Test your knowledge: Information Theory

You are playing "Twenty Questions". This is a guessing game of yes/no questions, where the goal is to guess what item you have in mind. Here's an example:

You: Is it bigger than a breadbox?
Opponent: No.
You: Is it alive?
Opponent: Yes.
You: Is it a mouse?
Opponent: No.

This game can be analyzed using Information Theory. We'll use a simple example, where your friend can choose from only three things to think about: "animal, vegetable, or mineral."

Your friend is biased towards "animal". She is 50% likely to choose animal, 25% likely to choose vegetable, and 25% likely to choose mineral.

Question 1: What is the uncertainty in this distribution in bits?

You'll need to remember the formula for uncertainty:

$$H(\{p_1, \ldots, p_N\}) = -\sum_{i=1}^{N} p_i \log p_i$$

where the *log* is base two. (So that log(1/8) is -3, because 1/8 is 2^{-3})

You also learned that the uncertainty of a distribution is related to the average number of YES/NO questions you need to ask, if you choose them optimally.

(To be super-correct: the uncertainty is a *lower bound* on the average number of questions; the actual average may be between H and H+1, in bits.) So they're close but not perfect.

Question 2: For the game above, devise a set of questions, and their order, that will achieve this optimal bound.

Remember: the questions can have the answer "YES" or "NO" only. You should come up with a "script". Once you've eliminated all but one possibility, you know the answer.

You will need to roleplay out the question-asking process for the different possible choices your opponent may make.

Finally, you will need to work out the average number of questions, given the possible choices your opponent makes, and the probability that he chooses the different options.

Is the average number of questions equal to the uncertainty in bits?

Question 3: Imagine that the pattern of probabilities above (1/2, 1/4, 1/4) for (animal, vegetable, mineral) is commonly chosen by women.

However, imagine that men prefer (1/4, 1/4, 1/2). If you're facing a man but using the script for a woman, how many questions would your previous script need?