CMPSCI 611 FALL '09: HOMEWORK 3
DUE 2:30 PM, NOVEMBER 17TH

- Collaborating with at most two other students is allowed but answers must be written independently. Solutions must be typed (ideally in Latex). Please write with whom you collaborated on your solutions. Unfortunately, you are not allowed to use material from the net or talk about the homework with anybody outside your collaboration group (aside from the lecturer or TA of course.)
- Solutions are due before 2:31 pm (according to the clock in the classroom) on the due date. Marks for solutions that are returned up to 24 hours late will be reduced by 25%. After that no credit will be given. This policy will be strictly enforced.
- Solutions can either
  - Be emailed to `mcgregor@cs.umass.edu` and `smurtagh@cs.umass.edu`. Use the subject line "CMPSCI 611 Homework 3" or
  - Be printed and handed in at the start of class.
- To get full marks, answers must be sufficiently detailed, supported with rigorous mathematical proofs, and be clearly explained.

Question 1 (15 marks)**.** You have a coin that has probability p of heads and probability $1-p$ of tails. You repeatedly flip the coin until you get a heads. Let X be the number of times the coin is flipped.

(1) Prove that $\mathbb{E}[X] = 1/p$.
(2) Prove the best upper bound you can for $\mathbb{P}[X \geq 10/p]$. You can assume $\mathbb{V}[X] = \frac{1-p}{p^2}$.

The manufacturer of your favorite cereal is running a promotion in which they place a coupon in each cereal box. There are n different types of coupon and the type of coupon in a box is chosen uniformly from these n types. Each week you buy a box of cereal until you have collected all the different coupons. Let Z be the number of boxes you end up buying.

(3) Prove that $\mathbb{E}[Z] = nH_n$ where $H_n = 1 + 1/2 + \ldots + 1/n$.
(4) Prove that $\mathbb{V}[Z] = n^2(\pi^2/6 + o(1)) - nH_n$. You can assume $\lim_{n\to\infty}\left(\sum_{1\leq i\leq n} i^{-2}\right) = \pi^2/6$.

Hint: Let Z_i be the number of the week on which the i-th new coupon is found and consider $X_i = Z_{i+1} - Z_i$.

Question 2 (10 marks)**.** Say we are given two lists L_1 and L_2, each containing n integers, and we want to determine if the two lists have the property that each integer occurs the same number of times in both lists. This problem could be solved by sorting the two lists in $O(n \log n)$ time, and then comparing the two sorted lists.

Describe a randomized technique for solving this problem based on verifying a polynomial identity. Demonstrate a bound on the probability of the resulting algorithm returning "yes" when the correct answer is "no" and a bound on the probability of returning "no" when the correct answer is "yes." Describe how to make these probabilities smaller than ϵ, for any given ϵ. Describe the running time of your algorithm, assuming that pairwise arithmetic operations on arbitrarily large integers can be performed in constant time.

Question 3 (10 marks)**.** For an undirected graph $G = (V, E)$ with n nodes and m edges, we define the induced subgraph on $X \subset V$, to be the graph $G[X]$ whose node set is X and whose edge set consists of all edges of G for which both ends lie in X. Design and analyze a randomized algorithm that, given G and $1 \leq k \leq n$, finds $X \subset V$ such that $G[X]$ has exactly k nodes and at least $\frac{mk(k-1)}{n(n-1)}$ edges. Your algorithm should have an expected running time that is polynomial and only output correct answers.

Question 4 (15 marks). After a late night working in the library, you decide it's time to go home and rest. You live k meters away from the library. You may assume that $k > 0$. Unfortunately, you're so tired that you find it hard to walk in a straight line. At each step you either move one meter further from home or one meter closer to home (with equal probability). You've drunk enough coffee to ensure that you can take $\frac{k^2}{1000 \log k}$ steps before falling asleep. Prove the smallest upper bound you can on the probability that you get home before falling asleep. [Hint: Try using a combination of the Chernoff bound and the Union Bound.]