CSIS0351/CSIS8601: Randomized Algorithms
Homework 1 **Due Date:** 16 Sept 2010

Rules: Discussion of the problems is permitted, but writing the assignment together is not (i.e. you are not allowed to see the actual pages of another student).

Please justify your arguments carefully. Unless otherwise stated, you may assume that there is an underlying probability space $(\Omega, \mathcal{F}, Pr)$, and all random variables take discrete values.

1. (5 points) Prove the union bound: Suppose $\{A_i\}_{i=1}^n$ is a collection of events. Then, $Pr(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n Pr(A_i)$.
   (Hint: Use induction and the fact $Pr(A) + Pr(B) = Pr(A \cup B) + Pr(A \cap B)$.)

2. (5 points) Prove the linearity of expectation: Suppose for each $1 \leq i \leq n$, X_i is a random variable and a_i is a real number. Then, $E[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i E[X_i]$.

3. (10 points) Suppose $X, Y : \Omega \to \mathbb{Z}$ are independent random variables taking integer values. In this question, we shall prove the basic result $E[XY] = E[X]E[Y]$.

   Prove from first principle (without using any results not proved in class) that $\sum_{\omega \in \Omega} X(\omega)Y(\omega)Pr(\omega) = \sum_{\omega_1 \in \Omega} X(\omega_1)Pr(\omega_1) \sum_{\omega_2 \in \Omega} Y(\omega_2)Pr(\omega_2)$.

4. (30 points) **Max 3-SAT.** Let ϕ be a 3-CNF formula with m clauses and n variables. Recall the basic randomized procedure mentioned in class for maximizing the number of satisfied clauses in the formula ϕ. In that procedure, each variable takes the value TRUE independently with probability $\frac{1}{2}$. We showed in class that the expected number of satisfied clauses is $\frac{7m}{8}$.

   The point of this question is to show that the algorithm finds an assignment satisfying a large proportion of clauses with at least a constant probability.

   (a) Let Y be the random variable denoting the number of unsatisfied clauses. Compute $E[Y]$.

   (b) Find a suitable upper bound for $Pr[Y > \frac{3m}{16}]$, and conclude that the randomized algorithm finds an assignment satisfying at least $\frac{13m}{16}$ clauses with probability at least $\frac{1}{3}$.

5. (50 points) **Max Cut.** Let $G = (V, E)$ be a graph. Recall the randomized algorithm mentioned in class for finding a cut $C \subset V$ for the graph G, namely, a point $v \in V$ is included in C independently with probability $\frac{1}{2}$. Assume that to make this decision takes 1 independent random bit for each point.

   Let $E(C) := \{\{u, v\} \in E : u \in C, v \in V \setminus C\}$ be the edges in the cut. It is shown in class that $E[|E(C)|] = \frac{|E|}{2}$. The goal of this question is to design another randomized algorithm with better guarantees. Let $0 < \epsilon < 1$ and $0 < \delta < 1$. We shall design a randomized algorithm that, with failure probability at most δ, returns a cut C such that $|E(C)| \geq (\frac{1-\epsilon}{2}) \cdot |E|$.

(a) Give an upper bound on the failure probability that the above randomized procedure returns a cut such that the number of edges $|E(C)|$ is less than $(\frac{1-\epsilon}{2}) \cdot |E|$.

(b) Show that by repeating the above randomized procedure, it is possible to obtain a better randomized algorithm with failure probability at most δ. Compute the number of independent random bits used by your algorithm.

(Hint: You might find the following inequality useful: for $0 < \epsilon < 1$, $1 + \epsilon \geq e^{\frac{\epsilon}{2}}$.)

6. **(25pt Extra Credit) Unifying Formula for Expectation.** Recall that the formula for the expectation of a random variable depends on the particular kind of random variables. For example,

(1) If X is a continuous real-valued random variable, then $E[X] = \int_{\mathbb{R}} x f(x) dx$, where $f(x) := \lim_{h \to 0} \frac{1}{h} \cdot Pr[x \leq X \leq x + h]$ is the probability density function of X.

(2) If X is a random variable taking non-negative integers, then $E[X] = \sum_{k=0}^{\infty} k \cdot Pr[X = k]$.

In this question, we investigate if it is possible to have a single formula for expectation that works for all kinds of real-valued random variables. It can be shown that if X is a continuous random variable taking non-negative reals, then $E[X] = \int_0^\infty Pr[X \geq t] dt$. (Try to verify this!) We show that this formula is valid for other kinds of random variables as well.

(a) (15 pt) Suppose X is a discrete random variable that takes only a countable number of non-negative values. Prove that $E[X] = \int_0^\infty Pr[X \geq t] dt$. (If you do not like to deal with infinite support, you may assume X only takes a finite number of values.)

(b) (10 pt) Assume that for all non-negative random variables X, $E[X] = \int_0^\infty Pr[X \geq t] dt$. Derive a similar formula for $E[X]$ when X is any real-valued random variable.