Birthday paradox

EG Among $n=35$ people, how likely is it that two have the same birthday?

$$1 - \underbrace{\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\left(1-\frac{3}{365}\right)\cdots\left(1-\frac{34}{365}\right)}_{\text{probability that all birthdays different}} = 81.4\%$$

$n=50$: 97.0% $n=70$: 99.9 %

$\Rightarrow$ collisions more frequent than one might expect

365 $\rightsquigarrow M$

probability of no collision among n selections from M choices :

$$\left(1-\frac{1}{M}\right)\left(1-\frac{2}{M}\right)\cdots\left(1-\frac{n-1}{M}\right)$$

$\approx e^{-1/M}$ $\approx e^{-2/M}$ $\approx e^{-(n-1)/M}$

$e^x e^y = e^{x+y}$

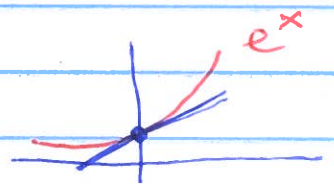
assume $n \ll M$

$$\approx e^{-(1+2+3+\cdots+(n-1))/M}$$

$$= e^{-\frac{(n-1)n}{2M}} \approx e^{-\frac{n^2}{2M}}$$

if $n \approx \sqrt{M}$ $\approx e^{-1/2} = 60.7\%$

(collision likely: 39.3%)



$e^x \approx 1+x$ for small x

$1 + 2 + \cdots + (n-1) = \frac{(n-1)n}{2}$

for hashes: output size b bits $\rightarrow M = 2^b$

among $\sqrt{M} = 2^{b/2}$ hashes a collision is likely

$\Rightarrow$ b needs to be twice as large as needed to prevent brute-forcing

AES-128
key size 128 bit
OK

SHA-2/3 : output size 256+ bits