LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

MAY 5, 2025

NETWORKED MINDS:
OPINION DYNAMICS AND COLLECTIVE INTELLIGENCE IN SOCIAL NETWORKS

THE CONDORCET JURY THEOREM

Adrian Haret
a.haret@lmu.de

Let’s start with a game!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

Privatbäckerei
WIMMER

Privat Bäckerei
wimmer

GUESS-THE-LOGO GAME

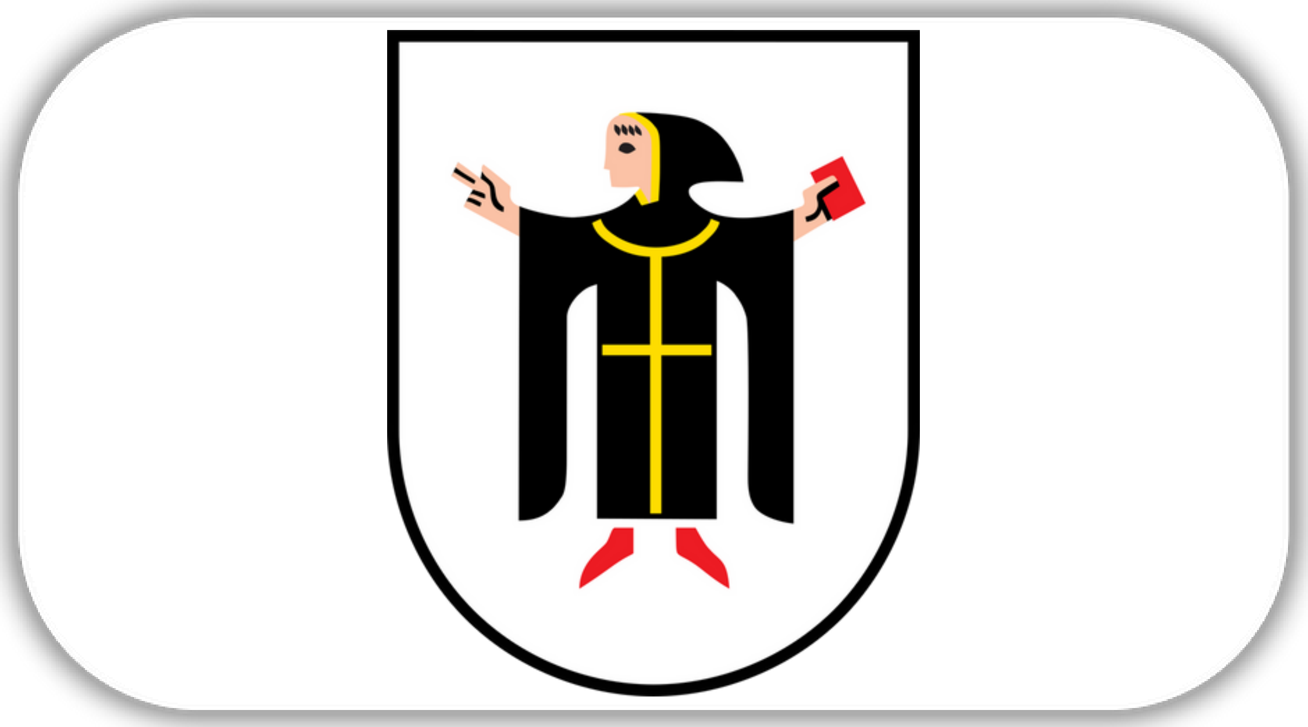
Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

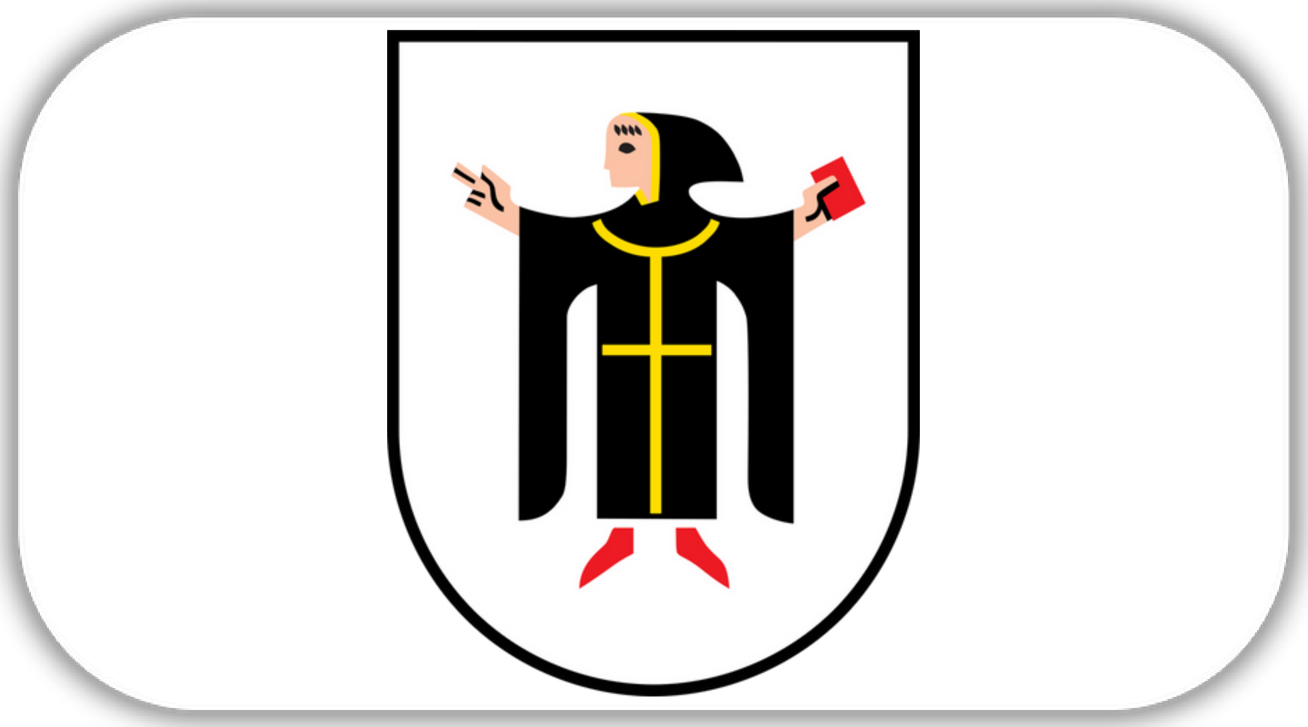
Keep track of your score!



GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!



GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

Keep track of your score!

GUESS-THE-LOGO GAME

Guess the correct version of the logo.

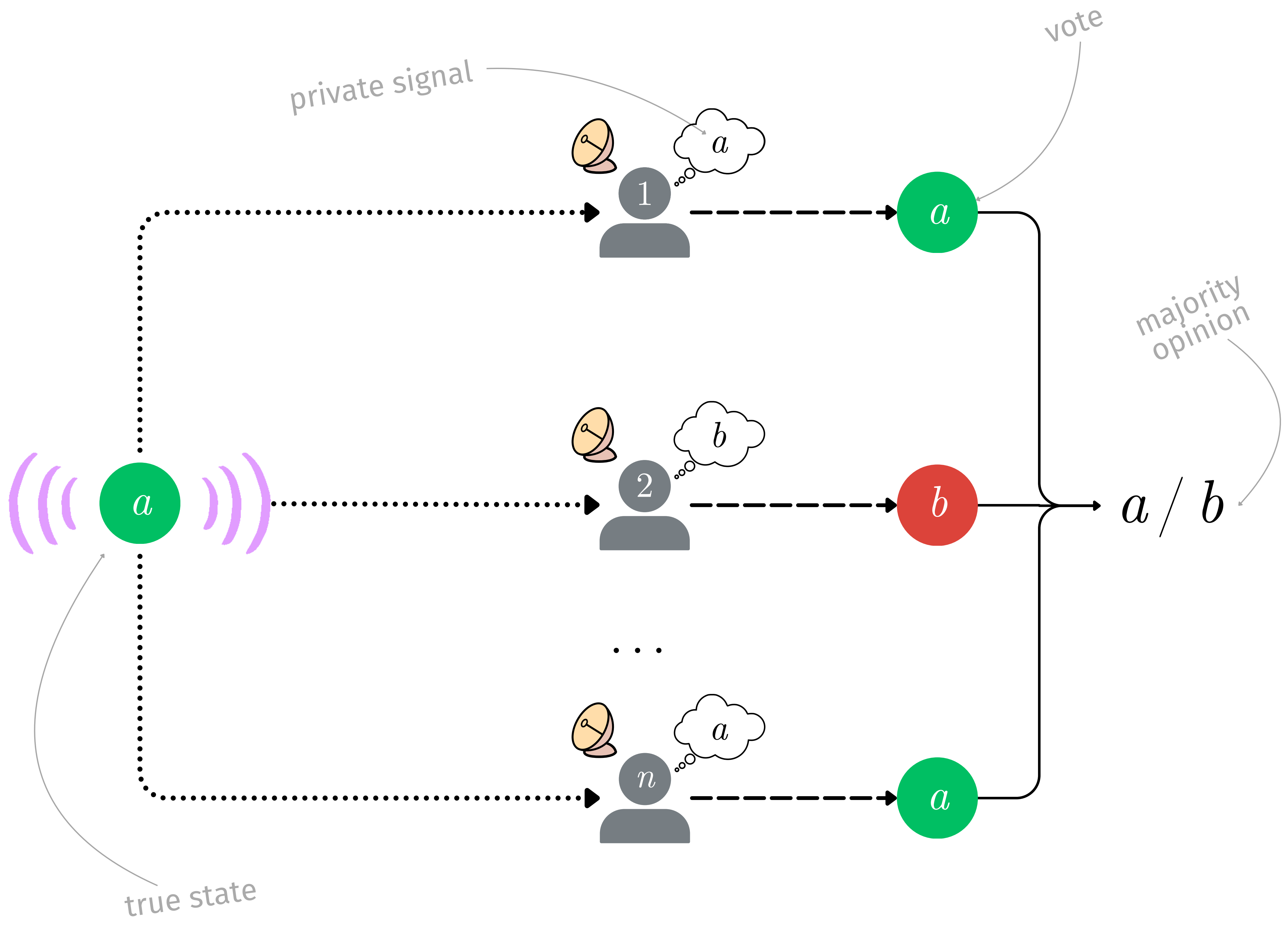
Keep track of your score!

How did you do? And how did the group do?

Here’s the model we’re working with.

AGENTS AS NOISY ESTIMATORS OF THE TRUTH

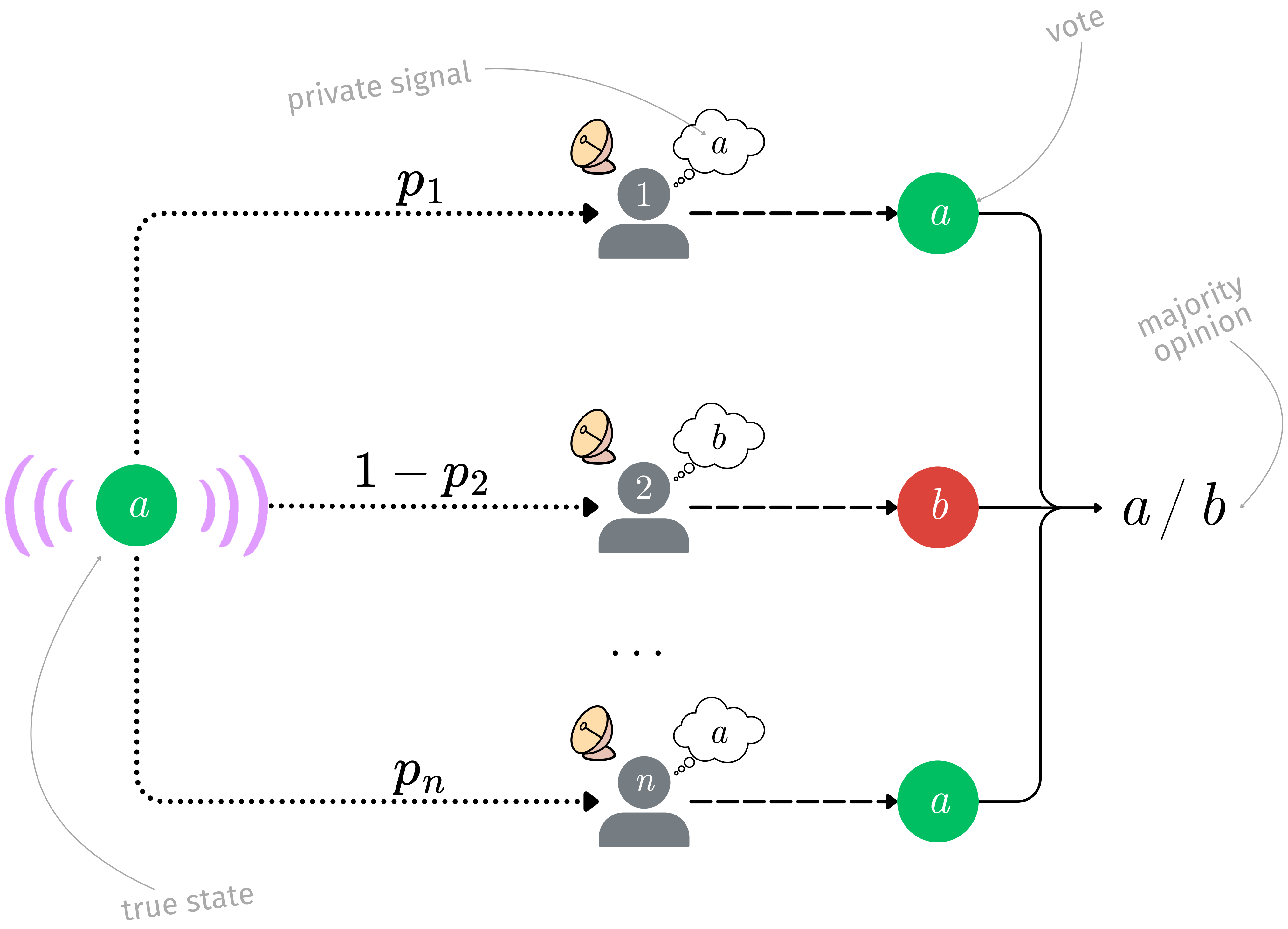
A number of *agents* vote on two *alternatives*, one of which is *correct*.



AGENTS AS NOISY ESTIMATORS OF THE TRUTH

A number of *agents* vote on two *alternatives*, one of which is *correct*.

Each agent has a specific *competence*, i.e., the probability of voting for the correct alternative.



It’s possible that everyone ends up voting for the wrong thing, e.g., if they get the wrong signal.

It’s possible that everyone ends up voting for the wrong thing, e.g., if they get the wrong signal. But how likely is this?…

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*.

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*. There are two *alternatives*, a and b, one of which is *correct*.

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*. There are two *alternatives*, a and b, one of which is *correct*. Each voter casts a *vote* for one of the alternatives.

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*. There are two *alternatives*, a and b, one of which is *correct*. Each voter casts a *vote* for one of the alternatives. We keep track of whether each voter i is correct using a *random variable* v_i:

$$v_i = \begin{cases} 1, & \text{if voter } i \text{ votes for the correct alternative,} \\ 0, & \text{otherwise.} \end{cases}$$

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*. There are two *alternatives*, a and b, one of which is *correct*. Each voter casts a *vote* for one of the alternatives. We keep track of whether each voter i is correct using a *random variable* v_i:

$$v_i = \begin{cases} 1, & \text{if voter } i \text{ votes for the correct alternative,} \\ 0, & \text{otherwise.} \end{cases}$$

The *profile* of votes is a vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ of the votes cast. The *majority outcome** is the alternative with the most votes.

*We assume n is odd to avoid ties.

NOTATION

There is a set $N = \{1, \ldots, n\}$ of *voters*. There are two *alternatives*, a and b, one of which is *correct*. Each voter casts a *vote* for one of the alternatives. We keep track of whether each voter i is correct using a *random variable* v_i:

$$v_i = \begin{cases} 1, & \text{if voter } i \text{ votes for the correct alternative,} \\ 0, & \text{otherwise.} \end{cases}$$

The *profile* of votes is a vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ of the votes cast. The *majority outcome** is the alternative with the most votes.

Each voter i has a *competence* p_i, which is their probability of voting correctly:

$$v_i = \begin{cases} 1, & \text{with probability } p_i, \\ 0, & \text{with probability } 1 - p_i. \end{cases}$$

*We assume n is odd to avoid ties.

JACOB BERNOULLI

May I humbly point out that the vote random variables v_i are called *Bernoulli variables*: $v_i = 1$ is *success*, $v_i = 0$ is *failure*.

JACOB BERNOULLI

May I humbly point out that the vote random variables v_i are called *Bernoulli variables*: $v_i = 1$ is *success*, $v_i = 0$ is *failure*.

The sum of the votes is also a random variable:

$$S_n = v_1 + \cdots + v_n.$$

S_n tracks the number of correct votes in a profile of n votes.

JACOB BERNOULLI

May I humbly point out that the vote random variables v_i are called *Bernoulli variables*: $v_i = 1$ is *success*, $v_i = 0$ is *failure*.

The sum of the votes is also a random variable:

$$S_n = v_1 + \cdots + v_n.$$

S_n tracks the number of correct votes in a profile of n votes.

Note that the majority outcome is correct exactly when $S_n > \lfloor n/2 \rfloor$.

THE MARQUIS DE CONDORCET

I want to make some assumptions!

ASSUMPTIONS

(Competence) Agents are better than random at being correct:

$$p_i > \frac{1}{2}, \text{ for any voter } i \in N.$$

ASSUMPTIONS

(Competence) Agents are better than random at being correct:

$$p_i > \frac{1}{2}, \text{ for any voter } i \in N.$$

(Equal Competence) All agents have the same competence:

$$p_i = p_j = p, \text{ for all voters } i, j \in N.$$

ASSUMPTIONS

(Competence) Agents are better than random at being correct:

$$p_i > \frac{1}{2}, \text{ for any voter } i \in N.$$

(Equal Competence) All agents have the same competence:

$$p_i = p_j = p, \text{ for all voters } i, j \in N.$$

(Independence) Voters vote independently of each other:

$$\Pr[v_i = x, v_j = y] = \Pr[v_i = x] \cdot \Pr[v_j = y], \text{ for all voters } i, j \in N.$$

THE MARQUIS DE CONDORCET

I claim that under these conditions, the majority tends to get it right!

What Condorcet means is that the majority vote

THE MARQUIS DE CONDORCET

Mark my words:

S_n

What Condorcet means is that the majority vote is correct

THE MARQUIS DE CONDORCET

Mark my words:

$$S_n > \lfloor n/2 \rfloor$$

What Condorcet means is that the majority vote is correct with high probability.

THE MARQUIS DE CONDORCET

Mark my words:

$$\Pr\left[S_n > \lfloor {}^{n}/_{2} \rfloor\right]$$

to the moon! ↗↗↗

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters with equal competence $p > 1/2$ that vote independently of each other, then, for any odd n, it holds that:

(1) Larger groups are more accurate than smaller groups.

THE CONDORCET JURY THEOREM (CJT)

THEOREM
For n voters with equal competence $p > 1/2$ that vote independently of each other, then, for any odd n, it holds that:

(1) $\Pr\left[S_{n+2} > \lfloor (n+2)/2 \rfloor\right] > \Pr\left[S_n > \lfloor n/2 \rfloor\right]$, and

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters with equal competence $p > 1/2$ that vote independently of each other, then, for any odd n, it holds that:

(1) $\Pr\left[S_{n+2} > \lfloor (n+2)/2 \rfloor\right] > \Pr\left[S_n > \lfloor n/2 \rfloor\right]$, and

(2) Groups are more accurate than their members.

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters with equal competence $p > 1/2$ that vote independently of each other, then, for any odd n, it holds that:

(1) $\Pr\left[S_{n+2} > \lfloor (n+2)/2 \rfloor\right] > \Pr\left[S_n > \lfloor n/2 \rfloor\right]$, and

(2) $\Pr\left[S_n > \lfloor n/2 \rfloor\right] \geq p$, and

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters with equal competence $p > 1/2$ that vote independently of each other, then, for any odd n, it holds that:

(1) $\Pr\Big[S_{n+2} > \lfloor (n+2)/2 \rfloor\Big] > \Pr\Big[S_n > \lfloor n/2 \rfloor\Big]$, and

(2) $\Pr\Big[S_n > \lfloor n/2 \rfloor\Big] \geq p$, and

(3) The probability of a correct decision approaches 1 as the group size increases.

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters with equal competence $p > {}^1\!/_2$ that vote independently of each other, then, for any odd n, it holds that:

(1) $\Pr\Big[S_{n+2} > \lfloor {}^{(n+2)}\!/_2 \rfloor\Big] > \Pr\Big[S_n > \lfloor {}^n\!/_2 \rfloor\Big]$, and

(2) $\Pr\Big[S_n > \lfloor {}^n\!/_2 \rfloor\Big] \geq p$, and

(3) $\lim_{n\to\infty} \Pr\Big[S_n > \lfloor {}^n\!/_2 \rfloor\Big] = 1.$

To prove this, we have to see how group accuracy depends on the accuracy of the members.

To prove this, we have to see how group accuracy depends on the accuracy of the members. Let's start simple.

ONE VOTER

The profile is $v = (v_1)$.

The probability of a correct decision is:

ONE VOTER

The profile is $\boldsymbol{v} = (v_1)$.

The probability of a correct decision is:

$$\begin{aligned} \Pr\Big[S_1 > 0\Big] &= \Pr\Big[v_1 = 1\Big] \\ &= p \\ &> {}^1\!/_2. \end{aligned}$$

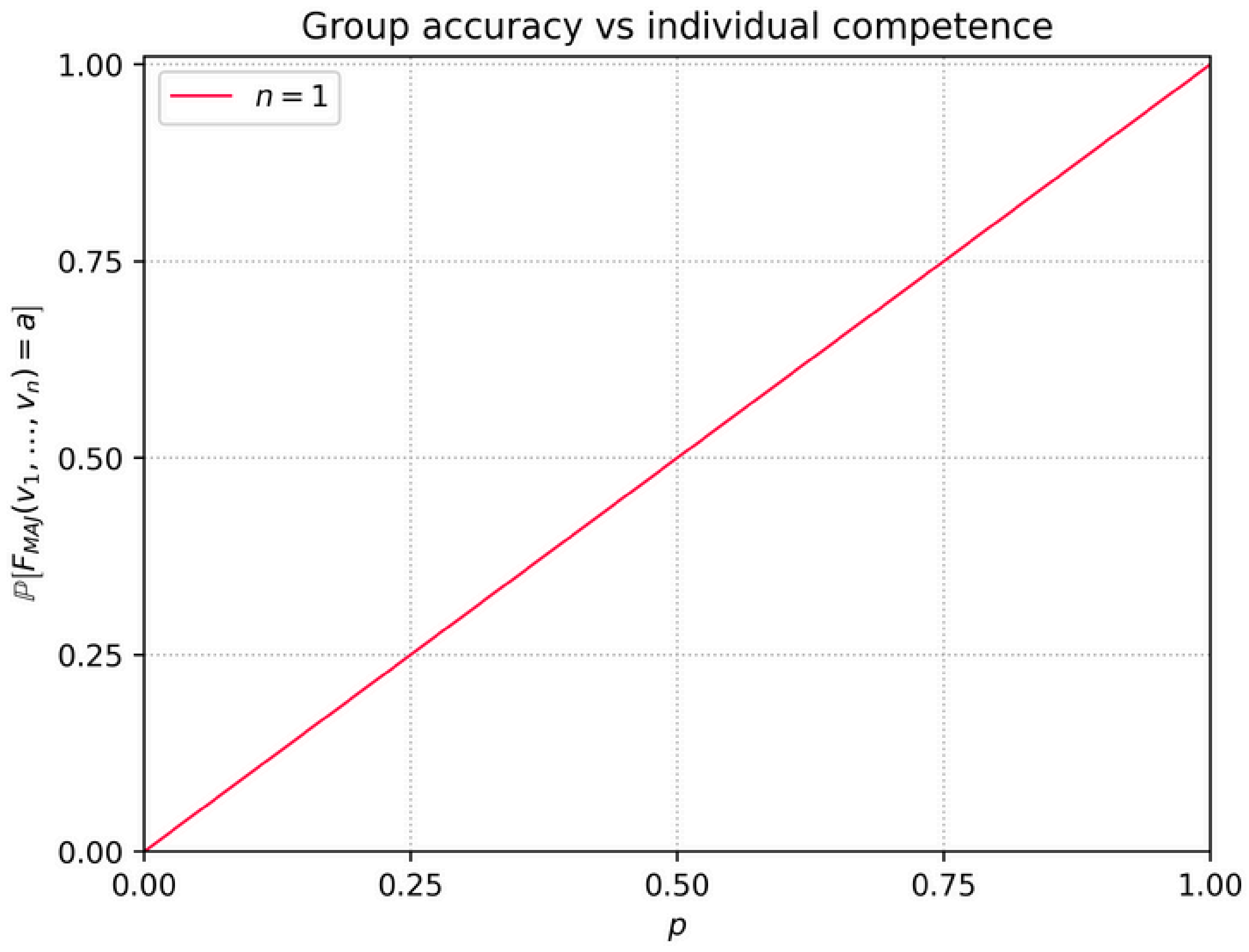
ONE VOTER

The profile is $v = (v_1)$.

The probability of a correct decision is:

$$\begin{aligned}\Pr\left[S_1 > 0\right] &= \Pr\left[v_1 = 1\right] \\ &= p \\ &> 1/2.\end{aligned}$$

As p grows, so does group accuracy.



TWO VOTERS

The profile is $v = (v_1, v_2)$.

TWO VOTERS

The profile is $v = (v_1, v_2)$.

Oh wait, we're not looking at this case.

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$\Pr\left[S_3 > 1\right] = \Pr\left[S_3 = 2 \text{ or } S_3 = 3\right]$$

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$\begin{aligned}\Pr\Big[S_3 > 1\Big] &= \Pr\Big[S_3 = 2 \text{ or } S_3 = 3\Big] \\ &= \Pr\Big[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\Big]\end{aligned}$$

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$
\begin{aligned}
\Pr\left[S_3 > 1\right] &= \Pr\left[S_3 = 2 \text{ or } S_3 = 3\right] \\
&= \Pr\left[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\right] \\
&= \Pr\left[\boldsymbol{v} = (1,1,0)\right] + \Pr\left[\boldsymbol{v} = (1,0,1)\right] + \Pr\left[\boldsymbol{v} = (0,1,1)\right] + \Pr\left[\boldsymbol{v} = (1,1,1)\right]
\end{aligned}
$$

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$
\begin{aligned}
\Pr\Big[S_3 > 1\Big] &= \Pr\Big[S_3 = 2 \text{ or } S_3 = 3\Big] \\
&= \Pr\Big[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\Big] \\
&= \Pr\Big[\boldsymbol{v} = (1,1,0)\Big] + \Pr\Big[\boldsymbol{v} = (1,0,1)\Big] + \Pr\Big[\boldsymbol{v} = (0,1,1)\Big] + \Pr\Big[\boldsymbol{v} = (1,1,1)\Big] \\
&= \Pr[v_1 = 1] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 0] + \\
&\quad\ \Pr[v_1 = 1] \cdot \Pr[v_2 = 0] \cdot \Pr[v_3 = 1] + \\
&\quad\quad \Pr[v_1 = 0] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 1] + \\
&\quad\quad\ \Pr[v_1 = 1] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 1]
\end{aligned}
$$

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

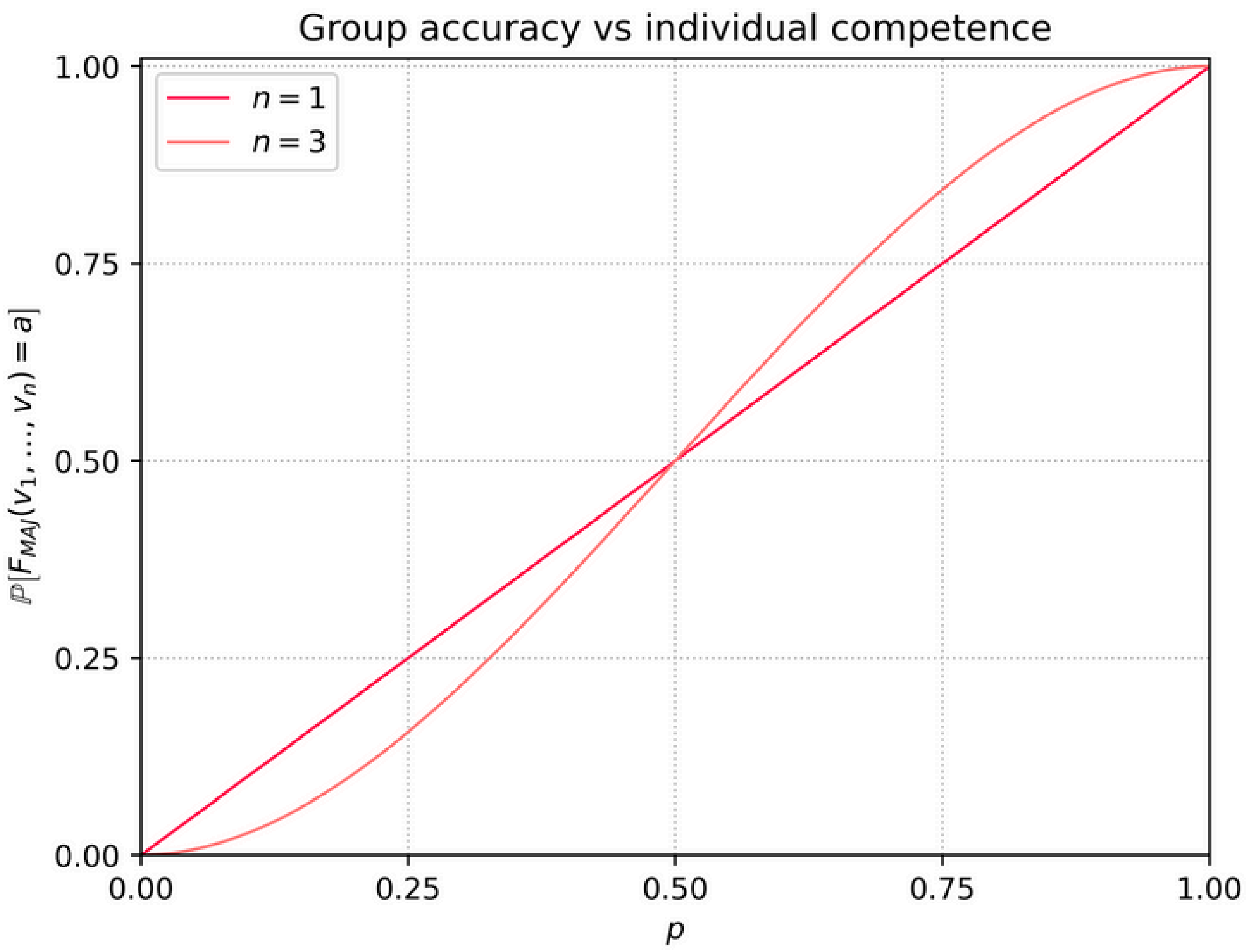
$$
\begin{aligned}
\Pr\Big[S_3 > 1\Big] &= \Pr\Big[S_3 = 2 \text{ or } S_3 = 3\Big] \\
&= \Pr\Big[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\Big] \\
&= \Pr\Big[\boldsymbol{v} = (1,1,0)\Big] + \Pr\Big[\boldsymbol{v} = (1,0,1)\Big] + \Pr\Big[\boldsymbol{v} = (0,1,1)\Big] + \Pr\Big[\boldsymbol{v} = (1,1,1)\Big] \\
&= \Pr[v_1 = 1] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 0] + \\
&\quad\ \Pr[v_1 = 1] \cdot \Pr[v_2 = 0] \cdot \Pr[v_3 = 1] + \\
&\qquad \Pr[v_1 = 0] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 1] + \\
&\qquad\ \Pr[v_1 = 1] \cdot \Pr[v_2 = 1] \cdot \Pr[v_3 = 1] \\
&= p \cdot p \cdot (1-p) + p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p + p \cdot p \cdot p \\
&= 3p^2(1-p) + p^3 \\
&> p.
\end{aligned}
$$

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$\begin{aligned}
\Pr\Big[S_3 > 1\Big] &= \Pr\Big[S_3 = 2 \text{ or } S_3 = 3\Big] \\
&= \Pr\Big[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\Big] \\
&= \Pr\Big[\boldsymbol{v} = (1,1,0)\Big] + \Pr\Big[\boldsymbol{v} = (1,0,1)\Big] + \Pr\Big[\boldsymbol{v} = (0,1,1)\Big] + \mathrm{P} \\
&= \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 0\big] + \\
&\qquad \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 0\big] \cdot \Pr\big[v_3 = 1\big] + \\
&\qquad\quad \Pr\big[v_1 = 0\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 1\big] + \\
&\qquad\qquad \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 1\big] \\
&= p \cdot p \cdot (1-p) + p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p + p \cdot p \cdot p \\
&= 3p^2(1-p) + p^3 \\
&> p.
\end{aligned}$$



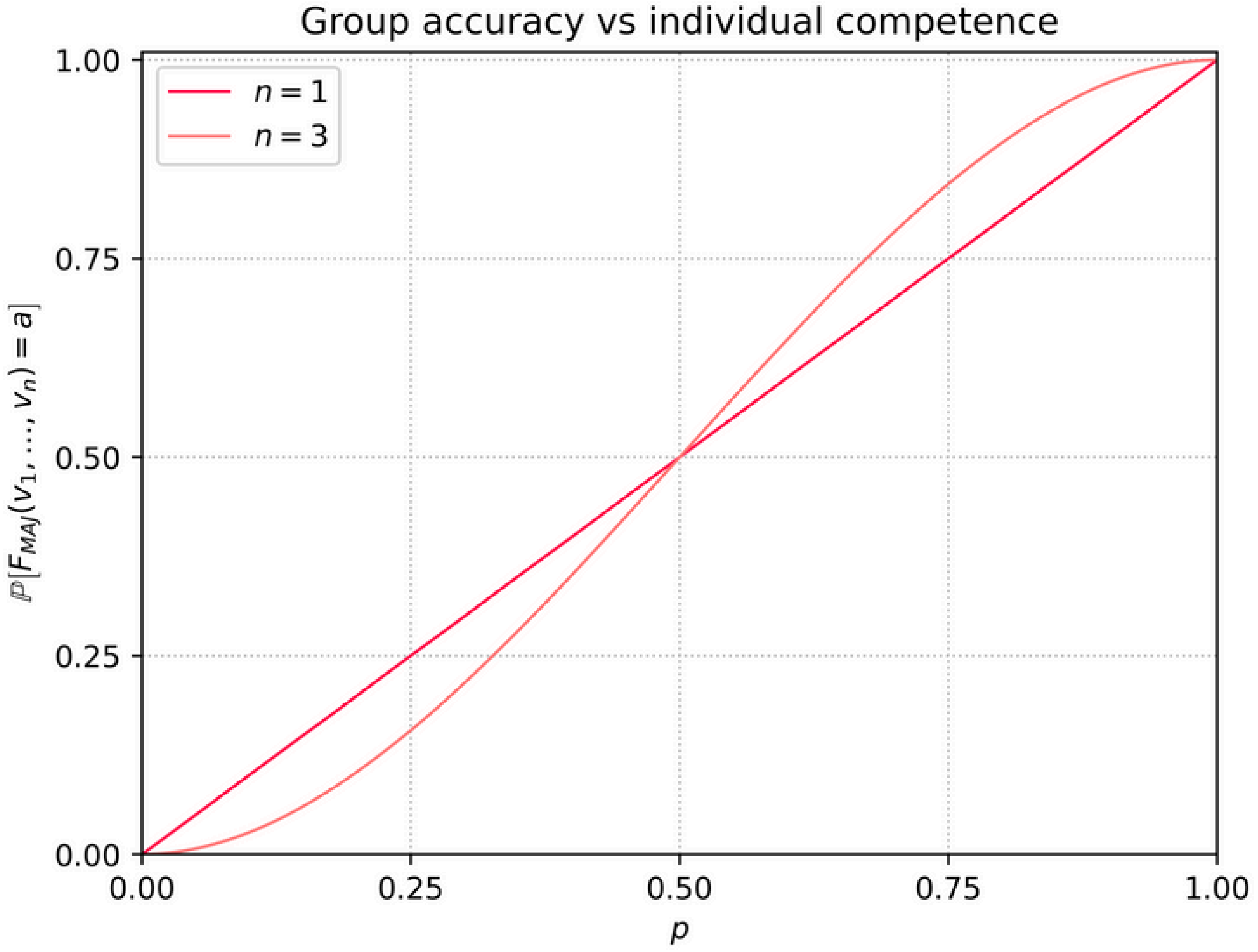
Again, as p grows, so does group accuracy.

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

The probability of a correct decision is:

$$
\begin{aligned}
\Pr\Big[S_3 > 1\Big] &= \Pr\Big[S_3 = 2 \text{ or } S_3 = 3\Big] \\
&= \Pr\Big[\boldsymbol{v} \text{ is one of } (1,1,0), (1,0,1), (0,1,1) \text{ or } (1,1,1)\Big] \\
&= \Pr\Big[\boldsymbol{v} = (1,1,0)\Big] + \Pr\Big[\boldsymbol{v} = (1,0,1)\Big] + \Pr\Big[\boldsymbol{v} = (0,1,1)\Big] + \mathrm{I} \\
&= \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 0\big] + \\
&\quad\ \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 0\big] \cdot \Pr\big[v_3 = 1\big] + \\
&\quad\ \Pr\big[v_1 = 0\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 1\big] + \\
&\quad\ \Pr\big[v_1 = 1\big] \cdot \Pr\big[v_2 = 1\big] \cdot \Pr\big[v_3 = 1\big] \\
&= p \cdot p \cdot (1-p) + p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p + p \cdot p \cdot p \\
&= 3p^2(1-p) + p^3 \\
&> p.
\end{aligned}
$$



Again, as p grows, so does group accuracy.

And a group of three voters is more accurate than a single voter!

FIVE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5)$.

The probability of a correct decision is:

$$\begin{aligned}
\Pr\Big[S_5 > 2\Big] &= \Pr\Big[S_5 = 3 \text{ or } S_5 = 4 \text{ or } S_5 = 5\Big] \\
&= \Pr\big[\boldsymbol{v} \text{ is either } (1,1,1,0,0),\ldots,(1,1,1,1,0),\ldots, \text{ or } (1,1,1,1,1)\big] \\
&\ldots \\
&= 10 \cdot p^3(1-p)^2 + 5 \cdot p^4(1-p) + p^5
\end{aligned}$$

FIVE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5)$.

The probability of a correct decision is:

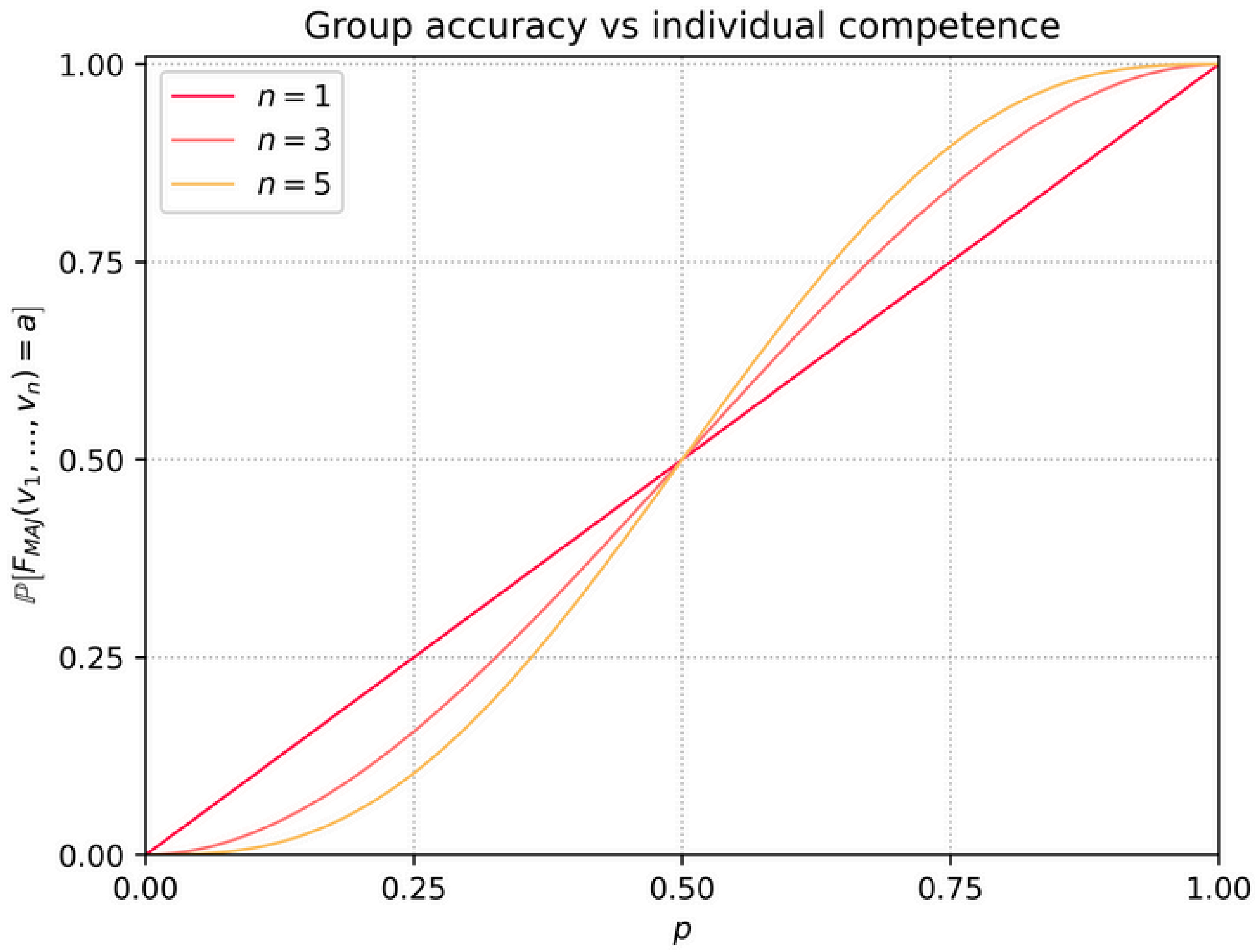
$$
\begin{aligned}
\Pr\Big[S_5 > 2\Big] &= \Pr\Big[S_5 = 3 \text{ or } S_5 = 4 \text{ or } S_5 = 5\Big] \\
&= \Pr\big[\boldsymbol{v} \text{ is either } (1,1,1,0,0), \ldots, (1,1,1,1,0), \ldots, \text{ or } (1,1,1,1,1)\big] \\
&\ldots \\
&= 10 \cdot p^3(1-p)^2 + 5 \cdot p^4(1-p) + p^5 \\
&= \binom{5}{3} p^3(1-p)^2 + \binom{5}{4} p^4(1-p) + \binom{5}{5} p^5
\end{aligned}
$$

FIVE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5)$.

The probability of a correct decision is:

$$\begin{aligned}
\Pr\Big[S_5 > 2\Big] &= \Pr\Big[S_5 = 3 \text{ or } S_5 = 4 \text{ or } S_5 = 5\Big] \\
&= \Pr\Big[\boldsymbol{v} \text{ is either } (1,1,1,0,0), \ldots, (1,1,1,1,0), \ldots, \text{ or } (1 \\
&\ldots \\
&= 10 \cdot p^3(1-p)^2 + 5 \cdot p^4(1-p) + p^5 \\
&= \binom{5}{3} p^3(1-p)^2 + \binom{5}{4} p^4(1-p) + \binom{5}{5} p^5
\end{aligned}$$



Again, as p grows, so does group accuracy.

And a group of five voters is more accurate than a group of three!

ANY ODD NUMBER OF VOTERS

The profile is $\boldsymbol{v} = (v_1, \ldots, v_n)$, for $n = 2k + 1$ and $k \geq 1$.

The probability of a correct decision is:

$$
\begin{aligned}
\Pr\Big[S_n > k\Big] &= \Pr\Big[S_n = k{+}1 \text{ or } \ldots \text{ or } S_n = n\Big] \\
&= \binom{n}{k+1} \cdot p^{k+1}(1-p)^{n-(k+1)} + \cdots + \binom{n}{n-1} \cdot p^{n-1}(1-p)^1 + \binom{n}{n} p^n \\
&= \sum_{i=k+1}^{n} \binom{n}{i} \cdot p^i (1-p)^{n-i}.
\end{aligned}
$$

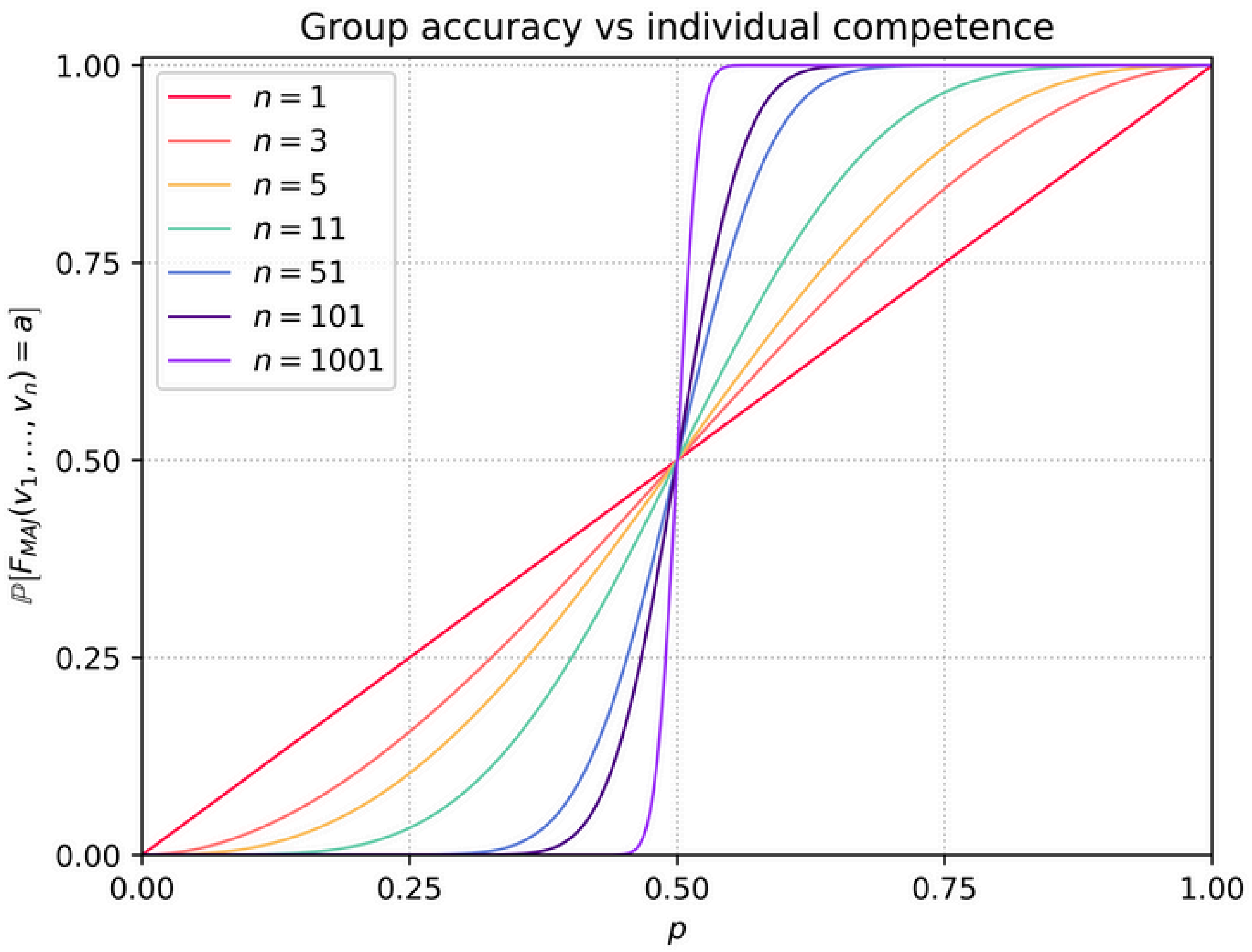
ANY ODD NUMBER OF VOTERS

The profile is $\boldsymbol{v} = (v_1, \ldots, v_n)$, for $n = 2k+1$ and $k \geq 1$.

The probability of a correct decision is:

$$
\begin{aligned}
\Pr\Big[S_n > k\Big] &= \Pr\Big[S_n = k{+}1 \text{ or } \ldots \text{ or } S_n = n\Big] \\
&= \binom{n}{k+1} \cdot p^{k+1}(1-p)^{n-(k+1)} + \cdots + \binom{n}{n-1} \cdot p^{n-1}(1-p)^1 + \binom{n}{n} p \\
&= \sum_{i=k+1}^{n} \binom{n}{i} \cdot p^i (1-p)^{n-i}.
\end{aligned}
$$

And it looks like the same reasoning applies: as n grows, so does group accuracy!



ANY ODD NUMBER OF VOTERS

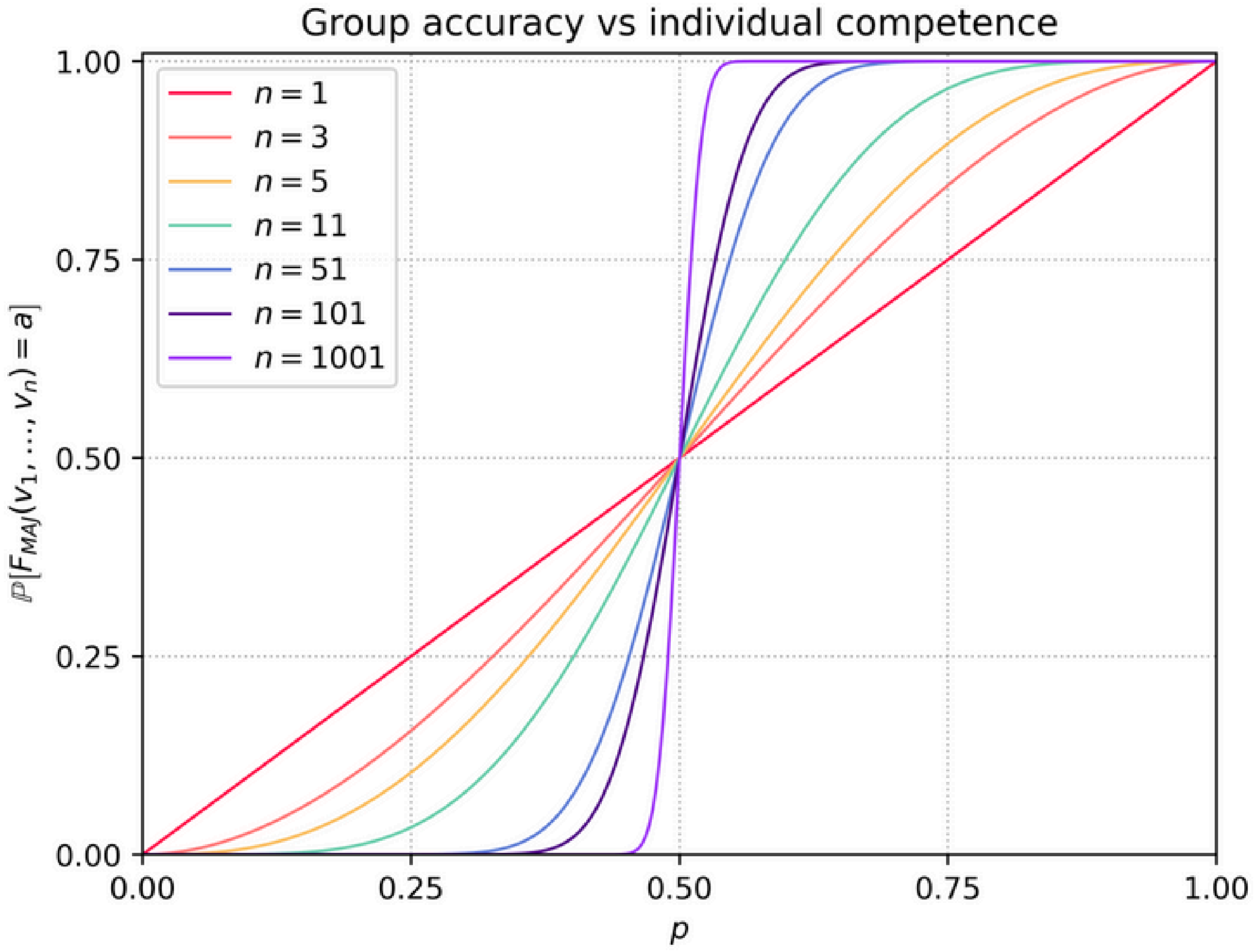
The profile is $\boldsymbol{v} = (v_1, \ldots, v_n)$, for $n = 2k+1$ and $k \geq 1$.

The probability of a correct decision is:

$$\begin{aligned}
\Pr\Big[S_n > k\Big] &= \Pr\Big[S_n = k{+}1 \text{ or } \ldots \text{ or } S_n = n\Big] \\
&= \binom{n}{k+1} \cdot p^{k+1}(1-p)^{n-(k+1)} + \cdots + \binom{n}{n-1} \cdot p^{n-1}(1-p)^1 + \binom{n}{n} p \\
&= \sum_{i=k+1}^{n} \binom{n}{i} \cdot p^i (1-p)^{n-i}.
\end{aligned}$$

And it looks like the same reasoning applies: as n grows, so does group accuracy!

But only as long as $p > 1/2$…



To prove that accuracy increases with group size, we derive a recurrence relation for the probability of a correct decision with $n + 2$ voters, given the probability of a correct decision with n voters.

To prove that accuracy increases with group size, we derive a recurrence relation for the probability of a correct decision with $n+2$ voters, given the probability of a correct decision with n voters.

Take $n = 5$.

FIVE VOTERS AND A CORRECT MAJORITY

We use a clever way of counting the cases that lead to a correct decision.

$$(\square, \square, \square, \square, \square)$$

FIVE VOTERS AND A CORRECT MAJORITY

We use a clever way of counting the cases that lead to a correct decision.

Separate the first two voters.

$$(\square, \square, \square, \square, \square)$$

FIVE VOTERS AND A CORRECT MAJORITY

We use a clever way of counting the cases that lead to a correct decision.

Separate the first two voters.

$$(\square, \square, \square, \square, \square)$$

If they're both wrong, the remaining three have to be correct.

$$(0, 0, 1, 1, 1)$$

FIVE VOTERS AND A CORRECT MAJORITY

We use a clever way of counting the cases that lead to a correct decision.

Separate the first two voters.

$$(\square, \square, \square, \square, \square)$$

If they're both wrong, the remaining three have to be correct.

$$(0, 0, 1, 1, 1)$$

If exactly one of them is correct, which can happen in two ways, at least two of the remaining voters have to be correct.

$$(1, 0, 1, 1, 1)\,(1, 0, 1, 1, 0)\,(1, 0, 1, 0, 1)\,(1, 0, 0, 1, 1)$$
$$(0, 1, 1, 1, 1)\,(0, 1, 1, 1, 0)\,(0, 1, 1, 0, 1)\,(0, 1, 0, 1, 1)$$

FIVE VOTERS AND A CORRECT MAJORITY

We use a clever way of counting the cases that lead to a correct decision.

Separate the first two voters.

$$(\square, \square, \square, \square, \square)$$

If they're both wrong, the remaining three have to be correct.

$$(0, 0, 1, 1, 1)$$

If exactly one of them is correct, which can happen in two ways, at least two of the remaining voters have to be correct.

$$(1, 0, 1, 1, 1)\,(1, 0, 1, 1, 0)\,(1, 0, 1, 0, 1)\,(1, 0, 0, 1, 1)$$
$$(0, 1, 1, 1, 1)\,(0, 1, 1, 1, 0)\,(0, 1, 1, 0, 1)\,(0, 1, 0, 1, 1)$$

If both of the first two voters are correct, at least one of the remaining voters has to be correct.

$$(1, 1, 1, 1, 1)$$
$$(1, 1, 1, 1, 0) \quad (1, 1, 1, 0, 1) \quad (1, 1, 0, 1, 1)$$
$$(1, 1, 1, 0, 0) \quad (1, 1, 0, 1, 0) \quad (1, 1, 0, 0, 1)$$

GENERAL RECURRENCE RELATION

The recurrence relation for five voters is thus:

$$\Pr\Big[S_5 > 2\Big] = (1-p)^2 \cdot \Pr\Big[S_3 > 2\Big] + 2p(1-p)^2 \cdot \Pr\Big[S_3 > 1\Big] + p^2 \cdot \Pr\Big[S_3 > 0\Big].$$

GENERAL RECURRENCE RELATION

The recurrence relation for five voters is thus:

$$\Pr\Big[S_5 > 2\Big] = (1-p)^2 \cdot \Pr\Big[S_3 > 2\Big] + 2p(1-p)^2 \cdot \Pr\Big[S_3 > 1\Big] + p^2 \cdot \Pr\Big[S_3 > 0\Big].$$

The general relation is:

$$\Pr\Big[S_{2k+3} > k{+}1\Big] = (1-p)^2 \cdot \Pr\Big[S_{2k+1} > k{+}1\Big] + 2p(1-p)^2 \cdot \Pr\Big[S_{2k+1} > k\Big] + p^2 \cdot \Pr\Big[S_{2k+1} > k{-}1\Big].$$

GENERAL RECURRENCE RELATION

The recurrence relation for five voters is thus:

$$\Pr\left[S_5 > 2\right] = (1-p)^2 \cdot \Pr\left[S_3 > 2\right] + 2p(1-p)^2 \cdot \Pr\left[S_3 > 1\right] + p^2 \cdot \Pr\left[S_3 > 0\right].$$

The general relation is:

$$\Pr\left[S_{2k+3} > k+1\right] = (1-p)^2 \cdot \Pr\left[S_{2k+1} > k+1\right] + 2p(1-p)^2 \cdot \Pr\left[S_{2k+1} > k\right] + p^2 \cdot \Pr\left[S_{2k+1} > k-1\right].$$

SOME REWRITING

Write:

$$\Pr\left[S_{2k+1} > k-1\right] = \Pr\left[S_{2k+1} > k\right] + \Pr\left[S_{2k+1} = k\right]$$

SOME REWRITING

Write:

$$\begin{aligned}\Pr\left[S_{2k+1} > k-1\right] &= \Pr\left[S_{2k+1} > k\right] + \Pr\left[S_{2k+1} = k\right] \\ &= \Pr\left[S_{2k+1} > k\right] + \binom{2k+1}{k} \cdot p^k (1-p)^{k+1}\end{aligned}$$

SOME REWRITING

Write:

$$\Pr\left[S_{2k+1} > k-1\right] = \Pr\left[S_{2k+1} > k\right] + \Pr\left[S_{2k+1} = k\right]$$

$$= \Pr\left[S_{2k+1} > k\right] + \binom{2k+1}{k} \cdot p^k (1-p)^{k+1}$$

$$\Pr\left[S_{2k+1} > k+1\right] = \Pr\left[S_{2k+1} > k\right] - \Pr\left[S_{2k+1} = k+1\right]$$

SOME REWRITING

Write:

$$\Pr\Big[S_{2k+1} > k-1\Big] = \Pr\Big[S_{2k+1} > k\Big] + \Pr\Big[S_{2k+1} = k\Big]$$

$$= \Pr\Big[S_{2k+1} > k\Big] + \binom{2k+1}{k} \cdot p^k (1-p)^{k+1}$$

$$\Pr\Big[S_{2k+1} > k+1\Big] = \Pr\Big[S_{2k+1} > k\Big] - \Pr\Big[S_{2k+1} = k+1\Big]$$

$$= \Pr\Big[S_{2k+1} > k\Big] - \binom{2k+1}{k+1} \cdot p^{k+1} (1-p)^k$$

SOME REWRITING

Write:

$$\Pr\left[S_{2k+1} > k-1\right] = \Pr\left[S_{2k+1} > k\right] + \Pr\left[S_{2k+1} = k\right]$$
$$= \Pr\left[S_{2k+1} > k\right] + \binom{2k+1}{k} \cdot p^k (1-p)^{k+1}$$
$$\Pr\left[S_{2k+1} > k+1\right] = \Pr\left[S_{2k+1} > k\right] - \Pr\left[S_{2k+1} = k+1\right]$$
$$= \Pr\left[S_{2k+1} > k\right] - \binom{2k+1}{k+1} \cdot p^{k+1} (1-p)^k$$

And:

$$\binom{2k+1}{k} = \binom{2k+1}{k+1} = c$$

Now let’s plug this into the recurrence relation.

PROOF OF CLAIM 1: ACCURACY INCREASES WITH SIZE

$$\Pr\left[S_{2k+3} > k+1\right] = (1-p)^2 \cdot \Pr\left[S_{2k+1} > k+1\right] + 2p(1-p)^2 \cdot \Pr\left[S_{2k+1} > k\right] + p^2 \cdot \Pr\left[S_{2k+1} > k-1\right]$$

PROOF OF CLAIM 1: ACCURACY INCREASES WITH SIZE

$$\Pr\left[S_{2k+3} > k+1\right] = (1-p)^2 \cdot \Pr\left[S_{2k+1} > k+1\right] + 2p(1-p)^2 \cdot \Pr\left[S_{2k+1} > k\right] + p^2 \cdot \Pr\left[S_{2k+1} > k-1\right]$$

$$\ldots$$

$$= \Pr\left[S_{2k+1} > k\right] + c \cdot p^{k+1} \cdot (1-p)^{k+1} \cdot (2p-1)$$

PROOF OF CLAIM 1: ACCURACY INCREASES WITH SIZE

$$
\begin{aligned}
\Pr\left[S_{2k+3} > k+1\right] &= (1-p)^2 \cdot \Pr\left[S_{2k+1} > k+1\right] + 2p(1-p)^2 \cdot \Pr\left[S_{2k+1} > k\right] + p^2 \cdot \Pr\left[S_{2k+1} > k-1\right] \\
&\dots \\
&= \Pr\left[S_{2k+1} > k\right] + c \cdot p^{k+1} \cdot (1-p)^{k+1} \cdot (2p-1) \\
&> \Pr\left[S_{2k+1} > k\right].
\end{aligned}
$$

PROOF OF CLAIM 2: GROUPS BETTER THAN MEMBERS

This follows from Claim 1, as single voters are just groups of size 1.

$$\begin{aligned} p &= \Pr\Big[S_1 > 0\Big] \\ &< \Pr\Big[S_3 > 1\Big] \\ &\cdots \\ &< \Pr\Big[S_n > \lfloor n/2 \rfloor\Big] \\ &\cdots \end{aligned}$$

The claim that in the limit accuracy is perfect follows from the *Law of Large Numbers*.

THE (WEAK) LAW OF LARGE NUMBERS

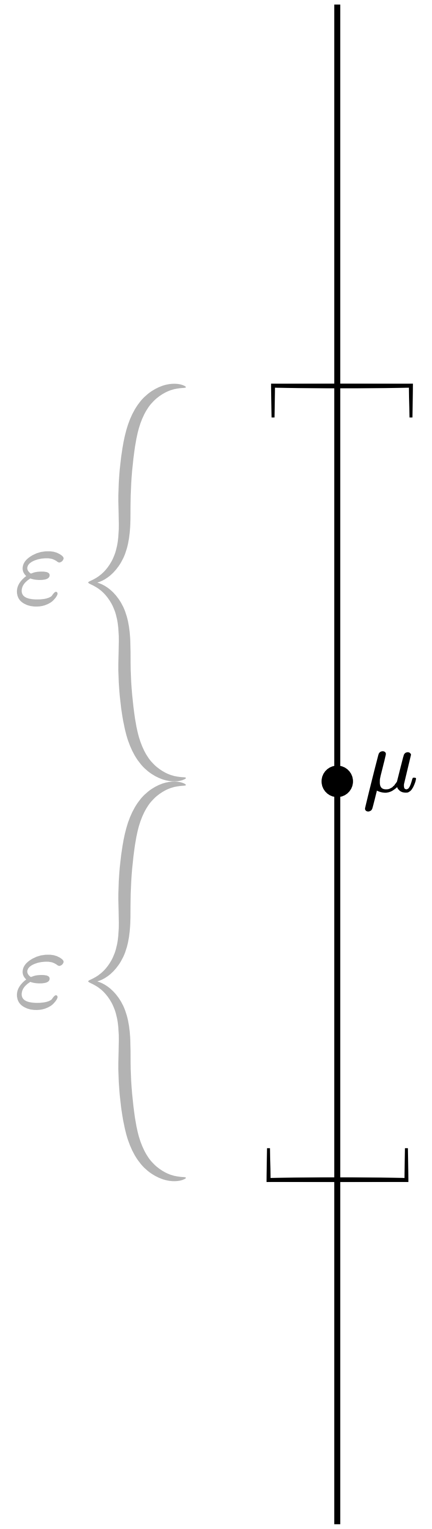
THEOREM

If $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables such that $\mathbb{E}[X_i] = \mu$, then, for any $\varepsilon > 0$, it holds that:

$$\lim_{n\to\infty} \Pr\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right] = 1.$$

THE (WEAK) LAW OF LARGE NUMBERS: INTUITION

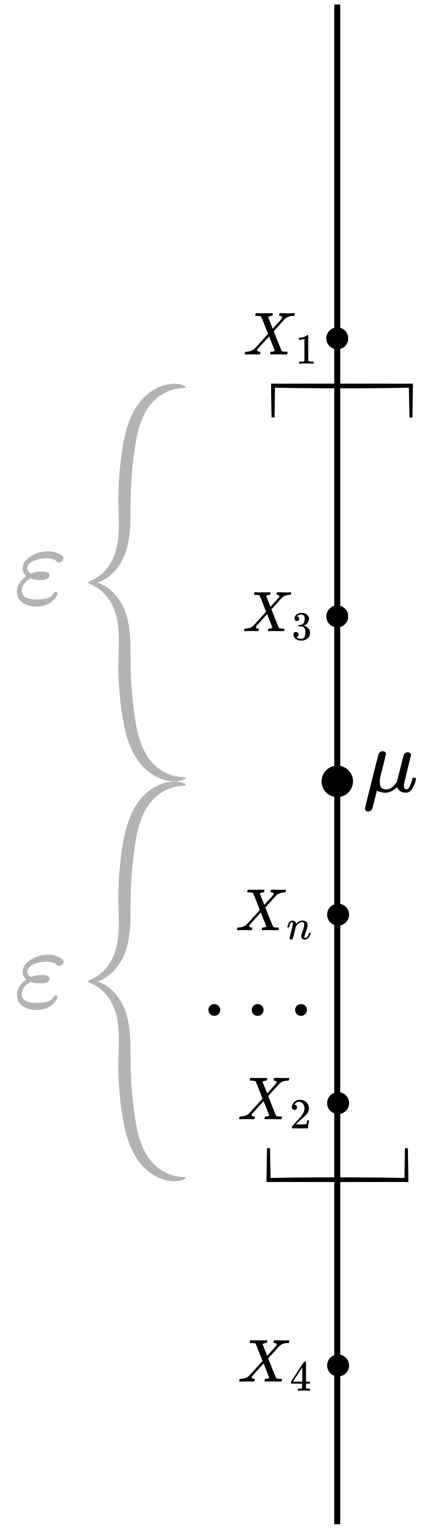
Pick an ε, as small as you want.



THE (WEAK) LAW OF LARGE NUMBERS: INTUITION

Pick an ε, as small as you want.

Sample n variables.

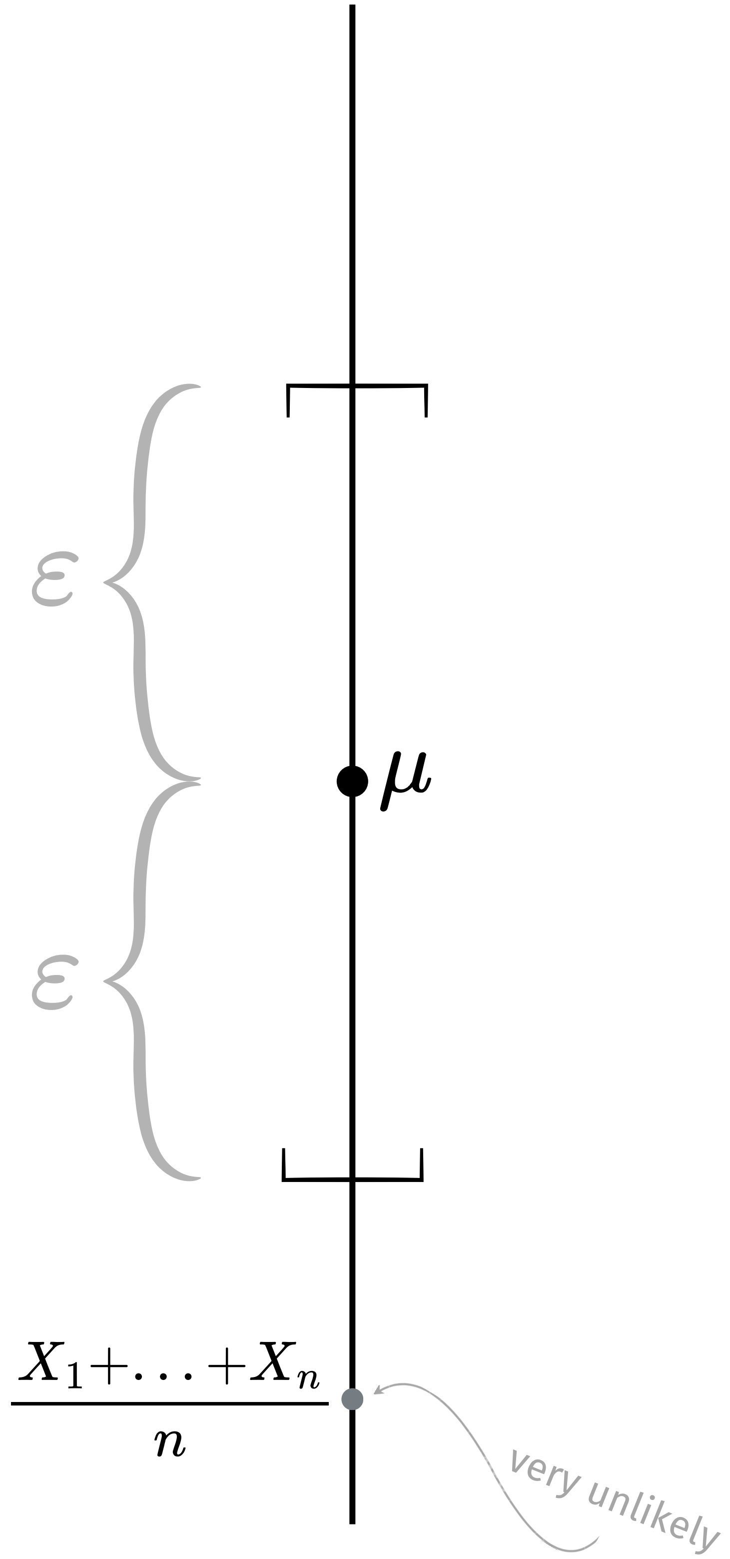


THE (WEAK) LAW OF LARGE NUMBERS: INTUITION

Pick an ε, as small as you want.

Sample n variables.

As n grows, it is overwhelmingly likely that the average of the sampled random variables falls within ε of the mean μ.

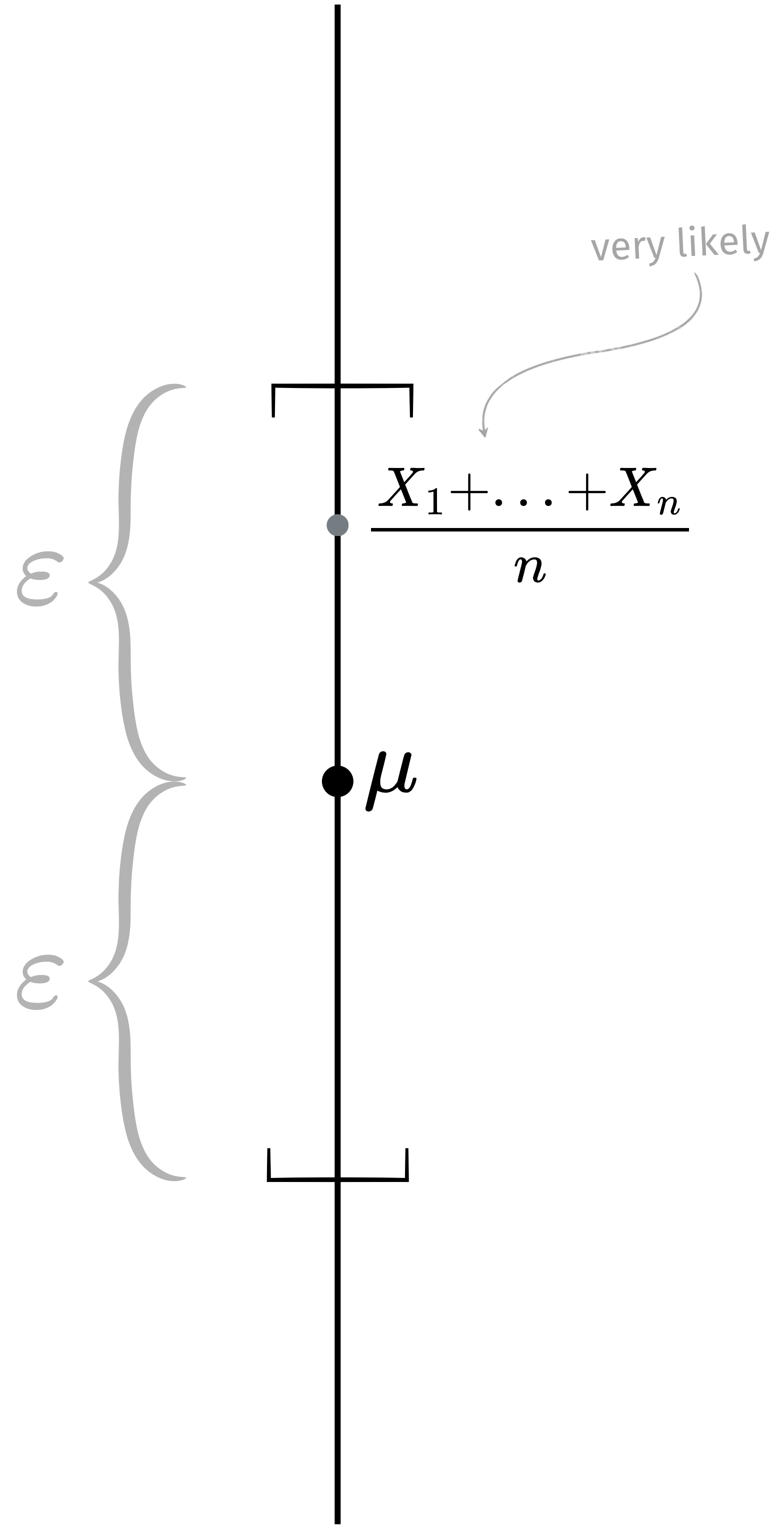


THE (WEAK) LAW OF LARGE NUMBERS: INTUITION

Pick an ε, as small as you want.

Sample n variables.

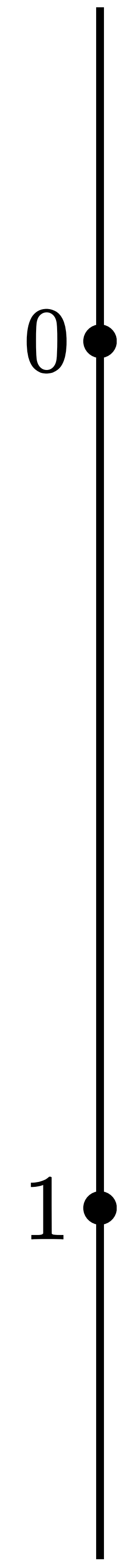
As n grows, it is overwhelmingly likely that the average of the sampled random variables falls within ε of the mean μ.



THE LAW OF LARGE NUMBERS: EXAMPLE

Consider some Bernoulli random variables that keep track of whether someone dies from some disease:

$$X_i = \begin{cases} 1, & \text{with probability } 0.02, \\ 0, & \text{with probability } 0.98, \end{cases}$$



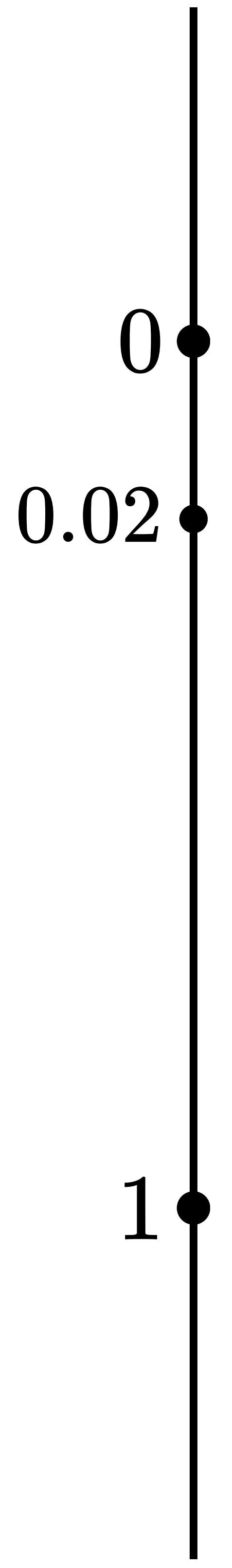
THE LAW OF LARGE NUMBERS: EXAMPLE

Consider some Bernoulli random variables that keep track of whether someone dies from some disease:

$$X_i = \begin{cases} 1, & \text{with probability } 0.02, \\ 0, & \text{with probability } 0.98, \end{cases}$$

The expected value of such a variable is:

$$\mathbb{E}[X_i] = 1 \cdot 0.02 + 0 \cdot 0.98 = 0.02.$$



THE LAW OF LARGE NUMBERS: EXAMPLE

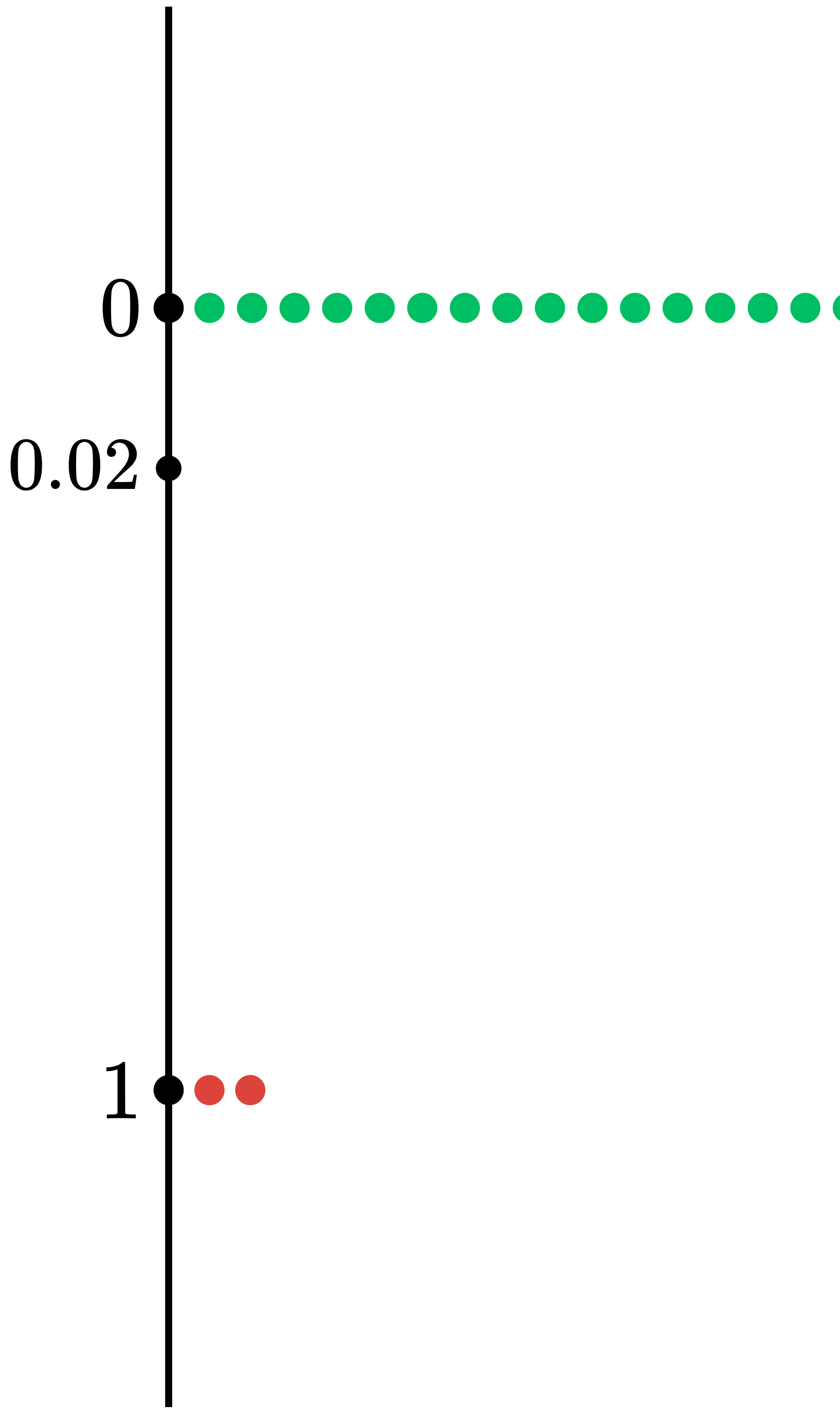
Consider some Bernoulli random variables that keep track of whether someone dies from some disease:

$$X_i = \begin{cases} 1, & \text{with probability } 0.02, \\ 0, & \text{with probability } 0.98, \end{cases}$$

The expected value of such a variable is:

$$\mathbb{E}[X_i] = 1 \cdot 0.02 + 0 \cdot 0.98 = 0.02.$$

If we sample a million such variables, we'd expect about 2% of them to take value 1.



THE LAW OF LARGE NUMBERS: EXAMPLE

Consider some Bernoulli random variables that keep track of whether someone dies from some disease:

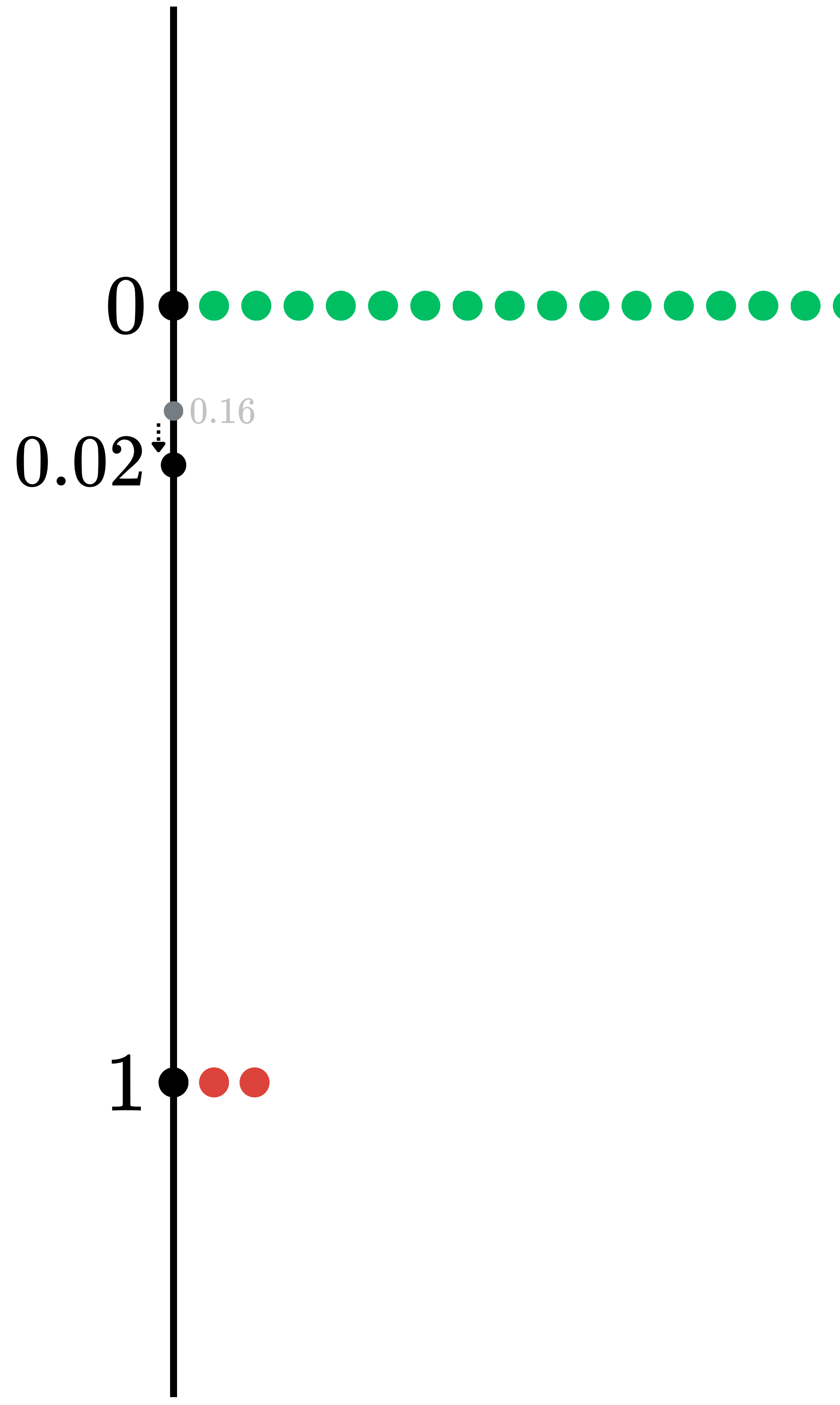
$$X_i = \begin{cases} 1, & \text{with probability } 0.02, \\ 0, & \text{with probability } 0.98, \end{cases}$$

The expected value of such a variable is:

$$\mathbb{E}[X_i] = 1 \cdot 0.02 + 0 \cdot 0.98 = 0.02.$$

If we sample a million such variables, we'd expect about 2% of them to take value 1.

Or, put differently: the average to be very close to 0.02.



BACK TO OUR VOTING SCENARIO

In our case, each independent random variable v_i keeps track of whether voter i votes correctly, with:

$$v_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p. \end{cases}$$

BACK TO OUR VOTING SCENARIO

In our case, each independent random variable v_i keeps track of whether voter i votes correctly, with:

$$v_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p. \end{cases}$$

The majority vote is correct when:

$$v_1 + v_2 + \cdots + v_n > \frac{n}{2}$$

BACK TO OUR VOTING SCENARIO

In our case, each independent random variable v_i keeps track of whether voter i votes correctly, with:

$$v_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p. \end{cases}$$

The majority vote is correct when:

$$v_1 + v_2 + \cdots + v_n > \frac{n}{2} \quad \text{iff} \quad \frac{v_1 + \ldots v_n}{n} > \frac{1}{2}.$$

PROOF OF CLAIM 3: ASYMPTOTIC ACCURACY

The *Law of Large Numbers* gives us that, as n grows, $(v_1 + \ldots v_n)/n$ gets very close to the expected value of the random variables v_i.

PROOF OF CLAIM 3: ASYMPTOTIC ACCURACY

The *Law of Large Numbers* gives us that, as n grows, $(v_1 + \ldots v_n)/n$ gets very close to the expected value of the random variables v_i.

The expected value (i.e., mean μ) is:

$$\begin{aligned} \mathbb{E}[v_i] &= 1 \cdot p + 0 \cdot (1-p) \\ &= p. \end{aligned}$$

PROOF OF CLAIM 3: ASYMPTOTIC ACCURACY

The *Law of Large Numbers* gives us that, as n grows, $(v_1 + \dots v_n)/n$ gets very close to the expected value of the random variables v_i.

The expected value (i.e., mean μ) is:

$$\begin{aligned} \mathbb{E}[v_i] &= 1 \cdot p + 0 \cdot (1 - p) \\ &= p. \end{aligned}$$

So, for very large n, with high probability:

$$\begin{aligned} \frac{v_1 + \cdots + v_n}{n} &\approx p \\ &> \frac{1}{2}. \end{aligned}$$

PROOF OF CLAIM 3: ASYMPTOTIC ACCURACY

The *Law of Large Numbers* gives us that, as n grows, $(v_1 + \ldots v_n)/n$ gets very close to the expected value of the random variables v_i.

The expected value (i.e., mean μ) is:

$$\begin{aligned} \mathbb{E}[v_i] &= 1 \cdot p + 0 \cdot (1 - p) \\ &= p. \end{aligned}$$

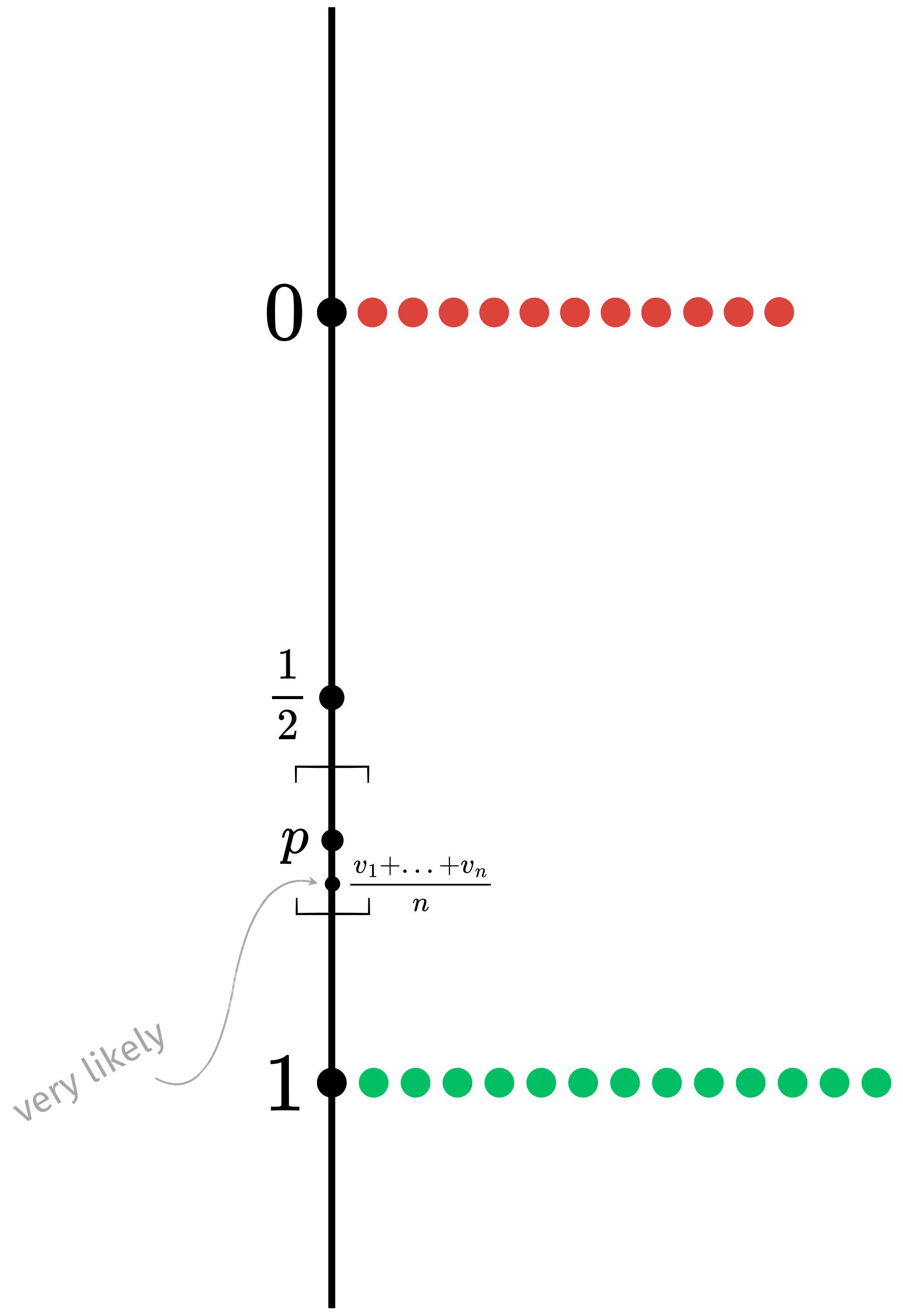
So, for very large n, with high probability:

$$\begin{aligned} \frac{v_1 + \cdots + v_n}{n} &\approx p \\ &> \frac{1}{2}. \end{aligned}$$

This can be made precise with an appropriate choice of ε in the *Law of Large Numbers*.

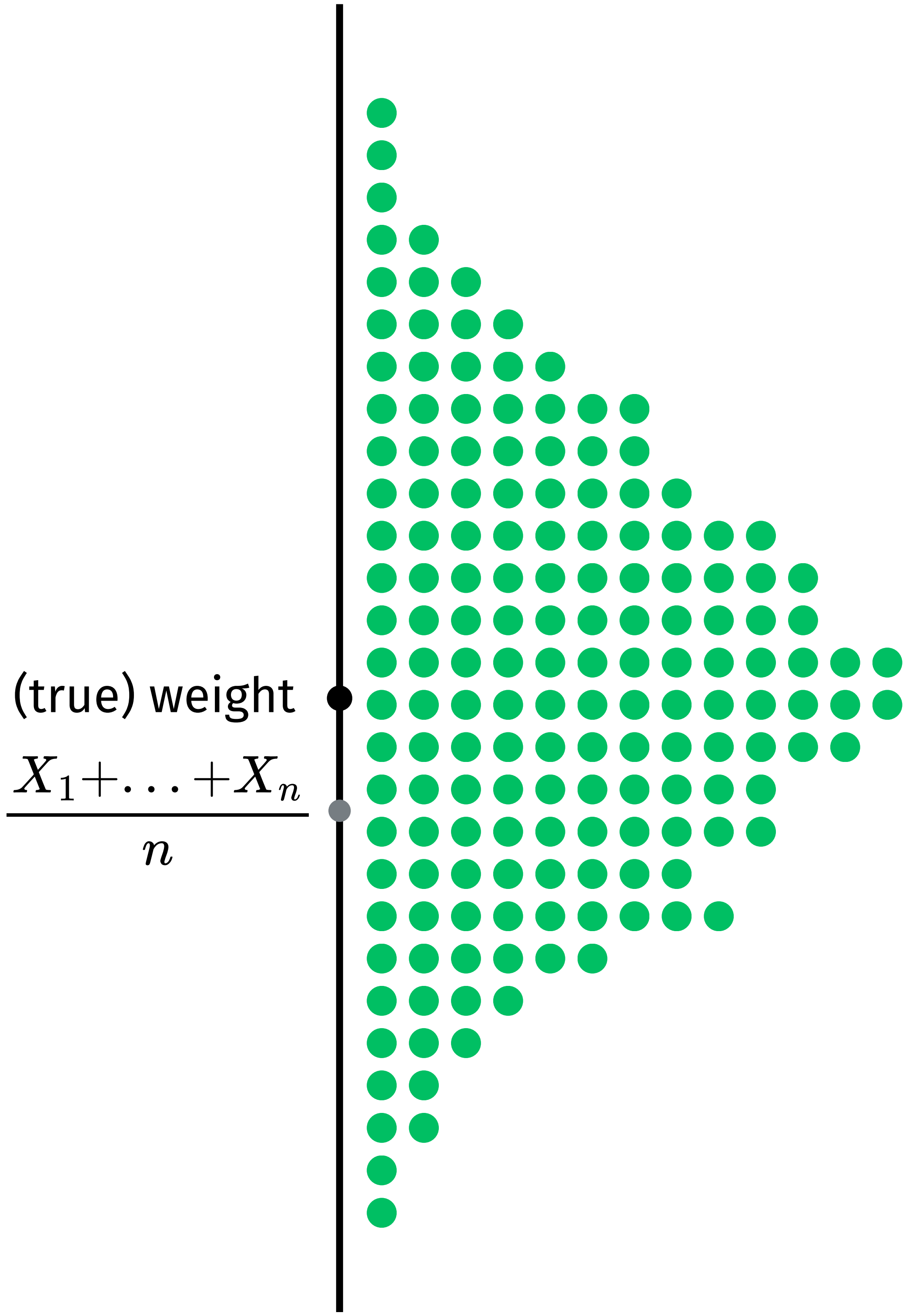
ASYMPTOTIC ACCURACY: INTUITION

The intuition is simple: in the long run, more people end up voting correctly than not.



FRANCIS GALTON

This probably also explains what happened at the country fair!



Let's sum up.

THE MARQUIS DE CONDORCET

Groups are better than their members.

The larger the group, the better.

In the limit, performance is perfect.

THE MARQUIS DE CONDORCET

Groups are better than their members.

The larger the group, the better.

In the limit, performance is perfect.

As long as people are better than random, and vote independently!