The inverse Banzhaf problem

Noga Alon * Paul H. Edelman †

Abstract

Let $\mathcal{F}$ be a family of subsets of the ground set $[n] = \{1, 2, \ldots, n\}$. For each $i \in [n]$ we let $p(\mathcal{F}, i)$ be the number of pairs of subsets that differ in the element i and exactly one of them is in $\mathcal{F}$. We interpret $p(\mathcal{F}, i)$ as the influence of that element. The normalized Banzhaf vector of $\mathcal{F}$, denoted $B(\mathcal{F})$, is the vector $(B(\mathcal{F}, 1), \ldots, B(\mathcal{F}, n))$, where $B(\mathcal{F}, i) = \frac{p(\mathcal{F}, i)}{p(\mathcal{F})}$ and $p(\mathcal{F})$ is the sum of all $p(\mathcal{F}, i)$. The Banzhaf vector has been studied in the context of measuring voting power in voting games as well as in Boolean circuit theory. In this paper we investigate which non-negative vectors of sum 1 can be closely approximated by Banzhaf vectors of simple voting games. In particular, we show that if a vector has most of its weight concentrated in $k < n$ coordinates, then it must be essentially the Banzhaf vector of some simple voting game with $n - k$ dummy voters.

1 Introduction

A fundamental question when analyzing a voting method is what is the distribution of power among the voters. The most common measure of power, the Banzhaf index, quantifies the percentage of power of a voter by its ability to alter the outcome, i.e., the probability that if that voter were to change its vote, the outcome would change. There are numerous theorems that show how to compute the Banzhaf index in various circumstances. This paper will be about the inverse problem: if we fix a vector of prospective percentages of power, can we find a voting method which will give us a good approximation to those desired powers ?

*Tel Aviv University, Tel Aviv 69978, Israel and IAS, Princeton, NJ, 08540, USA. Research supported in part by an ERC Advanced grant, by a USA-Israel BSF grant, by NSF grant CCF 0832797 and by the Ambrose Monell Foundation. Email: nogaa@tau.ac.il

†Department of Mathematics and the Law School, Vanderbilt University, Nashville, TN 37203. Email: paul.edelman@vanderbilt.edu

The problem of designing a voting system to achieve a certain distribution of power is not a purely theoretical one. In New York State there are a number of counties whose governmental structure is produced in just this way. Each town in the county is assigned a representative, but that representative is assigned a weighted vote in such a way that the representative's Banzhaf index is close to the percentage of the population of the county living in the particular town. These computations are re-done every ten years after the official census of the county is taken.[1]

The work that has been done on the inverse Banzhaf problem has been computational in nature. Iterative algorithms have been designed that take as input a vector in the standard simplex and a threshold value, and output a weighted voting game whose Banzhaf vector is within the threshold of the given one. See Laruelle and Widgren [6], Sutter [7] and Aziz, Paterson and Leech [1] for examples of such iterative algorithms which seem to work well in practice. Nevertheless, there is a fundamental problem in analyzing any such algorithm fully. There are no theoretical results that give *a priori* estimates of what thresholds are achievable. Without some such bound, it is hard to know when an iterative algorithm has converged sufficiently.

When the number of voters is small, it is clear that one can not closely approximate every power vector. There are just not enough different voting games to get close to every vector. But, one might think that as the number of voters increases one can closely approximate every vector. In this paper we show this to be false by exhibiting vectors in the standard n-simplex for all n that can not be well approximated. Our focus will be on those vectors where the power is concentrated mainly on some strict subset of the voters, say k out of n. We will show that if the power is concentrated on only k of the n voters, then it can not be closely approximated unless it is essentially the power vector of a voting game with only k voters.

The Banzhaf index is usually defined in terms of a simple voting game. We will adopt a somewhat more general setting here. For positive integers $k < n$, let $[n] = \{1, 2, \ldots, n\}$, and $(k, n] = \{k+1, \ldots, n\}$. For a set X, let 2^X denote the family of all subsets of X. Let $\mathcal{F}$ be a family of subsets of $[n]$. For each i, $1 \leq i \leq n$, let

$$p(\mathcal{F}, i) = |\{A \subseteq [n] - \{i\} : \ |\{A, A \cup \{i\}\} \cap \mathcal{F}| = 1\}|.$$

be the number of pairs of subsets of $[n]$ that differ in one element, so that exactly one of

[1]While this form of representation was found to be unconstitutional for Nassau County, New York in *Jackson v. Nassau County Board of Supervisors*, 818 F.Supp. 509(1993), that holding was not binding outside of the Eastern District of New York. Weighted voting continues to be used elsewhere in the state, e.g., Essex County (see www.co.essex.ny.us/bos.asp) and Washington County (www.co.washington.ny.us/Departments/bos/bos_wgt.htm).

them lies in $\mathcal{F}$. Let $p(\mathcal{F}) = \sum_{i=1}^{n} p(\mathcal{F}, i)$. If $p(\mathcal{F}) > 0$, the (normalized) *Banzhaf vector* of $\mathcal{F}$, denoted $B(\mathcal{F})$, is the vector $(B(\mathcal{F}, 1), B(\mathcal{F}, 2), \ldots, B(\mathcal{F}, n))$, where $B(\mathcal{F}, i) = \frac{p(\mathcal{F}, i)}{p(\mathcal{F})}$.

A collection $\mathcal{W} \subset 2^X$ is called a *simple voting game* if it satisfies the following three conditions

1. $X \in \mathcal{W}$;
2. $\emptyset \notin \mathcal{W}$;
3. Whenever $A \subseteq B \subseteq X$ and $A \in \mathcal{W}$, then $B \in \mathcal{W}$.

The elements of X are called the *voters* and the sets in $\mathcal{W}$ are referred to as *winning coalitions*. If $|X| = n$ then we will refer to $\mathcal{W}$ as an n-game. A voter x is a *dummy* if there are no winning coalitions $A \in \mathcal{W}$ such that $x \in A$ and $A - \{x\} \notin \mathcal{W}$. That is to say, x is a dummy if $p(\mathcal{W}, x) = 0$. Note also that for any n-game $\mathcal{W}$ we have that $p(\mathcal{W}) > 0$.

For an n-game $\mathcal{W} \subseteq 2^{[n]}$ and for $A \subseteq [k]$, define $\mathcal{W}_A = \{B \subseteq (k, n] : \ A \cup B \in \mathcal{W}\}$. It is easy to check that $\mathcal{W}_A$ is one of three types; $\mathcal{W}_A = \emptyset$, $\mathcal{W}_A = 2^{(k,n]}$ or $\mathcal{W}_A$ is an $n-k$-game. Call an n-game $\mathcal{W}$ *k-pure* if for every $A \subseteq [k]$, $\mathcal{W}_A$ is either empty or equal to $2^{(k,n]}$. Equivalently, $\mathcal{W}$ is k-pure if all of the voters in $(k, n]$ are dummies. Clearly, if $\mathcal{W}$ is k-pure, then the family $\mathcal{V} \subseteq 2^{[k]}$ of all sets $A \subseteq [k]$ for which $\mathcal{W}_A = 2^{(k,n]}$ is itself a k-game. Moreover, in this case the vector consisting of the first k coordinates of $B(\mathcal{W})$ is equal to $B(\mathcal{V})$, and the last $n - k$ coordinates of $B(\mathcal{W})$ vanish. Our main result is the following theorem, that shows that if almost all the weight of a Banzhaf vector of an n-game is concentrated in k coordinates, then its Banzhaf vector is close to a Banzhaf vector of a k-pure game, i.e., a game with $n - k$ dummy voters.

Theorem 1.1 *Let $n > k$ be positive integers, let $\epsilon < \frac{1}{k+1}$ be a positive real, and let $\mathcal{W} \subseteq 2^{[n]}$ be an n-game. If $\sum_{i=k+1}^{n} B(\mathcal{W}, i) \leq \epsilon$, then there exists a k-pure n-game $\mathcal{W}'$ so that*

$$||B(\mathcal{W}') - B(\mathcal{W})||_1 = \sum_{i=1}^{n} |B(\mathcal{W}', i) - B(\mathcal{W}, i)| \leq \frac{(2k+1)\epsilon}{1-(k+1)\epsilon} + \epsilon.$$

This shows that the significant part of the Banzhaf vector of any n-game in which most of the weight is concentrated in the first k coordinates is essentially equal to the Banzhaf vector of a k-game. Thus, for example, for $k = 2$ and $\epsilon = 0.01$, the theorem implies that if for an n-game $\mathcal{W}$, $\sum_{i=3}^{n} B(\mathcal{W}, i) \leq 0.01$, then the two dimensional vector $(B(\mathcal{W}, 1), B(\mathcal{W}, 2))$ lies within ℓ_1-distance smaller than $1/16$ of one of the vectors $(1, 0)$, $(0, 1)$ or $(0.5, 0.5)$, since these are the only vectors in the plane that are realizable as Banzhaf vectors. In other words, if two voters together share almost all the power in a voting scheme, and none of them is a near dictator, then they must have almost the same power.

2 The proof

In this section we present the proof of the main result. The basic idea is to show that if almost all weight of the Banzhaf vector is concentrated in the first k coordinates, then it is possible to add or delete a relatively small number of coalitions to the game to get a k-pure game. The result follows by showing that such a small number of modifications cannot change the Banzhaf vector significantly. We proceed with the details.

Proof of Theorem 1.1: Let $\mathcal{W} \subseteq 2^{[n]}$ be an n-game, and assume that $\sum_{i=k+1}^{n} B(\mathcal{W}, i) \leq \epsilon$, where $0 < \epsilon < \frac{1}{k+1}$. Therefore

$$\sum_{i=k+1}^{n} p(\mathcal{W}, i) \leq \epsilon p(\mathcal{W}). \tag{1}$$

Suppose that for $i \in (k, n]$ and $B \subset [n]$ we have that $i \notin B,\ B \notin \mathcal{W}$, but $B \cup i \in \mathcal{W}$. Let $A = B \cap [1, k]$ and $B' = B - A$. Then $i \notin B',\ B' \notin \mathcal{W}_A$, but $B' \cup i \in \mathcal{W}_A$. This correspondence implies that for every i, $k+1 \leq i \leq n$,

$$p(\mathcal{W}, i) = \sum_{A \subseteq [k]} p(\mathcal{W}_A, i). \tag{2}$$

By the well-known edge-isoperimetric inequality for the cube (see [4], [2], [5]), for every family $\mathcal{F} \subseteq 2^{[n]}$,

$$p(\mathcal{F}) \geq |\mathcal{F}|(n - \log_2 |\mathcal{F}|).$$

In particular, if $|\mathcal{F}| \leq 2^{n-1}$, $p(\mathcal{F}) \geq |\mathcal{F}|$. Since one of $\mathcal{F}$ and $\overline{\mathcal{F}}$, the complement of $\mathcal{F}$, has less than half of all possible subsets and $p(\mathcal{F}) = p(\overline{\mathcal{F}})$, it follows that

$$p(\mathcal{F}) \geq \min\{|\mathcal{F}|, |\overline{\mathcal{F}}|\}. \tag{3}$$

Combining (1) and (2), and applying (3) to each of the collections $\mathcal{W}_A$, we conclude that

$$\epsilon p(\mathcal{W}) \geq \sum_{i=k+1}^{n} p(\mathcal{W}, i) = \sum_{i=k+1}^{n} \sum_{A \subseteq [k]} p(\mathcal{W}_A, i) = \sum_{A \subseteq [k]} p(\mathcal{W}_A) \geq \sum_{A \subseteq [k]} \min\{|\mathcal{W}_A|, |\overline{\mathcal{W}_A}|\}. \tag{4}$$

Let $\mathcal{W}' \subseteq 2^{[n]}$ be the family obtained from $\mathcal{W}$ by defining, for every $A \subseteq [k]$, $\mathcal{W}'_A = \emptyset$ if $|\mathcal{W}_A| \leq |\overline{\mathcal{W}_A}|$, and $\mathcal{W}'_A = 2^{(k,n]}$ if $|\mathcal{W}_A| > |\overline{\mathcal{W}_A}|$. It is not difficult to check that $\mathcal{W}'$ is a collection satisfying axiom 3 of a simple voting game, and it is obviously k-pure. We will now establish some further inequalities to show that $\mathcal{W}'$ is neither empty nor all of $2^{[n]}$, and hence is an n-game.

By (4), the complex $\mathcal{W}'$ is obtained from $\mathcal{W}$ by removing or adding at most $\epsilon p(\mathcal{W})$ sets. The crucial observation is that for every fixed i, the quantity $p(\mathcal{W}, i)$ can change by at most 1 with the addition or deletion of a single set to $\mathcal{W}$. It thus follows that for each i, $1 \leq i \leq k$,

$$p(\mathcal{W}, i) - \epsilon p(\mathcal{W}) \leq p(\mathcal{W}', i) \leq p(\mathcal{W}, i) + \epsilon p(\mathcal{W}).$$

By summing over all i, $1 \leq i \leq k$, and by noting that for each $i \geq k+1$, $p(\mathcal{W}', i) = 0$ and that $\sum_{i=k+1}^{n} p(\mathcal{W}, i) \leq \epsilon p(\mathcal{W})$ this implies that

$$[1 - (k+1)\epsilon]p(\mathcal{W}) \leq p(\mathcal{W}') \leq p(\mathcal{W}) + k\epsilon p(\mathcal{W}) < [1 + (k+1)\epsilon]p(\mathcal{W})$$

Note, in particular, that since $\epsilon < \frac{1}{k+1}$ this implies that $p(\mathcal{W}') > 0$, showing that $\mathcal{W}'$ cannot be the collection of all sets or empty. Thus $\mathcal{W}'$ is an n-game.

It remains to show that the last two inequalities imply that the two vectors $B(\mathcal{W}')$ and $B(\mathcal{W})$ are close to each other in the ℓ_1-norm. Indeed, by the above inequalities, for every i, $1 \leq i \leq k$

$$\frac{p(\mathcal{W}, i) - \epsilon p(\mathcal{W})}{[1 + (k+1)\epsilon]p(\mathcal{W})} \leq \frac{p(\mathcal{W}', i)}{p(\mathcal{W}')} \leq \frac{p(\mathcal{W}, i) + \epsilon p(\mathcal{W})}{[1 - (k+1)\epsilon]p(\mathcal{W})},$$

that is

$$\frac{B(\mathcal{W}, i)}{1 + (k+1)\epsilon} - \frac{\epsilon}{1 + (k+1)\epsilon} \leq B(\mathcal{W}', i) \leq \frac{B(\mathcal{W}, i)}{1 - (k+1)\epsilon} + \frac{\epsilon}{1 - (k+1)\epsilon}.$$

Therefore, for each $1 \leq i \leq k$,

$$-B(\mathcal{W}, i)\frac{(k+1)\epsilon}{1 + (k+1)\epsilon} - \frac{\epsilon}{1 + (k+1)\epsilon} \leq B(\mathcal{W}', i) - B(\mathcal{W}, i)$$

$$\leq B(\mathcal{W}, i)\frac{(k+1)\epsilon}{1 - (k+1)\epsilon} + \frac{\epsilon}{1 - (k+1)\epsilon},$$

and thus

$$|B(\mathcal{W}', i) - B(\mathcal{W}, i)| \leq \frac{(k+1)\epsilon}{1 - (k+1)\epsilon} B(\mathcal{W}, i) + \frac{\epsilon}{1 - (k+1)\epsilon}. \tag{5}$$

Summing (5) over all $1 \leq i \leq k$, and using the fact that $\sum_{i=1}^{k} B(\mathcal{W}, i) \leq 1$ we conclude that

$$\sum_{i=1}^{k} |B(\mathcal{W}', i) - B(\mathcal{W}, i)| \leq \frac{(2k+1)\epsilon}{1 - (k+1)\epsilon}.$$

Since $\sum_{i=k+1}^{n} |B(\mathcal{W}', i) - B(\mathcal{W}, i)| \leq \epsilon$ it follows that $||B(\mathcal{W}') - B(\mathcal{W})||_1 \leq \frac{(2k+1)\epsilon}{1-(k+1)\epsilon} + \epsilon$, completing the proof. ■

3 Concluding remarks

We have seen that some vectors of distribution of power cannot be approximated well by Banzhaf vectors of simple voting games. In some voting schemes, the voters are distributed into regions, and we may be interested in the distribution of power among the regions, rather than among the individual voters. In the next simple proposition we observe that if every region is large, we can always approximate well any distribution of power among the regions.

Proposition 3.1 *Let r be an integer, let $\epsilon > 0$ be a real, and let $(b_1, b_2, \ldots, b_r)$ be a vector of probabilities, that is, $b_i \geq 0$ for all i and $\sum_{i=1}^{r} b_i = 1$. Let $[n] = N_1 \cup N_2 \cup \ldots \cup N_r$ be a partition of $[n] = \{1, 2, \ldots, n\}$ into r pairwise disjoint sets N_i so that $|N_i| > \frac{r}{\epsilon}$ for all i. Then there is an n-game $\mathcal{W}$ so that for $B_i = \frac{\sum_{j \in N_i} p(\mathcal{W},j)}{p(\mathcal{W})}$, $\sum_{i=1}^{r} |B_i - b_i| \leq \epsilon$.*

Proof: Define $t = \lceil \frac{r}{\epsilon} \rceil$. For each i, $1 \leq i \leq r$, let t_i be an integer satisfying $\lfloor b_i t \rfloor \leq t_i \leq \lceil b_i t \rceil$, so that $\sum_{i=1}^{r} t_i = t$. Let R_i be an arbitrary subset of cardinality t_i of N_i, which will be called the set of representatives of N_i. The set $R = \cup_{i=1}^{r} R_i$ is the set of all representatives. The n-game $\mathcal{W}$ consists of all subsets of $[n]$ that contain at least $t/2$ elements of R. (In fact, any symmetric condition on the set of representatives, for example, containing at least one representative, will do). By symmetry, each element $j \in R$ has the same Banzhaf index $p(\mathcal{W}, j)$, while clearly for each $j \in [n] - R$, $p(\mathcal{W}, j) = 0$. It follows that for each i, $B_i = \frac{t_i}{t}$, implying that $|B_i - b_i| \leq \frac{1}{t} < \frac{\epsilon}{r}$, and completing the proof. ■

Note that by our main result, in order to achieve a good approximation for any desired distribution of powers, we sometimes need large sets of representatives for each region; if each region is only allowed to have a single representative and all the power is distributed among them, then some vectors will not admit good approximations, even if the number of regions is large.

This paper begins the study of what vectors in the standard simplex can be closely approximated by Banzhaf vectors of simple voting games. We have shown that vectors in which most weight is concentrated in a small number of coordinates can only be closely approximated if they are essentially the Banzhaf vectors of simple voting games on a smaller ground set. However, we still know little of how the Banzhaf vectors are distributed throughout the simplex. The technique used in this paper looks to be unable to provide a complete solution of that more general question.

References

[1] H. Aziz, M. Paterson and D. Leech, Efficient algorithm for designing weighted voting games, Multitopic Conference, 2007, INMIC 2007, IEEE International (2007), 1–6.

[2] A.J. Bernstein, Maximally connected arrays on the n-cube, SIAM J. Appl. Math. 15 (1967) 1485-1489.

[3] D. S. Felsenthal and M. Machover, The Measurement of Voting Power: Theory and Practice, Problems and Paradoxes, Elgar(1998).

[4] L. H. Harper, Optimal assignments of numbers to vertices, J. Soc. Indust. Appl. Math. 12 (1964), 131–135.

[5] S. Hart, A note on the edges of the n-cube, Discrete Math. 14 (1976), 157–163.

[6] A. Laruelle and M. Widgren, Is the allocation of voting power among EU states fair?, Public Choice 94 (1998), 317-339.

[7] M. Sutter, Fair allocation and re-weighting of votes and voting power in the EU before and after the next enlargement, J. Theor. Politics 12 (2000), 433–449.