{ Csc 80030 | Lecture 7 }

Probabilistic Analysis & Randomized Algorithms

Hoeteck Wee · hoeteck@cs.qc.edu

PART 1 | power of two choices

Balls-and-Bins with Multiple Choices

- n balls thrown into n bins
- balls arrive sequentially, pick d random bins, place in least loaded
- previously, $d = 1$: max load $O(\log n / \log \log n)$
- TODAY $d = 2$: max load $O(\log \log n)$
  - arbitrary tie breaking
  - sampling *with* replacement
  - recent work: Talwar and Wieder [STOC 2007]; Godfrey [SODA 2008]

Analysis (step one)

- STEP ONE: a graph representation
  - n vertices (one for each bin)
  - $n/8$ edges: for each ball, connect the two random bins with an edge
  - ignore ordering of the edges for now
- CLAIM 1: w.p. $1 - O(1/n)$, every connected component in the graph has size $O(\log n)$
- CLAIM 2: w.p. $1 - O(1/n)$, the average degree in every induced subgraph is at most 6.
- FACT: $(n/k)^k \leq \binom{n}{k} \leq (ne/k)^k$

Analysis (step one)

- CLAIM 1: w.p. $1 - O(1/n)$, every connected component in the graph has size $O(\log n)$
- PROOF

  connected component of size $\geq k$ means

  $\exists$ a subset of k vertices with $\geq k-1$ internal edges

$$
\begin{aligned}
& \sum_{|S|=k, |E|=k-1} \Pr[E \text{ is a set of } k-1 \text{ edges with both end-points in } S] \\
\leq & \sum_{|S|=k, |E|=k-1} ((k/n)^2)^{k-1} \\
= & \frac{n^2}{k^2}\binom{n}{k}\binom{n/8}{k-1}\left(\frac{k}{n}\right)^{2k} \leq \frac{n^2}{k^2}\left(\frac{ne}{k}\right)^k\left(\frac{ne}{8k}\right)^k\left(\frac{k}{n}\right)^{2k} \leq n^2\left(\frac{e^2}{8}\right)^k
\end{aligned}
$$

Analysis (step one)

- CLAIM 2: w.p. $1 - O(1/n)$, the average degree in every induced subgraph is at most 6.
- PROOF

The complement is, $\exists$ a subset of k vertices with $> 3k$ internal edges.

$$
\begin{aligned}
& \sum_{k=1}^{n} \sum_{|S|=k,|E|=3k} \Pr[E \text{ is a set of } 3k \text{ edges with both end-points in } S] \\
\leq \quad & \sum_{k=1}^{n} \binom{n}{k} \binom{n/8}{3k} \left(\frac{k}{n}\right)^{2\cdot 3k} \leq \sum_{k=1}^{n} \left(\frac{ne}{k}\right)^{k} \left(\frac{ne}{24k}\right)^{3k} \left(\frac{k}{n}\right)^{6k} = \sum_{k=1}^{n} \left(\frac{e^4}{24^3} \cdot \frac{k^2}{n^2}\right)^{k} \\
\leq \quad & \sum_{k=1}^{\Theta(\log n)} \frac{O(\log^2 n)}{n^2} + \sum_{k \leq n} \frac{1}{n^2} = O(1/n)
\end{aligned}
$$

Analysis (step one)

- STEP ONE: a graph representation
  - n vertices (one for each bin)
  - $n/8$ edges: for each ball, connect the two random bins with an edge
- CLAIM 1: every connected component has size $O(\log n)$
- CLAIM 2: the average degree in every induced subgraph is at most 6.
- NEXT, iterative vertex removal

Analysis (step two)

- STEP TWO: iterative vertex removal
- at each iteraction, remove all vertices (& incident edges) with deg ≤ 12
- OBSERVATION 1: remove at least $1/2$ the vertices in each iteration.
  - CLAIM 2 says average deg is ≤ 6, so less than $1/2$ have deg > 12.
- OBSERVATION 2: terminates in $\log \log n + O(1)$ iterations.
  - CLAIM 1 says initially, every connected component is size $O(\log n)$
  - apply previous observation to each connected component.
- CLAIM 3: The load of a bin (vertex) removed in iteration i is $\leq 13i$.

Analysis (step two)

- CLAIM 3: The load of a bin (vertex) removed in iteration i is $\leq 13i$.
- PROOF by induction.
  - Base case: trivially follows from deg bound.
  - Analyze load of bin v (vertex) removed in iteration $i + 1$,

    3 types of incident edges:

    I. ball does not land in v

    II. deleted in iterations $1, \ldots, i$ $\longleftarrow$ any number of such edges

    III. deleted in iteration $i + 1$ $\longleftarrow$ at most 12 of these
  - Exploit ordering: examine loads right after the last type II edge (v, w).
    - w is deleted in iteration $\leq i$, so bin w has load at most $13i$.
    - Ball lands in bin v means bin v now has load $\leq 13i + 1$.
    - At most 12 type III edges after that

From $n/8$ balls to n balls

- CLAIM: Suppose when tossing $n/8$ balls into n bins, the max load is L w.p. at least $1 - \delta$. Then, when tossing n balls into n bins, the max load is $8L$ w.p. at least $1 - 8\delta$.
- PROOF:
  - View tossing n balls as running 8 phases of tossing $n/8$ balls
  - By a union bound, prob. that max load exceeds L in any of the 8 phases is at most 8δ.

THE END | next, ...