1. Recall that a *priority search tree* is a binary tree in which every node has both a *search key* and a *priority*, arranged so that the tree is simultaneously a binary search tree for the keys and a min-heap for the priorities. A ***heater*** is a priority search tree in which the *priorities* are given by the user, and the *search keys* are distributed uniformly and independently at random in the real interval $[0, 1]$. Intuitively, a heater is a sort of anti-treap.

   The following problems consider an n-node heater T whose priorities are the integers from 1 to n. We identify nodes in T by their *priorities*; thus, "node 5" means the node in T with *priority* 5. For example, the min-heap property implies that node 1 is the root of T. Finally, let i and j be integers with $1 \le i < j \le n$.

   (a) What is the *exact* expected depth of node j in an n-node heater? Answering the following subproblems will help you:

   i. Prove that in a random permutation of the $(i+1)$-element set $\{1, 2, \dots, i, j\}$, elements i and j are adjacent with probability $2/(i+1)$.

   ii. Prove that node i is an ancestor of node j with probability $2/(i+1)$. *[Hint: Use the previous question!]*

   iii. What is the probability that node i is a *descendant* of node j? *[Hint: Do **not** use the previous question!]*

Solution: We follow the proposed outline:

i. Fix a permutation of the subset $\{1, 2, \dots, i\}$. There are exactly $i+1$ places to insert j into this permutation, exactly two of which are adjacent to i. Each possibility is equally likely. It follows that in a random permutation, i and j are adjacent with probability $2/(i+1)$.

ii. Recall from class that a node x is an ancestor of node y in a priority search tree if and only if, among all nodes with search keys between *key*(x) and *key*(y), node x has the smallest priority. Thus, node i is an ancestor of node j if and only if, when we sort the nodes by their search keys, nothing in the set $\{1, 2, \dots, i-1\}$ appears between node i and node j. Equivalently, in the permutation of nodes $\{1, 2, \dots, i, j\}$ induced by the search keys, nodes i and j are adjacent. It follows that i is an ancestor of j with probability $2/(i+1)$

iii. Node i cannot be a descendant of node j, because a heater is a min-heap. The probability is zero.

The depth of a node is equal to the number of proper ancestors. Thus, the expected depth of node j can be computed using the usual sum-of-indicators analysis, as follows:

$$\begin{aligned} \mathrm{E}[\#\text{proper ancestors of } j] &= \sum_{i=1}^{n} \Pr[i \text{ is a proper ancestor of } j] \\ &= \sum_{i=1}^{j-1} \frac{2}{i+1} \;=\; \mathbf{2H_j - 2} \;=\; \Theta(\log j) \end{aligned}$$

■

Rubric: 6 points = 2 for part i. + 2 for part ii. + 1 for part iii. + 1 for conclusion (½ for $\Theta(\log j)$; no points for $\Theta(\log n)$). These are not the only correct proofs for i. and ii.

(b) Describe and analyze an algorithm to insert a new item into an n-node heater.

Solution: The algorithm is identical to the algorithm for inserting into a treap. First, insert a new vertex with a random search key, using the textbook algorithm for inserting into a binary tree. Then assign this new node the desired priority and rotate it upward to fix the heap property.

The running time of the algorithm is proportional to the depth of the new node *before* its priority is assigned. If we set the new priority to ∞, this depth is unchanged, but the second phase of the algorithm would do nothing. The new node would have the largest priority in the heater, and so by part (a), its expected depth is $2H_{n+1} - 2 = O(\log n)$. We conclude that the expected running time of our insertion algorithm is $O(\log n)$. ■

Rubric: 2 points = 1 for algorithm + 1 for analysis. No analysis credit for just writing $O(\log n)$, or pointing to the lecture notes (which analyze *treaps*, not *heaters*).

(c) Describe and analyze an algorithm to delete the smallest priority (the root) from an n-node heater.

Solution: We essentially run the insertion algorithm backwards, just as we do for treaps: First rotate the node to be deleted downward until it becomes a leaf (implicitly increasing its priority to ∞), and then discard that leaf.

The running time is proportional to the depth of the former root just before we discard it. The analysis in part (a) implies that the expected depth of this leaf is $O(\log n)$. ■

Rubric: 2 points = 1 point for algorithm + 1 point for analysis.

2. Suppose we generate a bit-string w by flipping a fair coin n times. Thus, each bit in w is equal to `0` or `1` with equal probability, and the bits in w are fully independent. A ***run of length ℓ*** in w is a substring of length ℓ in which all bits are equal. For example, the string `0100011101` contains ***three*** runs of length 3, starting at the third, fourth, and seventh bits.

   (a) Suppose n is a power of 2. Show that the expected number of runs of length $\lg n + 1$ is $1 - o(1)$. (Here "lg" is standard shorthand for log-base-2.)

   Solution: Let X_i be an indicator random variable which takes value 1 if there is a run of $\lg n + 1$ starting from index $i \in \{1, \ldots, n\}$ and takes value 0 otherwise. Note that for any index $1 \le i \le n - \lg n$, we have that

$$\mathbb{E}[X_i] = \left(\frac{1}{2}\right)^{\log n} = \frac{1}{n},$$

   because if a run starts from the ith index, the next $\lg n$ bits must be the same as the ith one. On the other hand, $\mathbb{E}[X_i] = 0$ for $n - \lg n < i \le n$ since there can be no runs of length $\lg n + 1$ starting at these indices. Hence, the expected number of runs of length $\lg n + 1$ is

$$(n - \lg n) \cdot \frac{1}{n} = 1 - \frac{\lg n}{n} = 1 - o(1).$$

   ■

   Rubric: 4 points = 2 points for the correctly computing the probability of a single run + 2 points for computing the correct expected number of runs.

   (b) Show that, for sufficiently large n, the probability that *every* run in w has length less than $\lfloor \lg n - 2 \lg \lg n \rfloor$ is less than $1/n$. *[Hint: Break w into disjoint substrings of length $\lfloor \lg n - 2 \lg \lg n \rfloor$ and use the following fact: The event that all bits in one substring are equal is independent of the event that all bits in any other substring are equal.]*

   Solution: Let $k = \lfloor \lg n - 2 \lg \log n \rfloor$. The event that every run has length less than k is the same as the event that there is no run of length k. Let us call this event $\mathcal{A}$. Instead of bounding the probability of this event directly, we will consider another event $\mathcal{B}$ which contains $\mathcal{A}$. For this, we break the string into disjoint consecutive substrings of length k. There are $\lfloor n/k \rfloor$ such substrings. For a string of length n to not contain a run of length k (i.e. the event $\mathcal{A}$ occurs), it is necessary that none of the substrings contains a run of length k. Denoting the latter event by $\mathcal{B}$, we note that $\mathcal{A} \subseteq \mathcal{B}$, and thus $\mathbb{P}[\mathcal{A}] \le \mathbb{P}[\mathcal{B}]$.

   The probability that a single substring does not contain a run is $1 - (1/2)^{k-1}$. Since the substrings are disjoint and independent, the probability that none of them contains a run is

$$\mathbb{P}[\mathcal{B}] = \left(1 - \left(\frac{1}{2}\right)^{k-1}\right)^{\lfloor n/k \rfloor}$$

Now, $k \leq \lg n - 2\lg n \lg n - 1$ and for sufficiently large n, we have that $\lfloor n/k \rfloor \leq \frac{n}{\lg n} - 1$. Thus,

$$\mathbb{P}[\mathcal{B}] \leq \left(1 - \left(\frac{1}{2}\right)^{\lg n - 2\lg\lg n}\right)^{n/\lg n - 1} = \left(1 - \frac{\lg^2 n}{n}\right)^{n/\lg n - 1}.$$

Using the inequality $1 + x \leq e^x$ which holds for all $x \in \mathbb{R}$, we can bound the above by the following

$$\begin{aligned}\mathbb{P}[\mathcal{B}] &\leq \left(\exp\left(-\frac{\lg^2 n}{n}\right)\right)^{n/\lg n} \cdot \left(1 - \frac{\lg^2 n}{n}\right)^{-1} \\ &= \exp\left(-\frac{\lg^2 n}{n}\left(\frac{n}{\lg n}\right)\right) \cdot \left(1 - \frac{\lg^2 n}{n}\right)^{-1} \\ &= \exp(-\lg n) \cdot \left(1 - \frac{\lg^2 n}{n}\right)^{-1} \\ &= \frac{1}{n^{1/\ln 2}} \cdot \left(1 - \frac{\lg^2 n}{n}\right)^{-1}\end{aligned}$$

Finally, note that $(1-x)^{-1} \leq 1 + 2x$ for any $0 \leq x \leq 1/2$. Thus, for n large enough so that $\lg^2 n/n \leq 1/2$ holds, we have

$$\mathbb{P}[\mathcal{B}] \leq \frac{1}{n^{1/\ln 2}} \cdot \left(1 + \frac{2\lg^2 n}{n}\right)$$

Since $1/\ln 2 > 1.4$, it follows that the above probability is at most $1/n$ for large enough n. ■

Rubric: 6 points = 2 points for correctly defining the event in terms of disjoint substrings, 2 point for correctly computing the probability of the event and 2 points for showing that it is less than $1/n$ for large enough n. This is not the only way of analyzing these probabilities.

3. Suppose we are given a coin that may or may not be biased, and we would like to compute an accurate *estimate* of the probability of heads. Specifically, if the actual unknown probability of heads is p, we would like to compute an estimate $\tilde{p}$ such that

$$\Pr[|\tilde{p} - p| > \varepsilon] < \delta$$

where ε is a given ***accuracy*** or ***error*** parameter, and δ is a given ***confidence*** parameter.

The following algorithm is a natural first attempt; here FLIP( ) returns the result of an independent flip of the unknown coin.

```
MEANESTIMATE(ε):
  count ← 0
  for i ← 1 to N
      if FLIP() = HEADS
          count ← count + 1
  return count/N
```

(a) Let $\tilde{p}$ denote the estimate returned by $MeanEstimate(\varepsilon)$. Prove that $\mathrm{E}[\tilde{p}] = p$.

Solution: Let $X_i = 1$ if the ith flip is HEADS and $X_i = 0$ if the ith flip is TAILS. The final value of *count* is $X = \sum_i X_i$, so linearity of expectation implies

$$\mathrm{E}[X] = \sum_{i=1}^{N} \Pr[X_i = 1] = Np.$$

Finally, $\tilde{p} = X/N$, so linearity of expectation implies $\mathrm{E}[\tilde{p}] = \mathrm{E}[X]/N = p$, as required. ∎

Rubric: 3 points.

(b) Prove that if we set $N = \lceil \alpha/\varepsilon^2 \rceil$ for some appropriate constant α, then $\Pr[|\tilde{p} - p| > \varepsilon] < 1/4$. *[Hint: Use Chebyshev's inequality.]*

Solution: The coin flips are pairwise independent (in fact, *fully* independent) so we can apply Chebyshev's inequality. Let X be the final value of *count*, and recall from part (a) that $\mu = \mathrm{E}[X] = Np$.

$$\begin{aligned}
\Pr[|\tilde{p} - p| > \varepsilon] &= \Pr[|X - \mu| > N\varepsilon] \\
&= \Pr[(X - \mu)^2 > N^2\varepsilon^2] \\
&< \frac{\mu}{N^2\varepsilon^2} = \frac{p}{N\varepsilon^2} && \text{[Chebyshev]}
\end{aligned}$$

Setting $N = \lceil 4/\varepsilon^2 \rceil$ implies $\Pr[|\tilde{p} - p| > \varepsilon] < p/4 \le 1/4$. ∎

Rubric: 3 points. We can't apply the form of Chebyshev's inequality given in the notes to $\tilde{p}$ directly, because $\tilde{p}$ is not a sub of indicators.

(c) We can increase the previous estimator's confidence by running it multiple times, independently, and returning the *median* of the resulting estimates.

```
MedianOfMeansEstimate(δ, ε):
  for j ← 1 to K
      estimate[j] ← MeanEstimate(ε)
  return Median(estimate[1..K])
```

Let p^* denote the estimate returned by MEDIANOFMEANSESTIMATE(δ, ε). Prove that if we set $N = \lceil \alpha/\varepsilon^2 \rceil$ (inside MEANESTIMATE) and $K = \lceil \beta \ln(1/\delta) \rceil$, for some appropriate constants α and β, then $\Pr[|p^* - p| > \varepsilon] < \delta$. *[Hint: Use Chernoff bounds.]*

Solution: For each index j, define an indicator variable

$$Y_j := [|estimate[j] - p| > \varepsilon].$$

Let $Y = \sum_j Y_j$ denote the number of bad mean estimates. Our analysis in part (b) implies that if we set $N = \lceil 4/\varepsilon^2 \rceil$ inside MEANESTIMATE), then $\Pr[Y_j = 1] < 1/4$ for all j and therefore $\mathrm{E}[Y] < K/4$.

The median estimate p^* is larger than $p + \varepsilon$ if and only if at least half of the mean estimates are larger than $p + \varepsilon$. Similarly, $p^* < p - \varepsilon$ if and only if at least half of the mean estimates are larger than $p + \varepsilon$. Thus,

$$\Pr[|p^* - p| > \varepsilon] \;\le\; \Pr[Y \ge K/2]$$

The indicator variables Y_j are mutually independent (because the coin flips inside MEANESTIMATE are mutually independent). However, we cannot apply Chernoff bounds directly to Y, because we would eventually need a *lower* bound on $\mathrm{E}[Y]$.

Let $Z_1, Z_2, \ldots, Z_d$ be mutually independent indicator variables where $\Pr[Z_i = 1] = 1/4$ for all i, and let $Z = \sum_{i=1}^d Z_i$. We immediately have

$$\Pr[Y \ge K/2] \;\le\; \Pr[Z \ge K/2];$$

intuitively, in any sequence of K independent coin flips, if we increase the probability that each coin comes up heads, we also increase the probability of getting at least $K/2$ heads.

Finally, we apply the Chernoff bound $\Pr[X \ge (1+\Delta)\mu] < \exp(-\Delta^2\mu/3)$ with $\mu = \mathrm{E}[Z] = K/4$ and $\Delta = 1$:[a]

$$\Pr[Z \ge K/2] \;=\; \Pr[Z \ge 2\mu] \;\le\; \exp(-\mu/3) \;=\; \exp(-K/12).$$

We conclude that if we set $K = \lceil 12 \ln(1/\delta) \rceil$, then $\Pr[|p^* - p| > \varepsilon] < \delta$, as required. ■

[a] Sorry, δ was already taken.

Rubric: 4 points. -1 for implicitly assuming that $\mathrm{E}[Y] = K/4$. A perfect solution must explicitly invoke the fact that the mean estimates are mutually independent. This is more detail than necessary for full credit.