Elementary “Birthday Paradox” Inequality

(Sorry if this is little bit clumsy, I’m just a cryptographer)

Feb 12, 2004 Markku-Juhani O. Saarinen <mjos@tcs.hut.fi>

We wish to find n (‘number of persons”) as a function of m (‘number of days in year”), so that probability of at least one match is $\frac{1}{2}$.

Since probability of each pair forming a match is $\frac{1}{m}$ and there are $\frac{n(n-1)}{2}$ pairs, the product probability of $\frac{1}{2}$ is reached at:

$$(1-\frac{1}{m})^{\frac{n(n-1)}{2}} = \frac{1}{2},$$

taking logs we get:

$$\frac{n(n-1)}{2}\ln(1-\frac{1}{m}) = -\ln 2$$

which can be written as a quadratic equation as a function of n:

$$n^2 - n + \frac{\ln 4}{\ln(1-\frac{1}{m})} = 0$$

The unique positive solution for this lies at:

$$n = \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{\ln 4}{\ln(1-\frac{1}{m})}}$$

However, this is cumbersome, and we would like to have an easy asymptotic expression for n. We start by bounding $\ln(1-\frac{1}{m})$.

Well-known Mercator series (a Taylor series for logs) is usually written as:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

this easily translates into

$$\ln(1-\frac{1}{x}) = -\frac{1}{x} - \frac{1}{2x^2} - \frac{1}{3x^3} - \frac{1}{4x^4} - \cdots$$

Inspection reveals that this monotonically increasing for $x > 2$ and the following bounds hold:

$$-\frac{1}{x} - \frac{1}{x^2} < \ln(1-\frac{1}{x}) < -\frac{1}{x}.$$

Substituting these into the original “exact” solution and observing trivially that $1/(\frac{1}{x}+\frac{1}{x^2}) = x^2/(x+1)$ we get:

$$\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{m^2}{m+1}\ln 4} \;<\; n \;<\; \frac{1}{2} + \sqrt{\frac{1}{4} + m\ln 4}$$

Removing those terms that vanish when $m \to \infty$, we clearly get:

$$n \approx \sqrt{\ln 4}\sqrt{m} \approx 1.1774\sqrt{m}.$$

NOTE: This derivation not entirely correct! The original assumption that birthday coincidence “events” are independent is not strictly true, it is just a very good approximation for large m.