Classic Riddler

18 September 2020

Riddle:

One of Ollie's favorite online games is Guess My Word. Each day, there is a secret word, and you try to guess it as efficiently as possible by typing in other words. After each guess, you are told whether the secret word is alphabetically before or after your guess. The game stops and congratulates you when you have guessed the secret word. For example, the secret word was recently "nuance," which Ollie arrived at with the following series of nine guesses: naan, vacuum, rabbi, papa, oasis, nuclear, nix, noxious, nuance.

Each secret word is randomly chosen from a dictionary with exactly 267,751 entries. If you have this dictionary memorized, and play the game as efficiently as possible, how many guesses should you expect to make to guess the secret word?

Solution:

The most efficient way to search in a game like this is a binary search. That means the first guess should be the middle word alphabetically. In the almost-certain event that this guess is wrong, then the search space is divided exactly in half; the "before" region has the same number of words as the "after" region. That leaves two possible words for the second guess: the middle of each of the "before" and "after" regions. Similarly, if the next guess is also wrong, there are now four regions and four possible third guesses. This process continues until the word is found, either by luck or by exhaustion of the search space (the dictionary). Each turn has twice as many possible guesses as the previous, until the exhaustive guess.

In this specific case, with 267,761 words, the first guess should be the 133,876th word alphabetically. If this is wrong, the next guess should be either of the 66,938th word (before) or the 200,813th word (after). On the 18th round, there are 131,072 possible guesses, for a cumulative total of 262,143. In the worst case, there would be at most 19 rounds, and there are only 5,608 available words left to guess in the 19th round. Denoting the number of possible guesses for round n as g_n, then we can write

$$g_n = \begin{cases} 2^{n-1} & 1 \le n \le 18 \\ 5,608 & n = 19 \end{cases}$$

This satisfies

$$\sum_{n=1}^{19} g_n = 267,751$$

The expected number of guesses E_g is of course just the average of these values, weighted by the number of guesses to get there:

$$E_g = \frac{1}{267,751} \sum_{n=1}^{19} n \cdot g_n$$

Evaluating this sum gives the (approximate) solution $\boxed{17.04}$.