Double Hashing Analysis

- Assume both hash functions look random
- Assume secondary probes are random
- Then all probes double hashing are independent: a probe doesn’t depend on a previous probe
- Much easier to analyze

Analysis

- What is probability of probing an occupied node?
- What is probability of probing an unoccupied node?
What is probability of an unsuccessful search... 
- ... first probing an occupied cell (but doesn’t match key)... 
- ... and then hitting an empty cell?

Expected # probes for unsuccessful search is

\[ = 1 \cdot \Pr[\text{one probe to empty cell}] + 2 \cdot \Pr[\text{one probe to occupied and one probe to empty cell}] + 3 \cdot \Pr[\text{two probes to occupied and one probe to empty cell}] + 4 \cdot \Pr[\text{three probes to occupied and one probe to empty cell}] + \cdots \]

=
To insert an item, we probe like an unsuccessful search for the item items, then insert it.

After each insert, the load changes

Expected time for each insert is $U_i = \frac{1}{m} = \frac{1}{i/m} = \frac{m}{i}$.

To search for an item, we do same number of probes as what took to insert it.

Hence $S_n$ is average over $\#$ probes to insert previous $i-1$ items.
The Scary Slide

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} U_{i-1} \\
= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 - \alpha_{i-1}} \\
= \frac{1}{n} \sum_{i=1}^{n} \frac{m}{m - 1 + 1} \\
= \frac{m}{n} \sum_{i=1}^{n} \frac{1}{m - 1 + 1} \\
= \frac{m}{n} \left( \frac{1}{m} + \frac{1}{m - 1} + \frac{1}{m - 2} + \cdots + \frac{1}{m - n + 2} + \frac{1}{m - n + 1} \right) \\
= \frac{m}{n} \left( \frac{1}{m} \cdot 1 + \frac{1}{m} \cdot \left( \frac{1}{m - 1} + \frac{1}{m - 2} + \cdots + \frac{1}{m - n} \right) \right) \\
= \frac{m}{n} (H_m - H_{m-n}) = \frac{m}{n} (\ln m - \ln(m-n)) = \frac{1}{\alpha_n} \ln \frac{1}{1 - \alpha_n}
\]

Summary

- \( S_n \approx \frac{1}{n} \ln \frac{1}{\alpha_n} \)
- \( U_n \approx \frac{1}{\alpha_n} \)
- If table (of any size) is 90% full
  \[
  S_n \approx \frac{1}{n} \ln \frac{1}{\alpha_n} \approx 2.56 \quad \text{Not bad!} \\
  U_n \approx \frac{1}{\alpha_n} \approx 10 \quad \text{Still constant}
  \]
- Compare with linear probing (analysis is harder):
  \[
  S_n \approx \frac{1}{2} \left( 1 + \frac{1}{1 - \alpha_n} \right) \\
  U_n \approx \frac{1}{2} \left( 1 + \left( \frac{1}{1 - \alpha_n} \right)^2 \right)
  \]
  At 90%, \( S_n \approx 5.5 \) and \( U_n \approx 50.5 \).