Algorithms, Games, and Networks February 19, 2013

Lecture 11

Lecturer: Ariel Procaccia *Scribe: Chris McComb*

1 Overview

We discuss several topics related to the concept of voting rules as Maximum Likelihood Estimators (MLEs). These topics include:

- Background and Motivation
- Condorcet's Noise Model and Solution Interpretations
- Young's Solution
- Selecting an MLE Top-Ranked Alternatives
- More Recent Work

This lecture does not cite any reading from the course textbook, but the following articles supplement the lecture.

- H.P. Young, "Condorcet's Theory of Voting". *America Political Science Review*, 1988.
- H.P. Young, "Optimal Voting Rules". *Journal of Economic Perspectives*, 1995.
- V. Conitzer, T. Sandholm, "Common Voting Rules as Maximum Likelihood Estimators". In Proceedings of the *Conference on Uncertainty in Artificial Intelligence*, 2005.

2 Background and Motivation

Condorcet viewed voting not merely as balancing subjective opinions, but as a search for a more objective truth. He believed that voters were *enlightened* in some way, having the ability to determine which alternative would most benefit society. Although this is not always the case in voting, it is a fairly realistic model of some processes, including pooling expert opinion, trial by jury, and human computation.

A more specific example of voting in human computation is EteRNA, an in-browser game developed by researchers at Carnegie Mellon University and Stanford University. In EteRNA, players design an RNA sequence to fit a target shape. At regular intervals, the players vote to select the most promising RNA sequences. The top eight sequences are then synthesized.

More broadly, EteRNA serves as an example for how to employ a voting system as a Maximum Likelihood Estimator (MLE). This concept will now explored in more detail.

3 Condorcet's Noise Model and Solution

In Condorcet's noise model, alternatives are presented to voters in a series of pairwise elections. There is a true ranking, and the preferences of individual voters are noisy approximations of this ranking. Each comparison is correct with probability $p > 1/2$. The goal of an election is to determine the true ranking.

In order to further illustrate this concept, let's consider an example. The results of an election with three alternatives are shown in Table 1.

Table 1: Voting Matrix for 3 Alternatives

	a	**b**	**c**
a	-	8	6
b	5	-	11
c	7	2	-

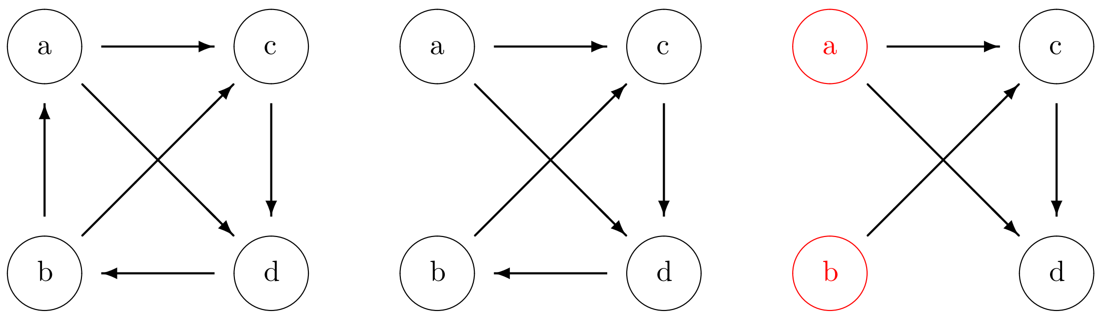
In this example, we see that $a \succ b$, $b \succ c$ and $c \succ a$. In other words, the preference profile is cyclic. Condorcet suggested to "successively delete the comparisons that have the least plurality". Using this suggestion, we choose to delete $c \succ a$. This makes the preference profile non-cyclic, resulting in a final ranking of $a \succ b \succ c$.

However, this method does not readily extend to voting matrices with more than 3 alternatives. Table 2 shows another voting matrix. In order of strength, the preferences from Table 2 are: $c \succ d$, $a \succ d$, $b \succ c$, $a \succ c$, $d \succ b$ and $b \succ a$. Therefore, these preferences are also cyclic in nature.

Table 2: Voting Matrix for 4 Alternatives

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

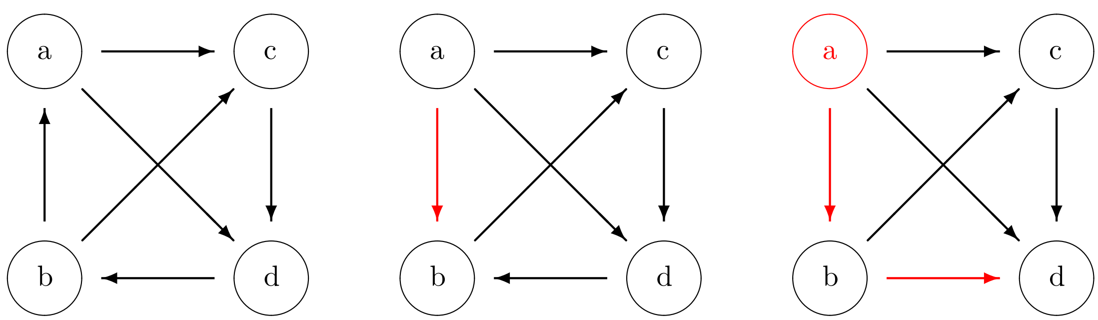
Following Condorcet's example, we should delete $b \succ a$ (see Figure 1(a)). However, this preference relationship is not part of the cycle, so this deletion does nothing to solve our dilemma (see Figure 1(b)). The next weakest preference is $d \succ b$. If we delete this preference, it *does* alter the cycle. Unfortunately, we are then left with an ambiguous set of preferences, indicating that either a or b is the winner (see Figure 1(c)). Obviously, the successive deletion of pairwise preferences is not a valid method for determining a ranking or selecting a winner.



(a) Unaltered Graph (b) $b \succ a$ deleted (c) $d \succ b$ deleted

Figure 1: Successive deletion of preferences for 4 alternatives

Perhaps Condorcet meant to suggest that pairwise preferences should be *reversed* instead of *deleted*? For the voting matrix in Table 2, we would first we reverse, $b \succ a$ (see Figure 2(b)). Since this doesn't result in a clear winner, we also reverse $d \succ b$ (see Figure 2(c)).



(a) Unaltered Graph (b) $b \succ a$ reversed (c) $d \succ b$ reversed

Figure 2: Successive reversal of preferences for 4 alternatives

This set of reversals reveals a ranking of $a \succ b \succ c \succ d$. If we count the number of votes in support $a \succ b$, $b \succ c$ and $c \succ d$, we see that 89 votes support this ranking. Had we only reversed $d \succ d$, the final ranking would have been $b \succ a \succ c \succ d$. This ranking would have had more support, amassing a total of 90 voters. For this reason, it appears that reversal is not the appropriate interpretation of Condorcet's suggestion either.
Considerable frustration with Condorcet's ambiguous suggestion has been stated in the literature, most notably in articles by Black (1958) and Todhunter (1949).

4 Young's Solution

Young decided to look at the noise model from a more probabilistic point of view. For instance, if we consider the voting matrix in Table 1, we can calculate the probability that the true ranking is $a \succ b \succ c$ as follows:

$$P(a \succ b \succ c) = \binom{13}{8} p^8 (1-p)^5 \cdot \binom{13}{6} p^6 (1-p)^7 \cdot \binom{13}{11} p^{11} (1-p)^2$$

Similarly, we can calculate the probability that the true ranking is $a \succ c \succ b$ as:

$$P(a \succ c \succ b) = \binom{13}{8} p^8 (1-p)^5 \cdot \binom{13}{6} p^6 (1-p)^7 \cdot \binom{13}{2} p^2 (1-p)^{11}$$

The most likely true ranking is the ranking that has maximum probability. Since $\binom{n}{k} = \binom{n}{n-k}$, all of the binomial coefficients in the above equations are identical. Also, since $p > 1/2$, a ranking with a larger exponent for p (and a small exponent for $1-p$) has a higher probability. We can use Bayes' Theorem to formalize this mathematically as follows:

$$P(\succ |M) = \frac{P(M| \succ) \cdot P(\succ)}{P(M)}$$

This equation can be greatly simplified if we recognize that $P(M)$ is a constant (since the voter matrix, M, is know). It can be further simplified by assuming uniform priors for the different rankings ($P(\succ) = 1/m!$). Therefore, to maximize $P(\succ |M)$ we need only maximize $P(M| \succ)$. We can do this by selecting the profile that minimizes the number of disagreements in votes on pairwise alternatives. This is a well-established voting rule know as the Kemeny Rule.

5 Selecting a MLE Top-Ranked Alternative

In many voting situations, the top-ranked alternative is of more interest than the ranking itself. In these cases, it might be tempting to find the MLE ranking, and then take the top alternative in the ranking. However, this alternative is not necessarily the most likely best alternative. Let us use the voting matrix from Table 3 for this example.

Table 3: Voting Matrix for 3 Alternatives

	a	**b**	**c**
a	-	23	29
b	37	-	29
c	31	31	-

In order for c to be the most likely best alternative, we must have $c \succ a$ and $c \succ b$. We can express the probability for these two comparisons as follows:

$$P(c \succ a|M_{ca}) = \frac{p^{31}(1-p)^{29}}{p^{31}(1-p)^{29} + p^{29}(1-p)^{31}}$$

$$P(c \succ b|M_{cb}) = \frac{p^{31}(1-p)^{29}}{p^{31}(1-p)^{29} + p^{29}(1-p)^{31}}$$

Since both of these must occur for c to win, the probability that c is the winner is the joint probability.

$$P(c|M_{ca} \cap M_{cb}) = P(c \succ a|M_{ca}) \cdot P(c \succ b|M_{cb}) = \frac{\left(p^{31}(1-p)^{29}\right)^2}{\left(p^{31}(1-p)^{29} + p^{29}(1-p)^{31}\right)^2}$$

We can state the probability that a or b is the winner in much the same way

$$P(a|M_{ac} \cap M_{ab}) = \frac{\left(p^{37}(1-p)^{23}p^{29}(1-p)^{31}\right)^2}{\left(p^{37}(1-p)^{23} + p^{23}(1-p)^{37}\right) \cdot \left(p^{29}(1-p)^{31} + p^{31}(1-p)^{29}\right)}$$

$$P(b|M_{ba} \cap M_{bc}) = \frac{\left(p^{23}(1-p)^{37}p^{29}(1-p)^{31}\right)^2}{\left(p^{37}(1-p)^{23} + p^{23}(1-p)^{37}\right) \cdot \left(p^{29}(1-p)^{31} + p^{31}(1-p)^{29}\right)}$$

Given the form of the above equations, it is apparent that the relative magnitudes of the probabilities depends heavily on the exact value of p. For the case where $p \approx 1$, we can reduce the above to:

$$P(a|M_{ac} \cap M_{ab}) \approx 0$$

$$P(b|M_{ba} \cap M_{bc}) \approx 0$$

$$P(c|M_{ca} \cap M_{cb}) \approx 1$$

Therefore, for the voting matrix in Table 3, when $p \approx 1$, the winner is alternative c.
For the case when $p \approx 1/2$, Young demonstrated that the Borda voting rule is a maximum likelihood estimator for the best alternative. For the voting matrix shown in Table 3, the Borda winner is b. Given a voting matrix, the Borda score can be quickly calculated

6 More Recent Work

In 2005, Conitzer and Sandholm published a paper further exploring voting rules in MLEs. Specifically, their work asked the question: which common voting rules have a noise model for which they are maximum likehood estimators of the true ranking (MLER) or of the true winner (MLEW). They presented two important theorems. The first theorem shows that every scoring rule is a MLEW, as follows:

Theorem 1 *Any scoring rule is a maximum likelihood estimator of the true winner.*

Proof: Assume that w is the true winner, and consider N voters. Every voter i assigns a ranking r to every alternative. A score s_r is then assigned based on the ranking.
The noise model is such that the probability that voter i ranks w in position $r_i(w)$ is proportional to $2^{s_{r_i(w)}}$, and all other alternatives are ranked randomly.
Therefore, the probability that matrix M results in w being a winner is:

$$P(M|w) \propto \prod_{i=1}^{N} 2^{s_{r_i(w)}} = 2^{\Sigma s_{r_i(w)}}$$

■

The other important theorem showed that the maximin voting rule is *not* an MLEW. Before beginning this proof, we will prove two lemmas.

Lemma 2 *If there exist preference profiles* $\succ^1$ *and* $\succ^2$ *such that* $f(\succ^1) = f(\succ^2) \neq f(\succ^3)$, *where* $\succ^3$ *is their union, then* f *is not an MLEW.*

Proof: All rankings are iid. If both $\succ^1$ $\succ^2$ maximize the probability of x, then their union must almost maximize the probability of x.

$$P(\succ^3 |x) = P(\succ^2 |x) \cdot P(\succ^1 |x)$$

■

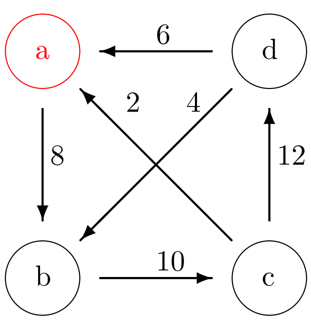
Lemma 3 *Any pairwise comparison graph whose weights are even-valued can be realized via votes.*

Proof: To increase the weight on edge (a, b), it is only necessary to add the votes $a \succ b \succ x_1 \succ ... \succ x_{m-2}$ and $x_{m-2} \succ ... \succ x_1 \succ a \succ b$ to the graph. This results in two votes being added to edge (a, b). ■

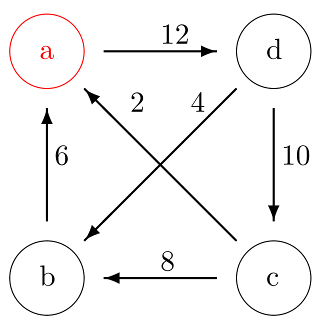
By Lemma 2, we can prove the second theorem by generating preference profiles such that $f(\succ^1) = f(\succ^2) \neq f(\succ^3)$, where $\succ^3$ is the union of $\succ^1$ and $\succ^2$, and f is the maximin voting rule. We will use Lemma 3 to generate these profiles.

Theorem 4 *Maximin is not an MLEW voting rule.*

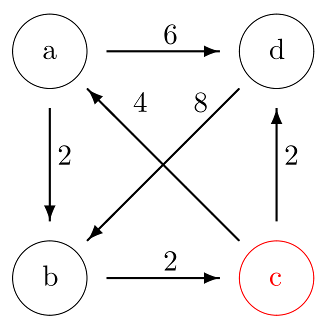
Proof: Let $\succ^1$ be as shown in Figure 3(a), $\succ^2$ be as shown in Figure 3(b), and $\succ^3$ (the union of $\succ^1$ and $\succ^2$) be as shown in Figure 3(c).



(a) Graph for $\succ^1$



(b) Graph for $\succ^2$



(c) Graph for $\succ^3$

Figure 3: Maximin is not MLEW

Profile $\succ^1$ results in winner a with the maximin voting rule, as does profile $\succ^2$. However, $\succ^3$ results in winner c. Therefore, maximin is not an MLEW voting rule. ■