Problem 1: (12 points) Prove the following statement, which essentially says bounds on all the moments of a random variable $Z$ and a bound on its tail behavior are equivalent. You may use moment bounds you find online or in a book for common distributions (such as the exponential distribution or gaussian distribution) without proof. Recall $\|Z\|_p = (\mathbb{E}|Z|^p)^{1/p}$.

Let $Z$ be a scalar random variable. Consider the following statements:

(1a) There exists $\sigma > 0$ s.t. $\forall p \geq 1$, $\|Z\|_p \leq C_1 \sigma \sqrt{p}$.

(1b) There exists $\sigma > 0$ s.t. $\forall \lambda > 0$, $\mathbb{P}(|Z| > \lambda) \leq C_2 e^{-C'_2 \lambda^2/\sigma^2}$.

(2a) There exists $K > 0$ s.t. $\forall p \geq 1$, $\|Z\|_p \leq C_3 K p$.

(2b) There exists $K > 0$ s.t. $\forall \lambda > 0$, $\mathbb{P}(|Z| > \lambda) \leq C_4 e^{-C'_4 \lambda/K}$.

(3a) There exist $\sigma, K > 0$ s.t. $\forall p \geq 1$, $\|Z\|_p \leq C_5 (\sigma \sqrt{p} + K p)$.

(3b) There exist $\sigma, K > 0$ s.t. $\forall \lambda > 0$, $\mathbb{P}(|Z| > \lambda) \leq C_6 (e^{-C'_6 \lambda^2/\sigma^2} + e^{-C'_6 \lambda/K})$.

Then 1a is equivalent to 1b, 2a is equivalent to 2b, and 3a is equivalent to 3b, where the constants $C'_i, C'_{i+1}, C''_{i+1}$ in each case change by at most some absolute constant factor. **Hint:** It is a fact that $\mathbb{E}|Z|^p = \int_0^\infty x^p \cdot p \cdot x^{p-1} \cdot \mathbb{P}(|Z| > x) dx$. You may use this fact without proof, though feel free to try proving it yourself as an exercise (try using integration by parts).

Problem 2: (10 points) In the first lecture we analyzed the following counter algorithm which counted up to $n$ using much less than $O(\log n)$ bits of space: initialize a counter $X$ to 1, and for every increment instruction, increment $X$ with probability $1/(1 + a)^X$. By averaging many such estimators, we obtained a $(1 + \varepsilon)$-approximation to $n$ with good probability. Here we will investigate a different way to obtain a good approximation. Imagine we still initialize $X$ to 1, but we increment it with probability $1/(1 + a)^X$ instead. (Note: our estimator for $n$ would have to change from $2^X - 1$ to something else; figure out what!)

**Question:** How small must $a$ be so that our estimate $\hat{n}$ of $n$ satisfies $|n - \hat{n}| \leq \varepsilon n$ with at least $9/10$ probability when we return the output of a single estimator instead of averaging many estimators as in class? Also derive a bound $S = S(n, \varepsilon)$ on the space (in bits) so that this algorithm uses at most $S$ space with at least $9/10$ probability by the end of the $n$ increments.
Problem 3: Here we will show some space lower bounds for streaming algorithms.

(a) (3 points) Consider the approximate counting problem from lecture 1. Show that any deterministic algorithm providing a 2-approximation for this problem must use \(\Omega(\log n)\) bits of space. (credit: intended solution suggested by Vsevolod Oparin)

(b) (7 points) In class we saw that there is no deterministic algorithm using \(o(n)\) bits of space which approximates the number of distinct elements in a stream up to a factor 2, where the stream tokens come from the universe \(\{1, \ldots, n\}\). Show that there is no randomized algorithm for the distinct elements problem using \(o(n)\) bits of space which solves the problem exactly with at least \(99/100\) probability on any input. Thus, any \(o(n)\) space algorithm must be both randomized and approximate. Hint: Come up with a compression/decompression scheme (as in class) that is randomized and works with good probability for a certain class of inputs. Conclude that there is a deterministic scheme which works for many inputs.

Problem 4: Recall the AMS sketch from class for \(F_2\) moment estimation: a random \(m \times n\) matrix \(\Pi\) with entries \(\pm 1/\sqrt{m}\) is drawn for \(m = O(1/\varepsilon^2)\), and \(\|x\|_2^2\) is estimated as \(\|\Pi x\|_2^2\). Then with at least \(2/3\) probability,

\[
(1 - \varepsilon)\|x\|_2^2 \leq \|\Pi x\|_2^2 \leq (1 + \varepsilon)\|x\|_2^2.
\]

(a) (5 points) Imagine picking \(\Pi\) differently: for each \(i \in \{1, \ldots, n\}\) we pick a uniformly random number \(h_i \in \{1, \ldots, m\}\). We then set \(\Pi_{hi,i} = \pm 1\) for each \(i \in \{1, \ldots, n\}\) (the sign is chosen uniformly at random from \(\{-1, 1\}\)), and all other entries of \(\Pi\) are set to 0. This \(\Pi\) has the advantage that in turnstile streams, we can process updates in constant time. Show that using this \(\Pi\) still satisfies the conditions of Equation 1 with \(2/3\) probability for \(m = O(1/\varepsilon^2)\).

(b) (5 points) Show that the matrix \(\Pi\) from Problem 3(a) can be specified using \(O(\log n)\) bits such that Equation 1 still holds with at least \(2/3\) probability, and so that given any \(i \in \{1, \ldots, n\}\), \(\Pi_{hi,i}\) and \(h_i\) can both be calculated in constant time. You may assume that standard machine word operations take constant time (arithmetic, mod, bitwise operations, and bitshifts). Hint: Consider a hash function that does some arithmetic modulo a prime \(p\) for some choice of \(p\).

Problem 5: An incoherent matrix \(\Pi \in \mathbb{R}^{m \times n}\) is such that each column \(\Pi^i\) of \(\Pi\) has unit \(\ell_2\) norm (i.e. \(\sum_{j=1}^m (\Pi^i_j)^2 = 1\)), and the dot products \(\langle \Pi^i, \Pi^j \rangle\) are all at most \(\varepsilon\) in magnitude for \(i \neq j\). In class we will show that any incoherent matrix can be obtained from codes. In particular, if \(C = \{C_1, \ldots, C_N\}\) is a collection of \(N\) vectors each in \([q]_t\) for some positive integers \(q, t\), then define \(\alpha \in (0, 1)\) as the maximum fraction of the \(t\) coordinates for which any distinct \(C_i, C_j\) agree. In class we will show that given such a code, one can construct an incoherent matrix with \(m = qt\), \(n = N\), and \(\varepsilon = \alpha\). We will also show, that for any \(N\)
and \(0 < \alpha < 1/2\), such codes exist with \(q = O(\alpha^{-1})\), \(t = O(\alpha^{-1}\log N)\) by picking random \(C_i\). This homework problem gives codes with smaller \(m = qt\) when \(\varepsilon\) is small.

Consider the finite field \(\mathbb{F}_q\) and consider all degree at most \(d\) polynomials \(p_1, \ldots, p_N \in \mathbb{F}_q[x]\) where \(N = q^{d+1}\). Define a code \(C_1, \ldots, C_N\) with \(t = q\) where the \(j\)th entry of \(C_i\) is the evaluation of \(p_i\) on the \(j\)th element of \(\mathbb{F}_q\) (so \(C_i\) is the evaluation table of \(p_i\)).

(a) (2 points) Give a bound on \(\alpha\) in terms of \(d, q\) (recall \(t = q\)).

(b) (6 points) Recall we need \(N \geq n\) to ensure there are enough \(C_i\)’s to form our matrix \(\Pi\). Show how to choose \(d, q\) so that \(N \geq n\) and \(\alpha \leq \varepsilon\), and show what this gives (in big-Oh notation) for \(m = qt\) being the number of rows over the incoherent matrix \(\Pi\) we obtain.

(c) (2 points) How small does \(\varepsilon\) need to be as a function of \(n\) for the codes from part (b) to give smaller \(m\) than the random codes from class (which give \(m = O(\varepsilon^{-2}\log N)\))?

OPEN PROBLEM: As mentioned above, an incoherent matrix with \(m = O(\varepsilon^{-2}\log n)\) exists. The solution to this problem provides another, incomparable bound. The lower bound is \(m = \Omega(\min\{n, \varepsilon^{-2}(\log n)/\log(1/\varepsilon)\})\) [1, Section 9]. Can the gap between upper and lower bounds be closed?

Problem 6: (1 point) How much time did you spend on this problem set? If you can remember the breakdown, please report this per problem. (sum of time spent solving problem and typing up your solution)

Extra credit problem: (3 points) The Chernoff bound states that if \(X_1, \ldots, X_n\) are independent and in \([0, 1]\), and \(X = \sum_i X_i\) and \(\mu = \mathbb{E}X\), then

\[
\forall \varepsilon > 0, \quad \mathbb{P}(X > (1 + \varepsilon)\mu) < \left(\frac{\varepsilon}{(1 + \varepsilon)^{1+\varepsilon}}\right)^\mu, \quad \forall \varepsilon > 0, \quad \mathbb{P}(X < (1 - \varepsilon)\mu) < \left(\frac{e^{-\varepsilon}}{(1 - \varepsilon)^{1-\varepsilon}}\right)^\mu.
\]

Clearly also if \(\varepsilon \geq 1\) then the second inequality can be improved to

\[
\forall \varepsilon \geq 1, \quad \mathbb{P}(X < (1 - \varepsilon)\mu) = 0.
\]

What do the above imply as an upper bound on \(\|X - \mu\|_p\)? Your solution to the extra credit problem does not count toward the page limit.

References