CPSC 536N Randomized Algorithms (Winter 2014-15, Term 2)
Assignment 1

Due: Monday January 26th, in class.

Question 1: Let X be a random variable taking values on the positive integers with $\Pr[X = x] = 2^{-x}$. Define the random variable Y by $Y = 2^X$. Use the Markov inequality to give an upper-bound $\Pr[Y > a]$. Your bound should be a function of a and should be less than 1 (for sufficiently large a).

Question 2: Consider a sequence of n unbiased coin flips. Let X be the length of the *longest* contiguous sequence of heads.

(a): Define $\ell = \lceil \log_2(1/\delta) + \log_2 n \rceil$. Show that $\Pr[X \geq \ell] \leq \delta$.

(b): Let $c \geq 1$ be arbitrary. Let $k = \log_2 n - O(\log_2 \log_2 n)$, where the constant inside the Big-O depends somehow on c. Show that, $\Pr[X < k] \leq n^{-c}$

Question 3: Let $X_1, ..., X_n$ be independent, geometric random variables with parameter $p = 1/2$. (The number of fair coin flips needed to see the first head. So $\Pr[X_1 = 1] = 1/2$, $\Pr[X_1 = 2] = 1/4$, etc.)

(a): Prove that $\mathrm{E}\left[e^{tX_i} \right] = \frac{e^t/2}{1-e^t/2}$ for all sufficiently small $t \geq 0$.

(b): Let $X = \sum_i X_i$. We will use the Chernoff-style method to prove a tail bound on X. Fix some $\delta = (0, 1)$. Prove that

$$\Pr[X \geq (1+\delta)2n] \;\leq\; \left(\frac{1+2\delta}{1+\delta}\right)^{-2(1+\delta)n} \cdot (1+2\delta)^n.$$

(c): OPTIONAL: For some constants $c_1, c_2 > 1$, prove that the upper bound from part (b) is at most

$$\begin{cases} \exp(-\delta^2 n/c_1) & \delta \in [0,1] \\ \exp(-\delta n/c_2) & \delta > 1 \end{cases}$$

Question 4: Let M be a matrix with m rows, n columns, every entry $M_{i,j} \in [0,1]$ and such that every row sums to r. (That is, $\sum_{j=1}^n M_{i,j} = r$ for all i.) Pick a vector $Y \in \{0,1\}^n$ uniformly at random. Let Z be the vector $M \cdot Y$. Let $\alpha = (r/2) + 3\sqrt{r \ln m}$. Prove that $\Pr[\max_i Z_i > \alpha] \leq 1/m$.

Question 5: Let $Z_1, ..., Z_n$ be independent, identically distributed random variables. The Z_i's all have the same expectation $\mathrm{E}[Z_i]$. It is often the case that we would like to estimate $\mathrm{E}[Z_i]$ from the sample $Z_1, ..., Z_n$.

If we assume that the Z_i's lie in a bounded interval then we can use the *average* $\sum_i Z_i/n$ to estimate $\mathrm{E}[Z_i]$ and use the Chernoff bound to show that this is a good estimate. But for this question we will **not** assume that the Z_i's lie in a bounded interval.

Instead, suppose we know that $\Pr[Z_i \geq t] \leq p$ for some t and some p. Let M be the **median**[1] of the Z_i's.

- **(a):** Assuming $p \in [0, 1/4]$, prove that $\Pr[M \geq t] \leq \exp(-n/100)$.
- **(b):** **OPTIONAL:** Assuming $p \in [0, 1/4]$, prove that $\Pr[M \geq t] \leq p^{n/c}$ for some constant $c > 1$.

[1] A median is a value M such that $|\{ i : Z_i \geq M \}| \geq n/2$ and $|\{ i : Z_i \leq M \}| \geq n/2$. If n is odd then M is unique so we can say "the median", but if n is even then it need not be unique and we should say "a median".