MATH 341 Homework 10

David Moore

November 20, 2009

4 Hagerup, T. and Rb, C. 1990. A guided tour of Chernoff bounds. *Information Processing Letters* 33, 6 (Feb. 1990), 305-308.

This expository paper provides accessible derivations of "the various inequalities collectively known as Chernoff bounds." These are bounds on the probability that a sum of independent random 0-1 indicator variables will take a value arbitrarily far from its expected value. Intuitively we can think of this as the probability of tossing a fair coin 100 times and producing only 10 heads (or 90 heads). The paper gives bounds in several different forms, under varying assumptions (that the variables are iid, for example), but the most general bound is that for any $S = X_1 + X_2 + \cdots + X_n$ where the X_i are independent random 0-1 variables (not necessarily identically distributed, so each X_i is 1 with probability p_i and 0 otherwise), then $\mathbb{P}(S \geq (1+\varepsilon)m) \leq e^{-\varepsilon^2 m/3}$, where $m = p_1 + p_2 + \cdots + p_n$ is the expected value of S. In other words, the probability of a sum of random variables being greater by some distance ε than its expected value decreases *exponentially* as the mean, which correlates roughly with the number of trials, increases (the paper also gives analagous results for the lower tail). This is a much stronger constraint than that given by, for example, Chebyshev's inequality, which though more generally applicable only guarantees a polynomial decrease.

The derivation works in part by writing $\mathbb{P}(S \geq (1+\varepsilon)m)$ in terms of the moment generating function of S, using Markov's inequality ($\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$) to derive $\mathbb{P}(S \geq (1+\varepsilon)m) \leq e^{-t(1+\varepsilon)m}\mathbb{E}[e^{tS}]$. This general bound holds no matter how S is distributed; if we then use the fact that S is a sum of random indicator variables to calculate $\mathbb{E}[e^{tS}] = e^{m(e^t-1)}$, and take $t = \ln(1+\varepsilon)$, we can derive the particular bound given above.

One application of these bounds is in computer science, in the analysis of probabilistic algorithms. Suppose we have an algorithm for some decision problem (i.e., the algorithm produces a "yes" or "no" answer) which gives a correct answer only slightly more than half of the time. We expect that if we run the algorithm many, many, times, it will become increasingly likely that the most common answer is in fact the correct answer. The Chernoff bound tells us that as long as the probability of an incorrect answer is less than $\frac{1}{2}$ on any particular trial, then the probability that the majority of trials will give incorrect answers decreases exponentially with the number of trials. We can use this to calculate the number of times we would need to run the algorithm to have an arbitrary degree of confidence in the result.