

Probability of Collision

Morgan Benavidez

July 2022

1 Probability of Collision in Hashing

Suppose we use a hash function h to hash n distinct keys into an array T of length m . Assuming simple uniform hashing, what is the expected number of collisions?

h = hash function

n = distinct keys

T = Array

$m = \text{len}(\text{Array})$

$$A = h(k) = r \quad (1)$$

$$B = h(1) = r \quad (2)$$

$$P(A) = \frac{1}{m} \quad (3)$$

$$P(B) = \frac{1}{m} \quad (4)$$

$$P(A) * P(B) = \frac{1}{m^2} \quad (5)$$

$$\sum_{r=1}^m \frac{1}{m^2} = \frac{1}{m} \quad (6)$$

The formula $n(n-1)/2$ for the number of pairs you can form from an n element set has many derivations.

One is to imagine a room with ' n ' people, each of whom shakes hands with everyone else. If you focus on just one person you see that she participates in $n-1$ handshakes. Since there are n people, that would lead to $n(n-1)$ handshakes.

$${}_nC_3 = \frac{n(n-1)}{2} \quad (7)$$

$${}_nC_3 \frac{1}{m} = \frac{n(n-1)}{2} * \frac{1}{m} \quad (8)$$

$$= \frac{n(n-1)}{2m} \quad (9)$$