The Birthday Paradox: Only 23 people need to be in the room before it is more likely than not that two people share a birthday. More generally, if there are m people and n possible birthdays then the probability that all m have different birthdays is:

$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\left(1-\frac{3}{n}\right)\cdots\left(1-\frac{m-1}{n}\right) = \prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right).$$

using that $1-\frac{k}{n} \simeq e^{-k/n}$ when k is small compared to n, we see that if m is small compared to n then

$$\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right) \simeq \prod_{j=1}^{m-1} e^{-j/n} = \exp\left\{-\sum_{j=1}^{m-1}\frac{j}{n}\right\} = e^{-m(m-1)/2n}$$

$\simeq e^{-m^2/2n}$ Hence the value for m at which the probability that m people all have different birthdays is $1/2$ is approximately given by the equation

$$\frac{m^2}{2n} = \log_e 2, \text{ or } m = \sqrt{2n\log_e 2}.$$

For the case $n = 365$, this approximation gives $m = 22.49$ to two decimal places, matching the exact calculation quite well. Let us consider each person one at a time, and let E_k be the event that the k^{th} person's birthday does not match any of the birthdays of the first $k-1$ people. Then the probability that the first k people fail to have distinct birthdays is: $P(\overline{E}_1 \cup \overline{E}_2 \cup \cdots \cup \overline{E}_k)$

$$\leq \sum_{i=1}^{k} P(\overline{E}_i) \leq \sum_{i=1}^{k}\frac{i-1}{n} = \frac{k(k-1)}{2n}.$$

If $k \leq \sqrt{n}$ this probability is less than $1/2$, so with $\lfloor\sqrt{n}\rfloor$ people the probability is at least $1/2$ that all birthdays will be distinct.

Now assume that the first $\lceil \sqrt{n} \rceil$ people all have distinct birthdays. Each person after that has probability at least $\sqrt{n}/n = 1/\sqrt{n}$ of having the same birthday as one of these first $\lceil \sqrt{n} \rceil$ people. Hence the probability that the next $\lceil \sqrt{n} \rceil$ people all have different birthdays than the first $\lceil \sqrt{n} \rceil$ people is at most $\left(1-\frac{1}{\sqrt{n}}\right)^{\lceil \sqrt{n} \rceil} < \frac{1}{e} < \frac{1}{2}$.

Hence, once there are $2\lceil \sqrt{n} \rceil$ people, the probability is at most $1/e$ that all birthdays will be distinct.

The Balls-and-Bins Model: The birthday paradox is an example of a more general mathematical framework that is often formulated in terms of balls and bins. We have m balls that are thrown into n bins, with the location of each ball chosen independently and uniformly at random from the n possibilities. What does the distribution of the balls in the bins look like? The question behind the birthday paradox is whether or not there is a bin with two balls.

When n balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \log_e n / \log_e \log_e n$ is at most $1/n$ for n sufficiently large.

The probability that bin 1 receives at least M balls is at most $\binom{n}{M}\left(\frac{1}{n}\right)^M$. This follows from a union bound; there are $\binom{n}{M}$ distinct sets of M balls, and for any set of M balls the probability that all land in bin 1 is $(1/n)^M$. We now use the inequalities $\binom{n}{M}\left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M$.

Here the second inequality is a consequence of the following general bound on factorials: since $\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = e^k$,

we have $k! > \left(\frac{k}{e}\right)^k$.

Applying a union bound again allows us to find that, for $M \geq 3 \log_e n / \log_e \log_e n$, the probability that any bin receives at least M balls is bounded above by

$$
\begin{aligned}
n\left(\frac{e}{M}\right)^M &\leq n\left(\frac{e \log_e \log_e n}{3 \log_e n}\right)^{3 \log_e n / \log_e \log_e n} \\
&\leq n\left(\frac{\log_e \log_e n}{\log_e n}\right)^{3 \log_e n / \log_e \log_e n} \\
&= e^{\log_e n}\left(e^{\log_e \log_e \log_e n - \log_e \log_e n}\right)^{3 \log_e n / \log_e \log_e n} \\
&= e^{-2 \log_e n + 3(\log_e n)(\log_e \log_e \log_e n)/\log_e \log_e n} \\
&\leq \frac{1}{n}
\end{aligned}
$$

for sufficiently large n.