Twenty Questions

Yuval Filmus

November 27, 2017

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 μ 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 μ 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 μ. The *optimal cost* of μ, 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 μ (assuming we only need to supply codewords for elements in the support of μ).

A distribution μ is *dyadic* if the probability of every element in the support of μ 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 μ is dyadic.

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

3. Show that $\mathcal{Q}$ is optimal iff for every dyadic μ there is a $\mathcal{Q}$-strategy with cost $H(\mu)$.
4. A collection $\mathcal{Q}$ is a *dyadic hitter* if for every non-constant dyadic μ there is a set $S \in \mathcal{Q}$ such that $\mu(\mathcal{Q}) = 1/2$.
   (a) Show that $\mathcal{Q}$ is optimal iff it is a dyadic hitter.
   (b) Show that in the definition of dyadic hitter, we can assume that μ has full support.

Let $\operatorname{Opt}(n)$ 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 $\overbrace{2^{-k}, \ldots, 2^{-k}}^{2^k-1}, \overbrace{2^{-k-1}, \ldots, 2^{-k-r}}^{r}, 2^{-k-r}$.
   (a) Suppose that $2^k = \alpha n$. Obtain a lower bound on $\operatorname{Opt}(n)$ by considering all $(k, n-2^k)$-almost-uniform distributions.
   (b) Optimizing over α, find the best C so that $\operatorname{Opt}(n) \geq C^{n-o(n)}$ for infinitely many n.
6. **Simple upper bound:** Let $\mathcal{Q} = \{A : A \subseteq [\lfloor n/2 \rfloor] \text{ or } A \supseteq [\lfloor n/2 \rfloor]\}$.
   Show that $\mathcal{Q}$ is optimal, and deduce an upper bound on $\operatorname{Opt}(n)$.
7. **Tight upper bound:** For a non-constant dyadic μ having full support, let $D(\mu) = \{S : \mu(S) = 1/2\}$. Suppose that there exists $p(n) > 0$ such that for each such μ, 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 μ.
   (b) Show that $\operatorname{Opt}(n) \leq n^{O(1)}/p(n)$.
   (c) Show that $\operatorname{Opt}(n) \leq C^{n+o(n)}$, where C is the constant from the lower bound. (Hard!)