CS 122 Fall 2010 HW 3, Due Weds. Oct. 20

This homework concerns the material on randomized algorithms. The homework is heavy on concept and understanding. This is the kind of homework that benefits from starting early and letting your brain work on it - as opposed to "recitation" or "application" types of homework that can be done at the last minute.

1. On page 731 of the book the second paragraph has the statement

$X = X_0 + X_1 + X_2 + ...$, where X_j is the expected number of steps spent by the algorithm in phase j.

I think there is an error in this statement. If you understand the analysis, you should be able to spot it. What is it, and why is it an error, how is it fixed?

2a. Suppose in the analysis of Randomized Select (S,k) we defined a central splitter to be one that is larger than at least one sixth of the elements and smaller than at least one sixth of the elements. Then what would the analysis show about the expected number of comparisons done by the algorithm? That is, before we showed it was at most $8n$. What is it now?

2b. Given your answer to part 2a, how would you answer the question asked in class: "Why 3/4?"

3. The analysis of Randomized Select (S,k) showed that the expected number of comparisons the algorithm does while in phase j is at most $2n(\frac{3}{4})^j$. This was done by considering the expected number of comparisions done until a central splitter is picked, using the fact that the algorithm will be out of phase j after a central splitter is picked.

Here are two questions:

3a. Must a central splitter be found before the algorithm leaves phase j? If no, give an example, and if yes, give a convincing argument.

3b. If the algorithm can leave phase j without picking a central splitter, how does that affect the above analysis, since it was based on seeing what happens when a central splitter is picked.

4. Read the discussion of randomized quicksort starting on p. 731 in the book. Then answer the following questions:

4a. Splitting a type j subproblem creates two disjoint subproblems of higher type. Is it important in the analysis of the expected running time of the algorithm that the subproblems are disjoint? Explain.

4b. Statement 13.20 on p. 733 says that the number of type j subproblems is at most $(\frac{4}{3})^{j+1}$. I think this needs more explanation. Give a more complete explanation for how this is derived. It only needs to be a few lines, but it has to have the right elements.

Optional: For extra credit, program the randomized Select algorithm or the randomized quicksort algorithm along with a counter that counts the number of comparisons. Then give it test data to confirm empirically the results we obtained analyticaly. Also see that the choice of S does not matter, by trying very differt types of input. For example, when S is completely sorted, or reverse sorted, or in random order.