Collaboration Statement: Andrew Koulogeorge and Eric Richardson

Problem 2: Handling Both Under and Over Estimates.

Solution.
Suppose we use the same strategy as in Problem 1 and construct a new estimate $\hat{est}$ by taking the minimum across m independent samples $est_1, est_2, ..., est_m$. In this scenario, there are two outcomes that result in us choosing a "bad" estimate: either all estimates were overestimates, or at least one estimate was an underestimate. Let X and Y denote these events, respectively. In particular, define

$$X = \bigwedge_{i=1}^{m} est_i \geq (1 + \varepsilon)N$$
$$Y = \bigvee_{i=1}^{m} est_i \leq (1 - \varepsilon)N$$

Since each sample is chosen independently, it follows that for any est_i, $\mathbf{Pr}[est_i \geq (1 + \varepsilon)N] \leq 0.9$. Therefore,

$$\mathbf{Pr}[X] = \prod_{i=1}^{m} \mathbf{Pr}[est_i \geq (1 + \varepsilon)N]$$
$$\leq \prod_{i=1}^{m} 0.9$$
$$= 0.9^m$$

Similarly, we have that $\mathbf{Pr}[est_i \leq (1 - \varepsilon)N] \leq 0.01$. We can apply union bound on the union of these events occurring:

$$\mathbf{Pr}[Y] = \sum_{i=1}^{m} \mathbf{Pr}[est_i \leq (1 - \varepsilon)N]$$
$$\leq \sum_{i=1}^{m} 0.01$$
$$= 0.01m$$

Let Z be the event that, using this strategy, that $\hat{est}$ is either an underestimate or an overestimate. Because X and Y are mutually exclusive events, it follows that $\mathbf{Pr}[Z] = \mathbf{Pr}[X \cup Y] = \mathbf{Pr}[X] + \mathbf{Pr}[Y]$. From the above, we have that $\mathbf{Pr}[Z] \leq 0.9^m + 0.01m$. Its important to note here what this inequality is telling us. We are bounding our algorithm's probability of

failure from above by a function of m that is larger than 1 when $m > 100$, which is a vacuous statement! Unlike in most other sampling applications we have seen, the probability of failure is not monotonically decreasing as we increase the number of samples. Instead, lets consider taking the minimum of this function, which corresponds to finding the number of samples m such that the bound on our our probability of failure is the tightest. Suppose $m = 20$. Then, we have

$$\mathbf{Pr}[Z] \leq 0.9^{20} + 0.01(20)$$
$$\approx 0.3216$$
$$< \frac{1}{3}$$

Thus, we can say that taking the minimum of 20 independent samples of est will result in a "bad" estimate $\hat{est}$ with probability less than $\frac{1}{3}$.

Lets now use $\hat{est}$ to find an even stronger estimate est^* by generating k independent estimates $\hat{est}_1 \dots \hat{est}_k$ and taking the median of those k estimates. We will follow the analysis laid out in class on boosting by DeepC closely. Note from the properties we derived from $\hat{est}_i$ that

$$\mathbf{Pr}[Z] = \mathbf{Pr}[\hat{est}_i \notin (1 \pm \varepsilon)N] < \frac{1}{3}$$

Lets define the two events A and B to be the events that $\hat{est}_i$ was an overestimate and underestimate respectively. Note that is if $\hat{est}_i$ is an underestimate, its surly not within our epsilon region of the mean. Same logic applies if $\hat{est}_i$ is an overestimate. From this, we can see then that:

$$\mathbf{Pr}[A] = \mathbf{Pr}[\hat{est}_i > (1 + \varepsilon)N]] \leq \mathbf{Pr}[Z] \leq \frac{1}{3}$$

$$\mathbf{Pr}[B] = \mathbf{Pr}[\hat{est}_i < (1 - \varepsilon)N]] \leq \mathbf{Pr}[Z] \leq \frac{1}{3}$$

Now, let us define two sets of Bernoulli random variables. Let W_i and Q_i be two random variable that takes the value 1 if $\hat{est}_i$ was an overestimate or $\hat{est}_i$ was an underestimate respectively. Let $W = \sum_{i=1}^{k} W_i$ and $Q = \sum_{i=1}^{k} Q_i$. By linearity of expectation, we see that $\mathbf{E}[W] = \mathbf{E}[Q] \leq \frac{k}{3}$. This tells us that the expected number of over or under estimates given k samples of $\hat{est}$ is at most $\frac{k}{3}$. Recall that we wish to ensure that est^* is within ε of the mean whp. Note that if our estimator est^* is an over/underestimate of the mean, there are is at least $\frac{k}{2}$ samples of $\hat{est}_i$ to its right/left. Thus, the probability of the estimator est^* being farther than ε from the mean is at least as large as W or Q being larger than $\frac{k}{2}$: (for what its worth, I am pretty sure this is an equality sign)

$$\mathbf{Pr}[\hat{est}_i > (1 + \varepsilon)N] \leq \mathbf{Pr}[W \geq \frac{k}{2}]]$$

2

$$\mathbf{Pr}[e\hat{s}t_i < (1-\varepsilon)N]] \leq \mathbf{Pr}[Q \geq \frac{k}{2}]$$

Since W and Q are sums of independent Bernoulli random variables, we can apply Chernoff to bound the probability of either of them from deviating too far from their mean.

$$\mathbf{Pr}[W \geq (1+\varepsilon)\mathbf{E}[W]] \leq \exp\frac{-\varepsilon^2\mathbf{E}[W]}{3}$$

$$\mathbf{Pr}[Q \geq (1+\varepsilon)\mathbf{E}[Q]] \leq \exp\frac{-\varepsilon^2\mathbf{E}[Q]}{3}$$

Let $\varepsilon = \frac{1}{2}$ and recall that the mean of W and Q is at most $\frac{T}{3}$

$$\mathbf{Pr}[W \geq \frac{T}{2}] \leq \exp\frac{-T}{36}$$

$$\mathbf{Pr}[Q \geq \frac{T}{2}] \leq \exp\frac{-T}{36}$$

$$\implies \mathbf{Pr}[A] = \mathbf{Pr}[e\hat{s}t_i > (1+\varepsilon)N] \leq \exp\frac{-T}{36}$$

$$\implies \mathbf{Pr}[B] = \mathbf{Pr}[e\hat{s}t_i < (1-\varepsilon)N] \leq \exp\frac{-T}{36}$$

$$\implies \exp\frac{-T}{36} < \frac{\delta}{2} \implies \frac{2}{\delta} < \exp\frac{T}{36} \implies T = O(\log\frac{1}{\delta})$$

Note that the event $est_i^* \notin (1\pm\varepsilon)N$ can be expressed as $\mathbf{Pr}[A\cup B]$. Since A and B are disjoint events (an estimate cant be both an over and an under estimate at the same time):

$$\mathbf{Pr}[est^* \notin (1\pm\varepsilon)N] = \mathbf{Pr}[A\cup B] = \mathbf{Pr}[A] + \mathbf{Pr}[B] < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

Thus, since we have shown that $\mathbf{Pr}[A]$ and $\mathbf{Pr}[B]$ are bound by δ when $T = O(\log\frac{1}{\delta})$, we have arrived at our (ε, δ) estimate. Note that we are taking $O(\log\frac{1}{\delta})$ number of samples of $e\hat{s}t$ while the question asks us to take $O(\log\frac{1}{\delta})$ number of samples of the original estimate est. Since we take a constant number of samples of est to construct our "min" estimator $e\hat{s}t$, this does not change our number of samples asymptotically. $\qquad\square$

break

For the sake of obtaining an upper bound, suppose $\mathbf{Pr}[W = 1] = \frac{1}{3}$. Then, $\mathbf{Exp}[W] = \frac{1}{3}$. Because the expected value of W is exactly equal to the statistic we wish to estimate, it follows that W is an unbiased estimator. Therefore, we can apply the Boosting/Median-of-Means theorem in order to find an (ε, δ)-estimator. Applying the theorem, we can achieve such an estimator by taking k independent samples of W, with

$$k = \frac{C\mathbf{Var}[W]}{\mathbf{Exp}[W]^2} \cdot \frac{1}{\varepsilon^2} \cdot \ln\left(\frac{2}{\delta}\right)$$

where C is some constant. Substituting our values, we have

$$k = \frac{C\frac{2}{9}}{\frac{1}{9}} \cdot \frac{1}{\varepsilon^2} \cdot \left(\ln\frac{2}{\delta} \right)$$

$$= \frac{2C}{\varepsilon^2} \cdot \ln\left(\frac{2}{\delta} \right)$$

$$= O\left(\ln\left(\frac{1}{\delta} \right) \right)$$

Note that each k requires 20 samples of our actual estimator. Since this is a constant factor, however, the asymptotic size of k does not change.

Problem 5: The better way of estimating bias of an unknown coin.

Solution. Note that the total number of tosses needed to get N heads can be expressed as the number of tosses between the $i - 1$th and ith head. Thus, we can compute $\mathbf{E}\left[Z\right] = \sum_{i=1}^{N} Z_i = \frac{N}{p^*}$ since each Z_i is a geometric random variable with probability p^* of succeeding $\implies \mathbf{E}\left[Z_i\right] = \frac{1}{p^*}$. $\qquad \square$

In order to prove that the probability that the number of tosses needed to get N heads concentrates "close" to $\frac{N}{p^*}$, we need to connect the events related to Z to events related to another random variable X and Y which we can represent as sums of Bernoulli random variables and thus bound their probability of deviating from the mean with Chernoff bounds. Note that its impossible to represent Z itself as a sum of a finite number of Bernoulli random variables since the image of Z is infinite.

Let A be the event that $Z < (1 - \varepsilon)\frac{N}{p^*}$ and let B be the event that $Z > (1 + \varepsilon)\frac{N}{p^*}$. Note that we wish to bound $\mathbb{P}\left[A \cup B\right] = \mathbb{P}\left[A\right] + \mathbb{P}\left[B\right]$ since A and B are disjoint events. Note that the union of these two events is identical to the event that Z does not lie within ε of its mean.

Let X be the number of heads in $\frac{(1 + \varepsilon)N}{p^*}$ tosses and let Y be the number of heads in $\frac{(1 - \varepsilon)N}{p^*}$ tosses and let Y be the number of heads in $\frac{(1 + \varepsilon)N}{p^*}$ tosses. We can see that $\mathbf{E}\left[X\right] = (1 + \varepsilon)N$ and $\mathbf{E}\left[Y\right] = (1 - \varepsilon)N$ since they are sums of Bernoulli random variables and we can again apply LoE.

Now for the crux of the argument; I claim that the probability it takes less than $(1 - \varepsilon)\dfrac{N}{p^*}$ tosses to see N is the same as the probability that we get at least N coins in $(1 - \varepsilon)\dfrac{N}{p^*}$ tosses. The same argument applies for the over estimate case: the probability it takes more than $(1 + \varepsilon)\dfrac{N}{p^*}$ tosses to see N is the same as the probability that we get less than N coins in $(1 + \varepsilon)\dfrac{N}{p^*}$ tosses.

$$\mathbf{Pr}[A] = \mathbf{Pr}[Z < (1 - \varepsilon)\frac{N}{p^*}] = \mathbf{Pr}[Y > N]$$

$$\mathbf{Pr}[B] = \mathbf{Pr}[Z > (1 + \varepsilon)\frac{N}{p^*}] = \mathbf{Pr}[X < N]$$

Since X and Y are both sums of independent Bernoulli random variables, we can apply Chernoff. Let $t_x = \dfrac{\varepsilon}{1 + \varepsilon}$ and $t_y = \dfrac{\varepsilon}{1 - \varepsilon}$:

$$\mathbf{Pr}[A] = \mathbf{Pr}[Y > N] = \mathbf{Pr}[Y > (1 + t_y)\mathbf{E}[Y]] \le \exp\frac{t_y^2 \mathbf{E}[X]}{3} = \exp\frac{-\varepsilon^2 N}{3(1 - \varepsilon)} < \exp\frac{-\varepsilon^2 N}{3}$$

$$\mathbf{Pr}[B] = \mathbf{Pr}[X < N] = \mathbf{Pr}[X < (1 - t_x)\mathbf{E}[X]] \le \exp\frac{t_x^2 \mathbf{E}[X]}{2} = \exp\frac{-\varepsilon^2 N}{2(1 + \varepsilon)} < \exp\frac{-\varepsilon^2 N}{3}$$

where both of the last inequalities follow from the fact that $\varepsilon \in (0, \frac{1}{2})$. Thus, we have shown that

$$\mathbf{Pr}[Z \notin (1 \pm \varepsilon)\mathbf{E}[Z]] = \mathbf{Pr}[A] + \mathbf{Pr}[B] < 2\exp\frac{-\varepsilon^2 N}{3}$$

$\square$

Qualitatively, we have now shown that if we flip our coin until we get more than $N = \dfrac{3}{\varepsilon^2}\log\dfrac{2}{\delta}$ heads, then the number of tosses it will take will concentrate around $\mathbf{E}[Z] = \dfrac{N}{p^*}$ w.h.p. Note that this holds $\forall\, \varepsilon \in (0, \frac{1}{2})$. In our algorithm, we are going to restrict $\forall\, \varepsilon \in (0, \frac{1}{4})$. Note that our result in part b still holds for this restricted range of values on ε. We can use this bound to now construct an estimator for p^* which will also concentrate around the true value of p^* w.h.p.

> **procedure** ESTIMATE $p^*(\varepsilon_0, \delta)$
>
>   Assert $\varepsilon_0 \in (0, \frac{1}{3})$
>
> $\varepsilon = \dfrac{\varepsilon_0}{1 + \varepsilon_0}$ (Always less than $\frac{1}{2} \implies$ bound holds)
>
>   Let $N > \dfrac{3}{\varepsilon^2}\log\dfrac{2}{\delta}$

By flipping our coin until we see greater than $N = \dfrac{3}{\varepsilon^2} \log \dfrac{2}{\delta}$ heads, we know:

$$\mathbf{Pr}[Z \in (1 \pm \varepsilon)\dfrac{N}{p^*}] \geq 1 - \delta$$

Lets define our estimate $\hat{p} = \dfrac{N}{Z}$. We know that with probability at least $1 - \delta$ (something large)

$$\implies Z \leq (1+\varepsilon)\dfrac{N}{p^*} \quad \& \quad Z \geq (1-\varepsilon)\dfrac{N}{p^*}$$

$$\implies p^* \leq (1+\varepsilon)\hat{p} \quad \& \quad p^* \geq (1-\varepsilon)\hat{p}$$

$$\implies \dfrac{1}{1+\varepsilon}p^* \leq \hat{p} \quad \& \quad \dfrac{1}{1-\varepsilon}p^* \geq \hat{p}$$

Note that the above inequalities hold $\forall \, \varepsilon \in (0, \dfrac{1}{2})$. Lets choose $\varepsilon \in (0, \dfrac{1}{4})$ and define $\varepsilon_0 = \dfrac{\varepsilon}{1-\varepsilon}$. NOTE that we have a degree of freedom over ε here because we proved a result about Z which holds true $\forall \, \varepsilon \in (0, \dfrac{1}{2})$ and so it also will hold true for a smaller subset of values for ε. With this clever construction of ε_0, we see that ε_0 lies between $(0, \dfrac{1}{3})$:

$$\varepsilon = 0 \implies \varepsilon_0 = \dfrac{\varepsilon}{1-\varepsilon} = 0$$

$$\varepsilon = \dfrac{1}{4} \implies \varepsilon_0 = \dfrac{\varepsilon}{1-\varepsilon} = \dfrac{1}{3}$$

Note that ε cant equal 0 or $\dfrac{1}{4}$ exactly but this analysis shows that $\varepsilon_0 \in (0, \dfrac{1}{3})$. Next, we see that when we restrict $\varepsilon \in (0, \dfrac{1}{4})$, we are able to bound the region where our estimate $\hat{p}$ will lie:

$$(1 - \varepsilon_0) \leq \dfrac{1}{1+\varepsilon}$$

$$\varepsilon = 0 \implies 1 \leq 1$$

$$\varepsilon = \dfrac{1}{4} \implies \dfrac{2}{3} \leq \dfrac{4}{5}$$

$$(1 + \varepsilon_0) \geq \dfrac{1}{1-\varepsilon}$$

$$\varepsilon = 0 \implies 1 \geq 1$$

$$\varepsilon = \frac{1}{4} \implies \frac{4}{3} \geq \frac{4}{3}$$

Since $p^* > 0$, we can multiply these inequalities and preserve the ordering:

$$(1 - \varepsilon_0)p^* \leq \frac{1}{1 + \varepsilon}p^* \leq \hat{p}$$

$$(1 + \varepsilon_0)p^* \geq \frac{1}{1 - \varepsilon}p^* \geq \hat{p}$$

Since the bounds on Z hold with probability $1 - \delta$, our estimate $\hat{p}$ falls in this range as well for any $\varepsilon_0 \in (0, \frac{1}{3})$

$$\implies (1 - \varepsilon_0)p^* \leq \hat{p} \leq (1 + \varepsilon_0)p^* \quad \square$$

Problem 7: A Question of Fairness.

Solution.
Let X_i be a Bernoulli random variable that takes the value 1 if the ith sample from U is selected and colored azure. Similarly, let Y_i be a Bernoulli random variable that takes the value 1 if the ith sample from U is selected and colored blue. We have

$$\mathbf{Exp}[X_i] = 1 \cdot \frac{s}{n} \cdot \frac{1}{2}$$
$$\mathbf{Exp}[Y_i] = 1 \cdot \frac{s}{n} \cdot \frac{1}{2}$$

Hence, $\mathbf{Exp}[X_i] = \mathbf{Exp}[Y_i] = \frac{s}{2n}$. Now, define X and Y to be the sum of all X_i and Y_i, respectively. We see that this summation results in the number of azure/blue elements in our sample R.

$$\mathbf{Exp}[X] = \mathbf{Exp}[Y] = \sum_{i=1}^{n} \frac{s}{2n}$$
$$= \frac{sn}{2n}$$
$$= \frac{s}{2}$$

Thus, we expect to see $\frac{s}{2}$ azure elements and $\frac{s}{2}$ blue elements in R. Now, define $Z = |X - Y|$. By Linearity of Expectation, $\mathbf{Exp}[Z] = 0$. Because X and Y are given by a summation of independent Bernoulli random variables, the Chernoff Bound tells us that the probability they diverge from their expectation by some factor ε decays exponentially in ε. Now, consider what needs to happen for Z to be greater than or equal to some

7

value t. In the general case, X must diverge from its expectation by at least $\dfrac{t}{q}$ for some $1 \leq q \leq t$, and Y must diverge (in the opposite direction as X) by $t - \dfrac{t}{q}$. Of course, this is not the only possible way for Z to be greater than t (for instance, $X = \mathbf{Exp}[X] + 2t$ and $Y = \mathbf{Exp}[Y] + t$), but Chernoff asserts that the probability of these events occurring are orders of magnitude smaller than the variations that may occur closer to the expectations of X and Y. In particular, the Chernoff Bound says that X and Y are heavily concentrated around their means. This implies that, due to the Chernoff-induced decay in probability as t increases, the case in which X and Y individually diverge from their means by the *smallest* amount (while still maintaining the property that $|X - Y| \geq t$) is exponentially more likely than any other case. In other words, this "minimal" event alone yields an asymptotically identical bound to the summation across all events that result in $|X - Y| \geq t$. This occurs precisely when $X \geq \mathbf{Exp}[X] + \dfrac{t}{2}$ and $Y \leq \mathbf{Exp}[Y] - \dfrac{t}{2}$, or vice versa.

$$\mathbf{Pr}[Z \geq t] \approx \mathbf{Pr}[X \geq \mathbf{Exp}[X] + \frac{t}{2} \text{ and } Y \leq \mathbf{Exp}[Y] - \frac{t}{2}]$$
$$= \mathbf{Pr}[X \geq \mathbf{Exp}[X] + \frac{t}{2}] \cdot \mathbf{Pr}[Y \leq \mathbf{Exp}[Y] - \frac{t}{2}] \quad \text{(since } X \text{ and } Y \text{ are independent)}$$

Suppose we fix $\varepsilon = \dfrac{t}{s}$. Then, substituting our values into the Chernoff Bound, we have

$$\mathbf{Pr}[X \geq (1 + \frac{t}{s})\frac{s}{2}] \cdot \mathbf{Pr}[Y \leq (1 - \frac{t}{s})\frac{s}{2}] \leq \exp -\frac{\left(\frac{t}{s}\right)^2 \frac{s}{2}}{3} \cdot \exp -\frac{\left(\frac{t}{s}\right)^2 \frac{s}{2}}{3}$$

$$\mathbf{Pr}[X \geq \frac{s}{2} + \frac{t}{2}] \cdot \mathbf{Pr}[Y \leq \frac{s}{2} - \frac{t}{2}] \leq \exp -\frac{2 \cdot \left(\frac{t}{s}\right)^2 \frac{s}{2}}{3}$$

$$= \exp -\frac{\frac{t^2}{s^2} \cdot s}{3}$$

$$= \exp -\frac{t^2}{3s}$$

Since we want this probability to be less than δ, we have

$$\exp -\frac{t^2}{3s} < \delta$$
$$-\frac{t^2}{3s} < \ln(\delta)$$
$$t^2 < -3s \cdot \ln(\delta)$$
$$t < \sqrt{3s \cdot \ln\left(\frac{1}{\delta}\right)}$$
$$t = O\left(\sqrt{s \cdot \ln\left(\frac{1}{\delta}\right)}\right)$$

Therefore, we have shown that $\mathbf{Pr}[Z \geq t]$ becomes smaller than δ when t is of size $O\left(\sqrt{s \cdot \ln\left(\frac{1}{\delta}\right)}\right)$.