Harvard CS 1360, Spring 2026
Instructor: Ariel Procaccia
Scribe: Taig Singh, Jakob de Raaij

The Epistemic Approach to Voting

—

Lecture 6

In some scenarios, the purpose of voting is not to aggregate subjective opinions but to collective try to uncover a hidden ground truth. This is known as the *epistemic approach*.

1 The Condorcet Jury Theorem

As an example of the epistemic approach, consider the jury in a criminal trial: There is an unknown ground-truth about whether the defendant is guilty or not. Each member of the jury makes the correct judgment with some probability. The hope is that if the jury is large enough, the members that decide correctly will outnumber the members that decide incorrectly, so that the majority vote among the judges will make the correct judgment.

Theorem 1 (Condorcet Jury Theorem (1785))**.** *Suppose that there are* 2 *alternatives (a correct and an incorrect one) and* n *voters, each of whom votes independently votes for the correct alternative with probability* $p > 1/2$*. Then, the probability that a majority of votes are correct approaches* 1 *as* $n \to \infty$*.*

This theorem provides a formal justification for the use of majority voting to uncover the underlying ground truth when comparing two options: Given that n is large and $p > 1/2$, the majority vote is almost certainly correct.

Proof. The result follows directly from the weak law of large numbers. Intuitively, it states that the observed average of sufficiently many random variables converges to the expectation μ.

> *Weak Law of Large Numbers.* Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. random variables with expectation μ. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any $\epsilon > 0$,
>
> $$\lim_{n\to\infty} \Pr\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1.$$

To apply this law to our context, let X_i be the outcome of voter i's vote, taking on 1 if they vote correctly and 0 if they do not. Then, the majority of n voters is correct if $\overline{X}_n > \frac{1}{2}$. It holds that $\mu = \mathbb{E}[X_i] = p$ for all voters i, so taking $\epsilon = p - \frac{1}{2}$, it follows from the law that the majority vote will converge to the correct alternative as $n \to \infty$. □

2 The Condorcet noise model

We now consider the case of 3 or more alternatives, in which the correct approach is less straight-forward.

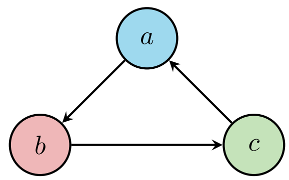
Definition 1 (Condorcet Noise Model)**.** Assume that there is a true ranking of alternatives. Under the *Condorcet noise model*, each voter evaluates every pair of alternatives independently and gets each comparison right with probability $p > 1/2$. The results are tallied in a *pairwise comparison matrix*: For any pair of alternatives a, b, the entry $M_{a,b}$ is the number of voters that rank a above b.

Condorcet proposed to find the true ranking by taking the majority opinion for each comparison. If a cycle forms, he suggested to find the "most probably" ranking by "successively deleting the comparisons that have the least plurality".

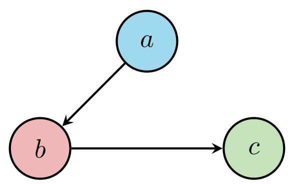
Example 1 (Condorcet's proposal for three alternatives)**.** Consider three alternatives a, b, c with the following pairwise comparison results.

	a	b	c
a	–	8	6
b	5	–	11
c	7	2	–

For each comparison between two alternatives, we display which alternative is preferred in the *pairwise majority graph* by drawing a directed edge from the alternative winning in the pairwise comparison to the alternative loosing in the pairwise comparison.



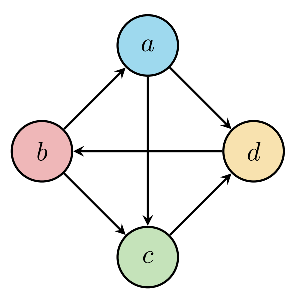
If there was no cycle, we could go with the majority opinion on every pairwise comparison by using the topological order of the vertices as our ranking. However, in this example, we have a cycle — the final ranking will need to disagree with one of the three pairwise comparisons. In the spirit of Condorcet's proposal, we remove the comparison that has the smallest margin; in this case this means removing $c \succ a$ (7 votes in favor 6 votes against).



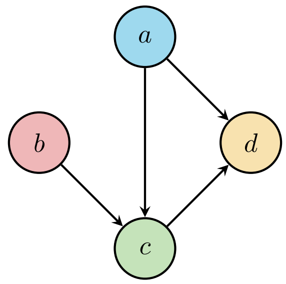
Ranking the vertices in topological order, we get $a \succ b \succ c$. This ranking is in agreement with all pairwise comparison majorities except $c \succ a$.

Example 2 (Condorcet's proposal for more than three alternatives)**.** Consider four alternatives a, b, c, d with the following pairwise comparison results and pairwise majority graph

	a	b	c	d
a	–	12	15	17
b	13	–	16	11
c	10	9	–	18
d	8	14	7	–

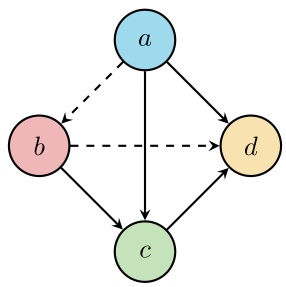


Again, we have cycles in our ranking. The order of strength of these comparisons (by plurality) is $c \succ d$, $a \succ d$, $b \succ c$, $a \succ c$, $d \succ b$, and $b \succ a$. One possible way to implement Condorcet's proposal is to start deleting these edges in the pairwise majority graph from weakest to strongest until there is no cycle left. Doing this, we first delete $b \succ a$, removing the cycle (a, b, d). However, this still leaves the cycle (b, c, d), so we next delete $d \succ b$.



Now, there are no cycles left. However, we can no longer compare a and b in the graph; they could both be the highest-ranked alternative without violating any of the edges.

To not run into this issue, let's consider a different interpretation of Condorcet's proposal: What if we *reversed* the edges in order of weakest to strongest, until no cycle remains? Thus, in a similar manner to before, we first reverse $b \succ a$. There still remains the cycle (b, c, d), so we next reverse $d \succ b$, to find the graph now without cycles:



We get the ranking $a \succ b \succ c \succ d$. Should this be our best guess for the correct ranking? The answer may be no: Adding up how many observed pairwise comparisons agree with this ranking, we get 89 votes. However, if we only reversed $d \succ b$ (and not $b \succ a$), we would have gotten $b \succ a \succ c \succ d$, agreeing with 90 comparisons! Reversing $b \succ a$ was not necessary to get rid of all cycles; doing that unnecessarily caused a violation of another pairwise comparison.

Frustrated with these shortcomings of any known concrete implementation of Condorcet's proposal, Isaac Todhunter (1820–1884) stated:

> "The obscurity and self-contradiction are without any parallel, so far as our experience of mathematical works extends ... no amount of examples can convey an adequate impression of the evils."

It was only a century later that an implementation of Condorcet's proposal was found that seems to have commonly been accepted as the correct interpretation:

Definition 2 (Young's solution)**.** Use the ranking under which the observed pairwise comparison matrix is most likely. That is, choose ranking π to maximize $\Pr[M|\pi]$ where M denote the pairwise comparison matrix.

Example 3 (Young's solution)**.** Let's reconsider the pairwise comparison matrix with 25 voters for four alternatives a, b, c, d from Example 2:

	a	b	c	d
a	–	12	15	17
b	13	–	16	11
c	10	9	–	18
d	8	14	7	–

Suppose the true ranking is $\pi = a \succ b \succ c \succ d$. Under the Condorcet Noise Model (recall, Definition 1), we get that the probability of observing M is

$$\begin{aligned}\Pr[M|\pi] = &\underbrace{\binom{25}{12}p^{12}(1-p)^{13}}_{a\succ b} \cdot \underbrace{\binom{25}{15}p^{15}(1-p)^{10}}_{a\succ c} \cdot \underbrace{\binom{25}{17}p^{17}(1-p)^{8}}_{a\succ d} \\ &\cdot \underbrace{\binom{25}{16}p^{16}(1-p)^{9}}_{b\succ c} \cdot \underbrace{\binom{25}{11}p^{11}(1-p)^{14}}_{b\succ d} \cdot \underbrace{\binom{25}{18}p^{18}(1-p)^{7}}_{c\succ d}.\end{aligned}$$

For each pair of alternatives, the corresponding term represents the probability of observing the preferences in M for this pair, if π is the true ranking. For example, for the alternatives a and b, the term $\binom{25}{12}p^{12}(1-p)^{13}$ is exactly the probability that 12 voters rank a above b and 13 rank b above a when the true ranking is $a \succ b$, as in π.

Similarly, if the true ranking was $\pi = b \succ a \succ c \succ d$, then

$$\begin{aligned}\Pr[M|\pi] = &\underbrace{\binom{25}{13}p^{13}(1-p)^{12}}_{b\succ a} \cdot \underbrace{\binom{25}{15}p^{15}(1-p)^{10}}_{a\succ c} \cdot \underbrace{\binom{25}{17}p^{17}(1-p)^{8}}_{a\succ d} \\ &\cdot \underbrace{\binom{25}{16}p^{16}(1-p)^{9}}_{b\succ c} \cdot \underbrace{\binom{25}{11}p^{11}(1-p)^{14}}_{b\succ d} \cdot \underbrace{\binom{25}{18}p^{18}(1-p)^{7}}_{c\succ d}.\end{aligned}$$

Note that all terms are the same as before except the first one, which is smaller: $\binom{25}{12} = \binom{25}{13}$ but $p^{12}(1-p)^{13} < p^{13}(1-p)^{12}$ (because $p > 1/2$). Thus, it is more likely that the true ranking is $b \succ a \succ c \succ d$ rather than $a \succ b \succ c \succ d$.

It is not a coincidence that Young's solution let to the ranking $b \succ a \succ c \succ d$, for which we observed that it agrees with 90 individual pairwise comparisons, over $a \succ b \succ c \succ d$, which agrees with 89 individual pairwise comparisons. Since $\binom{n}{k} = \binom{n}{n-k}$, the binomial coefficients in $\Pr[M|\pi]$ are going to be the same for all π, so only the exponents of p and $1-p$ matter. However, these exponents are exactly how many individual pairwise comparisons agreed and disagreed with π, respectively, for each pair of alternatives! We thus get that

$$\Pr[M|\pi] \propto p^{\#\text{ agree}}(1-p)^{\#\text{ disagree}}.$$

Since $p > 1/2$, Young's solution will always output the ranking that agrees with as many of the voters' pairwise comparisons as possible.

3 The Mallows model

In the Condorcet Noise Model, we assumed that every voter gets each pairwise comparison right independently with probability p. This allows (an actually, makes it very likely for a large number of candidates or p close to $1/2$) that a voter reports pairwise preferences that include a cycle, $a \succ b \succ c \succ c$, so that their preferences adhere by no consistent ranking. This seems fairly unrealistic. Thus, we now consider a related model in which any voter is are always guaranteed to have a consistent ranking. The solutions we obtain will be very similar to Young's model.

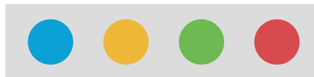
Definition 3 (Kendall Tau Distance)**.** The *Kendall tau distance* between two rankings σ and σ' is defined as

$$d_{KT}(\sigma, \sigma') = |\{(a, b) : a \succ_\sigma b \text{ and } b \succ_{\sigma'} a\}|$$

i.e., the number of pairs of elements on which the two rankinsgs disagree. Equivalently, this is the minimum number of swaps of neighboring elements needed to convert one ranking into the other, sometimes referred to as "bubble sort distance."

Example 4 (Kendall Tau Distance)**.** Consider two rankings of four alternatives:

$\sigma = a \succ b \succ c \succ d$



$\sigma' = a \succ d \succ c \succ b$

The Kendall tau distance between these two rankings is 3, as there is disagreements on 3 pairs: (b, c), (b, d), and (c, d). Equivalently, if we wanted to convert σ into σ' via (buble-sort like) swaps of neighboring elements, we optimally need three flips: $c \leftrightarrow d$, $b \leftrightarrow d$, and then $b \leftrightarrow c$.

We will now use this distance to define our probabilistic model.

Definition 4 (Mallows Model)**.** Assume that there is a true ranking of alternatives π. Under the *Mallows model* with parameter $\phi \in (0, 1]$, the probability that a voter has ranking σ is:

$$\Pr[\sigma|\pi] = \frac{\phi^{d_K(\sigma,\pi)}}{\sum_\tau \phi^{d_K(\tau,\pi)}}.$$

In the Mallows model, the probability of a ranking σ decreases exponentially with the number of disagreements that σ has with π. The denominator, $\sum_\tau \phi^{d_K(\tau,\pi)}$, is merely for normalization. Observe that all rankings are equally likely when $\phi = 1$, while the probability of observing ranking π goes to 1 as $\phi \to 0$.

The Mallows problem actually is equivalent to a 'rejection sampling' variant of the Condorcet noise model for $\phi = \frac{1-p}{p}$, where we sample a ranking from the Condorcet noise model until we obtain a ranking with no cycles. We won't formally prove this, but to gain some intuition on this, let's consider the extreme cases: If $p = 0.5$, voters decide their pairwise comparisons in the Condorcet noise model uniformly at random, which correspond to the case $\phi = 1$ where all rankings are equally likely. If $p \to 1$, voters obtain π with probability approaching 1, which matches what we observed for $\phi \to 0$.

In the spirit of Young's solution, we may now wonder: What is the probability of observing a voting profile $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_n)$, where voter i holds ranking σ_i, given that the true ranking is π? Applying the definition of Mallows model, we find that

$$\Pr[\boldsymbol{\sigma}|\pi] = \prod_{i \in N} \frac{\phi^{d_{KT}(\sigma_i,\pi)}}{Z_\phi} = \frac{\phi^{\sum_{i\in N} d_{KT}(\sigma_i,\pi)}}{(Z_\phi)^n}.$$

where $Z_\phi = \sum_\tau \phi^{d_{KT}(\tau,\pi)}$ is the normalizing constant in the denominator. Since the normalizing constant is the same for all possible underlying rankings π, we find that for any $\phi \in (0, 1)$, the π that maximizes the probability of observing $\boldsymbol{\sigma}$ is the π that has the smallest Kendall tau distance summed for all voters, $\sum_{i \in N} d_{KT}(\sigma_i, \pi)$.

Definition 5 (Kemeny rule)**.** The *Kemeny rule* returns the ranking π that minimizes the *Kemeny score*

$$\sum_{i \in N} d_{KT}(\sigma_i, \pi).$$

For $\phi \in (0, 1)$, this is equivalent to returning the ranking π that maximizes the MLE $\Pr[\boldsymbol{\sigma}|\pi]$ under the Mallows model.

Unfortunately, there is a big caveat to finding this ranking: computing it is intractable.

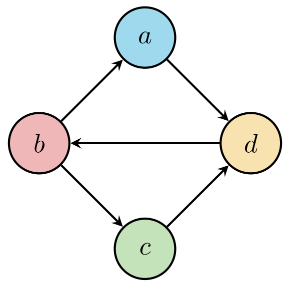
Theorem 2. *Deciding whether there exists a ranking π with Kemeny score at most k is **NP**-complete.*[1]

[1] In fact, determining the winner, i.e., the top-ranked candidate, of the Kemeny voting rule is believed to be even harder than **NP**, under common complexity-theoretic assumptions.

Proof Sketch. To show that this problem is **NP**-hard, we can reduce from the Minimum Feedback Arc Set problem:

> *Minimum Feedback Arc Set.* Given a directed graph $G = (V, E)$ and a number $\ell \in \mathbb{N}$, is it possible to delete ℓ edges from G in a way the make it acyclic?

To describe the reduction, we start from an instance of the Minimum Feedback Arc Set Problem: We are given a directed graph $G = (V, E)$ and threshold $\ell \in \mathbb{N}$. We treat the vertex set V as the alternatives and construct $2|E|$ voters with rankings over those alternatives. In particular, for each $(x, y) \in E$, create a pair of voters $v_{(x,y)}$ and $w_{(x,y)}$ that agree on $x \succ t$ but disagree on any other pair of alternatives. For example, for the graph below, we can construct a voting profile as shown:



$v_{b,a}$	$w_{b,a}$
b	d
a	c
c	b
d	a

$v_{a,c}$	$w_{a,c}$
a	d
c	b
b	a
d	c

$v_{d,b}$	$w_{d,b}$
d	c
b	a
a	d
c	b

$v_{b,c}$	$w_{b,c}$
b	d
c	a
a	b
d	c

$v_{c,d}$	$w_{c,d}$
c	b
d	a
a	c
b	d

For example, consider the edge (b, a). Voters $v_{b,a}$ and $w_{b,a}$ agree on $b \succ a$, but disagree on any other pairwise comparison.

Now, we claim that there is an acyclic subgraph that deletes k edges if and only if there is a ranking that disagrees with $2k$ pairwise comparison of individual voters beyond the "inevitable disagreement". By "inevitable disagreement" we mean that for any given pairwise comparison made by a ranking, within every pair of voters one voter will agree with the comparison and the other will disagree with it, except for the pair of voters corresponding to this pairwise comparison in the ranking. For example, if a ranking made the choice that $b \succ a$, then in the pair of voters constructed above, for each $(x, y) \in E$, one of $v_{x,y}$ and $w_{x,y}$ will agree while the other will disagree, except if $(x, y) = (b, a)$ or $(x, y) = (a, b)$, in which case both voters would agree with $b \succ a$ in the earlier case and disagree with both voters in the latter. For any pair of voters (x, y), we say that the ranking agrees with the pair of voters $v_{x,y}$, $w_{x,y}$ if $x \succ y$, else we say it disagrees with the pair. Thus, any ranking will have "inevitable disagreement" accounting for a Kemeny score of $\text{ID} = \binom{|V|}{2} \cdot |E| - |E|$, with an additional $2k$ for the k pairs of voters with which the ranking disagrees. Thus, the more precise claim is that there is an acyclic subgraph that deletes k edges if and only if there is a ranking with Kemeny score at most $\text{ID} + 2k$.

We will now justify the above statement: If you can delete k edges to obtain an acyclic subgraph, then we will take the ranking that is consistent with the remaining edges (that is, the vertices in topological order); this is always possible because the graph is acyclic. This ranking will agree with all the pairwise comparisons exceeding "inevitable disagreement," except for the pairs of voters that correspond to the k deleted edges, so we get a Kemeny score of at most $\text{ID} + 2k$. For the other direction, if there is a ranking that disagrees with at most $\text{ID} + 2k$ of the pairwise comparisons, we then delete the k edges of the graph that corresponds to the pairs with which the ranking disagrees, accounting for the Kemeny score exceeding "inevitable disagreement". Now, the remaining subgraph must be acyclic, because the edges that remain agree with all pairs of voters on the ranking we started from, which is acyclic. □

In practical instances, we can often compute the Kemeny score of a ranking with an integer linear program (ILP). While ILPs are intractable in the worst case, there are efficient solvers that can handle ILPs with thousands of variables.

To obtain the ILP, we let $x(a,b) = 1$ iff a is ranked above b, for any pair of alternatives a and b. We let $w(a,b) = |\{i \in N : a \succ_{\sigma_i} b\}|$ be the number of voters that prefer a to b. We denote by A the set of all alternatives. The ILP then is

$$\begin{aligned} \min \quad & \sum_{(a,b)\in A^2} x(a,b) \cdot w(b,a), \\ \text{s.t.} \quad & x(a,b) + x(b,a) = 1 \quad \forall a,b \in A \\ & x(a,b) + x(b,c) + x(c,a) \leq 2 \quad \forall a,b,c \in A \\ & x(a,b) \in \{0,1\} \quad \forall a,b \in A \end{aligned}$$

The objective function corresponds to the Kemeny score, i.e., the number of voters that disagree with each comparison $a \succ b$ if we decide to rank a and b that way. The first constraint ensures that the ranking is *complete*: we are choosing exactly one of $a \succ b$ or $b \succ a$. The second constraint makes sure that the ranking is *transitive*: there are no three alternatives such that $a \succ b \succ c \succ a$. It turns out that completeness and transitivity suffice to make sure that the $x(a,b)$ correspond to a valid ranking. The third constraint just makes sure that $x(a,b)$ is a binary variable.