2. One Rule to Rule Them All?

Adrian Haret

a.haret@uva.nl

June 8, 2022
Outline

1. The Optimal Voting Rule for Two Alternatives

2. Voting Rules as Maximum Likelihood Estimators (MLEs)
Recall the CJT

Theorem: The Condorcet Jury Theorem (CJT)

If agents have the same accuracy $p > \frac{1}{2}$ and vote independently of each other, then, for odd $n$, it holds that:

(LIB) *Larger is better*, i.e., accuracy of the group improves as its size grows:

$$\Pr[F_{\text{MAJ}}(v_1, \ldots, v_{n+2}) = a] > \Pr[F_{\text{MAJ}}(v_1, \ldots, v_n) = a].$$

(GBI) *Groups are better than individuals*, i.e., accuracy of the group better than that of any of its members:

$$\Pr[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] > \Pr[v_i = a], \text{ for } n \geq 3.$$

(ASY) The accuracy of the group approaches 1 *asymptotically*:

$$\lim_{n \to \infty} \Pr[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] = 1.$$
Relaxing the equal competence assumption

What happens if we drop the assumption that all agents have the same competence?
1. The Optimal Voting Rule for Two Alternatives
The model

agents, aka voters \( N = \{1, \ldots, n\} \)

\( n \) typically assumed odd

two alternatives \( A = \{a, b\} \)

the correct alternative \( a \)

voter \( i \)'s vote \( v_i \in A \)

\( i \)'s guess of the right answer

voter \( i \)'s competence \( p_i = \mathbb{P}[v_i = a], \text{ for } p_i \in [0, 1] \)

the probability that voter \( i \) gets the right answer

profile of votes \( \mathbf{v} = (v_1, \ldots, v_n) \)

voting rule \( F(\mathbf{v}) \mapsto x \in A \)

majority rule \( F_{\text{MAJ}}(\mathbf{v}) = x, \text{ such that } v_i = x \text{ for a strict majority of voters in } N \)
agents, aka voters $N = \{1, \ldots, n\}$

$n$ typically assumed odd

two alternatives $A = \{a, b\}$

the correct alternative $a$

voter $i$’s vote $v_i \in A$

$i$’s guess of the right answer

voter $i$’s competence $p_i = \mathbb{P}[v_i = a], \text{ for } p_i \in [0, 1]$

the probability that voter $i$ gets the right answer

profile of competences $p = (p_1, \ldots, p_n)$, assuming wlog that $p_1 \geq \cdots \geq p_n$

profile of votes $v = (v_1, \ldots, v_n)$

voting rule $F(v) \mapsto x \in A$

majority rule $F_{\text{MAJ}}(v) = x$, such that $v_i = x$ for a strict majority of voters in $N$
We need to recalculate

**Observation: Probability of correct majority for** \( n = 3 \)

\[
\mathbb{P}[F_{MAJ}(v_1, v_2, v_3) = a] = \mathbb{P}[v \in \{aab, aba, baa, aaa\}]
\]

\[
= p_1p_2(1 - p_2) + p_1(1 - p_2)p_3 + (1 - p_1)p_2p_3 + p_1p_2p_3
\]

\[
= p_1p_2 + p_2p_3 + p_1p_3 - 2p_1p_2p_3.
\]
We need to recalculate

Observation: Probability of correct majority for \( n = 3 \)

\[
P[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = P[v \in \{aab, aba, baa, aaa\}] \\
= p_1p_2(1 - p_2) + p_1(1 - p_2)p_3 + (1 - p_1)p_2p_3 + p_1p_2p_3 \\
= p_1p_2 + p_2p_3 + p_1p_3 - 2p_1p_2p_3.
\]

Observation: Probability of correct majority for any \( n \)

\[
P[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] = \sum_{\text{correct majorities } C, |C| > \frac{n}{2}} \left( \prod_{i \in C} p_i \right) \left( \prod_{N \setminus C} (1 - p_i) \right)
\]
Some examples

**Condorcet**

What can happen with different competences?

Sacré bleu! The group is not necessarily better than its members also, larger groups are not necessarily better (LIB). Jacob Paroush and the asymptotic claim (ASY) fail as well!

\[ \angle \beta = (0.9, 0.6, 0.55) : \]

\[ P[F_{MAJ}(v_1, \ldots, v_5)] = a \approx 0.77 \]

\[ \angle \beta \approx (0.9, 0.6, 0.55, 0.55, 0.55) : \]

\[ P[F_{MAJ}(v_1, \ldots, v_n)] = a \approx 0.76 \]

\[ \lim_{n \to \infty} P[F_{MAJ}(v_1, \ldots, v_n)] \neq 1. \]

Some examples

**Condorcet**
What can happen with different competences?

Sacré bleu! The group is not necessarily better than its members

\[ p = (0.9, 0.6, 0.55): \]
\[ P[F_{MAJ}(v_1, v_2, v_3) = a] = 0.77 \]

Some examples

**Condorcet**
What can happen with different competences?

Sacré bleu! The group is not necessarily better than its members

Also, larger groups are not necessarily better (LIB).

\[ \mathbf{p} = (0.9, 0.6, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.77 \]

\[ \mathbf{p} = (0.9, 0.6, 0.55, 0.55, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_5) = a] = 0.76 \]
Some examples

**Condorcet**

What can happen with different competences?

Sacré bleu! The group is not necessarily better than its members.

Also, larger groups are not necessarily better (LIB).

**Jacob Paroush**

And the asymptotic claim (ASY) fails as well!

\[ p = (0.9, 0.6, 0.55) : \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.77 \]

\[ p = (0.9, 0.6, 0.55, 0.55, 0.55) : \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_5) = a] = 0.76 \]

If \( p_i \)'s approach \( \frac{1}{2} \) very fast:

\[ \lim_{n \to \infty} \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] \neq 1. \]

Recovers (ASY): bounding competences away from $\frac{1}{2}$

Theorem: Paroush (1998)

If votes are independent, competences are $p = (p_1, \ldots, p_n)$ and, for any $i$, there is $\epsilon > 0$ such that $p_i > \frac{1}{2} + \epsilon$, then, for odd $n$, it holds that:

$$(\text{ASY}) \lim_{n \to \infty} \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] = 1.$$
Recovering (ASY): bounding competences away from $1/2$

**Theorem: Paroush (1998)**

If votes are independent, competences are $p = (p_1, \ldots, p_n)$ and, for any $i$, there is $\varepsilon > 0$ such that $p_i > \frac{1}{2} + \varepsilon$, then, for odd $n$, it holds that:

$$(ASY) \lim_{n \to \infty} \mathbb{P}[F_{MAJ}(v_1, \ldots, v_n) = a] = 1.$$  

**Proof**

- For $0 < \delta < 1 - p_i$, we say that $\tilde{p} = (p_1, \ldots, p_i + \delta, \ldots, p_n)$ is an improvement of $p = (p_1, \ldots, p_i, \ldots, p_n)$.
- (Lemma 1) If $\tilde{p}$ is an improvement of $p$, then:

$$\mathbb{P}[F_{MAJ}(\text{"}v\text{ with }\tilde{p}\text{"}) = a] > \mathbb{P}[F_{MAJ}(\text{"}v\text{ with }p\text{"}) = a].$$

- Proof of this is “straightforward”.
- We can get any $p = (p_1, \ldots, p_n)$ from $p' = (\frac{1}{2} + \varepsilon, \ldots, \frac{1}{2} + \varepsilon)$ in $n$ improvements.
- But we already know that $\mathbb{P}[F_{MAJ}(\text{"}v\text{ with }p'\text{"}) = a] = 1$, from the (standard) CJT.
Recovering (ASY): keeping average competence above $1/2$


If votes are independent, competences are $p = (p_1, \ldots, p_n)$ and $\frac{p_1 + \cdots + p_n}{n} = \bar{p} > \frac{1}{2}$, then, for odd $n$, it holds that:

\[
(\text{ASY}) \lim_{n \to \infty} P \left[ F_{\text{MAJ}}(v_1, \ldots, v_n) = a \right] = 1.
\]
Recovering (ASY): keeping average competence above $\frac{1}{2}$


If votes are independent, competences are $p = (p_1, \ldots, p_n)$ and $\frac{p_1 + \cdots + p_n}{n} = \overline{p} > \frac{1}{2}$, then, for odd $n$, it holds that:

$\lim_{n \to \infty} \mathbb{P} [F_{\text{MAJ}}(v_1, \ldots, v_n) = a] = 1.$

Proof: (idea)

> Look at polynomial for $\mathbb{P} [F_{\text{MAJ}}(v_1, \ldots, v_n) = a]$ as a function in the $p_i$’s.

> Use derivatives with respect to $p_i$ to understand how to maximize it.

---

Recovering (ASY): keeping average competence above $\frac{1}{2}$

**Theorem: Owen, Grofman, Feld (1989)**

If votes are independent, competences are $p = (p_1, \ldots, p_n)$ and $\frac{p_1 + \cdots + p_n}{n} = \bar{p} > \frac{1}{2}$, then, for odd $n$, it holds that:

$$\text{(ASY)} \lim_{n \to \infty} \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_n) = a] = 1.$$  

**Proof: (idea)**

- Look at polynomial for $\mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_n) = a]$ as a function in the $p_i$’s.
- Use derivatives with respect to $p_i$ to understand how to maximize it.

**Observation: Strange things still happen**

Can still get $> \frac{1}{2}$ group competence even with < $\frac{1}{2}$ average voter competence, and vice-versa:

- If $p = (1, 0.28, 0.28)$, then $\bar{p} = 0.52$ and $\mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.48$.
- If $p = (0.72, 0.72, 0)$, we get $\bar{p} = 0.48$ and $\mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.52$.
It strikes me that in these examples we can do better.

> \( p = (0.9, 0.6, 0.55) \):

\[
\mathbb{P}[F_{\text{maj}}(v_1, v_2, v_3) = a] = 0.77
\]

> \( p = (0.9, 0.6, 0.55, 0.55, 0.55) \):

\[
\mathbb{P}[F_{\text{maj}}(v_1, \ldots, v_5) = a] = 0.76
\]
Weird examples revisited

Shmuel Nitzan
It strikes me that in these examples we can do better.

Just let the most competent voter decide.

\[ p = (0.9, 0.6, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.77 \]

\[ p = (0.9, 0.6, 0.55, 0.55, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_5) = a] = 0.76 \]
Weird examples revisited

Shmuel Nitzan

It strikes me that in these examples we can do better. Just let the most competent voter decide. Instead of the majority.

\[ p = (0.9, 0.6, 0.55): \]
\[ P[F_{MAJ}(v_1, v_2, v_3) = a] = 0.77 \]
\[ p = (0.9, 0.6, 0.55, 0.55, 0.55): \]
\[ P[F_{MAJ}(v_1, \ldots, v_5) = a] = 0.76 \]
Shmuel Nitzan
It strikes me that in these examples we can do better.

Just let the most competent voter decide.

Instead of the majority.

Condorcet
Mon dieu!

\[ p = (0.9, 0.6, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, v_2, v_3) = a] = 0.77 \]

\[ p = (0.9, 0.6, 0.55, 0.55, 0.55): \]
\[ \mathbb{P}[F_{\text{MAJ}}(v_1, \ldots, v_5) = a] = 0.76 \]
Beyond majority rule?

What if we play around with the voting rule?
The model + expert rule

agents, aka voters \( N = \{1, \ldots, n\} \)
n typically assumed odd

two alternatives \( A = \{a, b\} \)
the correct alternative \( a \)
voter \( i \)'s vote \( v_i \in A \)
\( i \)'s guess of the right answer
voter \( i \)'s competence \( p_i = \Pr[v_i = a], \text{ for } p_i \in [0, 1] \)
the probability that voter \( i \) gets the right answer
profile of competences \( p = (p_1, \ldots, p_n), \text{ assuming wlog that } p_1 \geq \cdots \geq p_n \)
profile of votes \( v = (v_1, \ldots, v_n) \)
voting rule \( F(v) \mapsto x \in A \)
majority rule \( F_{MAJ}(v) = x, \text{ such that } v_i = x \text{ for a strict majority of voters in } N \)
The model + expert rule

- **agents, aka voters** \( N = \{1, \ldots, n\} \)
- \( n \) typically assumed odd
- **two alternatives** \( A = \{a, b\} \)
- **the correct alternative** \( a \)
- **voter \( i \)'s vote** \( v_i \in A \)
- \( i \)'s guess of the right answer
- **voter \( i \)'s competence** \( p_i = \mathbb{P}[v_i = a] \), for \( p_i \in [0, 1] \)
- the probability that voter \( i \) gets the right answer
- **profile of competences** \( p = (p_1, \ldots, p_n) \), assuming wlog that \( p_1 \geq \cdots \geq p_n \)
- **profile of votes** \( v = (v_1, \ldots, v_n) \)
- **voting rule** \( F(v) \mapsto x \in A \)
- **majority rule** \( F_{\text{MAJ}}(v) = x \), such that \( v_i = x \) for a strict majority of voters in \( N \)
- **expert rule** \( F_{\text{EXP}}(v) = v_1 \)
- generally, the choice of the most accurate agent
When is it preferable to use the expert rule instead of majority rule?
**Observation:** For $n = 3$

Expert rule beats majority rule precisely when:

\[
\mathbb{P}[F_{\text{EXP}}(\mathbf{v}) = a] > \mathbb{P}[F_{\text{MAJ}}(\mathbf{v}) = a] \quad \text{iff} \quad p_1 > p_1p_2 + p_2p_3 + p_1p_3 - 2p_1p_2p_3
\]

\[
p_1(1 - p_2 - p_3 + p_2p_3) > p_2p_3(1 - p_1) \quad \text{iff} \quad p_1(1 - p_1)(1 - p_3) > p_2p_3(1 - p_1)
\]

\[
\frac{p_1}{1 - p_1} > \frac{p_2}{1 - p_2} \frac{p_3}{1 - p_3} \quad \text{iff} \quad \ln \frac{p_1}{1 - p_1} > \ln \frac{p_2}{1 - p_2} + \ln \frac{p_3}{1 - p_3}.
\]
Observation: For $n = 3$

Expert rule beats majority rule precisely when:

$$\mathbb{P}[F_{\text{EXP}}(v) = a] > \mathbb{P}[F_{\text{MAJ}}(v) = a] \quad \text{iff} \quad p_1 > p_1 p_2 + p_2 p_3 + p_1 p_3 - 2p_1 p_2 p_3$$

$$p_1(1 - p_2 - p_3 + p_2 p_3) > p_2 p_3(1 - p_1) \quad \text{iff} \quad p_1(1 - p_2)(1 - p_3) > p_2 p_3(1 - p_1)$$

$$\frac{p_1}{1 - p_1} > \frac{p_2}{1 - p_2} \frac{p_3}{1 - p_3} \quad \text{iff} \quad \ln \frac{p_1}{1 - p_1} > \ln \frac{p_2}{1 - p_2} + \ln \frac{p_3}{1 - p_3}.$$


For any $n \in \mathbb{N}$, if $\ln \frac{p_1}{1 - p_1} > \ln \frac{p_2}{1 - p_2} + \cdots + \ln \frac{p_n}{1 - p_n}$, then the expert rule beats majority rule.

Weighted voting rules

agents, aka voters \( N = \{1, \ldots, n\} \)
\( n \) typically assumed odd

two alternatives \( A = \{a, b\} \), of which \( a \) is the correct one

voter \( i \)'s competence \( p_i = \mathbb{P}[x_i = 1] \), for \( p_i \in [0, 1] \)
the probability that voter \( i \) gets the right answer
agents, aka voters \( N = \{1, \ldots, n\} \)

\( n \) typically assumed odd

two alternatives \( A = \{a, b\}, \) of which \( a \) is the correct one

voter \( i \)’s vote \( x_i = \begin{cases} 1, & \text{if } i \text{ votes for } a, \\ -1, & \text{otherwise}. \end{cases} \)

voter \( i \)’s competence \( p_i = \mathbb{P}[x_i = 1], \) for \( p_i \in [0, 1] \)

the probability that voter \( i \) gets the right answer

agent \( i \)’s weight \( w_i \in \mathbb{R} \)

profile of votes \( x = (x_1, \ldots, x_n) \)

weighted voting rule \( F_w(x) = \begin{cases} a, & \text{if } w_1x_1 + \cdots + w_nx_n > 0, \\ b, & \text{otherwise.} \end{cases} \)
Weighted voting generalizes voting rules we’ve seen before

**Observation**

- the majority rule $F_{\text{MAJ}}$ is given by $w = (1, \ldots, 1)$
- the expert rule $F_{\text{EXP}}$ is given by $w = (1, 0, \ldots, 0)$
Choosing weights

What weights make sense, given the individual competences?
Intuitively, the more competent an agent the more weight it should get.
Weighted voting in action

Lloyd Shapley
Intuitively, the more competent an agent the more weight it should get.

Bernard Grofman
Say the weight is equal to the accuracy.

Shmuel Nitzan
We can do better!

Jacob Paroush
Give better agents even more weight!

\[ p = (0.9, 0.6, 0.55), w = (0.9, 0.6, 0.55) \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( F_w(x) )</th>
<th>prob. of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((-1, -1, -1))</td>
<td>(-1)</td>
<td>0.018</td>
</tr>
<tr>
<td>((-1, -1, 1))</td>
<td>(-1)</td>
<td>0.022</td>
</tr>
<tr>
<td>((-1, 1, -1))</td>
<td>(-1)</td>
<td>0.027</td>
</tr>
<tr>
<td>((-1, 1, 1))</td>
<td>1</td>
<td>0.033</td>
</tr>
<tr>
<td>((1, -1, -1))</td>
<td>(-1)</td>
<td>0.162</td>
</tr>
<tr>
<td>((1, -1, 1))</td>
<td>1</td>
<td>0.198</td>
</tr>
<tr>
<td>((1, 1, -1))</td>
<td>1</td>
<td>0.243</td>
</tr>
<tr>
<td>((1, 1, 1))</td>
<td>1</td>
<td>0.297</td>
</tr>
</tbody>
</table>

\[ \mathbb{P}[F_w(x) = 1] = 0.77 \]
Weighted voting in action

Lloyd Shapley
Intuitively, the more competent an agent the more weight it should get.

Bernard Grofman
Say the weight is equal to the accuracy.

Shmuel Nitzan
We can do better!

\[ p = (0.9, 0.6, 0.55), \; w = (0.9, 0.6, 0.55) \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( F_w(x) )</th>
<th>prob. of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((-1, -1, -1))</td>
<td>(-1)</td>
<td>0.018</td>
</tr>
<tr>
<td>((-1, -1, 1))</td>
<td>(-1)</td>
<td>0.022</td>
</tr>
<tr>
<td>((-1, 1, -1))</td>
<td>(-1)</td>
<td>0.027</td>
</tr>
<tr>
<td>((-1, 1, 1))</td>
<td>1</td>
<td>0.033</td>
</tr>
<tr>
<td>((1, -1, -1))</td>
<td>(-1)</td>
<td>0.162</td>
</tr>
<tr>
<td>((1, -1, 1))</td>
<td>1</td>
<td>0.198</td>
</tr>
<tr>
<td>((1, 1, -1))</td>
<td>1</td>
<td>0.243</td>
</tr>
<tr>
<td>((1, 1, 1))</td>
<td>1</td>
<td>0.297</td>
</tr>
</tbody>
</table>

\( P[F_w(x) = 1] = 0.77 \)
Weighted voting in action

Lloyd Shapley
Intuitively, the more competent an agent the more weight it should get.

Bernard Grofman
Say the weight is equal to the accuracy.

Shmuel Nitzan
We can do better!

Jacob Paroush
Give better agents even more weight!

\[ p = (0.9, 0.6, 0.55), \ w = (0.9, 0.6, 0.55) \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( F_w(x) )</th>
<th>prob. of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((-1, -1, -1))</td>
<td>-1</td>
<td>0.018</td>
</tr>
<tr>
<td>((-1, -1, 1))</td>
<td>-1</td>
<td>0.022</td>
</tr>
<tr>
<td>((-1, 1, -1))</td>
<td>-1</td>
<td>0.027</td>
</tr>
<tr>
<td>((-1, 1, 1))</td>
<td>1</td>
<td>0.033</td>
</tr>
<tr>
<td>((1, -1, -1))</td>
<td>-1</td>
<td>0.162</td>
</tr>
<tr>
<td>((1, -1, 1))</td>
<td>1</td>
<td>0.198</td>
</tr>
<tr>
<td>((1, 1, -1))</td>
<td>1</td>
<td>0.243</td>
</tr>
<tr>
<td>((1, 1, 1))</td>
<td>1</td>
<td>0.297</td>
</tr>
</tbody>
</table>

\[ \mathbb{P}[F_w(x) = 1] = 0.77 \]
Weighted voting in action

Lloyd Shapley
Intuitively, the more competent an agent the more weight it should get.

Bernard Grofman
Say the weight is equal to the accuracy.

Shmuel Nitzan
We can do better!

Jacob Paroush
Give better agents even more weight!

\[ p = (0.9, 0.6, 0.55), w = (1.2, 0.6, 0.55) \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( F_w(x) )</th>
<th>prob. of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(−1, −1, −1)</td>
<td>−1</td>
<td>0.018</td>
</tr>
<tr>
<td>(−1, −1, 1)</td>
<td>−1</td>
<td>0.022</td>
</tr>
<tr>
<td>(−1, 1, −1)</td>
<td>−1</td>
<td>0.027</td>
</tr>
<tr>
<td>(−1, 1, 1)</td>
<td>−1</td>
<td>0.033</td>
</tr>
<tr>
<td>(1, −1, −1)</td>
<td>1</td>
<td>0.162</td>
</tr>
<tr>
<td>(1, −1, 1)</td>
<td>1</td>
<td>0.198</td>
</tr>
<tr>
<td>(1, 1, −1)</td>
<td>1</td>
<td>0.243</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>1</td>
<td>0.297</td>
</tr>
</tbody>
</table>

\[ \mathbb{P}[F_w(x) = 1] = 0.9 \]
What a good rule looks like

Shmuel Nitzan

Note that $F_w(-x) = -F_w(x)$.

$F_w(-x_1, \ldots, -x_n) = -F_w(x_1, \ldots, x_n)$. 
What a good rule looks like

Shmuel Nitzan
Note that $F_w(-x) = -F_w(x)$.

Jacob Paroush
Thus, $F_w$ partitions the set of profiles into two equally sized and symmetric sets of profiles.

$Lloyd Shapley [w/ Liam Neeson]$ This can be achieved with a particular set of weights.

\[ F_w(-x_1, \ldots, -x_n) = -F_w(x_1, \ldots, x_n). \]

\[ \downarrow \]

\[ F_w(x) = 1 \text{ iff } F_w(-x) = -1 \]
What a good rule looks like

**Shmuel Nitzan**

Note that $F_w(-x) = -F_w(x)$.

**Jacob Paroush**

Thus, $F_w$ partitions the set of profiles into two equally sized and symmetric sets of profiles.

**Bernard Grofman**

We want to choose $w$ such that the probability of a profile $x$ that gets things right is higher than the probability of $-x$, which will be wrong.

\[ F_w(-x_1, \ldots, -x_n) = -F_w(x_1, \ldots, x_n). \]

\[ \Downarrow \]

\[ F_w(x) = 1 \text{ iff } F_w(-x) = -1 \]

\[ \Downarrow \]

\[ \text{?} w \text{ such that } \mathbb{P}[F_w(x) = 1] > \mathbb{P}[F_w(-x) = -1] \]
What a good rule looks like

Shmuel Nitzan
Note that $F_w(-x) = -F_w(x)$.

Jacob Paroush
Thus, $F_w$ partitions the set of profiles into two equally sized and symmetric sets of profiles.

Bernard Grofman
We want to choose $w$ such that the probability of a profile $x$ that gets things right is higher than the probability of $-x$, which will be wrong.

Lloyd Shapley [w/ Liam Neeson]
This can be achieved with a particular set of weights.

$F_w(-x_1, \ldots, -x_n) = -F_w(x_1, \ldots, x_n)$.

\[ \therefore \]

$F_w(x) = 1 \text{ iff } F_w(-x) = -1$

\[ \therefore \]

? $w$ such that $\mathbb{P}[F_w(x) = 1] > \mathbb{P}[F_w(-x) = -1]$
The optimal voting rule


If votes are independent and competences are \( p = (p_1, \ldots, p_n) \), then one optimal rule is \( F_{w^*} \), with:

\[
w^* = (\ln \frac{p_1}{1-p_1}, \ldots, \ln \frac{p_n}{1-p_n}).
\]
The optimal voting rule


If votes are independent and competences are $p = (p_1, \ldots, p_n)$, then one optimal rule is $F_{w^*}$, with:

$$w^* = (\ln \frac{p_1}{1-p_1}, \ldots, \ln \frac{p_n}{1-p_n}).$$

**Proof**

> If $x = (x_1, \ldots, x_n)$, then for $x$ such that $F_{w^*}(x) = 1$ we get:

$$\ln \frac{p_1}{1-p_1} x_1 + \cdots + \ln \frac{p_n}{1-p_n} x_n > 0$$

iff

$$\sum_{i \text{ such that } x_i = 1} \ln \frac{p_i}{1-p_i} > \sum_{i \text{ such that } x_i = -1} \ln \frac{p_i}{1-p_i}$$

iff

$$\prod_{i \text{ such that } x_i = 1} \frac{p_i}{1-p_i} > \prod_{i \text{ such that } x_i = -1} \frac{p_i}{1-p_i}$$

iff

$$\prod_{i \text{ such that } x_i = 1} p_i \prod_{i \text{ such that } x_i = -1} (1-p_i) > \prod_{i \text{ such that } x_i = -1} p_i \prod_{i \text{ such that } x_i = 1} (1-p_i)$$

iff

$$\mathbb{P}[\text{we get } x] > \mathbb{P}[\text{we get } -x]$$

---


What did we learn?

Condorcet

So we should give up one-person-one-vote?…


What did we learn?

**Condorcet**

So we should give up one-person-one-vote?…

**Ruth Ben-Yashar**

Well in practice we might not know the $p_i$’s. In which case might be better to go with majority.

---

What did we learn?

**Condorcet**

So we should give up one-person-one-vote?...

**Ruth Ben-Yashar**

Well in practice we might not know the $p_i$’s. In which case might be better to go with majority.

**Eyal Baharad**

Perhaps we can estimate the $p_i$’s.

---


What did we learn?

Condorcet

So we should give up one-person-one-vote?...

Ruth Ben-Yashar

Well in practice we might not know the $p_i$’s. In which case might be better to go with majority.

Eyal Baharad

Perhaps we can estimate the $p_i$’s.

Manon Revel

Even if we do know the $p_i$’s, we can still do pretty well by using the majority rule...


What did we learn?

Condorcet

So we should give up one-person-one-vote?...

Ruth Ben-Yashar

Well in practice we might not know the $p_i$'s. In which case might be better to go with majority.

Eyal Baharad

Perhaps we can estimate the $p_i$’s.

Manon Revel

Even if we do know the $p_i$’s, we can still do pretty well by using the majority rule...

over a carefully selected subset of the voters.


This is all fine and dandy

But what if we have more than two alternatives?
2. Voting Rules as Maximum Likelihood Estimators (MLEs)
More than two alternatives?

**Condorcet**

Everyone knows that when there are more than two alternatives all hell breaks loose.

Ordering alternatives by pairwise majority contests can lead to cycles.

\[ \sigma_1 : a \succ b \succ c \]
\[ \sigma_2 : b \succ c \succ a \]
\[ \sigma_3 : c \succ a \succ b. \]

\[ F(\sigma_1, \sigma_2, \sigma_3) =? \]
More than two alternatives?

**Condorcet**
Everyone knows that when there are more than two alternatives all hell breaks loose.

Ordering alternatives by pairwise majority contests can lead to cycles.

**Borda**
Sure. But, mon ami, there are other voting rules out there…

\[
\sigma_1 : a \succ b \succ c
\]
\[
\sigma_2 : b \succ c \succ a
\]
\[
\sigma_3 : c \succ a \succ b.
\]

\[F(\sigma_1, \sigma_2, \sigma_3) = ?\]
Everyone knows that when there are more than two alternatives all hell breaks loose.

Ordering alternatives by pairwise majority contests can lead to cycles.

Sure. But, mon ami, there are other voting rules out there...

Borda! My arch-nemesis!

\[ \sigma_1 : a \succ b \succ c \]
\[ \sigma_2 : b \succ c \succ a \]
\[ \sigma_3 : c \succ a \succ b. \]

\[ F(\sigma_1, \sigma_2, \sigma_3) = ? \]
The model

agents, aka voters \( N = \{1, \ldots, n\} \)

\( n \) typically assumed odd

alternatives \( A = \{a, b, c, \ldots\} \), with \( |A| = m \)

agent \( i \)'s preference \( \sigma_i \), linear order on \( A \)

write as words: \( \sigma_i = (a \succ b \succ c \succ \ldots) \sim \sigma_i = (abc\ldots) \)

correct order \( \sigma^* \)

prob. that \( i \) gets \( \sigma_i \) \( P[\sigma_i | \sigma^*] \in [0,1] \)

typically said to be given by a 'noise model', of which more shortly

profile of votes \( \sigma = (\sigma_1, \ldots, \sigma_n) \)

social choice rule \( F(\sigma) \mapsto A', \text{ with } \emptyset \subset A' \subset A \)

social preference rule \( F(\sigma) \mapsto \sigma, \text{ where } \sigma \text{ is a linear order on } A \)
Think of $\sigma_i$ as an independent ‘noisy estimate’ of the true ranking $\sigma^*$. This process is described by a noise model that determines $P[\sigma_i \mid \sigma^*]$. Typically, though, we don’t have access to the truth. We only see the individual estimates $\sigma_i$. But if we know the noise model we can get a good idea of what caused them. Using Bayes’ theorem!
Think of $\sigma_i$ as an independent ‘noisy estimate’ of the true ranking $\sigma^*$. This process is described by a noise model that determines $\mathbb{P}[\sigma_i | \sigma^*]$. Typically, though, we don’t have access to the truth. We only see the individual estimates $\sigma_i$. Using Bayes’ theorem!
Think of $\sigma_i$ as an independent 'noisy estimate' of the true ranking $\sigma^*$.

This process is described by a noise model that determines $P[\sigma_i | \sigma^*]$. Typically, though, we don’t have access to the truth. We only see the individual estimates $\sigma_i$.

But if we know the noise model we can get a good idea of what caused them.
H. Peyton Young

Think of $\sigma_i$ as an independent ‘noisy estimate’ of the true ranking $\sigma^*$. This process is described by a noise model that determines $P[\sigma_i \mid \sigma^*]$.

Arkadii Slinko

Typically, though, we don’t have access to the truth. We only see the individual estimates $\sigma_i$.

Edith Elkind

But if we know the noise model we can get a good idea of what caused them. Using Bayes’ theorem!
Bayes’ Theorem

**Theorem: Bayes’ Theorem**

\[ P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \]
Bayes’ Theorem

Theorem: Bayes’ Theorem

\[ P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \]
Bayes’ Theorem

**Theorem: Bayes’ Theorem**

\[
P[A | B] = \frac{P[B | A] \cdot P[A]}{P[B]}.
\]

Grant Sanderson (aka 3blue1brown)

You can see a really nice video about Bayes’ Theorem on my YouTube channel.

https://www.youtube.com/watch?v=HZGCoVF3YvM
Finding the likeliest ranking

Rev. Thomas Bayes

Apply my theorem to calculate the likelihood that some $\sigma$ is the true ranking—given the voting profile $\sigma = (\sigma_1, \ldots, \sigma_n)$. 

$$P[\sigma | \sigma] = \frac{P[\sigma | \sigma] \cdot P[\sigma]}{P[\sigma]}$$

$$= \left( \prod_{i \in N} P[\sigma_i | \sigma] \right) \cdot \frac{P[\sigma]}{P[\sigma]}$$
Finding the likeliest ranking

Apply my theorem to calculate the likelihood that some $\sigma$ is the true ranking—given the voting profile $\sigma = (\sigma_1, \ldots, \sigma_n)$.

Assume that, in the absence of any additional information, all rankings $\sigma$ and profiles $\sigma$ are equally likely to occur.

$$P(\sigma | \sigma) = \frac{P(\sigma | \sigma) \cdot P(\sigma)}{P(\sigma)}$$

$$= \left( \prod_{i \in N} P(\sigma_i | \sigma) \right) \cdot \frac{P(\sigma)}{P(\sigma)}$$

$$= \left( \prod_{i \in N} P(\sigma_i | \sigma) \right) \cdot c.$$
Finding the likeliest ranking

Rev. Thomas Bayes

Apply my theorem to calculate the likelihood that some \( \sigma \) is the true ranking—given the voting profile \( \sigma = (\sigma_1, \ldots, \sigma_n) \).

Assume that, in the absence of any additional information, all rankings \( \sigma \) and profiles \( \sigma \) are equally likely to occur.

Then finding the likeliest ranking, given \( \sigma \), is equivalent to finding \( \sigma \) that maximizes \( \prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] \).

\[
\mathbb{P}[\sigma | \sigma] = \frac{\mathbb{P}[\sigma | \sigma] \cdot \mathbb{P}[\sigma]}{\mathbb{P}[\sigma]}
= \left( \prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] \right) \cdot \frac{\mathbb{P}[\sigma]}{\mathbb{P}[\sigma]}
= \left( \prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] \right) \cdot c.
\]
Finding the likeliest ranking

Rev. Thomas Bayes

Apply my theorem to calculate the likelihood that some $\sigma$ is the true ranking—given the voting profile $\sigma = (\sigma_1, \ldots, \sigma_n)$.

Assume that, in the absence of any additional information, all rankings $\sigma$ and profiles $\sigma$ are equally likely to occur.

Then finding the likeliest ranking, given $\sigma$, is equivalent to finding $\sigma$ that maximizes $\prod_{i \in N} P[\sigma_i | \sigma]$.

\[
P[\sigma | \sigma] = \frac{P[\sigma | \sigma] \cdot P[\sigma]}{P[\sigma]} = \left( \prod_{i \in N} P[\sigma_i | \sigma] \right) \cdot \frac{P[\sigma]}{P[\sigma]}
\]

\[
= \left( \prod_{i \in N} P[\sigma_i | \sigma] \right) \cdot c.
\]

We want to find:

$$\arg\max_{\sigma} \prod_{i \in N} P[\sigma_i | \sigma]$$
Recovering the likeliest ranking via a voting rule

H. Peyton Young

Recall, though, that we can aggregate rankings using a voting rule.

Wouldn’t it be cool if the aggregated result turned out to be the same as the most likely ranking?

I very much like where this is going.

Edith Elkind

Will depend on the voting rule. And the noise model.

\[ \sigma^* \]

\[ F(\sigma) \]

\[ \sigma_1 \quad \sigma_2 \quad \ldots \quad \sigma_n \]
H. Peyton Young

Recall, though, that we can aggregate rankings using a voting rule.

Condorcet

Wouldn’t it be cool if the aggregated result turned out to be the same as the most likely ranking?
Recalling, though, that we can aggregate rankings using a voting rule.

**H. Peyton Young**

Wouldn’t it be cool if the aggregated result turned out to be the same as the most likely ranking?

I very much like where this is going.

**Condorcet**

Noise model:

$$\sigma^* 
\sigma_1 \rightarrow \sigma_2 \rightarrow \ldots \rightarrow \sigma_n$$

Voting rule:

$$F(\sigma)$$
Recovering the likeliest ranking via a voting rule

Recall, though, that we can aggregate rankings using a voting rule.

Wouldn’t it be cool if the aggregated result turned out to be the same as the most likely ranking?

I very much like where this is going.

Will depend on the voting rule. And the noise model.
Voting rules as maximum likelihood estimators (MLEs)

Definition: Elkind & Slinko (2016)

A social preference function $F$ is a maximum likelihood estimator (MLE) for a noise model that generates probabilities $P[\sigma_i | \sigma]$, if for every $n \in \mathbb{N}_{>0}$ and voting profile $\sigma = (\sigma_1, \ldots, \sigma_n)$, it holds that:

$$F(\sigma) = \arg\max_{\sigma} \prod_{i \in N} P[\sigma_i | \sigma].$$

We want to ask now

Under what noise models do known voting rules end up being MLEs?
agents, aka voters $N = \{1, \ldots, n\}$

$n$ typically assumed odd

alternatives $A = \{a, b, c, \ldots\}$, with $|A| = m$

agent $i$’s preference $\sigma_i$, linear order on $A$

we write linear orders as words: $a \succ b \succ c \succ \cdots \sim abc\ldots$

Kendall-tau distance $d_\tau(\sigma_i, \sigma_j) = |\{(x, y) \in A \times A | x \succ_i y \text{ and } y \succ_j x\}|$

that is, the number of pairwise disagreements between $\sigma_i$ and $\sigma_j$

e.g., $d_\tau(abc, cab) = 2$
agents, aka voters \(\{1, \ldots, n\}\)

\(n\) typically assumed odd

alternatives \(\{a, b, c, \ldots\}\), with \(|A| = m\)

agent \(i\)’s preference \(\sigma_i\), linear order on \(A\)

we write linear orders as words: \(a \succ b \succ c \succ \cdots \mapsto abc\ldots\)

Kendall-tau distance \(d_\tau\)

\[d_\tau(\sigma_i, \sigma_j) = |\{(x, y) \in A \times A \mid x \succ_i y \text{ and } y \succ_j x\}|\]

that is, the number of pairwise disagreements between \(\sigma_i\) and \(\sigma_j\)

e.g., \(d_\tau(abc, cab) = |\{(a, c), (b, c)\}| = 2\)
The Mallows noise model

Definition: Mallows (1957)

If $\sigma^*$ is the true ranking over $m$ alternatives, then for $\phi \in (0, 1)$ the probability of getting a ranking $\sigma$ is:

$$P[\sigma | \sigma^*] = \frac{1}{Z} \phi^{d_\tau(\sigma, \sigma^*)},$$

where $Z = (1 + \phi)(1 + \phi + \phi^2) \ldots (1 + \phi + \cdots + \phi^{m-1})$.

Observation: Note about the Mallows model

- probability decays as $\sigma$ is further away from $\sigma^*$
- for $\phi = 1$ we get the uniform distribution
- for $\phi = 0$ we get $P[\sigma | \sigma^*] = \begin{cases} 1, & \text{if } \sigma = \sigma^*, \\ 0, & \text{otherwise.} \end{cases}$
The MLE under the Mallows noise model

Clive L. Mallows
Plug the parameters of my noise model into the MLE expression.

\[ \prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] = \prod_{i \in N} \left( \frac{1}{Z} \phi^{d_\tau}(\sigma_i, \sigma) \right) \]
The MLE under the Mallows noise model

Clive L. Mallows
Plug the parameters of my noise model into the MLE expression.

Note that we can ignore the $Z'$s
The MLE under the Mallows noise model

Clive L. Mallows
Plug the parameters of my noise model into the MLE expression.

H. Peyton Young
The optimal rankings end up being those that minimize the sum of Kendall tau distances to all rankings in $\sigma$.

$\prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] = \prod_{i \in N} \left( \frac{1}{Z} \phi^{d_{\tau}(\sigma_i, \sigma)} \right)$

$\propto \prod_{i \in N} \phi^{d_{\tau}(\sigma_i, \sigma)}$  

$= \phi \sum_{i \in N} d_{\tau}(\sigma_i, \sigma)$

Since $\phi \in (0, 1)$, we have that:

$\arg\max_{\sigma} \prod_{i \in N} \mathbb{P}[\sigma_i | \sigma] = \arg\min_{\sigma} \sum_{i \in N} d_{\tau}(\sigma_i, \sigma)$
The MLE under the Mallows noise model

Clive L. Mallows
Plug the parameters of my noise model into the MLE expression.

Note that we can ignore the $Z’s$.

H. Peyton Young
The optimal rankings end up being those that minimize the sum of Kendall tau distances to all rankings in $\sigma$.

John G. Kemeny
Hey that’s my rule!

\[
\prod_{i \in N} \mathbb{P}[^{\sigma_i} | ^{\sigma}] = \prod_{i \in N} \left( \frac{1}{Z} \phi^{d_{\tau}(\sigma_i, \sigma)} \right)
\propto \prod_{i \in N} \phi^{d_{\tau}(\sigma_i, \sigma)}
= \phi \sum_{i \in N} d_{\tau}(\sigma_i, \sigma)
\]

Since $\phi \in (0, 1)$, we have that:

\[
\arg\max_{\sigma} \prod_{i \in N} \mathbb{P}[^\sigma_i | ^\sigma] = \arg\min_{\sigma} \sum_{i \in N} d_{\tau}(\sigma_i, \sigma)
= F_{Kemeny}(\sigma)
\]
We’ve just shown the following result

**Theorem: Young (1988)**

The Kemeny social preference rule defined as:

\[ F_{\text{Kemeny}}(\sigma) = \arg\min_{\sigma} \sum_{i \in N} d_{\tau}(\sigma_i, \sigma) \]

is an MLE for the Mallows noise model.

---

Other rules?

What about other rules?
The model + scoring rules

agents, aka voters \( N = \{1, \ldots, n\} \)

\( n \) typically assumed odd

alternatives \( A = \{a, b, c, \ldots\} \), with \( |A| = m \)

agent \( i \)'s preference \( \sigma_i \), linear order on \( A \)

we write linear orders as words: \( a \succ b \succ c \succ \cdots \sim abc \cdots \)

profile of votes \( \sigma = (\sigma_1, \ldots, \sigma_n) \)

social choice rule \( F(\sigma) \mapsto A' \), with \( \emptyset \subset A' \subseteq A \)

social preference rule \( F(\sigma) \mapsto \sigma \), where \( \sigma \) is a linear order on \( A \)
The model + scoring rules

agents, aka voters \( N = \{1, \ldots, n\} \)
n typically assumed odd

alternatives \( A = \{a, b, c, \ldots\}, \text{with } |A| = m \)

agent \( i \)'s preference \( \sigma_i, \text{linear order on } A \)

we write linear orders as words: \( a \succ b \succ c \succ \cdots \Rightarrow abc \ldots \)

profile of votes \( \sigma = (\sigma_1, \ldots, \sigma_n) \)

social choice rule \( F(\sigma) \mapsto A', \text{with } \emptyset \subset A' \subset A \)

social preference rule \( F(\sigma) \mapsto \sigma, \text{where } \sigma \text{ is a linear order on } A \)

position of \( x \in A \) in \( \sigma_i \) \( \text{pos}_i(x) \in \{1, \ldots, m\}, \text{where } x \text{ shows up in } \sigma_i \)

scoring vector \( s \)
\( s = (s_1, \ldots, s_m), \text{for } s_i \in \mathbb{R}_{\geq 0} \)

with \( s(1) \geq \cdots \geq s(m) \), at least one inequality strict

score of \( x \in A \) in \( \sigma_i \) \( s_i(x) = s(\text{pos}_i(x)) \)

scoring rule \( F_s \)
\( F_s(\sigma) = \arg\max_{x \in A} \sum_{i \in N} s_i(x) \)
The Borda rule

Mes amis! Let’s give candidates scores from 0 upwards.

\[ s_{Borda} = (m - 1, \ldots, 1, 0) \]

\[ \sigma = (abcd, bcad, adcb) \]

Total Borda scores:

- \( a \mapsto 3 + 1 + 3 = 7 \)
- \( b \mapsto 2 + 3 + 0 = 5 \)
- \( c \mapsto 1 + 2 + 1 = 4 \)
- \( d \mapsto 0 + 0 + 2 = 2 \)

Borda winner is \( a \).
Finding the likeliest alternative

Note that now we want to hit on the likeliest alternative.

Rev. Thomas Bayes
Note that now we want to hit on the likeliest alternative. Not the likeliest ranking.
Finding the likeliest alternative

Rev. Thomas Bayes

Note that now we want to hit on the likeliest alternative. Not the likeliest ranking. But the same argument applies.

We want to find:

$$\arg\max_{x \in A} \prod_{i \in N} \mathbb{P}[\sigma_i \mid x].$$
Rev. Thomas Bayes

Note that now we want to hit on the likeliest alternative.

Not the likeliest ranking.

But the same argument applies.

A noise model in this case gives the probability of seeing $\sigma_i$ given alternative $x$.

We want to find:

$$\arg\max_{x \in A} \prod_{i \in N} \mathbb{P}[\sigma_i | x].$$
The further down the ranking, the less likely

Vincent Conitzer
If $a$ is the true alternative, then it should be more likely to show up first than second.
The further down the ranking, the less likely

Vincent Conitzer
If $a$ is the true alternative, then it should be more likely to show up first than second.

And more likely to show up second than third, etc.
The further down the ranking, the less likely

**Vincent Conitzer**

If $a$ is the true alternative, then it should be more likely to show up first than second.

And more likely to show up second than third, etc.

\[
\begin{align*}
\mathbb{P}[abc \mid a] &= \mathbb{P}[acb \mid a] \propto 2^2 \\
\mathbb{P}[bac \mid a] &= \mathbb{P}[cab \mid a] \propto 2^1 \\
\mathbb{P}[bca \mid a] &= \mathbb{P}[cba \mid a] \propto 2^0
\end{align*}
\]
The further down the ranking, the less likely

Vincent Conitzer
If $a$ is the true alternative, then it should be more likely to show up first than second.

And more likely to show up second than third, etc.

Tuomas Sandholm
Let’s say twice more likely.

\[
\begin{align*}
P[abc \mid a] &= P[acb \mid a] \propto 2^2 \\
P[bac \mid a] &= P[cab \mid a] \propto 2^1 \\
P[bc a \mid a] &= P[cb a \mid a] \propto 2^0
\end{align*}
\]
Definition: Conitzer & Sandholm (2005)

If \( x^* \) is the true alternative, then the probability of getting a ranking \( \sigma_i \) on \( m \) alternatives is:

\[
P[\sigma \mid x^*] \propto 2^{m - \text{pos}_i(x^*)}.
\]
A position-based noise model

**Definition: Conitzer & Sandholm (2005)**

If $x^*$ is the true alternative, then the probability of getting a ranking $\sigma_i$ on $m$ alternatives is:

$$P[\sigma | x^*] \propto 2^{m-\text{pos}_i(x^*)}.$$ 

**Observation**

- the terms can be normalized to get proper probabilities, but the normalization factor doesn’t matter for the general argument

---

The MLE under the Conitzer-Sandholm noise model

Vincent Conitzer
Plug the parameters of this noise model into the MLE expression.

\[
\prod_{i \in N} \mathbb{P}[\sigma_i \mid x] \propto \prod_{i \in N} \left(2^{m - \text{pos}_i(x)}\right) \\
= 2^{\sum_{i \in N} (m - \text{pos}_i(x))}
\]
The MLE under the Conitzer-Sandholm noise model

Vincent Conitzer
Plug the parameters of this noise model into the MLE expression.

Tuomas Sandholm
The optimal alternatives those that maximize the sum of position-based scores to rankings in $\sigma$.

$$\prod_{i \in N} \mathbb{P}[\sigma_i \mid x] \propto \prod_{i \in N} \left(2^{m - \text{pos}_i(x)}\right) = 2 \sum_{i \in N} \left(m - \text{pos}_i(x)\right)$$
The MLE under the Conitzer-Sandholm noise model

Vincent Conitzer
Plug the parameters of this noise model into the MLE expression.

Tuomas Sandholm
The optimal alternatives those that maximize the sum of position-based scores to rankings in $\sigma$.

Hey that’s my rule!

$$\prod_{i \in N} \mathbb{P}[\sigma_i \mid x] \propto \prod_{i \in N} \left(2^{m - \text{pos}_i(x)}\right) = 2\sum_{i \in N} (m - \text{pos}_i(x))$$
We’ve just shown the following result:

**Theorem: Conitzer & Sandholm (2005)**

The Borda social choice rule* defined as:

\[ F_{\text{Borda}}(\sigma) = \arg\max_{x \in A} \sum_{i \in N} (m - \text{pos}_i(x)) \]

is an MLE for the Conitzer-Sandholm noise model.

* Output is a set of alternatives, not a ranking.

---

For scoring rules, more generally

This idea works more generally, for any scoring rule.

\[ P[\sigma_i \mid x^*] \propto 2^{s_i(x^*)} \]

For scoring rules, more generally

Vincent Conitzer
This idea works more generally, for any scoring rule.

Tuomas Sandholm
Just take the exponent to be the score of $x^*$ in $\sigma_i$.

$$P[\sigma_i | x^*] \propto 2^{s_i(x^*)}$$

For scoring rules, more generally

**Vincent Conitzer**
This idea works more generally, for any scoring rule.

**Tuomas Sandholm**
Just take the exponent to be the score of $x^*$ in $\sigma_i$.

**Vincent Conitzer**
We also this doesn’t work for certain (types of) rules! There are structural properties that mean there’s no noise model that makes these rule optimal.

\[
\mathbb{P}[\sigma_i \mid x^*] \propto 2^{s_i(x^*)}
\]

---

For scoring rules, more generally

Vincent Conitzer
This idea works more generally, for any scoring rule.

Tuomas Sandholm
Just take the exponent to be the score of \( x^* \) in \( \sigma_i \).

Vincent Conitzer
We also this doesn’t work for certain (types of) rules! There are structural properties that mean there’s no noise model that makes these rule optimal.

Tuomas Sandholm
Check out our paper!

\[
\mathbb{P}[\sigma_i \mid x^*] \propto 2^{s_i(x^*)}
\]

Summing up + an existential question

Vincent Conitzer

Much ink has been spilled on these topics.


Summing up + an existential question

Vincent Conitzer

Much ink has been spilled on these topics.

Tuomas Sandholm

That is, seeing voting rules as techniques for estimating the truth, given the noise model.


Summing up + an existential question

Vincent Conitzer
Much ink has been spilled on these topics.

Tuomas Sandholm
That is, seeing voting rules as techniques for estimating the truth, given the noise model.

Marcus Pivato
In other words: give me the voting rule, and I give you the noise model for which this rule is optimal.

---


Summing up + an existential question

Vincent Conitzer
Much ink has been spilled on these topics.

Tuomas Sandholm
That is, seeing voting rules as techniques for estimating the truth, given the noise model.

Marcus Pivato
In other words: give me the voting rule, and I give you the noise model for which this rule is optimal.

This is completely backwards.

References:
Summing up + an existential question

Vincent Conitzer
Much ink has been spilled on these topics.

Tuomas Sandholm
That is, seeing voting rules as techniques for estimating the truth, given the noise model.

Marcus Pivato
In other words: give me the voting rule, and I give you the noise model for which this rule is optimal.

This is completely backwards.

We should begin with a model of voter belief formation, and then design a statistical instrument to extract the truth from these beliefs as accurately as possible.

