Homework Out: April 18
Due Date: April 25, midnight
Reminder: this homework is for extra credit! No late days. The HW contains some exercises (fairly simple problems to check you are on board with the concepts; dont submit your solutions), and problems (for which you should submit your solutions, and which will be graded). Some problems have sub-parts that are exercises. For this problem set, its OK to work with others. (Groups of 2, maybe 3 max.) That being said, please think about the problems yourself before talking to others. Please cite all sources you use, and people you work with. The expectation is that you try and solve these problems yourself, rather than looking online explicitly for answers. Submissions due at beginning of class on the due date. Please check the Piazza for details on submitting your *LaTeXed* solutions.

Problems

1. **(A Counter, and the Median-of-Means Estimator.).** Here is a way of maintaining an approximate counter. (Call this the basic counter.) Start with $X \leftarrow 0$. When an element arrives, increment X by 1 with probability 2^{-X}. When queried, return $N := 2^X - 1$.

   (a) Suppose the actual count is n, show that $\mathbb{E}[N] = n$, and $\text{Var}(N) = \frac{n(n-1)}{2}$.

   Since its variance is large, average k independent basic counters $N_1, N_2, \ldots, N_k$, and output the sample average $\hat{N} := \frac{1}{k}\sum_i N_i$. Call this the k-mean counter.

   (b) (Do not submit) Show that
   $$\mathbb{P}[\hat{N} \notin (1 \pm \epsilon)n] \leq \frac{1}{2\epsilon^2 k}.$$
   Hence using $k = \frac{1}{P2\epsilon^2\delta}$ counters can make the failure probability at most δ. (Said in other words, your error is less than ϵn with confidence $1 - \delta$.) Heres a way to use only $K = O(\frac{1}{\epsilon^2}\log\frac{1}{\delta})$ counters to get the same answer (and the approach is useful in many different contexts beyond this one):

   Take a collection of $\ell = 10log\frac{1}{\delta}$ independent k_0-mean counters, where $k_0 = \frac{4}{\epsilon^2}$. Output the median M of these ℓ counters.

   (c) Prove that $\mathbb{P}[M \notin (1 \pm \epsilon)n] \leq \delta$. *Hint: what must happen for the the median to be too high? What is the chance of that?*

2. **(I Stream, You Stream.).** In data streaming model, suppose we denote frequency vector by $x = (x_1, x_2, ..., x_D) \in \mathbb{Z}^D_{\geq 0}$ where x_i counts the number of occurences of element $i \in [D]$ seen so far. We want a streaming algorithm that stores information about the stream so that when it is eventually queried with some index $q \in [D]$, returns a value $\hat{x}_q$ that is $\approx x_q$ with probability at least $1 - \delta$. We dont want to store x explicitly, we want to use less space.

   Consider the following algorithm:

   Keep a global hash function $g : [D] \to [d]$, and also d counters $C_1, C_2, \ldots, C_d$ (initially zero), each with its own hash function $h_i : [D] \to \{-1, +1\}$. If you see element $e \in [D]$, first hash it using the global hash function g to get the bucket number $g(e)$, and then update
   $$C_{g(e)} \leftarrow C_{g(e)} + h_{g(e)}(e).$$
   When faced with the query q, output $A(q) := h_{g(q)}(q)C_{g(q)}$.

   Assume that g and the h_is are all independently picked, and each hash function is itself 2-universal.

   (a) Show that $E[A(q)] = x_q$.

   (b) Show that the variance of the estimate $A(q) = \frac{1}{d}(F_2 - x_q^2) \leq F_2/d$.

   (c) Show that if we set $d = \frac{1}{\epsilon^2\delta}$, our estimate $A(q)$ satisfies
   $$\mathbb{P}[A(q) \in x_q \pm \epsilon\sqrt{F_2}] \geq 1 - \delta.$$
   Recall that $F_2 := \sum_i x_i^2 = ||x||_2^2$ and hence the error term is $\epsilon||x||_2$.

(d) Finally, consider an extension of this idea: maintain t independent copies of the above data structure. On a query for q, if the answers from the individual copies of the data structure are $A_1(q), A_2(q), ..., A_t(q)$, return the median $M(q)$ of these t answers. Show that with $t = c_1 \log \frac{1+}{\delta}$, and $d = \frac{c_2}{\epsilon^2}$, where c_1, c_2 are constants you can choose, you get $M(q) \in x_q \pm \epsilon||x||_2$ with probability $1 - \delta$.

3. **(Chernoff meets Matrices).** In Lecture 18 we mentioned a very general theorem about matrix-valued Chernoff bounds for symmetric matrices. In this problem well take the first steps towards it. Assume eigenvalues are numbered so that $\lambda_1 \geq \ldots \lambda_n$. Well prove:

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be independent symmetric $d \times t$ matrices, and $S_n = \sum_i X_)i$. Then, for any $t \geq 0$ and $\ell \in \mathbb{R}$,*

$$\mathbb{P}[\lambda_1(S_n) \geq \ell] \leq d \cdot e^{-t\ell} \cdot \prod_{i=1}^{n} \lambda_1(\mathbb{E}[e^{tX_i}]).$$

$$\mathbb{P}[\lambda_d(S_n) \leq -\ell] \leq d \cdot e^{-t\ell} \cdot \prod_{i=1}^{n} \lambda_1(\mathbb{E}[e^{-tX_i}]).$$

Recall that $\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$. You may use the following without proof:

i. $\text{tr}(A) = \sum_i \lambda_i(A)$
ii. $\lambda_i(e^A) = e^{\lambda_i(A)}$
iii. The Golden-Thompson inequality: $\text{tr}(e^{A+B}) \leq \text{tr}(e^A \cdot e^B)$.
iv. For PSD matrices A and B, $\text{tr}(AB) \leq \text{tr}(A) \cdot \text{tr}(B)$.
v. Expectations and trace commute: $\mathbb{E}[\text{tr}(A)] = \text{tr}(\mathbb{E}[A])$.

(a) Show that for any $t \geq 0$,

$$\mathbb{P}[\lambda_1(S_n) \geq \ell] \leq \mathbb{P}[\text{tr}(e^{tS_n}) \geq e^{t\ell}] \leq e^{-t\ell} \cdot \mathbb{E}[\text{tr}(s^{tS_n}].$$

(b) Show that

$$\mathbb{E}_{X_1,\ldots,X_n}[\text{tr}(e^{tS_n})] \leq \mathbb{E}_{X_1,\ldots,X_{n-1}}[\text{tr}(e^{tS_{n-1}})] \cdot \lambda_1(\mathbb{E}[e^{tX_n}]).$$

Hint: why can you use (iv) above even if X_n isn't PSD?

(c) Use the previous two parts to prove the first statement of the theorem.

(d) Use the same arguments on $(-S_n) = \sum_i(-X_i)$ to prove the other part.