EXERCISE FOR CSE202 – WEEK 7

Consider hashing with separate chaining under the uniformity assumption. Let n be the number of hashed keys, m the size of the table and $\alpha = n/m$ its filling ratio. The aim of this exercise is to bound the probability that the number of keys in a list is much larger than α.

Question 1. *Show that the probability that the ith slot has exactly k keys is*

$$q_k = \binom{n}{k}\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}.$$

Solution : One can think of it as a sequence of n head/tails, k of which are hashed to i. There are $\binom{n}{k}$ such sequences and each has probability $1/m^k(1-1/m)^{n-k}$, due to the uniformity assumption. ☐

Question 2. *Show that an upper bound is*

$$q_k \le e^{-\alpha}(\alpha e/k)^k.$$

[Indication: use the bounds from the lecture.]

Solution : Using the bound of the lecture on the binomial coefficients gives

$$q_k \le \frac{n^n}{k^k(n-k)^{n-k}}\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}.$$

This rewrites as

$$q_k \le \frac{\alpha^k}{k^k}\left(\frac{n-\alpha}{n-k}\right)^{n-k} = \frac{\alpha^k}{k^k}\left(\frac{n-k+k-\alpha}{n-k}\right)^{n-k} = \frac{\alpha^k}{k^k}\left(1+\frac{k-\alpha}{n-k}\right)^{n-k}.$$

Finally, using the inequality $(1+x/m)^m \le e^x$ gives

$$q_k \le e^{-\alpha}\left(\frac{\alpha e}{k}\right)^k.$$

☐

Question 3. *If $s_k = e^{-\alpha}(\alpha e/k)^k$, observe that for any positive integer j, $s_{k+j}/s_k < (\alpha e/k)^j$ and deduce that the probability p_k that the ith slot has $k = t\alpha$ keys* or more *(with $t > e$) is bounded by*

$$\left(\frac{e}{t}\right)^k \frac{e^{-\alpha}}{1-e/t}$$

and thus decreases extremely fast with k.

Solution : The first inequality is a direct consequence of $k+j > k$:

$$\frac{s_{k+j}}{s_k} = (\alpha e)^j \frac{k^k}{(k+j)^{k+j}} < (\alpha e)^j \frac{k^k}{k^{k+j}} = \frac{(\alpha e)^j}{k^j}.$$

Thus the probability bound follows from

$$p_k = \sum_{j\ge 0} q_{k+j} \le \sum_{j\ge 0} s_{k+j} \le e^{-\alpha}\left(\frac{\alpha e}{k}\right)^k \sum_{j\ge 0}\left(\frac{\alpha e}{k}\right)^j = e^{-\alpha}\left(\frac{\alpha e}{k}\right)^k \frac{1}{1-\frac{\alpha e}{k}},$$

where the last geometric series is convergent since $k > e\alpha$. The final result is obtained by expressing k in terms of t. $\square$

For instance, with $\alpha = 3$ and $t = 3$, this bound on the probability is 0.22, while with $\alpha = 3$ and $t = 3.6$, it is less than 1%.