Arrow's Theorem, Noise Stability, and Dictator Testing (lecture notes)

Course: Analysis of Boolean Functions, Autumn 2025, University of Chicago
Instructor: William Hoza (`williamhoza@uchicago.edu`)

1 The spoiler effect

In a typical election, each voter picks their favorite candidate, and whichever candidate gets the most votes wins. This is called "the plurality voting system." It's quite a sensible system if there are only two candidates. However, when there are three or more candidates, the plurality voting system suffers from the so-called *spoiler effect*, which is when a candidate doesn't get enough votes to win, but they do get enough votes to change the outcome of the election.

For example, suppose a club is deciding when to hold meetings. Most people are pushing for afternoon meetings, but a few people say they would prefer morning meetings, so they decide to vote on the matter. Crucially, there are *three* possible meeting times listed on the ballot: 8am, 2pm, and 2:15pm. 33% of participants vote for 2pm meetings, 33% of participants vote for 2:15pm meetings, and 34% of participants vote for 8am meetings. Therefore, meetings are scheduled for 8am. This outcome seems somehow wrong.

Several alternative voting systems have been proposed to try to address this issue. Here are a few examples.

- **Instant-runoff voting**: Each voter ranks the candidates. Whichever candidate is ranked first by the fewest voters is eliminated. We repeat until only one candidate remains; that candidate is the winner.
- **Borda count voting**: Each voter ranks the candidates. Whoever the voter ranks last receives zero points, second-to-last receives one point, etc. Whichever candidate receives the most points wins the election.
- **Copeland-Llull voting**: Each voter ranks the candidates. For each pair of candidates, whichever candidate is preferred by a majority of voters receives a point. Whichever candidate receives the most points wins the election.

Do any of these voting methods eliminate the spoiler effect? Let's get a bit more mathematical. We'll focus on the case that there are three candidates, denoted a, b, and c. Let S_3 denote the set of permutations on $\{a, b, c\}$. A ranked voting system can be modeled as a function $F\colon S_3^n \to S_3$. If σ_i is voter i's ranking of the three candidates, then $F(\sigma_1, \sigma_2, \dots, \sigma_n)$ is how "society as a whole" ranks the candidates.

To avoid the spoiler effect, ideally, we would like to find a ranked voting system that satisfies the *independence of irrelevant alternatives* criterion. Basically, this criterion says that the quality of candidate c shouldn't have any effect on whether candidate a beats candidate b. In joke form, it can be expressed as follows, paraphrasing Sidney Morgenbesser.

> **Waiter:** "Would you like apple pie, blueberry pie, or cherry pie?"
>
> **Customer:** "I'll have apple pie, please."
>
> **Waiter:** "Are you sure? The cherries are excellent this time of year."
>
> **Customer:** "Oh! In that case, I'll have blueberry pie."

Mathematically, let us use the following notation. If $\sigma \in S_3$ and p, q are two distinct candidates

$p, q \in \{a, b, c\}$, then define

$$\sigma^{pq} = \begin{cases} +1 & \text{if } p \text{ is ranked higher than } q \text{ under } \sigma \\ -1 & \text{otherwise.} \end{cases}$$

If $x = (\sigma_1, \ldots, \sigma_n) \in S_3^n$, then we define $x^{pq} = (\sigma_1^{pq}, \ldots, \sigma_n^{pq}) \in \{\pm 1\}^n$. The independence of irrelevant alternatives condition can be stated as follows.

Definition 1.1 (Independence of Irrelevant Alternatives)**.** Let $F \colon S_3^n \to S_3$. We say that F satisfies *independence of irrelevant alternatives* if there exist functions $f, g, h \colon \{\pm 1\}^n \to \{\pm 1\}$ such that for every $x \in S_3^n$, we have

$$F(x)^{ab} = f(x^{ab})$$
$$F(x)^{bc} = g(x^{bc})$$
$$F(x)^{ca} = h(x^{ca}).$$

Arrow's theorem says that unfortunately, there is no good way to satisfy independence of irrelevant alternatives. To be precise, we have the following.

Theorem 1.2 (Arrow's Theorem)**.** *Let $F \colon S_3^n \to S_3$. Assume that F satisfies independence of irrelevant alternatives. Then at least one of the following holds:*

- *There exists $i \in [n]$ such that for every $\sigma_1, \ldots, \sigma_n \in S_3$, we have $F(\sigma_1, \ldots, \sigma_n) = \sigma_i$.*
- *There exists $\sigma \in S_3$ such that $F(\sigma, \sigma, \ldots, \sigma) \neq \sigma$.*

The first possible conclusion says that voter i has complete control of the outcome of the "election" and is effectively a dictator. The second possible conclusion says that even when the voters hold an opinion unanimously, society as a whole might hold the opposite opinion. Neither option is appealing.

2 Condorcet's paradox

Let's play the role of the skeptic and try to come up with a counterexample to Arrow's theorem. The most natural choice for the functions f, g, h in the definition of independence of irrelevant alternatives would be $f = g = h = \mathsf{Maj}_n$. That is, if a majority of voters prefers candidate p over candidate q, then we should consider society as a whole to prefer candidate p over candidate q. What's wrong with this proposal?

The difficulty is that "society" can sometimes have *cyclic preferences.* For example, suppose there are 3 voters with the following preferences.

- Voter 1: $a \succ b \succ c$
- Voter 2: $b \succ c \succ a$
- Voter 3: $c \succ a \succ b$.

Then "society" prefers a over b, and b over c, and c over a. Each individual voter is rational, but society as a whole is irrational. This is Condorcet's paradox. It demonstrates that there is no well-defined $F \colon S_3^n \to S_3$ corresponding to the choice $f = g = h = \mathsf{Maj}_n$.

To prove Arrow's theorem, we will prove that Condorcet's paradox still occurs even if we use a different function f in place of Maj_n. Essentially the only way to guarantee that Condorcet's paradox never occurs is to use the function $f(x) \equiv x_i$ for some i. In this context, the function $f(x) = x_i$ is called the i-th "dictator function."

3 Noise stability

The proof of Arrow's theorem is based on analyzing the *noise stability* of f.

Definition 3.1 (Correlated strings)**.** Let $\rho \in [-1,1]$ and let x and y be jointly distributed random variables, each taking values in $\{\pm 1\}^n$. We say that x and y are *ρ-correlated* if:

- The n random variables $(x_1, y_1), \dots, (x_n, y_n)$ are independent.
- For each i, we have $\mathbb{E}[x_i] = \mathbb{E}[y_i] = 0$ and $\mathbb{E}[x_i \cdot y_i] = \rho$.

Equivalently, we can imagine first sampling $x \in \{\pm 1\}^n$ uniformly at random. Then, for each coordinate independently, we let y_i be x_i with probability $(1+\rho)/2$ and $-x_i$ with probability $(1-\rho)/2$.

Definition 3.2 (Noise stability)**.** Let $\rho \in [-1,1]$ and $f\colon \{\pm 1\}^n \to \mathbb{R}$. The *noise stability of f at ρ* is given by

$$\mathrm{Stab}_\rho[f] = \mathop{\mathbb{E}}_{\substack{(x,y)\\ \rho\text{-correlated}}}[f(x) \cdot f(y)].$$

Lemma 3.3. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. Then*

$$\Pr_{x \in S_3^n}[f(x^{ab}) = f(x^{bc}) = f(x^{ca})] = \frac{1}{4} + \frac{3}{4}\mathrm{Stab}_{-1/3}[f].$$

Note that the condition "$f(x^{ab}) = f(x^{bc}) = f(x^{ca})$" describes Condorcet's paradox.

Proof. Let $\mathsf{EQ}_3\colon \{\pm 1\}^3 \to \{0,1\}$ indicate whether the given three bits are all equal. The Fourier expansion of EQ_3 is

$$\begin{aligned}\mathsf{EQ}_3(y_1,y_2,y_3) &= \left(\frac{1+y_1}{2}\right)\cdot\left(\frac{1+y_2}{2}\right)\cdot\left(\frac{1+y_3}{2}\right) + \left(\frac{1-y_1}{2}\right)\cdot\left(\frac{1-y_2}{2}\right)\cdot\left(\frac{1-y_3}{2}\right)\\ &= \frac{1}{4} + \frac{1}{4}y_1y_2 + \frac{1}{4}y_1y_3 + \frac{1}{4}y_2y_3.\end{aligned}$$

Therefore,

$$\Pr[f(x^{ab}) = f(x^{bc}) = f(x^{ca})] = \frac{1}{4} + \frac{1}{4}\mathbb{E}[f(x^{ab})f(x^{bc})] + \frac{1}{4}\mathbb{E}[f(x^{ab})f(x^{ca})] + \frac{1}{4}\mathbb{E}[f(x^{ab})f(x^{ca})].$$

The triples $(x_i^{ab}, x_i^{bc}, x_i^{ca})$ are independent from one i to the next. Each triple is selected uniformly at random from $\{\pm 1\}^3 \setminus \{(+1,+1,+1),(-1,-1,-1)\}$. Consequently, each individual bit is uniform random, and

$$\mathbb{E}[x_i^{ab}x_i^{bc}] = \mathbb{E}[x_i^{bc}x_i^{ca}] = \mathbb{E}[x_i^{ab}x_i^{ca}] = \frac{2-4}{6} = -\frac{1}{3}. \qquad \square$$

The next step is a Fourier formula for noise stability. We use the following notation.

Definition 3.4 (Level-k weight)**.** Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ and let $0 \le k \le n$. We define

$$W^k[f] = \sum_{\substack{S \subseteq [n]\\ |S| = k}} \widehat{f}(S)^2.$$

Lemma 3.5 (Fourier formula for noise stability)**.** *For any $f\colon \{\pm 1\}^n \to \mathbb{R}$ and any $\rho \in [-1,1]$, we have*

$$\mathrm{Stab}_\rho[f] = \sum_{k=0}^{n} \rho^k \cdot W^k[f].$$

Proof.

$$\mathrm{Stab}_\rho[f] = \mathbb{E}[f(x)\cdot f(y)] = \sum_{S,T\subseteq[n]} \widehat{f}(S)\cdot\widehat{f}(T)\cdot\mathbb{E}[x^S\cdot y^T] = \sum_{S\subseteq[n]}\widehat{f}(S)^2\cdot\prod_{i\in S}\mathbb{E}[x_i\cdot y_i]. \qquad \square$$

In the case that f is $\{\pm1\}$-valued, there is way to make the formula a bit easier to remember / interpret. By Parseval's theorem, $\sum_{S\subseteq[n]}\widehat{f}(S)^2 = 1$. Therefore, we can interpret $\widehat{f}(S)^2$ as a probability.

Definition 3.6 (Spectral sample)**.** Let $f\colon\{\pm1\}^n\to\{\pm1\}$. Then $\mathcal{S}_f$ is the distribution over subsets of $[n]$ in which $\Pr[S] = \widehat{f}(S)^2$.

The formula for noise stability becomes $\mathrm{Stab}_\rho[f] = \mathbb{E}_{S\sim\mathcal{S}_f}[\rho^{|S|}]$.

Corollary 3.7. *Let $f\colon\{\pm1\}^n\to\{\pm1\}$. Sample $x\in S_3^n$ uniformly at random. Then*

$$\Pr[f(x^{ab}) = f(x^{bc}) = f(x^{ca})] \geq \frac{2}{9}\cdot(1-W^1[f]).$$

Proof.

$$\begin{aligned}
\frac{1}{4}+\frac{3}{4}\mathrm{Stab}_{-1/3}[f] &= \frac{1}{4}+\frac{3}{4}\sum_{k=0}^{n}\left(-\frac{1}{3}\right)^k\cdot W^k[f] \\
&\geq \frac{1}{4}+\frac{3}{4}\cdot\left(-\frac{1}{3}\right)\cdot W^1[f]+\frac{3}{4}\cdot\left(-\frac{1}{3}\right)^3\cdot\sum_{k=3}^{n}W^k[f] \\
&\geq \frac{1}{4}-\frac{1}{4}\cdot W^1[f]-\frac{1}{36}\cdot(1-W^1[f]) \\
&= \frac{2}{9}-\frac{2}{9}\cdot W^1[f].
\end{aligned} \qquad \square$$

Corollary 3.8. *Let $f\colon\{\pm1\}^n\to\{\pm1\}$. Suppose that for every $x\in S_3^n$, the values $f(x^{ab})$, $f(x^{bc})$, $f(x^{ca})$ are not all equal. Then there is some i such that $f\equiv\chi_i$ or $f\equiv-\chi_i$.*

Proof. By Corollary 3.7, we have $W^1[f]=1$. By Parseval's theorem, this implies that f has the form $f(x)=\sum_{i=1}^n\widehat{f}(i)\cdot x_i$. Plugging in $x_i=\mathrm{sign}(\widehat{f}(i))$, we see that $\sum_{i=1}^n|\widehat{f}(i)| = 1 = \sum_{i=1}^n\widehat{f}(i)^2$. This implies that $\widehat{f}(i)=\pm1$ for some i and $\widehat{f}(i)=0$ for all other i. $\square$

Proof of Arrow's Theorem. Assume that $F(\sigma,\sigma,\dots,\sigma)=\sigma$ for every $\sigma\in S_3$. Let f,g,h be the functions in the definition of independence of irrelevant alternatives. Since F is well-defined, Condorcet's paradox never occurs, i.e., for every $x\in S_3^n$, the values $f(x^{ab}),g(x^{bc}),h(x^{ca})$ are not all equal.

Let us show that $f\equiv g\equiv h$. For any $y\in\{\pm1\}^n$, there exists $x\in S_3^n$ such that $x^{ab}=y$, $x^{bc}=-y$, and $x^{ca}=(g(-y),g(-y),\dots,g(-y))$. Therefore, the three values $f(y)$, $g(-y)$, and $h(g(-y),g(-y),\dots,g(-y))$ are not all equal. Because of the assumption $F(\sigma,\sigma,\dots,\sigma)=\sigma$, we must have $h(g(-y),g(-y),\dots,g(-y)) = g(-y)$. Therefore, $f(y)=-g(-y)$. By symmetry, we also have $h(y)=-g(-y)$, hence $f(y)=h(y)$. Again, symmetry gives us $f(y)=g(y)=h(y)$.

From here, Corollary 3.8 tells us that we either have $f\equiv\chi_{\{i\}}$ or else $f\equiv-\chi_{\{i\}}$. The second case would violate the assumption $F(\sigma,\sigma,\dots,\sigma)=\sigma$, so $f\equiv\chi_{\{i\}}$, which implies that $F(\sigma_1,\dots,\sigma_n)=\sigma_i$ for every $(\sigma_1,\dots,\sigma_n)\in S_3^n$. $\square$

4 Dictator testing

When Arrow gives you lemons, make lemonade.

Theorem 4.1 (Dictator testing)**.** *Suppose we have query access to an unknown function $f\colon\{\pm1\}^n\to\{\pm1\}$. There is a randomized algorithm that makes 3 queries to f and has the following properties.*

- *If f is a dictator function, then the algorithm accepts with probability* 1.
- *If f is ε-far from every dictator function, then the algorithm rejects with probability* $\Omega(\varepsilon)$.

Proof. With probability 1/2, perform the BLR linearity test. Otherwise, we do the following, inspired by Arrow's theorem.

- Sample $x \in S_3^n$ uniformly at random.
- If $f(x^{ab}) = f(x^{bc}) = f(x^{ca})$ (Condorcet's paradox), reject. Otherwise, accept.

If f is a dictator function, clearly the algorithm accepts with probability 1. Conversely, suppose the algorithm accepts with probability $1 - \delta$. Then the probability of Condorcet's paradox is at most 2δ. By Corollary 3.7, this implies $\frac{2}{9} \cdot (1 - W^1[f]) \leq 2\delta$, i.e., $W^1[f] \geq 1 - 9\delta$. Meanwhile, the BLR test accepts with probability at least $1 - 2\delta$, hence there is some $S \subseteq [n]$ such that $\widehat{f}(S) \geq 1 - 4\delta$. If $|S|$ were bigger than 1, then we would have

$$1 = \sum_{k=0}^{n} W^k[f] \geq (1 - 9\delta) + (1 - 4\delta)^2 = 2 - 17\delta + 16\delta^2,$$

which is a contradiction for small enough δ. Therefore, f is $O(\delta)$-close to a dictator function. $\square$

It turns out that dictator testing is a key ingredient in the famous PCP Theorem.