Last time
• Finished routing on hypercube
• Balls-and-Bins model

Today
• Application of Balls-and-Bins model: Bucket Sort
• Poisson distribution
• Poisson approximation
Bucket Sort

- **Given:** $n$ integers from range $[r]$.
- **If** $r \leq n$, we can sort in $O(n)$ time
  - Use possible values as buckets
  - Keep a linked list for each bucket.
  - Make a pass over the list and put each element in the right bucket
  - Concatenate the lists.
- **What if** $r > n$? (Suppose for simplicity that $n$ divides $r$.)

**Theorem**

| If $n$ integers are chosen u.i.r. from range $[r]$, they can be sorted in expected time $O(n)$. |

- **Expectation is over randomness in choice of integers:** Bucket Sort is deterministic.
Bucket Sort

- **Idea:** Break the range \([r]\) into \(n\) buckets.
  The expected # of elements in each bucket is 1.
  We can easily sort all buckets (say, using Insertion Sort)

<table>
<thead>
<tr>
<th>Algorithm. Input: integers (a_1, \ldots, a_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Make linked lists for buckets (B_1, \ldots, B_n).</td>
</tr>
<tr>
<td>2. For each (i \in [n]), let (j = \left\lfloor \frac{a_i \cdot n}{r} \right\rfloor) and add (a_i) to (B_j).</td>
</tr>
<tr>
<td>3. Sort all buckets using Insertion Sort.</td>
</tr>
<tr>
<td>4. Output the concatenation of (B_1, \ldots, B_n)</td>
</tr>
</tbody>
</table>

- Steps 1, 2, and 4 can be implemented to run in \(O(n)\) time.

**Lemma**

Step 3 runs in expected time \(O(n)\).
Bucket Sort: Analysis

**Lemma**

Step 3 (sorting the buckets) runs in expected time $O(n)$.

**Proof:** Buckets are bins, elements are balls.

- Let $X_j = \#$ of elements that land in bucket $B_j$, for $j \in [n]$.
- Time to sort $B_j$ is: $\leq c \cdot X_j^2$ for some constant $c$.
- Expected run time of Step 3:

  $$\leq \mathbb{E}
  \left[
  \sum_{j \in [n]} cX_j^2
  \right] = c \cdot \sum_{j \in [n]} \mathbb{E}[X_j^2] = cn \cdot \mathbb{E}[X_1^2]$$

  *by linearity of expectation*
  *by symmetry*
The number of empty bins

*m balls into n bins*

- The probability that bin 1 is empty is

- Expected number of empty bins

  \[ X = \text{the number of empty bins} \]

  \[ X_i = \text{indicator for the event that Bin } i \text{ is empty, } i \in [n] \]

  \[ \mathbb{E}[X] = \]

  \[ e^{-x}(1 - x^2) \leq 1 - x \leq e^{-x} \quad \text{for } |x| \leq 1 \]
The number of bins with $k$ balls

$m$ balls into $n$ bins, $k$ is a small constant

- The probability $p_k$ that bin 1 has $k$ balls is

$$p_k =$$
A Poisson random variable with parameter \( \mu \) is given by the following distribution on \( j = 0, 1, 2, \ldots \):

\[
\Pr[X = j] = \frac{e^{-\mu} \mu^j}{j!}
\]

Check that probabilities sum to 1:

\[
\sum_{j=0}^{\infty} \Pr[X = j] = \sum_{j=0}^{\infty} \frac{e^{-\mu} \mu^j}{j!} =
\]

The expectation of a Poisson R.V. \( X \) is

\[
\mathbb{E}[X] =
\]

\[
\text{var}[X] = \mu \quad \text{(See Ex. 5.5)}
\]
Independent Poisson RVs

Theorem
Let $X$ and $Y$ be independent Poisson RVs with means $\mu_X$ and $\mu_Y$. Then $X + Y$ is a Poisson RV with mean $\mu_X + \mu_Y$. 
Theorem. Let $X$ be a Poisson RV with mean $\mu$.

• (upper tail, additive) If $x > 0$, then
  \[
  \Pr[X \geq \mu + x] \leq \frac{e^{-\mu} (e\mu)^x}{x^x}.
  \]

• (lower tail, additive) If $x < \mu$, then
  \[
  \Pr[X \leq x] \leq \frac{e^{-\mu} (e\mu)^x}{x^x}.
  \]

• (upper tail, multiplicative) For any $\delta > 0$,
  \[
  \Pr[X \geq (1 + \delta)\mu] \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu.
  \]

• (lower tail, multiplicative) For any $\delta \in (0,1)$,
  \[
  \Pr[X \leq (1 - \delta)\mu] \leq \left( \frac{e^\delta}{(1 - \delta)^{1-\delta}} \right)^\mu.
  \]
Poisson Distribution is Limit of Binomial Distribution

**Theorem**

Let $X_n \sim \text{Bin}(n, p)$, where $p$ is a function of $n$ and $\lim_{n \to \infty} np = \mu$, a constant independent of $n$. Then, for all fixed $k$,

$$\lim_{n \to \infty} \Pr[X_n = k] = \frac{e^{-\mu} \mu^k}{k!}.$$

- Applies to balls-and-bins if $m = nc$. 

Sofya Raskhodnikova; Randomness in Computing