E0234: Assignment 5 Solutions

It is highly recommended you do not google for the answers to the questions below. You can discuss with your friends, but then mention that in your submission. The writing should solely be your own.

1. Let $X_1, \ldots, X_n$ be independent indicator random variables, and let $p_i := \Pr[X_i = 1]$. Let U and L be any positive reals such that $U \geq \sum_{i=1}^{n} p_i \geq L$. Let $X = \sum_{i=1}^{n} X_i$. Prove

   (a) $\Pr\left[X \geq (1+\delta)U\right] \leq e^{-U((1+\delta)\ln(1+\delta)-\delta)}$

   (b) $\Pr\left[X \leq (1-\delta)L\right] \leq e^{-L\delta^2/2}$

   Solution sketch:

   Follows from the proof of Theorem 4.2 and 4.2 in MR.

2. We are given a collection of n sets $S_1, \ldots, S_n$ where each S_i is a subset of $U = \{1, 2, \ldots, n\}$. A 2-coloring is an assignment $c : U \to \{-1, +1\}$ to the elements of U. The *discrepancy* of a 2-coloring is $\max_{j=1}^{n} \sum_{i \in S_j} c(i)$. What can you say about the discrepancy of a random coloring which independently assigns each element $+1$ or -1 with probability $1/2$? In particular, give a confidence bound (in terms of n) in which the discrepancy lies with probability $> 99\%$.

   Solution sketch:

   For $\{1, -1\}$ random variable, we get the following from the Chernoff bound.

$$\Pr[c(S_i) \geq \Delta] \leq e^{-\frac{\Delta^2}{|S_i|}} \leq e^{-\frac{\Delta^2}{n}}$$

   Using union bound, we have the following.

$$\Pr[\exists i, |c(S_i)| \geq \Delta] \leq ne^{-\frac{\Delta^2}{n}}$$

   Hence, $\max_{j=1}^{n} \sum_{i \in S_j} c(i)$ is at most $O(\sqrt{n \log n})$ within 99% probability.

3. (a) If Z is a random variable with $\mathbf{Exp}[Z] = 0$ and $|Z| \leq 1$ with probability 1, then prove Z is 1-subgaussian.

   Solution sketch:

$$
\begin{aligned}
\mathbf{Exp}[e^{tZ}] &= \mathbf{Exp}[1 + tZ + \frac{t^2Z^2}{2!} + \dots] \\
&\leq 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \dots \\
&= e^{t^2/2}
\end{aligned}
$$

(b) If Z is a σ-subgaussian random variable, prove that $\mathbf{Exp}[Z] = 0$ and $\mathbf{Var}[Z] \leq \sigma^2$.
Solution sketch:

$$
\begin{aligned}
&\mathbf{Exp}[e^{tZ}] \leq e^{\frac{t^2\sigma^2}{2}} \\
\Rightarrow &1 + t\mathbf{Exp}[Z] + \frac{t^2}{2}\mathbf{Exp}[Z^2] + \frac{t^3}{3!}\mathbf{Exp}[Z^3] \dots \leq e^{\frac{t^2\sigma^2}{2}} \\
\Rightarrow &\mathbf{Exp}[Z] + t\mathbf{Exp}[Z^2] + \frac{t^2}{3}\mathbf{Exp}[Z^3] + \dots \leq 2te^{\frac{t^2\sigma^2}{2}} && \text{[Differentiating w.r.t } t\text{]} \\
\Rightarrow &\mathbf{Exp}[Z] \leq 0 && \text{[Putting } t = 0\text{]} \\
\Rightarrow &\mathbf{Exp}[Z] = 0 && \text{[Since Z is subgaussian, -Z is subgaussian too]}
\end{aligned}
$$

4. **Johnson-Lindenstrauss Lower Bound.** Consider the following $(n+1)$ points in n dimensions: $\{0, e_1, \dots, e_n\}$ where e_i is the ith unit vector with 1 in the ith position and 0 elsewhere. Consider a mapping of these $(n+1)$ vectors to k dimensions such that the pairwise distances are preserved to a $(1 \pm \varepsilon)$-multiplicative factor. What is the best *lower bound* you can prove on k?

   Solution sketch: See here and here.