In this worksheet we will explore a different interpretation of entropy.

The twenty questions game is a game between two players, Alice and Bob. Bob chooses a distribution $\mu$ over $[n]$ and sends it to Alice. Bob then draws an element $x \sim \mu$, and Alice’s task is to find $x$ using Yes/No questions. The game proceeds in rounds. At each round, Alice asks Bob a question of the form “$x \in S$?”, and Bob answers truthfully. The game ends when Alice knows $x$, that is, when there is a unique element in the support of $\mu$ which is consistent with all of Bob’s answers. The cost of a strategy for Alice is the average number of questions that Alice asks until she knows $x$, where the average is taken with respect to $\mu$. The optimal cost of $\mu$, denoted $c(\mu)$, is the minimal cost of a strategy for Alice.

1. Show that $c(\mu)$ equals $T(\mu)$, the cost of a Huffman code for $\mu$ (assuming we only need to supply codewords for elements in the support of $\mu$).

   A distribution $\mu$ is dyadic if the probability of every element in the support of $\mu$ is of the form $2^{-m}$ (for integer $m$). It is constant if $\mu(x) = 1$ for some $x$.

2. Show that $c(\mu) = H(\mu)$ iff $\mu$ is dyadic.

Let $Q$ be a collection of subsets of $[n]$. A $Q$-strategy for Alice is one in which she always asks questions of the form “$x \in S$?” for $S \in Q$. A collection $Q$ is optimal (for $n$) if for every $\mu$ there is a $Q$-strategy with cost $c(\mu)$.

3. Show that $Q$ is optimal iff for every dyadic $\mu$ there is a $Q$-strategy with cost $H(\mu)$.

4. A collection $Q$ is a dyadic hitter if for every non-constant dyadic $\mu$ there is a set $S \in Q$ such that $\mu(S) = 1/2$.

   (a) Show that $Q$ is optimal iff it is a dyadic hitter.
   (b) Show that in the definition of dyadic hitter, we can assume that $\mu$ has full support.

Let Opt($n$) be the minimal size of an optimal set of questions for $n$.

5. **Lower bound:** A $(k, r)$-almost-uniform distribution is one in which the probabilities of the elements are $2^{k-1}, \ldots, 2^{k-r}, 2^{k-r}, \ldots, 2^{k-1}, 2^{-r}$.

   (a) Suppose that $2^k = \alpha n$. Obtain a lower bound on Opt($n$) by considering all $(k, n-2^k)$-almost-uniform distributions.
   (b) Optimizing over $\alpha$, find the best $C$ so that Opt($n$) $\geq C^{n-o(n)}$ for infinitely many $n$.

6. **Simple upper bound:** Let $Q = \{A: A \subseteq \lfloor n/2 \rfloor \text{ or } A \supseteq \lceil n/2 \rceil\}$.

   Show that $Q$ is optimal, and deduce an upper bound on Opt($n$).

7. **Tight upper bound:** For a non-constant dyadic $\mu$ having full support, let $D(\mu) = \{|S: \mu(S) = 1/2\}$. Suppose that there exists $p(n) > 0$ such that for each such $\mu$, there exists $i$ such that $|D(\mu) \cap \binom{[n]}{i}| \geq p(n)\binom{n}{i}$.

   (a) Find a reasonable upper bound on the number of non-constant dyadic $\mu$.
   (b) Show that Opt($n$) $\leq n^{O(1)}/p(n)$.
   (c) Show that Opt($n$) $\leq C^{n+o(n)}$, where $C$ is the constant from the lower bound. (Hard!)