class recap

Lecture #1: Randomization, probability, hashing

Lecture #2: Advanced hashing and concentration bounds

Lecture #3: Sketching, searching, and the curse of dimensionality

- Homework #2 due next Thursday
- Project preproposals due next Sunday

class project

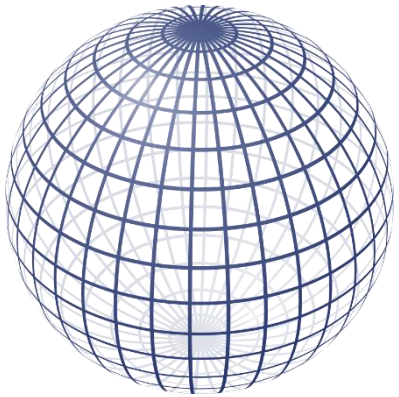
- 40% of the course grade
- Project **must** be done in pairs (with one other student)
- Use discussion board to find a partner if you don't have one
- Makes sense to look through some topics first so you can pair up based on joint interests
- Project must involve algorithm design & analysis in a non-trivial way
- Can include modeling, implementation, experiments
- You will generate a 10-page report by week 8 of the quarter
- It will be posted to the class discussion board
- You will write short evaluations of each others' reports

- Preproposal:
  - Include at least two citations to papers you'll be drawing from
  - Should contain a description of the problem you will be studying, the application domain if it's relevant, and your plan of attack (how will you say something interesting?)
  - New models? New algorithms? Implementation? Evaluation on specific data sets? Etc.

EXERCISE

Is it concentrated? [why or why not?]

#1: Choose a uniformly random vector $X \in \mathbb{R}^n$ with $\|X\| = \sqrt{X_1^2 + X_2^2 + \cdots + X_n^2} = 1$



What is $\mathbb{E}[X_1^2]$?

What is the typical value of the maximum: $\max(|X_1|, |X_2|, \ldots, |X_n|)$?

Is it concentrated? [why or why not?]

#2: Rich get richer? Suppose we have N people. Everyone starts with 1 dollar.

We assign N^2 more dollars in rounds.

***i*th round:** If person j already has n_j dollars, we give them the ith dollar with probability

$$\frac{n_j}{i-1}$$

i.e., with probability to the proportional the amount of money they already have.

Let X_i be the amount of money person i ends up with.

What is the typical value of X_1? Is X_1 concentrated?

What is the typical value of $\max(X_1, X_2, \dots, X_n)$? Is it concentrated?