Exercise 5.2 Wheel of Fortune

Task

You are shown N cards, each of which cover one letter. Each letter has been independently chosen from the same distribution, and you are given the distribution ($p_0, p_1, \ldots, p_{25}$). You get to choose one letter from the alphabet, say you choose letter number i. Now every position in the hidden string where letter i occurs (if any) are uncovered. Your goal is to learn (on average) as much information as possible on the hidden string.

Solution

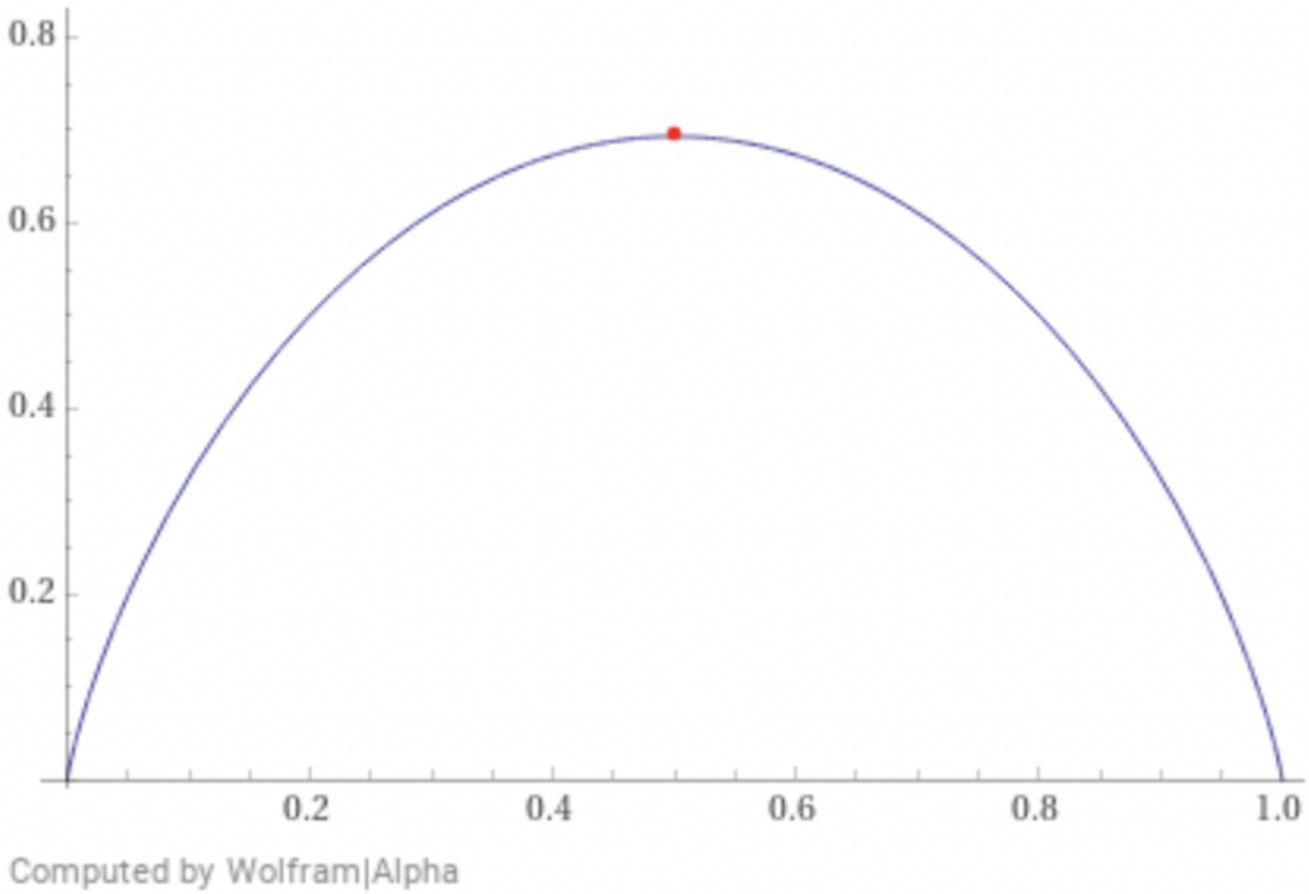
1. Now, if your guess is letter nr. i, how many bits of information will you learn on average from playing the game (as a function of p_i and N)?

We would learn:

$$f(p_i) = N \times \left(p_i \times log(\frac{1}{p_i}) + (1 - p_i) \times log(\frac{1}{1 - p_i}) \right)$$

2. What strategy does your result suggest for choosing your guess, given frequencies $p_0, \ldots, p_{25}$ as in English?

$f(p_i)$ reaches its maximum in $p_i = \frac{1}{2}$, so we want to pick letters whose distribution is as close to $\frac{1}{2}$ as possible. In English that would be the most common letters: E (12.49%), T (9.28%), A (8.04%), etc.



3. Based on this, does it make sense that players in real life choose the most frequent letter(s)? Why or why not?

Based on the observation in the previous step it makes sense that players in real life choose the most common letters first. It would lead them to learning the most about the hidden phrase on average.

4. Would this be a good strategy no matter what the frequencies were?

It would be a good strategy in most cases. It wouldn't be in the case that all frequencies are the same - then there is no "most-frequent" letter, so people would have to guess blindly.

Also, this strategy works assuming that the people do not have any additional information. e.g. If there are multiple words, and they have spacing, and you have a two-letter word in the middle of other longer lords, you could guess that it is either 'is' or 'of'. Then it would make more sense to try to guess either 'o' or 'i' as a first vowel instead of 'e', even if they aren't as common.