All problems are of equal value.


2. (Multiplicative Chernoff Bound). Let $X_1, \ldots, X_n$ be independent random variables taking values over the continuous interval $[0, 1]$. Let $X = \sum_{i} X_i$. Show the following.

   (a) For $r \in (-\infty, \ln 2]$, prove that $\mathbb{E}[e^{rX}] \leq e^{r\mathbb{E}[X] + r^2 \mathbb{E}[X]}$, where you may use (without proof) that $1 + x \leq e^x$ for all $x \in \mathbb{R}$, and $e^x \leq 1 + x + x^2$ for $x \leq \ln 2$.

   (b) Explain how the above used the independence of the $X_i$.

   (c) Apply Markov’s inequality ($\Pr[Y \geq a] \leq \mathbb{E}[Y]/a$) to $e^{rX}$, and optimize over $r$, to conclude that:

      i. For $0 \leq \epsilon \leq \ln 4$, $\Pr[X \geq (1 + \epsilon)\mathbb{E}[X]] \leq e^{-\epsilon^2 \mathbb{E}[X]/4}$.
      ii. For $\epsilon \geq \ln 4$, $\Pr[X \geq (1 + \epsilon)\mathbb{E}[X]] \leq 2^{-\epsilon \mathbb{E}[X]/2}$.
      iii. For $0 \leq \epsilon \leq 1$, $\Pr[X \leq (1 - \epsilon)\mathbb{E}[X]] \leq e^{-\epsilon^2 \mathbb{E}[X]/4}$.
      iv. (Additive Chernoff Bound) For $\epsilon \geq 0$, $\Pr[|X - \mathbb{E}[X]| \geq \epsilon \cdot n] \leq 2e^{-\epsilon^2 n/4}$.

   Note: The additive Chernoff bound suffices for applications such as estimating the errors in polling, but the multiplicative bound is in general stronger and often needed (e.g. consider $\mathbb{E}[X] = \lg n$ and the resulting bound for $\Pr[X \geq 2\mathbb{E}[X]]$). Note also that the above omits one range of parameters, where one can show that $\Pr[X \geq (1 + \epsilon)\mathbb{E}[X]] \leq e^{-(1+\epsilon)\ln(1+\epsilon)\mathbb{E}[X]/4}$ if $\epsilon \geq 1$.

3. Let $G$ be an undirected graph $G = (V, E)$ with $n$ vertices and $m$ edges. We wish to partition the vertex set $V$ into $k$ disjoint sets $V = V_1 \cup V_2 \cup \cdots \cup V_k$, while minimizing the total number of conflicting vertices, where two vertices are in conflict if they are adjacent in $G$ and belong to the same set in the partition.

   (a) Give an $O(n)$ time randomized algorithm that outputs a partition, where the expected number of conflicts is at most $\frac{m}{k}$.

   (b) Give an $O(n + m)$ time deterministic algorithm that outputs a partition, where the number of conflicts is at most $\frac{m}{k}$.

   Hint: Use a greedy approach.