Problem Set 2

Randomized Algorithms

Due Wednesday, September 30

1. [Wainwright.] Let $X_1, \ldots, X_n \sim N(0,1)$ for some $n \geq 10$, and let $Z = \max_i X_i$.

   (a) Show that
   $$\frac{1}{2} \leq \frac{\mathbb{E}[Z]}{\sqrt{\log n}} \leq 3$$

   (b) Show that
   $$\frac{\mathbb{E}[Z]}{\sqrt{2 \log n}} = 1 - o(1)$$
   as $n \to \infty$.

2. [Wainwright.] Let X_1 and X_2 be zero-mean subgaussians with parameters σ_1 and σ_2, respectively.

   (a) Show that if X_1 and X_2 are independent, then $X_1 X_2$ is subgamma. What are the parameters in terms of σ_1 and σ_2?

   (b) Show that in general, without assuming independence, $X_1 + X_2$ is subgaussians with parameter $2\sqrt{\sigma_1^2 + \sigma_2^2}$.

3. [Karger.] Consider a sequence of n unbiased coin flips. Consider 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 dont look random.

4. Negative Association.

   (a) Let $X_1, \ldots, X_n$ be independent but not necessarily identically distributed random variables. Let $\sigma_1, \ldots, \sigma_n$ be drawn from a permutation distribution on $[n]$. Are the variables $Y_i = X_{\sigma_i}$ negatively associated?

   (b) Recall the following algorithm from class for estimating the mean of an unknown random variable X with mean μ and variance σ^2. Given $n = mB$ samples $x_1, \ldots, x_n$, choose $m = O(\log(1/\delta))$ blocks of size $O(1/\epsilon^2)$. Output

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

   We showed that the result is within $\epsilon\sigma$ of μ with probability $1-\delta$. Now, suppose that our sample $x_1, \ldots, x_n$ were not independent, but negatively associated. Would the same result hold?

5. [Karger.] In class we proved that the two-choices approach improves the maximum load to $O(\log\log n)$. A generalization is that choosing the least loaded of d choices reduces the maximum load to $O(\log_d \log n)$. Explain what changes to the proof are needed to derive this result. Give only the diffs; do not bother writing a complete proof.

6. Consider events $E_1, \ldots, E_n$ and $Q_1, \ldots, Q_n$ such that

   $$\Pr[Q_1] = 1$$
   $$\Pr[E_i] \leq p \text{ for all } i$$
   $$\Pr[\overline{Q}_{i+1} \mid Q_i] \leq \Pr[E_{i+1} \mid Q_i] \text{ for all } i$$

   Show that

   $$\Pr[\overline{Q}_i] \leq np \text{ for all } i.$$