CS 909, (Randomized Algorithms) Assignment Sheet 1 (Due Feb 13, 03)

1. Justify the following observation. I have to toss a fair coin $\log n$ times to obtain the first head with probability $\geq 1 - 1/n$. To obtain $\log n$ heads, I have to toss the coin $c \log n$ times for an appropriate c.

2. Consider the following game. You are given an n sided dice that comes up with a number in the range $k \in [1..n]$ uniformly at random. Subsequently you are given a k sided dice - this continues till you have a $1 - \mathit{sided}$ dice and the game ends. What is the expected number of times that you have to roll the dice before the game ends ?
   What can you say about the number of times one has to roll the dice so that the game ends with probability $1 - 1/n$?

3. Repeat the calculations for the selection algorithm for the case when the success probability $1 - \epsilon$ is given as an input. (You have to calculate the size of the sample required to obtain this probability of success to complete the algorithm in one round).

4. Suppose T is an ordered table of n keys x_i , $1 \leq i \leq n$ drawn uniformly from $(0, 1)$. Instead of doing the conventional binary search, we use the following approach.

   > Given key y, we make the first probe at the position $s_1 = \lceil y \cdot n \rceil$. If $y = x_{s_1}$, we are through. Else if $y > x_{s_1}$, we recursively search for y among the keys $(x_{s_1} \ldots x_n)$
   > else recursively search for y among the keys $(x_1 \ldots x_{s_1}$.

   At any stage when we search for y in a range $(x_l \ldots x_r$, we probe the position $l + \lceil \frac{(y - x_l)(r - l)}{x_r - x_l} \rceil$. We are interested in determining the expected number of probes required by the above searching algorithm.

   In order to somewhat simplify the analysis, we modify the algorithm as follows. In round i, we partition the input into $n^{1/2^i}$ sized blocks and try to locate the block that contains y and recursively search within that block. In the i-th round, if the block containing y is $(x_l \ldots x_r)$, then we probe the position $s_i = l + \lceil \frac{(y - x_l)(r - l)}{x_r - x_l} \rceil$. We then try to locate the $n^{1/2^i}$-sized block by sequentially probing every $n^{1/2^i}$-th element starting from s_i.

   Analyze the expected number of probes required. (Analyze the expected number of probes in each round using Chebychev's inequality).

5. Complete the proof of Lemma 4 of Lecture 1, viz., $p_k = \frac{|\mathbf{A}^{(k)}||\mathbf{B}_{(k)}|}{\sum_{l=1}^{n} |\mathbf{A}^{(l)}||\mathbf{B}_{(l)}|}$ for $k = 1, \ldots, n$, minimizes $E\left(||\mathbf{AB} - \mathbf{CR}||_F^2\right)$.

6. For obtaining an approximate bound on the number of distinct elements, we used the random variable $Z = min_i Z_i$, where $Z_i = (a_i a + b) \text{mod } p$. Can you come up with another random variable which will give us a similar approximation ? Prove it rigorously.
   Bonus exercise : If we choose Z to be the k-th order statistics of Z_i's for some $k > 1$, can we obtain superior bounds ?