CSCE 658: Randomized Algorithms

Lecture 10

Samson Zhou

CountSketch Error Analysis

- $s(i) \cdot c_a = s(i) \cdot s(i) \cdot f_i + \sum_{j \neq i,\text{ with } j:h(j)=a} s(i) \cdot s(j) \cdot f_j$
- Since $s(i) \in \{-1, +1\}$, we have $s(i) \cdot s(i) = 1$
- What is the expectation of the error term for f_i?

CountSketch Error Analysis

- $s(i) \cdot c_a = s(i) \cdot s(i) \cdot f_i + \sum_{j \neq i, \text{ with } j: h(j) = a} s(i) \cdot s(j) \cdot f_j$
- What is the expectation of the error term for f_i?
- $\mathrm{E}\left[\sum_{j \neq i, \text{ with } j: h(j) = a} s(i) \cdot s(j) \cdot f_j\right] = \Sigma_{j \neq i} \mathrm{E}\left[s(i) \cdot s(j) \cdot f_j \cdot I_{h(j) = h(i)}\right]$

CountSketch Error Analysis

- $s(i) \cdot c_a = s(i) \cdot s(i) \cdot f_i + \sum_{j \neq i, \text{ with } j:h(j)=a} s(i) \cdot s(j) \cdot f_j$
- What is the expectation of the error term for f_i?
- $\mathrm{E}\left[\sum_{j \neq i, \text{ with } j:h(j)=a} s(i) \cdot s(j) \cdot f_j\right] = 0$

CountSketch Error Analysis

- $s(i) \cdot c_a = s(i) \cdot s(i) \cdot f_i + \sum_{j \neq i, \text{ with } j : h(j) = a} s(i) \cdot s(j) \cdot f_j$
- What is the expectation of the error term for f_i?
- $\mathrm{E}\left[\sum_{j \neq i, \text{ with } j : h(j) = a} s(i) \cdot s(j) \cdot f_j\right] = 0$
- What is the variance of the error term for f_i?

CountSketch Error Analysis

- Variance is at most the 2^{nd} moment of the error term
- $\mathrm{E}\left[\left(\sum_{j \neq i,\ \text{with}\ j: h(j)=a} s(i) \cdot s(j) \cdot f_j\right)^2\right]$

CountSketch Error Analysis

- Variance is at most the 2nd moment of the error term
- $\mathrm{E}\left[\left(\sum_{j\neq i,\text{ with } j:h(j)=a} s(i)\cdot s(j)\cdot f_j\right)^2\right]=\sum_{j\neq i}\mathrm{E}\left[\left|f_j\right|^2\cdot I_{h(j)=h(i)}\right]$

CountSketch Error Analysis

- Variance is at most the 2nd moment of the error term
- $\mathrm{E}\left[\left(\sum_{j\neq i,\ \text{with}\ j:h(j)=a} s(i)\cdot s(j)\cdot f_j\right)^2\right] = \sum_{j\neq i}\mathrm{E}\left[\left|f_j\right|^2\cdot I_{h(j)=h(i)}\right]$
$$= \sum_{j\neq i}\mathrm{E}\left[I_{h(j)=h(i)}\right]\cdot\left|f_j\right|^2$$

CountSketch Error Analysis

- Variance is at most the 2nd moment of the error term
- $\mathrm{E}\left[\left(\sum_{j\neq i,\text{ with } j:h(j)=a} s(i)\cdot s(j)\cdot f_j\right)^2\right] = \sum_{j\neq i}\mathrm{E}\left[\left|f_j\right|^2 \cdot I_{h(j)=h(i)}\right]$

$$= \sum_{j\neq i}\mathrm{E}\left[I_{h(j)=h(i)}\right]\cdot\left|f_j\right|^2$$

$$= \sum_{j\neq i}\Pr[h(j)=h(i)]\cdot\left|f_j\right|^2$$

CountSketch Error Analysis

- Variance is at most the 2nd moment of the error term
- $\mathrm{E}\left[\left(\sum_{j \neq i,\text{ with } j: h(j)=a} s(i) \cdot s(j) \cdot f_j\right)^2\right] = \sum_{j \neq i} \mathrm{E}\left[\left|f_j\right|^2 \cdot I_{h(j)=h(i)}\right]$

$$= \sum_{j \neq i} \mathrm{E}\left[I_{h(j)=h(i)}\right] \cdot \left|f_j\right|^2$$

$$= \sum_{j \neq i} \Pr[h(j) = h(i)] \cdot \left|f_j\right|^2$$

$$= \sum_{j \neq i} \frac{1}{b} \cdot \left|f_j\right|^2 \leq \frac{\|f\|_2^2}{b}$$

CountSketch Error Analysis

- Variance is at most the 2nd moment of the error term
- $\mathrm{E}\left[\left(\sum_{j\neq i,\text{ with } j:h(j)=a} s(i)\cdot s(j)\cdot f_j\right)^2\right] = \sum_{j\neq i}\mathrm{E}\left[\left|f_j\right|^2\cdot I_{h(j)=h(i)}\right]$

$$= \sum_{j\neq i}\mathrm{E}\left[I_{h(j)=h(i)}\right]\cdot\left|f_j\right|^2$$

$$= \sum_{j\neq i}\Pr[h(j)=h(i)]\cdot\left|f_j\right|^2$$

$$= \sum_{j\neq i}\frac{1}{b}\cdot\left|f_j\right|^2 \leq \frac{\|f\|_2^2}{b}$$

- Set $b=\frac{81k^2}{\varepsilon^2}$, then the variance is at most $\frac{\varepsilon^2\|f\|_2^2}{81k^2}$

CountSketch Error Analysis

- Set $b = \frac{81k^2}{\varepsilon^2}$, then the variance is at most $\frac{\varepsilon^2 \|f\|_2^2}{81k^2}$

- By Chebyshev's inequality, the error for f_i is at most $\frac{\varepsilon}{3k} \|f\|_2$ with probability at least $\frac{2}{3}$

- How to ensure accuracy for all $i \in [n]$?

CountSketch Error Analysis

- By Chebyshev's inequality, the error for f_i is at most $\frac{\varepsilon}{3k}\|f\|_2$ with probability at least $\frac{2}{3}$

- How to ensure accuracy for all $i \in [n]$?

- Repeat $\ell := O(\log n)$ times to get estimates $e_1, \ldots, e_\ell$ for each $i \in [n]$ and set $\widehat{f_i} = \mathrm{median}(\mathrm{e}_1, \ldots, \mathrm{e}_\ell)$

CountSketch Error Analysis

- Claim: For all estimated frequencies $\widehat{f_i}$ by CountSketch, we have

$$f_i - \frac{\varepsilon \|f\|_2}{3k} \leq \widehat{f_i} \leq f_i + \frac{\varepsilon \|f\|_2}{3k}$$

CountSketch Summary

- CountSketch solves the L_2 heavy-hitters problem: Given a set S of m elements from $[n]$ that induces a frequency vector $f \in R^n$ and a threshold parameter $\varepsilon \in (0, 1)$, output a list that includes:
  - The items from $[n]$ that have frequency at least $\varepsilon \cdot \|f\|_2$
  - No items with frequency less than $\frac{\varepsilon}{2} \cdot \|f\|_2$

- Space usage: $O\left(\frac{1}{\varepsilon^2} \log^2 n\right)$ bits of space

L_2 Estimation

- Goal: Given a set S of m elements from $[n]$ that induces a frequency vector $f \in R^n$ and an accuracy parameter $\varepsilon \in (0,1)$, output a $(1+\varepsilon)$-approximation to $\|f\|_2$

- Find Z such that $(1-\varepsilon)\cdot\|f\|_2 \le Z \le (1+\varepsilon)\cdot\|f\|_2$

- Find Z' such that $(1-\varepsilon)\cdot\|f\|_2^2 \le Z' \le (1+\varepsilon)\cdot\|f\|_2^2$

F_2 Moment Estimation

- Goal: Find Z' such that $(1-\varepsilon) \cdot \|f\|_2^2 \leq Z' \leq (1+\varepsilon) \cdot \|f\|_2^2$

F_2 Moment Estimation

- Goal: Find Z' such that $(1-\varepsilon)\cdot\|f\|_2^2 \le Z' \le (1+\varepsilon)\cdot\|f\|_2^2$

1 7 7 7 3 7 7 1 4 1 1 1 1 5 1 1 7 1 7 5 1 7 7

F_2 Moment Estimation

- Goal: Find Z' such that $(1-\varepsilon)\cdot\|f\|_2^2 \le Z' \le (1+\varepsilon)\cdot\|f\|_2^2$

1 7 7 7 3 7 7 1 4 1 1 1 1 5 1 1 7 1 7 5 1 7 7

f_1	f_2	f_3	f_4	f_5	f_6	f_7
10	0	1	1	2	0	9

Johnson-Lindenstrauss Lemma

- Distributional Johnson-Lindenstrauss Lemma: Given $\Pi \in R^{m \times n}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$, then for any $x \in R^n$ and setting $y = \Pi x$, then with probability at least $1 - \delta$

$$(1 - \varepsilon)\|x\|_2 \leq \|y\|_2 \leq (1 + \varepsilon)\|x\|_2$$

F_2 Moment Estimation

- Algorithm: Generate $\Pi \in R^{m \times n}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}} N(0,1)$. Set $g = \Pi \cdot f$

- Whenever there is an update to a coordinate of f, update g

F_2 Moment Estimation

- Algorithm: Generate $\Pi \in R^{m\times n}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$. Set $g = \Pi \cdot f$

1 7 7 7 3 7 7 1 4 1 1 1 1 5 1 1 7 1 7 5 1 7 7

- Whenever there is an update to a coordinate of f, update g

F_2 Moment Estimation

- Algorithm: Generate $\Pi \in R^{m\times n}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$. Set $g = \Pi \cdot f$

1 7 7 7 3 7 7 1 4 1 1 1 1 5 1 1 7 1 1 7 5 1 7 7

- Whenever there is an update to a coordinate of f, update g
- $f = f + e_1$
- $f = f + e_7$
- $f = f + e_7$

F_2 Moment Estimation

- Algorithm: Generate $\Pi \in R^{m\times n}$ with $m = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ and each entry drawn from $\frac{1}{\sqrt{m}}N(0,1)$. Set $g = \Pi \cdot f$

1 7 7 7 3 7 7 1 4 1 1 1 1 5 1 1 7 1 7 5 1 7 7

- Whenever there is an update to a coordinate of f, update g
- $f = f + e_1, g = g + \Pi e_1$
- $f = f + e_7, g = g + \Pi e_7$
- $f = f + e_7, g = g + \Pi e_7$

AMS Algorithm

- Generate a random sign vector $s \in \{-1, +1\}^n$
- Maintain $Z = \langle s, f \rangle$
- Output $W := Z^2$

1

1

2

1

2

1

1

1

2

1

1

2

2

2

1

AMS Algorithm

- What values of Z did you get?
- $Z = \langle s, f \rangle = s_1 f_1 + s_2 f_2 + \cdots + s_n f_n$
- What values of W did you get?
- $W = Z^2 = \sum_{i,j} s_i s_j f_i f_j$

AMS Algorithm

- What values of W did you get?
- $W = Z^2 = \sum_{i,j} s_i s_j f_i f_j$

f_1	f_2	f_3	f_4	f_5	f_6	f_7
9	6	0	0	0	0	0

AMS Algorithm

- What is $\mathrm{E}[W]$?
- $Z = \langle s, f \rangle = s_1 f_1 + s_2 f_2 + \cdots + s_n f_n$
- $W = Z^2 = \sum_{i,j} s_i s_j f_i f_j$
- $\mathrm{E}[W] = \sum_{i,j} \mathrm{E}\left[s_i s_j f_i f_j\right] = \sum_i \mathrm{E}\left[f_i^2\right] = \|f\|_2^2$

AMS Algorithm

- What is $\mathrm{Var}[W]$?
- $Z = \langle s, f \rangle = s_1 f_1 + s_2 f_2 + \cdots + s_n f_n$
- $W^2 = Z^4 = \sum_{a,b,c,d} s_a s_b s_c s_d f_a f_b f_c f_d$
- $\mathrm{E}[W^2] = \sum_{a,b,c,d} \mathrm{E}[s_a s_b s_c s_d f_a f_b f_c f_d] = \sum_i \mathrm{E}\left[f_i^4\right] + 6 \sum_{i \neq j} \mathrm{E}\left[f_i^2 f_j^2\right] \leq 6\|f\|_2^4$

AMS Algorithm

- By Chebyshev's inequality, W will be a 9-approximation to $\|f\|_2^2$ with probability $\frac{2}{3}$

AMS Algorithm

- How to get $(1+\varepsilon)$-approximation?

- Repeat $O\left(\frac{1}{\varepsilon^2}\right)$ times and take the average

AMS Algorithm

- Space of algorithm: $O\left(\frac{1}{\varepsilon^2}\right)$ words of space or $O\left(\frac{1}{\varepsilon^2}\log m\right)$ bits of space