Hashing

Universality

A hash function $h : U \to [m]$ maps values from a key universe U into values in $m = [0...m-1]$.
Universal hashing is the concept of generating a random *universal* hash function h, so when we pick two distinct keys $x, y \in U$, the probability of collision is:

$$Pr_h[h(x) = h(y)] \leq 1/m \quad or \quad Pr_h[h(x) = h(y)] \leq c/m$$

Application

Universal - Hash tables with chaining. When we want lookups in expected constant time, $1 + |L(h(x)|$. If we have used a universal hash function, we can expect the buckets to be of size $|S|/m$. (We have the indicator varialbe $I(y)$. It is the number of collisions with a new key $x \notin S$).

$$E[L(h(x))] = E[\sum_{y\in S} I(y)] = \sum_{y\in S} E[I(y)] = \sum_{y\in S} E[h(x) = h(y)] = |S| \cdot \frac{1}{m}$$

Strong universality

A stronger condition known as pairwise independence or strong universality is when given two distinct keys $x, y \in U$ hash to values r and q respectively with probability $1/m^2$. If it is strongly universal, it implies it is also universal, as

$$Pr[h(x) = h(y)] = \sum_{q\in[m]} Pr[h(x) = q \wedge h(y) = q] = m/m^2 = 1/m$$

Proof that two keys are hashed individually and each key is hashed uniformly into $[m]$. Uniformly as each pair has exactly $1/m^2$, and there are m values of r for each q.
Independence (calculate $P[A|B] = \frac{P[A]*P[B]}{P[B]}$).

Application

Strongly universal - coordinated sampling, important in handling of big data and machine learning. We can define a set $S_{h,t}(A)$ from a set A, a strongly universal hash function h and a threshold t. The size is $|A| \cdot t/m$ as a strongly universal hash function means that values are uniformly mapped to $[m]$.
We can say something about unions and intersections between two sets by multiplying with m/t.

Chebyshev's inequality?

$$Pr[|X - \mu| \geq q\sigma_x] \leq 1/q^2$$

for $q > 0$. Says something about that in any probability distribution, "nearly all" values are close to the mean.

Implementations

Multiply-mod-prime:
Universal (with $c = 1$) where $h_{a,b} : [u] \to [m]$:

$$h_{a,b}(x) = ((ax + b) \bmod p) \bmod m$$

Strongly universal where $h_{a,b} : [p] \to [p]$:

$$h_{a,b}(x) = (ax + b) \bmod p$$

Multiply-shift:
Universal (with $c = 2$) where $h_a : [2^w] \to [2^d]$:

$$h_a(x) = \lfloor (ax \bmod 2^w)/2^{w-l} \rfloor$$

Strongly universal where $h_{a,b} : [2^w] \to [2^l]$:

$$h_{a,b}(x) = (ax + b)[w' - l, w']$$

and $w' \geq w + l - 1$