Clueless Voting

The minutes of a talk given to the Trinity Mathematical Society by
Professor Imre Leader on 10 October, 2005.

LaTeXed by Paul Smith, T.C.
Please send comments and corrections to pjs59@cam.

We will be looking at situations in which there is a group of people who are going to vote on an issue but none has any idea whom or what to vote for.

We will be focusing on the following example. There are 1000 people in a room, each of whom has a piece of paper stuck to her forehead, with either a 0 or a 1 written on it. The 0s and 1s occur independently and with equal probability. Everyone has to vote on the parity of the sum of all the 0s and 1s in the room.

Everyone has a 50-50 chance of guessing correctly — they all can see what everyone else's number is, but they have no idea what their own is.

It seems obvious that however everyone chooses to vote, there is no way one can do better than have a 50% chance of getting the correct result.

Digression 1: Simpson's Paradox

Consider the following contingency table of wins for the Pittsburgh Pirates baseball team.

	Pittsburgh	Atlanta
Rain	30%	60%
No rain	20%	50%

The Pirates win more often in Pittsburgh when it rains, and they win more often in Atlanta when it rains. So, if someone said just before a match, "it's going to rain," would we be happy?

The answer is "not necessarily". Let's look at the table again, this time with the numbers of matches added in.

	Pittsburgh	Atlanta
Rain	30% **1000000**	60% **100**
No rain	20% **100**	50% **1000000**

So the situation is this:

Raining: we're almost certainly in Pittsburgh, so the chance of winning is about 30% (in fact, it's something like $(30 + \epsilon)\%$, where ϵ is small).

Not raining: we're almost certainly in Atlanta, so the chance of winning is only very slightly less than 50%.

That means that overall we're much better off if it's dry. That seems odd, but it's because the variables Pittsburgh/Atlanta and Rain/No rain are not independent.

Clueless Voting

Lesson from Digression 1: independence is very counter-intuitive. Although in our election no single person's vote matters, we cannot conclude that overall there is no way of bettering a 50% chance of getting the result right.

So, can we in fact do better than 50%?

Consider the system where every person votes according to the following rule:

> **"Pretend I have the same number (0 or 1) as the majority of the others, and vote accordingly."**

What happens, for example, when there are 530 1s and 470 0s?

People with 1s	**People with 0s**
See more 1s than 0s, so vote as if they have a 1	See more 1s than 0s, so vote as if they have a 1
$\Rightarrow$ All correct	$\Rightarrow$ All wrong

Thus, all 530 people with 1s vote correctly, and all 470 people with 0s vote wrongly. In this example, they get it right.

When does this method fail? The only way it can possibly fail is if there are exactly the same number of 0s and 1s. The chance of this happening is

$$\binom{1000}{500}\frac{1}{2^{1000}} \sim \frac{1}{\sqrt{500\pi}} \approx \frac{1}{40}.$$

(In general, if there are $2n$ people, the chance is

$$\binom{2n}{n}\frac{1}{2^{2n}} \sim \frac{1}{\sqrt{n\pi}}.)$$

That's really rather good. Under this system, the voters only get it wrong if the 0s and 1s are split evenly among the voters, and this only happens about $1/\sqrt{n\pi}$ of the time. But it's still reasonable to ask: can we do better?

The answer is "no" if we're restricted to *symmetric* voting systems — that is, systems where everybody follows the same rule to determine how they vote.

Asymmetric Systems

Consider the following *asymmetric* system. For simplicity we'll assume there are $2^n - 1$ players, in this case 1023, but it works equally well with any number.

First split the players up into blocks: the first block contains just player 1, the second block contains players 2 and 3, the third block contains players 4 to 7, etc. up to the tenth block which contains players 512 to 1023. Now use the following rules:

1	2 3	4 5 6 7	......
□	□ □	□ □ □ □	
"Pretend I have a 0, and vote accordingly."	**"If player 1 votes correctly, cancel out. Otherwise, pretend our block total is even, and vote accordingly."**	**"If either of the first two block totals is even, cancel out. Otherwise, pretend our block total is even and vote accordingly."**	

This continues, so the last block will vote to cancel each other out if any of the previous nine blocks have an even block total, and if not they pretend their block total is even and vote accordingly.

Players 2 and 3 know whether player 1 has voted correctly, because they can see whether or not she has a 0 on her forehead (however, they *don't* know what the correct answer is) — and similarly for the larger blocks.

What do we mean by "cancel out"? Well, when we know that the players who have already voted have done so correctly we want to make sure we don't ruin that. So in this case half of the block pretends the block total is even and votes accordingly, and the other half pretends the block total is odd and votes accordingly. This way the particular block in question has no effect on the outcome of the election.

When does this system fail? If player 1 has a 0 on her forehead, we win — all other players cancel each other out. If the total of the second block is even we also win, because their votes cancel out player 1's, and all the other players cancel each other out. In fact, the only way we can lose is if all ten blocks have an odd total, in which case every single player votes incorrectly.

The chance of failure is therefore

$$\underbrace{\frac{1}{2}\cdot\frac{1}{2}\cdot\dots\cdot\frac{1}{2}}_{\times 10} = \frac{1}{1024} \sim \frac{1}{n}.$$

By allowing asymmetric voting, we've improved the chance of failure to just $1/n$. But can we do even better? Once again, the answer is "no" if we are using standard voting systems. However, the answer changes if we allow ourselves the luxury of *weighted* voting.

Weighted Voting

Under a weighted voting system, each player is allowed to choose a weight for his vote (and it need not be predetermined, she can choose when she votes).

The system that we're going to look at is very similar to the above asymmetric system, except now all our blocks are of size 1. Each block then uses *exactly* the same system as the blocks above, but now they have the choice of how to weight their vote, and that is determined according to the scheme below.

1	2	3	
☐	☐	☐	……
"Pretend I have a 0, and vote with weight 1."	**"If player 1 votes correctly, use weight 0.00001. Otherwise, pretend I have a 0, and use weight 2."**	**"If either of the first two players has a 0, use weight 0.00001. Otherwise, vote with weight 4."**	

If player k is the first player to have a 0 on her forehead, all subsequent players will vote with weight 0.00001 (which is an arbitrary, silly small number), and their vote will have no effect on the outcome. Player k will have voted with weight 2^{k-1}, which will beat all the previous players' votes. The only way we can lose is if every single player has a 1, and the chance of this happening is just $1/2^n$.

Again, this is the best possible result.

Digression 2: The chance of always being ahead

A natural question to ask following on from symmetric voting is the following: if n is odd, and there is exactly one more 1 than 0, what is the chance the 1s are always ahead in the voting?

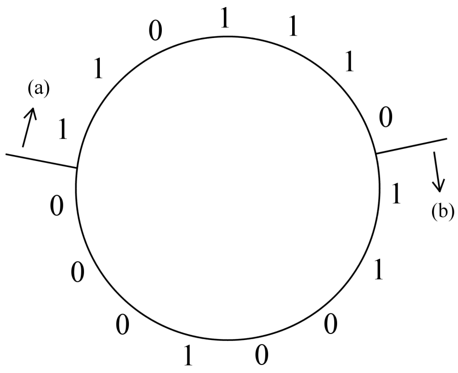
To simplify matters, let's look at $n = 7$. Suppose the voting is 1001101. Consider the cyclical permutations of that, and look for which of those the 1s are always ahead.

1001101	✗
1100110	✗
0110011	✗
1011001	✗
1101100	✓
0110110	✗
0011011	✗

So the chance here is 1/7. If we could prove that this held in general — that among all cyclical permutations of a binary string containing one more 1 than 0, we are always ahead in exactly one of them — then we'd be done, because all binary strings of this form can be split into distinct sets of size n with elements equal up to cyclical permutations.

The proof of the claim is by "moonwalking" (after a Trinity interview question set by Béla Bollobás in the mid 1970s).

Arrange the string in a circle, as in the diagram. First we prove there is at most one point on the circle such that starting at that point and moving round the cricle, the 1s are always ahead in the voting. Suppose there are two such points, (a) and (b). Then between (a) and (b) (clockwise) the 1s are always ahead, and between (b) and (a) (clockwise) the 1s are always ahead, so in each sector there is at least one more 1 than 0. But there are exactly $(n+1)$ 1s and n 0s, a contradiction.



Now we prove there is such a point. Start anywhere on the circle and move round until we've passed an equal number of 0s and 1s, or, if that doesn't occur, until we return to the starting point. In the latter case the 1s must have always been ahead, so we are done. In the former case, keep repeating this until either we hit a point we've already visited or the 1s are ahead for an entire loop. This time, in the latter case we're done as before, and in the former case we've made an integer number of revolutions between the two visits to that point, so the total number of 1s passed must be more the number of 0s. But each time we stopped we had passed an equal number of 0s and 1s, a contradiction, proving the claim.

Final Thoughts

What we have done up to now is this: we have n people, and, for each of the subsets of size $n-1$ of them (of which there are of course n) we have a voter who sees that subset and can vote based on that.

What happens if we see less? Suppose we still have n people, but this time we have $\binom{n}{3}$ voters, each of whom can see exactly 3 people. How well can this collection of $\binom{n}{3}$ voters do? More generally, say we have $\binom{n}{k}$ voters, each of whom sees a k-set of the people. How large should k be, in terms of n, before there is a voting strategy that is right nearly all the time (i.e. with probability of success tending to 1 as n gets large). (We know that $k = n-1$ will do. How small can k be?)