Paradoxe des anniversaires

Si on réunit 23 personnes, alors il y a une probabilité d'environ 50% d'avoir 2 personnes qui sont nées le même jour.

Avec 57 personnes, la probabilité grimpe à 99%.

Paradoxe : vérité mathématique qui contredit l'intuition

Calcul : On suppose que toutes les dates d'anniversaires sont équiprobable
i.e. $Pr(\text{être né à la date } d) = \frac{1}{365}$

Soient k personnes. Les dates d'annivers des k personnes sont vues comme un tuple $(d_1, \dots, d_k) \in \{356\}^k$.

→

nombre de configurations possibles :

$$N = 365^k \qquad (1)$$

nombre de configurations où tous le monde est né un jour différents

$$365 \cdot 364 \cdots (365-k+1) \qquad (2)$$

$$\Rightarrow Pr(\text{tout le monde soit né un jour différent}) = \frac{(2)}{(1)} = \frac{365 \cdot 364 \cdots (365-k+1)}{365^k} = \frac{365!}{365^k (365-k)!}$$

$$\Rightarrow Pr(\text{au moins 2 personnes soient nées le même jour}) = 1 - \frac{(2)}{(1)} = 1 - \frac{365!}{365^k (365-k)!}$$

$k = 23 \Rightarrow p = 50.73\%$

$k = 60 \Rightarrow p = 99.41\%$

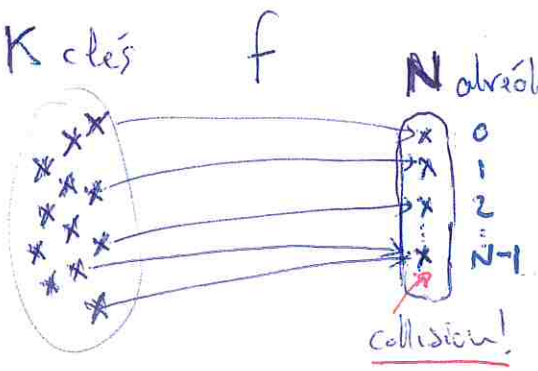
Probabilité de collision

Soit la fonction de hachage

$$f : C \longrightarrow \{0, \dots, N-1\}, \quad \text{avec } |C| = K$$

(clés : C ; alvéoles : $\{0, \dots, N-1\}$)

Si $K \geq N$, alors il y a forcément collision



car la fonction f ne peut pas être injective.

Pour K et N grands, on peut estimer

$$\Pr(\text{pas de collision}) = \overline{P}(K,N) = \frac{N \cdot (N-1) \cdots (N-K+1)}{N^K}$$

$$= \frac{N}{N} \cdot \frac{N-1}{N} \cdots \frac{(N-K+1)}{N}$$

$$= \prod_{i=1}^{K-1} \frac{N-i}{N} \approx e^{-\frac{K(K-1)}{2N}}$$

$$\Pr(\text{au moins 1 collision}) = 1 - \overline{P}(K,N)$$

$$= 1 - \prod_{i=1}^{K-1} \frac{N-K}{N} \approx 1 - e^{-\frac{K(K-1)}{2N}}$$

→

Exemple : Stocker $K = 1000$ clés avec $< 1\%$ chance de collision

$$P(\text{collision}) < 0.01 \quad \text{ssi}$$

$$1 - e^{-\frac{K(K-1)}{2N}} < 0.01 \quad \text{ssi}$$

$$-e^{-K(K-1)/2N} < -0.99 \quad \text{ssi}$$

$$e^{-K(K-1)/2N} > 0.99 \quad \text{ssi}$$

$$-\frac{K(K-1)}{2N} > \ln(0.99) \quad \text{ssi}$$

...

$$2N \underbrace{\ln(0.99)}_{<0} < -K(K-1) \quad \text{ssi}$$

$$N > -\frac{K(K-1)}{2\ln(0.99)} \approx 4.97 \cdot 10^7$$

↑ division par <0 change le sens de l'inég.

Ce qui représente une importante mémoire

$\Rightarrow$ On doit gérer les collisions