CMPSCI 611: Advanced Algorithms
Lecture 17: Balls and Bins and Schwartz-Zippel

Andrew McGregor

Last Compiled: November 2, 2020
Balls and Bins

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

- **Birthday Paradox**: How large can $m$ be such that all bins have at most one ball? Applications: Picking IDs without coordination in a Multi-Agent System.
- **Coupon Collecting**: How large must $m$ be such that all bins get at least one ball?
- **Load Balancing**: What is the maximum number of balls that fall into the same bin? Application: Assigning jobs to different machines without overloading any machine.
Birthday Paradox

Lemma
\[ P \left[ \text{first } m \text{ balls fall in distinct bins} \right] \leq e^{-m(m-1)/(2n)}. \]

Proof.

- Let \( A_i \) be event that the \( i \)-th ball lands in a bin not containing any of the first \( i - 1 \) balls.
- \[ P \left[ \cap_{1 \leq i \leq m} A_i \right] = P[A_1]\, P[A_2|A_1]\, \ldots \, P[A_m| \cap_{1 \leq i \leq m-1} A_i] \]
- \[ P[A_i| \cap_{1 \leq j \leq i-1} A_j] = 1 - (i - 1)/n \]
- Putting it together and using \( \sum_{1 \leq i \leq a} i = (a + 1)a/2 \):

\[
P \left[ \cap_{1 \leq i \leq m} A_i \right] = \prod_{1 \leq i \leq m} \left( 1 - \frac{i - 1}{n} \right) \leq e^{-\sum_{i=1}^{m} \frac{i-1}{n}} = e^{-m(m-1)/(2n)}
\]

With \( n = 365 \) and \( m = 29 \), probability \( < e^{-1} \). Tighter analysis possible.
Coupon Collecting

Suppose you throw \( r \) balls into \( n \) bins. If each ball is equally likely to land in each bin, how large does \( r \) need to be such that a ball lands in every bin with probability at least \( 1 - 1/n \). We’ll show \( r = 2n \ln n \) are sufficient.

- Let \( A_i \) be the event that the \( i \)th bin is empty after \( r \) balls are thrown. Then,

\[
\mathbb{P}[A_i] = (1 - 1/n)^r = (1 - 1/n)^{2n \ln n} \leq e^{-2 \ln n} = 1/n^2
\]

- Then \( A_1 \cup A_2 \cup \ldots \cup A_n \) is the event that there is an empty bin:

\[
\mathbb{P}[A_1 \cup A_2 \cup \ldots \cup A_n] \leq \mathbb{P}[A_1] + \mathbb{P}[A_2] + \ldots + \mathbb{P}[A_n] = n \times 1/n^2 = 1/n
\]
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 1?

- If $X_1$ is the number of balls that land in bin 1 then $X_1$ is a binomial distribution with $m$ trials and $p = 1/n$.

- **Lemma:** $P(X_1 \geq k) \leq \binom{m}{k} p^k$.

- If $m/n = 1$ and $k = 2 \log n$,

  
  \[ P(X_1 \geq k) \leq \binom{m}{k} p^k \leq \frac{m^k}{k!} \cdot \left(\frac{1}{n}\right)^k = \left(\frac{m}{n}\right)^k / k! = 1/k! \leq 1/2^k = 1/n^2 \]

- Same analysis applies to $X_2, X_3, \ldots$, i.e., the number of balls in bins 2, 3, \ldots. Hence, no bin has more than $k = 2 \log n$ balls in it with probability at least $1 - 1/n$. 

---

6/11
Lemma

Let $X$ be the number of heads observed when we toss $m$ coins each with probability of heads equal to $p$. Then $\Pr[X \geq k] \leq \binom{m}{k} p^k$.

- Let $S_1, S_2, \ldots S_{\binom{m}{k}}$ be all subsets of $[m]$ with exactly $k$ elements.

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

  where $A_S$ is the event that for all $i \in S$, the $i$th coin toss is heads.

- 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,

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

Balls and Bins and Birthdays and Coupons

Schwartz-Zippel
Checking Polynomial Multiplication via Schwartz-Zippel

Problem

Given three $n$ variable polynomials $P_1, P_2, P_3$. Can you test if

$$P_1(x_1, \ldots, x_n) \times P_2(x_1, \ldots, x_n) = P_3(x_1, \ldots, x_n)$$

faster than multiplying the polynomials? Equivalently, is

$$Q(x_1, \ldots, x_n) = P_1(x_1, \ldots, x_n) \times P_2(x_1, \ldots, x_n) - P_3(x_1, \ldots, x_n)$$

zero for all $x_1, \ldots, x_n$?

Theorem (Schwartz-Zippel)

Let $Q(x_1, \ldots, x_n)$ be a non-zero multivariate polynomial of total degree $d$. Fix any finite set of values $S$ and let $r_1, \ldots, r_n$ be chosen independently and uniformly at random from $S$. Then,

$$\mathbb{P}[Q(r_1, \ldots, r_n) = 0] \leq d/|S|$$
Schwartz-Zippel Proof

- Induction on $n$: For $n = 1$, because $Q$ has at most $d$ roots,
  \[ \mathbb{P}[Q(r_1) = 0] \leq d/|S| \]
  
- For induction step define $Q_i$ for $0 \leq i \leq k$:
  \[ Q(x_1, \ldots, x_n) = \sum_{i=0}^{k} x_1^i Q_i(x_2, \ldots, x_n) \]
  where $k$ is maximum such that $Q_k(x_2, \ldots, x_n) \neq 0$

- Since total degree of $Q_k$ is at most $d - k$,
  \[ \mathbb{P}[Q_k(r_2, \ldots, r_n) = 0] \leq (d - k)/|S| \]

- Consider $q(x) = Q(x, r_2, \ldots, r_n)$,
  \[ \mathbb{P}[q(r_1) = 0|Q_k(r_2, \ldots, r_n) \neq 0] \leq k/|S| \]

- Putting together gives $\mathbb{P}[Q(r_1, \ldots, r_n) = 0]$ at most
  \[ \mathbb{P}[Q_k(r_2, \ldots, r_n) = 0] + \mathbb{P}[q(r_1) = 0|Q_k(r_2, \ldots, r_n) \neq 0] \leq d/|S| \]
  where we used $\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \cap B^c] \leq \mathbb{P}[B] + \mathbb{P}[A|B^c]$