Math 261B

Notes Lecture 7 - 31 Jan 2012

Probabilistic Recurrences

Example: k-selection: We wish to select the kth smallest number out of a set S of n distinct numbers.

Algorithm: Pick a random element $r \in S$. Compare r with the elements in $S \setminus \{r\}$, and partition into $L = \{y \in S \mid y < r\}$, $U = \{y \in S \mid y > r\}$. If $|L| \geq k$ we get a subproblem on L. If $|L| = k-1$, we return r and are done. If $|L| < k-1$, we solve the subproblem on U with a new $k' = k - 1 - |L|$. The amount of work needed to solve the problem is $T(n) = n + T(H(n))$ where $H(n)$ is the number of elements in the subproblem. Note that $H(n)$ is a random variable. In fact it is a sequence of stochastic random variables. In general we wish to solve

$$T(x) = a(x) + T(H(x)).$$

We use a tool from Karp: First, solve the deterministic recurrence relation $u(x) = a(x) + u(m((x))$ where $m(x) \geq E(|H(x)|)$ is an upper bound on the expectation. In our example with k-selection, we will be solving $u(n) = n + u(\frac{3}{4}n)$. Taking $m(x) = \frac{3}{4}n$, , then

$$u(x) = n + \frac{3}{4}n + \left(\frac{3}{4}\right)^2 n + \cdots = \left(\sum_{i=1}^{\infty} \left(\frac{3}{4}\right)^i\right) n = 4n.$$

In general

$$u(x) = \sum a(m^{(i)}(x)) \text{ where } m^{(i)}(x) = m(m^{i-1}(x)).$$

We have left to show that $m(x) = \frac{3}{4}n \geq E[H(n)]$. To compute $E[H(n)]$, consider the elements of S to be $s_1 < s_2 < \cdots < s_n$. Then the expectation is computed considering the sizes of L and U for $r = s_1$, $r = s_2$, etc. Then we see that

$$E[H(n)] = \frac{1}{n}\left((n-1) + \cdots + (n-k+1) + k + \cdots + (n-1)\right) = \frac{1}{n}\left(\binom{n}{2} - \binom{n-k+1}{2} + \binom{n}{2} - \binom{k}{2}\right).$$

We wish to minimize $\binom{n-k}{2} + \binom{k}{2}$, which occurs at $k = \frac{n}{2}$. This value gives $\frac{1}{n}(2 \cdot \frac{n^2}{2} - \frac{n^2}{4}) = \frac{3}{4}n$.

Theorem. *(Karp)*

$$P(T(x) > u(x) + ta(x)) \leq \left(\frac{m(x)}{x}\right)^t$$

provided $a(x)$, $m(x)$, and $\frac{m(x)}{x}$ are all non-decreasing.

For k-selection, this says

$$P(T(n) > 4n + tn) \leq \left(\frac{3}{4}\right)^t.$$

We remark that this is close to the best bound that can possibly be attained. To see this, we want to get a lower bound on $P(T(n) > 4n + tn)$. Define a "bad splitter" if $\frac{n}{|U|} \geq \log\log n$ of if $\frac{n}{|L|} \geq \log\log n$. Then $P(\text{pick a bad splitter}) = \dfrac{2}{\log\log n}$ so $P(\text{ bad splitter in } \log\log n \text{ consecutive rounds}) = \left(\dfrac{2}{\log\log n}\right)^{\log\log n}$. Therefore

$$T(\text{bad splitter in } \log\log n \text{ rounds }) \geq n + n\left(1 - \frac{1}{\log\log n}\right) + n\left(1 - \frac{1}{\log\log n}\right)^2 + \cdots + n\left(1 - \frac{1}{\log\log n}\right)^{\log\log n}$$
$$= \Omega(n \log\log n),$$

recalling that $(1 - \frac{1}{n})^n \to \frac{1}{e}$ as $n \to \infty$. Taking $t = \log\log n$, the upper bound we get is $\left(\frac{3}{4}\right)^{\log\log n}$ and the lower bound is $\left(\dfrac{2}{\log\log n}\right)^{\log\log n}$. The lower bound can be expressed as $c^{\log\log n \cdot \log\log\log n}$. So these bounds are very close.

We will now prove the theorem for the case of k-selection. We will proceed by induction. This is true for $t = 0$. Then

$$\begin{aligned} P(T(x) > 4n + tn) &= P(n + T(H(x)) > 4n + tn) \\ &= P(T(H(x)) > 3n + tn) \\ &= \frac{1}{n}\left(\sum_i P(T(x_i) > 3n + tn)\right) \\ &= \frac{1}{n}\sum_i P\left(T(x_i) > 4x_i + x_i\left(\frac{(t+3)n - 4x_i}{x_i}\right)\right) \\ &\leq \frac{1}{n}\sum_i \left(\frac{3}{4}\right)^{\frac{(t+3)n}{x_i} - 4} \end{aligned}$$

by the induction hypothesis. We want to show that this is bounded above by $\left(\frac{3}{4}\right)^t$, so we will be done if we can show that $\left(\frac{3}{4}\right)^{(t+2)n/x-4} \leq \frac{x}{n}\left(\frac{3}{4}\right)^{t-1}$ or in other words, if $\left(\frac{3}{4}\right)^{(t+2)n/x} \leq \frac{x}{n}\left(\frac{3}{4}\right)^{t+3}$. Set $y = \frac{n}{x} > 1$, then this is true if and only if $\left(\frac{3}{4}\right)^{(y-1)(t+3)} \leq \frac{1}{y}$. This can be verified by calculus, so we are done.