CS5314 Randomized Algorithms

Homework 2
Due: 13:10, Nov 16, 2010 (before class)

1. (20%) Let X be a binomial random variable with parameters n and p, so that X can be expressed as the sum of n indicators $X_1, X_2, \ldots, X_n$, with $\Pr(X_i) = p$. In this question, we shall give an alternative derivation for its variance.

   (a) Show that
   $$\mathrm{E}[X^2] = \sum_{i=1}^{n} \mathrm{E}[X \cdot X_i] = \sum_{i=1}^{n} \Pr(X_i = 1)\mathrm{E}[X \mid X_i = 1].$$

   (b) Compute the value of $\mathrm{E}[X \mid X_i = 1]$.

   (c) Using the results of parts (a) and (b), show that $\mathrm{Var}(X) = np(1-p)$.

2. (20%) Suppose that we flip a fair coin n times to obtain n random bits. Consider all $m = \binom{n}{2}$ pairs of these bits in some order. Let Y_i be the exclusive-or of the ith pair of bits, and let $Y = \sum_{i=1}^{m} Y_i$.

   (a) Show that each Y_i is 0 with probability 1/2 and 1 with probability 1/2.

   (b) Show that for any Y_i and Y_j, $\mathrm{E}[Y_i Y_j] = \mathrm{E}[Y_i]\mathrm{E}[Y_j]$.

   (c) Using the result of part (b), show that $\mathrm{Cov}(Y_i, Y_j) = 0$.

   (d) Using Chebyshev's inequality, show that $\Pr(|Y - E[Y]| \geq n) \leq 1/8$.

3. (20%) A fixed point of a permutation $\pi : [1, n] \to [1, n]$ is a value for which $\pi(x) = x$. Find the variance in the number of fixed points of a permutation chosen uniformly at random from all permutations.

   Hint: Let X_i be an indicator such that $X_i = 1$ if $\pi(i) = i$. Then, $\sum_{i=1}^{n} X_i$ is the number of fixed points. You cannot use linearity to find $\mathrm{Var}[\sum_{i=1}^{n} X_i]$, but you can calculate it directly.

4. (20%) Suppose you are given a biased coin that has Pr(head) $= p$. Also, suppose that we know $p \geq a$, for some fixed a. Now, consider flipping the coin n times and let n_H be the number of times a head comes up. Naturally, we would estimate p by the value $\tilde{p} = n_H/n$.

   (a) Show that for any $\epsilon \in (0, 1)$,
   $$\Pr\Big(|p - \tilde{p}| > \epsilon p\Big) < \exp\left(\frac{-na\epsilon^2}{2}\right) + \exp\left(\frac{-na\epsilon^2}{3}\right)$$

   (b) Show that for any $\epsilon \in (0, 1)$, if
   $$n > \frac{3\ln(2/\delta)}{a\epsilon^2},$$
   then
   $$\Pr\Big(|p - \tilde{p}| > \epsilon p\Big) < \delta.$$

5. (20%) Let $X_1, X_2, ..., X_n$ be independent Poisson trials such that $\Pr(X_i) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathrm{E}[X]$. During the class, we have learnt that for any $\delta > 0$,

$$\Pr(X \geq (1+\delta)\mu) < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

In fact, the above inequality holds for the *weighted* sum of Poisson trials. Precisely, let $a_1, ..., a_n$ be real numbers in $[0,1]$. Let $W = \sum_{i=1}^{n} a_i X_i$ and $\nu = \mathrm{E}[W]$. Then, for any $\delta > 0$,

$$\Pr(W \geq (1+\delta)\nu) < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\nu$$

   (a) Show that the above bound is correct.

   (b) Prove a similar bound for the probability $\Pr(W \leq (1-\delta)\nu)$ for any $0 < \delta < 1$.

6. (Further studies: No marks) For a random variable X with standard deviation $\sigma(X)$, and any positive real number t, Chebyshev's inequality implies that

$$\Pr(X - \mathrm{E}[X] \geq t\sigma(X)) \leq \Pr(|X - \mathrm{E}[X]| \geq t\sigma(X)) \leq \frac{1}{t^2}.$$

Indeed, we can bound $\Pr(X - \mathrm{E}[X] \geq t\sigma(X))$ slightly tighter. Show that

$$\Pr(X - \mathrm{E}[X] \geq t\sigma(X)) \leq \frac{1}{1+t^2}.$$

7. (Further studies: No marks) Previously, we have shown the expected time for the Randomized Quicksort to sort n numbers is $O(n \log n)$. Here, we want to prove further that the algorithm runs in $O(n \log n)$ time with high probability. Consider the following view of Randomized Quicksort. Every point in the algorithm where it decides on a pivot element is called a *node*. Suppose the size of the set to be sorted at a particular node is s. The node is called *good* if the pivot element divides the set into two parts, each of size not exceeding $2s/3$. Otherwise, the node is called *bad*.

The nodes can be thought of as forming a tree in which the root node has the whole set to be sorted and its children have the two sets formed after the first pivot step and so on.

   (a) Show that the number of good nodes in any path from the root to a leaf in this tree is not greater than $c \log_2 n$, where c is some positive constant.

   (b) Show that, with high probability (greater than $1 - 1/n^2$), the number of nodes in a given root to a leaf path of the tree is not greater than $c' \log_2 n$, where c' is another constant. (*Hint:* If c' is much larger than c, there must be many bad nodes on the path... Will that be likely to occur?)

   (c) Show that, with high probability (greater than $1 - 1/n$), the number of nodes in the *longest* root to leaf path is not greater than $c' \log_2 n$. (*Hint:* How many nodes are there in the tree?)

   (d) Use your answers to show that the running time of Randomized Quicksort is $O(n \log n)$ with probability at least $1 - 1/n$.