Birthday Paradox

How many people must be there in a room to make the probability 100% that two people in the room have same birthday?
Answer: 367 (since there are 366 possible birthdays, including February 29).
The above question was simple. Try the below question yourself.

How many people must be there in a room to make the probability 50% that two people in the room have same birthday?
Answer: 23
The number is surprisingly very low. In fact, we need only 70 people to make the probability 99.9 %.

Let us discuss the generalized formula.

What is the probability that two persons among n have same birthday?
Let the probability that two people in a room with n have same birthday be P(same). P(Same) can be easily evaluated in terms of P(different) where P(different) is the probability that all of them have different birthday.

P(same) = 1 – P(different)

P(different) can be written as 1 x (364/365) x (363/365) x (362/365) x …. x (1 – (n-1)/365)

How did we get the above expression?
Persons from first to last can get birthdays in following order for all birthdays to be distinct:
The first person can have any birthday among 365
The second person should have a birthday which is not same as first person
The third person should have a birthday which is not same as first two persons.
…………….
……………
The n’th person should have a birthday which is not same as any of the earlier considered (n-1) persons.

Approximation of above expression

The above expression can be approximated using Taylor's Series.

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots$$

provides a first-order approximation for ex for x << 1:

$$e^x \approx 1 + x.$$

To apply this approximation to the first expression derived for p(different), set x = -a / 365. Thus,

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

The above expression derived for p(different) can be written as
1 x (1 – 1/365) x (1 – 2/365) x (1 – 3/365) x …. x (1 – (n-1)/365)

By putting the value of 1 – a/365 as $e^{-a/365}$, we get following.

$$\approx 1 \times e^{-1/365} \times e^{-2/365} \cdots e^{-(n-1)/365}$$
$$= 1 \times e^{-(1+2+\cdots+(n-1))/365}$$
$$= e^{-(n(n-1)/2)/365}.$$

Therefore,

p(same) = 1- p(different) $\approx 1 - e^{-n(n-1)/(2\times 365)}$.

An even coarser approximation is given by

p(same) $\approx 1 - e^{-n^2/(2\times 365)}$,

By taking Log on both sides, we get the reverse formula.

$$n \approx \sqrt{2 \times 365 \ln\left(\frac{1}{1-p(same)}\right)}.$$

Using the above approximate formula, we can approximate number of people for a given probability. For example the following C++ function find() returns the smallest n for which the probability is greater than the given p.

C++ Implementation of approximate formula.

The following is C++ program to approximate number of people for a given probability.

```
// C++ program to approximate number of people in Birthd
// problem
#include <cmath>
#include <iostream>
using namespace std;

// Returns approximate number of people for a given prob
int find(double p)
{
    return ceil(sqrt(2*365*log(1/(1-p))));
}

int main()
{
   cout << find(0.70);
}
```

Output:

```
30
```

Source:
http://en.wikipedia.org/wiki/Birthday_problem

Applications:
1) Birthday Paradox is generally discussed with hashing to show importance of collision handling even for a small set of keys.
2) Birthday Attack