Unigram model

For a vocabulary $V = \{v_k\}_k$, consider a unigram model $\mathbf{p} = [p_k]_{k=0}^{|V|-1}$ where n_k is the number of observations of v_k, and $p_k = \frac{n_k}{\sum_k n_k}$.

Prove that this $\mathbf{p}$ is optimal - it maximizes the probability of the set of observations.

Consider a scenario where $V = \{\text{“apple”}, \text{“banana”}\}$ and the following are observed:

$$[\text{“apple”}, \text{“apple”}, \text{“apple”}, \text{“apple”}, \text{“apple”}, \text{“apple”}, \text{“banana”}, \text{“banana”}, \text{“banana”}, \text{“banana”}]$$

Plot the probability of the observations under the unigram model, as a function of p_{apple} ($= 1 - p_{\text{banana}}$). Use `matplotlib`.

You should turn in a document (`.txt`, `.md`, or `.pdf`) answering all of the **red** items above. You should also turn in Python scripts (`.py`) for *each* of the **blue** items. Unless otherwise specified, you may use only `numpy` and the standard library.