Problem Set 5

Sublinear Algorithms

Due Tuesday, October 13

1. Let X_1 and X_2 be mean zero subgaussian random variables with variance proxies σ_1^2 and σ_2^2 respectively.

   (a) Show that $X_1 + X_2$ is subgaussian with variance proxy $4\max(\sigma_1^2, \sigma_2^2)$, regardless of whether X_1 and X_2 are independent.

   (b) If X_1 and X_2 are independent, show that X_1X_2 is subgamma and specify the parameters in terms of σ_1^2 and σ_2^2.

2. You have a collection of n biased coins. Coin i comes up heads with probability $p_i > 0$. You go through the coins, one by one, and flip each coin until it comes up heads. The goal of this problem is to analyze the total number of coin flips T until you finish.

   (a) Let T_i be the number of times you flip coin i. What are the mean and variance of T_i?

   (b) Show that T_i is subgamma$(O(1/p_i^2), O(1/p_i))$, where the first argument is the variance proxy and the second is the exponential scale.

   (c) Show that T is subgamma, and give a bound on its parameters.

   (d) Suppose that $p_i = 1/i^2$. Give a bound on $\Pr[T > 10n]$.

3. The *Morris counter* is a randomized algorithm for approximate counting in small space. Given a stream of unknown length N, we would like to estimate N in $o(\log N)$ space. The Morris counter works as follows:

   Let $a \leq 1$ be a parameter to be determined later. We keep a single counter X initialized to zero. On each successive item, we increment X with probability $(1+a)^{-X}$. The space required for this counter is of course $\log X$. Our estimate of N at the end is $\widehat{N} := (1+a)^X/a - \frac{1}{a}$.

   (a) Define Z_i to be the number of items, under an infinitely long stream, for which $X = i$. Show that
   $$\mathbb{E}[Z_i] = (1+a)^i$$
   and that Z_i is subgamma with good parameters.

   (b) Let T_t be the number of updates until $X > t$. What is $\mathbb{E}[T_t]$? Show that T_i is subgamma$(O((1+a)^{2t}/a), (1+a)^t)$.

(c) Show that, for $t \geq 1/a$ and $a < \epsilon^2/\log(2/\delta)$,

$$|T_t - \mathbb{E}[T_t]| \lesssim \epsilon\,\mathbb{E}[T_t]$$

with probability $1 - \delta$.

(d) Now consider the Morris counter. Note that the final value of the counter X is distributed as the minimum t such that $T_t \geq N$. Show that, for $N \geq O(1/a)$ and $a < \epsilon^2/\log(2/\delta)$,

$$\Pr[\widehat{N} \notin (1 \pm O(\epsilon))N] \leq 2\delta.$$

(e) Show that the Morris counter, if the parameter a is chosen appropriately, with probability $1-\delta$ uses space $S = O(\log\log N + \log(1/\epsilon) + \log\log(1/\delta))$ and outputs an $\widehat{N} \in (1 \pm \epsilon)N$ for all $N \geq 2^S$.[1]

[1] For $N < 2^S$, one can keep a regular counter in the same space. (Optional) Does the Morris counter estimate such N well?