# Administrative

1) Prerequisites:
   - Probability
   - Linear algebra
   - Mathematical maturity!!!
2) Grading
   - 2–3 homeworks 50%
   - Project: 50%
3) Google Group
   - Write your e-mails
   - For private e-mails: ilyaraz@microsoft.com

---

What is the class about?

Algorithms through geometric tools.

"Geometry" is not the school geometry

Boundary between "geometry" and combinatorics, probability and analysis is sort of fuzzy.

Topics:

- Sketching / streaming
- Dimension reduction + NLA
- Similarity search
- SDPs for graphs
- Spectral graph algorithms.
- Metric embeddings & applications
- Distance oracles
- Discrepancy minimization

---

Crash course in probability.

1) Random variables

$\Omega$ - probability space

$P(.)$ - probability measure.

$\Omega$ - finite or countable $\quad p_1, p_2, \dots$

$\mathbb{R} \qquad f(t) \qquad P(A) = \int_A f(t)\,dt.$

2) Independence: $P(A \cap B) = P(A) \cap P(B)$

$X, Y \qquad \{X \leq t\} \qquad \{Y \leq s\}.$

(mention k-indep.).

3) Expectation

$$E[X] = \cancel{\sum_i x_i \cdot p_i} \; \sum_v v \cdot \Pr[X=v]$$

$$E[X] = \int v\, f(v)\, dv$$

4) Linearity of expectation

$$E[X+Y] = E[X] + E[Y]$$

$$E[XY] = E[X]E[Y]$$

for indep.

5) Markov:

$X \geq 0 \qquad t \geq 0$

$$\Pr[X \geq t] \leq \frac{E[X]}{t}$$

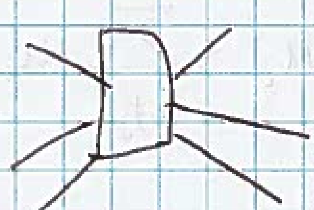
6) Variance: $\mathrm{Var}[X] = E[X^2] - (E[X])^2 \geq 0.$

7) Chebyshev:

$\mathrm{Var}[X+Y] =$
$= \mathrm{Var}[X] + \mathrm{Var}[Y].$

$$\Pr[|X - E[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

Router problem:



packets

src IP

dst IP

payload.

IS there a DDOS attack

$$\sum_{ip} \# \text{times}(ip)^2 = \|u\|_2^2$$

$$u_{ip} = \# \text{times}(u_{ip})$$

$$i_1, \ldots, i_m \in \{1, \ldots, n\}$$

$$u_i \mathrel{+}= 1$$

---

Idea: maintain $X = \sum_i \sigma_i u_i$ $\quad X^2$ is an estimator.

$\sigma_1, \ldots, \sigma_n \in \{\pm 1\}$ uniform independent.

next $i$ comes

$X = X + \sigma_i$. $\quad$ NB: can't store $\sigma_i$'s, will need hash functions.

Analysis:

$$E[X^2] = E\left[\left(\sum_i \sigma_i u_i\right)^2\right] = \sum_i u_i^2 + \sum_{i \neq j} u_i u_j E[\sigma_i \sigma_j] =$$

$$= \|u\|_2^2$$

On average correct.

$$E[X^4] = \sum_i u_i^4 + 3\sum_{i \neq j} u_i^2 u_j^2$$

$$Var[X^2] \leq E[X^4] \leq 4\|u\|_2^4$$

Chebyshev:

$$Pr\left[\left|X^2 - \|u\|_2^2\right| \geq t \cdot \|u\|_2^2\right] \leq \frac{Var[X^2]}{t^2} = \frac{4\|u\|_2^2}{t^2\|u\|_2^2}$$

$$Pr\left[\left|X^2 - F_2\right| \geq 7 \cdot F_2\right] \leq 0.1$$

---

Tug-of-War ++

$$Z = \frac{X_1^2 + \ldots + X_k^2}{k}$$

$$E[Z] = F_2$$

$$Var[Z] \leq O(F_2^2/k)$$

$$Pr\left[|Z - F_2| \geq t \cdot F_2\right] \leq O\left(\frac{1}{t^2 k}\right)$$

$t \sim \varepsilon \qquad k \sim \frac{1}{\varepsilon^2}$

$$\frac{1}{\varepsilon^2}\ \overbrace{\boxed{\quad \pm 1 \quad}}^{n}$$