18 Feb 2026

Applications of 2-Universal Hash Functions.

Hash family/function: probability distrib on

$\{$functions $\chi \to B\}$

(χ: keys, B: buckets)

$|\chi| = N = n^{O(1)}$

$|B| = n$

Def. A hash family is k-universal if

$\forall$ k distinct keys $x_1, x_2, \ldots, x_k$

the k-tuple of buckets $h(x_1), h(x_2), \ldots, h(x_k)$

is uniformly distributed over B^k.

"k-wise independent": $h(x_1), \ldots, h(x_k)$ are mutually independent of each other.

$k=2$ — universal hash families have many applications in algs & data structures.

If $|B| = n = p$ (prime number) and $|\chi| \le p$,

choose any one-to-one mapping from χ to $\mathbb{F}_p$ (integers mod p)

$\| \{0, 1, 2, \ldots, p-1\}$

then sampling $a, b \in \mathbb{F}_p$ unif. at random

and defining $h_{a,b}(x) = ax + b \pmod p$

yields a 2-univ hash family.

If $|B| = p$ prime, $|X| \le p^d$ $(d \in \mathbb{N})$

identify X with a subset of

$\mathbb{F}_p^d = \{0, \dots, p-1\}^d$

vectors of dimension d

with all operations (addition, scalar mult, dot prod) interpreted modulo p.

If we sample $a \in \mathbb{F}_p^d$, $b \in \mathbb{F}_p$

indep, unif random, then

$$h_{a,b}(x) = \langle a, x \rangle + b \pmod p$$

$\langle a, x \rangle$ ← inner product $\sum_{i=1}^{d} a_i x_i$

Why 2-universal?

For $x \neq y$, $\in \mathbb{F}^d$ $\Rightarrow$ $\left(h_{a,b}(x), h_{a,b}(y)\right) \in \mathbb{F}_p^2$.

Suppose WLOG $x_1 \neq y_1$.

Game plan: for any fixed $(a_2, \dots, a_d) \in \mathbb{F}_p^{d-1}$

as a_1 & b range over $\mathbb{F}_p^2$

$\left(h_{a,b}(x), h_{a,b}(y)\right)$ -- -- $\mathbb{F}_p^2$

If

$$a_1x_1 + a_2x_2 + \cdots + a_dx_d + b = a_1'x_1 + a_2x_2 + \cdots + a_dx_d + b'$$

$$a_1y_1 + a_2y_2 + \cdots + a_dy_d + b = a_1'y_1 + a_2y_2 + \cdots + b'$$

$$a_1x_1 + b = a_1'x_1 + b'$$

$$a_1y_1 + b = a_1'y_1 + b'$$

$$a_1(x_1 - y_1) = a_1'(x_1 - y_1)$$

$$(a_1 - a_1')\cdot(x_1 - y_1) \equiv 0 \mod p.$$

$\therefore$ p divides $a_1 - a_1'$

$\Rightarrow$ $a_1 = a_1' \mod p.$

$a_1x_1 + b = a_1'x_1 + b' \Rightarrow b = b' \mod p$

$(a_1, b) \longmapsto \left(h_{a,b}(x), h_{a,b}(y)\right)$ is injective from $\mathbb{F}_p^2 \to \mathbb{F}_p^2$.

$\therefore$ bijective

$\therefore$ $(h_{a,b}(x), h_{a,b}(y)) \in \mathbb{F}_p^2$ uniform rand.

Why not 3-universal?

E.g. $p > 2$

$x = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ $y = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$$h_{a,b}(x) + h_{a,b}(y) = (2a_1 + b) + (2a_2 + b)$$

$$h_{a,b}(z) = a_1 + a_2 + b$$

$$h_{a,b}(z) - h_{a,b}(x) = a_2 - a_1$$

$$h_{a,b}(y) - h_{a,b}(z) = a_2 - a_1$$

$$h_{a,b}(x) + h_{a,b}(y) = 2\, h_{a,b}(z)$$

$\Rightarrow$ $h_{a,b}(z)$ is uniquely determined by $h_{a,b}(x)$ & $h_{a,b}(y)$, not independent of them.

What about $x, y \in \{0, 1, \dots, p-1\}$ $p > 2$

$$h_{a,b}(x) = ax + b \pmod 2$$

Is this a 2-univ hash family from $\mathbb{F}_p^2$ to $\{0,1\}$?

It is not!

$$h_{a,b}(0) = b \pmod 2$$

$$h_{a,b}(2) = 2a + b \pmod 2$$
$$\equiv b \pmod 2$$

On any set of n elements we can construct a 2-univ $\{0,1\}$-valued hash family using the inner product hash family

on $\mathbb{F}_2^d$ where $d \geq \lceil \log_2(n) \rceil$.

Dictionary data struct. stores (key, value) pairs and offers INSERT(x,v), LOOKUP(x), DELETE(x) all in $O(1)$ time in a model where operations on words of size $O(\log n)$ bits take $O(1)$ time.

Chain hashing. Suppose we know at initialization that we will never store more than $m = O(n)$ key-value pairs in dictionary. We sample hash function h with n buckets from a 2-univ family. Each bucket stores linked list of (x,v) pairs (initially empty). "chain hashing"

INSERT(x,v): let $b \leftarrow h(x)$
if (x,v') already stored at b, overwrite v' with v.
else append (x,v) to list stored at b.

LOOKUP(x): let $b \leftarrow h(x)$
search list stored at b to find value assoc to x.

DELETE(x): let $b \leftarrow h(x)$
search list stored at b, delete the pair assoc to x if found.

Time

Complexity of each operation:

- $O(1)$ to evaluate $h(x)$.
- $O(s_b)$ where s_b is # of (key, value) pairs stored in bucket b.

s_b is a random variable.

$$\forall b \in B \quad \mathbb{E}[s_b] \leq \frac{m}{n} = O(1).$$

↑ linearity of expectation.
Each of $\leq m$ keys has exactly $\frac{1}{n}$ prob. of being in b.

Complexity of LOOKUP(x) depends on

$$\mathbb{E}[s_{h(x)}].$$

$$= \sum_{b \in B} \Pr(h(x) = b) \cdot \mathbb{E}[s_b \mid h(x) = b]$$

For all b,

$$\mathbb{E}[s_b \mid h(x) = b] = 1 + \sum_{y \neq x} \Pr(h(y) = b \mid h(x) = b)$$

↑ keys other than x inserted into dictionary.

$$\leq 1 + \frac{n-1}{n} = O(1).$$