Determining who's who

Adam Bargteil[*] Univ. of MD, College Park
Alex Chan[†] Univ. of MD, College Park
William Gasarch[‡] Univ. of MD, College Park

1 This is Not Quite Adam's Honors Thesis

This document is a paper-in-progress that grew out of Adam Bartgeil's Honors thesis.

2 Introduction

Given a group of n people, where $a < n$ answer arbitrarily (hereafter, randomly) and the rest answer truthfully, determine who is random and who is honest. Only questions directed at one person and asking about the nature of one other person are permitted. The object is to ask as few questions as possible. We denote the number of questions by $C(a, n)$.

We show that for a general a and n if $a < \frac{n}{2}$ then $C(a, n) = n + a - 1$.

3 General Upper Bound

Theorem 3.1 *If $a < \frac{n}{2}$ then $C(a, n) \leq n + a - 1$.*

Proof: We will eumerate the persons $A_1, \ldots, A_n$. We will ask $A_2, \ldots, A_n$ whether A_1 is random until one of the following occur

1. a persons say that the A_1 is non-random. Then A_1 must be non-random, otherwise we would have $a + 1$ random people. This is not possible if their are at most a random people.

   Suppose that m people said A_1 is random, then these m people are random. But, we must still establish the nature of the remaining $n-m-1$ remaining people. We do so by asking A_1 about these people. This gives us of $a+m$ questions to determine the nature of A_1 and $n-m-1$ questions to determine the rest. Giving a total of $a+m+n-m-1 = n+a-1$ questions to determine the nature of $A_1, \ldots, A_n$

2. The number of people who have said A_1 is random exceeds the number who said he was non-random. In this case, if m people said A_1 is random, then $(m - 1)$ people said he was non-random. If A_1 is random, then so are the $m - 1$ people who said otherwise. Similarly, if A_1 is non-random, then the m people who said he was random must be

[*]Dept. of C.S., University of CA at Berkeley., Berkeley, CA 94720, U.S.A. Supported in part by NSF grant CCR–9732692. (Email: adamb@cs.berkeley.edu.)
[†]Dept. of C.S., University of MD., College Park, MD 20742, U.S.A. (Email: alexchan@cs.umd.edu.)
[‡]Dept. of C.S. and Inst. for Adv. Comp. Stud., University of MD., College Park, MD 20742, U.S.A. Supported in part by NSF grant CCR–9732692. (Email: gasarch@cs.umd.edu.)

random. It must be, then, that in the group of $2m$ people $(A_1, \ldots, A_{2m})$ at least m of them are random.

So, of the $n-2m$ people left, the non-random people must be in the majority. Indeed there will be at most $a-m$ random people. They can be sorted out recursively in $n-2m+a-m-1$ questions. We then ask one of the non-random people from this group about A_1. We then ask about the $\leq m$ people who answered correctly about A_1. The whole process then takes $(2m-1)+(n-2m+a-m-1)+m+1 = n+a-1$ questions.

Note that either case 1 or case 2 must occur. ∎

4 General Lower Bound

Theorem 4.1 *If $a < \frac{n}{2}$, then $C(a,n) \geq n+a-1$.*

Proof: We shall give a strategy of answers such that after the first $n+a-2$ questions there are at least two assignments of $A_1, \ldots, A_n$ to {random, non-random} which agrees with all the answers given.

For Stage 1 (the first $a-1$ questions), the answer is that the person asked about is random.

Notation:

Let $\mathcal{V}$ be the set of people who have been asked questions or asked about.

Let $\mathcal{W}$ be the set of people who have not been asked questions or asked about. ($\mathcal{W} = \mathcal{V}^C$).

Let G be the graph whose vertices are $\mathcal{V}$ and has the edge (u,v) if u was asked about v.

Let $G_1, \ldots, G_r$ be the connected components of G.

Let $\mathcal{V}_1, \ldots, \mathcal{V}_r$ be the vertices in the subgraphs $G_1, \ldots, G_r$.

Let i be the question index $i = 1, \ldots, n+a-2$.

Let $\mathcal{V}[i]$ = the set of people asked about in questions $a, \ldots, i$, and is defined for $i = a, \ldots, n+a-1$.

Initialize $\mathcal{V}^0 = \emptyset$.

For Stage 2 (the next $n-1$ questions), we answer:

- yes if we asked about a member of $\mathcal{W}$
- no if we asked about a member of $\mathcal{V} \backslash \mathcal{V}^0$
- yes if we asked about a member of $\mathcal{V}$

We then update $\mathcal{V}^0$ as follows: If we asked about $v \in \mathcal{V}_m$, $v \notin \mathcal{V}[i-1]$ and $\forall u \in \mathcal{V}_m$, $u \neq v \rightarrow u \in \mathcal{V}[i-1]$, we set $\mathcal{V}^0 = \mathcal{V}^0 + v$. That is, if we if we ask about v for the first time in stage 2 and we have asked about all $u \neq v \in \mathcal{V}_m$ in stage 2, we add v to $\mathcal{V}^0$.

Upon completetion of stage 2, note that $\mathcal{V}^0[i]$ contains at most one vertex from any $\mathcal{V}_m$, $m = 1, \ldots, r$.

Let $\mathcal{V}^{0\prime}$ be defined as follows: for each $m = 1, \ldots, r$ if not all the people in $\mathcal{V}_m$ have not been asked about in Stage 2, we select an arbitrary member of $\mathcal{V}_m$ not asked about in Stage 2 and add them to $\mathcal{V}^{0\prime}$.

By construction, $\mathcal{V}^0$ and $\mathcal{V}^{0\prime}$ are disjoint, and each $\mathcal{V}_m$ contains exactly one element of $\mathcal{V}^0 \cup \mathcal{V}^{0\prime}$.

Now consider the partition

- A_n is non-random if $A_n \in \mathcal{W} \cup \mathcal{V}^0 \cup \mathcal{V}^{0\prime}$
- A_n is random if $A_n \in \mathcal{V} \backslash (\mathcal{V}^0 \cup \mathcal{V}^{0\prime})$

Consider the questions in Stage 1: $s \in \mathcal{W}$ does not participate in these questions.

$\forall s \in \mathcal{V}^0 \cup \mathcal{V}^{0\prime}$, $s \in \mathcal{V}_m$ was asked about $s' \in \mathcal{V}_m$ so $s' \notin \mathcal{V}_m$ so the answer of random was correct.

Now consider the questions in Stage 2: we answered that a person was non-random if they were in $\mathcal{W} \cup \mathcal{V}^0$. If they were in $\mathcal{V}^{0\prime}$, then they were not asked about, so these answers agree with the partition.

All random people are in some $\mathcal{V}_m$, but there is always 1 non-random person in each $\mathcal{V}_m$. So, the total number of random people is:

$$\sum_{m=1}^{r} (|\mathcal{V}_m| - 1) \leq a - 1$$

Let $s \notin \mathcal{V}[n + a - 2]$ such that $s \in \mathcal{W} \cup \mathcal{V}^{0\prime}$. This is possible because there is at least one person not asked about in Stage 2, this person is either in $\mathcal{W}$ or is in some $\mathcal{V}_m$. If $s \notin \mathcal{V}^{0\prime}$ then $\exists s' \in \mathcal{V}_m \cap \mathcal{V}^{0\prime}$ and we set $s = s'$

Notice that s is non-random. If we set $s =$random, we are still consistent with all answers since all answers in Stage 1 about s said that he was random and there were no answers about s in Stage 2. We still have a majority of non-random people, and so this new partition is just as good as the initial one given. ∎

5 $a \geq \frac{n}{2}$

If $a \geq \frac{n}{2}$ then no matter how many questions are asked, the exact nature of the people cannot be found. However, we can limit the number of possible assignments of the group to $\{random, non-random\}$. We will let $P(a, n)$ represent the maximum number of possibile assignments that there will be once we have asked as many questions as we like. Certainly, for $a < \frac{n}{2}$, $P(a, n) = 1$. Since the number of questions asked is not a function of (a, n), for $a \geq \frac{n}{2}$ let $C(a, n)$ denote the maximum number of questions asked by our algorithm on any inputs of size (a, n). The next theorem shows that even for $a \geq \frac{n}{2}$, the number of possibilities is not that large.

Theorem 5.1 $P(a, n) \leq 1 + \left\lfloor \frac{a}{n-a} \right\rfloor$

Proof: We begin by asking each of $A_2, \ldots, A_n$ if A_1 is random until one of the following occur:

1. a persons say that the A_1 is non-random. Then A_1 must be non-random, otherwise we would have $a + 1$ random people. This is not possible if their are at most a random people.

   Suppose that m people said A_1 is random, then these m people are random. But, we must still establish the nature of the remaining $n-m-1$ remaining people. We do so by asking A_1 about these people. This gives us of $a+m$ questions to determine the nature of A_1 and $n-m-1$ questions to determine the rest. Giving a total of $a+m+n-m-1 = n+a-1$ questions to determine the nature of $A_1, \ldots, A_n$. This gives us one possibility.

2. We have asked A_n about A_1. Let x be the number of people who have said that A_1 is non-random. Then $n - x - 1$ people have said that A_1 is random. We have two cases:

(a) $x \le n-a-2$. Then at least $n-(n-a-2)-1 = a+1$ people have said that A_1 is random. At least one of these people is non-random, so A_1 must be random. Also, the x people who said A_1 was non-random are also random. So we can throw out these random people and after asking $n-1$ questions can recurse on the remaining $n-x-1$ people, of which at most $a-x-1$ are random. This gives us $P(a-x-1, n-x-1)$ possibilities and at most $(n-1)+C(a-x-1, n-x-1)$ questions.

(b) $n-a-1 \le x \le a-1$. Here we do not know the nature of A_1, but notice that we do know that either A_1 is non-random OR A_1 is random. If A_1 is non-random, then we do as above (1), giving us one possibility. If A_1 is random, then we do as above (2a), giving us $P(a-x-1, n-x-1)$ possibilities. So, we have $1+P(a-x-1, n-x-1)$ possibilities. However, we have asked $(n-1)+x+C(a-x-1, b-x-1)$ questions.

So, we have, as our final recurences:

- $P(a,n) \le \max$
  - 1
  - $\max_{0\le x\le n-a-2} P(a-x-1, n-x-1)$
  - $1+\max_{n-a-1\le x\le a-1} P(a-x-1, n-x-1)$
- $C(a,n) = \max$
  - $n+a-1$
  - $(n-1)+\max_{0\le x\le n-a-2} C(a-x-1, n-x-1)$
  - $(n-1)+\max_{n-a-1\le x\le a-1}(x+C(a-x-1, n-x-1))$

Now we will show, by induction on $a+n$ that $P(a,n) \le 1+\left\lfloor \frac{a}{n-a} \right\rfloor$. Base cases:

- $a+n=1$. Then $a=0$ and $n=1$. We then have one possibility.
- $a < \frac{n}{2}$. Then we also will have only one possibility.

Assume that $\forall a, n[a+n < k -> P(a,n) \le 1+\left\lfloor \frac{a}{n-a} \right\rfloor$. Now consider $a+n=k$. Let $a \ge \frac{n}{2}$. Now, by inductive hypothesis, $P(a-x-1, n-x-1) < P(a-x, n-x)$. So the recurrence is maximized when x is minimized, yielding

$$P(a,n) = \max(1, P(a-1, n-1), 1+P(2a-n, a)) =$$

$$1+P(2a-n, a) = 1+1+\left\lfloor \frac{2a-n}{a-2a+n} \right\rfloor =$$

$$1+\left\lfloor \frac{2a-n+n-a}{n-a} \right\rfloor = 1+\left\lfloor \frac{a}{n-a} \right\rfloor$$

∎

Lemma 5.2 *For $a < n-1$,*

(i) $C(a-1, n-1) < C(a,n)$

(ii) $x + C(a-x, n-x) \leq C(a,n)$

Proof: For $1 \leq a < \frac{n}{2}$, $C(a-1, n-1) = (n-1) + (a-1) - 1 < n + a - 1 = C(a,n)$. For $a \geq \frac{n}{2}$, (i) holds because the algorithm on parameters (a,n) may recurse on parameters $(a-1, n-1)$. (i) implies (ii) by induction on x. If $x + C(a-x, n-x) \leq C(a,n)$, then $(x+1) + C(a-(x+1), n-(x+1)) \leq x + C(a-x, n-x) \leq C(a,n)$. ∎

Theorem 5.3 *For* $\frac{n}{2} \leq a < n-1, C(a,n) = (n-1) + C(a-1, n-1)$.

Proof:

$$\begin{aligned}
C(a,n) &= \max \left\{ \begin{array}{l} n+a-1, \\ (n-1) + \max_{0 \leq x \leq n-a-2} C(a-x-1, n-x-1), \\ (n-1) + \max_{n-a-1 \leq x \leq a-1} (x + C(a-x-1, n-x-1)) \end{array} \right\} \\
&= \max \left\{ \begin{array}{l} n+a-1, \\ (n-1) + C(a-1, n-1), \\ (n-1) + (n-a-1) + C(a-(n-a-1)-1, n-(n-a-1)-1) \end{array} \right\} \\
&= (n-1) + C(a-1, n-1).
\end{aligned}$$

All steps follow from application of the lemma above. ∎

Corollary 5.4 *For* $\frac{n}{2} \leq a < n-1, C(a,n) = 2a(2n-a-3) - \frac{1}{2}(n-1)(3n-8)$

Proof:

$$\begin{aligned}
C(a,n) &= (n-1) + (n-2) + \ldots + (n-(2a-n+1)) + C(a-(2a-n+1), n-(2a-n+1)) \\
&= \sum_{i=1}^{2a-n+1} (n-i) + (3n-3a-3) \\
&= (2a-n+1)n - \frac{(2a-n+1)(2a-n+2)}{2} + (3n-3a-3) \\
&= 4an - 2a^2 - 6a - \frac{3}{2}n^2 + \frac{11}{2}n - 4.
\end{aligned}$$

∎

6 Lower Bound on $P(a,n)$

Theorem 6.1 $P(a,n) \geq 1 + \left\lfloor \frac{a}{n-a} \right\rfloor$

Proof: Divide the people into $\left\lfloor \frac{n}{n-a} \right\rfloor$ groups of size $n-a$. There may be some people not placed in a group. These people will answer randomly. For the rest of the people, they will answer that anyone in their group is non-random, and that everyone else is random. So, each person will think that his group is non-random and that every other group is random. The questioner will not be able to tell which group is non-random, and since there are $\left\lfloor \frac{n}{n-a} \right\rfloor$ groups, there are $\left\lfloor \frac{n}{n-a} \right\rfloor = \left\lfloor \frac{a}{n-a} + 1 \right\rfloor = 1 + \left\lfloor \frac{a}{n-a} \right\rfloor$ possibilities. ∎