Applied Algorithms Lec 7: Assignment 1 and Probability

Sam McCauley

September 27, 2024

Williams College

Admin

- Colloquium today relates closely to applied algorithms
- Sign up today for the “what I did last summer” colloquiums over the next two weeks (you should have gotten an email)

Homework 1 back

- Code grades usually were A (or close)
- Many students did not do as well on the more theoretical part
- If you get a C on 10% of the homework, and an A on 90% of the homework, that does still average in the A range
- Going into the Assignment: bear in mind that the theory is a significant part of the course! Don’t put it off until your code is perfect

Admin: Assignment 1

- Assignment 1 out around 4PM today
- Get started soon!
- Not as finicky as Homework 1 or Homework 2

Admin: Assignments vs Homeworks

- Assignments are *individual*. Don't share code or ideas!
- I won't give hints, but I can help with some basic questions about the problem setup. I'll hold office hours as normal for any questions.
- Extensions only if necessary (serious illness, etc.)
- In terms of collaboration, treat it like a take-home test!
- Questions?

Hashing

Hash Table

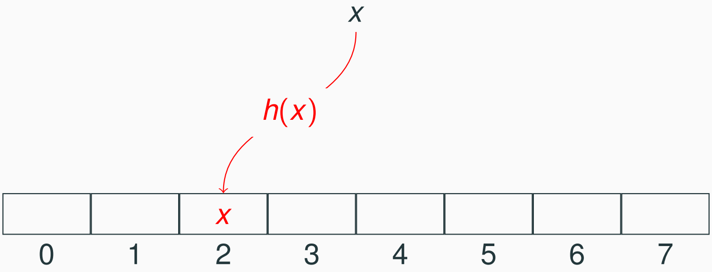
- Who has seen a hash table before? (Like the back end: a table/array with a hash function)
- Make a sad face if you have not seen it before or are very uncomfortable with it
- Plan: Assignment 1 has some very light hashing on it. Let's review hash tables. We'll need them for the rest of the class anyway
- We are *not analyzing* hash tables.
  - yet
- I just want you to know the basics

Hash tables

- Back end of many data structures (most notably python dictionaries)
- Store a set of n items
- Takes $O(n)$ space
- Can insert new items, delete old items, find if a given item is in the set, all in $O(1)$ time*
- Exceptionally practical

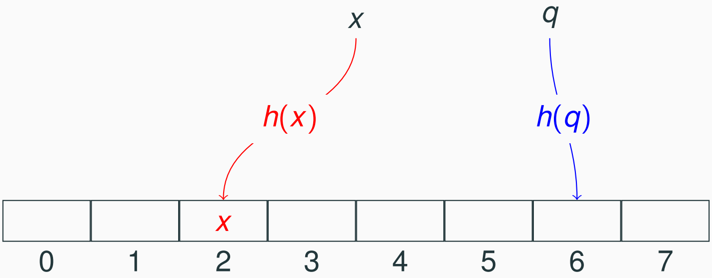
How a hash table works

- Start with (say) $2n$ slots
- The hash function maps each item to one of the $2n$ slots
- To insert an item: store it in the mapped slot



How a hash table works

- The hash function maps each item to one of the $2n$ slots
- To query: look in the mapped slot for the item



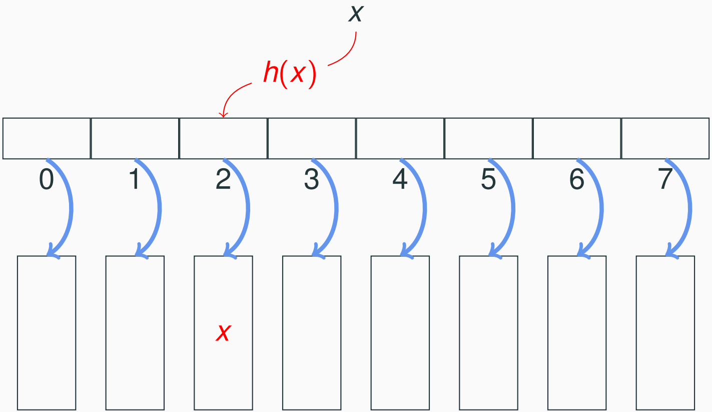
A hash table always gives the correct answer

- Any questions as to why?
- What is the lookup time so far if h can be evaluated in $O(1)$ time?
- Answer: $O(1)$. Just evaluate the hash and check if the item is stored there
- What is the potential issue we haven't talked about?

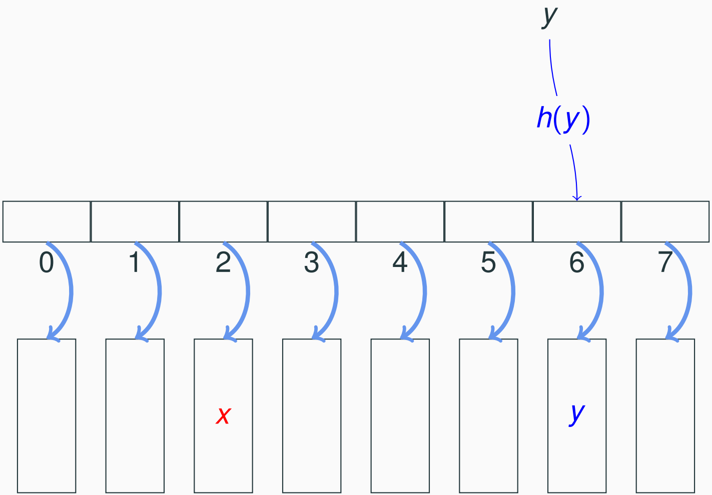
Collisions!

- Several items might hash to the same location. How can we resolve this?
  - Keep a data structure for all items that hash to the same location
  - Let's call each location a "bucket"

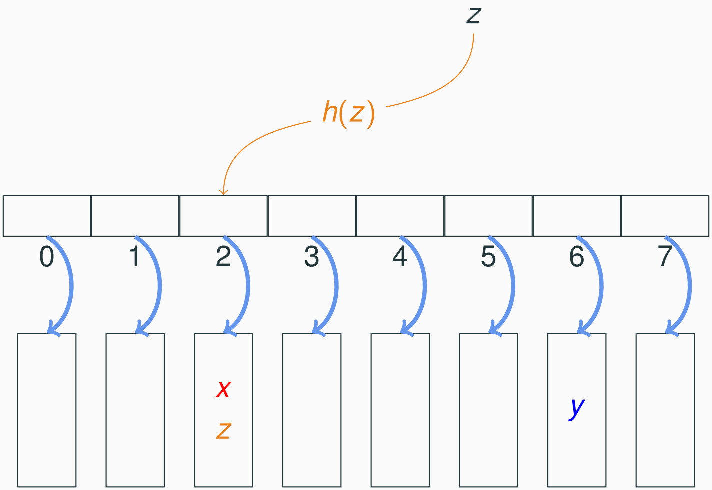
Handling Collisions with Buckets



Handling Collisions with Buckets

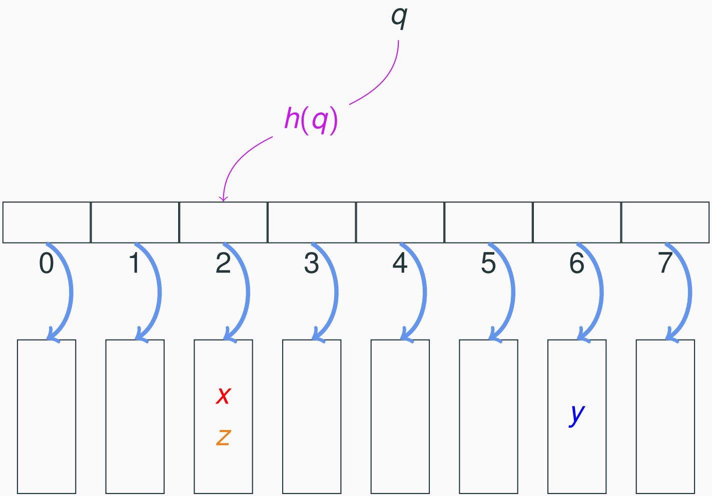


Handling Collisions with Buckets



Querying with Buckets

- Need to search through the whole bucket



Buckets

- Thought question: what are some good data structures for storing buckets? What are the tradeoffs of using each?
- Need to insert items in the bucket, look up an item in the bucket
- If buckets are *small*, and we use a reasonable data structure to store the items, get good performance

Assignment 1 Topic (On Other Slides)

Probability

Moving forward

- Mountain day messes with the schedule substantially
- I'd normally go over the Homework 3 topic next Friday
- Instead we'll do it Tuesday. So, I want to do a quick introduction to probability today
- Not a topic on Assignment 1!
- *Not a topic on Assignment 1!* It does use hashing, but you won't analyze that part of it

Quick Survey

- Who’s seen bloom filters before?
- Who’s seen cuckoo hashing before?
- Who’s done any probability math before?
- Who’s done any probabilistic algorithms analysis before?
- Who’s heard of streaming algorithms before?
  - (Not streaming like Netflix.)

Probability Takeaways

What I want you to know by the end of this section of the course:

1. Definition of probability/basic calculations
2. Determine if two events are independent
3. Calculate expectation
4. Linearity of expectation
5. Difference between “concentration bounds” vs expected performance

Definition of Probability

- Defined over a set of possible *outcomes* (often called the *sample space*)
- An *event* is a subset of the outcomes
- $$\Pr[\text{Event } E] = \frac{\#\text{ outcomes in the event}}{\text{Total }\#\text{ of outcomes}}$$
- Formal definition probability generally applies weights to the events (in which case the definition of probability is the weight of outcomes in the event, divided by total weight of all events). We will usually have equal-weight events.

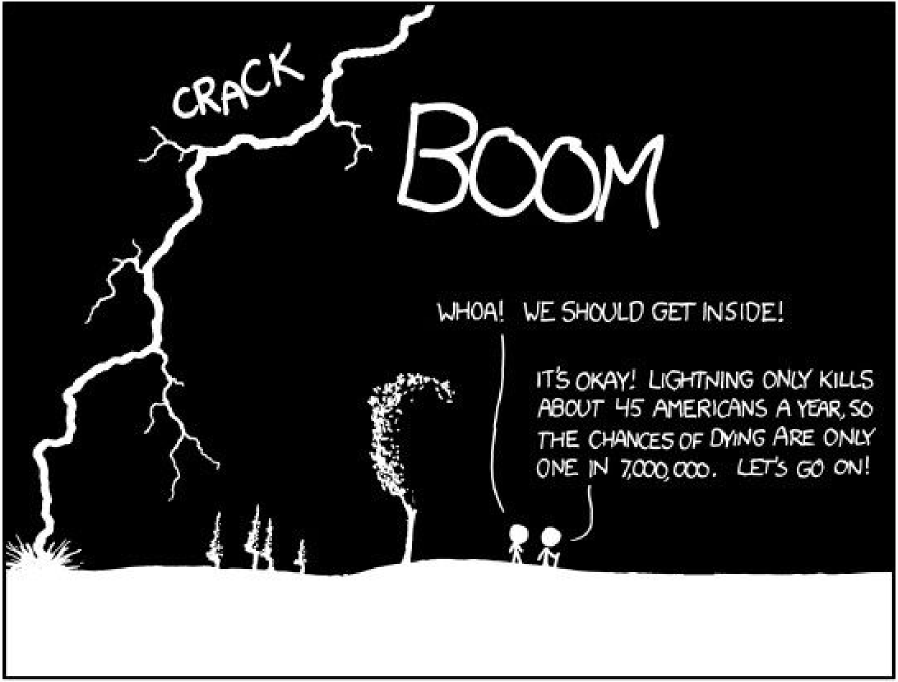
Probability Calculation Examples

- Let's say I roll a 20-sided die. What is the probability that an even number comes up?
- Answer: $10/20 = 1/2$.
- Let's say I flip a coin 10 times. What is the probability of getting exactly 5 heads? (Let's do this on the board)
- $\binom{10}{5}/2^{10}$ Which is $\frac{10!}{5!5!2^{10}} \approx .246$
- (In a couple lectures we'll see tools to estimate probabilities without lots of large numbers all over the place.)

30 Second Stretch

Conditional Probability

- Sometimes we want to calculate the probability of an event, when we already have some *partial information* about the outcome
- Specifically: want to calculate the probability of event E_1, already knowing that the outcome is in E_2. Denoted $\Pr[E_1 | E_2]$.
- Example: let's say I'm playing cards with a 52-card deck. I have already drawn three cards; all three were clubs. What is the probability that the fourth card is a club?
- $\Pr[\text{draw 4 clubs} \mid \text{first three cards were a club}]$
- How many outcomes are there for the fourth card? How many of them are a club?
- 49 outcomes. 10 of them are clubs. Probability: $10/49$.



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT *STATISTIC* IS ONE IN SIX.

Conditional Probability Second Example

- Let’s say that I have one child who is a boy (in a very simplified model where each child is a boy or girl according to a coin flip). What is the probability that when I have a second child, it winds up being a boy?
- What is your answer intuitively?
- Outcomes for my children: BB BG GB GG
- Outcomes consistent with "the second one is a boy": BB GB
- Probability $1/4$

Conditional Probability Third Example

Conditional probability can be a bit unintuitive at times! Break it down to be sure you're getting the right answer. (This is why we go over the formal definitions)

- Let's say that I have two children. One of them is a boy. What is the probability that both of them are boys?
- Outcomes for my children: BB BG GB GG
- Outcomes consistent with "one of them is a boy": BB BG GB
- Probability that both of them are boys: $1/3$
- Rephrasing the question: Let's say I have two children. They are not both girls. What is the probability that they are both boys?

Independence

- Idea: two events are independent if one does not have any impact on the other
- Example: let's say I flip a (fair) coin twice. Let E_1 be the event that the first flip is heads, and E_2 be the event that the second flip is heads. E_1 and E_2 are independent.
- Formal definition: E_1 and E_2 are independent if $\Pr[E_1 \mid E_2] = \Pr[E_1]$ and $\Pr[E_2 \mid E_1] = \Pr[E_2]$.
  - Knowing the outcome of one does not affect the probability of the other!
- I'll generally ask you if things are independent *intuitively* rather than asking for a proof. But, let's look at a couple classic examples of independence and how this definition works with them.

Proving Independence

Example: let's say I flip a (fair) coin twice. Let E_1 be the event that the first flip is heads, and E_2 be the event that the second flip is heads. E_1 and E_2 are independent.

Definition 1

E_1 and E_2 are independent if $\Pr[E_1 \mid E_2] = \Pr[E_1]$ and $\Pr[E_2 \mid E_1] = \Pr[E_2]$.

All possible outcomes: HH HT TH TT

Proving Independence

Example: let's say I flip a (fair) coin twice. Let E_1 be the event that the first flip is heads, and E_2 be the event that the second flip is heads. E_1 and E_2 are independent.

Definition 2

E_1 and E_2 are independent if $\Pr[E_1 \mid E_2] = \Pr[E_1]$ and $\Pr[E_2 \mid E_1] = \Pr[E_2]$.

All possible outcomes: HH HT TH TT

$$\Pr[E_2] = 1/2$$

Proving Independence

Example: let's say I flip a (fair) coin twice. Let E_1 be the event that the first flip is heads, and E_2 be the event that the second flip is heads. E_1 and E_2 are independent.

Definition 3

E_1 and E_2 are independent if $\Pr[E_1 \mid E_2] = \Pr[E_1]$ and $\Pr[E_2 \mid E_1] = \Pr[E_2]$.

All possible outcomes: $\boxed{HH}$ $\boxed{HT}$ TH TT

$$\Pr[E_2 \mid E_1]$$

Proving Independence

Example: let's say I flip a (fair) coin twice. Let E_1 be the event that the first flip is heads, and E_2 be the event that the second flip is heads. E_1 and E_2 are independent.

Definition 4

E_1 and E_2 are independent if $\Pr[E_1 \mid E_2] = \Pr[E_1]$ and $\Pr[E_2 \mid E_1] = \Pr[E_2]$.

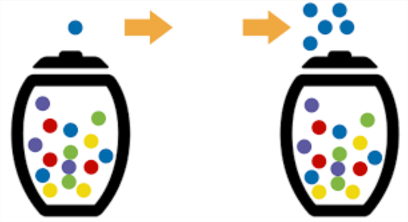
All possible outcomes: $\boxed{HH}$ $\boxed{HT}$ TH TT

$$\Pr[E_2 \mid E_1] = 1/2$$

Again: we'll normally be looking at this intuitively.

Independence: Examples

- Let's say I have a bag of balls, half of which are black, and half of which are white. I take a ball out of the bag and look at what color it is. Then I take another ball out of the bag and look at what color it is.
  - Event 1: The first ball is white.
  - Event 2: The second ball is black
  - Are these events independent?



Independence: Examples

- Let’s say I have a bag of balls, half of which are black, and half of which are white. I take a ball out of the bag and look at what color it is. Then I take another ball out of the bag and look at what color it is.
  - Event 1: The first ball is white.
  - Event 2: The second ball is black
  - Are these events independent?

- No! If the first ball is white, there will be more black balls then white balls remaining in the bag for the next draw.

Independence: Examples

- Let's say I shuffle a deck of cards and look at the top card. I replace the card and shuffle the deck again and look at the top card. Is the event that the first card is red, and the event that the second card is red, independent?
- Yes
- Let's say I roll a 20-sided die. Is the event that the resulting number is a multiple of 3 independent of the event that the result is even?
  - Yes. (Multiples of 3 are: 3 6 9 12 15 18; half of these are even.)
- What about a 21-sided die?
  - No! (Multiples of 3 are: 3 6 9 12 15 18 21; $3/7$ of these are even.)

Independence: Examples

- Let's say I use a *random* hash function *h* to store a set of elements in a hash table
- Are the following to events independent?
  - The first bucket of the hash table contains no elements
  - The second bucket of the hash table contains no elements
- No! Since the first bucket contains no elements, the remaining elements are *slightly* more likely to hash to the second bucket
- For intuition: let's say I knew that all but the last bucket of the hash table contained 0 elements. Then I would know exactly the contents of the last bucket

Why independence is useful

- If A and B are independent, then $\Pr[A \text{ and } B] = \Pr(A) \cdot \Pr(B)$.
- What is the probability of flipping 10 heads in a row? All 10 are independent, so $1/2^{10}$.

Why independence is useful

- Let's say you're in a class of n students. Every day the professor asks a student to explain the previous night's reading (the student is chosen by rolling an n-sided die). What is the probability that you won't be chosen after all k lectures in the course?
- Probability (not being chosen on one day) is $(1 - 1/n)$
- Probability (not being chosen after k days) is $(1 - 1/n)^k$
- Side note: we can put this in a more readable form.
  $(1 - 1/n)^k = ((1 - 1/n)^n)^{k/n} \approx 1/e^{k/n}$