Problem Set 5

Hash Functions, Sub-Guassian Random Variables, Heoffding Bound

1. For any family of hash functions from X to Y where $|X| = m$, $|Y| = n$, show that for h chosen at random from the family, there exists a pair of elements $x, x' \in X$ such that

$$\mathbf{Pr}[h(x) = h(x')] \geq \frac{1}{n} - \frac{1}{m},$$

irrespective of the distribution according to which h is chosen.

2. (a) Let X and Y be numbers that are chosen independently and uniformly at random from $\mathbb{Z}_{n+1}$. Let $Z = X + Y \bmod (n+1)$. Show that X, Y, Z are pairwise independent but not independent.

   (b) Extend this example to give a collection of random variables that are k-wise independent but not $(k+1)$-wise independent.

3. Suppose we are given m vectors $\vec{v}_1, \ldots, \vec{v}_m \in \{0,1\}^\ell$ such that any k of the m vectors are linearly independent modulo 2. Let $\vec{v}_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,\ell})$. Let $\vec{u}$ be chosen uniformly at random from $\{0,1\}^\ell$ and let $X_i = \sum_{i=1}^{\ell} v_{i,j} u_j \bmod 2$. Show that the X_i are uniform k-wise independent bits.

4. Suppose that Alice and Bob secretly agree on a hash function h from a 2-universal family $\mathcal{H} = \{h : \mathcal{M} \to \mathbb{Z}_p\}$ of hash functions, where p is a prime. Later, Alice sends a message m to Bob over the Internet, where $m \in \mathcal{M}$. She authenticates this message to Bob by also sending an authentication tag $t = h(m)$, and Bob checks that the pair (m, t) he receives satisfies $t = h(m)$. Suppose that an adversary intercepts (m, t) en route and tries to fool Bob by replacing the pair with a different pair (m', t'). Argue that the probability that the adversary succeeds in fooling Bob into accepting (m', t') is at most $1/p$, no matter how much computing power the adversary has, even if the adversary knows the family $\mathcal{H}$ of hash functions used.

5. Prove the following variants of Hoeffding bound.

   (a) Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbf{E}[X_i] = \mu_i$ and $\mathbf{Pr}[a_i \leq X_i \leq b_i] = 1$ (i.e., X_i is a bounded random variable with lower bound a_i and upper bound b_i) where a_i, b_i are constants. Then

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq \epsilon\right] \leq 2e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i - a_i)^2}.$$

   Hint: Use the Heoffding inequality discussed in class and the fact that bounded random variables are sub-Gaussian (show this!).

   (b) Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbf{E}[X_i] = \mu$ for all $i \in [n]$ and $\mathbf{Pr}[a \leq X_i \leq b] = 1$. Then

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$