Problem Set 3

Randomized Algorithms

Due Friday, September 20

1. [Karger.] Consider a sequence of n unbiased coin flips, and look at the length of the longest contiguous sequence of heads.

   (a) Show that you are unlikely to see a sequence of length $c + \log_2 n$ for $c > 1$ (give a decreasing bound as a function of c).

   (b) Show that with high probability you will see a sequence of length $\log_2 n - O(\log_2 \log_2 n)$. Note: this observation can be used to detect cheating. When told to fake a random sequence of coin tosses, most humans will avoid creating runs of this length under the mistaken assumption that they don't look random.

2. Suppose we have sample access to a random variable X with unknown mean μ and variance σ^2, and we want to find estimate μ to within $\pm\epsilon\sigma$. That is, we sample $x_1, \ldots, x_n \sim X$ and want to use the samples to produce an estimate $\widehat{\mu}$ of μ such that

$$|\widehat{\mu} - \mu| \leq \epsilon\sigma$$

   with probability $1 - \delta$.

   (a) Show that the sample mean $\overline{x} = \frac{1}{n}(\sum_i x_i)$ works as long as $n > O(\frac{1}{\epsilon^2\delta})$.

   (b) Show that this is optimal: for every setting of ϵ and δ, there exists a distribution X for which the sample mean requires $\Omega(\frac{1}{\epsilon^2\delta})$ samples to be $\epsilon\sigma$-close with probability $1 - \delta$. **Hint (rot13):** K pna or ovanel.

(c) Now consider the median-of-means estimator: we split n into $m = O(\log \frac{1}{\delta})$ blocks, each of length $B = O(1/\epsilon^2)$, and output

$$\widehat{\mu} := \operatorname*{median}_{i \in [m]} \operatorname*{mean}_{j \in [B]} x_{B(i-1)+j}.$$

Show that this satisfies the guarantee with $n = O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$.

(d) Now, suppose that our sample $x_1, \ldots, x_n$ were not independent, but negatively associated. Would the same result hold?

3. Let X_1, X_2 be independent but not necessarily identically distributed random variables. Let σ_1, σ_2 be a random permutation on $\{1, 2\}$. Are the variables $Y_i = X_{\sigma_i}$ negatively associated?