A law of large numbers for weighted plurality

Joe Neeman

Abstract Consider an election between \( k \) candidates in which each voter votes randomly (but not necessarily independently) for a single candidate, and suppose that there is a single candidate that every voter prefers (in the sense that each voter is more likely to vote for this special candidate than any other candidate). Suppose we have a voting rule that takes all of the votes and produces a single outcome and suppose that each individual voter has little effect on the outcome of the voting rule. If the voting rule is a weighted plurality, then we show that with high probability, the preferred candidate will win the election. Conversely, we show that this statement fails for all other reasonable voting rules. This result is an extension of one by Häggström, Kalai and Mossel, who proved the above in the case \( k = 2 \).

1 Introduction

Consider an election between two candidates in which the voters vote independently at random, but with some small bias towards the first candidate. When the winner of the election is determined by a simple majority, Condorcet’s Jury Theorem (1725) implies that if there are many voters then the first candidate will win with overwhelming probability. What if we consider a voting rule other than a simple majority? Russo’s zero-one law (1982) can be interpreted as saying that for any voting rule under which each voter has only a small effect on the outcome, the same phenomenon occurs: as the number of voters increases, the probability that the first candidate will win converges to 1. From now on, we will refer to this property as aggregation of information: a voting rule aggregates information if it turns many small biases (in the same direction) into a large bias.
The assumption that every voter has only a small effect on the outcome has become reasonably common in the theory of social choice. Besides the obvious social desirability of such voting rules, their mathematical properties have proved useful for establishing quantitative versions of Arrow’s Theorem (1950, 2002) and the Gibbard–Satterthwaite Theorem (1973, 2010). Moreover, Kalai (2004) showed that when voters vote independently, information aggregates if and only individual power diminishes.

When the voters are not statistically independent, the phenomenon of information aggregation becomes more complicated. It is no longer true, in particular, that every reasonable voting rule aggregates well. In fact, Häggström et al. (2006) show that the only voting rules which aggregate information under arbitrary joint distributions of votes are weighted majority rules. The main contribution of this work is to extend their result to the case of more than two candidates.

2 Definitions and results

Throughout this work, we will consider only neutral voting rules, i.e., rules that do not have a built-in preference for any candidate. This is a common assumption, and its definition is standard (see, e.g. Brams and Fishburn 2002). In what follows, the notation \([k]_n\) stands for the set \([1, \ldots, k]\).

**Definition 2.1** A function \(f : [k]^n \rightarrow [k]\) is neutral if \(f(\sigma(x)) = \sigma(f(x))\) for all \(x \in [k]^n\) and all permutations \(\sigma\) on \([k]\), where \(\sigma(x)_i = \sigma(x_i)\).

Note that our voting rules are defined on \([k]^n\); that is, each voter votes for a single candidate. In other work on voting functions, voters are often allowed to supply a linear order on all of the candidates.

**Example 2.2** When \(k = 2\) and \(n\) is odd, then the simple majority function (for which \(f(x) = 1\) if \(#\{i : x_i = 1\} > #\{i : x_i = 0\}\)) is neutral. On the other hand, if \(n\) is even then in order to fully specify the simple majority function, we need to say what happens in the case of a tie; the choice of tie-breaking rule will determine whether the resulting function is neutral. For example, if we define \(f(x) = x_1\) for every tied configuration \(x\), then \(f\) is neutral. On the other hand, if \(f(x) = 1\) for every tied configuration \(x\), then \(f\) is not neutral.

The example can be extended to \(k \geq 3\). In this case, consider the tie-breaking rule \(f(x) = x_i\) where \(i \in [k]\) is the smallest possible number for which \(x_i\) is equal to one of the tied alternatives. This tie-breaking rule is neutral, and it is more natural than setting \(f(x) = x_1\) because it guarantees that the output of \(f\) is one of the tied alternatives.

In a later section, we will also consider monotone voting rules: a voting rule is monotone if extra votes cannot hurt a candidate’s cause.

**Definition 2.3** A function \(f : [k]^n \rightarrow [k]\) is monotone if for every \(x \in [k]^n\) and every \(i \in [n]\),

\[
f(x_1, \ldots, x_{i-1}, f(x), x_{i+1}, \ldots, x_n) = f(x).
\]
In other words, if voter $i$ changes their vote to the winning candidate, then that candidate will still win.

2.1 Weighted plurality functions

Plurality is the voting rule which simply chooses the most popular candidate. A weighted plurality rule is similar, except that the voters have weights and the popularity of a candidate is defined as the sum of the weights of their supporters. To be mathematically precise, we make the following definition:

**Definition 2.4** A function $f: [k]^n \to [k]$ is a *weighted plurality function* if there exist weights $w_1, \ldots, w_n \in \mathbb{R}_{\geq 0}$ such that $\sum_i w_i = 1$ and for all $a, b \in [k]$, $f(x) = a$ implies that

$$\sum_{i: x_i = a} w_i \geq \sum_{i: x_i = b} w_i.$$

Note that the above definition does not prescribe a particular behavior if a tie occurs between two alternatives. If the weights are chosen so that ties never occur, then the weighted plurality function is clearly neutral. Moreover, for any set of weights we can construct a neutral weighted plurality function with those weights by following the tie-breaking rule outlined in Example 2.2. Note also that every weighted plurality function is monotone, regardless of the tie-breaking rule.

2.2 The influence of a voter

The final notion that we need before stating our result is a way to quantify the power of a single voter. When $k = 2$ and the voting rule is a weighted majority, the notion of a voter’s power goes back to Penrose (1946) and Banzhaf (1964). The *Banzhaf power index* of voter $i$ – a well-established measure of voter power – is defined as the probability that voter $i$ casts the deciding vote in an election where all other voters vote uniformly at random.

Definitions of voter power are much less well-established when voters vote non-independently and non-uniformly. In the case $k = 2$, Häggström et al. defined the *effect* of voter $i$ by

$$\tilde{e}_i(f, P) = P(f(X) = 1 \mid X_i = 1) - P(f(X) = 1 \mid X_i = 0).$$

This definition is perhaps best understood with a Bayesian interpretation: if $\tilde{e}_i(f, P)$ is large, then a pollster can accurately predict the election result simply by observing voter $i$’s vote. Note that if the $X_i$ are independently and uniformly distributed under $P$, then $\tilde{e}_i$ coincides with the Banzhaf power index.

For general $k$, we propose the following extension of Häggström et al.’s definition:

**Definition 2.5** Let $f$ be a function $[k]^n \to [k]$ and fix a probability distribution $P$ on $[k]^n$. The *effect of voter $i$* is
where \( X \) is a random variable distributed according to \( P \).

Note that when \( k = 2 \), \( e_i(f, P) = 2\tilde{e}_i(f, P) \); hence \( e_i \) generalizes the definition from Häggström et al. (2006). Observe also that \( e_i(f, P) \) is closely related to the correlation between the vote of voter \( i \) and the outcome of the election:

\[
P(f(X) = a | X_i = a) - P(f(X) = a | X_i \neq a) = \frac{\text{Cov}(\mathbb{1}_{\{f = a\}}, \mathbb{1}_{\{X_i = a\}})}{P(X_i = a)P(X_i \neq a)} \geq 4\text{Cov}(\mathbb{1}_{\{f = a\}}, \mathbb{1}_{\{X_i = a\}}) \]

and so

\[
e_i(f, P) \geq 4 \sum_{a=1}^{k} \text{Cov}(\mathbb{1}_{\{f = a\}}, \mathbb{1}_{\{X_i = a\}}).
\]

In particular, this shows that \( e_i(f, P) \) has an interpretation in terms of election predictability: if \( e_i(f, P) \) is large then knowing voter \( i \)'s intentions lets us predict the winner.

**Example 2.6** The simplest example of \( e_i(f, P) \) is when \( P \) is a product measure (i.e. the \( X_i \) are independent) and the function \( f \) does not depend on its \( i \)th coordinate; in that case, \( P(f(X) = a | X_i = a) = P(f(X) = a | X_i \neq a) \) for all \( a \) and so \( e_i(f, P) = 0 \). On the other hand, if \( P \) is a distribution such that \( X_1 = X_2 = \cdots = X_n \) with probability 1, and if \( f \) is a plurality function, then \( P(f(X) = a | X_i = a) = 1 \) for all \( a \), while \( P(f(X) = a | X_i \neq a) = 0 \); hence, \( e_i(f, P) = 1 \) for all \( i \).

For a less trivial example, suppose that the \( X_i \) are independent and uniformly distributed on \([k]\). Let \( f \) be an unweighted plurality function. Then the Central Limit Theorem implies that \( e_i(f, P) = O(\frac{1}{\sqrt{n}}) \) as \( n \to \infty \).

On the other hand, suppose that \( f \) is still an unweighted plurality function and the \( X_i \) are independent, but now \( P(X_i = 1) > P(X_i = a) + \delta \) for some \( \delta > 0 \) and all \( j \neq 1 \). Then Hoeffding’s inequality implies that \( P(f(X) = 1 | X_i) \geq 1 - 2\exp(-\delta^2n/4) \) for sufficiently large \( n \), regardless of the value of \( X_i \). In particular, this implies that \( e_i(f, P) = O(\exp(-\delta^2n/4)) \). Compared to the case where the \( X_i \) are uniformly distributed, this demonstrates that \( e_i(f, P) \) can depend strongly on \( P \), even when \( P \) is restricted to being a product measure.

Let us close this section with a remark about computing and estimating \( e_i(f, P) \). First of all, computing \( e_i(f, P) \) for an arbitrary \( f \) and \( P \) would naïvely require enumerating all possible voting configurations; this is clearly computationally intractable. More realistically, \( e_i(f, P) \) can be accurately estimated by drawing samples from \( P \): by the Central Limit Theorem, \( m \) samples will determine \( e_i(f, P) \) to within an error of \( O(1/\sqrt{m}) \).
2.3 The main result

We present two main results, which are essentially converses of one another. First of all, every low-effect weighted plurality function aggregates information well:

**Theorem 2.7** For every \( \delta > 0 \) and \( \epsilon > 0 \), there is a \( \tau > 0 \) such that the following holds: let \( f \) be a weighted plurality function \( f \) with weights \( w_i \) and suppose that \( P \) is a probability distribution on \([k]^n\) such that \( e_i(f, P) \leq \tau \). If there is a set \( A \subset [k] \) of candidates such that

\[
\sum_{i=1}^{n} w_i P(X_i = a) \geq \sum_{i=1}^{n} w_i P(X_i = b) + \delta
\]

for all \( i \in [n] \), all \( a \in A \) and all \( b \notin A \), then \( P(f(X) \in A) \geq 1 - \epsilon \).

Theorem 2.7 admits an important special case: if there is a single candidate \( a \) such that for every voter \( i \) and every other candidate \( b \),

\[
P(X_i = a) \geq P(X_i = b) + \delta,
\]

then the Theorem holds with \( A = \{a\} \); in particular, \( P(f(X) \in A) \geq 1 - \epsilon \).

Our second main result is a converse to Theorem 2.7: if a function \( f \) aggregates information well for every distribution \( P \) under which \( e_i(f, P) \) is small, then it must be a weighted plurality function. In other words, Theorem 2.7 is false for every function that is not a weighted plurality.

**Theorem 2.8** If a neutral function \( f \) is not a weighted plurality function then there exists a probability distribution \( P \) on \([k]^n\) such that \( P(X_i = 2) > P(X_i = 1) \) for all \( i \in [n] \) but \( P(f(X) = 1) = 1 \) (and hence \( e_i(f, P) = 0 \) for all \( i \)).

We remark that Theorem 2.8 is constructive in the sense that we can give an algorithm (based on solving a linear program) which either constructs some weights \( w_i \) witnessing the fact that \( f \) is a weighted plurality, or constructs a probability distribution \( P \) which witnesses the Theorem’s claim.

3 Discussion

We must be careful not to overstate the social implications of Theorems 2.7 and 2.8. In particular, we do not wish to infer from Theorem 2.8 that weighted plurality functions are the only “reasonable” voting rules. Instead, we put our results forward as evidence for the futility of studying voting rules according to their worst-case behavior over arbitrary voting distributions. Indeed, the probability distributions produced by Theorem 2.8 can be extremely unrealistic, as we will show by example.

**Example 3.1** Consider a two-candidate electoral college system on three states, each of which has three voters. In this system, each state selects a candidate by a simple majority vote and the overall winner is the candidate who wins the most states. Now
consider all voting configurations with the property that candidate two wins one state by three votes to none, while losing the other two states by two votes to one; let \( P \) be the uniform probability distribution over all such voting configurations.

The \( P \) defined above satisfies Theorem 2.8, since candidate one always wins under \( P \) even though each individual voter chooses candidate two with probability \( \frac{5}{9} \). On the other hand, this probability distribution \( P \) is clearly an unrealistic voting model since the voters’ opinions depend on one another in a very odd way.

One possible way around this problem is to consider a more restricted question: is there a family of “reasonable” probability distributions on votes and a family of “reasonable” voting rules such that for every admissible distribution and every admissible voting rule, low influence implies information aggregation? Under this framework, Theorem 2.8 implies that if we consider all probability distributions then our class of admissible rules must only contain weighted majority functions. One expects that as the class of admissible distributions is tightened, the class of admissible voting rules can be expanded. The tradeoff between the size of the class of distributions and the size of the class of functions is crucial. In particular, one would like the class of admissible distributions to contain all reasonable voter dependencies while ensuring that the class of admissible voting rules contains all those that one might want to use.

Let us single out one concrete instance of the problem above. Since Theorem 2.8 closes the case in which all probability distributions are admissible, we will ask about the other extreme: is there a family of distributions under which every neutral, monotone function \( f \) with small voter effects aggregates information? To pose this question more precisely, let us say that a family of probability distributions \( \mathcal{P} \) aggregates information if for every \( \delta > 0 \) and \( \epsilon > 0 \), there is a \( \tau > 0 \) such that for every \( P \in \mathcal{P} \)

\[
P(X_i = 1) \geq P(X_i = b) + \delta \quad \text{for all } i \in [n] \text{ and all } b \in \{2, \ldots, k\}
\]

and for every neutral, monotone \( f : [k]^n \to [k] \) satisfying

\[
e_i(f, P) \leq \tau \quad \text{for all } i \in [n],
\]

it holds that

\[
P(f(X) = 1) \geq 1 - \epsilon.
\]

Our discussion in the introduction implies that if \( \mathcal{P} \) is the set of probability distributions over statistically independent voters, then \( \mathcal{P} \) aggregates information. But what about a less trivial family of distributions? The only result that we know of is negative: Häggström et al. (2006) showed by counterexample that the family of distributions satisfying the Fortuin-Kasteleyn-Ginibre (FKG) inequality (Fortuin et al. 1971) does not aggregate information. Distributions satisfying the FKG inequality have been well-studied in both statistical physics and economics (where it has been sometimes called the “affiliation inequality” Milgrom and Weber 1982). For example, although they do not guarantee aggregation of information as we defined it, FKG distributions do imply a sharp threshold property (Graham and Grimmett 2006), which is related.
We are led, therefore, to the following open question: is there a set \( \mathcal{P} \) of probability distributions which is sufficiently restricted so that it aggregates information, but sufficiently rich so that it includes reasonable models of voter dependence?

4 The proof

The proof of Theorem 2.7 follows the argument from Häggström et al. (2006) closely; nevertheless, we will show the complete proof in order to keep this presentation self-contained.

**Proof of Theorem 2.7** Suppose that \( f \) is a weighted plurality function with weights \( w_i \). The first step is to show that \( f \) is “correlated” in some sense with each voter: define \( p_{ia} = P(X_i = a) \) and let \( W_a \) be the (random) weight assigned to candidate \( a : W_a = \sum_{i: X_i = a} w_i. \) Then

\[
\mathbb{E} \sum_{i=1}^{n} w_i \sum_{a=1}^{k} 1_{\{f(X)=a\}} (1_{\{X_i=a\}} - p_{ia})
\]

\[
= \mathbb{E} \left( \sum_{i=1}^{n} \sum_{a=1}^{k} w_i 1_{\{f(X)=a\}} 1_{\{X_i=a\}} - \sum_{i=1}^{n} \sum_{a=1}^{k} 1_{\{f(X)=a\}} w_i p_{ia} \right)
\]

\[
= \mathbb{E} \sum_{i=1}^{n} \sum_{a=1}^{k} w_i 1_{\{f(X)=a\}} 1_{\{X_i=a\}} - \sum_{a=1}^{k} P(f = a) \mathbb{E} W_a. \quad (1)
\]

Now, let \( \alpha_a = P(f = a) \) and set \( \tilde{\alpha}_a = \alpha_a / (\sum_{a \in A} \alpha_a) \) for \( a \in A \) and \( \tilde{\alpha}_a = 0 \) otherwise. The first term of \( (1) \) is just

\[
\mathbb{E} \sum_{i=1}^{n} \sum_{b=1}^{k} w_i 1_{\{f(X)=b\}} 1_{\{X_i=b\}} = \mathbb{E} \sum_{b=1}^{k} 1_{\{f(X)=b\}} W_b
\]

\[
\geq \mathbb{E} \sum_{b=1}^{k} 1_{\{f(X)=b\}} \sum_{a \in A} \tilde{\alpha}_a W_a
\]

\[
= \sum_{a \in A} \tilde{\alpha}_a \mathbb{E} W_a, \quad (2)
\]

where the inequality follows because the winning alternative always has at least as much weight as any other alternative; i.e. \( 1_{\{f(X)=a\}} W_a \geq 1_{\{f(X)=a\}} W_b \) for any \( b \in \{k\} \). This is where we require \( f \) to be a weighted plurality function.

Recall the Theorem’s assumption that

\[
\sum_{i=1}^{n} w_i P(X_i = a) \geq \sum_{i=1}^{n} w_i P(X_i = b) + \delta
\]
for all \( a \in A \) and all \( b \not\in A \). With our definition of \( W_a \), this assumption is equivalent to

\[
\min_{a \in A} \mathbb{E}W_a \geq \max_{b \not\in A} \mathbb{E}W_b + \delta.
\]

If we set, therefore, \( W^* = \max_{b \not\in A} \mathbb{E}W_b \) then plugging (2) into (1) implies that

\[
(1) \geq \sum_{a \in A} \tilde{\alpha}_a \mathbb{E}W_a - \sum_{b=1}^k \alpha_b \mathbb{E}W_b
\]

\[
= \sum_{a \in A} (\tilde{\alpha}_a - \alpha_a) \mathbb{E}W_a - \sum_{b \not\in A} \alpha_b \mathbb{E}W_b
\]

\[
\geq (W^* + \delta) \sum_{a \in A} (\tilde{\alpha}_a - \alpha_a) - W^* \sum_{b \not\in A} \alpha_b.
\]

Noting that \( \sum_{a \in A} \tilde{\alpha}_a = \sum_{b=1}^k \alpha_b = 1 \), we obtain

\[
(1) \geq \delta \sum_{j \in A} (\tilde{\alpha}_j - \alpha_j)
\]

\[
= \delta P(f \not\in A).
\]

Recalling that \( e_i(f, P) \geq 4 \sum_a \text{Cov}(1 \{ f = a \}, 1 \{ X_i = a \}) \), we have

\[
\delta P(f \not\in A) \leq (1) = \mathbb{E} \sum_{i=1}^n w_i \sum_{a=1}^k 1 \{ f(X) = a \}(1 \{ X_i = a \} - p_{ia})
\]

\[
\leq \frac{1}{4} \sum_i w_i e_i(f, P)
\]

\[
\leq \frac{\tau}{4}
\]

and this establishes the theorem once we take \( \tau \) small enough that \( \epsilon \geq \tau / (4\delta) \).

The proof of Theorem 2.8 follows the idea of Häggström et al. (2006), in that we use linear programming duality to find a witness for \( f \) being a weighted plurality function. However, the details of the proof are quite different, since Häggström et al. (2006) uses a well-known linear program (the fractional vertex cover of a hypergraph) which does not extend beyond \( k = 2 \).

The proof idea is this: we will write down a linear program and its dual. If the primal program has a large enough value it will turn out that \( f \) is a weighted plurality function. Otherwise, the dual has a small value and the dual variables witness the claim of Theorem 2.8. In particular, note that this proof provides the algorithm that we mentioned after the statement of Theorem 2.8.
First we make a trivial observation that will simplify our linear program considerably: if a function is neutral, it is easier to check whether it is a weighted plurality because it is not necessary to try all possible combinations of \(a, b \in [k]\):  

**Proposition 4.1** Suppose \(f : [k]^n \to [k]\) is neutral. Then \(f\) is a weighted plurality if and only if there exist weights \(w_1, \ldots, w_n \in \mathbb{R}\) such that \(f(x) = 1\) implies that \[
\sum_{i : x_i = 1} w_i \geq \sum_{i : x_i = 2} w_i.
\]

Next, we will write down a linear program for checking whether a given neutral function \(f\) is a weighted plurality. The variables for this program are \(t_+\) and \(t_-; w_i\) for each \(i \in [n]\) and \(g_x\) for each \(x \in [k]^n\) for which \(f(x) = 1\). In standard form, the primal program is the following:

maximize \(t_+ - t_-\)  
subject to \(g_x \geq 0\) for all \(x \in [k]^n\) such that \(f(x) = 1\)  
\(w_i \geq 0\) for all \(i \in [n]\)  
\(t_+ \geq 0\) and \(t_- \geq 0\)  
\(\sum_{i=1}^{n} w_i = 1\)  
\(\sum_{i : x_i = 1} w_i - \sum_{i : x_i = 2} w_i - g_x - (t_+ - t_-) = 0\) for all \(x \in [k]^n\) with \(f(x) = 1\).

The linear program above may seem complicated at first glance, but all of the intuition behind it is contained in the following Proposition and its proof:

**Proposition 4.2** Let \(t^*\) be the value of the above linear program. If \(t^* \geq 0\) then \(f\) is a weighted plurality function.

**Proof** Let \(w_i, g_x, t_+\) and \(t_-\) be feasible points such that \(t_+ - t_- \geq 0\). Then, for all \(x\) with \(f(x) = 1\),

\[
\sum_{i : x_i = 1} w_i - \sum_{i : x_i = 2} w_i = g_x + (t_+ - t_-) \geq 0
\]

and so \(f\) satisfies the conditions of Proposition 4.1. \(\square\)

Now consider the dual program; since we have written the primal program in standard form, the dual program is easy to write down. Let the dual variables be \(a_+\) and \(a_-\), and \(q_x\) for all \(x\) such that \(f(x) = 1\). Then the dual program is:
minimize \( a_+ - a_- \)
subject to
\[
\sum_{x: f(x)=1} q_x \leq -1
\]
\[
\sum_{x: f(x)=1} (\mathbb{1}_{\{x_i=1\}} - \mathbb{1}_{\{x_i=2\}})q_x + (a_+ - a_-) \geq 0 \text{ for all } i \in [n]
\]
\[
q_x \leq 0 \text{ for all } x \text{ such that } f(x) = 1
\]
\[
a_+ \leq 0 \text{ and } a_- \leq 0.
\]

**Proposition 4.3** Let \( a^* \) be the value of the above dual program. If \( a^* < 0 \) then there exists a probability distribution on \([k]^n\) such that \( P(X_i = 2) > P(X_i = 1) \) for all \( i \) but \( f(X) = 1 \) almost surely.

**Proof** Choose a feasible point with \( a_+ - a_- < 0 \) and for every \( x \) with \( f(x) = 1 \), define \( p_x = q_x / (\sum_{x: f(x)=1} q_x) \). Then \( p_x \geq 0 \) and \( \sum_{x: f(x)=1} p_x = 1 \), so we can define a probability distribution by \( P(X = x) = p_x \) when \( f(x) = 1 \) and \( P(X = x) = 0 \) otherwise. Under this distribution, \( f(X) = 1 \) with probability 1. On the other hand, with \( a_+ - a_- < 0 \) the constraints of the dual program imply that
\[
\sum_{x: f(x)=1} \mathbb{1}_{\{x_i=1\}}q_x > \sum_{x: f(x)=1} \mathbb{1}_{\{x_i=2\}}q_x
\]
for all \( i \). Thus,
\[
P(X_i = 1) = \sum_{x: f(x)=1} \mathbb{1}_{\{x_i=1\}}p_x < \sum_{x: f(x)=1} \mathbb{1}_{\{x_i=2\}}p_x = P(X_i = 2)
\]
for all \( i \).

To conclude the proof of Theorem 2.8, note that both the primal and dual programs are feasible and bounded and so \( a^* = t^* \). Hence, either the hypothesis of Proposition 4.2 or the hypothesis of Proposition 4.3 must be satisfied. In the former case, \( f \) is a weighted plurality function; in the latter case, there is a probability distribution \( P \) witnessing the claim of Theorem 2.8.

**Acknowledgements** The author would like to thank Elchanan Mossel for suggesting this problem and providing fruitful discussions. He also thanks the Managing Editor and the two anonymous referees for helping him to substantially improve the presentation.

**References**

Arrow K (1950) A difficulty in the theory of social welfare. J Polit Econ 58:328–346
Banzhaf III JF (1964) Weighted voting doesn’t work: a mathematical analysis. Rutgers L Rev 19:317
Brams SJ, Fishburn PC (2002) Voting procedures. Handbook of social choice and welfare, vol 1. Elsevier, Amsterdam, pp 173–236
Fortuin CM, Kasteleyn PW, Ginibre J (1971) Correlation inequalities on some partially ordered sets. Commun Mathe Phys 22(2):89–103
Gibbard A (1973) Manipulation of voting schemes: a general result. Econometrica 41(4):587–601
A law of large numbers

Graham BT, Grimmett GR (2006) Influence and sharp-threshold theorems for monotonic measures. Ann Probab 34(5):1726–1745
Häggström O, Kalai G, Mossel E (2006) A law of large numbers for weighted majority. Adv Appl Math 37(1):112–123
Isaksson M, Kindler G, Mossel E (2010) The geometry of manipulation—a quantitative proof of the gilbard satterthwaite theorem. In: Foundations of computer science (FOCS), pp 319–328
Kalai G (2002) A Fourier-theoretic perspective on the Condorcet paradox and Arrow’s theorem. Adv Appl Math 29(3):412–426
Jean-Antoine-Nicolas Condorcet (1785) Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix. De l’Imprimerie Royale
Kalai G (2004) Social indeterminacy. Econometrica 72:1565–1581
Milgrom PR, Weber RJ (1982) A theory of auctions and competitive bidding. Econometrica 50:1089–1122
Penrose LS (1946) The elementary statistics of majority voting. J R Stat Soc 109(1):53–57
Russo L (1982) An approximate zero-one law. Prob Theory Relat Fields 61(1):129–139