



Hyper Log Log



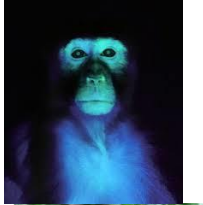
How many different types of monkeys are there?



Monkey Stack (sorted list, linked array, etc.)

- 1) Grabs monkey
- 2) Checks stack for identical monkeys
- 3) If no identical monkeys adds monkey to stack

Takes too long



Monkey Shelf (HashMap)

- 1) monkeys placed on the shelf are instantly sorted.
- 2) Slots for each monkey need to be configured beforehand

Takes too much storage



coinflips

Probability of all heads in

$$1 \text{ tosses} - \frac{1}{2^1}$$

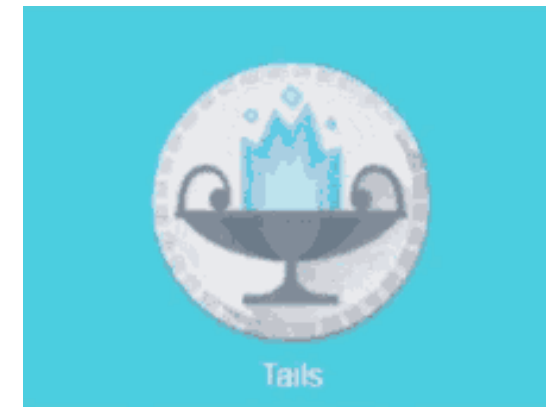
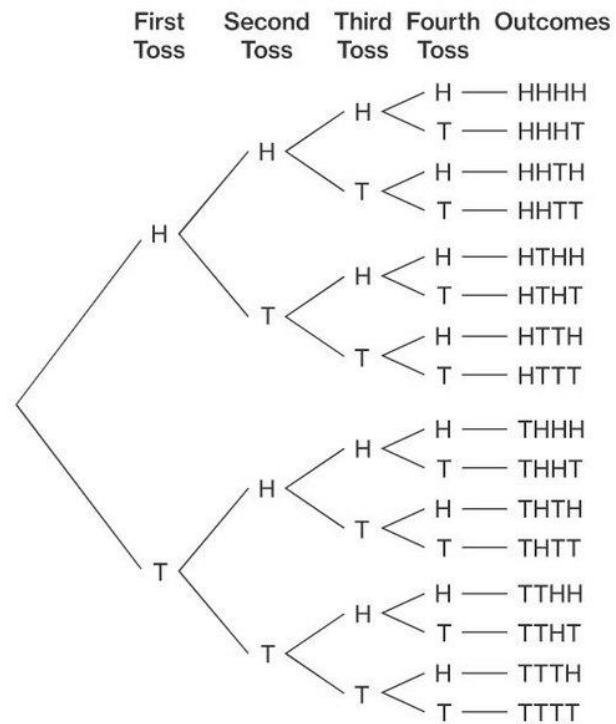
$$2 \text{ tosses} - \frac{1}{2^2}$$

$$3 \text{ tosses} - \frac{1}{2^3}$$

$$4 \text{ tosses} - \frac{1}{2^4}$$

$$5 \text{ tosses} - \frac{1}{2^5}$$

Etc...



Magic Monkey Meatgrinder(Hash Function)

- 1) Generates random binary sequence for each monkey
- 2) Will generate same set for same monkey



10100110101011011001



00100111101011011010



00100111101011011010

Magic Monkey Meatgrinder(Hash Function)

- 1) Run monkey through meat grinder
- 2) Count amount of zeros in a row
- 3) Repeat for every monkey and Keep track of highest amount of zeros on scorecard



10100110101011010000

4 zeros

4



01101010100001001100

2 zeros

4



01101010100010000000

7 zeros

7

Using scorecard to estimate monkeys scanned

$\frac{1}{2^7}$ chance of happening
01101010100010000000 7

We can approximate that we scanned 128 different types of monkeys

$\frac{1}{2^{10}}$ chance of happening
11001110110000000000 10

We can approximate that we scanned 1024 different types of monkeys

What if we just get unlucky?

11001101010000000000000000000000 15

index

15

1100110101110101011101001

11101101110000000000000000

1101110101110101011101000

index

2

0110110101111000100000110

0101010101110101001101000

index

0001100100010101011101000

3

index

3

1001110101110101011101010

1010110101111000101111000

100111010101010101011101100

Take harmonic mean of all scorecards

$$\frac{4}{\frac{1}{15} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}} = 3.252$$

Real world application

