CS124 Lecture 12 March 4, 2020

Birthday Paradox

$O(\sqrt{days})$

k people in room

p(two have same birthday) → instead, 1 − p(two don't have same birthday)

k	prob
1	0
2	$(1 - 1/365)$
3	$(1 - 1/365) \times (1 - 2/365)$
4	$''' \times (1 - 3/365)$

$$\prod_{i=1}^{k-1} (1 - i/365)$$

$k = 23$ (more likely than not pair has same bday)

Balls and Bins

m bins, n balls

What fraction of the bins are empty?

$\text{prob(bin \#1 is empty)} = (1 - 1/m)^n \approx e^{-n/m}$

Linearity of Expectations

$E[\#\text{empty bins}] = m(1 - 1/n)^m \approx me^{-n/m}$

What is the expected # of bins with 1 ball?

$\Pr(\text{bin \#1 gets 1 ball}) : \binom{n}{1} \frac{1}{m} (1 - 1/m)^{n-1}$

Hash Functions

$H: \mathcal{U} \to \{0, \dots, m-1\}$

Good hash functions look random

Usage: distribute items so you can look them

Password checking

- prevent bad passwords
- small space complexity (w/ some mistakes)

rather tell people to not use a safe pass than to use unsafe pass

m bits

0	~~0~~ 1	0	0	~~0~~ 1	0	0	~~0~~ 1

H:

$H(x) \to 0$ to $m-1$

set that bit to 1

if $0 \to$ okay!

$1 \to$ not okay, might have been a bad password

n bad passwords, m bits in Hash

$p(\text{say bad even though password is good}) = \dfrac{\#1s}{m} = 1 - e^{-n/m}$

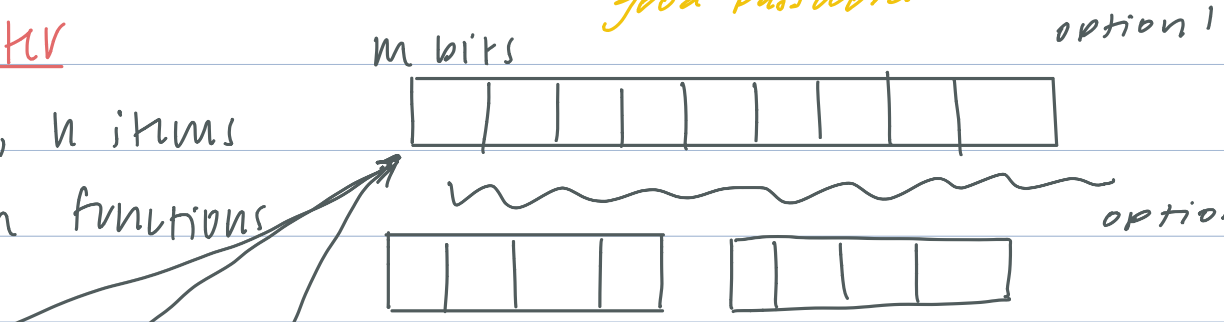
$p(\text{good}) = \dfrac{m e^{-n/m}}{m} = e^{-n/m}$

Bloom Filter

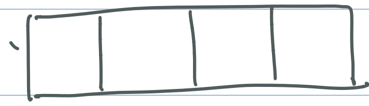
good password: at least 1 0

m bits, n items

k hash functions



$X \to h_1(X), h_2(X), \ldots h_k(X)$



$p(\text{reject good password}) = (1 - e^{-n/(m/k)}) = (1 - e^{-nk/m})^k$

$k = \frac{m}{n} \log 2 \to$ optimal # of hash functions

Using 1 byte/item, false prob down to 2%!