Algorithms in the Real World (15-853), Fall 06
Assignment #5 **Due: 15 Nov 06**

You can look up material on the web and books, but you cannot look up solutions to the given problems. You can work in groups, but must write up the answers individually, and must write down your collaborators' names.

Problem 1: Polynomials (15pts)
The fundamental theorem of polynomials states that any polynomial of degree d over any field $\mathbb{F}$ has at most d roots. (Note: whenever we say "degree d", we mean "degree *at most* d".)

a. Use the theorem to prove that if two different degree-d polynomials $P(x)$ and $Q(x)$ satisfy $P(x_i) = Q(x_i)$ for $d+1$ *distinct* points $x_1, x_2, \ldots, x_{d+1}$, then $P(x) = Q(x)$ for all $x \in \mathbb{F}$ (and not just on these $d+1$ points).

Hence, this proves the claim that *there is <u>at most</u> one degree d polynomial passing through $d+1$ points* (x_i, y_i). (I.e., at most one degree-d polynomial satisfying $P(x_i) = y_i$ for $i = 1, 2, \ldots, d+1$.)

b. Suppose we are given $d+1$ pairs $(x_i, y_i) \in \mathbb{F}^2$, and we want to figure out a degree-d polynomial $P(x)$ satisfying $P(x_i) = y_i$. We guess $P(x) = c_0 + c_1 x^1 + \cdots + c_d x^d$, and treating the $d+1$ coefficients c_i as *variables*, we write down $d+1$ linear equalities $P(x_i) = y_i$.

This gives us $d+1$ linear equalities in $d+1$ variables (the coefficients c_i), which we can potentially solve to get the values for the c_i's.

E.g., If we are given the 3 points $(1,1), (2,2), (3,2)$ and want to find a degree-2 polynomial $P(x)$ passing through these points. Write down these 3 linear equalities, and solve then to find the polynomial $P(x)$.

c. In general, the $d+1$ equalities in $d+1$ variables give us a linear system $A\vec{c} = \vec{y}$. Hence we can solve for the coefficients by $\vec{c} = A^{-1}\vec{y}$, *as long as the matrix A is invertible.* Show that the matrix A generated by the process above is *always* invertible.

d. Hence, infer that there is *exactly one* degree-d polynomial passing through $d+1$ points.

Do one of the following two problems (2 and 3).

Chernoff Bounds. A convenient form of the Chernoff-Hoeffding large deviation bounds states the following:

Let $X_1, X_2, \ldots, X_t$ be independent random variables with each $X_i \in [0,1]$. Let $X = \sum_i X_i$, and $\mu = E[X]$. Then for any $\beta \geq 0$,

$$\mathbf{Pr}[X \geq (1+\beta)\mu] \leq e^{-\frac{\beta^2}{2+\beta}\mu} \tag{1}$$

$$\mathbf{Pr}[X \leq (1-\beta)\mu] \leq e^{-\frac{\beta^2}{3}\mu} \tag{2}$$

We often call it a "Chernoff bound" for brevity.

Problem 2: Chernoff Bounds (20pt)
Consider n independent and identical $\{0,1\}$ random variables X_i with $\mathbf{Pr}[X_i = 1] = 1/n = 1 - \mathbf{Pr}[X_i = 0]$.

a. If $X = \sum_i X_i$, what is $\mu = E[X]$? What is the variance σ^2 of the random variable X?

b. Use Chebychev's inequality

$$\mathbf{Pr}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2} \tag{3}$$

to find the lowest value λ_1 such that $\mathbf{Pr}[X > \mu + \lambda_1] \leq 1/2n$?

c. Now use Chernoff bounds to give the lowest value λ_2 you can find such that $\mathbf{Pr}[X > \mu + \lambda_2] \leq 1/2n$? You should compare this answer with the bound λ_1.

d. Consider throwing n balls into n bins independently and uniformly, so that each ball has a $1/n$ chance of falling into any fixed bin. Use the analysis above, and the trivial union bound (which says that for any events $\mathcal{B}_j$, $\mathbf{Pr}[\cup_j \mathcal{B}_j] \leq \sum_j \mathbf{Pr}[\mathcal{B}_j]$) to argue that with probability at least $1/2$, every bin has at most $O(\log n)$ balls in it.

(This shows that if we use a random hash function and hash n keys into a table of size n, with probability at least 50%, we will have at most $O(\log n)$ keys hashing into any location.

Problem 3: Deriving Your Own Bounds (20pt)
Suppose you are now given independent and identical random variables $Y_1, Y_2, \ldots, Y_t$ with $Y = \sum_i Y_i$, and each r.v. $Y_i \sim N(0,1)$ is distributed as a standard Normal (a.k.a. Gaussian). In this question, we want to derive a Chernoff bound for the sum of Gaussians.

a. Recall that if $Z_1 \sim N(\mu_1, \sigma_1^2)$ and $Z_2 \sim N(\mu_2, \sigma_2^2)$, then $Z_1 + Z_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. The density function of the normal distribution $N(\mu, \sigma^2)$ is

$$f(z) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4}$$

Use the above facts to give the best bound you can on $\mathbf{Pr}[Y \geq \lambda]$ for $\lambda > 0$, where $Y_i \sim N(0,1)$ and $Y = \sum_{i=1}^{t} Y_i$.

b. If $Y_i \sim N(0,1)$, then derive the moment generating function (mgf) $E[e^{sY_i}]$ for the r.v. Y_i.

c. In class, you saw how the derivation of a Chernoff bound is fairly standard (use Markov's inequality on $\mathbf{Pr}[e^{sY} \geq e^{s\lambda}]$, etc.) up to the point where you plug in the mgf, after which you try to optimize the parameter $s > 0$ to obtain the tightest upper bound you can get on $\mathbf{Pr}[Y \geq \lambda]$.

Use the same idea to derive a Chernoff bound for the sum of n independent Gaussians. How does your Chernoff bound compare to the bound from the first subpart of this question?