Yoo Jin (Grace) LEE
CSCI 104L
Prof. Redekopp
HW 9

PROBLEM 3 (Analysis of Hash Tables w/ Chaining)

Part A (7%): How many buckets end up entirely empty in expectation?

each bucket has $1/m$ probability of being filled in by a hash function (hash function maps to random location)

SO: in order for a bucket to be empty, it would be $(1-\frac{1}{m})$ by a single function.

But for a bucket to be empty, $(1-\frac{1}{m})$ should occur for each hash function, leading to $(1-\frac{1}{m})^m$

And since the probability of A bucket being empty is $(1-\frac{1}{m})^m$, according to linearity of expectation, we would multiply this by # of buckets, $\boxed{m(1-\frac{1}{m})^m}$ yielding the # of expected # of empty buckets.

Part B (13%): What's the fullest that a bucket will typically become.

- Probability that k keys get mapped to this bucket. $= (\frac{1}{m})^k$
  → // a key gets mapped to this bucket $= \frac{1}{m}$
- Probability that m-k keys gets / does not get mapped to this bucket. $(1-\frac{1}{m})^{m-k}$
- Probability that ANY set of k keys get mapped to this bucket.

union Bound: $\text{Prob}(\sum_{i=1}^{m} E_i) \le \sum_{i=1}^{m} \text{Prob}(E_i)$.

$$: \binom{m}{k}\left(\frac{1}{m}\right)^k\left(1-\frac{1}{m}\right)^{m-k}.$$

* Find a value for k.

$$\hookrightarrow \frac{\cancel{m^k}}{2^{k-1}}\left(\frac{1}{\cancel{m}}\right)^k\left(1-\frac{1}{m}\right)^{m-k} \le \frac{1}{m^2}$$

$$\frac{1}{2^{k-1}}\left(\frac{m-1}{m}\right)^{m-k} \le \frac{1}{m^2} = \frac{1}{2^{k-1}}\left(\frac{(m-1)^m}{(m-1)^k}\right)\left(\frac{m^k}{m^m}\right) = \frac{1}{m^2}$$

$$\Rightarrow \left(\frac{2}{2^k}\right)\left(\frac{m^k}{(m-1)^k}\right) = \frac{1}{m^2}\left(\frac{m^m}{(m-1)^m}\right) \Rightarrow 2\left(\frac{m}{2(m-1)}\right)^k = \frac{m^{m-2}}{(m-1)^m}$$

take log of both sides.

$$\Rightarrow \log 2 + k\log\left(\frac{m}{2(m-1)}\right) = \log m^{m-2} - \log(m-1)^m$$

$$\Rightarrow \log 2 + k\log\left(\frac{m}{2m-2}\right) = (m-2)\log m - \cancel{\log}\, m\log(m-1)$$

$$\Rightarrow k\log\left(\frac{m}{2m-2}\right) = (m-2)\log m - m\log(m-1) - \log 2.$$

$$k = \frac{(m-2)\log m - m\log(m-1) - \log 2}{\log m - \log(2m-2)}$$

← Fullest # of keys possible in a bucket.