The Birthday Problem

Andreas Klappenecker
The Birthday Problem

What is the probability $p_{\text{uni}}$ that among a group of $m$ people, at least two share the same birthday?
Solution

Let's solve the problem for arbitrary planets. Let's assume that the m people live on a planet that has n days per year. Then

\[
\frac{n(n-1) \cdots (n-m+1)}{n^m}
\]

is the probability that no two share a birthday, so

\[
p_{uni} = 1 - \frac{n(n-1) \cdots (n-m+1)}{n^m} = 1 - \prod_{i=1}^{m-1} \left(1 - \frac{i}{n}\right),
\]

assuming that m \leq n and the birthdays are independent and uniformly distributed.
Lower Bound

Since $1-x \leq \exp(-x)$ holds for all real numbers $x$, we have

$$p_{uni} = 1 - \prod_{i=1}^{m-1} \left(1 - \frac{i}{n}\right)$$

$$\geq 1 - \exp\left(-\frac{\sum_{i=1}^{m-1} i}{n}\right) = 1 - \exp\left(-\frac{(m-1)m}{2n}\right).$$
Consequence

Therefore, if we consider \( m \geq \frac{1}{2} \left(1 + \sqrt{1 - 8n \ln \delta}\right) \) people, where \( \delta \) is a real number in the range \( 0 < \delta \leq 1 \), then the probability \( p_{uni} \) that at least two of them have a common birthday satisfies \( p_{uni} \geq 1 - \delta \). For example, when \( n = 365 \), we have

\[
\begin{array}{c|cccc}
  m & 23 & 42 & 59 & 72 \\
p_{uni} & 0.5 & 0.9 & 0.99 & 0.999 \\
\end{array}
\]
The Flaw

There are fewer births on weekends than during the week.

There are fewer births on July 4 than on other days in July.

There are significant seasonal variations.

=> Birthdays are not uniformly distributed.
Nonuniform Birthday Problem

Let $p_k$ denote the probability that a person is born on the $k$-th day of the year, where $1 \leq k \leq n$. Then the probability $p_{nu}$ that among $m$ people at least two have the same birthday using the distribution $(p_1, p_2, \ldots, p_n)$ of birthdays is given by

$$p_{nu} = 1 - e_m(p_1, p_2, \ldots, p_n),$$

where $e_m$ denotes the $m$-th elementary symmetric function,

$$e_m(x_1, \ldots, x_n) = \sum_{1 \leq j_1 < j_2 < \cdots < j_m \leq n} x_{j_1} x_{j_2} \cdots x_{j_m}.$$
Relation

Any probability distribution majorizes the uniform distribution,

\[(1/n, 1/n, \ldots, 1/n) \prec (p_1, p_2, \ldots, p_n),\]

which means that the sum of the \( k \) largest probabilities in \( \{p_1, \ldots, p_n\} \) is at least \( k/n \) for all \( k \) in the range \( 1 \leq k \leq n \). Since the elementary symmetric functions are Schur-concave (meaning that they are monotonically decreasing with respect to the relation \( \prec \)), it follows that \( e_m(1/n, 1/n, \ldots, 1/n) \geq e_m(p_1, p_2, \ldots, p_n) \).
Relation

Therefore, we can conclude that

\[ p_{\text{uni}} = 1 - \frac{n(n - 1) \cdots (n - m + 1)}{n^m} \]

\[ = 1 - e_m(1/n, 1/n, \ldots, 1/n) \]

\[ \leq 1 - e_m(p_1, p_2, \ldots, p_n) = p_{\text{nu}}. \]
One can show the following relation between uniform and nonuniform distribution case:

\[
p_{\text{uni}} = 1 - \frac{n(n-1) \cdots (n-m+1)}{n^m}
= 1 - e_m(1/n, 1/n, \ldots, 1/n)
\leq 1 - e_m(p_1, p_2, \ldots, p_n) = p_{\text{nu}},
\]

as \( e_m \) is a so-called Schur-concave function.
References

- J. Buchmann, Introduction to Cryptography, Springer, 2004
- J. Michael Steele, The Cauchy Schwarz Master Class, Cambridge University Press, 2004