1 Hash tables

Hash tables with universal hash families guarantee an expected runtime of \( O(1) \) for the INSERT, SEARCH, and DELETE operations. What is the meaning of “expected”? This shows the hash family from lecture can be labeled by the optimal number of bits (\( \log M \)) for the hash family.

- It is an average over the choices of the adversary who picks the elements in the table.
- It is an average over the choices of the algorithm who picks the hash function from the hash family.

In order to conclude an expected runtime of \( O(1) \) for hash table operations, we assumed the following two happen in one or some order:

- The adversary picks elements \( u, v, \ldots \), for the hash table.
- The algorithm picks a hash function from the hash family.

In what order do these happen?

- Algorithm first, and then the adversary.
- Adversary first, and then algorithm.
- It does not matter.

2 Bit lengths

Suppose that there is a toy box with \( N \) toys in it. You have a label printer that can print arbitrary strings of \( 0 \)s and \( 1 \)s in such a way that each toy gets a unique label, what can be said about the longest label’s length?

- \( \Omega(\log n) \)
- \( \Omega(\log M) \)
- \( \Omega(\log N) \)
- \( \Theta(\log n) \)
- \( \Theta(\log M) \)
- \( \Theta(\log N) \)

Correct

As a remark, for any labeling scheme, the same lower bound of \( \Omega(\log n) \) applies even to the average label length, not just the longest label length.

If you produce labels in a way that minimizes the longest label’s length, what is this minimum?

- \( \Omega(\log n) \)
- \( \Omega(\log N) \)
- \( \Theta(\log n) \)
- \( \Theta(\log N) \)

Correct

For any toy box consists of all functions from \( 0, \ldots, M−1 \) to \( 0, \ldots, n−1 \), what is the minimum longest label’s length?

- \( \Omega(\log n) \)
- \( \Omega(\log M) \)
- \( \Theta(\log n) \)
- \( \Theta(\log M) \)

Correct

How about \( M > 1000 \) (the universe size) is a prime number. If we pick \( n = \{1, \ldots, M−1\} \), independently and uniformly at random, what is the expected label length, not just the longest label length.

- \( \Omega(\log n) \)
- \( \Theta(\log n) \)
- \( \Omega(\log M) \)
- \( \Theta(\log M) \)

Correct

If you produce labels in a way that minimizes the average label length, what is this minimum?

- \( \Omega(\log n) \)
- \( \Theta(\log n) \)
- \( \Theta(\log M) \)

Correct

3 Modular arithmetic

Suppose that \( M = 1000 \) (the universe size) is a prime number. If we pick \( n = \{1, \ldots, M−1\} \) and \( b = \{0, \ldots, M−1\} \), independently and uniformly at random, what is the probability that \( n \mod b \neq 0 \)?

- \( \frac{1}{29} \)
- \( \frac{1}{30} \)
- \( \frac{1}{31} \)
- \( \frac{1}{32} \)
- \( \frac{1}{33} \)

Correct

In fact, for any pair of distinct elements \( x, y \) in the universe, \( x = a \times x \mod M \) and \( y = a \times y \mod M \) are uniformly distributed amongst all distinct pairs. How many elements of \( \{0, \ldots, M−1\} \) are equal to \( 0 \mod a \)?

- \( |\{n \mod a \mid n = 0\}| \)
- \( |\{n \mod a \mid n = 1\}| \)
- \( |\{n \mod a \mid a \leq n\}| \)
- \( \frac{1}{M/a} \)

Correct

In fact, for any \( a \), the number of elements of \( \{0, \ldots, M−1\} \) equal to \( 0 \mod a \) is \( \lfloor n/M \rfloor \).

Let \( n = 1000 \) pick \( a \) uniformly at random from \( \{0, \ldots, M−1\} \) and \( n \not\equiv 0 \mod a \). What is the chance that \( n \mod a = 0 \)?

- \( \frac{1}{200} \)
- \( \frac{1}{250} \)
- \( \frac{1}{299} \)
- \( \frac{1}{300} \)
- \( \frac{1}{301} \)

Correct

You can verify that this answer above is always \( \geq \frac{1}{2} \). The same answer holds as an upper bound if we changed a from 0 to any other element in \( \{0, \ldots, M−1\} \).

4 Hash family size

Suppose that we have a universe of size \( M \), and our hash table size is \( n \). If \( n \leq M \), what is the minimum size of a universal hash family?

- \( \frac{M}{n} \)
- \( \frac{n}{M} \)
- \( \frac{M}{n} - 1 \)
- \( M - 1 \)

Correct

Suppose now that \( M = 2^{100} \) and we have a non-uniform hash family of \( 1 \leq t \leq 20 \). Let \( H \) be one of the hash functions in \( H \). Since \( M = 2^{100} \), any \( h \) is a \( t \)-distinct elements \( \{x, \ldots, x^t\} \) in the universe to the same bucket (by the pigeonhole principle). What can be said about \( \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \)?

- \( \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{t} \)
- \( \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{20} \)
- \( \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{2} \)
- \( \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{M} \)

Correct

This means that if \( M \) is uneven, then

\[
\frac{1}{n} \sum_{h \in H} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{M} \frac{1}{|\{x, \ldots, x^t\}|} \geq \frac{1}{M} \]

or in other words \( |\{x, \ldots, x^t\}| \leq \frac{M}{n} \). What can be said about the minimum longest \( L(H) \)?

- \( \frac{M}{n} \)
- \( \frac{20}{n} \)
- \( \frac{100}{n} \)
- \( \frac{1000}{n} \)

Correct

For the universal hash family from lecture, how many bits do we need to label the hash functions, if we minimize the longest label’s length?

- \( \Omega(\log n) \)
- \( \Omega(\log M) \)
- \( \Theta(\log n) \)
- \( \Theta(\log M) \)

Correct

This shows the hash family from lecture can be labeled by the optimal number of bits \( \Theta(\log M) \) when \( M = 2^{100} \).