Example 0.1

Suppose you get to play the following variant of the bonus round in the TV show Wheel of Fortune (Lykkehjulet): 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 $(p_0, \ldots, p_{25})$. 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.

Of course this is a very crude model of Wheel of Fortune since we only consider single letter frequencies, but we want something that is feasible to analyze.

In the real-life version of the game, people tend to choose the most frequent letters as their guesses. Let's try to see what information theory has to say about this. Suppose we adopt the convention that Shannon used when defining Entropy: if you know that some event occurs with probability p, and you learn that this event did indeed occur, you have learnt $\log(1/p)$ bits of information.

1. *Now, if your guess is letter number i, how many bits of information will you learn on average from playing the game (as a function of p_i and N)? Hint: note that you learn something for every position in the hidden string, namely either that letter i occurred here, or that it did not occur.*

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

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

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

Solution:

1. Suppose we guess letter i. For each letter i revealed in the message, we learn $\log\left(\frac{1}{p_i}\right)$ bits of information. In the message, we expect to see Np_i ocurrences of letter i.

   Then, from the same guess, we also learn information from those letters which are not revealed. Each of these reveals $\log\frac{1}{1-p_i}$ bits of information, and we will expect to see $N(1-p_i)$ such un-revealed letters.

   So in total, we should expect $Np_i \log\frac{1}{p_i} + N(1-p_i) = N\left(p_i \log\frac{1}{p_i} + (1-p_i)\log\frac{1}{1-p_i}\right)$ bits of information per guess.

2. In terms of strategy, this suggests that the way to reveal the most information per guess is to maximize $f(x) = -x\log x - (1-x)\log(1-x)$, where x is the letter frequency.
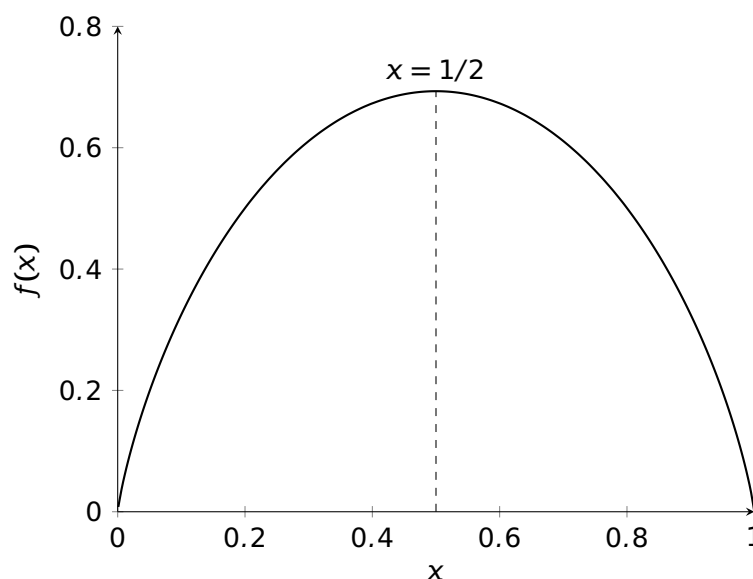
We use calculus or consult the graph on the previous page to find the function is maximised at $x = \frac{1}{2}$. So to maximize information per guess we should pick the letters which have p_i as close to $\frac{1}{2}$ as possible.

3. The strategy above lends itself well to English, since every letter has frequency less than $\frac{1}{2}$. This means the strategy in English Wheel of Fortune is to pick the most common letters, and that the typical strategy is the correct one.

   Suppose for the sake of contradiction, that i, j are letters, with $p_i < p_j$, and that the entropy, $f(p_i) > f(p_j)$. This must mean that f is decreasing between p_i and p_j, and we know our function f decreases on $[\frac{1}{2}, 1]$. Therefore both p_i, p_j are greater or equal to $\frac{1}{2}$. This contradicts the axioms for a valid propbability distribution, and therefore we cannot have a letter with lower frequency but higher entropy.