# CMPSCI 240: Reasoning about Uncertainty

## Lecture 19: Balls and Bins and Hashing

Andrew McGregor

University of Massachusetts

Last Compiled: April 11, 2017

# Outline

1 Balls into Bins

# Balls and Bins

Throw $m$ balls into $n$ bins where each throw is independent.

- How large must $m$ be such that it is likely there exists a bin with at least two balls? (Birthday Paradox)

  $$P(\text{all bins have at most one ball}) = \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \ldots \times \frac{n-m+1}{n}$$

  e.g., for $n = 365$ and $m \geq 23$ there is a greater than $1/2$ chance that there exists a bin with two or more balls.

- How large must $m$ be such that it is likely that all bins get at least one ball? (Coupon Collecting)

  $$P(\text{there exists an empty bin}) \leq ne^{-m/n}$$

  e.g., for $m = 2n \ln n$ this probability is at most $1/n$.

# Balls and Bins

Two useful facts:

- Union bound: If we have a set of events $A_1, A_2, \ldots, A_n$ then

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n)$$

- Binomial Distribution: If we throw $m$ balls and the probability of landing in a specific bin is $p$ then the number of balls in this bin is a binomial distribution with $m$ trials and probability of success $p$.

# Balls and Bins: Load Balancing

Throw $m$ balls into $n$ bins where each throw is independent.

- How full is the fullest bin? This has applications to load balancing.
- What's the probability that $k$ or more items land in bin $j$?
- If $X$ is the number of balls that land in bin $j$ then $X$ is a binomial distribution with $m$ trials and $p = 1/n$.
- Lemma: $P(X \geq k) \leq \binom{m}{k} p^k$.
- If $m/n = 1$ and $k = 2\log n$,

$$P(X \geq k) \leq \binom{m}{k} p^k \leq \frac{m^k}{k!} \cdot \left(\frac{1}{n}\right)^k = \left(\frac{m}{n}\right)^k / k! = 1/k! \leq 1/2^k = 1/n^2$$

and hence no bin has more than $k = 2\log n$ balls in it with probability at least $1 - 1/n$.

# Proof of Lemma

- Suppose we toss $m$ coins with probability of heads equal to $p$.
- For any set $S \subseteq [m]$, let $A_S$ be the event that $i$th coin toss was heads for all $i \in S$.
- Let $S_1, S_2, \ldots S_{\binom{m}{k}}$ be all subsets of $[m]$ with exactly $k$ elements.

$$P(A_{S_j}) = p^k$$

- Then $A_{S_1} \cup A_{S_2} \cup \ldots \cup A_{S_{\binom{m}{k}}}$ is the event you get $k$ or more heads.
- Hence,

$$P(k \text{ or more heads}) = P(A_{S_1} \cup A_{S_2} \cup \ldots \cup A_{S_{\binom{m}{k}}}) \leq \sum_{j=1}^{\binom{m}{k}} P(A_{S_j}) = \binom{m}{k} p^k$$