Algorithmic Game Theory, Summer 2015 — Lecture 7 (4 pages)

Social Choice

Instructor: Xiaohui Bei

1 Introduction

Social choice is essentially the study of voting. Leaving aside the question of strategic behavior, we want to know how to aggregate a set of different people's preference together in a sensible way. The foundation of social choice theory is build on the following Condorcet's paradox.

Example 7.1 (Condorcet Paradox). *Assuming that there is a election with three candidates A, B and C. One third of the voters prefer $A > B > C$, one third of the voters prefer $B > C > A$, and one third prefer $C > A > B$. Who should win the election?*

Suppose that candidate A wins, then a counter argument would be that C should win instead, because a majority (two thirds) of the voters prefer C over A. The same argument also holds for the other two possibilities.

2 Model

A social choice problem consists of:

- a finite set of possible *outcomes* or *alternatives* O
- a set of voters $N = \{1, 2, \ldots, n\}$
- a (strict) *preference* over the outcomes for every agent: linear orders

Definition 7.2. *A* linear order *or* preference *over a set of outcomes O is a set of binary relations $\succ$ that are* total *and* transitive*.*

- *total: for every pair of outcomes $a \neq b$, either $a \succ b$ or $b \succ a$*
- *transitive: $a \succ b$ and $b \succ c$ implies $a \succ c$.*

Definition 7.3. *Given a finite set of outcomes O, a set of agents $N = \{1, 2, \ldots, n\}$, and the set of preferences L over outcomes.*

- *A* social choice function *(or* voting rule*) is a function $C : L^n \mapsto O$.*
- *A* social welfare function *(or* social welfare ordering*) is a function $W : L^n \mapsto L$.*

Example 7.4. *Examples of some simple social choice functions.*

- **Plurality.** *The outcome that is ranked at the very top by the most voters wins.*
- **Plurality with elimination.** *If an outcome is ranked 1st by a majority of the voters, it is selected as the winner. Otherwise, eliminate the outcome that is ranked 1st by the least number of voters. Repeat the above procedure until a winner is selected.*
- **Borda count.** *A voter gives 0 point to the very bottom outcome in his preference, 1 point to the outcome that is second to last, 2 points to the outcome that is third to last, and so on. Add up all points from all voters and the outcome with most points wins.*

A concept that is of particular interest in called *Condorcet consistency.*

Definition 7.5. *If there is an outcome o such that for every other outcome o', there is a majority of voters that prefer o to o', then o is called a* Condorcet winner*. A social choice function that always choose the Condorcet winner when there is one is said to have* Condorcet consistency*.*

Notice that a Condorcet winner may not always exist. A simple counter-example is the Condorcet paradox example that we discussed earlier, in which there is a cycle of outcomes that defeats one another.

3 Axioms

The goal of social choice is to design social choice functions or social welfare functions in some reasonable way. In Economics, such "reasonableness" is often described by a set of axioms. Below are some axioms on social welfare functions that are commonly studied.

Notation. We will use $[\succ] = (\succ_1, \ldots, \succ_n)$ to denote a preference profile, which is an element in L^n, and use $\succ_W$ as a short form of $W([\succ])$.

Definition 7.6 (Pareto Efficiency (PE))**.** *W is Pareto efficient if for any $o_1, o_2 \in O$, $\forall i, o_1 \succ_i o_2$ implies that $o_1 \succ_W o_2$.*

In other words, when all voters prefer one outcome over another outcome, the social welfare function must also agree with this order.

Definition 7.7 (Independence of Irrelevant Alternatives (IIA))**.** *W is independent of irrelevant alternatives if, for any two preference profiles $[\succ], [\succ'] \in L^n$ and any pair $o_1, o_2 \in O$, $\forall i, o_1 \succ_i o_2$ iff $o_1 \succ'_i o_2$ implies $o_1 \succ_W o_2$ iff $o_1 \succ'_W o_2$.*

The IIA condition means that the returned ordering between two outcomes depends only on the relative ordering of these two outcomes given by each voter.

Definition 7.8 (Non-dictatorship)**.** *W is not dictatorial if there does not exist a voter i such that $\succ_W = \succ_i$.*

Non-dictatorship means there does not exist a single voter whose preference always determines the result of the social welfare function.

Theorem 7.9 (Arrow's Impossibility Theorem, 1951)**.** *There does not exist a social welfare function over three or more outcomes that is Pareto efficient, independent of irrelevant alternatives and not dictatorial.*

Proof. We will assume a function W which is both PE and IIA, and show that W must be dictatorial.

The proof consists of four steps. In the following, we call a preference profile *o-extreme* if the outcome o is either at the very top or the very bottom of every voter's preference.

- **Step 1: If $[\succ]$ is o-extreme, then o will be at either the very top or the very bottom of $\succ_W$ as well.**

  Assume by contradiction that there exists an o-extreme preference profile $[\succ]$, and there exists $o_1, o_2 \in O$ such that $o_1 \succ_W o \succ_W o_2$. Now let's modify $[\succ]$ in the following way: for every voter i, switch the position of o_1 and o_2 if $o_1 \succ_i o_2$, otherwise leave the preference unchanged. Denote the new preference profile $[\succ']$. We know that in $[\succ']$ every voter ranks o_2 above o_1. Note that the relative order between o_1 and o remains unchanged for every voter from $[\succ]$ to $[\succ']$, hence by IIA we know $o_1 \succ'_W o$. The same is true for o and o_2, hence we also have $o \succ'_W o_2$. Then by transitivity we have $o_1 \succ'_W o_2$. This contradicts PE since we have $o_2 \succ'_i o_1$ for every voter i.

- **Step 2: There exists an o-extreme preference profile $[\succ]$ and a voter i^*, such that by only changing the rank of o in $\succ_{i^*}$ from the very bottom to the very top, in $\succ_W$ outcome o will also change from the very bottom to the very top.**

  Consider an arbitrary preference profile in which o is ranked at the very bottom in every voter's preference. By PE we know o will also be at the very bottom in the social ranking. Now we change only the ranking of o from bottom to top in each voter's preference one by one. Because at the end when all voters rank o as their top choice, o will also be at the top in the social ranking. Hence there must exists a voter i^*, such that when it is i^*'s turn to change o from the bottom to the top, the social ranking $\succ_W$ will also change o from the bottom to the top.

- **Step 3: i^* is a dictator over any pair of outcomes o_1, o_2 not involving o.**

  Let $[\succ]$ be an arbitrary preference profile in which $o_1 \succ_{i^*} o_2$, we want to show that $o_1 \succ_W o_2$. Denote by $[\succ^1]$ the preference profile just before i^* changes the rank of o, and denote by $[\succ^2]$ the preference profile right after i^* puts o at the top of his ranking. Now let's modify $[\succ]$ in the following way: first for every voter $i \neq i^*$, change the rank of o to the same rank as in $[\succ^1]$ (which is either at the very top or the very bottom), then for voter i^*, move o to an arbitrary place between o_1 and o_2, such that $o_1 \succ_{i^*} o \succ_{i^*} o_2$. Denote this new preference profile by $[\succ']$. Note that these modifications do not change the relative order of o_1 and o_2 in any voter's preference. Hence the relative order of o_1 and o_2 should be the same in $\succ_W$ and $\succ'_W$.

  Now consider o_1 and o, notice that when compare $[\succ^1]$ to $[\succ']$, their relative rankings are the same for every voter. Since $o_1 \succ^1_W o$, by IIA, we must have $o_1 \succ'_W o$ as well. Using the same argument, observe that the relative order between o and o_2 does not change for every voter in $[\succ^2]$ and $[\succ']$, and $o \succ^2_W o_2$. Thus we have $o \succ'_W o_2$. Finally, by transitivity, we have $o_1 \succ'_W o_2$, and thus $o_1 \succ_W o_2$.

- **Step 4: i^* is a dictator over any pair of outcomes o_1, o.**

  Consider another outcome $o_2 \neq o_1, o$. By the argument in Step 2 and 3, we know there exists a voter i^{**} who is a dictator over any pair of outcomes not involving o_2. Hence i^{**} is a dictator over o and o_1. Observe that in the case discussed in Step 2, there is a scenario in which voter i^* can affect the order between o and o_1 in $\succ_W$, thus we must have $i^{**} = i^*$.

$\square$

4 Social Choice Function

Arrow's impossibility theorem tells us that one cannot design a voting rule that satisfies certain basic axioms simultaneously. It might be thought that the problem lies in the fact that a social welfare function requires one to output a complete preference list, which might be too demanding. If instead one only wants to select a winner, as does a social choice function, then such paradox would be resolved. Unfortunately, in the following we will show that this is not the case.

Before talking about the results, we need to first refine the axioms for the social choice function setting. Some of them, such as PE and IIA, cannot be easily translated to versions for social choice functions. But as we will see, there are closely related notions that are well defined for the social choice function setting.

Definition 7.10 (Weak Pareto Efficiency). *A social choice function C is weakly Pareto efficient if for any preference profile $[\succ]$, $C([\succ])$ never output an outcome o when there exists another outcome o' such that $\forall i, o' \succ_i o$.*

Weak Pareto efficiency indicates that a dominated outcome should never be chosen.

Definition 7.11 (Monotonicity). *C is monotonic if, for any outcome $o \in O$ and two preference profiles $[\succ], [\succ']$ that satisfy $\forall i \in N, o' \in O$, $o \succ_i o'$ implies $o \succ_i' o'$, then $C([\succ]) = o$ implies $C([\succ']) = o$.*

In other words, with a monotonic social choice function, the winner outcome for a preference profile will remain the winner if we only increase the support of this outcome.

Definition 7.12 (Non-dictatorship). *C is not dictatorial if there does not exist a voter i such that C always outputs the top choice in i's preference ordering.*

Theorem 7.13 (Muller-Satterthwaite, 1977). *There does not exist a social choice function that is weakly Pareto efficient, monotonic and not dictatorial.*

The Muller-Satterthwaite theorem tells us that social choice functions are not more benign than social welfare functions.

The intuition behind the proof is to design a procedure that uses a social choice function to determine the relative social ordering between two outcomes. Then by repeating such procedure on all pairs of outcomes, we can construct a full social welfare function using this technique, and Arrow's impossibility theorem for social welfare functions can be translated to the desired claim that we need for social choice functions. The full proof is left as an exercise for the reader.

Final Remark. Social choice theory studies the problem of preference aggregation when each voter's preference is known. To consider the case where these preferences are unknown gives rise to the field of mechanism design, which we will discuss in the next lecture.