En el document adjunt de Stanford on se'ns explica el càlcul de similitud de documents mitjançant signatures minhash, es defineixen com han de ser les funcions hash:

- Triar n funcions hash aleatòries.
- No hi hagi un gran nombre de col·lisions, és a dir, dues funcions amb les mateixes columnes (realitzen la mateixa funció de hash).

Basant-nos amb aquests paràmetres, el document de Stanford adjunt, i recerca per part nostre sobre com escollir unes funcions hash correctes, hem arribat a la conclusió de que les funcions escollides en els exemples donats són funcions conegudes com Universal Hashing Functions.

Una Universal Hashing Function, fa referència a un conjunt de funcions hash aleatòries les quals compleixen la següent propietat matemàtica: **(ax + b) mod m.** Aquesta propietat matemàtica té les següents característiques:

- a és un enter senar més petit que m.
- x és el paràmetre a aplicar la funció hash.
- b és un nombre aleatori més petit que m.
- m en el nostre cas és el nombre total de shingles.

Aquest fet assegura un nivell de col·lisió relativament baix, encara que els documents (paràmetres d'entrada a comparar) siguin escollits per un adversari.

En l'exemple donat pel document de Stanford es realitza el càlcul minhash amb dues funcions de hash. Nosaltres per augmentar la precisió i el nombre de tests, realitzarem el càlcul min hash amb n funcions hash. Aprofitant les dues primeres donades tindrem:

1. (a1*x + 1) mod m.
2. (a2*x + 1) mod m.
3. (a3*x + 1) mod m.
4. (a4*x + 1) mod m.
   ...
5. (an*x + 1) mod m.

Tal com s'ha dit abans, partint de les funcions de l'exemple i el càlcul de funcions hash de tipus Universal Hashing, calcularem n funcions aleatòries modificant el paràmetre que multiplica la x. Aquest paràmetre serà un nombre imparell aleatori i diferent entre tots ells, el qual serà calculat pel nostre algorisme.

D'aquesta manera obtenim una manera eficient de calcular funcions de hash amb un nombre relativament petit de col·lisions.

https://en.wikipedia.org/wiki/Universal_hashing