CPSC 536N: Randomized Algorithms 2014-15 Term 2

Lecture 14

Prof. Nick Harvey *University of British Columbia*

1 Streaming Algorithms for Estimating ℓ_2

In this lecture we return to the topic of streaming algorithms that we saw in Lecture 9. We will change the notation to be more consistent with the streaming literature.

The input is a sequence $(a_1, a_2, \ldots, a_m)$ of indices with each $a_i \in [n]$. The ***frequency vector*** is $f \in \mathbb{Z}^n$, where

$$f_j \;=\; |\{\; i \,:\, a_i = j \;\}| \;=\; \text{number of occurrences of } j \text{ in the stream.} \tag{1}$$

The goal is to output some statistic of f, such as a norm $\|f\|_p$, while using $O(\log(n) + \log(m))$ bits or words of space. For notational simplicity we will assume that $m = \text{poly}(n)$ so that $\log m = \Theta(\log n)$.

Today we will again consider the problem of estimating the ℓ_2-norm, namely $\|f\|_2 = (\sum_j f_j^2)^{1/2}$. In Lecture 9 we discussed an algorithm for estimating ℓ_2 using the Johnson-Lindenstrauss theorem. At that time we could not formally implement the algorithm in $O(\log n)$ space. Today we will give a similar algorithm that provably uses $O(\log n)$ space.

The algorithm discussed in Lecture 9 is shown in Algorithm 1; in that lecture, the entries of L were chosen to be independent Gaussians (scaled to have variance $1/t$). The issue with this approach is that storing this L takes $\Omega(tn)$ space, which is too much. Intuitively, we'd like to generate L using some sort of "pseudorandom" source, for which we can regenerate entries on demand. Our hash functions from Lecture 13 will allow us to do exactly that.

Algorithm 1: The generic streaming algorithm for estimating $\|f\|_2^2$.

```
Create the random matrix $L$ of size $t \times n$
Initialize $y \in \mathbb{R}^t$ to zero
for $i = 1, \ldots, m$ do
    Receive the symbol $j = a_i \in [n]$
    Add the $j^{\text{th}}$ column of $L$ to $y$
    ▷ Invariant: $y = Lf$
Output $\|y\|_2^2$
```

1.1 The Basic Estimate: $t = 1$

For simplicity let us start off by considering the case $t = 1$, so the matrix L becomes just a row vector. The entries of L will be generated by our hash function constructed in Lecture 13, which can be summarized as follows.

Corollary 1 *For any $n \geq 1$, there is a hash function $h = h_s : [n] \to \{-1, 1\}$, where the seed s is a uniformly random bit string of length $O(k \log n)$, such that*

$$\Pr_s\left[\, h_s(u_1) = v_1 \;\wedge\; h_s(u_k) = v_k \,\right] \;=\; 2^{-k}$$

for all distinct $u_1, \ldots, u_k \in [n]$ *and* $v_1, \ldots, v_k \in \{-1, 1\}$.

To generate the vector L, we simply pick a random seed s then define

$$L_j \;=\; h_s(j) \qquad \forall j \in [n]. \tag{2}$$

Whereas previously we used random Gaussians for the entries of L, we now use random signs (which is also helpful for saving space). Our hash function h is k-wise independent, but let us not yet specify the value of k. Later we will see what value is needed by the analysis.

The intuition behind the analysis is the same as for the Johnson-Lindenstrauss transform, discussed in Lecture 7. Recall that we stated the following algebraic property of variance.

Fact 2 *Let* $G_1, \ldots, G_d$ *be mutually independent random variables with finite variance. Let* $\sigma_1, \ldots, \sigma_d \in \mathbb{R}$ *be arbitrary. Then* $\mathrm{Var}\left[\sum_j \sigma_j G_j\right] = \sum_j \sigma_j^2 \,\mathrm{Var}\left[G_j\right]$.

The assumption of mutual independence is stronger than necessary. Corollary 8 implies that the same is true with only pairwise independence. So let us now restate the fact in a form that is useful for our purposes.

Fact 3 *Let* $L \in \mathbb{R}^n$ *be a random vector whose coordinates are pairwise independent and satisfy* $\mathrm{Var}\left[L_j\right] = 1$ *for all* j. *Let* $f \in \mathbb{R}^n$ *be arbitrary. Then* $\mathrm{Var}\left[Lf\right] = \sum_j f_j^2 = \|f\|_2^2$.

Moreover, our hash function's output is an unbiased random sign: $\Pr\left[h(i) = 1\right] = \Pr\left[h(i) = -1\right] = 1/2$. It follows that every entry of L has expectation zero, and variance equal to 1. So our vector L satisfies the hypotheses of Fact 3.

Algorithm's output is unbiased. Since we're in the case $t = 1$, the algorithm outputs the value y^2. We now claim that this value is correct in expectation.

We have just argued that the entries of L have expectation zero. By linearity, $\mathrm{E}\left[y\right] = 0$ as well. So, by Fact 3,

$$\|f\|_2^2 \;=\; \mathrm{Var}\left[y\right] \;=\; \mathrm{E}\left[y^2\right] - \underbrace{\mathrm{E}\left[y\right]^2}_{=0} \;=\; \mathrm{E}\left[y^2\right]. \tag{3}$$

1.1.1 Concentration of y^2

In order to show that y^2 provides a good estimate of $\|f\|_2^2$, we need to show that y^2 is concentrated around its expectation. The analogous step of our Johnson-Lindenstrauss analysis required no effort because already we knew that y had a Gaussian distribution. Today some more effort is necessary.

Since we don't have mutual independence we cannot use a Chernoff bound. In scenarios without much independence, a good option for showing concentration is Chebyshev's inequality, which appears in the appendix as Theorem 9. This yields

$$\Pr\left[\,|y^2 - \mathrm{E}\left[y^2\right]| \geq z\,\right] \;\leq\; \mathrm{Var}\left[y^2\right]/z^2. \tag{4}$$

The trouble is that we now have to analyze $\mathrm{Var}\left[\, y^2 \,\right]$, which is somewhat unpleasant as it involves fourth-powers of y:

$$\begin{aligned}\mathrm{Var}\left[\, y^2 \,\right] &= \mathrm{E}\left[\, y^4 \,\right] - \mathrm{E}\left[\, y^2 \,\right]^2 \;\leq\; \mathrm{E}\left[\, y^4 \,\right] \;=\; \mathrm{E}\left[\, \Big(\sum_{j \in [n]} L_j f_j \Big)^4 \,\right] \\ &= \sum_{j_1, j_2, j_3, j_4 \in [n]} \mathrm{E}\left[\, L_{j_1} L_{j_2} L_{j_3} L_{j_4} \,\right] f_{j_1} f_{j_2} f_{j_3} f_{j_4}. \end{aligned} \tag{5}$$

(Note that the indices $j_1, \ldots, j_4$ need not be distinct.) In order to apply the "expectation-of-product equals product-of-expectations" rule to (5), we now decide to set $k = 4$ so that the entries of L are 4-wise independent. $k = 4$

Claim 4 $\mathrm{Var}\left[\, y^2 \,\right] \leq 3 \left\| f \right\|_2^4$.

Plugging this and (3) into (4), we obtain

$$\Pr\left[\, |y^2 - \left\| f \right\|_2^2| \geq z \,\right] \;\leq\; 3 \left\| f \right\|_2^4 / z^2.$$

If we wanted multiplicative error of $1 + \epsilon$, we would take $z = \epsilon \left\| f \right\|_2^2$, which gives

$$\Pr\left[\; |y^2 - \left\| f \right\|_2^2| \;\geq\; \epsilon \left\| f \right\|_2^2 \;\right] \;\leq\; 3/\epsilon^2.$$

Unfortunately this is only useful when $\epsilon > \sqrt{3}$. To handle ϵ close to zero, we will need to average many such estimates which is exactly the purpose of the case $t > 1$.

1.2 The Actual Estimate: $t > 1$

We now let L be a $t \times n$ matrix for which each row is generated as in Section 1.1. We pick mutually independent random seeds $s_1, \ldots, s_t$ and construct the hash functions $h_{s_1}, \ldots, h_{s_t}$ using Corollary 1. We then define

$$L_{i,j} \;=\; h_{s_i}(j)/\sqrt{t} \qquad \forall i \in [t],\, j \in [n]. \tag{6}$$

The algorithm's output is the vector $y = Lf$.

Each coordinate of y can be analyzed as in Section 1.1, after incorporating the scaling factor $1/\sqrt{t}$. Equation (3) and Claim 4 become

$$\begin{aligned}\mathrm{E}\left[\, y_i^2 \,\right] &= \left\| f \right\|_2^2 / t \\ \mathrm{Var}\left[\, y_i^2 \,\right] &\leq 3 \left\| f \right\|_2^4 / t^2.\end{aligned}$$

Linearity of expectation and Corollary 8 imply

$$\begin{aligned}\mathrm{E}\left[\, \textstyle\sum_{i=1}^t y_i^2 \,\right] &= \left\| f \right\|_2^2 \\ \mathrm{Var}\left[\, \textstyle\sum_{i=1}^t y_i^2 \,\right] &\leq 3 \left\| f \right\|_2^4 / t.\end{aligned}$$

Plugging these into Chebyshev's inequality and taking $z = \epsilon \left\| f \right\|_2^2$, we get

$$\Pr\left[\; \left| \textstyle\sum_{i=1}^t y_i^2 - \left\| f \right\|_2^2 \right| \;\geq\; \epsilon \left\| f \right\|_2^2 \;\right] \;\leq\; \frac{\mathrm{Var}\left[\, \sum_{i=1}^t y_i^2 \,\right]}{z^2} \;\leq\; \frac{3 \left\| f \right\|_2^4 / t}{(\epsilon \left\| f \right\|_2^2)^2} \;=\; \frac{3}{t\epsilon^2}.$$

Thus, taking $t = 3/(\delta\epsilon^2)$, we obtain

$$\Pr\left[\; \left| \textstyle\sum_{i=1}^t y_i^2 - \left\| f \right\|_2^2 \right| \;\geq\; \epsilon \left\| f \right\|_2^2 \;\right] \;\leq\; \delta.$$

1.2.1 Space Analysis.

Let us now consider how much space the algorithm needs in order to obtain an estimate that achieves $(1+\epsilon)$-multiplicative error with failure probability δ.

The vector y. There are t coordinates, each of which uses $O(\log n)$ bits.

The hash functions $h_{s_1}, \ldots, h_{s_t}$. To represent the hash function h_{s_i} we only need to store the random seed s_i. By Corollary 1, each seed uses $O(k \log n)$ bits of space, which is $O(\log n)$ since $k = 4$.

Thus, the total space is $O(t \log n) = O(\log(n)/\delta\epsilon^2)$ bits.

A Review of Variance

Let us review ***variance*** and related notions that should be familiar from an introductory probability course. The variance of a random variable X is

$$\mathrm{Var}\left[\,X\,\right] \;=\; \mathrm{E}\left[\,\big(X - \mathrm{E}\left[\,X\,\right]\big)^2\,\right] \;=\; \mathrm{E}\left[\,X^2\,\right] - \mathrm{E}\left[\,X\,\right]^2.$$

(For some random variables, the variance may be undefined or infinite.)

The ***covariance*** between two random variables X and Y is

$$\mathrm{Cov}\left[\,X, Y\,\right] \;=\; \mathrm{E}\left[\,\big(X - \mathrm{E}\left[\,X\,\right]\big)\big(Y - \mathrm{E}\left[\,Y\,\right]\big)\,\right] \;=\; \mathrm{E}\left[\,XY\,\right] - \mathrm{E}\left[\,X\,\right]\mathrm{E}\left[\,Y\,\right].$$

This gives some measure of the correlation between X and Y.

Here are some properties of variance and covariance that follow from the definitions by simple calculations.

Claim 5 *If X and Y are pairwise independent then* $\mathrm{Cov}\left[\,X, Y\,\right] = 0$.

PROOF: We showed in Lecture 13 that pairwise independent random variables satisfy $\mathrm{E}\left[\,XY\,\right] - \mathrm{E}\left[\,X\,\right]\mathrm{E}\left[\,Y\,\right]$. □

References: Mitzenmacher-Upfal Corollary 3.4.

Claim 6 $\mathrm{Var}\left[\,X + Y\,\right] = \mathrm{Var}\left[\,X\,\right] + \mathrm{Var}\left[\,Y\,\right] + 2 \cdot \mathrm{Cov}\left[\,X, Y\,\right]$.

References: Mitzenmacher-Upfal Theorem 3.2.

More generally, induction shows

Claim 7 *Let $X_1, \ldots, X_n$ be arbitrary random variables. Then*

$$\mathrm{Var}\left[\,\sum_{i=1}^{n} X_i\,\right] \;=\; \sum_{i=1}^{n} \mathrm{Var}\left[\,X_i\,\right] + 2 \sum_{i=1}^{n} \sum_{j>i} \mathrm{Cov}\left[\,X_i, X_j\,\right].$$

Corollary 8 *Let $X_1, \ldots, X_n$ be pairwise independent random variables. Then*

$$\mathrm{Var}\left[\,\sum_{i=1}^{n} X_i\,\right] \;=\; \sum_{i=1}^{n} \mathrm{Var}\left[\,X_i\,\right].$$

References: Mitzenmacher-Upfal Theorem 3.5, Motwani-Raghavan Lemma 3.4, Durrett Theorem 2.2.1.

B Chebyshev's Inequality

Chebyshev's inequality you've also presumably seen before. It is a 1-line consequence of Markov's inequality.

Theorem 9 *For any $z > 0$,*

$$\Pr\left[\,\left|X - \mathrm{E}\left[\,X\,\right]\right| \geq z\,\right] \;\leq\; \frac{\mathrm{Var}\left[\,X\,\right]}{z^2}.$$

PROOF:

$$\Pr\left[\,\left|X - \mathrm{E}\left[\,X\,\right]\right| \geq z\,\right] \;=\; \Pr\left[\left(X - \mathrm{E}\left[\,X\,\right]\right)^2 \geq z^2\right] \;\leq\; \frac{\mathrm{E}\left[\left(X - \mathrm{E}\left[\,X\,\right]\right)^2\right]}{z^2} \;=\; \frac{\mathrm{Var}\left[\,X\,\right]}{z^2},$$

where the inequality is by Markov's inequality. □

References: Durrett Theorem 1.6.4.

B.1 Chebyshev for sums of pairwise independent RVs

Chebyshev is often useful for showing concentration for sums of pairwise independent random variables. Suppose that $X_1, \ldots, X_n$ are pairwise independent and identically distributed. Then, by Corollary 8,

$$\mathrm{Var}\left[\,\textstyle\sum_{i=1}^n X_i\,\right] \;=\; \sum_{i=1}^n \mathrm{Var}\left[\,X_i\,\right] \;=\; nv,$$

where $\mathrm{Var}\left[\,X_i\,\right] \leq v$ for each i. So, by Chebyshev's inequality,

$$\Pr\left[\,\left|\textstyle\sum_{i=1}^n X_i - \mathrm{E}\left[\,\sum_{i=1}^n X_i\,\right]\right| > z\,\right] \;\leq\; \frac{\mathrm{Var}\left[\,\sum_{i=1}^n X_i\,\right]}{z^2} \;\leq\; nv/z^2.$$

References: Mitzenmacher-Upfal Section 13.2.