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
\[ \Omega \text{ - probability space} \]
\[ P(.) \text{ - probability measure} \]
\[ \Omega \text{ - finite } p_1, p_2, \ldots \]
or countable
\[ \mathbb{R} \]
\[ f(t) \]
\[ P(A) = \int_A f(t) \, dt. \]

2) Independence:
\[ P(A \cap B) = P(A) \cap P(B) \]
\[ X, Y \]
\[ \{X \leq t\} \cap \{Y \leq s\} \]
(mention k-indep.)
3) Expectation

\[ E[X] = \sum \nu \cdot \Pr[X = \nu] \]

\[ E[X] = \int \nu f(\nu) d\nu \]

4) Linearity of expectation

\[ E[X + Y] = E[X] + E[Y] \]

\[ E[XY] = E[X]E[Y] \]

for indep.

5) Markov:

\[ X \geq 0 \quad t > 0 \]

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

6) Variance:

\[ \text{Var}[X] = E[X^2] - (E[X])^2 \geq 0 \]

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \]

7) Chebyshev:

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

Router problem:

- packets
- src IP
- dst IP
- payload

Is there a DDoS attack?
\[
\sum_{i \in p} \# \text{times}\ (u_{i}) = \|u\|_2^2
\]

\[
u_{i} = \# \text{times}\ (u_{i})
\]

\[i_1, \ldots, i_m \in \{1, \ldots, n\}\]

\[
u_i + = 1
\]

\[
\text{Idea: maintain } X = \sum \sigma_i u_{i} \quad \text{where } \sigma_i \text{ is an estimator.}
\]

\[\sigma_1, \ldots, \sigma_n \in \{\pm 1\} \text{ uniform independent.}\]

\[\text{next } i \text{ comes}\]

\[X = X + \sigma_i. \quad \text{NB: can't store } \sigma_i \text{ is}
\]

\[\text{will need hash functions.}\]

\[\text{Analysis:}\]

\[
E[X^2] = E\left[\left(\sum \sigma_i u_{i}\right)^2\right] = \sum \sigma^2_i u_i^2 + \sum u_i u_j E[\sigma_i \sigma_j] = \|u\|_2^2
\]

\[\text{On average correct.}\]
\[ E[X^n] = \sum_i u_i^n + 3 \sum_{i \neq j} u_i^2 u_j^2 \]
\[ \text{Var} [X^2] \leq E[X^4] \leq \| u \|^4 \]

Chebyshev:
\[ \Pr \left[ |X^2 - \| u \|^2| \geq \epsilon \right] \leq \frac{\text{Var} [X^2]}{\epsilon^2} = \frac{\| u \|^2}{\epsilon^2 \| u \|^2} \]

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

\[ T \text{ug - of } \text{Var} \text{ + +} \]
\[ Z = \frac{X_1^2 + \ldots + X_k^2}{k} \]

\[ E[Z] = F_2 \]
\[ \text{Var} [Z] \leq O \left( \frac{F_2}{k} \right) \]
\[ \Pr \left[ |Z - F_2| \geq \epsilon \cdot F_2 \right] \leq O \left( \frac{1}{\epsilon^2 k} \right) \]

\[ t \sim \epsilon \quad k \sim \frac{1}{\epsilon^2} \]

\[ \begin{array}{c}
\frac{1}{\epsilon^2} \\
\pm 1
\end{array} \]