Birthday Paradox

or

Birthday Problem

Birthday paradox

- In a group of 23 randomly chosen people, at least two will share a birthday with probability at least 50%. If there are 30, the probability is around 70%.

- Finding two people with the same birthday is the same thing as finding a collision for this particular hash function.

- The probability that all 23 people have different birthdays is

$$1 \times (1-\frac{1}{365})(1-\frac{2}{365})\ldots(1-\frac{22}{365}) = 0.493$$

Therefore, the probability of at least two having the same birthday is p(23)=1- 0.493=0.507

(see next page p(r) for different number of people r)

- More generally, suppose we have N objects, where N is large. There are r people, and each chooses an object. Then

$$P(\text{there is a match}) \approx 1 - e^{-r^2/2N}$$

r	P(r)
10	11.7%
20	41.1%
23	50.7%
30	70.6%
50	97.0%
57	99.0%
100	99.99997%
200	99.999999999999999999999999999998%
300	$(100 - (6 \times 10^{-80}))\%$
350	$(100 - (3 \times 10^{-129}))\%$
365	$(100 - (1.45 \times 10^{-155}))\%$
366	100%

Derivation of $P(r)$

and its approximation

1

$p(r) = Pr($at least one birthday match among r people$) = 1 - \bar{p}(r)$

$\bar{p}(r) = Pr($all different birthdays among r people$)$

$$\bar{p}(r) = 1 \times \frac{364}{365} \times \cdots \times \frac{365-r+1}{365}$$

$$= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{r-1}{365}\right)$$

② Approximations

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

for $|x| \ll 1$

$$e^x \approx 1 + x$$

$$\bar{p}(r) \simeq 1 \times e^{-\frac{1}{365}} \times \cdots \times e^{-\frac{r-1}{365}} = e^{\frac{-r(r-1)}{2} \cdot \frac{1}{365}}$$

$$\simeq e^{-\frac{r^2}{(2)(365)}}$$

$\therefore\ P(r) \simeq 1 - e^{-\frac{r^2}{(2\times 365)}}$

③ For r people, N birthdays,

how many r will have $P_r(r) \simeq \frac{1}{2}$?

<Sol> $P(r) = \frac{1}{2} \Rightarrow \bar{P}(r) = \frac{1}{2} \Rightarrow e^{-\frac{r^2}{2N}} = \frac{1}{2}$

$\Rightarrow \frac{r^2}{2N} = \ln 2 \Rightarrow r^2 = N \ln 4$

$\Rightarrow r \simeq 1.177\sqrt{N} = O(\sqrt{N})$

4 Application in Hash

$h(x) = y$

$h : \{0,1\}^* \longrightarrow \{0,1\}^{160}$

How many hashes, we will have a collision?

i.e. we randomly choose

$$x_1, x_2, \cdots, x_r \in \{0,1\}^*$$

$\exists$ some $i \neq j$, s.t. $h(x_i) = h(x_j)$

<sol>

$$r = O(\sqrt{N}) = O(\sqrt{2^{160}}) = O(2^{80})$$