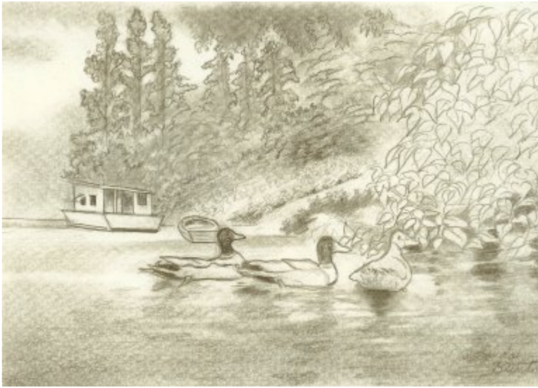
Sketching Streams

Sumit Ganguly

IIT Kanpur



Data Stream Model

Stream is a sequence of records

- Arrives fast, continuously.
- Not enough main memory to store stream.
- Too fast to store on secondary storage with random access. May be stored as a log file for later mining.


Example Applications

- Network switch data (Distr. Denial of Service brewing?)
- Sensor networks (intrusion?)
- Satellite data (storm? flashflood?)
- Others: web-usage, financial market, etc.

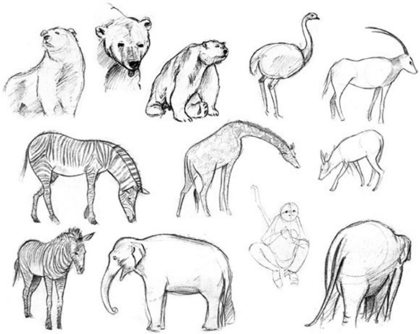
Data Stream Processing Model

- Low space data structure: Sub-linear/ poly-logarithmic in stream size.
- Process each arriving record efficiently to match fast arrival speeds.
- Online Processing: input record is processed as it arrives.
- Streaming Model: Online, sub-linear space and time processing.
- Other Models: not in this talk.
  - Semi-Streaming: Stores data in sequential order. Multiple passes are allowed.

This talk

Some algorithmic techniques have evolved for data stream processing. We will see some important ones:

Linear Sketching, Dimensionality Reduction.



Not in this talk

Sampling from Data Streams: Not covered

Data Stream Model

- Domain of items $[n] = \{1, 2, \ldots, n\}$.
- n is known but very large : IP-addresses, pairs of IP-addresses—2^{64}.
- Insert-Delete Streams: Sequence of updates (*item*, *change in frequency*) $\equiv (i, v)$.

$$(1,1)\ (4,1)\ (5,3)\ (7,1)\ (5,-1)\ (5,2)\ (7,2)\ (6,1)\ (1,-1)\ldots$$

Frequency Vector

	0	0	0	1	4	1	3	0	0	...
Items	1	2	3	4	5	6	7	8	9	

Frequency Vector of Stream

$(1,1)\ (4,1)\ (5,3)\ (7,1)\ (5,-1)\ (5,2)\ (7,2)\ (6,1)\ (1,-1)\ldots$

Frequency Vector

Items	1	2	3	4	5	6	7	8	9	
	0	0	0	1	4	1	3	0	0	...

Incremental view:

1. Initially $f = 0$.
2. When (i, v) arrives:

$$f_i := f_i + v \quad .$$

Global view:

$$f_i = \sum_{(i,v) \in \text{ stream}} v, \quad i \in [n] \quad .$$

Data Streaming: Algorithmic Model

- Single pass over stream (Online algorithm).
- Sublinear storage: n^{α} ($\alpha < 1$) or, better poly-logarithmic in n.
  - Units of storage: bits.
- Fast processing per arriving stream record.
  - Approximate processing (almost always necessary).
  - Randomized computation (almost always necessary).

Independence in Probability: Revisited

- Independence: Random variables $\{X_1, X_2, \ldots, X_n\}$ are independent if their joint probability (density) function is the product of individual probability (density) function.
- Computational Problems:
  - Design $h : [n] \rightarrow \{0, 1\}$ so that $\{h(1), \ldots, h(n)\}$ are independent. All constructions require $\Omega(n)$ random bits.
  - High randomness and storage.
  - Algorithms may not always require full independence.
  - Approximate independence often suffices.

Limited Independence

$\{X_1, X_2, \ldots, X_n\}$ are k-wise independent if the joint distribution of any k variables is the product of their individual distributions.

$$\Pr\left\{X_{i_1} = a_1 \wedge X_{i_2} = a_2 \wedge \ldots \wedge X_{i_k} = a_k\right\} = \Pr\left\{X_{i_1} = a_1\right\} \Pr\left\{X_{i_2} = a_2\right\} \ldots \Pr\left\{X_{i_k} = a_k\right\} \quad .$$

for any k distinct indices $1 \leq i_1, i_2, \ldots, i_k \leq n$ and $a_1 \in \mathrm{support}(X_{i_1}), \ldots, a_k \in \mathrm{support}(X_{i_k})$.

- Product of expectation of any k distinct variables is the product of individual expectations.
- k-wise independence implies $k-1$-wise indep.

k-wise independent hash functions

[Wegman Carter JCSS 81]

- $\mathcal{H}$ is a finite family of functions mapping $[n] \to [m]$.
- Pick random member $h \in \mathcal{H}$ with prob. $1/|\mathcal{H}|$.
- $\mathcal{H}$ is k-wise independent if $\{h(x_1), \ldots, h(x_n)\}$ are k-wise independent.
- Equivalently, for distinct $x_1, x_2, \ldots, x_k \in [n]$ and $b_1, \ldots, b_k \in [m]$ not necessarily distinct,

$$\Pr_{h \in \mathcal{H}} \{(h(x_1) = b_1) \wedge (h(x_2) = b_2) \ldots \wedge (h(x_k) = b_k)\}$$

$$=$$

$$\Pr_{h \in \mathcal{H}} \{h(x_1) = b_1\} \cdot \Pr_{h \in \mathcal{H}} \{h(x_2 = b_2)\} \cdots \times \Pr_{h \in \mathcal{H}} \{h(x_k) = b_k\} \quad .$$

Hash Family: Degree $k-1$ polynomials

- $\mathbb{F}$ is a finite field of size at least n.
- $\mathcal{H}_k$: set of all k-tuples from $\mathbb{F}$. So $|\mathcal{H}_k| = |\mathbb{F}|^k$.
- Interpret a k-tuple $(a_0, \ldots, a_{k-1})$ as a degree $k-1$ polynomial $p(x)$ over $\mathbb{F}$:

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{k-1} x^{k-1} \quad .$$

- The family $\mathcal{H}_k$ is k-wise independent.

Space and Randomness

- $|\mathcal{H}_k| = |\mathbb{F}|^k$.
- Requires $k \log|\mathbb{F}|$ bits to store a polynomial from $\mathcal{H}_k$.
- Randomness required: choose $a_0, \ldots, a_k$ at random–$k \log|\mathbb{F}|$ random bits.
- $h(\cdot)$ can be computed in time $O(k)$ field operations $(+, \cdot)$.
- Special Case. $\mathcal{H}_2$: space of affine functions over $\mathbb{F}$

$$h(x) = a_0 + a_1 x, \quad a_0, a_1 \in F \quad .$$

Pair-wise independence.
"Pair-wise independence and Derandomization", Luby and Wigderson (web)

Frequency Moment Estimation

- Frequency moment defined as

$$F_p = \sum_{i \in [n]} |f_i|^k \quad .$$

  $p \in \mathbb{R}$ and non-negative.
- The problem of estimating frequency moments has played an important role in data stream computations.
- F_0 is the number of distinct elements in the stream

$$F_0 = \sum_{i \in [n]} |f_i|^0 = \left| \{ i : f_i \neq 0 \} \right| \quad .$$

F_2 Estimation Problem

[Alon Matias Szegedy: STOC'96, JCSS '98.]

- Deterministically estimating F_2 to within $1 \pm 1/16$ requires $\Omega(n)$ space.
- Modified problem: Given ϵ and δ, design an algorithm that returns $\hat{F}_2$ satisfying

$$|\hat{F}_2 - F_2| \leq \epsilon F_2 \text{ with prob. } 1 - \delta \ .$$

Linear Sketch

- Let $\xi : [n] \rightarrow \{-1, +1\}$
  - $\xi(\cdot)$ four-wise independent hash function.
  - Maps to ± 1 with equal probability.
- Implementation: Choose h at random from the family of cubic polynomials over $\mathbb{F}_{2^r}$, where, $n \leq 2^r < 2n$.

$$\xi(u) = \begin{cases} 1 & \text{if last bit of } h(u) = 1 \\ -1 & \text{otherwise.} \end{cases}$$

- A sketch is a linear combination

$$X = \sum_{i=1}^{n} f_i \xi(i) \quad .$$

- Updating sketch in presence of stream updates:

$$\textsc{UpdateSketch}(i, v) : \quad X := X + v \cdot \xi(i) \quad .$$

Sketches

Sketch: $\sum_i f_i \xi(i)$, $\xi : [n] \to \{-1, +1\}$ four wise independent.

$$\mathbb{E}\left[\xi(i)\right] = (-1)\frac{1}{2} + (1)\frac{1}{2} = 0$$

We now calculate $\mathbb{E}\left[X^2\right]$.

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} f_i \xi(i)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} f_i^2 (\xi(i))^2 + 2 \sum_{1 \le i < j \le n} f_i f_j \xi(i) \xi(j)\right] \\
&= \sum_{i=1}^{n} f_i^2 \mathbb{E}\left[(\xi(i))^2\right] + 2 \sum_{1 \le i < j \le n} f_i f_j \mathbb{E}\left[\xi(i)\xi(j)\right]
\end{aligned}$$

using linearity of expectation.

Sketch: Expectation

- We have shown that

$$\mathbb{E}\left[X^2\right] = \sum_{i=1}^{n} f_i^2 \mathbb{E}\left[(\xi(i))^2\right] + 2 \sum_{1 \le i < j \le n} f_i f_j \mathbb{E}\left[\xi(i)\xi(j)\right] \quad .$$

- Now, $(\xi(i))^2 = 1$, and by pair-wise independence, if $i \neq j$,

$$\mathbb{E}\left[\xi(i)\xi(j)\right] = \mathbb{E}\left[\xi(i)\right]\mathbb{E}\left[\xi(j)\right] = 0 \cdot 0 = 0 \quad .$$

- Therefore, we get an unbiased estimator.

$$\mathbb{E}\left[X^2\right] = \sum_{i=1}^{n} f_i^2 = F_2 \quad .$$

Sketch:Variance

$$\mathbb{E}\left[X^4\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} f_i \xi(i)\right)^4\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} f_i \xi(i) \sum_{j=1}^{n} f_j \xi(j) \sum_{k=1}^{n} f_k \xi(k) \sum_{l=1}^{n} f_l \xi(l)\right]$$

Expanding

$$\mathbb{E}\left[X^4\right] = \mathbb{E}\Bigg[\sum_{i=1}^{n} f_i^4 \xi(i)^4 + \sum_{i \neq j} 4 f_i^3 f_j (\xi(i))^3 \xi(j)$$

$$+ \sum_{i,j \text{ distinct}} 6 f_i^2 f_j^2 \xi(i)^2 \xi(j)^2 + \sum_{i,j,k \text{ distinct}} 12 f_i^2 f_j f_k \xi(i)^2 \xi(j) \xi(k)$$

$$+ \sum_{i,j,k,l \text{ distinct}} 4! f_i f_j f_k f_l \xi(i) \xi(j) \xi(k) \xi(l)\Bigg]$$

Sketch: Variance

Using linearity of expectation

$$\mathbb{E}\left[X^4\right] = \sum_{i=1}^{n} f_i^4 \mathbb{E}\left[\xi(i)^4\right] + \sum_{i \neq j} 4 f_i^3 f_j \mathbb{E}\left[(\xi(i))^3 \xi(j)\right]$$
$$+ \sum_{i,j \text{ distinct}} 6 f_i^2 f_j^2 \mathbb{E}\left[\xi(i)^2 \xi(j)^2\right] + \sum_{i,j,k \text{ distinct}} 12 f_i^2 f_j f_k \mathbb{E}\left[\xi(i)^2 \xi(j) \xi(k)\right]$$
$$+ \sum_{i,j,k,l \text{ distinct}} 4! f_i f_j f_k f_l \mathbb{E}\left[\xi(i)\xi(j)\xi(k)\xi(l)\right]$$

$\xi(j)$'s are 4-wise independent. So expectation of pair-wise, three-wise or four-wise products of $\xi(j)$'s are the product of the corresponding expectations.

Sketch: Variance

Using linearity of expectation

$$\mathbb{E}\left[X^4\right] = \sum_{i=1}^{n} f_i^4 \mathbb{E}\left[\xi(i)^4\right] + \sum_{i \neq j} 4 f_i^3 f_j \mathbb{E}\left[(\xi(i))^3 \xi(j)\right]$$
$$+ \sum_{i,j \text{ distinct}} 6 f_i^2 f_j^2 \mathbb{E}\left[\xi(i)^2 \xi(j)^2\right] + \sum_{i,j,k \text{ distinct}} 12 f_i^2 f_j f_k \mathbb{E}\left[\xi(i)^2 \xi(j) \xi(k)\right]$$
$$+ \sum_{i,j,k,l \text{ distinct}} 4! f_i f_j f_k f_l \mathbb{E}\left[\xi(i)\xi(j)\xi(k)\xi(l)\right]$$

$\xi(j)$'s are 4-wise independent. So expectation of pair-wise, three-wise or four-wise products of $\xi(j)$'s are the product of the corresponding expectations. So, for $\{i, j, k, l\}$ distinct

$$\xi(i)^2 = \xi(i)^4 = 1$$
$$\mathbb{E}\left[\xi(i)^3 \xi(j)\right] = \mathbb{E}\left[\xi(i)\xi(j)\right] = \mathbb{E}\left[\xi(i)\right] \mathbb{E}\left[\xi(j)\right] = 0 \cdot 0 = 0$$
$$\mathbb{E}\left[\xi(i)^2 \xi(j) \xi(k)\right] = \mathbb{E}\left[\xi(j)\xi(k)\right] = 0$$
$$\mathbb{E}\left[\xi(i)\xi(j)\xi(k)\xi(l)\right] = \mathbb{E}\left[\xi(i)\right] \mathbb{E}\left[\xi(j)\right] \mathbb{E}\left[\xi(k)\right] \mathbb{E}\left[\xi(l)\right] = 0 \cdot 0 \cdot 0 \cdot 0 = 0 \quad .$$

Sketches: Variance contd.

From Last Slide

$$\mathbb{E}\left[X^4\right] = \sum_{i=1}^{n} f_i^4 \mathbb{E}\left[\xi(i)^4\right] + \sum_{i,j \text{ distinct}} 4 f_i^3 f_j \mathbb{E}\left[(\xi(i))^3 \xi(j)\right]$$
$$+ \sum_{i \neq j} 6 f_i^2 f_j^2 \mathbb{E}\left[\xi(i)^2 \xi(j)^2\right] + \sum_{i,j,k \text{ distinct}} 12 f_i^2 f_j f_k \mathbb{E}\left[\xi(i)^2 \xi(j) \xi(k)\right]$$
$$+ \sum_{i,j,k,l \text{ distinct}} 4! f_i f_j f_k f_l \mathbb{E}\left[\xi(i)\xi(j)\xi(k)\xi(l)\right]$$

$$\mathbb{E}\left[X^4\right] = \sum_{i=1}^{n} f_i^4 + \sum_{i,j \text{ distinct}} 6 f_i^2 f_j^2 \leq 3 \left(\sum_{i=1}^{n} f_i^2\right)^2$$
$$\leq 3 F_2^2 \quad .$$

Designing estimator for F_2

AMS Sketch: $X = \sum_{i \in [n]} f_i \xi(i), \quad \xi : [n] \to \{1, -1\}$ 4-wise indep. .

$$\mathbb{E}\left[X^2\right] = F_2 \ .$$

$$\mathrm{Var}\left[X^2\right] = \mathbb{E}\left[X^4\right] - (\mathbb{E}\left[X\right])^2 \leq 3F_2^2 - F_2^2 = 2F_2^2$$

- We can use Chebychev's inequality (Recall)

$$\Pr\left\{\left|Y - \mathbb{E}\left[Y\right]\right| > t\right\} < \frac{\mathrm{Var}\left[Y\right]}{t^2} \ .$$

for any real valued variable Y.

Designing estimator for F_2 contd.

- Need a random variable Y with expectation F_2 and variance at most $\epsilon^2 F_2^2/8$.
- Why? Then, by Chebychev's inequality, we would have,

$$\Pr\{|Y - F_2| > \epsilon F_2\} \leq \frac{\mathrm{Var}[Y]}{\epsilon^2 F_2^2} \leq \frac{1}{8} \ .$$

- Keep t independent sketches $X_1, X_2, \ldots X_t$. Return averages of squares: $Y = \left(X_1^2 + \ldots + X_t^2\right)/t$.
- Taking average preserves expectation, by linearity of expectation and X_i^2 are *i.d.* So $\mathbb{E}[Y] = \mathbb{E}\left[X_1^2\right] = F_2$.
- Since, X_i^2's are independent, variance of their sum is the sum of their variances. So,

$$\mathrm{Var}[Y] = \frac{1}{t^2} t \mathrm{Var}[X_1^2] = 2F_2^2/t \ .$$

Estimator for F_2

- Let $t = 16/\epsilon^2$. Then $\mathbb{E}[Y] = F_2$ and
$$\mathrm{Var}\left[Y\right] \leq \epsilon^2 F_2^2/8 \ .$$
- Therefore, by Chebychev's inequality
$$\Pr\left\{\left|Y - F_2\right| \leq \epsilon F_2\right\} \geq \frac{7}{8} \ .$$
- We now use a standard argument for boosting confidence.

Boosting confidence from constant $> 1/2$ to $1 - \delta$

- Let A be a randomized algorithm.
- On input I, correct value is $Y(I)$.
- Suppose A on input I returns (random) numeric value $\hat{Y}(I)$. and the following guarantee:

$$\Pr\left\{|\hat{Y}(I) - Y(I)| < \epsilon Y(I)\right\} \geq \frac{7}{8}$$

- To boost confidence to $1 - \delta$, run A independently on I $s = O(\log \frac{1}{\delta})$ times to obtain

$$\hat{Y}_1(I), \ldots, \hat{Y}_s(I) \quad .$$

- Now return

$$\mathrm{med}\{\hat{Y}_1(I), \ldots, \hat{Y}_s(I)\}$$

Boosting using Median: Analysis

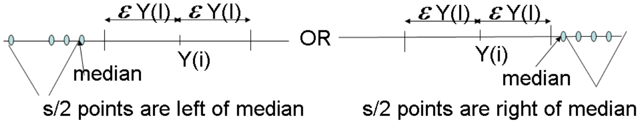
- $X_i = 0$ if the jth run of A gives a "good answer" and is 1 otherwise.

$$X_j = \begin{cases} 0 & \text{if} \quad |\hat{Y}_j(I) - Y(I)| < \epsilon Y(I) \\ 1 & \text{otherwise.} \end{cases}$$

$$\Pr\{X_j = 1\} \leq \frac{1}{8}$$

- Let $X = X_1 + X_2 + \ldots + X_s$: count number of "bad" answers.
- $\mathbb{E}[X] \leq s/8$.
- The event $\left|\text{med}(\hat{Y}_1(I), \ldots, \hat{Y}_k(I)) - Y(I)\right| > \epsilon Y(I)$ implies

$$X \geq \frac{s}{2} \quad .$$



Boosting with median: Analysis

Chernoff's bound

Let $X_1, \ldots, X_t$ be independent random variables taking values from $\{0, 1\}$ with $\mathbb{E}[X_i] = p_i$. Let $X = X_1 + X_2 + \ldots + X_t$ and $\mu = p_1 + \ldots + p_t$. Then, for $0 < \epsilon < 1$,

$$\Pr\{X > (1+\epsilon)\mu\} < e^{-\mu\epsilon^2/3}$$
$$\Pr\{X < (1-\epsilon)\mu\} < e^{-\mu\epsilon^2/2} \quad .$$

- By Chernoff's bound, with high probability, X should concentrate close to $\mathbb{E}[X] = s/8$.

$$\Pr\{X \geq s/2\} \leq \Pr\{X \geq s/4\} \leq e^{-s/24} \quad .$$

This is at most δ if $s = O(\log \frac{1}{\delta})$.

AMS F_2 estimation algorithm

- Maintain s groups of t independent sketches X_j^r, $j = 1, 2, \ldots, t$, $r = 1, 2, \ldots, s$, $t = 16/\epsilon^2$ and $s = O(\log(1/\delta))$.
- In each group r, take average

$$Y_r = \text{avg}_{j=1}^{t} (X_j^r)^2, \quad r = 1, 2, \ldots s \ .$$

- Return median of the averages

$$\hat{F}_2 = \text{med}_{r=1}^{s} Y_j \ .$$

- Property:

$$\Pr\left\{ |\hat{F}_2 - F_2| < \epsilon F_2 \right\} \geq 1 - \delta \ .$$

AMS: Resources consumed

Space:

- Let $|f_i| \leq m$. Each sketch $\sum_i f_i \xi(i)$ can be stored in $\log(mn)$ bits.
- Space = $O(\frac{1}{\epsilon^2} \log(1/\delta)) \times \log(mn)$.

Time to process stream update (i, v):

- Each sketch is updated.
- Requires evaluating degree 3 polynomial over $\mathbb{F}$: $O(1)$ simple field operations.

Randomness:

- Each sketch requires $4 \log n$ random bits.

A Dimensionality Reduction View

- Suppose we keep $s = O(\log m)$ groups.
- Sketch as a map: $f \in \mathbb{R}^n$ to $\mathrm{sk}(f) \in \mathbb{R}^{O(\epsilon^{-2} \log(m))}$.
- m streams with frequency vectors $f^1, \ldots, f^m$.
- Sketch is linear: therefore,

$$sk(f^i - f^j) = sk(f^i) - sk(f^j) \ .$$

- So with probability $1 - \frac{1}{8m^2}\left(\frac{m^2}{2} + m\right) \geq 7/8$, we have

$$\|f^i - f^j\|_2 \in (1 \pm \epsilon)\text{Medavg}(sk(f^i) - sk(f^j)), \forall i, j.$$
$$\|f^i\| \in (1 \pm \epsilon)\text{Medavg}(sk(f^i)), \forall i$$

- Medavg is not ℓ_2 norm.

Dimensionality Reduction: Metric Space view

- A discrete metric space (X, d_X): X is a finite set of points, $d_X(x, y)$ gives distance between points x and y in X. d_X function satisfies metric properties.
- (X, d_X) embeds into (Y, d_Y) with distortion D if there exists $f : X \to Y$ and a scaling constant c such that

  $$c \cdot d_X(x, y) \leq d_Y(f(x), f(y)) \leq c \cdot D \cdot d_X(x, y), \quad \forall x, y \in X \quad .$$

Well-known embeddability results

- [Bourgain] Every metric space can be embedded into ℓ_2 (any ℓ_p) with $O(\log n)$ distortion.
- [Johnson-Lindenstrauss(J-L)] There exists a randomized mapping $f : \mathbb{R}^n \to \mathbb{R}^t$, $t = O(\epsilon^{-2} \log m)$ s.t. for any set S of m points from $\mathbb{R}^n$

$$(1-\epsilon)\|x-y\|_2 \leq \|f(x)-f(y)\|_2 \leq \|x-y\|_2, \forall x, y \in S \ .$$

- $(1+\epsilon)$-distortion for arbitrary ϵ: known to be impossible for ℓ_p to ℓ_q metric.

Non-embeddability doesn't imply non-estimation

Following is still possible:

- there is a randomized function $f : \mathbb{R}^n \to \mathbb{R}^t$, $t = O(1/\epsilon^2 \log m)$ s.t. for any set S from $\mathbb{R}^n$ having m points,

$$\|x - y\|_p \in (1 \pm \epsilon) d'(f(x), f(y)), \quad \forall x, y \in S$$

  with probability $7/8$.
- But d' is not a metric.

Usefulness of Embeddability

- ϵ-distortion implies: nearest neighbors are approximately preserved.
- k-d trees and other ℓ_2-based geometric data structures can be used in much fewer dimensions.
- Time complexity of most geometric algorithms, including NN, is exponential in dimension.
- A basic step in reducing this "curse of dimensionality".

Normal Distribution

- Gaussian distribution (Normal distribution): $X \sim N(\mu, \sigma^2)$.
- $\mathbb{E}[X] = \mu$, $\mathrm{Var}\left[X\right] = \sigma^2$.
- Probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad .$$

- Standard Normal distribution: $N(0, 1)$.
- Stability: Sum of independent normally distributed variates is normally distributed. $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$, X_i's independent. Then,

$$X_1 + \ldots + X_k \sim N(\mu_1 + \ldots + \mu_k, \sigma_1^2 + \ldots + \sigma_k^2) \quad .$$

Gamma distribution

- Gamma(k, θ), $k =$ Gamma parameter, $\theta =$ scale factor (non-negative).
- Pdf: $f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta}$.
- $\mathbb{E}[X] = k\theta$.
- If $X \sim N(0, \sigma^2)$, then, $X^2 \sim$ Gamma$(1/2, 2\sigma^2)$.
- Scaling Property: If $X \sim$ Gamma(k, θ), then, $aX \sim$ Gamma$(k, a\theta)$.
- Sum of Gamma variates is Gamma distributed **if** scale factors are same.
  Let $X_i \sim$ Gamma(k_i, θ) and independent. Then,

$$X_1 + \ldots + X_r \sim \text{Gamma}(k_1 + k_2 + \ldots + k_r, \theta) .$$

Application to estimating F_2: Gaussian sketches

- Let $\xi(j) \sim N(0,1)$ for $j \in [n]$.
- $\xi(j)$'s are (fully) independent. Ignore randomness/space/time required for now.
- Consider sketch
$$X = \sum_{i=1}^{n} f_i \xi(i) \quad .$$
- By stability property of normal distr.
$$X \sim N(0, F_2) \quad .$$
- Problem reduces to: Estimate variance of X.

Gaussian sketches

- Let $X_1, X_2, \ldots, X_t$ be independent Gaussian sketches.
- Define
$$Y = X_1^2 + \ldots + X_t^2 \ .$$
- Each $X_j^2 \ \sim \ \text{Gamma}\,(1/2, 2F_2)$. Therefore,
$$Y \ \sim \ \text{Gamma}(t/2, 2F_2) \ .$$
- $\mathbb{E}\left[Y\right] = tF_2$.
- Need Tail probabilities:
$$\Pr\left\{Y > (1+\epsilon)F_2\right\} \text{ and } \Pr\left\{Y < (1-\epsilon)F_2\right\}.$$

Tail Bounds for Gamma Distribution

Property. Let $Y \sim \mathit{Gamma}(t, \theta)$. Then, for $\epsilon < 1$,

$$\Pr\{Y \in (1 \pm \epsilon)\mathbb{E}[Y]\} \leq \frac{2e^{-\epsilon^2 t/6}}{\epsilon\sqrt{2\pi(t-1)}}.$$

- Let $Y = (X_1^2 + \ldots + X_{2t}^2)/t \sim \text{Gamma}(t, F_2/t)$.
- Let $t = O(\epsilon^{-2}\log(m))$.
- By concentration property,

$$Y \in (1 \pm \epsilon)F_2 \text{ with prob. } 1 - \frac{1}{8m^2} \quad .$$

Another view of mapping: J-L Lemma

- $t \times n$ matrix A, entries $z_{i,j}$ drawn from $N(0,1)$ i.i.d.

$$A = \begin{bmatrix} z_{1,1} & z_{1,2} & \dots & z_{1,n} \\ z_{2,1} & z_{2,1} & \dots & z_{2,n} \\ & \vdots & \vdots & \\ z_{t,1} & z_{t,2} & \dots & z_{t,n} \end{bmatrix}$$

- $x \in \mathbb{R}^n$, $x \mapsto Ax$, $\|Ax\|_2 \in (1 \pm \epsilon)\|x\|_2$ with prob. $1 - 1/m^{O(1)}$.
- By linearity, $A(x - y) = Ax - Ay$.
- Let $t = O(\epsilon^{-2} \log m)$. For any set S of m points,

$$\|Ax - Ay\|_2 \in (1 \pm \epsilon)\|x - y\|_2, \quad \forall x, y \in S$$

with probability $1 - 1/m^2$.

Other Applications of Sketching

- Estimating ℓ_p norms for $0 < p < 2$.
- Heavy Hitters: HH_p^{ϕ}
  - If $|f|_i > \phi \|f\|_p$, then, $i \in \mathrm{HH}_p^{\phi}$.
  - Parameter ϕ': If $|f|_i < \phi' \|f\|_p$, then, $i \notin \mathrm{HH}_p^{\phi,\phi'}$.
- Estimating ℓ_p norms for $p > 2$.

Conclusion (Sketching Streams)

THANK YOU!