Scalable ML

10605-10805

Space Complexity of Frequency Moments

Barnabás Póczos

The space complexity of approximating the frequency moments

Noga Alon, Yossi Matias, Mario Szegedy

Journal of Computer and System Sciences
Volume 58, Issue 1, February 1999, Pages 137-147

Problem Statement

We are given a sequence $A = [a_1, a_2, \ldots, a_m]$

$a_i \in \{1, 2, \ldots, n\}, \forall 1 \leq i \leq m$

The elements $a_1, s_2, \ldots, a_m$ are given one by one, and they cannot be stored in the memory or hard drive.

Example: $A = [1, 2, 1, 1, 3, 2, 1, 1]$,
$a_i \in \{1, 2, 3\}$, $m = 8$, $n = 3$

Problem Statement

Let m_i denote how many times we oberved $i \in \{1, \ldots, n\}$ in $A = [a_1, \ldots, a_m]$

$$m_i = |\{j : a_j = i, 1 \le j \le m\}|$$

Observation: $\sum_{i=1}^{n} m_i = m$

Type	Number
1	m_1
2	m_2
…	…
i	m_i
…	…
n	m_n
Sum	**m**

Problem Statement

Goal: Estimate $F_k = \sum_{i=1}^{n} m_i^k$ frequency moments.

$$F_0 = \sum_{i=1}^{n} \underbrace{m_i^0}_{1} = n$$

$$F_1 = \sum_{i=1}^{n} m_i = m$$

$$F_2 = \sum_{i=1}^{n} m_i^2$$

$\ldots$

$$F_k = \sum_{i=1}^{n} m_i^k$$

$\ldots$

$$F_\infty^* = \max_{1 \le i \le n} m_i$$

The Naïve Method

If we have at least $n \log m$ bit of memory,
then calculating $F_k = \sum_{i=1}^{n} m_i^k$ is easy:

* Calculate $m_1, m_2, \ldots, m_n$ from $A = [a_1, \ldots, a_m]$

* Each can be stored on $\log m$ bits storage since $m_i \leq m$

* We have n of these numbers

* This together requires $n \log m$ bits of memory.

Surprising Fact

Informal (somewhat) surprising statement:

F_k can be estimated randomly with using only

$$O\left(n^{1-1/k}(\log n + \log m)\right) \qquad \text{bits}$$

Formal Statement

Theorem 1: For every $\left.\begin{array}{r} k \geq 1 \\ \lambda > 0 \\ \epsilon > 0 \end{array}\right\}$

$\Rightarrow$ There exists a randomized algorithm $\mathcal{A}$ such that

given a sequence $A = [a_1, \ldots, a_m]$ (where $a_i \in N \doteq \{1, 2, \ldots, n\}$)

$\star$ in one pass $\mathcal{A}$ computes random $Y \in \mathbb{R}$ so that

$$Pr\Big(|Y - F_k| > \lambda F_k\Big) < \epsilon \qquad (*1)$$

$\star$ $\mathcal{A}$ is using only

$$O\left(\frac{k \log(1/\epsilon)}{\lambda^2} n^{1-1/k} (\log n + \log m)\right)$$ memory bits.

(*2)

Main Idea of the Algorithm

Main idea to estimate F_k:

Define a random variable whose

- expected value is F_k
- variance is relatively small

Preliminaries

Let $\lambda > 0$, $\epsilon > 0$ be fixed

This 12 is only 8 in the paper,

but I can only see the proof with 12

Let $S_1 \doteq \frac{12kn^{1-1/k}}{\lambda^2}$

Let $S_2 \doteq 2\log(1/\epsilon)$

We want to prove that $\mathcal{A}$ will only use

$$O\left(\underbrace{\frac{k\log(1/\epsilon)}{\lambda^2}n^{1-1/k}}_{O(S_1S_2)}(\log n + \log m)\right) \quad \text{memory bits.}$$

Here we assume m is known in advance, i.e. $\mathcal{A}$ can use m.

The paper also discusses the case when m is not known in advance, i.e. $\mathcal{A}$ cannot use m.

The Random Algorithm to Calculate Y

The Algorithm $\mathcal{A}$ will sample $\{X_{ij}\}$ i.i.d. random variables.

$$1 \le j \le S_1$$
$$1 \le i \le S_2$$

From these X_{ij} random variables we will caculate $Y_1, Y_2, \ldots, Y_{S_2}$ random variables

$$\begin{pmatrix} X_{1,1} & \ldots & X_{1,S_1} \\ \vdots & & \vdots \\ X_{i,1} & \ldots & X_{i,S_1} \\ \vdots & & \vdots \\ X_{S_2,1} & \ldots & X_{S_2,S_1} \end{pmatrix} \begin{array}{l} \Rightarrow Y_1 \doteq \frac{1}{S_1}\left(X_{1,1} + \ldots + X_{1,S_1}\right) \\ \Rightarrow Y_i \doteq \frac{1}{S_1}\left(X_{i,1} + \ldots + X_{i,S_1}\right) \\ \Rightarrow Y_{S_2} \doteq \frac{1}{S_1}\left(X_{S_2,1} + \ldots + X_{S_2,S_1}\right) \end{array}$$

Let $Y \doteq \text{median}(Y_1, Y_2, \ldots, Y_{S_2})$

We will show that Y can be used to estimate F_k, i.e.

$$Pr\Big(|Y - F_k| > \lambda F_k\Big) < \epsilon$$

Proof of Main Steps

To compute $X_{i,j}$ for a given i and j, we will use $O(\log n + \log m)$ bits

Therefore, to compute $\{X_{i,j}\}_{i=1,\, j=1}^{S_2,\; S_1}$, the algorithm $\mathcal{A}$ will use

$$O\left(S_1 S_2 (\log n + \log m)\right) \text{ bits}$$

and this is what we had to prove for (*2) part of Theorem 1.

The remaining parts are (i) to show how to compute X_{ij} and (ii) to show that

$$Pr\Big(|Y - F_k| > \lambda F_k\Big) < \epsilon$$

The Random Algorithm to Calculate X_{ij}

Let $p \sim U[1, 2, \ldots, m]$

$\Rightarrow a_p$ is a random member of $A = [a_1, a_2, \ldots, a_m]$

Suppose $a_p = l \in \{1, 2, \ldots, n\}$

Let $r = |\{q : n \geq q \geq p, a_q = l\}|$

$= $ num of occurancies of l in A following a_p

Example: $A = [1, 2, 1, 1, 3, 2, 1, 1]$

$\Rightarrow r = |\{4, 7, 8\}| = 3$

$p = 4$

Let $X \doteq m(r^k - (r-1)^k)$

The Random Algorithm to Calculate X_{ij}

This is what we know so far:

$$r = |\{q : n \geq q \geq p, a_q = l\}|$$
$$= \text{num of occurancies of } l \text{ in } A \text{ following } a_p$$

$$X = m(r^k - (r-1)^k)$$

Remark:

To be able to calculate $X = m(r^k - (r-1)^k)$, we need

$\log n$ bits to store $a_p = l \in \{1, 2, \ldots, n\}$.
(a_p is used to caclulate r)

$\log m$ bits to store r $(1 \leq r \leq m)$

That is $\log m + \log n$ bits together.

This is what we wanted to prove for (*2) part of Theorem 1

Example

Example: $A = [a_1, a_2, \ldots, a_m]$

$$= [1, 2, 1, 1, 3, 2, 1, 1]$$

$$m_1 = 5$$

$$m_2 = 2 \qquad F_k = 5^k + 2^k + 1^k$$

$$m_3 = 1$$

$$\begin{array}{lllll}
\text{If } p = 1 & \Rightarrow & a_1 = 1 & \Rightarrow & r = |1, 3, 4, 7, 8| = 5 \\
\text{If } p = 2 & \Rightarrow & a_2 = 2 & \Rightarrow & r = |2, 6| = 2 \\
\text{If } p = 3 & \Rightarrow & a_3 = 1 & \Rightarrow & r = |3, 4, 7, 8| = 4 \\
\text{If } p = 4 & \Rightarrow & a_4 = 1 & \Rightarrow & r = |4, 7, 8| = 3 \\
\text{If } p = 5 & \Rightarrow & a_5 = 3 & \Rightarrow & r = |5| = 1 \\
\text{If } p = 6 & \Rightarrow & a_6 = 2 & \Rightarrow & r = |6| = 1 \\
\text{If } p = 7 & \Rightarrow & a_7 = 1 & \Rightarrow & r = |7, 8| = 2 \\
\text{If } p = 8 & \Rightarrow & a_8 = 1 & \Rightarrow & r = |8| = 1
\end{array}$$

Proof of Theorem 1

Example (Continued)

We will caclulate $\mathbb{E}[X]$ and $Var(X)$.

By definition, $X = \underbrace{m(r^k - (r-1)^k)}_{f(r)}$

Therefore,

$$\mathbb{E}[X] = \frac{1}{8}[f(5) + f(2) + f(4) + f(3) + f(1) + f(1) + f(2) + f(1)]$$

$$= \frac{1}{8}\begin{pmatrix} f(1) + f(2) + f(3) + f(4) + f(5) \\ +f(1) + f(2) \\ +f(1) \end{pmatrix}$$

n rows

m elements

$p \sim U[\{1, 2, \ldots, 8\}]$

Expected Value of X

More generally,

$$\mathbb{E}[X] = \frac{1}{m}\begin{pmatrix} f(1)+f(2)+\ldots+f(m_1) \\ +f(1)+f(2)+\ldots+f(m_2) \\ \vdots \\ +f(1)+f(2)+\ldots+f(m_i) \\ \vdots \\ +f(1)+f(2)+\ldots+f(m_n) \end{pmatrix}$$

n rows

m elements

$p \sim U[\{1,2,\ldots,m\}]$

$$= \frac{m}{m}\begin{pmatrix} 1^k+(2^k-1^k)+\ldots+(m_1^k-(m_1-1)^k) \\ +1^k+(2^k-1^k)+\ldots+(m_2^k-(m_2-1)^k) \\ \vdots \\ +1^k+(2^k-1^k)+\ldots+(m_i^k-(m_i-1)^k) \\ \vdots \\ +1^k+(2^k-1^k)+\ldots+(m_n^k-(m_n-1)^k) \end{pmatrix} = \sum_{i=1}^{n} m_i^k = F_k$$

We have proved that X is unbiased estimator of F_k, i.e. $\mathbb{E}[X] = F_k$

Preliminaries

Lemma 1 [Bound on $a^k - b^k$]

If $0 \le b \le a$, then $a^k - b^k \le (a-b)ka^{k-1}$

Proof of Lemma 1

$$
\begin{aligned}
a^k - b^k &= (a-b)(a^{k-1} + a^{k-2}b + \ldots + ab^{k-2} + b^{k-1}) \\
&\le (a-b)(ka^{k-1})
\end{aligned}
$$

Since $0 \le b \le a$

This finishes the proof of Lemma 1. Q.E.D.

Lemma 2

Lemma 2

If $1 \leq a$, then $(a^k - (a-1)^k)^2 \leq ka^{2k-1} - k(a-1)^{2k-1}$

Proof

$$\begin{aligned}(a^k - (a-1)^k)^2 &= (a^k - (a-1)^k)(a^k - (a-1)^k) \\ &\leq ka^{k-1}(a^k - (a-1)^k) \\ &= ka^{2k-1} - ka^{k-1}(a-1)^k \\ &\leq ka^{2k-1} - k(a-1)^{2k-1}\end{aligned}$$

From Lemma 1[If $0 \leq b \leq a, \Rightarrow a^k - b^k \leq (a-b)ka^{k-1}$]

This finishes the proof of Lemma 2. Q.E.D

Variance of X

Let us calculate and bound the variance of X.

$$Var(X) = \mathbb{E}[X^2] - \underbrace{\mathbb{E}^2[X]}_{F_k^2}$$

Lemma 3 [Bound on E[X²]]

$$\begin{aligned}\mathbb{E}[X^2] &\leq kmF_{2k-1} \\ &= kF_1F_{2k-1}\end{aligned}$$

Proof of Lemma 3

By definition,

$$X = \underbrace{m(r^k - (r-1)^k)}_{f(r)}$$

Therefore,

$$X^2 = \underbrace{m^2(r^k - (r-1)^k)^2}_{f^2(r)}$$

Expected Value of X^2

We know $$X^2 = \underbrace{m^2(r^k - (r-1)^k)^2}_{f^2(r)}$$

Therefore,

$$\mathbb{E}[X^2] = \frac{1}{m}\begin{pmatrix} f^2(1) + f^2(2) + \ldots + f^2(m_1) \\ +f^2(1) + f^2(2) + \ldots + f^2(m_2) \\ \vdots \\ +f^2(1) + f^2(2) + \ldots + f^2(m_i) \\ \vdots \\ +f^2(1) + f^2(2) + \ldots + f^2(m_n) \end{pmatrix}$$

$$= \frac{m^2}{m}\begin{pmatrix} 1^{2k} + (2^k - 1^k)^2 + \ldots + (m_1^k - (m_1 - 1)^k)^2 \\ +1^{2k} + (2^k - 1^k)^2 + \ldots + (m_2^k - (m_2 - 1)^k)^2 \\ \vdots \\ +1^{2k} + (2^k - 1^k)^2 + \ldots + (m_i^k - (m_i - 1)^k)^2 \\ \vdots \\ +1^{2k} + (2^k - 1^k)^2 + \ldots + (m_n^k - (m_n - 1)^k)^2 \end{pmatrix}$$

Expected Value of X^2

Therefore from Lemma 2

[If $1 \leq a$, then $(a^k - (a-1)^k)^2 \leq ka^{2k-1} - k(a-1)^{2k-1}$]

$$\mathbb{E}[X^2] = m \begin{pmatrix} \underbrace{1^{2k}}_{\leq k1^{2k-1}} + \underbrace{(2^k - 1^k)^2}_{\leq k2^{2k-1} - k1^{2k-1}} + \ldots + \underbrace{(m_1^k - (m_1 - 1)^k)^2}_{\leq km_1^{2k-1} - k(m_1 - 1)^{2k-1}} \\ \vdots \\ \underbrace{1^{2k}}_{\leq k1^{2k-1}} + \underbrace{(2^k - 1^k)^2}_{\leq k2^{2k-1} - k1^{2k-1}} + \ldots + \underbrace{(m_i^k - (m_i - 1)^k)^2}_{\leq km_i^{2k-1} - k(m_i - 1)^{2k-1}} \\ \vdots \\ \underbrace{1^{2k}}_{\leq k1^{2k-1}} + \underbrace{(2^k - 1^k)^2}_{\leq k2^{2k-1} - k1^{2k-1}} + \ldots + \underbrace{(m_n^k - (m_n - 1)^k)^2}_{\leq km_n^{2k-1} - k(m - 1)^{2k-1}} \end{pmatrix}$$

$$\leq m(km_1^{2k-1} + \ldots + km_n^{2k-1})$$

$$= mkF_{2k-1}$$

$$= kF_1 F_{2k-1}$$

This completes the proof of Lemma 3. Q.E.D

Power Mean Inequality

Lemma 4

If $$\left.\begin{array}{r} m_1 \geq 0 \\ m_2 \geq 0 \\ \vdots \\ m_n \geq 0 \\ k \geq 1 \end{array}\right\} \Rightarrow \sum_{i=1}^{n} m_i \leq n\Big(\sum_{i=1}^{n} m_i^k/n\Big)^{1/k}$$

Proof of Lemma 4

According to the power-mean inequality,

If $k \geq 1$, then $$\sum_{i=1}^{n} \frac{1}{n} m_i \leq \Big(\sum_{i=1}^{n} \frac{1}{n} m_i^k\Big)^{1/k}$$

See, e.g. Cauchy-Schwarz master class, Problem 8.3

This finishes the proof of Lemma 4. Q.E.D

Variance of X

By Lemma 3, we have that $\mathbb{E}[X^2] \leq kmF_{2k-1} = kF_1F_{2k-1}$

Let us bound the r.h.s

Lemma 5 (Bound on E[X^2])

$$\text{If} \left.\begin{array}{r} m_1 \geq 0 \\ m_2 \geq 0 \\ \vdots \\ m_n \geq 0 \\ k \geq 1 \end{array}\right\} \Rightarrow F_1 F_{2k-1} \leq n^{1-1/k} F_k^2$$

Proof of Lemma 5

We need to prove:

$$\underbrace{\left(\sum_{i=1}^{n} m_i\right)}_{F_1}\underbrace{\left(\sum_{i=1}^{n} m_i^{2k-1}\right)}_{F_{2k-1}} \leq n^{1-1/k}\underbrace{\left(\sum_{i=1}^{n} m_i^k\right)^2}_{F_k^2}$$

Proof of Lemma 5 (Continued)

Observation 1

Let $M \doteq \max\limits_{1 \le i \le n} m_i$

Since $M^k = \max\limits_{1 \le i \le n} m_i^k$

$$\le \sum_{1 \le i \le n} m_i^k$$

therefore, $M \le (\sum\limits_{i=1}^{n} m_i^k)^{1/k}$

Proof of Lemma 5 (Continued)

We need to prove: $\left(\sum_{i=1}^{n} m_i\right)\left(\sum_{i=1}^{n} m_i^{2k-1}\right) \leq n^{1-1/k}\left(\sum_{i=1}^{n} m_i^k\right)^2$

$$\left(\sum_{i=1}^{n} m_i\right)\left(\sum_{i=1}^{n} m_i^{2k-1}\right) \leq \left(\sum_{i=1}^{n} m_i\right)\left(\sum_{i=1}^{n} M^{k-1} m_i^k\right)$$

Since $M \doteq \max_{1 \leq i \leq n} m_i$

$$= \left(\sum_{i=1}^{n} m_i\right) \underbrace{M^{k-1}}_{\leq \left(\sum_{i=1}^{n} m_i^k\right)^{\frac{k-1}{k}}} \left(\sum_{i=1}^{n} m_i^k\right)$$

Since $M \leq (\sum_{i=1}^{n} m_i^k)^{1/k}$ by Observation 1

$$\leq \left(\sum_{i=1}^{n} m_i\right)\left(\sum_{i=1}^{n} m_i^k\right)^{\frac{k-1}{k}}\left(\sum_{i=1}^{n} m_i^k\right)$$

Proof of Lemma 5 (Continued)

This is what we know so far:

$$\Big(\sum_{i=1}^{n} m_i\Big)\Big(\sum_{i=1}^{n} m_i^{2k-1}\Big) \le \Big(\sum_{i=1}^{n} m_i\Big)\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{\frac{k-1}{k}}\Big(\sum_{i=1}^{n} m_i^{k}\Big)$$

$$= \underbrace{\Big(\sum_{i=1}^{n} m_i\Big)}_{\le n\left(\sum_{i=1}^{n} m_i^k/n\right)^{1/k}}\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{\frac{2k-1}{k}}$$

By Lemma 4

$\sum_{i=1}^{n} m_i \le n\left(\sum_{i=1}^{n} m_i^k/n\right)^{1/k}$

$$\le n\Big(\sum_{i=1}^{n} m_i^{k}/n\Big)^{1/k}\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{\frac{2k-1}{k}}$$

$$= n^{1-1/k}\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{1/k}\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{\frac{2k-1}{k}}$$

$$= n^{1-1/k}\Big(\sum_{i=1}^{n} m_i^{k}\Big)^{2}$$

This is what we had to prove for Lemma 5. Q.E.D

Summary

So far we know

$$\mathbb{E}[X_{ij}] = F_k \qquad\qquad \forall i,j \quad \begin{array}{c} 1 \le j \le S_1 \\ 1 \le i \le S_2 \end{array}$$

$$\mathbb{E}[X_{ij}^2] \le kF_1F_{2k-1} \le kn^{1-1/k}F_k^2$$

By Definition,

$$\begin{pmatrix} X_{1,1} & \dots & X_{1,S_1} \\ \vdots & & \vdots \\ X_{i,1} & \dots & X_{i,S_1} \\ \vdots & & \vdots \\ X_{S_2,1} & \dots & X_{S_2,S_1} \end{pmatrix} \Rightarrow Y_i \doteq \frac{1}{S_1}\left(X_{i,1} + \ldots + X_{i,S_1}\right)$$

Therefore,

$$\mathbb{E}[Y_i] = \frac{1}{S_1}\sum_{j=1}^{S_1}\mathbb{E}[X_{i,j}] = F_k$$

That is, Y_i is unbiased estimator of F_k.

Variance of Y_i

Let us bound the variance of Y_i

Lemma 6

$$\text{Var } [Y_i] \leq \tfrac{1}{S_1}\mathbb{E}[X^2]$$

Proof

$$\text{Var}[Y_i] = \text{Var}\left[\frac{1}{S_1}\sum_{j=1}^{S_1} X_{ij}\right]$$

$$= \frac{1}{S_1^2}\sum_{j=1}^{S_1}\underbrace{\text{Var}[X_{ij}]}_{\mathbb{E}[X^2]-F_k^2} \qquad \text{Since } \mathbb{E}[X] = F_k$$

$$= \frac{1}{S_1}(\mathbb{E}[X^2] - F_k^2)$$

$$\leq \frac{1}{S_1}\mathbb{E}[X^2]$$

This is what we had to prove for Lemma 6. Q.E.D

Variance of Y_i

Lemma 7

$$\text{Var }[Y_i] \le \frac{k}{S_1} n^{1-1/k} F_k^2$$

Proof of Lemma 7

$$\text{Var }[Y_i] \le \frac{1}{S_1}\mathbb{E}[X^2]$$

By Lemma 6

$$\le \frac{1}{S_1} k F_1 F_{2k-1}$$

By Lemma 3: $\mathbb{E}[X^2] \le k F_1 F_{2k-1}$

$$\le \frac{k}{S_1} n^{1-1/k} F_k^2$$

By Lemma 5: $F_1 F_{2k-1} \le n^{1-1/k} F_k^2$

This completes the proof of Lemma 7. Q.E.D

Concentration Bound on Y_i

Chebyshev's inequality:

If $a > 0$, then $Pr(|Y - \mathbb{E}[Y]| > a) \leq \dfrac{\text{Var}[Y]}{a^2}$

Lemma 8

Let i be fixed ($1 \leq i \leq S_2$). We have that $Pr(|Y_i - F_k| > \lambda F_k) \leq \dfrac{1}{12}$

Proof of Lemma 8 From Chebyshev's inequality, we have that

For all fixed i:

$$Pr(|Y_i - \underbrace{\mathbb{E}[Y]}_{F_k}| > \underbrace{\lambda F_k}_{a}) \leq \frac{\text{Var}[Y_i]}{\lambda^2 F_k^2}$$

$$\leq \frac{\frac{k}{S_1} n^{1-1/k} F_k^2}{\lambda^2 F_k^2}$$

Since Var $[Y_i] \leq \frac{k}{S_1} n^{1-1/k} F_k^2$ by Lemma 7

$$= \frac{k n^{1-1/k}}{\lambda^2 S_1}$$

Since $S_1 \doteq \dfrac{12 k n^{1-1/k}}{\lambda^2}$

$$\leq \frac{1}{12}$$

This completes the proof of Lemma 8. Q.E.D

Concentration Bound on Y

We know from Lemma 8 that

For all $1 \leq i \leq S_2$, we have that $Pr(|Y_i - F_k| > \lambda F_k) \leq \frac{1}{12}$

This is a bound on the probability that (for a fixed i) Y_i deviates from $\mathbb{E}[Y_i] = F_k$ more than λF_k

Now let us bound the probability that at least $S_2/2$ among $\{Y_1, Y_2, \ldots, Y_{S_2}\}$ deviates from F_k more than λF_k

We will prove that this probability is $\leq \epsilon$

Since $Y \doteq \text{median}(Y_1, Y_2, \ldots, Y_{S_2})$, this will imply that

$$Pr\Big(|Y - F_k| > \lambda F_k\Big) \leq \epsilon$$

This will complete the proof of Theorem 1

Chernoff Bound

Lemma [Chernoff, 1952, Tail bound for binomial distribution]

Let B be a binomial random variable $B \sim B(n, p)$

Let $0 < p < \delta < 1$

Then

$$Pr(B > n\delta) \leq \exp\left(-n \underbrace{\left[\delta \log \frac{\delta}{p} + (1-\delta) \log \frac{1-\delta}{1-p}\right]}_{\geq 0}\right)$$

Proof: [Gyorfi et al, Distribution Free Theory of Nonparametric Regression, Page 592]

Application of Chernoff Bound

Let $Z_i \doteq 1_{\{|Y_i - F_k| \geq \lambda F_k\}} \in \{0,1\}$

We already know from Lemma 8 that

$$\mathbb{E}[Z_i] = \mathbb{P}(|Y_i - F_k| > \lambda F_k) \leq \frac{1}{12}$$

Observation:

$$\sum_{i=1}^{S_2} Z_i \sim Binomial(S_2, p) \quad \text{where } p = \mathbb{E}[Z_i] \leq \tfrac{1}{12}$$

To see that (*1) part of Theorem 1 holds, we need to prove that

$$Pr\left(\sum_{i=1}^{S_2} Z_i > \frac{S_2}{2}\right) \leq \epsilon$$

Application of Chernoff Bound

From Chernoff $Pr(B > n\delta) \leq \exp\left(-n\left[\delta \log\frac{\delta}{p} + (1-\delta)\log\frac{1-\delta}{1-p}\right]\right)$

Let $B \doteq \sum_{i=1}^{S_2} Z_i$, $n \doteq S_2$, $\delta \doteq 1/2$ in the Chernoff bound

$$Pr\left(\sum_{i=1}^{S_2} Z_i > \frac{S_2}{2}\right) \leq \exp\left(-S_2\left[\frac{1}{2}\log\frac{1}{2p} + \frac{1}{2}\log\frac{1}{2(1-p)}\right]\right)$$

Since $S_2 = 2\log(1/\epsilon)$

$-S_2 = 2\log(\epsilon)$

$$= \exp\left((\log\epsilon)\left[\log\frac{1}{2p} + \log\frac{1}{2(1-p)}\right]\right)$$

$$= \exp\left((\log\epsilon)\left[\log\frac{1}{4p(1-p)}\right]\right)$$

If $\log\frac{1}{4p(1-p)} > 1$

i.e $\frac{1}{4p(1-p)} > e$

i.e $1 > e4p(1-p)$

$$\leq \exp(\log\epsilon)$$

$$= \epsilon$$

This holds if $p < \frac{1}{12}$. **This completes the proof of Theorem 1 Q.E.D.**

Thanks for your Attention! ☺

Another Surprising Fact

We know that

F_k can be estimated randomly with using only

$$O\left(n^{1-1/k}(\log n + \log m)\right) \quad \text{bits}$$

Theorem:

F_2 can be estimated randomly with using only

$$O\Big((\log n + \log m)\Big) \quad \text{bits}$$