CS295 Introduction to Algorithmic Game Theory

Instructor: Ioannis Panageas Scribed by: Ryuto Kitagawa

Lecture 14. Social Choice Theory

1 Introduction

1.1 Definitions

The following is a set of definitions to prepare the reader for the rest of the topics.

Definition 1.1 *Social choice theory is defined by:*

- *Set I of n voters*
- *Set A of m candidates*
- *Each voter has some preference of the candidates as a set L, which is a permutation of A*

As an example, there could be 3 voters with 3 candidates A, B, C. Voter i, denoted as V_i, may have some preference C, A, B.

Definition 1.2 *A social choice function is defined by the following function*

$$f : L \times \cdots \times L \to A.$$

where the input is a set of preferences from voters, and the result is a single candidate.

Definition 1.3 *A social welfare function is defined by the following function*

$$f : L \times \cdots \times L \to L.$$

where the input is a set of preferences from voters, and the result is an ordering of candidates.

1.2 Desirable Properties

Given the social welfare function, there are two desirable properties the function desires: unanimity and independence of irrelevant alternatives.

Definition 1.4 *A social welfare function F satisfies unanimity if*

$$F(>, >, \cdots, >) => ,$$

for all $> \in L$.

In other words, if all voters prefer candidate a to candidate b, then the social welfare outcome should also prefer a to b.

Unanimity should be desired since the implication is the social welfare outcome reflects the desires of the voters. If all voters prefer b to a, then the social welfare function preferring a to b would not be representative of the voters.

Definition 1.5 *A social welfare function F satisfies independence of irrelevant alternatives if for any pair of candidates $a, b \in A$ and any pair of preferences $>_1, \ldots, >_n$ and $\overline{>}_1, \ldots, \overline{>}_n$ with $>= F(>_1, \ldots, >_n)$ and $\overline{>} = F(\overline{>}_1, \ldots, \overline{>}_n)$,*

$$a <_i b \iff a \overline{<}_i b \text{ for all } i \rightarrow a < b \iff a \overline{<} b.$$

More intuitively, if an outside irrelevant factor within the preference changes, then the relative preferences should not change. Suppose candidate a is preferred over candidate b. If a new candidate c appears, who is still not preferred to a, then the voter will not pick candidate c, since it should not be factored.

independence of irrelevant alternatives ensures the voters are remaining truthful in their votes. If for example a voter prefers a over b, they are always incentivized to vote truthfully. However, if candidate c enters, and the voter has a preference in the following order a, c, b, it is possible that if candidate b and c are significantly more popular than a, then the voter may become incentivized to support c, being untruthful.

2 Dictatorship

2.1 Arrow's Theorem

First, allow us to define a dictatorship.

Definition 2.1 *A voter i is a dictator if*

$$F(>_1, \ldots, >_n) =>_i,$$

for all $>\in L$.

In other words, if the social function (choice or welfare) is decided entirely by a single voter. Arrow's Theorem states that obtaining a social welfare function that satisfies unanimity and independence of irrelevant alternatives means it is a dictatorship.

Theorem 2.1 *Every social welfare function over a set of more than 2 candidates (—A— ¿ 3), that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.*

To begin the intuition of the proof, assume there's a social welfare function F that satisfies unanimity and independence of irrelevant alternatives. Consider two sets of preference profiles

$(>_1, \cdots >_n)$ and $(\overline{>}_1, \ldots \overline{>}_n)$ such that $F(>_1, \cdots >_n) => $ and $F(\overline{>}_1, \ldots \overline{>}_n) = \overline{>}$, and $a, b, c, d \in A$ such that

$$\text{for all voters } i,\ a <_i b \iff c \overline{<}_i d.$$

Since unanimity is fulfilled, then the output of the social welfare function would give us

$$a < b \iff c \overline{<} d.$$

The two outputs of the social welfare function can be merged in such a way where we claim that $c < a$ and $b < d$. This ensures that the two cases of the inequality above is fulfilled.

Now we must prove that a dictator exists within this ordering. Suppose there is a set of preference profiles π where where for the first i preferences prefer candidate a to b. More formally

$$a >_j b \iff j \leq i.$$

Suppose that i^* is the pivoting index; we need to show that i^* is a dictator. First, assume that $c \neq d$ and $c >_{i^*} d$. We need to show need to show that $c > d$ where $>= F(>_1, \ldots, >_n)$. Consider some candidate $e \neq c, d$.

When $i < i^*$, have e be at the most prioritized candidate, and when $i > i^*$, have e be the least prioritized candidate. If $i = i^*$, then have the ordering such that $d < e < c$. Adding this candidate satisfies the independence of irrelevant alternatives since the relative ordering of c and d do not change. We can conclude $c > e$ and $e > d$, which tells us that $c > d$, which proves that i^* is a dictatorship.

2.2 Gibbard Satterthwaite Theorem

Definition 2.2 *A social choice function f is monotone if $a = f(<_1, \ldots, <_i, \ldots, <_n)$ and $b = f(<_1, \ldots, <'_i, \ldots, <_n)$ implies that*

$$b <_i a \text{ and } a <'_i b.$$

In other words, a social choice function is monotone if changing a voter's preference, i, will result in a different outcome. In other words, they are the pivotal voter. This results in another negative result which goes hand in hand with Arrow's Theorem.

Theorem 2.2 *Let f be a monotone social choice function onto A with $|A| \geq 3$, then f is a dictatorship.*

To prove the statement, suppose we are given a social choice function f that is monotone and onto. We first construct a social welfare function from this social choice function, then use Arrow's Theorem to complete the proof. Note that with the social choice function, we can decide which of two candidates are preferred easily by moving those candidates to the front of the preference profile (maintaining their relative ordering) and then passing that to the social choice function.

To generalize, suppose we have a subset of candidates S and an ordering preference $<$. Suppose we denote $<^S$ to be the preference profile by considering S candidates. Formally:

- If $a, b \in S$ then $a < b \iff a <^S B$
- If $a, b \notin S$ then $a < b \iff a <^S B$
- $a \notin S$ and $b \in S$ then $a <^S b$

Suppose we have a social welfare function F such that $<= F(<_1, \ldots, <_n)$ where $a < b$ if and only if $S = \{a, b\}$ and $a <^S b$. This is a basic extension of the social choice function. If we can then show that this function F satisfies unanimity and independence of irrelevant alternatives, then we can use Arrow's Theorem to conclude that there is a dictatorship in f when f is monotone.

Onto means that there is some preference profile such that any of the candidates could potentially win. More formally, this can be written as

$$f(<'_1, \ldots, <'_n) = a.$$

We can replace $<'_i$ with $<^S_i$ instead without loss of generality. Notice that if $b \notin S$, then it will not be chosen by the function, so long as S is not $\emptyset$. In other words, for any S and $<_1, \ldots, <_n$, we have $f(<^S_1, \ldots, <^S_n) \in S$.

We need to prove that F satisfies unanimity and independence of irrelevant alternatives. For unanimity, assume that $b <_i a$ for all i. We know that if $S = \{a, b\}$, that the result of the social choice function would be a for any two candidates. Therefore, this satisfies unanimity directly. Next to show that F is independent of irrelevant alternatives, suppose that $b <_i a \iff b <'_i a$. If we set $S = \{a, b\}$, then we know that placing the preference profiles into the social choice function will give us the same result for both $<_i$ and $<^*_i$.

3 Circumventing Dictatorship

Due to the issue of dictatorship being unavoidable according to Arrow's Theorem and Gibbard Satterthwaite Theorem. Therefore, a simple way to avoid the possibility of dictatorship is through the introduction of randomization to the system. The fastest and simplest way to achieve this is through a simple lottery system. For example, a single voter is chosen at random, and their preference is decided for the social welfare function and social choice function. Intuitively, it is clear to see that the outcome is could never be rigged such that a single player is the sole decider of the social functions. However, the natural question that arises is how to measure the performance of the mechanism and what the guarantees are.

3.1 Positional Scoring Based Rules

Definition 3.1 *Let n be the number of voters and m the number of candidates. Each voter i has preference $>_i$. A positional scoring rule is defined by a vector of non-negative real numbers $a = (a_1, \ldots, a_n)$ so that the score of candidate x is given by*

$$sc(x, >) = \sum_{i=1}^{n} a_{>_i}(x).$$

Intuitively, the score for some candidate is the sum of all the points each voter gives it. There are several scoring methodologies that exist for different purposes.

- Plurality: A voter's top candidate is given a single point, and no other candidates are awarded points by that voter. This is similar to a ballot box, where you are only able to place a single candidate's name in the poll.
- Borda: Each of the candidates are given points awarded to their position in the voter's ranking. The top candidate for a voter would get $m - 1$ points, the second best candidate would get $m - 2$, and so on and so forth.
- Veto: All of the candidates are given a single point except for the least favored candidate by the voter. Notice that this is the antithesis of plurality, i.e. the candidate with the least votes in veto is the same as voting with plurality for the least favored candidate.

The overall goal of the positional scoring based system is to design a scoring system that is incentive compatible and is having a system where the candidate with the most votes is the winner.