Birthday Problem

Yukang Shen

University of California, Santa Barbara

May 7, 2014

Start with a question

In a room of n people, How many people do we need to make sure that at least two of them have the same birthday?

Start with a question

In a room of n people, How many people do we need to make sure that at least two of them have the same birthday?

Pigeonhole principle: 366 people!

Another question

How about the number of people that there is very high probability (say over 50%) at least two of them have the same birthday?

Another question

How about the number of people that there is very high probability (say over 50%) at least two of them have the same birthday?

183? No!

Another question

How about the number of people that there is very high probability (say over 50%) at least two of them have the same birthday?

183? No!

We only need 23 people!

Prove

If $P(A)$ is the probability of at least two people in the room having the same birthday, it may be simpler to calculate $P(A')$, the probability of there not being any two people having the same birthday. Then, because A and A' are the only two possibilities and are also mutually exclusive:

$$P(A) = 1 - P(A')$$

Prove

If $P(A)$ is the probability of at least two people in the room having the same birthday, it may be simpler to calculate $P(A')$, the probability of there not being any two people having the same birthday. Then, because A and A' are the only two possibilities and are also mutually exclusive:

$$P(A) = 1 - P(A')$$

$$P(\textit{At least } 1 \textit{ same birthday}) = 1 - P(\textit{No same birthday})$$

If there are two people, the chance that they do *not* have the same birthday is

$$\frac{364}{365}$$

If there are two people, the chance that they do *not* have the same birthday is

$$\frac{364}{365}$$

So the change that they *do* have the same birthday is:

$$1 - \frac{364}{365} \approx 0.28\%$$

If there are three people, you and 2 others, the chance that *neither of the other two* shares *your specific birthday* is

$$\frac{364}{365} \times \frac{364}{365}$$

If there are three people, you and 2 others, the chance that *neither of the other two* shares *your specific birthday* is

$$\frac{364}{365} \times \frac{364}{365}$$

and so the chance that no one else shares *your* birthday is:

$$1 - \frac{364}{365} \times \frac{364}{365} \approx 0.55\%$$

However, the other two might have the same birthday, not equal to yours. The chance that all 3 people have different birthdays is

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365}$$

However, the other two might have the same birthday, not equal to yours. The chance that all 3 people have different birthdays is

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365}$$

hence, the probability that *not all three* birthdays are distinct(at least two share the same birthday is)

$$1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \approx 0.82\%$$

Continuing this way, we see that in the group of 23 people, the chance that at least two share the same birthday is:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-22}{365}) \approx 50.7\%!$$

How about for n people?

How about for n people?

If $n > 365$, by pigeonhole principle, $P(A) = 1$

How about for n people?

If $n > 365$, by pigeonhole principle, $P(A) = 1$

If $n \leq 365$, we can get the formula:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-n+1}{365})$$

How about for n people?

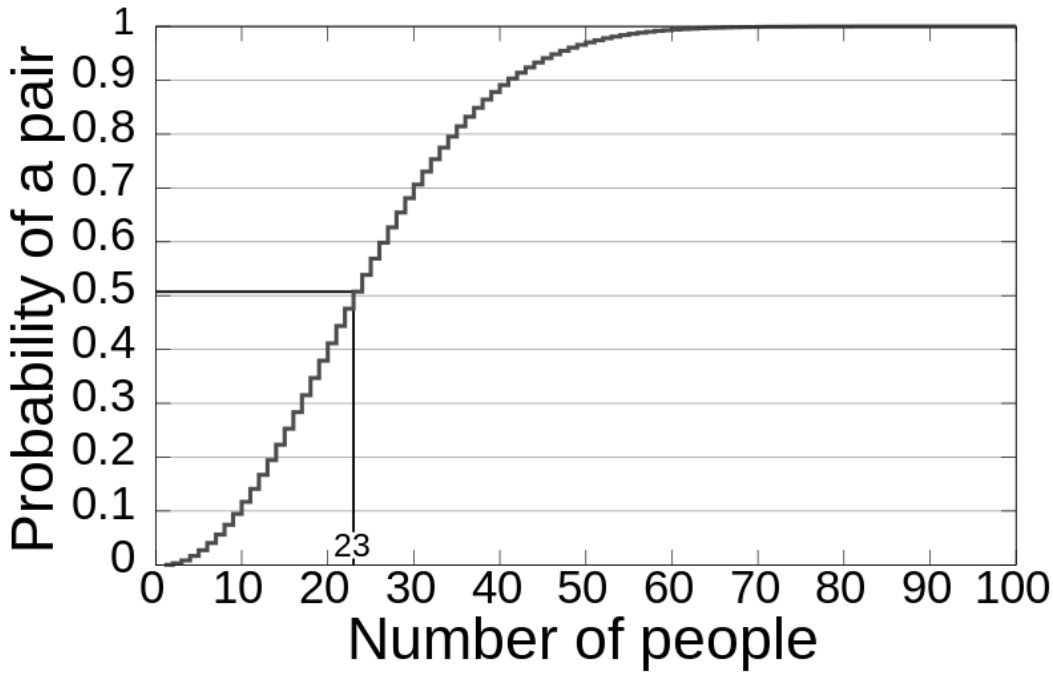
If $n > 365$, by pigeonhole principle, $P(A) = 1$

If $n \leq 365$, we can get the formula:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-n+1}{365})$$

$$1 - \frac{365!}{365^n(365-n)!}$$

n	P(n)
5	2.7%
10	11.7%
20	41.1%
23	50.7%
40	89.1%
50	97.0%
70	99.7%
100	99.99997%
366	1



Calculation!

Let’s try 23 in our calculator using that formula!

Calculation!

Let's try 23 in our calculator using that formula!

OOPS!

Calculation!

Let's try 23 in our calculator using that formula!

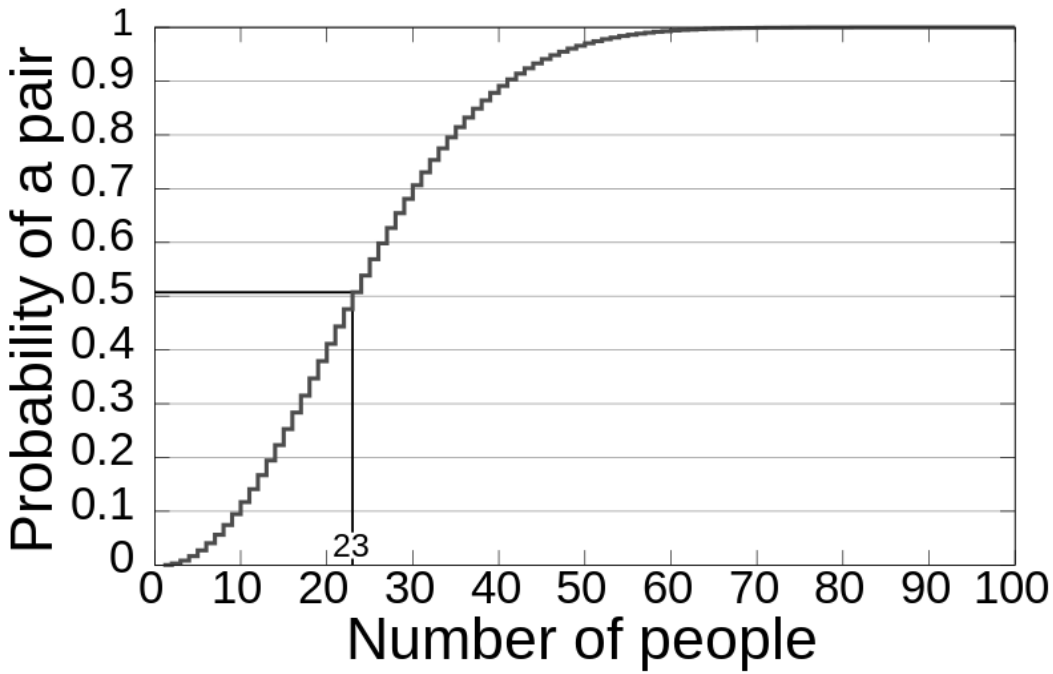
OOPS!

365! is soooo big!

Can we do better?

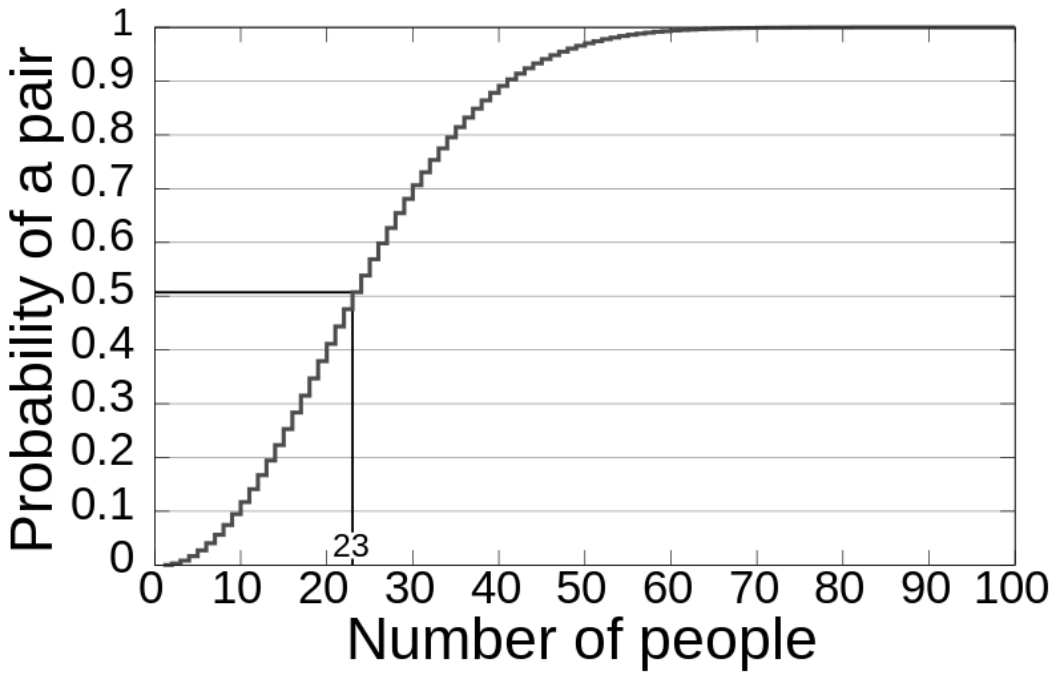
Can we do better?

First , let's review the graph:



Can we do better?

First , let's review the graph:



Like a graph of exponential function?

Let's start from this formula:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-22}{365})$$

Let's start from this formula:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-22}{365})$$

$$= 1 - 1 \times (1 - \frac{1}{365}) \times (1 - \frac{2}{365}) \times ... \times (1 - \frac{22}{365})$$

Let's start from this formula:

$$1 - (\frac{365}{365} \times \frac{365-1}{365} \times ... \times \frac{365-22}{365})$$

$$= 1 - 1 \times (1 - \frac{1}{365}) \times (1 - \frac{2}{365}) \times ... \times (1 - \frac{22}{365})$$

Looking at each term $(1 - \frac{a}{365})$, we can use first-order Taylor expansion for it!

When x is close to 0, a first-order Taylor approximation for e^x is:

$$e^x \approx 1 + \frac{x}{1!} = 1 + x$$

When x is close to 0, a first-order Taylor approximation for e^x is:

$$e^x \approx 1 + \frac{x}{1!} = 1 + x$$

so

$$1 - \frac{1}{365} \approx e^{-1/365}$$

When x is close to 0, a first-order Taylor approximation for e^x is:

$$e^x \approx 1 + \frac{x}{1!} = 1 + x$$

so

$$1 - \frac{1}{365} \approx e^{-1/365}$$

Then we can rewrite the formula to

$$P(A) \approx 1 - 1 \times e^{-1/365} \times e^{-2/365} \times ... \times e^{-22/365}$$

$$\approx 1 - e^{(-1-2-3-...-22)/365}$$

$$\approx 1 - e^{-(1+2+3+...+22)/365}$$

Now remember that adding the numbers 1 to $n = n(n+1)/2$, Adding 1 to 22 is $(22 \times 23)/2$ so we get:

$$P(A) \approx 1 - e^{-((23 \cdot 22)/2 \cdot 365)}$$

$$= 50.0\% \approx 50.7\%$$

How about for n people now?

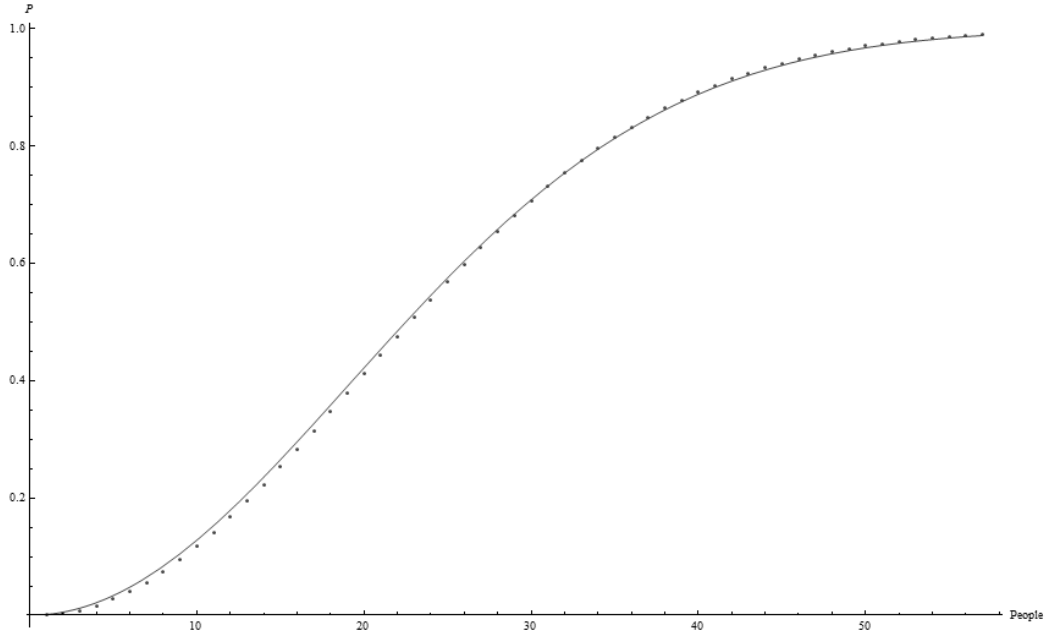
$$P(A) \approx 1 - e^{-(n\cdot(n-1)/2\cdot 365)}$$

$$\approx 1 - e^{-(n^2/2\cdot 365)}$$

How about for n people now?

$$P(A) \approx 1 - e^{-(n\cdot(n-1)/2\cdot 365)}$$

$$\approx 1 - e^{-(n^2/2\cdot 365)}$$



Let's generalize the formula to picking n people from T total items (instead of 365):

$$P(n, T) \approx 1 - e^{-(n^2/2 \cdot T)}$$

If we choose a probability m (like 50% chance of a match) and solve for n:

$$1 - m \approx e^{-(n^2/2 \cdot T)}$$

$$-2ln(1 - m) \cdot T \approx n^2$$

$$n \approx \sqrt{-2ln(1 - m)} \cdot \sqrt{T}$$

Birthday attack

Definition A birthday attack is a type of cryptographic attack that exploits the mathematics behind the birthday problem in probability theory. This attack can be used to abuse communication between two or more parties. The attack depends on the higher likelihood of collisions found between random attack attempts and a fixed degree of permutations (pigeonholes), as described in the birthday problem/paradox.

First, let's rewrite the function to calculate the P(n,T):

$$P(n,T) = \begin{cases} 1 - \prod_{i=1}^{T-1}(1 - \frac{i}{n}), & \text{if } T \leq n \\ 1, & \text{if } T > n \end{cases}$$

Consider a 64-bit hash function: It has 2^{64} possible different outputs, if we want to make a collision with 100% possibility. We need $2^{64} + 1 \approx 10^{19}$ attacks, but base on the theorem of birthday problem, we only need $2^{32} \approx 10^{9}$ to have 50% probability to attack successfully. It's like we have half income but square root the cost of trial, it is very significant for the big numbers! In order to increase the possibility of collision we can make 2^{32} attacks a round and attack for 10 times, which we need in total 10^{10} attacks but the probability of collision now is higher than 99.9% ! That is very ideal success rate! The number of attack is 1/1,000,000,000 of original method!

Thanks for listening

Questions?