Project Euler: Problem 323

> **Problem.** Let $y_1, y_2, \ldots$ be a sequence of random unsigned 32-bit integers that are uniformly chosen. Define a sequence (x_k) such that $x_1 = y_1$ and $x_k = x_{k-1} \mid y_k$ (bitwise OR).
>
> It can be seen that there will eventually be a smallest index N such that $x_N = 2^{32} - 1$. Find the expected value of N to 10 digits after the decimal point.

Solution. When I first saw this problem, I thought it was harder than it looked, but once you start thinking in terms of random variables, it actually becomes quite easy.

Definition. Define a *bit random variable* y to be a discrete random variable which takes on the values 0 and 1 with equal probability $1/2$.

We can decompose each y_k into bit random variables, which we shall notate as follows:

$$y_k = (y_k[0], y_k[1], \ldots, y_k[31]),$$

where $y_k[i]$ represents the ith bit of y_k. This is motivated by the fact that, since each y_k is uniformly distributed, we can treat it as a combination of independent, similarly uniform, random variables. We can decompose each x_k in the same fashion.

There are two important observations to make:

- The probability that $x_k = 2^{32} - 1 = 11\ldots1_2$ can be found by taking the intersections of the events of the individual bits being 1. In other words,

$$P(x_k = 2^{32} - 1) = \prod_i P(x_k[i] = 1).$$

By symmetry, we can simplify this further by noticing that the probability that any one of these bits are 1 is the exact same. So,

$$\prod_i P(x_k[i] = 1) = P(x_k[0] = 1)^{32}.$$

- The probability $P(x_k[i] = 1)$ is rather easy to calculate. Observe that
$$P(x_k[i] = 1) = 1 - P(x_k[i] = 0) = 1 - \frac{1}{2^k}.$$

With this, the expected value of N, the smallest index, is easy to find. The probability that some index n is the smallest such index is the difference between the probability that x_n is $2^{32} - 1$ and the probability that x_{n-1} is $2^{32} - 1$. Thus

$$E[N] = \sum_{n=1}^{\infty} n \left((1 - 2^{-n})^{32} - (1 - 2^{-(n-1)})^{32} \right).$$

One can very easily truncate this to about 50 terms to get the answer to 10 decimal places.