Algebraic Voting Theory

Michael Orrison

Harvey Mudd College

Collaborators and Sounding Boards

- Don Saari (UC Irvine)
- Anna Bargagliotti (University of Memphis)
- Steven Brams (NYU)
- Brian Lawson (Santa Monica College)
- Zajj Daugherty '05
- Alex Eustis '06
- Mike Hansen '07
- Marie Jameson '07
- Gregory Minton '08
- Stephen Lee '10
- Jen Townsend '10 (Scripps)
- Aaron Meyers '10 (Bucknell)
- Sarah Wolff '10 (Colorado College)
- Angela Wu '10 (Swarthmore)

Voting Paradoxes

Voting–Preferences

Example

Eleven voters have the following preferences:

2 ABC 3 ACB 4 BCA 2 CBA.

We will call this voting data the profile.

Change of Perspective

Focus on the procedure, not the preferences, because "...rather than reflecting the views of the voters, it is entirely possible for an election outcome to more accurately reflect the choice of an election procedure." (Donald Saari, *Chaotic Elections!*)

Let's Vote!

Preferences

2 ABC	3 ACB	4 BCA	2 CBA

Plurality: Vote for Favorite

A: 5 points B: 4 points C: 2 points $A > B > C$

Anti-Plurality: Vote for Top Two Favorites

A: 5 points B: 8 points C: 9 points $C > B > A$

Borda Count: 1 Point for First, $\frac{1}{2}$ Point for Second

A: 5 points B: 6 points C: $5\frac{1}{2}$ points $B > C > A$

Algebraic Perspective

Positional Voting with Three Candidates

Weighting Vector: $w = [1, s, 0]^t \in \mathbb{R}^3$

- 1st: 1 point
- 2nd: s points, $0 \leq s \leq 1$
- 3rd: 0 points

Tally Matrix: $T_{\mathbf{w}} : \mathbb{R}^{3!} \to \mathbb{R}^3$

$$T_{\mathbf{w}}(\mathbf{p}) = \begin{bmatrix} 1 & 1 & s & 0 & s & 0 \\ s & 0 & 1 & 1 & 0 & s \\ 0 & s & 0 & s & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 0 \\ 4 \\ 0 \\ 2 \end{bmatrix} \begin{matrix} \text{ABC} \\ \text{ACB} \\ \text{BAC} \\ \text{BCA} \\ \text{CAB} \\ \text{CBA} \end{matrix} = \begin{bmatrix} 5 \\ 4+4s \\ 2+7s \end{bmatrix} \begin{matrix} \text{A} \\ \text{B} \\ \text{C} \end{matrix} = \mathbf{r}$$

Linear Algebra

Tally Matrices

In general, we have a weighting vector $\mathbf{w} = [w_1, \ldots, w_n]^t \in \mathbb{R}^n$ and

$$T_{\mathbf{w}} : \mathbb{R}^{n!} \to \mathbb{R}^n.$$

Profile Space Decomposition

The effective space of $T_{\mathbf{w}}$ is $E(\mathbf{w}) = (\ker(T_{\mathbf{w}}))^{\perp}$. Note that

$$\mathbb{R}^{n!} = E(\mathbf{w}) \oplus \ker(T_{\mathbf{w}}).$$

Questions

What is the dimension of $E(\mathbf{w})$? Given $\mathbf{w}$ and $\mathbf{x}$, what is $E(\mathbf{w}) \cap E(\mathbf{x})$?

Change of Perspective

Profiles

We can think of our profile

$$\mathbf{p} = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 4 \\ 0 \\ 2 \end{bmatrix} \begin{matrix} \text{ABC} \\ \text{ACB} \\ \text{BAC} \\ \text{BCA} \\ \text{CAB} \\ \text{CBA} \end{matrix}$$

as an element of the group ring $\mathbb{R}S_3$:

$$\mathbf{p} = 2e + 3(23) + 0(12) + 4(123) + 0(132) + 2(13).$$

Change of Perspective

Tally Matrices

We can think of our tally $T_{\mathbf{w}}(\mathbf{p})$ as the result of **p** acting on **w**:

$$T_{\mathbf{w}}(\mathbf{p}) = \begin{bmatrix} 1 & 1 & s & 0 & s & 0 \\ s & 0 & 1 & 1 & 0 & s \\ 0 & s & 0 & s & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 0 \\ 4 \\ 0 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ s \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \\ s \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 1 \\ s \end{bmatrix} + 2 \begin{bmatrix} 0 \\ s \\ 1 \end{bmatrix}$$

$$= (2e + 3(23) + 4(123) + 2(13)) \cdot \begin{bmatrix} 1 \\ s \\ 0 \end{bmatrix} = \mathbf{p} \cdot \mathbf{w}.$$

Representation Theory

We have elements of $\mathbb{R}S_n$ (i.e., profiles) acting as linear transformations on the vector space $\mathbb{R}^n$:

$$\rho : \mathbb{R}S_n \to \mathrm{End}(\mathbb{R}^n) \cong \mathbb{R}^{n \times n}.$$

This opens the door to using tools and insights from the **representation theory** of the symmetric group.

Theorems

Equivalent Weighting Vectors

Defintion

Two nonzero weighting vectors $\mathbf{w}, \mathbf{x} \in \mathbb{R}^n$ are equivalent ($\mathbf{w} \sim \mathbf{x}$) if and only if there exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha > 0$ and $\mathbf{x} = \alpha\mathbf{w} + \beta\mathbf{1}$.

Example

$$[3, 2, 1]^t \sim [2, 1, 0]^t \sim [1, 1/2, 0]^t \sim [1, 0, -1]^t$$

Sum-zero Weighting Vectors

For convenience, we will usually assume that the entries of our weighting vectors sum to zero, i.e., our weighting vectors are sum-zero vectors.

Key Insight

If $\mathbf{w} \neq \mathbf{0}$ is sum-zero, then $E(\mathbf{w})$ is an irreducible $\mathbb{R}S_n$-module. In fact, $E(\mathbf{w}) \cong S^{(n-1,1)}$.

Results

Theorem (Saari)

Let $n \geq 2$, and let $\mathbf{w}$ and $\mathbf{x}$ be nonzero weighting vectors in $\mathbb{R}^n$. The ordinal rankings of $T_{\mathbf{w}}(\mathbf{p})$ and $T_{\mathbf{x}}(\mathbf{p})$ will be the same for all $\mathbf{p} \in \mathbb{R}^{n!}$ if and only if $\mathbf{w} \sim \mathbf{x}$.

Theorem

If $\mathbf{w}$ and $\mathbf{x}$ are nonzero sum-zero weighting vectors in $\mathbb{R}^n$, then $E(\mathbf{w}) = E(\mathbf{x})$ if and only if $\mathbf{w} \sim \mathbf{x}$. Moreover, if $E(\mathbf{w}) \neq E(\mathbf{x})$, then $E(\mathbf{w}) \cap E(\mathbf{x}) = \{\mathbf{0}\}$.

Theorem

If $\mathbf{w}$ and $\mathbf{x}$ are nonzero sum-zero weighting vectors in $\mathbb{R}^n$, then $\mathbf{w} \perp \mathbf{x}$ if and only if $E(\mathbf{w}) \perp E(\mathbf{x})$.

Results

Theorem

Let $n \geq 2$, and suppose $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \subset \mathbb{R}^n$ is a linearly independent set of sum-zero weighting vectors. If $\mathbf{r}_1, \ldots, \mathbf{r}_k$ are any k sum-zero results vectors in $\mathbb{R}^n$, then there exist infinitely many profiles $\mathbf{p} \in \mathbb{R}^{n!}$ such that $T_{\mathbf{w}_i}(\mathbf{p}) = \mathbf{r}_i$ for all $1 \leq i \leq k$.

In other words...

For a fixed profile $\mathbf{p}$, as long as our weighting vectors are different enough, there need not be any relationship whatsoever among the results of each election.

Key to the Proof

A theorem by Burnside says that every linear transformation from an irreducible module to itself can be realized as the action of some element (i.e., a profile) in $\mathbb{R}S_n$.

Why the Borda Count is Special

Pairwise Voting

Ordered Pairs

Assign points to each ordered pair of candidates, then use this information to determine a winner.

Example of the Pairs Matrix

$$P_2(\mathbf{p}) = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 0 \\ 4 \\ 0 \\ 2 \end{bmatrix} \begin{matrix} \text{ABC} \\ \text{ACB} \\ \text{BAC} \\ \text{BCA} \\ \text{CAB} \\ \text{CBA} \end{matrix} = \begin{bmatrix} 5 \\ 6 \\ 5 \\ 6 \\ 6 \\ 5 \end{bmatrix} \begin{matrix} \text{AB} \\ \text{BA} \\ \text{AC} \\ \text{CA} \\ \text{BC} \\ \text{CB} \end{matrix}$$

Voting Connection

Some voting procedures (e.g., Copeland) depend only on $P_2(\mathbf{p})$.

Pairwise and Positional Voting

Question

How are pairwise and positional voting methods related?

Definition

Let T and T' be linear transformations defined on the same vector space V. We say that T is recoverable from T' if there exists a linear transformation R such that $T = R \circ T'$.

Theorem (Saari)

A tally map $T_{\mathbf{w}} : \mathbb{R}^{n!} \to \mathbb{R}^{n}$ *is recoverable from the pairs map* $P_2 : \mathbb{R}^{n!} \to \mathbb{R}^{n(n-1)}$ *if and only if* $\mathbf{w}$ *is equivalent to the Borda count* $[n-1, n-2, \ldots, 1, 0]$.

Key to Our Proof

$E(T_w) \cong S^{(n-1,1)}$ and $E(P_2) \cong S^{(n)} \oplus S^{(n-1,1)} \oplus S^{(n-2,1,1)}$.

Counting Questions

To find the number of times each candidate is ranked above a $(k-1)$-element subset of other candidates, use the weighting vector

$$\mathbf{b}_k = \left[\binom{n-1}{k-1}, \binom{n-2}{k-1}, \ldots, \binom{1}{k-1}, \binom{0}{k-1} \right].$$

This is a generalization of the Borda count (which is $\mathbf{b}_2$).

Example

If $n = 4$, then $\mathbf{b}_1 = [1, 1, 1, 1]$, $\mathbf{b}_2 = [3, 2, 1, 0]$, $\mathbf{b}_3 = [3, 1, 0, 0]$, and $\mathbf{b}_4 = [1, 0, 0, 0]$.

Generalized Specialness

k-wise Maps

Generalize the pairwise map P_2 to create the *k-wise* map $P_k : \mathbb{R}^{n!} \to \mathbb{R}^{(n)_k}$ where P_k counts the number of times each ordered k-tuple of candidates is actually ranked in that order by a voter.

Theorem

Let $n \geq 2$ and let $\mathbf{w} \in \mathbb{R}^n$ be a weighting vector. The map $T_{\mathbf{w}}$ is recoverable from the k-wise map P_k if and only if $\mathbf{w}$ is a linear combination of $\mathbf{b}_1, \ldots, \mathbf{b}_k$.

Definition

We say that a weighting vector is k-Borda if it is a linear combination of $\mathbf{b}_1, \ldots, \mathbf{b}_k$.

Orthogonal Bases

Applying Gram-Schmidt to the $\mathbf{b}_i$ for small values of n yields:

$n = 2$: $\mathbf{c}_1 = [1, 1]$, $\mathbf{c}_2 = [1, -1]$

$n = 3$: $\mathbf{c}_1 = [1, 1, 1]$, $\mathbf{c}_2 = [2, 0, -2]$, and $\mathbf{c}_3 = [1, -2, 1]$.

$n = 4$: $\mathbf{c}_1 = [1, 1, 1, 1]$, $\mathbf{c}_2 = [3, 1, -1, -3]$, $\mathbf{c}_3 = [3, -3, -3, 3]$, and $\mathbf{c}_4 = [1, -3, 3, -1]$.

Theorem

A weighting vector for n candidates is $(n-1)$-Borda if and only if it is orthogonal to the nth row of Pascal's triangle with alternating signs.

Proof.

Focus on the inverses of so-called Pascal matrices. □

Pascal Matrices

If $n = 5$, then we are interested in the following Pascal matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 1 & 3 & 3 & 1 & 0 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}.$$

Its inverse looks just like itself but with alternating signs:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}.$$

Tests of Uniformity

Profiles

Ask m people to fully rank n alternatives from most preferred to least preferred, and encode the resulting data as a profile $\mathbf{p} \in \mathbb{R}^{n!}$.

Example

If $n = 3$, and the rankings of the alternatives A, B, C are ordered lexicographically, then the profile

$$\mathbf{p} = [10, 15, 2, 7, 9, 21]^t \in \mathbb{R}^6$$

encodes the situation where 10 judges chose the ranking ABC, 15 chose ACB, 2 chose BAC, and so on.

Data from a Distribution

We imagine that the data is being generated using a probability distribution P defined on the permutations of the alternatives.

We want to test the null hypothesis H_0 that P is the uniform distribution. A natural starting point is the estimated *probabilities vector*

$$\widehat{P} = (1/m)\mathbf{p}.$$

If $\widehat{P}$ is far from the vector $(1/n!)[1, \ldots, 1]^t$, then we would reject H_0.

In general, given a subspace S that is orthogonal to $[1, \ldots, 1]^t$, we'll compute the projection of $\widehat{P}$ onto S, and we'll use the value

$$mn! \|\widehat{P}^S\|^2$$

as a test statistic.

Linear Summary Statistics

The *marginals* summary statistic computes, for each alternative, the proportion of times an alternative is ranked first, second, third, and so on.

The *means* summary statistic computes the average rank of obtained by each alternative.

The *pairs* summary statistic computes for each ordered pair (A_i, A_j) of alternatives, the proportion of voters who ranked A_i above A_j.

Key Insight

The linear maps associated with the means, marginals, and pairs summary statistics described above are module homomorphisms. Futhermore, we can use their effective spaces (which are submodules of the data space $\mathbb{R}^{n!}$) to create our subspace S.

Matrices

Linear summary statistics may easily be realized by multiplying $\widehat{P}$ by a suitable matrix. For example, when $m = 3$, let

$$M_{\text{mns}} = \begin{bmatrix} 1 & 1 & 2 & 3 & 2 & 3 \\ 2 & 3 & 1 & 1 & 3 & 2 \\ 3 & 2 & 3 & 2 & 1 & 1 \end{bmatrix}.$$

Then $M_{\text{mns}}\widehat{P}$ encodes the average rank of each alternative.

Key Insight

The highly structured row spaces of these matrices form the effective spaces of the associated linear maps.

Decomposition

If $n \geq 3$, then the effective spaces of the means, marginals, and pairs maps are related by an orthogonal decomposition

$$\mathbb{R}^{n!} = W_1 \oplus W_2 \oplus W_3 \oplus W_4 \oplus W_5$$

into $\mathbb{R}S_n$-submodules such that

1. W_1 is the space spanned by the all-ones vector,
2. $W_1 \oplus W_2$ is the effective space for the means,
3. $W_1 \oplus W_2 \oplus W_3$ is the effective space for the marginals, and
4. $W_1 \oplus W_2 \oplus W_4$ is the effective space for the pairs.

Key Insight

The effective spaces for the means, marginals, and pairs summary statistics have some of the W_i in common. Thus the results of one test could have implications for the other tests.

Examples of Disagreement

Let $m = 3$, let $\alpha = .05$, and consider the data vector

$$\mathbf{d} = \begin{bmatrix} 6 \\ 10 \\ 6 \\ 10 \\ 14 \\ 14 \end{bmatrix} \begin{matrix} ABC \\ ACB \\ BAC \\ BCA \\ CAB \\ CBA \end{matrix}$$

for the three alternatives A, B, and C. When using the means test, the p-value is 0.0408, thus we reject the null hypothesis.

On the other hand, the p-values for the marginals test and pairs test are 0.1712 and 0.0937, respectively, thus we fail to reject the null hypothesis when using the marginals and pairs tests.

Finding Examples is Now Easy

The results above become less surprising once we see that $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2$, where $\mathbf{d}_i \in W_i$, and $\mathbf{d}_1 = [10, 10, 10, 10, 10, 10]^t$ and $\mathbf{d}_2 = [-4, 0, -4, 0, 4, 4]^t$. Thus, the data vector $\mathbf{d}$ is composed of vectors in just W_1 and W_2, which together form the effective space of the means summary statistic.

The spaces W_3 and W_4 are not needed to construct $\mathbf{d}$. Because they are necessary to form the effective spaces of the marginals and pairs summary statistics, however, this explains the larger p-values for the associated tests.

Other Examples

Marginals

The data vector $\mathbf{d} = [8, 16, 6, 18, 10, 8]^t$ rejects the null hypothesis for the marginals test, but not for the means or pairs tests. The *p*-values for the means, marginals, and pairs tests that are 0.8338, 0.0375, and 0.8232, respectively.

Pairs

The data vector $\mathbf{d} = [15, 8, 7, 16, 17, 9]^t$ rejects the null hypothesis for the pairs test, but not for the means or marginals tests. The *p*-values for the means, marginals, and pairs test are 0.8465, 0.9876, and 0.0396, respectively.

Connections and New Directions

Connections

Approval Voting

These ideas are applicable to approval voting where there are several weighting vectors being used at once: $[1, 0, 0, 0, \ldots, 0]^t, [1, 1, 0, 0, \ldots, 0]^t, [1, 1, 1, 0, \ldots, 0]^t, \ldots.$

Partial Rankings

These ideas may be extended to partially ranked data, in which case we have nontrivial analogues of the Borda count.

Extending Condorcet's Criterion

We can focus k candidates at a time and get different "k-winners" for different values of k.

New Directions

Dropping Candidates

How can we use this algebraic framework to help us better understand what happens when candidates drop out of an election?

Voting for Committees

When it comes to voting for committees, what do these techniques have to offer? What changes?

Resources

- Spectral analysis of the Supreme Court (with B. Lawson and D. Uminsky), Mathematics Magazine 79 (2006).
- Dead Heat: The 2006 Public Choice Society Election (with S. Brams and M. Hansen), Public Choice 128 (2006).
- Borda meets Pascal (with M. Jameson and G. Minton), Math Horizons 16 (2008).
- Voting, the symmetric group, and representation theory (with Z. Daugherty, A. Eustis, and G. Minton), The American Mathematical Monthly 116 (2009).
- Linear rank tests of uniformity: Understanding inconsistent outcomes and the construction of new tests (with A. Bargagliotti), Journal of Nonparametric Statistics 24 (2012).
- Generalized Condorcet winners (with A. Meyers, J. Townsend, S. Wolff, and A. Wu), Social Choice and Welfare (2013).

Take Home Message

Looking at voting theory from an algebraic perspective is gratifying and illuminating. Doing so gives rise to new techniques, surprising insights, and interesting questions.