Twenty (simple) questions

Yuval Dagan, **Yuval Filmus**, Ariel Gabizon, Shay Moran

March 1, 2017

1 Introduction

The twenty questions game is a two-player game. Alice chooses an element $x \in \{1, \ldots, n\}$ according to a distribution μ, and Bob, who knows μ, identifies x using Yes/No questions. Bob's goal is to identify x using as few questions as possible on average (the average being taken according to μ). Let us denote the optimal number of questions for a distribution μ by $T(\mu)$.

Here are a few facts about $T(\mu)$:

1. $T(\mu) \geq H(\mu)$, where $H(\mu)$ is the binary entropy of μ.
2. $T(\mu) < H(\mu) + 1$.
3. $T(\mu)$ can be found efficiently using Huffman's algorithm.

Indeed, a strategy for Bob corresponds to a prefix code, and the average number of questions is the same as the average length of a codeword.

The twenty questions game is the starting point of *combinatorial search theory*, which studies many different scenarios, for example what happens when Bob is allowed to lie. Our interest lies elsewhere.

The questions that Bob asks in the optimal strategy can be arbitrary. What happens if we restrict the set of questions that Bob is allowed to use? If Bob is only allowed to use questions of the form "$x < i$?" (that is, Bob's strategy is a binary search tree), then Bob always has a strategy whose cost is at most $H(\mu)+2$. This is a classical result due to Gilbert–Moore, Shannon–Fano–Elias and Horibe. This almost matches the performance of Shannon–Fano codes, which use arbitrary questions and achieve a cost of at most $H(\mu) + 1$.

Our central question is: *Are there "simple" sets of questions with "good" performance?* To formalize this questions, we will need a few definitions.

Let $\mathcal{Q}$ be a set of questions over $\{1, \ldots, n\}$, i.e., a subset of $2^{\{1,\ldots,n\}}$. We define:

- The *redundancy* of $\mathcal{Q}$, denoted $r_H(\mathcal{Q})$, is the infimal r such that for any distribution μ over $\{1, \ldots, n\}$ there is a strategy, using only questions from μ, which asks at most $H(\mu) + r$ questions.
- The *prolixity* of $\mathcal{Q}$, denoted $r_T(\mathcal{Q})$, is defined analogously but with respect to $T(\mu)$.

This explains how we measure quality. We will quantify simplicity by the *size* of the set of questions, though qualitatively we will also be interested in how natural a set of questions is, and how efficient it is to answer questions from the set and to construct good strategies. Given n, r, we define:

- $u_H(n, r)$ is the minimal size of a set of questions $\mathcal{Q}$ such that $r_H(\mathcal{Q}) \le r$.
- $u_T(n, r)$ is the minimal size of a set of questions $\mathcal{Q}$ such that $r_T(\mathcal{Q}) \le r$.

In our paper we:

1. Determine the order of growth of $u_H(n, r)$ for all r.
2. Determine the order of growth of $u_T(n, 0)$ for infinitely many n, and obtain good bounds for all n.
3. Essentially determine the order of growth of $u_T(n, r)$ for all r.

We also have a few other results, which we probably won't have time to mention.

2 Achieving small redundancy

We start our exploration by considering $u_H(n, 1)$. Note that 1 is the optimal number here: a distribution almost concentrated on a single element x has entropy roughly 0, but requires at least one question to identify x.

As we noted above, $r_H(\{x < i\}) = 2$, the lower bound following from distributions almost concentrated on $x \neq 1, n$. There are two algorithms for constructing strategies achieving this bound: Gilbert–Moore / Shannon–Fano–Elias, which is related to arithmetic coding, and Rissanen–Horibe, which is based on the *weight balancing* heuristic:

1. Ask the question "$x < i$?" which partitions $\{1, \ldots, n\}$ into two sets whose measures with respect to μ are as close to $1/2$ as possible.
2. Condition μ on the answer, and repeat if necessary.

Any set of questions which achieves redundancy 1 must include all questions of the form "$x = i$?", in order to handle distributions almost concentrated on i. We show that allowing both types of questions achieves redundancy 1:

$$r_H(\{x < i, x = i\}) = 1.$$

This shows that $n \le u_H(n, 1) \le 2n - 3$.

The strategy which shows this is a very natural modification of the weight balancing heuristic:

1. If the maximum probability element i has probability at least 0.3, ask the question "$x = i$?".
2. Otherwise, ask the question "$x < i$?" which partitions $\{1, \ldots, n\}$ into two sets whose measures with respect to μ are as close to $1/2$ as possible.
3. Condition μ on the answer, and repeat if necessary.

Thresholds in some interval around 0.3 also work. The proof that this strategy achieves redundancy 1 is a challenging exercise.

As a corollary, we obtain the following result for integer $r \ge 1$:

$$u_H(r) = \Theta(rn^{1/r}).$$

(It turns out that for real r, the answer depends only on $\lfloor r \rfloor$.)

We won't prove this in full, but only give some hints:

Upper bound Think of x as encoded in base $n^{1/r}$. Determine the digits of x one by one, for a total redundancy of r.

Lower bound Consider distributions almost concentrated at a single element. For each $x \in \{1, \ldots, n\}$, there is a set of at most r question/answer pairs that identify it, hence $\binom{2|\mathcal{Q}|}{\leq r} \geq n$.

We mention in passing that an algorithm inspired by Gilbert–Moore shows that

$$u_T(\{i \leq x \leq j\}) \leq 1/2.$$

3 Optimal sets of questions

We continue our exploration with $u_T(n, 0)$. A set of questions $\mathcal{Q}$ is *optimal* if it has zero prolixity: $r_T(\mathcal{Q}) = 0$. The set of all questions is optimal. Can we improve on that?

Here is what we show:

1. Let $n \approx 1.25 \cdot 2^m$. Then $u_T(n, 0) \geq 1.25^{n-o(n)}$.
2. For all n, $u_T(n, 0) \geq 1.23^{n-o(n)}$.
3. For all n, $u_T(n, 0) \leq 1.25^{n+o(n)}$.

Notice that for $n \approx 1.25 \cdot 2^m$, we have determined $u_T(n, 0)$ up to subexponential factors. We explain below where 1.25 comes from.

Before delving further into $u_T(n, 0)$, let us mention what happens when we allow a small prolixity:

$$u_T(n, r) = \Theta((rn)^{\Theta(1/r)}).$$

Thus if we allow constant prolixity, for whatever constant, then we can make do with only a polynomial number of questions; and we can make do with a subexponential number of questions as long as $r = \omega(1/n)$.

Back to $u_T(n, 0)$. It seems hard at first to say anything about the space of *all* distributions. The following two simple observations will make our life much easier:

Definition. A distribution is *dyadic* if the probability of each element is zero or an integer power of 2.

Observation 1. A set of questions $\mathcal{Q}$ is optimal iff for every dyadic distribution μ there is a decision tree using only questions from $\mathcal{Q}$ whose cost is $H(\mu)$.

Observation 2. A set of questions $\mathcal{Q}$ is optimal iff for every non-constant distribution μ it contains a questions A such that $\mu(A) = 1/2$. We say that A *splits* μ.

Observation 1 follows by "rounding" an arbitrary distribution to the dyadic distribution corresponding to its optimal decision tree. Observation 2 follows from Observation 1 and the chain rule for entropy.

We can now prove the lower bound on the size of an optimal set of questions. Let $n = 2.5 \cdot 2^m$, and consider all permutations of the distribution

$$\underbrace{2^{-m}, \ldots, 2^{-m}}_{2^m - 1 \text{ times}}, 2^{-m-1}, 2^{-m-2}, \ldots, 2^{-m-1.5 \cdot 2^m}, 2^{-m-1.5 \cdot 2^m}.$$

The only questions splitting this distribution are those containing 2^{m-1} out of the first $2^m - 1$ elements (or their complements), and each such question only handles so many distributions. A simple calculation shows that 0.4 is indeed the optimal fraction of elements in the first part of the distribution, and yields the lower bound.

The upper bound is a bit more complicated. The idea is to show that for each non-constant distribution μ there is a question size i and at least $\binom{n}{i}/1.25^{n+o(n)}$ questions of that size that split μ; this is another challenging exercise. The same number 1.25 shows up again, since it is obtained by solving the same optimization problem.

Given this, we construct an optimal set of questions by choosing roughly 1.25^n questions of each size at random. The union bound shows that with high probability, these questions split every non-constant dyadic distribution, and so we're done by Observation 2. Here it is very helpful that there are only about n^n dyadic distributions.

The construction we have just sketched is not very satisfying: it is not explicit, and constructing an optimal strategy takes exponential time. Indeed, even answering a question would require an exponentially large table!

The best deterministic construction we know of uses $O(\sqrt{2}^n)$ questions: all subsets and all supersets of an arbitrary set A of size $n/2$. We show that this set works using Observation 2. Let μ be a non-constant dyadic distribution. We consider three cases:

1. $\mu(A) = 1/2$. In that case, A splits μ.
2. $\mu(A) > 1/2$. Arrange the elements of A in weakly decreasing order of probability. It is not hard to show that some prefix of that order has probability exactly $1/2$.
3. $\mu(A) < 1/2$. This is similar to the preceding case.

This not only shows that our set of questions is optimal, but also gives an efficient way of constructing an optimal decision tree.

4 Odds and ends

Finally, let us mention one more result from our paper. When discussing $u_H(n, 1)$, we have seen that the hardest distributions are those almost concentrated on a single element. But these distributions aren't very interesting! What happens if we forbid such elements?

Formally, let $r_H(\mathcal{Q}, \epsilon)$ be the redundancy of $\mathcal{Q}$ on distributions in which all elements have probability at most ϵ, and let $r_H(\mathcal{Q}, 0) = \inf_{\epsilon>0} r_H(\mathcal{Q}, \epsilon)$. Using an algorithm inspired by Gilbert–Moore (and similar to the one mentioned above using queries "$i \le x \le j$?"), we show that

$$0.5011 < r_H(\{x < i, x = i\}, 0) < 0.586.$$

5 Open questions

The three most interesting open questions are:

1. Find an explicit optimal set of questions of size $1.25^{n+o(n)}$.
2. Determine $u_T(n, 0)$ for all n; we believe that the answer depends on $\{\log_2 n\}$.
3. Determine the first-order asymptotics of $u_H(n, 1)$.