Condorcet Quintets — A Proof from the Book

Here's my favorite problem in voting theory.

> *Suppose we have an election in which voters provide ranked preferences over the candidates. For what k can we always choose k winners, such that no loser is ranked above all winners by a majority of voters?*

The question is a riff on *Condorcet's paradox*, which famously says that this can't be done with $k = 1$. Even though the paradox has been around for almost 250 years, this natural twist on it was posed only 15 years ago by Elkind, Lang, and Saffidine (ELS).

When I tell someone about this problem, it's fun to ask them to guess the answer, because it can be a window into the way they think. People well versed in the litany of negative results in voting theory often have the instinct that k should be nearly as large as the number of candidates m, maybe $\Omega(m)$. Friends who work in quantum computing are a little more optimistic, and might suggest $O(\sqrt{m})$ is possible. Others who think a lot about graphs (or have recently had the phrase *dominating sets* whispered in their ear) might guess $O(\log m)$, which ELS proved in their paper.

But it turns out you can do a lot better. You can get away with $k = 5$, no matter the number of candidates.

Before explaining how, a little history. In a paper a couple years ago with Charikar, Lassota, Vetta, and Wang (CLRVW), we pointed out that work by Jiang, Munagala, and Wang (JMW) on *approximately stable committees* already implies, as a black box, that $k = 32$ suffices. By generalizing some of their ideas, we got k down to 6, and Song, Nguyen, and Lin (SNL) improved it to 5. ELS also showed examples which require k to be at least 3, so the answer is 3, 4, or 5. Figuring out which is optimal remains open.

The upper bound papers (JMW, CLRVW, SNL) all work similarly: construct a "good" distribution over candidates, then use the probabilistic method to show that sampling from it has a chance of producing a Condorcet winning set. JMW use distributions called *stable lotteries*, introduced by Cheng, Jiang, Munagala, and Wang. CLRVW generalize these to what were later called *g-stable lotteries*. SNL use a distribution derived from a *Lindahl equilibrium*, and give a much simpler proof.

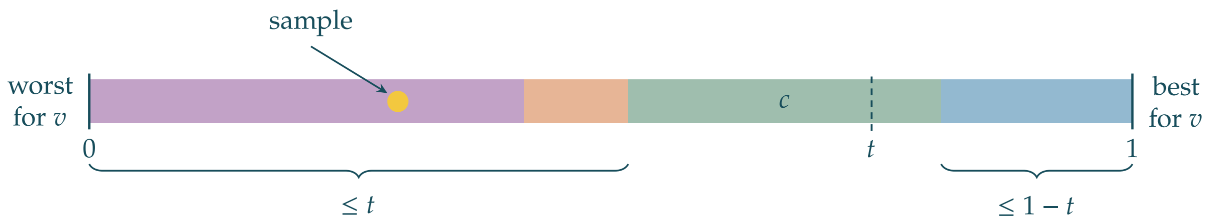
In a recent short note, we explained that since the existence of Lindahl equilibria is proved via Kakutani's fixed point theorem, peeling off that outer layer makes the proof even easier. But actually, the onion has another layer: if you peel off Kakutani, you get an amazingly simple and elegant proof using Brouwer's fixed point theorem. This version was suggested by a frontier model (as a peace offering for failing to solve the full problem), but I want to emphasize that it is just a rearrangement of the core ideas from SNL.

So without further ado, here's the proof.

Suppose we have an election in which, for every set S of at most k candidates, there is a candidate a_S whom at least an α fraction of voters rank above every candidate in S. Our goal is to derive the best upper bound on α as a function of k, and show that for $k \geq 5$, $\alpha < \frac{1}{2}$.

Given a distribution D over candidates, let $\widehat{D}$ be the distribution of a_S when $S \sim D^k$ (k iid samples from D). Because a_S is fixed by S, and the probability of each S is continuous as a function of D, $D \mapsto \widehat{D}$ is a continuous map from the set of distributions (a simplex) to itself. Brouwer's theorem tells us that this map has a fixed point D. In other words, if we sample $S \sim D^k$ and set $a = a_S$, then a also has distribution D (though it isn't independent of S).

Fix a voter v with preference $\succ_v$. To understand the probability that v prefers a to all of S, picture their ranking as an interval. Give each candidate a block whose width is their probability under D, and arrange the blocks from left to right, from least to most preferred, filling $[0, 1]$. Sampling from D is then just choosing a uniform point on the interval and seeing whose block it lands in.



Now fix some $t \in (0, 1)$, and let c be the candidate whose block occupies the point t on the interval. The blocks above and below c have total mass at most $1 - t$ and t respectively. If $a \succ_v S$, then either $a \succ_v c$, or $c \succ_v S$. Since a has distribution D and S has distribution D^k, these two events happen with probability at most $1 - t$ and t^k respectively. By a union bound,

$$\Pr_{S \sim D^k}[a \succ_v S] \leq \Pr_{a \sim D}[a \succ_v c] + \Pr_{S \sim D^k}[c \succ_v S] \leq (1 - t) + t^k.$$

Averaging over voters, the left-hand side is at least α, so $\alpha \leq 1 - t + t^k$. To get the strongest bound, set $t = k^{-1/(k-1)}$, giving $\alpha \leq 1 - (k-1)k^{-k/(k-1)}$, which is less than $1/2$ for $k \geq 5$.

Reflecting on points made by Terry Tao and others online, I found myself wondering what would have happened if Elkind, Lang, and Saffidine had posed their question 15 or 20 years later. It feels plausible that a frontier model would have given them a proof like the one above or even solved the problem outright, but maybe there was value in struggling with it more than we needed to.

Without that struggle, we might not have thought to generalize stable lotteries, or developed the visual language for interpreting randomized voting rules that grew alongside that work. Our theory of Lindahl equilibria would be less complete without SNL, which deepened our understanding of how to adapt the idea to settings with ordinal preferences. Their work also helped us understand what our generalization really should have been (in a way that was key to some new work, incidentally related to my last blog post). And selfishly, I would have missed out on the joy of working on a cool problem with brilliant collaborators and friends.

Would that theory have been developed anyway for other problems? What does all of this say about how we should handle what's coming? Like the question at the start of this blog, I don't have a complete answer, but there's an old theorem that I think could be a useful tool:

> *(Folklore) It's not just the destination, it's the journey and the friends we made along the way.*

On that note, thanks to Adrian and Alex for introducing me to this problem, and Moses, Kangning, Edith, Emin, Jamie, Thanh, Ariel, and Zohar for many wonderful conversations about it over the years.