CPS630: Homework 2

Your submission must be written up in LaTeX.

Problem 1: Suppose there are n jobs. There is a parameter $k > 1$. Each job has a processing time of one step with probability p and k steps with probability $1-p$, independent of other jobs. Suppose these jobs are randomly assigned to m processors so that there are exactly n/m jobs on each processor. Give a high probability upper and lower bound on when all jobs will be completed. Assume m divides n for simplicity.

Problem 2: Consider the following randomized permutation routing algorithm on a hypercube. For each source node i and destination $\pi(i)$, permute the bits in i at random, and fix the bits to $\pi(i)$ in this order. In other words, the packet carries the (random) order in which bits need to be fixed. An intermediate node scans the bits for the destination of the packet in this order and fixes the next bit where the current node differs from the destination node. The random permutation of bits is independent for different sources. Note further that unlike Valiant's scheme, there is no intermediate destination for the packet. Give an instance where this scheme has expected congestion $2^{\Omega(n)}$, where n is the number of dimensions (or bits).

Problem 3: Suppose balls are thrown at random into n bins. Show that if $c_1\sqrt{n}$ balls are thrown for some constant c_1, then the probability that no two balls land in the same bin is at most $1/e$. Likewise show that if $c_2\sqrt{n}$ balls are thrown, the probability that no two balls land in the same bin is at least $1/2$. Make these constants c_1 and c_2 as tight as possible. You can use the following facts:

$$e^{-x} \geq 1 - x \qquad \text{and}$$
$$e^{-x-x^2} \leq 1 - x \qquad \text{for } x \leq 1/2$$

Problem 4: Suppose $X_1, X_2, \ldots, X_n$ are a collection of n independent Geometrically distributed random variables with mean 2. Let $X = \sum_{i=1}^{n} X_i$. Using the idea in the proof of Chernoff bounds done in class, derive a bound on $\Pr[X \geq (1+\delta)(2n)]$ where $\delta > 0$. Note that each X_i is not bounded, so you have to do the math for this case from scratch.

Problem 5: Use Chebychev's inequality to prove the weak law of large numbers: Given independent and identically distributed random variables $X_1, X_2, \ldots,$ with mean μ and standard deviation σ, then for any constant $\epsilon > 0$:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Problem 6: Consider the Set Cover problem where we are given a universe U of n elements, and sets $S_1, S_2, \ldots, S_m \subseteq U$. The goal is to choose the smallest number of sets from these such that the union of the chosen sets is U. We can write an integer program for this problem by having variable $x(i)$ for set S_i, which is set to 1 if S_i is chosen to be in our collection, and 0 otherwise. The goal therefore is to minimize $\sum_{i=1}^{m} x(i)$. The constraints are that for every element $p \in U$, we pick at least one set that contains p (*i.e.*, p is *covered* by the collection). In other words:

$$\sum_{i:p \in S_i} x(i) \geq 1 \quad \forall p \in U$$

This defines a valid integer program. We can solve its linear relaxation to obtain a fractional solution. We round the solution as follows: For every set S_i, pick S_i independently with probability $x(i)$ to be included in our collection. Here $x(i)$ is the value of the variable in the optimal fractional solution. Show the following:

1. Show that the expected size of this collection is at most the size of the optimal cover.

2. For any element p, write the expression for the probability that none of the sets containing this element belongs to the collection, in terms of the $x(i)$ values.

3. Show that the expression from the previous step is at most $1/e$ for all values of $x(i)$. This means that the probability that a particular element p is not covered by the collection is at most $1/e$.

4. Let the collection of sets we choose be denoted C_1. We repeat the random process independently to choose collections $C_2, C_3, \ldots C_{2 \log n}$, and output the union of these collections as the final answer. Show that the output is a valid set cover with probability at least $1 - 1/n$.

5. What is the expected size of the finally output set cover in terms of OPT, the size of the optimal set cover?