Searching with a Lie Using Only Comparison Questions

Duncan Innes

Department of Computer Science & Statistics
University Of Rhode Island

Abstract

S. M. Ulam presented the following problem. If one person picks a number from one to one million and the other person could ask yes or no questions, how many questions would be required to find the number with certainty if the opponent were allowed to lie once or twice. Joel Spencer found that by using a weight balancing strategy, if $n \leq 5/8 * 2^k/(k+1)$ and there is only one lie, then a number in an interval of length n can be found with at most k questions. This still means either 25 or 26 questions will be needed when n=1,000,000 and only questions of the form "Is x>c" are allowed. However, it will be shown that, by using a somewhat modified algorithm, if $n \leq 27/128 * 2^k/(k+1)$, then the number can be found in k-2 questions, no matter what the number or when the lie was told. This result means that the answer to Ulam's question, when only 1 lie is used and only comparison questions are allowed, is, in fact, twenty-five.

Introduction

S. M. Ulam[5] posed the following problem. One person picks a number between 1 and 1 million. How many questions will it take to guess the number if the responder can lie once or twice? A weight balancing strategy where the remaining possible lies are taken in to account has been investigated by Rivest et al.[3] and Ravikumar and Lakshmanan[2] and they determined bounds for both the continuous and discrete cases. Spencer[4] considers the discrete problem when only comparison questions are allowed and arrives at a tighter bound. He shows that, for a number chosen in the interval {1, 2, 3, ... , n-1,n}, if $n \leq 5/8 * 2^k/(k+1)$ the number can be found in k questions using only questions of the form "Is x>24?" and there is only one lie. Spencer concludes by saying "it seems very difficult to determine whether the answer to Ulam's original problem is 25 or 26." Pelc[1] shows that if arbitrary yes-no questions, like "Is $x \in [5,8] \cup [12,16]$?" the answer is, in fact, 25.

This paper will establish that the answer to Ulam's question is 25, even if restricted to only comparison questions. It will present an algorithm, based on weight balancing, to accomplish this and provides a tighter bound for the general case.

Weight Balancing

The weight balancing algorithm works much like a binary search except numbers are assigned weights and the interval is then split, so as to best balance the weights. The weight of a number is the possible number of ways a lie could be told about it in future queries. When there is only 1 lie allowed, a number which is inconsistent with one of the answers so far would have a weight of 1. With t questions remaining a number, which is consistent with all the answers given, could have a lie told about it on any of the remaining questions or not at all. Consequently it will be assigned a weight of t+1. To determine the value of c to be used in the question "Is x>c?", the weights of the state created by a "no" answer and the weight of the state created by a "yes" answer are calculated. The value that most nearly balances these two weights is then chosen as the split point. This process continues until the number in question has been found or there are no more questions allowed. In the first case the questioner has 'won' the game or 'solved' the problem. In the latter, the responder is considered to have 'won'. At any time, the set of potential numbers (those about which less than 2 lies have been told) fall into three intervals. The main group is the interval containing numbers that are consistent with all the answers so far and have a weight of t+1 each, where t is the number of queries left. The side groups are the intervals on etiher side of the main group, that contain numbers about which one lie may have been told and have a weight of one each.

It should be noted that with comparison questions, a perfect split of the weights is not always possible, particularly when the split point is in the main group or between a side group and the main group. The weight resulting from a "yes" answer and from a "no" answer may differ (as Spencer points out) by as much as t-1, when t questions remain. Using membership questions, it is possible (as Pelc shows) to include enough members of the side group to maintain a more even balance. With only comparison questions, however, it is not so easy and the unevenness of the weights can add up so that in some cases one more question may be needed.

0-8186-2812-X/92 $03.00 © 1992 IEEE

The Modified Algorithm

Employing a slightly different strategy, a tighter bound than that presented by Spencer can be found:

If $n \leq \frac{27}{128}\,\frac{2^k}{k+1}$*, then the number can be found in k-2 questions.*

The modified algorithm requires using weight balancing until there are seven questions left. At this point, a table of positions that yield wins in only 5 more questions is consulted. It will be shown that if n is within the bound above then all the positions that can be reached when there are seven questions left can be solved with only 5 questions. The first phase also employs an "alternating" strategy. According to this strategy, if there are 2 possible values for c that split the weights equally well, then the decision is based on the previous split. If the previous split favored the lower numbers(i.e the weight of the no set was greater than the weight of the yes set) then the next split should favor the higher numbers. Similarly, if the previous split favored the higher numbers the new split point should be chosen to favor the lower numbers. Notice that the prior balancing is only considered when the appropriate split is midway between 2 integers.

First, consider the result of the weight balancing, where w_t is the weight with t questions left(counting backwards), i is the size of the smallest side group, j is the size of the main group and k is the size of the other side group. Weight states are described with the triple (i,j,k). It should be remembered, to avoid confusion, that when t=7 a table of patterns that lead to solutions in only 5 questions is consulted and there are not really 7 questions remaining.

To approach the proof, we will look at the results of the 2 most recent splits.

Lemmas:

1) If the last split(with t+1 questions left) was in a side group,

$$1)\ w_t \leq \frac{w_{t+1}}{2} + \frac{1}{2}$$

Since the weight of each element in the side group is 1, the worst split between groups would have a difference of 1 and differ from a perfect split by 1/2.

2) If the last split was in the main group then

$$2.1)\ w_t \leq \frac{w_{t+1}}{2} + \frac{t}{2} \text{ and}$$

$$2.2)\ (t+2)(i+j)+k \leq w_{t+1}$$

2.1) If you move the split one unit over, there is a gain of t+1, the weight of an element in the main group when t questions remain, and a loss of 1, the weight of an element in the side group. Thus, the worst split possible has a difference of t between the weights and must be within t/2 of a perfect split.

2.2) The old main group would have been i+j(since i≤k, j+k as a main group would be more restrictive) so that you get intervals of (x,i+j,k), where $x \geq w_t-(t+1)i-j-t$, in order to balnce the weights and the main group would have a weight of (t+2)(i+j).

3) If the last 2 splits were in side groups then

$$3)\ w_t \leq \frac{w_{t+2}}{4} + \frac{3}{4}$$

From (1) $w_t \leq \frac{w_{t+1}}{2} + \frac{1}{2}$ and $w_{t+1} \leq \frac{w_{t+2}}{2} + \frac{1}{2}$

4) If the last 2 splits were in main groups then (2.2) must be true and

$$4.1)\ w_t < \frac{w_{t+2}}{4} + \frac{t+1}{4} + \frac{t}{2} \text{ and}$$

$$4.2)\ (t+3)(i+j+k) \leq w_{t+2} \text{ or}$$

$$(t+3)(w_t-ti-t)+k \leq w_{t+2}$$

4.1) From (2) $w_t \leq \frac{w_{t+1}}{2} + \frac{t}{2}$ and $w_{t+1} \leq \frac{w_{t+2}}{2} + \frac{t+1}{2}$

4.2) On the previous split you would have (x, i+j, k), where $x \geq w_t-(t+1)i-j-t$, so that on the second previous split either k or x would be added to the main group. If k is added, then main group would be i+j+k with a weight of (t+3)(i+j+k). If x is added, then the main group would be at least equal to $w_t-(t+1)i-j-t+i+j=w_t-ti-t$, and would have a weight of at least $(t+3)(w_t-ti-t)$with k still in a sidegroup.

5) If the position with t questions remaining resulted from a side group split followed by a main group split then (2.2) must be true and

$$5)\ w_t \leq \frac{w_{t+2}}{4} + \frac{1}{4} + \frac{t}{2}$$

From (2) $w_t \leq \frac{w_{t+1}}{2} + \frac{t}{2}$ and from (1) $w_{t+1} \leq \frac{w_{t+2}}{2} + \frac{1}{2}$

6) If the position with t questions remaining resulted from a main group split followed by a side group split then

$$6)\ w_t \leq \frac{w_{t+2}}{4} + \frac{t+1}{4} + \frac{1}{2}$$

From (1) $w_t \leq \frac{w_{t+1}}{2} + \frac{1}{2}$ and from (2) $w_{t+1} \leq \frac{w_{t+2}}{2} + \frac{t+1}{2}$

7) With t questions remaining the following inequality holds:

$$w_t < w_k 2^{-k+t} + t + 1$$

Proof is by induction. After 1 question, when t=k-1, we get from lemma 2 that

$$w_{k-1} \le \frac{w_k}{2} + \frac{k-1}{2}$$

which satisfies the inequality. Now, assume the statement is true for all questions until there are only t questions left. Then, with t+1 questions remaining,

$$w_{t+1} < w_k 2^{-k+t+1} + t + 2.$$

Hence, from (2),

$$w_t < w_k 2^{-k+t} + \frac{t+2}{2} + \frac{t}{2} = w_k 2^{-k+t} + t + 1,$$

which was to be shown.

Letting $n = \frac{27}{128} * \frac{2^k}{k+1}$. Since $w_k = n(k+1)$, we get,

$w_9 < 118$
$w_8 < 63$
$w_7 < 35$

Depending on the splits on the previous 2 questions we get the following:

mm - 2 main group splits
$w_7 < 35$
sm - 1 side group split followed by a main group split
$w_7 < 33.25$
ms - 1 main group split followed by a side group split
$w_7 < 32$
ss - 2 side group splits
$w_7 < 30.25$

If j =0, the previous split was either a side group split or the favorable result of a split at the boundary of a main group. Therefore, $w_7 < 32$. Since the lie has been exposed, only a binary search is needed and 5 questions is sufficient to find it among 31 elements.

The table below shows all of the positions of weight ≤35 that will be able to be solved in 5 more questions(rather then 7). The positions such that their weight, with 7 questions left, is less than 35, that do not appear in the table, can be broken up into 4 groups:

A = {(i, 0, k): 32<i+k<35}
B ={(12,1,13), (6,1,20), (5,1,21), (12,1,14), (13,1,13), (9,1,17), (7,2,11), (4,3,6), (4,3,5), (12,1,12), (5,1,20), (4,3,5), (4,3,4)}
C = {(2,1,24), (0,3,10), (1,3,9)}
D = {(0,3,9)}

Those in group A, where j=0, must have been the result of a side group split or the favorable result of a main group split and as such would have to have a weight of less than 32,

	j=0	j=1	j=2	j=3	j=4
t=5	all (i,0,k)such that i+k≤32	all (i,1,k)such that i+k≤26 **except** (12,1,12),(5,1,20),(12,1,13), (6,1,20),(5,1,21),(12,1,14), (13,1,13),(9,1,17), (2,1,24)	(6,2,14), (10,2,10) and all (i,2,k)such that i+k≤19 **except** (8,2,11),(7,2,11),(7,2,12), (4,2,15),(1,2,18)	a few where i+k>11 and all (i,3,k)such that i+k≤11 **except** (0,3,9),(0,3,10),(0,3,11) (1,3,10),(1,3,9),(2,3,9) (4,3,7),(4,3,6)(4,3,5) *(4,3,4)	several where i+k>3 and all (i,4,k)such that i+k≤3
t=4	all (i,0,k)such that i+k≤16	all (i,1,k)such that i+k≤11 and i≤4 **except** (2,1,9)	all (i,2,k) such that i=0 and k≤6 all (i,2,k) such that i≤3 and k≤3		
t=3	all (i,0,k)such that i+k≤8	all (i,1,k)such that i+k≤4 **except** (2,1,2)	(0,2,0)		
t=2	(i,0,k) such that i+k≤4	(0,1,1)			
t=1	(1,0,1), (0,0,1), (0,0,2)	(0,1,0)			

Patterns (i,j,k) where i≤k that will always result in a win for the questioner in t questions

eliminating the losers listed above.

Each of those in group B have a weight, w_7=33 or 34 and must have been the result of a main group split. However, by lemma 2.2, this could not happen. (5,1,21), for example must be the split of some (x,6,21), which would have a weight,greater than 63. ≥63.

Each of those in group C have a weight, w_7=34 and herefore must have resulted from 2 main group splits. But by lemma 4.2, that is also impossible. The smallest weight state (2,1,24), for instance, could have come from would be (11,3,24) and its smallest possible predecesor, (x,14,24) would be too large to have occurred.

To show that (0,3,9) cannot be obtained you must go back 2 questions and reevaluate the position. (0,3,9) must have come from (23,3,9) which must have come from (23,3,64). This can be shown to be impossible in the same manner as above, by starting from t=9.

Notice that (0,3,9) could not be the result of 2 splits of (78,3,9), because of the "alternating" strategy. (78,3,9) would be split so that a "yes" answer results in (23,3,9) and a weight of 59, while a "no" answer results in (55,0,3) of weight 58. Consequently, the split of (23,3,9) would then result in (23,1,2) or (1,2,9) and not (0,3,9) and (23,0,3) so that the "no" answer now results in a larger weight state.

Conclusion

Thus, with $n \le \frac{27}{128} * \frac{2^k}{k+1}$, when k-7 questions have been asked and answered, the only possible patterns that remain allow the solution to be found in 5 more questions.

Setting k=27, we know that, if n≤1,011,126, twenty-five questions are sufficient. Therefore, the answer to Ulam's problem, when only 1 lie is permitted and comparison questions must be used, is twenty-five.

Acknowledgement

I wish to express considerable thanks to Bala Ravikumar for his enthusiasm, support, direction and encouragement.

References

[1] A. Pelc, "Solution of Ulam's problem on searching with a lie," *Journal of Combinatorial Theory* Series A

[2] B. Ravikumar and K.B. Lakshaman, "Coping with known patterns of lies in a search game," *Theoretical Computer Science* 33(1984)

[3] R.L. Rivest et al., "Coping with errors in binary search procedures," *Journal of Comput. System Science* 20 (1980)

[4] J. Spencer, "Guess a number - with Lying," *Mathematics Magazine* 57 No. 2 (1984)

[5] S.M. Ulam, *Adventures of a Mathematician*, Scribner, New York, 1976