STAT 206A: Social Choice and Networks	Fall 2010
Manipulation and GS Theorem — October 21	
Lecturer: Elchanan Mossel	*Scribe: Kristen Woyach*

In this lecture we cover manipulation by a single voter: whether a single voter can lie about his preference and produce an outcome that is better for him. We look at manipulation in the case of binary voting and ranking multiple preferences. We will then cover the Gibbard-Satterthwaite (GS) Theorem, which says that under certain conditions, if the aggregation function is not manipulable then it must be the dictator function.

1 Manipulation for binary voters

Consider the following scenario:

- n voters each choose from $\{-1, 1\}$
- x_i is the vote of the ith voter.
- outcome function $f : \{-1, 1\}^n \to \{-1, 1\}$

We will define manipulation by a single voter in the following way:

Definition 1.1. *f is manipulable by voter 1 if there exists $x_2 \dots x_n$ such that:*

$$f(1, x_2, \dots, x_n) = -1 \quad \textit{and} \quad f(-1, x_2, \dots, x_n) = 1$$

If voter 1 has a true preference of 1, then he must vote -1 in order to get his preferred outcome. So, voter 1 has incentive to lie about his true preference.

Notice that the majority function is an example of a function that is not manipulable: any voter voting against their true preference can only make the outcome less likely to be their preference. This example can be generalized to say something about the relationship between manipulability and monotonicity:

Claim 1.2. *f is manipulable if and only if f is* not *monotone*

Proof:

- if f is monotone, then it is not manipulable:

  Assume by contradiction that f is manipulable. Then

  $$-1 = f(1, x_2, \dots, x_n) \not\geq_1 f(-1, x_2, \dots, x_n) = 1$$

  which contradicts monotonicity.

- if f is not manipulable, then it is monotone:

  If f is non-manipulable then it is monotone with respect to flip of any signle coordinate. We know that this condition implies monotonicity of f.

□

2 Manipulation for voters with ≥ 3 alternatives

When there are more than 2 alternatives, the situation is more complicated. Consider the following example, reminiscent of the 2000 presidential election:

There are three possible rankings of the options $\{a, b, c\}$ with the probability of desiring that ranking shown below:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \begin{pmatrix} b \\ c \\ a \end{pmatrix} \quad \begin{pmatrix} c \\ b \\ a \end{pmatrix}$$
$$45\% \quad 40\% \quad 15\%$$

(This corresponds roughly to the situation in 2000 with a = Bush, b = Gore, and c = Nader). Assume we use a plurality function to determine the winner. If everyone votes according to their true preferences, then Bush will win. However, if the Nader supporters realize that they can never win, they may choose to lie about their preferences and vote for Gore instead. Now Gore will win, which is a better alternative for the Nader supporters than Bush.

We will assume that we want voters to vote according to their actual preferences (truthful voting), and we do not want voters to try to game the election by voting creatively. So, we want to understand under what aggregation functions no voter will be able to manipulate the election. To do this, we need a number of definitions:

Definition 2.1. *A* **social choice function** *(SCF) is a function*

$$F : S(k)^n \to [k]$$

which takes in all rankings from the voters and computes a winner.

Definition 2.2. *F is* **manipulable** *by voter 1 if*

$$\exists(\sigma_1, \sigma_{-1}), (\sigma'_1, \sigma_{-1}) : \sigma_1(F(\sigma'_1, \sigma_{-1})) > \sigma_1(F(\sigma_1, \sigma_{-1}))$$

where σ_1 is the ranking of voter 1, σ_{-1} is all of the rankings of all voters except voter 1 ($\sigma_{-1} = \sigma_2 \ldots \sigma_n$) and $\sigma_1(F)$ is the rank of the outcome of F in the preference of the first voter.

In other words, there must be a ranking that is different from voter 1's true ranking that would make the SCF return a value that is more preferable to voter 1. In the 2000 election example, Gore is a better option than Bush for the Nader voters, so it is in their interest to rank Gore above Nader despite actually preferring Nader.

Notice that the inequality in this definition is strict: we do not care about cases when switching the ranking will result in an equivalently good outcome.

Definition 2.3. *F satisfies* **unanimity** *If*

$$\forall i, a \text{ is the top alternative of } \sigma_i$$

Then

$$F(\sigma) = F(\sigma_1, \ldots, \sigma_n) = a$$

Definition 2.4. *F is* **neutral** *if the function is fair among all alternatives:*

$$\forall\ \sigma' \in S(a,b,\ldots,k)\ \text{ and }\ \sigma \in S(a,b,\ldots,k) : F(\sigma'\ \sigma) = \sigma' F(\sigma)$$

Definition 2.5. *F is* **strategy-proof** *if F is not manipulable*

Definition 2.6. *The ith dictator function is the function described by*

$$F(\sigma) = \textit{top alternative}(\sigma_i)$$

A remark was made that we might be able fix the manipulability of the plurality function in the last example by implementing a run-off election. This means that the first election chooses the two alternatives with the highest number of votes and then there is a second election between these two. The winner is the majority winner of the second election. However, this example too is manipulable. Consider the following case:

$$\underset{31\%}{\begin{pmatrix} a \\ c \\ b \end{pmatrix}} \quad \underset{29\%}{\begin{pmatrix} b \\ a \\ c \end{pmatrix}} \quad \underset{40\%}{\begin{pmatrix} c \\ b \\ a \end{pmatrix}} \quad \overset{a, c \text{ go to run-off}}{\longrightarrow} \quad \underset{60\%}{\begin{pmatrix} a \\ c \end{pmatrix}} \quad \underset{40\%}{\begin{pmatrix} c \\ a \end{pmatrix}}$$

In the first election, a and c receive the most votes, so they are part of the run-off. In the runoff, a holds the most votes, so it is declared the winner. However, if 3% of the (c, b, a) preference voters choose to lie in the initial election and vote for (b, a, c) instead, then return to their true vote in the run-off, we have the following situation:

$$\underset{31\%}{\begin{pmatrix} a \\ c \\ b \end{pmatrix}} \quad \underset{32\%}{\begin{pmatrix} b \\ a \\ c \end{pmatrix}} \quad \underset{37\%}{\begin{pmatrix} c \\ b \\ a \end{pmatrix}} \quad \overset{b, c \text{ go to run-off}}{\longrightarrow} \quad \underset{29\%}{\begin{pmatrix} b \\ c \end{pmatrix}} \quad \underset{71\%}{\begin{pmatrix} c \\ b \end{pmatrix}}$$

Now, b and c are in the run-off, where c wins by a landslide. Therefore, there is incentive for a number of voters to not be truthful in the first round.

With these definitions in hand, the rest of the lecture is devoted to stating and proving the GS theorem.

3 Gibbard-Satterthwaite Theorem

Theorem 3.1 (Gibbard-Satterthwaite (GS) Theorem). *If $F : S(k)^n \to [k]$ is onto and strategy-proof with $k \geq 3$, then F is a dictator function.*

A function is **onto** if all alternatives can be the winner. This theorem is true in particular if F is neutral. The original proof was a derivative of the proof of Arrow's theorem [1, 2]. The proof we will follow here comes from Svensson [3]. We will need a couple of lemmas in order to complete this proof:

Lemma 3.2 (Monotonicity). *If F is strategy-proof and $F(\sigma) = a$, and τ satisfies for all alternative x and all voters i that*

$$\sigma_i(a) \geq \sigma_i(x) \Rightarrow \tau_i(a) \geq \tau_i(x)$$

Then $F(\tau) = a$

Note that $\sigma = (\sigma_1 \ldots \sigma_n)$ and $\tau = (\tau_1 \ldots \tau_1)$.

Proof: It suffices to prove this lemma assuming that $\sigma_i = \tau_i$ for all $i > 1$ because we can change voters one by one to get any desired configuration.

Assume by contradiction that $F(\tau) = b \neq a$

Then $\sigma_1(b) \leq \sigma_1(a)$. (otherwise if $\sigma_1(b) > \sigma_1(a)$ then voter 1 would want to cheat and vote τ.)

Then $\tau_1(b) \leq \tau_1(a)$ by the conditions of the lemma.

Which implies that $F(\tau_1) = a$ (otherwise, a voter with preference τ_1 will vote σ_1 to get a be the winner) □

Lemma 3.3 (Pareto). *If for all i, $\sigma_i(a) > \sigma_i(b)$ and F is strategy-proof and onto,*

Then $F(\sigma) \neq b$

Proof: Assume by contradiction that $F(\sigma) = b$

Since F is onto, there exists τ such that $F(\tau) = a$

Let

$$\sigma'_i = \begin{pmatrix} a \\ b \\ \text{ordered like } \sigma_i \end{pmatrix}$$

By Lemma 3.2 applied to the b preferences, $F(\sigma') = F(\sigma) = b$

But by Lemma 3.2 applied to the a preferences, $F(\sigma') = F(\tau) = a$ as well.

This is a contradiction, so $F(\sigma) = a$ □

With these two lemmas, we will first prove the GS theorem for two voters and then we will prove it for the general case.

Proof (GS theorem with 2 voters): Let there be two voters, with option ranking u and v as follows:

$$u = \begin{pmatrix} a \\ b \\ others \end{pmatrix}, v = \begin{pmatrix} b \\ a \\ others \end{pmatrix}$$

From Lemma 3.3, we have $F(u, v) \in \{a, b\}$. Without loss of generality, we assume $F(u, v) = a$.

First we prove that for all v' with b at the top of the ordering, $F(u, v') = a$. This is true because if there exists an ordering

$$v' = \begin{pmatrix} b \\ others \end{pmatrix}$$

such that $F(u, v') = b$, then the second voter will lie about his true ordering and vote v', and therefore F would not be strategy-proof. In particular this is true for v' which ranks a at the bottom.

Now, by Lemma 3.2, for all u' with a at the top of the ranking, and all v'', $F(u', v'') = a$.

To finish, we need to show that it does not matter that it is option a that is on top of the ranking:

Let A_1 = set of alternatives x such that if x is at the top of voter 1's ranking, the outcome is x.

Let A_2 = set of alternatives y such that if y is at the top of voter 2's ranking, the outcome is y.

Notice that $A_1 \cup A_2 = [k]$ because any of the options can be the outcome. We now argue that A_2 is empty. Indeed, let $b \neq a$ then if b is at the top of voter 2 and a is at the top of voter 1 then the outcome is a. Therefore A_2 cannot contain any element different than a. Fixing $b \neq a$ the same argument implies that a cannot belong to A_2 so so $A_1 = [k]$ and $A_2 = \emptyset$. Therefore F must be a dictator. □

Now we can build on this to get the general case. We will call the 2-voter case of the GS theorem Lemma 3.4. Now we prove the ≥ 3 voter case:

Proof: We will do this proof by induction on the number of voters n.

Assume that F is strategy-proof and onto.

Let $g(u, v) = F(u, v, v, \ldots, v)$. We need to prove that g is onto and strategy-proof.

- g is onto:

  Lemma 3.3 implies that that $F(u, u, \ldots, u)$ = top alternative of u, so $g(u, u)$ = top alternative of u and so g must also be onto.

- g is strategy-proof:

  - Voter 1 cannot manipulate g by definition
  - Voter 2 cannot manipulate g because otherwise there must exist a u, v, v' such that $v(g(u, v')) > v(g(u, v))$. Define $u_k = (u, k \times v', (n - k - 1) \times v)$. For voter k to manipulate g, there must exist a k where $v(g(u_{k+1})) > v(g(u_k))$ which is a contradiction to the fact that F is strategy proof.
  - Therefore, g is strategy proof.

With g strategy-proof and onto, by Lemma 3.4 g must be a dictator. Now we prove that F is also a dictator.

- if g is a dictator on voter 1, then by the monotonicity lemma, F is also a dictator on voter 1.

- Assume g is a dictator on voter 2.

  Fix $u*$ and look at $h(v_2, \ldots, v_n) = F(u*, v_2, \ldots, v_n)$. h is onto and strategy-proof, so it is dictatorial. Without loss of generality, assume 2 is the dictator and fix $v_3, \ldots, v_n$.

  Then $z(u, v) = F(u, v, v_3, \ldots, v_n)$ is onto and strategy-proof and voter 1 cannot be the dictator. So, z is a dictator on voter 2 implies that F is a dictator on voter 2.

□

References

[1] Gibbard, 1973 A. Gibbard, "Manipulation of voting schemes: a general result," Econometrica 41 (1973), pp. 587601.

[2] Satterthwaite, 1975 M.A. Satterthwaite, "Strategy-proofness and Arrows conditions: existence and correspondence theorems for voting procedures and social welfare functions," Journal of Economic Theory 10 (1975), pp. 187217.

[3] Lars-Gunnar Svensson, 1999. "The Proof of the Gibbard-Satterthwaite Theorem Revisited", preprint.