Harvard CS 1360, Spring 2025
Instructor: Ariel Procaccia
Scribe: Taig Singh

The Epistemic Approach to Voting

—

Lecture 6

For Condorcet, the purpose of voting is not merely to balance subjective opinions. It is a collective quest for the truth. Enlightened voters try to judge which alternative best serves society. This is an arguable model of political elections, but there are certainly settings where the ground-truth assumption holds true. For example, consider the example of a criminal trial. There is a ground-truth about whether the defendant is guilty or not guilty, and the jurors are trying to uncover that truth.

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

Implication: This theorem provides a formal justification for the use of majority voting in truth-seeking contexts, reinforcing its reliability when n is large and $p > \frac{1}{2}$.

Proof. The result follows directly from the weak law of large numbers which is written as follows: Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. random variables with expectation μ. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1.$$

Now, applying 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 $E[X_i] = p$ for all voters i. Further, it follows that the majority is correct if $\overline{X}_n > \frac{1}{2}$. Then, taking $\epsilon = p - \frac{1}{2}$, it follows from the law that the majority vote will converge to the correct alternative as $n \to \infty$. □

We now consider the case of 3 or more alternatives.

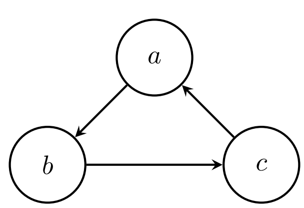
Definition 2 (The Condorcet Noise Model)**.** In this model, there is a true ranking of alternatives. Each voter evaluates every pair of alternatives independently and gets the comparison right with probability $p > \frac{1}{2}$. The results are tallied in a voting matrix, where there is a 1 if the voter thinks the row alternative is better than the column alternative, and a 0 otherwise.

Condorcet's proposal: Find the "most probable" ranking by taking the majority opinion for each comparison; if a cycle forms, "successively delete the comparisons that have the least plurality".

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

	a	b	c
a	–	8	6
b	5	–	11
c	7	2	–

Taking the majority opinion for each comparison creates the following ranking:

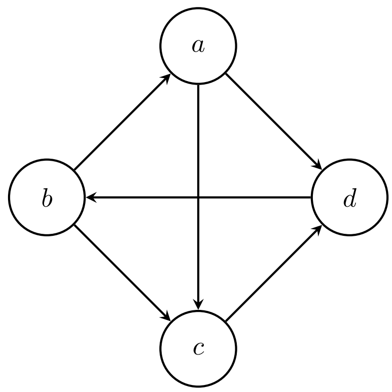


Since we have a cycle, we remove the comparison with the least plurality. In this case, we would remove $c \succ a$ (7 votes), and thus we get the ranking $a \succ b \succ c$.

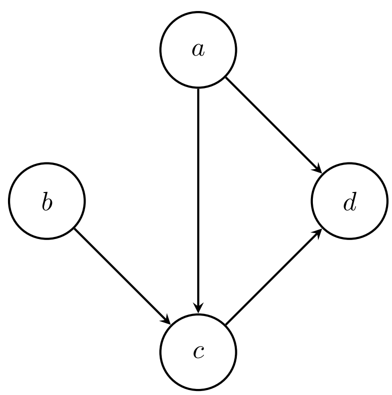
Example 2 (Condorcet's "Solution")**.** Consider the following voting matrix for four alternatives a, b, c, d:

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

Taking the majority opinion for each comparison creates the following ranking:

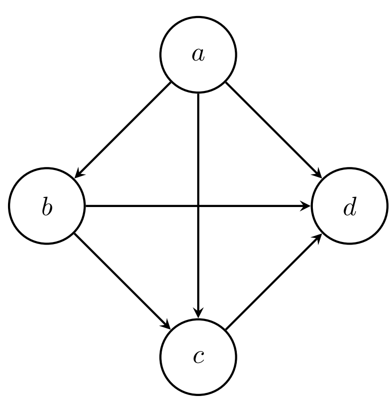


Note that we have cycles in our ranking. The order of strength of these comparisons (by plurality) is $c \succ d$, $a \succ d$, $b \succ c$, $a \succ c$, $d \succ b$, and $b \succ a$. Condorcet suggests deleting $b \succ a$, which still leaves a cycle. But, deleting $d \succ b$ creates ambiguity:



Should we rank a above b or b above a?

Did Condorcet mean that we should reverse the weakest comparisons? If we reverse $b \succ a$ we still have cycles, and if we then reverse $d \succ b$, we have no ambiguity:



We now get the ranking of $a \succ b \succ c \succ d$. But is this the correct ranking? Adding up the votes for how many voters support this ranking, we get 89 votes. However, if we initially only reversed $d \succ b$ (and not $b \succ a$), we would have gotten $b \succ a \succ c \succ d$ with a support of 90 votes!

Isaac Todhunter (1820–1884) critiqued Condorcet's method, stating: "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."

Definition 3 (Young's Solution)**.** Let M denote the matrix of votes and π the true ranking. Young proposes that we find the MLE for π given M, that is, the π that maximizes $P(M|\pi)$.

Example 3 (Young's Solution)**.** Consider the following voting matrix among 13 voters for four alternatives a, b, c:

	a	b	c
a	–	8	6
b	5	–	11
c	7	2	–

Suppose the true ranking is $a \succ_\pi b \succ_\pi c$. Then, we get that

$$P(M|\pi) = \binom{13}{8} p^8(1-p)^5 \cdot \binom{13}{6} p^6(1-p)^7 \cdot \binom{13}{11} p^{11}(1-p)^2$$

Here, the first term represents the probability that 8 voters rank a above b, the second term represents the probability that 6 voters rank a above c, and the third term represents the probability that 11 voters rank b above c. Similarly, if the true $a \succ_\pi c \succ_\pi b$ then

$$P(M|\pi) = \binom{13}{8} p^8(1-p)^5 \cdot \binom{13}{6} p^6(1-p)^7 \cdot \binom{13}{2} p^2(1-p)^{11}$$

Note that the first and second terms are the same as before, but the third term is necessarily smaller. Indeed, $\binom{13}{2} = \binom{13}{11}$ but $p^2(1-p)^{11} < p^{11}(1-p)^2$ because $p > \frac{1}{2}$. Thus, it is more likely that the true ranking is $a \succ_\pi b \succ_\pi c$ rather than $a \succ_\pi c \succ_\pi b$.

Note that for a single voting matrix, the binomial coefficients in these probabilities will actually always be identical. Therefore, all that matters is the exponents on p. Thus, we get a general form of

$$\Pr[M \mid \pi] \propto p^{\#\text{correct comparisons}}(1-p)^{\#\text{incorrect comparisons}}.$$

To maximize this, you will always want to maximize the number of agreements and minimize the number of disagreements because $p > \frac{1}{2}$.

Note that it is more plausible to think about the case where voters actually have rankings rather than potentially acyclic preferences. We will now introduce this model and see how the solution is essentially the same. But first, we must introduce the Kendall tau distance.

Definition 4 (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 pairwise disagreements between the two rankings. This distance can be interpreted as the number of swaps required to convert one ranking into the other, often referred to as the "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 are 3 disagreements of pairs. This can be seen by a bubble sort process by starting at σ, then making the following flips: $c \leftrightarrow d$, $b \leftrightarrow d$, $b \leftrightarrow c$. The resulting ranking is σ', and since we made 3 swaps, we know that the Kendall tau distance is just 3.

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

Definition 5 (Mallows Model)**.** The Mallows model is parameterized by $\phi \in (0, 1]$. The probability of a voter having a ranking σ given the true ranking π is:

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

Note that the more disagreements that σ has with π, the less likely it is that a voter will have the ranking σ. Additionally, when $\phi = 1$, all rankings are equally likely; as $\phi \to 0$, rankings closer to π become exponentially more likely.

Further, note that this is equivalent to the Condorcet noise model, where a cycle forces the process to "restart," with $\phi = \frac{1-p}{p}$. To gain some intuition on this, consider $p = 0.5$. Then, voters pick rankings at random, which makes sense because in this case $\phi = 1$ and so all rankings are equally likely. If $p \to 1$ then $\phi \to 0$, in which case the probability concentrates around the correct ranking for which the Kendall tau distance is 0.

Definition 6 (Kemeny Rule)**.** So what is the probability of observing profile σ given true ranking π? If we denote $Z_\phi = \sum_\tau \phi^{d_{KT}(\tau,\pi)}$, which is the normalizing constant in the denominator of our probability above, then

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

Thus, the MLE is clearly the Kemeny Rule: Given a preference profile σ, return a ranking π that minimizes

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

This is equivalent to what Condorcet wanted!

Unfortunately, there is a big caveat with finding this score...

Theorem 1. *Computing the optimal Kemeny score is NP-complete.*

Proof Idea. We wish to answer the following question: is there a ranking that achieves a Kemeny score (sum of Kendall Tau distances) of at most t for some given threshold t.

We will reduce to this problem from the Minimum Feedback Arc Set Problem, which asks whether, given a directed graph $G = (V, E)$ and $L \in \mathbb{N}$, there exists a subset $F \subseteq E$ such that $|F| \leq L$ and $(V, E \setminus F)$ is acyclic.

So, start with an instance of the Minimum Feedback Arc Set Problem, that is, we are given a directed graph $G = (V, E)$ and threshold $L \in \mathbb{N}$. We will treat the vertex set V as our alternatives and will construct $2|E|$ voters on those alternatives. For each edge in E, create a pair of voters that agree on the corresponding ordered pair of alternatives and disagree on everything else. More specifically, if the directed edge (a, b) exists in E and $V = \{a, b, c, d\}$, then we could construct a pair of voters with the following rankings:

1	2
a	d
b	c
c	a
d	b

Note that both of these voters agree on $a \succ b$, but they disagree on every other pairwise comparison. In this way, we continue constructing a pair of voters corresponding to every single edge in E.

Now, we claim that there is an acyclic subgraph that deletes k edges if and only if there is a ranking that (beyond inevitable disagreements) disagrees with k pairs of voters. By "beyond inevitable disagreements," we mean that for any given pairwise comparison made by a ranking and within every pair of voters, unless that pair of voters corresponds to that pairwise comparison in the ranking, one voter will agree with the comparison, and the other will disagree with it. For example, if a ranking made the choice that $b \succ c$, then in the pair of voters constructed above, one voter (1) will agree with the ranking, and the other voter (2) will disagree with the ranking, as the pair of voters corresponds to the ranking $a \succ b$. In this way, the only way you can do better with regard to the disagreements among voters is based on the distinguished alternatives that define the pairs of voters.

We will now justify the above statement. If you have an acyclic subgraph that deletes k edges, then we will take the ranking that is consistent with the remaining edges. Note that this is possible because the graph is acyclic. This ranking will agree with all the pairs of voters except for the pairs of voters that correspond to the k deleted edges, and so this direction is satisfied. For the other direction, if there is a ranking that disagrees with k of the pairs, then delete the k edges of the graph that corresponds to those pairs, and now the remaining subgraph must be acyclic because everything that remains agrees with the ranking we started from, which is acyclic. □

In practice, Kemeny computation is typically formulated as an integer linear program (ILP): For every $a, b \in A$, define:

- $x(a, b) = 1$ iff a is ranked above b,
- $w(a, b) = |\{i \in N : a \succ_{\sigma_i} b\}|$, how many voters prefer a to b

The ILP is:

$$\text{Minimize } \sum_{(a,b)} x(a, b)w(b, a),$$

subject to:

- For all distinct $a, b \in A$, $x(a, b) + x(b, a) = 1$,
- For all distinct $a, b, c \in A$, $x(a, b) + x(b, c) + x(c, a) \leq 2$,
- $x(a, b) \in \{0, 1\}$ for all distinct $a, b \in A$.

Here, the objective function is the penalty of the number of voters that disagree with the comparison $a \succ b$ across all pairs (a, b). The first constraint ensures we are choosing exactly one of $a \succ b$ or $b \succ a$. The second constraint makes sure we have no situation where $a \succ b \succ c \succ a$, and so transitivity is enforced. The third constraint just makes sense that $x(a, b)$ is a binary variable.