Fall 2014 Randomized Algorithms	Deadline: Oct 22, 2014

Problem Set 2

Prof. Friedrich Eisenbrand — *Assistant: Chidambaram*

1. Let X be a random variable with mean μ and variance σ^2. The statement $\mathbf{P}[|X-\mu| \geq a] \leq \sigma^2/a^2$ is called Chebyshev's inequality and is easily derived from Markov's inequality. Prove the "one-sided" version of it, i.e.,

$$\mathbf{P}[X \geq \mu + a] \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

[4 pts]

2. Consider a collection of $X_1, \ldots, X_n$ of n independent integers chosen uniformly from the set $\{0, 1, 2\}$. Let $X = \sum_{i=1}^{n} X_i$ and $0 < \delta < 1$. Derive a Chernoff bound for $\mathbf{P}[X \geq (1+\delta)n]$ and $\mathbf{P}[X \leq (1-\delta)n]$. **[4 pts]**

3. We can often make better estimates on the expected value of a random variable when we have access to higher moment information. In the first part of the exercise we will obtain good estimates on the absolute value of any random variable using the fourth moment. In the second part we will use this to prove a discrepancy result.

   (a) Using the information that for all $x \geq 0$ and $q > 0$,

   $$x \geq \frac{3\sqrt{3}}{2\sqrt{q}}\left(x^2 - \frac{x^4}{q}\right),$$

   prove that for any random variable S,

   $$\mathbf{E}[|S|] \geq \frac{(\mathbf{E}[S^2])^{3/2}}{(\mathbf{E}[S^4])^{1/2}}.$$

   (b) Let $X_1, \ldots, X_n$ be a collection of 4-wise independent random variables taking -1 or 1 with equal probability. Let $\epsilon_1, \ldots, \epsilon_n \in \mathbb{R}$. Show that if $S = \epsilon_1 X_1 + \cdots + \epsilon_n X_n$ then

   $$\sqrt{\frac{\epsilon_1^2 + \ldots \epsilon_n^2}{3}} \leq \mathbf{E}[|S|] \leq \sqrt{\epsilon_1^2 + \ldots \epsilon_n^2}.$$

[6 pts]

4. Suppose that we can obtain independent samples $X_1, X_2, \ldots$ of a random variable X and that we want to use these samples to estimate $\mathbf{E}[X]$. Using

t samples, we use $\left(\sum_{i=1}^{t} X_i\right)/t$ for our estimate of $\mathbf{E}[X]$. We want the estimate to be within $\epsilon\mathbf{E}[X]$ from the true value of $\mathbf{E}[X]$ with probability at least $1-\delta$. We may not be able to use Chernoff's bound directly to bound how good our estimate is if X is not a $0-1$ random variable (as we saw in the lecture), and we do not know its moment generating function. We develop an alternative approach that requires only having a bound on the variance of X. Let $r = \sqrt{\mathbf{Var}[X]}/\mathbf{E}[X]$.

(a) Show using Chebyshev's inequality that $O(r^2/\epsilon^2\delta)$ samples are sufficient to solve the problem.

(b) Suppose that we need only a weak estimate that is within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$ with probability at least $3/4$. Argue that $O(r^2/\epsilon^2)$ samples are enough for this weak estimate.

(c) Show that, by taking the median of $O(\log(1/\delta))$weak estimates, we can obtain an estimate within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$ with probability at least $1-\delta$. Conclude that we need only $O((r^2\log(1/\delta)))/\epsilon^2)$.

[6 pts]

5. Consider a set S of n points on a line. Pick a point uniformly at random from S. Repeat this procedure at total of r times and let R be the set of points that was picked at least once. Show that with constant probability every interval of points in S defined by consecutive points in R is of size at most $O(n\log r/r)$. **[5 pts]**