6.856
Problem Set 9
Nov. 5th, 2002

1. Run k times and take the median. For the median to be out of range, $k/2$ estimates must have deviated by more than ϵn. In expectation, only $X = k/4$ estimates are out of range. So by Chernoff, the median is out of range with probability less than: $Pr[X > k/2] < e^{-k/12} < \delta \Rightarrow k = O(\log \frac{1}{\delta})$.

2. Our scheme will be to take the mean of n samples and repeat this k times. At the end, take the median of the k means. Let X be the sum of n independent samples from this distribution, μ be the mean, and σ be the standard deviation. The independence allows us to sum the variances so that $\sigma_X = \sigma\sqrt{n}$. Chebyshev gives $Pr[|X - n\mu| > t\sigma\sqrt{n}] < \frac{1}{t^2} \Rightarrow Pr[|X - n\mu| > 2\mu\sqrt{n}] < \frac{1}{2^2} = \frac{1}{4}$ for t=2, since $\mu > \sigma$. Thus we have $Pr[|X/n - \mu| > \frac{2}{\sqrt{n}}\mu] < \frac{1}{4}$, which will be an ϵ approximation if we have $2/\sqrt{n} < \epsilon \Rightarrow n = \text{poly}(\frac{1}{\epsilon})$ samples. Now we use this as the blackbox algorithm of problem 1 to get an error probability of $< \delta$ by taking the median of $k = O(\log \frac{1}{\delta})$ estimates. Thus, the total number of samples we need is $O(poly(\frac{1}{\epsilon}) \log \frac{1}{\delta})$.

3. Everything is as in the original Cohen algorithm, except that the rule for deleting an edge is no longer valid. But we modify it so that everything continues to work as follows. Maintain d separate counts for each edge. If the edge is traversed as the i-th edge from the source of a DFS, increment the i-th count. Delete the edge once we have $\mu_{\epsilon\frac{\delta}{n}}$ for each of the d counts. By this rule, any vertex at distance $\leq d - 1$ from the terminal vertex of that edge will receive $\mu_{\epsilon\frac{\delta}{n}}$ hits before the edge is cut. Any vertex further away doesn't care, because no message can reach it by traversing the edge in question anyways. Thus the algorithm is correct with this modification, and we merely have a constant d factor blowup in running time, giving us the original asymptotic running time as desired.