TTIC/CMSC 31150 Mathematical Toolkit **Spring 2013**

Homework 1

Madhur Tulsiani Due: May 13, 2013

Note: *You may discuss these problems in groups. However, you must write up your own solutions and mention the names of the people in your group. Also, please do mention any books or papers you refer to.*

1. **Pairwise-Independence.** Recall that for computing the variance of a random variable $Z = X_1 + \cdots + X_n$, we could say that $\mathsf{Var}[Z] = \mathsf{Var}[X_1] + \cdots + \mathsf{Var}[X_n]$ as long as *each pair* of random variables X_i, X_j were independent. Such a collection of random variables is called pairwise independent. Here we give a way of generating $2^t - 1$ pairwise independent bits from t completely independent bits. This is often very useful in de-randomizing algorithms.

   Let $X_1, \ldots, X_t$ be independent random variables, each of which is 0 with probability 1/2 and 1 with probability 1/2. Let $[t]$ denote the set $\{1, \ldots, t\}$. For each non-empty subset S of $[t]$, define the random variable Y_S as

   $$Y_S \; = \; \sum_{i \in S} X_i \mod 2 \, .$$

   (a) Show that the $2^t - 1$ random variables $\{Y_S\}_{S \subseteq [t], S \neq \emptyset}$ are pairwise independent.

   (b) Show that the collection of $\{Y_S\}_{S \subseteq [t], S \neq \emptyset}$ is *not* 3-wise independent i.e., there exist three sets S_1, S_2 and S_3 such that Y_{S_1}, Y_{S_2} and Y_{S_3} are not mutually independent.

   (c) Do the variables $\{Y_S\}_{S \subseteq [t], S \neq \emptyset}$ remain pairwise independent if each X_i is 1 with probability p and 0 with probability $1 - p$ for some $p \neq 1/2$?

2. **Better than Chebyshev?** Recall that for a real-valued random variable X with mean μ and variance σ^2, Chebyshev's inequality shows that

   $$\mathbb{P}[|X - \mu| \geq c] \; \leq \; \frac{\sigma^2}{c^2} \, .$$

   Note that the above bound does not say anything when $c \leq \sigma$. Prove the following one-sided variant of Chebyshev's inequality for any real-valued random variable with mean μ and variance σ^2:

   $$\mathbb{P}[X - \mu \geq c] \; \leq \; \frac{\sigma^2}{c^2 + \sigma^2} \, .$$

   Note that this bound is meaningful even when $c \in [0, \sigma]$.
   (**Hint**: Bound the probability that $\mathbb{P}[X + t - \mu \geq c + t]$ and optimize the bound over t.)

3. **Dominating sets.** Given a graph $G = (V, E)$ and a set $U \subseteq V$, a set S is said to be a *dominating set* for U, if for each $i \in U$, S contains i or some neighbor of i. For a graph G with n vertices, let U be a subset of vertices such that all vertices in U have degree at least d. Show that there exists a dominating set for U of size at most $n \cdot \left(\frac{1+\ln(d+1)}{(d+1)}\right)$.

4. **Poisson Random Variables.** A Poisson random variable with parameter λ is a random variable X which takes values in $\{0\} \cup \mathbb{N}$ with probabilities

$$\mathbb{P}[X = i] \;=\; e^{-\lambda} \cdot \frac{\lambda^i}{i!}\,.$$

   (a) Let X be a Poisson random variable with parameter λ. Show that $\mathbb{E}[X] = \lambda$ and $\mathsf{Var}[X] = \lambda$.

   (b) Let $Z = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are independent Poisson random variables with parameter λ. Show that Z is also a Poisson random variable with parameter $n\lambda$.

   (c) Show that for $\varepsilon \in (0,1)$, we have

$$\mathbb{P}[Z \geq (1+\varepsilon)\cdot n\lambda] \;\leq\; e^{-\varepsilon^2 \cdot n\lambda/2}\,.$$

   (d) Show that this gives a tail inequality for *any* Poisson random variable. In particular, for any Poisson random variable with parameter λ, and $\varepsilon \in (0,1)$, we have that $\mathbb{P}[Z \geq (1+\varepsilon)\cdot \lambda] \;\leq\; e^{-\varepsilon^2 \cdot \lambda/2}$.

5. **More on the power of two choices.**

   (a) We discussed in class that if n balls are thrown into n bins such that each ball probes two random bins and goes into the least loaded one, then the maximum load is $\frac{\ln\ln n}{\ln 2} + O(1)$ with high probability. We defined a sequence β_i such that the number of bins with at least i balls was at most β_i with high probability.

   We will now show that the maximum load is *at least* $\frac{\ln\ln n}{\ln 2} - O(1)$ with high probability. Consider first throwing $n/2$ balls, then another $n/4$, then another $n/8$ and so on. Define a sequence α_i with the property that after throwing $n \cdot (1 - 1/2^i)$ balls, the number of bins with load at least i is at least α_i with high probability (you will need to prove this for your choice of α_i). Use this to show that the maximum load is at least $\frac{\ln\ln n}{\ln 2} - O(1)$ with high probability.

   (b) Suppose that there is a bug in the implementation of the load balancing scheme with two choices such that at each step, after probing two random bins, the ball is put into the one that is *more loaded*. Show that if we throw n balls this way, the maximum load is still $O\left(\frac{\log n}{\log\log n}\right)$ with high probability.