] - \mathbb{E}\left[(1+c)^{X_n}\right]^2 \\
&= 1 + c(c+2)\left[\frac{2n + cn(n-1)}{2}\right] - (1+nc)^2 \\
&= 1 + \frac{c^2+2c}{2}(2n + cn^2 - cn) - 1 - 2nc - n^2c^2 \\
&= c^3 \frac{n(n-1)}{2} \quad \text{(simplification by expansion)}
\end{aligned} \tag{5}$$

Application of Chebyshev's Inequality

Now, we can simply apply Chebyshev's inequality to get the required bound on the error

$$\begin{aligned}
\mathbb{P}\left[|\tilde{n} - n| \geq \epsilon n\right] &= \mathbb{P}\left[\left|\frac{(1+c)^X - 1}{c} - n\right| \geq \epsilon n\right] \\
&= \mathbb{P}\left[\left|(1+c)^X - 1 - cn\right| \geq \epsilon cn\right] \\
&= \mathbb{P}\left[\left|(1+c)^X - \mathbb{E}\left[(1+c)^X\right]\right| \geq \epsilon cn\right] \\
&\leq \frac{\textsc{Var}\left[(1+c)^X\right]}{(\epsilon cn)^2} \\
&= \frac{c^3 n(n-1)}{2\epsilon^2 c^2 n^2} \approx \frac{c}{2\epsilon^2}
\end{aligned} \tag{6}$$

Discussion on the value of c

We want to find how small c should be so that the estimator $\tilde{n}$ satisfies $|\tilde{n} - n| \leq \epsilon n$ with probability at least 0.9. So, we have

$$\begin{aligned} &\mathbb{P}\left[|\tilde{n} - n| \leq \epsilon n\right] \geq 0.9 \\ \implies &\mathbb{P}\left[|\tilde{n} - n| \geq \epsilon n\right] \leq 0.1 \\ &\implies \frac{c}{2\epsilon^2} \leq 0.1 \quad \text{or} \quad c \leq 0.2\epsilon^2 \end{aligned} \tag{7}$$

Note: This actually suggests that c should be less than 1 as opposed to the given constraint $c > 1$, which makes sense, since a smaller c corresponds to *more* frequent increments of the counter, which decreases the error.

A bound for $S(n)$

We want to obtain an upper bound on $S(n)$, the space required by the algorithm to store the counter value. We see that

$$(1+c)^{\mathbb{E}[X_n]} \leq \mathbb{E}\left[(1+c)^{X_n}\right] = 1 + nc \quad \text{by Jensen's Inequality} \tag{8}$$

So we have

$$\begin{aligned} &(1+c)^{\mathbb{E}[X_n]} \leq 1 + nc \\ \implies &\log\left((1+c)^{\mathbb{E}[X_n]}\right) \leq \log(1+nc) \\ \implies &\mathbb{E}[X_n] \leq \frac{\log(1+nc)}{\log(1+c)} \leq \frac{1}{c}\log(1+nc) \quad \because c < 1 \end{aligned} \tag{9}$$

Though the base of the logarithm does not matter, we consider log base 2. This suggests that after n rounds, the expected value of the counter is $O(\log n)$, which means that the space required to store the counter value is $O(\log \log n)$ in expectation. If we consider c to be a variable, we require $O(\log \log (nc)^{\frac{1}{c}})$ space.

Solution 2.

We are given a set of independent integers $X_1, X_2, \ldots, X_n$, drawn uniformly from $\{0, 1, 2\}$. Let us find the expected value of the random variable X_i $(0 \leq i \leq n)$.

$$\mathbb{E}[X_i] = \sum_{x=0}^{2} x\ \mathbb{P}[X_i = x] = \sum_{x=0}^{2} x \cdot \frac{1}{3} = 0 + \frac{1}{3} + \frac{2}{3} = 1 \tag{10}$$

Given $X = \sum_{i=1}^{n} X_i$, we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} 1 = n = \mu \quad \text{(say)} \tag{11}$$

To derive a Chernoff bound, we find the moment generating functions of X_i $(0 \leq i \leq n)$ and X.

$$M_{X_i}(t) = \mathbb{E}\left[e^{tX_i}\right] = \sum_{x=0}^{2} e^{tx}\ \mathbb{P}[X_i = x] = \frac{1}{3}\left(1 + e^t + e^{2t}\right) \tag{12}$$

$$\begin{aligned} M_X(t) &= \prod_{i=1}^{n} M_{X_i}(t) \quad \text{(by independence)} \\ &= \prod_{i=1}^{n} \frac{1}{3}\left(1 + e^t + e^{2t}\right) = \left(\frac{1 + e^t + e^{2t}}{3}\right)^{\mu} \quad \because \mu = n \end{aligned} \tag{13}$$

We now apply the following inequality derived from Markov's inequality (covered in class)

$$\mathbb{P}[X \geq a] = \mathbb{P}\left[e^{tX} \geq e^{ta}\right] \leq \frac{M_X(t)}{e^{ta}} \quad (t > 0) \tag{14}$$

For our case, we get

$$\mathbb{P}[X \geq a] \leq \frac{M_X(t)}{e^{ta}} \leq \frac{\left(1 + e^t + e^{2t}\right)^\mu}{3^\mu e^{ta}} \tag{15}$$

Now, we set $a = (1+\delta)\mu$ (note $\mu = n$) and $t = \ln(1+\delta)$ to get the Chernoff bound

$$\begin{aligned}
\mathbb{P}[X \geq (1+\delta)\mu] &\leq \frac{\left(1 + e^{\ln(1+\delta)} + e^{2\ln(1+\delta)}\right)^\mu}{3^\mu e^{\mu(1+\delta)\ln(1+\delta)}} \\
&= \left(\frac{1 + (1+\delta) + (1+\delta)^2}{3(1+\delta)^{(1+\delta)}}\right)^\mu \\
&= \left(\frac{1 + \delta + \frac{\delta^2}{3}}{(1+\delta)^{(1+\delta)}}\right)^\mu \\
&\leq \left(\frac{e^{\delta + \frac{\delta^2}{3}}}{(1+\delta)^{(1+\delta)}}\right)^\mu \qquad \because 1 + x \leq e^x \ \forall x \in \mathbb{R}
\end{aligned} \tag{16}$$

Similarly, we use the following inequality for the lower tail

$$\begin{aligned}
\mathbb{P}[X \leq a] = \mathbb{P}\left[e^{tX} \leq e^{ta}\right] &\leq \frac{M_X(t)}{e^{ta}} \quad (t < 0) \\
&= \frac{\left(1 + e^t + e^{2t}\right)^\mu}{3^\mu e^{ta}}
\end{aligned} \tag{17}$$

Now, we set for $\delta < 1$, $a = (1-\delta)\mu$ and $t = \ln(1-\delta)$ to get the Chernoff bound

$$\begin{aligned}
\mathbb{P}[X \leq (1-\delta)\mu] &\leq \frac{\left(1 + e^{\ln(1-\delta)} + e^{2\ln(1-\delta)}\right)^\mu}{3^\mu e^{\mu(1-\delta)\ln(1-\delta)}} \\
&= \left(\frac{1 + (1-\delta) + (1-\delta)^2}{3(1-\delta)^{(1-\delta)}}\right)^\mu \\
&\leq \left(\frac{e^{-\delta + \frac{\delta^2}{3}}}{(1-\delta)^{(1-\delta)}}\right)^\mu \qquad \because 1 + x \leq e^x \ \forall x \in \mathbb{R}
\end{aligned} \tag{18}$$

(16) and (18) give us the required Chernoff bounds for the upper and lower tails of the X.

Solution 3.

We sample a subset $S \subseteq A$ from a set A of unknown size such that $|S| = n$ uniformly at random. We are interested in the expected number of elements in A that hash to a value at most t for a given $t < m$ where $h \sim \mathcal{H}$ is a hash function drawn uniformly at random from a pairwise independent family.

Expected number of elements in S that hash to at most t

Let X_i $(1 \leq i \leq n)$ be the indicator random variable that the i^{th} element hashes to a value at most t. We have

$$X = \sum_{i=1}^{n} X_i \tag{19}$$

$$\begin{aligned}
\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] &= \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mathbb{P}[h(a_i) \leq t] \\
&= \sum_{i=1}^{n} \mathbb{P}\left[\bigcup_{j=1}^{t} h(a_i) = j\right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{t} \mathbb{P}[h(a_i) = j] \quad \text{(disjoint events)} \\
&= \frac{nt}{m} \quad \text{(by universality)}
\end{aligned} \tag{20}$$

Using the fact that pairwise independence is a stonger condition, and implies universality. Thus, the expected number of elements in S that hash to a value at most t is $\frac{nt}{m}$.

Using the sampling strategy as an estimator for $|A|$

Let us say we sample some n elements from A. We derive an estimator $\tilde{A}$ of $|A|$ as follows

$$\tilde{A} = \frac{m}{t} X \tag{21}$$

By (20), we have

$$\mathbb{E}\left[\tilde{A}\right] = \mathbb{E}\left[\frac{m}{t} X\right] = \frac{m}{t}\mathbb{E}\left[X\right] = n \tag{22}$$

Next, we find the variance of $\tilde{A}$

$$\begin{aligned}
\textsc{Var}\left[\tilde{A}\right] = \textsc{Var}\left[\frac{m}{t} X\right] &= \left(\frac{m}{t}\right)^2 \textsc{Var}\left[X\right] \\
&= \left(\frac{m}{t}\right)^2 \textsc{Var}\left[\sum_{i=1}^{n} X_i\right] \\
&\leq \left(\frac{m}{t}\right)^2 \sum_{i=1}^{n} \textsc{Var}\left[X_i\right] \\
&= \left(\frac{m}{t}\right)^2 n \left(\frac{t}{m}\right)\left(1 - \frac{t}{m}\right) \\
&= \frac{n(1-t)}{t} \leq \frac{n}{t} \quad \text{(if you will)}
\end{aligned} \tag{23}$$

We can now bound the error of our estimator $\tilde{A}$ using Chebyshev's inequality

$$\begin{aligned}
\mathbb{P}\left[\left|\tilde{A} - \mathbb{E}\left[\tilde{A}\right]\right| \geq \epsilon n\right] &\leq \frac{\textsc{Var}\left[\tilde{A}\right]}{(\epsilon n)^2} \\
&\leq \frac{1-t}{\epsilon^2 n t}
\end{aligned} \tag{24}$$

For a given $\delta < 1$, we can find the minimum value of t such that

$$\begin{aligned}
&\mathbb{P}\left[\left|\tilde{A} - \mathbb{E}\left[\tilde{A}\right]\right| \geq \epsilon n\right] \leq \delta \\
&\implies \frac{1-t}{\epsilon^2 n t} \leq \delta \implies t \geq \frac{1}{1 + \epsilon^2 n \delta}
\end{aligned} \tag{25}$$

(25) suggests that for a given (ϵ, δ) pair, we can choose a $t \geq \frac{1}{1+\epsilon^2 n \delta}$ to ensure that our estimator $\tilde{A}$ is *good*.

Estimating $|A \cup B|$

Assuming $A, B \subseteq U$ for some universe U of elements, let us say we sample n_a elements from A and n_b elements from B. By the above arguments, we can derive estimators $\tilde{Z}_A$ and $\tilde{Z}_B$ for $|A|$ and $|B|$ as

$$\tilde{Z}_A = \frac{m}{t} X_a \quad \text{and} \quad \tilde{Z}_B = \frac{m}{t} X_b \tag{26}$$

where X_a and X_b are the number of elements in A and B that hash to a value at most t. We can now derive estimators for $|A \cup B|$ as follows

$$\tilde{Z}_{A \cup B} = \tilde{Z}_A + \tilde{Z}_B \tag{27}$$

We can now find the expectation and variance of $\tilde{Z}_{A \cup B}$ and apply Chebyshev's inequality to get a bound on the error of our estimator.

$$\mathbb{E}\left[\tilde{Z}_{A \cup B}\right] = \mathbb{E}\left[\tilde{Z}_A\right] + \mathbb{E}\left[\tilde{Z}_B\right] = n_a + n_b \tag{28}$$

$$\textsc{Var}\left[\tilde{Z}_{A \cup B}\right] = \textsc{Var}\left[\tilde{Z}_A\right] + \textsc{Var}\left[\tilde{Z}_B\right] \leq \frac{(n_a + n_b)(1-t)}{t} \tag{29}$$

We can now apply Chebyshev's inequality to get a bound on the error of our estimator $\tilde{Z}_{A\cup B}$.

$$\begin{aligned}
\mathbb{P}\left[\left|\tilde{Z}_{A\cup B} - \mathbb{E}\left[\tilde{Z}_{A\cup B}\right]\right| \geq \epsilon(n_a+n_b)\right] &\leq \frac{\text{VAR}\left[\tilde{Z}_{A\cup B}\right]}{(\epsilon(n_a+n_b))^2} \\
&\leq \frac{(1-t)}{t\epsilon^2(n_a+n_b)} \leq \delta \\
\implies t &\geq \frac{1}{1+\epsilon^2(n_a+n_b)\delta}
\end{aligned} \tag{30}$$

Note: I guess I do not still understand the question in a sense - if we already know n, we can always say that my estimate for $|A|$ is n.

Solution 4.

By definition, a hash function is called Universal if it is drawn from a family $h \sim \mathcal{H}$ such that

$$\forall\ x_1 \neq x_2 \in U,\ \mathbb{P}\left[h(x_1) = h(x_2)\right] \leq \frac{1}{m} \tag{31}$$

and strongly Universal if[1]

$$\forall\ x_1 \neq x_2 \in U,\ \mathbb{P}\left[h(x_1) = y_1 \cap h(x_2) = y_2\right] = \frac{1}{m^2} \quad \text{for some } y_1, y_2 \in [m] \tag{32}$$

We want to find an example of a hash function which is Universal, but not strongly Universal. Let us say $U = \mathbb{Z}$. Let m be the size of the hash table and p be a sufficiently large prime. Consider the following family $\mathcal{H}$ of hash functions $h : \mathbb{Z} \to [m]$

$$\mathcal{H} = \{h_a(x) = ax \bmod p \bmod m \mid a = 1, 2, \ldots, p-1\} \tag{33}$$

Consider any $h \sim \mathcal{H}$ chosen uniformly at random. Let us find the probability for any $x \in U, j \in [m]$ that

$$\begin{aligned}
\mathbb{P}\left[h(x) = j\right] &= \sum_{i=1}^{p-1} \mathbb{P}\left[h_i(x) = j\right] \mathbb{P}\left[h = h_i\right] \\
&= \frac{1}{p-1} \sum_{i=1}^{p-1} \mathbb{P}\left[h_i(x) = j\right] \quad h \text{ is chosen uniformly} \\
&= \frac{1}{p-1} \cdot (p-1) \cdot \left(\frac{p}{m} \cdot \frac{1}{p}\right) = \frac{1}{m}
\end{aligned} \tag{34}$$

since for any h_i, the probability that $h_i(x) = j$ is that the $\bmod p$ operation maps ax to one of the $\frac{p}{m}$ locations that are congruent to j mod m. We have, for any $x_1 \neq x_2 \in \mathbb{Z}$,

$$\begin{aligned}
\mathbb{P}\left[h_a(x_1) = h_a(x_2)\right] &= \sum_{i=0}^{m-1} \mathbb{P}\left[h_a(x_2) = i \mid h_a(x_1) = i\right] \mathbb{P}\left[h_a(x_1) = i\right] \\
&= \sum_{i=0}^{m-1} \left(\frac{p}{m} \cdot \frac{1}{p}\right) \frac{1}{m} = \frac{1}{m}
\end{aligned} \tag{35}$$

where we use the fact that collisions can only occur due to clashses in $\bmod m$. Since p is a large prime, it is impossible that $a(x_1 - x_2) \bmod p \equiv 0$ for $x_1 \neq x_2$. Thus, $\mathcal{H}$ is Universal. However, we have

$$\mathbb{P}\left[h_a(x_1) = y_1 \cap h_a(x_2) = y_2\right] = \mathbb{P}\left[h_a(x_2) = y_2 \mid h_a(x_1) = y_1\right] \mathbb{P}\left[h_a(x_1) = y_1\right] \tag{36}$$

So, we solve the following equation

$$\begin{aligned}
&ax_1 \bmod p \bmod m = y_1 \\
\implies &ax_1 \bmod p \equiv y_1 \bmod m \\
\implies &ax_1 \equiv y_1 \bmod m \bmod p \\
\implies &a \equiv x_1^{-1} y_1 \bmod m \bmod p
\end{aligned} \tag{37}$$

[1]**Note:** The notation $[m]$ here denotes the set $\{0, 1, \ldots, m-1\}$ as opposed to the usual $\{1, 2, \ldots, m\}$.

where z^{-1} denotes the modular inverse of z modulo p. Since a anyway lies between 1 and $p-1$, this suggests that x_1 and y_1 uniquely determine a. Similarly, there can only be one value of $1 \leq a < p$ for which $h_a(x_2) = y_2$. Thus,

$$\begin{aligned}\mathbb{P}\left[h_a(x_1) = y_1 \cap h_a(x_2) = y_2\right] &= \mathbb{P}\left[h_a(x_2) = y_2 \mid h_a(x_1) = y_1\right]\mathbb{P}\left[h_a(x_1) = y_1\right] \\ &= \frac{1}{p-1} \cdot \frac{1}{m} < \frac{1}{m^2} \quad \because p > m\end{aligned} \tag{38}$$

Thus, $\mathcal{H}$ is Universal, but not strongly Universal.

Solution 5.

Note: Used different variables than the ones given in the question.

We asume that answers given by different people are independent. Let us use the random variables X_i $(1 \leq i \leq n)$

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ person answers yes} \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

with $\mathbb{P}\left[X_i = 1\right] = p$. Let our estimate for the fraction p be

$$\tilde{p} = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{40}$$

where X is the number of people who want their president impeached and n is the total number of people surveyed. We first find the expectation and variance of $\tilde{p}$. We have

$$\mathbb{E}\left[\tilde{p}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n} p = p \tag{41}$$

$$\begin{aligned}\text{VAR}\left[\tilde{p}\right] = \text{VAR}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\text{VAR}\left[X_i\right] \quad \text{(by independence)} \\ &= \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}\end{aligned} \tag{42}$$

We want the estimator $\tilde{p}$ to satisfy $\mathbb{P}\left[|\tilde{p} - p| \leq \epsilon p\right] \geq 1 - \delta$. We apply Chebyshev's inequality to get the required bound on the error

$$\begin{aligned}\mathbb{P}\left[|\tilde{p} - p| \geq \epsilon p\right] &= \mathbb{P}\left[|\tilde{p} - \mathbb{E}\left[\tilde{p}\right]| \geq \epsilon p\right] \\ &\leq \frac{\text{VAR}\left[\tilde{p}\right]}{(\epsilon p)^2} \\ &= \frac{p(1-p)}{n(\epsilon p)^2} = \frac{1-p}{n\epsilon^2 p}\end{aligned} \tag{43}$$

We want

$$\begin{aligned}&\mathbb{P}\left[|\tilde{p} - p| \leq \epsilon p\right] \geq 1 - \delta \\ \implies &\mathbb{P}\left[|\tilde{p} - p| \geq \epsilon p\right] \leq \delta\end{aligned} \tag{44}$$

We can now find the minimum number of people n we need to survey to ensure that our estimator $\tilde{p}$ satisfies the required error bound. We can find n upto varying degrees of looseness as follows

$$\begin{aligned}&\frac{1-p}{n\epsilon^2 p} \leq \frac{e^{-p}}{n\epsilon^2 p} \leq \frac{1}{n\epsilon^2 p} \leq \delta \\ \implies &n \geq \frac{1}{\delta\epsilon^2 p} \geq \frac{e^{-p}}{\delta\epsilon^2 p} \geq \frac{1-p}{\delta\epsilon^2 p}\end{aligned} \tag{45}$$

The inequalities above present the minimum number of people n we must to ensure that our estimator $\tilde{p}$ satisfies the required error bound.

References

1. Lecture Notes on *Approximate Counting using Morris Counter*, CSE5223 (Winter 2024), Dr. Rajiv Raman