Comp 610 Lecture 5: Selection and Priority Queues Continued

We completed the analysis of our algorithm to find the median (or the kth largest) which was expected to use 3n/2 +o(n) comparisons. We will apply the following variant of the Chrnoff Bounds.

The probability that Bin(m,p) differs from its expected value by more than t, is at most e to the $(-t^2/2mp)$ for any t<mp.

Our algorithm first selected a random subset S of $n/\log^2 n$ elements. It then sorted this set using $O(|S| \log |S|) = o(n)$ comparisons and found the $((k-n/\log n)/\log^2 n)$th element l and the $((k+n/\log n)/\log^2 n)$th element u of S. By comparing each element to a randomly chosen element of {u,l}, and to both if necessary, it split the elements other than u and l into the set L of those smaller than l, the set H of those bigger than u and the set M of those lying between the two pivots. We remark that if l<1 (respectively u>n) we set l=1 (respectively u=n).

If |L|<k and |M|< n/root(log n) and |H|<n-k then it used our deterministic algorithm to find and return the $(k-|L|)^{th}$ element of M in o(n) steps. Otherwise it sorted the entire set using n log n comparisons and returned the kth element. We have seen t that the expected number of comparisons used by the algorithm was 3n/2+o(n), unless it actually resorted to sorting all n elements. We show now that the probability this occurred was $o(n^{-2})$. We can and do assume n is large enough to satisfy certain lower bounds in the proof.

We let m=|S| and consider choose the elements of S, one by one, indexing them in the order they are chosen as $s_1,\ldots,s_m$. Each s_i is chosen uniformly from those elements not yet in S.

We bound first the probability that S has fewer than $((k-n/\log n)/\log^2 n)$ -1elements smaller than the k^{th} element. Letting A_i be the set of s_j with j<i which are smaller than the kth element, we see that the probability that s_i is less than the k^{th} element is precisely $(k-|A_i|)/(n-i)$. If $|A_i|$ is at most $(k-n/\log n)/\log^2 n$ -1 this is at least $p=k(1-1/\log^2 n)/n$. So, while A_i is less than $(k-n/\log n)/\log^2 n$-1, the probability that s_i is less than k is at least p. If follows that the probability that S has fewer than $((k-n/\log n)/\log^2 n)$ -1elements less than k is at most the probability that Bin(m,p) is at most $((k-n/\log n)/\log^2 n)$ -1
Since the expectation of a sum is the sum of the expectations, the expectation of $Bin(m,p)=mp>(k/\log^2 n)(1-1/\log^2 n)$. So, the probability that Bin(m,p) is at most $((k-n/\log n)/\log^2 n)$ -1 is at most the probability that Bin(m,p) is lower than its expected value by at last $n/\log^3 n -1-k/\log^4 n> n/2\log^3 n$. We let $t=n/2\log^3 n$.

If t is more than mp, then we are asking for the probability that S has a negative number of elements less than k, which has probability 0. Otherwise, Applying the

Chernoff bound we obtain that this probability is at most e to the ($-t^2/2mp$). This is less than e to the ($n/8\log^6 n$) which is $o(n^{-2})$.

We bound next the probability that S has more than $((k+n/\log n)/\log^2 n)$ elements smaller than the k^{th} element. We can assume k is more than $n/\log^3 n$ as otherwise the probability that this happens is zero. there is nothing to prove. Letting A_i be the set of s_j with j<i which are smaller than the kth element, we see that the probability that s_i is less than the k^{th} element is precisely $(k-|A_i|)/(n-i)$. This is at most $p=k/(1-1/\log^2 n)n$. So, the probability that S has more than $((k+n/\log n)/\log^2 n)$ -1elements less than k is at most the probability that Bin(m,p) exceeds $((k+n/\log n)/\log^2 n)$ -1
Since the expectation of a sum is the sum of the expectations, the expectation of $Bin(m,p)=mp<(k/\log^2 n)(1+2/\log^2 n)$. So, the probability that Bin(m,p) is at least $((k+n/\log n)/\log^2 n)$ -1 is at most the probability that Bin(m,p) exceeds its expected value by at last $n/\log^3 n$ -1-$2k/\log^4 n> n/2\log^3 n$. We let t be the minimum of the expected value of Bin(m,p) or $n/2\log^3 n$. So, by our bound on k, t is at least $n/2\log^3 n$.

Applying the Chernoff bound we obtain that this probability is at most
e to the ($-t^2/8mp$) this is less than e to the ($n/8\log^6 n$) which is $o(n^{-2})$.

We bound next the probability that M has more than n/root(log n) elements. If l is the i^{th} element, this implies that letting B_i be the i^{th} through the $(i+ n/root(\log n))^{th}$ element, S contains at most $2n/\log^3 n$ elements of B_i. To complete the proof we need only show that the probability this occurs for a particular B_i is $o(n^{-3})$ as then the probability it occurs for some i is, by the subadditivity of probabilities $o(n^{-2})$.
Letting C_i be the set of s_j with j<i which are in B_i, we see that the probability that s_i is in B_i is precisely $((n/root(\log n)-|C_i|)/(n-i)$. While $|C_i|<2n/\log^3 n$ this is at least $p=(1-1/\log n)/root(\log n)$. So, the probability that S has fewer than $2n/\log^3 n$ elements in B_i is at most the probability that Bin(m,p) is less than $2n/\log^3 n$.
Since the expectation of a sum is the sum of the expectations, the expectation of $Bin(m,p)=mp>(1-1/\log n)n/\log^{5/2} n>n/2\log^{5/2} n$. So, the probability that Bin(m,p) is at most $2n/\log^3 n$ is at most the probability that Bin(m,p) is less than half of its expected value

Applying the Chernoff bound we obtain that this probability is $o(n^{-3})$.

Our discussion of heap-building bounds was Section 4 (bottom of page 20 and tops of page 21 and 22) of the paper of Zhentao and Reed to which there is a link on the webpage.