ECS222a Graduate Algorithms
Midterm Preparation Questions

Do not hand in the solutions to these problems. But since the midterm questions will be related to these problems, it would be wise to think about them.

We will discuss the solutions in class on Monday, along with reviewing other material which might be helpful for the midterm. Of course, you should look over the lecture topics, the readings, and the "questions to think about".

1. Which of the following probability formulas hold only for independent random variables a and b, and which hold for any pair of random variables a and b:

$$\Pr[a = 1 \text{ and } b = 1] = \Pr[a = 1]\Pr[b = 1] \quad (1)$$
$$E[a + b] = E[a] + E[b] \quad (2)$$
$$E[ab] = E[a]E[b] \quad (3)$$
$$\Pr[a = 1 \text{ or } b = 1] = 1 - \Pr[a \neq 1]\Pr[b \neq 1] \quad (4)$$
$$\Pr[a = 1 \text{ or } b = 1] \leq \Pr[a = 1] + \Pr[b = 1] \quad (5)$$

2. In this problem we will get a somewhat better bound than we did in lecture on the probability that our universal hash function

$$h(x) = ((ax + b) \mod p) \mod m$$

puts too many items in one bucket. Consider again hashing m items into a table of size m.
a) What was the expected number of collisions?
b) Argue that Pr[any bucket contains $\geq 2\sqrt{n}$ items] $\leq$ Pr[number of collisions $\geq 3n$].
c) Use Markov's inequality to get an upper bound on Pr[number of collisions $\geq 3n$].

3. Get a worst-case upper bound on the average length of the path from the root to any node (leaf or internal) in a binary tree, where each node may have zero, one or two children.

4. We are given a set of n items in *sorted* order. We want to build a skip list to contain these items. Explain, in a few sentences, or by giving very high level pseudo-code, how to construct the skip list in expected $O(n)$ time. Indicate why the construction should be $O(n)$ and not $O(nlgn)$.