CIS 121 — Data Structures and Algorithms
Homework Assignment 8

Assigned: November 17, 2020 **Due:** November 23, 2020

Note: The homework is due **electronically on Gradescope** on November 23, 2020 by 11:59 pm ET. For late submissions, please refer to the Late Submission Policy on the course webpage. You may submit this assignment up to 2 days late.

A. **Gradescope**: You must select the appropriate pages on Gradescope. Gradescope makes this easy for you: before you submit, it asks you to associate pages with the homework questions. Forgetting to do so will incur a 5% penalty, which cannot be argued against after the fact.

B. **LaTeX**: You must use the LaTex template provided on the course website, or a 5% penalty will be incurred. Handwritten solutions or solutions not typeset in LaTex will not be accepted.

C. **Solutions**: Please write concise and clear solutions; you will get only a partial credit for correct solutions that are either unnecessarily long or not clear. Please refer to the Written Homework Guidelines for all the requirements. Piazza will also contain a complete sample solution in a pinned post.

D. **Algorithms**: Whenever you present an algorithm, your answer must include 3 separate sections. Please see Piazza for an example complete solution.

1. A precise description of your algorithm in English. No pseudocode, no code.
2. Proof of correctness of your algorithm
3. Analysis of the running time complexity of your algorithm

E. **Collaboration**: You are allowed to discuss **ideas** for solving homework problems in groups of up to 3 people but *you must write your solutions independently*. Also, you must write on your homework the names of the people with whom you discussed. For more on the collaboration policy, please see the course webpage.

F. **Outside Resources**: Finally, you are not allowed to use *any* material outside of the class notes and the textbook. Any violation of this policy may seriously affect your grade in the class. If you're unsure if something violates our policy, please ask.

1. [20 pts] Hashing with Chaining

Suppose we have an initially empty hash table $A[1..n]$ with n slots that resolves collisions through chaining. Moreover, suppose we insert exactly n elements into this table such that each element inserted has a probability $\frac{1}{n}$ to be put into slot $A[i]$ *independently of all other elements.* Let R_i be the random variable that counts the number of elements that get hashed to slot $A[i]$.

(a) Give an expression in terms of n and k for $Pr(R_i = k)$ (i.e., the probability that k elements get hashed to slot $A[i]$), where $0 \le k \le n$. *For this problem, your final answer may involve any combinatorial expressions (e.g., factorial, "choose" notation, etc.).*

(b) Prove that

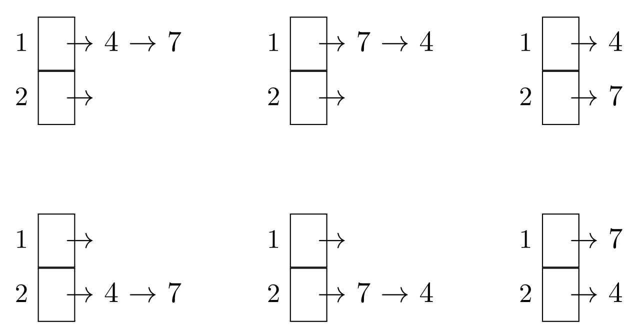
$$Pr(R_i = k) \le \left(\frac{4}{k}\right)^k, \qquad 0 \le k \le n$$

Note, you may find the inequality $\binom{n}{k} \le \left(\frac{n \cdot e}{k}\right)^k$ useful. Here $e \approx 2.718$ is Euler's number, and you can assume the convention that $\left(\frac{4}{0}\right)^0 = 1$.

2. [15 pts] How many hashes?

Suppose we have an initially empty hash table $A[1..m]$ with m slots that resolves collisions through chaining. You may assume that that each element is equally likely to be inserted into each of these m slots independently of all other elements (simple uniform hashing assumption). Given that we insert n distinct elements into the hash table described above, what is the total number of distinct hash tables that could result? *For this problem, your final answer may involve any combinatorial expressions (e.g., factorial, "choose" notation, etc.).*

For example, if $m = 2$ and $n = 2$ and the elements inserted are 4 and 7, then there are six possible hash-tables:



3. [15 pts] Tracing through Hashing

Consider inserting the keys $3, 12, 23, 39, 18, 52, 83, 46, 29$ (in that order) into a hash table of length $m = 12$ using open addressing with the auxiliary hash function $h'(k) = k$. Give the resulting hash table after inserting these keys using:

a. linear probing

b. quadratic probing with $c_1 = 3$ and $c_2 = 2$

c. double hashing with $h_1(k) = k$ and $h_2(k) = 1 + (k \bmod (m-2))$

[**0 pts**] **Feedback:** No feedback form for this assignment. Feel free to reach out to the Head TAs if you have any comments/concerns.