COT 6936: Topics in Algorithms

Giri Narasimhan
ECS 254A / EC 2443; Phone: x3748
giri@cs.fiu.edu
http://www.cs.fiu.edu/~giri/teach/COT6936_S10.html
https://online.cis.fiu.edu/portal/course/view.php?id=427

Birthday Paradox

Probability that m balls are put in distinct bins is

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)=\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right).$$

Birthday Paradox

- To achieve probability ≥ ½, we need:
  - $m^2/2n \geq \ln 2$
  - m ≥ sqrt{2n ln 2}
- In a room with at least 23 people, the probability that at least two people have the same birthday is more than ½.

Balls and Bins Model

- Throw m balls into n bins
- Location of each ball chosen independently and uniformly at random
- Interesting questions to ask
  - How many balls in a bin on the average?
  - How many bins are empty?
  - How many balls in the fullest bin?
  - How many bins are expected to have > 1 ball in it?
- Applications: Hashing with Chaining

Birthday Paradox

1/12/10 COT 6936 4

Average Size of a Chain in Hash Table

- Let N = # of possible hash values
- Let k = # items stored in the hash table
- Probability that exactly i out of k items hash to the same value is

$$p_i = \binom{k}{i}(N-1)^{k-i}N^{-k}.$$

2/11/10 COT 6936 5

Average Search Time

Unsuccessful Search:

$$\begin{aligned} A &= \sum_i (i+1)p_i = \sum_i \binom{k}{i}(i+1)(N-1)^{k-i}N^{-k} \\ &= \sum_i \binom{k}{i} i(N-1)^{k-i}N^{-k} + \sum_i \binom{k}{i}(N-1)^{k-i}N^{-k} \\ &= \sum_i k\binom{k-1}{i-1}(N-1)^{k-i}N^{-k} + 1 \\ &= kN^{-k}\sum_i \binom{k-1}{i}(N-1)^{k-i-1} + 1 \\ &= kN^{-k}N^{k-1} + 1 = 1 + k/N \end{aligned}$$

Successful Search:

$$A' = \sum_{i,j} j q_{ij} = 1 + \frac{k-1}{2N}$$

2/11/10 COT 6936 6

Maximum Load

- Prob that a bin has at least j balls is

$$\binom{n}{j}\left(\frac{1}{n}\right)^j \le \frac{1}{j!} \le \left(\frac{e}{j}\right)^j$$

Maximum Load: most balls in any bin

- Prob that one of n bins has at least j = (3 ln n / ln ln n) balls is

$$\begin{aligned} n\left(\frac{e}{j}\right)^j &\le n\left(\frac{e \ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n} \\ &\le n\left(\frac{\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n} \\ &= e^{\ln n}(e^{\ln\ln\ln n - \ln\ln n})^{3\ln n/\ln\ln n} \\ &= e^{-2\ln n + 3(\ln n)(\ln\ln\ln n)/\ln\ln n} \\ &\le \frac{1}{n} \end{aligned}$$

Power of Two Choices

- Each ball comes with d = 2 possible bins, each chosen independently at random
- Ball is placed in the least full bin among the d choices
  - ties broken arbitrarily
- MAGICALLY, with high prob:
  - MAX LOAD = ln ln n / ln 2 + O(1)
  - Down from Θ(ln n / ln ln n) (when d = 1)
  - In general, when d ≥ 2,
    - MAX LOAD = ln ln n / ln d + Θ(1)

Applications

- Hashing with 2-way chaining
  - 2 hash function applied to each data item
  - Item inserted in shorter of two chains
- Dynamic Resource Allocation
  - Choosing a server among servers in a network
  - Choosing a disk to store an entity
  - Choosing a printer to serve a print job