Problem Set 5

Randomized Algorithms

Due Tuesday, November 14

1. 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. 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. Suppose you can sample elements $x_1, \ldots, x_m$ from a mean zero subgaussian variable X, and would like to compute its *kurtosis*
   $$\kappa = \frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]^2}.$$

   (a) Consider the algorithm that computes the empirical fourth and second moments $\frac{\sum x_i^4}{m}$ and $\frac{\sum x_i^2}{m}$, and outputs the former divided by the latter squared. How large must m be for this to be an (ϵ, δ) approximation to κ?

(b) Construct a different algorithm that achieves an (ϵ, δ) approximation to κ with fewer measurements.

4. For a hash family $\mathcal{H}$ from $[U]$ to $[n]$, and a set of items $S \subset [U]$, let $X(\mathcal{H}, S)$ be the random variable denoting the load in the first bin:

$$X := |\{i \in S \mid h(i) = 1\}|$$

as a distribution over $h \in \mathcal{H}$. Further, let $f(\mathcal{H}, S)$ denote the expected max load in any bin:

$$f(\mathcal{H}, S) := \mathop{\mathbb{E}}_{h \in \mathcal{H}} \max_{j \in [n]} |\{i \in S \mid h(i) = j\}|.$$

(a) For any $t \geq 1$, and for any k-wise independent hash family $\mathcal{H}$ with $k = O(1)$, and any set S with $|S| = n$, show that

$$\Pr[X \geq t] \lesssim 1/t^k.$$

Hint: bound $\mathbb{E}[X^k]$.

(b) Show that for a k-wise independent family $\mathcal{H}$, $k = O(1)$, that

$$f(\mathcal{H}, S) \lesssim n^{1/k}$$

for any S with $|S| = n$.

(c) Show that there exists a pairwise independent hash family $\mathcal{H}$ and set S with $|S| = n$ such that

$$f(\mathcal{H}, S) \gtrsim \sqrt{n}.$$