Scribe Notes for Mossel's CS294-063/Stat206A, October 7th

Gregory Valiant

October 14, 2010

This lecture covered *Arrow's Impossibility Theorem* and its generalization *Wilson's Theorem.* Informally, this theorem states that if we have $n \geq 3$ voters trying to rank $k \geq 3$ alternatives, then the dictator function is the only way to aggregate each voter's ranking to yield a complete ranking of the alternatives that satisfy:

- the relative positions of alternatives a and b depend only on each voter's relative ranking of a and b,
- if all voters prefer a to b, then a must be above b in the final ranking.

To motivate the intuition behind Arrow's theorem, we consider Condorcet's Paradox, defined by the Marquis de Condorcet in 1785 in his *Essay on the Application of Analysis to the Probability of Majority Decisions* (the same essay that outlined his "Jury Theorem").

Example 1 (Condorcet's Paradox)**.** *Consider the preferences of voters v_1, v_2, and v_3 for alternatives a, b, c, where v_1 ranks $a > b > c$, v_2 ranks $b > c > a$, and v_3 ranks $c > a > b$. A majority $(2/3)$ of the voters rank $a > b$, but similarly, $2/3$ of the voters also rank $b > c$ and $c > a$, and thus there seems to be no rational ranking of a, b, c in a manner consistent with the voters' preferences.*

Before formally stating Arrow's Impossibility Theorem, we will need some notation and definitions. For n voters ranking k alternatives, let the ranking submitted by voter i be denoted $\sigma_i \in S(k)$, and let $\sigma := (\sigma_1, \ldots, \sigma_n)$ denote the list of rankings given by the n voters.

We will now focus on the instance when $k = 3$, and we will denote the three alternatives as a, b, and c. For ease of notation, we will map each voter's ranking, σ_i to a tuple $(x_i, y_i, z_i) \in \{-1, 1\}^3$, where $x_i = 1$ if $\sigma_i(a) > \sigma_i(b)$, and -1 otherwise. Similarly, $y_i = 1$ if $\sigma_i(b) > \sigma_i(c)$ and -1 otherwise, and $z_i = 1$ if $\sigma_i(c) > \sigma_i(a)$, and -1 otherwise. Finally, we will let $x := (x_1, \ldots, x_n)$, $y := (y_1, \ldots, y_n)$, and $z := (z_1, \ldots, z_n)$.

Remark 2. *Note that (x_i, y_i, z_i) corresponds to a σ_i if, and only if, $(x_i, y_i, z_i) \in \{-1, 1\}^3 \setminus \{(1, 1, 1), (-1, -1, -1)\}$.*

We now state the main definitions:

Definition 3. *A* constitution *is a map $F : S(3)^n \to \{-1, 1\}^3$.*

The first coordinate of the image of F is the a vs. b outcome, the second is the b vs. c outcome and the third the c vs. a coordinate.

We now define three basic properties that, intuitively, are reasonable guidelines that we might hope "good" constitutions satisfy.

Definition 4. *A constitution F is* Transitive *if, for all sets of rankings σ, $F(\sigma) \in \{-1,1\}^3 \setminus \{(1,1,1),(-1,-1,-1)\}$; that is, F is transitive if for all σ, $F(\sigma)$ is a proper ranking of the alternatives.*

Definition 5. *A constitution F is* Independent of Irrelevant Alternatives (IIA) *if there exist functions $f,g,h : \{-1,1\}^n \to \{-1,1\}$ such that for all σ, we have $F(\sigma) = (f(x(\sigma)), g(y(\sigma)), h(z(\sigma)))$.*

Definition 6. *A constitution F satisfies* Unanimity *if $\sigma_1 = \sigma_2 = \ldots = \sigma_n \Rightarrow F(\sigma) = \sigma_1$.*

Example 7.

- *The* dictator *function $F(\sigma) = \sigma_i$ clearly satisfies Transitivity, IIA, and Unanimity.*

- *The* majority *function $F(\sigma) = (Maj(x), Maj(y), Maj(z))$, where $Maj(v) = 1$ if v contains at least as many 1s as -1s, satisfies IIA and Unanimity, but, as Condorcet's Paradox demonstrates, does not satisfy Transitivity.*

- *The function $F(\sigma) = \tau$, where τ is the most frequently occurring permutation in σ satisfies Transitivity, Unanimity, but not IIA, as can be seen by considering the outcomes corresponding to $\sigma = ((a,b,c),(a,b,c),(b,a,c),(b,c,a))$ and $\sigma = ((a,b,c),(a,c,b),(b,a,c),(b,a,c))$. $F(\sigma) = (1,1,-1) \neq (-1,1,-1) = F(\sigma')$, yet in both sets of rankings, $a > b$ for the first two voters, and $b < a$ for the second two voters, but the first coordinate of $F(\sigma)$ and $F(\sigma')$ differ, thus F is not IIA.*

We now state Arrow's "Impossibility" Theorem:

Theorem 1 (Arrow's "Impossibility" Theorem)**.** *Any constitution F on $k \geq 3$ alternatives that is* transitive, IIA, *and satisfies* unanimity *is a* dictator *function; that is there exists some $i \in \{1,\ldots,n\}$ such that for all σ, $F(\sigma) = \sigma_i$.*

The following definition will be helpful in our proof of Arrow's theorem.

Definition 8. *Voter* 1 *is* pivotal *for $f : \{-1,1\}^n \to \{-1,1\}$ (denoted $I_1(f) > 0$)if there exist some $x_2,\ldots,x_n$ such that $f(-1,x_2,\ldots,x_n) \neq f(1,x_2,\ldots,x_n)$; we say voter i is pivotal if the analogous statement holds.*

Note that saying that voter i is pivotal for f exactly corresponds to saying that for the function f, the i^{th} variable has nonzero influence.

The proof of Arrow's theorem follows easily from the following lemma, due to Barbera;

Lemma 9 (Barbera '82)**.** *Any IIA constitution $F = (f,g,h)$ on 3 alternatives that has $I_1(f) > 0$ and $I_2(g) > 0$ is non-transitive.*

Proof. Since $I_1(f) > 0$ and $I_2(g) > 0$, there exist $x_2,\ldots,x_n$ and $y_1,y_3,y_4,\ldots,y_n$ such that

$$f(1,x_2,\ldots,x_n) \neq f(-1,x_2,\ldots,x_n) \text{ and } g(y_1,1,y_3,\ldots,y_n) \neq g(y_1,-1,y_3,\ldots,y_n).$$

Let $v = h(-y_1,-x_2,-x_3,\ldots,-x_n)$, and note that we can choose x_1, y_2 such that $f(x_1,x_2,\ldots,x_n) = g(y_1,y_2,\ldots,y_n) = v$. To conclude, note that the rankings given by $x = (x_1,\ldots,x_n), y = (y_1,\ldots,y_n)$, and $z := (-y_1,-x_2,-x_3,\ldots,-x_n)$ are valid rankings, since for all i $(x_i,y_i,z_i) \notin \{(-1,-1,-1),(1,1,1)\}$, yet $f(x) = g(y) = h(z)$, thus F is not transitive. □

Proof of Theorem 1: We first prove the theorem for the case that there are 3 alternatives. Since F is IIA, by assumption, without loss of generality let $F = (f, g, h)$. Let $I(f) := \{\text{pivotal voters for } f\}$. Since F satisfies Unanimity, none of f, g, or h can be constant functions, and thus $I(f), I(g), I(h)$, are all nonempty. Assume for the sake of contradiction that it is not the case that $|I(f)| = 1$ and $I(f) = I(g) = I(h)$, then there is a pair of voters $i \neq j$ which each are, respectively, pivotal for functions f, g or f, h or g, h; by Barbera's lemma F is not transitive, a contradiction, thus we conclude that $F(\sigma) = G(\sigma_i)$, for some function G. By the unanimity condition, G must be the identity function, so $F(\sigma) = \sigma_i$, as desired.

We now prove the theorem for $k > 3$ alternatives. First observe that given a transitive IIA function F for $k > 3$ alternatives that satisfies unanimity, the restriction of F to any subset of just 3 alternatives will be a transitive IIA function on three alternatives that satisfies unanimity, and thus, from the above, we know that the restriction of F must be a dictator function. All that remains is to show that for all pairs of alternatives $(a, b), (a', b')$, for distinct a, b, a', b', the *same* dictator decides the relative position of a, b as a', b'. To see this, let voter i be the dictator that decides the relative positions of a, b, a', and voter j the dictator that decides the relative positions of a, b, b', and note that $i = j$, because the relative positions of a, b can not be decided by different dictators. □

Given Arrow's theorem, a natural direction is to relax our notions of a *reasonable* aggregation function. Along these lines, a natural question is:

What happens if we remove the unanimity constraint?

The first easy observation is that we only used the unanimity assumption in our proof of Arrow's theorem in two places; concluding that $I(f), I(g), I(h)$ are all nonempty, and in the final step where we say that since $F(\sigma) = G(\sigma_i)$ it must be the case that G is the identity function. If, instead of assuming that F satisfies unanimity, we assume that for all pairs of alternatives (a, b), there exists voter rankings σ, σ' such that $a > b$ in $F(\sigma)$ but $a < b$ in $F(\sigma')$, then we can still conclude that $I(f), I(g), I(h)$ are nonempty, and the proof goes through to yield that $F(\sigma) = G(\sigma_i)$ for some voter i and some function G. What functions G can we use without violating the IIA condition?

Proposition 10. *Given constitution $F(\sigma) = G(\sigma_i)$ that is IIA and transitive, and for which for every pair of alternatives (a, b), there exists some σ_i, σ_i' for which $G(\sigma_i)$ ranks $a > b$ and $G(\sigma_i')$ ranks $a < b$, then either $G(\sigma_i) = \sigma_i$, or $G(\sigma_i) = -\sigma_i$, where $-\sigma_i$ denotes the "reverse" of ranking σ_i.*

Proof. Note that $G : \{-1, 1\}^{|S(k)|} \to \{-1, 1\}^{|S(k)|}$. Additionally, since F, and thus G is transitive, we can write $G = (g_1, \ldots, g_{|S(k)|})$, where $g_i : \{-1, 1\} \to \{-1, 1\}$. Note that the conditions of the proposition now imply that g_i can not be the constant function, and thus $g_i = \pm Id$. Phrased such, the claim now amounts to showing that for all i, j, $g_i = g_j$ (ie either they are all the identity function, or all (-1) times the identity function). Assume for the sake of contradiction that there exist i, j s.t. $g_i = -g_j$.

For clarity of notation, we replace the subscripts i, j by the pair of alternatives to which they refer, thus $g_{a,b}$ indicates the relative ranking of a, b. Thus we have two pairs (a, b) and (c, d) such that $g_{a,b} = -g_{c,d}$. We now claim that we can find a triple r, s, t s.t. $g_{r,s} = -g_{s,t}$. Indeed, consider $g_{a,b}$ and $g_{c,d}$, and assume without loss of generality that $a \neq d$. If $c = b$, then we have found such a triple. Otherwise, consider $g_{a,b}$ and $g_{b,d}$; if $g_{a,b} = -g_{b,d}$, then we have found such a triple, otherwise we must have $g_{b,d} = -g_{c,d}$.

To conclude, given $g_{r,s} = -g_{s,t}$, without loss of generality assume that $g_{r,s} = Id$ and that $g_{s,t} = -Id$. We now consider the two cases that $g_{r,t} = Id$ and $g_{r,t} = -Id$. First, consider $g_{r,t} = Id$: consider $G((t > r > s))$: in the result, it must be that $r > s$, $s > t$ and $t > r$, which is not a valid ordering. Similarly, if $g_{r,t} = -Id$: consider $G((s > t > r))$: in the result, it must be that $s > r$, $t > s$ and $r > t$, which is not a valid ordering. □

We now characterize the set of constitutions that are IIA and transitive (and drop the condition that every pair a, b of alternatives can be ranked in both relative orderings $a > b$ and $b > a$).

Definition 11. *For a constitution F, we write $A >_F B$ if for all σ and all alternatives $a \in A$ and $b \in B$, it holds that $F(\sigma)$ ranks a above b.*

Theorem 2 (Wilson, '72, Mossel '10)**.** *A constitution F on k alternatives satisfies IIA and transitivity if, and only if there exists a partition of the alternatives into sets $A_1, \ldots, A_s$ such that:*

- *$A_1 >_F A_2 >_F \ldots >_F A_s$,*
- *If $|A_r| > 2$ then F restricted to A_r is a dictator on some voter j, in that $F^{A_r}(\sigma) = \pm\sigma_i^{A_r}$, where the superscript A_r denotes the restriction to the alternatives in A_r.*

Clearly any function of the above form is IIA and transitive, so it remains to prove that if F is IIA and transitive, then it has the claimed form. The following definitions will be helpful in our proof of Wilson's theorem:

Definition 12. *For a constitution F, and two alternatives a, b write $a >_F b$ if, for all σ, $F(\sigma)$ ranks $a > b$. Write $a \sim_F b$ if there exist σ, σ' such that $F(\sigma)$ ranks $a > b$ and $F(\sigma')$ ranks $a < b$.*

Lemma 13. *For a transitive and IIA function F, if there exists two sets of voter rankings σ, σ' for which $F(\sigma)$ ranks $a > b$, and $F(\sigma')$ ranks $b > c$ then there exists a set of voter ranking τ such that in $F(\tau)$, $a > c$.*

Proof. Letting $x, y \in \{-1, 1\}^n$ denote the vectors of relative preferences between a, b and b, c, respectively, consider the voter rankings in which $x = x_\sigma$, and $y = y_{\sigma'}$. We can extend these preference lists into a set of valid rankings by setting the relative preferences between a, c to be $z = -x$, and the preferences between all other pairs to be some arbitrary unanimous ranking. Thus we have constructed a set of voter preferences τ which agrees with σ on the relative ranking $a > b$ and agrees with σ' on the relative ranking $b > c$. □

Corollary 14. *For a transitive, IIA function F, the relations $>_F$, and $\sim_F$ are transitive. Additionally, if $a >_F b$ and $a \sim_F c$ and $b \sim_F d$ then $c >_F d$..*

Proof. The transitivity of $>_F$ is obvious, and the transitivity of $\sim_F$ follows immediately from Lemma 13. If $a >_F b$ and $a \sim_F c$, then $c >_F b$, since otherwise, given an instance σ for which $F(\sigma)$ ranks $c < b$, by Lemma 13 we can compose it with an instance for which $a < c$, yielding an instance in which $a < b$, contradiction $a >_F b$. Thus if $a >_F b$, $a \sim_F c$, $b \sim_F d$ then $c >_F b$. Applying the same argument to $c >_F b$ and $b \sim_F d$ yields that $c >_F d$, as desired. □

The proof of Wilson's theorem now follows easily from the above corollary and Arrow's theorem. *Proof of Theorem 2:* We first leverage Corollary 14 to show that there exists a partition of the alternatives into sets $A_1 >_F A_2 >_F \ldots >_F A_s$. Indeed, for a given alternative a, let $A := \{b :$

$a \sim_F b\}$. For $a' \not{i}nA$, define the set A' analogously, and note that since $a' \notin A$, without loss of generality we may assume that $a <_F a'$. By Corollary 14, $A <_F A'$, and thus the construction of the partitions is well-defined. To conclude, we note that for every pair of alternatives a, b that lie in the same partition, from the definition of $\sim_F$, there exist outcomes for which F ranks $a > b$ and $b > a$, and thus we may apply Arrow's theorem to the restriction of F to each partition, yielding that the restriction of F to a partition A_i is a dictator function $G(\sigma_i)$, and thus by Proposition 10, the restriction of F to any partition is either the dictator function σ_i or $-\sigma_i$, as claimed. □

We concluded with a final remark that if voters don't need to provide strict ordering, and instead can indicate ties, then one-sided versions of Arrow's and Wilson's theorem hold—though considering such a general settings seems to only obfuscate the interesting characterizations.