# Homework 2 - Solutions
# Fundamental Algorithms, Fall 2001
# Professor Yap

October 19, 2001

1. (a) (10 points) Since $H_n = \ln n + \Theta(1)$ it is enough to show that $H_n \to \infty$ as $n \to \infty$.

   For $n = 2^k$ we have:

$$\begin{aligned}
H_{2^k} &= \sum_{i=1}^{2^k} \frac{1}{i} \\
&= 1 + \frac{1}{2} + (\frac{1}{3} + \frac{1}{4}) + (\frac{1}{5} + \ldots + \frac{1}{8}) + \ldots + (\frac{1}{2^{k-1}+1} + \ldots + \frac{1}{2^k}) \\
&\geq 1 + \frac{1}{2} + (\frac{1}{4} + \frac{1}{4}) + (\frac{1}{8} + \ldots + \frac{1}{8}) + \ldots + (\frac{1}{2^k} + \ldots + \frac{1}{2^k}) \\
&= 1 + \frac{1}{2} + +2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + \ldots + 2^{k-1} \cdot \frac{1}{2^k} = 1 + \frac{k}{2}.
\end{aligned}$$

   We want to prove that $H_n$ is unbounded. Let M be any positive number. We show that

$$H_n > M$$

   for $n \geq n_0$ where $n_0$ is a natural number which we will compute. Let $k_0 = \lceil 2M \rceil$ and $n_0 = 2^{k_0}$. Then we have

$$H_{n_0} \geq \frac{k_0}{2} \geq M.$$

   We know also that $H_{n+1} = H_n + \frac{1}{n+1}, \forall n \geq 1$ so it follows $H_{n+1} \geq H_n, \forall n \geq 1$. Using this observation we get:

$$H_n \geq H_{n_0} \geq M, \forall n \geq n_0.$$

   Hence, we proved that $H_n$ is unbounded and defines an increasing sequence. That means exactly $H_n \to \infty$ as $n \to \infty$.

   (b) (10 points) We split $H_n$ as follows:

$$\begin{aligned} H_n &= \left(\sum_{i=1}^{\lfloor n^{\frac{1}{2}} \rfloor} \frac{1}{i}\right) + \left(\sum_{\lfloor n^{\frac{1}{2}} \rfloor + 1}^{n} \frac{1}{i}\right) \\ &\leq (1 + \ldots + 1) + \left(\frac{1}{n^{\frac{1}{2}}} + \ldots + \frac{1}{n^{\frac{1}{2}}}\right) \\ &< n^{\frac{1}{2}} + \frac{n}{n^{\frac{1}{2}}} \\ &= 2 \cdot n^{\frac{1}{2}}. \end{aligned}$$

Hence $H_n = O(n^{\frac{1}{2}})$.

2. (a) (5 points) It is easy to compute the probability $\Pr(A_n|B)$, this is just $\frac{1}{2^n}$. But what you need is the "opposite", $\Pr(B|A_n)$. In general, for any two events $A$ and $B$, if you need to compute $Pr(B|A)$ from $Pr(A|B)$, you must use Bayes's theorem:

   **Theorem 1** *(Bayes) For two events $A$ and $B$, both with nonzero probability, we have:*

   $$Pr(B \mid A) = \frac{Pr(B)Pr(A \mid B)}{Pr(B)Pr(A \mid B) + Pr(\bar{B})Pr(A \mid \bar{B})}.$$

   [It is very easy to remember how to derive this theorem; this will also help you remember Bayes' formula above.] Returning to our problem, we have : $Pr(B) = \frac{1}{2}$, $Pr(A_n \mid B) = 1$, $Pr(\bar{B}) = \frac{1}{2}$ and $Pr(A_n \mid \bar{B}) = \frac{1}{2^n}$. Applying Bayes's theorem we get:

   $$Pr(B \mid A_n) = \frac{\frac{1}{2} \cdot 1}{\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2^n}} = \frac{2^n}{2^n + 1}.$$

   (b) (5 points) After we flip a coin $n$ times and get head each time, we want to conclude that the coin is fake with probability 99%. In part (a), we computed the probability that the coin is fake if we had only heads after $n$ flips. So we just compute a minimal n such that this probability is at least 99/100. This means $Pr(B \mid A_n) \geq \frac{99}{100}$, or

   $$\frac{2^n}{2^n + 1} \geq \frac{99}{100}$$

   and we get $2^n \geq 99$ so $n = 7$.

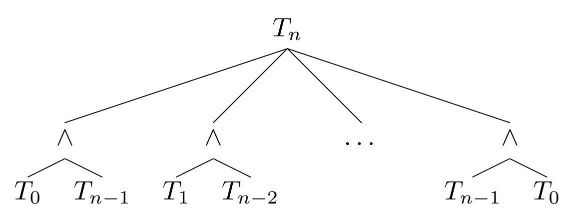
   This is essentially it. But actually, there are two cases to consider in the overall analysis. We have solved the "hard" case but there is also an easy case: while trying to flip the coin 7 times, if you see a tail, you can immediately stop flipping and conclude that the coin is the fair coin (this conclusion is 100% correct). But if you do not see any tails after the 7th flip, the above calculations say that you can conclude that the coin is biased, with only 1% chance of error.

3. (10+5+10 points) The probability space in the analysis of Quicksort, *for a fixed input of size $n$*, is the set of all complete runs of the algorithm. Let $S_n$ be the sample space for this probability space.

   Each run of Quicksort is determined by the positions of the pivots during the execution of Quicksort. At the first step, the last element is taken as pivot and can be placed on each of the n positions in the array of size n. Let the first pivot be placed on the k-th position. Then the Quicksort will run for an array of size $k-1$ and on another array of size $n-k$. If $\omega$ is a complete run and $\omega_1$ is the run on the array of size $k-1$, $\omega_2$ is the run on the other array of size $n-k$ then we have:

$$Pr(\omega) = \frac{1}{n} Pr(\omega_1) Pr(\omega_2).$$

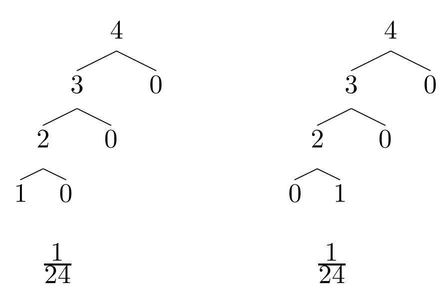
   Let us represent the entire space of possibilities by a single tree $T_n$. (The set $S_n$ is "embedded" inside $T_n$, as we shall see.) When $n = 0$ or $n = 1$, there are no pivots to choose, and thus $T_0$ and $T_1$ consists of just a single node with no children. For $n > 1$, the tree $T_n$ is built recursively from $T_0, \ldots, T_{n-1}$ as follows: at the root, there are $n$ choices for the pivots and once the pivot $i$ is chosen, we have two children corresponding to $T_{i-1}$ and $T_{n-i}$. Here is the picture:

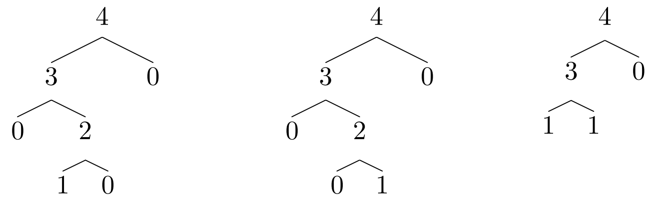



   Each $\omega \in S_n$ can be viewed as a subtree of $T_n$ in which, at each root of a subtree $T_k$, we choose a pivot between 1 and $k$; this gives a child of $T_k$ with 2 children – we must choose BOTH children of this node, and repeat the process recursively.

   For instance, when $n = 4$, we enumerate the following subtrees of $S_4$:

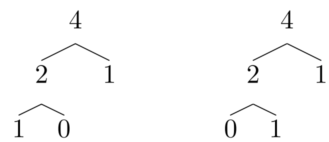
   $k = 4$: the pivot is in the last position:





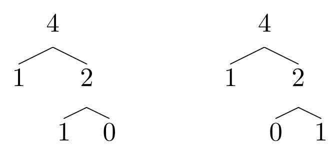

$\frac{1}{24}$ $\qquad$ $\frac{1}{24}$ $\qquad$ $\frac{1}{12}$

$k = 3$: If the pivot is in the third position,



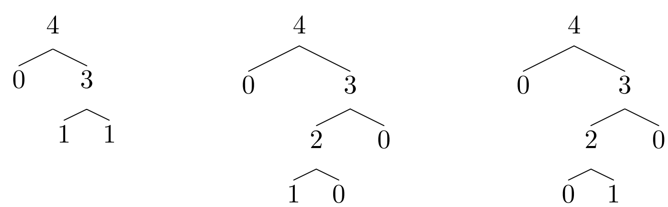

$\frac{1}{8}$ $\qquad$ $\frac{1}{8}$

$k = 2$: If the pivot is in the second position,

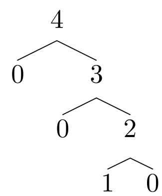



$\frac{1}{8}$ $\qquad$ $\frac{1}{8}$

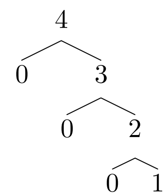
$k = 1$: If the pivot is in the first position,




$\frac{1}{12}$ $\qquad$ $\frac{1}{24}$ $\qquad$ $\frac{1}{24}$




$\frac{1}{24}$




$\frac{1}{24}$

Hence, we see that $|S_4| = 14$ when $n = 4$.

Let's try to find upper and lower bounds on the size of $|S_n|$. We denote $C_n := |S_n|$. We get that:

$$C_n = \sum_{i=1}^{n} C_{i-1}C_{n-i}$$

since a run is determined by the position $i$ of the first pivot and the runs on the resulting subarrays. We have: $C_0 = C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14$. These numbers $C_n$ are called the Catalan numbers, and they also count the number of binary trees with exactly $n$ nodes. An exact formula is also known:

$$C_n = \binom{2n}{n}\frac{1}{n+1}.$$

However, in this problem we do not assume you know this fact, and only want you to estimate upper and lower bounds on $C_n$.

We observe that for $n \geq 2$ $C_n \geq 2^{\frac{n}{2}}$ (and this is true for 2,3,4,5). We prove that by induction. We assume $C_i \geq 2^{\frac{i}{2}}$ $\forall 1 < i < n$. Hence

$$C_n \geq \sum_{i=3}^{n-2} 2^{\frac{n-1}{2}} \geq (n-4)2^{\frac{n-1}{2}} \geq 2^{\frac{n}{2}}. \text{ (We take only } n \geq 6)$$

For an upper bound we prove by induction that $C_n \leq n^n$. It is true for $n = 1$. We have:

$$C_n \leq \sum_{i=1}^{i=n}(i-1)^{i-1}(n-i)^{n-i} \leq \sum_{i=1}^{i=n} n^{i-1}n^{n-i} = n^n.$$

4. (10 points) We want to prove that Randomize-in-Place(A) produces a uniform random permutation, that means at termination this algorithm produces every permutation with probability $\frac{1}{n!}$. So, let $\sigma = (x_1, x_2, \ldots, x_n)$ be any permutation of $[1, \ldots, n]$. Let $A_i$ be the event $A[i] = x_i, \forall i = 1 \ldots n$. Using the following formula:

$$Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = Pr(A_1) \cdot Pr(A_2 \mid A_1) \cdots Pr(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

for any collection of events $A_1, \ldots, A_{n-1}$, we get

$$Pr(A = \sigma) = Pr(A[1] = x_1, \ldots A[n] = x_n) = Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = \frac{1}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{1} = \frac{1}{n!}.$$

5. (10 points) Let $A_i$ be the event that $P[i]$ is unique. Using the formula:

   $Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = Pr(A_1) \cdot Pr(A_2 \mid A_1) \cdots Pr(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$

   for any collection of events $A_1, \ldots, A_{n-1}$, we get

$$\begin{aligned} Pr(\text{all elements in P are unique}) &= Pr(A_1 \cap A_2 \cap \ldots \cap A_n) \\ &= Pr(A_1) \cdot Pr(A_2 \mid A_1) \cdots Pr(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \\ &= \frac{n^3}{n^3} \cdot \frac{n^3 - 1}{n^3} \cdots \frac{n^3 - i}{n^3} \cdots \frac{n^3 - (n-1)}{n^3}. \end{aligned}$$

   Now let's try to prove that the right side is at least $1 - \frac{1}{n}$. First, let us observe that $1 - \frac{i}{n^3} \geq 1 - \frac{n}{n^3} = 1 - \frac{1}{n^2}$ for all $i = 1 \ldots n-1$. So

$$Pr(\text{all elements in P are unique}) \geq \prod_{i=1}^{n-1} (1 - \frac{1}{n^2}).$$

   Now we use the following remark: for $a, b \geq 0$, we have $(1-a)(1-b) \geq (1-a-b)$. From this we can deduce easily that for all $a_1, \ldots, a_n \geq 0$, we have $(1-a_1)\cdots(1-a_n) \geq (1 - a_1 - \ldots - a_n)$. Hence, we get:

$$\begin{aligned} Pr(\text{all elements in P are unique}) &\geq \prod_{i=1}^{n-1} (1 - \frac{1}{n^2}) \\ &\geq 1 - \sum_{i=1}^{n-1} \frac{1}{n^2} \\ &= 1 - \frac{n-1}{n^2} \\ &\geq 1 - \frac{n}{n^2} = 1 - \frac{1}{n}. \end{aligned}$$

**Additional Questions**

1 The idea is similarly to the one in the first problem. We are interested only in the case $c \leq \frac{1}{2}$ since for $c \geq \frac{1}{2}$ we have $H_n \in O(n^{\frac{1}{2}}) \subseteq O(n^c)$. Take $\ell$ the smallest number such that $\ell c \geq 1$( i.e. $\ell = \lceil \frac{1}{c} \rceil$. The idea is to split the sum in $\ell$ parts as follows:

$$
\begin{aligned}
H_n &= \Big(\sum_{i=1}^{\lfloor n^c \rfloor} \frac{1}{i}\Big) + \Big(\sum_{\lfloor n^c \rfloor + 1}^{\lfloor n^{2c} \rfloor} \frac{1}{i}\Big) + \ldots + \Big(\sum_{\lfloor n^{(\ell-1)c} \rfloor + 1}^{n} \frac{1}{i}\Big) \\
&\leq n^c + \frac{n^{2c}}{n^c} + \ldots + \frac{n^{\ell c}}{n^{(\ell-1)c}} \\
&= \ell \cdot n^c.
\end{aligned}
$$

Since $\ell$ is constant, we have proved $H_n \in O(n^c)$.

2 Here is a trick: use the fact that

$$
\frac{1}{k^2} < \frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}.
$$

Then

$$
\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k^2} &< 1 + \sum_{k=2}^{\infty} \left( \frac{1}{k-1} - \frac{1}{k} \right) \\
&= 1 + \Big(\frac{1}{1} - \frac{1}{2}\Big) + \Big(\frac{1}{2} - \frac{1}{3}\Big) + \Big(\frac{1}{3} - \frac{1}{4}\Big) + \cdots \\
&= 2.
\end{aligned}
$$

If you want to use calculus, we could also use approximation by integrals of a sum (textbook, page 1067). We take $f(x) = \frac{1}{x^2}$ and we get:

$$
\sum_{k=2}^{n} \frac{1}{k^2} \leq \int_1^n \frac{dx}{x^2} = -\frac{1}{x} \,\Big|_1^n = 1 - \frac{1}{n} \leq 1.
$$

Hence $\sum_{k=1}^{n} \frac{1}{k^2} \leq 1 + 1 = 2$.

3 (a) We use the following approximations. On the one hand we have:

$$
\sum_{k=1}^{n} k^r \leq \sum_{k=1}^{n} n^r = n^{r+1}.
$$

On the other we have:

$$
\begin{aligned}
\sum_{k=1}^{n} k^r &= 1 + 2^r + \cdots + \lfloor \frac{n}{2} \rfloor^r + \lceil \frac{n}{2} \rceil^r + \cdots + n^r \\
&\geq \sum_{k=\lceil \frac{n}{2} \rceil}^{n} \Big(\frac{n}{2}\Big)^r \\
&= \Big(n - \lceil \frac{n}{2} \rceil + 1\Big)\Big(\frac{n}{2}\Big)^r
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{n}{2}(\frac{n}{2})^r \\
&= (\frac{n}{2})^{r+1} \\
&= \frac{n^{r+1}}{2^{r+1}}
\end{aligned}
$$

Since $2^{r+1}$ is constant, we get

$$\sum_{k=1}^{n} k^r = \Theta(n^{r+1}).$$

Another way to solve the problem is to use calculus. In this case we take $f(x) = x^r$. We get:

$$\int_0^n x^r dx \leq \sum_{k=1}^{n} k^r \leq \int_1^{n+1} x^r dx.$$

We compute the integrals and get:

$$\frac{n^{r+1}}{r+1} \leq \sum_{k=1}^{n} k^r \leq \frac{(n+1)^{r+1} - 1}{r+1}.$$

Both bounds are in $\Theta(n^{r+1})$.

(b) We use the same idea as above.
For an upper bound we have:

$$\sum_{k=1}^{n} \ln^s k \leq \sum_{k=1}^{n} \ln^s n = n \ln^s n.$$

Let's find now a lower bound. We have:

$$
\begin{aligned}
\sum_{k=1}^{n} \ln^s k &= \ln^s 1 + \ln^s 2 + \cdots + \ln^s \lfloor \frac{n}{2} \rfloor + \ln^s \lceil \frac{n}{2} \rceil + \cdots + \ln^s n \\
&\geq \sum_{k=\lceil \frac{n}{2} \rceil} \ln^s (\frac{n}{2}) \\
&= (n - \lceil \frac{n}{2} \rceil + 1) \ln^s (\frac{n}{2}) \\
&\geq \frac{n}{2} \ln^s (\frac{n}{2}) \\
&= \frac{n}{2} \ln^s n - \frac{n}{2} \ln^s 2
\end{aligned}
$$

Both bounds are in $\Theta(n \ln^s n)$.

Another solution with calculus would be:
We take $f(x) = \ln^s(x)$ and first we compute the indefinite integral $I_s = \int \ln^s x dx$. We use integration by parts:

$$I_s = \int (x)' \ln^s x dx = x \ln^s x - \int x \frac{s}{x} \ln^{s-1} x dx = x \ln^s x - sI_{s-1}.$$

For $s = 1$ $I_1 = x \ln x - x$ and then by induction we get:

$$I_s = x \ln^s x - sx \ln^{s-1} x + s(s-1)x \ln^{s-2} x - \cdots + (-1)^s s! x.$$

We denote by $F(y) = \int_1^y \ln^s x dx = [x \ln^s x - sx \ln^{s-1} x + s(s-1)x \ln^{s-2} x - \cdots + (-1)^s s! x] |_1^y = y \ln^s y - sy \ln^{s-1} y + s(s-1)y \ln^{s-2} y - \cdots + (-1)^s s! y - (-1)^s s!$. We observe that $F(n) \in \Theta(n \ln^s n)$. For our sum we get:

$$F(n) = \int_1^n \ln^s x dx \leq \sum_{k=2}^{n} \ln^s k \leq \int_2^{n+1} \ln^s x dx = F(n+1) - F(2).$$

Hence our sum is in $\Theta(n \ln^s n)$.

(c) We try to find an upper bound and a lower bound for $\sum_{k=1}^{n} k^r \ln^s k$.
For an upper bound:

$$\sum_{k=1}^{n} k^r \ln^s k \leq \sum_{k=1}^{n} n^r \ln^s n = n^{r+1} \ln^s n.$$

For a lower bound:

$$\begin{aligned}
\sum_{k=1}^{n} k^r \ln^s k &= 1^r \ln^s 1 + \cdots + (\lfloor \frac{n}{2} \rfloor)^r \ln^s(\lfloor \frac{n}{2} \rfloor) + \\
&+ (\lceil \frac{n}{2} \rceil)^r \ln^s(\lceil \frac{n}{2} \rceil) + \cdots + n^r \ln^s n \\
&\geq \sum_{k=\lceil \frac{n}{2} \rceil} (\frac{n}{2})^r \ln^s(\frac{n}{2}) \\
&= (n - \lceil \frac{n}{2} \rceil + 1)(\frac{n}{2})^r \ln^s(\frac{n}{2}) \\
&\geq (\frac{n}{2})^{r+1} \ln^s(\frac{n}{2})
\end{aligned}$$

Both bounds are in $\Theta(n^{r+1} \ln^s n)$.
Again the problem can be solved using calculus. We proceed similarly as in (b), the only change is the computation of the integral.
We denote: $I_s = \int x^r \ln^s x dx$. We have:

$$I_s = \int (\frac{x^{r+1}}{r+1})' \ln^s x dx = \frac{x^{r+1}}{r+1} \ln^s x - \int \frac{x^{r+1}}{r+1} \frac{s}{x} \ln^{s-1} x dx = \frac{x^{r+1}}{r+1} \ln^s x - \frac{s}{r+1} I_{s-1}.$$

We have $I_1 = \frac{x^{r+1}}{r+1} \ln x - \frac{x^{r+1}}{(r+1)^2}$. By induction one can get an exact formula for $I_s$. The dominant term is $x^{r+1} \ln^s x$. The same argument as above gives that our sum is in $\Theta(n^{r+1} \ln^s n)$.