



planetmath.org

Math for the people, by the people.

approximating the birthday problem

Canonical name	ApproximatingTheBirthdayProblem
Date of creation	2013-03-22 16:09:56
Last modified on	2013-03-22 16:09:56
Owner	Algeboy (12884)
Last modified by	Algeboy (12884)
Numerical id	9
Author	Algeboy (12884)
Entry type	Derivation
Classification	msc 60C05
Classification	msc 05A10
Classification	msc 60-00

A straightforward generalization of the birthday problem can be stated as:

Given N possible outcomes, how many random samples must be taken so that with a probability of at least $1/2$ there are two equal samples?

Answer: Approximately $1.2\sqrt{N}$ samples must be taken.

So in the typical *birthday problem* setting the $N = 365$ – the number of days in the typical year, and the result is that 23 people in a group gives at least a 50% chance of two people with a common birthday. This result can be obtained by a concise computation of the probabilities for various sizes of the group until the proper value of 23 is found, but this approach does not lend itself to finding the value of 23 efficiently.

The result however has an easy to compute approximation of $1.2\sqrt{N}$, so in the case of $N = 365$, the result is approximated by $1.2\sqrt{365} \approx 23$. We now validate the estimate.

Recall the probability of selecting a duplicate from N possible outcomes within k choices is

$$p(N, k) = \binom{N}{k} \frac{1}{N^k} = \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{k-1}{N}\right).$$

As

$$e^x = \lim_{m \rightarrow \infty} \left(1 + \frac{x}{m}\right)^m$$

we can roughly approximate the formal product above by:

$$p(N, k) \approx e^{-1/N - 2/N - \cdots - k/N} = e^{-\binom{k+1}{2} \frac{1}{N}}.$$

Now we are after a formula in which given N and the desired probability of $1/2$ we can estimate k . So we solve for k in the equation:

$$\frac{1}{2} = p(N, k) = e^{-\binom{k}{2} \frac{1}{N}} = e^{-\frac{k(k-1)}{2N}};$$

$$-2N \ln \frac{1}{2} = k^2 - k;$$

$$\frac{1}{4} + 2N \ln 2 = k^2 - k + \frac{1}{4} = \left(k - \frac{1}{2}\right)^2;$$

$$k = \frac{1}{2} + \sqrt{\frac{1}{4} + 2N \ln 2}. \quad (1)$$

A less precise but easier to compute approximation comes from simplifying our original estimate further:

$$p(N, k) \approx e^{-\frac{k(k-1)}{2N}} \approx e^{-k^2/2N}.$$

Then set $1/2 = e^{-\frac{k(k-1)}{2N}} \approx e^{-k^2/2N}$ and once again solve for k to find:

$$k = \sqrt{2N \ln 2} \approx 1.2\sqrt{N}.$$

We also see that if we wish to have an outcome with probability $\varepsilon > 0$ of no matches then we can estimate the sample size by:

$$\varepsilon = p(N, k) \approx e^{-k^2/2N}; \quad k \approx \sqrt{2N \ln \frac{1}{\varepsilon}}.$$

So the number of samples required to find a match with probability $\varepsilon > 0$ is approximately

$$k = \sqrt{2N \frac{1}{1 - \varepsilon}}.$$

References

- [1] Wikipedia, *Birthday paradox*, http://en.wikipedia.org/w/index.php?title=Birthday_paradox&oldid=69569553. Accessed 14 August, 2006.